

音意を用いた類義語のニュアンスを 推定する手法の提案

長村里穂^{†1} 芋野美紗子^{†2} 土屋誠司^{†3} 渡部広一^{†4}

概要: 日本語においては、辞書で意味を調べても使い方や意味の区別がつかない類義語が多く存在する。日本語を母国語にする人は、様々な類義語を感覚的に使い分けているが、それらを明確に説明することは難しい。人とロボットが会話を行う上では、ロボットも人間と同様に類義語の使い分けやニュアンスを理解していることが望ましい。本研究では、類義語における発音の違いや、音声そのものの意味を表す「音意」という概念を用いて、類義語のニュアンスの推定を行う。

Proposal of Judgement Nuances of Synonyms Using Phonetic-Nuances

RIHO NAGAMURA^{†1} MISAKO IMONO^{†2}
SEIJI TSUCHIYA^{†2} HIROKAZU WATABE^{†2}

Abstract: In Japanese language there are some synonyms which you can't distinguish when you look them up in the dictionary. Japanese people separately use synonyms separately sensibility. However, it is difficult to explain clearly these differences. When people and robots communicate each other, it is better that both of them understand the differences of nuances of synonyms. The purpose of this research is estimating nuances of Japanese synonyms using difference of pronunciation and "Phonetic Nuance" based on Japanese archaic words.

1. はじめに

近年、情報処理技術は発達し、高性能化、多機能化した情報システムを個人が利用する機会が増えており、専門知識を持たないユーザが単純な操作でコンピュータを制御できることが求められる。そのためには、コンピュータが自然言語による命令を認識し、実行できるということが望ましい。例えば、ユーザがコンピュータに言葉を話したり文章を入力するだけで、システムがあらゆる要求に応えられたり、システムからユーザに向けて言葉を話したり文章を提示できるという技術である。正確な自然言語処理が可能となれば、コンピュータがユーザにとって利用しやすいものとなるだけでなく、あらゆる要求に対してより高度な応答ができることが見込まれる。

コンピュータが正確な自然言語処理を実現するには、言葉や文章の意味を認識するだけでなく、それらのニュアンスも認識できることが望ましいと考えられる。ここで、ニュアンスとは、広辞苑によれば「言葉の根底にある直感的な意味」であり、その言葉の表面的な意味とは別に定義できる感覚的な概念である。人は言葉や文章によって意思表示をする際、伝えたいニュアンスに合致する語句を選び、表現していると考えられる。用いる語句を選ぶということ

を日常的に意識している人は少ないと思われるが、例えば、「修正」と「是正」という2つの類義語を比較したとき、どのような場面でどちらの語句を用いるか、あるいは伝えたいニュアンスに近い語句はどちらかということ、人は瞬間的に選択していると考えられる。広辞苑によれば、「修正」とは、「よくないところを直して正しくすること」とあり、「是正」とは、「悪い点を改めただすこと」とある。日本語を母国語とする人は、この2つの類義語の使い分けやニュアンスを感覚的に理解している。一方で、両者の広辞苑の意味文はほぼ同じであり、日本語を母国語としない人やコンピュータがこの2語の使い分けやニュアンスを理解することは難しいと考えられる。そこで、あるアルゴリズムによって類義語の使い分けやニュアンスを表現する技術が実現できれば、自然言語処理の技術の向上に繋がると考えられる。また、それはコンピュータと人間との円滑なコミュニケーションの実現や、日本語の教育の現場など幅広い分野において応用できると考えられる。

そこで本研究では、「音意^[1]」という、日本語の音（読み）に着目した考え方をを用いることで、類義語において国語辞書には載っていないニュアンスを推定する手法を提案する。音（読み）に着目した理由として、言葉1つ1つには音（読み）があり、それによりニュアンスが生まれていると考え

^{†1} 同志社大学大学院理工学研究科
Graduate School of Science and Engineering, Doshisha University
^{†2} 同志社大学 研究開発推進機構
Doshisha University Organization for Research Initiatives and Development

^{†3} ^{†4} 同志社大学 理工学部
Faculty of Science and Engineering, Doshisha University

たからである。また、音意とは、発声や音声そのものの固有の働きや意味を示す考え方である。詳細は2章で述べる。そして、提案手法を再現するために構築するシステムの概要図を図1に示す。また、図1内にあるデータベースや読み仮名の差については4章で詳しく述べる。

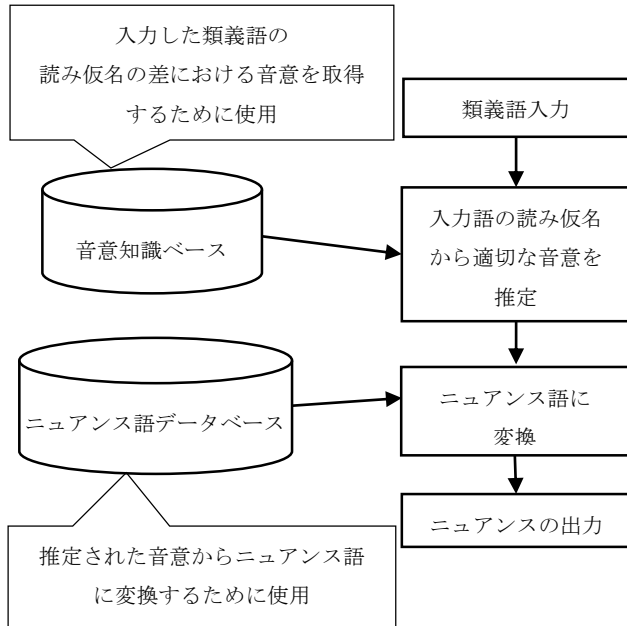


図1 提案手法の概要図

2. 音意

音意¹⁾とは、発音（音声）そのものの固有の働きや意味を示す考え方である。人類の言葉のルーツは猿の鳴き声であり、属全員が理解しあえる共通の鳴き声を持ち喜怒哀楽を表現しているとされる。すなわち、猿の個々の鳴き声には、「固有の働きや意味」があるということを示している。鳴き声にある「固有の働きや意味」を根幹として人類の言葉は進化したと言われており、人類の言葉の音にも「固有の働きや意味」があると考えられる。つまり、言葉を作り出す音（音声）の1つ1つの音節は、「固有の働きや意味」を含む鳴き声から、意味を通ずる用語に組み立てられていったと考えることができる。例えば「山」という言葉は、日本語では「ヤマ」、中国語では「サン」、英語では「マウンテン」、と発音するが、その音には「山」の形態を表す必然的な理由が介在して、言語が成り立っているのではないかと考えられる。日本語の場合は、その山を「岳（タケ）」や「峰（ミネ）」などと呼ぶ言葉遣いもあるが、これらの「山」を言い表したそれぞれの音にも、「山の形態を表現する固有の働きや意味」があると考えられる。そのような音声の意味として「音意」が文献¹⁾で定義されている。

また、音意を定義するに至った背景や、本研究にて用いるにふさわしいとの考えに至った背景として、日本語の古語や古語から現代語の変遷についての考察が文献¹⁾内にある。日本列島における言葉の使われ方を観察すると、それ

ぞれの地域にはその地域だけに共通する言語帯（すなわち、方言）がある。各地域の方言の言葉遣いは、口伝により人から人へ伝わり、且つ他の地域の方言との重なり合いの繰り返しで変動を重ねつつ広がった。つまり、言葉の音声や、その音声を持つ固有の働きや意味（音意）の伝播により方言は変動し、現代に至るまで伝えられてきたと考えられている。歴史的に見て、方言の口伝や変動という現象が少なく、言語変化があまりなかった地域に特定できる場所は、日本列島の端、つまり東北と南九州であると考えられる。加えて、日本列島は海に閉ざされた島国であるため、世界的にも外国の言語が侵入する機会是非常に少なかったと判断できる。よって、この東北・南九州の両地帯は世界の言語史の観点から見ても希少な古語温存地帯であると判断できる。中でも、南九州地方の伝承される南九語の語法を日本最古の歴史本である古事記に照らし合わせると、容易に解読できることが明らかとなっている。そして、南九語の語法は、平安時代の文語体と著しく共通していることも知られている。つまり、その言葉遣いは文語体の源流を現代に伝えているという仮説を立てることができ、現代の日本語群の中でも最も古い形態を留める言語の1つであると考えられる。最も古い形態を留める言語が現代にまで残っていることや、我々が日常的に使っている現代語がこのような歴史的背景を持ち、変遷を辿って成り立っていることを考慮すると、確かに音意は存在し、言葉の一音一音の音節には「固有の働きや意味（音意）」が介在して言葉が成立しているということが考えられるのである。

以上のような歴史的背景から、日本語の起源には「音意」が介在している可能性が極めて高く、文献では各音の音意について論じられている。文献では、日本語の50音に加えて、濁音、半濁音、拗音を合わせた114音についての音意について述べられているが、本研究ではその中の103音を取り扱う。その理由は、音意の定義の中には「を（ヲ、ウオ）」音のように、現代語の名詞、動詞、形容詞などの自立語に出現しない音についての音意も存在するためである。また、「ん（ン）」音については、この文献では音意が定義されていないため、本研究においては取り扱わない。表1に音意の例の一部を示す。

表1 音意の例

音	音意
ア(a)	浮上, 出現, 存在, 自我, 出生, 感動, 天…
イ(i)	著しい, 集中, 一点, 一極, 目立つ, 反発…
ウ(u)	広がる, 多数, 大きい, 多量, 拡大, 多い…
エ(e)	会う, 合う, 接する, 挨拶, 笑う, 絵…
オ(o)	合体, 合着, 強力, 繋ぐ, 結ぶ, 突出…
⋮	⋮

表1に示すように、1音に対して音意は複数定義されている。最も多いもので「ア」音の10語、最小で「ヅ」音等

の2語である。また、文献には、「ア」とは自然界において、ある象体（現象と実体）が出現し、その存在を明らかにすることであり、それは「浮上」、「出現」、「存在」といった意味を示すとある。この音意が適用されている語の例として「有る」、「あつ（感動詞）」等が当てはまる。「有る」とは、象体の存在を確認したときに用いる語であり、「ア」の音意である「浮上」、「出現」、「存在」が当てはまる。「あつ（感動詞）」とは、物事に気づいたときや、物事に対して新しい理解が生まれた際に発する言葉であり、「浮上」、「出現」の音意が当てはまる。このように音意が適用されている語句の例を表2に示す。また、音意と音意が適用されている語句の情報を元に、音意のデータベース（3.3節で後述）を作成する。

表2 音意が適用されている語の例

音	例となる語
ア(a)	有る, あつ (感動詞), 雨, 遊ぶ, 明日
イ(i)	射る, 今, 糸, 行く, 田舎, 家, 石, 衣裳
ウ(u)	上, 歌, 植える, 裏, 嬉しい, 腕
エ(e)	縁, 得る, 笑顔, 選ぶ, 偉い, 絵
オ(o)	緒, 尾, 負う, 豆腐, 降りる, 男, 女, 親

3. 関連技術

3.1 概念ベース

概念ベース^[2]とは、電子化された複数の国語辞書や新聞記事などからある一定のルールに従って自動的に構築された知識ベースであり、約9万語が登録されている。概念ベースにおいて概念Aはm個の属性 a_i と重み w_i (>0) の対によって定義されており、一般的には次のように表される。

$$\text{概念}A = \{(a_1, w_1), (a_2, w_2), \dots, (a_m, w_m)\} \quad (1)$$

属性の数mは概念によって異なり、属性 a_i も概念ベースの中で1つの概念として定義されている。概念ベースの具体的な例を表3に示す。

本研究では、4.1節で述べる国語辞書内の意味キーワードの語群と音意の候補に関連度計算方式^[3]を適用する際に、この概念ベースを用いる。関連度計算方式については次の節で述べる。

表3 概念ベースの例

概念	属性
医者	(医師,0.34)(患者,0.11)(病院,0.08)...
病院	(医院,0.25)(手術,0.18)(施設,0.04)...
治す	(治療,0.43)(医療,0.21)(病気,0.13)...
⋮	⋮

3.2 関連度計算方式

関連度計算方式^{[3][4]}とは、概念ベースに定義されている2つの概念間の関連の強さを定量的に表現する手法である。関連度は0.0から1.0の間の実数値で表され、概念間の関連が強いほど大きな数値となる。例えば概念「自動車」に

対して、「車」、「自転車」、「馬」の関連の強さは表4に示す通りとなり、コンピュータは「自転車」と関連が強いのは3つの内、「車」であるということ判断できる。

本研究で構築するシステムでは、入力語を国語辞書で調べた結果の意味文と、入力語の音意の意味の近さを比較する際にこの技術を導入する。また、次項に関連度計算方式を用いるために必要な一致度、および重み比率付き関連度計算方式について述べる。

表4 関連度計算方式の例

基準概念	対象概念	関連度
自動車	車	0.40
	自転車	0.18
	馬	0.02

4. 提案手法

本研究では、入力された2語の類義語における音（読み）の違いを用いて、入力語の本来の意味と近い音意を選出し、選出された音意に対応するニュアンス語を出力する。ある類義語の組における違いの1つに、音（読み）の違いが挙げられる。よって、語句と語句の音（読み）の違いに音意の考え方を取り入れた本手法により、類義語のニュアンスの違いを明確化できると考えた。提案手法の流れを以下の図2に示す。

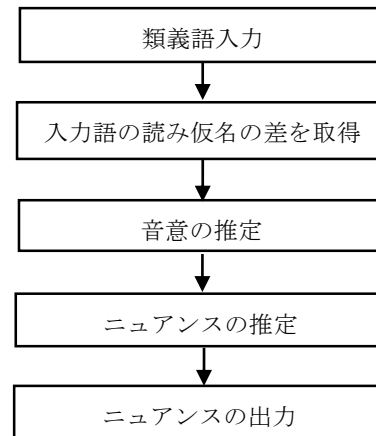


図2 提案手法の流れ

4.1 読み仮名の差の取得

類義語においては、その読み仮名（音）の差にニュアンスの違いがあると考えられるので、読み仮名の差に音意の概念を当てはめることで類義語のニュアンスを推定できると考えた。よって本研究では、入力した類義語における読み仮名の差から音意を推定する過程を構築する。

まず、類義語（名詞、動詞、形容詞、形容動詞）を2語入力し、それらの読み仮名の差をそれぞれ取得する。尚、本システムにおいては入力する類義語の読み仮名は、直接入力するか、漢字入力した類義語を国語辞書^{[5][6][7]}で検索し、辞書内に記載されている読み仮名を参照することで取得できる。また、読み仮名の差とは、2語の入力語の読み仮名において異なる部分の読みのことである。これは、2語の

読み仮名（ひらがな）をそれぞれ比べることで容易に取得できる。2語の間で読み仮名が同じ部分については削除し、読み仮名が異なる部分についての音意のみを選出する。表5に具体例を示す。例として、「自己・自身」、「制圧・鎮圧」、「削除・抹消」の類義語の組を挙げる。「自己・自身」の場合、読み仮名はそれぞれ「じこ・じしん」となるので、これらの共通の読み仮名である「じ」をそれぞれの読みから削除したものを読み仮名の差として抽出する。つまり、「自己・自身」の読み仮名の差は「こ・し」となる。本来であれば、「自己・自身」の読み仮名の差は「こ・しん」となるが、「ん」の音意は文献^[4]内で定義されておらず本研究では取り扱わないので、「自己・自身」の読み仮名の差は「こ・し」となる。「制圧・鎮圧」の類義語の組も同様に読み仮名の差を抽出すると、それぞれ「せい・ち」となる。「削除・抹消」の類義語の組は、共通の読み仮名が1文字も存在しないので、読み仮名の差は類義語そのものの読み仮名と同じとなる。

表5 読み仮名の差を示す例

類義語の組	読み仮名	読み仮名の差
自己 自身	じこ じしん	こ し
制圧 鎮圧	せいあつ ちんあつ	せい ち
削除 抹消	さくじよ まっしょう	さくじよ まっしょう

また、「っ」以外の小文字があるとき、その前のひらがなの大文字と合わせて1文字とし、音意知識ベースの参照する。尚、読み仮名について、「っ」と「ん」に関しては音意が定義されていないため、この過程においてこれらの読み仮名を検出した際はこれらの読みを削除する。また、「りゃ」音、「りゅ」音、「りょ」音についても音意として定義されていないが、文献^[4]によれば、このような2音で1音とする音は、その2音それぞれの音意が重なったものとして定義されている。例えば、「きゃ」という音意は、「生（き）の象体を対外的に行使する」と定義されている。「生（き）の象体」とは「き」の音意そのもののことであり、「対外的に行使する」とは「や」の音意として定義されている。よって、「ゃ」等が含まれる拗音は、それぞれの音を直音としてみなして音意を取得しても良いと考えられる。これを踏まえて、本研究において「りゃ」、「りゅ」、「りょ」音については、それぞれ「り」と「や」、「り」と「ゆ」、「り」と「よ」の音意を取得する。

4.2 音意の推定

4.1節で述べたように読み仮名の差を取得した後、読み仮名の差に対する音意を全て取得する。入力例として「従属」と「隷属」を挙げると、従属の「じゅ」と「う」からそれぞれの音意を取得し、隷属の「れ」から「い」の音意を取

得する。その際に音意知識ベース（4.2.1項で後述）を参照する。

4.2.1 音意知識ベース

音意知識ベースとは、音（かな）と音（カナ）とその音意を格納した知識ベースである。2節や表1に示したような、日本語の50音に加えて濁音、半濁音、拗音を合わせた103音それぞれの音意をデータベースに格納し、音意知識ベースとする。2章で述べたように、音意は1つの音（音声）に対して複数定義されているので、これらを各音別に整理し、データベース化した。音意知識ベースの一部を以下の表6に示す。

表6 音意知識ベースの一部

音（かな）	音（カナ）	音意
あ	ア	浮上
あ	ア	出現
⋮	⋮	⋮
か	カ	構う
⋮	⋮	⋮
しゃ	シャ	表面
⋮	⋮	⋮

4.2.2 音意知識ベースと国語辞書による音意の推定

音意は1つの音に対して複数の語が定義されているので、どの音意が適切かを選出する必要がある。そこで、入力語を国語辞書で検索し、入力語2語に対する意味キーワードをそれぞれ取得する。意味キーワードとは、国語辞書内にある語句に対する意味文を形態素解析し、名詞、動詞、形容詞等の自立語のみを抽出した語である。取得した意味キーワードの語群と、前述した順序に従って取得した音意について、総当りで関連度計算を行う。計算の結果、国語辞書内の意味キーワードの語群との関連度が最も高い音意を1語選出する。尚、本研究で用いる国語辞書は、約9万9000語の意味や説明文が格納された広辞苑、スーパー大宇林、岩波国語辞書のデータを用いて作成されている。

入力語において読み仮名の差に対する音意を全て取得したあと、適切な音意を1つに絞る。そのために、国語辞書内のキーワード語群と、取得した音意の候補との関連度計算を行う。入力語が国語辞書内の見出し語として複数定義されている場合は、その全ての見出し語に対する意味キーワード語群を取得し、それらと音意の候補全てと総当りで関連度計算を行う。そして、国語辞書内のキーワード語群との関連度が最も高い音意を1語選出する。この操作により、入力した類義語本来の意味と、意味的に最も距離が近い音意が選出される。

4.3 ニュアンスの推定

4.2.2項で述べた手法により選出された音意を、ニュアンス語に変換する。1つの音意には1語～3語のニュアンス語が定義されており、選出された音意に対応したニュアンス

ス語を参照し、ニュアンスとして出力する。

4.3.1 ニュアンス語データベース

ニュアンスとは、広辞苑によれば「言葉の根底にある直感的な意味」であり、その言葉の表面的な意味とは別に定義できる感覚的な概念である。音意として定義されている語句において、直感や感覚を示すものは少なく、言葉のニュアンスを示す語句として不適切である場合が多いと考えられる。例えば、「あ」の音意の1つに「浮上」があるが、「浮上」がある言葉のニュアンスを表す語句としては考えにくく、感覚的な概念としても捉えにくい。そこで、本研究では音意からニュアンスを示す語に変換する工程を構築する。そのために、「ニュアンス語」というものを新たに定義する。

ニュアンス語とは、本研究では言葉のニュアンスを示す語としてふさわしい感覚・知覚語¹⁴や状態を表す動詞として定義する。感覚・知覚語とは、人の五感に基づく「硬い」、「柔らかい」等の感覚や知覚を、形容詞や形容動詞で表現した語のことであり、言葉の感覚的な概念であるニュアンスを表現する語として適していると考えられる。また、状態を表す動詞や「〇〇的」という表現も、言葉のニュアンスを示す語として適切ではないかと考えられる。例として、「目立っている（目立った）」や、「攻撃的な」という表現があるが、これらは感覚・知覚語ではないが言葉のニュアンスを表す語として想定される語である。一般に、言語におけるニュアンスを表現する方法として、感覚、知覚に基づく表現が多いと考えられるので、本研究では感覚・知覚語¹⁴と状態を表す動詞「〇〇的」という表現をニュアンス語として扱う。表7にニュアンス語データベースの一部を示す。

表7 ニュアンス語データベースの一部

音	音意	ニュアンス語
あ	出現	目立った
あ	出現	明るい
⋮	⋮	⋮
か	構う	嬉しさ
⋮	⋮	⋮
しゃ	表面	明るい
⋮	⋮	⋮

4.3.2 音意に夜ニュアンスの推定

4.2節で述べた方法により、入力した類義語の読みにおける音意を推定したあと、4.1.4項で述べたニュアンス語データベースを用いて、音意からニュアンス語へ変換する。これにより、入力した2語の類義語における言葉の根底にある直感的な意味であるニュアンスを音（読み仮名）の差の音意により判別した最終出力が得られる。類義語「従属」と「隷属」について、表8に音意からニュアンス語への変換の例を示す。入力語「従属」の意味キーワード語群と、

「従属」と「隷属」の読み仮名の差である「じゅ」と「う」における全ての音意との関連度計算を行った結果、意味キーワード語群と最も関連度の高い音意は「大きい」である。そして、ニュアンス語データベースより、「従属」に対応するニュアンス語は「大きい」と「多い」となる。

表8 音意からニュアンス語への変換

入力した類義語	類義語の本来の意味と最も関連度の高い音意	音意に対応するニュアンス語
従属	大きい	大きい, 多い
隷属	威	威圧的な, 強い

5. 評価

本研究における、音意を用いた類義語のニュアンスの推定について評価を行った。提案手法に基づいたシステムの出力結果を被験者3人で評価した。

5.1 評価方法

本研究の評価方法を述べる。システムに入力された類義語に対して、選出されたニュアンス語がそれらの類義語のニュアンスを表現できているかを評価する。被験者3人に、入力した類義語自体の意味も考慮しつつシステムの出力結果を目視にて正解、どちらともいえない、不正解、わからない、の4段階で評価してもらった。尚、出力されたニュアンス語のうち、被験者が正解だと判断したニュアンス語の個数が、不正解と判断したニュアンス語よりも多い場合は○、少ない場合は×、同じ個数の場合は△と評価してもらった。また、「わからない」という評価は、被験者が類義語そのものの意味がわからない場合や、出力された音意やニュアンスが正しいかどうか判断しかねると被験者が思った場合に選択してもらった。また、評価に用いる類義語は、類義語辞書内に登録された類義語から無作為に200語(100セット)選出し、それらの類義語を本システムに入力して評価を行う。

また、200語全体の評価と共に、類義語同士で音が異なる個数によって類義語を分類した場合の評価も行う。分類方法は、1音違い、2音違い、3音(以上)違いの3種類である。表9に例を示す。類義語辞書に登録された類義語から無作為に200語(100セット)選出し、それら全ての評価を行った後、表9に示したように類義語を分類し、その分類ごとに評価結果を集計する。無作為に選出した200語(100セット)の類義語のうち、1音違いの類義語は40語(20セット)、2音違いの類義語は120語(60セット)、3音以上違う類義語は40語(20セット)存在した。

表9 類義語の分類の例

異なる音の個数	語例
1音	自己・自身, 苛虐・呵責…
2音	従属・隷属, 修正・是正…
3音(以上)	削除・抹消, 熟考・深慮…

5.2 評価結果

本手法の評価結果を以下に示す。5.1 節で述べた評価方法により評価を行った。全体の評価は以下の図のようになった。また、以下の図における「○」は「正解」、「△」は「どちらともいえない」、「×」は「不正解」、「不明」は「わからない」を示す。全体評価の結果を以下の表 10 に示す。また、5.1 節で述べたように、類義語どうしの音の違いの個数による分類での評価結果も表 10 に示す。尚、表 10 における被験者 3 人の評価結果は、正解、どちらともいえない、不正解、不明の各評価の個数を合計し、母数で割った割合を算出している。

また、実際の評価例を表 11 に示す。例えば、「従属」という言葉のニュアンス語は音意と同じく「大きい」と出力される。この結果は、「隷属」よりも大きい意味合いで使われると捉えることができる。実際に、「従属」とは「強力なもの、中心となるものにつき従うこと」であり、「隷属」とは「強力なもの、中心となるものにつき従って言いなりになること」であるので、この場合の音意によるニュアンス語の出力結果は被験者 3 人ともに正解と評価された。失敗例としては、「搜索」と「探索」の類義語が挙げられる。本手法により「搜索」のニュアンス語は「近い」、「探索」のニュアンス語は「最高の、最高な」が得られたが、これはどの被験者からも不正解の評価となった。

表 10 各評価結果

評価	割合 (%)			
	全体評価	1 音違い	2 音違い	3 音(以上)違い
○	39.5	25.8	45.0	39.2
△	19.7	10.8	21.1	21.7
×	35.5	56.7	28.6	37.5
不明	4.83	6.7	5.3	1.7

表 11 実際の評価例

類義語の組	音意	ニュアンス語	評価
従属	大きい	大きい, 多い	○
隷属	威	威圧的な, 強い	○
搜索	近い	近い	×
探索	最高	最高の, 最高な	×
考え	構う	嬉しさ	×
思慮	増大	増える	×
会談	議論	強い	○
会議	静止	静かな	○

6. 考察

本手法により 2 つの類義語におけるニュアンスの違いを推定した結果についての様々な観点からの考察を述べる。

6.1 音の違いの個数による評価について

評価する類義語の組み合わせで、類義語を 1 音違い、2 音違い、3 音 (以上) 違いに分類して評価を行った。その結果、1 音違いの類義語における精度が最も低く、2 音違いに

おける精度が最も高くなった。

1 音違いの類義語における精度が低くなった原因として、候補となる音意が減少することによって、正解が得られにくくなったことが考えられる。本手法では、1 音違いの類義語はその 1 音の音意のみを選出したため、音によっては得られる音意が 2 語だけの場合があった。あ行の音意は 7~10 語定義されているが、それ以外の音の音意は 2~5 語程度なので、異なる 1 音が「あ行」以外の音の場合は正解となる音意の候補を選出できない可能性が高まる。それに伴い、最終的なニュアンス語も正しく出力されない可能性も高まったと考えられる。

また、1 音違いの類義語よりも 2 音違い及び 3 音 (以上) 違いの類義語の方が高い精度が得られたことにより、類義語同士の音の違いが多い方が適切なニュアンスを選出できると考えられる。また、類義語同士で全ての音 (読み) が異なる場合でも、28 語中 13 語 (約 46%) について正解が得られた。全ての音 (読み) が異なる場合でも、ある程度の精度が得られていることから、音 (読み) そのものにニュアンスの情報が含まれていると考えられる。今後は、類義語に限らず知識ベースに存在しない未知の語句についても、読み仮名の情報があれば意味やニュアンスを音意によって推定できる可能性を示せたと考えられる。

6.2 音の違いの個数による評価について

評価する類義語の組み合わせで、類義語を 1 音違い、2 音違い、3 音 (以上) 違いに分類して評価を行った。その結果、1 音違いの類義語における精度が最も低く、2 音違いにおける精度が最も高くなった。

1 音違いの類義語における精度が低くなった原因として、候補となる音意が減少することによって、正解が得られにくくなったことが考えられる。本手法では、1 音違いの類義語はその 1 音の音意のみを選出したため、音によっては得られる音意が 2 語だけの場合があった。あ行の音意は 7~10 語定義されているが、それ以外の音の音意は 2~5 語程度なので、異なる 1 音が「あ行」以外の音の場合は正解となる音意の候補を選出できない可能性が高まる。それに伴い、最終的なニュアンス語も正しく出力されない可能性も高まったと考えられる。

また、1 音違いの類義語よりも 2 音違い及び 3 音 (以上) 違いの類義語の方が高い精度が得られたことにより、類義語同士の音の違いが多い方が適切なニュアンスを選出できると考えられる。また、類義語同士で全ての音 (読み) が異なる場合でも、28 語中 13 語 (約 46%) について正解が得られた。全ての音 (読み) が異なる場合でも、ある程度の精度が得られていることから、音 (読み) そのものにニュアンスの情報が含まれていると考えられる。今後は、類義語に限らず知識ベースに存在しない未知の語句についても、読み仮名の情報があれば意味やニュアンスを音意によって推定できる可能性を示せたと考えられる。

6.3 負の意味を持つ語について

負の意味とは、否定的な意味や、マイナスイメージのある意味を指す。例えば、「不法」や「違法」といった語句が挙げられる。負の意味を持つ語句に対する類義語も存在するが、本手法における負の意味を持つ類義語に対する評価は正解率が20%にとどまった。特に、「不法」と「違法」の類義語の組では、全ての被験者がどちらともいえない、不正解と評価した。この原因は、音意の定義に問題があるからであると考えられる。まず、「い」音や「ふ」音には負の意味を示す音意が定義されていない。「い」音に関しては負の意味に近い「異なる」という音意が定義されているが、「違法」の国語辞書内の意味キーワードの語群とは関連度が低かったため、音意として選出されなかった。「異なる」という言葉（音意）は「違う」、「違い」といった別の言葉に言い換えることができるので、似た表現の言葉も音意として登録、定義することで「違法」といった「違」を伴った負の意味を示す語句については本手法による精度向上が期待できると考えられる。

7. おわりに

本研究では、音意という概念を用いて、類義語のニュアンスを区別、推定する手法を提案し、これを再現したシステムを構築することにより評価を行った。その結果、本研究で提案した手法では、無作為に選出した類義語200語（100組）の約40%の類義語に対してニュアンスを区別することができた。また、類義語どうしの読み仮名の差の個数が多い方が、高い精度が得られたことから、類義語に限らず知識ベースに存在しない未知の語句についても、読み仮名の情報があれば意味やニュアンスを音意によって推定できる可能性を示せたと考える。本手法では、ニュアンス語の選択方法や国語辞書の構成方法などに問題があったと考えられるため、今後はこれらの精錬を行う必要があると言える。

謝辞 本研究の一部は、科学研究費補助金（若手研究（B）24700215）の補助を受けて行った。

参考文献

- [1] 飯野布志夫 “言葉の起こり” (鳥影社 2013).
- [2] 奥村紀之, 土屋誠司, 渡部広一, 河岡司, 「概念間の関連度計算のための大規模概念ベースの構築」, 自然言語処理, Vol.14, No.5, pp.41-64, 2007.
- [3] 渡部広一, 奥村紀之, 河岡司, “概念の意味属性と共起情報を用いた関連度系計算方式”, 言語処理学会, 自然言語処理 Vol.13(2006), No.1 P53-74.
- [4] 奥村紀之, 荒木孝允, 渡部広一, 河岡司, “概念属性の動的評価に基づく概念関連度計算方式”, 情報処理学会, E-033, pp.223-226, 2006.
- [5] 松村明, 大辞林第二版, 株式会社三省堂, 1995.
- [6] 新村出(編), “広辞苑”, 岩波書店, 1992
- [7] 西尾実, 岩淵悦太郎, 水谷静夫, “岩波国語辞典第五版”, 岩波書店, 1994.
- [8] 堀口敦史, 渡部広一, 河岡司, 「常識的感覚判断システムの構築」, 情報処理学会研究報告, 130-6, pp.31-36, 2002.