

シソーラスの自動構築を目指した
名詞概念の顕現性計算機構に関する研究

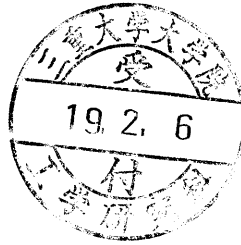
平成 18 年 度

三重大学大学院工学研究科
博士前期課程 情報工学専攻

森 田 悠 介

修士論文

シソーラスの自動構築を目指した 名詞概念の顕現性計算機構に関する研究



平成 18 年度修了

三重大学大学院 工学研究科

博士前期課程 情報工学専攻

森田 悠介

要旨

柔軟な言語処理を実現するためには、比喩の処理が不可欠である。比喩的關係を計算機処理するためには、概念とその特徴（属性値）の關係を記述した知識ベースの整備が重要である。

比喩の計算処理を目的とした知識ベースに関する先行研究では、新聞記事コーパス中の形容詞・形容動詞-名詞間の連体修飾關係に基づく共起知識に基づいて知識ベースを自動構築する手法や、World Wide Web (WWW) を利用した適合フィードバックによって、知識ベースを洗練する手法などが提案されている。しかしながら、知識の網羅性に関する問題や、概念の典型性を示す属性値（顕現属性値）の認識において、人間による判断と自動処理による判断に差異がみられるという問題が指摘されている。

そこで本論文では、知識ベースの網羅性を確保するために WWW から直接知識を抽出する手法を提案する。また、人間判断と自動処理判断との差異を解消するために、人間判断による知識ベースの分析を行い、顕現属性値の多段識別手法を提案する。

WWW からの知識の直接抽出手法では、「～な+名詞」「～い+名詞」という形容詞・形容動詞の連体修飾における活用形の規則性を利用することで WWW から大量の知識を抽出することに成功した。新聞記事を知識源とした知識ベースとの比較評価によって、網羅性において優位性が確保できることが明らかになった。

人間判断との差異解消においては、心理学実験によって収集した人間判断に基づく知識ベースを詳細に分析し、人間が顕現属性値を複数レベルで認識していることがわかった。この知見に基づき、形容詞・形容動詞の連体修飾における諸機能を利用して、顕現属性値を多段階に識別する手法を提案した。人間判断による顕現性の識別結果と提案手法による識別結果を比較評価した結果、提案手法が、人間の判断に近い識別性能を有することが明らかとなった。

目次

1. 序論.....	1
2. 従来知識ベース構築手法と問題点.....	4
2. 1. テキストコーパスからの知識の抽出手法.....	4
2. 2. 知識ベースの精緻化手法.....	4
2. 3. 従来手法の問題点.....	5
3. WWWからの概念属性値の獲得と洗練.....	6
3. 1. 提案手法.....	6
3. 2. 実験と評価.....	12
3. 3. 考察.....	14
4. 顕現属性値の多段階識別モデルの基礎的検討.....	16
4. 1. 予備調査.....	16
4. 2. 顕現属性値の多段階識別モデル.....	18
4. 3. 分析と考察.....	20
5. 結論.....	24

謝辞

参考文献

1. 序論

1.1. 背景

我々が日常で使用する言語には、必ずといってよいほど比喩が含まれている。比喩とは、ある概念を異なる概念を用いて説明、もしくは強調する修辭的手法の一つであり、心理学や認知言語学など、様々な分野で研究対象として重要視されている[1]~[3]。それは、自然言語処理の分野においても例外ではない。柔軟な言語処理を実現するためには、比喩表現の理解は必要不可欠である。例えば、対話処理では、比喩理解によって発話数の削減や、より柔軟な対話制御が期待できる[4]。また、質問応答では、曖昧な質問文に対して、効果的に的確な回答をする方法として、比喩指標「のような」を用いた回答文を作成する手法が考えられる[5]。

次に、上記の一例を示す。

「ダルビッシュ有とはどんな投手ですか？」という曖昧な質問に対し、「松坂大輔のような投手です」と回答することで、「松坂大輔」の持つ特徴（属性値）である「すばらしい」「優秀な」によって「ダルビッシュ有は優秀ですばらしい投手だ」、といった比喩解釈がなされる。このように、比喩を利用することで、質問文に対して的確な回答を返すと共に、発話数を削減することが可能となる。

上記のような比喩解釈を、属性値に基づいた比喩（属性比喩）解釈という。属性比喩解釈を実現するにあたり、概念と属性値の関係を正確に把握し、属性値に基づく属性比喩の解釈過程を理解することが必要となる。属性比喩の解釈過程は、比喩表現において、例えられる概念（target 概念）の持つ特徴（属性値）が、例える概念（source 概念）が持つ典型的な特徴（顕現属性値）によってクローズアップされることによる、と説明できる[7][8][10][11]。例えば、「リンゴ」という概念から「リンゴのような頬」という属性比喩を考えた時、その意味は「赤い頬」や「丸い頬」を指す。これは、source 概念である「リンゴ」が持つ「赤い」や「丸い」といった顕現属性値が target 概念である「頬」の特徴としてクローズアップされることによって連想されるのである。ここで、概念を顕現属性値の集合で表現した知識ベース（知識ベース）の構築がなされていれば、属性比喩の解釈を行う上でクローズアップされるべき属性値の判定が容易になる。したがって、知識ベースの構築は、計算機上での属性比喩解釈に寄与するものと考えられる。

1.2. 研究の目的

本論文では属性比喩解釈を目的とした知識ベース構築について述べるものとする。

従来、知識ベースの構築は、被験者を用いた心理学実験により構築する手法が主流であった[6][7]。この手法は精度の高い知識が得られる一方、汎用的な大規模知識ベースの構築

を目指した場合、多数の被験者数の確保や膨大な時間を必要とするため、コスト面において大きな制限がある。

この問題に対し、梶井らは、新聞記事コーパスから抽出した形容詞・形容動詞一名詞間の連体修飾関係を用いて概念を表現する属性値集合を確率的に表現するモデルを提案している[8]。さらに、獲得してきた概念と属性値から、比喩的修飾を示す比較表現を生成し、World Wide Web (WWW) を知識源として利用することで、比較表現の生成能力を検証する。これにより、各属性値の顕現属性値としての適合性を判定している。そして、WWW から取得した属性値の頻度情報に基づき、判定結果に応じて属性値集合を再構成し、フィードバックを行うことで、知識ベースを自動的に洗練している。このようにして、梶井らは低コストでの知識ベースの構築に成功している[9]。しかしその一方で、彼らの手法では、次のような問題があった。

梶井らの手法では、概念や属性値といった基本的な知識の獲得を新聞記事に依存している。新聞記事は社会の出来事について事実や解説を広く伝えるための刊行物であり、知識コーパスとしては知識に偏りがあることが考えられる。従って、人間が一般的に利用する知識と比較すると、獲得される知識が少なく、知識の網羅性が不足しているという問題があった。

また、知識ベースの精緻化をするにあたって、自動処理によって判定された顕現属性値が、顕現属性値として妥当であるにもかかわらず、心理学実験に基づく顕現属性値と比較すると、属性値の顕現性の判定基準に差異がみられるという問題が指摘されている。

本研究では、比喩解釈を目指す上で理想的な、人間を対象として構築した知識ベースに近い、精度の高い知識ベースを自動構築すべく、上記で述べた問題点の解決手法について提案する。

知識の網羅性に関する問題に対しては、WWW を知識源として利用し、WWW から直接知識を抽出する手法を提案する。この場合、WWW 中の文章は文体の自由度が高いため、高い精度で知識抽出を行うには、形態素解析や係り受け解析の解析誤りの影響を回避する手段が必要となる。しかし、WWW のテキスト情報は日々増加する傾向にあり、膨大な情報を有している。それらすべての情報に対し解析誤りの回避を目的とした処理を施しているには膨大な計算量が必要となる [9]。そこで、提案手法では、snippet (検索エンジンを用いて検索した結果得られるテキスト情報) を利用することで、WWW から必要とする表現 (連体修飾関係) を含むテキスト情報を確保する。それによって、上記に示すような処理の計算量を抑えつつ、高い網羅性を持つ知識ベースを構築する手法を提案する。具体的には定型表現を含むテキスト情報を文字列照合により WWW から収集する。このように、WWW の利用を、定型表現の文字列照合のみに限定することで、WWW は新聞記事に勝る網羅性を持つ知識源として利用することが可能となる[9][19]。次に、収集されたテキスト情報に対し形態素解析、並びにその誤り解析といった最低限の処理を施すことで連体修飾関係を抽出し、その連体修飾関係に尤度付けすることで最も基本となる知識ベース (基本知

識ベース)を自動構築する。さらに、基本知識ベースに対し梶井らの手法と同様に、WWWを用いた精緻化を行うことで網羅性の高い知識ベースを構築する。

また、人間の判断する顕現属性値と自動処理により判定された顕現属性値との差異を解消するために、人間判断による知識ベースの分析を行い、顕現属性値を自動的に多段階に識別する手法を提案する。人間が判断する顕現属性値と自動的に判定する顕現属性値の間に差異が生じる要因として次のようなものが考えられる。人間は典型性の強さに応じて顕現属性値にレベルを付与して区別しており、心理学実験においては、より典型性の強い顕現属性値を選択的に提示しているため、自動判定と人間判断の間に差異が生じる可能性が考えられる。そこで、本論文では、顕現属性値に対してレベル識別が可能であると仮定する。次に、与えられた顕現属性値集合に対して、顕現属性値の典型性判定に関して観点の異なる表現の生成可能性を考慮することによって顕現属性値を多段階に識別する手法を提案する。

1.3. 論文の構成

本節では、本論文の構成について述べる。

2章では従来の知識ベース構築手法について概観する。まず、形容詞・形容動詞一名詞間の連体修飾関係を用いた概念－属性値情報の抽出手法について述べる。次に、文章コーパスから概念－属性値情報を抽出してすることで構成された基本知識ベースへの、WWWを用いた精緻化について説明する。最後に、従来手法の問題点について再記する。

3章ではWWWを知識源とした知識ベースの構築手法について説明し、提案手法の妥当性を確認するための先行研究との比較実験と評価・考察を行った。1節で提案手法について述べ、2節で提案手法の妥当性を確認するための従来手法との比較実験と評価を行う。3節ではそれに対して考察する。

4章では、顕現属性値の多段階識別モデルに関して基礎的検討を行う。1節では、予備調査として、人間判断に基づく顕現属性値を収集し、それらが複数の集合に区別できるかどうかを検証し、人間の顕現属性値を多段階識別可能性について考察する。2節では、WWWから獲得してきた形容詞・形容動詞(属性値)－名詞(概念)の共起関係と表現パターンを用いて比喩的修飾機能および述定機能を持つ表現を生成する。そして、WWW上における各表現の出現分布を調べることで共起関係の適合性判定を行い、判定結果の組み合わせに基づいて共起関係にレベルを付与することを試みる。3節では、人間判断に基づく顕現属性値の多段階識別の結果と、WWWを利用した適合性判定に基づく顕現属性値の多段階識別の結果を比較し、顕現属性値の多段階識別自動化の可能性について吟味する。

最後に5章で、本研究の結論について述べる。

2. 従来の知識ベース構築手法と問題点

本章では、従来の知識ベース構築手法の概要とその問題点について述べる。以下、1節でテキストコーパスからの知識の抽出手法について述べ、2節では比喩解釈を目的とした知識の精緻化手法について略述する。さらに3節では従来手法の問題点を挙げる。

2.1. テキストコーパスからの知識の抽出手法

本節では、知識源となるテキストコーパスから大規模な知識を自動抽出する従来手法について述べる[8]。

対象となるテキストコーパスを形態素解析し、それにより得られた結果から形容詞・形容動詞（属性値）－名詞（概念）の共起関係と出現頻度を抽出する。抽出された情報の知識ベース化を図る。

例えば、対象となるテキストコーパス中に“赤い花”という表現が存在した場合、そのテキストを形態素解析することで、“花（概念）－赤い（属性値）”という共起関係が抽出され、この結果に出現頻度を付加することで“花＝{赤い#1}”のような知識ベースが得られる。さらに、異なる文章から“花－赤い”“花－白い”、“沼－小さい”といった共起関係が抽出された場合、知識ベースは、“花＝{赤い#2, 白い#1}, 沼＝{小さい#1}”に更新される。

2.2. 知識ベースの精緻化手法

本節では、前節で構築した基本知識ベースに対し、比喩解釈を目的としたWWWを利用した属性値の適合性判定及び判定結果へのフィードバック手法[9]について略述する。

属性比喩の解釈過程は、1章でも述べたとおり、target 概念の持つ属性値が、source 概念が持つ顕現属性値によってクローズアップされることによる、と説明できる。しかし、概念間でクローズアップされるべき属性値が正しく選択されなければ、正しい比喩解釈がなされない。従って、知識ベースを構築する上で、属性値の顕現性を正しく判定することは重要な意味を持つ。

適合性判定では属性値が、対象となる概念の顕現属性値であるかどうかを判定する。概念と顕現属性値を用いれば、「リンゴのような赤い」や「蜂のように忙しい」といった、例示や比喩を意味する比較表現を生成可能であると考えられる。このことから、概念と属性値を用いて比較表現を生成し、その比較表現が一般的に利用されているかどうかを調べることで、属性値の顕現性の有無を判定することができる。比較表現の構成時に用いた「～のような」「～のように」といった定型パターンを指標表現（ground）と呼ぶ。

次に、判定結果に応じて、WWW から得られる頻度情報に基づいて属性値集合を再構成

し、フィードバックを行う。フィードバックは、テキストコーパスから獲得してきた属性値集合（基本知識ベース）の、頻度情報に基づく属性値ランキングに歪みが生じている場合に、それを補正するための処理である。適合性判定にて属性値が不適合と判定された場合、基本知識ベースの、属性値ランキングに歪みが生じている可能性が考えられる。ここで、適合性判定の結果を反映することで、対象属性値のランキングを補正すれば、属性比喩解釈における、概念対の属性値のクローズアップも補正することにつながるため、比喩解釈の性能向上に貢献することが可能となる。

適合性判定により、不適合と判定された属性値は、対象概念の顕現属性値ではないと判断できる。したがって、該当属性値の尤度を下げ、属性値集合の再構成を行うことで、対象概念の属性値ランキングを補正することができる。

2.3. 従来手法の問題点

2.1での知識の自動抽出手法では、基本的な知識獲得を新聞記事コーパスに依存している。1章でも述べたとおり、新聞記事は社会の出来事について事実や解説を広く伝えるための刊行物であり、著者は新聞記者であり、知識コーパスとしては知識に偏りがあることが考えられる。例えば、10年分の新聞記事(注1)を知識源としてその中から2.1節の手法を用いて知識を抽出した場合「レモン」という概念の顕現属性値であると考えられる「黄色い」や「酸っぱい」といった属性値が獲得されない。つまり、新聞記事を知識源として利用した場合、知識ベースを構築する上で、必要な知識が獲得できない。知識の網羅性が不足しているといった問題が挙げられる。

上記の問題を解決するためには、より膨大でかつ分野依存性の低い文章集合であるWWWを知識源として利用し、WWWから直接知識を抽出する手法を提案する必要がある。WWWから直接知識を抽出する手法については3章にて詳述する。

また、上記とは異なる問題として、概念の顕現属性値の認識において、人間による顕現属性値の判断結果と自動処理による顕現属性値の判定結果に、その判定基準において差異がみられるという問題が指摘されている。このような問題が生じる原因として、人間は顕現属性値を典型性の強さに応じて多段階に識別しており、心理学実験においては、より典型的な顕現属性値を選択的に提示しているため、システムの判定する顕現属性値と人間が判断する顕現属性値の間に差異が生じる可能性が考えられる。例えば、「風」という概念においては、従来手法を用いることで、システムが判定した顕現属性値は、人間が顕現属性値として判断している「冷たい」「強い」といった属性値はすべて網羅できている。しかし、システムが顕現属性値であると判定した属性値数が多すぎるため、結果的にシステムの判定精度の低下が報告されている。

(注1)：毎日新聞CD-ROM版1994～2003年、約111万記事（約1,046万文）を用いた。

この問題を解決するにあたって、人間を対象として顕現属性値を獲得してくることで、実際に人間が顕現属性を多段階に識別しているのかの調査を行った。その結果、人間が顕現属性値を複数レベルで多段階に認識していることがわかった。この知見に基づき、形容詞・形容動詞の連体修飾における諸機能を利用して、顕現属性値を多段階に識別する手法を提案した。この手法についての詳細は4章で述べる。

3. WWWからの概念属性値獲得と洗練

本章では、知識ベースの網羅性を確保するために、WWWから直接知識を抽出する手法について述べる。1節で提案手法について詳述し、2節で提案手法の妥当性を確認するために先行研究との比較実験と評価を行い、3節ではそれに対して考察を行う。

3.1. 提案手法

本節では、提案手法について詳述する。本手法は、従来手法と同様に基本知識ベースの構築部と知識ベースの精緻化部といった2つの処理部から構成される。以下、各処理部について説明する。

3.1.1. 基本知識ベースの構築

本節では、基本知識ベースの構築部について詳述する。知識ベース構築部は(1)フレーズ生成・検索、(2)snippet取得、(3)形態素解析、(4)概念-属性値情報取得、(5)頻度情報取得、(6)属性値ランキング、からなる(図1)。

以下、各処理過程について順を追って説明する。

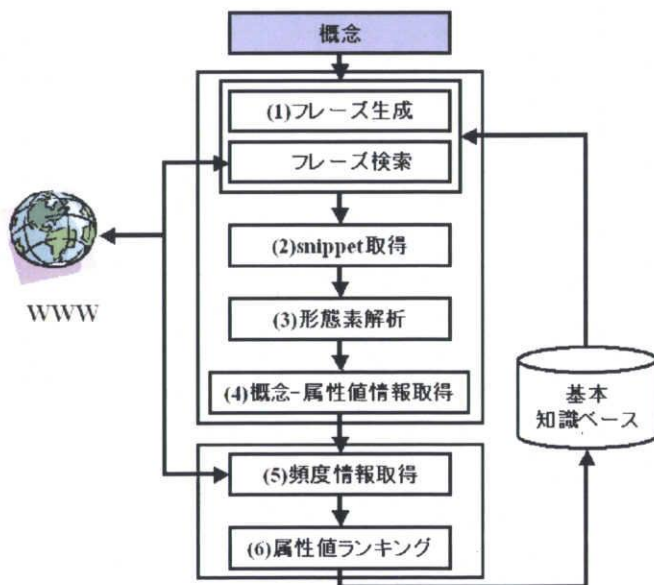


図1 基本知識ベースの構築手法概要

(1) フレーズ生成・検索

形容詞・形容動詞と名詞の連体修飾関係では、多くの場合、「～な＋名詞」、「～い＋名詞」というパターンで出現する。そこで、入力クエリ（名詞）に対して「な＋クエリ」「い＋クエリ」という定型表現を生成し、これを WWW 上で検索する(注2)。例えば、クエリとして「リンゴ」が与えられた場合、「いリンゴ」「なリンゴ」という定型表現を生成し、WWW 検索を行う。

(2) snippet 取得

検索した結果を走査し、検索したフレーズを含む snippet を収集する。

(3) 形態素解析

収集した snippet を形態素解析し、形容詞・形容動詞（以下、連体修飾語句）と名詞の連接を連体修飾関係として抽出する。「リンゴ」の例では、「赤いリンゴ」や「大きなリンゴ」などの連接が取り出せる。

(4) 概念－属性値情報取得

(3)で得られた連体修飾関係による語の共起を、概念－属性の関係として抽出する。神崎らは、連体という用法の観点から、連体修飾関係の解析・抽出を人手で行っている[12][13]。これは、連体修飾語句が必ずしも概念の特徴を表すものではなく、顕現属性か否かの判定を自動で行うことは難しいためである。

提案手法では、後述する適合性判定およびフィードバックによってこの問題を解決する。

(5) 頻度情報取得

(4)で抽出した「属性値＋対象概念」を用いて、再び連体修飾関係「連体修飾語句＋名詞」を生成する。生成した連体修飾関係を再度検索エンジンで検索することにより得られた検索件数を頻度情報として、各属性値に付加する。その結果構築された属性値集合から、頻度情報が閾値以下となった属性値を削除することで、形態素解析誤りなどで抽出してきた連体修飾語句を削除することができる。

(6) 属性値ランキング

最後に、その頻度情報を各属性値の尤度とし、尤度に基づき概念ごとに属性値集合のランキングを行う。その結果得られたものを基本知識ベースとする。

(注2)：検索には、検索エンジン、yahoo! Japan, <http://www.yahoo.co.jp> を用いた。

3.1.2. 基本知識ベースの精緻化

基本知識ベースの精緻化部は、前章で略述した適合性判定並びにフィードバックを用いた知識ベース構築手法[9]を拡張することで実現する。基本的には、3.1.1で構築した基本知識ベースに対し、「リンゴのように赤い」「蜂のように忙しい」といった比喻や例示を意味する比較表現を生成し、これらの表現が一般的に利用されているか否かを、WWWを用いて調べることによって、連体修飾としての妥当性を自動的に検証・補正するという考え方に基づいている。

以下、基本知識ベース精緻化の流れ(図2)を説明する。

適合性判定の処理は、梶井らの手法と同様、2.1で構築した知識ベースからの(1)要素の取得、(2)比較表現生成、(3)表現検索、(4)適合性判定、の4つの処理で構成され、判定結果に対するフィードバックは、(2)比較表現生成、(3)表現検索、(5)再ランキング、から成る。

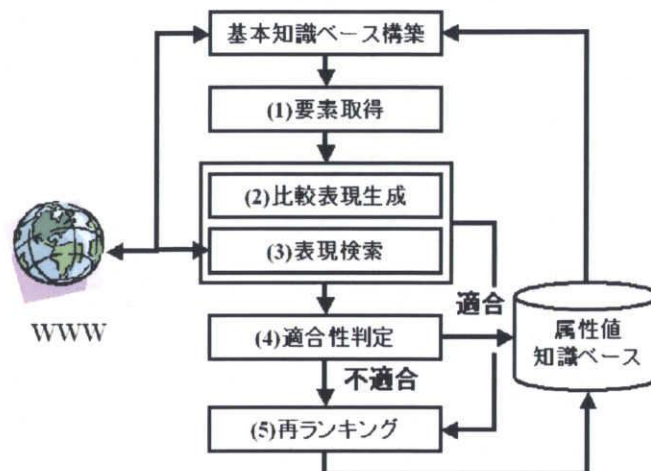


図2 基本知識ベースの精緻化手法概要

(1) 要素の取得

基本知識ベースから、比較表現生成の対象となる概念(名詞)と、その概念の属性値(連体修飾語句)を取り出す。

(2) 比較表現生成

まず「X_ground_γ」という定型パターンを用意しておく。このときのgroundは指標表現であり、指標表現とは、比較表現を構成する「~のように」といった定型表現を指す。また、Xには概念(名詞)が、γには属性値(連体修飾語句)をそれぞれ当てはめて「リンゴのように赤い」のような比較表現を生成する。

使用した指標表現(ground)は、中村の調査による「比喻指標実現形」35種[14]のうち、比

較表現を生成できるもの9種類(注3)の中から、梶井らが使用した「のように」、「のような」に加え、「みたいな」「みたいに」の計4種類を使用した。

(3) 表現検索

(2)で生成した比較表現をクエリとして WWW 検索を行う。検索の結果得られた件数を、(2)で生成された比較表現の頻度として取得する。

(4) 適合性判定

(3)の結果、得られた頻度が閾値 α 以上であれば、属性値は適合と判定される。頻度が α 未満であれば、属性値は不適合と判定される。このとき、閾値 α は、対象属性が顕現性の高い属性値として適合か否かを判定する基準である。本研究では閾値 $\alpha=15$ に設定している。閾値をこれ以下にすると知識ベースの精度が失われ、これ以上にすると必要以上に属性値が削除され、網羅性が損なわれる可能性がある。

(5) 再ランキング

属性値集合の再ランキングを行う。WWW から取得した比較表現の頻度情報を用いて、属性値の尤度の再計算を行う。属性値の尤度計算は、適合性判定の時点で閾値 α 以上となったものとそうでないものにそれぞれ異なった計算式を用いることで、基本知識ベース構築の際に行った属性値ランキングでうまれたランキングの歪みが補正できる。以下に、属性値集合の再ランキングにおける、尤度計算について説明する。

[属性値集合再ランキングのための尤度計算]

概念 X に対して、属性値である用言を w_i とすると、再ランキングをするにあたって、 X を対象とした w_i の尤度は $F(w_i|X)$ によって算出される。この $F(w_i|X)$ は、概念 X について、 w_i が、同概念を対象とした属性値集合の中でどの程度頻繁に使用されるのかを示す指標 $pf(w_i|X)$ と、 w_i を持つ属性値とする概念が X 以外にどの程度存在するのかを示す指標 $icf(w_i|X)$ の積で表され、式(1-1)で求められる。

$$F(w_i | X) = pf(w_i | X) \times icf(w_i) \quad (1-1)$$

(注 3): 上位から順に「のような」、「のように」、「みたいな」、「みたいに」、「に似て」、「に似た」、「のごとく」、「のごとき」、「じみた」[8]

このとき、 w_i に関して、基本知識ベース K の構成時に WWW から獲得してきた頻度情報を $f(w_i|X,K)$ で表し、比較表現「 X_ground_y 」を用いて WWW から得られる頻度を $f(w_i|X,W)$ で表す。

$f(w_i|X,W)$ が適合性判定の閾値 α 以上 ($f(w_i|X,W) \geq \alpha$) であった場合、 w_i は顕現属性値と判断される。このとき、指標 $pf(w_i|X)$ は比較表現をクエリとすることで得られた頻度情報を用いた補正值で表し、式(1-2)で表される。

$$pf(w_i | X) = \frac{f(w_i | X, W)}{\delta(X)} \quad (1-2)$$

ここで $\delta(X)$ は、概念ごとに決まるパラメータであり、顕現属性値と判断された属性値集合の中で最小頻度となるものがそれにあたり、式(1-3)で求められる。

$$\delta(X) = \min(f(w_{i1} | X, W), \dots, f(w_{in} | X, W)) \geq \alpha \quad (1-3)$$

一方、適合性判定において適合と判定されなかった属性値は、顕現属性値ではない属性値（一般属性値）とみなされる。このとき、指標 $pf(w_i|X)$ は、基本知識ベース構築時に得られた頻度情報を用いた補正值で表し、式(1-4)で求められる。

$$pf(w_i | X) = \frac{f(w_i | X, K)}{\delta(X)} \quad (1-4)$$

このとき $\delta(X)$ は、基本知識ベース構築時の最大頻度であったものがそれにあたり、式(1-5)で表される。

$$\delta(X) = \max(f(w_{i1} | X, K), \dots, f(w_{in} | X, K)) \geq \alpha \quad (1-5)$$

$icf(w_i|X)$ は、知識ベースに記録された全概念数 N と、 w_i が属する概念数 $cf(w_i|X)$ の関係を正規化した値であり、式(1-6)にて求められる。

$$icf(wi) = \log \frac{N}{cf(wi)} + 1 \quad (1-6)$$

3.2. 実験と評価

提案手法の有効性を検証するために、実験評価を行った。以下、実験と評価について詳述する。

3.2.1. 実験環境

実験では、提案手法および、ベースラインとして、榊井らの手法を用いた。まず、提案手法およびベースラインを用いて概念ベースを構築する。構築した知識ベースは、提案手法によるものを CB_w 、ベースラインによるものを CB_n とする。次に、10組の概念（表1）[7]について、それぞれの知識ベースを参照し、参照された属性値集合を人手で評価し、その結果を比較検討した。

表1 実験に用いた概念

概念
部屋, 番犬, 風船, チーター, 湖 鏡, 冷蔵庫, 鬼, 風, 流れ星

両知識ベースから参照した属性値集合は、以下のように評価した。5名の被験者（成人男性）に対して10組の属性値集合を提示し、3段階の評価ポイントを付与してもらった。

- 属性値が概念の顕現属性値として妥当である場合：1 pt
- 属性値が概念の属性値としては妥当であるが、顕現属性値としては妥当でない場合：0 pt
- 属性値が概念の属性値として妥当でない場合：-1 pt

ポイント付与終了後、各属性値の合計ポイントを求め、合計ポイントが閾値 β 以上である場合、その属性値が顕現属性値であると見なし、属性値集合における顕現属性値の割合を適合率とした。

ところで、提案手法は、新聞記事と比較してはるかに大規模な WWW を知識源とすることで、獲得可能な顕現属性値の網羅性を高めようとするものである。したがって、本手法の優位性を検証するためには、従来手法に対する網羅性を評価すればよい。

一般に、網羅性を評価するためには、正解データと処理結果を比較する必要がある。ところが、本論文で議論するような属性値集合の正解データを理想的な状態であらかじめ作成することは難しい。そこで、適合率に基づいて網羅性を考慮できる相対再現率 (Relative Recall) という尺度を用いた。

3.2.2. 相対再現率

相対再現率 (Relative Recall) は、二つの概念ベース A, B 間の適合率を利用して求められ [15], 網羅性を間接的に評価することができる。したがって、相対再現率は、正確な再現率を把握できない状況では有効である。相対再現率 $R_{a,b}$ は以下の式で求めることができる。

$$R_{a,b} = \frac{R_a}{R_b} = \frac{C_a}{C_b} \quad (2-1)$$

- $R_{a,b}$: 概念ベース B から得られる概念ベース A の相対再現率
- R_a : A の再現率
- R_b : B の再現率
- C_a : A に含まれる顕現属性値数
- C_b : B に含まれる顕現属性値数

ここでの、 C_a , C_b は知識ベース CB_w , CB_n に含まれる顕現属性値の総数を指す。ここで、それぞれの属性値の適合率 P を利用することで、 $C \approx P \times |A|$ と示すことが可能となる。したがって、上記の式(2-1)を(2-2)に書き換えることができる。

$$R_{a,b} = \frac{C_a}{C_b} = \frac{P_a \times |A|}{P_b \times |B|} \quad (2-2)$$

- P_a : 知識ベース A の適合率
- P_b : 知識ベース B の適合率
- $|A|$: A の属性値総数
- $|B|$: B の属性値総数

相対再現率 $R_{a,b}$ を求めることで、比較対象の網羅性の性能比較が可能である。

3.2.3. 評価結果

提案手法およびベースラインについて、閾値 β を 0~5 まで変動させた場合の相対再現率の変化を図 3 に示す。(◆)は提案手法、(▲)はベースラインを示す。提案手法の相対再現率は全ての閾値 (β) において 1 を上回り、ベースラインよりも網羅性が高いことを示し

た。提案手法の相対再現率は、 $\beta=2$ の時に最大で 1.714 であり、ベースラインの相対再現率は $\beta=4$ の時に最大 0.917 であった。

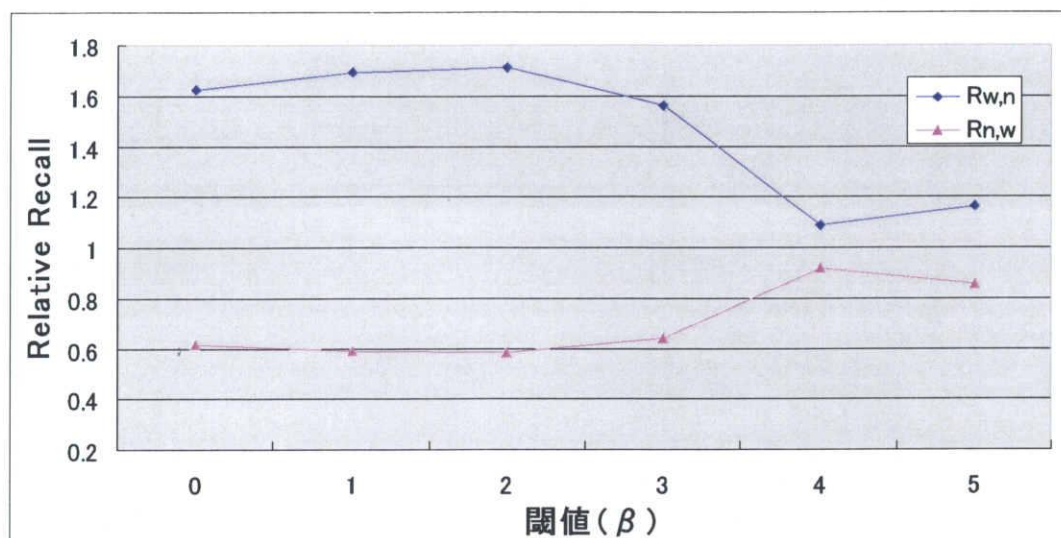


図3 閾値の変化による相対再現率の推移

3.3. 考察

評価結果について考察する。

β が 0~3 では、相対再現率の差が大きい。 β の値が小さいということは、一般性が大きくない顕現属性値（最も強いと言えない典型的特長）を考慮することを意味する。したがって、この結果は、提案手法が、ベースラインと比較して、属性値獲得性能においてより高い網羅性を確保していることを示す。

β が 4~5 では相対再現率の差は小さくなっている。 β の値が大きいということは、より一般性の高い顕現属性値（最も強い典型的特徴）であることを意味する。したがって、この結果は、一般性の高い顕現属性値の獲得性能については、提案手法とベースラインの間に大きな差がないことを示している。このことは、一般性の高い顕現属性値については、新聞記事を知識源とした場合でも十分獲得可能であることも意味する。

また、適合率についてみると、提案手法 (P_w) では 0.52、ベースライン (P_n) では 0.42 であり、意外にも適合率においても若干の優位性が見られた。提案手法が名詞-形容詞・形容動詞の連体修飾関係に基づいて知識を抽出する点はベースラインと同じであるから、基本的な適合率に差は生じないはずである。にもかかわらず適合率に差が生じた理由として、ベースラインと比べて基本知識ベースの網羅性が高いため、ベースラインでは得られなかった一般性の低い顕現属性値が獲得できるようになったことが適合率にも反映された

と考えられる。

次に、提案手法の処理誤りについて述べる。

基本知識ベースとしては抽出できたものの、適合性判定よって顕現属性値として獲得できなかった例が存在した。表2では、「番犬」や「流れ星」の顕現属性値は全く獲得できなかったことがわかる。しかし、これらの概念に対応する基本知識ベースを調べてみると、「流れ星」に対して「ロマンチックな」、「綺麗な」、「番犬」に対して「賢い」、「忠実な」など、顕現属性値として妥当と思われる属性値が含まれていた。これらの属性値を獲得するためには、適合性判定における判定基準の設定において、WWW 以外にも判定用知識源を用いる方法や、適合性判定の閾値の動的決定、比較表現の記述頻度以外にも根拠を求めるなどの工夫が必要である。後者の検討手法については次章にて詳述する。

表2 獲得された顕現属性値の比較

概念	CBw に含まれる顕現属性値	CBn に含まれる顕現属性値
部屋	プライベートな	
番犬		
風船	<u>軽い</u>	<u>軽い</u>
チーター	速い	優美な
湖	<u>静かな</u> , <u>穏やかな</u> (<u>おだやかな</u>), <u>深い</u> , <u>青い</u> , 美しい, <u>綺麗な</u> (<u>きれいな</u>), <u>巨大な</u> , <u>波静かな</u>	<u>大きい</u> , <u>静かな</u> , <u>深い</u> , <u>青い</u> , <u>広い</u>
鏡	<u>美しい</u> , <u>綺麗な</u> (<u>きれいな</u>), <u>キレイな</u> , 不思議な, 透明な	<u>美しい</u> , <u>きれいな</u>
冷蔵庫		重い, 大きい
鬼	強い, 恐い, <u>悪い</u> , 強力な, 大きい, 冷酷な, <u>醜い</u> , 巨大な, <u>真っ赤な</u> , 凶悪な	<u>怖い</u> , <u>恐ろしい</u> , <u>悪い</u> , <u>赤い</u>
風	<u>優しい</u> (<u>やさしい</u>), <u>爽やかな</u> (<u>さわやかな</u>), 自由な, 軽やかな, 冷たい, <u>気まぐれな</u> , <u>清々しい</u> , 自然な, <u>爽快な</u> , <u>涼しい</u> , 涼やかな, 気持ちいい, <u>心地よい</u>	<u>優しい</u> (<u>やさしい</u>), <u>爽やかな</u> (<u>さわやかな</u>), 自由な, <u>心地良い</u> , <u>爽快な</u> , <u>清々しい</u> (<u>すがすがしい</u>), <u>気まぐれな</u> , <u>涼しい</u>
流れ星		

4. 顕現属性値の多段階識別モデルの基礎的検討

本章では、顕現属性値の多段階識別モデルについて検討する。

以下、1節では連想実験によって得た顕現属性値のレベル区別可能性に関する予備調査について述べ、2節では特定のフレームを利用した顕現属性値のレベル付与方法について述べる。3節では、予備実験による結果とパターンと WWW を利用した結果とを比較し、顕現属性値のレベル付与可能性について考察する。

4.1. 予備調査

本節では、顕現属性値集合は多段的に識別することができるという仮説の正当性を確認するために被験者を用いた予備調査を実施したので、調査について詳述する。

4.1.1. 調査環境

調査において、以下のような環境を設定した。

- 調査対象 : 大学生 15 名
- 調査データ : 抽象度の低い名詞概念 30 組
- 調査方法 : 口頭で 30 組の概念[7][16][17]を順に提示し、思いつく特徴（属性値）を 10 秒以内に答えてもらう。

無制限に思いつく特徴を提示した場合、典型性の高い属性値ばかりが収集できるとはいえない。本実験で獲得対象となるのは、対象概念の特徴を説明できる典型性の高い特徴（顕現属性値）である。このような特徴は、複雑な推論を経ずに瞬間的に想起可能であると考えられるので、回答までに要する時間はできるだけ短時間に設定することが望ましい。提示概念を理解して特徴を想起し、その特徴を発話するまでに必要な時間と、複数の特徴が想起された場合に要する時間を勘案し、本調査において、制限時間を 10 秒に設定した。

4.1.2. 調査結果

前節で説明した調査結果について述べる。得られた顕現属性値について、提示した回答者数に関する度数分布を概念の区別なしにみると図 4 のようになる。

この度数分布が複数の分布集合から合成されていると仮定して、境界となる度数を変化させて Wilcoxon の順位和検定を行い、有意な境界が存在するかどうかを調べた。その結果、回答者数 5~6 ($p<0.05$) と 10~11 ($p<0.01$) において有意差が見られた。

さらに、各概念を区別して同様の検定を行った結果、回答者数 5 および 10 付近の 2 箇所では有意差が認められた。一例として「リンゴ」に関する結果を図 5 に示す。この例では、{赤

い}, {丸い, …, 甘い}, {すっぱい, …} の間に有意差が見られる。

これらの結果は、顕現属性値の一般性（多くの人間が一致して想起する可能性）にはある程度レベル差が存在することを示唆している。すなわち、顕現属性値にレベルを付与して区別できる可能性は高いと考察することができる。

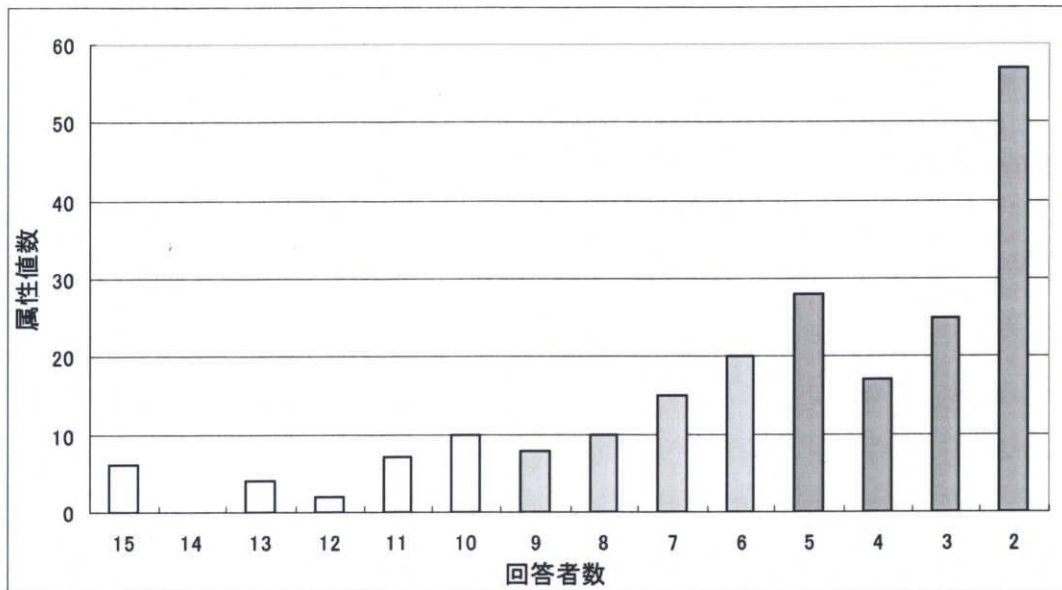


図4 連想実験に基づく顕現属性値毎の回答人数

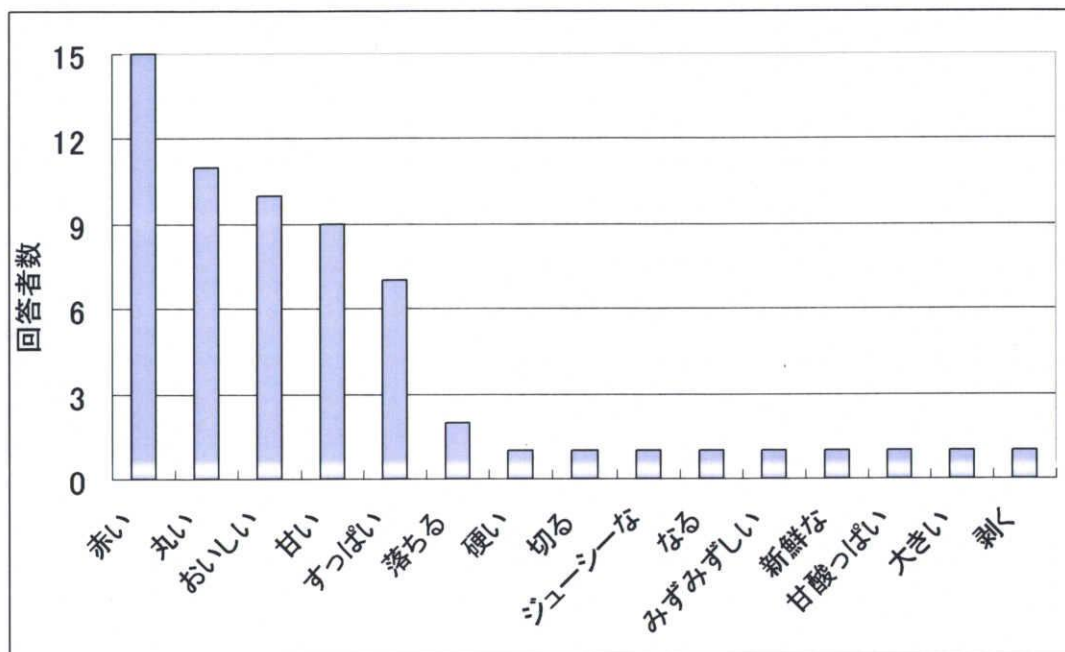


図5 概念「リンゴ」に対する顕現属性値

4.2. 顕現属性値の多段階識別モデル

本章では、与えられた名詞概念とそれを表現する属性値集合に対して、WWW における属性値の適合性判定方法を工夫することによって、顕現属性値集合を多段階に区別する方法について検討する。

4.2.1. 基本的な考え方

2章でも述べたが、梶井ら[8]の手法では、形容詞・形容動詞と名詞の連体修飾関係の共起を基本的な概念 - 属性値関係（例えば、「カラス - 黒い」とみなしている（連体修飾における装定機能の利用）。さらに、これらの共起関係を用いて「カラスのように黒い～」という比喻表現（比較表現）が成立するならば、属性値「黒い」は、概念「カラス」を“喩える概念” 足らしめる強い典型性を示す顕現属性値であると判断できる（連体修飾における比喩的修飾機能の利用）。すなわち、この考え方を利用することで、強い顕現属性値としての適合性判定を行っている。この考え方は3.1.2の適合性判定でも利用されている。

これに加えて、上述の属性値について、「カラスは黒いです。」という叙述表現（断定表現）を考えてみる。この表現が成立するならば、属性値「黒い」は、概念「カラス」の主要な特徴、「カラス」を“喩える概念” としないまでも、一般属性の中でも比較的典型性の高い顕現属性値であるといえる（用言の述定機能の利用）。この考え方を利用することで、許容範囲の広い適合性判定ができる（図6）。

比喩的表現の生成能力および叙述表現の生成能力を評価し、その結果を統合することによって、以下のような推論が可能である。

1. 叙述表現が成立し、かつ比喩表現も成立する。
 - 強い顕現属性値の可能性を強く支持 (S)
2. 叙述表現は成立しないが、比喩表現は成立する
 - 強い顕現属性値の可能性を支持 (A)
3. 叙述表現は成立するが、比喩表現は成立しない
 - 顕現属性値の可能性を支持 (B)

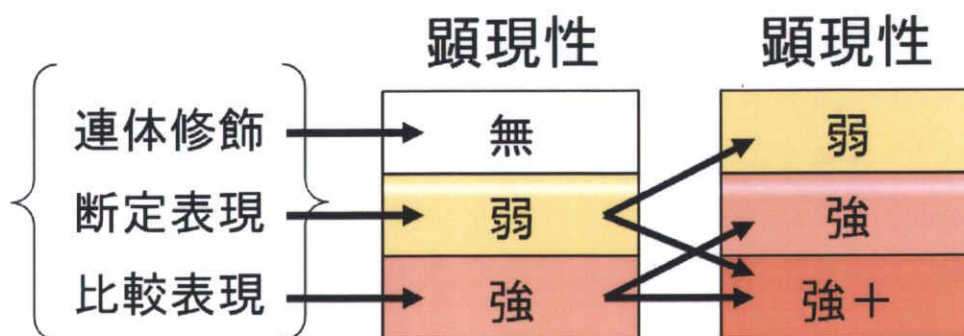


図6 複数表現を利用した顕現属性値の識別

4.2.2. 複数表現を利用した多段階識別モデル

前項で述べた考え方にに基づき、顕現属性値を多段階に識別し、顕現性の強さに応じてレベルを付与するモデルの検討を行う。

1. 対象コーパスから連体修飾関係にある形容詞・形容動詞 - 名詞の共起 $C(X, \gamma)$ を抽出する。
2. 共起関係 C を「 X は γ です」というパターンに適用し、述定関係にある表現を生成し、WWWにおける出現頻度を取得する。
3. 尺度閾値 α によって適合性判定を行い、頻度に基づいてランキングした後リスト1に加える。
4. 共起関係 C を「 X のように γ 」「 X のような γ 」というパターンに適用し、比喩関係にある表現を生成し、WWWにおける出現頻度を取得する。
5. 頻度閾値 β によって適合性判定を行い、頻度に基づいてランキングした後リスト2に加える。
6. リスト1とリスト2を比較し、3段階のレベル（両方適合=S, 比喩適合=A, 述定適合=B）を付与する。

図7では「リンゴ」という概念が処理対象である。頻度閾値 α 、頻度閾値 β は $\alpha=0.1$, $\beta=15$ として考える。形容詞・形容動詞 - 名詞の共起関係に基づいて、属性値 (γ) が抽出される。

まず、「 X は γ です」というパターンの X に「リンゴ」、 γ に「赤い」「甘い」「美味しい」「爽やかな」などを適用して「リンゴは赤いです」「リンゴは美味しいです」のような述定関係にある表現を生成する。これらの表現を WWW 検索し、得られた検索件数に基づいてランキングした時の順位の逆数 (RR 値) が閾値以上と判定された「美味しい」や「赤い」は適合と判定され、リスト1に加えらる。

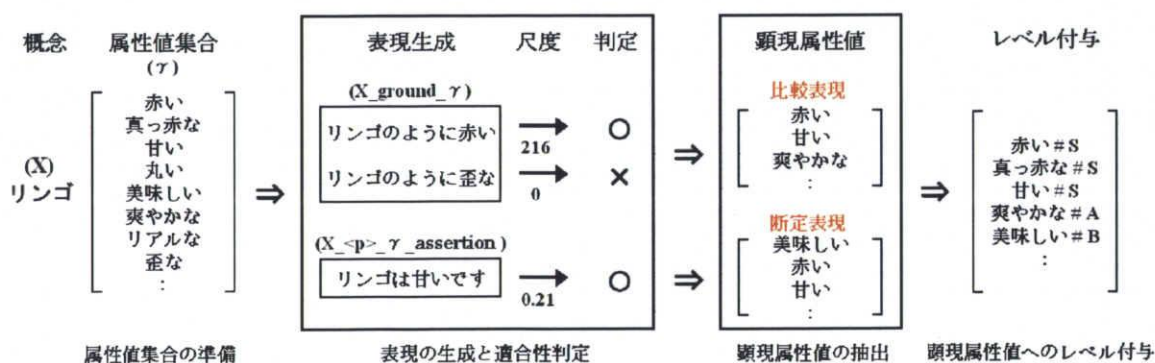


図7 多段階識別モデルの適用例

次に、「Xのように γ 」「Xのような γ 」というパターンから上記と同様に「リンゴのように甘い」「リンゴのように赤い」といった比喩関係にある表現を生成し、WWWにおける出現頻度を取得する。この時出現頻度が閾値以上であったものを適合と判定し、リスト2に加える。

最後にリスト1とリスト2を比較し、両方のリストに存在する属性値「赤い」にはS、リスト2のみに存在する属性値「爽やかな」にはA、リスト1のみに存在する属性値「美味しい」にはBのレベルを付与する。

4.3. 分析と考察

本節では、名詞概念に対する顕現属性値について、2節で行った連想実験によるデータを用いてレベルを区別した結果（人間判断データ）と、3節で行った識別モデルによってレベルを区別した結果（識別モデルデータ）を比較し、その一致性について考察する。

3節では、識別モデルによって顕現属性値を3分割し、それぞれにS、A、Bのレベルを付与した。これにあわせて、2節で述べた連想実験の結果に対しても、支持率の高い集合から順に、S、A、B、Cのレベル付与を行った。ただし、両者のレベルは単に順位尺度として対応するだけのもので、同じ定義を示すものではない。そのため、データによってはCレベルが存在する可能性がある。

人間判断データの各レベルと識別モデルデータのレベルの間で対応のあるもののみの比較結果を見る。対応のない属性値を対象としない理由は、対象が取れていない属性値はデータ不足等の理由から対応が見られていない可能性があり、今評価の対象とすることは適切ではないと考えられるからである。

比較の結果は以下のものであった。人間判断でSレベルと付与されたもののうち、識別モデルがどう識別したかという観点から考える。比較結果を表3に示す。

人間判断でSレベルと認識されたもののうち、同じくSレベルと識別したものは52.0%、Aレベルと認識されたものは44.0%、Bレベルと認識されたものは4.0%であった。

人間判断でAレベルであったもののうち、同じくAレベルと認識したものは54.6%、Sレベルと認識されたものは40.9%、Bレベルと認識されたものは4.5%であった。

人間判断でBレベルであったもののうち、同じくBレベルと認識したものは45.5%、Sレベルと認識されたものは0.0%、Aレベルと認識されたものは54.5%であった。

人間判断でCレベルであったもののうち、Sレベルと認識したものは0件、Aレベルは0件、Bレベルは4件であり、人間がCレベルと判断したものに至っては考察から除外するものとする。

上記の比較結果について考察する。まず、議論を簡単にするために、人間判断に基づいて付与されたレベルに対して、識別モデルによって付与されたレベルがどの程度一致しているかについて述べる。

人間判断データの各レベルと識別モデルデータのレベルの相関性を確認することで、レベル付与結果の一致性を確認する。相関性の確認には Kendall の順位相関係数を利用した。

両データ間の相関係数 τ を算出すると、その値は最大で 0.768、最小で 0.220 の値を示した。 τ の値が揺れるのは、属性値へのレベル付与が 3 段階で行われており、これを順位尺度と見なした場合順位が多数存在するためである。Kendall の順位相関係数についての詳細は付録を見ていただきたい。 τ が最小値の場合に両データ間の相関性について検定を行ったところ、両データ間に相関性が確認された ($p < 0.05$)。また、 τ の平均値を算出すべくランダムで相関係数 τ の値を 10 パターン算出したものを図 8 に示す。これを見ていただくと、主に 0.5 付近に値が集中していることが分かる。この場合の τ の平均値は 0.512 であり、この場合にも両データ間に相関性が確認された ($p < 0.01$)。このことから、人間判断によって区別されるレベルについては、おおむね識別モデルでも同様に区別されると考えてよい。

次に、両データを比較した際の、再現率（網羅性）と適合率（精度）について考えてみる。再現率(Recall)と適合率(Precision)はそれぞれ以下の式で求めることができる。

$$Recall = \frac{\text{両データで対象となるレベル付与が一致した属性値数}}{\text{人間が対象となるレベルを付与した属性値の数}} \quad (3-1)$$

$$Precision = \frac{\text{両データで対象となるレベル付与が一致した属性値数}}{\text{識別モデルが対象となるレベルを付与した属性値の数}} \quad (3-2)$$

まず、再現率の結果を見ると、S レベルは 0.52、A レベルは 0.55、B レベルは 0.45 と、いずれのレベルも 0.5 程度の値を示している。適合率の結果について見ると、S レベルは 0.59、A レベルは 0.34、B レベルは 0.83 という値を示している（図 9）。網羅性、適合率共に人間の判断を代替することを考えた場合には、今後も改良の余地が残る結果である。

このような結果が得られた理由として、以下の二つの可能性が考えられる。

ひとつは識別モデルの判定性能が人間判断と比較して過大評価する傾向が強いという可能性である（表 3）。もうひとつは、人間判断に用いたデータがまだ十分でないため、人間が判断する属性値の判別基準に揺れが生じるため、現状では上記のような結果が得られている可能性である。

前者の場合、今後判定基準の補正を考えるのであれば、3.3 でも述べているように、WWW 以外にも判定用知識源を用いる方法や、適合性判定時の閾値設定基準の見直し、今回用いた表現の記述頻度以外にも根拠を求めるなどの工夫などが必要である。

後者の場合の可能性を正確に確かめるためには、連想実験における被験者数を増やして人間判断によるデータ規模を拡大していく必要がある。

また、人間判断データ同士の比較調査を行い、その再現率と適合率を求めることで、識別モデルデータの限界を見極めていく必要もあると考える。

表3 人間判断データと識別モデルデータの比較結果

		識別モデル		
人間判断	Rank	S	A	B
	S	13	9	0
	A	11	12	12
	B	1	1	10

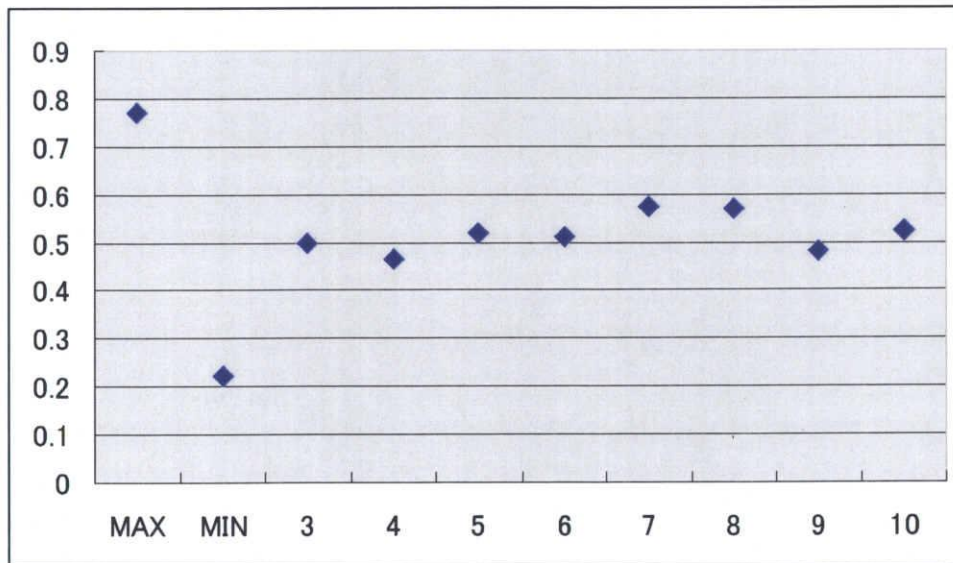


図8 相関係数のランダムサンプリング

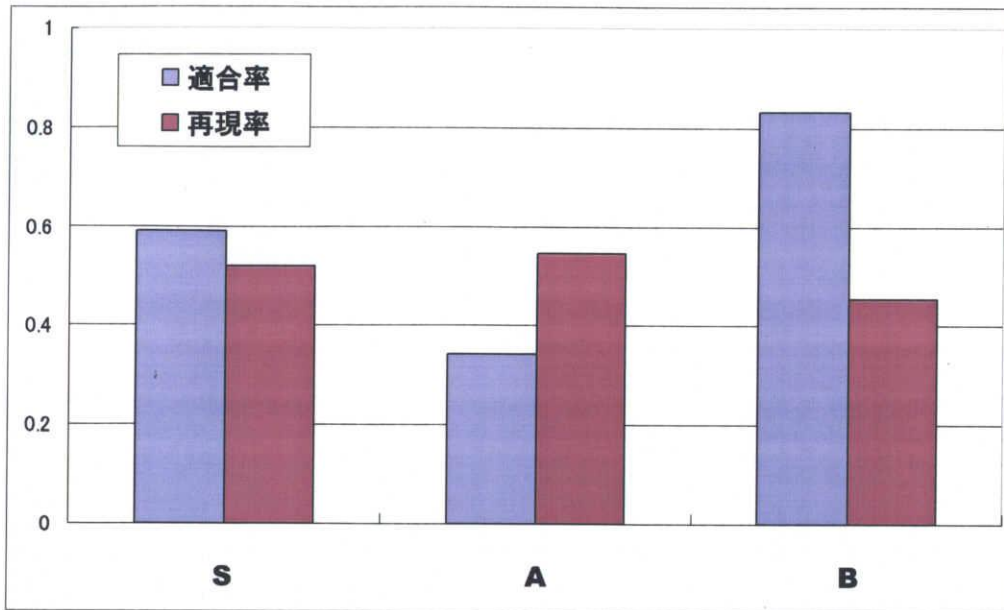


図9 人間判断を基準としたレベル毎の再現率と適合率

5. 結論

本論文では、WWW を知識源として知識ベースを構築する手法について提案し、また、顕現属性値に対してレベル識別が可能であるという仮定し、与えられた顕現属性値の集合に対して、観点の異なる表現の生成可能性を考慮することによって複数のレベルに区別する方法を検討した。

WWW からの知識の直接抽出手法では、WWW から必要とする表現を含むテキスト情報を確保しておくことで、計算量を抑えつつ高い網羅性を持つ知識ベースを構築することに成功した。実験による検証の結果、獲得できる顕現属性値については、従来手法に対して高い網羅性を確保できることが確認でき、さらに、適合率についても優位性が認められた。

顕現属性値の多段階識別手法については、まず、回答時間を短く限定した連想実験を実施して人間判断による顕現属性値データを収集し、それらの分析による予備調査を行うことで顕現属性値が 3 つ程度の集合に有意に区別できることを確認した。次に、名詞概念とその顕現属性値を用いて複数の異なる表現（比喩的修飾表現あるいは叙述文）の生成を試み、それらの生成可能性を評価し、評価結果の組み合わせによって顕現属性値を 3 つのレベルに区別する識別モデルを考えた。人間判断に基づくレベル付与結果と識別モデルによるレベル付与結果を比較し、その一致性に関して分析したところ、概して一致性が確認できた。

今後は、実験規模を拡大し、より詳細な考察を進め、今回試みたような定型表現の生成可能性に基づく顕現性レベルの識別可能性の検証を進めていく。今回行った統計的分析を自動的に行える仕組みについても考察を進めたい。例えば、EM アルゴリズム[18]のような統計的アプローチの応用などが考えられる。

謝辞

本研究を遂行するにあたり、日頃からご指導、ご鞭撻頂きました井須尚紀教授、河合敦夫助教授、榊井文人助手に深く御礼申し上げます。また、多岐にわたり便宜を図って頂きました岡保子事務官並びに、田中みゆき事務官に深く感謝します。特に、榊井文人助手には、普段の研究から論文執筆等、多大なご尽力を頂くと共に、研究効率の悪い著者を、時に優しく、時に厳しくご指導頂きました。さらに、それ以外の面においても、教官である以上に人生の先輩として、様々な助言や今後の人生の参考となるお言葉を頂くことができ、研究室での生活をより有意義なものとして過ごすことができ、感謝の言葉に絶えません。

最後になりましたが、知識不足の否めない著者に対して、確率統計学はじめ、様々なアドバイスをいただきました石田健二君、著者が行き詰った時等に相談にのって頂きました今岡裕貴君はじめ、ここに書ききることは決してできませんが、人工知能研究室で同じ時間を共有してきた先輩、同輩、後輩の全員に心より感謝致します。

研究室に配属されてから3年間、誠にありがとうございました。

参考文献

- [1] G. Lakoff, and M. Johnson: *Metaphors We Live by*, The University of Chicago Press, Chicago, IL, 1988.
- [2] 山梨正明, 比喩と理解, 東京大学出版会, 1998.
- [3] 芳賀純, 安増生, *メタファーの心理学*, 誠信書房, 1990.
- [4] 池ヶ谷有希, 野口靖浩, 鈴木夕紀子, 伊藤敏彦, 小西達裕, 近藤真, 高木朗, 中島秀之, 伊藤幸宏: “対話文脈を利用した構文・意味解析手法の検討”, *人工知能学会第18回全国大会講演論文集*, pp.3E2-10, 2004.
- [5] 榊井文人, 森田あすか, 福本淳一: “比喩指標を利用した曖昧な質問への応答”, *信学技法 NLC2002-39*, vol.102, No.414, pp.7-12, 2002.10.
- [6] 岡本潤, 石崎俊: “概念辞書の構築と概念空間の定量化---連想実験による概念空間の抽出”, *情処研報*, NL130-11, pp81-88, 1999.
- [7] 今井豊, 石崎俊: “比喩理解における顕著な属性の発見手法”, *自然言語処理*, Vol.6, No.5, pp.27-42.1999.
- [8] 榊井文人, 福本淳一, 椎野努, 河合敦夫: “確率的尺度を用いた比喩性検出手法”, *自然言語処理*, Vol9, No5, pp.71-92, 2002.
- [9] 榊井文人, 福本淳一, 荒木健治: “比喩解釈を目的とする World Wide Web を利用した属性値の適合性判定手法とそのフィードバック”, *電子情報通信学会論文誌*, Vol.J89-D, No.4, pp.860-870, 2006.
- [10] 今井豊, 石崎俊: “比喩理解のための顕現属性の幾何学的抽出法”, *人工知能誌*, Vol.15, No.2, pp.309-320.2000.
- [11] 岩山真, 徳永健伸, 田中穂積, “比喩を含む言語理解における顕現性の役割”, *人工知能誌*, Vol.6, No.5, pp.647-681.1991.
- [12] 神崎享子: “連体修飾関係を結ぶ形容詞類と名詞”, *計量国語学*, Vol.21, No.2, pp.53-68, 1997.
- [13] 神崎享子, 井佐原均: “形容詞類の連体用法にみられる連用的な意味”, *計量国語学*, Vol.22, No.2, pp.51-65, 1999.
- [14] 中村明: “比喩表現の理解と分類”, 共立出版, 1997.
- [15] Patrick Pantel, Deepak Ravichandran, and Eduard Hovy: “Towards terascale knowledge acquisition”, In *Proceedings of 20th International Conference on Computational Linguistics: COLING-2004*, pp. 771-777, 2004.
- [16] 平井文人: “いろのえほん”, PHP 研究所, 2001.
- [17] 岡本健: “あか あお きいろ”, 小林芳郎 (監修), ひかりのくに株式会社, 1997.
- [18] 北研二: “確率的言語モデル”, *言語と計算* 4, 東京大学出版会, pp.41-46, 1999.
- [19] 関口洋一, 山本和英: “Web コーパスの提案”, *情報学研報*, IPSJ-SIG-FI, pp.123-130, 2003.