

属性情報を用いた比喩表現における概念の類似性の発見手法

今井 豊 石崎 俊

慶應義塾大学 政策・メディア研究科
〒 252 神奈川県藤沢市遠藤 5322
yim@sfc.keio.ac.jp, ishizaki@sfc.keio.ac.jp

本論文では、二つの名詞概念からなる比喩表現における顕著な属性を自動的に発見する手法を提案する。まず、概念から連想される属性について調べる連想実験を行い、次に、その結果に基づく属性の束をつくってSD法の実験を行う。そして、SD法実験の評定値をパラメータとして用いるニューラルネットワークを使用して、二つの概念に共通でしかも顕現性の高い属性を抽出する。この手法では多数の属性間の顕現性に関する数値的な順序づけが行われるので、多様な概念の組み合わせを含む「TはVだ」の形の比喩表現に対して適用できる。この手法を用いたシステムCOFFSによる実行例を示し、その有効性を検証する。

A Method for Finding Similarity between Concepts in a Metaphorical Expression Using Attribute Information

Yutaka IMAI and Shun ISHIZAKI
Graduate School of Media and Governance, Keio University
3522 Endo, Fujisawa-shi, Kanagawa, 252 Japan
yim@sfc.keio.ac.jp, ishizaki@sfc.keio.ac.jp

In this paper, we propose a method which automatically finds salient features in a metaphorical expression which consists of two noun concepts. First, we prepared a bundle of features based on the result of human association experiments on attributes of concepts, and, using the bundle, we implemented SD(Semantic Differential) Method experiments. Then we extracted common salient features by using a neural network mechanism and the result of the SD Method experiments for the parameters of the mechanism. Since this mechanism has numerical information on the concerning attributes, we can apply this method to analyzing a metaphorical expression, "T is V", on which various combinations of topic and vehicle concepts were given. We show examples calculated by the COFFS system and verify its effectiveness.

1 はじめに

コンピュータの自然言語理解機能は柔軟性を高めて向上しているが、字義通りでない文に対する理解機能については人間と比べてまだ十分に備わっていない。例えば、慣用的でない比喩表現に出会ったとき、人間はそこに用いられている概念から連想されるイメージによって意味をとらえることができる。ここでは、いくつかの共通の属性が組み合わされて比喩表現の意味が成り立っていると考えられる。し

たがって、属性が見立ての対象となる比喩の理解をコンピュータが実現するためには、属性を表す多数の状態概念の中から、与えられた二つの名詞概念に共通の顕著な属性を自動的に発見する技術が重要な要素になると考えられる。本論文では、任意に与えられた二つの名詞概念で「TはVだ」と表現するときの共通の顕著な属性を発見する手法について述べる。また、定量的な基準を設けて顕現性(saliency)の高い属性を絞り込む方法について検討する。具体的には、連想実験に基づいて構成される属性の束を

用いて SD 法 (Semantic Differential) 実験を行い、その結果を入力データとして用いるニューラルネットワークの計算モデルによって行う。

本論文における用語の説明をする。比喩文「TはVだ」において、T(Topic)を被喩辞、V(Vehicle)を喩辞と呼ぶ。本論文で扱う比喩はこの形の隠喩である。

2 比喩理解に関する最近の研究

コンピュータによる比喩理解に関する研究は最近多く見られる。Fass[1]は、比喩理解において概念の階層構造の中で共通の上位概念を持つものに着目し、動詞や名詞に関する類似した対応関係を発見している。Martin[2]は、比喩文を字義通りの文と同様に扱う立場をとり、比喩文に関する明示的な知識をあらかじめ知識ベースに与えている。そして新しい比喩にも対応できるように既知の比喩を拡張できるようにしている。我々の研究では、概念の階層構造や比喩文に関する知識はシステムにあらかじめ与えられておらず、概念を独立に提示された場合の、属性の束に対する SD 法実験の評定結果だけが与えられている。

土井ら [3] は、階層型のニューラルネットワークを用いて属性層の中から比喩の意味を選択する方法を検討しているが、顕著な属性を自動的に抽出することまでは行っていない。諏訪ら [4] は、隠喩理解の類推的アプローチとして、プリミティブマッチング法という効率的に類比決定を行う手法を示した。その研究は、属性ではなく構造・関係が見立ての対象となる比喩を扱っている。森ら [5] は、状況理論から展開された視点を導入した比喩理解のモデル化を行っているが、そこでは、比喩の属性に対する数値的な順序づけはなされていない。岩山 [6] の提案する計算モデルは、情報理論に基づく定式化によって喩辞の属性を選択している。そしてベイジアンネットワークを用いて被喩辞の概念変化も扱っている。内海ら [7] は、隠喩解釈のために特別な手続きを用いることなく文脈に応じた隠喩解釈を行う、関連性に基づく言語解釈モデルを提案している。岩山と内海らの研究では、喩辞の属性情報として、属性名とその重要度や、属性値集合 (属性値とその確率の対) が人手によって与えられている。内山ら [8] は、SD 法実験を行ってその結果から顕現性を計算しているが、被喩辞と喩辞の別々の評定値をもとに比喩表現の顕現性を計算しているわけではない。

本論文では、比喩を構成する概念の組み合わせの変化に対し柔軟に対応でき、共通の顕著な属性を発見する方法について述べる。

3 比喩の特徴発見手法

比喩の特徴発見の準備として二つの認知心理実験を行う。一つ目は概念の属性に関する連想実験で、二つ目は連想実験の結果を参考にして行う SD 法の実験である。そして、SD 法の実験結果を 3.3 節で説明する比喩特徴発見システム COFFS への入力データとして用いる。

3.1 属性の連想実験

文献 [9] で述べられている連想実験システムを使用して、概念の属性に関する連想実験を行った。具体的には、『類語新辞典』[10]の中の「自然」「人物」「物品」という三つのカテゴリーから、小学生にわかりやすいと思われる 90 個の概念を用意し、大学生の被験者 45 人を 6 つのグループに分けて一人につき 15 個の概念を独立に提示した。被験者には提示された概念から連想される属性を自由に記述してもらい、そのとき連想する属性の個数に制限を設けなかった。表 1 に、我々が後で比喩を構成する概念として有用であると考えた 10 個の概念に関する結果をまとめた。自由記述であったので、多様な属性があげられたが、被験者の間で一致度の高い属性も見られた。例えば「チーター」に対しては、ほぼ全員が「速い」をあげていた。

表 1: 連想実験によって得られた主な属性例 (提示概念下の括弧内の数字は連想した属性の平均個数を示す)

提示概念	被験者の 1/3 以上によって連想された属性
部屋 (11.5)	広い、狭い、汚い、きれいな、暖かい
番犬 (8.4)	恐い、うるさい、噛む、吠える、強い、大きい
風船 (7.2)	ふわふわした、軽い、色とりどり、割れる、飛ぶ、赤い、子供っぽい、やわらかい
チーター (5.9)	速い、スマート、美しい、鋭い
湖 (5.6)	静かな、きれいな、深い、冷たい、広い、大きい
鏡 (5.4)	映る、丸い、反射する、割れる、神秘的、光る
冷蔵庫 (5.3)	冷たい、大きい、四角い、白い、保存する、重い
鬼 (4.6)	恐い、強い、赤い、でかい、悪い、角がある
風 (4.3)	強い、冷たい
流れ星 (4.1)	速い、ロマンチック、願い事、美しい

3.2 SD 法

SD 法は、実験心理学で従来から行われている手法であり、例えば多数の形容詞を意味空間上に配置し、概念の情緒・感覚的な意味の定量的な分析を行うときに用いられている [11][12]。対の意味になる形容詞を両端に置いて、提示される概念に対する当てはまり具合を 7 段階か 5 段階の尺度で評定することが多い。本 SD 法実験で両極概念対は、連想実験の結果を参考にして「広い-狭い」「遅い-速い」「きれいな-汚い」「寒い-暑い」など 36 個 (ペア) 用意し、すべての概念に共通の属性の束とした。また分析対象の概念として、連想実験に用いた概念のうちの表 1 に示されている 10 個の概念を用いた。今回は下の「風」のように文脈をつけずに概念を独立にディスプレイ上に提示した。



評定尺度として-3 から+3 までの 7 段階を用いることにし、欄外に「全く当てはまらない」という項目を設けた。提示概念によってはある尺度が「全く当てはまらない」と考えられる場合があるが、この別項目を設けることによって、評定値 0 に対して「尺度の中間に位置する」という意味を明確に持たせることができる。例えば、「風」に対し「非常に速い」というイメージを持っていれば評定値を+3 とする。図 1 に、大学生 22 人による評定の分布例を示す。提示概念は「風」で、横軸に-3(遅い)←→+3(速い)の評定値、縦軸に被験者全員に対する人数の相対頻度をとった。

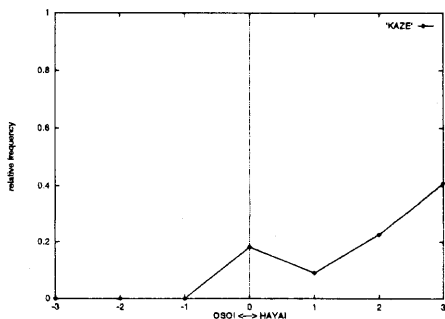


図 1: 評定値の分布例
(「遅い-速い」に関する「風」の評定値分布)

3.3 特徴発見の方法

図 2 に、SD 法の結果に基づく比喩特徴発見システム COFFS (COmmon Features Finder System) の構造を示す。SD 法に用いる両極概念対の個数を m とするとき、 $m+2$ 個のノードを持ち、 f_1 から f_m は両極概念対のノードを表し、 T と V はそれぞれ被喩辞 (Topic) と喩辞 (Vehicle) のノードを表す。

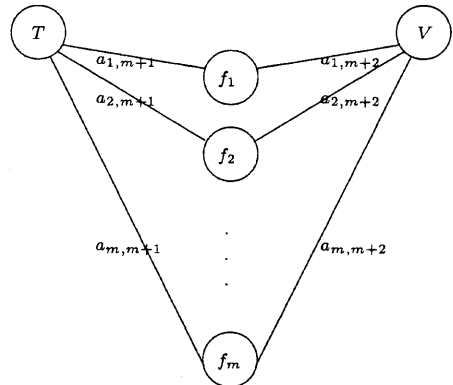


図 2: COFFS の構造

T は被喩辞, V は喩辞, f_i は属性.
 $a_{i,j}$ は遷移行列の要素.

このネットワークを用いて、次のような定式化を行う [13]。

$$\begin{aligned} \mathbf{x}(t+1) &= A * \mathbf{x}(t) + \mathbf{c} \quad (t = 0, 1, 2, \dots) \quad (1) \\ \mathbf{x}(0) &= \mathbf{c} \quad (2) \end{aligned}$$

ここで $\mathbf{x}(t)$ はノード ($f_1, f_2, \dots, f_m, T, V$) の時刻 t における活性値を表す。 f_i について、活性値が大きいほどその属性の顕現性が高いと考えられる。

$m+2$ 個の要素を持つ定常入力 \mathbf{c} の i 番目 ($i = 1, 2, \dots, m$) の要素には、SD 法実験の f_i に関する T と V の評定値分布の形状の類似度が反映される。 \mathbf{c} の i 番目 ($i = 1, 2, \dots, m$) の要素は、 f_i に関する T と V の分布の差の自乗和を計算し、その逆数をとって正規化したものとする。また、「 T は V だ」において T よりも V の顕現性の方が重要であると考えて、 T に対応する $m+1$ 番目の要素は 1、 V に対応する $m+2$ 番目の要素は 2 とする。

$(m+2) \times (m+2)$ の遷移行列 A には、SD 法実験の T と V に関する平均評定値の符号の一致不一致と平均評定値の絶対値の大きさが反映される。

Aの要素は次のようにして決める。まず「TはVだ」が与えられたときの $f_i(i=1,2,\dots,m)$ に対するTの平均評定値 $M_{T,i}$ とVの平均評定値 $M_{V,i}$ を求める。 $M_{T,i}$ と $M_{V,i}$ が同符号である場合には、下のように絶対値を代入する。 $M_{T,i}$ と $M_{V,i}$ が異符号である場合には、 $|M_{T,i}|$ と $|M_{V,i}|$ の箇所をともに0にする。ただし、対角成分は1とし、リンクの存在しない要素は0とする。

$$\left(\begin{array}{cccc|cccc} 1 & & & & & & & 0 \\ & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & \vdots & \\ & & & & & & |M_{T,i}| & |M_{V,i}| \\ 0 & & & & & & \vdots & \vdots \\ \hline & & & & & & 1 & 0 \\ \dots & & & M_{T,i} & \dots & & 1 & 0 \\ \dots & & & M_{V,i} & \dots & & 0 & 1 \end{array} \right)$$

このような行列を収束のために正規化する。ところで図2のCOFFSの構造を見ると、T,Vと f_i とはリンクの個数に大きな差があることがわかる。これは、比喩文における被喩辞・喩辞とそれを構成する属性との相違を反映したものであるが、収束のために行列を正規化する(この場合、行和の最大値ですべての要素を割って、行和の最大値を1にする)と、TとVのノードに対応する $m+1$ 行目と $m+2$ 行目に比べて f_i のノードに対応する $1\sim m$ 行目の値が非常に小さくなってしまふ。そこで行列の要素間で数値のバランスをとることを考え、 $1\sim m$ 行目と $m+1$ 行目・ $m+2$ 行目とで別々に正規化を行ったものを選択行列Aとする。こうすることによって、後ですべてのノードの活性値を比較することが妥当になると考えられる。

以下、計算式(1)(2)に従って反復計算し、 x が収束したときの f_1 から f_m の活性値を調べ、顕著な属性を発見する。また、 f_i とTとの活性値の比を調べて、特に顕現性の高い属性を抽出する。

4 実行例

「風船は流れ星だ」という比喩文を例にして実行結果を説明する。この場合、被喩辞T(Topic)が「風船」で、喩辞V(Vehicle)が「流れ星」である。まず、SD法実験(被験者は大学生22人)の評定値の分布をもとに、両極概念対 $f_i(i=1,2,\dots,36)$ に対してTとVの分布の差を $-3\sim+3$ についてとり、その自乗和の逆数を計算して正規化する。この値が大きいくほど評定値分布の形状が類似しているといえる。

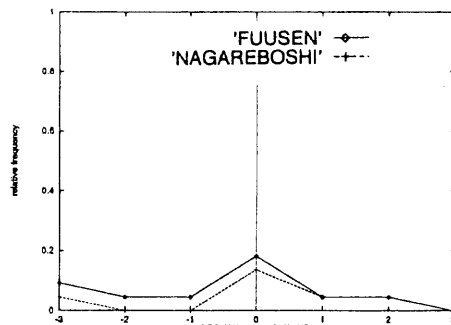


図3: 「風船は流れ星だ」における処理前1位「愚かな-賢い」についての評定値分布

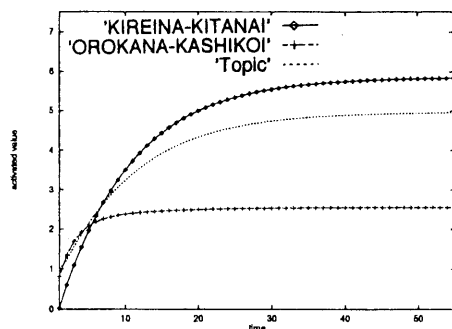


図4: 活性値の収束の様子

図3には、COFFSによる処理を行う前の「風船」と「流れ星」の形状が最も類似している両極概念対「愚かな-賢い」についての評定値分布が示されている。図3を見ると、確かに形状はよく類似しているが、縦軸(相対頻度)の値がかなり小さいことがわかる。これは、被験者の多くが「風船」と「流れ星」のいずれに対しても「愚かな-賢い」という評定尺度は「全く当てはまらない」と考えていたことを示している。また横軸(評定値)について見ても、どちらかの端に偏っているとはいえないので、顕現性が低いといえる。

そこで、TとVの平均評定値の符号が一致しているときにその平均評定値の絶対値を要素に持つ行列を用いて(1)(2)式による反復計算を行った。図4に、すべてのノードの活性値が収束するまでの「きれいな-汚い」と「愚かな-賢い」およびTopic「風船」の活性値を示す。収束後の活性値を調べると、36個の概念対の中で「きれいな-汚い」という概念対の活性値(5.86)が最も大きかった。そして平均評定値の絶対値の符号から「きれいな」が最も顕現性の高い属性であるという結果を得た。図5には、「きれいな-汚い」についての「風船」と「流れ星」の評定値分布が示されている。これを図3と比較すると、

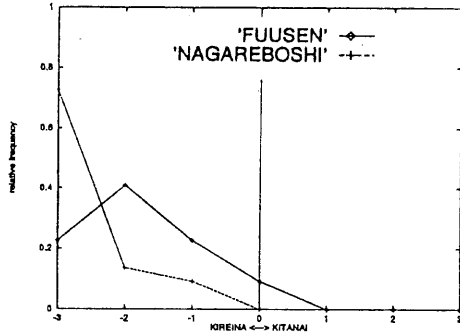


図5: 「風船は流れ星だ」における処理後1位「きれいな-汚い」についての評定値分布

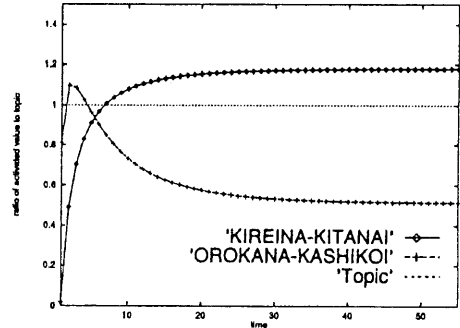


図6: Topic に対する活性値比の収束の様子

表2: 比喩文ごとの顕著な属性の例

比喩文	活性値の大きい属性・上位5個
「風船は流れ星だ」	1. きれいな 2. 美しい 3. 高い 4. 夢のある 5. 丸い
「風船は鏡だ」	1. 割れやすい 2. きれいな 3. 透き通った 4. 弱い 5. 美しい
「チーターは流れ星だ」	1. 速い 2. 美しい 3. カッコいい 4. 鋭い 5. 独立的
「チーターは風だ」	1. 速い 2. 神秘的 3. 軽い 4. カッコいい 5. 美しい
「チーターは鬼だ」	1. 強い 2. 恐い 3. 速い 4. 独立的 5. 大きい
「番犬は鏡だ」	1. 強い 2. 大きい 3. うるさい 4. 恐い 5. 重い
「番犬は鏡だ」	1. 役に立つ 2. 忠実な 3. 天切な 4. ありがたい 5. 鋭い
「流れ星は風だ」	1. 神秘的 2. 速い 3. 透き通った 4. 美しい 5. きれいな
「湖は鏡だ」	1. きれいな 2. 早らな 3. 役に立つ 4. 大切な 5. 冷たい
「部屋は冷蔵庫だ」	1. 四角い 2. 必要不可欠な 3. ありがたい 4. 大切な 5. 役に立つ

「きれいな」の方に偏っていて、しかも相対頻度の高い分布であることがわかる。これによって「きれいな」の顕現性が示されている。活性値の大きい上位5個から得られた属性をあげると、「きれいな」「美しい」「高い」「夢のある」「丸い」となる。

表2には、「風船は流れ星だ」「風船は鏡だ」など10個の比喩文についての実行結果が示されている。これを見ると、TとVの組み合わせによって、上位にくる属性に違いがあることがわかる。例えば、「チーターは流れ星だ」では「速い」「美しい」がそれぞれ1位、2位であるが、「チーターは鬼だ」では「強い」「恐い」がそれぞれ1位、2位になっている。これは直観と合致しているといえる。しかし、「部屋は冷蔵庫だ」では、SD法実験に用いた両極概念対の中に「寒い-暑い」「暖かい-涼しい」などがあるにも関わらず、「四角い」「必要不可欠な」などの人間が比喩として読むときにはあまり意識されないものが属性としてあげられている。これは、SD法

実験で文脈情報がなく独立に提示された「部屋」に対して「寒い」「涼しい」の方に評定した被験者が少ないことによる。「部屋」は暑さ・寒さに対して中立的であり、文脈の影響を強く受ける概念であるためと考えられる。

次に、文献[14]では扱わなかった問題で、抽出すべき属性と必ずしも抽出する必要がない属性を識別する基準について考える。ただし、あらかじめ与えられた両極概念対を属性の集合としてとらえることにする。実際には、あらかじめもっている属性の個数が数十個では多数の概念をカバーすることができないが、ここではその大規模化の問題には触れない。

さて、(1)式の説明で、収束のために行列Aを正規化するとき、 T, V と f_i とで別々に正規化を行って要素全体のバランスをとることを述べたが、これを利用して、 f_i の活性値と $T(\text{Topic})$ の活性値との比較を行う。Topicの活性値に対する f_i の活性値の比(Topicに対する活性値比)が大きい両極概念対ほど、与えられた比喩文の理解に寄与する属性を提供していると考えられ、評定値の偏っている方の端にある形容語の顕現性が高いといえることができる。

図6には、「きれいな-汚い」と「愚かな-賢い」のTopic(「風船」)に対する活性値比が描かれている。「きれいな-汚い」の初期値はほとんど0に近いが、収束時には約1.2になっている。このように収束時の活性値比が1より大きい概念対を特に顕著な属性と考えることにすると、この場合は平均評定値の符号から「きれいな」を顕現性の特に高い属性であるとみなすことができる。逆に、「愚かな-賢い」は、始めのうちは1を越えているが収束時には約0.5になっているので、顕著な属性として寄与していないといえる。

表3には、活性値比が1以上であることによって抽出された顕現性の特に高い属性が表2と同じ10個の比喩文について示されている。これによって、

表3: Topic に対する活性値比 (1 以上)
によって選択された属性

比喩文	被喩辞の活性値よりも大きい 活性値を持った属性
「風船は 流れ星だ」	きれいな、美しい、高い、 夢のある
「風船は鏡だ」	割れやすい
「チーターは 流れ星だ」	速い
「チーターは 風だ」	速い
「チーターは 鬼だ」	強い
「番天は鬼だ」	強い
「番天は鏡だ」	役に立つ
「流れ星は風だ」	神秘的、速い
「湖は 鏡だ」	きれいな、平ちな、役に立つ、 大切な、冷たい
「部屋は 冷蔵庫だ」	四角い、必要不可欠な、 ありがたい、大切な、役に立つ

属性が絞り込まれていることがわかる。このように、Topic に対する活性値比が 1 以上という定量的な基準によって、有効な少数個の属性に絞り込むことが可能になる。

5 おわりに

本論文では、まず、概念の属性に関する連想実験によって比喩文を構成する概念の属性を両極概念対として表現した。そして、両極概念対を用いて SD 法実験を行い、その結果とニューラルネットワークシステムを組み合わせることで比喩文における二つの概念に共通の顕著な属性を自動的に抽出した。その手法をいろいろな概念の組み合わせに対して適用し、直観的にはほぼ妥当な結果を得た。また、収束時の活性値の比を用いて属性の絞り込みを行った。

今後の課題としては、「部屋は冷蔵庫だ」のように、喩辞の顕著な属性に対して被喩辞が中立的である場合の比喩の特徴を抽出することがあげられる。また、あらかじめ与える属性の集合をどのように構成するかという問題にも取り組む必要がある。それから、文脈情報を扱うことも重要な課題である。

謝辞

本研究で行われた連想実験では、富士ゼロックス株式会社の大熊智子氏と慶應義塾大学環境情報学部 4 年の岡本潤君に大変お世話になりました。ここに感謝の意を表します。また、心理実験に多くの協力をいただいた慶應義塾大学環境情報学部の自然言語論 II の履修者の方々に感謝致します。

なお、本研究は文部省科研費一般研究 B 「言語の状況依存性の認知モデルと文脈理解システムの研究」の援助を受けて行われました。

参考文献

- [1] Fass, D.: met*: A Method for Discriminating Metonymy and Metaphor by Computer, *Computational Linguistics*, Vol.17, No.1, pp.49-90 (1991).
- [2] Martin, J.: Computer Understanding of Conventional Metaphoric Language, *Cognitive Science*, Vol.16, pp.233-270 (1992).
- [3] 土井晃一, 佐川浩彦, 田中英彦: ニューラルネットワークを用いた隠層理解, 「学習のパラダイムとその応用」シンポジウム論文集, pp.1-10(1989).
- [4] 諏訪正樹, 元田浩: プリミティブマッチング法: 隠層的見立て知識の学習法, 人工知能学会誌, Vol.9, No.3, pp.417-425(1994).
- [5] 森辰則, 中川裕志: 意味マッチングによる比喩理解モデル, 情報処理学会論文誌, Vol.32, No.3, pp.345-353(1991).
- [6] 岩山真: 計算機による比喩理解モデルに関する研究, 東京工業大学学位論文 (1992).
- [7] 内海彰, 菅原道夫: 関連性理論を用いた文脈の中の隠層解釈の計算モデル, 情報処理学会論文誌, Vol.37, No.6, pp.1017-1029(1996).
- [8] 内山将夫, 板橋秀一: 被喩詞の意味と比喩表現の意味との違いを示す指標, 自然言語処理, Vol.3, No.3, pp.31-51(1996).
- [9] 大熊智子, 石崎俊: 認知実験に基づく概念辞書の構築と検索, 情報処理学会自然言語処理研究会研究報告 112-18(1996).
- [10] 大野晋, 浜西正人: 類語新辞典, 角川書店 (1981).
- [11] 楠見孝: 比喩の処理過程と意味構造, 風間書房 (1995).
- [12] 石崎俊, 田中茂範, 今井むつみ: 柔軟な意味解析のための概念空間の定量化, 情報処理学会自然言語処理研究会研究報告 100-3 (1994).
- [13] Hasida, K. et al.: A Connectionist Approach to the Generation of Abstracts, Kempen, G. (Ed.), *Natural Language Generation*, pp.149-156, Martinus Nijhoff Publishers(1987).
- [14] 今井豊, 石崎俊: 比喩理解における顕著な属性の抽出, 電子情報通信学会技術研究報告 NLC95-80 (1996).
- [15] 石崎俊, 田中茂範, 今井むつみ: 対話文理解のための解析手法と認知意味論的モデルの研究, 文部省科研費重点領域「音声対話」研究成果報告書 pp.187-194 (1996).
- [16] 岩下豊彦: SD 法によるイメージの測定, 川島書店 (1983).
- [17] 大森隆司, 山岸信雄: ニューラルネットワークによる概念記憶の表現と操作のモデル化, 計測自動制御学会論文集 101A-4(1994).
- [18] 諏訪正樹, 岩山真: 比喩の計算モデル, 情報処理学会誌, Vol.34, No.5, pp.566-575(1993).