

概念連想を用いた直喩表現の解釈

奥田裕也[†] 吉村枝里子^{††} 土屋誠司^{††} 渡部広一^{††}

近年コンピュータが発展し、コンピュータがない生活は考えられないぐらい世の中に浸透している。そしてコンピュータは、人間のパートナーとなることが期待されている。しかし、コンピュータの機能が多岐にわたり、使いにくくなっているのも事実である。もしもコンピュータに、人間と同じようなコミュニケーションができれば、コンピュータを誰でも容易に扱うことができると考えられる。本稿では、比喩表現をコンピュータに理解させる手法を提案する。

Interpretation the Metaphor Expression based on Concept Association

Yuya Okuda[†], Eriko Yoshimura^{††}
, Seiji Tsuchiya^{††} and Hirokazu Watabe^{††}

Computers have developed recently. It's hard to imagine life without computers nowadays. Computers are expected partners in modern society. But it is hard to take a conversation for computers and to use very easily. If computers can take a conversation with people, we will be able to use computers very easily. We have to communicate with computers. In this paper, a method of metaphorical understanding is proposed. Metaphorical is one of conversation expressions and we frequently use metaphorical expressions in conversation. If computers can understand metaphorical expressions, we will be able to communicate with computers.

1. はじめに

近年、コンピュータの発展と共に人工物の知的化が進められてきている。すなわち、コンピュータが人間にとって単なる計算機械や道具ではなく、良きパートナーとして機能することが求められてきている。そのためには、人間同士が日常的に行っているコミュニケーションの仕組みをモデル化し、それをコンピュータと人間の言語インタフェースとして構築・実現することが重要な課題となる。そして、コミュニケーションを行うには、コンピュータに人間の意図を理解し、曖昧な表現にも適宜対応させることによって円滑な会話や常識的で柔軟な判断をさせる必要がある。

人間はあいまいな情報を受け取った場合でも、適宜に解釈し、適切に会話を進め、適切な行動をとることができる。これは、人間が長年の経験から得られた情報によって「常識」を確立しているからである。人間が日常的に用いている常識には、大きさ・長さ・重さ・速さなど比較的数値に直し易い量に関するものや、クリスマス・夕方といった時間に関するもの、病院には医者がいる、診察やお見舞いをするところという場所に関するもの、さらには、林檎は赤いや丸いといった五感で感じる感覚や、赤ちゃんは可愛いといった一般的な印象である知覚などがある。

コミュニケーション手段のひとつ、会話には様々な表現方法があり、その中には比喩表現という表現方法がある。比喩表現とは、一方の単語の特徴を別の単語に置き換える手法である。詳細については、2章で記述する。人間は会話や文章のなかに比喩表現が用いられていても、「常識」を基に状況に応じて意味を解釈することができる。例えば、会話の中で「林檎のような頬」という比喩表現を相手が用いた場合、我々はまず、人間の持つ「常識」から林檎は「赤い、球状、おいしい」という単語を連想する。次に「赤い、球状、おいしい」から頬に適した語を「常識」によって判断する。頬というのは、顔の一部であり「球状、おいしい」と表現するのは間違いであると判断できる。これは頬には、色の概念が存在することや、味の概念は存在せず、また頬は球状ではないという「常識」が、人間には構築されているからである。

しかし、コンピュータには「常識」がなく、人間と同じように比喩表現の意味解釈を行うことが困難である。また現在コンピュータに一般名詞から感覚的に連想される感覚語（赤い、甘い等）と知覚から連想される知覚語（可愛い、恐ろしいなど）を連想するメカニズム[1]が存在する。しかし感覚語と知覚語が登録されている一般名詞の数が限定されているため多量の名詞を扱うことは困難である。

そこで本稿では、Web から自動構築した大規模格フレーム[2]を用い、名詞に対する

[†] 同志社大学大学院工学研究科
Graduate School of Engineering, Doshisha University
^{††} 同志社大学理工学部
Faculty of Science and Engineering, Doshisha University

感覚語と知覚語を取得し、Web上で利用される頻度を重みとして付与することで、比喩表現の文を意味の通じる自然な文へと、変換するシステムの構築を行う。意味の通じる自然な文に変換し理解することで、人と同じ「常識」を持つことができたと言える。

2. 比喩表現

比喩表現とは、物の性質を同じ性質をもつ別の言葉に置き換えて表現する技法である。例えば、「氷のような性格」という比喩表現では、「氷」という言葉を用いて「冷たい」を表現している。「氷」に「冷たい」という概念が存在するため比喩表現として成立する。このように、同じ性質をもつ別の言葉に置き換えて表現することが比喩表現である。

比喩表現は強調の目的で用いられる。例えば「皿のような目」と表現されると、これは単に「丸い目」だけではなく、「(すごく)丸い目」をイメージする事ができる。これは我々が「目」と「皿」を想像する場合、きれいな円形を想像するからである。このように、より強調して物事を伝える場合に比喩表現は用いられる。

また比喩表現は、ある物を同じ性質のある別の物に例えて、相手に伝えようとする場合に用いられる。例えば「10センチぐらいの大きさ」と表現された場合、すぐにどれぐらいの大きさかイメージする事ができない。しかし「林檎ぐらいの大きさ」と表現であれば、我々は容易にイメージする事ができる。これは、我々が日常的に林檎の大きさを「常識」として持っているからである。このように相手にわかりやすく物事を伝える場合に比喩表現は用いられる。比喩表現には、直喩表現と隠喩表現がある。直喩表現とは、単語と単語が「みたいな」や「のような」で接続される比喩表現である。「地獄みたいな生活」や「りんごのような頬」のように用いられる。隠喩表現とは、「地獄の生活」のように「のような」などの比喩特有の形式ではなく、比喩であることを明示しない形式である。

比喩表現では、例えられる概念を被喩詞、例える概念を喩詞と呼ぶ。例えば「氷のような性格」の場合、被喩詞は「性格」であり、喩詞は「氷」である。顕現性落差理論 (Ortony 1979) では、被喩詞の性質よりも喩詞の顕現性の高い性質を優先されるべきだとしている。投稿する論文の作成から、論文が掲載された論文誌が出版されるまでの流れは、次の通りである。

3. 先行研究

比喩表現の理解を目的とした研究として、岩山らの研究がある[3]。彼らは、ひとつの名詞をその名詞の特性を示す属性と、属性の詳細を示す属性値の集合で表現している。名詞の属性と属性値の例を図1に示す。

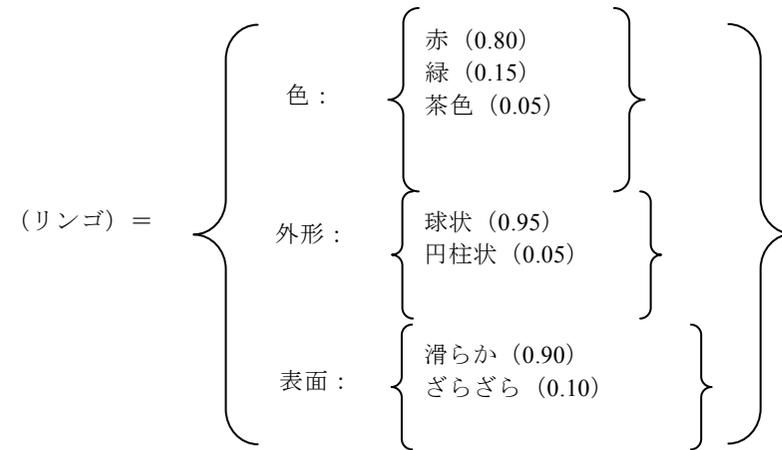


図1 属性の例

図1の「色」「外形」「表面」が属性名、「赤」「緑」「球状」「滑らか」などが属性値である。属性値の右の数値は、名詞が属性値を持つ確率を表している。ある名詞の顕現性の高い属性値、すなわちその名詞が持つ典型的な性質が、喩詞から被喩詞に移されることにより、比喩表現を理解する。

これらの属性値や属性値の確率などは、心理実験により人手で作成したもので、大規模な解析システムを構築するのは困難である。そこで本稿では、属性値およびその確率を、Webから自動構築した大規模格フレームと関連度計算方式[4]を用いて定義し、膨大な名詞に対応させる。

4. 研究概要

本稿で提案するシステムの前提条件とシステムの概要について記述する。

4.1 前提条件

本稿で取り扱う比喩表現は、比喩表現のなかでも一般的な直喩表現に限定する。また固有名詞を用いた直喩表現は、対象外とする。固有名詞の場合、連想する語が人によって変化するためである。

4.2 システム概要

比喩表現の文を意味の通じる自然な文に変換するシステム、比喩表現理解システムの概要を記述する。比喩表現理解システムとは、入力された比喩表現の文を自然な文

に変換するシステムである。例えば「林檎のような頬」と入力されると「林檎」、「頬」から共通して連想される語を取得し、「赤い頬」と出力するシステムである。このように直訳のできる、直接的な表現の意味の通じる文に変換することによって、コンピュータにも比喻表現を理解することができ、円滑な会話や正確な文書要約が可能になると考えられる。なぜなら直訳ができれば、話し相手が何を伝えたいのか、筆者はこの一文で何を表現したいのかが容易に読み取ることができるからである。本システムは、Web から自動構築した大規模格フレームにより連想される形容詞とその重みを取得し、五感知識ベース[1]、知覚知識ベース[1]と概念ベース、関連度計算方式を用いて出力語の絞り込みを行い、間接的表現である比喻表現の入力文を、直接的表現の文に変換し出力をするシステムである。

5. 関連技術

本章では、システムを構築する上で重要となる関連技術について記述する。

5.1 Web から自動構築した大規模格フレーム

Web から自動構築した大規模格フレーム（以下、格フレームとする）とは、動詞とその動詞に関係する名詞を用法ごとに整理したものである。この格フレームは、Web 上の約 5 億文の日本語テキストから自動的に構築されている。また、格フレームに含まれる動詞の数は約 5 万語である。

この格フレームを用いることで、動詞からその動詞に結びつく名詞、格、頻度を取得できる。また、名詞からその名詞に結びつく動詞、格、頻度を取得することも可能である。格フレームの頻度とは Web 上において、その名詞と動詞が存在する回数を指す。

5.2 概念ベース

「概念ベース」とは複数の国語辞書や新聞から自立語を抽出して自動構築した、概念を表す語の知識ベースである。概念ベースには現在約 12 万語の概念が登録されている。

ある概念 A を、 A と関連が強いと考えられる「属性」（概念を特徴付ける語） a_i と、属性の重要度を表す「重み」（概念を属性がどの程度特徴付けているかを 0.0 から 1.0 の値で表す） w_i の集合で定義する。概念 A の属性数を n 個とすると式 (1) のように表せる。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_n, w_n)\} \quad (1)$$

w_i は 1 に近いほど概念にとってより重要な属性である。ここで、 a_i を概念 A の一次属性と呼ぶ。 a_i もまた概念ベースに登録された語である

ため、 a_i の属性も同様に導くことができる。これを概念 A の二次属性と呼ぶ。一つの概念は N 次属性までをその構成要素として持つ。

5.3 関連度計算方式

関連度計算方式[5]とは、概念ベースで定義された任意の概念と概念の関連の強さを、定量的に評価する手法である。それぞれの概念が持っている属性とその重みによって関連度計算は行われ、その結果は関連度という数値で表すことができる。関連度の値は 0.0～1.0 の実数値をとり、値が高いほど関連の深い語であることを意味する(表 1)。表 1 の例では、「本屋」に対して、「書店、店、学校」の中で「書店」が「本屋」と最も関連が強いということが数値によって判定することができる。

表 1 関連度の例

基準概念	対象概念	関連度
本屋	書店	0.791
	店	0.211
	学校	0.002

5.4 シソーラス

シソーラス[6]とは広く同義語、類義語を整理したもので、約 2700 の意味属性の上位下位関係、全体部分関係を木構造で示すものである。約 13 万語が登録されており、親子・兄弟関係を持つ語についてはその関係も記述している。

6. 比喻表現理解システム

比喻表現理解システムは、4.1 節で述べたように、入力文は固有名詞を含まない直喩表現に限定する。

比喻表現理解システムは、はじめに喩詞と被喩詞に相当する単語を抜き出す。次に格フレームにより、抜き出した二語に接続される語を取得する。取得した語を候補語とし、候補語を関連度計算方式によって絞り込みを行い、結果を出力する。

6.1 喩詞・被喩詞の取得

比喻表現理解システムに喩文が入力されるとまず喩詞、被喩詞を取得する。入力文を直喩表現と限定しているため、「のような」又は「みたいな」の前後の単語を取得する。直喩表現には、他にも「的な」や「風な」などにより表現される直喩表現なども存在するが、今回一般的な表現である「のような」又は「みたいな」の表現に限定する。

6.2 格フレームによる形容詞の取得

形容詞取得を行う方法は喩詞、被喩詞を取得後、格フレームを用いて喩詞、被喩詞を名詞に設定し、それぞれ接続される動詞と件数を取得する。このとき、接続される格は指定せず、接続される動詞をすべて取得する（以下、取得語とする）。名詞を修飾

する形容詞は、名詞と形容詞を入れ替えて格助詞を接続させることにより、主語・述語の関係としても成立する。例えば、「赤いりんご」の場合、格助詞“が”を接続させることにより「りんごが赤い」というように主語・述語の関係として成立する。

次に取得語が、五感知識ベースと知覚知識ベースの同義語の単語と表記一致する場合、感覚語または知覚語に変換する。

次に茶笥を用いて、取得語を形容詞と形容動詞に絞り込む。形容動詞は形容詞に変換する。

6.3 候補語の決定

喩詞、被喩詞のそれぞれの取得語から、表記一致の場合候補語とする。また、取得語数は、喩詞の取得語数を2倍にして足し合わせる。2倍にするのは、顕現性落差理論より比喩表現では、喩詞の性質がより優先されるからである。

6.4 関連度計算方式による候補語の重み変更

格フレームによる語の取得は、Web上の約5億文の日本語テキストを参照しているため、取得語には雑音も多く含まれてしまう。そこで候補語を、関連度計算方式を用いて絞り込む。

6.4.1 関連度計算方式の利用

関連度計算方式は、候補語を獲得した後に、喩詞の単語と候補語の関連度を用いる。関連度計算方式を用いることで、喩詞または被喩詞の特徴を表す固有の形容詞と、全ての単語に接続されるような形容詞を判別することができる。例えば「赤い」や「青い」は具体物の中でも限定された単語にしか接続しないが、「多い」や「少ない」といった形容詞は、具体物には必ず接続される。こういった語は格フレームによって全て取得するため、雑音も多くなる。この雑音を除去するために関連度計算方式を用いる。

6.4.2 関連度の検証

形容詞と名詞の関連度の分布を検証するため、1つの名詞に対し10個の形容詞を組み合わせそれぞれの関連度を取得した。それらを名詞10語分用意した。すなわち100組の名詞と形容詞の組み合わせの関連度を取得した。1つの名詞に対し組み合わせる形容詞のなかには、関連性のある形容詞を2,3個含めてある。関連性のある形容詞は、本研究室に構築されている感覚・知覚判断システム^[1]に登録されている代表語を使用する。代表語とは、あらかじめ特定の名詞を抽出したもので、代表語には連想される感覚語・知覚語がデータベース化されている。例えば、「林檎」の場合「赤い、丸い」や「鬼」の場合「怖い、恐ろしい」などが登録されている。これらを用いて形容詞と名詞の関連度の検証を行う。表2に関連性のある名詞と形容詞の組み合わせ例と、関連性のない名詞と形容詞の組み合わせの例を示す。また図2に関連性のある名詞と形容詞の組み合わせ24組の関連度の分布、図3に関連性のない名詞と形容詞の組み合わせ76組の関連度の分布を示す。

表2 組み合わせ例

関連のある組み合わせ	関連のない組み合わせ
林檎—赤い	林檎—重い
砂糖—甘い	砂糖—広い
空—青い	空—甘い
鉄—硬い	鉄—眩しい
雪—白い	雪—赤い

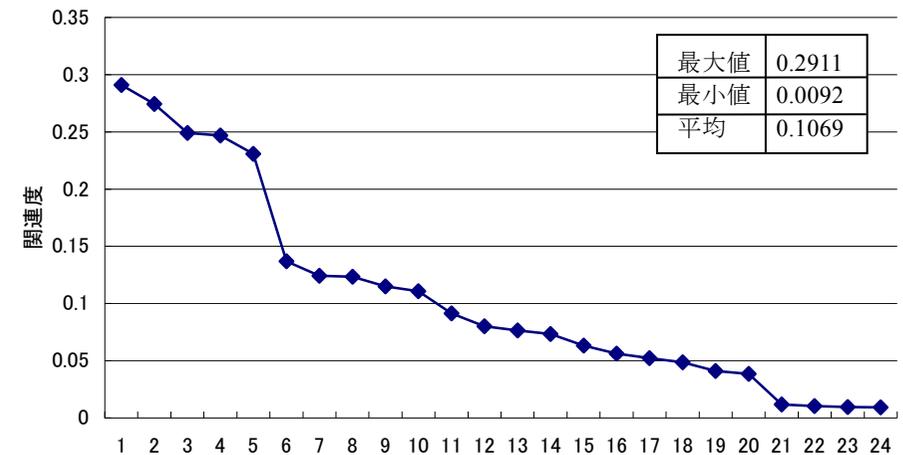


図2 関連のある組み合わせの関連度

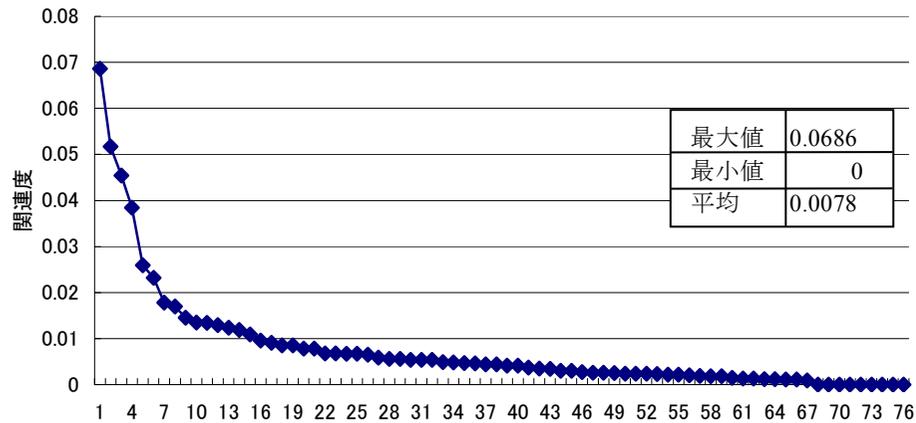


図 3 関連のない組み合わせの関連度

図 2, 図 3 より, 関連のある名詞と形容詞の組み合わせ数み合わせの関連度は, 0.0092 から 0.2911 となる. 一方関連のない名詞と形容詞の組み合わせ数み合わせの関連度は 0 から 0.0078 となった.

6.4.3 重み変更

6.4.2 節の結果より喩詞と候補語の関連度が 0.10 以上の場合は候補語の重みに関連度を掛け合わせる. 0.10 以下の場合は候補語の重みを 0 とする.

6.5 出力方法

出力は候補語に被喩詞を接続する. しかし, 候補語が複数存在する場合, 一番重みの高い候補語のみとする.

6.6

特殊表現とは, 被喩詞の意味的特性だけでは, 変換が困難な表現のことである. 例えば喩詞が「山」である比喻表現, 「山のような仕事」である. 「山」から連想される形容詞として「高い」があり, 「山のような波」の場合「高い波」と解釈できる. しかし, 「山のような仕事」の場合, 「仕事」の量を表し, 「多い仕事」と解釈する. 今回取り扱う特殊表現を表 3 に示す.

表 3 特殊表現

喩詞	例文 (意味)
山	山のような仕事 (多い仕事), 山のような波 (高い波)
嵐	嵐のような日々 (劇的な日々), 嵐のような速度 (速い速度)
夢	夢のような日々 (すばらしい日々)

喩詞が「山」の場合, 格フレームにより接続される語を取得するのではなく, 取得語を「高い」, 「多い」, 「大きい」の 3 つとする. 比喻表現において喩詞に「山」を用いる場合, 「高い」, 「多い」, 「大きい」の 3 つである. これより, もし喩詞が「山」の場合, 喩詞の取得語を「高い, 多い, 大きい」とし, 重みを 0.333 とする.

喩詞が「嵐」の場合は, まず被喩詞の親ノードをシソーラスにより取得し, 被喩詞の語が, 時間ノードに含まれている場合, 「嵐のような」を「劇的な」に変換して出力を行う. 例えば「嵐のような日々」の場合, 「日々」は時間ノードに含まれるため「劇的な日々」として出力を行う. もし被喩詞の語が, 時間ノードに含まれない場合は, 取得語を「速い」とし, 喩詞が「山」のときと同様の処理を行う.

喩詞が「夢」の場合は, 格フレームにより接続される語を取得するのではなく, 取得語を「すばらしい」とする. 比喻表現において喩詞に「夢」を用いる場合, 「すばらしい」の 1 つである. これより, もし喩詞が「夢」の場合, 喩詞の取得語を「すばらしい」とし, 重みを 1.000 とする.

7. 評価

比喻表現「AのようなB」形式の入力文を合計 105 文用意し, システムの評価を行った. 入力文は, アンケートにより取得した.

評価は, 1 つの入力文に対して, 出力結果が正しい場合は 2 点, どちらとも言えない場合は 1 点, 正しくない場合は 0 点とし, 被験者 5 名により評価した. 1 つの入力文に対して, 被験者 5 名の合計点が 7 点~10 点の場合は「○」, 4 点~6 点の場合は「△」, 0 点~3 点の場合「×」とする. 図 4 に評価結果のグラフを示す.

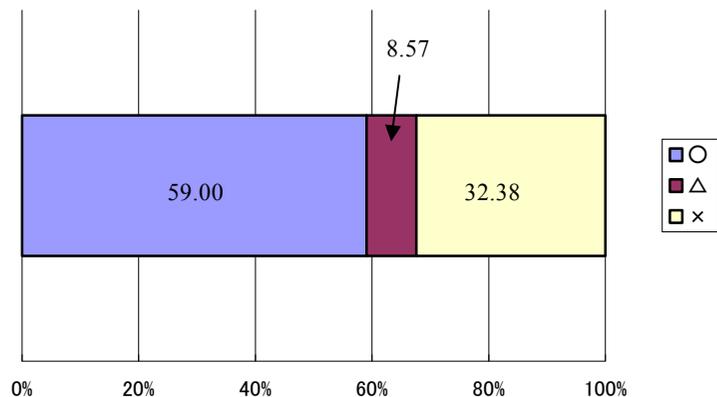


図 4 評価結果

8. 考察

7章の評価結果より、本システムでは精度 59.00%という結果になった。精度が低くなってしまった原因に、形容詞の多義語があげられる。

本システムでは、五感知識ベースと知覚知識ベースにより同意の形容詞を一つにまとめる処理を行っている。そのときに多義の形容詞をうまく処理できていない。例えば、「パソコン」という語から格フレームにより「うまい」という語を取得できるが、五感知識ベースにより「美味しい」と変換されてしまう。「うまい」という語は、「パソコン」という語に接続されることで技術的に優れているという意味の「うまい」で用いられるが、この場合「美味しい」と同じ意味の「うまい」と判断され、取得語の「うまい」を「美味しい」と変えてしまった。形容詞の多義語の処理を行う必要がある。

また被験者 5 名による評価の後、出力結果が「○」以外の 46 文中の候補語上位 5 件に正解と思われる出力があるかを質問したところ、15 文に候補語上位 5 件に正解と思われる語が含まれているという回答を得た。

入力文「火のような夕焼け」の場合、「美しい夕焼け」と出力される。しかし候補語上位 5 件のなかに正解と思われる「赤い」が含まれている。比喩表現は喩詞の特徴を顕著に表すため、今回は喩詞と候補語の関連度のみを用いた。しかし、今回の結果

より、被喩詞と候補語の関連度を用いて、バランス良く重み付けを行う必要があると感じられる。よって新たな候補語の重み付け手法を考案する必要があると考えられる。

具体的には、「食べ物」や「生き物」のように単語の種類を識別する方法である。例えば食べ物以外の単語に「美味しい」という語があれば削除するといったものである。単語をいくつかの種類に分類し、接続される特有の形容詞を絞り込むことにより、正確な出力結果を得ることができると考えられる。

8.1 おわりに

本稿は比喩表現のなかの直喩表現である「AのようなB」という表現の意味解釈を目的に、Web から自動構築した大規模格フレームと本研究室に構築されている概念ベースを用いる手法を提案した。関連度計算方式を用いることにより、直喩表現の文の意味の通る自然な文に変換することができた。

本稿による直喩表現の文を意味の通る自然な文に変換する手法の提案によって、コンピュータが話者の伝えたい内容を正確に判断でき、コンピュータとの自然な会話に繋がると考えられる。

謝辞 本研究の一部は、科学研究費補助金（若手研究（B）21700241）の補助を受けて行った。

参考文献

- [1]米谷彩, 渡部広一, 河岡司, “常識的知覚判断システムの構築”, 第 17 回人工知能学会全国大会論文集, 3C1-07, 2003.
- [2]河原大輔, 黒橋禎夫, “高機能計算環境を用いた Web からの大規模格フレーム構築”, 情報処理学会自然言語処理研究会資料, 2006-NL-171-12, pp.67-73, 2006.
- [3]岩山真, 徳永健伸, 田中徳積, “比喩を含む言語理解における顕現性の役割” 人工知能学会誌, Vol.6, No.5, pp.674-680, 1991.
- [4]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計算方式”, 自然言語処理, Vol.13, No.1, pp.53-74, 2006.
- [5]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.
- [6]NTT コミュニケーション科学研究所監修, 「日本語語彙体系」, 岩波書店, 1997.