

平成22年6月2日現在

研究種目：基盤研究（B）  
 研究期間：2007～2009  
 課題番号：19330156  
 研究課題名（和文） 比喩理解と比喩生成の心理学的メカニズムに関する実験的・理論的研究  
 研究課題名（英文） An experimental and theoretical study on psychological mechanism of metaphor understanding and metaphor generation

研究代表者  
 中川 正宣（NAKAGAWA MASANORI）  
 東京工業大学・大学院社会理工学研究科・教授  
 研究者番号：40155685

研究成果の概要（和文）：大規模言語データの統計解析に基づき、「AのようなB」という形式で表現される比喩（直喩）の比喩理解と比喩生成に関する理論モデル（コンピュータ上に実装される計算モデル）を構築する。さらに、比喩理解と比喩生成モデルのシミュレーション結果と心理学実験結果を比較し、モデルの心理学的妥当性を検証する。これらの手続きをとうして比喩理解・比喩生成の心理学的メカニズムを理論的かつ実験的に明らかにした。

研究成果の概要（英文）：Theoretical (computational) models concerning comprehension and generation of metaphors, in form of “A like B”, were constructed based on statistical analysis of large-scale language data. Furthermore, the validity of the models was verified by comparing between simulation results of the models and psychological experiment results. Through these procedures, the psychological mechanism of metaphor comprehension and metaphor generation was elucidated theoretically and experimentally.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	5,700,000	1,710,000	7,410,000
2008年度	4,200,000	1,260,000	5,460,000
2009年度	3,500,000	1,050,000	4,550,000
年度			
年度			
総計	13,400,000	4,020,000	17,420,000

研究分野：社会科学

科研費の分科・細目：心理学・実験心理学

キーワード：言語、比喩理解、比喩生成

## 1. 研究開始当初の背景

比喩理解過程を説明する理論モデルとしては、Ortony(1979)のSalience imbalance modelを採用し、モデル構築を行う。この理論では、直喩は「喩える語」との関連が強く「喩えられ

る語」との関連が中程度の属性値が、比喩文中の「喩える語」により顕在化することによって理解が成立すると説明されている。「パレリーナは蝶のようだ」という比喩表現を例として説明すると、「蝶」という概念と関連

が強く、「バレリーナ」という概念との関連が中程度である、「ひらひらした」という属性値が、比喩文の「喩える語」である「蝶」によって顕在化されることで「バレリーナは蝶のようだ」という比喩は理解される。

Ortony (1979) の理論に基づいて作成された比喩理解モデルの計算モデルとしては概念変化モデル (岩山・徳永・田中 1991) がある。しかし、このモデルでは前提条件として、概念が「属性」、「属性値」、「属性値確率」(例: {りんご: 色 {赤い, 0.8} {緑色の, 0.15} {黄色の, 0.05}}) で表されている必要がある。心理評定法を用いて概念と属性値の関連強度を測定し、その結果に基づくモデル (内海 2000、Nakagawa et al. 2004 他) もいくつか作成されている。しかし、被験者の負担を考慮すると、一般的に用いられる比喩を網羅するようなモデルを心理評定法にのみ基づいて構築することは事実上困難である。上記のようないくつかの課題を解決するため本研究では人間が使用する比喩を網羅するモデルの構築を前提に、このような知識構造として、言語統計解析を用いて推定された概念と属性値の関連強度を用いる。また、ダイナミックな属性値間の相互関係を含め、人間が行う比喩理解過程を質的にもより正確に表現するため、ニューラルネットワークを用いた比喩理解モデルの構築を行う。さらに、モデルのシミュレーション結果と心理学実験の結果を比較することで、構築したモデルの心理学的妥当性の検証を行い、比喩理解のメカニズムとはどのようなものであるかを明らかにする。

比喩の生成過程を扱った心理学的検討としては、久野(1996)や楠見(2003)による知見がある。久野(1996)は、比喩の題材となる名詞の親近性やイメージ性が、比喩生成の困難さや理解容易性、良好度などの自己評価に影響を与えることを示した。これらの研究は、主に心理学実験に基づいている。しかし、膨大な言語知識についていちいち同様の心理学実験を行うことは現実的に困難である。

すなわち、これまでに提案された比喩生成システムの多くが抱えている問題、そして比喩生成メカニズムの解明を困難にしている問題は、人間の膨大な言語知識がいかに表現され、利用されるか、という点である。本研究では、まず、確率的言語知識構造をもとにしたニューラルネットワークを構築し、「形容詞+名詞」という入力から「Aのような名詞」という比喩表現を生成する計算モデルを構成する。

さらに、この場合も比喩理解のモデルと同様、モデルのシミュレーション結果と心理学実験の結果を比較することで、構築したモデルの心理学的妥当性の検証を行い、比喩生成のメカニズ

ムとはどのようなものであるかを明らかにする。

## 2. 研究の目的

本研究の目的は、大規模言語データの統計解析に基づく心理学実験と理論モデル(コンピュータ上に実装される計算モデル)を用いて、比喩理解と比喩生成の認知的メカニズムを量的にも質的にも十分、明らかにすることである。

人は文学的表現に限らず、科学的記述や日常生活での会話においても、様々な種類の比喩を使用している。例えば、車を運転することを「ハンドルを握る」と表現したり、とても疲れているときに「死にそうに疲れている」と言ったりする。「AのようなB」、「AはBのようだ」という形式で表される直喩に限定しても、「雲のような犬」、「雪のような肌」、「バレリーナは蝶のようだ」などの様々な表現を、人はごく日常的に理解し、生成する。そのため比喩理解・生成を説明する理論は、人間が比喩理解・生成において扱う多くの概念を十分に網羅することで、人間が行う比喩理解・生成過程を量的に説明できることが必要である。さらに、心理学的妥当性を確保することにより、人間が行う理解・生成過程を質的に説明できることも望まれる。本研究では、直喩(特に「AのようなB」と表現される比喩)について、人間が行っている理解・生成過程を量的にも質的にも表現することができる理論モデルを構築する。具体的には、大規模言語データの統計解析を用いて作成された確率的知識構造に基づいてコンピュータ上に計算モデルを構築する。さらにモデルのシミュレーション結果と心理学実験の結果と比較し、そのモデルの心理学的妥当性を検証する。最終的には、これらの実験的・理論的方法を通じて、比喩理解と比喩生成の認知的メカニズムを量的にも質的にも十分、明らかにする。

## 3. 研究の方法

### (1) 言語統計解析

まず基本的言語データの統計解析から確率的な概念構造を構築する。

言語統計解析では、Latent Semantic Analysis (LSA) (Deerwester et. al, 1990)がしばしば用いられているが、共起頻度に0が多い場合、データスパースネス問題が発生する。また機能語に対する重みが大きくなってしまいう問題点もある。機能語の問題に対する対処方法として tf-idf 法 (文書を特徴付けるような高頻度語を重み付ける方法) や

stop word list (分析対象としない単語のリスト) を用意するなどの方法が挙げられるが、これらはいずれも事前に人手による恣意的な設定を必要とする。

これらの問題を解決する方法として、当該研究では Pereira のモデル(1993)や Probabilistic Latent Semantic Indexing (PLSI) (Hofmann 1999) と同型の言語統計解析 (Kameya & Sato 2005) を用いる。この言語統計解析では、単語  $N_i$  (名詞) と単語  $A_j$  (形容詞) の共起確率がいくつかの潜在意味クラスを介在して決定されると仮定し (図3参照)、潜在意味クラス  $C_k$  の確率 ( $P(C_k)$ )、潜在意味クラスが与えられたときの単語  $N_i$  の条件付確率 ( $P(N_i | C_k)$ )、潜在意味クラスが与えられたときの単語  $A_j$  の条件付確率 ( $P(A_j | C_k)$ ) の値を、EM アルゴリズムを用いてデータの対数尤度  $P(D | \theta) = \sum_{i,j} N(N_i, A_j) \log P(N_i, A_j)$  ( $\theta$  は  $P(C_k)$ 、 $P(N_i | C_k)$ 、 $P(A_j | C_k)$  をまとめたパラメータベクトル、 $N(N_i, A_j)$  は名詞  $N_i$  と形容詞  $A_j$  の共起頻度) が最大になるように推定する。EM アルゴリズムでは E-step と M-step を対数尤度が収束するまで繰り返すことにより  $\sum_i P(N_i | C_k) = 1$ 、 $\sum_j P(A_j | C_k) = 1$  の制約を満たしながらパラメータの推定を行う。この解析法は統計理論、情報理論に基づいたものであり、LSA の抱える問題点を回避できる。

本研究では、毎日新聞10年分 (1993年~2002年) から、CaboCha (工藤・松本 2002) を用いて作成された形容詞 (2736語) と名詞 (14807語) の係り受け頻度データを分析に用いる。このように、新聞コーパスに対して言語統計解析を行うことで、人間が保持するであろう確率的知識構造を作成する。

次に、言語統計解析の結果と人間の言語知識の分類との対応を心理学実験で検討する。具体的な方法としては、「形容詞 3 語+名詞 1 語」からなる名詞句 (例:「かわいく幼くいたいけな子ども」) から連想される単語を回答させる記憶連想課題を被験者に課した。この題材となる句には 3 種類のバージョンを用意しており、(1) 全ての形容詞、名詞が共通の潜在意味クラス 1 つと強く結びつく語で構成された名詞句 (以下、これを 1 クラス題材と呼ぶ)、(2) 形容詞 3 語が共通する 1 つの潜在意味クラス、名詞は別の潜在意味クラスと強く結びつく語からなる名詞句 (2 クラス題材)、(3) 各単語がそれぞれ別々の潜在意味クラスと強く結びつく語からなる名詞句 (4 クラス題材) の 3 種類の題材が被験者に提示された。これらの題材は、種類によって関連する潜在意味クラスの数異なっている。もし、潜在意味クラスに基づく分類が、人間の

言語知識上の分類と対応するのであれば、連想の刺激となる名詞句と関連する潜在意味クラスの数が多いほど、連想し得る単語のバリエーションも限られ、回答も容易になると予想される。名詞句と関連する潜在意味クラスの数が多くなれば、それだけ多様な単語を連想させてしまい、回答が困難になると考えられる (図 4)。回答のしやすさ、という観点で見れば、潜在意味クラスの数が多い方が有利であるが、回答の斬新さという観点から見た場合、4 クラス題材の方が、多様な単語を想起させる分、ありきたりではない斬新な回答が出現する可能性も高くなる。これに対し、1 クラス題材は連想の範囲が狭い分、ありきたりな回答が出現しやすくなる。

## (2) 比喩理解モデルの構築

比喩理解においては、概念と属性値がどのような関連強度を持つかという情報が重要である。そこで本研究の理論モデルでは、これらの関連強度を、言語統計解析結果を用いて推定する。言語統計解析によって推定された確率 ( $P$  (潜在意味クラス)、 $P$  (名詞 | 潜在意味クラス)、 $P$  (形容詞 | 潜在意味クラス)) から、Bayes の定理を用いて名詞が与えられたときの形容詞の条件付き確率  $P$  (形容詞 | 名詞) を計算する。これらの値が比喩表現で用いられる概念 (名詞) と属性値 (形容詞) の関連強度を表していると仮定する。モデルで扱う属性値は、比喩文 (例: 少女のようなメロディー) で用いられている概念 (少女、メロディー) が与えられたときの形容詞の条件付確率 ( $P$  (形容詞 | 概念)) の値に基づき、「喩えられる語 (メロディー)」と関連が強く「喩える語 (少女)」との関連が中程度の形容詞 (勇壮な、流麗な、雄大な) と、「喩える語 (少女)」と関連が強く「喩えられる語 (メロディー)」との関連性が中程度の形容詞 (幼い、元気な、可愛い) を選択する。ニューラルネットワークモデルに基づく比喩理解モデルのアーキテクチャを上述の例に即して図5に示す。ニューラルネットワークとは、人間の脳の中の神経細胞 (ニューロン) が互いに結合して情報処理を行う働きを模擬したノードを多数配置し、互いに結合させることによってコンピュータ上で演算処理を行う数理的モデルである。この比喩理解モデルは概念ノードと属性値ノードの2種類のノードから構成されている。概念ノードは入力ノードであり、各ノードは比喩で用いられている概念 (「喩えられる語 (メロディー)」、「喩える語 (少女)」) を表している。また、属性値ノードは出力ノードであり、各ノードは属性値 (勇壮な、流麗な、雄大な、

幼い、元気な、可愛い)を意味する。

### (3) 比喩理解モデルの心理学的妥当性の検証

モデルでのシミュレーションと対応する題材の属性値について測定する。たとえば「メロディー」のイメージ、「少女」のイメージ、比喩文で表現された「少女のようなメロディー」のイメージの各々が持つ属性値、「幼い」「可愛い」「元気な」「勇壮な」「流麗な」「雄大な」という属性値の強度を、心理評定法を用いて測定する。測定は質問紙形式で行い、尺度は7件法(1(全くあてはまらない)~4(どちらでもない)~7(非常にあてはまる))を用いる。

最後に、モデルのシミュレーション結果と比喩理解の心理学実験結果とを比較し、モデルの心理学的妥当性を検証する。最後に、これらの手続きをどうして比喩理解の基本メカニズムを理論的かつ実証的に明らかにする。

### (4) 比喩生成モデルの構築

確率的概念構造を用いて、比喩生成の基本モデルを構築する。

確率的言語知識構造をもとにしたニューラルネットワークを構築し、「形容詞+名詞」という入力から「Aのような名詞」という比喩表現を生成する計算モデルを実現する。モデルの構造としては、入力層、中間層、出力層の3層からなる階層型のニューラルネットワークになっており、入力層、出力層の各ノードはそれぞれ単語(名詞、形容詞)に対応付けられる。残る中間層の各ノードは、言語統計解析上での潜在意味クラス、あるいは人間の概念カテゴリに対応付けられる。各層のノード間を結ぶリンクの結合強度には、言語統計解析から得られた、単語と潜在意味クラス間の条件付確率を用いている。

### (5) 比喩生成モデルの心理学的妥当性の検証

比喩表現を回答させる比喩生成実験を実施する。上記記憶連想実験との違いは、題材として提示した名詞句(かわいく幼くいたいけな子ども)に対して「AのようなB」という比喩を回答するよう教示する。

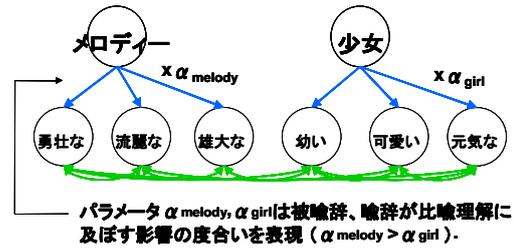
## 4. 研究成果

### (1) 比喩理解モデル

言語統計解析結果に基づき推定された条件付確率( $P(\text{属性値} | \text{概念})$ )に基づいて比喩理解過程を表現するニューラルネットワークモデルを構築した(図1)。モデルは、概念ノードと属性値ノードからなっており、概念ノードの各ノードは比喩文における喩辞、被喩辞を表現し、属性値ノードの各ノードは一つの属性値を表現する(図1参照)。属性値ノード間には結合加重が存在し、比喩理解過程における属性値間の相

互作用を表現する。

モデルシミュレーションでは、入力値の違いにより、被喩辞、喩辞と属性値との関連強度



- 概念(「少女」「メロディー」)から各属性値への結合強度は、概念が与えられたときの属性値の確率( $P(\text{属性値} | \text{概念})$ )で推定される
- 属性値ノード間の結合強度は、属性値間の相関係数で推定される(比喩理解過程における、属性値間の相互作用を表現)

図1: 比喩理解モデルのアーキテクチャ

度と、比喩文が与えられたときの被喩辞と属性値との関連強度を出力する。被喩辞と属性値の関連強度から比喩文が与えられた時の被喩辞と属性値との関連強度への変化により比喩理解を表現する。シミュレーション結果と心理学実験結果を比較したところ、上述の変化に関し同様の变化傾向を得ることができ、属性値間の相互作用を持つモデルの妥当性を検証した。

### (2) 比喩生成モデル

言語統計解析により推定された潜在クラスを用いて、比喩生成の計算モデルを構築した(図2)。モデルは3層からなっており、入力層の各ノードは形容詞または名詞を、中間層の各ノードは潜在クラスを、出力層の各ノードは名詞を表現する。

シミュレーション結果と心理実験結果ともに、題材の属する潜在意味クラスの数が少な

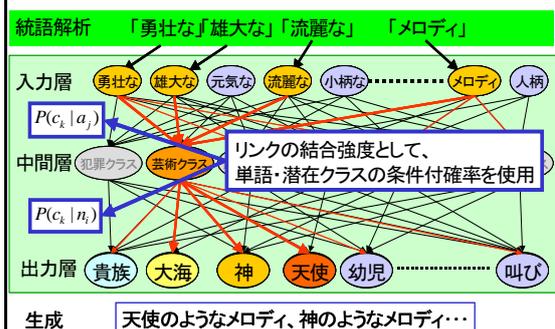


図2: 比喩生成の計算モデル

ければイメージしやすいが、斬新さの乏しい比喩が生成され、多くなればイメージしにくい斬新さの高い比喩が生成されるという結果を得ることができた。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 3 件)

- ① 寺井あすか, 中川正宣 (2010) 確率的概念構造に基づく比喩理解の計算モデル—日本語コーパスの統計解析を用いて—。『認知科学』, 日本認知科学会, Vol.17, No.1, 129-142, 査読有
- ② 阿部慶賀, 中川正宣 (2007), 言語統計解析に基づく確率的言語知識構造の構築と心理学実験への応用。『認知科学』, 日本認知科学会, Vol.14, No.1, 91-117, 査読有
- ③ Terai A., Nakagawa M. (2007) "A Neural Network Model of Metaphor Understanding with Dynamic Interaction based on a Statistical Language Analysis: Targeting a Human-like Model", International Journal of Neural Systems, Vol.17, No.4, p265-274, 査読有

[学会発表] (計 8 件)

- ① Kuroda K. (2009.12.4) Pattern lattice as a model of linguistic knowledge and performance, The 23<sup>rd</sup> Pacific Asia Conference on language, information and Computation, Hong Kong
- ② Terai A., Nakagawa M. (2009.9.16) A Neural Network Model of Metaphor Generation with Dynamic Interaction. International Conference on Artificial Neural Networks 2009, Limassol, Cyprus
- ③ 寺井あすか, 中川正宣 (2009.9.11) "特徴間の相互作用を持つ比喩理解の計算モデル: 日本語コーパスの統計解析を用いて"、日本認知科学会第 26 回大会発表論文集, 慶応義塾大学
- ④ 寺井あすか, 中川正宣 (2008.9.7) "2 段階からなる比喩理解の計算モデルの構築"、日本認知科学会第 25 回大会発表論文集, 同志社大学
- ⑤ Terai A., Nakagawa M. (2009.7.30) Experimental Examination of Feature Emergence in Metaphor Understanding, the 31st Meeting of the Cognitive Science Society, Amsterdam, Neatherlands

- ⑥ Terai A., Nakagawa M. (2008.9.3) A Corpus-based Computational Model of Metaphor Understanding Incorporating Dynamic Interaction. International Conference on Artificial Neural Networks 2008, Prague, Czech
- ⑦ Terai A., Nakagawa M. (2007.9.12) "A Computational Model of Metaphor Understanding Consisting of Two Processes", International Conference on Artificial Neural Networks 2007, Porto, Portugal
- ⑧ 寺井あすか, 中川正宣 (2007.9.5) "複数の言語コーパスを用いた比喩理解の計算モデル"、日本認知科学会第 24 回大会発表論文集, 成城大学

[図書] (計 1 件)

- ① 中川正宣, 寺井あすか, 阿部慶賀 (2007), 第 17 章 比喩理解と比喩生成のニューラルネットワークモデル. 楠見孝(編) メタファー研究の最前線, ひつじ書房, 345-367

## 6. 研究組織

(1) 研究代表者

中川 正宣 (NAKAGAWA MASANORI)  
東京工業大学・大学院社会理工学研究科・教授  
研究者番号: 40155685

(2) 研究分担者

黒田 航 (KURODA KOU)  
独立行政法人情報通信研究機構・知識創成コミュニケーション研究センター・長期専攻研究員  
研究者番号: 30425764

中本 敬子 (NAKAMOTO KEIKO)

文教大学・教育学部・講師  
研究者番号: 50329033

(3) 連携研究者

前川 眞一 (MAEKAWA SHINICHI)  
東京工業大学・大学院社会理工学研究科・教授  
研究者番号: 70190288

赤間 啓之 (AKAMA HIROYUKI)

東京工業大学・大学院社会理工学研究科・准教授  
研究者番号: 60242301

寺井 あすか (TERAI ASUKA)  
東京工業大学・大学院情報理工学研究科・21 世  
紀 COE 研究員  
研究者番号：70422540