

# 連想概念辞書を用いたメトニミー意味理解システムの構築

須賀智秋

慶應義塾大学大学院

政策・メディア研究科

〒252-8520 神奈川県藤沢市遠藤 5322

[sugat@sfc.keio.ac.jp](mailto:sugat@sfc.keio.ac.jp)

石崎俊

慶應義塾大学

環境情報学部

〒252-8520 神奈川県藤沢市遠藤 5322

[ishizaki@sfc.keio.ac.jp](mailto:ishizaki@sfc.keio.ac.jp)

## 概要

本研究は、コンピューター上でのメトニミー意味理解の実現を目的とする。連想概念辞書と日本語語彙体系を用いたメトニミー意味理解システムを構築した。意味理解に際しては、喩えられる語の被喩辞の候補を抽出するだけだったこれまでの研究に対し、喩えられる語の被喩辞の抽出と共に、被喩辞と喩える語の喩辞の関係を動詞で表すという新しい手法を考案し、実際に意味理解文を生成することで、先行研究よりも正確で分かりやすい意味理解を実現した。実際に60文の解析を行った結果を、20名の男女に評価してもらい、その結果、正答率85.5%という数値を得て、手法の有用性を確認した。

## 1. はじめに

メトニミー（換喩）とは、「隣接性に基づく比喩」であり、「一升瓶を飲む」「ネクタイがコーヒーをこぼす」などがその例として挙げられる。コンピューターでメトニミー理解を行う数理モデルは幾つか提案されているが、既存の概念データを利用する事を前提としたものは未だ少ない。

本研究では、実データである連想概念辞書を利用したメトニミー理解モデルを提案し、これまで実現されなかった喩辞、被喩辞間の関係を動詞によって表し、両者を結びつけるシステムを実装した。この手法は従来の統計的手法より一歩深い意味理解を実現したといえる。

## 2. メトニミー

メトニミーは、ある対象を指示するために、それと隣接するたとえる語を用いる比喩である。換喩が依拠する主な隣接関係には次のものがある。

(a) 顕著な対象で空間的隣接対象を指す

例えば部分で全体（「ポニーテール」でその髪型の「女の子」）、容器で内容物（「ボトル」でそれに入った「酒」）、場所や建物で機関（「ワシントン、ホワイトハウス」で合衆国政府）を指す。

(b) 顕著な事象で時間的隣接事象を指す

例えば、結果で原因（「涙を流す」で「泣く」）、原因で結果（「杯を傾ける」で「酒を飲む」）を指す。

こうしたメトニミーの理解や生成は、文脈情報と知識（場面やその時間的連続であるスクリプト）

に支えられている。（楠見[1995]、山梨[1988]）

本研究では、以下の隣接関係に基づくメトニミーを扱った。

表1 本研究で扱うメトニミーの隣接関係

(1) 製作者—製品	バッハを聴く。 ポルシェに乗る。
(2) 主体—付属物	白バイが犯人を逮捕した。 赤シャツが笑う。
(3) 機関—責任者	ジーコが勝った！ 大学はそれを認めないだろう。
(4) 容器—中身	一升瓶を飲み干す。 あのレストランはおいしい。

## 3. 連想概念辞書を用いたメトニミー意味理解システム

連想概念辞書を用いたメトニミー意味理解システムを構築した。連想概念辞書の他には、日本語語彙体系【2】を利用した。また、構文解析器、係り受け解析器には、ChaSen【8】、Cabocha【9】を用いた。

### 3.1 連想概念辞書

石崎研究室では人間の連想を基にした連想概念辞書を作成している。具体的には、名詞を刺激語として連想実験を行い、人間が日常利用している知識を連想概念辞書として構造化し、また、刺激語と連想語の2つの概念間の距離の定量化を行っている。本研究では、刺激語一語につき50人分の連想データを元に作成した辞書を用いた。なお、2006年1月12日現在で、刺激語は880語、連想語は26万語である。

連想実験は、オンラインシステムで行い、データは自動的に収集される。被験者は、1つの刺激語に対してそれぞれ「上位概念」「下位概念」「部分・材料」「属性」「類義語」「動作」「環境」の7つの課題に関して連想を行い、連想語を入力する。

### 3.2 メトニミー意味理解システム

本システムは4つの小さなシステムによって構成されている。

図1はメトニミー意味理解システムの概念図である。本研究では以下の手順で解析を行う。

(1) 解析文の主語と述語を抽出する。その上で、動詞の制約充足により、

メトニミーであるかどうかの判断を行う。メトニミーでないと判断された場合は、処理から抜ける。

(2) 述語との制約充足とあわせて、連想概念辞書の連想距離を元に、被喩辞の抽出を行う。

(3) さらに被喩辞の動作概念で喩辞が意味的に満たすものを抽出し、両者の関係を表す動詞とする。

(4) 意味理解文生成部で、喩辞、被喩辞、両者の関係を表す動詞を結びつけ、理解文を出力する。

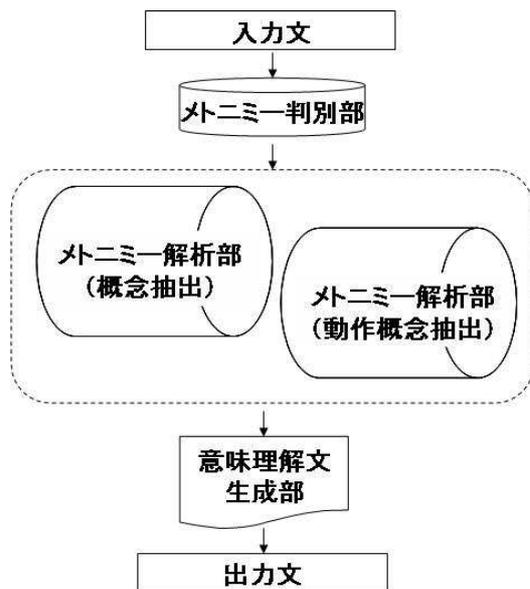


図1 メトニミー意味理解システム概念図

例として、「あのネクタイがコーヒーをこぼした」というメトニミー文を解析することを考える。

#### (1) メトニミー判別部

人間はメトニミーを、名詞と述語の情況内的な辻褄(深谷、田中[1996])があわないことを契機としてメトニミーであると理解する。本研究では、文法的に解析文が成立するかどうかをその判断根拠として利用した。解析文「あの赤いネクタイがコーヒーをこぼした」が入力され、主語「ネクタイ」と述語「こぼす」が抽出される。次に、日本語語彙体系を元に作成したデータベースから、「ネクタイ」と「こぼす」の意味情報を参照し、「こぼす」の主格は「人もしくは動物」とあるという情報から文法的に成立しないという判断をし、メトニミー文であると判別する。

#### (2) 概念抽出部

喩辞(この場合は「ネクタイ」と隣接関係にある概念で、かつ、述語(この場合は「こぼす」)の制約充足を満たす中で、最も連想距離の短いもの(つまり一番連想されやすいもの)を、被喩辞の候補として提示する。その際、連想概念辞書の以下の項目をチェックする。・刺激語がネクタイの部分材料概念にある場合、刺激語がネクタイの環境概念にある場合、ある語の部分材料概念としてネクタイが連想語になっている場合、ある語の環境概念としてネクタイが連想語になっている場合。

#### (3) 動作概念抽出部

「ネクタイ」と「サラリーマン」の間にある動作関係は、「ネクタイ」に対してなされるものであり、かつ、「サラリーマン」がその主格の条件を満たすものである必要がある。そのため、1.連想概念辞書より「ネクタイ」の動作概念を抽出する。2.連想距離の近いものから順に、意味情報を参照しつつ、被喩辞、すなわち「サラリーマン」がその動詞の主格としての条件を満たすかチェックしてゆく。という手順で解析を行う。

また、メトニミーの隣接性の種類により、喩辞と被喩辞の関係があらかじめ特定できる場合、例えば、喩辞が瓶やコップなどの容器の場合などは「<喩辞>に入っている<被喩辞>」など、あらかじめ限定的に動詞を定めておく手法を取った。具体的には以下の表の通りである。さらに、動作概念が抽出できない場合は、格助詞「の」でそれを補った。

表2 喩辞と被喩辞との関係をあらかじめ定めたケース

隣接性		喩辞と被喩辞の関係
容器-中身	瓶、コップなど	<喩辞>に入っている<被喩辞>
	皿など	<喩辞>に乗っている<被喩辞>
企業-製品		<喩辞>が作った<被喩辞>

#### (4) 意味理解文生成部

これまでの喩辞、被喩辞、両者の関係を表す動作概念を結びつけ、意味理解文として出力する。

出力文の述語形態は被喩辞の「状態」を表す表現を用いた。「サラリーマン」の場合は、「ネクタイを締めたサラリーマン」となる。日本語語彙体系の構文情報を利用し、新たに状態を表す場合の構文情報を付与し、それに従い文章を自動生成する仕組みである。

## 4. 実験と考察

### 4.1 実験

本研究で提案した手法の有用性を調べるためにメトニミー文(60文)について解析を行い、被験者にその意味理解文を評価してもらった。60文は表1に基づき全体的にバランスが取れるように、先行研究や文献等から収集した。また、評価基準以下である。

表3 判断基準

◎	自分が想定した意味理解文と全く一緒だった。
○	自分が想定した意味理解文とは、少し違ったが、ほぼ同じだった。
△	自分が想定した意味理解文とは違ったが、この文も「意味理解文として」理解は出来る。
×	元の文の意図を理解できていないとは思えない。

1 厳密に言うと、「企業」がメーカーではない場合は、「作る」という動詞は正確な表現ではないが、第1近似のものとして、今回は「作る」を採用した。正確を期するのであれば、企業概念系を新たに作成し、「メーカーは作る」、「販売会社は売る」など

被験者の評価に基づき重み付けをして (3, 2, 1, 0)、平均値を求めた。また平均評価値 1 (△) 以上を「正解」とし、2 (○) 以上を「人間と同程度の意味理解」とした。

#### 4. 2 実験結果

実験結果 (一部) を表 4、評価結果を表 5 に示す。なお 7 文に関しては本研究の立場では「メトニミーとして扱わない」として評価の結果には入れなかった<sup>2</sup>。結果、全 53 文に対して、正答率 88.7%、人間と同程度以上のもの 57.8%、平均評価値 2.01 という数字を得た。喩辞と被喩辞との関係を動詞で表し、意味理解文を生成するという試みがこれまでなされてこなかったことを考え、また、そのような難解な状況の中で、平均評価値が 2 (すなわち被験者がほとんど自分と同じだと判断したレベル) を越えていることを考えれば、この結果は十分に評価に値するといえよう。

表 4 実験結果<sup>3</sup> (一部)

意味理解文	評価値
(1) 製作者—製品	
村上春樹 (が書いた本) を読む。	2.8
バッハ (が作曲した音楽) を聴く。	2.8
ボルシェ (が作った車) に乗る。	2.6
ヒッチコック (が撮った映画) を観る。	2.59
ドコモ (が作った携帯電話) で話す。	2.1
ティファニー (が作った指輪) を贈る。	1.8
(2) 主体—付属物	
白バイ (に乗った警官) が違反者を逮捕する。	2.9
厚化粧 (を塗った女性) が来る。	2.7
赤シャツ (を着た人) が笑う。	2.6
黒帯 (を締めた有段者) が敵に勝った。	2.55
セーラー服 (を着た女子高生) が歩く。	2.4
眼鏡 (をかけた学者) が説明した。	1.5
パーマ (をかけたおばさん) が歩く。	1.35
(3) 機関—責任者	
大学 (に通っている教授) はそれを認めないだろう。	0.95
(4) 容器—中身	
あのレストラン (が出す料理) はおいしい。	2.8
ペットボトル (に入っている飲み物) を飲む。	2.7

の細かい設定が必要である。

<sup>2</sup> 例えば、「テーブルを拭く」などの慣用的に意味が成立してしまっている、日本語彙大系で動詞の意味情報と被喩辞が矛盾しないもの、などを省いた。

<sup>3</sup> カッコ内がシステムが出力時に付加した内容である。例えば、「村上春樹 (が書いた本) を読む。」は、入力「村上春樹を読む」に対して、「村上春樹が書いた本を読む」という出力をしたということを意味する。

一升瓶 (に入っている酒) を飲み干す。	2.65
上野公園 (に存在している桜) が満開だ。	2.15
頭 (から垂れている髪の毛) を刈る。	1.65
永田町 (に行く政治家) が慌てている。	1.6
お皿 (に乗っている刺身) をたいらげる。	1.3
カルシウム (の牛乳) を飲む。	0.7

表 5 評価結果

カテゴリー	正答率 (評価値1以上)	人間と同程度の出力 (評価値2以上)	平均評価値
判別部ではじかれたもの	0% (0/3)	0% (0/3)	—
製作者—製品	92.9%(13/14)	71.4%(10/14)	2.18
主体—付属物	100%(13/13)	69.2%(9/13)	2.26
機関—責任者	50%(1/2)	0%(0/2)	1.45
容器—中身	95.2%(20/21)	47.6%(10/21)	1.8
全体	88.7% (47/53)	54.7%(29/53)	2.01

#### 4. 3 考察

メトニミーであるのにメトニミーでないと判別されてしまった文が 3 文ある。原因は複合的な表現に対応できなかった (「永田町はパニックに陥った」こと、意味的な矛盾に気付けなかった (ブッシュがイラクを爆撃した) こと、などである。

平均評価値、正答率ともに、「製作者—製品」、「主体—付属物」と比べ、「機関—責任者」「容器—中身」が落ち込んでいるのが分かる。「機関—責任者」では、「責任者」を明確に絞り込むことが出来なかった<sup>4</sup>。「責任者」を絞り込むアルゴリズムが必要である。また被験者に行った評価アンケートから、「機関で責任者を表す」場合 (例えば、「ホワイトハウスは何も言っていない。」の「ホワイトハウス」で合衆国大統領をあらわす) などは、必ずしも「人」ではなく、漠然とした権力などを想起することが多いことが分かった。

「容器—中身」については、喩辞と被喩辞の関係を表す動詞の抽出が困難で、時には unnecessary 場合まで強引に抽出してしまい (頭から生えている髪の毛、一升瓶に入っている蓋、など)、不自然な出力をしてしまったことが、評価値が伸びなかった理由であろう。喩辞と比喩辞の関係の再整理の後、細かいところでのルール設定が必要だと思われる。ただし、「容器—中身」に関して言えば、評価値 1 を超えるものは全体の 95.2% であり、ほとんどのケースで最低限の意味理解は達成したと考えて良い。

一方、「製作者—製品」、「主体—付属物」に関しては、個々のケースで考慮すべき点がありつつも、正答率、平均評価値ともに概ね良い結果であるといえる。

また、被喩辞の抽出の段階で、例えば、「皿を平らげる」を「皿の上の刺身を平らげる」と限定するなど、人間が想定している以上の下位概念

<sup>4</sup> 例えば、「大学」を「大学に通っている教授」と出力した。

まで絞り込んでしまうことで、逆に人間の理解と遠くなってしまふ場合もあった。人間はメトニミーを認知すると、被喩辞を一意に絞り込めるレベルまでの連想を行う。例えば、被験者は「皿を平らげた」というメトニミーの場合、「皿の上にある何らかの料理」と捉え、それ以外に判断材料がない場合は、「サラダ」、「ピラフ」、「チャーハン」、「刺身」など「料理の下位概念」のことは考えなかった。与えられた条件では、それ以上の絞り込みは不可能だからである。

ところが、本システムは、連想距離によって、さらにその先の絞り込みが可能であるので、被験者が「皿を平らげる」という一文だけでは連想しない「刺身」までたどり着き、意味理解文として出力を行った。そこにズレが生じたのである。(結果、評価値は1.3であった)

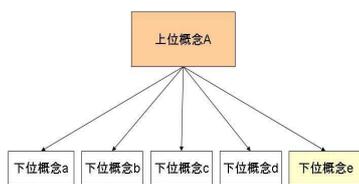


図2 上位概念と下位概念

上位概念Aで連想が止まっていた被験者には、下位概念eまでの連想は不自然に感じられ、結果的に高い評価値は得られなかったことが分かる。

逆に、「白バイ」などの顕在性の高いものからは、人間もシステムと同じ「警官」という連想を行った。「バイク」と比較すると分かりやすいが、「白バイ」はより下位の概念までの絞り込みが可能だからである。「黒帯」(＝有段者)も同様である。よって評価値も高くなった(2.9)と考えられる。被験者と本システムの違いは、人間が大きく顕在性に拠り、特に特徴がない場合などは、例えば「何かの料理」など、曖昧な状況で連想を止めていたのに対し、本システムは、状況に限らず、1つの答えを常に見つけ出したところにある。

顕現性を表す数値を顕現値とし閾値を定めることが出来れば、例えば閾値以下の顕現値を持つ概念であれば、概念部の上位概念(「皿」の場合は、「刺身」「カレー」「チャーハン」などの上位概念である「食べ物」)を、被喩辞として出力するなどの工夫が考えられ、評価値や精度の向上が期待できる。

## 5. 課題と展望

本研究では、喩辞と被喩辞との関係を動詞によって表すという新しい手法を提案し、高い成果を上げた。しかし、同時に、喩辞と被喩辞の間に動詞が介在しない場合、喩辞と被喩辞を入れ替えると良い場合などがあるこ

5 「皿」とリンクしている概念は非常に多いが、連想距離はどれも一様に近く突出した概念は見られなかった(その中でシステムは僅差で近かった「刺身」を抽出した)。この状態のことを本研究では「顕現性が低い」とする。逆に「白バイ」(＝警官)のように、その他の概念と比較して突出して距離の短いものがある状態のことを「顕現性が高い」とする。

とが分かった。それぞれの場合をシステムが判断し、より柔軟な意味理解文の出力を目指すことが、今後の課題のひとつとして挙げられる。

また、連想概念辞書は現在、刺激語850語、連想語が約26万語あるが、世の中に存在するメトニミー表現は無数に存在する。将来的に、コンピューターとの自然の会話を見据えるのであれば、連想概念辞書だけでなく、コーパスやweb情報などからの情報をうまく組み合わせ、より精度の高い、広範な意味理解に適用できる理解システムの構築が必要である。喩辞と被喩辞との関係性の絞り込み、また、被喩辞の絞り込み等の際、連想概念辞書とコーパスなどを併用して使うことで、より精度の向上が期待される。

## 6. おわりに

最後に、本研究を総括する。

本研究では、連想概念辞書を用いたメトニミー意味理解システムの構築を行った。1語につき50名の被験者を集め連想実験を行い、固有名詞の連想辞書を作成し、日本語語彙体系などと併用して、意味理解を行った。意味理解に際しては、被喩辞の候補を抽出するだけだったこれまでの研究に対し、喩辞の抽出と共に、被喩辞と喩辞の関係を動詞で表すという新しい手法を取り入れ、先行研究よりも分かりやすく、正しい意味理解を実現した。また、字義通りの文か、メトニミーか、という判定も行い、その正答率は96%であった。実際に53文の解析を行った結果を、20名の男女に評価してもらい、結果、正答率88.7%、また平均評価値2.01(「2」が被験者が「ほぼ自分と同じ意味理解をしている」と判断した数値)という数値を得て、メトニミー意味理解において高い成果を上げるとともに、本手法の有用性を示した。

## 参考文献

- [1] George Lakoff (池上嘉彦,河上誓作,他=訳), “認知意味論”, 紀伊国屋書店, 1987
- [2] 池原悟、宮崎正弘、白井諭、横男昭男、中岩浩己、小倉健太郎, “日本語語彙体系”, 岩波書店, 1999
- [3] 谷口一美, “認知意味論の新展開”, 研究社, 2003
- [4] 山梨正明, “比喩と理解”, 東京大学出版社, 1988
- [5] 内山将夫, “統計的手法による換喩解析” 自然言語処理, 2000
- [6] 楠見孝, “メタファ研究の総括, 21世紀に向けて: 認知心理学の立場から”, 日本認知言語学会論文集, 2002
- 楠見孝, “類似性と近似性: 人間の認知の特徴について” 人工知能学会誌, 2002
- [7] 村田真樹, 山本専, 黒橋禎夫, 井佐原均, “名詞句「AのB」「AB」の用例を利用した換喩解析” 人工知能学会誌, 2000
- [8] ChaSen <http://chasen.naist.jp/hiki/ChaSen>
- [9] Cabocha <http://chasen.org/~taku/software/cabocha/>

## 謝辞

※この研究は総務省委託研究「戦略的情報通信研究開発推進制度(SCOPE)」により実施したものです。