

統計的手法による換喩の解釈

内山 将夫[†] 村田 真樹[†] 馬 青[†]
内元 清貴[†] 井佐原 均[†]

本稿では、換喩を統計的に解釈する方法を述べた。換喩とは、喩える言葉(喩詞)と喩えられる言葉(被喩詞)との連想に基づいた比喩である。たとえば、「漱石を読む」という換喩は、「漱石の小説を読む」というように解釈できる。この場合、喩詞である「漱石」と被喩詞である「(漱石の)小説」との間には、「作者-作品」という連想関係が成立する。本稿では、以下の方針で換喩を解釈することを試みた。

- (1) 「名詞 A , 格助詞 R , 述語 V 」というタイプの換喩が与えられたとき、与えられた喩詞 A から連想される名詞群を求めるためにコーパスを利用する。
- (2) 連想された名詞群のなかから、与えられた視点 (R, V) に適合するような名詞を被喩詞として統計的に選択する。

その結果、コーパスが連想名詞の供給源として有効なことが例証され、かつ、提案手法を用いることにより、喩詞から連想された名詞群の中から、換喩の視点に適合する名詞を被喩詞として選択できることが分かった。また、提案手法による換喩解析の正解率は、厳しい評価を適用した場合には 0.47 であり、緩い評価を適用した場合には 0.65 であった。これらは提案手法が換喩の解析に有効であることを示している。

キーワード: 比喩, 換喩, コーパス, 連想, 視点, 統計

Statistical Approach to the Interpretation of Metonymy

MASAO UTIYAMA[†], MASAKI MURATA[†], QING MA[†],
KIYOTAKA UCHIMOTO[†] and HITOSHI ISAHARA[†]

This paper describes a statistical approach to the interpretation of metonymy. In metonymy, the name of one thing (*the source*) is substituted for that of another related to it (*the target*). For example, in *Souseki wo yomu* (read a Souseki), the source 'a Souseki' is substituted for the target 'a novel written by Souseki.' In this case, they are related by an *Artist for Artform* relation. The method in this paper follows the procedure below in interpreting a metonymy.

- (1) Given a metonymy 'noun A , case-marker R , predicate V ,' nouns related to the source A are collected in a corpus.
- (2) From the collected nouns, a candidate for the target that satisfies the constraints imposed by R and V is selected by applying a statistical criterion.

The method was tested experimentally. It was shown that a corpus is valuable for extracting nouns that are related to a given source and it was also shown that the proposed statistical criterion can select a good target from the extracted nouns. The precision of the experiment, when based on a rigorous judgment, was 0.47 and when based on a less rigorous judgment it was 0.65. The effectiveness of the proposed method has thus been demonstrated.

KeyWords: *figure of speech, metonymy, corpus, association, view, statistical approach*

1 はじめに

本稿では、比喩の一種である換喩を統計的に解釈する方法を述べる。

比喩は大別すると、直喩・隠喩的なものと換喩的なものとに分けられる(李 1990)。まず、直喩・隠喩的な比喩とは、喩えるもの(喩詞)と喩えられるもの(被喩詞)との類似性に基づいた比喩である。たとえば、「あの男は狼のようだ」という直喩、あるいは、「あの男は狼だ」という隠喩は、喩詞である「狼」と被喩詞である「あの男」との間の何らかの類似性(獐猛さなど)に基づいている。ここで、直喩と隠喩との違いは、直喩が比喩であることを言語的に明示するのに対して、隠喩はそのようなことを明示しない点にある。一方、換喩的な比喩とは、喩詞と被喩詞との連想関係に基づいた比喩である。たとえば、「漱石を読む」という換喩は、「漱石の小説を読む」というように解釈できる。この場合、喩詞である「漱石」と被喩詞である「(漱石の)小説」との間には、「作者-作品」という連想関係が成立する(山梨 1988)。

比喩の処理は、検出と解釈の2段階に分けて考えることができる。まず、比喩の検出とは、与えられた言語表現が比喩であるかどうかを判定する処理である。次に、比喩の解釈とは、与えられた言語表現が比喩であるとして、その比喩の非字義的な表現から字義的な表現を求める処理である。たとえば、比喩の検出の段階では、「漱石を読む」が比喩であり、「小説を読む」が比喩でないことを区別する。また、比喩の解釈の段階では、既に比喩であることが分かっている「漱石を読む」という非字義的な表現から、「漱石の小説を読む」という字義的な表現を導出する。

本稿では、直喩・隠喩的なものと換喩的なものとに大別できる比喩のうちで、換喩を対象とする。また、換喩の検出と解釈のうちでは、換喩の解釈を対象とする。なお、本稿の対象をこのようにした理由は、まず、第1に、直喩や隠喩や換喩などは、上述のように、一応区別できるものであるので、それらを別々のものとして、そのうちの一つを研究対象とすることは可能であるからである。次に、換喩の解釈を対象とする理由は、換喩の解釈は換喩のみを考慮すれば実現可能なのに対して、換喩の検出は直喩や隠喩なども考慮しなければ実現不可能なためである。すなわち、換喩を検出するには、まず、比喩を検出し、その後でその比喩が換喩かどうかを検出しなければならないので、換喩検出を直喩や隠喩と別々に研究することは困難であるのに対して、換喩の解釈の場合には、既に換喩が与えられたものとすれば、他の比喩のことは考慮せずに独立に研究できるためである。

本稿では、換喩のなかでも、「名詞 A , 格助詞 R , 述語 V 」というタイプの換喩を対象とする。そして、以下の方針に基づいて、換喩を解釈する。

- (1) 「 A, R, V 」というタイプの換喩が与えられたとき、与えられた喩詞 A から連想される名詞群を求めるためにコーパスを利用する(3章)。

† 郵政省通信総合研究所, Communications Research Laboratory, Ministry of Posts and Telecommunications

- (2) 連想された名詞群のなかから、与えられた視点 (R, V) に適合するような名詞を被喩詞として統計的に選択する (4章).

たとえば、「一升瓶を飲む」という換喩が与えられたとすると、喩詞である「一升瓶」から連想される名詞として「酒, 栓,...」をコーパスから求め、その中から「を飲む」という視点に適合する「酒」を被喩詞として選択する。一方、「一升瓶を開ける」という換喩に対しては、「一升瓶」から連想される名詞群は同じであるが、被喩詞としては「栓」を選択する。

上述の (1) と (2) は本稿の手法を特徴付けるものである。そして、これらは (山本, 村田, 長尾 1998) の方法を発展させたものと考えられることができる。

まず, (1) については、これまでの換喩の研究としては、連想される (名詞とは限らない) 単語群を求めるために、意味ネットワークや規則などを利用したものがある (Iverson and Helmreich 1992; Bouaud, Bachimont, and Zweigenbaum 1996; Fass 1988) が、そのような知識は人手で構築するのが困難であるという欠点がある。それに対して、コーパスを利用すれば、意味ネットワークのような知識を人手で構築する必要はない。そのため、コーパスを利用すれば、相当多くの換喩を解析できる可能性がある。すなわち、コーパスに基づく手法の方が、意味ネットワークなどに基づく手法よりも、広い範囲の換喩を解析できる可能性が高い。

なお, (山本他 1998) は、名詞 A から連想される名詞の候補として、「名詞 A の名詞 B 」における B と、「名詞 A 名詞 B 」における B を用いていたが、本稿では、(i) 「名詞 A の名詞 B 」における B と、(ii) 名詞 A と同一文中に出現した名詞 B とを連想される名詞の候補に用いる¹。(ii) を用いることにより、(山本他 1998) の方法ではカバーできない名詞を連想の候補として利用できることが期待できる。

次に, (2) については、換喩の解釈を絞り込むための情報源として換喩の視点 (R, V) を利用していると考えられる。このような絞り込みは、従来の研究では、意味ネットワークや規則により実現されてきたが、本稿では、コーパスにおける統計情報を利用して実現する。

なお, (山本他 1998) は、換喩の解釈を絞り込むために、与えられた述語の格フレーム (R, V) に適合する名詞のうちで喩詞 A との共起頻度が最大のもを被喩詞として選ぶという方法を用いている。しかし、全ての述語について格フレームが利用できるとは限らないので、本稿では格フレームを利用せず、統計的手法に基づいて被喩詞を選択する手法を提案する。なお、格フレームが利用できる場合には、その格フレームに適合する名詞のみを候補として、本稿で提案する手法を適応すれば良いので、本稿で提案する手法と共に格フレームを利用することは容易である²。

1 (i) における名詞の候補は (ii) における候補に含まれるが、4章で述べる統計的尺度の計算において別扱いを受ける。

2 (山本他 1998) では、本稿と同様に、換喩の解釈のみを対象にしているが、入力される換喩としては、「名詞 A_1 , 格助詞 R_1 , 名詞 A_2 , 格助詞 R_2, \dots , 名詞 A_n , 格助詞 R_n , 述語 V 」を想定している。そして、その入力に含まれる名詞のなかで述語 V の格の選択制限に合致しないものを喩詞と特定し、その喩詞の被喩詞を求めている。たとえば、「私が漱石を読む」という換喩の場合には、「漱石」が「読む」の選択制限を満たさないことを特定し、「漱石」の被喩詞として「小説」を求めている。一方、本稿では、喩詞が特定済みの入力を想定している。つまり、入力としては、「漱石を読む」のようなものを想定している。この点では、(山本他 1998) の方法の方が優れている。このような喩詞の特定は今後の課題で

以下、2章では、換喩の種類と本稿の対象とする換喩について述べ、3章では、喩詞に関連する名詞群をコーパスから求めるときに使う共起関係について述べ、4章では、被喩詞らしさの統計的尺度について述べる。そして、5章において、提案尺度の有効性を実験により調べ、6章で、その結果を考察する。7章は結論である。なお、付録の表5から表9には、提案尺度に基づいて換喩を解釈した結果がある。

2 換喩の種類と対象とする換喩

換喩の種類を区別するものとして、構文的区別、意味的区別、文脈的区別を考える。そして、それぞれの場合において、本稿で対象としている換喩の特徴を述べる。

2.1 構文的区別

構文的な区別とは、換喩の表現形式による区別のことである。

まず、1項だけで、たとえば、名詞だけで、換喩となる場合が考えられる。この例³としては、「コニヤック」という地名が「コニヤック産のブランデー」という産物を表すというものがあ
る。ただし、この例のように、名詞1語だけで換喩という場合は、語源を考えれば換喩ということであり、現在の用法としては、換喩というよりは、語義の多義性であると言えるであろう。

次に、2項関係により換喩の解釈が決まるものがある。これは、たとえば、「名詞 A 、格助詞 R 、述語 V 」という形⁴で、かつ、 A が換喩かどうかは R と V のみに依存するものである。そのような例としては「漱石を読む」がある。この例は、「漱石の小説を読む」と解釈できるが、そのように解釈できる理由は、「漱石」は人であり、「人を読む」ことはできないので、その他の解釈として、蓋然性が高い、「漱石の小説を読む」が選択されると考えられる。

3項関係としては、「名詞 A_1 、格助詞 R_1 、名詞 A_2 、格助詞 R_2 、述語 V 」のようなものが考えられる。この形の換喩の場合には、 A_1 が換喩かどうかは、 R_1 と V だけでなく、 A_2 と R_2 にも依存する。このような例としては、「(観光ブームで) 祇園に笑顔が戻った」のようなものがある。この例では「 $A_1 =$ 祇園、 $R_1 =$ に、 $A_2 =$ 笑顔、 $R_2 =$ が、 $V =$ 戻った」であり、その解釈は、「祇園の人々に笑顔が戻った」というものである。この場合には、 A_1 の「祇園」が喩詞であり、その被喩詞は「(祇園の) 人々」である。ここで、この喩詞と被喩詞との関係は、 A_2 の「笑顔」がなければ成立しないものである。これは、たとえば、「祇園に太郎が戻った」という文では、「祇園という場所に太郎が戻った」という字義通りの解釈となることから分かる。

項数の他の構文的区別としては、換喩において、喩詞がどのような表現形式をしているかがある。上記では、喩詞の表現としては名詞しか考えていなかったが、その他に、「唇を噛むこと」

ある。ただし、本稿の方法に加えて、格フレームを利用できれば、(山本他 1998)と同様の方法を使うことにより、喩詞を特定できる。

³ 本稿における換喩の例は(山梨 1988)または(山本他 1998)からの引用もしくは著者らの作例である。

⁴ 格助詞は項数に含めない。

が「くやしき」を喩えていたり、「バンドの穴が一つ縮むこと」が「やせたこと」を喩えている場合がある。このように、換喩において、喩詞の表現形式は名詞とは限らない。

上記のように、構文的な区別としては様々なものが考えられるが、本稿で対象とするものは、「A,R,V」という単純なものである。

2.2 意味的区別

意味的な区別とは、喩詞と被喩詞との意味的な関係による換喩の区別のことである。

このような関係としては、表1に示すように、「作者-作品」「主体-手段」「容器-中味」「主体-付属物」のようなものがある(山梨 1988)。

表1 換喩における意味的な関係の例

意味的關係	換喩の例	解釈
作者-作品	漱石を読む	漱石の小説を読む
主体-手段	クラリネットを笑う	クラリネットの奏者を笑う
容器-中味	鍋を食べる	鍋の料理を食べる
主体-付属物	詰め襟が歩く	詰め襟の学生が歩く
主体-近接物	ハムサンドが勘定を払う	ハムサンドの客が勘定を払う

本稿での換喩の解釈においては、このような意味的な区別はせずに、単に、被喩詞としての名詞を同定することを目標とする。たとえば、「一升瓶を飲む」については、「一升瓶の酒を飲む」のように、「一升瓶」の被喩詞として「酒」を取り出すことを目標とするが、「一升瓶」と「酒」の間にある「容器と中味」という意味的な関係について推定することは解釈には含めない。その理由は、本稿では、換喩の被喩詞となる名詞をコーパス中から求めるのだが、そのコーパスには、「一升瓶の酒」というような事例はあっても、その事例における「一升瓶」と「酒」との間の意味関係については何の情報もないからである。そのため、そのような意味関係を推定するのは今後の課題とする。なお、そのような意味関係の推定のためには、(Kurohashi and Sakai 1999)の方法が利用できるであろう。

2.3 文脈的区別

ここで述べる文脈的区別とは、ある換喩が発話されたとき、その換喩の解釈を一意に決めるために必要な文脈の程度による区別のことである。

まず、解釈を一意に決めるために、特に文脈を必要としない換喩がある。これは、その換喩の解釈に一般的な知識のみを必要とするようなものである。そのような例としては「漱石を読む」がある。この換喩を解釈するために必要な知識は、「漱石の小説」というものが存在するという知識のみであり、かつ、そのような知識は常識的な知識であるので、この換喩を一意に解

積するためには、特に文脈は必要ない。

一方、一意に解釈するためには、それが発話された状況までも知らないとならない換喩がある。たとえば、「黒がいい」という換喩の場合、洋服売場においては、「黒のスラックスがいい」という意味かもしれないし、囲碁の対局においては、「黒石を持つ人の情勢がいい」という意味かもしれない。

このように換喩の解釈に必要な文脈の程度は様々であるが、本稿が対象とする換喩は、一意に解釈するために文脈が不要な換喩である。

3 喩詞と関連する名詞群を求めるときに使う共起関係

コーパスにおける単語間の共起頻度を利用することにより、互いに意味的に関連がある単語対を抽出できる (Church and Hanks 1989; Brown, Pietra, deSouza, Lai, and Mercer 1992, など)。このことに基づき、喩詞から連想される名詞群を求めためにコーパスを利用する。

本稿では、特に、「の」による共起関係と「同一文内」における共起関係とに基づいて喩詞と関連する名詞群を求める。ただし、「の」による共起関係とは、係り受け関係にある名詞 A と B とが「 A の B 」の形で出現した場合を言う。さらに、 A と「の」による共起関係にある名詞としては、 A に「の」を介して後接する名詞のみを利用し、前接する名詞は利用しない。たとえば、「 A の B 」と「 X の A 」とがあった場合には、 A と「の」による共起関係にある名詞は B のみである。一方、「同一文内」における共起関係では、「の」による共起関係のような非対称性はない。たとえば、上例と同様に、 A と B とが同一文内にあり、 X と A とが同一文内にある場合には、 A の共起名詞としては、 A との相対的な出現順序に関わらず B と X とが抽出される⁵。

ここで、被喩詞としての妥当性という観点から、それぞれの共起関係により求められる名詞群を比べると、一般的に言って、「の」による共起関係による名詞群の方が優れている。なぜなら、「の」による共起関係にある名詞は、「同一文内」における共起関係にある名詞と違い、二つの名詞間に係り受け関係という構文的な関係が成立している。つまり、構文的な関係という観点からは、「の」による共起関係にある名詞の方が、「同一文内」における共起関係にある名詞よりも明らかに強い。そのため、意味的な関係についても「の」による共起関係にある名詞の方

⁵ 「の」による共起関係が非対称であるのに対して、「同一文内」における共起関係が対称である理由は以下の通りである。まず、「の」による共起関係が非対称である理由は、喩詞と被喩詞との間にある連想の非対称性と、「の」による共起関係における出現頻度の非対称性がよく一致すると考えたからである。たとえば、「漱石を読む」という換喩では、喩詞「漱石」と被喩詞「小説」との間には、「漱石」から「小説」を連想するのであって、「小説」から「漱石」を連想するのではないという非対称性がある。そして、「の」による共起関係についても、出現頻度に関して、このような非対称性がある。たとえば、5章の実験に用いたコーパスでは、「漱石の小説」の頻度は14であるが、「小説の漱石」の頻度は0である。このことから、喩詞「漱石」から連想される名詞群を求めるときには、「漱石の B 」における B を利用すれば良いと考えられる。我々は、このようなことが一般にも成立すると予想し、喩詞から連想される名詞群を精度良く求めるために、「の」による共起関係を非対称なものとした。一方、「同一文内」における共起関係については、そのような連想の非対称性は顕著ではないと考えた。更に、以下で述べるように、「同一文内」における共起関係は、共起名詞のカバー率を上げることを目的として利用しているため、なるべく多くの共起名詞を拾うためには対称の方が都合が良いため、対称的な共起関係を利用した。

が強いと考えることができるからである。たとえば、表1では、様々な意味的な関係が「の」による共起関係として表現されている。これは、「の」による共起関係が、意味的な関係が強い名詞と起こることを例示している。なお、「の」による共起関係の有用性は、(村田 長尾 1997)でも確認されている。

一方、「同一文内」における共起関係にある名詞群は、「の」による共起関係にある名詞群よりも数が多い。実際、「の」による共起関係にある名詞群は、「同一文内」における共起関係にある名詞群の部分集合である。そのため、「の」による共起関係でカバーできないような連想についても「同一文内」における共起名詞によりカバーできる可能性がある。

このように、「の」による共起関係と「同一文内」における共起関係とは、互いに、意味的な関連の強さとカバー率において性質が異なる関係であると言える。そのため、これらを区別して喩詞と共起関係にある名詞群を求める。

たとえば、付録の表5から表9には、5章の実験で使われた換喩について、その喩詞と共起関係にある名詞群を載せてある。ここで、それらの表において、「AのB」とあるのは、喩詞と「の」による共起関係にある名詞群であり、「A近B」とあるのは、喩詞と「同一文内」における共起関係にある名詞群である。なお、それらの名詞群は、喩詞との共起頻度が多いものから最大10個が並べられている。ただし、共起名詞の数が10個以下のものについては、全ての共起名詞が共起頻度の降順に記載されている。なお、各名詞の右下の添字は、その名詞と喩詞との、その共起関係における共起頻度である。

これらの表のうちで、表6にある、10番の換喩「顔を剃る」を見ると、喩詞「顔」と、「の」による共起関係にある名詞群としては、「表情、部分、前、輪郭、しわ、大きさ、筋肉、色、アップ、形」があるが、「同一文内」による共起関係にある名詞群としては、「日本、人、前、女性、首相、目、自分、男、手、東京」がある。これらを比べたとき、「の」による共起関係にある名詞群の方が、「顔」との意味的な関係が強い(明瞭である)と言える。このことは、その他の換喩の例についても言えると考えられる。一方、12番の換喩「アデランスが歩く」を見ると、「同一文内」における共起名詞として「かつら、アートネイチャー、東京、新宿、髪、男性、女性、アルシンド、メーカー、大手」があるが、「の」による共起関係にある名詞は存在しない⁶。このことは、「同一文内」における共起関係が、連想される名詞のカバー率を上げるために有用であることを示している。これらのことから分かるように、二つの共起関係は性質が異なるものである。そのため、共起名詞を得るときには、二つの関係を区別しなければならないと言える。

6 この理由の一つは、5.1節でも述べるが、「の」による共起関係が成立するかどうかの判定を厳しくしていることにもある。

4 被喩詞らしさの統計的尺度

本章では、与えられた名詞群から、換喩の被喩詞として適当な名詞を選択するための統計的尺度について述べる。

与えられた換喩が「名詞 A , 格助詞 R , 述語 V 」のとき、 A と関係 Q にある被喩詞を B とすると、「 A, R, V 」は、「 A, Q, B, R, V 」の省略形であると考えられる (山本他 1998)。たとえば、「漱石を読む」という換喩は、「漱石の小説を読む」という字義的な表現の省略形であると考えられる。ただし、「 $A =$ 漱石」「 $Q =$ の」「 $B =$ 小説」「 $R =$ を」「 $V =$ 読む」である。これらの関係を図 1 に示す。

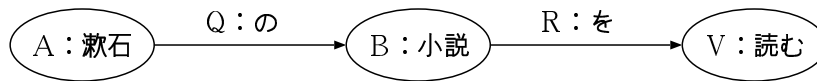


図 1 「 A, Q, B, R, V 」における依存構造

このことを統計的に表現するために、換喩「 A, R, V 」に対して、関係 Q を用いたときの被喩詞 B の尤もらしさを表す尺度を

$$L_Q(B|A, R, V) \doteq \Pr(B|A, Q, R, V) \quad (1)$$

により定義する。そして、 B の被喩詞としての尤もらしさを表す尺度を

$$M(B|A, R, V) \doteq \max_Q L_Q(B|A, R, V) \quad (2)$$

と定義する。尺度 M が大きい名詞ほど、被喩詞として尤もらしい。ここで、 $\Pr(\dots)$ は、そのような事象が生じる確率である。また、 Q は、本稿の場合には、「の」による共起関係か、「同一文内」での共起関係であり、 B は、 A と関係 Q にあるような名詞の一つである。

(1) 式を計算するために、以下のような変形をする。

$$\begin{aligned} L_Q(B|A, R, V) &= \frac{\Pr(B|A, Q, R, V)}{\Pr(A, Q, B, R, V)} \\ &= \frac{\Pr(A, Q, R, V)}{\Pr(A, Q, B) \Pr(R, V|A, Q, B)} \\ &= \frac{\Pr(A, Q) \Pr(R, V|A, Q)}{\Pr(A, Q, B) \Pr(R, V|B)} \\ &\simeq \frac{\Pr(A, Q) \Pr(R, V)}{\Pr(A, Q, B) \Pr(B, R, V)} \frac{1}{\Pr(R, V)} \end{aligned} \quad (3)$$

上式において、4行目から5行目への近似は、形態素間の依存構造を根拠としている。すなわち、「 A, Q, B, R, V 」には、図1のような依存構造を考えることができることから、 A, Q と R, V との関係は密接でないと考えられる。そのため、 A, Q と R, V とは確率的に独立と近似し、4行目から5行目において、 $\Pr(R, V|A, Q, B) \simeq \Pr(R, V|B)$ および $\Pr(R, V|A, Q) \simeq \Pr(R, V)$ とした。

(3) 式の要素式を以下のように定義する⁷。

$$\Pr(A, Q, B) \doteq \frac{f(A, Q, B)}{N_0}. \quad (4)$$

$$\Pr(A, Q) \doteq \frac{f(A, Q)}{N_1}. \quad (5)$$

$$\Pr(B) \doteq \frac{f(B)}{N_2}. \quad (6)$$

$$\Pr(B, R, V) \doteq \begin{cases} \frac{f(B, R, V)}{N_3} & \text{if } f(B, R, V) > 0, \\ \frac{\sum_{C \in \text{Classes}(B)} \Pr(B|C) f(C, R, V)}{N_3} & \text{otherwise.} \end{cases} \quad (7)$$

$$\Pr(B|C) \doteq \frac{f(B)/|\text{Classes}(B)|}{f(C)}. \quad (8)$$

ただし、 $f(\dots)$ は当該事象の頻度であり、 $N_i (0 \leq i \leq 3)$ は確率の総和が1となるように定める正規化定数であり、 $\text{Classes}(B)$ は B が所属する (たとえばシソーラスにおける) 意味的なクラスの集合である。なお、 N_i は定数であり、さらに、(3) 式の $\Pr(R, V)$ についても、換喩「 A, R, V 」が与えられているという条件下では定数である。よって、これらの値は、 $L_Q(B|A, R, V)$ の大小の比較には影響を与えない。そのため、5章の実験では、簡単のため、 $N_i = 1, \Pr(R, V) = 1$ として (3) 式を計算している。

(4) 式と (7) 式とを比べると、(4) 式では、 $f(A, Q, B)$ のみを利用しているため、 A と関係 Q にないような名詞 B は、 $f(A, Q, B) = 0$ であるので、無視される。一方、(7) 式では、 $f(B, R, V) = 0$ であるような名詞 B についても、 B の所属するクラスを利用することにより、0 より大きい確率を付与する。このようにした理由は、まず、(4) 式については、 A と B とは、喩詞と被喩詞という、ある程度強い連想関係にあるはずなので、共起頻度が0であるような名詞は無視しても良いと考えたからである。それに比べて、(7) 式では、 B と V とは必ずしも強い連想関係にあるとは限らないので、 $f(B, R, V) = 0$ であっても、 B が被喩詞として不適当とは限らない可能性がある。そのため、 $f(B, R, V) = 0$ のようなものについても、 B の意味的なクラスを利用することにより、確率を付与することにした。

(8) 式は、意味的クラス C が名詞 B として出現する確率である。これは、もし、名詞 B に多義性がなく、一つのクラス C にしか所属しないとすれば、名詞 B の頻度 $f(B)$ を、クラス C

⁷ 本節で後述する頻度の計数法に従うと、(4) 式と (5) 式については、 $\Pr(A, Q) = \sum_B \Pr(A, Q, B)$ である。しかし、(6) 式と (7) 式については、 $\Pr(B) \neq \sum_{R, V} \Pr(B, R, V)$ である。このように、(6) 式と (7) 式には確率の観点からは不整合がある。この不整合性を失くすためには、 $\Pr(B)$ を $\sum_{R, V} \Pr(B, R, V)$ と定義すれば良いのだが、それを計算するのはコストが高いため、(6) 式のように定義した。

の頻度 $f(C)$ で割れば求めることができるが、名詞 B が多義で複数のクラスに所属する場合には、名詞 B はそれらのクラスに等確率で所属するものとして、 $f(B)$ を $|Classes(B)|$ で割った頻度を $f(C)$ で割ったものを $\Pr(B|C)$ とする。

なお、各事象の頻度は以下のようにして求めた。まず、(4)式における $f(A, Q, B)$ の計数法は、詳しくは5.1節で述べるが、基本的には、コーパス中における、そのような事象を直接計数する。たとえば、 Q が「の」による共起関係の場合は、たとえば、「漱石、の、小説」のような共起を計数するし、「同一文内」による共起関係の場合は、たとえば、「漱石」と「小説」が同一文内で共起した回数を数える。一方、(5)式における $f(A, Q)$ の場合は、 $f(A, Q) \doteq \sum_B f(A, Q, B)$ と定義する。つまり、 $f(A, Q)$ は、 A と関係 Q により共起した名詞 B について、 $f(A, Q, B)$ を足し合わせた値である。更に、(6)式の $f(B)$ の場合には、単に、コーパス中における B の出現頻度を数える。また、(7)式の $f(B, R, V)$ については、詳しくは5.1節で述べるが、基本的には、コーパス中における「 B, R, V 」という形態素列の個数を数える。以上をまとめると、 $f(A, Q, B)$ 、 $f(B)$ 、 $f(B, R, V)$ は、コーパス中における事象を計数し、 $f(A, Q)$ については、 $f(A, Q) = \sum_B f(A, Q, B)$ と定義する、ということである。

また、各クラスの頻度 $f(C)$ は、そのクラスに属する名詞 B の頻度により、次のようにして求めた。

$$f(C) = \sum_{B \in C} \frac{f(B)}{|Classes(B)|}.$$

ここで、 $f(B)$ を $|Classes(B)|$ で割っているのは、(8)式の説明で述べたように、 B が複数の意味的クラスに属する場合に、 $f(B)$ を各クラスに等分割するためである。また、 $f(C, R, V)$ も同様にして求めた。すなわち、

$$f(C, R, V) = \sum_{B \in C} \frac{f(B, R, V)}{|Classes(B)|}.$$

以上、(4)式から(8)式までを利用することにより、(3)式を計算し、(2)式により、名詞の被喩詞らしさを求める。

これらの式の適用例として、たとえば、「森鷗外を読む」という換喩における被喩詞の候補として、「小説」と「戯曲」があるとする。ここで、「 $A =$ 森鷗外」「 $R =$ を」「 $V =$ 読む」である。また、 Q は、「の」による共起関係か、「同一文内」での共起関係であり、 B は「小説」か「戯曲」である。このとき、それぞれの頻度を、 $f(A, Q, B)$ などで分類して表示すると以下の通りである。

$f(A, Q, B)$

$$f(\text{森鷗外, の, 小説}) = 4,$$

$$f(\text{森鷗外, の, 戯曲}) = 2,$$

$$\begin{aligned} f(\text{森鷗外, 同一文内, 小説}) &= 10, \\ f(\text{森鷗外, 同一文内, 戯曲}) &= 2, \end{aligned}$$

$f(A, Q)$

$$\begin{aligned} f(\text{森鷗外, の}) &= 18, \\ f(\text{森鷗外, 同一文内}) &= 1311, \end{aligned}$$

$f(B, R, V)$

$$\begin{aligned} f(\text{小説, を, 読む}) &= 88, \\ f(\text{戯曲, を, 読む}) &= 4, \end{aligned}$$

$f(B)$

$$\begin{aligned} f(\text{小説}) &= 7545, \\ f(\text{戯曲}) &= 781. \end{aligned}$$

ただし、これらの頻度は、5章の実験で用いられたコーパスにおいて、5.1節の方法により頻度を集計した結果である。

次に、 L の大小比較に関係のない $N_0, N_1, N_2, N_3, \Pr(R, V)$ を 1 として、 L を計算すると、(3) 式は、 $f(B, R, V) > 0$ のときには、

$$L_Q(B|A, R, V) = \frac{f(A, Q, B)}{f(A, Q)} \frac{f(B, R, V)}{f(B)}. \quad (9)$$

のように簡単化できるので、

$$\begin{aligned} L_{\text{小説}|\text{森鷗外, を, 読む}} &= \frac{f(\text{森鷗外, の, 小説})}{f(\text{森鷗外, の})} \frac{f(\text{小説, を, 読む})}{f(\text{小説})} \\ &= \frac{4}{18} \frac{88}{7545} = 2.59 \times 10^{-3}, \\ L_{\text{同一文内}(\text{小説}|\text{森鷗外, を, 読む})} &= \frac{f(\text{森鷗外, 同一文内, 小説})}{f(\text{森鷗外, 同一文内})} \frac{f(\text{小説, を, 読む})}{f(\text{小説})} \\ &= \frac{10}{1311} \frac{88}{7545} = 8.90 \times 10^{-5}, \\ L_{\text{戯曲}|\text{森鷗外, を, 読む}} &= \frac{f(\text{森鷗外, の, 戯曲})}{f(\text{森鷗外, の})} \frac{f(\text{戯曲, を, 読む})}{f(\text{戯曲})} \\ &= \frac{2}{18} \frac{4}{781} = 5.69 \times 10^{-4}, \\ L_{\text{同一文内}(\text{戯曲}|\text{森鷗外, を, 読む})} &= \frac{f(\text{森鷗外, 同一文内, 戯曲})}{f(\text{森鷗外, 同一文内})} \frac{f(\text{戯曲, を, 読む})}{f(\text{戯曲})} \\ &= \frac{2}{1311} \frac{4}{781} = 7.81 \times 10^{-6}. \end{aligned}$$

これより,

$$\begin{aligned}
 & M(\text{小説} \mid \text{森鷗外, を, 読む}) \\
 &= \max\{L_{\text{小説}}(\text{小説} \mid \text{森鷗外, を, 読む}), L_{\text{同一文内}}(\text{小説} \mid \text{森鷗外, を, 読む})\} \\
 &= 2.59 \times 10^{-3}, \\
 & M(\text{戯曲} \mid \text{森鷗外, を, 読む}) \\
 &= \max\{L_{\text{戯曲}}(\text{戯曲} \mid \text{森鷗外, を, 読む}), L_{\text{同一文内}}(\text{戯曲} \mid \text{森鷗外, を, 読む})\} \\
 &= 5.69 \times 10^{-4}
 \end{aligned}$$

となり、「森鷗外を読む」の被喩詞としては、「小説」の方が「戯曲」よりも適当となる。この結果は妥当であると考えられる。なぜなら、森鷗外は、確かに戯曲も書いているが、それよりも、「舞姫」「青年」「高瀬舟」などの小説で有名であるからである。

5 実験

実験材料, 実験方法, 実験結果について順に述べる。

5.1 実験材料

換喩 実験に用いた換喩は、(山梨 1988)において例文として採用されているものを、「名詞 A 、格助詞 R 、述語 V 」の形に適合するように変形した 33 例⁸に、著者らによる作例 1 例⁹を加えた 34 例である¹⁰。なお、これらの例では、格助詞としては、「が」と「を」しか出現していないが、その他の格助詞についても提案手法は適用できるし、もっと一般に、任意の 2 項関係についても提案手法は適用できる。

コーパス 単語間の共起頻度を計数するためのコーパスとしては、「CD-毎日新聞」の 91 年度版から 97 年度版の 7 年間分を用いた。このコーパスを自動処理により一文単位に分けたあとで、茶筌 version 2.0b6(松本, 北内, 山下, 平野 1999)により形態素解析し、共起頻度を計数した。

ここで、(4)式における $\text{Pr}(A, Q, B)$ を計算するときに必要な、名詞間の共起頻度を求めるとき、「同一文内」における共起関係については、特別な処理をすることなく、単に頻度を計数

8 実験に用いたのは、2.1節で述べたように、2 項関係により換喩の解釈が決まる例文のうちで、喩詞が名詞であるようなものである。

9 著者らの作例は、換喩解釈における視点の役割を例示するために加えた。

10 (山本他 1998) は 35 例で実験をしているが、それらにおける異なり述語数は 18 であるので、実質上は似たタイプの換喩における実験であると言える。それに対して本稿の実験では、述語の異なり数が 31 であるので、取扱う換喩のタイプ数としては、本稿の実験の方が多くなる。また、(山本他 1998) は (山梨 1988) から換喩の例文を選んでいるのだが、(山梨 1988) にある全ての換喩を選んでいるわけではない。それに対して、本稿では、(山梨 1988) において、完全な文として成立するような例 40 例の中から、2.1節の規準に合致する 32 例 ($32/40=0.80$ 、ここで 32 例のうち 1 例は二つの換喩を含んでいたため、それを 2 例に分け、最終的に 33 例の換喩を得た) を選んだものであるため、本稿の実験は、換喩のタイプを良く網羅するものであると考えられる。

した。しかし、「の」による共起関係については、名詞間に係り受け関係が高確率で成立するような共起関係を同定することを目的として、「非名詞, 名詞 A, の, 名詞 B, 非名詞」という単語列からのみ、AとBの共起頻度を計数した。なお、ここでの名詞には、茶筌にとっての未知語を含む。

また、(7)式における $Pr(B, R, V)$ を計算するときに必要な、「名詞, 格助詞, 述語」の共起頻度を求めるときには、格助詞が「を」の場合には、単に、「名詞, を, 述語」という単語列の生起頻度をもって共起頻度としたが、格助詞が「が」の場合には、「が」だけでなく、「は」「も」「の」を間に狭んだ名詞と述語の連続、すなわち、「名詞, は, 述語」や「名詞, も, 述語」や「名詞, の, 述語」についても、「が」による共起関係の頻度とした。たとえば、「僕が行く」「僕は行く」「僕も行く」という例が、それぞれ1例ずつあった場合には、「僕, が, 行く」の頻度を3にした。このようにした理由は、「を」については、「を」だけで共起頻度が十分に利用できるが、「が」については、「が」だけでは十分な共起頻度が得られないためである。なお、構文解析の結果に基づいた係り受け情報を利用すれば、「が」についても十分な共起頻度が得られる可能性はあるが、今回は構文解析をしなかったため、上記のような措置により、共起頻度情報を得た。

名詞の意味的クラス (7)式を計算するためには、各名詞の意味的なクラスが必要である。そのクラスとしては、分類語彙表増補版(国立国語研究所 1996)における上位3桁の分類番号を用いた。これによるクラスの種類の、全体では、90種であり、名詞に限れば、43種である。なお、本稿では、名詞のクラスのみを用いた。

分類語彙表増補版には約85,000語が記載されているが、これに記載されていない名詞については、もし、その名詞の細分類が、茶筌における、「人名」「組織」「地域」「数」のいずれかである場合には、以下の分類番号を割当てた。

細分類	分類番号	分類語彙表にない名詞例
人名	1.20, 1.21, 1.22, 1.23, 1.24	佐藤, 田中, 鈴木
組織	1.26, 1.27, 1.28	自民党, 筑波大, 日本相撲協会
地域	1.25	大阪, 神戸, 兵庫
数	1.19	一一〇, 一万八千, 七十一

ただし、これらの細分類に該当する名詞については、それに関連する全ての分類番号を割当てた。たとえば、「佐藤」という人名の場合には、「1.20, 1.21, 1.22, 1.23, 1.24」の全てを分類番号として割当てた。つまり、「佐藤」は複数の意味的クラスに所属する多義的な名詞ということである。

また、品詞細分類が上記四つ以外の名詞については、それが分類語彙表で未登録名詞の場合には、その名詞自体を一つのクラスとした。

5.2 実験方法と実験結果

各換喩について、喩詞と、「の」または「同一文内」での共起関係にある名詞群を得て、それぞれの名詞について、(2)式の尺度 M を計算した。

その結果を付録の表 5 から表 9 に示す。これらの中で、表 5 と表 6 は正解例であり、表 7、表 8、表 9 は不正解例である。また、解析結果の一部を表 2 に示す。

表 2 解析結果の一部

1.	一升瓶を飲む
被喩詞	○酒 _{10,1} ^近 (2.6×10^{-3}) ビール _{6,6} ^近 (5.0×10^{-4}) 日本酒 _{7,4} ^近 (3.9×10^{-4}) ぶどう酒 _{1,79} ^近 ワイン _{2,26} ^近 牛乳 _{2,26} ^近 ジュース _{2,26} ^近 ウイスキー _{2,26} ^近 地酒 _{1,79} ^近 薬 _{1,79} ^近
A の B	栓 ₁ ラベル ₁
A 近 B	酒 ₁₀ ビール瓶 ₁₀ 瓶 ₁₀ 日本酒 ₇ 空 ₇ ビール ₆ 手 ₆ 足 ₄ 量 ₄ 程度 ₄
21.	押し入れをかきまわす → 押し入れの中味をかきまわす
被喩詞	△奥 _{17,1} ^{の,*} (2.5×10^{-7}) 天袋 _{5,2} ^{の,*} (2.3×10^{-7}) 天井 _{2,4} ^{の,*} (9.2×10^{-8}) 戸 _{2,4} ^{の,*} 上段 _{2,4} ^{の,*} 中段 _{2,4} ^{の,*} 金庫 _{1,11} ^{の,*} 相手 _{1,393} ^近 和室 _{17,5} ^{近,*} 前 _{3,3} ^{の,*}
A の B	奥 ₁₇ 天袋 ₅ 前 ₃ 天井 ₂ 戸 ₂ 中段 ₂ 上段 ₂ 布団 ₂ 整理 ₂ 中身 ₂
A 近 B	自宅 ₂₅ 奥 ₂₃ 調べ ₂₂ 遺体 ₁₉ 和室 ₁₇ 容疑 ₁₇ 部屋 ₁₆ 疑い ₁₆ 布団 ₁₄ 県 ₁₃
27.	ベルイマンを見る
被喩詞	×息子 _{1,1} ^の (8.1×10^{-4}) 心情 _{1,1} ^の (4.9×10^{-4}) ○映画 _{10,2} ^近 (4.7×10^{-4}) 父 _{1,1} ^の 脚本 _{1,1} ^の 作品 _{6,4} ^近 夢 _{1,26} ^近 秘蔵 _{1,1} ^{つ子,*} シーン _{1,26} ^近 風景 _{1,26} ^近
A の B	息子 ₁ 心情 ₁ 父 ₁ 脚本 ₁ 秘蔵 ₁ 子 ₁
A 近 B	監督 ₁₉ 映画 ₁₀ イングマール ₈ 作品 ₆ スウェーデン ₅ 脚本 ₄ 息子 ₃ 野 ₃ アメリカ ₃ 時代 ₃

表 2 に示すように、各解析結果の第 1 行には、解析対象の換喩が番号と共に記述されている。また、解析対象の換喩の横には「→」で示された換喩の解釈がある。ただし、この解釈は、第 2 行で示される「被喩詞」のなかに、(山梨 1988) で想定されている正解がない場合にのみ記述されている。次に、「被喩詞」の行には、喩詞とのコーパス中での共起名詞が、(2)式の尺度 M について、その降順に上位から最大 10 個並んでいる。ここで、第 1 番目の名詞については、○/△/×が付与されているが、これらの意味は、○は当該の名詞が(山梨 1988) で想定されている正解に合致する例であり、△は(山梨 1988) では想定されていないが意味的には成立可能な例¹¹であり、×は完全な間違い例である。さらに、第 1 番目の名詞が△か×の場合には、それ以降に最初に現れた名詞で被喩詞として適当なものに○か△を付けている。なお、○/△/×の判断は著者らによる。また、各名詞は、

名詞_{頻度, 順位}^{共起関係,*} (尺度 M の値)

のように表現されている。ただし、*や尺度 M の値は特別の場合(後述)にしか表示しない。ここで、まず、「共起関係」とは、(2)式の尺度 M を計算した際に選ばれた共起関係であり、それ

¹¹ △を付けられた例の許容度は様々である。特に、表 8 の 26 番、表 9 の 33,34 番の例は×に近いとも言える。

が「の」の場合には「の」による共起関係であることを示し、「近」の場合には「同一文内」における共起関係であることを示す。また、「頻度」とは、その共起関係における喩詞との共起頻度であり、「順位」とは、その共起関係における、その共起頻度の名詞の順位である。なお、同頻度の名詞は同順位である。たとえば、「名詞₁,名詞₂,名詞₃,名詞₄,名詞₅,名詞₆」の頻度が、それぞれ、「7,5,5,3,3,1」であるとすると、それぞれの順位は、「1,2,2,4,4,6」である。また、名詞の右肩に*が付いているものは、その名詞と換喩の述語との共起頻度が0であったため、(7)式において意味的クラスが利用されたことを示す。たとえば、21番の「押し入れをかきまわす」では「奥をかきまわす」などの頻度が0であったことを示す。さらに、上位3位までの名詞と○または△が付いた名詞については、(2)式の尺度 M の値を括弧内に示す。最後に、「A の B」の行にある名詞群は、喩詞と「の」による共起関係にある名詞群であり、「A 近 B」の行にある名詞群は、喩詞と「同一文内」における共起関係にある名詞群である。なお、これらの名詞群は、共起頻度の降順に上位から最大10個並んでいる。また、各名詞の右下の添字は、その名詞のその共起関係における共起頻度である。ただし、共起頻度が同じ名詞群については、それらを尺度 M の値により降順にソートして表示している。

実験結果の解釈

全体的な精度 (2)式の尺度 M の全体的な精度を調べるために、表5から表9までの被喩詞の第1候補について、○△×を数えた。その結果を表3に示す。この表によると、○のみを正解とする厳しい評価では、正解率は $16/34 \approx 0.47$ であり、○と△を正解とする緩い評価では、正解率は $(16+6)/34 \approx 0.65$ である。これらの数値は、扱う対象が換喩という従来あまり解析の対象とされていない現象であることを考えると、高い値であると言える。すなわち、提案手法は、換喩の解析に対して有効な手法であると考えられる。ただし、これらの値は少数例についてのものであるので、定量的に確定的なことを言うためには、更に大規模な実験が必要である。このことは本節で以下で述べることに同様に言える。なお、実験結果の定性的な解釈については、6章で考察する。

表3 提案尺度の全体的な精度

○	△	×	計
16	6	12	34

二つの共起関係を使った効果 二つの共起関係を使った効果を見るために、「の」による共起関係にある名詞のみを被喩詞の候補とした場合と「同一文内」での共起関係にある名詞のみを被喩詞の候補とした場合とについて、それぞれの共起名詞について尺度 M を計算し、被喩詞の第

表 4 共起関係毎の精度

	○	△	×	計
「の」による共起関係	14	8	12	34
「同一文内」における共起関係	12	8	14	34

1 候補を得た。そのときの○△×の数を表 4に示す。

まず、表 3と表 4とを比べると、二つの共起関係を利用した表 3の結果の方が、それぞれ一つだけの共起関係を利用した表 4の結果のどちらよりも優れていることが分かる。これから、二つの共起関係を同時に用いることが有効であると言える¹²。

次に、表 4における二つの結果を比べると、「の」による共起関係の精度が若干良いことが分かる。一方、「同一文内」における共起関係のカバー率が高いことは、たとえば、表 5の 1 番の換喩「一升瓶を飲む」における「酒」や表 6の 12 番の換喩「アデランスが歩く」における「男性」など、「の」による共起関係に出現していないような名詞でも、「同一文内」における共起関係を利用することにより被喩詞として選択できることからわかる。これらのことは、3章で述べたことを例証している。

意味的クラスの有効性 (7) 式では、被喩詞と述語との共起頻度のスパース性を考慮して、意味的クラスを導入した。そのことは、ある程度は有効であったと考える。その理由は以下の通りである。

まず、表 5から表 9までを見ると、尺度 M の計算において意味的クラスが使用された名詞については、* が右肩に付されている。これから、被喩詞の第 1 候補で、意味的クラスが使用されたものは、5 例あり (13,18,19,21,30 番の換喩)、そのうち、○が 1 例、△が 2 例、×が 2 例である。一方、もし、意味的クラスを使用しない場合¹³には、この 5 例について、△が 3 例、×が 2 例となる。

次に、各換喩ごとに、1 位から 10 位までの候補について、○か△の数の大小を、意味的クラスを使用した場合と使用しない場合とで比べる。たとえば、7 番の「平安神宮が満開」の場合には、1 位から 10 位までを順に示すと、意味的クラスを使用した場合には「○桜、△ハナショウブ、△花、池、柳、枝、幕、○さくら、△植物、花びら」が候補であり、意味的クラスを使用しない場合には「○桜、△ハナショウブ、△花、幕、今年、春、今、人、日、一つ」が候補である。したがって、意味的クラスを使用した場合の方が○か△の数が多い。

このように、各換喩ごとに○か△の数を比較すると、意味的クラスを使用した場合の方が○

12 (山本他 1998)と同様に「名詞 A の名詞 B」における B と「名詞 A 名詞 B」における B とを名詞 A の被喩詞の候補に使った場合には、全体の精度は、○=12, △=10, ×=12, であり、「名詞 A の名詞 B」における B のみを被喩詞の候補とした場合の精度は、表 4における「の」による共起関係の場合と同じであり、「名詞 A 名詞 B」における B のみを被喩詞の候補とした場合の精度は、○=3, △=13, ×=18, であった。このことから、「名詞 A 名詞 B」における B を被喩詞の候補とすることは、有効ではないと言える。

13 $\Pr(B, R, V)$ の計算に (7) 式の $\frac{f(B, R, V)}{N_3}$ のみを利用し、 $f(B, R, V) = 0$ のときには、 $\Pr(B, R, V) = 0$ とした場合。

か△の数が多い換喩は12例であり、意味的クラスを使用しない場合の方が○か△の数が多い換喩は5例である。この結果から、片側検定により符合検定をすると、有意水準2.5%で、意味的クラスを使用した場合の方が、○か△の数が多い換喩が多いと言える。これより、意味的クラスの使用は有効であると言える。

6 考察

6.1 連想名詞の供給源としてのコーパスの有用性

コーパスが連想名詞の供給源として有効なことを第1章で述べた。そのことは、付録の表5から表9までに例証されていると考える。なぜなら、本実験においては、様々な名詞を喩詞として利用したが、それらの大半において妥当な名詞が連想されていることが表5から表9を見れば分かるからである¹⁴。

6.2 視点を考慮した連想

換喩を解析するためには、喩詞から連想された名詞の中から、換喩の(格助詞と述語という)視点に適合する名詞を(被喩詞として)選択する必要があると第1章で述べた。その規準を、(2)式の尺度 M は満たすと言える。そのことを端的に示す例が、1番と2番の換喩の対、および、3番と4番の換喩の対である。まず、1番の「一升瓶を飲む」という換喩と2番の「一升瓶を開ける」という換喩を見ると、同じ喩詞であっても、異なる被喩詞として「酒」と「栓」が選ばれている。次に、3番の換喩「鍋が煮える」と4番の換喩「鍋を食べる」を見ると、それぞれ、「出汁」と「料理」が選択されている。また、その他の例についても、被喩詞の候補として選択されているのは、格助詞を介して述語に継がるような例がほとんどなので、尺度 M は、換喩の視点に適合する名詞を被喩詞として選択していると言える。

6.3 不正解例の分析

不正解となった原因について分析し、それらの誤りを改善する可能性について述べる。

頻度が少ないことによる不正解例

表7には、頻度が少ないことによる不正解例を載せてある。

まず、喩詞の頻度が少ない場合として、17番の「藁草履が来る」がある。これは、喩詞である「藁草履」がコーパス中で1回も出現しなかったため解析に失敗した例である。このように喩詞自体の頻度が少ないような例は本手法では解析できない。このような単語を含む換喩を解

¹⁴ このことを数値的に示すことは難しいが、5.2章の実験における、○と△の数を一応は妥当な連想の数であるとすると、少くとも、 $(16 + 6)/34 \approx 0.65$ は妥当な連想であると言える

析するためには、より大規模なコーパスが必要であろう。ただし、「藁草履」の場合には、「藁草履」の頻度は0であるが、「わら草履」の頻度は0ではない。そのため、このような表記のゆれを吸収することができれば、比較的頻度が少ない被喩詞を含む換喩についても解析できる可能性がある。あるいは、「藁草履」でなく「草履」を喩詞とみなして換喩の解析をすることも考えられるが、これらを試みるのは今後の課題である。

次に、喩詞と被喩詞との共起頻度が少ない例として、18,19,20番の換喩がある。たとえば、19番の「傘が行く」は「傘をさした人が行く」などと解釈できるが、「傘」と「人」などとの共起頻度は、「傘」と「先」とか「下」とか「柄」とかとの共起頻度と比べれば少ない¹⁵。そのため、被喩詞として優先されない。このような例を提案手法で解析することはできない。

ただし、18,19,20番の換喩に限って言えば、被喩詞が全て「人」であると特徴付けることができる。このことは、「人」の場合には、たとえ、喩詞との関連性が低くても被喩詞となりうることを示していると解釈できる。もし、この解釈が正しければ、「人」に類するものが被喩詞の候補としてある場合には、それを優先するようにすれば、これらの換喩を解釈できる可能性がある。ただし、このことを検証するのは今後の課題である。

最後に、被喩詞と述語との共起頻度が少ない例として、21,22,23,24番がある。たとえば、23番の「川が氾濫」は「川の水が氾濫」と解釈できるが、「水」と「氾濫」の共起頻度は、「水路」や「河川」や「ダム」と「氾濫」との共起頻度よりも小さい¹⁶。そのため被喩詞として優先されない。このような例も提案手法で解析することはできない。

ただし、21,22,23,24番の換喩に限れば、述語が場所を対象格として取りうると特徴付けることができる。そして、これらについては、換喩として与えられた入力、現在ではそのまま字義的な表現として通用すると言える。すなわち、これらの換喩は語源的には換喩であっても、現在の用法としては既に換喩ではなく、慣用的に場所を対象格に取り得るといえる。そのため、これらは換喩として解析する必要はないと考える。ただし、与えられた入力を換喩として解釈すべきかどうか決めるためには、その入力が換喩かどうかを検出する必要がある。

解析に文脈が必要な例

表8にある25,26,27番の換喩は、述語の対象格が漠然としているため、被喩詞を決めることができない例である。このような例を解析するためには、2.3節で述べたように、換喩が発話された文脈を考慮する必要がある。

その他の原因による不正解例

表9には、その他の原因による不正解例を載せる。

15 より正確に言えば、(3)式における $\frac{\text{Pr}(A,Q,B)}{\text{Pr}(A,Q)}$ が小さい。

16 より正確に言えば、(3)式における $\frac{\text{Pr}(B,R,V)}{\text{Pr}(B)}$ が小さい。

まず、28,29,30番は、被喩詞の一般化が足りない例である。たとえば、29番の「大阪がげんなり」は「大阪の人々がげんなり」と解釈できるが、被喩詞の候補としては「委員、弁護士、業者,...」となっている。これらの候補はいずれも「人」または「人々」の下位語であるので、これらの候補を適切に一般化できれば、この換喩を解釈できる可能性がある。それと同様なことが28,30番にも言える。

次に、31,32番は、「名詞、が、述語」という共起関係を求めるために、5.1節で述べたように、「名詞、は、述語」などの共起関係も利用したために生じた間違いである。つまり、31番の場合には「ドラマは興奮」という共起関係を利用し、32番の場合には「今日は行く」という共起関係を利用したための間違いである。このようなことを回避するためには、名詞と述語の、「が」や「は」を介した共起頻度をとるときには、その名詞が主格となっている場合についてのみ共起頻度を取らなければならない。そうするためには、構文解析を利用して共起頻度を求める必要があるであろう。なお、現在の方法により共起頻度を求めた場合には、「が」のみを利用した場合の解析結果が、その他のものも利用した場合と比べて良くないことは予備実験で確かめてある。

最後に、33,34番は、換喩を解釈した結果が依然として換喩的な意味を持っている場合である。たとえば、33番の「理論が主張」の解析結果は「党が(理論を)主張」であるが、「党」の背後には、更に「人」が暗黙の内にいると考えられる。このような例を解釈するためには、解析結果が換喩的な意味を持つかどうかを調べ、もしそれが換喩的な意味を持つならば、それをもう一度解釈する必要がある。

7 おわりに

本稿では、比喩の一種である換喩を統計的に解釈する方法について述べた。

そして、換喩のなかでも、「名詞 A , 格助詞 R , 述語 V 」というタイプの換喩を対象とし、以下の方針に基づいて、換喩を解析することを試みた。

- (1) 「 A, R, V 」というタイプの換喩が与えられたとき、与えられた喩詞 A から連想される名詞群を求めるためにコーパスを利用する。
- (2) 連想された名詞群のなかから、与えられた視点 (R, V) に適合するような名詞を被喩詞として統計的に選択する。

その結果、コーパスが連想名詞の供給源として有効なことが例証され、かつ、提案手法を用いることにより、喩詞から連想された名詞群の中から、換喩の(格助詞と述語という)視点に適合する名詞を(被喩詞として)選択できることが分かった。また、提案手法による換喩解析の精度は、扱う対象が換喩という従来あまり解析の対象とされていない現象であることを考えると、高い値であると我々は判断した。

また、実験結果の不正解例を分析した結果、

- 喩詞と被喩詞や被喩詞と述語の共起頻度が少ない例

- 解釈に文脈が必要な例
- 被喩詞の適切な一般化が必要な例

などがあることが分かった。今後は、そのような問題も解決できるように提案手法を拡張していきたい。

謝辞

本稿に対して有益なコメントを下された筑波大学山本幹雄助教授に感謝する。

付録

表 5 正解例 (表 6 に続く)

1.	一升瓶を飲む 被喩詞 ○酒 _{10,1} ^近 (2.6×10^{-3}) ビール _{6,6} ^近 (5.0×10^{-4}) 日本酒 _{7,4} ^近 (3.9×10^{-4}) ぶどう酒 _{1,79} ^近 ワイン _{2,26} ^近 牛乳 _{2,26} ^近 ジュース _{2,26} ^近 ウイスキー _{2,26} ^近 地酒 _{1,79} ^近 薬 _{1,79} ^近 A の B 栓 ₁ ラベル ₁ A 近 B 酒 ₁₀ ビール瓶 ₁₀ 瓶 ₁₀ 日本酒 ₇ 空 ₇ ビール ₆ 手 ₆ 足 ₄ 量 ₄ 程度 ₄
2.	一升瓶を開ける (著者らの作例) 被喩詞 ○栓 _{1,1} ^近 (8.4×10^{-3}) 冷蔵庫 _{2,26} ^近 (5.8×10^{-5}) 箱 _{1,79} ^近 (2.0×10^{-5}) 缶 _{3,17} ^近 瓶 _{10,1} ^近 住宅 _{2,26} ^{近,*} 番屋 _{2,26} ^{近,*} ビール瓶 _{10,1} ^{近,*} 袋 _{1,79} ^近 ワイン _{2,26} ^近 A の B 栓 ₁ ラベル ₁ A 近 B 瓶 ₁₀ ビール瓶 ₁₀ 酒 ₁₀ 空 ₇ 日本酒 ₇ ビール ₆ 手 ₆ ケース ₄ 足 ₄ 事務所 ₄
3.	鍋が煮える 被喩詞 ○出汁 _{1,5} ^の (1.6×10^{-2}) 具 _{1,5} ^の (2.6×10^{-5}) みそ汁 _{4,186} ^近 (5.4×10^{-6}) 芋 _{2,426} ^近 めん _{12,30} ^近 コンニャク _{2,426} ^近 野菜 _{20,11} ^近 魚 _{10,39} ^近 たれ _{1,5} ^{の,*} 料理 _{78,2} ^{近,*} A の B 季節 ₄ 底 ₃ ふた ₂ 材料 ₂ 出汁 ₁ 具 ₁ たれ ₁ 人気 ₁ 柄 ₁ 周囲 ₁ A 近 B なべ ₁₁₁ 料理 ₇₈ 火 ₃₄ 県 ₃₂ 東京 ₂₇ 前 ₂₃ 水 ₂₃ スープ ₂₂ 肉 ₂₂ 神 ₂₂
4.	鍋を食べる 被喩詞 ○料理 _{78,2} ^近 (1.7×10^{-4}) 肉 _{22,8} ^近 (1.3×10^{-4}) シャブシャブ _{2,426} ^近 (6.5×10^{-5}) なべ _{111,1} ^近 ご飯 _{4,186} ^近 うどん _近 刺し身 _近 ごはん _{3,262} ^近 ちゃんこ _{10,39} ^近 スープ _{22,8} ^近 A の B 季節 ₄ 底 ₃ 材料 ₂ ふた ₂ たれ ₁ 具 ₁ 残骸 ₁ セット ₁ 取っ手 ₁ 柄 ₁ A 近 B なべ ₁₁₁ 料理 ₇₈ 火 ₃₄ 県 ₃₂ 東京 ₂₇ 水 ₂₃ 前 ₂₃ 肉 ₂₂ スープ ₂₂ 神 ₂₂
5.	池が枯れる 被喩詞 ○水 _{63,1} ^の (9.8×10^{-5}) 水草 _{17,230} ^近 (7.7×10^{-6}) わき水 _{11,415} ^近 (6.4×10^{-6}) 葉 _{30,85} ^近 下流 _{1,35} ^の 土壌 _{1,35} ^の 井戸 _{25,117} ^近 水源 _{14,299} ^近 井戸水 _{2,2310} ^近 木 _{25,117} ^近 A の B 水 ₆₃ 周り ₂₁ 底 ₂₀ ほとり ₁₄ 搜索 ₁₄ 周囲 ₁₃ 端 ₉ 水面 ₇ 近く ₇ そば ₅ A 近 B 水 ₂₅₃ 県 ₁₉₆ 大阪 ₁₁₈ 前 ₉₈ 近く ₉₆ 調整 ₉₆ 東京 ₉₁ 上曾根 ₈₄ 午後 ₈₂ 公園 ₇₉
6.	湖が満ちる 被喩詞 ○水 _{13,2} ^の (5.5×10^{-6}) 美しさ _{1,34} ^の (3.4×10^{-6}) 汚染 _{5,6} ^{の,*} (1.2×10^{-6}) 青 _{2,17} ^{の,*} 香 _{3,773} ^近 空気 _{4,562} ^近 月 _{19,41} ^近 干満 _{1,34} ^{の,*} 波 _{1,34} ^{の,*} 水質 _{4,9} ^{の,*} A の B ほとり ₁₄ 水 ₁₃ 水位 ₈ 国 ₇ 底 ₆ 汚染 ₅ 面積 ₅ 主 ₅ 水質 ₄ 伝説 ₄ A 近 B 水 ₉₆ 県 ₈₈ 川 ₈₅ 白鳥 ₇₄ 琵琶湖 ₆₉ 海 ₆₂ 森 ₅₁ 日本 ₅₁ 河川 ₄₈ 前 ₄₃
7.	平安神宮が満開 被喩詞 ○桜 _{1,2} ^の (2.1×10^{-3}) ハナシヨウブ _{1,100} ^近 (2.9×10^{-5}) 花 _{2,45} ^近 (5.9×10^{-6}) 池 _{2,1} ^{の,*} 柳 _{2,45} ^{近,*} 枝 _{2,45} ^{近,*} 幕 _{1,100} ^近 さくら _{1,100} ^{近,*} 植物 _{1,100} ^{近,*} 花びら _{1,100} ^{近,*} A の B 池 ₂ 桜 ₁ 祭り ₁ 大鳥居 ₁ ヤエベニダレサクラ ₁ A 近 B 京都 ₄₀ 左京 ₂₂ 写真 ₁₂ 御所 ₁₁ 時代 ₉ 行列 ₇ 記念 ₇ 鳳輦 ₇ 平安 ₇ 午前 ₆
8.	皿を平らげる 被喩詞 ○料理 _{1,3} ^の (3.5×10^{-6}) ステーキ _{5,167} ^近 (2.0×10^{-6}) カツ _{3,351} ^近 (1.5×10^{-6}) スパゲティ _{2,588} ^近 ケーキ _{5,167} ^近 えさ _{1,3} ^{の,*} ごちそう _{1,1158} ^近 大根 _{1,1158} ^近 杯 _{1,1158} ^近 スープ _{14,28} ^{近,*} A の B 前 ₃ 破片 ₃ 料理 ₁ えさ ₁ 枚数 ₁ 展示 ₁ 山 ₁ 羊かん ₁ 取り合わせ ₁ 雨 ₁ A 近 B 料理 ₄₆ 絵 ₃₀ 手 ₂₈ 屋敷 ₂₈ 写真 ₂₇ 水 ₂₇ 食器 ₂₇ 展示 ₂₆ ガラス ₂₅ 前 ₂₄

各解析結果の第 1 行は当該の換喩である。「被喩詞」の行は (2) 式の値で共起名詞を降順に上位から最大 10 個並べたものである。「A の B」および「A 近 B」には、喩詞と、「の」による共起関係および「同一文内」における共起関係にある名詞が共起頻度の降順に上位から最大 10 個だけ並んでいる。詳しくは 5.2 節を参照のこと。

表 6 正解例 (表 5 の続き)

9.	庭を掃く 被喩詞 A の B A 近 B	○落ち葉 ^{1,51} (2.0×10^{-5}) 木立 ^{1,51} (1.6×10^{-5}) 砂 ^{5,11} (7.4×10^{-6}) 水面 ^{1,51} 花びら ^{4,768} 居間 ^{9,283} 廊下 ^{5,607} 池 ^{10,4} 落葉 ^{1,287} ごみ ^{16,127} 手入れ ²⁰ 隅 ¹⁶ 片隅 ¹² 池 ¹⁰ 花 ¹⁰ 木 ⁹ 千草 ⁸ 土 ⁷ 緑 ⁶ 草むしり ⁶ 花 ¹³⁵ 家 ¹³⁵ 自宅 ¹³⁰ 木 ⁸⁵ 県 ⁸⁵ 前 ⁸³ 人 ⁷⁴ 日本 ⁵⁸ 水 ⁵⁶ 自分 ⁵⁴
10.	顔を剃る 被喩詞 A の B A 近 B	○ヒゲ ^{1,82} (5.5×10^{-6}) 頭 ^{300,14} (6.1×10^{-7}) 髪 ^{88,157} (5.9×10^{-7}) 筋肉 ^{10,7} ヘア ^{10,246} 皮 ^{5,16} 汗 ^{4,22} 骨格 ^{3,34} 皮膚 ^{2,48} 骨 ^{4,22} 表情 ⁴⁶ 部分 ²¹ 前 ²⁰ 輪郭 ¹⁸ しわ ¹⁷ 大きさ ¹¹ 筋肉 ¹⁰ 色 ¹⁰ アップ ⁸ 形 ⁸ 日本 ⁸⁰⁵ 人 ⁷³⁷ 前 ⁵³⁴ 女性 ⁴²⁶ 首相 ⁴⁰⁹ 目 ⁴⁰³ 自分 ³⁸⁸ 男 ³⁸⁶ 手 ³⁷⁹ 東京 ³⁷³
11.	頭を刈る 被喩詞 A の B A 近 B	○毛 ^{8,16} (5.6×10^{-6}) 髪の毛 ^{1,63} (3.1×10^{-6}) 髪 ^{1,63} (2.9×10^{-6}) 骨 ^{204,1} 草 ^{6,359} 上 ^{18,9} 芋 ^{1,63} 細胞 ^{1,63} 一部 ^{7,17} 頭髪 ^{4,492} 骨 ²⁰⁴ 回転 ⁴⁹ 片隅 ³⁶ てっぺん ³⁶ 部分 ²⁷ 体操 ²⁶ 切り替え ²² 隅 ²¹ 上 ¹⁸ 後ろ ¹⁷ 死亡 ⁶²³ 前 ⁵¹⁴ 日本 ⁵⁰⁵ 県 ⁴⁶⁰ 人 ⁴⁵³ 午後 ⁴²⁸ 問題 ⁴¹³ 午前 ³⁸⁴ 会社 ³⁷¹ 自分 ³⁵²
12.	アデランスが 悪く 被喩詞 A の B A 近 B	○男性 ^{4,6} (2.8×10^{-6}) 男 ^{1,70} (1.0×10^{-6}) 人間 ^{1,70} (8.5×10^{-7}) 女性 ^{4,6} 人 ^{2,28} 文学 ^{1,70} 子供 ^{2,28} かつら ^{11,1} 成人 ^{2,28} 若者 ^{2,28} かつら ¹¹ アートネイチャー ⁸ 東京 ⁶ 新宿 ⁵ 髪 ⁵ 男性 ⁴ 女性 ⁴ アルシンド ⁴ メーカー ⁴ 大手 ⁴
13.	赤シャツを 嫌う 被喩詞 A の B A 近 B	○やつ ^{1,1} (2.5×10^{-5}) 進行 ^{1,8} (2.1×10^{-6}) 話 ^{2,7} (6.2×10^{-7}) 舞台 ^{1,8} 赴任 ^{1,8} 完結 ^{1,8} 会話 ^{1,8} 仮名 ^{1,8} 辞表 ^{1,8} 制裁 ^{1,8} やつ ¹ 坊っちゃん ⁶ 夏目 ³ 漱石 ³ 野 ³ 校長 ³ 山嵐 ³ 話 ² 進行 ¹ 舞台 ¹ 完結 ¹
14.	厚化粧が 来る 被喩詞 A の B A 近 B	○おばさん ^{1,27} (2.1×10^{-5}) 女性 ^{1,2} (1.7×10^{-5}) マダム ^{1,2} (1.6×10^{-5}) 人 ^{5,2} 裏 ^{2,1} P T A ^{1,2} 仕事 ^{1,27} 昼 ^{1,27} 知人 ^{1,27} 下 ^{1,2} 裏 ² 女性 ¹ マダム ¹ P T A ¹ 下 ¹ イメージ ¹ 顔 ¹ オバタリアン ¹ 女性 ⁶ 人 ⁵ 化粧 ³ 口紅 ³ 大阪 ³ 顔 ³ むかし ² 記事 ² メーク ² 女 ²
15.	三島を 読む 被喩詞 A の B A 近 B	○小説 ^{2,3} (6.0×10^{-4}) 作品 ^{2,3} (1.4×10^{-4}) 戯曲 ^{1,10} (1.3×10^{-4}) 書簡 ^{1,10} 本 ^{8,214} 原稿 ^{9,177} 台本 ^{2,121} 名 ^{1,10} 本書 ^{4,567} 文学 ^{50,13} 郷 ¹ 死 ³ 小説 ² 作品 ² 連打 ² 葬儀 ² 死後 ² 遺族 ² 左前打 ² 戯曲 ¹ 由紀夫 ²⁴⁸ 県 ²²⁰ 静岡 ¹⁶² 大阪 ⁸⁶ 東京 ⁷⁶ 作家 ⁶⁹ 日本 ⁶⁵ 午後 ⁶² 作品 ⁵⁴ 会 ⁵³
16.	マーラーを 聴く 被喩詞 A の B A 近 B	○音楽 ^{2,2} (9.1×10^{-4}) 交響曲 ^{13,1} (3.5×10^{-4}) 歌曲 ^{2,2} (1.3×10^{-4}) 演奏 ^{20,4} 楽想 ^{1,200} 長調 ^{3,53} 楽章 ^{6,19} 歌 ^{6,19} 他 ^{1,4} 難曲 ^{1,4} 交響曲 ¹³ 音楽 ² 歌曲 ² 他 ¹ 難曲 ¹ 名作 ¹ 作品 ¹ 抒情 ¹ 叙情 ¹ 妻 ¹ 交響曲 ⁴⁹ 指揮 ²⁶ 音楽 ²⁵ 演奏 ²⁰ 作曲 ¹² ウィーン ¹² ブルックナー ¹⁰ フィル ⁹ 響 ⁸ 作品 ⁸

表 7 頻度が少ないことによる不正解例

原因	喩詞の頻度が少ない。
17. 被喩詞 A の B A 近 B	藁草履 が 来る → 藁草履を履いた人が来る
原因	喩詞と被喩詞との共起頻度が少ない。
18. 被喩詞 A の B A 近 B	仮説 が 説明 → 仮説の提唱者が説明 ×実証 ^{4,1} (2.5 × 10 ⁻⁵) 域 ^{2,2} (1.8 × 10 ⁻⁵) 裏付け ^{1,4} (6.2 × 10 ⁻⁶) 立証 ^{1,4} 提示 ^{1,4} 検証 ^{1,4} △助教 ^{7,36} もとの ^{1,4} 博士 ^{12,14} 理由 ^{9,25} 実証 ⁴ 域 ² 一つ ² 裏付け ¹ 提示 ¹ 立証 ¹ 検証 ¹ もと ¹ 研究 ⁴⁰ 日本 ³¹ 大胆 ²⁵ 前 ²³ 実験 ²⁰ 発表 ¹⁶ 実証 ¹⁵ 一つ ¹⁵ 教授 ¹⁴ 科学 ¹³
19. 被喩詞 A の B A 近 B	傘 が 行く → 傘をさした人が行く ×先 ^{25,2} (2.4 × 10 ⁻⁶) 下 ^{113,1} (1.6 × 10 ⁻⁶) 柄 ^{8,4} (1.0 × 10 ⁻⁶) △自分 ^{26,26} (8.9 × 10 ⁻⁷) 列車 ^{19,51} 花 ^{13,3} 形 ^{3,5} お化け ^{1,13} 目 ^{14,84} 優勝旗 ^{1,1697} 下 ¹¹³ 先 ²⁵ 花 ¹³ 柄 ⁸ 形 ³ 骨 ³ 直径 ³ 裏側 ³ 列 ² 保護 ² 核 ²⁵⁹ 下 ¹³⁸ 日本 ¹²⁰ 米国 ¹⁰⁶ 雨 ⁹¹ 米 ⁶⁵ ビニール ⁵⁶ 人 ⁵² 安全 ⁴⁸ サリン ⁴⁶
20. 被喩詞 A の B A 近 B	スローガン が 決める → スローガンの提唱者が決める ×中身 ^{1,5} (5.9 × 10 ⁻⁶) 素材 ^{1,5} (4.6 × 10 ⁻⁶) 下 ^{12,1} (4.6 × 10 ⁻⁶) もとの ^{10,2} △国民 ^{61,8} (2.6 × 10 ⁻⁶) 政府 ^{33,26} 政権 ^{63,6} 大統領 ^{1,5} 正体 ^{1,5} 党 ^{39,13} 下 ¹² もと ¹⁰ 域 ³ 一つ ³ 中身 ¹ 素材 ¹ 大統領 ¹ 正体 ¹ 開発 ¹ 文字 ¹ 政治 ⁸³ 運動 ⁷⁵ 大会 ⁷⁰ 日本 ⁶⁶ 改革 ⁶⁴ 政権 ⁶³ 選挙 ⁶³ 国民 ⁶¹ 政策 ⁴⁶ 大統領 ⁴³
原因	被喩詞と述語の共起頻度が少ない。
21. 被喩詞 A の B A 近 B	押し入れ を かきまわす → 押し入れの中味をかきまわす △奥 ^{17,1} (2.5 × 10 ⁻⁷) 天袋 ^{5,2} (2.3 × 10 ⁻⁷) 天井 ^{2,4} (9.2 × 10 ⁻⁸) 戸 ^{2,4} 上段 ^{2,4} 中段 ^{2,4} 金庫 ^{1,11} 相手 ^{1,393} 和室 ^{17,5} 前 ^{3,3} 奥 ¹⁷ 天袋 ⁵ 前 ³ 天井 ² 戸 ² 中段 ² 上段 ² 布団 ² 整理 ² 中身 ² 自宅 ²⁵ 奥 ²³ 調べ ²² 遺体 ¹⁹ 和室 ¹⁷ 容疑 ¹⁷ 部屋 ¹⁶ 疑い ¹⁶ 布団 ¹⁴ 県 ¹³
22. 被喩詞 A の B A 近 B	海 が 干上がる ×陸棚 ^{1,10609} (8.6 × 10 ⁻⁶) 川 ^{315,6} (1.2 × 10 ⁻⁶) 湖 ^{62,239} (9.9 × 10 ⁻⁷) 河川 ^{52,323} 東京湾 ^{34,590} ○水 ^{18,11} (4.9 × 10 ⁻⁷) 海水 ^{53,313} 列島 ^{62,239} 景色 ^{6,35} 水路 ^{10,2212} 家 ¹⁹² 幸 ⁹¹ 向こう ⁸⁸ 底 ⁴¹ 色 ³⁹ 男 ³⁴ 記念日 ²⁶ 生物 ²¹ 汚染 ²⁰ 魚 ¹⁹ 日本 ⁷³⁶ 県 ⁶⁴² 山 ⁴⁷³ 空 ³⁸⁹ 前 ³²⁵ 川 ³¹⁵ 大阪 ²⁹⁰ 土佐 ²⁸⁹ 家 ²⁸² 世界 ²⁷⁹
23. 被喩詞 A の B A 近 B	川 が 氾濫 → 川の水が氾濫 ×水路 ^{32,272} (7.9 × 10 ⁻⁷) 河川 ^{120,22} (7.7 × 10 ⁻⁷) ダム ^{138,20} (6.3 × 10 ⁻⁷) 名前の ^{5,22} 黄河 ^{3,3734} 遊び ^{1,90} 流れ ^{127,1} せせらぎ ^{22,3} 情報 ^{34,241} 土手 ^{11,6} 流れ ¹²⁷ 水 ¹²⁶ せせらぎ ²² ほとり ²⁰ 上流 ¹³ 土手 ¹¹ 向こう ¹⁰ 汚染 ⁹ 底 ⁹ 浄化 ⁹ 水 ⁵⁰⁶ 県 ⁴⁴⁰ 海 ³¹⁵ 日本 ²⁸⁷ 山 ²³³ 近く ²¹⁴ 流域 ²⁰⁸ 流れ ²⁰⁵ 自然 ¹⁹⁴ 人 ¹⁹¹
24. 被喩詞 A の B A 近 B	テーブル を 拭く → テーブルのほこりを拭く ×グラス ^{11,215} (8.0 × 10 ⁻⁷) 畳 ^{6,513} (1.3 × 10 ⁻⁷) △角 ^{5,3} (7.5 × 10 ⁻⁸) 脚 ^{3,6} 下 ^{29,1} 汗 ^{1,2640} 印象 ^{5,633} ナブキン ^{1,14} 足 ^{1,14} 裏 ^{4,827} 下 ²⁹ 上 ⁶ 角 ⁵ 周り ⁵ 中央 ⁴ 脚 ³ 横 ³ 真ん中 ³ 端 ³ 縁 ² 交渉 ¹⁴⁹ いす ⁷⁴ 前 ⁷⁴ 席 ⁷⁰ 料理 ⁶⁸ 問題 ⁶⁶ 東京 ⁵⁸ 食事 ⁵⁷ タイム ⁵⁵ 人 ⁵⁵

表 8 述語の対象格が漠然としていることが原因の不正解例

25.	クラリネットを笑う → クラリネットの奏者を笑う
被喩詞	×うそ _{1,339} ^近 (2.7×10^{-7}) △管楽器 _{1,4} ^{の,*} (2.4×10^{-7}) 作品 _{1,4} ^{の,*} (2.1×10^{-7}) 誤り _{1,339} ^近 △音 _{2,1} ^{の,*} (1.9×10^{-7}) かわり _{1,4} ^{の,*} オレンジ _{2,1} ^{の,*} △ジェームス _{1,4} ^{の,*} (1.3×10^{-7}) リチャード _{1,4} ^{の,*} ボール _{1,4} ^{の,*}
A の B	音 ₂ オレンジ ₂ ストルツマン ₂ 管楽器 ₁ 作品 ₁ かわり ₁ ボール ₁ リチャード ₁ ジェームス ₁ 音色 ₁
A 近 B	ピアノ ₄₁ 奏者 ₃₅ 演奏 ₃₁ バイオリン ₂₆ フルート ₁₈ 楽器 ₁₇ 作品 ₁₇ 音楽 ₁₆ チェロ ₁₅ 協奏曲 ₁₅
26.	黒がいい → 黒石を持つ人の情勢が良い(囲碁)
被喩詞	△ジャージー _{3,27} ^の (3.7×10^{-5}) ブーツ _{2,44} ^の (2.2×10^{-5}) 幸先 _{3,2127} ^近 (1.9×10^{-5}) エプロン _{2,44} ^の ツートンカラー _{6,4} ^{の,*} スーツ _{11,1} ^{の,*} ライン _{3,27} ^{の,*} 制服 _{3,27} ^の 組み合わせ _{4,18} ^の イキ _{2,2947} ^近 スーツ ₁₁ ボールペン ₁₀ ジャンパー ₇ ツートンカラー ₆ ズボン ₆ 一団 ₆ 勝ち ₆ モノトーン ₅ ブラウス ₅ 帽子 ₅
A の B	白 ₁₁₁₈ 煙 ₃₀₇ 赤 ₂₈₃ 姿 ₂₆₀ 色 ₂₄₇ 趙 ₂₄₀ 手 ₁₇₈ 日本 ₁₆₀ 鷺 ₁₅₀ 中央 ₁₃₃
A 近 B	
27.	ベルイマンを見る
被喩詞	×息子 _{1,1} ^の (8.1×10^{-4}) 心情 _{1,1} ^の (4.9×10^{-4}) ○映画 _{10,2} ^近 (4.7×10^{-4}) 父 _{1,1} ^の 脚本 _{1,1} 作品 _{6,4} ^近 夢 _{1,26} ^近 秘蔵っ子 _{1,1} ^{の,*} シーン _{1,26} ^近 風景 _{1,26} ^近
A の B	息子 ₁ 心情 ₁ 父 ₁ 脚本 ₁ 秘蔵っ子 ₁
A 近 B	監督 ₁₉ 映画 ₁₀ イングマル ₈ 作品 ₆ スウェーデン ₅ 脚本 ₄ 息子 ₃ 野 ₃ アメリカ ₃ 時代 ₃

表 9 その他の原因による不正解例

原因	被喩詞の一般化が足りない。
28.	どんぶりが好き → どんぶりの料理が好き
被喩詞	×具の _{1,1} (1.3 × 10 ⁻⁴) △ラーメン _{10,3} (4.3 × 10 ⁻⁵) 大盛り _{1,121} (4.2 × 10 ⁻⁵) キンピラ _{1,121} めん類 _{1,121} みそ汁 _{5,10} 水 _{1,1} 飯 _{8,5} お茶漬 _{1,121} スパゲティ _{1,121}
A の B	具 ₁ 水 ₁ ふた ₁ 中身 ₁ もと ₁ ツユ ₁
A 近 B	丼 ₂₁ ご飯 ₁₂ ラーメン ₁₀ 親子 ₁₀ 飯 ₈ メニュー ₇ うどん ₆ そば ₆ 鉢 ₆ みそ汁 ₅
29.	大阪がげんなり → 大阪の人々がげんなり
被喩詞	△委員 _{772,192} (1.7 × 10 ⁻⁸) 弁護士 _{17,7} (4.1 × 10 ⁻⁹) 業者 _{11,22} (2.7 × 10 ⁻⁹) 社長 _{3643,14}
A の B	原告 _{5,49} 高校生 _{5,49} メンバー _{4,69} メーカー _{3,87} 商人 _{3,87} ビジネスマン _{3,87}
A 近 B	街 ₃₂ ホテル ₂₆ 人 ₂₅ 会社 ₂₁ ベッドタウン ₂₀ 町 ₁₈ 弁護士 ₁₇ 文化 ₁₇ ザ ₁₆ 都心 ₁₅ 東京 ₇₃₆₀ 会社 ₆₂₃₁ 府警 ₅₆₄₁ 容疑 ₅₅₂₇ 事件 ₅₅₂₆ 市内 ₅₂₆₇ 会 ₅₀₃₈ 午後 ₄₇₇₂ 午前 ₄₆₁₄ 県 ₄₄₃₃
30.	赤バットが行く → 赤バットを持った人が行く
被喩詞	△川上 _{3,1} (2.8 × 10 ⁻⁵) 大下 _{6,3} (1.6 × 10 ⁻⁶) 自身 _{1,11} (1.3 × 10 ⁻⁶) バット _{9,2} 時代 _{4,4} 豊田 _{1,11} 和久 _{1,11} 並木 _{1,11} 猛 _{1,11} 光 _{1,11}
A の B	川上 ₃
A 近 B	川上 ₁₁ バット ₉ 大下 ₆ 時代 ₄ 野球 ₄ 青 ₃ 戦後 ₂ 戦前 ₂ プロ ₂ 人気 ₂
原因	ハをガに置き換えたのが良くなかった。
31.	学生服が興奮
被喩詞	×ドラマ _{1,2} (1.3 × 10 ⁻⁵) ○男 _{1,2} (3.7 × 10 ⁻⁶) 同級生 _{2,36} (3.1 × 10 ⁻⁶) ナイン _{1,100} 学生 _{1,2} 会場 _{2,36} 丈 _{1,2} コンテスト _{1,100} 学校 _{1,2} 生徒 _{7,2}
A の B	上着 ₂ ドラマ ₁ 男 ₁ 学生 ₁ 丈 ₁ 学校 ₁ 色 ₁ 汚れ ₁ スポン ₁ 代金 ₁
A 近 B	姿 ₁₇ 生徒 ₇ 男子 ₇ 詰め襟 ₇ 県 ₆ ボタン ₆ 学生 ₅ 会 ₅ 中学 ₅ 学校 ₅
32.	蓑が行く → 蓑を着た人が行く
被喩詞	×今日 _{1,20} (2.1 × 10 ⁻⁶) 目 _{1,20} (1.3 × 10 ⁻⁶) みの _{6,1} (9.4 × 10 ⁻⁷) 実 _{1,20} 姿 _{4,2} △芭蕉 _{2,6} (3.2 × 10 ⁻⁷) 喜之助 _{2,6} 形 _{2,6} 入り口 _{1,20} 水 _{1,20}
A の B	
A 近 B	みの ₆ 姿 ₄ ガサ ₄ 火 ₃ 時雨 ₃ 喜之助 ₂ 芭蕉 ₂ 形 ₂ 葉 ₂ 笠 ₂
原因	もう一回換喩を解釈しなければならない。
33.	理論が主張 → 理論の提唱者が主張
被喩詞	△党 _{1,3} (5.7 × 10 ⁻⁵) 一つ _{4,1} (2.5 × 10 ⁻⁵) 両方 _{1,3} (1.6 × 10 ⁻⁵) 従来 _{19,201} △卒業生 _{1,3} (6.7 × 10 ⁻⁶) 世界 _{1,3} 日本 _{150,4} 確立 _{1,3} 米国 _{52,30} 双方 _{3,1499}
A の B	一つ ₄ 学習 ₂ 党 ₁ 両方 ₁ 卒業生 ₁ 世界 ₁ 確立 ₁ もと ₁ 最後 ₁ 手本 ₁
A 近 B	経済 ₂₄₃ 研究 ₁₆₆ 教授 ₁₅₆ 日本 ₁₅₀ トウ ₁₃₃ 社会主義 ₁₀₃ 問題 ₁₀₂ 小平 ₉₉ 政治 ₉₇ 世界 ₉₁
34.	霞が関が動揺
被喩詞	△システム _{3,15} (1.1 × 10 ⁻⁵) 秩序 _{1,39} (3.5 × 10 ⁻⁶) ○官僚 _{30,1} (2.9 × 10 ⁻⁶) 一種 _{1,39} 行内 _{1,2151} 官庁 _{8,5} 法務省 _{7,6} 国民 _{45,33} 役人 _{5,9} 市場 _{11,263}
A の B	官僚 ₃₀ 弁護士会館 ₁₈ 解体 ₁₅ 論理 ₁₀ 官庁 ₈ 法務省 ₇ 文化 ₆ 各省庁 ₆ 役人 ₅ 常識 ₅
A 近 B	東京 ₆₂₆ 千代田 ₁₈₁ 官僚 ₁₇₈ ビル ₁₂₅ 改革 ₁₂₄ 午後 ₁₁₅ 官庁 ₁₁₂ 会 ₁₀₀ 大蔵 ₉₆ 日本 ₉₀

参考文献

- Bouaud, J., Bachimont, B., and Zweigenbaum, P. (1996). "Processing Metonymy: a Domain-Model Heuristic Graph Traversal Approach." In *COLING-96*, pp. 137-142.
- Brown, P. F., Pietra, V. J. D., deSouza, P. V., Lai, J. C., and Mercer, R. L. (1992). "Class-Based n-gram Models of Natural Language." *Computational Linguistics*, **18** (4), 467-479.
- Church, K. W. and Hanks, P. (1989). "Word Association Norms, Mutual Information, and Lexicography." In *ACL-89*, pp. 76-83.
- Fass, D. (1988). "Metonymy and Metaphor: What's the Difference?." In *COLING-88*, pp. 177-181.
- Iverson, E. and Helmreich, S. (1992). "Metallel: An Integrated Approach to Non-Literal Phrase Interpretation." *Computational Intelligence*, **8** (3), 477-493.
- 国立国語研究所 (1996). "「分類語彙表」形式による語彙分類表 (増補版).".
- Kurohashi, S. and Sakai, Y. (1999). "Semantic Analysis of Japanese Noun Phrases: A New Approach to Dictionary-Based Understanding." In *ACL-99*, pp. 481-488.
- 松本裕治, 北内啓, 山下達雄, 平野善隆 (1999). "日本語形態素解析システム『茶筌』 version 2.0 使用説明書." 奈良先端科学技術大学院大学.
- 村田真樹 長尾真 (1997). "意味的制約を用いた日本語名詞における間接照応解析." 自然言語処理, **4** (2), 41-56.
- 山本専, 村田真樹, 長尾真 (1998). "用例による換喩の解析." 言語処理学会第4回年次大会発表論文集, pp. 606-609.
- 山梨正明 (1988). 比喩と理解, 4章. 東京大学出版会.
- 李徳奉 (1990). "言語学における比喩研究の最近の動向から." 芳賀純 子安増生 (編), メタファーの心理学, 6章. 誠信書房.

略歴

- 内山 将夫: 筑波大学第三学群情報学類卒業 (1992). 筑波大学大学院工学研究科博士課程修了 (1997). 博士 (工学). 信州大学工学部電気電子工学科助手 (1997). 郵政省通信総合研究所非常勤職員 (1999).
- 村田 真樹: 1993年京都大学工学部卒業. 1995年同大学院修士課程修了. 1997年同大学院博士課程修了, 博士 (工学). 同年, 京都大学にて日本学術振興会リサーチ・アソシエイト. 1998年郵政省通信総合研究所入所. 研究官. 自然言語処理, 機械翻訳, 情報検索の研究に従事. 言語処理学会, 情報処理学会, ACL, 各会員.

馬 青： 1983年北京航空航天大学自動制御学部卒業。1987年筑波大学大学院理工学研究科修士課程修了。1990年同大学院理工学研究科博士課程修了。工学博士。1990～93年株式会社小野測器勤務。1993年郵政省通信総合研究所入所，主任研究官。人工神経回路網モデル，知識表現，自然言語処理の研究に従事。日本神経回路学会，言語処理学会，電子情報通信学会，各会員。

内元 清貴： 1994年京都大学工学部卒業。1996年同大学院修士課程修了。同年郵政省通信総合研究所入所，郵政技官。自然言語処理の研究に従事。言語処理学会，情報処理学会，ACL，各会員。

井佐原 均： 1978年京都大学工学部電気工学第二学科卒業。1980年同大学院修士課程修了。博士（工学）。同年通商産業省電子技術総合研究所入所。1995年郵政省通信総合研究所関西支所知的機能研究室室長。自然言語処理，機械翻訳の研究に従事。言語処理学会，情報処理学会，人工知能学会，日本認知科学会，ACL，各会員。