

比喩表現コーパスの構築と問題点

-言語学の立場から-

伊藤 薫
京都大学大学院

ito.kaoru.38r@st.kyoto-u.ac.jp

1 はじめに

比喩表現は、機械翻訳をはじめとした自然言語処理の様々な分野で重要である。また近年、意味論の研究でメタファー、メトニミーは主要な研究テーマとなっている。その一方で、比喩表現は直喩を除き形式が定まっておらず、自動的に大量のデータを集めることが困難である。したがって、ある程度の規模を持った比喩表現の用例を資源として公開するのは、多くの研究者にとって有益であると言える。

本発表では、主に言語学の立場から修辭表現コーパスの必要性和価値、および修辭表現コーパスを構築する上で生じる問題と、限定的ではあるがその解決方法について述べる。なお、修辭表現にはメタファー、メトニミー以外にも様々なものが存在するが、本発表では収集すべきデータとして比喩表現の中でも中心的地位を占めるこの二つを念頭に置いている。また、比喩表現であることを示すタグは、人手によってつけることを想定している。実際にコーパスの構築をする上では、著作権など法律上の問題も生じると考えられるが、その点については考慮せずまずは理想像を提示することにする。

2 これまでの比喩表現データ

管見の限り、これまで比喩表現の例を電磁的方法でまとめたデータとして公開したものは存在しない。比喩表現を含む修辭表現の例は『辞典』という形で公にされてきた。(e.g. 榛谷 [1]、中村 [2][3]、野内 [4][5]、佐藤ら [6]) これらの本には『辞典』という名を關してはいるものの、各項目の意味的な定義だけでなく、様々なテキストから実際の用例を収集・引用しており、部分的に修辭表現例集としての機能も持っている。大まかにいえば、これらは次の二つの編纂方法に分けられる。

- (i) ある喩えの対象がどのように表現されているかについて記述していく方法。つまり Richards(1936) に倣っていえば、ある趣意 (tenor) がどのような媒体 (vehicle) によって喩えられているかを中心として記述する方法。
- (ii) ある修辭技法がどのような性質であるかについて記述していく方法。いわば「修辭学用語辞典」のような存在。

例えば、榛谷 [1] は (i) の方法を取っており、「比喩表現辞典」という名を冠しながらも直喩のみを対象に「愛」「哀歌」「哀愁」……といった趣意ごとに収集している。中村 [2] は五分の程度を (i) に割り、残りを (ii) の記述に当てている。(ii) の方法を取っているのは中村 [3]、野内 [4][5]、佐藤ら [6] であるが、これらは主に網羅している項目の数と項目あたりの記述量の点で異なっている。歴史的に比喩研究はレトリック¹の下位分野であり、文体や議論法などの項目も含めた場合には、当然ながら一項目あたりの記述量は減り、収録される実際の用例も減少する。

これらの編纂方法から用途を推測すると、(i) は文章を書いたり話をする上で何か喩えたいものがあるときに、適切な喩えを収集された例を参考にして導くもの、(ii) はどのような修辭技法があるかを概念的に把握するためだと考えられる。当然ながら、これらの中に収集された例はあくまで補助的 (特に (ii) の場合) なので、現在の言語研究や機械学習の訓練データとして用いるには不便な点が多く存在する。本発表では、これらの持つ不都合な点を、近年の比喩研究の動向を踏まえながら明らかにする。

¹ここでいう「レトリック」とは、佐藤 [8] の用いているのと同じ意味で、「説得する表現の技術」と「芸術的表現の技術」を含む研究分野である。

3 近年の比喩研究の動向

本節では、近年の比喩研究の動向の中で、比喩表現コーパスを構築する上で特筆すべき点を紹介する。

3.1 比喩表現の同定基準

一つ目としては、比喩表現を同定する手法が確立されつつあることが挙げられる。

これまでの研究では、ある表現をメタファー、メトニミーとして認めるかどうかは研究者によって様々であったため、メタファーやメトニミーを量的に研究することは困難であった。例えば、もともとはメタファーであったものの、慣習化が進みもはや一般の話し手にはメタファーだと捉えられなくなった、いわゆる死喩 (dead metaphor) をメタファーとして認めるかどうかは、研究者や研究目的によって異なってくる。また、修辞表現の研究者が多くの場合、特にメタファー研究において、自分が何をメタファーとしているかを明示して来なかったことが問題視されていた。この問題に対し、Pragglejaz Group² は”metaphor identification procedure(MIP)”を提唱している [9]。MIP の内容は以下のとおりである。

1. Read the entire text-discourse to establish a general understanding of the meaning.
2. Determine the lexical units in the text-discourse
3. (a) For each lexical unit in the text, establish its meaning in context, that is, how it applies to an entity, relation, or attribute in the situation evoked by the text (contextual meaning). Take into account what comes before and after the lexical unit.
(b) For each lexical unit, determine if it has a more basic contemporary meaning in other contexts than the one in the given context. For our purposes, basic meanings tend to be
 - More concrete; what they evoke is easier to imagine, see, hear, feel, smell, and taste.
 - Related to bodily action.

²”Pragglejaz Group”は、以下のメンバーの頭文字を取って名付けられた研究者のグループである: Peter Crisp, Ray Gibbs, Alan Cienki, Gerard Steen, Graham Low, Lynne Cameron, Elena Semino, Joseph Grady, Alice Deignan, Zoltan Kövecses.

- More precise (as opposed to vague)
- Historically older.

Basic meanings are not necessarily the most frequent meanings of the lexical unit.

- (c) If the lexical unit has a more basic current-contemporary meaning in other contexts than the given context, decide whether the contextual meaning contrasts with the basic meaning but can be understood in comparison with it.

4. If yes, mark the lexical unit as metaphorical.

MIP は、より幅広い語をメタファーとして捉えるアプローチである。ここでは代表的なものとして MIP を挙げたが、これを発展させた MIPVU[10](VU は Steen の所属する *Vrije Universiteit* の略) など、別の基準も存在する。いずれにせよ、少なくともメタファーに関しては、研究者が自分の研究において、何を以ってメタファーとするかをはっきりさせる動きが活発であり、ある程度の基準ができていと見てよいだろう。

3.2 言語使用の重視

また、近年の研究の動向として、実際の言語使用におけるメタファーやメトニミーの研究が盛んであることが挙げられる。例えば、Deignan, Littlemore and Semino は、現代のメタファー研究者を次の二つに分けている [11]。

- those who research metaphorical patterns of thought; and
- those who research metaphorical language, and seek explanation for it.

前者は言語を静的で均一なものとして捉えており、どのような思考のパターンがメタファー表現に反映されているかについて、作例を中心に研究が進められてきた。これに対し、Deignan らはジャンルやレジスターがどのようなメタファーを使用するかに影響を与えることを、異なる参加者による科学についての談話の比較等を通じて示しており、実際の言語使用におけるメタファー研究の重要性を説いている。

3.3 談話レベルの観察

3.2 で紹介した Deignan らの主張は、言語使用に含まれるレジスターやジャンルといった、言語使用にお

ける言語外の要素、つまりコンテキストが比喩表現を研究する上で重要だということを示していた。しかし、実際の言語使用における用例を研究することの重要性は他にも存在する。

ここでは、テキスト内の他の要素が比喩表現の理解に関わっていることを示しておく。3.2で述べた思考のパターンとしてのメタファー研究では、文レベルの作例が主な研究対象であった。しかし、談話³レベルを考慮しなければ、比喩表現の理解の説明として不十分な点が出てくる場合がある。例えば、次の諷諭(連続する一連のメタファー)のような場合は、文レベルの観察ではメタファーかどうかを判断できない。

- (1) 「これが全国規模の会社の強みっていうやつだよ。なにしろ、ウチの店は重点戦略店舗になっているから、営業管理部のほうでも優先的に回すようにしてくれてるしな」

セール終了の打ち上げで酔っぱらって帰ってきたお父さんは、上機嫌に言った。

「ジューテンセンリヤクテンポって、なに?」

「要するに、アレだ、本社が大事にしてくれるお店っていうことだよ。ここの陣地をぶんどるまで、どんどんタマでも兵隊でも送り込んで、なにがなんでも戦争に勝つぞーっ、ってな」

(重松清「いいものあげる」『ロング・ロング・アゴー』所収)

(1)では、太字部分の「陣地」「ぶんどる」「タマ」「兵隊」「戦争」がビジネスに関連のある意味を表す一連のメタファーとして用いられているが、これらは「重点戦略店舗」の説明であること、つまりテキストの前方を考慮しなければメタファーであることを判別できない。

また、局所的に見てメトニミーだと思われるような表現も、テキストの他の部分で別の意味で用いることを明示している場合がある。例えば、(2)では、この部分のみを読んだ場合「赤シャツ」が文中の「男」を指すメトニミーとして用いられているという解釈が妥当であると考えられる。

- (2) 君釣りに行きませんかと赤シャツがおれに聞いた。赤シャツは気味の悪いように優しい声を出す男である。まるで男だか女だか分りゃしない。男なら男らしい声を出すもんだ。ことに大学卒業生じゃないか。物理学校でさえおれくらいな声が出るのに、文学士がこれじゃ見っともない。

³ここでいう「談話」は、単に文より長い言語単位のことを指す。

(夏目漱石『坊っちゃん』)

しかし、(2)より前に、「赤シャツ」はあだ名であることを明示する部分がある(3)。

- (3) 今日学校へ行ってみんなにあだなをつけてやった。校長は狸、教頭は赤シャツ、英語の教師はうらなり、数学は山嵐、画学はのだいに。今にいろいろな事を書いてやる。

(夏目漱石『坊っちゃん』)

(2)では、「赤シャツ」は教頭を指すあだ名であることが明示されている。この教頭は年中赤シャツを着ていることが別の部分で述べられているので、当然このあだ名をつける際にはメトニミーが関わっているが、「赤シャツ」が教頭を指しうる⁴ということは読み手にとって明らかたため、推測のみによって解釈されるような、言語学でよく例に出されるメトニミーとは理解過程が異なるといえよう。

以上の例のように、比喩表現の理解には文を超えたレベルで観察しなければ見落とししてしまう事実が存在する。特に「赤シャツ」の例は一つの作品を通じて語の理解に関わる問題であり、これだけの規模の作例をすることは困難である。この点でも、実際の言語使用から比喩表現の用例を収集し蓄積しておくことは有意義だといえる。

4 比喩表現コーパスに望むこと

本説では、前節までの内容を踏まえ比喩表現コーパスにとって必要であると考えられる仕様について箇条書きの形で述べる。

・一つのテキストを通して比喩のタグ付けをすること

³で述べた通り、近年の比喩研究では実際の言語使用が重視されている。つまり、比喩表現の含まれる文だけでなく、前後のテキストやコンテキストも重要となる。したがって、2で述べたような編纂の方法では言語使用における比喩を研究する際に用いるには不都合である。

・データとした談話やテキストは、全文を閲覧可能であること

⁴ちなみに、『坊っちゃん』に登場する「赤シャツ」という語が教頭を指すか字義通りの赤シャツかを手作業で調べたところ、「赤シャツ」全体の生起頻度が168であり、字義通りに用いられているのは4例であった。また、あだ名であると明示された部分は3番目の「赤シャツ」であり、これ以前の「赤シャツ」2例は全て字義通りの意味であった。

現在言語学で用いられているコーパスの多くは、コーパス利用者が全文にアクセスすることを許していない。しかし、3.3 で述べた通り、ある部分だけを見れば推測に頼って意味を捉えなければならないような比喻表現に見えても、テキスト中ですでにそれが何を指すか明示されている場合が存在する (2)。したがって、著作権の問題はあるものの、一つのテキスト全体に利用者がアクセス可能であることが望ましい。

- ・データのメタ情報にアクセス可能なこと

3.2 で述べたように、ジャンルやレジスターはメタファーの使用に大きく関わる。したがって、データのメタ情報へアクセス可能にしておく必要がある。また、これに関連して、収集するテキストのジャンルやレジスターに偏りが無い方が望ましいといえる。

- ・判断の揺れるものについては、できるだけ比喻であると認めること

MIP の部分でも触れられていたが、研究者や研究目的によって何をメタファーとするか (メトニミーもであるが) は異なってきてしまう。人手でも比喻表現の収集は難しいため、再現率 (recall) を上げる方略の方がよいらる。

- ・オープンであること

これと関連して、何を以って比喻表現とするかの基準が異なるため、利用者が必要としている例とそれ以外に分けられることが望ましい。自由にタグを編集できるように、できるだけオープンにするべきだろう。

- ・比喻の判断基準を明示すること

3.1 で述べたように、近年メタファーを同定する基準について明示する動きが出ている。MIP, MIPVU や独自の基準等によって、何を以ってメタファーやメトニミー、他の比喻表現としているかについて、基準が明示されていることが望ましい。

参考文献

- [1] 榛谷 泰明 (編) 1988. 『レトリカ - 比喻表現事典』 白水社: 東京 .
- [2] 中村明. 1995. 『比喻表現辞典』 角川書店: 東京 .
- [3] 中村明. 2007. 『日本語の文体・レトリック辞典』 東京堂出版: 東京 .
- [4] 野内良三. 1998. 『レトリック辞典』 国書刊行会: 東京 .
- [5] 野内良三. 2005. 『日本語修辞辞典』 国書刊行会: 東京 .
- [6] 佐藤信夫・佐々木健一・松尾大. 2006. 『レトリック事典』 東京: 大修館書店 .
- [7] Richards, I A. 1936. *The Philosophy of Rhetoric.*: New York and London Oxford University Press.
- [8] 佐藤信夫. 1992. 『レトリック感覚』 講談社: 東京.
- [9] Praggeljaz Group. 2007. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol* 22(1) 1-39.
- [10] Steen, G. J., Aletta G. D., Herrmann, J. B., Kaal, A. A., and Krennmayr, T. 2010. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU.*: Amsterdam John Benjamins.
- [11] Deignan, A., Littlemore, J., and Semino, E. 2013. *Figurative Language, Genre and Register.*: Cambridge Cambridge University Press.