

日本語の用言相当慣用表現の 意味空間における分布図

亀井 真一郎 田村 真子 村木 一至

NEC 情報メディア研究所

1 はじめに

近年、言語処理技術の発達により広範囲の文書を解析できるようになっているが、解釈の品質をより一層高めるには、単語単位の辞書だけでなく、特定の語句の決まった組み合わせにより意味の定まる句表現 (慣用表現) の扱いを充実させる必要がある。

従来、日本語の慣用表現には様々な種類の存在が指摘されている [1, 2, 3, 4, 5]。関連する用語も「熟語」「連語」「換喩」「比喩」「機能動詞表現」など数多く、それらの間の関係は明確でなかった。そこで我々は、特に用言句相当表現を中心にして、意味的な成り立ちの観点から慣用句を分類し、慣用表現の「分布図」を作成した [6]。例えば「あぐらをかく」という句は慣用的と感じられる読みを二つ持つ ((1) 足を組んだ座り方、(2) いい気な態度でいる様) が、我々の分布図を用いれば、それらの違いが図中の位置として視覚化できる。実際、この分布図を元にして慣用句の大規模収集に役立てることができた [7, 8, 9]。本稿では、この慣用表現の意味空間における分布図について説明する。

2 日本語慣用表現の複雑さ

日本語における用言句相当の慣用表現とは、「体言 + 格助詞 + 用言」の形式を持つ表現のうち、その句全体の意味を単純に構成要素の基本義の合成に還元できない表現や、句を構成する体言と用言の組合せが固定的に定まっている表現を指す。以下、本稿で「慣用表現」とは上記の「用現句相当慣用表現」を指すものとする。

日本語の慣用表現には、様々な種類が存在する。一般に「慣用句」という用語からまず思い浮かべるのは、悪い組織から抜ける、という意味の「足を洗う」のような、句全体が全く新規の意味をもつ表現であろう。しかしながら日本語にはこのタイプ以外にも慣用的と感じられる表現が多種類存在する。

表現「電話をかける」の場合、体言部は「電話すること」という通常の意味であるが、動詞「かける」を用いる点が慣用的である。「かける」の主たる意味として、電話を相手に通じさせる行為を指し、英語で「make」で表現されるような意味は、通常は考えられない。

表現「声はずむ」の場合、体言部「声」は用言部

「はずむ」の主体の意味制約 (「物理的具体物」) を破っている点が慣用的である。

表現「目をつぶる」の句全体の意味は、体言部「目」の意味も用言部「つぶる」の意味も、その基本義を保っていると考えられる。しかしながら「閉じる」という意味の動詞「つぶる」は、その目的語が「目」の場合だけに使われるという点が慣用的である。またさらに「目をつぶる」という句は、全体で「見てみないふりをする」という新たな慣用的な意味をもっている。

このように、日本語の慣用表現には多数の種類があり、従来それらの関係は明確でなく、複数の異なる「慣用的」意味を持つ場合の把握が困難であった。

3 慣用表現の意味空間分布

3.1 意味空間における3次元表示

日本語の多種類の慣用表現の関係を明確に把握するため、慣用句の成立要件として次の3つの要因を考えて作成した分布図が表1である。

1. 体言部 (x 軸) : 体言の意味が基本義か ($x=0$)、派生義か ($x=1$) に関する軸。
2. 用言部 (y 軸) : 用言の格スロットが広い語義を受け付けるか ($y=0$)、特定の語類のみを許すか ($y=1$)、単一の語だけと結合するか ($y=2$) という、体言と用言の結合に関する軸。
3. 表現全体 (z 軸) : 表現全体が新たな意味を生じることに関する軸。

それぞれを意味空間の3つの座標軸になぞらえて、慣用句の種類を便宜的に (010) 類、(020) 類、等と名付けた。表中、(000) 類は一般句であり、それ以外が種々の慣用句の類を表している。

3.2 各慣用句類の特徴

この分布図の各類には次のような特徴がある。

1. (000) 類 [通常句] : 体言と用言の通常の組合せからなるもので慣用表現ではない。例にある「足を洗う」「電話がある」「電話をかける」は文字通りの意味である。例えば「壁に電話をかける」の場合の「電話をかける」はここに分類される。

表 1: 日本語の動詞句相当慣用表現の分布図

		体言部 (⇒ x 軸)		表現全体 (z 軸)
		基本義	派生義	
用言部 格スロットの語義制約 (↓ y 軸)	一般義	(000) 類: <u>通常句 (literal)</u> 頭を使う 株が上がる 足を洗う (しのぎを削る) 電話がある 電話をかける 首をひねる	(100) 類: <u>換喩</u> 頭を使う 株が上がる	z 類: <u>比喩的慣用句</u> 株が上がる 足を洗う しのぎを削る 首をひねる
	特定類	(010) 類: <u>制約不整合</u> 意見をおつける 声ははずむ	(110) 類: <u>換喩 + 不整合</u> 腹をこわす	
	単一語	(020) 類: <u>特定類 (a)</u> 医者 (内科 ...) にかかる 便り (連絡 ...) がある <u>特定類 (b)</u> 睡眠 (休息 ...) をとる 攻撃 (圧迫 ...) を受ける	(120) 類: <u>換喩 + 特定類</u> 電話がある	
		(030) 類: <u>単一語 (a)</u> あぐらをかく 茶を点てる <u>単一語 (b)</u> 功を奏する 端を発する <u>単一語 (c)</u> 命を取りとめる 目をつぶる 口をつぐむ	(130) 類: <u>換喩 + 単一語</u> 口をきく 電話をかける	あぐらをかく 口をきく 端を発する 目をつぶる 口をつぐむ

- (010) 類 [制約不整合]: 用言の語義制約と整合しない体言の組み合わせが固定化したものである。例えば「声ははずむ」では「はずむ」の制約「物理的具体物」と整合しない「声」が結合している。
- (020) 類 [特定類]: 特定の体言の一群と結合することで用言の意味が一つに定まる点に特徴をもつ。動詞「かかる」は多数の意味を持つが、「医者」の類 (内科 / 外科 etc.) と組合わさることで意味が一意に定まる (特定類 (a))。同類として、ある種の動作性体言と結び付く特定の動詞が助動詞的な役割をする一群がある (特定類 (b))。例えば「攻撃を受ける」の「受ける」は受身の助動詞と同様に働き、全体で「攻撃される」とほぼ同義となる。このように元の実質的な意味が薄らいだ動詞は機能動詞と呼ばれる [1]。
- (030) 類 [単一語]: 体言がほぼ唯一に固定されている一群である。「あぐらをかく」の「かく」がこの意に用いられるのはこの表現に限られる (単一語 (a))。同様に、「奏する」には「天皇に申し上げる」「演奏する」「うまく成し遂げる」の意味があるが、「うまく成し遂げる」の意味の場合の目的語は「功」のみである点が慣用的である (単一語 (b))。また「口をつぐむ」の場合、「閉じる」意味の「つぐむ」が、目的語として「口」だけを伴う点が特殊である (単一語 (c))。
- (100) 類, (110) 類, (120) 類, (130) 類 [換喩]: 「頭を使う」という句を「考える」の意で用いるとき、体言「頭」は換喩の意味である「頭の機能」を表している点で、通常句と区別される。
- z 類 [比喩的慣用句]: 構成要素の元の意味はほとんど残っておらず、全体が新しい意味をもつ典型的な慣用句である。中には「しのぎを削る」のように元々存在した文字通りの読みが失われ、慣用句としてのみ機能するようになった例もある。「首をひねる」という句はその動作が シグナルとして用いられるところから「不思議がる」の意となる。このような象徴的動作に起因した表現も比喩的慣用句の一つである。

また、直喩の中で慣用的に用いられる表現も「後ろ髪をひかれる (ような名残惜しい思い)」「手にとる (ようによくわかる)」など多数ある。直喩と慣用句とは以下のような「直喩→暗喩→比喩的慣用句」という過程をたどって関係するものと考えられる。

個々の表現により下記のどのレベルにまで独立度が進んでいるかに差があるが、その度合いを z 軸の座標に相当させることで、これらの表現も比喩的慣用句の中で分類表示することもできる。つまり下記の 1. 例え 2. 直

喩 3. 暗喩 4. 独立 を Z 軸の座標 = 1、2、3、4 に対応づけて表示し、どのレベルの用法が可能であるかを視覚的に把握することができる。

段階	表現	直喩マーカ	様態	事柄
1. 例え	しのぎを削る	ような	激しい	戦い
2. 直喩	しのぎを削る	ような	—	戦い
3. 暗喩	しのぎを削る	—	—	戦い
4. 独立	しのぎを削る	—	—	—

3.3 多種の慣用的意味の表示

上述したように、一つの句が複数の意味を表現する場合がある。我々の分布図でそれらがどのように表示されるかを説明する。

例えば「頭を使う」の場合、サッカーのヘッディングのように物理的に頭で何かを行なう意味は通常句 (000) 類に配置される。これに対し、「考える」の意の「頭を使う」は「頭」という語で「思考力」といった「頭の機能」を意味しており、かつ用言部「使う」は元の意味を保持していると考えられるから、体言が派生義である (100) 類に配置される。表 1 では、(000) 類と (100) 類とは横に並んでいるが、これがこの二つの意味の関係を視覚的に表現している。

表現「足を洗う」の場合、文字通りの意味 ((000) 類) と、表現全体がもつ新たな意味 (Z 類) とがある。表 1 では (000) 類と Z 類とを横に並べて、これら二つの意味の関係が視覚的に把握できるようにしている。

表現「電話がある」の場合、電話機が存在することを意味する場合 (000) 類に配置されるが、「電話という通信を受ける」という意味は (120) 類に配置される。この (120) 類は体言部の意味が派生義であり、用言部も特定の意味グループに属する体言とだけ結びつくという特徴をもっている。同様に「電話をかける」の場合、電話機を壁などにかけることを意味する場合 (000) 類に配置されるが、「電話という通信を発する」という意味は (130) 類に配置される。この (130) 類は体言部の意味が派生義であり、用言部がほぼ単一の体言とだけ結びつくという特徴をもっている。このような表現の場合、通常句 (000) 類と慣用句 (120) 類または (130) 類は、分布図で斜めに配置されており、句のもつ複数の意味が視覚的に把握できる。

表現「あぐらをかく」の場合、(1) 足を組んで座る、(2) いい気な態度でいるという二つの慣用的な読みを持つ。前者が慣用的であるのは、用言部「かく」の特殊性による。体言部「あぐら」は基本義であるが、組んで座ることを「かく」という動詞で表現する部分が特殊で慣用的用法と認定される。この意味は (030) 類に配置される。後者は、句全体の意味が新たに生じている点が慣用的であり、Z 類に配置される。これら二者の関係を視覚

的に把握しやすいように、分布図では横に並んで表現している。同様な、二つの慣用的意味を持つ句の例としては「目をつぶる」「口をつぐむ」などがある。

表現「口をきく」の場合 (1) 言葉を話す (2) 仲介する という二つの意味を持つ。前者は「口」で「言葉」を意味する点で、名詞部分がその派生した意味で使われており、かつ「きく」という動詞を「はなす」という意味に用いる特殊性と合わせて慣用的であると言える。この意味は (130) 類に配置される。後者は前者の「言葉を話す」意味を特殊な条件に適用し、全体として「仲介する」という意味に固定化している点が慣用的である。この意味は Z 類に配置される。

4 慣用表現を構成する要素の特徴

慣用表現を構成する要素、特に用言部に関しては以下のような特徴を見出した。

1. 和語動詞：慣用句を構成する動詞は大部分が和語動詞である。和語動詞の使われ方としては「連絡をとる (=する)」「攻撃を受ける (=される)」といった機能動詞表現の出現頻度が高い。
2. 漢語動詞：漢語動詞 (サ変動詞) は「(汚名を) 返上する」等ごく少数に限られる。
3. 和語化漢語動詞：特定類あるいは単一の体言と結びつく表現には、「愛す/愛する型」「信じる/信ずる型」「熱す/熱する型」活用の「和語化漢語動詞」が多く見られる。

慣用表現の構成要素として、和語動詞 (通常の五段、一段の動詞) が多く、漢語動詞 (いわゆるサ変動詞) が少ないのは、古くからのやまと言葉の言い回しが定着して今日まで生き残ったものが慣用句と認識されていることを示している。「連絡をとる (=連絡する)」の「とる」のような機能動詞表現と呼ばれる表現の存在は、漢語を日本語にとり入れる際に、「動作性の名詞 + する」というサ変動詞形成の型と、「動作性の名詞 + 助詞 + 和語動詞」という機能動詞表現形成の型の二つがあったためであると推測される。

また和語化漢語動詞 (ex. 発する、奏する) が多いのは、漢文の原典を読み下した表現が定着したものと推測することができる。和語化漢語動詞は、愛する (サ変活用) / 愛す (五段活用) のように、漢語系と和語系の二種類の活用をもっている点から、一般の漢語動詞 (サ変動詞) よりも日本語化が一層進んでいると考えられる。一般の漢語動詞を構成要素用言とする慣用表現が少ないのは、慣用表現の多くが漢語動詞を日本語にとり入れる際に生じたからであると考えれば理解できる。少数の例外 (「(汚名を) 返上する」等) は、直接漢文の原典に起源をもつものと推測される。

5 分布図の応用

5.1 表現処理の効率化・高品質化

慣用句が意味的なまとまりである以上、純粋な統計的処理だけによって慣用句か否かを判断することはできず、表現収集に際しては最終的には人間の判断が必要である。人間の目視による判断には、時間がかかり、また判断のゆれ、不注意によるものなどが伴うから、表現の大量収集の効率化は大きな課題である。

表1の分布図は、慣用句の種々の要因が視覚的に把握できるので、慣用表現の大量収集を効率化する強力な基礎データとして機能する。例として「医者にかかる」という表現を考察する。動詞「かかる」には「(医者などに) 診てもらう」という意味があるが、動詞「かかる」は多義性を持ち、医者およびその類の語との組合せの時にのみ「診てもらう」の意味になるので分布図中(020)類に配置される。

ここで注意すべきは、慣用句として処理するか否かの判断である。処理方法は、実際の辞書における用言の多義性と、格スロットの制約に用いる体言の意味分類の体系とを考慮して判断する必要がある。(020)類に配置された表現は、名詞の意味類が定義され、用言の格の意味制約として処理できれば、そうしても良い。あるいは慣用表現として表層を名指す処理も可能である。いずれにせよ、各表現が意味空間で占める位置を特定し、処理方法と辞書情報を明確化することが、この分布図を用いる意義である。

5.2 日英対訳辞書の構築

実際に我々は、上述の分布図と、前節で述べた構成要素の特徴とを元に基本的には人間の目視に基づいて日本語の慣用表現を効率的に大量収集し、これに英語対訳を付与し、翻訳のための情報を記述して辞書化を行ない([9])、システムに実装してその有効性を確認した([8])。現在までに約3万の慣用表現を得ている。

英語対訳付与および翻訳情報記述の際には、各慣用表現とそれに対応する英語表現との句構成の差の観点から、慣用表現を2つのタイプに分類した[7]。

一つは、日本語と英語とで各構成要素ごとに対応をとる必要があるタイプである。日本語の句構成「体言＋格助詞＋用言」の体言部に修飾語が付き得る表現で、かつ英語の句構成が「動詞／形容詞＋名詞」の形式である場合、体言、用言それぞれの対応を保持する必要がある。例えば「(ひどい) うそをつく → tell a (terrible) lie」や、「(彼の) 弱点をつく → tough him on a sore spot」などである。特に後者の場合、日本語の体言修飾句が英語で動詞の目的語となり、文中で占める位置が異

なっている。

もう一つは日英で表現全体の対応をとれば十分なタイプである。例えば「袖を通す → wear」「(彼に) 電話をかける → telephone (him)」のように、体言部に修飾要素がつかないか、英訳語が一語で表現される場合には、句表現全体で日英対応がついていれば充分である。

この日英の構造の対応を考慮した分類は、日本語見出しを収集する際にも利用できる。日本語と英語とで構文形の異なる表現(ex. ペンキをぬる ↔ paint)は、日本語・英語単独で見た場合特に慣用的でなくとも、処理上は本稿で述べた慣用表現と同等の機構で処理できる。

6 おわりに

用言句相当の慣用表現をその意味的な成り立ちの観点から分類し、全体を分布図の形に可視化した。種々の慣用的表現を図中の位置の違いとして表現でき、慣用表現を収集する際の品質を高めることができる。この分布図を基本として、実際に日英対訳支援システムのための慣用表現辞書の日本語表現の収集、対訳付与を行なった。現時点で約3万表現の日英対訳慣用表現辞書が得られている。それをシステムに実装して動作・確認を行なった。これにより本分類の有効性を検証できた。今後は、この分布図と実際に得られた慣用表現辞書を、日本語の動詞の語義選択に必要な名詞の意味分類(シソーラス)構築などの意味分析に活用してゆきたい。

参考文献

- [1] 村木 新次郎：「日本語の機能動詞表現をめぐる」国立国語研究所報告 65, pp.17-75, 1980.
- [2] 宮地 裕：『慣用句の意味と用法』, 明治書院, 1982.
- [3] 奥 雅博：「日本語慣用表現の分析と日英翻訳への適用」, 情処研究会資料, 87-NL-62-2, 1987.
- [4] 首藤 公昭 他：「日本語の慣用表現について」情処研究会資料, 87-NL-66-1, 1987.
- [5] 森田 良行 他 編：「ケーススタディ 日本語の語彙」, 桜楓社, 1989.
- [6] 亀井：「日本語の用言句相当慣用表現の分類とその応用」情報処理学会第47回全国大会. 1993.
- [7] 田村 他：「用言句相当慣用表現の日英対訳の型の分類とその応用」情報処理学会第50回全国大会. 1995
- [8] Yamabana, et al. "Interactive Machine-Aided Translation Reconsidered - Interactive Disambiguation in TWP -", Proc. of NLP'95, pp.368-376. 1995.
- [9] 田村 他：「日英機械翻訳のための大規模慣用表現辞書の構築」言語処理学会第2回年次大会. 1996