

3軸モデルによる慣用表現の分類

6L-2

野村直之 高橋一裕

日本電気C&C情報研究所

1 はじめに

言語学・工学の諸分野では慣用句および類似の術語の指し示す概念は一定していないように見える。(たとえば文献[1],[2],[3]を比較。)市販の辞典によれば慣用句とは「二語以上が結合し、または相応じて用いられ、その全体がある固定した意味を表すもの。『油を売る』…(文献[8])」とされ、また idiom とは、「全体の意味が各単語の意味を組み合わせて得られるのではなく、独特なもの;例 “in the soup” 『困って』(文献[5])」とされる。これらは、いずれも計算言語学の言葉で言えば「構成性(compositionality)が成立しないこと」に注目した定義になっている。しかし、現在工学分野では明らかにこの定義とは異なる扱いもみられる。(文献[2])そこで、本稿では慣用性を互いに独立な3軸によって定義する枠組を提案する。

慣用表現の定式化は意味解析システムの解析精度を高め、さらに解析結果の意味構造を適切に定義することに貢献する。特に、ある語彙が慣用句であるか否かの線引きを迫られるという意味で『大規模な』自然言語処理システムの構築に対して、有益な示唆を与えるものと考えられる。この他、自然言語からの知識獲得や、外国語学習者のためのCAIシステムの効率的な設計や教材内容そのものへの応用が考えられる。

2 慣用表現を定式化する3軸モデル

従来からの慣用表現の各種の定義や分類を分析・整理し実データで検証した結果、表層の現象として非常に多数の種類があるかに見える慣用表現を第1図に示すような3軸にまとめることができるという結論を得た。

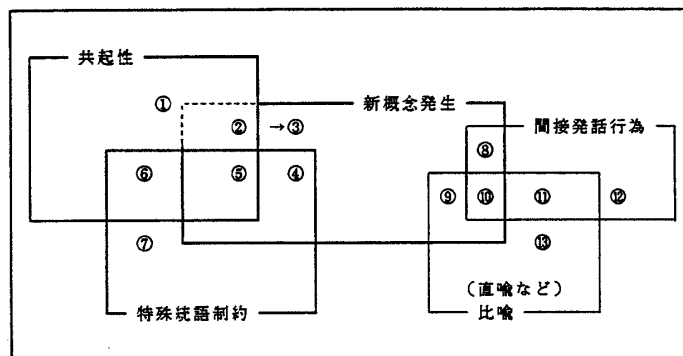
ここで「新概念の発生」とは前節に記した「構成性の不成立」に等しく、個々の構成要素を通常のフレームに結びつけることにより句全体としての意味(概念)を計算することができないことを基準とする。「虎になる(周囲をかえりみず酒を飲んでいい気持ちで酔っ払う)」等が典型例である。

一方、「特殊な統語的制約」とは、成句を構成する単語の品詞が許す通常の文法的振る舞いと異なる振る舞いが生じていることを基準とする。

例:「*なった虎」「?酔うと虎にも天狗にもなる」「*虎にはなる」「OK立派な虎になる」

たとえば、上例のように関係節化、副助詞挿入、構成要素の並列構文化などのテストにおいて、独立した単語の集まりならば許されるはずの統語的実現形が許されない。しかし、全体で1単語となっているわけではない。

「共起性」は、複数の形態素の間で、互いに同時に出現しやすい度合いを基準とする。ここでは「虎になる」のように完全に固定化した組み合わせは除外し、統計的な共起頻度が高いことが形態素間で観測される現象を対象とする。



第1図 慣用表現を定式化するモデルの分類

3 3軸に基づく慣用句の分類

以下、様々な慣用句の具体例を照らしながら第1図の3軸モデルを検証する。

共起性のみ①には、名詞が連続して出現する現象が該当するとみられる。たとえば英語の “political situation” に対応する「政治情勢」において、「政治」という単語と「情勢」という単語の組み合わせは「政治状況」「*政治場合」の組合せよりも高い共起性をもつ。この例において、なぜ共起しやすいかを論理的に説明するのは困難である。また、名詞の連続に関する通常の言語制約以外のものはもたず、構成単語のセットから計算不可能な新概念が発生しているわけではない。

次に、②であるが、組み合わせが固定した共起性を第1図の共起性のカテゴリからははずしたことから、これに属する例は存在しないものと予想される。すなわち特殊な統語的振る舞いの起こらない、①の例に示した複合語のような共起において新概念が発生した場合は共起は固定されまい、次に述べる③の1単語の慣用句というカテゴリに入ると考えられるからである。

共起が完全に固定化するか、あるいは省略が起って1単語となり、決して分離せずに現れるようになったのが③である。きまりことばに分類されるものや(例:「有り難う(く思います)」「お疲れさまでした」)、連語・名詞句(例:『根まわし』『艶持ち』『犬猿の仲』)が該当する。

大多数のいわゆる慣用句は④に該当すると考えられる。体(かた)言葉の慣用句を考えてみても、「口が悪い(品の悪い言葉で遠慮無く批判する)」「口をはさむ(会話に割り込む)」「口にする(試食する/考えを言葉に出す)」等、各構成要素間に完全に固定した組み合わせがある。また各要素は互いに分離して現れるが(例:「口にもせずに」)統語操作には制約がある(例: *「はさんだ口」)。また、いわゆる日本語の拡張付属語の多くは、共起の組み合わせが固定していて新概念を生じ、且つ1単語でもなく一般の句とも違う統語的振る舞いをする(例:「たことがある」)。

⑤に属す例として、残りの一部の拡張付属語があげられる。たとえば、「場合がある」「ことがある」という可能性の意味を表す助動詞は、構成要素間の高い相互共起性と特殊な統語制約（「*あったこと」「ok こともある」）とを備えている。この他、国広（文献[9]）による形式的慣用句なども該当する。

例：「AはAで（晴れの日が晴れの日で）」。

「AはAとして（その話はその話として）」句としての特別な概念を派生してはいるが、構成要素の助詞類には固定化した共起は無い。一方で自立語部分の同一性という特殊なシンタックスを要求する。

⑥には代動詞類が該当する。名詞に依存して表層語の定まるタイプの代動詞類は「新概念の発生」ではなく、一部構成要素の語義が消失し、一部は本来の語義がそのまま残ったものとみなすことができる。さらに構成要素の並列構文化等において特殊な統語的ふるまいをすることから、これらの代動詞類は⑥のカテゴリに分類するのが妥当と考えられる。

例：攻撃を/考察を [加える]

⑦は、慣用句の範疇からははずれるとみなされるが、強いてあげれば連体埋め込み文の主名詞となったときに「が」格や「を」格を埋めない関係体言「中」「後」（文献[6]）を分類することができる。

4 慣用句定義の周辺

慣用句には比喩から派生して用法が固定化したとみられるものが多数存在する。しかし直喩（例：その話は夢のようだ）の大半は明らかに用法が固定化していないものであり、比喩の全てが慣用句というわけではない。そこでこのような観点で慣用句の定義の境界線を考察した。以下、第1図の右側部分を参照して説明する。

比喩にも何種類かのカテゴリが考えられるが、ここでは、比喩とはあるモノ概念のもつ属性を別のモノ概念を用いて表現したものと定義する。
例：彼女の頬はリンゴだ。彼女の心は海だ。（第1図⑩）
用いられる語の組み合わせが固定化し、表面に現れない属性表現が全体の語義となった場合、それは「新概念発生」に該当する。

「間接発話行為」の典型例は、「あそこに居酒屋があるね」という文によって「今から一杯行こうよ」という勧誘の意味を伝達するという類である。直接の言葉面に現れない内容が真の意味であることを特徴とするのであるから、用法が固定化した場合は「新概念の発生」に結びつくと考えられる。

例：彼は頭が痛い。（私にとって困る）（第1図⑩）
この他の領域に分類した例を第2図に示す。

5 慣用的な意味と構成的な意味との切分け

解析システムが慣用句を認定する際にはまず形態素解析モジュールで全ての構成要素が揃っていることを検出し（完備性のチェック）、さらに何種類かの解析モジュールによって、その語彙のもつ構文的・意味的な制約を検査することによって、すべての必要条件の充足を調べる。

さらに曖昧性がある場合、これらの間の選択に有効な探索条件としては、次のようなものを考えている。

- 1) 格パターンの異なり、
- 2) 格スロットの選択制約、
- 3) 統語制約という必要条件の違反、
- 4) 並列パターンの検出
- 5) 共起コストの比較、

- 領域⑧ → 情けは人のためならず。（自分自身のためである）
領域⑨ → 顔から火が出る思いだ。（その場で非常に恥ずかしい）
根も葉も無い話だ。（何の根拠ももたない話し）
領域⑩ → 彼は頭が痛い。（私にとって困った奴だ）
領域⑪ → 私は理じゃない。（私を信用してくれ）
領域⑫ → あそこに居酒屋があるね。（お酒を飲みたい）
領域⑬ → 彼云の頬はリンゴだ。彼女の心は海だ。

第2図 比喩・間接発話行為の関わる分類の具体例

- i) 「我々がその件を考えたことがある。」 → 慣用句の語義「経験」
「その件に関して我々が考えたことがある。」 → 曖昧性残存
ii) 「一度口に出した言葉は2度と元に戻らない。」 → 慣用句の語義
「固い牛肉を一巨喉から口に出してから飲み込んだ。」 → 構成的意味

第3図 慣用句を構成単語を含む文の曖昧性解消例

第3図に挙げた3つの例について説明すると、まず i) の成句『ことがある』は1)の格パターンの異なりによってある程度意味を絞ることができる。即ち、経験という場合にのみ『を』格をとり得ることから、『を』格が出現している上の例の語義が1つに決まる。次に、ii)の成句『口に出す』は2)の選択制約を対象格について適用することにより、より尤もらしい語義を選択することができる。

これらの比較的強い制約だけでは、慣用的な意味と通常の構成的な意味との切り分けが十分に行えなかった場合、「構成要素が全部揃っているからデフォルトで慣用的な意味のほうに決める」というヒューリスティックスが有効であると予想される。

6 おわりに

いわゆる慣用句の定義を3軸で捉える枠組を提案し、多種類見出された表層の現象をこの枠組に基づいて分類した。さらに意味の曖昧性解消の条件について考察した。この結果、慣用句の語彙の知識表現の種類をいわずらに増加させることなく、3軸の素性の組み合わせによって多種類の言語現象を定式化できるという見通しが得られた。また分類された具体例によれば、①③④⑥⑦のカテゴリが十分多数の語彙をもっていることが示唆された。

今後は、これらの分析結果に基づいて具体的な知識表現および曖昧性解消手順を設計していく。

謝辞

本研究の機会を与えて下さった日本電気C&C情報研究所村木一至課長に感謝します。

参考文献

- [1] 宮地：慣用句の意味と用法，明治書院，1982
- [2] 首藤：日本語における固定的複合表現，
文部省科研費特定研究#63101005，1989
- [3] 小倉他：慣用句を利用した形態素解析情報収集法，
情処第40回大会，1990
- [4] 山梨：比喩と言語理解，東大出版会，1988
- [5] 市山他：意味表現における関係概念，
情処第40回大会，1990
- [6] 野村他：機械翻訳システムPIVOTにおける日本語格フレームモデル，情処第38回大会，1990
- [7] 稲村，Stein他：ランダムハウス英和辞典
パーソナル版第4刷，小学館，1980
- [8] 新村：広辞苑第3版，岩波，1983
- [9] 国広他：日本語百科大事典，大修館，1986