

# 定型から逸脱した言語表現の分析

土屋智行

京都大学大学院 / 日本学術振興会特別研究員

tsuchiya@hi.h.kyoto-u.ac.jp

## 1 はじめに

### 1.1 定型からの逸脱現象

特定の言語コミュニティにおいて一定の定型性と認知度を保つ言語表現は、そのまま全体として発話される場合もあれば、パロディ表現など一定の逸脱を伴って発話される場合も存在する。その逸脱は内容語、機能語に関わらず、発話文脈と創造性に任せて比較的自由におこなわれている。たとえば「学問に王道なし」という本来の言語形式から、下の例のような逸脱用法がおこなわれる。

- (1) 学問に近道はない
- (2) ダイエットに王道はない
- (3) 英語学習に王道なし

人間はこのような逸脱した言語表現を受け取ったとき、即座に元の言語表現が何であったかを照合し、かつ元の言語表現からの意味的な差異を認識できる。では、自由な逸脱が行われる膨大な言語表現の中で、元の言語表現を特定するための有効な手法は何であろうか。

本稿では、このような定型から逸脱した言語表現の傾向を記述することで、逸脱の範囲とその傾向を確認する。同時に、元の言語表現から継承される超語彙的パターンの存在を指摘し、逸脱用法の抽出と生成に関わる具体的な展望を述べる。

### 1.2 超語彙的パターンを想定した分析の必要性

複数の語が既成的に定着している表現の例として慣用句が挙げられる。[2]は対象とする言語表現の依存構造を解析し、慣用句と照合するものを抽出する手法を用いている。この手法では、形態素の挿入などの形態的变化には高い確率で対応できているものの、内容

語の入れ替えが生じた場合、抽出が困難であるという問題が生じる。今回取り扱う逸脱用法の場合、原則的にあらゆる逸脱パターンが想定されうる。同時に、元の語がどれだけ割合で残されているかで違いがあるとはいえ、どのような逸脱パターンであっても、元の言語表現を喚起する可能性が存在する。つまり、元の言語表現の喚起を可能とするかという点において、定型的な表現の逸脱用法には、慣用句の形態的变化にあるような内容語・機能語の明確な貢献度の違いが存在していない。

このような特徴を持つ「逸脱」という言語現象において、元の言語表現を特定するためには、まずはその範囲の実態を把握する必要がある。今回は[3]のパターン束分析の手法を用いることで、逸脱用法の実態と、元の言語表現の内部構造の分析を試みる。これには、想定されるあらゆる超語彙的パターンへ探索の対象を拡大することで、依存構造解析によって生じるような情報の損失<sup>1</sup>および探索すべきパターンの不足を回避する狙いがある。

## 2 逸脱用法の分析

### 2.1 分析の手法

本稿の分析では、まず複数の定型的かつ日本語話者にとって認知度の高い表現を選び、それらの逸脱用法を収集した。その後、それぞれの逸脱用法のパターンを具体的に見ていくことで、その範囲を分析した。範囲の分析には二通りある。ひとつは、「どれだけ拡張が容認されるのか」という広がり分析であるのに対し、もうひとつは「元の言語表現がどれだけ維持されているのか」という保持分析である。

まず本稿では「学問に王道なし」のような「NP1にNP2なし」パターン4例を調査対象とした。

<sup>1</sup>依存構造解析の情報損失については[1]を参照。

- (4) 学問に王道なし，触らぬ神にたたりなし，死人に口なし，看板に偽りなし

分析の対象としたコーパスは，Sketch Engine の Jp-WaC コーパスである．全体のサイズは約4億語で，ウェブページ上の日本語を収集しているため，逸脱用法は比較的多く観察される．まずは扱う用例の一部を空所化したパターンを作成し，各々のパターンに当てはまる事例をコーパスから収集した．空所化のパターンは元の表現の語を2以上保持している全ての表現とし，1語のみしか保持していないものは収集の対象とはしていない．その後収集した事例に対し，当該の用例の逸脱用法であると判断されるものを人手で抽出した．

「逸脱であるか否か」の判定であるが，元の言語表現を想起できるかという点を焦点とした．具体的には，元の言語表現が用いられる語用論的な文脈が共通している場合や，その文脈を想定していると感じた場合とした．たとえば5の例は「学問に王道なし」と継続的な努力の必要性を説く文脈という点で共通している．対して6の例は，その文脈を想定した上で，その逆の主張をさらに行っていると考える．最終的な判定に関してアドホックな部分が残されていることは否定できないが，逸脱用法は元の言語表現の文脈を前提としているという点を考慮した上で，このような判定基準とした．

- (5) 勉学に王道はない  
(6) 謝罪法に王道あり

## 2.2 どれだけ逸脱が行われるか

収集の結果は以下のとおりである．逸脱用法は，最終的に全体で90例<sup>2</sup>確認することができた．用例の右に表示されている括弧内( )の数字は，収集できた事例の数を表している．

学問に王道なし(34例)

- (a) 内容語の拡張:[学問 ダイエット, 子育て, 営業, 恋愛, 等][王道 近道, 横道]  
(b) 助詞挿入:学問に王道{は|が}ない

<sup>2</sup>参考までに，元の言語表現のコーパス内での生起頻度を以下に示しておく．

学問に王道なし(10例)  
触らぬ神にたたりなし(21例)  
死人に口なし(14例)  
看板に偽りなし(6例)

触らぬ神にたたりなし(15例)

- (a) 内容語の拡張:神 猫, 機械, ○○, 等  
(b) 句の省略:たたりなし 何とやら, 何とか  
(c) 音韻変化:たたり 当たり  
(d) その他:「触らぬ祟り神化」した人々

死人に口なし(4例)

- (a) 内容語の拡張:[死人 ほとけ][なし あり]  
(b) 音韻変化:口 グッチ

看板に偽りなし(37例)

- (a) 内容語の拡張:[看板 キャッチコピー, 肩書き, タイトル, 等][偽り はずれ, 嘘偽り][なし ない, あり]  
(b) 助詞挿入:看板に偽り{は|が|の}ない  
(c) その他:メル・ギブソンにはずれなし

「N1にN2なし」の表現の場合，他にも助詞の挿入や内容語の拡張が広く観察できる．また前半部や後半部にあたる句の省略や，いわゆる駄洒落にあたる音韻の変化も見られる．加えて興味深いことに，調査の対象とはしていないが，一定の定型性と認知度をもつと判断されるような事例も多く観察できた．7にその例を示す．

- (7) 職業に貴賤なし，構造改革に聖域なし，空き腹にまずいものなし  
(8) 乗客にけが人なし，かばんに財布なし

これらの表現は，一定の高さの定型性と認知度を持つ言語表現から，特定の発話文脈を表す抽象的な言語表現が新たに発話され，定着する可能性を示唆している．しかし，「N1にN2なし」の言語形式は，必ずしも7のように定型性が高くかつ認知度の高い表現に限られるものではなく，8のように，定型的な表現ではあまり使われないような言語表現も観察される．この要因については，今後具体的な分析と考察を重ねる必要がある．

## 2.3 どれだけ元の表現が保持されているか

次に，収集した事例を Pattern Lattice Builder (以下 PLB) を用いて，元の言語表現のパターンがどれだけ保持されているかを分析した．本稿では，紙面の都合上「死人に口なし」のパターン束を用いて説明する．

PLB では図 1 のように、空所化された個々のパターンのより下位の（空所が 1 箇所少ない）パターンとの接続数を表示する。今回は空所が 2 箇所存在するパターンと、それが接続している下位のパターン（ノード）との数を対象とした。

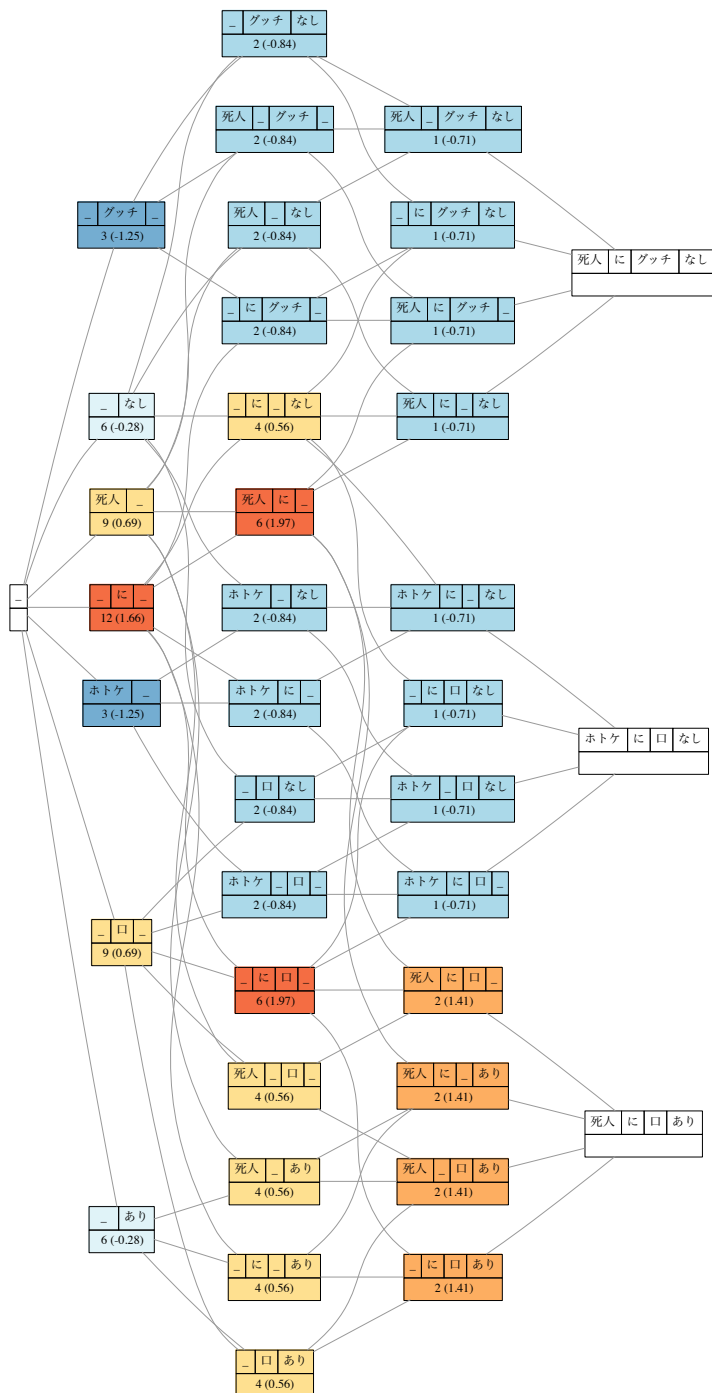


図 1: 「死人に口なし」のパターン束

以下では全体の語数が同じで、逸脱用法の事例数が近似している「学問に王道なし」と「看板に偽りなし」を見ていく。以下に元の言語表現の空所化のパターン

および接続ノード数を示す。

学問に王道なし		看板に偽りなし	
Pattern	Nodes	Pattern	Nodes
XY 王道なし	34	XY 偽りなし	8
X に Y なし	43	X に Y なし	25
X に王道 Y	72	X に偽り Y	64
学問 XY なし	0	看板 XY なし	0
学問 X 王道 Y	9	看板 X 偽り Y	53
学問に XY	12	看板に XY	53

表が示すように、「X に Y なし」や「X に王道 Y」「X に偽り Y」のパターンが比較的多く観察できるのに対し、「学問 XY なし」「看板 XY なし」のパターンの逸脱用法が観察されなかった。あらゆるパターンが想定されるにも関わらず、出現する事例は分散しないことが確認できる。

## 2.4 分析のまとめ

以上から、定型性と認知度の高い言語表現に対し、様々な逸脱用法が観察される一方で、話者にとって逸脱しやすいパターンとそうではないパターンで偏りが生じていることが確認できる。これは、聞き手側にとって元の定型表現を喚起しやすいパターンが存在しているということもできる。では、この超語彙的パターンの存在は理論的にどのような含意があると考えられるだろうか。

逸脱用法に頻繁に現れるパターンは、そうではないパターンに比べて元の定型表現を喚起する可能性が高いということは、そのようなパターンは元の定型表現にとって重要なパターンであり、それだけ元の定型表現に意味的に貢献している可能性を示唆している。(意味的に貢献していないパターンが定型表現そのものにとって重要であるとは通常考えにくい。)

## 3 今後の課題と展望

### 3.1 逸脱の範囲と意味的な乖離

逸脱用法のパリエーションに関して、意味的な側面からの分析の可能性が残されている。今回観察した逸脱用法に用いられる語の多くは、元の言語表現との共通点を見出すことが可能である。例えば「学問」に対し「数学」は意味的に上位と下位の関係に位置する。

「学問」と「ダイエット」は意味的に大きく異なるが、「Xに王道なし」という言語表現に埋め込まれた語として考えた場合、両者は 継続した努力を必要とする行為 という点で共通した意味的機能を持つ。

また、今回は特定のパターンの逸脱事例が結果的に観察されなかったが、それは人間による逸脱用法の生成をおこなう際にそのパターンを用いるのが相対的に難しいことを示しているに過ぎず、このパターンによる逸脱が全く行われないうことは言えない。実際に人間にとって「学問 XY なし」あるいは「看板 XY なし」のパターンの逸脱用法の生成は比較的容易である。また、図 1 にもあるように、「死人に口なし」の場合は「死人 XY なし」の逸脱用法が確認できる。

今後定型性と認知度の高い言語表現の逸脱用法を観察することで、個々の表現の逸脱用法の出現パターンの予測を行うことは可能となるだろう。ただし、上に挙げたように、逸脱に用いられる語と元の語との乖離を具体的に記述する必要がある。これには、まだ具体的な手段は明確ではないが、逸脱に用いられる語と元の語の意味だけではなく使用実態も含めた記述が必要であると考えられる。

逸脱に用いられる語の選択制限のモデル化は、将来解決すべき大きな課題となるだろう。

### 3.2 パターン束では収集できない事例

もう一点具体的な問題として、スクランプリング現象が挙げられる。今回は具体的な分析までは行っていないものではないが、「N1 で N2 を V」の形式の逸脱用法では 9 に対して 10 の例のように語または句同士が入れ替わっているものがある。

(9) 血で血を洗う、真綿で首を絞める

(10) 血を血で洗う、首を真綿で絞める

本稿での分析の枠組みは、2 語以上の配列から逸脱用法を抽出する方法であるため、たとえ使用されている語が同じであっても、語の配列が元の言語表現と異なれば抽出が困難である。したがって、例に挙げたように「真綿 / で」の配列「首 / を」の配列が元の言語表現と共通している場合は抽出が可能となる。しかし、たとえば 11 のような事例が存在した場合は、たとえ共通した語が出現していても抽出が困難となる。一方人間はこのような言語表現を見たときには、元の言語表現である「看板に偽りなし」を喚起できる。

(11) 偽り看板が出ている

この問題には、先ほどのそれぞれの超語彙的パターンの意味的な貢献度に加えて、語単独の意味的貢献度にも配慮する必要があるだろう。逸脱用法は内容語・機能語の区別なく様々なパターンが想定されるが、内容語と機能語の根本的な意味の違いの存在は否定できず、意味的な貢献度とはまた別のものと考えなければならない。この点については、定性的な研究に大きく依頼するところである。

### 参考文献

- [1] K. KURODA. Arguments for Parallel Distributed Parsing. *Proceedings of the 24th PACLIC*, pages 455–462.
- [2] 橋本力, 佐藤理史, and 宇津呂武仁. 依存構造照合に基づく慣用句自動検出. 言語処理学会第 12 回年次大会発表論文集, pages 829–832, 2006.
- [3] 黒田航 and 長谷部陽一郎. Pattern Lattice を使った (ヒトの) 言語知識と処理のモデル化. 言語処理学会第 15 回大会発表論文集, pages 670–673, 2009.