

国立国語研究所学術情報リポジトリ

Characteristic Elements of Words and Their Frequencies: The Case of Tsuno ‘Horns’

メタデータ	言語: jpn 出版者: 公開日: 2018-01-29 キーワード (Ja): キーワード (En): 作成者: 加藤, 祥, KATO, Sachi メールアドレス: 所属:
URL	https://doi.org/10.15084/00001412

特徴的な要素と用例頻度の関係

——角を例とした一考察——

加藤 祥

国立国語研究所 コーパス開発センター 非常勤研究員

要旨

コーパスの頻度情報は有用なデータであり、COBUILD やウィズダム英和辞典などの辞書に語や意味の重要度の指標として活用されている。ある対象物に関する様々な要素のうち重要なものは、テキストにおいて高頻度で言及されている可能性が高い。動物の身体部位語の頻度を調査したところ、ある動物において特徴的と考えられる角のような要素の頻度が高い傾向が見られた。また、対象物の有する要素とその頻度分布情報から、対象物を認識することも可能という実験結果も得られた。我々は対照する他物との差異となり得る特徴的な要素に着目し、それらが高頻度であることを期待する。しかし、高頻度であると期待される要素が、必ずしも高頻度で言及されていない場合がある。たとえば、それぞれ馬と人との差異として角を有するユニコーンと鬼を見ると、ユニコーンの角は期待通りの高頻度で言及されるが、鬼の角は頻度が低い。期待される頻度と実頻度に差の生じる一因は、用例において比喩表現に現れていた。外観上特徴的な要素は、形状を表す喩辞として用いられる傾向がある。ゆえに、固定的なイメージがない場合には比喩表現として用いられにくい。また、対照されやすい他動物が被喩辞となる比喩表現では、差異となる要素こそあえて言及する必要がない。このように、特徴的な要素と用例頻度の関係には、比喩表現のような表現形式が関わるため、頻度情報を用いる際には考慮が必要である*。

キーワード：コーパス、用例頻度、特徴的要素、比喩、角

1. はじめに

ある対象物において特徴的と考えられる要素は、コーパスにおいて高頻度に取得可能なものであろうか。テキストから対象物情報を集めるにあたり、コーパスは大変有用である。Sinclair(1991)が大量のテキストの有用性を示した通り、大規模なテキストコーパスからは、量的なデータを得ることができる。また、たとえば McEnergy and Hardie (2012) は頻度情報の有用性を示しており、コーパスの量的情報は様々な活用されている。COBUILD やウィズダム英和辞典をはじめとする多くの辞書は、コーパスの頻度情報に基づき、語や語義の重要度をも提供している。語や語義の重要度が頻度情報として取得できると考えられているためである。この種の頻度の利用法は、多義語の場合に語義相互間の意味関係が捉え難くなるという指摘もある(国広 2000 など)が、このような辞書を使用すれば、見出し語に菱形や星形で示した重要度として頻度情報を確認することができるので、重要な語や語義を効率的に学習することが可能となるだろう。

* 本稿は、13th International Cognitive Linguistic Conference (2015, 於 Northumbria University) での発表内容「The effect of metaphor on frequency of usage; Horns are mentioned more for unicorns, less for devils」を基に、大幅なデータの追加による再調査を行い、文章化したものである。本研究は JSPS 科研費 26770156 の助成を受けている。また、本研究の一部は国立国語研究所コーパス開発センターの共同研究プロジェクト「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」(2016–2021 年度) の成果である。

このように、ある対象物において重要と考えられる要素は、言及される機会の多いことが予想される。たとえば、加藤ほか (2015) の調査では、兎に関する用例において、耳が最高頻度の要素であった。反対に、目立つこともなく意識されることの少ない兎のヒゲのような要素は、ほとんど用例が取得できなかった。ある対象物に関する様々な要素の頻度差は、対象物において特徴的であるかどうかという点に関わることが考えられる。よって、ある対象物について調査する際、コーパスにおいて用例頻度の高い語が、ある対象物の特徴的な要素であることが期待される。

それでは、言語化される要素、すなわちコーパスから取得される用例の頻度は、一般的な知識として特徴的であると期待される通りなのだろうか。実際の用例頻度と認識 (期待する頻度) に差があるとすれば、なぜその差は生じるのだろうか。

本稿では、言及頻度と特徴的な要素との関係を確認することを目指す。そのため、対象物として動物を取り上げ、その外観的要素について、コーパスにおける言及頻度の調査を行うこととした。調査の結果、特徴的な要素は高頻度で言及されていることが期待された。このことを検証するため、我々は上位頻度の要素とその用例頻度に基づいて対象物が同定できるのかという実験を行った。この実験によって、特徴的と考えられる要素が期待通りの頻度で言及されていないと考えられる場合が見つかった。そこで、実際のコーパス頻度と期待された頻度が異なっていた対象物とその要素についての用例を分析し、何がこの現象の起因となっているのか、対象物の特徴的な要素と言及頻度の関係を考察した。この結果、特徴的な要素が言及されないという現象が見られる場合、比喩表現の用例に特性が見られるとわかった。

2. 動物とその外観的な身体部位の出現頻度

まず、動物とそれらの身体部位について、コーパスの出現頻度を調べる。動物に関する外観的な情報は、先行研究における実験を用いた調査結果がその有用性を示している。10 種類の国語辞書の語釈において、200 種類の動物について 5 冊以上の辞書に記述のあった要素を調査した結果では、36.7% と最も多くの要素が取得されたのは形態情報であった (加藤 2015)。動物であれば耳・口・尾などの外観的な要素に関する記述がそれにあたる。McRae et al. (2005) の連想調査結果の動物について見ると、「visual-form and surface」と分類された要素 (feature) が 31.3% と最も多い。連想実験においても辞書記述に類似した結果が得られているということであり、外観的に特徴的な情報をもっとも連想されやすかったと考えられる。これらの結果から、各動物と外観的な身体部位の用例を収集することとした。

2.1 データ

大規模 Web コーパスの Google n-gram と Google Japanese n-gram を使い、英語と日本語の動物の身体部位用例を収集した。用例の収集にあたっては、英語は 40 種、日本語は 140 種の動物と 55 種の外観的な身体部位を組み合わせた。たとえば、「犬の足」や「猫の肉球」などである。英語で 2,853,348 例 (動物平均 71,334 例) と日本語で 3,511,421 例 (動物平均 25,082 例) を収集した。

これらの用例からは、概ね特徴的と考える身体部位が高い頻度で得られるという結果が得られ

た。たとえば象の身体部位頻度（102,282例）は、高頻度順に耳が21%、背が18%、鼻が17%などとなっている。耳と背と鼻は象の特徴的要素といえよう。また、他動物との差異となるような要素が多く記述される傾向が考えられた¹。たとえば、ユニコーンは角のある馬という架空生物である。ユニコーンの用例では、角が馬との差異として多くの割合で現れている。コーパス用例におけるユニコーンの身体部位は、ほとんどが角である。日本語のユニコーンの用例では95%が角であった（後掲の図5参照）。

但し英語では、ユニコーンの身体部位用例は67%が角であり、英語と日本語でユニコーンの角の言及割合に差が見られる。最高頻度で言及されている点は同様であり、他動物との差異となり得る特徴的な要素の言及頻度が高いという傾向は、日本語にのみ特有な現象ではないと考えられるが、普遍的な現象ではない可能性が残ったため、次節で日英語コーパス間の差異について確認しておく。事例の確認には、『現代日本語書き言葉均衡コーパス（BCCWJ）』を用いる。

2.2 データの言語横断的な普遍性

ユニコーンの身体部位の言及頻度を見たとき、日英語で角の割合に差があった。特徴的な要素の言及頻度が高くなる傾向に言語による差が生じているようであれば、ある要素を特徴的と考えるかどうかという判断に言語横断的な普遍性がないという可能性も考えられる。英語と日本語のコーパスから取得した用例の違いを確認したい。

結論として、英語と日本語のコーパスから取得した用例に差が見られる場合でも、身体部位の言及頻度やその部位ごとの順序は、両言語のコーパスにおいて多くの動物で類似している。たとえば、日英語のどちらの言語でも、カメレオンの高頻度言及部位は舌と目である。高頻度の要素は英語も日本語も類似しているといえ、ある対象物に関する外観的な要素のうち、特徴的な要素の判断には普遍性がありそうである。

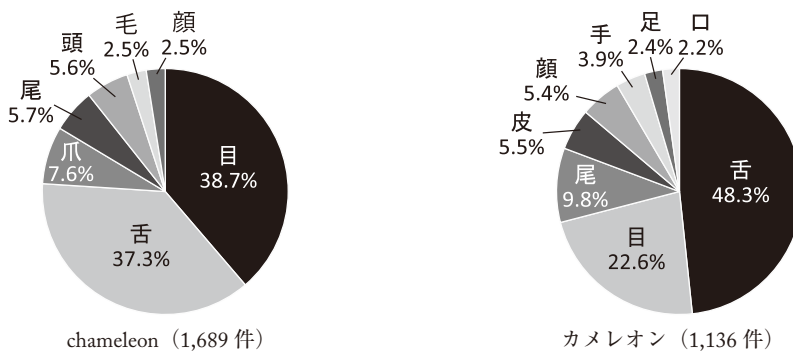


図1 chameleonとカメレオンの部位頻度分布

¹ 加藤（2015）は、対象物認識のために、親カテゴリのプロトタイプや同カテゴリの他メンバーとの差異の記述が有用となることを示した。差異となる部分は、対象物の有する特徴的な要素であることが考えられる。

それでは、両言語で差異が見られる動物はどのような身体部位がどのように異なっているのか。ユニコーンと同様に、両言語で言及部位や頻度の異なる例を探した。

結果、両言語で言及部位や頻度が異なるのは、英語と日本語で指示対象が全く同じ範囲ではない場合か、対象物が言語で異なっている場合であるとわかった。たとえば、どちらの言語でも概ね同じ語として対応しているようであっても、日本語の鶏は概ね鳥そのもの（食用を含む）を指し、英語の chicken が概ね食用の鳥の肉を指すので、両言語における鶏の身体部位の言及頻度は、主に食べるものとして言及するかそのものについても言及するかで異なってくる。これは鮭（と salmon）や鮪（と tuna）などでも同様である。

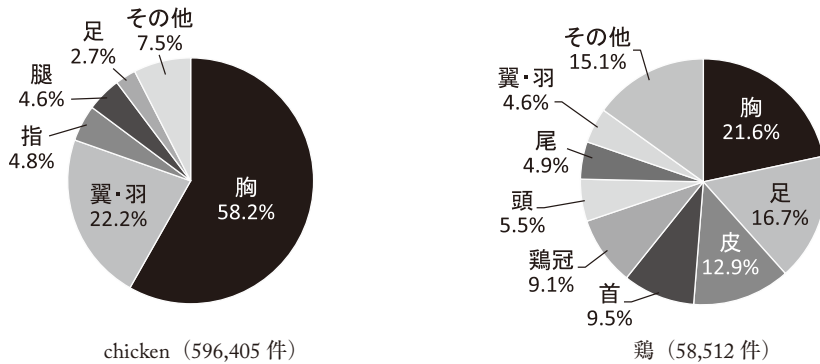


図2 chicken と鶏の部位頻度分布

また、架空生物では、文化差によって要素の頻度が異なるといえる。人魚は、日本語で鱗（43%）が高頻度であるほか、足や髪が上位頻度で見られるのに対し、英語ではいずれも言及されていない（0%）。人口に膾炙した物語内容が異なるなどにより、母語話者の有するイメージに差があるものと考えられる。このことの反映として、日本の人魚の絵では鱗が描かれるが、人魚姫の絵や彫刻には鱗がない傾向がある。

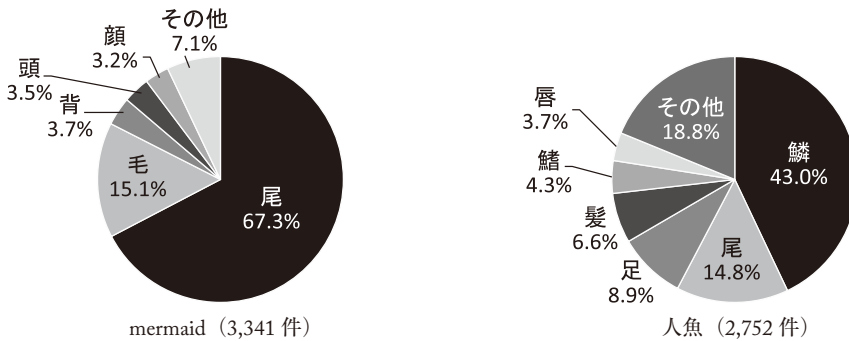


図3 mermaid と人魚の部位頻度分布

用例を確認すると、このほかの架空生物についても、日本語のユニコーンは海外小説の翻訳作品に現れることが多い(例 12)という出現傾向が見られ、ユニコーンのイメージが欧米ほどには浸透していない可能性が考えられる。人魚やユニコーンなどは架空生物であるため、文化差の点において言及頻度に差が生じやすいようである。

- (1) 「狼男がユニコーンを殺すなんてことあり得るの？」とハリーは聞いてみた。
「あいつらはそんなに速くない。ユニコーンを捕まえるのはたやすいことじゃない。強い魔力を持った生き物なんじゃよ。ユニコーンが怪我したなんてこたあ、俺はいままで聞いたことがないな」

(BCCWJ Sample ID: LBr9_00032, J・K・ローリング (著) / 松岡佑子 (訳)
「ハリー・ポッターと賢者の石」)

なお、このような日英語コーパスに見られる言及頻度差が、必ずしも現代人の感覚に当てはまっているともいえないという問題がある。日本語コーパスの用例頻度が、必ずしも日本語話者に期待される頻度ではないからである。以下では日英語コーパスと日本語話者の期待する頻度の関係を確認しておきたい。

2.3 用例頻度と日本語話者の認識

現代の日本人は日本語コーパスなどの頻度情報が常に正しいとは感じておらず、たとえば英語コーパスの頻度情報を自然と感じる可能性が考えられる。Yahoo! クラウドソーシング²上にて、1,000名の日本語話者(20歳以上)に同じ動物に関する日本語と英語の動物身体部位頻度グラフ(図 1, 2, 3 参照)を提示し、どちらが対象物の頻度分布と考えるか選択させる実験を行った。結果、表 1 (次頁)に示したように、日本語話者は日本語グラフを選ばないこともあった。

表 1 に示す部位分布頻度の相関(積率・順位)のように、日英語で部位分布頻度が異なる場合もある。順位相関について、前節で見た鶏/chicken(順位相関 0.48)や人魚/mermaid(順位相関 0.36)、同様に文化差が大きいと推測される天使/angel(順位相関 0.52)では低い傾向が見られる。また、積率相関は、鬼/devil(積率相関 0.45)において低い。しかし、その他の一般的な動物に関しては、日英語で類似した(順位相関 0.75 以上)部位分布の頻度であると思われる。

しかし、部位分布の頻度が類似していても、一般的な動物で英語のグラフが選択される傾向が確認された。用例が過去に書かれたものであるという性質からも、コーパスから取得された用例頻度と現代日本人の認識に差の生じている可能性がある。但し、狼や竜のような童話に出てくる頻度の高い動物や、犬のように親しみがあると考えられる動物では日本語が選択されている。日本語話者に馴染みのある対象物に関しては、期待される頻度の用例が取得できるものとする。

² BCCWJ においては、たとえば Sample ID: LBr9_00032 が 11 件 / 68 件を占めており、同著者の他作品サンプルも取得されているなどの偏りが見られる。

表1 日英語頻度グラフの選好割合と日英語部位頻度分布の相関

	英語選択	日本語選択	積率相関	順位相関
狼 /wolf	11.1%	88.9%	0.93	0.76
竜 /dragon	19.7%	80.3%	0.94	0.72
鶏 /chicken	20.7%	79.3%	0.91	0.48
犬 /dog	25.8%	74.2%	0.98	0.81
天使 /angel	47.2%	52.8%	0.84	0.52
人魚 /mermaid	65.8%	34.2%	0.81	0.36
猫 /cat	67.6%	32.4%	0.93	0.80
鬼 /devil	85.7%	14.3%	0.45	0.64
平均	43.0%	57.0%	0.85	0.64

3. 特徴的的身体部位による対象物同定実験

3.1 実験

さて、特徴的な身体部位は高頻度なのだろうか。高頻度な身体部位に着目することで、我々は対象物を同定できるのだろうか。高頻度の特徴的要素から、その要素を有する対象物が連想できるのだろうか。

コーパスの頻度情報を用い、どの程度対象物が同定できるか、同定できるとすれば高頻度の要素に着目するのか、あるいは他のどのような要素に着目するのかを調査した。2節で収集したデータから作成した身体部位頻度グラフ(図1～3の右図参照)を提示し、同頻度グラフから対象物を同定する実験を行う。実験データは日本語コーパスデータのみとし、Google Japanese n-gramから得た主に単語親密度が上位である動物の身体部位データを用いた。

実験協力者は、1,000人のYahoo!日本語IDを有した20歳以上の男女をYahoo!クラウドソーシングで募集した。実験協力者には各2種類の動物のデータを提示して以下の2つの回答を依頼した。たとえば、「カバ／アザラシ」の身体部位の言及頻度グラフを提示し、「カバ」はどちらであると思うか、何を根拠に「カバ」と判断したのかを訊ね、実験協力者はグラフの番号(A)と判断に有用な部位やその比率(B)について答える。

Q(A): 対象物は2つのグラフのどちらか選ぶ。

Q(B): 同定にあたりどの身体部位(と比率)が判断に有用であったか答える。

これらの問いにより、目立つ身体部位を見れば対象物が同定可能であるのか(A)、また対象物を認識するにあたってどの要素(身体部位もしくは比率)が注目されるのか(B)を調査する。提示した円グラフ(正答)は、カバ・鶴などの12種類(内訳は表2を参照。対照した動物例については後述の表3を参照)であり、各動物における上位頻度の要素(割合)は表2に示した通りである。

また、2.3節の結果から、調査対象の親密度が回答に影響する可能性を考慮した。そのため、

本実験では対象物として親密度の非常に高い動物を主として扱うこととした。表2に示した12種類の単語親密度（天野・近藤1999, 天野ほか2008）の平均は、5.89（max：馬6.38, min：カバ3.53）である³。

表2 調査動物における上位頻度の要素（上位10位までの要素を各6種以上、網掛け部は10%以上）

	カバ	鶴	猿	フクロウ	マグロ	鬼	狸	馬	ペンギン	ヒラメ	犬	猫
用例数 (100%)	2,817 (100%)	8,739 (100%)	27,207 (100%)	7,943 (100%)	22,429 (100%)	119,006 (100%)	13,685 (100%)	131,934 (100%)	6,232 (100%)	2,048 (100%)	164,886 (100%)	551,746 (100%)
尾	1.6%	0.3%	12.2%	1.1%	27.9%	0.0%	35.0%	9.9%	1.7%	0.0%	8.8%	9.1%
手	0.0%	2.6%	23.4%	0.5%	5.0%	4.2%	1.1%	0.8%	3.6%	0.0%	4.8%	30.9%
耳	8.3%	0.3%	4.0%	4.7%	0.0%	0.1%	1.6%	19.5%	0.8%	1.4%	6.9%	3.7%
目	3.1%	4.4%	6.0%	44.4%	6.4%	11.5%	4.2%	5.3%	20.3%	41.7%	8.3%	15.7%
角	1.1%	0.0%	0.0%	0.0%	0.0%	3.9%	0.3%	0.3%	0.0%	0.0%	0.0%	0.0%
背	14.0%	3.6%	1.1%	0.9%	2.7%	2.1%	2.3%	5.8%	5.9%	0.0%	2.4%	1.0%
翼・羽	0.0%	23.7%	0.0%	15.0%	0.0%	0.0%	0.0%	0.0%	13.8%	0.0%	0.0%	0.1%
毛	0.0%	2.6%	5.7%	0.0%	0.0%	0.1%	6.3%	8.4%	3.7%	0.0%	12.7%	11.5%
顔	8.3%	4.9%	19.7%	14.4%	1.0%	7.0%	6.8%	6.0%	12.3%	15.4%	8.8%	3.9%
皮	2.4%	0.9%	0.7%	0.0%	6.9%	0.0%	32.9%	1.2%	1.8%	9.3%	1.3%	1.0%
頭	4.0%	9.9%	7.8%	7.2%	34.9%	1.0%	2.1%	5.7%	7.9%	6.7%	10.6%	1.2%
爪	0.0%	0.3%	0.4%	2.3%	0.0%	31.0%	0.0%	0.5%	0.5%	0.0%	2.1%	5.2%
首	0.7%	13.1%	1.1%	5.9%	0.1%	31.6%	0.4%	6.5%	2.2%	0.0%	4.7%	1.6%
足	4.8%	11.2%	2.2%	2.0%	0.1%	2.2%	1.3%	1.2%	16.0%	0.0%	7.5%	2.9%
鼻	4.6%	0.8%	0.4%	0.3%	0.0%	0.2%	0.2%	1.7%	0.0%	0.0%	5.4%	0.8%
口	27.8%	2.2%	1.0%	0.3%	0.6%	0.6%	0.4%	4.0%	2.5%	7.4%	2.9%	1.0%
嘴	0.0%	3.1%	0.0%	1.2%	0.0%	0.0%	0.0%	0.0%	1.6%	0.0%	0.0%	0.0%
肉球	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.6%	3.6%

3.2 結果

3.2.1 要素の言及頻度分布から対象物を同定可能か

平均 64.8%（max：カバ 94.3%, min：鬼 14.9%）の実験協力者が身体部位の頻度分布から対象物を同定可能であった。ほとんどの動物で、半数以上の実験協力者が身体部位の言及頻度から対象物を同定できたという結果である。但し、提示したグラフが2種類であることから、頻度分布の類似性が正答率に関係することが考えられる。そこで、表3には、対象物の正答率とともに、提示した2種類の動物ペア（正答・対照用）の身体部位頻度分布について、積率相関・順位相関をあわせて示した。

³ 後述する実験結果においても、単語親密度の低いカバ（3.53）が最も高い正答率であったことから、調査対象とした動物の単語親密度が回答に影響した可能性は低いとも考えられるが、単語親密度と用例に現れる文化的な馴染みは同種のものではないため、参考として考慮した。

表3 提示した2種類の動物の部位頻度分布相関と正答率

	積率相関	順位相関	正答率
カバ (アザラシ)	0.23	0.56	94.3%
鶴 (カモメ)	0.77	0.62	91.5%
狸 (リス)	0.67	0.60	84.1%
犬 (猪)	0.50	0.70	81.5%
猿 (ゴリラ)	0.69	0.79	79.1%
猫 (クマ)	0.81	0.90	63.5%
マグロ (ヘビ)	0.70	0.68	63.3%
フクロウ (カラス)	0.54	0.83	62.9%
馬 (カンガルー)	0.26	0.67	62.5%
ヒラメ (シャチ)	0.41	0.68	54.2%
ペンギン (インコ)	0.67	0.81	25.4%
鬼 (鹿)	0.09	0.68	14.9%
平均	0.53	0.71	64.8%

たとえば猫とクマのペアの積率相関は0.81、順位相関は0.90であり、どちらも高い相関があるため類似した頻度分布に見えろと考えられるが、正答率は63.5%と平均的であることがわかる。また、正答率が14.9%と極端に低くなった鬼と鹿のペアについては、積率相関は0.09、順位相関は0.68であり、類似した頻度分布とは言い難いが、正答は得られなかった。鬼と鹿のペアも、頻度分布の類似度としては、正答率が94.3%と最高値であったカバとアザラシのペア（積率相関0.23・順位相関0.56）と大差がないように見える。よって、頻度分布の類似性が必ず正答率に関係しているとは言い難い。それでは、言及頻度の分布が類似していないにもかかわらず、鬼のように多くが誤認されることとなった動物は、何が誤認の原因となったのだろうか。

3.2.2 対象物の認識にあたってどのような要素が注目されるか

注目された身体部位は対象物の特徴的要素と考えられる。判断の根拠として記述された身体部位名の回答数を図4に示す。正答・誤答ともに、鶴の首（68.7%）やヒラメの目（74.3%）など、判断時に着目される身体部位が集中する傾向があるとわかる。鶴の首（13.1%：表2）もヒラメの目（41.7%：表2）もともに高頻度の身体部位でもある。提示されたグラフにおいて高頻度の身体部位は着目しやすいはずであり、また頻度が高いことから特徴的と考えられる身体部位に注目し、対象物の同定を行っていたと推測される。各々の動物において着目される身体部位は、特徴的要素に集中する傾向があるといえよう。また、特徴的要素は他動物との差異になり得るので、高頻度の身体部位を見ることで概ね対象物が認識できると期待される。

着目された部位を見ると（図4）、正答率の高いカバ（表3参照、以下同様：94.3%）はほとんどが口（図4参照、以下同様：84.1%）である。同様に、半数以上が正答したフクロウ（同：62.9%）は目（同：73.0%）、馬（同：62.5%）はたてがみ（同：60.5%）が着目された。特徴的要素はコーパスにおいて高頻度な用例であり、実験協力者もまず高頻度項目に目を引かれ、特徴的要素に着目すると考えられる。

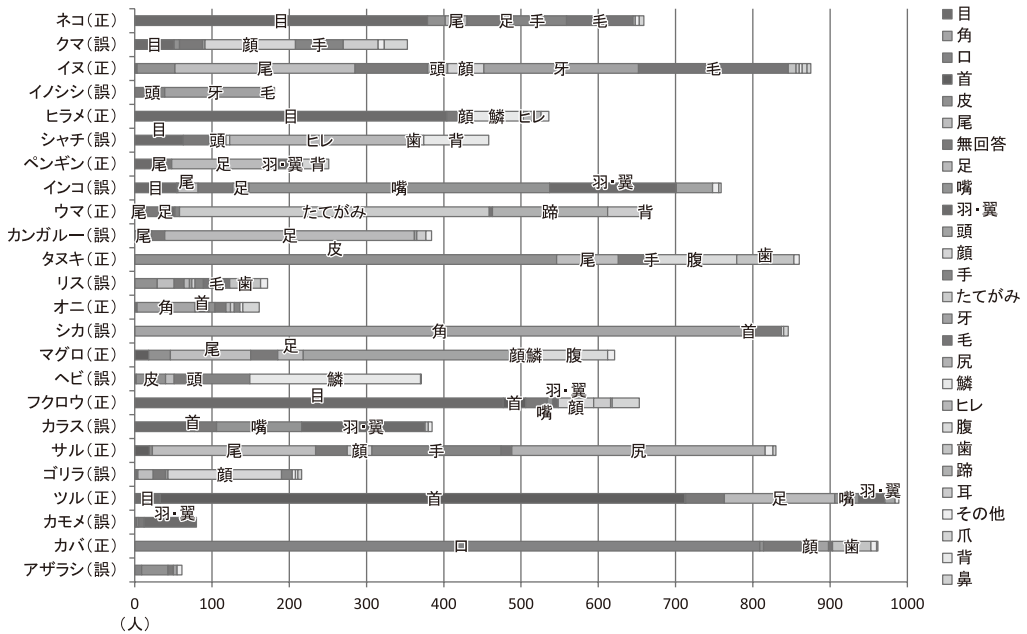


図4 対象物同定のために着目された身体部位 (回答数)

しかし、馬の身体部位では、着目されたのはたてがみであるが、高頻度の身体部位は耳（表2: 19.5%）と尾（同：9.9%）である。実験協力者は、これらの部位からは馬を他の動物と区別することができなかったため、頻度の高くないたてがみ（表2にはないが4.5%）に着目したのだと考えられる。また、馬はたてがみのような他動物との区別に有用な特徴的要素が見つかった例であるが、ペンギンとインコのペアの場合は、どちらの鳥も特徴的外観要素がなく、鳥でほぼすべての特徴が類似して言及されるため、区別に有用な特徴的要素が見つからない。頻度割合に若干の差はあっても、2種の鳥の差異が身体部位のみからは明確でない。このことによって、ペンギンの正答率が低い（表3：25.4%）という結果が現れているものと考えられる。

ある対象物の有する要素の言及頻度情報から対象物を認識しようとする場合、他との区別が可能な特徴的要素を探し、対象物を認識する手がかりとする。このような他との区別が一般に可能となる特徴的要素が高頻度である可能性は高いが、必ずしも高頻度の要素であるとは限らない。さて、低正答率（表3：14.9%）の鬼の場合も、鬼の角は鬼において高頻度な部位ではない（後掲の図5参照：3.9%）が、85.8%（図4）の実験協力者が特徴的要素として角に着目していた。しかし、対照する動物（実験では鹿）についても角が高頻度で言及されていたために、高頻度であることを期待して誤答が増加したと考えられる。それでは、なぜ鬼の角は期待される高頻度で言及されていなかったのだろうか。はたして、鬼の角は対照する動物との差異となり得ないのだろうか。特徴的な要素ではないのだろうか。

4. 「特徴的な」要素としての「角」用例

4.1 頻度情報と対象物認識の差

日本語母語話者は一般に角を持った鬼のイメージを有していると考えられる。鬼は300年ほど前から角が描かれる傾向にあり、描画実験(保田2011)でも、他の要素や形状は様々であるため鬼が画一的なイメージとは言い難いが、角は必ず描かれているという結果が見られた。しかし、コーパスの用例では鬼の角は目立たない(図5左)。用例においては「首(31.6%)」「爪(31.0%)」の頻度が高く、「角(3.9%)」の頻度は決して高くない。一般的な知識とコーパスの頻度データに差のある例といえる。鬼の誤答は鹿であったため、実験協力者はグラフの類似度ではなく鹿の角が高頻度(63.1%, 図5中央)であることに着目して鹿のグラフを鬼のものと同判断したのである(誤答4の場合の93.3%が「角」に着目したと回答した)。

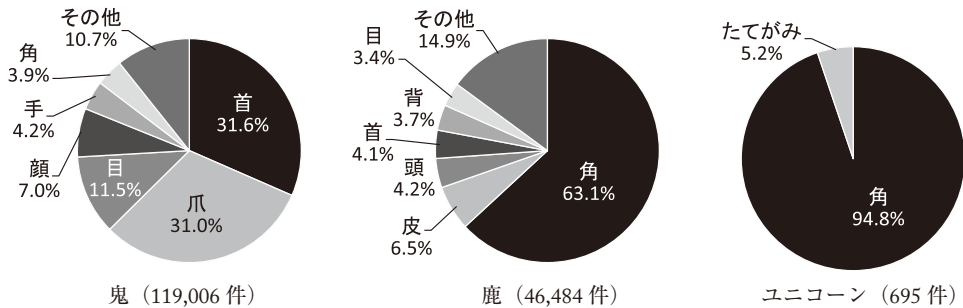


図5 鬼, 鹿, ユニコーンの部位頻度分布

加藤ほか(2015)は、動物の身体部位におけるテキスト世界と現実世界の差異についての分析結果として、3種類の差異とプロトタイプ効果を報告した。テキストと現実世界に生じる差異には、対象物が属するカテゴリにおけるカテゴリメンバーとの違いの度合いが関わっている。たとえば鬼の耳が高頻度で言及されるのは、他哺乳類と対照したときに顕著な特徴だからである。鬼のヒゲが言及されにくいのは、他哺乳類と違いが見受けられないからである。

既存の動物に何かを加えることで架空生物とする例においては、既存の動物と架空生物の区別は加えられた要素である。ユニコーンは角を持った馬である。そこで、ユニコーンの用例において、角が当然馬との差異として大きな割合(94.8%, 図5右)で言及されている。ユニコーンは角の他の部位が馬と共通のため、馬と共通する部位はほとんど言及されない。鬼の角の場合も、ユニコーンと馬同様、人とは角において異なるため、角の言及頻度はユニコーンに類して高頻度であることが期待される。とすれば、角の性質においてユニコーンと鬼の違いは何であろうか。なぜ特徴的要素が期待された頻度で言及されていない場合があるのだろうか。

⁴ 正答の場合に「角」に着目したという回答には「鬼の角はこの程度の割合だと思う」「角ばかりではないはず」などのコメントを加えた記述が散見され、高頻度であることに着目したのではないことが示される傾向があった。

4.2 動物的特徴的要素としての角

角はユニコーンや鬼の特徴的的身体部位と考えられるが、その他の動物においても典型的な身体部位となり得る。2.1 節のデータを集計し、動物の身体部位ごとの頻度を見ると、目（全身体部位データの 10.0%）、尾（以下同 9.7%）、翼（9.2%）など様々な動物において一般的な部位が頻出する。角（3.3%）の出現頻度も上位であり、角を有する動物においては特徴的な身体部位であるため、言及されやすいことがわかる。しかし、角は多くの動物が有しているのではない。牛や鹿のような一部の動物に特徴的な部位であるといえる。日本語でも英語でも、角（horn または antler）と共起する動物は、上位 10 位までの動物で用例の 95% を占めている。日英語ともに、鹿（deer, stag, moose, elk）、牛（bull, ox, cow, cattle）、羊（sheep, ram）、ヤギ、サイ、鬼（devil）、ユニコーン、兎（rabbit, hare, bunny）が角と共起する頻度の高い動物である（英語では野牛（buffalo, bison）と羚羊（antelope）、日本語では竜とキリン⁵が加わるという点が異なる）。また、日英語のどちらでも角との共起頻度の高かったユニコーン、兎（角のある兎としてジャッカロップ、角兎と呼ばれるものがある）、鬼は架空生物もしくはあり得ない意味である。これらの 3 種は現実の生物（馬、兎、人間）にそれぞれ角が追加された動物であるといえる。本節では、コーパスデータに基づいてユニコーンの角も確認、対照しながら、鬼の角の分析を試みたい。

用例の収集にあたっては、『現代日本語書き言葉均衡コーパス』（以降 BCCWJ）を主に用い、用例数の少ない場合などには『国語研日本語ウェブコーパス』（以降 NWJC）を補助として用いた。BCCWJ から取得した「角（語彙素・語彙素読みツノ）」の用例は 1,539 件であるが、誤解析⁶を排除し、604 件（394 サンプル）を角（ツノ）の用例として得た。意味分類は人手により前後 200 語の文脈から行ったが、キー（角）の前後 200 語から判断が困難な場合は「特定不能」とした。また、それぞれの意味分類で頻度を算出した。また、比喩性の判断についても同様の文脈から行い、喩辞となる鬼の前後 200 語から判断が困難な場合は「特定不能」とした。比喩性は、中村（1977）に挙げられた A 型把握（指標と文脈上の概念移行）、B 型把握（文脈における結合関係）、C 型把握（文脈との無縁性）のそれぞれにより判定した。判定にあたっては、同書の言語形式としての類型も参照した。

結果、BCCWJ における角と共起する動物としては、鹿（9.3%）、牛（9.0%）、カブトムシ（5.0%）、サイ（2.0%）、鬼（1.7%）、羊（1.5%）、ユニコーン（1.3%）が得られた。また、NWJC から「名詞+の（語彙素・助詞）+角（語彙素「角」、語彙素読み「ツノ）」として取得した用例は 1,817 例あり、うち「動物の角」は 659 例であった。上位頻度 10 位までの動物は、鹿（15.0%）、トナカイ（14.9%）、サイ（10.3%）、鬼（7.1%）、カブトムシ（6.7%）、牛（6.2%）、悪魔（5.6%）、羊（5.3%）、竜（4.4%）、ユニコーン（2.7%）であった。BCCWJ・NWJC とともに、2.1 節のデータ（Google n-gram）と同様の動物が角と共起している用例が取得できた。

また、BCCWJ から取得された特定動物の角と判断できない用例としては、何らかの材料（3%）、

⁵ 本稿の調査において ossicone（horn-like または antler-like protuberances）は扱っていない。

⁶ 語彙素「角」は「カド」となっている誤解析が多いため、「角（語彙素「角」、語彙素読み「カド）」となっている用例についても、人手により「ツノ」と判断された用例を抽出した。

泡立てたものの形状の比喩 (3%, 例 (2)), その他形状の比喩 (2%, 例 (3), (4)) などであった。特定動物の角でない場合の角の用例は, 材料のほか形状の比喩として用いられる例の頻度が高い。

- (2) グラニュー糖を 3 ~ 4 回に分けて加え、ピンと角が立つまでさらに泡立てます (写真 6)。
(BCCWJ Sample ID: PB15_00067)
- (3) 剛毛をオールバックになでつけているが、平定できない部分はピンと角のように両側におたっている。
(BCCWJ Sample ID: LBi9_00178)
- (4) 陸雄は、その間、小型筒型の懐中電燈二個を買い、それを両脇に恰度牛の角のような恰好で取付ける布製の鉢巻をつくった。
(BCCWJ Sample ID: LBT3_00047)

もっとも、角に限らず、動物の特徴的な身体部位が比喩表現に用いられるときは、形に着目されることが多い。たとえば、象において特徴的な部位である鼻は主にその形が着目され、形状に基づいた一般的な名詞となっている場合もある (例 (5), (6), (7))。

- (5) この形から象の鼻星雲と呼ばれることもあります。
(NWJC <http://webryblog.biglobe.ne.jp/3/4/34a824b06a/3.html>)
- (6) 「象鼻山」や「駱駝山」など動物の姿をした珍しい岩山があり、観光の名所となっています。
(NWJC <http://blog.kikuchisan.net/archives/200909-1.html>)
- (7) お鼻が長いからゾウ虫。毒も無く、刺すことも無い、子供たちにも安心な昆虫さんです。
(NWJC <http://himehimehime5848.blog59.fc2.com/blog-date-201005.html>)

4.2.1 ユニコーンの角

BCCWJ におけるユニコーンの用例数は 68 件である⁷。比喩に用いられた例は見つからなかったが、NWJC⁸ からは他動物と同様に、特徴的な身体部位として、形状に関する比喩表現が得られている (例 (8), (9))。

- (8) オープンにして空を見上げた時、否応なしに飛び込んでくる、あの、ユニコーンの角のような長いアンテナを、チビたこれに換えた。
(NWJC <http://miniconvertible.seesaa.net/category/210229-1.html>)
- (9) そんなエビの正面。でも、これはユビナガスジエビではなく、アシナガスジエビ。ユニコーンみたいな角が何気に格好良かったり。
(NWJC <http://babura.blog52.fc2.com/category7-1.html>)

ユニコーンの角は形状に関する比喩表現として用いられていることから、架空生物ではあるが、その形状は一般的に固定しているものと考えられる。すなわち、特徴的な要素が一般的に固定し

⁷ BCCWJ から取得可能なユニコーン (語彙素) の用例 (68 件) 中、身体部位と共起する例には「角」のほか「(の尻尾の) 毛」「たてがみ」もあるが、これらは特定ファンタジー作品に関するテキストに限定されている。

⁸ NWJC のユニコーン (語彙素・普通名詞) の用例は 38,907 件取得されたが、大部分が特定バンド名 (固有名詞) であり架空生物 (普通名詞) ではないため、ここでは分類等を行わず、参考用例のみの取得に用いた。

たイメージである場合には、架空生物であっても形状を喩えるのに用いることができる。同様に、ジャッカロープ⁹（角兎）の角についても、固定的なイメージが一般的であるとわかる用例が取得されている（例（10））。架空生物の角であるために他動物とは異なるとはいえない。

- (10) もちろんジャッカロープのような立派な角のほうが、発見されたときもドラマチックであり理想的なのですが、おそらくジャングルのような森の中で生活していると考えられる小型の哺乳類に、そんなゴージャスな角が生えるとは考えにくいことは確かです。

(NWJC <http://umafan.blog72.fc2.com/category5-1.html>)

特徴的な要素として一般に認識されている場合には、比喩表現としてその形状が固定的に用いられると考えられる。

4.2.2 鬼の角

鬼の身体部位の用例として角の頻度は高くない（図5左）が、なぜ角が言及されないのかは明らかでない。鬼の用例の詳細を確認し、鬼と角との関係を考察したい。

鬼（語彙素）の用例はBCCWJから2,750例収集できた。用例の分類は、前述の角と同様に行った。用例を分類した結果、鬼用例として取得された例（鬼でない例を含む）の17.4%が比喩表現であり（次頁表4）、そのうち人間を被喩辞とする比喩表現が416例（鬼の比喩表現の86.8%）という高い割合を占めていた。以下に比喩表現の例を示す。

- (11) 「おのれ！」 政実の顔は一瞬にして鬼の形相となった。襲撃を命じた者は分かっている。
(BCCWJ Sample ID: PB19_00147)
- (12) 「営業三課の課長は鬼のようにしごくという厳しい面と、穏やかに長所を指摘するという面を持ち合わせている。
(BCCWJ Sample ID: LBk8_00016)
- (13) ほくが作ったパスワードクラッカーでも破るのに何週間もかかる鬼みたいなパスワードですよ。
(BCCWJ Sample ID: PM32_00104)
- (14) 「わが子をここまで追い込んだこの一徹…（ゆっくりと顔を上げながら）世間並みの親らしいことは何一つしてやらず、ひたすら鬼となってきましたわい」
(BCCWJ Sample ID: PB57_00193)

(11) は怒りのため人が鬼の形相になったことを意味し、(12) は職務のために鬼となることを示し、(13) は人知を超えた能力を鬼と呼び、(14) は厳しいことを意味する。このように、人間は怒りや威厳で鬼と呼ばれるようになる傾向がある。またその際、人は比喩的に角が生えているのだと考えられる。BCCWJから取得された鬼の用例のうち、「鬼＋助詞＋動詞」の例は882例あり、

⁹ ジャッカロープの角については、以下の用例などが「シカそっくりの角」と鹿の角に喩えている。

(i) ジャッカ・ロープは特にカウボーイの間では有名で、1934年7月にワイオミング州ダグラスでキャンプをしていたカウボーイがウサギの頭にシカそっくりの角が生えている生物を目撃し捕まえようと後を追うと群を成していたという都市伝説まであるんだ。

(NWJC <http://karapaia.livedoor.biz/archives/51427844.html>)

このうち「鬼となる」「鬼になる」「鬼に変わる」などの人が鬼に変化する例が 134 例 (15.2%) ある。これらに次ぐ「鬼がいる」「鬼はいる」「鬼がおる」などの存在に関する動詞の例が 53 例 (6.0%) であることから、変化する動詞とのコロケーション頻度が非常に高いことがわかる。しかし、人が鬼になる例は多いが、人に角が生える例は取得しにくい¹⁰。「鬼の角」は人が鬼になる (比喩表現) ことにどのように関係しているのだろうか。

表 4 鬼の用例分類 (BCCWJ 頻度 20 以上)

意味	頻度	
固有名詞・作品固有	823	30.1%
鬼 (一般)	687	25.1%
比喩表現	479	17.5%
鬼でない (吸血鬼など)	271	9.9%
鬼ごっこ	155	5.7%
慣用表現	106	3.9%
節分	69	2.5%
接頭語	36	1.3%
その他	111	4.1%
計	2,737	100.0%

鬼においても他動物と同様に、顔のような外観的な身体部位は、人間を被喩辞とする比喩表現において用いられるとき、その形状などが着目される (例 (15), (16))。しかし、鬼の角が比喩表現に用いられる場合は、主に人の頭部に加わった突起部分として、頭部全体の形状について用いられる傾向がある (例 (17), (18))。もしくは、形状を表す別の喩辞が並列的に示される (例 (19))。角は鬼にとって特徴的な部位であるが、そのものの形において特徴的ではないのだと考えられる。

前掲の保田 (2011) の実験結果においても、描かれた鬼の角は、牛や鹿、カモシカのものなど様々で形も本数も定まっていなかった。鬼には角がなくてはならないが、角らしきものがあればよく、その形や数は単一でなくてよいということであろう。

¹⁰ BCCWJ において、人らしきものに角が生える例はあったが、宗教的な文脈 (例 (ii)) か物語文脈 (例 (iii)) に限定された。(語彙素「角」+後方共起 (前方共起は 0 件) 10 語以内語彙素「生える」28 件, 語彙素「角」+後方共起 (前方共起は 0 件) 10 語以内語彙素「生やす」15 件中。なお, BCCWJ に語彙素「角」+語彙素「出る」・「出す」は 2 件に留まり, また人を被喩辞としていた例はない。このことから、角は予め有しているものが出るのではなく、なかったものが新規に生えるものという性質を持っていると考えられる。)

(ii) 怒っていて、怖い形相の人は、鬼のように見える。角が生えているようにも、口が裂けているようにも、牙が生えているようにも見える。その鬼の姿が、あの世では実在化してくる。心がそうなら、そうなってくるのだ。
(BCCWJ Sample ID: OB6X_00038 大川隆法「神秘の法」)

(iii) あれほどうつくしかった女房がいきなり、身のたけ八尺 (約二、四メートル) もの鬼に変じた。両眼は朱をとかして鏡の表面にながしたようで、口のはしには、上下くいちがった歯がむきだされている。口は耳のつけ根までかっとさげ、眉はうるしを百回もぬりかさねたようにもりあがって、ひたいをかくし、髪の毛から、うろこをかぶったつのが、によっきとはえているではないか。

(BCCWJ SampleID: LBjn_00004 木暮正夫「平家の怨霊」)

架空生物は実在の何かからイメージされるため、似た何かとの差異となる特徴的部位を持つが、身体部位はその形において特徴的ではない場合があるのだといえる。よって、鬼の角は形状に関して喩辞となることが少なく、言及頻度は低くなる。反対に、4.2.1 で見たように、ユニコーン（やジャッカローブ）の角のイメージは概ね固定的なので、鬼のようにイメージが固定的ではない角とは異なっている可能性が考えられ、用例にもそのように現れていたのである。

(15) 水はたちまち私のバケツに溢れていった。姑は鬼のような顔でバケツの水をにらんでいた。

(BCCWJ Sample ID: LBo8_00009)

(16) 鬼のような顔の少尉は、ぬいた刀をひっさげたまま本部に走っていく。

(BCCWJ Sample ID: LBdn_00003)

(17) 横向くと、たんこぶが盛り上がって、ええ、まるで鬼の角のような……（これはいいすぎ）

(NWJC <http://kikufuji.blog81.fc2.com/blog-category-12.html>)

(18) でも、確かにリスの耳は、鬼の角みたいかも(*^w^*)なんて思っちゃいました〜♪

(NWJC <http://hide1201.blog66.fc2.com/blog-entry-563.html>)

(19) 4月16日、茶色い、タケノコのような、鬼の角のような芽を出しました。

(NWJC <http://www.blogmura.com/profile/00987773.html>)

4.3 ユニコーンの角と鬼の角

鬼とユニコーンの角は、どちらも架空生物の特徴的な身体部位であり、基となる動物に角が加わったという点で類似している。これらの点においてユニコーンと鬼は同種であると考えられるが、ユニコーンの角の言及率と鬼の角の言及率には大きな差がある。ユニコーンは馬との差異となる角の頻度比率が高いが、鬼は人との差異となる角の頻度比率が低い。鬼において最も特徴的な部位であり、高い頻度で言及されることの期待される角が、実際にはそれほど言及されないという現象がある。

このように、高頻度で言及されていることの期待される要素が、実際の用例頻度として取得しにくい理由のいくつかは、比喩表現としての出現状況に現れていた。

まず、ユニコーンと鬼では、角が比喩表現においてどのように用いられているのかが異なっている。ユニコーンの角は、他の動物一般の特徴的要素と同様に、形状に関する喩辞として用いられる傾向が見られる。すなわち、固定的な形状のイメージが一般に持たれているということでもある。これに対し鬼の角は、比喩表現において人の頭部全体の形状を喩辞とするか、あるいは同形状の喩辞が並列されるなど、鬼の角として固定的な形状が焦点となっていない傾向が見られる。鬼の角は描画調査などでも形状の多様性が示されており、用例からも角単独の形状のイメージが固定的ではないことが明らかとなった。そのため、鬼の角は形状を喩辞とする比喩表現に用いられず、言及頻度が上がらない。

次に、鬼が人の喩辞として用いられる例は高頻度であっても、ユニコーンが馬の喩辞として用いられる例は本稿の調査において見つからなかったという違いがあった。この違いは、「人が鬼

になる」は頻繁に成立するが、「馬がユニコーンになる」はまず成立しないためであると考えられる。馬とユニコーンは別の生物であるが、人は鬼に変化する場合が多々ある。このように付加ではなく全体が変化するため、部分的な要素がとりたてて喩辞となることはない。もちろん、人が鬼となる場合には、鬼の特徴的な要素として角の生じることが必要である。人と鬼の外観的な差異は角である。しかし、鬼に角のあることは明らかでもあるため、また喩表現をはじめしければ人が鬼に変化するため、角について言及する必要がないといえる。

5. まとめ

ある対象物に関して重要な要素は、高頻度で言及されている可能性が高い。ある対象物に関して高頻度で言及される要素は、当該対象物において特徴的であるというよりも、他との差異となる要素であるために言及される傾向が見られる。それゆえに、高頻度で言及される要素は重要な要素であるとも考えられる。たとえば、ユニコーンは角のある馬である。よって、ユニコーンの角は馬との違いという点において特徴的な要素であるといえる。実際に、大規模コーパスにおけるユニコーンの身体部位の言及頻度を調査すると、そのほとんどが角である。角が最も言及されていることから、ユニコーンの特徴的な身体部位は角であるともいえよう。

我々が言及頻度からある対象物を認識しようとするとき、まず注目するのは高頻度で言及される要素である。鶴の首やヒラメの目など、高頻度の身体部位が対象物の特徴的な要素として着目される。反対に、我々はある対象物における要素の言及頻度を考えるにあたり、特徴的要素が高頻度で言及されると期待している。よって、対象物において特徴的と考えられる要素が低頻度であると、頻度情報から異なるカテゴリの誤った対象物を認識する。たとえば鬼は、人との違いとして、あるいは着目される部位として、角を持っている。しかし、コーパスにおける鬼の角の頻度はごく僅かであり、特徴的（他との区別に有用）な要素として期待される言及頻度ではない。そのため、身体部位の頻度分布から鬼を同定することは困難となっていた。

対象物のある要素が高頻度で言及されるはずだという我々の認識と、実際の言及頻度の差は、当該対象物を喩辞として用いる場合の喩表現としての用例差として現れていた。一般に、特徴的な要素は形状に関する喩辞として用いられやすいが、形状が固定的でない場合には喩辞として用いにくい。架空生物であってもユニコーンの角は形状に固定的なイメージがあるが、鬼の角の形状は一定ではないため、形状の喩辞にはなりにくい。また、差異が着目されるはずの対照される物こそが被喩辞となり、被喩辞が喩辞としての対象物そのものに変化する場合、部分的な変化は言及されにくい。変化過程において、差異たる特徴的要素が生じるのではあるが、かえって存在が明らかであるために言及する必要はないことにもなる。鬼とユニコーンは、それぞれ人と馬に角が加わった動物である。しかし用例において、人は鬼になるが、馬はユニコーンにならないという違いがあった。また、人が鬼になる際には角が当然生じているため、当該喩文脈では角が言及される必要はないといえる。

本稿の調査により、ある対象物に関する用例において、高頻度であることが期待される要素の言及が少ない場合のあること、その場合には喩表現に特性の見られることがわかった。コーバ

スを用いた頻度情報の利用に際しては、重要な要素が高頻度で言及されない場合について考慮する必要があると考えられる。また、比喩表現をはじめとする表現手法が、テキストから取得される情報に及ぼす影響についての整理を進めることが重要であろう。

参考文献

- 天野成昭・笠原要・近藤公久（編）（2008）『日本語の語彙特性 第9巻 単語親密度増補』東京：三省堂。
天野成昭・近藤公久（編）（1999）『日本語の語彙特性 第1巻 単語親密度』東京：三省堂。
加藤祥（2015）「テキストからの対象物認識に有用な記述内容—動物を例に—」『国立国語研究所論集』9: 23-50。
加藤祥・岡本雅史・荒牧英治（2015）「テキスト世界と現実世界の差異—動物の部位分布における3つのプロトタイプ効果—」山梨正明（編）『認知言語学論考』12: 81-113. 東京：ひつじ書房。
国広哲弥（2000）「日本語動詞の多義体系（3）」『神奈川大学言語研究』22: 1-12。
McEnery, Tony and Andrew Hardie (2012) *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
McRae, Ken, George S. Cree, Mark S. Seidenberg and Chris McNorgan (2005) Semantic feature production norms for a large set of living and nonliving things. *Behaviour Research Methods, Instruments & Computers* 37(4): 547-559.
中村明（1977）『比喩表現の理論と分類』（国立国語研究所報告 57）東京：秀英出版。
Sinclair, John (1991) *Corpus, concordance, collocation*. Oxford: Oxford University Press.
保田祥（2011）「名詞の百科事典的意味の抽出方法とその有用性；内省・描画実験・コーパス調査」博士論文、神戸大学。

例文出典

- Google n-gram: <https://catalog.ldc.upenn.edu/LDC2006T13>
Google Japanese n-gram: <https://catalog.ldc.upenn.edu/LDC2009T08>
Kudo, Taku and Hideto Kazawa (2009) Japanese Web N-gram Version 1. Linguistic Data Consortium.
国立国語研究所『現代日本語書き言葉均衡コーパス』「中納言」<https://chunagon.ninjal.ac.jp/>（2017年6月27日確認）

例文出典（Web サイト）

- 国立国語研究所『国語研日本語 Web コーパス』「梵天」<http://bonten.ninjal.ac.jp/>（2017年6月27日確認）

Characteristic Elements of Words and Their Frequencies: The Case of *Tsuno* ‘Horns’

KATO Sachi

Adjunct Researcher, Center for Corpus Development, NINJAL

Abstract

Many dictionaries, such as the Collins COBUILD English language dictionary and WISDOM English-Japanese Dictionary, use corpus frequency data as the basis for determining the importance of words or word meanings. Based on the corpus frequency data, we assume that the most characteristic elements of an object tend to be mentioned frequently in corpora. In this study, we investigated the use of words that describe animal body parts and their frequencies. If the characteristic attribute of a target animal has a high frequency in the corpora, we would be able to guess the target animal. For example, we expected *tsuno* ‘horn,’ a word that distinguishes one animal type from another, to be used frequently. In the case of unicorns, we found that its horn was mentioned frequently, as it distinguishes a unicorn from a horse. However, the horns of *oni* ‘devil’ were mentioned less frequently, even though it is a feature that distinguishes *oni* from human beings. Upon analysis of the corpora, it was revealed that *oni* are often used as metaphors for human beings. By contrast, unicorns are not used as metaphors for horses. Moreover, *oni* horns do not have the fixed image that unicorn horns do as a metaphor for its form. Our results lead to the conclusion that the tendency for the most characteristic feature of an object not to be mentioned is the effect of metaphors.

Key words: corpora, usage frequency, characteristic elements, metaphor, horns