

「犬」と「イヌ」と「いぬ」

～日本語表記の違いによる動物の部位分布～

保田祥[†] 岡本雅史[‡] 荒牧英治^{†*}[†] 東京大学 知の構造化センター[‡] 清泉女子大学 *JST さきがけ「情報環境と人」

1. はじめに

テキストでは、ある対象が異なる表記を用いて記述されることがある。特に日本語テキストは、漢字・カタカナ・ひらがななどの異なる文字種により、多くの表記ゆれが存在する。表記ゆれは実際に言語処理アプリケーションを構築する際に大きな問題となるため、これまで、表記ゆれを吸収するための研究がなされてきた。しかし、表記の異なる語を同じ語とみなしてよいのだろうか？ 本稿では動物名の表記(たとえば「犬」と「イヌ」と「いぬ」)を題材とし、身体部位の用例(「犬の尾」「イヌの尾」「いぬの尾」)を収集することで表記の差を調査した。調査は2点からなる。

まず、日本語表記の違いによる動物の部位分布の異同を確かめた。表記の差があっても対象物が同じ動物であるとすれば、同じ部位が同じ頻度で出現すると期待される。しかし、実際には、表記ごとに部位の出現分布は大きく異なる場合が多いことが明らかとなった(4.1節)。たとえば、漢字表記の「犬」には高頻度で共起する「毛」が、ひらがな表記の「いぬ」にはめったに見られない。

次に、表記の違いを引き起こす原因について考察を行った。従来から、表記の違いはジャンルの違いに起因すると説明されてきた。例えば、カタカナが学術的(啓蒙書・翻訳書や自然科学関係の書物、また法律など「大辞林」)、ひらがなが易しい文章(漢字の知識に乏しい人々などが用いる「大辞林」)などといわれてきたが、本稿では、表記の差がジャンルの違いによってだけでは完全には説明できず、むしろ、動物の持つ多様な側面(本稿では「役割」と呼ぶ)、例えば、¹〈生物〉、〈ペット〉や〈食べ物〉のうちどこに焦点があたっているかが特定の表記を選好している(4.2節)。

なお、〈鰻〉²のように概ね〈食べ物〉として扱われるような動物は、いずれの表記においてもほぼ同じ

¹ 食材や料理名などの下位語を含む役割を〈食べ物〉のように表す。

² 「鰻」と「ウナギ」と「うなぎ」のような表記の違いを考慮しない場合、〈鰻〉と表す。

部位分布を示しており、先の仮説を補強する(4.3節)。

2. 先行研究：表記の違いによって生じる差異

国立国語研究所(1983)は、新聞に現れる34,477語の約14%で表記にゆれが認められることを示す。実際、動物名の表記は、漢字・カタカナ・ひらがなのいずれもが使用されている場合が多い。動物名については、「ヒラガナとカタカナの対立」として「うさぎ」「ウサギ」の表記例を示し、「漢字とカタカナの対立」として、「山羊」「ヤギ」の表記例を示しているが、「動植物名³」である場合にカタカナを用いるため、表記の違いが発生しているという。

表記と表現の関係として、佐竹(1980)が、表記の差で「表現性」が生じることを言うほか、中村(2007)は「ことばの意義をふくらませる修辞技法」としての「字装法」を説き、文字に音と意味の二重の働きをさせる例⁴を挙げる。表記が異なることで、伝わる内容が異なる可能性が考えられる。

山仲(2006)は、ひらがなを「話し言葉を中心にするやわらかい語り口調」、カタカナを「漢文訓読に用いる硬い言葉や言い回しでストレートに物事を描写する」と説明している。日本語史との関係から、ひらがなは、年少向けなどやさしい文章に現れやすく、カタカナは、表音目的や学術的分野で用いられやすいという表記の性質の違いが生じている可能性がある。

また、国立国語研究所(1961)は、同音語の判別条件のうち、語の意味に関する部分として、その語をとりまく外部的な条件に「文脈の違い」と「話題(位相)の違い」を挙げている。

³ 文部省指導の内閣告示(1946～1981)に、動植物名をカタカナで表記することが示された影響も考えられる。

⁴ 中村(2007)によれば、谷崎純一郎の『盲目物語』の極端な書きの多さは、「いかにも目の不自由な者が訥々と語る雰囲気漂わせることに成功したとされる」。このほか、カタカナ表記では、伊藤聖の「漢字表記にそなわっていた重々しさが剥奪される例や、小島信夫の「諷刺的な文字操作」、安岡正太郎の異例表記による「感覚的に伝える」表現効果の例などが示される。

3. コーパスと調査方法

動物 (<犬><猫>など 200 種) とその外観的な身体部位 (<足><手>など 100 部位) について、組み合わせの用例 (「犬の尾」, 「猫の手」など) の使用頻度を、Web 大規模コーパス (Google n-gram) を用いて調査した。このとき、各動物について、異なる表記のそれぞれを調査対象とした。たとえば<犬>については、「犬」「イヌ」「いぬ」のように、漢字・カタカナ・ひらがな表記の三種類を調査対象とした。

動物の選定に関しては、<犬><猫>のようなレベルの動物を対象とした。よって、「秋田犬」「土佐犬」などの複合語 (下位語) も「犬」として扱った。

<テントウムシ>のように漢字表記の種類が多い場合は、「瓢虫」「天道虫」「紅娘」と、漢字表記の調査対象を増やした。

また、用例が固有名詞である場合 (各々の組み合わせ用例を確認し、社名・店名・書籍タイトル・ブログ名などが用例の頻度の上位に現れる場合) は、当該動物を考察対象外とした⁵。

4. 調査結果

コーパスから、動物名と身体部位名の組み合わせの用例が、3,944,601 例取得できた。調査結果から、表記の違いによって動物の部位分布に違いが生じるのかを検証し (4.1 節)、部位分布の原因を探り (4.2 節)、例外を確認することで、部位分布の差について考察を行う (4.3 節)。

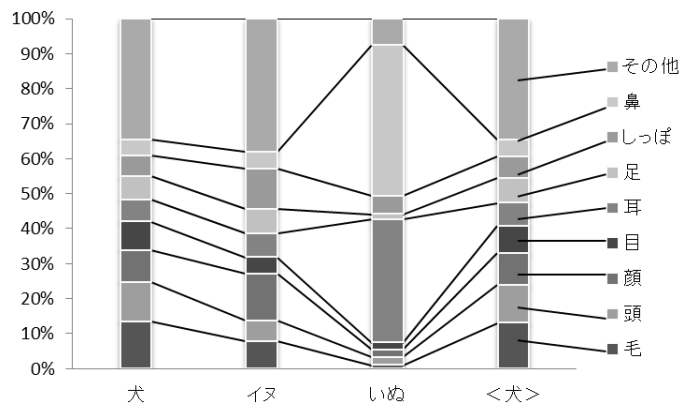
4.1. 表記によって部位分布が異なる

表記が異なっても、同一の動物であれば、出現する部位とその頻度が概ね同じであることが期待される。しかし、出現する部位の分布が異なることと、特定の表記にのみ出現しやすい部位もあることが明らかとなった。

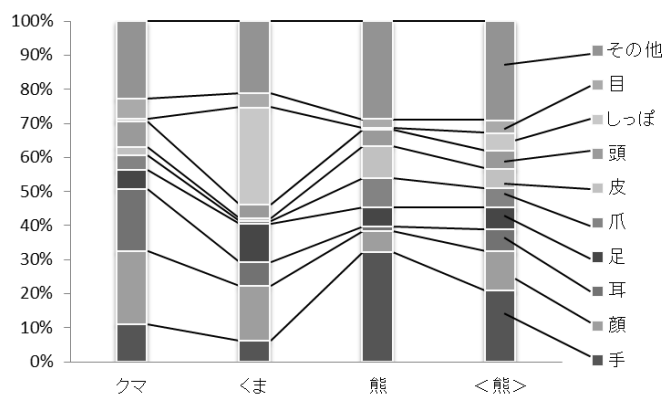
まず、図 1 の<犬>では、「犬」(153,040 例)・「イヌ」(4,256 例)・「いぬ」(3,874 例) のそれぞれの表記で、出現する部位の分布が異なっていることがわかる。また、ある表記にのみ出現しやすい部位があることもわかる。「いぬ」表記では、鼻が「いぬ」用例全体の半数近くを占める部位であるが、「イヌ」と「犬」で

は鼻は頻度上位の部位ではない。

なお、「犬」「イヌ」「いぬ」の各表記における各部位の出現頻度を検定すると、「イヌ」の顔・しっぽ・図内には表示されていないが、歯・背中、「いぬ」では耳・鼻が有意に出現する ($p < 0.01$)⁶。



【図 1. <犬>の部位分布】



【図 2. <熊>の部位分布】

他の動物においても、「くま」(図 2)における「しっぽ」(6,013 例の 33%), 「ネコ」における「おしり」(81,364 例の 12%) など特異的に出現しやすい部位がある。以上のように、表記によって部位分布は不均一である。

なお、本研究の統計は、付録の Web サイトにて閲覧可能である。

4.2. 動物の部位分布の発生原因

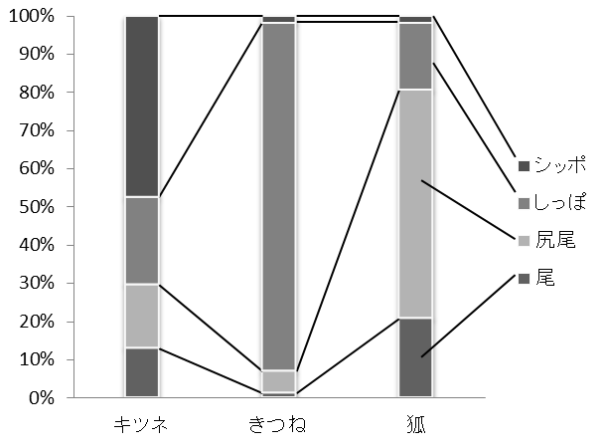
では、何が原因で部位分布の違いが発生するのだろうか?

⁵固有名詞の影響が強く現れている場合も考察対象としなかった。たとえば、「サソリの爪」は固有名詞の影響が強いと考えられる。ゲーム「ファイナルファンタジーXI (株)スクウェア・エニックス」における「上質な大サソリの爪」のようなアイテム名の影響が考えられるためである。このほか、「山羊の頭」は、ローリングストーンズのアルバム邦題「山羊の頭のスープ」, 「猿の手」は、W.W. ジェイコブズの著作「猿の手」の影響が大きいことなどが推測された。

⁶「犬」「イヌ」「いぬ」の総数 (表記ゆれを吸収した結果) との比較を行った。

同表記の選好

ひらがなの動物名であれば、部位分布を見ても、ひらがな表記の部位や幼児語と結びつく可能性がある。例えば、〈狐〉においては、図3のように、「キツネ」がカタカナのシッポと共起し、「きつね」がひらがなの「しっぽ」とよく共起している。しかし、〈犬〉においては、カタカナの「イヌ」がひらがなの「しっぽ」と共起しており（図1）、一般性を持ち得ない。



【図3. 〈狐〉と〈しっぽ〉】

ジャンルによる選好

自然科学分野における動物ならば、カタカナ表記が選好されるといった、ジャンルの違いが影響している可能性がある。ジャンルによる分析を行うために「現代日本語書き言葉均衡コーパス（以下、BCCWJ）」を用い、動物名の出現するサブコーパスとジャンルを調査した。

この結果、例えば〈犬〉では、漢字表記はどの分野にも表れるが、カタカナ表記が、自然科学・産業・文学・歴史分野でのみ現れ、ひらがな表記が書籍とブログ・Yahoo! 知恵袋のサブコーパスに限られるという偏りが見られた。当然、ジャンル毎に出現する部位にも差があるため、ジャンルと部位分布は深く関連すると考えられる。但し、どの表記にどのような部位が突出するか個々の動物に依存し、ジャンルによる一般的な説明も困難である。

役割による選好

〈熊〉であれば、〈食べ物〉や〈生物〉や〈害獣〉など、〈猫〉であれば、〈愛玩動物〉や〈肉食獣〉や〈三味線の材料〉などのように、動物は多くの役割を有しており、文脈によりそれらのいずれかに注目していると考えられる。このような多様な役割が、いずれかの表記に割り振られ、部位分布の差異をつくっている可能性がある。そこで、「役割」という観点から、

次節にて考察を行う。

4.3. 〈役割〉と部位分布

〈役割〉が単一なら部位分布は偏らない

例外的に、役割が少ない動物がある。例えば、〈鰻〉などは、一般の人間が接するのはほぼ〈食べ物〉としての〈鰻〉である。もし、役割が部位分布の偏りを生んでいるならば、役割が1つしかない動物については、部位分布は偏らないはずである。

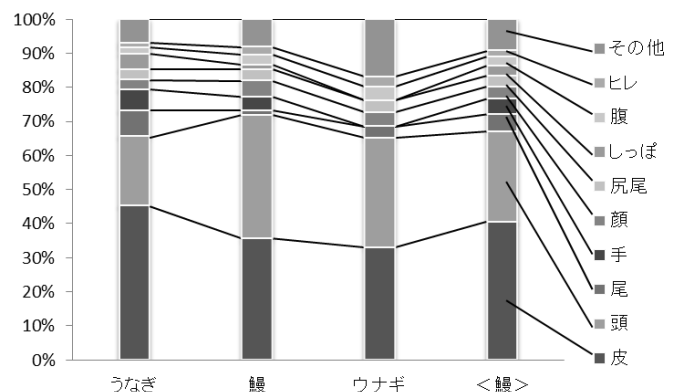
図4に〈鰻〉の部位分布を示す。他の動物（図1や図2）とは異なり、表記による部位分布の差が少なくなっている。同様の性質は、〈鰻〉以外にも、〈秋刀魚〉〈蛸〉など、役割が〈食べ物〉しかない動物に共通している。

以上のように、役割の単一である場合は、部位分布の偏りが見られないことがわかる。

なお、役割が実際に単一であるかについては以下の用例に見られるように厳密な議論が必要となる。

- 1) アツアツのご飯が鰻の皮を柔らかく保ってくれる。
- 2) ウナギを半分に切り、フライパンに、ウナギの皮を下にして乗せます。
- 3) うなぎの皮の表面がふつふついつてきたら、身の部分にひっくり返して
(以上、すべて Google 検索による)

これらでは、1が調理済みの食材であり、漢字を選好している。また、2が調理前の食材である（カタカナを選好）。3はどちらとも言い切れない（ひらがなを選好）。このように〈食べ物〉といっても、さらに細分化した役割の存在があるが、部位分布の大きな差を生じさせるまでには至っていない。



【図4. 〈鰻〉の部位分布】

<役割>が多ければ部位分布は大きく偏る

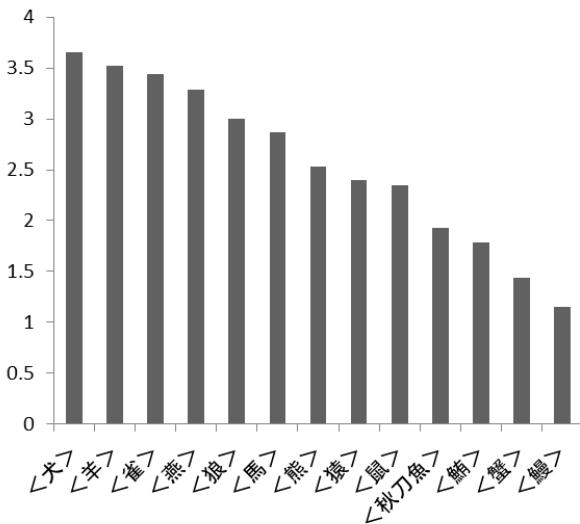
では、逆に<役割>が多い場合は、部位分布は偏るのであるか。これを調べるため、動物ごとに部位分布の差を調査した。

調査は、「犬」「イヌ」「いぬ」を合計した<犬>における部位の出現率と個々の表記の出現率の差を調べた。この際、全く用例の取得できなかった表記がある動物や、特定の表記で固有表現の頻度が高い動物、慣用句のみの突出する動物などを除いた。

結果を図5に示す。縦軸の値が大きければ大きいほど、部位分布の偏りが大きい動物であるといえる。結果、<犬>をはじめ、<羊>や<雀>のような動物が部位分布の偏りが大きいことが分かった。これらの動物は、我々が日常的に目にする動物であり、多様な役割を持つと考えられる。

一方、<鰻>をはじめ、<蟹>や<鮪>のように、<食べ物>の役割のみが多いであろうと考えられる動物は部位分布の偏りが小さい。

このように、役割の多少が、部位分布の差として観察されており、部位分布と役割の関連性を示唆する。



【図5. 表記の違いによる動物の部位分布差】

Y 軸は以下に定める部位分布の偏りの大きさを示す：N 種類の表記を持ち、M 種類の部位が出現する動物について表記 $i (i=1, \dots, N)$ 中の部位 $j (j=1, \dots, M)$ の出現頻度の割合を b_{ij} 、表記を区別しない場合の部位 j の割合を \bar{b}_j とすると

$$\text{部位分布の偏りの大きさ} = \sqrt{\frac{M}{N} \sum_{i=1}^N \sum_{j=1}^M (b_{ij} - \bar{b}_j)^2}$$

5. まとめ

動物の身体部位を調査することにより、日本語テキストでは、表記の違いが大きな差異を生むことが明らか

かとなった。また、部位分布の差は、従来言われているような共起語や出現ジャンルのみの影響でなく、その動物の役割と大きく関連した現象であることがわかった。表記ゆれの安易な吸収は、この役割に関する情報を失ってしまうとも言える。本稿の考察をもとに、緻密な表記ゆれの研究が今後望まれる。

謝辞：本研究は、JST 戦略的創造研究推進事業（さがけタイプ）「情報環境と人」及び、科研費補助金（若手研究 A「表記ゆれ及びそれに類する現象の包括的言語処理に関する研究」）による。

データの整形や検定については、東京大学知の構造化センターの馬場一貴氏、増川佐知子氏の両氏にご協力いただいた。

参考文献

国立国語研究所 (1961) 『同音語の研究』
国立国語研究所 (1983) 『現代表記のゆれ』
佐竹秀雄 (1980) 「日本語の表記の表現性」『国文学解釈と鑑賞』45(8) 至文堂
中村明 (2007) 『日本語の文体・レトリック辞典』東京堂出版
山口仲美 (2006) 『日本語の歴史』岩波書店
松村明編 (2010) 『大辞林』第三版 三省堂
Taku Kudo, Hideto Kazawa, “Web Japanese N-gram Version 1”, published by Gengo Shigen Kyokai
国立国語研究所『日本語書き言葉均衡コーパス』(<https://chunagon.ninjal.ac.jp/>)

付録

本研究のデータは以下のサイトで閲覧可能である。

<http://mednlp.jp/~baba/animal/>
animal: ウサギ (44620 records found)

