

象は鼻が長いか —テキストから取得される対象物情報—

加藤 祥 (国立国語研究所コーパス開発センター) †

Does an Elephant Have a Long Nose? Features of Entities Acquired from Texts

Sachi Kato (National Institute for Japanese Language and Linguistics)

要旨

本稿は、対象物に関する情報について、コーパスから取得可能な内容・頻度と、対象物の説明文に見られる内容・頻度・順序を調査し、テキストから取得される情報の特性について考察を行う。特徴的な身体部位を有すると考えられる象をとりあげ、その調査結果を報告する。まず、現代日本語書き言葉均衡コーパス (BCCWJ) の象の用例から、取得可能な情報を調査した。また、「対象物をまったく知らない人に説明する」条件指示によりクラウドソーシング実験を行い、一般的な作文テキストを収集した。これらのテキストを分析した結果、象が大きいことと象の鼻が長いことは高頻度かつ早い順序で言及されやすいが、象の鼻の長さがどの程度かは言及されにくいとわかった。対象物認識に重要視される外観的特徴情報は、身体部位が「長い」「大きい」などの形容表現に前提的文化的知識が期待されやすく、既存のテキストのみからでは対象物のイメージが獲得しにくいといえる。

1. はじめに

テキスト情報からのみで対象物を認識するのは困難な傾向がある¹。すなわち、我々が日常的にテキストから知識を獲得する例は多いが、正しくテキスト内容を認識できていないとは限らない。知識のない読み手に対してどのような記述をすれば情報が適確に伝わるかという問題がある。

本稿は、対象物を説明するにあたり、特徴と考えられる情報がどのように言語化 (記述) されるのか調査する。まず、用例としてコーパスから取得可能な特徴情報 (内容・頻度) を調査することで、言及されやすい情報を整理する。次に、対象物を説明する作文を被験者実験によって収集し、対象物を効果的効率的に説明するためには、どのような情報をどのような順序で記述する傾向があるのか分析する。具体的には、象を対象とした調査を行い、象に関する記述から取得できる象についての特徴的な情報は何であるのか、また、象の鼻が長い、耳が大きいというような特徴的な情報がどのように取得できるか、あるいは取得しにくい情報は何かを考える。

2. 関連研究と本研究

国語辞書における意味は、対象物を説明するにあたって様々な内容が記述されたものと考えられる。しかし同時に、国語辞書の記述は必ずしも十分なものではないと指摘されて

† yasuda-s@ninjal.ac.jp

¹ 加藤 (近刊) では、対象物についての各種テキスト (辞書語釈、被験者によって求められた情報、コーパスから取得した用例) を用いた対象物 (知識率の高い動物) の同定実験を行っている。この実験結果では、いずれのテキストでも平均的に半数程度の正答率に留まっており、テキストのみから対象物を認識することの、ある種の困難さを示している。

きた (Fillmore & Atkins, 1994 など)。では, どのような記述が理想的な対象物の説明となるのか。

たとえば, 國廣 (1997) は「辞書の意味記述」に求める項目を示した。一般的な国語辞書の記述に現れにくいものとして, 「語義的位置 (語彙体系の中の位置)」「語義の対義的定義 (対義語を示す)」「現象素² (認められる場合には図示)」「用例³ (広く実例を観察した上で適当にまとめる)」「連想 (動物名であればその動物の習性或故事来歴など (百科的知識))」が挙げられている。但し, これらの項目は国語辞書の意味記述の場合に限るため, 辞書のほかのテキストからも同様に得にくい情報とは言い難いであろう。

また, 辞書の意味とは異なる百科事典的知識 (folk-knowledge; Wierzbicka 1996) として Natural Semantic Meta language (NSM) theory (e.g., Goddard and Wierzbicka, 2014) による記述がある。Wierzbicka (1985) の dog の例では, dog が認識可能な形や形態的な特徴を持たないため, 必要十分な特性ではなく特徴的な特性のリストによって概念が定義されるとする。この際, dog の認識可能な特徴は振る舞い (特に吠える・唸る・尾を振る) であり, dog は「人とともに生き, 献身的で従順, 信頼し得る仲間, よき学習者, 勤勉な労働者である」というような, 人との関係において概念化される。しかし, 人との関係が一般的に薄い動物であれば, この種の情報が記述として得にくい可能性もある。

そのほか, コーパスを用いた辞書の語釈の記述として, Sinclair が編集主幹を務めた学習者用辞書の COBUILD (1987~) では, 語の意味は顕著だと見なされた最小限の細目 (Sinclair 1992) とされ, コーパスに近い例文を掲載する試みが為されている (COBUILD 2009, p. xi)。

以上のような対象物に関する記述において, ある対象物を説明するにあたり特徴的な情報が適確に記述されているのかという検証は行われにくい。

加藤 (近刊) は, 対象物の認識に有用な情報はどのようなものかという観点で, 辞書語釈やコーパスなどのテキストを用い, テキスト内の対象物認識に有用な情報を被験者実験によって調査した。この調査において対象物の認識に必要とされた記述は, 主に読み手の経験や知識を喚起する情報と, 提示された情報によって設定されるカテゴリに属する他メンバーとの差異に関する情報であった。記述されている情報は, 予め読み手の保有している知識と合致した場合には有用な情報となる。また反対に対象物に関する知識が読み手に不足している場合には, 対象物の認識に親カテゴリのプロトタイプとの差異の記述が有用であり, あるいは誤認を避けるために他メンバーとの差別化の可能な記述が有用であった。

しかし, コーパスの利用などによりテキストから取得できる情報には, その内容に限らず, 頻度や記述順序という情報もある。対象物について説明するにあたり, 何が特徴的な情報としてどのように記述されるかという問題が残っている。そこで本稿は, まず既存の説明文として国語辞書 10 種類の語釈を収集し, 次にコーパスから対象物の用例を取得して対象物に関する情報がそれぞれどのような頻度で得られるのかを調べるとともに, 同一の対象物に関する 100 以上の説明文章を作文実験によって収集し, 情報内容の出現頻度と記述順序を調査することとした。

² 國廣 (1994) は, 現象素を「人間の認知作用を通して, ひとまとまりをなすものとして把握された現象」と呼ぶ。

³ 「適切な用例が見付かるとは言い難いという問題がある」と指摘する。

3. 調査

対象物を説明する際、辞書の語釈であれば外観に関する情報が記述されやすい⁴。そこで、Google日本語n-gramにおける動物の身体部位の用例頻度を調査したところ⁵、象（異表記を含む）については「背⁶36%（固有名詞を含む）」「鼻 21%」「耳 10%」と割合の高い部位が上位3種ある（図1）という結果が得られた。象は外観的に特徴的な属性を有しているため、特徴が記述されやすいと考えられる。以上により、本稿の調査の対象として象を用いる。

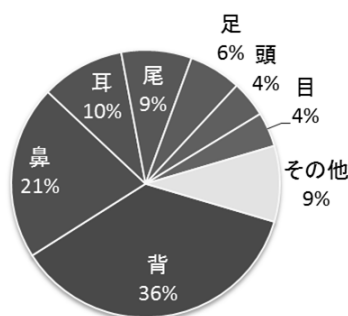


図1 Google 日本語 n-gram における象の身体部位用例分布

調査データとして、国語辞書（3.1）、コーパス（3.2）、作文実験（3.3）を用いる。以下の節にそれぞれの調査結果を示す。

3. 1 国語辞書

象の説明例として、まず国語辞書の語釈から得られる情報をみておきたい。

国語辞書 10 種類（表 1）の語釈における「象」項目の記述内容とその提示順序を調査した。平均 66 文字（min : 14 文字, max : 136 文字）を得た。

表 1 データを取得した国語辞書

辞書	三省堂国語	新明解国語	岩波国語	明鏡国語	新選国語	集英社国語	角川国語	新潮現代	大辞林	デイリー国語
出版社	三省堂	三省堂	岩波書店	大修館書店	小学館	集英社	角川書店	新潮社	三省堂	三省堂
版	5 版	6 版	5 版	初版	7 版	2 版	新版	2 版	Web 版	3 版
項目数	76,000	75,000	62,000	70,000	83,000	92,000	75,000	79,000	260,000	70,000
字数(象)	65 文字	39 文字	66 文字	108 文字	80 文字	54 文字	52 文字	45 文字	136 文字	14 文字

⁴ 加藤（近刊）では、国語辞書 10 種類から動物 200 種類の語釈を収集し、どのような種類の記述があるかまとめている。以下の表から、形態情報（外観に関する情報）が 9 割近くの動物で記述されており、形態情報の記述される割合が高いとわかる。語釈文においては形態情報が重要視されると考えられる。

補表 国語辞書における動物語釈の分類別記述（加藤 近刊による）

	分類	形態	生態	人間との関係	その他
当該分類の記述がある割合(200 種類中)	96.0%	87.5%	82.0%	52.5%	44.5%
各語釈における当該分類の記述割合(平均)	25.6%	36.7%	24.4%	23.3%	15.8%

⁵ 身体部位の用例頻度は外観的な情報と均衡しないが、特徴的な身体部位は言及されやすい傾向がある（加藤ほか 近刊）。

⁶ Google 日本語 n-gram では、「象（異表記を含む）の背」用例の 26%が「象の背に乗っ」であった。後述する 3.2 の表 3 でも「(背に) 乗る」が全用例（3%）である。背が身体部位として特徴的とは言い難い。

記述内容とその提示順序を表 2 に示す。平均 5.9 種類の内容 (min : 2, max : 9) が得られた。提示順序は内容毎に出現順を数えている。

まず、内容について、大型であることは 10 種全ての辞書で記述されていた。鼻が長いことについても 10 種全てに記述があったが、「長い」という形容詞の他に「ものをつかめる」「自由に動かせる」のような鼻についての記述があった辞書は 4 種類にとどまったため、表 2 では詳細の有無で別内容として示してある。

表 2 国語辞書における「象」項目の記述内容数とその順序 (上位)

内容	記述有辞書数	1 番目	2 番目	3 番目	4 番目	5 番目
大型であること	10	3	5	1	0	0
象牙に関して	7	0	0	1	2	3
哺乳類	6	5	1	0	0	0
鼻が「長い」(詳細なし)	6	0	2	0	3	1
種類の別があるなど	5	0	0	1	1	0
生息地	4	2	0	1	0	0

次に、情報の提示順序をみると、まず 1 番目に、哺乳類であること (5 種類)、大型であること (3 種類) と「アジアアフリカに」生息すること (2 種類) が記述されていた。2 番目には、大型であること (5 種類)、鼻が長いこと (2 種類) が見られる。大型であることは 1~3 番目で 9 種類、「鼻が長い」に関しては 2~5 番目までで 10 種類と、前半に記述されやすい傾向があった。国語辞書においては、大型であることと鼻の長いことが、内容としても順序としても特徴的であると読み取れる。

3. 2 コーパス

現代日本語書き言葉均衡コーパス (BCCWJ) より取得した象の用例から得られる象に関する情報を分類し、コーパスからどのような情報が取得できるのか調査した。用例の収集には中納言⁷を用い、語彙素「象」について前後 300 文字の文脈を取得した。

語彙素「象」の検索を行うと、1,323 件がヒットする。このうち、動物の「象」についての用例は 1,050 件 (サンプル数では 349 件) と判断された。これらの用例の整理を、作業者の判断によって行った。同内容と考えられる例 ((1)(2)のような例) を意味内容によってまとめた ((1)(2)をまとめて(3)とした例)。以下に挙げる例の下線は著者による。

- (1) しかし、与えると命がのびるので動物園の人たちは悲しみやつらさをじいっと耐え、心を鬼にして食べるものを与えなかったのです。やがて、象は何十日も食べ物を口でできず、とうとう飢えて死んでいったのでした。

(LBg9_00083 : 石森史郎『Once upon a time in...』⁸)

⁷ 中納言 1.1.0 (<https://chunagon.ninjal.ac.jp/>) 短単位データ 1.0, 長単位データ 1.0 を使用した。

⁸ 用例の出典は、(サンプル ID : 著者名『タイトル』(またはサブコーパス名)) と記す。

(2) 私も「かわいそうなゾウ」

戦争中動物園をつぶさなくてはいけなくて動物達を毒殺したそうです。でもゾウは死ななくてしかたがなく餓死させたそうです。(OC12_03193 : Yahoo!知恵袋)

(3) 戦時中, 上野動物園で餓死させられた。(意味的な用例として(1)(2)などをまとめた例)

以上のような作業により, 1,314 種類の意味的な用例が取得できた。この作業にあたっては, 上記(3)のように数件の用例を 1 種類にまとめた場合や, 1 件の用例から 2 種類以上の意味的な用例が取得される場合がある。なお, コーパスから取得した用例は, 基本的に象を説明する文でないか, 完結した文章でないこともあるため, 内容の提示順については本調査の対象外とした。

BCCWJ における象の意味的な用例 1,314 種類を内容で分類すると, 1%以上の割合で見られた内容には表 3 の種類が見られた。

表 3 BCCWJ における内容別用例分類結果出現割合上位 (1%以上)

内容	出現割合	内容	出現割合	内容	出現割合
固有(象?)	20.7%	場所(国・動物園)	5.7%	歴史(祖先・来歴)	4.9%
共起(並列)	4.0%	造形(かたどったもの)	3.8%	飼育する(人が)	3.7%
大きいこと ⁹	3.3%	比喩 ¹⁰	3.1%	乗る(人が)	3.0%
象牙(密猟含)	2.9%	訓練する(人が)	2.6%	種類(下位カテゴリ)	2.6%
鼻について	2.2%	伝説(英雄譚・歴史)	2.1%	共起(対照)	1.8%
重いこと ¹¹	1.5%	性質	1.4%	食べる(量・種類)	1.4%
例示	1.4%	メディア(経験取得)	1.2%		

まず, コーパスデータの中には, 動物の象であることが擬人化などにより曖昧な固有の

⁹ 以下の注 9 も同様であるが, 比喩・例示と別項目に分類した例にも, 大きさに関して喩える例や, 大きなものとして例示している例が見られる。以下のような用例を「大きいこと」として扱うと, 全体の 4.6% が大きさに関する意味的な用例であるといえる。

(補例 1) ゾウをのんだウワバミになったような, 変な気分になってしまう。だから, やめよう。

(LBhn_00019 : 荻原規子『これは王国のかぎ』)

¹⁰ 比喩用例として分類した用例のターゲットドメインによる細分類は以下である。

形状	大きさ	動作	耳	鼻	様態	情景	不明
1.4%	0.9%	0.4%	0.2%	0.2%	0.1%	0.1%	0.1%

比喩用例において「耳」「鼻」が着目されることから, 象は一般に「耳」と「鼻」が特徴的と考えられている可能性が考えられる。

¹¹ 注 7 と同様に, 比喩・例示と別項目に分類した例にも, 重さに関して喩える例や, 重いものとして例示している例が見られる。以下のような用例を「重いこと」として扱うと, 全体の 2.7% が重さに関する意味的な用例であるといえる。

(補例 2) 入ってる辞書的にはキャノンがよかったのですが, 象が踏んでも壊れない (←筆箱だって?) 頑丈さと, なんと言っても電子辞書シェア No. 1 と言うことで, カシオになりました。

(OY05_06688 : Yahoo!ブログ, 原文ママ)

象用例が多く現れ、20.7%がこの種と分類された。本稿では、以下の(4)(5)のような例は固有の象と判断し、その他への細分類を行わなかった。

- (4) それから白い象は大急ぎでドアに鍵をかけ、鍵はドアマットの下に押し込み、森のほうへととっとと駆けてゆきました。もちろん人の声が聞こえたのは反対の方向へ。

(LBIn_00034 : C・ネストリンガー作/松島富美代訳『象さんの素敵な生活』)

- (5) 大きな湖を見わたして暮らそうと、ババールがつくった「セレストビル」。学校や病院や図書館、そして映画やお芝居を楽しめる「たのしみのやかた」もある、りっぱな都です。ぞうたちが、みんな楽しく平和に暮らすババールの国。

(PM51_00768 : 『月刊MOE』2005年9月号)

このほかの取得可能な象に関する要素としては、見ることのできる場所(国や生息地域、動物園名、出現メディアなど)、形を知ることのできるもの(模ったもの)、人との関係(飼育・訓練を行うこと、乗ること、象牙をとることなど)、歴史(祖先や来歴)と伝説、カテゴリ(並列・対照して共起するもの)が主となった。

上位で出現する内容を見るに、対象物そのものについては、「大きいこと」「重いこと」「鼻」が特徴的な情報として取得できている。

3. 3 作文実験

「対象物をまったく知らない人に説明する」という条件提示によって、象の説明文を作文する実験を行った。クラウドソーシングを用いたタイピング入力による作文の取得を行った¹²。実験協力者は、Yahoo!クラウドソーシングに登録している15歳以上の男女114名で、150文字以上200文字程度の分量を目安にするよう教示して作文を行った。

結果、平均185文字(max:248文字, min:150文字)の114説明文を得た。オンライン実験の特性上、Wikipediaや辞書類のコピー&ペーストも見られたが、文字数の範囲に貼り付けた部分が各々異なることや、文字数や文末表現などの調整が行われていることを鑑み、すべて調査対象とした。記述内容は1文あたり平均8.2(min:4, max:13)の要素が得られた。

表4に記述割合が上位(25%以上)であった内容とその現れた順位を示す。形容表現については、その説明の有無に別があるため、内訳を示した。半数以上の実験協力者が記述した内容は、鼻が長いこと(96%:「鼻について65%」,「鼻が長いことのみ(詳述なし)44%」,「長いこと+鼻について(後述追記)47%」,「鼻の長さについて(詳述あり)4%」), 大型であること(73%:「大型であることのみ(詳述なし)7%」,「大型であること(詳述あり)66%」), 耳が大きいこと(65%:「耳が大きいことのみ(詳述なし)61%」,「耳の大きさについて(詳述あり)4%」)の3種類であった。象について説明する際、「鼻が長い」「大型」「耳が大きい」ことは重要な要素であると考えられる。

¹² クラウドソーシング実験の前に、手書き作文を取得する実験を行った。実験協力者は3名(20代~50代の男女)で、1回につき5分間の作文を行った。同様に記述を繰り返すことを4回行った。解答用紙は都度回収し、同内容を記述する要請などの条件提示は行っていない。得られた解答数は、3人分×4回の12説明文である。平均299文字(max:448文字, min:170文字)を得た。この結果により、200文字程度と文字数の目安を設定した。

また、記述された順序としても、1 番目に「鼻が長い (39%)」「大型である (30%)」、2 番目に「耳が大きい (24%)」「哺乳類である (18%)」が出現しやすかったという傾向が見られる。

表 4 作文実験における「象」の記述内容とその記述順序 (上位)

記述要素	記述あり	1 番目	2 番目	3 番目	4 番目	5 番目	6 番目	7 番目	8 番目	9 番目
「長い」鼻	96%	39%	18%	19%	8%	9%	1%	2%	0%	1%
(後述追記あり)	47%	20%	11%	8%	2%	5%	0%	1%	0%	1%
(詳述なし)	44%	18%	5%	10%	6%	4%	1%	1%	0%	0%
(詳述あり)	4%	2%	2%	1%	0%	0%	0%	0%	0%	0%
「大型」である	73%	30%	12%	17%	6%	2%	3%	2%	1%	0%
(詳述なし)	7%	4%	2%	1%	1%	0%	0%	0%	0%	0%
(詳述あり)	66%	27%	11%	16%	5%	2%	3%	2%	1%	0%
「大きな」耳	66%	4%	24%	12%	12%	6%	4%	1%	3%	0%
(後述追記あり)	1%	0%	0%	0%	0%	0%	1%	0%	0%	0%
(詳述なし)	61%	4%	23%	12%	12%	5%	3%	1%	3%	0%
(詳述あり)	4%	1%	1%	1%	0%	1%	0%	0%	0%	0%
鼻について	65%	0%	6%	8%	13%	9%	12%	6%	4%	4%
象牙について	47%	0%	1%	5%	6%	5%	11%	8%	7%	2%
哺乳類	35%	11%	18%	4%	4%	0%	0%	0%	0%	0%
生息地	35%	10%	5%	3%	4%	3%	5%	1%	2%	0%
重さについて	31%	0%	4%	4%	12%	8%	2%	2%	0%	0%
動物園にいる	31%	0%	1%	1%	1%	3%	3%	7%	5%	5%
草食である	27%	0%	1%	6%	4%	4%	4%	1%	4%	1%
水浴びをする	27%	0%	0%	1%	4%	11%	2%	3%	0%	1%

4. 考察：象の鼻はどのように長いか

3で得たデータから、テキストに記述される情報からとくに象の鼻の長さがどのように取得されたか見ることで、象の鼻の長さがテキストからどう得られるのか考察する。

4. 1 象の鼻は「長い」

象の「鼻が長い」ことについては、ほぼ全ての種類のテキストから記述が得られた。辞書においては10種全てで、コーパスにおいては対象物そのものについての要素として最頻出(2.2%)で、象の説明作文においては96%で、記述があった。作文で記述される順序を見ても、1番目であることが最も多く(39%)、3番目までには75%が記述される。象の「鼻が長い」ことは、象の形態的な特徴として言及されやすい要素であるといえよう。但し、作文データの詳細を見てみると、具体的な形態の説明や長さを示す記述(比喩表現、例示など)が加えられていたのは4%(以下の(6)(7)など)のみであり、鼻についての詳細説明があった例は47%(以下の(8)(9)など)あるが、残る44%では、その長さの記述が全くない(以下の(10)など)。

- (6) 鼻がホース状で長く牙が左右の口角にある。
 (7) 鼻が長いのが特徴で、立っていても地面に届く程に長い。
 (8) その長い鼻を使って器用に水を飲んだり、高いところにある果実を取る。
 (9) 鼻は器用に動かすことができ、餌を口に運んだり水を飲むことも出来ます。
 (10) 鼻の長い動物である。

また、コーパスから取得した用例は以下のようなものがあつた。(10)に近い(11)(12)のような鼻の長さのみの例や、(8)(9)に類し(13)のように説明の加わる例も見られる。この(13)における「ニュルニュルッと、私の手元めがけて伸びて」くるという鼻の情報は、(6)(7)と同じく具体的な形態を認識することに役立つと考えられる。

- (11) 校長先生に紹介されて、壇の上にあがった上野先生は、ゆっくりと、静かな声で、ぞうの話をはじめました。「ぞうさんは、食べ物をちょうだいと、長い鼻をのぼしながら死にました。(後略) (LBkn_00031: 矢崎節夫『先生のピアノが歌った』)
- (12) 長い鼻がどこか象を思わせる愛敬のある顔が、のぞき込んだ。驚くほど英語がうまい。「どうセカネ目当てだろう。案内なんか知らない」と、いったんは断わつたが、あまりのしつこさに根負けして、とうとう物乞いのガイドで市内の名所を見てまわるはめになった。(LBa3_00045: 五島昭『インドの大地で』)
- (13) 「あなたがミッキー? こんにちは」 息を切らしながら駆け寄る私の前に、突き出されたのは、なんと、ゾウの長〜い鼻!! 輸送用の檻の隙間からニュルニュルッと、私の手元めがけて伸びてきます。(LBs4_00063: 坂本小百合『ゾウが泣いた日』)

しかし、象の鼻は「長い」のであるが、どの程度長いのかという詳細情報がテキストからは得にくい。但し、(14)のように、比喩表現に用いられている場合などには、喩えたものの知識がある場合、具体的な情報の得られる可能性がある。

- (14) だから、医者はお腹だけでなく、必ずからだ全体を診察するのだ。鼻だけを触って、ゾウは蛇のように長い動物だといった寓話もある。木を見て森を見なければ、誤診の道をたどることにもなりかねない。(LBm4_00049: 奈良信雄『名医があかす「病気のたどり方」事典』)

4. 2 象の鼻はどのくらい「長い」のか

今回行った調査では、辞書・コーパス・作文のすべてのテキストで、象の鼻に関して具体的な数値(メートルなど)や比較対象などの記述があつたのは(15)のみであつた。

- (15) 現在の大人のアフリカゾウの鼻の長さは三メートル近くあります。ゾウの鼻が、だんだん長くなってきたのは確かなのですが、どうして長くなったのかという科学的な理由は、現在でもわかっていません。(LBqn_00035: 久道健三『かがくなぜどうして』二年生)

国語辞書では50%が、作文実験においては44.2%が、「長い」とのみ記述しており、具体

的に詳細を示そうとする記述はなかった。これは、象の鼻が「長い」とのみいう場合、比較対象が一般的に予測されるとの前提で記述されているためと考えられる。

たとえば、象の属するあるカテゴリ（アフリカ獣上目）には、同じくハネジネズミやツチブタ（図2）などの「鼻が長い」と評せられるメンバーが含まれている。象をはじめこれらの動物はそれぞれ鼻の長さが異なるが、どれも「長い」と評され得る。しかし、これらはその名前からもそれぞれネズミやブタのようなカテゴリが想定され、ネズミカテゴリやブタカテゴリにおいて「鼻が長い」という他メンバーと異なる特徴を有しているであろう。

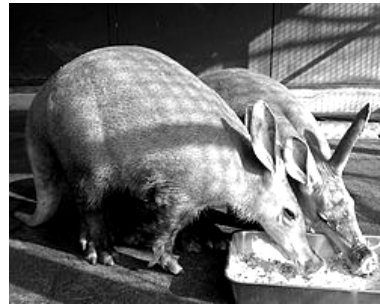


図2 ハネジネズミとツチブタ

<http://ja.wikipedia.org/wiki/ハネジネズミ> より

<http://ja.wikipedia.org/wiki/ツチブタ> より

しかし、辞書では「鼻が長い」と同率を占めた「大型」な動物であることが、作文の73%で記述されていた。大きさについては、「鼻が長い」と異なり、具体的な数値や陸生動物最大であることなどの詳細情報が66%で記述されており、「大型」であることの説明が加えられている割合が高い。「大型」は属するカテゴリ内においてもメンバーの差異として大小をいうことがあるため、一般的に「大型」というものが前提的に想定しにくい可能性が考えられる。大きさについては具体的な情報が必要と判断される場合が多いといえる¹³。

また、身体部位については、言語活動を行う人間も有している部位である場合、言及がなければ人間の部位を比較対象として想定することができるため、あえて正確な記述が必要ない可能性もある。しかし、象の「鼻が長い」ことや「耳が大きい」ことは、人間と比較するに差が大きい。テキストからのみ象の鼻の長さを明確に認識することは困難であろう。

5. まとめ

テキストから対象物に関して得られる情報として、コーパスから取得できる用例の頻度を見ると、場所情報と人間との関係情報が上位となっている（3.2 参照）。また、対象物の説明を試みた場合、特徴的と考えられる形状情報が記述されやすい。とくに形状の情報が一番目に記述されやすく、次いで場所や人間との関係が記述されるという傾向がある（3.1, 3.3 参照）。

動物の象に関するテキストにおいて、全体的な大きさ（「大型」）については説明に補足的な情報が加わっていることが多く（本稿の作文実験では66%）、具体的に程度を説明しようという傾向が見られた。しかし、特徴的部位の長さや大きさは、一般的な程度認識が期待され、具体的な記述が得にくいという結果が見られた。「大型」「鼻」はコーパス・説明文ともに頻度としては上位であるが、補足的な情報は得にくく（半数以下の割合）、具体的な程度は得にくいのである。

¹³ 「鼻が長い」「大型」に続いて高頻度で記述されていたのは「耳が大きい」の65%であるが、その大きさについての詳細は4%にとどまっていた。すなわち、特徴的な身体部位についての「大きい」という形容は、鼻についての「長い」同様、一般的な程度が前提的に期待されている可能性がある。

よって、象の鼻の長さがどの程度であるかという情報は、テキストから得にくいといえる。これは、文化的に標準と考えられる長さや大きさなどが、前提的に必要とされるためであると考えられる。今後、文化的背景の異なる相手への情報伝達において、説明文に何を記述すべきか応用可能性を考えたい。

謝 辞

本研究は JSPS 科研費 26770156 の助成を受けたものである。

文 献

- Goddard, Cliff. and Wierzbicka, Anna. (2014) *Words and Meanings*. Oxford: Oxford University Press.
- Fillmore, Charles. J. and Atkins, Beryl. T. Sue. (1994) “Starting where the dictionaries stop: The challenge for computational lexicography.” In B. T. S. Atkins and A. Zampolli, eds., *Computational Approaches to the Lexicon*, Oxford: Oxford University Press. pp. 349–393.
- 加藤祥 (近刊) 「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— (仮)」『国立国語研究所論集』9
- 加藤祥, 岡本雅史, 荒牧英治 (近刊) 「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, ひつじ書房.
- 国広哲也 (1997) 『理想の国語辞典』, 大修館書店.
- Maekawa, Kikuo, Yamazaki, Makoto., Ogiso, Toshinobu., Maruyama, Takehiko., Ogura, Hideki., Kashino, Wakako., Koiso, Hanae., Yamaguchi, Masaya., Tanaka, Makiro., and Den, Yasuharu.(2014) “Balanced corpus of contemporary written Japanese.” *Language Resources and Evaluation* 48 (2): 345-371 (DOI10.1007/s10579-013-9261-0).
- Sinclair, John. (1992) “Trust the text.” In Davies, M. and L. Ravelli, eds., *Advances in Systemic Linguistics: Recent Theory and Practice*, London: Pinter. pp. 5–19.
- Wierzbicka, Anna (1985) *Lexicography and Conceptual Analysis*. Ann Arbor, MI: Karoma Publishers, Inc.
- Wierzbicka, Anna (1986) *Semantics: Prime and Universals*. Oxford: Oxford University Press.

資料

- 現代日本語書き言葉均衡コーパス (国立国語研究所)
- 三省堂国語辞典 (5 版), 新明解国語辞典 (6 版), 岩波国語辞典 (5 版), 明鏡国語辞典 (初版), 新選国語辞典 (7 版), 集英社国語辞典 (2 版), 角川国語辞典 (新版), 新潮現代国語辞典 (2 版), 大辞林 (3.0 : Web 更新版), デイリー国語辞典 (3 版), COBUILD (2009)
- Kudo, Taku, and Hideto Kazawa. (2007) “Web Japanese N-gram Version 1”, Gengo Shigen Kyokai.

関連 URL

- 現代日本語書き言葉均衡コーパス (国立国語研究所)
- http://www.ninjal.ac.jp/corpus_center/bccwj/
- コーパス検索アプリケーション「中納言」1.1.0, 短単位データ 1.0, 長単位データ 1.0
- <https://chunagon.ninjal.ac.jp/>
- Yahoo! クラウドソーシング <http://crowdsourcing.yahoo.co.jp/>