

「中納言」と「梵天」

浅原 正幸 (国立国語研究所コーパス開発センター) *

加藤 祥 (国立国語研究所コーパス開発センター)

‘Chunagon’ and ‘BonTen’

Masayuki Asahara (National Institute for Japanese Language and Linguistics)

Sachi Kato (National Institute for Japanese Language and Linguistics)

要旨

ジュウシマツはスズメ目カエデチョウ科のスズメに似た鳥であり、観賞用に飼育される。ジュウシマツはその毛並みによって「中納言」「大納言」「梵天」「千代田」「千代田梵天」などに分類される。本稿では、その中で「中納言」と「梵天」の差異について検討する。

1. はじめに

本稿は『言語資源活用ワークショップ 2016』予稿集原稿のテンプレートである。引用などの形式については、本稿に形式に合わせるのが望ましい。本稿ではジュウシマツの形態分類について、多様な観点から検討を行う。

2. ジュウシマツ

2.1 コーパスにおけるジュウシマツ

まず「ジュウシマツ」と「十姉妹」について各コーパスについて度数調査を試みる。調査対象として以下の五つのコーパスを用いる：『国語研日本語ウェブコーパス (NWJC)』(Asahara et al. 2014)、『現代日本語書き言葉均衡コーパス (BCCWJ)』(Maekawa et al. 2014)、『日本語歴史コーパス (CHJ)』(国立国語研究所 2016)、『日本語話し言葉コーパス (CSJ)』(Maekawa et al. 2000)、『多言語母語の日本語学習者横断コーパス (I-JAS)』(迫田久美子ほか 2016)。

検索系として、NWJC は検索系『梵天』(Asahara et al. 2016) を用い、それ以外のコーパスはコーパス検索アプリケーション『中納言』を用いる。

表 1 に各コーパスに出現する「ジュウシマツ」と「十姉妹」の度数を示す。

本調査の重要な点として、NWJC と検索系『梵天』における度数の概念が他のコーパス異なる点がある。NWJC はコーパス構成時にページ単位の異なりをとった後、文単位の異なりをとり、なおかつ『梵天』は一文中に複数件出現する場合には最左要素のみを数えており、UI が出力する件数は異なり件数でも延べ件数でもない。そのまま頻度として用いることはできないので注意すること。

BCCWJ においては、「ジュウシマツ」が 2 事例、「十姉妹」が 32 事例あった。書き言葉においては「十姉妹」表記が多い ($p < 0.005$) ことがわかる。

* masayu-a@ninja.ac.jp

表1 「ジュウシマツ」「十姉妹」の度数表

	ジュウシマツ	十姉妹
NWJC	809	1732
BCCWJ	2	32
CHJ	0	2
CSJ	22	0
I-JAS	0	0

CHJ においては、太陽コーパスに「十姉妹」表記が2事例出現した。

CSJ においては「ジュウシマツ」が22事例出現した。音声言語の書き起こし時にカタカナ表記で記述するよう統制していることの表れだと考える。

I-JAS においては残念ながら「ジュウシマツ」や「十姉妹」が出現しなかった。このことから日本語学習者はあまりジュウシマツに興味がないことがうかがえる。

なお、「十四松」表記はNWJCのみにみられた(54事例)。

2.2 ジュウシマツの語釈

「中納言」と「梵天」の形態分類以前に、日本語学習者は「ジュウシマツ」に興味ないことがうかがえる。そこで、われわれは「ジュウシマツ」の語釈について検討する。

(加藤祥 2015) は、語釈文を呈示して対象を同定できるかどうかのさまざまな実験を行った。以下行った三つの実験について示す。

【実験1】

実験協力者は日本語母語話者の30代～50代の男女(日本語母語話者)20名であり、実験室で行った。動物名は「この動物」などとマスクした。

まず、実験協力者は、提示した記述から何についての説明であるのかを読み取り、テキストの示す対象物を回答する。これにより、正答率の評価を行う。

たとえば、ライオンでは、「ネコ科の猛獣である。」「黄褐色である。雄はたてがみがある。」「アフリカに分布している。」「百獣の王と呼ばれる。」という情報が提示される。実験協力者は、解答するために有用だった情報にマークする。

次に、正答を実験協力者に提示する。この際、当該動物に関する知識がなければ正答はあり得ないため、実験協力者の当該動物の知識有無を確認し、知識率を評価する。よく知っていた場合(100%)、自信はないが知っていた場合(50%)、まったく知らない場合(0%)の三段階解答とした。誤答の場合、実験協力者は正答するために不足だった(記述があれば正答できたはずの)情報を自由に記述する。

表2に実験1の結果を示す。日本語母語話者は「ジュウシマツ」という動物をよく知っている(知識率80%)が、語釈文だけではジュウシマツを同定できないことがわかった。

【実験2】

実験1では被験者から「不足だった情報」についても確認した。次に辞書の語釈に不足していたとされる情報を追加することで、対象物の認識が可能となるのか確かめる。

表 2 辞書語積による「ライオン」「ジュウシマツ」の対象同定実験

	正答率	知識率	分類	形態	生態	人間との関係	その他
ライオン	100%	100%	10%	40%	10%		40%
ジュウシマツ	25%	80%	0%	25%		75%	

対象物を認識するために求められた情報は有用か。先の実験 1 と同条件で実験を行うが、提示する情報に実験 1 で得られた情報を追加する。追加する情報は、複数人が記述した内容を整理したものである。動物名はマスクし、「この動物」などとした。実験室において対面で実施した。

実験協力者は日本語母語話者の 30 代～50 代の男女（日本語母語話者）20 名である。実験協力者は、提示した記述から何についての説明であるのかを読み取り、テキストの示す対象物を回答し、解答に有用だった情報にマークする。

表 3 にジュウシマツの実験 1 と実験 2 の正答率を示す。追加情報を与えた結果、正答率が下がることが確認された。

表 3 辞書語積と追加情報による「ジュウシマツ」の対象同定実験

	知識率	実験 1 正答率 (辞書語積のみ)	実験 2 正答率 (辞書語積+追加情報)
ジュウシマツ	80%	25%	15%

【実験 3】

次に BCCWJ と Google 日本語 N-gram (Kudo and Kazawa 2007) から収集した意味的用例をデータとして用いた。実験は、実験 1・2 と同様の手法によって行う。動物名はマスクし、「この動物」などとした。実験協力者は、提示した記述から何についての説明であるのかを読み取り、テキストの示す対象物を回答する。これにより、正答率の評価を行う。また、実験協力者は、解答するために有用だった情報にマークする。

本実験においては、Yahoo! クラウドソーシングを用いて募集した実験協力者（15 歳以上の男女）1,000 名の回答を得た。

表 4 に結果を示す。コーパス用例を示しても「ジュウシマツ」が当てられないことがわかる。

表 4 辞書語積と追加情報による「ジュウシマツ」の対象同定実験

	正答率
BCCWJ	8.2%
Google N-gram	12.9%

三つの実験をまとめると表 5 のようになる。さまざまな情報を提示しても「ジュウシマツ」を同定することが難しいことがうかがえる。(´; ω; `)

表5 「ジュウシマツ」の対象同定実験まとめ

対象物	知識率	分類名	辞書語釈 (実験室)	求められた情報 (実験室)	BCCWJ (実験室)	BCCWJ (CS)	Google (CS)	平均
ジュウシマツ	80%	正答率	25%	15%	0%	8%	13%	12%
		分類	0%	-	0%	15%	52%	17%
		形態	25%	45%	0%	27%	17%	23%
		生態	-	10%	0%	30%	11%	13%
		人間との関係	75%	65%	0%	35%	20%	39%
		その他	-	10%	0%	2%	31%	11%

2.3 ジュウシマツを同定するために必要な情報

では、対象物同定において誤答の割合が多い場合はどのような原因によるか。図1にジュウシマツの回答を示す。図1では、過半数を上回る60%がブンチョウの誤答である。

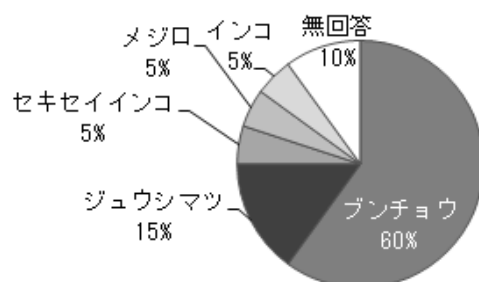


図1 実験2におけるジュウシマツ（正答）と誤答

ブンチョウとジュウシマツは、概ね色味が少ないほか、外観上に似た特徴があるのではない。実験2では、ジュウシマツの「形態」として、「スズメよりやや小さく小形。羽色は豊富であるが、主に白く、茶の不定紋がある。」という情報が提示されている。「主に白く、茶の不定紋がある」は実験1で求められて追加した情報である。これらの情報はブンチョウの外観とそぐわず、ジュウシマツの知識があればブンチョウとの差別化が可能となることが期待された。

また、「人間との関係」における「飼い鳥」「手乗りにもできる」などの情報から「ペットの小鳥」というカテゴリが想定されたと推測される。そこで、「ペットの小鳥（主に白く、茶の不定紋がある）」カテゴリにおいて、限定的か曖昧な知識しか有していなかった場合、同カテゴリのプロトタイプ

(Rosch 1975a,b, Rosch and Mervis 1975, Rosch et al. 1976, Rosch 1978, 1981) としてブンチョウと回答した可能性が考えられる。ジュウシマツの記述から想定されたカテゴリのプロ

トタイプがブンチョウであった実験協力者の割合が高かったのであろう。

対象物によっては、コーパスから取得した情報からの認識が困難で、有用な情報がコーパスからは取得しにくいと見える場合もある。実際にコーパスでは対象物認識のための情報が不足していたのか。あるいは何が正答の邪魔となったのか。誤答の際に有用とされた情報を見ておきたい。

表6は、対象物がジュウシマツ（正答率は8.2%：BCCWJ、12.9%：Google 日本語 n-gram）であった意味用例について、それぞれのコーパスで有用とされた情報の上位を、正答誤答別に示した（提示は文以上の単位で行ったが、表では大意のみ簡略化して記述する）。BCCWJではブンチョウ（誤答）が60%と突出し、Google 日本語 n-gramではブンチョウ（誤答）が37%、次いでインコが34%と分散されている。コーパスの意味用例からジュウシマツを同定した実験（実験3）の回答は図2に示す。

表6を見るに、誤答の際に有用とされたのはどちらのコーパスでも「手乗りにもなる」という情報が目立つ。また、Google 日本語 n-gramの提示情報においては「昔飼った」という情報の次いでいることが着目される。ジュウシマツの分析と同じく、「ペットの小鳥」カテゴリを想定したとき、「手乗り」という情報によっては典型例として想起されたブンチョウは排除されない。

表6 コーパス別対象物認識に有用とされた意味用例（ジュウシマツ上位・複数回答）

Google 日本語 n-gram				BCCWJ			
意味用例	利用 ⁽¹⁾ (正誤)	正答 (12.9%)	誤答 (87.1%)	意味用例	利用 ⁽²⁾ (正誤)	正答 (8.2%)	誤答 (91.8%)
歌に文法の本	正	53.5%	10.6%	日本で作り出した	正	63.4%	12.7%
小型鳥・フィンチ	正	51.9%	30.8%	手乗りにもなる	誤	37.8%	73.6%
手乗りにもなる	誤	26.4%	64.6%	飼い鳥	-	36.6%	46.3%
複数飼い	-	19.4%	19.2%	つぼ巢	-	35.4%	29.5%
昔飼った人が多い	誤	19.4%	38.2%	11cm	正	26.8%	7.6%
同じ餌であわせ飼い	-	17.8%	18.6%	子育て上手	正	25.6%	7.7%
小斑	正	17.1%	4.9%	多品種	-	14.6%	13.8%
展覧会	-	14.7%	8.8%	女子供が珍重	-	3.7%	7.4%

しかし、Google 日本語 n-gram から得られた多くの「昔飼った」という「人間との関係」情報によって個別の経験知識が喚起され、実際に国内飼育数が多いと考えられるインコが活性化されることになったと考えられる。Google 日本語 n-gram から取得された情報は個別の経験知識に関わる種類の情報が多いこともあり、対象物の認識にあたって利用するにはカテゴリの絞り込みがミスリードされる危険を慮る必要があるだろう。

なお、正答で有用とされる割合の高かった「茶の小斑がある（Google 日本語 n-gram）」「日本で作り出された（BCCWJ）」「子育て上手（BCCWJ）」などは、ジュウシマツを「ペット

(1) 有意水準 0.1% 以下で頻度に有意差がある。

(2) 有意水準 0.1% 以下で頻度に有意差がある。

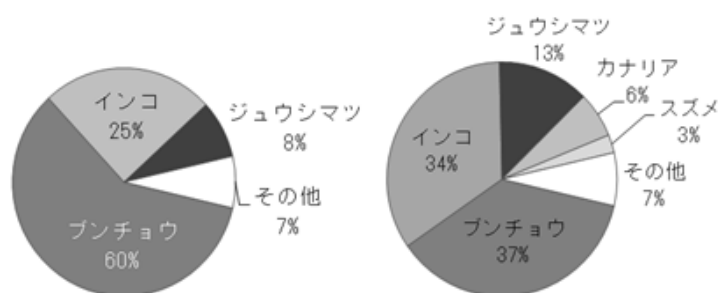


図2 実験3におけるジュウシマツ（正答）と誤答（左：BCCWJ、右：Google 日本語 n-gram）

の小鳥」カテゴリ内で差別化するに有用な情報であったといえる。しかし、誤答を招いた「昔飼った人が多い」のような一般的に得やすい経験とは反対に、これらは実験協力者個々人の有していた知識と合致した場合にのみ有用であったともいえる。

3. 「中納言」と「梵天」の違い

図3に本稿のそもそもの目的である「中納言」と「梵天」の違いについて示す。「中納言」は後頭部の毛が逆立っているジュウシマツである。「梵天」は頭の毛が巻き上がっているジュウシマツである。前からみると区別がつきにくい。



図3 「中納言」と「梵天」の違い

この違いについてコーパス調査を行う。2.1節と同様にコーパスの度数評価を行う。表7に結果を示す。

表7 「中納言」「梵天」の度数表

	中納言	梵天
NWJC	3475	5731
BCCWJ	142	60
CHJ	500	13
CSJ	1	0
I-JAS	0	0

以下に実際に得られた用例について示す。

NWJC に出現する用例を示す際には、NWJC を出典とするのではなく、元の URL を出典とすることが望ましい。リンク先をクリックして、元データが消されていた場合には論文に掲載しないほうがよい。

- (1) 本書には、朱雀院の女四宮、■■■■、頭中将、巢■■■位、内侍典侍、大納言、■■■■の今の上、一品の宮の宣旨、帥中納言上、帥中納言、少将の上、蔵人弁の上、■■■■、少納言、左京大夫、阿闍梨、左京尼といった人物名とその略歴が記されており、巢■■巻の復元資料としてきわめて貴重

【出典】 <http://www.nijl.ac.jp/~koen/200807keizu.htm>

- (2) 本書には、朱雀院の女四宮、源三位、頭中将、巢守三位、内侍典侍、大納言、源三位の今の上、一品の宮の宣旨、帥中納言上、帥中納言、少将の上、蔵人弁の上、三河守、少納言、左京大夫、阿闍梨、左京尼といった人物名とその略歴が記されており、巢守巻の復元資料としてきわめて貴重

【出典】 <http://www.nijl.ac.jp/~koen/200807keizu.htm>

尚、NWJC は、個人情報保護の観点から (1) のように品詞が「名詞-固有名詞-人名-姓」「名詞-固有名詞-人名-名」は■でマスクしている。元のテキストを復元するためには、(2) のようにリンク先から情報を得ること。

BCCWJ に出現する用例を示すには、BCCWJ・サンプル ID・著者名・書名を出典として示すのが望ましい。

- (3) 覚恵は中納言阿闍梨宗恵という

【出典】 BCCWJ サンプル ID：PB29_00611 津本陽 (著) 『弥陀の橋は』

CHJ に出現する用例を示すには、CHJ・サンプル ID・書名を出典として示すのが望ましい。

- (4) 中納言まゐりたまひて、御扇奉らせたまふに、

【出典】 CHJ サンプル ID：20-枕草 1001.00098 『枕草子』

CSJ に出現する用例を示すには、CSJ・講演 ID を出典として示すのが望ましい。

- (5) 二十例の中納言殿おわしますとて経営し合えりその他の五例のうち

【出典】 CSJ 講演 ID：A02F0708

I-JAS に出現する用例を示すには、I-JAS・サンプル ID を出典として示すのが望ましい。

- (6) ケーキとまあ自分の手袋ですね

【出典】 I-JAS サンプル ID：JJJ12-I

I-JAS を利用の際には JJJ12-I が誰なのかなど、必要以上に詮索するのは望ましくない。

4. おわりに

本稿ではジュウシマツ「中納言」と「梵天」の形態分類を試みた。最後にジュウシマツを「結び雁金」風にしたものを示す (図 4)

左から「結び中納言」「結び梵天」「結び千代田」と呼ぶ。「千代田」は胸の毛が巻き上がっているジュウシマツである。

ほかに、頭のぐるりと胸の毛が巻き上がっている「大納言」や全身の羽が巻き上がっている「キング」がある。羽が巻き上がっている分、飛翔能力が低いという欠点もあるが、品評会では



図4 「結び雁金」風ジュウシマツ

高く評価されている。

謝 辞

本研究は国立国語研究所のプロジェクト「多文化共生社会における日本語教育研究」および科研費基盤(A)「海外連携による日本語学習者コーパスの構築－研究と構築の有機的な繋がりに基づいて－」による成果『I-JAS』を利用して行われたものである。

文 献

- Masayuki Asahara, Kikuo Maekawa, Mizuho Imada, Sachi Kato, and Hikari Konishi (2014). “Archiving and Analysing Techniques of the Ultra-large-scale Web-based Corpus Project of NINJAL, Japan.” *Alexandria*, 25:1-2, pp. 129-148.
- Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den (2014). “Balanced Corpus of Contemporary Written Japanese.” *Language Resources and Evaluation*, 48, pp. 345-371.
- 国立国語研究所 (2016). 『日本語歴史コーパス』, (バージョン 2016.10, 中納言バージョン 2.2.1) <https://chunagon.ninjal.ac.jp/> (2016年10月26日確認).
- Kikuo Maekawa, Hanae Koiso, Sadaoki Furui, and Hitoshi Isahara (2000). “Spontaneous Speech Corpus of Japanese.” *Proceedings of LREC-2000 (Second International Conference on Language Resources and Evaluation)* Vol. 2., pp. 947-952.
- 迫田久美子・小西円・佐々木藍子・須賀和香子・細井陽子 (2016). 「多言語母語の日本語学習者横断コーパス」 国立国語研究所プロジェクトレビュー, 6:3, pp. 93-110.
- Masayuki Asahara, Kazuya Kawahara, Hideto Masuoka, Yasuko Ohba, Yuki Torii, Toru Morii, Yuki Tanaka, Kikuo Maekawa, Sachi Kato, and Hikari Konishi (2016). “Bon-Ten’ Corpus Concordance System for NINJAL Web Japanese Corpus.” *Proceedings of COLING-2016. Demo Session. (To Appear)*.
- 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, pp. 23-50.
- Taku Kudo, and Hideto Kazawa (2007). *Web Japanese N-gram Version 1*. 言語資源協会.
- Eleanor H. Rosch (1975a). “Cognitive reference points.” *Cognitive Psychology*, 7:4,

pp. 532–547.

Eleanor H. Rosch (1975b). “Cognitive representtation of sementic categories.” *Journal of Experimental Psychology*, 104:3, pp. 192–233.

Eleanor H. Rosch, and Carolyn B. Mervis (1975). “Family resemblances: Studies in the internal structure of categories.” *Cognitive Psychology*, 7:4, pp. 573–605.

Eleanor H. Rosch, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem (1976). “Basic objects in natural categories.” *Cognitive Psychology*, 8:3, pp. 382–439.

Eleanor H. Rosch (1978). *Cognition and Categorization.*, Chap.Principles of Categorization pp. 27–48. NJ, Hillsdale: Lawrence Erlbaum.

Eleanor H. Rosch (1981). “Categorication of Natural Objects.” *Annual Review of Psychology*, 32, pp. 89–113.

関連 URL

コーパス検索アプリケーション『中納言』 <https://chunagon.ninjal.ac.jp/>

『国語研日本語ウェブコーパス』検索系『梵天』 <http://bonten.ninjal.ac.jp/>