語の表記の違いによって想起される語の傾向が異なる
Effects of Orthographic Variations on Word Recall Patterns in Japanese

加藤 祥[†]，浅原 正幸[‡]
Sachi Kato, Masayuki Asahara
[†]目白大学，[‡]国立国語研究所
Mejiro University, National Institute for Japanese Language and Linguistics
s.kato@mejiro.ac.jp

概要
日本語では，同じ語であっても，様々な表記を選択することが可能であり，表記ゆれが生じる．表記ゆれの原因は文脈や用法において説明される傾向にあるが，表記の違いによる読み手の認識の違いは明らかではない．そこで 9 語の名詞を調査対象とし，語（名詞）を刺激とした一般的な日本語話者の想起語の異同を調査した．想起される語は表記によって異なる傾向があり，受容や生産の傾向との関係が示唆される．

キーワード：表記，名詞，想起語，コーパス

1. はじめに

日本語では，同じ語と考えられる場合であっても，様々な表記を選択することが可能である．漢字，ひらがな，カタカナの 3 種類をはじめ，場合によってはローマ字や算用数字，記号などを用いた表現ができる．同文内においても異表記が出現するような例や，同段落内でそれぞれ複数出現するような例も散見される．しかし，異なる文字列から構成される場合，読み手には本当に同一の語と認識されているかという疑問がある．そこで語（名詞）を刺激とした一般的な日本語話者の想起語を収集し，表記の違いにより読み手の認識の傾向が異なるのか調査した．コーパス用例における表記・共起分布と想起語を対照し，さらに想起語の親密度も確認した．

2. 関連研究と本研究

一般的に，同語と考えられる用例の表記差は，表記ゆれとして一つの見出し語に吸収される傾向にある．同一の見出し語に属する単位語がことなる文字列から構成される場合を「表記ゆれ」とされる[1]．表記ゆれは，新聞に現れる 34,477 語の約 14%[1]，『現代日本語書き言葉均衡コーパス』における 13%[2]の語に生じる．

表記ゆれの発生原因として，文部省指導の内閣告示（1946～1981）に従い，動植物名はカタカナ表記にする[1]傾向や，ひらがな表記が「話し言葉を中心にするやわらかい語り口調」で，カタカナ表記が「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[3]傾向などが考えられている．このほか，「ことばの意義をふくらませる修辞技法」として「字装法」がある[4]．代表表記ではない表記を用いるなどの表記の使い分けは，表記の特性を用いた意味的な違いにつながる可能性があるといえる．実際に，日本語コーパスを用いた表記に関する調査[5]では，異表記の場合には文脈や共起語が異なる傾向が示される．また，語義付きコーパスを用いた調査[6]では，多義語の語義の読み分けに表記の使い分けが関わる可能性が見られる．しかし，コーパスのみを用いた調査においては，たとえばひらがな表記であると「やわらかい」印象が伝わっているのかなど，異なる文字列の語の単位そのものが読み手に与える影響が確認できない．

3. 調査

名詞 9 語（鬼・犬・兎・猫・鶏・椅子・氷・雷・芸術）の漢字（読みを付した），ひらがな，カタカナの各表記を刺激として提示した．それぞれ 100 名の実験協力者が「思いついた語」を異なる 5 種類ずつ記述した．

日本語 ID を有する 18 歳以上の実験協力者を Yahoo! クラウドソーシング（https://crowdsourcing.yahoo.co.jp/）で募集した．但し，不特定多数を対象とした記述式回答調査は一文字回答をはじめとする不適切な回答が多く集まる可能性が高いため，試行を実施した結果，回答の得られる可能性が高いと期待された約 400 名の協力者に調査協力依頼を行うこととした．なお，9 割程度が有効な回答と考えられたため，無効回答の排除は行わなかった．

また，調査対象の 9 語について，『現代日本語書き言葉均衡コーパス』における表記ゆれおよび，表記別の共起語や意味的な用法を確かめた．共起語や用法の確認は，人手により前後の文脈から情報を取得し，判断し

た．用例によっては複数の情報を取得している．

4. 結果

4.1 『現代日本語書き言葉均衡コーパス』における表記分布と表記別の用例・共起分布

まず，調査対象とした名詞 9 語の『現代日本語書き言葉均衡コーパス』における表記分布を表 1 に示す．漢字表記の用いられる名詞は，一般に漢字表記の頻度が 9 割程度を占める傾向が見られる．しかし，「兎」は，カタカナ表記が 51.4%と半数以上を占め，次いでひらがな表記が 18.5%用いられている．また，「鶏」はカタカナ表記が 17.8%，「椅子」はひらがな表記が 14.1%である．音読みをはじめとする別の読みとの区別（「兎」は「兎に角」「兎も角」のような副詞にも用いられる例が多く，「鶏」は「鶏肉」「鶏冠」などの別の読みがある）や，複合語で漢字が続く場合の区別（「車椅子」「長椅子」など）があると漢字で表記されない場合があるといえる．

表 1 『現代日本語書き言葉均衡コーパス』における調査対象語の表記分布

	鬼	犬	兎	猫	鶏	椅子	氷	雷	芸術
漢字	1799	7787	280	5356	1616	4749	1905	991	5441
	91.4%	91.7%	18.5%	76.1%	75.4%	77.2%	96.4%	91.3%	98.3%
ひらがな	94	86	454	380	145	866	16	58	0
	4.8%	1.0%	29.9%	5.4%	6.8%	14.1%	0.8%	5.3%	0.0%
カタカナ	75	594	779	1301	381	539	16	30	1
	3.8%	7.0%	51.4%	18.5%	17.8%	8.8%	0.8%	2.8%	0.0%
その他語形・旧字など	0	28	4	0	1	0	39	7	93
	0.0%	0.3%	0.3%	0.0%	0.0%	0.0%	2.0%	0.6%	1.7%
計	1968	8495	1517	7037	2143	6154	1976	1086	5535
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

次に，調査対象語の表記と用例の分布を確かめておく．表 2 には表 1 で示した「鬼」の用例について，表記別の分布を示す．

表 2 『現代日本語書き言葉均衡コーパス』の「鬼」の用例分布（上位頻度）

用例	鬼	おに	オニ	総計
架空生物（比喩を含む）	1151	12	9	1172
固有名詞	247	45	3	295
遊戯（追いかける役）	77	34	18	129
接頭	33	0	18	51
鬼は外（節分）	41	1	8	50
心を鬼にする	42	0	2	44
～の鬼	32	0	0	32
副詞	29	1	2	32
鬼の首（慣用句）	28	0	1	29
金棒（慣用句）	29	0	0	29
霊魂	18	0	4	22
鬼が笑う（慣用句）	15	0	2	17
誤解析	16	0	0	16
総計	1799	94	75	1968

ひらがなやカタカナ表記では，いわゆる架空生物の種としての「鬼」ではない「鬼」の用例の割合が高い傾向が見られる．特に，節分の用例として「オニは外」が目立つほか，副詞用法や慣用句でカタカナ表記を使用する傾向があるといえる．表 3 には「椅子」の表記別の共起表現分布を示す． 共起表現は，1 つの用例に対し，複数が取得された場合がある．たとえば，漢字の続く複合語でひらがな表記「車いす」の頻度が高い（「車」共起用例 1,090 件中 425 件が「車いす」（表 3 における 425 件），「車椅子」は 609 件）．ひらがな表記で「乗る」が共起しやすいのは，「車いす」の影響による．「車」と「乗る」は 1 例から収集された例が多い．その他，カタカナ表記では「テーブル」というカタカナ語の同種の表記の選好が推測されるが，表記による大きな共起表現の分布の差異，用例の分布は見られない．但し，「背」や「腰掛ける」「腰を下ろす」「立つ」などの動作と共起しやすいのは漢字表記であるという特徴が見られる．

一般的な表記が漢字である名詞については，ひらがな表記やカタカナ表記が用いられる場合，一般的でない用例であることを示すか，あるいは誤読を避けるための区別の影響が見られるものの，表記差に

より特段の違いが現れるということはない.

表 3　『現代日本語書き言葉均衡コーパス』の「椅子」の共起分布（上位頻度）

共起表現	椅子	いす	イス	計
車（車椅子も含む）	609	425	56	1090
座る（座り込む・坐る等を含む）	838	96	89	1023
腰掛ける（腰（を）かける・こしかける等を含む）	208	21	16	245
腰を下ろす（腰を（椅子に）下ろす等も含む）	209	3	3	215
テーブル	163	14	35	212
長（長椅子）	188	10	3	201
立つ・立ち上がる	168	9	8	185
机	114	22	17	153
置く	122	10	12	144
かける（腰をかけるは含まない）	113	7	7	127
肘掛（肘掛椅子も含む）	103	3	1	107
誤解析	0	32	73	105
背（背もたれは含まない）	90	5	4	99
乗る	65	26	3	94
凭れる（背もたれは含まない）	74	3	3	80
～の椅子（慣用句のみ・実際の椅子でない）	67	2	5	74
引く	54	4	2	60

4.2　表記別の想起語

以下では，想起語の収集結果を示す．オンライン上の調査であるため機器によるかな漢字変換の可能性が考えられるが，想起語において同表記の選好傾向は見られなかった．また，カタカナ表記であるために生態情報や身体部位などが表れるなど，動物名であることによる特徴も確認されていない．しかし，表記による連想語の頻度分布には差が確認される．

表 4 は「鬼」の例である．また，表 5 には，抽象名詞の「芸術」の例を示す．想起された語の表記別の上位語を示している．

表 4　＜鬼＞から想起された語
（表記別上位語例）

鬼		おに		オニ	
金棒	38	節分	36	節分	47
節分	30	桃太郎	27	赤	26
桃太郎	27	金棒	26	金棒	23
赤	26	赤	26	怖い	22
怖い	25	青	17	桃太郎	21

表 4 では，概ね類似したが語が上位で現れているものの，漢字で「金棒」（38 件／表記総計 87 件），カタカナで「節分」（47 件／表記総計 113 件）が突出する．「鬼に金棒」の慣用句や「金棒」との共起については，コーパス（表 2）でカタカナやひらがなの表記は見られていないが，想起語としてはどの表記においても取得されている．ひらがな表記では色彩語として「赤」に加え「青」も上位で取得されている．

表 5　＜芸術＞から想起された語
（表記別上位語例）

芸術		げいじゅつ		ゲイジュツ	
絵画	35	絵画	49	絵画	46
彫刻	26	音楽	26	秋	29
秋	22	彫刻	26	爆発	28
美術館	20	爆発	23	音楽	25
爆発	19	秋	20	彫刻	24
音楽	16	美術館	18	美術館	16
絵	14	アート	12	岡本太郎	13
アート	10	作品	11	美術	13
ピカソ	8	美術	11	アート	10
才能	7	絵	10	絵	10

表 5 の「芸術」では，漢字で「絵画」（35 件／表記総計 130 件）や音楽（16 件／表記総計 67 件）な

どの下位カテゴリ名が出にくいが，想起語が他表記（ひらがな 209 種・カタカナ 211 種）よりも多様で（248 種）ある．また，カタカナで「爆発」（28 件／表記総計 70 件）「岡本太郎」（13 件／26 件）の頻度が高い．

このほか「犬」では，漢字で「散歩」（41 件／表記総計 104 件）や「吠える」「走る」などの動詞が出やすく，ひらがなでは「猫」（25 件／表記総計 102 件）が出にくい傾向が見られた．

表 6 では「椅子」の例について，上位頻度の想起語とその親密度[7]および表記別の頻度を示す．コーパス（表 3 参照）では確認されないが，ひらがなで「背もたれ」（3 件／表記総計 32 件）や「足・脚」（6 件／同 31 件）などの構成要素が出にくく，同カテゴリの家具類の想起では「テーブル」では表記差がないものの，漢字で「机（27 件／表記総計 112 件）」が出にくい傾向が確認される．

漢字でない表記で想起しやすい「机」や「学校」は，受容（読む・聞く）において親密度が高い．カタカナと漢字表記で出やすい「背もたれ」は音声（話す・聞く）において親密度が高い．また，「座る」「腰掛ける」などの動作はカタカナで想起しにくい（表 6）ようである．「腰掛ける」では知っているに対し受容・生産のいずれにおいても親密度が低い（表 6）．一般的な表記が漢字表記である場合，動作の想起は漢字表記と関連する可能性がある．

表 6　＜椅子＞から想起された語（上位語例）の親密度と表記別頻度

	知っている	書く	読む	話す	聞く	椅子	いす	イス	計
座る	1.87	0.92	1.32	1.35	1.51	62	62	47	171
机	2.11	0.94	1.24	1.19	1.30	27	38	47	112
テーブル	1.91	0.95	1.07	1.40	1.46	21	24	24	69
ソファー	1.99	0.59	1.06	1.07	1.09	11	14	17	42
木	1.93	0.74	1.02	1.06	1.13	10	13	9	32
背もたれ	1.83	-0.03	0.39	0.94	1.05	17	3	12	32
足・脚	2.01	1.06	1.41	1.42	1.35	13	6	12	31
家具	1.89	0.77	1.22	1.26	1.25	7	10	11	28
学校	1.96	0.89	1.46	1.52	1.60	4	11	7	22
座椅子	1.78	0.15	0.56	0.61	0.67	5	8	6	19
クッション	2.14	0.25	0.57	0.80	0.78	4	6	8	18
ベンチ	2.04	0.33	0.96	0.80	0.99	6	6	6	18
腰掛ける	1.99	0.14	0.63	0.71	0.84	6	7	2	15

5.　おわりに

本稿では，表記が異なって提示された場合に想起される語で異なる傾向のあることを確認した．漢字表記は名詞一般において最も使用されやすい傾向があるため，慣用句や動詞なども多様に想起されやすい．カタカナ表記は，節分のような特定行事やキャッチコピーをはじめとする，音声に関わる想起語の頻度が高い傾向があり，聴覚的な経験の影響が考えられる．ひらがな表記は，日常的な経験に関わる想起語の頻度が高い傾向にある．

謝辞

本研究は科研費 22H00663 および国語研共同研究プロジェクト「アノテーションを用いた実証的計算心理言語学」の支援を受けた。

文献

[1] 国立国語研究所（1983）,現代表記のゆれ.
[2] 小椋秀樹（2012）”コーパスに基づく現代語表記のゆれの調査—BCCWJ コアデータを資料として—”,第 1 回コーパス日本語学ワークショップ, pp.321-328.
[3] 山口仲美（2006）,日本語の歴史.
[4] 中村明（2007）,日本語の文体・レトリック辞典.
[5] 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”, 認知言語学論考,12, pp.81-113.
[6] 加藤祥（2019）”表記によって異なる語義を読み取るのか—多表記多義語実態調査の試み—”,ことばと文字,12, pp.86-94.
[7] 浅原正幸（2022）” クラウドソーシングによる単語親密度データの構築 (2021 年版)”, 言語処理学会 第 28 回年次大会 発表論文集, pp.800-805.