

2つの異形態動詞「Xじる」と「Xずる」のジャンル
ごとの分布とコロケーション特徴：
「現代日本語書き言葉均衡コーパス」(BCCWJ)を利用した研究

金城 克哉

2つの異形態動詞「Xじる」と「Xずる」のジャンルごとの分布と
コロケーション特徴：
「現代日本語書き言葉均衡コーパス」(BCCWJ)を利用した研究

金城 克哉

0. はじめに

次のデータ例 (1) および (2) が示すように、日本語には「生じる」と「生ずる」のように異形態でありながら同じ意味を表す一群の動詞がある：

- (1) 退職前に義務者である配偶者が死亡した場合には、勤務先の規定による死亡退職金の受給権者に対して、元配偶者が支払を請求できるのかという問題も生じる。

『離婚判例ガイド』

- (2) すべての都道府県にまず一選挙区を配分してしまうと、残りの選挙区をいかに人口に正確に比例して配分しようとも、選挙区を過大に配分された都道府県と、過小に配分された都道府県との間にかなりの格差が生ずる。

『プロセス演習憲法』

これらの動詞を本稿では「Xじる」・「Xずる」とする。「X」は漢字一文字の場合もあれば、「重んずる／重んじる」のように、漢字と平仮名から成る場合もあるが、ここでは等しく「X」で表す。

管見ではこの「Xじる」と「Xずる」の差異や使用および分布状況について調査を行っている研究がなく、これらの動詞がどのように用いられているかの実態が十分に明らかになっていないとは言えない。「Xじる」と「Xずる」が同等の意味を表し、相互に互換性があるのであれば、双方の出現頻度にはそれほど差は見られないであろう。しかしながら用法に意味以外の語用論的・文体的な影響があるとすれば、それは出現頻度やある種の文体やジャンルの特徴となって現れるに違いないと考える。

1. コーパス研究と本稿のリサーチクエスション

まず、コーパス研究について概観する。Chomskyが言語研究の目標を普遍文法の解明に

設定して以来、理論言語学では主に議論を構築する際に、実際の言語データを収集してそれをもとに議論がなされるというケースはそれほど多くはなかった。Chomsky が非文法的とした*Colorless green ideas sleep furiouslyのような例に限らず、文法理論の中核を成す統語論の議論は、そのほとんどが研究者の内省によって作成した文をもとに構成されてきた。しかしながら、このような文法性の判断は言語事実に反するデータを是としたり、研究者の個人語 (idiolect) による判断がなされたりする恐れもある (赤野2006)。このような言語理論の内部の議論との比較で言うならば、コーパスの果たす役割は特定の文法理論の構築に貢献するというよりは、より「経験主義的な言語記述」(齋藤・他2005: 4) を目指した一つの方法論ということができるだろう。研究という側面から言えば、具体的には当該言語の運用サンプルを収集し、コロケーション情報 (後述のセクション4考察を参照) や文法情報、文体情報、社会言語学的情報など様々な情報が得られる点にある (大曾2006)。一方、言語教育という視点からコーパスを見ると、実際の言語の使用状況が把握できることで、それが教材開発や教材分析へとつながる (石川2008) という利点もある。また学習者の産出したデータをもとに誤用の傾向を分析するばかりでなく、学習者自らが主体的に学ぶという学びの変化へとつながっていく可能性を持つと言える。

このようなコーパスを用いた言語研究はコーパス言語学 (corpus linguistics) と呼ばれ、「コンピュータで処理可能な電子コーパスを検索して言語分析・記述を行う言語学一般」を指すとされる (齋藤・他2005: 3)。言語研究では従来、理論言語学の核を成すとされる音韻・形態・統語・意味のそれぞれの「論」があり、それとは別に社会言語学や歴史言語学などの名称が用いられてきた。それに対して「コーパス言語学」という名称は比較的新しいもので、「コーパス」という言語資料を名に冠するという点で従来の手法とは異なっており、パーソナル・コンピュータの普及・操作性や容量の向上といった背景も手伝って1990年代から広まってきたものである (田野村2010)。

本稿で用いたものと同じコーパスである『現代日本語書き言葉均衡コーパス』(BCCWJ) を利用した研究には、類義表現の「～てならない」「～てたまらない」「～てしかたない」「～てしようがない」に関して、前接語彙とジャンルという2つの観点からその使い分けを分析した鄭・小池・船橋 (2009) の研究、基本オノマトペの提唱を行なっている黄 (2010) の研究、接続助詞ガとケレド類の機能について通時的側面から調査を行った宮内 (2010) の研究、通常同等の表現として扱われる「くらい」と「ぐらい」の使用上の差を明らかにした金城 (2011) の研究などがある。また、本稿と同様にジャンル別の語彙の特徴を探った研究に内田・藤井 (2009) がある。内田・藤井 (2009) は『BCCWJ領域内公開データ2008年度版』の語彙頻度表からジャンル別に上位100語を用いてクラスター分析を適用し、「動詞・名詞ともに教科書が書籍とは異なった傾向を示す」(2009: 443) もの、「教科書の中では『国語』が書籍コーパスと近い振る舞い」をすることを明らかにしている。たとえば、ジャンルごとの特徴語として「産業」のジャンルでは「伸びる」と「売る」が抽出され、

それぞれFrameNetではExpansionとCommerce-sellというフレームを喚起するとされる。

本稿でのリサーチクエスションは以下のとおりである：(1) 「現代日本語書き言葉均衡コーパス」(BCCWJ)において「Xじる」と「Xずる」動詞の出現頻度を調査し、高頻度語を特定する。(2) 「Xじる」と「Xずる」動詞の分布に特徴的な点があるか、クラスター分析を用いて日本十進分類法(NDC)におけるジャンルごと(例：総記・哲学・歴史など)の分布状況を把握する。(3) 分布状態に特徴が見られた場合、ジャンルを特徴づける動詞のコロケーションを調べ、考察を加える。

2. データと分析手順

『現代日本語書き言葉均衡コーパス』(BCCWJ)は、大学共同利用機関法人人間文化研究機構国立国語研究所と文部科学省科学研究費特定領域研究「日本語コーパス」プロジェクトが共同で開発した現代日本語の書き言葉を代表するコーパスであり、約1億480万語からなる書き言葉を対象としたコーパスである(詳しくは前川(2009)および前川・山崎(2009)を参照のこと)。本稿ではこの「現代日本語書き言葉均衡コーパス」(BCCWJ)がインターネット上で利用できる少納言(<http://www.kotonoha.gr.jp/shonagon/>)という検索サイトを用いて、書籍(1971～2005年、22,058件、約6,270万語)の項目から「Xじる」と「Xずる」という動詞の終止形を機械的に収集し、データマイニングソフトウェアであるKH Coder¹を用いて粗頻度表を作成した(表1・2)。それぞれのジャンルではコーパスの大きさが異なるため、100万語あたりの調整頻度に換算し²、「Xじる」・「Xずる」それぞれの高頻度語20語について、「総記」・「歴史」・「哲学」といったジャンルごとの出現頻度を求めた。統計処理を行う際には、データを(Excelなどを用いて)頻度表の形で整理するが、その頻度表では「列方向(縦方向)に入るデータを変数、行方向(横方向)に入るデータをケース(または「サンプル」「検体)」と呼ぶ(石川・他2011:166)。今回は上位20語をケース、ジャンルを変数として変数を分類する変数クラスター分析を行った^{3・4}。

3. 結果

表1で「Xずる」動詞の総出現数を比較すると「生ずる」(944)の出現頻度が最も高く、以下「信ずる」(307)、「感ずる」(301)、「通ずる」(264)、「論ずる」(245)などの順となっている。また、表2で示した「Xじる」動詞では、「感じる」(3,555)が最も多く、以下「生じる」(2,463)、「信じる」(978)、「通じる」(909)、「演じる」(393)などの順となっている。上位20語で見ると、「Xじる」動詞を100とした場合、「Xずる」動詞の割合は約32となっており、一口に異形態といっても出現頻度に差があることが確認された。

表1 「Xずる」動詞のジャンル毎の粗頻度表

	総記	哲学	歴史	社科学	自科学	技・工	産業	芸・美	言語	文学	分類無	計
1 生ずる	15	74	40	544	96	65	21	24	11	53	1	944
2 信ずる	7	88	63	35	6	4	3	10	3	84	4	307
3 感ずる	3	53	39	54	17	9	7	16	3	100	0	301
4 通ずる	4	39	62	49	10	7	8	21	6	57	1	264
5 論ずる	10	35	31	88	12	14	9	10	7	26	3	245
6 準ずる	5	1	6	133	16	34	5	4	7	11	1	223
7 命ずる	0	16	18	96	2	19	7	7	1	49	2	217
8 応ずる	2	13	6	54	5	5	3	8	3	31	2	132
9 演ずる	5	6	9	23	3	5	2	23	2	18	2	98
10 案ずる	2	4	8	5	3	0	0	3	0	61	4	90
11 重んずる	1	14	6	13	4	1	0	6	3	21	0	69
12 禁ずる	2	6	13	21	6	0	1	1	1	14	2	67
13 転ずる	0	6	11	13	2	1	0	2	1	13	1	50
14 投ずる	0	1	8	5	4	0	0	1	0	15	1	35
15 報ずる	3	4	6	5	0	1	0	3	0	11	0	33
16 任ずる	2	2	5	7	0	0	1	0	0	13	0	30
17 減ずる	2	3	2	9	6	3	0	0	0	4	0	29
18 断ずる	1	2	3	8	1	0	0	0	0	9	0	24
19 殉ずる	0	3	7	6	0	0	0	0	1	7	0	24
20 長ずる	0	2	4	3	1	0	0	4	0	8	0	22
計	64	372	347	1171	194	168	67	143	49	605	24	3203

表2 「Xじる」動詞のジャンル毎の粗頻度表

	総記	哲学	歴史	社科学	自科学	技・工	産業	芸・美	言語	文学	分類無	計
1 感じる	63	380	195	637	399	257	97	276	70	1070	111	3555
2 生じる	43	194	62	851	623	280	91	70	62	174	13	2463
3 信じる	11	174	75	155	33	16	10	47	29	422	6	978
4 通じる	12	48	93	187	48	20	39	61	31	361	9	909
5 演じる	8	17	28	84	16	11	15	108	7	90	9	393
6 論じる	22	46	31	132	29	22	14	23	10	32	1	362
7 応じる	4	15	26	115	7	7	12	8	5	134	4	337
8 命じる	3	15	24	64	8	6	2	12	2	102	2	240
9 重んじる	1	20	24	39	4	8	1	5	4	37	4	147
10 転じる	2	9	10	40	3	5	4	6	2	29	0	110
11 恥じる	2	6	7	14	2	3	1	3	2	45	1	86
12 封じる	0	0	5	4	2	0	3	9	1	52	0	76
13 禁じる	2	8	11	19	5	2	0	4	1	8	1	61
14 投じる	1	4	9	15	1	2	1	6	0	13	0	52
15 案じる	0	2	1	3	1	0	0	1	0	42	1	51
16 準じる	0	1	3	18	10	0	1	4	3	2	1	43
17 興じる	0	1	11	9	0	0	2	1	0	15	2	41
18 報じる	2	0	4	10	0	0	4	5	1	9	1	36
19 動じる	0	1	0	1	2	0	0	1	1	26	3	35
20 奉じる	1	3	4	4	3	4	0	1	0	7	0	27
計	177	944	623	2401	1196	643	297	651	231	2670	169	10002

図1と図2の樹形図ではグラフの上方へ向かって段階的にジャンルの融合が進んでいく様子が示されている。双方ともクラスター数を3に設定してある。横に引かれた破線はカッティングポイント（クラスター数に大別するための分割線）（石川・他2011：180）である。20の「Xずる」動詞をケースとしてクラスター分けをした図1の顕著な特徴は「哲学」と「社会科学」が第2・第3クラスターとしてそれぞれ1つの群を成している点である。一方、「Xじる」動詞20を用いたクラスター分けでも「哲学」が1つのクラスターを成しており、また図1では独立していた「社会科学」がここでは「技術・工学」さらに「言語」、「自然科学」と融合して1つのクラスターとなっている様子が見て取れる。

このような樹形図は、いかにグループ分けが成されているのかが直感的に分かるように視覚化する手段としては優れているが、個々のグループがどのような特徴を持っているのかという点にまでは踏み込んで示すことができない。それを補うのが表3と表4である。これらの表では変数（本稿では「総記」「自然科学」などのジャンル）とクラスターの関係性が示されている。変数別最大クラスターは薄いグレー、変数別最小クラスターは濃いグレーの網掛けで示す（石川・他2011：181）。「生ずる」の項を横方向に見てみると、第1クラスターに最小値が、第3クラスターに最大値が示されているが、このことより、「生ずる」という動詞は「総記」その他がまとめられた第1クラスターに関しては関与の割合が低いことを、逆に「社会科学」のみで成り立つ第3クラスターでは33,945という高い関与が示されていることがわかる。

4. 考察

以下、主に「Xずる」をとりあげて見ていく。表3の第3クラスターは「社会科学」という1つのジャンルのみから成り立っており、このクラスターを特徴づける語には「生ずる」・「準ずる」・「命ずる」の3つがあることがわかった。この中の「生ずる」を詳しく見てみよう。社会科学での「生ずる」の頻度は544、そのうち「～が生ずる」（188例 35%）、「～を生ずる」（148例 27%）、「～から生ずる」（74例 14%）、「～に生ずる」（45例 8%）となっており、自動詞としての用例が最も多いことがわかった。また、この自動詞の用例でのコロケーションを調べてみると、表5のように「問題が生じる」の用例が最も多く、それに「効力が生ずる」、「必要が生ずる」が続いている。

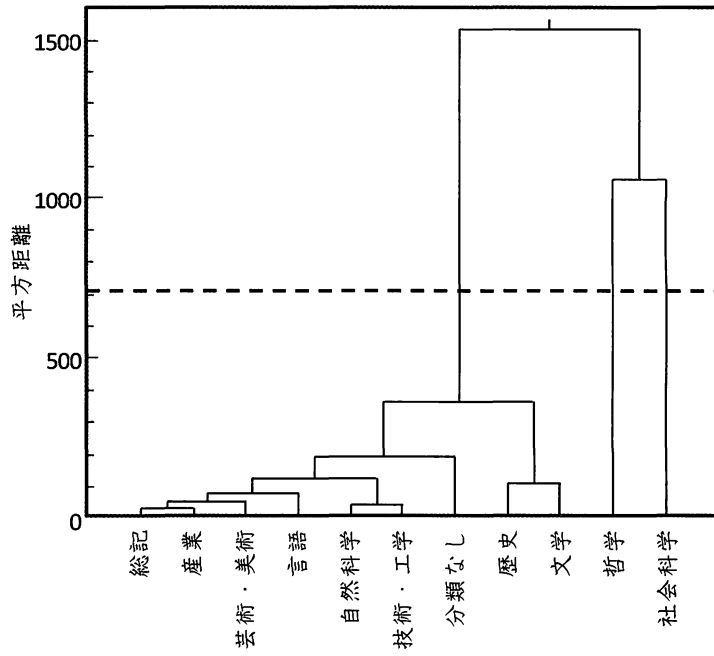


図1 樹形図（「Xずる」による分類）

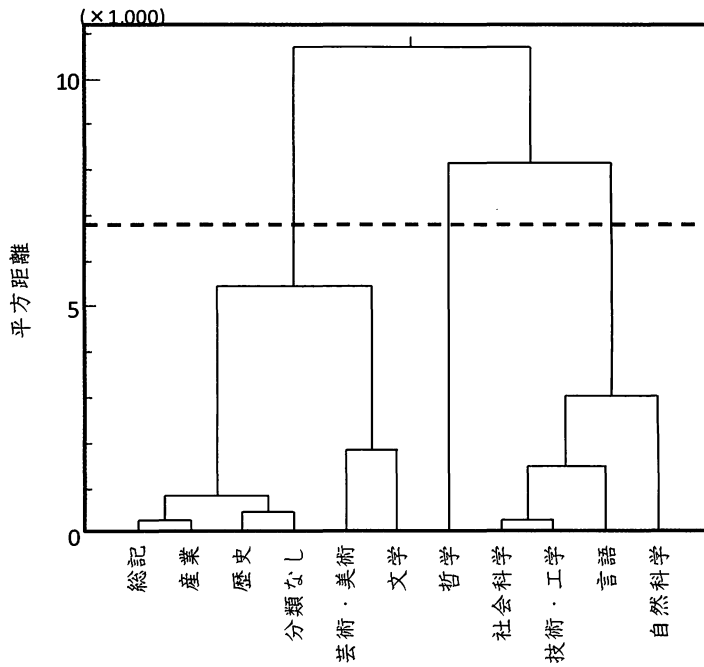


図2 樹形図（「Xじる」による分類）

表3 クラスター間平均 (Xずる)

	: 検体別最大クラスター		
	: 検体別最小クラスター		
	第1クラスター	第2クラスター	第3クラスター
生ずる	7.477	22.060	33.945
信ずる	3.404	26.234	2.184
感ずる	3.302	15.800	3.369
通ずる	3.763	11.626	3.058
論ずる	3.260	10.434	5.491
準ずる	2.407	0.298	8.299
命ずる	1.832	4.770	5.990
応ずる	1.385	3.875	3.369
演ずる	1.709	1.789	1.435
案ずる	1.111	1.192	0.312
重んずる	0.901	4.174	0.811
禁ずる	0.835	1.789	1.310
転ずる	0.581	1.789	0.811
投ずる	0.431	0.298	0.312
報ずる	0.477	1.192	0.312
任ずる	0.364	0.596	0.437
減ずる	0.340	0.894	0.562
断ずる	0.212	0.596	0.499
殉ずる	0.301	0.894	0.374
長ずる	0.276	0.596	0.187
群内項目	1 総記 3 歴史 5 自然科学 6 技術・工学 7 産業 8 芸術・美術 9 言語 10 文学 11 分類なし	2 哲学	4 社会科学

それに対応する「生じる」も表4の第3クラスターを特徴づける語の1つとなっているが、この場合は社会科学だけに限らず、自然科学、技術・工学、言語が1つのクラスターとなっている点が「生ずる」の場合とは異なっている。ここでも社会科学・自然科学・技術・工学・言語の各分野における「生じる」の頻度1,816例のうち「～が生じる」(937例 52%)という自動詞の用例が最も多く、それに「～に生じる」(234例 13%)、「～を生じる」(188例 10%)が続く。「生ずる」の場合には自動詞と他動詞の用例差がそれほどなかったのに比べ、「生じる」の場合には自動詞の用例数が他動詞の用例数を大きく上回っていることがわかる。コロケーションでは、「問題」と共起する例が最も多い点は「生ずる」の場合と同じであるが、「生ずる」の場合にはみられなかった「変化が生じる」というコロケーションが頻出していることが特徴としてあげられる。次にデータ例をあげる：

- (3) 弁済の効果に関しては、次のような問題が生ずることがある。(1) 弁済の充当 X のYに対する債務が一つである場合… 『基本民法』⁵

表4 クラスタ間平均 (Xじる)

	: 検体別最大クラスター		
	: 検体別最小クラスター		
	第1クラスター	第2クラスター	第3クラスター
感じる	48.649	113.282	51.202
生じる	16.520	57.834	62.317
信じる	11.741	51.871	10.668
通じる	13.948	14.309	12.343
演じる	8.394	5.068	3.908
論じる	4.896	13.713	6.268
応じる	4.199	4.472	3.444
命じる	3.074	4.472	1.991
重んじる	1.798	5.962	1.968
転じる	1.331	2.683	1.384
封じる	1.401	0.000	0.357
禁じる	0.815	2.385	0.789
投じる	0.825	1.192	0.356
案じる	0.707	0.596	0.085
準じる	0.393	0.298	1.316
興じる	0.810	0.298	0.140
報じる	0.893	0.000	0.374
奉じる	0.335	0.894	0.345
減じる	0.314	0.000	0.455
断じる	0.169	0.000	0.104
群内項目	1 総記 3 歴史 7 産業 8 芸術・美術 10 文学 11 分類なし	2 哲学	4 社会科学 5 自然科学 6 技術・工学 9 言語

表5 「～が生ずる」

1 問題	21
2 効力	12
3 必要	11
4 欠損	7
5 権	6
6 効果	6
7 疑問	5
8 差異	5
9 支障	5
10 関係	4
11 既判力	4
12 損害	4
13 影響	3
14 差	3
15 状態	3

表6 「～が生じる」

1 問題	55
2 変化	24
3 必要	20
4 差	17
5 性	17
6 支障	16
7 違い	13
8 義務	13
9 事態	13
10 障害	13
11 疑問	12
12 など	11
13 関係	10
14 差異	10
15 効果	9

次に「準ずる」に前接する語を調べてみると、「これ」「これら」「それ」といった指示語が前接するデータ例が全体（133例）の半数を上回る72例（54.14%）あった：

- (4) 新規公開申請会社は、証券取引所や日本証券業協会の規則により、これに準ずる監査を受けることを義務づけられています。 『株式公開の知識』

表7 「～に準ずる」に前接する語

1	これ	42
2	これら	27
3	役員	6
4	号	5
5	会社	4
6	それ	3
7	会報	3
8	法	3
9	貨幣	2
10	金	2

社会科学を特徴づける3つ目の「Xずる」動詞は「命ずる」である。96例中「～を命ずる」は85例（88.5%）、「～に命ずる」は5例（5.2%）であった。目的語として「を」に前接するのは「～ことを命ずる」（11例 12.9%）が最も多く、以下、「勤務を命ずる」（8例 9.4%）、「～旨を命ずる」（7例 8.2%）の順となっている。一方、「命じる」の場合は「哲学」のクラスターを特徴づける語の1つとなっており、こちらも目的語を取る例が多いが（15例中7例 46.7%）、「命ずる」とは異なり次に多い例は引用の「と」に導かれる例である（15例中5例 33.3%）。

- (5) 採用時に「栄養士として採用するが、場合によっては他の業務を命ずることもある」と労働条件通知書に明記しておくのも1つの方法でしょう。

『社会福祉施設のための労働基準法Q&A』

以上、3つの「Xずる」動詞、「生ずる」「準ずる」「命ずる」が社会科学の分野に特徴的であること、また対応する「Xじる」動詞とは異なる独特のコロケーションがあることが明らかとなった。社会科学は自然科学とその方法論や研究対象の独自性で比較・対照されることが多いが、本論では自然科学には「Xずる」動詞がそれほど見られないことが明らかとなった。文体の特徴というものはいずれの動詞を用いるかに大きく左右されるが、社会科学と自然科学ではその論述の仕方、「(対象について) どのように述べるか」が大きく異なることがわかった。また、自然科学よりは社会科学に近い哲学の分野も17の「Xずる」

動詞が特徴的に見られることがわかった。

「Xずる」動詞は全体的に擬古的な印象を読む側に与える。例えば「問題が起こる」よりも「問題が生じる」のほうが、また「問題が生じる」よりはさらに「問題が生ずる」の方が堅く古めかしいという印象を与える。文体特徴についてはさらに調べる必要があるだろうが、少なくとも「Xずる」動詞の使用によって社会科学や哲学のジャンルの文章はその他のジャンルよりも少し古めかしく厳しい印象を与えることになっていると言えそうである。

5. おわりに

本稿では2つの異形態動詞「Xずる」と「Xじる」のジャンルごとの分布とコロケーション特徴を探った。(1)「現代日本語書き言葉均衡コーパス」(BCCWJ)において「Xずる」と「Xじる」動詞の出現頻度を調査し、「Xずる」動詞では「生ずる」から「長ずる」までの高頻度語を特定した。また「Xじる」動詞では「感じる」から「奉じる」までの20の高頻度語が明らかとなった。(2)「Xずる」動詞と「Xじる」動詞の分布では、クラスター分析を用いることで日本十進分類法(NDC)における「社会科学」と「哲学」の2つの分野が「Xずる」動詞によって特徴づけられることがわかった。さらに(3)動詞のコロケーションでは、社会科学の群に顕著な3つの動詞「生ずる」「準ずる」「命ずる」のうち、「生ずる」では自動詞の用法、特に「問題が生ずる」という用例が最も多く見られ、また「準ずる」では指示代名詞「これ」や「これら」とのコロケーションが全体の用例の半数以上あることが明らかとなった。

今回明らかになった最も重要な点は「Xずる」と「Xじる」はどちらを用いても同じというわけではなく、用いられるジャンル・文体に特徴があること、すなわち少なくとも自然科学の分野では「Xずる」動詞の使用は「Xじる」ほど一般的ではないという点である。今後はこれらの動詞だけに限らず、広い意味でジャンルごとの特徴、例えば社会科学を特徴づける文体などの研究が進めば、「社会科学の文体は自然科学の文体とどのように異なるのか」という点がさらに明らかになると思われる。

また、今回は時間的な制約があり十分な分析が行えなかったが、BCCWJは「現代」に焦点を合せてはいるが、年代的にかなり幅のあるデータを収集するという方法をとっている。査読者からの指摘にもあったが、今後はBCCWJのそういった利点を生かし、通時的な(史的な)使用状況の推移なども考慮に入れた研究へ発展させたいと考える(通時的な側面をとりあげたコーパス研究には松田(2008)がある。興味のある方はそちらも参照されたい)。

謝辞

本稿をまとめるにあたり、2名の査読者から大変示唆に富む有益なコメントをいただきました。ここに記して感謝申し上げます。

注

1. KH Coderは立命館大学の樋口耕一氏によって開発されたフリーのソフトウェアで、本稿ではVersion 2.beta29を用いた。KH Coderでは形態素解析器として茶筌とMeCabのいずれかが選択できるようになっているが、今回は茶筌を利用した。
2. 各ジャンルの総語数はインターネット上には示されていないが、以下の総文字数と割合を参考にし、例えば「総記」は6270万語の3.37%と捉え、おおまかな数字を求めた。尚、表8のデータは現代日本語書き言葉均衡コーパスのためのサンプル抽出より転記したものである。

表8 各ジャンルと総文字数

	文字数	%
0. 総記	1,636,414,548	3.37%
1. 哲学	2,597,610,813	5.35%
2. 歴史	4,301,204,340	8.86%
3. 社会科学	12,408,321,943	25.56%
4. 自然科学	5,069,594,034	10.44%
5. 技術工学	4,615,929,967	9.51%
6. 産業	2,196,387,437	4.53%
7. 芸術	3,258,432,447	6.71%
8. 言語	888,800,128	1.83%
9. 文学	9,341,275,486	19.25%
n. 分類なし	2,225,954,208	4.59%
合計	48,539,925,351	100.00%

3. 分析には「多変量解析システムSeagull-Stat」を用いた。これは早狩進氏（元青森県環境保健センター、現NPO法人グリーンシティ常務理事）の開発によるマクロ・プログラムである（石川・他2011）。
4. 『現代日本語書き言葉均衡コーパス』（BCCWJ: Balanced Corpus of Contemporary Written Japanese）には書籍以外にも10種のデータを含む。書籍をひとまとめにしたものと他の10種のクラスター分析結果を参考として次に示す：

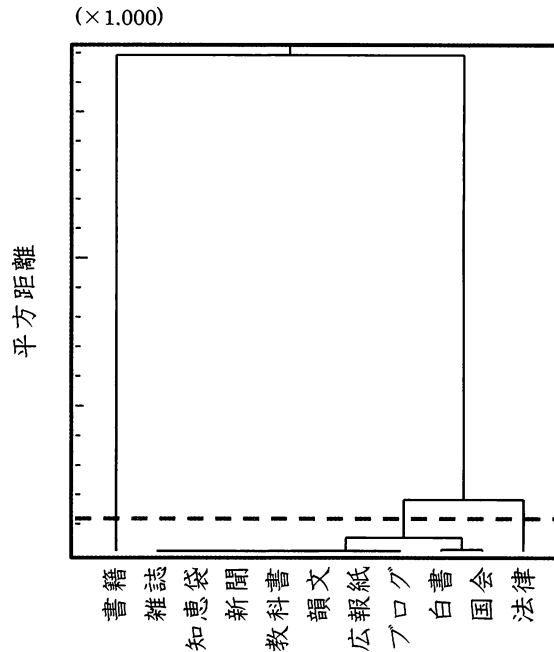


図3 樹形図（書籍と他のジャンルの分

図3が示すように、書籍は他の10種とは明らかに異なる1つの群を成しているため、本稿では書籍内の分野別頻度に注目することとした。

- インターネット上で利用できる「少納言」では語句を検索すると、その検索語の前後の文脈や出典なども同時に表示されるようになっている。香港うではこの利点を生かし、本文中に引いたデータ例は全てBCCWJのコーパスからの引用例を用い、「少納言」で示された出典を明記した。

参考文献

- 赤野一郎 (2006) 「英語コーパス言語学と英語教育」『日本語教育』130, 11-21.
- 石川慎一郎 (2008) 『英語コーパスと言語教育：データとしてのテキスト』大修館書店
- 石川慎一郎・前田忠彦・山崎誠編 (2011) 『言語研究のための統計入門』くろしお出版
- 内田諭・藤井聖子 (2009) 「『日本語コーパス』における語彙のジャンル別特徴：クラスター分析とフレームの観点から」『言語処理学会第15回年次大会発表論文集』, 442-445.
- 大曾美恵子 (2006) 「日本語コーパスと日本語教育」『日本語教育』130, 3-10.
- 金城克哉 (2011) 「コーパス分析に基づく『～にくい』・『～づらい』表現の研究」琉球大学留学生センター『留学生教育』8, 19-35.
- 黄慧 (2009) 「オノマトペの基本語彙に関する一考察—『現代日本語書き言葉均衡コーパス (2009モニター公開版)』を用いて」『東京外国語大学日本研究教育年報』15, 17-39.

- 齋藤俊雄・赤野一郎・中村純作 (2005) 『英語コーパス言語学—基礎と実践』(改訂新版)
研究社
- 田野村忠温 (2010) 「コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発」
『人口知能学会誌』24(5), 647-655.
- 鄭 惠先・小池 真理・船橋 瑞貴 (2009) 「『現代日本語書き言葉均衡コーパス』に見られる
「～てならない」「～てたまらない」「～てしかたない」「～てしようがない」の使い分け
—日本語学習者に対する指導への応用」『北海道大学留学生センター紀要』13, 4-21.
- 前川喜久雄 (2009) 「代表性を有する大規模日本語書き言葉コーパスの構築」『人工知能
学会誌』24(5), 616-622.
- 前川喜久雄・山崎誠 (2009) 「『現代日本語書き言葉均衡コーパス』」『国文学：解釈と
鑑賞』74(1), 6-14.
- 松田謙次郎 (編) (2008) 『国会会議録を使った日本語研究』ひつじ書房
- 宮内 佐夜香 (2010) 「通時的变化を背景とした接続助詞ガとケレド類の機能についての調
査—『現代日本語書き言葉均衡コーパス』を資料として—」『都大論究』47, 1-15.

参考ウェブサイト

- 「少納言」 <http://www.kotonoha.gr.jp/shonagon/>
「現代日本語書き言葉均衡コーパスのためのサンプル抽出」
http://www.ninjal.ac.jp/kotonoha/ex_3.html
「KHCoder」 <http://khc.sourceforge.net/index.html>

『現代日本語書き言葉均衡コーパス』からのデータ出典

- LS憲法研究会 (編) (2004) 『プロセス演習憲法』信山社出版
- 大村敦志 (2004) 『基本民法』有斐閣
- 加藤昌春・松野雄一郎 (1996) 『株式公開の知識』日本経済新聞社
- 東京都社会福祉協議会 (2001) 『社会福祉施設のための労働基準法Q&A』
- 二宮周平, 榊原富士子 (2005) 『離婚判例ガイド』有斐閣

Abstract

A study of the two allomorphic verbs *X-jiru* and *X-zuru* based on BCCWJ:

Katsuya Kinjo

This paper examines two allomorphic verbs *X-jiru* and *X-zuru* based on Balanced Corpus of Contemporary Written Japanese (BCCWJ). These two forms have been treated as if they appear in any contexts interchangeably; however, close examination by applying the cluster analysis to the frequency of them in BCCWJ, it will be shown that their appearance is rather biased: *X-zuru* characterizes the genres of social science and philosophy. On the other hand, the examination of verbal collocations shows that among the three verbs *shoo-zuru* (to occur), *jun-zuru* (to follow, to be based on), *mei-zuru* (to order), *shoo-zuru* is most often used as an intransitive verb and it is preceded by the word *mondai* (problem). Moreover, collocation reveals that in more than 50% of the usage of *jun-zuru*, it is preceded by pronouns such as *kore* (this) or *korera* (these). Also it will be discussed that the use of *X-zuru* forms is one of the reasons that the style found in these two genres makes the readers archaic/ authoritarian/ dogmatic. The research results seem to indicate that this stylistic characteristic is avoided in the natural science genre where *X-zuru* forms rarely occur.