



新聞における外来語ジャンル別出現状況の調査：朝日新聞「聞蔵IIビジュアル」を用いた計量的調査

鄧, 琪

(Citation)

統計数理研究所共同研究レポート, 414:145-156

(Issue Date)

2019-03

(Resource Type)

journal article

(Version)

Accepted Manuscript

(URL)

<https://hdl.handle.net/20.500.14094/90007760>



新聞における外来語ジャンル別出現状況の調査

—朝日新聞「聞蔵 II ビジュアル」を用いた計量的調査—

鄧 琪(神戸大学 大学院生)

dengqixq123@gmail.com

A survey on appearance of katakana loan-words by genre in Japanese newspaper

—Quantitative survey using the Asahi Newspaper—

DENG Qi (Kobe University, Graduate Student)

概要

本研究は、朝日新聞「聞蔵 II ビジュアル」を用い、各内容ジャンルにおける外来語使用を調査する。具体的には、(1)どのぐらい外来語を使っているかという延べ語数の観点、(2)どのぐらい幅広く外来語を使っているかという異なり語数の観点、(3)どのような外来語を多く使っているかという上位語の観点、という3つの観点に絞って調査を行った。その結果、(1)各ジャンルの外来語の頻度にはかなり大きな違いが存在し、スポーツ、経済、科学ジャンルなどで外来語が多く使用されていることと、同じく延べ語数の計量を行った先行研究のジャンル順位とほぼ一致していること、(2)延べ語数の場合とは大きく異なり、文化・芸能、生活、経済などのジャンルで外来語が幅広く使用されていること、また、延べ語数と異なり語数の2つの観点を同時に評価した場合、生活ジャンルにおいて外来語が最も多くかつ幅広く使用されていること、(3)一定の特徴度基準を満たす語について言えば、スポーツ、生活ジャンルが多く、オピニオン、政治が少ないこと、また、いわゆるジャンル特徴語の中には、ジャンルに関連した専門用語だけでなく、一般用語も混在していることが明らかになった。

キーワード

外来語, 新聞, 内容ジャンル

1. はじめに

下記は新聞のスポーツジャンルの記事の1例である。

18年現在、小学生の軟式チームは全国に約1万1500、競技者は約20万人いる。地方には少人数のチームもあるが、宗像豊巳専務理事は「将来の野球界を担う小学生がすでにひじ、肩を壊している。小学生でけがをした選手の一定数が後に再発するというデータもあり、待ったなしの時期

に來ている」と、ルールの必要性について説明した。

この日の評議員会では、球数制限導入のほかに、「1日の練習は3時間以内」「1年間での試合数は100試合以内」などとするガイドラインも、各チームに周知されることになった。宗像専務理事は「硬式よりも球が軽く軟らかい軟式は、安全神話も根強かった。勝利主義から育成主義へと転換し、大切な子どもたちが健康な体で野球が楽しめるようにしていきたい」と話した。（朝日新聞 2019 年 2 月 14 日）

上記の例に明らかなように、日本の新聞には多くの外来語が使用されている。しかしながら、ここで注意すべきことが 2 つある。1 つ目は、このように外来語が多く使用されるのは、スポーツといった特定の内容ジャンルに限られるのかどうかということである。多くの新聞は様々な内容ジャンルによって紙面が構成されているが、スポーツ以外の他のジャンルにおいても外来語が多く使用されているかどうかは必ずしも明らかでない。

2 つ目は、外来語が多いといっても、延べ語数が多いのか、あるいは異なり語数が多いのかが明確でないということである。たとえば上記の例について言うと、外来語の延べ語数は 6 となるが、実際には「チーム」が 3 回反復使用されているため、異なり語数は 4 となる。新聞に外来語が多いとしても、延べ語数と異なり語数の関係がどうなっているかは必ずしも明らかでない。そこで、本研究はこの点を計量的に明らかにしようとする。

2. 先行研究

本研究と関係する先行研究は多いが、以下、新聞媒体の各内容ジャンルにおける外来語使用という観点に関わる先行研究を概観する。

田中(2006)は公共的な媒体における外来語の実態を把握するため、新聞、省庁発行の白書、地方自治体発行の広報紙という 3 つの公共媒体をデータとして調査を行った。具体的調査対象として、新聞は 2003 年発行の『毎日新聞』朝夕刊を、白書は 2003 年発行の国会や閣議に提出される白書(電子媒体と紙媒体との双方が公開されているもの全 35 冊)を、広報紙は 2003 年発行のものから無作為に 61 号を、取り上げた。その結果、(1)国立国語研究所(1994)年の雑誌に対する調査の結果と比べると、公共性の高い 3 つの媒体における外来語の比率は低いことがわかった。(2)また、新聞の各面種の外来語比率を見てみると、スポーツ(5.3) > 芸能(4.4) > 経済(4.2) = 科学(4.2) > 家庭(3.7) > 社会(2.5) > 文化(2.4) > 国際(2.0)という順になった。(3)白書の各内容ジャンルの外来語比率について、情報(6.9%) > 外交(2.9%) > 国土(2.8%) = 全体(2.8%) = 経済(2.8%) > 教育(2.6%) > 福祉(2.3%) > 環境(2.2%) > 安全(1.6%) > 農林(1.3%)という順になった。(4)新聞と白書において外来語比率の高い分野が目立ち、他の分野ではあまり多く使われない外来語については、それぞれなくてはならない概念を表しており、それが報道や報告に取り上げられた結果、公共性の高い新聞や白書にも顔を出すことになる述べた。

桐生(2007)は、新聞の各面種の外来語使用実態の特徴を明らかにするため、『CD—毎日新聞』を用いて分析を行った。具体的には、『CD—毎日新聞』(日外アソシエーツ)1994 年～2003

年の 10 年分に収録される記事データのうち、大阪版の記事を除いた東京版の記事全文である。総語数は 452, 307, 883 語である。特徴語の抽出にあたっては、特化係数を用いる(面種 j における語 i の特化係数 = 面種 j における語 i の割合 / 新聞全体における語 i の割合)。その結果、(1) 各面種に含まれる外来語の分量について、スポーツ > 芸能 > 国際 > 総合 > 科学 > 経済 > 家庭 > 特集 > 読書 > 文化 > 社会 > 一面 > 三面 > 解説 > 社説 > 二面という順になること、(2) 各面種の高頻度外来語に占める特徴語の割合について、スポーツ・芸能・科学・家庭・経済などは特徴語の割合が 50% を超え、特徴的専門的な面種である一方、社説・解説・一面・特集・総合・二面・三面などは特徴語の割合が 20% に満たなく、平均的一般的な面種であること、(3) 各語の特徴語となる面種について、「スポーツ」の特徴語の大多数が「面種専用の語」である一方、「三面」「社説」「一面」などの特徴語は複数の面種にまたがる特徴語が多いこと、がわかった。

先行研究はそれぞれ重要な知見を示したが、(1) 語数を計量した研究は延べ語数を中心で、異なり語数を調べたものが少ない、(2) 語数を調べた研究と特徴語を調べた研究に 2 分されており、両者を同時に調べた研究は多くない、という制約を残っている。そこで本研究は、(1') 延べ語数と異なり語数をそれぞれ調べて、外来語の語数を計量し、(2') 合わせて特徴語の分析を組み合わせたい。

3. リサーチデザインと手法

3.1 研究目的と RQ

新聞記事を取り上げ、外来語が多いという見解はしばしば示されるが、新聞に含まれる各ジャンルにおいてどの程度外来語が使用されているかは必ずしも明らかでない。ゆえに、本研究は新聞の各内容ジャンルにおける外来語使用状況を調査することで、ジャンルによる外来語使用状況の違いを明らかにすることを目指す。なお、本研究は、(1) どのぐらい外来語を使っているかという延べ語数の観点、(2) どのぐらい幅広く外来語を使っているかという異なり語数の観点、(3) どのような外来語を多く使っているかという上位語の観点、という 3 つの観点で調査を行う。

上記の目的に従い、以下のような 3 つのリサーチクエスチョンを設定した。

RQ1 延べ語数に注目した場合、外来語の出現数はジャンルによってどう異なるか？

RQ2 異なり語数に注目した場合、外来語の出現数はジャンルによってどう異なるか？

RQ3 上位語に注目した場合、多く使用されている外来語はジャンルによってどう異なるか？

3.2 データ

本研究は、朝日新聞社の「聞蔵 II ビジュアル」を用いる。「聞蔵 II ビジュアル」は 1879 年の創刊号から今日までの記事・広告が検索できる国内最大級のオンライン新聞記事データベースである。このデータベースでは、検索画面から詳細検索を選ぶことで、新聞の 17 種のジャンルを指定して調査を行うことができる。今回は 16 種の区分の中から、新聞記事としては特殊であると思われる家庭、歌壇・俳壇、Be、グローブ、マリオンの 5 種、および、内容ジャンルが統一していない総合

面を除き、他の 11 種を分析の対象とした。調査対象年代は 2017 年 3 月から 2018 年 10 月までの 20 ヶ月間である。サンプリング方法として、内容ジャンルごとに一ヶ月に一本の記事をランダムに取得し、全体で 20 本*10 種類*10 ヶ月=2000 本のデータを取った。総語数は 71790 語となった。

3.3 手法

RQ1(延べ語数)については、KH Coder を使い、各内容ジャンルに出現するすべての外来語を抽出する。なお、比較する 11 種のジャンルの総語数が異なることから、10 万語あたりの調整頻度で比較を行う。

RQ2(異なり語数)については、各内容ジャンルにおける外来語の種類を数える。

RQ3(特徴語)については、それぞれの外来語について、特定ジャンル内の頻度とその他全ジャンルにおける頻度を加味し、対数尤度比統計量を計算する。そして、対数尤度比が 10 以上になるものを分野特徴語として抽出する。

なお、KH Coder で語彙分析を行うと、内蔵されている辞書の基準と照合して外来語を抽出することができるが、辞書にないものとしては未知語に分類される。今回の研究では、未知語に分類されたカタカナ文字列については、可能な範囲で筆者が目視で確認し、外来語と認定された場合は分析に加える方針を取った。

4. 結果と考察

4.1 RQ1 延べ語数

各内容ジャンルに出現する外来語の延べ語数を調査したところ、以下のような結果が得られた。なお、先行研究の結果と比較できるよう、田中(2006)と桐生(2007)の結果を合わせて示す。

表 1 各ジャンルにおける外来語の延べ語数

ジャンル	延べ語数	順位	田中(2006)	桐生(2007)
スポーツ	7022	1	1	1
生活	5923	2		
経済	5688	3	3	6
科学・医療	5671	4	3(科学)	5(科学)
文化・芸能	5365	5	2(芸能) 7(文化)	2(芸能) 10(文化)
教育	4520	6		
社会	4154	7	6	11
国際	3364	8	8	3
読書	2897	9		9
オピニオン	2809	10		
政治	1873	11		

表 1 からわかることは以下の 3 点である。1 点目は、外来語の数が最も多いスポーツと最も少ない政治の間には 3.7 倍の差があり、ジャンルによる外来語の数の違いは予想以上に大きいということである。このことは、内容ジャンルを考慮せず、一概に、「新聞には外来語が多い」といった判断を行うことが必ずしも適切ではないことを示している。2 点目は、2 つの先行研究の内、ジャンル順位は田中(2006)とはほぼ一致しているが、桐生(2007)とはかなり異なっていたということである。この点については、田中(2006)が本研究と同じ延べ語数で外来語の数を調査したのに対し、桐生(2007)では文字種に基づく、異なる計量法を採用していたためと考えられる。3 点目は、そうした違いがあるにも関わらず、外来語がスポーツ面で最も多用され、経済や科学がそれに次ぐことはすべての研究で一致しているということである。つまり、新聞における外来語はすべてのジャンルにおいて均衡的に出現するのではなく、スポーツをはじめとする特定のジャンルに集中的に出現するものであると言えるだろう。

4.2 RQ2 異なり語数

RQ1 では、各内容ジャンルにおける外来語の延べ語数の順位は先行研究の順位とほとんど一致していることを確認した。しかしながら、ここで注目すべきは、異なり語数に注目した研究がほとんどなかったということである。そこで、異なり語数に注目したところ、以下の結果を得た。なお、先述べたように、先行研究の結果と比較できるよう、田中(2006)と桐生(2007)との結果を合わせて示す。

表 2 各ジャンルにおける外来語の異なり語数

ジャンル	異なり語数	異なり語数 順位	延べ語数 順位	田中(2006)	桐生(2007)
文化・芸能	187	1	5	2(芸能) 7(文化)	2(芸能) 10(文化)
生活	186	2	2		
経済	179	3	3	3	6
スポーツ	177	4	1	1	1
教育	169	5	6		
社会	160	6	7	6	11
科学・医療	146	7	4	3(科学)	5(科学)
国際	138	8	8	8	3
読書	119	9	9		9
オピニオン	80	10	10		
政治	79	11	11		

ここで注目すべきことは、ここまで述べてきた研究と異なる結果が示されたということである。すでに述べたように、延べ語数を単位とすれば、スポーツ、経済、科学などにおいて外来語が多用さ

れると言えるが、異なり語数に注目すると、文化・芸能、生活、経済などにおいて外来語が広く使用されていると言える。

では、延べ語数に基づく順位と異なり語数に基づく順位を同時に観察すると、どのようなものであろうか。以上の 2 つの観点を整理すると、以下のような図が得られる。

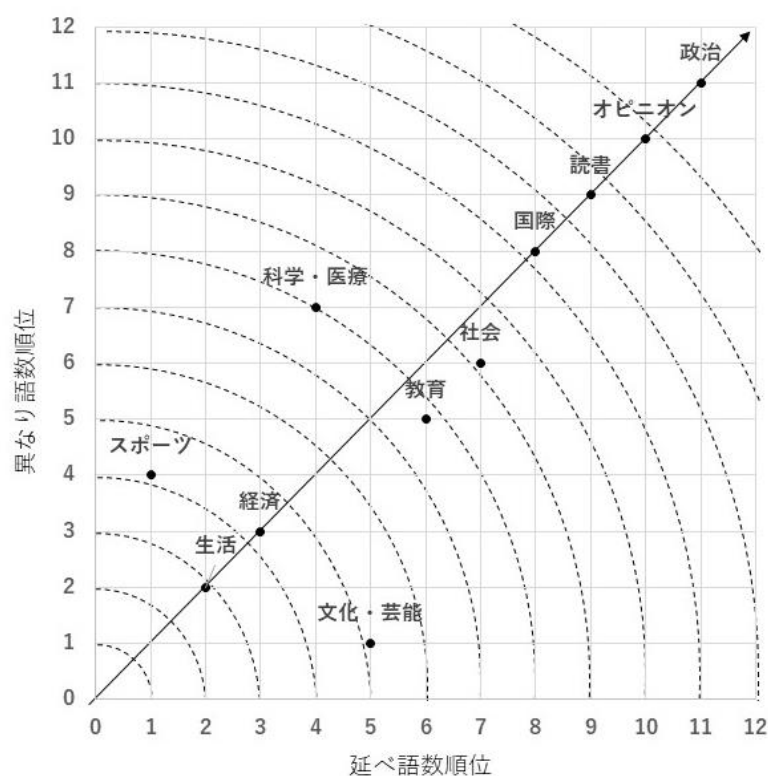


図 1 延べ語数順位と異なり語数順位に基づく 11 ジャンルの位置づけ

図 1 で明らかなように、教育、社会、国際、読書、オピニオン、政治という相対的に外来語頻度が低い 6 ジャンルについては、延べ語数順位と異なり語数の順位がほぼ一致している。一方、相対的に外来語出現頻度の高い 5 ジャンルについては、延べ語数と異なり語数で食い違いが見られる。仮に、いずれの順位においても総合的に上位を占めるもの、つまり、上記の図において原点に最も近いものを、特に外来語が顕著に使用されるジャンルであると言うならば、生活ジャンルがそれに合致すると言えるだろう。

4.3 RQ3 上位語

以上、RQ1 と RQ2 を通して、外来語の出現状況はジャンルによって大きな違いがあることを確認した。では、各ジャンルにおいてそれぞれ特徴づける外来語にはどのようなものがあるのだろうか。そこで、対数尤度比が 10 以上を基準として特徴語を抽出した結果、以下のような表が得られた。

表 3 ジャンルごとの外来語特徴語

ジャンル	外来語	内頻度	全頻度	対数 尤度比	外来語	内頻度	全頻度	対数 尤度比
経済 (21 語)	ローン	160	0	51.69	ブレーキ	88	6	15.53
	カレー	131	0	42.29	データ	204	54	14.17
	スプレー	131	0	42.29	カレーライス	44	0	14.09
	バー	131	3	32.26	デフレ	44	0	14.09
	カード	160	9	30.81	ヘリコプター	44	0	14.09
	アルミ	88	0	28.19	ストア	44	0	14.09
	コンビニ	88	0	28.19	グループ	160	37	12.92
	スーパー	146	9	27.02	カメラ	117	20	12.28
	メーカー	117	5	25.30	サービス	117	20	12.28
	メニュー	58	0	18.79	トップ	131	28	11.52
	カット	73	2	18.28				
政治 (3 語)	カジノ	194	3	52.75				
	サポーター	69	2	17.79				
	オール	56	2	13.60				
社会 (11 語)	リムジン	150	0	50.25	イメージ	27	19	18.98
	サッカー	164	9	33.19	カメラ	137	17	18.98
	レンタル	96	0	31.97	テープ	55	0	18.27
	バイク	82	0	27.40	ブラジャー	55	0	18.27
	ブロック	82	0	27.40	サラダ	41	0	13.70
	デモ	109	8	20.29				
国際 (8 語)	メディア	193	17	30.67	マクロ	45	0	14.21
	ニュース	134	6	27.39	スピーチ	45	0	14.21
	インフラ	74	2	18.48	ミサイル	60	2	14.15
	インフレ	74	2	18.48	メッセージ	104	12	14.01
文化・芸能 (23 語)	アート	239	3	55.63	ショー	55	2	11.14
	コレクション	129	0	36.12	ディレクター	55	2	11.14
	ブランド	165	6	31.02	ドア	55	2	11.14
	コンペ	92	0	25.80	アングル	37	0	10.32
	デザイナー	92	0	25.80	インスタレーション	37	0	10.32
	ファッション	73	0	20.64	スタジオ	37	0	10.32
	ミュージシャン	73	0	20.64	ソナタ	37	0	10.32
	アトリエ	55	0	15.48	バンド	37	0	10.32
	モヘア	55	0	15.48	ビエンナーレ	37	0	10.32

	レジ	55	0	15.48	フェスティバル	37	0	10.32
	フェイクファー	55	0	15.48	ブラック	37	0	10.32
	テーマ	110	9	15.27				
科学・医療 (19 語)	ワクチン	412	0	144.90	スパコン	100	5	22.92
	ウイルス	362	0	127.33	パワー	87	6	17.24
	センサー	387	5	116.53	マウス	62	2	16.77
	ゲノム	225	0	79.01	トラブル	62	3	14.04
	インフルエンザ	150	0	52.66	パー	37	0	13.16
	リスク	175	3	49.86	レーダー	50	2	12.78
	データ	312	38	47.50	バイオ	75	8	12.35
	リチウム	100	0	35.11	マス	75	8	12.35
	ラット	87	0	30.72	プロジェクト	50	3	10.39
	ポスト	75	0	26.33				
教育 (13 語)	インターン	247	0	87.33	アンケート	111	13	17.69
	アプリ	173	2	54.01	ジャーナリスト	49	0	17.46
	スタッフ	185	9	41.79	リスト	49	0	17.46
	クラス	111	2	33.02	シート	62	2	16.66
	トランスジェンダー	86	0	30.56	ノート	111	17	14.39
	インターンシップ	86	2	24.77	ワン	37	0	13.09
	アルバイト	86	2	24.77				
スポーツ (39 語)	チーム	414	42	48.56	トレーニング	59	1	11.55
	ジャンプ	178	0	47.72	リレー	59	1	11.55
	リーグ	197	4	39.42	シーズン	79	6	10.70
	メダル	158	1	36.29	ボール	79	6	10.70
	アイスホッケー	118	0	31.81	スクラップブック	39	0	10.60
	ファン	118	0	31.81	スコア	39	0	10.60
	スポーツ	316	42	31.27	セカンド	39	0	10.60
	プロ	158	4	29.97	デマンド	39	0	10.60
	バドミントン	79	0	21.21	ノルディックスキー	39	0	10.60
	マラソン	79	0	21.21	ハーフ	39	0	10.60
	シュート	79	0	21.21	ブラインド	39	0	10.60
	ゴール	99	4	16.36	ブルペン	39	0	10.60
	プレー	99	4	16.36	マインド	39	0	10.60
	コーチ	79	1	16.35	ユニバーシアード	39	0	10.60
	ドーピング	79	1	16.35	レース	39	0	10.60
	エース	59	0	15.90	カーブ	39	0	10.60

	ナショナル	59	0	15.90	シンプル	39	0	10.60
	ゲーム	99	7	13.38	タイムリー	39	0	10.60
	ハンド	59	1	11.55	アウェー	39	0	10.60
	ボルト	59	1	11.55				
生活 (36 語)	リハビリ	139	0	40.34	マグネット	52	0	15.12
	ケア	156	2	39.05	セミナー	69	3	12.86
	シングル	104	0	30.25	ポスター	69	3	12.86
	マザー	87	0	25.21	カップル	52	2	10.79
	マップ	87	0	25.21	ハードル	52	2	10.79
	ロボット	121	5	23.58	ウェット	35	0	10.08
	スーツ	104	3	21.59	ガイドライン	35	0	10.08
	デイサービス	69	0	20.17	ゲイ	35	0	10.08
	フローチャート	69	0	20.17	ゲスト	35	0	10.08
	パンフレット	87	2	19.97	パート	35	0	10.08
	ライフ	87	2	19.97	ボンベ	35	0	10.08
	デザイン	87	2	19.97	ママ	35	0	10.08
	ボランティア	87	3	17.17	マンション	35	0	10.08
	トーク	69	2	15.33	ローリング	35	0	10.08
	ベッド	69	2	15.33	ワークショップ	35	0	10.08
	ホームページ	69	2	15.33	アワード	35	0	10.08
	キット	52	0	15.12	ダイエット	35	0	10.08
	グッズ	52	0	15.12	メール	104	17	10.02
オピニオン (7 語)	バス	103	0	26.89	ホモ	62	0	16.14
	ライン	144	4	25.86	シナリオ	41	0	10.76
	タイム	144	7	22.06	トラフ	41	0	10.76
	スポーツ	246	48	17.49				
読書 (10 語)	ノンフィクション	109	0	33.81	ポケット	62	2	14.50
	シリーズ	78	0	24.15	カクテル	47	0	14.49
	ポピュリズム	78	0	24.15	タレント	47	0	14.49
	プレゼント	62	0	19.32	ペン	47	0	14.49
	モビリティ	62	0	19.32	リトル	47	0	14.49

ここで注目すべきことは 2 つがある。1 つ目は、分野によって基準を満たす特徴語の数が大きく異なることである。具体的には、多い順にスポーツ(39 語)＞生活(36 語)＞文化・芸能(23 語)＞経済(22 語)＞科学・医療(19 語)＞教育(13 語)＞社会(11 語)＞読書(10 語)＞国際(8 語)＞オピニオン(7 語)＞政治(3 語)となる。このことは、特定の外来語を特徴的に使用するジャンルと

そうでないジャンルが新聞の中に混在していることを示す。また、特徴語の多い順は RQ2 で得られた総合順位とほとんど一致していることがわかった。たとえば、これまでの調査で、延べ語数順位と異なり語数順位のいずれにおいても上位となった生活について言えば、特徴語の数も多いことを示す。また、総合順位最も低かった政治ジャンルについて言えば、特徴語も最も少ない。

2 つ目は、ジャンル特徴語の中には、いわゆる専門用語と思われるものと必ずしも分野との関係が明確でないものが含まれているということである。たとえば、経済ジャンルについて言えば、「ローン」や「カード」のような経済や金融に直接関係した語を存在したと同時に、「カレー」、「スーパー」、「サービス」のような経済と直接関係していなく、どの分野においても広く使用される語も含まれている。また、教育ジャンルにも、「クラス」や「ノート」のような教育ジャンルと直結している語が含まれていると同時に、「スタッフ」、「アプリ」、「アルバイト」などのような教育ジャンルとの関係性が希薄の語も多く含まれている。今回のデータから、テーマと直接に関係した語よりもむしろテーマとの関係がより拡張的な語がそれぞれのジャンルに特徴づけられていることがわかった。

5. まとめ

本研究は朝日新聞「聞蔵 II ビジュアル」を用い、日本語外来語の新聞におけるジャンル別出現状況を調査した。RQ1(延べ語数)では、各ジャンルの外来語の頻度にはかなり大きな違いが存在し、ジャンルなどで外来語が多く使用されていることと、同じく延べ語数の計量を行った先行研究のジャンル順位とはほぼ一致していることが確認された。RQ2(異なり語数)では、延べ語数の場合とは大きく異なり、文化・芸能、生活、経済などのジャンルで外来語が幅広く使用されていることが明らかになった。また、延べ語数と異なり語数の 2 つの観点を同時に評価した場合、生活ジャンルにおいて外来語が最も多くかつ幅広く使用されていることが示された。最後に、RQ3 では、スポーツ、生活ジャンルが多く、オピニオン、政治が少ないことが明らかになった。また、いわゆるジャンル特徴語の中には、ジャンルに関連した専門用語だけでなく、一般用語も混在していることが明らかになった。

今回の研究は日本語の新聞における外来語の出現状況について、多くの知見をもたらしたと言える。ただし、制約も残される。1 つ目は、調査に使用したのは朝日新聞データベースのみであったということである。これについては、他の新聞データベースを合わせて調査することで、今回得られた結果が再現性を持つかどうかを確認する必要がある。2 つ目は、外来語の出現状況の調査にとどまり、それぞれの外来語が具体的な文脈でどう使用されているかを調査していないという点である。この点については、特に高頻度の外来語を取りあげ、具体的な文脈の中でそれらがどのような意味機能を果たしているかを質的に研究していく必要があるだろう。これらの点は今後の研究で対応したい。

引用文献

桐生りか(2007)「新聞の面種と外来語」『公共媒体の外来語:「外来語」言い換え提案を支える調査研究』国立国語研究所。

田中牧郎(2006)「現代社会における外来語の実態」『新「ことば」シリーズ 19 外来語と現代社会』
国立国語研究所.

