『日本語話し言葉コーパス（CSJ）』の異なる講演タイプにおける外来語の質的分析 —言語外的および言語内的指標を用いた外来語分類の試み—

久屋 愛実（オックスフォード大学）†

A Qualitative Analysis of Loanwords in Different Speech Styles in the Corpus of Spontaneous Japanese (CSJ): Classifying Loanwords Based on Extra-/Intra-Linguistic Factors

Aimi Kuya (Faculty of Linguistics, Philology and Phonetics, University of Oxford)

要旨

本稿では、レジスター横断性やジャンル横断性に留意して『日本語話し言葉コーパス(CSJ)』から「基本度」（水谷 1964）の高い外来語を抽出し、それらの語彙的特徴を記述する。分析の結果、レジスター横断的かつジャンル横断的である最も基本度の高い語群は、それ以外の語群よりも抽象的あるいは多義的な意味を表す語の割合が高く、普通名詞（一般）以外の品詞の割合が高い傾向にあった。

１．はじめに

コーパスを使った語彙研究においては、語の「基本度」（水谷 1964）を頻度により捉えるのが最も一般的であろう。通時的コーパスを使う場合は、頻度の経年的増減を追うことによって基本語化した語彙を取り出すことが可能である（金 2011、田中 2014）。しかし、共時的コーパスを扱う場合は頻度の経年的増減が捉えられないため、広範囲に分布するかどうかを示す「散らばり度」（水谷 1964）が語の基本度をはかる指標として有効である。本稿は、共時的コーパスである『日本語話し言葉コーパス（以下、CSJ）』に出現する外来語を、異なるレジスターやジャンルにまたがって分布する語かという観点から分類し、特定のレジスターやジャンルに左右されない、「いわば無性格な語群」（田中 1973）を抽出する。こうした「無性格語」は、他と比べてより基本的な語彙であると考えられるが、これらがどのような語彙的特徴をもつのかについても考察する。

２．語の散らばり度に基づいた「無性格語」の抽出

本稿では、CSJ[1]の学会講演と模擬講演部分から抽出した外来語の分析を行う。水谷（1964: 10）が指摘するように、例えば雑誌における語の散らばり度は、「あるいは一編ずつの記事、あるいは雑誌の一冊ずつ、あるいは小説・随筆・論説のような記事分類の別」によって求められる。これに倣えば、CSJ における語の散らばり度は、文章別、講演別、学会種や講演テーマ別（ジャンル別）、講演のタイプ別（レジスター別）、あるいは講演者別など、あらゆる単位からはかることが可能である。本稿では、このうち講演タイプの別（レジスター）と学会種・講演テーマの別（ジャンル）の 2 指標を用いる。

† aimi.kuya@ling-phil.ox.ac.uk

[1] CSJ の概要については国立国語研究所（2006）を参照されたい。

２．１　レジスター横断性

表 1 は、CSJ の学会講演（Academic Presentation Speech、以下 A）と模擬講演（Simulated Public Speaking、以下 S）における異なり語数・延べ語数とその比率を語種ごとに示したものである。外来語のみに関して言えば、その割合は異なり語数・延べ語数ともに模擬講演より学会講演で高い。また、外来語の異なり語数は学会講演（3555 語）よりも模擬講演（4229 語）のほうが多いものの、延べ語数でみると学会講演（100428 語）が模擬講演（67863 語）の 1.5 倍にもなり、学会講演では外来語の一語あたりの平均出現度数が高いことがわかる。

表 1：CSJ 学会講演と模擬講演における語種別の頻度と比率

		外	漢	和	混	固	記号	その他 (空白・不明等)	総計
異なり	学会講演（A）	3555	8773	5222	507	2564	759	2283	23663
		15.0%	37.1%	22.1%	2.1%	10.8%	3.2%	9.6%	100.0%
	模擬講演（S）	4229	11660	9386	1004	4407	226	2464	33376
		12.7%	34.9%	28.1%	3.0%	13.2%	0.7%	7.4%	100.0%
延べ	学会講演（A）	100428	691117	1087123	22938	20818	9990	54504	1986918
		5.1%	34.8%	54.7%	1.2%	1.0%	0.5%	2.7%	100.0%
	模擬講演（S）	67863	470064	1348909	27969	41500	1863	47014	2005182
		3.4%	23.4%	67.3%	1.4%	2.1%	0.1%	2.3%	100.0%

UniDic 短単位による[2]。品詞が「空白、記号、助詞、助動詞」となるものは含まない。

表 2：レジスター横断性

	学会講演（3555 異なり語）			総計
		模擬講演（4229 異なり語）		
	特徴語 A	共通語	特徴語 S	
外来語の異なり語数	1735	1820	2409	5964
外来語の延べ語数	20740	133382	14169	168291
一語あたりの平均度数	12	73	6	28

こうした違いは、学会講演と模擬講演という異なるレジスターで出現する外来語が完全に同質ではないことに起因すると思われる。表 1 の外来語の中には両レジスターで重複して出現するものもあればそうでないものもあり、それぞれのふるまいが異なる可能性があるからである。そこで、表 1 で抽出した外来語を、学会講演（A）にのみ出現する「特徴語 A」、模擬講演（S）にのみ出現する「特徴語 S」、どちらにも共通で出現する「共通語」の 3 種に再分類してみる。散らばり度の観点からすると、共通語は 2 つの特徴語に比べて「レジスター横断性」が高い。分類の結果、表 2 に示す通り、学会・模擬講演を統合したときの外来語の異なり語数は 5964 語で、このうち特徴語 A の 1735 語、特徴語 S の 2409 語を除くと、共通語は 1820 語にまで減少する。つまり、5964 語のうち約 7 割がどちらかひと

[2]) UniDic 体系の CSJ 短単位データは、現在国立国語研究所が整備中である。今回は同研究所の許可を得て公開前のものを分析に利用したため、今後一般に公開されるデータを用いた分析とは結果が異なる可能性がある。（本データは 2014 年 11 月時点のもの）

つのレジスターにしか出現しない特徴語であることがわかる。さらに、一語あたりの平均出現度数は特徴語 A が 12 回で、特徴語 S（6 回）の 2 倍にもなる。一方、共通語の一語あたりの平均出現度数は両レジスター全体で 73 回であり、2 つの特徴語よりも圧倒的に高い。このことから見ても、この 3 つのカテゴリーは区分して論じたほうがよさそうである。

２．２　ジャンル横断性

次に、ジャンル横断的に分布する広範囲語かどうか、抽出した 5964 の外来語それぞれの「ジャンル横断性」をはかる。ここでは、学会講演における 13 の学会種、模擬講演における 12 の講演テーマをジャンル数とみなす。まず、表 3 のとおり、それぞれの外来語が講演タイプごとにいくつのジャンルに出現したかを求め、整理した。ジャンル横断性の序列は、表の色分けされた区分に従って行った。ジャンル横断性は、色なし部分が最も低く、薄い網掛け部分がその中間で、濃い網掛け部分が最も高い。

表 3：出現ジャンル数別にみた外来語 5964 語

			模擬講演				
			共通語			特徴語 A	
		出現ジャンル数	1~4 テーマ	5~8 テーマ	9~12 テーマ	なし	総計
学会講演	共通語	1~4 学会	893	306	147	1655	3001
		5~9 学会	154	105	142	79	480
		10~13 学会	7	15	51	1	74
	特徴語 S	なし	2175	206	28		2409
		総計	3229	632	368	1735	5964

散らばり度：低い（色なし）　中間　高い

表 4：ジャンル横断性と特徴語・共通語の別

ジャンル横断性	特徴語 A	共通語	特徴語 S	総計
高い	1	208	28	237
中間	79	719	206	1004
低い	1655	893	2175	4723
総計	1735	1820	2409	5964

表 4 は表 3 を色別にまとめ、先にみた特徴語・共通語の別を加えて分類しなおしたものである。その結果、ジャンル横断性が高い 237 語、中間レベルの 1004 語、ジャンル横断性が低い 4723 語に分かれた。このうち、ジャンル横断性が高い 237 語を「ジャンル横断性の高い語」または「ジャンル広範囲語」と定め、さらなる分析に利用する。ジャンル広範囲語は、特徴語 A（1 語）、共通語（208 語）、特徴語 S（28 語）の 3 つにさらに分けられる。

以下にこれら全ての語彙を示す（五十音順）。特徴語 A（1 語：「コンテキスト」）と特徴語 S（28 語：「エアロビック」～「ロープ」）は、個々のレジスターにおいてはジャンル横断性が高いが、レジスター横断的な語彙ではないため、あくまでもそれぞれのレジスターに限り広く分布している「キー・ワード」[3]（田中 1973）でしかない。これらを除いた

[3] 田中（1973）によれば、ある文章の頻度調査において頻度順位の比較的上位に来る語彙のうち、特定の文章や文献の性格に関わらず現れうる「無性格語」を排除すると「キー・ワードすなわち、『いかにも、その文章らしい単語』」が残るとする。

残りの共通語（208 語：「アイディア」～「ワールド」）が、ジャンル横断性だけでなくレジスター横断性も高いことから、特定のレジスターやジャンルに左右されない、本コーパスの「無性格語」と見ることができる。

ジャンル広範囲語全 237 語

特徴語 A：（1 語）
コンテキスト

共通語：（208 語）（＝無性格語）
アイディア、アウト、アクセス、アクセント、アップ、アドバイス、アナウンサー、アプローチ、アルバイト、アンド、イコール、イベント、イメージ、イン、インターネット、インタビュー、ウイーク、ウインドー、エネルギー、エピソード、エレベーター、エンジン、オーケー、オーバー、オープン、オフ、オブ、オフィス、オレンジ、カー、カード、ガイド、カウント、カット、カバー、カメラ、カラー、ガラス、キー、ギャップ、キャラクター、キロ、クラシック、クラス、グラフ、クリア、グループ、ケース、ゲーム、コース、コーヒー、コピー、コミュニケーション、コメント、コントロール、コンピューター、ザ、サービス、サイクル、サイズ、サイン、サポート、サン、シート、シーン、システム、ジャンル、シンボル、スーパー、スクリーン、スケジュール、スター、スタート、スタイル、ストーリー、ストップ、ストレス、スピーチ、スピード、スペース、スポーツ、スムーズ、スリー、ゼロ、センス、センター、センチ、ソフト、ターゲット、タイトル、タイプ、タイミング、タイム、ダウン、ダブル、チーム、チェック、チャンス、チャンネル、ツー（< *two*）、ツー（< *to*）、データ、データーベース、テープ、テーブル、テーマ、テキスト、デザイン、デジタル、テスト、テレビ、ドア、トップ、トラック、トラブル、ドラマ、トレーニング、ナンバー、ニュー、ニュース、ネット、ネットワーク、ノー、ノート、パーセント、ハード、ハイ、バス、パソコン、パターン、バック、バラエティー、バランス、パンフレット、ピーク、ビジネス、ヒット、ビデオ、ピンク、ヒント、ファースト、ファイブ、ファミリー、プラス、プラン、フリー、フル、ブルー、プロ、プログラム、プロジェクト、プロセス、ブロック、ペア、ページ、ベース、ペース、ペーパー、ベスト、ベッド、ポイント、ホーム、ボール、ボタン、ボックス、ボランティア、マーク、マイク、マイナス、マシン、マスコミ、マナー、マニュアル、ミス、ミリ、メーター、メートル、メール、メーン、メッセージ、メニュー、メモ、メリット、メンバー、モデル、モニター、ユニーク、ライフ、ライブ、ライン、ラジオ、ラベル、ランク、リアル、リーダー、リード、リスト、リズム、リラックス、ルーム、ルール、レコード、レストラン、レベル、ワーク、ワード、ワープロ、ワールド

特徴語 S：（28 語）
エアロビック、オーナー、クーラー、グッズ、ゴールデン、シャワー、ジャングル、ジョギング、スープ、スカート、スナック、ズボン、デザート、テント、バイク、バッグ、ハンバーグ、フルーツ、プロデューサー、マージャン、マラソン、ミネラル、メダル、リゾート、リフレッシュ、レース、レンタル、ロープ

３．無性格語の意味特性

ここでは、前節で抽出した無性格語の意味的特徴を調べるため、『分類語彙表―増補改訂版』（国立国語研究所 2004）の分類に従って意味分類を行う。手順は、各外来語に付与された UniDic の語彙素 ID を主キーとして分類語彙表から分類語彙表番号を割り出し[4]、その中の「部門」番号に基づいて 5 項目｛1 抽象的関係、2 人間活動の主体、3 人間活動-精神および行為、4 生産物および道具、5 自然物および自然現象｝に分類する、というものである。ただし、多義語の場合は、ひとつの語彙素 ID に対して複数の分類語彙表番号が割り当てられており（小木曽・中村 2011）、結果として異なる複数の「部門」番号を有することがある。そのような語彙素には、複数の意味分野を持つという意味で「多義」という 6 つ目の分類名を新たに付与した。最後に、分類語彙表において対応する語彙素 ID が見つけられない場合は、その語彙素が分類語彙表に収録されていないという意味で「未収録」とい

[4] 国立国語研究所コーパス開発センター「形態論情報データベース」（小木曽・中村 2014）上の辞書データと分類語彙表データを利用した。

う 7 つ目の分類名を付与した。なお、分類語彙表の採用語は、「現代の日常生活で普通に用いられる語を中心に、各種語彙調査の結果その他から選定」され、原版にあった語も含めて「見慣れない専門用語や古語・方言、また社会生活上使用を遠慮すべき語の類は除いている」（国立国語研究所 2004: 3）。よって、ここで「未収録」に区分された語彙は、あくまでも増補改訂版の作業時に上記条件に当てはまらないと判断されたものであり、当時から約 10 年経った現在の感覚とは異なる可能性がある。

表 5：ジャンル広範囲語（237 語）の意味分類

	1 抽象的関係	2 人間活動-主体	3 人間活動-精神・行為	4 生産物・道具	5 自然物・自然現象	多義	未収録	総計
特徴語 A							1	1
共通語（＝無性格語）	49	12	54	31	3	46	13	208
特徴語 S	1	3	5	12	4	2	1	28
総計	50	15	59	43	7	48	15	237
延べ語数（両レジスターの合計）	21290	2208	18446	8465	501	15717	5892	72519
一語あたりの平均度数	**426**	147	**313**	197	72	**327**	**393**	306

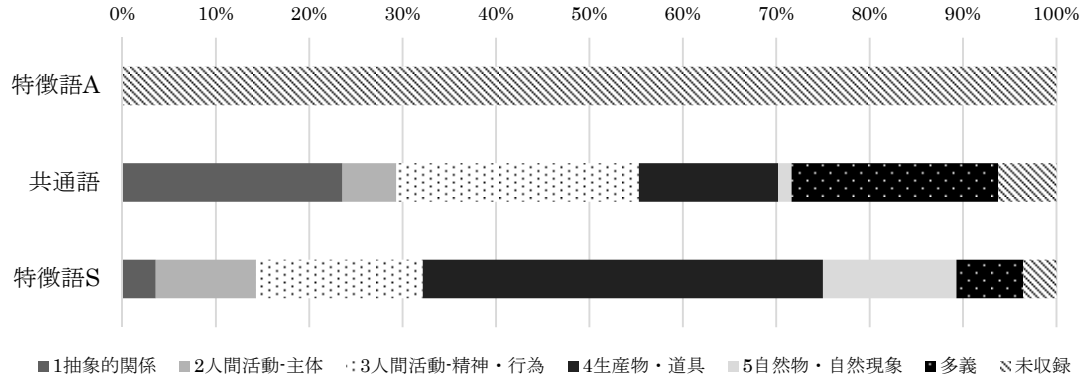

図 1：ジャンル広範囲語（237 語）の意味分類比率

表 5 は、無性格語を含むジャンル広範囲語 237 語の意味分類を示したものである。図 1 はそれを百分率に直したものである。特徴語 A は「コンテキスト」一語で、未収録語に分類されている。共通語（＝無性格語）と特徴語 S とを比較すると、特徴語 S よりも共通語で「1 抽象的関係」、「3 人間活動（精神・行為）」、「多義」の割合が高く、特に「1 抽象的関係」と「多義」は特徴語 S―共通語間の比率差が著しい。一方、「2 人間活動（主体）」、「4 生産物・道具」、「5 自然物・自然現象」の割合は特徴語 S よりも共通語で低く、特に「4 生産物・道具」と「5 自然物・自然現象」は特徴語 S―共通語間の比率差が著しい。なお、7 つの意味分類のうち、一語あたりの平均出現度数は「1 抽象的関係」、「未収録」、「多義」、「3 人間活動（精神・行為）」の順に高く、「1 抽象的関係」、「3 人間活動（精神・行為）」、「多義」の割合が高い共通語（208 語）には比較的高頻度の語彙が多く含まれていることがわかる。一方、一語あたりの平均出現度数が相対的に低いのは「5 自然物・

自然現象」、「2 人間活動（主体）」、「4 生産物・道具」であり、「4 生産物・道具」や「5 自然物・自然現象」の割合が高い特徴語 S（28 語）には、ジャンル広範囲語でありながら比較的低頻度の語彙が多く含まれていることがわかる。

４．無性格語の品詞特性

次に、無性格語の品詞的特徴を調べるため、無性格語を含むジャンル広範囲語 237 語を、UniDic の品詞分類に基づいて分類し、表 6 に示した。図 2 ではそれを百分率で示している。

表 6：ジャンル広範囲語（237 語）の品詞分類

	名-普-一般	名-普-サ変可能	名-普-サ変形状詞可能	名-普-形状詞可能	名-普-助数詞可能	名詞-数詞	形状詞-一般	総計
特徴語 A	1							1
共通語（＝無性格語）	138	46	3	10	7	1	3	208
特徴語 S	25	3						28
総計	164	49	3	10	7	1	3	237
延べ語数（両レジスターの合計）	51658	11374	335	1598	5282	2098	174	72519
一語あたりの平均度数	315	232	112	160	**755**	**2098**	58	306

*UniDic では品詞情報が語形 ID に紐づけされるため、語彙素 ID が複数の品詞情報を持つ場合がある。ここでは「オフ」と「ノート」が名-普-一般または名-普-サ変可能であった。今回は語彙素 ID でカウントするために、サ変用法が実際に確認できた前者を名-普-サ変可能、サ変用法が確認できなかった後者を名-普-一般として 1 つの品詞にまとめた。

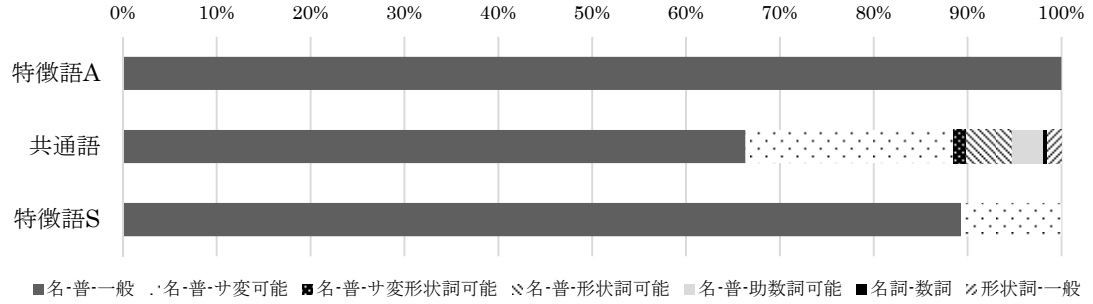

図 2：ジャンル広範囲語（237 語）の品詞比率

特徴語 A は「コンテキスト」一語で、普通名詞（一般）である。共通語と特徴語 S を比較すると、特徴語 S では「エアロビック」などの普通名詞（一般）が圧倒的多数で、「ジョギング」などの普通名詞（サ変可能）は 1 割程度である。それに対し、共通語では普通名詞（一般）が 7 割に満たず、「アクセス」などの普通名詞（サ変可能）、「オープン」などの普通名詞（サ変形状詞可能）、「イコール」などの普通名詞（形状詞可能）、「キロ」などの普通名詞（助数詞可能）、「ゼロ」などの数詞、「スムーズ」などの形状詞などが合わせて 3 割以上を占めており、普通名詞（一般）以外の品詞の割合が比較的高い。なお、7 つの品詞分類のうち、一語あたりの平均出現度数が圧倒的に高いのは「数詞」で、「ゼロ」一語で 2098 延べ語数に達する。その次に普通名詞（助数詞可能）が続き、数詞や助数詞系の語彙は少ない異なり語数がかなりの高頻度で使われていることがわかる。一方、形状詞、普通名詞（サ変形状詞可能）、普通名詞（形状詞可能）など、形状詞系は一語あたりの平均出現度数が相対的に低く、ジャンル横断的ではあるものの比較的低頻度である。

よって、頻度を基本語抽出の基準とすると、数詞や助数詞系は抽出されやすいが、形状詞系の品詞は抽出されにくい場合もあるかもしれない。

５．レジスター偏重度

最後に、無性格語 208 語についてレジスター別の出現度数を調べ、レジスターによる出現頻度の偏りのない、無性格語のなかでもさらに「無性格な」語を特定する。レジスター効果による偏りは、各語彙の「学会講演占有率（A%）」で評価する。学会講演占有率とは、学会講演における PMW（百万語当たりの出現度数）が、学会講演における PMW と模擬講演における PMW の合計の何%を占めているかを表す値である。例えば表 7 にあるように、外来語「データ」の PMW は学会講演で 4141/100428*1000000=2084、模擬講演で 85/67863*1000000=42 となり、学会講演占有率は 2084/(2084+42)=0.9801 となる。こうして求めた値をもとに、学会講演占有率が 75%より大きいものを学会講演（A）に偏って出現する「共通語（A 偏重型）」、25%より小さいものを模擬講演（S）に偏って出現する「共通語（S 偏重型）」、それ以外（25%以上 75%以下のも）を「共通語（AS 共通型）」に分類していった。その結果、表 8 に示すように A 偏重型は 48 語、AS 共通型は 109 語、S 偏重型は 51 語となった。特定の講演タイプに偏って出現している偏重型よりも、両講演タイプで同程度に出現する AS 共通型が共通語（＝無性格語）の中でもさらに「無性格な語」といえるだろう。

表 7：学会講演度数占有率（A%）に基づく共通語（＝無性格語）の下位分類作業の例

語彙素	語彙素 ID	粗頻度（学会）	粗頻度（模擬）	PMW（学会）	PMW（模擬）	学会講演占有率（A%）	語の分類名
データ	25819	4141	85	2084	42	98.01%	共通語 A 偏重型
クラス	10431	605	315	304	157	65.97%	共通語 AS 共通型
テーマ	25515	198	1432	100	714	12.25%	共通語 S 偏重型

表 8：レジスター偏重度別に見た共通語（＝無性格語）の内訳

特徴語 A	共通語（＝無性格語）			特徴語 S	総計
	A 偏重型	AS 共通型	S 偏重型		
A%=100	100>A%>75	75≧A%≧25	25>A%>0	A%=0	
	48 語	109 語	51 語		
1 語	208 語			28 語	237 語

表 9 は、本分析のまとめとして、今回扱ったジャンル広範囲語全 237 語を、これまでにみてきた意味分類、品詞分類、レジスター偏重度の 3 指標に基づいて分類したものである（五十音順、*や**は普通名詞（一般）以外の品詞であることを示す）。一方のレジスターにのみ出現する特徴語のうち、特徴語 A は「コンテキスト」1 語のみで、特徴語 S は「クーラー」、「シャワー」など具体物を示す語が多い。これ以外の、両レジスターに出現する共通語 208 語を「無性格語」と呼んだ。そのうち、レジスター偏重度の高い A 偏重型 48 語と S 偏重型 51 語を除くと、無性格語のなかでもレジスター偏重度の低い、さらに「無性格な」AS 共通型 109 語が特定できる（網掛け部分）。無性格語は総じて抽象的な語が多いが、A 偏重型では「アプローチ*」、「データ」、「データーベース」や「パーセント**」などの助数詞系など、学術分野と関連の深そうな語が目立つ。一方、S 偏重型は、「キャラク

ター」、「ファミリー」、「アルバイト*」など、より日常的な分野と関連の深そうな語が目立つ。

表 9：ジャンル広範囲語全 237 語の分類（まとめ）

	1 抽象的関係	2 人間活動-:主体	3 人間活動-精神・行為	4 生産物・道具	5 自然物・自然現象	多義	未収録	計
特徴A							1 コンテキスト	1
共通A偏重型	16 アプローチ* オフ* システム ゼロ** タイミング チャンネル ツー（< *two*） データ データーベース パターン プロセス ペア ページ** ベース ランク* レベル	1 アナウンサー	11 アクセント グラフ コントロール* サポート* テキスト テスト* プログラム* プロジェクト マーク* リスト* ルール	3 キー マイク ラベル		11 イコール* カウント* カバー* グループ ターゲット ネットワーク ピーク プラス* ブロック* マイナス* モデル	6 アンド オブ ツー（< *to*） パーセント** ミリ** ワード	48
共通AS共通型	29 アップ* ギャップ サイクル* サイズ シート ジャンル シンボル スタイル ストップ* スピード スペース スムーズ* スリー タイム チャンス デジタル* ニュー ハイ* バランス* ファイブ フル* ベスト* ポイント* メーン メリット ユニーク* ライン リアル* リード*	4 オフィス ガイド* スター モニター*	27 アイディア アウト イベント イメージ* イン インタビュー* ゲーム コミュニケーション* コメント* サイン* ストレス スピーチ* センス デザイン* トレーニング* ニュース ノー ヒント プラン フリー* マナー マニュアル ミス* メール* メッセージ メモ* ワーク	15 ウインドー エンジン カー カード カメラ ガラス コンピューター スクリーン テープ テーブル ネット ビデオ ペーパー ボタン マシン	3 オレンジ ピンク ブルー*	26 エネルギー オーバー* オープン* カラー クラス クリア* ケース サービス* シーン センター ソフト* タイプ ダウン* ダブル チェック* トップ トラブル ナンバー バック* バラエティー ヒット* ファースト ボックス ライフ リーダー リズム	5 アクセス* キロ** ザ サン メートル**	109
共通S偏重型	4 キャラクター スケジュール スタート* ペース	7 チーム ファミリー プロ ボランティア メンバー レストラン	16 アドバイス* アルバイト* エピソード オーケー* クラシック* コピー*	13 エレベーター コーヒー テレビ ドア トラック バス		9 インターネット カット* コース スーパー ノート ハード*	2 ウイーク センチ**	51

		ワールド	ストーリー スポーツ* タイトル テーマ ドラマ ビジネス マスコミ メニュー ライブ リラックス*	パソコン パンフレット ベッド メーター ラジオ ルーム ワープロ		ホーム ボール レコード		
特徴S	1	3	5	12	4	2	1	28
	リフレッシュ*	オーナー スナック プロデューサー	エアロビック ジョギング* マラソン レース レンタル*	クーラー シャワー スープ スカート ズボン デザート テント バイク バッグ ハンバーグ メダル ロープ	ゴールデン ジャングル フルーツ ミネラル	マージャン リゾート	グッズ	
計	50	15	59	43	7	48	15	237

*サ変／形状詞可能名詞・形状詞系、**助数詞可能名詞・数詞系

６．まとめ

以上、本稿では、「無性格な」外来語を抽出し、その語彙的特徴についてみてきた。その際、高頻度語を特定するだけではレジスターやジャンルの影響を排除できないため、レジスター横断性・ジャンル横断性という散らばり度に留意した。さらにレジスター偏重度を調べ、無性格語のなかでもレジスターによる出現度数の偏りが少ない語を特定した。このようにして抽出した無性格語は基本度が高く、他のコーパス調査の結果とも整合性が高いのではないかと推測される。

分析の結果、ジャンル横断性もレジスター横断性も高い無性格語は、他の語群と比べて「1 抽象的関係」「3 人間活動（精神・行為）」「多義」の割合が高い反面、「4 具体物・道具」「5 自然物・自然現象」の割合は著しく低かった。表 9 を見ると、「4 生産物・道具」は主に具体語が分類されていることから、その割合が相対的に低いということは、裏を返せば、対立する抽象語の割合が高いということでもある。これは、明治後期において基本語化した漢語の 3 類型の一つとして「抽象概念を表す語」を挙げた田中（2014）の考察と共通する部分がある（ただし、「基本語」や「抽象的」の定義は完全に同じではない）。具体的な意味を持つ語よりも抽象的な意味を持つ語のほうが使用頻度や使用範囲が拡大しやすいということは直観的にも理解しやすい。金（2011）は新聞において通時的増加傾向を見せる外来語は抽象名詞に多いとし、その一例である「ケース」が意味範囲を拡大させながら類義語のなかで出現率を伸ばしていることを指摘したが、抽象的な意味を持つ語にはこうした意味範囲の拡大、あるいは変化を通じて使用頻度や使用範囲を拡大させる潜在性があるのかもしれない。

品詞に関しては、無性格語は、それ以外の語群と比べて、サ変可能名詞や形状詞可能名詞などといった普通名詞（一般）以外の品詞を多く含むことがわかった。この傾向も明治後期以降基本語化した漢語と類似している（田中 2012）。このことは、外来語が名詞だけではなく動詞系や形状詞・形容詞系といった品詞カテゴリーにおいても広がりを見せてい

ることを示唆するものである。しかし、これを確かめるには、個々の用法を吟味してサ変動詞用法や形状詞用法のみを取り出し、そうした用法が実際にどれほどあるのかをみなければならない。そうした側面を調べるために、久屋（2014）では、サ変可能名詞である「サポート」、「イメージ」、「キープ」、「マスター」、「スタート」などのサ変動詞用法だけを取り出し、これら外来語に対応する既存類義語である和語動詞や漢語サ変動詞用法との量的関係を調べた。その結果、既存語に対する外来語の使用率が若年層を中心に増加していることが明らかになった。

今回抽出した基本度の高い外来語の語彙的特徴は、明治後期以降に基本語化したかつての借用語である漢語のそれと類似する部分がある。ということは、こうした語彙的特徴は、外来語に限らずあらゆる語種にとって基本語化の重要な要素である可能性がある。いずれにせよ、こうした外来語の広がりが、同じような語彙的特徴を持つ漢語や和語にどういった影響を及ぼしているのかについては、外来語・漢語・和語の語種全体を巨視的に眺めた研究が望まれるところである。この点に関しては今後の課題とする。

謝　辞

本稿で分析に利用した CSJ および分類語彙表関連データは、筆者が国立国語研究所に特別共同利用研究員として滞在していた期間（2014 年 9 月～現在）に、同研究所の許可を得て使用させていただいたものである。ここに感謝申し上げる。

文　献

小木曽智信、中村壮範（2011）「『現代日本語書き言葉均衡コーパス』形態論情報データベースの設計と実装 改訂版」国立国語研究所内部報告書 LR-CCG-10-06

小木曽智信、中村壮範（2014）「『現代日本語書き言葉均衡コーパス』形態論情報アノテーション支援システムの設計・実装・運用」『自然言語処理 21:2』, pp.301-332

金愛蘭（2011）「20 世紀後半の新聞語彙における外来語の基本語化」『阪大日本語研究―別冊 3』

久屋愛実（2014）「外来語の共時的分布パターンの一般化に向けた予備的考察」『韓国日本語学会第 30 回国際学術発表大会予稿集』, pp.156-165

国立国語研究所（2004）『分類語彙表―増補改訂版』大日本図書

国立国語研究所（2006）『日本語話し言葉コーパスの構築法』

田中章夫（1973）「自動抄録処理におけるキー・ワードの性格」『電子計算機による国語研究 V』, pp.141-184, 国立国語研究所

田中牧郎（2012）「明治後期から大正期の語彙レベルと語種―『太陽コーパス』の形態素解析データによる」田中牧郎ほか（2012）『近代語コーパス設計のための文献言語研究 成果報告書』国立国語研究所共同研究報告 12-03

田中牧郎（2014）「明治後期における漢語の基本語化」『第 6 回コーパス日本語学ワークショップ予稿集』, pp.193-200

水谷静夫（1964）「語の基本度」『現代雑誌九十種の用語用字―第三分冊（分析）』, pp.7-51, 国立国語研究所