

## 第5章 「外来語」言い換え提案対象語の頻度推移類型 ——読売新聞・毎日新聞14年分を調査して——

柏野 和佳子

### 1 はじめに

「外来語」言い換え提案は、日本語に十分定着していない、分かりにくい外来語を対象として行われた。本稿では、その対象語の新聞14年間にわたる使用状況を考察する。「外来語」言い換え提案は、第1回から第4回まで、合計4回に分けて行われたが、そのうちの第1～2回の提案対象語(109語)の読売新聞1986～2001年における調査結果は、既に柏野・他(2005)で報告した。ここでは語の出現率(=出現度数÷収録文字数)の推移に着目し、その類型と社会調査の「理解率」とを用いて、外来語の使用実態について分析した。

本稿では、調査対象語を全4回分の176語(詳細分析は低頻度語を除いた156語)に増やし、検討を進める。その際、語の出現率の推移を類型化する手法に焦点を絞り、これを「頻度推移類型」と呼ぶ。また、新聞調査から得られる頻度推移類型の意義を明確化するため、新聞調査から得られる頻度情報の特性を分析する。すなわち、新聞を調査に使用する上での問題として、新聞の特性、及び、対象語の同定について考察する。それとともに、新聞社固有の偏りの有無を調べるため、読売新聞と毎日新聞の2紙による調査結果を比較する。

なお、柏野・他(2005)に対し、今回の調査範囲は次のとおり拡大、変更している。

- ・読売新聞に、毎日新聞を新たに加える。
- ・調査範囲を両方の記事データベースのそろそろ1991年以降とする。
- ・最近の外来語の出現率の推移をとらえるために調査範囲を2004年までに広げる。

頻度推移類型に基づく分析は、経年調査による語彙調査に位置づけられるものである。本稿では、最初に、外来語の使用状況をとらえるために経年調査による語彙調査を行うことの意義について述べる。次に、頻度調査においては、対象データと、得ようとする数値の吟味が欠かせないため、3節で調査に用いた新聞記事データの概要を述べ、4節で調査上の留意点を述べる。続く5節で使用度数の総計を示した後、6節において頻度推移類型に基づく外来語分析結果を詳述する。そして、7節で、読売、毎日の2紙の調査結果の比較から新聞社固有の偏りについて言及する。

なお、6節に述べる語別の頻度推移のグラフは、176語すべてについて、本報告書の第1部に収録した語別の「調査データ」の中に「新聞の頻度推移」として掲載した。

### 2 経年調査による語彙調査の意義

語の使用状況をとらえるためには、語彙調査、なかでも使用頻度の調査はたいへん有効である。これまでに国立国語研究所(1962, 1964)をはじめ、各種の語彙調査が行われてきた。それらの多くは調査時の使用状況の把握を目的にしていたため、対象年を1年以内に限定するものがほとんどであった。しかしながら、語の使用状況の把握には、時間的推移という視点もまた重要である。

経年調査による語彙調査には、雑誌用語の10年ごとの変遷調査(国立国語研究所1987)、明治37年～昭和24年の国定読本の語彙調査(国立国語研究所1997)、雑誌用語の語種構成を約40年の間隔において比較した分析(山崎・小沼2004)がある。しかしながらいずれも規模は限られている。

電子テキスト化された大量の各種新聞記事データベースが研究に利用できるようになって以来、これを用いて時間的推移を大規模に分析する研究が現れてきた。毎日新聞や朝日新聞におけるカタカナ語の月次変動の分析(久野他1998; 野崎他1998)、毎日新聞における時系列変動と文字使用の面差の分析(久野2000)、毎日新聞と朝日新聞における時系列変動と文字使用の紙面差の分析(久野他2002)、毎日新聞における出現頻度の時間的変化の語種による分析(山口他2004)などである。たとえば山口他(2004)では、毎日新聞9年分を対象に調査し、外来語は他の語種よりも変動係数の平均、分散が大きいことを報告している。

そして外来語の特徴として、高頻度の形態素において、頻度変化の大きい語が多いと考えられることを指摘している。ほかに、新聞記事データベースを対象にした大規模な頻度調査に朝日新聞14年分の頻度調査(NTT2000)などもある。

新しく次々に登場する外来語の使用実態を論じる上で、時間的変化を経年調査により外来語の盛衰としてとらえることの意義は大きい。そこで、新聞データを用いて、語の使用推移、すなわち出現率の推移に着目して語の使用状況をとらえることを試みた。

### 3 読売新聞・毎日新聞の電子化データについて

本調査には、読売新聞と毎日新聞の1991年から2004年までの14年分の記事データを使用した。その概要を記す。

#### 3.1 読売新聞

読売新聞記事データの概要は次のとおりである。

- ・読売新聞社の記事データベースに蓄積してあるもののうち、著作権が読売新聞社に無いものを除いた全記事<sup>1</sup>。
- ・東京本社発行記事に加え、地方支社（大阪，西部，中部）発行記事が入っている。
- ・1999年以降はさらに県版（地域版）の記事が入っている。
- ・新聞紙面記事との違いは、次の2点である。

①見出しをデータベース用に再編集している。

②紙面上で見開き2ページにまたがってレイアウトされた記事をデータベースに採録する際には、ページごとに分けて採録せず、1つの記事としてまとめて右ページにある記事として採録している。そして、左ページにあるものとして、見出しのみ、もしくは見出しと本文の一部のみを採録している。その結果、左ページ入力分は右ページ分との二重採録になっている。

今回の調査では、記事データのうち、本文のみを用いた。読売新聞記事データには本文の他に、見出しやキーワード（統制語及び自由語）、分類コード、面種コードなどの書誌情報も入っているが、これらは用いていない。調査に用いた本文の文字数のグラフを図1に示す。図1より1999年を境にデータ量が大きく増加している。これは、1999年より県版（地域版）記事が加わったことによるものとみられる。

#### 3.2 毎日新聞

毎日新聞記事データ<sup>2</sup>は、毎日新聞の東京・大阪本社の朝夕刊最終版を対象とした全文記事データである。今回の調査では、記事データのうち、本文のみを用いた。毎日新聞記事データには本文の他に、見出しやキーワード、掲載面種別コードなどの書誌情報も入っているが、これらは用いていない。調査に用いた本文の文字数のグラフを図2に示す。読売新聞に比べると、1991年から2004年までの間でさほどデータ量に大きな変動は見られない。

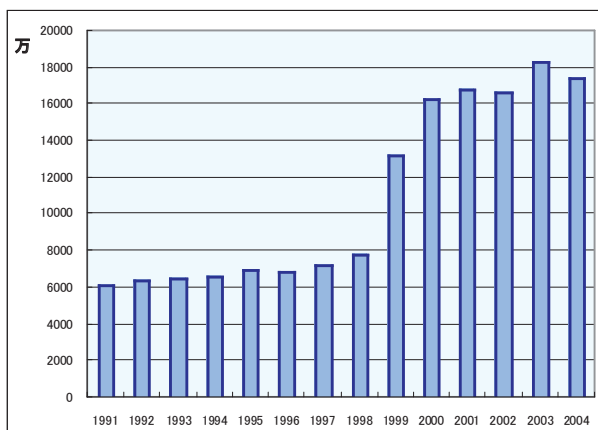


図1 読売新聞記事データ文字数

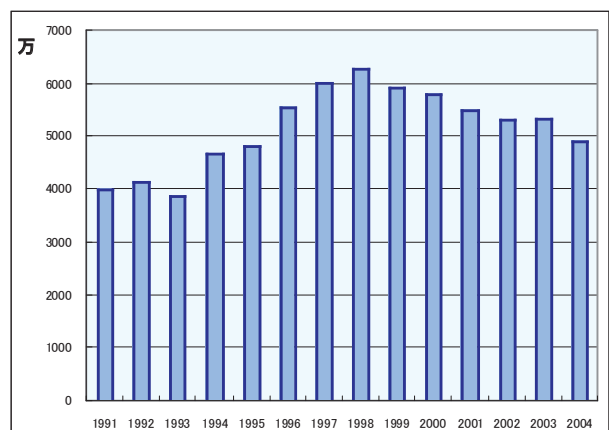


図2 毎日新聞記事データ文字数

## 4 調査上の問題点とその対処

新聞データを調査に用いる上で問題となる点、および、その対処を述べる。

### 4.1 データ量増加の影響

読売新聞に顕著であるが、年により記事データの量が異なるという問題がある。このため、より使用状況を的確に表す数値として、出現率、すなわち使用度数を収録単語数で正規化した値を用いることにした。ただし、平均単語長は年によらずほぼ一定とみなせるので、本稿では収録単語数の代わりに、より算出の容易な収録文字数を用いて正規化することとし、次の式によって、年ごとの出現率を算出した。

$$\text{出現率} = \text{使用度数} \div \text{収録文字数}$$

### 4.2 重複用例の影響

柏野・他(2005)で述べたように、重複記事と、それにより抽出される重複用例の扱いには注意を要する。このような重複記事は、特に読売新聞に多くみられる。

読売新聞では、たとえば、本社版、支社版、地方版とで記事が重複するものを、度数を数える際にその重複を排除して数えるべきかが問題になる。しかし、もし重複を用例単位で除こうとすれば調査母体が対象語ごとに異なってしまう、記事単位で除こうとすれば調査の対象記事データがかなり限られてしまう。重複の数を調べた結果、多い場合でも全体の約1%以下であったため、今回の調査では重複は無視できると判断した。

実際のところ、重複の排除は判断に迷うものが多く、作業は困難である。たとえば掲載日が同じで、内容がほとんど同一であり、発行だけが違うようであれば比較的重複であるとの判断がしやすいが、そのようなものは少ない。地方、地域によって日を変えて掲載されている場合も多く、掲載日だけでは判断できない。また、文や記事レベルにおいて重複の度合いが様々であり、異なる部分が多くなるほど重複と判断しがたくなっていく。文レベルでは、全く同一のもの、句読点のみ違うもの、表記、語句が少しだけ違うものから、語句がかなり違うものまである。記事レベルでは、ほとんど同じものもあれば、一方が詳細記事のためそちらだけに存在する文が多くあり、記事全体での重複文は実際には少ないというものまである。同じテーマで次々に新しい記事が書かれる場合にも文や語句単位で重複は多く生じるが、この場合と問題にする重複記事とを区別する必要もある。

なお、このような重複は読売新聞、毎日新聞のほかにも、日経新聞にもあることが報告されている(Aizawa2003)。今後、Aizawa(2003)のように、機械的に重複を調べる研究も進んでくるであろうが、新聞記事の編集過程にまで立ち入った議論をする必要もあると思われる。

### 4.3 対象語の同定の問題

柏野・他(2005)で述べた2つの問題点を再掲する<sup>3</sup>。

しばしば指摘されているように外来語には表記のゆれのあることが多い。そこで考えられる表記のゆれを想定して検索することで対処した。

例：アイデンティティー 全2462件

「アイデンティティー」1987件

「アイデンティティ」472件

「アイデンティーター」1件

「アイデンティ」1件

「アイデンテティ」1件

また、今回は、文字列マッチングのみによって用例を抽出したが、部分的に文字列が同じである別語の用例が抽出されることがある。これには人手によって排除することによって対処した。

例：スクリーニング 全311件

「ハウスクリーニング」

「ガラススクリーニング」など71件を排除

なお、形態素解析プログラムを使用したとしても、人手による見直しは必要と考えられる。

## 5 使用度数の総計

対象176語について読売新聞および、毎日新聞における各年の使用度数、出現率、および使用度数の総計を出した。

14年間の使用度数の総計が1万を超えたのは次の語のみである。度数の総計とともに多い順に示すと、読売新聞「ケア31,219, シェア15,981, ベンチャー15,883, アクセス14,879, セクター13,930, ビジョン13,725, コミュニティー12,876, ガイドライン11,246」, 毎日新聞「ケア11,415, シェア10,292」である。読売新聞では118語, 毎日新聞では130語の使用度数の総計が1,000にも満たなかった。そのうち60にも満たなかったものがそれぞれ次のとおりである。(添えてある数字は使用度数の総計。)

**読売新聞：**リードタイム4, エンフォースメント10, バックオフィス17, サマリー18, ハーモナイゼーション18, セットバック28, カスタムメイド30, アクセシビリティ33, カウンターパート35, オーガナイザー44, ログイン50, パブリックインボルブメント52, ボトルネック56, オーナーシップ57

**毎日新聞：**リードタイム6, エンフォースメント10, パブリックインボルブメント18, セットバック20, ノンステップバス21, スtockヤード22, サプライサイド25, サマリー27, ログイン30, リデュース31, バックオフィス33, ボトルネック37, ハーモナイゼーション39, アクセシビリティ44, オーナーシップ55, モビリティ52, オフサイトセンター56, カウンターパート56

一般に分かりにくいと思われる外来語は、おおむね、新聞において、まだその使用は多くない語であることを、使用度数の総計によりまずは確認することができる。

## 6 「頻度推移類型」に基づく分析

14年間における出現率の増減傾向を分類した頻度推移類型に基づき、対象語を分析する。

### 6.1 頻度推移類型の6つの型による分類

はじめに、本稿で設定する頻度推移類型の6分類<sup>4</sup>とその判断基準とを表1に示す。

表1 頻度推移類型の判断基準

分類	判断基準
突出型	特定の年のみ増加している点が顕著なもの
増加型	期間内をならずと右上がり
減少型	期間内をならずと右下がり
山型	3年以上の増加後3年以上の減少が顕著なもの
水平型	すべて出現率 $0.5(\times 10^{-6})$ 以上かつ期間内をならずと平ら
非特異型	上記のいずれにも分類されなかったもの (出現率 $0.5(\times 10^{-6})$ 未満の値があり、一定の傾向のないもの)

6分類のうち、「増加型, 減少型, 山型, 水平型, 非特異型」の5つは、語の生命サイクルとしてとらえることができる。一般に、語の生命のサイクルとして、登場して、出現が増加し、定着すれば出現推移が水平になり、また、その語が廃れていくにあたっては出現が減少していく、という経過をたどると考えら

れる。使用が安定しその出現期間が長い語は一定の調査期間内においては「水平型」となり、新語は「増加型」、廃れつつある語は「減少型」としてとらえられよう。「山型」は、一時的に非常によく使われる語が示す型だと思われる。出現が増えた後、安定せずに減少に転じる、ということが短期間のうちに起こることで山型になると考えられる。また、登場したものの、低頻度のまま、生命サイクルに乗る前の状況が続く場合もあり得る。そのようなものは一定の傾向、すなわち、特異性がないものとして「非特異型」と呼ぶこととした。

一方、もう1つの「突出型」は、語の生命サイクルとは別軸の分類とみるべきものである。その語がどの生命サイクルに位置しているかを問わずに、特定の年のみ増加している点が目立つという特徴によってとらえる型である。語の生命サイクルの軸と重ねて、「増加型、減少型、山型、水平型、非特異型」の5分類を突出の有無によって10分類に細分するという方法も検討したが、今回の対象語については、突出のある点を優先し、それをひとくりに扱う方が対象語の特徴を見通しやすくと考え、一つの分類項目としてたてた。つまり、判断基準のうちではこれのみが優先され、その他には優先順位はない。

読売新聞、毎日新聞、いずれかが度数60未満であった20語（前節参照）は、度数が少なすぎて推移傾向をとらえ難いため分析対象外にした。よって、対象語156語は、最多度数「ケア 読売：31,219・毎日：11,415」から最少度数「プライオリティー 読売：79・毎日：65」までである。このうち、度数60以上であっても、出現率が $0.5(\times 10^{-6})$ 未満で推移しているような場合には、その出現率の変化は、度数の多いものに比べて非常に小さなものになる。しかしながら、たとえ小さくても増減傾向をとらえることが語の使用状況の把握には意味があると考え、出現率の大小に関わらず、すべての語に対して推移の型を相対的に導き出して分類した。実際に、「突出型、増加型、減少型、山型」については、いずれにも $0.5(\times 10^{-6})$ 未満の値を含むものも分類している。問題は、低頻度で特異性のない「非特異型」と、高頻度で出現率が高いまま平らに推移する「水平型」との区別である。両者を客観的に区別するものとして、「出現率 $0.5(\times 10^{-6})$ 」との数値判断を導入した。なお、この「 $0.5(\times 10^{-6})$ 」という数値は、経験的に導き出したものである。

次に、対象語156語について読売新聞、毎日新聞別に、14年間の出現率の増減傾向によって上記6分類を行った結果の語数を表2に示す。集計は各語の分類が両紙で一致しているか一致していないかを区別して行った（そのため、型別の合計語数は156語ではなく、197語）。「突出型、山型」については、ピーク時が大きく違う場合は、その違いを重視し、不一致として集計している。また、「非特異型」についても、推移傾向が大きく異なる場合は、同様に不一致として集計している。その結果、対象語156語のうち、両紙での分類が一致したものが約7割の106語、不一致であったものが約3割の50語であるとの結果が得られた。この2紙間の異同については、7節で再び取り上げる。

型別の集計結果で目立つ点は、「山型」が多いという点である。この点を含め、6.2節以降に調査結果を型別に分析する。以下、例を示す図の各グラフは、増減の推移を分かりやすくするため、出現率の大きさによってY軸の目盛り幅を調整している。

表2 頻度推移類型の分類結果

分類	語数	両紙一致の語数		両紙不一致の語数*	
		読売・毎日	読売	毎日	
突出型	24	10	4 (2)	12 (2)	
増加型	36	20	9	7	
減少型	26	17	9	0	
山型	65	37	16 (4)	16 (4)	
水平型	15	12	1	2	
非特異型	31	10	11 (3)	13 (3)	
合計(語)	197	106	50 (9)	50 (9)	

\*: 突出型、山型は、ピーク時が大きく異なる場合は不一致として集計。また、非特異型も推移の傾向が大きく異なる場合は不一致として集計。( )内はその語数。

## 6.2 突出型

出現率が他と比べて非常に突出している年のあるものがある。ある特定の年に出現率が高くなるのは、その年に盛んに取り上げられた話題や政策に関わる語であり、ニュース性が高かったためと見られる。新聞は、重要だと判断したことをニュースとして伝えることを主目的とした文書である。新聞のニュースについて佐藤(1991)は、「その日の相対的な重要度によって、ニュースを価値判断し、秩序だった情報に再構築して、その日の最重要ニュースを最低限押さえられる」ようになっているものであると述べている。大きいニュースは必ず大きく取り上げられる。また、何度もニュースとなるような事柄は必ず何度も取り上げられる。図3に示す「セーフガード」は貿易摩擦問題、とりわけ輸入制限についての議論が沸騰した2001年の出現率が突出している。また、図4に示す「ライフライン」は阪神・淡路大震災のあった1995年の出現率が突出している。

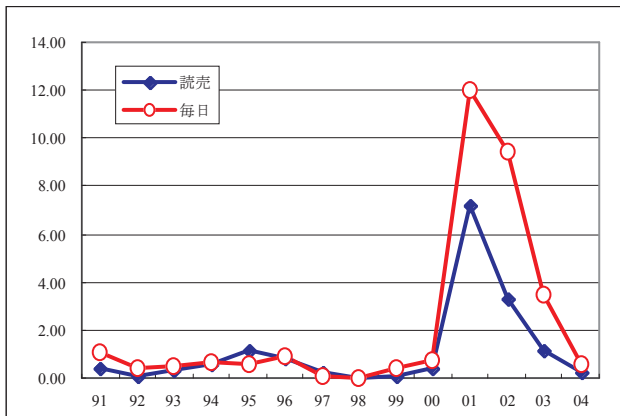


図3 「セーフガード」出現率( $\times 10^{-6}$ )

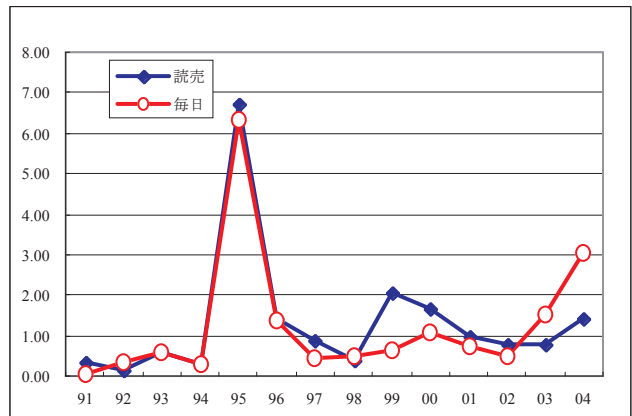


図4 「ライフライン」出現率( $\times 10^{-6}$ )

突出がより顕著な25語をこの突出型に分類した。分類した語の一覧を度数の多い順<sup>5</sup>に表3に示す。

表3 突出型に分類した語

見出し	両紙一致の分類		両紙不一致の分類		総度数	
	読売・毎日	読売	毎日	読売	毎日	
ガイドライン	突出型			11246	7682	
ドナー	—	減少型	突出型	5089	4944	
ダンピング	突出型			3517	2790	
セーフガード	突出型			2335	1644	
ライフライン	突出型			1973	894	
インサイダー	突出型			1278	1555	
アジェンダ	突出型			620	529	
マンパワー	—	非特異型	突出型	609	309	
メンタルヘルス	—	突出型	非特異型	566	219	
アーカイブ	—	増加型	突出型	408	220	
ノーマライゼーション	—	非特異型	突出型	390	209	
アクションプログラム	突出型			382	226	
ポートフォリオ	—	突出型	非特異型	360	191	
スクリーニング	—	突出型	突出型	354	136	
サーベイランス	—	山型	突出型	309	331	
コミット	—	減少型	突出型	266	300	
レシピエント	—	減少型	突出型	263	225	
タスクフォース	突出型			260	234	
サムターン	突出型			255	62	
ソフトランディング	—	非特異型	突出型	247	267	

リターナブル	—	突出型	突出型	132	90
メディカルチェック	—	非特異型	突出型	123	73
フレームワーク	突出型			103	108
デジタルデバインド	—	非特異型	突出型	89	201

### 6.3 増加型

右肩上がりに大きく増加傾向を示すものを増加型とした。今回の14年間の調査期間中、増加傾向を維持し続けた語は、表4に示すとおりである。

表4 増加型に分類した語

見出し	両紙一致の分類		両紙不一致の分類		総度数	
	読売・毎日	読売	毎日	読売	毎日	
セクター	—	増加型	水平型	13930	2925	
コミュニティー	増加型			12876	4111	
ワークショップ	—	増加型	山型	5040	1566	
シンクタンク	—	増加型	山型	3726	2526	
リニューアル	—	山型	増加型	2820	1464	
セキュリティ	増加型			2445	1860	
ハイブリッド	増加型			2113	1266	
ビオトープ	—	増加型	山型	1958	304	
インターンシップ	増加型			1647	396	
リリース	増加型			1131	1354	
ハザードマップ	—	山型	増加型	1024	154	
コンソーシアム	増加型			990	330	
ユニバーサルデザイン	増加型			903	220	
ツール	増加型			835	797	
リバウンド	—	山型	増加型	791	424	
バイオマス	増加型			760	329	
ヒートアイランド	増加型			701	286	
スキル	増加型			539	273	
ポジティブ	—	山型	増加型	535	453	
ドクトリン	—	減少型	増加型	532	522	
ガバナンス	増加型			490	387	
コラボレーション	増加型			478	390	
コンプライアンス	増加型			467	218	
セカンドオピニオン	—	増加型	山型	439	213	
モチベーション	増加型			437	301	
アーカイブ	—	増加型	突出型	408	220	
ネグレクト	—	増加型	山型	397	122	
サプリメント	増加型			378	423	
トレーサビリティ	—	増加型	山型	376	114	
インタラクティブ	—	山型	増加型	294	1726	
インキュベーション	増加型			275	90	
リユース	増加型			265	93	
ワンストップ	増加型			257	111	
ユニバーサルサービス	増加型			236	211	
エンパワーメント	—	増加型	非特異型	209	108	
フリーランス	—	減少型	増加型	97	130	

外来語は新しい概念を表すために導入されることが多い。その新しい概念の普及とともに、それを表す外来語の使用が増えていくということがしばしば観察される。そして、使用が増えていくという局面にある語がこの増加型を示すと言える。たとえば、「サプリメント」(図5)は今回追加調査を行った2002年以降に急激に増加し、ここ最近の急速な普及状況が確認できる。その一方で、特に新しい概念を表す外来語ではなくとも、増加型としてとらえられるものもある。たとえば、「スキル」(図6)である。「スキル」とは、訓練によって身につけることができる、技術上の能力のことである。その事柄自体の新聞での取り上げられ方が増加している、というよりは、その事柄を「スキル」と表現することが新聞上で増加している点がとらえられている、と言えるのではないであろうか。

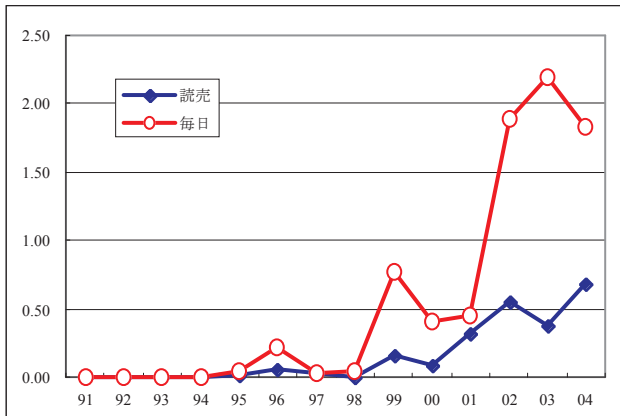


図5 「サプリメント」出現率(×10<sup>-6</sup>)

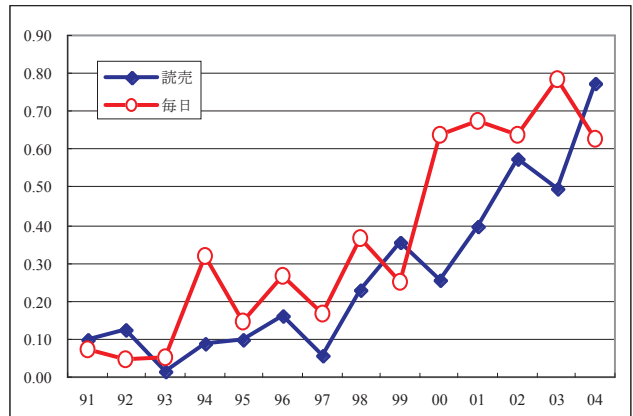


図6 「スキル」出現率(×10<sup>-6</sup>)

#### 6.4 減少型

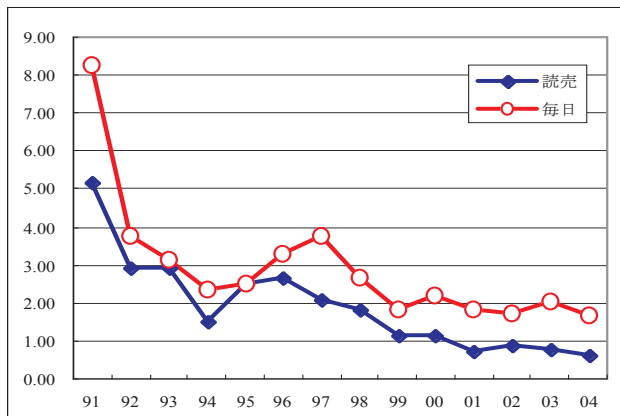


図7 「イニシアチブ」出現率(×10<sup>-6</sup>)

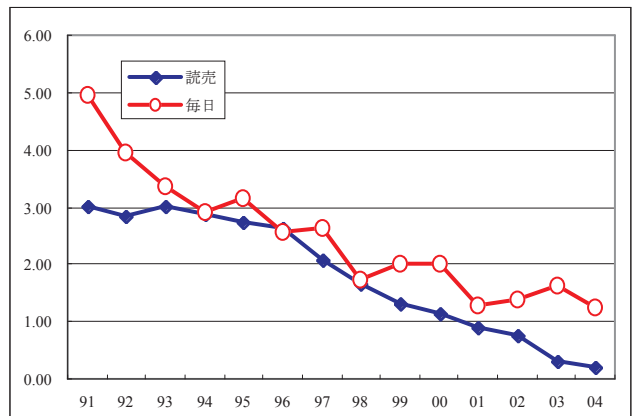


図8 「コンセンサス」出現率(×10<sup>-6</sup>)

一般には外来語の使用は増えているという印象があるが、外来語は短命であることが少なくない。水平型に進む前の段階で失速してしまうようである。「アメニティー」の減少については関根(2003)に言及があり、「フレックスタイム」は柏野・他(2005)で取り上げた。これらは、一時期新聞を賑わせたが、その後話題に上ることが徐々に少なくなり、新聞での使用が減少していると考えられる。

ここでは、減少型の例として「イニシアチブ」と「コンセンサス」を図7、図8に示す。どちらもはっきりとした右下がりの推移が認められる。表5に減少型に分類した語の一覧を示す。

表5 減少型に分類した語

見出し	両紙一致の分類		両紙不一致の分類		総度数	
	読売・毎日		読売	毎日	読売	毎日
ドナー	—		減少型	突出型	5089	4944
バックアップ	減少型				4953	2495



アナリスト	—	減少型	山型	3156	1863
イニシアチブ	減少型			2322	2041
オブザーバー	減少型			2249	1562
コンセンサス	減少型			2156	1722
インフォームドコンセント	減少型			1885	1119
ミッション	減少型			1816	1449
コミュニケ	減少型			1003	931
ボーダーレス	減少型			945	615
プレゼンス	減少型			881	564
フォローアップ	減少型			821	560
アメニティー	減少型			661	282
ドクトリン	—	減少型	増加型	532	522
モラトリアム	減少型			513	493
コミットメント	減少型			481	398
フレックスタイム	減少型			407	191
ライフサイクル	—	減少型	非特異型	396	200
キャピタルゲイン	減少型			327	200
コミット	—	減少型	突出型	266	300
レシピエント	—	減少型	突出型	263	225
ステレオタイプ	—	減少型	非特異型	226	201
ケーススタディー	減少型			208	130
オピニオンリーダー	—	減少型	山型	177	181
プロトタイプ	減少型			103	105
フリーランス	—	減少型	増加型	97	130

### 6.5 山型

今回、最も多くの語を分類したのが、この山型であった。山型というのは、調査期間内のうちに、増加の局面も見せ、減少の局面も見せ、かつ、短期間のうちにその両方を見せているという点において特徴のある型である。この山型が多かったということが、今回の調査対象語の大きな特徴を表していると言える。

実は、第1～2回の対象語(109語)について読売新聞1986～2001年における調査をした柏野・他(2005)では、増加型が44語ともっとも多く、次いで、水平型17語、減少(右下がり)型13語、減少(山・山の後平ら)型14語、という順に続いていた。なぜ今回の調査で山型が多くなったのか。追加した第3～4回の対象語67語のうちで山型は18語であり、追加によって特別に山型が増えたわけではない。今回山型が多くなった大きな原因として、柏野・他(2005)で増加型に分類した語のうち、延長した期間の2002～2004年に増加から減少に転じた語が26語もあったことが挙げられる。たとえば図9に示す「バリアフリー」<sup>6</sup>や図10に示す「ベンチャー」である。

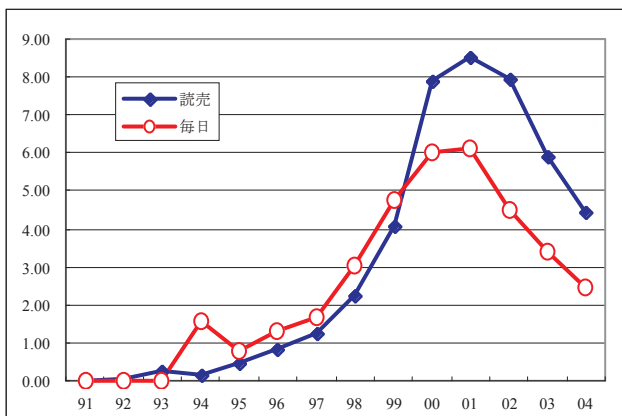


図9 「バリアフリー」出現率(×10<sup>-6</sup>)

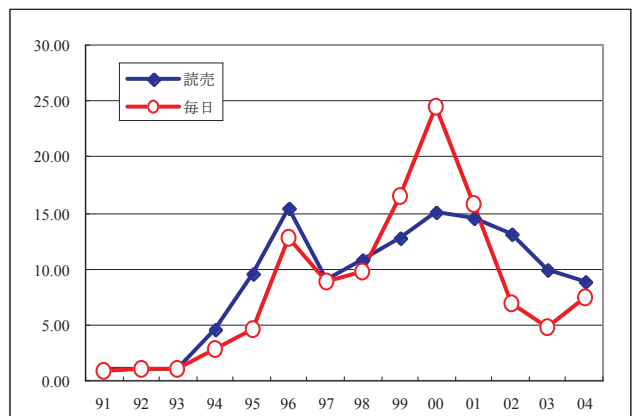


図10 「ベンチャー」出現率(×10<sup>-6</sup>)

山型とひとくくりに行っている中で、増加から減少に転じる時点は語によって異なる。たとえば、図11の「マルチメディア」は1994年～1996年に、図12の「アカウントビリティー」は1996年～1998年に増加から減少に転じている点がとらえられる。

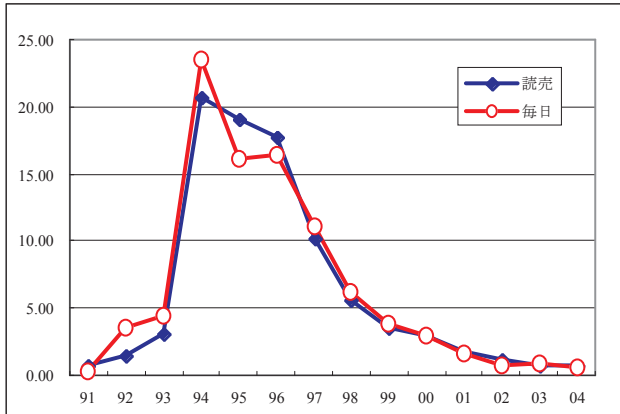


図11 「マルチメディア」出現率(×10<sup>-6</sup>)

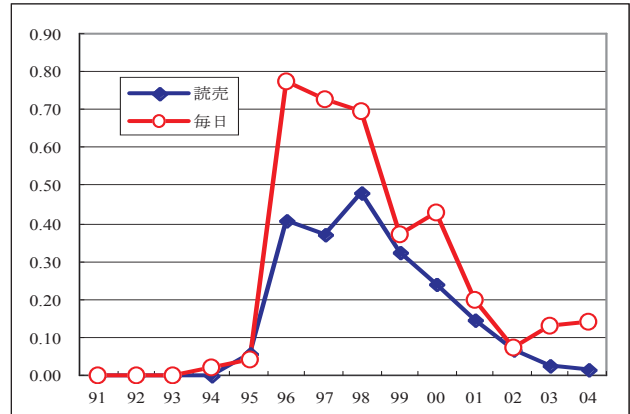


図12 「アカウントビリティー」出現率(×10<sup>-6</sup>)

以下、山型に分類した65語の一覧を表6に示す。

表6 山型に分類した語

見出し	両紙一致の分類		両紙不一致の分類		総度数	
	読売・毎日	読売	毎日	読売	毎日	
ケア	山型			31219	11415	
ベンチャー	山型			15883	6599	
アクセス	—	山型	水平型	14879	8278	
グローバル	山型			9319	6195	
マルチメディア	山型			7181	4782	
バリアフリー	山型			6845	1991	
デイサービス	山型			6478	1204	
インフラ	山型			5843	3456	
ワークショップ	—	増加型	山型	5040	1566	
マネージメント	山型			4892	2265	
シフト	山型			4156	3451	
マーケティング	—	水平型	山型	3806	2349	
シンクタンク	—	増加型	山型	3726	2526	
マクロ	山型			3712	2737	
アナリスト	—	減少型	山型	3156	1863	
リニューアル	—	山型	増加型	2820	1464	
アセスメント	山型			2807	959	
リアルタイム	山型			2578	1369	
バーチャル	山型			2344	1388	
コンテンツ	山型			2170	1989	
ビオトープ	—	増加型	山型	1958	304	
バイオテクノロジー	山型			1826	649	
セーフティネット	山型			1352	861	
ワークシェアリング	山型			1340	511	
モニタリング	山型			1295	471	
ミスマッチ	山型			1121	508	
ナノテクノロジー	山型			1098	161	

ワーキンググループ	—	山型	非特異型	1095	528
リーフレット	—	山型	非特異型	1027	208
ハザードマップ	—	山型	増加型	1024	154
スキーム	山型			991	789
アミューズメント	山型			944	531
プレゼンテーション	山型			914	473
オペレーション	—	非特異型	山型	837	568
グローバル化	山型			830	902
リバウンド	—	山型	増加型	791	424
ソリューション	—	山型	山型	767	349
マスタープラン	山型			752	144
ポジティブ	—	山型	増加型	535	453
トラウマ	山型			521	568
ランドデザイン	山型			513	251
オンデマンド	山型			471	213
デポジット	—	山型	非特異型	467	201
アウトソーシング	山型			451	265
セカンドオピニオン	—	増加型	山型	439	213
モラルハザード	山型			407	597
ゼロエミッション	—	山型	山型	405	171
リテラシー	山型			402	173
ネグレクト	—	増加型	山型	397	122
トレーサビリティ	—	増加型	山型	376	114
コンポスト	山型			372	178
スケールメリット	—	非特異型	山型	361	194
パブリックコメント	—	山型	山型	339	75
スクーリング	—	非特異型	山型	329	149
サーベイランス	—	山型	突出型	309	331
アイドルングストップ	—	山型	非特異型	299	81
インタラクティブ	—	山型	増加型	294	1726
モータリゼーション	山型			274	132
ロードプライシング	山型			270	83
クライアント	山型			230	232
アカウントビリティ	山型			223	210
タイムラグ	—	山型	非特異型	202	140
キャッチアップ	山型			187	146
オピニオンリーダー	—	減少型	山型	177	181
フィルタリング	—	山型	山型	126	66

## 6.6 水平型

水平型とは、そもそも、今回の調査対象外となる、定着した語が示す類型であろう。たとえば、調査対象外の「テーマ」は定着した外来語の一つと考えられるが、その毎日新聞の出現率の推移を参考に調べた結果を図13に示す。この「テーマ」のような出現率の推移をもつ語が水平型の典型例と言えるだろう。それら定着したものに連続するものとして、今回、「すべて出現率 $0.5(\times 10^{-6})$ 以上かつ出現期間内をならずと平ら」という判断基準のもと、図14に示す「シミュレーション」など15語を「水平型」に分類した。語の一覧を表7に示す。

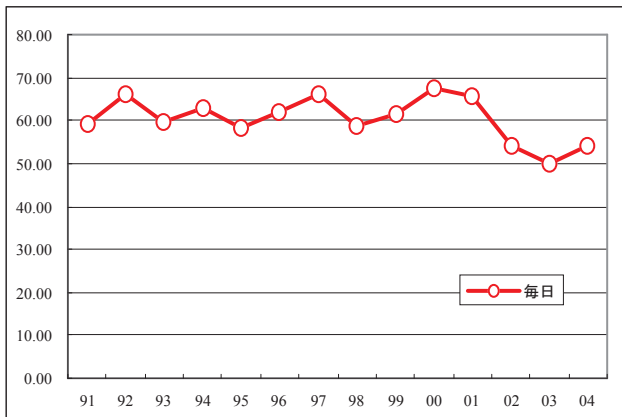


図13 「テーマ」出現率(×10<sup>-6</sup>)

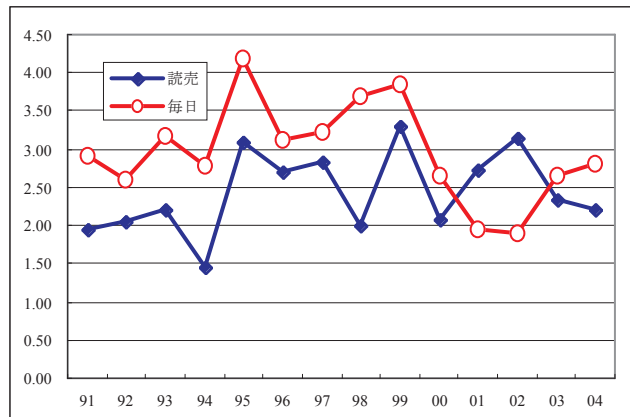


図14 「シミュレーション」出現率(×10<sup>-6</sup>)

表7 水平型に分類した語

見出し	両紙一致の分類		両紙不一致の分類		総度数	
	読売・毎日		読売	毎日	読売	毎日
シェア	水平型				15981	10292
アクセス	—		山型	水平型	14879	8278
セクター	—		増加型	水平型	13930	2925
ビジョン	水平型				13725	7639
シミュレーション	水平型				3826	2155
マーケティング	—		水平型	山型	3806	2349
パートナーシップ	水平型				3455	2554
インパクト	水平型				3189	2204
スタンス	水平型				3043	2457
アイデンティティ	水平型				2792	1880
フロンティア	水平型				2666	1057
コンセプト	水平型				2597	2337
コア	水平型				1755	793
ライブラリー	水平型				1598	748
トレンド	水平型				1334	1162

「出現率0.5(×10<sup>-6</sup>)以上」という数値は、今回の対象語の中から水平型を分類しようとした場合に妥当と判断した値である。もっと出現率の高い語を対象に、語の定着の程度を厳しく認定しようとする場合は、さらに高い数値に設定することが検討されることになろう。このあたりの吟味は今後の課題としたい。

## 6.7 非特異型

「度数60未満」の20語は度数が少なすぎて推移の傾向をとらえ難いため分析対象外にしたが、「出現率が0.5(×10<sup>-6</sup>)未満」という境界はそれに続くものであり、出現率の推移に傾向のあるものと、ないものとが混在している。その「出現率が0.5(×10<sup>-6</sup>)未満」の値を含んで推移するもののうち、特徴的な突出が認められず、また、増減の推移に一定した傾向が認められないものを非特異型に分類した。つまり、この分類は、言うなれば「その他」である。この非特異型には表8に示すとおり、31語を分類した。31語というのは少ない数ではない。非特異型が31語もあったという点は、分かりにくい外来語を調査対象にした今回の分類結果の特徴の一つであると言えよう。非特異型の分類例として、図15に「イノベーション」を、図16に「デフォルト」を示す。

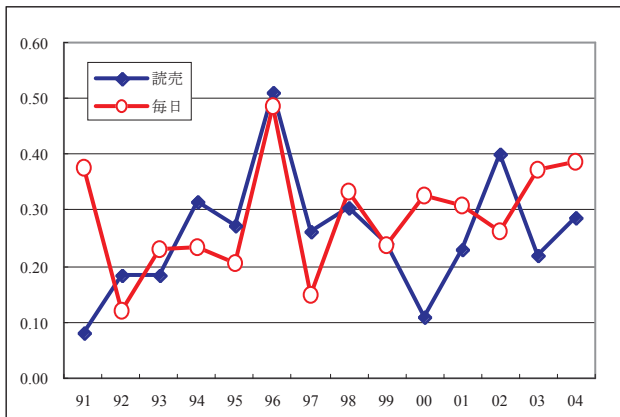


図15 「イノベーション」出現率(×10<sup>-6</sup>)

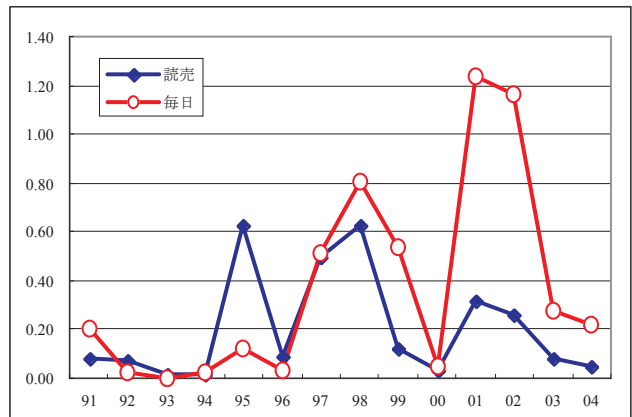


図16 「デフォルト」出現率(×10<sup>-6</sup>)

表8 非特異型に分類した語

見出し	両紙一致の分類		両紙不一致の分類		総度数	
	読売・毎日	読売	毎日	読売	毎日	
ワーキンググループ	—	山型	非特異型	1095	528	
リーフレット	—	山型	非特異型	1027	208	
オペレーション	—	非特異型	山型	837	568	
マンパワー	—	非特異型	突出型	609	309	
メンタルヘルス	—	突出型	非特異型	566	219	
デポジット	—	山型	非特異型	467	201	
インセンティブ	—	非特異型	非特異型	435	301	
ライフサイクル	—	減少型	非特異型	396	200	
イノベーション	非特異型			393	210	
ノーマライゼーション	—	非特異型	突出型	390	209	
コージェネレーション	非特異型			385	178	
スケールメリット	—	非特異型	山型	361	194	
ポートフォリオ	—	突出型	非特異型	360	191	
シーズ	—	非特異型	非特異型	342	80	
スクーリング	—	非特異型	山型	329	149	
アイドリングストップ	—	山型	非特異型	299	81	
デフォルト	非特異型			288	291	
ソフトランディング	—	非特異型	突出型	247	267	
ポテンシャル	非特異型			237	181	
デリバリー	非特異型			236	112	
ステレオタイプ	—	減少型	非特異型	226	201	
エンパワーメント	—	増加型	非特異型	209	108	
タイムラグ	—	山型	非特異型	202	140	
コンファレンス	非特異型			170	82	
センサス	非特異型			163	66	
フェロウシップ	非特異型			157	72	
メディカルチェック	—	非特異型	突出型	123	73	
タスク	—	非特異型	非特異型	105	73	
デジタルデバイド	—	非特異型	突出型	89	201	
ブレイクスルー	非特異型			85	63	
プライオリティー	非特異型			79	65	

## 7 新聞間の分類の一致と不一致

最後に、読売新聞と毎日新聞とで分類が一致しなかった50語について説明する。不一致の原因を調べたところ、新聞個別の事情が原因であることがわかった。その例として、「ポートフォリオ」を図17に、「インタラクティブ」を図18に示す。「ポートフォリオ」は、読売新聞にのみ、1996年を中心とした突出がある。読売新聞には、この年を中心に前後3年にわたり、「ニッセイ投資顧問チーフポートフォリオマネージャー」の連載コラムがあった。そのためこのように毎日新聞とは異なる型を示していた。「インタラクティブ」は、読売新聞が低く平らに推移しているのに対し、毎日新聞が急激に右上がりに推移する増加型を示している。これは、1999年より毎日新聞の新しいホームページが「毎日インタラクティブ」という名前になり、紙面におけるその使用が一気に増加したことによるものであった。

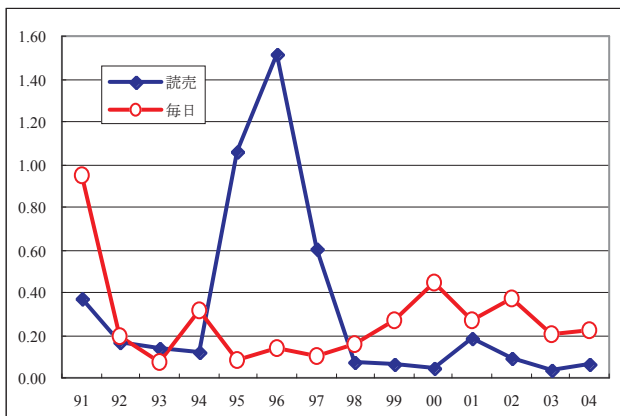


図17 「ポートフォリオ」出現率(×10<sup>-6</sup>)

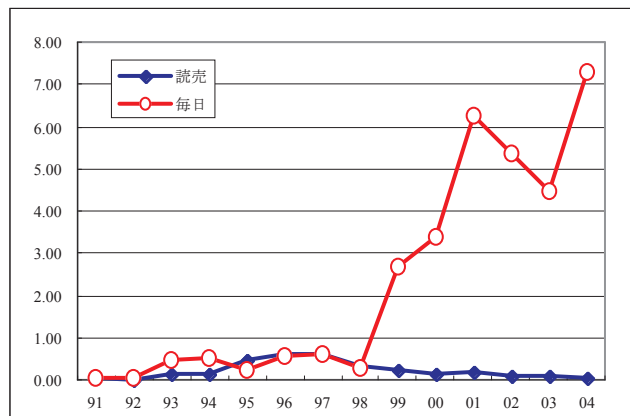


図18 「インタラクティブ」出現率(×10<sup>-6</sup>)

このように、一方の新聞に何らかの特別な事情がある場合、2紙の分類が不一致となっていた<sup>7</sup>。そのようなものが50語、全体の約3割あった。逆に、106語、全体の約7割は分類が一致した。つまり、約3割程度は新聞固有の偏りがあり、逆に約7割は新聞というメディアにおける使用状況と言える、という結果である。これらの数字が大きい小さいかは人によって判断が分かれるかもしれない。別の新聞を調べれば、不一致の割合がさらに増えることも十分予想される。しかし、今回、7割程度は新聞固有の偏りのない結果が得られるという一つの結果が確認できたことの意義は大きいと考える。

## 8 まとめ

読売新聞と毎日新聞の1991年から2004年までの14年分の記事データを用いて、出現率の増減傾向をとらえる頻度推移類型を提案し、「外来語」言い換え提案の対象となった176語のうち、低頻度語を除いた156語について分析した結果を報告した。

頻度推移類型は合計6分類とした。出現率の相対的な推移を語の生命のサイクルとしてとらえる、増加型、水平型、減少型、山型、水平型、非特異型の5つに、生命サイクルとは別軸に区別する突出型を加えた分類である。外来語156語を分析した結果、山型を示す語が65語もあり、もっとも多かった。山型というのは、増加の局面と減少の局面とを短期間に見せるという点において特徴のある型であり、比較的短命な語が示しやすい型である。今回の調査対象語にはそういった短命傾向を示す語が多かったと言える。

さらに、今回、読売新聞と毎日新聞の2紙を調査し、調査をした156語のうち、106語の分類が両紙で一致し、残りの50語の分類が一致しなかったという結果を得た。ここから、約3割程度は新聞による偏りがあること、逆に約7割は新聞というメディアにおける使用状況の確認が得られるということが分かった。

本稿では、頻度推移類型に基づく分析方法が、語の使用状況や、その語の特徴をとらえる上で有効であることを示した。この手法は、ほかの外来語、あるいは、外来語以外の語にも適用できるはずである。今後、対象語をさらに広げて調査を進めていきたい。また、分類の妥当性、客観性についてもさらに検討を進めたい。

## 注

- 1 読売新聞社「読売新聞記事データ集」(日外アソシエーツ)として1987年版より市販されている(<http://www.nichigai.co.jp/sales/corpus.html>)。本調査では、2002年～2004年のみ市販のものを使用し、2001年までは読売新聞社から「外来語」言い換え提案を行った委員会の調査用に提供を受けた記事データを使用した。そのため、2004年までの期間は、市販されているものよりも著作権関係での排除記事が少ないものを使用した調査結果になっている。
- 2 毎日新聞社「CD-毎日新聞データ集」(日外アソシエーツ)として1991年版より市販されている(<http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html>)。本調査でもこれを使用した。
- 3 この検索件数は柏野・他(2005)によるため、読売新聞1986～2001年のものである。
- 4 柏野・他(2005)では大きく4分類を設けて分析したが、分類名、判断基準を再検討し、新たに6分類とした。
- 5 度数の多い順は読売新聞の度数によるものとする。以下、同様。
- 6 福祉関連の「ケア」「デイサービス」も2000年に右下がりの減少に転じており、「バリアフリー」と同じく前回の増加型から今回、山型に分類しなおしている。
- 7 不一致の50語すべてについての原因分析結果は、別途報告する予定である。

## 参考文献

- Akiko Aizawa (2003) Analysis of Source Identified Text Corpora: Exploring the Statistics of Reused Text and the Authorship, *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (ACL 2003), 383-390.
- N T T コミュニケーション科学基礎研究所 (2000) 『日本語の語彙特性 第Ⅱ期』三省堂書店
- 柏野和佳子・山口昌也・桐生りか・田中牧郎 (2005) 「新聞記事データベースを利用した外来語の出現率の推移調査」『自然言語処理』12 (4), 97-116
- 国立国語研究所 (1962) 『現代雑誌九十種の用語用字 第1分冊 総記・語彙表』秀英出版
- 国立国語研究所 (1964) 『現代雑誌九十種の用語用字 第3分冊 分析』秀英出版
- 国立国語研究所 (1997) 『国定読本用語総覧 CD-ROM版』三省堂書店
- 佐藤宏秀 (1991) 「H I A S K (ハイアスク) 一朝日新聞社」『日本語学』10(8), 95-108, 明治書院
- 関根健一 (2003) 「新聞記事の中のカタカナ語」『日本語学』22(8), 30-39, 明治書院
- 野崎浩成・久野雅樹・横山詔一 (1998) 「新聞電子メディアを用いたカタカナ語の使用実態調査—朝日新聞における月次変化の分析—」『日本行動計量学会第26回大会発表論文抄録集』77-78
- 久野雅樹・野崎浩成・横山詔一 (1998) 「カタカナ語使用の季節変動に関する分析」『日本行動計量学会第26回大会発表論文抄録集』79-80
- 久野雅樹 (2000) 「新聞の用字の面による変動と時系列変動」『自然言語処理』7(2), 45-61
- 久野雅樹・横山詔一・野崎浩成 (2002) 「毎日新聞と朝日新聞における文字使用の差」『計量国語学』23(6), 277-295
- 山口昌也・茂木俊伸・桐生りか・田中牧郎 (2004) 「語種との関係に基づいた新聞記事における語彙の時間的変化分析」『第13回社会言語科学大会発表論文集』113-116, 山口昌也「新聞記事における語彙の時間的変化分析—語種の観点から—」として本書第3部第4章に収録
- 山崎誠・小沼悦 (2004) 「現代雑誌における語種構成」『言語処理学会第10回年次大会発表論文集』670-673

## 謝辞

使用度数調査、並びに、集計表とグラフの作成といった本研究の調査データの作成は、研究補佐員の西部みちるさんの尽力によるところが大きい。ここに記して感謝する。