

# 国立国語研究所学術情報リポジトリ

現代書き言葉における外来語の共時的分布：  
「ケース」を事例として

メタデータ	言語: Japanese 出版者: 公開日: 2015-10-30 キーワード (Ja): キーワード (En): loanwords, lexical variation, language-external factors, corpus 作成者: 久屋, 愛実, KUYA, Aimi メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/00000511">https://doi.org/10.15084/00000511</a>

## 現代書き言葉における外来語の共時的分布： 「ケース」を事例として

久屋 愛実

オックスフォード大学 博士課程／国立国語研究所 外来研究員 [–2013.02]

### 要旨

本稿の目的は、外来語「ケース」を事例としてその共時的分布を明らかにし、外来語使用における言語外的要因を特定することである。本稿では国立国語研究所が2011年に公開した『現代日本語書き言葉均衡コーパス』(BCCWJ)を利用し、外来語と既存語の語彙バリエーションの研究を経験的に行う。分析の結果、書き手の生年代、最終学歴、媒体種、共起述語の種類について統計的有意差が認められ、これらの要因が外来語の生起に影響していることがわかった。若年層ほど「ケース」の生起率が上昇していることから、既存語（「事例、場合、例」）から外来語への語彙使用の変化（言語変化）が起こっていることが確認された。また、外来語の生起率は書き手の学歴の上昇に伴い低下し、より公共性の高い媒体として特徴づけられる書籍において低いことがわかった。このことは、ある社会的カテゴリーや環境ではまだ浸透度が低いという、日本語における外来語の現状をあらわしていると思われる\*。

**キーワード：**外来語、語彙バリエーション、言語外的要因、コーパス

### 1. はじめに

本研究の目的は、「ケース」を事例として、同様の意味をもつ和語や漢語（以下、単に「既存語」）の存在にもかかわらず使用の増加を見せている外来語<sup>1</sup>の共時的分布をコーパスデータに基づいて明らかにすることである。名詞系外来語が既存語の存在にもかかわらず「基本語化」した言語内的要因については金（2011）で明らかにされている<sup>2</sup>。一方、本研究は、言語外的側面から外

\* 本稿は第3回コーパス日本語学ワークショップ(2013年2月28日、国立国語研究所)で発表した論文を加筆・修正したものである。本稿で分析したデータは、筆者が2012年11月から2013年2月までの期間中に国立国語研究所に外来研究員として滞在していた際、BCCWJから収集したものである。なお、一部BCCWJに収録されていない属性データについては、筆者が研究所の許可を得て収集・利用した。滞在接受入れて頂いた同研究所所長の影山太郎先生、言語資源研究系・系長の前川喜久雄先生、受け入れ教官として滞在中様々な面でご指導ご鞭撻頂いた田中牧郎先生に感謝の意を表したい。なお、データ抽出に関しては、同研究所コーパス開発センターの中村荘範さん（マンパワージャパン株式会社）に技術的なご指導・ご支援を頂いた。また、本稿の準備段階で貴重なコメントをして頂いた同研究所研究員の金愛蘭さんと南部智史さん、上記ワークショップでコメント頂いた方々にもこの場を借りてお礼申し上げる。

<sup>1</sup> Myers-Scotton (2006) は、*Cultural borrowing* と *Core borrowing* という借用語の概念区分を示している。前者は受け入れ言語側の既存語彙では言い表すことができないような事物や概念を表現するために借用され、科学や技術の分野でその多くの例を見ることができる。それに対して、後者は受け入れ言語側の語彙に類似の表現が存在するにもかかわらず、借用されるものである。本稿では、日本語におけるこうした外来語のタイプの違いを認識したうえで、後者、つまり既存語をもつ外来語に焦点をあてて調査することを前提としている。

<sup>2</sup> 金によれば、「基本語化」とは、当該語が日本語語彙の周辺部から中心部へと移行して基本語彙（一定の言語使用域において広範囲・高頻度に用いられる語彙）へと仲間入りすることである。金は既存語をもつ2つの外来語「トラブル」と「ケース」を挙げ、両者が新聞における基本語としての地位をどのように確立していったのかを、意味・用法の側面から既存語と比較しながら通時的に分析している。

来語増加のメカニズムを説明し、日本語における外来語の役割をとらえ直すことに主な重点を置く。

金 (2011) は、20 世紀後半の通時的新聞コーパスにおいて、「ケース」が一貫した増加を見せていることを明らかにした (図 1)。そこで、本稿では、国立国語研究所が 2011 年に公開した『現代日本語書き言葉均衡コーパス』(BCCWJ)<sup>3</sup>を用いて、図 1 で生起率のピークを迎えている 2000 年以降 (2001-2005 年) における「ケース」の共時的分布について調査する。具体的には、使用者の属性や媒体種などの言語外的要因の影響について、クロス表を用いた分析を行う。また、先行研究 (金 2011) で取り上げられた、語彙が出現する文法形式やコロケーションなどの言語内的要因の影響についても検証する。

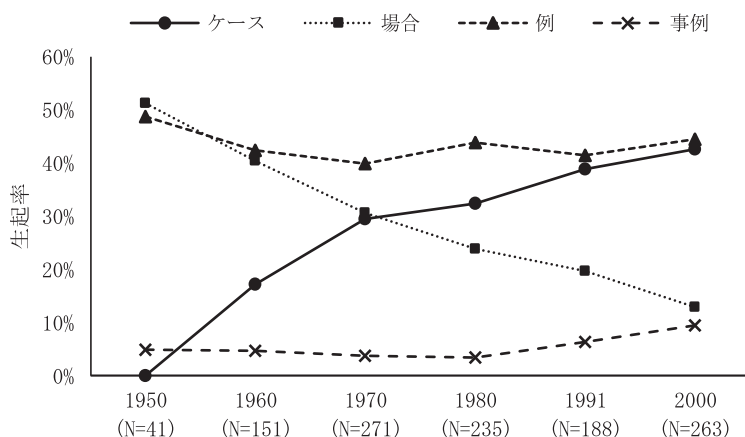


図 1 新聞の発行年別に見た「ケース」と類義語の生起率の変動<sup>4</sup> (金 2011: 108)

外来語に関するこれまでの大規模な社会言語学的研究は、意識調査にとどまることが多かった。その原因は、外来語という出現頻度の低い語種を使用者の属性情報とともに効率よく収集できる大規模コーパスの不在にあったのではないと思われる。本稿では、BCCWJ を利用することで、外来語のような低頻度の語種でも一定量のサンプルを得ることができ、同時にサンプルのタグから書き手の属性情報 (生年代、性別など) も抽出することができた。これにより、実際の言語資料に基づいた外来語の社会言語学的研究を試みている。こうした研究により、先行研究とあわせてより幅広い視野で外来語をとらえることが可能になるとと思われる。

<sup>3</sup>『現代日本語書き言葉均衡コーパス』の概要、利用の手引等は、[http://www.ninjal.ac.jp/corpus\\_center/bccwj/](http://www.ninjal.ac.jp/corpus_center/bccwj/)を参照。

<sup>4</sup> グラフは、金 (2011: 107) の表 1 a 「ケースと類義語の出現度数」の数値を、筆者が 4 語の出現度数の合計を 100 としたときの生起率に置き換えたものである。よって金の出現率 (100 万字あたり) のグラフとは若干異なる。

## 2. 方法論

### 2.1 バリエーション研究のアプローチ

本稿では、バリエーション研究の手法を語彙レベルで適用し、外来語を「特定の文脈において既存語と交替可能な社会言語学的語彙変異形 (sociolinguistic lexical variant)」として扱う。ただし、語彙をバリエーション研究の枠組みで扱うことは、そう単純なことではない。その理由は、語彙には意味というものが介在してくるところにある。そもそも、ある要素をバリエーションとして扱うことができるのは、変異形の間で交替が起こっても意味の変化が生じないという条件を満たす場合である。その典型的な例は音韻交替であるが、語彙交替の場合は、それぞれの語彙が複数の意味や慣用的表現をもち、意味を共有しないことがあるため、予備段階として変異形の間で意味の変化が生じないような交替環境を特定することが必要である。そこで本稿では、辞書的な類義性をベースとして、それぞれの類義語が文章中で同じような機能を果たす場合にそれらを変異形と認めることとする。さらに、文法構造やコロケーションの観点から分析対象を絞って変異形の等価性を保つことを試みる。

### 2.2 分析対象となる環境

先行研究（金 2011）から、「ケース」の類義語として「場合」「例」「事例」の3語を採用した。以下、「ケース」とこれら既存語の交替が可能な環境を定義していく。

金（2011）は、「ケース」がいわゆる「コト」に代表される形式名詞的な用法をもつことを指摘し、その中でも、叙述文が表現する内容を客観的な事柄として名詞化する「客観的同格連体名詞」としての用法を手がかりに、これら4語の類義関係を認めている。そこで、本稿ではまず、4語が機能的等価性をもつのは、文中で形式名詞として「ある内容を客観的な事柄として名詞化する」とき、と定義する。これにより、金の言う(1)「場合」の「仮定条件」的用法、(2)「提題」的用法、(3)「例」の単なる「例示」的用法、(4)各語に特有な慣用的用法は分析の対象外となる（下記の例はいずれも金（2011: 94-95）より引用）。

- (1) 賃上げが難しい場合は雇用延長など別のテーマで交渉する“選択”の時代になったと問題提起している。
- (2) アサガオの場合、
- (3) 例として、～を例に挙げ、
- (4) ケース・バイ・ケース、場合によっては、そんなこと言ってる場合かつ、例の悪名高い作家、例によって無言で打つ

形式名詞はふつう修飾語をとるが、金（2011）によれば、これら4語は「コト」よりもやや具体性のある名詞であるため、とくに修飾語をとらなくてもよいという。これら4語がとりうる形式は4つあり、(5)修飾部をとらず単独で、(6)合成語の構成要素として、(7)名詞句における被修飾語として、(8)連体修飾節構造における被修飾語として文中に現れうる。以下、それぞれ代表として「ケース」の用例を挙げる（それぞれ金（2011: 96, 96-97, 98-99, 101）より引用）。

- (5) しかし、女性の平均賃金は男性より低いいため、男女の賠償額にケースによっては 1000 万円近い差が生じている。
- (6) ケース 1, テストケース, レアケース, 重症ケース, 脳死・虐待ケース
- (7) いじめなどのケース, マドンナさんのケース, 京都市のケース, 今回のようなケース, 4 件のケース, 悪質なケース, 初めてのケース, 似たケース, いろんなケース
- (8) a. ネット先進国の米国でも、ネット関連企業は苦戦するケースが少なくない。  
 b. (略) まず母親に『癒(いや)し』が必要なケースも多い」と話す。  
 c. (略) しつけの域を超えて繰り返される暴力・ネグレクトが原因のケースに絞って調べた。

ここで、20 世紀後半の新聞コーパスにおける「ケース」の生起率を形式ごとに通時的(発行年別)に概観する。図 2～5 は、金(2011: 109-111 表 2a-4a)の調査結果(各語の出現度数)を筆者が全語彙の出現度数の合計を 100 としたときの各語の生起率(%)に置き換えたものである<sup>5</sup>。図 2～5 を概観すると、4 つの形式のうち、20 世紀後半の新聞コーパスにおいて「ケース」の生起率が一貫した増加を見せているのは例文(7)(8)でそれぞれ見た名詞句構造(図 4)と連体修飾節構造(図 5)においてであることがわかる。以上のことをふまえて、本稿では名詞句または連体修飾節構造において出現している「ケース」およびその既存語のみを分析対象とした。

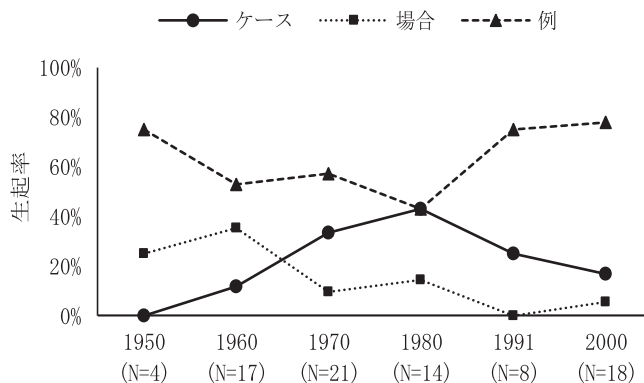


図 2 単独用法

<sup>5</sup>「事例」の発行年ごとの出現率は用例数が少ないことから金(2011)では省略されているため、ここでも同じく省略されている。

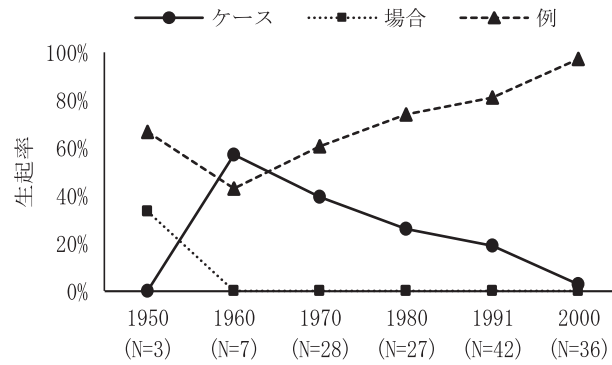


図3 合成語の要素としての用法

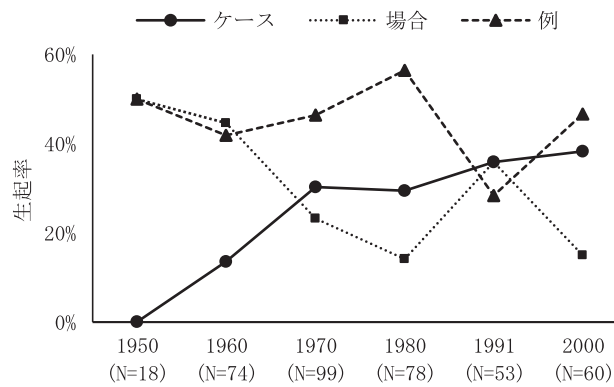


図4 名詞句用法

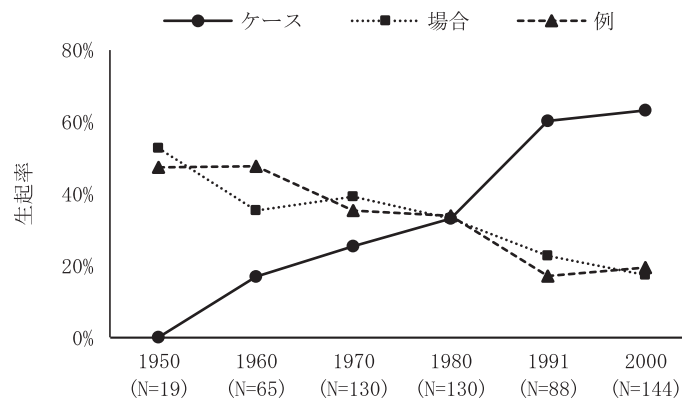


図5 連体修飾節用法

さらに、コロケーションという観点からも4語の等価性を保つために、後続する格と述語の種類を絞った。金(2011)は、連体修飾節構造において「ケース」が多少・有無・生起・増減・想定・報告・限定・異同・規定・関与・比較の意味をもつ述語表現と結びつき、このうち最も出現頻度が高いのは有無・多少の意味をもつ述語であることを指摘した。ここで、有無・多少を含むこれらの多くがガ格と結びつく述語であることに注目し、分析対象を、ガ格をとまなう多少・有無・生起・増減・想定・報告の意味をもつ述語と共起しているものに限定した(共起述語の意味分類ごとの具体例の一部を(9)に挙げる)。つまり、前述の(8c)のように二格などをとまなう例は、分析対象とならない。ただし、ガ格は(8b)のように助詞のハ・モとしても頻繁に現れうるため、これら2つの助詞が後続する場合も含めた。これで、(10)の通り、機能・構造(形式)・コロケーションの3つの側面からコントロールされた語彙交替環境が特定できた。実際に分析対象となったサンプルの具体例を、共起述語の意味分類とともに(11)に示す。

(9) 共起述語の意味分類ごとの具体例(金(2011)より引用)

- 多少(多い・少ない・ほとんどだ・珍しい)
- 有無(ある・ない・見られる・認められない)
- 生起(起きる)
- 増減(増える・減る)
- 想定(想定される・考えられる・予想される)
- 報告(挙げられる・紹介される・報じられる)

(10) 分析対象となる語彙交替環境のモデル

- a. 機能：形式名詞として「ある内容を客観的事柄として名詞化する」とき
- b. 構造(形式)：名詞句または連体修飾節
- c. コロケーション：{ケース／場合／例／事例} + {ガ格／ハ／モ} + {多少／有無／生起／増減／想定／報告の意味をもつ述語}

(11) 分析対象となったサンプルの例(「ケース」, 連体修飾節の場合)

( ) 内は BCCWJ のサンプル ID

- a. また回線料も現状の専用線や VAN と比べると安くできるケースが多い。[多少] (PB10\_00063)
- b. 手順7の写真の方が自然なケースもありますので、[有無] (PB10\_00093)
- c. 最近では、こうした危機管理以前の「安全管理」に問題あるケースが続発している。[生起] (PB13\_00434)
- d. 最近では地方銀行が「株主優遇定期作成優待券」を贈呈したり、地元の名産品を贈るケースも増えています。[増減] (PB23\_00254)
- e. 銀行で借り直すことになるケースも考えられる [想定] (PB13\_00381)
- f. ネグレクトが疑われるというケースも報告されている。[報告] (PB13\_00249)



## 2.3 『現代日本語書き言葉均衡コーパス』(BCCWJ)

本稿で用いるのは、国立国語研究所が2011年に公開した『現代日本語書き言葉均衡コーパス』(以下、BCCWJ)の出版サブコーパスである。出版サブコーパスは、BCCWJの3つのサブコーパス(出版サブコーパス、図書館サブコーパス、特定目的サブコーパス)のうちのひとつであり、2001年から2005年までの間に国内で出版された書籍・雑誌・新聞を母集団とし、そこからランダムにサンプリングされたデータおよそ3,600万語(短単位)からなるコーパスである。このうち、書籍は2,954万語、雑誌は569万語、新聞は88万語のデータからなる(山崎他2012)。ただし、新聞データは書籍と雑誌に比べて極端に規模が小さいため、今回は分析の対象外とした。また、データには固定長と可変長があるが、なるべくたくさんのデータから調べるために分量の多い可変長データ(書籍と雑誌あわせておよそ3,500万語)を利用した。この規模であれば、他の語種より相対的に出現頻度が低い外来語でも、定量的調査に耐えうるサンプル量を抽出できると思われる。また、全てではないが、出版サブコーパスのサンプルには年齢や性別など、社会言語学的調査に必要な書き手の属性情報がタグ付けされていることも、本コーパスを利用した理由のひとつである。

ちなみに、今回言語外的要因として想定するのは、書き手の生年、性別、最終学歴、海外滞在歴、媒体種、ジャンル、販売対象(書籍のみ)などであるが、このうち、書き手の生年、性別、媒体種、ジャンル、販売対象についてはBCCWJに収録されているタグ情報を利用する。一方、書き手の最終学歴と海外滞在歴についてはBCCWJには収録されておらず、国立国語研究所により2006年から2010年にかけて行われた著者へのアンケート調査の情報を、同研究所の許可を得て利用した。

## 2.4 抽出サンプルの概要

分析対象となるサンプルを抽出した結果、2.2節(10)で設定した条件にあてはまるサンプルは、「ケース」363件(20.2%)、「事例」92件(5.1%)、「場合」1010件(56.2%)、「例」331件(18.4%)の合計1796件であった(表1)。なお、今回抽出したのは、著者が特定できており、著者の生年・性別・最終学歴の属性タグ全てが揃っているサンプルだけである。また、書き手自身の言葉のみを分析対象とするために、引用タグを利用して引用文も除外している。

表1 分析対象となる4語のサンプル数と生起率<sup>6</sup>

	ケース	事例	場合	例	合計
度数	363	92	1010	331	1796
%	20.2%	5.1%	56.2%	18.4%	100.0%

ここで、1796件のサンプルの質を把握するために、書き手の属性別にデータを概観する(表2)。

<sup>6</sup> 久屋(2013)では、学歴情報の利用と引用文の除外を行っていないため、今回のサンプル抽出結果とは数値が異なる。



BCCWJにおける著者の生年情報は10年刻みで公表されており、著者の生年は1880年代<sup>7</sup>から1970年代と、ちょうど100年間にまたがる。ただし、データのおよそ95%は1950年代をピークに1920年代から1960年代生まれの書き手によるものに集中している。性別は男女の割合に偏りがあり、女性の書き手によるサンプルは全体の1割に満たないことがわかる。最終学歴については、大学院（修士）卒以上の書き手によるサンプルが48.8%とほぼ半数にのぼり、大学卒の書き手によるサンプル（46.3%）とあわせると全体の約95%を占めている。媒体種は、もともとのコーパス規模の違いもあり、書籍からのサンプルがほとんど（95.7%）を占める。雑誌は4語すべての用例数をあわせても77例（4.3%）と少ない。

表2 サンプルにおける書き手の属性分布

生年代	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	合計
度数	1	2	4	32	144	283	432	539	310	49	1796
%	0.1%	0.1%	0.2%	1.8%	8.0%	15.8%	24.1%	30.0%	17.3%	2.7%	100.0%
性別	女	男									合計
度数	164	1632									1796
%	9.1%	90.9%									100.0%
最終学歴	中学	高校	高専	短大	大学	大学院					合計
度数	13	58	9	8	832	876					1796
%	0.7%	3.2%	0.5%	0.4%	46.3%	48.8%					100.0%
職業	A 管理	B 専門・技術	D 販売	E サービス	F 保安	G 農林漁業	その他	無回答	無職	合計	
度数	100	1429	6	2	2	1	144	78	34	1796	
%	5.6%	79.6%	0.3%	0.1%	0.1%	0.1%	8.0%	4.3%	1.9%	100.0%	
媒体種	書籍	雑誌									合計
度数	1719	77									1796
%	95.7%	4.3%									100.0%

最後に著者の職業について言及しておく。職業情報は、最終学歴などと同様にBCCWJには収録されておらず、国立国語研究所によって実施された著者へのアンケート調査によるものであり、これも研究所の許可を得て使用した。職業に関する回答は、自由記述のため分類しづらいものが多かったことから以下分析には使わないが、データの質を概観するには有用である。回答は、『日本標準職業分類（平成21年12月統計基準設定）』（総務省統計局2009）<sup>8</sup>の大分類項目に基づき分類を試みた<sup>9</sup>。1796のサンプルのうち、その他（8.0%）、無回答（4.3%）、無職（定年退職がほとんど）（1.9%）を除くと、ほとんどが「A 管理的職業従事者」（5.6%）と「B 専門的・技術的職

<sup>7</sup> 日本では明治維新からすでに英語との言語接触が始まっており、この頃から英語教育が盛んになり、ピジンの存在もあったという（Irwin 2011）。このことから、1880年代に生まれた著者も外来語使用の可能性があると考え、本稿ではデータに含めることとした。ちなみに、出版コーパスには再版本も含まれ、特に生年代の早い著者による出版物は再版本である。

<sup>8</sup> 総務省統計局ホームページから閲覧可能（<http://www.stat.go.jp>）。

<sup>9</sup> 複数記述のものは一番目にあるもの、時系列で説明があるものは最新のものを選び、退職して当時無職であっても過去の職業について記述がある場合はその職業を採用した。自由記述のため分類の際、判断がつきにくいもの（単に「公務員」、「会社員」とあるものなど）は「その他」扱いにした。

業従事者」(79.6%)といった、いわゆる「ノンマニュアル」層によるものであることがわかる。「A 管理的職業従事者」は国・地方公共団体、会社、団体におけるいわゆる「管理職」で、「B 専門的・技術的職業従事者」は専門的技術を要する様々な分野の職種を指す。「B 専門的・技術的職業従事者」では、教員、著述家・記者・編集者、保健医療従事者、法務従事者の順で多く、それぞれデータ全体の 42.8%, 13.6%, 8.9%, 4.3% を占める。さらに、教員、保健医療従事者、法務従事者のカテゴリーにおいては、大学教員、医師、弁護士がそれぞれ圧倒的多数（いずれも各カテゴリーの約 9 割）を占めており、A の管理職とあわせても書き手の職業は一部のカテゴリーに集中していることがわかる。

### 3. 要因ごとの分析結果（言語外的要因）

以下、外来語「ケース」の出現に影響を与えると想定される言語外的要因について、クロス表を用いた分析を行う。なお、本稿の目的は外来語が出現する要因を検証することなので、分析をしやすくするために「場合・例・事例」の 3 語をまとめて「既存語」とし、クロス表分析の際には「ケース」対「既存語」という単純な 2 項対立で見ていく。

#### 3.1 書き手の生年と言語変化

金（2011）の研究により、20 世紀後半の新聞における「ケース」の生起率が、特に連体修飾節・名詞句構造において大きく増加していることがわかっている（2.2 節、図 2～5）。この増加が起こった理由のひとつとしては、既存語から外来語への語彙使用の変化（言語変化）が起こっている可能性が挙げられる。だとすれば、「ケース」の生起率に、書き手の生年による段階的な差が存在するのではないか、という予測がたつ。類義語のうち外来語を最も使いたいという意識に年齢による段階的な差があったことは田中（2007）でも指摘されているが、本稿では意識ではなく実際の言語使用においてこうした差を観察することができるかどうか検証を行う。

以上をふまえて、収集した 1796 件のサンプルについて、書き手の生年という観点から外来語「ケース」と既存語の生起率を整理し、生年による段階的な差が認められるかを検証する。図 6 より、書き手の生年代が上がるにつれて「ケース」の生起率が次第に上昇していることがわかる。「ケース」は 1910 年代生まれの書き手によって初めて使用され、1920 年代生まれで大幅な増加を見せた後横ばいになるが、1960 年代生まれで再度大幅な増加を見せている。「例」は 1910 年代生まれ以降減少を続け、1940 年代生まれの書き手で「ケース」の生起率に追い抜かれている。「場合」は世代によるゆれはあるものの常に 4 語の中で最も高い生起率を保っており、「事例」は常に最も低い生起率で全体を通してあまり変動がない。

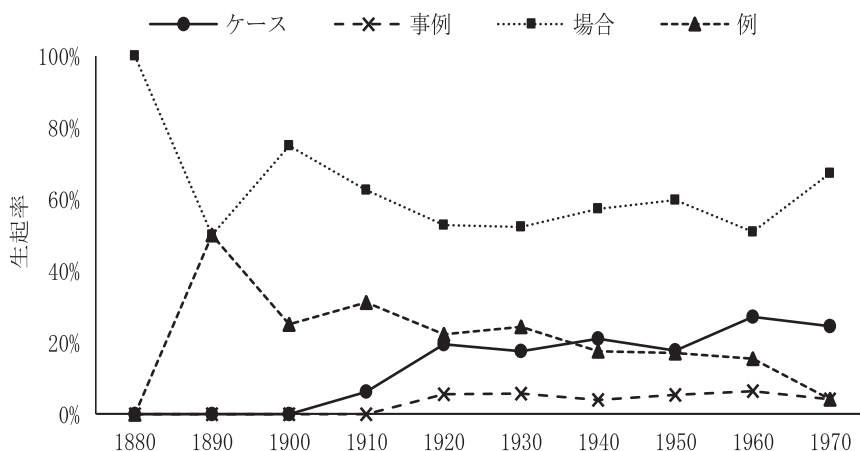


図6 書き手の生年代別に見た「ケース／事例／場合／例」の生起率

ここで、便宜的に書き手を生年代が最も早いグループ（1880–1930年代）、中間のグループ（1940–1950年代）、そして最も遅いグループ（1960–1970年代）の3つに区分し、カテゴリーごとに「ケース」と既存語の生起率をしてみる（表3）。この結果、生年が上がる（世代が若くなる）につれて、17.2%、19.3%、26.7%と、既存語群に替わって外来語がより用いられており、生年による語彙使用の見かけ上の変化（change in apparent time）が認められる。カイ2乗検定<sup>10</sup>の結果、「ケース」の生起率は生年代と関連があることが統計的に認められ（ $\chi^2=12.72$ , d.f.=2,  $p<0.01$ ）、外来語志向には年齢による段階的な差があるという意識調査の結果（田中 2007）と一致した。図6や表3の分布から総合的に判断すると、これは特に若い年代（1960–1970年代生まれ）が「ケース」の生起率を上げていることに起因すると思われる。

表3 生年代の違いと「ケース／既存語」の生起率

		1880–1930	1940–1950	1960–1970	合計
ケース	度数	80	187	96	363
	%	17.2%	19.3%	26.7%	20.2%
既存語	度数	386	784	263	1433
	%	82.8%	80.7%	73.3%	79.8%
合計	度数	466	971	359	1796
	%	100.0%	100.0%	100.0%	100.0%

$$\chi^2=12.72, \text{ d.f.}=2, p<0.01$$

### 3.2 書き手の性別

バリエーション研究において、女性のほうが変化をリードするということがよく言われる。これを今回の調査語である「ケース」にあてはめると、女性のほうが外来語をより使うことが予測

<sup>10</sup> 検定には SPSS ver. 21 を使用した。

される<sup>11</sup>。そこで、書き手の性別という観点から「ケース」と既存語の生起率を見ていく。表4から、男性（20.1%）よりも女性グループ（21.3%）において「ケース」の生起率が若干高いことが読み取れるものの、これは統計的に有意な差ではなかった。これで、外来語「ケース」の生起率において性差は認められないことがわかった。また、図7は性別ごとの既存語の内訳を示したもののだが、ここでも3語の全体的な分布パターンは男女であり差が見られない。

表4 性別と「ケース／既存語」の生起率

		女	男	合計
ケース	度数	35	328	363
	%	21.3%	20.1%	20.2%
既存語	度数	129	1304	1433
	%	78.7%	79.9%	79.8%
合計	度数	164	1632	1796
	%	100.0%	100.0%	100.0%

$$\chi^2=0.14, \text{ d.f.}=1, p=0.71$$

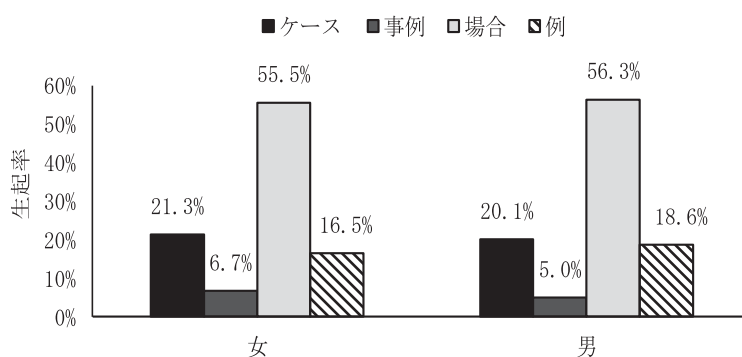


図7 性別ごとに見た「ケース／事例／場合／例」の生起率

### 3.3 書き手の最終学歴

次に、書き手の最終学歴を、中学—短大（中学，高校，高専，短大），大学，大学院（修士卒以上）の3つに区分し、カテゴリーごとに「ケース」と既存語の生起率を見てみる。表5から、書き手の最終学歴が上昇するにつれて外来語の生起率が28.4%，23.6%，16.2%と段階的に低下しており、これは統計的有意差が認められた（ $\chi^2=18.14, \text{ d.f.}=2, p<0.001$ ）。既存語それぞれの内訳（図8）を見ると、最終学歴が上がるにつれて「場合」「例」の生起率が徐々に上昇しており、特に「例」の生起率は、大学院卒において「ケース」の生起率を上回る。「事例」はすべての学歴カテゴリーで生起率が低く、カテゴリー間でさほど差はないようである。このことから、書き手の最終学歴が高くなるほど、外来語よりも既存語「場合」「例」が好まれ、特に大学院卒では「ケース」よりも「例」のほうが好まれることがわかる。

<sup>11</sup> ファッションや美容，スポーツ関連語など，特定の性別のみが使うことの多い特徴的な語彙ではなく，ニュートラルな語であることが前提である。

表5 最終学歴の違いと「ケース／既存語」の生起率

		中学—短大	大学	大学院	合計
ケース	度数	25	196	142	363
	%	28.4%	23.6%	16.2%	20.2%
既存語	度数	63	636	734	1433
	%	71.6%	76.4%	83.8%	79.8%
合計	度数	88	832	876	1796
	%	100.0%	100.0%	100.0%	100.0%

$$\chi^2=18.14, \text{ d.f.}=2, p<0.001$$

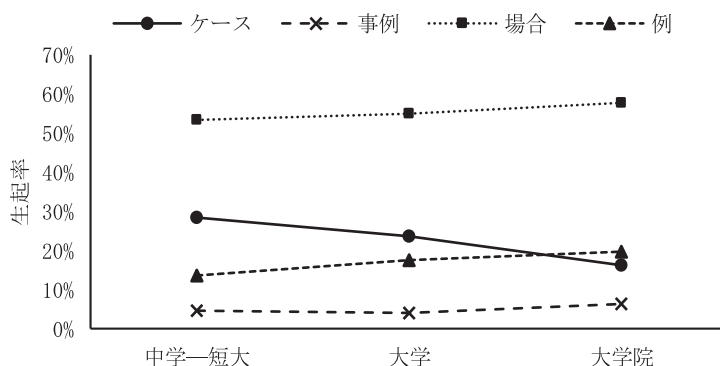


図8 最終学歴別に見た「ケース／事例／場合／例」の生起率

### 3.4 書き手の海外滞在歴（出生地または長期滞在地として）

『外来語に関する意識調査』（国立国語研究所編 2004）のデータから、海外渡航経験のある人はいない人よりも外来語に対して肯定的な意識を示し、外来語や略語の知識や使用も多いと回答していることがわかっている。例えば、表6では、海外渡航経験のある人のうち、42.0%が外来語の増加に対して「好ましい・まあ好ましい」と答えているのに対し、海外渡航経験のない人では34.3%となっている。このことから、海外での滞在歴は外来語の使用率に影響があると予測でき、今回検証することとした。

表6 海外渡航歴と外来語に対する意識（国立国語研究所編 2004）<sup>12</sup>

過去10年の海外渡航歴（旅行／滞在または両方）	ある	ない
回答者数	1137	1935
A) 今以上に外来語が増えることについての意識：「好ましい・まあ好ましい」と答えた割合 (%)	42.0	34.3
B) 外来語や略語の知識：「知っている方だと思う」と答えた割合 (%)	30.1	15.9
C) 外来語や略語の使用：「使う方だと思う」と答えた割合 (%)	21.6	11.9

<sup>12</sup> 回答者の海外渡航歴は『外来語に関する意識調査』（国立国語研究所編 2004）には掲載されていないが、研究所の許可を得て元データである集計表の情報から作成した。

BCCWJ では、書き手の出生地と、出生地以外で長期滞在し言語形成に影響を与えたと感じる場所を著者へのアンケートにより別途調査している。分析では、出生地または言語形成に影響を与えた場所のうち、どちらかの項目に海外の地名の記述があれば「海外滞在歴あり」<sup>13</sup>、どちらにも海外の地名がなければ「海外滞在歴なし」とし、どちらの項目にも記述自体がない8サンプル（未回答）は分析から外した。表7から、「ケース」の生起率は、海外滞在歴はないと答えたグループ（19.8%）よりも海外滞在歴があると答えたグループ（24.4%）で高いものの、これは統計的に有意な差ではなかった。既存語の内訳（図9）を見ると、海外滞在歴がないグループで「場合」の生起率が若干高いものの、あまり目だった差とは言えないようである。

表7 海外滞在歴と「ケース／既存語」の生起率

		ある	ない	合計
ケース	度数	39	323	362
	%	24.4%	19.8%	20.2%
既存語	度数	121	1305	1426
	%	75.6%	80.2%	79.8%
合計	度数	160	1628	1788
	%	100.0%	100.0%	100.0%

$$\chi^2=1.86, \text{ d.f.}=1, p=0.17$$

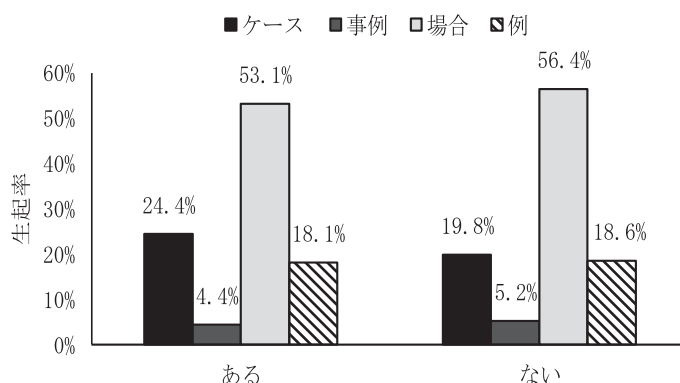


図9 海外滞在歴別に見た「ケース／事例／場合／例」の生起率

### 3.5 媒体種

次に、異なる媒体間で「ケース」の生起率に差があるかについて検討する。表8から、「ケース」の生起率は書籍（19.6%）よりも雑誌（33.8%）においてかなり高いことがわかり、これはカイ2乗検定により有意差が認められた（ $\chi^2=9.17$ , d.f.=1,  $p<0.01$ ）。既存語の内訳（図10）を見ると、特に「場合」の生起率が書籍よりも雑誌において低く、「事例」と「例」の生起率は書籍—雑誌間で大きな違いは見られない。以上のことから、雑誌では「ケース」の生起率が高い反面、既存語のうち「場合」の生起率が低いことがわかる。

<sup>13</sup> 非英語圏を含む。

表 8 媒体の違いと「ケース／既存語」の生起率

		書籍	雑誌	合計
ケース	度数	337	26	363
	%	19.6%	33.8%	20.2%
既存語	度数	1382	51	1433
	%	80.4%	66.2%	79.8%
合計	度数	1719	77	1796
	%	100.0%	100.0%	100.0%

$\chi^2=9.17$ , d.f.=1,  $p<0.01$

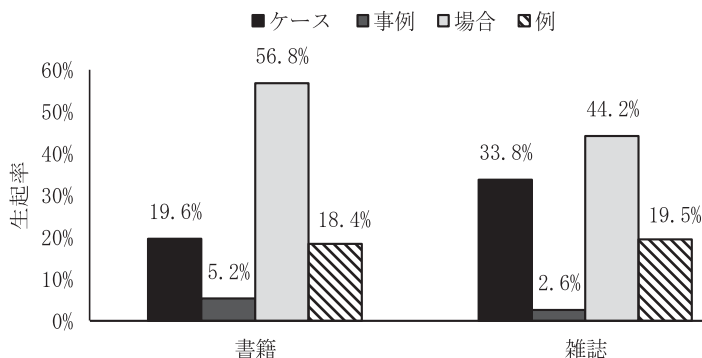


図 10 媒体別に見た「ケース／事例／場合／例」の生起率

では、なぜ媒体間で差が出たのだろうか。考えられる可能性は以下の3つである。第1に、書き手の属性が影響していることが考えられる。まず、カイ2乗検定により「ケース」の生起率との関連が認められた生年代について考えてみる。表9で、書籍と雑誌のそれぞれについて書き手の生年代の分布を見たところ、雑誌のデータでは37.7%が1960年代以降に生まれた若年層の書き手であることがわかる。一方、書籍では若年層の割合が19.2%と、雑誌と比べて18.5ポイントも低い。このことは、若年層ほど外来語の生起率が高いという語彙使用の変化の影響が、書籍よりも雑誌において大きいことをあらわしている。次に、同じくカイ2乗検定により「ケース」の生起率との関連が認められた最終学歴について考えてみる。表10で、書籍と雑誌のそれぞれについて書き手の最終学歴の分布を見たところ、雑誌よりも書籍において中学—短大卒の割合が4.4ポイント低く、逆に大学院卒以上の割合は2.1ポイント高くなっていることがわかった。よって、高学歴層ほど外来語の生起率が低いという影響が、雑誌よりも書籍において大きいように思われるが、その影響は生年代の影響ほど大きくはないと考えられる。

表 9 媒体別に見た書き手の生年代の分布

	1880-1930	1940-1950	1960-1970	合計
書籍	26.7%	54.1%	19.2%	100.0%
雑誌	9.1%	53.2%	37.7%	100.0%



表 10 媒体別に見た書き手の最終学歴の分布

	中学—短大	大学	大学院	合計
書籍	4.7%	46.4%	48.9%	100.0%
雑誌	9.1%	44.2%	46.8%	100.0%

第2に、媒体そのものの特徴が影響している可能性を考えてみる。そもそも書籍と雑誌の媒体としての違いはどこにあるのだろうか。田中（2012）は、外来語の比率がBCCWJの「雑誌」「知恵袋」「ブログ」の3媒体で高く、「図書館書籍」「出版書籍」「新聞」の3媒体で低いことを指摘した。田中は、前者と後者の媒体の違いについて、前者を「私的で日常的な話題」「娯楽性」「仲間内への情報提供」という性質、後者を「公共的で社会的な話題」「まとまった内容を不特定多数の人に伝える」性質と結びつけて説明を試みている。この考えに基づけば、書籍と雑誌の差は「公共性」の違いとしてとらえることができ、外来語の生起率が雑誌という媒体で高いということは、田中の指摘する通り、外来語が公共性の高い場面よりも、より「私的な」場面での語彙として機能していることを示しているのかもしれない<sup>14</sup>。

また、書籍と違い、雑誌は短期間で売り上げを伸ばすために、すぐに読者の目をひくような個性的・魅力的な存在である必要がある。加えて、常に新しい情報を提供していくという特徴があるため、目新しさという側面ももちあわせていなければならない。こういった特徴のために、雑誌という媒体は、「スタイリッシュな・おしゃれな・かっこいい・斬新な・目新しい」というようなイメージと結びつきやすいと思われる。一方、梁（2012）によれば、日本語における外来語という語種のもつプラスイメージとして一番多かったものは、「かっこいい」すなわち「洗練されている」という評価だったという。また、国立国語研究所編（2004）の『外来語に関する意識調査』によると、外来語を使用することの良い点として、「新しさ（28.2%）」、「しゃれた感じ（22.1%）」など、外来語の与えるスタイリッシュな機能が上位に挙がっている（図11）。ここで今一度、雑誌において外来語の生起率が高い理由を考えるならば、雑誌という媒体のもつ「スタイリッシュさ」というイメージが、同じく「かっこよさ」「新しさ」「しゃれた感じ」という「洗練された」イメージをもつ外来語によって体现しやすいことが背景にあると考えることができなだろうか。

<sup>14</sup> 久屋（2013）では、「ケース」の生起率は書籍よりも「知恵袋」で低く、田中（2012）の示した外来語の一般的な傾向と一致しなかった。この点に関しては、両媒体の「公共性」の違いだけではなく、ジャンルや内容にまで踏み込んだ比較が必要である。

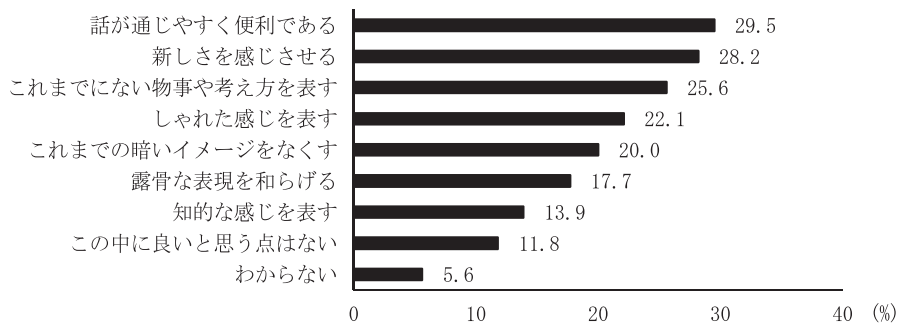


図 11 外来語を使うことの良いと思う点 (2 つまでの複数回答, N=3087, 回答計 = 174.2%)  
(国立国語研究所編 2004: 17)

最後に、媒体間でのサンプルサイズに大きな違いがあることも考慮する必要がある。書籍コーパスからのサンプルが 1719 件であるのに対し、雑誌コーパスからのサンプル数は 77 件しかない。そのため、雑誌においては少しの度数の差でも全体の割合の差として出やすいという側面もあるのかもしれない。

### 3.6 ジャンル

補足として、各媒体に含まれるジャンルについて考える。表 11 と表 12 は、BCCWJ に収録された「ジャンル 1」のタグ<sup>15</sup>を利用して、書籍と雑誌それぞれに含まれるジャンル分布を示したものである。表 11 を見ると、平均 (19.6%) より「ケース」の生起率が高いのは、(高い順に)「6 産業 (29.9%)」、「7 芸術・美術 (26.8%)」、「9 文学 (24.4%)」、「1 哲学 (23.3%)」、「3 社会科学 (21.9%)」の 5 ジャンルであった。一方、表 12 を見ると、「ケース」の生起率が平均 (33.8%) よりも高いジャンルは「政治・経済・商業 (80.0%)」、「工業 (53.8%)」、「総合 (36.8%)」の 3 ジャンルであった。書籍と雑誌のジャンル分類は同一でなく、両者を比較して一般的な傾向を読み取るのは難しいため、今回は記述のみにとどめる。また、全体的にセルの度数が少ないため、統計的検定も控えた。データ規模の拡大やジャンルカテゴリーを媒体間で統一することが今後の課題となる。

表 11 書籍コーパスのジャンル内訳

		0 総記	1 哲学	2 歴史	3 社会科学	4 自然科学	5 技術・工学	6 産業	7 芸術・美術	8 言語	9 文学	分類なし	合計
		3.4%	5.2%	6.3%	41.4%	15.9%	7.7%	7.4%	3.3%	6.5%	2.4%	0.5%	100.0%
ケース	度数	11	21	18	156	44	15	38	15	8	10	1	337
	%	18.6%	23.3%	16.7%	21.9%	16.1%	11.3%	29.9%	26.8%	7.2%	24.4%	11.1%	19.6%
既存語	度数	48	69	90	556	229	118	89	41	103	31	8	1382
	%	81.4%	76.7%	83.3%	78.1%	83.9%	88.7%	70.1%	73.2%	92.8%	75.6%	88.9%	80.4%
合計	度数	59	90	108	712	273	133	127	56	111	41	9	1719
	%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

<sup>15</sup> これは、日本十進分類法 (NDC) の第 1 次区分 (類) を用いたものである。

表 12 雑誌コーパスのジャンル内訳

		教育 ・学芸 11.7%	厚生 ・医療 11.7%	工業 16.9%	産業 3.9%	政治・経 済・商業 6.5%	総合 49.4%	合計 100.0%
ケース	度数	0	0	7	1	4	14	26
	%	0.0%	0.0%	53.8%	33.3%	80.0%	36.8%	33.8%
既存語	度数	9	9	6	2	1	24	51
	%	100.0%	100.0%	46.2%	66.7%	20.0%	63.2%	66.2%
合計	度数	9	9	13	3	5	38	77
	%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

### 3.7 販売対象（書籍のみ）

誰を対象に書かれた言葉であるかは、そのスタイルを決定づける重要な要素のひとつとなりうる。販売対象コードは、書籍コーパスにおいてのみ C コード（図書分類コード）の 1 桁目としてタグ付けされており、10 のカテゴリー<sup>16</sup>に分けられる。そこで販売対象別に「ケース」の生起率の分布を調べたところ、表 13 のような分布となった。セルの値が 5 以下である「8 児童（書）」と「9 雑誌扱い」を除くと、「ケース」の生起率が平均（19.6%）よりも高いカテゴリーは、「2 実用（書）（32.2%）」、「0 一般（書）（26.3%）」で、「3 専門（書）」では生起率は最も低い（12.6%）。度数の少ないセルがあることからカイ 2 乗検定は控えたが、一般書と専門書はどちらも書籍全体で同程度の割合を占める（それぞれ 34.5%, 35.5%）ことを考慮すると、単純に比較して「ケース」の生起率は相対的に一般書で高く専門書で低い傾向にあることが言えそうである。

表 13 販売対象の違いと「ケース／既存語」の生起率（書籍のみ）

		0 一般	1 教養	2 実用	3 専門	8 児童	9 雑誌 扱い	不明	合計
		34.5%	10.2%	8.5%	35.5%	0.1%	0.6%	10.6%	100.0%
ケース	度数	156	26	47	77	0	5	26	337
	%	26.3%	14.8%	32.2%	12.6%	0.0%	50.0%	14.3%	19.6%
既存語	度数	437	150	99	533	2	5	156	1382
	%	73.7%	85.2%	67.8%	87.4%	100.0%	50.0%	85.7%	80.4%
合計	度数	593	176	146	610	2	10	182	1719
	%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

## 4. 要因ごとの分析結果（言語内的要因）

本節では、金（2011）で取り上げられた「ケース」の基本語化に関わる言語内的要因のうち、「ケース」が出現する文法形式と、共起述語（金（2011）では「主節述語」）の種類の 2 要因について検証する。

<sup>16</sup> 分類項目は以下の通り。0 一般、1 教養、2 実用、3 専門、4（欠番）、5 婦人、6 学参Ⅰ（小中）、7 学参Ⅱ（高校）、8 児童、9 雑誌扱い（国立国語研究所コーパス開発センター 2011）

#### 4.1 形式差

本稿では、「ケース」が出現しうる形式区分のうち、名詞句と連体修飾節構造のみを分析対象とした。金（2011）によれば「ケース」は特に連体修飾節構造において多用されることでその使用を伸ばしてきたそうだが、今回、「ケース」の生起率に形式による差があるかについて検証する。表 14 から、「ケース」の生起率は名詞句において 15.2%、連体修飾節において 20.8% となっており、金の指摘する通り、連体修飾節における生起率の方が高い。ただし、統計的有意差は認められなかった。さらに、形式ごとに既存語の内訳（図 12）を見ると、「場合」の生起率が名詞句（26.3%）よりも連体修飾節（59.9%）で特に高く、「例」と「事例」の生起率は連体修飾節よりも名詞句で高いことがわかる。これにより、「外来語」対「既存語」という対立構造では両形式で統計的有意差は認められないが、既存語の中で形式による使い分けが見られるようである。

表 14 形式の違いと「ケース／既存語」の生起率

		名詞句	連体修飾節	合計
ケース	度数	30	333	363
	%	15.2%	20.8%	20.2%
既存語	度数	168	1265	1433
	%	84.8%	79.2%	79.8%
合計	度数	198	1598	1796
	%	100.0%	100.0%	100.0%

$$\chi^2=3.53, \text{ d.f.}=1, p=0.06$$

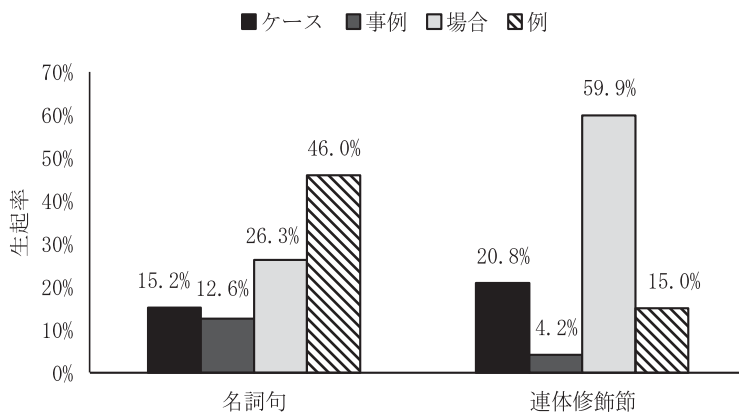


図 12 形式別に見た「ケース／事例／場合／例」の内訳

#### 4.2 共起述語の種類

最後に、2.2 節で見た分類に基づいて、共起述語の意味の違いによる「ケース」の生起率について検討する。表 15 から、「ケース」の生起率は、増減（58.0%）と生起（55.1%）の意味をあらわす述語との共起でかなり高く、以下、想定（26.7%）、多少（23.6%）、有無（15.1%）、報告（12.9%）と続く。カイ 2 乗検定の結果、これは統計的有意差が認められ（ $\chi^2=104.77, \text{ d.f.}=5, p<0.001$ ）、共起

述語の種類は「ケース」の生起に関連があることがわかった。

また、4語の内訳（図13）を見ると、「増減」と「生起」の述語は「ケース」との共起率が圧倒的に高く、「想定」、「多少」、「有無」の述語は「場合」との共起率が圧倒的に高い。「報告」の述語との共起率が相対的に高いのは「事例」と「例」である。このことから、これら4語の中で共起述語による使い分けがされているようである。

表15 共起述語の意味分類と「ケース／既存語」の生起率

		増減	生起	想定	多少	有無	報告	合計
ケース	度数	29	27	8	134	161	4	363
	%	58.0%	55.1%	26.7%	23.6%	15.1%	12.9%	20.2%
既存語	度数	21	22	22	433	908	27	1433
	%	42.0%	44.9%	73.3%	76.4%	84.9%	87.1%	79.8%
合計	度数	50	49	30	567	1069	31	1796
	%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

$$\chi^2=104.77, \text{ d.f.}=5, p<0.001$$

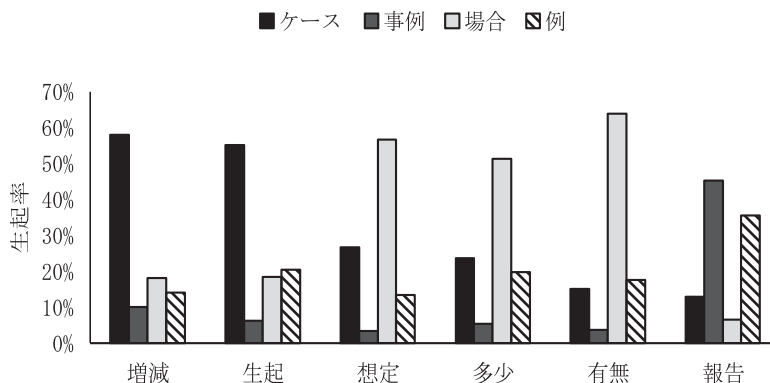


図13 共起述語の意味分類別に見た「ケース／事例／場合／例」の内訳

## 5. 結論

本稿では、「ケース」を事例として、外来語と既存語の語彙交替に影響を与えうる要因のうち、特に言語外的要因に焦点をあてて検証を行った。語彙を社会言語学的変異形として扱うことには方法論的な問題がいくつかあるが、先行研究（金2011）から語彙交替が可能な環境をある程度特定することと、BCCWJの利用によりターゲットとなる語彙を書き手の属性情報とともに効率よく収集することで問題の解決を試みた。これにより、本稿では、意識調査にとどまっていた外来語の社会言語学的調査を、実際の言語資料に基づいて行うことができた。

分析の結果、「ケース」の生起率に影響がある言語外的要因として統計的有意差が認められたのは、書き手の生年代と、最終学歴、媒体種であることがわかった。生年代については、若い世代ほど「ケース」の生起率に上昇が見られたことで、既存語から外来語への語彙使用の変化が認

められた。これは、外来語志向に年齢による差が見られるという意識調査の結果（田中 2007）を補足する結果となった。書き手の最終学歴については、学歴が高くなるほど「ケース」の生起率が低下しており、特に大学院卒の書き手においては他の学歴カテゴリーと異なり「例」の生起率を下回ることがわかった。媒体種の違いによる生起率の差については、書籍よりも雑誌において「ケース」の生起率が高かったが、これには書き手の生年代と学歴が少なからず影響している可能性を指摘した。また、書籍と雑誌の媒体差の本質が何であるかは、媒体の「公共性」の度合いや媒体と結びつくイメージの違いについてより詳しく特徴づけることで見えてくる可能性も指摘した。ジャンル分布と販売対象の差については、サンプル規模が小さかったことと、両媒体で統一されたカテゴリー分けがされておらず一般的な傾向を導くことができず、さらなる調査が今後の課題となった。

一方、「ケース」の生起率に影響がある言語内要因として統計的有意差が認められたのは、共起述語の種類で、「ケース」の生起率は特に「増減」「生起」の意味をもつ述語と共起するときに圧倒的に高く、コロケーションの要素も重要な要因のひとつであることがわかった。これにより、先行研究（金 2011）について付加的な検証を行うことができた。

最後に、外来語が日本語においてどのような地位を占めているのかについて考察したい。本稿では、「ケース」の生起率が雑誌において高いことから、外来語のもつ「私的な日常語性」（田中 2012）が指摘された。このことから、外来語「ケース」は、より日常的な領域や場面と結びつき、書籍のような公共性が高い媒体においてはまだまだ浸透率が低い語種であると考えられる。このような外来語の「日常語性」は、「ケース」の生起率が書籍のうち専門書よりも一般書において高い傾向にあったことや、学歴の高い書き手ほど外来語「ケース」の使用率が減少することとも関連がある可能性がある。

今後の課題としては、ロジスティック回帰分析などにより、今回検証した各々の要因が外来語の生起率にどう影響しているかを他要因の影響を考慮した上で検証する必要がある。また、本研究で得られた傾向が外来語一般に拡張できるのかどうか、「ケース」以外の外来語も調査することにより事例研究を積み重ねていく必要がある。それにより、日本語において外来語が果たす社会言語学的役割の全体像が見えてくるだろう。

## 参考文献

- Irwin, Mark (2011) *Loanwords in Japanese*. Amsterdam: John Benjamins.  
 陣内正敬・田中牧郎・相澤正夫（編）（2012）『外来語研究の新展開』東京：おうふう。  
 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。大阪大学大学院文学研究科日本語学講座。  
 国立国語研究所（編）（2004）『外来語に関する意識調査（全国調査）』。  
 国立国語研究所コーパス開発センター（2011）『『現代日本語書き言葉均衡コーパス』利用の手引（第 1.0 版）』。  
 久屋愛実（2013）「外来語使用における言語外的要因の分析—書き言葉コーパスの利用可能性—」『第 3 回コーパス日本語学ワークショップ予稿集』183-192。  
 Myers-Scotton, Carol (2006) *Multiple voices: An introduction to bilingualism*. Malden, MA: Wiley Blackwell.  
 総務省統計局（2009）『日本標準職業分類（平成 21 年 12 月統計基準設定）』  
<http://www.stat.go.jp/index/seido/shokugyou/21index.htm>（2013 年 6 月 1 日参照）

- 田中牧郎 (2007) 「漢語・和語と比較した外来語に対する意識」『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—』(国立国語研究所報告 126), 302–310.
- 田中牧郎 (2012) 「国語教育における外来語—コーパスによる類型化を通して—」陣内正敬・田中牧郎・相澤正夫 (編), 224–242.
- 山崎誠・小椋秀樹・小沼悦・柏野和佳子・佐野大樹・高田智和・富士池優美・間瀬洋子・丸山岳彦・山口昌也 (2012) 「研究活動・成果の総括：データ班 代表性を有する現代日本語書籍コーパスの構築」『特定領域「日本語コーパス」』平成 22 年度公開ワークショップ (研究成果報告会) 予稿集』149–156.
- 梁敏鎬 (2012) 「日本語と韓国語の外来語の受容意識—イメージ調査の分析—」陣内正敬・田中牧郎・相澤正夫 (編), 148–167.

## Synchronic Distribution of Loanwords in Contemporary Written Japanese: A Case Study of *keesu* ('case')

KUYA Aimi

Ph.D. Candidate, University of Oxford / Visiting Researcher, NINJAL [–2013.02]

### Abstract

This paper aims to investigate the synchronic distribution of the loanword *keesu* ('case') in contemporary Japanese and to examine language-external effects on its use. This is an empirical investigation into lexical variation between loanwords and their native equivalents, using the Balanced Corpus of Contemporary Written Japanese (BCCWJ) (National Institute for Japanese Language and Linguistics 2011). The statistical analyses show that the author's birth year and level of education, the type of media, and the type of predicate co-occurring with the targeted lexical item all have an impact on the use of this loanword. The tendency for the younger generation to use the loanword more shows that there has been a lexical change over time in favor of *keesu* as opposed to its Japanese counterparts, such as *jirei*, *baai* and *rei*. The loanword is disfavored by the most educated authors and in books, which seem to be associated with more formal settings. This may imply that the use of loanwords has not yet been established fully in certain social categories or environments.

**Key words:** loanwords, lexical variation, language-external factors, corpus