

## コーパスを用いた日本語研究に関する一考察

庵 功雄

一橋大学国際教育センター教授

### 1. はじめに

現代日本語書き言葉均衡コーパス（BCCWJ）の公開など、近年大規模なコーパスが整備されてきている。それにともなって、「コーパスを調べれば何とかなる」といった「コーパス万能論」とでも言うべき考え方が増えてきているが、実際には適切な使い方をしない限り、コーパスを用いて意味のある研究を行うことは難しい。本稿では、筆者がこれまでに発表したコーパスを用いた研究をケーススタディとして紹介しながら、コーパスを用いた日本語研究の要点を述べる。最後に、コーパスの可能性と限界についても触れる。

### 2. コーパスはなぜ必要か？

コーパスはどのような場合に有力なのだろうか。そのことを知っておくと、コーパスの限界も見えてくる。

#### 2.1 母語話者が持っている文法能力

母語話者は、母語について以下の能力(文法能力 grammatical competence)を持っている（もちろん、このことは母語について説明できるということとは無関係である）。

- (1)a. 母語話者は母語の任意の文が文法的かどうかを判断できる。
- b. 母語話者は（モニターができる環境下では）文法的な文のみを産出する。

このことから、文法性のみを論じる場合にはコーパスは不要であると

言える（コーパスでの出現頻度からある形式の文法性を論じることはできない）。

## 2.2 頻度が問題となる場合

文法現象の中には文法性だけでは論じられない問題がある。つまり、ある表現が「言えるか言えないか」ではなく、「よく使われるかあまり使われないか」を問題にする必要がある場合がある。そうした場合には、母語話者の内省は無力で、コーパスを利用する必要がある（cf. 庵 2009a、森 2011）。

これ以外にも、シラバスを策定する際などにもコーパスは重要な役割を果たす。

## 3. コーパスの利用の仕方

ここでは、コーパスの利用法について述べる。

### 3.1 研究手法

コーパスを用いた研究には大きく分けて2つの方法がある。

第一は、Corpus-driven approach（CDA）で、これは、事前に特定の見込みなどを持たずにコーパスを検索し、出てきた結果から帰納的に結論を導くという方法である。

第二は、Corpus-based approach（CBA）で、これは、事前にある一定の目算を立てて、その論証のためにコーパスを利用するという方法である。

2つの方法の主な長所と短所は次の通りである。

	長所	短所
CDA	取りかかりやすい	結果を分析する言語学的な能力を要求される度合いが高い
CBA	調査結果が無駄になる可能性は低い	仮説を立てられるまで調査に取りかけられない

表1 CDA と CBA

CDA の立場から書かれた論文に中俣(2011)がある。一方、4.1 で取り上げる庵(1995)は CBA の立場から書かれたものである。

### 3.2 コーパスを用いた日本語研究のいろいろ

コーパスを使った調査・研究には、狭義の文法研究以外にもいろいろなものがある。

#### 3.2.1 大規模コーパスを用いた調査・研究

まず、母語話者の大規模コーパスを用いた調査・研究について述べる。

第一に挙げられるのは、ある項目の用例を全て取り上げるタイプの調査(悉皆(しっかい)調査)である。これは古くは国立国語研究所(1964)などで行われたものだが、BCCWJ の登場でこうした調査がやりやすくなっている。こうした調査の例に森(2011a)がある。最新の研究に田(2013)がある。4.2 で取り上げる庵・宮部(2013)もこのタイプの研究である。

第二に挙げられるのは、ある項目のジャンルによる違いを扱うものである。BCCWJ を使って書きことばの中でのジャンル差を見ることができし、BCCWJ と CSJ (日本語話し言葉コーパス)、名大会話コーパスなどを比べることで、書きことばと話しことばの差を見ることが出来る。このタイプの調査の例に小西(2011)がある。

第三に挙げられるのは、ある形式の許容度などを経年的に調査するものである。このタイプの研究には松田(2008)などがある。また、太陽コーパスや明六雑誌コーパスといった国立国語研究所が刊行している近代語コーパスを用いて近代からの日本語の変化をコーパスを用いて研究することも可能になってきた。これについて詳しくは田中(2013)を参照。

#### 3.2.2 学習者コーパスを用いた調査・研究

次に、学習者コーパスを用いた調査・研究について述べる。

第一に挙げられるのは、学習レベルによる違いを考えるものである。このタイプの研究としては山内(2009)や橋本(2011)が挙げられる。

第二に挙げられるのは、母語の違いや母語の影響を考えるものである。このタイプの代表的な研究に奥野(2005)がある。また、最新の研究とし

て三井(2013)がある。

最後に挙げられるのは、母語の転移ということに踏み込んで考えるタイプの研究で、代表例に張(2011)がある。同書で主張されている「三位一体」の習得研究というのは、次の3つを総合した研究のことである。

(2) a. 対照研究

b. 誤用観察 (学習者コーパスの利用)

c. 習得研究

この三者の例に中国語話者の「も」の習得を論じた張(2010)、中俣(2010, 2013)がある。

#### 4. ケーススタディ

ここでは、コーパスを用いた日本語研究の具体例として、筆者のこれまでの研究の中でコーパスを用いたものを取り上げて解説する。

##### 4.1 CBA の立場の研究 (ガ～シタイとヲ～シタイ)

最初に取り上げる論文の書誌は次の通りである。

・庵 功雄(1995)「ガ～シタイとヲ～シタイ—格標示のゆれ—」『日本語教育』86

###### 4.1.1 問題となる現象

「～をV」を「～\_\_Vたい」にした場合、「\_\_」の部分に入る助詞は「が」だとされていた(ex. 水をを飲む→水がを飲みたい)。しかし、(3)(4)の「を」を「が」に変えるのは難しい。

(3) ロンドンでは世界選手権のリベンジを果たしたい。(朝日新聞朝刊 2012. 1. 1)

(4) 昨日までの世界はもうない。たとえ危険があっても、勇気を持ちたい。新たな展開を期待し、わくわくしながら——。(朝日新聞朝刊 2012. 1. 1)

実例を集めたり、作例してみると、「Vたい」の前は「を」が多い／「を」でないと不自然という場合が多いことに気づく。

#### 4.1.2 どのようにコーパスを利用するか

この論文は CBA の立場に立ち、以下のリサーチクエスション (RQ) を立てて、それをコーパスを使って論証することを目指したものである。

- (5) RQ: 「Vたい」(ただし、Vは「を」を取る動詞)の前の助詞は「が」ではなく、「を」であるのが無標である。

#### 4.1.3 調査の結果

調査の結果は次の通りである。この結果から、(5)の RQ は証明された。

「～が」「～を」と共に使われる動詞  
「～が」とよく使われる動詞  
「～を」とよく使われる動詞

	意 味 分 野	ガ	ヲ		意 味 分 野	ヲ	ガ		意 味 分 野	ヲ	ガ
する	2.34	10	69	テモラウ		132	0	言う	2.31	20	8
食べる	2.33	10	7	する	2.34	69	10	作る	2.38	20	0
言う	2.31	8	20	テミル		65	1	テイク		19	0
知る	2.306	6	23	テヤル		64	0	考える	2.306	18	0
飲む	2.33	5	1	サセル		43	0	送る	2.152	14	0
				見る	2.309	41	2	テシマウ		14	0
				テオク		38	0	避ける	2.35	13	1
				聞く	2.309	27	1	テイル		11	0
				テイタダク		27	0	見せる	2.309	10	0
				知る		23	6				

\* 「意味分野」は国立国語研究所(1964)『分類語彙表』の意味分類による

#### 4.2 大規模コーパスを用いた研究(漢語サ変動詞の使役)

次に取り上げる論文の書誌は次の通りである。

・庵 功雄・宮部真由美(2013)「二字漢語動名詞の使用実態に関する報告―「中納言」を用いて―」『一橋大学国際教育センター紀要』4号、pp. 97-108、一橋大学

##### 4.2.1 問題となる現象

この論文では漢語サ変動詞の「する」と「させる」について取り上げた。その背景には次のような問題がある。

初級で教える項目を減らすべきだという議論があり (ex. 庵 2009b、山内 2009)、その中でも使役は不要だという意見が強い (cf. 庵 2012b, 2013、岩田 2012、森 2012)。こうした流れの中で、漢語サ変動詞の場合の「させる」について考えたのがこの論文である。

もう 1 つは、定延 (2000) が「使役余剰」と呼ぶ現象についてである。使役余剰というのは、例えば、次の文で「実現する」で自足的であるはずなのに、「実現される」も使われる (しかも、母語話者の許容度は後者の方が高い) という現象である。

- (6) 01 年の自民党総裁選では、党内のアウトサイダーだった小泉氏が、消費税を上げないといって勝った。民主党も、消費税を上げないという小沢一郎氏の戦略で政権交代を実現した。

(2012. 5. 10 朝日新聞朝刊)

- (7) 2009 年、「コンクリートから人へ」を掲げたマニフェストで民主党が政権交代をを実現させた。(2012. 9. 16 朝日新聞朝刊)

#### 4.2.2 コーパスの使い方

今回の論文におけるコーパスの使い方は「中納言」を使った悉皆調査である。

まず、旧日本語能力試験の出題基準にある 1 級～4 級の語彙をコーパスにし、それを「茶まめ」にかけて形態素解析し、語種と品詞に関する情報を取り出す。その中で、品詞情報が「名詞－普通名詞－サ変可能」であるもの全てを調査対象とした。

#### 4.2.3 検索の工夫

検索に関しては、単に語彙素「為る」で検索すると、書字形として「する」と「させる」が区別されないので、書字形としての「する」と「させる」を区別するように検索した。

#### 4.2.4 結果とその含意

この論文の結果は次の 2 点において意味を持つ。

#### A. 「使役」について

まず、「使役」については、「させる」の頻度が高い語は大部分、ヲ格に非情物をとる動詞である。逆に言うと、ヲ格に有情物をとる動詞はほとんど「させる」の上位には現れない。一般に、使役のヲ格は有情物が来るのが典型であり、これが使役と他動詞の最も大きな違いである (cf. 庵 2012a、前田 2003)。

(8) 太郎が看板を廊下に立てた。(他動詞)

(9) 先生が太郎を廊下に立たせた。(使役)

それにもかかわらず、使役のヲ格が非情物であるというのは、「漢語+させる」が「使役」ではなく、「他動詞」として使われていることを示している<sup>1</sup>。つまり、「漢語+させる」の中心的な用法は「使役」ではなく、「他動詞」なのである (cf. 森 2012)。

#### B. 「使役余剰」について

一方、「使役余剰」について言うと、「使役余剰」は頻度から見ても一般的な現象であることがわかった。これは、「使役」について、これまでの「ビリヤードモデル」ではなく、「カビ生えモデル」を採用すべきことを示唆しているように思われる (cf. 定延 2000)。

### 5. 論文化する際の留意点

ここでは、以上のようなコーパスの使い方を受けて、特に、投稿論文 (学会誌論文) を執筆する際に留意すべき点について述べる。4 で見たことは論文の骨格であるが、これを投稿論文に仕上げるには、投稿論文の作法にしたがって「お化粧」をしていく必要がある。

---

<sup>1</sup> 次のように使役形が他動詞の代わりに使われることは和語でもあるが、その頻度は高くない。

- 太郎は新宿に向けて車を走らせた。
- 次郎が懐中電灯を光らせた。
- 梅雨前線が雨を降らせる。

### 5.1 問題提起

この部分では、論文で扱う対象を簡潔に示すことが必要である。読者（査読者）を「読む気にさせる」ことができるかがポイントになる。

### 5.2 先行研究

この部分で重要なのは、本論で扱う論文のみを取り上げるということである。原則として、実際に本論で議論の対象としない論文は取り上げてはいけない。逆に言うと、この部分で取り上げた論文については、本論で何らかのコメントをする（＝論文の中に位置づける）ことが必要である。

### 5.3 先行研究の取り上げ方

投稿論文は何らかの意味で「オリジナリティ」を持つはずなので、先行研究を全面的に肯定してしまっては論文にならない。したがって、先行研究の批判をすることが必要なのだが、ここで重要なのは先行研究の「揚げ足を取る」ことではない。必要なのは、「先行研究の議論にしたがって実例を分析すると、～という問題が生じる」という書き方である。合わせて必要なのは、「本論の分析を取れば、先行研究で説明できる現象に加えて、先行研究では説明できなかった現象も説明できる」ことを示すことである。

### 5.4 現象の記述（コーパス調査の結果）

この部分では、コーパスの調査の結果を提示する。この際に重要なのは、論証したい内容に合わせて調査結果を提示することである。そして、そうした提示をするためには、事前に RQ を明確に立てておく必要がある。

CDA の場合は、言わばコーパス調査の結果から RQ を導き出す形になる。これは不可能なわけではないし、コーパス言語学の方法論としてはこちらの方がまっとうであるとも言えるが、実際には、コーパスの検索結果から適切な RQ を導き出すには、相当の言語学的な分析力が必要であり、それを十分に持っていない人が CDA をとると、有意な差を見逃し



たり、「数えただけ」の結果になることが多い。

ただし、CBA でも、事前に適切な見込みを立てられることが必要であり、そのためには、「内省」ができること、あるいは、それに匹敵する程度のデータを見る必要がある。

## 5.5 現象の説明

この部分は理論的な研究では最も必要とされる部分である。記述的な研究では 5.4 の部分の方が相対的には重要であるが、それでも、学習者への説明ということを考えると、現象をどのように理解するかということは重要である。

## 5.6 参考文献

現在の投稿論文では、これは「引用文献」と考える必要がある。5.2 でも述べたように、大原則は、「本文で引用した論文は必ず引用文献に挙げる。引用文献に挙げた論文は必ず本文で引用する」ということである。

なお、特に博士論文を書くまでの段階の場合は、自分の論文を引用する必要がある。この際には次のことに注意する。

\*\*\*\*\*日本語教育学会大会応募規定より抜粋\*\*\*\*\*

### <引用文献の挙げ方>

本文で言及した論文および発表に重要な関連を持つ先行研究などがある場合は要旨にその文献を挙げてください。文献は要旨の分量に含まれません。(中略) 文献を挙げる際には以下の情報を入れてください。

著者名，出版年，論文名，雑誌名／書名，号数，出版社名

(例) 教育花子(2009)「英語のオノマトペ」『世界のオノマトペ』○×出版

※応募者自身の論文であっても、発表の内容に関係する場合には引用してください。その際、次のような言及の仕方をすることによって、執筆者が特定されないようにしてください。

(例) ○田中(2010)で{述べられている／指摘されている}のように、

...

×田中(2010)で{述べた／指摘した}のように，…

(「<論文名>で～したように」という表現は(執筆者が特定できるので)  
使わないでください。)

※応募時において公刊されている文献のみを挙げてください(応募時  
において「印刷中」「投稿中」などの文献は挙げないでください)。

\*\*\*\*\*  
\*\*\*\*\*

## 6. まとめ—コーパスでできることとできないこと—

以上、コーパスを用いた研究について述べてきたが、その内容を踏まえて、コーパスでできることとできないことについてまとめる。

### A. 用例数が多い現象についてはコーパスは有効(であることが多い)

一般に、用例数が多い現象については、コーパスは有効(であることが多い)。今回取り上げた2つの現象はそういう意味でコーパスを用いた研究に向いていると言える。

### B. 用例数が少ない現象についてはコーパスだけでは有意な一般化は困難(なことが多い)

一方で、用例数が少ない現象については、コーパスはあまり有効ではない。特に、用例数が1ケタのものなどについては、「誤用」か「正用」かの判断が難しい(cf. 張 2013)。こうした場合、他のコーパスを用いてデータ数を増やす、アンケート調査や文法性判断テストを行う、など他の調査法を合わせて用いることが必要である(cf. 森 2011b)。

### C. 「内省」ができる能力(「変数」を変えて作例ができる能力)が必要

コーパスを使うようになると、全てデータで考えられる(あるいは、考えるべきだ)と思われがちであるが、実は、コーパスを適切に使う上で重要なのは、正しく「内省」ができる能力である。「内省」の優れているところは、変数を自由に変えて考えることができる点にあり、「内省」が自由にできる能力は、検索の結果得られたデータを分析する際に重要

な役割を果たす。コーパスを使う際に重要なのは、一次的な検索結果であるよりも、そのデータをいかに自分が論証しようとする結果に合わせて検索し直すかという点にある。そして、その際に重要になるのは「内省」ができる能力である。非母語話者の場合であっても、優れた内省力を持つ母語話者をインフォーマントにすれば、同様の研究は可能である。

D. 「話しことば」については、コーパスでできることはかなり限られている

コーパスが普及したといっても、それは「書きことば」についてである。「話しことば」については、コーパス（だけ）でできることはまだ限られている。例えば、BCCWJ は日本語における最初の「均衡コーパス」（「代表性」を持つコーパス）であるが、「話しことば」において現存するコーパスはどれもこうした意味の「代表性」を持つものとは言い難い。Halliday は言語テキストは次の3つの変数の組み合わせで形成されると述べている<sup>2</sup>が、話しことばの場合、これらの変数をコントロールしてコーパスを作ることは難しい。

(10) Field : 何について

Tenor : 誰が誰に対して

Mode : どのような方法で (ex. 話しことば、書きことば、メール、チャット)

このことは、「話しことばの文法」をコーパス（だけ）を用いて研究するということが「原理的に」非常に難しいということを示している。

## 参考文献

- 庵 功雄(2009a)「推量の「でしょう」に関する一考察—日本語教育文法の視点から—」『日本語教育』142
- 庵 功雄(2009b)「地域日本語教育と日本語教育文法:「やさしい日本語」という観点から」『人文・自然研究』3、一橋大学

---

<sup>2</sup> この3つを合わせて「レジスター（言語使用域）」と言う (cf. Halliday1994)。なお、「レジスター」という語については、このほかにもさまざまな考え方がある (cf. Allwood 2013)。

- 庵 功雄(2012a)『新しい日本語学入門 (第 2 版)』スリーエーネットワーク
- 庵 功雄(2012b)「文法シラバス改訂のための一試案—ボイスの場合—」『日本語／日本語教育研究』3、ココ出版
- 庵 功雄(2013)「「使役 (態)」に言及せずに「使役表現」を教えるには—1 つの「教授法」—」『日本語／日本語教育研究』4、ココ出版
- 岩田一成(2012)「初級教材における使役の「偏り」と使用実態」『日本語／日本語教育研究』3、ココ出版
- 奥野由紀子(2005)『第二言語習得過程における言語転移の研究—日本語学習者による「の」の過剰使用を対象に』風間書房
- 国立国語研究所(1964)『現代雑誌九十種の用語用字 第三分冊：分析』秀英出版
- 小西 円(2011)「使用傾向を記述する—伝聞の「ソウダ」を例に—」森・庵編(2011)所収
- 定延利之(2000)『認知言語論』大修館書店
- 田中牧郎(2013)『近代書き言葉はこうしてできた (そうだったんだ！日本語)』岩波書店
- 張 志剛(2013)「現代日本語の二字漢語動詞の自他」一橋大学言語社会研究科博士論文
- 張 麟声(2010)「「同類」の「も」と対応する中国語の諸形式との対照研究」『中国語話者のための日本語教育研究』創刊号、日中言語文化出版社
- 張 麟声(2011)『新版中国語話者のための日本語教育研究入門』日中言語文化出版社
- 田 昊(2013)「「言いさし」の「けど」類の使用実態に関する一考察」『日本語教育』156
- 中俣尚己(2010)「学習者の「も」の使用状況—「同類」の「も」の不使用に注目して—」『中国語話者のための日本語教育研究』創刊号、日中言語文化出版社
- 中俣尚己(2011)「コーパス・ドライブン・アプローチによる日本語教育文法研究—「てある」と「ておく」を例として—」森・庵編(2011)所

収

- 中俣尚己(2013)「中国語話者による「も」構文の習得—「A も B も P」「A も P、B も P」構文に注目して—」『日本語教育』156
- 橋本直幸(2011)「学習者コーパスから見る超級日本語学習者の言語特徴—2つの観点から—」森・庵編(2011)所収
- 前田直子(2003)「10 使役 コロンブスの卵—卵を立たせる」庵 功雄・日高水穂・前田直子・山田敏弘・大和シゲミ『やさしい日本語のしくみ』くろしお出版
- 松田謙次郎(2008)『国会会議録を使った日本語研究』ひつじ書房
- 三井さや花(2013)「英語母語話者による日本語名詞の複数形の産出について—英語と日本語の複数認識のずれ—」『日本語教育』154
- 森 篤嗣(2011a)「『現代日本語書き言葉均衡コーパス』コアデータにおける初級文法項目の出現頻度」森・庵編(2011)所収
- 森 篤嗣(2011b)「着点を表す助詞「に」と「へ」における日本語母語話者の言語使用について」森・庵編(2011)所収
- 森 篤嗣(2012)「使役における体系と現実の言語使用—日本語教育文法の視点から—」『日本語文法』12(1)、pp. 3-19、日本語文法学会
- 森 篤嗣・庵 功雄編(2011)『日本語教育文法のための多様なアプローチ』ひつじ書房
- 山内博之(2009)『プロフィシエンシーからみた日本語教育文法』ひつじ書房
- Allwood, J.(2013)“Resister.” Chappelle, C.A.(ed.) *The encyclopedia of applied linguistics*. Wiley-Blackwell.
- Halliday, M.A.K.(1994) *An introduction to functional grammar*.(2<sup>nd</sup> Ed.) Edward Arnold.