

国立国語研究所学術情報リポジトリ

言語政策に役立つ，コーパスを用いた語彙表・漢字表等の作成と活用

メタデータ	言語: jpn 出版者: 公開日: 2020-06-29 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://doi.org/10.15084/00002858

言語政策に役立つ,コーパスを用いた 語彙表・漢字表等の作成と活用

田中 牧郎・相澤 正夫・斎藤 達哉・棚橋 尚子・近藤 明日子・
河内 昭浩・鈴木 一史・平山 允子

国立国語研究所内部報告書 (LR-CCG-10-07)

言語政策に役立つ，コーパスを用いた 語彙表・漢字表等の作成と活用

田中	牧郎
相澤	正夫
斎藤	達哉
棚橋	尚子
近藤	明日子
河内	昭浩
鈴木	一史
平山	允子

平成23年2月

©2011 大学共同利用機関法人 人間文化研究機構 国立国語研究所

まえがき

この報告書は、文部科学省科学研究費特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21世紀の日本語研究の基盤整備」（領域代表：前川喜久雄）の中に設けられた計画研究班のひとつである言語政策班が進めてきた、「言語政策に役立つ、コーパスを用いた語彙表・漢字表等の作成と活用」という研究の成果について、研究期間（2006年度～2010年度）を終了するにあたって報告するものである。

私たち言語政策班は、従来コーパスがほとんど使われていなかった分野である国語政策と国語教育の分野にコーパスを導入することを目指した。この二つの分野での共通の課題であり、コーパスを導入する効果が大きいと予想された語彙と漢字の問題を主に取り上げて、コーパスに基づく語彙表・漢字表等の設計、コーパス及び語彙表・漢字表の国語政策課題・国語教育課題への活用、という二段階の研究を進めてきた。

研究期間のちょうど折り返し点（2008年9月）で中間報告書を発行したが、そのときは、コーパスや語彙表・漢字表の設計に関する研究を中心に報告した。今回の最終報告書では、中間報告以後の研究を主に収録することとし、第1章でコーパスや語彙表・漢字表の設計についてまとめた後、第2章から第4章でそれらを国語政策・国語教育の諸課題にどのように活用していくことができるのかについて研究を行った。すべて、この報告書のために書き下ろしたものである。

依拠できる先行研究が乏しい中で、模索と錯誤を繰り返しながら進めてきた研究であったため、研究成果も十分に完結したものにはならなかった。しかし、こうして報告書にまとめてみると、国語政策研究や国語教育研究を、言語の実態把握に基づいた確実な議論を踏まえて行う方向へと展開させるのに、コーパスの導入はまことに大きな力をもたらさだろうとあらためて強く予感させてくれる。本特定領域研究で構築した『現代日本語書き言葉均衡コーパス』は、いままさに完成しようとしている段階で、本格的な利用はこれから始まるものである。このコーパスが、国語政策研究・国語教育研究の分野に与える影響も、きわめて大きなものになるに違いない。本報告書が、新しい時代の国語政策・国語教育を研究する人たちにとって、少しでも参考になることを願っている。

なお、本研究を進めるにあたっては、多くの方々にお世話になった。まず、教科書のコーパスを作成するにあたり、社団法人教科書協会、一般社団法人教学図書協会、各教科書を刊行している出版社には、それぞれひとかたならぬご配慮を賜った。コーパス化にご許諾をいただいた上に、著作権者に連絡を取る仲介など面倒なお願いに快く応じてくださった。コーパス化へのご理解をお願いした著作権者の多くの方々から、ご許可をいただいたことありがたかった。コーパスと語彙表・漢字表の作成においては、研究分担者・連携研究者・研究協力者のほかに、多くの補助者の方々に国立国語研究所に来ていただき緻密な作業を担当していただいたことで、信頼できるデータにしていけることができた。さらに、班会議やワークショップに参加していただき、数々の活発な意見や貴重な助言をくださった、研究者・教師・教材開発者・政策担当者の方々の激励も、研究を進める大きな原動力となった。研究期間の様々な局面でお世話になったすべての方々に、厚くお礼を申し上げたい。

この特定領域研究の企画が持ち上がった段階では、国語政策と国語教育の分野をそれぞれ扱うことができればよいと漠然とは思い描いていたが、この二つ

を組み合わせて一つの計画研究班にすることまでは、思いが及ばなかった。あれこれと思い悩んでいた私に、二つの分野を関連づけて扱う班を作る提案をされたのは、故野村敏夫氏であった。言語政策班の生みの親であり、研究の舵取りをしてくださった野村氏は、研究期間の折り返し点を目前にしたところで急逝された。コーパスが形をなし、語彙表・漢字表が十分なものに近づいてくるほどに、国語政策と国語教育の分野が連携し合って研究することの意義をあらためて確認する思いである。野村氏が見通されていた新しい研究段階に向かってさらに努力を続けていくことが、ご恩に報いることになるに違いない。

2011年2月
田中牧郎

目次

まえがき（田中牧郎）	1
------------------	---

第1章：コーパスと語彙表の作成

第1節 教科書コーパス（田中牧郎・近藤明日子・平山允子）	7
第2節 教科書コーパス語彙表（田中牧郎・近藤明日子）	55
第3節 BCCWJ 主要コーパス語彙表（田中牧郎・近藤明日子）	65
第4節 学校・社会対照語彙表（田中牧郎・近藤明日子）	69

第2章：国語政策でのコーパスの活用

第1節 語彙レベルに基づく重要語彙リストの作成 一国語政策・国語教育での活用のために―（田中牧郎）	77
第2節 難解用語の抽出と序列化におけるコーパスの利用 一医療用語を例に―（田中牧郎・近藤明日子）	89
第3節 漢字政策に役立つ漢字表のあり方 一固有名に使われる漢字の検討のために―（相澤正夫）	107
第4節 BCCWJによる「NDC ジャンル別漢字出現頻度表」の分析（斎藤達哉）	127

第3章：国語教育でのコーパスの活用

第1節	中学校・高校教科書の教科特徴語リストの作成（近藤明日子）	145
第2節	「分類重要語彙リスト」の作成による教科教育と語彙教育の関連づけ（田中牧郎）	153
第3節	作文コーパスからみる生徒の使用語彙（鈴木一史）	171
第4節	漢字教育におけるコーパスの活用（棚橋尚子）	181

第4章：国語の教材研究と実践事例

第1節	「少年の日の思い出」の語彙指導（田中牧郎）	195
第2節	コーパスを使った語彙単元学習（鈴木一史）	205
第3節	「テーマ型作文」の教材開発 — 「テーマ語彙集」の活用による — （河内昭浩）	215
第4節	他教科における漢字指導実践 — 社会科（小学校第6学年）の事例 — （棚橋尚子）	237
	「言語政策班最終成果 CD-ROM」について	247
	成果物一覧	251

第1章

コーパスと語彙表の作成

第1章 第1節

教科書コーパス

田中 牧郎・近藤 明日子・平山 允子

I. 教科書コーパスの基本設計

1. 教科書コーパス作成の目的

特定領域研究「日本語コーパス」言語政策班では、国語政策や国語教育の分野にコーパスを導入する研究を進めてきた。特に、コーパスに基づく語彙表や漢字表を作成し、これを介して、政策や教育上の課題とコーパスとを結びつけることに力を入れてきた。現在の日本社会において、国語政策や国語教育の現状を見ると、社会において使用されている語彙や漢字、あるいは学校で実際に教えられている語彙や漢字の実態が十分に把握されていないという問題があるからである。

社会において使用されている語彙や漢字の実態は、『現代日本語書き言葉均衡コーパス』(BCCWJ)によって十分に把握できると考えられ、これに基づく語彙表・漢字表を作成し、活用した研究を進めてきた。一方、学校で教えられている語彙や漢字の実態は教科書によって把握するのが望まれるだろう。BCCWJには教科書データが含まれるが、これはサンプリングされたものである。教科書は体系性が強いいため、学校で教えられている語彙や漢字を正しく把握するには、サンプリングデータではなく全文データが必要になる。そこで、小学校・中学校・高等学校で使われている検定教科書を対象とした全文コーパスを作成し、そのコーパスをもとにした語彙表・漢字表を設計することにした。

2. 教科書コーパスの対象

まず、対象とする年次について、現在の学校で使われている教科書を対象としたいところだが、BCCWJとの比較ということも考慮し、BCCWJの母集団コーパス部分が対象とする下限である2005年度に使用された教科書を対象とすることにした。

次に、対象とする学年と教科であるが、小学校・中学校・高等学校で教えられている全体を把握したいので、全学年・全教科を対象とすることにした。ただし、専門に分化した高等学校の一部の科目（「農業」「商業」などいわゆる専門科目とされているもの）は対象外とした。予算や研究期間などを考慮して、作成実務が確実にこなせる分量を見積もり、各学年・各教科1種ずつとした。その1種は、なるべく多くの学校、児童・生徒に使われているものとするべく、できるだけ発行部数の多いものから順次、出版元の教科書会社に依頼し、利用許諾が得られた教科書から、対象に定めていった。

このようにして教科書コーパスの対象に定めた教科書は、表1の144冊である。

表1 教科書コーパスが対象とした教科書

校種	教科	使用学年	書名	発行会社
小	国語	1	こくご 一上 かざぐるま	光村図書出版
小	国語	1	こくご 一下 ともだち	光村図書出版
小	国語	2	こくご 二上 たんぼぼ	光村図書出版
小	国語	2	こくご 二下 赤とんぼ	光村図書出版
小	国語	3	国語 三上 わかば	光村図書出版
小	国語	3	国語 三下 あおぞら	光村図書出版
小	国語	4	国語 四上 かがやき	光村図書出版
小	国語	4	国語 四下 はばたき	光村図書出版
小	国語	5	国語 五上 銀河	光村図書出版
小	国語	5	国語 五下 大地	光村図書出版
小	国語	6	国語 六上 創造	光村図書出版
小	国語	6	国語 六下 希望	光村図書出版
小	国語	1	しよしゃ 一ねん	光村図書出版
小	国語	2	しよしゃ 二年	光村図書出版
小	国語	3	書写 三年	光村図書出版
小	国語	4	書写 四年	光村図書出版
小	国語	5	書写 五年	光村図書出版
小	国語	6	書写 六年	光村図書出版
小	数学	1	新編 あたらしいさんすう 1	東京書籍
小	数学	2	新編 新しい算数 2上	東京書籍
小	数学	2	新編 新しい算数 2下	東京書籍
小	数学	3	新編 新しい算数 3上	東京書籍
小	数学	3	新編 新しい算数 3下	東京書籍
小	数学	4	新編 新しい算数 4上	東京書籍
小	数学	4	新編 新しい算数 4下	東京書籍
小	数学	5	新編 新しい算数 5上	東京書籍
小	数学	5	新編 新しい算数 5下	東京書籍
小	数学	6	新編 新しい算数 6上	東京書籍
小	数学	6	新編 新しい算数 6下	東京書籍
小	理科	3	新編 新しい理科 3	東京書籍
小	理科	4	新編 新しい理科 4上	東京書籍
小	理科	4	新編 新しい理科 4下	東京書籍
小	理科	5	新編 新しい理科 5上	東京書籍
小	理科	5	新編 新しい理科 5下	東京書籍
小	理科	6	新編 新しい理科 6上	東京書籍
小	理科	6	新編 新しい理科 6下	東京書籍
小	社会	3・4	新編 新しい社会 3・4上	東京書籍
小	社会	3・4	新編 新しい社会 3・4下	東京書籍

小	社会	5	新編 新しい社会 5上	東京書籍
小	社会	5	新編 新しい社会 5下	東京書籍
小	社会	6	新編 新しい社会 6上	東京書籍
小	社会	6	新編 新しい社会 6下	東京書籍
小	技術家庭	5・6	新編 新しい家庭 5・6	東京書籍
小	芸術	1	小学生のおんがく 1	教育芸術社
小	芸術	2	小学生の音楽 2	教育芸術社
小	芸術	3	小学生の音楽 3	教育芸術社
小	芸術	4	小学生の音楽 4	教育芸術社
小	芸術	5	小学生の音楽 5	教育芸術社
小	芸術	6	小学生の音楽 6	教育芸術社
小	芸術	1・2	ずがこうさく 1・2上 みつけたよ	日本文教出版
小	芸術	1・2	ずがこうさく 1・2下 たのしいな	日本文教出版
小	芸術	3・4	図画工作 3・4上 ためしながら	日本文教出版
小	芸術	3・4	図画工作 3・4下 たしかめながら	日本文教出版
小	芸術	5・6	図画工作 5・6上 自信をもって	日本文教出版
小	芸術	5・6	図画工作 5・6下 思いを広げて	日本文教出版
小	保健体育	3・4	新・みんなのほけん 3・4年	学習研究社
小	保健体育	5・6	新・みんなの保健 5・6年	学習研究社
小	生活	1・2	新編 あたらしいせいかつ 1・2上	東京書籍
小	生活	1・2	新編 あたらしいせいかつ 1・2下	東京書籍
中	国語	1	国語 1	光村図書出版
中	国語	2	国語 2	光村図書出版
中	国語	3	国語 3	光村図書出版
中	国語	1	中学書写 一年	光村図書出版
中	国語	2・3	中学書写 二・三年	光村図書出版
中	数学	1	新しい数学 1	東京書籍
中	数学	2	新しい数学 2	東京書籍
中	数学	3	新しい数学 3	東京書籍
中	理科	1・2	新しい科学 1分野上	東京書籍
中	理科	2・3	新しい科学 1分野下	東京書籍
中	理科	1・2	新しい科学 2分野上	東京書籍
中	理科	2・3	新しい科学 2分野下	東京書籍
中	社会	1・2	新しい社会 地理	東京書籍
中	社会	1・2	新しい社会 歴史	東京書籍
中	社会	3	新しい社会 公民	東京書籍
中	外国語	1	NEW HORIZON English Course 1	東京書籍
中	外国語	2	NEW HORIZON English Course 2	東京書籍
中	外国語	3	NEW HORIZON English Course 3	東京書籍

中	技術家庭	1・2・3	新しい技術・家庭 技術分野	東京書籍
中	技術家庭	1・2・3	新しい技術・家庭 家庭分野	東京書籍
中	芸術	1	中学生の音楽 1	教育芸術社
中	芸術	2・3	中学生の音楽 2・3 上	教育芸術社
中	芸術	2・3	中学生の音楽 2・3 下	教育芸術社
中	芸術	1・2・3	中学生の器楽	教育芸術社
中	芸術	1	美術1 感じる心で	日本文教出版
中	芸術	2・3	美術2・3上 創造の喜び	日本文教出版
中	芸術	2・3	美術2・3下 美術の広がり	日本文教出版
中	保健体育	1・2・3	中学保健体育	学習研究社
高	国語		国語表現Ⅰ	教育出版
高	国語		国語表現Ⅱ	教育出版
高	国語		国語総合	教育出版
高	国語		現代文	教育出版
高	国語		高等学校 古典 古文編	第一学習社
高	国語		高等学校 古典 漢文編	第一学習社
高	国語		高等学校 古典講読 大鏡 源氏物語 史記	第一学習社
高	数学		数学基礎	実教出版
高	数学		数学Ⅰ	数研出版
高	数学		数学Ⅱ	数研出版
高	数学		数学Ⅲ	数研出版
高	数学		数学A	数研出版
高	数学		数学B	数研出版
高	数学		数学C	数研出版
高	理科		理科基礎 自然のすがた・科学の見かた	東京書籍
高	理科		高等学校 理科総合A	新興出版社啓林館
高	理科		新編理科総合B	東京書籍
高	理科		高等学校物理Ⅰ	数研出版
高	理科		高等学校物理Ⅱ	数研出版
高	理科		高等学校化学Ⅰ	新興出版社啓林館
高	理科		高等学校化学Ⅱ	数研出版
高	理科		高等学校生物Ⅰ	第一学習社
高	理科		高等学校生物Ⅱ	第一学習社
高	理科		高等学校地学Ⅰ	新興出版社啓林館
高	理科		高等学校地学Ⅱ	新興出版社啓林館
高	社会		要説世界史	山川出版社
高	社会		詳説 世界史	山川出版社
高	社会		現代の日本史	山川出版社
高	社会		詳説日本史	山川出版社
高	社会		高等学校 新地理A 最新版	帝国書院

高	社会		新詳地理B 最新版	帝国書院
高	社会		現代社会	東京書籍
高	社会		倫理	東京書籍
高	社会		政治・経済	東京書籍
高	外国語		Hello there! Oral Communication I	東京書籍
高	外国語		empathy oral communication II	教育出版
高	外国語		Crown English Series I	三省堂
高	外国語		Crown English Series II	三省堂
高	外国語		Crown English Reading	三省堂
高	外国語		POLESTAR Writing Course	数研出版
高	技術家庭		家庭基礎 自分らしい生き方とパートナーシップ	実教出版
高	技術家庭		生活技術	教育図書
高	芸術		音楽I Tutti	教育出版
高	芸術		音楽II Tutti	教育出版
高	芸術		音楽III	教育出版
高	芸術		高校美術1	日本文教出版
高	芸術		高校美術2	日本文教出版
高	芸術		高校美術3	日本文教出版
高	芸術		高等学校 工芸I	日本文教出版
高	芸術		高等学校 工芸II	日本文教出版
高	芸術		書道 I	東京書籍
高	芸術		書道 II	東京書籍
高	芸術		書道 III	東京書籍
高	保健体育		現代保健体育	大修館書店
高	情報		Create information 新版情報A	実教出版
高	情報		Information & Solution 新版情報B	実教出版
高	情報		Communication & Collaboration 新版情報C	実教出版

表1の「教科」は、小学校・中学校・高校の各学習指導要領（平成10～11年文部省告示、平成15年一部改正）に定める教科に基づいて設定されている。「教科書コーパス」の教科と各学習指導要領の教科の対応関係を表2に示す。

表2 教科の対応関係

教科書コーパス	学習指導要領		
	小学校	中学校	高等学校
国語	国語	国語	国語
数学	算数	数学	数学
理科	理科	理科	理科
社会	社会	社会	地理歴史
			公民
外国語		外国語	外国語
技術家庭	家庭	技術・家庭	家庭
芸術	音楽	音楽	芸術
	図画工作	美術	
保健体育	体育	保健体育	保健体育
情報			情報
生活	生活		

3. コーパスの形式と公開

全文コーパスとしての教科書コーパスは XML による構造化テキストとして作成した。その文字入力と構造化タグの仕様は、本節のⅡに掲載した。この仕様は、特定領域研究「日本語コーパス」データ班が定める BCCWJ のそれにしている部分が多いが、教科書コーパス作成の目的や、想定される利用法などに鑑み、一部に変更を加えている。

こうして作成した教科書コーパスは、全文であるため、不特定多数の人への全体の公開は、教科書の著作権者から許諾が得にくいと考えられる。そこで、教科書コーパス自体は一般には公開しないこととし、国立国語研究所内に蓄積し、これを用いた語彙表・漢字表は一般に公開することとして、著作権者から許諾を得た。なお、BCCWJ の非母集団サブコーパスに含まれる教科書サンプルは、教科書コーパスからランダムにサンプルを抽出したもののうち、著作権上の問題がないと判断したものである。

Ⅱ. 教科書コーパスの文字入力・タグ仕様

ここでは、「教科書コーパス」の文字入力・タグ仕様について記述する。

教科書コーパスの文字入力・タグ仕様は、山口ほか（2008）に定めた BCCWJ の電子化フォーマットの仕様に基づき、その一部を改変した。

1. 文字入力仕様

教科書コーパスの文字入力仕様は、山口ほか（2008）の「第2章 文字入力仕様」に準拠した。

2. タグ仕様

教科書コーパスは XML を用いて記述した。XML タグの仕様は、山口ほか（2008）の「第3章 タグ仕様」に基づいて、教科書コーパスの入力に適したタグセットとなるよう、以下のタグを省略・追加した。

省略したタグ

abstract, authorsData, blockEnd, contents, cursive, delete, info, insert, list, listItem, orphanedTitle, paragraph, profile, quotation, quote, source, speaker, speech, table, titleBlock, verse, verseLine

追加したタグ

root, skippedSpan, skippedBlock, subRuby, supplement

また、タグの一部については定義の変更を行った。よって、山口ほか（2008）にあげるタグと同名のタグであっても、定義が異なる場合がある。

このようにして定義された教科書コーパスのタグ 32 種を表 3 に示す。

これらのタグについて、以下で詳しく説明する。参照の便宜上、説明はタグのアルファベット順に掲載する。なお、形式化の例示に際し、当該タグの説明に必要なないタグは適宜省略する。

表3 教科書コーパス タグ一覧

■言語的な階層構造を記述する要素	
book	教科書1冊に相当するまとまりを表す。
cluster	検定教科書の目次に示された章題・節題等により包括される、ひとまとまりの文書要素を表す。
title	cluster 要素や supplement 要素の内容を代表する記述を表す。
sentence	文に相当するまとまりを表す。
■特定の言語構造を記述する要素	
□図表に関する要素	
figureBlock	figure 要素と caption 要素をまとめて表す。
figure	図表・写真・絵等を表す。
caption	図表・写真・絵等に付されたタイトルや説明の文章を表す。
□注記に関する要素	
noteMarker	注を参照させるためのマーカーを表す。
noteBody	脚注・後注等、本文と区別して記述される注記を表す。
noteBodyInline	行間または行内に前後の本文と別の形態で示された注記を表す。
□引用・著作権処理に関する要素	
citation	他文献からの引用要素を表す。
copyright	citation 要素以外で、著作権処理が別途必要であった文書要素を表す。
rejectedBlock	著作権者の使用許諾が得られず削除対象となった文書要素を表す。
□『現代日本語書き言葉均衡コーパス』収録サンプルに関する要素	
sample	『現代日本語書き言葉均衡コーパス』のサンプルとされた文書要素を表す。
article	sample 要素内の、同一著者による、同一テーマのひとまとまりの文書要素を表す。
sampling	サンプリングポイントに関する情報を表す。
□その他の要素	
supplement	教科書の主文（主要な学習内容）とは異なる書式を施され、かつ主文に対して特定の機能を有し、その機能ごとに包括されるひとまとまりの文書要素を表す。
skippedBlock	教科書コーパスの語彙表の作成に際し、語彙調査対象外となる文書要素を表す。
■文字・表記に関する要素	
missingCharacter	≡で表現されている文字が JIS X 0213:2004 で規定されている文字以外の文字であることを示す。
surrogatePair	≡で表現されている文字が、JIS X 0213:2004 で規定されているがサロゲートペアを使用する文字であることを示す。
image	規格外のため入力できない記号・絵文字等を表す。
correction	原文の誤植を訂正したことを表す。
ruby	横書きの場合本文上側、縦書きの場合本文右側の行間にルビを付された本文中の文字を表す。
subRuby	横書きの場合本文下側、縦書きの場合本文左側の行間にルビを付された本文中の文字を表す。
enclosedCharacter	連続ラベル・参照ラベル、または特定の語の略記号として機能する囲み付きの文字列を表す。
superScript	数式・化学式等に用いられる上付き文字を表す。
subscript	数式・化学式等に用いられる下付き文字を表す。
fraction	分数部分のうち、マークアップしなければ分数である範囲が誤って解釈されてしまう恐れのある部分を表す。
root	√のかかった部分のうち、マークアップしなければ√のかかる範囲が誤って解釈されてしまう恐れのある部分を表す。
br	論理改行を表す。
rejectedSpan	文中にあって入力対象外となった数式・化学反応式・外国語等の部分を表す。
skippedSpan	教科書コーパスの語彙表作成に際し、語彙調査対象外となる文中の文字列を表す。

article

sample 要素内の同一著者（単著では一人、共著では複数）による、同一テーマのひとまとまりの文書要素を表す。

属性

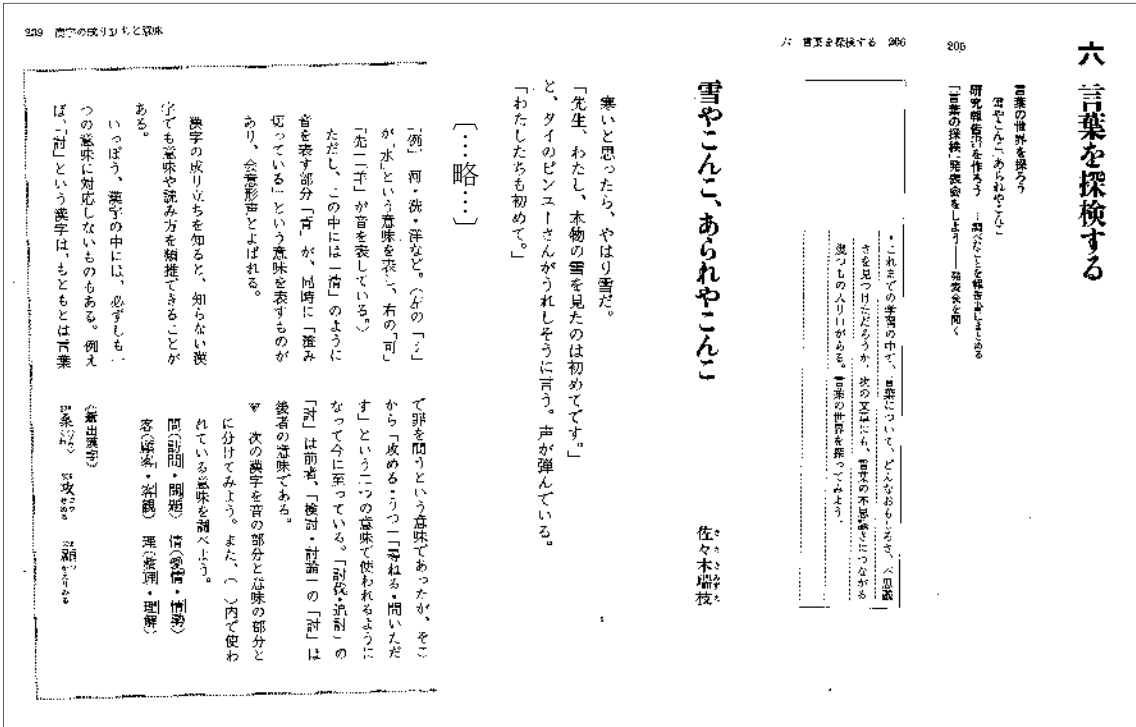
isWholeArticle 属性に、本来収録すべき文書範囲全体が収録できた否かを示す。"true"は本来収録すべき文書範囲全体が収録できたことを、"false"は収録文字数の制限などにより全体が収録できなかったことを示す。

inSample 属性に、sample 要素内の要素であるか否かを示す。"yes"は sample 要素内の要素であることを示す。

accID 属性に、book 要素内での article 要素・citation 要素・copyright 要素の通し番号を示す。

形式化例

原文



『国語 1』(光村図書出版 2005) p.205, 239

形式化

```
<sample sampleID="OT02_00019" type="variableLength">
<article isWholeArticle="false" inSample="yes" accID="21">
六 言葉を探検する
言葉の世界を探ろう
雪やこんこ、あられやこんこ
研究報告書を作ろう—調べたことを報告書にまとめる
「言葉の探検」発表会をしよう—発表会を開く
・これまでの学習の中で、言葉について、どんなおもしろさ、不思議さを見つけたら
うか。次の文章にも、言葉の不思議さにつながる幾つもの入り口がある。言葉の世界を
探ってみよう。
<citation>
雪やこんこ、あられやこんこ
佐々木瑞枝
寒いと思ったら、やはり雪だ。
「先生、わたし、本物の雪を見たのは初めてです。」
<sampling type="start" />と、タイのピンユーさんがうれしそうに言う。声が弾んでい
る。
「わたしたちも初めて。」
[…略…]
</citation>
[…略…]
《新出漢字》
2 3 8 桑
(ソウ)
くわ
2 3 9 攻
コウ
せめる
2 3 9 顧
コ
かえりみる
</article>
</sample>
```

book

教科書一冊を表す。

属性

bibID 属性に、教科書 1 冊ごとの ID を示す。

形式化例

省略。

br

論理改行を表す。空要素である。

属性

なし。

形式化例

原文

三線は、^{沖縄}や^{奄美}地方などで使われる、三味線と似た構造の楽器です。一般に^{ばん}胴の表面にニシキヘビの皮が張られているのが特徴です。ちなみに三味線には元来猫や犬の皮が張られていましたが、現在では合成皮革も多く使われています。

中国から渡来した三弦が三線となり、それがさらに変化して三味線になったものと考えられています。だとすれば、三線は三味線の「^{せん}先輩」ということとなります。三線は現在でも沖縄地方の音楽で重要な役割を果たしています。特に民謡の伴奏楽器としては欠かすことのできない楽器といえましょう。

『中学生の器楽』（教育芸術社 2005） p.44

形式化

[…略…] ちなみに三味線には元来猫や犬の皮が張られていましたが、現在では合成皮革も多く使われています。

中国から渡来した三弦が三線となり、それがさらに変化して三味線になったものと考えられています。だとすれば、三線は三味線の「先輩」ということとなります。三線は現在でも沖縄地方の音楽で重要な役割を果たしています。特に民謡の伴奏楽器としては欠かすことのできない楽器といえましょう。

caption

図表・写真・絵等に付されたタイトルや説明の文章を表す。

属性

なし。

形式化例

原文



『新しい社会 公民』（東京書籍 2005）p.101

形式化

```
<figureBlock>  
<figure/>  
<caption>  
↑ 現金支払機を設置したコンビニ店  
</caption>  
</figureBlock>
```

citation

book 要素内で、book 要素の著者以外の著者が著した文書要素を表す。

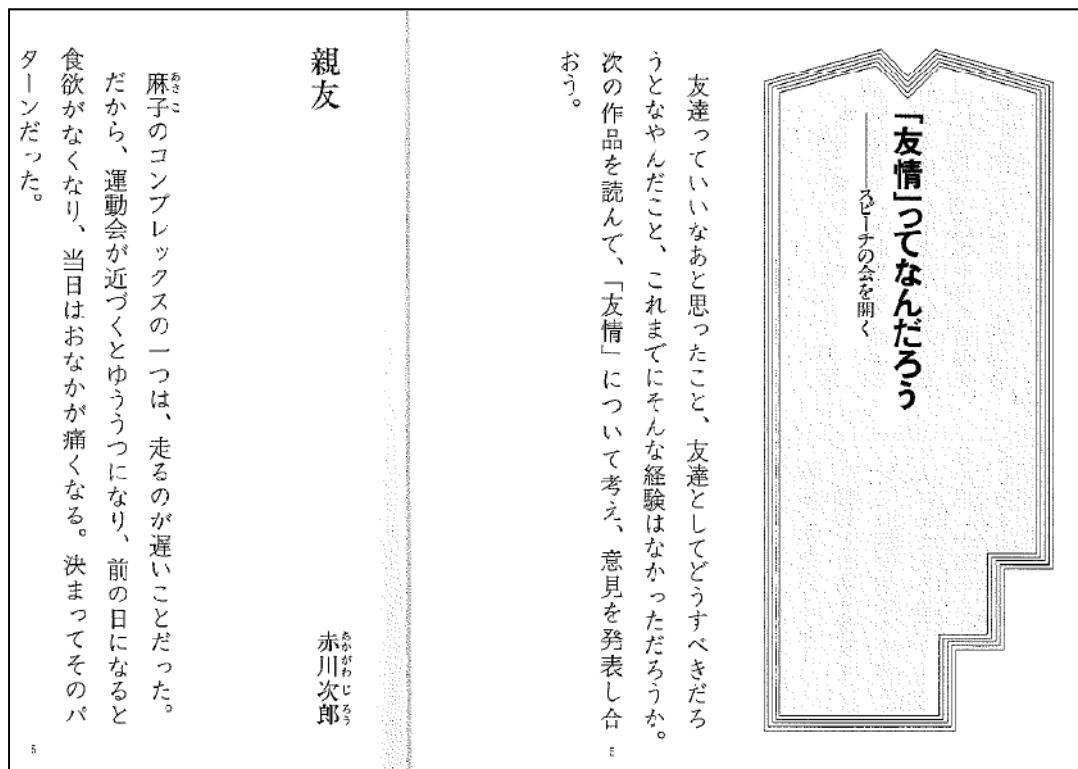
属性

inSample 属性に、sample 要素内の要素であるか否かを示す。"yes"は sample 要素内の要素であることを、"no"は sample 要素外の要素であることを示す。

accID 属性に、book 要素内での article 要素・citation 要素・copyright 要素の通し番号を示す。

形式化例

原文



『国語 1』（光村図書出版 2005）pp.24-25

形式化

「友情」ってなんだろう
—スピーチの会を開く
友達っていいなあと思ったこと、友達としてどうすべきだろうとなやんだこと、これまでにそんな経験はなかっただろうか。次の作品を読んで、「友情」について考え、意見を発表し合おう。
<citation inSample="yes" accID="6">
親友
赤川次郎
麻子のコンプレックスの一つは、走るのが遅いことだった。
だから、運動会が近づくとゆううつになり、前の日になると食欲がなくなり、当日はおなかが痛くなる。決まってそのパターンだった。
[…略…]
</citation>

cluster

検定教科書の目次に示された章題・節題等により包括される、ひとまとまりの文書を表す。cluster 要素の入れ子構造により、文書の論理的な階層構造を表現する。cluster 要素の

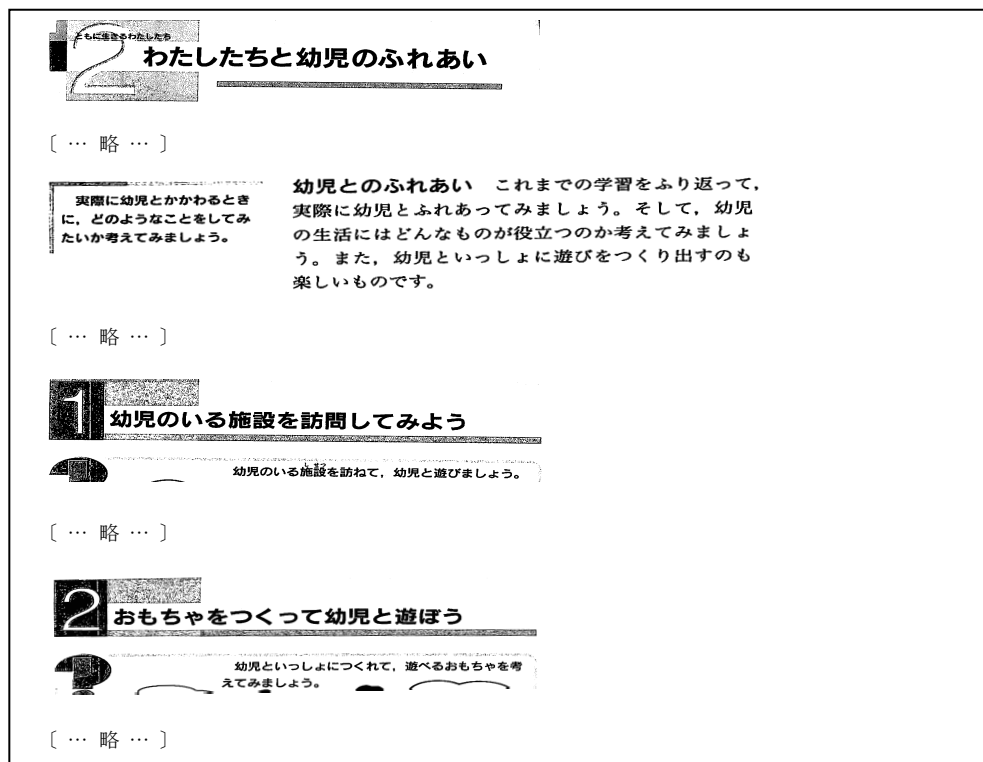
最下位は、各教科書の目次の最下位と原則一致する。

属性

なし。

形式化例

原文



『新しい技術・家庭 家庭分野』（東京書籍 2005）p.158, 160, 162

形式化

```
<cluster>
<title>
  2 わたしたちと幼児のふれあい
</title>
  [...]
  実際に幼児とかかわるときに、どのようなことをしてみたいか考えてみましょう。
  [...]
</cluster>
<cluster>
<title>
  1 幼児のいる施設を訪問してみよう
</title>
  幼児のいる施設を訪ねて、幼児と遊びましょう。
  [...]
</cluster>
<cluster>
<title>
  2 おもちゃをつくって幼児と遊ぼう
</title>
  幼児といっしょにつくって、遊べるおもちゃを考えてみましょう。
  [...]
</cluster>
</cluster>
```

copyright

book 要素内で、book 要素と同一の著者が著した文書であるが、別途著作権処理が必要であった要素を表す。

属性

inSample 属性に、sample 要素内の要素であるか否かを示す。"yes"は sample 要素内の要素であることを、"no"は sample 要素外の要素であることを示す。

accID 属性に、book 要素内での article 要素・citation 要素・copyright 要素の通し番号を示す。

形式化例

省略

correction

原文の誤植を訂正したことを表す。type 属性の値が"erratum"、"omission"の場合、訂正したテキストを要素とする。type 属性の値が"excess"の場合、空要素である。

属性

`type` 属性に、誤植の種類を示す。"erratum"は誤字を（形式化例 1 参照）、"excess"は衍字を（形式化例 2 参照）、"omission"は脱字を（形式化例 3 参照）示す。

`originalText` 属性に、訂正前のテキストを示す。`type` 属性の値が"omission"の場合、`originalText` 属性は不要。

形式化例 1

原文

化石燃料・再生可能エネルギーの有効利用に向けた研究

『理科基礎 自然のすがた・科学の見かた』（東京書籍 2006） p.195

形式化

化石燃<correction type="erratum" originalText="科">料</correction>・再生可能エネルギーの有効利用に向けた研究

形式化例 2

原文

とくに周囲の部分にとぎれのないように注意してして入れる

『高等学校 地学Ⅱ』（啓林館 2005） p.242

形式化

とくに周囲の部分にとぎれのないように注意して<correction type="excess" originalText="し" /><correction type="excess" originalText="て" />入れる

形式化例 3

原文

外液の濃度を急激に変化させることによって菌体を破裂させ、

『高等学校 生物Ⅱ』（第一学習社 2006） p.15

形式化

外液の濃度を急激に変化<correction type="omission">さ</correction>せることによって菌体を破裂させ、

enclosedCharacter

連続ラベル・参照ラベル、または特定の語の略記号として機能する囲み付きの文字列を表す。

連続ラベル・参照ラベルとは、以下のようなものを指し、多くは丸付きの数字や仮名・アルファベット等で示される（形式化例 1 参照）。

- ・連続する項に対して順に与えられる番号・記号等。
- ・離れた要素同士を参照させるために与えられる同一の番号・記号等。

略記号とは、多くの場合特定の語の冒頭一文字を取り、これに囲みを付ける等の強調表示を行って、その語の代用表記として用いるようなものを指す（形式化例2参照）。

属性

なし。

形式化例 1

原文

- ③ 次の組み立てにしたがって、レポートの後半部分を書こう。
- ① どこで生まれたか。どのような子どもだったか。
- ② 将来何になりたかったか。そして結局何になったか。



『NEW HORIZON English Course 3』（東京書籍 2005） p.70

形式化

● 次の組み立てにしたがって、レポートの後半部分を書こう。
 <enclosedCharacter> 1 </enclosedCharacter>どこで生まれたか。どのような子どもだったか。
 <enclosedCharacter> 2 </enclosedCharacter>将来何になりたかったか。そして結局何になったか。

形式化例 2

原文

⑩ 図10 電気用図記号と回路図 電流計・電圧計の \Rightarrow は、直流を表す図記号。直流・交流の区別をしない場合は、 \textcircled{A} 、 \textcircled{V} と表す。

『新しい科学 1 分野上』（東京書籍 2005） p.91

形式化

[…略…] 直流・交流の区別をしない場合は、<enclosedCharacter> A </enclosedCharacter>、<enclosedCharacter> V </enclosedCharacter> と表す。

figure

図表・写真・絵等を表す。図表等は原則として入力対象外であるため、多くの場合、空要素である（形式化例1参照）。ただし、図表等の内部に特別多く文章が存在する場合、この文章は入力対象とし、figure 要素として示す（形式化例2参照）。

属性

なし。

形式化例 1

原文



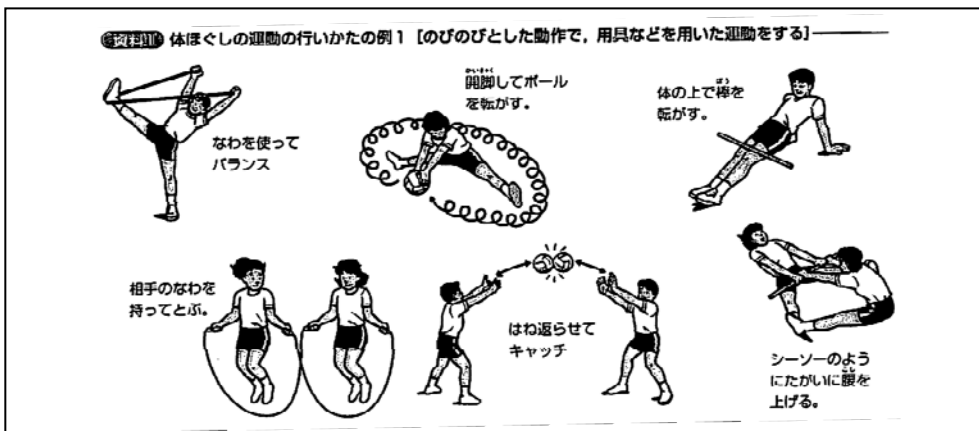
『新しい社会 公民』(東京書籍 2005) p.101

形式化

```
<figureBlock>  
<figure/>  
<caption>  
↑ 現金支払機を設置したコンビニ店  
</caption>  
</figureBlock>
```

形式化例 2

原文



『中学保健体育』(学習研究社 2005) p.22

形式化

```
<figureBlock>
<figure>
なわを使ってバランス
開脚してボールを転がす。
体の上で棒を転がす。
相手のなわを持ってとぶ。
はね返らせてキャッチ
シーソーのようにたがいに腰を上げる。
</figure>
<caption>
資料1 体ほぐしの運動の行いかたの例1 [のびのびとした動作で、用具などを用いた
運動をする]
</caption>
</figureBlock>
```

figureBlock

figure 要素として示される図表・写真・絵等と、それに付随し caption 要素として示される文章をまとめて表す。なお、figureBlock 要素は、原紙での図表等の正確な位置を必ずしも表すものではない。

属性

なし。

形式化例

原文



『新しい社会 公民』（東京書籍 2005）p.101

形式化

```
<figureBlock>
<figure/>
<caption>
↑ 現金支払機を設置したコンビニ店
</caption>
</figureBlock>
```

fraction

次に挙げるような、マークアップしなければ分数である範囲が誤って解釈されてしまう恐れのある分数部分を表す。

- ・ 帯分数における真分数部分（形式化例 1 参照）
 - ・ 分母と分子の片方または両方が、括弧等で囲まれていない 2 文字以上の文字列や数式から成っている場合（形式化例 2 参照）
 - ・ 分数の直前または直後に、分数に続く形で式が存在している場合（形式化例 3 参照）
- 上記以外の分数部分はマークアップしない（形式化例 4 参照）。

属性

なし。

形式化例 1

原文

ウスターソース … 8 mℓ (小 1 $\frac{3}{5}$)

『新しい技術・家庭 家庭分野』（東京書籍 2005） p.38

形式化

ウスターソース… 8 m ℓ (小 1 <fraction> 3 / 5 </fraction>)

形式化例 2

原文

速度 [m/秒] = $\frac{\text{移動距離 [m]}}{\text{移動するのにかった時間 [秒]}}$

『新しい科学 1 分野下』（東京書籍 2005） p.37

形式化

速度 [m/秒] = <fraction>移動距離 [m] / 移動するのにかった時間 [秒] </fraction>

形式化例 3

原文

$$V = \frac{1}{3}Sh$$

『新しい数学 1』（東京書籍 2005） p.168

形式化

$$V = \langle \text{fraction} \rangle 1 / 3 \langle / \text{fraction} \rangle S h$$

形式化例 4

原文

じ高さにして、液面
の下のところを1め
もりの $\frac{1}{10}$ まで読みと
る。

『新しい科学 1 分野上』（東京書籍 2005） p.61

形式化

[…略…] 液面の下のところを1めもりの $1 / 10$ まで読みとる。

image

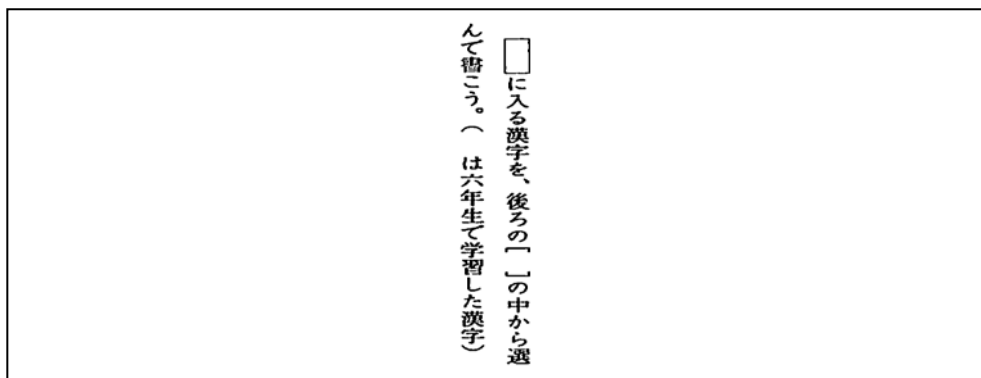
JIS X 0213:2004 がない記号・絵文字を表す。また、記号・絵文字ではなく、絵と見なされるものの中で、教科書の凡例等で学習上の意味が特に明記されているものについても、image 要素として表す（形式化例 2 参照）。空要素である。

属性

description 属性に、形状等の説明を示す。

形式化例 1

原文



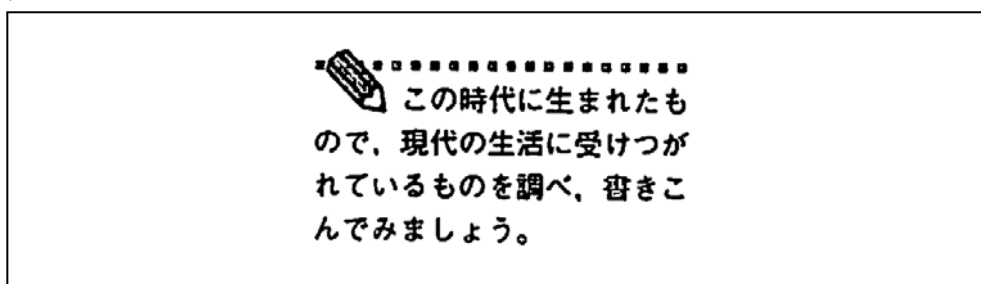
『国語 2』(光村図書出版 2005) p.225

形式化

<image description="枠付き空欄" />に入る漢字を、後ろの [] の中から選んで書こう。[…略…]

形式化例 2

原文



『新しい社会 歴史』(東京書籍 2005) p.64

形式化

<image description="えんぴつマーク" />この時代に生まれたもので、現代の生活に受けつがれているものを調べ、書きこんでみましょう。

missingCharacter

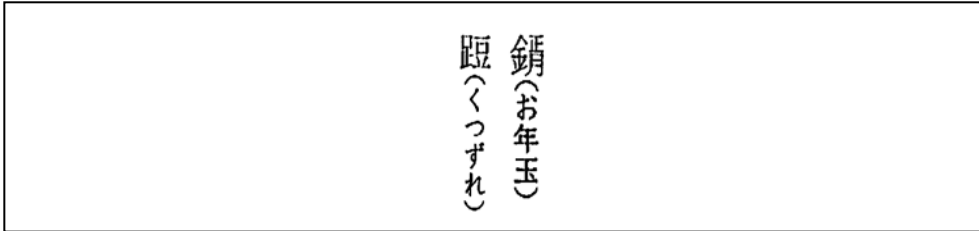
≡で表現されている文字が、JISX0213:2004 で規定されている文字以外の文字であることを示す。missingCharacter 要素で表したのは、漢字・仮名・ラテン文字等の文字である。JISX0213:2004 で規定されていない記号・絵文字は、image 要素で表す。字形・字体の扱いは、JISX0213:2004 の規格に従うことを原則とする。ただし、JISX0213:2004 の規格では規定されている文字に置き換えられる字形・字体であっても、区別することに意味があると考えられるものについては、missingCharacter 要素で表す (形式化例 3~4 参照)。

属性

description 属性に、字体についての説明を示す。

形式化例 1

原文



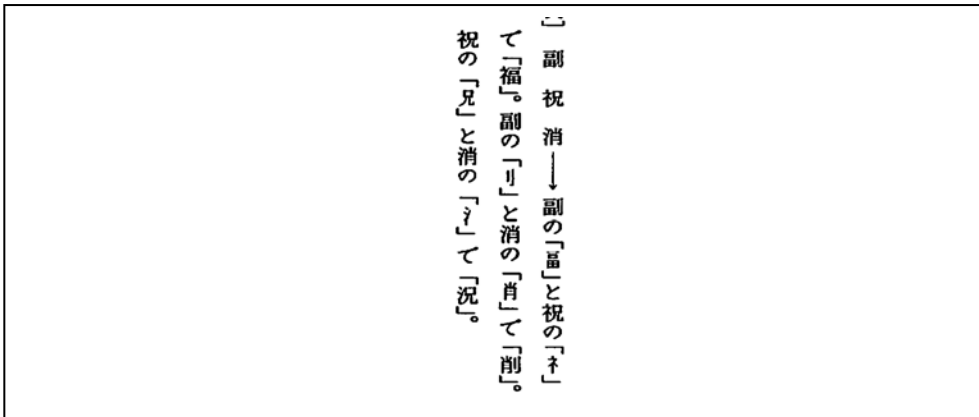
『国語 3』(光村図書出版 2005) p.226

形式化

```
<missingCharacter description="金偏+右上に「正」+右下に「月」">= </missingCharacter> (お年玉)  
<missingCharacter description="足偏+豆">= </missingCharacter> (くつずれ)
```

形式化例 2

原文



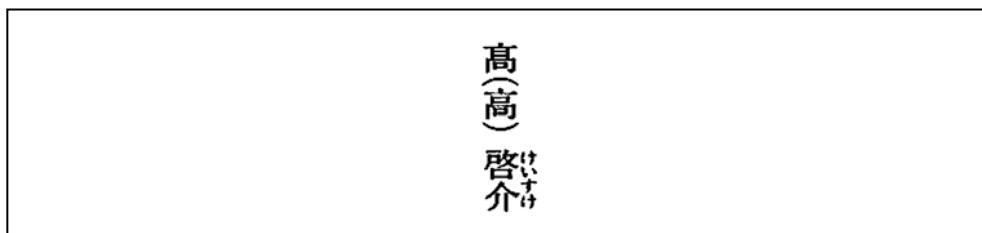
『国語 2』(光村図書出版 2005) p.37

形式化

```
[…略…] 副 祝 消 → 副の「<missingCharacter description="「副」の左側">= </missingCharacter>」と祝の「ネ」で「福」。副の「冫」と消の「肖」で「削」。祝の「兄」と消の「冫」で「況」。
```

形式化例 3

原文



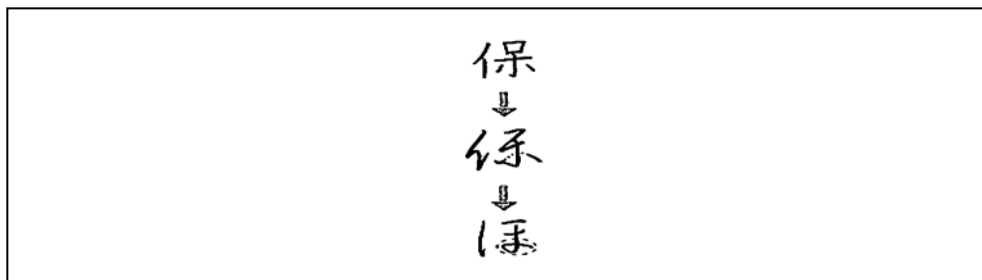
『中学書写 二・三年』（光村図書出版 2005） p.37

形式化

```
<missingCharacter description="はしご高"> = </missingCharacter> (高) 啓介
```

形式化例 4

原文



『中学書写 二・三年』（光村図書出版 2005） p.27

形式化

```
保→<missingCharacter description="「保」の草書"> = </missingCharacter>→ほ
```

noteBody

脚注・後注等、本文と区別して記述される注記を表す。傍注等のインラインの注記については、noteBodyInline 要素として別途マークアップする。

属性

なし。

形式化例

原文

*3 駅には、役人が乗りつ
ぐための馬などが用意されて
いました。

『新しい社会 歴史』（東京書籍 2005） p.37

形式化

```
<noteBody>  
* 3 駅には、役人が乗りつぐための馬などが用意されてい  
</noteBody>
```

noteBodyInline

行間に（または行内に前後の本文と別の形態で）示された注記を表す。多くは注記対象となる本文行内の要素の傍にルビのような形で示される。空要素である。

属性

text 属性に、注記部分の文字列を示す。

info 属性に、注記部分の文字列に **enclosedCharacter** 要素、**image** 要素等の文字情報要素に相当するものが含まれる場合、その情報を示す（形式化例 4 参照）。書式は以下の通り。

要素名:属性名=属性値:当該要素の text 属性値での開始文字位置-当該要素の text 属性値に対する文字数

複数の文字情報要素が共起する場合、:で連結して併記する。

形式化例 1

原文

とおるさんは、^{いわて}岩手県の伝統産業と近代
^{pp.190}工業を調べるという^{してん}視点で、なにを調べ

『新しい社会 地理』（東京書籍 2005） p.68

形式化

```
とおるさんは、岩手県の伝統産業と近代工業<noteBodyInline text="→ p. 190" />を調  
べるという視点で、[…略…]
```

形式化例 2

原文

自己理解のためには、心の働きについて知っておくことも役立つ。その代表的なものとして、まず^{complex}コンプレックスがあげられる。コンプレックスとは、

『倫理』（東京書籍 2006） p.17

形式化

自己理解のためには、心の働きについて知っておくことも役立つ。その代表的なものとして、まずコンプレックス<noteBodyInline text="complex"/>があげられる。
[…略…]

形式化例 3

原文

柿主 さればこそ、顔を（かくした。）かくいた。あの柿の木のかげへかくれたを、（よくよく）ようよう見れば、人ではないと見えた。

『国語 六（上） 創造』（光村図書出版 2006） p.121

形式化

柿主 さればこそ、顔を（かくした。）かくいた。<noteBodyInline text="（かくした。）"/>あの柿の木のかけへかくれたを、ようよう<noteBodyInline text="（よくよく）"/>見れば、人ではないと見えた。

形式化例 4

原文

さいしき あるいはせんこく へき が そうしよくこ ふん
彩色あるいは線刻された壁画を持つ装飾古墳
(→口絵④)

『詳説日本史』（山川出版社 2006） p.25

形式化

```
彩色あるいは線刻された壁画を持つ装飾古墳<noteBodyInline text=" (→口絵4) " info="enclosedCharacter:5-1"/>
```

noteMarker

注を参照させるためのマーカーを表す。多くは、出現順を示す番号や「*」等の記号が用いられている。本文行内の要素に対し、肩付き・下付き等の形式で付される要素をマークアップの対象とする。空要素である。

属性

text 属性に、注記部分の文字列を示す。

info 属性に、注記部分の文字列に enclosedCharacter 要素、image 要素等の文字情報要素に相当するものが含まれる場合、その情報を示す（形式化例 2 参照）。書式は以下の通り。

要素名:属性名=属性値:当該要素の text 属性値での開始文字位置-当該要素の text 属性値に対する文字数

複数の文字情報要素が共起する場合、:で連結して併記する。

形式化例 1

原文

地方は多くの国に区分されて、その国々には役所(国府)
が置かれました。^{*2} 国には中央の貴族が国司として派遣され、
その下で地方の豪族が郡司^{くんと}に任ぜられて政治を行いました。
都と地方を結ぶ道路も整えられ、役人の往来のために駅^{*3}
が設けられました。

『新しい社会 歴史』（東京書籍 2005） p.37

形式化

```
地方は多くの国に区分されて、その国々には役所（国府）が置かれました<noteMarker text=" * 2 " />。国には中央の貴族が国司として派遣され、その下で地方の豪族が郡司に任ぜられて政治を行いました。都と地方を結ぶ道路も整えられ、役人の往来のために駅<noteMarker text=" * 3 " />が設けられました。
```

形式化例 2

原文

水、ブドウ糖、^{↑2}酸素、二酸化炭素など、

『新しい科学 1 分野上』（東京書籍 2005） p.59

形式化

水、ブドウ糖<noteMarker text="2" info="image:description=旗:1-0"/>、酸素、二酸化炭素など、

rejectedBlock

削除対象となったブロック要素の存在を表す。

rejectedBlock 要素とするのは次の①②の 2 種類である。①は空要素、②は削除対象を要素とする。

- ①book 要素内において、著作権者の使用許諾が得られず、削除対象となった要素（形式化例 1 参照）。
- ②sample 要素内において、『現代日本語書き言葉均衡コーパス』のサンプリングの基準では削除対象となる要素（形式化例 2 参照）。

属性

type 属性に、削除対象要素の種別を示す。"copywrite"は著作権者の使用許諾が得られず削除対象となった要素を示す。それ以外の削除対象要素の場合、type 属性は不要。

形式化例 1

原文

<p>モーセの十戒^{じゅうがい} 1～4までは神に対する宗教的義務を示し、5～10は人間社会における道徳的義務を示している。ユダヤ教では、この十戒を中心に数多くの律法が制定された。</p>	<h1>削除対象要素</h1>
---	-----------------

『倫理』（東京書籍 2006） p.49

形式化

<p>モーセの十戒 1～4までは神に対する宗教的義務を示し、5～10は人間社会における道徳的義務を示している。ユダヤ教では、この十戒を中心に数多くの律法が制定された。 <rejectedBlock type="copyright" /></p>
--

形式化例 2

原文

<p style="text-align: right;">読みを深めるために</p> <p>今は昔、摂津の国のわたりより、盗みせむがために京に上りける男の、日のいまだ明かかりければ、羅城門の下に立ち隠れて立てりけるに…… （昔、摂津の国の辺りから、盗みをはたらくために上京してきた男が、まだ日が明るかったので、羅城門の下に隠れるようにして立っていたところ……）</p>
--

『国語総合』（教育出版 2006） p.37

形式化

読みを深めるために

<rejectedBlock>

今は昔、摂津の国のわたりより、盗みせむがために京に上りける男の、日のいまだ明かりければ、羅城門の下に立ち隠れて立てりけるに…

</rejectedBlock>

(昔、摂津の国の辺りから、盗みをはたらくために上京してきた男が、まだ日が明るかったので、羅城門の下に隠れるようにして立っていたところ…)

rejectedSpan

文中にあって入力対象外となった数式・化学反応式・外国語等のインライン要素の存在を表す。空要素である。

属性

なし。

形式化例

原文

この和 S も $\sum_{n=1}^{\infty} a_n$ と書き表す。

『数学Ⅲ』（数研出版 2006）p.38

形式化

この和 S も<rejectedSpan />と書き表す。

root

√のかかった部分のうち、マークアップしなければ√のかかる範囲が誤って解釈されてしまう恐れのある部分を表す。

属性

なし。

形式化例

原文

縦 $\sqrt{2}$ cm, 横 $\sqrt{5}$ cmの長方形の面積を求めるとき、 $\sqrt{2 \times 5}$ と計算してもよいでしょうか。

『新しい数学3』（東京書籍 2005） p.15

形式化

縦 $\sqrt{2}$ cm, 横 $\sqrt{5}$ cmの長方形の面積を求めるとき、 $\sqrt{2 \times 5}$ と計算してもよいでしょうか。

ruby

横書きの場合本文に対して上側、縦書きの場合本文に対して右側の行間にルビを付された本文中の文字を表す。ルビとしたのは、次のようなものである。

- ・振り仮名（形式化例 1～2 参照）
- ・別語による言い換え（形式化例 3 参照）

属性

rubyText 属性に、本文要素に対して行間に小書きされたルビ部分の文字列を示す。

形式化例 1

原文

戯画は、^{へいあん}平安時代になると^{えまきもの}絵巻物に描かれるよう
になりました。中でも^と鳥^ば羽^{そう}僧^{じょう}正作と伝えられる

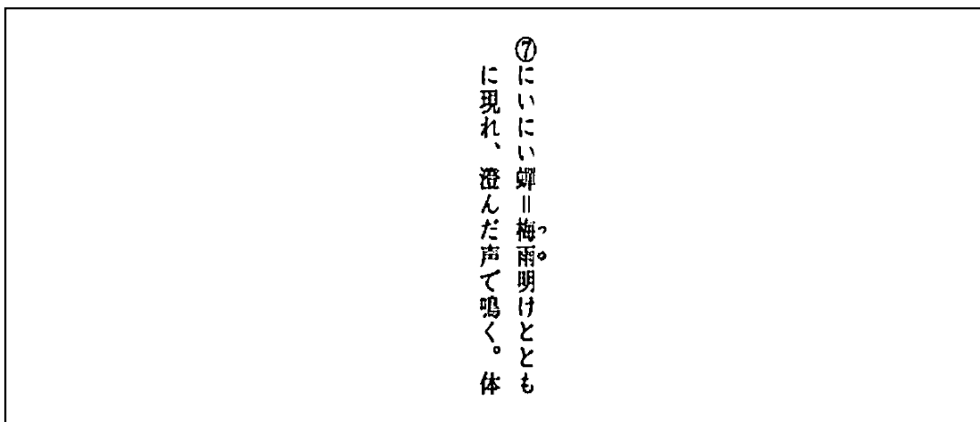
『美術2・3上 創造の喜び』（日本文教出版 2005） p.5

形式化

戯画は、 $\langle\text{ruby rubyText="へい"}\rangle$ 平 $\langle\text{ruby rubyText="あん"}\rangle$ 安時代になると $\langle\text{ruby rubyText="と"}\rangle$ 鳥 $\langle\text{ruby rubyText="ば"}\rangle$ 羽 $\langle\text{ruby rubyText="そう"}\rangle$ 僧 $\langle\text{ruby rubyText="じょう"}\rangle$ 正 $\langle\text{ruby}\rangle$ 作と伝えられる […略…]

形式化例 2

原文



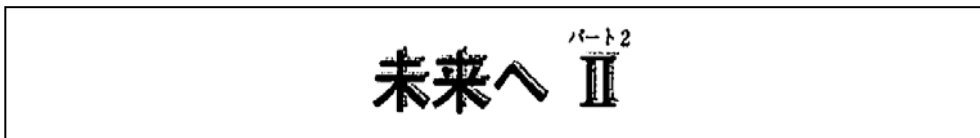
『国語 3』（光村図書出版 2005） p.107

形式化

[…略…] <ruby rubyText="つゆ">梅雨</ruby>明けとともに現れ、澄んだ声で鳴く。[…略…]

形式化例 3

原文



『中学生の音楽 2・3 下』（教育芸術社 2005） p.56

形式化

未来へ<ruby rubyText="パート 2">Ⅱ</ruby>

sample

『現代日本語書き言葉均衡コーパス』のサンプルとされた文書要素を表す。

属性

sampleID 属性に、サンプルごとに付した ID を示す。

type 属性にサンプルの種類を示す。"variableLength"は可変長サンプルを示す。

形式化

```
<sample sampleID="OT02_00019" type="variableLength">
<article isWholeArticle="false" inSample="yes" accID="21">
六 言葉を探検する
言葉の世界を探ろう
雪やこんこ、あられやこんこ
研究報告書を作ろう—調べたことを報告書にまとめる
「言葉の探検」発表会をしよう—発表会を開く
・これまでの学習の中で、言葉について、どんなおもしろさ、不思議さを見つけたら
うか。次の文章にも、言葉の不思議さにつながる幾つもの入り口がある。言葉の世界を
探ってみよう。
<citation>
雪やこんこ、あられやこんこ
佐々木瑞枝
寒いと思ったら、やはり雪だ。
「先生、わたし、本物の雪を見たのは初めてです。」
<sampling type="start" />と、タイのピンユーさんがうれしそうに言う。声が弾んでい
る。
「わたしたちも初めて。」
[...略...]
</citation>
[...略...]
《新出漢字》
2 3 8 桑
(ソウ)
くわ
2 3 9 攻
コウ
せめる
2 3 9 顧
コ
かえりみる
</article>
</sample>
```

sampling

サンプリングポイントに関する情報を表す。サンプル抽出の基準点となった一文字の直前に、空要素として入力する。

属性

type 属性に、サンプリングポイントの種別を示す。"start"はサンプル抽出基準点を示す。

形式化例

sample 要素の形式化例を参照。

sentence

文に相当するまとまりを表す。原則として、句点等の表記上の手がかりに基づき自動認定する。要素認定の詳細については、山口ほか（2008）pp.133-134 を参照。

属性

type 属性に、sentence 要素の種別を示す。"quasi"は文区切り文字以外の基準に基づいて認定された sentence 要素であることを示す。

形式化例

原文

1. 運動は楽しい
体をのびのびと自由に動かすことは、とても気持ちがよく、楽しいことです。この、体を動かす楽しさが、自分の力を試したり、自然に親しんだりすることと結びつくと、いっそう豊かなものになり、運動の楽しさに発展します。

『中学保健体育』（学習研究社 2005）p.6

形式化

```
<sentence type="quasi">1. 運動は楽しい</sentence>  
<sentence> 体をのびのびと自由に動かすことは、とても気持ちがよく、楽しいこと  
です。 </sentence><sentence>この、体を動かす楽しさが、自分の力を試したり、自然に親  
しんだりすることと結びつくと、いっそう豊かなものになり、運動の楽しさに発展しま  
す。 </sentence>
```

skippedBlock

教科書コーパスの語彙表の作成に際し、語彙調査の対象外とするブロック要素を表す。skippedBlock 要素と認定したのは、次のようなものである。

- ・近世以前に著された文章（漢文の読み下し文や古典文法の説明等で使用される語句も含む）（形式化例 1 参照）
- ・漢文
- ・近代以後に著された文語文
- ・方言
- ・漢字やその音訓のリスト（形式化例 2 参照）
- ・重要語句のリスト（形式化例 3 参照）

属性

type 属性に、要素の種別を示す。"kobun"は近世以前に著された文章を、"kanbun"は漢文を、"etc"はその他の種類を示す。

形式化例 1

原文

読みを深めるために
今は昔、摂津の国のわたりより、盗みせむがために京に上りける男の、日のいまだ明かりければ、羅城門の下に立ち隠れて立てりけるに……
(昔、摂津の国の辺りから、盗みをはたらくために上京してきた男が、まだ日が明るかったので、羅城門の下に隠れるようにして立っていたところ……)

『国語総合』(教育出版 2006) p.37

形式化

読みを深めるために

<skippedBlock type="kobun">

今は昔、摂津の国のわたりより、盗みせむがために京に上りける男の、日のいまだ明かりければ、羅城門の下に立ち隠れて立てりけるに……

</skippedBlock>

(昔、摂津の国の辺りから、盗みをはたらくために上京してきた男が、まだ日が明るかったので、羅城門の下に隠れるようにして立っていたところ……)

形式化例 2

原文

70 器(うつわ) 72 承る(うけたまわる)
70 陶(とう) 71 賓(ヒン) 72 謙(ケン)

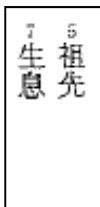
『国語 3』(光村図書出版 2005) p.73

形式化

```
《新出漢字》  
<skippedBlock type="etc">  
7 0 陶トウ  
7 1 賓ヒン  
7 2 謙ケン  
</skippedBlock>  
《新出音訓》  
<skippedBlock type="etc">  
7 0 器（うつわ）  
7 2 承る（うけたまわる）  
</skippedBlock>
```

形式化例 3

原文



『国語 1 』(光村図書出版 2005) p.50

形式化

```
<skippedBlock type="etc">  
5 祖先  
7 生息  
</skippedBlock>
```

skippedSpan

教科書コーパスの語彙表の作成に際し、語彙調査の対象外とするインライン要素を表す。
skippedSpan 要素と認定したのは、次のようなものである。

- ・近世以前に著された文章（漢文の読み下し文や古典文法の説明等で使用される語句も含む）（形式化例 1 参照）
- ・漢文（形式化例 2 参照）
- ・近代以後に著された文語文
- ・方言

属性

type 属性に、要素の種別を示す。"kobun"は近世以前に著された文章を、"kanbun"は漢文を、"etc"はその他の種類を示す。

形式化例 1

原文

⑱ 頭身の毛も太る ぞつ
とするほどの恐怖感の
形容。『今昔物語集』
卷二十七に、「頭身の
毛太るやうにおぼえけ
れば」という表現があ
る。

『国語総合』（教育出版 2006） p.30

形式化

1 8 頭身の毛も太る ぞつとするほどの恐怖感の形容。『今昔物語集』卷二十七に、「<skippedSpan type="kobun">頭身の毛太るやうにおぼえければ</skippedSpan>」という表現がある。

形式化例 2

原文

(2)	(1)
以 _レ 子 _レ 之 _レ 矛 _レ 陷 _レ 子 _レ 之 _レ 楯 _レ 何 _レ 如 _レ 。	無 _レ 不 _レ 陷 _レ 也。
(疑問)	(二重否定)
同・	270
8	ペー
8	ジ
9	7
行	8
	行

『国語総合』（教育出版 2006） p.271

形式化

(1)	<skippedSpan type="kanbun">無不陷也。</skippedSpan>	(二重否定)	270	ペー	ジ
				7	8
				7	8
(2)	<skippedSpan type="kanbun">以子之矛陷子之楯何如。</skippedSpan>	(疑問)	同・		
				8	9
				8	9
				行	行

subRuby

横書きの場合本文に対して下側、縦書きの場合本文に対して左側の行間に振り仮名を付された本文中の文字を表す。

属性

rubyText 属性に、本文要素に対して行間に小書きされたルビ部分の文字列を示す。

形式化例

原文

小国を統一した高句麗が occurred.

『新しい社会 歴史』(東京書籍 2005) p.19

形式化

小国を統一した<subRuby rubyText="こう"><ruby rubyText="コ">高</ruby></subRuby><subRuby rubyText="く"><ruby rubyText="グ">句</ruby></subRuby><subRuby rubyText="り"><ruby rubyText="リョ">麗</ruby></subRuby>が occurred.

subScript

数式・化学式等に用いられる下付き文字を表す。

属性

なし。

形式化例

原文

残った原子は、Hが2個とOが1個なので、水(H₂O)ができていると予想できます。実際に、このとき、水ができています。

『新しい科学 1分野下』(東京書籍 2005) p.26

形式化

残った原子は、Hが2個とOが1個なので、水(H<subScript>2</subScript>O)ができていると予想できます。実際に、このとき、水ができています。

superScript

数式・化学式等に用いられる上付き文字を表す。

属性

なし。

形式化例

原文

1.4², 1.5² を計算してみましょう。

『新しい数学3』（東京書籍 2005）p.9

形式化

1. 4², 1. 5² を計算してみましょう。

supplement

検定教科書の主文（主要な学習内容）とは異なる書式を施され、外形的に主文とは区別され、かつ主文に対して特定の機能を有し、その機能ごとに包括される、ひとまとまりの文書を表す。

次に、supplement 要素の例をあげる。図中、枠線で囲んで示した①～③が supplement 要素で、枠線で囲まれていない部分が主文（主要な学習内容）である。①～③は、主文とは外形的に区別され、かつ主文に対して特定の機能を有する、ひとまとまりの文書と見なされる。このうち、①②は主文への導入という機能を持つと見なされ、type 属性が"orientation"の supplement 要素となる。また、③は、主文に対して情報を補足する機能を持つと見なされ、type 属性が"advance"の supplement 要素となる。

①

②

③

100

『新しい社会 公民』（東京書籍 2005）p.100

属性

type 属性に以下の種別を示す。

- orientation （形式化例 1～3 参照）

主要な学習内容に対して、内容の理解を促す機能を持つ部分を表す。例えば、章や節の冒頭にある導入部分や末尾にあるまとめ部分、および問題・課題・実験・実習部分等が、これに相当する。

- advance （形式化例 4～5 参照）

主要な学習内容に対して、情報を補足し発展的理解を促す機能を持つ部分を表す。例えば、囲み記事として表されるコラム・参考資料等が、これに相当する。

- newWords (形式化例 6 参照)
 主要な学習内容に新しく登場した語句・漢字を一覧にした部分を表す。
- keyWords (形式化例 7 参照)
 主要な学習内容に登場した重要な語句を一覧にした部分を表す。

形式化例 1

原文

※主文に対する気づきを促す部分。



『新しい科学 2分野上』(東京書籍 2005) p.54

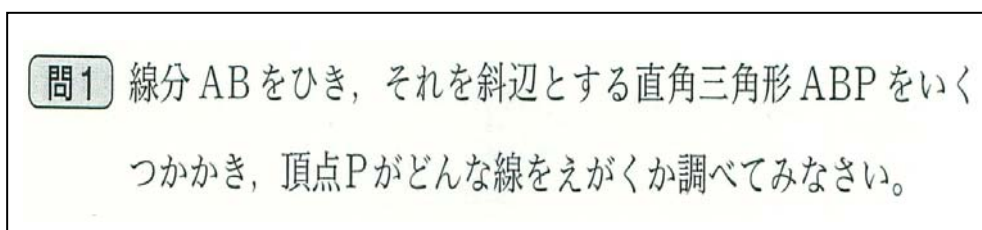
形式化

```
<supplement type="orientation">
  現在生きている生物に近い種類の化石なら、そのころの環境を推理できるかな。
</supplement>
```

形式化例 2

原文

※問題部分。



『新しい数学 2』(東京書籍 2005) p.140

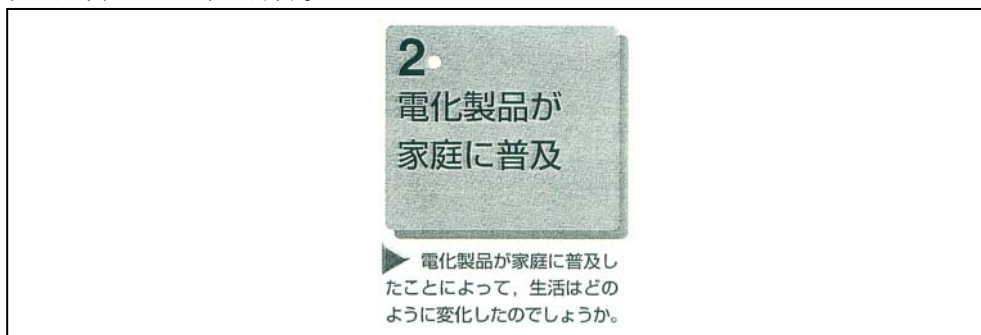
形式化

```
<supplement type="orientation">
<title>
問 1
</title>
線分 AB をひき、それを斜辺とする直角三角形 ABP をいくつかかき、頂点 P がど
んな線をえがくか調べてみなさい。
</supplement>
```

形式化例 3

原文

※節の冒頭にある導入部分。



『新しい社会 公民』（東京書籍 2005） p.10

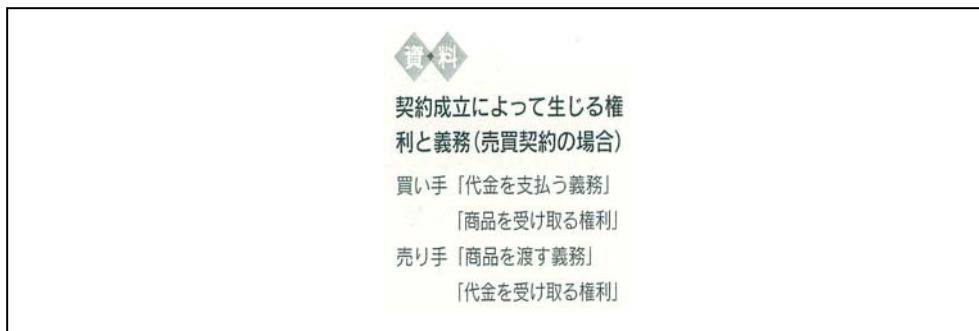
形式化

```
<cluster>
<title>
2. 電化製品が家庭に普及
</title>
<supplement type="orientation">
▶ 電化製品が家庭に普及したことによって、生活はどのように変化したのでしょうか。
</supplement>
[…略…]
</cluster>
```

形式化例 4

原文

※主文に添えられた資料部分。



『新しい技術・家庭 家庭分野』（東京書籍 2005） p.190

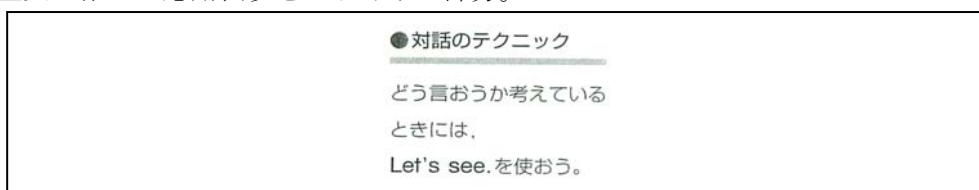
形式化

```
<supplement type="advance">
<title>
資料
</title>
契約成立によって生じる権利と義務（売買契約の場合）
買い手
「代金を支払う義務」
「商品を受け取る権利」
売り手
「商品を渡す義務」
「代金を受け取る権利」
</supplement>
```

形式化例 5

原文

※主文に添えて応用表現をとりあげた部分。



『NEW HORIZON English Course 2』（東京書籍 2005） p.3

形式化

```
<supplement type="advance">
<title>
●対話のテクニック
</title>
どう言おうか考えているときには、 L e t ' s   s e e . を使おう。
</supplement>
```

形式化例 6

原文

※主文に出現する新出漢字・新出音訓を列挙した部分。



『国語 3』(光村図書出版 2005) p.73

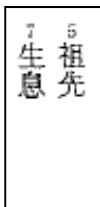
形式化

```
<supplement type="newWords">
<title>
《新出漢字》
</title>
7 0 陶トウ
7 1 賓ヒン
7 2 謙ケン
</supplement>
<supplement type="newWords">
<title>
《新出音訓》
</title>
7 0 器 (うつわ)
7 2 承る (うけたまわる)
</supplement>
```

形式化例 7

原文

※主文に出現する重要語句を列挙した部分。



『国語 1』(光村図書出版 2005) p.50

形式化

```
<supplement type="keyWords">
5 祖先
7 生息
</supplement>
```


surrogatePair

= で表現されている文字が、JIS X 0213:2004 で規定されているがサロゲートペアを使用する文字であることを示す。

属性

x0213 属性に、JIS X 0213:2004 の面区点番号を示す。
unicode 属性に、Unicode4.0 の 16 進コードを示す。

形式化例

原文

人	人	イ
(ひと)	(ひとやね)	(にんべん)
⋮	⋮	⋮
人	今・会・余・令	体・作・休・僧

『国語 1』(光村図書出版 2005) p.47

形式化

```
イ (にんべん) … 体・作・休・僧  
<surrogatePair x0213="2-01-22" unicode="U+201A2">= </surrogatePair> (ひとやね) …  
今・会・余・令  
人 (ひと) … 人
```

title

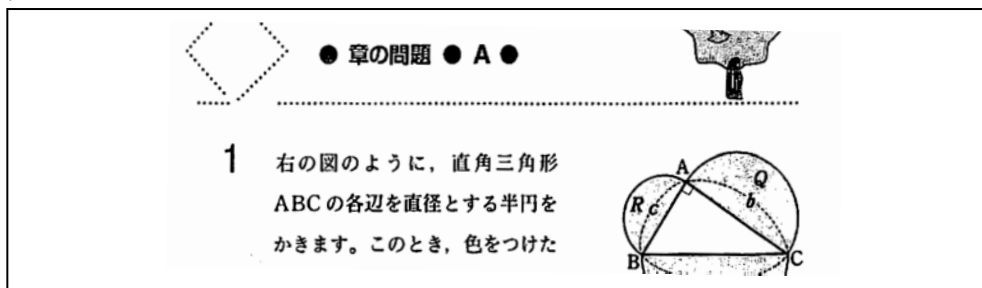
cluster 要素や supplement 要素の内容を代表する記述を表す。本文とは視覚的に区別され、文書の階層構造を示すための標識として機能する。title 要素の典型的な例は、文書における章・節等のタイトルである。

属性

なし。

形式化例

原文



『新しい数学3』（東京書籍 2005） p.113

形式化

```
<supplement type="orientation">  
<title>  
● 章の問題 ● A ●  
</title>  
1 右の図のように、直角三角形ABCの各辺を直径とする半円をかきます。このとき、  
色をつけた [...] <br>  
</supplement>
```

文献

山口昌也・高田智和・北村雅則・間淵洋子・小林正行・西部みちる（2008）．『現代日本語書き言葉均衡コーパス』における電子化フォーマット ver.2.0』，特定領域研究「日本語コーパス」平成 19 年度研究成果報告書（JC-D-07-03），特定領域研究「日本語コーパス」データ班.

第1章 第2節

教科書コーパス語彙表

田中 牧郎・近藤 明日子

1. 「教科書コーパス語彙表」とは何か

「教科書コーパス語彙表」は、特定領域研究「日本語コーパス」言語政策班で構築した「教科書コーパス」（第1章第1節参照）に出現する語彙に関するデータを収録した電子ファイルである。この語彙表によって、語彙素・語種・品詞などの語彙情報について、小学校・中学校・高等学校の教科書における全体像を把握することができるようになる。

本語彙表によって、「教科書コーパス」に用いられている語彙について、全校種・全学年・全教科を総合した頻度を知ることができれば、校種別・学年別・教科別の頻度を知ることが出来る。また、『現代日本語書き言葉均衡コーパス』（BCCWJ）の流通実態サブコーパス（図書館サブコーパス）の固定長サンプル（LB_FL）の語彙頻度も示した。こうした情報をもとに、「教科書コーパス」の各種の語彙頻度と図書館サブコーパスの語彙頻度とを統計的に比較することができる。具体的には、校種・学年・教科別あるいは教科書全体の「特徴度」、すなわち、各語がそれぞれの校種や学年や教科あるいは教科書においてどの程度特徴的であるかの数値を示した。

本語彙表を用いることで、教科書の語彙を様々な角度からながめることができるようになり、校種間・学年間・教科間で比較をしたり、一般社会で使われる語彙と比較したりすることができるようになる。言語政策研究・教育政策研究・言語教育研究の基礎資料の一つとして幅広く活用されることを期待している。

念のために付け加えるが、本語彙表が対象にしているのは、一般には非公開の部分も含めた「教科書コーパス」全体であり、BCCWJの非母集団コーパスの一部として一般公開する教科書サンプルの部分だけを対象にするものではない。したがって、本語彙表には、公開版の教科書サンプルにはない語彙も含まれている。

本語彙表は、本報告書の付録CD-ROM及び特定領域研究「日本語コーパス」研究成果報告DVDに収録し公開する。ファイル名は「教科書.txt」、形式はタブ区切りテキストファイル、文字符号化方式はUTF-16LE（BOM付き）、改行コードはCR+LFである。

2以下に記す、本語彙表の仕様については、本語彙表に添付する説明文書「教科書_README.pdf」の内容とおおむね同じものである。

2. 語彙調査の概要

本語彙表の作成にあたりおこなった、語彙調査の概要は以下の通りである。

2. 1 対象テキスト

調査の対象としたテキストは以下の2種類である。

- (1) 「教科書コーパス」2010年12月9日版(非公開)から、skippedSpan要素・skippedBlock要素およびtype属性値がnewWordsまたはkeyWordsのsupplement要素を除いた部分。

- (2) 『現代日本語書き言葉均衡コーパス』(BCCWJ) 2010年12月9日版(非公開)のうち、図書館サブコーパスに含まれる書籍の固定長サンプル群(LB_FL)、計10,640サンプル。

2. 2 テキストの形態素解析

調査対象テキストの形態素解析結果として、2010年12月9日時点で国立国語研究所内のデータベースに集積されているデータを利用した。これは、形態素解析辞書 UniDic¹ (MeCab版) による解析結果に基づくデータであるが、人手による修正の程度については、テキストごとに以下のように異なる。

- ・調査対象テキスト(1)の小学校教科書部分

UniDic による解析結果に対して、人手により全体を2度チェックし、解析の誤りを修正した。

- ・調査対象テキスト(1)の中学校・高校教科書部分

UniDic による解析結果に対する人手による修正はほとんどおこなっていない。

- ・調査対象テキスト(2)

UniDic による解析結果に対して、人手により一部をチェックし、解析の誤りを修正した。

以上のことから、解析精度はテキストにより異なることや、解析結果には解析の誤りが一部含まれ、それに基づく本語彙表にもその誤りが含まれていることに十分留意する必要がある。

2. 3 語の単位

語の単位には UniDic の解析単位である「短単位」を用い、1短単位を1語とした。

2. 4 同語異語判別

解析結果の同語異語判別は、UniDic により各短単位に付与される属性のうち、「語彙素読み」「語彙素」「語彙素細分類」「語種」「品詞」「活用型」の6属性を用い、これらの属性値がすべて一致するものを同語とし、一つの見出し語のもとにまとめた。逆に、6属性のうち、一つでも属性値が異なれば別語と認定した。表1に例としてあげた①～⑦の語では、①と②は「語彙素」、②と③は「活用型」、④と⑤は「語彙素細分類」、⑥と⑦は「品詞」がそれぞれ異なるため別語と認定した。

表1 別語の例

	語彙素読み	語彙素	語彙素細分類	語種	品詞	活用型
①	アラワス	著わす		和	動詞-一般	五段-サ行
②	アラワス	表わす		和	動詞-一般	五段-サ行
③	アラワス	表わす		和	動詞-一般	下一段-サ行
④	オール	オール	all	外	名詞-普通名詞-一般	
⑤	オール	オール	oar	外	名詞-普通名詞-一般	
⑥	アマリ	余り		和	形状詞-一般	
⑦	アマリ	余り		和	副詞	

¹ <http://download.unidic.org/>

2. 5 収録対象語の選定

本語彙表に収録する見出し語は、調査対象テキスト(1)に 1 回以上出現し、かつ「品詞」属性の大分類が「名詞・代名詞・形状詞・連体詞・副詞・接続詞・感動詞・動詞・形容詞・接頭辞・接尾辞」のいずれかであるものとした。助詞・助動詞・記号類・未知語等は掲載対象外とした。

3. 語彙表で使用するカテゴリー

本語彙表で調査対象資料(1)に関するデータを示す際に用いたカテゴリーには以下のようなものがある。() 内にカテゴリーの略称を示す。

- ・校種

小学校 (小)・中学校 (中)・高校 (高) の 3 種がある。

- ・学年

小学校から高校までを 3 学年ごとに分けたものを学年とした。小学校前半 (小_前)・小学校後半 (小_後)・中学校 (中)・高校 (高) の 4 種がある。

- ・教科

「教科書コーパス」の教科種別に従い、国語・数学・理科・社会・外国語・技術家庭・芸術・保健体育・情報・生活の 10 種がある。「教科書コーパス」の教科種別は、小学校・中学校・高校の各学習指導要領（平成 10～11 年文部省告示、平成 15 年一部改正）に定める教科に基づいて設定されている。「教科書コーパス」の教科と各学習指導要領の教科の対応関係を表 2 に示す。

表 2 教科の対応関係

教科書コーパス	学習指導要領		
	小学校	中学校	高等学校
国語	国語	国語	国語
数学	算数	数学	数学
理科	理科	理科	理科
社会	社会	社会	地理歴史 公民
外国語		外国語	外国語
技術家庭	家庭	技術・家庭	家庭
芸術	音楽 図画工作	音楽 美術	芸術
保健体育	体育	保健体育	保健体育
情報			情報
生活	生活		

4. 語彙表の構成

本語彙表は、以下の 204 列から構成される。また、先頭行は列名データである。

- 1 ID_教科書 …本語彙表の掲載された各見出し語に一意に付した番号（語彙素読みの五十音順）
- 2 ID_全体 …本語彙表および言語政策班の作成した「学校・社会対照語彙表」「BCCWJ 語彙表」に掲載された各見出し語に一意に付した番号（語彙素読みの五十音順）
- 3 語彙素読み …UniDic の語彙素読み

4	語彙素	…UniDic の語彙素
5	語彙素細分類	…UniDic の語彙素細分類
6	語種	…UniDic の語種
7	品詞	…UniDic の品詞
8	品詞_大分類	…UniDic の品詞の大分類
9	活用型	…UniDic の活用型
10	度数_小中高_全教科	…調査対象テキスト(1)での度数
11	使用率_小中高_全教科	…調査対象テキスト(1)での使用率 ²
12	特徴度_小中高_全教科	…調査対象テキスト(1)での特徴度 ³
13	度数_小_全教科	…調査対象テキスト(1)の小学校の教科書部分での度数
14	使用率_小_全教科	…調査対象テキスト(1)の小学校の教科書部分での使用率
15	特徴度_小_全教科	…調査対象テキスト(1)の小学校の教科書部分での特徴度
16	度数_中高_全教科	…調査対象テキスト(1)の中学校・高校の教科書部分での度数
17	使用率_中高_全教科	…調査対象テキスト(1)の中学校・高校の教科書部分での使用率
18	特徴度_中高_全教科	…調査対象テキスト(1)の中学校・高校の教科書部分での特徴度
19	度数_小_前_全教科	…調査対象テキスト(1)の小学校前半の教科書部分での度数
20	使用率_小_前_全教科	…調査対象テキスト(1)の小学校前半の教科書部分での使用率
21	特徴度_小_前_全教科	…調査対象テキスト(1)の小学校前半の教科書部分での特徴度
22	度数_小_後_全教科	…調査対象テキスト(1)の小学校後半の教科書部分での度数
23	使用率_小_後_全教科	…調査対象テキスト(1)の小学校後半の教科書部分での使用率
24	特徴度_小_後_全教科	…調査対象テキスト(1)の小学校後半の教科書部分での特徴度
25	度数_中_全教科	…調査対象テキスト(1)の中学校の教科書部分での度数
26	使用率_中_全教科	…調査対象テキスト(1)の中学校の教科書部分での使用率
27	特徴度_中_全教科	…調査対象テキスト(1)の中学校の教科書部分での特徴度
28	度数_高_全教科	…調査対象テキスト(1)の高校の教科書部分での度数
29	使用率_高_全教科	…調査対象テキスト(1)の高校の教科書部分での使用率
30	特徴度_高_全教科	…調査対象テキスト(1)の高校の教科書部分での特徴度
31	度数_小中高_国語	…調査対象テキスト(1)の国語の教科書部分での度数
32	使用率_小中高_国語	…調査対象テキスト(1)の国語の教科書部分での使用率
33	特徴度_小中高_国語	…調査対象テキスト(1)の国語の教科書部分での特徴度
34	度数_小中高_数学	…調査対象テキスト(1)の数学の教科書部分での度数
35	使用率_小中高_数学	…調査対象テキスト(1)の数学の教科書部分での使用率
36	特徴度_小中高_数学	…調査対象テキスト(1)の数学の教科書部分での特徴度
37	度数_小中高_理科	…調査対象テキスト(1)の理科の教科書部分での度数
38	使用率_小中高_理科	…調査対象テキスト(1)の理科の教科書部分での使用率
39	特徴度_小中高_理科	…調査対象テキスト(1)の理科の教科書部分での特徴度
40	度数_小中高_社会	…調査対象テキスト(1)の社会の教科書部分での度数
41	使用率_小中高_社会	…調査対象テキスト(1)の社会の教科書部分での使用率
42	特徴度_小中高_社会	…調査対象テキスト(1)の社会の教科書部分での特徴度
43	度数_小中高_外国語	…調査対象テキスト(1)の外国語の教科書部分での度数
44	使用率_小中高_外国語	…調査対象テキスト(1)の外国語の教科書部分での使用率
45	特徴度_小中高_外国語	…調査対象テキスト(1)の外国語の教科書部分での特徴度
46	度数_小中高_技術家庭	…調査対象テキスト(1)の技術家庭の教科書部分での度数
47	使用率_小中高_技術家庭	…調査対象テキスト(1)の技術家庭の教科書部分での使用率
48	特徴度_小中高_技術家庭	…調査対象テキスト(1)の技術家庭の教科書部分での特徴度
49	度数_小中高_芸術	…調査対象テキスト(1)の芸術の教科書部分での度数
50	使用率_小中高_芸術	…調査対象テキスト(1)の芸術の教科書部分での使用率
51	特徴度_小中高_芸術	…調査対象テキスト(1)の芸術の教科書部分での特徴度
52	度数_小中高_保健体育	…調査対象テキスト(1)の保健体育の教科書部分での度数
53	使用率_小中高_保健体育	…調査対象テキスト(1)の保健体育の教科書部分での使用率
54	特徴度_小中高_保健体育	…調査対象テキスト(1)の保健体育の教科書部分での特徴度
55	度数_小中高_情報	…調査対象テキスト(1)の情報教科書部分での度数
56	使用率_小中高_情報	…調査対象テキスト(1)の情報教科書部分での使用率
57	特徴度_小中高_情報	…調査対象テキスト(1)の情報教科書部分での特徴度
58	度数_小中高_生活	…調査対象テキスト(1)の生活の教科書部分での度数
59	使用率_小中高_生活	…調査対象テキスト(1)の生活の教科書部分での使用率
60	特徴度_小中高_生活	…調査対象テキスト(1)の生活の教科書部分での特徴度
61	度数_小_国語	…調査対象テキスト(1)の小学校の国語の教科書部分での度数

² 付記 1 参照

³ 付記 2 参照

184	度数_高_社会	…調査対象テキスト(1)の高校の社会の教科書部分での度数
185	使用率_高_社会	…調査対象テキスト(1)の高校の社会の教科書部分での使用率
186	特徴度_高_社会	…調査対象テキスト(1)の高校の社会の教科書部分での特徴度
187	度数_高_外国語	…調査対象テキスト(1)の高校の外国語の教科書部分での度数
188	使用率_高_外国語	…調査対象テキスト(1)の高校の外国語の教科書部分での使用率
189	特徴度_高_外国語	…調査対象テキスト(1)の高校の外国語の教科書部分での特徴度
190	度数_高_技術家庭	…調査対象テキスト(1)の高校の技術家庭の教科書部分での度数
191	使用率_高_技術家庭	…調査対象テキスト(1)の高校の技術家庭の教科書部分での使用率
192	特徴度_高_技術家庭	…調査対象テキスト(1)の高校の技術家庭の教科書部分での特徴度
193	度数_高_芸術	…調査対象テキスト(1)の高校の芸術の教科書部分での度数
194	使用率_高_芸術	…調査対象テキスト(1)の高校の芸術の教科書部分での使用率
195	特徴度_高_芸術	…調査対象テキスト(1)の高校の芸術の教科書部分での特徴度
196	度数_高_保健体育	…調査対象テキスト(1)の高校の保健体育の教科書部分での度数
197	使用率_高_保健体育	…調査対象テキスト(1)の高校の保健体育の教科書部分での使用率
198	特徴度_高_保健体育	…調査対象テキスト(1)の高校の保健体育の教科書部分での特徴度
199	度数_高_情報	…調査対象テキスト(1)の高校の情報の教科書部分での度数
200	使用率_高_情報	…調査対象テキスト(1)の高校の情報の教科書部分での使用率
201	特徴度_高_情報	…調査対象テキスト(1)の高校の情報の教科書部分での特徴度
202	初出学年	…当該見出し語が調査対象テキスト(1)で初めて出現した学年の略称
203	度数_LB_FL	…調査対象テキスト(2)での度数
204	使用率_LB_FL	…調査対象テキスト(2)での使用率

5. 延べ語数・異なり語数

本語彙表に掲載された見出し語について、調査対象テキスト(1)での延べ語数・異なり語数を校種・学年・教科別に示すと表3のようになる。

表3 延べ語数・異なり語数

	小学校+中学校+高校		小学校		中学校+高校	
	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数
全教科	2,518,486	50,329	286,449	13,682	2,232,037	48,077
国語	372,493	23,770	76,556	7,642	295,937	21,825
数学	238,916	5,144	64,298	2,261	174,618	4,209
理科	619,508	14,463	39,392	2,698	580,116	14,097
社会	760,397	27,174	63,076	6,200	697,321	26,192
外国語	51,374	5,693	0	0	51,374	5,693
技術家庭	171,516	10,488	12,772	1,774	158,744	10,273
芸術	155,452	15,115	20,616	3,543	134,836	14,070
保健体育	75,154	6,047	7,872	1,549	67,282	5,800
情報	71,809	4,342	0	0	71,809	4,342
生活	1,867	539	1,867	539	0	0

	小学校前半		小学校後半		中学校		高校	
	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数
全教科	67,191	5,384	219,258	12,238	451,245	22,609	1,780,792	43,882
国語	25,877	3,569	50,679	6,360	96,116	11,824	199,821	17,680
数学	19,042	1,085	45,256	1,963	49,067	1,853	125,551	3,704
理科	5,065	802	34,327	2,535	65,146	4,427	514,970	13,484
社会	6,634	1,298	56,442	5,923	104,205	9,491	593,116	24,859
外国語	0	0	0	0	13,450	2,051	37,924	5,109
技術家庭	0	0	12,772	1,774	63,847	6,140	94,897	8,107
芸術	8,706	1,854	11,910	2,686	35,553	5,832	99,283	12,160
保健体育	0	0	7,872	1,549	23,861	3,622	43,421	4,551
情報	0	0	0	0	0	0	71,809	4,342
生活	1,867	539	0	0	0	0	0	0

また、調査対象テキスト(2)の延べ語数は 3,938,696、異なり語数は 86,002 である⁴。

付記 1 使用率について

使用率とは、当該資料・カテゴリーにおける延べ語数に対する当該語の度数の割合を千分率（単位%）で示したものである。使用率算出に使用した各資料・カテゴリーの延べ語数は本文 5 を参照のこと。

付記 2 特徴度について

特徴度とは、当該資料における当該語が、他の資料（参照資料）と比べて出現度数の点でどの程度特徴的であるかを示す値である。特徴度には対数尤度比を補正した数値を用い、以下の式によって算出した（Kilgarriff, 2001 ; 内山・中條・山本・井佐原, 2004）。

$$2(\text{alna} + \text{blnb} + \text{clnc} + \text{dln}d - (\text{a} + \text{b})\ln(\text{a} + \text{b}) - (\text{a} + \text{c})\ln(\text{a} + \text{c}) - (\text{b} + \text{d})\ln(\text{b} + \text{d}) - (\text{c} + \text{d})\ln(\text{c} + \text{d}) + (\text{a} + \text{b} + \text{c} + \text{d})\ln(\text{a} + \text{b} + \text{c} + \text{d}))^5$$

- a : 当該資料での当該語の度数
- b : 参照資料での当該語の度数
- c : 当該資料の延べ語数 - a
- d : 参照資料の延べ語数 - b

※ただし、 $ad - bc < 0$ の場合の場合、-1 を乗じる補正を行う。

特徴度が 0 であれば、当該資料と参照資料で当該語の出現の程度は等しい。特徴度が正の値で、かつ値が高ければ高いほど、当該資料において高頻度という意味で特徴的な語と見なされる。逆に、特徴度が負の値で、かつ値が低ければ低いほど、当該資料において低頻度という意味で特徴的な語と見なされる。

また、正の値の特徴度の有意水準とその臨界値は以下の通りである（高見、2003）。

有意水準	0.1 (10%)	0.05 (5%)	0.01 (1%)	0.005 (0.5%)	0.001 (0.1%)
臨界値	2.71	3.84	6.63	7.88	10.83

⁴ ただし、本語彙表では調査対象テキスト(2)に出現しても調査対象テキスト(1)には出現しない見出し語は収録していないため、本語彙表の調査対象テキスト(2)に関するデータを集計してもこの延べ語数・異なり語数には達しないことに留意する必要がある。

⁵ \ln は自然対数を表す。a または b が 0 の場合、 alna または blnb を 0 として値を算出した（高見、2003）。

本語彙表では、調査対象テキスト(2)を参照資料として特徴度を算出した。

文献

- 内山将夫、中條清美、山本英子、井佐原均 (2004) 「英語教育のための分野特徴単語の選定尺度の比較」 自然言語処理, 11-3, pp.165-197.
- 高見敏子 (2003) 「「高級紙語」と「大衆紙語」の corpus-driven な特定法」 (北海道大学) 大学院国際広報メディア研究科・言語文化部紀要, 44, pp.73-105.
- 田中牧郎、近藤明日子、平山允子 (2011) 「教科書コーパス」 本報告書第 1 章第 1 節.
- Adam Kilgarriff (2001) "Comparing corpora" *International Journal of Corpus Linguistics*, 6-1, pp.1-37.

第1章 第3節

BCCWJ 主要コーパス語彙表

田中 牧郎・近藤 明日子

1. 「BCCWJ 主要コーパス語彙表」とは何か

「BCCWJ 主要コーパス語彙表」は、『現代日本語書き言葉均衡コーパス』(BCCWJ)の一部を構成するサンプル群に出現する語彙に関するデータを収録した電子ファイルである。この語彙表によって、語彙素・語種・品詞などの語彙情報について、現代社会の様々な媒体における全体像を把握することができるようになる。

本語彙表によって、BCCWJ に用いられている語彙について、各種媒体別の頻度や、頻度に基づく語彙のレベルなどを知ることができる。本語彙表に収録される各サブコーパスの語彙頻度や語彙レベルを相互に比較対照することもでき、媒体間の語彙のありようの違いについて確かめることもできる。

本語彙表には、BCCWJ のサブコーパスのうち、小学校から高等学校までの学校教育での語彙を考える際に重要だと思われる媒体、学校教育や国語政策に役立てる目的で現代日本語の語彙のありようを考える際に特に有益な情報を提供してくれそうな媒体、あわせて六種の語彙情報を収録した。その六種とは、次のものである。

- (1)流通実態 (図書館) サブコーパスに含まれる書籍の固定長サンプル (LB_FL)
- (2)生産実態 (出版) サブコーパスに含まれる書籍の固定長サンプル (PB_FL)
- (3)生産実態 (出版) サブコーパスに含まれる雑誌の固定長サンプル (PM_FL)
- (4)生産実態 (出版) サブコーパスに含まれる新聞の固定長サンプル (PN_FL)
- (5)非母集団 (特定目的) サブコーパスに含まれる Yahoo!知恵袋の可変長サンプル (OC_VL)
- (6)非母集団 (特定目的) サブコーパスに含まれる Yahoo!ブログの可変長サンプル (OY_VL)

このうち、(1)~(4)は、学校教育や国語政策での語彙を考える際に特に重要だと思われる媒体のサブコーパスである。可変長ではなく固定長サンプルを採用したのは、語彙頻度という計量的観点での扱いに適しているからである。(5)(6)は、書き言葉コーパスである BCCWJ の中で最も話し言葉的な性格を持っているのではないかと予想される媒体として採用した。話し言葉的性格を帯びる媒体での語彙頻度や語彙レベルは、(1)~(4)の教育上重視される書き言葉媒体での語彙頻度や語彙レベルとの比較対象として重要だと考えたことによる。なお、(5)(6)は固定長サンプルが存在しないので、可変長サンプルを用いた。

本語彙表は、本報告書の付録 CD-ROM 及び特定領域研究「日本語コーパス」研究成果報告書 DVD に収録し公開する。ファイル名は「BCCWJ.txt」、形式はタブ区切りテキストファイル、文字符号化方式は UTF-16LE (BOM 付き)、改行コードは CR+LF である。

2 以下に記す、本語彙表の仕様については、本語彙表に添付する説明文書「BCCWJ_README.pdf」の内容とおおむね同じものである。

2. 語彙調査の概要

本語彙表の作成にあたりおこなった、語彙調査の概要は以下の通りである。

2. 1 調査対象テキスト

調査の対象としたテキストは BCCWJ の 2010 年 12 月 9 日版（非公開）のうち、以下の 6 種類である。

- (1) 流通実態(図書館)サブコーパスに含まれる書籍の固定長サンプル(LB_FL)、計 10,640 サンプル。
- (2) 生産実態(出版)サブコーパスに含まれる書籍の固定長サンプル(PB_FL)、計 10,277 サンプル。
- (3) 生産実態(出版)サブコーパスに含まれる雑誌の固定長サンプル(PM_FL)、計 2,439 サンプル。
- (4) 生産実態(出版)サブコーパスに含まれる新聞の固定長サンプル(PN_FL)、計 1,489 サンプル。
- (5) 非母集団(特定目的)サブコーパスに含まれる Yahoo!知恵袋の可変長サンプル(OC_VL)、計 45,725 サンプル。
- (6) 非母集団(特定目的)サブコーパスに含まれる Yahoo!ブログの可変長サンプル(OY_VL)、計 52,680 サンプル。

2. 2 テキストの形態素解析

対象テキストの形態素解析結果として、2010 年 12 月 9 日時点で国立国語研究所内のデータベースに集積されているデータを利用した。これは、形態素解析辞書 UniDic¹ (MeCab 版) による解析結果に基づくデータであるが、人手による修正の程度がテキストごとに異なるため、解析精度もまたテキストごとに異なることに留意する必要がある。また、解析結果には解析の誤りが一部含まれ、それに基づく本語彙表にもその誤りが含まれていることにも十分留意する必要がある。

2. 3 語の単位

語の単位には UniDic の解析単位である「短単位」を用い、1 短単位を 1 語とした。

2. 4 同語異語判別

解析結果の同語異語判別は、UniDic により各短単位に付与される属性のうち、「語彙素読み」「語彙素」「語彙素細分類」「語種」「品詞」「活用例」の 6 属性を用い、これらの属性値がすべて一致するものを同語とし、一つの見出し語のもとにまとめた。逆に、6 属性のうち、一つでも属性値が異なれば別語と認定した。表 1 に例としてあげた①～⑦の語では、①と②は「語彙素」、②と③は「活用例」、④と⑤は「語彙素細分類」、⑥と⑦は「品詞」がそれぞれ異なるため別語と認定した。

表 1 別語の例

	語彙素読み	語彙素	語彙素細分類	語種	品詞	活用例
①	アラワス	著わす		和	動詞・一般	五段・サ行
②	アラワス	表わす		和	動詞・一般	五段・サ行
③	アラワス	表わす		和	動詞・一般	下一段・サ行
④	オール	オール	all	外	名詞・普通名詞・一般	
⑤	オール	オール	oar	外	名詞・普通名詞・一般	
⑥	アマリ	余り		和	形状詞・一般	
⑦	アマリ	余り		和	副詞	

¹ <http://download.unidic.org/>

2. 5 収録対象語の選定

本語彙表に収録する見出し語は、調査対象テキスト(1)～(6)のいずれかに1回以上出現し、かつ「品詞」属性の大分類が「名詞・代名詞・形状詞・連体詞・副詞・接続詞・感動詞・動詞・形容詞・接頭辞・接尾辞」のいずれかであるものとした。助詞・助動詞・記号類・未知語等は掲載対象外とした。

3. 語彙表の構成

本語彙表は、以下の33列から構成される。また、先頭行は列名データである。

1	ID_BCCWJ	…本語彙表の掲載された各見出し語に一意に付した番号(語彙素読みの五十音順)
2	ID_全体	…本語彙表および言語政策班の作成した「教科書コーパス語彙表」「学校・社会対照語彙表」に掲載された各見出し語に一意に付した番号(語彙素読みの五十音順)
3	語彙素読み	…UniDicの語彙素読み
4	語彙素	…UniDicの語彙素
5	語彙素細分類	…UniDicの語彙素細分類
6	語種	…UniDicの語種
7	品詞	…UniDicの品詞
8	品詞_大分類	…UniDicの品詞の大分類
9	活用型	…UniDicの活用型
10	度数_LB_FL	…調査対象テキスト(1)での度数
11	度数_PB_FL	…調査対象テキスト(2)での度数
12	度数_PM_FL	…調査対象テキスト(3)での度数
13	度数_PN_FL	…調査対象テキスト(4)での度数
14	度数_OC_VL	…調査対象テキスト(5)での度数
15	度数_OY_VL	…調査対象テキスト(6)での度数
16	使用率_LB_FL	…調査対象テキスト(1)での使用率 ²
17	使用率_PB_FL	…調査対象テキスト(2)での使用率
18	使用率_PM_FL	…調査対象テキスト(3)での使用率
19	使用率_PN_FL	…調査対象テキスト(4)での使用率
20	使用率_OC_VL	…調査対象テキスト(5)での使用率
21	使用率_OY_VL	…調査対象テキスト(6)での使用率
22	レベル_LB_FL	…調査対象テキスト(1)でのレベル ³
23	レベル_PB_FL	…調査対象テキスト(2)でのレベル
24	レベル_PM_FL	…調査対象テキスト(3)でのレベル
25	レベル_PN_FL	…調査対象テキスト(4)でのレベル
26	レベル_OC_VL	…調査対象テキスト(5)でのレベル
27	レベル_OY_VL	…調査対象テキスト(6)でのレベル
28	サンプル数_LB_FL	…調査対象テキスト(1)での出現サンプル数
29	サンプル数_PB_FL	…調査対象テキスト(2)での出現サンプル数
30	サンプル数_PM_FL	…調査対象テキスト(3)での出現サンプル数
31	サンプル数_PN_FL	…調査対象テキスト(4)での出現サンプル数
32	サンプル数_OC_VL	…調査対象テキスト(5)での出現サンプル数
33	サンプル数_OY_VL	…調査対象テキスト(6)での出現サンプル数

4. 延べ語数・異なり語数

本語彙表に掲載された見出し語130,437語について延べ語数・異なり語数をテキスト・レベル別に示すと表2のようになる。

² 付記1参照

³ 付記2参照

表2 調査対象テキストの延べ語数・異なり語数

	LB_FL		PB_FL		PM_FL	
	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数
全体	3,938,696	86,002	3,903,395	82,784	896,988	45,900
レベル a	3,074,655	4,177	3,045,639	3,842	700,831	4,336
レベル b	395,994	6,330	391,312	5,609	92,353	5,293
レベル c	242,911	11,595	239,221	10,506	51,085	7,493
レベル d	118,642	14,176	124,601	14,290	37,925	13,984
レベル e	106,494	49,724	102,622	48,537	14,794	14,794

	PN_FL		OC_VL		OY_VL	
	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数
全体	624,020	35,727	2,762,864	49,809	6,127,125	76,823
レベル a	486,976	3,420	2,155,871	2,071	4,779,106	3,441
レベル b	63,784	4,045	275,758	2,776	617,945	4,724
レベル c	40,018	6,941	165,957	5,122	372,114	8,406
レベル d	20,607	8,686	83,349	7,062	181,482	10,285
レベル e	12,635	12,635	81,929	32,778	176,478	49,967

付記1 使用率について

使用率とは、当該テキストの延べ語数に対する当該語の度数の割合を千分率（単位‰）で示したものである。使用率算出に使用した各テキストの延べ語数は本文4を参照のこと。

付記2 レベルについて

レベルとは、調査対象テキスト(1)～(6)のそれぞれに出現する見出し語について、度数降順の累積使用率（カバー率）により以下のようにa～eの5段階に分けたものである（田中、2011）。

レベル	カバー率（累積使用率）
a	0 ～ 78%
b	～ 88%
c	～ 94%
d	～ 97%
e	～ 100%

文献

田中牧郎（2011）「語彙レベルに基づく重要語彙リストの作成—国語政策・国語教育での活用のために—」 本報告書第2章第1節。

第1章 第4節

学校・社会対照語彙表

田中 牧郎・近藤 明日子

1. 「学校・社会対照語彙表」とは何か

「学校・社会対照語彙表」は、特定領域研究「日本語コーパス」言語政策班で構築した「教科書コーパス」（本報告書第1章第1節参照）の中学校・高校教科書部分に出現する語彙を中心に、『現代日本語書き言葉均衡コーパス』（BCCWJ）に出現する語彙を加えて、それらに関するデータを収録した電子ファイルである。具体的には、「教科書コーパス」をもとに作成した「教科書コーパス語彙表」（本報告書第1章第2節参照）に収録した語彙頻度の情報と、BCCWJをもとに作成した「BCCWJ主要コーパス語彙表」（本報告書第1章第3節参照）に収録した語彙頻度や語彙レベルの情報のうち、特に活用されることが多いと思われるものを抜き出して比較対照できるようにまとめ直したものである。

中等教育などが目標とすることの一つに、一般的な社会生活で必要とされる語彙力を身につけさせるということが挙げられよう。一般的な社会生活で必要とされる語彙は、現代語の書き言葉を代表できる形に設計された BCCWJ に反映していると考えられることから、この BCCWJ で頻度が高いものは、社会的な重要語彙である可能性が高い（本報告書第2章第1節参照）。一方、中等教育において現実に教えられている語彙は教科書に反映していると考えられ、その語彙は教科の学習を通して習得されると考えられる。その両者を比較対照できるように作成したのが、「学校・社会対照語彙表」である。

本語彙表の具体的な使用目的には、例えば次のようなことが想定できる。まず、社会的な重要語彙が教科書で取り上げられているか否か、取り上げられている場合はどの教科で主に学習するのかといったことを確かめるときに利用できる。また、各教科で学習される語彙について、他教科で学習されることがあるか否か、学習されることがある場合、それはどの教科であるかを調べることができる。さらに、本語彙表には「分類語彙表 増補改訂版」データベースの番号を収録しているので、社会で必要とされる重要な語彙を学校教育でどのように扱うかについて、意味という観点を取り込んで検討を深めていくことも期待できる。他にも、学校教育における語彙の扱いを考える基礎資料として、この語彙表を活用する方向は色々と考えられるだろう¹。

本語彙表は、本報告書の付録 CD-ROM 及び特定領域研究「日本語コーパス」研究成果報告書 DVD に収録し公開する。形式はタブ区切りテキストファイル、文字符号化方式は UTF-16LE（BOM 付き）、改行コードは CR+LF である。

本語彙表は、「分類語彙表 増補改訂版」データベースに関するデータの扱いの違いにより、「分類語彙表番号統合」版（ファイル名「学校社会対照_統合.txt」）と「分類語彙表番号分割」版（ファイル名「学校社会対照_分割.txt」）の2種がある。本語彙表の見出し語一つに対し同語と認められる「分類語彙表 増補改訂版」データベースのレコードが複数ある場合、「分類語彙表番号統合」版では複数レコードを一つのレコードに集約して示すのに対し、

¹ 本報告書第3章第2節では、その活用例のひとつとして、教科教育と語彙教育の関連づけについて検討した。

「分類語彙表分割」版では複数レコードを複数レコードのまま示すものである。

2以下に記す、本語彙表の仕様については、本語彙表に添付する説明文書「学校社会対照_TXT_README.pdf」の内容とおおむね同じものである。

2. 語彙調査の概要

本語彙表の作成にあたりおこなった、語彙調査の概要は以下の通りである。

2. 1 調査対象テキスト

調査の対象としたテキストは以下の(1)～(7)の7種類である。

- (1) 「教科書コーパス」2010年12月9日版(非公開)から、skippedSpan 要素・skippedBlock 要素および type 属性値が newWords または keyWords の supplement 要素を除いた部分。
- (2) 『現代日本語書き言葉均衡コーパス』(BCCWJ) 2010年12月9日版(非公開)のうち、流通実態(図書館)サブコーパスに含まれる書籍の固定長サンプル(LB_FL)、計10,640サンプル。
- (3) BCCWJの2010年12月9日版のうち、生産実態(出版)サブコーパスに含まれる書籍の固定長サンプル(PB_FL)、計10,277サンプル。
- (4) BCCWJの2010年12月9日版のうち、生産実態(出版)サブコーパスに含まれる雑誌の固定長サンプル(PM_FL)、計2,439サンプル。
- (5) BCCWJの2010年12月9日版のうち、生産実態(出版)サブコーパスに含まれる新聞の固定長サンプル(PN_FL)、計1,489サンプル。
- (6) BCCWJの2010年12月9日版のうち、非母集団(特定目的)サブコーパスに含まれるYahoo!知恵袋の可変長サンプル(OC_VL)、計45,725サンプル。
- (7) BCCWJの2010年12月9日版のうち、非母集団(特定目的)サブコーパスに含まれるYahoo!ブログの可変長サンプル(OY_VL)、計52,680サンプル。

2. 2 テキストの形態素解析

調査対象テキストの形態素解析結果として、2010年12月9日時点で国立国語研究所内のデータベースに集積されているデータを利用した。これは、形態素解析辞書 UniDic² (MeCab版)による解析結果に基づくデータであるが、人手による修正の程度がテキストごとに異なるため、解析精度もまたテキストごとに異なることに留意する必要がある。また、解析結果には解析の誤りが一部含まれ、それに基づく本語彙表にもその誤りが含まれていることにも十分留意する必要がある。

2. 3 語の単位

語の単位には UniDic の解析単位である「短単位」を用い、1短単位を1語とした。

2. 4 同語異語判別

解析結果の同語異語判別は、UniDicにより各短単位に付与される属性のうち、「語彙素読み」「語彙素」「語彙素細分類」「語種」「品詞」「活用型」の6属性を用い、これらの属性値がすべて一致するものを同語とし、一つの見出し語のもとにまとめた。逆に、6属性のうち、一つでも属性値が異なれば別語と認定した。表1に例としてあげた①～⑦の語では、①と

² <http://download.unidic.org/>

②は「語彙素」、②と③は「活用型」、④と⑤は「語彙素細分類」、⑥と⑦は「品詞」がそれぞれ異なるため別語と認定した。

表 1 別語の例

	語彙素読み	語彙素	語彙素細分類	語種	品詞	活用型
①	アラワス	著わす		和	動詞-一般	五段-サ行
②	アラワス	表わす		和	動詞-一般	五段-サ行
③	アラワス	表わす		和	動詞-一般	下一段-サ行
④	オール	オール	all	外	名詞-普通名詞-一般	
⑤	オール	オール	oar	外	名詞-普通名詞-一般	
⑥	アマリ	余り		和	形状詞-一般	
⑦	アマリ	余り		和	副詞	

2. 5 収録対象語の選定

本語彙表に収録する見出し語は、調査対象テキストの(1)の中学校・高校教科書部分あるいは調査対象テキスト(2)に1回以上出現し、かつ「品詞」属性の大分類が「名詞・代名詞・形状詞・連体詞・副詞・接続詞・感動詞・動詞・形容詞・接頭辞・接尾辞」のいずれかであるものとした。助詞・助動詞・記号類・未知語等は掲載対象外とした。

3. 語彙表で使用するカテゴリー

本語彙表で「教科書コーパス」のデータを示す際に用いたカテゴリーには以下のようなものがある。() 内にカテゴリーの略称を示す。

- ・ 学年

小学校から高校までを 3 学年ごとに分けたものを学年とした。小学校前半 (小_前)・小学校後半 (小_後)・中学校 (中)・高校 (高) の 4 種がある。

- ・ 教科

教科書コーパスの教科種別に従い、国語 (国)・数学 (数)・理科 (理)・社会 (社)・外国語 (外)・技術家庭 (技)・芸術 (芸)・保健体育 (保)・情報 (情)・生活 (生) の 10 種がある。「教科書コーパス」の教科種別は、小学校・中学校・高校の各学習指導要領 (平成 10~11 年文部省告示、平成 15 年一部改正) に定める教科に基づいて設定されている。「教科書コーパス」の教科と各学習指導要領の教科の対応関係を表 2 に示す。

表2 教科の対応関係

教科書コーパス	学習指導要領		
	小学校	中学校	高等学校
国語	国語	国語	国語
数学	算数	数学	数学
理科	理科	理科	理科
社会	社会	社会	地理歴史
			公民
外国語		外国語	外国語
技術家庭	家庭	技術・家庭	家庭
芸術	音楽	音楽	芸術
	図画工作	美術	
保健体育	体育	保健体育	保健体育
情報			情報
生活	生活		

また、媒体に関するカテゴリとして、調査対象テキスト(1)の中学校・高校教科書部分を「教科書(教)」、調査対象テキスト(2)を「書籍(書)」と設定する。

4. 語彙表の構成

本語彙表は、以下の29～30列から構成される。また、先頭行は列名データである。

- | | | |
|----|------------|---|
| 1 | ID_学校社会 | …本語彙表の掲載された各見出し語に一意に付した番号(語彙素読みの五十音順) |
| 2 | ID_全体 | …本語彙表および言語政策班の作成した「教科書コーパス語彙表」「BCCWJ 主要コーパス語彙表」に掲載された各見出し語に一意に付した番号(語彙素読みの五十音順) |
| 3 | 語彙素読み | …UniDicの語彙素読み |
| 4 | 語彙素 | …UniDicの語彙素 |
| 5 | 語彙素細分類 | …UniDicの語彙素細分類 |
| 6 | 語種 | …UniDicの語種 |
| 7 | 品詞 | …UniDicの品詞 |
| 8 | 品詞_大分類 | …UniDicの品詞の大分類 |
| 9 | 活用型 | …UniDicの活用型 |
| 10 | 初出学年 | …当該見出し語が調査対象テキスト(1)で初めて出現した学年の略称 |
| 11 | 度数_中高_全教科 | …調査対象テキスト(1)の中学校・高校の教科書部分での度数 |
| 12 | 度数_中高_国語 | …調査対象テキスト(1)の中学校・高校の国語の教科書部分での度数 |
| 13 | 度数_中高_数学 | …調査対象テキスト(1)の中学校・高校の数学の教科書部分での度数 |
| 14 | 度数_中高_理科 | …調査対象テキスト(1)の中学校・高校の理科の教科書部分での度数 |
| 15 | 度数_中高_社会 | …調査対象テキスト(1)の中学校・高校の社会の教科書部分での度数 |
| 16 | 度数_中高_外国語 | …調査対象テキスト(1)の中学校・高校の外国語の教科書部分での度数 |
| 17 | 度数_中高_技術家庭 | …調査対象テキスト(1)の中学校・高校の技術家庭の教科書部分での度数 |
| 18 | 度数_中高_芸術 | …調査対象テキスト(1)の中学校・高校の芸術の教科書部分での度数 |
| 19 | 度数_中高_保健体育 | …調査対象テキスト(1)の中学校・高校の保健体育の教科書部分での度数 |
| 20 | 度数_中高_情報 | …調査対象テキスト(1)の中学校・高校の情報の教科書部分での度数 |
| 21 | レベル_LB_FL | …調査対象テキスト(2)でのレベル ³ |
| 22 | レベル_PB_FL | …調査対象テキスト(3)でのレベル |
| 23 | レベル_PM_FL | …調査対象テキスト(4)でのレベル |
| 24 | レベル_PN_FL | …調査対象テキスト(5)でのレベル |
| 25 | レベル_OC_VL | …調査対象テキスト(6)でのレベル |
| 26 | レベル_OY_VL | …調査対象テキスト(7)でのレベル |

³ 付記1参照

- 27 特徴媒体 …当該見出し語が「教科書」「書籍」のいずれで特徴語となるかをその略称で示すもの。調査対象テキスト(1)の中学校・高校教科書部分を当該資料、調査対象テキスト(2)を参照資料とした場合の特徴度⁴が 6.63 (p>.01) より大きい場合、「教科書」の特徴語とみなし、-6.63未満の場合、「書籍」の特徴語とみなす。
- 28 特徴教科 …当該見出し語が特徴語となる教科の略称を示すもの。調査対象資料(1)の中学校・高校教科書の各教科部分を当該資料、調査対象資料(2)を参照資料とした場合の特徴度が 6.63 (p>.01) より大きい場合、当該教科の特徴語とみなす。

※以下、「分類語彙表番号統合」版にのみある列

- 29 分類語彙表番号_分類語彙表見出し …当該見出し語と同語と認められる「分類語彙表 増補改訂版」データベース⁵のレコードの分類語彙表番号⁶と見出しを_で連結するもの。複数のレコードが同語と認められる場合、レコード間を;で連結する。

※以下、「分類語彙表番号分割」版にのみある列

- 29 分類語彙表番号 …当該見出し語と同語と認められる「分類語彙表 増補改訂版」データベースのレコードの分類語彙表番号
- 30 分類語彙表見出し …当該見出し語と同語と認められる「分類語彙表 増補改訂版」データベースのレコードの見出し

5. 延べ語数・異なり語数

本語彙表に収録された見出し語 95,286 語のうち、調査対象テキスト(1)の中学校・高校教科書部分に出現するものは 48,077 語である。それらの語について、調査対象テキスト(1)の中学校・高校教科書部分での延べ語数・異なり語数を教科別に示すと表 3 のようになる。

表 3 延べ語数・異なり語数

	中学校+高校	
	延べ語数	異なり語数
全教科	2,232,037	48,077
国語	295,937	21,825
数学	174,618	4,209
理科	580,116	14,097
社会	697,321	26,192
外国語	51,374	5,693
技術家庭	158,744	10,273
芸術	134,836	14,070
保健体育	67,282	5,800
情報	71,809	4,342
生活	0	0

⁴ 付記 2 参照

⁵ <http://www.ninjal.ac.jp/products-k/kanko/goihyo/>

⁶ 「分類語彙表 増補改訂版」データベースのレコードを構成する項目のうち、分類番号・段落番号・小段落番号・語番号を-で連結したもの。

付記1 レベルについて

レベルとは、調査対象テキスト(2)~(7)のそれぞれに出現する見出し語について、度数降順の累積使用率(カバー率)により以下のようにa~eの5段階に分けたものである(田中、2011)。

レベル	累積使用率(カバー率)
a	0 ~ 78%
b	~ 88%
c	~ 94%
d	~ 97%
e	~ 100%

付記2 特徴度について

特徴度とは、当該資料における当該語が、他の資料(参照資料)と比べて出現度数の点でどの程度特徴的であるかを示す値である。特徴度には対数尤度比を補正した数値を用い、以下の式によって算出した(Kilgarriff、2001; 内山・中條・山本・井佐原、2004)。

$$2(\text{alna}+\text{blnb}+\text{clnc}+\text{dln d}-(\text{a}+\text{b})\ln(\text{a}+\text{b})-(\text{a}+\text{c})\ln(\text{a}+\text{c})-(\text{b}+\text{d})\ln(\text{b}+\text{d})-(\text{c}+\text{d})\ln(\text{c}+\text{d})+(\text{a}+\text{b}+\text{c}+\text{d})\ln(\text{a}+\text{b}+\text{c}+\text{d}))^7$$

- a: 当該資料での当該語の度数
- b: 参照資料での当該語の度数
- c: 当該資料の延べ語数-a
- d: 参照資料の延べ語数-b

※ただし、 $ad-bc < 0$ の場合の場合、-1 を乗じる補正を行う。

特徴度が0であれば、当該資料と参照資料で当該語の出現の程度は等しい。特徴度が正の値で、かつ値が高ければ高いほど、当該資料において高頻度という意味で特徴的な語と見なされる。逆に、特徴度が負の値で、かつ値が低ければ低いほど、当該資料において低頻度という意味で特徴的な語と見なされる。

また、正の値の特徴度の有意水準とその臨界値は以下の通りである(高見、2003)。

有意水準	0.1 (10%)	0.05 (5%)	0.01 (1%)	0.005 (0.5%)	0.001 (0.1%)
臨界値	2.71	3.84	6.63	7.88	10.83

文献

内山将夫、中條清美、山本英子、井佐原均(2004)「英語教育のための分野特徴単語の選定尺度の比較」自然言語処理, 11-3, pp.165-197.

高見敏子(2003)「「高級紙語」と「大衆紙語」の corpus-driven な特定法」(北海道大学)大学院国際広報メディア研究科・言語文化部紀要, 44, pp.73-105.

田中牧郎(2011)「語彙レベルに基づく重要語彙リストの作成—国語政策・国語教育での活用のために—」本報告書第2章第1節.

田中牧郎、近藤明日子、平山允子(2011)「教科書コーパス」本報告書第1章第1節.

Adam Kilgarriff(2001) "Comparing corpora" *International Journal of Corpus Linguistics*, 6-1, pp.1-37.

⁷ ln は自然対数を表す。a または b が 0 の場合、alna または blnb を 0 とし値を算出した(高見、2003)。

第2章

国語政策でのコーパスの活用

第2章 第1節

語彙レベルに基づく重要語彙リストの作成

—国語政策・国語教育での活用のために—

田中 牧郎

1. はじめに

特定領域研究「日本語コーパス」言語政策班が対象としている国語政策や国語教育の研究は、行政や教育の分野での日本語に規範を与え、その規範に基づいた教育を行うことに貢献することを目指している。こうした規範の設定とそれに基づく教育を円滑に進めるためには、重要な媒体での言語の実態を正しく把握しておくことが望まれる。

言語の実態を把握しておくべき重要な媒体とは何であろうか。それはどんな課題に取り組むかによって変わってこようが、『現代日本語書き言葉均衡コーパス』(BCCWJ)に収録されている書き言葉を代表する種々のサブコーパスは、国語政策や国語教育における多くの課題に対応できるのではないかと予想される。

言語政策班では、第一に語彙を、第二に語彙との関連から漢字を取り上げて、国語政策と国語教育の分野にコーパスを導入する研究を進めてきた。その研究によって得た指標のひとつとして「語彙レベル」があげられる。語彙レベルとは、各サブコーパスの語彙頻度に基づいて、語の重要度や日常度などを知る指標として用いることができるものである。語彙レベルの設定と付与の方法は、本報告書第1章第3節、同第4節でも述べたが、研究の結果得られた語彙レベルは、公開する「BCCWJ 主要コーパス語彙表」及び「学校・社会対照語彙表」に収録し、広く利用できるようにした。

本節は、この語彙レベルを重要語彙を抽出するための指標として扱う方法を検討し、重要語彙リストを作成する手順について報告するものである。

2. BCCWJによる語彙レベルの設定

2. 1. 利用するサブコーパス

言語政策班が作成した「BCCWJ 主要コーパス語彙表」(第1章第3節参照)及び「学校・社会対照語彙表」(第1章第4節参照)には、まず、母集団から無作為に抽出したサンプルからなる書き言葉としての代表性が高いサブコーパス、すなわち書籍(図書館)、書籍(出版)、雑誌、新聞の4つの媒体の語彙頻度に基づく語彙レベルが収録されている。次のものである。

- (1) 流通実態(図書館)サブコーパスに含まれる書籍の固定長サンプル(LB_FL)
- (2) 生産実態(出版)サブコーパスに含まれる書籍の固定長サンプル(PB_FL)
- (3) 生産実態(出版)サブコーパスに含まれる雑誌の固定長サンプル(PM_FL)
- (4) 生産実態(出版)サブコーパスに含まれる新聞の固定長サンプル(PN_FL)

これらのサブコーパスによく使われる語彙は、重要度が高いものではないかと予想され

る。いずれも、可変長（VL）ではなく固定長（FL）を採用したのは、語彙頻度を扱う計量的研究により適切だと考えたからである。

「BCCWJ 主要コーパス語彙表」及び「学校・社会対照語彙表」には、また、次の二つのサブコーパスでの語彙頻度に基づく語彙レベルも収録されている。

- (5) 非母集団（特定目的）サブコーパスに含まれる Yahoo!知恵袋の可変長サンプル (OC_VL)
- (6) 非母集団（特定目的）サブコーパスに含まれる Yahoo!ブログの可変長サンプル (OY_VL)

この二つのサブコーパスは、インターネット上の書き言葉媒体からサンプルが採られているものであり、話題が卑近であったり、私的な性質を帯びやすかったり、十分な推敲を経ていなかったりと、BCCWJ の中では最も日常的性格を有するものだと考えられる。したがって、これらのサブコーパスによく使われる語彙は、日常度が高いものではないかと予想される。なおこれら 2 つのサブコーパスには固定長サンプル（FL）がないため可変長サンプル（VL）を用いた。

2. 2. カバー率に基づくレベル分け

各サブコーパスにおける語彙レベルの設定の手順は、次のようにして行った。まず形態素解析辞書 UniDic で形態素解析を施し短単位で集計し語彙頻度表を作成した。その語彙を度数順に並べ、上位のものから度数を累積していき、その累積度数が当該サブコーパスにおける延べ語数の何パーセントを占めるかという「カバー率」（累積使用率）を算出し、このカバー率に一定の基準を設けて 5 つに区画した。その区画ひとつひとつを語彙レベルと扱った。このカバー率を基準とすることによって、規模の異なるサブコーパスであっても相互に均質なレベル分けが可能になる。カバー率によるレベル分けの区画と、各サブコーパスにおけるレベル別語数をまとめたものが表 1、表 2 である（第 1 章第 3 節、同第 4 節にも掲出）。

表 1 カバー率によるレベル分けの基準

レベル	カバー率（累積使用率）
a	0 ～ 78%
b	～ 88%
c	～ 94%
d	～ 97%
e	～ 100%

表2 各サブコーパスのレベル別語数

	LB_FL		PB_FL		PM_FL	
	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数
全体	3,938,696	86,002	3,903,395	82,784	896,988	45,900
レベル a	3,074,655	4,177	3,045,639	3,842	700,831	4,336
レベル b	395,994	6,330	391,312	5,609	92,353	5,293
レベル c	242,911	11,595	239,221	10,506	51,085	7,493
レベル d	118,642	14,176	124,601	14,290	37,925	13,984
レベル e	106,494	49,724	102,622	48,537	14,794	14,794

	PN_FL		OC_VL		OY_VL	
	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数
全体	624,020	35,727	2,762,864	49,809	6,127,125	76,823
レベル a	486,976	3,420	2,155,871	2,071	4,779,106	3,441
レベル b	63,784	4,045	275,758	2,776	617,945	4,724
レベル c	40,018	6,941	165,957	5,122	372,114	8,406
レベル d	20,607	8,686	83,349	7,062	181,482	10,285
レベル e	12,635	12,635	81,929	32,778	176,478	49,967

表2を概観してみよう。まず「全体」の行で「延べ語数」を見ると、OY（ブログ）¹がもっとも多く、LB（図書館書籍）とPB（出版書籍）がほぼ同量でこれに次ぐ。それらに次いでOC（知恵袋）が多く、以上の4つのサブコーパスは250万語以上の大規模なものである。一方、PM（雑誌）とPN（新聞）は100万語以下で、比較的小規模なものである。「異なり語数」も、延べ語数に応じて規模に大小があるが、次の点が目を引く。すなわち、延べ語数が600万語を超えるOY（ブログ）よりも、延べ語数が400万語弱のLB（図書館書籍）・PB（出版書籍）の方が、異なり語数が多いという点である。このことは、LB（図書館書籍）・PB（出版書籍）の語彙の豊かさをものがたるものだと考えられる。

レベルごとの延べ語数・異なり語数を見ても、おおまかには、全体の場合とほぼ同じような傾向が認められる。しかし細かく見ると、サブコーパスごとに各レベルに属する語数の多寡に特徴があるように見受けられる。以下では、サブコーパス間で、レベルと語彙の関係にどのような相違があるのかについて、少し詳しく考察してみたい。

3. 語彙レベルから見た各サブコーパス語彙の性格

3. 1. 語種の観点からの比較

はじめに、各サブコーパスのレベルによる語彙の分布から語彙全体の性格を考えるために、語種の観点から分析する。

表3・表4・表5は、各レベルでの「和語率」「漢語率」「外来語率」を、サブコーパス間で比較できるように一覧にしたものである。これらの数値は、付属語と記号を除くすべての語彙を対象に、異なり語数における和語・漢語・外来語・混種語の構成比率を算出した

¹ 以下、各サブコーパスを記号で表示する場合、FL（固定長）、VL（可変長）の区別は省略する。

結果をもとに、雑多な語彙を含む混種語率以外の数値を表にまとめたものである。

表3 各レベルにおける和語率（サブコーパス間の比較）

レベル	LB_FL	PB_FL	PM_FL	PN_FL	OC_VL	OY_FL
a	40.2%	35.6%	33.9%	28.6%	40.5%	40.2%
b	32.2%	28.9%	29.4%	27.2%	32.3%	32.7%
c	34.2%	32.1%	32.3%	29.3%	31.3%	34.3%
d	35.3%	32.5%	36.1%	34.0%	34.9%	36.5%
e	34.3%	33.4%	37.9%	38.3%	38.6%	36.9%
全体	34.6%	32.8%	34.9%	32.9%	36.9%	36.4%

表4 各レベルにおける漢語率（サブコーパス間の比較）

レベル	LB_FL	PB_FL	PM_FL	PN_FL	OC_VL	OY_FL
a	52.1%	55.3%	45.4%	62.5%	46.0%	44.7%
b	54.5%	56.1%	47.1%	58.2%	47.6%	46.3%
c	50.0%	50.9%	44.5%	53.2%	46.7%	44.5%
d	46.3%	48.0%	41.3%	47.4%	42.7%	42.4%
e	43.8%	44.4%	40.0%	42.1%	38.5%	40.2%
全体	46.7%	47.6%	42.7%	50.1%	41.1%	41.8%

表5 各レベルにおける外来語率（サブコーパス間の比較）

レベル	LB_FL	PB_FL	PM_FL	PN_FL	OC_VL	OY_FL
a	5.7%	7.2%	19.0%	7.3%	11.2%	13.2%
b	10.6%	12.7%	21.2%	12.1%	17.7%	18.3%
c	12.5%	13.9%	19.2%	13.9%	18.6%	17.9%
d	14.1%	15.8%	18.6%	14.6%	19.1%	17.3%
e	16.9%	17.3%	17.2%	15.0%	18.0%	17.6%
全体	14.5%	15.5%	18.7%	13.4%	17.9%	17.4%

表3の「和語率」は、全体として、OC（知恵袋）、OY（ブログ）で高く、PB（出版書籍）とPN（新聞）で低い。レベル別で見ると、高頻度から中頻度の語彙であるレベル a・b・c の部分では、LB（図書館書籍）の数値が比較的高くなり、低頻度であるレベル e の部分では、PM（雑誌）とPN（新聞）の数値が比較的高くなる。和語は、一般に日常語性の高いものが多いと言われ、一般的な語彙では日常的な語が高頻度から中頻度のところに多いとされているので、これが全体として多いOC（知恵袋）とOC（ブログ）は日常的な語彙が多く使われていること、これが高頻度・中頻度のレベルに多いLB（図書館書籍）は一般的な語彙のありようをよく反映していることなどが、読み取れよう。

次に、表4の「漢語率」を見ていこう。漢語率は全体として、PN（新聞）で最も高く、PB（出版書籍）がこれに次ぐ。一方、OC（知恵袋）、OY（ブログ）で低い。これは、先に見た和語率とちょうど反対の現象であり、和語が少ないものは漢語が多くなっていると見ることができる。ただ、和語率の場合、高い数値を示すサブコーパスと低い数値を示すサ

ブコーパスとの差は4ポイント程度だったものが、漢語率の場合、その差が8~9ポイントと大きいという違いがある。レベル別に見ると、高頻度から中頻度であるレベル a・b・c において、全体でも漢語率の高かった PB (出版書籍) と PN (新聞) の数値がより一層高くなっていることが目を引く。また、全体では漢語率は必ずしも高くなかった LB (図書館書籍) が、レベル a・b・c においてそれが比較的高くなっていることも注目される。漢語は、一般に文章語性の高いものが多いと言われるので、これが全般に多い PB (出版書籍) と PN (新聞) は文章語的な語彙が多く使われていること、これが高頻度・中頻度のレベルに多い LB (図書館書籍) は、文章語もよく取り込んでいるということを示すことができる。

最後に、表5の「外来語率」を検討する。外来語率は全体として、PM (雑誌) で最も高く、OC (知恵袋) と OY (ブログ) がこれに次ぐ。外来語は、一般に、新しい語彙、感覚的な語彙が多いと言われるので、これらのサブコーパスがそうした語彙を受け入れやすい側面を持っていることを示している。これらのサブコーパスは、レベル b・c といった中頻度のレベルでも外来語率が高く、特に、PM (雑誌) は高頻度のレベル a においてもその比率が極めて高く、語彙の基本的な部分にまで新しく感覚的なものを取り込んでいると見ることができよう。PN (新聞)、LB (図書館書籍)、PB (出版書籍) は、全体的に外来語率は低い、例頻度であるレベル d・e の部分ではその率が高くなる傾向が見られ、語彙の周辺的な部分には新しく感覚的な語彙が入り込んでいると思われる。その中で、LB (図書館書籍) だけが、高頻度であるレベル a における外来語率が極めて低いことは、語彙の基本的な部分に新しく感覚的なものが入り込みにくく、安定していることを示唆するものだと考えられる。

以上、語種の観点から見てきたことを総合し、各サブコーパスの語彙の性格をまとめると、次のようなことが指摘できよう。

- ・LB (図書館書籍) は、一般的な語彙のありようを反映し、文章語をよく取り込み、語彙の基本的な部分が安定している。
- ・PB (出版書籍) は、文章語的な語彙を多く含み、語彙の周辺的な部分には、新しく感覚的な語彙も取り込んでいる。
- ・PM (雑誌) は、語彙の基本的な部分にまで、新しく感覚的なものを取り込んでいる。
- ・PN (新聞) は、文章語的な語彙が特に多く、語彙の基本的な部分にまで、それを取り込んでいる。
- ・OC (知恵袋) と OY (ブログ) は、ともに、日常的な語彙が多くを占め、新しく感覚的な語彙も取り込んでいる。

3. 2. レベル a における特有語の分析

(1) 分析の方法

各サブコーパスの語彙の特徴を、語彙レベルとの関連でより深く分析してみよう。今度、できるだけ具体的な語彙に即して考察できるように、レベル a における特有語を扱うことにする。ここでいう「レベル a における特有語」というのは、あるサブコーパスでレベル a に属する語彙のうち、他のどのサブコーパスでもレベル a に属していないものである。例えば、LB (図書館書籍) でレベル a とされる 4,177 語のうち、他の 5 つのサブコーパスのいずれにおいてもレベル a になっていない語を、「LB 特有語」と扱うのである。各サブコーパスのレベル a における特有語を見ることで、各サブコーパスでよく使われる語でありながら他のサブコーパスではさほど多くは使われていない語を浮かび上がらせ

ることができ、各サブコーパスの語彙の性格がより明確に浮き彫りになると考えられる。

表 6 レベル a における特有語（品詞別の語数）

	名詞	動詞	形容詞	形状詞	副詞	連体詞	接続詞	感動詞	代名詞	接辞	計
LB 特有語	230	79	4	12	16	0	0	6	2	16	365
PB 特有語	177	4	1	2	0	0	0	0	0	5	189
PM 特有語	615	24	6	29	6	0	0	0	0	24	704
PN 特有語	473	23	1	3	0	0	0	0	0	26	526
OC 特有語	67	5	3	1	0	0	0	0	1	0	77
OY 特有語	224	20	6	4	17	0	2	12	1	14	300

表 7 レベル a における特有語（品詞別の比率）

	名詞	動詞	形容詞	形状詞	副詞	連体詞	接続詞	感動詞	代名詞	接辞	計
LB 特有語	63.0%	21.6%	1.1%	3.3%	4.4%	0.0%	0.0%	1.6%	0.5%	4.4%	100%
PB 特有語	93.7%	2.1%	0.5%	1.1%	0.0%	0.0%	0.0%	0.0%	0.0%	2.6%	100%
PM 特有語	87.4%	3.4%	0.9%	4.1%	0.9%	0.0%	0.0%	0.0%	0.0%	3.4%	100%
PN 特有語	89.9%	4.4%	0.2%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	4.9%	100%
OC 特有語	87.0%	6.5%	3.9%	1.3%	0.0%	0.0%	0.0%	0.0%	1.3%	0.0%	100%
OY 特有語	74.7%	6.7%	2.0%	1.3%	5.7%	0.0%	0.7%	4.0%	0.3%	4.7%	100%

（2）品詞別の概観

表 6、表 7 は各サブコーパスのレベル a における特有語の語数と比率を品詞別に集計したものである。品詞の枠組は、UniDic の「大分類」にしたがい、「接頭辞」と「接尾辞」はひとつにまとめて「接辞」とした。

表 6 から、PM（雑誌）と PN（新聞）には特有語が多く、OC（知恵袋）と PB（出版書籍）にはそれが少ないことが見て取れる。表 7 からは、どのサブコーパスの特有語も名詞が非常に高い比率を占めていることが分かる。これは、そもそも品詞の中では名詞が最も語数が多いことに関わるものだが、例えば LB（図書館書籍）のレベル a 全体では名詞の比率は約 65.1%、同じく PB（出版書籍）では約 67.5%、PM（雑誌）では約 71.3%、PN（新聞）では約 73.5%、OC（知恵袋）では約 63.8%、OY（ブログ）では約 66.6%であることに照らせば、特有語が名詞に集中する度合いは、LB（図書館書籍）の場合を除いて、極めて高いといつてよい。一般に、名詞は、文章の主題や話題に深く関わるものだと考えられるから、各サブコーパスの特有語に名詞が多いのは、主題となる事柄や話題として取り上げられる事物が、媒体によって大きく異なっていることをものがたっているのではないかと思われる。

特有語には名詞が多いという全体的傾向から外れる現象として、「LB 特有語」では名詞の比率が低いことが注目される。名詞が低くなっている分、動詞が高くなっていると見られ、その高さは他のサブコーパスと比較して群を抜いている。この点は、LB（図書館書籍）の語彙の性格を考える際に重要である。

(3) LB 特有語

「LB 特有語」(図書館書籍特有語)には、動詞が多いという目立った特徴があった。その動詞語彙を度数の高い方から 10 語あげよう。

候う、窺う、捕まえる、ぶつかる、傷つける、引き摺る、指差す、傾ける、怒鳴る、出でる

多くは日常場面でよく使われる基本的な動詞であり、一部に、古風な語(「候う」「出でる」)が混じる。どちらも、文学作品には多用される語彙だと考えられ、文学の占める割合が高い LB では、このような語彙が高頻度のレベル a に入ってくるのだと考えられる。副詞では「忽ち」「何しろ」「況して」、形容詞・形状詞では「素早い」「浅い」「奇妙」「密か」などが度数が高いが、これらも同類であろう。

名詞の度数上位 20 語は下のようのものであり、やはり多くが基本的な語であり、人や動植物など具体物を示すものが多い。

女房、詩人、父、地獄、巻、谷、奴隷、冗談、黒人、商人、奉行、二三、警部、移民、熊、竹、徒、翻訳、遠慮、国王

(4) PB 特有語

「PB 特有語」(出版書籍特有語)は、名詞が 90%を超える。その度数上位 20 語は次の通りである。

施策、取り消し、頻度、検出、算入、潜在、阻害、徴収、防御、間接、共済、主任、史料、税率、抽象、保管、曲線、社債、データベース、受容

LB 特有語に多かった具体物を表す語は見られず、ほとんどすべてが抽象概念を表す語である。これらは文学的な文章ではなく実用的な文章で多用されるものだと考えられ、LB に比べ文学よりも実用文の占める割合が高い PB では、このような語彙が高頻度のレベル a に入ってくるのであろう。

(5) PM 特有語

「PM 特有語」(雑誌特有語)は、語数が最も多く、やはり、名詞が多い。

スカート、ニット、渋谷、ジャケット、本紙、カジュアル、ベルト、プリント、ネックレス、収納、連載、ヘア、リング、モチーフ、デニム、小物、着こなし、コーディネート、-犬(けん)、ドレス

上記の語は、度数上位の 20 語(名詞以外も含む)であるが、ファッションにかかわる語が非常に多く、特定の話題の語彙に偏っている様子が見て取れる。さらに下位にはファッション以外の語彙も多くなるが、雑誌のテーマに関わる話題の語彙が多くを占めている。雑誌に特有語が最も多いのは、雑誌というものが特定の分野に焦点を当てて編集されているからだと思われる。

(6) PN 特有語

「PN 特有語」(新聞特有語)も名詞に多い。度数上位 20 語(名詞以外も含む)は次のものである。

会談、減、同日、衆院、参院、見通し、議長、高知、-区、与党、-面、幹事、市長、調べ、同市、拉致、アテネ、公明、懸念、盛り込む

政治や時事に関わる語彙や地名が目立つ。新聞という媒体が日々丁寧に取り上げ、他の媒体では扱われることがそれほど多くない話題が反映している語彙だと考えられる。

(7) OC 特有語

「OC 特有語」(知恵袋特有語)は全体の語数が最も少ないが、レベル a の異なり語数が際立って少ないこととも関連していよう。それは、よく使われる語彙に他の媒体に比べた特徴が薄いということの意味する。他の特有語の場合と同じ方法で、度数上位 20 語をあげると次の通りである。

落札、ID、発送、存知、口座、何方、振り込み、入金、手数、インストール、カテ、中古、浮気、振り込む、履歴、ウィン、ダウンロード、不倫、排卵、揭示

インターネット上に集まった人たちが比較的短い文章で、Q & A のやり取りを行うという場でこそ話題になりやすい語があがっていると見られよう。

(8) OY 特有語

「OY 特有語」(ブログ特有語)は、名詞の比率が他の媒体に比べて低くなり、感動詞や副詞の比率が高くなっているのが特徴である。度数上位 20 語は次の通りであり、感動詞や形容詞も多く含まれている。

ブログ、本日、眠い、久々、いや、目茶、先週、んっ、来週、御早う、インデックス、めでたい、過ぎ、レーザー、やばい、土産、捲る、まじ、チョコ、クリニック

日記あるいは語りに取り上げられやすい話題や、インターネット上で記したり語ったりする際に使われやすい語彙があがっていると見られよう。

(9) まとめ

特有語のうち度数の高い語彙には、それぞれのサブコーパスの語彙の特徴が極端な形で現れていると考えられる。語彙全体としては、各サブコーパス間で共通する部分が多いことに留意した上で、述べてきたような性格を各サブコーパスの語彙が持つことを踏まえて、その活用方法を考えていくことが必要である。

3. 3. サブコーパスの語彙レベルの利用に向けて

以上、各サブコーパスにおける語彙の性格を、語彙レベルとの関わりから明らかにしてきた。こうした各サブコーパスの語彙レベルを、国語政策や国語教育でどのように利用することが考えられるだろうか。

第一に、政策や教育の種々の局面でよりどころとすることができる、規範的な語彙リストが作成できると便利である。この規範的な語彙リストには、語彙が豊かで安定していること、現代語彙の一般的なありようを映していること、書き言葉の語彙を反映していることなどが、要件となろう。先に考察した各サブコーパスの語彙の特徴を見比べて、この要件に合致するサブコーパスを求めれば、LB (図書館書籍) が最適だと言えるだろう。書き言葉の語彙のありようを映し出しているという点では、PB (出版書籍)、PN (新聞) も要件を満たすが、語彙の豊かさと安定、現代語彙の一般的なありようの反映という 2 点で、LB (図書館書籍) に及ばないと考えられる。したがって、LB (図書館書籍) の語彙レベルを、政策や教育に活用するための「重要度」の指標として設定し、これに基づいた語彙リストを作ることが有益であろう。

ただし、LB (図書館書籍) の語彙レベルだけで、重要度が十分に把握できるというわけではない。LB の語彙レベルをあくまでもひとつの分かりやすい目安として扱い、必要に応じて、他のサブコーパスの語彙レベルを援用していくのが現実的なやり方であろう。

例えば、日常的な語彙が反映していると見られる、OC (知恵袋) と OY (ブログ) は日

常度の指標として使える面があろう。特に、特有語が少なく媒体としての特徴が希薄な OC（知恵袋）の語彙レベルは、日常度の指標として最適なものではないかと考えられる。利用例をひとつあげれば、LB（図書館書籍）の語彙レベルによって重要度が高いとされた語彙のうち、OC（知恵袋）の語彙レベルを援用して日常度が高いとされるものを除外すれば、日常生活の中では身に付きにくい学習すべき重要語彙を特定するのに役立つと考えられる。ほかに、社会における種々の実務で必要になる重要語彙を抽出したいときには、PB（出版書籍）の語彙レベルを援用したり、時事的な重要語彙には、PN（新聞）の語彙レベルを援用したりすることなども、想定できよう。

4. LB_FN（図書館書籍）の語彙レベルに基づく重要語彙リストの作成

4. 1. 重要語と扱う範囲

LB（図書館書籍）の語彙レベルを重要度と扱う場合、a・b・c・d・eの順で重要度が高いということになるが、ある水準までを「重要語」として区画できると便利である。例えば、重要語の範囲が過不足なく書き表せる範囲を常用漢字の適用範囲と想定して、漢字政策の目安を設定したり、重要語の範囲を語彙学習の到達目標としたりすることが考えられる。その範囲は、政策や教育の対象や目的によって拡大したり縮小したりするのは当然だが、本報告書の別の節で進める研究のための作業仮説として、一般的な重要語の範囲を定めてみたい。

従来、国語政策に活用される一般的な語彙リストというものはないが、国語教育に活用する目的ではいくつかの語彙リストが提供されている。そのうち、最も影響力があり、語数も多いものに、阪本一郎による『新教育基本語彙』（阪本 1984）がある。この語彙リストは、小学生・中学生段階で教育すべき語彙として、小学生段階約 10,000 語、中学生段階約 10,000 語、合計約 20,000 語を具体的に示している。このリストにならうなら、義務教育期間の中学校終了段階での学習目標として約 20,000 語を重要語として設定することが考えられよう²。表 2 の LB（図書館書籍）にまとめた各レベルの所属語数を累積していくと、レベル a から c までで約 22,000 語となるので、ここまでを重要語の範囲として扱うことにしてみよう。別の考え方として、レベル a・b の約 11,000 語を小学校終了段階の目標としたり、レベル a から d までの約 36,000 語を高等学校終了段階の目標としたりと、いくつかの作業仮説を立てる可能性なども想定できよう。ここでは、試験的な作業として 22,000 語のリストを作って、それを利用する方法を考えていこうとするのである。

4. 2. 重要語のふるい分け

LB（図書館書籍）の語彙レベルが a・b・c の約 22,000 語を重要語として抽出した語彙リストを見ていくと、中学生に対してわざわざ取り立てて学習させるまでもないような平易な語が混じっていたり、反対に、中学生にはまだ不必要ではないかと思われるような専門的な語が混じっていたりすることがある。次は、このリストから名詞（UniDic の大分類による）を抽出し五十音順に並べた冒頭部分である。固有名詞には*を付けた。

ア、亜、アー、アーケード、アーサー*、アーチ、アーチスト、アート、アーノルド*、

² 阪本（1984）は、「教育基本語彙」を示したものであり、本稿の「重要語」とは異なる概念である。したがって、阪本に合わせて約 20,000 語のリストを作るということには、実はあまり意味がないのかもしれない。ただ、他に参照すべきより適切な語彙リストがないので、ここでは阪本（1984）を持ち出した。

アーム、アーモンド、アーリー、アール*、RNA、RB、愛、アイ、藍、アイ*、アイアン、ISO*、ISDN、IMF、合鍵、アイカワ*、哀願、愛敬、愛犬、アイコ*、愛好、愛国、アイコン、挨拶、アイザワ*、IC、ICPO*、哀愁、相性、愛称、愛情、愛人、アイス、合図、アイゼン、愛想、アイゾウ*、間、アイダ*、間柄、アイチ*、愛着、アイヅ*、相槌、相手、アイディア、IT、ID、アイテム、アイデンティティ、愛読、アイドル、アイヌ*、アイノスケ*、アイハラ*、アイビー、IBM*、愛撫、相棒、合間、愛用、哀楽、アイルランド*、アイロン、アインシュタイン*、アウシュビッツ*、アウト、青、・・・

平易な語は、日常度を反映していると考えられるOC(知恵袋)の語彙レベルが高頻度のレベルになっていると考えられる。そこで、OCのレベルがa・bのものを、上記のリストから除外してみると次のようになる。

ア、亜、アー、アーケード、アーサー*、アーチ、アート、アーノルド*、アーム、アーモンド、アーリー、アール*、RNA、RB、藍、アイアン、ISO*、ISDN、IMF、合鍵、アイカワ*、哀願、愛敬、愛犬、アイコ*、愛好、愛国、アイザワ*、ICPO*、哀愁、愛称、愛人、合図、アイゼン、愛想、アイゾウ*、アイダ*、間柄、愛着、アイヅ*、相槌、IT、アイテム、アイデンティティ、愛読、アイヌ*、アイノスケ*、アイハラ*、アイビー、IBM*、愛撫、相棒、合間、愛用、哀楽、アイルランド*、アイロン、アインシュタイン*、アウシュビッツ*、・・・

「愛」「挨拶」「アイス」「間」「相手」「アウト」「青」等が消えている。また、専門的な語がリストに入る場合、特定のサンプルで主題になっていることで高頻度になっているためではないかと思われる。そこで、「BCCWJ主要コーパス語彙表」(第1章第3節参照)にある「サンプル数_LB_FL」の列が1、2、3のいずれかであるものを除外すると、次のようになる。

ア、亜、アー、アーケード、アーサー*、アーチ、アート、アーノルド*、アーム、アーモンド、アーリー、アール*、RNA、藍、アイアン、ISO*、ISDN、IMF、合鍵、アイカワ*、哀願、愛敬、愛犬、アイコ*、愛好、愛国、アイザワ*、哀愁、愛称、愛人、合図、愛想、アイダ*、間柄、愛着、アイヅ*、相槌、IT、アイテム、アイデンティティ、愛読、アイヌ*、アイハラ*、アイビー、IBM*、愛撫、相棒、合間、愛用、哀楽、アイルランド*、アイロン、アインシュタイン*、アウシュビッツ*、・・・

「RB」「ICPO」「アイゼン」「アイゾウ」「アイノスケ」等が消えている。このように、LB(図書館書籍)の語彙レベルを重要度の基本的な指標としながらも、OC(知恵袋)の語彙レベルを日常度の指標、LBのサンプル数を専門度の指標として援用するなどして、重要語リストの妥当性を向上させていくことができると考えられる。

LBの語彙レベル(重要度の指標)、OCの語彙レベル(日常度の指標)、LBのサンプル数(専門度の指標)の3つの指標によってふり分けられた重要語は、約17,000語のリストとなる。こうした語彙レベルを中心とした、コーパスの語彙頻度に基づく種々の情報をどのように組み合わせれば抽出していけば最適な重要語リストになるのかという点については、さらに研究が必要である。

5. おわりに

語彙レベルをはじめとする各コーパスの語彙頻度をもとに、国語政策・国語教育のより

どころとなるような重要語彙リストを作成する考え方と具体的手順について述べてきた。こうして作成した重要語彙リストを、国語政策や国語教育の諸課題にどのように役立てていくのかについては、個々の課題に対応させた別の研究が必要である。

例えば、田中（2009）では、難解な用語を平易に言い換える課題や、常用漢字に入れるべき（入れなくてもよい）漢字を検討する課題という、2つの国語政策課題に語彙レベルを活用する方向を示した。その方向について、BCCWJの完成データを用いて、あらためて検証してみる必要がある。

国語教育については、重要語彙リストを基盤に中等教育での語彙教育のあり方を検討する課題に取り組んでいる。本報告書の第3章第2節の「分類重要語彙リスト」作成の試み、同じく第4章第1節の語彙指導の教材研究での活用は、その課題についての現段階での報告である。

文献

阪本一郎（1984）『新教育基本語彙』，学芸図書。

田中牧郎（2009）「言語政策に役立つ、コーパスを用いた語彙表・漢字表などの作成と活用」人工知能学会誌, 24-5, 人工知能学会, pp.665-67.

第2章 第2節

難解用語の抽出と序列化におけるコーパスの利用

—医療用語を例に—

田中 牧郎・近藤 明日子

1. 難解用語の言語問題へのコーパスの活用

同じ言語社会に属し同じ言語を使っているが、用語が難解であるために情報伝達に障害が生じている問題がある。この「難解用語の言語問題」は、高度な情報化と専門化が進む一方で、専門性を伴う事柄についても非専門家である素人が意思決定に参加する機会が多くなってきている近年の日本社会において、重要性が増してきている。問題の解明と改善を目指して、行政¹、司法²、医療³などの分野で取り組みが行われ、筆者のひとり田中はこれらの取り組みにかかわった⁴。この言語問題の解明と改善策の策定のためには、当該分野の語彙の実態を把握することが求められるが、これまでそれは十分に行われてこなかった。

この、難解用語の実態把握の段階で、大規模コーパスを役立てることができないだろうか。特に、膨大な語彙の中から問題となる難解用語を抽出し、その重要度を序列化することにコーパスの語彙頻度が活用できないだろうか。

本稿では、難解な医療用語を抽出し重要度で序列化する作業にコーパスを活用する方法について考えてみたい。本稿は、国立国語研究所「病院の言葉」委員会の活動のために実施した語彙抽出作業に基づく部分があり、田中・金・桐生・近藤（2008）、国立国語研究所（2009）、田中（2009）などで述べた内容と部分的に重複するが、それらで述べたことを総合し、その後の研究も加えてまとめ直したい。

2. 語彙頻度に基づく難解な専門用語の抽出

2. 1 考え方

難解用語の言語問題の中心には、難解な専門用語の存在があり、そうした難解な専門用語を特定する作業が必要になる。コーパスから専門用語を抽出する研究は、自然言語処理分野で進んでおり、Uchimoto ほか（2000）、中川ほか（2003）、久保ほか（2010）などの成果がある。それらは専門論文等での情報処理を主目的とするもので、専門家間の情報伝達における重要な専門用語の抽出が主たる関心の対象となっている。本稿が扱っているような非専門家にとって難解な専門用語の抽出は、こうした成果を参照しつつも別の視点も取り入れた独自の検討が必要になる。

¹ 国立国語研究所「外来語」委員会による「『外来語』言い換え提案」（国立国語研究所 2006、<http://www.ninjal.ac.jp/gairaigo/>）。

² 日本弁護士連合会による「法廷用語の日常語化プロジェクト」（日本弁護士連合会 2008）

³ 国立国語研究所「病院の言葉」委員会による「『病院の言葉』を分かりやすくする提案」（国立国語研究所 2009、<http://www.ninjal.ac.jp/byoin/>）

⁴ これまで取り組んだ「難解用語の言語問題」への具体的対応については、田中・相澤（2010）にまとめた。

上記の先行研究のうち、Uchimoto ほか（2000）では、論文データベースと新聞データベースとで語彙頻度を比較し、久保ほか（2010）では、女性学の論文誌と女性学以外の分野の論文誌とで語彙頻度を比較し、それぞれ特定の分野の専門用語を抽出する方法を示している。本稿が目指す、非専門家にとって難解な医療分野の専門用語の抽出においても、異なる分野のコーパス同士の語彙頻度の比較が役立つと考えられる。本稿の目的にとって有効な頻度比較が行えるコーパスの組み合わせは、次のようなものであろう。

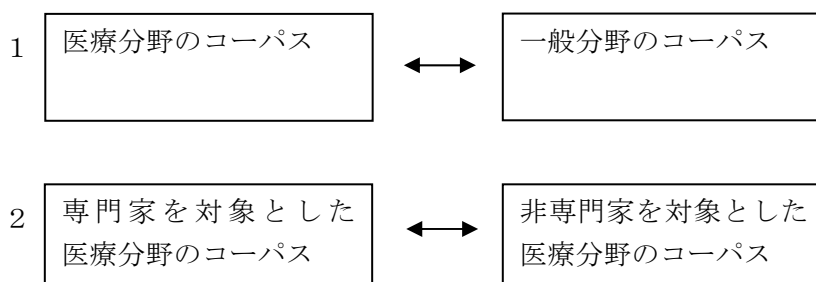


図1 コーパスの語彙頻度の比較

1は、久保ほか（2010）と類似の方法であり、他の一般的な分野と比較したときの医療専門用語を抽出することができると考えられる。一方、2は、Uchimoto ほか（2000）と類似し、専門家を読者に想定する文章と非専門家を読者に想定する文章とを比較することで、非専門家にとってなじみのない医療用語を抽出することができるだろう。1だけでは、専門家と非専門家の違いという観点での語彙比較ができず、また、2だけでは、医療分野とそれ以外の分野の違いという観点での語彙比較ができない。1と2を組み合わせることで、非専門家にとってなじみのない医療専門用語を抽出することができるのではないかと予想される。そこで、まず1を実施し、その後2を行うことを試みた。

2. 2 医療分野のコーパスと一般分野のコーパスの語彙比較

(1) コーパスの作成

まず、医療について書かれた文章として下の三種を収集し、医療分野のコーパスを作成した。①雑誌、②新聞、③ウェブの三つを選択したのは、これらが情報を得る媒体として多くの人にとって参照しやすいものだと考えられるからである。

①医療雑誌 33種 10箇月分

『雑誌総カタログ』（メディアリサーチセンター）において大分類「厚生・医療」に配されている雑誌のうち、発行部数が10,000部以上の月刊誌をリスト化し、小分類（「内科」「リハビリテーション」「医療管理」など）のバランスに配慮して、33種を対象にした。2006年9月から2007年6月までの10箇月分について、紙媒体の雑誌をスキャナ入力や手入力によって電子化し、約1,370万語のコーパスを作成した。

②新聞の医療記事 全国紙1種 23箇月分

ウェブ上で医療記事をひとまとめにして公開している全国紙のうち、よく読まれているものとして、「読売オンライン」（読売新聞社）の「医療・介護」を対象にダウンロードしてコーパスを作成した。このサイトの記事は多くが読売新聞紙上にも

掲載されているものである。2006年1月～2007年11月の記事を対象にしたところ、約260万語の規模となった。

③ウェブ上で製薬会社が提供する医療情報 60社分

ウェブ上で医療情報を提供するサイトは多いが、良質な情報を組織的に提供しているものとして、製薬会社が患者向けに提供している情報を対象とすることにした。具体的には日本薬剤師会のサイトのリンク集に挙げられた製薬会社のうち60社のサイトからダウンロードし、約250万語のコーパスとした。

以上の①②③を合わせた約1,890万語を、医療分野のコーパスとして扱った。

医療以外の分野を対象とする一般分野のコーパスには、特定領域研究「日本語コーパス」のデータ班が作成した『現代日本語書き言葉均衡コーパス』(以下「BCCWJ」と称する)の書籍データを用いることにした。2007年11月段階で電子化が済んでいた、生産実態(出版)・流通実態(図書館)・ベストセラーの各サブコーパスの書籍データあわせて約2,320万語分である。

どの程度の規模のコーパスを使えば目的とする語彙抽出が達せられるのかははっきりしないが、今回は、BCCWJの書籍データの量を目安にして、予算や時間の範囲内でこの量にできるだけ近い分量の医療分野のコーパスを作ることを目指した⁵。

(2) 語彙頻度の比較による医療専門用語の抽出

上記の二つのコーパス(医療分野のコーパス、一般分野のコーパス)に対して形態素解析を施した。解析器はChaSen、解析辞書はUniDic1.3.6(非公開バージョン)を用いた。UniDicの解析単位である短単位は、単純語など短い単位を基本としているので、合成語が多い専門用語の分析には適していない面があり、長い単位での集計が求められる。ところが、長単位での形態素解析は、本特定領域研究の電子化辞書班・データ班での研究で環境が整えられる見通しであるものの、現段階では容易に使える状態にはなっていない。そこで、短単位で切り出されたものを長い単位にまとめる規則(例:熱中/症→熱中症、起立/性/低/血圧→起立性低血圧)を独自に立て、擬似的な長単位で集計する作業を行った。

なお、頻度比較を行う計算方法については、久保ほか(2010)やUchimotoほか(2000)では複雑な計算式を用いているが、ここではより簡潔でコーパス言語学の分野でもよく用いられている「対数尤度比」(log-likelihood ratio: LLR)を用いることにした。この指標については、本報告書の第3章第1節「中学校・高校教科書の教科特徴語リストの作成」(近藤明日子)で詳しく述べているので、参照してほしい。

医療分野のコーパスと一般分野のコーパスを合わせた全語彙(異なり語数で約72万語)について、それぞれのコーパスでの頻度を求め、次の計算式によって、対数尤度比を算出した。

$$LLR=2(\text{alna}+\text{blnb}+\text{clnc}+\text{dlnd}-(\text{a+b})\ln(\text{a+b})-(\text{a+c})\ln(\text{a+c})-(\text{b+d})\ln(\text{b+d})-(\text{c+d})\ln(\text{c+d})+(\text{a+b+c+d})\ln(\text{a+b+c+d}))$$

a: 医療分野のコーパスでの単語Wの頻度 b: 一般分野のコーパスでの単語Wの頻度
c: 医療分野のコーパスの延べ語数-a d: 一般分野のコーパスの延べ語数-b
ただし、単語Wの医療分野のコーパスでの使用率が一般分野のコーパスでの使用率より低

⁵ この医療分野のコーパスは、国立国語研究所「病院の言葉」委員会の活動に使う目的で作成したものであり、コーパス自体の公開の予定はない。

い場合、×(-1)の補正を行う。

LLR>0 で LLR の高い語ほど医療分野のコーパスにおいて特徴的な語と見なされる。ここでは、LLR \geq 10 という基準を立て、これを満たす約 24,000 語を医療の専門用語として抽出した。これは、全体の約 72 万語のうち約 3.3%を占める。

(3) 抽出された医療専門用語の概観

抽出された医療専門用語がどのようなものであるかを具体的に見てみたいが、ここでは語彙素読みの語頭が「ア」の 558 語を取り上げて概観してみたい。558 語の中には、コーパス作成時の OCR による読み誤り（「問題」[正しくは「問題」] など）や、自動形態素解析による解析誤り（「アイソビエトザイム」[正しくは「アイソザイム」、「ソ」が「ソビエト」と誤って認識されたため）などが混入しているのでこれらを取り除き、また、「愛子さん」「愛知県」「愛育病院」などのように、人名・地名・組織名といった固有名詞も除外した（ただし、薬剤の商品名は除外しなかった）。残った 428 語について、意味分野⁶に分類して一覧にすると、次の通りである。

A 身体部位など (34 語)

赤色度、赤血球、赤血球数、赤み、アキレス腱、握力、顎関節、顎口腔領域、顎下腺、顎堤、顎堤粘膜、足裏、足裏マッサージ、足腰、足先、足背、足背動脈、足全体、足底、足底板、足底部、味蕾、足部、足指、足指回し、頭位、頭側、後壁、後腹膜、肋、肋骨、脂性肌、脂腺、アライメント

B 病気や症状など (113 語)

RA、ISD、ITC、IPF、アウトブレイク、アカシジア、亜急性、亜急性期、亜急性硬化性全脳炎、亜急性甲状腺炎、明細胞腺癌、悪性化、悪性、悪性化、悪性型、悪性胸膜中皮腫、悪性黒色腫、悪性疾患、悪性腫瘍、悪性症候群、悪性新生物、悪性度、悪性脳腫瘍、悪性貧血、悪性リンパ腫、亜型、亜型ウイルス感染、顎関節症、アジア風邪、足壊疽、足潰瘍、足型、足関節、足首、脚症候群、アジソン病、アシドーシス、足白癬、足病変、足ブロック、アスピリン潰瘍、アスピリン喘息、アスピリン喘息患者、アスピリン不耐症、アスペルガ、汗過敏症、汗疹、亜脱臼、悪化、悪化因子、悪化防止、悪化要因、圧迫、圧迫骨折、圧迫力、アテローム、アテローム血栓性梗塞、後首淋巴節、後大脳動脈、後縦靭帯、後囊、アトピー、アトピー克服、アトピー克服体験記、アトピー症状、アトピー性疾患、アトピー性皮膚炎、アトピー性皮膚炎患者、アトピー素因、アトピー体質、アトピー治療、アトピー肌、後鼻漏、後負荷、アナフィラキシー、アナフィラキシー様症状、アナフィラキシー様反応、アナフィラキシー症状、アナフィラキシーショック、アナフィラキシー反応、脂血症、脂肺肝、アメーバ性大腸炎、アルコール依存症、アルコール性、アルコール性肝炎、アルコール性肝障害、アルツハイマー、アルツハイマー型、アルツハイマー型認知症、アルツハイマー病、アレルギー、アレルギー科、アレルギー患者、アレルギー機序、アレルギー検査、アレルギー疾患、アレルギー症状、アレルギー性、アレルギー性刺草疹、アレルギー性結膜炎、アレルギー性結膜疾患、アレルギー性疾患、アレルギー性接触皮膚炎、アレルギー性鼻炎、アレルギー性皮膚炎、アレルギー素因、アレルギー体質、アレルギー反応、アレルギー物質、アレルギーマーチ、アレルギー歴、安静時狭心症

C 診療や検査など (33 語)

アートセラピー、アームダウン法、RIA、IMT、ICU、ICU患者、アイシング、アイソトープ、亜鉛不足、青色光、顎関節症治療、足心健康法、足切断、足底板療法、温シップ、圧接、圧迫止血、圧迫療法、圧力計、アニマルセラピー、アルコール消毒、アルコール制限、アルコール注入、アルコール綿、アルテプラゼ、アルテプラゼ静注療法、アロマセラピー、アロマセラピスト、アロママッサージ、安静、安静時、アンチエイジング、アンチエイジング医学

⁶ この意味分野の分類は「病院の言葉」委員会での語彙選定作業を行う際に設定した枠組みである。

D 物質や薬剤など (80 語)

RAS、RAS抑制薬、亜鉛、悪玉、悪玉菌、悪玉コレステロール、アクトス、アクネ菌、アクブラ、麻杏、アシクロビル、アジュバント、アスピリン、アスピリンイラクサシン、アスベスト、アセチルコリン、アセチルコリン受容体、アセチルサリチル酸、アセトアミノフェン、アセトアルデヒド、アディポネクチン、アデホビル、アデホビルピボキシル、アドリアシン、アトルバスタチン、アドレナリン、アナストロゾール、アバスタチン、アベロックス、アベロックス錠、アミノ酸、アミノ酸製剤、アミノフィリン、アミノ酪酸、編目状静脈瘤、アミラーゼ、アミロイド、アミロイド蛋白、アムロジピン、アモキシシリン、アリシン、アリセプト、アリミデックス、アルカロイド、アルキル化剤、アルケラン、アルコール製剤、アルドステロン、アルドステロン系、アルブミン、アルブミン製剤、アルブミン値、アルブミン尿、アレルゲン、アレルゲン除去、アレンドロン酸、アロプリノール、アロマシン、アロマターゼ、アロマターゼ阻害剤、アロマターゼ阻害薬、アンジオテンシノーゲン、アンジオテンシン、アンジオテンシン、アンジオテンシン、アンジオテンシン系、アンジオテンシン阻害薬、アンジオテンシンツウ、アンジオテンシン変換酵素、アンジオテンシン変換酵素阻害薬、アンスラサイクリン、アンスラサイクリン系、安息香酸リザトリプタン、アントシアニン、アントロゲン、アンピシリン、アンブル、アンブル入り風邪薬、アンペック座剤、アンモニア

E 理念や制度など (59 語)

ISO規格、ISBN、INR、IC、ICF、ICM、ICT、ICD、IDF、アウトカム、アウトカム評価、アウトソーシング、アウトソーシング先、赤ちゃんポスト、アクションプラン、アシステッド、預かり金、アスベスト新法、アセスメント、アセスメント結果、アセスメントシール、アセスメントツール、アセスメント能力、アドヒアランス、アドボカシー、アドボカシー室、アドボケイト、アメニティー、アルバイト診療、安心安全、安心感、安全安心、安全域、安全確認、安全確保、安全管理、安全管理指針、安全管理指針検討作業部会、安全管理体制、安全教育、安全使用、安全性、安全性確保、安全性向上、安全性試験、安全性情報、安全対策、安全配慮、安全配慮義務、安全配慮義務違反、安全部会、安全面、安定化、安定期、安定供給、安定性、安定性試験、安楽、安楽死

F 健康や生活習慣など (60 語)

アームレスト、アイケア、アイスパック、アイスマッサージ、アイビロー、アイマーク、青背、青背魚、赤ちゃん、アクアゲイン、アクアゲル、アクアビクス、揚げ物、朝御飯、朝晩、朝夕食後、味付け、足元作り、足湯、足浴、アセロラ、遊び編、温まる、温める、厚着、温感、温感タイプ、温経湯、アフターケア、脂、油、脂っこい、脂分、脂身、亜麻仁油、甘味、甘味料、アラーム、アラーム設定、洗い上がり、洗い過ぎ、洗い出す、洗い流す、洗い流す、アルカリ、アルカリ化、アルカリ性、アルカリ性単純泉、歩き、歩く、アルコール、アルコール摂取、アルコール等、アルコールベース、アルコール量、アロエ、アロマ、アロマオイル、阿波踊り、阿波踊り体操

G 医療と直接関わらないもの (49 語)

相性、IT化、IT化推進、ITシステム、IT新改革戦略、IT戦略本部、挙がる、秋頃、明き店舗、悪影響、悪質、アクシデント、悪循環、アクセス、アクセス数、アクティブシニア、揚げ、上げ下げ、浅い、アジる、頭打ち、当たり、圧、アップ、アップ-、宛て先、当て嵌まる、後押し、アドバイス、アピール、アフター、アプローチ、誤る、予め、泡状、合わせ持つ、併せる、泡立てる、案、安易、安価、アンケート、アンケート結果、アンケート調査、アンケート調査結果、餡こ、アンコールシリーズ、アンパランス、安否確認

まず、A身体部位、B病気や症状、C診療や検査、D物質や薬剤などを表す医学用語もしくは医療用語と言うべきものが多数抽出されている。また、E医療についての理念や制度を示す語彙や、F健康や生活習慣を表す語彙も多く、これらは医学用語ではないが、医療の場面でよく使われる医療用語と見なしてよいものであろう。このように、二つのコーパスの語彙頻度を比較する方法によって、医療に深く関わる語彙が多数抽出できている様子が確認される。

ところが、G医療と直接関わらないものも49語ある点に注意が必要である。これらの語彙は、例えば、「IT化」「IT化推進」などであればIT化の動きが、他の分野に比して

医療界で顕著であったり、「相性」「悪影響」「悪質」「悪循環」などは、医療において特に問題になりやすい概念であったりといった何らかの理由から、他の分野に比べて医療の分野で用いられやすいのだと考えられる。

2. 3 専門家対象の医療分野のコーパスと非専門家対象の医療分野のコーパスの語彙比較

(1) 医療分野のコーパスの二類別

次に、2.1において2として図示した、専門家を対象とした医療分野のコーパスと非専門家を対象とした医療分野のコーパスとで、語彙頻度を比較する。まず、2.2で記したコーパスのうち、医療分野のコーパスを次の二つに分けた。

- a 専門家を読者とする医療分野のコーパス 約 1,010 万語
 - ・医療雑誌のうち、医師・看護師など医療の専門家を読者に想定しているもの（例：『日経メディカル』『エキスパートナース』など）
- b 非専門家を読者とする医療分野のコーパス 約 880 万語
 - ・医療雑誌のうち、一般の人々や患者を読者に想定しているもの（例：『きょうの健康』『がんサポート』）
 - ・新聞の医療記事
 - ・ウェブ上での医療情報

雑誌は読者層を明確にして編集されており、aとbに分けることができる。また、新聞は専門家も非専門家も読むが、今回対象とした読売新聞は一般の人々に広く読まれることを想定して編集されている。さらに、ウェブ上での製薬会社の情報提供は、医療者向けと患者など非医療者向けとに分けて行われており、今回対象としたのは後者のものである。

(2) 語彙頻度の比較による専門度の設定

2.2(2)で行った形態素解析と集計の結果得られた語彙頻度データを用いて医療の専門用語と認定した約 24,000 語について、上記の a と b の二つのコーパス間で語彙頻度を比較した。比較の指標には、やはり 2.2(2)で記した対数尤度比 (LLR) を用いた。この数値が高いほど、非専門家向けの医療情報媒体よりも専門家向けの医療情報媒体によく使われていることを示しており、それだけ専門度が高いのではないかと考えられる。表 1 は、専門度を 5 段階のレベルに区切って整理してみたものである。レベル分けにあたっては、各区画の所属語数がほぼ均等になるようにした。これによって、対象にしている約 24,000 語すべてに 1～5 の専門度の情報が付与できる。

表1 専門度の設定

専門度	LLR の区間	語数
1	～ -12.1	5,102
2	-12.1 ～ -0.3	4,808
3	-0.3 ～ 4.4	4,969
4	4.4 ～ 14.0	4,740
5	14.0 ～	4,716
計		24,335

(3) 専門度の概観

上記のようにして設定された専門度が、各語に実際にどのように付与されたかを概観してみたい。2.2 (3) で取り上げたものと同じ語彙素読みの語頭が「ア」であるもの 428 語について、専門度と意味分野をクロスし、それぞれのセルに所属する語数を集計すると表 2 のようになり、その比率を示したのが表 3 である。

表 2・表 3 によれば、「E 理念や制度など」に専門度の高い語が多く、反対に「F 健康や生活習慣など」には低い語が多いという特徴が顕著である。また、「A 身体部位など」「B 病気や症状など」「C 診療や検査など」「D 物質や薬剤など」は、専門度の高い語から低い語まで多様な段階の語があることもわかる。

表 2 専門度と意味分野（アで始まる語彙の語数）

専門度	A 身体部位など	B 病気や症状など	C 診療や検査など	D 物質や薬剤など	E 理念や制度など	F 健康や生活習慣など	G 医療と関わらないもの
1	8	31	7	22	2	26	11
2	7	29	9	18	2	19	5
3	8	23	4	12	12	8	7
4	7	11	8	17	11	2	11
5	4	19	5	11	32	5	15
計	34	113	33	80	59	60	49

表 3 専門度と意味分野（アで始まる語彙の比率）

専門度	A 身体部位など	B 病気や症状など	C 診療や検査など	D 物質や薬剤など	E 理念や制度など	F 健康や生活習慣など	G 医療と関わらないもの
1	23.5%	27.4%	21.2%	27.5%	3.4%	43.3%	22.4%
2	20.6%	25.7%	27.3%	22.5%	3.4%	31.7%	10.2%
3	23.5%	20.3%	12.1%	15.0%	20.3%	13.3%	14.3%
4	20.6%	9.7%	24.2%	21.3%	18.6%	3.3%	22.4%
5	11.8%	16.8%	15.2%	13.8%	54.2%	8.3%	30.6%
計	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

表 2 で示したもののうち「B 病気や症状など」に属する 113 語を例に、専門度別にリスト化して示すと下の通りである。

○専門度 1 (31 語)

赤み、亜急性硬化性全脳炎、悪性、悪性型、悪性度、悪性リンパ腫、アジソン病、汗疹（あせも）、悪化、圧迫、後縦靭帯、後囊、アトピー、アトピー克服、アトピー克服体験記、アトピー症状、アトピー性疾患、アトピー性皮膚炎、アトピー体質、アトピー肌、脂血症、アルコール依存症、アルツハイマー病、アレルギー、アレルギー症状、アレルギー性疾患、アレルギー性鼻炎、アレルギー体質、アレルギー反応、アレルギーマーチ、安静時狭心症

○専門度 2 (29 語)

亜急性甲状腺炎、明細胞腺癌、悪性化、悪性胸膜中皮腫、悪性黒色腫、悪性腫瘍、悪性貧血、亜型、亜型ウイルス感染、アジア風邪、足型、亜脱臼、悪化防止、悪化要因、圧迫骨折、アテローム、アテローム血栓性梗塞、後大脳動脈、アトピー素因、アルコール性肝障害、アルツハイマー、アレルギー検査、アレルギー疾患、アレルギー性、アレルギー性結膜炎、アレルギー性皮膚炎、アレルギー素因、アレルギー物質

○専門度 3 (23 語)

RA、ITC、アカシジア、悪性症候群、悪性新生物、悪性脳腫瘍、顎関節症、足関節、足白癬、アスピ

リン潰瘍、アスピリン不耐症、悪化因子、圧迫力、アトピー性皮膚炎患者、アトピー治療、後鼻漏、アナフィラキシー様症状、アルコール性、アルコール性肝炎、アルツハイマー型、アルツハイマー型認知症、アレルギー患者、アレルギー機序

○専門度4(11語)

I P F、亜急性、悪性疾患、足壊疽、脚症候群、アシドーシス、後首淋巴節、後負荷、脂肺肝、アレルギー科、アレルギー性結膜疾患

○専門度5(19語)

I S D、アウトブレイク、亜急性期、足潰瘍、足病変、脚ブロック、アスピリン喘息、アスピリン喘息患者、アスペルガ、汗過敏症、アナフィラキシー、アナフィラキシー様反応、アナフィラキシー症状、アナフィラキシーショック、アナフィラキシー反応、アメーバ性大腸炎、アレルギー性刺草疹、アレルギー性接触皮膚炎、アレルギー歴

専門度が低い語彙には、例えば、「アトピー」や「アレルギー」を含む語が多いのに対して、専門度が高い語彙には、「アスピリン」や「アナフィラキシー」を含む語が目立つ。「アトピー」「アレルギー」と「アスピリン」「アナフィラキシー」とを、経験的な内省によって比較すると、前者はより一般的であり後者はより専門的であるという違いが感じられる。したがって、ここで試みた専門度の付与は妥当な結果を導いている面があると言ってよいであろう。

3. 語彙頻度に基づく重要度の設定

3. 1 考え方

難解な専門用語であっても、それが非専門家の目や耳に触れる機会がないのであれば、難解用語の言語問題を生じることはない。非専門家がそれについて理解を求められる場合に問題が表面化する。したがって、難解な専門用語の中でも、非専門家が接する機会が多いものがこの問題における重要語ということになる。このような考え方から、非専門家を読者に想定した医療情報媒体(2.3(1)のb)においてよく使われる語彙を重要語として扱うことにした。

ここでは「よく使われる」ということを、使われる回数の多さ(度数)と、使われる場面の広さ(記事数)との二つの指標でとらえることにした。度数だけでは、特定の場面に偏って使われる語彙が過剰に評価されてしまうおそれがあり、記事数だけでは、繰り返し使われる大事な語が十分評価されないおそれがあるからである。ここでいう「記事」とは、雑誌や新聞では各記事、ウェブでは一つのページに表示される情報のひとまとまりを指す。

3. 2 医療分野のコーパスを用いた実践

2.3(1)のbの非専門家対象の医療記事において、各語の使用度数と使用記事数をもとに、表4・表5のように重要度のレベルを設定した。この作業は、医療分野のコーパスで2回以上使われ、かつ非専門家対象の医療分野のコーパスで1回以上使われている約90,000語を対象に行った。

表4 使用度数から見た重要度

重要度 (度数)	使用度数 の区間	語数
1	1	17,865
2	2	22,573
3	3~5	20,865
4	6~16	14,916
5	17~	13,465
計		89,684

表5 使用記事数から見た重要度

重要度 (記事数)	使用記事数 の区間	語数
1	1	30,028
2	2	19,785
3	3~4	13,816
4	5~12	13,502
5	13~	12,553
計		89,684

上記の二つの指標によって、どのような意味分野の語にどのレベルの重要度が付与されたかを具体的に見ていこう。2.2 (3) で取り上げた、抽出された専門用語のうち、語彙素読みの語頭が「ア」のもの428語を事例に検討したい。二種の重要度と意味分野をクロスし、それぞれのセルに属する語数を集計したものが、表6(度数から見た重要度)・表7(記事数から見た重要度)である。表8・表9は、それぞれの比率を示したものである。なお、428語のうち74語は非専門家向けの医療コーパスには出現しなかったため、これらの表に含まれるのは354語である。

表6 重要度(度数)と意味分野(アで始まる語彙の語数)

重要度 (度数)	A 身体部 位など	B 病気や 症状など	C 診療や 検査など	D 物質や 薬剤など	E 理念や 制度など	F 健康や生 活習慣など	G 医療と関わ らないもの
1	2	8	2	7	4	2	3
2	3	4	3	2	5	0	2
3	2	16	2	15	8	6	3
4	9	39	9	18	11	23	5
5	13	34	7	24	7	25	31
計	29	101	23	66	35	56	44

表7 重要度(記事数)と意味分野(アで始まる語彙の語数)

重要度 (度数)	A 身体部 位など	B 病気や 症状など	C 診療や 検査など	D 物質や 薬剤など	E 理念や制 度など	F 健康や生 活習慣など	G 医療と関わ らないもの
1	3	15	4	11	6	3	3
2	5	8	3	4	5	2	2
3	2	23	3	18	8	10	4
4	9	28	9	13	10	18	5
5	10	27	4	20	6	23	30
計	29	101	23	66	35	56	44

表 8 重要度（度数）と意味分野（アで始まる語彙の比率）

重要度 (度数)	A 身体部 位など	B 病気や 症状など	C 診療や 検査など	D 物質や 薬剤など	E 理念や 制度など	F 健康や生 活習慣など	G 医療と関わ らないもの
1	6.9%	7.9%	8.7%	10.6%	11.4%	3.6%	6.8%
2	10.3%	4.0%	13.0%	3.0%	14.3%	0.0%	4.5%
3	6.9%	15.8%	8.7%	22.7%	22.9%	10.7%	6.8%
4	31.0%	38.6%	39.1%	27.3%	31.4%	41.1%	11.4%
5	44.8%	33.7%	30.4%	36.4%	20.0%	44.6%	70.5%
計	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

表 9 重要度（記事数）と意味分野（アで始まる語彙の比率）

重要度 (度数)	A 身体部 位など	B 病気や 症状など	C 診療や 検査など	D 物質や 薬剤など	E 理念や 制度など	F 健康や生 活習慣など	G 医療と関わ らないもの
1	10.3%	14.9%	17.4%	16.7%	17.1%	5.4%	6.8%
2	17.2%	7.9%	13.0%	6.1%	14.3%	3.6%	4.5%
3	6.9%	22.8%	13.0%	27.3%	22.9%	17.9%	9.1%
4	31.0%	27.7%	39.1%	19.7%	28.6%	32.1%	11.4%
5	34.5%	26.7%	17.4%	30.3%	17.1%	41.1%	68.2%
計	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

度数による重要度においても記事数による重要度においても、全般に重要度が低い語彙よりも重要度が高い語彙が多いことが見て取れる⁷。これは、専門用語として抽出された語彙の多くは、非専門家が読むものにも頻出する重要な語彙であるということを示すものである。意味分野ごとに見ても、すべての意味分野で重要度1・2よりも重要度4・5の方が語数が多く、特に「A身体部位など」「B病気や症状など」「D物質や薬剤など」「F健康や生活習慣など」でその傾向が強い。「C診療や検査など」「E理念や制度など」は、重要度の低い語や中程度の語も少なくなく、この二つの意味分野は、他の意味分野に比べると非専門家が目にする機会が少ないということを示している。

表 10 は、「B病気や症状など」の 101 語について、二種の重要度をクロスし具体的な語彙をリスト化したものである。この表において、例えば、もっとも右下のセル（度数・記事数ともに重要度5）には、「悪性腫瘍」「汗疹（あせも）」「アトピー」「アルツハイマー」「アレルギー」など、一般によく知られていそうな病気や症状を表す語が並んでいる。左上方のセル（レベル1または2）には「アスベルガ」「アカンジア」のように、あまりなじみのない病名や症状名が目立つ。また、「アレルギー」を含む複合語を見ても、もっとも右下のセルには「アレルギー疾患」「アレルギー症状」「アレルギー性」「アレルギー性疾患」「アレルギー性鼻炎」「アレルギー体質」「アレルギー反応」のように、非専門家でもよく見聞きしそうな語が多いが、左上方のセルには「アレルギー機序」「アレルギー性結膜疾患」「アレルギー歴」など、比較的なじみの薄そうな語が並んでいる。このように、非専門家の目に触れる機会が多いという重要さの観点からは、表の右下のものほど重要度が高く、表の左上のものほど重要度が低いというに見なしていくことができよう。

⁷ 度数による重要度と記事数による重要度を比較すると、度数による重要度の方が、高い重要度に集中する傾向がより強い。

表 10 重要度（度数・記事数）別の語彙リスト（B 病気や症状など）

	記事数 1	2	3	4	5
度数 1	亜急性、アスピリン潰瘍、アスペルガ、アナフィラキシー反応、脂肺肝、アレルギー機序、アレルギー性結膜疾患、アレルギー歴				
2	アスピリン不耐症、アナフィラキシー症状	アカシジア、足壊疽			
3	RA、ITC、アルコール性肝炎	悪性症候群、足潰瘍、後鼻漏、アナフィラキシー様症状	悪性脳腫瘍、足型、アシドーシス、足白癬、圧迫力、アトピー治療、アレルギー患者、アレルギー物質	悪性新生物	
4	明細胞腺癌、亜型ウイルス感染	悪性型、亜脱臼	亜急性硬化性全脳炎、亜急性甲状腺炎、悪性胸膜中皮腫、悪性貧血、アジソン病、足病変、アテローム血栓性梗塞、後大脳動脈、後縦靭帯、アトピー性疾患、アトピー性皮膚炎患者、アルコール性、アレルギー素因、安静時狭心症	悪性化、悪性黒色腫、悪性疾患、亜型、アジア風邪、足関節、脚症候群、悪化防止、悪化要因、アテローム、アトピー克服体験記、アトピー症状、アトピー素因、アナフィラキシー、アルコール性肝障害、アルツハイマー型、アレルギー科、アレルギー検査、アレルギー性皮膚炎、アレルギーマーチ	
5			後囊	顎関節症、アスピリン喘息、悪化因子、アトピー肌、アナフィラキシーショック、アレルギー性結膜炎	赤み、悪性、悪性腫瘍、悪性度、悪性リンパ腫、汗疹（あせも）、悪化、圧迫、圧迫骨折、アトピー、アトピー克服、アトピー性皮膚炎、アトピー体質、脂血症、アルコール依存症、アルツハイマー、アルツハイマー型認知症、アルツハイマー病、アレルギー、アレルギー疾患、アレルギー症状、アレルギー性、アレルギー性疾患、アレルギー性鼻炎、アレルギー体質、アレルギー反応

4. 抽出・序列化された用語の評価

4. 1 評価に用いるデータ

ここまでにおいてコーパスを用いて試みた、医療専門用語の抽出と専門度の付与（2）、重要度の付与（3）について、コーパス以外のデータを用いて評価を行ってみたい。評価の手段には、国立国語研究所「病院の言葉」委員会の活動のために実施した質問調査のデータを用いる。「病院の言葉」委員会は、患者にとって重要でありながら分かりにくい語を取り上げて分かりやすくする工夫について提案したが、その活動の過程で、医療用語 100 語について、国民に対する認知度等の調査と、医療者に対する必要度等の調査を実施した⁸。このうち、国民に対する認知度等の調査を専門用語抽出と専門度付与の評価に、医療者に対する必要度等の調査を重要度付与の評価に用いてみたい。

4. 2 専門度の評価

まず、国民に対する認知度等の調査結果と、2 で実施した専門用語抽出と専門度付与の結果とを対照してみたい。国民に対する調査は、医療用語 100 語について、20 歳以上の男女 10,811 人に対して、年齢と居住地（都道府県）とにバランスをとって協力を依頼し、4,276 人から回答が得られたものである。インターネットを介して次のような質問を行い、「ある」と回答した人が占める比率を「認知率」として集計した。

質問：あなたは〇〇という言葉を見たり聞いたりしたことがありますか。次の二つから選んでください。 → （1）ある （2）ない

この調査にかけた 100 語のうち、「かかりつけ医」「カテーテル」「振戦」「ノロウイルス」「ポリープ」「レシピエント」「DIC」「MRSA」の 8 語は、今回のコーパスによる語彙分析では取得できなかった。その理由は、コーパスの形態素解析に用いた解析辞書 UniDic に、解析を実施した 2007 年段階ではこれらの語が登録されていなかったことによるのではないかと思われる。現在の UniDic には登録されている可能性はあるが、解析辞書に登録されていない語はコーパスによる語彙分析の対象にできないことに注意が必要である。

残りの 92 語について、調査による認知率と、2 で実施した専門用語抽出と専門度付与の結果との関係をまとめると表 11 のようになる（数字は語数）。

表 11 認知率と専門度（語数）

認知率	専門用語	非専門用語	専門度 5	専門度 4	専門度 3	専門度 2	専門度 1
0-10%	2		2				
10-20%	5		2	1	1	1	
20-30%	7		4			3	
30-40%	1		1				
40-50%	3	1				1	2
50-60%	8		2	1	2	2	1
60-70%	6		1			2	3
70-80%	7		6	1			
80-90%	11	1	2	1		4	4
90-100%	39	1	1	1	2	2	33
計	89	3	21	5	5	15	43

⁸ 調査の方法と結果の詳細は <http://www.ninjal.ac.jp/byoin/tyosa/> で公開している。

「病院の言葉」委員会という専門家集団が「分かりにくい語」として選んだ92語のうち89語(96.7%)が、コーパスの語彙頻度によって抽出した専門用語と一致していることが確認できたことは、この抽出方法が一般の人には分かりにくい専門用語をよく抽出できていることを示している。

この方法では専門用語として抽出されなかった「非専門用語」が3語あるが、それは「うっ血」「クオリティーオブライフ」「ショック」である。このうち、「うっ血」は、医療分野では平仮名と漢字が混ざり書きされるこの表記が普通だが、形態素解析ではこの表記が認識されず、「鬱血」と漢字表記されたもののみが認識され「うっ血」と交ぜ書きされたものは正しく認識されなかったという事情がある。「鬱血」という漢字表記語だけの頻度だと非専門用語と判定されてしまうのである。また、「ショック」は、医療用語としての「血液の循環がうまくいかなくなって、場合によっては命の危険があること」という意味とは別に、「びっくりすること」「衝撃を受けること」などの意味で一般用語としてよく用いられるために、語義を区別しない形態素解析では非専門用語とされたものである。それから、「クオリティーオブライフ」は、医療分野では略語の「QOL」の形で用いられるのが通常で、片仮名の語形はあまり用いられないことによるものである。いずれも明確な事情があって抽出できなかったものであり、こうした事情がないものはすべて抽出可能である。

表11で専門度の5段階を示した「専門度5」から「専門度1」までの分布を見ると、概して、認知率が低いところには専門度の高い語が多く、認知率が高いところには専門度の低い語が比較的多い、という傾向が認められる。コーパスの語彙頻度によって付与した専門度の高低が質問調査による認知率の高低に対応することが確認でき、この方法による専門度付与の結果がおおむね妥当なものであるということが分かる。

一方で、認知率70-80%のところでは専門度5に集中している点や、認知率80-90%、90-100%のところでは専門度が5または4と高いものが見られる点、あるいは認知率40-50%、50-60%のところでは専門度1の語がある点など(表中の太線で囲んだ部分)、一部に全体の傾向に外れるものもあり、語彙頻度による専門度が、質問調査による認知率とそのままでは対応しない面があることにも注意が必要である。

表12は、表11の専門用語89語について、具体的な語を表示してリスト化したものである。認知率が最低レベルで専門度が最高レベルのセルには「EBM」「クリニカルパス」といった一般の人にはほとんどなじみのない語が配され、反対に認知率が最高レベルで専門度が最低レベルのセルには「化学療法」「臨床試験」など比較的なじみのありそうな語が配されている。一方、同じ専門度5であっても認知率はかなり高いものもあり、専門度1であっても認知率が低いものもある。認知率と専門度とが対応していない部分として指摘した表11の太線部分を見ても、そこに属する語彙に特定の事情を指摘することは難しい⁹。このような例外的な語が少ない理由については、さらに検討が求められるところである。

⁹ 認知率が極めて高く専門度5の「リスク」「ガイドライン」は、医療用語とは別に一般用語としても用いられるという事情が指摘できるが、他はそのような事情は指摘できない。

表 12 認知率と専門度（語彙リスト）

認知率	専門度 5	専門度 4	専門度 3	専門度 2	専門度 1
0-10%	E B M、クリニカルパス				
10-20%	C O P D、Q O L	イレウス	集学的治療	寛解	
20-30%	エビデンス、せん妄、プライマリーケア、ADL			日和見感染、間質性肺炎、H b A 1 c	
30-40%	ターミナルケア				
40-50%				虚血性心疾患	浸潤、生検
50-60%	重篤、誤嚥	予後	塞栓、緩和ケア	統合失調症、耐性	ネフローゼ症候群
60-70%	コンプライアンス			狭窄、治験	P E T、対症療法、腫瘍マーカー
70-80%	敗血症、インフォームドコンセント、グループホーム、既往歴、肺水腫、川崎病	抗生剤			
80-90%	介護老人保健施設、ガイドライン	術後合併症		髄膜炎、膠原病、頓服、CT	セカンドオピニオン、肉腫、慢性腎不全、ホスピス
90-100%	リスク	院内感染	尊厳死、MRI	黄だん、悪性腫瘍	化学療法、臨床試験、悪性リンパ腫、壊死、抗体、ステロイド、狭心症、血栓、インスリン、血糖、腎不全、自律神経失調症、肝硬変、動脈硬化、潰瘍、合併症、ぜん息、脳死、炎症、メタボリックシンドローム、腫瘍、免疫、がん、心筋梗塞、抗がん剤、白血病、副作用、糖尿病、うつ病、熱中症、貧血、ウイルス、インフルエンザ

4. 3 重要度の評価

次に、医療者に対する必要度等の調査結果と、3で行った語彙頻度による重要度の付与の結果とを対照してみたい。4.2で記したのと同様にUniDicに収録されていなかったなどの理由で、「かかりつけ医」「カテーテル」「振戦」「ノロウイルス」「ポリープ」「レシピエント」「DIC」「MRSA」の8語はコーパスの語彙解析で取得できなかったため、考察できない。

医療者に対する必要度等の調査は、国民に対する調査と同じ100語について¹⁰、医師3,000人と看護師・薬剤師1,280人に依頼し、医師685人、看護師・薬剤師1,002人から回答が得られたものである。インターネットを介して次のような質問を行い、回答された選択肢を(1)1点、(2)2点、(3)3点、(4)4点と得点化して、その平均を「必要度」として

¹⁰ 正確に言うと1語だけ異なっている。「術後合併症」は国民調査にのみ、「ケアプラン」は医療者調査にのみかけている。

算出した。

質問：あなたの仕事の場で、以下の言葉を、患者やその家族に理解してもらうことは必要ですか。次の4段階でご回答ください。

- (1) 全く必要でない (2) あまり必要でない
(3) やや必要である (4) 大いに必要である

必要度を0.2ポイントずつの区間（最も低い区間のみ0.3ポイント）に分け、医師の場合（表13）と看護師・薬剤師の場合（表14）それぞれで、コーパスの語彙頻度による重要度との対応を整理した（数字は語数）。この表の重要度は、3で記した、度数の重要度と記事数の重要度の二種の数値を合算したものであり、2ポイントずつを一区画とした。

表13 医師による必要度判定と重要度の関係（語数）

医師の必要度	重要度 9-10	重要度 7-8	重要度 5-6	重要度 3-4	重要度 1-2	一般向け媒体になし
3.6-3.8	4					
3.4-3.6	12					
3.2-3.4	28	3				
3.0-3.2	16	1	1			
2.8-3.0	15	3				1
2.6-2.8	3	3	1		1	
2.3-2.6						
計	78	10	2	0	1	1

表14 看護師・薬剤師による必要度判定と重要度の関係（語数）

看護師・薬剤師の必要度	重要度 9-10	重要度 7-8	重要度 5-6	重要度 3-4	重要度 1-2	一般向け媒体になし
3.6-3.8	1					
3.4-3.6	9	1				
3.2-3.4	24	1				
3.0-3.2	26	3				
2.8-3.0	9	2	1			1
2.6-2.8	6	1	1			
2.3-2.6	3	2			1	
	78	10	2	0	1	1

表13・表14いずれにおいても、全体に重要度9-10という最も高いレベルの語が多く¹¹、コーパスの語彙頻度による重要度の序列化が必ずしも有効に機能していないのではないかと思わせる。しかし、これは「病院の言葉」委員会による選定が、最重要語の選定をしていることを示していると考えられるべきだろう。現に、表13・表14の中で必要度と重要度の相関を見ると、どちらの表においても、必要度の高いところほど重要度の高い語の比率が高く、必要度の低いところほど重要度の高くない語の比率が高くなるという、全体的な傾向は認められる。このことから、語彙頻度によって付与した重要度は、専門家が判断する必要度と相関する面があると考えてよさそうである。もっと重要度の低い語もあわせて評価することができれば、語彙頻度による重要度の序列化の有効性が確かめられる可能性がある

¹¹ 表9と表10を比較すると、看護師・薬剤師よりも医師の方が必要度の高い部分に集中しており、看護師・薬剤師よりも医師の方が、多くの語について患者の理解する必要度が高いと考えていることが分かる。

ろう。

表 13 の医師の場合について必要度と重要度の関係について、具体的な語彙を表に配属させてリスト化すると、表 15 のようになる¹²。

表 15 医師による必要度判定と重要度の関係（語彙リスト）

医師の必要度	重要度 9-10	重要度 7-8	重要度 5-6	重要度 3-4	重要度 1-2	一般向け媒体になし
3.6-3.8	合併症、抗生剤、糖尿病、副作用					
3.4-3.6	悪性腫瘍、炎症、がん、血糖、誤嚥、腫瘍、対症療法、貧血、予後、リスク、CT、MRI					
3.2-3.4	インスリン、インフォームドコンセント、インフルエンザ、ウイルス、うつ病、潰瘍、化学療法、肝硬変、狭窄、狭心症、虚血性心疾患、血栓、抗がん剤、腫瘍マーカー、重篤、ショック、心筋梗塞、腎不全、ステロイド、生検、ぜん息、ターミナルケア、動脈硬化、敗血症、慢性腎不全、メタボリックシンドローム、HbA1c、QOL	イレウス、クオリティオブライフ、頓服				
3.0-3.2	悪性リンパ腫、院内感染、壊死、黄だん、間質性肺炎、緩和ケア、既往歴、抗体、浸潤、髄膜炎、セカンドオピニオン、せん妄、尊厳死、耐性、ホスピス、免疫	日和見感染	肺水腫			
2.8-3.0	エビデンス、介護老人保健施設、ガイドライン、寛解、グループホーム、膠原病、自律神経失調症、治験、肉腫、ネフローゼ症候群、脳死、白血病、臨床試験、COPD、PET	集学的治療、塞栓、ADL				うっ血
2.6-2.8	ケアプラン、統合失調症、熱中症	コンプライアンス、プライマリケア、EBM	クリニカルパス		川崎病	
2.3-2.6						

一般向け媒体にない「うっ血」の特殊な事情については前述した。重要度が極めて低い「川崎病」は、この病気の発見者を指す「川崎」を都市名の「川崎」と誤認することで、この病気を公害病であると誤解する人がどの程度いるのかを調べる、特別な目的をもっていたものである¹³。また、重要度が高くない「クリニカルパス」は、登場してまもない医療安全ツールを指すこの語について¹⁴、その普及の重要性をにらんで「病院の言葉」委員会が選定したものである。同じく重要度の高くない「肺水腫」も、それほど頻繁に話題になる病気ではないといったような事情があるのではないかと考えられる。このように例外的に低い重要度となる語には、何らかの事情が想定できそうである。

表 15 で大部分を占める重要度 9-10 の語彙について、医師の必要度が最高レベルである「合

¹² 看護師・薬剤師の場合も、類似した語彙リストになるため、ここへの掲載は省略した。

¹³ 「川崎病」とは、主に乳幼児に見られる、全身の血管に炎症が起きる病気のことである。

¹⁴ 「クリニカルパス」とは、診療内容をスケジュール化して分かりやすく示したものである。

併症」「抗生剤」「糖尿病」「副作用」と、それが最低レベルである「ケアプラン」「統合失調症」「熱中症」を比べると、確かに、前者の語彙の方が後者の語彙よりも重要な語彙であるという気がする。前者は、医学あるいは医療の基本的な重要概念としての性格が強いものに対して、後者は、医学や医療の中心からは外れたり個別の病気を表すものだったりして、基本的な重要概念としての性格があまり強くないのではないだろうか。同じく病名であっても「糖尿病」は、生活習慣病として患者やその予備群も多く医療の対象として極めて重要視されるのに対して、「統合失調症」「熱中症」は、重要視される度合いが比較的低いのではないかと考えられる。しかし、こうした重要度の差異は、コーパスの語彙頻度に必ずしも反映するわけではないとも考えられ、語彙頻度に基づく重要度の設定に限界があることを思わせるものだろう。

5. おわりに

以上、難解用語の言語問題に取り組むための基盤として、非専門家にとって難解でありながら重要な語彙を、大規模コーパスの語彙頻度を用いて抽出し、序列化する方法について、医療用語を例に試行した結果について報告した。

「非専門家にとって難解でありながら重要な語彙」という場合、「難解さ」へのアプローチと「重要さ」へのアプローチの二つが必要である。本稿では、「難解さ」については、タイプの異なるコーパスを作成して相互に語彙を比較することで、専門用語の抽出と専門度による序列化を試みた。また、「重要さ」については、非専門家を対象にした媒体での語彙頻度の高さという観点から、重要度による序列化を試みた。その結果、全体としては、目的とする語彙を抽出し、抽出した語彙を序列化することに、おおむね成功しているところが確かめられた。難解用語の言語問題に取り組む基盤として、コーパスの語彙頻度による基礎データの整備が有望であるという見通しが得られたと言ってよいだろう。

しかし一方、目的外の語彙が抽出されたり、序列化が十分にできないところがあったりと、コーパスの語彙頻度に基づくデータだけでは、色々と課題が残ることも明らかになった。その課題には、まず、コーパスの質や量あるいは解析の精度を向上させること、あるいは統計指標を工夫することなど、コーパス言語学の範囲内で研究すべきところが大きい。一方、「難解さ」「重要さ」の問題のすべてが語彙頻度に反映するわけではないため、今回試みたような方法だけでは、限界があることも確かである。その限界を見きわめた上で、コーパス以外の方法と役割を分担したり関連づけたりしながら、研究を進めていくことが強く求められよう。

文献

久保順子、辻慶太、杉本重雄（2010）「異なる学問分野のコーパスを利用した専門用語抽出手法の提案」情報知識学会誌, 20-1, pp.15-31.

国立国語研究所「外来語」委員会（2006）『分かりやすく伝える 外来語言い換え手引き』, ぎょうせい.

国立国語研究所「病院の言葉」委員会（2009）『病院の言葉を分かりやすく—工夫の提案—』, 勁草書房.

田中牧郎（2009）「言語政策に役立つコーパスを用いた語彙表・漢字表などの作成と活用」人工知能学会誌, 24-5, pp.665-672.

- 田中牧郎、相澤正夫（2010）「難解用語の言語問題への具体的対応—「外来語」と「病院の言葉」を分かりやすくする提案—」社会言語科学, 13-1, pp.95-108.
- 田中牧郎、金愛蘭、桐生りか、近藤明日子（2008）「コーパスによる難解語・重要語の抽出—医療用語を例に—」社会言語科学会第21回大会発表論文集, pp.296-299.
- 中川裕志、湯本紘彰、森辰則（2003）「出現頻度と接続頻度に基づく専門用語抽出」自然言語処理, 10-1, pp.27-45.
- 日本弁護士連合会（2008）『裁判員時代の法廷用語—法廷用語の日常語化に関するPT最終報告書』, 三省堂.
- Kiyotaka Uchimoto, Satoshi Sekine, Masaki Murata, Hiromi Ozaku, and Hitoshi Isahara (2000) "Term Recognition Using Corpora from Different Fields" *Terminology*, 6-2, Corpus-Based of the Corpus Terminology Linguistics, pp.233–256.

第2章 第3節

漢字政策に役立つ漢字表のあり方

—固有名に使われる漢字の検討のために—

相澤 正夫

1. はじめに

人名・地名など固有名に使われる漢字は、それ以外の一般語に使われる漢字と区別して扱う必要がある。漢字政策の検討のための基礎資料となる漢字表も、この区別を反映させたものでなければ、実はあまり役に立たない。

固有名と一般語は語彙レベルにおける区別の一つであるから、固有名に使われる漢字と一般語に使われる漢字を区別して扱うためには、まず、この両者を区別して扱えるような語彙調査を実施し、それに基づいて語彙表を作成する必要がある。このような語彙表には固有名と一般語を区別した形で個々の語が列挙されるが、それら個々の語の表記に使われた漢字について、今度は逆に個々の漢字（字種）を単位としてまとめあげれば、漢字ごとにどんな固有名で使われているのか、あるいはどんな一般語で使われているのかが分かる便利な漢字表となる。国立国語研究所『現代新聞の漢字』（1976）所収の「第Ⅱ表 用語例表」は、その一つの具体例である。また、国立国語研究所『『現代雑誌の語彙調査』に基づく漢字音訓一覧表』（2005）は、近年の語彙調査に基づき同様の形式を踏襲して作成された漢字表である。

特定領域研究「日本語コーパス」では、大規模な『現代日本語書き言葉均衡コーパス』（以下、BCCWJ）を開発する一方で、それと連携しながら詳細な「形態論情報データベース」の構築を推進しており、最終的には、上述のような便利な漢字表の作成に必要な情報が、ほぼ自動的に取得できるようになる見通しである。

本稿は、そのような本格的なコーパス活用のための準備として、既に整備されているBCCWJのコアデータを利用して行った試行的調査の報告である。固有名に使われる漢字を検討するためには、「教育用漢字」「常用漢字」「表外漢字」など既存の漢字政策カテゴリにとらわれることなく、全ての漢字について、固有名での使用率、本稿で提案する「固有名表記への傾斜度」を、基礎的な数値情報として把握しておくべきことを述べたい。

2. 漢字ごとの「固有名表記への傾斜度」

一般に、固有名に使われる漢字というと、「阪」や「岡」のように「固有名詞専用字」と呼ばれるような漢字に目が向きがちである。「岡」については「岡っ引き」という一般語も思い付くが、あまり使用頻度の高い語とは言えず、これらの漢字の固有名での使用率が100パーセントか限りなくそれに近いことは容易に想像がつく。

それでは、「田中」という姓の「田」や「中」は、どうであろうか。どちらも小学校第1学年で学習する「教育用漢字」であり、一般語にもよく使われていそうな漢字であるが、実際はどのようなのであろうか。試みに『『現代雑誌の語彙調査』に基づく漢字音訓一覧表』（2005）から、該当部分を引用すると、次の表1に示す「田」「中」のとおりである。

表1 「田」と「中」の固有名での使用状況

見出し	総合 頻度	音 訓	頻 度	語 例
田	1413	デン	17	水田(7) 田圃(10).
		た	32	小田急(22) 田んぼ(7) 田圃(3).
		(いなか) ▲	15	田舎▲(15).
		人名	947	飯田(12) [姓] 池田(33) [姓] 石田(22) [姓] 岩田(15) [姓] 上田(13) [姓] 植田(8) [姓] 内田(15) [姓] 太田(19) [姓] 岡田(22) [姓] 奥田(7) [姓] 小田(8) [姓] 織田(9) [姓] 鎌田(8) [姓] 川田(12) [姓] 久保田(9) [姓] 額田(7) [姓] 黒田(15) [姓] 郷田(11) [姓] 坂田(10) [姓] 真田(10) [姓] 沢田(6) [姓] 澤田(4) [姓] 高田(29) [姓] 竹田(10) [姓] 武田(21) [姓] 田島(9) [姓] 柴田(20) [姓] 島田(16) [姓] 杉田(8) [姓] 田中(84) [姓] 田原(8) [姓] 田宮(6) [姓] 田村(15) [姓] 塚田(9) [姓] 津田(7) [姓] 土田(7) [姓] 寺田(7) [姓] 戸田(8) [姓] 豊田(8) [姓] 中田(16) [姓] 永田(11) [姓] 成田(8) [姓] 西田(10) [姓] 野田(12) [姓] 羽田(11) [姓] 原田(23) [姓] 平田(9) [姓] 福田(18) [姓] 藤田(17) [姓] 本田(18) [姓] 前田(25) [姓] 増田(12) [姓] 松田(14) [姓] 宮田(16) [姓] 村田(9) [姓] 森田(14) [姓] 安田(11) [姓] 山田(57) [姓] 吉田(72) [姓] 和田(17) [姓].
地名	402	秋田(46) 飯田(7) 上田(7) 梅田(18) 太田(10) 大田(8) 小田原(15) 神田(13) 神田神保町(16) 下田(10) 吹田(11) 世田谷(44) 高田(7) 高田馬場(10) 千代田(64) 豊田(13) 十和田(7) 成田(38) 羽田(16) 町田(16) 三田(12) 早稲田(14).		
中	3196	チュウ	1572	海中(9) 空中(11) 最中(9) 集中(82) 初・中級(2) しよつ中(1) 水中(7) 中(1) [中心の意] 中(1) [中国地方] 中(128) [世界~] 中(532) [募集~] 中央(150) 中・大型(1) 中華(26) 中学(43) 中間(22) 中期(11) 中級(15) 中継(10) 中古(59) 中止(18) 中旬(25) 中・初級(1) 中心(199) 中世(11) 中性(7) 中絶(7) 中途(10) 中年(14) 中盤(12) 中火(8) 中部(28) 中路(12) 途中(58) 日中(12) 夢中(16) 連中(14).
		なか	921	ただ中(5) 只中(1) 背中(34) 中(764) 中々(1) 中庭(10) 中身(24) 中味(1) 中リール(9) まん中(5) 真ん中(15) 真中(2) 夜中(21) 世の中(29).
		人名	349	田中(84) [姓] 中井(14) [姓] 中尾(11) [姓] 中川(22) [姓] 中沢(10) [姓] 中島(24) [姓] 中嶋(10) [姓] 中田(16) [姓] 中西(19) [姓] 中野(24) [姓] 中原(18) [姓] 中村(80) [姓] 中山(8) [姓] 山中(9) [姓].
		地名	354	中央(72) 中京(15) 中国(20) [~地方] 中国(178) 中禅寺(7) 中(22) 中野(17) 中村(9) 府中(7) 山中(7).

『『現代雑誌の語彙調査』に基づく漢字音訓一覧表』(2005)より

総合頻度は、「田」が1413、「中」が3196と、両者の間には大きな差があり、総合頻度の順位も「中」の方が「田」よりもかなり上位にくることは明らかである。単漢字の集計で分かるのはこのレベルであるが、表1のように、一般語については音と訓(熟字訓も含む)

の区分で、また、固有名については人名と地名の区分で整理すると、個々の漢字の個性がより鮮明に見えてくる。ここでの関心事である一般語と固有名という対立する2項の実際使用面での勢力関係をみると、「田」は固有名の方に、「中」は一般語の方にそれぞれ明らかに傾斜していることが見て取れる。

ここでは、個々の漢字が固有名の表記に使われる割合を、次に示す計算式で算出し、その数値を「固有名表記への傾斜度」と呼ぶことにする。

$$\text{固有名表記への傾斜度} = [\text{人名・地名等の固有名における出現頻度}] / [\text{総合頻度}]$$

この計算式で「田」と「中」の固有名表記への傾斜度を算出してみよう。

$$\text{「田」} \quad (\text{人名 } 947 + \text{地名 } 402) / (\text{総合頻度 } 1413) = 0.9547$$

$$\text{「中」} \quad (\text{人名 } 349 + \text{地名 } 354) / (\text{総合頻度 } 3196) = 0.2200$$

このように実際に数値化してみると、「田」が大きく固有名表記に傾斜した漢字であるのに対して、「中」は固有名よりはむしろ一般語表記に傾斜した漢字であることが分かり、両者の実際使用面での違いが浮き彫りになる。本稿では、以下で BCCWJ のコアデータを利用して作成した漢字表（頻度表）の中に、固有名表記への傾斜度（＝固有名割合、固有名使用率）の項目を用意し、総合頻度順位や漢字政策カテゴリーとの関わりを観察する。

3. BCCWJ コアデータに基づく漢字表

試行的に分析対象として利用するのは、BCCWJ の「コア」に格納するサンプルのうちの「固定長」部分である。サンプルの抽出対象としたジャンルは、「新聞、雑誌、書籍、白書、Yahoo!知恵袋」である。作成した漢字表（頻度表）の仕様は、次のとおりである。

- 順位
- 種類 「常用漢字表」(1981) の表内漢字 教育, 常用
表外漢字 人A, 人B, 人C,
第1, 第2, 第3, 第4
- 字種
- 総合頻度
- 一般語頻度合計 総合頻度から固有名に使われた数を引いた頻度
- 固有名頻度合計 固有名に使われた頻度
- 固有名割合 「固有名頻度合計÷総合頻度×100」で算出
- 地名頻度 地名で使われた頻度
- 人名頻度 人名(姓・名・あだ名・しこ名等)で使われた頻度
- 固有名一般頻度 地名・人名以外の固有名(元号・商品名・組織名等)で使われた頻度

上記の仕様のうち、「種類」の欄には従来の漢字政策カテゴリーの略称を示した。「常用漢字表(1981)」に掲げられた表内漢字のうち、「教育用漢字」を特に「教育」として示し、それ以外を「常用」とした。表外漢字の区分は少々複雑であるが、概略、次に示すとおりである。

- 人A 表外漢字字体表（2000.12）が答申される以前に、人名用漢字として認められていた漢字。（285字）
- 人B 「戸籍法施行規則」（2004.9 改正）の「別表第二 漢字の表」（人名用漢字の一覧表）の「一」にある774字から「人A」を除いた漢字。（489字）
- 人C 「戸籍法施行規則」（2004.9 改正）の「別表第二 漢字の表」（人名用漢字の一覧表）の「二」にある漢字。（209字）
- 第1 人名用漢字（人A：245字，人B：454字，人C：10字）を除いたJIS第1水準に属する表外漢字。
- 第2 人名用漢字（人A：40字，人B：32字，人C：116字）を除いたJIS第2水準に属する表外漢字。
- 第3 人名用漢字（人B：3字，人C：83字）を除いたJIS第3水準に属する表外漢字。
- 第4 JIS第4水準に属する漢字。

また、ここでは「固有名」の下位区分を「地名」「人名」及びそれ以外の「固有名一般」とした。したがって、「固有名頻度」は、「地名頻度」「人名頻度」「固有名一般頻度」の合計となる。

調査の結果、出現した漢字の総合頻度の合計（延べ字数）は261,223字、字種数（異なり字数）は2,677字であった。本稿の末尾に、上記の仕様によって作成した漢字表のうち、総合頻度上位499位までを「別表」として掲げた。

ちなみに、第2節で話題にした4つの漢字について確認すると、「田」は総合頻度559で99位、「中」は総合頻度1722で8位と、両者の順位差は大きいものの、いずれも上位100位までに入っている。本稿の関心事である「固有名割合」については、「田」93.74%、「中」31.30%と、『現代雑誌の語彙調査』に基づく漢字音訓一覧表』（2005）における「田」95.47%、「中」22.00%と、おおむねよく似た傾向を示している。「田」は固有名の方に大きく傾斜し、「中」は一般語の方に傾斜していることが、改めて確認されたとと言えるだろう。

一方、漢字政策カテゴリーで「人B」に分類される「阪」は総合頻度171で422位、「岡」は総合頻度152で472位と、いずれも「別表」では後ろの方になるが、全体の500位以内に入り、かなり上位に位置している。「固有名割合」も、「阪」99.42%、「岡」99.34%と、十分に「固有名専用字」と言ってよい数値を示している（第5節の表2も参照）。

4. 固有名表記から見た漢字表データの概観的分析

合計（延べ字数）261,223字、字種数（異なり字数）2,677字の漢字表データに基づき、総合頻度を出現順位のランク別（100位ごと）に示したのが図1である。横軸の数字は、「1」が1～100位、「2」が101～200位、以下同様に続き、「27」が2601～2677位である。グラフには、一般語頻度合計と固有名頻度合計を内訳として示した。

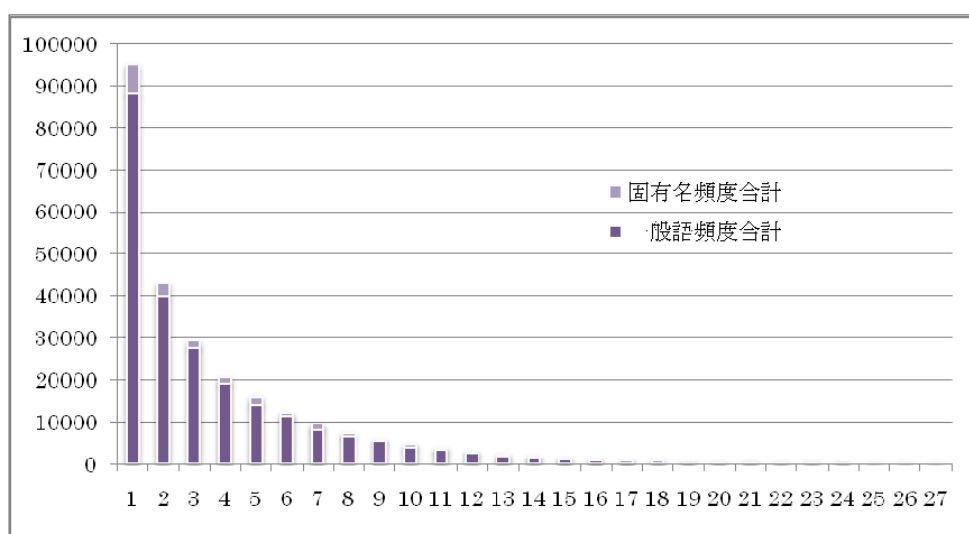


図1 総合頻度（出現順位ランク別）

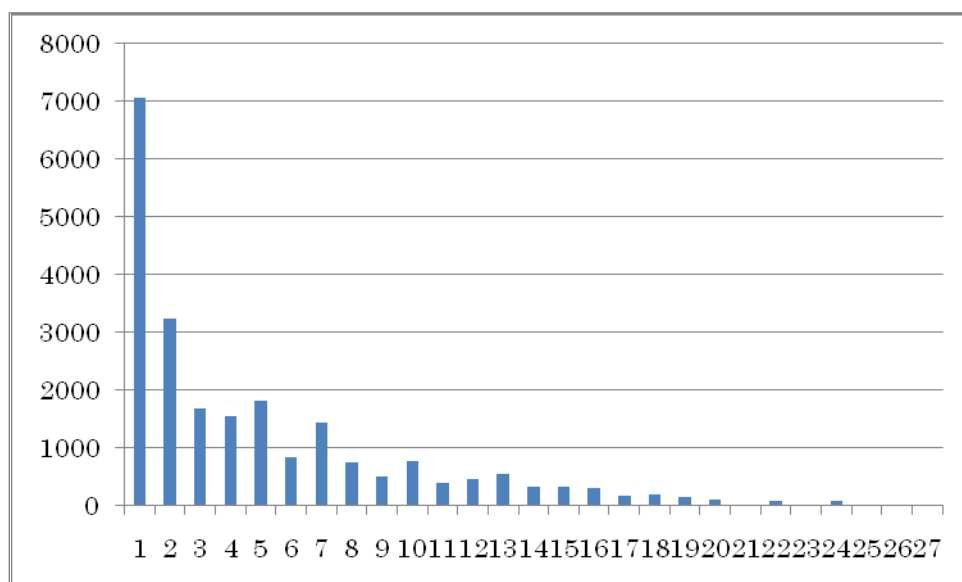


図2 固有名称頻度（出現順位ランク別）

図1から分かるように、総合頻度は出現順位ランク500～600位あたりまでは急激に減少するが、それ以下ではきれいなカーブを描いて緩やかに減少していく。図1から固有名称頻度合計だけを取り出したのが図2である。固有名称頻度は300位あたりまでは総合頻度と同様に急激に減少するが、それ以降はかなりの凸凹を見せながらきわめて緩やかに減少していく。

一方、「固有名称割合」を出現順位ランク別（100位ごと）に示したのが、図3である。2000位以下で大きく乱高下を見せるのは、1ランク（100位あたり）の頻度が100を割り込むために、特定の漢字の特徴が反映されやすいことよると考えられる。全体の傾向をつかむために、順位2000位までについて500位ごとのランク別に集計したのが、図4である。

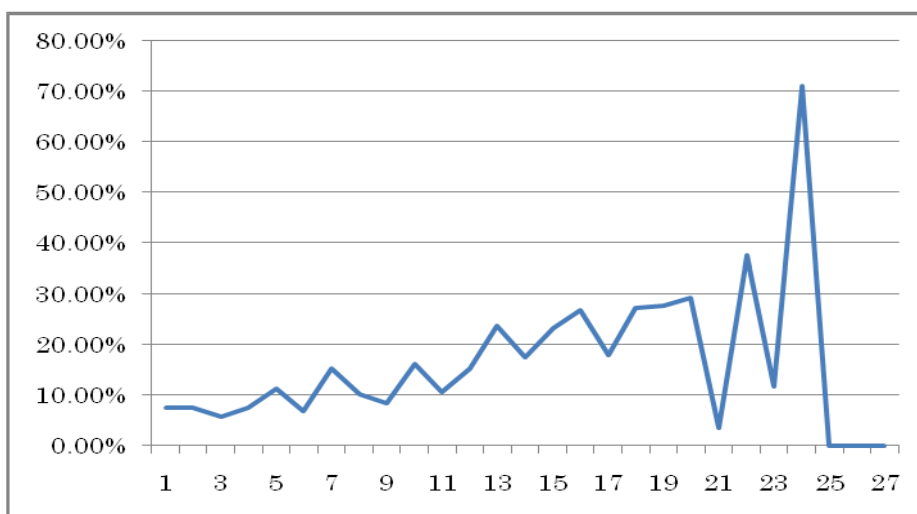


図3 固有名割合 (出現順位ランク別)

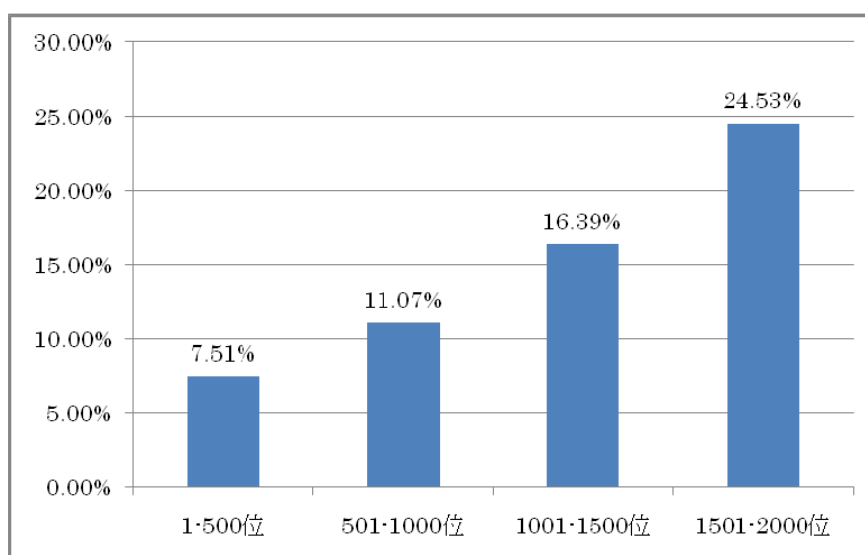


図4 固有名割合 (出現順位ランク別 (500 位ごと))

図4から、今回のBCCWJコアデータに基づく漢字表データで見ると、全体の傾向として上位2000位までは固有名での使用率が高まっていくことが分かる。ここから、「出現頻度が下位の漢字グループになるほど、固有名表記への傾斜度が高くなる」という大まかな実態把握ができると思われる。なお、上位2000位までの漢字の固有名割合の平均は、8.71%であった。おおよそ「10回使われたうちの1回弱が固有名表記」ということになる。

参考までに、出現順位のランク別(500位ごと)に各漢字政策カテゴリーの出現状況を見たのが、図5である。

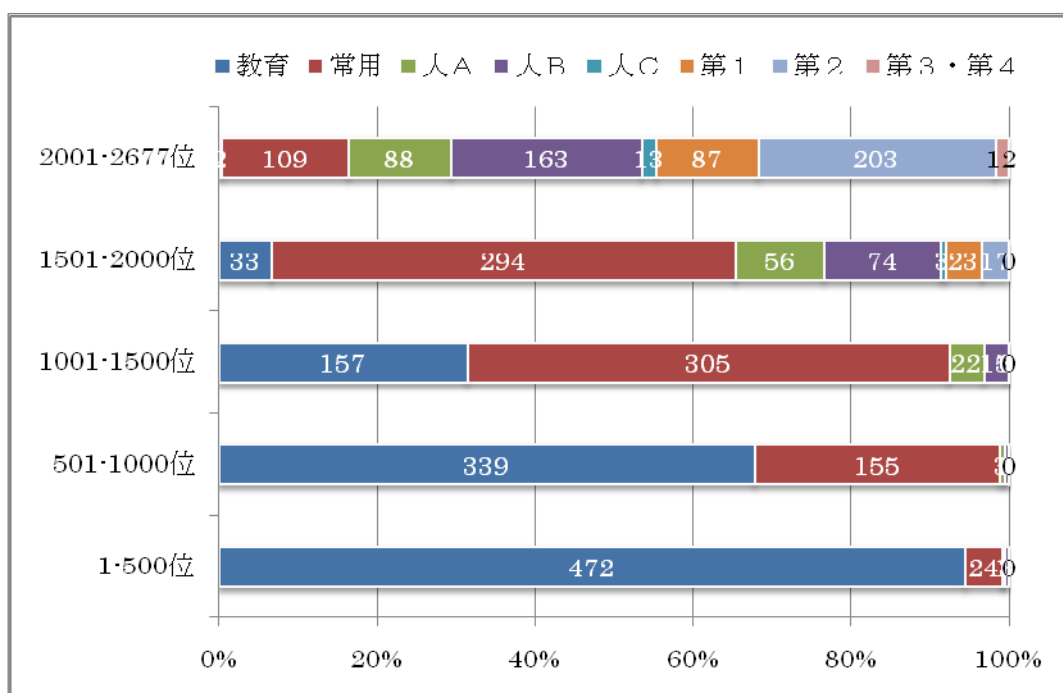


図5 漢字政策カテゴリーの割合（出現順位ランク別（500位ごと））

図5から、漢字の出現順位の観点から見ると、既存の漢字政策カテゴリーにおける字種選定がおおむね妥当であったことが分かる。頻度上位を中心に「教育用漢字」が根幹部分を形成し、それを補うように分厚くその他の「常用漢字（1981）」が分布し、さらにその周辺に各種の「表外漢字」が出現している。

ところで、「改定常用漢字表（2010）」は、基本的に出現頻度を重視しながら追加字種の選定を行っていることから、図5で上位に食い込んでいる表外漢字は、改めて検討の対象となり結果的に採用となった可能性が高い。これについては、第5節で取り上げる。

5. 固有名表記への傾斜度と漢字政策カテゴリー

個々の漢字を固有名表記への傾斜度の観点から見ると、例えば、「田」は93.74%、「中」は31.30%のように、ゼロパーセントから100パーセントまでの間の様々な段階に位置するものと予想される。今回の全2,677字がこのようなスケールのどの位置にどのくらい出現するのか、10パーセントごとに区切って実際の分布状況を見たのが、図6である。

まず、左右の両端を見ると、「ゼロ%」すなわち固有名表記に全く使われていない漢字が1,572語で全体の58.7%、逆に「100%」すなわち固有名表記にしか使われていない漢字が230語で全体の8.6%となっている。残りの875語が全体の32.7%を占め、固有名使用率は「1%台」から「90%台」までの間に散らばって分布する。図6からも分かるように、全体としては、真ん中よりもかなり左側に偏った分布状況になっている。

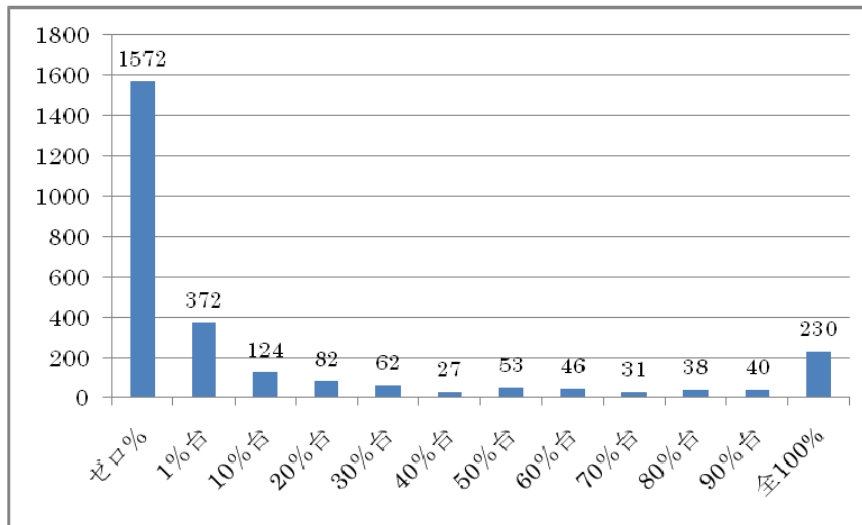


図6 固有名使用率ランキングから見た漢字の出現状況（全体）

図6（全体）から、言語政策カテゴリーのうち、「教育」「常用」（すなわち「常用漢字表（1981）」の表内漢字）を取り出して同様のグラフに示したのが図7、また、「人A」「人B」（すなわち「常用漢字表（1981）」の表外漢字の一部、詳細は第3節を参照）を取り出して示したのが図8である。

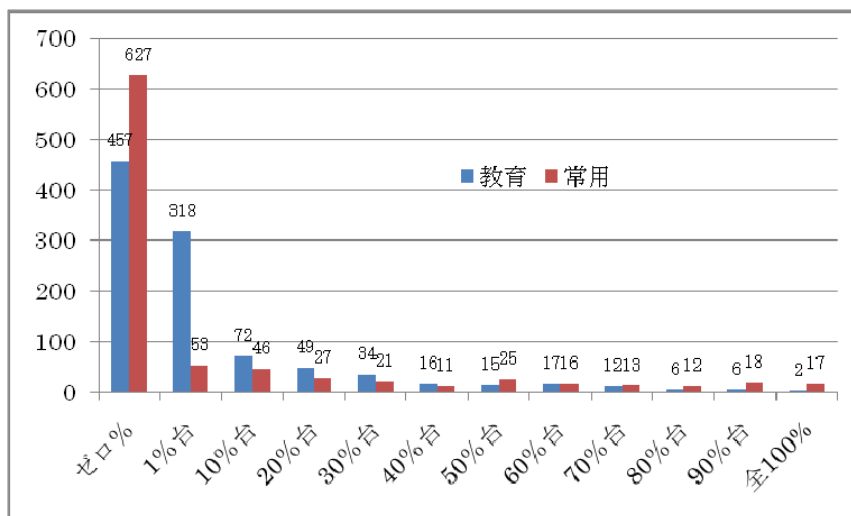


図7 固有名使用率から見た漢字の出現状況（教育・常用）

図7を見ると、「教育」「常用」は、左端に近い「ゼロ%」「1%台」あたりに集中して分布していることが目を引くが、そこから右側では少しずつ減少してそのまま「100%」に至る。漢字政策としての「教育」「常用」の「固有名詞を対象とするものではない」という基本方針の反映が見られるが、程度差はあれ一定数の「固有名表記にも使われる漢字」が含まれていることも、無視することのできない実際の使用実態と見るべきである。

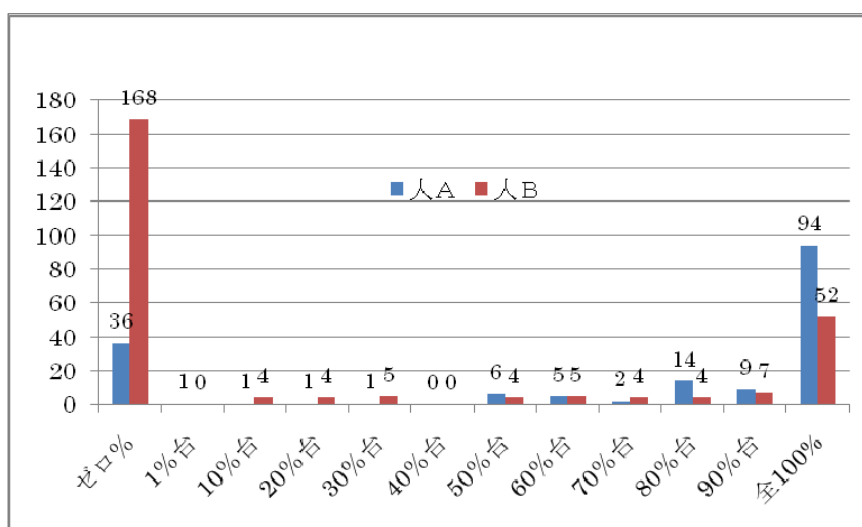


図8 固有名使用率から見た漢字の出現状況（人A・人B）

図8は、左端の「ゼロ%」と右端の「100%」のところに極端に両極に分かれて分布している点で、図7とは明らかに違った様相を見せている。漢字政策カテゴリーで「人名用漢字」に分類される漢字グループだけに、固有名表記については性格付けがはっきりしているものと思われる。詳しく見ると、「人A」が右端の「100%」に多いのに対して、「人B」はむしろ左端の「ゼロ%」の方に目立っている。「人名用漢字」とはいいながら、「人B」には実際には人名に使われていない（あるいは使われそうもない）漢字が数多く含まれていることを予想させる結果となっている。

第4節でも触れたように、「改定常用漢字表（2010）」は、基本的に出現頻度を重視しながら追加字種の選定を行っていることから、各種の漢字調査で頻度上位に登場する表外漢字の中には、結果的に採用となったものが多い。表2は、「改定常用漢字表（2010）」の追加字種のうち、今回の調査で総合頻度の多い方から上位15位までを、参考までに抽出したものである。

漢字政策カテゴリー別に見ると、「人A」から「藤、奈、鹿、亀」、人Bから「阪、岡、韓、誰、頃、曾、鍋、闇、塞」、第1から「狙」、第2から「拉」がリストに入っている。「人A」の漢字は、いずれも固有名割合がきわめて高いが、「藤、亀」が人名に、「奈、鹿」が地名に偏るといった首肯できる特徴も観察される。「人B」の漢字は、固有名割合の高い「阪、岡、韓、曾」と、反対に極端に低い「誰、頃、鍋、闇、塞」に二分される。「人B」の漢字は、使用頻度の高い点では共通するが、実際には性格の異なるものが混在しているという事実の一端が確認できる。残る二つの「狙」と「拉」は、固有名表記からは距離のある漢字と見られる。このように、「固有名表記」、さらに「人名表記」「地名表記」「それ以外」のように詳細に見ていくと、個々の漢字の特性が自ずからあぶり出されてくる。

表2 「改定常用漢字表（2010）」の追加字種（総合頻度上位15位）

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名割合	地名頻度	人名 頻度	固有名 一般頻度
419	人A	藤	172	6	166	96.51%	8	157	1
422	人B	阪	171	1	170	99.42%	162	4	4
472	人B	岡	152	1	151	99.34%	112	39	0
488	人B	韓	144	30	114	79.17%	113	1	0
721	人B	誰	79	79	0	0.00%	0	0	0
871	第1	狙	56	56	0	0.00%	0	0	0
982	人A	奈	43	0	43	100.00%	30	13	0
1113	人B	頃	32	32	0	0.00%	0	0	0
1149	人B	曾	29	1	28	96.55%	6	22	0
1199	第2	拉	26	26	0	0.00%	0	0	0
1217	人A	鹿	25	4	21	84.00%	20	1	0
1236	人B	鍋	24	21	3	12.50%	0	2	1
1267	人A	亀	22	1	21	95.45%	0	21	0
1291	人B	闇	21	21	0	0.00%	0	0	0
1291	人B	塞	21	21	0	0.00%	0	0	0

6. おわりに

本稿では、BCCWJ コアデータに基づく漢字表を事例として、個々の漢字が固有名表記に使われている頻度や割合を、基礎的な数値情報として漢字表に掲載することの有用性について述べた。これにより、既存の言語政策カテゴリーから離れて、全ての漢字を現実世界で実際に行われている固有名表記の観点から眺めることが可能となる。このことの意義は、思いのほか大きいものと思われる。

例えば、総合頻度が上位の基本的な教育用漢字の中にも、固有名表記に使われることにより総合頻度ポイントを稼いでいるものが少なくない。この事実を確認することは、初等教育段階における漢字教育・漢字指導に対して示唆的ではなかろうか。国語科では、基本的に漢字政策カテゴリーの優先順位に沿って一般語の漢字表記を教えるにしても、社会科、特に地理や歴史においては地名・人名など固有名の漢字表記を、それとして積極的に教えることが十分に考えられる。漢字教育の領域に固有名表記の観点を導入することは、検討に値する今後の重要課題と思われる。

漢字政策の面での新たな展開である「改定常用漢字表（2010）」は、その「基本的な性格」の一項で「固有名詞を対象とするものではない」としながらも、「ただし、固有名詞の中でも特に公共性の高い都道府県名に用いる漢字及びそれに準ずる漢字は例外として扱う」として、固有名詞の扱いに一定の柔軟な姿勢を見せた。これにより、「一般の社会生活における漢字使用の目安としての改定常用漢字表」の今後の検証には、固有名の漢字表記の実態

把握が不可欠という条件が加わったことになる。

テレビ放送における漢字表記の問題は、柴田実（2005）も指摘するように、報道メディアの性格上、画面で提供する重要情報の大部分が、数値と地名・人名などの固有名によって占められていることに起因する。一般の社会生活に浸透した、きわめて公共性の高いメディアとして、固有名に使う漢字をどのように考えればよいかは重大な問題であり、ここでも固有名の漢字表記の実態把握が強く求められている。

以上、広範な視野に立って今後の漢字政策を考える上で役に立つ漢字表は、固有名に使われる漢字の検討に十分に耐えうるものでなければならない。今回は BCCWJ コアデータにより試行的調査を行ったが、BCCWJ の本体について同様の調査を行い、固有名の漢字表記の実態把握を十全に行えるようにすることが、大きな課題として残されている。

本稿で導入した「固有名表記への傾斜度」という概念については、全ての漢字をこのような観点から眺めることにより、これまで見逃されてきた漢字一般の新たな側面の解明につながるのではないかと期待される。そのためには、例えば、土屋（1984）、同（1987）のような、固有名と一般語の区別に関する本質的な議論を踏まえることが不可欠であり、固有名研究それ自体のさらなる進展が俟たれるところである。

付記

BCCWJ コアデータから漢字表を作成する段階で、小椋秀樹氏（特定領域研究「日本語コーパス」データ班）の全面的な助力を得た。末尾ながら記して感謝申し上げる。

文献

- 相澤正夫，小椋秀樹，斎藤達哉（2008）「第6章 漢字表の設計と活用」田中ほか（2008）所収。
- 国立国語研究所（1976）『現代新聞の漢字』（国立国語研究所報告 56）。
- 国立国語研究所（2005）『『現代雑誌の語彙調査』に基づく漢字音訓一覧表』。
- 柴田実（2005）「第5章 放送と漢字」『朝倉漢字講座4 漢字と社会』朝倉書店。
- 田中牧郎，相澤正夫，斎藤達哉，棚橋尚子，小椋秀樹，鈴木一史，近藤明日子，平山允子，金愛蘭，桐生りか（2008）『言語政策に役立つ，コーパスを用いた語彙表・漢字表等の作成と活用』（特定領域研究「日本語コーパス」言語政策班中間報告書）。
- 土屋俊（1984）「固有名を他の一般名辞から区別するための条件について」千葉大学 人文研究，13，千葉大学人文学部，（土屋俊（2008）に再収）。
- 土屋俊（1987）「指示詞としての固有名詞」国文学 解釈と観賞，52-2，至文堂，（土屋俊（2008）に再収）。
- 土屋俊（2008）『土屋俊 言語・哲学コレクション 第1巻 真の包括的な言語の科学』，くろしお出版。

別表 BCCWJコアデータに基づく漢字表(総合頻度1位~499位)

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名 割合	地名 頻度	人名 頻度	固有名 一般 頻度
1	教育	日	2843	1967	876	30.81%	769	25	82
2	教育	人	2767	2747	20	0.72%	2	17	1
3	教育	年	2503	2502	1	0.04%	0	1	0
4	教育	一	2237	2059	178	7.96%	11	165	2
5	教育	大	2076	1688	388	18.69%	222	93	73
6	教育	国	2064	1546	518	25.10%	453	7	58
7	教育	会	1747	1745	2	0.11%	0	2	0
8	教育	中	1722	1183	539	31.30%	272	234	33
9	教育	者	1664	1663	1	0.06%	0	1	0
10	教育	本	1547	685	862	55.72%	768	93	1
11	教育	業	1422	1422	0	0.00%	0	0	0
12	教育	事	1380	1380	0	0.00%	0	0	0
13	教育	十	1367	1349	18	1.32%	11	6	1
14	教育	生	1346	1316	30	2.23%	7	21	2
15	教育	二	1280	1241	39	3.05%	1	36	2
16	教育	行	1279	1261	18	1.41%	0	18	0
17	教育	出	1245	1237	8	0.64%	1	5	2
18	教育	上	1225	1151	74	6.04%	34	40	0
19	教育	間	1149	1136	13	1.13%	3	10	0
20	教育	時	1138	1136	2	0.18%	0	0	2
21	教育	地	1136	1131	5	0.44%	3	2	0
22	教育	的	1133	1133	0	0.00%	0	0	0
23	教育	自	1109	1015	94	8.48%	3	0	91
24	教育	分	1107	1088	19	1.72%	17	2	0
25	教育	月	1106	1078	28	2.53%	10	15	3
26	教育	同	1105	1100	5	0.45%	1	0	4
27	教育	合	1091	1084	7	0.64%	0	7	0
28	教育	子	1082	753	329	30.41%	22	307	0
29	教育	学	1069	1033	36	3.37%	0	36	0
30	教育	見	1067	1030	37	3.47%	26	11	0
31	教育	長	1017	950	67	6.59%	40	22	5
32	教育	方	967	947	20	2.07%	9	11	0
33	教育	対	966	966	0	0.00%	0	0	0
34	教育	三	955	810	145	15.18%	33	75	37
35	教育	後	948	907	41	4.32%	20	21	0
36	教育	入	946	941	5	0.53%	3	2	0
37	教育	市	934	889	45	4.82%	18	27	0
38	教育	前	928	894	34	3.66%	6	28	0
39	教育	発	924	924	0	0.00%	0	0	0
40	教育	内	906	835	71	7.84%	22	49	0
41	教育	部	896	848	48	5.36%	22	26	0
42	教育	高	893	797	96	10.75%	28	64	4
43	教育	場	884	867	17	1.92%	7	10	0
44	教育	社	878	873	5	0.57%	0	1	4
45	教育	用	875	873	2	0.23%	2	0	0
46	教育	手	834	823	11	1.32%	10	1	0
47	教育	体	820	817	3	0.37%	0	0	3
48	教育	実	807	792	15	1.86%	0	10	5
49	教育	化	799	799	0	0.00%	0	0	0
50	教育	関	792	739	53	6.69%	40	12	1
51	教育	力	770	749	21	2.73%	0	21	0
52	教育	動	755	752	3	0.40%	0	3	0
53	教育	金	749	667	82	10.95%	6	76	0
54	教育	理	741	732	9	1.21%	0	9	0
55	教育	作	735	723	12	1.63%	3	8	1

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名 割合	地名 頻度	人名 頻度	固有名 一般 頻度
56	教育	開	723	723	0	0.00%	0	0	0
57	教育	成	719	520	199	27.68%	7	23	169
58	教育	性	718	718	0	0.00%	0	0	0
59	教育	定	717	715	2	0.28%	0	2	0
60	教育	目	709	700	9	1.27%	7	2	0
61	教育	度	707	707	0	0.00%	0	0	0
62	教育	新	696	612	84	12.07%	52	18	14
63	教育	代	685	653	32	4.67%	18	14	0
64	教育	政	682	672	10	1.47%	0	5	5
65	教育	今	677	665	12	1.77%	1	11	0
66	教育	員	671	671	0	0.00%	0	0	0
66	教育	気	671	671	0	0.00%	0	0	0
66	教育	外	671	669	2	0.30%	0	2	0
69	教育	立	665	654	11	1.65%	9	2	0
70	教育	通	663	653	10	1.51%	6	3	1
71	教育	民	657	553	104	15.83%	0	5	99
72	教育	全	649	634	15	2.31%	3	4	8
73	教育	取	647	632	15	2.32%	15	0	0
74	教育	回	635	635	0	0.00%	0	0	0
75	教育	要	632	630	2	0.32%	0	2	0
76	教育	機	629	628	1	0.16%	0	1	0
77	教育	明	623	540	83	13.32%	14	37	32
78	教育	五	622	605	17	2.73%	3	13	1
79	教育	保	621	569	52	8.37%	5	42	5
80	教育	家	620	605	15	2.42%	0	14	1
81	教育	当	619	619	0	0.00%	0	0	0
82	教育	円	617	611	6	0.97%	1	5	0
83	教育	言	616	616	0	0.00%	0	0	0
84	教育	道	614	504	110	17.92%	74	31	5
85	教育	現	610	610	0	0.00%	0	0	0
86	教育	四	593	572	21	3.54%	8	12	1
87	教育	表	592	591	1	0.17%	1	0	0
87	教育	所	592	571	21	3.55%	5	16	0
89	教育	東	591	81	510	86.29%	451	20	39
90	教育	主	589	583	6	1.02%	0	6	0
91	教育	調	583	579	4	0.69%	4	0	0
92	教育	相	579	567	12	2.07%	6	6	0
93	教育	小	577	382	195	33.80%	30	165	0
94	教育	報	571	571	0	0.00%	0	0	0
95	教育	議	570	570	0	0.00%	0	0	0
95	教育	経	570	556	14	2.46%	1	2	11
97	教育	問	565	565	0	0.00%	0	0	0
98	教育	法	564	558	6	1.06%	0	5	1
99	教育	田	559	35	524	93.74%	121	396	7
100	教育	物	553	551	2	0.36%	0	2	0
101	教育	約	550	550	0	0.00%	0	0	0
102	教育	思	546	546	0	0.00%	0	0	0
103	教育	下	543	481	62	11.42%	18	35	9
104	教育	意	542	542	0	0.00%	0	0	0
105	教育	米	541	387	154	28.47%	149	5	0
106	教育	戦	538	538	0	0.00%	0	0	0
107	教育	公	531	525	6	1.13%	2	3	1
107	教育	山	531	160	371	69.87%	121	249	1
109	教育	九	530	479	51	9.62%	47	1	3
110	教育	数	523	521	2	0.38%	1	1	0
111	教育	進	519	507	12	2.31%	2	9	1
112	教育	選	517	517	0	0.00%	0	0	0

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名 割合	地名 頻度	人名 頻度	固有名 一般 頻度
113	教育	女	511	498	13	2.54%	0	13	0
114	教育	品	509	508	1	0.20%	1	0	0
114	教育	万	509	505	4	0.79%	0	3	1
116	教育	害	507	507	0	0.00%	0	0	0
117	教育	最	503	501	2	0.40%	0	2	0
118	教育	設	501	500	1	0.20%	0	1	0
119	教育	持	498	498	0	0.00%	0	0	0
120	教育	多	492	469	23	4.67%	20	3	0
121	教育	話	487	487	0	0.00%	0	0	0
122	教育	組	481	477	4	0.83%	0	4	0
122	教育	加	481	444	37	7.69%	7	30	0
124	教育	教	480	476	4	0.83%	0	2	2
125	教育	受	478	478	0	0.00%	0	0	0
125	教育	安	478	422	56	11.72%	30	19	7
127	教育	不	475	471	4	0.84%	0	4	0
128	教育	活	471	471	0	0.00%	0	0	0
129	教育	期	470	470	0	0.00%	0	0	0
130	教育	京	467	18	449	96.15%	436	7	6
131	教育	連	462	384	78	16.88%	20	0	58
132	教育	書	461	461	0	0.00%	0	0	0
133	教育	心	460	455	5	1.09%	0	5	0
134	教育	以	456	455	1	0.22%	1	0	0
135	教育	産	455	444	11	2.42%	0	0	11
136	教育	利	454	432	22	4.85%	2	14	6
137	教育	県	453	451	2	0.44%	0	2	0
138	教育	支	452	452	0	0.00%	0	0	0
138	教育	知	452	364	88	19.47%	76	12	0
140	教育	題	451	451	0	0.00%	0	0	0
141	教育	務	450	444	6	1.33%	0	6	0
142	教育	平	449	199	250	55.68%	30	48	172
143	教育	名	441	326	115	26.08%	103	2	10
144	教育	続	435	435	0	0.00%	0	0	0
145	教育	決	434	434	0	0.00%	0	0	0
146	教育	等	433	433	0	0.00%	0	0	0
146	教育	指	433	431	2	0.46%	2	0	0
146	教育	来	433	427	6	1.39%	0	5	1
149	教育	査	429	429	0	0.00%	0	0	0
149	教育	野	429	222	207	48.25%	68	136	3
151	教育	情	426	426	0	0.00%	0	0	0
151	教育	六	426	423	3	0.70%	1	2	0
153	教育	売	424	424	0	0.00%	0	0	0
154	教育	点	421	421	0	0.00%	0	0	0
155	教育	強	420	419	1	0.24%	0	1	0
156	教育	向	419	416	3	0.72%	2	1	0
157	教育	面	418	418	0	0.00%	0	0	0
158	教育	百	414	410	4	0.97%	1	3	0
159	教育	北	413	118	295	71.43%	271	14	10
160	教育	制	411	411	0	0.00%	0	0	0
161	教育	料	402	402	0	0.00%	0	0	0
161	教育	感	402	399	3	0.75%	0	0	3
163	教育	世	401	381	20	4.99%	12	8	0
164	教育	計	399	398	1	0.25%	0	1	0
165	教育	区	398	398	0	0.00%	0	0	0
166	教育	水	397	355	42	10.58%	25	17	0
167	教育	能	396	384	12	3.03%	8	4	0
167	教育	都	396	334	62	15.66%	56	6	0
169	教育	資	395	395	0	0.00%	0	0	0

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名 割合	地名 頻度	人名 頻度	固有名 一般 頻度
170	教育	党	394	394	0	0.00%	0	0	0
171	教育	件	391	391	0	0.00%	0	0	0
172	教育	交	389	389	0	0.00%	0	0	0
172	教育	八	389	360	29	7.46%	17	10	2
174	常用	施	385	383	2	0.52%	0	2	0
175	教育	原	379	223	156	41.16%	37	117	2
176	教育	策	376	374	2	0.53%	0	2	0
176	教育	総	376	362	14	3.72%	0	14	0
178	教育	正	375	285	90	24.00%	0	72	18
179	教育	文	374	348	26	6.95%	4	18	4
180	教育	初	373	371	2	0.54%	1	1	0
181	教育	重	372	352	20	5.38%	15	5	0
182	教育	先	371	371	0	0.00%	0	0	0
182	教育	画	371	371	0	0.00%	0	0	0
184	教育	使	369	369	0	0.00%	0	0	0
185	教育	記	367	361	6	1.63%	0	6	0
186	教育	営	364	364	0	0.00%	0	0	0
187	教育	考	359	359	0	0.00%	0	0	0
188	教育	第	358	358	0	0.00%	0	0	0
189	教育	少	354	354	0	0.00%	0	0	0
189	教育	院	354	351	3	0.85%	3	0	0
191	教育	団	353	348	5	1.42%	0	1	4
192	教育	在	352	349	3	0.85%	0	3	0
193	教育	川	346	50	296	85.55%	129	160	7
194	教育	近	345	291	54	15.65%	3	34	17
195	教育	権	341	341	0	0.00%	0	0	0
195	教育	店	341	337	4	1.17%	4	0	0
195	教育	信	341	320	21	6.16%	4	15	2
198	教育	際	340	339	1	0.29%	0	1	0
199	常用	込	339	339	0	0.00%	0	0	0
199	教育	備	339	338	1	0.29%	1	0	0
199	教育	協	339	334	5	1.47%	0	4	1
202	教育	解	338	338	0	0.00%	0	0	0
202	教育	首	338	338	0	0.00%	0	0	0
202	教育	町	338	244	94	27.81%	81	13	0
205	教育	治	337	278	59	17.51%	3	31	25
206	教育	千	336	269	67	19.94%	47	20	0
207	教育	共	332	325	7	2.11%	0	1	6
208	教育	容	331	329	2	0.60%	0	2	0
208	教育	無	331	328	3	0.91%	2	1	0
208	教育	七	331	321	10	3.02%	2	8	0
211	教育	校	330	330	0	0.00%	0	0	0
211	教育	男	330	294	36	10.91%	0	36	0
213	教育	集	329	329	0	0.00%	0	0	0
213	教育	付	329	329	0	0.00%	0	0	0
215	教育	特	327	327	0	0.00%	0	0	0
216	教育	整	325	325	0	0.00%	0	0	0
217	教育	食	324	320	4	1.23%	0	0	4
217	教育	基	324	316	8	2.47%	0	8	0
219	教育	必	323	323	0	0.00%	0	0	0
219	教育	結	323	321	2	0.62%	0	2	0
221	教育	私	321	321	0	0.00%	0	0	0
221	教育	木	321	144	177	55.14%	23	153	1
223	教育	育	319	319	0	0.00%	0	0	0
223	教育	置	319	319	0	0.00%	0	0	0
223	教育	府	319	303	16	5.02%	15	1	0
226	教育	麥	317	317	0	0.00%	0	0	0

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名 割合	地名 頻度	人名 頻度	固有名 一般 頻度
227	教育	参	314	314	0	0.00%	0	0	0
228	教育	氏	313	311	2	0.64%	0	2	0
229	教育	係	312	312	0	0.00%	0	0	0
229	教育	果	312	310	2	0.64%	0	1	1
231	教育	神	310	202	108	34.84%	89	19	0
232	教育	改	309	309	0	0.00%	0	0	0
233	教育	元	308	284	24	7.79%	0	19	5
234	教育	再	306	306	0	0.00%	0	0	0
234	教育	増	306	300	6	1.96%	0	6	0
236	教育	認	305	305	0	0.00%	0	0	0
236	教育	職	305	305	0	0.00%	0	0	0
238	教育	別	304	289	15	4.93%	12	3	0
239	教育	域	303	303	0	0.00%	0	0	0
240	教育	切	302	301	1	0.33%	1	0	0
240	教育	電	302	295	7	2.32%	0	0	7
240	教育	海	302	211	91	30.13%	82	9	0
243	教育	凵	298	298	0	0.00%	0	0	0
243	教育	濟	298	297	1	0.34%	0	0	1
243	教育	省	298	296	2	0.67%	0	2	0
246	教育	和	297	174	123	41.41%	30	33	60
247	教育	勝	296	250	46	15.54%	15	30	1
248	教育	統	295	295	0	0.00%	0	0	0
248	教育	研	295	290	5	1.69%	0	0	5
250	教育	身	294	292	2	0.68%	0	2	0
251	教育	農	293	291	2	0.68%	0	0	2
251	教育	住	293	275	18	6.14%	5	2	11
253	常用	企	292	292	0	0.00%	0	0	0
254	教育	疑	291	291	0	0.00%	0	0	0
255	教育	流	290	288	2	0.69%	1	1	0
255	教育	界	290	283	7	2.41%	7	0	0
257	教育	確	289	289	0	0.00%	0	0	0
258	常用	援	288	288	0	0.00%	0	0	0
259	教育	運	287	287	0	0.00%	0	0	0
260	教育	告	285	285	0	0.00%	0	0	0
261	教育	判	284	284	0	0.00%	0	0	0
261	教育	得	284	282	2	0.70%	2	0	0
261	教育	広	284	226	58	20.42%	41	17	0
264	教育	防	283	283	0	0.00%	0	0	0
265	教育	車	282	282	0	0.00%	0	0	0
265	教育	島	282	132	150	53.19%	87	59	4
267	教育	引	281	281	0	0.00%	0	0	0
268	教育	村	280	146	134	47.86%	12	121	1
269	教育	語	279	279	0	0.00%	0	0	0
270	教育	減	278	278	0	0.00%	0	0	0
271	教育	費	277	277	0	0.00%	0	0	0
271	教育	状	277	277	0	0.00%	0	0	0
271	教育	半	277	274	3	1.08%	2	0	1
274	教育	有	276	251	25	9.06%	13	12	0
275	教育	局	275	275	0	0.00%	0	0	0
276	教育	示	273	273	0	0.00%	0	0	0
277	教育	予	271	271	0	0.00%	0	0	0
277	教育	次	271	241	30	11.07%	0	30	0
277	教育	真	271	216	55	20.30%	9	46	0
280	教育	屋	270	165	105	38.89%	101	4	0
281	教育	究	266	265	1	0.38%	0	1	0
281	教育	応	266	260	6	2.26%	0	0	6
283	教育	価	263	263	0	0.00%	0	0	0

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名 割合	地名 頻度	人名 頻度	固有名 一般 頻度
283	教育	求	263	263	0	0.00%	0	0	0
285	教育	術	262	262	0	0.00%	0	0	0
286	教育	空	261	247	14	5.36%	5	7	2
287	教育	口	260	204	56	21.54%	9	47	0
288	教育	間	259	259	0	0.00%	0	0	0
289	教育	台	257	175	82	31.91%	81	1	0
290	教育	何	253	253	0	0.00%	0	0	0
290	教育	工	253	242	11	4.35%	0	8	3
292	教育	打	252	252	0	0.00%	0	0	0
292	教育	割	252	252	0	0.00%	0	0	0
292	教育	演	252	247	5	1.98%	0	0	5
295	教育	率	245	245	0	0.00%	0	0	0
295	教育	建	245	241	4	1.63%	0	4	0
297	教育	投	244	244	0	0.00%	0	0	0
297	教育	死	244	244	0	0.00%	0	0	0
299	教育	検	243	243	0	0.00%	0	0	0
299	教育	軍	243	243	0	0.00%	0	0	0
301	教育	任	241	240	1	0.41%	0	1	0
302	教育	争	240	240	0	0.00%	0	0	0
302	教育	朝	240	97	143	59.58%	134	8	1
304	教育	材	239	239	0	0.00%	0	0	0
305	教育	反	238	238	0	0.00%	0	0	0
306	教育	線	237	234	3	1.27%	0	0	3
307	常用	歳	234	230	4	1.71%	4	0	0
308	教育	収	233	233	0	0.00%	0	0	0
308	教育	西	233	74	159	68.24%	90	50	19
310	教育	始	232	232	0	0.00%	0	0	0
310	教育	各	232	232	0	0.00%	0	0	0
312	教育	働	231	231	0	0.00%	0	0	0
313	教育	可	229	227	2	0.87%	0	2	0
314	教育	着	226	226	0	0.00%	0	0	0
314	教育	論	226	226	0	0.00%	0	0	0
314	教育	役	226	224	2	0.88%	0	2	0
314	教育	味	226	222	4	1.77%	1	3	0
314	教育	比	226	219	7	3.10%	6	1	0
319	教育	楽	223	211	12	5.38%	5	7	0
319	教育	直	223	202	21	9.42%	1	20	0
321	常用	及	222	222	0	0.00%	0	0	0
322	教育	提	221	221	0	0.00%	0	0	0
323	教育	供	219	218	1	0.46%	1	0	0
324	教育	構	218	218	0	0.00%	0	0	0
325	教育	境	216	216	0	0.00%	0	0	0
325	教育	止	216	216	0	0.00%	0	0	0
327	教育	格	215	214	1	0.47%	0	1	0
327	教育	由	215	188	27	12.56%	8	19	0
329	教育	展	214	213	1	0.47%	0	0	1
330	教育	病	213	213	0	0.00%	0	0	0
330	常用	彼	213	213	0	0.00%	0	0	0
330	教育	過	213	213	0	0.00%	0	0	0
330	教育	象	213	213	0	0.00%	0	0	0
330	教育	石	213	110	103	48.36%	25	76	2
335	教育	億	212	212	0	0.00%	0	0	0
335	教育	談	212	206	6	2.83%	6	0	0
335	教育	造	212	204	8	3.77%	0	8	0
335	教育	美	212	121	91	42.92%	11	77	3
339	教育	障	211	211	0	0.00%	0	0	0
339	教育	起	211	210	1	0.47%	0	1	0

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名 割合	地名 頻度	人名 頻度	固有名 一般 頻度
341	教育	推	210	210	0	0.00%	0	0	0
342	常用	違	209	209	0	0.00%	0	0	0
343	教育	領	208	208	0	0.00%	0	0	0
343	教育	医	208	208	0	0.00%	0	0	0
343	教育	張	208	150	58	27.88%	6	52	0
346	教育	説	207	207	0	0.00%	0	0	0
346	教育	勞	207	207	0	0.00%	0	0	0
346	教育	導	207	206	1	0.48%	0	0	1
346	教育	額	207	205	2	0.97%	2	0	0
346	教育	良	207	137	70	33.82%	22	48	0
351	教育	消	206	206	0	0.00%	0	0	0
351	教育	配	206	206	0	0.00%	0	0	0
351	常用	被	206	206	0	0.00%	0	0	0
354	常用	療	204	204	0	0.00%	0	0	0
354	教育	担	204	204	0	0.00%	0	0	0
354	教育	技	204	204	0	0.00%	0	0	0
357	教育	殺	203	203	0	0.00%	0	0	0
357	教育	科	203	201	2	0.99%	2	0	0
359	教育	残	202	202	0	0.00%	0	0	0
359	教育	形	202	191	11	5.45%	5	6	0
361	常用	環	201	201	0	0.00%	0	0	0
361	教育	様	201	201	0	0.00%	0	0	0
361	教育	古	201	92	109	54.23%	99	10	0
364	教育	案	200	200	0	0.00%	0	0	0
364	教育	親	200	199	1	0.50%	0	1	0
364	教育	種	200	190	10	5.00%	9	1	0
367	教育	然	199	197	2	1.01%	0	2	0
368	教育	伝	198	198	0	0.00%	0	0	0
369	教育	乘	197	197	0	0.00%	0	0	0
369	教育	昨	197	197	0	0.00%	0	0	0
369	教育	常	197	190	7	3.55%	7	0	0
372	教育	量	196	195	1	0.51%	0	1	0
372	教育	位	196	191	5	2.55%	0	5	0
374	教育	器	195	194	1	0.51%	1	0	0
375	教育	断	194	194	0	0.00%	0	0	0
375	教育	税	194	194	0	0.00%	0	0	0
377	教育	負	193	193	0	0.00%	0	0	0
377	教育	松	193	23	170	88.08%	23	131	16
379	教育	護	192	185	7	3.65%	2	5	0
379	教育	土	192	172	20	10.42%	3	17	0
379	常用	介	192	140	52	27.08%	0	52	0
382	教育	宅	191	185	6	3.14%	2	4	0
382	常用	井	191	13	178	93.19%	37	136	5
384	教育	側	190	190	0	0.00%	0	0	0
384	教育	福	190	76	114	60.00%	94	15	5
386	教育	終	189	189	0	0.00%	0	0	0
386	教育	仕	189	188	1	0.53%	0	1	0
386	教育	風	189	180	9	4.76%	3	6	0
386	教育	義	189	144	45	23.81%	2	43	0
386	教育	林	189	142	47	24.87%	1	40	6
391	教育	午	188	188	0	0.00%	0	0	0
391	教育	委	188	188	0	0.00%	0	0	0
391	教育	個	188	188	0	0.00%	0	0	0
391	教育	規	188	185	3	1.60%	0	3	0
395	教育	館	187	163	24	12.83%	19	5	0
396	教育	送	184	184	0	0.00%	0	0	0
397	教育	官	183	183	0	0.00%	0	0	0

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名 割合	地名 頻度	人名 頻度	固有名 一般 頻度
397	教育	管	183	183	0	0.00%	0	0	0
399	教育	観	182	179	3	1.65%	0	3	0
400	教育	両	181	181	0	0.00%	0	0	0
401	教育	限	180	180	0	0.00%	0	0	0
401	教育	足	180	172	8	4.44%	5	3	0
403	教育	放	179	174	5	2.79%	0	1	4
403	教育	夫	179	143	36	20.11%	0	36	0
403	教育	愛	179	115	64	35.75%	60	4	0
406	教育	式	178	178	0	0.00%	0	0	0
406	教育	製	178	178	0	0.00%	0	0	0
408	教育	算	177	177	0	0.00%	0	0	0
408	教育	災	177	177	0	0.00%	0	0	0
408	常用	影	177	175	2	1.13%	1	1	0
411	教育	失	176	176	0	0.00%	0	0	0
412	教育	注	175	175	0	0.00%	0	0	0
412	教育	転	175	175	0	0.00%	0	0	0
412	教育	質	175	175	0	0.00%	0	0	0
412	教育	視	175	175	0	0.00%	0	0	0
412	教育	積	175	172	3	1.71%	0	3	0
417	教育	路	174	145	29	16.67%	23	6	0
418	教育	例	173	173	0	0.00%	0	0	0
419	教育	商	172	172	0	0.00%	0	0	0
419	教育	頭	172	170	2	1.16%	2	0	0
419	人A	藤	172	6	166	96.51%	8	157	1
422	教育	段	171	171	0	0.00%	0	0	0
422	教育	効	171	170	1	0.58%	0	1	0
422	教育	優	171	170	1	0.58%	0	1	0
422	教育	球	171	170	1	0.58%	1	0	0
422	人B	阪	171	1	170	99.42%	162	4	4
427	教育	客	170	170	0	0.00%	0	0	0
428	教育	葉	169	138	31	18.34%	27	4	0
428	教育	白	169	127	42	24.85%	18	18	6
430	教育	返	167	167	0	0.00%	0	0	0
430	教育	待	167	167	0	0.00%	0	0	0
430	教育	健	167	139	28	16.77%	0	28	0
433	教育	買	166	166	0	0.00%	0	0	0
433	教育	拳	166	166	0	0.00%	0	0	0
433	教育	落	166	164	2	1.20%	0	2	0
433	教育	早	166	127	39	23.49%	7	22	10
437	教育	派	164	164	0	0.00%	0	0	0
437	教育	識	164	164	0	0.00%	0	0	0
437	教育	南	164	110	54	32.93%	42	12	0
440	教育	険	162	162	0	0.00%	0	0	0
440	教育	想	162	162	0	0.00%	0	0	0
440	常用	監	162	162	0	0.00%	0	0	0
440	教育	勢	162	152	10	6.17%	7	2	1
440	教育	門	162	140	22	13.58%	9	13	0
445	教育	助	161	150	11	6.83%	0	11	0
446	教育	写	160	160	0	0.00%	0	0	0
446	教育	若	160	149	11	6.88%	3	8	0
448	教育	型	159	159	0	0.00%	0	0	0
448	教育	証	159	149	10	6.29%	0	0	10
450	教育	好	158	154	4	2.53%	1	3	0
450	教育	士	158	143	15	9.49%	9	5	1
450	常用	郎	158	0	158	100.00%	0	158	0
453	教育	態	157	157	0	0.00%	0	0	0
453	教育	裁	157	157	0	0.00%	0	0	0

順位	種類	字種	総合頻度	一般語 頻度合計	固有名 頻度合計	固有名 割合	地名 頻度	人名 頻度	固有名 一般 頻度
453	常用	与	157	154	3	1.91%	1	2	0
456	教育	命	156	156	0	0.00%	0	0	0
456	教育	念	156	156	0	0.00%	0	0	0
456	教育	字	156	156	0	0.00%	0	0	0
456	常用	販	156	156	0	0.00%	0	0	0
456	教育	警	156	156	0	0.00%	0	0	0
456	教育	難	156	156	0	0.00%	0	0	0
456	教育	訪	156	155	1	0.64%	1	0	0
456	教育	室	156	139	17	10.90%	11	6	0
456	常用	鮮	156	24	132	84.62%	132	0	0
465	教育	移	155	154	1	0.65%	0	1	0
465	教育	達	155	144	11	7.10%	1	10	0
465	教育	森	155	96	59	38.06%	9	50	0
465	教育	州	155	67	88	56.77%	88	0	0
469	常用	債	154	154	0	0.00%	0	0	0
469	教育	番	154	149	5	3.25%	3	2	0
471	教育	処	153	153	0	0.00%	0	0	0
472	教育	準	152	152	0	0.00%	0	0	0
472	常用	況	152	152	0	0.00%	0	0	0
472	人B	岡	152	1	151	99.34%	112	39	0
475	教育	復	151	151	0	0.00%	0	0	0
476	教育	末	150	149	1	0.67%	0	1	0
476	教育	天	150	119	31	20.67%	22	5	4
478	教育	給	149	149	0	0.00%	0	0	0
479	教育	験	148	148	0	0.00%	0	0	0
479	教育	悪	148	148	0	0.00%	0	0	0
479	教育	銀	148	115	33	22.30%	15	2	16
482	教育	素	147	147	0	0.00%	0	0	0
482	教育	試	147	147	0	0.00%	0	0	0
482	教育	声	147	147	0	0.00%	0	0	0
482	教育	音	147	139	8	5.44%	5	2	1
486	教育	評	145	145	0	0.00%	0	0	0
486	教育	戸	145	69	76	52.41%	67	9	0
488	常用	訴	144	144	0	0.00%	0	0	0
488	教育	母	144	143	1	0.69%	1	0	0
488	人B	韓	144	30	114	79.17%	113	1	0
491	常用	含	143	143	0	0.00%	0	0	0
491	教育	姿	143	143	0	0.00%	0	0	0
491	教育	景	143	132	11	7.69%	0	11	0
494	教育	追	142	142	0	0.00%	0	0	0
494	教育	幹	142	136	6	4.23%	0	6	0
496	教育	習	141	141	0	0.00%	0	0	0
496	教育	族	141	141	0	0.00%	0	0	0
496	教育	接	141	141	0	0.00%	0	0	0
499	常用	離	140	140	0	0.00%	0	0	0
499	教育	低	140	140	0	0.00%	0	0	0
499	教育	階	140	137	3	2.14%	1	2	0
499	常用	崎	140	2	138	98.57%	52	86	0

第2章 第4節

BCCWJによる「NDC ジャンル別漢字出現頻度表」の分析

斎藤 達哉

1. はじめに

本稿は、『現代日本語書き言葉均衡コーパス』(BCCWJ)を資料として、各ジャンルの特徴漢字の抽出を試みるものである。

また、「改定常用漢字表」(2010年11月30日に内閣告示「常用漢字表」となったもの)で新たに追加された漢字がどのジャンルに頻出するものかについても調査し、「常用」の範囲の広がった方向についても考えたい。

2. 使用コーパス

本調査は、BCCWJ領域内公開データ(2009年度版)に収められたデータのうち、生産実態サブコーパス書籍(PB)の固定長サンプル(PB_FL)を使用して行った。以下で、「書籍サンプル」あるいは「PB」と称するのは、このPB固定長サンプルのことである。

PBのサンプルファイルの使用に当たっては、XMLファイルを使用し、各ファイルの<sampling type="start"/>タグと<sampling type="end"/>タグによって囲まれた1000文字分を対象とした。

なお、各サンプルファイルから1000文字を抽出する作業は、言語政策班・近藤明日子氏(国立国語研究所)作成のPerlスクリプトの提供を受けたことにより効率的に行うことができた。

3. 日本十進分類法(NDC)の「類」による書籍ジャンル分け

本稿では、図書館等で図書の分類に使用されている「日本十進分類法」(NDC)の第1次区分である「類」を基にした11分類を「ジャンル」と称することにする。

BCCWJでは、XMLファイルのファイル名の左から4桁目に、国立国会図書館で付与されたNDCの第1次区分(類)を「0~9, n」として表示する決まりとなっている。これを丸山(2009)によって示すと以下のとおりである。

「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	「n」=分類なし	

なお、11分類中の「n」(分類なし)は、2005年10月時点で国立国会図書館によってNDCが付与されていなかった場合であり、実際には0~9のいずれかに該当する内容の書籍が混在していると考えられるべきものである。

4. 書籍(PB固定長サンプル)の漢字

今回対象とした書籍サンプルに含まれている漢字の字種数及び字数は、表1のとおりである。BCCWJのサンプリングの方針上、各ジャンルのサンプル数は均一ではなく、それに伴って調査対象とした字数(句読点、記号も含む)も均一ではない。

表1 PB_FLにおける漢字の字種数・字数・含有率

	サンプル数	調査対象字数	字種数	漢字字数	漢字含有率
0 総記	251	251,000	2,242	74,530	29.7%
1 哲学	536	536,000	2,889	171,681	32.0%
2 歴史	682	682,000	3,479	269,514	39.5%
3 社会科学	2,267	2,267,000	3,446	898,053	39.6%
4 自然科学	615	615,000	2,523	217,397	35.3%
5 技術・工学	618	618,000	2,596	202,657	32.8%
6 産業	334	334,000	2,353	119,593	35.8%
7 芸術・美術	524	524,000	2,875	158,780	30.3%
8 言語	153	153,000	2,252	50,833	33.2%
9 文学	2,243	2,243,000	4,000	648,111	28.9%
n 分類なし	262	262,000	2,084	67,420	25.7%
PB_FL 合計	8,485	8,485,000	4,815	2,878,569	33.9%

漢字の「字種数」は、PB全体では4,815字種であり、ジャンル別に見ると「9 文学」が4,000字種と最も多くの漢字を含有していた。また、いずれのジャンルでも2,000字種を下回ることはない。

「漢字含有率」は、漢字字数を調査対象字数で割ったもので、PB全体で見ると33.9%であった。最も高いのは「3 社会科学」の39.6%、最も低いのは「9 文学」の28.9%（「n 分類なし」を含めると25.7%）であり、ジャンルによっては10%程度の差が見られる。

5. 「NDCジャンル別漢字出現頻度表」の作成

PBの各ジャンルの漢字について、出現頻度の高いものから順に配列したものが、付録CD-ROMに収めた「NDCジャンル別漢字出現頻度表」（以下「漢字表」と称する）である。

漢字表は、ジャンル別の頻度表である「PB_0_総記.xls」「PB_1_哲学.xls」「PB_2_歴史.xls」「PB_3_社会科学.xls」「PB_4_自然科学.xls」「PB_5_技術・工学.xls」「PB_6_産業.xls」「PB_7_芸術・美術.xls」「PB_8_言語.xls」「PB_9_文学.xls」「PB_n_分類なし.xls」、各ジャンルを合計した総合頻度表である「PB 総計（度数）.xls」、及び漢字の出現するサンプル数の一覧表である「PB 総計（サンプル数）.xls」の3種13ファイルから成っている。

5.1 ジャンル別の頻度表の項目について

ジャンル別頻度表である「PB_0_総記.xls」～「PB_9_文学.xls」及び「PB_n_分類なし.xls」では、それぞれ次の項目を表示している。

[順位]	……	ジャンル内での度数による出現順位
[漢字]	……	見出しとなる漢字字種
[度数]	……	ジャンル内での度数
[度数%]	……	[度数]を全漢字数で割算したものを百分率で表示
[累積度数%]	……	[度数%]を出現順位順に加算
[殊別]	……	「改定常用漢字表」との関連を表示（7.2で述べる）
[配当学年]	……	小学校での配当学年
[総合順位]	……	全ジャンルを合計したPB全体での出現順位
[順位差]	……	[総合順位]から[順位]を減算して算出

- [順位差ランク] …… [順位差] について A～K の 11 段階にランク分け
- A 総合順位に比べジャンルでの順位が +1001 以上
 - B 総合順位に比べジャンルでの順位が +1000～+751
 - C 総合順位に比べジャンルでの順位が +750～+501
 - D 総合順位に比べジャンルでの順位が +500～+251
 - E 総合順位に比べジャンルでの順位が +250～+1
 - F 総合順位に比べジャンルでの順位が ±0
 - G 総合順位に比べジャンルでの順位が -1～-250
 - H 総合順位に比べジャンルでの順位が -251～-500
 - I 総合順位に比べジャンルでの順位が -501～-750
 - J 総合順位に比べジャンルでの順位が -751～-1000
 - K 総合順位に比べジャンルでの順位が -1001 以上
- [サンプル数] …… ジャンル内における、見出し漢字を含むサンプルファイル数
- [サンプルカバー率]
- …… [サンプル数] をそのジャンルのサンプルファイル数で割算したものを百分率で表示
- [サンプルカバー率 (全体)]
- …… ジャンルを分けて PB 全体でのサンプルファイルのカバー率
 - [サンプルカバー率差]
 - …… [サンプルカバー率 (全体)] から [サンプルカバー率] を減算して算出

5.2 総合頻度表の項目について

総合頻度表である「PB 総計 (度数) .xls」は、次の項目を表示している。

- [総合度数順位]
- …… [度数合計] による順位 (ジャンルを分けて PB 全体での度数を合計した場合の出現順位)
- [漢字] …… 見出しとなる漢字字種
- [度数合計] …… [PB_0] ～ [PB_9] 及び [PB_n] に示した度数の合計
- [度数%] …… [合計] を全漢字数で割算したものを百分率で表示
- [累積度数%] …… 度数%を出現順位順に加算
- [殊別] …… 「改定常用漢字表」との関連を表示 (7.2 で述べる)
- [配当学年] …… 小学校での配当学年
- [PB_0] …… 総記ジャンルでの度数
- [PB_1] …… 哲学ジャンルでの度数
- [PB_2] …… 歴史ジャンルでの度数
- [PB_3] …… 社会科学ジャンルでの度数
- [PB_4] …… 自然科学ジャンルでの度数
- [PB_5] …… 技術・工学ジャンルでの度数
- [PB_6] …… 産業ジャンルでの度数
- [PB_7] …… 芸術・美術ジャンルでの度数
- [PB_8] …… 言語ジャンルでの度数

- [PB_9] …… 文学ジャンルでの度数
- [PB_n] …… 分類なしでの度数

5.3 サンプル数一覧表の項目について

漢字の出現するサンプル数一覧表である「PB 総計 (サンプル数) .xls」は、次の項目を表示している。

- [サンプルカバー率総合順位]
 - …… [サンプルカバー率 (全体)] による順位 (ジャンルを分けて PB 全体でのサンプルのカバー率を見た場合の順位)
- [漢字] …… 見出しとなる漢字字種
- [サンプルカバー率 (全体)]
 - …… [出現サンプルファイル数合計] を PB 全体のサンプル数で割算したものを百分率で表示
- [殊別] …… 「改定常用漢字表」との関連を表示 (7.2 で述べる)
- [配当学年] …… 小学校での配当学年
- [出現サンプル数合計]
 - …… [PB_0] ~ [PB_9] 及び [PB_n] のサンプル数の合計
- [PB_0] …… 総記ジャンルでの出現サンプル数
- [PB_1] …… 哲学ジャンルでの出現サンプル数
- [PB_2] …… 歴史ジャンルでの出現サンプル数
- [PB_3] …… 社会科学ジャンルでの出現サンプル数
- [PB_4] …… 自然科学ジャンルでの出現サンプル数
- [PB_5] …… 技術・工学ジャンルでの出現サンプル数
- [PB_6] …… 産業ジャンルでの出現サンプル数
- [PB_7] …… 芸術・美術ジャンルでの出現サンプル数
- [PB_8] …… 言語ジャンルでの出現サンプル数
- [PB_9] …… 文学ジャンルでの出現サンプル数
- [PB_n] …… 分類なしでの出現サンプル数

6. 各ジャンルの特徴漢字の抽出の試み

ジャンルの特徴漢字の抽出は、PB 全体での漢字の出現傾向と当該ジャンルでの漢字の出現傾向との差によって見出せるのではないか。今回作成した漢字表の中で、漢字の出現傾向の比較材料としては、漢字出現の「順位差」「サンプルカバー率」の二つがある。

以下では、ジャンルでの特徴漢字の指標として、[順位差] と [サンプルカバー率差] とのどちらが有効であるかについて具体例を挙げながら検討した上で、特徴漢字の抽出を試みたい。

6.1 [順位差] についての検討

「常用漢字表」の改定に当たって文化庁が行った漢字出現頻度数調査では、出現度数に基づく順位によって漢字が配列されている。そこで、BCCWJ の PB についても、順位での比較が有効であるか検討してみたい。

ジャンル別頻度表「PB_0_総記.xls」「PB_1_哲学.xls」「PB_2_歴史.xls」「PB_3_社会科学.xls」

「PB_4_自然科学.xls」「PB_5_技術・工学.xls」「PB_6_産業.xls」「PB_7_芸術・美術.xls」「PB_8_言語.xls」「PB_9_文学.xls」「PB_n_分類なし.xls」では、PB 全体での〔総合順位〕から、当該ジャンルでの〔順位〕を減算して〔順位差〕を出している。〔順位差〕のプラス数値はジャンルでの〔順位〕が高く、マイナス数値はジャンルでの〔順位〕が低いことになる。

さらに、〔順位差〕を基にしてA～Kの11段階にランク分けをした結果を〔順位差ランク〕に示している。順位差ランクA～Eはジャンルでの順位の方が高い場合、ランクFはジャンルでの順位と総合順位とが同じ場合、ランクG～Kは総合での順位の方が高い場合ということになる。順位差ランクAの漢字は、BCCWJ全体での順位よりも、そのジャンルでの順位の方が1000位以上高位になっている。

この〔順位差〕が大きいことをもって、特徴漢字と考えてよいだろうか。例えば、サンプル数2267、総漢字数898,053と最も規模が大きい「3 社会科学」で順位差ランクがAのものについて上位10を示すと表2のようになる。

表2 「3 社会科学」における順位差ランクAの漢字（上位10）

漢字	順位差 ランク	順位差	順位	総合順位	頻度数	度数%
鄧	A	1083	1888	2971	12	0.001%
滌	A	1033	1888	2921	12	0.001%
湃	A	1050	1964	3014	10	0.001%
櫻	A	1001	2013	3014	9	0.001%
孜	A	1072	2062	3134	8	0.001%
旛	A	1008	2062	3070	8	0.001%
歴	A	1137	2062	3199	8	0.001%
叭	A	1008	2062	3070	8	0.001%
榜	A	1009	2125	3134	7	0.001%
錮	A	1016	2183	3199	6	0.001%
嬰	A	1016	2183	3199	6	0.001%

「3 社会科学」ジャンルでの〔順位〕は、〔総合順位〕と比べて1000位以上高位になっているが、〔度数〕〔度数%〕は極めて小さい数字となっており、「鄧」「滌」「湃」などを「3 社会科学」ジャンルの特徴漢字とすることはできない。

文化庁の『漢字出現頻度数調査(3)』では、扱われた漢字数が約4,900万漢字という大規模データであるが、本稿で使用したPBは290万弱である。さらにPBを11ジャンルに分けた場合、ジャンルによっては漢字数が約5万となってしまう。漢字数が少なくなるほど、どこをサンプリングしたかに左右され、集計結果に一般性を持たせることが難しくなることが予想される。以上から、順位及び順位差で比較することは難しいと判断せざるをえない。

6.2 「サンプルカバー率差」についての検討

BCCWJのPBは文化庁の『漢字出現頻度数調査(3)』よりも規模が小さい。したがって、BCCWJの漢字の使用実態を分析する場合、単純に順位（順位差）を用いるのではなく、何

らかの補正を行う必要がある。

PB では、書籍のジャンル別の生産実態に沿ったサンプリングをしているというのが他に類を見ない特徴である。そこで、ジャンルごとの割合が長所であることに着目し、漢字の出現についても、出現サンプルの割合が比較材料にならないかを検討してみたい。

ジャンル別頻度表「PB_0_総記.xls」「PB_1_哲学.xls」「PB_2_歴史.xls」「PB_3_社会科学.xls」「PB_4_自然科学.xls」「PB_5_技術・工学.xls」「PB_6_産業.xls」「PB_7_芸術・美術.xls」「PB_8_言語.xls」「PB_9_文学.xls」「PB_n_分類なし.xls」では、BCCWJ 全体での [サンプルカバー率 (全体)] から、そのジャンルでの [サンプルカバー率] を減算して算出して [サンプルカバー率差] を出している。

この中では、同頻度の漢字であっても一部のサンプルだけに偏って出現する場合と、ある程度の範囲のサンプルに分散して出現する場合とがある。例として、「4 自然科学」ジャンル (サンプル数 615, 総漢字数 217,397) での「医」と「方」を表 3 に示した。

表 3 「4 自然科学」における「医」「方」のサンプル数の違い

順位	漢字	頻度数	度数%	サンプル数	サンプルカバー率	サンプルカバー率 (全体)	サンプルカバー率差
18	医	1062	0.489%	218	35.4%	8.9%	26.5%
19	方	1061	0.488%	414	67.3%	57.5%	9.9%

「医」は [度数] 1062, [度数%] 0.489%, 「方」は [度数] 1061, [度数%] 0.488% で、頻度の上ではほぼ同じに見える。しかし、出現する [サンプル数] は「医」が 214, 「方」が 414 と大きな開きが見られる。「4 自然科学」ジャンル全体のサンプル数は 615 であるから、[サンプルカバー率] は「医」が 35.4%, 「方」が 67.3% になる。この数字を見ると、「方」が約 7 割のファイルに分布しており、「医」よりも「4 自然科学」ジャンルで遭遇することの多い漢字であるように見えるが果たしてそうであろうか。

BCCWJ 全体でのサンプルカバー率 ([サンプルカバー率 (全体)]) を見ると、「医」は 8.9%, 「方」は 57.5% であることが分かる。「方」は BCCWJ 全体で見ても半数以上のサンプルに出現し、「4 自然科学」ジャンルとの差は 9.9% である。一方、「医」は BCCWJ 全体では 8.9% で、「4 自然科学」ジャンルになると 26.5% 上昇して 35.4% となる。

このように見ると「医」の方が「4 自然科学」ジャンルで使用が上昇する漢字ということになる。この視点で、先の「3 社会科学」ジャンルについて [サンプルカバー率差] の大きいもの上位 10 を抽出してみたものが表 4 である。[サンプルカバー率差] が大きいものは、結果的には [度数] [順位] 等も高いものでもあることが分かる。

表4 「3 社会科学」における[サンプルカバー率差]の高い漢字(上位10)

漢字	サンプルカバー率差	サンプル数	サンプルカバー率	サンプルカバー率(全体)	順位差ランク	順位差	順位	総合順位	頻度数	度数%
業	22.7%	1197	52.8%	30.1%	E	24	9	33	6079	0.677%
社	20.2%	1138	50.2%	30.0%	E	30	20	50	4826	0.537%
定	19.9%	1366	60.3%	40.4%	E	18	19	37	4920	0.548%
的	19.3%	1838	81.1%	61.8%	E	4	3	7	8378	0.933%
対	19.1%	1473	65.0%	45.9%	E	18	30	48	4039	0.450%
要	18.7%	1461	64.4%	45.7%	E	23	29	52	4047	0.451%
基	18.0%	886	39.1%	21.0%	E	89	100	189	2018	0.225%
関	17.7%	1353	59.7%	42.0%	E	22	35	57	3779	0.421%
制	17.6%	849	37.5%	19.9%	E	90	74	164	2489	0.277%
会	16.2%	1433	63.2%	47.0%	E	17	6	23	6487	0.722%
法	16.2%	1110	49.0%	32.8%	E	25	17	42	5191	0.578%

このようなことから、以下では、[サンプルカバー率差]が高い漢字を基に、そのジャンルでの「特徴漢字」の抽出を試みる。

6.3 各ジャンルの特徴漢字の抽出結果

[サンプルカバー率差]の高い漢字を、そのジャンルでの「特徴漢字」と位置付けると、表5(本稿末尾に掲載)に示す漢字を抽出することができる。

表5では、「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」「n 分類なし」の11ジャンル各々での[サンプルカバー率差]が高い漢字(10.0%以上)を抽出して示した。これらの漢字が、各ジャンルでの特徴漢字ということになる。いずれも、現行の「常用漢字表」(2010年11月30日・内閣告示)における常用漢字である(「9 文学」に太字で示した「俺」「誰」は2010年の改定で新たに加わった漢字)。

各ジャンルの特徴漢字の傾向について、正確な分析はその漢字を使用した語を示した上でないといけないが、以下ではおおよその傾向について触れておくことにする。

「0 総記」「n 分類なし」は、内容が多岐にわたるジャンルといえるが、共通して「使」が入っている。

「1 哲学」では、「感」「心」「意」「想」「思」などの心的事項に関連すると思われる漢字、「神」「聖」「仏」などの宗教に関連すると思われる漢字が見られる。

「2 歴史」では、「五」「三」「六」「二」「七」「八」「四」のように数詞を構成すると思われる漢字が目立っている。

「3 社会科学」の特徴漢字は、経済、政治、法律などの分野で使われるもの、社会人となって生活する上で常用中の常用とも位置付けられよう。

「4 自然科学」では、「療」「病」「患」「症」「医」などに始まり、医療用語と関連すると思われる漢字が目立っている。

「5 技術・工学」では、「使」「用」「製」「設」「加」「作」「作」などの動作を表わす語と関連すると思われる漢字が目立つ(「使」は「0 総記」「n 分類なし」にも見られる)。

「6 産業」は漢字の特徴がつかみにくい¹。

「7 芸術・美術」では、「楽」「演」「曲」などの音楽と関連すると思われる漢字が目立っている。

「8 言語」は「語」「文」「意」「詞」などが見られる。「9 文学」では「彼」「僕」「俺」「私」「誰」「男」「女」「母」「子」「父」などの人を呼び表わすのに関連する漢字が目立つ。

以上のように、各ジャンルの特徴漢字は、字種の多少の重なりはあるものの、おおよその傾向として差が見られることが分かる。

7. 「改定常用漢字表」(2010年)との比較

7.1 「改定常用漢字表」に至る漢字の出入り

2005年3月30日、文部科学大臣は文化審議会に対して、「敬語に関する具体的な指針の作成について」及び「情報化時代に対応する漢字政策の在り方について」を諮問した。後者の諮問を受けて、「常用漢字表」(1981年内閣告示)の見直しが行われることになり、文化審議会国語分科会では漢字小委員会を設置して審議し、2度の意見募集を経た上で、2010年6月7日に「改定常用漢字表」2,136字を文部科学大臣に答申した。その後「改定常用漢字表」は、2010年11月30日に内閣告示となり現在に至っている²。

以下では、見直しが始まるまでからの漢字の出入りについて時系列に記した。この漢字の出入りに関する情報は、漢字表の「[殊別]」欄にも付与している。

7.1.1 「常用漢字表」(1981年)の改定の基礎資料

常用漢字表の字種見直しでは、まず、漢字小委員会によって追加及び削除の候補漢字が選定された。この作業では、文化庁文化部国語課による大規模の漢字出現頻度調査(以下、文化庁調査と呼ぶ)が基礎資料として使用された。文化庁調査の概要について、文化審議会答申「改定常用漢字表」(2010年6月7日)からの引用を表6に示す。

表6 文化庁調査一覧(「改定常用漢字表」, (9)ページによる)

	対象総漢字数	調査対象としたデータ
A 漢字出現頻度数調査(3)※1	49,072,315	書籍860冊分の凸版組版データ
B 上記Aの第2部調査	3,290,795	Aのうち教科書分の抽出データ
C 漢字出現頻度数調査(新聞)※2	3,674,613	朝日新聞2か月分の紙面データ
D 漢字出現頻度数調査(新聞)※2	3,428,829	読売新聞2か月分の紙面データ
E 漢字出現頻度数調査(ウェブサイト)	1,390,997,102	ウェブサイト調査の抽出データ

※1 Aの調査対象総文字数は「169,050,703」。また、Bとは別に、第3部として月刊誌4誌の抽出調査も実施している。これらの組版データは、いずれも平成16年、17年、18年に凸版印刷が作成したものである。

※2 C, Dは、いずれも平成18年10月1日～11月30日までの朝刊・夕刊の最終版を調査したデータである。

※3 調査全体の漢字数は「3,128,38,952」。このうち「電子掲示板サイトにおける投稿本文」のデータを除いたもの。

7.1.2 第1次・字種候補素案

2008年5月12日、第21回漢字小委員会で「第1次・字種候補素案」として、追加候補274字、削除候補6字(銚, 錐, 勺, 斤, 匆, 脹)が示された。なお、追加候補は、次の二

¹ 前ページに述べたように、その漢字がどのような語に使われているかを検討する必要があるが、これについては稿を改めたい。

² 「改定常用漢字表」は、2010年11月30日以降は「常用漢字表」となったが、本稿では従来のものと区別するために「改定常用漢字表」と称することにする。

つのグループに区分されていた。

- ・「本表に入れる可能性のある候補漢字」220字
- ・「別表に入れる可能性のある候補漢字」54字

7.1.3 第2次・字種候補案

2008年6月16日、第23回漢字小委員会で「第2次・字種候補案」として、追加候補188字が示された。この段階で、追加候補から脱落した字は、以下の86字であった。

叩、嘘、噂、濡、笠、嬉、朋、覗、撫、庄、溜、鷹、揃、頷、攪、翔、喋、嚙、洩、禄、栗、馴、駕、鴨、淵、駿、賭、蘭、胡、蘇、狼、蝶、搔、惚、蒼、腿、菩、吊、雀、樽、壺、祀、卿、歪、棲、釜、毅、磯、桶、柿、揆、蹠、躄、鷺、憐、猥、萌、媚、寵、秤、撥、遡、謳、套、刹、蔓、醬、疼、賤、顛、捏、糊、饅、倦、屏、毀、恍、幹、膠、誼、疇、謗、乖、戴、誹、綬（下線を付さない7字は、最終的に追加された。）

なお、削除候補は5字（銑、錐、勺、匆、脹）が示された。（常用漢字からの削除候補は、以後変更がない。）

7.1.4 「新常用漢字表(仮称)」に関する試案

2008年9月22日、第25回漢字小委員会で、追加候補に4字（刹、遡、椎、賭）が追加され、1字（蒙）が削除された。この結果、追加候補は191字となった。

7.1.5 「改定常用漢字表」に関する試案

2009年3月～4月に、一般国民及び各府省等を対象とした意見募集を実施し、そこで寄せられた意見を踏まえて再度の見直しを行った結果、「改定常用漢字表」に関する試案では、追加9字（柿、哺、楷、睦、釜、錮、賂、勾、毀）、削除4字（聘、憚、哨、諜）の修正を行い、結果的に196字が追加字種となった。

さらに、2009年11月～12月に2回目の意見募集を実施し、寄せられた意見を精査した上で更に検討を加えたが、答申「改定常用漢字表」でも、この196字の追加字種はそのまま踏襲されている。

7.2 「NDCジャンル別漢字出現頻度表」と「改定常用漢字表」

「NDCジャンル別漢字出現頻度表」（漢字表）では、[殊別]欄に「改定常用漢字表」に関連する情報を付与している。

[常用]

「当用漢字表」（1946年）から「常用漢字表」（1981年）を経て「改定常用漢字表」に継続して入った1,850字について、漢字表の[殊別]欄では「常用」と記した。

[常用・旧]

上記「常用」の旧字体について、漢字表の[殊別]欄では「常用・旧」と記した。

[常用1981]

「常用漢字表」（1981年）から新たに追加された95字について、漢字表の[殊別]欄では「常用1981」と記した。

[常用1981・旧]

上記「常用1981」の旧字体について、漢字表の[殊別]欄では、「常用1981・旧」と記した。

〔常用 2010〕

「改定常用漢字表」(2010年)から新たに追加された196字について、漢字表の〔殊別〕欄では追加漢字を「常用 2010」と記した。なお、196字は、次のものである。

挨, 曖, 宛, 嵐, 畏, 萎, 椅, 彙, 茨, 咽, 淫, 唄, 鬱, 怨, 媛, 艶, 旺,
岡, 臆, 俺, 苛, 牙, 瓦, 楷, 漬, 諧, 崖, 蓋, 骸, 柿, 顎, 葛, 釜, 鎌,
韓, 玩, 伎, 亀, 毀, 畿, 臼, 嗅, 巾, 僅, 錦, 惧, 串, 窟, 熊, 詣, 憬,
稽, 隙, 桁, 拳, 鍵, 舷, 股, 虎, 錮, 勾, 梗, 喉, 乞, 傲, 駒, 頃, 痕,
沙, 挫, 采, 塞, 埼, 柵, 刹, 拶, 斬, 恣, 摯, 餌, 鹿, 叱, 嫉, 腫, 呪,
袖, 羞, 蹴, 懂, 拭, 尻, 苾, 腎, 須, 裾, 凄, 醒, 脊, 戚, 煎, 羨, 腺,
詮, 箋, 膳, 狙, 邈, 曾, 爽, 瘦, 踪, 捉, 遜, 汰, 唾, 堆, 戴, 誰, 旦,
綻, 緻, 耐, 貼, 嘲, 抄, 椎, 爪, 鶴, 諦, 溺, 填, 妬, 賭, 藤, 瞳, 枳,
頓, 貪, 井, 那, 奈, 梨, 謎, 鍋, 勺, 虹, 捻, 罵, 剝, 箸, 汜, 汎, 阪,
斑, 眉, 膝, 肘, 訃, 阜, 蔽, 餅, 璧, 蔑, 哺, 蜂, 貌, 頰, 睦, 勃, 昧,
枕, 蜜, 冥, 麵, 冶, 弥, 闍, 喻, 湧, 妖, 瘍, 沃, 拉, 辣, 藍, 璃, 慄,
侶, 瞭, 瑠, 呂, 賂, 弄, 籠, 麓, 脇

〔常用 2010・旧〕

上記「常用 2010」の旧字体について、漢字表の〔殊別〕欄では「常用 2010・旧」と記した。

〔候補 2010〕

「改定常用漢字表」に至る検討過程で追加候補に挙げられたが、結果的に追加されなかった字種は以下のものである。漢字表の〔殊別〕欄では「候補 2010」と記した。

叩, 噓, 嚙, 濡, 笠, 嬉, 朋, 視, 撫, 庄, 溜, 鷹, 揃, 頷, 摑, 翔, 喋,
嚙, 洩, 祿, 栗, 馴, 駕, 鴨, 淵, 駿, 蘭, 胡, 蘇, 狼, 蝶, 搔, 惚, 蒼,
腿, 菩, 吊, 雀, 樽, 壺, 祀, 卿, 歪, 棲, 毅, 磯, 桶, 揆, 踏, 躊, 驚,
憐, 猥, 萌, 媚, 寵, 秤, 撥, 謳, 套, 蔓, 醬, 疼, 賤, 顛, 捏, 糊, 饅,
倦, 屏, 恍, 幹, 膠, 誼, 疇, 謗, 乖, 誹, 綬, 蒙, 聘, 憚, 哨, 諜

〔削除 2010〕

「常用漢字表」(1981年)の表内字のうち、「改定常用漢字表」(2010年)に入らなかった字種は以下の5字である。漢字表の〔殊別〕欄では「削除 2010」と記した。

勺, 錘, 銑, 脹, 匆

7.3 追加漢字のジャンル

「改定常用漢字表」で新たに追加された漢字やその検討段階で候補に挙げられた漢字は、どのジャンルに多く使われるのであろうか。

表7では、各ジャンルで用いられた〔常用 2010〕〔常用 2010・旧〕の漢字のうち、[サンプルカバー率]が0.5%以上のものを列記した。また、追加漢字となった字種数が多い順にジャンルを示すと、「9 文学」「3 歴史」「8 言語」「7 芸術・美術」「4 自然科学」「1 哲学」「0 総記」「8 産業」「n 分類なし」「5 技術・工学」となる。数だけを見ると人文系の中でも「9 文学」「3 歴史」「8 言語」での追加漢字が目立つ。

表7 各ジャンルでのサンプルカバー率が0.5%以上の追加漢字（0.5%以上）

ジャンル	〔常用2010〕（〔常用2010・旧〕）	
	字種数	字種
0 総記	19	呂 凄 曾 須 貼 挨 詣 璧 梨 鎌 爽 艶 伎 汎 捻 戴 餅 瘦 詮
1 哲学	24	誰 頃 弥 捉 窟 侶 味 鎌 嵐 葛 畏 乞 湧 臆 諦 挫 呪 詣 喻 醒
2 歴史	47	藤 頃 岡 鎌 阪 鹿 呂 韓 那 弥 熊 脇 亀 葛 籠 呪 瓦 斬 宛 挨 虎 蓋 麓 (曾) 塞 梨 釜 痕 憧 嵐 串 湧 茨 勃 戴 曾 崖 蔑 錦 栃 遡 諳 畏 蜂 侶 怨 緻
3 社会科学	3	韓 捉 綻
4 自然科学	26	腫 瘍 腎 腺 塞 梗 斑 潰 脊 椎 膝 梨 咽 尻 哺 須 煎 萎 痕 葛 肘 蜂 餅 頓 埼 桁 貼 串 勾 湧 抄 煎 臼 汎
5 技術・工学	13	鍋 芯 須 柿 桁 貼 串 勾 湧 抄 煎 臼 汎
6 産業	18	岡 鍵 俺 那 捉 斑 諦 鶴 韓 潰 挨 釜 狙 惧 脇 椅 璧 瓦
7 芸術・美術	29	藤 頃 岡 阪 狙 稽 膝 駒 斬 唄 弥 凄 鶴 呪 鹿 隙 挨 拶 伎 尻 壁 憧 僅 骸 梨 玩 錦 旺 緻
8 言語	37	韓 囊 捉 藤 瞭 枕 遡 頓 井 尻 畏 侶 堆 貪 恣 遜 曖 沙 唄 諧 栃 蔽 戴 崎 艶 裾 麵 汰 井 慄 膝 瑠 汎 瘦 冶 璃 弄 鑰 籠 虎 拭 誰 藤 奈 岡 椅 閣 那 瞳 眉 沙 膝 弥 勾 斬 喉 脇 枕 鎌 籠 虎 袖 狙 揆 揆 眺 旦 股 淫 拳 顎 妖 潰 駒 爪 剥 熊 枕 鎌 籠 虎 芯 柵 瘦 箸 懂 崖 虹 溺 羞 彌 臆 脛 瞭 釜 貪 拉
n 分類なし	14	鍋 貼 狙 呪 鍵 巾 釜 爽 須 瓦 爪 桁 踪 睦

また、表8では〔候補2010〕の漢字のうち、〔サンプルカバー率〕が0.5%以上のものを列記した。「9 文学」「3 歴史」「8 言語」での追加候補であった漢字が目立つ。

表8 各ジャンルでのサンプルカバー率が0.5%以上の追加候補漢字（0.5%以上）

ジャンル	〔候補2010〕	
	字種数	字種
0 総記	1	吊
1 哲学	7	菩 蒙 祀 蘇 朋 倦 疇
2 歴史	17	笠 蘇 禄 祀 嬉 庄 卿 蒙 磯 挨 鴨 樽 駿 淵 毅 菩 蔓
3 社会科学	0	
4 自然科学	6	疼 膠 溜 蔓 倦 棲
5 技術・工学	3	吊 醬 鴨
6 産業	4	揃 祀 翔 摺
7 芸術・美術	9	揃 嬉 雀 磯 叩 馴 蘇 朋 媚
8 言語	15	疇 蘇 撥 菩 嘘 栗 禄 顛 賤 蒼 誼 諜 綬 籠 聘
9 文学	35	叩 嬉 濡 頷 視 撫 嚙 溜 嘘 摺 嚙 栗 庄 喋 蘭 腿 狼 胡 揃 惚 洩 翔 歪 蒼 馴 搔 嘘 萌 鷹 壺 躄 駕 蝶 憐 憚
n 分類なし	5	揃 狼 秤 醬 鴨

今回の改定では、字種数の上では、「9 文学」「3 歴史」「8 言語」のサンプルカバー率の高い漢字が目立つ。しかしこれをもって、漢字の追加が3ジャンル中心であったと断じるのは早計である。漢字使用が従来の常用漢字の範囲内で収まっているジャンルかも知れないからである。表7で11ジャンル全部に追加漢字が見られることに注目すれば、今回の追加漢字が特定の偏りを示していないと判断するのが妥当であろうと考える。

8. まとめ

本稿では、現代日本語書き言葉均衡コーパス（BCCWJ）を資料として、各ジャンルの特徴漢字の抽出を試みた。〔サンプルカバー率差〕を基にして抽出した結果、表5に示した各ジャンルの特徴漢字を抽出することができた、各ジャンルの特徴漢字の傾向は、6.3で述べたとおり、おおよその差が見られた。

また、「常用漢字表」に新たに追加された漢字のジャンルについては、表7のように11ジャンル全部に〔サンプルカバー率〕の高い追加漢字が見られる。今回の追加漢字は特定

の偏りを示していないと判断するのが妥当であろう。

表5 各ジャンルの「サンプルカバー率差」上位（10.0%以上）の漢字（ジャンルの特徴漢字）

「0 総記」

漢字	サンプルカバー率差	順位
作	15.6%	15
使	15.2%	38
用	13.2%	14
数	13.0%	17
書	12.4%	10
示	12.0%	50
設	11.9%	64
利	11.9%	82
報	11.3%	79
選	11.1%	105
文	11.1%	21
初	10.7%	115
字	10.6%	97
読	10.5%	121
違	10.1%	163

「1 哲学」

漢字	サンプルカバー率差	順位
世	20.8%	37
感	20.7%	40
神	20.3%	11
心	19.5%	22
教	16.6%	13
自	16.3%	3
意	15.5%	34
考	13.4%	38
聖	13.4%	145
在	13.1%	75
身	13.1%	63
言	12.7%	8
想	12.6%	138
的	12.4%	5
思	12.3%	12
仏	12.0%	99
界	11.9%	66
性	11.9%	31
人	11.8%	1
命	11.4%	93
愛	11.1%	112
何	10.9%	62
私	10.9%	14
生	10.4%	4
語	10.4%	86
相	10.2%	81
観	10.2%	145
霊	10.1%	147

「2 歴史」

漢字	サンプルカバー率差	順位	漢字	サンプルカバー率差	順位
五	27.8%	26	南	12.0%	240
三	27.4%	11	府	11.9%	164
国	26.5%	7	都	11.8%	136
九	25.5%	36	古	11.6%	174
地	25.3%	10	住	11.1%	244
六	25.0%	50	所	11.0%	62
二	24.6%	6	婦	11.0%	291
七	24.0%	65	在	10.7%	142
八	24.0%	46	野	10.7%	132
四	22.4%	37	木	10.6%	172
日	20.6%	4	城	10.6%	216
代	20.5%	31	海	10.5%	92
年	20.1%	5	市	10.4%	107
山	20.1%	28	勢	10.4%	261
東	19.0%	75	小	10.3%	61
西	18.3%	119	周	10.3%	288
月	17.8%	33	台	10.3%	226
戦	17.7%	39	記	10.2%	102
北	17.4%	135	紀	10.1%	236
政	17.4%	51	公	10.1%	158
史	17.4%	125	武	10.0%	188
田	17.1%	54	武	10.0%	349
世	17.0%	56			
民	16.8%	79			
名	16.7%	48			
時	16.7%	12			
川	15.9%	85			
近	15.8%	91			
後	15.7%	25			
十	15.2%	14			
和	14.8%	112			
長	14.7%	38			
新	14.7%	55			
土	14.7%	160			
道	14.5%	58			
軍	14.1%	57			
天	14.0%	99			
朝	14.0%	170			
治	13.7%	117			
百	13.6%	277			
建	13.6%	163			
平	13.4%	108			
家	12.7%	22			
本	12.6%	9			
大	12.5%	3			
中	12.5%	8			
町	12.5%	118			
前	12.3%	32			
文	12.2%	59			
島	12.1%	105			

「3 社会科学」

漢字	サンプルカバー率差	順位	漢字	サンプルカバー率差	順位
業	22.7%	9	限	10.9%	200
社	20.2%	20	活	10.8%	91
定	19.9%	19	認	10.8%	153
的	19.3%	3	企	10.8%	125
対	19.1%	30	現	10.7%	71
要	18.7%	29	当	10.4%	38
基	18.0%	100	營	10.3%	185
関	17.7%	35	得	10.3%	119
制	17.6%	74	理	10.2%	34
会	16.2%	6	特	10.2%	118
法	16.2%	17	担	10.1%	252
資	16.0%	48	任	10.1%	189
経	15.8%	70	実	10.1%	39
務	15.6%	62			
係	15.4%	137			
保	15.3%	43			
支	15.1%	105			
問	14.3%	80			
規	13.8%	140			
権	13.8%	46			
題	13.7%	101			
成	13.7%	41			
民	13.4%	69			
金	13.3%	23			
第	13.3%	60			
点	12.9%	136			
利	12.9%	81			
決	12.9%	114			
者	12.8%	5			
等	12.7%	97			
価	12.6%	77			
政	12.6%	54			
済	12.5%	152			
義	12.4%	164			
方	12.2%	27			
主	12.2%	55			
期	12.1%	85			
産	12.1%	72			
必	12.0%	102			
提	12.0%	204			
例	12.0%	183			
員	11.9%	86			
有	11.9%	115			
用	11.8%	31			
年	11.8%	4			
意	11.6%	64			
以	11.5%	87			
条	11.4%	94			
設	11.4%	132			
発	11.3%	58			

「4 自然科学」

漢字	サンプル カバー率 差	順位	漢字	サンプル カバー率 差	順位
療	33.3%	10	吸	10.9%	226
病	30.8%	20	低	10.8%	177
患	28.8%	40	高	10.5%	53
症	28.4%	43	含	10.4%	272
医	26.5%	18	能	10.3%	76
性	25.7%	5	変	10.2%	69
状	23.2%	47	器	10.1%	240
血	21.5%	34	食	10.0%	31
質	18.5%	79	痛	10.0%	173
化	17.8%	28	管	10.0%	170
量	17.5%	83			
科	16.4%	109			
多	16.3%	49			
薬	16.0%	57			
体	16.0%	8			
診	15.9%	119			
発	15.7%	38			
効	15.1%	109			
学	14.8%	14			
治	14.6%	37			
果	14.4%	97			
因	13.9%	194			
害	13.9%	123			
経	13.7%	61			
酸	13.6%	114			
液	13.5%	213			
態	13.4%	117			
胞	13.3%	99			
感	13.1%	58			
障	13.0%	180			
重	12.9%	98			
検	12.6%	146			
類	12.6%	126			
健	12.5%	172			
活	12.5%	93			
常	12.4%	109			
的	12.3%	6			
剂	12.2%	234			
臓	12.1%	278			
細	11.7%	67			
疾	11.7%	313			
種	11.7%	131			
法	11.6%	33			
期	11.3%	84			
増	11.2%	186			
素	11.0%	126			
起	11.0%	135			
康	10.9%	298			
院	10.9%	133			
脳	10.9%	165			

「5 技術・工学」

漢字	サンプル カバー率 差	順位
使	23.3%	32
用	19.3%	4
製	15.2%	112
設	15.2%	38
料	14.1%	50
加	13.9%	82
量	13.8%	131
工	13.7%	61
品	13.3%	57
材	13.2%	112
入	13.1%	18
図	12.5%	47
作	12.1%	27
造	11.3%	135
切	10.8%	67
技	10.5%	168
定	10.5%	17
電	10.2%	60
器	10.1%	146
構	10.1%	159

「6 産業」

漢字	サンプル カバー率 差	順位	漢字	サンプル カバー率 差	順位
本	52.1%	18	活	15.1%	106
見	51.0%	45	無	14.9%	229
多	47.0%	57	理	14.7%	33
高	46.1%	36	時	14.1%	21
人	45.7%	2	考	14.1%	94
手	44.0%	38	点	13.9%	154
的	36.0%	8	供	13.8%	270
以	33.5%	99	進	13.4%	222
代	33.0%	88	原	13.3%	136
発	32.9%	47	総	13.1%	410
方	30.2%	24	今	13.0%	148
大	28.4%	4	向	12.9%	133
話	28.4%	113	指	12.7%	222
部	27.2%	73	上	12.7%	15
間	22.9%	16	際	12.6%	258
家	22.0%	102	持	12.5%	118
仕	21.6%	140	細	12.5%	389
気	21.0%	68	食	12.4%	86
付	21.0%	237	規	12.2%	276
状	20.9%	196	解	11.9%	180
来	20.4%	123	明	11.9%	104
増	20.3%	222	場	11.8%	6
題	20.0%	180	独	11.8%	461
全	19.8%	85	最	11.7%	84
現	19.3%	79	過	11.7%	273
先	19.2%	175	構	11.6%	262
決	19.2%	178	不	11.3%	117
然	18.9%	296	起	11.2%	314
面	18.5%	196	知	11.2%	152
所	18.4%	101	予	11.1%	230
続	18.2%	185	例	11.1%	286
前	18.0%	51	設	11.0%	157
直	18.0%	212	次	10.9%	211
国	18.0%	49	返	10.9%	367
水	17.7%	96	温	10.7%	384
意	17.4%	79	環	10.6%	358
特	17.4%	109	楽	10.5%	331
主	17.2%	95	減	10.4%	310
数	17.1%	90	接	10.3%	309
備	17.0%	250	与	10.3%	399
相	17.0%	188	保	10.3%	144
結	16.8%	144	管	10.2%	238
下	16.6%	64	積	10.1%	481
初	16.4%	250	毎	10.1%	457
近	15.9%	170	帰	10.0%	488
少	15.6%	157	義	10.0%	502
可	15.5%	168			
自	15.3%	27			
重	15.2%	172			
会	15.2%	28			

「7 芸術・美術」

漢字	サンプル カバー率 差	順位
楽	18.0%	39
感	17.1%	44
演	14.2%	88
選	12.4%	51
曲	12.3%	100
作	11.8%	18
戦	10.7%	48
打	10.5%	117
勝	10.5%	71
描	10.5%	204

「8 言語」

漢字	サンプル カバー率 差	順位
語	53.9%	1
文	35.0%	3
意	30.8%	11
味	27.2%	33
表	24.7%	25
詞	23.5%	36
使	22.3%	38
言	21.8%	2
書	19.7%	17
名	19.6%	34
英	17.8%	80
字	17.8%	50
代	17.4%	49
象	17.4%	130
音	17.3%	12
方	17.1%	14
単	17.0%	109
形	15.8%	57
異	14.8%	172
際	14.7%	162
説	14.3%	118
考	14.3%	57
例	14.0%	64
点	13.6%	120
違	13.5%	123
読	13.2%	66
学	13.1%	16
明	12.7%	48
古	12.5%	129
葉	12.5%	68
現	12.2%	31
本	11.7%	5
感	11.6%	70
句	11.4%	277
辞	11.3%	175
記	11.0%	81
様	10.6%	121
世	10.1%	99
的	10.1%	8

「9 文学」

漢字	サンプル カバー率 差	順位
声	32.6%	46
彼	29.2%	14
顔	27.0%	45
話	24.8%	29
僕	24.3%	120
屋	24.0%	69
気	23.9%	11
男	23.0%	50
女	22.7%	16
母	21.5%	84
言	21.3%	4
死	21.2%	86
笑	20.9%	82
思	20.6%	7
十	19.7%	32
郎	19.4%	110
何	18.1%	21
見	18.0%	3
口	17.5%	51
俺	17.5%	126
夜	16.9%	142
手	16.5%	8
私	15.4%	28
二	15.3%	19
美	15.2%	102
兵	14.7%	210
子	14.7%	6
父	14.4%	93
軍	13.7%	190
殺	13.4%	172
首	13.3%	209
家	13.1%	34
太	12.9%	135
葉	12.5%	106
返	12.3%	130
息	12.2%	192
真	12.2%	97
誰	12.1%	191
王	11.8%	264
飛	11.8%	195
歩	11.2%	153
食	11.0%	116
隊	10.9%	260
着	10.5%	123
石	10.5%	222
警	10.4%	260
聞	10.3%	55
胸	10.1%	339
頭	10.1%	112

「n 分類なし」

漢字	サンプル カバー率 差	順位
使	23.6%	15
選	17.6%	35
入	11.3%	4
画	11.3%	31
色	10.4%	66

文献

文化審議会(2010), 『改定常用漢字表』(2010年6月7日, 文化審議会答申).
文化庁文化部国語課(2007), 『漢字出現頻度数調査(3)』, 文化庁文化部国語課.
丸山岳彦(2009), 『現代日本語書き言葉均衡コーパス』領域内公開データ(2009年度版)書誌
情報・サンプル情報・著者情報について』『現代日本語書き言葉均衡コーパス』領域内公
開データ(2009年度版)DVD-R.

関連 URL

特定領域研究「日本語コーパス」ホームページ : <http://www.tokuteicorpus.jp/>
KOTONOHA ホームページ : <http://www2.kokken.go.jp/kotonoha/>
文化審議会答申「改定常用漢字表」 : <http://www.bunka.go.jp/>

第3章

国語教育でのコーパスの活用

第3章 第1節

中学校・高校教科書の教科特徴語リストの作成

近藤 明日子

1. はじめに

本稿筆者は、学校教育での言語活動の充実を図るための基礎的な資料とすべく、コーパスを利用した教科書の教科特徴語の抽出について、中学校教科書を例とした試行を行ってきた（近藤、2008a；2008b；2009）。教科特徴語とは、当該教科において特徴的に高頻度に出現する語彙のことであり、当該教科の語彙指導において重要な位置を占める教科特徴語のリスト化は、言語活動の充実に欠かせないものとする。

本稿では、これまでの試行の成果としての教科特徴語リストの作成について、その作成方法と仕様について報告する。

2. 教科特徴語リストの作成方針

教科特徴語リストの作成にあたっては、語彙指導の基本的な資料となり、かつ語彙指導の現場で実用性のあるものとするため、以下の方針に基づくこととする。

2. 1 リスト作成対象の校種

中学校や高校のある教科の教科書といった、特定分野の特徴語を抽出する方法として、その特定分野のコーパス（対象コーパス）とその比較対象とするコーパス（参照コーパス）を用意し、参照コーパスよりも対象コーパスで偏って高頻度に出現する語を特徴語として抽出する方法がある。本稿もこの方法に従って各教科の特徴語の抽出を行う。

本稿で教科特徴語とするものは、「児童・生徒が日常生活で触れる書き言葉と比較して、当該教科の教科書で特に偏って高頻度に出現する語」である。これを抽出するためには、対象コーパスとして児童・生徒が学校で使用する教科書の言葉を代表するコーパスを、参照コーパスとして児童・生徒が日常生活で触れる書き言葉を代表するコーパスを用意する必要がある。

そこで、まず対象コーパスとして、特定領域研究「日本語コーパス」言語政策班が構築した「教科書コーパス」を利用する。「教科書コーパス」は2005年度に小学校・中学校・高等学校で用いられた検定教科書（各学年・各教科1種ずつ）を対象とした全文コーパスであり、児童・生徒が学校で使用する教科書の言葉を代表するものとして設計されたコーパスである¹。この「教科書コーパス」を校種・教科別に分割し、それぞれを対象コーパスとする。

一方、参照コーパスとして利用できるものに、『現代日本語書き言葉均衡コーパス』（BCCWJ）²の図書館サブコーパスがある。これは、公共図書館の収蔵図書をもとにしたコーパスで、ある程度広い範囲に流通したことが確認されている書き言葉を代表するものとして設計されているものである。これを「児童・生徒が日常生活で触れる書き言葉」

¹ 「教科書コーパス」の詳細については、田中・近藤・平山（2011）を参照のこと。

² <http://www.ninjal.ac.jp/kotonoha/>

に相当するものと想定し、参照コーパスとすることが考えられる。ただし、一口に「児童・生徒」とは言っても年齢の幅は広い。年齢の高い高校生ならば、その日常生活は社会人のそれに近く、BCCWJ 図書館サブコーパスを「日常生活で触れる書き言葉」の代表とすることに妥当性もあろう。しかし、低年齢になればなるほど、「日常生活で触れる書き言葉」と BCCWJ 図書館サブコーパスとの間には乖離が生じ、小学生ともなれば BCCWJ 図書館サブコーパスを「日常生活で触れる書き言葉」の代表とすることにはかなり無理がある。よって、本稿では BCCWJ 図書館サブコーパスを「日常生活で触れる書き言葉」の代表として想定できるのは、高校生および中学生までとし、BCCWJ 図書館サブコーパスを参照コーパスとする方法での教科特徴語の抽出もまた高校生および中学生の使用する教科書からのみ行うこととする。

2. 2 教科特徴語抽出の対象とする教科書の文書要素

対象コーパスすなわち中学校・高校の各教科の教科書の内部は、その機能から「主要学習部分」「補助」「発展」「図表」「注」「引用」の文書要素に分けることができる(近藤、2010)。これらの文書要素のうち、教科書の主幹となるのは、各単元で学習する主な内容を説明する機能を持つ「主要学習部分」要素である。基本的なリスト作成のために、この「主要学習部分」要素に出現する語彙から教科特徴語を抽出する。よって、対象コーパスは中学校・高校の各教科の教科書全体ではなく、「主要学習部分」要素の部分とする。

2. 3 語の単位

リスト作成に際し、コーパスの形態素解析は形態素解析辞書 UniDic³を使って行う。短単位⁴を解析単位とする UniDic を用いた解析では、例えば「全自動洗濯機」のような合成名詞は「全／自動／洗濯／機」のように短い単位に分割される。本稿で抽出する教科特徴語は、語彙指導での利用を目的としている面からも、また合成名詞の多い専門用語を多く含むという面からも、合成名詞は切らずに一つの単位としたままのほうがよいと考える。そこで、UniDic による解析結果から、一定の条件をみたす短単位連続を合成名詞に近似するものと見なし、1 単位として再構成し、語の単位として利用する。

2. 4 収録語の選定

教科特徴語は、対象コーパスに出現する各語について、対象コーパスでの度数と参照コーパスでの度数から算出される特徴度(後述)の値に基づき抽出する。しかし、特徴度がある水準より大きい語すべてを教科特徴語として機械的にリストに収録するにはいろいろな問題がある。例えば、形態素解析や単位再構成の誤りによって実際にはコーパスに出現しない語が教科特徴語として抽出される場合がある。このようなものは手作業での修正が必要である。また、教科特徴語として抽出された語のなかには、教科書の特定の題材に由来するゆえに、当該教科において一般性の低いものが混在する。こうした語は基本的なリストでの重要度は低いと考え、本稿で作成するリストからは除外したい。そのためには、教科での一般性が低いかな否かを人によって判断する必要がある。

そのような人手によるリストの整備を経て得られる教科特徴語リストでも、その掲載語数が多すぎるとは実用的なリストとは言えないであろう。よって、校種・教科ごとに作成す

³ <http://download.unidic.org/>

⁴ 短単位の規定の詳細は小椋・小磯・富士池・宮内・原(2010)を参照のこと。

るリストに掲載する語数に上限を設ける。

3. リスト作成の方法

ここでは、2で述べた方針によるリスト作成の具体的な方法について説明する。

3. 1 使用コーパス

まず、対象コーパスとして利用する「教科書コーパス」⁵から中学校・高校教科書部分を取り出し、それを「教科書コーパス」での教科分類に従い、「国語・数学・理科・社会・外国語・技術家庭・芸術・保健体育・情報」の9種（中学は情報がないので8種）のコーパスに分割する。この17種のコーパスから、`citation`・`figureBlock`・`noteBody`・`skippedSpan`・`skippedBlock`・`supplement`の各要素を除外して得られる「主要学習部分」要素のみを対象コーパスとして用意する。

次に、参照コーパスとして、BCCWJ 図書館サブコーパスに収録予定の固定長サンプル(LB_FL)、計10,640サンプル⁶を用意する。

3. 2 形態素解析と同語異語判別

17種の対象コーパスと1種の参照コーパスは、形態素解析辞書UniDic (MeCab版)の最新版(非公開)を用いて形態素解析する⁷。ただし、合成名詞は切らずに一つの単位とするために、解析結果から、一定の条件をみたす短単位連続⁸を合成名詞に近似するものと見なし、1単位として再構成する⁹。

短単位および再構成された単位の同語異語判別は、UniDicによって付与される属性¹⁰のうち「語彙素読み」「語彙素」「語彙素細分類」「語種」「品詞」「活用型」を用い、これらの値がすべて一致するものを同語と見なし、一つの見出し語のもとにまとめる。再構成した単位の場合は、「語彙素読み」「語彙素」属性の値には、構成前の短単位の属性値を結合したものをを用いる¹¹。「語彙素細分類」「語種」「活用型」属性の値には、構成前の短単位の値を+で結合したものをを用いる。「品詞」属性の値は、構成前の短単位が「接頭辞」のみからなるものは「接頭辞」、「接尾辞」のみからなるものは「接尾辞」、最後尾の短単位が活用型「形

⁵ 2010年12月9日版(非公開)を利用する。

⁶ 2010年12月9日版(非公開)を利用する。

⁷ この形態素解析結果は国立国語研究所内のBCCWJのための形態論情報データベースで管理されている。このデータベースの2010年12月9日時点のデータを利用する。

⁸ UniDicによって付与される品詞属性値が「名詞-普通名詞」「名詞-固有名詞」「接頭辞」「接尾辞」「形状詞-一般」「形状詞-タリ」のいずれかで始まる短単位が複数連続するものを1単位として再構成する。ただし、その短単位連続の先頭に品詞属性値が「接尾辞」で始まる単位が位置する場合と、末尾に品詞属性値が「接頭辞」で始まる単位が位置する場合は、それぞれを短単位連続から切り出し、別の単位と認定する。

⁹ 合成名詞等を長まま1単位とする言語単位として、BCCWJで採用されている長単位がある。将来的には、教科特徴語の抽出で長単位による形態素解析結果を利用することを予定している。長単位の規程の詳細は小椋・小磯・富士池・宮内・原(2010)を参照のこと。

¹⁰ UniDicの付与する属性の詳細について、UniDic同梱のマニュアルを参照のこと。

¹¹ 「語彙素読み」の値については、合成名詞となることで連濁等の語形変化が起きるものは、短単位の「語彙素読み」を単純に結合した値を用いるのではなく、合成名詞として一般的なものに修正したものをを用いる。例えば、「カザン(火山)」と「ハイ(灰)」で構成される合成名詞の「語彙素読み」は「カザンハイ」ではなく「カザンバイ」に修正した値を用いる。また、「語彙素」の値については、再構成前の短単位では人名・地名等の固有名詞の多くが片仮名表記となるが、合成名詞の「語彙素」では文脈での表記に修正した値を用いる。例えば、「シュシ」と「学(ガク)」で構成される合成名詞の「語彙素」は「シュシ学」ではなく「朱子学」に修正した値を用いる。

容詞」の接尾辞からなるものは「形容詞」、それ以外は「合成名詞」とする。

このようにして得られた対象コーパス・参照コーパスの語彙のうち、UniDicの付与する品詞属性の値が「名詞・代名詞・形状詞・連体詞・副詞・接続詞・感動詞・動詞・形容詞・接頭辞・接尾辞」のいずれかで始まる見出し語と、短単位連続から再構成した単位からなる見出し語を対象に教科特徴語の抽出を行う（助詞・助動詞・記号類は対象外とする）。これらの見出し語の異なり語数・延べ語数を示すと表1のようになる。

3. 3 特徴度の算出

次に、各対象コーパスの語彙について、対象コーパスでの度数と参照コーパスでの度数を比較し、対象コーパスに偏って高頻度に出現する程度（以下、「特徴度」）を数値化する。特徴度の指標とする統計値として、対数尤度比（log-likelihood ratio、 G^2 ）を用いる。対数尤度比は、英語学において特徴語の抽出のための指標として一定の評価を得ているものである（石川、2008、p.99）。

対象コーパスに出現する語 W の対数尤度比は、次の式[1]によって求めることができる（Kilgarriff、2001）。

$$G^2 = 2(a \ln a + b \ln b + c \ln c + d \ln d - (a+b) \ln(a+b) - (a+c) \ln(a+c) - (b+d) \ln(b+d) - (c+d) \ln(c+d) + (a+b+c+d) \ln(a+b+c+d)) \quad \dots[1]^{12}$$

- a : 対象コーパスでの語 W の度数
- b : 参照コーパスでの語 W の度数
- c : 対象コーパスの延べ語数 $-a$
- d : 参照コーパスの延べ語数 $-b$

さらに、単語 W の対象コーパスでの使用率が参照コーパスでの使用率より低い場合（ $ad-bc < 0$ の場合）、対数尤度比に -1 を乗じる補正（内山・中條・山本・井佐原、2004）を行った値を単語 W の特徴度とする。特徴度は、単語 W が参照コーパスに比べて対象コーパスでより高頻度に出現する場合、正の値をとり、高頻度に出現する偏りの程度が大きいほど大きい値をとる。

3. 4 リスト収録語の選定

まず、特徴度が 10.83 より大きい語（ $p < .001$ ）を有意に偏って高頻度に出現する語と見なし（高見、2003）、各校種・教科の特徴語としていったん抽出する。リスト収録候補として抽出された見出し語の異なり語数・延べ語数を表2に示す。

表 1 対象コーパス・参照コーパスの異なり語数・延べ語数

		異なり語数	延べ語数	
対象コーパス	中学校	国語	6,113	30,806
		数学	970	13,012
		理科	2,463	17,703
		社会	7,607	38,746
		外国語	975	3,135
		技術家庭	2,650	12,685
		芸術	4,025	17,092
		保健体育	1,668	6,740
		高校	国語	6,716
	数学		2,430	46,181
	理科		15,408	192,616
	社会		35,787	293,080
	外国語		1,398	3,674
	技術家庭	7,202	42,766	
芸術	10,648	54,976		
保健体育	3,755	19,123		
情報	3,855	28,295		
参照コーパス		256,186	3,452,846	

¹² \ln は自然対数を表す。また、 a または b が 0 の場合、 $a \ln a$ または $b \ln b$ が 0 と見なし対数尤度比を算出する（高見、2003）。

次に、このリスト収録候補の見出し語を一覧し、解析の誤りを含むことが疑われるものについて、出現した文脈を確認する。その結果、1つの見出し語に所属するすべての用例が同一の見出し語に修正されることが確認された場合、修正後の見出し語をリスト収録候補に追加し、修正前の見出し語は削除する。例えば、見出し語「【常備漢字表】(合成名詞)¹³」に所属する用例はすべて正しくは「【常用漢字表】(合成名詞)」に所属するべきものである。この場合、リスト収録候補から「【常備漢字表】(合成名詞)」を削除し、代わりに「【常用漢字表】(合成名詞)」を追加する。しかし、1つの見出し語に所属する用例が複数の見出し語に分かれて修正される場合は、修正後の見出し語をリスト収録候補に追加することはせず、修正前の見出し語の削除のみおこなう。これは、修正後の見出し語の特徴度が10.83より大きいかどうか推定が困難なためである。例えば、「【はあ】(感動詞-一般)」に所属する用例は、正しくは「【ハ】(記号-一般)」「【は】(助詞-係助詞)」のいずれかに所属するべきものである。この場合、リスト収録候補から「【はあ】(感動詞-一般)」の削除のみ行い、「【ハ】(記号-一般)」「【は】(助詞-係助詞)」の追加は行わない。そして、このリスト修正作業で新たに追加された見出し語が既存の見出し語と重複する場合は、特徴度がもっとも高い見出し語を残し、その他のものは削除する¹⁴。

この修正作業を経たリスト収録候補から、次の①～⑤の条件にあてはまる見出し語を削除する。

- ① 上述の修正作業の結果、品詞が助詞・助動詞・記号類等のリスト対象外のものとなるもの
- ② 数詞
- ③ 人名・地名
例：【ボブ】、【横田選手】、【北海道】、【アフリカ大陸】
- ④ 合成名詞に近似するものとして再構成された単位のうち、語とは見なせないもの
例：【其々平行】、【全て合同】、【殆ど自覚症状】
- ⑤ 引用文や例文等の題材に由来し、当該教科での一般性が低いと考えられるもの
例：【金魚】、【菊人形】、【洗面台】(以上、中学の国語の特徴語の収録候補)、【大仏】、【格闘ゲーム】、【ハンバーガーショップ】(以上、中学の英語の特徴語の収録候補)

表 2 リスト収録候補の見出し語の異なり語数・延べ語数

		異なり語数	延べ語数
中学校	国語	739	13,957
	数学	480	10,023
	理科	718	10,472
	社会	1,357	19,474
	外国語	291	1,552
	技術家庭	784	7,231
	芸術	757	7,481
	保健体育	444	3,754
高校	国語	1,085	14,998
	数学	1,013	35,606
	理科	5,178	138,826
	社会	5,209	172,077
	外国語	301	1,125
	技術家庭	1,359	24,437
	芸術	1,831	26,736
	保健体育	686	10,173
	情報	1,064	18,830

¹³ 以下、見出し語は【 】内に語彙素、()内に品詞を示して表記する。

¹⁴ この際、度数を合計する等の調整は行わなかった。

4. リストの収録語数と教科書の語彙カバー率

以上の整備を終えたリスト収録候補の見出し語数が 500 語を超える校種・教科については、特徴度降順上位 500 位までの語をリストに収録する。この結果、校種・教科別の教科特徴語リストに収録される見出し語の異なり語数・延べ語数を表 3 に示す。また、表 1 に示した校種・教科別の対象コーパス全体での異なり語数・延べ語数に対する、表 3 の割合（語彙カバー率）を示したものが表 4 である。

表 3 教科特徴語リストの見出し語の異なり語数・延べ語数

		異なり語数	延べ語数
中学校	国語	500	11,351
	数学	381	7,741
	理科	500	8,502
	社会	500	13,443
	外国語	109	750
	技術家庭	500	6,804
	芸術	506	6,188
	保健体育	434	3,584
高校	国語	511	9,824
	数学	500	26,202
	理科	500	78,289
	社会	500	86,499
	外国語	29	168
	技術家庭	531	18,565
	芸術	500	17,456
	保健体育	500	8,977
	情報	503	12,973

表 4 教科書全体の語数に対する教科特徴語リストの見出し語の異なり語数・延べ語数の割合

		異なり語数	延べ語数
中学校	国語	8.2%	36.8%
	数学	39.3%	59.5%
	理科	20.3%	48.0%
	社会	6.6%	34.7%
	外国語	11.2%	23.9%
	技術家庭	18.9%	53.6%
	芸術	12.6%	36.2%
	保健体育	26.0%	53.2%
高校	国語	7.6%	29.2%
	数学	20.6%	56.7%
	理科	3.2%	40.6%
	社会	1.4%	29.5%
	外国語	2.1%	4.6%
	技術家庭	7.4%	43.4%
	芸術	4.7%	31.8%
	保健体育	13.3%	46.9%
	情報	13.0%	45.8%

表 4 からわかるように、語彙カバー率は校種・教科により差がある。異なり語数で見ると、リスト収録語数に上限を設けたことにより、対象コーパス全体での語数が多い校種・教科については語彙カバー率が低くなっている。また、外国語ではリスト収録候補の語数が少なかったことに加え、収録候補から主に 3.4 に述べた①⑤の条件により多くの語が削除されたため、リスト収録語数そのものが少なくなり、結果、語彙カバー率も低くなっている。

5. リストの構成

教科特徴語リストは校種・教科ごとに計 17 種作成し、本報告書の付録 CD-ROM に収録する。ファイル形式は Microsoft 社の表計算ソフト Excel の Excel 97-Excel 2003 ブック形式（.xls）とする。表 5 に例として中学校の数学の教科特徴語リストの一部を示す。

表 5 中学校の数学の教科特徴語リスト(一部)

ID	語彙素読み	語彙素	語種	品詞	活用型	説明	度数	特徴度	度数_LB_FL	レベル_LB_FL
1	アタイ	値	和	名詞-普通名詞-サ変可能			75	507.06	217	II
2	アツマリ	集まり	和	名詞-普通名詞-一般			7	34.43	57	III
3	アテハマル	当て嵌まる	和	動詞-一般	五段-ラ行-一般		7	33.15	63	III
4	アラウス	表わす	和	動詞-一般	五段-サ行		107	561.85	729	I
5	アラウス	表わす	和	動詞-一般	下一段-サ行	「表せる」	5	32.92	16	IV
6	アル	或る	和	連体詞			19	14.69	1796	I
7	イコウ	移項	漢	名詞-普通名詞-サ変可能			5	55.85	0	
8	イチ	位置	漢	名詞-普通名詞-サ変可能			20	49.09	638	I
9	イッテイ	一定	漢	名詞-普通名詞-サ変形状詞可能			13	41.08	278	II
10	イッパン	一般	漢	名詞-普通名詞-一般			18	50.28	473	I
11	イドウ	移動	漢	名詞-普通名詞-サ変可能			12	35.45	288	II
12	イフゴウ	異符号	漢+漢	合成名詞			3	33.51	0	
13	イレカエル	入れ替える	和	動詞-一般	下一段-ア行		4	20.04	31	IV
14	イロイロ	色々	和	形状詞-一般			36	132.59	577	I
15	インスウ	因数	漢	名詞-普通名詞-一般			5	55.85	0	
16	インスウブン	因数分解	漢+漢	合成名詞			10	100.91	2	V
17	ウエ	上	和	名詞-普通名詞-副詞可能			73	120.33	3783	I
18	ウヘン	右辺	漢	名詞-普通名詞-一般			4	21.6	25	IV
19	ウラガエス	裏返す	和	動詞-一般	五段-サ行		3	12.94	34	III
20	エイカク	鋭角	漢	名詞-普通名詞-一般			3	22.96	5	V

全リストが表 5 同様の構成をとる。リストの各列の詳細については、リストに添付する説明文書「教科特徴語_README.pdf」を参照してほしいが、最右列の「レベル_LB_FL」についてここで説明する。「レベル_LB_FL」は LB_FL に出現する見出し語について、その度数降順の累積使用率により I～V の 5 段階に分けたものである¹⁵。レベルの設定基準と各レベルに所属する見出し語の LB_FL における度数および異なり語数を表 6 に示す。

表 6 LB_FL におけるレベルの設定

レベル	累積使用率	度数	異なり語数
I	0 ～ 60%	158,650 ～ 348	1,108
II	～ 70%	347 ～ 126	1,689
III	～ 80%	125 ～ 33	5,778
IV	～ 90%	32 ～ 6	28,236
V	～ 100%	5 ～ 1	219,375

レベル I に所属する見出し語は LB_FL での度数が高く、レベル V になるほど度数が低くなる。つまり、レベルは「生徒が日常生活で触れる書き言葉において、頻繁に使われる程度」、言い換えれば「生徒の日常生活での馴染み度」を示す指標ともなるものである。語彙指導において「馴染み度」を把握することは重要との考えに基づき、本リストにこのレベルの情報を付与する。なお、「レベル_LB_FL」が空値の見出し語は、LB_FL での度数が 0 の語である。これはレベル V の語よりもさらに「馴染み度」の低い語と言える。

6. おわりに

以上、中学校・高校教科書から校種・教科別に教科特徴語を抽出し、整備した上で教科特徴語リストを作成した。コーパスという大規模な言語資源を活用し、客観的な指標に基づき抽出された教科特徴語に、人手による選別作業を加えることで、基本的かつ実用的なリストが作成できたと考える。本リストの基本性・実用性の検証や語彙指導の現場での活

¹⁵ 累積使用率による語彙のレベル分けについては田中（2011）を参照した。

用方法の検討については今後の課題としたい。

文献

石川慎一郎 (2008) 『英語コーパスと言語教育』, 大修館書店.

内山将夫、中條清美、山本英子、井佐原均 (2004) 「英語教育のための分野特徴単語の選定尺度の比較」, 自然言語処理, 11-3, pp.165-197.

(<http://www2.nict.go.jp/x/x161/members/mutiyama/pdf/chara.pdf> よりダウンロード可能)

小椋秀樹、小磯花絵、富士池優美、宮内佐夜香、原裕 (2010) 『特定領域研究「日本語コーパス」平成 21 年度研究成果報告書 『現代日本語書き言葉均衡コーパス』形態論情報規定集 第 3 版』.

近藤明日子 (2008a) 「中学校教科書の教科別特徴語の抽出 —理科を例として—」特定領域研究「日本語コーパス」平成 19 年度公開ワークショップ (研究成果報告会) 予稿集, pp.181-186.

近藤明日子 (2008b) 「中学校教科書の教科別特徴語の抽出」特定領域研究「日本語コーパス」言語政策班中間報告書 言語政策に役立つ, コーパスを用いた語彙表・漢字表等の作成と活用, pp.111-114.

近藤明日子 (2009) 「中学校教科書の教科特徴語の抽出と考察 —『現代日本語書き言葉均衡コーパス』の語彙との比較から—」特定領域研究「日本語コーパス」平成 20 年度公開ワークショップ (研究成果報告会) 予稿集, pp.117-122.

近藤明日子 (2010) 「検定教科書の語彙分析 —主要学習部分とその他の部分との比較から—」特定領域研究「日本語コーパス」平成 22 年度全体会議予稿集, pp.209-216.

高見敏子 (2003) 「「高級紙語」と「大衆紙語」の corpus-driven な特定法」(北海道大学) 大学院国際広報メディア研究科・言語文化部紀要, 44, pp.73-105.

(http://www.hucc.hokudai.ac.jp/~p16537/papers/Takami_2003_MLC.pdf よりダウンロード可能)

田中牧郎 (2011) 「語彙レベルに基づく重要語彙リストの作成—国語政策・国語教育での活用のために—」本報告書第 2 章第 1 節.

田中牧郎、近藤明日子、平山允子 (2011) 「教科書コーパス」本報告書第 1 章第 1 節.

Adam Kilgariff (2001) "Comparing corpora" *International Journal of Corpus Linguistics*, 6-1, pp.1-37. (<http://www.kilgariff.co.uk/Publications/2001-K-CompCorpIJCL.pdf> よりダウンロード可能)

第3章 第2節

「分類重要語彙リスト」の作成による 教科教育と語彙教育の関連づけ

田中 牧郎

1. 背景と目的

中等教育の目的の一つに、社会生活で必要となる語彙を身につけさせるということがあげられよう。この目的に添った語彙教育のあり方を検討するための基礎的な資料として、「教科書コーパス」に基づく語彙頻度表と『現代日本語書き言葉均衡コーパス』(以下、BCCWJ と称する)に基づく語彙頻度表とを比較した、「学校・社会対照語彙表」を作成した(本報告書第1章第4節参照)。

「学校・社会対照語彙表」は、学校で教えられている教科書の語彙と、社会で使われている書き言葉の語彙とを、様々な観点から比較対照できる。教科書における各教科の語彙は、その教科で扱われている知識項目を反映していると考えられる。一方、現代の書き言葉を代表できるように設計されたBCCWJの語彙は、社会で必要とされる知識項目を映し出していると考えられる。BCCWJの中でも母集団からの無作為抽出によって作成されたサブコーパスは、一般社会の書き言葉を反映しているという利点が特に高い。とりわけ、公共図書館の多くに所蔵されている図書が無作為に対象にして作成された「図書館コーパス」は、公共性の観点で価値のある本が集められているという性質から、社会で必要とされる知識項目をもっとも反映しているサブコーパスであると見てよいであろう。教科書における知識項目と社会で必要とされる知識項目とを語彙を通して比較対照することで、教科教育の現状を把握し問題点を指摘することができるのではないだろうか。

一方、国語教育の関心事の一つに語彙の教育がある。国語科では漢字指導を中心に多くの時間が語彙教育にあてられている¹。しかしながら、国語科における語彙教育の目的や方法は必ずしも十分に確立しているとは言えないのではないか。もちろん、常用漢字の読み書きができるようになる、多くの語彙を理解し使えるようになる、意味用法や語感への感覚を鋭くして文脈や場面に応じた語の使い分けができるようになるなど、重要な目的が掲げられ、そうした目的を達するために様々な活動が行われてはいる。しかし、日本語を使って社会生活を営むのに十分な語彙力を養成するための語彙教育という視点で、目的や方法が体系化されているかという点、現状はそうではないと言わざるを得ないのではないか。その現状の背景には、社会生活で必要な日本語の語彙というものが体系的に把握できていないために、目標設定や方法の体系化に具体性が伴いにくいという問題があったのではないかとと思われる²。

語彙というものは、知識項目としてとらえる場合は教科教育の対象となり、言語項目と

¹ 漢字指導は語彙教育とは区別されることが一般的かもしれないが、漢字は語を表記するのが基本的な機能であるので、広い意味では語彙教育に属すると考えられる。本報告書の第3章第4節及び第4章第4節の棚橋尚子の論考からもそのことはうかがえる。

² この点は、本特定領域の開始時点で領域代表の前川喜久雄が指摘しており、本領域で作成するコーパスができれば、国語教育の語彙の議論に相当の進歩をもたらすと見通していた(前川2006)。

してとらえる場合は国語科における語彙教育の対象になるという、二つの側面を持っている。現在の学校教育では、この二側面への目配りと関連づけが十分にできていないのではないか。確かに、教科教育で重視される語彙と国語科で重視される語彙は重ならないものが多いかもしれない。しかし、本稿でこれから見ていくように、教科で学習される知識としての語彙が、国語科の言語事項としての語彙や語彙力と密接に関連している場合も多い。語彙の二側面を有機的に関連づける視点を持つことは重要であろう。

我々が作成した「学校・社会対照語彙表」は、教科教育と語彙教育の関係を考えるのに、極めて有益な情報源となるのではないかと思われる。この語彙表をもとに、教科の語彙と社会における重要な語彙を比較対照するのに使いやすい形式に編集した「分類重要語彙リスト」を作成し、これを活用して語彙を分類・分析し、学校教育における語彙の扱い方について検討していくことが、本稿の目的である。

2. 「学校・社会対照語彙表」から「分類重要語彙リスト」へ

2. 1 「学校・社会対照語彙表」に基づく重要語彙の抽出と分類

学校教育における語彙を考える場合、個々の語がどのような意味分野に属しているかという情報を整理しておくことは有効である。「学校・社会対照語彙表」には、国立国語研究所編『分類語彙表 増補改訂版』の意味番号が収録されており、これを用いて、語の意味分野をとらえることができる。『分類語彙表 増補改訂版』は、多義語には複数の意味番号を与えているが、ここでは、別の意味番号が当たっている場合は別語と扱って分析を行う。

社会で必要とされる重要語彙を「図書館コーパス」(LB)によってとらえる場合、このコーパスにおける頻度が一定以上あるという基準でその重要語彙を抽出することが考えられる。これは、第2章第1節で語彙レベルに基づいて「重要語彙リスト」の範囲を設定した考え方と同じである。ここでは、図書館コーパスの語彙レベルが、a・b・cの三段階である22,102語を重要語彙として抽出した。複数の意味番号を持つ多義語を語義ごとに別語扱いして集計した30,693語を以下の分析対象とする。

教科書の語彙については、教科教育との関連でこれを見ていくために、教科特徴語に注目する。中学校・高等学校の教科特徴語については、第1章第2節でその考え方と抽出方法を述べ、第3章第1節で特徴語のリストを示した通りであるが、ここでは、短単位に基づいた「学校・社会対照語彙表」における特徴語彙を利用する。また、中学校・高等学校の教科書全体と図書館コーパスの書籍全体とを比較した、媒体特徴語についてもあわせて注目していくことにしたい。

2. 2 「分類重要語彙リスト」の作成

以上のような考え方で、社会的な重要語彙を教科特徴語・媒体特徴語との関連で一覧できるように、意味番号(分類項目)順に整理したリストを作成し、「分類重要語彙リスト」と称することにした。名称に含む「分類」には、意味、重要度レベル、教科、媒体の4つの観点で類別可能なものという意図を込めている。本節の末尾に「別表」として冒頭から4つの細目の部分を掲載した。この別表を例に表の見方を説明しよう。

まず、分類項目番号と分類項目名の後の「レベル」の列で図書館コーパスのa・b・cの三つの重要度に分類する。そしてレベルごとに、「国語」「数学」「理科」「社会」「外国語」「技術家庭」「芸術」「保健体育」「情報」の9教科それぞれの教科書の特徴語をそれぞれの列に

リストアップする。この特徴語は、中学校高等学校の各教科の教科書の語彙頻度を図書館コーパス全体のそれと比較して抽出したものである（抽出の基準は第1章第4節を参照）。その次の「教科書」の列には、教科書全体と書籍（図書館コーパス）全体で語彙頻度を比較した場合の教科書の特徴語、「書籍」の列は、同じ方式で比較した書籍の特徴語を示している。「無特徴」という列には、どの教科の特徴語にもならず、教科書や書籍の特徴語にもならないものを示す。以上の「国語」から「無特徴」までの12の列のいずれかに、a・b・cの3つのレベルに属するすべての語が配されることになる。なお、複数の教科の特徴語になったり、教科と媒体の双方で特徴語となったりする場合もあるが、その場合は同じ語が複数の列にあがることになる。

別表のうち冒頭の「1.1000 事柄」の分類項目の語彙を見てみよう。この分類項目には全体にレベルaの語彙が多く、基本的な意味を担う分類項目だと思われるが、教科特徴語にもレベルaの語彙が多い。国語の特徴語には、「物」「事」「物事」「出来事」など、事柄をありのままとらえる場合に骨組みとなる意味を持つと考えられる語が多く、数学の特徴語には「分（ぶん）」「例」のように、事柄を数量や事例としてとらえる場合に中心的な役割を担うと考えられる語が目立つ。このように、「分類重要語彙リスト」を用いることで、教科ごとの特徴語の様子を見わたして比較していくことができる。

例えば、国語の特徴語が、他教科の特徴語にもなっているかどうかを見ると、「事柄」は数学・外国語の特徴語になっており、「項目」は技術家庭・保健体育・情報のそれになっており、ともに多くの教科でよく使われことが分かる。一方、「物」は芸術、「事」は保健体育、「出来事」は外国語のそれぞれもう一教科のみで特徴語になっており、「物事」は他のどの教科でも特徴語になっていない。

次にレベルbの教科特徴語に目を移すと、「事物」「事象」「事変」といった硬い漢語が教科の特徴語となっていることがわかる。これらの語彙は日常会話などふだんの生活を通じて習得されるものではなく、読み書きを通して学習して身につけることができるようになるものであると思われる。その学習の機会が、「事物」は国語で、「事象」は数学や理科で、「事変」は社会で多く与えられるのではないかということが、この表から推測できる。

別表の右の方にある「書籍」の列には書籍の特徴語が多く抽出されているが、レベルaだけでなくbやcにも多く、教科書にはあまり使われない重要語彙が広く存在していることが確かめられる。「無特徴」の列に掲出される語は、教科書と書籍のどちらかに偏って出現するものではなく、また教科書では特定の教科に偏っていないものである。

このように、「分類重要語彙リスト」によれば、図書館コーパスによる重要度のレベルa・b・cの語彙すべてについて、『分類語彙表 増補改訂版』の意味番号順に教科や媒体に特徴的な語彙を対照しながら見ていくことができるのである。なお、『分類語彙表 増補改訂版』に収録されていない語や、コーパスに対する形態素解析結果と『分類語彙表 増補改訂版』との関連づけができなかった語については「分類重要語彙リスト」には収録されていない。

3. 「分類重要語彙リスト」の概観

3. 1 品詞の観点から

2で作成した「分類重要語彙リスト」を概観すると、重要語彙がどの意味分類項目や教科に多いのか、また少ないのかといった分布の様子を知ることができる。

表1・表2は、『分類語彙表 増補改訂版』の品詞分類である「1.体の類」「2.用の類」「3.

相の類「4.その他」の分類枠によって、重要語彙 30,693 語（多義語を別語として扱った語数）を分類した結果の語数（表 1）と比率（表 2）を示したものである。

表 1 品詞別の特徴語の概観（語数）

	全体	教科特徴語										媒体特徴語	
		国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報	教科書	書籍	
1. 体の類	18314	1398	760	2069	2831	782	1885	1331	1178	1071	3380	8102	
2. 用の類	4970	284	238	433	465	173	402	278	306	188	469	2889	
3. 相の類	2802	100	74	180	183	71	147	178	121	83	208	1902	
4. その他	273	12	17	14	7	7	1	8	5	13	12	242	
小計	26359	1794	1089	2696	3486	1033	2435	1795	1610	1355	4069	13135	
該当なし	4334	187	49	170	397	111	91	218	55	66	332	2474	
合計	30693	1981	1138	2866	3883	1144	2526	2013	1665	1421	4401	15609	

表 1 における「小計」は、1～4.の全品詞を合算した語数、「該当なし」は、『分類語彙表増補改訂版』に掲載されていない語、もしくは技術的な事情によって関連づけができなかった語である。また「合計」の行には、「小計」と「該当なし」とを合算した語数を示した。表 2 に示す比率は、1～4.のそれぞれの品詞に属する全体の語数の中で各教科や各ジャンルの特徴語数が占める比率を示したものである。「小計」「該当なし」「合計」についても、同様の方法で計算した比率を示した。

表 2 品詞別の特徴語の概観（比率）

	全体	教科特徴語										媒体特徴語	
		国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報	教科書	書籍	
1. 体の類	100.0%	<u>7.6%</u>	4.1%	<u>11.3%</u>	<u>15.5%</u>	4.3%	<u>10.3%</u>	<u>7.3%</u>	<u>6.4%</u>	<u>5.8%</u>	18.5%	44.2%	
2. 用の類	100.0%	<u>5.7%</u>	4.8%	<u>8.7%</u>	<u>9.4%</u>	3.5%	<u>8.1%</u>	<u>5.6%</u>	<u>6.2%</u>	3.8%	9.4%	58.1%	
3. 相の類	100.0%	3.6%	2.6%	<u>6.4%</u>	<u>6.5%</u>	2.5%	<u>5.2%</u>	<u>6.4%</u>	4.3%	3.0%	7.4%	67.9%	
4. その他	100.0%	4.4%	<u>6.2%</u>	<u>5.1%</u>	2.6%	2.6%	0.4%	2.9%	1.8%	4.8%	4.4%	88.6%	
小計	100.0%	<u>6.8%</u>	4.1%	<u>10.2%</u>	<u>13.2%</u>	3.9%	<u>9.2%</u>	<u>6.8%</u>	<u>6.1%</u>	<u>5.1%</u>	15.4%	49.8%	
該当なし	100.0%	4.3%	1.1%	3.9%	9.2%	2.6%	2.1%	5.0%	1.3%	1.5%	7.7%	57.1%	
合計	100.0%	6.5%	3.7%	9.3%	12.7%	3.7%	8.2%	6.6%	5.4%	4.6%	14.3%	50.9%	

表 2 において各教科の特徴語の占める比率を見てみよう。数値の入った各セルにおいて、10%を超えるセルに網掛け、5%を超える数値に下線をそれぞれ施した。「小計」で 10%を超えるのは社会と理科であり、この 2 教科は「1.体の類」の数値も 10%を超えている。これらは特徴語が特に多い教科だと見ることができる。また、技術家庭も「小計」では 10%に達していないものの「1.体の類」では 10%を超えており、特徴語がかなり多いと見ることができる。さらに、「小計」では、理科・社会・技術家庭のほか国語・芸術・保健体育・情報において「小計」が 5%を超えており、これら 3 教科も特徴語が比較的多い教科と言ってよいであろう。一方、外国語・数学は、「小計」が 5%に達しておらず、品詞別でも 5%を超える場合があまりなく、これらは特徴語が少ない教科とすることができる。

同じく表 2 において媒体特徴語の比率を見てみよう。全体に書籍特徴語の方の数値が圧

倒的に高いが、これは書籍の方がコーパスの規模が大きく語彙の範囲も広いことによるものであると考えられる。そうした全体的な傾向の中で、「1.体の類」は教科書特徴語の数値が比較的高い点が注目される。

3. 2 意味分野の観点から

次に、意味分野による教科特徴語・ジャンル特徴語の分布を確認していこう。表3・表4は、『分類語彙表 増補改訂版』の意味分野である「.1 抽象的關係」「.2 人間活動の主体」「.3 精神および行為」「.4 生産物および用具」「.5 自然物および自然現象」という分類枠によって、重要語彙 30,693 語を分類した結果の語数（表3）と比率（表4）を示したものである。この表の数値は「1.体の類」「2.用の類」「3.相の類」それぞれの各意味分野（「2.用の類」「3.相の類」には「2.人間活動の主体」「4.生産物および用具」はない）を合算したものであり、「4.その他」は意味分野に分けられていないので表1・表2の数値をそのまま示した。「小計」「該当なし」「合計」等の算出方法は、表1・表2の場合と同様である。

表3 意味分野別の特徴語の概観（語数）

	全体	教科特徴語										媒体特徴語	
		国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報	教科書	書籍	
.1 抽象的關係	9196	552	680	1316	1292	330	897	666	600	608	1669	4335	
.2 人間活動の主体	2321	168	27	50	488	89	76	115	78	41	304	1282	
.3 精神および行為	9684	815	274	514	1335	499	839	623	603	557	1272	5011	
.4 生産物および用具	2051	115	61	219	157	63	304	196	84	92	280	965	
.5 自然物および自然現象	2834	132	30	583	207	45	318	187	240	44	532	1300	
4. その他	273	12	17	14	7	7	1	8	5	13	12	242	
小 計	26359	1794	1089	2696	3486	1033	2435	1795	1610	1355	4069	13135	
該当なし	4334	187	49	170	397	111	91	218	55	66	332	2474	
合 計	30693	1981	1138	2866	3883	1144	2526	2013	1665	1421	4401	15609	

表4 意味分野別の特徴語の概観（比率）

	全体	教科特徴語										媒体特徴語	
		国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報	教科書	書籍	
.1 抽象的關係	100.0%	<u>6.0%</u>	<u>7.4%</u>	14.3%	14.0%	3.6%	<u>9.8%</u>	<u>7.2%</u>	<u>6.5%</u>	<u>6.6%</u>	18.1%	47.1%	
.2 人間活動の主体	100.0%	<u>7.2%</u>	1.2%	2.2%	21.0%	3.8%	3.3%	<u>5.0%</u>	3.4%	1.8%	13.1%	55.2%	
.3 精神および行為	100.0%	<u>8.4%</u>	2.8%	<u>5.3%</u>	13.8%	<u>5.2%</u>	<u>8.7%</u>	<u>6.4%</u>	<u>6.2%</u>	<u>5.8%</u>	13.1%	51.7%	
.4 生産物および用具	100.0%	<u>5.6%</u>	3.0%	10.7%	<u>7.7%</u>	3.1%	14.8%	<u>9.6%</u>	4.1%	4.5%	13.7%	47.1%	
.5 自然物および自然現象	100.0%	4.7%	1.1%	20.6%	<u>7.3%</u>	1.6%	11.2%	<u>6.6%</u>	<u>8.5%</u>	1.6%	18.8%	45.9%	
4. その他	100.0%	4.4%	<u>6.2%</u>	<u>5.1%</u>	2.6%	2.6%	0.4%	2.9%	1.8%	4.8%	4.4%	88.6%	
小 計	100.0%	<u>6.8%</u>	4.1%	10.2%	13.2%	3.9%	<u>9.2%</u>	<u>6.8%</u>	<u>6.1%</u>	<u>5.1%</u>	15.4%	49.8%	
該当なし	100.0%	4.3%	1.1%	3.9%	9.2%	2.6%	2.1%	5.0%	1.3%	1.5%	7.7%	57.1%	
合 計	100.0%	6.5%	3.7%	9.3%	12.7%	3.7%	8.2%	6.6%	5.4%	4.6%	14.3%	50.9%	

表4において各教科の特徴語の占める比率を見ていきたいが、表2の場合と同じく、数値の入った各セルにおいて、10%を超えるセルに網掛け、5%を超える数字に下線を施した。

表 2 の観察から特徴語の多い教科であることが判明した社会・理科および技術家庭を見ると、意味分野によって数値の高低がずいぶん異なっていることが分かる。社会では「2 人間活動の主体」の数値がとりわけ高く、次いで「1 抽象的關係」「3 精神および行為」が高い。一方、理科では、「2 人間活動の主体」「3 精神および行為」の数値は低く、「5 自然物および自然現象」が特に高く、「1 抽象的關係」「4 生産物および用具」がこれに次ぐ。また、技術家庭は理科の数値に近いが、「4 生産物および用具」が最も高いなど、理科とは異なる面もある。表 2 から、特徴語が比較的多い教科とされた国語・芸術・保健体育・情報について表 4 の数値を見ると、国語・芸術は意味分野による相違が目立たないのに対して、保健体育では「2 人間活動の主体」と「4 生産物および用具」が低く、情報ではほかに「5 自然物および自然現象」も低い。特徴語が少ない、外国語・数学については、外国語はどの意味分野の数値も低いのに対して、数学は「1 抽象的關係」の数値がやや高い。このように、教科によってどの意味分野の特徴語が多いか少ないかは、大きく異なっているのである。

次に、表 4 において教科書と書籍の媒体特徴語の比率を見ると、全般に書籍特徴語の数値が高いのは、表 2 の品詞の場合と同様だが、「1 抽象的關係」「5 自然物および自然現象」の 2 つの意味分野で、教科書特徴語の数値が比較的高くなっている。

3. 3 大分類の観点から

『分類語彙表 増補改訂版』は、品詞と意味分野を組み合わせ、「1.1 体の類 抽象的關係」のようにして大分類の枠組を設定している。表 5・表 6 は、その大分類の枠組にしたがって、各教科の特徴語と各ジャンルの特徴語とを、表 1～表 4 と同じ要領でまとめたものである。

表 5 大分類別の特徴語の概観（語数）

	全体	教科特徴語									媒体特徴語	
		国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	教科書	書籍
1.1 体の類_抽象的關係	5500	382	497	948	947	229	633	453	396	464	1278	2172
1.2 人間活動の主体	2321	168	27	50	488	89	76	115	78	41	304	1282
1.3 人間活動-精神及び行為	6282	620	153	333	1051	361	590	421	406	437	1032	2821
1.4 生産物および用具	2051	115	61	219	157	63	304	196	84	92	280	965
1.5 自然物および自然現象	2160	113	22	519	188	40	282	146	214	37	486	862
2.1 用の類_抽象的關係	2214	107	119	248	212	57	173	122	131	89	243	1190
2.3 用の類_精神および行為	2416	165	112	150	240	114	210	138	159	96	198	1499
2.5 用の類_自然現象	340	12	7	35	13	2	19	18	16	3	28	200
3.1 相の類_抽象的關係	1482	63	64	120	133	44	91	91	73	55	148	973
3.3 相の類_精神および行為	986	30	9	31	44	24	39	64	38	24	42	691
3.5 相の類_自然現象	334	7	1	29	6	3	17	23	10	4	18	238
4. その他	273	12	17	14	7	7	1	8	5	13	12	242
小 計	26359	1794	1089	2696	3486	1033	2435	1795	1610	1355	4069	13135
該当なし	4334	187	49	170	397	111	91	218	55	66	332	2474
合 計	30693	1981	1138	2866	3883	1144	2526	2013	1665	1421	4401	15609

表 6 大分類別の特徴語の概観（比率）

	全体	教科特徴語									媒体特徴語		
		国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	教科書	書籍	
1.1	体の類_抽象的關係	100.0%	<u>6.9%</u>	<u>9.0%</u>	17.2%	17.2%	4.2%	11.5%	<u>8.2%</u>	<u>7.2%</u>	<u>8.4%</u>	23.2%	39.5%
1.2	人間活動の主体	100.0%	<u>7.2%</u>	1.2%	2.2%	21.0%	3.8%	3.3%	<u>5.0%</u>	3.4%	1.8%	13.1%	55.2%
1.3	人間活動_精神及び行為	100.0%	<u>9.9%</u>	2.4%	<u>5.3%</u>	<u>16.7%</u>	<u>5.7%</u>	<u>9.4%</u>	<u>6.7%</u>	<u>6.5%</u>	<u>7.0%</u>	16.4%	44.9%
1.4	生産物および用具	100.0%	<u>5.6%</u>	3.0%	10.7%	<u>7.7%</u>	3.1%	14.8%	<u>9.6%</u>	4.1%	4.5%	13.7%	47.1%
1.5	自然物および自然現象	100.0%	<u>5.2%</u>	1.0%	24.0%	<u>8.7%</u>	1.9%	13.1%	<u>6.8%</u>	<u>9.9%</u>	1.7%	22.5%	39.9%
2.1	用の類_抽象的關係	100.0%	4.8%	<u>5.4%</u>	11.2%	<u>9.6%</u>	2.6%	<u>7.8%</u>	<u>5.5%</u>	<u>5.9%</u>	4.0%	11.0%	53.7%
2.3	用の類_精神および行為	100.0%	<u>6.8%</u>	4.6%	<u>6.2%</u>	<u>9.9%</u>	4.7%	<u>8.7%</u>	<u>5.7%</u>	<u>6.6%</u>	4.0%	8.2%	62.0%
2.5	用の類_自然現象	100.0%	3.5%	2.1%	<u>10.3%</u>	3.8%	0.6%	<u>5.6%</u>	<u>5.3%</u>	4.7%	0.9%	8.2%	58.8%
3.1	相の類_抽象的關係	100.0%	4.3%	4.3%	<u>8.1%</u>	<u>9.0%</u>	3.0%	<u>6.1%</u>	<u>6.1%</u>	4.9%	3.7%	10.0%	65.7%
3.3	相の類_精神および行為	100.0%	3.0%	0.9%	3.1%	4.5%	2.4%	4.0%	<u>6.5%</u>	3.9%	2.4%	4.3%	70.1%
3.5	相の類_自然現象	100.0%	2.1%	0.3%	<u>8.7%</u>	1.8%	0.9%	<u>5.1%</u>	<u>6.9%</u>	3.0%	1.2%	5.4%	71.3%
4.	その他	100.0%	4.4%	<u>6.2%</u>	<u>5.1%</u>	2.6%	2.6%	0.4%	2.9%	1.8%	4.8%	4.4%	88.6%
	小 計	100.0%	<u>6.8%</u>	4.1%	<u>10.2%</u>	<u>13.2%</u>	3.9%	<u>9.2%</u>	<u>6.8%</u>	<u>6.1%</u>	<u>5.1%</u>	15.4%	49.8%
	該当なし	100.0%	0.7%	0.2%	0.6%	1.5%	0.4%	0.3%	0.8%	0.2%	0.3%	1.3%	9.4%
	合 計	100.0%	<u>7.5%</u>	4.3%	<u>10.9%</u>	<u>14.7%</u>	4.3%	<u>9.6%</u>	<u>7.6%</u>	<u>6.3%</u>	<u>5.4%</u>	16.7%	59.2%

表 6 では、表 2・表 4 と同じように、10%を超えるセルに網掛け、5%を超える数字に下線を施した。例えば、特徴語が特に多い理科と社会を見ると、「1.1 体の類_抽象的關係」はどちらの教科も多く「1.2 人間活動の主体」「1.3 人間活動_精神及び行為」は社会に多いが理科には少なく、「1.4 生産物および用具」「1.5 自然物および自然現象」「2.5 用の類_自然現象」は理科に多いが社会に少ないことが分かる。また、特徴語の多さがほぼ同程度（「小計」でともに 6.8%）の国語と芸術とを比較すると、国語は 1.1～1.5 の体の類の数値が比較的高いものに対して、芸術は「4.その他」を除くすべての大分類で比較的高いという相違も興味深い。

3. 4 中項目の観点から—「1.1 体の類 抽象的關係」について—

『分類語彙表 増補改訂版』は、各大分類の下位に、それぞれ 5 から 10 の中項目を立てている。この中項目の状況を考察することで、語彙と教科や媒体の関係について、より具体的に把握できると考えられる。以下に『分類語彙表 増補改訂版』の冒頭の大分類である「1.1 体の類 抽象的關係」に属する中項目を事例に、その様相を見てみよう。

「1.1 体の類 抽象的關係」の大分類の下位にある中項目は、「1.10 事柄」「1.11 類」「1.12 存在」「1.13 様相」「1.14 力」「1.15 作用」「1.16 時間」「1.17 空間」「1.18 形」「1.19 量」の 10 項目である。これらの中項目について、表 1～表 6 と同じ要領で特徴語の数と比率を整理したものが、表 7・表 8 である。「計」の行は、この 10 項目を合算したものの、すなわち「1.1 体の類 抽象的關係」の全体の語数と比率を示した。

表7 中項目（1.1体の類_抽象的關係）の特徴語（語数）

		全体	教科特徴語									媒体特徴語	
			国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報	教科書	書籍
1.10	事柄	194	22	19	21	15	9	18	13	11	21	32	109
1.11	類	443	36	35	74	74	23	55	46	43	51	113	167
1.12	存在	302	24	13	35	78	9	50	18	24	29	74	88
1.13	様相	473	27	22	59	52	11	69	43	43	29	80	210
1.14	力	74	0	2	18	21	0	11	1	11	5	27	29
1.15	作用	1290	46	56	249	290	42	145	69	99	97	318	409
1.16	時間	785	80	28	51	129	49	42	50	37	34	107	392
1.17	空間	705	40	75	146	126	29	52	61	26	33	179	287
1.18	形	224	14	30	56	18	10	25	30	8	19	57	71
1.19	量	1010	93	217	239	144	47	166	122	94	146	291	410
	計	5500	382	497	948	947	229	633	453	396	464	1278	2172

表8 中項目（1.1体の類_抽象的關係）の特徴語（比率）

		全体	教科特徴語									媒体特徴語	
			国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報	教科書	書籍
1.10	事柄	100.0%	11.3%	9.8%	10.8%	7.7%	4.6%	9.3%	6.7%	5.7%	10.8%	16.5%	56.2%
1.11	類	100.0%	8.1%	7.9%	16.7%	16.7%	5.2%	12.4%	10.4%	9.7%	11.5%	25.5%	37.7%
1.12	存在	100.0%	7.9%	4.3%	11.6%	25.8%	3.0%	16.6%	6.0%	7.9%	9.6%	24.5%	29.1%
1.13	様相	100.0%	5.7%	4.7%	12.5%	11.0%	2.3%	14.6%	9.1%	9.1%	6.1%	16.9%	44.4%
1.14	力	100.0%	0.0%	2.7%	24.3%	28.4%	0.0%	14.9%	1.4%	14.9%	6.8%	36.5%	39.2%
1.15	作用	100.0%	3.6%	4.3%	19.3%	22.5%	3.3%	11.2%	5.3%	7.7%	7.5%	24.7%	31.7%
1.16	時間	100.0%	10.2%	3.6%	6.5%	16.4%	6.2%	5.4%	6.4%	4.7%	4.3%	13.6%	49.9%
1.17	空間	100.0%	5.7%	10.6%	20.7%	17.9%	4.1%	7.4%	8.7%	3.7%	4.7%	25.4%	40.7%
1.18	形	100.0%	6.3%	13.4%	25.0%	8.0%	4.5%	11.2%	13.4%	3.6%	8.5%	25.4%	31.7%
1.19	量	100.0%	9.2%	21.5%	23.7%	14.3%	4.7%	16.4%	12.1%	9.3%	14.5%	28.8%	40.6%
	計	100.0%	6.9%	9.0%	17.2%	17.2%	4.2%	11.5%	8.2%	7.2%	8.4%	23.2%	39.5%

表8において各教科の特徴語が占める比率を見ていこう。各数値が20%を超えるセルに濃い網掛け、10%を超えるセルに薄い網掛けを施した。「1.14 力」において社会と理科の特徴語が目立って多く、「1.12 存在」「1.15 作用」では社会の特徴語の多さが際立ち、「1.17 空間」「1.18 形」「1.19 量」で理科の特徴語が特に多いことなどが分かる。ほかにも教科によってどの中項目に特徴語が多いのか少ないのかが異なっており、教科と中項目との関連について様々な考察ができそうである。

媒体の特徴語を見ると、教科書特徴語の数値が比較的高い中項目として「1.14 力」「1.19 量」などが指摘でき、書籍特徴語の数値が高い中項目として「1.10 事柄」「1.16 時間」があげられる。書籍特徴語が多い中項目は国語の特徴語が多いという対応があることにも気付かれる。

中項目は意味分野をより具体的に細分したものであり、どの中項目が担う意味概念がどの教科と関わりが深いかなどについて、ひとつひとつ確認していくことができる。

3.5 まとめ

以上、『分類語彙表 増補改訂版』の分類枠に沿って教科や媒体の特徴語を概観してきた。その結果、教科によって特徴語の多い少ないに大きな差異があること、教科によってどの品詞やどの意味分野（大分類・中項目）に特徴語が多いか少ないかが大きく異なることが、明らかになった。そして、各教科について意味分野との関わりも詳しく見ていくことで教科の語彙の特徴を確認していくことができることも分かった。以下では、中項目の内部に分け入ってさらに詳しく考察を進めていきたい。

4. 「分類重要語彙リスト」の中項目における語彙の具体相

— 「1.14 力」の語彙の分析—

4.1 「1.14 力」の語彙の概要

3の表7・表8で示した「分類重要語彙リスト」の大分類「1.1 体の類・抽象的關係」における「1.14 力」の中項目を例に、語彙の具体的様相を観察してみよう。「1.14 力」は、所属語数が最も少なく、その中項目の全体を見わたすことが比較的容易であり、同時に個々の語の観察も詳しく行えるため、語彙の様相がとらえやすいのではないかと考えられる。また、全体に特徴語が最も多い教科である理科と社会の両科目に特徴語が同程度分布し、典型例と見なすこともできそうである。表9は、「分類重要語彙リスト」の当該の中項目の部分を示したものである。

表9 分類重要語彙リスト (1.14 力)

分類項目	レベル	教科特徴語									媒体特徴語		無特徴	
		国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報	教科書	書籍		
1.1400	力	a	—	—	力、エネルギー	エネルギー、実力	—	エネルギー	—	力、エネルギー	—	力、エネルギー	パワー、強力、実力、強烈	—
		b	—	—	—	主力	—	—	—	全力	—	—	威力、迫力、全力	自力、無力
		c	—	—	強、弱	総力、弱体	—	—	—	—	—	強、総力、弱	パンチ、魔力、非力（ひりき）、非力（ひりよく）、脆弱、痛烈、激烈、劇烈、熾烈	無敵、最強、他力、余力
1.1401	弾力・動力・圧力など	a	—	入力	抵抗、入力、圧力	抵抗、圧力	—	抵抗、入力	—	抵抗、血圧	入力	抵抗、入力、圧力	血圧	—
		b	—	気圧	反発、摩擦、重力、動力、電力、気圧	反発、摩擦、動力、電力、圧迫	—	動力、電力、出力	—	圧迫	出力	反発、摩擦、重力、動力、電力、出力、圧迫、気圧	—	—
		c	—	—	引力、火力、高圧、電圧	入力、火力	—	弾力、火力、電圧	弾力	—	電圧	引力、火力、高圧、電圧	応力、テンション、眼圧	—

1.1402	物力・権力・体力など	a	—	—	—	権力、勢力	—	能力	—	能力、体力	能力	権力、勢力、体力	手、暴力	—
		b	—	—	気力	兵力、戦力、武力	—	労力	—	活力、視力	—	武力	精力、視力	—
		c	—	—	—	国力	—	—	—	筋力	人手	国力、筋力	資力、腕力	—
1.1403	勢い	a	—	—	—	—	—	—	—	—	—	—	勢い	—
		b	—	—	—	—	—	—	—	—	—	—	威勢、攻勢	—
		c	—	—	—	優勢	—	—	—	—	—	優勢	—	劣勢

表 9 について、四つの分類項目ごとに横に見ていくと、それぞれ次のような傾向が指摘できる。

- (1) 「1.1400 力」は、教科特徴語や教科書特徴語が少なく、書籍特徴語が多い。
- (2) 「1.1401 弾力・動力・圧力など」は、教科特徴語や教科書特徴語が多く、書籍特徴語が少ない。
- (3) 「1.1402 物力・権力・体力など」は、教科特徴語や教科書特徴語も、書籍特徴語も、どちらもある程度見られる。
- (4) 「1.1403 勢い」は、全体として語彙が多くないが、教科特徴語や教科書特徴語が非常に少ない。

このように、分類項目によって、特徴語の様相はかなり大きな相違があることが分かる。

どのような意味を表す語彙かによって、使われやすい媒体や教科に特徴が現れやすいことをものがたてよう。3 で、語彙全体を広く見わたしたときに見られた意味分野によって教科の特色があるという事実は、中項目という狭い範囲の語彙の内部でも確かめられるのである。以下、書籍特徴語が多い「1.1400 力」と、教科特徴語や教科書特徴語が多い「1.1401 弾力・動力・圧力など」の 2 つの分類項目を取り上げて分析しよう。

4. 2 「1.1400 力」の分類項目

4. 2. 1 概要

この分類項目には、力そのものを意味するもっとも基本的な語や、1.1401 以下の 3 つの分類項目には分類しにくい、ある性質を帯びた力や力の性質を意味する語が属している。そして、この項目には、教科特徴語や教科書特徴語が少なく、書籍特徴語が多いという顕著な傾向がある。

4. 2. 2 レベル a の教科特徴語・教科書特徴語

まず、教科特徴語・教科書特徴語の方から見ていこう。教科や教科書の特徴語は、全体の語数は少ないものの、レベル a に力そのものを表すもっとも基本的な語だと考えられる、「力 (ちから)」「エネルギー」の 2 語があり、ともに複数の教科で特徴語となっていることが目を引く。力についての根本概念を表す語彙が教科で学習される語彙であるということが確認できる。このうち、特徴語となる教科が 4 つともっとも多い「エネルギー」は、この分類項目の重要語彙のなかで、おそらく教育上最も重要な語であろう。この語の使われ方について具体的に見てみよう。

まず、この語の社会一般での意味を確認するために、国語辞典の記述を見よう。

エネルギー

- ①物事をなしとげる気力・活力。精力。「仕事で一を消耗する」「若い一」
- ②物体が物理的な仕事をするのできる能力。力学的エネルギー（運動エネルギーと位置エネルギー）のほか、化学・電磁気・熱・光・原子などの各エネルギーがある。さらに相対性理論によれば、質量そのものもエネルギーの一形態である。
- ③「エネルギー資源」の略。

エネルギー資源

産業・運輸・消費生活などに不可欠な動力源。石炭・石油・天然ガス・水力・原子力・太陽熱など。
(『デジタル大辞泉』小学館)

「エネルギー」の語義③については略されない語形である「エネルギー資源」の記述もあげた。『デジタル大辞泉』以外の辞書でも、だいたいこの3つの語義をあげている。図書館コーパスでこの語の用例を調べても、ほぼこの3つの意味に分けられる。外来語でありながら、日本語に十分定着しており、意味の輪郭が明確になっている語だと考えられる。

次に「エネルギー」という語が特徴語となっている教科（理科・社会・技術家庭・保健体育）での教科書の用例を確認していきたい。「分類重要語彙リスト」から教科書の用例を直接引くことはできないが、BCCWJの非母集団サブコーパスに取り込んで公開されている教科書データを検索することで、教科書の用例の一部を見ることができる³。

まず、理科では、「運動エネルギー」「位置エネルギー」「放射エネルギー」「イオン化エネルギー」など、物理的な仕事を可能にする力としての意味、すなわち国語辞典の②の意味で幅広く使われている。これは自然科学の学術用語を基盤に持つ使われ方である。一方、社会では、「電力などの二次エネルギー」「再生可能なエネルギー」「省エネルギー」といった、生活や生産活動に必要な動力資源の意味で多彩に使われており、これは国語辞典の③の意味すなわち「エネルギー資源」にあたる。また、保健体育では、「単位時間に発揮されるエネルギー量」のように②と思われる例のほか、「自分の関心のあることにエネルギーを注いでいる」といった、人の身体や心の活力の意味、すなわち国語辞典の①の意味の例も目に付く。そして、技術家庭では、「電気・ガス・石油などのエネルギー」「わたしたちに必要なエネルギーのとり方」のように③と①の二つの意味がともに多い。このように、「エネルギー」という語のカバーするいくつかの意味を教科が分担するようにして、学校での教科学習に取り込まれていると見られる。

以上の語義と教科の関係をまとめると次のようになる。

語義① 保健体育、技術家庭

語義② 理科、保健体育

語義③ 社会、技術家庭

下線は、学習指導要領にその語義でのエネルギーが取り上げられているものである。自然科学における原理や現象を理科で学ばせ、エネルギー資源が社会でどのような役割を果たし、実際の生活にどのように役立っているのかは社会と技術家庭で学ばせるのが指導要領の考え方であると思われる。教科書の語彙の実態もそれを反映しているが、保健体育のように指導要領に記述がない教科で特徴語になっている場合や、語義①のように指導要領

³ 「分類重要語彙リスト」の基本情報となる教科特徴語や媒体特徴語の抽出に用いた「教科書コーパス」は、教科書や教材作品の著作権者との合意によりその全体は公開できない。教科書コーパスの一部のデータをBCCWJの非母集団サブコーパスに含めて公開している。

に記述がない語彙が教科特徴語になっている場合もある。語彙の実態は、このように指導要領と対応する場合もあれば、対応しない場合もあるのである⁴。

語義①は一般語としての意味、語義②は科学の専門用語としての意味、語義③は社会的に話題になりやすい意味というように性格づけを行うことができそうであり、生徒はこれらの語義のすべてを教科書で経験することになる。生徒は教科書や教科の学習を通してこうした「エネルギー」という言葉に繰り返し触れるわけだが、その過程を経て「エネルギー」という語が表す概念がどのようなものであるかについて自らの中に確立させていくのではないかと考えられる。エネルギーをどう教えるかについては、原理や具体的現象としては、理科・社会・技術家庭の教科教育論の中で考えられてきたと思われるが、これを語彙として見ることによって、より広い視野から概念のありようや生徒の学習機会の広がりについて見わたすことが可能になり、これまでには見えにくかったことが見えてくるのではないかと思う。

4. 2. 3 レベルb・cの教科特徴語・教科書特徴語

「1.1400 力」の分類項目には、レベルb・cにおいて理科と社会の特徴語となっている語がいくつかある。まず、「強」「弱」が理科の特徴語になっているが、「強塩基」「弱酸性」など、化学物質にかかわる複合語の構成要素として接頭辞的に用いられる例が多い。また、社会の特徴語になっている「主力」「総力」は、「アメリカを主力とする連合軍」「総力戦」といった軍事の分野で多用され、同じく「弱体」は「経済基盤の弱体なイタリア」「政権の弱体化」など国家の分野で用いられている。これらの語について国語辞典には次のようにある（いずれも『デジタル大辞泉』）。

きょう【強】

① [名] 強いこと。また、強いもの。⇔弱。

② [接尾] 数量を表す語に付いて、実際はその数よりも少し多いことを表す。数の端数を切り捨てたときに用いる。「五キロ―」「九割―」⇔弱

じゃく【弱】

① [名] よわいこと。また、よわいもの。「こたつの目盛りを―にする」「―アルカリ性」⇔強。

② [接尾] 端数を切り上げたとき、数を表す語の下に付けて用いる。「五〇〇人―の聴衆」「二〇万円―の給料」⇔強

しゅりよく【主力】

① 出せる力のうちのおもな部分。おもな力。「語学に―を注ぐ」

② 中心となって力を発揮するもの。主要な戦力・勢力。「チームの―選手」「当社の―商品」

そうりよく【総力】

集団・国家などの全体の力。また、あるかぎりの力。「―をあげて取り組む」「―を結集する」

じゃくたい【弱体】

⁴ 「学校・社会対照語彙表」（第1章第4節参照）で各教科における「エネルギー」の出現頻度を見ると、理科が桁違いに多く、社会と技術家庭もかなり多いが、それらに比較すれば保健体育は多くない。指導要領に取り上げられるような語彙はもちろん重要だが、そうでない語彙や語義についても重要なものは多いと考えられ、「分類重要語彙リスト」にはそのようなものも多くあがっている。

[名・形動]

① 弱いからだ。

② 組織としての体制や体質が弱いこと。また、そのさま。「一な営業部門」

多くの語が2つの意味をあげる中で、「総力」のみが単一の意味をあげるが、これも「また」の前後で別の意味に分け、「①あるかぎりの力、②集団・国家などの全体の力」のように書き換えることもできるだろう。そうすれば、すべて二義の語ということになる。「強」と「弱」について理科で学習する意味と国語辞典の意味区分とを突き合わせると、理科で用いられる意味はともに国語辞典の①の意味である「強いこと」「よわいこと」に相当すると考えられる。国語辞典に反映する、強いこと・弱いことという一般的な意味のうち、特定の分野に結びついた意味用法で使われた例が教科書には頻繁に登場していると考えられる。「主力」「総力」「弱体」についても、国語辞典が記す2つの意味のうち一方の意味（いずれも②の意味。「総力」は書き換えた②の意味）が教科書に用いられ、特定の分野（軍事や国家）に結びついた意味用法が教科に頻出していると見ることができる。以上の様相をまとめると次のようになる。

強	①	理科	②	—
弱	①	理科	②	—
主力	①	—	②	社会
総力	①	—	②	社会
弱体	①	—	②	社会

これらの語彙は、「エネルギー」と違って、専門用語であったり、社会的に話題になりやすい意味を担ったりするものではないため、教科学習の中ではほとんど気に止められることもないのではないかと考えられる。しかし、一般社会でよく使われる語彙として習得が望まれるものであることに異論はないであろう。これらの語彙に触れる機会が理科や社会で特に多いのであるから、その機会を生かすような語彙指導の工夫が行われると効果があるに違いない。「分類重要語彙リスト」を用いることで、これまでは教科の語彙だとは思われていなかった語彙が、実は特定の教科に特徴的なものであるという事実を数多く指摘していくことが可能になる。こうした語彙を生徒に定着させていくことは、国語科の言語事項における語彙力を養う活動に属するものだと考えられる。国語科が担当する語彙教育においては、個々の語がよく使われる教科での意味・用法あるいは用例に配慮した、教材開発や指導法の検討が求められるのではないだろうか。

4. 2. 4 書籍特徴語について

語数の多い書籍特徴語には、ある性質を帯びた力を意味する語（「強力」「実力」「威力」「迫力」「全力」「魔力」）や、力の性質を表す語（「強烈」「脆弱」「痛烈」「激烈」「熾烈」など）が目立つ。「パワー」「パンチ」のような、一般によく使われ意味も多様な外来語もある。こうした語彙は、教科書にはあまり使われないので、教科学習とは別の、例えば読書などを通して習得することが求められる語彙ということになるだろう。このような語彙についてどのように指導していくかを考えることも、国語科が担当する語彙教育の課題として重要である⁵。

⁵ 書籍に特徴的な語彙をどのように語彙教育に取り込んでいくかを具体的に考える際に、BCCWJ 内部の語彙表を作成して語彙の分類を行う研究が役立っていくだろう。

4. 2. 5 まとめ

ここまで、「1.1400 力」の分類項目に属する語彙を例に、「分類重要語彙リスト」から何が分かり、それを教科教育や語彙教育の検討にどのようにつなげていくことができるかについて、具体的に考えてきた。その要点は、教科や媒体の観点から語彙を分類できること、その分類の結果に基づいて、これまでには十分検討できなかったことについて具体的な議論を始めることができるようになるのではないかということである。この点について今度は「1.1401 弾力・動力・圧力など」の分類項目を取り上げて、さらに考察を進めていこう。

4. 3 「1.1401 弾力・動力・圧力など」の分類項目

4. 3. 1 概要

教科特徴語や教科書特徴語が多く、書籍特徴語が少ないこの分類項目の語彙は、生徒は教科学習を通じて習得する機会が多いのではないかと考えられる。その具体的なありようを、特徴語が特に多い教科である、理科と社会を例に観察してみよう。理科には13語、社会には9語の特徴語があるが、そのうち「抵抗」「圧力」「反発」「摩擦」「動力」「電力」「火力」の7語は、両方に共通している。

4. 3. 2 理科の特徴語

理科の特徴語であって社会の特徴語でないものは、「入力」「重力」「気圧」「引力」「高圧」「電圧」の6語であるが、これらは国語辞典に記載された一般に用いられる意味と理科の教科書における意味とが一致している。「重力」を例に取れば、『デジタル大辞泉』には「地球上の物体が地球から受ける引力で、物体の重さの原因となっている力。地球の万有引力と地球自転による遠心力との合力」とあり、この意味は、理科の教科書に見られる「物体にはたらく重力」「火星の表面の重力」「重力加速度」などの例の意味と一致している。他の語についても、例は省略するが、「高圧」を除いて、一般に使われている意味と教科書で使われている意味とが一致していることが確かめられる⁶。このようなものは、多くの生徒にとって、語とそれが表す概念の全体とが教科の学習を通して習得されていくのではないかと考えられる。

4. 3. 3 社会の特徴語

一方、社会の特徴語であって理科の特徴語でないものは「圧迫」「人力」の二語である。例えば「圧迫」は、『デジタル大辞泉』には、「① 強くおしつけること」「② 武力や権力などで押さえつけること」「③ 押さえつけて規模を縮小させること」の3つの意味が記されているが、社会の教科書には、「異民族の圧迫」「モンゴル帝国は一時イスラーム教を圧迫した」など②の意味と、「経済を圧迫する」「生活は圧迫され」など③の意味とで使われる場合が多く、①の意味は稀である。これは、4.2 で見た社会の特徴語「主力」「総力」「弱体」などの場合と同じで、語の持つ複数の意味のうちある特定の意味が教科書に使われているものである⁷。

⁶ 「高圧」のみは、『デジタル大辞泉』の3番目の語義に「権力などでおさえつけること」という、理科の教科書の意味とは違う意味が記載されていて例外となるが、他の語はすべて、教科書の意味と国語辞典の意味とが合致する。

⁷ もう一つの例「人力」も、社会の教科書には「人力車」という語が多いが、これは「人力」の持つ複数の意味の一つにあたるものである。

4. 3. 4 理科と社会の共通の特徴語

では、理科と社会の両方でもともに特徴語となる7語（「抵抗」「圧力」「反発」「摩擦」「動力」「電力」「火力」）はどうだろうか。まずは、「抵抗」を例に見てみよう。この語は、『デジタル大辞泉』には次のようにある。

ていこう【抵抗】[名] (スル)

- ①外部から加わる力に対して、はむかうこと。さからうこと。「権力に一する」「大手資本の進出に地元の商店会が一する」
- ②すなおに受け入れがたい気持ち。反発する気持ち。「相手の態度に一を感じる」「一人で入るには一がある」
- ③流体中を運動する物体が流れから受ける、運動方向と逆向きの力。
- ④「電気抵抗」の略。

でんきていこう【電気抵抗】

電流の流れにくさを表す量。電位差を電流で割ったもの。単位はオーム。抵抗。

理科の教科書に用いられている「抵抗」は、「空気の抵抗」「水の抵抗」「電気抵抗」といった例が多く、国語辞典の③または④の意味にあたる、自然科学の専門用語を基盤にもつものである。一方、社会の教科書には、「独立に抵抗するフランス人入植者」「抵抗権」など国語辞典の①の意味で用いられている場合が多い。このように、同じ「抵抗」でも、理科と社会とでは用いられる意味が異なっており、それぞれ、一般に用いられる「抵抗」の多様な意味の一部を分担していると言える。

「抵抗」という語を習得した人は、①②のような一般語として理解しまた使用することが多いだろうが、③④のような科学の専門用語の意味との連続性は意識しているのではないかと考えられる。流体中の物体が流れから受ける反対方向への力という意味は、例えば、進もうする自身が社会や他人から受ける反対方向の力というように、①②の意味の「抵抗」においても感じ取ることがあるのではないか。一般語と専門語の意味を有機的に結びつけることができるようになることで、語彙力が豊かになる面があるだろう。理科において③④の意味の「抵抗」を学習する際に、①②の意味と関連づけながら「抵抗」の概念を学習者の中に確立できれば、語彙学習の効果は格段に増進しよう。

表 10 理科と社会に共通の特徴語の意味と教科書の用例

	国語辞典（デジタル大辞泉）の用例	理科の主な用例	社会の主な用例
抵抗	1 外部から加わる力に対して、はむかうこと。さからうこと。「権力に一する」「大手資本の進出に地元の商店会が一する」		独立に抵抗するフランス人、抵抗権
	2 すなおに受け入れがたい気持ち。反発する気持ち。「相手の態度に一を感じる」「一人で入るには一がある」		
	3 流体中を運動する物体が流れから受ける、運動方向と逆向きの力。	空気の抵抗、水の抵抗	
	4 「電気抵抗」の略。	電気抵抗	

圧力	1 押さえつける力。気体・液体または固体が、ある面を境にして、その両側から垂直に押し合う力。単位はパスカルのほか、アト、水銀柱メートルなどを用いる。	空気に圧力を加える、重力による圧力、温度や圧力の変化	
	2 威圧して服従させようとする力。「大国の一に屈する」		欧米列強の圧力、裁判所に圧力をかけ、圧力団体
反発	1 他人の言動などを受け入れないで、強く否定すること。また、その気持ち。「一を買う」「運命に一する」		親米政権への反発、教皇庁の擄取に反発する、民衆の反発
	2 他からの力をはねかえすこと。はねかえること。「磁石の同じ極どうしは一する」	髪の毛が反発しあって逆立つ、磁石を置くと反発力をうみ、反発係数	
	3 値下がりしていた相場が一転して値上がりすること。⇔反落。		
摩擦	1 物と物とがすれ合うこと。また、こすり合わせること。「肌を一して暖をとる」「乾布一」		
	2 人間の社会関係で、二者の間に意見や感情の食い違いによって起こる、不一致・不和・抵抗・紛争など。軋轢(あつれき)。「貿易一」		競争と摩擦、貿易摩擦、日米摩擦
	3 互いに接触している二つの物体のうち、一方が運動しようとするとき、または運動しつつあるとき、その接触面に運動を妨げようとする力が働く現象。また、その力。相対速度により運動摩擦・静止摩擦に、運動状態により滑り摩擦・転がり摩擦などに分けられる。	燃焼・摩擦・気体の圧縮、摩擦力、摩擦熱	
動力	水力・電力など、天然に存在するエネルギーを原動機によって機械を動かす力に変えたもの。原動力。	火の動力、電気を動力に変える、水力に変わる動力源	水車の動力、動力機関が発明され、大型動力船
電力	電流が単位時間にする仕事量。直流では電圧と電流の積に等しい。単位はワット。	発電機の電力、電力の単位にはワットが使われる	電力を風力でまかなう、電力会社
火力	1 火の燃える勢い。また、燃える火の強さ。「一が弱い」	火力発電の原理、火力発電をしてみよう	火力発電の増加、中国の火力発電所
	2 銃砲などの火器の威力。「敵は一においてまさる」		

「抵抗」も含め、理科と社会の双方の特徴語になっている 7 語について、国語辞典の記述と教科書で多用されるタイプの用例とが対照できるようにまとめると、表 10 のようになる（教科書の用例は教科書コーパスの公開部分だけでなく、非公開部分からも引いた）。

「抵抗」「圧力」「反発」「摩擦」のような複数の語義に分かれるものについては、理科と社会とで異なる意味を分担している様子が明瞭に見て取れる。「圧力」「反発」「摩擦」いずれも、自然科学の専門用語としての意味が理科で、一般用語としての意味が社会で学習さ

れる。これはちょうど「抵抗」の場合と同様であり、専門用語と一般用語とをうまく関連づけて学習させることができるような教育上の工夫が求められていこう。一方、「動力」「電力」のような語義が一つだけの場合は理科と社会とで同じ意味で用いられている。

4. 3. 5 まとめ

以上、「1.1401 弾力・動力・圧力など」の分類項目に属する語彙のうち、理科と社会の特徴語になっているものについて、「分類重要語彙リスト」から分かることをもとに、教育上どのような工夫を行うことが考えられるかについて、述べてみた。その要点は、理科で学ぶような科学の専門用語の意味と、社会科でよく使われる一般用語としての意味が有機的に結びつくような語彙指導が望まれるのではないかということである。

5. おわりに

以上見てきたように、「分類重要語彙リスト」を用いることで、重要度・意味・教科・媒体などの観点から語彙を分類しながら考察することが可能になる。その分類と考察を通して、学校教育における語彙の扱いについて具体的な検討が進むことが期待できる。

文献

前川喜久雄（2006）「特定領域研究『日本語コーパス』がめざすもの」，特定領域「日本語コーパス」平成 18 年度全体会議予稿集，pp.1-8.

別表「分類重要語彙リスト」の例

番号	細目	レベル	国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報	教科書	書籍	無特徴
1.1000	事柄	a	物(もの)〔サ変可能〕、事(こと)、事柄、物事、項目、出来事	分(ぶん)、事柄、項、例	例、現象	事件	事柄、例、出来事	分(ぶん)、項目、例、事故	対象、物(もの)〔サ変可能〕、例、変	事、項目、項、例、事故	対象、項目、例、現象	事柄、項目、項、例、現象	者(もの)、分(ぶん)、やつ、一(ひと)〔サ変可能〕、対象、物(もの)〔一般〕、事(こと)、事情、件、旨(むね)、あれこれ、ケース、事態、事件、変〔本能寺の〜〕、大変〔形状詞〕、大変〔副詞〕、騒ぎ	事項
		b	事物	事象	事象	事変							物件、条項、前項、条、本件、由(よし)、儀、さた、異変、パニック、スキヤンダル、奇跡、奇蹟	余儀
		c											アイテム、件(くだり)、万事、一式(いっしき)、案件、大(だい)〔じ〕、ひと事、ただ事	修事、時事、風物
1.1010	こそあど、他	a	どれ、他者	どれ、どちら、どちら	これ、どれ、どちら、どちら	他方	どちら、どちら				別	他方	これ、こちら、それ、そちら、か、あれ、あちら、いづれ、何(なに)に、何(なん)、何事、前者、後者、方(ほう)、別	
		b											何者	
		c											そ、何(なに)に	別個
1.1030	真偽・是非	a	現実、実際	真(しん)、正(せい)、不良	正(せい)	真理、本格		実態		実際	実際	正(せい)	真(しん)、真実、誠(まこと)、本当、本来、事実、現実、実態、生(なま)〔〜の声〕、不思議、当然〔形状詞〕、当然〔副詞〕、まとも	
		b	是非	非、否	公式	尊敬	尊敬		神秘		不正	公式、尊敬	ほんま、リアリティー、試し、ミステリ、是、当	ありのまま
		c	是非	真偽	虚(け)仮(か)レギユラー	是非					虚偽	虚偽(こけ)、是非	正銘、オカルト、不可	白黒、実地、史実、虚(きよ)、当否、善悪、よしあし
1.1040	本体・代理	a	外(ほか)、補助、例、資料、データ、材料	正(せい)、例、資料、データ、材料、主(しゆ)	正(せい)、例、資料、データ、材料、主(しゆ)	主体、外(ほか)、主体、外(ほか)、資料	他(た)、例	主体、他(た)、例、資料、データ、素材、材料、主(しゆ)	代表、例、資料、素材、材料	主体、補助、例、資料、材料	本人、例、資料、データ、主(しゆ)	主体、外(ほか)、正(せい)、代表、例、資料、データ、素材、主(しゆ)	自体、自身、本質、本物、実(じつ)、仮、本人、代理、代わり、替わり、候補	
		b		偽(にせ)、標本	個体、本体、複製、模(も)型	人体、模(も)型、史料	ボディ、メーン	本体、人体、データベース	本体、人体、模(も)型	本体、人体	本体、媒体、実物、複製、標本、データベース	個体、本体、人体、複製、標本、模(も)型	自、正体、機体、車体、ボディ、偽物(にせもの)、判例、支流、メーン	代善(だいい)、サンプル、主流
		c	掛け替え			自他、掛け替え		掛け替え				ハンクアップ	掛け替え	部外、得(え)体、神(かみ)体、船(ふね)体、控(か)え、サブ、前(まへ)例、本(ほん)流

第3章 第3節

作文コーパスからみる生徒の使用語彙

鈴木 一史

1. はじめに

教育現場で語彙が問題とされるとき、多くは「漢字」の問題であり、「理解力」の問題であった。知っていなければならない言葉や知ってほしい言葉などを教科書に出てくるたびに、教師は選定して学習させる。その結果、定着しているかどうかについてはペーパー・テスト形式で行う。これも一つの方法であることは間違いない。しかし、「理解語彙」と「使用語彙」という二つの側面で捉えるときに、「表現語彙」の伸長についてはなかなか図られていないのではないか。そもそも、表現語彙の伸長・拡充をどのようにして測るのかについては、教師の経験によるところが大きかったように思う。

しかし、このプロジェクトにより、語彙のレベルや教科書語彙の分析などが進み、使用語彙についてもある程度の指標が得られてきた。そこで、生徒の作文を『現代日本語書き言葉均衡コーパス』(BCCWJ)から得られたいくつかの指標によって分析することで、生徒の発達過程が分かるのではないかと考えた。

学年が上がるに従って文章力も上がると考えられる。その時の「文章力」とはなにか、なにが変化しているのかをとらえたい。生徒の表現語彙の発達や傾向が分かることで、それらを伸ばす次のステップに進めることになると思う。本稿では、生徒作文をデータベース化し、生徒の使用語彙の分析から使用語彙の傾向をとらえることが目的である。

2. 調査データ

2. 1 学校・社会対照語彙表

言語政策班では、BCCWJの語彙表に加える情報として、頻度によるレベル分け、教科書特徴語、教科書初出学年を付加してきた。これらの情報によって、コーパスが教育分野で有効活用されると考えられるからである。(分析方法とデータは本報告書の他の章節を参照のこと。)

本稿では、図書館コーパスをもとにしたレベル分け、及び教科書特徴語について扱う。

2. 2 作文コーパス

本校(東京大学教育学部附属中等教育学校)は中等教育学校であるために、生徒は中学1年(1年生)から高校3年(6年生)まで、一貫した教育がおこなわれている。1学年は、1クラス40人で3クラスあり、120人が在籍している。また、男女の比率は50:50である。国語の教員は6人いて、担当分野や担当学年が固定しているわけではない。2-2-2-制のカリキュラムを取っているため、2年ごとに担当が変わる。国語科すべての教員が6年間で1度は1人の生徒と接するため、特定の学年が特定の教員の影響を強く受けることが少ない。このような状況の中で、同時期に共通の課題で作文を書かせデータベース化することは、各学年の特徴をとらえるには最適であると考え、この調査を行った。

調査対象は中学1年から高校2年。調査は、平成21年1月に、冬休みの宿題として課した。課題は、「年末・年始の行事をふまえて、文化について述べよ」である。原稿用紙を配

付し、字数は 800 字程度とした。手書きで提出させた。

時間を区切った提出物ではないので、他文献やネットからの引用も考えられ、分析時に考慮すべき課題である。しかし、手書き提出にしたために、調べたとしても最終的に自分で書きなおす必要があり、その点では生徒自身の言葉であるにとらえてよいだろう。

手書き作文の解析方法は以下の手順で行った。手書き文字を電子データ化した。文章表現上どのような間違いを犯しやすいか分析するため、表記ミスなどもそのまま入力。次に全文検索ソフトで検索できるように、タグ付けを行う。解析ソフトは「茶まめ」と形態素解析辞書 UniDic1.3.12 を使用した。これにより、文章を単語に区切り、品詞情報、語種情報が付与される。

5 年生の作文データ分析に、レベルと初出学年、特徴教科のデータを関連付けると表 1 のような語彙表ができる。これは 5 年生のデータの一部分である。「年末・年始…」という課題から、「正月」という語が何度も使われていることが分かる。

表 1 生徒使用語彙表例

語彙素読み	語彙素	語彙素の カウント	品詞	語種	レベル LB FL	初出学年	特徴教科
ジョウ	城	1	接尾辞-名詞的-一般	漢	a	小後	
ジョウ	場	2	接尾辞-名詞的-一般	漢	a	小前	保
ジョウ	状	146	接尾辞-名詞的-一般	漢	a	小前	理芸
ショウカイ	紹介	3	名詞-普通名詞-サ変可能	漢	a	小前	国外
ショウガイ	障害	1	名詞-普通名詞-サ変可能	漢	a	小後	技保
ショウガク	小学	3	名詞-普通名詞-一般	漢	a	小前	
ショウガツ	正月	160	名詞-普通名詞-一般	漢	a	小前	
ショウギョウ	商業	4	名詞-普通名詞-一般	漢	a	小後	社
ジョウキョウ	状況	3	名詞-普通名詞-一般	漢	a	中	保情
ショウゲキ	衝撃	1	名詞-普通名詞-一般	漢	a	中	
ジョウケン	条件	1	名詞-普通名詞-一般	漢	a	小後	数理社技 保情
ショウコ	証拠	1	名詞-普通名詞-一般	漢	a	小前	
ショウゴ	正午	1	名詞-普通名詞-副詞可能	漢	b	小前	
ショウシ	少子	3	名詞-普通名詞-一般	漢	c	中	社技

3. 作文コーパスの分析

3. 1 データ概要

作文コーパスの全データは以下のとおりである。「1 年」が 1 年生全体の作文データであり、以下同様に学年が上がっていく。「ALL」は全ての作文の合計である。学年により人数に異なりがあるのは、宿題として課したために、提出した人数による。人数にばらつきがあるために、これからのデータ分析は主に「異なり語数」によって行う。

表 2 作文コーパス概要

	人数	延べ語数	一人当たり	異なり語数	異なり語数 /延べ語数	句点の数	一文の平均 語数
1年	118	64,285	544.8	3310	5.15%	2729	23.6
2年	103	54,745	531.5	3488	6.37%	2412	22.7
3年	66	37,205	563.7	2765	7.43%	1183	31.4
4年	84	45,976	547.3	3396	7.39%	1859	24.7
5年	81	47,707	589.0	3559	7.46%	1964	24.3
ALL	452	249,918	552.9	7389	2.96%	10147	24.6

一人あたりの語数は、531.5 語から 589.0 語であり、全体では 552.9 語である。800 字程度の作文を課し、おおよそ 550 語程度で作文を仕上げてきたことになる。

異なり語数は、2765 語から 3559 語である。3 年生は人数も少なく、また学年の特質も影響していると思われるために、以下の分析データでも 3 年生のみ特殊な状態が生じているように見える。1 学年当たりおおよそ 3000 語強の語彙で文章を書いていることがうかがえる。しかし、全体をみると 7000 語を超えている。これは、学年によって重なっている語があるために単純な合計ではない。しかし逆に、学年別の分析よりはるかに多いということは、多様な語彙を感じさせる。

次に、異なり語数を延べ語数で割ることによって、語彙の広がりを見る。1 年生では 5.15% であるが、徐々に数値が上がり、5 年では 7.46% である。同じ年末年始について書く場合でも、1 年生はお年玉など発想が同様になる傾向があるが、学年が上がると、多様な語彙を使って年末年始について表現していることがうかがえる。語彙の多様さは発想の多様さにもつながり、同じ年末年始の過ごし方でも、捉え方が多様になってきていると考えられる。

一文の長さは、20 字強。学年が上がると、若干長くなっているが、それほど大きな差はない。つまり、一文の長さが文章の巧拙に影響するわけではないようである。

3. 2 語種分析

作文コーパスの語彙を異なり語数で学年・語種別にカウントしたものが以下の表 3 である。学年が上がるに従って抽象的な言葉、つまり漢語を多く使うようになっていないかということである。

表 3 作文コーパス語種分析

語種	和語	漢語	外来語	記号	固有	不明	混成
1年	50.4%	36.4%	4.4%	1.3%	3.7%	0.5%	2.8%
2年	49.9%	37.1%	3.8%	1.2%	4.6%	0.5%	2.4%
3年	51.1%	37.4%	4.3%	1.2%	3.0%	0.4%	2.4%
4年	47.7%	39.4%	4.7%	1.1%	3.5%	0.4%	2.8%
5年	46.5%	41.2%	4.5%	1.2%	3.3%	0.3%	2.8%
ALL	44.1%	40.0%	5.9%	1.1%	5.0%	0.4%	2.7%

表 3 から、和語の割合がかなり減ってきているのに対し、漢語の割合は増えているように思う。そこで、抽象的な言葉の比率を調べるために、「和語」と「漢語・外来語」について、グラフ化した。

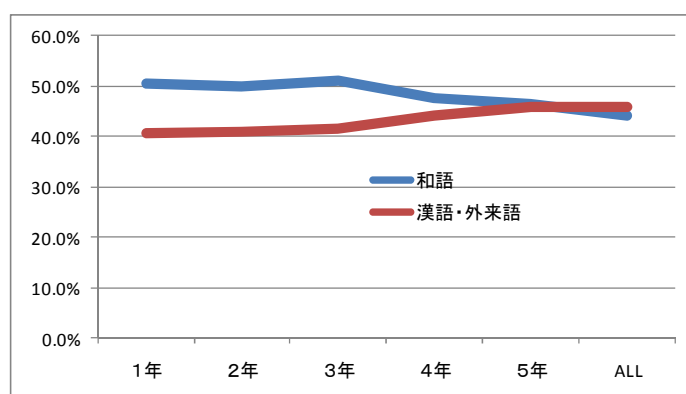


図1 和語と漢語・外来語の使用変化

4年生までは和語の使用率が高いが、5年生で和語と漢語・外来語の比率がほぼ同じになる。そして、1年から5年まで徐々に和語の比率が下がっていき、反対に漢語・外来語の比率は上がっている。ここから、学年が上がるにつれて段々と抽象的な言葉や概念を表す言葉が増えていることがうかがえる。

ALL では完全に逆転している。つまり、各学年で和語については重なる言葉が多く、漢語については、異なる言葉を使っているということである。たとえば、「私」や「降る」などはどの学年でも使うが、「酒宴」や「儀礼」などは、個別に使う生徒がいるということである。

3. 3 レベル分析

作文コーパスの語彙表と「学校・社会対照語彙表」を関係づけ、レベルデータを付与し、それぞれについて異なり語数の数をカウントしたものが表4であり、それをグラフ化したものが、図2である。

表4 学年別レベル分け

レベル	a	b	c	d	e
1年	48.0%	18.6%	13.4%	8.1%	11.9%
2年	47.5%	18.8%	13.6%	7.7%	12.5%
3年	50.7%	18.0%	13.5%	7.0%	10.8%
4年	48.1%	18.9%	13.5%	7.6%	11.9%
5年	48.9%	18.3%	13.5%	7.3%	12.0%
ALL	36.1%	22.1%	17.4%	9.6%	14.8%

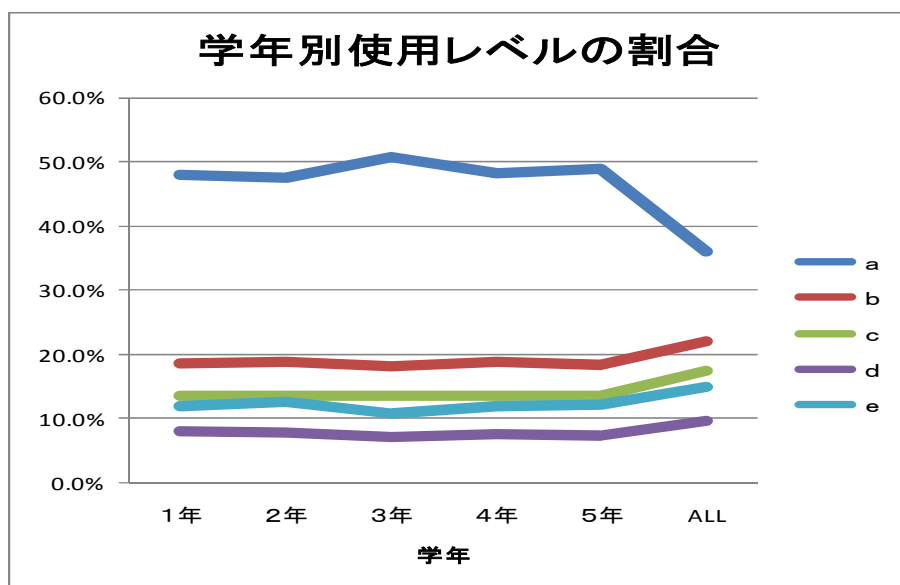


図2 学年別使用レベルの割合

全ての学年でaが50%近く使われ、文章を成り立たせている。aからeの使用率は1年から5年まで、ほとんど同じである。文章の特性として、aが減るから文章の発達が見られ、deを使用すると高度な文章であるとはいえないようである。

しかし、全体（ALL）を分析すると、aの数値が格段に下がり、その他の全てのレベルが確実に上昇している。これは、aという基礎的な言葉は、全ての学年で同様に出現するために、1年から5年までほとんどaの語彙は増えていかないことが示されている。つまり、学年を問わず重複語彙が多いということである。その反面、レベルb c d eの語彙は、重複はするものの、いろいろな言葉を使っていることが分かる。全学年で同じテーマで作文を書かせていることを考えると、同じテーマや題材に対しても、様々な語彙で考えて書いていることがうかがえる。これは、異なり語数が、1年から5年までほとんど同じ数であったものが、全学年で見ると倍以上になっていることと合わせて考えると、異なり語数の伸長は主にレベルb c d eを中心に起こっていることが分かる。また、語種変化から考えると、その語彙は漢語や外来語のレベルの高い語が使用語彙として増えているようである。

3. 4 初出学年分析

次に、作文コーパスの使用語彙がいつ学習したものであるか、すべきものであるかについて、初出学年のデータを踏まえて考察する。

表5 生徒使用語彙の初出学年

初出学年	小学校前期	小学校後期	中学校	高校
1年	42.9%	28.2%	18.6%	10.3%
2年	44.1%	27.3%	18.4%	10.2%
3年	45.7%	26.7%	17.9%	9.6%
4年	41.3%	27.4%	20.2%	11.1%
5年	39.7%	27.5%	20.8%	11.9%
ALL	32.4%	29.2%	24.1%	14.3%

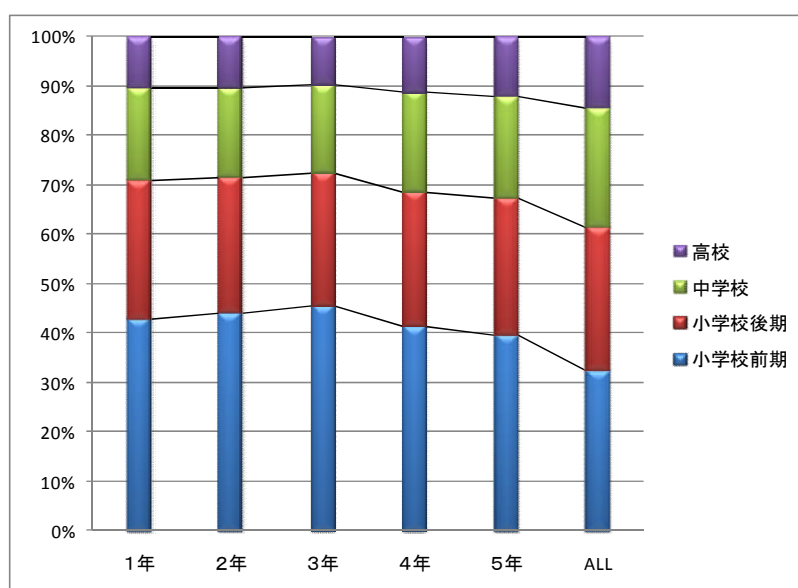


図3 生徒使用語彙の初出学年推移

学年を追ってみてみると、小学校前期で習う言葉は1年から3年まで上昇し、その後低くなっている。これは教科書に出てきた言葉が、自分の使用語彙になるまでにはしばらくかかることを意味している。しかし、小学校前期の語彙は、基礎的な語彙であり使用頻度の高い語彙であると考えられる。全学年データで小学校前期語彙は2312語、そのうちレベルaは1592語、約70%である。前項でみたように、レベルaは重複が多い。それでも3年生までは使用語彙が増えていることが分かる。逆に4年生からは減っていく。基礎的な語彙よりも新しく獲得した語彙を使う傾向がみられる。そして、全体では小学校前期の使用語彙は32.4%。これは全体のレベルaの使用率36.1%に近い。

小学校出現語彙に比べて増えているのが、中学・高校で習う語彙である。習った言葉を徐々に自分のものとして使用していく様子が見えてくる。

3.5 教科特徴語分析

最後に、教科書に特徴的な語彙との比較をする。これは、一般の書籍に比べて、教科書に特徴的な語彙ということは、学校で習うことで身につけて、しかも使えるようになってきていることを意味する。基本的に言葉は国語で習い、それを使う練習をしながら、使用語彙が増えていくと考えられる。また、そうであるからこそ、国語科の授業の中で漢字練習などが多くなされることとなっている。他教科で漢字練習や語彙テストなどはあまり行われていないだろう。

これを検討するために、教科特徴語と使用語彙との関係を調べたものが表6である。教科特徴語については、複数教科にわたって特徴度が高いものは除き、各教科単独で数値が高く有意な語彙をカウントした。

表 6 使用語彙の教科特徴語の割合

教科特徴	国語	社会	数学	理科	英語	技術	芸術	保健	情報
1年	3.4%	6.0%	1.0%	2.2%	1.7%	4.1%	2.6%	0.9%	0.8%
2年	3.6%	6.2%	1.0%	2.6%	1.7%	3.5%	3.1%	1.5%	0.9%
3年	3.4%	5.6%	1.3%	1.8%	1.7%	3.6%	2.9%	1.4%	1.2%
4年	3.0%	6.3%	1.1%	2.0%	1.7%	3.5%	2.5%	1.6%	0.8%
5年	6.4%	3.1%	1.1%	2.4%	1.7%	2.8%	2.5%	1.5%	1.0%
ALL	3.0%	6.4%	0.8%	2.5%	1.4%	3.1%	2.5%	1.4%	1.1%

表 6 の全学年データをグラフにしたものが、図 4 であるが、その際に、語彙全体にどれだけ教科特徴語が含まれているのかを比較するために、横に並べてグラフ化した。全学年で使用されている語彙の中で、各教科の特徴語の割合を示したものが図 4 の ALL である。

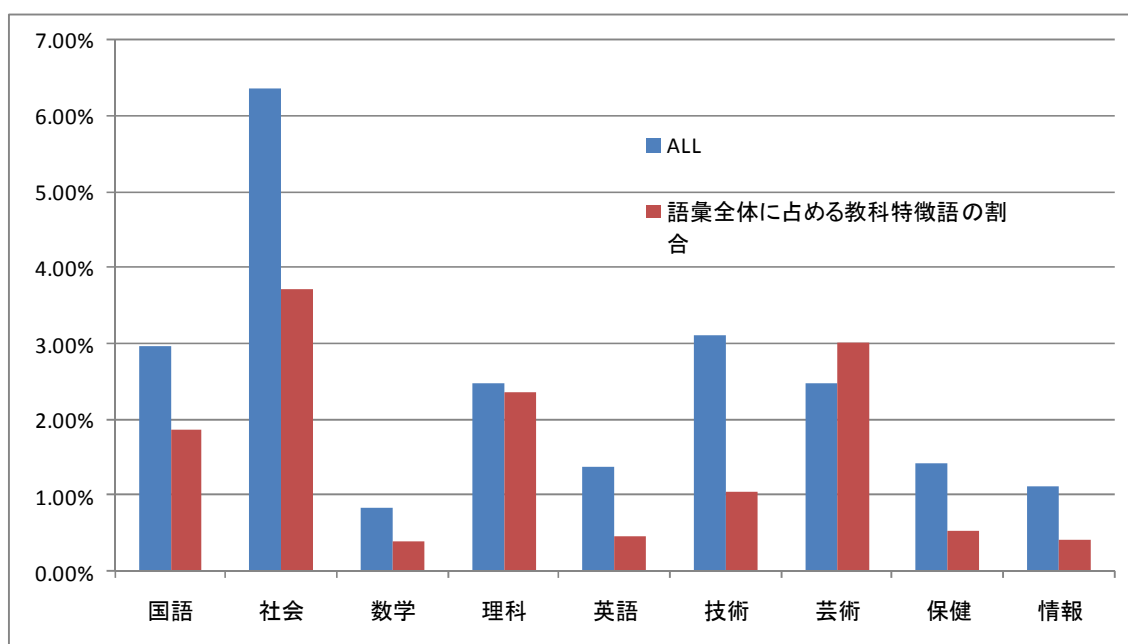


図 4 生徒使用語彙と語彙全体に占める教科特徴語比較

この図をみると、圧倒的に社会の特徴語が多い。社会科で習った言葉を使って文章を書いていることが分かる。この結果は、課題が「文化について」というテーマと評論文のような文種であったことも影響していると考えられる。しかし、語彙全体に占める割合の倍近い語彙を使用しているということは、評論文などを書く力は、社会科的語彙の拡充が必要であることが分かる。また、社会科での語彙の拡充により、それが概念や考える力などを形成し、文章力につながるのではないかと。社会科のテストなど、別の方面からの検証も必要である。

次に多いのが、技術科である。技術科は「技術・家庭」科であり、おせちなどの食文化にかかわることを学習するために、このような結果になっていると考えられる。具体的な語は以下のものである。(全学年データで、技術科だけに特徴的なレベル e の語彙)

田作り、取り分ける、酢の物、挟む、魚肉、取り皿、干し柿、レトルト、八宝、ごまめ、きと、グラタン、満たす、汲み、である。

このことから、語彙の学習は国語科だけで行うのでは、なかなか広がって行かないことが分かる。生徒は様々な学習を通して言葉を広げ、生活の中で生かしながら、自分の使用

語彙として定着させている。したがって、語彙の拡充は国語科の教科書の中だけで完結してしまうのではなく、様々な教科や生活と結びつけながら行うことで、考えも広がっていく。そのような観点でこれからの学習方法の開発を行う必要がある。

次に、学年ごとの変化を調べるために、図5を作製した。

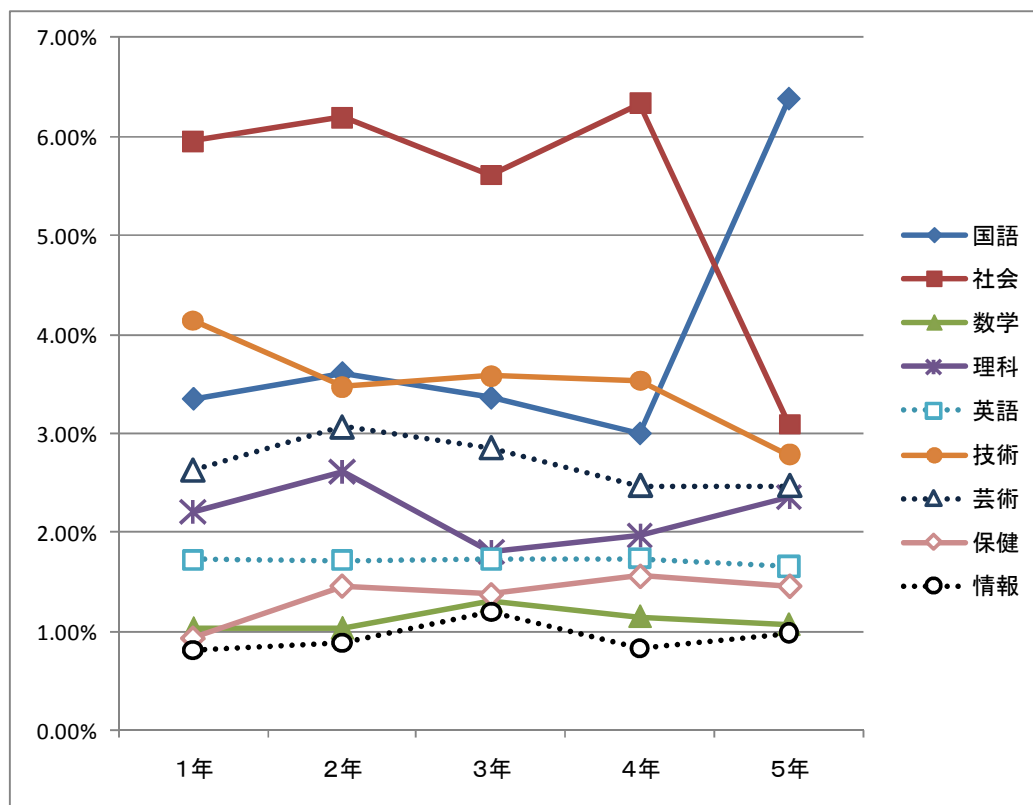


図5 各教科特徴語の学年推移

1年から4年までは、社会科や技術科の特徴語彙が多いなど、全学年で見たときと同様の傾向がみられる。特筆すべきは5年生の国語と社会の関係である。5年生では、国語特徴語彙が飛びぬけて多くなり、社会特徴語彙が一気に減っている。全体を100としているので、片方が増えれば片方が減ることになるが、他の教科はそれほど大きな変動がないので、この二つの教科の特性が影響したものと考えられる。

表7は国語にのみ特徴的な語彙の中で、全てのレベルeを抜き出したものである。これらの言葉は、一般にあまり頻度が高くないが、教科書には特徴的に出てくる語であり、しかも使用語彙となっている語である。5年生になってこれらの言葉が特徴的に出てくるといことは、5年生の使用語彙に関してはかなり国語科に負うところが大きく、学習によって身につけていく様子が見えてくる。

表7 生徒使用語彙*国語特徴語*レベル e

語彙素読み	語彙素	語彙素の カウント	品詞	語種	レベル LB_FL	初出学年	特徴教科
ナナクサ	七草	13	名詞-普通名詞-一般	和	e	中	国
モチヨル	持ち寄る	3	動詞-一般	和	e	小前	国
アヤツル	操る	1	動詞-一般	和	e	高	国
ジュウタイ	重態	1	名詞-普通名詞-一般	漢	e	高	国
トシノクレ	年の暮れ	1	名詞-普通名詞-一般	和	e	高	国
ハレガマシイ	晴れがましい	1	形容詞-一般	和	e	中	国
ヒモジイ	ひもじい	1	形容詞-一般	混	e	中	国
ホームルーム	ホームルーム	1	名詞-普通名詞-一般	外	e	中	国
ミンブ	民部	1	名詞-普通名詞-一般	漢	e	高	国
モギ	裳着	1	名詞-普通名詞-一般	和	e	高	国
シュウ	終	1	接頭辞	漢	e	中	国

国語科のカリキュラムとして、4年生までが必修単位「国語総合」を履修することになっており、5年生から「現代文」や「古典」などを選択することになっている。このことを考えると、5年生での使用語彙として抽象度も上がっていることがうかがえる。国語科が担う責任は大きい。

4年生までの必修単位では、全教科との連携や生活や生きることに結び付いた学習が望まれ、5年からは一般的にはそれほど多く目にするような抽象的な語彙を積極的に学習の中に取り入れていく必要があり、そのことが学習者にとって知識を活性化し、使用語彙の広がりへと結びついていっている。

4. まとめ

語彙レベルや教科特徴語など、生徒の表現語彙を分析する際にそれらを使用することで、生徒の使用語彙の発達や伸長がとらえられた。

今後の課題として二つある。一つは、分析の観点を増やすことであり、それには分類語彙表のような、語彙を意味という観点で分析したものが有効である。「学校・社会対照語彙表」には、分類語彙表番号も付加されているので、それらを使うことでさらに研究が深まると考える。

もうひとつは、この使用語彙を、どのようにしたら効果的に伸ばせるかという教育プログラムの問題である。それには、具体的な一つ一つの語彙についての分析が必要であろう。使用語彙の拡充が生徒の概念形成を助け伸ばすことになる。

第3章 第4節

漢字教育におけるコーパスの活用

棚橋 尚子

1. はじめに

文字、特に漢字の習得は学習者の学力向上に欠かせないものである。OECDのPISA調査や全国学力・学習状況調査において学習者の読解力、ひいては「書く力」の不振が言われて久しい。その理由としてカリキュラムの問題、授業内容の問題などさまざまな分析がなされている。どの研究者も指摘していないことではあるが、私は日本語における複雑な表記体系がそれらの伸長を阻んでいる可能性があると考えている。

ある学級の6年生児童に漢字学習への意識を聞いたところ、半数が漢字学習を「嫌い」と答えた。前掲の調査においては、「学力の二極化」も問題になっているが、これらの回答実態からは漢字学習に対する学習者の意識そのものが二極化している事態が看取できる。また、高等学校教員からの聞き取り調査で、芥川の「羅生門」を高等学校用のテキストではなく、小学生向けの総ルビの文庫本を使って授業したところ、受験学力の低いといわれるその高校の生徒たちが主題に迫る活発な意見交換を行った事例を耳にした。同様の事例は小野瀬（1999）も小学生を対象とした調査実験の結果、以下のように述べている。

文章から直接読み取れる浅い処理で済むタイプの問題ではすべて、ルビなし群の方が正答者が多かったのだが、全体の読み取りに加えて、その深部を想像したり見通したりすることが必要になる質問では、総ルビクラスの誤答者の方が少ないという結果を得た。

小野瀬の実験は、ルビの有無が読解にどのように影響するかを解明したものである。先の高等学校教師の実践や小野瀬の実験は、漢字が読み書きできないこと、できても不十分なことが文脈の意味理解を阻害していることを示唆している。このことは、漢字の習得が教科内容の学習にいかに重要かを示している。

これらの示唆を踏まえ、本稿は、現状の漢字指導の問題点を指摘したうえで、コーパスの活用でひらけていく漢字指導の地平を述べ、学習者の学力向上の一助とすることを目的とする。

2. 漢字指導の実態

2. 1 学習指導要領と漢字

明治維新以来国語国字問題は多方面で論議されたが、教育との関係でみると学校教育で習得した漢字が社会における漢字使用の実態と連動していないために学校教育での漢字習得が社会生活では「役立たない」という点が大きな問題であった。また、音訓や送り仮名等表記にまつわる規則の不統一も学習の阻害要因となっていた。昭和13年に発足した国語協会の会長である近衛文麿はその設立総会で、「わが国語国字が世界にその比を見ないほど複雑にしてかつ不規則でありますので、この学習負担の過重なことは今更申述べるまでもありません」とし、国語国字の整理を児童の学習負担の軽減という観点から教育上の喫緊の課題として挙げた。（旧言語問題談話会同人（1937））

このような問題は戦後一気に改革され、学校と社会における漢字使用の実態の断絶も含め、ある程度解決されることになる。教育漢字は当用漢字別表として社会での使用が前提となる当用漢字の中から選定され、音訓や送り仮名も統一された。教育漢字の字数は、当用漢字制定時には 881 字、その後常用漢字の改定に連動し、996 字となり現在は 1006 字である。学習指導要領では、この 1006 字について義務教育修了時までには書けるようになることを求めている。ここで「書ける」という点に焦点を当てて考えると、書きは読みより習熟に時間がかかるということが問題として挙げられる。このような漢字の読みと書きについての分離的な考え、つまり読みは書きに先行するという考え方が学習指導要領上に明示されたのは昭和 43 年度版のもの以降からである。昭和 43 年改訂の学習指導要領では第 1 学年では、読み 70 字ぐらゐに対し、書きは 40 字ぐらゐ、第 6 学年では読みは 881 字ぐらゐ、書きは 800 字ぐらゐと規定していた。しかしながら、書きまでを指導する漢字そのものの指定がなかったため、実態としては配当された漢字のすべてを小学校で書けるようにするための指導が行われていた。この事態に一線を引いたのが平成 10 年告示の学習指導要領である。

平成 10 年の学習指導要領では、書きは配当の次の学年までにかけて漸次できるようにすると定めている。従前の字数指定ではなく、書けるようになるための期間を繰り下げることを通して、漢字の習得範囲と段階を明らかにしたのである。この考え方は平成 20 年告示の学習指導要領にも引き継がれたが、平成 20 年告示学習指導要領では、従前と比較し文や文章の中で漢字を「使える」ことが重視された。さらに中学校での「書き」については平成 10 年度より早い段階である 2 年生で教育漢字 1006 字の習得が終わるようにとし、3 年次をその習熟にあてるようにした。一方、常用漢字については、中学校では読めるようにすることを求めている。しかしながら、教科書では「新出漢字」として教科書会社が独自に常用漢字の学年配当を行っているため、かなりの常用漢字を中学校で書きまで含め学習しているという実態がある。

2. 2 字形学習に傾く小学校漢字指導の実態

表 1 漢字配当数と国語科の授業時間数

	漢字配当数	授業時間数
第 1 学年	80	306
第 2 学年	160	315
第 3 学年	200	245
第 4 学年	200	245
第 5 学年	185	175
第 6 学年	181	175

漢字の習得は、学習を重ねて自学ができるようになるまでの期間の指導が重要である。この期間に字形の原則や音訓、筆順などの漢字の基礎を丁寧に習得させ、漢字を覚えることの達成感や読み書きにおける漢字の有用性を指導することがのちの漢字学習の成否に影響する。さて、小学校の漢字指導は、当然のことながら字形指導が中心となる。初めて出会う漢字の形を覚えられないことには目の前にある文章を読み、内容を理解していくことができないからである。しかしながら実際には、国語の授業時数と配当漢字数の関係、そして前述した自学能力のある程度の伸長もあって、学年が上がるほど漢字の指導は学習者の主体性に任されることになる。そして、とめ、はね、はらいといった細部を意識した評価が繰り返される。このことは、ともすれば漢字嫌いの学習者を生み出す危険性をはらんでいる。現職教員 251 名を対象とした棚橋（1998）の調査によれば、「テストの際、細部への目配り（とめ、はね、はらいなど）をよくしたか」という設問に対し、「よくした」と回答した教員は 67.5%、「たまにした」と回答した教員

は 27.8%であり、ほとんどの教員が字形細部にこだわって漢字を採点している状況だと判断できる。「てへんははねる」「うしへんははねない」「もんがまえに入った耳の 5 画目は、6 画目を突き抜けない」「女の 2 画目は 3 画目よりやや上に出す（または出さない）」など、細かい基準が示される。問題は教師によってその基準が異なっていることである。

大学生に漢字の採点について考えさせる授業¹をすると、多くの学生の非常に高い「標準意識」に驚かされる。常用漢字表および改訂常用漢字表の前書きには、手書きの許容が示されているが、その範囲の書字を大学生に採点させる²と、ほとんどの学生が多くの「正しくないもの」として採点する。しかし、どこを間違いととるかは各学生によってかなり異なる。

たとえば、以下の漢字テストでは、学生の採点は 6 点から 16 点までの範囲を形成した。

①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑲	⑳	
る あらため	い いつさい	お おくやま	つ はけんし	かん いちねん	こ ことば	う うかい	が がつぎせ	ひ ひかくす	よ よ	か かぜのた	こ こがい									
改 改める	一 一切	奥 奥山	保 保健室	一 一年間	言 言葉	学 学級会	比 比較する	風 風の使い	戸 戸外											
ひ ひび	か か	あ あさける	え え	お おほれる	お おんなのこ	き きのした	さ さ	り りくじょう	さ さけのむ											
比 比喩	嗅 嗅ぐ	嘲 嘲ける	餌 餌	溺 溺れる	女 女の子	木 木の下	空	陸 陸上	酒 酒を飲む											

採点してみました。

次の平仮名を漢字に直しなさい・字べてあげなさい

月日	名前	学年	採点	氏名

図1 学生が採点した漢字テスト

これをさらにグループで相談させると「厳しい採点に傾く」実態となる。厳しい採点をする学生は、漢字の細部について標準意識が高く、他の学生は「この人は漢字に自信があるのだな」と認識するようである。

グループでの話し合いの結果

を代表者に発表させると「そこも間違いなのか。すごい。」というような「称賛の声」が上がる。そのあとに常用漢字表の前書きを見せることで、自分たちの意識の中にあった漢字の正誤に関する基準が、実は経験則にのっとりたものにしかすぎなかったことを理解するのである。

学生たちの意見によれば、彼らの正誤基準は教師の評価の結果というより漢字ドリルや教科書の書体がどのようなであったかということらしい。しかし、同じ教科書体であっても教科書によって活字は異なる。端的な例では、平仮名の「そ」の始線を離すか続けるかは受講者全体をちょうど半数に分けることになる。これは、学生たちがどの教科書を使って学習したかによる結果である。そのようなことを話し合っていくと、学生たちは徐々に漢字は書体によって代表化されているにすぎない「字の骨格」であることに気付いていく。そもそも教師の中にも常用漢字表の前書きを目にしたことがない者は少なくないだろう。

¹ 奈良教育大学における教養科目「文字の世界」の中で毎年実施。本稿における資料は平成 22 年度のもの（受講者 46 名）。

² このテストにおいて、明らかな誤りは下段⑱「嘲る」の送り仮名のみである。

特に小学校の教師は、中学校、高等学校とは異なり全教科について指導していかなくてはならない。教師自身も学生たちと同様、自身の経験、もしくは指導する教科書の書体に左右された字形指導を行っていくであろうことは推測にかたくない。問題はそのことが漢字嫌いの子どもたちを作っていく可能性が大きいことである。

小林（1981）は漢字を国語科内に「閉じていく」ことの不合理性をすでに30年も前に説いている。「国語の時間に、国語科の教材としてなじむ文種で、それに現れる字種で漢字の知識を与え、読み書き出来るようにさせるのは基礎的基本的なところに立ち是非必要なだけけれども、現代の日常の国語生活で用いられる語彙や漢字の字種を背景とした場合には（あるいは更に文化の総体を背景にした場合には）、多くの語彙を与えていく、ルビ付きでよいから多くの漢字に視覚で触れさせ慣れさせていく必要があるという仕方、考え方を採りたい。」とし、日常生活に生きる漢字指導の在り方に言及している。小林は上記引用の中で「漢字の知識」「読み書き」について触れ、漢字学習の基礎として字形を習得させることの重要性を解いてはいるが、細部にこだわらないことの重要性もまた、別の文献で繰り返し述べている。（小林（1998））

2. 3 漢字から離れる子どもたち

前述したように漢字に対する児童の意識は二極化している。このことは、国立政策研究所（2008）の「特定の課題に関する調査」でも結果として現われている。「特定の課題に関する調査」は、小学校第4学年から中学校第3学年を対象に行った調査で、学習指導要領改訂に際して行う「教育課程実施状況調査」のうち、特に課題だと思われる点に内容を絞ったものである。2008年の調査で国語は、「漢字」と「長文記述」が対象となり、同時に質問紙による意識調査を行っている。以下の表2は「設問1（1）漢字の学習が好きだ」に答えた結果である。³

表2 特定の課題に関する調査における漢字の好悪について

回答状況	そう思う		どちらか といえば そう思う		どちらか といえば そう思わない		そう思わない		分からない		その他		無回答	
	人数	割合 (%)	人数	割合 (%)	人数	割合 (%)	人数	割合 (%)	人数	割合 (%)	人数	割合 (%)	人数	割合 (%)
小学4年	849	26.6	1,149	35.9	648	20.3	363	11.4	177	5.5	0	0	11	0.3
小学5年	725	22.0	1,177	35.7	817	24.8	432	13.1	146	4.4	0	0	2	0.1
小学6年	664	20.5	1,093	33.7	856	26.4	493	15.2	139	4.3	1	0	0	0
中学1年	547	17.1	1,049	32.8	912	28.5	569	17.8	121	3.8	0	0	2	0.1
中学2年	479	14.8	988	30.6	912	28.2	718	22.2	132	4.1	0	0	2	0.1
中学3年	562	18.6	959	31.7	797	26.3	596	26.3	107	3.5	2	0.1	3	0.1

³ 掲載にあたってグラフとともに棚橋が成形した。その際、合計欄に当たる「全体」の項目は削除した。

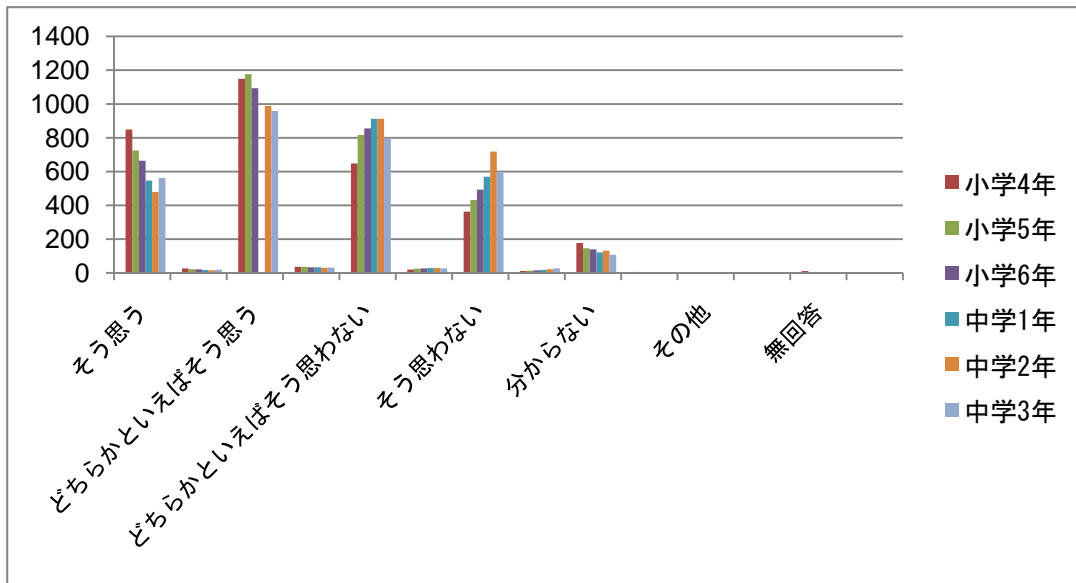


図2 特定の課題に関する調査における漢字の好悪について（人）

「二極化」の詳細を見ていくと、学年が上がるにつれて「そう思う（好きだ）」という回答者は減少し、逆に「そう思わない（嫌いだ）」という回答者が増加する傾向にあることが理解できる。

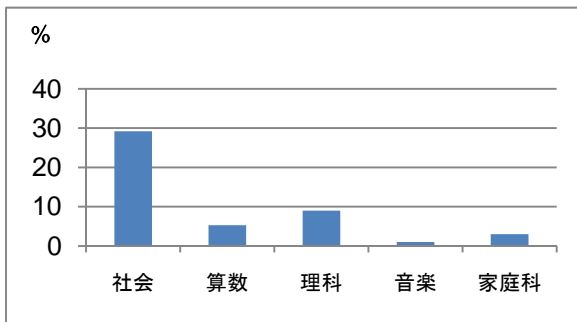


図3 漢字が多くて嫌いな教科（％）

また、棚橋（1999）が、小学校第6学年の児童301名を対象に行った調査では、漢字が多くて嫌だと思ふ教科があると回答した児童は、左の図に示すように社会29.2％、算数5.3％、理科9.0％、音楽1.0％、家庭科3.0％であった。

この調査の際には、同時に漢字テストを行い、その結果と「漢字が多くて嫌いな教科」との関係も調べた。図3はその結果である。

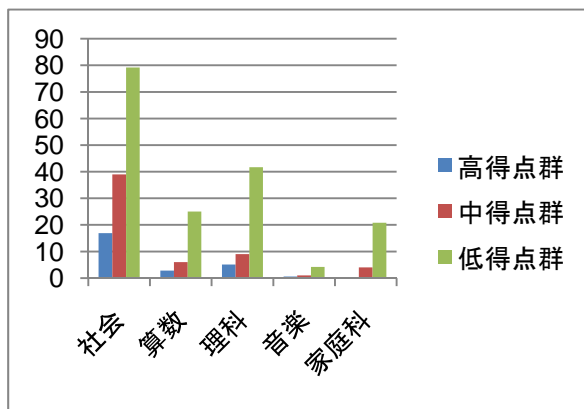


図4 得点別「漢字が多くて嫌いな教科」（人）

テストの結果は得点分布を勘案し、高得点群、中得点群、低得点群の3層に分けた。図をみると、漢字の得点が低いほど「漢字のためにその教科の学習を嫌い」とする児童がいずれの教科でも多くなっている。この調査は、漢字の習得度と教科内容の定着度を調べたものではなく、あくまでも漢字の習得度と、教科内容に対する「好き嫌い」を問うたものである。ただ、この結果からは、漢字の習得が教科の学力とかなり強い相関にあるということが言えるだろう。

3. 漢字教育の目指す方向とコーパス

3. 1 漢字指導の目指すもの

平成 20 年告示の学習指導要領が示したように、漢字は「使えるようになる」ことが重要である。それは、教科書に提示された形の熟語だけでなく、その漢字を含む多くの語句が読み書きできるということである。教育現場ではよく「教えたのに、使えない。」という言葉が聞かれるが、ある意味それは当然である。字形指導に力点を置いていけば、単漢字および教科書に載っている語句程度しか認識できるようにはならないのである。

使えるようにするためには使えるようにするための指導が必要であるし、語彙を広げるには語彙を広げるための指導が必要である。漢字をできるようにするためには、先に引用した小林（1981）でも「多くの語彙を与えていく、ルビ付きでよいから多くの漢字に視覚で触れさせ慣れさせていく必要がある」と述べていたように、繰り返し漢字熟語に出合わせる必要がある。見慣れることで漢字が習得できることについては、棚橋（2007）の「漢字の読み書きにおけるルビの有効性」を解明する実験でも証明されている⁴。この実験で棚橋は児童を実験群と対象群とに分け、実験群の児童には調査のために作成した教科書テキストを与え、授業を受けさせた。作成したテキスト本文のうち 10 語は教科書とは異なるルビ付き漢字表記に改めた。対象群の児童については通常通り教科書で学習させた。単元後にその 10 語句の書きとりを行った結果、読みばかりか書きについてもルビ付きでその漢字を見慣れさせるようにした実験群の児童の結果が対象群の児童の結果を上回るようになった。漢字を何回も見るとは、その漢字が読めるようになるばかりでなく、書けるようになることにもつながっていく。これらのことを踏まえて考えると、これからの漢字指導はより多くの漢字語句に計画的に出合わせるが必要になってくると言える。

3. 2 「見慣れること」を組織するために

3. 2. 1 BCCWJ の初等教育における活用可能性の限界

漢字は、戦前戦後一貫して国語科において指導してきた。しかしながら、国語科での指導には実際のところ限界も多い。もっとも大きな点は、国語科の教材配列が学習指導要領に即した内容を有する文学作品や説明的文章を中心に組み立てられていて、漢字および漢字熟語は教材の採用状況によって偶然に提示される点にある。かつて輿水（1971）の示した読解の「基本的指導過程」における漢字学習の位置づけの特徴の一つには、「漢字を、文章を読む前に取り出して教えてしまわない—センテンス・メソッドの原則を守る。」とあるが、⁵国語科教科書において、その立場を貫こうとするとかなり無理な事態が引き起こされる。輿水らの提案後、学校教育は過激な受験戦争、「落ちこぼれ」の増加などの問題から「ゆとり」を重視する方向に傾き、学習指導要領が改訂されるたびに「教育内容の精選」、「厳選」、と指導事項が縮小される方向に向かった。国語について言えば、単元数が少なくなり、それとともに国語科教科書の総語彙数も少なくなった。漢字の場合も、従前は読み替え漢字の音訓もすべて教科書本文で提示していた A 社教科書では、低学年を除いて、一つの漢字が提示されると音訓すべての読みを提示するようになった。つまり、教科書語彙数の減

⁴ 調査対象 小学校第 3 学年 449 名、第 4 学年 424 名、第 5 学年 495 名、第 6 学年 331 名、計 1,699 名。

⁵ 先述の棚橋（1998）の調査では、教師の多くは読解単元の前に漢字を取り出して指導する実態がある。

少から読み替え漢字を含む漢字語句の提示が困難になったのである。このような実態のある国語科の授業で漢字を生活に即した形で教えるのは、はなから問題があると言える。

平成 20 年告示の学習指導要領では、国語科改訂の要点として「読書活動の充実」を挙げている。NHK 放送文化研究所（2006）によれば、2005 年当時小学生が平日新聞を読む時間は平均 1 分、雑誌・マンガ・本でも平均 18 分であった。近年は「朝の読書活動」などの全国的な取組みもあり、特に小学生については読書傾向が拡大しているようだが、それだけで計画的な漢字習得ができるかは少々疑問がある。

本研究では当初、児童に多くの漢字語句に出会わせることを目的として、『現代日本語書き言葉均衡コーパス』（BCCWJ）を活用して学習用漢字表を作ることを構想していた。社会の漢字使用実態にあわせ、学年別配当漢字を含む語句と文章とを提示することで、学習者に社会生活につながる漢字指導をしようとしたのである。しかしながら、そこには看過しがたい問題が存在した。一つは BCCWJ で扱っているテキストのほとんどが内容、用語、論述形式ともに小学生の言語実態から乖離しており、小学生用の漢字表を作るという点からは無理があることである。

以下の表 3、4 は BCCWJ を使って第 6 学年に配当された教育漢字の「権」と「衆」の使用例を前・後文脈とともに提示順に上から 10 例並べたものである。一見して小学生向けの文章としては内容が専門的または高度すぎると判断できる。

表 3 BCCWJ による「権」の検索結果

前文脈	検索文字	後文脈
ものであるが、この私案もまた第一章を「皇	権	」として、「天皇」の規定から始まり、その
戦争に与してきた。それが、小泉 - 安倍自公政	権	がこれまでとってきた道であり、福田政権は
知られる。吉備なり筑紫なりの軍勢を大和政	権	が朝鮮での戦闘に動員するといっても、それ
する叫びであった。蒋介石はファシスト政	権	であり、この中国型全体主義は毛沢東のそれ
に認められた精神的損害に対する慰謝料請求	権	といった権利がこれにあたる。保存行為と
則精説 295 など)、そこでは暖簾は営業	権	と称され、企業の将来期待しうる超過収益力
か。新九郎は考え込んだ。いま頼芸政	権	には中心となる人物がいない。守護と守護
いただきます。それは、戸塚ゴルフクラブの会員	権	に関する件でございますが、税務の観点から
ある。国際人権だけでなく、人権や国家主	権	の侵害の問題を含む安保条約や N A T O の空
契約関係がない場合にも発生しうる。留置	権	の基本的性質として、①留置権者は債権全額

表 4 BCCWJ による「衆」の検索結果

前文脈	検索文字	後文脈
を推進させた。副次的効果として、IT の大	衆	化が進み、情報通信関連機器やソフトウェア
中だから、明日の朝まで待てないか」お側	衆	が云った。「洪水のおそれがございます。
農村青少年を農業研修生としてアメリカ合	衆	国に 2 年間で 1 年間派遣する 2 つのプログラ
りまして、露天のお風呂なんかに土地の芸者	衆	とか、若いモデルさんとかが入っている写真を
えなおすと、張儀が、「勝負はみせかけの	衆	寡では決してないといったのは、おまえではな
言っている一面もある。新聞の読者という大	衆	に対して、文学的な名文の分析と研究をして
とともに慈悲の利他行を重んじ、広く一切の	衆	生の救済を目指して献身する菩薩((2))
既にこの問題、	衆	議院でも相当突っ込んだ論議をされておりま
「なにをしておる！」怒声をあげて葛城	衆	の人垣に飛び込む蓮長の頬に、拳が炸裂した
公	衆	衛生学は社会医学に属するものでございます

もう一つは、一つ目の問題とも関係するが、これらの生データを見せたとしても小学生の興味関心には堪えないということである。小学生に提示して授業をする場合、情意面の指導（意欲の喚起）を無視することはできない。したがって、BCCWJは小学生を対象とした初等教育においては、教員の指導用資料としては活用できてもそのまま児童用の教材とするのは難しいと判断した。

3. 2. 2 教科書の活用を視野に入れた教科書コーパスの援用

言語政策班では、教育に生かすために平成17年度に出版された全教科書⁶のテキストに基づいて「教科書コーパス」（本報告書第1章第1節参照）を作成した。この教科書コーパスを援用することで、小学校の漢字指導について有効な提言ができることがわかった。

先に児童生徒の読書時間の少なさについて述べた。この点について2000年のPISA調査における質問紙法では、家庭における読書環境についても問うているが、それによれば日本の家庭における蔵書数は決して多くなく、そもそも子どもたちは本を目にする場面が少ないということが顕わになった。（国立教育政策研究所（2001））

そこで、国語科以外の教科書を活用して漢字語句の習熟を図ることを構想した。教科書は日本のすべての児童が手にする書籍であり、家庭や学校の蔵書環境に関係なく手にすることができる。さらに、国語科以外の教科書においては、国語科よりも提示される漢字、および漢字語句が多く存在する。次ページの表5、表6は第6学年に配当された漢字が社会科、理科の教科書にどれくらいの頻度で提示されるか、社会科と理科の頻度の降順に並べ替えたものである。20位までを示したが、この範囲の漢字であれば、国語科より特に社会科は圧倒的に頻出する状態にあると理解できる。

このような「他教科に頻出する漢字」は、国語科では学年別配当漢字として、「話すこと・聞くこと」、「書くこと」、「読むこと」といった国語科カリキュラムの本質とは関係なく意図的に提示している場合が多い。配当されている事実があるからいわば強引に提示されるのである。教育漢字の学年配当の基本方針は昭和27年に発足した漢字学習研究会で決めた基準によっている。その基準をとりまとめた文部省（1957）は、漢字の提出時期について「他教科の学習に必要な漢字は、その教科の進度を考えて出す。」としており、表に示された漢字の一部は、教科内容とのかかわりを強く持つ漢字だと判断できる。

一方、表7は、社会科において提示頻度が一位であった「権」について、教科書上にどのような語句が使用されているかを示したものである。表を見ると、国語科では「権利」と「人権」の二つの語句のみの提示であるのに対して、社会科の異なり語数は11になる。

社会科で提示された11語句のうち、学習指導要領上に明示された語句は、権利—「日本国憲法は、国家の理想、天皇の地位、国民としての権利及び義務など国家や国民生活の基本を定めていること。」、主権—「我が国の政治の働きについて、次のことを調査したり資料を活用したりして調べ、国民主権と関連付けて政治は国民生活の安定と向上を図るために大切な働きをしていること、現在の我が国の民主政治は日本国憲法の基本的な考え方に基づいていることを考えるようにする。」、参政権—「また、イの「国民としての権利及び義務」については、参政権、納税の義務などを取り上げること。」の三語句であり社会科の指導内容の中核にかかわる語句だと言える。このような語句については社会科での漢字学習が妥当である。その理由は、まず「見慣れ」の機会をより多く持つこと、さらに

⁶ 小、中、高等学校の全教科対象。各教科とも原則的に採択率1位の教科書をサンプルとした。

表5 第6学年の漢字頻度—国語 - 社会

No	字種	全教科	国語	社会
1	権	104	4	95
2	皇	81	3	78
3	憲	75	2	73
4	幕	66	2	64
5	将	60	5	52
6	城	69	9	51
7	域	90	4	50
8	遺	68	23	42
9	展	77	8	38
10	聖	38	1	37
11	障	38	3	30
12	貴	29	4	25
13	忠	26	1	25
14	敬	26	2	24
15	探	46	10	23
16	割	92	12	22
17	源	44	5	21
18	濟	19	3	16
19	衆	21	1	16
20	革	16	1	15

表6 第6学年の漢字頻度—国語 - 理科

No	字種	全教科	国語	理科
1	層	83	1	78
2	磁	68	2	64
3	灰	51	3	48
4	吸	80	10	28
5	砂	39	5	24
6	肺	33	2	23
7	呼	50	10	22
8	蒸	24	2	20
9	臓	44	1	19
10	割	92	12	17
11	針	70	3	11
12	骨	23	9	10
13	域	90	4	9
14	宇	17	3	8
15	筋	16	3	8
16	宙	17	3	8
17	棒	26	2	8
18	株	9	1	7
19	収	17	3	6
20	卵	14	3	5

表7 第6学年社会科における提示語句

熟語および漢字	全教科	国語	社会
権利	38	3	34
人権	22	1	19
権	14	0	12
主権	9	0	9
民権	9	0	9
執権	3	0	3
法権	3	0	3
権限	2	0	2
参政権	2	0	2
権力	1	0	1
実権	1	0	1

漢字の学習が教科内容の理解を強化すると考えられることによる。そのような立場から本研究では、小学校6学年児童を対象にした実践を行った。それについては本報告書第4章第4節に述べる。

4. おわりに

本研究に携わった5年間は試行錯誤の連続であった。もっとも苦心した部分はBCCWJの漢字使用実態を漢字習得・語彙習得の過程にある初等教育に連動させようとした点にあった。レベルこそ違うが、戦前、小学校での漢字習得が社会では役立たないと言われたような事態が、そこには存在した。使われている漢字は同じでも、語彙のレベル差が存在し、提示された文脈は小学生には十分に把握できない場合が多かったからである。しかし、言語政策班で教科書コーパスを作成していく中で、他教科教科書の漢字習得上のテキストとしての価値が浮上してきた。

もともと、配当された漢字を教科書コーパスで検索していて、教科書の中では「宇」が「宇宙」という語句ぐらいしか作りえないことに気付いたことが、きっかけだったと感じる。逆に言えば「宇」は、教科（この場合理科）と結びついた漢字と言えるのではないか。班会議での発表を通し、それぞれの教科にはそれぞれの専門と結びついた「特徴的な漢字がある」という点が見えてきた。この点は言語政策班の近藤明日子氏の「教科特徴語」（本報告書第3章第1節）の考え方に示唆を受けた部分が多い。本研究の過程では、「特徴漢字」についての論考を学会発表した⁷が、データ分析などテクニカルな面で近藤氏にお世話になったことをここに記したい。

さて、教科と結びついた漢字の存在は、もともと文部省サイドの方針でもあったため、教師間でも経験的には認識されていた事実である。しかし、今回、コーパスを使って、頻度や特徴度を調査していくと、思っていた以上に国語科以外の他教科で頻出する漢字が見られた。

図5からは、現職教員の他教科の漢字提示についての認識のありようが見られる。母集団が少ない調査であるので、確定的な判断は難しいが、3分の2の教師が、他教科において頻出する漢字があることを認識している。

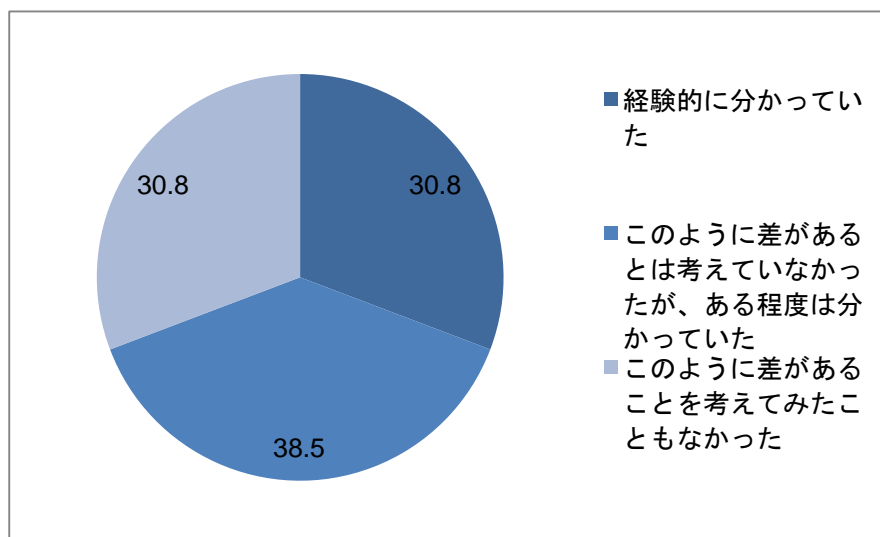


図5 社会科で多く提示される漢字についての教師の認識⁷

私が提案するように、他教科において漢字指導をすることについては賛否両論があった。詳しくは本報告書第4章第4節で述べるが、当初予想していたよりも肯定的な見解が多か

⁷ 現職小学校教員13名に対する質問紙調査。本論文の表5の一部を見せて尋ねている。

ったという点は、本研究をさらに進めようとしていく上で少し頼もしく感じられる点であった。参考のため以下に自由記述を原文のまま掲載する。

- ・画数が多く難しい漢字が多いので、意味があると思う。しかし、授業時間等を考えると継続的に指導に組み込んでいくことは難しいのではないだろうか。
- ・国語のように指導するとなると筆順や熟語作りや短文作りとなり、かなりの時間を要するため、「国語のように」とは難しいと思います。よって国語では学習した漢字が提出されている時に指導者が子どもたちに例示することは大切だと考えています。漢字が使える場面を示したり使う方法を示したりして気付かせることで、漢字を用いる子どもに育っていきと思っています。また、教科性のある漢字や社会生活でよく使われる漢字は提出学年ではなくても漢字で示してふりがなをつけるようにしています。いわゆる「ませ書き」はしないようにしています。
- ・必要があれば指導するが、通常、国語科で先に出てくれば国語科で指導という形が多いように思う。社会科で先に出てくれば、書き順や意味は指導するが、熟語や文章作りなどはしていなかった。国語科以外の教科で指導するほうが、定着するとは思いますが、国語科の学習と重複するので、時間が限られた今の状況では、なかなかできにくいようにも思う。
- ・初めて学習する時、社会科で出てくる場合は、国語と同様に取り扱い、くわしく教える必要がある。小学校は学級担任制なのでそれが可能であるし、それを意識して指導にあたっていきたい。
- ・各教科により、指導する内容の重要性が違うので、国語以外の教科のときに漢字を詳しく指導する必要はないと思う。
- ・漢字の学習を主眼に置いたとき、教科の枠をこえて漢字の学習を行うことは必要だと考える。また、子どもにとっても漢字が定着しやすい。ただ、時間の確保が難しく、なかなかできていないというのが現状である。
- ・社会科の学習で理解させるためには、指導が必要と考えます。
- ・度々出てくるのであれば、漢字の意味などを知っておくことは良いことだと思うし、子どもから知りたいと思ってくれればなお良いと思います。
- ・子どもたちが知識を習得するという点では、国語以外でも指導する必要性はあると思います。
- ・単に漢字を学習するだけでなく、日本文化を学習するという点でも、その時々に出てくる漢字を教えることは必要だと思います。
- ・国語科のように漢字だけを取り上げ学習するのは、時間的に難しいのでは。
- ・子どもたちの定着を図るのであれば仕方がないのでします。

さて、小学校の場合、教科の専門性が高まるのは高学年の学習においてである。今後は、小学校5,6年の全教科におけるカリキュラムを漢字指導との関係の中で考察していきたい。さらに、第5学年、第6学年に配当された漢字の教科割り当てを検討し、学校教育の現場に有用な資料として提案していこうと考えている。

文献

- 小野瀬雅人（1999）『振り仮名の教育心理学的研究』、野間教育研究所、p.69.
- 旧言語問題談話会同人（1937）（棚橋注；特定署名なし）「新しい国語協会の生れるまで」国語運動、1巻1号、国語協会、p.90.
- 棚橋尚子（1998）「主体的な言葉の学び手の育成に向けて一語彙指導としての漢字教育を考える一」国語教育研究、329、日本国語教育学会。

- 小林一仁（1981）『漢字指導の基礎研究』，明治図書，pp.12-13.
- 小林一仁（1998）『バツをつけない漢字指導』，大修館書店.
- 国立教育政策研究所（2008）「特定の課題に関する調査（国語）集計結果（PDF）」
<http://www.nier.go.jp/kaihatsu/tokutei/04002010000007002.pdf>
- 棚橋尚子（1999）「漢字ルビに対する児童の反応—実態調査を通して—」国語教育探究，第12号，国語教育探究の会.
- 興水実（1971）「漢字学習指導改造の視点と課題」国語教育研究所編『漢字の読み書き分離学習』，明治図書，p14.
- 棚橋尚子（2007）「漢字習得におけるルビの有効性の解明」『奈良教育大学国文—教育と研究—』30，奈良教育大学国文学会.
- NHK 放送文化研究所編（2006）『日本人の生活調査・2005 NHK 国民生活時間調査』，NHK 出版，p.96.
- 国立教育政策研究所（2001）『生きるための知識と技能 OECD 生徒の学習到達度調査（PISA）2000年調査国際結果報告書』，ぎょうせい.
- 棚橋尚子（2009）「教科に特徴的な漢字に関する考察—他教科における漢字指導の可能性—」国語科教育，67，全国大学国語教育学会.
- 文部省（1957）『教育漢字の学年配当 漢字学習実験調査報告』，教育出版，p.8.

第4章

国語の教材研究と実践事例

第4章 第1節

「少年の日の思い出」の語彙指導

田中 牧郎

1. 背景と目的

『現代日本語書き言葉均衡コーパス』や「教科書コーパス」あるいはこれらのコーパスを活用する関連ツールは、国語教科書の教材研究や国語科の授業実践にも、直接役立てることができるのではないかと考えられる。本節では、中学校の国語教材として多くの教科書に採られている「少年の日の思い出」（ヘルマン=ヘッセ作、高橋健二訳）の語彙指導についての教材研究において、コーパスや語彙表を使うことを検討してみたい。

「少年の日の思い出」は、国定読本の頃から国語教科書に採られている作品で、いわゆる定番教材として、現在も中学校 1 年生用の国語教科書の多くに掲載されている。少年時代に蝶の採集に夢中になった「僕」が、隣に住む少年の蝶を盗んでしまい、その罪を償うことができないまま、自分の蝶の収集を粉々にしてしまったことを大人になった時点から物語る内容である。ヘルマン=ヘッセの作品としては著名なものではないが、高橋健二の日本語訳は、登場人物の心情を読み取らせる読解教材として高く評価され、長い間使われ続けている。

こうした定番教材は、教師や教材開発者によって様々な角度から研究が重ねられ、指導の実績も豊富であり、国語教育界に指導上の様々な知見が蓄積されてきているはずである。しかし、教材化されたこの作品を、語彙指導という観点から教科書や教師用指導書で見ると、学習すべき語句あるいは指導すべき語句として指示されているものについて、語句の何を学習すべきなのか、どのような観点で指導すべきなのかが明示されていないものが多い。また、新出漢字として取り上げられている漢字や語句についても、この作品でそれを取り立てることの必然性が分かりにくい場合がある。さらには、この作品を使った学習活動の中心にあると思われる、登場人物の心情の読解と語彙との関わりについても、教科書や指導書の指摘だけでは不十分に思われる部分がある。つまり、定番教材であっても、語彙指導についての知見は誰でも十分に使いやすいようには体系化されているわけではないように思われる。本特定領域によってコーパスが整備され、現代日本語の語彙を体系的に把握することが可能になりつつあるいま、語彙指導の教材研究の知見を体系化し新しい段階に進めることができないだろうか。

以上のような背景と問題意識のもと、「少年の日の思い出」の語彙指導に関する教材研究に、コーパスを活用する方法を考えようというのが、本節の直接の目的である。以下で具体的に試みるのは、次のような教材研究である。

- (1) 教材本文を形態素解析したものに、第 1 章第 4 節に記した「学校・社会対照語彙表」の情報を関連づけた、「少年の日の思い出_語彙表」を作成する。
- (2) 「少年の日の思い出_語彙表」をもとに、教科書や指導書に指示されている「注意する語句」を検証する。
- (3) 「少年の日の思い出_語彙表」をもとに、教科書や指導書に指示されている「新出

漢字」を検証する。

- (4) 「少年の日の思い出」における類語指導のための教材研究に、コーパスの用例分析を導入する。

以下、この順に具体的に述べていく。

2. 「少年の日の思い出」の語彙表作成

まず、「少年の日の思い出」の作品本文に対して形態素解析辞書 UniDic を用いて形態素解析を行い、短単位で語彙頻度表を作成した。作品の本文には、光村図書刊行の中学校 1 年生用の国語教科書（2006 年度版）の指導書に添付されている CD-ROM 収録の電子テキストを用いた¹。集計の結果、延べ語数 3,898、異なり語数 782 となった（記号・付属語・固有名詞等も含む）。この語彙頻度表に、教科書で「注意する語句」「新出漢字」と指示されている情報を加え、さらに「学校・社会対照語彙表」における、図書館サブコーパス（固定長データ、LB_FL）の語彙レベル、知恵袋サブコーパス（可変長データ、OC_VL）の語彙レベル、特徴教科の情報を加えた。こうした情報が増補された語彙表を「少年の日の思い出_語彙表」と称することとし、五十音順に配列したものの冒頭 1 ページ分を別表として本節末に掲げた。

この「少年の日の思い出_語彙表」の見方を説明しよう。列の左側から順に、「語彙素読み」「語彙素」「品詞」「語種」は、語彙の基本情報として、UniDic に登録されている情報をそのまま掲出した。その次の「度数」の列は、「少年の日の思い出」に出現する度数（ID 1 の「間」の場合は 3 回出現する）を示している。そして「語句指導」の列には、光村図書の教科書で「注意する語句」としているものには「注意」（「生憎」「依然」など）、「新出漢字」を含む語句としているものには「漢字」（掲出部分にはないが「獲物」「丘」など）と記した。さらに、「レベル_LB」の列には LB（図書館書籍）の語彙レベル、「レベル_OC」には OC（Yahoo!知恵袋）の語彙レベルを、それぞれ a～e の 5 段階（当該サブコーパスに 1 回も使われていない場合は「-」）で示した。「特徴教科」の列には、「教科書コーパス」の各教科の頻度を LB（図書館書籍）と比較した場合に当該語が特徴語となる教科名の略称を記した。こうした語彙表をつくることで、教材全体の語彙を把握し、一般の語彙や他教科の語彙との関係にも目配りできる、語彙指導の基礎資料とすることができる。

3. 「注意する語句」とされる語句の検証

3. 1 「注意する語句」の語彙表

光村図書の中学校国語教科書は、教材本文の掲載される各ページの下方に脚注欄を用意し、そこに「注意する語句」を本文から抜き書きして記す工夫を施している²。ここに掲出された語句については、教科書の凡例に「意味を調べ、使い方に慣れよう」とあり、指導

¹ 光村図書の中学校国語教科書は「教科書コーパス」の対象にもなっており、「少年の日の思い出」もこれに含まれる。「教科書コーパス」は 2005 年度版を使っており、この年度の教科書では「少年の日の思い出」は「好きな作品を選び自ら取り組もう」とされる位置に掲載され、教師が教室で扱う教材からは外されているため、十分な教材化が行われていない。学習内容が増加した 2006 年度版では、この作品を本格的な教材に格上げし語彙指導の内容も盛り込まれている。本節では、教材化された「少年の日の思い出」を扱いたいので、2006 年度版を用いた。

² この方式で教材中で語彙指導するものを明示するのは、多くの教科書で共通している。

書に意味の説明が掲載されているので、授業時などに取り立て指導される語句だと考えられる。「少年の日の思い出_語彙表」を用いて、教科書に指示されている「注意する語句」の妥当性の検証を試みよう。

「注意する語句」は、「少年の日の思い出_語彙表」の「語句指導」の列に「注意」とあるものがそれに相当する。全部で24語あるが³、そのすべてを抽出し、第1キーを「レベル_LB」、第2キーを「レベル_OC」、第3キーを「語彙素読み」として並べ替えたのが表1である。この表を一見すると、品詞は様々であり、語種は和語と漢語にわたっており、多様な語句が「注意する語句」に選ばれている様子が見て取れる。

ここでは、「レベル_LB」と「レベル_OC」に注目したい。この「レベル」とは何であり、どのような利用法が考えられるかについては、第2章第1節に詳しく記したが、その要点を記すと次のようになる。

- ・公共図書館による選書を経た書籍での語彙頻度を反映した「レベル_LB」は、国語政策・国語教育における一般的な「重要度」の指標と扱うことができる。
- ・書き言葉コーパスである BCCWJ の中では、最もくだけたスタイルで身近な内容を話題にすることも多いインターネット掲示板での語彙頻度を反映した「レベル_OC」は、「日常度」の指標と見ることができる。
- ・上記のことを踏まえ、「レベル_LB」の a・b・c の語彙を「重要語」の範囲と扱い（重要度は、a・b・c の順に高い）、「レベル_OC」の a・b の語彙を「日常語」と扱う作業仮説を立てることができる。

表1の24語の2種のレベルを見ると、次のような傾向があることに気付かれる。

- ・「レベル_LB」は、a・b・c が大部分で、これに外れるものは4語のみである。
- ・「レベル_OC」は、c・d・e・ーが大部分で、これに外れるものは2語のみである。

このことは、「注意する語句」に選ばれている語句は、おおむね「重要語」でありかつ「日常語」ではないという性質を持つものであることを示していよう。

³ 他に、「身にしみる」「・・・よしもない」「例の」「・・・を盾に」の4語が「注意する語句」とされているが、いずれも UniDic の短単位よりも長い単位であり、この単位で取り出していないため、この表には入っていない。

表1 「少年の日の思い出」の「注意する語句」

ID	語彙素読み	語彙素	品詞	語種	度数	語句指導	レベル_LB	レベル_OC	特徴教科
596	ビミョウ	微妙	形状詞-一般	漢	3	注意	a	a	芸
555	ハタシテ	果たして	副詞	和	1	注意	a	c	
690	ムスウ	無数	形状詞-一般	漢	1	注意	a	d	
746	ヨクボウ	欲望	名詞-普通名詞-一般	漢	2	注意	a	d	
748	ヨミガエル	蘇る	動詞-一般	和	1	注意	a	d	
39	イゼン	依然	副詞	漢	1	注意	a	e	
587	ビショウ	微笑	名詞-普通名詞-サ変可能	漢	1	注意	a	e	
229	コクハク	告白	名詞-普通名詞-サ変可能	漢	1	注意	b	a	
214	ケッテン	欠点	名詞-普通名詞-一般	漢	1	注意	b	c	
733	ユウワク	誘惑	名詞-普通名詞-サ変可能	漢	1	注意	b	c	
187	キョクド	極度	名詞-普通名詞-形状詞可能	漢	1	注意	b	d	
211	ケツカン	欠陥	名詞-普通名詞-一般	漢	1	注意	b	d	技
308	シンピ	神秘	名詞-普通名詞-形状詞可能	漢	1	注意	b	d	芸
3	アイニク	生憎	副詞	和	1	注意	b	e	
259	サシズメ	差詰	副詞	和	1	注意	c	-	
768	レイタン	冷淡	名詞-普通名詞-形状詞可能	漢	1	注意	c	-	国
53	イロアセル	色褪せる	動詞-一般	和	1	注意	c	d	
474	トリコ	虜	名詞-普通名詞-一般	和	1	注意	c	d	
609	ヒンジャク	貧弱	名詞-普通名詞-形状詞可能	漢	1	注意	c	d	
689	ムサボル	貪る	動詞-一般	和	1	注意	c	e	
527	ネホリ	根掘り	名詞-普通名詞-一般	和	1	注意	d	d	
565	ハホリ	葉掘り	名詞-普通名詞-一般	和	1	注意	d	d	
389	タンショウ	嘆賞	名詞-普通名詞-サ変可能	漢	1	注意	e	-	
767	レイゼン	冷然	形状詞-タリ	漢	1	注意	e	-	

3. 2 「少年の日の思い出」の全語彙の中での「注意する語句」の位置

2 で作成した「少年の日の思い出_語彙表」の全語彙の中で、3 で見た「注意する語句」はどのような位置にあるのだろうか。重要度と日常度の指標として用いているレベル_LBとレベル_OC とを掛け合わせて、その語数（異なり語数）を整理したものが表 2 である。表 2 で LB と OC を掛け合わせた各セルの右上の細字の数値が全語彙の語数、左下の太字の数値が「注意する語句」の語数を示している。

表 2 を全語彙の語数について全体的に見ると、まず、LB においても OC においても、レベル a が最も語数が多く、b・c・d・e・ーと進むにつれて少なくなっている。そして、表全体の左下方よりも右上方に数字が偏在しており、相対的に OC のレベルよりも LB のレベルの方が強く作用していることが分かる。「少年の日の思い出」の語彙は、日常度よりも重要度と強く相関すると言い換えてもよいだろう。

表2 全語彙と「注意する語句」の語彙レベル（右上：全語彙 左下：注意する語句）

OC LB	a		b		c		d		e		-	
	a	<u>1</u>	349		97	<u>1</u>	45	<u>3</u>	7	<u>1</u>	4	
b	<u>1</u>	4		20	<u>2</u>	37	<u>3</u>	27	<u>1</u>	18		3
c						4	<u>3</u>	16	<u>1</u>	35	<u>2</u>	7
d							<u>2</u>	4		11		3
e				1						5	<u>2</u>	10
-										2		21

表3 「注意する語句」が占める比率

OC LB	a	b	c	d	e	-
	a	0.3%		2.2%	42.9%	25.0%
b	25.0%		5.4%	11.1%	5.6%	
c				18.8%	2.9%	28.6%
d				50.0%		
e						20.0%
-						

次に、表2で「注意する語句」の分布に目を移すと、これも表全体の右上方を中心にかなり広い範囲に広がっていることが分かる。全語彙の中での位置を見るために、2つのレベルを掛け合わせた各セルにおいて、注意する語句が全語彙の何パーセントを占めるかを計算したものが、表3である。表3によれば、LB（重要度の指標）ではレベルa・b・cが多く、OC（日常度の指標）ではレベルd・e・-の比率が高いことが見て取れる。このことから、注意すべき語句は、比率から見ても、重要語であって日常語でないものが選ばれているということが確かめられる。

しかし、一部に、そうした傾向から外れるのではないかとと思われる語がある。表2の中の数字に下線を引いたものがそれにあたる。まず、OCのレベルがaのものが2語（「微妙」「告白」）ある。これらは作品中で何か特別の意味を持っていたり、重要な箇所用いられていたりのかもしれない。教科書や指導書などにはその旨は明確に記されていない（「告白」にはやや詳しい説明が指導書にある）が、もし、そのような特別な語句なのであれば、学習や指導の際に注意すべき観点を明示することが望まれるのではないか。それから、LBのレベルがeのものが2語（「嘆賞」「冷然」）ある。これは、中学生にはかなり難解

な語彙であると思われ、他の重要語のようにこの語句を深く指導するのは難しいように思われる。こうした難解な語彙を「注意する語句」に含める事情は、教科書や指導書に明示することが望まれよう。

4. 「新出漢字」の検証

中学校の学習指導要領には、小学校の「学年別漢字配当」(1006字)のような具体的な定めがなく、どの学年でどの漢字を指導するかには公的な規範はない。一方、漢字を漢字だけ取り出して学習させるのではなく、できるだけ教材作品中に使われている漢字を「新出漢字」として順次学習させる方法を取っているのは、小学校と同様である。このため、各教科書会社では、各学年に漢字を配当し、その漢字を教材に割り振り、通常の教材に出てこない漢字については、漢字を学習するための教材を用意し、そこでまとめて扱うようにしている。例えば、光村図書版では、第1学年に300字、第2学年に350字、第3学年に289字を配分し、中学3年生までに常用漢字1945字の全体を学習できるようにしている。したがって、どの学年のどの教材でどの漢字を新出漢字として割り当てるかは、漢字指導の基本的問題として重要である。

ところで、教材の中で新出漢字を指導する際は、その漢字が表記している語例とともに指導される。したがって、重要語句の表記に用いられている箇所でも新出漢字扱いができれば、漢字指導・語彙指導の両面において効果は高まるだろう。反対にあまり重要でない語句の中で新出漢字を扱っていると、その効果は低くなるだろう。そこで、「少年の日の思い出」で新出漢字が使われている語彙を観察してみることにしたい。

表4は、新出漢字が採られている語彙を別表の「少年の日の思い出」語彙表から抽出したもので、語彙素の列に下線を付けた文字が新出漢字にあたる。この表で「レベル_{LB}」の列を見ると、a・b・cがほとんどで、全体として、重要語から新出漢字が採られている実態が確かめられ、よく配慮されていると見ることができよう。ただし、「難癖」「胴乱」は、語彙レベルから見ると重要語の範囲を逸脱しており、問題があるのではないかと思われる。「癖」の字は、3年生の教材「握手」(井上ひさし著)に「右の人さし指をびんと立てるのが癖だった」とある、「癖」の語(LBのレベルa)などで新出漢字を扱うのがよいのではないだろうか⁴。また、「胴」は、レベルcまでの語彙が教科書の教材中がないが、2年生の教材「走れメロス」(太宰治著)に出てくる「胴震い」(LBのレベルe)の方が、物として分かりにくい「胴乱」よりも適切であるように思われる。あるいは、より一般的な語である「胴体」(LBのレベルc)などを、漢字教材に登場させることもよいかもしれない。

⁴ 後に見るように、この物語の転換点にあたる重要な一節の中に「難癖」は登場しており、この作品中で注目すべき語句であるという配慮から、採られている可能性はある。

表4 「少年の日の思い出」の「新出漢字」が用いられる語彙

ID	語彙素読み	語彙素	品詞	語種	度数	語句指導	レベル _LB	レベル _OC	特徴 教科
193	キンチョウ	緊張	名詞-普通名詞-サ変可能	漢	1	漢字	a	a	
260	サス	刺す	動詞-一般	和	1	漢字	a	a	国
553	ハズカシイ	恥ずかしい	形容詞-一般	和	1	漢字	a	a	
742	ヨウチ	幼稚	形状詞-一般	漢	1	漢字	a	a	技保
289	ジマン	自慢	名詞-普通名詞-サ変可能	漢	1	漢字	a	b	
439	トウ	塔	名詞-普通名詞-一般	漢	1	漢字	a	c	
82	オカ	丘	名詞-普通名詞-一般	和	1	漢字	a	d	
301	ショサイ	書斎	名詞-普通名詞-一般	漢	1	漢字	b	-	
185	キュウカ	休暇	名詞-普通名詞-一般	漢	1	漢字	b	b	
164	カンテイ	鑑定	名詞-普通名詞-サ変可能	漢	1	漢字	b	c	
262	サトル	悟る	動詞-一般	和	3	漢字	b	c	
730	ユウギ	遊戯	名詞-普通名詞-サ変可能	漢	1	漢字	b	d	
77	エモノ	獲物	名詞-普通名詞-一般	和	4	漢字	b	e	
390	タンネン	丹念	形状詞-一般	漢	1	漢字	b	e	
421	テイスル	呈する	動詞-一般	混	1	漢字	b	e	
727	ユウガ	優雅	名詞-普通名詞-形状詞可能	漢	1	漢字	b	e	芸
338	セン	栓	名詞-普通名詞-一般	漢	1	漢字	c	c	理情
407	ツグナイ	償い	名詞-普通名詞-一般	和	1	漢字	c	d	
525	ネツレツ	熱烈	形状詞-一般	漢	1	漢字	c	e	
509	ナンクセ	難癖	名詞-普通名詞-一般	混	1	漢字	e	e	
444	ドウラン	胴乱	名詞-普通名詞-一般	漢	1	漢字	-	-	国

このような新出漢字の割り振り作業は、教科書教材全体の語彙表を作成した上で、取り上げる必要のある漢字のリストと突き合わせて検討していくことで、より適切な扱いができるようになるであろう。従来は、こうした語彙表がない状態で、教科書編集者の経験と手作業に基づいて行われていたと思われる。コーパスに基づく語彙表類が整備されることで、こうした作業の質と効率を向上させていくことが期待できよう。

5. 類義語の指導 — 「欠陥」と「欠点」—

国語科における語彙指導では、類義語を相互に比較し、意味やニュアンスの違いを考えさせ語感を養っていくことも重視されている。表1に示した「注意する語句」の中にも、そのような観点で選ばれたのではないと思われる語がある。「欠点」と「欠陥」である。この二語は、教材本文では次のように近接した箇所では使われている⁵。

この少年に、コムラサキを見せた。彼は、専門家らしくそれを鑑定し、その珍しいことを認め、二十ペニヒぐらいの現金の値打ちはある、と値踏みした。しかし、それから、彼は難癖をつけ始め、展翅のしかたが悪いとか、右の触角が曲がっているとか、左の触角がのびているとか言い、そのうえ、足が二本欠けているという、もっともな欠陥を発見した。僕は、その欠点をたいしたものとは考えなかったが、こっぴどい批評家のため、自分の獲物に対する喜びはかなり傷つけられた。それで、僕は、二度と

⁵ 注4で取り上げた「難癖」もこの箇所に出現している。

彼に獲物を見せなかった。

自分の収集であるコムラサキを見せた僕に対して、少年がその難点を指摘し、それよって僕が傷つく場面であり、後に起こる重大な事件（僕が少年の収集である蝶を壊してすぐに謝罪しなかったこと）に向かって物語が転換する重要な場面である。僕の収集であるコムラサキの難点を、少年の視点からは「欠陥」、僕の視点からは「欠点」と表現し分けられている。この二語の意味やニュアンスの違いを考えさせることと、物語の読解とを関連づけた指導が行われることが期待される箇所である。

星（2010）は、これを見事に実践した報告であるが、生徒と作業をした結果を図1のような板書にまとめている。これは辞書と内省によって引き出したものであるが、語彙表やコーパスからこのような分析・整理ができると、客観性が増し、信頼性も高まり、応用の範囲が広まるのではないかと思われる。

【欠陥】		【欠点】
大きい	↔	小さい
重大な失敗	↔	軽い失敗
改善できないもの	↔	改善できるもの
物・機械的	↔	生物・人間的
少し冷たい感じ	↔	少し温かい感じ

図1 星（2010）による授業実践

まず、「教科書コーパス語彙表」で、中学校・高校の教科別にこの二語の頻度を調べると表5のようになる。

表5 「教科書コーパス語彙表」における使用度数（中学校・高校）

読み	語彙素	全教科	国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報
ケツカン	欠陥	33	1	0	0	19	0	11	0	1	1
ケツテン	欠点	49	11	1	18	4	1	7	2	0	5

「欠陥」は社会と技術家庭に多く、特徴語を表示した表によれば技術家庭科の特徴語となっている。一方、「欠点」は国語と理科に多い。教科書の用例を見ると、社会では組織や制度について、技術家庭では製品について「欠陥」と言うものが多い。一方、国語では人物について、理科では物質や素材について「欠点」と言うものが多い。このように、教科書の語彙表と教科書のコーパスからだけでも、二語の相違はかなりの程度明らかにできる。

二語の違いをより明確に明らかにするには、BCCWJの豊富な用例を分析するのが望ましい。BCCWJの検索デモサイトで「欠陥」「欠点」を引くと、ともに300件以上の用例が得られる。それらは、次のような意味構成要素に分析できる。

A	ノ(ガ)・ニ	B	トイウ	{欠陥・欠点}	ガ・ヲ	C	スル
[主体]		[内容]				[状態・動作]	

次の例における実線部がA（主体）、点線部がB（内容）、波線部がC（状態・動作）に相当する。A・B・Cに相当する語句を抽出し整理すると、表6のようにまとめられる。

製品に欠陥があってリコールとなった場合（書籍）
鉄はさびやすいという欠点がある（教科書）

表6 「欠陥」と「欠点」の意味構成要素

	A (主体)	B (内容)	C (状態・動作)
欠陥	製品、構造物、制度、・・・ (ニが多い)	(表現されにくい)	ある、ない、持つ、・・・
欠点	人、物、制度、・・・ (ノ・ガが多い)	(表現されやすい)	ある、ない、持つ、指摘する、自覚する、気付く、改善する、直す、・・・

この表から、「欠陥」は、製造物が完全でないところについて言い、そこを直すことはできないような固定的な性質を帯びているのに対して、「欠点」は、人や物の十分でないところについて言い、その内容によっては直すことができるような性質を帯びていると総括することができよう。この総括は、図1に示した星(2010)に合致する部分もあるが、合致しない部分もある。従来は、主に辞書や内省で行われてきた類義語の語彙指導に、コーパスの用例を手段として加えることで、指導現場はより活発なものになるのではないだろうか。

6. おわりに

見てきたように、従来の教材研究の成果の蓄積にコーパスを加えることによって、教材研究はより活発化していくのではないかと考える。多くの教材の語彙指導の教材研究にコーパスや語彙表を導入し試行を重ねていくことが、まずは求められるだろう。

文献

星由紀枝(2010)『『言い換え』と『対比』で語句・語彙を増やす』, 教育科学国語教育, 719, 明治図書, pp.72-75.

別表 少年の日の思い出 語彙表

ID	語彙素読み	語彙素	品詞	語種	度数	語句 指導	レベル _LB	レベル _OC	特徴教科
1	アイダ	間	名詞-普通名詞-副詞可能	和	3		a	a	
2	アイツ	彼奴	代名詞	和	1		a	c	
3	アイニク	生憎	副詞	和	1	注意	b	e	
4	アウ	会う	動詞-一般	和	1		a	a	外
5	アオイ	青い	形容詞-一般	和	2		a	b	
6	アカチャ	赤茶	名詞-普通名詞-一般	混	1		d	e	
7	アガル	上がる	動詞-一般	和	4		a	a	
8	アカルイ	明るい	形容詞-一般	和	2		a	a	理芸保情
9	アクトク	悪徳	名詞-普通名詞-一般	漢	1		c	c	
10	アゲサゲ	上げ下げ	名詞-普通名詞-サ変可能	和	1		d	d	
11	アケル	開ける	動詞-一般	和	1		a	a	
12	アサ	朝	名詞-普通名詞-副詞可能	和	2		a	a	
13	アシ	足	名詞-普通名詞-一般	和	1		a	a	
14	アジワウ	味わう	動詞-一般	和	1		a	c	国芸保
15	アセ	汗	名詞-普通名詞-一般	和	1		a	a	保
16	アツガミ	厚紙	名詞-普通名詞-一般	和	2		d	d	国数理技芸保
17	アツカン	悪漢	名詞-普通名詞-一般	漢	1		e	-	
18	アツメ	集め	名詞-普通名詞-一般	和	2		c	d	
19	アツメル	集める	動詞-一般	和	1		a	b	国理社
20	アト	後	名詞-普通名詞-副詞可能	和	1		a	a	
21	アト	跡	名詞-普通名詞-一般	和	1		a	c	
22	アナドル	侮る	動詞-一般	和	1		c	e	
23	アノ	彼の	連体詞	和	6		a	a	
24	アマリ	余り	名詞-普通名詞-副詞可能	和	1		a	c	数社
25	アミ	網	名詞-普通名詞-一般	和	1		b	c	
26	アラノ	荒野	名詞-普通名詞-一般	和	1		c	-	
27	アラユル	あらゆる	連体詞	和	1		a	c	
28	アル	或る	連体詞	和	1		a	a	数外
29	アル	有る	動詞-非自立可能	和	11		a	a	数
30	アルイハ	或いは	接続詞	和	1		a	a	
31	アルク	歩く	動詞-一般	和	1		a	a	
32	アワス	合わす	動詞-一般	和	1		c	d	
33	イウ	言う	動詞-一般	和	42		a	a	
34	イウ	言う	動詞-一般	和	1		a	a	
35	イエ	家	名詞-普通名詞-一般	和	3		a	a	
36	イキ	息	名詞-普通名詞-一般	和	1		a	b	芸
37	イク	幾	名詞-数詞	和	1		a	a	国
38	イク	行く	動詞-非自立可能	和	7		a	a	
39	イゼン	依然	副詞	漢	1	注意	a	e	
40	インソグ	急ぐ	動詞-一般	和	1		a	b	
41	イタム	傷む	動詞-一般	和	1		a	a	
42	イチ	一	名詞-数詞	漢	4		a	a	国数理社外技芸保情
43	イチメン	一面	名詞-普通名詞-副詞可能	漢	1		a	c	
44	イツサイ	一切	名詞-普通名詞-副詞可能	漢	1		a	a	
45	イマ	今	名詞-普通名詞-副詞可能	和	2		a	a	国
46	イモウト	妹	名詞-普通名詞-一般	和	1		a	a	国
47	イリ	入り	名詞-普通名詞-一般	和	1		a	a	
48	イリグチ	入り口	名詞-普通名詞-一般	和	2		a	b	
49	イリマジル	入り交じる	動詞-一般	和	1		c	e	
50	イル	居る	動詞-非自立可能	和	60		a	a	
51	イレル	入れる	動詞-一般	和	2		a	a	理外技
52	イロ	色	名詞-普通名詞-一般	和	6		a	a	理技芸情
53	イロアセル	色褪せる	動詞-一般	和	1	注意	c	d	
54	イロイロ	色々	形状詞-一般	和	1		a	a	国数理技芸保情
55	イワ	岩	名詞-普通名詞-一般	和	1		a	c	
56	ウエ	上	名詞-普通名詞-副詞可能	和	12		a	a	国数芸
57	ウケツグ	受け継ぐ	動詞-一般	和	1		b	d	理社技芸
58	ウケル	受ける	動詞-一般	和	1		a	a	社保
59	ウシナウ	失う	動詞-一般	和	1		a	b	
60	ウシロ	後ろ	名詞-普通名詞-一般	和	1		a	a	
61	ウスイ	薄い	形容詞-一般	和	1		a	a	理技
62	ウチ	内	名詞-普通名詞-副詞可能	和	6		a	a	外
63	ウチアケル	打ち明ける	動詞-一般	和	1		b	c	
64	ウチガワ	内側	名詞-普通名詞-一般	和	1		a	b	理
65	ウチコム	打ち込む	動詞-一般	和	1		b	c	

第4章 第2節

コーパスを使った語彙単元学習

鈴木 一史

1. はじめに

現在の国語科の学習指導要領は、「A 話すこと・聞くこと」、「B 書くこと」「C 読むこと」の4領域、「言語事項」の1事項に分かれている。言語事項には語句について書かれており、各領域を通して身につけることとされている。そして、教科書もそれに準じて作られ、例えば小説を読むということを通して、そこに出てくる語句の意味や関連した語彙を学習することになる。したがって、語彙そのものに特化した学習活動はなかなか組み立てていないのが実情である。

小学校段階では、語句と漢字は同化しており、漢字の練習として「ドリル学習」は多く行われている。ドリル学習とは、漢字の読み書きに始まる反復学習であり、授業時間すべてを使って行うというよりは、授業の始めの10分や、朝学習の時間などに行うことが多い。つまり、小学校でも語句に関しての特化した授業や単元というのはあまり組み立てていないと言える。

平成24年実施の中学校学習指導要領では、「言語事項」が〔伝統的な言語文化と国語の特質に関する事項〕と変わり、以下のように書かれている。

(1) 「A 話すこと・聞くこと」、「B 書くこと」及び「C 読むこと」の指導を通して、次の事項について指導する。

イ 言葉の特徴やきまりに関する事項

(エ) 単語の類別について理解し、指示語や接続詞及びこれらと同じような働きをもつ語句などに注意すること。

(オ) 比喩(ゆ)や反復などの表現の技法について理解すること。

このように学習すべきことははっきりしているが、その内容や方法については、授業者に任されている。もしくは、教科書教材の中で適宜行うことになる。

それでは、今まで語彙単元として特化した授業はなかったか。管見によれば、いくつかのすぐれた実践が行われており、それは国語科本来の姿である「ことば」について真剣に取り組む姿である。次章で先人の三つの実践の紹介と、自分の実践記録を述べる。しかしそれはまだコーパスがなかったころの実践である。コーパスがなくても、優れた実践は可能である。しかし、コーパスがあることによって、さらに多くのことばに関しての単元が実践することができるかもしれない。

そこで、コーパスを用いた試みの実践を提案する。もちろん学習指導要領にあるように、「敬語」「類語」「接続語」などの実践である。しかし、同じ実践でもコーパスがあることによってどのように変わるのか。そのことを実際の授業プログラムを作りながら検討していきたい。

2. 今までの語彙単元学習

2. 1 今までの語彙単元

今までの実践に学ぶことは多く、言語を中心に据えた活動として学習者の言語感覚や語彙力・興味関心など総合的言語力を育成するのに役立っていることは間違いない。そこで、本実践もこれらの実践に習うのであるが、言語データベースである「コーパス」の充実により、言語研究のみならず、言語教育・国語科教育に十分資する状態になりつつある。よって、これらの先人の単元学習をコーパスを使うことでどのような可能性が見えてくるのかを、本単元学習でのいくつかの実践をもとに考えていきたい。

2. 2 教科書分析の実践

一つめは、大村（1983）である。この実践は、教科書を一冊の本として読むことで、教科書から「ことば」ということばを抜き出して分類し、まとめ、発表する授業である。その中で、大村氏は「この学習は、分類の出来上がり、結果に重点を置かず、一連の作業をする、そのことを大切にしたい。」と述べている。これこそまさに新学習指導要領の重点項目である「言語活動」そのものである。そして、その活動の中心には、言語そのものが置かれている。

2. 3 ことばの収集の実践

二つめは町田（1992）である。この実践は、身近な場所にあることばとして、教科書のみならず、新聞・雑誌・広告・読んでいる本・テレビやラジオなどからことばを収集して分類し発表している。発表の記録からは、分類方法として「語種」や「品詞」など様々な観点があげられている。

2. 4 類語レトリックの実践

三つめは金子（1999）の広告のことばの実践である。この実践では、広告のことばを広く集めることで、表現のレトリックの分析という「読むこと」の指導から、書く立場に立って様々にことばを操って使ってみる「書くこと」の指導まで行われている。言語活動の単元として、言語を集め分類・整理し、表現していく活動が織り込まれている。

2. 5 類語学習の実践

今までの語彙単元を踏まえつつ、新しいシソーラスを使った授業を以前実践した（鈴木、2005）。その時の指導案が以下の表である。

表1 シソーラス使用による語彙単元学習指導案

	指導事項	学習活動	留意事項	時間
導入	・ ことばをいろいろと変えても同じような意味になることを知らせる。	・ 本字の学習として、ことばをさまざまに変換させることを知る。	・ 年間計画の中で、どの部分をやっているのか、何をこれから行うのかについて確認させる。	5

展 開	<ul style="list-style-type: none"> ・プリント配付 ・個人でプリントに書かれている文章の一部を文章の大意を変えずに言いかえるように指示。 ・同じ文を選んだグループごとに集まり、文意が変わっていないか確認させる。 ・グループ内でことばのニュアンスの違いについて書き出せる。 	<ul style="list-style-type: none"> ・個人個人で配られたプリントに書かれている文を読み、どの文の言いかえが可能か考える。 ・自分なりに言いかえのたくさんできそうな文を選び、かえてみる。 ・選んだ文章ごとにグループを作り、どのようなことばにかえられたかお互いに見合う。 ・グループごとに六つ以上のことばに言いかえて、文章を作り直す。 ・他にもないかどうかを辞書類を使って確かめる。 ・それぞれのことばについて、どのようにニュアンスが違うかを説明する。 ・それぞれのことばについてどのようなシチュエーションで使うかについて考え、文脈を考えさせる。 	<ul style="list-style-type: none"> ・大意を変えずに言いかえるとはどういうことかについて、例を挙げながら説明する。 ・ことばの言いかえには、品詞を変える、和語と漢語の変換、カタカナ語に直す、擬音を使うなどがあることを指示。 ・品詞分類や変換がうまくいかない生徒に対して、期間巡視をしつつ、変換の種類について、個別指導をする。 ・例文は十以上を用意するも、一つの文を三人以上は選択するように調整。 ・グループにそれぞれの記事を持ち寄った後、文意についてチェックする。 ・ニュアンスの違いの説明について、文章として説明したり、長い文章を作って使い方を説明したりすることを説明。 ・グループごとに回り、書き方のわからないところに具体的な指示をする。 	30
ま と め	<ul style="list-style-type: none"> ・グループごとにことばの言いかえとニュアンスの違いを確認させる。 ・ニュアンスの違いが正当かどうか聞き取る。 	<ul style="list-style-type: none"> ・班ごとに元の文章と言いかえた文章と使い方の違いについて発表する。 ・それぞれのニュアンスの違いを聞き、その違いは正しいかどうかについて意見交換する。 ・次回は実際の文章でどのように使われているかを確かめることを確認する。 	<ul style="list-style-type: none"> ・発表の仕方を指示。元の文章を言い、変えた文章を書き、それぞれがどのように違いがあるのかを説明させる。 ・発表された内容について、自分なりに検討し、自分の使い方とあっているか考えるようにさせる。 ・ノートにメモをさせることで、他者が考えたことばの種類や広がりをもとにさせる。 	15

2. 6 語彙単元の問題点

これら四つの実践は、それぞれが教育的効果という点では優れた実践であり、新しい試みの実践である。また、学習指導要領に沿って学習内容も精選され、学習も進んでいる様子もわかる。問題は、教材開発をする際に、授業者の経験だけに頼っていることである。当然経験豊かな授業者は、学習者の実態に即して有効な教材化をなす。しかし、それでは敷衍して多くの学習者に使うことができない。また、それが他の語彙である場合に有効かどうかが分かりにくい。

たとえば3.2の教科書分析の例では、授業者がきちんと教科書からすべての学習語彙を事前に抜き出す必要があり、その分析まで行う必要がある。これは「教科書コーパス」(本報告書第1章第1節参照)を用いることによって、全学年や全教科まで視野を広げて検討す

ることができ、どのようなことばが学習者にとって必要か、どのようなことばが学校の学習にとって必要か、といったことを考慮して語彙の選択をすることができる。

3.3 や 3.4 のことばの収集の学習では、まさにコーパスが有効であろう。今まで教師は教材化するために、道を歩いていてもことばや看板や広告などに注意して、少しずつ言葉を集めていた。また、新聞などのスクラップブックを作ったり、同じテーマの広告の切り抜きを集めたりしてきた。それは教師にとって楽しみでもあり、教師自身の言語観を養うことにもつながっていたかもしれない。しかし、コーパスがあることによって、さまざまなことばを集めるときに、授業者は事前に集める語彙や範囲をコーパスによって狭め、学習に適した範囲を示すことができる。このことは教科書との関連など学習の効率化につながる大きな方法論の変化である。

以下の章では、コーパスを使った学習についての実践授業プログラムを提示し、その可能性と課題に迫る。

3. コーパスを使った単元学習

3. 1 語彙教育へのコーパスの方向性

語彙教育にコーパスを取り入れる際には、二つのアプローチがある。以下の図はその二つを図示したものである。

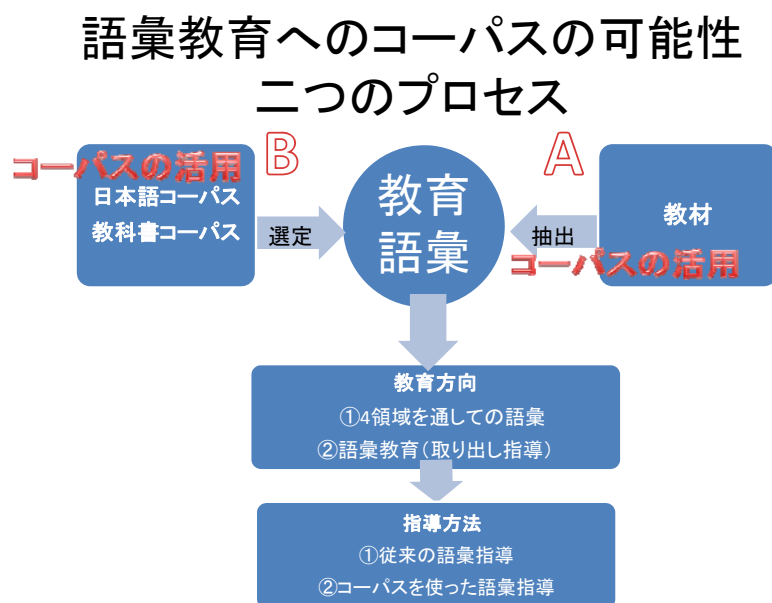


図1 コーパス活用のアプローチ

一つは、教科書教材を中心とする既存の教材を学習する際に、それらの教材にふくまれる語彙がどのようなものかを分析する方向。例えば説明文教材などで、学習者が理解しにくい語彙は何か、小説教材などで類語として知っていなければならない語は何かなどを分析し教材化するために使用する方向である。

もう一つは、コーパスそのものの分析によって、学ぶべき言葉を抽出し、語彙単元を作る方法である。この章では、教科書教材の分析ではなく、語彙単元の作成をし、そこにコ

一パスがどのようにかかわってくるのかについて考察する。

3. 2 単元設定の理由

文章表現に際して、語彙の選択の幅を広げ、語感をさらに磨くために、この単元を設定する。学習者の身の回りにある様々なところからことばを集め、自分を取り巻く言語環境を見直すことで、自分の言語表現に役立てつつ、語彙や言語感覚を豊かにすることを目標にしている。このような単元学習は、先に述べた優れた授業実践がある。この授業実践の延長線にあるものとして本実践を提案する。

中学段階で文章表現を習っている学習者は、課題作文のような課題把握を主眼とした作文や、意見文のように意見や主張を、根拠をあげて論じる作文を書く練習をしてきている。また、多くの説明文や文学的文章に触れることで、語彙や語感を磨きつつある。しかし、表現語彙の幅や語のニュアンスについてはまだ曖昧な部分を残している。自分で一度書きあげた文章をさらに良くしようとする意欲や努力、またどのようにすればそれが可能になるかなどの方法は未熟であるといえる。そのような学習者に対して、自分が使う語彙に対しての意識を高め、最終的には語のニュアンスをつかんで巧みに使えるように指導する必要がある。そのための方法として、語彙を中心とした単元を組む。

以下、三つの単元活動を述べる。一つは青空文庫の検索、一つは教科書の検索、もう一つは言語コーパスの検索。観点や目的に従って、それぞれの言語データベースをどのように活用するかは異なる。また本単元学習以外にも、多様な活用が可能である。ここでは、以上三つの言語データベースを検索することを中心とした授業実践を、学習指導の実際として紹介する。

3. 3 比喩表現の学習

3. 3. 1 教科書に出てくる比喩表現

教科書には比喩表現が多く使われ、学習目標にもなっている。国語科の教科書では「…のような…」が多いが、国語以外では、「このような…」、「次のような…」と出てくる。具体的な例として、以下のような文がある。

- ・ 自意識に目覚めていった少女時代。地獄のような引き揚げを経験した娘時代。そのどれもが、かけがえのない日々
- ・ 日常が遠く感じられ、まるで仙人になったような気がする。
- ・ 胸が締めつけられるような思いが込み上げてくるのである。
- ・ 母は、竹を割ったような性格だ。
- ・ 息をのむような驚きをそのままに伝えるフォト・エッセイ集。

これらの言葉について、辞書的な意味ではなく、文脈から考えさせたり捉えさせて語彙を広げていきたい。そのために「青空文庫」での比喩を教材として使用した。

ネット上の文学作品データベースに「青空文庫」がある。その全作品をデータベース化し、言語データベース全文検索ソフト「ひまわり」で検索する。2009年で1万作品近く挙げられている。ただし、著作権の消滅した作品のみであるため、ことばの使用例としては古い。特に現代の学習者にとっては使わないことばが多く見つかり、また普段使っていることばがあまり見つからないという現象が起こる。それでも、作家が使っていることばであるという点で、十分ことばの学習には役立つものである。

3. 3. 2 生徒研究テーマ

授業では、生徒一人ひとりが次のようなテーマを持って調べた。

- ・「優しさ」という言葉を調べて
- ・「くぐる」による変化のきっかけ
- ・「持っている」の前につく言葉
- ・「全然」の時代変化
- ・「暑さ」「寒さ」についての変化
- ・「失敬」という言葉
- ・「擬人」「比喻」について
- ・真珠の象徴するもの
- ・効果音について
- ・「さびしい」について
- ・時間を表す言葉
- ・太宰治の作品と時代背景、心理状況

表記の違いや比喻の使用例、作家による差異と様々な観点でことばを見つけようとする姿勢がうかがえるとともに、それらの違いを学習することが可能な教材である。一例として、「比喻」の学習方法と成果について紹介する。

「ひまわり」でキーワードに「真珠」を入れる。そして、「後文脈」に「よう（ようだ・ような・ように 等）」を入れて検索する。前後 15 文字で検索をすると、図のような画面で、検索結果が表示される（図 2）。

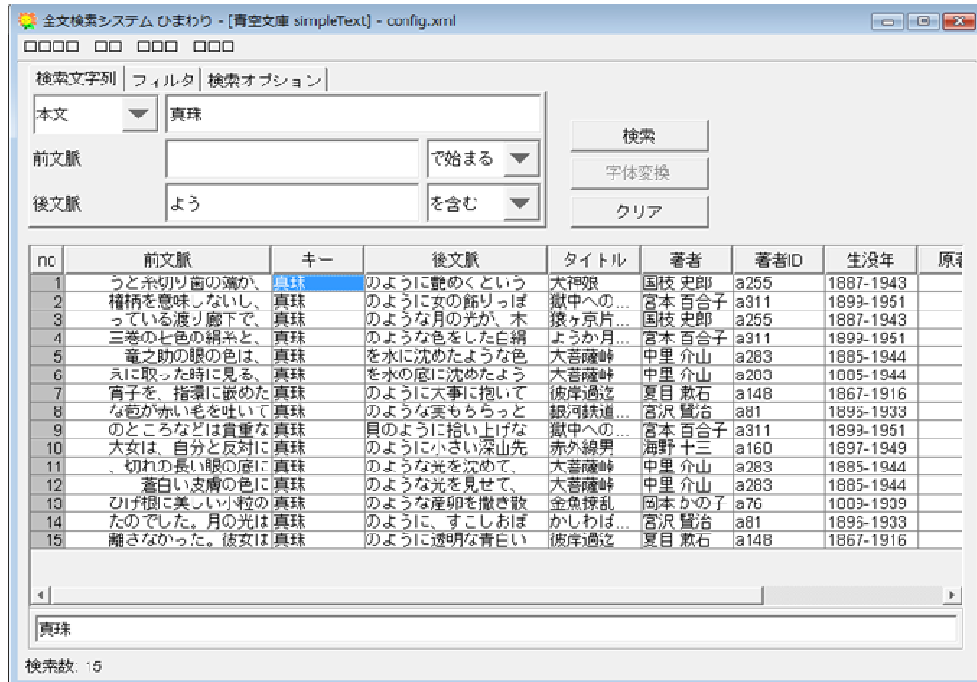


図 2 「ひまわり」検索例：「真珠」

ここから、学習者は、真珠が「大切なもの」「高価なもの」「光」などの比喻として使われていることを読み取る。また、「声」について、どのような例え方をされているかを調べた学習者もいる。以下に数例あげると、「胡弓を弾くような響のある淋しい声」「金の鈴を

振るような制多伽童子の音声」「くすぐられるような雪ちゃんの笑い声」「やはり囁の呟くような、ぶつぶつ云う声」「おかあさんと同じような、やさしいいい声」「浪花節語りのような妙に張りのある声」「老婆はつぶやくような、うめくような声で」、などが検索の結果上がってくる。また、「すごく」などは、現在では、程度のはなはだしい様子一般に用いられているが、青空文庫で検索すると、「マイナス面での用法が多い」ということも発見されている。このように集めて分類した結果を、表などにまとめて発表して、自分たちの感覚との共通点や差異などを話し合う。

比喩表現は、たくさんの用例を集めることで、そのことばがたとえているものや状況について、感覚的に学ぶことができる。学習材として、授業者が検索して比喩一覧にして提示することもできる。

3. 4 教科書コーパスを使った教材開発

教科書に出てくることばについて、大村はま実践では国語の教科書から「ことば」を抽出したが、ここではコーパスの特性を用いて、他教科での使い方を調べる。そうすることで、他教科とのつながりやことばの使い方の差異を明確にすることができる。国語の教科書に出てくる、中心的な概念を表す名詞や付帯的状态を表す副詞などが、他教科でどのように出てくるかについては、学習者は自覚的ではない。しかし、一人の学習者の中では、必ず統合されるべき語彙であるために、これらを調べて整理しまとめることで、語彙の面から他教科の学習に役立てることができる。学習材の一つとして「接続詞」について教科書の事例を検討する。

接続詞を適切に使えるようにするために、接続詞の用法について学習する。学習者は高校段階では、かなり幅広く接続詞を用いて文章を書く。教科書にも多くの接続詞が出現する。「或いは」「一方」「及び」「が」「かつ」「から」「けれど」「さて」「さらに」「さりとて」「されど」「されば」「しかし」「しかも」「即ち」「そして」「ただ」「ただし」「で」「では」「と」「ないし」「なお」「また」「もしくは」「もともと」などである。

「および」に限って調べると、国語では「接続詞」として学習し、文と文をつなぐことばとして学習する。しかし、その使用方法は練習するであろうか。その使用例はどうであろうか。

社会科や数学や理科の教科書にも「および」が出てきており、それは文章理解に重要な役割を担っている場合もある。

社会

- ・思想および良心の自由...第19条
 - ・社会福祉，社会保障及および公衆衛生の向上及び増進に努めなければならない
- など、条文を中心にして「および」が使われる。

理科

- ・マグネシウムが酸化されたと同時に，酸素および塩素がそれぞれ還元されたことになる
- ・互いに反応する気体および生成する気体の体積比

数学

- ・閉区間で連続な関数は，その閉区間で最大値および最小値をもつ。
 - ・半円 $y = b + \sqrt{r^2 - x^2}$ と x 軸および2直線 $x = -r$ ， $x = r$ で囲まれた部分
- これらの教科に出てくる「および」はまるで違った様相を呈していることがうかがえる。

しかし、学習者は一人であり、「および」も一つである。そうであるならば、国語科として他教科に目配りした言語学習の授業を設定する必要がある。

国語の教科書に出てくることばについては、接続詞に限らず他教科との関連を調べ、文例を調べる必要がある。これが、コーパスによって可能となる。もしくは、「接続詞語彙表」が文例とともに挙げられていればそれでも良い。

国語の教科書に出てくる様々な名詞については、更に多岐にわたった使い方がなされる。これらのことばを国語科の授業として行う際に、他教科ではどうなっているのかという視点を持つことで、生徒の中の語彙習得の一端を知ることになる。一人の生徒がそのことばにどのように出会っていくのか、それは国語科だけの問題ではない。と同時に、国語科の授業を中心に考えなければいけない問題である。

3. 5 言語コーパスを使った類語指導の実際

類義語の学習は、今までは教科書に出てきた言葉について、「他にどんな言葉があるか考えよう（調べよう）」という学習であった。類語の差異についての学習はあまり行われていない。それは、内省によるものが多く、多くの文脈から抽象化するということがなかなかできなかったからである。もちろん言語の内省は母語話者である以上大切な方法である。しかし言語についての学習途上者は他の多くの文章に触れることで、言語感覚を豊かにしていくと考える。

「ことばの意味」（平凡社ライブラリー）には、類義語について多く記載されている。そこでそのうちのいくつかを選んで、使用方法の差を考えさせた。はじめにクラス全体で「サワル」と「フレル」について考え、差異の観点を出し合った。たとえば、「強弱」や「対象」

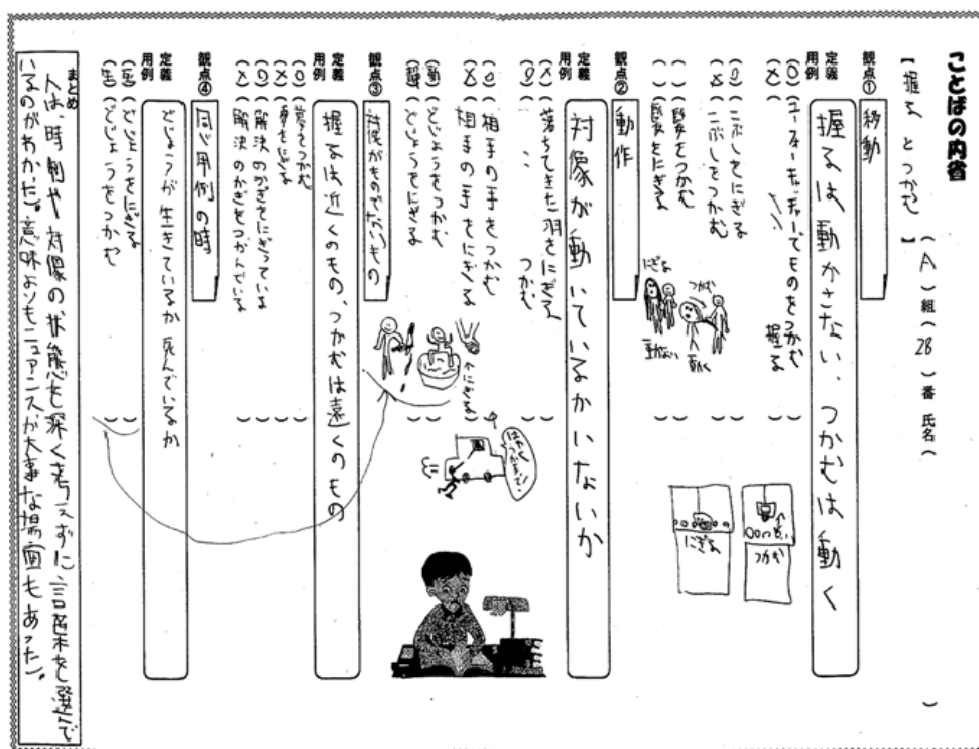


図3 「握るとつかむ」の生徒作品

などいくつかの観点があげられる。次に、「オモウ」「カンガエル」、「サケル」「ヨケル」、「ツカム」「ニギル」について、差異を考えさせた。その時の生徒のワークシートの一例である(図3)。

ことばの差異を考えるときに、まずは「内省」をする。「使う」「使える」「使えない」「あまり使わない」等で分類する。しかし、その感覚が本当に良いかどうかを調べなければならない。調べる方法は二つ。巧みな言語使用者に聞くか、実際に使われているかどうか、である。言語使用者に聞くことはインタビューなどを使うことがある。もう一つの方法は、言語コーパスを検索する作業を入れることで、差異を考えさせることである。つまり、「ツカム」と「ニギル」について、対象の差を考えたい場合、全文検索ソフト「ひまわり」を使って、キーワードに「ツカム」「ニギル」を入れる。(この時は、漢字やひらがなで入力する)そして、前文脈に「を」を入れることで、「・・・をにぎる(つかむ)」という文章が抽出される。それによって、違いを考えていくことができるのである。「を」入れた場合は、キーワードの「対象」の差異を導き出すことができる。

もう一つの例は、「考える」と「思う」の差異である。実際の学習者のワークシートが以下のものである。

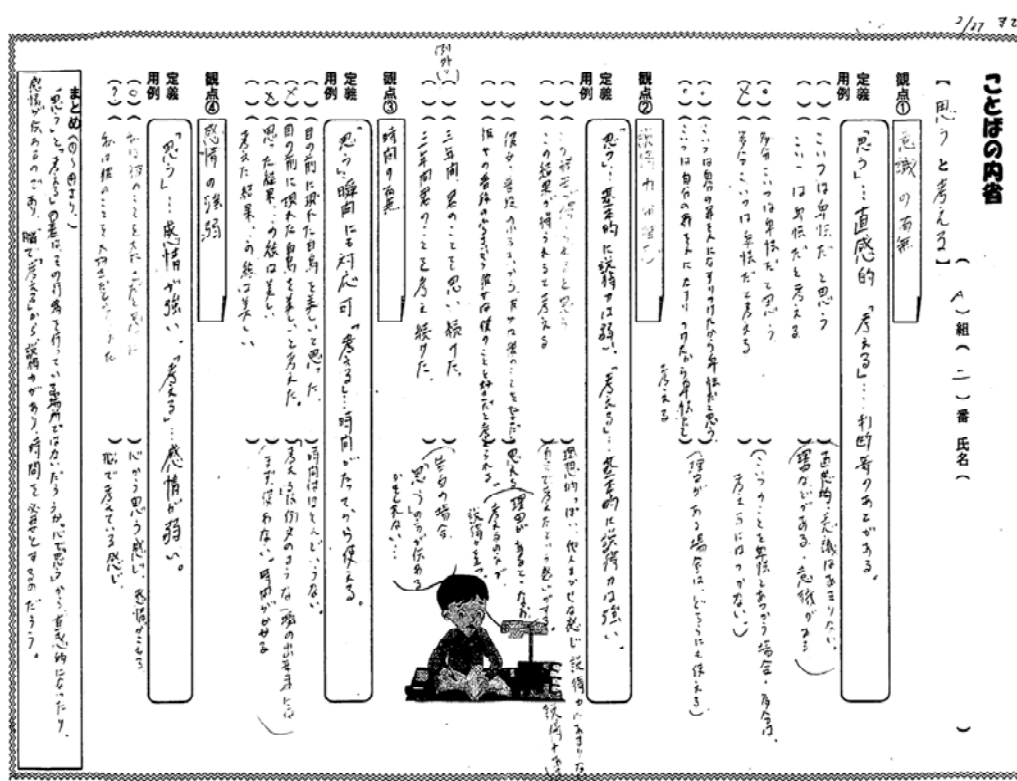


図4 「思うと考える」の生徒作品

これを考える際に、授業者が前もって、良い文例をコーパスで探しておくことによって、学習者はスムーズな学習に取り組むことができる。今まではこの文例を教師が「創作」していたが、それを実際の文章からたくさん引用することができるようになった。

問題は、生徒自身が使えないことである。コーパスの多くの文例の中には、規範的でないものも多く含まれているので、教師が事前に抽出しておくことが大切である。

3. 6 評価について

評価の方法は、集めたことばや分類・整理の方法などをワークシートに記入することで、評価することができる。しかし、評価の観点、具体的な言葉一つ一つについて、考えていかなければならない。たとえば、「比喩」については、「真珠の象徴性」というテーマは生徒が決めたものである。したがって授業者は生徒とともに、それらの言葉を収集しつつ、評価しなければならない。これは一人ひとりのテーマが違う場合、授業者の作業量も大きくなる。しかし、テーマを決める際には、ある程度の方向性や系統性があるために、テーマ決定時に授業者が助言することで、学習材たりうる言葉を提示することができる。それによって、集められたことば、分類整理の仕方を評価することが可能となる。しかし、時には授業者の想定を超えた言葉を選び成果を上げる学習者もいる。そこがこの言語データベースを使う際の教材化の難点でもあり、面白味でもある。

- ① ことばを集めて整理し、自分なりのまとまりでとらえられているか。(ワークシート・発表)
- ② 自分の言葉を見直し、表現に生かそうとしているか。(推敲)

4. おわりに

学習者を取り巻く言語環境をどのようにしたら、学習材として成立させられるか。この大きな問題に様々な授業実践がなされてきた。まとめて言えば、ことばを集め、整理し、分類し、分析し、表現するという一連の作業になる。学ぶということは学習者自身がその研究のプロセスを辿るということに他ならない。そのような実践を行いたいと願って、先人の実践を踏まえつつ、新たな言語データベース・コーパスを使った実践を試みている。

学習者は言語環境とともに情報機器の使用にも長けてきており、今後このような機器を使った学習が進むのではないかと考える。しかし、課題もある。コーパス自体の設計や構築の問題とそれをどう教育に生かしていくかという問題である。本実践では、コーパス使用の一端であるが、筆者の力不足によりコーパスの持つ可能性をあまり引き出せなかった。今後の課題として、教育にこれら膨大なデータをどのように生かし、どのように学習者に提示することで学習材となしうるのか。コーパスの設計とともに、教育的使用に耐えうるコーパスを構築し、豊かな授業プログラムを構築していく必要がある。

文献

- 大村はま (1983) 「ことば—こんな意味が、こんな意味も」大村はま国語教室 9, 筑摩書房, pp.101-111.
- 金子守 (1999) 『総合的学習に生きる広告の読み方・生かし方』, 東洋館出版社.
- 鈴木一史 (2005) 「語彙拡充におけるシソーラスの役割」東京大学教育学部附属論集, 48, pp.149-156.
- 町田守弘 (1992) 「身近なことばを集める—ワードハンティングの実践」単元学習の新展開 VI, 東洋館出版社, pp.232-245.

第4章 第3節

「テーマ型作文」の教材開発

—「テーマ語彙集」の活用による—

河内 昭浩

1. はじめに

本稿は、作文・小論文のための「テーマ語彙集」を基礎資料とした、「テーマ型作文」の教材開発の方法について述べたものである。「テーマ語彙集」とは、筆者が、作文・小論文のテーマとなる語彙を収集し、その語彙の、「教科書コーパス」（本報告書第1章第1節参照）及び『現代日本語書き言葉均衡コーパス』¹における出現数を示したものである。本稿に別表として掲載している。

筆者はこれまで高等学校において、コーパスを活用した作文指導の授業実践を重ねてきた²。今回作成した「テーマ語彙集」を、今後の、コーパスを活用した、体系的な作文指導と教材開発のための基礎データにしたいと考えている。

ここではまず「テーマ語彙集」の設計と特徴について述べ、その上でこれまでに行ってきた教材化の方法を述べ、加えて今後の展望について述べていきたい。

2. 「テーマ語彙」について

高校生が書く作文・小論文には大きく二つの種類がある。それは、あるテーマを与えられ、そのテーマについての考えを述べる「テーマ型作文」と、課題文や図表の内容を読解し、その内容についての考えを述べる「課題文型作文」の二つである。本稿で提示する語彙は、主に前者の「テーマ型作文」のテーマとなる語彙（以下、「テーマ語彙」と呼ぶ）である。

また「テーマ語彙」についても大きく二つの種類があると言える。それは、そのテーマに対する専門的知識が求められるもの（例えば「インフォームド・コンセント」について）と、幅広い解釈と論点が可能なもの（例えば「望ましい医療」について）の二つである。前者は個別の時事用語であることが多く、後者は前者のカテゴリーに相当する語であることが多い。これまでの実践では主に後者の語彙を取り上げて、コーパス内のデータを作文指導に活用してきた。しかし前者についても、特定の分野の語彙が、どの教科の教科書に、またどの程度一般社会で使用されているかということも、今後の授業や教材開発のための有効なデータになると考え、「テーマ語彙集」に加えている。

「テーマ型作文」においては、「テーマ語彙」に対する発想力がまず必要になる。「発想」することが個々の知識や経験に委ねられるがために、従来この「発想」という作文の初期動作においては、生徒に「空白」が与えられるのみであることが多かった（例えば、「思い

¹ 学習活動での利用を考慮して、本稿では『現代日本語書き言葉均衡コーパス』モニター公開データ（2009年度版）での検索数を示した。

² 河内（2010）など。

ついたことを書き出してみよう」など)。しかし自由に発想せよと指示するだけで、過去の知識や経験が必ずしも開かれるものではない。知識や経験の多くは、それを獲得した個々の場面に閉ざされている場合が多く、作文を書くという場面で生徒の脳裏にいつでも登場するものではない。

教科というジャンル、そして教科書という書物が、「発想」の扉を開く。生徒は日々、様々な教科書で学習し、多くの知識を得ている。これまで個々の教科内の暗記事項でしかなかった知識が、作文を書く上で有益であると知ることによって、生徒は負担感を感じずに作文を書くことができる。閉ざされた知識と経験が、「教科書コーパス」を活用した作文活動を通じて生徒内で結ばれていく。これはこれまでの実践で得た確かな実感である。

また『現代日本語書き言葉均衡コーパス』内の文例は、生徒が過去の知識と経験を超え、視野を広げるために有効に機能する。ただし膨大なデータの中から生徒に有益なデータを取り出す方法が確立できていないため、今回は、検索ソフト「ひまわり」による検索数を挙げるのみとした。

本稿では「テーマ語彙集」を基礎資料とした、コーパスを活用した作文指導の教材化の試案を示している。「テーマ語彙」の前後の文脈に共起する語を抽出し、発想の補助資料として活用している。また、発想を文にしていくうえでの手本として、文例を示している。今後、様々な「テーマ語彙」を対象に教材を開発していきたいと考えている。

なお用語としての「作文」と「小論文」については、高等学校では、「作文」は主観的に気持ちを述べたもの、「小論文」は他者に分かるように論理的に述べたものと区別している。本語彙集はそのどちらにも資するものであると考え、「作文・小論文」と併記している。

3. 「テーマ語彙集」の作成

「テーマ語彙」を以下のものから抽出した。

- (1) 『全国大学小論文入試全出題 2002～2006』（旺文社）
- (2) 『小論文キーワードファイル』（第一学習社）
- (3) 『大学入試現代文キーワード』（桐原書店）

(1) には、2002年度から2006年度までの、計5年間の大学入試の小論文の問題が収められている。本書は出題年度を更新して毎年出版されているが、コーパスに収められている文献が2005年までのものであることを考慮して、この年度のものを採用した。今回は国立大学の入試問題全2084題の内、「テーマ型作文」（本書では「課題論述型」と呼ぶ）に区分されている471題から、テーマの重なりなどを除き、315の「テーマ語彙」を選定した。

「テーマ語彙」を次のように選定した。

[問題文] 「遊び」について、あなたの考えたことを述べよ。

→ 「テーマ語彙」： 遊び

ただしこのように、問題文から明瞭に「テーマ語彙」を抜き出せる問題ばかりではない。そこで下記の基準を設定して、語彙を抽出することにした。

- a) 発想の主体となる部分のみを「テーマ語彙」として抽出する。
 [問題文] 日本在住の外国人に対する日本語教育において留意すべきことを述べなさい。
 →「テーマ語彙」 日本語教育
- b) テーマとなる語彙が併記されている場合、いずれも「テーマ語彙」として抽出する。
 [問題文] 人間の感覚と芸術について自由に論じなさい。
 →「テーマ語彙」 感覚 芸術
- c) 語によっては、コーパスから関連語彙や文例を導きやすい同概念の語に改める。
 [問題文] 「はたらくこと」というテーマで自由に論じなさい。
 →「テーマ語彙」 労働

(2) には、「生活・社会」、「環境」などの10のカテゴリーのもとに150の語彙がキーワードとして掲載されている。ここではそのカテゴリー及びキーワードを示す語を「テーマ語彙」として選定した。またその際には前述の基準に沿って語彙表に加えた。本書のキーワードには、専門分野固有の用語が多い。従来、作文・小論文において必要とされる語彙はこれである。

(3) には、現代文の読解に必要とされる500の語彙が、「対義語」「必修語」などのカテゴリーのもとに収められている。ここでは、「主題語」としてくくられた52の語彙を「テーマ語彙」に加えた。本書で「主題語」とされる語（「文化」、「国際化」など）は、専門分野を問わず様々な観点から論じることができるものが多い。これまで筆者が授業実践で扱った語彙はこれに当たる。

以上(1)～(3)の文献から抽出した計473の「テーマ語彙」の、「教科書コーパス」及び『現代日本語書き言葉均衡コーパス』モニター公開データでの出現数を示したのが、別表の「テーマ語彙集」である。尚、表中の(1),(2),(3)は上記引用文献を指し、記号「*」はその文献での「テーマ語彙」の出現を示している。

4. 特徴のある「テーマ語彙」と教材化の指針

本項では、「テーマ語彙集」の語彙の特徴について述べる。以下、教科別の度数をもとに、特徴のある「テーマ語彙」を列挙する。また、その語彙を用いた「テーマ型作文」の教材開発の指針を示す。

4. 1 多教科型語彙

様々な教科で見られる「テーマ型語彙」を「多教科型語彙」と呼ぶこととする。主なものは以下の通りである。

10教科（全教科）で見られる「テーマ語彙」： 学校

9教科（生活以外の教科）で見られる「テーマ語彙」：

知 心 日本 食 死 子ども 時間 声 病 人間 地域 情報 生活 女性
 大学 地方 船 基本 生産 青 都市 薬 自然 自由 消費 波 文化 家族
 科学 歴史 身体 人口 健康 歯 運動 音楽 表現 効率 発見 地震 現代
 手紙 感覚 スポーツ 空間 ネットワーク 授業 学習 創造 インターネット

教科 宇宙 階段 エネルギー 未来 観察 コンピュータ 便利 地球 学び
角度

8教科で見られる「テーマ語彙」:

アルファベット 椅子 現代社会 コマーシャル 地球温暖化 グリーン ラジオ
科学技術 温暖化 瞬間 若者 あいさつ コミュニケーション 電子メール 環境
問題 おおきくなる 生きる 差別 基礎 言語 文明 障害 精神 文学 日本語
メディア 日本人 自己 宗教 英語 伝統 教育 事件 保存 個人 近代 資源
農業 民族 企業 食品 労働 戦争 政府 文字 生物 実験 環境

これらの「多教科型語彙」には、多義性を有するものや、概念を示す語が多い。様々な視点・論点が可能であり、作文へのステップとしての話し合い活動にも適した「テーマ語彙」である。学習者の視野を拡充させるために、また論点を整理させるために、その「テーマ語彙」の前後の共起語が有効に活用できると考えられる。また教科書の文例が、作文の書き出しや結文の良き手本となる。教材化の具体例については次項で述べる。

4. 2 単教科型語彙

「多教科型語彙」とは反対に、特定の教科書にのみ見られる「テーマ語彙」を「単教科型語彙」と呼ぶこととする。単一の教科にのみ見られる「単教科型語彙」を以下に挙げる。

国語: 推薦入試 メディアリテラシー 少年犯罪 医学生 オタク ねつ造 学級崩壊
人工知能 キレる 総合的な学習の時間 アンドロイド 良医

数学: 引きこもり

理科: 重力加速度 地学 石鹼

社会: 地域紛争 外国人労働者 住民投票 憲法改正 デモクラシー 行政改革
代替エネルギー 財政危機 国際貢献 教育基本法 京都議定書 公共交通
イラク戦争 高速交通 環境アセスメント デジタル・デバイド 環境税
尊厳死 ホームレス 助産婦 自然観察 バイオエシックス 裁判員制度
市町村合併 社会基盤 優生学 先端医療 説明責任 犯罪被害者 地産地消
塩害 予防医学 定住外国人 企業倫理 一寸法師

外国語: ナノテクノロジー 養護教諭

技術家庭: 児童虐待 孤食 エコクッキング 在宅福祉 ギャンブル 情報開示

芸術: 五線譜

保健体育: 熱中症 有酸素運動 心肺蘇生法 PTSD 一気飲み

セカンド・オピニオン 院内感染 カルテ開示

情報: 知的財産権 電子政府 ヒューマノイド コラボレーション

一見して、個別の事案を指す語彙が多いことが分かる。これらの「テーマ語彙」の教材化においては、教科書の記述が知識の補充に役立つことが分かるようにしていくことが望ましい。教科書の文例が、語義や具体的事例を知るうえで役に立つであろう。また、例えば社会科でのみ見られる「地域紛争」などについては、教科書の具体的事例を整理して示

すことで、学習者にとって有益な教材とすることができると考えられる。

4. 3 教科書特徴語彙

『現代日本語書き言葉均衡コーパス』モニター公開データと比較して、「教科書コーパス」の出現数に特徴のある「テーマ語彙」を「教科書特徴語彙」と呼ぶこととする。『現代日本語書き言葉均衡コーパス』モニター公開データより、「教科書コーパス」の全教科総計の度数が高い「テーマ語彙」は以下の通りである。

二酸化炭素 衣服 天体 20世紀 高分子化合物 万有引力 重力加速度
ソフトウェア 化石燃料 五線譜 エコクッキング 熱中症 スペースシャトル
メディアリテラシー 電子辞書 コンピュータ・ウイルス ナノテクノロジー

これらの語彙は、理科や技術家庭、社会、情報の教科書に特徴的に多く見られる。国語以外の各教科でそれらのテーマについて正しく知識を身につけ、かつそのテーマについて思考を深めることができれば、それがそのまま作文・小論文へとつながる。教材開発においては、各教科での「教科特徴語彙」の指導・取扱いへの提言も意識したい。

また反対に、『現代日本語書き言葉均衡コーパス』モニター公開データには見られるが、「教科書コーパス」には見られない「テーマ語彙」もある。以下に挙げるのは、「教科書コーパス」には一つも見られない「テーマ語彙」である。

ブログ サプリメント 守秘義務 ニート 島嶼 学力低下 大学入試
トップ・ダウン 医療過誤 消臭 中高一貫 ゆとり教育 海洋開発 機能材料
モデル生物 大学入学資格 レジ袋 フリースクール 拡大生産者責任
知識ベースシステム 非正規雇用 タテ社会 エコビジネス 医師不足 へき地医療
PISA 三歳児神話 文化相対主義 メディテーション AO入試 キャリア教育
大学入試センター試験 短縮語

これらの語彙については、『現代日本語書き言葉均衡コーパス』を活用して教材化を図る必要がある。また「テーマ語彙」はもともと大学入試等で高校生に思考を求める語彙である。それらが教科書には見られないということも、教科書の検証の材料にすることも可能であろう。

5. 「多教科型語彙」を用いた「テーマ型作文」の教材開発

筆者はこれまで、「多教科型語彙」をテーマとして、コーパスを活用した作文指導の授業実践を行ってきた。本項ではこれまでの実践をふまえ、「多教科語彙」におけるコーパスを活用した授業の流れや、ワークシート作成の過程を、具体例とともに示していく。

5. 1 問題文・テーマ語彙の設定

具体例として、問題文と「テーマ語彙」を次のように設定する。

問題文：「スポーツ」について考えたことを述べなさい。

テーマ語彙：スポーツ

「教科書コーパス」中の「スポーツ」の出現数は、568 である。9 教科にその語彙を見ることができる。また教科別の度数は以下の通りである。

国語	数学	理科	社会	外国語	技術家庭	芸術	保健体育	情報
23	9	9	52	15	19	6	432	3

保健体育に数の偏りがあるものの、幅広い教科に見られ、様々な観点を生徒に提示できると考えられる。また、「スポーツ」は生徒にとって思考しやすいテーマであり、作文に対する負担感も少ないと思われる。

5. 2 自由発想

まず、生徒に自由な発想を促す。

(ワークシート例)

Step1 「スポーツ」という言葉から思い浮かぶことを挙げてみましょう。

--	--	--	--	--

(予想される生徒の発想)

・野球 ・サッカー ・バスケットボール ・スポーツ観戦 ・体育 など

生徒は自由発想において、まず個人の体験から発想していく。多くの高校生にとってスポーツは極めて身近なものである。身近な体験から発想できる事柄は、逆に個人の体験から離れることができず、画一的な発想になりやすい。「スポーツ」という語彙から様々な視点・論点を見出すことは生徒にとって容易ではない。また、こうした「スポーツ」のような発想しやすい「テーマ語彙」で話し合いや発表活動を行うと、授業は大いに活性化する。ただしそれは、身近な体験を出し合うものになり、他者の意見が発想の拡充にはつながるものにはなりにくい。

そこで、「教科書コーパス」から導くことのできる「スポーツ」の共起語を生徒に提示することにする。

5. 3 共起語の抽出

「教科書コーパス」中の「スポーツ」を、検索ソフト「ひまわり」で検索する。

そして検索画面に現れる前後の文脈から、生徒の発想の拡充に役立つと思われる語句を共起語として選ぶ。「スポーツ」を含む文脈の中にある語句が、生徒の発想の拡充に役立つからである。共起語の選定については、現状では、筆者の教師経験による個人的判断に頼っている。今後、客観的な共起語の選定方法を案出していきたい。尚、共起語の選定において意識していることは次の点である。

- ・様々な教科から選定する。
- ・生徒が過去の経験や教科書での学習事項を思い起こすとと思われる語句を選ぶ。
- ・教科書本文のデータを閲覧し、文例としても生かせる語句を選ぶ。

(検索例)

no	前文脈	キー	後文脈
183	の例 2. オリンピックと国際交流	スポーツ	の国際交流はさまざまなかたちでおこなわれ
184	より大きいことがわかる。ランニングは	スポーツ	の基本です。陸上競技の選手だけでなく、野
185	るスポーツの意義と発展 地域における	スポーツ	の奨励は、これまで住民どうしのつながりの
186	」 「ささえる」 「しらべる」ということは、	スポーツ	の思想、規則（ルール）や技術、施設や用具
187	な角度から調べてみましょう。 2 競技	スポーツ	の意義学習の目標◆競技スポーツの
188	ができるようになる。 1. 地域における	スポーツ	の意義と発展 地域におけるスポーツの
189	ています。ここでは、現代社会における	スポーツ	の意義や必要性、あるいはその歴史・文化的な
190	っています。このようなことから、生涯	スポーツ	の意義や重要性がさげばれているのです。
191	国際理解 学習の目標 ◆文化としての	スポーツ	の意義を民族スポーツの例をあげながら説明
192	ツライフ 学習の目標 ◆地域における	スポーツ	の意義を説明できるようになる。◆身近な
193	して、解決することも必要でしょう。勉強や	スポーツ	の成績を上げることがストレスの原因になっ
194	目的をはっきりとさせよう 健康のため、	スポーツ	の技能を高めるためなど、目的をはっきりと
195	る。図3 古代中国の健康体操 2. スポーツ	スポーツ	の技術とルールの変遷 スポーツは、こ
196	的行為でもあります。わが国でも、競技	スポーツ	の振興を国の重要施策に位置づけ、ほかの国
197	ことができるのです。また、このような	スポーツ	の文化的内容は、私たちがどのような欲求を
198	966年)をきっかけにして、地域における	スポーツ	の施設、指導者、プログラムなどの整備がす
199	たちに喜び、感動、勇気を与えてくれる競技	スポーツ	の本質は、対戦相手どうしが同意したルール
200	スポーツの歴史を調べてみよう	スポーツ	の歴史は、その内容がさまざまです。まず、
201	ざまな観点から話しあってみましょう。	スポーツ	の歴史を調べてみよう スポーツの歴史
202	うになる。◆各ライフステージにひろがる	スポーツ	の特徴や高校期のスポーツのあり方について
203	るかを説明できるようになる。 1. 競技	スポーツ	の現状 世界中のほとんどの国や地域が
204	技スポーツの意義 学習の目標 ◆競技	スポーツ	の現状について理解できるようになる。◆
205	できるようになる。◆身近な地域における	スポーツ	の現状を調べ、その内容を整理して自分のス
206	展開されています(図1)。このような地域	スポーツ	の発展は、高校生にとってもけっして無縁な
207	したのかを話しあってみましょう。 3	スポーツ	の社会的役割学習の目標◆スポーツ

(ワークシート例)

Step2 教科書の「スポーツ」にかかわる語句（共起語）を参考にしましょう。

テレビ	新聞	ドリンク	オリンピック	文化
ウェア	生涯	国際理解	健康	番組
トレーニング	ルール	熱中	生涯	競争

関心のある語句（共起語）に○を付けましょう。

上記の共起語を示した後、再び「スポーツ」から発想できることを発表させる。共起語から連想を深めたことを発表する生徒や、教科書という発想の手がかりを得て、上記の語句以外のことを思い浮かべることができる生徒も現れるであろう。教科書を手がかりに、生徒の過去の知識や経験が開かれるはずである。

5. 4 文例の選定

語彙から得た発想をもとに、生徒は自らの考えをまとめ、文章化していく。その作業の支援に「教科書コーパス」中の文例を用いる。先に示した共起語をもとに、文例と思考のヒントを生徒に示していく。

文例を提示する上で重要なことは、これまでの実践経験で言えば、教科書の文例に対してできるだけ解説等を加えないことである。解説を加えすぎると、教師のもの見方に生徒を誘導してしまう恐れがある。また、提示した文例にとどまらず、その後生徒が、実際に教科書を手にとって作文の参考にするといった学習活動に発展するようにしたい。そのための思考のヒントとなるような助言を文例に書き添えるとよいと考える。

(ワークシート例)

Step3 「スポーツ」と「〇〇」について考えてみましょう。

「〇〇」に言葉を入れて、スポーツとの関係について考えてみましょう。

(例) 「スポーツ」と「(メディア)」

スポーツを、「見る」「知る」ことについて考えてみましょう。

僕がテレビの長所として考えたのは、「みんなと話題が共有できる」ことだ。一つの番組を話題に、友達との会話が弾む。また、出会ったばかりの人とも、ニュースやスポーツ中継など、テレビで知った共通の話題で会話のきっかけを作り、打ち解けていくことが可能だ。(中学 国語)

多くの人々が、テレビを通じて「スポーツ」観戦を楽しんでいます。「スポーツ」と「テレビ」の関係について考えてみましょう。また教科書の文例には、テレビを通じて知った「スポーツ」の話題が友達との会話のきっかけになるとあります。「スポーツ」と「友達」や「人間関係」について考えてみるのもよいかもしれません。

たとえば、通信技術が発達し、通信網が整備されたことで、外国のニュースやスポーツの試合も、衛星中継を通じて、現地とほぼ同時に見ることが日常的になった。このような通信技術の発達は、現代世界にどのような変化をもたらしているのだろう。(高校 社会)

通信技術の発達で、外国のスポーツも楽しむことができるようになりました。「スポーツ」と「通信技術の発達」について考えてみましょう。あるいは「日本のスポーツ」と「外国のスポーツ」について考えてみるのもよいかもしれません。

また外国のスポーツを、テレビ中継ではなく、インターネットを通じてパソコン上で楽しむ人も増えています。「スポーツ」と「インターネット」といった視点で考えることもできます。

5. 5 表現力向上のための文例

教科書の文例を、生徒の表現力向上に役立てることもできる。

(ワークシート例)

Step4 教科書の表現を参考にしよう (特に下線部)。

- ・地域におけるスポーツの奨励 (高校 保健)
- ・生涯スポーツの意義や重要性がさげばれている (高校 保健)
- ・平和なスポーツの祭典であるオリンピック大会 (中学 国語)

こうした文例を模範とすることで、生徒の表現力は向上していく。教科書の記述が表現の参考になることを気づかせることができる。

5. 6 『現代日本語書き言葉均衡コーパス』の活用

生徒の発想をより一層広げるために、『現代日本語書き言葉均衡コーパス』のデータを

活用する。『現代日本語書き言葉均衡コーパス』には、教科書には掲載されていない、新しい視点で考えさせることができる内容が含まれている。「スポーツ」の共起語として、例えば次のようなものがある。

(共起語)

ドーピング	くじ	アナウンサー	振興基金	文科省
-------	----	--------	------	-----

また文例には以下のようなものがある。

(文例)

文部省の「体育・スポーツ施設現況調査」(平成8年10月)によると、スポーツ施設は全国で約25万8,000か所あり、そのうち学校体育施設が約16万600か所、公共スポーツ施設が6万5,500か所、民間スポーツ施設が約3万1,900か所である。
(『我が国の文教施策』文科省)

こうした『現代日本語書き言葉均衡コーパス』から得た共起語や文例を、「教科書コーパス」の場合と同様にワークシート化していく。『現代日本語書き言葉均衡コーパス』を活用することで、生徒は作文に個別の事例や具体的な数値などを織り込むことができるようになり、いわゆるステレオタイプではない、独自性のある、かつ説得力のある作文を書けるようになる。テーマによっては、教科書と一般社会での扱われ方と比較をさせるといった授業を展開すると、より思考を深められるであろう。

また以前、モニター版を授業で生徒に使用させたことがあったが、生徒は大変関心を持ち、作文の補助として活用していた。生徒自身に語彙や文例を選ばせて、生徒ともに教材を作っていくことができれば、大変有意義な、コーパスを活用した単元型の授業が構築できる。ただそのためには、データ量を制限したり、教育的なフィルターをかけるなどの措置が必要になる。そうした、コーパスを教育で活用するためのソフトやツールの開発が、今後の国語教育におけるコーパス活用における大きな課題であると考えている。

6. おわりに—今後の展望—

本稿では、これまでの実践で行ってきた、多教科型のテーマ語彙の教材化の過程を示した。今後は「単教科型語彙」や「教科書特徴語彙」を生かした教材を作成していきたい。そしてそれらをまとめ、体系的な「作文・小論文のためのコーパス・ワーク集」にしていきたいと考えている。

「テーマ語彙集」は、作文指導に役立てることを主な目的としている。しかし、このようにコーパスから意図を持って語彙を抽出し、その頻度を眺めると、そこに様々な国語教育の研究課題が見えてくる。例えば、新学習指導要領では、各教科において道徳教育を充実させることが改訂の要点の一つとされている。「テーマ語彙集」をもとに、教科書における「道徳」や「しつけ」の扱われ方を論じることも可能であろう。今後、様々な課題を設定して論じていきたいと考えている。

また現場の教師としては、コーパスを教材として、授業で生徒に直接活用させたいという気持ちは強い。それには前述したように、教育用のフィルターをかけたデータ、あるいは

はフィルターをかけるためのソフトやツールが必要である。多くの方々に協力を得ながら、国語教育の分野にコーパスを広めていくために、ぜひともそうした開発に取り組んでいきたい。

文献

河内昭浩（2010）「作文指導におけるコーパスの活用—高等学校での小論文指導を通して—」 解釈, 56 卷 5・6 号, 解釈学会, pp.27-36.

別表 テーマ語彙集

(1)～(3)の「*」は、それぞれの文献でのテーマ語彙の出現を示す。
 (1)『全国大学小論文読解全出題 2002～2006』(旺文社)
 (2)『小論文キーワードブレイク』(第一学習社)
 (3)『大学入試現代文キーワード』(桐原書店)

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コーパス	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWJ モニター
AO入試					0											1
eコマース	AOニユウシ eコマース		*		8				6					2		18
ISO	ISO		*		30			7	7		15	1				278
IT革命	ITカクメイ	*	*		19				14		1			4		62
NGO	NGO		*		79				72	1	1		5			270
NPO	NPO		*		23				16		6		1			306
ODA	ODA		*		58	1			55				2			649
PISA	PISA		*		0											2
PTSD	PTSD		*		2								2			46
QOL	QOL		*		5				4				1			36
あいさつ	アイサツ	*			157	105			15	19	4	8	2	1	3	2,261
相手の立場	アイテノタチバ	*			15	5			4	1	3		2			57
アイデンティティ	アイデンティティ		*		36	8			24	1	2		1			412
青	アオ	*			1,060	228	52	226	320	6	70	127	13	18		10,962
遊び	アソビ	*		*	356	123	7		25	11	141	41	8			3,035
新しさ	アタラシサ	*			4	1		1	1	1		1				65
アナログ	アナログ	*			89			3	4		2			80		228
アニメ	アニメ	*			78	1		2	8	6	14	28		19		501
アルコール	アルコール	*			202	1		149	2	1	4		45			856
アルファベット	アルファベット	*			38	4	5	5	9	2	3	5		5		136
安全	アンゼン	*	*		795	19	2	70								10,098
アンドロイド	アンドロイド	*			1	1										4
安楽死	アンラクシ	*	*		9	3		1	5							42
医学生	イガクセイ	*			2	2										70
生きる	イキル	*			296	69		16	131	4	40	20	7	9		2,209
生きる力	イキルチカラ	*			7				2		1	1		3		106
育児	イクジ	*	*		85	2	1		21		60		1			1,280
医師	イシ	*			122	31		15	23	4	2		47			3,043
医師不足	イシブツク		*		0											3
いじめ	イジメ		*		29	8			15		5	1				1,212
椅子	イス	*			39	8	1		2	1	7	14	5	1		1,798
一気飲み	イツキノミ	*			2								2			6
一寸法師	イツスンボウシ	*			1				1							2

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コアノズ	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モ二ター
遺伝子	イデンス	*			702	11		600	49	1	32		8	1		1,471
遺伝子組み換え	イデンス/クミカエ	*	*		27			1	22		1		3			144
衣服	イフク	*			544	21		12	37		462	4	8			420
異文化	イブンカ	*			41	7			24	3		5	2			76
イベント	イベン	*			47	11		2	13	6	8	3	4			684
違法駐車	イホウチユウシヤ	*			2						1		1			54
癒し	イヤシ	*	*		6	2						4				525
イラク戦争	イラクセンソウ		*		7				7							82
医療	イリョウ	*	*		315	23		54	101	3	29		93	12		5,148
医療過誤	イリョウカゴ		*		0											42
医療費	イリョウヒ	*			17				7		2		8			618
インターネット	インタ-ネット	*	*		694	84	1	94	180	16	112	7	15	185		2,286
院内感染	インナイカンセン		*		1								1			4
インフルエンザ・コンセント	インフオー-ムド・コンセント	*	*		6	2			3				1			71
宇宙	ウチュウ	*		*	613	73	2	420	81	1	11	17	6	2		2,157
宇宙開発	ウチュウカイハツ		*		17			11	5			1				168
運動	ウンドウ	*			3,237	62	40	1,453	932	12	99	28	610	1		6,638
永遠	エイエン	*			46	18		3	19			6				823
英語	エイゴ	*			510	67	11	12	54	342	4	14		6		3,155
映像	エイノウ	*			133	31		11	25		11	41	2	12		1,102
エコクッキング	エコクッキング	*			5						5					0
エコビジネス	エコビジネス		*		0											4
エコロジー	エコロジー		*		5	3					2					105
エネルギー	エネルギ-	*			2,854	43	1	2,178	318	1	276	11	25	1		4,269
エラー	エラー	*			4				3			1				381
塩害	エンガイ	*			1				1							15
古い	オイ			*	59	22		1	16	2	10	5	3			751
大きくなる	オオキクナル	*			248	5	39	179	9		8	2	2	4		522
汚染	オセン		*		393	25		58	159	7	38	2	104			2,401
オン層	オンソウ		*		90	4		55	22	3	1	1	4			149
オタク	オタク	*			1	1										153
おもちゃ	オモチャ	*			80	12	2	23	3		35	5				330
音楽	オンガク	*			635	51	1	9	72	12	7	454	7	22		3,619
温室効果ガス	オンシツコウカガス	*			27	1		7	14	2	2		1			124
温暖化	オンダンカ	*	*		145	3	2	43	80	7	7	1	2			583
介護	カイゴ	*	*		203	1		1	65	5	119		10	2		3,064
外国人労働者	ガイコクジンロウドウシヤ		*		42				42							132
介護保険	カイゴホケン		*		36				18	17			1			587
階段	カイダン	*			83	17	16	17	6	3	16	4	1	3		1,535
海洋開発	カイヨウカイハツ		*		0											33

テーマ/語彙	ふりがな	(1)	(2)	(3)	教科書 コア/久	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モーター
科学	カガク	*	*	*	783	58	21	311	301	9	43	18	13	9		7,032
科学技術	カガクキジュツ	*	*		144	7	3	54	56	2	18	1	3			2,633
科学者	カガクシヤ	*			102	7	4	72	15	3		1				403
格差	カクサ	*	*		191	3			168		4		1	15		1,236
学習	ガクシユウ	*			1,612	697	32	180	250	5	190	56	148	54		3,287
学習指導要領	ガクシユウシドウヨウリョウ	*	*		2					1	1					264
拡大生産者責任	カクダイセイサンシヤセキニン	*			0											8
角度	カクド	*			259	26	93	90	6	6	6	16	12	4		631
核兵器	カクヘイキ	*	*		97	9			87			1				400
学力低下	ガクリョクテイカ	*	*		0											57
化石燃料	カセキネンリョウ	*			89	12		50	24	1	1		1			72
家族	カゾク	*	*	*	1,046	134	2	2	210	10	626	19	42	1		7,405
学級崩壊	ガクキユウホウカイ	*	*		1	1										65
学校	ガッコウ	*			1,414	303	124	76	536	83	115	70	57	44	6	17,280
活字	カツジ	*	*		58	33			23		1	1				292
家庭科	カテイカ	*			36	3					33					47
からくり	カラカリ	*			14	1	2	2	2		1	5		1		93
カルテ開示	カルテカイジ	*	*		1	1							1			1
過労死	カロウシ	*			7	2			4				1			62
感覚	カンカク	*	*		233	55	2	73	47	3	14	29	8	2		2,785
環境	カンキョウ	*	*	*	2,341	92		597	857	15	465	82	203	30		15,792
環境アセスメント	カンキョウアセスメント	*			6				6							34
環境汚染	カンキョウオセン	*			51	2		5	17	3	9	1	14			182
環境税	カンキョウゼイ	*			4				4							41
環境保全	カンキョウホゼン	*	*		58	1		10	24		19		4			1,124
環境ホルモン	カンキョウホルモン	*	*		18			6	5				7			74
環境問題	カンキョウモンダイ	*			242	27		29	137	1	33	5	7	3		650
観察	カンサツ	*			1,483	89	10	1,226	48	4	47	41	14	4		2,726
患者	カンシヤノケア	*			121	16		14	50	1	3		35	2		4,720
鑑賞	カンシヨウ	*			4			1			2	1				170
感染症	カンセンシヨウ	*	*		112	2		6	7	1	3		103			289
願望	ガンボウ	*			21	12		1	5	3						462
偽	ギ	*	*		81	17	23	2	21		4	4		10		1,739
飢餓	キガ	*			19	2			16		1					140
企業	キギョウ	*			990	7		11	873	1	55	4	6	33		21,455
企業倫理	キギョウリンリ	*	*		1				1							5
規制緩和	キセイカンロウ	*	*		14				13		1					477
基礎	キソ	*			339	23	22	53	187		24	13	12	5		4,026
喫煙	キツエン	*			142		1			7	2		132			392
機能材料	キノウザイリョウ	*			0											21

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コア/入	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モ二ター
基本	キホン	*			1,147	103	91	165	409	66	160	62	32	59		11,401
基本的人権	キホンテキジンケン		*		102	3			94		5					143
義務教育	ギムキョウイク		*		21				20		1					281
虐待	ギヤクタイ	*			34	1			2		31					533
キャリアクター	キヤラクター	*			16	1			5	4	3	3				312
キャリア教育	キヤリアキョウイク		*		0											1
ギャンブル	ギャンブル	*			1						1					188
給食	キョウシヨク	*			38	11	9	1	12		5					619
教育	キョウイク		*		580	39		7	400	10	63	16	29	16		14,964
教育基本法	キョウイクキホンホウ		*		8				8							120
教科	キョウカ	*			239	59	30	34	71	17	11	4	8	5		1,806
境界	キョウカイ	*			290	7	17	180	83		2	1				790
共生	キョウセイ	*		*	95	13		14	40	40	20	8				292
行政改革	ギョウセイカイカク		*		15				15							415
競争	キョウソウ	*			213	4		24	159	1	3		21	1		3,579
京都議定書	キョウトギテイシヨ		*		8				8							57
教養	キョウヨウ	*			31	6		2	16		6	1				839
キレル	キレル		*		1	1										44
金属	キンゾク	*			1,047	15	4	794	51		73	105	5			1,566
近代	キンダイ			*	697	48	5	27	539	1	5	71	1			3,064
金融	キンユウ		*		342	1	2		322		15			2		6,228
空間	クワンカン			*	455	33	47	131	43	1	88	109	2	1		2,654
粟	クスリ	*			988	115	4	434	137	5	44	43	204	2		9,852
グリーン	グリーン	*	*		128	2		20	18	54	16	8	7	3		674
グローバルゼーション	グローバルゼーション	*	*		3				2				1			196
クローン	クローン	*	*		62	27		7	23	2	3					82
ケア	ケア	*			31	5	1	5	6		11		3			951
ケアマネージャー	ケアマネージャー		*		3				2		1					3
敬語	ケイゴ	*	*		109	108			1							285
芸術	ゲイジュツ			*	264	45		5	105	5		93	2	9		2,149
携帯電話	ケイタイデンワ	*			116	10		15	20	6	9		6	50		782
言語	ガンゴ			*	360	129	2	7	149	11	26	8		28		2,973
健康	ケンコウ	*			841	20	1	14	68	13	149	2	566	8		5,054
原子力	ガンシリョク	*	*		155	9		72	70	1	2			1		3,571
原子力発電	ガンシリョクハツデン	*			94	6		43	41	1	2			1		834
幻想	ガンソウ	*			43	15			5			23				465
現代	ガンダイ			*	691	174	20	49	292	4	43	80	28	1		3,353
現代社会	ガンダイシャカイ		*		52	8	3		23	1	2	3	11	1		199
憲法	ケンポウ				765	5			747		9		4			3,396
憲法改正	ケンポウカイセイ		*		33				33							131

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コーパス	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWJ モ二ター
権力	ケンリョク			*	283	18		1	263			1				2,184
公共交通	コウキョウコウツウ	*			7				7							178
抗菌	コウキン	*			6			4			2					57
考古学	コウコガク	*			21	3		1	16			1				288
高速交通	コウソクコウツウ	*			7				7							25
公的年金	コウテキネンキン	*	*		6				4		2					252
高分子化合物	コウブンスイカゴウブツ	*			75	1		74								6
効率	コウリツ	*			251	17	3	95	52	1	29	3	5	46		3,150
高齢社会	コウレイシヤカイ	*	*		74				43		25		6			239
声	コエ	*			1,270	592	5	24	188	30	34	252	23	122		21,571
国際化	コクサイカイ	*		*	106	7		2	91		5	1				615
国際協力	コクサイキョウリョク	*			46	3		1	39		1		2			510
国際貢献	コクサイコウケン	*	*		9				9							125
国際連合	コクサイレンゴウ	*	*		149	1			143		4		1			162
国民健康保険	コクミンケンコウホケン	*	*		6				4				2			212
心	ココロ	*			6,155	1,832	330	836	1,900	55	342	373	437	50		58,300
孤食	コシヨク		*		8						8					19
個人	コジン	*			632	83		10	237	3	111	9	78	101		7,952
個性	コセイ	*			153	17		1	47	1	39	41	7			1,164
五線譜	ゴセンフ	*			12							12				6
子育て	コソダテ	*			73	1		4	9		56		3			1,073
古代中国	コダイチュウゴク	*			8	2		1	3			1	1			55
古典	コテン	*		*	258	136	2	6	60			54				792
言葉遣い	コトバツカイ	*			42	39			1					2		232
子ども	コドモ	*		*	1,467	291	36	61	393	27	547	71	37	4		24,522
コマージュアル	コマージュアル	*			63	3	3		40	5	2	1	7	2		191
ゴミ	ゴミ	*			52	9		5	4	1	33					1,093
コミュニケーション	コミュニケーション	*			209	51		4	14	18	32	9	7	74		1,191
娯楽	ゴラク	*			56	7			43		3			3		391
コラボレーション	コラボレーション	*			2									2		40
コンクール	コンクール	*			12	3				1		1		7		127
コンサート	コンサート	*			41	13			7	9	1	8		2		405
コンテスト	コンテスト	*			15			1	2	1	9			2		92
コンピュータ	コンピュータ	*			1,268	41	59	170	136	12	272	77	17	484		2,352
コンピュータ・ウイルス	コンピュータ・ウイルス	*	*		23				2		6			15		21
サイエンス	サイエンス	*			5				3	2						169
財政危機	ザイセイキキ	*	*		10				10							43
在宅介護	ザイタクカイゴ	*			4				2		1		1			73
在宅福祉	ザイタクフクシ	*	*		4						4					78
裁判員制度	サイバンインセイド	*	*		3				3							9

テーマ/語彙	ふりがな	(1)	(2)	(3)	教科書 コア/ハズ	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モ二ター
雑種	ザツシユ		*		24			23	1							119
砂漠化	サバクカ		*		55	3		5	45	1	1					74
サブリメント	サブリメント	*			0											132
差別	サベツ	*			324	12		1	283	5	14	2	2	5		1,939
サミット	サミット	*	*		36				33		3					429
さようなら	サヨウナラ	*			25	12			2	1		10				120
産業構造	サンギョウコゾウ	*	*		44	4			32		1			7		469
産業廃棄物	サンギョウハイキヤプツ	*	*		28			2	12			1	13			351
三歳足神話	サンサイジンシワ	*			0											2
酸性雨	サンセイトウ	*	*		112	3		44	56	1	3	1	4			163
死	シ		*	*	1,385	492	2	96	565	9	46	41	128	6		24,837
ジェンダー	ジェンダー		*		21	2			7		12					198
歯科	シカ	*			9	1							8			475
自我	ジガ		*	*	35	4			24	1	4		2			479
時間	ジカン			*	2,220	325	223	699	312	68	270	38	130	155		24,226
死刑	シケイ		*		58	17			38	1		2				687
資源	シゲン		*		698	30		170	343	1	129	14	10	1		3,568
事件	ジケン	*	*		599	71	2	15	485		7	2	13	4		10,840
自己	ジコ		*	*	507	100		41	225	21	50	14	38	18		6,103
自殺	ジサツ		*	*	26	11		2	10		1		2			1,842
思春期	シシュンキ	*			48			1	2		3		42			275
市場経済	シジョウケイザイ		*		79				78		1					361
地震	ジシン	*			608	9	5	482	70	2	12	2	15	11		3,404
自然	シゼン	*		*	2,136	192	190	589	793	10	79	231	48	4		10,382
自然観察	シゼンカンサツ	*			3				3							44
自然災害	シゼンサイガイ	*			41	1		7	19		7		6	1		210
思想	シノウ	*			648	99		7	523			14	3	2		3,427
市町村合併	シチヨウソンガンツパヘイ	*			2				2							116
しつけ	シツケ	*			88	13		11	16	1	43	2	1			1,502
実験	シツケン	*			2,212	36	26	2,017	82		24	6	6	15		2,984
実用	ジツヨウ	*			122	5	5	59	30		4	8		11		1,120
児童虐待	ジトウギヤクタイ	*	*	*	15						15					96
資本主義	シホシシユギ			*	216	34			182							856
市民社会	シミンシヤカイ			*	30	1			29							115
社会基盤	シヤカイキバン	*			2				2							46
社会的役割	シヤカイテキヤクワリ	*			4				1		1		2			24
社会保障	シヤカイホシヨウ	*	*		121				87		33		1			713
自由	ジユウ	*		*	1,663	148	8	185	1,058	14	86	84	44	36		9,014
宗教	シユウキョウ	*		*	510	17		6	471	1	1	10	2	2		3,751
住民投票	ジユウミントウヒョウ		*	*	41				41							168

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コアノズ	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モーター
重力加速度	ジュウリョクカクソクド	*			63			63								1
儒教	ジュウキョウ	*			82	5			77							231
授業	ジュウギョウ	*			147	48	2	7	28	20	9	5	20	8		2,101
出生前診断	シュウシヨウマエジンダン		*		4				3		1					8
出生率	シュウセイリツ	*			39	2		1	34		2					278
守秘義務	シュヒギム	*			0											122
瞬間	シュンカン	*			149	37	6	85	2	1	5	8	5			2,097
循環	ジュンカン	*	*		291	6	20	161	38		15	2	49			1,297
循環型社会	ジュンカンカタシヤカイ	*			13			3	6		4					127
障書	ショウガク	*			374	32		27	117	1	91	1	89	16		7,050
生涯学習	ショウガクシユウ	*			32				22	1	3			6		201
少子化	ショウシカ	*	*		46	2			23	1	18			2		309
省資源	ショウシゲン	*			19			4	10		4	1				106
少子高齢化	ショウシコウレイカ	*			9				8		1					75
消臭	ショウシュウ	*			0											42
少年犯罪	ショウネンハンザイ	*	*		3	3										82
消費	ショウヒ		*	*	1,219	31	7	194	519	3	418	3	34	10		8,439
消費者	ショウヒシヤ	*			524	11		64	187		230	1	21	10		3,530
情報	ショウホウ	*	*		3,327	361	5	303	680	49	481	21	80	1,347		18,387
情報開示	ショウホウカクシ	*			1						1					96
情報公開	ショウホウコウカイ	*	*		44				41				1	2		357
情報通信	ショウホウツウシン	*			133	2		16	22		17	1		75		866
情報保護	ショウホウホゴ	*	*		9				5				2	2		141
情報リテラシー	ショウホウリテラシー	*	*		4				3		1					48
食	ショク	*			5,068	549	29	978	1,059	109	1,917	70	344	13		37,123
職人	ショクニン	*			80	25		7	39			9				971
食品添加物	ショクヒンテンカブツ	*			49			15			32		2			83
食品	ショクヒン/アンゼンセイ	*			1,005	27		95	66	2	718	1	95	1		2,350
食糧	ショクリョウ	*	*		88	17		14	55	1	1					1,192
食糧自給率	ショクリョウジキョウリツ	*			2	1			1							16
助産婦	ジョサンブ	*			3				3							51
女性	ジョセイ	*	*		594	104	2	9	314	28	71	22	42	2		13,574
自立	ジリツ	*			231	32			99	1	89	1	9			1,333
自律	ジリツ	*			44	5		21	11		5	1	1	1		428
進化	ジンカ	*			352	19	1	279	44		4	3		2		614
人権	ジンケン	*			388	8			354		18	6		2		1,511
人口	ジンコウ	*	*	*	983	37	35	20	797	3	49	3	16	24		5,608
人工知能	ジンコウチノウ	*	*		1	1										53
身体	シントイ	*			213	26	3	10	59	2	62	12	37	2		4,856
心肺蘇生法	シンパインセイホウ	*			9								9			16

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コアノズ	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モ二ター
シンボル	シンボル	*			27	1			15			10	1			360
水害	スイガイ	*			16				15		1					317
水質汚濁	スイツツオダク	*			35			4	17		2		12			452
推薦入試	スイセンニュウシ	*	*		4	4										12
睡眠	スイミン	*			48	2			2	1	17		26			647
数学	スウガク	*			370	8	283	45	25	7				2		1,076
スクールカウンセラー	スクールカウンセラー	*	*		3	1							2			54
ストレス	ストレス	*	*		128				7		5		109	7		1,393
スペースシャトル	スペースシャトル	*	*		22			14	3		4		1			18
スポーツ	スポーツ	*			568	23	9	9	52	15	19	6	432	3		2,889
生活	セイカツ	*			4,330	364	24	428	1,555	30	1,306	179	382	62		18,659
生活習慣病	セイカツシユウカンビョウ	*	*		64				1		9		54			77
生活リズム	セイカツリズム	*			10						7		3			28
生産	セイサン	*			1,543	27	11	190	1,175	3	69	36	19	13		11,476
政治	セイジ	*	*		1,575	69		6	1,488		2	4	2	4		8,598
精神	セイシン	*		*	384	62		2	219	2	36	23	38	2		6,749
生体	セイトイ	*			109	1		99	5				3	1		308
青銅器	セイドウキ	*			37	3		6	17			11				77
政府	セイフ	*	*		1,418	13	1	8	1,368	1	18		5	4		16,596
制服	セイフク	*			78	57			9	1	10	1				483
生物	セイブツ	*			1,949	56	2	1,738	91	10	36	3	13			2,578
生物多様性	セイブツタヨウセイ	*	*		17			12		4	1					53
性別役割	セイベツヤクワリ	*			12				5		7					58
生命	セimei	*		*	491	39		244	118	1	28	30	31			2,956
生命現象	セimeiゲンシヨウ	*			13			9	1			3				27
セカンド・オピニオン	セカンド・オピニオン	*			1								1			25
セキユリテイ	セキユリテイ	*	*		27						2			25		470
世間	セケン		*		52	39			8	4		1				1,891
世帯	セタイ	*	*		114	4	7		22		81					3,097
石炭	セッケン	*			2			2								149
説明責任	セツメイセキニン	*			1				1							65
全員一致	ゼンインイツチ	*			19	18		1								28
戦争	センソウ	*			1,373	97		5	1,247	7	2	13	1	1		6,889
先端医療	センタンイリョウ	*			2				2							9
臓器移植	ゾウキイシヨク	*	*		29	5		1	19				4			55
総合的な学習の時間	ソウゴウテキナガクシユウノジカン	*	*		1	1										37
創造	ソウゾウ	*			140	23	2	6	53	3	10	27	3	13		1,784
疎外	ソガイ	*		*	12	8			4							136
ソフトウエア	ソフトウエア	*			314		11		3		154	1		145		258
尊厳死	ソンゲンシ	*	*		4				4							14

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コアノズ	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モ二ター
大学	ダイガク	*			395	56	4	37	236	9	9	34	4	6		13,304
大学入学資格	ダイガクニユウガクシカク		*		0											12
大学入試	ダイガクニユウシ	*			0											54
大学入試センター試験	ダイガクニユウシセンターシケン	*	*		0											1
大気汚染	タイキオウゼン	*			124	2		19	76	2	6	1	18			455
大衆	タイシュウ		*	*	160	4			154			2				968
代替エネルギー	ダイタイエネルギー				13				13							71
タテ社会	タテシヤカイ		*		0											5
単語	タンゴ	*			138	99	8		1	23	2	3		2		559
短縮語	タンシュクゴ	*			0											1
男女	タンジョウ	*	*		270	21			115	11	104	2	17			2,546
男女共同参画	タンジョウキョウドウサンカク	*			39				12		27					106
知	チ			*	3,609	1,149	138	528	1,003	76	231	186	139	159		57,764
地域	チイキ	*			3,203	45	2	319	2,353	15	322	46	86	15		20,756
地域社会	チイキシヤカイ	*			58	1			30		21		6			721
地域紛争	チイキフンソウ	*	*		49				49							73
地学	チガク	*			18			18								51
地球	チキユウ	*	*		2,335	130	41	1,484	554	15	45	44	17	5		3,131
地球温暖化	チキユウオンドンカ	*			92	2	2	23	55	5	3	1	1			386
地産地消	チサンチシヨウ	*			1				1							17
知識ベースシステム	チシキベースシステム	*			0											7
知的財産権	チテキサイサンケン	*	*		31									31		75
知的障害	チテキシヨウガイ	*			4				2		2					227
地方	チホウ	*			1,080	80	11	108	792	4	31	30	22	2		12,686
中高一貫	チュウコウイツカン	*	*		0											41
著作権	チョウサクケン	*			127			2	11		29	2		83		577
定住外国人	テイジユウガイコクジン	*	*		1				1							7
手紙	テガミ	*			393	256	2	9	52	33	14	19	3	5		2,987
テクノロジー	テクノロジー	*			65	2		17	21	6	9	10				348
デジタル	デジタル	*			99	10		34	38	2	5	8		2		970
デジタル・デバイス	デジタル・デバイス	*	*		5				5							5
デジタルカメラ	デジタルカメラ	*			26	2		10	5	2	3	4				48
デッサン	デッサン	*			33	6			1			26				67
デモクラシー	デモクラシー	*			31				31							125
電子辞書	デンシジショ	*			13					1				12		11
電子政府	デンシセイフ	*	*		3									3		27
電子メール	デンシメール	*			209	15	2	8	16	1	54		1	112		233
伝染病	デンセンビョウ	*			21	2		9	9				1			125
天体	テンタイ	*			275	13	5	249	7					1		174
伝達	デンタツ	*			261	33		58	18		54	21	5	72		1,054

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コーパス	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モ二ター
伝統	デントウ			*	523	37	4	11	317	4	23	125	2			2,414
島嶼	トウシヨ	*			0											67
道徳	トウトク	*			115	11			100	1	1	2				1,052
読書	トクショ	*			99	78	4		8		1	4	4			592
都市	トシ	*		*	1,301	51	16	41	1,124	7	27	21	9	5		10,518
トップ・ダウン	トップ・ダウン	*			0											51
ナノテック/ロジ	ナノテック/ロジ		*		6					6						5
波	ナミ	*			1,914	103	11	1,463	167	1	17	87	7	58		8,793
南北問題	ナンボクモンダイ		*		33				32		1					50
ニート	ニート	*	*		0											79
二酸化炭素	ニサンカタンソ	*			681	5	14	568	60	5	16		13			383
20世紀	ニジュウセイキ	*			244	2	4	38	166	3	3	28				147
日記	ニツキ	*			129	91			20	8	4	6				1,321
日本	ニホン	*			7,829	638	62	613	5,410	245	230	542	64	25		56,113
日本人	ニホンジン	*	*		28				9			19				131
日本語	ニホンゴ	*	*		407	169	3	7	53	144	4	18		9		2,514
日本語教育	ニホンゴキョウイク	*	*		6	2			4							52
日本人	ニホンジン	*	*		482	81	1	4	329	17	27	17	6			6,268
日本文化	ニホンブンカ	*			27	6			16	4	1					201
人間	ニンゲン	*			1,947	495	18	222	807	16	126	135	66	62		20,197
人間らしさ	ニンゲンらしさ	*			3	1			1			1				26
ねっ造	ネツゾウ	*			1	1										112
熱帯雨林	ネッタイリン	*	*		34	1		5	21	6			1			50
熱中症	ネッチユウシヨウ	*			16								16			12
ネットワーク	ネットワーク	*			414	9	1	28	72	1	52	2	3	246		2,380
年金	ネンキン	*			111				89	1	21					6,644
農業	ノウギョウ	*			746	22		26	687	2	6	1	1	1		5,822
脳死	ノウシ	*	*		26	5			16				5			151
ノーベル賞	ノーベルシヨウ	*	*		7			2	5							110
ノーマライゼーション	ノーマライゼーション	*	*		20				11		8		1			46
菌	ハ	*	*		278	54	7	68	37	1	24	13	73	1		4,313
バイオエシックス	バイオエシックス	*	*		3				3							10
バイオテック/ロジ	バイオテック/ロジ	*	*		43			17	17	9						168
ハイテック	ハイテック	*	*		19				15			1	1	2		261
恥	ハジ	*	*		87	72			10	4		1				2,540
発見	ハツケン	*			833	95	33	489	131	6	18	38	19	4		3,641
バブル崩壊	バブルホウカイ	*	*		15	1			14							147
バランピック	バランピック	*	*		5	1			1				3			19
バリアフリー	バリアフリー	*	*		63	3			25	1	14		13	7		173
犯罪	ハンザイ	*	*		90	6			49		3		5	27		3,552

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コアノズ	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モ二ター
犯罪被害者	ハンザイヒガイシヤ		*		1				1							35
万有引力	バンユウインリョク	*			78	4	4	65	5							15
ピオトーブ	ピオトーブ	*			5			4	2							7
干潟	ヒガタ		*		16	1	1	7	5					3		71
引きこもり	ヒキコモリ	*	*		1		1									111
美術	ビジュツ	*			557	13	3		168	7		355		11		1,917
非正規雇用	ヒセイキヨウウ	*	*		0											6
ヒトゲノム	ヒトゲノム	*	*		17	1	5		10					1		42
批評	ヒビョウ	*			66	59		1	5			1				548
非暴力	ヒボウリョク	*			18	11			7							42
肥満	ヒマン	*			14				2		4		8			344
ヒューマノイド	ヒューマノイド	*			1									1		15
病院	ビョウイン	*			111	31		4	26	11	7		27	5		6,530
表現	ヒョウゲン	*			2,121	706	7	71	135	321	61	560	15	245		5,092
平等	ヒョウドウ	*		*	299	27	1		239	8	22	1		1		1,255
貧富	ヒンフ	*			37	2			35							99
ファンタジー	ファンタジー	*			2						1	1				87
風景	フウケイ			*	195	46		10	69	2	4	60		4		1,391
福祉	フクシ		*		380	13	1		238	4	109		15			4,845
仏像	ブツゾウ	*			48	8	1	31				8				166
物理学	ブツリガク	*			132	7	6	85	33	1						340
不登校	フトウコウ	*	*		4				2		2					334
船	フネ	*			723	112	15	88	464	4	3	28	5	4		11,583
フリースクール	フリースクール		*		0											10
フリーター	フリーター	*	*		15	1			3		6		5			210
古さ	フルサ	*			2	1		1								66
ブログ	ブログ		*		0											166
文化	ブンカ	*		*	1,835	111	4	17	1,456	22	77	85	43	20		8,297
文学	ブンガク	*		*	399	153	3	43	183	3	1	12		1		2,951
文化財	ブンカザイ	*			26	3			17		5	1				715
文化相対主義	ブンカソウタイシユギ		*		0											2
文明	ブンメイ			*	371	40	12	30	272	4	3	7	3			1,658
へき地医療	ヘキチイリョウ	*			0											3
偏見	ヘンケン	*			31	2		1	25		2		1			367
便利	ベンリ	*			283	30	22	37	90	2	56	10	10	26		1,264
ホームステイ	ホームステイ	*			16				5	11						60
ホームレス	ホームレス	*			3				3							137
保健医療	ホケンイリョウ	*			4						3		1			161
保健室	ホケンシツ	*			7	3				1	1		2			49
保存	ホゾン	*			621	33		270	72	2	89	2	9	144		2,550

テーマ語彙	ふりがな	(1)	(2)	(3)	教科書 コアノズ	国語	数学	理科	社会	外国語	技術 家庭	芸術	保健 体育	情報	生活	BCCWU モーター
ボランテニア	ボランテニア	*			181	31			64	5	56		19	6		1,435
本物	ホンモノ	*			27	15		3	1		2	4		2		908
摩擦	マサツ	*			273	1		206	61		5					531
まちづくり	マチヅクリ	*			67		1		55		11					392
学び	マナビ	*			311	32	33	16	74	3	21	100	28	4		721
未来	ミライ	*			233	93	1	28	50	13	14	26	3	5		1,635
民主主義	ミンシュユギ			*	192	14			178							1,033
民族	ミンゾク			*	953	23	1	2	852	2	7	41	25			3,079
民俗	ミンゾク			*	42	9			26			6	1			339
メディア	メディア	*			425	144		2	130	1	48	21	8	71		2,080
メディアリテラシー	メディアリテラシー	*			4	4										1
メディアーション	メディアーション	*			0											2
文字	モジ	*			1,605	469	258	45	195	16	90	236		296		4,865
もったいない	モッタイナイ	*			8	3		1			4					396
モデル生物	モデルセイブツ	*			0											13
モバイル	モバイル	*	*		13			1	3		1			8		63
模倣	モホウ	*			25	5		1	8		2	7	2			215
モラル	モラル	*	*		50	32			3		2			13		265
薬害	ヤクガイ	*	*		10				2				8			77
野菜	ヤサイ	*			331	27		21	73	2	201	3	4			2,105
病	ヤマイ	*		*	1,255	262	9	130	252	24	108	10	451	9		20,496
有酸素運動	ユウサンウンドウ	*			13								13			74
優生学	ユウセイガク	*			2				2							13
ゆとり教育	ユトリキョウイク	*	*		0											36
ユニバーサルデザイン	ユニバーサルデザイン	*	*		24	4			4		5		4	7		31
ユビキタス	ユビキタス	*	*		5				2					3		16
養護教諭	ヨウゴキョウユ	*			1					1						58
余暇	ヨカ			*	97	2	1		69	1	9		15			411
欲望	ヨクボウ			*	43	13		4	26							740
予防医学	ヨクボウイガク		*		1				1							12
ラジオ	ラジオ	*			136	18		36	52	1	7	7	7	8		1,275
理科	リカ	*			115	6	4	79	20	5			1			359
リハビリテーション	リハビリテーション	*	*		7						3		4			217
略語	リヤクゴ	*			6	4		1	1							183
良医	リョウイ	*			1	1										2
歴史	レキシ	*		*	995	193	13	114	587	8	18	52	4	6		6,960
レジ袋	レジブクロ	*			0											11
労働	ロウトウ	*		*	1,338	53		2	1,068	2	137	2	57	17		12,690
ロボット	ロボット	*	*		154	19	5		11	7	48	11		53		594
若者	ワカモノ	*	*		151	50		1	63	8	7	2	18	2		2,433

第4章 第4節

他教科における漢字指導実践

—社会科（小学校第6学年）の事例—

棚橋 尚子

1. はじめに

本報告書第3章第4節において、国語以外の教科における漢字指導の提案を行った。この「指導」は、単に新出の漢字に留意させるといった程度のものではなく、教科内容に深くかかわり、その教科の教科書において頻出する、または割合的に多く提示される（教科特徴度の高い）漢字を、現在国語科で行っているような指導法で習得させようというものである。棚橋（2009）は、その場合の課題としてこのような方法を「指導者たちが無理なく受け入れられるか」という点を指摘した。本稿は、実際に小学校第6学年の社会科においておこなった漢字指導実践の報告と考察について述べることを目的とするが、それに先立ち、現職教員から寄せられた意見について検討していく。

2. 現職教員に対する質問紙調査について

2. 1 概要

2. 1. 1 調査の目的

国語以外の教科において漢字指導を行うことについて現職教員の意向を聴取する。

2. 1. 2 調査対象・期間

奈良県下の小学校教員（13名）、平成22年12月

2. 1. 3 質問方法・項目

紙媒体のアンケートによる。質問項目は以下のとおり。

◆新出漢字の学習について実態をお聞かせください。

1. 国語以外の教科で漢字を学習させることがありますか。

- ① ある（「2.」へ）
- ② ない（「3.」へ※念のため「2.」の選択肢をご確認ください。）

2. 「1.」で「①ある」と答えた方にお聞きします。

どの程度の学習ですか（複数回答可）。

- ① 国語と同様、当該漢字の音訓、意味、熟語などを学習させる。（「4.」へ）
- ② 繰り返し書かせて練習させる。
- ③ 国語でも学習した漢字であることを教師の発言などで印象づける。
- ④ その他

◆「1.」で「②ない」と答えた方と、「2.」で①以外の選択肢を選ばれた方にお尋ねします。

3. 国語以外の教科で漢字を扱わないと考える理由をお答えください（複数回答可）。

- ① 学習指導要領上、漢字学習は「国語」で行うことになっているから。
- ② 国語以外の教科では漢字の知識は特に必要がないから。

- ③ すでに国語科で学習済みであり重複するから。
- ④ 国語以外の教科で漢字を学習するのは紛らわしいから。
- ⑤ その他

◆平成17年度版教科書では、第6学年の配当漢字に国語科と社会科との中で、提出頻度に次のような実態が見られます。(各教科書とも教科シェア1位の社のもの)

漢字	社会頻度	国語頻度
権	95	6
皇	77	6
憲	73	4
幕	64	4
将	52	7
聖	37	3

4. このような実態についてご存知でしたか。
- ① 経験的に分かっていた。
 - ② このように差があるとは考えていなかったが、ある程度は分かっていた。
 - ③ このように差があることを考えてみたこともなかった。
 - ④ その他

5. 上記のような「教科性のある漢字」について国語科以外の教科で現在の国語のように指導することについてどうお考えになりますか。以下にお書きください。

2. 1. 4 結果と考察

A 国語科以外での漢字指導の実態

質問1.「国語以外の教科で漢字を学習させることがありますか。」という問いに対しては全員の教員が「ある」という回答をした。しかし、その内実は図1に示すように、国語科

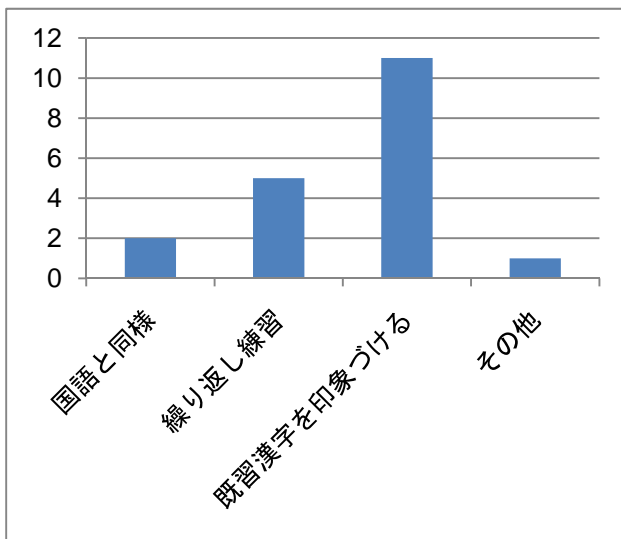


図1 他教科における漢字指導の実態 (人)

でおこなう漢字指導とは異なった質のものである。本稿で提案しようとする指導は、グラフの端にある「国語と同様」という指導であり、少ない母集団ではあるものの、現状ではあまりおこなわれていない指導法である。

しかしながら、このたび13名のうち「国語科同様」の指導をおこなっているという回答者が2名もあったことに正直なところ驚いている。

自由記述を求めた質問5.「教科性のある漢字」を国語科以外の教科で現在の国語のように指導することについてどう考えるか」の記述¹を見ていくと、

「社会科の学習で理解させるためには、指導が必要と考えます。」や「初めて学習する時、社会科で出てくる場合は、国語と同様に取り扱い、くわしく教える必要がある。小学校は学級担任制なのでそれが可能であるし、それを意識して指導にあたっていきたい。」という意見がみられた。

その一方で、「漢字の学習を主眼に置いたとき、教科の枠をこえて漢字の学習を行うこと

¹ アンケートの回答者表記は書かれたままを掲載する。

は必要だと考える。また、子どもにとっても漢字が定着しやすい。ただ、時間の確保が難しく、なかなかできていないというのが現状である。」といったように必要性を認めながらも時間的な問題で難しいという指摘や「各教科により、指導する内容の重要性が違うので、国語以外の教科のときに漢字を詳しく指導する必要はないと思う。」と、カリキュラム構成の点から「国語は国語、他教科は他教科」と指摘する旨の意見もみられた。時間的な問題の指摘は4名ほどがおこなっており、これらの意見を総合して考えると、漢字指導が教科の内容理解につながるという考えの教師は少なくないと考えられる。

B 国語科以外の教科で漢字指導をおこなわない理由

アンケートの自由記述からは、国語科以外の教科で漢字指導をおこなわない理由は「時間的制約」だと読み取れた。質問3では選択式で国語科以外の教科で漢字指導をおこなわ

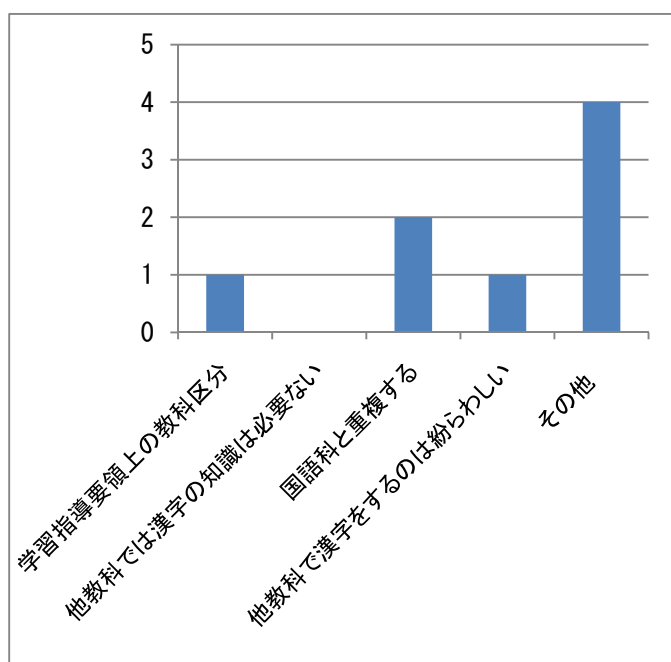


図2 他教科で漢字指導を行わない理由(人)

ない理由を聞いている。それによると、「国語以外の教科で漢字指導は必要ない」とする回答はゼロであり、回答したすべての教師が、教科内容の理解と漢字の習得とを関連させて考えていることが看取できた。「学習指導要領上、漢字学習は「国語」で行うことになっている」という意見や「すでに国語科で学習済みであり重複するから」という回答も少数見られたが、逆に教科別配当が実施されれば、授業内容の重複の問題はなくなるわけである。

こういった意見を踏まえると、学年別漢字配当表を教科別配当にする意味が見出せることが言える。

C 漢字の「教科性」に対する意識

図3は、社会科の頻出漢字を国語科教科書での提示数と比べることを基にして、回答を求めたものである。第3章第4節でも述べたように、教師たちは国語以外の教科において教科内容とかかわって多くの漢字が提示されることを経験的に理解している。理解しながら、「時間的制約」「教科的制約」を受け、実態としては国語科同等の漢字指導ができないというのが現状だと判断してよいように思われる。紙幅の関係ですべての自由記述は載せられないが、特徴的な意見として以下のようなものがあった。

- ・必要があれば指導するが、通常、国語科で先に出てくれば国語科で指導という形が多いように思う。社会科で先に出てくれば、書き順や意味は指導するが、熟語や文章作りなどはしていなかった。国語科以外の教科で指導するほうが、定着すると思うが、国語科の学習と重複するので、時間が限られた今の状況では、なかなかできにくいようにも思う。

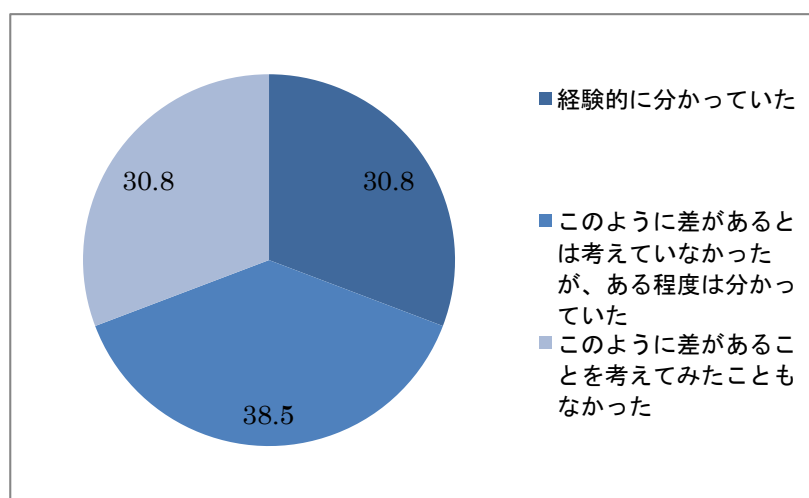


図 3 他教科で頻出する漢字があることがわかっていたか (%)

- ・漢字の学習を主眼に置いたとき、教科の枠をこえて漢字の学習を行うことは必要だと考える。また、子どもにとっても漢字が定着しやすい。ただ、時間の確保が難しく、なかなかできていないというのが現状である。
- ・子どもたちの定着を図るのであれば仕方がないのであります。

上掲の記述は他教科の中で漢字の学習をすることが、漢字の定着そのものに効果があるとするものである。当該教科における漢字学習は、教科内容の理解を促進するだけでなく、漢字の習得に役立つとの判断を呼び起こすことのできる記述だと言える。

3. 社会科における「憲」の学習

3. 1 実践の概要

実際の授業は平成 22 年 12 月 15 日に第 6 学年の社会科で実施した。私が学習指導案を書き、担当教師と打ち合わせをおこなったのち実践に移った。

3. 1. 1 授業者等の情報

対象クラス 奈良県御所市立御所小学校 6 年 3 組 (29 名)
授業者 伊藤輔教諭

3. 1. 2 単元と目標

単元 新しい憲法ができた (大阪書籍)

- 目標
1. 日本国憲法の内容を理解し、戦後すぐに制定されたことに対し考えを持つことができる。
 2. 「憲」の意味を理解し、漢字の定着を図る。

3. 2 実践の記録

以下、授業記録を載せる。発言は教師 (T)、児童 (C) とも意味を崩さない程度に書きまとめた。下線部分は漢字指導に関する部分である。

T: それでは社会の教科書の新しいところを勉強します。112 ページ。なんて書いてある？

C: 新しい憲法ができた。

T: うん。今日は新しい憲法ができたということで、憲法がどんな考え方で作られたのか、その憲法がどうやってできてきたのか、日本が、戦争が終わって負けた後に復活していく原因になったということを勉強していきたいと思います。では 112 ページ、113 ページをあけて下さい。今日はひとりで黙読してもらいます。ただし、年号に線を引っぱってほしいです。

(3 分黙読)

T: そうしたら確認していきましょうか。はい、1946 年、線ひっぱってくれた人？ だいたいの人が引っぱれたんちゃうかな。はい、手をおろして。で、右側のページへ行って、1951 年に線が引っぱれた人？

C: 1947 年。

T: 1947 年？ どこにある？

C: 左下。

T: うん、左下。日本国憲法の三原則のところ、1947 年って出てるね。なるほど。これは、日本国憲法に関するところで、公布と読みます。それで、日本国憲法に関しては実はもう一つ大事な年号で 1947 年、日本国憲法施行っていうのがあります。公布と施行というのは少し違います。公布ってどういう言葉か説明できる人はいますか。

C: 使っていないけど、こんなことをしますよみたいな発表。

T: なるほど。それに対して施行って言われたら、どうなる？

C: その日から使う。

T: なるほど。施行っていうのは実際にその日に行いますよということですね。「行」っていう字が入ってるからな。なるほど。ほかにも 1951 年サンフランシスコ講和会議、平和条約調印とかあるね。ほかの年、見つけた人いますか？

C: 1945 年。

T: 1945 年。どこにありますか？

C: 右の真ん中と、真ん中と下の間に。

T: なるほど。1945 年、インドネシアの独立宣言のところやな？ ここもあったんやね。で、国際連合のところにも年号があった。なるほど。ほか、どうですか？

C: 1950 年。

T: 1950 年はどこですか？

C: 朝鮮戦争と日本のところ。

T: 朝鮮戦争と日本のところの上のほうに 1950 年ってありますよね。はい、では年号はないけど年はわかるというところがありますか。

C: 翌年。

T: うん。翌年というところ。それはどこに書いてあるかな？

C: 右のページの真ん中らへん。

T: 48 か国と平和条約を結び、翌年に…。これのことやな。ということは、1951 年の翌年で 1952 年っていうことになるよね？ 実は、こんなふうに、このページには年号がたくさん書いてあります。ただ、今日は特にこの上の二つの年について考えていきたいと思います。1946 年（現憲法公布）、47 年（現憲法施行）ということは…。戦争が終わった年はいつやったかな？

C: 1945 年 8 月 15 日。

T: 1945 年 8 月 15 日に戦争が終わりました。ということは、45 年に終わって、その一年後ってことやな。一年後には、もう新しい憲法が決まったんですよ。みんなは、憲法について勉強するのは、これが初めてでしたか？

C: 明治時代に大日本帝国憲法があった。

T: 覚えていますか、みなさん。それで、少し思い出してほしいんねんけど、憲法ってどんなものでしたか？

C: ルール。

C: この国のルールで、一番強いルール。

A

T: 一番強いルール。おおっ、最強のルール。なるほど。そういう考えかたもできる。実は一番基になる法律なんですよと、ということです。で、なぜ憲法が基になる法律なのか、分かる？これはね、とっても大事なことやねんけど、「憲」という字が、この漢字自体に、「基になる」という意味があるからなんです。

T: プリント渡します。まず、名前書いてや。

T: (「憲」の成り立ちを見せて)これなんていう漢字が分かります？

C: 憲法の「憲」。

T: おお、その通り。この「憲」っていう字の昔の形は、こういう形になってん。ちなみに、はい、この部分は、何表しているか分かる？

C: 目。

T: 目やね、目のことを表していますよ。では、この部分は。

C: 心。

T: そうやねん。これ心を指します。これね、ふたが目と心をおおっているんです。人間ってさあ、何かを考えると

きに、目とか心で考えるやんか。それを、ルールっていうもので隠してしまうっていうことや。なっ、憲法の「憲」っていうのはそういうふうに形ができていますよ。まずそれを確認しますね。その次に、「憲」っていう字をまず書いてもらいたいと思います。はい、手、挙げて。順番にいくよ。大きく書くよ。筆順確認します。先生に合わせていってください。はい、いくよ。いち、にい、さん…(教師と一緒に筆順を空書する。)はい。筆順確認できたら、ワークシートに書いてごらん。

T: 読みかたは、「けん」やな。意味、ていねいな字で書いていこうな。一番基になるという意味です。

T: はい、それではですね、書けた人は顔をあげてください。それでは、日本国憲法の実際の中身についてみていきたいと思います。では、まず教科書 112 ページから確認していきたいと思います。

(社会科の指導が続く。)

T: はい。日本国憲法の三原則といわれるものがあります。三つ大事なことがあるんです。その三つ大事なことが、これから先、読んでもらう部分にはいっています。それがどこか分かったら、線をひきながら聞いてください。はいでは、このところ読んでくれる人？

(音読)

T はい。では、そこのところ言ってみてください。

C: 国民が主権者となり、天皇は国や国民のまとまりの象徴であるとされました。

C: すべての国民の基本的な人権を保障。

C: 軍隊を持たず永久に戦争をしないこと。

T: まず、この三つを書いてください。丁寧に書いてね。

T: 書けましたか？書けた人は顔をあげてください。最初のほうにも少し話したんですが、1945年8月15日に戦争が終わりました。で、そこからすぐに、日本の憲法が46年に決まり、47年には始まっていきました。なぜこんなにすぐに日本国憲法を作らないといけなかったのか。一番基になるという意味があるんやったらそんなに簡単に決めていいのかな。(中略)三原則、国民主権、基本的人権の尊重、戦争放



図4 漢字の指導場面

棄もしくは平和主義。これは結局、何が言いたかったのか。

C：戦争に負けた。

C：日本が平和になる。

T：うん、きっとそうやろうね。日本が平和になりましたよ、これからはもう大丈夫ですよってことを早く言わなきゃあかんかったわけやな。でもそのためにはちゃんと考えなあかんこともたくさんあったやろうけど、それが日本の国民だけでよかったんですか？—まわりの全世界の人にも平和になったってことをアピールしなければいけなかったんですよ。また詳しくは、今度説明しますが、サンフランシスコ講和会議とか、そこによって日本が独立したこととか、国際連合に加盟したこととかってというのは、結局日本が平和になったからできたことで、それがもし平和になってなかったら、世界は認めてくれないってことなんです。

B

T：では、ワークシートの下の方の「今日学習した日本国憲法の公布についてあなたの考えたことを自由に書いてみよう」をしてみましょう。



図5 ワークシートに記入する児童

3. 3 漢字指導場面

授業記録の中には、2か所漢字指導を目的とした場面があった。それは、AとBの囲みの部分である。Aの部分でおこなった指導は、通常の国語の時間における漢字指導の内容とほぼ同じであり、漢字の成り立ち、読み、字形と筆順、意味を確認し、一度練習をさせている。授業を構想する際に、意味の確認が本時の学習の中心事項である「憲法」の意味理解につながるように考慮しており、その点は担当教師が効果的に授業を進めうまくいったと考える。

また、Bの部分は、「憲法」を必然的に書く場を与えるねらいで設定した学習であり、次ページに掲載するワークシートの下半分を使って自分の考えを書かせるようにした。憲法の公布について書くことになり、自然に習熟できることを目的としている。漢字そのものがまだ理解できていなければ、ワークシートの上部を見ればその漢字が表示されている。さらに、この実践では、Aの部分もBの部分もワークシートを活用しながら授業が進めていけるようにした。

社会科ワークシート 12月15日(水)

氏名

憲 読み テン 意味 憲

◆今日の学習

◆日本国憲法の三原則を書こう。

<input type="text" value="国民主権"/>
<input type="text" value="基本的人権の尊重"/>
<input type="text" value="戦争放棄(平和主義)"/>

日本分県占領図

http://happy-an.teacup.com/shutte/572.html

◆今日学習した日本国憲法の公布についてあなたの考えたことを自由に書いてみよう。

<input type="text" value="「憲」という漢字の成り立ちは深いと思った。"/>
<input type="text" value="日本がすぐに戦争を放棄してくれたから、今の平和な日本があるのかなと思った。"/>
<input type="text" value="そんなすぐに憲法を変えていいのかと思った。"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>

図6 ワークシートの実際

ワークシートの自由記述欄について「憲法」と漢字で書いているかについては図7に示す実態であった。

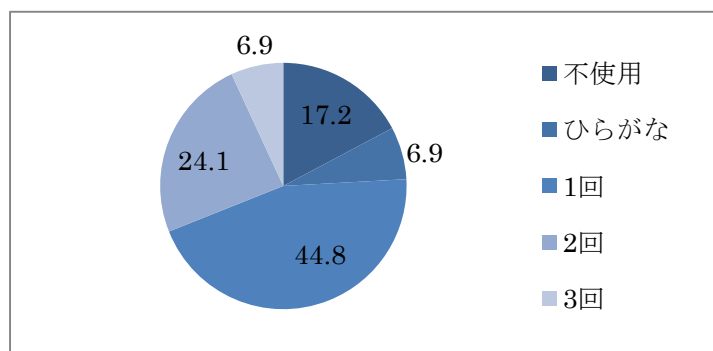


図7 ワークシートの自由記述部分での「憲」の使用 (%)

この結果を見る限りおおむね児童は漢字を使って「憲法」と記述をしていることが理解できる。今回は、対照実験を実施しておらず、確定的なことは述べられないが、教科で中心となる語句と漢字指導を結びつけていくことは漢字習得の上で有効であると言えそうである。また、今回の授業では、内容と授業進捗の関係で自由記述にかける時間が少ししか取れなかった。この点が改善できれば、さらに自然な形での漢字練習ができると考えられる。

4. 授業に対する児童の反応

本実践では、授業後、その効果をみるために以下のようなアンケートを実施した²。

◆今日の授業では、社会科の中で漢字の勉強をしました。このことについてみなさんの意見を聞かせてください。

(1) 漢字の勉強をすることで授業の内容はよく分かりましたか。当てはまるものの番号に○をつけてください。

①よく分かった。 11人 ②まあまあ分かった。 9人
③それほど分からなかった。 6人 ④分からなかった。 2人

(2) 国語以外の教科で漢字の勉強をすることについてどのように思いますか。自分の考えを書いてください。

◆あなたは漢字の勉強が好きですか。当てはまるものの番号に○をつけてください。

①とても好き。 5人 ②まあまあ好き。 9人
②あまり好きではない。 8人 ④きらい。 6人

4. 1 漢字学習と内容理解との関係

本アンケートの実施にあたって一人の児童が、「正直に書いていいの。」という発言をした。

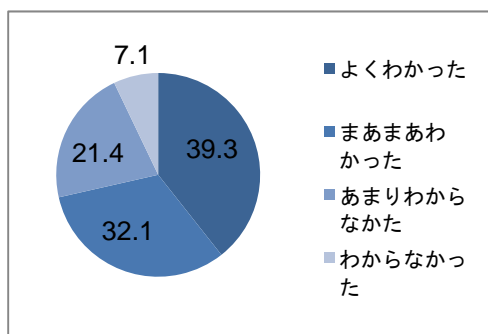


図8 漢字学習と内容理解 (%)

それに対し、指導の教員も参観していた私も「もちろん、そのほうがいいです。」と念を押したため、本アンケートには児童の実感が反映されているものと考えられる。「よくわかった」「まあまあわかった」と答えた児童は全体の7割である。この数値を高いとみるか低いとみるかは前述のように対照的な調査ができていないので明言はできないが、漢字を学んだことが内容理解の一助にはなっていると判断してもよさそうである。しかし、児童の自由記述を検討すると、

そこにはかなりの抵抗感があることが分かる。以下は児童の自由記述の全体である。否定的な感想の部分に下線を付す。

- ① 漢字以外の勉強をしてほしいと思った。
- ② 色々な漢字がわかっておもしろいと思う。
- ③ 例えば算数の問題文の意、国語の文章の意図がわかりやすくなった。

² 実際のアンケートは総ルビ表記とした。

- ④ 社会、楽しみにしてたから、ふつうに社会やりたかった。漢字、あんま好きじゃないです。社会は社会でやりたかったです。あんま、こうゆうのやりたくないです。
- ⑤ 不思議な感じがするけど、新鮮だし知識も増えるし楽しい。
- ⑥ かんじのべんきょうはきれいなのでしたくないです。社会はノートをたくさんかきたいです。もうしたくないです。
- ⑦ 算数の時間にやってもらえるとうれしいです。
- ⑧ あまり国語以外でやることはないのよかったです。
- ⑨ 漢字は好きではないのであまりこのまない。
- ⑩ それはそれであるいみいとおもいました。でも漢字は好きじゃないのでちょっといやでした。めんどくさくはありません。
- ⑪ めんどっちい、おもしろい、おもしろい、漢字うっとうしい。
- ⑫ 社会の時間に漢字の勉強するなんていいかいだったと思う。
- ⑬ 初めは、社会やのになんでやろと思ったのに、「憲」という漢字の意味をして、すごくよかったです。
- ⑭ 国語以外も好きだと思います。でも、算数はちょっと苦手です。
- ⑮ 社会の時間だから社会がしたかった。
- ⑯ 国語以外かんじをやるなんて、びっくりしました。
- ⑰ あまりなかったのでじゃっかん変だった。おかしいと思う。
- ⑱ 難しすぎて分からないけど好き。
- ⑲ 社会でよく出てくる漢字をくわしく勉強することで、その単語の意味が分かって良いと思いました。
- ⑳ すごと思う。
- ㉑ いわかんがあった。
- ㉒ たまに社会の時間にそう言うこともいいと思う。
- ㉓ 別にいいと思います。その漢字を習うことによって、その教科につながるなら、いいと思いました。
- ㉔ 別にいいと思った。社会がきれいだから。
- ㉕ へんだと思った。ふつうかんじのべんきょうで、かんじをするから。

否定的な感想には、二つの類型があった。一つは、漢字そのものの学習が嫌いだということである。これについてはアンケート項目の(2)に関連項目があるが、漢字嫌いの児童が学級の約半数に及ぶことがわかる。もう一つは社会科で漢字を学習することについての「違和感」である。むしろ、2で述べた教師の結果より学習の主体者である児童自身のほうが、抵抗感が高いことが印象的であった。

5. おわりに

今回、実際に社会科の中で漢字学習をおこなう授業の実践を検討することで、教科内容の理解に資する学習、漢字の習熟が期待できる学習が組織できる手ごたえを感じた。それは、現職教員の多くが、それが望ましいと考えていることからわかる。今回の実践は、研究の位置づけの中では、予備的なものとも言えるが、できるだけ早く学年別漢字配当表を教科配当することを考案し、小学校全体の教科カリキュラムを漢字習得の観点から見直していきたいと考える。最後に実践に協力いただいた奈良県御所市立御所小学校と伊藤輔氏に謝意を表す。

文献

棚橋尚子(2009)「教科に特徴的な漢字に関する考察—他教科における漢字指導の可能性—」国語科教育, 67集, 全国大学国語教育学会, p.18.

「言語政策班最終成果 CD-ROM」について

この報告書とあわせて、「特定領域研究『日本語コーパス』言語政策班最終成果 CD-ROM」を作成し配布します。

特定領域研究「日本語コーパス」言語政策班最終成果 CD-ROM

「言語政策に役立つ、コーパスを用いた語彙表・漢字表等の作成と活用」

田中牧郎 相澤正夫 斎藤達哉 棚橋尚子 近藤明日子 河内昭浩 鈴木一史 平山允子

平成 23 年 2 月

文部科学省科学研究費特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21 世紀日本語研究の基盤整備」言語政策班

CD-ROM には、この報告書と同じ内容の PDF ファイルと、次の 5 種類のデータが入っています。

(1) 教科書コーパス語彙表

言語政策班が作成した「教科書コーパス」に用いられている語彙を対象に、語彙素・品詞・語種などの情報について、校種・学年・教科別の頻度を収録した語彙表です。『現代日本語書き言葉均衡コーパス』(BCCWJ) の図書館書籍(固定長)と対比したときの各教科の特徴度も示しています。

(2) BCCWJ 主要コーパス語彙表

BCCWJ のうち、図書館書籍(固定長)、出版書籍(固定長)、雑誌(固定長)、新聞(固定長)、知恵袋(可変長)、ブログ(可変長)の 6 つのサブコーパスの語彙頻度を収録した語彙表です。各サブコーパスに対してカバー率の基準を適用して、5 つに区画した語彙レベルを設定してあります。サブコーパス間のレベルを比較したり、語彙を分類したりすることができます。

(3) 学校・社会対照語彙表

「教科書コーパス語彙表」に収録した語彙頻度・特徴度の情報と、BCCWJ をもとに作成した「BCCWJ 主要コーパス語彙表」に収録した語彙頻度・語彙レベルのうち、特に中等教育での活用を念頭において、必要な情報をまとめ直した語彙表です。

(4) 教科特徴語リスト

中学校・高等学校の教科書の各教科において特徴的な語彙に関するデータを収録したりリストです。短単位を一定のルールで組み合わせて合成名詞等を一語に認定する特別の単位を設定し、教科の特徴が反映しやすくなるような配慮をしています。

(5) NDC ジャンル別漢字出現頻度表

BCCWJ の出版書籍(固定長)のサブコーパスを用いて、NDC(日本十進分類法)に基づ

くジャンルごとに、漢字の出現頻度をまとめた漢字表です。NDCの10分類にNDCが付与されない場合を加えた11のジャンルごとに、出現度数、出現度数%、累積度数%、出現サンプル数、出現サンプルカバー率などの情報を示しています。

CD-ROMのフォルダ構成は、次の通りです。

-- README.txt	・CD-ROMの説明書
-- 報告書.pdf	・言語政策班報告書
--1_教科書/	[教科書コーパス語彙表]
-- 教科書_README.pdf	・教科書コーパス語彙表の説明書
-- 教科書.txt	・教科書コーパス語彙表
--2_BCCWJ/	[BCCWJ主要コーパス語彙表]
-- BCCWJ_README.pdf	・BCCWJ主要コーパス語彙表の説明書
-- BCCWJ.txt	・BCCWJ主要コーパス語彙表
--3_学校社会対照/	[学校・社会対照語彙表]
-- 学校社会対照_TXT_README.pdf	・学校・社会対照語彙表 [TXT形式] の説明書
-- 学校社会対照_PDF_README.pdf	・学校・社会対照語彙表 [PDF形式] の説明書
-- 学校社会対照_統合.txt	・学校・社会対照語彙表 [TXT形式] 分類語彙表番号統合版
-- 学校社会対照_分割.txt	・学校・社会対照語彙表 [TXT形式] 分類語彙表番号分割版
-- 学校社会対照_ア.pdf	・学校・社会対照語彙表 [PDF形式] のうちア行で始まる語
-- 学校社会対照_カ.pdf	・学校・社会対照語彙表 [PDF形式] のうちカ・ガ行で始まる語
-- 学校社会対照_サ.pdf	・学校・社会対照語彙表 [PDF形式] のうちサ・ザ行で始まる語
-- 学校社会対照_タナ.pdf	・学校・社会対照語彙表 [PDF形式] のうちタ・ダ・ナ行で始まる語
-- 学校社会対照_ハ.pdf	・学校・社会対照語彙表 [PDF形式] のうちハ・パ・バ行で始まる語
-- 学校社会対照_マヤラワ.pdf	・学校・社会対照語彙表 [PDF形式] のうちマ・ヤ・ラ・ワ行で始まる語
--4_教科特徴語/	[教科特徴語リスト]
-- 教科特徴語_README.pdf	・教科特徴語リストの説明書
-- 教科特徴語_中_国語.xls	・中学校の国語の教科特徴語リスト
-- 教科特徴語_中_数学.xls	・中学校の数学の教科特徴語リスト
-- 教科特徴語_中_理科.xls	・中学校の理科の教科特徴語リスト
-- 教科特徴語_中_社会.xls	・中学校の社会の教科特徴語リスト
-- 教科特徴語_中_外国語.xls	・中学校の外国語の教科特徴語リスト
-- 教科特徴語_中_技術家庭.xls	・中学校の技術家庭の教科特徴語リスト
-- 教科特徴語_中_芸術.xls	・中学校の芸術の教科特徴語リスト
-- 教科特徴語_中_保健体育.xls	・中学校の保健体育の教科特徴語リスト
-- 教科特徴語_高_国語.xls	・高校の国語の教科特徴語リスト
-- 教科特徴語_高_数学.xls	・高校の数学の教科特徴語リスト
-- 教科特徴語_高_理科.xls	・高校の理科の教科特徴語リスト
-- 教科特徴語_高_社会.xls	・高校の社会の教科特徴語リスト
-- 教科特徴語_高_外国語.xls	・高校の外国語の教科特徴語リスト
-- 教科特徴語_高_技術家庭.xls	・高校の技術家庭の教科特徴語リスト
-- 教科特徴語_高_芸術.xls	・高校の芸術の教科特徴語リスト

	-- 教科特徴語_高_保健体育.xls	・高校の保健体育の教科特徴語リスト
	-- 教科特徴語_高_情報.xls	・高校の情報の教科特徴語リスト
	--5_ジャンル別漢字/	[NDC ジャンル別漢字出現頻度表]
	-- ジャンル別漢字_README.pdf	・NDC ジャンル別漢字出現頻度表の説明書
	-- PB_0_総記.xls	・NDC 「0 総記」の漢字出現頻度表
	-- PB_1_哲学.xls	・NDC 「1 哲学」の漢字出現頻度表
	-- PB_2_歴史.xls	・NDC 「2 歴史」の漢字出現頻度表
	-- PB_3_社会科学.xls	・NDC 「3 社会科学」の漢字出現頻度表
	-- PB_4_自然科学.xls	・NDC 「4 自然科学」の漢字出現頻度表
	-- PB_5_技術・工学.xls	・NDC 「5 技術・工学」の漢字出現頻度表
	-- PB_6_産業.xls	・NDC 「6 産業」の漢字出現頻度表
	-- PB_7_芸術・美術.xls	・NDC 「7 芸術・美術」の漢字出現頻度表
	-- PB_8_言語.xls	・NDC 「8 言語」の漢字出現頻度表
	-- PB_9_文学.xls	・NDC 「9 文学」の漢字出現頻度表
	-- PB_n_分類なし.xls	・NDC 「n 分類なし」の漢字出現頻度表
	-- PB 総計 (度数) .xls	・各ジャンルを合計した総合漢字出現頻度表
	-- PB 総計 (サンプル数) .xls	・漢字ごとの出現サンプル数の一覧表

注意：

○CD-ROM に収録するデータの著作権は、特定領域研究「日本語コーパス」言語政策班の研究代表者・研究分担者・連携研究者・研究協力者、および国立国語研究所が有しています。

○CD-ROM に収録するデータの再配布は禁じます。

○CD-ROM に収録するデータを再利用される場合、必ず事前にご連絡ください。

連絡先： 国立国語研究所コーパス開発センター・田中牧郎

mtanaka@ninjal.ac.jp 042-540-4300 (代表電話)

○この CD-ROM に収録する内容を引用される場合は、必ず出典を明示してください。

出典の表示例：

- ・「教科書コーパス語彙表」（特定領域研究「日本語コーパス」言語政策班、国立国語研究所）
- ・特定領域研究「日本語コーパス」言語政策班（国立国語研究所）による「BCCWJ 主要コーパス語彙表」に収録される「語彙レベル」

○言語政策班の成果物についての最新の情報は、しばらくの間、下記のサイトで提供していきます。

<http://www2.ninjal.ac.jp/tokuteiseisaku/>

成果物一覧

2006年4月～2011年2月 筆頭著者 50音順

◆論文（査読付き）

- ・小椋秀樹・相澤正夫（2007）「現代雑誌 70 誌における漢字の使用実態と常用漢字表 一 国語施策へのコーパス活用に向けた基礎調査一」, 『日本語科学』 22, pp.125-146, 国立国語研究所
- ・河内昭浩（2010）「作文指導におけるコーパスの活用一高等学校での小論文指導を通して一」, 『解釈』 56-5・6, pp.27-36, 解釈学会
- ・鈴木一史（2008）「漢字の読み書き問題の通過率に影響を及ぼす要因について一中学三年生の調査をもとに一」, 『解釈』 638・639, pp.9-17, 解釈学会
- ・棚橋尚子（2010）「教科に特徴的な漢字に関する考察一他教科における漢字指導の可能性一」, 『国語科教育』 67, pp.11-18, 全国大学国語教育学会

◆論文（「日本語コーパス」特集号・依頼）

- ・田中牧郎（2009）「言語政策に役立つ、コーパスを用いた語彙表・漢字表などの作成と活用」, 『人工知能学会誌』 24-5, pp.665-672, 人工知能学会

◆論文

- ・河内昭浩（2009）「作文指導における「教科書コーパス」の活用一マッピングと共起語の援用一」, 『国語教室』 90, pp.29-33, 大修館書店
- ・河内昭浩（2010）「他教科の教科書を使って作文を書く一コーパスを活用した作文指導」, 『月刊国語教育』 30-6, pp.78-81, 東京法令出版
- ・近藤明日子・田中牧郎（2008）「学校教科書の語彙一語種を観点として一」, 『日本語学』 27-10, pp.26-35, 明治書院
- ・鈴木一史（2007）「教科書コーパスによる国語科学習語彙の選定方法の検討」, 『東京大学教育学部附属論集』 51, 東京大学教育学部附属中等教育学校
- ・鈴木一史（2008）「国語教育における『語彙』指導」, 『日本語学』 27-10, pp.16-25, 明治書院
- ・鈴木一史（2009）「文例参照による敬語の指導」, 『東京大学教育学部附属論集』 52, pp.101-108, 東京大学教育学部附属中等教育学校
- ・鈴木一史（2009）「コーパスを用いた授業の可能性」, 『月刊国語教育』 356, pp.34-37, 東京法令
- ・棚橋尚子（2007）「漢字学習におけるルビの有効性」, 『奈良教育大学国文一研究と教育一』 30, pp.77-89, 奈良教育大学
- ・棚橋尚子（2009）「漢字指導の課題と方向」, 『月刊国語教育』 2009年12月号, pp.42-45, 東京法令
- ・棚橋尚子（2010）「文字の学習指導の方法」, 全国大学国語教育学会編『新たな時代を拓く中学校高等学校国語科教育研究』 pp.117-121, 学芸図書

- ・田中牧郎（2006）「学習用漢字辞典と常用漢字表」、『日本語学』25-11，pp.204- 213，明治書院
- ・田中牧郎（2010）「教科語彙を辞書記述に取り込むために」，荻野綱男（編）『コーパスを利用した国語辞典編集法の研究』，pp.101-111，「日本語コーパス」辞書編集班

◆招待講演

- ・田中牧郎（2009）「コーパスの語彙頻度を用いた教育語彙の検討」，韓国日本語学会第20回国際学術発表会，2009.9.19，建国大学校
- ・田中牧郎（2010）「一般語彙リストの作成について」，イタリア日本語教育協会第4回日本語・日本語教育シンポジウム，2010.3.26，ローマ大学

◆口頭発表・ポスター発表（予稿集あり）

- ・相澤正夫・小椋秀樹（2008）「白書コーパスに基づく常用漢字の使用実態調査」、『社会言語科学会第21回大会発表論文集』，pp.206-209，社会言語科学会
- ・河内昭浩（2010）「語彙を豊かにする作文指導—「コーパス」を活用した実践と今後の教材開発に向けて—」、『全国大学国語教育学会 国語教育研究 第119回鳴門大会研究発表要旨集』，pp.186-189，全国大学国語教育学会
- ・金愛蘭・桐生りか・近藤明日子・田中牧郎（2008）「『一般向け専門用語』抽出の試み—医療用語を例に—」、『日本語学会2008年度春季大会予稿集』，pp.199-206，日本語学会
- ・佐野大樹・田中牧郎・丸山岳彦（2010）「『病院の言葉』の類型の推測とモデル化—『現代日本語書き言葉均衡コーパス』における語の使用度数を用いた一考察—」、『日本言語学会第140回大会予稿集』，pp.370-375，日本言語学会
- ・田中牧郎・金愛蘭・桐生りか・近藤明日子（2008）「コーパスによる難解語・重要語の抽出—医療用語を例に—」、『社会言語科学会第21回大会発表論文集』，pp.296-299，社会言語科学会
- ・田中牧郎・近藤明日子・河内昭浩・鈴木一史・棚橋尚子（2010）「『学校・社会対照語彙表』の作成と活用」、『日本語学会2010年度秋季大会予稿集』，pp.249-254，日本語学会

◆口頭発表・ポスター発表（予稿集なし）

- ・河内昭浩（2010）「国語教育におけるコーパスの活用—新しい授業と教材開発に向けて—」，2010年度群馬大学語文学会，2010.8.22，群馬大学
- ・鈴木一史（2009）「生徒作文に見られる語彙の発達」，第41回解釈学会全国大会，2009.8.26，同志社女子大学
- ・鈴木一史（2010）「使用語彙の発達過程と日本語コーパスの可能性」，日本国語教育学会大学部会，2010.12.11，筑波大学附属中学校

◆特定領域研究報告会（予稿集あり）

- ・相澤正夫・小椋秀樹・斎藤達哉（2008）「常用漢字表の見直しに役立つ漢字表の作成と提供」、『特定領域研究「日本語コーパス」平成20年度全体会議予稿集』，pp.121-128
- ・河内昭浩（2010）「コーパスを活用した授業の実際—高等学校における作文指導を中心に—」、『特定領域研究「日本語コーパス」平成21年度公開ワークショップ（研究成果

報告会) 予稿集』, pp.147-154

- ・近藤明日子 (2008) 「中学校教科書の教科別特徴語の抽出—理科を例として—」, 『特定領域研究「日本語コーパス」平成 19 年度公開ワークショップ (研究成果報告会) 予稿集』, pp.181-186
- ・近藤明日子 (2009) 「中学校教科書の教科特徴語の抽出と考察—『現代日本語書き言葉均衡コーパス』の語彙との比較から—」『特定領域研究「日本語コーパス」平成 20 年度公開ワークショップ (研究成果報告会) 予稿集』, pp.117-122
- ・近藤明日子 (2010) 「検定教科書の語彙分析—主要学習部分とその他の部分との比較から—」, 『特定領域「日本語コーパス」平成 22 年度全体会議予稿集』, pp.209-216
- ・斎藤達哉 (2009) 「コーパスを用いた公共性の高い文章における表記改善への視点」, 『特定領域研究「日本語コーパス」平成 20 年度公開ワークショップ (研究成果報告会) 予稿集』, pp.109-122
- ・鈴木一史 (2007) 「国語教育と語彙指導」『特定領域研究「日本語コーパス」平成 18 年度公開ワークショップ (研究成果報告会) 予稿集』, pp.165-174
- ・鈴木一史 (2010) 「コーパスを用いたレベル分けと学習者の実態」, 『特定領域研究「日本語コーパス」平成 21 年度公開ワークショップ (研究成果報告会) 予稿集』, pp.155-160
- ・田中牧郎 (2009) 「語彙政策とコーパス—医療用語を例に—」『特定領域研究「日本語コーパス」平成 20 年度公開ワークショップ (研究成果報告会) 予稿集』, pp.191-198
- ・田中牧郎 (2010) 「国語教育で語彙レベルを使うために」, 『特定領域研究「日本語コーパス」平成 21 年度公開ワークショップ (研究成果報告会) 予稿集』, pp.205-210
- ・田中牧郎・近藤明日子・河内昭浩・鈴木一史・棚橋尚子 (2010) 「<学校の語彙>と<社会の語彙>—「教科書コーパス」と「流通実態サブコーパス」の比較—」, 『特定領域「日本語コーパス」平成 22 年度全体会議予稿集』, pp.171-178
- ・田中牧郎・近藤明日子・平山允子 (2008) 「均衡コーパスに基づく語彙のレベル分け」『特定領域研究「日本語コーパス」平成 19 年度公開ワークショップ (研究成果報告会) 予稿集』, pp.27-32
- ・棚橋尚子 (2010) 「教科に特徴的な漢字に関する考察」, 『特定領域研究「日本語コーパス」平成 21 年度全体会議予稿集』, pp.137-144
- ・棚橋尚子 (2007) 「漢字習得調査に見られる問題点の整理」『特定領域研究「日本語コーパス」平成 19 年度全体会議予稿集』, pp.103-106
- ・棚橋尚子 (2010) 「他教科における教科特徴漢字の指導」, 『特定領域研究「日本語コーパス」平成 21 年度公開ワークショップ (研究成果報告会) 予稿集』, pp.211-214

◆審議会等への提出資料

- ・国立国語研究所 (2008) 「『現代日本語書き言葉均衡コーパス』に基づく漢字音訓一覧表」, 文化審議会国語分科会漢字小委員会漢字ワーキンググループ用資料, 2008.6.2
- ・国立国語研究所 (2008) 「『俺』等の表記について (頻度と内訳)」, 文化審議会国語分科会漢字小委員会用資料, 2008.7.2
- ・国立国語研究所 (2008) 「BCCWJ に基づく表記の実態」, 文化審議会国語分科会漢字小委員会漢字ワーキンググループ用資料, 2008.8.26

◆報告書

- ・田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・小椋秀樹・鈴木一史・近藤明日子・平山允子・金愛蘭・桐生りか（2008）『特定領域研究「日本語コーパス」言語政策班中間報告書 言語政策に役立つ，コーパスを用いた語彙表・漢字表等の作成と活用』

◆語彙表

- ・田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子（2009）『「教科書コーパス」中学校教科書語彙表（短単位）（1）』
- ・田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子（2009）『「教科書コーパス」中学校教科書語彙表（短単位）（2）』
- ・田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子（2009）『「教科書コーパス」中学校教科書語彙表（疑似長単位）（1）』
- ・田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子（2009）『「教科書コーパス」中学校教科書語彙表（疑似長単位）（2）』
- ・田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子（2009）『「教科書コーパス」中学校教科書国語科特徴語彙表』
- ・田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・鈴木一史・近藤明日子・平山允子（2009）『「教科書コーパス」中学校教科書社会科特徴語彙表』



国立国語研究所

