T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	コーパスに基づく辞書記述の精緻化の研究
Title(English)	
著者(和文)	柏野和佳子
Author(English)	Wakako Kashino
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第10284号, 授与年月日:2016年7月31日, 学位の種別:課程博士, 審査員:高村　大也,新田　克己,寺野　隆雄,三宅　美博,奥村　学
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第10284号, Conferred date:2016/7/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Powered by T2R2 (Science Tokyo Research Repository)

コーパスに基づく辞書記述の精緻化の研究

柏野和佳子
東京工業大学　総合理工学研究科　知能システム科学専攻

目次

第 1 章 序論 1
1.1 本論文の背景と目的 1
1.1.1 本論文の主題 1
1.1.2 本論文が対象とする辞書 2
1.1.3 辞書に記載すべき情報と本論文で着目する情報 3
1.1.4 辞書構築における用例収集と分析の重要性 7
1.2 本論文の構成 8
第 2 章 従来の辞書とその問題点 11
2.1 国語辞典とその問題点 11
2.1.1 主な国語辞典 11
2.1.2 国語辞典の問題点 14
2.2 計算機用辞書とその問題点 27
2.2.1 主な計算機用辞書 27
2.2.2 計算機用辞書の問題点 27
2.3 本論文で取り上げる問題 28
第 3 章 コーパスを用いた辞書記述 31
3.1 コーパスの概要 31
3.1.1 コーパスとは 31
3.1.2 主要な英語コーパス 32
3.1.3 『現代日本語書き言葉均衡コーパス(BCCWJ)』構築に至る背景 33
3.1.4 『現代日本語書き言葉均衡コーパス(BCCWJ)』の概要 34
3.1.5 BCCWJ 以降の主要な日本語コーパス 42
3.1.6 コーパス言語学 43
3.2 コーパスに基づく辞書研究の概要 44
3.2.1 コーパスベース英語辞書の研究 44
3.2.2 コーパスベース英和辞書の研究 45
3.2.3 コーパスベース国語辞書の研究 47
3.3 「コーパスベース国語辞書」の提案 52

3.3.1 コーパスに基づくことによる問題解決の見通し 52
3.3.2 提案する辞書記述に対する要件 54
3.3.3 提案する辞書記述により期待される効果と新規性 54
第 4 章 コーパスに基づく見出し語選定と語義区分の精緻化 57
4.1 「古風な語」に関する見出し語選定 57
4.1.1 「古風な語」の現行辞典での扱い 59
4.1.2 「古風な語」の現行の国語辞典における記述 60
4.1.3 「古風な語」のコーパス分析 65
4.1.4 考察 69
4.1.5 見出し語選定の手順 69
4.1.6 コーパスからの自動抽出の試行 70
4.2 語義の区分 71
4.2.1 現行の国語辞典における多義語記述の調査と分析 71
4.2.2 語義の区分の構造化の比較 72
4.2.3 語義の区分と用例提示の比較 74
4.2.4 考察 78
4.2.5 語義の区分の手順 79
4.2.6 下位区分の判断基準の適用例 82
4.3 多義構造の記述 95
4.3.1 『IPAL 名詞辞書』における多義語 96
4.3.2 比喩に基づく意味の拡張関係 98
4.3.3 考察 99
4.3.4 多義構造の辞書記述の手順 110
4.3.5 記述の試行 112
第 5 章 コーパスに基づく辞書記述内容の精緻化 117
5.1 多義性解消のためのコロケーションの詳述 117
5.1.1 コロケーションの多義性 118
5.1.2 コロケーションの重複の分析 119
5.1.3 考察 124
5.1.4 コロケーションの多義性を解消するための辞書記述の手順 130

5.1.5 辞書情報による重複記載の多義性解消の確認 132
5.2 使用域記述のための文体アノテーションと「古風な語」の使用域の詳述 138
5.2.1 使用域の辞書記述のための文体分類 138
5.2.2 文体分類の指標の設計 139
5.2.3 文体分類のアノテーション作業 144
5.2.4 文体アノテーション結果の考察 148
5.2.5 文体アノテーション結果からの語の抽出の試行 159
5.2.6 「古風な語」の使用域の詳述 161
5.2.7 「古風な語」の用法の分類 162
5.2.8 「古風な語」の考察 166
5.2.9 コーパスベースの使用域の辞書記述の手順 168
5.2.10 コーパスベースの使用域の辞書記述の試行 169
5.3 外来語の時間的使用推移の詳述 172
5.3.1 調査の問題点への対処 173
5.3.2 出現率の調査 176
5.3.3 出現率の推移の分析 177
5.3.4 考察 180
5.3.5 外来語の出現率の推移を示す辞書記述の手順 186
5.3.6 分類型の安定性の確認 190
第 6 章 結論 207
謝辞 213
参考文献 215
関連発表論文 243

第1章 序論

1.1 本論文の背景と目的

1.1.1 本論文の主題

人間による文章の理解・作成や計算機による言語処理において，辞書は重要な役割を果たす。したがって，辞書に記載すべき情報や，記載の仕方を検討することは重要な課題である。

これまでは，辞書は個別の目的に沿って作られるのが通例であった。これまでの国語辞書を大別すれば，人間のための辞書と計算機のための辞書に大別され，さらに，例えば計算機のための辞書は，翻訳用の辞書，要約用の辞書，マイニング用の辞書，などに分けられる。後に議論するように，これらが個別に構築されてきた背景には，その構築作業が主に人手に依存しており，多大な作業量が不可避であったことが指摘される。

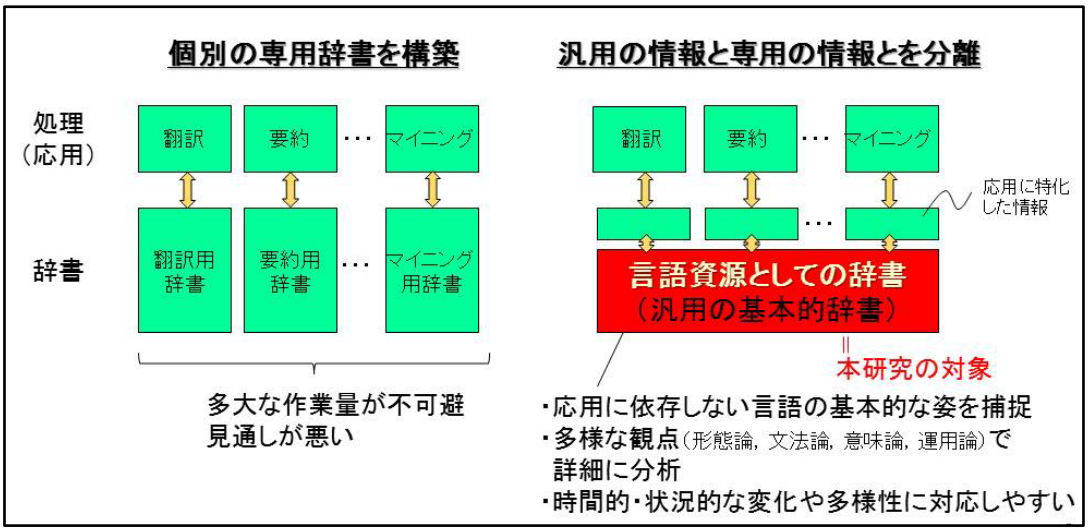


図 1.1 個別の専用辞書と汎用辞書

しかし近年，大規模な言語コーパスが整備されてきたことによって，辞書構築に新たな方法論を適用できる可能性が生まれている。本論文では，個別の目的に沿って辞書を構築するのではな

く，図 1.1 に示すように，汎用の情報，即ち言語がもつ共通の性質をとらえた情報と，専用の情報，即ち個別の目的に特有の情報とを峻別することを提案する。このような構造をとることで，見通し良く辞書を構築できると考えられる。

その上で，本論文では，汎用の基本的辞書の構築に焦点を当て，大規模な言語コーパスを活用した新たな辞書構築の方法を検討する。このような汎用の基本辞書の考察やその体系的な構築は，応用上重要であるばかりでなく，言語学上も重要な課題である。

一般に，辞書の構築は，見出し語選定，語義の区分と記述，見出し語／語義レベル別の情報の記述，という 3 つの工程に大別される。そこで，本論文では，まずは見出し語選定・語義の区分という辞書記述の骨格に対して，コーパスに基づき客観化・明示化することを一つ目の目的とする。次に，見出し語／語義レベル別の情報の記述に対し，コーパスに基づく新たな辞書記述方法を提案し，辞書記述を精緻化することを二つ目の目的とする。さらに，いかに精緻な記述法を考案したとしても，これが実際的に有意義なものであるためには，辞書作成の実務として実施可能なものでなくてはならない。そこで，提案する記述方法に基づく辞書記述の作業手順を具体的に示すことを三つ目の目的とする。

1.1.2 本論文が対象とする辞書

本論文では，汎用的な言語資源となる「コーパスベース国語辞書」の構築のための議論を行う。ここで言語資源とは，言語のありようを可能な限り精緻に体系化し明示化したものを指す。

これまでに市販されている各種の国語辞典も，ここで言う汎用の言語資源と捉えることができる。そこで本論文では，従来の国語辞典の記述および構築過程と対比しながらコーパス活用について議論する。なお，これらの国語辞典は，本来人間が参照するものとして作られているが，国語辞典を電子化したものを様々な言語処理に適用した例も多い[鶴丸ほか 1991，松本ほか 1996，笠原ほか 2004，土屋ほか 2002，鍜治ほか 2003，笹野，黒橋 2008，新里ほか 2011 など]。したがって，国語辞典の記述を精緻化すれば，これを参照する人間にとって有用である可能性があるほか，言語処理の精度向上も期待される。

一方で，計算機による言語処理に用いることを目的とした辞書も数多く構築されてきた。これらは特定の用途を想定したものと，そうでないものとに大別される。前者を専用辞書，後者を汎用辞書と呼ぶことにする。専用辞書には，かな漢字変換や形態素解析といった抽象度の低い処理から，意味検索や自動翻訳などのような抽象度の高い処理を目的とするものまで，様々なものが存在する。汎用辞書の例としては，情報処理振興事業協会(IPA) 技術センターによる『IPAL』

[1987,1990,1996]，日本電子化辞書研究所による『EDR』[1993]，NTT コミュニケーション科学基礎研究所による『日本語語彙大系』[1997]などがあげられる。

近年，抽象度の高い言語処理の研究の進展に伴い，個々の用途における辞書を一から構築することはしばしば困難であることが認識されてきた。その困難さの本質は，言語の様相の複雑さ，その捉え方の難しさそのものに起因していると考えられる。つまり，形態論，文法論，意味論，運用論といった多様な観点からの言語の分析が必要であることや，言語に時間的・状況的な変化や多様性が存在することが，その困難さの要因となっている。このような観点から，本論文では，特定の用途を前提とするのではなく，まずは言語そのものの姿を捉えることを主眼とした「言語資源としての辞書」の構築に着目する。

1.1.3 辞書に記載すべき情報と本論文で着目する情報

では，汎用の言語資源としての辞書に記述されるべき情報は何であるか。

市販の国語辞典には，見出し語，その表記，品詞，意味，用法，用例などが記述されるが，主に母語話者の利用が想定されていること，また，紙幅の制約などがあることから記載される情報は厳選され，その記述は簡潔である。一方，計算機用辞書においては，辞書に取り込める言語情報を豊富に集め，整理し，記載する工夫が行われてきた。これまでの計算機用辞書に記載される情報は，宮崎[1995]を参考に，次の通り，《メタ情報》《形態情報》《統語情報》《意味情報》の 4 つに大別できる。これを表 1.1 に示す。

《メタ情報》とは，見出し語をメタ的に位置づけるような情報のことを呼ぶものである。例えば，基本語，基礎語，あるいは，重要語といった区別である。一般語，専門語，固有名，方言，古語といった区別もある。また，語や共起語の頻度情報，社会言語調査などによる，理解率，使用率や単語親密度，といった統計情報もある。

《形態情報》は，語のありさまについての情報である。音韻，アクセント，連濁化などの読み情報や，字種別表記，標準表記，送り仮名，表記のゆれ，よくある誤用表記などの表記情報がある。また，語種情報(和語，漢語，外来語，混種語)，語構成情報(語基，転成語，派生語，複合語など)や，外来語の原形などもある。

《統語情報》は，語が文中で担う文法的な役割に関する情報である。品詞，活用型に加え，動詞であれば自他の区別などの情報がある。動詞などの活用語には文型(あるいは，格フレーム [1]，

[1] 格フレームとは，用言と，その用言がとる格要素の関係を記述したものである[河原・黒橋 2005]。1970 年代に Fillmore が提唱した格文法に由来する[フィルモア 1975]。

結合価 [2])が記述される。また，文例のほか，文ではなく句レベルの統語情報として共起情報(コロケーション [3])が記述される。慣用表現(定型句，慣用句など)も統語情報の 1 つとして捉える場合もあるが，形態情報に位置づけられる場合もある。さらに，述語の有する文法カテゴリーが記述される。テンス，アスペクト [4]，また，ヴォイス，ムード [5]というカテゴリーがある。

《意味情報》は，その語がもつ意味や，その語の意味的な位置づけに関する情報である。意味を表す言葉に，語義，語釈，釈義，意味記述，基本義，派生義や，意味属性，意義素，意味素性，意味概念などがあるように，その記述方法，記述形式は多様である。意味的位置づけ情報には，使用域(位相とも呼ばれる)の情報や，関連語情報がある。使用域の情報は，男女・年齢・時代・地域・職業・専門や，書き言葉と話し言葉などの相違に応じた言葉の違いを示す。また，待遇(敬語)，評価(プラス，マイナス)，感情(皮肉，軽蔑など)なども時に記述される。関連語は，同義語，類義語，対語，上位語，下位語などである。連想語や，「春，夏，秋，冬」といった組となる語の記述もある。

表 1.1　辞書の記載情報

≪メタ情報≫	≪形態情報≫
基本語，基礎語，重要語といった区別	音韻，アクセント，連濁化などの読み情報
一般語，専門語，固有名，方言，**古語といった区別**	表記，送り仮名，誤用表記などの表記情報
頻度情報	語種情報（和語，漢語，外来語，混種語）
理解率，使用率，単語親密度	語構成情報（語基，転成語，派生語，複合語）
≪統語情報≫	≪意味情報≫
品詞，活用型，動詞であれば自他の区別	**語義**（語釈，釈義，意味記述とも呼ばれる）
文型（あるいは，格フレーム，結合価）	基本義，派生義，**多義の構造**
文例，**共起情報（コロケーション）**	意味属性，意義素，意味素性，意味概念
慣用表現	**使用域**（**位相**とも呼ばれる）
動詞のテンス，アスペクト，ヴォイス，ムード	関連語（同義語，類義語，対語など），連想語

[2] 結合価とは，石綿らによればある原子が他のいくつかの原子と結びついて分子をつくる考え方を，動詞がいくつかの成分を伴うことに応用したものである。フランスの言語学者テニエールが 1959 年に発表し，その後ドイツに入って発達したという[石綿，荻野 1983，石綿 1999]。

[3] コロケーションとは，「慣習的に共に用いられる(２語以上)の語」[Firth1951]，「語と語が共に出現していること」[Sinclair1966]と定義されている[福島 2007]。

[4] テンスとは，成立時と基準時との時間的前後関係を捉えるものである。動詞の述語文だけでなく，形容詞，名詞の述語文にもある。アスペクトとは，動詞の表す動作の過程において，どの局面に焦点を当てるかを捉えるものである。

[5] ヴォイスとは，動詞が表す動作・作用が他に及ぶのか，他からその助作・作用をうけるのかなどを表す動詞の形態的変化を捉えるものである。ムードは，聞き手や述べる内容に対する話し手の心的態度を表す動詞の語形変化を扱うものである。

本論文で提案する辞書記述では，これらの記載情報を，見出し語のレベルと語義ごとのレベルとに構造化して記述する。この構造を図 1.2 に示す。また，本論文では，表 1.1 の情報のうち，特に表中に太字で示した情報に着目することとする。これらの情報の位置づけを図 1.2 中にも太字で示した。本論文でこれら取り上げるべき情報とその理由を以下，A)～E)に述べる。

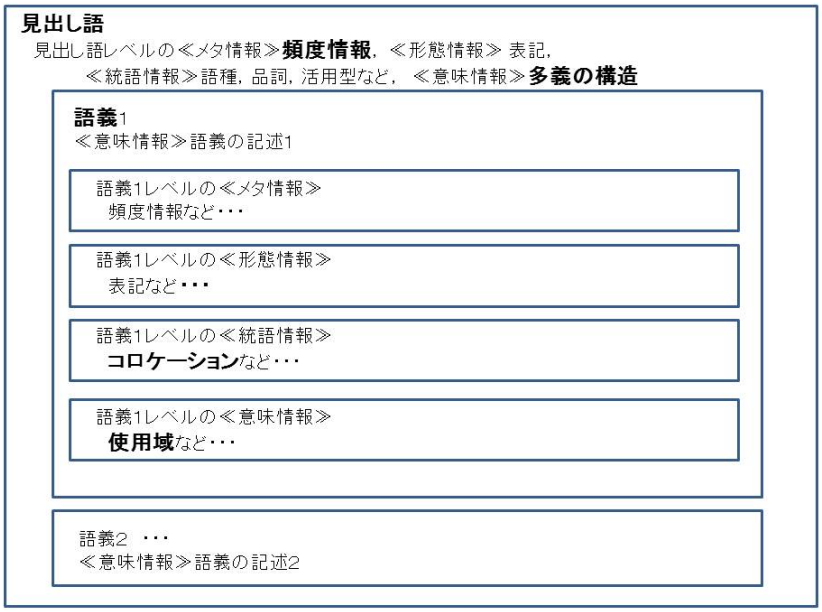


図 1.2 辞書の構造と記載情報

A) 見出し語の選定

見出し語の選定は，辞書構築にとって最も基本的な課題である。WWW をはじめとする電子化されたテキストの収集が容易になったことを背景として，電子化辞書における見出し語については，特に選択することなく単にできるだけ多くの語を集めればよいとする考え方もあるだろう。しかし，そのような場合にも，暗黙のうちに語の収集範囲が仮定され，そのことが辞書の有効性に予期せぬ限定を与える可能性がある。例えば，現代語の辞書を構築しようとして，単純に古語を排除してしまうと，現代語と古語の間に位置して現代でも使われる「古風な語」の見落としにつながる。そこで本論文では，見出し語の選定は明示的に取り扱うべき問題であると考え，まず，見出し語の選定に着目する。

B) 見出し語の頻度情報の記述

従来の辞書において，頻度情報の記述の試みはなされていない。それは，よりどころとする頻

度情報がなかったためである。しかしながら，もし，あれば，高頻度語はすなわち，基本語や重要語であるとの情報になり得る。また，分野別の頻度情報が示せるのであれば，その語の使用実態把握が可能になる。さらに，時系列の頻度情報が示せれば，語の流行り廃りや，時流に合った表現かどうかなどの情報を示せる。なお，十分な頻度情報が得られる場合は，語単位で示すと同時に，語義単位でも記述するのが望ましい。

C) 語義の区分と多義構造の記述

語義の区分と記述は，言うまでもなく辞書記述の根幹に相当する重要な情報である。とりわけ，多義語の場合，従来の辞書からは，語義を区分した結果しかわからず，区分間の差異が不明瞭であるという問題があった。基本義と派生義とで整理し捉えようとする議論[国広 1982a, 1982b, 1997 など]はあったものの，意味や用法どうしの関係を体系的に記述した例は，これまでほとんど見られなかった。辞書に体系的に記述する事例はまだなかった。本論文では，多義語を異なる意味の単なる寄せ集めとして扱うのではなく，多義構造(語義間の意味的関係)をもち得るものと捉える。方法，および区分した結果の間にある意味的関係の体系的記述に着目する。

D) コロケーションの記述

語を理解するためには豊富な用例が欠かせない。特に，用例のうち，語と語の共起関係を記述するコロケーションの豊富な記載は，語の理解や文の適切な生成に役立つ。例えば，『岩波国語辞典 第七版』(岩波書店，2009 年)では「車」の自動車の用例としては「—の両輪」の 1 例のみが掲載されている。これに対し，『IPAL』の「車」の自動車のコロケーションの用例には，①車を動くものとしてみるもの:「—が故障する」「—がスピードを出す」「—が混む」など 36 例，②車を物体としてみるもの:「—が壊れる」「—を潰す」「—を洗う」など 23 例，③車を内部空間があるものとしてみるもの:「—を／から降りる」「—に／へ入る」「—に／へ乗せる」など 14 例，④車を工業生産物としてみるもの:「—を生産する」「—増産する」「—減産する」など 6 例，合計 79 例の記載がある。これらを比較すると，コロケーションを適切に記述することで，語の使われ方，使い方をよりよく知ることができることがわかる。

E) 使用域の記述

さらに，使用域の記述に着目する。英語の学習者用の辞書には，使用域や位相情報を記載したものが見られるが，日本の国語辞典や計算機用辞書では，「使用域」や「位相」に関する記述は

乏しいのが現状である。しかし，男女・年齢・時代・地域・職業・専門や，書き言葉と話し言葉などの相違に応じた言葉の違いに関する記述は，例えば文の生成において文脈にあった適切な語選択を行うためには必要になる記述である。

1.1.4 辞書構築における用例収集と分析の重要性

前節で，汎用の言語資源としての辞書にどのような情報が記述されるべきかと，本論文で特に重要なものとして着目する情報 A)～E)について述べた。本論文で着目する情報は，これまでの辞書には精緻には記述されていない情報であり，その記述の方法論は確立していない。しかし，辞書構築において，単に必要な情報を検討するだけではなく，実現可能な記述の方法論を明らかにすることは，辞書学的観点および実用的観点の両面において極めて重要である。そこで，本論文では，前節に述べたそれぞれの情報の記述方法を第 4 章と第 5 章において，具体的に議論する。

その方法論の特徴は，コーパスを活用する点にある。即ち，コーパスからの用例収集と用例分析に基づいて辞書を記述する。この作業の流れは以下の通りとなる。

① 用例を収集し，「見出し語」を選ぶ。さらに「見出し語」の用例を収集する。
② 用例を分析し，「語義」を認定し，「語義」を記述する。
③ 用例を分析し，《メタ情報》《形態情報》《統語情報》《意味情報》を抽出し，抽出した各種情報を，「見出し語」レベルと「語義」レベルのものとに分けて記述する。

コーパスが整備される以前の従来の国語辞典の構築の場合，この用例収集と分析とは，編集者の不断の努力の積み重ねによる用例採取や，編集者の見識，力量に大きく負うところの考察で行われていた。しかしながら，用例収集に際し，人間が目を通せるテキストのジャンルや量には限界がある。また，用例分析を主観に頼って行うことにも限界がある。その問題を解決するために，本論文では，コーパスに基づく用例収集と用例分析を導入する。また，本論文で提案する辞書記述では，紙幅等の制約をほとんど考慮する必要がない電子化辞書を想定して，詳細な情報を盛り込む方針とする。

1.2　本論文の構成

まず，第 2 章で，先行研究と最近の研究を通し，従来の国語辞典と計算機用辞書の問題点を述べる。

次に，第 3 章で，指摘した問題に対する解決方法を検討する。まず，3.1 節において，近年構築され利用が可能になった『現代日本語書き言葉均衡コーパス(Balanced Corpus of Contemporary Written Japanese，略称 BCCWJ)』を中心にコーパスの概要を述べる。次に，3.2 節において，コーパス分析に基づくことでどのような辞書記述の改善が期待できるかを先行研究や最近の研究の事例により検討する。これらを踏まえ，3.3 節において，本論文で提案する辞書記述方法の基本設計を行い，提案法の新規性と期待される効果について述べる。

続いて，第 4 章と第 5 章で，1.1.3 節において本論文で特に重要な情報として掲げたものを順に取り上げて具体的に議論する。

第 4 章では，辞書記述の骨格となる見出し語選定と語義の区分に関わる検討を行う。即ち，まず，4.1 節において**見出し語選定**に関する問題点を指摘し，現代語と古語の間に位置する「古風な語」を例に，コーパスの頻度情報を使った選定方法について述べる。次に，4.2 節において**語義の区分**について論じる。従来の「国語辞典」の多義語の記述の分析を述べ，多義語の記述に編集者の主観的判断が反映されている現状を示す。その上で，客観的に区分する方法として，コーパス分析に基づく語義の区分方法を示す。最後に，4.3 節において**多義構造の記述**について論じる。多義構造(語義間の意味的関係)の明示的記述方法を示し，記述の客観性が向上し得ることを示す。

第 5 章では，構造化された辞書の記述内容に関わる検討を行う。まず 5.1 節では，**コロケーション(共起情報)の記述**について論じる。従来，計算機による言語理解において多義性解消は主要な問題の 1 つと認識されてきた。しかし従来は，主として語の多義性が注目されており，複数の語にまたがる多義についてはほとんど検討されていなかった。例えば「骨が折れる」には骨折するという意味と苦労するという意味とがあるが，後者の意味は「骨」「折れる」のいずれの語の多義としても捉えにくく，語と語とが結び付いたコロケーションの多義性という観点から捉えるのが適当である。5.1 節では，まずそのことを指摘し，さらに，コロケーションの多義性の要因が述語(動詞／形容詞)による多義性か，名詞やコロケーション自体の多義性によるものかを見極める必要があることを述べ，そのための辞書記述を提案する。

次に，5.2 節では，**使用域の記述**について論じる。より自然な言語生成を行うには，内容や文脈に相応しい適切な語を選択することが必要である。しかし，従来の辞書には必ずしも十分な位相情報が記述されていなかった。この問題に対し，まず，5.3 節のはじめにおいて，使用域のうち，文体情報と言われるものを辞書記述に盛り込むことを可能にするためには，コーパスの側に文体情報を付与しておくことが有用であることを指摘し，その具体的な情報付与の方法を提案する。その次に，「古風な語」を例に，コーパスから得られる出典情報を活用して辞書記述を行う方法を提案する。即ち古風な語の場合，(1)古典の引用での使用，(2)明治期から戦前までの使用，(3)時代・歴史小説での使用，(4)現代文脈での使用，という 4 分類に基づく辞書記述が，従来の辞書記述に比べて語の文体的特徴をより豊富に記述できることを示す。

最後に，5.3 節において**頻度情報による時間的使用推移の記述**について論じる。まず，従来の辞書に頻度情報の記述が不足していることによる問題点と頻度情報の有用性を指摘する。次に，外来語を例に，新聞記事データベースより得られる頻度情報により時間的使用推移を明らかにし，それを辞書に記述することで，それぞれの語が流行語であるのか，廃れつつある語であるのか，安定して使われている語であるのかといった区別ができるようになることを示す。

第 6 章で本論文の結論を述べる。

第2章 従来の辞書とその問題点

本章では，現在市販されている代表的な国語辞典と，これまでに開発された主な計算機用辞書とを取り上げ，それらの問題点を指摘する[柏野 2009]。

2.1 国語辞典とその問題点

2.1.1 主な国語辞典

市販の主な国語辞典の 2014 年 12 月時点の最新版を対象や規模別に分類して，収録語数を表示し，以下に示す(辞典名の五十音順)。なお，以降，略称を用いて言及する辞典については，その略称を『』を付して併せて表示する。

小学生向け(2〜4 万語)

くもんの学習 小学国語辞典 改訂第四版 (くもん出版，2011 年) 25,000 語

三省堂 例解小学国語辞典 第六版(三省堂，2014 年)35,500 語

新レインボー小学国語辞典 改訂第4版(学習研究社，2010 年)37,000 語

チャレンジ 小学国語辞典 第六版(ベネッセコーポレーション，2014 年)35,100 語

中学生・高校生向け(3〜6 万語)

角川必携国語辞典(角川書店，1995 年)52,000 語

ベネッセ表現・読解国語辞典(ベネッセコーポレーション，2003 年)35,000 語:『表現』

ベネッセ新修国語辞典 第二版(ベネッセコーポレーション，2012 年)47,000 語

例解新国語辞典 第八版(三省堂，2011 年)57,000 語:『例解』

小型(6〜9 万語)

岩波国語辞典 第七版 新版(岩波書店，2011 年)65,000 語:『岩国』

現代国語例解辞典 第四版(小学館, 2005 年)69,000 語

学研現代新国語辞典改訂第五版(学習研究社 2012,)75,600 語

三省堂国語辞典 第七版(三省堂, 2013 年)82,000 語:『三国』

新選国語辞典 第九版(小学館, 2011 年)90,320 語:『新選』

新潮現代国語辞典 第二版(新潮社, 2000 年)79,000 語:『新潮』

新明解国語辞典 第七版(三省堂, 2011 年)77,500 語:『新明解』

明鏡国語辞典 第二版(大修館書店, 2010 年)70,000 語:『明鏡』

中型(10～30 万語)

学研国語大辞典 第二版(学習研究社, 1988 年)102,000 語

広辞苑 第六版(岩波書店, 2008 年)240,000 語

大辞泉 第二版(小学館, 2012 年)250,000 語

大辞林 第三版(三省堂, 2006 年)238,000 語

日本国語大辞典 精選版(小学館, 2005 年)300,000 語

大型(50 万語)

日本国語大辞典 第 2 版 全 13 巻(小学館, 2003 年)500,000 語:『日国』

以上の辞典の収録語数の差は, 基本語, 基礎語を十分にカバーした上で, 何をどこまで収録するかによるものである。基本語と基礎語は, 言語研究や言語教育上では次のように区別される。

基本語 ： 統計的な調査により, 使用頻度が高く, 使用範囲の広い語を選定。客観的, 帰納的。

基礎語 ： 日常言語生活に必要な語を選定。主観的, 演繹的。

基礎語のはじまりは, 1930 年, イギリスの言語心理学者 C.K.Ogden の提案した 850 語の Basic English と言われる。この考えを汲んで『Longman Dictionary of Contemporary English』[Procter 1978] では, 7 万 5 千以上の見出し語の説明が 2,000 の基礎語(the Longman Defining Vocabulary)のみで記述されている。日本語には, Ogden に習い, 英文学者の土居光知が選定した 1,100 語の『基礎日本語』[土居 1943]がある。しかし, 現在からこれを見ると「はかま」など, 日常

では使われなくなっている語を含んでいる。現代日本語については，基礎語の考え方に基づき，初級レベルの日本語学習者が学ぶべき語彙を選定した『日本語教育基本 2570 語』[玉村 1987]などがある。

日本語の基本語としては，国語教育のために選定された『新教育基本語彙』[阪本 1984](約 2 万語)や，日本語教育のために選定した『日本語教育のための基本語彙調査』[国立国語研究所 1984]の「基本六千語」と，そのうちのより基本的な「基本二千語」がよく知られている。

また，以上の観点とは異なり，理解できる語彙，使用することができる語彙として基本語を捉えることもある。例えば 12 歳の理解語彙は 2 万 5 千語，20 歳では 4 万 8 千語と見積もられている[阪本 1955]。

以下には，基本語，基礎語という用語は出てこないが，各国語辞典の説明や出版社の紹介文によると，おおむね次のように見出し語が収録されている。

小・中・高校生向け：学習上必要な語を収録
小型：現代生活に必要な語，使用頻度の高い語を収録
中型：現代語と百科項目を収録
大型：国語項目と百科項目を収録，20 世紀までの日本語の集大成

中型以上は専門用語や固有名詞などのいわゆる百科項目を収録することにより，収録語数が大きく増えている。しかし，収録語数の差の要因は百科項目の有無だけではない。学習上必要な語，現代生活に必要な語，現代語，国語項目という段階において，いくつもの観点による収録語の差異が生じている。

まず，古語や方言の収録に大きな差がある。『日国』は，見出し語と異形を含めた方言の収録語数は約 14 万語という。次に，複合語や派生語の収録語数の違いも大きい。例えば，「生活設計」という複合語は中型以上のどの辞典にもあるが，小型以下では「生活」の小見出しにあるもの，「生活」の用例にあるもの，何も記述のないものとに分かれる。また，「行き止まり」という名詞形はたいていどの辞典にもあるが，「行き止まる」の動詞形は中型以上にしかない。

同様に，慣用句，ことわざ，四字熟語などの決まった表現や，それらに準ずる連語，句なども，規模の大きい辞典ほど広く収録するが，小型以下では辞典ごとに収録傾向が異なり，見出しとして立項するか用例として扱うかという差もある。また，成句として定着していないようなものの収録も，辞典間の差が大きい。例えば「赤くなる」「青くなる」を見出しや小見出しにしているのは，『表現』，

『例解』,『三国』,『新明解』と，中・大型辞典であり，限られる。用例にも挙げない辞典も少なくない。より口語的なものになると，中・大型辞典にないものを，小型のみが収録する場合もある。例えば，「それ」のつく連語や句として，「それはそうと」を見出しや小見出しにあげる辞典は多いが，「それはさておき」になると減っていき，「それはないでしょう」まで載せているのは『三国』,『新明解』のみである。

規模よりも辞典ごとの傾向の違いが強いものには，新語，俗語，外来語の収録がある。特に，新語，俗語の収録に積極的なのは小型の『三国』である。初版(1960 年)は 57,000 語であったものが，現在の七版に版を重ねるたびに，新語を中心に，毎回約 4,000 語を増補している。他方，新語や俗語の収録に比較的慎重な『岩国』は，初版(1963 年)57,000 語，現在七版であるという点は『三国』と偶然にも同じであるが，版を重ねるたびに増補した語は約 1,000 語であり，両者の増補のペースの違いは，新語採用の積極性の違いを反映している。

収録語の内訳を載せている国語辞典はほとんどないが,『新選』には明記されている。それを参考として表 2.1 に示す。

表 2.1 収録語の内訳(『新選』)

総語数＝90,320 語
一般語＝76,536 語(84.7%)
固有名詞(人名・地名・書名など)＝1,430 語(1.6%)
慣用句(故事・成語を含む)＝3,846 語(4.3%)
古語(枕詞を含む)＝4,791 語(5.3%)
漢字字母(漢字解説の見出し)＝3,602 語(4.0%)
その他(誤り語形など)＝115 語(0.1%)

2.1.2 国語辞典の問題点

現代の国語辞典では，見出し語を五十音順に並べ，その表記，品詞，意味，用法，用例などが記述されるが，各項目の記載方法のどれをとっても，辞典間でさまざまな考え方や立場の違いによる差異がある。また，それぞれに問題が指摘されている。ここでは，現代の国語辞典の記述における論点を先行研究を踏まえて示す。

2.1.2.1 見出し語

見坊[1990]は，辞書に載せる語は，①実際の使用例が多いこと，②語の質的評価の高いこと，③何度も見かける語であること，④その辞書の性格・規模・使用者層にあうことを選定の基準とし，「自分がそのことばを知っているかどうか、とか、あるいは、そのことばを好ましく思うかどうか、とかの個人的評価もあとまわしにする。」と述べている。倉島[2008]は，「筆者は実務の経験から、一定の段階まで客観的基準で選定した後は、主観的な判断で選定してよいし、現実にそうせざるを得ないところがあると考えている。」と述べ，「見出し語の全体を見渡しつつ、適切な主観的な判断によって選ばれた項目を加えることにより、その辞書の見出し語に個性的な特徴をもたせることができると考えるのである。」と述べる。また柴田[1987]は，「辞書を編集していて、語を新しく採録することよりも不要語を削ることのほうがはるかに難しい。知っている語は削れない。歴史的役割が終ったといっても、すぐに消してはあとで困る。そんなことを考えたら一語も削ることはできない。どんな語でも、見出し項目から追放することは、編集者にとって最も苦しい仕事である。」と述べる。

このように，見出し語の立て方，すなわち，収録語の範囲の定め方が，国語辞典における基本的論点の一つになる。この問題はさまざまな観点から議論されている。例えば，大塚[2006]は，見出し語選定に際し，単位の長さと幅についての選択が問題になることを指摘する。矢澤[2009]は，「～に越したことはない」は，「に／越し／た／こと／は／ない」という六つの語から成り，「越し」は動詞「越す」だから，辞書では「越す」の項目に出ているだろうという分析を経る困難さを指摘し，「に越したことはない」がこのままの形，あるいは「こした」「こしたこと」くらいで引ければ便利であると提案する。井上[2009b]は，例えば，「～にしては」のような連語が見出しにない問題を指摘する。佐竹[2000]は「おかたい」「おやすくない」などの見出しの必要性を指摘する。これについては，相澤[2006]も「お茶」はあって「おビール」がないのは妥当だが，社会状況の変化を考えると出会ったことばそのままの形でひけることも検討していくべきと述べている。また，宇野ほか[1971]で野元菊雄は，「「くださる」はあるのに一番多く使われかつ不規則の「ください」が見出しにない辞書があるのは不満だ。これでは外国人には用が足りない。」との問題を指摘する。複合語については，「構成要素から全体の意味が比較的容易におしはかれる語は辞書にのらないという一般的な傾向が示されている」[石井 1989]が，どこまで載せるべきかの十分な議論はまだない。

さらに，収録語の範囲以外にも問題がある。例えば国広[1997]は，同音異義語の語義とするか，単一語の多義とするかの辞典による違いを指摘し，その極端な例として，『三国』では，「図る，計る，測る，量る，諮る，謀る」が別項目扱いであるのに対し，『岩国』では全部を同一項目で多義扱いにしている点を取り上げている。これに関し，竈岡[2006]は，漢字が違うと分ける傾向もあるが，

逆に表記が同じであれば，たとえ語源が異なっていても分けない傾向が強い点(例えば，「離島」[島を離れること／離れ島])を問題にし，同形・同音異義語の見直しも必要であると問題提起をしている。

2.1.2.2 表記

おおむねどの辞典も内閣告示の「常用漢字表」「現代仮名遣い」「送り仮名の付け方」に基づいた表記を示し，一見，議論の余地は少ないようにも思われる。しかしながら，倉島[2008]が，「日本語には厳密な意味での正書法が存在していない。」ことにより，「決まりの解釈の相違や運用の仕方の相違によって、いく通りもの表記が可能になる。」と指摘するように，国語辞典の表記の例示が国語辞典によって異なる場合が少なくはない。例えば，異なる漢字表記がある場合に，一括表示をして使い分けを示さない辞典もあれば，個々の使い分けを詳細に示す辞典もある。さらに，「標準表記」を明示するものもある。『新選』では，「また」に【また】というひらがな表記を，「または」に【または・又は】という漢字表記とひらがな表記を併記する。『新潮』は，見出しそのものについて，和語をひらがなで，漢語と外来語とをカタカナで示し，語種を明示する。

笹原[2012]は，国語辞典と，漢和辞書，特殊辞書と比較すると，表記の記載例はおのおのの役割と目的によって収録範囲や対象が大いに異なっていることを明らかにしている。特殊辞書とは，例えば『当て字・当て読み漢字表現辞典』(笹原宏之編，三省堂，2010 年)である。国語辞典では「さだめ」について常用漢字表に従った表記「定め」のみを記載するが，『当て字・当て読み漢字表現辞典』は「運命」，「宿命」等を記載しているといった例が挙げられている。国語辞典が表記に関しどこまで記述主義的に実例を示していくかは今後の課題の一つである。

2.1.2.3 品詞

品詞は，多くの辞典は学校文法に基づいた分類を凡例や付録に示し，それを付与する。しかしながら，品詞だけで語の挙動の振る舞いを説明しきれない場合がある。『岩国』のように，付録に学校文法とは異なる体系の品詞体系の概説を載せ，学校文法よりも詳細な品詞を付与する辞典もある。例えば，いわゆる「形容動詞」は「～ナ＋名詞」と接続するもの(例：静かな時)と「～ノ／～ナ＋名詞」と接続するもの(例：静寂の中／静寂な街)とに区別する。さらにこれを進めるものとして，水谷，星野[1994]は，無活用語(名詞から副詞)を 100 以上に細分類する試みを提案しているが，国語辞典では採用していない。

2.1.2.4 語義の区分・記述

語義の区分・記述(意味記述)の問題点は様々な観点より活発に議論されている。最近では語佐藤, 夏目[2014, 2015]が新しい日本語辞書定義文型を提案しているが, 中野[1985]には 1970 年代から 1980 年代前半までの議論の多くが示されている。まず, 次の暮らしの手帖社[1971]の「国語の辞書をテストする」のような利用者側からの批判が取り上げられている。

1. 意味を説明するその文章が, りっぱな美しい日本語でないことが多い。
2. その文章が短すぎて, なんのことかよくわからないのも多い。
3. やたらに, あとを見よ, これを見よ, は不親切である。
4. いちばんよく使われる意味を, はじめに出してほしい。
5. 用例, 引例はその言葉のニュアンスをつかむために, 役立つことが多いが, そのためには, 古文漢文などをのせるのは, 意味がない。

宮島[1972]は辞典における語義記述の不完全さについて次のように述べている。

1. 辞典における意味の説明は, あまりに短く不完全である。単語の意味, 用法における中間的なもの, あいまいなものを排除して, 断定的に一面的にのべる傾向をもちがちだ。
2. 辞典の説明は結論だけで, その結論に到達する根拠をしめしていない。
3. 辞典の見出し語は一般に語形の順に並べられており, (中略)意味の体系的な関連がとかれていない。単語の意味は孤立してあるのではなく, これをとりかこむほかの単語によって相対的に制約されている。他の単語との共通点と区別とをあきらかにできないような説明では不完全である。

さらに続けて, 杉本監修[1982]は国語辞書 10 冊について比較検討し, 説明文に現れた差別意識の問題, 主観的な解説の有無, 内容の誤り, 不十分で紛らわしい説明の指摘等を批判的に報告している。鈴木[1983, 1985]もまた, 小型の国語辞典 8 冊を選び,「用法・用例に対する配慮」や「語義は正確・妥当・的確か」等を分析し, 批判している。

このように, 辞書を利用する側から批判がなされる一方, 辞書を編集する側からの議論も重ねられてきている。見坊は,「國語辭書の欠點」[1951],「辞書における定義づけの批判」[1952a],「国語辞書の盲点」[1952b],「現代語辞書の批判と編集の実際」[1953a],「辞書における定義の問題」

[1953b],『辞書をつくる 現代の日本語』[1976a],『ことばの海をゆく』[1976b],『辞書と日本語』[1977]等において，収録すべき見出し語，語義の区分・記述，辞書編集の方法論等を議論する。語義記述，すなわち言葉の定義について，それらのうち見坊[1953b]では次のように述べている。

1. 定義ということばの論理的な規定は，概念の内容を明確に限定し，ほかの同類の概念との差を区別することであるが，辞書においては，コトバの意味をハッキリさせることぐらいの気持である。
2. 一般の定義においては，かえって，むずかしい用語を作り，大衆の日常的用語と手を切ることに学問の尊さをみとめようとするかの傾向さえ見られた．ちかごろはこうした悪い傾向がへってきて，たいへん喜ばしい。
3. 文学者でさえ，簡潔・厳密を要求すれば，数式的表現をとりたくなる。
4. やさしい見出し語の概念とか，ふくみとかいうものは，直観的にはよくわかっているのである。ただそれを客観的にあらわにしてみせることがむずかしいのである。
5. 国語辞書は，専門語を一般人にやさしくとりつぐという役目をもつべきである。だから国語辞典の編集者は，一般読者になりかわって十分にかみくだき，現代人の現代的感覚にうったえて，よくなっとくのいくように 解説することが大切なのである。

さらに，辞書を編集する側から，水谷[1961],山田[1962]の議論があり，これらは見坊[1977]でも大きく取り上げられている。水谷[1961]は次のように述べている。

1. 基礎語ほど語釈が難しく，いずれは循環が断ち切れない。「赤い」「甘い」など，その感覚を引き起こす経験を直接に与える方が，百万言の解説にまさるとさえ思われる語もある。
2. 意味は指されるものではない。「こと」と「もの」は意味を区別するが同じ対象を指すことはあり得る。対応のさせ方に意味を認めるべき。しかし，そこには認識の態度と独立に語釈を行い難いという問題がある。しかし，語釈なるものを本義と適用範囲とに分けて考えるとき，本義の部分では，語釈者の認識上(人生観・世界観など)の立ち場の違いをある程度避けることができる。
3. 語釈中に用いる object language と metalanguage とのレベルの相違を(たとえ気づいても)無視して，同列の形式で語釈を与えてきた点が問題である。(例:「銀世界」の語釈で「・・・の景色を言う語」とする部分は metalanguage になるがその区別は示されていない。)

4. 言い換えは一概に悪いと決めつけることはできない。語釈は本質において言い換えである。しかし，語の言い換えは往々にして最も稚拙な語釈である。語の意味の区分を言い換え語の異同に寄らず，論理的な包含関係によって決めることはできる。
5. 意味の違いとは何か。どんな条件が満たされた時，語義の別を立てるか，何によって辞書の語釈の番号を打つか，という問題もある。主観による内省で意味の記述はできない。観察可能な外的条件を突き止め，それとの対応を記述すべきである。しかし，客観性の保証は依然問題として残る。日常言語は由来客観的でないがその認識活動には客観性が要求される点が言語研究の難しいところである。

見坊[1976a]は上記 1.の循環すなわち「堂堂めぐり」は 2 種あるとし，次のように述べている。

1. 第一の堂堂めぐりは技術論にぞくする堂堂めぐりである(例:「長い」を「短いの反対。」とし，「短い」を「長いの反対。」とするような誤り)。これは，適当は方法で予防することができる。両方を別な共通の基準から定義することによって救うことができる。
2. 第二の堂堂めぐりは本質論にぞくする堂堂めぐりである(例:「もの」「こと」「ところ」などの定義づけの道具専門のことばの定義。)。これは必然的なものであって，予防することができない。定義づけの道具専門のことばは，数学で言えば公理にあたるような存在であって，いちだんとやさしい用語でこれらを定義することはできない。だから、この段階で堂堂めぐりがおこるのは，さけられないことである。
3. 第一は，いちだんとやさしいことばで定義することはほどほどにして，見出しの概念を適当な用語体系で徹底的に説明すること。
4. 第二は，抽象の度合いやむずかしさの程度に応じて、ことばを幾種類かの段階に分け、同じ段階の用語を使って定義をすること。

見坊[1977]は，「堂堂めぐり」の解決にあたり，水谷[1961]は上記 3.による方法をとり，自分は 4.による方法をとっていると述べている。

山田[1962]は，現行の国語辞典の問題として「漢語が大いに不足している」点を挙げ，現代語を「落ち穂拾いのようにひろい上げて補充して進歩を誇る」ことを批判する。そして，「辞書史を古く遡ることによって、近代の辞書の欠陥を十分に探索すべき」と述べ，日本語の歴史的記述を主にした，歴史主義的な国語辞書の追及を主張している。これに対し，見坊[1977]は，その意義を認

めながらも，自身の小型辞書の立場から受け止めると，「実例による裏づけは、過去、現在、将来のあらゆる辞書を通じて、常にもっとも必要でたいせつな条件」であると反論している。

1980 年代後半以降も，語義記述を中心とした辞書論は続いている。野村[2006]は，基礎日本語のカバー率という観点から，小学生，中学生，一般向けの国語辞典を比較し，全体で基礎語 1,100 語のカバー率は 3 分の 1 強であり，小学，中学，一般の順に当然小さくなることを示した。なお「基礎語をメタ言語として意味の記述をおこなうことは，困難ではあるが，不可能ではない」と述べている。

さらに，語義の区分・記述の態度については，大きく 3 つの議論がある。1 つ目は，意味用法をまとめて説明するか，細分して説明するかという態度の違いである。山田[2005]はこれを「中核的意味」を求める態度と，「用法列挙」を求める態度の違いであると説明し，「取る」は，『岩国』が 7 個にまとめているのに対し，『明鏡』は 59 個に細分する例を挙げている。

2 つ目は，語義の説明においては，編著者の主観性の多少による態度の違いである。独特の説明では『新明解』がよく知られる。例えば，「たらこ」を「普通，一腹がサイドカーのように並んでおり，焼いたり酢の物にしたりして食べる」と説明するように，独特のたとえを用いたり，生活上の実感に踏み込んだ説明をしたりする。

3 つ目は，規範的であるか記述的(あるいは慣用的)であるかという態度の違いである[見坊 1990, 倉島 1997a, 2008, 神永 2013][6]。これは主に，新用法，誤用，特例などの扱い方に現れる。例えば，一般に辞典は誤用による新用法の認定には慎重であるが，『新選』は「さわり」の区分の一つに「曲などの出だしの部分」を最近の用法として認めて挙げており，他辞書が「聞かせどころ」のみ挙げているのとは対照的である。

この 3 つ目については，1980 年代以前の議論に戻るが，見坊の「かがみ論」が重要な論点になっている。「かがみ論」は見坊[1953a]をはじめ，多くの見坊の論考で繰り返し言及されているが，ここでは見坊ほか[1982]『三省堂国語辞典』第三版の序文をひく。これは 2014 年刊の第七版の序文にも採録されている。

> 辞書は〝かがみ〟である。――これは、著者の変わらぬ信条であります。
>
> 辞書は、ことばを写す〝鏡〟であります。同時に、

6 見坊[1990]では，規範主義：正しくない語形・意味を一切みとめず、雅馴なことばの登載につとめる立場，記述主義：正しくない語形・意味であっても、一般化したとみとめられるものは、積極的に取り入れ、または、誤りであることを注記などの形で示す立場，と，説明している。

辞書は、ことばを正す〝鑑(かがみ)〟であります。

〝鏡〟と〝鑑〟の両面のどちらに重きをおくか、どう取り合わせるか、それは辞書の性格によってさまざまでありましょう。ただ、時代のことばと連動する性格を持つ小型国語辞書としては、ことばの変化した部分については〝鏡〟としてすばやく写し出すべきだと考えます。〝鑑〟としてどう扱うかは、写し出したものを処理する段階で判断すべき問題でありましょう。

そのことばを見出しに立てる、ということがまず大切です。

そのことばが社会にあることを知り、次に、そのことばが辞書にないことを知る——新しい見出しが辞書に立つまでには、この二つの手続きがどうしても必要です。そして、その手続きを可能にする方法はただ一つ、用例を採集することであります。

本節の最後に多義語についての議論を取り上げる[7]。多義語は「関連性が認められる複数の意味を有す語」と定義される。多義語の量的な傾向を分析するもの[大塚 1995a, 1995b, 2013a, 山崎 2011, 姜紅 2012]や, 多義の語義の区分を分析するもの[丹保 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 下村, 矢澤 2009, 山田 2012], 記述や配列を問題にするもの[国広 1982b, 1997, 1998, 2006, 森田 2000]などがある。多義語の配列は, 多くの小型の国語辞典は, 通常使用されているもの, 頻度の高いもの, 一般的なものから配列する「頻度型」であり, 中型・大型の国語辞典は, 時代を追って記述する「語史型」である[国広 1982b, 沖森 2006]。これらに対し, 国広[1998b]は,「これはひとえに使用者の検索の便宜を考えてのことである。しかしこの配列順は一方では、語義相互間の意味的関連性をまったく無視したものとなっている。多義は意味的な関連性を保ちながら派生してゆくのが普通である。」と述べ, 配列は意味的関連性に基づいて行うべきであると主張し, 国広[1982b, 1997, 2006]では意味的関係の辞書記述を試みている。森田[2000]も同様に,「辞書の語義記述として理想を言えば、多義語の場合、その語の意味領域全体を覆う基本義を捉えるとともに、史的に原義からどのような軌跡を辿って今日の多義化が進んでいったかを、基本義と絡めてその実相を示し、個々の意味枝が全体として有機的に捉え得るようにする。」と述べている。

しかしながら, 国語辞典で意味的関係の記述を取り入れるものはまだ見られない。だが, 日本語学習者向けには, 近年, 荒川洋平(編著)[2011]『日本語多義語学習辞典 名詞編』(アルク), 今井新悟(編著)[2011]『日本語多義語学習辞典 形容詞・副詞編』(アルク), 森山新(編著)[2012]

[7] 認知言語学による議論は 4.3 節で言及する。

『日本語多義語学習辞典 動詞編』(アルク)といった辞書が出ている。これらは多義の構造をネットワーク図で示すという記述方法が採られている。

2.1.2.5 用例

国語辞典の用例の重要性は早くから議論されている。まず，倉島[2008]は次の様に用例を大きく2つに整理して取り上げている。

1 語義分析のための用例
2 辞書に記載される用例
　①存在例
　②初出例
　③典型的使用例
　④記述を補強する用例

「1 語義分析のための用例」に関しての議論は見坊[1976a，1990]にくわしい。見坊[1976a]は「辞書の方法にとって今後重要なことは、今まで自明のこととしてほとんどかえりみられなかった同語別語の識別の問題、ならびに従来未開拓のまま打ち過ぎてきた意味の分析と記述の問題である。この2つを正しく発展させるためには、まず用例を採集し、用例に即して作業と考察を進めることが必要である。」とその重要性を述べている。

「2 辞書に記載される用例」に関しては，例えば林[1983]は次のように3種あると説明する。

第一の用例　比較的短い語の組み合せで，それが組合せとしてしばしば用いられる，標準的な，また場合によっては特殊な形式であることを示すものである。
　例：帽子をかぶる

第二の用例　その語を用いる具体的場面が示されるものである。世間の新聞雑誌等に現れる実例の適切なものを求めて，意味のとりやすい「用例」を「作例」する。

第三の用例　言語活動の個別実現の跡である「実際使用例」。その語の用法の条件ばかりでなく，その使用の時期を歴史的に示すことになる。

また，築島[1983]は，「過去の国語においては、確実な文献から正確且つ豊富な語例を集めることが、妥当な語釈語源説を記述するための必須の条件である。」と述べる。さらに森田[1989]は，次のように説明する。

> 用例提示の目的としては、大きく分けて、(1)意味理解の補助手段として、(2)語彙的知識の獲得のために、(3)用法上の特性を知るため、の三つがある。特徴的な工夫としては，例えば，『学研国語大辞典』は実例提示が多いことで知られ，『日国』は，「百万の用例」をうたい，豊富な用例，また，時代をさかのぼった用例提示を行っている。

このように，用例の収集・記述は重要であることの議論が重ねられてきた一方，従来の国語辞典に対しては用例の記述不足が指摘されている。鈴木[1983，1985]は国語辞典と和英辞典を比較し，「和英辞典は言語視野が広く、日本語の文例が極めて豊富である。国語辞典は足もとにも及ばない。」と述べ，例えば「考える」について国語辞典は数例のみ，和英辞典は小型でも 28 例(講談社和英辞典)，大型だと80 例(研究社新和英大辞典)の記載があることを指摘している。ゆえに，「国語辞典は的確な語義に適切な用例を配することで有用な国語辞典が誕生する。」と述べている。

また，少なさと同時に用例の表示方法にも問題があるとの指摘もある。井上[2009b]は「用例もその一つで，一般の国語辞典では，用例は次のようにフレーズの形で簡単に示されるのが普通です。」と言い，「正確」の用例が「―を期す」「時間に―な人」とだけ示されている例を挙げている。そして，それに対し，日本語学習者向けの辞書では，できるだけ多くの用例を，できるだけ実際の使われ方に即して示すことが重視されているとし，文化庁編『外国人のための基本語用例辞典』(第3 版，1990 年)や，姫野昌子監修『日本語表現活用辞典』(研究社，2004 年)を挙げている。現在，後者についてはその改訂版の姫野昌子監修『研究社　日本語コロケーション辞典』(研究社，2012 年)が出ている。大幅に増補改訂され，20 万を超える連結例，及び，4 万 5 千を越える例文が収録され，日本語教育の現場はもとより，広く文章作成時に参考になる辞典となっている。

倉島[1997b]の意見はやや異なり，国語辞典には出典・所在をより詳しく表示し長く引用する用例と，従来のようなより簡潔により短く工夫した作例の両方を記述するのがよいとしている。

日本語教育用に豊富な例文を文型とともに記載する辞書もある。グループジャマシイ編『教師と学習者のための日本語文型辞典』(くろしお出版，1998 年)である。「～てもいい」「それはそうと」といった連語を数多く収録し，接尾辞，文末表現，副詞と文末の呼応などについて詳しく記述して

ある。また,「国語辞典」ではわからない，日常の「話し言葉」や「慣用句」を理解して使いこなすためのものとして，山根智恵監修『研究社 日本語口語表現辞典』(研究社，2013 年)がある。用例が「会話例」であることが特徴的である。

なお，用例収集や記述の難しさについては，水谷[2012]が，「理想的な辞書を夢みると，すべての釈義に確かな実例が示してあってほしい。しかし，適切な用例を，しかも短いものを探すのは，なかなか大変である。」「良い辞書を作るのは，基本的に，どれだけ良い用例を持っているかの勝負。」と述べている。また，飯間[2013a，2013b]には街中の看板等からも用例収集に努める積極的な「ワードハンティング」の実際が報告されている。

2.1.2.6 用法

沖森[2006]が「近年は語釈に意味を記すのに加えて、その語の使い方などについても詳しく記すことが多くなってきた。」と述べているように，用法に関し，辞典によって様々な解説が工夫されている。文法説明，語源，語誌，位相，誤用，文型，ニュアンス，類義語の使い分け，数え方などの記載がある[大塚 2013b，2014]。その他，『新明解』や『表現』には重要語の区別がある。『三国』や『大辞林』にはアクセント表示がある。また，学習用辞典には，コラムや図解などもある。

これら用法の記述に関しては，表現や発信のための情報が不十分になりがちであるという問題点が多く指摘されている。例えば，宇野ほか[1971]の中で阪倉篤義は，「従来の辞書は、ほとんどが読むための辞書として、即ち理解用に、つくられている。しかし書くための辞書、即ち表現用の辞書も必要で、この面を多少でも兼ねる意味で、類義語や、語感の違いの説明は有用であろう。」と述べる。鈴木[1985]は，「従来の国語辞典は日本語の部品辞典である。どの部品と部品を組合わせて完成品(センテンス)を作るかという図式は，ほとんど示されていない。」「文章を書くのに役立つこと，という機能に著しく欠けている。」と指摘する。玉村[1995]は，「上下・親疎・利害・好悪・褒晴れ(けはれ)などを基準にした表現」がわかるような「位相・語感」情報の重要性と説いている。砂川[1998]は，「文体・位相にかかわる問題を辞典で記述すると、使用実態としてよりも使用の規範を示したものとして受けとられてしまうため、日本人向けの辞典では一般に非常に控えめな記述しかなされていない。」「しかし、使用の実態を知りたいという欲求がある。」と指摘する。佐竹[2000]は，「近年、国語辞書は、ことばの意味を調べるだけだけでなく、表現の可否をチェックするためにも使われる。そのようなときに問題になるのが用法に関する情報であり、それはもはや用例を示す形式だけでは不十分である。用法を知るには、その語についての文法情報だけでなく、それが話しことば的か書きことば的か、ことばとしてどの程度の丁寧さをもっているか、また、一般的な語か

専門的な語かといった、さまざまな情報が必要である。それらに関する情報をいかした形で用例に示すことが使いやすさにつながると考える。」と述べている。井上[2009b]もまた，表現の使用条件を過不足なく説明するのは難しく，一般の国語辞典にはあまり書かれていないと述べる。しかし，日本語学習者向けの辞書の中には各見出し語の使用条件を大まかに示すものがあるとし，その辞典を紹介している。それは，南昌妊ほか編『일본어 표현사전(日本語表現辞典)』(韓国：NEXUS JAPANESE)である。その辞典では，各見出し語について「文(文章)」「話(会話)」「男(男性)」「女(女性)」「F(フォーマル)」「in(インフォーマル)」という項目を立て，「○(使用)」「△(制限的使用)」「×(不使用)」のどれに当たるかを示してあるという。例えば，並列を表す「および」は，「文○ 話△ 男○ 女○ F○ in×」のように書かれているという。

以上述べた通り，長い間国語辞典には表現・発信のための記述が不足していると指摘され続けているが，いまだ解決にはいたっていない。一方，表現用辞書としては，国語辞典とは別に，シソーラス(Thesaurus)[8]や類語辞典が存在する [山崎 2009b]。シソーラスは，一定の意味分類の枠に従い，見出し語を分類・配列し，語彙の全体像を体系的に記述するものである。分類枠の中に見出し語が並ぶだけのもの，見出し語に語釈や例文が付いているものなどがある。類語辞典は，意味の似た語を集め，それらの違いがわかるように説明するものである。そこで山田[2006]は「一般辞書が目指すように，個々の語の意味・用法が十分に記述され，シソーラスが追求するように，他の語との関連性についての記述が含まれる。」ような，「意味から引く辞書」を提案している。この着想に近い国語辞典はまだ存在していないが，英英辞典には存在する。『ロングマンEssential Activator英英辞典』(ピアソンエデュケーション，2 訂版，2008 年)である[宮畑 2009b]。日常生活の 21 場面別に表現を知ることができるようになっている。また，表現の出発点となる物事や概念となる 750 語のキーワードが見出し語として並んでいる。例えば，「頭が痛くて体調が悪い」ことを伝えるには，PAINをひき，次にDOCTORの見出し語に飛ぶと，そこから痛みに関する様々な表現の選択肢が得られるようになっている。それら表現のための情報記述を国語辞典にどう取り入れるとよいかが課題である。新しい国語辞典への切り口となり得る着想と言える。

その他，日本の文化，日本特有の事物，日本事情，慣用表現に関する理解を助ける情報の必要性も指摘されている。玉村[1995]は，「雀の涙」は，「ごくわずかなもののたとえ」では不十分であり，「予算・給金など金銭について大変少量であることをたとえる。」という説明がいるとしている。潘[2006]は，「料亭」は，「政治要人などが出入りするところで，普通のサラリーマンぐらいの人は入れ

[8] イギリスのロジェが著した Thesaurus of English words and phrases(1852年) により広まったとされる。ロジェのシソーラスは版を重ね，現在多くのロジェの名を冠したシソーラスが刊行されている。

ない場所である。」といった例を挙げている。曺[2006]は,「立てひざ」は,「日本は行儀が悪いが、韓国ではチマ・チョゴリを着たときの立てひざは女性の正座。」といった例を挙げている。

また, 野田[2011]は, 日本語を母語としない人たちが「聞く」「話す」「読む」「書く」というコミュニケーション活動をするのに役立つ辞書を作るという観点から必要な辞書記述として, ①聞き間違えていても調べたい語句が調べられる, ②個々の状況に合ったあいづちなどの音声が聞ける, ③固有名詞を含め、あらゆる語句の意味が示される, ④依頼や感謝などのメールに使う文章例が示される, という点を挙げている。

国語辞典に十分にはない辞書記述を試みるものに, 品詞別の辞書がある。小泉保, 船城道雄, 田畠治, 仁田義雄, 塚本秀樹編『日本語基本動詞用法辞典』(大修館書店, 1989 年)や荻野孝野, 小林正博, 井佐原均編『日本語動詞の結合価』(三省堂, 2003 年)などの動詞辞書には, 例えば, 次のような情報が記載されている[丸山 2009a]。

1. どのような格成分と結びつくか。
   例: 「読む」〈人〉が〈文章など〉を読む
   「着く」〈人・乗り物など〉が〈場所〉に着く
2. テイル形, 可能形, 受身形, 使役形, 意志形などをとるか。
3. テイル形の意味は何か。
   例: 「食べている」は＜動きの最中＞,「倒れている」は＜結果の持続＞

飛田良文, 浅田秀子『現代形容詞用法辞典』(東京堂出版, 1991 年)や, 飛田良文, 浅田秀子『現代副詞用法辞典』(東京堂出版, 1994 年)には, ニュアンスの詳細な解説や, プラス・マイナスいずれの評価を表すか等が記述されている。オノマトペ(擬音語・擬態語)に関しては, 多くの辞典がある。天沼寧編『擬音語・擬態語辞典』(東京堂出版, 1974 年), 浅野鶴子編『擬音語・擬態語辞典』(角川書店, 1978 年), 阿刀田稔子, 星野和子『擬音語擬態語使い方辞典』(第 2 版;創拓社, 1995 年), 飛田良文, 浅田秀子『現代擬音語擬態語用法辞典』(東京堂出版, 2002 年), 山口仲美編『暮らしのことば 擬音・擬態語辞典』(講談社, 2003 年), 小野正弘編『日本語オノマトペ辞典 擬音語・擬態語 4500』(小学館, 2007 年)などである。

森田良行には, 図示を豊富に用いている『基礎日本語辞典』(角川書店, 1989 年)をはじめ,『日本語の類義表現辞典』(東京堂出版, 2006 年),『助詞・助動詞の辞典』(東京堂出版, 2007 年),

『動詞・形容詞・副詞の事典』(東京堂出版, 2008 年), 『日本語の慣用表現辞典』(東京堂出版, 2010 年)などがある。

2.2 計算機用辞書とその問題点

2.2.1 主な計算機用辞書

1980 年代後半より, 情報処理振興事業協会(IPA) 技術センターにて, 言語学者, 国語学者を中心とする研究グループが結成され, 次の辞書が構築された[情報処理振興事業協会 1997b]。

『計算機用日本語基本動詞辞書 IPAL(Basic Verbs)』(1987 年): 和語動詞 861 語

『計算機用日本語形容詞辞書 IPAL(Basic Adjectives)』(1990 年): 基本形容詞 136 語

『計算機用日本語基本名詞辞書 IPAL(Basic Nouns)』(1996 年): 文法的に特徴あるものを優先した基本名詞 1,081 語

これらはいずれも収録語数は少ないが, 収録された語については, 詳細な形態, 統語, 意味情報が広く網羅されている点に特徴がある。一連の『IPAL』は, 従来の国語辞典では省略されているような情報を漏らすことなく記載し, 言語処理の精度を上げようと試作された辞書である。

同時期に, 国家プロジェクトとして日本電子化辞書研究所(EDR: Japan Electronic Dictionary Resarch Institute) が設立され, 日本電子化辞書研究所『EDR』(1993 年)として, 日本語辞書 27 万語ほか, 英語辞書, 対訳辞書, 概念辞書, 共起辞書が構築された[日本電子化辞書研究所 1993]。意味概念は 40 万項目である。後に, EDR に基づき,「名詞・格助詞・動詞」を対象にした約 11,000 語のコロケーション辞典『日本語動詞の結合価』[荻野ほか編 2003]も構築された。

そして, EDR に続いて, NTT『日本語語彙大系』(1997 年)が構築された。これは, NTT の日英機械翻訳システムで用いられている計算機用辞書を再編集したものであり, 約 40 万語の日本語辞書が 3,000 種類の意味属性で分類されている。また, 14,000 件の文型パターンが収録されているものである[NTTコミュニケーション科学研究所監修 1997]。

汎用的な計算機辞書は, 以上の, 『IPAL』, 『EDR』, 『日本語語彙大系』が主要なものである。

2.2.2 計算機用辞書の問題点

上記の計算機用辞書には, それぞれに問題点が指摘される。

IPAL は，人手の分析により言語学的精密さを追求したものであるが，人手に頼ったがゆえに収録語数が少ないという問題がある。同程度の分析を行いながら，汎用と言える水準にまで語数を増やすのは非常に困難であると推察される。

EDR は，上記のように収録語数は多く，意味概念も極めて詳細に細分化されている。しかし，意味概念の認定と配列が構造的に整理されていないため，例えば，正確な意味処理や，文生成における適切な語選択が困難になりやすいという問題がある。

NTT 語彙大系は，収録語数や文型パターンが豊富な点が長所である。名詞の意味の階層構造は有用なシソーラス情報である。しかし，EDR 同様に，規模が大きすぎるため意味処理が困難であるという問題がある。また，表記の使い分けや使用域・位相などの情報は必ずしも十分に記述されていなかった。

問題解決の一つの試みとして，これら三つの辞書以降，開発されたものに，『Lexeed』[笠原ほか 2004]がある。笠原ほか[2004]は，意味処理を行うには，辞書が大規模であれば整合性を保つことが困難であるため，むしろ語彙を絞り，基本語に限定した「基本語彙知識ベース」を構築する必要性を主張し，意味データベース『Lexeed』を構築した。『Lexeed』は，『学研国語大辞典』を母集団とし，『日本語の語彙特性』[天野，近藤 1999]で単語親密度 5 以上である 2.8 万語を基本語として選定している。この基本語の新聞記事 14 年分のカバー率の調査結果は，延べ語数で約 72%，異なり語数で約 14%であったと言う。笠原らは，基本語のカバー率は異なり語数ではかなり低いが，延べ語数では 7 割もの高いカバー率があったことから，未知語を基本語で近似するなどの方法を検討すれば，日本語処理への基本語彙知識ベースの適用は有効と言えるだろう，と結論している。『Lexeed』の 2.8 万語というのは，国語辞典でいえば，「小・中・高校生向け：学習上必要な語を収録」という規模に相当する。『Lexeed』は心理学的な知見に基づいて設計されているため，例えばマニュアルのわかりやすさを判定するなど，言語の読みやすさや理解しやすさを定量的に評価する必要がある教育やユーザインタフェース設計などの分野での応用に適していると言われている[永田ほか 2008]。

2.3 本論文で取り上げる問題

本論文では，汎用的な言語資源となる「コーパスベース国語辞書」の構築を目指すことを目標としている。2.2 節で従来の計算機用辞書の問題解決の試みとして示した『Lexeed』は，国語辞典をベースに，心理学的調査結果を加えることにより新しい辞書を構築した事例であった。本論文

では，国語辞典をベースに，コーパス分析結果を加えることにより新しい辞書を構築する方法を提案する。

2.1 節で述べた従来の国語辞典の主な問題を整理すると，次の通りである。

見出し語の選定や語義の区分・配列が主観的であり，多義語の意味関係が不明瞭であり，必ずしも理解しやすいものになっていないこと，漢字・ひらがな・カタカナといった表記の判断にゆれがあり，標準的な表記が必ずしも明確でないこと，品詞の分類方法に諸説あるが，どの分類方法にも，品詞だけで語の挙動の振る舞いを説明しきれない場合があること，用例が限定されがちであること，使用域や位相の情報が不十分になりがちであること，などが問題点であった。

これらのうち，本論文で解決すべき重要な課題を次の 5 点に絞り込む。

① 見出し語の選定を客観的・明示的に。
② 語義の区分を客観的・明示的に。
③ 多義語の意味関係を客観的・明示的に。
④ 用例を網羅的に。
⑤ 使用域や位相の情報を具体的に。

第3章 コーパスを用いた辞書記述

本章では，本論文で提案する辞書記述方法を概説する。まず，3.1 節で，本論文で主な情報源として活用する『現代日本語書き言葉均衡コーパス(BCCWJ)』を中心としてコーパスの概要を述べる。次に，3.2 節で，BCCWJ 等のコーパスによって辞書構築のための用例収集がどのように充実し得るのかを述べる。さらに，3.3 節で，コーパスに基づく英語辞書構築の先行事例も踏まえ，本論文で提案する辞書記述の新規性と，その提案する辞書記述により期待される効果について説明する。

3.1 コーパスの概要

3.1.1 コーパスとは

コーパスとは，実際に使われた書き言葉や話し言葉を収集したデータベースのことである。主に言語研究目的に用いられ，電子化されているのが普通である。

コーパスという語は，狭義に用いる場合と広義に用いる場合とがある。もっとも狭義のコーパスは，言語の多様性を分析するための「均衡コーパス」(balanced corpus)である。言語の多様性を捉えるためには，コーパスが言語の使用状況を偏りなく代表している必要がある。これを言語使用の代表性(representativeness)と言う。代表性を備えるために，綿密な設計(コーパスデザイン)が行われ，各種ジャンルのテキストよりバランスよくサンプルを抽出する工夫が行われる。それらは「サンプルコーパス」とも呼ばれるが，そのうちバランスに留意して収集されたものが均衡コーパスである。また，狭義のコーパスでは，研究に必要な品詞情報などのさまざまな情報が付与されるのが普通である。このように，質，量に対し言語研究の観点から格別の注意を払い，明確な設計の基に構築されたものが狭義のコーパスである。

一方，代表性や付与情報に関する綿密な設計を行うことなく，テキストを単に収集したものもコーパスと呼ばれることがある。これを広義のコーパスと言う。例えば，新聞記事のデータベース，特

定の作家や作品のテキストを集めたもの，あるいは，Web 上の電子テキストを集めたものなどである。

その他，利用目的や，収集した言語の特徴に応じ，コーパスはさまざまに呼び分けられる。例えば，常に変化する言語を監視しながら，新しい情報を加えていくことを目的に，大量のテキストの集積を続けるものは「モニターコーパス」と呼ばれる。さらに，汎用コーパス，特殊目的コーパス，書き言葉コーパス，話し言葉コーパス，共時コーパス，通時コーパス，学習者コーパス，対訳コーパス，などと呼び分けられることがある[齊藤ほか編 2005]。

3.1.2 主要な英語コーパス

明確なコーパスデザインに基づいて構築された世界初のコーパスは，1964年にアメリカの Brown 大学で公開された Brown Corpus (1961-1964年)[9]である。このコーパスでは，1サンプル2,000語のテキストが30ジャンルから合計500サンプル集められ，収録語数100万語であった。本コーパスは代表性を備えており，続くコーパス構築のモデルとなった。Brown Corpus と同一デザインで構築されたコーパスには，LOB Corpus (1970-1978年，イギリス英語100万語)，FLOB Corpus (1991-1996年，イギリス英語100万語)[10]，FROWN Corpus (1992-1996年，アメリカ英語100万語)[11]があり，それら“Brown Family” の各コーパスは，ICAME[12]より入手可能になっている。

1995年には，Oxford University Press を中心とした6機関から成るBNC コンソーシアムにより，1980-1993年の間の1億語を収録したBNC (British National Corpus)[13]が公開された。これには，書き言葉だけではなく話し言葉も含められ，現代イギリス英語を代表するさまざまなテキストがバランスよく収録されている。また，同じく1995年に，Collins とBirmingham大学によってBoE (Bank of English)[14]が公開された。BoEは辞書の編纂を目的とするモニターコーパスであり，代表性の確保よりもむしろコーパスサイズを継続的に増やしていくことに重点が置かれている。公開当初は2億語であったが，現時点では4億5千万語を超えている。

1990年代よりBNCのアメリカ英語版として，ANC (American National Corpus)の構築が進められ，途中，約2,200万語が有料にて，約1,500万語が無料にて公開されたが，当初目的とされた1億語

9 http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/
10 http://www.helsinki.fi/varieng/CoRD/corpora/FLOB/index.html
11 http://www.helsinki.fi/varieng/CoRD/corpora/FROWN/
12 http://clu.uni.no/icame/
13 http://www.natcorp.ox.ac.uk/
14 http://www.titania.bham.ac.uk/

の構築は終えていないようである[15]。それに代わるものとして位置づけられているのが，COCA (Corpus of Contemporary American English)[16]である。COCAはアメリカのBrigham Young大学のMark Daviesが構築・運営するアメリカ英語のモニターコーパスである。1990年より毎年2000万語ずつ追加される計画になっており，2015年12月時点で5億3千万語を越えたデータ[17]が公開されている。

3.1.3 『現代日本語書き言葉均衡コーパス（BCCWJ）』構築に至る背景

アメリカでは 1964 年に Brown Corpus が公開されていたのに比べ，日本における本格的な均衡コーパスの構築は遅れていた[後藤 1995，丸山，田野村 2007]。その理由の 1 つとして，まだその頃の日本では，研究用データは自らの着眼点で収集してこそ独自の研究が可能になるという意識が強く，多くの人でデータを共有し，その共有データで研究を行うという発想が乏しかったのではないかと考えられる。また，コンピュータやコンピュータ上での言語処理の普及が欧米よりやや遅れていたことや，著作権処理が壁になったことの影響もあると考えられる。

しかし，日本でも 1990 年代に入り，まずは広義のコーパス，つまりある量のテキストを収集したものが言語研究者や言語処理研究者の間で言語資料として使われ始めた。入手が比較的しやすい新聞記事データベースの利用が広まっていったほか，主に言語研究者の間では，新潮文庫の 100 作品を収録した CD-ROM がよく使われ，言語処理研究者の間では，その頃公開された，ATR 音声・対訳テキストデータ，EDR 日本語コーパス，RWC テキストデータベース，IPA コーパス[橋本 1995]といったものが使われはじめた。

これらのうち，新聞記事データベースは，新聞社各社により，現在なお，毎年，新聞記事データが公開されている。その結果，現時点において，時系列調査が可能なコーパスとしても位置づけられているものである。

1997 年には，著作権の切れた著作を集めた『青空文庫』がインターネット上に開設され，今なお広く利用されている。インターネットの普及により，そういったさまざまな電子化テキストが作られ，公開されてきている。

さらに，2000 年以降には，大規模な量と，利便性の高さから，Web ページのテキストそのものが言語研究や言語処理研究に利用されるようになり，その傾向は現在ますます強くなっている[例え

[15] 2016 年 1 月現在，ANC へのサイトへのアクセスは不可能になっている。
[16] http://corpus.byu.edu/coca/
[17] http://corpus.byu.edu/coca/help/texts_e.asp

ば，田野村 2003，2011，関口，山本 2003，河原，黒橋 2006，荻野 2006]。だが，Web ページは母集団の確定の困難さや，質の確保，検索エンジンのくせ，コピーの重複などに課題があるとも指摘されている[田野村 2000，2011，田中 2003]。この他，国会会議録が Web 上で公開され，言語研究に利用されている[松田編 2008]。

このように，日本語研究において徐々にコーパスへの期待が高まる中，国立国語研究所は，2004 年に『日本語話し言葉コーパス (Corpus of Spontaneous Japanese，略称 CSJ) 』を，2005 年に『太陽コーパス』を公開した。そして 2011 年 8 月に BCCWJ を公開した。

3.1.4 『現代日本語書き言葉均衡コーパス（BCCWJ）』の概要

3.1.4.1 BCCWJ の設計方針と構成

BCCWJ の主な設計方針は次の 3 点であり，これらは従来の日本語コーパスには見られない顕著な特徴となっている[前川 2008，2011，山崎 2009a，前川編 2013，Maekawa et al. 2014]。

(1) 母集団を厳密に定義し，コーパスがその母集団の統計的な縮図になり，母集団に対し代表性をもつように設計する。
(2) 言語研究や教育，政策，辞書編集，自然言語処理などで幅広く活用できるよう汎用的に設計する。
(3) 広く公開可能になるよう，収録する著作物について利用許諾を得る。

BCCWJ は図 3.1 に示す通り，3 つのサブコーパスから構成される。1 つめの出版サブコーパスは，書き言葉が生産される実態を把握するためのものである。2001～2005 年に出版された全ての書籍，雑誌，新聞が母集団とされ，そこから抽出したサンプルから成る。

図 3.1 BCCWJ の構成

2 つめの図書館サブコーパスは，書き言葉が流通している実態を把握するためのものである。1986 年から 2005 年までの 20 年間に発行された書籍のうち，東京都内の 13 自治体以上の公共図書館で共通に所蔵されていた書籍が母集団とされ，そこから抽出したサンプルから成る。

そして，3 つめの特定目的サブコーパスは，日本語にとって重要でありながら上記 2 つのサブコーパスには含まれにくいデータ等が特定目的のために収録されるものである。「白書」「教科書」「広報紙」「法律」は公的な性格の強い書き言葉であり，言語政策に関わる基礎資料を提供することが期待できる。「ベストセラー」は，多くの一般読者に受容された書き言葉の実態を反映する資料として考えることができる。「Yahoo!知恵袋」「Yahoo!ブログ」は一般人が書いた書き言葉であり，Web ページ上の書き言葉がもつ変異のありさまを捉えることができる。短歌・俳句・詩を含む「韻文」は，現代日本語の書き言葉の重要な一部を構成するものとして採録された。国会における会議での発言を書き起こした「国会会議録」は，話し言葉を記録した書き言葉として採録された。

各コーパスは「サンプル」と呼ぶ文章単位で収録される。統計的な言語調査にも，幅広いテキスト研究にも利用できるよう，収録するサンプルの長さは 2 種類設計された。1 つは，1 サンプルの長さを 1,000 字とする「固定長サンプル」である。母集団からの抽出比が統計的な意味を持ち，語彙表や漢字表などの作成に適する。もう 1 つは 1 サンプルの長さを最長で 1 万字以内とし，章や節などの文章のまとまりを 1 サンプルとする「可変長サンプル」である。テキストの論理構造の把握やテキスト内での役割を持った要素の分析などに適する。

3.1.4.2 BCCWJ のサンプリングの設計と結果

出版サブコーパス・図書館サブコーパスのサンプリングの設計と，取得結果[丸山ほか 2011，山崎編 2014]について，NDC(日本十進分類法)別に，表 3.1，表 3.2 に示す。設計時の構成比は，母集団の推定文字数による NDC(日本十進分類法)別の割合を求めたものである。実際はさらに発行年ごとにも層別しているが，ここでは省略する。なお，可変長サンプル語数は推定である。いずれの表においても，構成比が設計時と最終結果とでほぼ同じであり，かつ，可変長サンプル語数の合計が，それぞれ目標の 3,500 万語と 3,000 万語を超えていることから，設計時の目的を十分に達成できたことが確認できる。

また，表 3.3 には，特定目的サブコーパスに収録されたメディアの種類，その対象期間，取得対象，取得したサンプル数，取得した可変長サンプルの推定語数を示す。

表 3.1 出版サブコーパスのサンプリングの設計と結果

	ジャンル	構成比 設計時	構成比 最終結果	取得した サンプル数	可変長サンプル語数
書籍	0.総記	2.50%	2.39%	363	833,197
	1.哲学	3.96%	4.02%	610	1,490,930
	2.歴史	6.57%	6.10%	926	2,447,545
	3.社会科学	18.95%	17.92%	2,721	7,194,570
	4.自然科学	7.74%	7.37%	1,119	2,646,734
	5.技術工学	7.05%	6.64%	1,008	2,447,023
	6.産業	3.35%	3.16%	480	1,232,742
	7.芸術	4.98%	4.79%	728	1,809,129
	8.言語	1.36%	1.30%	198	466,008
	9.文学	14.27%	16.84%	2,557	7,625,880
	n.記録なし	3.40%	3.31%	502	1,347,602
	小計	74.14%	78.34%	11,212	29,541,361
雑誌	1.総合	11.34%	11.76%	1,786	4,111,719
	2.教育	1.34%	1.27%	193	472,600
	3.政治	0.70%	0.75%	114	208,197
	4.産業	0.17%	0.16%	25	33,200
	5.工業	2.24%	2.13%	323	790,200
	6.厚生	0.28%	0.28%	42	71,569
	小計	16.06%	16.35%	2,483	5,687,485
新聞	全国紙	3.69%	3.62%	550	345,956
	ブロック紙	1.98%	2.01%	305	162,057
	地方紙	4.13%	4.18%	635	356,351
	小計	9.80%	9.81%	1,490	864,364
	合計	100.00%	100.00%	1,490	36,093,211

表 3.2 図書館サブコーパスのサンプリングの設計と結果

	ジャンル	構成比 設計時	構成比 最終結果	取得した サンプル数	可変長サンプル語数
書籍	0.総記	2.10%	2.21%	249	601,669
	1.哲学	4.90%	4.98%	560	1,466,585
	2.歴史	10.47%	10.08%	1,133	3,056,778
	3.社会科学	18.69%	19.52%	2,195	5,716,463
	4.自然科学	6.33%	5.90%	663	1,682,878
	5.技術工学	6.58%	6.14%	690	1,616,570
	6.産業	3.53%	3.38%	380	955,392
	7.芸術	8.47%	7.98%	897	2,167,036
	8.言語	2.00%	1.93%	217	427,326
	9.文学	32.34%	33.49%	3,765	11,212,003
	n.記録なし	4.61%	4.39%	493	1,150,711
	合計	100.00%	100.00%	11,242	30,053,412

表 3.3 特定目的サブコーパスの構成

メディア	対象期間	取得対象	取得したサンプル数	可変長サンプル語数
白書	1976年～2005年	1,006冊	1,500	約490万語
教科書	2005年～2007年	145冊	483	約120万語
広報紙	2008年	100自治体	355	約400万語
法律	1976年～2005年	718法律	346	約100万語
ベストセラー	1976年～2005年	951冊	1,696	約447万語
Yahoo! 知恵袋	2004年～2005年	約312万質問	91,450	約1,030万語
Yahoo! ブログ	2008年～2009年	約346万記事	52,680	約1,030万語
韻文	1980年～2005年	130冊	253	約20万語
国会会議録	1976年～2005年	32,925会議	159	約510万語

3.1.4.3 BCCWJ の主な付与情報

BCCWJ の収録テキストには，様々な検索が可能なように以下のような情報が付与されている[山口ほか 2011，山崎編 2014]。

(a) **書誌情報** 出典(書名，著者名，出版社，出版年，ジャンル)に関する情報

(b) **文字情報** ルビ，原文の誤植，JIS 外字，数式等の上付き・下付き文字などの情報

(c) **文書構造情報** 言語的な階層構造(記事，段落，文，見出し，引用，箇条書きなど)に関する情報

(d) **サンプリング情報** 固定長サンプルにおけるサンプル範囲の情報

また，BCCWJ は，短単位(意味を持つ最小の単位を結合させる，または結合させないことによって認定)と長単位(文節を規定に基づいて分割する，または分割しないことによって認定)という 2 種類の言語単位によって解析されている[小椋ほか 2011，小椋，冨士池 2011，小木曽，中村 2011]。解析された言語単位には検索の便のため語彙素(見出し語形)，書字形(表記)，品詞，語種などの形態論情報が付与されている。

3.1.4.4 BCCWJ の利用方法

BCCWJは，DVDの有償配布のほか，コンコーダンス機能をもつ全文検索サイト『少納言』[18]と，短単位検索サイト『中納言』[19]にて一般公開されている。『中納言』では，「短単位検索」「長単位検索」「文字列検索」ができる。

[18] http://www.kotonoha.gr.jp/shonagon/
[19] https://chunagon.ninjal.ac.jp/

また，レキシカルプロファイリングという手法を用いたコーパス検索ツールとして，『NINJAL-LWP for BCCWJ』(NLB)[20]が一般公開されている[パルデシ，赤瀬川 2011，2012]。『NINJAL-LWP for BCCWJ』(NLB)とは，国立国語研究所の共同研究プロジェクト「日本語学習者用基本動詞用法ハンドブックの作成」において，基本動詞ハンドブック執筆用に開発され，一般にも広く公開されているツールである。名詞や動詞などの内容語の共起関係や文法的振る舞いを網羅的に一覧表示できるのが最大の特長になっている。

ほかにも，BCCWJ の語彙・文法情報を他のコーパスと比較し参照できるツールが構築されている[千葉 2011]

3.1.4.5 BCCWJ の収録語の頻度

『現代日本語書き言葉均衡コーパス』短単位語彙表(ver.1.0)(以降，『BCCWJ短単位語彙表』)をはじめとするBCCWJの語彙表が一般に公開されている[21]。BCCWJ全体の頻度，BCCWJ全体での 100 万語当たりの頻度(pmw)，ほか，各レジスタ別の頻度が示されている。この『BCCWJ短単位語彙表』によると，BCCWJには異なりの合計で 185,136 語の短単位の語が収録され，頻度の最上位は頻度 5,061,558 の「の（格助詞）」であり，最下位 5,104 位には頻度 1 の語が 32,695 語あることがわかる。図 3.2 に，『BCCWJ短単位語彙表』収録語を頻度の降順に並べたグラフを示す。

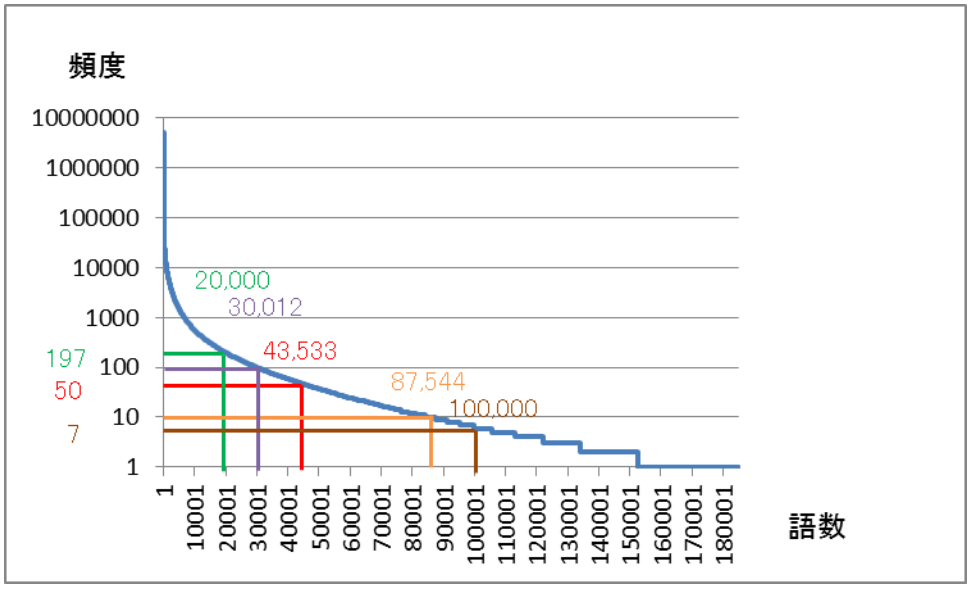


図 3.2『BCCWJ 短単位語彙表』収録語の頻度とその語数

[20] http://nlb.ninjal.ac.jp/

[21] http://pj.ninjal.ac.jp/corpus_center/bccwj/freq-list.html

図 3.2 に，次の情報がわかるよう書き加えている。また，以下では()に 100 万語当たりの頻度(pmw)を示す。

- 緑：上位 20,000 語は，頻度 197 以上の語である。(1.88)
- 紫：頻度 100 以上の語は，上位 30,012 語である。(0.96)
- 赤：頻度 50 以上の語は，上位 43,533 語である。(0.48)
- 橙：頻度 10 以上の語は，上位 87,544 語である。(0.10)
- 茶：上位 100,000 語は，頻度 7 以上である。(0.07)

先に，2.1.1 項にて，一般に国語辞典の収録語数は，小学生向け(2〜4 万語)，中学生・高校生向け(3〜6 万語)，小型(6〜9 万語)であることを述べた。また，国語教育のために選定された『新教育基本語彙』[阪本 1984]は約 2 万語であり，12 歳の理解語彙は 2 万 5 千語，20 歳では 4 万 8 千語と見積もられている[阪本 1955]ことを述べた。例えば，「上位 20,000 語は頻度 197 以上」ということは，BCCWJ で基本的な 2 万語レベルでの議論をする際に，少なくとも約 200 例以上の用例が得られるということになる。一般に理解されていると思われる語のレベルでの議論には約 50 例以上が，さらに小型国語辞典に収録されている語のレベルでの議論には，約 10 例以上が得られるコーパスである，ということである。10 例以上あれば意味・用法がある程度確認ができ，50 例以上あれば，多義語の意味分類が十分可能になるといえる。

さらに，10 未満の頻度が必ずしも辞書の議論に参考にならないというわけでもない。図 3.2 でわかる通り，頻度 10 未満の語も数多く収録されている。その中には，国語辞典に収録されているような語が BCCWJ では頻度が少ないという場合も含まれ，ただちに低頻度語すべてが辞書の対象外となるわけではない。また，実は，頻度 0 と 1 の境は大きく，1 例でも用例が収録されていれば，辞書記述に参考になる場合が少なくない。用例数 1 については，見坊[1990]が次のように述べている。

> 人は、気のつかないものについては考えることができない。用例数 1 の最大の価値は、そのことばの存在に気がつくようにさせてくれるところにある。"1"を手がかりにして、経験を重ねたり（追加採集をしたり）、類推や連想で考察を深めたり、また発展させたりすることができる。

以上のことより，BCCWJ の頻度情報は，辞書記述の十分な参考になり得るものであることが確認された。同時に，BCCWJ の頻度情報により，小型国語辞典に収録されている語に相当する頻度は 10，すなわち，0.10pmw がめやすになることが確認された。

3.1.4.6 BCCWJ の収録語と小型国語辞典の見出し語との関係

BCCWJ の収録語と小型国語辞典の見出し語との関係とを次の方法により調べる。

(1) BCCWJ には出版サブコーパス(PB)と図書館サブコーパス(LB)という同規模の書籍のサブコーパスがある(3.1.4.1 目参照)。そこで，出版サブコーパスから得られる収録語リストに対し，図書館サブコーパスから得られる収録語リストと，小型国語辞典の見出し語リストとのカバー率を比較する。

① 小型国語辞典『岩国第五版』[22]の見出し語 61,171 語の「国語辞典見出し語リスト」を作成する。

② 『長単位語彙表データ（頻度 2 以上）』[23]より，岩国の見出し数 61,171 にあわせ，BCCWJ全体の頻度 35 以上 63,338 語のリストを作成する。

③ そのリストの中より，BCCWJ の出版サブコーパス(PB)の頻度 1 以上の 61,634 語の「PB 収録語リスト」を作成する。また，図書館サブコーパス(LB)の頻度 1 以上の「LB 収録語リスト」を作成する。

④ 「PB収録語リスト」に対し，「LB収録語リスト」のカバー率と，「国語辞典見出し語リスト」のカバー率を出す [24]

(2) BCCWJ の全体(長単位異なり 2,434,619 語)の収録語における，小型国語辞典の見出し語リストのカバー率を出す。

出版サブコーパス(PB)のカバー率

次の表 3.4 の結果が得られる。

---

22 『GSK2010-A 岩波国語辞典第五版タグ付きコーパス 2004』を利用。
http://www.gsk.or.jp/catalog/gsk2010-a/

23 http://pj.ninjal.ac.jp/corpus_center/bccwj/freq-list.html

24 カタカナ表記列による一致調査であり，同音異義や単位認定の差異を考慮した同語異語判定はできていない。

表 3.4 「PB 収録語リスト」との一致数とカバー率

リスト	一致数	カバー率(%)
LB収録語リスト	60,119	97.5
国語辞典見出し語リスト	28,677	46.5

表 3.4 に示す通り，「LB収録語リスト」のカバー率が高い。LBもPBも，同じ長単位で解析され，単位がそろっている分一致率が高くなるのは当然なところもあるが，「LB収録語リスト」にあって「国語辞典見出し語リスト」にないものがどのような語であるかがわかる。頻度上位からみていくと，例えば，「ている」「という」「ではない」という複合辞と呼ばれるもの，「実施する」「使用する」などのサ変動詞，「お金」「お話」や「基本的」「積極的」「世界中」など接頭辞，接辞がついている語，「あんた」「ですから」などの話し言葉，「驚き」「美しさ」などの転成名詞，「一月」「父親」「社長」「出品者」「事務所」「外国人」「大匙」「電話番号」などの複合語などが岩国では見出しになっていないことが確認できる。すなわち，コーパスを活用することで，従来の国語辞典では漏れている語を見出し語として選定できることが確認できる。なお，形態素辞書としては，BCCWJから得られる語彙情報を反映させた辞書『UniDic』[25]がすでに構築され，広く公開され様々な言語処理に活用されている。

BCCWJ 全体における小型国語辞典のカバー率

結果を表 3.5 に示す。この結果は，BCCWJ 全体を使えば，岩国の見出し語 61,171 のうち，53,975 語(88.2%)の見出し語の用例を BCCWJ から得られることを表す。

表 3.5 「国語辞典見出し語リスト」の一致数とカバー率

語彙表	一致数	カバー率(%)
全体 2,434,619	53,975	88.2

BCCWJ で頻度 0 となる 7,196 語のうち 50 語(アイウエオ順の先頭より)を以下に示す。日常的に見かけにくい語であることがわかる。

愛吟，哀絶，愛想尽かし，相対ずく，間食い，間(の)宿・合(の)宿，合符，合札，曖昧屋，相持ち，相宿，相嫁，アウタルキー，アウフヘーベン，敢えず，青緡，青立ち，青電車，青道心，仰ける，青蠅・蒼蠅，青ばむ，青瓢箪，青膨れ・青脹れ，青やか，閼伽棚，赤切

25 https://osdn.jp/projects/unidic/

符，赤ゲット，赤行嚢，赤砂糖，赤襷，赤提灯，赤電車，赤茄子，茜雲，赤の飯，垢光(り)，赤間石，赤芽芋，明(か)り先，上がり高，上がり段，上がり端，上がり物，揚り屋，アカンベ，亜灌木，秋落ち，空(き)殻・明き殻，秋蚕

これら50語について，現在国立国語研究所にて構築中の200億語規模の『国語研日本語ウェブコーパス』(NWJC)で用例検索を試してみる。その結果，上記で網掛けした4語以外46語は用例が得られる[26](全体に対する推定カバー率 99.1%に相当)。

以上の調査結果より，小型国語辞典規模の辞書を構築する際には，9割近くはBCCWJで用例が得られること，また，コーパスを1億語規模だったものを200億語規模に増やすことにより，用例のカバー率が大幅に向上することがわかる

3.1.5 BCCWJ以降の主要な日本語コーパス

近年，欧米と同じく，日本でもWebデータを用いたコーパス構築のメリットが認識され，それに関する研究が増加し，Webコーパスが次々に構築されている。

はじめに，4.1億語のJpWaC [Srdanović et al.2008，スルダノヴィッチ，仁科2008]が構築され，続けて，100億語のJpTenTen11 [Baisa et al.2012]が構築された。JpTenTen11は，TenTen 群の1つとして2011年に構築された日本語のWebコーパスであり，Sketch Engine [Kilgarriff and Tugwell 2002，Kilgarriff et al.2004，Kilgarriff and Kosem 2012][27]に搭載されている[Srdanovic et al. 2008，スルダノヴィッチほか2013]。Sketch Engineは，コンコーダンス機能に加え，レキシカルプロファイリング手法を用いた共起・文法関係の検索・分析機能を持つ。検索語の複数のコロケーション・文法関係の情報を短時間にまとめた形で表示することができる。なお，Sketch Engine は，多言語に対応しており，辞書学関係の最新の国際会議EURALEX2014[28]や，eLex2013[29]には，Sketch Engine を活用する研究成果が多数発表されている。そのうちの一つ，Benko[2014]には，“Sketch Engine (SkE) has become a standard tool in numerous lexicographic projects, as well as in various areas of corpus-based linguistic research.”と記述されている。

[26] NWJC 公開前の2016/05/24 時点の検索結果による。この時点，インデックスの再作成中であり，検索漏れが生じている。よって，完成後には今回頻度0となった4語の用例も得られる可能性がある。
[27] https://the.sketchengine.co.uk
[28] http://euralex2014.eurac.edu/en/callforpapers/Pages/default.aspx
[29] http://eki.ee/elex2013/conf-proceedings/

また，11億語の筑波Webコーパス(TWC)が構築され[今井ほか2013]，コーパス検索ツール『NINJAL-LWP for Tsukuba Web Corpus』(NLT)[30]として，NLBとともに，一般に公開されている。

さらに，国立国語研究所においても200億語規模の『国語研日本語ウェブコーパス』(NWJC)[31]が構築された[浅原ほか2014a，浅原ほか2016]。

上述のWebコーパス以外には，学習者コーパス(例えば，「日本語学習者作文コーパス」[32])，国立国語研究所の「通時コーパス」[33]，Webから収集したテキストから得た格関係を整理したデータ(「京都大学格フレーム」[34])，単語Nグラムのデータ(「Web日本語Nグラム」[35])等が構築され，公開されている[丸山2013]。

3.1.6 コーパス言語学

英語研究において，コーパスの構築，普及とともに，コーパスを用いた言語研究が「コーパス言語学」と呼ばれている [バイバーほか2003，齊藤ほか編2005，マケナリーほか2014]。齊藤ほか編[2005]に，次のように説明されている。

> ‘corpus linguistics’ という術語は，Aarts & Meijs (eds.), *Corpus Linguistics* (1984) の出版以後広く一般に使われ始め，1990年前後に定着した歴史の浅い術語である [Leech 1992].

一方，日本語研究においてコーパス言語学的な研究は，宮島[2007]によれば，国立国語研究所が1948年の設立以降に行った，各種語彙調査において実践されていたと言う。しかし，「コーパス」という言葉が用いられはじめたのは，英語研究でその語が定着した1990年代以降のことであった[例えば，後藤1995]。その後，BCCWJ等のコーパスやWebの利用，また，各種コーパスツールの整備，公開[36]により，コーパスを用いた日本語研究は急速な進展を見せている[37]。コーパスによ

30 http://nlt.tsukuba.lagoinst.info/
31 http://pj.ninjal.ac.jp/corpus_center/nwjc/
32 http://sakubun.jpn.org/howto/
33 http://historicalcorpus.jp/
34 http://www.gsk.or.jp/catalog/gsk2008-b/
35 http://www.gsk.or.jp/catalog/gsk2007-c/
36 例えば，GREP 検索・テキストビューアの「KWIC Finder」(http://www.vector.co.jp/soft/win95/util/se184313.html)，分析ツールの「KH Coder」(http://khc.sourceforge.net/)，全文検索システムの「ひまわり」(http://www2.ninjal.ac.jp/lrc/)，タグ付きコーパスを管理・検索ツールの「茶器」(http://osdn.jp/projects/chaki/)，コロケーション抽出ツールの「茶漉」(http://tell.cla.purdue.edu/chakoshi/public.html)がある。

る日本語学研究の進展については，後藤[2007]，丸山，田野村[2007]，田野村[2009a]，田野村ほか[2010]，石川[2012a][38]，田野村編[2014]にくわしく記されている。コーパス言語学は，言語学，言語教育，自然言語処理など，幅広い研究分野に大きな影響を与えている[前川編2013]。

3.2　コーパスに基づく辞書研究の概要

3.2.1 コーパスベース英語辞書の研究

1964年にアメリカで Brown Corpus が公開されて以降，各種大規模コーパスが構築，公開され，コーパスを活用した英語辞書の構築が盛んに行われるようになった[石川2004，2012b，井上2005，投野2013]。欧米では古くから辞書学という学問が成立し，辞書編集についての議論が重ねられ，その流れの中でコーパスの辞書編集への議論が増えてきていた[ハートマン編1984，ランドウ1998，Béjoint2000，Landau2001，カウイー2003，ジャクソン2004等]。

最初のコーパスに基づく英語辞書は，『コウビルド英語辞典』(Collins COBUILD English Dictionary)の初版(1987年)である[Sinclair 1991]。そして，1990年以降，イギリスのCobuild，Longman，Oxford，Cambridge，Macmillan各社の学習者用辞書はすべてコーパスベースをうたっている。例えば，石川[2004]はLongmanの『ロングマン現代英英辞典』(第4版; 2003年:LDOCE4)を取り上げ，そのコーパスに基づく特徴として，①Spoken ／ Written の上位3000語を赤の見出しで記載，②語義の頻度順表記を記載，③共起例・類義語例を豊富に記載，④学習者コーパスの誤用分析に基づ語法注記を記載，をあげている。

Atkins and Rundell [2008]では，コーパスからの情報抽出から辞書編集までの工程が解説された。さらに，Rundell and Kilgarriff [2011]，Kilgarriff and Kosem [2012]，Rundell [2012]では，コーパスに基づく辞書編集におけるコンピュータの役目がくわしく解説されている[39]。具体的な最近の応用事例は，例えば，Atkins et al. [2010]のDANTEプロジェクトやKosem et al. [2014]のスロヴェニア語の辞書構築として示されている。これらには，コーパスに基づく辞書編集を支えるソフトとして，3.1.5項で言及したSketch Engineに加え，GDEX tool [Kilgarriff et al.2008, Kosem et al. 2011, Kosem et al. 2013]の有用性や，Cook et al. [2013]の新語や新語義の分析成果が述べら

---

37 http://pj.ninjal.ac.jp/corpus_center/bccwj/list.html に，BCCWJ を用いた研究業績の一覧が掲載されている。(635 件＜更新日時:2016 年 3 月 30 日＞)

38 石川[2012a]は，英語と日本語の両方を扱う。

39 他にも，Rundell[2012]が収録されている，Granger and Paquot eds.[2012]，及び，Fuertes-Olivera and Bergenholtzm eds. [2011] 等に，コンピュータを活用する辞書編集の議論が収録されている[Caruso 2013]。

れている。コーパスから用例を得るだけでなく，用例の意味分析を行った成果を辞書編集に取り込めるようになっていることを示している。Rundell [2012]では，今や，コンピュータの処理能力は上がり，コーパス構築にかかるコストは下がり，編集ソフトの活用で編集コストは下がり，改訂の速度は上がっていると述べられている。そして最後に，辞書編集者にとっての環境がほぼ整い，「We live in interesting times.」という言葉で結ばれている。実際に，ここ最近の英語の学習者用辞書は頻繁に改訂されている。そして，その形態は書籍のみならず，CD版，DVD版，オンライン版，電子書籍版など，多様化しながら進化を続けている。

ジャクソン[2004]は，コンピュータを活用したコーパスに基づく辞書編集は，従来の辞書編集と制作の方法を変えてきたが，辞書項目を記述する辞書編集者が不要になったわけではないと述べていた。辞書編集者の必要性はまだ変わらないであろうが，英語辞書の編集の現場では，ここ10年で半自動編集化が加速していると言えるようである[Cook 2014]。

文体情報に関しては，Biber [1988]，Nakamura [1995]，Biber, and Conrad [2009]の一連の研究がある。それらは，話し言葉と書き言葉，あるいは，小説，散文，新聞，といったレジスタの違いよって，文法項目や，語彙項目に差異がみられることを議論している。Rundell and Kilgarriff [2011]では，50億語のWebコーパスにフォーマリティのラベルをつけ，その分類の正確さを確かめながら辞書記述に取り込む試みが報告されている。

3.2.2 コーパスベース英和辞書の研究

近年では，日本の英和辞典にコーパスの活用が広がっている。コーパスに基づく辞書記述を最初に取り入れた英和辞典は『ウィズダム英和辞典』(三省堂，2002年)である。『ウィズダム英和辞典』の編集には数千万語の自社コーパスが用いられている。続けて，BNCと2千万語の自社コーパスが用いられた『ユース・プログレッシブ英和辞典』(小学館，2004年)，ケンブリッジ・インターナショナルコーパス(約6億語)が用いられた『ケンブリッジ英英和辞典』(小学館，2004年)が出ている。

make /meɪk/ [語源は「粘土を練って(家などを)作る」] ((名) maker, making)
動 (~s /-s/; made /meɪd/; made; making)
コーパス頻度ランク make＋名
① sense (→ 成句 make (...) SENSE)
② decision (↓ 他 7 第 1 例)
③ money (↓ 他 14 第 1 例)
④ difference (↓ 他 4 第 1 例)
⑤ mistake (↓ 他 7 第 2 例)
! make は「…を作る」が基本の意の 動 だが，目的語に 動 の名詞形を伴うと「〈特定の行為〉をする」ことを表し，この用法が頻度の上で多くを占める．また，比喩的に「〈抽象的な状態〉を生じさせる」,「〈金など〉を得る」などの意のほか，種々の慣用的表現でもしばしば用いられる．
他【作る】 1 a 〈人が〉«材料で»〈物〉を作る，制作[製

図 3.3『ウィズダム英和辞典』の「コーパス頻度ランク」

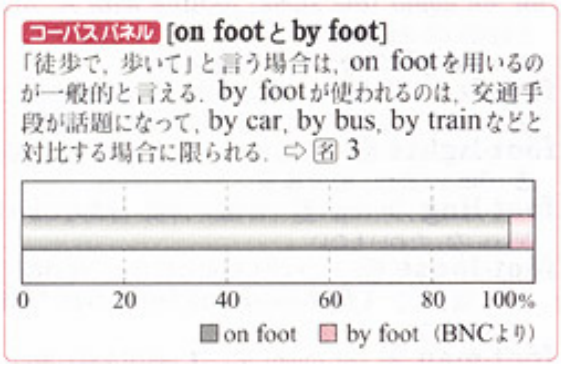


図 3.4『ユース・プログレッシブ英和辞典』の「コーパスパネル」

コーパス情報を明記している辞書記述例を示す。図3.3は，コーパスから抽出した特徴ある共起表現を記述した例である。また，図3.4はコーパス分析によって明らかにできた前置詞の使い分けの記述例である。『ルミナス英和辞典』(第2版；研究社，2005年)には「コーパス・キーワード」欄が新設され，動詞・句動詞と名詞、動詞と副詞などの結びつきをまとめて示されている。

その他にも，2007年に出た『ロングマン英和辞典』(桐原書店)は，Longman Corpus Network に加え，5千万語のロングマン現代日本語コーパスが用いられて編纂されていると言う。このロングマン現代日本語コーパスの一部には，構築当時に入手可能であったコーパスとして『日本語話し言葉コーパス(CSJ)』(国立国語研究所)が収録され利用されていると言う。

その後も，コーパスベースであることをうたう英和辞典は増えている。最新版の『ウィズダム英和辞典 第3版』(三省堂，2012年)のほか，『ジーニアス英和辞典』(第5版；大修館書店，2014年)，『プログレッシブ英和中辞典』(第5版；小学館，2012年)，『オーレックス英和辞典』(第2版；旺文社，2013年)など，いずれもコーパスベースをうたっている。

塚本[2004]は，コーパスベースの英和辞典に対し，「コーパスの発達は，頻度に基づく言語情報，話し言葉と書き言葉の違いなどを提示することを可能にし，学習辞典の編纂に画期的な成果をあげている。」と総括する。その一方，用例の記述に関しては意味と構造とをできるだけまとめて示すべきである，コーパスから得られる情報の中でさらに重要なものが明示されるべき，といった課題を指摘している。

3.2.3 コーパスベース国語辞書の研究

英語の学習者用辞書や和英・英和辞典に対し，日本の国語辞典の編集に，現時点，コーパスベースをうたうものは現れていない。BCCWJ構築時には特定領域研究「日本語コーパス」辞書編集班が組織され，BCCWJを基にした新しい形の国語辞書のサンプルを作るという目標が掲げられたが，作成には至らなかった。しかし，その時に，コーパスに基づく辞書編集の議論は重ねられた [40]。また，ほかにも日本語の辞書記述に，コーパス分析を活かそうとする議論は徐々に活発になってきている。

以下，はじめに辞書編集班での議論を取り上げ，続けて辞書の記述項目別にコーパス分析を行っている先行研究を概観する。

3.2.3.1 辞書編集班の成果

辞書編集班による主な報告は次の 4 つである。(1)コロケーション情報を抽出する基礎データとしては BCCWJ では量が不足している。そこで，Web のデータの資料性を確認した。ただし，Web では，①同文のコピーが繰り返され，頻度情報に偏りが生じている，②校正を経ない直接の書き込みによる誤字・脱字が頻繁にある，という問題があることを指摘した[荻野 2009，荻野編 2010，荻野ほか 2011]。(2)複合辞の研究では，BCCWJ を基にした複合辞資料集ができあがった。従来の辞書に載っている語だけではなく，日常的に使われながらも見落とされていた複合辞を多くあげることができた[近藤ほか 2011]。(3)国語辞典の意味記述，意味区分，動詞の自他用法の認定に対し，アンケート及び，BCCWJ を利用した実態調査を行い，記述の妥当性を検証した。漢語サ変動詞の自他については「自他両用漢語動詞辞典」を完成させた[下村，矢澤 2009，楊，矢澤 2010，荻野ほか 2011]。(4)従来の辞書や先行研究で問題となる動詞と名詞の格関係や，オノマトペの記述に関し，BCCWJ から得られる用例を使って辞書を改訂する方法論を検討した[丸山 2009，

[40] プロジェクトの成果報告書として，荻野綱男編『コーパスを利用した国語辞典編集法の研究』が，『特定領域研究「日本語コーパス」平成 18 年度研究成果報告書』[荻野編 2007]から『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』[荻野編 2010]まで計 4 冊刊行された。

2010a, 星野 2009, 2010]。基礎的な動詞の格情報の記述には BCCWJ の有効性を示せたが, オノマトペは BCCWJ だけでは用例収集が不十分になることを指摘した。

3.2.3.2 見出し語

佐藤[2012]は, 小型国語辞典の場合は, 見出し語に「日常的に使われる語彙」という集合が想定されいるとし,「素材情報を収集する情報源として、「日常的に使われる語彙を十分にカバーする」コーパスを用いる必要がある。『現代日本語書き言葉均衡コーパス』のような均衡コーパスは、この条件を満たす。」「コンピュータによる自動編集では、見出し語の取捨選択に人間の常識的判断を利用できないため、代表性を持つ見出し語集合の源泉を、使用する情報源の代表性に求めざるを得ない。」と述べ, BCCWJ のような代表性をもつコーパス利用の効果を述べている。

金[2006a, 2006b, 2011]は, 20 世紀後半の新聞記事を資料として「通時的新聞コーパス」を作成し,「抽象的な意味を表す外来語の基本語化」という現象の実態を明らかにした。基本語化した外来語は, 詳細に辞書記述すべき語であると言える。

近藤[2009], 田中[2010], 立山[2014]は, 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」や BCCWJ を用いて, 教科において特徴的に高頻度に出現する語彙を抽出している。これらもまた, 積極的に辞書に収録すべき語彙と言える。

3.2.3.3 表記

BCCWJ を用いた表記の実態分析がある。これは辞書記述に反映できる情報に成り得るものである。小椋[2012]は, 語種別には和語に, 品詞別では動詞に, 平仮名と漢字の対立によるゆれや, 異なる漢字の対立によるゆれが多くみられると述べている。また, 統語的複合動詞等の後項動詞の語表記のゆれや, 外来語の表記のゆれの実態を明らかにしている[小椋 2013, 2014a, 2014b, 2015]。

柏野, 奥村[2012]は, 和語や漢語などの非外来語の中にカタカナ表記が多用される語を明らかにし, その表記の使用実態を辞書記述に取り入れることを提案している。

3.2.3.4 品詞

楊, 矢澤真人[2010]は, 従来国語辞典でゆれの大きい漢語動詞の自他認定について, BCCWJ を用いて客観的に判定できることを示すものである。

石川[2011]は，BCCWJ を使い，従来の国語辞典では高頻度ながらも立項されない複合動詞に着目し，「出す」「出る」の複合動詞を例に，辞書情報の分析と記述例とを示した。例えば，複合動詞の頻度は「出す」の場合は本動詞の 2 倍ほど多いのに対し，「出る」は本動詞の 1 割程度にすぎないこと，補助動詞部分は一般には仮名書きにされがちだと思われているが，むしろ「～出す」「～出る」と漢字を当てられることが圧倒的に多いこと，補助動詞の結合は上位語の 5～10 語に偏りがあるということ(思い出す，生み出す，申し出る，流れ出る等)，複合動詞としての格パターンを抽出・記述ができることである。つまり，従来の国語辞典をコーパス分析によって充実させるには，その一つに複合動詞を取り上げるとよいという提案がなされている。

石塚[2013]は，「子育て」「値下がり」のような「名詞十和語動詞連用形」型複合名詞の品詞が「名詞」か「名詞・サ変」かに着目する。現行の 3 種の国語辞典の記述を調査した結果，その品詞認定は 8 割以上が不一致であったこと，かつ，一致している場合でも，コーパスで得られる用例の 9 割以上が「サ変」用法であるにも関わらず「サ変」の品詞認定がされていない，といった問題点を指摘し，辞書編集のコーパスの実態調査の有効性を論じている。

3.2.3.5 語義の区分・記述

加藤[1998]は，国定読本のデータを用いて「くらい」「ぐらい」を比較し，「くらい」は 4 つある語義にわたって用例が得られるが，「ぐらい」は特定の語義に偏りがあると指摘し，コーパス分析による意味記述の可能性を示した。そのために，均一な単位に分割されたコーパスの構築の必要性を訴えた。現在，均一な単位に分割されたコーパスは BCCWJ の実現によって達成されている。

星野[2009，2010]は，国語辞典では十分に意味記述されていなかったオノマトペを対象に，BCCWJ から意味記述に必要な情報を得て，辞書記述に反映させる方法を示している。

コーパスを用いることで客観的な区分を行うことを試みるものに，下村，矢澤[2009]がある。

また，類義表現の差異をコーパス分析によって明らかにする研究も多くある。これらも辞書記述につながる研究である。例えば，趙[2014]では，副詞の類義表現を対象に，BCCWJ の文学データと新聞データを比較し，文体差を明らかにしている。ほかにも，前田[2002]，金城[2011]，杉本[2012]，前坊[2009，2012，2014]など，類義表現の差異分析を進める論考は多く出ている。

3.2.3.6 用例

コーパスからの用例抽出，用例分析の議論の中心は「コロケーション」の研究であると言える。コロケーションの研究については多くの議論，分析がある。BCCWJ 構築以前と以後とに分けて述べる。

BCCWJ構築以前のコロケーション研究の概観や定義は，大曽，滝沢[2003]，荻野ほか[2005]，荻野，荻野[2007]，木村[2007]，野田[2007]，福島[2007]，村木[2007]，山田[2007]，兪[2009]，真田[2012]，堀編[2012]，スルダノヴィッチ[2013a]などにくわしく説明されている。基本的には，コロケーションとは，「慣習的に共に用いられる(2語以上)の語」[Firth1951]，「語と語が共に出現していること」[Sinclair1966]と定義されている[福島2007]。しかし，場合により広くも狭くも定義される。例えば，野田[2007]は助詞の結びつきまで含めて広いが，村木[2007]は助詞は含めずに狭い。また，山田[2007]は慣用句を含めて広いが，村木[2007]は慣用句は区別して狭い。

続いて，BCCWJ構築以後の研究について述べる。前川[2009]は，BCCWJのおよそ30年の幅がある点を用いたコロケーションの通時的研究を示す。例えば，「すべからく」の意味変化や，「自己」と共起する語の傾向と変化を示している。田野村[2009b]は，コーパスの有効な活用が見込まれる一つとして，コロケーションの抽出と記述に着目している。まず，言語形式の共起関係に関する情報は広範で雑多であり，そのすべてが有用であるわけではないことを指摘する。つまり，コーパスを使いさえすれば有用なコロケーション情報が容易に得られるということではなく，コーパスからコロケーション情報を抽出するための有効な手法を模索・案出することの重要性を述べている。その手法として，田野村[2009b]では，「共起語連鎖分析」が提案されている。中心語の前後の文脈を分け，各語の出現形に基づいて語連鎖を認定するという方法である。このとき，単一の語だけでなく，句の形も含める。これにより，言い回しを含めた共起表現の抽出が可能になっている。「制限」と「制約」，「希望」と「願望」のコロケーション辞書の試作例を示すことにより，コーパス分析に基づくコロケーション辞典作成の可能性を示している。

カルヴェッティ[2011]とStrafella，松本[2013]は，日伊辞典のための日本語のコロケーションの記述をBCCWJを用いて行うという議論である。カルヴェッティ[2011]は，BCCWJの貢献を次の通り挙げている。(1)従来の対訳辞典に収録されている用例に比べ，「現代の一般の日本人が日常使う時に必要とする日本語」というレベルに近づく。(2)一貫性のある形態素解析結果が利用できることは編集のさまざまな作業を楽にする。(3)用例から統語関係を認定することが容易である。(4)位相の明示的な情報はないが，出典の提示と広い文脈の観察でどのような場面で使用されているかの見当は付けられる。また，課題としては，均衡性がうたわれているが，語によってはジャンルに偏り

が感じられる場合があり，代表性について十分に検証してほしいことが指摘されている。Strafella，松本[2013]もまた，現行のイタリアの日本語辞典に不足する現代日本語の用例をBCCWJから得られるとし，コロケーションの抽出と記述とを行っている。

真田[2012]は，散文，『日本語動詞の結合価』[荻野ほか編2003]，BCCWJの3つの資料を用いて，6 つの動詞が伴う助詞の使用頻度を結合価の分布の観点から分析している。従来言われていた，動詞と助詞の結びつきのパターンを明らかにするだけではなく，量的な分布を明らかにすべきとの課題に対し，コーパスを用いることで解決を図る試みになっている。

スルダノヴィッチ[2013a]は，「形容詞+名詞」の組み合わせによるコロケーションに焦点を当て，BCCWJ，JpTenTen11，という2つのコーパスを用い，Sketch Engineを活用し，形容詞と名詞のコロケーション辞書の作成方法を提示した。英語辞書の構築にすでに広く使われているSketch Engine(3.1.5項)の日本語版を日本語辞書の構築へ導入することの利点が述べられている。

なお，コロケーションの自動抽出に関しての最新の研究には，尾本，北[1996]，園田ほか[2012, 2014]，高山ほか[2014]がある。高山ほか[2014]では，係り受け構造に基づき日本語コロケーションを抽出する手法を提案している。現段階では対象は論文であり，論文でよく使われている言い回しの抽出を行っている。

最後に，コロケーションの研究ではなく，コロケーションの資料集に言及する。BCCWJを直接活用してまとめたコロケーション資料集がすでに存在する。中俣[2014]である。BCCWJと7つの日本語教科書をもとに，初級向けの文法項目をまとめたものである。格助詞や助動詞などの文法項目と内容語のコロケーションを「文法コロケーション」と呼び，その「文法コロケーション」の生きた使用状況が包括的に，網羅的にまとめられている。今後の日本語教育やコーパス研究に資する資料となっている。

3.2.3.7 用法

後藤[2001]は，各種国語辞典に注記されている「文章語」「俗語」「雅語」「古語」などの文体レベルの記載に着目し，コーパス分析による客観的記述の可能性を検討している。各種国語辞典で注記のつく語を対象に，インターネット上で収集した，公文書，学術論文，エッセー，小説などの硬軟差のあるテキストの出現率を調査した。その結果，例えば，「問う」は「文章語」と注記する辞典としない辞典とがあるが，テキストの出現率は硬い文章に偏りのあることが確認できること，「訪れる」は多くの辞典が「文章語」あるいは「雅語的」と注記するが，テキストには硬軟偏りなく出現し

ていることなどがわかり，コーパス分析を用いることで，文体レベルの注記を客観的に行える可能性を示唆している。

BCCWJには，書籍，雑誌，新聞，Web文書といったレジスタの異なるサブコーパスが収録されているため，その差異を辞書に記述していくことが可能である。BCCWJに収録されるサンプルテキストの文体を計量的に考察する試みには間淵ほか[2010]や小磯ほか[2011]などがある。

動詞の格情報のコーパス分析には，丸山[2010，2011]や服部[2015]がある。丸山[2010，2011]は，BCCWJ を用いて，助詞の用法をその頻度別に分析し，従来の辞書の記述の見直しを行っている。服部[2015]は，主に 60 年間の国会会議録のデータを用いて，二字漢語動詞のニ格とヲ格が大きく意味役割を換えることなく交替する例のあるものについて，その格助詞の選択傾向の変化を探索している。一般傾向としてニ格からヲ格への推移が見られる一方，反対にヲ格からニ格への推移が見られる動詞が少なくとも 6 語あったことが報告されている。このことは辞書記述に有用な情報となる。

3.3 「コーパスベース国語辞書」の提案

3.3.1 コーパスに基づくことによる問題解決の見通し

3.2 節で概観した通り，コーパスに基づく辞書研究は，コーパスの整備とともに，ますます盛んになっている。 コーパスに基づく主な利点を整理して示すと，次の 3 点と言えよう。

(1)客観的・明示的になる：頻度情報や実例に基づき，辞書情報の抽出と記述とが客観的・明示的に行えるようになる。

(2)網羅的になる：生きた用例を可能な限り網羅的に収集できるようになる。

(3)具体的になる：資料の幅が広がり，語の多様性を具体的に捉えることができる。ジャンルあるいはレジスタなどと呼ばれる，言葉の使用域が異なる場合，それを具体的に辞書記述に反映できる(例えば，話し言葉的か，書き言葉的か，フォーマルか，インフォーマルか)。

以上の利点は，以下の，本論文で取り上げる問題(2.3 節)の解決を可能にするものと考える。

① 見出し語の選定を客観的・明示的に。

② 語義の区分を客観的・明示的に。

③ 多義語の意味関係を客観的・明示的に。

④ 用例を網羅的に。

⑤ 使用域や位相の情報を具体的に。

さらに，時系列のコーパスを活用することにより，従来まだ議論されていなかった，「語の使用頻度の経年変化」の辞書記述を可能にできると考える。そこで，解決すべき問題に次を加える。

⑥ 「語の使用頻度の経年変化」を具体的に。

①，②，③は，辞書記述の骨格となる見出し語選定と語義区分に関わる問題である。④，⑤，⑥は，辞書記述の内容にかかわる問題である。これらの問題に対し，本論文では次の通り，着目する対象と論点を設定する。

(1) 辞書記述の骨格となる見出し語選定と語義区分について

① 現代語で注意深く扱うべき「古風な語」に着目し，見出し語の選定を客観的・明示的に行う方法を示す。

② 名詞に着目し，名詞を下位区分する際の客観的・明示的な区分方法を示す。

③ 名詞の多義語に着目し，その意味関係を客観的・明示的な記述方法を示す。

(2) 辞書記述の内容について

④ 用例を網羅的に収集した場合に生じるコロケーションの多義性の問題に着目し，それを解決するための記述方法を示す。

⑤ 書籍テキストに着目し，語の使用域や位相の情報を具体的に行うために，書籍テキストへの文体アノテーションの方法，および，書籍テキストに出現する「古風な語」の使用域の記述方法を示す。

⑥ 経年変化が比較的捉えやすい外来語に着目し，外来語の使用頻度の経年変化に基づく辞書記述案を示す。

3.3.2 提案する辞書記述に対する要件

本論文では，3.3.1 項であげた問題設定 6 点を提案する辞書記述に対する要件とする。表 3.6 の各点に着目して，それぞれの要件を満たす辞書の構築を目指す。

表 3.6　提案する辞書記述が満たすべき要件

(1) 辞書記述の骨格となる見出し語選定と語義区分について
要件 1-1 見出し語の選定基準が客観的・明示的であること
要件 1-2 語義の区分が客観的・明示的であること
要件 1-3 多義語の意味関係の構造が客観的・明示的であること
(2) 辞書記述の内容について
要件 2-1 多義性をもつコロケーションの意味を区別するための情報が得られること
要件 2-2 語の使用域・位相の情報が得られること
要件 2-3 語の使用頻度の経年変化の情報が得られること

3.3.3 提案する辞書記述により期待される効果と新規性

要件 1-1 が満たされることにより，既に 2.1.3 項で述べたような，対象とする語の取捨選択が暗黙のうちに行われることによる問題を回避し，必要な語を漏らさず収録することが可能となる。また生成の際の適切な語選択を可能にする。

要件 1-2 が満たされることにより，多義性の解消が容易となる。例えば，ある名詞について，サ変動詞用法の有無で語義が区分されていることが明示されていれば，サ変動詞用法かどうかを判断材料として多義性を解消できる。また，用法の有無の正確な記述は，より適切な生成を可能にする。

要件 1-3 も，多義性の解消において重要な項目である。多くの多義語において，複数の意味の間には，ある種の関係が存在している。これを記述することで，そもそも多義がどのように生じているかの仕組みを明示化することができる。例えば，鍋という語には，容器の意味から中身の食べ物の意味への拡張が生じている。これが辞書に記述されていれば，具体物に結び付く動詞と可食物に結び付く動詞とを判別することで，容易に多義性を解消できると考えられる。

要件 2-1 は，多義性の解消において重要な項目である。多義性の解消の重要な手がかりの 1 つはほかの語とのコロケーションであるが，コロケーションを考慮してもなお多義性が解消できな

い例も多く存在する。これに対し，コロケーションとほかの語，例えば修飾語との結びつきまでを考慮することで，多義性を解消できると期待される。同時に詳細で豊富なコロケーション例は，適切なコロケーションの生成にも役立つ。

要件 2-2 は，言語生成において重要な項目であり，これが満たされることにより，文脈に相応しい語を精緻に選択することができる。解析時にはその情報を手がかりに使用域・位相の特定が可能になる。

要件 2-3 が満たされることにより，語の流行り廃りなどの情報が得られ，言語生成において，文脈や目的に応じた効果的な語を用いることが可能となる。

以上，述べたような辞書記述の検討例はこれまでに無く，それぞれが新規な提案である。提案の具体論を以降，第 4 章と第 5 章とで述べる。

第4章 コーパスに基づく見出し語選定と語義区分の精緻化

本章では，辞書記述の骨格となる，見出し語選定と語義区分に関わる事項について論じる。まず，4.1 節で，現代語と古語の間に位置する「古風な語」を例に，コーパスの頻度情報を使った見出し語の選定方法について提案する。次に，4.2 節において，現行の国語辞典における多義語記述の実態を踏まえて，「名詞」を対象に，コーパスから得られる用例分析に基づく名詞の区分方法を提案する。さらに，4.3 節において，名詞の多義語を対象に，コーパスから得られる用例から多義構造(語義間の意味的関係)を捉え，その多義構造を明示的に記述する方法を提案する。

4.1 「古風な語」に関する見出し語選定

本節では，「古さ」を帯びながら現代語として用いられている語に着目する。そのような語を「古風な語」と呼び，次の二点を満たすものと定義する。

(a) 「時代・歴史小説」を含めて現代で使用が見られる。

(b) 明治期以前，あるいは，戦前までの使用が見られる。

(a)は，現代ではほとんど使われなくなっている古語と区別するものである。(b)は「古風な語」の「古さ」の範囲を定めるものである。本節では，現代語と古語との境と一般にされている明治期以前までを 1 つの区切りにする。また，戦前と戦後とで文体変化が大きいと考えられるため，明治期から戦前までという区切りも設ける。しかしながら，一般には，戦前までさかのぼらずとも，事物の入れ替わりや，流行の入れ替わりにより，減っていったもの，なくなっていったものに「古さ」を感じることは多い。例えば，「ポケベル」「黒電話」「ワープロ」「こたつ」などである。こういった，近年急速に古さを感じるようになっている一連の語の分析も辞書記述の 1 つの課題と考えるが，本節で取り上げる「古風な語」は，戦前までさかのぼって「古さ」を捉えることとし，それ以外とは区別する。

「古風な語」に注目する理由は，三点ある。一点目は，現代語の中で用いられる「古風な語」は少なくないにも関わらず，「古語」にまぎれ辞書記述に取り上げ損なってしまう危険性のあるものであること。二点目は，その「古風な語」には，文語の活用形をもつなど，その文法的な扱いに注意の必要なものがあること。三点目は，「古風」という文体的特徴を的確かつ，効果的に用いることができるよう，十分な用法説明が必要な語であるということ。以上三点の理由より，「古風な語」は現代語の辞書構築の際に，見出し語として収録すべき語であると考える。

「古風な語」には，例えば，「さ【然】」がある。これは，「状態・様子がそうだという意を表す語。」(『岩波国語辞典第六版』岩波書店，2000 年，以下『岩国』)であり，現代では，「さほど」「さまで」「さばかり」「さしも」「さも」…のように結合して用いられる。その 1 つ，「さもありなん」(そうなるのがもっともだ)は，「さも」＋「あり」＋文語助動詞「ぬ」の未然形「な」＋文語助動詞「む」である「ん」，から成る連語である。枕草子(128 段)に，「大口また、長さよりは口ひろければさもありなむ」と使われている。一方，BCCWJ には，全体で 34 件の用例があり，いずれも，現代文脈での使用である。「まさか和久さんが指導員として復帰してるなんて思わなかったから。でも、**さもありなん**、という気もする。」(君塚良一(1950 年代生まれ)／丹後達臣，『踊る大捜査線スペシャル』扶桑社，1998 年)などである。同じように，「なきにしもあらず」「いわずもがな」「推して知るべし」…など，現代文脈で用いられる文語調の表現はほかにもあり，BCCWJ の現代文脈でそれぞれの用例を得ることができる。

「古風な語」は，これまでにも現代日本語における特徴的な語として着目されてきた。実際，多くの国語辞典では，現代文脈で使われる古さを帯びている語については，「古語」とはせず，「古語的」「古風」「雅語」「文語」「文語的」といった注記が付され，収録されている。しかし，それらの対象語を横断的に俯瞰することはこれまで行われていなかった。

以上の問題を解決するために，本節では，まずは電子化版が市販されている『CD-ROM 岩波日本語表現辞典―国語・漢字・類語―』(岩波書店，2002 年)収録の『岩国』に「古語的」「古風」と注記されている語を用い，現在刊行されている国語辞典で「古風な語」がどのように取り上げられているかを横断的に俯瞰する。次に，現代語のコーパスである BCCWJ に収録されている約 3,000 万語分の書籍テキストを用いて，その使用実態を分析し[柏野，奥村 2010，柏野，奥村 2011，柏野，奥村 2015]，それに基づき，見出し語の収録方法を提案する。

4.1.1 「古風な語」の現行辞典での扱い

4.1.1.1 現行の国語辞典での扱い

多くの国語辞典で,「古風な語」に関し,「古風」あるいは,「古語的」「雅語」「文語」「文語的」といった注記を付す記述が試みられている。積極的に「古風な語」の収録と記述が試みられているものは『三省堂国語辞典第六版』(三省堂, 2008 年, 以下『三国』)である。巻頭にある「この辞書の使い方」(p.15)には, 次の注意書きが記されている。

> 注意　古語ではないかと思われることばが、ずいぶん〔文〕や〔雅〕として出ていますが、これらは評論文〔新聞のコラムをふくむ〕・小説、新聞・雑誌の短歌欄や俳句欄などで実際に使われている現代語です。

そして,〔文〕とは文章語〔現代語のうち、文章などに使われる、話しことばとの差の大きいことば〕であり,〔雅〕とは雅語の略であり,「短歌・俳句・歌詞などで、現代でも使われるみやびやかなことばや詩的なおもむきのあることば」であると説明されている。さらに,「年配者の話しことば、小説のせりふ、落語・時代劇などに出てくる古めかしいことばは、〔古風〕として示しました。」とも記されている。つまり,『三国』では, 現代語として扱うべき「古風な語」を「文章語, 雅語, 古風」と分類して取り上げ, 記述しているということである。

一方,『新選国語辞典第八版』(小学館, 2001 年, 以下『新選』)のように,「古語」を現代文脈で使われる語だけに収録対象を絞っていないものもある。巻頭にある「この辞典を使う人のために」(p.6)には,「中学校・高等学校の国語科学習に必要な基本的な古語」は収録したと説明されている。

それら以外の国語辞典においては,「古風な語」について, 何らかの注記は付されているものの, その扱いについての明確な定義や説明はない。宮島[1977]が「「文章語」「文語」「雅語」など, それぞれの辞書で呼んでいるものの内容が, はたして一致しているのかどうかは, 凡例にかいていないのでわからない」と指摘したのをはじめ, 注記の用語や定義が明確でないことはしばしば指摘されてきた[遠藤 1988, 後藤 2000, 前坊 2009]。

4.1.1.2 現行の英語辞典での扱い

英語辞典では, 1970 年代頃より, さまざまな文体や使用域をどうしたら辞書で最もよく示し得るかという考察がはじまったと言われている[Hartmann 1983]。Cowie[2002]によると, 文体や使用域

に関する情報が「ラベル」(label)として英語辞典の記述に初めて導入されたのは『The Oxford Advanced Learner's Dictionary of Current English』(Oxford Univ Press, 3 版, 1974 年)であり, 体系的に定義されたラベルが導入されたのは『Longman Dictionary of Contemporary English』(Longman, 初版, 1978 年)であると言う。以降, 多くの英語辞典において, さまざまなラベルが活用されている。本論文で焦点をあてている「古風」に近いラベルは,「Time」に関わるものである。Hünig[2003][41]によると,「Time」に関わるラベルには,「dated」,「old-fashioned」,「old use」,「archaic」の 4 つがある。それらは,「比較的最近まで使用され, 現在高齢者に用いられている」「近現代英語では使われない」「過去の世紀に使用された」などを示すものとして用いられている。しかしながら, それら用語と定義, 付与される語は辞典間に差異のあることが指摘されている[Fedorova 2003, Hünig 2003, Ptaszynski 2010, Sakwa 2011]。

英語辞典の編集は早くからコーパスに基づいていると言われているが(3.2.1 項), コーパスに基づいてこれらのラベルを体系的に付与することを論じた文献は見当たらない[42]。

4.1.2 「古風な語」の現行の国語辞典における記述

4.1.2.1 対象の国語辞典

現行の国語辞典の「古風な語」に対する付与情報の記述の現状を把握するため, 次の 5 種の国語辞典を取り上げ,「古風な語」の見出し語としての採否, および, ラベル・注記の有無とその内容とを比較した。まず, 表 4.1 に, 対象とした辞典名と, 各辞典において「古風な語」に該当すると思われる語に付されていたラベル・注記情報の一覧を示す。表 4.1 を見ると, ラベルや注記情報は似てはいるがそれぞれに特徴のあることがわかる。例えば,『新選』にはほかで「雅語(的)」とされるものに相当するものは設けられていない。

表 4.1 調査対象の国語辞典と「古風な語」への注記一覧

国語辞典:略称	ラベル	語釈や注記
『岩波国語辞典』(第6版)2000年:岩国	なし	古語的・古風・雅語的・文語(的)
『学研現代新国語辞典』(第4版)2007年:学研	〔古〕(古語)・〔文〕(文章語)・〔雅〕(雅語)	古風・文語的
『明鏡国語辞典』(第2版)2010年:明鏡	〔古〕(古語)・〔文〕(文語)・〔雅〕(雅語)	古い・古風・文語的
『三省堂国語辞典』(第6版)2007年:三国	〔古風〕・〔文〕(文語)・〔雅〕(雅語)	文語的
『新選国語辞典』(第9版)2011年:新選	〔古語〕・〔文章語〕	古めかしい

[41] 調査対象は次の 4 つの辞書である。Collins Cobuild English Dictionary(1995,2nd), Longman Dictionary of Contemporary English(1995,3rd), Oxford Advanced Learners´ Dictionary of Current English(1995,5th), Cambridge International Dictionary of English(1995,1st)。

[42] Michael Rundell(2012)“Labels in Dictionaries” (http://trac.sketchengine.co.uk/wiki/AK/CourseNotes#)

4.1.2.2 調査対象語の選定

調査対象とした国語辞典のうち，唯一電子化版が市販されているのが『岩国』である。そこで，『CD-ROM 岩波 日本語表現辞典―国語・漢字・類語―』(岩波書店，2002 年)収録の『岩国』(第六版 [43])を利用して，付された注記を手掛かりに調査対象語を選定し，それらの語について，ほかの 4 辞書の記載を確認することとした。

先の表 4.1 に示した通り，『岩国』では，「古語的」「古風」「雅語的」「文語(的)」の 4 つの注記が付されている。本論文では，これらのうち，「古い」ということだけを表す指標と思われる「古語的」「古風」に着目した。「雅語」には風雅な趣があることや，和歌などに用いられる語という側面があり，「文語」「文章語」には主に文章に用いられる語という側面があるため，今回の調査には含めないこととした。

「古語的」や「古風」という注記は，例えば次の『岩国』の記述例に示すように，下位区分された特定の語義にだけ付されるもの(例：「いたい」)もあれば，語に付されるもの(例：「あいやく」「ころおい」)もある(引用中の太字表示は本論文著者による)[44]。また，「古風」の場合は，実際は，「古風」「既に古風」「やや古風」や，「古風な言い方」といったように注記の仕方に差異が見られる(例：「いでたち」「あいやく」「ころおい」「かまえて」)。本論文では，これらを区別せず「古風」とあるものをすべてひとくくりにして調査候補語とする。

『岩国』(第六版)より：

いた - い

一【痛い】神経に耐えがたいほど強い刺激を受けた感じだ。

①刃物で手を切る、虫歯がうずく等、外力・病気で肉体や精神が苦しい。(略)

②しまったと思うほど手ひどい打撃を受けたり、弱点を鋭く突かれたりして、つらい。(略)

二《「―・く」の形で》はなはだしく。ひどく。「自分の不明を―・く恥じる」「―・く感心した」▽一とは別語源か。**古語的**。

いでたち【出(で)立(ち)】

①(外出する時の)身なり。装い。「たいそうな―だ」

②旅立ち。しゅったつ。 ▽**古風**。

---

43 紙版は第七版新版(2011)が最新である。また，本調査時以降，『学研』は第 5 版(2012)，『三国』は第七版(2013)が最新である。

44 「いでたち」と「かまえて」の注記は，語につくのか②のみにつくのかがわからないが，これらの場合は，いずれの語も①も②も「古語」としてすでにその用法があったものなので，語につく注記と考える。

あいやく【相役】
同じ役(についている者)。▽**既に古風**。

ころおい【頃おい】
その折。「晩秋の—」。ころあい。「—を見て訪ねる」 ▽**やや古風**。

かまえて【構えて】
《副詞的に》①待ちうけて。用意して。心にかけて。
②決して。「—油断するな」 ▽**古風な言い方**。

具体的には,『岩国』には,「古語的」と注記される語は,「いたく【痛く】(～する)」「いとど」など 16 語あり,「古風」と注記される語は,「あいやく【相役】」,「あとげつ【後月】」など, 151 語あった。それらの語について最新版の第七版の記述を確認したところ,「古語的」と注記のあった「いやちこ」,「古風」と注記のあった「かえり【回り】」「じする【治する】」「みやばら【宮腹】」の合計 4 語は, 第七版未収録語となっていた。現代語辞書の見出し語として取り上げずともよいという判断のもと収録から外されたものと考えられる。そこで, 以上の語を除いた,「古語的」15 語(後掲 4.1.3 項の表 4.4 の「見出し」を参照),「古風」147 語(後掲 4.1.3 項の表 4.5 の「見出し」を参照)を本論文の調査対象語として選定した。

4.1.2.3 採否と注記付与状況の比較

調査対象語の辞典における見出し語としての採否, および, 調査対象語へのラベル・注記の有無とその内容とを比較した。

はじめに, 表 4.2 に『岩国』で「古語的」の注記のあるものより 3 例の, 表 4.3 に『岩国』で「古風」の注記のあるものより 5 例の調査結果を示す。表 4.2 と表 4.3 には, 5 種の辞典に見出し語として採用されている数, そのうち, 古さについてのラベルや注記が付与されている数, ラベルや注記の内容, 語釈を示した。調査対象語に対して, 辞典によっての見出し語としての採否や注記付与に差異が見られる。例えば, 表 4.2 の「まがまがしい【禍々・枉々・曲々・凶々しい】」(災いをもたらしそうだ。いまわしい。不吉だ。『岩国』)と表 4.3 の「よしなに【良しなに】」(よろしく。いいぐあいになるように。『岩国』)は, 5 種の辞典とも見出し語に採用としている点は一致しているが, 注記が付与されているのは 2 種の辞典のみである。

次に, 調査対象語すべてについての採用数と注記付与数の調査結果を後掲 4.1.3 項の表 4.4 と表 4.5 とに示す。表 4.4, 表 4.5 には, 半数以上の辞典が見出し語として採用し, かつ, 注記をつ

けていた語については○印を付している。それぞれの合計欄に示した通り，○印のついた語は，『岩国』で「古語的」と扱っている 15 語のうち 6 語であり，「古風」と扱っている語 148 語のうち 62 語であり，いずれも半分以下であった。辞典により編集方針が異なるため，採否や注記の付与に違いが生じるのは当然のことではある。しかしながら，本論文で構築を目指す「コーパスベース国語辞書」は，辞書編集者の主観的洞察の重要性を認識しつつも，見出し語の採否や注記の付与をはじめとする辞書記述全般において，コーパスを用いた使用実態の分析に基づく客観的指標を導入することを主眼とする。そこで，次節で BCCWJ を用いた「古風な語」の使用実態の分析を行う。

表 4.2 採否と注記付与状況の例(古語的)

見出し	表記	採用数	注記付与数	辞典	見出し無	ラベル	語釈前の注記	語釈	付記としての注記
めずらか	【珍か】	5	5	岩国				珍しいさま。	▽古語的。
				学研		〔文〕		めずらしいようす。	
				明鏡			〔やや古風な言い方で〕	めずらしいさま。	
				三国		〔雅〕		めずらしい感じだ。	
				新選		〔文章語〕		めずらしいようす。	
まがまがしい	【禍々・枉々・曲々・凶々しい】	5	2	岩国				災いをもたらしそうだ。いまわしい。不吉だ。	▽古語的。
				学研		―	―	悪い事が起こりそうでいまわしい。不吉だ。	―
				明鏡		―	―	悪いことが起こりそうに思われるさま。不吉だ。	―
				三国		〔文〕		わざわいをもたらすような感じだ。不吉だ。	
				新選		―	―	悪い事が起こりそうで、いやな感じである。不吉である。「―出来事」	―
やわか	【弥猛】	2	2	岩国				《あとに反語や否定的推量の表現を伴って》なんとしてでも。「―勝たでおくべき」。どうして。「―劣ることがあろう」	▽古語的。
				学研	無				
				明鏡	無				
				三国	無				
				新選		〔文章語〕		①どうして。「―負けるべき」②よもや。「―とどくまい」	

表 4.3 採否と注記付与状況の例(古風)

見出し	表記	採用数	注記付与数	辞典	見出し無	ラベル	語釈前の注記	語釈	付記としての注記
じっしょう	【実正】	5	5	岩国				まちがいないこと。いつわりのないこと。真実。「右ー なり」	▽古風な語。
				学研		〔文〕		確かなこと。まちがいのないこと。	
				明鏡		〔文〕		まちがいないこと。確かなこと。「ー明白」	
				三国		〔文〕		まちがいないこと。	
				新選				まちがいないこと。真実。	[参考]古めかしい言い方。
つくもがみ	【〈九十九〉髪・〈江浦草〉髪】	4	4	岩国				老女の白髪のこと。	▽「江浦草つくも」は植物のフトイの異名。白髪がフトイに似るからともいう。やや古風。
				学研		〔文〕		老女の白髪。	
				明鏡		〔古〕		年老いた女性の白髪。	
				三国	無				
				新選		〔古語〕		老女のしらが。	
よしなに	【良しなに】	5	2	岩国				よろしく。いいぐあいになるように。「―取り計らえ」	▽かなり古風。
				学研		〔古風な言い方〕		よいように。よろしく。	
				明鏡				よい具合になるように。よいように。よろしく。	
				三国				いいように。よろしく。「どうぞーお伝えください」	
				新選				いいように。よろしく。「ーとりはからえ」	
だいふ	【×乃父】	2	2	岩国				①父がわが子に対して自分を指す語。▽なんじ(=乃)の父の意。②一般に、父。	▽(1)(2)ともに既に古風。
				学研	無				
				明鏡	無				
				三国	無				
				新選		〔文章語〕		[「乃(なんじ)の父」の意]父が子に対して自分をいう語。おまえの父。	
たまざん	【玉算・△珠算】	5	1	岩国				そろばんを使ってする計算。しゅざん。	▽既に古風。
				学研		―	―	そろばんを使ってする計算。しゅざん。珠算(しゅざん)。	―
				明鏡		―	―	⇒しゅざん。	―
				三国		―	―	しゅざん。	―
				新選		―	―	しゅざん。	―

4.1.3 「古風な語」のコーパス分析

「古風な語」の現代書き言葉における使用実態を把握するために，BCCWJ[45]の使用頻度を調査した。全文検索システム『ひまわり』[46]を用いて，「BCCWJ 領域内公開データ 2009」の「図書館サブコーパス」[47](約 3,000 万語)における調査対象語の使用頻度を調査した[柏野，奥村 2010，柏野，奥村 2011，柏野，奥村 2015)。4.1.2.2 目では，「古語的」15 語，「古風」147 語を調査対象語としたが，ここでは，別語や別の意味用法の用例と紛れ，該当用例の判別が困難であった次の語は，検索の対象外とした。

●「古語的」15 語より対象外とした語

こうべ【首・頭】，つつみ【堤】，ぶにん【補任】

●「古風」147 語より対象外とした語(23 語)

いっそ，う，うつ【打つ】，くにびと【国人】，さと【里】，じきげ【直下】，しも【下】，じゃ(ぢや)，しょせい【書生】，ぜんぶ【全部】，そち【其方】，それ【其(れ)】，たいじん【大人】，たいぜい【大勢】，つ【唾】，つかさ【司・官】，であう【出合う・出会う】，とうじ【当時】，とも，の，むやく【無益】，やうち【家内】，やくだい【薬代】

つまり，使用頻度調査の対象語は，「古語的」12 語と，「古風」124 語である[48]。これらのBCCWJにおける使用頻度を求めた結果を，次の表 4.4，表 4.5 に示す。「使用有」の欄に「—」のある語は，上記で非調査対象とした語である。「○」のある語が使用頻度の得られたものであり，その数は「使用頻度」の欄に示した。使用頻度が得られたのは，表 4.4，表 4.5 の最終行に示す通り，「古語的」12 語のうちでは 7 語，「古風」124 語のうちでは 76 語であった。先の 3.1.4.5 目で述べた通り，一般の小型国語辞典に収録される語のめやすとなる使用頻度は 0.10(pmw)であり，約 3,000 万語規模のコーパスの場合は使用頻度 3 以上ということになる。3 を超えた語は 60 語であった。また，頻度

[45] BCCWJ を本調査で用いた主な理由は次の 2 点である。(1)現代日本語書き言葉の均衡がとられたコーパスであり，1 億語の規模があるため，現代語を対象とする国語辞典の記述でおさえるべき用例，用法を調べることができる。(2)著作権処理済であるため，国語辞典の例文作成時の参考にできる。

[46] http://www2.ninjal.ac.jp/lrc/

[47] 1986 年から 2005 年までの 20 年間に発行された書籍のうち，東京都内の 13 自治体以上の公共図書館で共通に所蔵されていた書籍が母集団とされ，そこから抽出したサンプルから成るサブコーパスである。

[48] 念のため第七版の未収録を理由に調査対象外とした 4 語についても使用頻度を調査したところ，「かえり【回り】」(回数・度数を表す，古風な助数詞。回かい。たび。)のみ，「蕎麦の三かえり」(藤村和夫，1930 年代生まれ，『蕎麦屋のしきたり』日本放送出版協会，2001 年)という用例があった。ほかは，使用頻度 0 であった。

が 0 だった語は 53 語であった。表 4.4, 表 4.5 を次に示す。いずれの表も以下の通り注記している。

※表中,「多義」の欄に「＊」を付してある語は,「古語的」あるいは「古風」の注記が多義のうちの 1 つの語義についていた語であることを示す。
※表中,「使用有」の欄に「—」のある語は, 使用頻度調査の対象外の語であることを示す。
※表中, 使用頻度が 3 を超えている語には, 使用頻度の欄に色をつけて強調している。

表 4.4「古語的」15 語の調査結果

	見出し	多義	表記	採用数	注記付与数	半数以上有	使用有	使用頻度
1	いたく	＊	【痛く】(△甚く)	5	5	○	○	66
2	めずらか		【珍か】	5	5	○	○	5
3	こうべ		【首・頭】	5	4	○	—	—
4	おおどか			5	2		○	1
5	まがまがしい		【禍々・曲々・枉々・凶々しい】	5	2		○	33
6	ついな		【追儺】	5	1			0
7	いとど			4	4	○	○	3
8	つ		【津】	4	4	○		0
9	やたけ		【弥猛】	4	4	○		0
10	しきしまのみち		【敷島の道】	4	2			0
11	やわか		【弥猛】	2	2		○	3
12	つつみ		【堤】	2	1		—	—
13	ぶにん		【補任】	2	1		—	—
14	おおみこと		【大御言】	1	1			0
15	おおみたから		【大御宝】	1	1		○	1
	「5辞典×15見出し」の合計			54/75	39/75	6/15	7/12	

表 4.5「古風」147 語の調査結果

	見出し	多義	表記	採用数	注記付与数	半数以上有	使用有	使用頻度
1	うせる	＊	【失せる】	5	5	○	○	214
2	じっしょう		【実正】	5	5	○		0
3	そち		【〈其方〉】	5	5	○	—	—
4	そなた		【〈其方〉】	5	5	○	○	434
5	つかさ		【△司・△官】	5	5	○	—	
6	とのご		【殿御】	5	5	○	○	7
7	ないふく		【内福】	5	5	○	○	3
8	にょしょう		【女性】	5	5	○		0
9	まかりならぬ		【△罷り成らぬ】	5	5	○	○	11
10	いかさま	＊	【〈如何〉様】	5	4	○	○	8
11	いな		【異な】	5	4	○	○	7
12	からめる	＊	【×搦める・絡める】	5	4	○		0
13	くちおしい		【口惜しい】	5	4	○	○	36
14	けそう		【懸想】	5	4	○	○	8
15	ごじん		【御仁】	5	4	○	○	41
16	こなた		【〈此方〉】	5	4	○	○	5
17	さ	＊	(△然)	5	4		○	38
18	じゃ	＊	ぢや	5	4	○	—	—
19	にんじょう		【△刃傷】	5	4	○	○	40
20	はたまた		【△将又】	5	4	○	○	65
21	まいる	＊	【参る】	5	4	○		0
22	ものども		【者共】	5	4	○	○	104
23	やくぎ		【役儀】	5	4	○	○	8
24	ゆうけい	＊	【夕景】	5	4	○	○	15
25	ゆめさら		【夢更】	5	4	○	○	1

表 4.5「古風」147 語の調査結果(続き)

	見出し		表記	採用数	注記付与数	半数以上有	使用有	使用頻度
26	よしない		【由無い】	5	4	○	○	2
27	ろうせい		【老生】	5	4	○		0
28	わい			5	4	○	○	159
29	う	＊		5	3	○	—	—
30	かくし		【隠し】	5	3	○		0
31	けっく	＊	【結句】	5	3	○	○	3
32	ころおい		【×頃おい】	5	3	○	○	2
33	せっかく	＊	【折角】	5	3	○		0
34	たいじん	＊	【大人】	5	3	○	—	—
35	とも	＊		5	3	○	—	—
36	なにさま	＊	【何様】	5	3	○		0
37	ならぬ	＊		5	3	○	○	5
38	にょにん		【女人】	5	3	○	○	171
39	ふうぎ		【風儀】	5	3	○	○	8
40	ほうばい		【×朋輩・△傍輩】	5	3	○	○	33
41	ほんに		【本に】	5	3	○	○	41
42	むやく		【無益】	5	3	○	—	—
43	いでたち		【出(で)立(ち)】	5	2		○	77
44	きこえる	＊	【聞(こ)える】	5	2		○	1
45	こくぼ		【国母】	5	2		○	3
46	しも	＊	【下】	5	2		—	—
47	だんこん		【男根】	5	2		○	41
48	どうやく		【同役】	5	2		○	13
49	なと			5	2		○	1
50	なんだ			5	2		○	21
51	ひきあけ		【引(き)明け】	5	2			0
52	みんりょく		【民力】	5	2		○	19
53	やくだい		【薬代】	5	2		—	—
54	よしなに		【良しなに】	5	2		○	7
55	さしまねく		【差(し)招く・×麾く】	5	1		○	3
56	さと	＊	【里】	5	1		—	—
57	しまつ	＊	【始末】	5	1			0
58	たまざん		【玉算・△珠算】	5	1			0
59	つけぶみ		【付け文】	5	1		○	5
60	ばうて		【場打て】	5	1			0
61	むさい			5	1		○	3
62	いと			4	4	○	○	37
63	おもんみる		【×惟る】	4	4	○		0
64	かまえて		【構えて】	4	4	○	○	1
65	だいじない		【大事無い】	4	4	○	○	6
66	つくもがみ		【〈九十九〉髪・〈江浦草〉髪】	4	4	○	○	3
67	かしこくも		【×畏くも】	4	3	○	○	5
68	かんばつ		【簡抜】	4	3	○		0
69	くにびと		【国人】	4	3	○	—	—
70	けいさい		【継妻】	4	3	○		0
71	すすどい			4	3	○		0
72	てんうん	＊	【天運】	4	3	○		0
73	あんずるに		【案ずるに・×按ずるに】	4	2		○	11
74	せわ	＊	【世話】	4	2		○	14
75	せんど		【先度】	4	2			0
76	いかん			4	1		○	446
77	いっこう	＊	【一向】	4	1		○	1
78	げせる		【解せる】	4	1		○	37
79	けんご	＊	【堅固】	4	1			0
80	しろっぽい	＊	【白っぽい】	4	1			0
81	ずんと			4	1		○	4
82	なにがさて		【何がさて】	4	1			0
83	うちつけ			3	3	○	○	1
84	こうじき		【高直】	3	3	○	○	1
85	ごのう		【御悩】	3	3	○		0
86	しゅんじょう		【春情】	3	3	○	○	3
87	て	＊		3	3	○	○	18

表 4.5「古風」147 語の調査結果(続き)

	見出し		表記	採用数	注記付与数	半数以上有	使用有	使用頻度
88	とうじ	＊	【当時】	3	3	○	—	—
89	の	＊	(のう)	3	3	○	—	—
90	よち		【×輿地】	3	3	○	○	8
91	よのぎ		【余の儀】	3	3	○	○	1
92	ろうだい		【老台】	3	3	○		0
93	あつかい	＊	【扱い】	3	2			0
94	あつかう	＊	【扱う】	3	2			0
95	おなじくは		【同じくは】	3	2		○	2
96	ぐんばい		【軍配】	3	2		○	1
97	ぜんぶ	＊	【全部】	3	2		—	—
98	たいぜい		【大勢】	3	2		—	—
99	ましい		【△継しい】	3	2			0
100	いずれ	＊	何れ・孰れ	3	1		○	5
101	かしこきあたり		【×畏き辺り】	3	1			0
102	しんずる		【進ずる】	3	1		○	15
103	であう	＊	【出合う・出会う】	3	1		—	—
104	ひきずり	＊	【引(き)摺り】	3	1			0
105	ぎじょう	＊	【議定】	2	2		○	7
106	これ	＊	此(れ)	2	2		○	2
107	さり	＊		2	2		○	8
108	しょげん		【諸×彦】	2	2		○	1
109	それ	＊	【其(れ)】	2	2		—	—
110	だいふ		【×乃父】	2	2			0
111	なにがな		【何がな】	2	2			0
112	のう	＊	(なう)	2	2		○	11
113	やうち		【家内】	2	2		—	—
114	あいやく		【相役】	2	1		○	1
115	あとげつ		【後月】	2	1			0
116	うつ	＊	打つ	2	1		—	—
117	おおみよ		【大△御代】	2	1			0
118	しゅっとう	＊	【出頭】	2	1		○	1
119	つけびと	＊	【付け人】	2	1		○	2
120	ておもい		【手重い】	2	1			0
121	どうぞ	＊		2	1		○	2
122	どうりゅう	＊	【同流】	2	1			0
123	とと		【△父】	2	1			0
124	ひとしい	＊	【等しい・均しい】	2	1		○	1
125	ひろいあるき	＊	【拾い歩き】	2	1			0
126	ひろう	＊	【拾う】	2	1			0
127	ぶじ	＊	【無事】	2	1			0
128	いかい			1	1		○	4
129	いきせき			1	1			0
130	いっそ	＊		1	1		—	—
131	かいしき			1	1			0
132	がな	＊		1	1			0
133	けっこう	＊	【結構】	1	1			0
134	ござなく		【御座無く】	1	1		○	3
135	こわつき		【声つき】	1	1			0
136	ざ			1	1		○	3
137	じきげ		【直下】	1	1		—	—
138	しだら			1	1		○	1
139	しょせい	＊	【書生】	1	1		—	—
140	しんがく	＊	【進学】	1	1			0
141	つ		【×唾】	1	1		—	—
142	つむ	＊	【摘む】	1	1		○	2
143	なぐさむ	＊	【慰む】	1	1			0
144	にえぎも		【煮え肝】	1	1			0
145	ははじゃびと		【母者人】	1	1			0
146	ほど	＊	【程】	1	1		○	4
147	よがら		【世柄】	1	1			0
	「5辞典×147見出し」の合計			521/735	346/735	62/147	76/124	

4.1.4 考察

前項の表 4.4, 表 4.5 にて,『岩国』から抽出した「古風な語」の使用頻度を示した。一般の小型国語辞典の収録のめやすとなる 0.10(pmw)(使用頻度 3)以上あった語は 44.1%と, 全体の約半分にも満たない数であったが, それらの多くは見出し語として採録されていることが確認できるが, 中には使用頻度があっても採録する辞書が少ない語もある。採録外としている辞書も積極的に採録すべき語であるといえる。また, 頻度が低い語(3 未満)については, 低いにも関わらず見出し語として採録する辞書もあれば, それゆえか見出し語としてあまり採録しない辞書もあり, 態度が大きく分かれていることがわかる。低頻度でも用例の得られる語は現代語として目にする機会のある語であると考え, 見出し語としての採録を検討すべき語であると考える。

一方, 今回の調査で頻度 0 であった語を, ただちに見出し語として採録しないと結論づけるのは難しい。現代語の辞典で取り扱う必要性の低い「古語」であるのか, あるいは, たまたま例がとれなかった「現代語」であるのかがわからないからである。BCCWJ において頻度 0 であることの意味は, 今後, 調査範囲とするコーパスを増やしたさらなる検討が必要であるだろう。

4.1.5 見出し語選定の手順

以上の考察をふまえ, 本項では次の通り, コーパス分析を利用する「古風な語」の見出し語選定の手順を提案する。

1. 複数の国語辞典より, 付与されている位相情報「古風, 古語的等」を手掛かりに, 見出し語候補を選定する。
2. 複数の国語辞典どうしでの採否状況を確認する。
3. 見出し語の頻度情報を現代語のコーパスである BCCWJ より得る。
   (1) 0.10(pmw)以上の頻度あり:見出し語として採録。
   (2) 0.10(pmw)未満の低頻度:採録していない国語辞典もある場合, 採否のボーダーとなるが, BCCWJ で低頻度でも用例の得られる語は現代語として目にする機会のある語であると考え, 採録。
   (3) 頻度 0:採否は別途検討。現代語の辞典で取り扱う必要性の低い「古語」であるのか, あるいは, たまたま例がとれなかった「現代語」であるのか別資料にあたる。

※頻度 0.10(pmw)は，BCCWJ において小型国語辞典収録語数のめやすとなる，100 万語当たりの頻度(3.1.4.5 目)。BCCWJ 図書館サブコーパスのみ使用の場合は 3 以上，BCCWJ 全体であれば 10 以上が相当。

本節では「古風な語」について検討したが，同様の方法で，「古風な語」以外にも，コーパスの頻度情報を用いた見出し語の選定が可能であると考える。

4.1.6 コーパスからの自動抽出の試行

「**要件 1-1** 見出し語の選定基準が客観的・明示的であること」により，漏れのない見出し語選定を行える方法として，BCCWJ のコーパスの頻度情報を用いた見出し語選定を提案した。本項では，提案手法以外の方法としてコーパスからの「古風な語」の自動抽出を試行する。

4.1.6.1 抽出方法

BCCWJ「図書館サブコーパス」の形態素解析結果より，「活用型：文語」となっている語を抽出し，リストを作成する。BCCWJ は短単位(意味を持つ最小の単位を結合させる，または結合させないことによって認定)と，長単位(文節を規定に基づいて分割する，または分割しないことによって認定)とで解析されている[小椋 2014，小椋ほか 2011，小椋，冨士池 2011]。そこで，短単位のリストと，長単位のリストの両方を作成する。

4.1.6.2 抽出結果

短単位の抽出リストの使用頻度上位は「べし：13,889，なり：5,421，たり：3,357 ，ごとし：2,048，り：1,888」であった。高頻度の助詞・助動詞の用例が多数得られることがまずは確認できた。また，「有り：618，然り：525，来たる：284，恐る：262」など，文語の活用形で用いられる語の抽出ができることも確認できた。

しかしながら，長単位リストとあわせてみても，文語形を含む見出し語そのものの形での抽出は自動では簡単ではないことが確認された。例えば，「止む無し」は 1 つの長単位として抽出できているが，そういう語は少ない。多くは，「なきにしもあらず」が「なき＋に＋しも＋あら＋ず」となっているように，長単位も短単位同様に複数に切れている。また，「さもありなん」のように誤解析されている場合もある(「なん」が「名詞【何】」と誤解析。正しくは文語助動詞「な＋む」)。いずれの場合も，

自動抽出のためには形態素解析の辞書整備が必要であるが，そのためには，結局，既存の国語辞典等にある見出し形をあらかじめ先に参照しなければならない，ということになる。

しかしながら，自動抽出した結果のリストの資料価値がないわけではない。例えば，抽出したリストの上位に「めでたし」がある。短単位の頻度 66(58 位)，長単位の頻度 73(61 位)のものである。この語は『岩国』の見出し語ではない。しかし，『広辞苑』や『大辞林』等では，「めでたし」や「めでたしめでたし」が見出し語として立項されている。BCCWJ の用例のほとんどは「めでたしめでたし」であり，単独の「めでたし」の用例もある。このような，BCCWJ において文語形で高頻度である語は，『岩国』のような小型国語辞典レベルでも収録すべき語と考える。このように，コーパスから得られる自動抽出リストは，従来の国語辞典での見出し語選定漏れを見つける参考になり得るものである。

4.1.6.3 抽出結果のまとめ

本項では，既存の国語辞典等によらず，コーパスから自動的に「古風な語」の抽出が可能であるかを検討した。その方法として，BCCWJ「図書館サブコーパス」の形態素解析結果より，「活用型：文語」となっている語を抽出し，リストを作成した。その結果，文語形を含む見出し語そのものの形での抽出は自動では簡単ではないことが確認された。しかしながら，「めでたし」の例のように，高頻度でありながら小型国語辞典の見出し語にない語を発見できることも示した。

4.2　語義の区分

本節において語義の区分について論じる。従来の「国語辞典」の多義語の記述の分析を述べ，多義語の記述に編集者の主観的判断が反映されている現状を示す。その上で，『IPAL 名詞辞書』の構築の際に筆者らが導入した，コーパス分析に基づく語義の区分方法を提案する。

4.2.1 現行の国語辞典における多義語記述の調査と分析

複数の意味や用法をもつ多義語の理解は，計算機による言語処理において大きな課題の 1 つである。例えば，情報検索や情報抽出などにおいて，語の多義性が支障になっており，多義性を解消して意味を特定するための手法の研究が進められてきている。そこで多義語の実態解明が必要になる。

多義語の辞書記述については従来から辞典間に差異のあることが指摘されていた。例えば，国広[1997]では，市販されている国語辞典どうしを比べると，区分する数や内容が一致しないことは多々あると述べている。そこで，本項では，多義語の実態解明のために，現行の国語辞典における多義語の意味記述を調査し，意味区分の構造化，意味区分の認定，用例提示という以上三点についての差異を考察する。使用した国語辞典を本稿で用いる略称とともに示すと，

『岩国』:『岩波国語辞典第六版』(岩波書店，2000 年),
『新明解』:『新明解国語辞典第六版』(三省堂，2004 年),
『新選』:『新選国語辞典第八版』(小学館，2001 年),
『明鏡』:『明鏡国語辞典』(大修館書店，2002 年),

以上 4 冊である。

調査対象は，動詞 100 語である。その 100 語は，『岩波国語辞典第五版』(岩波書店，1994 年)の多義語に基づく意味タグが付与された『毎日新聞記事データ』(3,000 記事)のテキストデータベース(新情報処理開発機構(RWCP))を利用し，そこから使用頻度の高い意味を含む上位のものより選定した。

4.2.2 語義の区分の構造化の比較

山田[2005]は『岩国』，『明鏡』，『新明解』3 つの辞典を比較した上で，次のことを指摘している。「辞書における意味記述の態度は，中核的意味やさらに抽象的な一般的意味を求めようとするかどうかによって大きく異なる。これらを求めようとする場合には，用法群が少ない数になり，そうでない場合には用法群の数が大きくなる傾向がある。」そして，《取る》の区分として，『新明解』は 8 個の中核的意味をあげ，『岩国』は 7 個の中核的意味(各々には 3 個ないし 8 個の用法群が対応させてある)をあげ，『明鏡』は中核的意味をたてず，59 個の用法群を列挙している，という差が認められることを示している。

そこで，動詞 100 語の各記述に対し，「階層」と「区分最大値」を求め，山田が指摘したような，辞典間の構造化の差異を定量的に確認することを試みた。まず，多義の「階層」数を出し，次に「階層別区分最大値」を出した。そして，「階層別区分最大値」の中での最大値を「区分最大値」とした。例えば，「(1) (ｱ) (ｲ) ，(2) (ｱ) (ｲ) (ｳ) (ｴ)，(3)」という記述の

場合は，(1)〜の階層の下に，(ｱ)〜という階層があるので，「階層」は 2 である。1 階層目は(3)まで，2 階層目は最大(ｴ)まであるので，「階層別区分最大値」は 3，4 になる。そして，この場合は，その 3 と 4 を比べ，大きい方の 4 が「最大区分値」になる。さらに，構造の詳細を把握するために，番号を列挙する「構造」表示も試みた。調査結果の一部として，頻度順位の上位 5 語と，「区分最大値」が『明鏡』で最大であった《かける》の結果，および，100 語分の「階層」と「区分最大値」の平均とを表 4.6 に示す。

本調査により，各辞典がどの程度階層化しているのか，あるいは，どの程度フラットに列挙しているのかを明示させることができた。表 4.6 にあげた一部分だけでも，『明鏡』の「区分最大値」の多さや，『岩国』のまとめあげる階層化が目に付く。100 語全般にわたってこの傾向は強く見られた。この傾向は何より，100 語分の平均値によく現れている。なお，『新明解』が『岩国』寄り，『新選』は『明鏡』寄り，ということも本調査結果から判明した。

表 4.6　頻度順位の上位 5 語と《かける》の記述の比較

頻度順位	語	岩国				新明解				新選				明鏡			
		階層	階層別区分最大値	区分最大値	構造	階層	階層別区分最大値	区分最大値	構造	階層	階層別区分最大値	区分最大値	構造	階層	階層別区分最大値	区分最大値	構造
1	する	2	8,4	8	1*(1,2),2*(1〜3),3,4,5,6*(1〜4),7!,8*(1〜4)	2	2,4	4	1*(1〜5),2*(1〜3)	2	3,8	8	1*(1〜8),2*(1〜7),3*(1〜7)	3	4,11,3	11	1*(1(1〜3),2〜6,7*(1,2)),2*(1,2!,3〜10,11!),3*(1〜4),4
2	いる	2	3,3	3	1,2*(1〜3),3	1	2	2	1,2	2	10,3	10	1!,2!,3!,4,5*(1〜3),6〜10	3	2,2,5	5	1*(1,2),2*(1*(1〜5),2)
3	なる	2	3,3	3	1!(1(1,2!,3),2!,3!)	2	3,4	4	1,2*(1〜4),3*(1,2!)	1	13	13	1〜13	2	2,15	15	1!(1*(1〜15),2*(1!,2,3))
4	ある	3	3,3,7	7	1!(1,2!(1〜7),3),2*(1!(1,2),2),3	2	4,5	5	1*(1〜3),2*(1〜5),3*(1,2),4*(1〜3)	2	13,3	13	1〜12,13*(1〜3)	3	2,18,2	18	1!(1〜18),2*(1〜4(1,2),5)
5	いう	2	3,7	7	1!(1,2),2,3*(1〜7)	2	2,3	3	1*(1,2),2*(1〜3)	2	2,10	10	1*(1〜3),2*(1〜10)	3	2,6,21	21	1*(1〜5,6*(1〜21)),2
32	かける	3	5,6	6	1!(1!(1〜4),2,3!(1〜3!),4!(1〜6)),2!(1〜3!(1〜5),4),3!(1!(1〜3),2!(1〜4),3!(1,2),4!(1,2),5),4	2	2,7	7	1*(1〜6,7!),2*(1,2)	3	3,28,4	28	1*(1〜8(1,2),9〜14!(1〜3),15,16!(1〜4),17〜28),2*(1,2),3	2	2,48	48	1*(1〜48),2*(1〜3)
						1	1	1	1	1	2	2	1,2				
100語平均		1.8		3.7		1.4		3.6		1.6		5.9		1.4		10.0	

4.2.3 語義の区分と用例提示の比較

次に，《得る》と《進む》という 2 語の具体的な記述例を通し，語義の区分と用例提示の差異について比較，分析する。

4.2.3.1 《得る》の場合

はじめに，語義の区分の差異を見るために，『岩国』の「中核的意味型」の記述に対し，『明鏡』の「用法群列挙型」の記述が顕著な例として，《得る》の 1 番目の区分を取り上げる。

『岩国』

(ア) 手に入れる。自分の物にする。「知識をえる」「志をえる」

(イ) 理解する。さとる。「その意をえない」

『明鏡』

① 有益な物事を手に入れる。自分のものにする。獲得する。「優勝して賞金を—」「協力者を得て事業を始める」「収入を—」「書物から情報を—」「権利を—」

② 好ましくない物事を身に受ける。「病を得て郷里に帰る」「罪を得て遠島の身となる」

③ あることがきっかけとなって、力や想念・霊感などが与えられる。それを自分のものとする。「激励のことばに力を—」「成功に自信を得て次作に取りかかる」「撤退もやむなしとの結論を—」「解散は必至という感触を—」「霊感を得て詩作に興じる」

④ 相手から許諾や賛同などを受ける。「会長から許可を—」「貴兄の御了承を得て大慶に存じます」「スタッフの信頼を—」

《得る》はヲ格に様々な名詞をとる。その点に『明鏡』は着目し，ヲ格にくる名詞群を「有益なこと」「好ましくないこと」「力や想念・霊感など」「許諾や賛同」という観点で 4 分類して用法を捉えている。それに対し，『岩国』は，このようなヲ格の分類は，あくまでもヲ格の分類に過ぎず，《得る》そのものの意味が分かれるものではない，という立場かと思われる。『明鏡』のようにヲ格で分類して用例を多くあげることは，個々の用法がより理解しやすいという効果があるだろう。一方，『岩国』のように，《得る》そのものの大きな

意味分類を示すことは,《得る》という動詞そのものの意味把握がしやすいという効果があるだろう。

次に，同じ《得る》の辞書記述を通し，用例提示の差異をみてみる。『岩国』では唯一，(ア)に対し，(イ)「理解する。さとる。」を分けているが，その(イ)にあげている「その意をえない」という典型と思われる用例が『明鏡』で取り上げられていないため，その用法が『明鏡』の①～④のどこに該当するのかが実ははっきりしない。

次に,『新選』の記述を引く。

『新選』

① 自分のものにする。手に入れる。

② とげる。「志を—」

『新選』では興味深いことに，②「とげる」を唯一①より分けているのだが，これは『岩国』の(ア)に同一に分類された用例 2 例を区分する態度である。『岩国』が(イ)として区別した用例は特別に取り上げず，明らかに記述態度の違いを見せている。また，この『新選』が区別にこだわった②の「志を—」という典型的と思われる用例がまた『明鏡』では取り上げられていない。これも『明鏡』では③であろうか。ちなみに,『新明解』では用例に注記を付与するという形で『岩国』と『新選』とがそれぞれにこだわった用法を明記するという記述態度をとっている。

『新明解』

(一)自分の物にする(手に入れる)。「成案(勢い・力・教訓・承認)を—／理解(支持)が得られる／時宜(当)を得た処置／所を—〔＝その人の希望通りの(に ふさわしい)仕事や地位につく〕／病を—〔＝病気にかかる〕／志を—〔＝所を得て、自分の抱負を実行することが出来る〕／信頼(信任)を—〔＝受ける〕／貴意を得たい〔＝ご承知いただきたい〕／ご出席を—〔＝出席していただく〕ことが出来ましたことは…／その意を得ない〔＝意味がよくわからない〕」

以上，見たように，《得る》の場合，語義の区分や用例提示は，辞典によって差異がある。用法，用例の網羅的記述という点において，従来の国語辞典には課題のあることがわかる。

4.2.3.2 《進む》の場合

次に，『岩国』と『明鏡』とで区分数はほぼ同じであった《進む》を比較する。各記述は次の通りである。

『岩国』

(勢いに乗って)前へ出る。

(1) 自分が向かっている方向に動く。

(ア) 前方へ行く。⇔退く・引く。「三歩―」「転じて東へ―」

(イ) 目標に向かって乗り出す。「芸能界に―」「理科系学科に―」

(ウ) 《「ー・んで…する」の形で》自分から積極的に…する。乗り気になって…する。「ー・んで勉強する」

(エ) 仕事がはかどる。「工事が着着と―(＝完成に向かう)」

(2) 標準・普通の動き方よりはやい。⇔遅れる。「時計が二分―」

(3) 度合が高まる。

(ア) 物事の程度や内容がよくなる。⇔遅れる。「ー・んだ技術」

(イ) 地位・階級があがる。「官位が―」

(ウ) 勢いが増す。⇔衰える。「食が―」(食欲が盛んになる)「病勢が―」(病気が悪化する)

『明鏡』

①前方に向かって移動する。前進する。「一歩前へ―」「船が波をけたてて―」「未来［目標］に向かって―」

②上の段階・地位に移行する。上がる。「高校から大学へ―」「将棋の道に―」「課長から部長に―」「準決勝に―」

③物事が(計画した通りに)進行する。「仕事が予定通りに―」「工事が―」「筆が―」「大いに食が―」

④物事の状態・程度がはなはだしくなる。進行する。「過疎化［相互理解・世代交代・活字離れ］が—」「病状［近視］が—」

⑤(他に先んじて)物事が好ましい方向に動く。発展する。進歩する。「文明が—」「公害対策が他県より一歩—・んでいる」「考え方が—・んでいる」

⑥《「気が—」の形で》積極的に…しようという気持ちになる。気乗りがする。下に打ち消しを伴うことが多い。「宮仕えは気が—・まない」

⑦時計の針が正しい時刻より先のほうを指すようになる。「この時計は五分—・んでいる」

⑧《「—・んで」の形で、副詞的に》みずから積極的に物事を行う。「—・んで事に当たる」「みずから—・んで応募する」

『岩国』は 3 個の中核的意味(1 には 4 個，3 には 3 個の用法群が対応)をあげ，『明鏡』は中核的意味をたてず，8 個の用法群を列挙している。構造化には違いが見られるが，区分の合計数では 1 しか差がない。しかし，このように同様の区分数でも，個々の記述には異なる点がある。以下，問題としたい 4 点を順に指摘する。

まず，『明鏡』の 1 つの区分が，『岩国』の 2 つの区分にまたがっているように見えるところがある。それが『明鏡』の②である。『明鏡』②の 4 例のうちの 2 例，「課長から部長に—」「準決勝に—」は，『岩国』(3)(イ)「官位が—」の類例と見られるが，「将棋の道に—」は『岩国』の(1)(イ)に分類される「芸能界に—」「理科系学科に—」の類例ではないだろうか。『明鏡』②の「高校から大学へ—」と『岩国』の(1) (イ)の「理科系学科に—」が同じ分類になるものであるのか異なる分類になるものであるのかは，意見が分かれるところかと思われる。「高校から大学へ—」は，段階が上がることに着目すれば，そのまま「課長から部長に—」などの類例となるが，「大学へ」というところを「理科系学科に」と同様，「目標に向かって乗り出す」と捉えることも可能であろう。

次に，『明鏡』と『岩国』と，釈義を読む限り，区分は対応しているものの，用例が異なる区分に分類されている箇所がある。それは，『明鏡』③と④，『岩国』(1)(エ)と(3) (ウ)である。『岩国』(3) (ウ)で，「食が—」と「病勢が—」は同じ分類である。ところが，『明鏡』では「大いに食が—」は③に，「病状［近視］が—」は④にと，別分類になっている。筆者の直感では『岩国』の分類が妥当に思えるが，試みに筆者以外の人に意見を求めると，『明鏡』の分類が妥当に思える，との意見もあった。ちなみに，『岩国』「病勢が—」「工事が着着と—(＝完成に向かう)」と，『明鏡』「病状［近視］が—」「工事が—」は対応がとれてお

り，そこに問題はない。ここで問題なのは「食が進む」をどう捉えたか，という点にある。「食が進む」を「食欲が増進している」ことと捉えれば『岩国』の分類になろう。一方，「食が進む」を，お皿が次々に空になっていくプロセス進行と捉えれば『明鏡』の分類になろう。

異なる点の 3 つ目は,『明鏡』⑦と『岩国』(2)の時計に関する釈義の幅の違いである。『岩国』が広く,『明鏡』は限定的な記述になっている。実際にその意味・用法が限られたものである場合は限定的な記述のほうがより事柄がはっきりして有効であろう。しかし，それに類する別の用例が存在する場合，扱えなくなってしまう。例えば，予定より早く運動会のプログラムが進行しているような場合,「プログラムが進んでいる」と言うのではないだろうか。これは『明鏡』⑦や『岩国』(2)の時計の例と同じ区分になる例ではないだろうか。そう思えばこれらの例の追加分類を受け入れやすい『岩国』の釈義が妥当にみえる。ただ，例えば，プログラムの用例は時計の比喩表現である，と見たとしよう。そうであれば，時計に限定した語釈のところにプログラムの用例も分類が可能であり，本義をはっきり示すことが重要であると考えれば『明鏡』の記述が妥当だと言えよう。

そもそも，比喩表現の用例は，柏野[2006]でも取り上げたように，程度によって本義の用例に連続して扱われる場合と，独立した区分をたて，別に扱われる場合とがある。《進む》にもそのような箇所がある。『明鏡』①の「未来［目標］に向かって―」は本義の用例と連続して扱われているが,『岩国』の(1) (イ)は，その『明鏡』の用例と同じ「目標に向かって」という語句を用いた釈義の独立した区分が立っているのである。このように，比喩表現の扱いは辞典によってゆれが出るところである。

4.2.4 考察

国語辞典における多義語記述の実態を把握するために，語義の区分の構造化，語義の区分，用例提示に着目して,『岩国』,『新明解』,『新選』,『明鏡』という 4 冊の国語辞典における多義語の記述の差異を調査分析した。

はじめに，語義の区分の構造化について調査した。動詞 100 語の意味区分の「階層」と「区分最大値」を求めて比較した。その結果,『明鏡』の「区分最大値」の多さ,『岩国』のまとめあげる「階層」の多さが顕著であった。『明鏡』のように用法群の列挙を追求するタイプの辞典では，事柄が細分化され数多くの意味区分が設けられていることが,『岩国』

のように中核的意味を求めるタイプの辞典では用法群がまとめられ階層化されていることが，確認できた。

次に，語義の区分，用例提示の差異について調査した。『岩国』,『新明解』の《得る》と《進む》の記述を取り上げ比較した。その結果,《得る》では，構造化の態度により，区分が異なることを確認した。また,『新選』,『明鏡』とあわせて 4 つの辞書記述を比較した結果，ある辞書で取り上げている用法・用例が，別の辞書にはない，といったことが確認された。つまり，用法・用例の網羅性という点において，従来の国語辞典には課題のあることが確認された。

《進む》の比較結果からは，そもそも語義の区分，適切な用例の選定が，辞書間で必ずしも一致するものではない，という事実が確認された。それは，一方が正しく他方が誤っている，というものではなく，観点の違いによるものであった。言葉の意味・用法の捉え方は，観点によって異なるものであり，一意に決まるものではない，という事実である。

以上，語義の区分の構造化，語義の区分，用例提示，いずれにおいても，現行の国語辞典には差異があることが確認された。この差異は，辞典の編集者の方針や判断が異なる以上，必然的に生じるものであると言えるだろう。

辞書を用いる目的により望ましい辞書記述は変わってくるため，あらゆる要求を満たす記述は実現困難であろう。しかしながら，用途に応じた方法論の構築は，言語処理，言語理解のために必要である。そのためには，さまざまな辞書記述の方法とその効果の分析が有効と考える。

4.2.5 語義の区分の手順

コーパス分析に基づく, 語義の区分方法について提案する。この方法は,『計算機用日本語基本名詞辞書 IPAL(Basic Nouns)』[情報処理振興事業協会 1996](以下,『IPAL 名詞辞書』) を構築する際に, 筆者らが導入したものである[桑畑, 橋本 1994, Kuwahata et al. 1995]。『IPAL 名詞辞書』では, 見出し語のもつ用法を意味だけからではなく統語的特徴からも分類し, これを「下位区分」と呼び, コーパスから得られる用例分析によって「下位区分」する基準を 12 通り設ける。

4.2.5.1 「慣用句」と下位区分

『IPAL 名詞辞書』では下位区分ごとに，下記の情報を記載する。

A. 形態情報

B. 意味情報

C. 述語の項としての用法

D. 連体修飾語としての用法

E. 連体被修飾語としての用法

F. 述語としての用法

G. サ変動詞としての用法

このほか，下位区分はしないが，見出し語の用例として記載するものとして，「慣用句」がある。「慣用句」は，(1)個々の語の意味の総和が当該句全体の意味にならない，(2)述語の活用形や使役化，否定化，丁寧化や，さらに修飾語の付加や省略の可能性などに制限が見られる，という特徴を持つ。例えば，次のものがそうである。

1)鼻　が　高い

この表現が「得意になっている」という意味で用いられている時は，「鼻」には実際の「鼻」の意味がない。このような場合を「慣用句」と呼ぶ。『IPAL 名詞辞書』では，1 つの見出し語につき 1 つの「慣用句」欄を設け，その欄にまとめて記載する。

一方，「鼻が大きい」「鼻がいい」の場合には，それぞれ，「動物の顔にあり，呼吸や発声を助ける器官」「『鼻』の機能・働き」のように，句全体の意味を個々の語の意味へ還元することができ，文法的なふるまいに制限がない結び付きである(これを「コロケーション」と呼ぶ)。そこで，これらは見出し語の一般的な用法であると認め，それ 1 つの下位区分をたて，先にあげた A から G までの情報の記述対象とする。

中には，「慣用句」とすべきなのか，見出し語の 1 つの「コロケーション」として下位区分をたてて記述すべきなのか，判断に迷うものがある。次にあげる見出し語「腕」の例は，そのようなものの 1 つである。

2)腕　が　いい

身体部位としての「腕」を指してはいないが，この時の「腕」には次のように複数のコロケーションをコーパスから得ることができる。

3)腕　が　確かだ／上がる／鈍る／立つ
4)腕　を　磨く／上げる／競う／ふるう

これらの結び付きから，「腕」に「技量」という意味を認めて，下位区分を 1 つ設け，3)，4)は「述語の項としての用法」として記述すべきであると判断できる。

「腕」の例とは違い，複数のコロケーションを持たず，コーパスからも，内省からも，結び付く述語が 3 つより少ないと判断されるものについては，「慣用句」とする。3 つは便宜的に定めた数である。次の見出し語「虫」の例では，「第六感」という意味を認めることもできそうであるが，結び付く語の数が二例と限られているため，「慣用句」と判断する。

5)虫　が　知らせる
6)虫　の　知らせ

4.2.5.2 「下位区分」の判断基準

辞書記述の際には，「慣用句」ではないという判断をした後に，複数ある用法を「下位区分」するが，実はこの判断は，「慣用句」であるか否かの判断よりも複雑である。結び付く述語の違いが顕著であれば，そこに見出し語の用法の違いを見出しやすいが，実際には述語にくる動詞や形容詞なども多義であるため，見出し語の用法が違っても結び付く述語に差はなく，それだけでは「下位区分」の判断がつかないものが多くある(コロケーション重複の問題は 5.1 節で論じる)。また，辞書を共同作業で執筆する場合，人によって下位区分数が多かったり少なかったりという問題も生じる。そこで，「下位区分」する際の基準を設け，判断の助けとする方法を提案する。「下位区分」の判断基準として用いるものを列挙すると，以下の 12 通りである(「NP0」は見出し語を表す)。

A 形式的な基準

1. 結び付く述語の違い
2. サ変動詞用法の有無の違い
3. 述語用法の有無の違い

4. 連体修飾語の違い(NP の／な NP0)

5. 連体被修飾語の違い(NP0 の／な NP)

6. 合成語の違い

B 意味的な基準

7. 意味カテゴリの違い

8. 意味素性の組み合わせの違い

9. 類義語や対語の組み合わせの違い

10. 比喩的表現か否かの違い

11. 総称か特化かの違い

12. 対訳の違い

これらは相互に関係しあっていて，どれか 1 つの基準だけで十分であるということは少なく，いくつかの基準を組み合わせて「下位区分」するものである。

4.2.6 下位区分の判断基準の適用例

「**要件 1-2** 語義の区分が客観的・明示的であること」を満たすために，用例や用法に着目して区分する12通りの「下位区分の判断基準」を設けた。その「下位区分の判断基準」によって弁別される適用例により基準の有効性を確認する。

4.2.6.1 適用方法

「下位区分の判断基準」を弁別する適用例として，下記の見出し語を用いる。

	見出し語
A 形式的な基準	
1. 結び付く述語の違い	根
2. サ変動詞用法の有無の違い	収穫
3. 述語用法の有無の違い	田舎
4. 連体修飾語の違い(NP の／な NP0)	点数
5. 連体被修飾語の違い(NP0 の／な NP)	窮屈
6. 合成語の違い	神経

B 意味的な基準

7. 意味カテゴリの違い	貝
8. 意味素性の組み合わせの違い	みかん
9. 類義語や対語の組み合わせの違い	悪質
10. 比喩的表現か否かの違い	雨
11. 総称か特化かの違い	動物, ごはん, 酒
12. 対訳の違い	応援

4.2.6.2 適用結果

A 形式的な基準による下位区分

1.「結び付く述語の違い」を用いる例

見出し語「根」についてコーパスから得られるコロケーションは, 次のように 3 タイプある。

1) 植物の成長に関するもの	根	がつく／が伸びる／が生える／が枯れる
2) つながりに関するもの	根	を絶やす／を断つ／が深い／をなくす
3) 人の性格に関するもの	根	が優しい／が明るい／が卑しい

そこで,「根」は次のように三つに下位区分する。

1. 植物を支え, 水分／養分を吸収するため, 地中にある器管。	1)
2. 抽象的な物事のもととなっている部分。	2)
3. 生まれ持った性格。	3)

一見すると, 述語の違いで 3 つの下位区分が判断できているかにみえるが, 実は, 2)と同じ述語が 1)の述語となって現れることもある。1)と 2)は, 述語の違い以上に,「(植物の)根」であるか,「(物事の)根」であるかという,「連体修飾語」の違いによる弁別が効く。また, 3)についても,「(人の)根」であるか否かによる弁別が可能である。よって, この場合は,「述語の違い」だけでなく,「連体修飾語の違い」も併用しての区分ということになる(コロケーションの重複の問題は 5.1 節で改めて論じる)。

2.「サ変動詞用法の有無」を用いる例

見出し語「収穫」については，次のようなコロケーションがコーパスから得られる。

1)ブドウの収穫　が　　始まる／終わる
2)大豆の収穫　が　　ある／ない／
　　多い／少ない／
　　減る／増える
3)今日の会議は収穫　が　ある／ない／
　　多い／少ない

2)と 3)とでは，「減る，増える」を除いて，「ある，ない，多い，少ない」など，結び付く述語に類似性がみられる。ここからは，2)は 1)よりも 3)に近いようにみえる。そこで，サ変動詞としての用法を確認する。

4)農夫がブドウを　　収獲する
　農家が大豆を　　収穫する

「収穫する」のヲ格にたつ名詞句は，「ブドウ」や「大豆」などの作物を表す名詞句である。これらの名詞句は，1)，2)の「収穫」の連体修飾語として現れるものである。一方，以下の 5)に示すように，3)の「収穫」と共起する名詞句を組み合わせても，サ変動詞用法は成り立たない。

5)*議長が会議を　　収穫する
　*今日の会議で知識を　収穫する

したがって，1)，2)の「収穫」と 3)の「収穫」とでは，結び付く述語に同じものがあっても，修飾成分が異なり，その文法的ふるまいに違いがあることがわかる。よって「収穫」は次のように 1)，2)と 3)の 2 つに下位区分する。

1. 農作物をとり入れること。または取り入れた農作物など。　1)　2)
2. ある事をして得た有益な結果。　3)

3.「述語用法の有無」を用いる例

続いて，見出し語「田舎」についてコーパスから得たコロケーション例をあげる。

1)田舎 に 生まれる／育つ

2)田舎 が ある／ない

3)田舎 へ 帰る／電話する

ここで，「とても」という程度副詞との共起関係を見ると，違いがはっきり見えてくる。

1)とても田舎 に 生まれる／育つ

2)*とても田舎 が ある／ない

3)*とても田舎 へ 帰る／電話する

つまり，「田舎」は，述語用法があるものとないものとに区分されることがわかる。ここでいう「述語用法」とは，次のような，程度副詞を伴って述語になる用法のことをいう。

4)私が住んでいるところはとても田舎だ

結局，「田舎」は以下のように 2 つに下位区分する。

例:「いなか【田舎】」

1. 都会から離れた所。 1) 4)

2. 故郷または出身地。 2) 3)

4.「結び付く述語と連体修飾語の違い」を用いる例

次に「点数」という見出し語について考察する。この「点数」という見出し語には，サ変動詞としての用法や述語用法がないため，先にあげた方法では下位区分することができない。このような場合に下位区分の判断基準となるのは，「結び付く述語の違い」と「連体修飾語の違い」である。コーパスから得られるコロケーションに，連体修飾語を付したまま示すと，次の通りである。

1)応募作品の点数 が 多い／少ない

2)絵画の点数 を 増やす／減らす

3)数学の点数 が 高い／低い

4)試験の点数 を 上げる／下げる

上の例では，1)，2)の「多い，少ない，増やす，減らす」と結び付く「点数」と，3)，4)の「高い，低い，上げる，下げる」と結び付く「点数」とでは，性格に違いがあると考えられる。また，それぞれの「点数」の連体修飾語にも，異なる性格の名詞句が現れている。1)，2)は修飾語に現れる名詞句が指示するものの個数を問題にするものであり，3)，4)は程度を問題にするものである。このような理由から「点数」については 1)，2)と 3)，4)の 2 つに下位区分することができる。

例:「てんすう【点数】」

1. 試験やゲームなどの成績を表した数。1) 2)
2. 商品・作品などの数。 3) 4)

5. 「被修飾語の違い」を用いる例

「窮屈」という見出し語は，格助詞ガやヲを介して述語と結び付くという用法がない。また，ほかの名詞の連体修飾を受ける用法もない。コーパスから得ることのできるコロケーションは，次のような a)連体用法，b)述語用法，c)格助詞ニを介して述語と結び付く用法のものである。

a1)窮屈なズボンははきたくない

a2)窮屈な式典は肩がこる

a3)窮屈な制度は変えていくべきだ

b1)ズボンが窮屈だ

b2)式典が窮屈だ

b3)制度が窮屈だ

c1)ズボンが窮屈になる

c2)式典が窮屈になる

c3)制度が窮屈になる

そこで，a)連体用法の時に現れる「被修飾語の違い」に着目する。b)述語用法や，c)格助詞ニを介して述語と結び付く用法の時にも違いとなって現れるのは，この「被修飾語」である。これを用いて次のように下位区分する。

例:「きゅうくつ【窮屈】」

1. 狭くて，思うように身動きができないさま。 1)

[被修飾語]—な服，—な部屋，—な姿勢

2. 堅苦しく，気詰まりに感じるさま。 2)

[被修飾語]—な式典，—な雰囲気，—な感じ

3. ぎりぎりでゆとりがないさま。 3)

[被修飾語]—な制度，—な制限，—な法，—な金利体系，—な暮らし，—な生活

6.「合成語の違い」を用いる例

見出し語「神経」の場合，結び付く述語にも違いがあるが，「神経」を語基として成立する合成語の違いによっても下位区分できる。合成語の例もまた，コーパスから多く抽出可能である。これを手がかりとする。

例:「しんけい【神経】」

1.「しんけい」とよばれる身体器官。

例: 神経が 麻痺する／を切る／を抜く

[合成語]—細胞，—中枢，—病，視—，自立—，運動—，中枢—，脳—，交感—

2.「しんけい」の機能・働き。

例: 神経 が弱る／を痛める／が過敏だ

[合成語]—疾患，—障害

3. 精神の働き，機能。

例: 神経 が細い／細かい

図太い神経 を している

[合成語]—安定剤，—過敏，—衰弱，—戦，無(む)—，—質，—性，—症

B 意味的な基準による下位区分

7.「意味カテゴリの違い」を用いる例

ここで，意味カテゴリというのは，「生物」「植物」「食べ物」「乗物」「衣服」「時間」「場所」などのような大きな意味分類を表す。コーパスから得た用例を分析し，名詞の属する意味カテゴリを判断することによって，下位区分をするという方法である。見出し語「貝」は，「生物」「食べ物」「物」などの意味カテゴリの違いで 3 つに下位区分できる。

例:「かい【貝】」

1. 硬い殻をもち，水中に生息する「カイ」という生き物。

例: 貝が棲む／プランクトンを食べる

貝を育てる

2.「カイ」という生き物の食用部分。

例: 貝がおいしい／新鮮だ

貝を食べる／焼く

3.「カイ」という生き物のからだを覆っている硬い殻。

例: 貝を拾う／細工する／削る

8.「意味素性の組み合わせの違い」を用いる例

名詞と述語が格助詞を介して結び付く時，その述語によって名詞の持つ様々な側面の 1 つに焦点が当てられる。例えば，「手紙を破る」という時には，「手紙」の物質的な側面に焦点が当てられている。これに対して，「手紙を読む」では「手紙」の持つ内容的な側面に焦点が当てられている。『IPAL 名詞辞書』における意味素性[青山 1989]は，ある見出し語が持つ下位区分ごとに，結び付く述語によって焦点があてられる側面とその範囲を捉えようとするものである。これにより，コーパスから得られる数多い用例を見通しよく整理することが可能になる。

例えば，「みかん」という見出し語では，「ミカン」という植物そのものと，「ミカン」という植物の果実部分とを表す場合が考えられる。両者とも，具体物としての側面に焦点を当てる述語として「売る」「買う」のように，同じ動詞が共起するため，〈CON〉(具体物)という素性が振られる。さらに両者とも，植物としての側面に焦点が当てられる述語と共起するため，〈PLA〉(植物)という素性が振られる。しかし，「みかん」が植物を表す場合には，「食べる」「甘い」のような述語との共起は考えにくいため，〈EDI〉(可食物)という意味素性は現れない。

両者の違いは，先にあげた意味カテゴリの違い，「植物」「食べ物」の違いで下位区分することができるが，意味素性の組合せからも，〈EDI〉の有無によって次のように区分することができる。

例:「みかん」

1.「みかん 1」「ミカン」と総称される植物。

1.1〈PLA:植物としての側面に焦点が当てられるもの〉

例: みかんを栽培する

みかんが枯れる

1.2〈CON:具体物としての側面に焦点が当てられるもの〉

例: みかんを買う

みかんを売る

2.「みかん 2」「ミカン」と総称される植物の果実部分。

2.1〈EDI:可食物としての側面に焦点が当てられるもの〉

例: みかんを食べる

みかんが甘い

2.2〈PLA:植物としての側面に焦点が当てられるもの〉

例: みかんがなる

みかんが熟れる

2.3〈CON:具体物としての側面に焦点が当てられるもの〉

例: みかんを買う

みかんを盛る

9.「類義語や対語の組合せの違い」を用いる例

見出し語「悪質」について，コーパスから得られる用例をあげる。

a1)最近のスリの手口はますます悪質になってきている

a2)悪質な事件が多発している

a3)悪質な石炭を使わないようにする

a4)最近，悪質の酒が出回っているという

先ほどの「窮屈」という見出し語と同様に，格助詞ガやヲを介して述語と結び付く用法がなく，ほかの名詞の連体修飾を受ける用法もない。よって，「被修飾語の違い」をみるわけだが，この見出し語の場合，類義語の「粗悪」と対語の「良質」を用いることで違いがより明らかになる。

b1)*そのスリの手口は　粗悪だ／良質ではない

b2)*その事件は　粗悪だ／良質ではない

b3)その石炭は　粗悪だ／良質ではない

b4)その酒は　粗悪だ／良質ではない

つまり，類義語に「粗悪」を，対語に「良質」を持つものと持たないものとに，次のように下位区分される。

例:「あくしつ【悪質】」

1. 行いがひどくたちの悪いさま。　1) 2)

2. 品質が悪いさま。　3) 4)

[類義語]粗悪

[対語]良質

10.「比喩的表現か否かの違い」を用いる例

コーパスから得られる用例を整理する観点の一つに，比喩表現か否かの判断がある。例えば，「あの人は本の虫だ」の「虫」や「先日の講演会では講師が壇上で涙ぐむ一幕があった」の「一幕」のように，ある名詞がその指されたものの性質・特徴とよく似た性質・特徴をもつ別のものを表すような用例である。これらは，比喩として使用されたと推測されるが，現代語ではかなり一般的に使われ，1 つの用法として定着していると判断してもよいものである。『IPAL 名詞辞書』では，このような名詞の表現を見出し語のもつ 1 つの用法として認め，下位区分としてたてる。例えば，次の「虫」「卵」「種」「壁」「雨」「薬」「せのび」のようなものである。

1)彼は本の虫だ

2)医者の卵と付き合う

3)争いの種になる

4)言葉の壁がある

5)火の雨が降りそそぐ

6)薬が効きすぎたようだ

7)少女がせのびをしてお化粧をする

1)～5)では，連体修飾語の「NP の」の部分が必須であり，これらの修飾語は，その見出し語が本来的な意味を表していないことを明示している。これらは，修飾語の違いとしても下位区分できる。一方，6)～7)では，文脈の助けがないと本来的な意味との区別は難しい。このような場合には，形式的な基準を用いて下位区分することはできないが，意味的な違いを認めて別の下位区分をたてる。このように，比喩的な意味を 1 つの下位区分として認める見出し語については，その比喩的な意味を表す下位区分と本来的な意味を表す下位区分とが大きく区別できるように区分する。

例えば，「雨」では，本来的な「雨」を表す下位区分と，5)で示したような「絶えまなくたくさん降りそそぐもの」という下位区分と二分する。そのため，前者の「雨」には，多様な述語が共起することになるが，これらは，次のように意味素性を分けることによって細分される。

例:「あめ【雨】」

1.「雨 1」広域に空から降ってくる水の粒。それが降っている状態。

1.1〈CON:具体物〉

例: 雨 がたまる／が冷たい／をよける

1.2〈LIQ:液体〉

例: 雨 が漏る／がしみ込む／に濡れる

1.3〈PHE:成立・消滅・変化する自然現象〉

例: 雨 が降る／にあう／になる

1.4〈NAT:存在・移動する自然現象〉

例: 雨 が来る／が近づく／を予報する

2.「雨 2」絶え間なくたくさん降り注ぐもの。

2.1〈CON:具体物〉

例: [血の／爆弾の]雨 が降る／が降り注ぐ／を浴びる

11.「総称か特化かの違い」を用いる例

コーパスで得られる用例のうち,「結び付く述語」に違いがあっても細かく下位区分をしない場合がある。例えば, 見出し語「時計」の下位区分は次のように 1 つにする。

例:「とけい【時計】」

1. 時刻を知るための機械。

例: 時計が止まる／進む／遅れる

時計を見る／探す

実は,「時計」には形状によって次のように「結び付く述語」に違いがある。

1)(掛け)時計 を かける／飾る

2)(置き)時計 を 置く／飾る

3)(懐中)時計 を 下げる

4)(腕)時計 を つける／はめる

この違いで下位区分しない理由は,「時計」が「総称」を表すものであるからである。これを形状差によって下位区分すれば, 辞書記述が冗長になる恐れがある。例えば, 見出し語「コップ」を「コップが割れる」といえる「ガラス製」と, いえない「紙製」とで下位区分をするようなことには, 辞書的な意味の弁別としての必要性を感じない。自然言語処理において, 形状が問題になることが予想されないわけではないが, 辞書的意味としては「総称」である名詞は, 大きく用法を捉える方が良いと考え, 形状差による下位区分は行わない。同じ理由で, 見出し語「薬」は,「飲み薬」「塗り薬」「貼り薬」で分けることをしない。物としての側面に焦点が当てられる時には,「飲む」「塗る」「貼る」という違う述語と結び付くが, 効果としての側面に焦点が当てられる時には,「効く」「作用する」「強い」「弱い」という同じ述語と結び付くことからも,「薬」を形状で区分する必要がないことがわかる。

しかし, 見出し語が「総称」を表す場合と,「特化」したものを表す場合とがある時には, その両者を下位区分する。以下に,「特化」で下位区分するものを 3 例示す。

例:「どうぶつ【動物】」

1.「ヒト」を含む生き物。

例: 火星には動物がいない

2.「ヒト」をのぞく生き物。

例: サファリパークには色々な動物がいる

例:「ごはん【御飯】」

1. 食事。

例: 今日は外でごはんを食べよう

2. 米を炊いて作る食物。

例: 一日一膳は御飯を食べる

例:「さけ【酒】」

1. アルコールを成分として含む飲料，又，それが身体に及ぼす作用。

例: いろいろな酒が置いてある

2. 米を原料とする，日本に古くから伝わるアルコール飲料，又，それが身体に及ぼす作用。

例: 酒を熱燗にする

上の例は 3 例とも，1 が「総称」であり，2 が「特化」である。

12.「類義語と対訳の違い」を用いる例

コーパスから得られた用例に基づき，見出し語「応援」を下位区分する結果を次に示す。

例:「おうえん【応援】」

1.1 他人の手助けをすること。〈ACT〉

例: 選挙の応援を要請する

1.2 他人の手助けをするひと。〈HUM〉

例: 引っ越しにたくさんの応援をよこす

2.1 声をかけたりしてチームや選手を元気づけること。〈ACT〉

例: チームの応援をする

2.2 声援を送っている人。〈HUM〉

例: 試合にたくさんの応援が集まった

実は，1 と 2 は「結び付く述語」や「連体修飾語」などにそれほどの差がなく，また，「意味素性の組合せ」も同じであるので，「相手を励まし，助ける行為」という下位区分を 1 つたてれば十分であるか迷う例である。しかし，以下のように「類義語」や「英訳」の違いに注意すれば，結局上のように 2 つに下位区分するのが適切であると結論できる。

a. 類義語が違う　　1. 支援，手助け
　　2. 声援，励まし
b. 英訳語が違う　　1. help
　　2. cheer

しかし，「対訳の違い」は，「下位区分の判断基準」の中でも適用度は低く，下位区分する絶対条件ではない。例えば，前にあげた「時計」は，形状によって以下のように英訳が分かれるものであった。

掛け時計: a (wall) clock
置時計: a (table) clock
懐中時計: a (pocket) watch
腕時計: a (wrist) watch

この「時計」の場合は先ほども述べた通り，「総称」であるものは 1 つの下位区分にするという方針にのっとり，「対訳の違い」は考慮しないで，「時刻を知るための機械」という下位区分を 1 つたてる。

4.2.6.3 適用結果のまとめ

以上，下位区分の基準が適用される例を通し，基準による弁別の可能性が確認できた。「結び付く述語の違い」だけで区分できる名詞は実際は少ないため，この基準表を上から順次照らし合わせながら，様々な情報を用いての総合的な判断を行い，区分することが有効である。基準があることにより，可能な限り客観的に区分を行うことができる。また，区分した結果を説明することが可能になることも大きな利点であると言える。

『IPAL 名詞辞書』の構築の際はこの基準を適用し，共同作業による辞書構築を行った。下位区分を説明可能に行うというこの方法は，均質な辞書を効率よく構築するために有効である。

4.3 多義構造の記述

多義語のもつ複数の意味や用法が無関係に存在するものではなく，互いに関係(多義構造)をもっていることに注意することは多義性解消のための辞書情報として極めて重要である。そこでコーパスから得られる用例から多義構造(語義間の意味的関係)を捉え，明示的に記述する方法を提案する。

多義語の辞書記述において，例えば基本義と派生義とで整理し捉えようとする議論[国広 1982, 1997 など]はあったものの，意味や用法どうしの関係を体系的に記述した例は，これまでほとんど見られなかった。4.2.2 項で述べた通り，一般の国語辞書などでは意味や用法を関連の強さに基づいて階層的に記述するという方法が試みられている。しかしこの方法では意味や用法どうしの関係が「関連の強さ」という一次元の量だけで測られているために，どのような種類の関係があるのかが全く明らかにならない。そればかりでなく，「関連の強さ」という尺度はその基準を明確にしづらく，人によってどれくらいの関連の度合で階層化させるかの判断にゆれが生じやすいという問題があった。

しかし，計算機で多義性を見通しよく扱う上では，その構造を明らかにすることが不可欠である。そこで，『IPAL名詞辞書』で導入した多義構造の記述方法を提案する[桑畑，本多 1996, Kuwahata et al. 1996, 柏野，本多 1998]。この提案は，比喩(メタファー，メトニミー，シネクドキー)に基づく意味の拡張関係に着目するというものである。もともと比喩とは修辞学の概念であり，文学的なテキストにおいて意外性や新奇性を出すために臨時的に用いられることが多いと考えられていた。しかしレイコフ，ジョンソン[1986]をきっかけとして，比喩がそれまで一般に考えられていた以上に日常言語に浸透していることが広く認識されるようになった。そしていわゆる認知言語学の研究者を中心に，多義構造を比喩に基づいて分析する試みが始められた [49]。それらの研究では典型例が示され，比喩が多義を支える原理であることについて述べられている。しかしながら，これまでそれを辞書記述に適用した例はなく，したがって実際にどれだけの語の多義がその原理で説明できるかについての定量的な研究はなされていなかった。そこで本論文では，『IPAL名詞辞

[49] その概要については山梨[1995]などを参照されたい。また，いわゆる認知言語学以前の研究としては，Ullmann[1969]，奥田[1967]，国広[1982]などがある。

書』に収録されている 560 の多義語全ての区分間の意味的関係を記述することで，定量的な分析を行い，記述方法の有効性を実証する。

まず，4.3.1 項で『IPAL 名詞辞書』に収録した語に多義がどれだけ存在するかを示す。続いて，4.3.2 項で，着目する 3 タイプの比喩(メタファー，メトニミー，シネクドキー)について説明する。そして，4.3.3 項で，比喩の 3 タイプ別に捉えられる多義構造について具体例を通し考察する。次に，4.3.4 項に辞書記述の手順を示し，4.3.5 項に本提案の記述の試行を述べる。

4.3.1 『IPAL 名詞辞書』における多義語

多義語の把握のために，『IPAL 名詞辞書』に収録されている語のうち，どのようなカテゴリの名詞がどれだけ多義であるかについて示す。

『IPAL 名詞辞書』では，語のもつ複数の意味・用法を意味的特徴や統語的特徴によって分類し，これを「区分」と呼んでいる。複数の区分をもつ語が多義語にあたる。収録されている名詞 1,081 語のうちの 560 語が，区分数が 2 以上の多義語である。その多義語には，区分数 2 のものから，最大 18 のもの(「め【目】」)まである。

最初に，収録されている名詞のカテゴリを表す名詞の分類を表 4.7 に示す。

表 4.7 名詞の分類とカテゴリ

分類 ()内は辞書記述で用いた分類記号	カテゴリ
＜ 文法的特徴に基づくもの＞	
サ変動詞用法のあるもの (a1, a2)	サ変
形容詞転成名詞 (b)	形容詞転成
形容動詞語幹といわれるもの (c)	形容動詞
関係性のあるもの(相対名詞といわれるものなど) (d5, d7)	相対
副詞用法のあるもの・自立性の低い名詞といわれるもの (d6)	副詞
「述語用法2 」のあるもの (e1, e2)	述語用法2
「連体被修飾用法2 」のあるもの (f)	被修飾2
動詞転成名詞 (g1)	動詞転成
＜ 意味的分類に基づくもの＞	
植物 (h2)	植物
食べ物 (h3)	食べ物
身体 (h4)	身体
動物 (h5)	動物
場所 (h6)	場所
建物・家具 (h7)	建物・家具
乗り物 (h9)	乗り物
洋装品 (hx)	洋装品
仕事・役割 (i1)	仕事・役割
自然物 (i5)	自然物
自然現象 (i6)	自然現象
組織・団体 (i7)	組織・団体

『IPAL 名詞

辞書』に収録されている語は，表 4.7 にあげた視点にたって選定されたものである。以下，この分類の省略名を名詞のカテゴリ名として用いることにする。各カテゴリに属する名詞がどれだけ多義であるかについて名詞の区分数を調べ，カテゴリごとに区分の平均数を計算した。その結果を表 4.8 に示す。

表 4.8 からは，「形容詞転成」「相対」「副詞」「身体」に属する名詞に，特に多義語が多いことがうかがえる。その次に，各カテゴリに属する名詞の例を表 4.9 に示す。( )は区分数である。各カテゴリに区分数の多いものから 1 例ずつ選び，末には区分数 1 の例をあげた。

表 4.8 カテゴリ別の「区分」の数

カテゴリ	区分（意味・用法）の数								平均
	1	2	3	4	5	6	7	8～	
サ変	74	50	15	7	1				1.7
形容詞転成	3	1	3	2	1	4		4	5.5
形容動詞	41	19	7	2		1			1.6
相対	6	15	5	7	9	1	2	12	4.8
副詞	5	4	4	3	1	2			2.8
述語用法2	75	27	13	3	2	1	3	1	1.8
被修飾2	59	40	10	1	3	2	2		1.8
動詞転成	8	9	3	1					1.9
植物	12	19	5	1			2		2.1
食べ物	13	10	3	2	1	1	1		2.2
身体	21	12	8	10	8	3	1	4	3.3
動物	16	16	8	3	2				2.1
場所	22	6	5	2	2			1	2.0
建物・家具	22	12	5	1					1.6
乗り物	17	5		1					1.3
洋装品	22	12	1	1	1				1.6
仕事・役割	40	22	7	6	1	1			1.8
自然物	26	10	7	5	1			1	2.0
自然現象	14	7	2	1	1		1		1.9
組織・団体	25	5	2		2				1.5
計	521	301	113	59	36	16	12	23	2.1

表 4.9 カテゴリ別の名詞の例

カテゴリ	名詞の例
サ変	展開(5), 支配(4), 営業(3), 開拓(2), 委託(1)
形容詞転成	重さ(13), 軽さ(12), 薄さ(11), 厚め(1)
形容動詞	きれい(6), 勝手(4), 鮮やか(3), 明らか(1)
相対	中(15), 上(14), かげ(13), 内(12), 形跡(1)
副詞	以上(6), 以下(5), 立場(4), 寸前(3), 有様(1)
述語用法2	筋(8), 色(7), 時間(6), 形(5), 味付け(1)
被修飾2	しるし(7), 癖(6), 心(5), 関係(4), 依頼(1)
動詞転成	走り(4), 付き合い(3), 生まれ(2), 急ぎ(1)
植物	種(7), 桜(4), 芋(3), 麻(2), 稲(1)
食べ物	茶(7), 水(6), 薬(5), クリーム(4), 牛乳(1)
身体	目(18), 頭(12), 手(10), 胸(8), 顔(7), 胃(1)
動物	羽(5), 魚(4), 生き物(3), 牛(2), おんどり(1)
場所	コース(8), 世界(5), 現場(4), 席(3), アジア(1)
建物・家具	舞台(4), 家(3), アパート(2), 居間(1)
乗り物	車(4), スキー(2), エレベーター(1)
洋装品	傘(5), クッション(4), リボン(3), 雨具(1)
仕事・役割	子(6), 役(5), 男(4), 主役(3), あなた(1)
自然物	日(9), 石(4), 足跡(3), 雨(2), 岩(1)
自然現象	火(7), 熱(5), 傷(4), 春(3), 霞(2), 火事(1)
組織・団体	国(5), グループ(3), 会(2), 会社(1)

4.3.2 比喩に基づく意味の拡張関係

意味の拡張関係を捉えるために着目する比喩には 3 タイプある。類似性によって意味を拡張する**メタファー(隠喩)**, 密接な関係にあるものや属性などの隣接関係によって意味を拡張する**メトニミー(換喩)**, カテゴリ全体とその一部という包摂関係によって意味を拡張する**シネクドキー(提喩)**という 3 タイプである。このとき, 隣接関係の一種である, いわゆる全体部分関係と, シネクドキーの包摂関係とは区別する。例えば,「て【手】」は,「着物の袖に手を通す」という時は, 肩より先の[全体](「て 01」[50])を指し,「手をたたいた」という時は手首より先の[部分](「て 02」)を指す。一方,「玉子(卵)を1ダース買った」という時の「たまご」は,「動物のメスが産む卵」のうち, 特に「鶏卵」を指す。この時, 両者の関係には違いがある。「手」の場合は,「て 02」以外の部分, 例えば「ひじ」などは「手」とは呼べない。しかしながら,「たまご」の場合は,「鶏卵」以外にも「金魚の卵, あひるの卵, かえるの卵」など「たまご」と呼べるものが複数ある。つまり,「手」の場合は, 物全体とその一部分との間に成り立つ全体部分関係であるが,「たまご」の場合は, カテゴリ間の包摂関係であるといっ

[50] 『IPAL 名詞辞書』では, 区分に 2 桁の通し番号をつけている。「て 01」という記法は, 見出し語「て」のうちの 1 番目の区分であることを示す。

た違いがある。よって，前者の関係をメトニミーの隣接関係の 1 つとし，後者の関係をシネクドキーとし，両者を区別して論じる [51]。

なお，拡張の方向性は，歴史的な派生順序にできるだけ沿うように捉える。即ち，語源的用法から順に解説している『日本国語大辞典』(小学館, 1976 年)(以下『日国』)によることを原則とする。しかし，IPAL の区分が『日国』の語義分けと一致しない場合や，『日国』の記述において派生順序が明確でない場合には，各区分の用例や意味を吟味して，拡張の方向性を判定する。また，外来語は，『日国』に記述がないものが多い。外来語の多義構造は，もとの言語の多義構造を保存する形で日本語に入ってきた場合と，日本語に入ってから語義が変化／拡張した場合とが考えられるが，どちらの場合に該当するか，また，もとの言語ではどういった語義の派生順序が認められるか，といった点は，厳密な確認が困難であるため，日本語としての用例分析によって拡張関係を捉えることとする。

4.3.3 考察

比喩の 3 タイプ別に具体例を示しながら，「比喩に基づく意味の拡張」と捉えられる多義構造について考察する [52]。そして，着目する比喩関係を表すための記号を定める。

4.3.3.1 メタファーに基づく拡張

メタファーと捉えられる用法の中には，明らかにメタファーであると意識されるものと，その用法が完全に慣習化しているために，メタファーであることがあまり意識されないものとがある。例えば，名詞「かた【肩】」には次のような用例がある。

(a) 後ろから肩をたたかれた。
(b) 洋服の肩がほつれる。
(c) その山の肩に，有名な山小屋がある。
(d) 文字の肩に印をつける。

---

[51] Lakoff and Johnson[1986]，山梨[1995]らはシネクドキーをメトニミーに含めている。佐藤[1992]は本論文と同様の区別をしている。

[52] 実際の辞書記述では例えば「01＝＞02」のように区分番号を記号でつないで記述している。本論文では「人の肩＝＞洋服の肩」というように，区分番号の代わりに各区分に記述されている用例を用いて例示する。

(a)は身体部位の「肩」を指す本来的な用法である。これに対して，(b)，(c)，(d)は「肩」という身体部位の物理的位置との類似から拡張してできたメタファーであると考えられるが，(b)は，(c)，(d)と比べるとメタファーであるという意識を持ちにくい。このとき，(b)は，「洋服の」という連体修飾語を除いて「肩がほつれる」といっても意味が通るが，(c)や(d)で同様に連体修飾語を除いて「肩に山小屋がある」，「肩に印をつける」といっては意味が通りにくくなるという違いがある。このことから，メタファーによって拡張したと捉えられる表現には，固定化の度合が強いものと弱いものとがあり，強いものは，連体修飾語を除いた表現も可能であることがわかる。よって，この統語的特徴を基準として固定化の度合で区別し，それぞれを「A ＝＞ B」，「A …＞ B」という記号で表す。どちらの場合も，左側の A から右側の B へ拡張することを表す。

(1) メタファー：連体修飾語が省けるほど固定化したもの

(記号) A ＝＞ B

(例) 肩：人の肩 ＝＞ 洋服の肩

皮：動物の皮 ＝＞ みかんの皮，まんじゅうの皮

こぶ：頭のこぶ ＝＞ らくだのこぶ，幹のこぶ

しわ：顔のしわ ＝＞ 服のしわ

クリーム：コーヒー用のクリーム ＝＞ 化粧品のクリーム

(2) メタファー：上記以外

(記号) A …＞ B

(例) 肩：人の肩 …＞ 山の肩，文字の肩

皮：動物の皮 …＞ 化けの皮

卵：動物の卵 …＞ 医者の卵

歯：赤ちゃんの歯 …＞ くしの歯

角：鹿の角 …＞ こんぺいとうのツノ

4.3.3.2 メトニミーに基づく拡張

メトニミーは隣接関係に基づいて意味を拡張するものである。その隣接関係にはいろいろなタイプがある。名詞「茶」は 7 つに区分されるが，それぞれがメトニミーの関係にある。以下に示す。

例:「ちゃ【茶】」

[区分 01] ツバキ科チャ節の常緑低木。

例:この畑にはたくさんの種類の茶がある。

[区分 02]「ちゃ 01」の葉を乾燥したり蒸したりして加工したもの。

例:静岡の実家で茶を生産している。

[区分 03]「ちゃ 02」に熱湯を注ぎ, こして入れたり, 粉末状の「ちゃ 02」に熱湯を注ぎ, 攪拌してたてる嗜好飲料。

例:食事の後の茶がおいしい。

[区分 04]「ちゃ 03」程度の軽い嗜好的な飲み物を飲みながら休憩すること。

例:さて一息ついた, お茶にしよう。

[区分 05]様々な嗜好的食べ物や, アルコールを含まない飲み物を味わい, 歓談する集まり。

例:叔母さんがお茶に招いてくれた。

[区分 06]「ちゃ 03」をたてて飲む作法。

例:茶をたしなむ。

[区分 07] 黒みがかった赤黄色。

例:靴とバックを茶に揃える。

このとき, [区分 01]から[区分 02]へ, また, [区分 02]から[区分 03]へは, どちらも［材料］と［加工品］の関係で, [区分 03]から[区分 04]へは［手段］と［行為］の関係で, [区分 04]から[区分 05]へは［行為］と［イベント］の関係で, 連続して拡張する。それから, [区分 02]から[区分 06]へは［手段］と［行為］の関係で, [区分 02]から[区分 07]へは［物］と［形状］の関係で拡張する。

その他, 4.3.2 項で名詞「手」を例にして述べた通り, 具体物の［全体］と［部分］の関係も隣接関係の 1 つである。全体部分関係による拡張は, 特に名詞の多義のうち顕著に見られるため, 全体部分関係とそれ以外の隣接関係とを区別して表示する。次の(3)と(4)に全体部分関係の例を示し, (5)にその他の隣接関係の例を示す。

(3)の「A ＞＞ B」, (4)の「A ＜＜ B」は, それぞれ「B が A の一部分である」「A が B の一部分である」という意味である。前者は全体から部分へ拡張する場合であり, 後者は逆に部分から全体へ拡張する場合である。(5)の「A［X］－＞［Y］B」はメタファーの場合と同様に, 左側の A から右側の B へ拡張することを表す。その時, どのような隣接関係に基づいているかを［X］, ［Y］に示す。

(3) メトニミー:全体部分関係1

(記号) A ＞＞ B

(例)　手:袖に手を通す ＞＞ 手 (hand) をたたく

足:足が長い ＞＞ 足 (foot) が大きい

貝:貝を掘りにいく ＞＞ 貝 (貝殻) でできた首飾り

(4) メトニミー:全体部分関係2

(記号) A ＜＜ B

(例)　花:花が開いた ＜＜ 花を 10 本買う

首:首 (neck) が細い ＜＜ 首 (head) をとる

(5) メトニミー:上記以外の隣接関係

(記号) A[X]－＞[Y]B

(例)　鍋:鍋で煮る［容器］ －＞［中身］鍋を食べる

肩:肩がこる［身体部位］ －＞［機能］いい肩をしている

リボン:リボンで結ぶ［物］－＞［形状］リボンに結ぶ

豚:豚を飼う［材料］ －＞［加工品］豚を 100 グラム買う

上記の[X]と[Y]は，A，B における当該名詞の用法そのものを記述したものではなく，あくまでもメトニミーを支える隣接関係がどのようなものであるかを記述するためのものである。上にあげた「豚を飼う[材料] －＞[加工品]豚を 100 グラム買う」という例で説明すれば，これは「豚を飼う」などと言う時の動物としての「豚」に「材料」を指す意味があるということを言っているわけではない。「豚を 100 グラム買う」と言うような，「豚」で「豚肉」を指す用法と見比べてみた場合に，動物としての「豚」の方をまず「材料」として捉え直すことができ，また，それをカットあるいはミンチした結果の「豚」の方を「加工品」として捉えることができるということを言っている。つまり，この時に両者の用法の間にあるメトニミーの拡張関係が「[材料] －＞ [加工品]」という種類の隣接関係によるものであることを示している。

4.3.3.3 メトニミーの隣接関係

メトニミーには全体部分関係以外にも隣接関係がある。幾つかよくあるパターンは知られている[Lakoff and Johnson 1986, Ostler and Atkins 1991, 山梨 1995]が，典型例以外のメトニミーを論じた研究はほとんどない。そこで『IPAL 名詞辞書』で捉えたメトニミーの拡張関係全体を見渡し，どのような隣接関係を取り上げたかについて述べる。『IPAL 名詞辞書』で全体部分関係以外に取り上げた隣接関係の種類は，実に 124 通りもあった。そのうち出現数 2 以上の隣接関係が 31 通りであった。その 31 通りの内訳を表 4.10 に示す。

表 4.10 で出現数 3 以上のところを見ると，いわゆる典型的な隣接関係が並んでいる。しかしながらそれら出現数が多いもののパターンは限られており，出現数 3 以上のものは 13 通りしかない。メトニミーの隣接関係について述べる時だれもが似たような典型例でしか説明しないのは，多く現れる隣接関係のパターンがそもそも限られていることによるものだと考えられる。

560 の多義語を分析した結果，全部で 126 通りもの隣接関係が認められ，そのほとんどは出現数が 1 であった [53]。これは，対象語を増やせば増やすほど隣接関係のパターンが増えていってしまうことを示している。つまり，このことからあらかじめ隣接関係のパターン全てを列挙して示すことは非常に困難であるということがわかる。これは人間の連想能力の範囲を確定する制約が発見されていない(あるいは存在しない)ことを示すものである。

53 出現数 1 のものには，例えば，「医者:医者に診てもらった［ある能力を持った人］—>［その人が中心的な役割を果たす機関］医者に行った」「日:日が沈む［物］—>［それが見える時間］日が長い」といったものがある。

表 4.10　隣接関係の例

隣接関係	出現数
[材料] —> [加工品]	17
例　木:子供が木に登っている —> 木で作った家に住んでいる	
[身体部位] —> [機能]	14
例　腕:あの人は腕がとても細い —> ゴルフの腕が上がる	
[手段] —> [行為]	8
例　口:飴玉を口に入れる —> 彼は口が達者だ	
[結果] —> [原因]	6
例　健全:彼の精神は健全だ —> この読み物は健全だ	
[物] —> [性質]	5
例　人:世の中にはいろいろな人がいる —> 彼は人が変わったようだ	
[方向] —> [位置]	4
例　北:部屋の窓は北に面している —> 北に住む人は寒さに強い	
[物] —> [形状]	4
例　顔:石けんで顔を洗う —> 強盗犯の顔が頭に焼きつく	
[原因] —> [結果]	4
例　大人:大人が子供達の面倒をみる —> もっと大人になれ	
[面] —> [方向]	3
例　正面:店の正面が海に面している —> 正面から風を受ける	
[方向] —> [収束先]	3
例　内側:ドアが内側に開く —> カーブを曲がる時体を内側に傾ける	
[物] —> [それが原因となる現象]	3
例　太陽:太陽が雲に隠れた —> 太陽がグラウンドに照りつけている	
[場所] —> [機関]	3
例　他:本棚を他に移す —> 他ではもっと安く売っている	
[行為] —> [行為者]	3
例　教授:華道の教授をする —> 教授に相談する	
[器官] —> [その使用]	3
例　目:他人の目を気にする —> 暖かい目で見守る	
[容器] —> [中身]	2
例　鼻／洟:鼻が赤い —> 洟をすする	
[面] —> [そこに付けるもの]	2
例　裏:写真の裏を見る —> 夏服だから裏はいらない	
[方向] —> [面]	2
例　外:このドアは外には開かない —> 箱の外を赤く塗る	
[方向] —> [場所]	2
例　背中:子供の背中をさする —> 背中で声がする	
[物] —> [色]	2
例　金:この指輪は金でできている —> 背景を金の絵の具でぬった	
[物] —> [場所]	2
例　湯:ガスコンロで湯を沸かす —> 近くの湯に行って来る	
[物] —> [機能]	2
例　クッション:ソファーにクッションを並べる —> この椅子はクッションが悪い	
[物] —> [それが作り出す周期]	2
例　月:月が満ち欠けする —> 会合のある月を尋ねる	
[物] —> [その使用]	2
例　塩:天日で塩を作る —> 魚に軽く塩をする	
[点] —> [線]	2
例　中央:部屋の中央に集まる —> 市の中央に川が流れている	
[出来事] —> [それをとりまく社会的な環境]	2
例　生まれ:東京で生まれた —> 彼女は生まれがよい	
[事柄] —> [それを対象とする学科]	2
例　社会:社会で通用する —> 社会が得意だった	
[作成] —> [作成物]	2
例　施設:体育館の施設が認可された —> 区の施設へ出掛ける	
[空間] —> [面]	2
例　内側:山の手線の内側に住む —> 箱の内側にアルミ箔を貼る	
[関係] —> [それを構成する項]	2
例　同志:恋人同志になる —> 同志が集まる	
[運搬システム] —> [運搬するもの]	2
例　地下鉄:地下鉄が通る —> 地下鉄が到着する	
[位置] —> [方向]	2
例　外:鉢植えは窓の外だ —> つま先を外に向ける	

4.3.3.4 シネクドキーに基づく拡張

シネクドキーの例を以下に示す。「A ⊃ B」という表示は「B は A に包摂される」, ないし「B は A というカテゴリーの成員である」という意味である。カテゴリ全体でその一部分を指す拡張の場合を「A ⊃ B」で示し, ある一部分でカテゴリ全体を指す拡張の場合を「A ⊂ B」で示す。

(6) シネクドキー1

(記号) A ⊃ B

(例) 卵:動物の卵 ⊃ 卵 (鶏卵) を1ダース買う

毛:体が毛でおおわれている ⊃ 毛(頭髪)を染める

毛:体が毛でおおわれている ⊃ 毛(羊毛)100%のセーター

粉:チョークが砕けて粉になった ⊃ 粉 (小麦粉) をこねる

天気:天気が変わりやすい ⊃ 今日は天気 (晴れ) だ

着物:着物 (衣服) を揃える ⊃ 着物 (和服) を着付ける

(7) シネクドキー2

(記号) A ⊂ B

(例) 酒:熱燗の酒(日本酒) ⊂ 世界各国の酒(アルコール飲料)

魚:白身の魚(食用) ⊂ 海の魚(魚類である水生動物)

上にあげた例のほか, 尺度や程度を表す名詞の区分間にも, シネクドキーの拡張関係を捉えることができる。例えば, 名詞「けんこう【健康】」は, 以下のように 3 つに区分される。

例:「けんこう【健康】」

[区分 01] (異常があるか, ないかという面から見た)心身の状態。

例:最近健康がすぐれません。

[区分 02] どこにも悪いところがなくて, 元気であること。

例:毎日を健康にすごす。

[区分 03] (あるものが)心身に悪影響を及ぼさないようなものであるさま。

例:この書物は極めて健康な読み物だと言える。

[区分 02]から[区分 03]へは，［結果］と［原因］というメトニミーの関係で拡張していると捉えられるので，ここでは問題にしない。ここでは，[区分 01]から[区分 02]への関係に着目する。[区分 01]は，心身の状態が良い場合も悪い場合も含むのに対し，[区分 02]は，心身の状態が良い場合だけに範囲を特定するものである。注目すべきことは，この時，「悪い場合」でなく，「良い場合」に範囲が特定されるという点である。もともと[区分 01]は，良い状態に重点を置いて，良いか悪いかを問題にするものであり，[区分 02]は，その重点が置かれている方へ範囲が特定されるものである。このように重点が置かれている方へ範囲が特定されるタイプもシネクドキーの 1 つと考える[54]。このような関係は「健康」のほかに，「奥行き」「スピード」など，尺度や程度を指す用法のある名詞の区分間に複数認められるため，(6)や(7)で示したシネクドキーとは区別して「尺度型」と呼ぶ。程度が強まって範囲が特定される関係を「尺度型1」として，「 A ⊃＋ B 」で示す。以下に「健康」以外の例を示す。

(8) シネクドキー：尺度型1

(記号) A ⊃＋ B

(例) 奥行き：この押入れの奥行きは浅い ⊃＋ この押し入れは奥行きがある

スピード：スピードが遅くなる ⊃＋ スピードがある

数：リンゴの数を数える ⊃＋ 数をこなす

女：救助の時は女が優先される ⊃＋ そのドレスは女を感じさせる

上に示した「尺度型1」は，範囲が特定される方向に拡張されるものであるが，その逆の向きの拡張関係が，尺度や程度を表す，形容詞からの転成名詞に認められる。例として転成名詞「広さ」をあげる。名詞「ひろさ【広さ】」は5つの区分をもつが，そのうちの[区分 01]と[区分 02]の意味記述と用例を示す。

例：「ひろさ【広さ】」

[区分 01] 幅や面積などの大きさの度合い。

例：グラウンドの広さを測る。

---

54 国広[1982]では，「体重のある人」は＜普通より重い人＞というプラス値を指す，といった例をあげて，「実力，人格，人物，距離，効率」などには「プラス値多義」という派生関係があると説明している。また，国広[1997]では，これを「プラス値派生義」と呼び，「トートロジーによる意味派生」と説明している。

[区分 02] 大きな幅や面積。

例:案内された部屋の広さに圧倒された。

「広さ」という名詞は，形容詞「広い」からの転成名詞であり，もともと重点は「広いこと」に置かれている。その「広いこと」を指すのが[区分 02]である。その程度を弱め，「広いか狭いか」という度合を指すまで範囲を広げるものが[区分 01]である。同じ関係が，「重さ」「厚さ」などにも認められる(「狭さ」「軽さ」「薄さ」などには，度合を指す用法はないため，この拡張関係はない)。

このように，程度が弱まって範囲が広がる関係を「尺度型2」として，「A＋⊂ B 」で示す。以下に，「広さ」以外の例を示す。

(9) シネクドキー:尺度型2

(記号) A ＋⊂ B

(例) 重さ:コンピュータの重さに耐えられるような頑丈な机を買った ＋⊂ 小包の重さを測る

厚さ:彼の胸の厚さにみんなが驚いた ＋⊂ ファイルの厚さを尋ねる

形容詞転成名詞の場合，もとの形容詞の意味に「目安ないし基準点を超えている」という面があるので，それと直接対応のとれる意味から，「目安を超過」という範囲特定が弱くなった意味が派生したと考えられる。一方，「尺度型1」の例で示した「健康」「奥行き」「スピード」といった名詞は，度合を指す意味がもともとあり，その意味から，重点のかかる方の程度が強まった意味が派生したと考えられる。

4.3.3.5 融合的な拡張

中には，メタファー，メトニミー，シネクドキーのうちのどれか 1 つのタイプに特定することが難しいと感じられる融合的な拡張例がある。山梨[1995]にはメタファー的な性質とメトニミー的な性質のいずれをもかね備えた融合的イディオムの例があげられている(例えば，「足を引っ張る」は，足が引っ張られれば体全体の動きも阻止されることにつながる点ではメトニミー的であり，この物理的な阻害が[邪魔される]という意味に拡張している点ではメタファー的であるといえる[山梨 1995])。

しかし，いずれかのタイプにできるだけ特定して記述する。例えば，山梨[1995]が融合的イディオムの例としてあげている「手に入る」や「手を引く」については，前者はメトニミー，後者はメタファーと特定する。「手に入る」や「手に入れる」という時の「手」の用法は，「物を手に取る」という時の

「手」の用法からの[保持]と[所有]という隣接関係によるメトニミーであると捉える。また,「手を引く」や「手を出す」という時の「手」の用法は, イベントへの関与を具体的なものとの関わりになぞらえたメタファーであると捉える。

4.3.3.6 区分外の意味概念を用いる場合

区分されている結果どうしを比べるだけでは区分間にある関係が捉えにくい場合がある。そのような場合, 区分外の意味概念を用いて記述するという方法をとる。「区分外の意味概念」とは, その名詞のもつ,『IPAL 名詞辞書』において独立した区分が与えられていない意味・用法を表す概念のことである。例えば名詞「ひな」は次のように 2 つに区分される。

例:「ひな」

[区分 01] 卵からかえって間もない鳥の子。

例:昨日ひなが 5 羽かえった。

[区分 02] ひな人形。

例:三月三日には女の子のために雛を飾ります。

この時, [区分 01]と[区分 02]とを比べるだけでは両者の間にある意味的関係が見えてこない。しかし, 区分外の用例として記述してある「雛形」や「雛菊」などの合成語を見ると,「ひな」には「小さなもの」を指す用法もあることがわかる。[区分 01]の「鳥の子」が小さいことから「小さなもの」を指す用法が派生し, さらに「人形の小さなもの」を指す[区分 02]の用法が派生したと考えれば, 両者の間にある意味的関係が見えてくる。このような時, 区分外の意味概念を立てて区分との関係を記述すれば(区分外の意味概念は{}でくくって表示する),「ひな」の場合は「01 ⊂ {小さなもの} ⊃ 02。」と書き表すことができ, その意味的関係がより明らかに説明できるようになる。

区分外の意味概念は恣意的に立てられるべきものではない。それを立てないと区分間の関係が捉えられない場合にのみ, 区分外の意味概念を立てる。例えば「ひな」の例のように, 区分として取り上げなかったまでもその意味概念をもつ用例が実際にあるものや, 次に示す「きかん」のように, かつてそのような意味をもって用いられたことがあるもののうちから立てる。

例:「きかん【機関】」

[区分 01] 国家や法人, 団体などが特定の事柄に対応するために設けた組織。

例:この事業は既存の機関には対応できない。

[区分 02] 火力・電力などで機械を動かす装置。

例:大型の機関では電源に畜電池を用いたものが多い。

＜関係＞{ある働きをするためのしかけのある装置} ＝＞ 01, 02。

＜参考＞『日本国語大辞典』(小学館)に,「1活動の装置を施したもの。しかけのある機械。からくり。」とある。

あるいは, 次の「テーマ」のように, 名詞のもつ意味から抽象化したものを区分外の意味概念として立てる。

例:「テーマ」

[区分 01] 芸術作品の, その作品によって表現されていることの中心をなすもの。

例:この本は人間社会の不条理をテーマとしている。

[区分 02] 論文や議論の主題。

例:卒業論文のテーマを変える。

＜関係＞ {主な題目} ⊃ 01, 02。

(01 は芸術作品に, 02 は論文や議論に対象を特定。)

ここで留意すべきことは, 1 つの大きな抽象化した意味概念を立ててシネクドキーとして記述することが名詞のほとんどについて可能であるということである。例えば, 4.3.3.1 目にメタファーの例としてあげた「皮:動物の皮 ＝＞ みかんの皮, まんじゅうの皮」は, 外面や表面を覆っている膜状のものという意味概念を立ててシネクドキーで捉えることも考えられる。しかし, 先にも述べた通り, 意味概念を立てるのは, それを立てないと区分間につながりがあるということが捉えられない場合のみである。「皮」のように区分間のつながりの具体的なあり方が, 区分どうしを見るだけで明確に捉えられる場合は, そのように捉えることを最優先にする。

4.3.3.7 その他の意味的関係

これまでに述べた比喩による意味の拡張は, 区分間の意味的な関係を指示対象間の関係として捉えるものである。この方針では, ほぼ同一の指示対象を持ちながら, それに対する話者の捉え方が異なるために別区分として認定されている場合や, 異なる指示対象に対して同一の捉え方

を適用していることが単一語の多義という形をとって現れる場合を適切に扱うことができない。そのような比喩によって捉えきれない多義については，できるだけパターン化した文章でその多義がどう捉えられるかを記述する [55]。例えば「人」には，[区分 01]「哺乳類霊長類目ヒト科に属する生物」(例：原始時代の人の化石が発見された)という用法(「ひと 01」)と，[区分 02]「高等な知識を備え，文化を生み，社会を形成する存在としての「ひと 01」」(例：世の中には色々な人がいる)という用法(「ひと 02」)とがある。これについては，「01 は 02 を特定の見方(生物学的見方)で捉えているもの」と記述する。

4.3.4 多義構造の辞書記述の手順

以上の考察を踏まえ，多義語を下位区分した区分どうしの関係を「多義構造」と呼び，その明示的記述方法を提案する。

まず，メタファー，メトニミー，シネクドキーで捉える意味的関係と用いる記号について，4.3.4.1 目から 4.3.4.4 目まで述べたものを表 4.11 にまとめて示す。

表 4.11　意味的関係の種類と記号

意味的関係	記号
(1)メタファー：連体修飾語が省けるほど固定化したもの	A ＝＞ B
(2)メタファー：(1)以外	A ...＞ B
(3)メトニミー：全体部分関係1	A ＞＞ B
(4)メトニミー：全体部分関係2	A ＜＜ B
(5)メトニミー：(3)(4)以外の隣接関係	A[X]－＞[Y]B
(6)シネクドキー1	A ⊃ B
(7)シネクドキー2	A ⊂ B
(8)シネクドキー：尺度型1	A ⊃＋ B
(9)シネクドキー：尺度型2	A ＋⊂ B

[55] くわしくは本多・桑畑[1997]。

表 4.11 で示す意味的関係の記号を辞書記述に用いる。「多義構造」の記述の手順は次の通りである。

1. 区分間の関係を，類似性によって意味を拡張する**メタファー(隠喩)**，密接な関係にあるものや属性などの隣接関係によって意味を拡張する**メトニミー(換喩)**，カテゴリ全体とその一部という包摂関係によって意味を拡張する**シネクドキー(提喩)**のいずれで捉えられるか判定する。
2. 該当する場合は，表 4.11 の(1)～(9)いずれであるか判定したうえで，その記号と，区分番号とを用いて関係を記述する。
3. 拡張の方向性は，歴史的な派生順序にできるだけ沿うように与える。即ち，語源的用法から順に解説している『日本国語大辞典』(小学館，1976 年)(以下『日国』)によることを原則とする。
4. 可能な限り単純明瞭に記述するために，3 タイプの比喩の融合と捉えられる場合も，いずれのタイプにできるだけ特定して記述する。
5. 比喩によって捉えきれない多義については，できるだけパターン化した文章でその多義がどう捉えられるかを記述する。

例えば，「のど」は 4 つに区分される。この区分間の意味的関係は[区分 02]と[区分 03]が[区分 01]のメトニミーであり，[区分 04]が[区分 02]のメタファーであるので，次のように記述する。

例：「のど」

[区分 01] 管の形をしている口の奥の部分。

例：のどに魚の骨がひっかかってしまった。

[区分 02] 頭と肩をつなぐ身体部分の前面。

例：喉に湿布をする。

[区分 03] 歌う声。

例：彼女の 自慢の喉を聞かせてもらいましょう。

[区分 04] 本の綴じ目の部分。

例：本ののどがこわれる。

＜関係＞ 01［物］ －＞ ［側面］02。 01［手段］ －＞ ［行為］03。 02 …＞ 04。

「あし【足，脚】」は，[区分 02]は[区分 01]と全体部分関係のメトニミーである。[区分 03]は[区分 01]と［身体部位］と［機能］の関係のメトニミーであり，[区分 04]は，その[区分 03]からの機能の類似性によるメタファーである。また、[区分 05]は[区分 01]からの形態的な類似性によるメタファーであるので，次のように記述する。

例：「あし【足，脚】」

[区分 01] 動物の胴に付属し、歩行や体を支える機能を果たす身体部分。

例：彼女は足が長い。

[区分 02] 「あし 01 」の一部で、くるぶしより下の部分。

例：彼は足が大きい。

[区分 03] 歩いたり、走ったりすること。

例：彼は足が速い。

[区分 04] 移動の手段となるもの。

例： 台風で市民の足が乱れる。

[区分 05] 椅子などの下部で、台の部分や上にある物を支える役割を果たす部分。

例： 机の脚が折れる。

＜関係＞ 01 ＞＞ 02。 01［身体部位］ －＞［機能］03 …＞ 04。 01 …＞ 05。

メタファー，メトニミー，シネクドキーでは説明できない拡張関係の例を最後に示す。

例：「けっさく【傑作】」

[区分 01] きわめてすぐれた作品。

例： 彼は画家として後世にたくさんの傑作を残した。

[区分 02] (思わず笑いたくなるほど)言動などがひどく滑稽なこと。

例： あいつがまた傑作をしでかした。

＜関係＞ 02 は，01 がアイロニーで使われたもの。

4.3.5 記述の試行

「**要件 1-3** 多義語の意味関係の構造が客観的・明示的であること」を満たす方法として，多義構造を比喩による意味の拡張という観点から捉え，明示的に記述する方法を提案した。実際にそ

の原理を用いて辞書に収録する多義語の説明ができるか，記述を試行する。

4.3.5.1 記述方法

『IPAL 名詞辞書』に収録されている 560 の多義語全ての区分間の意味的関係を記述することで，記述方法の有効性を確認する。

4.3.5.2 記述結果

1 つの「区分」，あるいは，1 つの「区分外の意味概念」どうしの関係を 1 と数えていくと，多義語 560 語の多義構造として記述した意味的関係の数は全部で 1,364 あった。

合計 1,364 の関係のうち，(1)～(9)の比喩に基づく意味的関係で捉えることのできた関係の内訳を，名詞のカテゴリ別に表 4.12 に示す。また比喩的関係だけでは捉えきれなかった関係の数を「(他)」として記す。

表 4.12 各カテゴリにおける意味的関係の数

カテゴリ	メタファー		メトニミー			シネクドキー				(他)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
サ変	53	18	3		7	34				8
形容詞転成	67				1	10		4	3	
形容動詞	12	1			8	21		2		12
相対	75	42			38	24	1			49
副詞	19	2			5	8		1		3
述語用法2	23	27		1	12	21		15		14
被修飾2	24	13		1	7	52		1		19
動詞転成	5	5			5	8				
植物		22	11	2	6	4				
食べ物	6	4			16	6	2			4
身体	25	56	10	4	47	2				17
動物	1	14	3	2	9	9	2			11
場所	5	9	2	1	5	19	1			3
建物・家具	1	12	1	1	6	3				3
乗り物		2		2	3	1				
洋装品	4	11			4	3				3
仕事・役割	15	9			12	20				22
自然物	3	15	2	1	18	5				12
自然現象	5	8			2	5		1		6
組織・団体	2	2			4	7				4
小計	345	272	32	15	215	262	6	24	3	190
合計	617		262			295				190

まず，比喩的関係として捉えた(1)〜(9)と，それ以外の(他)の合計を比較してみる。(1)〜(9)の合計が 1,174 であり，(他)の合計は 190 である。つまり，全体の約 86% を比喩的関係として捉えることができたということである。このことは，本論文のアプローチが多義構造を記述するのに有効であったことを示すものである。

続いて，関係ごとに合計の数字を比較してみる。3 つのタイプ別に合計すると，メタファーが 617，メトニミーが 262，シネクドキーが 295 になる。メタファーが全体のおよそ半数を占め，メトニミーとシネクドキーが残りを分け合う形になっている。これは，この 3 つの比喩のうちいずれを無視しても名詞の多義が適切に捉えられないことを示すと考えられる。

次に比喩の方向性に着目すると，興味深い非対称性が見られる。シネクドキーの場合，指示対象が狭まる方向の(6)が広がる方向の(7)に比べて圧倒的に多い。このような非対称性は，同じシネクドキーの尺度型の(8)と(9)や，メトニミーの全体部分関係の(3)と(4)にも指摘することができる。

また，表を名詞のカテゴリ別に見ると，機能する比喩のタイプが名詞のカテゴリによって多少違っていることがわかる。例えば，「形容詞」はメタファーが圧倒的に多い。「サ変」や「動詞転成」はメトニミーが少なく，他 2 つの関係が多い。一方，「身体」「自然物」はシネクドキーが少なく，他 2 つの関係が多い。さらに，メタファーが特に少ないというカテゴリがないのに対して，(8)や(9)の尺度型が関係する名詞のカテゴリは限られていることも，特徴的な点としてあげることができよう。

本節で提案する多義構造の記述方法に基づいて実際に辞書記述を行った結果，収録した語のもつ多義構造の約 86% を記述することができた。さらに，その記述を分析することによって，多義構造の特徴を定量的に明らかにできた。

4.3.5.3 記述結果のまとめ

多義語 560 語の多義構造として記述した意味的関係の数は全部で 1,364 あり，そのうちの約 86% を提案手法で記述できることを示した。多義語を異なる意味の単なる寄せ集めとして扱うのではなく，構造をもったものとして辞書に体系的に記述することにより，計算機による意味処理を系統的に進めることができると期待できる。即ち，語の解釈の候補となる意味の間にある多義構造を特定することによって，多義性解消に必要な本質的情報を絞り込むことができる。

例えば，名詞「クリーム」には，「食物」を指す意味と「人の肌や靴などの塗るもの」を指す意味とがある。本論文で提案した方法によれば，これらの意味の間に，可食物から非可食物へというメタファーに基づく関係があることが辞書記述からわかるため，可食物か非可食物かが，この多義性を解消するための本質的情報であることがわかる。メトニミーやシネクドキーの場合も同様であり，

メトニミーの場合には，語の指示対象を捉える観点の異なりが，またシネクドキーの場合には，語の指示対象を限定する範囲が本質的情報であることがわかる。従来は，多義性解消のための方法として，一律に，格パターンや用例に基づく類推が適用されていたが[Kurohashi 1993，梁 1996]，提案法によれば，本質的情報の絞り込みが可能であるために，従来法よりも見通しよく多義性解消を行える可能性があると考えられる。

第5章 コーパスに基づく辞書記述内容の精緻化

本章では，コーパスから得られる情報を活用し，辞書の記述内容そのものを精緻化する方法を提案する。5.1 節では，コーパスから網羅的に抽出したコロケーション(共起情報)の記述における多義性という問題を取り上げ，多義性のあるコロケーションの辞書記述を新たに提案する。5.2 節では，使用域の記述を提案する。そのために，まずは書籍という同一のレジスタの中に存在する書籍サンプルの多様な文体的特徴を捉えることを目的に，BCCWJ に収録されている書籍サンプルを対象にして，文体分類をアノテーションする方法を論じる。そのうえで，「古風な語」を対象に，その用例の使用域を明らかにし，それを辞書記述に取り込む方法を提案する。5.3 節では，外来語を例に，時系列の調査が可能な新聞記事データベースより頻度情報を得て，時間的使用推移を明らかにし，それを辞書記述に取り込む方法を提案する。

5.1 多義性解消のためのコロケーションの詳述

本節では，コーパスから抽出したコロケーション(共起情報)の記述における多義性という問題を取り上げ，多義性のあるコロケーションの辞書記述を新たに提案する。

語と語の結び付きに関する詳細な辞書情報は有用である。例えば，自然言語理解において問題となる多義性の解消に対して有効な情報源となることが期待される。また，生成時にも適切な組み合わせを選択するための情報源になる。しかし，これまでは，各辞書に記載されているどの情報をどのように利用すれば多義性を解消して解析できるか，あるいは，どのような情報を用いれば多義性を生じさせない表現を生成できるかという道筋は必ずしも明らかではなかった。そこで本節では，文の多義性をもたらす要因の 1 つである「コロケーションの多義性」に着目し，多義性解消のためのコロケーションの辞書記述方法を提案する。コロケーションとは，一般に，一文中に複数の語が密接に結び付いて現れる関係をいうが，本論文では，「名詞句＋格助詞＋述語」の密接な結び付きをコロケーションと呼ぶ。

まず，5.1.1 項でコロケーションの多義性について概説する。続けて，5.1.2 項で『IPAL 名詞辞書』におけるコロケーションの重複記載を分析する。次に，5.1.3 項で重複するコロケーションを区別する情報について考察する。5.1.4 項で辞書記述手順を示し，最後に 5.1.5 項で辞書記述による多義性の解消について確認する。

5.1.1 コロケーションの多義性

従来，多義性解消の試みは，主として語の多義性が着目されてきたが，語と語とが結び付いたコロケーションの多義性という観点からの検討も重要である。これを，梁[1996]の研究を例にあげて説明する。

梁は『IPAL 動詞辞書』，『IPAL 形容詞辞書』，『IPAL 名詞辞書』を統合して利用するために，辞書の用例中に現れる語と各辞書の見出し語間に語義レベルのリンクを張った。このとき作成されたリンク情報は，例えば「鍋＜ 01 ＞が割れる＜ 01 ＞ 」といったものである(＜＞内の数字は『IPAL』において見出し語の意味，用法を区分する区分番号を表す)。「鍋が割れる」という場合，「鍋」という名詞には，器を指す意味＜ 01 ＞ と，中身の料理を指す意味＜ 02 ＞ とがあり，また「割れる」という動詞には，物が壊れる意味＜ 01 ＞ と，意見などがいくつかに分かれる意味＜ 02 ＞ とがある。両方の語とも多義であるが，梁が与えたリンク情報は多義性の解消が図られ，語の区分番号が示されている。即ちその「鍋」は器を指すものであり，その「割れる」は物が壊れる意味のものであることがリンク情報によって判別できるようになっている。このように語の多義性が解消されていれば，それを利用して計算機による意味の特定が容易になる。ところが，梁のリンク情報は各々の名詞と述語(動詞／形容詞)のみに着目するものであったために，名詞と述語の組合せであるコロケーションに多義性があるとき，そのコロケーションの多義性を排除しきれていないという問題があった。例えば，「骨が折れる」という場合には，「骨」という語に人間の骨を指す意味と，傘などの骨を指す意味とがあり，さらに句全体で慣用的に「苦労する」という意味を指す場合もあるため，名詞と述語とを結び付けるリンク情報だけでは，人の骨が折れるのか，傘が壊れるのか，また苦労して何かをするのか，いずれの意味にも特定することができない。「骨が折れる」のような場合，各語にある多義性の解消に先立ち，「骨が折れる」というコロケーション自体にある多義性を解消する方法を検討する必要がある。

『IPAL 名詞辞書』において「骨が折れる」のように多義性を持つコロケーションは，重複して記載されている。そこで，コロケーションの多義性を解消するための辞書情報を，同じ名詞と述語の組

合せが重複して記載されている例の分析を通して検証した。次の 5.1.2 項にその具体的な内容を示す。

5.1.2 コロケーションの重複の分析

5.1.2.1 コロケーションの重複記載の例

『IPAL名詞辞書』には，コロケーションに関する情報として，見出し語に結び付く述語だけではなく，述語のとる文型や，見出し語が名詞句を形成する場合の，名詞句内の先行句や後行句などの情報も記載してある[井口ほか 1996]。『IPAL名詞辞書』に記載されているコロケーションは延べ43,176例，慣用表現は延べ1,168例である。重複するコロケーションを抽出した結果，重複例は延べで10,415例，異なりで4,578例であった [56]。平均して2〜3個重複しているという計算になる。

重複記載の例として，以下に見出し語「つの【角】」の例を示す。「つの【角】」は，「[区分01]動物の頭部に生えている，固く尖ったもの。例:雄鹿が角を突き合わせて戦う。」と，「[区分 02]ものの突き出ている部分。例:かたつむりがツノを出す。」との 2 つに区分される。『IPAL名詞辞書』では，以下の例に示すようにコロケーションに関する情報を意味素性 [57]別に収録している [58]。

[区分 01] [項の用法 1/2] CON

ニ	角ガ	ある，ない。
	角ガ	曲がる，折れる。
ガ(鹿の／…)	角ヲ	切る，折る，曲げる。
ガ	角ヲ	出す。
ガ，ト	角ヲ	突き合わせる。
ガ，ヲ	角デ	突く，刺す。
ガ，ヲ(動物の／—)	角デ	作る。
	角ガ	大きい，小さい，長い，短い。
	角ガ	立派だ。

[56] 抽出データは，1997 年 2 月 14 日時点の『IPAL 名詞辞書』(正式公開前の版)を利用している。

[57] 意味素性とは，結び付く述語によって焦点が当てられる名詞の側面をいう。『IPAL 名詞辞書』では英文字 3 字で略して表す。例えば，〈CON〉は〈concrete〉の略で物質としての側面を持つものという意味素性であり，〈PHE〉は〈phenomenon〉の略で自然現象，天候現象，生理現象のうち，その発生/変化/消滅に関する意味素性である.詳しくは青山[1996]を参照されたい。

[58] 「/…」という記号は，他にも同様の連体修飾表現がありうることを示す。

[区分 01] [項の用法 2/2] PHE

角ガ　　生える，伸びる，抜ける。

角ヲ　　生やす。

[区分 02] [項の用法 1/1] CON

(その虫)ニ　　角ガ　　ある，ない。

(こんぺいとうの／…)角ガ　欠ける。

(かたつむり)ガ　角ヲ　　出す，隠す，引っ込める。

また，区分以外の用法として，《慣用表現》欄には以下のように記述されている。

[慣用句等]

角 (を) 突き合わせる(=仲が悪く，争い対立する)。

角 を折る (=自分の意志を強く主張することを控え，協調的な態度をとる)。

角 を出す (=嫉妬する)

角 {が生える／を生やす}(=嫉妬する)

角 を矯めて牛を殺す

以上の記述のうち，「角ガある」「角ガない」は区分 01 と区分 02 とに重複し，「角が生える」「角ヲ生やす」は，区分 01 と慣用表現とに重複している。また「角ヲ出す」は，区分 01，区分 02，慣用表現にわたって重複している。なお，区分 02 には，「他の格助詞」と「他の格に立つ名詞句」の例として，「(その虫)二」や「(かたつむり)ガ」といった，区分 01 や慣用表現とは区別される記述がある。

コロケーションの多義性解消を考えるうえでは，どのような仕組みで多義性が生じているかを明らかにすることが必須である。そこで，上記のような辞書記述から同じ名詞と述語の組合せの重複記載をすべて抽出し，それらの重複記載のパターンを分析してその要因を考察した。これらを 5.1.2.2 目と 5.1.2.3 目に示す。

5.1.2.2 コロケーションの重複記載のパターン

『IPAL 名詞辞書』では，コロケーションは見出し語の区分ごとに，意味素性別に記載されているので重複記載のパターンは，区分が異なる場合と同じ場合の 2 通り，素性が異なる場合と同じ

場合の 2 通りがあり，計 4 通りのパターンがある。表 5.1 に，これら重複記載の 4 パターンと出現数を例とともに示す。合計で 7,623 例あった。

表 5.1 重複記載の 4 パターン

a. 異区同素(異なる区分で同じ素性と重複)	3,608例
(例) [区分01] ACT(敵軍の)攻撃が始まる	
[区分02] ACT(議長への)攻撃が始まる	
b. 異区異素(異なる区分で異なる素性と重複)	2,260例
(例) [区分01] MEA (大豆の)収穫が多い	
[区分02] GRA (会議の)収穫が多い。	
c. 同区同素(同じ区分で同じ素性と重複)	821例
(例) [区分01] CON 自転車(のサドル)ガ高い	
[区分01] CON 自転車(の価格)ガ高い	
d. 同区異素(同じ区分で異なる素性と重複)	934例
(例) [区分01] CON (シャワーデ)ヲ流す	
[区分01] LIQ (額ニ)汗ヲ流す	
合計	7,623例

5.1.2.3 コロケーション重複の要因

表 5.1 に示した通り，コロケーションの重複記載は，名詞が異区分のときも同区分のときにも生じている。このとき，述語(動詞／形容詞)を見ると，名詞が異区分のときには，述語に多義性のあるものとないものとがあり，名詞が同区分のときには，いずれの述語も多義性があるものばかりであった。つまり，コロケーションの重複には大きく分けて，以下の 3 つの場合がある [59]。

(A) 名詞に多義性があり，述語に多義性はない場合(異区分のとき)

(B) 名詞にも，述語にも多義性がある場合(異区分のとき)

(C) 名詞に多義性はなく，述語に多義性がある場合(同区分のとき)

[59] 名詞と述語双方に多義性がない場合にはコロケーションの重複はない。

以下，それぞれの場合について，重複記載の要因をあげる。

まず，(A)の場合には，次の 2 つの要因が考えられる。

要因(A)-1 多義である名詞が同じ素性を持つため，素性に特有の同じ述語と結び付く。

見出し語の各区分に同じ意味素性がある場合，「その素性に特有の述語」が重複する可能性は高い。例えば，〈LOC〉(場所)であれば「行く，戻る，広い，狭い」，〈ACT〉(行為)であれば，「始まる，終わる，止まる，早い，遅い」といったものが素性に特有の述語である。表 5.1 であげた「(敵軍の)攻撃が始まる」，「(議長への)攻撃が始まる」の重複例は，「攻撃」の両区分に同じく〈ACT〉固有の述語が現れているために生じた重複例である。

中には 2 つ以上の素性に特有の述語もある。例えば，〈MEA〉は計れるもの，〈GRA〉は計れないものに，という違いはあるものの，どちらもその量の多少に焦点を当てる意味素性である。表 5.1 の「(大豆の)収穫が多い」「(会議の)収穫が多い」の例で示した「多い」という述語は，〈MEA〉と〈GRA〉のどちらの素性にも特有の述語であるため，重複が生じている。

要因(A)-2 述語の抽象度が高いため，同じ述語と結び付く。

例えば，「ある」「ない」「する」「なる」といった述語は抽象度が高いため，見出し語が多義であるとき，それぞれに結び付きやすく，その結果，重複する数も多くなる。「ある」の例を示そう。「ある」には「存在する」という抽象度の高い意味がある。一方，見出し語「こぶ」には，「皮膚の隆起した部分」という意味と，「木や物の表面の，隆起した部分」という意味とがあり，どちらも存在するという意味の「ある」と結び付くため，「こぶガある」というコロケーションが重複する。

次に，(B)の場合には，次の要因が考えられる。

要因(B)-1 名詞と述語の組合せであるコロケーション自体が比喩的意味を持つため，同じように結び付く。

例えば，見出し語「あせ【汗】」の場合は，区分 01「暑いときやひどく緊張したときなどに，体内から皮膚上に出てくる水分」と，区分 02「物の表面にできた結露」との 2 つに区分される。このとき，区

分 02 は，区分 01 のメタファーによって意味が拡張したものだと捉えられる(第 4 章 4.3 節を参照)。このように区分間にメタファーの関係があると，コロケーションが重複しやすい[60]。区分 02 に記載したコロケーションは，「(ハム)ニ汗が出る」と「(ミカン)ガ汗ヲかく」の 2 例であるが，「汗が出る」，「汗ヲかく」のいずれも区分 01 にその記載がある。さらに，「汗をかく」全体で「尽力する」という慣用的な意味もあるため，慣用表現の例とも重複する。

区分間にメタファーの関係があるものの例をさらにあげると，見出し語「あめ【雨】」は区分 01「大気中の水蒸気が上空で冷やされ，水滴となって落ちてくるもの。また，それが降り続いている状況」と，区分 02「上から絶え間なくたくさん降り注ぐもの」とに分かれ，区分 02 の「～ニ火の／血の雨が降る。降り注ぐ」が，区分 01 の「雨が降る」「～カラ～ヘ／ニ雨が降り注ぐ」と重複する。また，見出し語「まど【窓】」は，区分 01「建物の内部から外をみたり明かりを採ったりするためのもの」と，区分 02「内と外の境界にあって，内部のものを外部に見せるところ」とに分かれ，区分 02 の「～ガ～ニ心の／話し合いの／窓ヲ開く，開ける，閉ざす」が，区分 01 の「～ガ窓ヲ開く，開ける，閉ざす」と重複している。

しかしながら，メタファーの関係があっても，コロケーションが重複しないものもある。見出し語「ふね【船，舟】」は，区分 01「人や荷物をのせて海や川などの水面を移動する乗り物」と区分 02「料理屋や旅館などで刺身や貝などをもりつけて食膳にだす。船の形をかたどった入れ物」の 2 つに区分されるが，メタファーである区分 02 のコロケーション「～ガ(刺身の)舟ヲ注文する」や「～ガ(刺身)ヲ舟ニ盛る」などは，区分 01 と重なるものではない。ただ，「舟が大きい，小さい」は区分 01 と 02 とに重複しているが，これは，〈CON〉に特有の述語であるので，先に述べた要因(A)-1 で説明される例に相当すると考える。

さらに，(C)の場合には，次の要因が考えられる。

要因(C)-1 多義である述語が同じ名詞と結び付く。

例えば，表 5.1 であげた重複例「(シャワーデ)汗ヲ流す」，と「(額ニ)汗を流す」では，述語の指す意味はまったく異なるのに，それが異なる述語ではなく，同じ 1 つの述語の多義によるものであるため，結果として重複が生じている。

[60] 比喩的多義には，メタファーの他にメトニミーとシネクドキーがあるが，メトニミーにはあまり重複はなく，シネクドキーにも，メタファーほど目立った重複は見られない。

また，要因(A)-2 であげた「ある」「ない」「する」「なる」といった述語は，非常に多義でもあるため，要因(B)-1 によるこれらの述語の重複も多くある[61]。例えば，「ある」には「存在する」という意味のほかに「発生する」という意味もある。よって，「便りガある」というコロケーションは，「(机の上などに)手紙などによる知らせが存在している」という意味のものと，「手紙などによる知らせが来るという出来事が発生する」という意味のものとが重複する。このとき，「便り」という名詞には「手紙などによる知らせ」という意味が 1 つあるだけで多義性はないが，意味素性に異なりがある。つまり，述語の多義によって名詞の持つ異なる側面に焦点が当たっている(前者は物〈CON〉としての側面に，後者は書くという行為〈ACT〉という側面に焦点が当たっている)。

5.1.3 考察

重複記載されるコロケーションを例に，その重複を区別する辞書情報について考察する。

例えば，「色が濃い」というコロケーションは，以下に示すように，見出し語「いろ【色】」の区分 01，区分 03，区分 05，区分 07 の用例として重複する。

例:「いろ【色】」

[区分 01] (赤・青・黄など)物に当たって反射した光線が，その波長の違いによって視覚的に区別されて感じとられるもの。

1) GRA (口紅の／…)色が濃い

[区分 03] 表情に現れる，その時々の気持ちや心の状態の反映。

2) APP あせりの／…色が濃い

[区分 05] [連体修飾部で表される]ある状況になることを感じさせる度合い。

3) GRA 敗戦の／…色が濃い

[区分 07] その集団独特の，考え方や行動の傾向。

4) GRA ハト派の／…色が濃い

1)の「濃い」は，『IPAL形容詞辞書』における区分 05(色や字などが白に対してよく目立つ)であり，2)～4)の「濃い」は区分 06(程度が大である)である[62]。1)から 4)に示した通り，「色が濃い」というコ

[61] 「ある」「ない」「する」「なる」といった述語は，抽象度が高く，多義であるため，これらの述語による重複は非常に多い。延べで「ある:812 例，ない:769 例，する:406 例，なる:321 例，計 2,308 例」であった。全体の延べは，10,427 例であるので，以上の 4 語だけでその約 5 分の 1 を占めていることになる。

[62] 意味素性〈APP〉は〈appearance〉の略で視覚チャンネルから主として得られる外的な印象に焦点が当てられるものである。

ロケーションは「色」という名詞の先行句部分に違いがある[63]ため，その先行句の違いが判別できればうまく多義性を解消することができるように見える。

『IPAL 名詞辞書』ではコロケーション記載欄とは別の[連体修飾用法 1]という欄に，コロケーションの先行句に関する情報として，連体修飾語の例を連体修飾のタイプによって分類し記載している[塩谷，斎藤 1996]。「色」の区分 01，03，05，07 の先行句に関する記載の抜粋を以下に示す。

[区分 01] 〈対象〉[CON｜TIM｜…]
信号の—，表紙の—，服の—，炎の—，花の—，夕暮れの—。
〈評価〉独特の—。
〈価値〉好みの—。
[区分 03] 〈部位〉顔の—，目の—。
〈具体化〉不安の—，警戒の—，反省の—。
[区分 05] 〈具体化〉敗戦の—，秋の—。
[区分 07] 〈具体化〉タカ派の—，保守の—。

実際のコーパスに「(〜の)色が濃い」というコロケーションが現れたとき，「色」の先行句についてコーパスに現れたものと，上記にあげた辞書の記載例との類似度を何らかの方法で計算することで，「色が濃い」の多義性を解消できると考えられる[64]。

試みに，「色が濃い」で『IPAコーパス』[65][橋本 1995]を検索すると，5)〜10)の実例が見つかる。

5)《老油は濃口しょうゆに似て【色が濃】い。》
6)《レモンティーの場合はレモンによって香味は増すが，紅茶の【色が濃】くなる。》
7)《イカ漁シーズンを控え，漁師にはあせりの【色が測い。》
8)《不便を強いられる市民には疲労の【色が濃】いが，そんな中で水道だけは二日中に復旧する兆しが見えてきた。》

[63] 先行句に( )のないものは何らかの先行句が必須であることを示す。
[64] 現在の分類名称は，必ずしも複数の区分に重複するコロケーションを判別させるための情報という観点から付与されたものではないため，分類名称から区分を判別するのは難しい。しかしながら，多義性が生じる可能性の高い区分どうしは，あらかじめ判別の手がかりになるような補助情報を与える方法を検討すべきであったと考える。例えば，[区分 03]，[区分 05]，[区分 07]にはいずれも同じく〈具体化〉と書いてあるだけだが，順番に「心情を表す名詞」「状況を表す名詞」「思想を持つ団体を表す名詞」といったような補助情報の記述が判別には役立つだろう。
[65] 新聞，文芸作品，マニュアルなど約 47 万文。

9)《スペインと米国による約四百年の植民地支配の【色が濃】いのだ。》

10)《全戸に配られたチラシには「決して強制ではありません」と記してあるが，行政区長や隣組長を総動員した寄付集めは，実質的に強制の【色が濃】くなる。》

5)～10)の「色」がどの区分の意味であるかを判別してみる。5)は「～の」という先行句がないことから唯一省略可能である区分 01 であり，6)は「紅茶」が具体物を示す名詞であることから，具体物が先行句例に示されている区分 01 である。また，7)は先行句例に同じものがある区分 03 であり，8)の「疲労」は，区分 03 の先行句例「あせり，不安，警戒，反省」に近いタイプの名詞だと考えられるので，区分 03 である。残りの 9)，10)は，「植民地支配」「強制」が 05 の先行句例「敗戦」に近いタイプの名詞だと考えられるので，区分 05 である。以上，すべての例について区分が判別できた。

続いて，5.1.1 項にあげた「骨が折れる」の記載を例に，その実例を 11)～15)に示す。

例:「ほね【骨】」

[慣用表現 1]　骨が折れる

[区分 01]　CON　(手の／足の／…)骨が折れる

〈所有者(人，高等動物)〉老人の—，魚の—，鳥の—。

〈部位〉体の部分頭の—，肩の—，足の—。

[区分 02]　CON　(傘の／…〉骨が折れる

〈全体部分(樹状に支えるものがある物)〉傘の—，扇の—

11)《荒井さんは左足の【骨が折れ】，約三カ月のけが》

12)《一色さんは頭の【骨が折れ】て重体》

13)《ロシアを説得することが一番【骨が折れ】た。》

14)《根のない話を虚空に描き出すのには【骨が折れ】る。》

15)《【骨が折れ】とるかも知れんな。》

11)，12)は，「左足の」「頭の」と，身体部位を表す先行句をともなっているので，区分 01 の意味だとすぐわかる。一方，13)，14)は，身体部位を表す先行句や，「傘」などの先行句をともなわないこと，さらに「一番」などの程度副詞が現れることから慣用表現であると判断できる。

15)は，14)同様に先行句はないが慣用表現ではない。区分 01 か区分 02 であるかの多義性は残るが，慣用表現でないことはわかる，それは，「骨が折れている」というテイル形から判断できる。「骨が折れる」の場合，慣用表現として用いられるときはテイル形をとりにくい。こういったテイル形の情報は『IPAL動詞辞書』では記載されていたものであるが『IPAL名詞辞書』では記載されていない。多義性を解消する辞書情報を考えるとき，このような情報，即ちアスペクト情報を記述することも必要であるといえる [66]。

先行句に関する辞書情報を考えるために，さらにもう一例あげる。「足跡を残す」の『IPAL 名詞辞書』の記載例とその実例を 16)～18)に示す。

例:「そくせき【足跡】」

[区分 01] RES ～ガ～ニ足跡を残す

〈場所〉裏庭の—，雪の上の—，河原の—。

〈関与者〉[ANI]ネコの—，犯人の—。

〈種類〉運動靴の—，パンプスの—。

[区分 03] RES ～ガ～ニ(輝かしい／偉大な／…)足跡を残す

〈主体〉先達の—，文人達の—。

〈関係事象〉青春の—。

〈時間〉五年間の—。

16)《中村校長らの不安をよそに二人は輝かしい【足跡を残】した。》

17)《しかも，選りに選って雪の夜，降り積もった裏庭から侵入して，かくも鮮明な【足跡を残】して去ったというわけか? 》

18)《時代感覚を鋭敏に反映する作品で昭和初年代の新詩運動に大きな【足跡を残】した。》

読みに関する情報があれば区分 01 は「あしあと」と読み，区分 03 は「そくせき」と読むのが普通であるので，両者を区別することは容易である。しかしながら計算機でテキストを処理する場合はテキスト上に現れるものと辞書情報とを手がかりに区別するほかない。16)は，「輝かしい」が区分

66 なお，IPAL 名詞辞書では，アスペクトの違いまで含む連用および連体の修飾情報に留意するという発想で意味素性がふられている。連用および連体の修飾情報については，橋本の用例分析と青山の本文執筆による橋本，青山[1992])がある。

03 のコロケーションの先行句例と合致することから区分 02 であることが簡単に判断できる。問題は，17)と 18)の判別である。

『IPAL 名詞辞書』では先行句の情報として，「～の」で連体修飾する用例は積極的に記載されているが，形容詞や「～な」などで連体修飾する用例はあまり記載されていなかった。そのため，17)に現れている「鮮明な」という先行句を辞書情報に照らし合わせてみても，17)が区分 01 か区分 03 かを判別することは難しい，区分 01 の先行句情報として，「鮮明な，はっきりとした，くっきりとした」といった用例の記載があれば，17)は 01 であることが判別できる。

18)の場合も辞書情報に「大きな」あるいは「小さな」といった先行句例がないので区分の判断ができない例である。しかも，「大きな，小さな」は，区分 01，02 のどちらの先行句としてもありえる表現であり判別する手がかりにはならない。実は，この場合は，「ニ格に立つ名詞句」の方を参照すればよい。『IPAL 動詞辞書』では，その動詞がとる格情報として格に立つ名詞のタイプを示していたが，『IPAL 名詞辞書』では，名詞のコロケーションの情報として，他の格に立つ名詞句例はいくつか例示するだけにとどめ，詳細な記載が行われていなかった。例えば，区分 01 に，ニ格に立つ名詞句のタイプは具体的な場所を表すものであるとの記載があれば，18)は，「新詩運動に」とあるので，区分 01 ではなく区分 02 であることがわかる。「足跡を残す」の場合は，先行句に「～の」の形の連体修飾句をとることはほとんどなく，形容詞や「～な」の形の連体修飾句か，ニ格に立つ名詞句(すなわち，述語の格情報)により多義性を解消できると考えられる。

以上「色が濃い」「骨が折れる」「足跡を残す」の例から，実例にある多義性を解消するためには「先行句」に関する辞書情報を用いて類似判断することの有用性を認めることができた。そして，そのとき，辞書情報として先行句に「～の」の連体修飾例を豊富に記載するだけでなく，それ以外の形の修飾用例も記載する必要があることがわかった。また，動詞辞書には記載されていたアスペクト情報やすべての名詞句の格についての詳細な情報も，コロケーションの多義性を解消するための辞書情報として欠かすことのできないものであったことも確認できた。さらに，副詞句に関する情報も多義性解消の手がかりとして記載すべきであることがわかった。

「先行句」をはじめとするコロケーションに付随する情報を参照することが多義性の解消に役立つことはわかったが，最後に，コロケーションに関する辞書情報を利用して多義性を解消することの限界について，今後検討すべき問題点の提起として述べる。

例えば，重複例のタイプによっては，多義性をなくすような辞書記述が最初から困難な場合がある。以下に例を示す。

19a) [区分 01] INT(この／…)一軒屋が広い

[区分 02] INT(その／…)一軒屋が広い

19b) [区分 01] LOC(新築の／…)一軒屋ヲ訪ねる

[区分 02] LOC(竹藪の中の／…)一軒屋ヲ訪ねる

19a)と 19b)の区分 01 は「独立した一戸建ての家」を指し，区分 02 は「周囲にほかの家がないところに，一軒だけ建っている家」を指すものである。名詞辞書では，19b)には両者の区分を区別できるような先行句を記述しているのに対し，19a)には区別のつかない先行句を記述している。実際のテキストにおいて「広い」といった述語と結び付くときは，「訪ねる」という述語とは違い，区別のつけやすい先行句をともないにくいため上記のような辞書記述に落ち着いてしまう傾向がある。

さらに，重複をうまく区別できる辞書記述ができたとしても，実際の文章中には型どおりには現れてはくれないため辞書情報が使えないという問題もある。次にその例を示す。

20a) [区分 01] RES～ガ(友人の／…)成功ヲうらやむ

[区分 02] RES～ガ(彼の／…)成功ヲうらやむ

20b) [区分 01] RES～ガ(実験の／計画の／…)成功ヲ祈る

[区分 02] RES～ガ(彼の／…)成功ヲうらやむ

20a)，20b)は，もともと区分 02 の意味を特定するのが困難な場合である。区分 01 は「物事が思い通りにいき，目的を達成すること」を指し，区分 02 は「富や地位を手に入れること」を指す。区分 01 の方は実際にコーパスをひくと，「企画の成功」「サミットの成功」「上映の成功」といった実例があり，20a)の先行句の記載例を訂正して，20b)のように記載すべきであることがわかるが，区分 02 の方は具体的な先行句が存在しにくい。『IPA コーパス』[橋本 1995]中に区分 02 の意味ととれるものが 1 例だけ見つかった。それは，次のものである。

《カプリ島に別荘を持つのはナポリ市民の夢だが，島の中心地に九室もある大きな別荘を購入したことは，ミケーレ弁護士の地位と【成功を】物語るものだ。》

このときの先行句は「ミケーレ弁護士」であるが，この先行句だけを見る限りでは区分 01 の可能性も否定できない。区分 02 の意味を特定するものは，「成功」に並列する「地位」という名詞からの連想や，「島の中心に大きな別荘を購入した」という文脈からの判断になる。コロケーションの出現

するテキスト全体から得られる文脈情報の参照方法を辞書情報に反映させるかという検討は，今後の課題として残されている。

5.1.4 コロケーションの多義性を解消するための辞書記述の手順

先行句に関する情報として記載されている「～の」の形で連体修飾する用例により多義性を解消できることが明らかになった。さらに，コーパスを分析することにより，『IPAL 名詞辞書』にはほとんど記載されていなかった，形容詞や「～な」の形で連体修飾する用例や，十分に記載されていなかったテイル形に関するアスペクト情報や副詞表現についての情報も，重複記載を区別するうえで有効な辞書情報であることがわかった。

そこで，コロケーションの多義性を解消するための辞書記述の手順について，次の通り提案する。

1. コーパスからコロケーションを抽出する。
2. コロケーションに重複がある場合，いずれによるものか特定する。

   要因(A)-1 多義である名詞が同じ素性を持つため，素性に特有の同じ述語と結び付く。

   要因(A)-2 述語の抽象度が高いため，同じ述語と結び付く。

   要因(B)-1 名詞と述語の組合せであるコロケーション自体が比喩的意味を持つため，同じように結び付く。

   要因(C)-1 多義である述語が同じ名詞と結び付く。
3. 「～の」の形で連体修飾する用例があれば，その用例[先行句]を詳細に記述する。
4. 形容詞や「～な」の形で連体修飾する用例があれば，その用例[先行句]を詳細に記述する。
5. 結び付く述語の格情報（[文型]と[格に立つ名詞句のタイプ]）を記述する。
6. テイル形に関する[アスペクト情報]や副詞表現[連用情報]についての情報抽出が可能であれば抽出し，それも詳細に記述する。

上記の方法により，BCCWJ より得られたコロケーション情報を利用し，「色が濃い」「骨が折れる」「足跡を残す」は次のように記述できる。

例：「いろ【色】」

「色が濃い」の重複の要因：

要因(A)-1　多義である名詞が同じ素性を持つため，素性に特有の同じ述語と結び付く。

[区分 01]　GRA　(口紅の／…)色が濃い

〈対象〉[CON｜TIM｜…]信号の—，表紙の—，服の—，炎の—，花の—，醤油の—，夕暮れの—，緑の—，葉の—。

[区分 03]　APP　あせりの／…色が濃い

〈部位〉顔の—，目の—。

〈具体化(心情を表す名詞)〉不安の—，警戒の—，反省の—，焦慮の—，憂慮の—，苦悩の—，傷心の—，いらだちの—。

[区分 05]　GRA　敗戦の／…色が濃い

〈具体化(状況を表す名詞)〉秋の—，自己防護の—，敗北の—，崩壊の—。

[区分 07]　GRA　ハト派の／…色が濃い

〈具体化(思想を持つ団体やその考え方・行動を表す名詞)〉タカ派の—，保守の—，過激なイスラム原理主義の—，徒弟制度の—，仏教国の—。

例：「ほね【骨】」

「骨が折れる」慣用表現との重複の要因：

要因(B)-1　名詞と述語の組合せであるコロケーション自体が比喩的意味を持つため，同じように結び付く。

「骨が折れる」区分 01 と 02 の重複の要因：

要因(A)-1　多義である名詞が同じ素性を持つため，素性に特有の同じ述語と結び付く。

[慣用表現 1]　[ABS]ニ骨が折れる

連用：一番，とても，たくさん，大分，ダントツで，もっと，少し，とてつもなく，すごく，なかなか，ちょっと。

＜テイル形＞無

[区分 01]　CON　(手の／足の／…)骨が折れる

〈所有者(人，高等動物)〉老人の—，魚の—，鳥の—。

〈部位〉体の部分頭の—，肩の—，足の—，足首の—，手首の—，鼻の—，腕の—，指の—。

連用:ポキポキ。

＜テイル形＞有

[区分 02] CON (傘の／…)骨が折れる

〈全体部分(樹状に支えるものがある物)〉傘の―, 扇の―

連用:ポキポキ。

＜テイル形＞有

例:「そくせき【足跡】」

「足跡を残す」の重複の要因:

要因(B)-1 名詞と述語の組合せであるコロケーション自体が比喩的意味を持つため, 同じように結びく。

[区分 01] RES [HUM／ANI]ガ[LOC]ニ(判別不能な／ひきずったような／きれいな／不審な／大きな／巨大な／…)足跡を残す

〈場所〉裏庭の―, 雪の上の―, 河原の―。

〈関与者〉[ANI]ネコの―, 犯人の―。

〈種類〉運動靴の―, パンプスの―。

＜テイル形＞有

[区分 03] RES [HUM]ガ[RES／TIM]ニ(輝かしい／偉大な／貴重な／重要な／確かな／多彩な／歴史的な／大きな／巨大な／…)足跡を残す

〈主体〉先達の―, 文人達の―。

〈関係事象〉青春の―。

〈時間〉五年間の―。

＜テイル形＞有

5.1.5 辞書情報による重複記載の多義性解消の確認

「**要件 2-1** 多義性をもつコロケーションの意味を区別するための情報が得られること」を満たす方法として, コロケーションの先行句などの重複記載を区別する情報を辞書に明示的に記述する方法を提案した。それら重複記載を区別する辞書情報により, 多義性の解消が可能である用例数を確認する。

5.1.5.1 確認方法

『IPAL 名詞辞書』に重複するコロケーションは，延べで 10,415 例，異なりで 4,578 例(全コロケーションのうち平均して 2〜3 個が重複例)である。この重複記載において，重複を区別できる情報をもつ割合を計る。この分析は，重複例どうしを比較することで行う。例えば，5.1.2.1 目にあげた「角を出す」は区分 01，区分 02，慣用表現の 3 カ所に重複記載があったが，まず，区分 01 の例と区分 02 の例とを比べ，次に，区分 01 の例と慣用表現との例を比べ，さらに区分 02 の例と慣用表現の例とを比べる，というように重複例を 1 組ずつ組み合わせて比較していく。比較する対象とした重複例の組合せ総数は，5.1.2.2 目の表 5.1 で示した通り 7,623 対である。

5.1.5.2 確認結果

重複記載を区別する情報として，次の 5 つに着目し，その例数を数えた。

a. 他の名詞句が立つ格助詞が違う(他格)
b. 他の名詞句が違う(他名)
c. 見出し語の先行句が違う(先行)
d. 見出し語の後行句が違う(後行)
e. 見出し語が立つ格助詞が違う(格助)

「他の名詞句」とは，「見出し語以外の格に立つ他の名詞句」のことである。排他的な分類ではないため，合計数は 7,623 例ではなく，10,339 例である。

表 5.2 重複記載を区別する違いの 5 種類

a. 他の名詞句が立つ格助詞が違う(他格)	2,421 例
(例) 02:CON～ガ～ニお**茶**ヲ入れる	
02:PRO～ガお**茶**ヲ入れる	
b. 他の名詞句が違う(他名)	1,145 例
(例) 02:HUM(控え室)ニ**家族**ガいる	
02:ROL(彼女)ニ**家族**ガいる	
c. 見出し語の先行句が違う(先行)	4,888 例
(例) 01:ACT～ガ(原野の)**開拓**ヲ行う	
02:ACT～ガ(技術の／新市場の)**開拓**ヲ行う	
d. 見出し語の後行句が違う(後行)	91 例
(例) 01:POT **のど**(の調子)ガいい	
03:FOR **のど**ガいい(＝歌う声がいい)	
e. 見出し語が立つ格助詞が違う(格助)	1,794 例
(例) 01:INT～ガ**舞台**ヲ／カラ出る	
01:SPA～ガ**舞台**ニ出る	

表 5.2 中の太字は「見出し語」を表す。表 5.2 より,「c. 見出し語の先行句が違う」だけで 47.3%の重複例が区別できるものであることがわかる。

しかし, 表 5.2 だけでは, a～e の違いがどういった組合せで出現しており, また, それらの違いがどういったパターンの重複を区別するものであるかが見えないため, a～e の違いを 24 の出現型に展開し, 出現数を集計した結果を表 5.3 に示す。

表 5.3 重複記載を区別する違いの出現型別出現数

「－」は各部分の内容が同じことを，「＋」は違うことを表す。ただし，「他格／他名」欄については，表 5.2 の「他格」に違いのある場合を「＋」とし，「他名」に違いのある場合を「×」で表す。

タイプ	重複記載を区別する違い 5 種類				
	他格／他名	先行	後行	格助	出現数(%)
1	－	－	－	－	1,187(15.6)
2	－	－	－	＋	77(1.0)
3	－	－	＋	－	20(0.3)
4	－	－	＋	＋	2(0.0)
5	－	＋	－	－	2,594(34.0)
6	－	＋	－	＋	162(2.1)
7	－	＋	＋	－	13(0.2)
8	－	＋	＋	＋	2(0.0)
9	＋	－	－	－	330(4.3)
10	＋	－	－	＋	533(7.0)
11	＋	－	＋	－	12(0.2)
12	＋	－	＋	＋	34〈0.4)
13	＋	＋	－	－	594(7.8)
14	＋	＋	－	＋	913(12.0)
15	＋	＋	＋	－	2(0.0)
16	＋	＋	＋	＋	3(0.0)
17	×	－	－	－	524(6.9)
18	×	－	－	＋	15(0.2)
19	×	－	＋	－	0(0.0)
20	×	－	＋	＋	1(0.0)
21	×	＋	－	－	551(7.2)
22	×	＋	－	＋	52(0.7)
23	×	＋	＋	－	2(0.0)
24	×	＋	＋	＋	0(0.0)
計					7,623(100.0)

表5.3から，一番多く重複しているタイプは2,594例存在するタイプ5であることがわかる。つまり，先行句だけに違いのある「寿命が長い」と「(電池の／…)寿命が長い」のようなものが一番多い，ということである。次に多いのは1,187例あるタイプ1である。これは重複記載を区別する違い(a～e)がまったくないコロケーションである。しかし，それは全体の15.6%であり，残りの84.4%(6,436例)は，何らかの違いが記載されているものであることがわかる。

次に，表5.3に示したタイプ1～タイプ24が，コロケーションの重複記載の4パターン(a～d)別にどのように出現しているかについて調べた結果を表5.4に示す。表5.4からは，コロケーションの重複が異区分にある場合と同区分にある場合とで，出現数の多いタイプが違っていることがわかる。特徴的なのは，タイプ5，タイプ13，タイプ17，タイプ21などは異区分の方の割合が高く，タイプ10，タイプ2，タイプ14などは同区分の割合が高いという点である[67]。

この特徴と，表 5.3 の内容を考え合わせると，先行句や他の名詞句に違いがあり，述語の文型に違いのないタイプが異区分に多く，逆に，述語の文型に違いのあるタイプは，同区分に多いといえる。ところが，5.1.2.3 目で述べた通り，コロケーションが異区分で重複しているときは，(A)名詞に多義性があり，述語に多義性はない場合か，(B)名詞にも，述語にも多義性がある場合かであった。また，同区分で重複しているときは，(C)名詞に多義性はなく，述語に多義性がある場合であった。

以上のことから，計算機でコロケーションの多義性を解消するためには，まずコロケーションの多義性が(A)，(B)，(C)のどの場合であるかを辞書情報を手がかりに判別することが有効であると考えられる。なぜなら，もし(A)もしくは(B)とわかれば，「先行句」や「他の名詞句」の辞書情報を利用して多義性の解消ができ，もし(C)とわかれば，述語の「文型」に関する辞書情報を利用して多義性の解消ができるということが，表 5.3 と表 5.4 の分析から示唆されるからである。

67 これらのうち特に特徴的なものがタイプ 5 とタイプ 10 である.たとえば，「(プロペラの)回転が速い」と「(頭の)回転が速い」のようなものがタイプ 5 であり，「～ガえびヲ食べる」「～ヲえびガ食べる」のようなものがタイプ 10 である。

表 5.4 重複記載パターンごとに見る 24 の出現型別出現数と割合 (%)

タイプ	重複記載の 4 パターン			
	a.異区同素	b.異区異素	c.同区同素	d.同区要素
1	819(22.7)	198(8.8)	0(0.0)	170(18.2)
2	7(0.2)	14(0.6)	43(5.2)	13(1.4)
3	9(0.2)	2(0.1)	6(0.7)	3(0.3)
4	0(0.0)	1(0.0)	1(0.1)	0(0.0)
5	1,698(47.1)	754(33.4)	25(3.0)	117(12.5)
6	65(0.7)	67(3.0)	22(2.7)	8(0.9)
7	8(0.2)	4(02)	0(0.0)	1(0.1)
8	1(0.0)	1(0.0)	0(0.0)	0(0.0)
9	63(1.7)	140(62)	57(7.0)	70(7.5)
10	27(0.7)	83(3.7)	298(36.3)	125(13.4)
11	2(0.1)	8(0.4)	2(0.2)	0(0.0)
12	4(0.1)	5(0.2)	19(2.3)	6(0.6)
13	190(5.3)	249(11.0)	61(7.4)	94(10.1)
14	114(3.2)	333(14.7)	248(30.2)	218(23.3)
15	0(0.0)	2(0.1)	0(0.0)	0(0.0)
16	0(0.0)	1(0.0)	0(0.0)	2(0.2)
17	375(10.4)	113(5.0)	5(0.6)	31(3.3)
18	1(0.0)	6(0.3)	4(0.5)	4(0.4)
19	0(0.0)	0(0.0)	0(0.0)	0(0.0)
20	0(0.0)	1(0.0)	0(0.0)	0(0.0)
21	220(6.1)	236(10.4)	30(3.7)	65(7.0)
22	5(0.1)	40(0.8)	0(0.0)	7(0.7)
23	0(0.0)	2(0.1)	0(0.0)	0(0.0)
24	0(0.0)	1(0.0)	0(0.0)	0(0.0)
計	3,608(100)	2,260(100)	821(100)	934(100.0)

5.1.5.3 確認結果のまとめ

『IPAL 名詞辞書』にすでに記載されている次の 5 つの情報により，重複対 7,623 例のうちの全体の 84.4%の重複が解消できるものであることを確認した。

a. 他の名詞句が立つ格助詞が違う(他格)
b. 他の名詞句が違う(他名)
c. 見出し語の先行句が違う(先行)
d. 見出し語の後行句が違う(後行)
e. 見出し語が立つ格助詞が違う(格助)

また，このうちの「c. 見出し語の先行句が違う(先行)」は，上記 5 種類の例数の合計 10,339 例のうち 47.3%を占めるものであることがわかった。これは，この情報が特に有効であることを表す。梁[1996]は『IPAL』辞書間のリンク情報を作成するときに多義性を解消する手段として文型情報のみを利用していたが，「先行句」の情報を用いればさらなる精度向上が図れたと思われる。

本論文では上記 a.～e.の 5 つの情報に加え，「テイル形に関するアスペクト情報や副詞表現」の記述も追加し，多義性を解消する情報を詳細に記述することを提案している。また，重複の要因も辞書に記載することを提案している。よって，本辞書記述案に沿って辞書の構築を進めれば，『IPAL 名詞辞書』の多義性解消精度 84.4%以上の効果が期待される。

5.2 使用域記述のための文体アノテーションと「古風な語」の使用域の詳述

本節では，使用域の記述を提案する。そのために，まずは書籍という同一のレジスタの中に存在する書籍サンプルの多様な文体的特徴を捉えることを目的に，BCCWJ に収録されている書籍サンプルを対象にして，コーパスに文体分類をアノテーションする方法を論じる。そのうえで，「古風な語」を対象に，コーパスから得られる用例の使用域を明らかにし，それを辞書記述に取り込む方法を提案する。

5.2.1 使用域の辞書記述のための文体分類

語の理解・把握のためには，その語の使用域をコーパスから明らかにし，辞書に記述することが重要だと考える。従来の辞書記述では不十分であったところである。

これまでに，作家等の個人の文体研究とは別に，言語的特徴からレジスタ別にテキストの文体を分析する研究が盛んになっている。例えば，英語コーパスを対象とした文体研究に，Biber [1988]，Nakamura [1995]，Biber and Conrad [2009]などがある。それらは，話し言葉と書き言葉，あるいは，小説，散文，新聞，といったレジスタの違いよって，文法項目や，語彙項目に差異がみられることを議論している。そうしたレジスタの差異を辞書記述に取り込むことも有用であろう。BCCWJ にも，書籍，雑誌，新聞，Web 文書といったレジスタの異なるサブコーパスが収録されているため，その差異を辞書に記述していくことは可能である。BCCWJ に収録されるサンプルテキストの文体を計量的に考察する試みは間淵ほか[2010]や小磯ほか[2011]などがある。

しかしながら，そういったレジスタの差異や，出典情報から得る差異ではなく，テキストそのものがもつ文体の差異の情報が利用できれば，さらなる詳細な使用域の記述が可能になると考える。例えば，難度の高い文体，低い文体，硬い文体，軟らかい文体など，どういった文体のテキストで多用される語であるかを示すことができれば，新しい辞書記述が可能になるだろう。

そこで，書籍という同一のレジスタの中に存在する書籍サンプルの多様な文体的特徴を捉えることを目的に，BCCWJ に収録されている書籍サンプルを対象にして，文体分類をアノテーションする方法を論じる。

5.2.2 文体分類の指標の設計

文体分類は，書籍サンプルが①か②のいずれのタイプであるかを先に判断し，それに沿って細分類をする。

① 文体判断が可能と判断されるもの，即ち，テキスト構造が単純(例：章節構造)なものを内容・表現の文体的特徴の印象判定により細分類する。(→5.2.2.1目)

② ①以外，文体判断が単純に判断し難いとされるもの，即ち，テキスト構造・紙面形式に特徴をもつものを選別，分類する。(→5.2.2.2目)

5.2.2.1 目に①の分類指標の設計を，5.2.2.2 目に②の分類指標の設計を示す。

5.2.2.1 内容・表現の文体的特徴を表す分類指標

コーパスへ文体情報を付与することの重要性は，EAGLES[1996]より議論されてきた。BCCWJ に収録されている書籍サンプルには，NDC(日本十進分類法)によるジャンルや，C コード(日本図

書コード)による販売対象，発売形態，また，著者情報，形態論情報などが付与されており，それらを利用して，半自動的に種々の観点から分類することは可能である。しかしながら，EAGLES [1996]がコーパスへ付与することが望ましいとあげる，(A)対象読者に想定される読解レベル(難易度)，(B)テキストの作成意図，(C)さまざまな文体情報，という 3 種に関する情報がほとんど付与されていない。わずかに，(A)について C コードの販売対象情報があるのみである。そこで，(A)～(C)を補えるものとして，「専門度，客観度，硬度，くだけ度，語りかけ性度」という 5 つの分類指標を設ける。以下，順次述べる。

(A)難易度

はじめは，「とても難しい・どちらかといえば難しい・ふつう・どちらかといえば易しい・とても易しい」の 5 段階評定による分類を行おうとした。しかしながら，試行を行ったところ，知らない分野や，親しみのない分野であればあるほど難しく感じられ，その逆は易しく感じられる，という，判定者の個人的な尺度による難易度判定になりやすい傾向がみられた。また，真ん中の「ふつう」の尺度が人によって大きくぶれる傾向がみられた。そこで，できる限り客観的にサンプルの難易度を測るために，難易度を直接問うのではなく，そのテキストが，どういった対象に向けて書かれていると判断されるか，という「専門度」を問うこととする。真ん中に「一般向き」をすえ，上には「専門家向き・やや専門家向き」という指標を設ける。下には年齢のスケールを導入するのがわかりやすいと考え，「中高生向き・小学生・幼児向き」という指標を設ける。次の通りである。

(a) **専門度**

1 専門家向き

読む前提に高度な専門知識が必要なもの

それを仕事にしているような人向きのもの

2 やや専門的な一般向き

読む前提に多少の専門知識が必要なもの

3 一般向き

特に専門的な内容ではないもの

専門的な内容であっても，読む前提に専門的知識を特に必要とせず，一般向きに書かれているもの

4 中高生向き

中高生向きに書かれているもの

専門性の有無にかかわらず，中高生でも読めそうなもの

5 小学生・幼児向き

このときの「専門性」はテキストを理解する上での「高度な知識の必要性」の有無と考える。例えば，パソコン関係では，技術者や研究者向きであれば「専門家向き」であり，やや高度な知識を必要とするパソコンを趣味にしている人向きであれば「やや専門的な一般向き」であり，家庭向きであれば「一般向き」と考える。

(B)作成意図

可能であれば，「論説，随筆，報告文，紀行文，手順書・・・」といったような体系的な分類案を作成し，それに基づいた指標の付与が理想であるが，指標の設計やその判断にかかる負荷が大きいことが予測されるため，そのような指標は設けない。代わりに，書き手の態度には「客観的」か「主観的」かの区別があると考え，その判断付与を行う。真ん中を設定せずに，「とても・やや客観的」か「とても・やや主観的」という指標を設ける。

(b) **客観度**

1 とても客観的

2 どちらかといえば客観的

3 どちらかといえば主観的

4 とても主観的

ここで「客観的」とは，主に，事実，観察，論証などが述べてあるものである。即ち，誰が読んでも納得できるよう著述されているものを想定する。反対に，「主観的」とは，主に，経験や感想などが述べてあるものである。なお，この客観度はノンフィクションと判断をしたテキストについてのみ付与する。つまり，小説等にはつけない。

(C)さまざまな文体情報

この情報は，EAGLES[1996]で定義が困難だと述べられているものである。複数のパラメータが議論されているが，標準は定まっていないと言う。しかし，例えば，外国語の学習者にとって重要

な情報になり得るものであるため，英語の学習者用辞書に載せる語レベルの文体情報(どのような文体で用いられる語であるか)を表すラベルの提案例が紹介されている。Joos[1961]の提案("frozen"，"formal"，"informal"，"colloquial"，"intimate")や，Halliday et al. [1964]の提案("colloquial"，"polite"，"casual"，"intimate"，"deferential")などである。つまり，EAGLES[1996]でコーパスに備えることが望ましいと議論されている「文体情報」とは，形式性，親疎性，口語性に関わる文体情報だと言えるだろう。ただし，これら英語の単語に付与されるものとして設計されたラベルを，そのまま日本語の書籍テキストの文体の分類指標に使うには無理がある。

日本語では，テキストの形式性，親疎性を"formal"(正式)か"informal"(非正式)かと表すより，「硬い」か「軟らかい」かと表すことが多い。よって，「硬度」を問うこととした。「硬い」とは，かしこまっている感じ，堅苦しい感じであり，「軟らかい」とは，かしこまっていない感じ，親しみやすい感じである。真ん中を設定せずに，「とても・やや硬い」か「とても・やや軟らかい」という指標を設けた。

(c) **-1 硬度**

1 とても硬い

2 どちらかといえば硬い

3 どちらかといえば軟らかい

4 とても軟らかい

また，日本語では，テキストの形式性，親疎性を"polite"(改まっている・丁寧である)か，"casual"(くだけている)かと表現することもある。しかし，ある程度整っていれば「改まっている」と感じられるが，その程度が「とても」であるのか「やや」であるのかの区別はつけ難い。だが，「くだけている」程度はそれに比べて答えやすい。そこで，「くだけ度」として 3 段階「とても・やや・いない(＝改まっている)」を問う指標を，形式性，親疎性を問う 2 つ目の指標として設けた。

(c) **-2 くだけ度**

1 とてもくだけている

2 どちらかといえばくだけている

3 くだけていない(＝改まっている)

口語性については，例えば，最近の英語辞書では，"spoken"(話し言葉)か"written"(書き言葉)かというラベル付与が行われることが多いようである。話し言葉と書き言葉とは，テキスト上大きく区別されるものである。しかし，書籍テキストはすべて書き言葉であるため，その文体には口語性を問う指標の設計は必要ないと考えるのが 1 つの考え方である。実際のところ，「えー」「あー」といったフィラーを含むことや，独り言を言っているような書きぶりにより，話し言葉的と感じられるテキストが書籍の中にもあるにはあるが，非常に少ない。よって，口語性を直接問うような指標は設けないこととした。しかしながら，書籍のサンプリングの過程において，口語性に関し，非常に特徴的な文体をもつテキストが多く目についた。それは，例えば，「あなた」や「みなさん」などの呼びかけ表現が用いられているもの，問いかけや相づちを求めるような文末表現が用いられているようなものである。特にハウツー本や，解説書に多くみられる，直接読者に意識的に語りかけ性るような文体である。それらを「直接的な語り」という表現スタイルをもつものと定義し[柏野 2010]，すべてのテキストに対し，「語りかけ性度」を問うこととした。3 段階「とても・やや・ない」の指標を設けた。

(c) **-3 語りかけ性度**

1 とても語りかけ性がある

2 どちらかといえば語りかけ性がある

3 特に語りかけ性はない

書籍テキストの文体情報を分類し得るものとして，以上 5 つの分類指標を設計する。

5.2.2.2 テキスト構造・紙面形式に特徴をもつものの分類指標

先の5.2.2.1目にあげた文体の分類指標を付与しがたいと考える一群の書籍サンプルがある。1つは，テキスト構造・紙面形式に目立つ特徴をもつものである。もう1つは，内容や表現に特定の特徴をもつものである。それらは，必要に応じて，その特徴による類型で整理分類をし，別途，その文体を吟味すべきと考えた。そこで，次のような分類指標を設け，該当するサンプルの選別，分類を行った。なお，いずれの分類指標も該当サンプルにはすべてに付与したため，複数の分類指標が付与されたサンプルも存在する。

［テキスト構造・紙面形式上の特徴］

(a) 対話系(対話，対談・座談，インタビュー，往復書簡，シナリオ，その他対話形式)

(b) 引用系(Q&A 形式, 投稿形式, その他引用編集形式)

(c) 視覚表現多用系(コマ割多用, 図解, その他写真やイラストの多用)

(d) データベースやリスト系(用語解説, 辞書形式, 見本・カタログ形式, その他リスト形式)

［内容や表現上の特定の特徴］

(e) 前書きや後書きである

(f) 明治時代より以前の古い言葉が多い

(g) 外国語が多い

(h) 数式やプログラミング言語などが多い

(i) 法律文が多い

(j) 教育現場で使いがたそうである [68]

(k) その他一定量の「本文」が認めがたい

5.2.3 文体分類のアノテーション作業

文体分類のアノテーション作業は, 次の概要の通り実施した。

●対象:BCCWJに収録されている図書館サブコーパス(10,551サンプル)の書籍サンプル[69]。

●1サンプルの範囲と長さ:サンプル全体の印象判定を行うため, 1サンプル全体を範囲とする。サンプルの一部より一定量の字数を揃えて抽出するようなことはあえてしない。1サンプルの平均はおよそ3,000語。

●作業者:言語データ構築経験有のおおよそ20～50代の女性, 延べ8名。

また, アノテーション作業は, 次の通り, 二段階に分けて実施した。

●作業段階:

[作業の一段階目] 典型例抽出を目的とする作業 [Kashino and Okumura 2010]

-目的:分類指標の検証及び基準の検討, 人手付与の作業上の問題点の検討, 典型例の抽出。
-態勢:判断のゆれを検証するために同一サンプルを作業者3人で判定。

---

68 例えば, 暴力的な描写や性的な描写を含むものを区別するための指標である。文体情報付与のための指標というよりは, コーパス活用のためのテキスト整理の指標として設けたものである。

69 1986 年から 2005 年までの 20 年間に発行された書籍のうち, 東京都内の 13 自治体以上の公共図書館で共通に所蔵されていた書籍を母集団とし, そこから抽出したサンプルから成るサブコーパスである。

−作業量:1セット約400〜500の書籍サンプルに対する作業を延べ約10日間で行う。
−判断:付与すべき指標の種類についてごく簡単な説明があるのみ。
−付与済テキスト数:3,324テキスト(全体の31.5%)。

[作業の二段階目] 全体付与を目的とする作業 (柏野ほか2012)

−目的:全10,551サンプルへの付与。
−態勢:1サンプル1人以上の判定と機械判定の相互参照。
−作業量:1セット約400〜500の書籍サンプルに対する作業を延べ約10日で行う。
−判断:判断事例付きマニュアルを参照。
−付与済テキスト数:BCCWJ図書館サブコーパス収録の全10,551サンプル。

なお，作業の一段階目では，「対象読者，主観的・客観的，硬軟と丁寧さ，話し言葉・書き言葉」という，5.2.2.1 目で示した最終形とは次の 2 点において異なる分類指標で判定作業の試行を行った。

1．(c)-1 硬度と(c)-2 くだけ度を，「硬くてくだけている」というテキストを想定せずに，次のようにまとめていた。

1 とても硬くて丁寧

2 どちらかといえば硬くて丁寧

3-1 どちらかといえば軟らかくて丁寧

3-2 どちらかといえば軟らかくてくだけている

4-1 とても軟らかくて丁寧

4-2 とても軟らかくてくだけている

2．(c)-3 語りかけ性度を問わず，話し言葉口調，おしゃべり口調のみられるテキストの区別を想定し，次のように問うた。

1 どちらかといえば書き言葉的

2 どちらかといえば話し言葉的

この一段階目の作業後に，「硬軟」と「丁寧さ」は似てはいるものの，異なる軸として見るべきことがわかり，指標を分けた。実際，全アノテーション終了後に「硬い」＋「くだけている」の両方が付与されるサンプルも得られ，初期の指標設計が不十分だったことがわかった。また，一段階目の作業を通し，作業者にとって「話し言葉的」と判断する基準がかなりあいまいであることがわかったため，書き言葉のテキストでは，「話し言葉的」であるか否かを問うのではなく，5.2.2.1 目で示した通り，直接的な「語りかけ性」を問う指標に改めた。

5.2.3.1 アノテーション作業における人手付与のゆれの検証

作業の一段階目終了後に，人手付与のゆれの検証を行った。

アノテーション作業結果より 161 テキスト(うち，主観的・客観的の対象となるノンフィクションは 49 テキスト)を抽出した。最初の 3 人による試行作業を Step1 と呼ぶこととする。次に，同じ 161 テキストについて，別の 2 人の作業者が，先の 3 人の判断結果を参照しながらアノテーションを行うという追加作業を試みた。これを Step2 と呼ぶこととする。

Step1，Step2 において，対象読者，主観的・客観的，硬軟と丁寧さ，話し言葉的・書き言葉的，それぞれについての作業者間の判断の一致度を，一致率，カッパ係数(一致率から偶然の一致率をひいたもの)，相関係数で求めた。その結果を表 5.5 に示す。

表 5.5　Step1, Step2 の作業者間の一致度

対象読者					
	Step1				Step2
作業者の組	1−2	2−3	3−1	Average	4−5
一致率	0.73	0.93	0.73	0.80	0.88
カッパ係数	0.17	0.55	0.18	0.30	0.58
相関	0.33	0.58	0.29	0.40	0.67
主観・客観					
	Step1				Step2
作業者の組	1−2	2−3	3−1	Average	4−5
一致率	0.49	0.33	0.24	0.35	0.47
カッパ係数	0.46	0.19	0.21	0.28	0.47
相関	0.61	0.29	0.48	0.46	0.70
硬軟					
	Step1				Step2
作業者の組	1−2	2−3	3−1	Average	4−5
一致率	0.37	0.25	0.16	0.26	0.71
カッパ係数	0.24	0.21	0.07	0.17	0.67
相関	0.59	0.58	0.37	0.51	0.79
話し・書き					
	Step1				Step2
作業者の組	1−2	2−3	3−1	Average	4−5
一致率	0.73	0.32	0.07	0.37	0.89
カッパ係数	0.03	0.02	0.00	0.02	0.62
相関	0.13	0.06	0.02	0.07	0.63

Step1 は，明確な判断基準や典型例の例示のない状況下での試行作業であったため，判断の一致は高くないことは予想してはいたが，実際に中～低度の一致であった。しかしな

がら，Step2 ではいずれの値も飛躍的に改善された。3 人の作業結果を見て判断するという Step2 は，マニュアル参照に近いことを疑似的に行ったと考えることができる。よって，このことから，適切なマニュアルを整備した上でアノテーション作業を進めることにより，アノテーション作業の質を高めることは十分可能であることが確認できた。

5.2.3.2 典型例の選定と検討

典型例を抽出するために，各分類指標の各選択肢において，3 人全員の判断が一致したテキスト，2 人が一致したテキストを同定した。分類指標別にその結果を表 5.6～表 5.9 に示す。

表 5.6「対象読者」の判定結果別テキスト数

	1.専門家向き		2.やや専門的な一般向き		3.一般向き		4.中高生向き		5.小学生・幼児向き	
全員一致	1	0.0%	23	0.9%	1904	71.3%	11	0.4%	2	0.1%
2人一致	6	0.2%	138	5.2%	503	18.8%	23	0.9%	6	0.2%
計	7	0.3%	161	6.0%	2407	90.1%	34	1.3%	8	0.3%

表 5.7「主観的・客観的」の判定結果別テキスト数

	1.とても客観的		2.どちらかといえば客観的		3.どちらかといえば主観的		4.とても主観的	
全員一致	21	1.3%	84	5.1%	70	4.2%	53	3.2%
2人一致	150	9.1%	578	34.9%	271	16.4%	105	6.3%
計	171	10.3%	662	40.0%	341	20.6%	158	9.5%

表 5.8「硬軟と丁寧さ」の判定結果別テキスト数

	1.とても硬くて丁寧		2.どちらかといえば硬くて丁寧		3-1.どちらかといえば軟らかくて丁寧	
全員一致	2	0.1%	250	9.4%	60	2.2%
2人一致	50	1.9%	707	26.5%	231	8.7%
小計	52	1.9%	957	35.8%	291	10.9%
	3-2.どちらかといえば軟らかくてくだけている		4-1.とても軟らかくて丁寧		4-2.とても軟らかくてくだけている	
全員一致	53	2.0%	4	0.1%	18	0.7%
2人一致	326	12.2%	34	1.3%	35	1.3%
小計	379	14.2%	38	1.4%	53	2.0%

表 5.9「直接的な語り性」の判定結果別テキスト数

	1.直接的な語り性あり		2.直接的な語り性なし	
全員一致	1	0.3%	311	82.7%
2人一致	20	5.3%	44	11.7%
計	21	5.6%	355	94.4%

表 5.6～表 5.9 を見ると，判定の一致したテキスト数が選択肢によって差のあることがわかる。「一般向き」や「直接的な語り性なし」といった，該当テキスト数がそもそも多数あるようなものは全員一致のテキスト数が多い。逆に該当するものが少なそうなもの，特に「専門家向き」，「小学生向き」「とても硬くて丁寧」「とても柔らかく丁寧」「直接的な語り性あり」は，全員一致のテキスト数が非常に少ない。そのようなものこそ典型例の抽出と分析の必要なものであるため，それらは 2 人一致のテキストからも典型例の候補を抽出した。そのようにすることにより，作業の二段階目に入る前に，全分類指標の典型例をそろえることができた。

5.2.4 文体アノテーション結果の考察

5.2.2 項に述べた通り，アノテーション作業は，書籍サンプルが①か②のいずれかのタイプであるかを先に判断し，それに沿って細分類をするという方法で行った。

①文体判断が可能と判断されるもの，即ち，テキスト構造が単純(例：章節構造)なものを内容・表現の文体的特徴の印象判定により細分類する。(→5.2.2.1 目)

②①以外，文体判断が単純に判断し難いとされるもの，即ち，テキスト構造・紙面形式に特徴をもつものを選別，分類する。(→5.2.2.2 目)

上記①には 10,551 の書籍サンプルのうち，8,887 サンプル(84%)が該当した。②には 1,664 サンプル(16%)が該当した。上記①の結果を 5.2.4.1 目から 5.2.4.3 目に，②の結果を 5.2.4.4 目から 5.2.4.5 目に示し，分析する。

5.2.4.1 5 つの分類指標の付与例

各分類指標が付与されたサンプルより，典型例と認められる例を示す。サンプルの出典は，BCCWJ のサンプル ID と書名とで示す。

(1)専門度：1 専門家向き(LBi4_00021『がんと遺伝子』)

E2F以外のRB結合タンパク質としては，転写因子RAX，T細胞が活性化するときに誘導されるIL‐2，GM‐CSF，HIV‐2などの転写を活性化する転写因子ElF‐1や先に述べた細胞周期を制御するサイクリンDなどがある．おもしろいことに，ElF‐1やサイクリンDのRB結合ドメインにはlargeT抗原やE1Aタンパク質と同じようにLXCXEというアミノ酸配列が存在する．また，RBタンパク質は骨格筋分化を支配する重要な遺伝子群MyoDファミリー(MyoD，myogenin，MRF4，myf‐5)の産物とも複合体を形成し筋分化にも関与しているらしい．

(2)専門度:4 中高生向き(LBf9_00090『超魔炎獄変』)

白く薄い空気のヴェールが、漂うように揺らめいている。
シャ…ァァン、シャラ…ァァン…。
闇を抜け、霧の中を渡る金属の響き。それは魔を覇する浄化の音。
響きに道を開けるかのようにすぅ…っと霧が左右に分かれた。
それは。
霧の中にたたずむそれは。
闇。
…いや。闇ではない。

(3)客観度:1 とても客観的(LBo3_00158『行政法要論』)

たとえば委員会の開催が「急施を要する場合」にあたるかどうかとか、公衆浴場の施設が「公衆衛生上不適切」かどうかは通常人の経験則によって十分判断できる事柄であるから、羈束裁量であって裁判所の終局的な判断に服すべきものとする。これに対し、外国人の在留期間の更新を適当と認めるに足る相当の理由があるかどうかは、出入国管理行政の責任者である法務大臣の政治的判断に委ねらるべきであり、また、原子炉の安全性の認定は高度の科学的専門技術的知見に基づく総合的判断であるから、行政庁の便宜裁量事項であり、その当否は裁判所の審理・判断にはなじまないとする(最判昭和五三年一〇月四日民集三二巻七号一二二三頁、同平成四年一〇月二九日民集四六巻七号一一七四頁)。

(4)客観度:4 とても主観的(LBo3_00132『教師をめざす若者たち』

どんなに上手な言葉を使っても、思っていないことを発すれば、子供に伝わらない。どんなに下手な言葉でも、心から伝えたいという愛情があれば、伝わるものであるということを信じることができました。この実感は日本でも通じる「教育の原則」であると思いました。
二日目、子供たちと綿花摘みを一緒にしました。敦煌の子供たちの手は「仕事をしている手」でした。

(5)硬度:1 とても硬い(LBi3_00033『現代法社会学入門』)

取引費用がゼロである場合には、法的ルールの内容のいかんを問わず、資源配分は効率的レベルとなるというコースの定理は、法的ルールによる権利の分配のあり方のいかんを問わず、取引費用ゼロの社会では効率性が実現されることを意味する。したがって、法的ルールの選択、つまり権利の分配は、この意味のコース的世界においては、もっぱら所得分配、つまり分配的正義の観点から判断されることになる。もちろん、現実の社会では取引費用がゼロではない。

(6)硬度:4 とても軟らかい(LBa4_00010『恐竜の世界をたずねて』)

恐竜が滅亡したわけや、恐竜たちのさいごのようすをしり、その原因をきわめるためには、恐竜の先祖のことをしらなくては、ほんとうのことがわかりません。

恐竜の先祖をしらべるには、ふるい時代につもった地層を、一枚一枚、したへしたへとしらべていかなければなりません。

このようにして恐竜の先祖をたずねていくと、中生代の三畳紀のはじめにいた、「テコドント」(図86)という、からだの長さが一メートルあまりの爬虫類にいきあたります。テコドントは、四本足であるき、走るときは二本足だったことがわかっています。恐竜の先祖は、このころから四本足または二本足の動物だったわけですね。

(7) くだけ度:1 とてもくだけている(LBf9_00067『男はオイ！女はハイ…』)

最近流行りの通信販売。例の新聞の日曜版の裏面などに、克明にズラリと商品が写真などで広告してあるやつ。あれをば何となく眺めているうちに、どうしても欲しくなった商品があった。

よし、こいつひとついってやれとばかりすぐ電話にとびついた。

「ハイ、こちら―です」と出たのは、耳ざわりだけでわかるアルバイトギャルの声。

「商品番号をおっしゃって下さい」

といわれて答える。

さらに「御住所と御名前、電話番号を郵便番号からどうぞ」ってんで、こいつにも律儀に返事をする。

(8) 語りかけ性度:1 とても語りかけ性がある(LBt1_00013『5 分間集中力トレーニング』)

精神的に疲れていると、「ああなったら、どうしよう」「こうなったら、どうしよう」と常に不安だらけになります。

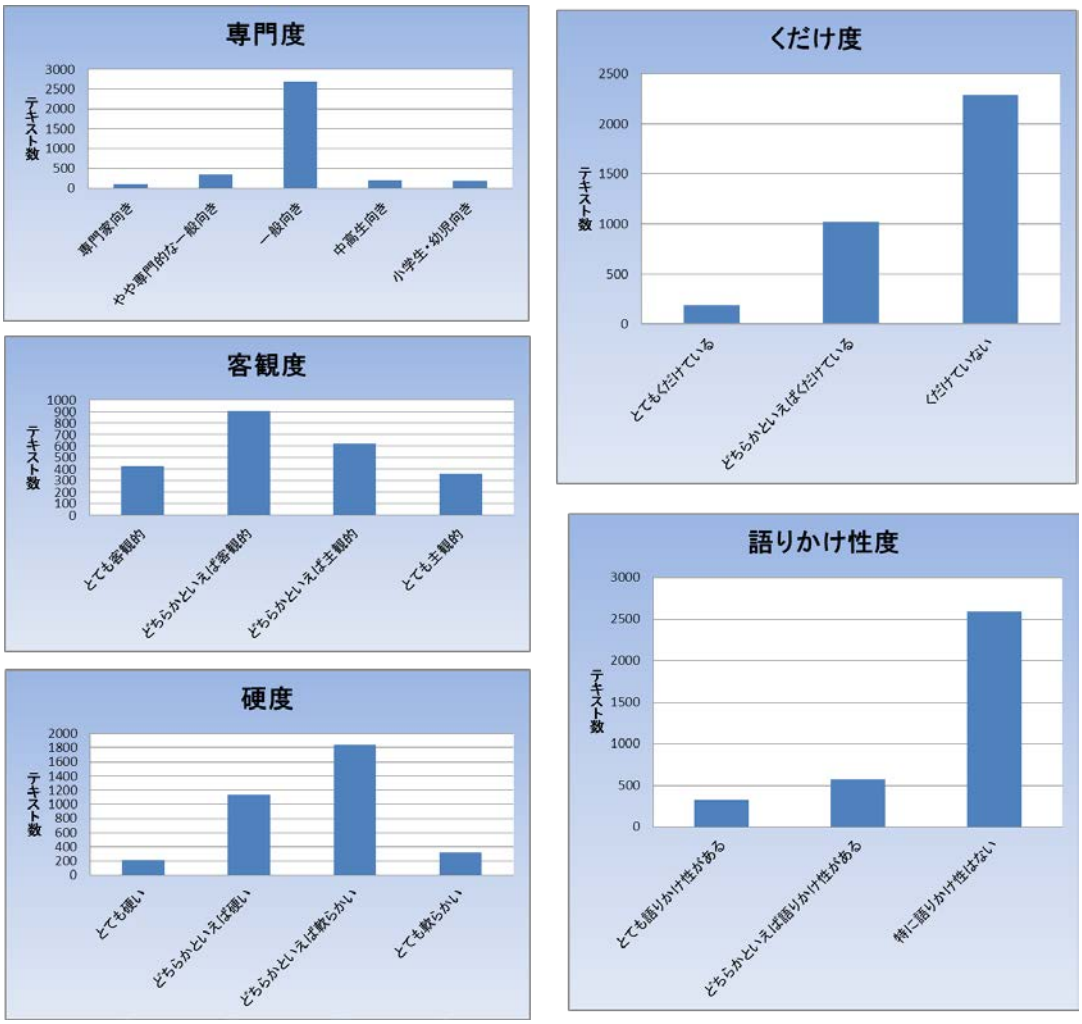
動物病院にいらっしゃる飼い主さんには、過剰な不安を抱えている人や心配性の人がとても多い。実はそれがペットの病気をさらに悪化させることになっていますが、そういう認識をお持ちの飼い主さんは、あまりいません。詳しい説明は避けますが、不安や心配性を放っておくと、動物の具合が悪くなり、当然それが自分にも返ってきます。

それでは、どうすればいいのでしょうか。

5.2.4.2 5 つの分類指標の付与結果

「構造的に単純なテキストタイプ」と判断され，5 つの分類指標が付与された，8,887 サンプル(全体の 84%)のアノテーション結果の内訳を，次の図 5.1(数値はいずれも，テキスト数)に示す。図 5.1 から，次のことがわかる。専門度は「一般向き」が圧倒的に多い。客観度は，「とても」と「どちらか」をあわせ，「主観的」よりも「客観的」の方が多い。硬度は，「とても」と「どちらか」をあわせ，「硬い」よりも「軟らかい」方が多く，そして，「硬い」も「軟らかい」も「とても」が付与されたものは少ない。くだけ度は，「とても」と「どちから」を足すと，全体の約 3 分の 1 に付与されたことがわかる。語りかけ性度は，「とても」と「どちから」を足すと，全体の約 4 分の 1 に付与されたことがわかる。くだけ度，語りかけ性度，いずれもその特徴を有するテキストが一定量あることが確認できる。

図 5.1 分類指標の付与結果



5.2.4.3 5 つの分類指標から捉える NDC 別書籍テキストの特徴

5 つの分類指標の付与結果を NDC 別に見ることで，NDC 別の書籍テキストの特徴を詳細に捉え得ることができる。まず，図書館サブコーパス収録サンプルの NDC 別の数と比率を図 5.2 に示す。収録サンプルは，「9.文学」と「5.社会科学」が多い。続けて，NDC 別の付与結果の内訳は表 5.10 に示す。表 5.10 にある通り，総記，言語，null(NDC 未付与)のテキスト数は少ないため，以降この 3 つ以外の NDC 別のテキストを分析対象とする。

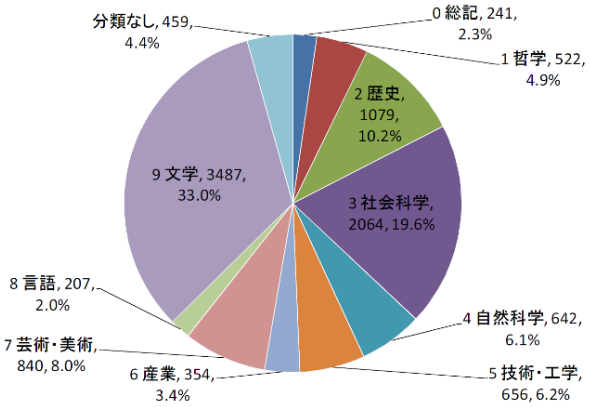


図 5.2　図書館サブコーパス収録サンプルの NDC 別の数と比率

(DVD 収録『『現代日本語書き言葉均衡コーパス』利用の手引第 1.0 版』(2011 年)より)

表 5.10 NDC 別付与済みテキスト数

番台	類	テキスト数
0番台	総記	8
100番台	哲学	209
200番台	歴史	346
300番台	社会科学	773
400番台	自然科学	196
500番台	技術，工学	178
600番台	産業	97
700番台	芸術，美術	204
800番台	言語	4
900番台	文学	1306
null	null	3
合計		3324

次に，NDC 別のテキストの分類指標の選択肢を点数化した。それを割合になおし，平均との差分を求めた。さらに，各分類指標がどちらに振れているかの尺度を次の通り求めた。おおよその傾向を捉えるために，選択肢別の重みづけは行わなかった。

「専門度」 (選択肢 1〜2 の和と 3〜5 の和との差分)

「客観度」 (選択肢 1〜2 の和と 3〜4 の和との差分)

「硬度」 (選択肢 1〜2 の和と 3〜4 の和との差分)

「丁寧度」 (選択肢 1,2,3-1,4-1 の和と 3-2,4-2 の和との差分)

「語りかけ性度」 (選択肢 1 と 2 の差分)

以下，求めた尺度の結果を NDC 別に図示する。

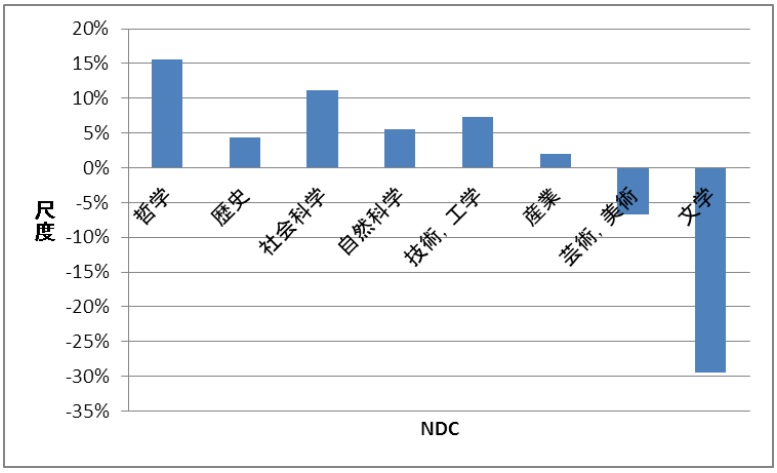


図 5.3 NDC 別専門度の違い

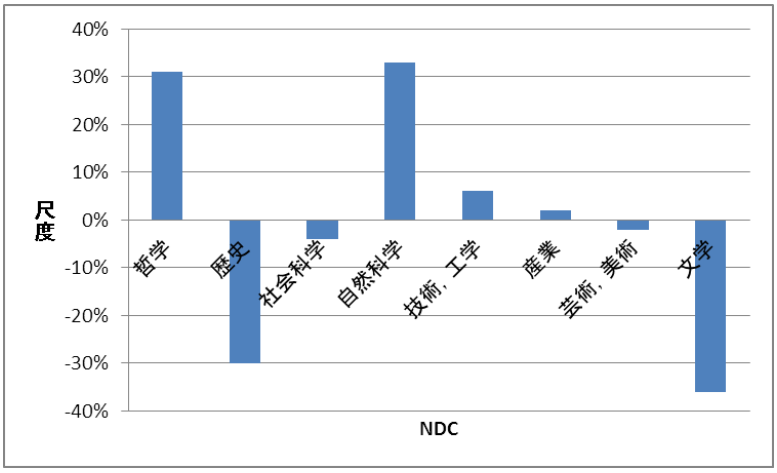


図 5.4 bNDC 別客観度の違い

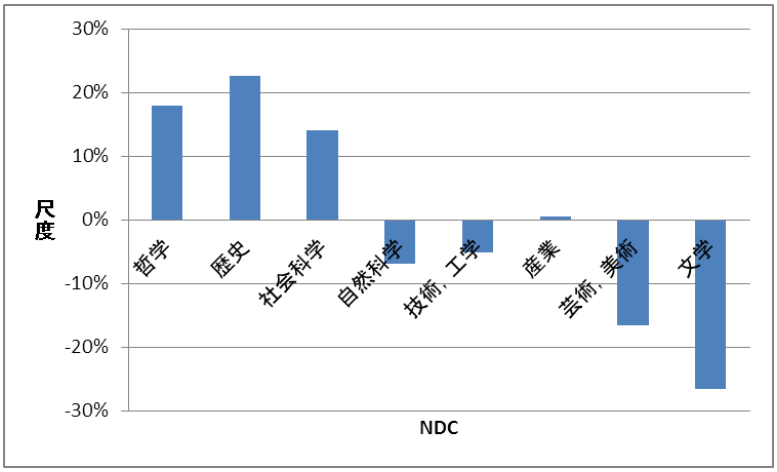


図 5.5 NDC 別硬度の違い

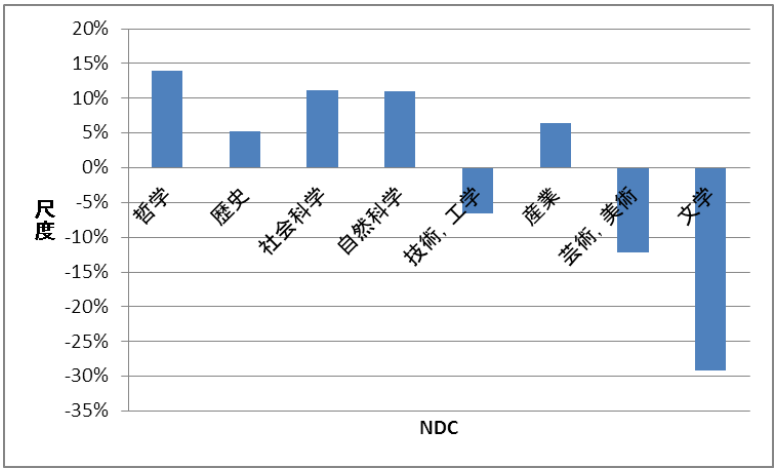


図 5.6 NDC 別丁寧度の違い

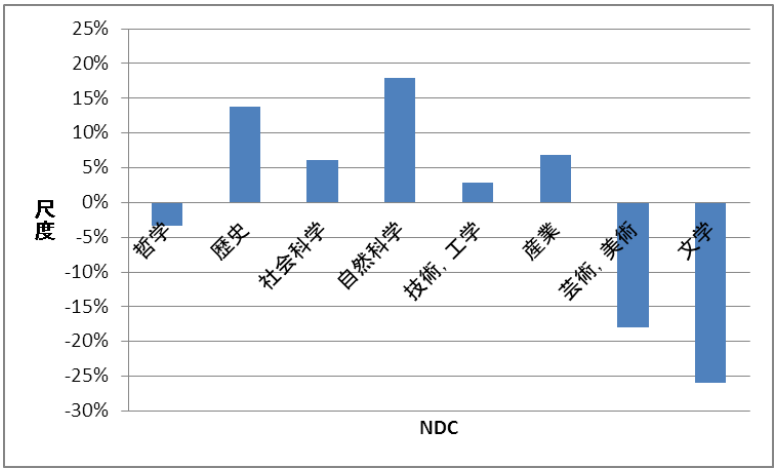


図 5.7 NDC 別語りかけ性度の違い

図 5.3～図 5.7 より, BCCWJ 図書館サブコーパスに収録された各 NDC の書籍テキストには，次のような特徴のあることを明らかにすることができた。

●専門度:文学以外の専門度が高いことは予測されたが，中では哲学がもっとも高い。また，芸術，美術の専門度が低いことは予測外。

●客観度:小説類は対象外であるため,「文学」はエッセー類のみ対象とした結果である。それらエッセー類は客観度が低いと判断されるのは予測通り。しかし，歴史もまた低いということが判明。

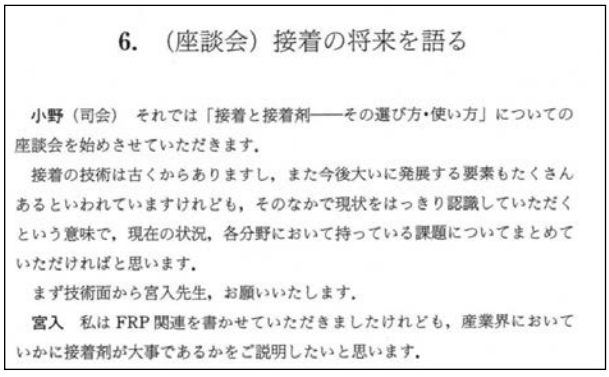
●硬度と丁寧度:両者は相関する傾向がみえるが，その中で，自然科学は硬度が低く丁寧度が高いという傾向をもつことが目立つ特徴。

●語りかけ性度:「丁寧度」と相関する傾向がみえる。自然科学や歴史の語り度が高いことが判明。また，哲学は丁寧度は高いが語り度は低めであり，逆に，技術，工学は丁寧度は低いが語り度は高めであるということが判明。

5.2.4.4 特徴的な類型のテキストの分類指標の付与例

「テキスト構造・紙面形式に特徴をもつもの」(5.2.2.2 目)と判断されて特徴的な類型の分類指標が付与されたサンプルより，典型例と認められる例を示す。サンプルの出典は，BCCWJ のサンプル ID と書名とで示す。

図 5.8 は，(a)対話系の座談の例であり，図 5.9 は，(a)対話系の対話の例である。



6. （座談会）接着の将来を語る

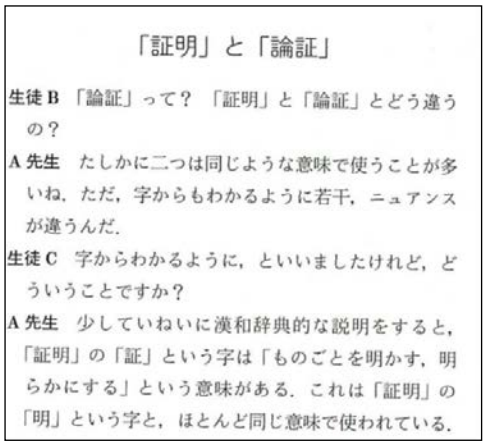
小野（司会） それでは「接着と接着剤――その選び方・使い方」についての座談会を始めさせていただきます。

接着の技術は古くからありますし，また今後大いに発展する要素もたくさんあるといわれていますけれども，そのなかで現状をはっきり認識していただくという意味で，現在の状況，各分野において持っている課題についてまとめていただければと思います。

まず技術面から宮入先生，お願いいたします。

宮入 私は FRP 関連を書かせていただきましたけれども，産業界においていかに接着剤が大事であるかをご説明したいと思います。

図 5.8 座談:『接着と接着剤』
(LBd5_00012)



「証明」と「論証」

生徒 B 「論証」って？ 「証明」と「論証」とどう違うの？

A 先生 たしかに二つは同じような意味で使うことが多いね，ただ，字からもわかるように若干，ニュアンスが違うんだ。

生徒 C 字からわかるように，といいましたけれど，どういうことですか？

A 先生 少していねいに漢和辞典的な説明をすると，「証明」の「証」という字は「ものごとを明かす，明らかにする」という意味がある。これは「証明」の「明」という字と，ほとんど同じ意味で使われている。

図 5.9 対話:『なぜ数学を学ぶのか』
(LBpn_00038)

図 5.10 は, (b)引用系の Q&A 形式の例であり, 図 5.11 は, (b)引用系のその他引用編集形式の例である。図 5.12 は, (c)視覚表現多用系のイラストの多用の例であり, 図 5.13 は, (c)視覚表現多用系のコマ割り多用の例である。

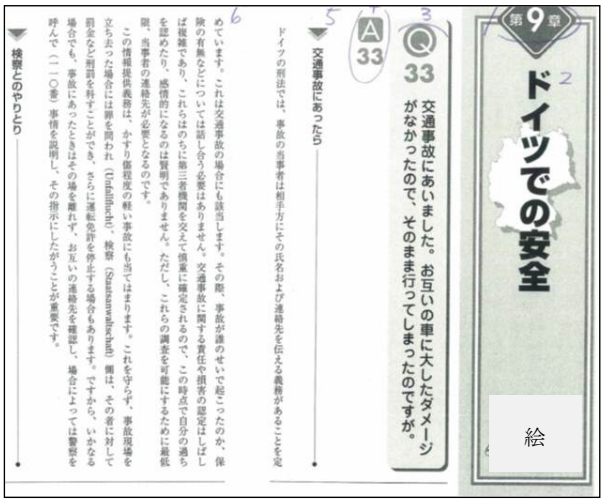


図 5.10 Q&A:『ドイツ暮らしの法律 Q&A』
(LBl8_00017)

図 5.11 引用:『試行錯誤の文章教室』
(LBr3_00037)

図 5.12 イラスト:『絵ときインテリア
ライティングの技法早わかり』
(LBb5_00011)

図 5.13 コマ割り:『東京で遊ぼ』
(LBsn_00022)

そして，図 5.14 は，(d)データベースやリスト系の辞書形式の例であり，図 5.15 は，(d)データベースやリスト系のカタログ形式の例である。

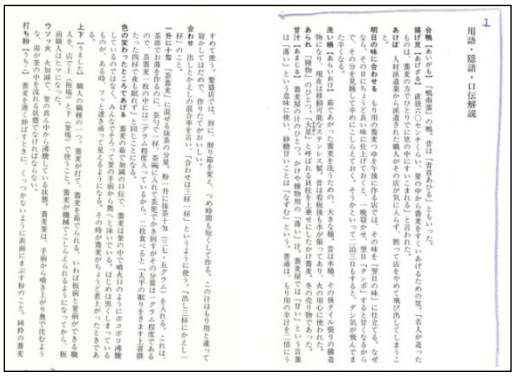


図 5.14　辞書:『蕎麦屋のしきたり』
(LBp6_00009)

図 5.15　カタログ:『熱帯魚・水草カタログ』
(LBj6_00025)

5.2.4.5 特徴的な類型のテキストの分類指標の付与結果

「テキスト構造・紙面形式に特徴をもつもの」と判断されて特徴的な類型の分類指標が付与された 1,664 サンプル(全体の 16%)に対する NDC 別分類指標の付与結果を表 5.11 に示す。分類指標は排他的ではないため合計は 1,664 を超える。図書館サブコーパス収録サンプルの NDC 別の数と比率は，先の図 5.2 に示した通り「9.文学」と「5.社会科学」が多い。よって，表 5.11 で「9.文学」「5.社会科学」が全体的に多いのは，もともとのサンプル数の比率の大きさに寄るところがある。しかしながら，図 5.16 の NDC 別分類指標の付与比率をみて見ると，収録サンプル比率とは異なる次のような特徴を確認することができる。

表 5.11　NDC 別分類指標の付与結果(1,664 テキスト)

NDC	サンプル数	(a)対話系	(b)引用系	(C)視覚表現多用系	(d)データベースやリスト系	(e)前書きや後書きである	(f)明治時代より以前の古い言葉が多い	(g)外国語が多い	(h)数式やプログラミング言語などが多い	(i)法律文が多い	(j)教育現場で使いがたそうである	(k)その他一定量の「本文」が認めがたい
0.総記	46	9	12	5	9	8	0	0	2	1	1	1
1.哲学	75	17	20	3	10	21	1	0	0	0	1	7
2.歴史	143	32	20	20	48	26	5	0	0	0	0	6
3.社会科学	355	112	68	10	66	54	3	0	0	13	17	31
4.自然科学	120	30	18	16	35	15	0	3	4	1	1	2
5.技術	180	18	22	57	71	13	0	1	1	3	0	11
6.産業	54	8	2	13	25	5	0	0	0	1	0	3
7.芸術	177	45	18	59	35	12	0	0	0	0	3	11
8.言語	86	11	14	1	39	7	0	16	1	0	0	5
9.文学	339	77	25	1	16	55	5	0	0	0	115	50
n.なし	89	9	10	30	26	5	1	1	0	0	3	5
計	1664	368	229	215	380	221	15	21	8	19	141	132

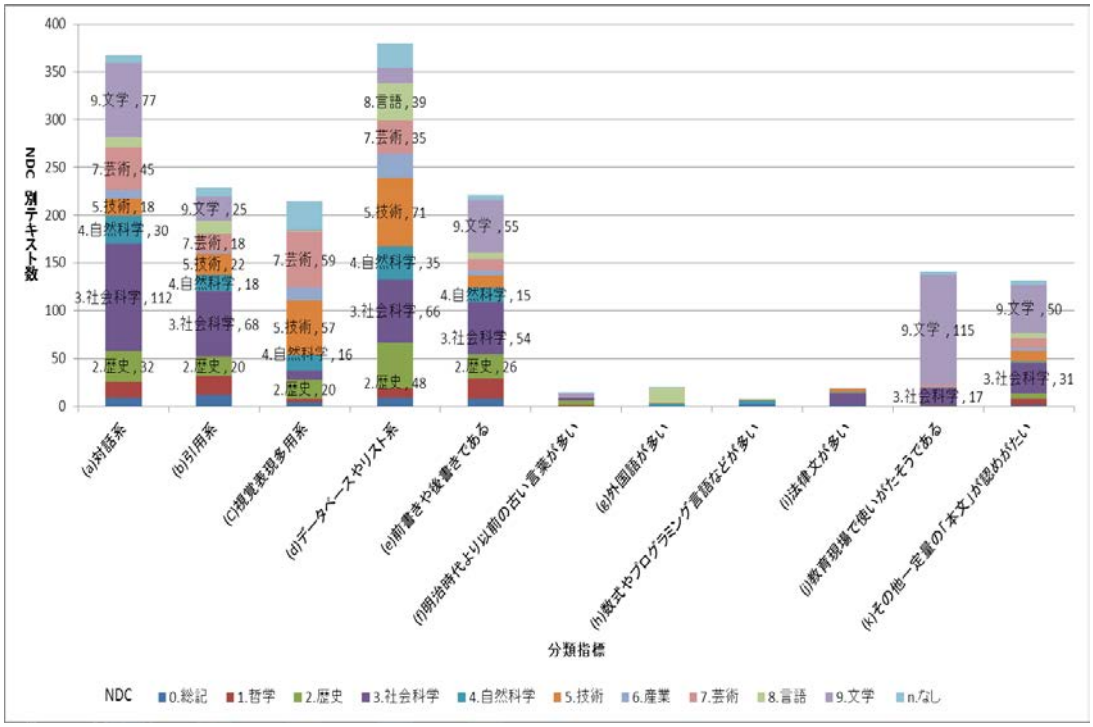


図 5.16 NDC 別分類指標の付与比率(1,664 テキスト)

・指標の(a)(b)(e)は，NDC の区別なく，広く用いられている形式である。

・指標の(c)は，「5.技術」「7.芸術」，「n.なし」に多い。これは「5.技術」にコンピュータのマニュアル等が多く，そこにキャプチャ画面が多用されていること，「7.芸術」に図画が多く提示されていること，「n.なし」にカタログ状の紙面が多いことに起因すると思われる。

・指標の(d)は，「6.産業」と「8.言語」に多い。「6.産業」には用語解説が，「8.言語」には辞書形式がそれぞれ多用されることによるものと考えられる。

・指標の(f)は，「3.歴史」が多くを占める。歴史を扱うテキストの中で古い言葉が多用されるからであろう。ただし，該当サンプル数はそもそも少ない。

・指標の(g)は，「8.言語」が大半を占める。外国語のテキストで、外国語が本文に入り込んでいるケースが多いためであろう。ただし，該当サンプル数はそもそも少ない。

・指標の(h)は，「0.総記」「4.自然科学」が大半を占める。前者にはコンピュータのプログラミング言語が，後者には数式が，それぞれ多用されているためであろう。

・指標の(i)は，「3.社会科学」の比率が高い。法学を含むこの NDC で，法律が多く引用されていることがうかがえる。

・指標の(j)は「9.文学」が非常に多くを占める。暴力的な描写や性的な描写を含む小説がこの NDC に入っているためである。

5.2.5 文体アノテーション結果からの語の抽出の試行

「**要件 2-2** 語の使用域・位相の情報が得られること」を満たすために，コーパスに文体アノテーションを行うことを提案した。そのコーパスより，文体的特徴をもつ語の抽出が可能であるかを試行する。

5.2.5.1 抽出方法

次の二つの方法により，その文体の特徴をもつ語の抽出を試みる。

方法1：定性的分析

Web アプリケーション「図書館コーパス検索ツール」を構築し[柏野，中村 2014]，それを利用し，特定の文体的特徴をもつテキスト群に含まれる文体的特徴をもつ語を抽出し，その結果を定性的にみる。

方法2：定量的分析

文体の分類指標に対して語彙素を特徴とした対応分析を行い，各指標と特徴的な語彙素の分布の対応を品詞ごとに定量的にみる[浅原ほか 2014b]。

5.2.5.2 抽出結果

定性的分析

文体情報付与済みコーパスから，特定の文体的特徴をもつテキストを検索し，形態論情報等を一覧できる文体情報検索ツールとして，Web アプリケーション「図書館コーパス検索ツール」を構築した[柏野，中村 2014]。このようなツールの先行事例としては，1 億語規模の BNC に対し，Lee[2001]が 70 種類(書き言葉 46 種，話し言葉 24 種)の言語使用域や，対象読者の性別や年齢層別，著者の属性等によって分類したものを検索可能にしたものがある。

「図書館コーパス検索ツール」では，特定の文体的特徴をもつテキスト群を指定でき，そのテキスト群に含まれるすべての語の品詞，語種，語彙素一覧や，平均文長，文末形式を抽出可能である。これを利用し，文体的特徴をもつ語を抽出する。

「とても硬い」と「とても軟らかい」というテキスト群より，「名詞-普通名詞-サ変可能」を抽出したものが，次の表 5.12 である。両者に重なって出現する語には網掛けをして表示する。表 5.12 を見ると，「とても硬い」と「とても軟らかい」では重なりが少ないことがわかる。「とても硬い」の方に多く現れる語には，「研究，教育，存在・・・」といった学術的な語が多くあり，「とても軟らかい」の方に多く現れる語には，「料理，結婚，食事・・・」といった日常生活で使用される語が多くある。同様に，「とても硬い」と「とても軟らかい」から「名詞-普通名詞-副詞可能」と「副詞」を抽出した結果を表 5.13 に，「接続詞」を抽出した結果を表 5.14 に示す。重なりは多いが，重なっていない語が，それぞれの文体でより使われやすい語と捉えることができる。

表 5.12　語彙素一覧より得た「名詞-普通名詞-サ変可能」の語

「とても硬い」名詞-普通名詞-サ変可能(上位 20 語):

物, 関係, 研究, 意味, 教育, 存在, 生活, 労働, 活動, 生産, 利用, 戦争, 運動, 規定, 著作, 裁判, 決定, 影響, 支配, 計画

「とても軟らかい」名詞-普通名詞-サ変可能(上位 20 語):

物, 話, 仕事, 心, 一緒, 電話, 生活, 関係, 意味, 料理, 結婚, 勉強, 食事, 説明, 涙, 存在, びっくり, 意識, 約束, 旅

表 5.13　語彙素一覧より得た「名詞-普通名詞-副詞可能」「副詞」の語

「とても硬い」名詞-普通名詞-副詞可能(上位 20 語):

ため, 場合, 中, 時, 所, 上(じょう), こう, 間, 他(ほか), 後(あと), 以上, そう, 例えば, 前, すべて, 上(うえ), 他(た), 内, 現在, 結果

「とても軟らかい」名詞-普通名詞-副詞可能(上位 20 語):

そう, 時, どう, 中, 所, もう, 今, 前, みな, ため, 上(うえ), 後(あと), 日, 二人, 内, 良く, こう, まま, ちょっと, 一人

表 5.14　語彙素一覧より得た「接続詞」の語

「とても硬い」接続詞:

また, しかし, 及び, そして, あるいは, 即ち, さらに, なお, しかも, かつ, ただし, 一方, ないし, ただ, もっとも, さて, もしくは, けれど, が

「とても軟らかい」接続詞:

そして, しかし, また, で, けれど, ただ, しかも, さて, あるいは, が, さらに, ただし, もっとも, 一方

定量的分析

浅原ほか[2014b]では，統計的手法を用いて，特定の文体に特徴的な語を明らかにした。例えば，次の通りである。

専門度：副詞「例えば」「最も」が「1.専門家向き」，「ずっと」「もう」が「5.小学生・幼児向き」
客観度：動詞「おける」「つく」が「1.とても客観的」，「やる」「くれる」「きく」が「4.とても主観的」
硬度：副詞「例えば」「さらに」「もっとも」が「2.どちらかといえば硬い」，「とても」「ずっと」「ちょっと」が「4.とても軟らかい」

また，浅原ほか[2014b]では，次の結論を得ている。

- 専門度・客観度については，対応分析によって得られた語彙が定性的な分析や作業者の判断記述などと一致する傾向がみられた。
- 硬度・くだけ度については，一部の特徴的な語彙素が得られているが，必ずしも作業者の判断基準としていた参考情報と一致しない部分がみられた。
- 語りかけ性度については語彙素のみによっては捉えられず，アノテーション作業においてより高度な認知的な判断が行われていることが示唆された。

5.2.5.3 抽出結果のまとめ

浅原ほか[2014b]により，必ずしも定性的な分析と，定量的な分析とが一致しないことを確認した。よって，「図書館コーパス検索ツール」より得られる文体的特徴をもつ語の一覧をそのまま辞書記述に反映させることはまだ課題が残ると言える。しかし，両者から得られる情報をあわせて吟味し，文体情報を辞書記述に取り込んでいくことは，これまでにない使用域の記述の実現につながると考えられる。

5.2.6 「古風な語」の使用域の詳述

第 4 章 4.1 節で取り上げた「古風な語」を再び取り上げる。4.1 節で述べた通り，次の 2 点を満たすものを「古風な語」と定義して呼ぶ。

(a) 「時代・歴史小説」を含めて現代で使用が見られる。
(b) 明治期以前，あるいは，戦前までの使用が見られる。

これより，現代語のコーパスである BCCWJ に収録されている約 3,000 万語分の書籍テキストから得られる用例について，その使用域を明らかにし，それを辞書記述に取り込む方法を提案する。

5.2.7 「古風な語」の用法の分類

「古風な語」の用法を分析するためには，コーパスから得られる用例がどういった文脈の用例であるのかを区別する必要がある。BCCWJ「図書館サブコーパス」より得られる用例は，その文脈により，執筆及び，記述対象の年代に着目して，大きく次の 4 つに分類することができる。

(1) 古典(江戸時代以前の文章)の引用での使用
(2) 明治期から戦前までの使用
(3) 時代・歴史小説での使用
(4) 現代文脈での使用

以下，4 分類の詳細を述べる。

(1)古典(江戸時代以前の文章)の引用での使用

BCCWJは現代語コーパスであるが，収録テキストに「非現代語」(BCCWJでは，明治元年より前に書かれた日本語と定義。)が若干混在している [70]。まとまった「非現代語」は収録テキスト対象外要素として収録しないのだが，一文単位でのテキストの完全収録を保証するために，インライン中に引用されているような「非現代語」は排除せず，そのまま収録している。本論文ではこれを「古典の引用」と呼ぶ。そのような引用中に出現する「古風な語」は「古語」としての使用例である。例えば，次に古典の引用中に現われる「あんずるに【案ずるに・按ずるに】」の例を示す(用例中に用いる，太字，下線及び，括弧内注記は本論文著者による)。

・仏御前は「つくづく物を**案ずるに**、娑婆の栄華は夢のうちの夢、たのしみさかえてもなにかせん」(「百二十句本」巻一〈義王出家〉)と言い、つづいて「一旦のたのしみにほこりて、後生を知らざらんことのかなしさに、今朝まぎれ出でて、かくなりてこそ参りたれ」(同前)と言って、かぶっていた衣をのけると、仏御前は、すでに尼姿になっていたのである。(中石孝，1920 年代生まれ，『平家れくいえむ紀行』新潮社，1999 年)

[70] BCCWJ に収録するテキストの抽出基準についての詳細は，柏野ほか[2009]を参照.

(2)明治期から戦前までの使用

BCCWJ では，明治期以降に執筆されたテキストは現代語のテキストであるとされ，明治期以降のテキストが収録されている。しかしながら，明治期から戦前までに執筆されたものは，例えば，旧仮名遣いを用いるなど，現代とは異なる印象を受ける文体のものが多い。また，「古風な語」が「古風」という意識なしで用いられている可能性があると考える。よって，この時期に執筆されているテキストは BCCWJ にそう多くは収録されていないが，現代文脈とは区別することとする。例えば次のようなテキストである。前者からは「いずれ【何れ】」の，後者からは「はたまた【将又】」の用例が得られる。

・遠野物語の中にも書いてある話は、同郡松崎村の寒戸といふ処の民家で、若い娘が梨の樹の下に草履を脱いで置いたまゝ、行方知れずになつたことがあつた。三十何年を過ぎて或時親類知音の者が其家に集まつて居るところへ、極めて老いさらぼうて其女が戻つて来た。どうして帰つて来たのかと尋ねると、あまりみんなに逢ひたかつたから一寸来た。それでは又行くと言つて、忽ち**何れ**へか走り去つてしまつた。(柳田國男，1870 年代生まれ，『柳田國男全集第 3 巻』筑摩書房，1997 年)

・然らば私の希ふ真の自由解放とは何だらう。云ふ迄もなく潜める天才を、偉大なる潜在能力を十二分に発揮させることに外ならぬ。それには発展の妨害となるものゝ総てをまず取除かねばならぬ。それは外的の圧迫だらうか、**はたまた**智識の不足だらうか、否、それらも全くなくはあるまい、併し其主たるものは矢張り我そのもの、天才の所有者、天才の宿れる宮なる我そのものである。(平塚雷鳥，1880 年代生まれ，『元始、女性は太陽であった　平塚らいてう自伝 1』大月書店，1992 年)

(※いずれも明治期の文章。)

(3)時代・歴史小説での使用

いわゆる「時代小説」「歴史小説」などと呼ばれる，江戸時代以前を舞台とする文芸作品(国内，国外を問わず)のテキストに「古風な語」が多く現れる。「時代小説」とは「古い時代の事件や人物に題材をとった通俗小説。」であり，「歴史小説」とは「過去の時代を舞台にとり、その時代の様相と人間とを描こうとする小説。(中略)単に過去の時代を背景にする時代小説とは異なる。」(以上，『広辞苑第』第六版，岩波書店)と，両者には異なる定義がされているが，本論文では特に両者を区別することはせず，まとめて，「時代・歴史小説」とひとくくりで扱うこととする。

石井[1986]は，例えば「おぬし，・・・でござるか」などは，「歴史小説なり時代小説なりに現はれるからと言つて，その小説の扱ふ時代の古代語と考へるのは，早計である。非現代語すなはち古代的言語を用ゐた作品においては，作家が古代的言語を創造し，読者がそれを享受する，といふ図式が想定できる。」と述べ，そういった享受と創造による「非現代語」が『源氏物語若紫』現代語訳や，日本文芸家協会『歴史ロマン傑作選』の会話文に多く現れることを調査分析し，報告している。本論文では，まさにこれも「古風な語」であると捉える。BCCWJ には「時代小説」「歴史小説」などのテキストが多数収録されており，そういった文脈で用いられる「古風な語」の用例が多く得られる。次に，「あんずるに【案ずるに・按ずるに】」と，「にょにん【女人】」の例を示す。

・成之の天才的な兵站事務の噂をききつけてのことであった。**按ずるに**、成之常に加賀藩の事務に従ひしも、其理財に老けたるの名、夙に朝廷に聞へしを以て、終に此事ありし也。(磯田道史，1970 年代生まれ，『武士の家計簿　「加賀藩御算用者」の幕末維新』新潮社，2003 年)

(※江戸時代を舞台にした時代・歴史小説。)

・廐戸には見せなかったが、河勝はいずれ山背に戻る運命にあることを覚悟していたようだった。

「そうか、河勝も里心がついたか、それもそうだ、河勝は山背に戻れば葛野の王者だ、屋形も大きく、そちにかしずく**女人**も多い、いつまでも吾に仕えてくれる、と思っていた吾が甘かった」(黒岩重吾，1920 年代生まれ，『聖徳太子　日と影の王子 1』文芸春秋，1990 年)

(※飛鳥時代を舞台にした時代・歴史小説。)

(4)現代文脈での使用

執筆時期が戦後であり，上記(3)に該当しない現代語の文章(口語体)を，本論文では「現代文脈」と呼んでいる。一部の見出し語に関しては，その現代文脈に出現する「古風な語」の用例が数多く得られている。例えば次のようなものである。

・ここで泣いては**いかん**、と咽喉の塊を懸命にのみ下しながら、(宮尾登美子，1920 年代生まれ，『朱夏』新潮社，1998 年)

・不審な気持ちが消え**失せ**て、とにかく言葉が交わしたかった。(浅倉卓弥，1960 年代生まれ，『雪の夜話』中央公論新社，2005 年)

・コミック誌から飛び出してきたような**いでたち**の男だ。(佐々木譲，1950 年代生まれ，『新宿のありふれた夜』角川書店，1997 年)

・セブン‐イレブンの店にほしいものがなければ、**いと**も単に、買いにこなくなります。(鈴木敏文|述;緒方知行|編,『商売の創造』講談社, 2003 年)

・恋なのか、忠義なのか、**はたまた**親子の恩愛なのか。(古井戸秀夫, 1950 年代生まれ,『歌舞伎』新潮社, 1992 年)

例えば,「ものども【者共】」には, 分類(1)(3)(4)に該当する次のような用例がある。

・(1)古典の引用： 壇の浦の合戦では、源平両軍は三十余町をへだててあい対し、いよいよ戦闘開始ということになったが、早い潮に流された平家の舟を梶原景時の舟が熊手でひっかけて、敵の首を数多く取る功名第一の働きから始まった。そして両軍あわせて鬨をつくり、それが静まると、平知度が大音声に、「天竺震旦にも、日本わが朝にも、雙びなき名将勇士といへども、運命尽きぬれば力及ばず、されども名こそ惜しけれ、東国の**者ども**に弱気見すな、何時の為にか命をば惜しむべき、軍ようせよ、**者ども**、只これのみぞ思ふ事よ」と全軍に宣した(同上・巻十一)。(阿部猛, 1920 年代生まれ,『鎌倉武士の世界』東京堂出版, 1994 年)

(※下線が古典の引用部分。上記「同上」とは『平家物語』。)

・(3)時代・歴史小説での使用： 久幸ははじめて深い笑みを見せ、「この臆病者が必ずお守り申す。ご安心あれ」といった。そして、「**者ども**続け！」と大声をあげると、私兵五百人を率いて、まっしぐらに大内勢に向かって行った。(童門冬二, 1920 年代生まれ,『小説毛利元就』PHP 研究所, 1996 年)

(※時代小説。この例は室町時代から戦国時代の時代設定。)

・(4)現代文脈での使用： 「なるほど、これは無用のことを申しました…さあ、**ものども**、引っ立てい！」ドラホマノフの命令を受けて、兵士たちがユリアスとパッシェンダールに縄をかける 。さすがに、この人数差では、パッシェンダールといえども抵抗のすべがなかった。(赤城毅, 1960 年代生まれ,『滅びの星の皇子』中央公論新社, 2001 年)

つまり,「ものども【者共】」は, BCCWJ において, 古語としての用例が現われる語であり, 時代小説や歴史小説では, その時代設定にあう語として使われる語であり, かつ, 現代文脈においても使われる語であることがわかる。「古風な語」としてひとくくりする中には, 現代語書き言葉においてこのように幅広い用法で出現する語のあることが, 本調査により確認できた。

5.2.8 「古風な語」の考察

5.2.8.1 使用頻度上位の語

BCCWJ が現代語書き言葉を収集したコーパスであるため，古典の引用の用例数はそもそも少ない。また，先に述べた通り，明治期から戦前までに執筆されたテキストも少ないため，その用例数も少ない。一方，時代小説や歴史小説のテキストは BCCWJ に多く収録されていることから，それらの用例数は多い。また，現代文脈から得られる用例も一定数以上得られている。

第 4 章の 4.1.3 項で頻度調査したもののうちより，使用頻度上位 9 語(「古語的」1 語(いたく)，「古風」8 語(いかん等)，第 4 章 4.1.3 項の表 4.4，表 4.5 を参照)を取り上げ，それらの用例を分類した結果を表 5.15 に示す。上から 5 語は，時代・歴史小説の用例の割合が多い順，次の 3 語は現代文脈の用例の割合が多い順である。そして，最後の 1 語が両者の用例の割合が拮抗していた語である。割合の高いところに色をつけて示している。

このように，BCCWJ において高頻度である「古風な語」は，時代・歴史小説の用例の割合が多いものと，現代文脈の用例の割合が多いものとがあった。同じ「古風な語」とくくるには，用法の傾向は大きく異なっている。つまり，時代・歴史小説の使用が多い，現代文脈の使用が多い，といったその語が使用されている文脈の特徴を辞書に明記し，その用例を具体的に記述することが，「古風な語」とひとくくりに説明するよりも，それぞれの語の文体的特徴を説明することができるようになるということである。

表 5.15 「古風」使用頻度上位 9 語の用例分類結果

見出し	古典		明治～戦前		時代・歴史		現代語		計
	頻度	%	頻度	%	頻度	%	頻度	%	
いかん	1	0.2	0	0.0	59	13.2	386	86.5	446
うせる	5	2.3	4	1.9	29	13.6	182	85.0	214
いたく	1	1.5	3	4.5	7	10.6	55	83.3	66
いでたち	0	0.0	0	0.0	13	16.9	64	83.1	77
はたまた	0	0.0	5	7.7	10	15.4	50	76.9	65
そなた	0	0.0	8	1.8	418	96.3	8	1.8	434
ものども	8	7.7	4	3.8	85	81.7	7	6.7	104
にょにん	16	9.4	9	5.3	102	59.6	44	25.7	171
わい	0	0.0	3	1.9	71	44.7	85	53.5	159

5.2.8.2 辞典間の記載に差異のある語

第 4 章の 4.1.3 項では，5 種の辞典間に，見出し語として採用されている数や，古さについてのラベルや注記が付与されている数が異なる語があることを述べた。第 4 章の表 4.2 の「まがまがしい【禍々・枉々・曲々・凶々しい】」と，表 4.3 の「よしなに【良しなに】」の 2 語は，見出し語として採用しているのは 5 辞書であるが，注記は 2 辞書ずつという語であった。また，表 4.2 の「やわか」は，見出しとして採用し，注記を付与しているのが 2 辞書のみ，という語であった。

第 4 章 4.1.3 項の表 4.4，表 4.5 に示した通り，BCCWJ には，「まがまがしい」に 33 例，「よしなに」に 7 例の使用がある。そのうち，2 例ずつを下記に引用する。

・それは途方もなく暗鬱な感じの建物で、**まがまがしい**気配にみち、見物するのは愉快な体験ではなかつた。(丸谷才一，1920 年代生まれ，『日本の名随筆』作品社，1988 年)

・「お気をおつけ下さい」 ヒロはいう。「今夜は、**まがまがしい**気配が満ちております。その心を感じるのです。ザック様のことですから何がおころうときっと大丈夫でしょうが…心配でなりません」(眉村卓，1930 年代生まれ，『迷宮物語』角川書店，1986 年)

・「よう冷えるなあ。こらたまらんワ。泰平堂、まあ**よしなに**調べてくれ」 声をかけて信濃は去っていった。(阿部牧郎，1930 年代生まれ，『出合茶屋』講談社，2003 年)

・「お世話をかけます、なにとぞ**よしなに**」 懐しさを顔一ぱいにみせてお茂の方は、頭を下げた。(竹内勇太郎，1920 年代生まれ，『甲府勤番帖』光風社出版，1992 年)

「まがまがしい」は 33 例中，5 例が時代・歴史小説の用例であったが，残り 28 例は上記 2 例のような現代文脈の使用であった。「よしなに」は，7 例中のすべての例が，上記 2 例のような時代・歴史小説の用例であった。どちらにも時代・歴史小説の用例のあることから，「古風な語」という注記を付与すべき語であることがまずはわかる。それに加え，「まがまがしい」は現代文脈での使用が多く，「よしなに」は時代・歴史小説での使用が多い語であるということもわかる。

続けて，2 辞書のみ見出し語に採用していた「やわか」の使用例をみてみる。BCCWJ から得られる用例 3 例中より，2 例を以下に示す。

・「この胴も手足も、蓬萊山から出た鉄を百年も磨き抜いてこしらえたものよ。項羽の豪刀をもってしても傷ひとつつかなんだ。**やわか**、おまえの妖糸ごときに引けは取らぬぞ」「そりゃ、どうも」(菊地秀行，1940 年代生まれ，『夜叉姫伝』祥伝社，1991 年)

・「いかなる御用とて、われらにおきかせあれ！　拙者、**やわか**島田虎之助の働きに劣りましょうや」と、斎藤弥九郎が、いかにも体調の悪いらしい島田虎之助をあごでさして、そのあごをお耀の方にぐいとつき出す。(山田風太郎, 1920 年代生まれ,『武蔵野水滸伝』富士見書房, 1993 年)

こういった用例に触れて国語辞典をひく読者がいることを想定すると,「やわか」は「古風な語」として現代語辞書に取り上げ, その用例を記述すべきであることがわかる。

このようにコーパスで使用例を確認することは, 辞書によって扱いに差異のある語の辞書記述の判断の参考になることを示した。

5.2.9 コーパスベースの使用域の辞書記述の手順

前 5.2.8 項で,「古風な語」の使用実態についてコーパスの分析結果を示した。本項にて, そのコーパスベースの辞書記述を提案する。

「古風な語」の用法は, (1)古典の引用での使用, (2)明治期から戦前までの使用, (3)時代・歴史小説での使用, (4)現代文脈での使用, と分類でき, かつ, その使用傾向が語の用法把握につながることを, 5.2.8.1 目と5.2.8.2 目とで示した。その際に, 特に, 分類(3)と(4)に相当する用法の具体的記述が重要であることを述べた。(3)の, 主に時代小説や歴史小説での使用は, 現代においてそれら小説を読む際の理解に欠かせないものである。現代語ではないと排除することなく, 辞書にとりあげ, 詳細に記述すべきものであろう。さらに, (4)の, 現代文脈での使用は, 一部の「古風な語」における顕著な特徴である。よって,「古風な語」の現代文脈における使用例がある場合は, それを具体的に辞書に記述すべきである。

そこで, 次の通り, コーパスベースの「古風な語」の使用域の辞書記述方法を提案する。

1. 現行の国語辞典類や, BCCWJ,『青空文庫』,『太陽』等のコーパスを利用し, 次の条件を満たす,「古風な語」を選定する。
   (a)「時代・歴史小説」を含めて現代で使用が見られる。
   (b) 明治期以前, あるいは, 戦前までの使用が見られる。
2. BCCWJ,『青空文庫』,『太陽』等のコーパスから,「古風な語」の使用頻度, 用例を得る。
3. 多義語の場合, 意味分析を行い, 意味別の使用頻度を得る。
4. 得られた用例を, (1)古典の引用での使用, (2)明治期から戦前までの使用, (3)時代・歴史小説での使用, (4)現代文脈での使用, に分類し, 各頻度を得る。

5. 分類別の使用頻度を参考に，中心的となる語義から順に配列する。

6. 用例は，用例の(1)〜(4)の分類傾向や，具体的な使用域がわかるよう明記する。

7. その他，表記情報や，使用者の性別・年代など，コーパスから抽出できた情報を明記する。

5.2.10 コーパスベースの使用域の辞書記述の試行

「**要件 2-2** 語の使用域・位相の情報が得られること」を満たすために，「古風な語」の使用域のわかる辞書記述方法を提案した。本提案による記述により，従来の国語辞典に比べて語の使用域の精緻化が可能になるかを確認する。

5.2.10.1 記述方法

「古風な語」の条件(a)，(b)を満たす「そなた【其方】」を例に，現行の国語辞典の記述(『岩国』)に対する，コーパスベースの記述案を示す。

5.2.10.2 記述結果

「そなた【其方】」を例に，現行の国語辞典の記述(『岩国』)に対するコーパスベースの記述案を次に示す。「そなた」は，上記，(a)，(b)の条件を満たす。

例：そなた【其方】

[『岩国』(第七版)]

①そちらのほう。そちら側の所。

②目下の相手を指す語。おまえ。なんじ。▽古風な言い方。

[コーパス分析]

・「そなた【其方】」はBCCWJで434の頻度があり，「古風な語」の中では高頻度である語である。

→見出し語の選定： 十分な頻度が得られるため見出し語として採録すべき語と判断。

・『岩国』①の該当用例はBCCWJでは頻度0。『青空文庫』には3例見つかる。BCCWJで得られる434の用例は，すべて『岩国』②の用法の該当例である。

→語義の選定・配列： ①と②を語義と認定。ただし，①の用法はBCCWJで0であり，現代語では，②の用法がより中心的な語義と認定できるため，語義の配列順を入れ替える。

・『岩国』②の用例分類は4.3節の表4の通り，ほとんどが「(3)時代・歴史小説での使用」である。国内の時代小説のほか，時代設定の古い翻訳小説の用例もある。現代文脈での使用は少ないが，ある。また，BCCWJ で得られた用例はすべて発話部分における使用。

→用例記述・見出し語の使用域：(3)時代・歴史小説での使用例を1番に。続けて，翻訳小説，現代文脈の使用例を記述。「発話で多用」を明記。発話については話者間の関係を明記。

・漢字表記「其方」の使用例は見つけられない。検索される「其方」は「そちら」「そのほう」の漢字表記の使用例と思われる。

→表記情報を注記。

[「コーパスベース国語辞書」記述案]

①多くは目下の相手をさす語。おまえ。なんじ。▽時代・歴史小説的。同等の相手をさすこともある。

・時代・歴史小説等の発話で多用。

「そなた、勘六どのを見舞って来やれ。さっきの飴の甘味はきつすぎます。よけい食べさせてはなりませぬ。急いでゆきゃれ」(山岡荘八『徳川家康』)

※後に家康を産む於大の方(目上)から，召使いの百合(目下)への発話。

「ほう、そなたもさようなことを考えておったのか？」豪族は大きく頷いた。(山田智彦『木曽義仲』)

※信濃国の各地から集まってきた豪族たち同士(同等)の対話。

・時代設定の古い小説(特に翻訳小説)における発話で使用。

「そなたは誰なのか、教えてくれ」。(井村君江『アーサー王物語』)

※アーサー王(目上)から，騎士であるトリストラム卿(目下)への発話。

「ねえ乳母や、そなたの申し条、もっともなことです。ただ怒りのあまり分別を失っていたのです」と言いました。(池田修『アラビアン・ナイト』)

※王女である姫(目上)から，老女の乳母(目下)への発話。

・現代文脈での使用は多くはないが，威厳や威圧があるよう造形された人物(多くは目上)からの発話を表すものとして使用。

「厚志よ」

舞が呼びかけると、厚志はビクリと身を震わせた。

「え、なに？」

「何を驚く？　そなたの名を呼んだだけだぞ」

(榊涼介『ガンパレード・マーチ 5121 小隊九州撤退戦』)

※小隊司令官である舞(目上)から，その下にいる厚志(目下)への発話。

②そちらのほう。そちら側の所。▽明治期頃まで。

・使用はまれ。

五月雨に四尺伸びたる女竹(めだけ)の、手水鉢(ちょうずばち)の上に蔽(おお)い重なりて、余れる一二本は高く軒に逼(せま)れば、風誘うたびに戸袋をすって椽(えん)の上にもはらはらと所択(えら)ばず緑りを滴(したた)らす。

「あすこに画がある」と葉巻の煙をぷっとそなたへ吹きやる。ぷっとそなたへ吹きやる。(夏目漱石『一夜』)

恋の淵・峯の薬師・百済の千塚など、通ひなれては、そなたへ足むくるもうとましきに、折しも秋なかば、汗にじむまで晴れわたりたる日を、たゞ一人、小さき麦稈帽子うち傾けて、家を出でつ。(折口 信夫『筬の音』)

※漢字表記「其方」の使用例は見つけにくい。該当表記例があっても，別見出し「そちら」「そのほう」との区別がつけがたい。

以上の記述案により，「そなた【其方】」の用法は，時代・歴史小説の発話文での使用が中心的な語であることが明確になる。多くの用例を示したことにより，現代の文章生成時の語選択においては，時代がかったセリフとなる効果のある語であることに留意すべき語であることがわかる。さらに，漢字表記の【其方】はほとんど使用例がなく，かつ，「そちら」「そのほう」とまぎれる可能性があるので，ひらがな表記を選択すべきこともわかる。

5.2.10.3 記述結果のまとめ

「そなた【其方】」の記述例を通し，従来の辞書記述よりも，文体的特徴や用例を辞書記述に反映させることは，語の理解はもとより，語の選択時の情報量を増やし，その利用価値を高めることが期待できることを確認した。

なお，中世・近世のコーパスがない状況においては，国語史研究の知見を参照することが有効である。例えば，「そなた」は中世より，主に上位の話し手が，下位の話し相手に対して使用する例の多い人称の 1 つと位置づけられている。中世から近世にかけての使用実態については，山崎[2004]，小島[1998]に詳細な分析があり，上位から下位に用いられ，時には対等の間柄でも用いられるものであると報告されている。「上位から下位」の例，「対等」の例がともにコーパスから得られたため，辞書記述案ではそれら先行研究の知見を活かし，「主に上位から下位へ」「時には対等」で用いられることがわかるよう注記することも有用である。

国立国語研究所では，将来的に上代から近代の作品をカバーする『日本語歴史コーパス(CHJ)』[71]の構築が進められている。それらコーパスの整備に伴い，コーパス分析の可能性が広がることが期待される。

記述方法の 3 にあげた意味判別に関しては，例えば，Kathuria，白井[2012]など，すでに自動化の研究が進んでいる。おそらく，4 にあげた当該語の用いられる文脈が分類(1)～(4)のいずれであるかの判別についても，今後，自動化が可能になると考えられる。

[71] http://www.ninjal.ac.jp/corpus_center/chj/

5.3　外来語の時間的使用推移の詳述

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

経年調査による語彙調査には，雑誌用語の 10 年ごとの変遷調査[国立国語研究所 1987]，明治 37 年～昭和 24 年の国定読本の語彙調査[国立国語研究所 1997]，雑誌用語の語種構成を約 40 年の間隔をおいて比較した分析[山崎，小沼 2004]がある。しかしながらいずれも規模はそう大きくはない。

電子テキスト化された大量の各種の新聞記事データベースが研究に利用できるようになって以来，これを用いて時間的使用推移を大規模に分析する研究が現れてきた。毎日新聞や朝日新聞におけるカタカナ語の月次変動の分析[久野ほか 1998，野崎ほか 1998]，毎日新聞における時系列変動と文字使用の面種差の分析[久野 2000]，毎日新聞と朝日新聞における時系列変動と文字使用の紙面差の分析[久野ほか 2002]，毎日新聞における出現頻度の時間的変化の語種による分析[山口ほか 2004]である。山口ほか[2004]では，毎日新聞 9 年分を対象に調査し，外来語はほかの語種よりも変動係数の平均，分散が大きいことを報告している。そして外来語の特徴として，高頻度の形態素において，頻度変化の大きい語が多いと考えられることを指摘している。これらのほか，新聞記事データベースを対象にした大規模な頻度調査に朝日新聞 14 年分の頻度調査[NTT 2000]もある。このように，新聞記事を用いた大規模な経年調査は最近になって増えているが，いずれも，文字や語，語種などについて，全体的な変動を分析したものであり，個々の語の使用推移に着目して各語の使用状況を捉えようとする分析はまだなかった [72]。特定の語に限れば，国語学において語の使用推移に着目する研究はないわけではない。しかしながら大規模な語彙調査によって時間的使用推移を研究するものはこれまで報告されていなかった。

国立国語研究所では 2002 年 8 月に「外来語」委員会を設置し，主に，省庁の行政白書に使用されている外来語のうち，一般への定着が不十分でその意味がわかりにくいと思われる語を対象

[72] 本調査の後には，大正以降の新聞社説資料を中心とした経年調査により，外来語の増加過程が典型的な成長現象に見られる「S 字カーブ」のパターンを描くことを明らかにした報告が出ている[橋本 2010]。

に，わかりやすく言い換えるなどの対応策を世の中に提案している[73]。本節では，同委員会が第1回，第2回の検討対象語として選定した 109 語について使用状況を定量的に明らかにする。即ち，その 109 語を対象に，16 年分の新聞記事データベースを用いて，出現率の推移に着目して語の時間的使用推移を捉える。なお，対象語すべての調査結果は 5.3 節末の附表 5.1 を参照されたい。

5.3.1 調査の問題点への対処

5.3.1.1 読売新聞記事データの概要

本調査には読売新聞の 1986 年から 2001 年までの 16 年分の記事データを使用した。データの概要は次の通りである。

- 読売新聞社の記事データベースに蓄積してあるもののうち，著作権が読売新聞社に無いものを除いた全記事。
- 1990 年以降は東京本社発行記事に加え，地方支社(大阪，西部，中部)発行記事が入っている。
- 1999 年以降はさらに県版(地域版)の記事が入っている。
- 新聞紙面記事との違いは，次の 2 点である。
    ① 見出しをデータベース用に再編集している。
    ② 紙面上で見開き 2 ページにまたがってレイアウトされた記事をデータベースに採録する際には，ページごとに分けて採録せず，1 つの記事としてまとめて右ページにある記事として採録している。そして，左ページにあるものとして，見出しのみ，もしくは見出しと本文の一部のみを採録している。その結果，左ページ入力分は右ページ分との二重採録になっている。

読売新聞記事データには本文や見出しのほかに，キーワード(統制語及び自由語)，分類コード，面種コードなどの書誌情報も入っているが，今回の調査で用いたのは本文のみである。調査に用いた本文の文字数のグラフを図 5.17 に示す。図 5.17 より記事データ量が年々増加したことが明らかである。特に，1989 年と，1999 年とを境に二度にわたってデータ量が大きく増加している。

[73] 「外来語」言い換え提案の詳細は，http://www.ninjal.ac.jp/gairaigo/。

これは，1990 年より支社発行記事が加わり，1999 年より県版(地域版)記事が加わったことによるものとみられる。しかし，最初の大きな増加は 1990 年ではなく，1989 年に起きている。それは，1990 年より地方支社発行記事が紙誌コードを分けて正式に登録されているが，これに先立ち，1989 年に地方支社関連記事が本社記事として，面種コードを分けて登録されたことによると推測される。

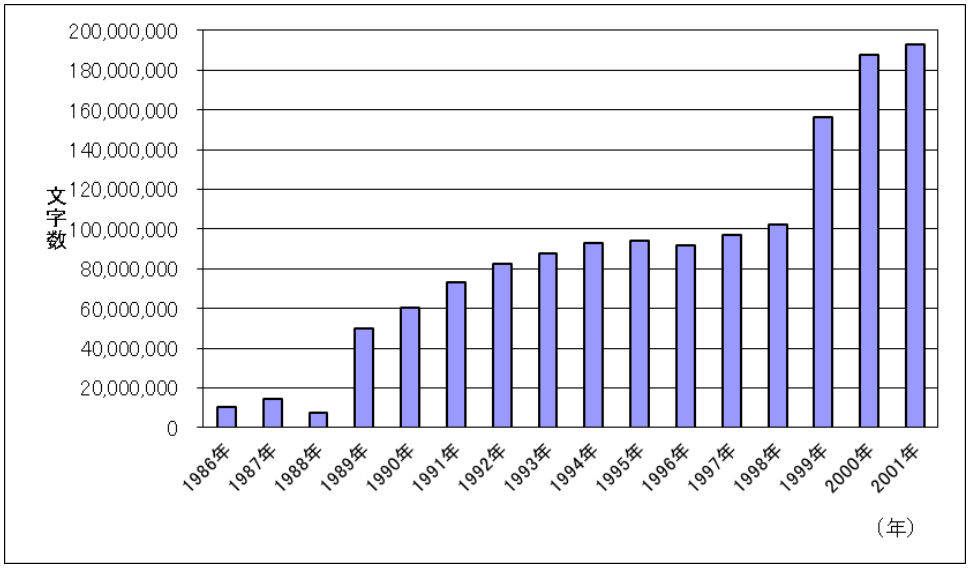


図 5.17 読売新聞記事データ文字数

5.3.1.2 使用状況の把握における問題点

はじめに，新聞記事の特徴といわれている事柄を押さえる。新聞記事はニュースを伝えることを目的とした文書である。新聞のニュースについて佐藤[1991]は，「その日の相対的な重要度によって，ニュースを価値判断し，秩序だった情報に再構築して，その日の最重要ニュースを最低限押さえられる」ようになっているものであると言っている。また，後藤[1995]は，新聞は「内容の上でかなり広い範囲が扱われてはいるが，政治，経済，社会の分野に大きな比重が置かれており，したがって，それに関連した単語が過大な割合で出現する。」と述べている。今回の調査において明らかにできる語の使用状況は，ニュース性の影響を受けているものであることを念頭に置く必要がある。今回の対象語のうちで，特にニュース性の影響を大きく受けていると考えられた語については 5.3.4 項で取り上げる。

(1) データ量増加の影響

読売新聞記事データを使用する際，使用状況を捉える上で問題になる点の 1 つ目は，年により記事データの量が異なる点である。このため本論文では，使用状況の指標として使用頻度だけではなく，より使用状況を的確に表す数値として，出現率，即ち使用頻度を収録単語数で正規化した値に着目する。ただし，平均単語長は年によらずほぼ一定とみなせるので，本論文では収録単語数の代わりに，より算出の容易な収録文字数を用いて正規化することとした。即ち，次の式によって，年ごとの出現率を算出した。

出現率＝使用頻度÷収録文字数

(2) 用例の重複

2 つ目の問題点は，重複記事と，それにより抽出される重複用例を重複して数えるかどうかである。ここでいう重複記事とは主に次の要因によって生じているものを指す。

① 本社発行と支社発行とがあるため。
② 複数の県版(地域版)があるため。
③ 見開きページで作られた記事をデータベースに登録する際に意図的に部分的に二重に採録するため。
④ 1 週間あるいは 1 か月のダイジェストとして再掲載するため。
⑤ 選挙広報，10 大ニュース募集，など，意図的に複数回掲載するため。

今回の対象語で重複例を調べたところ，多いところでは「ケア」で 23,512 件中，236 件(約 1%)，「シェア」で 13,637 件中，91 件(約 0.6%)見つかった。この数字を大きいと見るか小さいと見るかの判断は難しい。しかしながら，もし重複を用例単位で除こうとすれば調査母体が対象語ごとに異なってしまう。記事単位で除こうとすれば調査の対象記事データがかなり限られてしまう。このため，本論文では，重複記事を特に排除しないこととした。

実際のところ，重複の排除は判断に迷うものが多く，作業は困難である。例えば掲載日が同じで，内容がほとんど同一であり，発行だけが違うようであれば比較的重複であるとの判断がしやすいが，そのようなものは少ない。地方，地域によって日を変えて掲載されている場合も多く，掲載日だけでは判断できない。また，重複の度合いが様々であり，異なる部分が多くなるほど重複と判断しがたくなってくる。文レベルでは，全く同一のもの，句読点のみ違うもの，表記，語句が少しだけ違うものから，語句がかなり違うものまである。記事レベルでは，ほとんど同じものもあれば，一方

が詳細記事のためそちらだけに存在する文が多くあり，記事全体での重複文は実際には少ない場合もある。同じテーマで次々に新しい記事が書かれる場合にも文や語句単位で重複は多く生じるが，この場合と問題にする重複記事とを区別する必要もある。

なお，このような重複は読売新聞に限らず，毎日新聞や日経新聞にあることも報告されている[Aizawa2003]。今後，Aizawa [2003]のように，機械的に重複を調べる研究も進んでくるであろうが，そもそも新聞記事における重複とは何であるかの議論も進める必要がある。

5.3.1.3 対象語の同定における問題点

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

例:アイデンティティー 全 2,462 件
「アイデンティティー」1,987 件
「アイデンティティ」472 件
「アイデンティーティー」1 件
「アイデンティー」1 件
「アイデンテティ」1 件

また，記事データに対し形態素解析を施さずに，文字列マッチングのみによって用例を抽出したこともあり，部分的に文字列が同じである別語の用例が抽出されることが多々あった。これには人手によって排除することによって対処した。

例:スクリーニング 全 311 件
「ハウスクリーニング」
「ガラスクリーニング」など 71 件を排除

5.3.2 出現率の調査

まず，対象 109 語について読売新聞記事データにおける各年の使用頻度，出現率，および使用頻度の総計を出した。

16 年間の使用頻度の総計が 1 万を超えたのは次の 6 語のみである。頻度の総計とともに示すと,「ケア 23,512, シェア 13,637, ビジョン 11,170, アクセス 10,673, ベンチャー10,558, ガイドライン 10,251」である(以下, 必要に応じて頻度の総計を語に添えて示すことにする)。半数以上の 65 語は使用頻度の総計が 1,000 にも満たなかった。そのうち 100 にも満たなかったものが 9 語ある。「ストックヤード 99, フィルタリング 87, タスク 67, ログイン 29, ハーモナイゼーション 22, サマリー13, バックオフィス 8, エンフォースメント 5, トレーサビリティー5[74]」である。一般にわかりにくいと思われる外来語は, おおむね, 新聞においてもその使用頻度は大きくない語であることを確認できた。

また, 109 語のうちの約半数にあたる 48 語が今回の調査期間内においての初出語 [75]であったことも確認できた。語が定着するまでには時間を要するものであると考えれば, 新しい語ほどわかりにくい語であると推察と推察される。

5.3.3 出現率の推移の分析

次に, 16 年間における出現率の推移を調べた。頻度 50 未満の 6 語(前項参照)は初出が 1989 年以降の新語であり頻度が少なすぎて推移の傾向を捉え難いため除外し, 残り 103 語について出現率の増減傾向を捉えた。103 語の頻度は, 最大 23,512 から最小 67 まで開きがあり, 頻度の小さなものの出現率の変化は, 頻度の大きなものに比べると非常に小さなものであった。しかしながら, たとえ小さくても増減傾向を捉えることが語の使用状況の把握には意味があると考え, 出現率の大小に関わらず, すべての語に対して推移の型を導き出した。

表 5.16 出現率推移の型による分類

分類	判断基準	語数
平行型	平ら:出現期間内をならすと平ら	17
増加型	右上がり:出現期間内をならすと増加	44
減少型	右下がり:出現期間内をならすと減少	13
	山:3 年以上の増加後 3 年以上の減少	10
	山の後平ら:増加, 減少, 平らが各 3 年以上で連続	4
凹凸型	突出:特定の年のみ増加	10
	上記以外	5

[74] このトレーサビリティーは BSE 問題に絡み, 調査対象後の 2002 年以降に使用が大きく増えている。
[75] 少なくとも 1986 年の使用頻度が 0 であれば 1987 年以降の初出と見なした。

出現率の増減傾向を分類した判断基準と型を表 5.16 に示す。判断基準は 7 通り設けたが，その一部をまとめて，分類の型は 4 通りにした。

本論文では，これら 4 つの型別に，各語の使用頻度と「理解率」とを用いて分析をした。「理解率」とは，文化庁の協力によって「外来語」委員会が表 5.17 に示す外来語の定着度を測る調査(定着度調査)を実施して得た数値であり，その調査結果は公表されている[76]。

表 5.17「定着度調査」の概要

母集団	全国 16 歳以上の男女個人
抽出方法	層化二段無作為抽出法
調査方法	調査員による個別面接聴取法
調査時期	2002 年 11 月から 2003 年 10 月
語数と人数	120 語については 750 人，210 語については 2115 人または 2118 人
調査実施	各回 30 語ずつ回数を分けて実施
回収率	平均で 70.4%

質問文は次の通りである。

a. あなたは，ここにあげた(1)から(30)の言葉を，聞いたこと，また見たことがありますか。

b. (聞いたこと，見たことが「ある」と答えた人に)それでは，その言葉の意味は分かりますか。

c. (言葉の意味が「わかる」「何となくわかる」と答えた人に)自分でその言葉を使ったことがありますか。

このうち，b の質問に対して「わかる」と答えた人の比率が「理解率」と呼ぶ数値である。

[76] 「定着度調査」の詳細は，http://www.ninjal.ac.jp/gairaigo/Yoron/ を参照。

最初に，4 つの型別に，そこに分類された語の理解率が占める比率を図 5.18 に示す。理解率は 4 段階(25％未満，25～50％未満，50～75％未満，75％以上)に分けて示している。

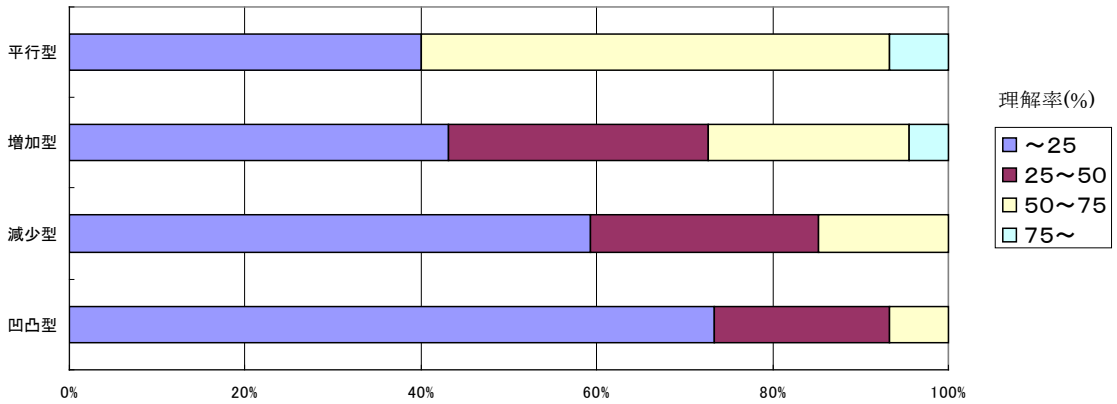


図 5.18 型別に見る理解率の比率

そもそも，今回の調査は一般にわかりにくいと思われる外来語を対象にしている[77]ため，おおむね調査結果の理解率は低かったが，それでも，理解率は各語によって異なり，また，推移の型ごとに理解率の高低が異なっていた。図 5.18 では平行型，増加型，減少型，凹凸型の順に，理解率の高い語の占める比率が小さくなっていることが確認できる。これを統計量で明らかにしたのが表 5.18 である。

表 5.18 型別に見る理解率の統計的特徴

	理解率の平均	理解率の標準偏差	変動係数(標準偏差／平均)
平行型	43.14	20.42	0.47
増加型	34.57	21.75	0.63
減少型	26.42	16.80	0.64
凹凸型	19.63	14.63	0.75

77 「外来語」委員会では，理解率が 75%未満の語を「分かりにくい外来語」として扱っている。中には，理解率が 75%を越えた語もあるが，その場合も，国民全体ではなく，60 歳以上に絞れば，75%未満であった。

表 5.18 より，理解率の平均は，平行型，増加型，減少型，凹凸型の順に高いことがわかる。また，それぞれの標準偏差の値から，型により理解率は有意に異なっていることがわかる。さらに，一番安定していると思われる平行型の変動係数が最も小さく，一番不安定だと思われる凹凸型の変動係数が最も大きくなっていることが確認できる。

次に，図 5.19 に，4 つの型に分類されている各語の理解率と，使用頻度とを示す。この図ではっきり確認できる点は，理解率と頻度の値がともに小さいところには，平行型の分布がほとんどない点[78]と，ともに大きいところには増加型と平行型のみが分布しているという点である。

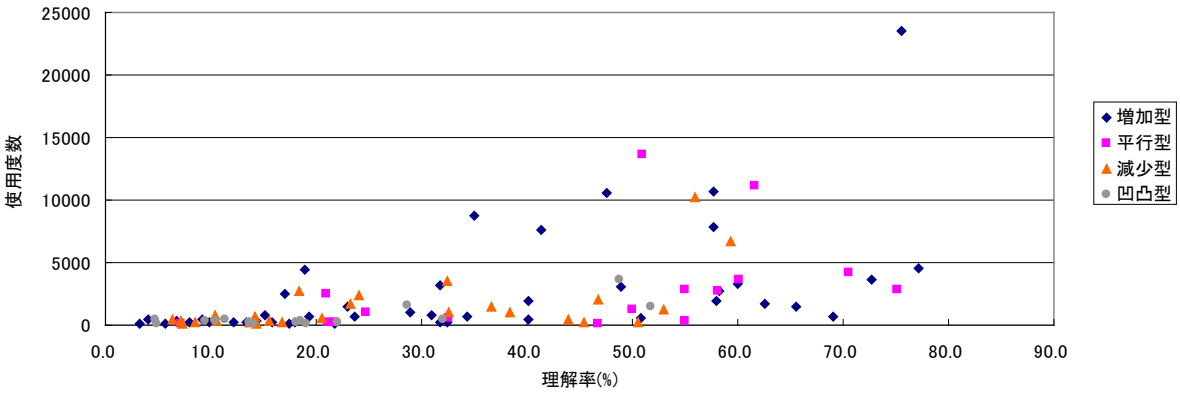


図 5.19 出現率推移の型別に見る理解率と使用頻度

5.3.4 考察

以下，型別に考察する。例を示す図 5.20～図 5.27 の各グラフは，推移をわかりやすくするため，出現率の大きさによって Y 軸の目盛り幅を調整している。

5.3.4.1 平行型

この型を示す語は全部で 17 語あるが，約半分の 8 語が頻度 2,000 以上で，残り半分の 9 語が頻度 100～1,300 であり，それぞれで異なる特徴をもつ。

前者に該当するのは，図 5.20 に示す「シェア 13,637」のほか，「ビジョン 11,170，バックアップ 4,154，マーケティング 2,748，アイデンティティー2,462」などである。これらは早くからある程度定着傾向を見せており，毎年同じように使用されてきていると見られる。おおむね理解率の調査でも

[78] 正確には，平行型としては例外的に，理解率 7.1%，度数 139 の「フェローシップ」の 1 語のみ理解率と度数の値がともに小さいところに位置している。このことは，5.3.4.1 目で言及する。

50%以上の高い数値がでている。ただし，この中では「アイデンティティー」のみ理解率が 20.9%と低い。この語は使用頻度や推移の型では定着傾向を示しつつも，語の表す概念の難しさによって，実際のところは日常語，生活語とは言いにくい語のようである。

後者に該当するのは，図 5.21 に示す「コア 1,034」のほか，「ライブラリー1,290，モータリゼーション 227，デリバリー168，フェローシップ 139」などである。理解率は「フェローシップ」のみ 7.1%と非常に低いが，それ以外は 20～50%未満であり，高くはないが，非常に低いというものでもない。これらは，目立って使われることはあまりないが，かといって，使われなくなる様子も見られない。継続的に使用されている語であると言える。

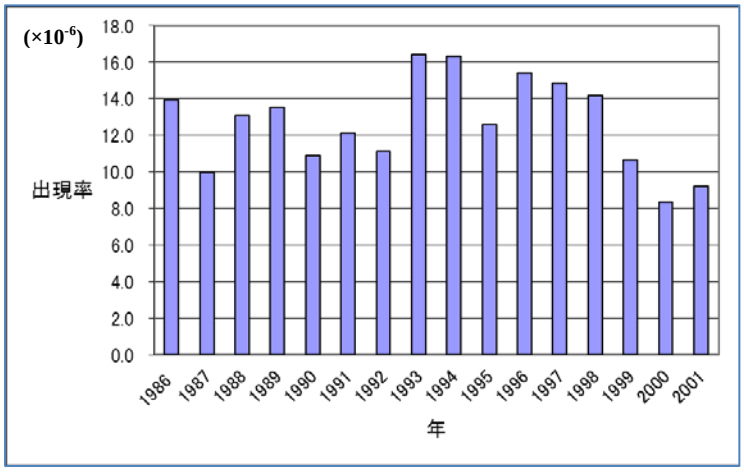


図 5.20 「シェア」出現率(×10^{-6})

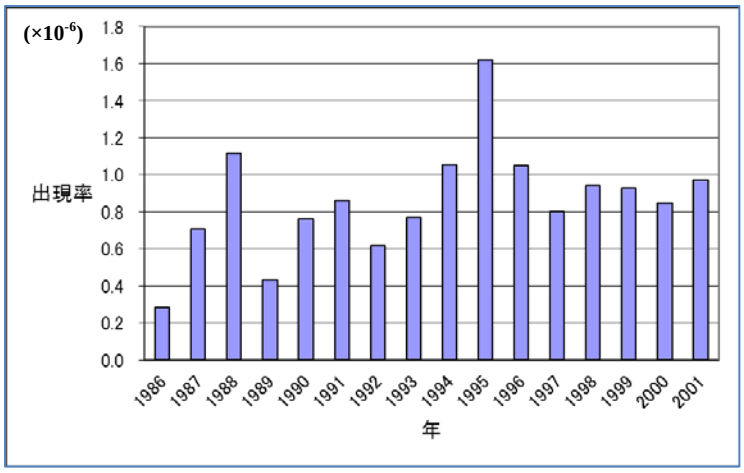


図 5.21 「コア」出現率(×10^{-6})

5.3.4.2 増加型

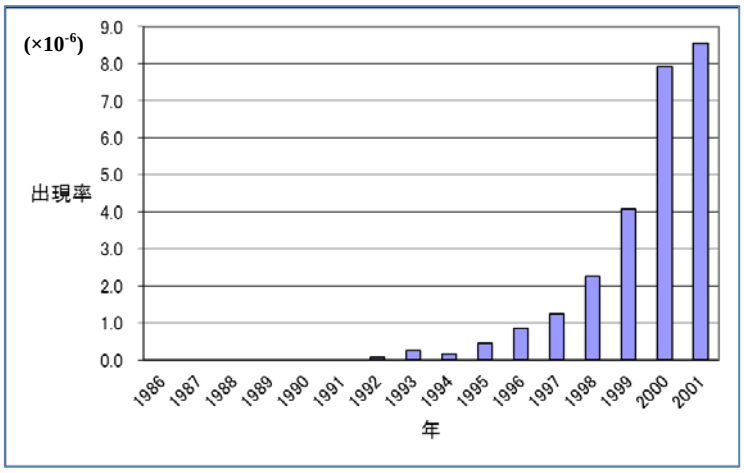


図 5.22 「バリアフリー」文字出現率(×10-^{6})

出現率が増加しているこのタイプにもっとも多くの語が分類された。最大頻度の「ケア 23,512」から「ストックヤード 99」まで 44 語あった。このうち，典型的な増加傾向を示す「バリアフリー3,656」の例を図 5.22 に示す。

この増加型を示す語は，近年，その語の表す概念が取り上げられることが多くなってきたと考えらえる。即ち，ニュース性の高い語であると考えられる。例えば，「ケア 23,512，デイサービス 4,585，バリアフリー3,656，ノーマライゼーション 277，ノンステップバス 217，セカンドオピニオン 205」といった，福祉や医療に関わる語がこの増加型であった。今回の対象語 109 語のうち，福祉や医療に関する語でこの増加型に入らなかったのは「インフォームドコンセント 1,732」(山型)，「メンタルヘルス 408，スクリーニング 311」(突出型)の 3 語のみである。「インフォームドコンセント」は 1997 年までは増加傾向を示しており，「スクリーニング」は BSE 問題に絡み 2001 年に突出して使用が増えていることをあわせて考えると，ほとんどが増加傾向を示したことになる。また，インターネットの普及とともに使用が増えてきていると思われる「アクセス 10,673，コンテンツ 1,422，アーカイブ 227」などもこの増加型を示した。

使用が増えているということは定着に向かっていると見ることもできる。ただし，中には定着が不十分なうちに，その使用だけが増加している語もある。理解率を見ると，頻度が 3,600 以上と大きい「デイサービス，ケア，バリアフリー」といった語は，順に 77.2%，75.6%，72.7%，と高いのに対して，頻度が 1,800 未満である「インフォームドコンセント，ノーマライゼーション，コンテンツ，アーカイブ」は順に 23.2%，12.2%，23.0%，8.0%，と低い。なお，「アクセス」は頻度が 10,673 と大きい割に理解率は 57.7%と，「デイサービス，ケア，バリアフリー」などと比較してそう高くない。

5.3.4.3 減少型

一般には外来語の使用は増えているという印象が強いが，今回の対象語の中には減少傾向を示す語があった。この 1990 年頃より減少しているものが「アメニティー576, フレックスタイム [79]492, サーベイランス 408」など 13 語，一度増えていき，その後減少に転じているものが「ガイドライン 10,251, マルチメディア 6,739, フィルタリング 87」など 10 語，一度増えて減った後に平らに推移しているものが「アナリスト 2,756, トレンド 1,234」など 4 語あった。図 5.23～図 5.25 に一例ずつ示す。

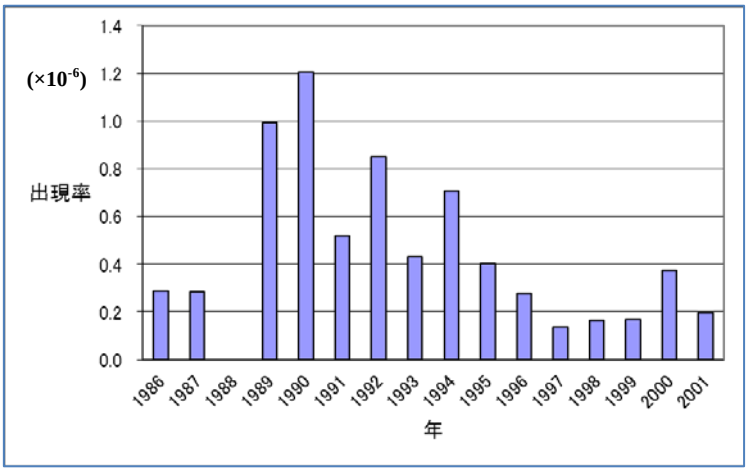


図 5.23「フレックスタイム」出現率(×10^{-6})

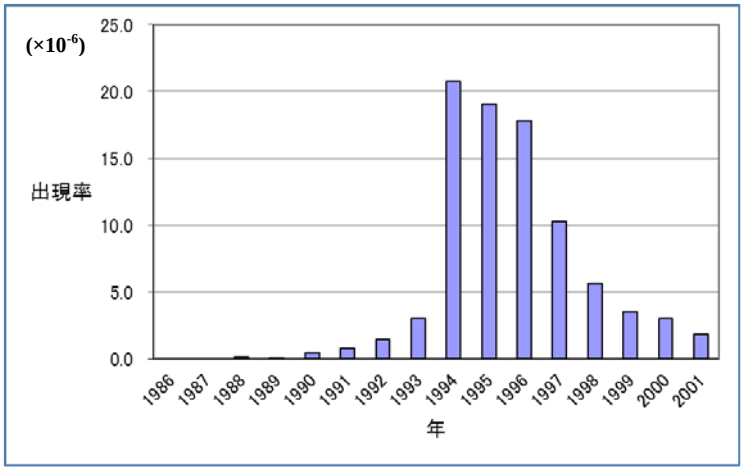


図 5.24「マルチメディア」出現率(×10^{-6})

[79] 「フレックスタイム」の場合は，念のため短縮形の「フレックス」のみの使用についても調査した。その結果，「フレックス」のみの使用頻度は全部で 47 と少なく，その使用を加えても「フレックスタイム」の出現率の推移の型はほとんど変わらなかった。短縮形が問題になりそうな語はこの減少型に関わらず対象語全体のうちに特にはなかった。

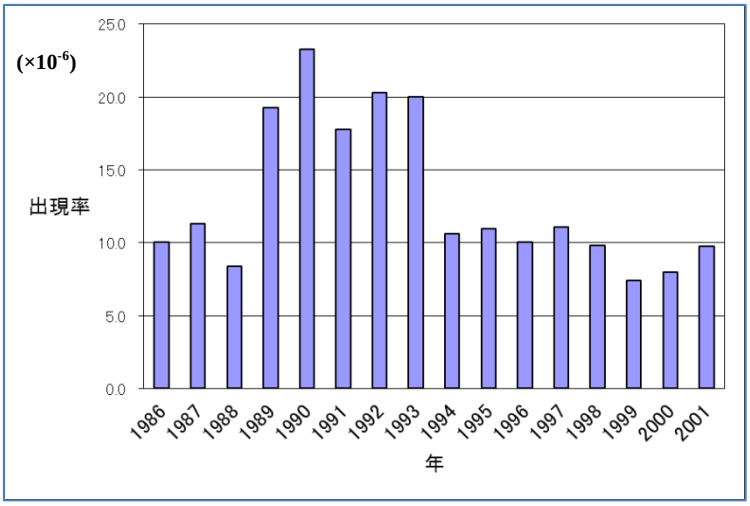


図 5.25「アナリスト」出現率(×10^{-6})

これらには，何らかの政策や運動などを契機として，一時期新聞を賑わせたが，その後話題に上ることが徐々に少なくなり減少している語が該当すると考えられる [80]。しかしながら，一時期減少型を示す語のすべてがそのまま定着せずに消えていくのではない。推移の型が増減後に平らになっているものも見られるように，ニュース性はなくなるものの，社会生活や日常生活においては少しずつ定着していく語もあると考えられる。平らな部分がそのまま安定的に推移していくのであれば，平行型として捉え直すべきものとも考えられる。

なお，今回ここに分類された語は，頻度が 2,000 以上ある語も含めて，ほとんどが理解率 50%に満たず，現時点においてはいずれの語も定着に至っていないと思われる。

5.3.4.4 凹凸型

これまでの型とは異なり，推移に連続した増減のないものがある。そのうち，出現率がほかと比べて非常に突出している年のあるものがある。細分類ではその突出がより顕著なものを突出型としてそれ以外のものと分けた。「ダンピング 3,691，インサイダー1,546，ライフライン 1,485，メンタルヘルス 408」など 10 語を突出型に，「コミットメント 508，ケーススタディ 183」など 6 語をそれ以外に分類した。各例を図 5.26，図 5.27 に示す。

ある特定の年に出現率が高くなるのは，その年に盛んに取り上げられた話題や政策に関わる語であり，ニュース性が高かったためと見られる。図 5.26 に示す「ライフライン」は阪神・淡路大震災のあった 1995 年の出現率が突出している。

[80] 「アメニティー」の減少については関根[2003]にも言及がある。

ここに分類された語は「ダンピング」のみ頻度が 2,000 を超えているだけで，総じて頻度が小さい。また，理解率を見ても 25～50%台にあるのが「ライフライン」や「ダンピング」など 4 語のみで，あとはいずれも 25%未満の低いものばかりである。減少型以上に，凹凸型に分類された語は定着を示していない。

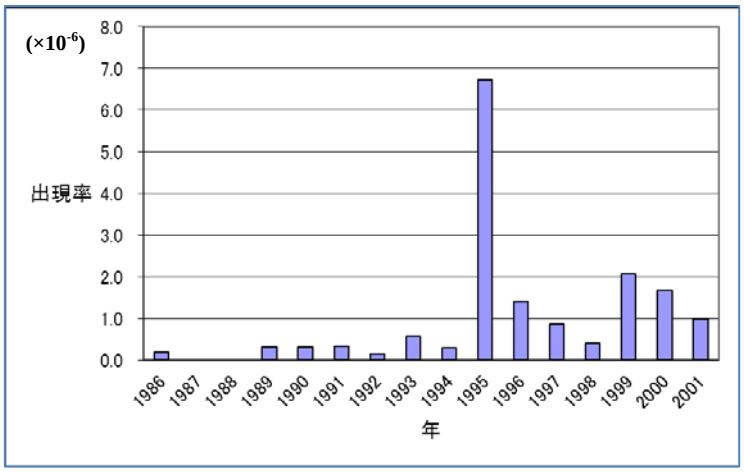


図 5.26「ライフライン」出現率(×10^{-6})

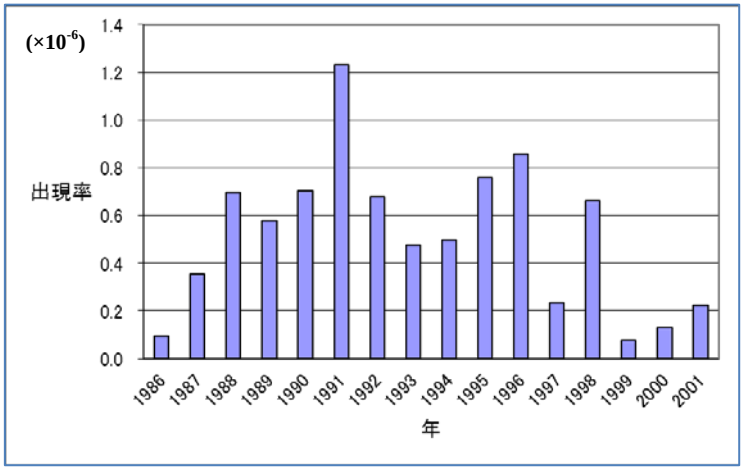


図 5.27「コミットメント」出現率(×10^{-6})

今回の対象語のうち，この凹凸型に分類された語は 15 語である。特定の事柄や事件に関連して偏って使用されるために，わかりにくい語としてあがってくることは想像しやすく，このような語が少なくなかったことも今回の調査対象語の特徴と言える。

5.3.4.5 初出年との関係

最後に，各語の初出年がいつであったかという観点から分析を試みる。ここでは頻度が少なかったため型別の分類外にしていた 6 語も加えて分析する。対象語 109 語のうち，1990 年以降に初出している語は 18 語である。初出時期を 1989 年以前と 1990 年以降の 2 つに分け，その時期別に型が占める比率によって作成したグラフを図 5.28 に示す。

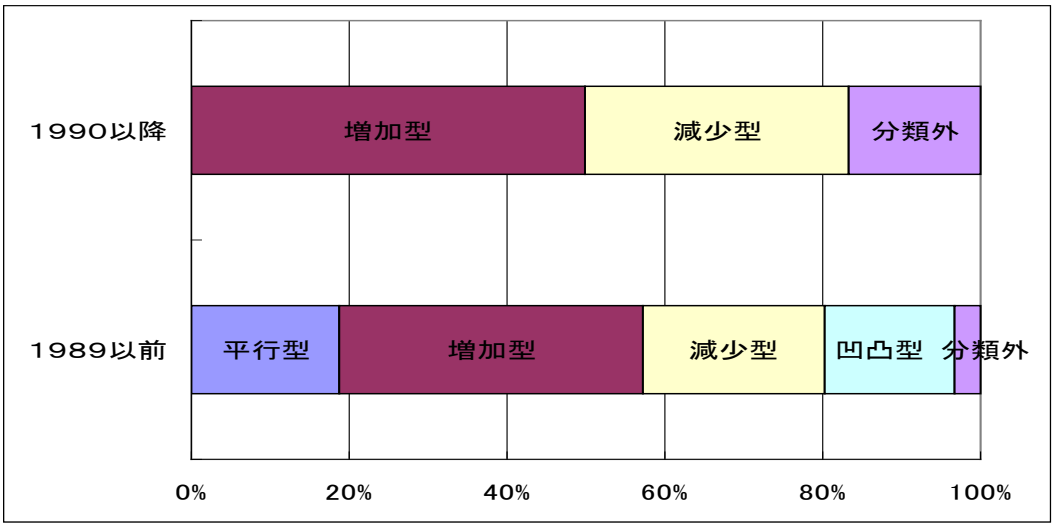


図 5.28 初出年時期別に見る型の比率

図 5.28 を見ると，使用が安定している平行型の語は，初出はすべて 1989 年以前であるという点は予想通りの結果と言えよう。1990 年以降の初出に増加型の語が多いという点もわかりやすい。ここで注目すべきは，1990 年以降の初出に減少型の語も多いという点であろう。これは今回調べた外来語の特徴の 1 つと言えよう。先の減少型の分析で，出現してすぐに定着せずに消えつつある，短命な語があると述べたが，その短命な語の占める割合がかなりあることをこの図は示している。なお，残りの凹凸型と分類外の語は使用頻度に差はあるものの，いずれも連続する推移の型がないものとして 1 つにくくることができるものである。それらの語は，初出年がいつかに関わらず全体的に存在している。この点も今回の対象語の特徴と言える。

5.3.5 外来語の出現率の推移を示す辞書記述の手順

語の出現率の推移の型を，平行型，増加型，減少型，凹凸型の 4 通りに大別した。語の生命サイクルは，一般に，出現，増加，定着，消滅の 5 段階で捉えらえる。古くからあり使用が安定している語は一定の調査期間内においては平行型となり，新語は増加型，廃れつつある語は減少型と

して捉えられよう。今回の調査では，それら 3 つの型を捉えることができた。その他，今回は，使用に偏りがあり，上記 3 つのいずれの型も示さないものがあり，それを凹凸型とした。

そこで，本論文では，辞書に，その語がある期間内において，「平行型，増加型，減少型，凹凸型」のいずれの型であるかと，そう判断した元となる使用推移のグラフを明記するという次の記述方法を提案する。

1. 経年のデータが蓄積している新聞データベース等を時間的使用推移の頻度情報を得て，使用推移のグラフを作成する。
2. 1で作成したグラフを基に，使用推移が次のいずれかの型であるかを判断する。

   平行型　平ら：出現期間内をならすと平ら

   増加型　右上がり：出現期間内をならすと増加

   減少型　右下がり：出現期間内をならすと減少

   山：3 年以上の増加後 3 年以上の減少

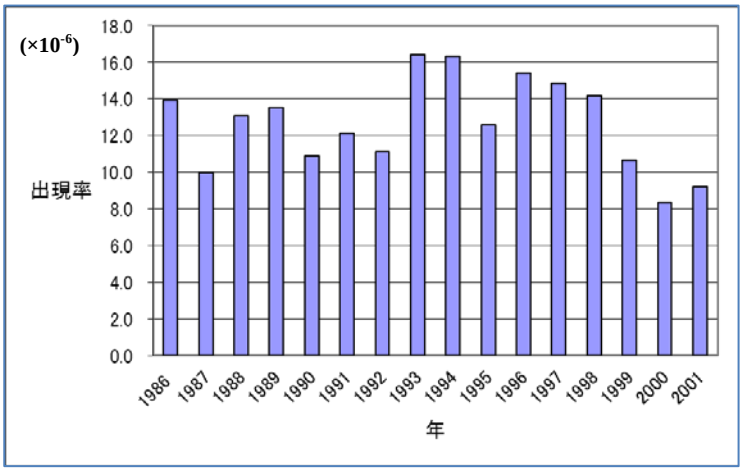
   山の後平ら：増加，減少，平らが各 3 年以上で連続

   凹凸型　突出：特定の年のみ増加

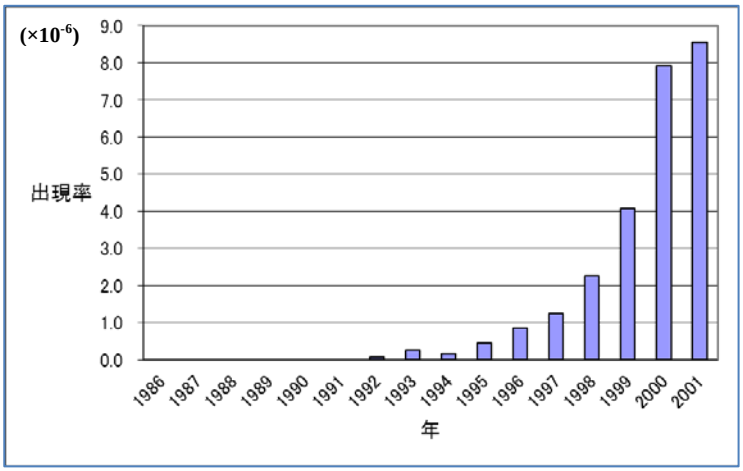
   上記以外
3. 2で判断した推移の型を辞書に明記するとともに，その判断の元となった頻度情報を記載する。頻度情報は，人が閲覧する用途では，閲覧用プログラムが必要に応じてグラフとして表示してもよい。

以下に，推移型に加え，頻度情報をグラフ表示した例を示す。

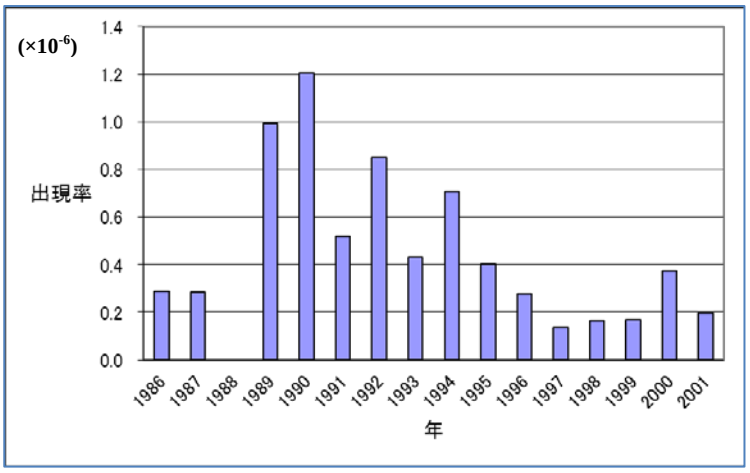
例 1：シェア　平行型



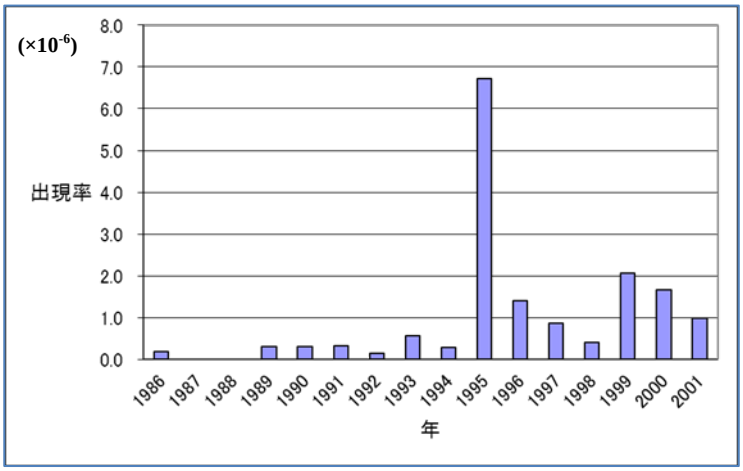
例 2:バリアフリー　増加型



例 3:フレックスタイム　減少型



例 4:ライフライン　凹凸型



なお，全ての語に対して推移型の情報を付与できるわけではない。今回の分析では，16 年間の頻度の総計が 30 未満の語については，明瞭な推移型の判定が困難であった。このような低頻度語については敢えて推移型を付与しないのが妥当であろう。

経年の頻度情報については，全ての語に付与することが可能と考えられる。頻度の計数は誤解析の修正を除けば自動的に行えるため，付与すること自体は現実的な作業で可能である。一方で，経年の頻度情報には次のような効用が期待できる。

・人が閲覧する場合：その語が生命サイクルのどの段階にあるかを理解することで，目的や意図に沿った語を選択したり，語が選択された意図を汲んだりすることができる。斬新な語で人を惹きつけたり，逆に，理解を得やすくするために新しすぎる言葉を避けたりすることができる。廃れつつある言葉であれば，別によりふさわしい言葉がないか考えるきっかけになる。

・言語処理に用いる場合： 上記，人の場合と同様に，精度の高い言語生成や深い言語理解に有用と考えられるほか，新語や死語の発見も可能である。

これらを勘案し，全ての語に頻度情報を付与することを基本方針としたい。

本論文では，一例として外来語のみを対象に，新聞記事コーパスを用いた分析と辞書記述を行った。今後，対象を外来語以外に，そして，調査対象を新聞記事以外に広げていくことが残された課題である。これに関して，大規模な時系列の Web コーパスを構築して新動詞の使用頻度の推移を捉える試みなどが注目される[宇野ほか 2009]。

5.3.6 分類型の安定性の確認

「**要件 2-3** 語の使用頻度の経年変化の情報が得られること」を満たすことは，特に語の流行性や話題性の把握に有用であることが期待されるが，その指標とする分類型の安定性を確認する。

5.3.6.1 確認方法

読売新聞記事と他紙とを比較し[柏野2007]，型分類(表5.16の4分類)による一致度合いと相関とを調べる。調査対象は次の通りとする。

- 読売新聞に，毎日新聞を新たに加える。
- 調査範囲を両方の記事データベースのそろう1991年以降にし，終わりを2004年までとする。
- 調査対象語を，「外来語」言い換え提案全4回分の176語とする。

追加で使用した毎日新聞の 1991 年から 2004 年までの 14 年分の記事データの概要を記す。毎日新聞記事データ[81]は，毎日新聞の東京・大阪本社の朝夕刊最終版を対象とした全文記事データである。今回の調査では，記事データのうち，本文のみを用いた。毎日新聞記事データには本文の他に，見出しやキーワード，掲載面種別コードなどの書誌情報も入っているが，これらは用いていない。調査に用いた本文の文字数のグラフを図 5.29 と図 5.30 に示す。読売新聞に比べると，1991 年から 2004 年までの間でさほどデータ量に大きな変動は見られない。

まず，対象 176 語について読売新聞および，毎日新聞における各年の使用頻度，出現率，および使用頻度の総計を出した。14 年間の使用頻度の総計が 1 万を超えたのは次の語のみである。頻度の総計とともに多い順に示すと，読売新聞「ケア 31,219，シェア 15,981，ベンチャー15,883，アクセス 14,879，セクター13,930，ビジョン 13,725，コミュニティー12,876，ガイドライン 11246」，毎日新聞「ケア 11,415，シェア 10,292」である。読売新聞では 118 語，毎日新聞では 130 語の使用頻度の総計が 1,000 未満であった。

81 毎日新聞社『CD-毎日新聞データ集』（日外アソシエーツ）として 1991 年版より市販されている (http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html)。本調査でもこれを使用した。

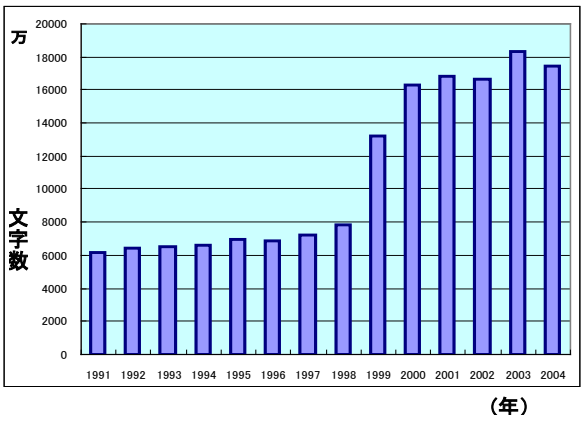


図 5.29 読売新聞記事データ文字数

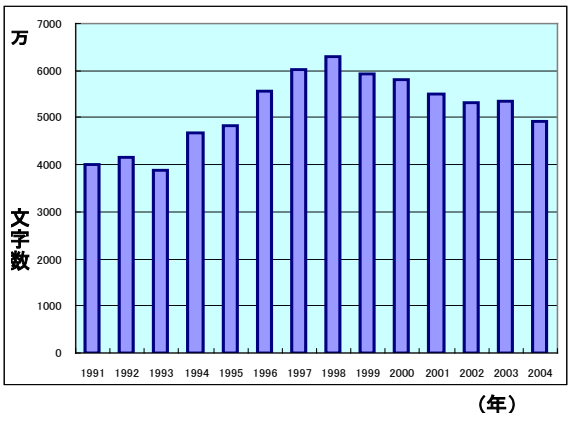


図 5.30 毎日新聞記事データ文字数

5.3.6.2 確認結果

読売新聞，毎日新聞，いずれかが頻度 60 未満であった 20 語は，頻度が少なすぎて推移傾向を捉え難いため分析対象外にした。よって，型分類する対象語は 156 語である。これらについて読売新聞，毎日新聞別に，14 年間の出現率の増減傾向によって 4 分類(平行型，増加型，減少型，凹凸型)を行った結果の語数を表 5.20 に示す。集計は各語の分類が両紙で一致しているか一致していないかを区別して行った。減少型，凹凸型については，ピーク時が大きく違う場合は，その違いを重視し，不一致として集計している。その結果，対象語 156 語のうち，両紙での分類が一致したものが 67.3%の 105 語，不一致であったものが 32.7%の 51 語であるとの結果が得られた。

表 5.20 頻度推移類型の分類結果

分類	両紙一致の語数	両紙不一致の語数	
	読売・毎日	読売	毎日
平行型	9	1	2
増加型	20	9	7
減少型	56	25	16
凹凸型	20	16	26
合計（語）	105	51	51

以下，両紙の一致例を型別に先に示す。その後，不一致例を取り上げる。図のグラフは，増減の推移をわかりやすくするため，出現率の大きさによって Y 軸の目盛り幅を調整している。

平行型

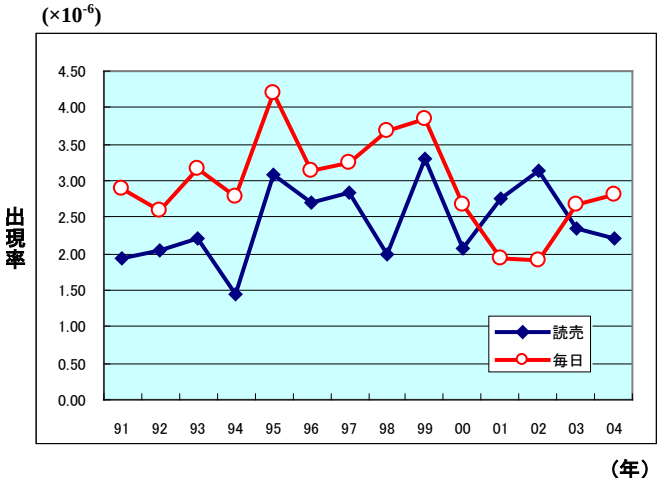


図 5.31 「シミュレーション」出現率(×10^{-6})

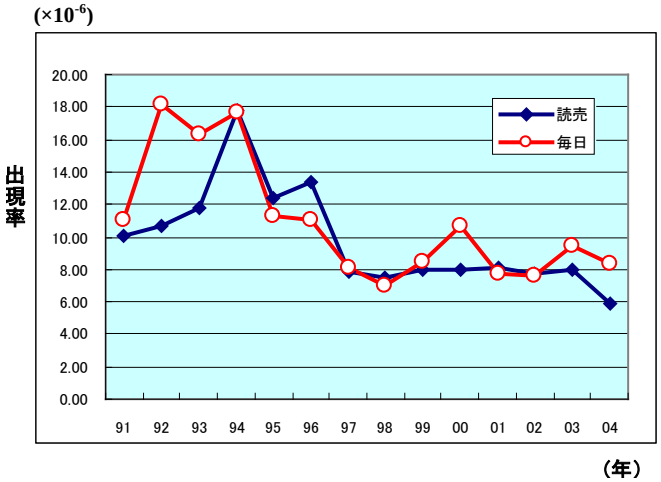


図 5.32 「ビジョン」出現率(×10^{-6})

増加型

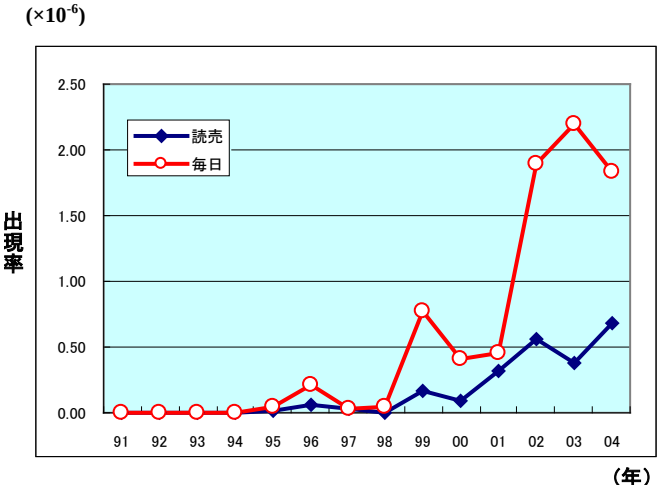


図 5.33 「サプリメント」出現率(×10^{-6})

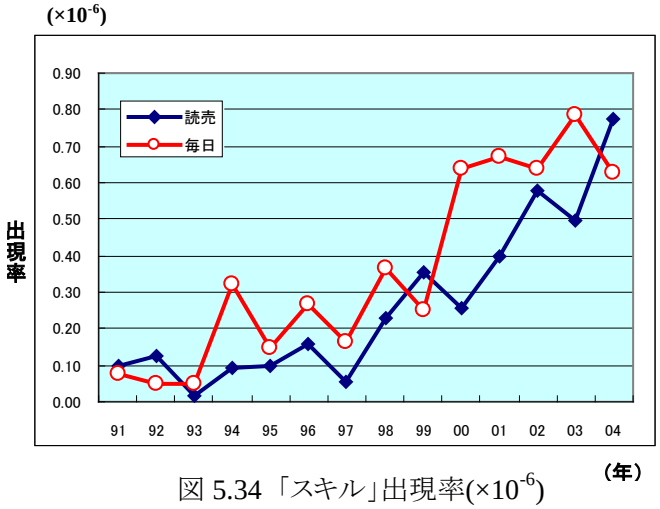


図 5.34「スキル」出現率(×10^{-6})

減少型

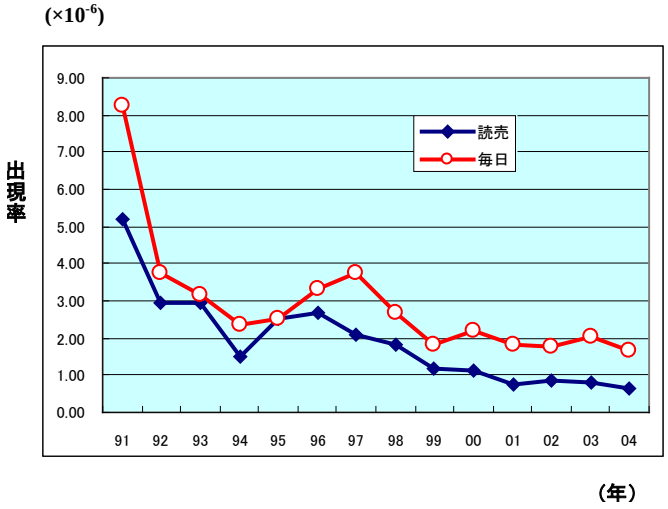


図 5.35「イニシアチブ」出現率(×10^{-6})

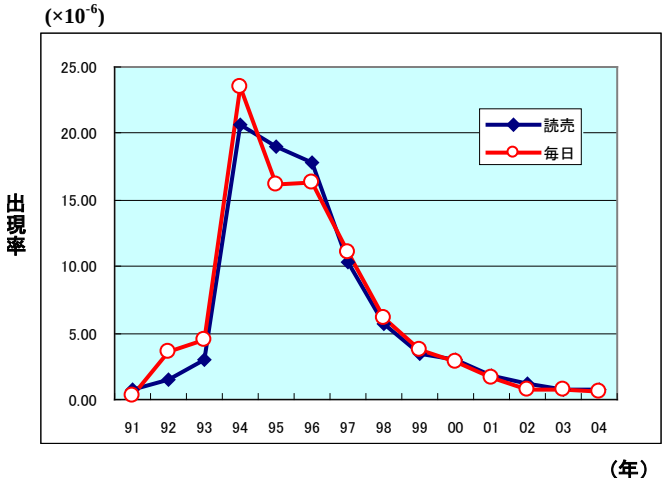


図 5.36「マルチメディア」出現率(×10^{-6})

突出型

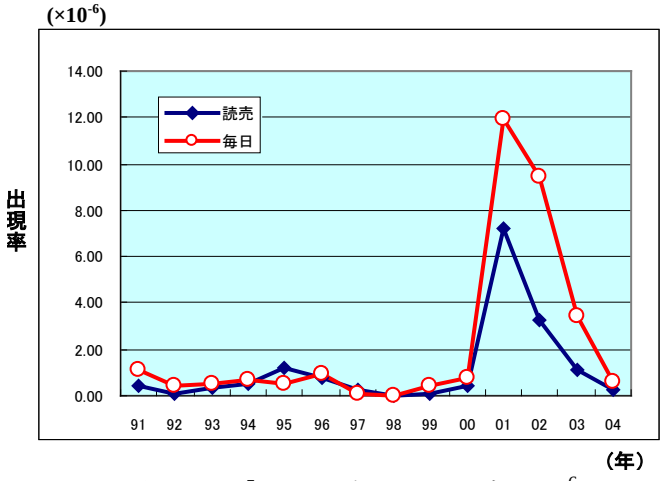


図 5.37 「セーフガード」出現率(×10^{-6})

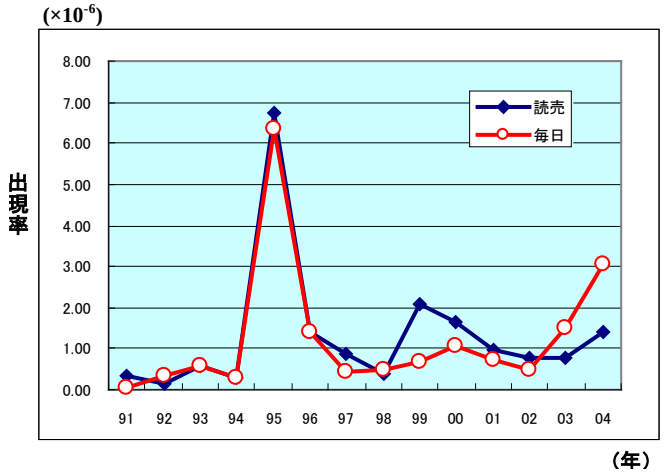


図 5.38 「ライフライン」出現率(×10^{-6})

不一致

読売新聞と毎日新聞とで分類が一致しなかった 50 語について述べる。不一致の原因を調べたところ，新聞個別の事情が原因であることがわかった。その例として，「ポートフォリオ」を図 5.39 に，「インタラクティブ」を図 5.40 に示す。「ポートフォリオ」は，読売新聞にのみ，1996 年を中心とした突出がある。読売新聞には，この年を中心に前後 3 年にわたり，「ニッセイ投資顧問チーフポートフォリオマネージャー」の連載コラムがあった。そのためこのように毎日新聞とは異なる型を示していた。「インタラクティブ」は，読売新聞が低く平らに推移しているのに対し，毎日新聞が急激に右上がりに推移する増加型を示している。これは，1999 年より毎日新聞の新しいホームページが「毎日インタラクティブ」という名前になり，紙面におけるその使用が一気に増加したことによるものであった。

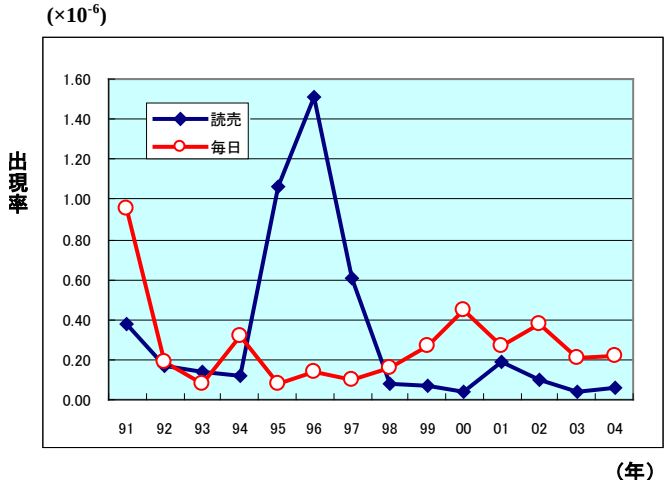


図 5.39「ポートフォリオ」出現率(×10^{-6})

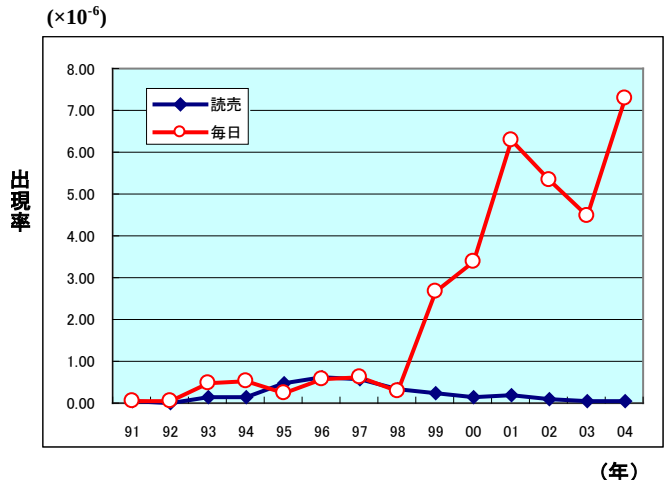


図 5.40「インタラクティブ」出現率(×10^{-6})

このように，一方の新聞に何らかの特別な事情がある場合，2 紙の型分類が不一致となっていた。そのようなものが 50 語，全体の約 3 割あった。逆に，106 語，全体の約 7 割は型分類が一致した。つまり，約 3 割程度は新聞固有の偏りがあり，逆に約 7 割は新聞というメディアにおける使用状況と言えそうである，という結果である。

5.3.6.3 2 紙の出現率間の相関

さらに，型をみた 156 語について，両紙の頻度間の相関を調べた。その結果は，次の通りである。

平均	0.57
0.5 以上	105 語
0.5 以上の割合	67.3%

0.5 以上を相関あり，とすれば，全体の 67.3%に相当する語が相関あり，となった。型分類と同じ値となる結果が得られた。

5.3.6.4 確認結果のまとめ

読売新聞と毎日新聞の二誌について，頻度 60 以上の外来語 156 語について頻度の推移の型を調べ，推移の型が普遍的に，安定的に得られるかを検証した。その結果，型分類で 67.9%の一致，相関関数 0.5 以上が 67.3% という結果が得られ，本論文の提案する型の普遍性，安定性が確認できた。しかしながら，新聞社固有の記事の影響を受ける場合も少なくはないことが確認された。別の新聞を調べれば，不一致の割合がさらに増えることも予想される。よって，複数ソースによる経年調査が推奨される。

ここ最近，ネットにおける検索語の回数の時系列推移を示すサイトが多く出現している。国会会議録を用いた経年調査も可能になっている[茂木 2008]。それらにより，まさに語の流行性や話題性が把握できるようになってきている。本節では新聞データベースにて時系列を分析したが，様々なソース別に時系列の推移が得られることがより充実した辞書情報を得るために望ましいと言える。

附表 1

調査対象語 109 語の一覧

(出現率の型の判断基準別に, 使用頻度の総計順にソート)

附表 5.1 調査対象語 109 語の一覧

語	初出	型の特徴	理解率	総計	異表記	別語
シェア	1986	平ら	50.9	13637	シェアー	シェアソン シェアロン シェア・スタジアム 等
ビジョン	1986	平ら	61.6	11170	ヴィジョン	デビジョン エキシビジョン スーパービジョン等
バックアップ	1986	平ら	70.6	4154	バック・アップ	
シフト	1986	平ら	60.2	3602		
インパクト	1986	平ら	75.1	2868		
パートナーシップ	1986	平ら	55.0	2867	パートナー・シップ	
マーケティング	1986	平ら	58.1	2748	マーケッティング	
アイデンティティー	1986	平ら	20.9	2462	アイデンティティ アイデンティーティー アイデンテティ 誤アイデンティー	
ライブラリー	1986	平ら	50.0	1290	ライブラリ	
コア	1986	平ら	24.7	1034		スコア コアラ ココア コアジ 等
マスタープラン	1986	平ら	32.6	584	マスター・プラン	
ライフサイクル	1986	平ら	55.0	366	ライフ・サイクル	
モータリゼーション	1986	平ら	21.7	227	モータライゼイション	
タイムラグ	1987	平ら	22.0	200	タイム・ラグ	
ポテンシャル	1986	平ら	21.1	173		
デリバリー	1986	平ら	46.7	168		
フェローシップ	1987	平ら	7.1	139		
ケア	1986	右上	75.6	23512	ケアー	ケアンズ マウナケア 等
アクセス	1986	右上	57.7	10673		
ベンチャー	1986	右上	47.5	10558	ヴェンチャー	アドベ(ヴェ)ンチャー ベンチャーズ
セクター	1986	右上	35.0	8696		
コミュニティー	1986	右上	57.7	7791	コミュニティ コミニティー コミニティ コミュニテイ 誤コミニュティ 誤コミニュティー	
グローバル	1986	右上	41.3	7560		
デイサービス	1986	右上	77.2	4585	デイ・サービス デーサービス デー・サービス ディサービス ディ・サービス	サニーデイサービス ホリデーサービス
インフラ	1986	右上	18.9	4406		メインフライト マインフランケン地方 等
バリアフリー	1992	右上	72.7	3656	バリア・フリー バリヤフリー バリアーフリー	
マネージメント	1986	右上	60.0	3307	マネジメント	
シンクタンク	1986	右上	31.8	3157	シンク・タンク	
ワークショップ	1986	右上	48.9	3045	ワーク・ショップ	
シミュレーション	1986	右上	58.3	2685	誤シュミレーション	
アセスメント	1986	右上	17.0	2519		
コンセプト	1986	右上	40.2	1948		
リアルタイム	1986	右上	58.0	1930	リアル・タイム	
リニューアル	1987	右上	62.5	1652		
セキュリティー	1986	右上	65.6	1481	セキュリティ セキュリテイ	
コンテンツ	1987	右上	23.0	1422	コンテント	
モニタリング	1986	右上	28.9	1032		
セーフティーネット	1991	右上	30.9	839	セーフティー・ネット セフティーネット セイフティーネット セーフティネット セーフティ・ネット セイフティ・ネット セイフティネット	
インターンシップ	1987	右上	15.1	804	誤インターシップ	
ミスマッチ	1986	右上	69.1	709	ミス・マッチ	
リーフレット	1986	右上	23.6	695		
ワーキンググループ	1986	右上	34.3	683	ワーキング・グループ	

附表 5.1 調査対象語 109 語の一覧(続き)

語	初出	型の特徴	理解率	総計	異表記	別語
グローバリゼーション	1987	右上	19.3	659	グローバリゼイション　グロバリゼーション	
プレゼンテーション	1986	右上	50.8	584		
コンソーシアム	1986	右上	4.1	499	コンソーシアム　コンソーシャム	
ポジティブ	1986	右上	40.1	419	ポジティヴ　誤ポジティブ	ポジティブオルガン　ポジティブリー
グランドデザイン	1987	右上	14.3	380	グランド・デザイン	
ゼロエミッション	1995	右上	6.7	285	ゼロ・エミッション	
ノーマライゼーション	1987	右上	12.2	277	ノーマライゼイション　ノーマリゼーション	
モチベーション	1989	右上	31.7	245	モティベーション	
コラボレーション	1987	右上	18.0	229	コラボレイション	
アーカイブ	1990	右上	8.0	227	アーカイヴ	
スケールメリット	1986	右上	13.4	223	スケール・メリット	
スクーリング	1986	右上	32.4	220	誤スクリーング	
ノンステップバス	1997	右上	45.4	217	ノン・ステップバス　ノンステップ・バス 誤ノンストップバス	
セカンドオピニオン	1991	右上	15.8	205		
シーズ	1987	右上	9.8	205		シーズン　ユリシーズ　オーバーシーズ メーシーズ　等
エンパワーメント	1993	右上	5.7	124	エンパワメント	
ユニバーサルサービス	1994	右上	21.7	112	ユニバーサル・サービス	
インキュベーション	1990	右上	3.3	111		
ストックヤード	1989	右上	17.4	99		
マクロ	1986	山平	32.4	3544		マクロクラン　マクロス等
アナリスト	1986	山平	18.4	2756		
バ(ヴァ)ーチャル	1990	山平	38.9	1937	バーチュアル　ヴァーチュアル	
トレンド	1986	山平	53.0	1234		血糖測定器アクトレンド
ガイドライン	1986	山	56.0	10251		
マルチメディア	1988	山	59.3	6739	マルティ・メディア　マルティメディア	
インフォームドコンセント	1988	山	23.2	1732	インフォームド・コンセント　誤インフォームドコセント	
スキーム	1989	山	10.4	849		
モラルハザード	1991	山	10.5	321	モラル・ハザード	
アウトソーシング	1992	山	14.2	318	アウト・ソーシング	
インタラクティブ	1987	山	8.5	262	インタラクティヴ	
アイドリングストップ	1996	山	50.5	188	アイドリング・ストップ	
キャッチアップ	1989	山	13.6	179	キャッチ・アップ	
フィルタリング	1996	山	7.3	87		
コンセンサス	1986	右下	24.0	2379		
オブザーバー	1986	右下	46.8	2013	オブザーヴァー	
コミュニケ	1986	右下	36.6	1432		
フォローアップ	1986	右下	38.4	1024	フォロー・アップ	
ボーダーレス	1989	右下	32.5	992	ボーダレス	
キャピタルゲイン	1986	右下	14.2	693	キャピタル・ゲイン	
アメニティー	1986	右下	20.6	576	アメニティ	
フレックスタイム	1986	右下	43.9	492	フレックス・タイム	
サーベイランス	1986	右下	6.4	408		
オンデマンド	1993	右下	7.2	357	オン・デマンド　オンディマンド　オン・ディマンド	

附表 5.1 調査対象語 109 語の一覧(続き)

語	初出	型の特徴	理解率	総計	異表記	別語
コミット	1986	右下	15.5	285		
オピニオンリーダー	1986	右下	16.8	205	オピニオン・リーダー	
プロトタイプ	1986	右下	14.3	169		
コミットメント	1986	ほか	9.2	508		
モラトリアム	1986	ほか	11.4	492		
インセンティブ	1986	ほか	10.4	346	インセンチブ　インセンティヴ	
イノベーション	1986	ほか	13.6	250	イノベイション　イノヴェーション	
ケーススタディー	1986	ほか	18.1	183	ケーススタディ　ケース・スタディ　ケース・スタディー	
ダンピング	1986	突出	48.8	3691		
インサイダー	1986	突出	28.6	1546		
ライフライン	1986	突出	51.8	1485	ライフ・ライン	
アジェンダ	1989	突出	4.7	465		ケアジェンダー
メンタルヘルス	1986	突出	32.0	408	メンタル・ヘルス	
スクリーニング	1986	突出	18.5	311		ハウスクリーニング　ガラスクリーニング　等
アクションプログラム	1986	突出	9.4	308	アクション・プログラム	
タスクフォース	1986	突出	4.9	164	タスク・フォース	
フレームワーク	1986	突出	14.0	103	フレーム・ワーク	
タスク	1989	突出	19.0	67		レタスクラブ　ロータスクーポン　データスクエア
ログイン	1989		23.4	29	ログ・イン	ダイアログ・イン・ザ・ダーク
ハーモナイゼーション	1989		5.8	22	ハーモナイゼイション　ハーモニゼーション	
サマリー	1990		11.6	13		人名等
バックオフィス	1991		7.8	8		
トレーサビリティー	1996		6.1	5	トレーサビリティ	
エンフォースメント	1989		3.4	5		

附表 2

調査対象語 156 語の分類結果一覧(5.3.6 項)

附表 5.1 平行型に分類した語

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

附表 5.2 増加型に分類した語

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

附表 5.3 減少型に分類した語

見出し	両紙一致の分類	両紙不一致の分類		総度数	
	読売・毎日	読売	毎日	読売	毎日
ケア	減少型			31219	11415
ベンチャー	減少型			15883	6599
アクセス	—	減少型	平行型	14879	8278
ビジョン	減少型			13725	7639
グローバル	減少型			9319	6195
マルチメディア	減少型			7181	4782
バリアフリー	減少型			6845	1991
デイサービス	減少型			6478	1204
インフラ	減少型			5843	3456
ドナー	—	減少型	凹凸型	5089	4944
ワークショップ	—	増加型	減少型	5040	1566
バックアップ	減少型			4953	2495
マネージメント	減少型			4892	2265
シフト	減少型			4156	3451
マーケティング	—	平行型	減少型	3806	2349
シンクタンク	—	増加型	減少型	3726	2526
マクロ	減少型			3712	2737
インパクト	減少型			3189	2204
アナリスト	—	減少型	減少型	3156	1863
リニューアル	—	減少型	増加型	2820	1464
アセスメント	減少型			2807	959
リアルタイム	減少型			2578	1369
バーチャル	減少型			2344	1388
イニシアチブ	減少型			2322	2041
オブザーバー	減少型			2249	1562
コンテンツ	減少型			2170	1989
コンセンサス	減少型			2156	1722
ビオトープ	—	増加型	減少型	1958	304
インフォームドコンセント	減少型			1885	1119
バイオテクノロジー	減少型			1826	649
ミッション	減少型			1816	1449
セーフティーネット	減少型			1352	861
ワークシェアリング	減少型			1340	511
モニタリング	減少型			1295	471
ミスマッチ	減少型			1121	508
ナノテクノロジー	減少型			1098	161
ワーキンググループ	—	減少型	凹凸型	1095	528
リーフレット	—	減少型	凹凸型	1027	208
ハザードマップ	—	減少型	増加型	1024	154
コミュニケ	減少型			1003	931
スキーム	減少型			991	789
ボーダーレス	減少型			945	615
アミューズメント	減少型			944	531
プレゼンテーション	減少型			914	473
プレゼンス	減少型			881	564
オペレーション	—	凹凸型	減少型	837	568
グローバリゼーション	減少型			830	902
フォローアップ	減少型			821	560

附表 5.4 減少型に分類した語(続き)

リバウンド	—	減少型	増加型	791	424
ソリューション	—	減少型	減少型	767	349
マスタープラン	減少型			752	144
アメニティー	減少型			661	282
ポジティブ	—	減少型	増加型	535	453
ドクトリン	—	減少型	増加型	532	522
トラウマ	減少型			521	568
モラトリアム	減少型			513	493
グランドデザイン	減少型			513	251
コミットメント	減少型			481	398
オンデマンド	減少型			471	213
デポジット	—	減少型	凹凸型	467	201
アウトソーシング	減少型			451	265
セカンドオピニオン	—	増加型	減少型	439	213
フレックスタイム	減少型			407	191
モラルハザード	減少型			407	597
ゼロエミッション	—	減少型	減少型	405	171
リテラシー	減少型			402	173
ネグレクト	—	増加型	減少型	397	122
ライフサイクル	—	減少型	凹凸型	396	200
トレーサビリティ	—	増加型	減少型	376	114
コンポスト	減少型			372	178
スケールメリット	—	凹凸型	減少型	361	194
パブリックコメント	—	減少型	減少型	339	75
スクーリング	—	凹凸型	減少型	329	149
キャピタルゲイン	減少型			327	200
サーベイランス	—	減少型	突出型	309	331
アイドリングストップ	—	減少型	凹凸型	299	81
インタラクティブ	—	減少型	増加型	294	1726
モータリゼーション	減少型			274	132
ロードプライシング	減少型			270	83
コミット	—	減少型	凹凸型	266	300
レシピエント	—	減少型	凹凸型	263	225
クライアント	減少型			230	232
ステレオタイプ	—	減少型	凹凸型	226	201
アカウンタビリティー	減少型			223	210
ケーススタディー	減少型			208	130
タイムラグ	—	減少型	凹凸型	202	140
キャッチアップ	減少型			187	146
オピニオンリーダー	—	減少型	減少型	177	181
フィルタリング	—	減少型	減少型	126	66
プロトタイプ	減少型			103	105
フリーランス	—	減少型	増加型	97	130

附表 5.4 凹凸型に分類した語

見出し	両紙一致の分類	両紙不一致の分類		総度数	
	読売・毎日	読売	毎日	読売	毎日
ガイドライン	凹凸型			11246	7682
ドナー	—	減少型	凹凸型	5089	4944
ダンピング	凹凸型			3517	2790
セーフガード	凹凸型			2335	1644
ライフライン	凹凸型			1973	894
コア	—	凹凸型	凹凸型	1755	793
インサイダー	凹凸型			1278	1555
ワーキンググループ	—	減少型	凹凸型	1095	528
リーフレット	—	減少型	凹凸型	1027	208
オペレーション	—	凹凸型	減少型	837	568
アジェンダ	凹凸型			620	529
マンパワー	—	凹凸型	凹凸型	609	309
メンタルヘルス	—	凹凸型	凹凸型	566	219
デポジット	—	減少型	凹凸型	467	201
インセンティブ	—	凹凸型	凹凸型	435	301
アーカイブ	—	増加型	凹凸型	408	220
ライフサイクル	—	減少型	凹凸型	396	200
イノベーション	凹凸型			393	210
ノーマライゼーション	—	凹凸型	凹凸型	390	209
コージェネレーション	凹凸型			385	178
アクションプログラム	凹凸型			382	226
スケールメリット	—	凹凸型	減少型	361	194
ポートフォリオ	—	凹凸型	凹凸型	360	191
スクリーニング	—	凹凸型	凹凸型	354	136
シーズ	—	凹凸型	凹凸型	342	80
スクーリング	—	凹凸型	減少型	329	149
サーベイランス	—	減少型	凹凸型	309	331
アイドリングストップ	—	減少型	凹凸型	299	81
デフォルト	凹凸型			288	291
コミット	—	減少型	凹凸型	266	300
レシピエント	—	減少型	凹凸型	263	225
タスクフォース	凹凸型			260	234
サムターン	凹凸型			255	62
ソフトランディング	—	凹凸型	凹凸型	247	267
ポテンシャル	凹凸型			237	181
デリバリー	凹凸型			236	112
ステレオタイプ	—	減少型	凹凸型	226	201
エンパワーメント	—	増加型	凹凸型	209	108
タイムラグ	—	減少型	凹凸型	202	140
コンファレンス	凹凸型			170	82
センサス	凹凸型			163	66
フェローシップ	凹凸型			157	72
リターナブル	—	凹凸型	凹凸型	132	90

附表 5.4 凹凸型に分類した語(続き)

見出し	両紙一致の分類	両紙不一致の分類		総度数	
	読売・毎日	読売	毎日	読売	毎日
メディカルチェック	—	凹凸型	凹凸型	123	73
タスク	—	凹凸型	凹凸型	105	73
フレームワーク	凹凸型			103	108
デジタルデバイド	—	凹凸型	凹凸型	89	201
ブレークスルー	凹凸型			85	63
プライオリティー	凹凸型			79	65

第6章 結論

本論文では，汎用の基本的辞書の構築に焦点を当て，大規模な言語コーパスを活用した新たな辞書構築の方法を論じた。具体的には，(1) 辞書構築の骨格となる見出し語選定，語義区分，多義構造の記述を客観化・明示化して行う方法を提案した。(2) 解析・生成処理の精度向上が特に期待できるコロケーション(共起情報)の記述，使用域の記述を精緻化するための辞書記述方法を提案した。(3) 提案する記述方法に基づく辞書記述の作業手順を具体的に示した。

人間の参照と機械での使用の両方を視野に入れた日本語の汎用辞書を，BCCWJ のような大規模な均衡コーパスを用いて構築する例はこれまでになく，その方法論についての具体的な議論も不十分であった。本論文は，この点に対する貢献を主眼としている。

コーパス整備以前は，辞書編集は編集者自身の不断の努力により用例収集を行わねばならなかった。しかし，人間が目を通せるテキストのジャンルや量には限界がある。本論文では，性質の異なる複数のコーパスを用い，それらの分析に立脚した記述方法を提案して，(1)客観的な記述，(2)網羅的な記述が可能になること，および，(3)使用域の具体的な記述が可能になることを示した。

第 2 章では，まず，従来の(人間用の)国語辞典について検討し，構築作業の主観依存性，非網羅性，暗黙の前提による情報の省略，を問題点として述べた。次に，従来の計算機用日本語辞書について検討し，非構造性，生成処理用の情報不足，などの問題点を指摘した。

第3章では，前章で指摘した問題点のコーパスによる解決可能性を検討した。まず，3.1節において，近年利用可能となった代表的コーパスであるBCCWJを中心に，コーパスとコーパス言語学の概要を述べた。次に，3.2 節において，コーパス分析に基づくことでどのような辞書記述の改善が期待できるかを先行研究，最近の研究の事例を通して検討した。それらを踏まえ，3.3 節において，本論文で提案する辞書記述方法の基本要件を示し，提案法の新規性と期待される効果について述べた。

第4章では，辞書記述の骨格となる辞書の見出し語選定と語義の区分に関わる検討を行った。即ち，まず，4.1 節において**見出し語選定**について論じた。現代語と古語の間に位置する「古風な

語」を例に,「**要件 1-1** 見出し語の選定基準が客観的・明示的であること」を満たす方法として, コーパスの頻度情報を使った見出し語選定方法を提案した。まず, 現行の国語辞典 5 種の調査によって, 古風な語の見出し語としての収録に差異のあることを指摘したが, これは, 見出し語としての採択が主観的・暗黙的に行われた結果であると考えられた。古風な語は現代語の中で用いられることが少なくないが, 文法的・文体的に注意を要す言葉であるため, 辞書に収録する必要性は高い。そこで, コーパスに用例が得られるかどうかを客観的基準として, 積極的に見出し語として収録すべきことを述べた。

次に, 4.2 節と4.3 節において**語義の区分と多義構造の記述**について論じた。はじめに, 従来の国語辞典の多義語の記述の分析を述べ, 多義語の記述に編集者の主観的判断が反映されていることを指摘した。その上で,『IPAL 名詞辞書』の構築の際に筆者らが導入した, コーパス分析に基づく語義の区分方法と, 多義構造(語義間の意味的関係)の明示的記述方法を示し, 記述の客観性が向上し得ることを示した。また, 記述の客観性を増すためにコーパス分析に基づく, 語義の区分方法を 4.2 節で, 語義間の意味的関係である多義構造の明示的記述方法を 4.3 節で提案した。

4.2 節では,「**要件 1-2** 語義の区分が客観的・明示的であること」を満たすために, 用例や用法に着目して区分する 12 通りの「下位区分の判断基準」を設け, 説明が可能な一貫した「下位区分」を行う方法を提案した。基準によって弁別可能な適用例により客観的に判断可能であることを確認した。

4.3 節では,「**要件 1-3** 多義語の意味関係の構造が客観的・明示的であること」を満たすために, 多義構造を比喩による意味の拡張という観点から捉え, 明示的に記述する方法を提案し, 多義語を異なる意味の単なる寄せ集めとして扱うのではなく, 構造をもったものとして辞書に体系的に記述することにより, 語をよりよく理解することが可能になることを示した。多義語 560 語について記述を試行した結果, 意味的関係全 1,364 のうちの 1,174 を比喩的関係として捉えて記述できた。つまり, 全体の約 86% の記述が可能であることを確かめた。研究当時, 辞書に比喩による意味の拡張関係を記述する試みは本研究が初めてであったが, 現在では『英語多義ネットワーク辞典』(2007, 小学館),『ベネッセ全訳古語辞典 改訂版』(ベネッセコーポレーション, 2007 年)等, 市販の辞典にも取り入れられていることは, 提案するアプローチの有用性を示すものと言えよう。

第 5 章では, 構造化された辞書の記述内容に関わる検討を行った。まず 5.1 節では, **コロケーション(共起情報)の記述**を取り上げ,「**要件 2-1** 多義性をもつコロケーションの意味を区別するため

の情報が得られること」を満たすために，コロケーションの多義性解消のための用例記述方法を提案した。従来，多義性解消の試みは，主として語の多義性が着目されてきたが，語と語とが結び付いたコロケーションの多義性という観点からの検討が重要であることを指摘し，その解決案を示した。コロケーションの多義性の要因が述語(動詞／形容詞)多義性による場合には，従来の方法から類推されるように文型情報を利用することが有効であることを示した。一方，コロケーションの多義性の要因が，述語ではなく，名詞やコロケーション自体の多義性によるものが多くあり，それらは文型情報のほかに，名詞句内の先行句に関する辞書情報の記述が多義性解消に有効であることを示した。提案手法により，『IPAL 名詞辞書』に重複するコロケーション(延べ 10,415 例，異なり 4,578 例)の重複対 7,623 例のうち 6,436 例，すなわち全体の 84.4%の多義性の解消ができることを確認した。詳細で豊富なコロケーションの記述は，適切なコロケーションの生成にも役立つことが期待されるものである。

5.2 節では，**使用域の記述**について論じた。まず，文体情報を使用域の辞書記述に取り込むために，文体的特徴を表す分類指標を設計し，コーパスに文体情報を付与するアノテーション方法を提案した。BCCWJ 図書館サブコーパス収録の全 10,551 サンプルを対象に，全サンプルにアノテーションを行った。テキスト構造が単純(例：章節構造)なもの(全体の 84%)については，内容・表現の文体的特徴により，専門度，客観度，硬度，くだけ度，および語りかけ性度，という 5 観点による分類指標を定め，主観的評定によって評価値を付与した。また，テキスト構造・紙面形式などの点で上記分類になじまないもの(全体の 16%)には別途，その特徴を表す別の指標(対話系，引用系，データベースやリスト系等)を設定した。それにより，多種多様な形式をもつサンプルがどの NDC にどの程度収録されているかなど，コーパス収録サンプルの全容を明らかにすることができた。これらアノテーションを手掛かりに，緻密な文体研究が可能になった。また，文体情報を付与したコーパスより，文体的特徴をもつ語の抽出が可能であるかを，定性的，定量的に検討した。その結果，必ずしも定性的な分析と，定量的な分析とは一致しない結果となった。しかし，両者から得られる情報をあわせて吟味し，文体情報を辞書記述に取り込んでいくことは，これまでにない使用域の記述の実現につながると考えられる。

次に，再び「古風な語」を取り上げ，(1)古典の引用での使用，(2)明治期から戦前までの使用，(3)時代・歴史小説での使用，(4)現代文脈での使用，という 4 分類に基づく辞書記述方法が，従来の「古風な語」をひとくくりにする辞書記述よりも，語の文体的特徴をより豊富に記述できることを示した。具体的に，その一例として「そなた【其方】」の記述例を示した。「**要件 2-2** 語の使用域・位

相の情報が得られること」を満たすことは，これまでの辞書構築でずっと理想論として議論されてきたが，コーパスを活用することで具体的に実現可能であることを本節において明示した。

最後に 5.3 節において**頻度情報による時間的使用推移の記述**について論じた。外来語を対象に，1986 年から 2001 年までの 16 年分の読売新聞記事データベースを用いて，語の使用推移，即ち出現率の推移に着目して語の使用状況を捉えた。語の出現率の推移の型を，大きく，平行型，増加型，減少型，凹凸型の4通りに分類し，使用状況を分析した。そして，その時間的使用推移の型を明示的に辞書に記述することを提案した。このような記述から，語の流行り廃りや，時流に合った表現かどうかなどの情報を得ることができ，人にとっても機械にとっても，精度の高い言語生成やニュアンスなども含めた深い言語理解の実現に有用であることを述べた。「**要件 2-3** 語の使用頻度の経年変化の情報が得られること」を満たすことは，特に語の流行性や話題性の把握に有用であることが期待される。そこで，その指標とする分類型の安定性を調べた。読売新聞，毎日新聞の二紙の記事データベースを用いて，外来語 156 語について，頻度推移 4 分類の型の一致度合いと相関とを調べた。その結果，型分類で 67.3%が一致した。また，相関関数 0.5 以上を相関ありとすれば，相関ありも同じく 67.3% という結果が得られ，本論文の提案する型の安定性を確認することができた。

以上述べたように，本論文では，従来の国語辞書および計算機用辞書の記述において，様々な制約によって十分な記述が行われていなかった側面，および編集者の主観的判断に強く依存していた側面のうち，言語学上あるいは応用上の観点から特に重要と考えられる点に順次光を当て，それぞれ，目的に応じた適切なコーパスを使用することによって，それらの辞書記述がどのように精緻化できるかを明らかにし，その手順と効果とを具体的に示した。

コーパスの意義は，単に用例を抽出できるだけではなく，統計的手法により，言語に関する定量的な指標を抽出できることにもある。客観的・網羅的言語分析に基づく見出し語の選定，語義の区分，意味的関係の記述，用例の記述，また，使用域の分類に基づく記述や時間的使用推移の記述により，従来の辞書の課題であった，母語話者利用を前提とする情報の省略や生成に必要な情報の不足などの問題が軽減された。また，辞書記述方法の手続き化により，プロの辞書編集者に頼らず，作業ベースでの辞書構築が可能になることにも実際上の意義がある。

このように，本論文では，掲げた要件を満たす辞書記述方法を提案した。しかし，本論文では，古風な語，外来語などといった限られた例を取り上げて，その範囲で議論したに過ぎない。これを足掛かりとして，より網羅的に検討し，数万語規模の実際の辞書を構築することは今後取り組むべ

き課題である。また，コーパスを用いた辞書記述の精緻化自体に関しても，なお課題が残っている。

その 1 つは，低頻度語の扱いである。BCCWJ の 1 億語の規模では，残念ながら頻度 0 という語も数多く存在する。低頻度語の用例抽出には数十億語以上のコーパスが必要とも言われている。コーパスの規模をさらに大きくすることや，日々更新される Web データからの情報抽出との併用は解決策として考えられる。一つの解決策として，3.1.5 項で述べた通り，国立国語研究所では 200 億語規模の『国語研日本語ウェブコーパス』(NWJC)が構築された[浅原ほか 2014a，浅原ほか 2016]。BCCWJ の規模では収集できない稀言語現象での情報収集のためには 100 倍の規模が必要だと算段され，当初 100 億語を超えることを目標に構築され，結果として 200 億語規模のものが構築されている。4.1.6.5 目において，BCCWJ で頻度 0 となる岩国の見出し語 50 語を例に検索したところ，そのうち 4 語以外はすべて用例が得られることを報告した。また，前川[2016]は，「書ける」の意味で「書けれる」を用いる例（二重可能あるいはレ足す言葉）は BCCWJ では頻度 0 であるが，NWJC では 100 以上ある，と報告し，稀言語現象の情報収集が可能になっていることを示している。しかしながら，前川は「～を泣く」「～を死ぬ」のようなヲ格の用例は，BCCWJ ではそれぞれ 1 例，3 例あったのに対し，NWJC では 0 であったことから，NWJC に均衡性はないため，単純に用例数が増えるわけでもないとも述べている。Web コーパスでは単純に用例数が増えないことは，BCCWJ が搭載されている NLB と，11 億語の筑波 Web コーパス(TWC)が搭載されている NLT を比較してもわかる。例えば，「押し合う」のヲ格の用例は，NLB7 例に対し，NLT は 38 例もあるが，その中で「プレートが押し合う」の用例はどちらも同数の 2 例である。よって，単にコーパスの規模拡大を求めていくだけではなく，これまでの国語辞典の編集者が行ってきたように，用例のとれそうなデータから選択的に用例を採取するなど，人手による洞察や分析も依然欠かせないと考えられる。

2 つめの課題は，用例記述の網羅性を示す記述方法の提案である。例えば，コロケーションの場合，結び付く可能性のある語がどれほどあるかは語によって大きく異なる。慣用的な言いまわしを含め，唯一の結び付きしかないものから，例えば，「私」や「欲しい」のように，かなりの語と結び付きが可能なものまである。これらを類型化し，記述の手続き化を進めたい。

3 つめの課題は，文体・位相情報の体系的な記述方法の提案である。本論文では「古風な語」に着目し，その文体・位相情報を精緻化する記述方法について論じた。他に，例えば，「話し言葉」か「書き言葉」かを区別することも重要な辞書情報である。英語辞書では，話し言葉コーパスも整

備され，それに基づく辞書記述が記載されている。日本語でも，まずは話し言葉コーパスを整備し，それに立脚した辞書記述方法を論じていきたい。

4 つめの課題は，意味・用法の自動分類である。本論文では，語に複数の意味や用法がある場合の分類，構造化は人手分析で行っている。一方で，意味・用法の分類や構造化を客観的に行う技術も盛んに研究されている。例えば，河原，黒橋[2006]は，Web 上の約 16 億文の日本語テキストから自動的に用例の意味分類を行い，格フレーム辞書を構築した。佐々木•新納[2007]は，用例を機械的に分類して語義を記述することを試みた。白井，中西，中村[2011]は，BCCWJ を用い，コーパスから単語の新語義を発見する手法を開発している。これらの統計的分析を取り入れて辞書記述の精緻化に活かすことは重要な課題である。

これらの課題の解決にあたっては，均衡コーパスをはじめどのようなコーパスも万能ではない。それぞれのコーパスで得やすい情報と得にくい情報をふまえて，目的に応じた既存のコーパスの選択，あるいは新規なコーパスの設計・構築，そして十分に活用することが重要である。

謝辞

本論文をまとめることができたのは，多くの皆様にご指導，ご助言を頂いたお陰です。ここに記して心より感謝を申し上げます。

奥村学教授は，博士課程においてご指導下さったばかりでなく，博士課程入学以前も含め大変長きにわたり見守り励まして下さいました。 本論文の審査において主査をお引き受け下さった高村大也准教授には，丁寧なご指導を賜りました。寺野隆雄教授，新田克己教授，三宅美博教授には，本論文の審査において貴重なご助言を賜りました。

古宮嘉那子准教授（現，茨城大学），横野光氏（現，(株)富士通研究所），笹野遼平助教をはじめとする奥村・高村ゼミの皆様は，いつも活発な議論をして下さいました。

辞書研究との出会いは東京女子大学にて水谷静夫教授と丸山直子教授にご指導を賜ったことによります。また学生時代，IBM 東京基礎研究所，田中穂積東京工業大学教授の研究室にて多くのことを学びました。

情報処理振興事業協会(IPA)に在職中，橋本三奈子氏（現，株式会社美味と健康），青山文啓桜美林大学教授，本多啓教授（現，神戸市外国語大学），村田賢一教授（元，帝京平成大学），外池俊幸名古屋大学教授をはじめとするプロジェクトメンバーと共同研究を行ったことが，本研究の発端となりました。棟上昭男理事（当時）をはじめとする IPA の皆様には多くの励ましを頂きました。IPAL 利用者の皆様，富士通株式会社・富士通研究所の皆様からも様々にご助言を頂きました。

国立国語研究所では，中野洋氏，佐渡島紗織教授（現，早稲田大学），中山恵利子阪南大学教授や，相澤正夫教授，加藤安彦氏，桐生りか氏（現，光村教育図書株式会社），田中牧郎教授（現，明治大学），茂木俊伸准教授（現，熊本大学），山口昌也准教授と外来語について様々に議論して頂きました。前川喜久雄教授，秋元祐哉氏（現，株式会社兼販），稲益佐知子氏（現，日本体育大学）， 浅原正幸准教授，小木曽智信准教授， 小椋秀樹准教授（現，立命館大学），小沼悦氏，加藤祥氏，小磯花絵准教授， 佐野大樹氏（現，Google Japan），高田智和准教授，田中牧郎教授（現，明治大学），田中弥生氏，中村壮範氏，間淵洋子氏（現，明治大学大学院），

丸山岳彦准教授（現，専修大学），山崎誠教授，山口昌也准教授，ほか多くの方々とともに，『現代日本語書き言葉均衡コーパス(BCCWJ)』のプロジェクトに参加する機会を得たことで，本研究が可能になりました。

BCCWJ への文体情報付与に関しては，さらに，徳永健伸東京工業大学教授，佐藤理史名古屋大学教授，飯田龍氏（現，情報通信研究機構）にご教示頂きました。

本研究における統計データの作成においては，西部みちる氏，立花幸子氏，山下智弥氏，田嶋明日香氏，小山田由紀氏，ほか多くの皆様にご協力を頂きました。

東京女子大学の星野和子氏， 仁科喜久子氏， 荻野孝野氏， 荻野紫穂氏， 牟田浩子氏，望主雅子氏，永井桃子氏，大塚裕子氏，村山尚子氏，山岡美保氏，大熊智子氏をはじめとする水谷・丸山ゼミ生にはいつも多くの励ましを頂きました。

計量国語学会，情報処理学会（自然言語処理研究会・人文科学とコンピュータ研究会），電子情報通信学会（言語理解とコミュニケーション研究会），言語処理学会，人工知能学会（ことば工学研究会）や，新情報処理開発機構（RWCP）における学会や委員会活動，また，共同研究を通し，多くの皆様からご助言を頂きました。

本研究は JSPS 科研費 11710229, 17720109, 18061007, 20520428, 23520572 の助成を受けました。

最後に，本研究の最初の投稿論文より現在まで議論を重ねてくれた夫の柏野邦夫に感謝します。

参考文献

[Aizawa 2003] Akiko Aizawa (2003) Analysis of Source Identified Text Corpora: Exploring the Statistics of Reused Text and the Authorship. *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (ACL2003), pp.383-390.

[Atkins and Rundell 2008]Sue Atkin, Michael Rundell (2008) *The Oxford Guide to Practical Lexicography.* Oxford: Oxford University Press.

[Atkins et al. 2010] Sue Atkins, Adam Kilgarriff,. Michael Rundell (2010) The Database of Analysed Texts of English (DANTE). Anne Dykstra, Tanneke Schoonheim.(eds) *Proceedings of the 14th European Association for Lexicography (EURALEX) International Congress*. Leeuwarden, The Netherlands, pp. 549-556.

[Baisa et al. 2012] Vít Baisa, Vít Suchomel, Jan Pomikálek (2012) Efficient web crawling for large text corpora. Adam Kilgarriff., Serge Sharoff (eds.) *Proceedings of the 7th Web as Corpus Workshop (WAC7)*, France, pp. 39-43.

[Bejoint 2000] Henri Bejoint (2000) *Modern lexicography: An introduction*. republication of Béjoint (1994), New York: Oxford University Press.

[Benko 2014] Vladimír Benko (2014) Compatible Sketch Grammars for Comparable Corpora. Andrea Abel, Chiara Vettori, Natascia Ralli (Eds.) *Proceedings of the 16th EUROPEAN ASSOCIATION FOR LEXICOGRAPHY (EURALEX) International Congress.* Bolzano, Italy, pp.417-430.

[Biber 1988] Douglas Biber (1988) *Variation across speech and writing.* Cambridge: Cambridge University Press.

[Biber and Conrad 2009] Douglas Biber, Susan Conrad (2009) *Register, Genre, and Style.* Cambridge: Cambridge University Press.

[Caruso 2013] Valeria Caruso (2013) From e-Lexicography to Electronic Lexicography: A Joint Review. *Lexikos*, 23, pp.585-610.

[Cook et al. 2013] Paul Cook, Jey Han Lau, Michael Rundell, Diana McCarthy, Timothy Baldwin (2013) A lexicographic appraisal of an automatic approach for detecting new word-senses. Iztok Kosem, Jelena Kallas, Polona Gantar, Simon Krek, Margit Langemets, Maria Tuulik (eds.) *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex2013 conference, 17-19 October, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, pp. 49-65.

[EAGLES 1996] EAGLES (1996) EAGLES Preliminary recommendation on Text Typology. EAGLES *Document EAG–TCWG–TTYP/P*, Version of Jun 1996. (http://www.ilc.cnr.it/EAGLES/texttyp/texttyp.html)

[Fedorova 2004] Irina V. Fedorova (2004) Style and usage labels in learners' dictionaries: Ways of optimization. *Proceedings of the 11th EUROPEAN ASSOCIATION FOR LEXICOGRAPHY (EURALEX) International Congress*, 6–10 July. pp. 265-272.

[Firth 1951] John Rupert Firth (1951) *Papers in linguistics, 1934-1951*. Oxford: Oxford University Press.

[Fuertes-Olivera and Bergenholtz eds. 2011] Pedro A. Fuertes-Olivera, Henning Bergenholtz (eds.) (2011) e-Lexicography: The Internet, Digital Initiatives and Lexicography. London & New York: Continuum.

[Granger and Paquot eds. 2012] Sylviane Granger, Magali Paquot (eds.) (2012) *Electronic Lexicography*, Oxford: Oxford University Press.

[Halliday et al. 1964] Michael Alexander Kirkwood Halliday, Angus McIntosh, Peter Strevens (1964) *The linguistic sciences and language teaching.* London: Longman.

[Hünig 2003] Wolfgang K. Hünig (2003) Style labels in monolingual English learners' dictionaries. Hubert Cuyckens, Thomas Berg, Rene Dirven, Klaus-Uwe Panther (eds) *Motivation in Language. Studies in Honor of Günter Radden.* Amsterdam:John Benjamins, pp.367–389.

[Joos 1961] Martin Joos (1961) *The five clocks.* New York: Harcourt Brace.

[Kashino and Okumura 2010] Wakako Kashino, Manabu Okumura (2010) An Approach toward Register Classification of Book Samples in the Balanced Corpus of Contemporary Written Japanese. *Proc. of the 24th Pacific Asia Conference on Language, Information and Computation (PACLIC24)*, pp.433-438.

[Kathuria, 白井 2012] Pulkit Kathuria, 白井清昭(2012) パラレルコーパスから自動獲得した用例に基づく語義曖昧性解消.『情報処理学会自然言語処理研究会報告』2012-NL-207(3), pp.1-8.

[Kilgarriff and Tugwell 2002] Adam Kilgarriff, David Tugwell (2002) Sketching words. Marie-Hélène Corréard (ed.) *Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins.* European Association for Lexicography (EURALEX), pp. 125-137.

[Kilgarriff et al. 2004] Adam Kilgarriff, Pavel Rychly, Pavel Smrž, David Tugwell (2004) The Sketch Engine. Geoffrey Williams, Sandra Vessier (eds.), *Proceedings of EUROPEAN ASSOCIATION FOR LEXICOGRAPHY (EURALEX)*, Lorient, France, pp.105–116.

[Kilgarriff et al. 2008] Adam Kilgarriff, Milos Husák, Katy McAdam, Michael Rundell, Pavel Rychly (2008) GDEX: Automatically finding good dictionary examples in a corpus. *Proceedings of the 13th EUROPEAN ASSOCIATION FOR LEXICOGRAPHY (EURALEX) international congress*. Barcelona: Institut Universitari de Linguistica Aplicada, Universitat Pompeu Fabra, pp. 425-432.

[Kilgarriff et al. 2010] Adam Kilgarriff, Vojtěch Kovář, Pavel Rychly. (2010) Tickbox Lexicography. Sylviane Granger, Magali Paquot. *ELexicography in the 21st century: New challenges, new applications.* Brussels: Presses universitaires de Louvain,. s. 411-418

[Kilgarriff and 2012] Adam Kilgarriff, Iztok Kosem (2012) Corpus tools for lexicographers. *Electronic lexicography*, pp.31-56.

[Kosem et al. 2011] Iztok Kosem, Milos Husák, Diana McCarthy (2011) GDEX for Slovene. Iztok Kosem, Karmen Kosem (eds.) *Electronic Lexicography in the 21st Century: New applications for new users, Proceedings of eLex 2011*, Bled, 10-12 November. Ljubljana: Trojina, Institute for Applied Slovene Studies, pp. 151-159.

[Kosem et al. 2013] Iztok Kosem, Polona Gantar, Simon Krek (2013) Automation of lexicographic work: an opportunity for both lexicographers and crowd-sourcing. Iztok Kosem, Jelena Kallas, Polona Gantar, Simon Krek, Margit Langemets, Maria Tuulik (eds.) *Electronic lexicography in the 21st Century: thinking outside the paper. Proceedings of the eLex 2013 conference*,17-19 October, Tallinn, Estonia. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, pp. 32-48.

[Kosem et al. 2014] Iztok Kosem, Polona Gantar, Nataša Logar, Simon Krek (2014) Automation of lexicographic work using general and specialized corpora: two case studies. Andrea Abel, Chiara Vettori, Natascia Ralli (eds.) *The user in focus: proceedings of the XVI EUROPEAN ASSOCIATION FOR LEXICOGRAPHY (EURALEX) International Congress*, 15-19 July, Bolzano/Bozen. Bolzano: Institute for Specialised Communication and Multilingualism, pp.355–364.

[Kuwahata et al. 1995] Wakako Kuwahata, Minako Hasimoto, Kenichi Murata (1995) Construction of IPAL-BN : IPA Lexicon of Basic Japanese Nouns. *Proc. of the Natural Language Processing Pacific Rim Symposium '95(NLPRS'95)*pp.137-142.

[Kuwahata et al. 1996] Wakako Kuwahata, Minako Hasimoto, Fumihiro Aoyama (1996) Senses of Polysemous Nouns : Building a Computational Lexicon of Basic Japanese Nouns. *Proc. of The 16th International Conference on Computational Linguistics (COLING1996)*, pp. 1082-1085.

[Landau 2001] Sidney I. Landau (2001) *Dictionaries: The Art and Craft of Lexicography.* second edition, Cambridge: Cambridge University Press.

[Lee 2001] David Y. W. Lee (2001) Genres, Registers, Text Types, Domains, and Styles. Clarifying the Concepts and Navigating a Path Through the BNC Jungle. *Language Learning & Technology*, 5(3), pp.37-72.

[Maekawa et al. 2014] Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, Yasuharu Den (2014) Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, pp.1-27.

[Nakamura 1995] Junsaku Nakamura (1995) Text Typology and Corpus: A Critical Review of Biber's Methodology. *English Corpus Studies*, 2, pp.75–90.

[Sakwa 2011] Lydia Namatende Sakwa (2011) Problems of Usage Labelling in English Lexicography. *LEXIKOS*, 21(1), pp.305–315.

[NTTコミュニケーション科学基礎研究所 2000] NTTコミュニケーション科学基礎研究所(2000)『日本語の語彙特性 第Ⅱ期』三省堂書店.

[NTTコミュニケーション科学研究所監修 1997] NTTコミュニケーション科学研究所(監修)(1997)『日本語語彙大系』岩波書店.

[Paul Procter ed. 1978] Paul Procter (ed.) (1978) Longman Dictionary of Contemporary English. Longman.

[Ptaszynski 2010] Marcin Overgaard Ptaszynski (2010) Theoretical Considerations for the Improvement of Usage Labelling in Dictionaries: A Combined Formal-Functional Approach. *International Journal of Lexicography*, 23(4), pp.411-442.

[Rundell 2012] Michael Rundell (2012) The road to automated lexicography: an editor's viewpoint. Sylviane Granger, Magali Paquot (eds.) *Electronic Lexicography.* Oxford: Oxford University Press. pp.15-30.

[Rundell and Kilgarriff 2011] Michael Rundell, Adam Kilgarriff (2011) Automating the creation of dictionaries: where will it all end? Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin (eds.) *A Taste for Corpora*. Amsterdam: Benjamins, pp. 257–281.

[Sinclair 1991] John Sinclair (1991) *Corpus Concordance Collocation*. Oxford: Oxford University Press.

[Strafella, 松本 2013] Elga Laura Strafella, 松本裕治(2013) 日伊コロケーション辞書の作成を目指す『現代日本語書き言葉均衡コーパス』からのコロケーションの検出と分析.『第 4 回コーパス日本語学ワークショップ予稿集』pp.325-330.

[Srdanovic et al. 2008] Irena Srdanovic, Tomaz Erjavec, Adam Kilgarriff (2008) A web corpus and word-sketches for Japanese.『自然言語処理』pp.137-159.

[相澤 2006] 相澤正夫(2006)「お茶」から「おビール」まで引ける"辞書"―利用者と提供媒体の見直しから―. 倉島節尚(編)(2006)『日本語辞書学の構築』おうふう.

[青木 1989] 青木和男(1989) 古語辞典のあり方 ―意味記述をめぐって―.『国文学解釈と鑑賞』54(1), 至文堂, pp.100-103.

[青山 1996] 青山文啓(1996) 意味素性.『情報処理振興事業協会計算機用日本語基本名詞辞書 IPAL(Basic Nouns)解説編』pp31-61.

[赤野 2000] 赤野一郎(2000) データ収集をめぐる闘い.『月刊言語』29(5), 大修館書店, pp.50-58.

[浅原ほか 2014a] 浅原正幸, 今田水穂, 保田祥, 小西光, 前川喜久雄(2014a) Web を母集団とした超大規模コーパスの開発:収集と組織化.『国立国語研究所論集』7, pp.1-26.

[浅原ほか 2014b] 浅原正幸, 加藤祥, 立花幸子, 柏野和佳子(2014b) 文体指標と語彙の対応分析.『第 6 回コーパス日本語学ワークショップ予稿集』pp.11-20.

[浅原ほか 2016] 浅原正幸, 河原一哉, 武井裕也, 舛岡英人, 大場寧子, 鳥井雪, 森井亨, 前川喜久雄, 加藤祥, 小西光(2016)『国語研日本語ウェブコーパス』の検索系.『言語処理学会第 22 回年次大会発表論文集』pp.1029-1032.

[天野, 近藤 1999] 天野成昭, 近藤公久(1999)『日本語の語彙特性』第 1 巻単語親密度, 三省堂.

[荒川編 2011] 荒川洋平(編著)(2011)『日本語多義語学習辞典 名詞編』アルク.

[飯間 2009] 飯間浩明(2009) 辞書の記述の舞台裏. 国立国語研究所『新「ことば」シリーズ 22 辞書を知る』pp.36-45.

[飯間 2013a] 飯間浩明(2013a)『辞書を編む』光文社.

[飯間 2013b] 飯間浩明(2013b)『『辞書に載る言葉はどこから探してくるのか?ワードハンティングの現場から』ディスカヴァー・トゥエンティワン.

[井口ほか 1996] 井口厚夫, 猪塚元, 桑畑(柏野)和佳子, 山下智弥(1996) 述語の項としての用法.『情報処理振興事業協会計算機用日本語基本名詞辞書 IPAL(Basic Nouns)解説編』pp.86-103.

[石井 1986] 石井久雄(1986)『古代的言語の享受と創造』(文部省昭和 60 年度科学研究費補助金による一般研究(C)研究報告書).

[石井 1989] 石井正彦(1989) 語構成.『国文学解釈と鑑賞』54(1), 至文堂, pp.37-45.

[石川 2004] 石川慎一郎(2004) Corpus, Dictionary, and Education：近刊 EFL 辞書に見る辞書編集の潮流.『Kobe English Language Teaching』19, pp.61-79.

[石川 2006] 石川慎一郎(2006) 言語コーパスからのコロケーション検出の手法―基礎的統計値について―.『統計数理研究所共同研究レポート』190, pp.1-14.

[石川 2011] 石川慎一郎(2011) 現代日本語書き言葉均衡コーパス(BCCWJ)における複合動詞～出すの量的分析.『統計数理研究所研究レポート』238, pp.15-34.

[石川 2012a] 石川慎一郎(2012a)『ベーシックコーパス言語学』ひつじ書房.

[石川 2012b] 石川慎一郎(2012b) 英語教育における辞書―コーパス準拠と学習者志向―『外国語教育学会紀要』15, pp.87-94.

[石塚 2013] 石塚直子(2013) 品詞認定に対するコーパス利用の貢献.『筑波日本語研究』18, pp.16-27.

[石綿, 荻野 1983] 石綿敏雄, 荻野孝野(1983) 結合価から見た日本文法.『文法と意味 I』朝倉書店, pp.81‐134.

[石綿 1999] 石綿敏雄(1999)『現代言語理論と格』ひつじ書房.

[井上 2001] 井上優(2001)〔問 13〕時々「～を注意する」という言い方を見聞きするのですが,「～に注意する」が正しい言い方ではないでしょうか。.『新「ことば」シリーズ 14 言葉に関する問答集―よくある「ことば」の質問―』, 国立国語研究所, pp.36-37.

[井上 2003] 井上永幸(2003) 辞書と辞書学のいま.『月刊言語』32(5), 大修館書店, pp.24-65.

[井上 2005] 井上永幸(2005) コーパスに基づく辞書編集. 齊藤俊雄, 中村純作, 赤野一郎(編)『英語コーパス言語学―基礎と実践』改定新版, 研究社, pp.207-228.

[井上, 有賀 2006] 井上優, 有賀千佳子(2006) これからの学習者用日本語辞書.『日本語学』25(8), 明治書院, pp.22–29.

[井上 2009a] 井上優(2009a) 辞書は何のためにある. 国立国語研究所『新「ことば」シリーズ 22 辞書を知る』pp.28-35.

[井上 2009b] 井上優(2009b) 日本語を学ぶための辞書. 国立国語研究所『新「ことば」シリーズ 22 辞書を知る』pp.104-107.

[今井編 2011] 今井新悟(編著)(2011)『日本語多義語学習辞典 形容詞・副詞編』アルク.

[今井ほか 2013] 今井新悟, 赤瀬川史朗, プラシャント・パルデシ(2013) 筑波ウェブコーパス検索ツール NLT の開発.『第 3 回コーパス日本語学ワークショップ予稿集』pp.199-206.

[内野ほか 1996] 内野一, 池原悟, 白井諭(1996) 弱抑制による連鎖共起表現の抽出とそれに基づく離散共起表現の抽出.『言語処理学会第 2 回年次大会発表論文集』pp.257-260.

[宇野ほか 1971] 宇野義方, 見坊豪紀, 阪倉篤義, 築島裕, 野元菊雄, 林大,宮地裕, 吉沢典男(1971) 国語辞書の条件 専門家八氏に聞く.『言語生活』235, pp.32-42.

[宇野ほか 2009] 宇野良子, 鍛治伸裕, 喜連川優(2009) 新動詞の認知言語学的分析: 大規模時系列ウェブコーパスと言語処理技術が可能にする言語のダイナミズム研究.『言語処理学会第 15 回年次大会発表論文集』pp.418-421.

[ウルマン 1969] S.ウルマン (池上嘉彦訳) (1969)『言語と意味』大修館書店(Ulman, S.(1962) *Semantics : An Introduction to the Science of Meaning*. Oxford: Basil Blackwell).

[遠藤 1988] 遠藤織枝(1988) 話しことばと書きことば―その使い分けの基準を考える―.『日本語学』3, 明治書院, pp.27-42.

[大島 2010] 大島中正(2010) 定義語からひく現代語辞典を構想する.『同大語彙研究』13, pp.1-8.

[大曽，滝沢 2003] 大曽美恵子，滝沢直宏(2003) コーパスによる日本語教育の研究―コロケーション及びその誤用を中心に.『日本語学』22(5), pp.234-244.

[大塚 1995a] 大塚みさ(1995a) 国語辞典における多義語の量的傾向.『計量国語学』19(8), pp.413-423.

[大塚 1995b] 大塚みさ(1995b)意味分野からみた辞書的な多義語について.『計量国語学』20(3), pp.112-122.

[大塚 2006] 大塚みさ(2006) 辞書の見出し語をどう選ぶか―単位の長さと幅をめぐって.『月刊言語』35(10), 大修館書店, pp.44-51.

[大塚 2009] 大塚みさ(2009) 連語とその周辺―国語辞典における扱いを中心に―.『歌子』17, 実践女子短期大学日本語コミュニケーション学科, pp.1-10.

[大塚 20013a] 大塚みさ(2013a) 多義語の出現環境とその辞書記述.『歌子』21, 実践女子短期大学日本語コミュニケーション学科, pp.155-164.

[大塚 2013b] 大塚みさ(2013b) 最近の国語辞書の新傾向『日本語学』32(2), 明治書院, pp.38-47.

[大塚 2014] 大塚みさ(2014) 国語辞典における言語の運用にかかわる情報－その現状との可能性『実践女子短大紀要』35, pp.85-94.

[小木曽，中村 2011] 小木曽智信，中村壮範(2011)『『現代日本語書き言葉均衡コーパス』形態論情報データベースの設計と実装 改訂版』, 国立国語研究所内部報告書(JC-U-10-01).

[荻野ほか編 2003] 荻野孝野，小林正博，井佐原均編(2003)『日本語動詞の結合価』三省堂.

[荻野ほか 2005] 荻野孝野，植田禎子，小林正博，井佐原均(2005) コーパスデータに基づく格助詞組み合わせレベルにおける結合価の実態と同音異表記判定での利用.『自然言語処理』12(4), pp.21-54.

[荻野 2006] 荻野綱男(2006) 検索エンジン Google の使い方と WWW コーパスによる日本語研究. 城生佰太郎博士還暦記念論文集編集委員会(編)『実験音声学と一般言語学』東京堂出版, pp.528-537.

[荻野編 2007] 荻野綱男(編)(2007)『特定領域研究「日本語コーパス」平成 18 年度研究成果報告書 コーパスを利用した国語辞典編集法の研究』.

[荻野，荻野 2007] 荻野綱男，荻野孝野(2007) 日本語のコロケーション研究の歴史―計量言語学、自然言語処理などを中心に―.『日本語学』26(12), 明治書院, pp.58-70.

[荻野 2009] 荻野綱男(2009) コーパスに含まれる「間違い」をめぐって―BCCWJ と WWW の比較―. 荻野綱男(編)『特定領域研究「日本語コーパス」平成 20 年度研究成果報告書 コーパスを利用した国語辞典編集法の研究』pp 1-13.

[荻野編 2010] 荻野綱男(編)(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 コーパスを利用した国語辞典編集法の研究』.

[荻野ほか 2011] 荻野綱男, 近藤泰弘, 矢澤真人, 丸山直子(2011) 研究活動・成果の総括:辞書編集班 コーパスを利用した国語辞典編集法の研究.『特定領域研究「日本語コーパス」平成 22 年度公開ワークショップ(研究成果報告会)予稿集』pp.191-198.

[沖森 2006] 沖森卓也(2006) 国語辞典の意味記述―語釈の示し方を中心に―. 倉島節尚(編)(2006)『日本語辞書学の構築』おうふう.

[奥田 1967] 奥田靖雄(1967) 語彙的な意味のあり方.『教育国語』, pp.6-16(奥田靖雄(1985)『ことばの研究・序説』pp.3-20).

[小椋ほか 2011] 小椋秀樹, 小磯花絵, 冨士池優美, 宮内佐夜香, 小西光, 原裕(2011)『『現代日本語書き言葉均衡コーパス』形態論情報規程集 第4版 』(上)(下), 国立国語研究所内部報告書 (LR-CCG-10-05-01, 02).

[小椋, 冨士池 2011] 小椋秀樹, 冨士池優美(2011) 第 4 章 形態論情報. 国立国語研究所『『現代日本語書き言葉均衡コーパス』利用の手引 第 1.0 版』, 4-1～4-35.

[小椋 2012] 小椋秀樹(2012) コーパスに基づく現代語表記のゆれの調査― BCCWJ コアデータを資料として―.『第 1 回コーパス日本語学ワークショップ予稿集』p.321-328.

[小椋 2013] 小椋秀樹(2013) 現代日本語における外来語表記のゆれ. 相澤正夫(編)『現代日本語の動態研究』おうふう, pp.151-171.

[小椋 2014a] 小椋秀樹(2014a) BCCWJ における複合動詞後項の表記の実態.『第 6 回コーパス日本語学ワークショップ予稿集』pp.113-120.

[小椋 2014b] 小椋秀樹(2014b) 外来語語末長音の表記のゆれについて.『論究日本文学』100, pp.195-208.

[小椋 2015] 小椋秀樹(2015) 外来語における[ei]の表記のゆれ.『第 8 回コーパス日本語学ワークショップ予稿集』pp.75-82.

[尾本, 北 1996] 尾本貴志, 北研二(1996) 距離反比例型スコアを導入したコロケーションの自動抽出法.『情報処理学会研究報告』NL-112, pp.75-82.

[カウイー2003] A.P. カウイー (赤須薫, 浦田和幸訳)(2003)『学習英英辞書の歴史 : パーマー、ホーンビーからコーパスの時代まで』研究社(Anthony Paul Cowie.(2002) *English Dictionaries for Foreign Learners. A History.* Oxford: Oxford University Press).

[笠原ほか 2004] 笠原要, 佐藤浩史, Francis Bond, 田中貴秋, 藤田早苗, 金杉友子, 天野昭成 (2004) 基本語意味データベース: Lexeed. の構築.『情報処理学会自然言語処理研究会報告』NL-159, pp.75–82.

[柏野, 本多 1998] 柏野和佳子, 本多啓(1998) IPAL 名詞辞書における多義構造の記述.『情報処理学会論文誌』39(9), pp.2603-2612.

[柏野 1998] 柏野和佳子(1998) 曖昧性解消過程解明のための多義語の分析.『国立国語研究所創立 50 周年記念研究発表会資料集』pp.69-72.

[柏野 1999] 柏野和佳子(1999) 解析と生成のための共起情報の記述方法.『言語処理学会第 5 回年次大会発表論文集』pp.205-208.

[柏野ほか 2005] 柏野和佳子, 山口昌也, 桐生りか(2005) 新聞記事データベースを利用した外来語の出現率の推移調査.『自然言語処理』12(4), pp. 97-116.

[柏野 2006a] 柏野和佳子(2006a) 国語辞典の釈義と用例の検討.『言語処理学会第 12 回年次大会発表論文集』 pp.8-11.

[柏野 2006b] 柏野和佳子(2006b) 国語辞典における多義語の意味区分の比較.『ことば工学研究会』23,pp.37-43.

[柏野 2007] 柏野和佳子(2007) 第5章 外来語言い換え提案対象語の頻度推移類型―読売新聞・毎日新聞 14 年分を調査して―.『国立国語研究所報告 126 公共媒体の外来語―「外来語」言い換え提案を支える調査研究―』pp.387-401.

[柏野 2009] 柏野和佳子(2009) 1.3 辞書. 言語処理学会編『言語処理学事典』共立出版, pp.80-85.

[柏野ほか 2009] 柏野和佳子, 稲益佐知子, 田中弥生, 秋元祐哉(2009) 第 4 章 対象外要素の排除指定.『特定領域研究「日本語コーパス」平成 20 年度研究成果報告書『現代日本語書き言葉均衡コーパス』における収録テキストの抽出手順と事例』pp. 66-88.

[柏野 2010] 柏野和佳子(2010)「直接的な語り」という表現スタイルをもつ書籍テキストの人手抽出の試み.『人工知能学会第 2 種研究会ことば工学研究会資料』35, pp.63-72.

[柏野, 奥村 2010] 柏野和佳子, 奥村学(2010) 国語辞典に「古い」と注記される語の現代書き言葉における使用傾向の調査.『情報処理学会 人文科学とコンピュータ研究会報告』88, pp. 59-70.

[柏野 2011] 柏野和佳子(2011) コーパスに基づく辞書づくり —これからの国語辞典はこう変わる—.『日本知能情報ファジイ学会誌』23(5), pp.705-713.

[柏野, 奥村 2011] 柏野和佳子, 奥村学(2011) 国語辞典に「古風」と注記される語の使用実態調査 —『現代日本語書き言葉均衡コーパス』を用いて—.『言語処理学会第 17 回年次大会発表論文集』pp.444-447.

[柏野ほか 2012] 柏野和佳子, 立花幸子, 保田祥, 飯田龍, 丸山岳彦, 奥村学, 佐藤理史, 徳永健伸, 大塚裕子, 佐渡島紗織, 椿本弥生, 沼田寛(2012) 書籍テキストへの文体情報付与の試み.『第 2 回コーパス日本語学ワークショップ予稿集』pp.155-164.

[柏野, 奥村 2012] 柏野和佳子, 奥村学(2012) 和語や漢語のカタカナ表記 —「現代日本語書き言葉均衡コーパス」の書籍における使用実態—.『計量国語学』28(4), pp.153-161.

[柏野, 中村2013] 柏野和佳子, 中村壮範(2013) 現代日本語書き言葉における非外来語のカタカナ表記事情.『第 4 回コーパス日本語学ワークショップ予稿集』pp.285-290.

[柏野, 中村 2014] 柏野和佳子, 中村壮範(2014) BCCWJ 図書館サブコーパスの文体情報検索ツールによるテキスト分析.『第 5 回コーパス日本語学ワークショップ予稿集』pp.171-180.

[柏野, 奥村 2015] 柏野和佳子, 奥村学(2015) 「コーパスベース国語辞典」構築のための「古風な語」の分析と記述.『自然言語処理』21(6), pp.1133-1161.

[鍛治ほか 2003] 鍛治伸裕, 河原大輔, 黒橋禎夫, 佐藤理史(2003) 格フレームの対応付けに基づく用言の言い換え.『自然言語処理』10(4), pp.67-81.

[加藤 1998] 加藤安彦(1998) 辞典とコーパス.『日本語学』17(12), pp.37-44.

[金出地 2004] 金出地真人, 徳久雅人, 村上仁一, 池原悟(2004) 結合価文法による動詞と名詞の訳語選択能力の評価.『自然言語処理』11(3), pp.149-164.

[金武 2000] 金武伸弥(2000)『『広辞苑』は信頼できるか 国語辞典 100 項目チェックランキング』講談社.

[金庭ほか 2005] 金庭久美子, 川村よし子, 前田ジョイス(2005) 日本語学習者のための電子辞書編纂の基礎調査.『第 10 回ヨーロッパ日本語教育シンポジウム』.

[金庭, 川村 2006] 金庭久美子, 川村よし子(2006) 日本語学習者のための電子辞書編纂—語の選定と意味の提示順序—.『日本語教育方法研究会誌』13(1), pp.26-27.

[神永 2013] 神永曉(2013) 辞典を改訂することの意義『日本語学』32(2), 明治書院, pp.26-36.

[カルヴェッティ 2011] パオロ・カルヴェッティ(2011) イタリア人向けの和伊辞典編纂における BCCWJ の貢献.『『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集』pp.217-226.

[河原, 黒橋 2006] 河原大輔, 黒橋禎夫(2006) 高性能計算環境を用いた Web からの大規模格フレーム構築.『情報処理学会自然言語処理研究会』171, pp.67-73.

[川村ほか 2000] 川村よし子, 北村達也, 保原麗(2000) EDR 電子化辞書を活用した日本語教育用辞書ツールの開発.『日本教育工学雑誌』24, pp.7-12.

[川村 2009] 川村よし子(2009)『チュウ太の虎の巻―日本語教育のためのインターネット活用術』くろしお出版.

[金 2006a] 金愛蘭(2006a) 外来語『トラブル』の基本語化―20 世紀後半の新聞記事における―.『日本語の研究』2(2), pp.18-33.

[金 2006b] 金愛蘭(2006b) 新聞の基本語『ケース』の意味・用法―類義語「事例」「例」「場合」との比較―.『計量国語学』25(5), pp.215-236.

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

[木村 2007] 木村睦子(2007) 格文法・結合価文法とコロケーション.『日本語学』26(12), 明治書院, pp. 28‐36.

[金城 2011] 金城克哉(2011) コーパス分析に基づく「～にくい」・「～づらい」表現の研究.『留学生教育:琉球大学留学生センター紀要』8, pp.19-35.

[国広 1982a] 国広哲弥(1982a)『ことばの意味3 辞書に書いてないこと』平凡社.

[国広 1982b] 国広哲弥(1982b)『意味論の方法』大修館書店.

[国広 1983] 国広哲弥(1983) 私の辞書論.『日本語学』2(6), pp.13-18.

[国広 1997] 国広哲弥(1997)『理想の国語辞典』大修館書店.

[国広 1999] 国広哲弥(1998) 現代日本語辞典の将来像.『日本語学』17(14), 明治書院, pp.4-11.

[国広 2000] 国広哲弥(2000) 規範主義と記述主義.『月刊言語』29(5), 大修館書店, pp.24-31.

[国広 2006] 国広哲弥(2006)『日本語の多義動詞―理想の国語辞典Ⅱ』大修館書店.

[久野ほか 1998] 久野雅樹, 野崎浩成, 横山詔一(1998) カタカナ語使用の季節変動に関する分析.『日本行動計量学会第 26 回大会発表論文抄録集』, pp.79-80.

[久野 2000] 久野雅樹(2000) 新聞の用字の面による変動と時系列変動.『自然言語処理』7(2),pp.45-61.

[久野ほか 2002] 久野雅樹, 横山詔一, 野崎浩成(2002) 毎日新聞と朝日新聞における文字使用の差.『計量国語学』23(6), pp.277-295.

[暮らしの手帖社 1971] 暮らしの手帖社(1971) 国語の辞書をテストする.『暮らしの手帖』10, 暮らしの手帖社, pp.104 - 114 .

[倉島 1997a] 倉島節尚(1997a) 国語辞典の編集. 新「ことば」シリーズ 5『辞書』文化庁, pp.49-62.

[倉島 1997b] 倉島長正(1997b)『「国語」と「国語辞典」の時代〈下〉―『日国』物語―』小学館.

[倉島 1998] 倉島節尚(1998) 編集者から見た国語辞典のこれから.『日本語学』17(14), 明治書院, pp.64-72.

[倉島 2000] 倉島節尚(2000) 語数をめぐるせめぎ合い.『月刊言語』29(5), 大修館書店, pp.64-75.

[倉島編 2006] 倉島節尚編(2006)『日本語辞書学の構築』おうふう.

[倉島 2006] 倉島節尚(2006) 日本語辞書学構築の礎を―序に代えて―. 倉島節尚(編)(2006)『日本語辞書学の構築』おうふう.

[倉島 2008] 倉島節尚(2008)『日本語辞書学への序章』大正大学出版会.

[桑畑, 橋本 1994] 桑畑(柏野)和佳子, 橋本三奈子(1994) 計算機用日本語基本名詞辞書における見出し語の下位区分.『情報処理学会研究報告自然言語処理』104, pp.1-8.

[桑畑, 本多 1996] 桑畑(柏野)和佳子,本多啓(1996) 区分間の意味的関係.『情報処理振興事業協会計算機用日本語基本名詞辞書 IPAL(Basic Nouns)解説編』pp.211-227.

[桑畑ほか 1998] 桑畑(柏野)和佳子, 橋本三奈子, 青山文啓(1998) IPAL 名詞辞書による多義性解消のためのコロケーションの分析.『情報処理学会論文誌』39(6), pp.1925-1934.

[見坊 1951] 見坊豪紀(1951) 國語辭書の欠點.『岩手大学学芸学部研究年報』3(1), pp.76-84.

[見坊 1952a] 見坊豪紀(1952a) 辞書における定義づけの批判.『岩手大学学芸学部研究年報』4(1), pp.50-59.

[見坊 1952b] 見坊豪紀(1952b) 国語辞書の盲点.『国語科教育』1, pp.78-97.

[見坊 1953a] 見坊豪紀(1953a) 現代語辞書の批判と編集の実際.『金田一博士古稀記念 言語民族論叢』三省堂, pp.27-50.

[見坊 1953b] 見坊豪紀(1953b) 辞書における定義の問題.『コトバの科学四 コトバと論理』, 中山書店, pp.55-76.

[見坊 1961] 見坊豪紀(1961) 辞書の姿.『言語生活』114, pp.16-25.

[見坊 1976a] 見坊豪紀(1976a)『辞書をつくる 現代の日本語』玉川大学出版部.

[見坊 1976b] 見坊豪紀(1976b)『ことばの海をゆく』朝日新聞社.

[見坊 1977] 見坊豪紀(1977)『辞書と日本語』玉川大学出版部.

[見坊 1982] 見坊豪紀, 金田一京助, 金田一春彦, 柴田武(1982)『三省堂国語辞典』第三版, 三省堂.

[見坊 1990] 見坊豪紀(1990)『日本語の用例採集法』南雲堂.

[小磯ほか 2011] 小磯花絵, 田中弥生, 小木曽智信, 近藤明日子(2011) 評定実験に基づくテキスト分類尺度の体系化の試み.『『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集』pp.47-52.

[国立国語研究所 1962] 国立国語研究所(1962)『現代雑誌九十種の用語用字 第 1 分冊総記語彙表』.

[国立国語研究所 1964] 国立国語研究所(1964)『現代雑誌九十種の用語用字第 3 分冊分析』.

[国立国語研究所 1971] 国立国語研究所(1971) 国立国語研究所資料集 7『動詞, 形容詞問題語用例集』秀英出版.

[国立国語研究所 1997] 国立国語研究所(1997)『国定読本用語総覧 CD-ROM 版』三省堂書店.

[国立国語研究所 1984] 国立国語研究所(1984)『日本語教育のための基本語彙調査』.

[国立国語研究所 1996] 国立国語研究所(1996)『「分類語彙表」形式による語彙分類表(増補版)』.

[国立国語研究所 2009] 国立国語研究所(2009)『新「ことば」シリーズ 22 辞書を知る』.

[小島 1998] 小島俊夫(1998)『日本敬語史研究 後期中世以降』(笠間書院).

[後藤 1993] 後藤斉(1993) 神話の比喩的用法について―コーパス言語学からのアプローチ―.『東北大学言語学論集』2, pp.

[後藤 1995] 後藤斉(1995) 言語研究のデータとしてのコーパスの概念について ―日本語のコーパス言語学のために―.『東北大学言語学論集』4, pp.71-87.

[後藤 2001] 後藤斉(2001) 日本語コーパス言語学と語の文体レベルに関する予備的考察.『東北大学文学研究科研究年報』50, pp.214-201.

[後藤2002] 後藤斉(2002) 慣用句と自由な語結合の間 ―「博する」を例にして―.『東北大学言語学論集』11, pp.1-8.

[後藤 2003] 後藤斉(2003)ウェブ辞書の位置づけ―その長所と短所.『月刊言語』32(5), 大修館書店, pp.61-65.

[後藤 2007] 後藤斉(2007) コーパス言語学と日本語研究.『日本語科学』22, pp. 47-58.

[近藤, 天野 2000] 近藤公久, 天野成昭(2000)「日本語の語彙特性」データベース:有効性と問題点.『電子情報通信学会技術研究報告 思考と言語』100(335), pp.1-8.

[近藤 2009] 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.

[近藤ほか 2011] 近藤泰弘, 坂野収, 多田知子, 岡田純子, 山元啓史(2011) BCCWJ 複合辞辞書の仕様・開発・評価.『特定領域研究「日本語コーパス」平成 22 年度公開ワークショップ(研究成果報告会)予稿集』pp.535-544.

[齊藤ほか編 2005] 齊藤俊雄, 中村純作, 赤野一郎編(2005)『英語コーパス言語学―基礎と実践』改訂新版, 研究社.

[阪本 1984] 阪本一郎(1984)『新教育基本語彙』学芸図書.

[阪本 1955] 阪本一郎(1955)『読みと作文の心理』牧書店.

[佐々木, 新納 2007] 佐々木稔, 新納浩幸(2007) 半教師有りクラスタリングを用いた語義数の推定と語義別用例の収集.『自然言語処理研究会報告』NL-76, pp.73-78.

[笹野, 黒橋 2008] 笹野遼平, 黒橋禎夫(2008) 自動獲得した名詞関係辞書に基づく共参照解析の高度化.『自然言語処理』15(5), pp.99-118.

[笹原 2012] 笹原宏之(2012) 辞書を編む.『エリイカ』3, 青土社, pp.87-98.

[佐竹 2000] 佐竹秀雄(2000) 使いやすさをめぐる闘い ―国語辞典篇―.『月刊言語』29(5), 大修館書店, pp.42-47.

[佐藤 1992] 佐藤信夫(1992)『レトリック感覚』講談社学術文庫版, 講談社.

[佐藤 2011] 佐藤宏(2011)国語辞典編集とコーパス.『『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集』pp.153-160.

[佐藤 2012] 佐藤理史(2012) 辞書の見出し語集合と代表性.『言語処理学会第 18 回年次大会発表論文集』pp.915-918.

[佐藤, 夏目 2014] 佐藤理史, 夏目和子(2014) 新しい日本語辞書定義文型の策定に向けて.『第 5 回コーパス日本語学ワークショップ予稿集』pp.153-160.

[佐藤, 夏目 2015] 佐藤理史, 夏目和子(2015) 新しい日本語辞書定義文型の策定に向けて(第二報).『第 7 回コーパス日本語学ワークショップ予稿集』pp.153-160.

[真田 2012] 真田治子(2012) 助詞の使用度数と結合価に関する計量的分析方法の検討.『経済学季報』62(2), 立正大学経済学会, pp.1-35.

[塩谷, 斎藤 1996] 塩谷英一郎, 斎藤初江(1996) 連体修飾語としての用法 1.『情報処理振興事業協会計算機用日本語基本名詞辞書 IPAL(Basic Nouns)解説編』pp.123-148.

[柴田 1987] 柴田武(1987) 辞書から消えた語、消える語.『言語生活』430, pp.28-35.

[下村, 矢澤 2009] 下村健司, 矢澤真人(2009) 国語辞典のブランチ分けと意味記述. 荻野綱男(編)『特定領域研究「日本語コーパス」平成 20 年度研究成果報告書 コーパスを利用した国語辞典編集法の研究』, pp. 23-33.

[情報処理振興事業協会 1987] 情報処理振興事業協会(1987)『計算機用日本語基本動詞辞書 IPAL(Basic Verbs)』.

[情報処理振興事業協会 1990] 情報処理振興事業協会(1990)『計算機用日本語形容詞辞書 IPAL(Basic Adjectives)』.

[情報処理振興事業協会 1996] 情報処理振興事業協会(1996)『計算機用日本語基本名詞辞書 IPAL(Basic Nouns)解説編』.

[情報処理振興事業協会 1997a] 情報処理振興事業協会(1997a)『計算機用日本語基本名詞辞書 IPAL(Basic Nouns)辞書編－改訂版－』.

[情報処理振興事業協会 1997b] 情報処理振興事業協会(1997b)『CD-ROM 版 計算機用日本語基本辞書 IPAL－動詞, 形容詞, 名詞－』.

[レイコフ, ジョンソン 1986] ジョージ レイコフ, マーク ジョンソン (渡部昇一ほか訳)(1986)『レトリックと人生』大修館書店(George Lakoff, Mark Johnson (1980) *Metaphors We Live By.* University of Chicago Press).

[白井ほか 2001] 白井清昭, 柏野和佳子, 橋本三奈子, 徳永健伸, 有田英一, 井佐原均, 荻野紫穂, 小船隆一, 高橋裕信, 長尾確, 橋田浩一, 村田真樹(2001) 岩波国語辞典を利用した語義タグ付きテキストデータベースの作成.『情報処理学会研究報告』NL-141, pp.117-124.

[白井ほか 2011] 白井清昭, 中西隆一郎, 中村誠(2011) 複数の観点から見た用例クラスタリングに基づく新語義の発見.『特定領域研究「日本語コーパス」平成 22 年度公開ワークショップ予稿集』pp.545-552.

[ジャクソン 2004] ハワード ジャクソン (南出康世, 石川慎一郎訳)(2004)『英語辞書学への招待』大修館書店(Howard Jackson (2002) *Lexicography: An Introduction*. Routledge).

[姜 2012] 姜紅(2012) コーパスに基づく多義語「甘い」の意味再分類及び語義分布調査.『第 1 回コーパス日本語学ワークショップ予稿集』pp.59-68.

[新里ほか 2011] 新里圭司, 鎌田浩司, 黒橋禎夫(2011) 同位語を利用した不在インデックス.『情報処理学会研究報告』NL-201, pp.1-7.

[杉本監修 1982] 杉本つとむ監修(1982)『国語辞書を読む』開拓社言語文化叢書.

[杉本 2012] 杉本武(2012) コーパスに見る類義表現―「気がつく」と「気づく」―.『文藝言語研究』言語篇 62, pp.37-51.

[鈴木 1983] 鈴木喬雄(1983) 国語辞書を採点する―杜会人の立場から―.『日本語学』2(6), pp.86-92.

[鈴木 1985] 鈴木喬雄(1985)『診断・国語辞典』日本評論社.

[砂川 1998] 砂川有里子(1998) 日本語学習者のための表現辞典.『日本語学』17(14), 明治書院, pp.45–53.

[スルダノヴィッチ, 仁科 2008] スルダノヴィッチ・イレーナ, 仁科喜久子(2008) コーパス検索ツール Sketch Engine の日本語版とその利用方法.『日本語科学』23, pp. 59-80.

[スルダノヴィッチ 2013] スルダノヴィッチ・イレーナ(2013) 大規模コーパスを用いた形容詞と名詞のコロケーションの記述的研究.『国立国語研究所論集』6, pp.135-161.

[スルダノヴィッチほか 2013] スルダノヴィッチ・イレーナ, スホメル・ヴィット, 小木曽智信, キルガリフ・アダム(2013) 百億語のコーパスを用いた日本語の語彙・文法情報のプロファイリング.『第 3 回コーパス日本語学ワークショップ予稿集』pp.229-238.

[関口, 山本 2003] 関口洋一, 山本和英(2003) Web コーパスの提案.『情報学基礎研究会報告』FI-98, pp.123-130.

[関根 2003] 関根健一(2003) 新聞記事の中のカタカナ語.『日本語学』22(8), pp.30-39.

[園田 2011] 園田匠(2011) 自然言語における特徴的表現の抽出に関する研究.『法政大学大学院紀要(理工学・工学研究科編)』55(http://hdl.handle.net/10114/9663).

[園田ほか 2012] 園田匠, 島田諭, 三浦孝夫(2012) 統計指標を用いた日本語コーパスからのコロケーション情報抽出と精度評価.『情報科学技術フォーラム講演論文集』11(2), pp.201-206.

[高橋 1989] 高橋太郎(1989) 言語の記述にとって用例とはなにか.『国文学解釈と鑑賞』54(1), 至文堂, pp.10-15.

[高原 2014] 高原真理(2014) 日本語学習辞書開発に伴う表記情報の掲載基準に関する一考察.『筑波大学留学生センター日本語教育論集』29, pp.59-69.

[高山ほか 2014] 高山宏規, 加藤芳秀, 大野誠寛, 松原茂樹(2014) 係り受け構造の共起性に基づく日本語コロケーションの自動抽出.『言語処理学会第 20 回年次大会』pp.856-859.

[滝沢 2003] 滝沢直宏(2003)「コロケーションに課される文法的制約ー「肩に雨が落ちてきた」と「お金は天から降ってこない」を例にしてー」平成 13 年度～平成 15 年度科学研究費補助金基盤研究(B)(2)(研究課題番号 13480069)中間報告論文集「日本語学習辞書編纂に向けた電子化コーパス利用によるコロケーション研究」pp.11-23.

[バイバーほか 2003] ダグラス バイバー, スーザン コンラッド, ランディ レッペン(齊藤俊雄ほか訳)(2003)『コーパス言語学―言語構造と用法の研究』南雲堂(Douglas Biber, Susan Conrad, Randi Reppen. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press).

[竹内 2005] 竹内孔一, 乾健太郎, 藤田篤, 竹内奈央, 阿部修也(2005) 分類の根拠を明示した動詞語彙概念構造の構築.『情報処理学会研究報告』NL-169, pp.123-130.

[立山 2014] 立山智絵(2014) 教科書コーパスにおける教科別特徴語.『第 5 回コーパス日本語学ワークショップ予稿集』pp.317-322.

[田中 2003] 田中ゆかり(2003) ネット検索は言語の研究に有用か.『日本語学臨時増刊号 コーパス言語学』22(5), 明治書院, pp.111-123.

[田中ほか 2004] 田中牧郎, 桐生りか, 茂木俊伸, 山口昌也(2004) 行政外来語の実態と問題点.『第 13 回社会言語科学会大会発表論文集』pp.109-112.

[田中 2006] 田中牧郎(2006) 努力するの定着とつとめるの意味変化―太陽コーパスを用いて―. 倉島節尚(編)(2006)『日本語辞書学の構築』おうふう, pp.223-303.

[田中 2010] 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)『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 コーパスを利用した国語辞典編集法の研究』pp 101-111.

[田中 2012] 田中牧郎(2012) 国語教育における外来語―コーパスによる類型化を通して―. 陣内正敬, 相澤正夫, 田中牧郎(編)『外来語研究の新展開』おうふう, pp.224-242.

[田野村 2000] 田野村忠温(2000) 電子メディアで用例を探す―インターネットの場合―.『日本語学』19(6), 明治書院, pp.25-34.

[田野村 2003] 田野村忠温(2003) コーパスによる文法の研究.『日本語学臨時増刊号 コーパス言語学』22(5), 明治書院, pp.174-186.

[田野村 2009a] 田野村忠温(2009a) コーパスを用いた日本語研究の精密化と新しい研究領域, 手法の開発.『人工知能学会誌』24(5), pp.647-655.

[田野村 2009b] 田野村忠温(2009b) コーパスからのコロケーション情報抽出―分析手法の検討とコロケーション辞典項目の試作.『阪大日本語研究』21, pp.21-41.

[田野村 2010] 田野村忠温, 服部匡, 杉本武, 石井正彦(2010) 特定領域研究「日本語コーパス」研究成果報告書『コーパス日本語学の新展開』.

[田野村 2011] 田野村忠温(2011) 大規模コーパスとしてのインターネット.『『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集』pp.83-88.

[玉村 1987] 玉村文郎(1987) 日本語教育基本 2570 語. 日本語教師養成通信講座『日本語の語彙意味(2)』,アルク, pp.100-114.

[玉村 1995] 玉村文郎(1995) 外国人のための日本語辞書構想.『言語』24(6), 大修館書店, pp.54–61.

[玉村 1998] 玉村文郎(1998) 辞典の将来.『日本語学』17(14), 明治書院, pp.12-21.

[丹保 1990] 丹保健一(1990) 多義語における語義配列について―形容詞語彙 (性状)をめぐって―.『三重大学教育学部研究紀要 人文・社会科学』41, pp.11-25.

[丹保 1991] 丹保健一(1991)多義語における語義の区切り方をめぐって― 位置形容詞「高い」「低い」「遠い」「近い」の場合―.『三重大学教育学部研究紀要 人文・社会科学』42, pp.11-32.

[丹保 1992] 丹保健一(1992) 多義語における語義の区切り方をめぐって―数量、実質、刺激を表す形容詞の場合―.『三重大学教育学部研究紀要 人文・社会科学』43, pp.1-20.

[丹保 1993] 丹保健一(1993) 辞書に見られる多義立項基準をめぐって―状態、時間、類型、程度、価値、性向形容詞の場合―.『三重大学教育学部研究紀要 人文・社会科学』44, pp.13-34.

[丹保 1994] 丹保健一(1994) 現代日本語形容詞の多義構造―形容詞多義語における第一義の条件(その 1)―.『三重大学教育学部研究紀要 人文・社会科学』45, pp.11-33.

[丹保 1995] 丹保健一(1995) 現代日本語形容詞の多義構造―形容詞多義語における第一義の条件(その 2)―.『三重大学教育学部研究紀要 人文・社会科学』46, pp.1-31.

[丹保 1996] 丹保健一(1996) 初級形容詞における第一義と使用頻度をめぐって.『三重大学教育学部研究紀要 人文・社会科学』47, pp.11-32.

[丹保 1997] 丹保健一(1997), 形容詞の連体,連用,終止用法の出現頻度と意味との関連性をめぐって―「高い」「広い」「寂しい」を例として.『三重大学教育学部研究紀要 人文・社会科学』48, pp.9-18.

[築島 1983] 築島裕(1983) 理想の国語辞書.『UP』12(02), 東大出版会.

[千葉 2011] 千葉庄寿(2011) BCCWJを用いた語彙・文法情報のプロファイリングとその応用.『特定領域研究「日本語コーパス」平成 22 年度公開ワークショップ予稿集』pp.439-442.

[塚本 2004] 塚本倫久(2004) 初級・中級者のための発信型学習英和辞典とコーパス.『文明21』12, 愛知大学国際コミュニケーション学会, pp.9-28.

[土屋ほか 2002] 土屋雅稔, 黒橋禎夫, 佐藤理史(2002) 辞書定義文の圧縮による定義表現パターンの発見.『人工知能学会誌』17(4),pp.420-430.

[曺 2006] 曺喜澈(2006) 二言語辞書の構築のための「日日辞書」のあり方―「新明解国語辞典」を中心として―. 倉島節尚(編)(2006)『日本語辞書学の構築』おうふう, pp.304-311.

[趙 2014] 趙恩英(2014) 現代日本語の類義関係にある副詞に関する計量的研究. 首都大学東京博士学位論文(http://hdl.handle.net/10748/6683).

[鷲岡 2006] 鷲岡昭夫(2006) 同形, 同音異義語の扱いについて. 倉島節尚編『日本語辞書学の構築』おうふう, pp.63-72.

[鶴丸ほか 1991] 鶴丸弘昭, 竹下克典, 伊丹克企, 柳川俊英, 吉田将(1991) 国語辞典情報を用いたシソーラスの作成について『情報処理学会研究報告』NL-83, pp.121-128.

[伝 2009] 伝康晴(2009) 多様な目的に適した形態素解析システム用電子化辞書.『人工知能学会誌』24(5), pp.640-646.

[投野 2000] 投野由紀夫(2000) 使いやすさをめぐる闘い ―英和辞典篇―『月刊言語』29(5), 大修館書店, pp.34-40.

[投野 2006] 投野由紀夫(2006)『投野由紀夫のコーパス超入門―コーパスでわかる英語学習のコツ』小学館.

[投野 2014] 投野由紀夫(2014) 第 3 章 言語教育とコーパス. 前川喜久雄(編)(2013)『講座日本語コーパス 1 コーパス入門』朝倉書店, pp.62-78.

[土居 1943] 土居光知(1943)『日本語の姿』改造社.

[永田 2008] 永田昌明, 藤田早苗, 平博順(2008) 汎用的な意味解析技術への挑戦.『NTT 技術ジャーナル』pp.24-27.

[中野 1985] 中野洋(1985) 3 語義記述法の問題点. 水谷静夫(編)『朝倉日本語新講座 文法と意味』4, 朝倉書店, pp.75-127.

[中道 1986] 中道真木男(1986) 国語辞典に見る類義語の記述.『日本語学』5(9), pp.34-46.

[新納, 井佐原 1995] 新納浩幸, 井佐原均(1995) 片方向の共起性による述語定型表現の自動抽出.『自然言語処理』2(3), pp.73-86.

[仁科 2000] 仁科喜久子(2000) オンライン教材『あすなろ』プロジェクト.『東工大留学生センター年報』5, pp.43-45.

[仁科ほか 2002] 仁科喜久子, 奥村学, 八木豊, 戸次徳久, 澤谷孝志, 傅亮, 杉本茂樹, 阿辺川武 (2002) 構文表示と多言語インターフェースを備えた日本語読解支援システムの開発.『言語処理学会第 8 回年次大会発表論文集』pp.228-231.

[仁科監 2012] 仁科喜久子(監), 鎌田美千子, 曹紅荃, 歌代崇史, 村岡貴子(編) (2012)『日本語学習支援の構築―言語教育・コーパス・システム開発』東京:凡人社.

[日本電子化辞書研究所 1993] 日本電子化辞書研究所(1993)『EDR電子化辞書仕様説明書』.

[野崎 ほか 1998] 野崎浩成, 久野雅樹, 横山詔一(1998) 新聞電子メディアを用いたカタカナ語の使用実態調査―朝日新聞における月次変化の分析―.『日本行動計量学会第 26 回大会発表論文抄録集』pp.77-78.

[野田 2007] 野田尚史(2007) 文法的なコロケーションと意味的なコロケーション.『日本語学』26(12), 明治書院, pp.18-27.

[野田 2011] 野田尚史(2011) コミュニケーションのための日本語学習用辞書の構想.『日本語／日本語教育研究』2, pp.5-32.

[野村 2006] 野村雅昭(2006) 基礎語で辞書の意味記述はできるか. 倉島節尚編『日本語辞書学の構築』おうふう, pp.17-36.

[延澤ほか 1996] 延澤志保, 堤純也, 中西正和(1996) 自然言語における有繋文字列の抽出.『言語処理学会第 2 回年次大会発表論文集』pp.181-184.

[ハートマン編 1984] R.R.K.ハートマン(編)(翻訳監修 木原研三, 加藤知己) (1984)『辞書学 : その原理と実際』三省堂(Hartmann, R.R.K.(ed.)(1983) *Lexicography: Principles and practice*. London: Academic Press).

[橋本,青山 1992] 橋本三奈子,青山文啓(1992) 形容詞の三つの用法:終止,連体,連用.『計量国語学』18(5), pp.201-214.

[橋本 1995] 橋本三奈子(1995) IPA コーパスにおけるテキストコーパスの作成 『情報処理学会全国大会講演論文集』51, pp.35-36.

[橋本ほか 1997] 橋本三奈子, 山下智弥, 桑畑(柏野)和佳子(1997) 文型テーブルを用いた統合辞書の試作.『ソフトウェア文書のための日本語処理の研究-13』情報処理振興事業協会.

[橋本 2010] 橋本和佳(2010)『名詞とそれを修飾する形容詞の関係』ひつじ書房.

[服部 2015] 服部匡(2015) 漢語動詞における格表示変化傾向の探索 ―ヲ格とニ格―.『第 8 回コーパス日本語学ワークショップ予稿集』pp.189-198.

[林 1983] 林大(1983) 辞書・用例・索引―つれづれのふみのくら(波) . 飛田良文(編)『言語生活』379, pp.82-87.

[パルデシ, 赤瀬川 2011] パルデシ・プラシャント, 赤瀬川史朗(2011) BCCWJ を活用した基本動詞ハンドブック作成 ―コーパスブラウジングシステム NINJAL-LWP の特長と機能―.『『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集』pp.205-216.

[パルデシ, 赤瀬川 2012] パルデシ・プラシャント, 赤瀬川史朗 (2012) レキシカルプロファイリング手法を用いた BCCWJ 検索ツール NINJAL-LWP とその研究事例.『日本言語学会第 144 回予稿集』pp.364-369.

[潘 2006] 潘鈞(2006) 中国人日本語学習者が求めている日本語辞書― 日本語教育の視点から.倉島節尚(編)(2006)『日本語辞書学の構築』おうふう, pp.239-254.

[飛田 2012] 飛田良文(2012) 辞書の正確さとは何か―「全然」の語義について―.『エリイカ』3, 青土社, pp.70-75.

[飛田編 2012] 飛田良文編(2012)『国立国語研究所「日本大語誌」構想の記録』港の人.

[姫野 2004] 姫野昌子(2004)『日本語表現活用辞典』研究社.

[姫野 2006] 姫野昌子(2006) 学習者のためのコロケーション辞典―『日本語表現活用辞典』の作成に際して『日本語学』25(8), 明治書院, pp.40-50.

[姫野監 2012] 姫野昌子(監), 柏崎雅世, 藤村知子, 鈴木智美(編) (2012)『研究社日本語コロケーション辞典』研究社.

[フィルモア 1975] チャールズ J. フィルモア(田中春美, 船城道雄訳) (1975)『格文法の原理―言語の意味と構造―』三省堂(Fillmore, C. J. (1968) *The Case for Case*. Emmon Bach,

Robert T. Harmes (eds.) *Universals in Linguistics Theory.* New York: Holt, Rinehart & Winston).

[深田 2007] 深田淳(2007) 日本語用例・コロケーション抽出システム『茶漉』.『日本語科学』22, pp.161-172.

[福島 2007] 福島一人(2007) 定義されたコロケーションとその有用性 : コロケーション辞典執筆の見地から.『情報研究』36, 文教大学, pp.253-282.

[文化庁 1995] 文化庁(1995)『言葉に関する問答集 総集編』国立印刷局.

[文化庁 1997] 文化庁(1997)『新「ことば」シリーズ 5 辞書』.

[星野 2009] 星野和子(2009) オノマトペの意味と用法―語義記述に必要な情報は何か―. 荻野綱男(編)『特定領域研究「日本語コーパス」平成 20 年度研究成果報告書 コーパスを利用した国語辞典編集法の研究』pp 44-73.

[星野 2010] 星野和子(2010) コーパスに見るオノマトペの用例と辞書の意味記述―ハ行畳語の表記・用法・意味の関係―. 荻野綱男(編)『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 コーパスを利用した国語辞典編集法の研究』pp 54-78.

[堀編 2012] 堀正広(編)(2012)『これからのコロケーション研究』ひつじ書房.

[本多, 桑畑 1997] 本多啓, 桑畑(柏野)和佳子(1997) 名詞の多義について－IPAL 名詞辞書のための研究から.『言語処理学会第 3 回年次大会発表論文集』pp.83-86.

[前川 2008] 前川喜久雄(2008) KONONOHA『現代日本語書き言葉均衡コーパス』の開発.『日本語の研究』4 (1), pp.82-95.

[前川 2011] 前川喜久雄(2011) 特定領域研究「日本語コーパス」と『現代日本語書き言葉均衡コーパス』.『『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集』pp.1-10.

[前川 2013] 前川喜久雄編(2013)『講座日本語コーパス 1 コーパス入門』朝倉書店.

[前川 2016] 前川喜久雄(2016) 言語研究のインフラ整備―日本語コーパスから見えてきたもの―.『国立国語研究所 第 9 回 NINJAL フォーラム』配布資料, pp1-2.

[前田 2002] 前田昭彦(2002) 類義表現の意味論的分析―「さっそく」と「すぐ(に)」―.『長崎大学留学生センター紀要』10, pp.24-40.

[前坊 2009] 前坊香菜子(2009) 語の文体的特徴に関する情報についての一考察―国語辞典と類語辞典の調査から―.『一橋日本語教育研究報告』3, pp. 50-60.

[前坊 2012] 前坊香菜子(2012) コーパスにおける「たぶん」「おそらく」の使用傾向の分析.『一橋日本語教育研究』1, pp. 49-60.

[前坊 2014] 前坊香菜子(2014)「必ず」「絶対」「きっと」の文体的特徴『現代日本語書き言葉均衡コーパス』の調査から.『一橋大学国際教育センター紀要』5, pp. 93-104.

[マケナリーほか 2014] トニー マケナリー, アンドリュー ハーディー (石川慎一郎訳) (2014)『概説コーパス言語学—手法・理論・実践』ひつじ書房(Tony McEnery, Andrew Hardie (2011) *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press).

[松井 1983] 松井栄一(1983)『国語辞典にない言葉言葉探しの旅の途上で』南雲堂.

[松井 1985] 松井栄一(1985)『続・国語辞典にない言葉 ことばの姿さまざま』南雲堂.

[松井 2002] 松井栄一(2002)『出逢った日本語・50 万語—辞書作り三代の軌跡』小学館.

[松井 2005] 松井栄一(2005)『国語辞典はこうして作る 理想の辞書をめざして』港の人, 新宿書房.

[松井 2012] 松井栄一(2012) 国語辞典編集の悩みと工夫.『エリイカ』3, 青土社, pp.58-63.

[松田編 2008] 松田謙次郎編(2008)『国会会議録を使った日本語研究』ひつじ書房.

[松本ほか 1996] 松本裕治, 須藤茂, 中山拓也, 平尾努(1996) 複数の言語資源からのシソーラスの構築.『情報処理学会研究報告』FI-42,pp.23-28.

[間淵ほか 2010] 間淵洋子, 柏野和佳子, 山口昌也, 高田智和(2010) コーパスを用いたテキスト分類指標の検討—BCCWJ の文書構造情報分析を中心に—.『言語処理学会第 16 回年次大会予稿集』pp.314-317.

[丸山, 田野村 2007] 丸山岳彦, 田野村忠温(2007) コーパス日本語学の射程.『日本語科学』22, pp.5-11.

[丸山 2009a] 丸山直子(2009a) 8 品詞別の辞書. 国立国語研究所『新「ことば」シリーズ 22 辞書を知る』pp.74-77.

[丸山 2009b] 丸山直子(2009b) 動詞の格情報とその記述法—国語辞書の例文とコーパス—. 荻野綱男(編)『特定領域研究「日本語コーパス」平成 20 年度研究成果報告書 コーパスを利用した国語辞典編集法の研究』pp 34-43.

[丸山 2010a] 丸山直子(2010a) 動詞の格情報とその記述法—国語辞書の例文とコーパス—. 荻野綱男(編)『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 コーパスを利用した国語辞典編集法の研究』pp 79-100.

[丸山 2010b] 丸山直子(2010b) 助詞「に」を伴う<役割>成分—コーパスに基づく分析—.『日本語文法』10(1), pp.71-87.

[丸山 2011] 丸山直子(2011) 動詞の格情報—国語辞書の記述とコーパス—.『東京女子大学日本文学』107, pp. 227-245.

[丸山ほか 2011] 丸山岳彦, 山崎誠, 柏野和佳子, 佐野大樹, 秋元祐哉,稲益佐知子,田中弥生, 大矢内夢子(2011)『『現代日本語書き言葉均衡コーパス』に含まれるサンプルおよび書誌情報の設計と実装』, 国立国語研究所内部報告書(JC-D-10-02).

[丸山 2013] 丸山岳彦(2013) 第 5 章 日本語コーパスの発展. 前川喜久雄編 『講座日本語コーパス 1 コーパス入門』朝倉書店, pp.105-133.

[水谷 1961] 水谷静夫(1961) 語釈—本格的辞書論の前座.『国語学』47, pp.11-23.

[水谷, 星野 1994] 水谷静夫, 星野和子(1994) 名詞から副詞まで—語類の新しい枠づけ—.『計量国語学』19(7), pp.331-340.

[水谷 2001] 水谷静夫(2001) 束論から見た語類立て.『計量国語学』23(3), pp.135-156.

[水谷 2007] 水谷静夫(2007) 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」辞書編集班, 丸山直子『特定領域研究「日本語コーパス」 平成 18 年度研究成果報告書 辞書編纂談義 —水谷静夫先生が語る—』.

[水谷 2012] 水谷静夫(2012) 理想の辞書を目指して.『エリイカ』3, 青土社, pp.64-69.

[宮崎 1995] 宮崎正弘(1995) 辞書の記述と利用－機械辞書の観点から－.『日本語学』14(4),明治書院, pp.52-61.

[宮島 1972] 宮島達夫(1972)『動詞の意味・用法の記述的研究』国立国語研究所報告 43, 秀英出版.

[宮島 1977] 宮島達夫(1977) 単語の文体的特徴.『村松明教授還暦記念国語学と国語史』, 明治書院, pp.871-903.

[宮島 2007] 宮島達夫(2007) 語彙調査からコーパスへ.『日本語科学』22, pp.29-46.

[宮畑 2009a] 宮畑一範(2009a) 16 表現のための辞書. 国立国語研究所『新「ことば」シリーズ 22 辞書を知る』pp.96-97.

[宮畑 2009b] 宮畑一範(2009b) 18 外国語を学ぶための辞書. 国立国語研究所『新「ことば」シリーズ 22 辞書を知る』pp.100-103.

[村木 1989] 村木新次郎(1989) 現代語辞典の輪郭.『国文学解釈と鑑賞』54(1), 至文堂, pp.66-72.

[村木 2007] 村木新次郎(2007) コロケーションとは何か.『日本語学』26(12), 明治書院, pp.4-17.

[村田ほか 2004] 村田真樹, 金丸敏幸, 井佐原均(2004) 複数の辞書の定義文の照合に基づく同義表現の自動獲得.『自然言語処理』11(5), pp.135-149.

[茂木 2008] 茂木俊伸(2008) 国会会議録における行政分野の外来語. 松田謙次郎(編)『国会会議録を使った日本語研究』ひつじ書房, pp.85-110.

[望月 1961] 望月久貴(1961) どんな辞書を選ぶか.『言語生活』114, pp.28-35.

[森田 1983] 森田良行(1983) 辞書の意味記述―その研究面と実用面―.『日本語学』2(6), pp.26-32.

[森田 1989] 森田良行(1989) 語彙論.『国文学解釈と鑑賞』54(1), pp.46-51.

[森田 2000] 森田良行(2000) 理想との戦い.『月刊言語』29(5), 大修館書店, pp.86-93.

[森山編 2012] 森山新(編著)(2012)『日本語多義語学習辞典 動詞編』アルク.

[矢澤 2007] 矢澤真人(2007) ユビキタス辞書の時代.『日本語学』26(8),明治書院, pp.58-66.

[矢澤 2009] 矢澤真人(2009) バリアフリーの国語辞典. 国立国語研究所『新「ことば」シリーズ 22 辞書を知る』pp.46-55.

[山口ほか 2004] 山口昌也, 茂木俊伸 ,桐生りか ,田中牧郎(2004) 語種との関係に基づいた新聞記事における語彙の時間的変化分析.『第 13 回社会言語科学会大会発表論文集』pp.113-116.

[山口ほか 2011] 山口昌也, 高田智和, 北村雅則, 間淵洋子, 大島一, 小林正行, 西部みちる(2011)『『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.2.2』, 国立国語研究所内部報告書(LR-CCG-10-04).

[山崎 2004] 山崎久之(2004)『増補補訂版 国語待遇表現体系の研究 近世編』武蔵野書院.

[山崎, 小沼 2004] 山崎誠,小沼悦(2004) 現代雑誌における語種構成. 『言語処理学会第 10 回年次大会発表論文集』pp.670-673.

[山崎 2009a] 山崎誠(2009a) 代表性を有する現代日本語書籍コーパスの構築.『人工知能学会誌』24(5), pp.623-631.

[山崎 2009b] 山崎誠(2009b) 15 シソーラス・類語辞典. 国立国語研究所『新「ことば」シリーズ 22 辞書を知る』pp.92-95.

[山崎 2011a] 山崎誠(2011a)第 2 章『現代日本語書き言葉均衡コーパス』の設計. 国立国語研究所『『現代日本語書き言葉均衡コーパス』利用の手引 第 1.0 版』, 2-1～2-8.

[山崎 2011b] 山崎誠(2011b)『現代日本語書き言葉均衡コーパス』の構築と活用.『『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集』pp.11-20.

[山崎編 2014] 山崎誠編(2014)『講座日本語コーパス 2 書きことばコーパス―設計と構築』朝倉書店.

[山下 2005] 山下喜代(2005) 日本語学習のための辞書―『漢語接辞用法辞典』の構想―.『早稲田日本語研究』14, pp.96-107.

[山田 1962] 山田俊雄(1962) 現代国語辞書の閉塞について.『文学』30(2), pp.113-123.

[山田 2002] 山田進(2002) 意味分類辞書.『國語學』53(1), pp.30-43.

[山田 2005] 山田進(2005) 辞書の意味記述.『レキシコンフォーラム』1,ひつじ書房, pp.47-63.

[山田 2006] 山田進(2006) 意味から引く辞書. 倉島節尚(編)(2006)『日本語辞書学の構築』おうふう.

[山田 2007] 山田進(2007) コロケーションの記述と名詞の意味分類.『日本語学』26(12), 明治書院, pp.48-57.

[山田 2012] 山田進(2012)「ゲットする」と「タッチする」: 外来語動詞の新用法.『聖心女子大学論叢』119, pp.147-170.

[山梨 1995] 山梨正明(1995)『認知文法論』ひつじ書房.

[山根監 2013] 山根智恵(監)(2013)『研究社 日本語口語表現辞典』研究社.

[山本 2013] 山本康一(2013) 電子辞書の意義と今後『日本語学』32(2), 明治書院, pp.48-61.

[兪 2009] 兪暁明(2009) 重言とコロケーション―その関連性と認め方―.『北陸大学紀要』33, pp.125-141.

[楊, 矢澤 2010] 楊ソルラン, 矢澤真人(2010) 漢語動詞の自他認定. 荻野綱男(編)『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 コーパスを利用した国語辞典編集法の研究』, pp.32-53.

[横山 2006] 横山晶一(2006) 新明解国語辞典の電子化. 倉島節尚(編)(2006)『日本語辞書学の構築』おうふう.

[吉田 2012] 吉田一彦(2012) 日本語教育における辞書―日本の大学に留学中の学習者を中心に―『外国語教育研究』外国語教育学会紀要 15, pp.121-126.

[ランドウ 1998] シドニー ランドウ (小島義郎, 増田秀夫, 高野嘉明訳)(1998)『辞書学のすべて』(Sidney I. Landau (1984) *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press).

[李ほか 2007] 李在鎬, 鈴木幸平, 永田由香, 黒田航, 井佐原均(2007) 動詞「流れる」の語形と意味の問題をめぐって.『計量国語学』26-2, p.64-74.

[梁 1996] 梁塵昇(1996) IPAL 辞書の自動的ハイパーテキスト化. 北陸先端科学技術大学院大学修士論文.

関連発表論文

【学術誌】

[桑畑ほか 1998] **桑畑(柏野)和佳子**, 橋本三奈子, 青山文啓(1998) IPAL 名詞辞書による多義性解消のためのコロケーションの分析.『情報処理学会論文誌』39(6), pp.1925-1934.

[柏野, 本多 1998] **柏野和佳子**, 本多啓(1998) IPAL 名詞辞書における多義構造の記述.『情報処理学会論文誌』39(9), pp.2603-2612.

[柏野ほか 2005] **柏野和佳子**, 山口昌也, 桐生りか(2005) 新聞記事データベースを利用した外来語の出現率の推移調査.『自然言語処理』12(4),pp. 97-116.

[柏野 2011] **柏野和佳子**(2011) コーパスに基づく辞書づくり —これからの国語辞典はこう変わる—.『日本知能情報ファジイ学会誌』23(5), pp.705-713.

[柏野, 奥村 2012] **柏野和佳子**, 奥村学(2012) 和語や漢語のカタカナ表記 —「現代日本語書き言葉均衡コーパス」の書籍における使用実態—.『計量国語学』28(4), pp.153-161.

[Maekawa et al. 2014] Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, **Wakako Kashino**, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, Yasuharu Den (2014) Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, pp.1-27.

[柏野, 奥村 2015] **柏野和佳子**, 奥村学(2015)「コーパスベース国語辞典」構築のための「古風な語」の分析と記述.『自然言語処理』21(6),pp.1133-1161.

【国際会議】

[Hasimoto et al. 1994] Minako Hasimoto, **Wakako Kuwahata,** Fumihiro Aoyama, Toshiyuki Tonoike (1994) Some Remarks on Ways to Compile Japanese Lexicons for Computers. *(SNLR'94)*, pp.115-122.

[Kuwahata et al.1995] **Wakako Kuwahata**, Minako Hasimoto, Kenichi Murata (1995) Construction of IPAL-BN : IPA Lexicon of Basic Japanese Nouns. *Proc. of the Natural Language Processing Pacific Rim Symposium '95(NLPRS'95)*pp.137-142.

[Kuwahata et al.1996] **Wakako Kuwahata**, Minako Hasimoto, Fumihiro Aoyama (1996) Senses of Polysemous Nouns : Building a Computational Lexicon of Basic Japanese Nouns. *Proc. of The 16th International Conference on Computational Linguistics (COLING1996)*, pp. 1082-1085.

[Hasida et al. 1998] Koiti Hasida, Hitoshi Isahara, Takenobu Tokunaga, Minako Hashimoto, Shiho Ogino, **Wakako Kashino**, Jun Toyoura, Hironobu Takahashi (1998) The RWC Text Databases. *Proc. of The First International Conference on Language Resource and Evaluation*, pp.457-461.

[Kashino and Okumura 2010] **Wakako Kashino**, Manabu Okumura (2010) An Approach toward Register Classification of Book Samples in the Balanced Corpus of Contemporary Written Japanese. *Proc. of the 24th Pacific Asia Conference on Language, Information and Computation (PACLIC24)*, pp.433-438.

[Maekawa et al. 2010] Kikuo Maekawa, Makoto Yamazaki, Takehiko Maruyama, Masaya Yamaguchi, Hideki Ogura, **Wakako Kashino**, Toshinobu Ogiso, Hanae Koiso, Yasuharu Den (2010) Design, Compilation, and Preliminary Analyses of Balanced Corpus of Contemporary Written Japanese. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC2010)*, pp. 1483-1486.

【その他】

[橋本ほか 1991] 橋本三奈子, **桑畑(柏野)和佳子**, 村田賢一(1991) 計算機用日本語基本名詞辞書の試作.『情報処理学会全国大会講演論文集』43, pp. 3.213-3.214.

[桑畑ほか 1992] **桑畑(柏野)和佳子**, 橋本三奈子, 外池俊幸, 村田賢一(1992) 計算機用日本語基本名詞辞書の概要.『情報処理学会全国大会講演論文集』45, pp. 3.203-3.204.

[橋本ほか 1992] 橋本三奈子, **桑畑(柏野)和佳子**, 青山文啓, 村田賢一(1992) 計算機用日本語基本名詞辞書における記述項目の構成.『情報処理学会全国大会講演論文集』45, pp. 3.201-3.202.

[桑畑ほか 1993a] **桑畑(柏野)和佳子**, 橋本三奈子, 村田賢一(1993a) 計算機用日本語辞書の開発.『情報処理学会研究報告人文科学とコンピュータ』42, pp. 27-34.

[桑畑ほか 1993b] **桑畑(柏野)和佳子**, 橋本三奈子, 村田賢一(1993b) 計算機用日本語基本名詞辞書の見出し語の選定.『電子情報通信学会大会講演論文集』1993, pp. 6.71.

[橋本ほか 1993] 橋本三奈子, **桑畑(柏野)和佳子**, 青山文啓, 村田賢一(1993) コーパスデータに基づく名詞コロケーションの辞書記述.『情報処理学会全国大会講演論文集』47, pp. 3.63-3.64.

[桑畑ほか 1994] **桑畑(柏野)和佳子**, 橋本三奈子, 井口厚夫, 猪塚元, 村田賢一(1994) 名詞辞書におけるコロケーションの記述.『情報処理学会全国大会講演論文集』49, pp. 145-146.

[桑畑, 橋本 1994] **桑畑(柏野)和佳子**, 橋本三奈子(1994) 計算機用日本語基本名詞辞書における見出し語の下位区分.『情報処理学会研究報告自然言語処理』104, pp.1-8.

[橋本ほか 1994] 橋本三奈子, **桑畑(柏野)和佳子**, 青山文啓, 村田賢一(1994) 名詞の比喩的表現とその統語的特徴.『情報処理学会全国大会講演論文集』49, pp. 139-140.

[桑畑ほか 1995] **桑畑(柏野)和佳子**, 橋本三奈子, 村田賢一(1995) 名詞辞書における下位区分間の意味的関係の記述.『情報処理学会全国大会講演論文集』51, pp. 43-44.

[桑畑, 橋本 1995] **桑畑(柏野)和佳子**, 橋本三奈子(1995) 名詞の下位区分間に見られる意味的関係の辞書記述.『情報処理学会研究報告自然言語処理』110, pp.29-34.

[桑畑ほか 1996] **桑畑(柏野)和佳子**, 橋本三奈子, 青山文啓(1996) IPAL ハイパーテキスト化のためのコロケーションの多義性の分析.『情報処理学会全国大会講演論文集』53, pp. 99-100.

[桑畑, 本多 1996] **桑畑(柏野)和佳子**, 本多啓(1996) 区分間の意味的関係.『情報処理振興事業協会計算機用日本語基本名詞辞書 IPAL(Basic Nouns)解説編』pp.211-227.

[橋本ほか 1996] 橋本三奈子, 岡部了也, **桑畑(柏野)和佳子**, 斉藤初江, 佐藤幸子, 中川建司, 山下智弥, 青山文啓(1996) 統合辞書における文型の記述.『情報処理学会全国大会講演論文集』53, pp. 97-98.

[橋本ほか 1997] 橋本三奈子, 山下智弥, **桑畑(柏野)和佳子**(1997) 文型テーブルを用いた統合辞書の試作.『ソフトウェア文書のための日本語処理の研究-13』情報処理振興事業協会.

[本多, 桑畑 1997] 本多啓, **桑畑(柏野)和佳子**(1997) 名詞の多義について —IPAL名詞辞書のための研究から.『言語処理学会第 3 回年次大会発表論文集』pp.83-86.

[柏野 1998] **柏野和佳子**(1998) 曖昧性解消過程解明のための多義語の分析.『国立国語研究所創立 50 周年記念研究発表会資料集』pp.69-72.

[柏野, 本多 1998] **柏野和佳子**, 本多啓(1998) 多義構造を辞典に書く.『日本語学』明治書院, 17(14), pp.54-63.

[柏野 1999a] **柏野和佳子**(1999a) 解析と生成のための共起情報の記述方法.『言語処理学会第 5 回年次大会発表論文集』pp.205-208.

[柏野 1999b] **柏野和佳子**(1999b) 共起情報の辞書記述方法—解析と生成の視点を同時にもつ記述方法の提案.『電子情報通信学会技術研究報告言語理解とコミュニケーション』99(88), pp.1-8.

[村田ほか 1999] 村田賢一, 石田直子, **柏野和佳子**, 常盤僚子, 西川賢哉(1999) 計算機用日本語生成辞書 IPAL(SURFACE/DEEP)の試作.『情報処理学会研究報告自然言語処理』130, pp.105-112.

[白井ほか 2001] 白井清昭, **柏野和佳子**, 橋本三奈子, 徳永健伸, 有田英一, 井佐原均, 荻野紫穂, 小船隆一, 高橋裕信, 長尾確, 橋田浩一, 村田真樹(2001) 岩波国語辞典を利用した語義タグ付きテキストデータベースの作成.『情報処理学会研究報告自然言語処理』141, pp.117-124.

[柏野 2005] **柏野和佳子**, 丸山岳彦, 稲益佐知子, 茂木俊伸(2005) 語の出現分布からみた月刊雑誌と新聞コーパスの特性調査 —用例収集資料としての多様性の検討—.『言語処理学会第 11 回年次大会発表論文集』pp.380-383.

[柏野 2006a] **柏野和佳子**(2006a) 国語辞典の釈義と用例の検討.『言語処理学会第 12 回年次大会発表論文集』pp.8-11.

[柏野 2006b] **柏野和佳子**(2006b) 国語辞典における多義語の意味区分の比較.『人工知能学会第 2 種研究会ことば工学研究会資料』23,pp.37-43.

[柏野 2006c] **柏野和佳子**(2006c) 書き言葉コーパスで探る日本語のありさま.『日本語学』明治書院, 25(9), pp.37-43.

[山崎ほか2006] 山崎誠, 前川喜久雄, 田中牧郎, 小椋秀樹, **柏野和佳子**,小磯花絵,間淵洋子, 丸山岳彦, 山口昌也, 秋元祐哉, 稲益佐知子, 吉田谷幸宏(2006) 代表性を有する現代

日本語書き言葉コーパスの設計.『言語処理学会第 12 回年次大会発表論文集』pp.440-443.

[柏野 2007a] **柏野和佳子**(2007a) 第5章 外来語言い換え提案対象語の頻度推移類型―読売新聞・毎日新聞 14 年分を調査して―.『国立国語研究所報告 126 公共媒体の外来語―「外来語」言い換え提案を支える調査研究―』pp.387-401.

[柏野 2007b] **柏野和佳子**(2007b) 国語辞典における多義語の意味記述の比較.『言語処理学会第 13 回年次大会発表論文集』pp.863-866.

[柏野 2007c] **柏野和佳子**(2007c) 新聞で捉える日本語の姿 ―記事データベースが明らかにする特徴とは.『新聞研究』日本新聞協会, 666, pp.21-24.

[柏野 2008] **柏野和佳子**(2008) 書籍の文章の多様性を捉える観点付与の設計 ―『現代日本語書き言葉均衡コーパス』の収録文章を対象に.『人工知能学会第 2 種研究会ことば工学研究会資料』30, pp.11-22.

[柏野ほか 2008] **柏野和佳子**, 丸山岳彦, 秋元祐哉, 稲益佐知子, 佐野大樹,田中弥生,山崎誠(2008). 書籍の生産実態を反映するサンプリング ―NDC ごとに取得したサンプルの多様性の分析.『言語処理学会第 14 回年次大会発表論文集』pp.939-942.

[柏野 2009] **柏野和佳子**(2009) 1.3 辞書. 言語処理学会編『言語処理学事典』共立出版, pp.80-85.

[柏野, 奥村 2009] **柏野和佳子**, 奥村学(2009) 和語や漢語のカタカナ表記 ―『現代日本語書き言葉均衡コーパス』における使用実態―.『計量国語学会第53回大会予稿集』pp.38-43.

[柏野ほか2009] **柏野和佳子**, 稲益佐知子, 田中弥生, 秋元祐哉(2009) 第 4 章 対象外要素の排除指定.『特定領域研究「日本語コーパス」平成 20 年度研究成果報告書『現代日本語書き言葉均衡コーパス』における収録テキストの抽出手順と事例』pp. 66-88.

[柏野 2010] **柏野和佳子**(2010)「直接的な語り」という表現スタイルをもつ書籍テキストの人手抽出の試み.『人工知能学会第 2 種研究会ことば工学研究会資料』35, pp.63-72.

[柏野, 奥村 2010] **柏野和佳子**, 奥村学(2010) 国語辞典に「古い」と注記される語の現代書き言葉における使用傾向の調査.『情報処理学会研究報告人文科学とコンピュータ』88, pp. 1-6.

[間淵ほか 2010] 間淵洋子, **柏野和佳子**, 山口昌也, 高田智和(2010) コーパスを用いたテキスト分類指標の検討―BCCWJ の文書構造情報分析を中心に―.『言語処理学会第 16 回年次大会予稿集』pp.314-317.

[柏野, 奥村 2011] **柏野和佳子**, 奥村学(2011) 国語辞典に「古風」と注記される語の使用実態調査 —『現代日本語書き言葉均衡コーパス』を用いて—.『言語処理学会第 17 回年次大会発表論文集』pp.444-447.

[丸山ほか 2011] 丸山岳彦, 山崎誠, **柏野和佳子**, 佐野大樹, 秋元祐哉,稲益佐知子,田中弥生, 大矢内夢子(2011)『『現代日本語書き言葉均衡コーパス』に含まれるサンプルおよび書誌情報の設計と実装』, 国立国語研究所内部報告書(JC-D-10-02).

[柏野, 奥村 2012a] **柏野和佳子**, 奥村学(2012a) 書籍テキストへの分類指標人手付与の試み —『現代日本語書き言葉均衡コーパス』の収録書籍を対象に—.『言語処理学会第 18 回年次大会発表論文集』pp.1260-1263.

[柏野, 奥村 2012b] **柏野和佳子**, 奥村学(2012b) 和語や漢語のカタカナ表記 —「現代日本語書き言葉均衡コーパス」の書籍における使用実態—.『計量国語学』28(4), pp.153-161.

[柏野ほか 2012a] **柏野和佳子**, 立花幸子, 保田祥, 丸山岳彦, 奥村学, 佐藤理史, 徳永健伸, 大塚裕子, 佐渡島紗織(2012a) テキストの硬さと軟らかさの考察 —『現代日本語書き言葉均衡コーパス』の収録書籍を対象に—.『第 1 回コーパス日本語学ワークショップ予稿集』pp.131-138.

[柏野ほか 2012b] **柏野和佳子**, 立花幸子, 保田祥(2012b). 書籍テキストをその形式, 内容, 表現に関わる特徴によって分類する.『人工知能学会第 2 種研究会ことば工学研究会資料』41, pp.21 -29.

[柏野ほか 2012c] **柏野和佳子**, 立花幸子, 保田祥, 飯田龍, 丸山岳彦, 奥村学, 佐藤理史, 徳永健伸, 大塚裕子, 佐渡島紗織, 椿本弥生, 沼田寛(2012c) 書籍テキストへの文体情報付与の試み.『第 2 回コーパス日本語学ワークショップ予稿集』pp.155-164.

[佐藤, 柏野 2012] 佐藤理史, **柏野和佳子**(2012) テキストの難易度に対する人間の判断と機械の判断.『第 1 回コーパス日本語学ワークショップ予稿集』pp.195-202.

[保田ほか 2012a] 保田祥, **柏野和佳子**, 立花幸子, 丸山岳彦(2012a).「「語り性」を有する書きことばの典型例の分析.『第 1 回コーパス日本語学ワークショップ予稿集』pp.139-146.

[保田ほか 2012b] 保田祥, **柏野和佳子**, 立花幸子(2012b). 総体として印象を与える表現 —「語りかけ性」を有すると判断する根拠.『人工知能学会第 2 種研究会ことば工学研究会資料』41, pp.3-10.

[保田ほか 2012c] 保田祥, **柏野和佳子**, 立花幸子, 丸山岳彦(2012c).「語りかけ性」を有すると判断される書きことばの表現.『第 2 回コーパス日本語学ワークショップ予稿集』pp.43-50.

[柏野，中村 2013] **柏野和佳子**，中村壮範(2013) 現代日本語書き言葉における非外来語のカタカナ表記事情.『第 4 回コーパス日本語学ワークショップ予稿集』pp.285-290.

[柏野ほか 2013] **柏野和佳子**，立花幸子，保田祥，飯田龍，丸山岳彦，奥村学， 佐藤理史，徳永健伸，大塚裕子，佐渡島紗織，椿本弥生，沼田寛(2013) BCCWJ 図書館サブコーパス全テキストへの文体情報付与結果の分析.『第 3 回コーパス日本語学ワークショップ予稿集』pp.63-70.

[保田ほか 2013a] 保田祥，**柏野和佳子**，立花幸子，丸山岳彦(2013a) 書きことばにおける「語りかけ性」は何のために用いられるのか.『第 3 回コーパス日本語学ワークショップ予稿集』pp.143-152.

[保田ほか 2013b] 保田祥，**柏野和佳子**，立花幸子，丸山岳彦(2013b) アノテーターコメントを用いた「語りかけ性」分析の試み ―頻度情報から捉え難いテキスト性質の解明に向けて―.『言語処理学会第 19 回年次大会発表論文集』pp.358-361.

[保田ほか 2013c] 保田祥，立花幸子，**柏野和佳子**，丸山岳彦(2013c) 「ベテランは足を保護する」が語りかけ性るとき.『第 4 回コーパス日本語学ワークショップ予稿集』pp.345-354.

[柏野，中村 2014] **柏野和佳子**，中村壮範(2014) BCCWJ 図書館サブコーパスの文体情報検索ツールによるテキスト分析.『第 5 回コーパス日本語学ワークショップ予稿集』pp.171-180.

[浅原ほか 2014] 浅原正幸，加藤祥，立花幸子，**柏野和佳子**(2014) 文体指標と語彙の対応分析.『第 6 回コーパス日本語学ワークショップ予稿集』pp.11-20.