

## 第5章 形態論情報

小椋 秀樹 富士池 優美

BCCWJには、他の章でも解説されているように種々のアノテーションが施されている。これらのアノテーションを利用することで、コーパスを活用した現代日本語の研究が、今後、大きく進展することが期待される。

これら種々のアノテーションのうち形態論情報については、BCCWJに格納したサンプルの言語的特徴の解明に適した「長単位」とコーパスからの用例収集に適した「短単位」の2種類の言語単位に解析し、それぞれの単位に見出し・品詞・語種等の情報を付与した。解析精度は長単位・短単位とも、データ全体に対して人手修正を行ったコアデータ<sup>1</sup>は99%以上、データの一部に対して人手修正を行ったコアデータ以外のデータ（非コアデータ）は98%以上である。

本章では、BCCWJ-DVD版、オンライン版（中納言）で形態論情報を活用しようとする研究者の用に資するため、形態論情報、特に長単位・短単位の認定基準を中心に、その概要を紹介する<sup>2</sup>。

### 5.1 BCCWJの言語単位

本節では、まず国立国語研究所がこれまでに行ってきた語彙調査における調査単位を概観し、続いてBCCWJの言語単位の設計方針、採用した長短2種類の言語単位の長所、短単位の自動解析に使用した形態素解析用辞書UniDicについて述べる。

#### 5.1.1 語彙調査の調査単位

国立国語研究所は、これまでに、マスメディアにおける書き言葉や話し言葉を中心に、合計10回の大規模な語彙調査を実施してきた。この語彙調査に当たっては、当然、語というものを規定することが必要となる。しかし、語の定義については研究者によって様々な立場があるため、語彙調査において語（調査単位）をどのように規定するかということは常に大きな問題となる。

国立国語研究所がこれまでに行った語彙調査では、調査単位の設計に当たって、語とは何かという本質的な議論の上に立って調査単位を設計するという立場は取っていない。それぞれの語彙調査の目的に応じて最もふさわしい単位を設計するという方針の下に、一貫して操作主義的な立場を取ってきた。そのため、表5.1に示すように、複数の調査単位が使われてきた<sup>3</sup>。

<sup>1</sup> コアデータについては、第2章を参照。

<sup>2</sup> BCCWJの形態論情報の詳細については、小椋他（2011）を参照。

<sup>3</sup> 単位の概略と例については、林（1982:582-583）、中野（1998:171-172）を基にした。

表 5-1: 国立国語研究所の語彙調査における主な調査単位

	単位の名称	語 彙 調 査 名
長い単位の系列	α 単位	現代の語彙調査・婦人雑誌の用語
	W 単位	高校教科書の語彙調査、中学校教科書の語彙調査
	長い単位	雑誌用語の変遷、テレビ放送の語彙調査
短い単位の系列	β 単位	現代の語彙調査・総合雑誌の用語、現代雑誌九十種の用語用字、雑誌 200 万字言語調査
	M 単位	高校教科書の語彙調査、中学校教科書の語彙調査

表 5-1 に挙げた各調査単位の概略と例とを次に示す。

### 【調査単位の概略】

- (1) 長い単位の系列 : 主として構文的な機能に着目して考えた単位。おおむね文節に相当する。

α 単 位 文節を基にした単位。「| 小学校 | 卒業 |」「| 男児用 | 外出着 |」のように長い語を分割する規定を設けている。

W 単 位 非活用語および活用語のうち終止・連体形、命令形、中止用法・修飾用法の連用形を 1 単位とする。また、それらに接続する付属語も 1 単位とする。

長い単位 文節に相当する単位。「テレビ放送の語彙調査」の長い単位は、複合辞を助詞・助動詞として扱っていること、人名・地名のほか書名・番組名・商品名なども固有名詞として扱っていることから、「雑誌用語の変遷」で採用した長い単位よりも長くなっている。

- (2) 短い単位の系列 : 主として言語の形態的な側面に着目して考えた単位。

β 単 位 原則として、現代語において意味を持つ最小の単位（最小単位）二つが、文節の範囲内で一次結合したものを 1 単位とする。

M 単 位 β 単位と同様に最小単位を基にした単位。漢語は、β 単位と同様に二つの最小単位が文節の範囲内で一次結合したものを 1 単位とするが、和語・外来語は 1 最小単位を 1 単位とする。

### 【調査単位の例】

- (1) 長い単位の系列

α 単位 : 型 紙 | どおり に | 裁断 し て | 外出 着 を | 作り まし た |

W 単位 : 型 紙 だおり | に | 裁断 し て | 外出 着 | を | 作り まし た |

長い単位 (雑誌用語の変遷) :

型 紙 だおり に | 裁断 し て | 外出 着 を | 作り まし た |

長い単位（テレビ放送の語彙調査）：

型紙どおりに|裁断して|外出着を|作りまし|た|  
その|問題について|検討している|

## (2) 短い単位の系列

β単位： 型紙|どおり|に|裁断|し|て|外出|着|を|作り|まし|た|

M単位： 型|紙|どおり|に|裁断|し|て|外出|着|を|作り|まし|た|

調査単位的设计に当たって操作主義的な立場を取ってきたのは、「必要以上に学術的な議論に深入りし、実際上の作業がすすまないことをおそれたため」（国立国語研究所1987:12）であり、「学者の数ほどもある「単語」の定義について、まず、意見を一致させてから、というのでは、見とおしがたたない。」（同:12）からである。

このような立場に対しては、当然のことながら「語というのは何なのか、調査のため便宜的に設けられた単位にすぎないのかという問題が残る。」（前田1985:740）という批判がある。確かに、語というものを定義しようとする以上、語とは何かという本質的な議論を積み重ねていくことは重要なことである。しかし、国立国語研究所（1987:12）に、「原則的にただしい定義に達したとしても、それが現実の単位きり作業に役立たないならば、無意味である。語彙調査というのは、現象の処理なのだから。」と述べられているように、語彙調査においては対象とする言語資料に現れた個々の事象を、的確に処理することも極めて重要である。このことから、これまでの語彙調査では、語とは何かという本質的な議論よりも、言語現象を的確に処理することを重視してきた。

このような立場で、各種の語彙調査を進めてきたことにより、「同じ資料の語彙調査を短単位と長単位との両方で行ってみてどのような違いが出てくるかを検討したことなどは、単位の区切り方を曖昧にしたまま「語彙調査」を行なうことに対する反省を促す」（前田1985:740）など、日本語の計量的な研究を進める上で先駆的な役割を果たしてきたと言える。国立国語研究所の語彙調査における調査単位的设计方針には批判もあるが、それにより現実の言語現象を的確に処理してきたことは、十分に意味があったと言える。

### 5.1.2 BCCWJの言語単位的设计方針

BCCWJの言語単位的设计に当たっては、語彙調査における調査単位的设计と同様の立場を取った。つまり、まずBCCWJを日本語研究に利用するために、どのような言語単位が必要か整理し、その上で設計方針を立て、その方針に基づいて言語単位を設計したのである。

このような立場を取ったのは、語とは何かという本質的な議論の重要性はもちろん認めるところではあるが、コーパス構築という実務を考えた場合、BCCWJに現れる言語現象を

的確に処理できる単位を設計することの方が、より重要であると考えたからである。このようにして大規模なコーパスを処理した結果をまとめておくことは、今後、言語単位論を進める上での基礎的な資料になると考えられる。

我々は、BCCWJの言語単位的设计方針として、次の三つを掲げた。

**方針 1:** コーパスに基づく用例収集、各ジャンルの言語的特徴の解明に適した単位を設計する。

コーパスの日本語研究への活用としてまず考えられるのは、コーパスから用例を集めることである。そのため、BCCWJを日本語研究で幅広く利用できるようにするには、用例収集に適した単位を設計する必要がある。

また BCCWJ は、新聞・雑誌・書籍といった複数の媒体を対象としたコーパスであり、内容も政治・経済・自然科学・文芸等と多岐にわたっている。このような BCCWJ の構成から、媒体別・ジャンル別の言語的な特徴を明らかにしていくことが重要な研究テーマになると考えられる。したがって、そのような分析に適した単位を設計することが必要になる。

**方針 2:** 『日本語話し言葉コーパス』と互換性のある形態論情報を設計する。

国立国語研究所が既に構築したコーパスとして、現代の話し言葉を対象とした『日本語話し言葉コーパス』(Corpus of Spontaneous Japanese、以下 CSJ とする。)がある<sup>4</sup>。CSJ、BCCWJ は共に現代日本語を対象とした大規模コーパスであり、日本語研究の立場からは、両コーパスを活用した現代日本語の話し言葉・書き言葉の研究を進めていくことが重要なテーマとなる。このような研究を進めるためには、CSJ と BCCWJ とを統一的に扱うことのできる互換性を持った単位を設計する必要がある。

**方針 3:** 国立国語研究所の語彙調査における知見を活用する。

国立国語研究所は、1949 年の『語彙調査 —現代新聞用語の一例—』以来、合計 10 回の語彙調査を実施した。その中で、調査単位的设计や言語現象の処理に関して、様々な知見を蓄積している。そこで、BCCWJ の言語単位的设计や単位認定の際に、これら語彙調査の知見を活用していく。語彙調査の結果は、日本語研究でも様々な活用されており、言語単位的设计等に語彙調査の知見を活用していくことは、BCCWJ を使った日本語研究を進めていくためにも有用であると考えられる。

### 5.1.3 BCCWJ の言語単位

以上の方針の下、BCCWJ の言語単位について検討した結果、次のような結論を得た。

BCCWJ の言語単位には、方針 1 で挙げた、用例収集・各ジャンルの言語的特徴の解明

---

<sup>4</sup> CSJ の言語単位の概要については、小椋 (2006) を参照。

という二つの利用目的に応じて、次に示す 2 種類を採用する。

- ① 用例収集を目的とした短単位
- ② 言語的特徴の解明を目的とした長単位

この短単位・長単位は、いずれも CSJ で採用した言語単位である。また短単位は国立国語研究所が行った現代雑誌九十種調査の  $\beta$  単位を、長単位はテレビ放送の語彙調査の長い単位を基に設計したものである。このようにして、CSJ との互換性の保持と、国立国語研究所の持つ語彙調査の知見の活用とを図る。なお、長単位・短単位認定規程は、CSJ の規程をそのまま用いるのではなく、書き言葉用に修正・拡張を行っている。CSJ の規程からの変更点については、5.4 節で述べる。

#### 5.1.4 長単位・短単位の長所

ここでは、長単位・短単位がコーパスの言語単位として、どのような長所を持つのかについて述べる。

##### (1) 長単位の長所

長単位の長所としては、次の 2 点が挙げられる。

**長所 1：**当該資料の性格を反映する特徴的な語を取り出しやすい。

一般に単位を短くすればするほど、取り出した単位はいわゆる基本的な語となる。短単位は基準が分かりやすくゆれが少ないため、用例収集を行う上では便利な単位であるが、合成語を構成要素に分割してしまうという問題点がある。これに対して長単位では、「国立国語研究所」「品詞比率」「分析する」のような合成語を 1 単位として認める。「を」「だ」のような付属語は単独で長単位とするのが原則であるが、「における」「ている」のような複合辞も付属語として 1 長単位としている。

コアデータを基に、どのような語と結合するかという点から、掲載媒体別の差異を見る。以下、「生活」という語を例に説明する。コアデータ約 100 万語中、「生活」は 508 例見られる。そのうち、「生活」単独で使われた例が 153 例、合成語の構成要素として使われた例が 355 例と、「生活」は合成語の構成要素として使われやすい傾向にあることが分かる。掲載媒体に注目して、「生活」を含む語を見てみよう。1 媒体のみに現れる語には、白書の「基礎的生活コスト」「国民生活選好度調査」「WHO 国際生活機能分類」「労働者生活」、新聞の「生活面子育て相談室係」、web の「残業生活する」「入院生活する」「週末泥沼生活」等、資料の内容・性格を反映したものが見られる。

「労働者生活」を「労働」と「者」と「生活」とに、「残業生活する」を「残業」と「生活」と「する」とに分割するのではなく、全体でひとつとして扱う長い単位を使うことで、各ジャンルの特徴的な語を把握することができる。

**長所 2:** 文脈に即した品詞が付与されている。

短単位には、「名詞-普通名詞-形状詞可能」のような曖昧性を持たせた品詞がある。これに対して、長単位では文脈に即して品詞を付与する方針を取り、「名詞-普通名詞-○○可能」といった品詞は設けず、その文脈での用法に基づき名詞・形状詞・副詞に判別する。「結果」を例にすると、短単位では一律に「名詞-普通名詞-副詞可能」という品詞が付与されるが、長単位では、「これらの結果に基づき」の場合は「名詞-普通名詞-一般」とし、「結果、様々な社会問題が発生し」の場合は「副詞」とする。

コアデータを基に、名詞・形状詞・副詞の判別を行った長単位データと判別を行わない短単位データとの品詞比率の差異を見ると、判別を行った場合、書籍の形状詞率、新聞の副詞率が判別を行わない場合よりも高くなっており、白書では判別による変化が少ないことがわかった(富士池他 2011)。これらは、用法に基づき、名詞・形状詞・副詞の判別をした結果、媒体の特徴がより明確になったものと考えられる。このように文脈に即して品詞を判別した長い単位を使うことで、構文的な機能を見る際に、より精密な分析が可能になる。

以上の二つの長所から、長単位は構文的な機能に着目した、各媒体・各ジャンルの資料的な性格を反映する単位であり、言語的特徴を解明するという目的にかなうものと言える。

## (2) 短単位の長所

短単位の長所としては、次の 2 点が挙げられる。

**長所 1:** 基準が分かりやすく、ゆれが少ない。

これは、短単位の基礎となる最小単位の認定に当たり、個人によって捉え方に幅のある要素を基準に持ち込んでいないことによる。

基準が分かりやすく、ゆれが少ないという短単位の長所は、作業効率の向上につながるだけでなく、コーパスの使いやすさにもつながる。基準が分かりやすければ、利用者が語を検索する際、どのように検索条件を指定すればよいか迷うことが少なくなる。また、ゆれの少なさ、つまりデータの精度の高さは、分析結果の確かさにもつながる。

**長所 2:** 取り出した単位が文脈から離れすぎない。

上で短単位はゆれが少ない単位であると述べたが、実は最もゆれが少ない単位は、短単位ではなく、その基礎となっている最小単位である。それにもかかわらず、最小単位を言語単位として採用しなかったのは、最小単位は文脈から離れすぎるため、日本語の研究に使いにくいからである。

例えば、短単位「気持ち」は「気」と「持ち」の二つの最小単位に分割することができる。もしこのような最小単位でコーパスが解析されていると、動詞「持つ」を検索した際

に、「荷物を持つ」などの「持つ」とともに、「気持ち」の「持ち」も検索結果として得られることになる。

しかし、動詞「持つ」の分析を行う際に、「気持ち」の「持ち」まで検索結果に含まれるのは望ましいとは言い難い。それは、実際の文脈の中では、動詞「持つ」として機能していないからである。したがって、コーパスから用例を収集し、分析することを考えた場合、正確に単位認定ができるとしても、最小単位のような単位では問題が多いということになる。

以上のように考えた場合、短単位は、基準の分かりやすさ、ゆれの少なさという条件を満たしつつ、用例を収集して分析を行うという利用目的にもかなう単位と言える。

### 5.1.5 形態素解析用辞書 UniDic について

BCCWJ は、1 億語からなる大規模なコーパスであるため、形態論情報の付与は自動解析システムにより行う。短単位解析には解析エンジン MeCab と形態素解析用辞書 UniDic<sup>5</sup> を、長単位解析には短単位解析結果から長単位を自動構成する解析器<sup>6</sup>を使う。ここでは、短単位解析で用いる形態素解析用辞書 UniDic についてその概要を紹介する。

短単位解析の解析用辞書に UniDic を採用したのは、UniDic の言語単位が CSJ の短単位とほぼ同じものであり、品詞等の情報についても BCCWJ と互換性を持つものであったことによる。

また、UniDic では、表記や語形の違いにかかわらず、同じ語であれば、同一の見出しを与えるという方針を取り、語を階層化した形で登録している。この階層構造の最上位を語彙素（国語辞典の見出しに相当）と呼んでおり、この語彙素の下に語形（語形の違いを区別する層）、更に語形の下に書字形（表記の違いを区別する層）という階層を設けている（図 5-1）。

語彙素	語形	書字形
矢張り	ヤハリ	やはり
		矢張り
	ヤッパリ	やっぱり
		矢っ張り
	ヤッパ	やっぱ

図 5-1: UniDic の階層構造の例

このような階層構造で登録した辞書を用いて、コーパスを形態素解析することによって、例えば、ある語について、どのような語形の変異や表記のゆれが、どの程度あるのかという情報を容易に得られるなど、日本語研究の可能性が格段に広がることが期待される。こ

<sup>5</sup> UniDic については、伝他 (2007) を参照。

<sup>6</sup> 長単位を自動構成する解析器については、Uchimoto 他(2007)を参照。

のことも UniDic を形態素解析用辞書として採用した理由である。

我々は、BCCWJ の構築を始めた 2006 年の時点で既に伝康晴氏が中心になって構築していた UniDic (見出し語:約 46,000 語) を基に、国語辞典や国立国語研究所の語彙調査等を基に作成されたデータ、そして BCCWJ の解析結果から UniDic の未登録語を追加していく作業を継続して行った。この作業の過程で、単位の認定、品詞情報等についてすり合わせを行い、現在では、UniDic の解析結果と本書で述べる BCCWJ の短単位、品詞情報等に違いがない状態となっている。

unidic-mecab 1.3.12 の解析精度を、以下の表 5-2 に示した。

表 5-2: UniDic の解析精度 (レジスター別)

	白書	書籍 (文学)	書籍 (文学以外)	新聞	Web (Y!知恵袋)	話し言葉 (CSJ)
単位境界	0.9992	0.9976	0.9975	0.9964	0.9947	0.9963
品詞	0.9957	0.9866	0.9896	0.9891	0.9841	0.9844
語彙素	0.9938	0.9857	0.9864	0.9864	0.9804	0.9795

表中の「単位境界」「品詞」「語彙素」の意味は以下のとおりである。

単位境界：単位境界が正解と一致するか否か。

品 詞：境界に加え、品詞・活用型・活用形が正解と一致するか否か。

語 彙 素：境界、品詞・活用型・活用形に加え、語彙素が正解と一致するか否か。

通常、形態素解析辞書は、品詞レベルまでで精度評価を行っている。UniDic は、それよりも厳しい語彙素レベルについても精度評価を行い、書き言葉については、全てのジャンルで 98%以上の精度を実現している。

UniDic は、BCCWJ の構築期間中に複数回バージョンアップを行った。BCCWJ のコアデータ・非コアデータの自動解析には、その時点における最新のバージョンを用いている。そのため、ジャンルや自動解析を行った時期によって使用した UniDic のバージョンが異なっている。

また、UniDic で自動解析を行った後、コアデータ・非コアデータとも人手修正を行った。コアデータは、データ全体に対して作業担当者を変えて 3 回にわたるデータチェックを行い、誤解析の発見と修正を行った。非コアデータは、コアデータに比べて規模が極めて大きいため、データ全体に対してチェックを行うことは不可能である。そこで、未知語に起因する誤解析と疑われる箇所を中心にチェックを行い、誤解析の発見と修正を行った。

以上、本節では、国立国語研究所の語彙調査における調査単位について概観した上で、BCCWJ の言語単位について設計方針等、その概要を述べた。5.2 節以降では、長単位と短単位の認定規程の概要を紹介するが、その際、以下の凡例に示した記号を用いて単位境界や単位のつなぎ目を示す。



《 凡 例 》

1. 各規程に示した例は、コーパスに現れた例又は作例である。
2. 文節・長単位・最小単位・短単位の境界を示すために次の記号を用いた。

文節の境界	……………		例：  国立国語研究所の
長単位の境界 (5.2節)	……………		例：  国立国語研究所   の
最小単位の境界	……………	/	例：/ 国 / 立 / 国 / 語 / 研 / 究 / 所 /
短単位の境界 (5.3節)	……………		例：  国立   国語   研究   所
当該規定で着目している箇所	…		例：  国立国語研究所の
3. 分割しないことを特に示す必要があるときには、次の記号を用いた。

文節・長単位のつなぎ目	……………	-	例：  からかわれて-ばかり-いる     大-丈夫   です
当該規定で着目している箇所	…	=	例：  からかわれて=ばかり=いる     大=丈夫   です
4. 着目している文節・長単位が分かりにくい場合は、当該箇所に下線を付した。

## 5.2 長単位

長単位は、言語の構文的な機能に着目して規定した言語単位である。長単位の認定は、文節の認定を行った上で、各文節の内部を規定に従って自立語部分と付属語部分とに分割していくという手順で行う。そのため、長単位の認定規程は、文節と長単位の二つの認定規程から成る。

以下、本節では文節認定規程・長単位認定規程のうち主要な規定を紹介する。また、長単位に付与する付加情報についても、その概要を述べる。

### 5.2.1 文節認定規程

文節の認定方法（区切り方）については、いわゆる学校文法によって広く知られているところである。ここでは、BCCWJの文節認定規程のうち、学校文法における文節の認定方法と異なる点をはじめ、特に注意すべき事項について概要を示す。

学校文法と異なる点としては、同格・並列の扱いが挙げられる。学校文法では、同格の関係にある要素、並列の関係にある要素は、以下のように切り離される。

**【例】**

〔同格〕 | 大江健三郎さんの | 長男 || 光さん |

〔並列〕 | 公正 || 妥当な | 実務慣行を | 集約した | ものという | 意味で |

しかし BCCWJ では、以下のとおり同格の関係にある要素、並列の関係にある要素を切り離さないこととした。

### 【例】

- 〔同格〕 | 大江健三郎さんの | 長男=光さん |  
| 東海汽船の | 支店長=・=重久さんは、 |  
| 機関誌=計量国語学・発行の | 年に |  
| 中国語日刊新聞=「=星島日報」 |
- 〔並列〕 | 公正=妥当な | 実務慣行を | 集約した | ものという | 意味で |  
| 麦=・=大豆=・=飼料作物の | 生産振興に | 資する | 水田の | 汎用化を |  
| 最も | 先進的な | 青森=・=岩手=・=秋田の | 北東北三県は、 |  
| 東京の | 郊外の | 市=町=村と | 言うか |

同格・並列の扱い以外で、特に注意すべき事項について、以下、その認定規定と例とを示す。

### 【句読点・空白に関する規定】

- (1) 句読点（句読点として用いられているカンマ・ピリオド・エクスクラメーションマーク・クエスチョンマーク、三点リーダー、並びにコロンを含む。）および空白の後ろで切る。

- 【例】 | 不合格には、 || 違いはないでしょうが。 ||  
| 十五歳少女が | 最年少記録 | 「エベレスト登頂」 | 三浦さん最高齢記録 | | | |
| その | 日に... ||  
| 第2部 || 森林 | 及び | 林業に関して | 講じた | 施策 |  
| 2 | 協力的自主国防推進： || 自主国防と | 米韓同盟が |  
ただし、文頭の空白の後ろでは切らない。

- 【例】 | =それは、 | 現実の | 世界情勢が |

### 【付属語に関する規定】

- (2) 助詞・助動詞・接尾辞連続（言いよどみの助詞・助動詞・接尾辞も含む。）の後ろで切る。助詞・助動詞には付録5-A・付録5-Bに挙げた複合辞を含む。

- 【例】 | 地域活動への || 参加、 | 地産地消といった || 小さな | 経済で || 充足感を ||  
| 得る | 社会と || なります。 | | | | | | | | |
| ネットワークが || 形成さ=れ=にくい || 状況が || 生じており、 |  
| その | 目的が || 個人に || 絞られ || 過ぎている || 傾向が || ある |

複合辞の中に副助詞など（言いよどみの助詞・助動詞も含む。）が挿入された場合も、文節認定の上では全体でひとつの複合辞と見なす。

- 【例】 | お友達には | からかわれて=ばかり=いる | 三枚目でもありました。 |

(2) - 1 次に挙げる付属語の後ろでは切らない。

①付録 5-C に挙げた連語、1 短単位として認定された「-が～」 「-の～」の中に現れる  
付属語

**【例】**

[連 語] | サイドの | ベルトが | お気に=入りの | ブーツは |  
[-が～] | そこが | 万が=- | 倒産すると |  
[-の～] | 皮を | よく | 亀の=子だわしで | こすって | 洗い |

②分割すると意味が不自然になるものの中に現れる付属語

**【例】** | しかたが=ない | | しょうが=ない |

**【主語・主題に関する規定】**

(3) 次に示すような付属語を伴わない主語・主題の後ろでは、文節を切らない。

**【例】** | 緑=あふれる | 風景の | 中に、 |  
| 心=洗われる | ような | ステージに |  
| 気持ち=悪いから、 |

**【敬語形式に関する規定】**

(4) 「お(ご)～する・できる・くださる・いただく・なさる・いたす・ねがう・もうし  
あげる・あそばす・になる」という形式の敬語表現は、全体を一続きとする。

**【例】** | ご理解と | ご協力の | ほど | よろしく | お=願い=申し上げます。 |  
| いかが | お過ごしでしたか、 | お=聞か=せ=ください。 |  
| 法廷にも | 全身ピンクづくめで | お=出まし=になる。 |

**【数を表す要素に関する規定】**

(5) 数を表す要素とその直前直後の要素とは切り離さない。

**【例】** | 昭和十三年=八月=八日の | 荒木文部大臣の | 発言や |  
平均値=三.〇六と	いうような	値に	なって	
日米韓=三国の	対応			
パチスロの	場合だったら	一箱=三万ぐらいなんですけど		
十年以上=前までは	(F ま)	規則合成って	いう	方式が
三十=～=五十代の	主婦を	対象に	行った	アンケートで、
ただし、直前の要素が数量の程度を表す場合は除く。

**【例】** | およそ || 十カ所で | 検問を | 受け、 | 旅券を | 確かめられた。 |  
| 笑うと | 同じ | 事を | 最低 || 3回は | 言います。 |

## 5.2.2 長単位認定規程

長単位は、長単位認定規程の各規定に基づいて文節を分割する（または分割しない）ことによって得られた要素を 1 単位とする言語単位である。以下、長単位認定規程のうち、主要な規定と例を挙げる。

### 【句読点・空白・改行に関する規定】

句読点・空白・改行に関する規定は、他の全ての規定に優先して適用される。

- (1) 句読点（句読点として用いられているカンマ・ピリオド・エクスクラメーションマーク・クエスチョンマーク、三点リーダー、並びにコロンを含む。）および空白は 1 長単位とする。

【例】 | 機動的 | に | 商業施設 | として | 活用する | 例 | など | も | ある || 。 ||  
| 米 | は | 湾岸戦争後 || 、 || 英 || 、 || 仏 | など | と | とも | に |  
| 十五歳少女 | が | 最年少記録 | 「 | エベレスト登頂 | 」 | 三浦さん最高齢  
記録 | その | 日 | に || ... ||  
| 2 || || 協力的自主国防推進 || : || 自主国防 | と | 米韓同盟 | が |  
| || それ | は | 、 | 現実 | の | 世界情勢 | が |

### 【記号に関する規定】

- (2) 記号は 1 長単位とする。

【例】 | 「 || = || 羨ましい | な || 」 ||  
| 与野党逆転 || ⇨ || 海部政権誕生 | と | の | 願望 |  
記号のうち中点については、原則として切り出さない。

【例】 | 平成 | 3 年度 | から | コンピュータ=・=ネットワーク | を | 利用し | 、 |  
豪商=・=山崎屋	の	与五郎	と	遊女=・=吾妻	と	の	恋	を
麦=・=大豆=・=飼料作物	の	生産振興	に	資する	水田	の		
D=・=N=・=A	(	A B C	=	深夜	3=・=二十	)		

- (2) - 1 語と同じ働きをする記号・記号連続およびそれらを含む結合体は、全体で 1 長単位とする。

【例】 | A || が || B || に | 特定 | の | 法律行為 | を | 指図し | た | 場合 |  
南青山	に	ある	敷地面積		2, 0 0 0=n<sup>2</sup>		の	土地	は	、
	P K O=地域訓練ワークショップ		の	開催	や					
一般会計	の		(= =) =内		は	0 3 年度当初予算				

### 【付属語に関する規定】

(3) 付属語（付録 5-A・付録 5-B に挙げた複合辞を含む。）は 1 長単位とする。

【例】 | 公害紛争処理法 || における || 公害紛争処理 || の || 手続 || は || , | 原則 || と  
して		紛争当事者		から		の		申請		によって		開始さ		れる		。					
その	目的		が		個人		に		絞ら		れ		過ぎ		ている		傾向		が		ある
「	やむ		を		得		ず		型	」		の		親同居未婚者							

ただし、それを 1 長単位とすると、動詞的・形容詞的・形状詞的接尾辞および用言・助動詞の終止形・連体形以外に続く名詞的接尾辞が切り出されることになる場合の付属語は切り出さない。

【例】 | ネットワーク | が | 形成さ=れ=にくい | 状況が | 生じ | ており | , |  
| どんな | 使わ=れ=方 | を | し | た | んだろう | 。 |

付属語を伴わない文節、および規定 (3) によって付属語を切り出した後に残った形式（おおよそ文節の自立語部分に相当する形式）に以下の規定を適用する。それによって得られた各形式を 1 長単位とする。

(4) 同格の関係にある体言連続、並列の関係にある体言連続は切り離さない。

#### 【例】

[同格] | 大江健三郎さん | の | 長男=光さん |  
| 機関誌=計量国語学-発行 | の | 年 | に |  
| 中国語日刊新聞=「=星島日報=」 |

[並列] | 公正=妥当 | な | 実務慣行 | を | 集約し | た | もの | という |  
| 東京 | の | 郊外 | の | 市=町=村 | と | 言う | か |

(5) 主語・主題の後ろで切る。

【例】 | 緑 || あふれる | 風景 | の | 中に、 |  
| 心 || 洗われる | よう | な | ステージ | に |  
| 気持ち || 悪い | から | 、 |

漢語形状詞の述部を持つ場合は、切らない。

【例】 | 持続=可能 | な | 発展 | の | ため |  
| センス=抜群 | の | クリエーター | だっ | た | んだ | な |

(6) 体言や副詞に形式的な意味の「する」「できる」「なさる」「いたす」が直接続く場合、体言・副詞と「する」「できる」「なさる」「いたす」とを切り離さない。

【例】 | まるで | 1つ | の | 光点 | が | 往復運動=し | ている | よう | に |  
| 私 | は | この | 予選 | を | 1位 | で | 通過=できる | と |

目 | を | きらきら=さ | せ | た |

「こう」「そう」「ああ」「どう」に「する」「できる」「なさる」「いたす」が直接続く場合は、切り離す。

【例】 | こう || し | た | サークル活動 | が | 盛ん | に | なる | 背景 |  
| 这样做 || し | ている | うち | に |

- (7) 体言+用言という形式のうち、『岩波国語辞典』第6版、『日本国語大辞典』第2版のいずれか一方で見出し語(連語としての見出し語は除く。)になっているものは、体言と用言とを切り離さない。

【例】 | 運転手 | は | さり気=なく | 答え | て | アクセル | を | ゆるめ | た |。 |

- (8) 数を表す要素を含む自立語は、以下の規定に基づき長単位を認定する。

- (8) - 1 数を表す要素は、単位の変わり目の後ろで切る。

【例】 | 平成 | 15年 || 9月 || 15日 || 午後 | 7時 || 33分 ||  
| 1m || 80cm |

- (8) - 2 数を表す要素の前で切る。

【例】 | 平成 || 15年 | 9月 | 15日 | 午後 || 7時 | 33分 |  
| 南青山 | に | ある | 敷地面積 || 2, 000㎡ | の | 土地 | は |、 |  
| おおむね || 十六歳以上 || 二十歳未満 | の | 者 | を | 収容 |

- (8) - 3 数を表す要素とそれに続く体言・接辞とは切り離さない。

【例】 | 残業時間 | が | 月 | 80時間=以上 | の | 者 | は |  
| 自家発電 | が | すぐ | に | 作動せ | ず |、 | 約1時間=停電 |。 |  
| 5年=連続=優勝 | を | 駒沢大 | が | 来年 |、 | 達成できる | か |

ただし、数を表す要素とそれに続く体言・接辞の連続体の後ろに、これを受ける形式的な意味の「する」「できる」「なさる」「いたす」がある場合は、数を表す要素の後ろで切り離す。

【例】 | 約1時間 || 停電した |。 |  
| 箱根駅伝 | で | 4年連続 || 優勝し | た | の | は | 5校目 |。 |

- (9) 付録5-Cに挙げた連語、1短単位として認定された「-が～」 「-の～」およびそれらを含む結合体は、全体で1長単位とする。

【例】  
[連語] | サイド | の | ベルト | が | お気に入り | の | ブーツ | は |  
[-が～] | そこ | が | 万が=一 | 倒産する | と |

〔一の～〕 | 皮 | を | よく | 亀の=子だわし | で | こすっ | て | 洗い |

以上の規定によって長単位を認定した例を次に示す。

| 平成 | 4年度 | に | 創設さ | れ | た | 定期借地権制度 | は |、 | 借地  
契約 | の | 更新 | が | なく |、 | 定め | られ | た | 契約期間 | で | 確定  
的 | に | 契約 | が | 終了する | 借地権制度 | である |。 | 貸し主 | ( |  
土地所有者 | ) | に | と | っ | て | は | 予定時期 | に | 土地 | の | 返還 | を |  
受ける | こと | が | 保証さ | れる | と | と | も | に |、 | 一定期間 | の | 地代  
収入 | が | 安定的 | に | 得ら | れ |、 | また |、 | 借り主 | に | と | っ | て | は |  
土地 | を | 取得する | より | も | 少ない | 負担 | で | 土地 | を | 利  
用できる | こと | から |、 | 双方 | に | と | っ | て | メリット | が | あり |、  
借地 | の | 供給拡大 | による | 土地 | の | 有効利用 | を | 促進する |  
もの | と | して | 期待さ | れ | ている |。 | 定期借地権 | に | は |、 | 一  
般定期借地権 |、 | 建物譲渡特約付借地権 |、 | 事業用借地権 | の | 3  
類型 | が | ある | ( | 図表 | 1 | - | 5 | - | 4 | ) |。 |

### 5.2.3 付加情報の概要

長単位認定規程によって認定された各単位に次に挙げる付加情報を付与する。

#### (1) 語彙素・語彙素読み

自立語の語彙素・語彙素読みは、同一語の活用変化・表記のゆれ（補助記号の有無を含む。）をグループ化するための情報である。

例えば、サ変動詞「構築する」の未然形「構築さ」、連用形「構築し」、終止形・連体形「構築する」には、いずれも同じ「コウチクスル【構築する】」という語彙素・語彙素読みが付与される。これによって、「構築さ」「構築し」「構築する」の各出現形が、ひとつの語（動詞「構築する」）の活用変化として扱われることになる。同様に、「打ち合わせ室」「打合せ室」という各出現形に対して「ウチアワセシツ【打ち合わせ室】」という同一の語彙素・語彙素読みが付与され、ひとつの語の表記のゆれとして扱われることになる。

出現形	語彙素読み	語彙素
構築さ	コウチクスル	構築する
構築し		
構築する		
打ち合わせ室	ウチアワセシツ	打ち合わせ室
打合せ室		

図 5-2: 長単位の語彙素・語彙素読みの例(1)

一方、表記以外のゆれ・音の転化・省略・融合等によって生じた異形態はグループ化しない。そのため、以下の図 5-3 に挙げた「あまり」と「あんまり」などについては、それぞれ異なる語彙素・語彙素読みが付与され、別語として扱われる。

出現形	語彙素読み	語彙素
あまり	アマリ	余り
余り		
あんまり	アンマリ	余り
ちょうふく 重複する	チョウフクスル	重複する
じゅうふく 重複する	ジュウフクスル	重複する
コンピューター	コンピューター	コンピューター
コンピュータ	コンピュータ	コンピュータ

図 5-3: 長単位の語彙素・語彙素読みの例(2)

付属語の語彙素は、同一語の活用変化・ゆれ・省略・融合等によって生じた異形態をグループ化するための情報である。



語彙素・語彙素読みは、原則としてコーパスに出現したすべての長単位に付与する。

出現形	語彙素読み	語彙素
からには	カラニハ	からには
からにゃ		
こととなる	コトナル	こととなる
ことと成る		

図 5-4: 長単位の語彙素・語彙素読みの例(3)

(2) 品詞等の情報

各単位に品詞を付与する。活用する語には、活用型・活用形を付与する。

長単位の品詞・活用型・活用形については、表 5-3 から表 5-5 に一覧した。

表 5-3: 長単位品詞一覧

名詞-普通名詞-一般	接続詞	接尾辞-形状詞的
名詞-固有名詞-一般	感動詞-一般	接尾辞-動詞的
名詞-固有名詞-人名-一般	感動詞-フィラー	接尾辞-形容詞的
名詞-固有名詞-人名-姓	動詞-一般	記号-一般
名詞-固有名詞-人名-名	形容詞-一般	記号-文字
名詞-固有名詞-地名-一般	助動詞	補助記号-一般
名詞-固有名詞-地名-国	助詞-格助詞	補助記号-句点
名詞-数詞	助詞-副助詞	補助記号-読点
名詞-助動詞語幹	助詞-係助詞	補助記号-括弧開
代名詞	助詞-接続助詞	補助記号-括弧閉
形状詞-一般	助詞-終助詞	補助記号-△△-一般
形状詞-タリ	助詞-準体助詞	補助記号-△△-顔文字
形状詞-助動詞語幹	接頭辞	空白
連体詞	接尾辞-名詞的-一般	
副詞	接尾辞-名詞的-助数詞	

表 5-4: 長単位活用型一覧

五段-〇行	助動詞-ラシイ	文語助動詞-ザマス
上一段-〇行	助動詞-レル	文語助動詞-ザンス
下一段-〇行	無変化型	文語助動詞-ジ
カ行変格	文語四段-〇行	文語助動詞-ズ
サ行変格	文語上一段-〇行	文語助動詞-タリ-完了
形容詞	文語上二段-〇行	文語助動詞-タリ-断定
助動詞-ジャ	文語下一段-〇行	文語助動詞-ツ
助動詞-タ	文語下二段-〇行	文語助動詞-ナリ-断定
助動詞-ダ	文語カ行変格	文語助動詞-ナリ-伝聞
助動詞-タイ	文語サ行変格	文語助動詞-ヌ
助動詞-デス	文語ナ行変格	文語助動詞-ベシ
助動詞-ドス	文語ラ行変格	文語助動詞-マシ
助動詞-ナイ	文語形容詞-ク	文語助動詞-マジ
助動詞-ナンダ	文語形容詞-シク	文語助動詞-ム
助動詞-ヌ	文語助動詞-キ	文語助動詞-ムズ
助動詞-ヘン	文語助動詞-ケム	文語助動詞-メリ
助動詞-マイ	文語助動詞-ケリ	文語助動詞-ラシ
助動詞-マス	文語助動詞-コス	文語助動詞-ラム
助動詞-ヤ	文語助動詞-ゴトシ	文語助動詞-リ
助動詞-ヤス		文語助動詞-ンス

表 5-5: 長単位活用形一覧

語幹-一般	連用形-融合	連体形-一般
語幹-サ	連用形-省略	連体形-〇音便
未然形-一般	連用形-ト	連体形-省略
未然形-サ	連用形-ニ	連体形-補助
未然形-セ	連用形-長音	仮定形-一般
未然形-撥音便	連用形-補助	仮定形-融合
未然形-補助	終止形-一般	已然形-一般
意志推量形	終止形-〇音便	已然形-補助
連用形-一般	終止形-融合	命令形
連用形-〇音便	終止形-補助	ク語法

### 5.3 短単位

短単位は、言語の形態的側面に着目して規定した言語単位である。短単位の認定に当たっては、まず現代語において意味を持つ最小の単位（最小単位）を規定する。その上で、最小単位を長単位の範囲内で短単位認定規程に基づいて結合させる（又は結合させない）ことにより、短単位を認定する。そのため、短単位の認定規程は、最小単位と短単位の二つの認定規程から成る。

以下、本節では、最小単位認定規程・短単位認定規程、および短単位に付与する付加情報について、その概要を述べる。

#### 5.3.1 最小単位認定規程

最小単位は、現代語において意味を持つ最小の言語単位のことである。

最小単位は、和語・漢語・外来語・記号・数・人名・地名の種類ごとに、以下の規定によって認定する。

なお、以下に述べる最小単位は、短単位を認定するために必要な概念として規定するものであり、BCCWJのデータに最小単位境界を示すことはしない。

(1) 和語

和語の最小単位は、以下のように認定する。

【例】 /母/親/ /青/白/ /いい/加/減/な/  
/本/箱/ /幾/人/ /オレンジ/色/  
/わたし/で/も/できる/ /読み/終わり/まし/た/

(2) 漢語

漢語（和製漢語を含む。）は、漢字1字で表されるものを1最小単位とする。

【例】 /白/紙/ /安/価/ /含/有/量/ /数/百/

(3) 外来語

外来語・外国語は原語で1単語になるものを1最小単位とする。

英語起源の外来語の最小単位の認定は『リーダーズ英和辞典』第2版（研究社）による。それ以外の言語を起源とする外来語については適宜判断する。

【例】 /カラー/コピー/ /レーザー/プリンター/  
/オレンジ/色/ /ビタミン/剤/

(4) 記号

記号は1文字に当たるものを1最小単位とする。

【例】 /表/A/ /図/B/ /U/ターン/ /V/リーグ/  
/甲/類/ /乙/種/

(4) -1 ローマ字を並べた略語は全体で1最小単位とする。ローマ字の間の中点・ピリオド等は1最小単位としない。

【例】 /OHP/ /OS/ /D・N・A/ /Ph. D. /

(5) 数

数字は1字に当たるものを1最小単位とする。

【例】 /一/億/語/ /七/百/五/十/万/語/  
/2/万/5/千/分/の/1/  
/0/4/2/-/5/4/0/-/4/3/0/0/

(6) 人名

人名は姓を1最小単位、名を1最小単位とする。

【例】 /星野/仙一/ /マツト/・/マートン/ /林/威助/

通称・雅号・しこ名（その略称も含む。）等は、次のように最小単位を認定する。

【例】 /琴奨菊/ /十返舎/一九/ /古今亭/志ん生/

(7) 地名

地名は、次の規定により最小単位を認定する。

- (7) - 1 行政区画を表す地名は「都・府・県・郡・市・区・町・村・字」を除いた部分をそれぞれ1最小単位とする。類概念を表す部分には、他の最小単位の認定規定を適用する。

【例】 /東京/都/北/区/西が丘/三/丁/目/九/番/十/四/号/

- (7) - 1 - 1 「北海道」は全体で1最小単位とする。

【例】 /北海道/夕張/郡/長沼/町/ /明日/の/北海道/の/天気/

- (7) - 1 - 2 市区内の小区分の「～町」は「～町」を含めて1最小単位とする。

【例】 /大阪/府/豊中/市/待兼山町/ /千代田/区/大手町/

- (7) - 2 外国の国名や行政区画名なども、日本のそれと同じに扱う。

【例】 /アメリカ/合/衆/国/ /南アフリカ/共/和/国/  
/中華/人/民/共/和/国/  
/カリフォルニア/州/ /広東/省/  
/メキシコ/シティー/ /ミズーリ/ステート/

以上の規定によって認定された最小単位を、短単位認定のために表 5-6 のように分類する。

表 5-6: 最小単位の分類

分類	例
一般	和語：山川 白い 話す 言葉 ...
	漢語：社会用 研究所 ...
	外来語：オレンジ ボックス アルゴリズム ...
付属要素	接頭的要素（付録5-D：接頭的要素に掲げたもの） ：相 <sup>お</sup> 御 各 <sup>ご</sup> 御 ...
	接尾的要素（付録5-E：接尾的要素に掲げたもの） ：致す っぽい 性的 ...
記号	A B ω イ ロ ア 甲 乙 丙 NHK JR ...
数	一 二 十 百 千 ... 幾 数 何
固有名	人名：星野 仙一 ジェフ ウィリアムス 橋 龍 ...
	地名：大阪 待兼山町 六甲 天六 ...
助詞・助動詞	た です ます か から て も ...

### 5.3.2 短単位認定規程

短単位は、長単位の中で最小単位が以下の規定に基づいて結合した（又は結合しない（これは0回結合と考える。））結合体である。

短単位の認定に関する規定は、表 5-6 に示した種類ごとに適用すべき規定が定められている。以下、それを示す。

#### (1) 一般

原則として、「一般」に分類した和語・漢語の最小単位二つの一次結合を 1 短単位とする。

【例】 | 母=親 | | 書き=言葉 | | 食べ=歩く | | 音=声 |  
| 無=口 |

「一般」に分類した外来語の最小単位のうち省略されたものは、和語・漢語の最小単位と同様に扱う。

【例】 | パソ=コン | | オートマ=車 | | 塩=ビ |

(1) - 1 以下のものは、3 最小単位以上の結合であっても全体で 1 短単位とする。

①三つ以上の最小単位から成る組織の名称等の略称

【例】 | 統=数=研 | | 奈=文=研 | | 日=経=連 |

②切る位置が明確でないもの、あるいは切った場合とひとまとめにした場合とで意味にずれがあるもの

【例】 | 大統領 | | 不可解 | | 明後日 | | 殺風景 |  
輸出入		国内外		町村長		原水爆		市町村長
大袈裟		大雑把		大丈夫		一辺倒		
十文字		二枚目		十八番				

③「-が～」 「-の～」 の体言句

【例】  
「-の～」 : | 日=の=丸 | | 床=の=間 | | 竹=の=子 |  
「-が～」 : | 天=が=下 | | 雁=が=音 | | 剣=が=峰 |

(1) - 2 以下に挙げるものは、1 最小単位を 1 短単位とする。

①外来語・外国語の最小単位

【例】 | オレンジ | 色 | | アウト | オブ | ドメイン |  
ただし、省略された外来語の最小単位との 1 次結合体は 1 短単位とする。  
【例】 | エア=コン | | マス=コミ | | デフレ=スパイラル |

②最小単位が三つ以上並列した場合の、それぞれの最小単位

【例】 | 衣 || 食 || 住 | | 松 || 竹 || 梅 | | 都 || 道 || 府 || 県 |

③名を表す部分と類概念を表す部分とが結合してできた固有名のうち、名を表す部分・類概念を表す部分が共に 1 最小単位である場合の、それぞれの最小単位

【例】 | さくら || 屋 | | のぞみ || 号 | | くない || 会 |  
ただし、名を表す部分が 1 字の漢語である場合は、その 1 次結合体を 1 短単位とする。  
【例】 | 阪=大 | | 仏=教 | | 李=朝 | | 壮=族 | | 礼=記 |

④感動詞

【例】 | はい | はい | | おい | おい | | どれ | どれ |

⑤規定 (1)、(1) - 1、(1) - 2 の①から③によって得られた短単位に、前または後ろから結合した最小単位

【例】 | 内閣 || 府 || | 副 || 大統領 | | 橋本 || 元 || 首相 |  
| 光 | ファイバー || 網 || | 自衛 || 隊 || | 国立 | 国語 | 研究 || 所 ||

⑥単独で文節を構成する最小単位

【例】 | やっぱり | これ | も | 一 | つ | の | | オレンジ | を | 食べる | 。 |  
| えーと | 、 | こちら | の | 場合 | でし | たら | ... | ... |

(2) 記号

記号は、1 最小単位を 1 短単位とする。

【例】 | 表 | A | | 図 | B | | J R | | N T T | | L . A . |

(3) 数

数は、以下の規定によって単位認定する。

(3) - 1 数は、ほかの最小単位と結合させない。

【例】 | 四 || 月 | の || 三十 || 日 | ぐらい |  
| 私 | が || 一 二 || 年 | 前 | まで | 住 | ん | で | い | た |

(3) - 2 数の間どうしの結合については、一・十・百・千の桁ごとに 1 短単位とする。

「万」「億」「兆」などの最小単位は、それだけで 1 短単位とする。小数部分は、1 最小単位を 1 短単位とする。

【例】 | 千 || 九 = 百 || 四 = 十 || 二 | 年 | 十 | 月 | 二 = 十 || 五 | 日 | 、 |  
毎年	何 = 十		億		円	も	の	都民	の	税金	を
都心	から	一	時間	半	どころ	か	、	三 = 、 = 四十	分	、	
地形	図	2	万		5 = 千	分	の	1			
0	4	2	-	5	4	0	-	4	3	0	0

(4) 固有名

固有名（人名・地名）は、1 最小単位を 1 短単位とする。

【例】

〔人 名〕 | 星野 | 仙一 | | マット | ・ | マートン | | 林 | 威助 | |  
| 琴奨菊 | | 十返舎 | 一 九 | | お千代 | |

〔国 名〕 | アメリカ | 合衆 | 国 | | ロシア | 共和 | 国 | |  
| 南アフリカ | 共和 | 国 | |

〔行政区画名〕 | 東京 | 都 | 立川 | 市 | 緑町 | 十 | 番 | 二 | 号 | |  
| 京都 | 市 | 上京 | 区 | 今出川 | 通 | 烏丸 | 東入る | |

〔地域名〕 | 九州 | 地方 | | 四国 | 地方 | | 北海道 | 地方 | |  
| 東海道 | | 山陰道 | |  
| 東 | ヨーロッパ | | 南 | アメリカ | |

〔地形名〕 | 生駒 | 山 | | 昭和 | 新山 | | サロマ | 湖 |  
 〔場所名〕 | 茨木 | 市 | 駅 | | さいたま | 新 | 都心 | 駅 |  
 | 山陽 | 本線 | | 大 | 江戸 | 線 |  
 | 東海道 | | 中山道 |  
 〔略 称〕 | ちとから | | 天六 |

(5) 付属要素

付属要素は、1 最小単位を 1 短単位とする。

【例】 | お || 母 || さん | | 見 || にくい |

(6) 助詞・助動詞

助詞・助動詞は、1 最小単位を 1 短単位とする。

【例】 | 統一 | 的 || な || 視点 || で || 切り || ましょう ||  
 | それ | に | つい | て | も | っとも | 示 | 唆 | に | 富 | む | の | は |

(6) - 1 1 短単位として認定された「-が～」 「-の～」の中の助詞「が」「の」は、助詞・助動詞として扱わない。

【例】

「-の～」 : | 日 = の = 丸 | | 床 = の = 間 | | 竹 = の = 子 |  
 「-が～」 : | 君 = が = 代 | | 万 = が = 一 |

以上が短単位認定規程における主要な規定である。その他、短単位の認定に当たって注意すべき事項について規定を示す。

(7) 可能動詞

可能動詞は、元になった五段活用動詞と同様に短単位を認定する。

【例】 | 読める | | 行ける | | 切り離せる | | 話し合える |

(7) - 1 ら抜き言葉は語末の「れる」を切り出さない。

【例】 | 着 = れる | | 来 = れる | | 食 = れる |  
 | 見 = れる | | 透かし見 = れる | | こじ開け = れる |

(8) 動詞「- (サ) ス」「- (サ) セル」

(8) - 1 「- (サ) ス」という形の動詞は、語末「ス」「サス」を助動詞としない。

【例】 | 言 = わ = す | | 書 = か = す | | 食 = べ = さす | | 受 = け = さす |



- (8) - 2 五段・サ変動詞の未然形+助動詞「セル」、五段・サ変以外の動詞の未然形+助動詞「サセル」に分析可能なものは、語末「セル」「サセル」を助動詞とする。

【例】 | 書か||せる | | 食べ||させる |

ただし、以下に挙げるものは、語末の「(サ)セル」を分割しない。

- ①五段・サ変動詞の未然形+助動詞「セル」、五段・サ変以外の動詞の未然形+助動詞「サセル」と分析できないもの。

【例】 | 見=せる | | 着=せる | | 乗=せる | | 寄=せる |

- ②元の動詞が文語動詞であるもの、口語動詞であっても、現代語ではほとんど使われないもの。

【例】 | くゆら=せる | | 遅ら=せる | | そばだた=せる |

- ③「(サ)セル」という形の複合動詞(連用形が名詞化したものも含む。)

【例】 | 言い聞か=せる | | 言い聞か=せ|続ける | | 読み聞か=せ|

- ④「(サ)セル」という形の動詞(複合動詞は除く。)が複合語を構成している場合。

【例】 | 食わ=せ=物 | | 人騒が=せ | | 人泣か=せ | | 番狂わ=せ |  
| 役者 | 泣か=せ |

以上の規定によって短単位を認定した例を次に示す。

|平成|4|年度|に|創設|さ|れ|た|定期|借地|権|制度|は|、|借地|契約|の|更新|が|なく|、|定め|られ|た|契約|期間|で|確定|的|に|契約|が|終了|する|借地|権|制度|で|ある|。|貸し主|( |土地|所有|者|) |に|とっ|て|は|予定|時期|に|土地|の|返還|を|受ける|こと|が|保証|さ|れる|と|とも|に|、|一定|期間|の|地代|収入|が|安定|的|に|得ら|れ|、|また|、|借り主|に|とっ|て|は|土地|を|取得|する|より|も|少ない|負担|で|土地|を|利用|できる|こと|から|、|双方|に|とっ|て|メリット|が|あり|、|借地|の|供給|拡大|に|よる|土地|の|有効|利用|を|促進|する|もの|と|し|て|期待|さ|れ|て|いる|。|定期|借地|権|に|は|、|一般|定期|借地|権|、|建物|譲渡|特約|付|借地|権|、|事業|用|借地|権|の|3|類型|が|ある|( |図表|1| - |5| - |4| ) |。|

### 5.3.3 付加情報の概要

短単位認定規程によって認定された各単位に、次に挙げる付加情報を付与する。

#### (1) 語彙素・語彙素読み

語彙素・語彙素読みは、同一語の活用変化・音の転化・ゆれ・省略・融合等によって生じた異形態や送り仮名の違い等の異表記をグループ化するための情報である。

例えば、動詞「取る」の未然形「取ら」、連用形「取り」、終止形・連体形「取る」には、いずれも同じ「トル【取る】」という語彙素・語彙素読みが付与される。これによって、「取ら」「取り」「取る」の各出現形がひとつの語（動詞「取る」）の活用変化として扱われることになる。同様に「打ち合わせ」「打合せ」という各出現形に対して「ウチアワセ【打ち合わせ】」という同一の語彙素・語彙素読みが付与され、ひとつの語の表記のゆれとして扱われることになる。

出現形	語彙素読み	語彙素
取ら	トル	取る
取り		
取る		
打ち合わせ	ウチアワセ	打ち合わせ
打合せ		

図 5-5: 短単位の語彙素・語彙素読みの例(1)

長単位の語彙素・語彙素読みでは、省略・融合等によって生じた異形態はグループ化しなかったが、短単位の語彙素・語彙素読みでは、以下の図 5-6 のように、それぞれ同じ語彙素・語彙素読みが付与され、同語として扱われる。

出現形	語彙素読み	語彙素
あまり	アマリ	余り
余り		
あんまり		
ちょうふく 重複	チョウフク	重複
じゅうふく 重複		
コンピューター	コンピューター	コンピューター
コンピュータ		

図 5-6: 短単位の語彙素・語彙素読みの例(2)

原則として、語彙素・語彙素読みは、コーパスに出現した全ての短単位に付与する。

(2) 品詞情報

各単位に品詞を付与する。活用する語には、活用型・活用形を付与する。

短単位の品詞・活用型・活用形については、表 5-7 から表 5-9 に一覧した。

表 5-7: 短単位品詞一覧

名詞-普通名詞-一般	連体詞	接尾辞-名詞的-サ変可能
名詞-普通名詞-サ変可能	副詞	接尾辞-名詞的-形状詞可能
名詞-普通名詞-形状詞可能	接続詞	接尾辞-名詞的-サ変形状詞可能
名詞-普通名詞-サ変形状詞可能	感動詞-一般	接尾辞-名詞的-副詞可能
名詞-普通名詞-副詞可能	感動詞-フィラー	接尾辞-名詞的-助数詞
名詞-普通名詞-助数詞可能	動詞-一般	接尾辞-形状詞的
名詞-固有名詞-一般	動詞-非自立可能	接尾辞-動詞的
名詞-固有名詞-人名-一般	形容詞-一般	接尾辞-形容詞的
名詞-固有名詞-人名-姓	形容詞-非自立可能	記号-一般
名詞-固有名詞-人名-名	助動詞	記号-文字
名詞-固有名詞-地名-一般	助詞-格助詞	補助記号-一般
名詞-固有名詞-地名-国	助詞-副助詞	補助記号-句点
名詞-数詞	助詞-係助詞	補助記号-読点
名詞-助動詞語幹	助詞-接続助詞	補助記号-括弧開
代名詞	助詞-終助詞	補助記号-括弧閉
形状詞-一般	助詞-準体助詞	補助記号-AA-一般
形状詞-タリ	接頭辞	補助記号-AA-顔文字
形状詞-助動詞語幹	接尾辞-名詞的-一般	空白

表 5-8: 短単位活用型一覧

五段-〇行	形容詞-一イ	文語ラ行変格
五段-カ行-一般	助動詞-ジャ	文語形容詞-ク-一般
五段-カ行-イク	助動詞-タ	文語形容詞-ク-多シ
五段-カ行-ユク	助動詞-タイ	文語形容詞-シク-シク
五段-マ行-一般	助動詞-ダ	文語形容詞-シク-ジク
五段-マ行-済ム	助動詞-デス	文語助動詞-キ
五段-ラ行-一般	助動詞-ドス	文語助動詞-ケム
五段-ラ行-アル	助動詞-ナイ	文語助動詞-ケリ
五段-ラ行-サル	助動詞-ナンダ	文語助動詞-コス
五段-ワア行-一般	助動詞-ス	文語助動詞-ゴトシ
五段-ワア行-〇ウ	助動詞-ヘン	文語助動詞-ザマス
上一段-〇行	助動詞-マイ	文語助動詞-ザンス
上一段-ラ行-一般	助動詞-マス	文語助動詞-ジ
上一段-ラ行-レル	助動詞-ヤ	文語助動詞-ズ
下一段-〇行	助動詞-ヤス	文語助動詞-タリ-完了
下一段-ア行-一般	助動詞-ラシイ	文語助動詞-タリ-断定
下一段-ア行-得ル	助動詞-レル	文語助動詞-ツ
下一段-サ行-一般	文語四段-〇行	文語助動詞-ナリ-伝聞
下一段-サ行-セル	文語四段-ハ行-一般	文語助動詞-ナリ-断定
下一段-ラ行-一般	文語四段-ハ行-〇ウ	文語助動詞-ヌ
下一段-ラ行-レル	文語四段-ハ行-イウ	文語助動詞-バシ
下一段-ラ行-呉レル	文語上一段-〇行	文語助動詞-マシ
カ行変格	文語上二段-〇行	文語助動詞-マジ
サ行変格-スル	文語下一段-カ行	文語助動詞-ム
サ行変格-ズル	文語下二段-〇行	文語助動詞-ムズ
サ行変格-為ル	文語下二段-ハ行-一般	文語助動詞-メリ
形容詞-一般	文語下二段-ハ行-経	文語助動詞-ラシ
形容詞-無イ	文語カ行変格	文語助動詞-ラム
形容詞-良イ-イイ	文語サ行変格-ス	文語助動詞-リ
形容詞-良イ-ヨイ	文語サ行変格-ズ	文語助動詞-ンス
形容詞-〇イ	文語ナ行変格	無変化型

表 5-9: 短単位活用形一覧

語幹-一般	連用形-融合	連体形-一般
語幹-サ	連用形-省略	連体形-○音便
未然形-一般	連用形-ト	連体形-省略
未然形-サ	連用形-ニ	連体形-補助
未然形-セ	連用形-長音	仮定形-一般
未然形-撥音便	連用形-補助	仮定形-融合
未然形-補助	終止形-一般	已然形-一般
意志推量形	終止形-○音便	已然形-補助
連用形-一般	終止形-融合	命令形
連用形-○音便	終止形-補助	ク語法

### (3) 語種情報

語種とは、語をその出自によって分類したもののことである。原則として、コーパスに出現したすべての短単位に付与する。

BCCWJ で付与した語種は、次のとおりである。

#### ①和語〔和〕

日本固有の語

【例】 暖かい 言葉 話す

#### ②漢語〔漢〕

近世以前に中国から入った語

【例】 音楽 国語 報告

和製漢語も漢語とする。

【例】 大根 返事

#### ③外来語〔外〕

欧米系の諸言語から入った語

【例】 ゲーム コーパス データ

上記のほか、以下のものも外来語とする。

##### a. 和製英語

【例】 アフレコ ナイター

##### b. 梵語等を中国で音訳した語に由来する語

【例】 阿羅漢 盂蘭盆 卒塔婆

c. アイヌ語から入った語

【例】 昆布 鮭 ラッコ

d. 中国以外のアジア諸国語から入った語

【例】 キムチ パッチ

e. 近代以降に中国から入った語

【例】 クーニャン シュウマイ メンツ

#### ④混種語〔混〕

和語・漢語・外来語のうち異なる2種類以上の語種の語が二つ以上結合した語。漢語・外来語であったものの末尾が活用するようになった語

【例】 塩ビ トラブル 本箱 力む

#### ⑤固有名〔固〕

人名・地名・商品名等。品詞が固有名詞となる語

【例】 大阪 星野 仙一 ソニー

#### ⑥記号〔記号〕

句読点・括弧などの補助記号や、箇条書きの項目名として使われた一字の片仮名などの記号。固有名以外のローマ字略語

【例】 ア イ A B OHP

### (4) 用法

用法とは、「名詞・普通名詞・形状詞可能」「名詞・普通名詞・副詞可能」「名詞・普通名詞・サ変形状詞可能」の各語が、実際に当該文脈で名詞・形状詞・副詞のどの品詞で用いられているのか、また「名詞・普通名詞・助数詞可能」の語が名詞・助数詞のどちらの品詞で用いられているのかを示す情報である。

BCCWJで付与した用法は、次のとおりである。

#### ①名詞

「名詞・普通名詞・形状詞可能」「名詞・普通名詞・サ変形状詞可能」「名詞・普通名詞・副詞可能」の語が当該文脈で名詞として使われている場合に付与。

【例】 寛容, 対話, 協力を重んじる異文化間交流  
ネットワーク担当の技術者が不足している

必要な場合には

## ②形状詞

「名詞-普通名詞-形状詞可能」「名詞-普通名詞-サ変形状詞可能」の語が当該文脈で形状詞として使われている場合に付与。

**【例】** それらに必要な施設の整備  
どの業種にも共通であるが

## ③副詞

「名詞-普通名詞-副詞可能」の語が当該文脈で副詞として使われている場合に付与。

**【例】** 笑福亭鶴笑氏が自ら考案した落語形式で  
一時騒然とした雰囲気にも包まれた

## ④助数詞

「名詞-普通名詞-助数詞可能」の語が当該文脈で助数詞として使われている場合に付与。

**【例】** その約6割を落札している  
前年と比べて1.8ポイント上昇している。

## 5.4 CSJからの変更点

5.1節で述べたように、BCCWJでは言語単位としてCSJと同じ長単位・短単位を採用した。しかし、長単位・短単位の認定規程は、CSJの規程をそのまま用いるのではなく、修正等を行っている。また付加情報についても、CSJとは異なるものとなっている。

そこで本節では、長単位・短単位および付加情報で、CSJから変更した箇所のうち、主な箇所について述べることにする。

### (1) 文節・長単位

文節・長単位の両方に関わる変更点としては、同格・並列の扱いがある。CSJでは学校文法と同様に、同格の関係にある要素、並列の関係にある要素を切り離していたが、BCCWJでは切り離さないこととした。

#### **【例】**

〔同格〕 | 大江健三郎さんの | 長男=光さん |

〔並列〕 | 公正=妥当な | 実務慣行を | 集約した | ものという | 意味で |

長単位認定規程の主な変更点としては、次の2点が挙げられる。1点目は、数量に関する規定である。CSJでは、以下に示すように、数量を表す要素は分割せず一続きとしていたが、長すぎるという指摘があった。

**【例】** | 1 m = 8 0 c m |

そこで、BCCWJでは以下のように、単位の変わり目の後ろで分割することとした。

【例】 | 1 m || 8 0 c m |

2点目は、係り受けが関係する規定の簡素化である。CSJでは「体言連続の一部分が連体修飾語を受けている場合、その後ろで切る」「2文節を受ける、若しくは2文節以上に係る接辞はその前後で切る」という規定があった。以下に例を示す。

【例】 | 項構造 | の | 曖昧性 || 解消 |  
| 円形劇場 | とか | 水路 || 等 |

これらは、語と語との係り受けを厳密に考えようとしたところから作られたものである。しかし実際に単位分割をする際には、体言連続の一部分が連体修飾語を受けているかどうかの判定が難しいものがある。そのため、特に判定が難しい「体言+以降、間(かん)、ごと、自体、達」という形式は、

【例】 | 住ん | での | 人=達 |

のように、体言と「達」などを切り離さないという例外規定を設ける等、煩雑な規定となっていた。このことが単位認定のゆれにつながっていたため、BCCWJでは規定を簡素化することとした。具体的には、体言連続の一部分が連体修飾語を受けていても、体言連続を分割することなく、以下のように一続きとした。

【例】 | 項構造 | の | 曖昧性=解消 |  
| 円形劇場 | とか | 水路=等 |

## (2) 最小単位・短単位

CSJの短単位や現代雑誌九十種調査のβ単位では、「一般」の外来語の最小単位も、和語・漢語と同様、二つの一次結合を1短単位としていた。例えば、「コールセンター」「オレンジ色」は共に1単位としていた。ただし、以下のような例外規定を設けた。

① 欧米語の冠詞・前置詞に当たるものは1最小単位を1短単位とする。

② β単位では最小単位二つの一次結合が7拍を超える場合、短単位では同じく10拍を超える場合、結合させずに1最小単位を1短単位

外来語の最小単位二つの一次結合を1短単位とすることについては、CSJの構築当初から和語・漢語に比べて長すぎるのではないかという指摘があった。このような指摘を踏まえ、上記②の拍数による例外規定を設けたが、10拍を超える場合としたことに言語学的な意味があるわけではなく、そういう意味でこの例外規定にも問題があった。

そこで、BCCWJでは「一般」の外来語の最小単位は、原則として1最小単位を1短単位とし、和語・漢語の最小単位とは異なる扱いにした。

【例】 | コール | センター | | オレンジ | 色 |

## (3) 付加情報

長単位・短単位とも品詞情報については、CSJの品詞から大幅な改定を行った。普通名詞を例にして、CSJの品詞体系とBCCWJの品詞体系とを比較すると、表5-10のとおりである。

表 5-10: CSJ と BCCWJ との品詞の比較 (名詞)

CSJ	BCCWJ (長単位)	BCCWJ (短単位)
名詞	名詞-普通名詞-一般	名詞-普通名詞-一般 名詞-普通名詞-サ変可能 名詞-普通名詞-形状詞可能 名詞-普通名詞-サ変形状詞可能 名詞-普通名詞-副詞可能 名詞-普通名詞-助数詞可能 名詞-助動詞語幹
名詞-固有名詞	名詞-固有名詞-一般 名詞-固有名詞-人名-一般 名詞-固有名詞-人名-姓 名詞-固有名詞-人名-名 名詞-固有名詞-地名-一般 名詞-固有名詞-地名-国	名詞-固有名詞-一般 名詞-固有名詞-人名-一般 名詞-固有名詞-人名-姓 名詞-固有名詞-人名-名 名詞-固有名詞-地名-一般 名詞-固有名詞-地名-国
名詞-数詞	名詞-数詞	名詞-数詞

表 5-10 に示したとおり、BCCWJ では固有名詞を細分化するとともに、短単位において「名詞-普通名詞-サ変可能」「名詞-普通名詞-形状詞可能」「名詞-普通名詞-副詞可能」「名詞-普通名詞-助数詞可能」のように普通名詞を細分化した上で、「○○可能」という曖昧性を持たせた品詞を設けた。

BCCWJ の細分化した品詞体系は、形態素解析用辞書 UniDic の品詞体系に準拠したものである。BCCWJ は、1 億語から成る大規模なコーパスであるため、形態論情報の付与は自動解析システムにより行った。短単位解析には解析エンジン MeCab と解析用辞書 UniDic を、長単位解析には短単位解析結果から長単位を自動構成する解析器を用いた。また、1 億語のうち約 100 万語 (コアデータ) については、自動解析後に人手修正を行い、解析精度 99% 以上の高精度なデータとし、形態素解析システムの学習用データとして用いた。

このような自動解析システムの利用等の観点から、UniDic に準拠する形で CSJ から品詞体系を大幅に改定したのである。なお、品詞体系の改定に当たっては、UniDic の品詞体系をそのまま採用するのではなく、UniDic への未登録語の新規追加作業、コーパス修正作業の中で UniDic についても一部に改定を加える形で BCCWJ と UniDic とに共通な品詞体系を実現した。

なお、曖昧性を持たせた品詞は、短単位において、普通名詞以外にも、動詞、形容詞、名詞的接尾辞において設定している (表 5-7 参照)。一方、長単位では、実際の文脈において名詞として使われているのか、形状詞として使われているのかなどを判断し、それに基づいて品詞を付与したので、「名詞-普通名詞-○○可能」等の曖昧性を持たせた品詞は設けていない。

品詞情報のほか、CSJ では付与されなかった語種に関する情報も BCCWJ では付与した。



## 5.5 終わりに

以上、本章では、BCCWJにおける長短2種類の言語単位の認定規程および付加情報について概略を述べた。

BCCWJでは、言語単位の設計に際し、①コーパスに基づく用例収集、各ジャンルの言語的特徴の解明に適した単位を設計する、②『日本語話し言葉コーパス』と互換性のある形態論情報を設計する、③国立国語研究所の語彙調査における知見を活用するという三つの方針を立てた。これら方針に沿って検討した結果、各ジャンルの言語的特徴の解明に適した長単位とコーパスに基づく用例収集短単位を採用した。

長単位・短単位はCSJで採用した言語単位であるが、既に述べたように認定規程については、書き言葉用に修正・拡張を行った。また、長単位・短単位の解析に自動解析システムを活用したことなどから、品詞情報については、CSJのものから大幅な改定を行った。

BCCWJの長単位・短単位解析は、目標としていた解析精度98%以上（コアデータについては99%以上）を達成しており、高精度な解析を実現できたと言える。

しかし、今後に残された課題もある。例えば、長単位の語彙素・語彙素読みの問題が挙げられる。語彙素・語彙素読みについては、表記や語形にかかわらず、同じ語であれば、同一の見出し（語彙素・語彙素読み）を付与するというのが基本的な方針であり、短単位については、そのとおりに設計されている。しかし長単位では、語形が異なる場合には、その語形に基づき異なる語彙素・語彙素読みを付与することとした。これにより、「あまり」と「あんまり」は、短単位では同一語と見なされるが、長単位では別語と見なされることとなった。

これは、短単位解析結果を基に長単位を自動構成する際に、「語形」「書字形」の情報を利用したことによる。短単位解析結果を基に長単位の語彙素・語彙素読みを自動構成するに当たっては、小椋他（2011）に述べるように種々の問題があったが、その中で現時点での最善の手法として「語形」「書字形」の情報を利用することとした。しかしながら、今後改善が必要な点である。

## 参考文献

- 小椋秀樹 (2006) 「第 3 章 形態論情報」『国立国語研究所報告 124 日本語話し言葉コーパスの構築法』,133-186.
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕 (2011a) 「『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)」国立国語研究所内部報告書 LR-CCG-10-05-01
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕 (2011b) 「『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 (下)」国立国語研究所内部報告書 LR-CCG-10-05-02.
- 国立国語研究所 (1987) 『国立国語研究所報告 89 雑誌用語の変遷』,秀英出版.
- 伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵 (2007) 「コーパス日本語学のための言語資源 — 形態素解析用電子化辞書の開発とその応用 —」『日本語科学』 22,101-122,国書刊行会.
- 中野洋 (1998) 「言語の統計」『岩波講座言語の科学 9 言語情報処理』,149-199,岩波書店.
- 林大監修 (1982) 『角川小辞典 9 図説日本語』,角川書店.
- 富士池優美・小西光・小椋秀樹・小木曾智信・小磯花絵 (2011) 「長単位に基づく媒体・カテゴリ間の品詞比率に関する分析」『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』,273-280.
- 前田富祺 (1985) 『国語語彙史研究』,明治書院.
- Uchimoto, K. and H. Isahara (2007). Morphological annotation of a large spontaneous speech corpus in Japanese, Proceedings of IJCAI, 1731-1737.

付録 5-A: 複合辞（助詞相当句）

語彙素読み	語彙素	品詞
カラシテ	からして	助詞-格助詞
カラスルト	からすると	助詞-格助詞
カラスレバ	からすれば	助詞-格助詞
サイニ	際に	助詞-格助詞
タメノ	ための	助詞-格助詞
トイウ	という	助詞-格助詞
トイッタ	といった	助詞-格助詞
トシテ	として	助詞-格助詞
ニアタッテ	にあたって	助詞-格助詞
ニアタリ	にあたり	助詞-格助詞
ニイタルマデ	に至るまで	助詞-格助詞
ニオイテ	において	助詞-格助詞
ニオケル	における	助詞-格助詞
ニカケテ	にかけて	助詞-格助詞
ニカンシテ	に関して	助詞-格助詞
ニカンスル	に関する	助詞-格助詞
ニサイシ	に際し	助詞-格助詞
ニサイシテ	に際して	助詞-格助詞
ニシテ	にして	助詞-格助詞
ニタイシ	に対し	助詞-格助詞
ニタイシテ	に対して	助詞-格助詞
ニタイスル	に対する	助詞-格助詞
ニツイテ	について	助詞-格助詞
ニツキ	につき	助詞-格助詞
ニトッテ	にとつて	助詞-格助詞
ニヨッテ	によつて	助詞-格助詞
ニヨリ	により	助詞-格助詞
ニヨル	による	助詞-格助詞
ニヨルト	によると	助詞-格助詞
ニヨレバ	によれば	助詞-格助詞
ニワタッテ	にわたつて	助詞-格助詞
ニワタリ	にわたり	助詞-格助詞
ニワタル	にわたる	助詞-格助詞
ヲツウジテ	を通じて	助詞-格助詞
ヲハジメ	をはじめ	助詞-格助詞
ヲメグル	をめぐる	助詞-格助詞
ヲモッテ	をもって	助詞-格助詞
ダケデナク	だけでなく	助詞-副助詞

語彙素読み	語彙素	品詞
ニカギラズ	に限らず	助詞-副助詞
ノミナラズ	のみならず	助詞-副助詞
トイエドモ	といえども	助詞-係助詞
トイッテモ	といつても	助詞-係助詞
トキタラ	ときたら	助詞-係助詞
ニイタッテハ	に至つては	助詞-係助詞
ニシタッテ	にしたつて	助詞-係助詞
ウエデ	上で	助詞-接続助詞
ウエニ	上に	助詞-接続助詞
ウエハ	上は	助詞-接続助詞
カトオモウト	かと思うと	助詞-接続助詞
カトオモッタラ	かと思ったら	助詞-接続助詞
ガハヤイカ	が早いか	助詞-接続助詞
カラトイッテ	からといつて	助詞-接続助詞
カラニハ	からには	助詞-接続助詞
タトコロ	たところ	助詞-接続助詞
タトコロデ	たところで	助詞-接続助詞
タメニ	ために	助詞-接続助詞
トシタラ	としたら	助詞-接続助詞
トシテモ	としても	助詞-接続助詞
トスレバ	とすれば	助詞-接続助詞
トドウジニ	と同時に	助詞-接続助詞
トトモニ	とともに	助詞-接続助詞
トハイエ	とはいへ	助詞-接続助詞
ニカカワラズ	に関わらず	助詞-接続助詞
ニシタガイ	にしたがい	助詞-接続助詞
ニシタガッテ	にしたがつて	助詞-接続助詞
ニシテハ	にしては	助詞-接続助詞
ニシテモ	にしても	助詞-接続助詞
ニシロ	にしろ	助詞-接続助詞
ニセヨ	にせよ	助詞-接続助詞
ニツレ	につれ	助詞-接続助詞
ニツレテ	につれて	助詞-接続助詞
ニモカカワラズ	にも関わらず	助詞-接続助詞
モノノ	ものの	助詞-接続助詞
ヤイナヤ	や否や	助詞-接続助詞
ワリニ	わりに	助詞-接続助詞

付録 5-B: 複合辞 (助動詞相当句)

語彙素読み	語彙素	品詞
カモシレナイ	かもしれない	助動詞
カモシレマセン	かもしれません	助動詞
コトガアル	ことがある	助動詞
コトガデキル	ことができる	助動詞
コトトナル	こととなる	助動詞
コトニスル	ことにする	助動詞
コトニナル	ことになる	助動詞
コトハナイ	ことはない	助動詞
コトモアル	こともある	助動詞
コトモナイ	こともない	助動詞
ザルヲエナイ	ざるを得ない	助動詞
シカナイ	しかない	助動詞
ズニハイラレナイ	ずにはいられない	助動詞
タライイ	たらいい	助動詞
ツツアル	つつある	助動詞
テアル	である	助動詞
デアル	である	助動詞
テイク	ていく	助動詞
テイタダク	ていただく	助動詞
テイル	ている	助動詞
テオク	ておく	助動詞
テオル	ておる	助動詞
テクダサル	てくださる	助動詞
テクル	てくる	助動詞
テクレル	てくれる	助動詞
テシカタガナイ	て仕方がない	助動詞
テシマウ	てしまう	助動詞
テシヨウガナイ	て仕様がな	助動詞

語彙素読み	語彙素	品詞
テタマラナイ	てたまらない	助動詞
デナイ	でない	助動詞
テナラナイ	てならない	助動詞
デハリマセン	ではありません	助動詞
テハイケナイ	てはいけない	助動詞
デハナイ	ではない	助動詞
テナナラナイ	てはならない	助動詞
テホシイ	てほしい	助動詞
テミル	てみる	助動詞
デモアル	でもある	助動詞
テモイイ	てもいい	助動詞
テモラウ	てもらう	助動詞
テヤル	てやる	助動詞
ナイデハイラレナイ	ないではいられない	助動詞
ナクテナナラナイ	なくてはならない	助動詞
ナケレバナラナイ	なければならない	助動詞
ニスギナイ	に過ぎない	助動詞
ニチガイナイ	に違いない	助動詞
ニホカナラナイ	にほかならない	助動詞
ネバナラナイ	ねばならない	助動詞
ノダ	のだ	助動詞
ノデアル	のである	助動詞
ノデス	のです	助動詞
ノデハナイ	のではない	助動詞
バイイ	ばいい	助動詞
マデモナイ	までもない	助動詞
ワケニハイカナイ	わけにはいかない	助動詞

付録 5-C: 連語

語彙素読み	語彙素	品詞
イッケンラクチャク	一件落着	名詞
トントンビョウシ	とんとん拍子	名詞
ニクマレグチ	憎まれ口	名詞
ヒトイチバイ	人一倍	名詞
ミタメ	見たい目	名詞
アマイモノギライ	甘い物嫌い	名詞・形状詞
オキニイリ	御気に入り	名詞・形状詞
クワズギライ	食わず嫌い	名詞・形状詞
タベズギライ	食べず嫌い	名詞・形状詞
マケズギライ	負けず嫌い	名詞・形状詞
イツモドオリ	何時も通り	名詞・副詞
イママデドオリ	今まで通り	名詞・副詞
カミヒトエ	紙一重	形状詞
コトバズクナ	言葉少な	形状詞
コトモナゲ	事も無気	形状詞
ワガモノガオ	我が物顔	形状詞
イイカゲン	良い加減	形状詞・副詞
シユタル	主たる	連体詞
アイカワラス	相変わらず	副詞
アオアオト	青々と	副詞
イカニモ	如何にも	副詞
イツカ	何時か	副詞
イマヤ	今や	副詞
カナラズシモ	必ずしも	副詞
クログロト	黒々と	副詞
コツゼント	忽然と	副詞
シキリト	頻りと	副詞
シキリニ	頻りに	副詞
シゼント	自然と	副詞
ジツハ	実は	副詞
スクナクトモ	少なくとも	副詞
ソコハカト	其処は彼と	副詞
ダンジテ	断じて	副詞
ドウシテ	どうして	副詞
ドウシテモ	どうしても	副詞
ドウニカ	どうにか	副詞
ドウニモ	どうにも	副詞
ドウニモコウニモ	どうにもこうにも	副詞
ナニヨリ	何より	副詞
ナンダカ	何だか	副詞
ナンデモ	何でも	副詞
ナント	何と	副詞
ナントカ	何とか	副詞

語彙素読み	語彙素	品詞
フカフカト	深々と	副詞
ベツニ	別に	副詞
ホウフツト	彷彿と	副詞
ホソボソト	細々と	副詞
モウゼント	猛然と	副詞
シカシナガラ	然しながら	接続詞
シタガッテ	従って	接続詞
スルト	すると	接続詞
ソウシテ	そうして	接続詞
ソレカラ	其れから	接続詞
ソレデ	其れで	接続詞
ソレデハ	其れでは	接続詞
ソレデモ	其れでも	接続詞
ソレトモ	其れとも	接続詞
ダガ	だが	接続詞
ダカラ	だから	接続詞
ダケレド	だけれど	接続詞
ダケレドモ	だけれども	接続詞
ダッタラ	だったら	接続詞
ダッテ	だって	接続詞
チナミニ	因みに	接続詞
ツテイウカ	って言うか	接続詞
デスガ	ですが	接続詞
デスカラ	ですから	接続詞
デスケレド	ですけれど	接続詞
デスケレドモ	ですけれども	接続詞
デハ	では	接続詞
デモ	でも	接続詞
トコロガ	所が	接続詞
トコロデ	所で	接続詞
ナノデ	なので	接続詞
ナノニ	なのに	接続詞
ナラビニ	並びに	接続詞
マタハ	又は	接続詞
ユエニ	故に	接続詞
ヨウスルニ	要するに	接続詞
ヨッテ	因って	接続詞
オメデトウ	御めでとう	感動詞
ゴメンナサイ	御免為さい	感動詞
スマン	済まん	感動詞
スママセン	済みません	感動詞
キモチワルガル	気持ち悪がる	動詞
カッコウヨイ	格好良い	形容詞

## 付録 5-D: 接頭的要素

語彙素読み	語彙素	品詞	注記
アイ	相	接頭辞	「相」と1最小単位との結合体が名詞である場合は除く。(相=乗り, 相=討ち)
オ	御	接頭辞	次に挙げるものは、後の部分と併せて1最小単位とする。[お足, おいた, お家(芸・流), お薄, おかか, お鏡, おかき, お陰, おかず, お河童, おかま, おかみ, おから, おかわ, お冠, 御形, おぐし, お好み(焼き), おこわ, お下げ(髪), お差し, おさつ, おざなり, おざぶ, おさん(どん), おしっこ, おしぼり, おしめ, おじや, おしゃぶり, お釈迦, お洒落, お節, お宅(代名詞), お多福, お陀仏, お玉, おつむ, お手(上げ・の物), おでき, おでまし, お転婆, お伽(話), お腹, お成り, お握り, お主, お寝しょ, お萩, おはこ(十八番の意), おはよう, お払い(箱), おひたし, お冷や, お袋, おふる, おまえ, おまけ, おませ, おまる, お巡り, お娘, おむすび, おむつ, お目見え, お漏らし, おやつ]
オン	御	接頭辞	次に挙げるものは、後の部分と併せて1最小単位とする。[御曹司, 御大, 御中, 御身]
カク	各	接頭辞	漢語の1最小単位と結合したものは除く。(各=国, 各=地)
コン	今	接頭辞	漢語の1最小単位と結合したものは除く。(今=回, 今=度)
ゴ	御	接頭辞	次に挙げるものは、後の部分と併せて1最小単位とする。[御形, 御供, 御所, 御新, 御仁, 御神火, 御前, 御饌, 御託, 御殿, 御伝, 御悩, 御飯, 御辺, 御免, 御覧, 御料, 御寮]
ショ	諸	接頭辞	漢語の1最小単位と結合したものは除く。(諸=国, 諸=所)
ゼン	全	接頭辞	漢語の1最小単位と結合したものは除く。(全=国, 全=社)
タイ	対	名詞-普通名詞-一般	漢語の1最小単位と結合したものは除く。(対=米, 対=人)
ホン	本	接頭辞	「この」の意。漢語の1最小単位と結合したものは除く。(本件)
ミ	御	接頭辞	次に挙げるものは後の部分と併せて1最小単位とする。[御生, 御門, 御溝, 御酒, 御籤, 御髪, 御座, 御食, 御子, 御輿, 御言, 御簾, 御衣, 御台, 御霊, 御堂, 御息所, 御幸, 御代]
ホノ	仄	接頭辞	「ほのか」「ほのめく」「ほのぼの」「ほのめかす」は除く。

## 付録 5-E: 接尾的要素

語彙素読み	語彙素	品詞	注記
アガリ	上がり	接尾辞-名詞的-一般	前にその職業・身分だった者の意。
アグネル	あぐねる	動詞-非自立可能	
アソバス	遊ばす	動詞-非自立可能	
アタウ	能う	動詞-非自立可能	動作・状態の継続・進行を表す。
アタリ	当たり	接尾辞-名詞的-副詞可能	
アテ	宛	接尾辞-名詞的-一般	名あての意。「名宛(人)」の「宛」は除く。
アテ	宛	接尾辞-名詞的-一般	「〜に対して」の意。
アリ	有り	名詞-普通名詞-一般	「大有り」「神在」「徒有り」「訳有り」の「有り(在り)」は除く。
アル	有る	動詞-非自立可能	動作・状態の継続・進行を表す。
イタス	致す	動詞-非自立可能	
イル	居る	動詞-非自立可能	
ウエ	上	接尾辞-名詞的-一般	
エル	得る	動詞-非自立可能	「〜することができる」の意。
オエル	終える	動詞-非自立可能	
オオセル	果せる	動詞-非自立可能	「すっかり終える」の意。
オクレル	遅れる	動詞-非自立可能	
オル	居る	動詞-非自立可能	動作・状態の継続・進行を表す。
オワス	御座す	動詞-非自立可能	動作・状態の継続・進行を表す。
オワル	終わる	動詞-非自立可能	
カ	化	接尾辞-名詞的-一般	漢語の1最小単位と結合したものは除く。(特=化, 液=化)
ガカル	がかる	接尾辞-動詞的	
カタ	方	接尾辞-名詞的-一般	「仕方」の「方」は除く。
ガタイ	難い	接尾辞-形容詞的	「有り難い」の「難い」は除く。
カタガタ	旁	接尾辞-名詞的-一般	
ガチ	勝ち	接尾辞-形状詞的	
ガテラ	がてら	接尾辞-名詞的-副詞可能	
カネル	兼ねる	接尾辞-動詞的	
ガル	がる	接尾辞-動詞的	助動詞「たがる」の「がる」は除く。
カワス	交わす	動詞-非自立可能	「互いに〜する」の意。
カン	間	接尾辞-名詞的-副詞可能	漢語の1最小単位と結合したものは除く。(空=間, 車=間)
ギミ	君	接尾辞-名詞的-一般	
キル	切る	動詞-非自立可能	「すっかり〜し終える」の意。
クサイ	臭い	接尾辞-形容詞的	「〜めいた感じがする」という意。望ましくない意を強める用法。「かび臭い」「焦げ臭い」の「くさい」は除く。
クダサル	下さる	動詞-非自立可能	
グルミ	ぐるみ	接尾辞-名詞的-一般	
クン	君	接尾辞-名詞的-一般	「同君」の「君」は除く。
ゲ	気	接尾辞-形状詞的	
ケイ	系	接尾辞-名詞的-一般	漢語の1最小単位と結合したものは除く。(文=系, 日=系)
ゴ	後	接尾辞-名詞的-一般	漢語の1最小単位と結合したものは除く。(戦=後, 老=後)
ゴ	御	接尾辞-名詞的-一般	
コト	事	名詞-普通名詞-一般	
ゴト	ごと	接尾辞-名詞的-副詞可能	「〜も一緒に」の意。
ゴト	毎	接尾辞-名詞的-一般	そのもの一つ一つ, その時その時の意。
コナス	熟す	動詞-非自立可能	「うまく〜する」の意。
サ	さ	接尾辞-名詞的-一般	「そうだ」「過ぎる」が接続するときの「なさ」「良さ」の「さ」, ケシ型形容詞に付く「さ」, 「憂さ」の「さ」は除く。
サス	さす	動詞-非自立可能	
サス	止す	動詞-非自立可能	
サマ	様	接尾辞-名詞的-一般	

語彙素読み	語彙素	品詞	注記
サン	さん	接尾辞-名詞的-一般	
ジ	時	接尾辞-名詞的-副詞可能	漢語の1最小単位と結合したものは除く。(戦=時)
シキ	式	接尾辞-名詞的-一般	形式・方法などの意。漢語の1最小単位と結合したものは除く。(洋=式, 正=式)
シナ	しな	接尾辞-名詞的-副詞可能	
ジミル	染みる	接尾辞-動詞的	
ジュウ	中	接尾辞-名詞的-副詞可能	
ジョウ	上	接尾辞-名詞的-副詞可能	漢語の1最小単位と結合したものは除く。(機=上, 車=上)
ジョウ	状	接尾辞-名詞的-一般	「～の形・有り様」の意。漢語の1最小単位と結合したものは除く。(液=状)
スギル	過ぎる	動詞-非自立可能	
ズク	尽く	接尾辞-名詞的-一般	
ズクメ	尽くめ	接尾辞-名詞的-一般	
スル	為る	動詞-非自立可能	漢語の1最小単位と結合したものは除く(対=する, 信=ずる)。「～んずる」という形式は除く(甘ん=ずる, 重ん=ずる)。
セイ	性	接尾辞-名詞的-一般	物事の性質・傾向の意。漢語の1最小単位と結合したものは除く。(特=性, 急=性)
ソウ	そう	形状詞-助動詞語幹	様態の助動詞「そうだ」の語幹に当たるもの。
ソウ	そう	名詞-助動詞語幹	伝聞の助動詞「そうだ」の語幹に当たるもの。
ソコナウ	損なう	動詞-非自立可能	
ソコネル	損ねる	動詞-非自立可能	
ソビレル	そびれる	動詞-非自立可能	
ソズル	損ずる	動詞-非自立可能	
タイ	対	名詞-普通名詞-一般	
ダス	出す	動詞-非自立可能	「～し始める」という意。
タチ	達	接尾辞-名詞的-一般	
タテマツル	奉る	動詞-非自立可能	
タマウ	給う	動詞-非自立可能	
ダラケ	だらけ	接尾辞-形状詞的	
チャン	ちゃん	接尾辞-名詞的-一般	
チュウ	中	接尾辞-名詞的-副詞可能	漢語の1最小単位と結合したものは除く。(空=中)
ツイデ	序で	名詞-普通名詞-一般	
ツキ	付き	接尾辞-名詞的-一般	「札付き」(知れわたっていること, 悪い評判が世間に広まっている人の意)は除く。
ツクス	尽くす	動詞-非自立可能	「すっかり～する」という意。
ツケル	付ける	動詞-非自立可能	習慣の意。
ツコ	っこ	接尾辞-名詞的-一般	「～すること」の意。
ツコ	っこ	接尾辞-名詞的-一般	「～比べ」「互いに～する」という意。
ツヅク	続く	動詞-非自立可能	「引き続く」「打ち続く」等, 動作継続の動詞に接続しないものは除く。
ツヅケル	続ける	動詞-非自立可能	「打つ続ける」等, 動作継続の動詞に接続しないものは除く。
ヅライ	辛い	接尾辞-形容詞的	
テキ	的	接尾辞-形状詞的	漢語の1最小単位と結合したものは除く。(人=的, 端=的)
デキル	出来る	動詞-非自立可能	
トウ	等	接尾辞-名詞的-一般	
ドウシ	同士	接尾辞-名詞的-一般	
トオス	通す	接尾辞-動詞的	
トオリ	通り	名詞-普通名詞-副詞可能	それと同じ状態であるという意。
ドノ	殿	接尾辞-名詞的-一般	
トモ	共	接尾辞-名詞的-副詞可能	全部の意。



語彙素読み	語彙素	品詞	注記
ドモ	共	接尾辞-名詞的-一般	
ナイ	内	接尾辞-名詞的-一般	漢語の1最小単位と結合したものは除く。(室=内, 社=内)
ナガラ	乍ら	接尾辞-名詞的-一般	
ナサル	為さる	動詞-非自立可能	
ナシ	無し	名詞-普通名詞-一般	「有る無し」「形無し」「底無し」「台無し」「人で無し」「人無し」「幕無し」「間無し」「道無し」「文無し」「休み無し」の「無し」は除く。
ナミ	並み	接尾辞-名詞的-一般	その類と同じ, 又は同じ程度であることを表す。
ナリ	形	接尾辞-名詞的-一般	そのもの相応である様の意。
ナリ	形	接尾辞-名詞的-一般	「~するまま」「~するに従う様」の意。
ナレル	慣れる	動詞-非自立可能	
ニクイ	難い	接尾辞-形容詞的	醜悪の意の「醜い」は除く。
ヌク	抜く	動詞-非自立可能	「終わりまでする」という意。
ハジメル	始める	動詞-非自立可能	
ハタス	果たす	動詞-非自立可能	「すっかり~し終える」の意。
ハテル	果てる	動詞-非自立可能	「すっかり~する」「~し終わる」という意。
ハナシ	放し	接尾辞-形状詞的	
バム	ばむ	接尾辞-動詞的	
ハン	版	名詞-普通名詞-一般	漢語の1最小単位と結合したものは除く。(新=版)
フウ	風	接尾辞-名詞的-一般	様子の意。漢語の1最小単位と結合したものは除く。(和=風, 古=風)
ブリ	振り	接尾辞-名詞的-一般	それだけの時間が過ぎたという意を表す。
ブリ	振り	接尾辞-名詞的-一般	様子・状態の意。
ブル	振る	接尾辞-動詞的	「そのように振る舞う」という意。
ブン	分	名詞-普通名詞-一般	
ポイ	ぼい	接尾辞-形容詞的	形容詞語幹に接続する「ぼい」は除く。「いがらっぼい」の「ぼい」は除く。
ポッチ	ぼっち	接尾辞-名詞的-一般	
マエ	前	名詞-普通名詞-副詞可能	
マクル	捲る	動詞-非自立可能	
マス	坐す	動詞-非自立可能	
マワリ	周り	接尾辞-名詞的-一般	
ミタイ	みたい	形状詞-助動詞語幹	
ムキ	向き	接尾辞-名詞的-一般	
ムケ	向け	接尾辞-名詞的-一般	
メ	奴	接尾辞-名詞的-一般	ののしる語。
メ	奴	接尾辞-名詞的-一般	謙そんの意。
メ	目	接尾辞-名詞的-一般	順序を表す。
メク	めく	接尾辞-動詞的	擬態語的なものの「めく」は除く。(きら=めく, ざわ=めく)
モウス	申す	動詞-非自立可能	
ヤガル	やがる	接尾辞-動詞的	
ヤスイ	易い	接尾辞-形容詞的	
ヨイ	良い	形容詞-非自立可能	
ヨウ	様	形状詞-助動詞語幹	助動詞「ようだ」の語幹に当たるもの。
ヨウ	様	接尾辞-名詞的-一般	方法の意。
ヨウ	用	接尾辞-名詞的-一般	漢語の1最小単位と結合したものは除く。(学=用)
ラ	等	接尾辞-名詞的-一般	複数を表す。
ラ	等	接尾辞-名詞的-一般	事物をおおよそに指す。
ラシイ	らしい	接尾辞-形容詞的	助動詞「らしい」は除く。
リュウ	流	接尾辞-名詞的-一般	流派の意。
ルイ	類	接尾辞-名詞的-一般	漢語の1最小単位と結合したものは除く。(人=類)
ワスレル	忘れる	動詞-非自立可能	
ワタル	渡る	動詞-非自立可能	「辺り一面に~する」という意。
ワタル	渡る	動詞-非自立可能	「徹底的に~する」という意。