

『現代日本語書き言葉均衡コーパス』と『分類語彙表』を利用した漢字3文字略熟語の抽出

著者	山崎 誠
雑誌名	言語資源活用ワークショップ発表論文集
巻	1
ページ	307-316
発行年	2017
URL	http://doi.org/10.15084/00001486

『現代日本語書き言葉均衡コーパス』と『分類語彙表』を利用した 漢字3文字略熟語の抽出

山崎 誠 (国立国語研究所研究系言語変化研究領域) [†]

Extraction of Clipped Compounds Comprised of Three Character Sino-Japanese Using “Balanced Corpus of Contemporary Written Japanese” and “Word List by Semantic Principles”

Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

要旨

「政財界」「国内外」などの漢字3字で構成される「略熟語」と呼ばれる形式は、先行研究が少なく実態が明らかでない。国語辞書にも掲載されることが少ない。本発表では、現代日本語にはどのような略熟語が存在するかを『現代日本語書き言葉均衡コーパス』(以下、BCCWJ)と『分類語彙表』を使って自動的に抽出することを試みた。具体的には、BCCWJから、前後が非漢字という条件で漢字3文字連続を抜き出し、それらを構成する漢語の頻度および分類語彙表における意味番号を付与したデータを作成した。そこから、出現頻度が一定以上で、構成要素となる漢語の分類番号が一致するものとして874語を抽出した。内訳は「政財界」タイプ656語、「国内外」タイプ297語、重複が79語であった。目視で確認したところ、抽出された3字漢語には、略熟語でないものも多く、精度を高めるにはさらに別の条件が必要であることが分かった。

1. はじめに

「政財界」「国内外」などの主に漢字3文字からなる複合語は、「略熟語¹」と呼ばれ、出現頻度はさほど多くないものの興味深い特徴を持っている。しかし、先行研究が少なくその実態は明らかではない。現象としての指摘は、国立国語研究所(1958: 12)(1959: 274-280)(1962: 19)(1985: 72-73)、野村(1974: 40)、村田(1998: 3,8)、玉村(2002: 200-223)(2005)、野村(2007: 207)、石井(2007)、修(2010)にあるが、詳しい分析を加えた文献は管見の限りでは玉村(2005)のみである²。本稿は現代日本語の略熟語の効率的な抽出方法を提案し、言語的な分析に資することを目的とする。

2. 略熟語とは

略熟語は、共通の要素を持つ2つ(以上)の語が複合する際に、共通の要素と、非共通の要素をつなげて作る複合語のことである。共通の要素は、非共通の要素の前に来ることもあり、後に来ることもある。ほとんどの略熟語は漢字3文字からなる漢語であるが、「小中学

[†] yamazaki (at) ninjal.ac.jp

¹ この名称は国立国語研究所(1985)を執筆した玉村文郎による命名と推測される。同書以前には名称を与えている文献は管見の限りでは見当たらない。

² 「略熟語」は「日本語研究・日本語教育文献データベース」(国立国語研究所)では1件もヒットしない(2017年1月31日検索)。また、CiNiiの全文検索(2017年1月31日検索)では2件ヒットするが1件は、「省略熟語」という別の概念の一部が文字列検索でヒットした例であり、もう1件は石井(2007)の引用であった。

校」のような漢字4文字や、「春夏物（はるなつもの）」のような和語も若干存在する。

略熟語は、略語と熟語の両方の性質を持っている³。2つ（以上）の語が複合する際に省略が起きるといふ点では、「東京大学」を「東大」と略すのと似ているが、略熟語は元となる2つ（以上）の語に共通の要素⁴を持つという点が特徴的である。すなわち、「祖父」と「祖母」を合わせて「祖父母」としたり、「登校」を「下校」を合わせて「登下校」としたりするような例である。「祖父母」の場合は、前要素が共通、「登下校」の場合は、後要素が共通である。この共通要素の位置に着目して略熟語には2つのタイプがあるとされている。1つは「祖父母」のように前要素が共通のタイプ、もう1つは「登下校」のように後要素が共通のタイプである。これらの語は、以下の様な過程を経て作られることが容易に見て取れる。

祖父 + 祖母 → 祖父母
登校 + 下校 → 登下校

国立国語研究所(1985)では、「祖父母」タイプ、「登下校」タイプをそれぞれ、XAB型、CDY型と呼び、玉村(2005)では、 α タイプ・ β タイプと呼んでいる。これらの名称はやや便宜的な名付けであり、名称から内容が推測しにくいいため、本稿では「祖父母」タイプを「前要素共通型（略熟語）」、「登下校」タイプを「後要素共通型（略熟語）」と呼ぶことにする。また、元となる語から略熟語が作られることを玉村(2002)(2005)に倣い、「縮約」と呼ぶことにする。

3. 略熟語の性質

これまで先行研究で挙げられた略熟語の例を通して、その性質を見てみよう。表1に先行研究に出て来た略熟語の例を挙げた。

表1：先行研究で挙げられた略熟語の例⁵

文献	前要素共通型	後要素共通型
国立国語研究所 (1958:12)	絹糸布 祖父母 養父母 国内外	給排水 魚獣肉 青少年 男女児 陶磁器 市町村長
国立国語研究所 (1959:277-278) ⁶	実父母 古金銀 好父母 祖父母 奉迎送 輸出入 養父子	鵜鴉猟 海陸軍 可否決 華士族 金銀貨 金銀鋳 区戸長 芸娼妓 原被告 紅白旗 佐尉官 上下院 上中等 静動産 大小豆 長次官 長次男 町村会 町村費 町村割 内外勤 動植物 府県会 府県官 府県社 府県庁 府県費 陸海軍

³ 「略熟語」という名称自体も略熟語である（略語+熟語→略熟語）。

⁴ 共通の要素のことを玉村(2005)では「軸字」と呼んでいる。

⁵ 掲出順は文献に現れた順である。

⁶ この文献は、明治10(1877)年11月1日から明治11(1878)年10月31日の郵便報知新聞を対象に行った語彙調査（サンプリング調査）で、標本延べ語数は約10万語である。表1に掲げるにあたり、字体を新字体に改めた。この調査では、三字漢語全体を分析しており、その中で「 $a \times (b + c)$ 」という構成を持つものが前要素共通型、「 $(a + b) \times c$ 」という構造を持つものが後要素共通型に相当する。なお、「+」は並列を、「 \times 」は修飾を、() は第一次の結合をそれぞれ表す。

国立国語研究所 (1962: 19)	絹糸布	給排水
野村(1974:40)	祖父母	重軽傷 陶磁器
国立国語研究所 (1985)	輸出入 転出入 移出入 流出入 国内外	許認可 乳幼児 転退職 部課長 政財界 給排水 冷暖房 校園長 本支店 送受信 預貯金 投融资 視聴覚
村田(1998: 8) ⁷		許認可 鋳工業 原材料 原燃料
玉村(2002: 222)	輸出入	判検事 (出入口 ⁸)
玉村(2005: 39-43)	輸出入 国内外 祖父母 養親子 関東西	内外傷 前後期 中近世 自他殺 死傷者 大公使 歩車道 理美容 地家裁 判検事 統廃合 行財政 乳幼児 編著書 青少年 編著書 理美容 耳鼻咽喉科医 ⁹ 陸海空軍 医歯薬大+進学ガイダンス 公私 立大+付属病院 京都+金銀糸+ 平箔+工業組合 海軍+兵經理機 関学校+受験準備要領 定期券+ 出入場+確認システム 医歯薬理 工大+受験 全国市町村長+会議
石井(2007)		乳幼児
修(2010: 49)	輸出入 移出入 流出入 国内外	行財政 出入り口 ¹⁰ 中高生 ¹¹ 乳幼児 入退室 入退場 入退院 緑黄色 中高年 預貯金 送受信 与野党 政財界 転退職 部課長 給排水 冷暖房 校園長 本支店 投融资 視聴覚 判検事

表 1 を概観すると、おおよそ以下のようなことが見て取れる。

- (1)ほとんどが漢字 3 文字である。「市町村長」や「陸海空軍」のような例もあるが、これらは 3 字漢語に比べると少ないと推測される。
- (2)前要素共通型が少なく、後要素共通型が多い。表 1 における前要素共通型の異なりは 15

⁷ この文献は、経済学入門書に現れる漢字三字の専門用語の調査である。専門用語全体では、異なり 1,015 語 (延べ 1,566 語) でそのうち漢字 3 字のものは 130 語である (村田 1998: 1)。

⁸ 玉村(2002: 222)では、「出入口」は、「出口」+「入口」の縮約なのか、「出入り」+「口」の複合なのかの判定がむずかしい」としている。

⁹ この語以降は玉村(2005: 41)で「多次縮約」と位置付けられ、3 語以上が縮約したり、2 語の縮約に他の成分が結合したりする例である。「+」記号は原文にあるとおりである。

¹⁰ 修(2010: 49)に「「出入り」+「口」の説もある」と注記あり。

¹¹ 修(2010: 49)に「稀にある 3 字熟語によるもの」と注記あり。

語であるのに対して、後要素共通型の異なりは、71語である（玉村2005の多次縮約の例を除く）。

- (3)語種は、略熟語及びそれらを構成する語も含めてほとんどが漢語である。「出入り口」が例外となるが、これには「出入り」＋「口」という、略熟語でない解釈も成り立つ。
- (4)略熟語になるときの並び順がほぼ決まっている。すなわち、「輸出入」はあるが、「輸入出」はない。同様に「送受信」はあるが「受送信」はない。表1では、国立国語研究所(1959)の例として「海陸軍」と「陸海軍」が挙げられているが、これは例外と見られる。
- (5)略熟語の元となる語同士は意味的に対義・類義関係にある。すなわち「冷暖房」であれば、その元となる「冷房」と「暖房」は対義語関係にあり、「預貯金」の元となる「預金」と「貯金」とは類義語関係にある。いずれにしても意味的に近い関係にある。

ほぼ同様のことが玉村(2002:222)でも指摘されている。そこでは、縮約の起きる条件として、以下のように述べられている。

この縮約は、①縮約される原単位の個々に、共通の語彙成分が同一位置に存すること、②原単位が近似の意味分野の語であること、③同じ場面で原単位が頻用されること、の3条件が揃ったときに生まれると考えられる（「人事」と「炊事」から「人炊事」は生まれないし、「月曜」と「月給」から「月曜給」は造れない）。

ただし、これらの条件は必要条件であって、これらの条件が満たされても、ただちに、略熟語が出来るわけではない。玉村(2005: 41-42)では以下のように述べている。

「主義」と「主張」の両語はよくセットで使われ、かつ「主」という漢字を共通に有している上、語義も近接している。従って略熟語になりやすいと考えられるが、現実には「主義張」や「主張義」という語は用いられていない。「結離婚」「心肺臓」「親反日」「恒惑星」なども現在までのところ縮約形を生み出していない。必要条件を具備している2成分、3成分であっても略熟語を生み出していないのは、何か十分条件を充たしていない点があるためであろう。

また、玉村(2005:44)では、略熟語が生まれる契機として主に文章論的な観点から、書記スペースの節約、元となる語が近接して使われている、元となる語が頻用されている等の条件が指摘されている。

本稿では、主に玉村(2002:222)の①と②の条件をもとに略熟語を自動的に抽出することを試みる。

4. 略熟語の抽出

4.1 データと方法

略熟語の抽出方法は大きく二通り考えられる。一つは玉村(2002)の条件①②を満たす2字漢語から人工的に略熟語を作り、それがコーパスに現れるかどうか（実際に使われているかどうか）を調べる方法、もう1つは、実際に使われた略熟語の候補を網羅的に収集し、そこから玉村(2002)の①②の条件に合うものを抜き出す方法である。前者は略熟語の可能性の最大を捉えたもので、実際に存在しないものも含まれてくる。仮に玉村(2002)の条件①②に当

ではまる語が 1,000 組あったとすると、それだけで 100 万通りの略熟語の候補ができることになる。処理の簡便さも考慮して今回は後者の方法を採用することにした。

データとして、『現代日本語書き言葉均衡コーパス¹²』（以下、BCCWJ）と『分類語彙表増補改定版データベース』（以下、分類語彙表）を利用する。BCCWJ（約 1 億語）の規模で略熟語を調査した例はないので、この規模で略熟語がどれくらい現れるのかをまず確かめるためである。分類語彙表は、意味の類似性を判定するために用いる。

略熟語を抽出する手順は以下のとおりである。

(1)BCCWJ-DVD に収録された長単位の TSV ファイルを用いて、書字形出現形を連結した文字列を作り、そこから、漢字 3 文字連続で前後が非漢字である文字列を抜き出す。その結果、延べ 2,698,888 文字列、異なり 309,652 文字列が得られた。これらが略熟語の候補¹³である。以降、便宜的に候補語と呼ぶ。

(2)309,652 の候補語に対して、略熟語であった場合に想定される略熟語を作る元となる 2 字漢語を 4 つ用意する（実際には重複があるので 3 つでよい）。

例えば、候補語が「教職員」であった場合、以下のようになる。

候補語	前要素共通型の場合	後要素共通型の場合
教職員	教職 教員	教員 職員

記号化して表すと、候補語が「ABC」であったとすると、「AB」「AC」「BC」の 3 つが必要になる。

候補語	前要素共通型の場合	後要素共通型の場合
ABC	AB AC	AC BC

その上で、候補語には、BCCWJ 長単位語彙表データ¹⁴からの頻度・読み・語種の情報を、元となる漢語候補のそれぞれには、BCCWJ 短単位語彙表データ¹⁵から頻度・読み・語種の情報と分類語彙表の分類番号¹⁶を付与した。

4.2 略熟語の判定

上記(2)で示した略熟語の候補語データのイメージを表 2 に示す。実際には表は横方向に展開するが、見やすさのため縦方向に表示している。

表 2 では結果のうち、いくつかのパターンを示した。「輸出入」は前要素共通型、「原材料」と「教職員」は後要素共通型であるが、略熟語と判定されるかどうか異なる。「屋内外」は両方の型に判定されたものである。

まず、「輸出入」は想定される元の漢語が「輸出」「輸入」「出入」となるが、「輸出」と「輸入」の分類番号が 1.3760 で一致するため、前要素共通型の略熟語と正しく判定される。次に、「原材料」は、想定される元の漢語が「原材」「原料」「材料」となり、「原料」の分類番号が 1.4100、「材料」の分類番号が 1.1040/1.4100 で、共通する 1.4100 があるため、後要素共通型と正しく判定される。「教職員」は、元となる漢語が「教職」「教員」「職員」であるが、「教職」の分類番号 1.3800、「教員」の分類番号 1.2410、「職員」の分類番号が

¹² 使用したバージョンは BCCWJ-NT ver 1.1 である。

¹³ もちろんこれらのほとんどは略熟語でない 3 字漢語である。出現頻度順では、「可能性」(13,577 回)、「日本人」(10,667 回)、「具体的」(9,251 回)と続く。頻度順で最初に現れる略熟語は 179 位の「青少年」(1,165 回)である。

¹⁴ BCCWJ_frequencylist_luw_ver1_0.tsv (http://pj.ninjal.ac.jp/corpus_center/bccwj/freq-list.html)

¹⁵ BCCWJ_frequencylist_suw_ver1_0.tsv (http://pj.ninjal.ac.jp/corpus_center/bccwj/freq-list.html)

¹⁶ 複数の分類項目に出現している場合は、それらを列挙した。

1.2400 で分類番号の一致はなく、略熟語とは判定されないケースである。ただし、分類番号の一致をやや緩和して上 4 桁までの一致でよいとするなら、この例は後要素共通型と正しく判定される。ただし、意味の一致の基準を緩めると適切でない例も混じってくるデメリットがある。「屋内外」は、想定される元の漢語が「屋内」「屋外」「内外」の3つである。「屋内」の分類番号が 1.1770, 「屋外」の分類番号が 1.1770, 「内外」の分類番号が 1.1770/1.1920 となっており、3 つ元の漢語の全てに 1.1770 が共通しているため、前要素共通型と後要素共通型のどちらにも当てはまると判定される（実際には前要素共通型である）。

このように、正しく判定されるもの、類義の条件を緩和すれば正しく判定されるもの、両方に当てはまるものがあるが、実は最も多かったのは、略熟語ではないものである。例えば、「不可能」は、元となる漢語の部類番号が、それぞれ「不可」1.1030/1.1332/1.1346, 「不能」1.1346/1.3421/3.1346/3.3421, 「可能」1.1346/3.1346 であり、1.1346 が共通のため、前要素共通型と判定され、かつ、1.1346 と 3.1346 が共通のため後要素共通型と判定されるが、実際にはそのどちらでもない。

表 2：略熟語の候補語データ（例）

候補語	「輸出入」	「原材料」	「教職員」	「屋内外」
出現頻度	224	482	344	8
長単位頻度	197	449	331	7
読み	ユシュツニュー	ゲンザイリョウ	キョウシヨクイン	オクナイガイ
語種	漢	漢	漢	漢
AB	輸出	原材	教職	屋内
AB 頻度	6147	95	161	541
AB 読み	ユシュツ	ゲンザイ	キョウシヨク	オクナイ
AB 語種	漢	漢	漢	漢
AB 分類番号	1.3760	no ¹⁷	1.3800	1.1770
AC	輸入	原料	教員	屋外
AC 頻度	8581	2053	1947	810
AC 読み	ユニュー	ゲンリョウ	キョウイン	オクガイ
AC 語種	漢	漢	漢	漢
AC 分類番号	1.3760	1.4100	1.2410	1.1770
BC	出入	材料	職員	内外
BC 頻度	84	8892	8601	1329
BC 読み	シュツニュー	ザイリョウ	シヨクイン	ナイガイ
BC 語種	漢	漢	漢	漢
BC 分類番号	1.1530	1.1040/1.4100	1.2400	1.1770/1.1920

¹⁷ 分類語彙表に一致する語がなかったことを意味する。

5 結果

略熟語の候補語データから、長単位頻度 5 以上、長単位の語種及び元となる漢語の語種が漢語であり、分類語彙表番号の 5 桁まで一致しているものという条件により以下の略熟語候補が抽出された。

前要素共通型 297 候補語，後要素共通型 656 候補語，重複 79 候補語

これらの全体を表 3，表 4 に示す。

表 3：前要素共通型と判定されたもの¹⁸ (297 候補語)

悪質性	医学科	異質性	一家族	一世代	一期生	<u>屋内外</u>	開港場	外周部	開幕戦
加減算	加減速	家族内	学級会	下流部	感覺性	感受性	元利金	義侠心	義兄弟
騎乗馬	疑問点	給食費	九千百	九百十	給付費	給与費	教育学	競走馬	競争馬
記録帳	近視眼	<u>近隣国</u>	<u>軍部隊</u>	計算上	計算量	系統図	競馬馬	現在世	減収額
減少額	減少量	減税額	現世代	合意点	公債金	構成図	<u>後世代</u>	構造図	硬直化
抗日戦	高年齢	鉱物油	後方部	<u>国内外</u>	五十六	五千十	五千百	<u>国家内</u>	五百十
五百六	在位中	在京中	最終期	<u>最初期</u>	在職中	在世中	在籍中	在宅中	最多勝
在任中	在米中	山岳地	山間地	三千百	三百十	残余金	地獄界	<u>室内外</u>	<u>実父母</u>
<u>市内外</u>	資本金	<u>社内外</u>	謝礼金	重罪犯	重心点	執着心	<u>終電車</u>	<u>重犯罪</u>	<u>銃砲弾</u>
主幹事	出資金	出場校	出走馬	種目別	種類別	<u>純利益</u>	<u>商工業</u>	上申書	<u>小中学</u>
商店会	上流部	諸行事	<u>食材料</u>	食料品	食糧品	女性史	女性性	人格性	信仰心
人口数	真实性	真理性	頭蓋骨	凶面上	<u>性差別</u>	<u>西南部</u>	精白米	<u>西北部</u>	全会員
全額国	<u>線形状</u>	<u>前後半</u>	全国紙	全国土	戦時中	全社員	全社会	全身部	<u>前世代</u>
全体会	全党員	<u>全部隊</u>	前方部	専門科	増加額	増加量	総監督	<u>増減額</u>	<u>増減税</u>
<u>草書体</u>	増税額	<u>俗世界</u>	<u>俗世間</u>	<u>祖父母</u>	対応等	耐火性	耐寒性	耐久性	<u>大祭典</u>
対策案	耐酸性	耐食性	耐震性	耐水性	耐熱性	単語数	男性性	地球上	地層上
地層中	地層面	<u>地盤面</u>	地表上	<u>地表面</u>	地方区	中央部	中間点	中間部	<u>忠義心</u>
中心核	中心間	中心軸	中心点	中心部	中枢部	<u>忠誠心</u>	中流部	長寿命	貯蓄金
<u>地理学</u>	地理上	<u>賃貸借</u>	<u>定員数</u>	停止車	適合度	<u>敵陣地</u>	適正度	電気灯	<u>電磁界</u>
<u>電磁気</u>	電磁場	電磁波	電話器	<u>同一種</u>	同一性	等価値	<u>同業種</u>	<u>同時期</u>	同質性
<u>同種類</u>	答申案	頭頂部	同等性	<u>東南部</u>	<u>東北部</u>	特殊性	<u>土日曜</u>	<u>内服用</u>	<u>南西部</u>
<u>南東部</u>	難破船	<u>難問題</u>	二十三	二千三	二千十	二千百	二百三	二百十	入院室
入国港	人間界	認定証	年度間	年度中	<u>農作業</u>	納入金	納付金	農牧業	排水水
配水管	配送車	配電線	八十九	発酵熟	八千百	八百十	半額分	分量	鼻孔下
病原体	<u>不可能</u>	<u>不正義</u>	復古調	<u>不適當</u>	<u>不満足</u>	墳丘墓	分極化	分散化	文章体
分水界	文明化	閉会式	閉校式	閉講式	変異性	変化形	返還金	返済金	変動性
方位角	防衛戦	法規制	法制度	<u>北西部</u>	<u>北東部</u>	<u>本図表</u>	<u>未完成</u>	<u>未決定</u>	民主政
<u>名演技</u>	<u>猛攻撃</u>	目的的	文科系	門下生	<u>役職員</u>	有害性	<u>輸出入</u>	<u>幼少時</u>	幼少年
養殖魚	幼年時	<u>養父母</u>	余剰分	預託金	来場所	<u>理科学</u>	理科系	<u>陸海軍</u>	理数科

¹⁸ 網掛けは、実際に前要素共通型と筆者が判断したもの。下線は、後要素共通型とも判定されたもの。

理数系 理想論 律令法 両側面 老朽化 六千十 六百十表4 後要素共通型と判定されたもの¹⁹ (656 候補語)

<u>圧痛点</u>	暗褐色	暗紅色	暗灰色	医学科	異業種	医局長	移住民	異常時	異世界
<u>一家族</u>	一画面	一議員	一機種	一業種	一原因	<u>一三墨</u>	一時代	一車種	一集団
一宗派	一手法	<u>一世代</u>	一側面	一大事	一断面	一昼夜	<u>一二墨</u>	一品種	一分派
一兵士	一方策	一方法	一要因	胃腸炎	一流派	一箇所	異邦人	<u>右左折</u>	英語学
<u>英数字</u>	英文学	英文名	英文字	英訳文	駅職員	絵地図	演奏者	王太子	<u>屋内外</u>
会会員	会会長	会議場	<u>海空軍</u>	外国語	会社員	海水中	灰白色	外壁面	家屋内
歌曲集	各委員	各一部	学院長	学園長	各会社	各議員	各機種	各業界	各業種
各産地	学識者	各社員	各宗派	各職員	各職種	各成員	格闘技	各党派	各品種
学部長	各流派	<u>加減算</u>	<u>加減速</u>	学校外	学校区	学校長	学校内	下背部	下腹部
歌謡曲	管区長	<u>管弦楽</u>	漢文学	<u>喜歌劇</u>	奇関数	義兄弟	既裁定	<u>騎乗馬</u>	義祖母
<u>客貨車</u>	客馬車	旧街道	旧刑法	旧憲法	旧国道	旧国名	旧社名	急性病	旧町名
<u>休廃止</u>	<u>給排水</u>	旧仏教	旧民法	共依存	<u>行財政</u>	強大国	<u>胸腹部</u>	強変化	局次長
<u>許認可</u>	<u>金銀貨</u>	近現代	金鉱山	金細工	金政権	<u>近隣国</u>	区隊長	<u>句読点</u>	<u>軍部隊</u>
傾向性	傾斜角	劇世界	原漢文	県議会	原原種	<u>原材料</u>	原資料	<u>現世代</u>	原地図
剣闘士	高確率	高気圧	後胸部	高金利	航空路	高血圧	<u>鉱工業</u>	高高度	高効率
<u>公社債</u>	好事例	<u>後世代</u>	<u>降積雪</u>	剛速球	広帯域	高体温	皇太子	高地価	高電圧
後頭部	高得点	高能率	高倍率	後半期	後半生	<u>好不況</u>	<u>好不調</u>	公法人	公法的
公民権	高利率	後近代	国語学	国史学	<u>国内外</u>	国文学	御高名	古社寺	御赦免
誤使用	個人性	<u>国家内</u>	国境内	国公法	<u>国公立</u>	今学期	紺無地	再逆転	再協議
<u>最初期</u>	再点検	再登録	再発見	再反論	最末期	雑海草	参議院	三事業	詩歌集
市議会	<u>資機材</u>	資金力	<u>四死球</u>	市場外	死傷者	市場内	<u>視触診</u>	次世代	<u>視知覚</u>
<u>視聴覚</u>	実作業	<u>室内外</u>	<u>実父母</u>	治天下	<u>市内外</u>	詩文学	詩文集	死亡者	弱小国
<u>社内外</u>	<u>受委託</u>	終楽章	習慣性	衆議院	<u>重軽傷</u>	終着駅	<u>終電車</u>	<u>重犯罪</u>	<u>銃砲声</u>
<u>銃砲弾</u>	<u>充放電</u>	終末期	終列車	儒学者	<u>祝祭日</u>	主原因	手術中	<u>出退勤</u>	術直後
<u>出融資</u>	主任務	<u>受発信</u>	<u>受発注</u>	主要因	純損益	<u>純利益</u>	小王国	小会社	<u>上下肢</u>
<u>商工業</u>	小住宅	上前端	小太鼓	<u>小中学</u>	商人道	上腹部	小部隊	小文字	小論文
諸王国	諸外国	諸学説	所管外	<u>食材料</u>	食事後	助数詞	諸大国	<u>初中級</u>	助動詞
所要員	諸流派	真意義	新街道	新改訳	新幹線	新刊本	新喜劇	新機種	新契約
新建築	新憲法	新国劇	心雑音	新施設	新車種	新宗教	新首都	新制作	新品种
新仏教	心理学	新路線	垂直線	水分量	水平面	数文字	正会員	聖画像	<u>政官界</u>
税金額	<u>政財界</u>	税財源	<u>性差別</u>	正社員	青少年	政省令	正職員	青壮年	<u>西南部</u>
性非行	<u>西北部</u>	<u>生没年</u>	正礼装	聖礼典	全音域	全学校	前下部	全画面	全艦隊
全教科	前胸部	前近代	<u>線形状</u>	<u>前後半</u>	<u>前後編</u>	<u>前後席</u>	<u>前後輪</u>	全色盲	前時代
全重量	全盛期	<u>前世代</u>	船隊長	全地域	前頭部	全半壊	前半期	<u>全部隊</u>	全領域
全歴史	線路上	素因数	躁鬱病	総雨量	<u>増改築</u>	総画数	<u>増減額</u>	総件数	<u>増減税</u>
総合計	総戸数	総字数	総重量	<u>贈収賄</u>	<u>送受信</u>	総首長	<u>草書体</u>	総大会	総定数

¹⁹ 網掛けは、実際に後要素共通型と筆者が判断したもの。下線は、前要素共通型とも判定されたもの。

総日数	総婦長	総容量	総連合	俗世界	俗世間	祖父母	村議会	大偉業	大異変
耐陰性	大運河	大宴会	大会戦	大街道	大学卒	大艦隊	堆厩肥	大苦戦	大激戦
大傑作	大決戦	対抗戦	大混戦	大祭典	大災厄	大戦争	大惨事	大地震	代執行
大集会	大小便	大接戦	大大名	大茶会	大帝国	大動乱	大納会	大破局	大爆笑
大迫力	大発会	大反乱	大部隊	大編隊	大暴風	多次元	他諸国	多品種	単独行
短母音	地下界	地盤面	地表面	忠義心	中高年	中上級	中心性	忠誠心	中長期
中低速	中毒性	中年期	町会長	町議会	町村会	町村長	町村民	長母音	直線上
著作者	地理学	珍回答	賃貸借	通行人	通常人	定員数	低確率	低気圧	定休日
低金利	低血圧	低山地	低湿地	低水温	低体温	低倍率	敵陣地	敵戦艦	敵本陣
適用例	鉄軌道	展延性	電磁界	電磁気	電磁力	転廃業	同一種	当会社	同機種
同基地	同業種	頭頸部	登下校	同公社	陶磁器	同時期	糖脂質	同種類	動植物
同女性	動静脈	東南端	東南部	統廃合	同比率	東北部	動名詞	投融资	同連盟
特定性	特別製	都市内	都市民	土日曜	内外装	内水面	内側面	内表面	内服用
南西端	南西部	南東端	南東部	南南東	南北朝	難問題	二三塁	二次元	入会費
入出金	入出力	乳幼児	妊産婦	熱容量	熱流量	年金額	年税額	年利率	農漁業
農漁村	農漁民	農工具	農耕地	農作業	農山村	農民兵	農用地	農林業	売買春
白砂糖	発着駅	半周期	反対論	悲喜劇	微苦笑	非合理	微震動	微振動	微積分
鼻濁音	秘伝書	非論理	不安定	部員数	風水害	不穏当	部会長	部課長	不可能
不完全	不協和	副主題	不公正	不作法	不正義	不誠実	不戦敗	部族長	不存在
部隊長	不注意	仏座像	物質量	物理性	不定休	不適當	不適法	不道德	不必要
不分明	不満足	部門内	不愉快	不用意	文書中	米海軍	米空軍	米陸軍	変異種
編著者	法医学	法規定	法原則	法制定	法哲学	北西端	北西部	北東部	母国語
歩車道	本街道	本原則	本個体	本事業	本事件	本自体	本支店	本数字	本凶表
本法案	本本堂	本問題	本訳書	本論文	未解決	未確定	未完成	未既婚	未決定
未公刊	未指定	未成熟	無香料	無際限	無彩色	無作為	無作法	無定形	無点灯
名演技	名義人	名女優	名選手	名文句	猛攻撃	猛反撃	役職員	薬理学	有彩色
養嗣子	幼少時	要相談	養祖母	陽電子	用筆法	養父母	予警報	預貯金	与野党
来訪客	理化学	理科学	利活用	陸海軍	離着陸	離転職	略系図	略地図	略礼装
榴散弾	領域内	両議院	領国内	両側面	両大国	林産業	例大祭	冷暖房	老病死
論争点	論文集	和漢文	和定食	和洋裁	和洋室				

表3で前要素共通型と判定された297語中、実際に前要素共通型であったのは、網掛けした9語のみであった(適合率約3%)。一方、表4で後要素共通型と判定された656候補語中、実際に後要素共通型であったのは網掛けした119語であった(適合率約18.1%)。

6. おわりに

本稿では、BCCWJと分類語彙表を用いて漢字3文字の略熟語の抽出を試みた。コーパスに実際に出現した漢字3文字列を候補語とし、そこから略熟語の元となる漢語を想定し、それらの意味分野の類似性を利用して略熟語の判定を行った場合、前要素共通型の略熟語は適合率約3%、後要素共通型の略熟語は適合率約18.1%の精度で抽出することができた。精度はかなり低いと言わざるを得ない。

今後の課題として、抽出方法の見直しがある。例えば、今回は漢字3文字列で前後が非漢

字という条件であったが、この場合、「駐停車禁止」のような漢字文字列に埋もれた略熟語が抽出できない。実際に「駐停車」は少納言では20件ヒットするが、今回の方法では出現頻度4、BCCWJ長単位語彙表での頻度は2だった。

意味分野の類似性を判断する分類語彙表も対義となる分類番号が一致していたり、1つだけずれていたりと今回のような調査のためには統一を欠いている。類義語・対義語辞書の併用などが考えられる。

また、後要素共通型と判定された候補語の中には「各」「旧」「新」「大」「不」「本」などの接頭辞を持つものが多く含まれており、これらの多くが略熟語ではないことから、別処理として検討する必要もあろう。

謝 辞

本研究は国立国語研究所コーパス開発センターのプロジェクト「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」により行われたものである。

文 献

- 石井正彦(2007)「略語」飛田良文他編『日本語学研究事典』171 東京：明治書院。
 国立国語研究所(1958)『現代雑誌の語彙調査 総合雑誌の用語 後編』東京：秀英出版。
 国立国語研究所(1959)『明治初期の新聞の用語』東京：秀英出版。
 国立国語研究所(1962)『現代雑誌九十種の用語用字 第一分冊：総記および語彙表』東京：秀英出版。
 国立国語研究所(1985)『語彙の研究と教育（下）』執筆：玉村文郎，東京：秀英出版。
 修徳健(2010)「略熟語の中日対照」『北京日本学術センター25周年記念シンポジウム世界日本学研究趨勢与合作論文集（予稿集）』49-50。
 玉村文郎(2002)「対照語彙論」『朝倉日本語新講座4 語彙・意味』208-235，東京：朝倉書店。
 玉村文郎（2005）「日本語とドイツ語の複合語の対照」Viktoria Eschbach-Szabo, Yoko Koyama-Siebert, Martina Ebi (Hg.) “Ibunka to no deai. Sekai no naka no Nihon to Doitsu” 33-45, BUNKA - WENHUA. Tübinger Ostasiatische Forschungen Bd. 11 (volume11) Münster:Lit Verlag.
 野村雅昭(1974)「三字漢語の構造」『電子計算機による国語研究VI』37-62。
 野村雅昭(2007)「語彙・文字」『国語と国文学』84(5), 202-210。
 村田年(1998)「経済学専門用語三字漢語の語構成—専門分野導入期の日本語教育の方法を探る—」『日本語と日本語教育』26, 1-11, 慶應義塾大学日本語・日本文化教育センター。

関連 URL

『現代日本語書き言葉均衡コーパス』

http://pj.ninjal.ac.jp/corpus_center/bccwj/

『現代日本語書き言葉均衡コーパス』語彙表

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

分類語彙表増補改訂版データベース

http://pj.ninjal.ac.jp/corpus_center/goihyo.html