



HOKKAIDO UNIVERSITY

Title	BCCWJ を用いた基幹慣用句の選定
Author(s)	呉, 琳
Citation	研究論集, 16, 99(左)-113(左)
Issue Date	2016-12-15
DOI	https://doi.org/10.14943/rjgsl.16.l99
Doc URL	https://hdl.handle.net/2115/63912
Type	departmental bulletin paper
File Information	16_009_wu.pdf



BCCWJ を用いた基幹慣用句の選定

呉 琳

要 旨

本研究の目的は、現代日本語における基幹慣用句を選定することである。「基幹慣用句」とは、林 (1971) が提唱した「基幹語彙」にのっとった概念で、調査された言語資料の中で、多方面にわたって高頻度で用いられている慣用句の集合を言う。

本研究の調査対象は、佐藤 (2007) と橋本・河原により選定された 926 句 (延べ 936 句) の慣用句である。調査資料は、『現代日本語書き言葉均衡コーパス (BCCWJ)』の書籍サブコーパスである。この書籍データは、図書館で利用される日本十進分類法 (NDC) に従い、0. 総記, 1. 哲学, 2. 歴史, 3. 社会科学, 4. 自然科学, 5. 技術・工学, 6. 産業, 7. 芸術・美術, 8. 言語, 9. 文学, n. 分類なしの 11 に下位分類されている。つまりは、書籍データにおけるジャンルごとの各慣用句の使用度数が調査可能である。その結果を基にして、慣用句がいくつのジャンルにわたって出現するかを広さの指標、使用度数の多少を深さの指標として設定し、複数のジャンルにわたって高頻度で使用される慣用句を基幹慣用句として選定する。

1. はじめに

Lin Wu (2016) と呉 (2016) では、佐藤 (2007) と橋本・河原が選定した 926 句の慣用句を調査対象として、『現代日本語書き言葉均衡コーパス』(以下においてはこのコーパスの英語名 “the Balanced Corpus of Contemporary Written Japanese” の略語「BCCWJ」を用いてこのコーパスを指すこととする) のうち、書籍に該当するデータを調査資料として、現代日本語における慣用句の使用度数を考察した¹。考察の際、書籍のジャンルという要因は検討の対象外だった

¹ Lin Wu (2016) で用いた BCCWJ モニター公開データ (2009 年度版) は、BCCWJ 一般公開版として構築中のデータの一部を研究利用目的で限定公開したものである。呉 (2016) では、より一般性の

ため、使用度数が多いが、ある特定のジャンルにしか使われない慣用句、あるいは、使用度数が少ないが、複数のジャンルにわたって広く使われる慣用句は存在するのか、もし存在するならば、どのくらいあるのか、といった課題が残された。本研究では、この課題に取り組むべく、基幹慣用句の選定を目的とし、慣用句の「広さ」と「深さ」を分析する。

具体的には、呉（2016）で得られた BCCWJ 書籍コーパスジャンルごとの慣用句の使用度数をデータとして利用する。慣用句がいくつのジャンルにわたって出現するかを表す「広さ」と、使用度数の多少を表す「深さ」を指標として、慣用句の基幹度を測る。

2. 基幹慣用句とは

本研究では、「基幹語彙」という概念にのっとり、慣用句という集合の中で基幹部として存在するものを「基幹慣用句」と呼ぶことにする。基幹慣用句の選定に先立ち、まず「基幹語彙」及びそれに関連する諸概念を整理しておきたい。

「基幹語彙」というのは、林（1971）による造語である。林（1971：2）が新聞の語彙調査から基本語彙を求めようとしたとき、それまでおおざっぱに基本語彙と言われた概念を細かく分けて、次の5つの概念を立てた。

- ①基礎語彙 意味の論理的分析によって求められた半人工的な語彙
- ②基本語彙 特定目的のための「○○基本語彙」
- ③基準語彙 標準的社会人としての生活に必要な語彙
- ④基調語彙 特定作品の基調を作るのに働く語彙
- ⑤基幹語彙 ある語集団の基幹部として存在する語彙

「基幹語彙」がほかの概念から区別される特徴として、林（1971：10-11）は以下のように述べている。

基礎語彙から基準語彙までは、いずれも、指定する一群の語によって、何か大きな働きをさせようとする功利性を、土台にもっている。基幹語彙は、こういう功利性を全く除去去ったところに成立する概念である。

（中略）基幹語彙とは、ある語集団の中に、その集団の骨格のような部分として、その集団をささえる基幹的部分として、現に存在する、語の部分集団を呼ぶ。基幹語彙が基礎語

高い知見を得るために、BCCWJ 一般公開版を調査資料として利用した。調査対象とする慣用句は 926 句であるが、そのうち多義性をもつものが 10 句認められたため、延べ数は 936 句である。

彙から基準語彙までのものと大いに違うところは、要請によって空に描き出される仮設的存在ではなく、現に実在する実体であることである。

林 (1971: 11) は、1つの語彙調査を縦に見るとともに、いくつもの語彙調査を横に見て、どこでもよく使われている語ほど、全資料内でそれだけ基幹的存在だという考え方を示した。つまり、「基幹語彙」は、調査された言語資料の中で、多方面にわたって高頻度で用いられている語の集合を言う。

本研究に言う「基幹慣用句」は、林 (1971) の「基幹語彙」に対する概念に従うゆえ、「基幹慣用句」が意味するもの、また、その選定に用いられた言語資料の中で、複数のジャンルにわたって高頻度で用いられる慣用句のことである。

3. 選定基準

3.1 林 (1971) による選定基準

林 (1971) では、1966 年の朝日・毎日・読売 3 紙を語彙調査した「新聞語彙調査」全データの 3 分の 1 に当たるデータ (長単位延べ 68 万、異なる長単位約 10 万) を使用し、度数 10 以上の 5,417 語を対象として、1,004 語の新聞基幹語彙を示した。

具体的には、話題別の 12 区分 (政治・外交・経済・労働・社会・国際・文化・地方・スポーツ・婦人家庭・芸能・広告) に注目して「層別」と呼び、10～12 層に使われる「極めて広い」語から、1～2 層にしか使われない「狭い」語まで、語が使われる幅の広さを 4 段階に分ける。また、各層での使用度数を語の使われ方の「深さ」とし、使用度数から 6 段階に分け、最も度数の高い階級を 6 点、最も度数の低い階級を 1 点として、12 層の階級点数の合計で「深さ」を設定し、「深い」、「中位」、「浅い」の 3 群に分ける。集計した結果、次の 5 区画に入る 1,004 語を新聞基幹語彙とした。

- 【A1】極めて幅が広く、深さも深いもの 162 語
- 【A2】極めて幅が広く、深さ中位のもの 229 語
- 【A3】極めて幅が広く、深さは浅いもの 198 語
- 【B1】かなり幅が広く、深さが深いもの 10 語
- 【B2】かなり幅が広く、深さ中位のもの 405 語

林 (1971: 27) は、【A1】、【A2】、【B1】に属する語が基幹度の高い語であること、【A3】に属する語は浅くても幅広く存在するため、やはり基幹度が高いと見るべきこと、そして、【B2】に属する語の基幹度も高く評価してよいことを主張している。

3.2 土屋（1992）による選定基準

一方、土屋(1992)は、高等学校の社会科教科書5冊(延べ語数211,244,異なり語数13,041)を対象として、基幹語彙を選定した。具体的には、教科書5科目(倫理社会,政治経済,日本史,世界史,地理)をそのまま採用して5層に分けて「広さ」とする。また、「深さ」は、使用度数を採らず、各教科書の累積使用率の平均値を計算し、各語の「深さ」の度合いとする。平均累積使用率の分布を見ることによって、上位何パーセントまでに含まれる語がどの程度あるかが分かる。平均累積使用率75%までを「深い」語,97.5%までを「中位」の語,100%までを「浅い」語として判断した。その結果,2,061語を社会科基幹語彙の候補として取り出した。

上記2つの先行研究における基幹語彙の選定基準では,1つの共通点が見られる。すなわち,語の使われ方の「広さ」と「深さ」に注目する点である。両者はともに語がいくつかの層にわたって出現することを「広さ」の尺度として使用するが,「深さ」の尺度においては,両者に食い違いが見られる。前者は,各語の使用度数を手掛かりにするのに対し,後者は各教科書の累積使用率の平均値を手掛かりにする。その理由として,後者は,生の使用度数は調査集団の大きさに左右されやすいからであると述べる。つまり,生の使用度数を「深さ」の尺度とすると,各分野で幅広く使われる低頻度の語が資料間の規模の差によって選定の対象から除外される可能性が高いからである。各語の累積使用率の平均値を計算して,それを「深さ」の尺度とすると,資料間の規模の差による問題をある程度補うことができるとと思われる(金2003:73-74)。

林(1971)が提示した「広さ」を測る基準と,土屋(1992)が提示した「深さ」を測る基準に従った実証的な研究として,金(2003)が挙げられる。金(2003)では,日本語と韓国語の「基幹語彙」が選定され,さらに比較語彙論の立場から両言語の対照研究がなされている。

3.3 本研究で用いた選定基準

これらの先行研究を踏まえ,本研究においても「広さ」と「深さ」の尺度を設定する。本研究は調査対象を慣用句という調査資料に出現した一部の語彙に絞ったため,土屋(1992)のような累積使用率の計算が難しいと思われる。したがって,「深さ」の尺度は林(1971)に基づきながら若干の調整を加えることにする。

調査資料としたBCCWJの書籍サブコーパスは,3種類のデータからなっている。同じ書籍データではあるが,それぞれ母集団が異なる。1つめは,出版サブコーパス・書籍で,2001年から2005年に日本国内で出版された全書籍からコーパス収録条件²で絞り込んだ約32万冊が

² 書き言葉コーパスの設計と目的に照らして,以下のような書籍は対象外とした。40ページ以下の書籍,ページ数情報のない書籍(ランダムサンプリングの際にページ情報を利用するため),官公庁刊行物のうち流通していないもの,学習試験図書,電子資料,地図資料,写真集,漫画などである(村田・山崎2011:87)。

母集団である。2 つめは、図書館サブコーパス・書籍で、東京都内の自治体ごとに ISBN により管理されている蔵書データを利用して、13 自治体以上に共通して所蔵されている書籍からコーパス収録条件で絞り込んだ約 38 万冊が母集団である。ISBN の付与が普及した 1986 年から 2005 年刊行の書籍を対象とした。3 つめは、1976 年から 2005 年版の『出版年鑑』及び『出版指標年報』に掲載されたベストセラーリスト上位 20 位に挙げられた書籍約 950 冊を対象としている（村田・山崎 2011：76-77）。

この書籍サブコーパスは、日本十進分類法（NDC）で 0. 総記，1. 哲学，2. 歴史，3. 社会科学，4. 自然科学，5. 技術・工学，6. 産業，7. 芸術・美術，8. 言語，9. 文学，n. 分類なしの 11 に分類されるが，呉（2016）では，ジャンルが明示された 1. 哲学から 9. 文学までを調査範囲とし，0. 総記と n. 分類なしは除くことにした。本研究では，呉（2016）で得られた各慣用句の使用度数に関するデータをそのまま使用する。

表 1 において，本調査で利用した中納言 BCCWJ（通常版）媒体ごとのジャンル別語数を示す³。

表 1 書籍サブコーパス媒体ごとのジャンル別語数

NDC	出版・書籍	図書館・書籍	特定目的・ ベストセラー	合 計
0. 総記	772,760	624,450	94,094	1,491,304
1. 哲学	1,507,725	1,484,705	309,214	3,301,644
2. 歴史	2,520,070	3,179,128	220,009	5,919,207
3. 社会科学	7,359,793	5,870,947	411,888	13,642,628
4. 自然科学	2,701,476	1,737,772	59,661	4,498,909
5. 技術・工学	2,342,557	1,641,476	94,681	4,078,714
6. 産業	1,234,687	950,369	40,347	2,225,403
7. 芸術・美術	1,708,758	2,170,715	197,911	4,077,384
8. 言語	492,975	447,808	42,053	982,836
9. 文学	6,802,589	11,147,690	2,188,986	20,139,265
n. 分類なし	1,108,893	1,122,803	83,417	2,315,113
合 計	28,552,283	30,377,863	3,742,261	62,672,407

調査資料の中では，9. 文学の語数が最も多く，8. 言語の語数が最も少ない。前者の語数が後者の 20 倍以上となり，9. 文学と 8. 言語における慣用句の使用度数も当然両者の大きさに左右されやすい。もしそこから得られた使用度数をそのまま使用すると，8. 言語のような語数の少ないジャンルで使われる低頻度の慣用句が選定の対象から除外されるかもしれない。

³ この語数は，コーパスの構築に際して使われている解析の言語単位である「短単位」で数えたものである。詳細は，<https://maro.ninjal.ac.jp/wiki/index.php?BCCWJ%2F%E7%9F%AD%E5%8D%98%E4%BD%8D%E8%AA%9E%E6%95%B0> を参照。

したがって、各ジャンルに現れる使用度数の重要性を等しくしなければならない。

その解決策として、各ジャンルにおける各慣用句の PMW 頻度⁴を算出し、この指標によって 5 段階に分ける。すなわち、10 回以上を最高グループとし、これを階級 4 とする。7～9 回が次のグループで、これを階級 3 とする。同様に、4～6 回を階級 2、1～3 回を階級 1、0 回を階級 0 とする。ジャンルごとに見る各階級に属する慣用句の数を一覧表にすると、表 2 になる。

表 2 ジャンルごとに見る各階級に属する慣用句の数

ジャンル 階級	BCCWJ における書籍データのジャンル								
	1. 哲学	2. 歴史	3. 社会科学	4. 自然科学	5. 技術・工学	6. 産業	7. 芸術・美術	8. 言語	9. 文学
4 (10 回以上)	20	18	15	12	11	16	21	16	45
3 (7～9 回)	13	10	7	5	6	6	8	12	25
2 (4～6 回)	41	28	12	20	22	26	30	35	55
1 (1～3 回)	340	447	306	183	183	218	330	278	465
0 (0 回)	522	433	596	716	714	670	547	595	346
合 計	936	936	936	936	936	936	936	936	936

次に、各慣用句に属する階級の数そのまま点数として付与する。つまり、階級 4 の慣用句は 4 点、階級 0 の慣用句は 0 点となる。そうすると、各慣用句はジャンルごとに点数をもち、最高点数は 4 点、最低点数は 0 点となる。各慣用句に対し、ジャンルごとの点数を合計すると、すべての慣用句は 0 点（いずれのジャンルにおいても階級 0 に属する慣用句）から 36 点（いずれのジャンルにおいても階級 4 に属する慣用句）までの点数が付与される。

それから、林（1971）を参考にし、「広さ」と「深さ」の尺度を設定する。「広さ」は 4 段階に分け、7～9 ジャンルを「かなり広い」、4～6 ジャンルを「広い」、2～3 ジャンルを「狭い」、0～1 ジャンルを「かなり狭い」とする。一方で、各慣用句に付与された点数をもとにして「深さ」を設定する。19～36 点までを「深い」、10～18 点までを「中位」、0～9 点までを「浅い」とする。

4. 調査結果の概要

表 3 に示すのは、慣用句の「広さ」と「深さ」の分布状況である。

以上の調査結果に基づき、林（1971）が指定した 5 つの区画に属する 199 句の慣用句は基幹慣用句として選定できる。なお、【B1】、【C1】、【C2】、【D1】、【D2】の慣用句がないことと、【B2】

⁴ PMW (per million words) 頻度とは、100 万語当たりの頻度である。

表3 慣用句の「広さ」と「深さ」の分布状況

広さ 深さ	A かなり広い (7～9 ジャンル)	B 広い (4～6 ジャンル)	C 狭い (2～3 ジャンル)	D かなり狭い (0～1 ジャンル)
1 深い (19～36 点)	【A1】 30 句	【B1】 0 句	【C1】 0 句	【D1】 0 句
2 中位 (10～18 点)	【A2】 74 句	【B2】 1 句	【C2】 0 句	【D2】 0 句
3 浅い (0～9 点)	【A3】 94 句	【B3】 205 句	【C3】 207 句	【D3】 325 句

の慣用句が少ないことは、使用度数の多い慣用句は、多くのジャンルにまたがって使われる傾向があり、少数のジャンルにおいて高頻度で使われることがない、ということの意味する。言い換えれば、よく使われる慣用句はジャンルごとの使用差がない。

一方で、【C3】と【D3】に半数以上の慣用句が含まれることと、【A3】、【B3】の慣用句が比較的少ないことは、使用度数の少ない慣用句は、特定のジャンルにおいて使われる傾向があり、多くのジャンルにわたって低頻度で使われることがあまりない、ということの意味する。言い換えれば、あまり使われない慣用句はジャンルごとの使用差がある⁵。

冒頭で言及した課題の結論を言うと、使用度数が多いが、少数のジャンルにしか現れない慣用句は【C1】と【D1】に属するもので、今回の調査では該当する慣用句が見当たらなかった。一方、使用度数が少ないが、複数のジャンルにわたって広く使用される慣用句は【A3】と【B3】に属するもので、合わせて 299 句あった。

選定された基幹慣用句を以下に示す⁶。多義性をもつ慣用句は、『分類語彙表——増補改訂版』⁷

⁵ 本研究の主旨は基幹慣用句の選定にあるため、【C3】と【D3】に属する慣用句のジャンルごとの使用差については言及するにとどめておく。今後、別の機会に譲り、詳しく論じたい。

⁶ ここに掲げた慣用句の形態と表記は、佐藤（2007）『基本慣用句五種対照表』の代表表記に従う。実際の用例検索では、たとえば「有無を言わせず」の用例として、「有無を言わせない/有無を言わせなかった/有無を言わせぬ」といったものも確認できた。ただし、使用できる正規表現のパターンに限られているため、データの収集方法には限界がある。特に、緊密性の低い慣用句に対しては、助詞が置き変わることや、間に別の要素が入ることも考えられるため、実際に収集できていないものがあるかもしれない。したがって、ここではデータをすべて予想通りに拾えているのではなく、原形に近いものしか拾えていないと言わざるを得ない。

⁷ 同書は、現代の日常社会で用いられる約 96,000 語を収録し、分類番号を用いてそれぞれの分類項目の体系的位置付けを示したところに特徴がある。分類の各項目には、たとえば「1.2340」のように類を整数位に置いて小数点以下 4 桁の分類番号が付されている。分類番号の整数位が類の区分に対応する。また部門は、小数点以下 1 桁目の数字で表される。さらに部門をより具体的に細分するものとして、中項目が設けられている。これは小数点以下 1 桁目と 2 桁目を合わせた部分に相当する。なお、慣用句コード付けの問題点と解決策は、呉（2016）を参照。

における該当の意味を表す分類番号を記して区別した。

【A1】 かなり幅が広く、深さが深いもの (30 句)

足を運ぶ、お目にかかる、顔を出す、気が付く、気に入る、気にする、気になる、気を配る、気を付ける、口にする (2.3100)、声を掛ける、異にする、力を入れる、手に入れる、手にする (2.3700)、手にする (2.3392)、手を出す、手を付ける、腹を立てる、間に合う (2.1660)、身に付く、身に付ける、耳にする、耳を傾ける、目にする、目に付く、目に入る、役に立つ、やむを得ず、やむを得ない。

【A2】 かなり幅が広く、深さ中位のもの (74 句)

相づちを打つ、あつけに取られる、案の定、いざという時、命を懸ける、多かれ少なかれ、後れを取る、思いも掛けない、思いも寄らない、掛け替えのない、気が利く、気が済む、気に掛かる、気に掛ける、脚光を浴びる、気を取られる、口にする (2.3331)、口を利く (2.3100)、口を出す、口を挟む、首をかしげる、首をひねる、工夫を凝らす、群を抜く、心を込める、端を発する、宙に浮く、注目を浴びる、手が掛かる、手が届く、手に負えない、手を合わせる、手を打つ (2.3084)、手を掛ける、手を貸す、手を差し伸べる、手を抜く、とてつもない、途方に暮れる、途方もない、取り返しがつかない、取りも直さず、流れをくむ、何が何でも、根を下ろす、念を押す、拍車をかける、話にならない、腹が立つ、枚挙にいとまがない、間に合う (2.1931)、まゆをひそめる、身にしみる、耳に入る、耳を貸す、耳を澄ます、実を結ぶ、身をもって、身を寄せる、胸を張る、目に浮かぶ、目に留まる、目に見えて、目を凝らす、目を付ける、目を通す、目を離す、目を引く、目を見張る、目をやる、物にする (2.3050)、物を言う、世を去る、訳は無い。

【A3】 かなり幅が広く、深さは浅いもの (94 句)

足を伸ばす、足を引っ張る、頭が下がる、頭に来る、頭を抱える、跡を絶たない、息が詰まる、息を吹き返す、意地が悪い、一目置く、意表をつく、嫌と言うほど、意を決する、有無を言わず、大目に見る、遅かれ早かれ、尾を引く、かたずをのむ、型にはまる、肩身が狭い、肩を並べる、気が強い、気が遠くなる、切っても切れない、気に食わない、肝に銘じる、気をもむ、口をそろえる、苦になる、首になる、心に留める、心を奪われる、心を砕く、腰を据える、心血を注ぐ、生計を立てる、精を出す、是が非でも、世話を焼く、背を向ける、先見の明、底を突く、だだをこねる、棚に上げる、知恵を絞る、手が込む、手が付けられない、手を入れる、手をこまねく、手を尽くす、手を引く、手を焼く、手を休める、頭角を現す、度肝を抜く、とどのつまり、長い目で見ると、何かにつけて、何はともあれ、波に乗る、名もない、似ても似つかない、ばつが悪い、幅を利かせる、歯を食い縛る、一役買う、日の目を見る、百も承知、変哲もない、骨が折れる、幕を開ける、幕を閉じる、磨きを掛ける、見切りをつける、身に覚えがない、見様見まね、身を引く、虫がいい、胸

をなで下ろす、目がない、目に触れる、目を奪う、目を掛ける、目を配る、目を背ける、目をつぶる、目を盗む、目を光らす、物にする (2.3700)、用を足す (2.5710)、らく印を押される、らちが明かない、レッテルをはる、我を忘れる。

【B2】 幅が広く、深さ中位のもの (1 句)

息をのむ。

5. 考察

以下、『分類語彙表 ― 増補改訂版』を意味分野の分類基準として利用し、基幹慣用句がどのような意味分野に分布するのかを考察する。

まず、類から見ると、「1. 体の類」が3句 (1.5%)、「2. 用の類」が155句 (77.9%)、「3. 相の類」が36句 (18.1%)、「4. その他の類」が5句 (2.5%) がある。つまり、動詞の類に含まれる慣用句の割合が最も高く、次いで形容詞・形容動詞・副詞・連体詞の類、一部の副詞・接統詞・感動詞の類、名詞の類の順となっている。

次いで、体・用・相三類に含まれる194句を部門から見ると、「.1 抽象的關係」が48句 (24.1%)、「.3 人間活動 ― 精神および行為」が142句 (71.4%)、「.5 自然物および自然現象」が4句 (2.0%) がある。一方、「.2 人間活動の主体」と「.4 生産物および用具」に含まれる慣用句がなかった。つまり、「.3 人間活動 ― 精神および行為」に含まれる慣用句の数が最も多く、その次は「.1 抽象的關係」、「.5 自然物および自然現象」の順となっている。図1に示すのは、基幹慣用句における類と部門の分布状況である。

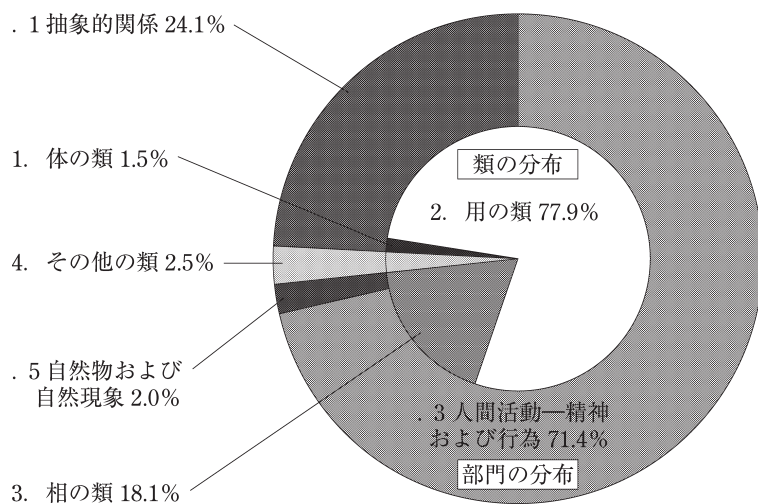


図1 基幹慣用句における類と部門の分布

最後に、中項目から基幹慣用句を観察する。各中項目に含まれる慣用句の数を一覧表に示すと、表 4 になる。

表 4 基幹慣用句における中項目の分布

部門 【総数】	中項目	数	部門 【総数】	中項目	数
.1 【48 句】	10 事柄	1	.5 【4 句】	50 自然	0
	11 類	6		51 物質	0
	12 存在	7		52 天地	0
	13 様相	7		53 生物	0
	14 力	1		54 植物	0
	15 作用	11		55 動物	0
	16 時間	7		56 身体	0
	17 空間	0		57 生命	4
	18 形	0	4. 【5 句】		5
	19 量	8	合 計		199
.3 【142 句】	30 心	95	.1 抽象的關係		
	31 言語	10	.3 人間活動—精神および行為		
	32 芸術	1	.5 自然物および自然現象		
	33 生活	11	4. その他の類		
	34 行為	7			
	35 交わり	5			
	36 待遇	9			
	37 経済	4			
	38 事業	0			

表 4 から「心」、「作用」、「生活」、「言語」に含まれる基幹慣用句が多いことが分かった。「心」に含まれる基幹慣用句が最も多い原因としては 2 つ考えられる。1 つめは、『分類語彙表——増補改訂版』のなかで収録された語彙自体、「心」を表すものが最も多いからである。林 (2009 : 41) によると、「心」を表す語が 11,879 語で、1 割強と分かる。2 つめは、慣用句全体として「心」を表すものが最も多いからである。これはまた、「気」、「心」、「腹」、「胸」といった語が慣用句の構成要素として多く使用されることに起因する。「腹」や「胸」は心が宿る場所で、「気」や「心」と同様に、心の状態や動きを表すとされる。この 2 点を踏まえると、基幹慣用句にも、当然「心」を表す慣用句が多く存在することとなるだろう。

以上、選定された基幹慣用句に対し、類・部門・中項目という 3 つの面から意味分野を考察した。全体として、動詞の基幹慣用句が多いのは、動詞の表現をより豊かにするという傾向が書籍において強く見られるからである。また、抽象的なものを表す基幹慣用句が多いのは、外界に存在する物体や景色といった具体的存在を用いて、思いや感情といった抽象的存在を表すという人間の思考パターンによるものと考えられる。

基幹慣用句と対立の立場に位置するのは、表 3 の【D3】に相当する一群の慣用句 (少数のジャ

ンルにしか低頻度で使用されない慣用句)で、0点の慣用句178句、1点の慣用句146句、2点の慣用句1句という内訳となっている。以下、これらの慣用句を掲げた上で、その性格を分析する。

【D3】 かなり幅が狭く、深さが浅いもの (325 句)

青筋を立てる、赤子の手をねじる、赤子の手をひねる、揚げ足を取る、あごで使う、あごを出す、足音を忍ばせる、足が付く、足が出る、足が鈍る、足が速い、足が棒になる、味も素っ気も無い、足元に火がく、足元を見る、足を棒にする、頭隠してしり隠さず、後は野となれ山となれ、穴があったら入りたい、危ない橋を渡る、あぶはち取らず、油を売る、油を絞る、網を張る、ありのはい出るすき間もない、合わせる顔がない、泡を食う、生き馬の目を抜く、息が切れる、息をつく暇もない、石にかじり付いても、石橋をたたいて渡る、一日の長、一矢を報いる、一杯食う、命の洗濯、今や遅しと、芋を洗うよう、いやでも応でも、因果を含める、上を下への大騒ぎ、牛の歩み、後ろ指を指される、腕が上がる、腕が鳴る、腕に覚えがある、腕によりを掛ける、打てば響く、うなぎの寝床、うの目たかの目、海の物とも山の物とも付かない、大きな顔をする、大きな口をきく、大目玉を食う、お株を奪う、お先棒を担ぐ、お茶の子さいさい、鬼の目にも涙、尾ひれを付ける、重荷を下ろす、恩に着せる、恩に着る、恩をあだで返す、飼い犬に手をかまれる、顔が売れる、顔がつぶれる、顔から火が出る、顔に泥を塗る、顔向けができない、顔を曇らせる、顔をつぶす、陰になりひなたになり、かさに懸かる、肩で風を切る、肩の荷を下ろす、肩をすぼめる、角が取れる、金に糸目をつけない、かぶとを脱ぐ、かまを掛ける、かゆい所に手が届く、我を折る、我を張る、気炎を上げる、気がとがめる、気をもめる、木で鼻をくくる、気骨が折れる、肝が据わる、肝が太い、興に乗る、気を落とす、悔いを残す、ぐうの音も出ない、空を切る、臭い物にふたをする、口がうまい、口が軽い、口が滑る、口が減らない、口から先に生まれる、口が悪い、くちばしが黄色い、くちばしを入れる、口も八丁手も八丁、口を利く (2.3522)、口を切る (2.1553)、首が回らない、雲をつく、軍門に下る、けがの功名、げたを預ける、けちが付く、犬猿の仲、けんもほろろ、口角泡を飛ばす、声を限りに、声をのむ、志を遂げる、腰を折る、事と次第によっては、事によると、言葉に尽くせない、言葉を濁す、細大漏らさず、三拍子そろろう、思案に暮れる、下にも置かない、地団太を踏む、十指に余る、しっぽを出す、しっぽをつかむ、春秋に富む、食指が動く、知らぬが仏、白羽の矢が立つ、しりが重い、しりに火が付く、しり目にかける、死力を尽くす、心臓が強い、すずめの涙、図に当たる、すねに傷を持つ、すねをかじる、隅に置けない、精が出る、世話が焼ける、相好を崩す、そでにする、大事を取る、たがが緩む、立つ瀬がない、手綱を引き締める、立て板に水、盾を突く、他人の空似、玉にきず、矯めつすがめつ、駄目を押す、力を落とす、血も涙もない、茶茶を入れる、調子に乗る (2.1660)、長足の進歩、ちょうちんを持つ、血わき肉踊る、月とすっぽん、土が付く、つむじを曲げる、つめに火をともし、つめのあかほど、面の皮が厚い、手が上がる、手がふさがる、手ぐすねを引く、手に汗を握る、手に乗る、手八丁口八丁、手を取る、

手を握る、堂に入る、毒にも薬にもならない、どこ吹く風、どじを踏む、取って付けたよう、トップをきる、とらの威を借るきつね、泥を吐く、度を失う、度を過ごす、泣きの涙、何はさておき、何は無くとも、名を売る、煮え湯を飲まされる、にしきを飾る、抜き足差し足、寝返りを打つ、願ったりかなったり、猫の手も借りたい、猫をかぶる、熱が冷める、熱しやすく冷めやすい、熱を入れる、年季を入れる、年貢の納め時、乗り掛かった船、のれんを分ける、ばかになる、歯が抜けたよう、馬脚を現す、白紙に戻す、薄氷を踏む、化けの皮がはがれる、はしにも棒にも掛からない、旗色が悪い、破竹の勢い、はちの巣をつついたよう、発破をかける、バトンを渡す、鼻息が荒い、鼻が利く、鼻が高い、鼻であしらう、鼻に付く(2.5040)、鼻を折る、歯にきぬを着せない、羽を伸ばす、歯の根が合わない、腹が据わる、腹が太い、腹の虫が治まらない、腹を探る、腹を据える、火が消えたよう、ひざを乗り出す、ひざを交える、額を集める、引込みがつかない、一泡吹かせる、人聞きが悪い、一肌脱ぐ、ひとみを凝らす、人目をはばかり、人を食う、日の出の勢い、不意を食う、風雲急を告げる、風前のともしび、不覚を取る、袋のねずみ、筆が立つ、筆を入れる、筆を加える、筆をふるう、船をこぐ、踏んだりけったり、へそを曲げる、へとも思わない、骨身にこたえる、骨身を惜しまない、骨身を削る、骨を埋める、魔が差す、的を射る、右から左、右に出る者がいない、水も漏らさぬ、水を打ったよう、みそを付ける、道草を食う、身に余る、耳が早い、耳にたこができる、耳に挟む、耳をそばだてる、耳をそろえる、脈が有る、見るに見兼ねて、虫が知らせる、虫ずが走る、虫の息、虫の知らせ、胸くそが悪い、胸が躍る、胸がすく、胸がつぶれる、胸が膨らむ、胸がふさがる、目が肥える、目が据わる、目が高い、芽が出る、目が回る、目から鼻へ抜ける、目と鼻の間、目に角を立てる、目に物見せる、目の上のこぶ、目の黒いうち、目の中に入れても痛くない、目鼻が付く、芽を吹く、目を細くする、もったいをつける、物は試し、焼きが回る、焼きを入れる、焼け石に水、やせても枯れても、矢の催促、やぶから棒、有終の美を飾る、湯水のように使う、弓を引く、欲の皮が突っ張る、横車を押す、横やりを入れる、夜目も寝ずに、夜も日も明けない、寄ると触ると、弱り目にたた目、らっぱを吹く、我も我もと。

【D3】の慣用句を類から見ると、「1. 体の類」が27句(8.3%)、「2. 用の類」が206句(63.4%)、「3. 相の類」が87句(26.8%)、「4. その他の類」が5句(1.5%)である。なお、各部門と中項目に含まれる慣用句の内訳は表5に示す。

呉(2016)が指摘したように、慣用句は全体的に、「1. 抽象的關係」と「3. 人間活動——精神および行為」という2部門には集中して分布するが、「2. 人間活動の主体」、「4. 生産物および用具」、「5. 自然物および自然現象」という3部門にはあまり分布しない。それゆえ、基幹慣用句も【D3】に属する慣用句も当然、表4と表5が示した如く、部門1と部門3に集中し、部門2、部門4と部門5には集中しないのである。

慣用句の構成要素からみると、文章語や現在では使われなくなる古い言い方が存在するのが特徴である。そのうち、「春秋に富む」、「とらの威を借るきつね」や「薄氷を踏む」など中国語

表5 【D3】に属する慣用句における中項目の分布

部門 【総数】	中項目	数	部門 【総数】	中項目	数
.1 【71句】	10 事柄	0	.4 【1句】	40 物品	0
	11 類	4		41 資材	0
	12 存在	9		42 衣料	0
	13 様相	19		43 食料	0
	14 力	3		44 住居	1
	15 作用	15		45 道具	0
	16 時間	12		46 機械	0
	17 空間	0		47 土地利用	0
	18 形	0	.5 【10句】	50 自然	2
	19 量	9		51 物質	0
.2 【0句】	20 人間	0		52 天地	0
	21 家族	0		53 生物	0
	22 仲間	0		54 植物	0
	23 人物	0		55 動物	0
	24 成員	0		56 身体	1
	25 公私	0		57 生命	7
	26 社会	0	4. 【5句】		5
	27 機関	0	合 計		325
.3 【238句】	30 心	106	.1 抽象的關係		
	31 言語	30	.2 人間活動の主体		
	32 芸術	3	.3 人間活動—精神および行為		
	33 生活	15	.4 生産物および用具		
	34 行為	29	.5 自然物および自然現象		
	35 交わり	18	4. その他の類		
	36 待遇	31			
	37 経済	6			
	38 事業	0			

に出典があるものもあれば、「軍門」、「三拍子」や「手綱」など特定の分野にしか用いられていない用語もある。また、「後ろ指」、「地団太」や「煮え湯」など慣用句にしか現れない語もある。【D3】に属する慣用句の使用度数が少なく、少数のジャンルにしか使われないのは、このような語を構成要素にもつからであろう。

この結論を裏付ける事例研究として、呉 (2010) と呉 (2013) がある。呉 (2010) と呉 (2013) は、類義の慣用句として「怒り」を表すものを取り上げ、コーパスに基づく使用実態調査と母語話者を対象にした使用意識のアンケート調査を行った結果、使用実態調査で使用度数の少ない慣用句、及び使用意識調査で馴染みの薄い慣用句には共通点が見られることを指摘した。すなわち、「怒髮冠を衝く」や「怒髮天を衝く」のように中国の古典に出自すること、または、「柳眉を逆立てる」や「眦を決する」のように現在では使われなくなる古い語が含まれることである。

一方、基幹慣用句（特に【A1】、【A2】に属する慣用句）の構成要素を見ると、このような語がほとんど含まれず、「顔」、「手」や「目」などの身体部位名詞と、「入れる」、「掛かる」、「掛ける」や「付ける」などの日常よく用いられる動詞が含まれる。ここから、慣用句の基幹度は、構成要素に左右されやすいことがうかがわれる。

6. おわりに

以上、BCCWJの書籍サブコーパスを利用しながら、慣用句がいくつかのジャンルにわたって出現するかを「広さ」、どのくらい使われるかを「深さ」として設定し、199句の基幹慣用句を選定した。

従来、語の基本度についてはいくつかの考え方がある。林（1971）が提唱した「基幹語彙」のほかには、意味・用法からみて重要と思われる語も基本的な語であるという考え方が示されている。この意味・用法を考えて選ぶ方法は、選定者の主観によって偏る可能性がある反面、満遍なく語を収集することができる（荻野2007：47-48）。これに反して、本研究ではより客観的に選定するため、慣用句の使用度数と使われるジャンルを選定の要因として考慮に入れたが、調査の便宜を図り、最初から調査対象を936句の慣用句に絞ったため、広く慣用句を集めたとは言えない。また、調査資料が書籍に限定されたため、資料によって偏りが出ること考えられる。

今後の課題として、書籍コーパスから得られた基幹慣用句を現代日本語における基幹慣用句の候補として扱う一方で、その他の資料における慣用句の使われ方や意味・用法からみて重要と思われる慣用句についても検討を試みる。このように、客観的な選定方法と主観的な選定方法を併用することによって、現代日本語における基幹慣用句を補いたい。

（ごりん・言語文学専攻）

参考文献

- 荻野綱男（2007）『現代日本語学入門』明治書院
 金直洙（2003）「日韓『基幹語彙』の比較研究——選定および意味分野別構造の分析を中心に——」『語彙研究』創刊号 pp. 67-84
 国立国語研究所編（2004）『分類語彙表——増補改訂版』国立国語研究所資料集14 大日本図書株式会社
 呉琳（2010）「アンケート調査からみた『怒り』を表す慣用句の使用頻度——中国語と対照しながら——」『日本語文学』第47輯韓国日本語文学会 pp. 141-157
 呉琳（2013）「『怒り』を表す慣用句の使用頻度の中日対照」『漢日語言対比研究論叢』第4輯北京大学出版社 pp. 244-259
 呉琳（2016）「BCCWJにおける慣用句の使用度数及び意味分野」『語彙研究』13号 pp. 43-50

- 佐藤理史編 (2007) 『基本慣用句五種対照表』名古屋大学大学院工学研究科佐藤理史研究室
- 土屋信一 (1992) 「基幹語彙の探索」『文化言語学 その提言と建設』三省堂 pp. 743-754
- 林四郎 (1971) 「語彙調査と基本語彙」『電子計算機による国語研究Ⅲ』国立国語研究所報告 39 秀英出版 pp. 1-35
- 村田年・山崎誠 (2011) 「『手』の慣用句を指標とした文章ジャンルの判別 ― 現代日本語書き言葉均衡コーパスを用いて ―」『日本語と日本語教育』39 慶應義塾大学日本語・日本文化教育センター紀要 pp. 75-88
- 林彦伶 (2009) 「『分類語彙表』における各意味分野の語数」『語彙研究』7号 pp. 40-47
- Lin Wu (2016) “Frequencies and Semantic Category Distribution of Idioms in Japanese” *Journal of the Graduate School of Letters, Hokkaido University*, Vol. 11, pp. 75-82

関連 URL

- 『現代日本語書き言葉均衡コーパス (BCCWJ)』中納言 <http://chunagon.ninjal.ac.jp/> (最終アクセス日 2016 年 7 月 25 日)
- 橋本力・河原大輔「OpenMWE for Japanese」〈<http://openmwe.osdn.jp/pukiwiki-j/#nab7f229>〉(最終アクセス日 2015 年 6 月 27 日)

付記 本研究は中国国家留学基金管理委员会「国家建設高水平大学公派研究生項目」の支援による研究成果の一部である。