『分類語彙表』の質的拡張の試み

『分類語彙表』の質的拡張の試み

山崎 誠（国立国語研究所研究系言語変化研究領域）†

An Attempt to Qualitative Expansion to the “Word List by Semantic Principles”

Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

要旨

『分類語彙表』は初版の刊行以来，日本語研究に利用されてきた。しかし，2004 年に増補改訂版が刊行されて以来，さらなる増補は行われていない。本稿は，『分類語彙表』を研究に利用する上で，もっとも重要な課題の一つである，不採録語を減らすという観点から，語彙の拡充の方法を分類体系の見直しを中心に検討し，試案を提示するものである。語彙の拡充の候補は以下のとおりである。(1) 助詞・助動詞などの機能語（2）固有名詞（固有表現）（3）外国語（4）メタ言語（5）句読点などの記号類（6）語断片（7）未知語。(1) ～（3）は，意味の付与が可能なもの，(4) 以降は，意味付与が可能でない（必要が無い）ものである。助詞・助動詞などの機能語は品詞相当と考え，0 番台を与える（例えば格助詞「が」に分類語彙表番号 0.1000 を与えるなど）。固有名詞（固有表現）は，現在の分類体系をできるだけ維持するのであれば，内包的表現の所属する分類項目に位置付けるのが妥当であろう（「アカデミー賞」「グラミー賞」は「1.3682 賞罰」に置くなど）。メタ言語的用法は意味分類には反映させず，「用法」という別フィールドで属性を記述する。また，句読点，語断片，未知語は意味付与が不要という属性を与えて区別することを考えている。以上のような拡張で，ほぼ全ての語に何らかの分類語彙表番号を与えることが可能となる。

キーワード：分類語彙表，分類体系，機能語，固有名詞，メタ言語，補助記号

1. はじめに

『分類語彙表』（1964 年初版，2004 年増補改訂版。以降，増補改訂版を単に『分類語彙表』，あるいは WLSP と称する）は，現代日本語のシソーラスとして日本語研究に利用されてきた。しかし，2004 年の増補改訂版の刊行以後，改訂が行われていない。したがって，現在までの間の新語や新用法には対応できていない。また分析においては，山崎・柏野（2017）で指摘されるような，文脈における多義語の意味をどう決定したらよいかという問題もある。本稿は，データを分析する際に，『分類語彙表』に載っていないため，分類語彙表番号（以下，「分類番号」とする）が付与出来なかった，ということをできるだけ減らすことを目的として『分類語彙表』の質的拡張のための方策を探るものである。

2. 先行研究

一般的なテキストで使用されていて『分類語彙表』に掲載されていない語としては，機能語（助詞・助動詞），固有名詞，外国語が挙げられる。これらについて，従来のシソーラスでの扱いを概観する。

2. 1 機能語

土居（1933）は，基礎語 1000 語に意味分類を施して提示したものであるが，その中に，

† yamazaki [AT] ninjal.ac.jp

機能語に該当する分類項目がある。「関係を表わす語[1]」として「て，で，の，に，を，は，も，や，か，へ，於（おい）て，よう，ほど，ばかり，故（ゆえ），から，より，さへ，ついて，以って，共（とも），まま，まで」が挙げられている。また，「語の尾に添える語」として「らしい，たい，ます，ない，ぬ，べき」[2]が見られる。

垣内（1936）は，『小學國語讀本』巻一～巻四に出現する語を分類したものであるが，機能語も分類されている。例えば，「9 関係[3]」に「ハ，ガ，ノ，ヲ」が，「43 結合」に「ト，モ，ヤ」が，「278 方向」に「カラ，ヨリ，マデ，へ，ニ」が，「461 尋問〔尋問の主題〕問題」に「カ，やら[4]」が，「465 識別」に「こそ」が，「476 推理」に「ノデ，ば」が，「535 肯定」に「ヨ，ネ，ナ，ゾ，デス，ダ，マス」が，「536 否定」に「ズ，ン，ない，まい，な」が，「602 欲求」に「たい，よう」が，「865 欲望」に「タイ」などである。

田島（2001）は，『分類語彙表』の拡張を試みた例である。ここでは，分類番号のピリオドより前の部分を変えることによって機能語等を位置付けている。例えば，助詞・助辞は，「8.xxx」，助動詞は「9.xxx」で表すとしている。

2．2　固有名詞

『分類語彙表』には「1.2390 固有人名」「1.2590 固有地名」という固有名詞を含む分類項目がすでにあるが，それ以外の固有表現はほとんど収録されていない。例外的に「古事記」「日本書紀」「万葉集」などの書名が見られる程度である。田島（2001）では，固有名詞は「15.xxx」に位置付けている。

2．3　外国語

日本語のシソーラスなので，外国語は直接対象にならず，収録しているものはない。田島（2001: 195）では，「日本語訳したものにコード付けする」としているが，「詳細については検討中とのことである。

3．分類番号が付かない語

どのような語に分類番号が付与されないのか，小調査を行った。『現代日本語書き言葉均衡コーパス』（以降，BCCWJ）からサンプルの語数が短単位で 3000 語[5]程度のサンプルを各レジスターから 1 つ選んだ。ただし，OV（特定目的・韻文），OC（特定目的・Yahoo!知恵袋）は，3000 語近いサンプルがなかったため，そこからは選んでいない。選ばれたサンプルを表 1 に示す。

表 1　BCCWJ から選んだサンプル

サンプル ID	レジスター	ジャンル	出版年	語数(全て)
LBo0_00012	図書館・書籍	0 総記/070/0236	2000	3000
OB4X_00182	特定目的･ベストセ	3 社会科学/368/0095	1994	2999
OL5X_00043	特定目的・法律	10 民事//	1999	3319
OM62_00002	特定目的・国会会	衆議院/特別委員会/災害対策	2002	3373
OP22_00010	特定目的・広報誌	関東地方/埼玉県/	2008	3012

[1] 引用にあたっては，旧字体は現行字体に，歴史的仮名遣いは現代仮名遣いに改めた。以下同じ。

[2] このほかにも「様（さま），君（くん）」などの接尾語もこの分類に含まれている。

[3] 「関係」の前にある「9」という数字は分類項目に付けられた一連番号である。原文は漢数字であるが，読みやすさのためにアラビア数字変えた。

[4] 平仮名表記と片仮名表記の違いは不明である。

[5] 補助記号・空白も含んだ短単位数。短単位数の情報は，「短単位語数 Excel データ(Version 1.1)」から得た。

OT22_00004	特定目的・教科書	理科/中/2	2005	3034
OW2X_00115	特定目的・白書	経済/労働経済白書（労働白	1981	3004
OY15_14045	特定目的・ブログ	趣味とスポーツ/趣味/その他	2008	3003
PB41_00178	出版・書籍	1 哲学/193/0316	2004	3005
PM41_00597	出版・雑誌	総合/一般/一般週刊誌	2004	3000
PN1k_00007	出版・新聞	地方紙/神戸新聞/	2001	3001

これらのサンプルに含まれる全短単位を『分類語彙表番号－UniDic 語彙素番号対応表』（wlsp2unidic）を使って，分類番号を付けた。wslp2unidic は，分類番号から UniDic の語彙素 ID への対応表であるが，そこから，UniDic の語彙素から分類番号を逆引きできるようにして，分類番号を付与した。

3．1　結果

表 2 に分類番号が付与された短単位と付与されなかった短単位の語数とその割合を示した。どのサンプルにおいても，5 割前後の短単位に分類番号が付かなかったことになる。

表 2　分類番号が付いた短単位と付かなかった短単位

サンプル ID	全短単位数	WLSP 番号有り(割合)	WLSP 番号無し(割合)
LBo0_00012	3000	1553 (0.518)	1447 (0.482)
OB4X_00182	2999	1561 (0.521)	1438 (0.479)
OL5X_00043	3319	2055 (0.619)	1264 (0.381)
OM62_00002	3373	1941 (0.575)	1432 (0.425)
OP22_00010	3012	1769 (0.587)	1243 (0.413)
OT22_00004	3034	1538 (0.507)	1496 (0.493)
OW2X_00115	3004	1864 (0.621)	1140 (0.379)
OY15_14045	3003	1353 (0.451)	1650 (0.549)
PB41_00178	3005	1496 (0.498)	1509 (0.502)
PM41_00597	3000	1282 (0.427)	1718 (0.573)
PN1k_00007	3001	1596 (0.532)	1405 (0.468)
全体	33750	18008 (0.536)	15742 (0.466)

表 3 は，全 11 サンプルにおいて分類番号が付かなかった短単位の品詞を頻度の降順に並べたものである。上位には助詞・助動詞と補助記号，空白が並ぶ。これらの合計は 14326 語であり，もし，これらの分類番号が付与されたら，全体の約 95.8%に付与できたことになる。さらに，固有名詞が 801 語あり，それらにも分類番号が付与されたとしたら，全体の約 98.2%をカバーすることになる。

表 3　分類番号が付かなかった短単位の品詞（頻度 4 以上）

順位	品詞	頻度
1	助詞-格助詞	5436
2	助動詞	1727
3	補助記号-読点	1599
4	助詞-係助詞	1201

5	助詞-接続助詞	1167
6	補助記号-句点	1027
7	空白	637
8	補助記号-括弧閉	450
9	補助記号-括弧開	446
10	補助記号-一般	409
11	名詞-普通名詞-一般	302
12	名詞-固有名詞-地名-一般	259
13	名詞-固有名詞-人名-名	258
14	名詞-固有名詞-人名-姓	216
15	助詞-副助詞	135
16	形状詞-助動詞語幹	103
17	助詞-終助詞	92
18	動詞-一般	70
19	名詞-固有名詞-人名-一般	47
20	名詞-数詞	26
21	名詞-固有名詞-一般	21
22	記号-一般	20
23	接尾辞-名詞的-一般	19
24	記号-文字	14
25	接尾辞-名詞的-副詞可能	12
26	未知語	8
27	名詞-普通名詞-サ変可能	8
28	形状詞-一般	7
29	接頭辞	5
30	副詞	4
31	接尾辞-名詞的-助数詞	4

4. 分類体系の拡張

4. 1　助詞・助動詞と補助記号

前節を受けて，助詞，助動詞と補助記号に分類番号を与えるのが効果的だと分かった。助詞・助動詞は品詞分類上，それぞれ 1 つの品詞と位置付けられるため，『分類語彙表』のシステム上，ピリオドより前の数字として表すのが妥当であろう。現在使用されているのは，1～4 の数字で，それぞれ「体の類」「用の類」「相の類」「その他の類」に割り当てられている。この中だけで言えば「その他の類」が候補になるが，「その他の類」の内部は，「4.11 接続」「4.30 感動」「4.31 判断」「4.32 呼び掛け」「4.33 挨拶」「4.50 動物の鳴き声」という分類になっており，接続詞，感動詞，一部の副詞に対応した内容になっている。そこで，使われていない「0」を助詞・助動詞に割り当てる方法が考えられる。下位分類をどうするかは今後の課題であるが，垣内（1936）のように，助詞・助動詞の機能・意味に対応させるのが妥当と思われる。なお，すでに一部の助詞・助動詞が『分類語彙表』に収録されているが，それらは，新設の 0 番台に移すことになろう。

補助記号や空白は語彙的な意味を持たないと考えられるため，分類番号がなくてもよい

のであるが，番号を持たないということを積極的に表すための記号を与えるということも考えられる。この場合の記号は数字ではなく，例えば「Z」のようなアルファベットが考えられる。例えば，「。」は，「Z.0001」のように記述する。

4．2　固有名詞（固有表現）

固有名詞の範囲は，狭く捉えると，人名および地名のみ，広く捉えると，組織名，動植物名，化学物質名，商品名，書名，曲名など命名により作られた語がすべて対象となる。固有名詞を『分類語彙表』上に位置付ける方法として 2 つが考えられる。1 つは，固有名詞だけを一箇所に集めて，その中の下位分類として体の類の各分類項目を利用するという方法と，固有名詞を外延的表現と捉え，その内包に当たる分類項目の中に位置付けるというものである。前者は固有名詞を一括して捉えることができるので便利であるが，分類体系を少なからず変更する必要がある。後者は，固有名詞がそれぞれの分類項目に散らばることになるが，分類項目の変更はほとんど必要ない。例えば，「1.3682 賞罰」に「ノーベル賞」があるが，ここに「アカデミー賞」「グラミー賞」「芥川賞」「直木賞」などを追加すればよい。その語が固有名詞であるかどうかという情報は，分類番号とは別のフィールドを設けてそこで表すことも検討されるべきであろう。このような分類番号には反映しにくい語の属性については，5 節の「今後の課題」で述べる。

4．3　外国語

日本語文脈中に現れた外国語は田島（2001）で述べているように，日本語に訳した場合に該当する分類番号を与えるのが妥当であろう。仮に「This is a pen.」という文があった場合には，This は，「1.1010 こそあど・他」，is は，新設の「0.xxx」，a は「3.1960[6]」，pen は「1.4530 文具」が該当する。

4．4　メタ言語

「「です」は丁寧語です。」における最初の「です」はメタ言語的用法であり，通常の助動詞である 2 つめの「です」とは違う用法である。このような場合，意味分類の違いではなく，用法の違いと考え，2 つの「です」を同じ分類番号を与えるほうが負担が少ない。もし，メタ言語的用法を取り込む場合には，別のフィールドで区別するのがよいであろう。

4．5　語断片・未知語

これも補助記号と同じように，語彙的な意味を持たないと考え，そのことを積極的に表す「Z」の記号を与える。また，未知語は調べてみて語の意味が分かれば，適切な分類番号が付くが，未知語の段階では，やはり「Z」の記号を与える。

5．今後の課題

助詞・助動詞に分類番号を与えることで，実質的にかなりの語に分類番号が付与されることになる。今後の課題としては，2004 年以降の新語，新用法の増補がある。

また，文脈における意味を考える場合，語用論的意味の扱いをどうするかを決めておく必要がある。例えば「漱石を読んだ。」という場合の「漱石」は漱石の作品をいう意味であるから，人名に対応する分類番号ではなく，作品に対応する分類番号を付与すべきかという問題である[7]。

分類番号は，意味の中でも語彙的意味（辞書的な意味）を捉えたものである。従って，語彙的意味以外の意味については分類番号以外の属性を作り，それで表すことが必要である。例えば，文体情報，プラス，マイナスなどの評価にかかわる属性，コノテーションの情報などは，分類語彙表番号のシステムとは違うアノテーション方式がよいのではないか。また，「メリケン粉」と「小麦粉」，「スチュワーデス」と「キャビンアテンダント」のような語の

---

6 『分類語彙表』には 3.1960 は存在しないが，この番号が妥当なものと考えられる。

7 この例では，「漱石」に作品の意味を与えてもよいかもしれないが，遅刻してきた人に向かって「ずいぶん早いね」という皮肉を言った場合，「早い」に「遅い」に対応する分類番号を与えることには抵抗があろう。

新旧や言い換えの関係なども分類番号には反映しにくいため，別の形での記述が必要となるだろう。

謝　辞

本研究は，JSPS 科研費 JP19K00655 の助成を受けたものです。

文　献

垣内松三（1938）「語彙の體系的分類」『國民言語文化體系・第三巻　基礎語彙學（上)』文學社.

土居光知（1933)『基礎日本語』六星館.

田島毓堂（2001）コード付けの基準,「名古屋大学文学部研究論集文学」47.

山崎誠，柏野和佳子（2017)『分類語彙表』の多義語に対する代表義情報のアノテーション,「言語処理学会第 23 回年次大会発表論文集」302-305.

資　料

短単位語数 Excel データ(Version 1.1)

https://pj.ninjal.ac.jp/corpus_center/bccwj/bcc-chu-suw.html　（2020 年 8 月 10 日閲覧）

分類語彙表番号－UniDic 語彙素番号対応表（wlsp2unidic）

https://github.com/masayu-a/wlsp2unidic