様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月 ２４ 日現在

機関番号：６２６１８
研究種目：基盤研究(C)
研究期間：2012～2014
課題番号：２４５２０５２０

研究課題名（和文）語彙分類の理論的整備に基づくシソーラスの改良に関する研究

研究課題名（英文）Study on the revision of thesaurus based on the theoretical improvement to the vocabulary classification

研究代表者
山崎　誠（Yamazaki, Makoto）

大学共同利用機関法人人間文化研究機構国立国語研究所・言語資源研究系・准教授

研究者番号：３０１８２４８９

交付決定額（研究期間全体）：（直接経費）　　3,800,000円

研究成果の概要（和文）：日本語研究におけるシソーラスのより一層の活用を図るため、人文系日本語研究者の間でもっとも普及している『分類語彙表増補改訂版』に研究に有益な情報を付与する作業を行った。多義語として複数の分類項目に出現している見出し語２７１７１語について、一定の基準に基づいて「代表義」を１つ決定し、その情報を付与した。作業結果は、２０１５年７月を目指してウェブ上で公開する予定である。これにより、意味解析上の精度が向上し、異なる分析結果の間の適切な比較が可能になることが期待される。また、旧版の分類語彙表との異動の比較を行い、結果の一部を「語彙研究」12号に発表した。

研究成果の概要（英文）：In order to promote the use of thesaurus in the Japanese linguistics, we annotated the useful information to the "Bunrui Goihyo (Word List by Semantic Principle, revised and enlarged edition)" which is the most popular thesaurus among Japanese linguists. Based on our original criteria, we chose the "representative meaning" of the 27,171 polysemic words which appears more than one semantic cluster in Bunrui Goihyo.The annotated thesaurus is due to be released in July 2015 on the web. Using this thesaurus, it will be possible to improve the precision of semantic analysis and make appropriate comparison between different analyses. We also examined the differences between old and new versions of Bunrui Goihyo. The result was reported in the Journal of the Study of Vocabulary Vol.12.

研究分野：日本語学

キーワード：シソーラス　語彙体系　分類語彙表

１．研究開始当初の背景

語彙研究のための基礎的な概念である語彙分類や語彙体系は、語彙の記述や比較には不可欠なものである。しかし、語彙分類の方法についての十分な研究は行われてこなかった。語彙を構成する語に備わっている属性、とりわけ、意味についてどのようなアプローチを採ればよいか、その点が吟味されないまま、直観をたよりに語彙分類が行われてきた経緯がある。

最近では、『現代日本語書き言葉均衡コーパス』（BCCWJ)をはじめ、電子化された大量の言語データ（コーパス）の利用が研究者に身近になり、語彙研究も行いやすくなってきた。ただ、語彙の記述や比較のための重要なツールであるシソーラスは、以下のような問題点があり、研究のニーズに十分に応えることができない状態である。

シソーラス利用上の問題点は以下のように整理することができる。

○技術的な問題点

・利用環境の問題：利用のためのツールが開発されていないため、意味情報の付与を手作業で行うか、あるいは、独自にプログラミングを行う必要がある。前者は非効率であり、後者はコンピュータに不慣れな多くの人文系研究者にとってハードルが高い。

・表記の問題：各見出し語に対してひとつの語形とひとつの表記しか与えられていないため、他の言語データとの照合（マッチング）を行う際、語形ないし表記が不一致のために照合に失敗する場合がある。

○内容的な問題点

・多義語の問題：シソーラスにおいて多義語は、複数の分類項目が並立的に示され、そのうちどの分類項目が当てはまるかを利用者がいちいち判断しなければならない。また、分類の仕方が背反的でなく、ひとつに決めがたい場合がある。

・未収録語の問題：収録されていない語に対して利用者が分類項目を指定しようとする場合、どのような手順・方法で所属する分類項目を与えてよいか判断が難しい。分類原理及び分類項目の概念規定が明文化されていないためである。

・編纂方針の問題：『分類語彙表』には人名・地名以外の固有名詞や専門用語がほとんど収録されていない。情報検索でさかんに利用される実用的なシソーラスとの接点が少ないため、学際的研究への発展につながらない。

・関連分野への応用の問題：日本語教育、古典、対照研究等関連分野での活用に関する観点が欠けているため、自分で加工しないと利用できない。

２．研究の目的

本研究では、１で挙げた問題点のうち、内容的な問題点のうち、多義語の問題を解決することを目的とする。

また、長年使われてきたシソーラスである『分類語彙表』の旧版と新版でどのような違いが生じているかについても調査した。

３．研究の方法

3.1　問題点と方法

『分類語彙表増補改訂版』（以下,『分類語彙表』と略す）では多義の語をそれぞれの分類項目に配置しているため、延べ語数で約 30%、異なり語数で約 17%の語が複数の分類項目に出現している。

しかし、『分類語彙表』を使って分析を行う際に複数ある分類項目のうちのどれを選んだらよいか適切な指針がない現状では、分析結果の妥当性や他の分析データとの比較に問題が生じる。

そこで、分類語彙表増補改訂版において、複数の分類項目に出現している語をすべて抜き出し、一定の作業基準に基づいて、そのうちどれか一つを代表義とする。

3.2　データ

作業対象となる見出し語の語数は以下のとおり。

	体	用	相	他
延べ	16555	8872	3918	310
異なり	8070	3915	2391	228

3.3　作業基準

基本義の決定が恣意的にならないように試行と合議を経て，以下の作業基準を作成した。これらには一定の優先順位を設け，基本義がなるべく客観的に決まるようにしている。

(1)　高頻度

使用頻度が高いものを基本義とする。直観的に使用頻度に差がありそうなものは内省で判断したが、コーパスでの頻度も参考にした。例えば,「切り出す」は、＜2.1571 切断＞、＜2.3100 言語活動＞、＜2.3810 農業・林業＞の 3 箇所に現れるが、『現代日本語書き言葉均衡コーパス』における語義の分布では、＜2.3100 言語活動＞が約 8 割を占める。

(2)　具体性（物理的）

抽象的なものより具体性（物理的）があるものを優先する。例えば、「傾く」は、＜2.1513 固定・傾き・転倒など＞と＜2.1583 進歩・衰退＞の２つの分類項目に現れるが、具体的な意味を持つ 2.1513 のほうを優先する。(1)のの高頻度との基準と具体性の基準とが両立しない場合がある。例えば、「あぶり出す」は、＜2.1210 出没＞と＜2.3842 炊事・調理＞の２箇所に出ているが、具体性を重視すれば「炊事・調理」であるが、この意味ではほとんど使われないと思われるため、高頻度である「出没」を選択した.

(3)　分類項目名と一致

分類項目と一致しているものがあれば、その分類項目を選ぶ。例えば、「上がる」は、以下の８つの分類項目に出現している。＜2.1540 上がり・下がり＞、＜2.1503 終了・中止・停止＞、＜2.1580 増減・補充＞、＜2.1651 終始＞、＜2.3000 心＞、＜2.3331 食

生活＞、＜2.3520 応接・送迎＞、＜2.3700 取得＞。これらの中で分類項目に「上がり」がある 2.1540 を選んだ。

(4)　類義語

同段落に類義があるものを選ぶ。『分類語彙表』は、分類項目に対応する意味領域の語を配列しているのであるが、その意味領域に対して中核的な語と周辺的な語がある。周辺的な語はどちらかというと、その回りに類義語が少なく、グループを形成していない場合がある。例えば、「意見する」の場合、＜2.3061 思考・意見・疑い＞の場合、「意見する」の所属する段落には「意見／考えがある、含むところがある、下心がある」が並んでいるが、＜2.3640 教育・養成＞には、「忠告する、忠言する、諷諫する、注意する」などが並んでいる。「意見する」の類義語のグループとしては後者がふさわしいため、後者の分類項目を選ぶ。

(5)　慣用句

『分類語彙表』には、慣用句も相当数収録されている。慣用句と字義通りの意味の両方がある場合は慣用句として意味を優先する。例えば「足をすくう／足をすくわれる」は、＜2.3392 手足の動作＞と＜2.3683 脅迫・中傷・愚弄など＞の２箇所にあるが、前者は字義どおりの意味、後者は慣用句としての意味である。このような場合は、慣用句の意味のほうを基本義とした。

(6)　辞書の第一義と一致

作業用に用いた『岩波国語辞典第５版』(以下、『岩国』と略す）の第一義と一致するものを選択。「しゃれる」は、＜2.3030 表情・態度＞と＜2.3332 衣生活＞の２箇所に出現しているが具体性や頻度では決め手を欠き、類義語もそれぞれ存在するため、『岩国』の第一義を優先することにした。「しゃれる」の語釈は以下のとおりである。「①気がきいている。あかぬけている。「―・れた家」。転じて物事に通じているような風(ふう)をする。なまいきな様子がみえる。「―・れたことをぬかすな」②しゃれ(1)を言う。③身なりを飾る。めかす。

(7)　複合語（後項）

複合語は原則として後項要素の意味を選ぶ。「追い出す」は＜2.1525 連れ・導き・追い・逃げなど＞と＜2.1531 出・出し＞の２つの分類項目にあるが、動作として「追う」よりも「出す」ほうに重点があると考え後者を選んだ。

ただし、前項要素の意味を選んだものもある。「泳ぎ回る」は、＜2.1523 巡回など＞と＜2.3374 スポーツ＞の２箇所に出現するが、回る動作よりもスポーツのほうを優先した。

(8)　段落出現順

段落内でより先頭に近いものを選ぶ。『分類語彙表』は分類項目がいくつかの段落から構成されている。その段落内での語の配列は、「なるべく意味・用法の広いほうから狭いほうへ配列しているが、必ずしも厳密ではない。」(『分類語彙表』P.4）とされている。そこで、段落内で前のほうに出現している語を基本義と考える基準を立てた。例えば、「こぼれる」は、＜2.1531・出し＞と＜2.1540 上がり・下がり＞の２か所に出現するが、それぞれの位置は＜出・出し＞では、5 行目（11 語め)、＜上がり・下がり＞では、1 行目（1 語め）になっているため、後者を優先した。

2.1531　出・出し

＜中略＞

02　　出る　出て来る
　　出過ぎる
　　はみでる　はみだす　あふれ出る　あふれ出す
　　出っ張る　出張る　突き出る
　　こぼれる　居こぼれる

2.1540　上がり・下がり

＜中略＞

16　　こぼれる　こぼす
　　落ちこぼれる　吹きこぼれる

(9)　その他

語によっては、より汎用的であると認められる語義を優先したものがある。「一蹴する」は、＜2.3133 会議・論議＞と＜2.3532 賛否＞の２箇所に現れるが前者は言語行為としての意味、後者は拒絶する意味であり、後者のほうがより基本義的と判断した。

基本義を決定する際には，次のように各基準の優先順を付けた．優先する順に，

高頻度＞分類項目名＞慣用句＞類義語＞岩国語義＞段落出現順

3.4　作業上の問題点

作業を行った結果、以下の問題点が明らかになった。

(1)語の同定

『分類語彙表』における語の特定の基準が曖昧なため、多義語か同音異義語かの区別を改めて行わなければならなかった。今回の作業では、まず見出しの語形とその読みとで機械的に語を特定したが、それでは語の同定に問題があるものがあった。3.3(3)に挙げた「上がる」の例は、見出しの形が「上がる」となっているものの中には基本義にふさわしいものがなかった。実は、「あがる」の形で別に立項されていた見出し語があり、それが基本義として最適であった。

また、表記により語を区別したほうがよいと思われるものがあった。例えば、「あがなう」は「買い求める」意味の＜2.3761 売買＞と「つぐなう」意味の＜2.3780 貸借＞があり、これらは表記上，「購う」「贖う」と書き分けられ、別語としている辞書もあることから、このようなものは多義語とはしなかった。

(2)　適切な分類項目の不在

「いかれる」は、「あたまの働きなどがまともでなくなる.」(『岩国』では②の意味）が基本義と思われるが、それに該当する分類項目がない。＜2.1583 進歩・衰退＞、＜2.3040 信念・努力・忍耐＞、＜2.3683 脅迫・中傷・愚弄など＞の中では 2.3040 がいちばん近い

ようであるが、同じ段落にある類義語は「惑溺する、耽溺する、おぼれる」であり、これは「人や物に心を奪われ，夢中になる.」(『スーパー大辞林 3.0』の③）に該当する。
(3)　品詞の違いの処理
　体の類（名詞）については、相の類（形容動詞）と重複して出現しているものが異なりで 1,269 語ある。このような語は意味的にはほぼ同一と考えられるが、本研究での多義に該当するため、どちらかに決定しなければならならず、そのための基準が必要になる。
3.5　新版と旧版の比較
　それぞれの電子データを 1 語ずつ比較し、異同の程度別に整理した。

４．研究成果
　作業結果を付与したデータは、2015 年 7 月にウェブ上で公開する予定である。また、中間報告として 2014 年度の計量国語学会大会（於東洋大学、2014 年 9 月 20 日）で発表を行った。
　新版と旧版の『分類語彙表』の異同については、「語彙研究」12 号に掲載するとともに、2015 年語彙研究会特別大会（於台湾大学、2015 年 3 月 7 日）で発表した。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 6 件）
1.田嶋毓堂（2015）『分類語彙表『分類語彙表』元版と新版のコード比較(中間報告),「語彙研究」12，pp.1-8.

2.山元啓史・村井源・ボル ホドシチェク（2014）二十一代集シソーラスのための漸近的語彙対応システムの開発,「じんもんこんシンポジウム 2014，人文科学とコンピュータシンポジウム論文集」2014-3, pp.157-162.

3.内山清子(2014)漫画の台詞やオノマトペにおける言語的特徴分析,「湘南工科大学紀要」48,pp.63-68.

4.山崎誠(2014)語彙研究に足りないもの－語彙研究の発展のために－,「語彙研究」11,pp.18-25.

5.田島毓堂(2014)意味分野別構造分析法の簡易化と電子辞書の増補・公開,「語彙研究」11，pp.26-39.

6.田島毓堂(2012)意味分野別語彙構造分析法における意味コードの使用法及び分類枠組についての提案：単語コードと語素コードによる分析(3)承前,「人間文化：愛知学院大学人間文化研究所紀要」27，pp.67-102.

〔学会発表〕（計 6 件）
1.田島毓堂, 国立国語研究所編『分類語彙表』元版と新版のコードの比較，語彙研究会特別大会（2015 年 3 月 7 日, 台北市・台湾大学（台湾））

2.山崎誠・柏野和佳子・内山清子・砂岡和子・田島毓堂・山元啓史・韓有錫・薛根洙,『分類語彙増補改定版』へのアノテーション－基本義の決定－，計量国語学会第 58 回大会（2014 年 9 月 20 日，東洋大学白山キャンパス（東京都文京区））

3.田島毓堂,『分類語彙表』元版と新版のコード比較（中間報告），語彙研究会大会（2014 年 9 月 6 日，駒澤大学深沢キャンパス（東京都世田谷区））

4.田島毓堂,『分類語彙表』元版・新版の比較の見通し，語彙研究会第 98 回例会（2013 年 12 月 21 日，愛知学院大学大学院栄サテライト（愛知県名古屋市））

5.山崎誠，共起語集合の頻度分布と語の属性との相関，第 3 回コーパス日本語学ワークショップ（2013 年 2 月 28 日，国立国語研究所（東京都立川市））

6.内山清子，論文の論理構造における分野基礎用語に関する分析，第 2 回コーパス日本語学ワークショップ（2012 年 9 月 7 日，国立国語研究所（東京都立川市））

６．研究組織
(1)研究代表者
山崎　誠（YAMAZAKI，Makoto）
大学共同利用機関法人人間文化研究機構
国立国語研究所・言語資源研究系・准教授
研究者番号：３０１８２４８９

(2)研究分担者
柏野　和佳子（KASHINO，Wakako）
大学共同利用機関法人人間文化研究機構
国立国語研究所・言語資源研究系・准教授
研究者番号：５０３１１１４７

田嶋　毓堂（TAJIMA，Ikudo）
名古屋大学名誉教授
研究者番号：２００８２３４９

山元　啓史（YAMAMOTO，Hirohumi）
東京工業大学留学生センター・准教授
研究者番号：３０２４１７５６

内山　清子（UCHIYAMA，Kiyoko）
湘南工科大学・工学部・准教授
研究者番号：２０４５８９７０

(3)連携研究者
砂岡　和子（SUNAOKA，Kazuko）

早稲田大学・政治経済学術院・教授
研究者番号：７０２５７２８６

(4)研究協力者

薛　根洙（SEOL, Guen-Soo）
全北大学校（韓国）・人文大学日語文学科・教授

韓　有錫（HAN, Yu-Suk）
東新大学校（韓国）・観光日本語学科・教授