

*Journal of Japan Society of Information and Knowledge*

# 情報知識学会誌

Vol.9 No.4 (Feb. 2000)

~~~~~ 目次 ~~~~~

## 特集 「シソーラスー領域知識の体系化」

|        |                                                   |    |
|--------|---------------------------------------------------|----|
| 巻頭言    | 特集 「シソーラスー領域知識の体系化」に当たって…石川徹也…                    | 1  |
| 寄稿論文   | Thesaurus 管見……………寺澤 芳雄…                           | 3  |
| 寄稿論文   | 情報処理研究とターミノロジーから見た『分類語彙表』<br>柏野 和佳子, 中野 洋, 石井 正彦… | 12 |
| 技術報告   | シソーラスと文献データベース……………永井 賢吉, 原田 郁子…                  | 29 |
| 論文     | 意味関係抽出手法統合による概念の体系化<br>伊東 千夏, 宇陀 則彦, 石塚 英弘, 藤原 譲… | 38 |
| 抄録     | ……………                                             | 49 |
| 書評     | ……………                                             | 50 |
| 用語解説   | ……………                                             | 52 |
| お知らせ   | ……………                                             | 53 |
| 投稿の手引き | ……………                                             | 55 |

~~~~~



情報知識学会

## 巻頭言

## 特集 「シソーラス—領域知識の体系化」に当たって

石川徹也

「シソーラス特集」編集委員長、図書館情報大学

日本における例えば「魚」に関する語彙は”豊富である”のに対して「肉」に関する語彙は”貧困である”。対して西欧においては全く逆であることはよく知られている。同じく「言葉」に関する語彙は日本においては貧困であり、西欧においては豊富である。事実、我が国において例えば、“語 (word)”、“用語 (term)”、“語彙 (vocabulary)”、さらには”辞書 (dictionary)”、“用語集 (glossary)”、“類義語集 (thesaurus)”等の差違の認識は一般の生活の中において依然無い。この訳を例えば木村尚三郎の著書「ヨーロッパからの発想」(講談社、昭和58年)の中で知ることができる。当起因は「覇権主義」に基づく「言語覇権現象」と理解できる。

「言語覇権現象」は特に西欧において多くの苦悩を生み出してきた。一方、当苦悩の解決に向け、“バベルの塔”を夢見て多くの試みと成果が生まれた。例えば、その一つに辞書学 (Lexicography) の発展があり、もう一つは普遍言語 (Universal language) の構築がある。辞書学はシソーラスの原典ともいわれる Roget, P.M. による「Thesaurus of English Words and Phrases」(初版:1852年)を生み、普遍言語の構築の試みは19世紀ウィーン学派において発芽し、現代科学の基礎をなす論理学に結実し、また Zamemof, L.L. による 에스ペラント語 (Esperanto, 1887年)を生んだ。これらの成果は現在尚発展している。Roget's Thesaurusは昨年(1998年)11年ぶりに改訂された。論理学は Chomsky, N. による形式文法 (Formal grammar, 1956年)を生み、現在の計算機および自然言語処理の発展をもたらし、エスペラント語に関してはその拡張と普及のために毎年世界大会が開催されている。

しかし、これらの成果にも係わらず、“民

族のアイデンティティ=固有言語種”の公式の下に、宗教・文化がもたらす紛争と並び、現在尚多くの苦悩が続いている。アルザス語民族、バスク語民族、ベルギーにおけるオランダ語系フラマン人とフランス語系ワロン人の軋轢、最近ではアルジェリアのベルベル人の反発等枚挙を見ない。

昭和40年代後半から出超国となった日本に、特にアメリカ政府は貿易摩擦回避の一貫として、我が国の科学技術関係の情報(特に、論文・特許等の技術資料)の英語での発信を求めた。このことに対して日本政府(「科学技術会議」)は1981年度から「日英科学技術文献の速報システムに関する研究」を科学技術振興調整費にて実施することを決定し、1985年までの4年間に渡り、日英機械翻訳の研究開発 (Mu - Project) が実施された。この成果を下に「JOIS-DB」の翻訳が今でも続いているが、昨今の Internet の普及に伴い、世界は再び我が国の Internet 上の情報に、とまどいと苛立ちを持ち始めている。英語を解読できるが容易に発信できない我々の Website の情報はほとんどが日本語である。言語種の違いによる情報流通の困難性が何をもたらすのか、日々体験しているが、初等教育の段階から英語教育を取り入れるべきであるとの議論があるものの、今、具体的な解決策を見いだせない我々は、同じく苦悩する”多言語国:EU”の努力を参考にする必要があると考えます。

EUは民族・言語・文化・政治風土の違い、さらに経済格差、過去の支配関係に起因する反目を越え”Global Village”として統合に向け進んでいる。この困難を乗り越え統合へ向ける心は、各国独立ではもはや”食べていけない”との危機感があるからであろう。し

かし、通貨統合がスタートしたものの、常に“崩壊への危うさ”があるという。その唯一の問題は、“コミュニケーションの難易問題であり、言語統一は不可能である”ということである<sup>[1]</sup>。ここ数年、年3、4回の頻度で渡欧する中で、当問題に起因する彼らの危機感を我が身のように感じさせられている。

EU15カ国において派生語を含め45言語あるという<sup>[2]</sup>。現在の公用語は11言語であるが、1997年7月にNATOに加盟したポーランド、チェコ、ハンガリー3カ国の例に見られるように、東欧諸国の加盟予定に伴い16言語が公用語になるという。この状況に対処すべくEU委員会は、“From the Tower of Babel to the Global Village”を標語に、“Language Insustry”を推進している<sup>[3]</sup>。その一例に、ECの時代から多言語機械翻訳システム:EUROTRAプロジェクトの研究開発、専門用語多言語対訳辞書DB:EURODICTAMの構築を進め、EU発足に伴い各国の生活・社会・産業関係の情報のDB化:ECHOプロジェクトを、さらに言語資源を統合し利用可能とするERLAプロジェクトを推進している。そして2000年を「EU言語年」と位置づけ約800万ユーロを投じEU人のマルチリンガル化策を打ち出した。

我々日本人は世界的には“言語(コミュニケーション)弱者”である。GDP第2位を誇る我が国として、必要ならば“日本語を学んで来い!”との気概も大切であるが、決して通用するものではない。対して国策は皆無に等しい。言語取得は基本的には個人の努力に帰着する問題であるが、EUの策を見るに“これで良いのか?”の疑念を抱く。今年になって音声・自然言語処理関係の研究・開発関係にたずさわる有志によって「言語資源共有機構(GSK)」が発足し、研究利用を目的に言語資源の共有化を進めだしたところである<sup>[4]</sup>。

未来学者でもあったArther C. Clarkeは著書「Profiles of the future, 1967」の中で、2020年代には「Logical language」が発明され使われるようになり、「Robots」とのコミュニケ

ーションが可能になると予測している。昨今の音声対話型機械翻訳システムの研究開発成果を見るに、当予測の実現を予感できるが、その過程において当学会の開与(貢献)もなければならぬのであろう。以上の理解の下に、「専門用語研究会」との関連も深い当学会(誌)において、言語資源の代表的な一つである「シソーラス」を通し考える機会を持つために、この度の特集を組んだ次第である。寄稿論文として、昨年11年ぶりに改訂されたRoget's Thesaurusに関して岐阜女子大学の寺澤芳雄先生に解説をいただき、わが国における代表的なシソーラスである国立国語研究所編纂の「分類語彙表」に関して元主任研究員、現大阪大学の石井先生にご寄稿をいただいた。また科学技術振興財団におけるDB事業の中心をなす「科学技術シソーラス」に関し永井氏に解説いただいた。投稿採録論文は1件である。寄稿・投稿いただいた方々に感謝する次第である。

尚、石井先生の前稿の共著者は柏野和佳子、中野洋さんである。当誌の編集途上にある昨年12月21日に中野洋さん(国立国語研究所)の悲しくも訃報が飛び込んできた。中野さんは「分類語彙表」の構築に中心を成した方である。「分類語彙表」は、特に我が国の自然言語処理研究になくてはならないツールである。これからも尚ご活躍いただきましたのになんとも残念でならない。中野さんご冥福をお祈りします。

## 参考文献

- [1] EC: The Multilingual Informtion Society(COM(95)486 final), 34p., 1995.11.
- [2] Hearn, P.M. and Button, D.F., Ed.: Language Industries Atlas, IOS Press, 406p.1994.
- [3] EC: Language and Technology, 26p.1996.
- [4] <http://tanaka-www.ca.titech.ac.jp/gsk/>

# Thesaurus 管見

## Thesaurus - Past and Present

寺澤 芳雄

岐阜女子大学 文学部

まず、中世から現代に至る thesaurus の発達を Roget's Thesaurus を中心に概観する。ついで、Roget's Thesaurus が近代科学発展期のキーワード「分類・組織」によって観念を分類・体系化し、これに基づいて英語の語彙組織を確立しようと試みたものであること、その際 Roget が F.Bacon, J.A.Comenius, J.Wilkins 等の構想から影響を受けたことを明らかにした。さらに、Roget's Thesaurus がその後の時代の要請に応え、どのような変容受けつつあるか、またその構想を応用した各種概念別分類辞典を取り上げて、thesaurus の今後の可能性を考察した。

### 1. Thesaurus とは何か

今日、一般に thesaurus というとき Roget's Thesaurus を連想することが多いが、では「thesaurus とは何か」ということになると、答は必ずしも明確ではないようである。この語の語源は、語形からも明らかのように、ギリシア語・ラテン語に由来する借入語で、結局は「宝物」の treasure と同一語源に遡る。そのことから窺えるように、thesaurus の本来の意味は「(神殿などの)宝物蔵」であった。そこから、まず比喩的に「知識の宝庫」の意が、ついで、「百科事典」や網羅的な「辞典」の意が生じた。「知識の宝庫」の用法は、Roger Lax & Frederick Smith 編の音楽宝典 *The Great Song Thesaurus*(Oxf.U.P., 1989<sup>2</sup>) など、最近の出版物の書名にも見られる。また「辞典」の例としては、英国における初期の羅英辞典 Thomas Cooper, *Thesaurus Linguae Romanae et Britannicae*(1565) や Robert Ainsworth, *Thesaurus Linguae Latinae compendarius; or ... Dictionary of the Latin Tongue* (1736) などがある。次に、単語や語句を集めた辞典ではあるが、見出し語をアルファベット順配列ではなく、意味すなわち語の表わす観念に従って配列した「概念分類辞典」「主題別分類類語辞典」の意味が発達する。1852年出版の Peter Mark Roget, *Thesaurus of English*

*Words and Phrases classified and arranged* の題名に見られるものが、この用法の最初の例であり、はじめにのべたとおり、今日狭義での thesaurus は、この Roget's Thesaurus あるいはこれをモデルにした類義語辞典を意味している。さらに thesaurus には、「(学問分野別)分類術語集」とか、最近ではコンピュータの情報検索用の標準言語である「シソーラス」などの用法も見られる。人文科学方面では、Oxford 大学出版局が最近出版された、*The Getty Art History Information Program Art and Architecture Thesaurus*(1990; 補遺 1992-) などの書名に見られる thesaurus は、コンピュータ用語の例であろう。

辞書の発達史に照らしても、thesaurus という語は、Roget 以前はある言語の(網羅的な)辞典すなわち dictionary の意味で使われることが多かったのである。上に Cooper, Ainsworth の例をあげたが、かつては Thesaurus といえは、Robert Estienne 父子が編集したアルファベット順配列の古典語辞典 *Thesaurus Linguae Latinae*(1532) および *Thesaurus Linguae Graecae*(1572) を指していた。最近でも Massimo Pallottino, *Thesaurus Linguae Etruscae* (1978-) などのような古代語(エトルリア語)の場合にも、dictionary と実質的区别なく用いられていることが少なくない。従って、the-

saurosを一義的に Roget 式概念別分類辞典の意味に限るのは、歴史的にも論理的にも根拠に乏しいといわなければならない。一方、概念別分類類語辞典を表わす語としては、thesaurusに限らず、promptorium(「貯蔵庫」)、(h)ortus vocabulorum(「ことばの苑」)、repository(「倉庫」)などが用いられたことがあり、また dictionary をこの意味で用いたものには、Robert Estienne, *Dictionnaire français-latin*(1539) や、John Withals, *Shorte Dictionarie for Yonge Begynners*(1553) などがあつた。

ところで辞書の歴史を辿ってみると、今日では常識的なアルファベット順配列の辞書の発達は意外に遅く、一般化するのには16世紀になってからであり、古代・中世の語彙集では主題別のほうがむしろ通例であつたようである。英国では、エンシャム修道院長で、文法学者としても知られるÆlfric(?955-?1020)のラテン語類語集では、農耕、政治、獣類名、昆虫名から航海用法に至る15の主題項目に分類している。また、13世紀半には、薬草をはじめとした植物名類語集(BL, MS Harley 978), 15世紀には羅英類語集(BL, MS Reg.17C xvii)で主題別に整理し、約1500語を収めたものや、32の主題項目を立て、欄外に挿絵を添えた図解語彙集(Lord Londesborough 所蔵写本)などが見られる。最後のものは現代のDudenの先駆とみることもできよう。

## 2. Roget と Roget's Thesaurus

ここで、Roget およびその Thesaurus について大要を記しておきたい。Peter Mark Roget(1779-1869)は、スイス・フランス系の英国の医者で、生理学をはじめ数学・光学・化学などの分野で業績を上げ、英国王立協会会員として、多方面の活躍をしていたが、70才を迎え学会関係の公職を退き自由の身になってから、僅か3-4年の間に完成したのが有名な *Thesaurus of English Words and Phrases, classified and arranged so as to facilitate the expressions of ideas and assist in literary composition* である。その初版がロングマン社

(Longman, Brown, Green, and Longmans) から出版されたのは、1852年。初版は僅か1000部、翌年の第2版および1855年の第3版も1500部に過ぎなかったが、父から息子、息子から孫へとロジェ家3代にわたる改訂を中心に、今日まで1世紀半たらずの間に出版された部数は3200万に及ぶという。聖書を別にすれば、一般書と辞書とを問わず驚異的な数であり、このこと自体が Roget's Thesaurus の並外れた有用性を端的に物語っているといえよう。(以下、Roget's Thesaurus に言及する場合、とくに断らない限り、便宜的にその最新版(Penguin Books 1998)によることにする。)

Roget's Thesaurus では、通例の類義語辞典とは異なり、人間・自然界などに関する諸概念を動・植物学などの分類組織(門・綱・目)にならって、6「分野」、39「部門」、990「見出し項目」(初版では1000見出し)という三層構造に分類記述し、包攝関係に基づく樹状図、意味論的な秩序立てによって、ある概念が普遍的な概念組織の中でどのような位置を占めるかを明らかにしている。すなわち、単語の連想関係を軸として人間の思考のモデルとなる構造を明らかにしようとした「主題別分類類義語辞典」ということができる。Roget 自身は初版の序文で次のように述べている。

普通の辞書の目的は単に語の意味を説明することにある。....本書で意図した目的はまさにその逆である。すなわち、一定の概念に対して、その概念を最も適切に表現するような語句を見つけることである。この目的のために、英語の語句はその音や綴り字に従ってではなく、その意味に従って分類配列されている。

次に、Roget の分類大綱を第2版(1853)によって次頁に示す。この分類は、第6分野の Affections が現行版で Emotion, religion and morality となっている点を除けば、初版以来今日までほぼ不変である。

PLAN OF CLASSIFICATION.

	Sect.		nos.	
I. ABSTRACT RELATIONS	1. EXISTENCE	- -	1 to	8
	2. RELATION	- -	9 -	24
	3. QUANTITY	- -	25 -	57
	4. ORDER	- -	58 -	83
	5. NUMBER	- -	84 -	105
	6. TIME	- -	106 -	139
	7. CHANGE	- -	140 -	152
	8. CAUSATION	- -	153 -	179
II. SPACE	1. GENERALLY	- -	180 -	191
	2. DIMENSIONS	- -	192 -	239
	3. FORM-	- -	240 -	263
	4. MOTION-	- -	264 -	315
III. MATTER	1. GENERALLY	- -	316 -	320
	2. INORGANIC	- -	321 -	356
	3. ORGANIC	- -	357 -	449
IV. INTELLECT	1. FORMATION OF IDEAS		450 -	515
	2. COMMUNICATION OF IDEAS		516 -	599
V. VOLITION	1. INDIVIDUAL	- -	600 -	736
	2. INTERSOCIAL	- -	737 -	819
VI. AFFECTIONS	1. GENERALLY	- -	820 -	826
	2. PERSONAL	- -	827 -	887
	3. SYMPATHETIC	- -	888 -	921
	4. MORAL	- -	922 -	975
	5. RELIGIOUS	- -	976 -	1000

### 3. Roget's Thesaurus の成立

Roget's Thesaurus における概念の分類・組織化という基本構想に直接的あるいは間接的影響を与えたものとして、近代的科学・哲学の基礎を築いた Francis Bacon(1561-1626)、モラヴィアの神学者・教育家 J.A. Comenius(1592-1670)、英国王立協会の創設会員の一人であった哲学者 John Wilkins(1614-72) を挙げることができよう。Bacon は、*Instauratio Magna* 『大革新』(1620) において人間の知識のすべてを再構成し、知識の体系化を試みるべく、外的自然・人間・外的自然への人間の働きかけという三大領域から、次々に下部領域への細分化を行ない、全体で 130 部門からなる知識の階層構造のモデルを構築した。Comenius は、*Janua Linguarum Reserata* 『諸言語の開かれた扉』(1631) において、世界の起源、元素・天体、動物、人間と人間の身体、感覚、

心・意志・感情から神・天使に至る 20 の主題に基づく百の区分を試み、また Wilkins は *An Essay Towards a Real Character and a Philosophical Language* 『真正文字および哲学的言語論』(1668) の第 2 部の中で、おそらく Bacon や Comenius の範疇を参照した上で、普遍的概念のプラトンの一覧表を提示した。これを Tom McArchur(1986) の要約を参考にして示すと次のとおりである。

1. 神
2. 元素・流星・石・金属
3. 植物・草木・花
4. 動物・魚・鳥
5. 身体部位
6. 量・大きさ・広さ・尺度
7. (自然力、習性、慣習、感覚、病気などの) 質

8. (精神的、肉体的、運行などの) 行動
9. (家事・家庭の) 関係
10. (民事、司法、軍事、教会などの)  
公的關係

Wilkins で注目すべきは、宇宙大綱の範疇にコードによる索引がつけられていること、また Roget がその Thesaurus の発想を Wilkins の「概念別辞典」の提案に負っていることである。

ところで、Roget が Thesaurus の編集・執筆に本格的に取り組んだのは、前に記したとおり、公職辞任後の数年間であったが、その母胎は 26 歳の頃から折にふれて書き留めた、40 数年にわたるメモである。当時の学者の常として、ラテン語のほか二、三のヨーロッパの言語に通じてはいたが、とくに言語学的素養をもたなかった Roget をして、あえてこの言語学的著作に駆り立てたものは何であったろうか。これは、生涯を通じ、公私の生活両面において、目の前にあるありとあらゆる事物をすべて分類・整理し、秩序ある組織に統合しなければ済まないという Roget の気質によるところが多かったのではないかと思われる。事実、Roget の生理学を始めとする研究は、主として分析したデータの分類・組織化にほかならず、一方 22 年間にわたり王立協会の書記として事務処理に当り、協会や学会の蔵書の整理に成果をあげたといわれている。しかしまた同時に、18 世紀後半から 19 世紀初頭における科学の進歩—地質学・生物学・化学・物理学などの諸分野に見られた長足の進歩が、結局、分類学に基づくものとするれば、Roget の Thesaurus はこの時代の申し子の一つであったということもできよう。ほかに二、三の例を挙げるなら、Karl Linné(1707-78) の植物分類法、D.I. Mendeleev(1834-1907) の元素の周期率の発見 (1869)、あるいは Jeremy Bentham(1748-1832) の人間行動の動機の分類など、この近代科学の発展期のキーワードはまさに「分類・組織」であった。Roget は、生涯にわたって Bentham の実利主義 (utilitarianism) の影響を受けたともいわれる。

#### 4. Roget's Thesaurus の具体例

ここで少し具体的な説明を現行版によって加えると、例えば、ふつうの英語辞書では dictionary は D の部、thesaurus は T の部のように、遠く離れた場所に記載される。これに対して、Roget では、第四部門 Intellect の "Means of communicating ideas" という分野の中にある 559 番の見出し語 Word の項に、dictionary と同種概念領域に属する lexicon, glossary, word list, vocabulary, thesaurus, wordhoard, concordance, index などが纏めて列記され、さらに類似の、あるいは連想関係にある概念領域の philology, linguistics, lexicology, lexicographer, compiler などが参照できるようになっているのである。

原著者の Roget が期待したこの辞書の引き方は、まず巻頭の分野別一覧表によって、自分の求める概念 (意味) の所属領域を同定し、その概念の項に集められているリストの中から適切な語句を見出すことである。たとえば、喜びの感情を表わす語を求める場合、一覧表で第六部門 Emotion, religion and morality の中の第二分野 Personal emotion に焦点を絞ると、そこに 824 Joy という見出し項目が見つかる。そこで本体の 824 の項を見ると、まず名詞で great[keen] pleasure, thrill, kick, delight, rapture, exhilaration, enchantment, ecstasy, ravishment など喜びの様々な程度を表わす語が、文語・口語・俗語など種々の位相にわたって列挙された後に、"意地悪さ" を含意する (喜びの) 感情を表わす schadenfreude, gloating, malice のような関連語が、さらに roses all the ways, days of wine and roses, palmy days (=honeymoon) のような慣用的句表現も見出され、名詞だけでも 100 を超える語句がリストになっている。また名詞のほか、80 を超える形容詞、50 を超える動詞の類義語・関連語が続く。さらにまた、次の見出し項目を見渡すと、喜びの対立概念を表わす suffering, painfulness, regret などの項目が目に入り、喜びの感情にプラス・マイナスの形で関わる夥しい語句を気の向くままに涉猟することができる仕組みになっている。我々は

この表現の泉から最も適切な語句が選べるだけでなく、同時に思いがけない思考の発展や転回の手掛りを与えられることもある。しかし一般読者は、おそらく概念の分類表から見出し項目を求めるのは迂遠と思うであろう。そこで、手っ取り早く引くためには、巻末にそのために添えられた索引(といっても、第2版では本体の260頁に対し索引175頁; 現行版では本体の680頁を上廻る)で、joyあるいは何か思い浮かぶ語句を引き、直ちに求める分類箇所の見出し項目を検索することができる。このような読者の要望に応え、1879年版で大幅に索引が補充され、その後も版を改めるとともに索引が増大し、ついには本体の頁を上廻ることになり、最近の版では読者は、まず索引から繙くよう勧められている。

## 5. 実用的な Thesaurus-Dictionary

たしかに、索引で直に見出し語を求める方が実際的なところから、ついに Thesaurus を解体して配列をアルファベット順に、つまりふつうの辞書形式に改めたものが、まず米国で、やがて英国でも現われるようになった。これは、古くは聖書を対象にした Richard Bernard, *Thesaurus Biblicus*(1644) と軌を一にするものである。しかし、最近多く見られるようになったアルファベット順配列にした Thesaurus が、Roget の初期の意図の本質的な部分を継承しているかどうか疑わしい。ここでは、もはや、人間の思考のモデルとなるべき概念の体系的構造は影をひそめてしまっているからである。

しかし、いずれにしても Thesaurus は、いわば、'選ばれた者の備忘録' のようなもので、相当程度その言語(英語)をマスターした人でないと十分使いこなせない憾みがある。ある概念をあらわす見出し項目について、連想関係にある類義語のリストが与えられているが、リストに挙げられた語については、その意味・用法(の区別)の説明がないからである。そこで、利用者の便を考え、選んだ語句を正しく用いるためにもう一度別の一般英語辞典で引き直す労を省こうとして、thesaurus

と dictionary を機械的に合本にしたものも出版されているが、これは結局単なる糊塗策に過ぎない。これに対して、とくに学習者の要望を満たすために、分類配列の本体に各類義語の意味・用法の記述を加え、有機的に一体化した thesaurus-dictionary が構想されており、その代表的な者に Tom McArthur, *Longman Lexicon of Contemporary English*(1981) がある。その内容について概観すると、基本的・一般的な語彙約1万5千語を取り上げ、これを次の14の意味領域(semantic field)に分類している:

- A 生・生物
- B 身体の機能・福祉
- C 人間・家族
- D 建物, 家, 家庭, 衣服
- E 飲食物・農耕
- F 感情・感覚;
- G 思考・伝達, 言語・文法
- H 物質, 器具, 電気, 武器
- I 科学・技術, 産業, 教育
- J 数, 度量衡, 金融, 商業
- K 娯楽, スポーツ, ゲーム
- L 空間・時間
- M 運動, 運輸, 旅行, 場所
- N 一般・抽象語

さらに、それぞれの意味領域を次々に下位区分する。例えば、C 人間と家族は

- C1 People
  - A1 Life and Living Things
  - A50 Animals Mammals
  - C212 Inheriting and Bequeathing
  - D60 Residence
  - D130 Clothes and Personal Belongings

から始めて

- C320 Religion and Beliefs
  - G1 Thinking, Judging, and Remembering
  - G14 Believing and Accepting

に至る16項に、さらにそれぞれの項がいくつかの下位区分されている。G1 Thinkingに



ついてみると

G1 *nouns*: mind, thought, and reasons  
 の項には mind, -minded, mentality, brain, thought, reason, rationality, cognition, intellect, think があげられ、それぞれの意味・用法が一々例文をあげて記述される。以下同様に

G2 *verbs*: thinking and reasoning  
 の項には think, reason, rationalize

G3 *verbs*: solving things and working things out  
 の項には solve, think out, reason out, work out

G4 *verbs*: considering and meditating  
 の項には consider, reflect, speculate, ponder, deliberate, muse, meditate, contemplate, entertain

G5 *nouns*: considering and meditating  
 の項には consideration, reflection, speculation, deliberation, meditation, contemplation

G6 *nouns*: reasoning and logic  
 の項には reasoning, logic, philosophy, psychology, psychologist

G7 *adjectives, etc.*: reasoning and logic  
 の項には mental(ly), intellectual(ly), rational(ly), irrational(ly), logical(ly), illogical(ly), abstract, conceptual(ly), philosophical(ly), psychological(ly), cognitive(ly), coherent(ly), incoherent(ly) [以下略] のようになっている。

## 6. Roget's Thesaurus に基づく多様な Thesaurus

Roget's Thesaurus に代表される「主題、概念別分類類語辞典」の構想はこれを軸としていろいろな方向に発展させることができるであろう。そのうちのいくつかを次に挙げておく。

- (1) 多くの言語について、各概念をあらわす語句を比較意味論的に編集した thesaurus:  
 [例] C.D. Buck, *A Dictionary of Selected Synonyms in the Principal Indo-European Languages. A Contribution to the History of Ideas.* (Univ. of Chicago Pr., 1949).

- (2) 英語の方言やアメリカ英語を対象とした thesaurus:

[例] Iseabail Macleod et al., *The Scots Thesaurus.* (Aberdeen U.P., 1990); L.V. Berry and M. Van Den Bark, *The American Thesaurus of Slang.* (T.Y. Crowell, 1942, 1952<sup>2</sup>)

- (3) Shakespeare や聖書など英語文化に大きな影響を与えてきた作家・作品を個別に対象とした thesaurus:

[例] Martin Spevack, *A Shakespeare Thesaurus* (Olms, 1993); A.C. Day, *Roget's Thesaurus of the Bible* (HarperCollins, 1992).

- (4) 通時的に各概念が英語発達の各期にどのような表現組織を持っていたかを明らかにする thesaurus:

[例] M.L. Samuels et al., *The Historical Thesaurus of English* (未刊); Jane Roberts and Christian Kay, *A Thesaurus of Old English* (King's Collage London, 1995).

- (1) の Buck では、1,428 の概念を 22 の範疇に分け、各概念をあらわす語を、印欧語族に属する 31 の言語 (英語が属するゲルマン系諸言語をはじめ、ギリシア語、ラテン語、ロマンス諸語、ケルト諸語、バルト諸語、スラブ諸語、サンスクリット語など) について対照させている。22 の範疇の内容は次のとおり:

Mankind: Sex, Age, Family Relationship; Animals; Parts of the Body, Bodily Functions and Conditions; Food and Drink, Cooking and Utensils; Clothing, Personal Adornment and Care, Dwelling, House, Furniture; Agriculture, Vegetation; Miscellaneous Physical Acts and those Pertaining to Special Arts; Motion, Locomotion, Transportation, Navigation; Possession, Property and Commerce; Spatial Relations: Place; Form, Size; Quantity and Number; Time; Sense Per-

ception; Emotion, Temperamental, Moral and Aesthetic Notions; Mind, Thought; Vocal Utterance, Speech, Reading and Writing; Territorial, Social and Political Divisions, Social Relations; Warfare; Law; Religion and Superstition.

(2) のスコットランド英語の thesaurus では、20,000 の語句が 15 範疇 (birds, wild animals, invertebrates; domestic animals; water life; plants; environment; water, sea, ships; farming; life cycle, family; physical states; food and drink; law; war, fighting, violence; architecture, building, trades; religion, superstition, education, festivals; emotions, character, social behavior), 312 項目に分類され、巻末には標準イギリス英語の単語による索引が加えられている。アメリカ俗語の thesaurus は、まず第 1 部 General Slang and Colloquialisms と第 2 部 Special Slang に 2 大別した上で、全体を 24 部門に、これをさらに下位区分している。しかし、各見出し語に類義の俗語・口語が羅列されるに止まり、俗語の位相や流行した時期などに関する情報は与えていない。

(3) Shakespeare や英訳聖書には各種の concordance があつたが、これと相補的關係をなす thesaurus が作られている。Shakespeare が用いた語の種類、いわゆる異なり語数は約 3 万語とされるが、これを 37 部門と 897 の細目に分け、すべての用語から検索できる索引を付けている。本体の 216 頁に対し、索引はこれを大幅に上回る 324 頁に及ぶ。これは Shakespeare の個人言語 (idiolect) を反映するものであるから、分類範疇や項目立ても Roget's Thesaurus のものと少なからず異なるのは当然であろう。大部門の範疇は次のとおり: the physical world・the elements; plants; animals; humans・family・friendship; life: birth・death; body・bodily functions; health・medicine; dwelling・furnishing; food; clothing; basic manual acts; sense perception; emotions; quali-

ties; mind・thought; socio-political structure; solidarity; power; opposition; warfare, law; trade・possession; occupation・tools; communication; arts・leisure; pastimes・games; religion・superstition; existence・relation; manner; quantity; time; space; motion; ships・navigation; names: persons and places; names: fictional characters.

最後の方の names(人名・地名) は約 1,500 に及び、その細目でも「イギリスの歴史的人名」は 377 に達している。この thesaurus から、Shakespeare が好んで使った色彩語は何か、喜びと悲しみの感情を表す形容詞・副詞・名詞・動詞はどういう分布を示しているか、など興味ある結果を引き出すことは可能だが、今はその余裕がない。聖書の場合も Shakespeare とは違った形で Roget の範疇とのずれが予想されるが、Day の thesaurus は原則的に Roget の分類を踏襲し、分類項目名などを随時、平易・明快な名称に改める程度に止めている。(例 Intrinsicity → Nature; Tergiversation → Change of mind; Fairy → Ghost) この thesaurus は聖書本文へのレファレンスに重点があるようで、概念別分類類語辞典としては類語リストが不十分である。むしろ前に言及した Richard Bernard の *Thesaurus Biblicus* を発展させる方向で考えるべきではないかと思う。英訳聖書の thesaurus については筆者も一つの計画をもっているが、別の機会に譲りたい。

(4) 1965 年以来、Glasgow 大学の M.L. Samuels を中心に編纂が進められてきた *Historical Thesaurus of English*(略 HTE) は、Roget's Thesaurus に史的原理を導入した最初の試みである。Oxford English Dictionary(1928, 1998<sup>2</sup>) や古期英語辞典を資料として、これらの辞典に記されている語の意味を分析し、これを歴史的に分類する。すなわち、OE(古期英語) から現代英語に至る英語の語彙を、廃語・廃義も含め網羅的に収録し、ある概念が英語史の各期にどのような言語表現であらわされていたかを記述するものである。例えば、sad は 1000-1450 年の期間は satisfied の項目

に、1315-1667年の期間は *steadfast, firm* の項に、1366年から現在までの期間には *sorrowful* の項に記載されることになる。これが完成すれば英語史研究のみならず、英語に関する 'history of ideas' の研究などにも資するところ多大と期待される。しかし、その完成が当初の20年計画から大幅に遅れているため、HTEの編集スタッフであった Roberts と Kay が主になって HTE のパイロット thesaurus として独立に *Thesaurus of Old English* (略 TOE) の編集を始めることになり、1995年に2巻本 (Vol. I: Introduction and Thesaurus; Vol. II: Index) の形で出版された。HTE と TOE の概念分類の大綱を対照して挙げておく。

HISTORICAL THESAURUS OF ENGLISH

SECTION I: THE EXTERNAL WORLD

- 01 The Earth
- 02 Life
- 03 Sensation and Perception
- 04 Matter
- 05 Existence
- 06 Relative Properties
- 07 The Supernatural

SECTION II: THE MIND

- 08 Mental Processes
- 09 Emotions
- 10 Judgement, Opinion
- 11 Aesthetics
- 12 Volition
- 13 Language
- 14 Endeavour
- 15 Possession

SECTION III: THE SOCIETY

- 16 Social Groups
- 17 Social Relationships
- 18 War and Peace
- 19 Government and Politics
- 20 The Law
- 21 Education
- 22 Institutional Religion
- 23 Communication, the Media
- 24 Travel and Transport
- 25 Work

26 Leisure

THESAURUS OF OLD ENGLISH

- 01 The Physical World
- 02 Life and Death
- 03 Matter and Measurement
- 04 Material Needs
- 05 Existence
- 06 Mental Faculties
- 07 Opinion
- 08 Emotion
- 09 Language and Communication
- 10 Possession
- 11 Action and Utility
- 12 Social Interaction
- 13 Peace and War
- 14 Law and Order
- 15 Property
- 16 Religion

例えば HTE に予定されていた意味範疇 *Aesthetics* に対応するものは、TOE の *Opinion* であり、TOE では 07.10 *Beauty, fairness, 07.10.02 Ugliness, 07.10.02.01 Deformed, ugly, 07.10.03 An adornment, decoration, ornament, 07.10.03.01 Unornamented, unadorned* のように下位分類されている。また、概念別にリストされる OE 語彙は、o(1例のみ), p(詩作品のみ), g(語註, 語彙集のみ), q(文証不十分) のような用法レベルの肩つき記号 (タグ) が与えられている。例えば、07.10 の項の *Lovely, beautiful, fair* を表わすと語して *geblēod, cȳme, cȳmlic, faeger, hīwbeorht, hīwlic, hlēortorht<sup>op</sup>, scīene, wlitebeorht<sup>p</sup>, wliteful<sup>o</sup>, wlitescīne, wlitig* が列挙されている。このような thesaurus によって、OE では、詩語は *sea, boat, God* などの概念をあらわす場合にとくに多いことが知られ、また、我々は、OE における意味の場がどのようなものであり、HTE 完成の暁には、それが ME (中期英語) でどのように変化するか、さらに OE で消失した語 (概念)、ME 以降に存続した語、ME 以降で新たに造語あるいは借入された語などとの関連において、OE lexicon の実態を窺うことが可能

となろう。なお、この TOE に続き、ME Thesaurus 関係のプロジェクトが同じく London 大学 King's College で、Louise Sylvester と Roberts によって計画されている。

このように、Shakespeare Thesaurus が Shakespeare の世界を、TOE がアングロサクソンの世界を、それぞれある限界内で再現してくれるように、HTE は英語民族の世界、その観念の発生と展開の過程を垣間見させてくれるであろう。

本稿執筆に当たり、拙稿 (1998) 「Roget's Thesaurus—過去と現在」『学鑑』(Vol.95 No.12, 16-19) を部分的に再録した。参照した文献は、本論中で挙げた資料のほか、主なものは次のとおりである。

## 参考文献

- [1] L.W. Collier and C.J. Kay. 1980-81. "The Historical Thesaurus of English", *Dictionaries* 2/3: 88-112.
- [2] K.H. van Dalen-Oskam et al. 1997. *Dictionaries of Medieval Germanic Languages: A Survey of Current Lexicographical Projects*. Brepols.
- [3] D.L. Emblen. 1970. *Peter Mark Roget: The World and the Man*. Longman.
- [4] 林 哲郎. 1968. 『英語辞書発達史』開文社.
- [5] Werner Hüllen. 1999. *English Dictionaries, 800-1700: The Topical Tradition*. Oxford U.P.
- [6] 小島義郎. 1999. 『英語辞書の変遷—英・米・日本を併せ見て』研究社.
- [7] Tom McArthur. 1986. *Worlds of Reference: lexicography, learning and language from the clay tablet to the computer*. Cambridge U.P. [光延明洋 (訳) 『辞書の世界史—粘土版からコンピュータまで』三省堂, 1991]
- [8] Jane Roberts. 1998. "A Thesaurus of Old English: One Snapshot of a Vanished World", *NOWELE* (Vol.33), 133-153.

- [9] Gabriel Stein. 1985. *The English Dictionary before Cawdrey*. Niemeyer.

(1999年10月22日受付)

## 著者紹介

### 寺澤芳雄

東京大学 大学院 修士課程修了。  
東京大学 人文科学研究科 教授、  
東京女子大学 現代文化学部 教授を経て、  
現在、東京大学名誉教授、  
岐阜女子大学文学部 教授。  
New Oxford English Dictionary 顧問委員

寄稿論文

## 情報処理研究とターミノロジーから見た『分類語彙表』 —分類の体系と専門語の扱い—

### "Bunrui Goihyo" as a resource of information processing -Compilation strategy and treatment of technical terms-

柏野 和佳子<sup>†</sup>, 中野 洋<sup>††</sup>, 石井 正彦<sup>†††</sup><sup>†</sup> 国立国語研究所 言語体系研究部、<sup>††</sup> 国立国語研究所 日本語教育センター<sup>†††</sup> 大阪大学 大学院 文学研究科

日本語の代表的シソーラスである国立国語研究所『分類語彙表』について、情報処理研究とターミノロジーの観点から、その特徴を論じる。はじめに、『分類語彙表』の体系について解説と分析とを行い、『分類語彙表』を情報処理に利用する場合の利点を明らかにする(柏野・中野)。次いで、ターミノロジーの立場から、『分類語彙表』における専門語の収載状況を調査し、専門語を含むシソーラスとしての利用可能性を検討するための基礎的な資料を提示する(石井)。

#### I. 『分類語彙表』の体系

『分類語彙表』は、日本語の代表的なシソーラスと言われている。印刷物の他に電子化されたものが一般に公開されているため、情報検索などの情報処理研究の利用も近年増加している。そこで、『分類語彙表』の体系について解説と分析とを行い、情報処理研究に利用する場合の利点を明らかにする。また、情報処理利用の進展に伴って明らかになってきた『分類語彙表』への要望を取り上げ、『分類語彙表』の将来像をさぐる。

##### 1. 利用状況

昭和39年3月に、国立国語研究所資料集6『分類語彙表』の初版<sup>[1]</sup>が刊行された。以来、増刷をかさね、現在は、平成11年に印刷された31版が出ている。これまで、さまざまな言語研究に利用されてきた。宮島達夫・小沼悦は、『分類語彙表』を言語研究に利用した論文136例を集め、解説している(表1参照)<sup>[2]</sup>。

1994年に、電子化されたFD版<sup>[3]</sup>が市販された後は、工学分野における自然言語処理研究等における利用も一気に広がった。表1には、日本語処理の道具(例えば辞書)としての

使用は取り上げられていない。工学分野において、『分類語彙表』を直接、間接に利用した研究も数多くある。情報検索に関する最近の利用研究としては、宮川和・他<sup>[4]</sup>がある。文書間の類似度を判定する際に、各文書に出現する格フレームの格要素間、および、動詞間の類似度の計算に利用されている。

表1. 『分類語彙表』を用いた言語研究の論文の内訳

分類	論文数(件)
1. 語彙体系	21
2. 作品の用語調査	41
3. 文法	22
4. 方言	3
5. 日本語史	5
6. 教育・発達	11
7. 言語情報処理	7
8. 類語群	14
9. 意味分類	12
合計	136

さらに、ここ1年の利用状況をみると、医学や建築学の利用もある。『分類語彙表』の研究利用のすそのはさらに広がっている。

## 2. 『分類語彙表』の概要

### 2.1 『分類語彙表』とは

『分類語彙表』の「まえがき」に、次のように述べられている。

ここに分類語彙表というのは、一般に一つの言語体系の中で、その語彙を構成する一つ一つの単語が、それぞれどのような意味で用いられるかを一覧できるように、単語が表わし得る意味の世界を分類して、その分類の各項にそれぞれの単語を配当したものである。

ここで特徴的なのは、『分類語彙表』は、事物や概念を分類したのではなく、語の意味の世界を分類したものであるという点である<sup>[6]</sup>。語の意味と、事物や概念とを区別すべき理由は、次のとおりである<sup>[6]</sup>。

- ① ある事物を表す概念があっても、それに相当する語がない場合があるため。例：「握り鮓・押し鮓」と、「ちらし鮓・蒸し鮓」とを区別する概念はあるが、それらを区別する語がない。
- ② 同じ事物を指す語があっても、語の意味は違う場合があるため。例：「大阪鮓」と「押し鮓」が同じものを指す場合もあるが、意味は違う。
- ③ 語も概念もあるのに、事物は実在しない場合があるため。例：「河童」や「竜」は実在しない。

なお、概念を分類したシソーラスには、EDR 概念辞書がある<sup>[7]</sup>。EDR 概念辞書では、語の意味に相当する「概念」（異なりで40万概念）が分類・体系化されている。

### 2.2 収載語

収載語数は、FD版によれば36780語である。これらの語は、次の通り選定されている。

- ① 国立国語研究所報告21『現代雑誌九十種の用語用字』<sup>[8]</sup>のうちから、高使用率の7000語を中心語彙として選定。（本表では\*印が付されている）それにつづく使用率をもつ、5000語をさらに追加。（ただし、固有名、記号の類は含まない）
- ② 阪本一郎著『教育基本語彙』<sup>[9]</sup>22500語のうち、①と重複しないもの。
- ③ ①、②には含まれないが、各項目に分けた後でその項目に入る基本的単語と思われるものが不足していた場合は、主観的に補完。

③の主観的増補では、項目をみて容易におもいつく語、所属項目の明白な語は、よほど重要と認めない限り、割愛されている。たとえば《楽器》の類であることが明白なくひちりき>は、とりあげられていない。また、組になっているような場合、常にすべてはあげられていない。たとえば<真北、真西>はあるが、<真南、真東>はない。

さらに、いわゆる専門語にあたる語は、積極的に収載されてはいない。しかしながら、中心の7000語に、裁縫用語や、野球や将棋の略語などのいくつかが入っていたため、それらと比較の上で、収載されている専門語もある。この、いわゆる専門語の収載状況に関しては、IIで改めて取り上げる。

### 2.3 収載語の単位

収載語の単位は、以下の3種類である。

- a. 2.2節に示した、選定①の資料から抽出した語彙：短単位（国立国語研究所の語彙調査におけるβ単位）  
例：自転車→「自転」と「車」とで別単位のものとしてそれぞれを収載
- b. 選定②および選定③によって補った語彙  
その1：短単位複合（β複合単位）  
例：自転車→「自転車」全体を1単位として収載

## c. 選定②および選定③によって補った語彙

その2：慣用語句の単位

例：しりもちをつく→「しりもちをつく」全体で1単位として収載

上記 a. の「β単位」について若干補足説明する。β単位というのは、『現代雑誌九十種の用語用字』の語彙調査において、語の基本度をはかることを目的として短めに規定された調査単位である。たとえば、「自転車」の場合は、「自転」と「車」とに分けられ、「自動車」の場合は「自動」と「車」とに分けられ、それぞれの単位によって使用率が調査される。その結果、「自転」「自動」に比べ、「車」の使用率が高く出れば、そこから「車」という語は基本度が高い、といったことが導き出せる。『分類語彙表』の収載語は、先にも述べた通り、『現代雑誌九十種の用語用字』の調査における高使用率の語より選定したものであるので、調査単位であった、そのβ単位によって語が収載されている。しかしながら、それより長めの表現単位を、ある程度採用しておくほうが実用的な目的にかなうとも考えられたため、上記 b. や c. の単位をたてて収載された語もある。

## 2.4 分類体系

『分類語彙表』の分類体系は階層構造になっている。よって、木の形が明示されていないが、その体系は木構造であるといえる。階層は固定されている。分類項目は、浅いもので3階層、深いもので5階層になっている。各項目には、その階層に応じて、3桁から5桁の分類番号が付されている。さらに、項目によっては、段落替えによって小グループに分けられており、その結果、一番深いものは6階層になっていることになる。そして、各語は、3階層目から6階層目に分類されている。

分類項目の1階層目は、品詞論的な4分類によるものであり、2階層目は、意味範囲の5部門によるものである。3階層目から5階層目は、さらなる意味範囲の細分類による。以下に、それぞれの詳細を示す。

## 2.4.1 1階層目：品詞論的な4分類

大分類として、はじめに、品詞論的に4分類されている。この4分類の番号が、分類項目の1桁目の番号になっている。

1. 体の類…名詞の仲間—《何、何ごと、何もの、どれ、だれ、いつ、どこ、いくつ》等の概念を表す語と、それらを問いとしたときの答えとなるべき語。
2. 用の類…動詞の仲間—《ある》に関するもののほか、《どうする、どうなる》等の答えとなるべき語。
3. 相の類…形容詞の仲間—《ない》に関するもののほか、《どう、どうだ、どんな、どんなに》等の答えとなるべき語。いわゆる形容動詞、連体詞、ある種の副詞を含む。
4. その他の類…その他の仲間—いわゆる接続詞、感動詞、ある種の副詞の類等。概念間の関係づけ、叙述間の関係づけ、感動、呼びかけ応答、判断・期待・仮定などの叙述態度の予告、待遇表現などを表す語。

しかしながら、品詞分類が優先されると、次のような語は、意味的に近くとも、離れてしまうことになる。

- (1) 名詞と複合サ変動詞語幹「変化」(体の類)「変化する」(用の類)
- (2) 名詞と形容詞・形容動詞語幹「美」(体の類)「美しい」(相の類)
- (3) 動詞の派生名詞と動詞「動き」(体の類)「動く」(用の類)
- (4) 形容詞・形容動詞の派生名詞と形容詞・形容動詞<sup>[註1]</sup>「高さ」(体の類)「高い」(相の類)

そこで、このような分離に対処するために、大分類以下を細分する際にはなるべく平行に行い、細分の番号をある程度一致させることが試みられている。たとえば、「光」に関しては次のように分類されている。

- 例：< 1.501 光 > 名詞「光，輝き」など  
 < 2.501 光 > 動詞「光る，輝く」など  
 < 3.501 光 > 形容詞や副詞「明るい，くっきり」など

つまり、品詞をまたいで関連する意味ごとの語を見比べたい場合は、分類番号がその手がかりになるよう、便がとられている。

このように、先に、品詞で分類するか、あるいは、意味で分類するかは、シソーラスの設計において、最初の大きな分かれ目であろう。他のシソーラスでは、たとえば、『分類語彙表』と同じく、先に品詞で分類されているものには、WordNet<sup>[10]</sup>がある。逆に、先に意味で分類されているものには、ロジェのシソーラス<sup>[11]</sup>や、角川類語新辞典<sup>[12]</sup>がある。先にどちらで分類する方がより望ましいかは、利用目的によるであろう。『分類語彙表』は、同じ品詞どうしでの語の比較のしやすさや、品詞別にどのような意味分布があるかの見通しの良さを優先させ、先に品詞論的に分類したものである。

## 2.4.2 2階層目：意味範囲の5部門

品詞論的分類の次に、意味範囲の違いにより、次の5部門がたてられている(ただし、5部門あるのは体の類のみ)。この5部門の番号が、分類項目の2桁目の番号である。

### 1 抽象的關係

(人間や自然のあり方のわく組)

体の類 用の類 相の類

### 2 人間活動の主体

体の類

### 3 人間活動—精神および行為

体の類 用の類 相の類

### 4 人間活動の生産物—結果および用具

体の類

### 5 自然—自然物および自然現象

体の類 用の類 相の類

## 2.4.3 3階層目から5階層目：細分類

さらに、各部門は、以下の項目の数だけ細分類されている。

体の類 : 5 4 5 項  
 用の類 : 1 4 5 項  
 相の類 : 8 5 項  
 その他の類 : 2 3 項  
 合計 : 7 9 8 項

(用の類、相の類の項目は、内容上ほぼ、体の類の項目に重なる。)

### (1) 項目の排列

基本的には、検索の便宜のために、互いに関連する項目は接続して排列されている。また、次の例のように、一般的総括的内容を持つ項目は、部分的な内容を持つ項目より先にあげられている。

- 例：1.57 身体に関する部分の総記  
 1.571~8 身体の各部位に関する記述

### (2) 項目の見出し

分類項目の見出しは、項目の全体を示すような代表単語のほか、単語の列挙、説明的な語句などによってつけられている。

- 例：1.251 家  
 1.265 店・旅館・病院・劇場など  
 1.275 国際機構

### (3) 項目の大きさ

各項目に収載する語の数は限定されていない。よって、項目によって収載されている語の数は異なり、項目の大きさには大小さまざまある。

- 例：語数 1 < 4.34 云々 > 云々  
 語数 2 < 1.1640 古今 > 古今 今昔  
 語数 258 < 3.15 変化・動き > <sup>[註2]</sup>  
 おのずから おのずと ひとりでに



自然 自然に そぞろ 何となく  
……(以下, 略)

(4) 一項目に収めた語の性質

一つの項目は, 同義類義の関係(待遇や語感をぬきにして, 同じ文脈で入れ替え可能であるもの)でまとめられている。

また, 一つの項目が自由連想による語群になることは, 極力避けられている。たとえば, 「ビール」についての連想による語群として, 次のものが考えられる。

例: <飲酒行動に関連するもの> 酒,  
スタウト, ウイスキー, 飲む, 酔う,  
一杯, あわ, ジョッキ, コップ,  
ほろにが, ホップ, 赤ら顔, ビヤホール  
しかし, 『分類語彙表』では, 「ビール」は,  
「酒・ウイスキー・スタウト」と同じグループ  
をなすものとして扱われ, 上記のうちの, 「飲む」  
や「ビヤホール」などとの関係は断たれている。

(5) 語の多重分類

多義についてはあまり考慮されておらず, ほとんどの語は, 一つの項目にしか掲げられていない。多重分類されているのは, たとえば, 複合概念のある語で, その場合, 構成部分のそれぞれが該当する分類項目にいれられている。

例: <言い直す> → < 2.312 言語 >  
< 2.1501 改定 >

ただし, 複合語でも, 異なる範ちゅうに属する語がそれぞれ独立して複合しているものは, それぞれの項目に分出されている。

例: <真善美> → < 1.1344 美醜・難易  
など >  
<真> → < 1.103 真・実 >  
<善> → < 1.3046 誠・徳・義・信・  
恩など >  
<美> → < 1.1344 美醜・難易など >

(6) 項目の多重分類

項目間の多重分類の可能性は明示されていない。たとえば, <出現> や <消滅> は <変化> の類に含めることが, あるいは, <生死> なども <変化> に含めることが考えられたということが, 「まえがき」に指摘されているのみで, 表中には, そのような多重分類についての明記はない。

2.4.4 体系のまとめ

以上, 分類体系の詳細を述べた。まとめを表2に示す。そして, 語数の内訳を表3に, 記述の一例を表4と表5に示す。

表2. 体系のまとめ

収載語数	36780 語
分類項目数	798 項
体系枠の形	木構造
階層数	固定(3~6階層) *ただし, 分類番号が付されてあるのは5階層目まで
ノード	第1ノードは品詞, 第2ノード以下は意味分類による (*必ずしも単語ではない)
1項目内の下位数	特に限定なし
分類単位	一語の意味レベル

3. 情報処理利用に応える『分類語彙表』

前述のように, 『分類語彙表』は近年計算機による情報処理にも利用されてきている。しかし, もともと人手による利用を想定して設計されたものであるために, 情報処理利用の観点からは, 補足すべき情報があることが明らかになってきた。そこで本章では, 『分類語彙表』の情報処理利用において指摘されてきた事項を列挙し, 他のシソーラスの内容とも比較検討しながら, 情報処理利用にも十分応え得る『分類語彙表』の将来像をさぐる。

表3. 語数の内訳 (2階層までの分類時)

	体	用	相	その他	合計
抽象的關係	6780	2153	2212	99	
人間活動主体	3272				
人間活動	9920	2158	1788		
生産物	3277				
自然	3735	468	653	265	
計	26984	4779	4653	364	36780

表4. 分類項目の記述例

2.	用の類
2.1	抽象的關係
2.111	關係
2.112	異同
2.113	包摂
2.120	存在
2.121	出現
2.122	成立・発生
2.123	仕上げ
2.124	残存・消滅
2.125	保有・除去

(1) 意味關係の明示

田中穂積らは、『分類語彙表』などの「従来のシソーラスは、階層化されたもの相互が意味的にどの様な關係にあるかが不明確で曖昧なことが多い。たとえば階層化されたもの相互が上位/下位關係にあるのか、それとも部分/全体關係にあるのかがはっきりしない」と指摘し、それらが自然言語意味処理には不十分であり、階層關係を明確にしたシソーラスの作成が重要であると主張した<sup>[13]</sup>。

人が使う場合は、意味關係を推察することが可能であるため、『分類語彙表』を含め、従来の日本語のシソーラスは、意味關係が明示化されていなかった。しかしながら田中らがいうように、計算機にそれを推察させることは困難であろう。

そのため、最近のシソーラスでは、意味關係の明示化がさまざまに試みられている。たとえば、荻野綱男の現代日本語名詞シソーラ

表5. 分類語の記述例

2. 用の類
2.1 抽象的關係
2.111 關係
1 *対する *属する 帰する *もとづく *従う *因る *抛る ちなむ *かかわる *關する *就く (就いて) *於く (於ける) *以つ *あずかる
2 *れる *られる
3 *役立つ 役立つ 用立てる
4 *ひびく 徹える 成りかわる なぞらえる
2.112 異同
1 *似る 似せる 似通う 類する あやかる
2 *違う 違える *変わる *異なる *ほかなる *反する もとる くいちがう たがう
3 即する *応ずる 準ずる *当たる *合う そぐう *副う
4 *兼ねる あてはめる あてはまる -ぐむ -らむ
5 *-じみる -ばむ -びる *-めく

ス<sup>[14]</sup>では、上位/下位關係と、部分/全体關係とが區別して明示されている。また、英語のシソーラスに、WordNet<sup>[12]</sup>がある。これは、語と語の間、また、語義と語義との間にある様々な意味關係をポインタとしてネットワーク状に張り巡らしたものである。ここでは意味關係を詳細にとらえることが試みられている。上位/下位關係の下位にあたるものは「部分」「属性」「機能」に分けてとらえられ、また、部分/全体關係の部分にあたるも

のは、「一部」「要素」「材料」に分けてとらえられ、階層化されている。また、「同義」だけでなく、「反義」もあわせて扱っている点も、『分類語彙表』などとは異なる点であろう。計算機処理の利用を想定するのであれば、このように詳細な関係の明示を検討する必要がある。

## (2) 意味素性の明示

渡辺靖彦らは、『分類語彙表』において同じ段落に属する名詞は、同じ意味的側面(意味素性)をもち得ることに着目した研究を行った。たとえば<1.2410 専門的・技術的職業>に分類されている「教授」「教官」「生徒」は、すべて「HUM(人間)」という意味素性をもち得る。その際に、「同じ段落に属する名詞でも非常に緩い類語関係しかなく、同じ意味素性を付与することが妥当ではない場合がある」ということを問題点としてあげ、<1.3830 運輸・交通>に分類されている「早馬」「小包」「新聞配達」の例を取り上げている<sup>[15]</sup>。

『分類語彙表』は、すべてにおいて、意味素性という観点で分類したわけではないため、このような問題が生じるのは、ある部分仕方がなかったといえる。しかしながら、分類されたものが結果的に同じ意味素性である部分を実際にはあるので、計算機利用を考えた時、分類されている語がどんな意味素性をもつ語の集合になっているかを明示することは有意義であると思われる。

なお、意味素性そのものを先に体系的に分類し作成した大規模シソーラスとしてNTT日本語語彙体系<sup>[16]</sup>がある。約3000の意味素性が体系的に分類され、その意味素性ごとに、30万語の単語が分類されている。

## (3) 視点の明示

長尾真編『自然言語処理』<sup>[17]</sup>では、「語の分類をする際には語に対する視点の問題を考慮する必要がある。すなわち言葉には別の視点からみれば別の意味の側面に焦点があたるといった性質がある。」という指摘がされ、次の例が示されている。

道具という視点：たわし、洗濯機

ほうき、掃除機

動力という視点：たわし、ほうき

洗濯機、掃除機

つまり、シソーラスが固定されていると、別の視点による意味関係がとらえられないということである。そのため、シソーラスは、理想的には視点別に動的に生成できることが望ましいということになる。そのようなシソーラスをめざすものとして、川村和美らの「語を種々の観点から分類した多次元シソーラス」がある<sup>[18]</sup>。これは、ひとつの語に観点を複数与え、観点ごとに動的にシソーラスを構築する方法を提案するものである。たとえば、「鳥」と「飛行機」について、次のような記述が提案されている。

動物＝生物 {動的属性：(動く),  
静的属性：(性別), …}

鳥＝動物 {動的属性：(動く(空中),  
利用・用途(食物, ペット),  
全体/部分(翼), …}

乗りもの＝人工物 {動的属性：(動く),  
利用・用途(移動), …}

飛行機＝乗りもの {動的属性：(動く(空中)),  
全体/部分(翼), …}

『分類語彙表』など、固定化された従来のシソーラスでは、「鳥」と「飛行機」は、それぞれ「生物」と「人工物」として分類されるため、離れてしまう。しかしながら、上記のように複数の観点によって記述してあれば、両者から「空中を動く」「翼を持つ」という共通点を取りだし、その観点でシソーラスを生成すれば、両者を近くに並べてとらえることが可能になる。

このように、川村らの提案する、各語について視点や観点を記述するという方法は、意味関係の明示や意味素性の情報明示にもつながり、上記(1)や(2)で取り上げた問題も同時に解決しようとするものといえる。

この他、『分類語彙表』に関して、現在、検討をはじめている課題について述べる。

①語の追加

現在、約8万5千語を目途に語の増補中である<sup>[註3]</sup>。

②多義の明示、用例の付与

語を分類しているため、多義語の場合、機械的にどの意味であるかを特定することが困難な場合がある。先に述べたとおり、視点や観点の記述が進めば解決されることではあるが、それに先立って、簡単な語釈の付与により多義を明示することや、用例の付与により意味の特定を容易にすることも検討している。

II.『分類語彙表』における専門語

既に述べたように、『分類語彙表』はあくまで一般語のシソーラスとして設計されたものであるが、情報検索などでの利用においては、専門語についての情報も必要になる。そこで以下では、『分類語彙表』にどれほどの専門語が含まれているのか、それら専門語はどのような意味分野に多いのか、また、どのような専門分野の専門語が多いのかを調査・列挙する。これは、専門語を含むシソーラスとしての『分類語彙表』の利用可能性を吟味するための基礎的な資料となろう。

1. 専門語の範囲

専門語の多くは名詞であり、動詞や形容詞など他の品詞の用語、また、句の形式の用語は少ない。したがって、『分類語彙表』の用・相・その他の類の語について、それが専門語か否かを問うことにはあまり意味がない。そこで、ここでは、体の類についてのみ、専門語か否かを調べることにする。

とはいえ、すべての専門分野をカバーすることは困難であり、また、必ずしも意味のあることとはいえないので、ここでは対象を科学技術系の専門語に限定する。具体的には、以下に示す文部省編『学術用語集』の理学系11分野、工学系8分野の用語を、専門語の集合として利用する(化学と電気工学は増訂版を、他は初版を用いた。詳しくは野村雅昭・石

井正彦<sup>[19][20]</sup>を参照されたい。なお、理学系・工学系の分類は青戸邦夫<sup>[21]</sup>に従った)。『学術用語集』は各分野の基本的な専門語を収めたもので、一般語を収めた『分類語彙表』との対照には都合がよい。高度な専門語はそもそも『分類語彙表』に載っていないと考えられるからである。

[理学系] 数学 天文学 物理学  
 気象学 海洋学 地震学 分光学  
 化学 動物学 植物学 遺伝学  
 [工学系] 機械工学 電気工学  
 土木工学 建築学 船舶工学  
 航空工学 計測工学 原子力工学

(下線部分は専門分野名の略号)

2. 専門語認定の方法

『分類語彙表』にある個々の語について、それが専門語であるか否かを判定することはそれほど簡単ではない。一般語の表すものを「意味」、それを通して専門語が表すものを「(科学的)概念」と呼ぶとすれば<sup>[22]</sup>、ある語の意味と概念との間には、少なくとも、以下の3つの場合が想定できる(語例の後の数字は分類番号、漢字は上に示した専門分野の略号)。

- A. 意味と概念とが(ほぼ)同じと考えられる場合  
 例:「隠花植物」(1.551-植),「地球観測年」(1.1630-天),「中央火口丘」(1.524-震地),「ヘモグロビン」(1.577-動化),「方解石」(1.5111-化海)など
- B. 意味と概念とは同じではないが、両者の間にある程度の共通性・有縁性があると考えられる場合  
 例:「厚板」(1.412-建),「圧力」(1.1401-1.193-土物機建計化),「仮説」(1.3071-1.3075-数),「クライマックス」(1.1652-植),「絶対」(1.1130-天)など
- C. 意味と概念とが同じではなく、また、両者の間に共通性・有縁性もないと考え

られる場合

例:「花器」(1.4511 - 植),「額面」(1.1911 - 動),「サイン」(1.3153 - 数) など

Aは、意味と概念との間に、詳しさを厳密さを除けば、ほとんど違いがない語である。したがって、国語辞典と専門用語辞典との記述も近いものになることがある。

### 隠花植物

花や種子を生じないで孢子で繁殖する植物の総称。しだ・こけなど。▽現在の分類では使わない。⇨顕花植物(『岩波国語辞典 第五版』1994)

顕花植物に対する用語。植物界を顕花植物と隠花植物とに2大別することは、正当ではないが、いろいろ便利な点もあるので一般に用いられている。孢子によって生殖する植物をいう。花を生じない植物をまとめた大群で種子植物以外の全植物を含む。例:シダ植物, コケ植物, 菌類, 藻類。(『最新植物用語辞典』廣川書店, 1965)

Bは、意味と概念とは異なるが、両者の間にある程度の共通性(類似性)・有縁性が看取される類である。国語辞典で「厚い板」「分の厚い板」などと説明される一般語の「厚板」と、「約6 mm以上の厚さの熱間圧延鋼板」(日本建築学会編『建築学用語辞典』岩波書店, 1993)と定義される建築学の「厚板」とは決して同じではないが、意味的主要部(類概念)の類似性(「板」), 意味的修飾部(種差)と主要部(類概念)との関係の共通性(「厚→板」)など、両者に重なる部分があることも確かであろう。

Cは、見かけ上は同じ語であるが、語源的には別語であったりして、一般語としての意味と専門語としての概念との間に確たる共通性・有縁性を見出せない類である。「花を生けるためのうつつわ」である一般語「花器」と、「花を構成する各器官」である植物学用語「花

器」との間に、とくに意味的主要部(類概念)における共通性は認められない。

さて、本稿では、『分類語彙表』にある語が上のA及びBの類であったとき、それを専門語と判断しよう。正確を期するならAのみを専門語とすべきであろうが、意味と概念との隔たりは相対的・連続的なものであり、AとBとの間に明確な境界線を引くことは困難である。Bには、意味と概念との差が小さいものから大きいものまでさまざまなものがあり得るが、ここでは、専門語の範囲をより広く認めることにしたい。

なお、意味と概念とが同じか重なっていても、語形が(部分的にでも)異なる場合は専門語とはしない(例:「勝手」(1.443) - 「勝手口」(建),「小数点」(1.3114) - 「小数部」(数),「頭蓋骨」(1.576) - 「頭蓋」(動))。

### 3.『分類語彙表』に専門語はどれほど含まれているか

『分類語彙表』収載の名詞(体の類)26984語のうち、上記19分野の『学術用語集』にもある、すなわち(本稿でいう)専門語であると考えられたものは、全部で3793語(14.1%)であり、ほぼ7語に1語が専門語という結果であった。このうち、理学系は2410語(8.9%), 工学系は2476語(9.2%)で、ほとんど差がなかった(理学系のみは1317語, 工学系のみは1383語, 理学系工学系共通は1093語)。7語に1語という専門語含有率は、前述した『分類語彙表』「まえがき」の「比較的専門特殊に属すると認める語も多くを捨てた」という記述に反して、決して小さい値ではない。これは、ここでいうBの類の多くが専門語と意識されていないためではないかと思われる。

表6. 専門語の量的な分布(大項目)

	分類語彙表の語数	理学系	%	工学系	%
1.1 抽象的關係	6780	916	13.5	763	11.3
1.2 主体	3272	42	1.3	116	3.5
1.3 人間活動	9920	318	3.2	377	3.8
1.4 生産物	3277	347	10.6	791	24.1
1.5 自然	3735	787	21.1	429	11.5

## 4. 専門語はどのような意味分野に多いか

### 4.1 大項目

大項目における専門語の量的な分布を表6に示す。語数で見ると、理学系は<1.1 抽象的關係>が最も多く、次いで<1.5 自然>が多い。工学系は<1.4 生産物>が最も多いが、<1.1 抽象的關係>も多い。『分類語彙表』の収載語数に対する割合(専門語含有率)で見ると、理学系では<1.5 自然>が最も高く、工学系では<1.4 生産物>が最も高い。理学系が<自然>、工学系が<生産物>というのは、両系の対象領域の違いを反映したものであろう。なお、両系とも<1.2 主体><1.3 人間活動>が少ないのは科学技術系だからで、社会・人文科学系であれば、これらの分野に属する専門語も多いはずである。

### 4.2 中項目

次いで、中項目における専門語の量的な分布を表7に示す。語数で見ると、理学系は<1.51 自然>が最も多く、工学系は<1.15 作用>が最も多い。<1.15 作用>は理学系でも多い。専門語含有率で見ると、理学系では<1.51 自然>が、工学系では<1.44 住居>が最も高い。これは、大項目の結果に沿うものだが、理学系でも<1.41 資材>の専門語含有率は高く、工学系でも<1.51 自然>は高い。

### 4.3 小項目

理学系で、語数から見た上位10位までの意味分野と、それに所属する専門語をあげる。

#### 1.5111 鉱物 52語

アスファルト(化), 油(化), 安山岩(震), 石(海), 雲母(化), エボナイト(化), エメラルド(化), 塩分(海), 塊炭(化), 花崗岩(化), 火山岩(震), 火成岩(震化), 化石(震), 軽石(海震化), ガソリン(化), 岩塩(化), 岩石(海), 煙(気), 原炭(化), 原油(化), 鉱石(化), 鉱物(震化), コークス

(化), コールタール(化), 磁石(物震化), 重油(化), 水晶(化), 煤(震化), 石英(化), 石炭(化), 石油(海化), 石灰(化), 石膏(化), ゼラチン(化), ダイヤモンド(化), 大理石(化), 土(化), 燈油(化), 泥(海), にがり(海化), 粘土(海化), 灰(化), 腐植土(植), 噴煙(震), 粉炭(化), 変成岩(震), 方解石(海化), 明礬(化), 無煙炭(化), 熔岩(震化), 汚れ(化), 緑青(化)

#### 1.574 膜・筋・神経・内臓 48語

胃(動), 鰓(動), 肝臓(動), 気管(動), 気管支(動), 気嚢(動), 筋肉(動), 頸動脈(動), 血管(動), 腱(動), 睪丸(動遺), 呼吸器(動), 骨膜(動), 鼓膜(動), 子宮(動), 消化器(動), 小腸(動), 小脳(動), 食道(動), 神経(動), 心臓(動), 十二指腸(動), 循環器(動), 静脈(動), 腎臓(動), 脾臓(動), 髄(植), 生殖器(動), 声帯(動), 脊髄(動), 胎盤(動), 胆嚢(動), 大腸(動), 大脳(動), 腸(動), 直腸(動), 動脈(動), 内臓(動), 粘膜(動), 脳(動), 肺(動), 扁桃腺(動), 膀胱(動), 盲腸(動), 網膜(物動), 卵巣(動遺), リンパ腺(動), 肋膜(動)

#### 1.5110 化学成分 47語

亜鉛(化), 亜硫酸(化), アルカリ(化), アルコール(化), アルミニウム(化), アンモニア(化), 硫黄(化), 一酸化炭素(化), 塩(震化), 塩酸(化), 塩素(化), 苛性ソーダ(化), カルシウム(化), 貴金属(化), 金(化), 金属(化), 銀(化), 軽金属(化), 元素(天化), 醋酸(化), 砂鉄(海), 酸(化), 酸素(化), 硝酸(化), 水銀(化), 水素(化), 錫(化), 青銅(化), タングステン(化), 炭酸(化), 炭素(化), 窒素(化), 鉄(化), 銅(化), ナトリウム(化), 鉛(化), ニッケル(化), 白金(化), 非金属(化), 砒素(化), 硼酸(化), マグネシウム(化), マンガン(化), ラジウム(化), 硫酸(化), 燐(化), 燐酸(化)

#### 1.510 自然・物体・物質 37語

イオン(物化), 陰極(物化), 引力(天物震光), 液体(震化), 音波(物), 還流(化),

表7. 専門語の量的な分布 (中項目)

分類語彙表の語数	理学系	%	工学系	%	分類語彙表の語数	理学系	%	工学系	%
1.10 こそあど	208	15	7.2	10	4.8	1.33 生活	1325	35	2.6
1.11 類	454	72	15.9	27	5.9	1.34 行為	517	4	0.8
1.12 有無	346	28	8.1	14	4.0	1.35 交わり	702	16	2.3
1.13 様相	455	55	12.1	40	8.8	1.36 人事	752	7	0.9
1.14 力	103	17	16.5	23	22.3	1.37 取得	851	13	1.5
1.15 作用	1750	262	15.0	291	16.6	1.38 産業	906	87	9.6
1.16 時間	1124	57	5.1	26	2.3	1.40 物品	143	8	5.6
1.17 空間	879	87	9.9	85	9.7	1.41 資材	399	95	23.8
1.18 形	396	111	28.0	81	20.5	1.42 衣服	428	29	6.8
1.19 量	1065	212	19.9	166	15.6	1.43 食料	457	43	9.4
1.20 人間	431	5	1.2	0		1.44 住居	451	32	7.1
1.21 家族	339	11	3.2	2	0.6	1.45 道具	817	68	8.3
1.22 相手	112	0		0		1.46 機械	328	51	15.5
1.23 人物	511	4	0.8	1	0.2	1.47 土地	254	21	8.3
1.24 職	916	5	0.5	38	4.1	1.50 刺激	336	58	17.3
1.25 区画	341	2	0.6	6	1.8	1.51 自然	834	305	36.6
1.26 社会	384	10	2.6	56	14.6	1.52 天地	497	113	22.7
1.27 機関	142	2	1.4	12	8.5	1.55 植物	510	73	14.3
1.28 団体	96	3	3.1	1	1.0	1.56 動物	384	23	6.0
1.30 心	2915	96	3.3	84	2.9	1.57 からだ	616	150	24.4
1.31 言語	1591	52	3.3	56	3.5	1.58 生命	558	65	11.6
1.32 芸術	361	8	2.2	4	1.1				

気体 (物化), 原子 (数物光化), 交流 (物震), 固体 (物震化), コロイド (物海化), 触媒 (物化), 磁場 (天物震), スパーク (光), 短波 (震), 弾性体 (震), 長波 (物気海震), 直流 (物震), 電極 (物震化), 電子 (天物化), 電磁波 (天物), 電波 (物), 電流 (物), 導体 (物化), 熱線 (物), 波動 (物震), 微粒子 (天), 不導体 (物), 粉末 (化), 物質 (物), 物体 (物), 分子 (数物光化), 放射能 (天物海震化), 放電 (物気化), 溶液 (化), 陽極 (物化), 流体 (物化)

圧力 (物化), 緯度 (天震), 重さ (物震), 重み (天物震), 温度 (物気), 荷重 (物震), 硬さ (物化), 感度 (物震光化植), 気圧 (気), 気温 (気), 輝度 (天物海光), 強度 (物震光化), 経度 (天震), 血圧 (動), 光度 (天物光), 硬度 (物化), 彩度 (光化), 視角 (物), 湿度 (物気化), 純度 (天物化), 常温 (化), 水圧 (海), 水温 (気海), 体重 (海), 強さ (天物震光), 電圧 (物), 電位 (物化), 濃度 (数光化), 比重 (物海震化), 密度 (数天物気海震化), 明度 (光化)

1.553 枝・葉・花など 37語

枝 (物光植), 雄しべ (植), 雄花 (植), 核 (数天震化動植遺), 果実 (植), 花粉 (植遺), 株 (植遺), 花卉 (植), 萼 (植), 茎 (植), 紅葉 (植), 莢 (植), 種子 (植), 子葉 (植), 樹脂 (物化), 樹皮 (植), 種 (植), 地下茎 (植), 蔓 (植), とげ (植), 根 (植), 年輪 (植), 葉 (植), 胚芽 (気), 花 (植), 花びら (植), 麦芽 (化), 複葉 (植), 節 (物震光化植), 弁 (動), 孢子 (天植遺), 松やに (化), 実 (植), 幹 (植), 芽 (動植), 雌しべ (植), 雌花 (植)

1.513 水 29語

泡 (化), 雨滴 (気), 海水 (海), かげろう (気), 乾燥 (化), 気泡 (震化), 下水 (海), 硬水 (化), しぶき (海), 霜 (気), 霜柱 (気), 重水 (海化), 樹氷 (気), 蒸気 (物化), 蒸溜水 (化), 水分 (物気化), 淡水 (海化), 脱水 (化植), 地下水 (震), 露 (気), つらら (気), 軟水 (化), 逃げ水 (気), 濁り水 (海), 初霜 (気), 晩霜 (気), 水 (化), 流水 (気海), 冷水 (海)

1.193 角度・軽重・寒暖など 31語

1.521 天体 29語

天の河 (天), 衛星 (天), 海王星 (天), 火星

(天震), 金星(天震), 銀河(天), 恒星(天), 新月(天), 新星(天), 水星(天震), 彗星(天), 星雲(天), 星座(天), 星団(天), 太陽(天震), 月(天震), 天体(天), 天王星(天), 土星(天), 伴星(天), 日(天), ほうき星(天), 星(天物), 北極星(天), 満月(天), 冥王星(天), 木星(天), 流星(天気), 惑星(天震)

1.181 点・線 28語

円弧(震), 回帰線(天震遺), 下弦(天), 軌跡(数震), 基線(天物震), 曲線(数), 弦(数天), 弧(数), 黒点(天), 子午線(天震), 斜線(数), 斜辺(数), 上弦(天), 垂線(数), 垂直線(天), 赤道(天震), 線(数), 線分(数), 対角線(数), 地平線(天震), 直線(数), 底辺(数), 点(数), 等温線(物気海震化), 等高線(気震), 班点(化), 辺(数物震), 放物線(数)

1.1912 整数・対数など 28語

因数(数), 函数(数天物), 奇数(数), 級数(数), 逆数(数), 偶数(数), 係数(数天), 公倍数(数), 公分母(数), 公約数(数), 根(数), 小数(数), 実数(数), 乗数(数), 除数(数), 正(数動植), 整数(数), 素数(数), 対数(数), 帯分数(数), 倍数(数), 負(数動植), 分子(数物光化), 分数(数), 分母(数), 変数(数植), 無理数(数), 約数(数)

1.5160 物質の変化 28語

汚染(海化遺), 化合(化), 還元(化植遺), 凝結(物気化), 凝固(天物海震化), 結晶(震化植), 結氷(海), 懸濁(物化), 酸化(化植), 煮沸(化), 昇華(物気化), 蒸発(天物気海化), 蒸溜(物化), 中和(物化), 沈澱(化), 電解(物化), 電離(天物光化), 凍結(気化), 爆破(震化), 爆発(物震化), 漂白(化), 腐蝕(物化), 沸騰(物化), 放射(天物気光化), 膨脹(震化), 融解(物気化), 雪どけ(気), 溶解(化)

また, 工学系の上位10位までの意味分野と, それに所属する専門語をあげる。

1.441 家屋 43語

東屋(建), アトリエ(建), 母家(建), 温室(建), 家屋(建), 格納庫(建航), 仮小屋(土), 宮殿(建), 議事堂(建), 庫裡(建), 公会堂(建), 校舎(建), 講堂(建), 米倉(船), 金堂(建), 蚕室(建), 車庫(機建), 書院(建), 食堂(建船), 書庫(建), スタンド(機電建), 倉庫(土建船), 建て物(建), 台(機電船), 鉄塔(電), 塔(建), 燈台(土建船), 土蔵(建), 長屋(建), 納屋(建), 拝殿(建), 離れ(建), 平屋(建), 兵舎(建), 方丈(建), 本館(建), 本殿(建), 本堂(建), 望楼(建), むろ(建), 物置き(建), 檣(建), 山小屋(建)

1.5111 鉱物 39語

赤土(建), アスファルト(機電土建), 油(建), 安山岩(機建), 石(土), 雲母(機電), エボナイト(機電船), 煙幕(船), 塊炭(機船), 花崗岩(機建原), 火成岩(建), 軽石(機土建), ガソリン(機), 原油(機土船), コークス(機), コールタール(機土建船), 錆(機船), 磁石(機電), 砂利(機土建), 重油(機), 水成岩(建), 煤(機), 砂(土建), 石炭(機), 石油(機), 石膏(土建), 玉砂利(建), 大理石(電建), 塵(建), 土(土), 燈油(機土航), 泥(建), 粘土(土建), 腐植土(土建), 粉炭(機船), 無煙炭(機船), 燃えさし(建), 湯垢(機船), 汚れ(船)

1.472 その他の土木施設 38語

暗渠(機電土建), 石畳(建), 井戸(土建), 運河(機土), 岸壁(機土船), 漁港(土船), 空港(機土建船航), 港湾(土), 棧橋(機土建船), 商港(船), 城(建), 浄水池(機土), 水門(機電土船), 捨て石(土建), 堰(機土計), 泉水(建), 溜め(機船), ため池(土), ダム(機電土), 築港(船), 貯水池(機電土建), 堤防(土), 転炉(機), 突堤(機土船), 飛び石(建), ドック(機船), 土手(建), 庭石(建), 不凍港(船), 噴水(建), プラットホーム(機建), 掘り抜き井戸(機土建), 防空壕(建), 防波堤(土船), みぞ(機建), 港(機船), やり水(建), 熔鉱炉(機船)



1.443 へや・廊下・床など 37語

暗室(機建船), 板の間(建), 居間(建), 縁側(建), 押入れ(建), 階段(建船), 回廊(建), 楽屋(建), 画廊(建), 客間(建), 玄関(建), 更衣室(建), 室(建), 書斎(建), 診察室(建船), 寝室(建), 事務室(建船), スタジオ(電建), 台所(建), 地階(土建), 茶室(建), 茶の間(建), 床の間(建), 土間(建), 納戸(建), 二階(建), 花道(建), 控え室(建船), 病室(建船), へや(建), 便所(建), 待合室(土建), 水屋(建), 浴室(建船), 浴場(建), 廊下(建), 渡り廊下(建船)

1.510 自然・物体・物質 37語

イオン(電原), 陰極(電), 陰電気(電), 引力(船), 音波(機電建), 還流(機原), 気体(機), 原子(電原), 交流(機電船), 固体(機), コロイド(機土), 周波(電), 触媒(電原), 磁気(電), 磁場(原), 静電気(電), 短波(電), 長波(電土), 直流(機電船), 電気(電), 電極(機電船原), 電子(電原), 電磁波(電), 電波(電), 電流(電船計), 導体(機電船), 熱線(電), 波動(機土建), 不導体(機電), 分子(機原), 放射能(電建計原), 放電(電船原), 溶液(原), 陽極(電), 陽電気(電), 流体(機船), 良導体(電)

1.5110 化学成分 37語

亜鉛(機建船原), アルカリ(機), アルマイト(建), アルミニウム(機船原), アンモニア(機), 硫黄(機原), 塩(機), 塩素(原), カルシウム(原), 金(原), 銀(原), 元素(電原), 砂鉄(機), 酸(機), 酸素(機原), 硝酸(機), 水銀(原), 水素(原), 錫(機原), 青銅(機建), ソーダ(機), タングステン(機原), 炭素(機電原), 窒素(機原), 鉄(機建原), 銅(建原), ナトリウム(原), 鉛(建原), ニッケル(機原), 白金(機原), 非金属(機), 砒素(原), マグネシウム(原), マンガン(原), ラジウム(原), 硫酸(機電), 燐(原)

1.471 道路橋 35語

安全地帯(土), 石橋(土), 架線(電), 環状線(土), 幹線(土), 街路(土建), 軌道(機

電土建), 下水(機土建), 坑道(原), 国道(土), 小みち(土), 棧橋(機土建船), 支線(電土), 車道(土建), 上水(土), 水道(建船), 水路(機電土), 線路(電土), 堅坑(土原), 単線(電土), 地下道(電土), 通路(機土船), 釣り橋(機土), 鉄橋(土), 通り(土建), トンネル(電土), 道路(土建), 土橋(土), 橋(土), 複線(土), 踏切り(電土), 歩道(土建), 本線(電土), 露地(建), 渡し(機)

1.454 農工具など 33語

アイロン(建), オール(機船), 鍵(機建船), 霧吹き(機), 釘ぬき(機建), 工具(機船), こて(機土建), 磁針(電土船), 錠(機建), じょうご(機), 鋤(機), スコップ(機土), スパナ(機建), 旋盤(機建船), 洗面器(機船), 槌(機建), 鶴嘴(機土建船), 梃子(機電計), 電熱器(機建船), 砥石(機建), ねじまわし(機建), はけ(建), 針(計), ハンマー(機土建), ふいご(機建船), ふるい(機土建), 噴霧器(機船), ブイ(機電船), ブラシ(機電計), ペンチ(機電建), 万力(機電建船), やつとこ(機電船), 漏斗(機)

1.3850 設備・作業・手当て・処理 32語

換気(機土建船原), 外装(電土船航), 給水(機電土船原), 給湯(建), 切り盛り(建), 工作(船), 採光(土), 仕上げ(機土船), 修繕(建), 照明(機電船), 処理(原), 整備(航), 設備(電船), 装甲(船), 操作(機電), 操縦(機船), 装置(機電船), 装備(機船), 造作(建), 暖房(機土), 通気(建), 点燈(電), 取り付け(機電土船), 配管(機電土建船計), 排水(機電土), 配水(土), 配線(機電土船), 配電(電), 保守(航), 冷蔵(船), 冷凍(機船), 冷房(土)

1.412 木・石・金 29語

厚板(建), アルマイト(建), 板(機土建船), 鋳物(機), 瓦(土建), ガラス(機), 鋼(機建船), コルク(機建), コンクリート(機土建), 合金(機船原), すりガラス(機土), スレート(電土建), 石材(土建), セ

メント(機土建), 銑鉄(機船), タイル(土建), たたき(機建), 垂木(建), 鉄筋(機土建), 陶器(機), 庭石(建), 羽目(土建), ビーム(機電船原), ブリキ(機建), 丸太(土建), 木材(土建船), 床板(機建船), 用材(土建), 煉瓦(機土)

表8 『分類語彙表』への収載状況

『分類語彙表』 収載の専門語	『学術用語集』 の語数		
化学	792	10067	7.9
物理学	464	3568	13.0
地震学	382	2324	16.4
動物学	370	2076	17.8
天文学	356	2137	16.7
数学	342	1593	21.5
植物学	336	2505	13.4
海洋学	239	2389	10.0
気象学	216	1728	12.5
遺伝学	190	1926	9.9
分光学	122	2112	5.8
(理学系計	3809	32425	11.7)
建築学	948	6228	15.2
機械工学	933	8885	10.5
船舶工学	666	8586	7.8
電気工学	626	9819	6.4
土木工学	558	5642	9.9
原子力工学	256	3706	6.9
計測工学	216	2394	9.0
航空工学	168	3059	5.5
(工学系計	4371	48319	9.0)

### 5. どのような専門分野の用語が多いか

表8によれば、語数で見ると、理学系では化学が最も多く、物理学がそれに次ぐ。しかし、各『学術用語集』の用語数に対する割合で見ると、数学が最も高い。工学系では、語数で見ても、用語数に対する割合で見ても、建築学が最も多く、機械工学がそれに次ぐ。

以下、各専門分野ごとに、最も多くの専門語が所属する意味分野と、その専門語をあげる。

#### [理学系]

##### 化学 1.5110 化学成分 46語

亜鉛, 亜硫酸, アルカリ, アルコール, アルミニウム, アンモニア, 硫黄, 一酸化炭素, 塩, 塩酸, 塩素, 苛性ソーダ, カルシウム, 貴金属, 金, 金属, 銀, 軽金属, 元素, 醋酸, 酸, 酸素, 硝酸, 水銀, 水素, 錫, 青銅, タングステン, 炭酸, 炭素, 窒素, 鉄, 銅, ナトリウム, 鉛, ニッケル, 白金, 非金属, 砒素, 硼酸, マグネシウム, マンガン, ラジウム, 硫酸, 燐, 燐酸

##### 物理学 1.510 自然・物体・物質 29語

イオン, 陰極, 引力, 音波, 気体, 原子, 交流, 固体, コロイド, 触媒, 磁場, 長波, 直流, 電極, 電子, 電磁波, 電波, 電流, 導体, 熱線, 波動, 不導体, 物質, 物体, 分子, 放射能, 放電, 陽極, 流体

##### 地震学 1.510 自然・物体・物質 12語

引力, 液体, 交流, 固体, 磁場, 短波, 弾性体, 長波, 直流, 電極, 波動, 放射能

##### 動物学 1.574 膜・筋・神経・内臓 47語

胃, 鰓, 肝臓, 気管, 気管支, 気嚢, 筋肉, 頸動脈, 血管, 腱, 辜丸, 呼吸器, 骨膜, 鼓膜, 子宮, 消化器, 小腸, 小脳, 食道, 神経, 心臓, 十二指腸, 循環器, 静脈, 腎臓, 膀胱, 生殖器, 声帯, 脊髄, 胎盤, 胆嚢, 大腸, 大脳, 腸, 直腸, 動脈, 内臓, 粘膜, 脳, 肺, 扁桃腺, 膀胱, 盲腸, 網膜, 卵巣, リンパ腺, 肋膜

##### 天文学 1.521 天体 29語

天の河, 衛星, 海王星, 火星, 金星, 銀河, 恒星, 新月, 新星, 水星, 彗星, 星雲, 星座, 星団, 太陽, 月, 天体, 天王星, 土星, 伴星, 日, ほうき星, 星, 北極星, 満月, 冥王星, 木星, 流星, 惑星

数学 1.1912 整数・対数など 28語

因数, 函数, 奇数, 級数, 逆数, 偶数, 係数, 公倍数, 公分母, 公約数, 根, 小数, 実数, 乗数, 除数, 正, 整数, 素数, 対数, 帯分数, 倍数, 負, 分子, 分数, 分母, 変数, 無理数, 約数

植物学 1.553 枝・葉・花など 32語

枝, 雄しべ, 雄花, 核, 果実, 花粉, 株, 花卉, 萼, 茎, 紅葉, 莢, 種子, 子葉, 樹皮, 種, 地下茎, 蔓, とげ, 根, 年輪, 葉, 花, 花びら, 複葉, 節, 胞子, 実, 幹, 芽, 雌しべ, 雌花

海洋学 1.526 海・島 24語

暗礁, 磯, 入江, 海, 沿岸, 遠洋, 海岸, 海峡, 海洋, 近海, 群島, 崎, 珊瑚礁, 島, 諸島, 深海, 大陸棚, 地峡, 内海, 浜, 半島, フィヨルド, 岬, 湾

気象学 1.5153 雨・雪 15語

雨足, あられ, 空梅雨, 霧雨, 降雨, 降雪, 積雪, つゆ, 初雪, 梅雨, 雹, 吹雪, みぞれ, 雪, 雷雨

遺伝学 1.583 生理 10語

遺伝, 去勢, 交配, 交尾, 混血, 受精, 生殖, 妊娠, 発酵, 風媒

分光学 1.501 光 10語

可視光線, 螢光, 光線, 光沢, コロナ, 紫外線, 照明, 赤外線, 発光, 燐光

[工学系]

建築学 1.441 家屋 39語

東屋, アトリエ, 母家, 温室, 家屋, 格納庫, 宮殿, 議事堂, 庫裡, 公会堂, 校舎, 講堂, 金堂, 蚕室, 車庫, 書院, 食堂, 書庫, スタンド, 倉庫, 建て物, 塔, 燈台, 土蔵, 長屋, 納屋, 拝殿, 離れ, 平屋, 兵舎, 方丈, 本館, 本殿, 本堂, 望楼, むろ, 物置き, 櫓, 山小屋  
機械工学 1.454 農工具など 29語  
オール, 鍵, 霧吹き, 釘ぬき, 工具, こて, 錠, じょうご, 鋤, スコップ, スパナ, 旋盤, 洗面器, 槌, 鶴嘴, 梃子, 電熱器, 砥石, ねじまわし, ハンマー, ふいご, ふるい, 噴

霧器, ブイ, ブラシ, ペンチ, 万力, やっこ, 漏斗

船舶工学 1.466 乗り物(海上) 24語

筏, 碇, 舵, 旗艦, 汽船, 漁船, 甲板, 商船, ジャンク, 戦艦, 船室, 船首, 潜水艦, 船体, 船尾, はしけ, 帆船, 船, 帆, 母艦, 母船, マスト, ヨット, ランチ

電気工学 1.510 自然・物体・物質 27語

イオン, 陰極, 陰電気, 音波, 原子, 交流, 周波, 触媒, 磁気, 静電気, 短波, 長波, 直流, 電気, 電極, 電子, 電磁波, 電波, 電流, 導体, 熱線, 不導体, 放射能, 放電, 陽極, 陽電気, 良導体

土木工学 1.471 道路・橋 30語

安全地帯, 石橋, 環状線, 幹線, 街路, 軌道, 下水, 国道, 小みち, 棧橋, 支線, 車道, 上水, 水路, 線路, 堅坑, 単線, 地下道, 通路, 釣り橋, 鉄橋, 通り, トンネル, 道路, 土橋, 橋, 複線, 踏切り, 歩道, 本線

原子力工学 1.5110 化学成分 26語

亜鉛, アルミニウム, 硫黄, 塩素, カルシウム, 金, 銀, 元素, 酸素, 水銀, 水素, 錫, タングステン, 炭素, 窒素, 鉄, 銅, ナトリウム, 鉛, ニッケル, 白金, 砒素, マグネシウム, マンガン, ラジウム, 燐

計測工学 1.1961 円・尺・坪・合・匁など 15語

アール, インチ, オンス, 海里, ガロン, キログラム, 光年, トン, ノット, バール, ポンド, マイル, メートル, ヤード, リットル

航空工学 1.467 乗り物(空中・宇宙)

9語  
気球, グライダー, 航空機, ジェット機, 飛行機, 飛行船, プロペラー, ヘリコプター, ロケット

註

- [註1] 形容詞・形容動詞の派生名詞については、「高さ」「長さ」など少数の慣用例のみが(体の類)で取り上げられ、たとえば、「快さ」「長たらしさ」等は、(相の類)で取り上げられた「快い」「長たらしい」等の一種の活用形とみなされ、掲載されていない。
- [註2] <変化・動き>の258語は、段落によって細分されている。
- [註3] 1996年3月時点の版を、モニター用に公開している。公刊は2001年の予定である。

引用文献

- [1] 国立国語研究所資料集6『分類語彙表』
- [2] 宮島達夫・小沼悦(1992)「言語研究におけるシソーラスの利用」国立国語研究所報告104『研究報告集13』秀英出版
- [3] 国立国語研究所編(1994)『分類語彙表 フロッピー版』秀英出版
- [4] 宮川和・徳永健伸・田中穂積(1998)「格フレームを用いた情報検索」言語処理学会第4回年次大会発表論文集,PP.112-115
- [5] 中野洋(1995)「分類語彙表の増補とその利用」言語処理学会第1回年次大会発表論文集,PP.141-144
- [6] 柴田武編(1988)「VII. 語と意味」『日本語百科大辞典』大修館書店,PP.371-445
- [7] 日本電子化辞書研究所(1993)『EDR 電子化辞書仕様説明書』
- [8] 国立国語研究所報告21(1962)『現代雑誌九十種の用語用字』秀英出版
- [9] 阪本一郎(1958)『教育基本語彙』牧書店

- [10] Miller,G.A.(1985)'Wordnet:A Dictionary Browser' in Information in Data, Proceedings of the First Conference of the UW Centre for the New Oxford Dictionary. Waterloo, Canada: University of Waterloo.
- [11] Peter Mark Roget (1852) "Roget's Thesaurus" First edition, Longman
- [12] 大野晋・浜西正人(1981)『角川類語新辞典』角川書店
- [13] 田中穂積・仁科喜久子(1987)「上位/下位関係シソーラス IMIMAPI の作成 (I)」情報処理学会 Vol.87, No.84
- [14] 荻野綱男(1993)『現代日本語名詞シソーラスから見た語彙の意味分類』平成4年度科研費補助金研究成果報告書
- [15] 渡辺靖彦・黒橋禎夫・長尾真(1992)「IPAL 辞書と分類語彙表を用いた単語意味辞書の作成」情報処理学会第45回全国大会論文集,6F-8,PP.213-214
- [16] NTTコミュニケーション科学研究所監修(1998)『日本語語彙体系』岩波書店
- [17] 長尾真編(1996)『自然言語処理』
- [18] 川村和美・片桐恭弘・宮崎正弘(1994)「語を種々の観点から分類した多次元シソーラス」信学技報 NLC-94-48,PP.33-40
- [19] 野村雅昭・石井正彦(1988)『学術用語語基連接表』昭和62年度科研費補助金研究成果報告書
- [20] 野村雅昭・石井正彦(1989)『学術用語語基表』昭和63年度科研費補助金研究成果報告書

[21] 青戸邦夫 (1995) 「戦後半世紀の学術用語  
標準化の歩み」『学術月報』48-4

[22] 宮島達夫 (1981) 『専門語の諸問題』国立  
国語研究所報告 68, 秀英出版

(1999年12月16日受付)

## 著者紹介

### 柏野和佳子

東京女子大学卒業。  
富士通株式会社, 情報処理振興事業協会  
(IPA) を経て, 現在, 国立国語研究所  
言語体系研究部第二研究室 研究員。  
現代日本語の意味・文法の計量的研究に  
興味を持つ。計量国語学会, 情報処理学会,  
言語処理学会等 会員。  
E-mail: waka@kokken.go.jp

### 中野洋

京都府立大学卒業。  
国立国語研究所 言語体系研究部長を経て,  
現在, 国立国語研究所 日本語教育センター長。  
コンピュータによる日本語研究に興味を持つ。  
国語学会, 計量国語学会, 言語処理学会等 会員。

### 石井正彦

東北大学大学院 文学研究科 修士課程修了。  
国立国語研究所 言語体系研究部 第二  
研究室長を経て, 現在, 大阪大学大学院  
文学研究科 助教授。  
現代日本語の語彙・文法・文章に興味を持つ。  
国語学会, 計量国語学会, 専門用語研究会等  
会員。  
E-mail: ishii@let.osaka-u.ac.jp

# シソーラスと文献データベース Thesauri and Bibliographic databases

永井 賢吉<sup>†</sup>, 原田 郁子<sup>††</sup>

<sup>†</sup> 科学技術振興事業団 情報事業部、<sup>††</sup> 科学技術振興事業団 文献情報部

JICST シソーラスについて、用語の表記、同義語の扱い、階層関係の設定等の用語統制に対する考え方を述べた。シソーラスで規定される用語間の関係は普通なものではなく、どのような収録範囲のデータベースを対象とするかによりシソーラス毎の違いが生じることを論じた。また、ネットワーク上に分散して存在する各種データベース等のデジタル・コンテンツの統合検索に用いること目的として、JSTにおいて整備を進めているデータベース検索用大規模電子辞書について紹介した。

## 1. はじめに

近年シソーラスといった場合、情報検索システムにおいて、索引作業と情報検索で使用するキーワードとなる用語を統制し、用語間の相互関係を管理することで検索効率を高めることを目的とした用語典拠リストを意味するようになっている<sup>[1]</sup>。JOISでサービスしているJICSTファイルにはJICSTシソーラスがあり、NLM(米国国立医学図書館)が提供するMEDLINEにはMeSHがあるというように、シソーラスとは、データベース作成のため、あるいは検索のためのツールとして特定のデータベースごとに編さんされるものといえる。

JSTにおいても1975年に第1版のJICST科学技術用語シソーラスを完成させるとともに、JICSTファイルの索引をこのシソーラスに基づく統制語方式に改め、今日に至っている<sup>[2]</sup>。シソーラスの用語統制には用語の表記に関するもの、同義語の扱い、用語の表す概念の限定等がある<sup>[3]</sup>。本稿では、JICSTシソーラスを紹介するとともに用語統制に対する考え方を述べる。シソーラスで規定される用語間の関係は普通なものと考えられる。しかし実際は、対象とするデータベースの収録分野を反映し、シソーラス毎の違いが生じる。MeSHと対照することによりこれらの違いについてもふれる。

一方で、全文検索システムの広がりとともに

にあいまい検索を実現するため、システムの中にシソーラスを組み込むことも行われるようになってきている<sup>[4]</sup>。この種のシソーラスは索引作業に使用しないものなので、検索用シソーラスと呼ぼう。JICSTシソーラスやMeSHといった従来のシソーラスと比べると検索用シソーラスは同義語辞書に近いものであり、用語のリンクはあるものの用語の表す概念間の関係を明示するものではない。現在JSTにおいてもこの種の用語辞書——データベース検索用大規模電子辞書——の整備を行っている。これについて第4章で述べる。

## 2. JICST シソーラスの概要

JICSTシソーラスは、JICST科学技術文献ファイルを始めとするJSTがJOISで提供しているデータベースの索引、検索のために作成されたもので、科学技術分野全般を網羅している。1975年に第1版を完成させた後、これまでに改訂を重ね、現在の1999年版は第6版になる。当初は33,000語余りだった登録語数も、現在の99シソーラスでは43,314語となっている(表1)。

表 1: 1999年版 JICST シソーラスの登録語数

登録語数	43,314 語
ディスクリプタ	37,544 語
非ディスクリプタ	5,770 語

## 水質汚濁(スイシツオダク)

## KA01

- NT 海洋汚濁
- 河川汚濁
- 湖沼汚濁
- BT 環境汚染
- ・ 汚染
- RT 海底堆積物
- 河成堆積物
- 建築排水設備
- 湖沼堆積物
- 水質基準
- 底生生物
- 富栄養化
- 水処理
- 水処理薬品

図1. JICST シソーラスの登録語の例

図1にJICSTシソーラスの用語を示す。すべての登録語(見出し語)を五十音順に配列し、それぞれの見出し語についてふりがな、カテゴリーコード、スコープノート、階層関係等の用語間の相互関係といった情報を表示している。登録語はディスクリプタと非ディスクリプタの2つに大別される。前者が索引、検索に用いる用語であり、後者はディスクリプタを案内するための用語である。第6版のディスクリプタ数、非ディスクリプタ数を表1に示す。主題カテゴリーコードはその用語の分野を示すもので1~3個付与されている。図1の「水質汚濁」の場合のKA01は環境分野の用語であることを示している。記号NTの後に表示された用語が下位語、記号BTの後に表示された用語が上位語であり、それぞれ見出し語に対してより狭い概念、広い概念の用語である。「水質汚濁」の場合は「海洋汚濁」、「河川汚濁」、「湖沼汚濁」、「地下水汚濁」の4語が下位語としてシソーラスに収録されていることになる。2段階以上階層の離れた用語は中点(・)を用いて表示される。図1の例では、「水質汚濁」の上位語には「環境汚染」があり、さらにその上位概念として「汚染」が登録されていることを示している。

記号RTの後にある用語が関連語であり、「水質汚濁」の例では9語が登録されている。非ディスクリプタは見出し語の前に記号\*を付けて表示しており、記号USEを使用して索引、検索に用いる用語であるディスクリプタを表示している。非ディスクリプタとの関連を持つディスクリプタからは、記号UFを使用してその用語を案内している。なお、記号UFはused forの略である。

このような表示方法により、JICSTシソーラスでは、どの用語からもその上位語、下位語、関連語、USE参照語を把握することができる。

シソーラスは定期的に用語の追加、削除、及び修正等のメンテナンスを行う必要がある。索引作業を進める中で、新たな用語追加の必要性や階層関係の不具合等が明らかになるからである。JICSTシソーラスの場合もこれまで5~6年に一度の割合で改訂を行い、科学技術の進展に伴って新たに発生した用語等を追加している。また改訂の際には、シソーラスにすでに登録されている用語であっても、その用語が使われなくなった場合など、索引作業の実情に合わせて削除、修正等を行ってきた。

JICSTシソーラスの改定作業は、登録語の主題カテゴリーコードにより担当部署を決めて実施する。新規用語は、改訂作業前の過去2年間程度の期間にディスクリプタとしては未登録ではあるが主題を表すために必要として索引された用語(自然語)や、標題からの抽出語のその期間の出現頻度統計データなどをもとに、既登録の用語や追加候補語間の相互関係、索引・検索用語としての意義などを総合的に考慮して選定する。既登録の用語についても、JST内外のユーザからの指摘を踏まえた階層関係の見なおし、スコープノートや主題カテゴリーコードの修正を行っている。

JSTにおいては、シソーラス改定が実施されても、JICST科学技術文献ファイル等の検索が最新のシソーラスの用語で行えるよう配慮している。このため、改定のたびに計算機処理により旧シソーラスに基づいて索引され

たファイルを新シソーラスの内容と整合が取れるように処理(バックファイルメンテナンス)を行っている。バックファイルメンテナンス作業は1981以降に蓄積された記事を対象とするため、半年以上の期間を必要とする負荷の高い作業である。

JSTの内外のシソーラスユーザからは常々、JICSTシソーラスの改定の周期を短縮すべきであるという声を頂いている。シソーラス改定周期の短縮においては、上述のバックファイルメンテナンス作業が最も大きな障害といえる。シソーラスに基づいて索引されたデータベースの扱いまでを含めて改定作業を考えなければならないことは、シソーラスが他の辞典・用語集とは異なる点である。

### 3. 用語統制の考え方

#### (1) 用語の表記

シソーラスの規格<sup>[3]</sup>ではシソーラスに登録する用語の表記については名詞または名詞句を用いることとなっている。またJICSTシソーラスの場合、JICSTファイルの索引に用いることから、使用可能文字種の制限がある。漢字についてはJIS第1水準漢字に制限され、ローマ数字は用いることができない。このため、JIS第2水準漢字を含む用語等については使用可能な漢字やひらがなに置き換えて登録している。第6版のシソーラス改定においては、このような制限のもとで、用語の表記が複数ある場合には当該分野で最も受け入れられている表記を用いることとした。シソーラスに対する馴染みやすさ、使いやすさを考慮してのことである。

しかし、JICSTシソーラスにおいては一般の専門用語ではない用語や、通常専門語としてはあまり用いられない言葉も登録されている。例えば職種別従事者という用語をシソーラスに登録し、個々の職業・職種を表す用語をその下位語としている。これは用語や概念構造などを整理するためのものであり、このような用語により同一概念に含まれるいくつかの用語をまとめて表示することができる。

複合語については、シソーラスの規格では原則としてシソーラスに登録しないこととなっている。これはシソーラスの規模を押さえるためのルールである。JICSTシソーラスにおいても1987年版シソーラスまでは極めて長い複合語はディスクリプタとしては登録しないとしていた。

用語の表す概念が分野によって異なる等の同形異義語については、見出し語に限定句を付けて区別している。JICSTシソーラスでは限定句は墨付きパーレン(【,】)で表示しているが、検索システムの中では通常の丸括弧に置き換えている。シソーラスの規格においては、限定句はシソーラス登録語内の同形異義語について付与することとされているが、JICSTシソーラスにおいては、またアルファベット三文字以内で示されるアクロニムやカナ表記の用語等、対応する同形異義語がシソーラス用語でない場合にも限定句を付けている。

#### (2) 同義語の扱い

シソーラスにおいては同一の概念を表す用語が複数存在する場合、索引・検索に用いるディスクリプタを選定する必要がある。ディスクリプタは、最もよく用いられる用語はどれか、その用語が示す概念がどれだけ特定できるか等を総合的に検討して選定される。例えば、「ワイン」と「ぶどう酒」を取り上げてみる。JICSTシソーラスにおいてはぶどう酒がディスクリプタとなっており、利用者からは「ワイン」を採用すべきではないかとの意見をいただいたことがある。前者は外来語のかな表記であり、広く一般的に普及しているといえる。しかし、ブドウ以外の果実から作られたワイン、例えばキウイ・ワインという用語もある。「ワイン」をディスクリプタとする場合には、索引及び検索においてその意味範囲を明確にするため、スコープノート等が必要になるであろう。すなわち、ワインよりもぶどう酒の方が特定性の高い用語といえる。

概して外来語のカナ表記にはこの傾向があるといえる。ディスクリプタとして選定され



なかった用語は、シソーラスに登録されないか、または非ディスクリプタとしてシソーラスに登録されるかまたはシソーラスに登録されず自然語のままとされるかのいずれかである。前者の場合のディスクリプタと非ディスクリプタの関係を優先関係という。優先関係を設定し、シソーラスに非ディスクリプタとして登録するか否かは、その表記からディスクリプタが容易に連想できるか、その表記の出現頻度、冊子体としたときの配列順を考慮している。物理分野や電気工学分野の用語に Electric field というのがあるが、その日本語表記には「電場」と「電界」がある。前者の方が用いられる範囲が広いこともあり、JICST シソーラスのディスクリプタとなっているが、「電界」はシソーラスに登録されていない。電界から電場が容易に連想できること、配列が近いことが、その理由と言える。

また JICST シソーラスにおいては慣用されている読み(ふりがな)が複数ある場合、いずれの読みからも用語が探せるようにシソーラスに登録している。このようなディスクリプタの異音同表記語は非ディスクリプタとして登録されており、記号☆を付して表示している。

分野によって用語の表記が異なる場合の調整は容易ではない。用語の担当者はそれぞれの分野で用いられている用語をディスクリプタと選定する傾向があることによる。JST においては、以前に「無重力」と「無重量」のどちらをディスクリプタとするかについて、物理分野と機械分野の担当者間で議論が交わされたという話もある。

☆胃腔(イコウ)

LS16

USE 胃腔(イクウ)

胃腔(イクウ)

LS16

UF 胃腔(イコウ)

## 図 2. 異音同表記語の例

一般にシソーラスにおける優先関係は用語の同義関係を示すものである。しかし、優先

関係の中には、索引及び検索における約束と捉えるべきであり、言語の持つ意味においての同義関係ではない場合がある。シソーラスの規格においては「上位語の使用」ということが定められている。これは極めて特定のな用語等で、索引する頻度が低い用語について、その上位概念の用語で代用するものである。「横浜 USE 神奈川」といった関係である。索引においては横浜を表すために神奈川を使い、検索者は神奈川についての文献からブラウジングにより適合文献を探すことになる。ブラウジングができないほどに出現頻度が高い場合にはディスクリプタとしてシソーラスに登録することになるが、そうでない場合には地理的上位概念である用語で代用することを示している。この関係は横浜をとらえる別な観点、例えば「横浜 USE 都会」といった関係が類推可能な場合に、利用者に正しいディスクリプタを案内する役目を有する。言いかえれば、索引規則を与える関係でもある。

JICST シソーラスにおいては、ディスクリプタの規模を制限するため、これまで索引頻度の低いディスクリプタを非ディスクリプタに移行し、その代わりに新語をディスクリプタとして登録するという改定を行ってきた。しかし、第 6 版の改定作業においては、このような関係を収録するツール(自然語と統制語の対応関係を収録)が別途用意されていたこともあり、改定規則において USE-UF 関係を同義関係に制限した。この結果、非ディスクリプタの語数が大幅に減少することとなった。ただし、改定作業においてこの規則が厳密に適用されたかというところではない。

JICST シソーラスは科学技術全般をその対象としているが、それでも周辺領域はある。主題カテゴリーコードの共通分野がそれである。会議の例で考えてみる。類似の概念を表す用語には「ワークショップ」、「パネルディスカッション」、「ミーティング」といった用語がある。JICST ファイルの索引を考えた場合、これらの概念を区別することにどれほどの意味があるだろうか。また、索引および検索においてこれらの概念を正しく使い分ける

ことができるだろうか。JICSTファイルの検索という目的に照らして見ると、これらの概念は会議に名寄せしたほうが便利であるとの結論に達する。専門用語であっても例外ではない。図3の「乗物」と「交通機関」の関係についても同様であろう。「乗物」といった場合、すぐに思いつくものは電車やバスといった用語であるが、船や馬車を意味する場合もある。これらの優先関係は同義関係ではなく索引規則を与える関係と考えるべきである。

\*乗物(ノリモノ)

MN01

USE 交通機関

交通機関(コウツウキカン)

MN01

UF 乗物

図3. 優先関係の例

### (3) 階層関係

階層関係とは、ディスクリプタ間の上位と下位を表すもので、ある語の上位語とは、その語よりもより広い意味を持つ概念を表し、ある語の下位語とは、その語よりもより狭い意味を持つ概念を表す。JICSTシソーラスにおいては、用語の上位-下位の関係は、常に成り立つ包摂関係の場合に設定することとしている。例えば、正面衝突と交通事故という用語を取り上げる。JICSTファイルでは、原子核の衝突現象の記述においても「正面衝突」という用語が出現し、索引に使う必要がある。しかし「正面衝突 BT 交通事故」という階層関係を設定した場合、正面衝突は交通事故の下位概念に限定されるため、原子核の衝突現象の主題表現には用いることができない。このような場合、どちらの正面衝突にも索引できるように配慮するため、階層関係は設定しない。しかしながら、このように階層関係を設定するには、「正面衝突」という概念が交通事故に限定されないことを知っていなければならない。概して不適切な階層関係は、シソーラス改定後の日々の索引作業の中で見つかることが多い。

また、階層関係が2段階以上となる場合、最上位語から見たとき、厳密な包摂関係とは

言えない用語がその階層に含まれることが少なからず生じる。例えば、「天体望遠鏡」の下位語に「電波望遠鏡」がある。この二つの間では厳密な包摂関係が成り立つ。また同様に「天体望遠鏡」の上位語は「望遠鏡」、更にその上位語は「光学器械」である。これらの間には包摂関係が成り立つが、電波望遠鏡は厳密には光学器械ではない。このため、用語関係の厳密性から考えると「望遠鏡」を「光学器械」の下位に置くことの問題点が考慮されるべきではあるが、検索ツールとしてのシソーラスの役割を考えた場合は、この2語の関係付けは必要とする考え方も否定できない。シソーラスで規定される階層関係は、その対象とするデータベースの検索における利便性が優先されるといえる。

このような階層関係における矛盾は、用語間の階層関係の設定がどのような観点でなされているかということから生じる。例えば「電磁波 NT 光」、「光 NT 太陽光」という階層関係を考えてみる。前者は物理学的な関係であるが、後者は発生源による細分化である。多段階の階層が異なる観点で設定されている場合には、最上位語からみた関係(この場合は「電磁波 NT 太陽光」)が不自然となることがあるといえる。

シソーラスの階層関係はその対象とするデータベースの収録範囲を反映するものである。これを見るためにJICSTシソーラスとNLM(米国国立医学図書館)のMeSHとを対照して見よう。MeSHは医学分野の専門シソーラスであるが、1960年に第1版が創刊され、1963年に第2版、その後は毎年改訂されている。2つの用語「虫垂炎・Appendicitis」と「救急車・Ambulances」の階層構造をそれぞれ図4、図5に示す。これを見ると、病名である「虫垂炎・Appendicitis」については、JICSTシソーラスとMeSHとでほぼ同様の構造を示しているのに対し、「救急車・Ambulances」の場合は、階層構造に異なる部分が見られる。JICSTシソーラスが自動車としての観点のみでの構造になっているのに対し、MeSHは医療サービスの階層も持たせ

ている。これは、MeSHによって索引されるデータベースにおいては、救急車を地域の医療体制といった問題を論じる文献の索引語の1つと見なしていることを意味する。もちろんJICST科学技術文献ファイルでは自動車の構造といった文献も対象としており、JICSTシソーラスでは救急車は自動車という観点の階層のみになっている。

虫垂炎(チュウスイエン)

- BT 腸炎
- ・ 炎症
  - .. 病気
  - ・ 腸疾患
  - .. 胃腸疾患
  - ... 消化器疾患
  - .... 病気

Diseases(疾患)

- Digestive system Diseases(消化器系疾患)
- Gastrointestinal Diseases(胃腸疾患)
- Intestinal Diseases(腸疾患)
- Cecal Diseases(盲腸疾患)
- Appendicitis(虫垂炎)

図4. 階層関係の比較(1)

上側がJICSTシソーラスの階層、下側がMeSHの階層(上位語から展開)である。

救急車(キュウキュウシャ)

- BT 特用車
- ・ 自動車

Technology, Industry, and Agriculture  
(技術, 産業, 農業)

Motor Vehicle(乗物)

Ambulances(救急車)

Health Care Facilities, Manpower, and Services(保健医療施設人的要員とサービス)

Health Services(保健医療サービス)

Emergency Medical Service(救急医療サービス)

Transportation of Patients(患者輸送)

Ambulances(救急車)

図5. 階層関係の比較(2)

上側がJICSTシソーラスの階層、下側がMeSHの階層(上位語から展開)である。

(4) 関連関係 関連関係とは、階層関係や優先関係以外の関係である。関連関係は、意味が部分的に重複している場合のように厳密な階層関係が成り立たない場合や、派生関係がある場合、また一方の語が他方の語を強く含意する場合などに付与する。

関連語も用語を案内するというシソーラスの機能の一つである。ある用語に対して、文字通りの意味での「関連語」がどの程度シソーラス中に含まれているのかを示し、適切な用語を案内する働きである。ある用語が上位語や下位語を持つ用語の場合はそれらから案内される上に、関連語からも案内されるので、より容易に適切な用語にたどりつくことができる。しかし階層関係を持たない孤立語の場合は、関連語からしか案内されない。従って、特に孤立語の場合は、関連語を付与することが推奨されている。

ヒト(ヒト)

LS05

RT 動物

ほ乳類

霊長類

化石人類

人間

妊婦

民族

図6. 「ヒト」の登録内容

また、本来ならば階層関係にある用語であっても、それが索引された時に対象を明確にするために、意図的に関連語として扱う場合もある。この観点から、「ヒト」という用語は、「動物」「ほ乳類」「霊長類」などからの上位-下位関係を外し、孤立語として扱っている(図6)。

#### 4. 分散型デジタルコンテンツ・統合システムと大規模電子辞書

インターネットの普及に伴い、ネットワーク上に多種多様な情報が氾濫する中、信頼できる情報の入手が困難になっている。このため、ネットワーク上に分散して存在する有用情報を統合的に案内し、あたかも1つのデータベースを利用しているかのように検索できるシステムが必要と考えられる。現在JSTにおいて開発を進めている分散型デジタル・コンテンツ統合システムは、このようなコンセプトに基づき、ネットワーク上に分散する有用なデータベース等を案内し、総合的に検索するシステムである。JSTではこのシステムをもとに科学技術情報をターゲットとしたポータルサイトを構築することを目指している。

分散型デジタル・コンテンツ統合システムにおいては、データベース毎に検索用語、言語が異なることによる検索の不具合を解消するため、英訳、同義語等が整備された大規模な辞書—データベース統合検索用大規模電子辞書—を組み込むことを予定している。

### (1) 大規模電子辞書

データベース統合検索用大規模電子辞書の構築にあたっては、各種科学技術関連データベースの検索に役立つ基本的な用語及び市販の辞書等に掲載されていない最新用語を網羅的に収集し、延べ150万語程度の規模の辞書の構築することを基本的考えとしている。用語の収集はJSTが所有しているJICST科学技術文献データベースの標題中に出現する用語から行うこととし、以下の技術的要件を設定している。

(イ) 辞書は日本語と英語の両方の見出し語を持つ

日英の用語がペアとなった辞書を分散型デジタルコンテンツ・統合システムに搭載することにより、日本語(英語)の検索用語からも英語(日本語)のデータベースを検索できることを目指している。

(ロ) 同義語が整理された辞書であること

従来のシソーラスにおいては同義語の整理とは、用語の表す概念が明確にした上で複数

の同義語の中から索引・検索に用いる用語をディスクリプタに設定することである。しかし、ネットワーク上に分散する各種データベースを統合検索するために用いる本辞書の場合はそのような用語統制はできない。表記の違いも含めて検索用語となりうるすべての同義語をグループ化する必要がある。ただし、日本語の送りがな、外来語の表記の違いや英語の見出し語の場合で単数・複数、語尾変化による表記の違いは異なり語としないこととしている。

(ハ) 見出し語同士で関連関係がある用語については3語程度の関連関係を付加する

(ニ) 上位語の付加(原則1階層のみ)

従来のシソーラスにおいては、その機能の1つとして、冊子体での利用を前提に用語を案内することが必要であり、そのため階層関係や関連関係を明示し、目的の用語に到達できるよう配慮している。本辞書は分散型デジタル・コンテンツ統合システムのフロントエンドや他の検索エンジンに組み込まれるものであり、このような冊子体で用語を探すための利用は想定していない。検索者が入力した用語をこれらの関連関係、上位関係により、検索語ある程度膨らませること、または検索候補語を表示させることを目指している。

### (2) 大規模辞書と従来のシソーラスの違い

(イ) 用語の意味限定、表示

従来のシソーラスは、あるデータベースの索引語がどのような意味で用いられるかの概念を明確に限定することが求められ、従ってそれを明示することが必要であった。このため、従来のシソーラスにおいてはスコープノート、限定句などにより概念の明確化を行っている。大規模辞書はネットワーク上に分散して存在する複数のデータベースの統合検索に利用するものであり、限定句により用語の表す概念を限定することはできない。また、スコープノートの付加も意味がない。大規模辞書においては、単に用語のリンクが張られているのみであり、同じ表記で意味の異なる

語(同形異義語)についても同様である。同形異義語については何らかの整理が必要であろうということで、内部的には分野を示すコードを持たせ、識別することとしている。

用語数が150万語程度の規模となるとその内容を表示させることが問題となる。従来のシソーラスのように冊子体を作成することは非現実的である。さらに、この辞書は検索用シソーラスであり索引には用いないため、索引ツールとしての参照機能は不用である。検索側での参照機能、例えばデータベース検索の際にどのような用語で実際に検索が行われるかを表示する機能等については、この辞書を組み込む検索システムの問題として、大規模辞書自体では参照機能を持たないこととしている。

#### (ロ) 階層関係

大規模辞書の階層関係は、用語をグループ化し、整理するためのものであり、従来のシソーラスで規定されているような深い階層関係を設定することは考えていない。これは、大規模辞書を索引には用いないことを前提としていることが大きい。シソーラスに基づく索引の場合、文献の表す内容に合った、より特定性の高い用語を索引することが必要である。このため、従来のシソーラスにおいては、適切な用語を案内するため、上位の概念から下位の概念へと系統的に整理することが必要であった。検索においては、このような索引の原則のもと、シソーラスを調べることにより、そのデータベースでは索引されている用語の表す概念の範囲がどのようなものかを知ることができる。

大規模辞書は索引における用語の使い方の異なる複数のデータベースの検索をサポートするためのものである。データベースにより階層関係が異なる場合があること、さらには統制語索引を採用していないデータベースも同時に検索する必要のあることから、大規模辞書では深い階層を用意しても、網羅的検索が保証できないといえる。もちろん検索の際に有用な場合もあるが、索引(すなわちデータベース)と独立した大規模辞書においては深

い階層はそれほど重要な意味を持たないといえる。

#### (ハ) 辞書のメンテナンス

従来のシソーラスについては、新規用語の出現や用語の使用法の変化に対応するため、メンテナンスが必要である。大規模辞書については、現時点ではメンテナンスを想定していない。これは以下の理由による。

- [1] 使われなくなった用語を排除する必要がない。
- [2] 新語は複合語が多い。
- [3] 用語関係をメンテナンスする基準が不明確。

現在辞書構築作業を進めているところであり、完成は平成12年3月の予定である。

## 5. おわりに

インターネットをはじめとする情報技術の近年の進歩により、情報検索を取り巻く環境も著しく変化している。従来の大規模データベースのみではなく、それぞれの研究機関がWeb上に独自に提供している小規模のデータベースにアクセスできるようになっている。これらのデータベースは検索にシソーラスを必要としない。また大規模データベースにおいても、検索コマンドやシソーラスを使わないことを前提としたインターネットからのアクセスが増加しつつある。検索ツールとしてのシソーラスの重要性を否定するものではないが、大規模データベースに適用され、索引者と検索者の橋渡しをする従来型のシソーラスの役割は相対的に低くなりつつあるのではないだろうか。従来型のシソーラスにおいては、索引における使い易さや対象となるデータベースの収録範囲を考慮して用語間の関係が定められている部分が少なからずある。今後は検索用シソーラスの必要性が大きくなると思われるが、この種のツールの構築に当たっては用語間の関係をどのように決めるべきかが今後の課題といえる。

## 参考文献

- [1] (財) データベース振興センター編「データベース白書 1998」
- [2] 斎藤和男, 坂上安彦, 戸塚隆之, 岡野弘行, 白山邦彦, 五味淵亘. 「JICST 科学技術用語シソーラス 1987年版の作成」 情報管理, Vol.29, No.11, p.952 - 964(1987).
- [3] JIS X 0901 「シソーラスの構築及びその作成方法」 (1991)
- [4] 全文検索システム協議会 97年度活動報告 < <http://www.asahi-net.or.jp/~zc7t-urb/dbtokyo98/db98.html> >

(1999年10月27日受付)

## 著者紹介

### 永井賢吉

1980年日本科学技術情報センター入社。  
1991年まで物理関係の索引業務に従事。  
科学技術庁へ出向の後、  
1993年から JICST シソーラス等のメンテナンス業務を担当。  
1999年より情報事業部。

### 原田郁子

1992年、日本科学技術情報センター入社。  
物理関係の索引業務に従事。  
1999年より JICST シソーラス等のメンテナンス業務他を担当。

論文

## 意味関係抽出手法統合による概念の体系化

## Structurelization of Concepts by Integration of Semantic Relationships

伊東 千夏<sup>†</sup>, 宇陀 則彦<sup>†</sup>, 石塚 英弘<sup>†</sup>, 藤原 譲<sup>††</sup><sup>†</sup>図書館情報大学<sup>††</sup>神奈川大学 理学部 情報科学科

本論文は知識資源の自己組織化手法の一つとして、複数の意味関係手法を統合した概念の体系化について述べる。本論文ではEDRの日本語専門用語単語辞書を対象に、造語規則を利用して階層関係を抽出するとともに対訳を利用して同義関係を抽出し、それぞれの結果を統合した。本論文では概念体系を構造としてとらえ、様々な原情報から複数の手法で意味関係を抽出し、構造を自己組織的に変化させて徐々に充実させた。本論文によって、構造中心の手法が概念の体系化に有効であることを示した。

## 1 はじめに

近年、辞書、シソーラス、フルテキストなどの各種コーパスやCD-ROMなどの電子本、さらにWebページなどの電子資料が急激に増加しており、電子資料を活用するための技術開発が盛んに行われている。しかしながら、電子資料をさらに高度に利用するためには、処理ツールの開発だけでなく、個々の電子資料を意味的に統合し、知識資源として利用できるよう組織化する必要がある。そのためには概念間の意味関係を原情報の意味を損なわない形式で体系化することが求められる。

概念体系を構築するためには意味関係の記述が必須であり、これまで多くの試みがなされてきた。それらは大きく人手による方法とコンピュータによる方法に分けられる。人手による方法は、意味関係を正確に記述できるが、時間と手間がかかる点が大きな欠点である。一方、コンピュータによる方法は、時間と手間はかからないものの、意味関係の正確な記述が困難である。どちらの方法も困難なことに変わりないが、増加する資料に応じて概念体系を更新したり、利用目的に応じて概念記述を柔軟に変化させることが必要になってきている現状を考えるとコンピュータによる方法を選択せざるをえない。

コンピュータによる自動化手法で問題になるのは、記述の詳細化とエラーの回避である。これらは意味記述の本質的問題であるので人手による方法でも同じ問題を抱えているが、コンピュータによる方法に対してより顕著に現れる。特にエラーには敏感で、少しでも間違った意味関係があると、構築された全ての概念体系および手法が否定されてしまう。あるレベルまでは詳細に記述できるものの、さらに詳細に記述しようとしたら、エラーを取り除こうとするとすぐに複雑な処理が要求され、修正処理が本処理より多くなり、本来の目的とはかけ離れていった。

概念体系は非常に多くの意味関係が複雑に結合しているため、一つの手法で意味関係を正確に記述するのは無理があり、全ての意味関係を抽出することはもちろん、エラーを含まないで抽出することは意味関係を限定したとしてもほとんど不可能である。これまでの研究は一つの手法だけで正確な記述を目指すことが多かった。ひとつの手法でうまく抽出できなければ、その手法と結果を捨てて、別の手法が新たに考えられた。このような処理と正確さを重視した手法では、ある一定のレベルまでしか記述できない。

本研究では概念体系は常に不完全という視

点に立ち、一つの手法だけで意味関係を記述しようとせず、複数の手法を組み合わせることによって徐々に概念体系を充実させることを考える。ある一つの手法で抽出された意味関係は最終結果ではなくある時点における概念体系であり、その概念体系は構造情報として次の概念体系構築に利用される。概念体系は常に意味に直結した構造としてとらえ、単純な処理結果とはみなさない。意味関係を構造として記述することによってはじめて、意味関係を断片的な集合ではなく、全ての意味関係が有機的に結合した概念体系として処理できる。このように概念に内在する意味に応じて自律的に概念を構造化することを自己組織化と呼ぶ [1] [2] [3]。

本研究は知識資源の自己組織化手法として、同義関係を抽出する C-TRAN 法 [1] [4]、主に階層関係を抽出する SS-KWEIC 法 [2] [5] [6] [7]、新しい関係自身を発見する SS-SANS 法 [8]、意味記述を利用する SANS 法などの意味関係抽出手法およびこれら意味関係手法を統合化する INTEGRAL 法を開発している。本論文は知識資源の自己組織化のひとつとして、EDRの日本語専門用語単語辞書 [9]を対象に、SS-KWEIC 法と C-TRAN 法を統合した INTEGRAL 法について述べる。本論文によって、構造中心の手法が概念体系構築に有効であることを示した。

## 2 意味関係の抽出

### 2.1 SS-KWEIC 法

SS-KWEIC 法 (Semantically Structured Key Word Element Index in terminological Context) は専門用語の造語規則に基づいて、概念の上位と下位の関係である階層関係を抽出する方法である。日本語では語基や接辞の組み合わせにより新しい単語が作成されることが多く、特に専門用語は用語の表す概念を明確に表現するため、新語を作るのではなく合成語により新しい用語が作成されることが多い。合成語の構成にも”山岳”のような並列や”

山々”のような重複、”山桜”のような修飾など複数の種類がある。

```

<語> ::= <単純語> | <合成語>
<合成語> ::= <複合語> | <派生語>
<複合語> ::= <語基> | <語基> + <複合語>
<派生語> ::= <接辞> + <合成語> |
               <合成語> + <接辞> |
               <合成語> + <接辞> + <合成語>
<接辞> ::= <接頭辞> | <接尾辞>
    
```

図 1: 造語規則

いずれの種類にしる、多くの専門用語では語基の数が多くなるにつれて規定する意味の範囲が狭くなり、より小さい概念を表現しているといえる。このことから、SS-KWEIC 法では語基が含まれている単語の長さ順に階層関係にあるとみなす。図 1 に BNF 記法による造語規則を示す。ただし、”+”は語の連結を意味する。階層関係を抽出するためには、文字列マッチングにより語基を含む単語を抽出する必要がある。このとき、語基をパターン文字列と呼ぶ。文字列マッチングの結果、パターン文字列を含む単語が下位語として抽出される。

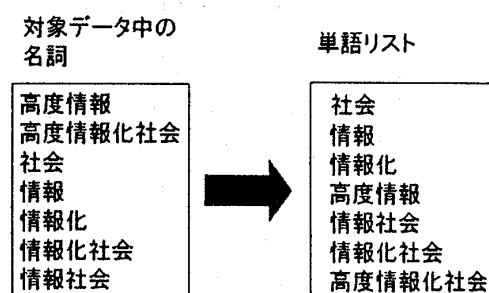


図 2: 単語リストの作成

次にプログラム処理について説明する。まず、単語リストを作成する。対象データより名詞単語を取り出し、文字列の長さを第1キー、文字コードを第2キーとして、ソートを行なう。ソートされた結果が単語リストとなる。このリスト中の単語は合成語の語基となりうるのでパターン文字列とし、パターン文字列を含む単語を探すために単語リスト内でマッチングを行う。なお、パターン文字列を含



む単語が下位語となるので、マッチングはパターン文字列よりも文字列の長い単語に対して行われる。文字列マッチングの結果、パターン文字列を含む単語が下位語として抽出される。抽出結果は左辺がパターン文字列となった単語、右辺がパターンを含む単語の2項関係として出力する。

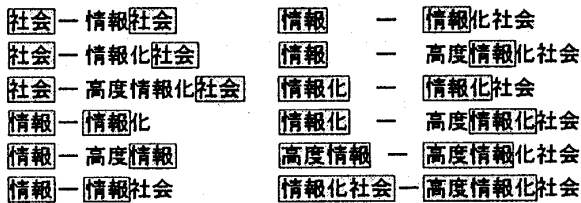


図3: 文字列のマッチング及び2項関係の作成

図3ではマッチングした全ての結果を2項関係として記述している。構造化する際の階層関係の記述は一番詳細な関係とし、推移的に記述する。そのため、階層を越える関係は削除する。

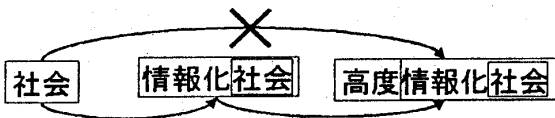


図4: 階層を越えた関係の削除

図4では、「社会」の下位語として「情報化社会」、「高度情報化社会」が抽出され、「情報化社会」の下位語として「高度情報化社会」が抽出されている。しかし「社会→高度情報化社会」は階層を越えた関係の記述であり、「社会→情報化社会、情報化社会→高度情報化社会」を用いて推移的に記述できるため、「社会→高度情報化社会」は削除する。階層を越えた関係の記述を削除することにより階層関係の抽出は終了する。

## 2.2 C-TRAN 法

C-TRAN 法 (Constrained Transitive Closure) は対訳を利用して同義関係の集合を抽出する手法である。同義関係とは、ほぼ同じ概念を持つ単語間の関係をいう。対訳を付与

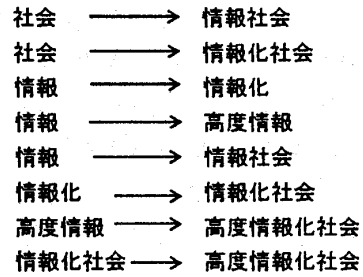


図5: 階層関係の抽出

する際には、同一またはほぼ同一の概念をもつ訳語が付与される。そこで、同じ訳語が付与されている単語を集めた集合に属する単語は同一の概念を表していると考え、同義関係として抽出する。

同義関係の抽出のために対象データから日本語と英単語の対訳リストを作成する。

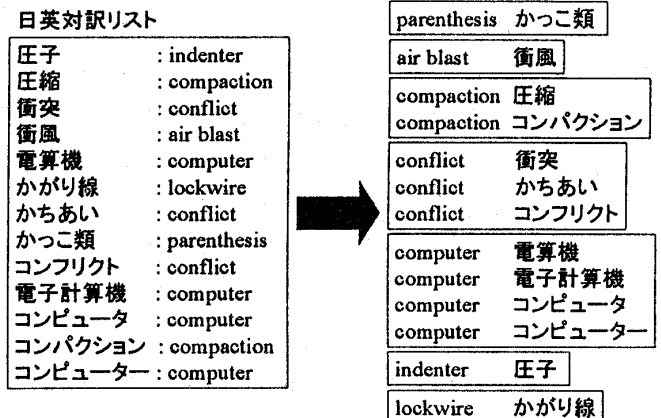


図6: 単語集合の作成

作成された単語 - 英単語の対訳リストから同じ英単語を持つ単語の集合を作る。同じ単語集合に属する単語は英単語を基準とした同じ概念をもつ単語と考えられる。このため同じ単語集合に属する単語を同義語集合として抽出する。

次に、表記ゆれによるノイズを減らす。同義語集合のなかに、{-, /, -, ·, -, ア, イ, ウ, エ, オ}の記号・文字を含む単語がある場合には、該当の記号・文字を除去する。記号・文字を除去した単語と一致する単語が同じ同義語集合内にある場合は、表記ゆれがあると判断する。表記ゆれのある単語群の中で代表となる単語

圧子	:	
圧縮	:	コンパクション
衝突	:	かちあい, コンフリクト
衝風	:	
電算機	:	電子計算機, コンピュータ, コンピューター
かがり線	:	かがり線
かちあい	:	衝突, コンフリクト
かっこ類	:	
コンフリクト	:	衝突, かちあい
電子計算機	:	電算機, コンピュータ, コンピューター
コンピュータ	:	電算機, 電子計算機, コンピューター
コンパクション	:	圧縮
コンピューター	:	電算機, 電子計算機, コンピューター

図 7: 同義関係の抽出

圧子	:	
圧縮	:	コンパクション
衝突	:	かちあい, コンフリクト
衝風	:	
電算機	:	電子計算機, コンピュータ, コンピューター
かがり線	:	かがり線
かちあい	:	衝突, コンフリクト
かっこ類	:	
コンフリクト	:	衝突, かちあい
電子計算機	:	電算機, コンピュータ, コンピューター
コンピュータ	:	電算機, 電子計算機, コンピューター
コンパクション	:	圧縮
コンピューター	:	→ コンピューター

図 8: ノイズの除去

を1語定め、そのほかの単語は代表となる単語を参照する。これにより表記ゆれをまとめられる。

### 3 意味関係の統合

以上の処理により SS-KWEIC 法、C-TRAN 法による階層関係、同義関係が抽出される。しかし、それぞれの抽出結果は独立したものであり、構造化されていない。そこで INTEGRAL 法 (Integration of Domain Established Knowledge) により SS-KWEIC 法と C-TRAN 法を統合し、抽出された階層関係と同義関係を構造化する。

SS-KWEIC 法、C-TRAN 法により「遠隔端末」、「リモート端末」の階層関係、同義関係を抽出した結果を図 9、図 10 に示す。SS-KWEIC 法は C-TRAN 法では抽出できない単語の階層関係の構造を抽出し、C-TRAN

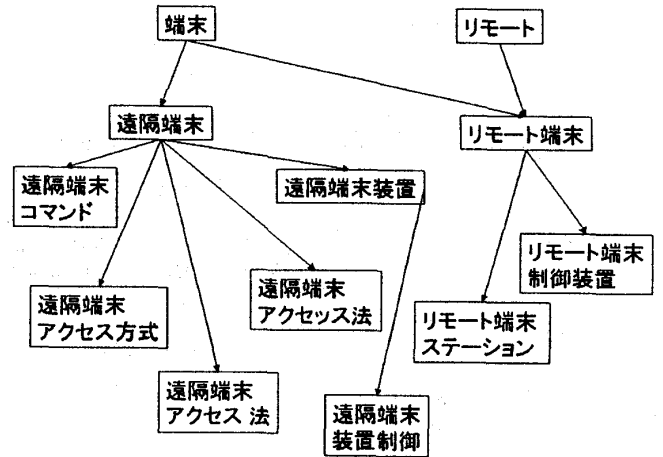


図 9: SS-KWEIC による階層関係の抽出

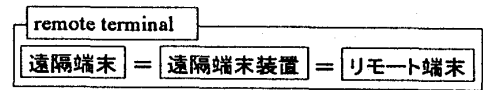


図 10: C-TRAN 法による同義関係の抽出

法では SS-KWEIC 法では抽出されない同義関係の集合を抽出する。

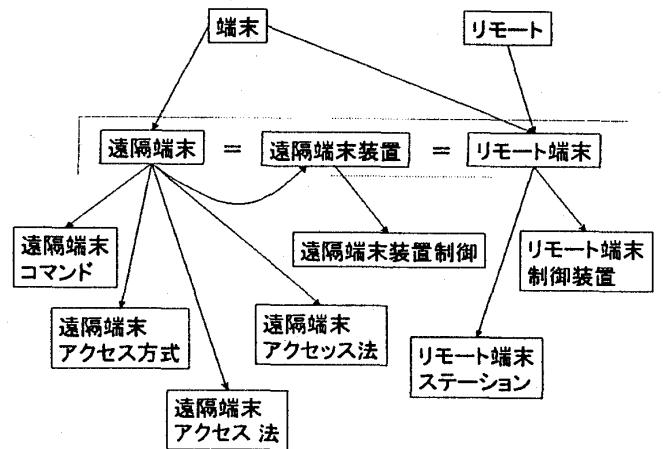


図 11: SS-KWEIC 法、C-TRAN 法による結果の組み合わせ

この2つの意味関係抽出結果を組み合わせたのが、図 11 である。これらを組み合わせたところ、「遠隔端末」と「遠隔端末装置」は階層、同義関係の両方の関係で抽出されていることがわかる。このように階層関係と同義関係を持つ場合、意味的には同義語だと判断できるので階層関係を消去し、同義関係を優先する。

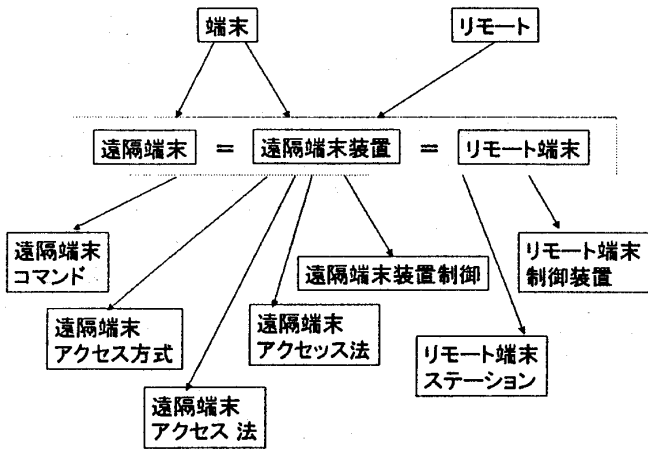


図 12: 統合結果

統合した結果が図 12 である。統合することにより同義関係にある「遠隔端末」、「遠隔端末装置」、「リモート端末」を1つのノードとしてとらえられる。そのノードを用いて上位・下位関係にある単語を参照すると、今までは参照できていなかった単語間の階層関係が抽出できる。図 12 では「リモート」は「遠隔端末」を下位語として、「遠隔端末」は「リモート端末制御装置」を下位語としてとらえている。

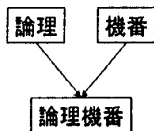


図 13: 論理機番 : SS-KWEIC 法による階層関係抽出結果

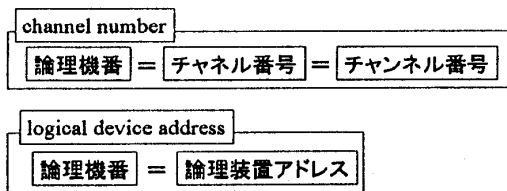


図 14: 論理機番 : C-TRAN 法による同義関係抽出結果

また、統合することにより抽出された階層関係が、どの概念に対して上位関係にあるかがわかるようになる。図 13、図 14 より「論

理機番」は、SS-KWEIC 法で階層関係を抽出した結果、上位語として「論理」、「機番」が抽出される。C-TRAN 法で同義関係を抽出した結果、「channel number」、「logical device address」を基準とした2つの同義語集合に属することから、「論理機番」には「channel number」、「logical device address」の2つの概念が存在することが明らかになる。SS-KWEIC 法、C-TRAN 法で抽出された結果を統合することにより、統合された構造の中で「論理機番」は、「channel number」、「logical device address」という2つの概念を表す単語であり、「channel number」、「logical device address」という2つの概念は、それぞれ「論理」、「機番」を上位関係に持つことがわかる。

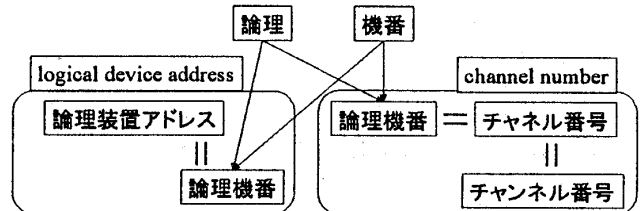


図 15: 「論理機番」: 統合した結果

以上のように、意味関係の統合を行なうことにより、1つの意味関係抽出法の結果だけでは抽出できなかった単語間の関係を抽出できる。また、1手法により抽出された結果の中で誤りが存在する場合、他の抽出結果と統合されることにより誤りが修正される。

## 4 実行環境および実データ

本研究の実行環境は、SUN Enterprise 3000(CPU: 250MHz × 2, メモリ: 512MB)、使用言語は C 言語である。処理に要した時間を表 2 に示す。

表 2: 処理時間

	処理時間	データ数
SS-KWEIC 法	32 時間	117,206
C-TRAN 法	1 時間	117,583
INTEGRAL 法	35 秒	234,789

(単語リスト)

- 9501 遠隔端末
- 33623 遠隔端末装置
- 33624 遠隔地バッチ
- 57806 遠隔端末装置制御
- 68335 遠隔端末アクセス法
- 77678 遠隔端末アクセス方式
- 77679 遠隔端末アクセス法

(データ形式)

ID 単語

<parent> 上位関係にある語の ID:上位関係にある語,

<children>下位関係にある語の ID:下位関係にある語,

(例)

9501 遠隔端末

<parent> 2042:端末,

<children>33623:遠隔端末装置, 57805:遠隔端末コマンド, 68335:遠隔端末アクセス法,  
77678:遠隔端末アクセス方式, 77679:遠隔端末アクセス法,

図 16: SS-KWEIC 法 : 単語リストおよび抽出結果

(単語リスト)

- |      |      |                            |
|------|------|----------------------------|
| 9500 | 遠隔測定 | remote measuring telemetry |
| 9501 | 遠隔端末 | remote terminal            |
| 9502 | 遠隔地側 | remote site                |

(データ形式)

ID 単語 英単語 同義語と抽出された語 1 同義語と抽出された語 2

(例)

- |      |      |                            |              |               |
|------|------|----------------------------|--------------|---------------|
| 9500 | 遠隔測定 | remote measuring telemetry |              |               |
| 9501 | 遠隔端末 | remote terminal            | 32970:リモート端末 | 33623:遠隔端末装置  |
| 9502 | 遠隔地側 | remote site                | 3487:遠隔側     | 45636:リモートサイト |

図 17: C-TRAN 法 : 単語リストおよび抽出結果

SS-KWEIC 法で用いた単語リストと抽出結果を図 16 に示す。単語には ID を付与しており、この ID は SS-KWEIC 法と C-TRAN 法の両方で識別子として用いられる。C-TRAN 法で用いた単語リストと抽出結果を図 17 に示す。単語リストは ID、単語、英単語から構成されている。データ出力時に同義語がない場合には空白にする。図 18 は SS-KWEIC 法と C-TRAN 法の統合結果である。統合結果は ID と単語から検索できる。

## 5 考察

### 5.1 処理結果の評価

SS-KWEIC 法と C-TRAN 法の処理対象には EDR1.5 版 日本語専門用語単語辞書 (情

報処理) のうち、重複を除いた名詞約 12 万語を用いた。SS-KWEIC 法により抽出を行なった結果を表 3 に示す。

以上の結果より、最上位語が下位語へのリンクを持つ単語の 1 割を占め、最下位語が上位語へのリンクを持つ単語数の 8 割を占めることから、最上位と最下位の間には存在する語が少ないことがわかる。このため、今回抽出された階層関係の構造の最上位語から最下位語までの距離は浅く、1 段および 2 段の距離が多い。段の階層が少ないのは、使用しているデータが専門用語を対象とし、一般用語を含んでいないため、本来あるはずの階層関係が抽出できないためである。SS-KWEIC 法は単語のみの情報で階層関係を抽出するため、本来あるべきはずの単語が使用するデータに含まれていないと大きな枠組みで階層関係を

(検索コマンド)

1. ID

2. word

enter number >> 1

No. >> 9501

(検索結果)

::: 9501 遠隔端末

<synonym>remote terminal\*\*{32970:リモート端末 33623:遠隔端末装置

<parent> 0

<children> 0}

<parent> 2042:端末

<children> 57805:遠隔端末コマンド 68335:遠隔端末アクセス法

77678:遠隔端末アクセス方式 77679:遠隔端末アクセス法

図 18: 統合結果

表 3: SS-KWEIC の処理結果

総単語数	118,748
総名詞単語数	117,747
名詞 (多義性による重複文字列を除く)	117,206
抽出された階層関係	414,840
1 単語あたりの階層関係 (最大)	3,135
1 単語あたりの階層関係 (最小)	0
1 単語あたりの階層関係 (平均)	3.5
下位語へのリンクを持つ単語数	26,475
上位語へのリンクを持つ単語数	109,731
下位語を持ち上位語を持たない単語数 (最上位語)	3,134
上位語を持ち下位語がない単語数 (最下位語)	86,390
上位語・下位語ともない単語数	4,341

抽出することになる。このことは構造化してはじめてわかることである。なお、最も多くリンクを持つ単語は最上位レベルに位置する「システム」で、3,135の直接の下位語を持つ。さらに下位まで合わせると6,801の下位語をもつ。

次に、抽出された階層関係の妥当性を調べるために、抽出された結果からランダムに1000語を選び出した。この1000語に対し上位関係、下位関係として抽出された単語が概念的に階層関係として正しいと人が判断できた場合、正しい結果であるとした。その結果、96%の関係が階層関係として妥当であると判断できた。確認の結果、「同期」の下位語として「非同期」が抽出されるなど単語間に何

らかの意味関係は存在するが、階層関係ではないと判断される結果や、「ロック」の下位語として「クロック」、「ブロック」が抽出されるなど全く関係のない単語を階層関係として抽出している結果は誤りとした。全く関係のない単語が階層関係として抽出される場合は、パターンおよびパターンにマッチングした単語がひらがな、カタカナのみの単語の時に多い。これは文節として正しく区切れない文字列とパターンがマッチしているためと思われる。これは精度を下げる原因でもあるため、改善をする必要がある。また、今回の抽出条件では接辞を無視しているため、「系」、「式」などの漢字一文字の単語の時に多くの間違いが発見された。接辞の処理は今後の課

題となる。

次に、C-TRAN法により抽出を行なった結果を表4に示す。

表4: C-TRANの処理結果

総名詞単語数	117,747
名詞 (多義性を考慮)	117,583
同義関係の数	15,726
1単語あたりの同義関係 (最大)	256
1単語あたりの同義関係 (最小)	0
1単語あたりの同義関係 (平均)	7.5
同義関係を持つ単語	57,838
同義関係を持たない単語	59,745
多義性あり	375
多義性なし	116,831

C-TRAN法では、同じ見出し語でも対訳が異なる場合には別の用語とした。そのため、C-TRAN法で用いた名詞はSS-KWEIC法で用いた名詞よりも数が多くなっている。EDRは、日本語に対応する英単語がない場合は、対訳を与えていない。そのため、上記の表で同義語なしのうち、18,899語は元データに対訳が含まれていない。C-TRAN法では使用するデータ作成機関が対訳を付与する基準、データに付与されている対訳数により、抽出できる同義関係数、結果の精度が変わってくるといえる。また、C-TRAN法は英単語を利用し同義関係を抽出しているため、英語の多義性によるノイズが生じる。だが今回は、専門用語辞書をデータとして使用しているため、英語の多義性によるノイズは少なかった。英語の多義性によるノイズは、SS-KWEIC法で抽出された階層関係などと統合した時に他の関係を参照することにより、消去できると考えられる。C-TRAN法は、同じ単語集合に属する単語を同義関係とするため、表記ゆれにある単語も同義関係として抽出する。しかし、同義関係抽出後に処理を行なうことにより、表記ゆれにある単語はまとめられて1つの単語として扱われる。

## 5.2 関連研究

人が利用することを前提としたシソーラスには分類語彙表 [10] や角川類義語新辞典 [11] があり、計算機を用いて意味処理を行なうことを目的としたシソーラス、辞書には、上位・下位関係を明確に記述した ISAMAP [12]、概念を体系化した EDR 概念辞書 [13]、IPAL 名詞辞書 [14]、日本語語彙大系 [15]、WordNet [16] などがある。これら既存のシソーラスや辞書を計算機によって処理を行なう場合の問題点は既に指摘されており [17] [18]、分類語彙表や角川類義語辞典に対しては記述が不明確で多義性を考慮していない、ISAMAP に対しては収録語彙数が少ない、EDR 概念辞書に対しては分類項目の大半が終端ノードに偏っている、IPAL 名詞辞書に対しては意味記述が不十分であるといった指摘がなされている。また、現在のシソーラスの共通の問題点として、上位・下位といった単一の観点から意味関係を記述しているため関係が不明確であり、観点によって変化する関係を表現できないということも指摘されている [17]。一方、意味関係の自動抽出を目的とした研究には、辞書の語義記述を利用した階層関係の抽出 [19] [20] [21]、コーパスの共起データを基に確率的手法を用いた単語間の類似度計算 [22]、コーパスを利用したシソーラスへの未知語登録 [23] [24] などが挙げられる。その他、複数の観点による意味関係の記述の提案もある [25]。

これら先行研究の傾向をみると、概念体系全体を記述したシソーラスや辞書は人手によって構築されており、逆に計算機によって構築されたシソーラスは部分的な概念体系の構築であることがわかる。また、人手によるシソーラスや辞書も完全ではなく、記述の詳細さや整合性に問題がある。人手により作成された辞書は概念の説明は詳細に記述されているが、関係の記述が不十分である。関係が記述されている場合でも記述の詳細さに偏りがある。シソーラスは、同義関係、階層関係を中心に語を意味関係を記述している。正確な意味処理を行なうためには、内部属性間の関係

や原因・結果などの論理関係など他の意味関係も記述する必要がある。計算機により構築されたシソーラスは原情報と処理を一つに限定し、処理結果を最終結果とみなしている。しかし、原情報には人手によるシソーラスや辞書、新聞や論文などのフルテキストなど様々な種類があるので、それぞれの質に応じて網羅的に意味関係を抽出する必要がある。また、一つの手法により全ての意味関係を抽出することはほとんど不可能であるので、抽出する意味関係に応じて抽出手法を開発する必要がある。このように、情報源は多種存在し、抽出する意味関係に応じ抽出手法も存在するのならば、一つの情報源、一つの手法による結果を最終結果とするのではなく、複数の意味関係の抽出結果をまとめることにより一つの手法単独の結果よりも、よい構造が構築できると考える。

本研究は処理ではなく構造を中心に概念体系を構築する。以下に構造化することの利点を述べる。複数の手法で抽出された意味関係は抽出された時点では断片的で単なる集合である。各集合は独立しており他の意味関係との関連を見出せない。そのため、複数の意味関係を組み合わせ関連を持たせる必要がある。しかし、組み合わせただけでは重複、矛盾する関係が存在するので、関係を整理統合し構造化する。構造を記述する方法として木構造とネットワーク構造が挙げられる。木構造は階層の記述には適するが、一つのノードから多次元の階層にまたがる関係の記述、同義、反義など他の意味関係を記述することができない。ネットワーク構造は階層を越えた記述、リンクに値を持たせた意味関係を記述することが可能である。また、リンクをたどることにより構造内の関係を全て参照できる。以上の点から本研究ではネットワーク構造による構造化を行なった。しかし、ネットワーク構造でも概念体系を構築するための構造として不十分であり、ノードの内部に意味関係がある場合は構造として記述できない。そのため、概念を体系化するための新たな構造を考える必要がある。

本論文ではSS-KWEIC法とC-TRAN法を用いて階層関係、同義関係を抽出し、統合することにより意味構造を作成した。2つの意味関係抽出結果を統合した意味構造でも、構造化を行なう利点を証明することができた。今後は他の原情報も用いて様々な意味関係を統合し、構造化する予定である。

## 6 おわりに

本論文は階層関係と同義関係をそれぞれ独立に抽出し、1つの構造として自己組織的に統合するINTEGRAL法について述べた。INTEGRAL法によってSS-KWEIC法とC-TRAN法の結果を構造として統合し、新しい意味関係を抽出できた。さらに、概念の多義性も除去できた。本論文によって構造中心の手法が概念の体系化に有効であることを示した。

- [1] Yuzuru Fujiwara, Ye Liu. The Homogenized Bipartite Model for Self Organization of Knowledge and Information. The Proceedings of the 48th FID Conference and Congress. Graz, Oct.1996-10, IFID 2(1), p.13-17(1998)
- [2] Jinjuan Lai, Hanxiong Chen, Yuzuru Fujiwara. An Information-Base System Based on the Self-Organization of Concepts Represented by Terms. Terminology. Vol.3, No.2, p.313-334(1996)
- [3] 藤原譲. 情報学基礎論の現状と展望 - 学習・思考機構と超脳計算機への応用 -. 情報知識学会誌. Vol.9, No.1, p.13-29(1999)
- [4] 頼静娟, 北川博之, 藤原譲. 概念構造表現のためのシソーラス自動構築と精度向上の研究. 情報処理学会第45回(平成

- 4年後期) 全国大会講演論文集 3. p.73-74(1992)
- [5] 頼静娟, 王晓晶, 藤原譲. 専門用語における階層関係及び関連関係抽出法. 情報知識学会 第2回研究報告会講演論文集. p.9-16(1994)
- [6] 情報処理用語大事典編集委員会編. 情報処理用語大事典. 東京, オーム社, 1992, 1305p.
- [7] 長尾真ほか編. 岩波情報科学辞典. 東京, 岩波書店, 1990, 953p.
- [8] Hikomaro Sano, Yuzuru Fujiwara. Automatic Assignment of Word Categories for Improved Facet Analysis of Titles and Indexes. J.Inf.Sci. Principles and Practice20, p.23-31(1994)
- [9] 日本電子化辞書研究所. "第11章 専門用語辞書(情報処理)". EDR 電子化辞書 2.0 版使用説明書. 東京, 日本電子化辞書研究所, 1999.(TR-006) < URL:http://www.ijnet.or.jp/edr >
- [10] 国立国語研究所. 分類語彙表. 東京, 秀英出版, 1982, 362p.
- [11] 大野晋, 浜西正人. 角川類語新辞典. 東京, 角川書店, 1981,932p.
- [12] 田中穂積, 仁科喜久子. 上位/下位関係シソーラス ISAMAP1 の作成 [I],[II]. 情報処理学会研究報告 自然言語処理研究会報告 No.64(87-NL-64). p.25-34, p.35-44(1987)
- [13] 日本電子化辞書研究所. EDR 電子化辞書 2.0 版使用説明書. 東京, 日本電子化辞書研究所, 1999.(TR-006) < URL:http://www.ijnet.or.jp/edr >
- [14] 情報処理振興事業協会技術センター. 計算機用日本語基本名詞辞書 IPAL( Basic Nouns ).
- < URL:http://www.ipa.go.jp/STC/NIHONGO/IPAL/>
- [15] 池原悟他編, 日本語語彙大系. 東京, 岩波書店, 1997. 5冊.
- [16] Fellbaum,C.,ed. WordNet: an electronic lexical database. Massachusetts. MIT Press, 1998, 423p.
- [17] 宮崎正弘, 池原悟, 横尾昭男, 白井諭. 日英機械翻訳のための意味属性体系. 情報処理学会研究報告 自然言語処理研究会報告 No.120(97-NL-120), p.29-36(1997)
- [18] 柏野和佳子, 本多啓. IPAL 名詞辞書における多義構造の記述. 情報処理学会論文誌. Vol.39, No.9, p.2603-2612(1998)
- [19] 荻野孝野, 横山晶一, 荻野綱男. シソーラス作成のための辞書関係語の抽出. 情報処理学会第39回(平成元年後期)全国大会 講演論文集. p.670-671(1989)
- [20] 横山晶一, 荻野孝野, 荻野綱男. 国語辞典に基づくシソーラスの計算機処理. 情報処理学会第39回(平成元年後期)全国大会 講演論文集. p.672-673(1989)
- [21] 鶴丸弘昭, 竹下克典, 伊丹克企, 柳川俊英, 吉田将. 国語辞典情報を用いたシソーラス作成について. 情報処理学会 自然言語処理研究会報告 No.83(91-NL-83). p.121-128(1991)
- [22] 本田岳夫, 奥村学. コーパスから二重確率的シソーラスを自動的に獲得する手法の提案. 情報処理学会研究報告 自然言語処理研究会報告 No.109(95-NL-109). p.33-38(1995)
- [23] 中山拓也, 松本祐治. シソーラスへの未知登録語の自動登録. 情報処理学会研究報告 自然言語処理研究会報告 No.120(97-NL-120). p.103-138(1997)



- [24] 浦本直彦. コーパスに基づくシソーラス  
-統計情報を用いた既存のシソーラスへ  
の未知語の配置-. 情報処理学会論文誌.  
Vol.37, No.12, p.2182-2189(1996)
- [25] 川村和美, 片桐康裕, 宮崎正弘. 語を種々  
の観点から分析した多次元シソーラス.  
電子情報通信学会技術研究報告. NLC-  
94-48, p.33-40(1995)

(1999年10月12日受付)

(2000年1月14日採録)

## 著者紹介

### 伊東 千夏 (学生会員)

現在、図書館情報大学大学院 修士課程  
在学中. デジタル図書館、情報の組織化と  
その応用に興味を持つ.  
情報科学技術協会 会員.  
Email: chinatsu@ulis.ac.jp

### 宇陀 則彦 (正会員)

筑波大学大学院 工学研究科 博士課程修了.  
工学博士.  
現在、図書館情報大学総合情報処理センター  
講師. 主にデジタル図書館の研究に従事.  
ACM, IEEE, 情報処理学会等 会員.  
Email: uda@ulis.ac.jp

### 石塚 英弘 (正会員)

東京大学大学院 理学系研究科 博士課程  
修了. 理学博士.  
現在、図書館情報大学 図書館情報学部 教授  
情報検索システム、電子図書システムなどに  
興味を持つ.  
情報処理学会, ACM 等 会員.  
Email: ishizuka@ulis.ac.jp

### 藤原 譲 (正会員)

東京大学大学 工学部 応用物理学科卒業  
理学博士.  
現在、神奈川大学 理学部情報科学科 教授  
情報学基礎理論、超脳計算機などに興味をもつ  
ACM, IEEE, ASIS, ACS、情報処理学会  
電子情報通信学会等、会員  
Email: yfuji@info.kanagawa-u.ac.jp

## Abstracts

### Thesaurus - Past and Present

Yoshio Terasawa

*J.J.S.I.K., Vol.9, No.4, pp.3-11(1999)*

First, the development of 'thesaurus' from medieval to present is briefly surveyed with special reference to Roget's thesaurus. The intention of Roget's Thesaurus is to categorize the ideas which are expressible by language and then to formulate English lexical system. It is also asserted that Roget owed his idea of thesaurus to F. Bacon, J.A. Comenius and J. Wilkins. Finally, we discuss what kind of modification Roget's Thesaurus has been receiving to meet the practical demands of the times, and thus examine several types of thesaurus in order to investigate the possibilities of further developments of thesaurus.

### "Bunrui Goiho" as a resource of information processing -Compilation strategy and treatment of technical terms-

Wakako Kashiwano, Takashi Nakano, Masahiko Ishii

*J.J.S.I.K., Vol.9, No.4, pp.11-28(1999)*

"Bunrui Goiho" (Word List By Semantic Principles) is one of the most widely used Japanese thesauruses compiled by The National Language Research Institute. This paper discusses its design policies as well as its real lexical descriptions both from an information processing viewpoint and from a viewpoint of terminology. Firstly, its word-classification strategy is discussed. The discussion clarifies the merits and limitations when the thesaurus is applied to information processing tasks such as information retrieval. This paper then focuses on treatment of technical terms. The extensive analysis of its lexical descriptions provides potential justification for its use as a technical-term thesaurus, although it is originally designed as a thesaurus of non-technical terms.

### Thesauri and Bibliographic databases

Kenkichi Nagai, Ikuko Harada

*J.J.S.I.K., Vol.9, No.4, pp.29-37(1999)*

On the JICST thesaurus, the way of vocabulary control such as representation of terms, the synonym control and setting of the hierarchical relation was described. There is the distinct feature in the semantic relations among thesauri. It is discussed that this feature is the outcome of the distinction in bibliographic databases to which thesauri are applied.

And we mentioned the large-scale electronic dictionary for searching databases which is developing in JST. The purpose of this dictionary is integrated retrieval of variously digital contents that distributed on network.

### Structurelization of Concepts by Integration of Semantic Relationships

Chinatsu Ito, Norihiko Uda, Hidehiro Ishizuka, Yuzuru Fujiwara

*J.J.S.I.K., Vol.9, No.4, pp.38-48(1999)*

This paper describes structuralization of concepts by integration of semantic relationships. Hierarchical relationships are extracted by pattern matching according to coinage rules. Synonymous relationships are extracted by a transitive rule between English words and Japanese ones. Extracted hierarchical relationships and synonymous ones are integrated as a structure, and new connections are discovered in the conceptual structure. The method is effective in structuring concepts in the real world.

## 書評

## 日本語形成の謎に迫る

澤田洋太郎 著. 新泉社. 1999年. 329p. 2,800円(本体)

IT 経営研究所 平田 周

世界の主要言語で、形成された道筋が明らかにされていない言語は日本語だけだとされる。しかも、日本語は、ある一つの言語から枝分かれして生まれたヨーロッパ系の諸語とは異なり、多くの系統の言語が流入して出来た混成語である。縄文人は、日本全国に広がって平和に豊かな生活をしてきたことが最近の遺跡発掘からますます確からしくなっている。一体、彼らはどんな言葉話を話していたのだろうか。それが日本語の元になったと考えてもおかしくない。著者は、アイヌの祖先こそ縄文時代の列島先住民だったと考え、アイヌ語は原日本語の面影を残しているという説をとる。

アイヌ語と日本語とは語順が同じであるから、単語さえ入れ替えれば簡単に翻訳される。その点は日本語と朝鮮語についても言える。文法的には同系の言語であるといえよう。

そのアイヌ語に由来すると思われる地名は、全国に広がっている。アイヌ研究の草分けである金田一京助博士は、本州各地の地名の意味が日本語としては理解できなくても、アイヌ語で考えるとその土地の地形などを意味する言葉だとわかるなどを豊富な事例で示した。

たとえば、アイヌ語で「シヨウ」とは「滝」を意味し、「ナイ」は「川」や「沢」を意味する。したがって、日本各地にある「庄内」という名の土地は、「シヨウナイ」すなわち「滝川」のことになる。

もちろん知名がアイヌ語由来のものだけではなく、中国や東南アジアはもとより、ペルシャやユダヤに起因すると思われるものまでさまざまである。そのなかで東海大学名誉教授の茂在寅男氏は、伊豆に多い「狩野」という名前は、古代ポリネシア語のカヌーに由来するもので、日本書記にも「伊豆で船を造らした」という記事のあることを指摘する。「安

房」「阿波」はポリネシア語で「港」という意味であり、そのほかにもこれまで難解とされていた「天磐船」「鳥之石楠船」などの名もポリネシア語で説明がつくという。

土地の名前の多くがアイヌ語に由来しているならば、縄文時代にアイヌ系の人々が日本各地に住んでいた。というよりは、縄文人はアイヌ語を話していたのかもしれないと想像してもおかしくない。近年、青森県の三内丸山遺跡や鹿児島県の上野山遺跡などの発掘が進められ、縄文時代のイメージは根本から覆された。狩猟採集の原始的な生活をしてきたと考えられてきたこれまでの縄文人の生活は、整然とした住居配置、大型の建物、そして一部には栽培すら行われていたという、これまでは想像もされなかった新しい縄文時代のイメージがつくられている。縄文人が採集生活をしてきたため、縄張り意識が強く、閉鎖的だったとされたこれまでの通説は、地域間で広く物資の交易を行い、きわめて豊かで平和な暮らしをしていたと思われるようになったのである。しかし、彼らがどんな言葉話を話していたのか。文字がなかったため(アイヌ語はいまでも文字がない)、遺跡は証明してくれない。

日本語では、「頭」のことを「アタマあるいはツブリ」「カシラ」「コウベ」という3とおりの言い方がある。「顔」の場合も「カオ」「ツラ」「オモテ」と3とおりある。「猿」も「サル」「エテ」「マシラ」、「蛇」にも「ヘビ」「オロチ」「クチナワ」という呼び方がある。山を意味する日本語には、「ヤマ」以外に「タケ(岳)」「ミネ(峰)」「モリ(杜、森)」と「サン(山)」がある。このように、日本語には3つ以上の異なる言語体系があったと思われる。日本民族はけっして単一な民族ではなく、3つ以上の人種と複数の文化から構成さ

れていたということであろう。

安本美典氏は、最初に流入した言語を「古極東アジア語系の言語」と名づけ、日本語祖語、朝鮮語祖語、アイヌ語祖語の共通語があったとし、それが日本語の文法的、音韻的特徴をつくり、若干の基礎語彙を残したと考える。そのあとに「インドネシア系、クメール系言語」が入り、基礎語彙となり、さらに下って「中国江南地方からのビルマ系言語」が流入してきて弥生時代の中心を形成し、主に身体語、数詞、代名詞、植物関係の語が広がったとする。西暦紀元以後、中国語(北京方言の祖語)が入り多くの文化語を生み出したというのである。(「日本語の成立」講談社現代新書 1978)

弥生時代が始まる BC 4 世紀以前の日本列島には原日本語というべきものがあり、そこに外部から新たな言語が流入してきて、7 世紀ごろほぼ上代日本語が成立したとみることは、大筋として誰も異存はないと著者は言う。

しかし、数では少ない流入者が文化水準や軍事、政治、経済などを力として、在来居住者を圧倒し、自分たちの言語を強制的に使用させたり、ある言語分野については自発的に新しい言葉や文法的な言い回しさをも迎合的に受け入れるということがありうると指摘している。すなわち、記録として残されているものは、支配層の言葉で綴られており、原住民の生活用語はそのままだったのではないかという疑問も提示する。

そのような例として、ヨーロッパで、ラテン系がゲルマン系を圧倒していったことをあげ、動物そのものの名前については飼育者である農民の使うゲルマン系の言語 cow, ox, sheep, pig などが使われるのに対して、その肉については beef, mutton, pork というようにそれを食べる支配者が使っていたフランス語系になっているという例を上げる。牛、羊、豚は動物と肉とで名称が違うが、鶏肉については動物も肉も chicken というゲルマン系のままであるのは、侵入者は鶏など見向きもしなかったのだろうかと言っている。これと似たようなことは、上代日本語の形成でも起き

ただろうという。

日本語の由来や歴史、その本質を解明しようとする努力のほとんどは、専門の言語学者ではない。医師、心理学者、文学者、海洋学者、大使館員など実にさまざまである(ちなみに著者は東京大学法学部政治学科を卒業後、高校社会科教師を勤め、定年退職後文筆活動をしている)。そして、非言語学者が打ち上げる「日本語論」の花火に対して、専門学者たちは「恣意的で方法論としてなっていない」と酷評はするものの、以後は完全無視の態度を貫き、それは非礼という以前の冷酷きわまる姿勢だと批判する。

その一因は、1865 年のパリの言語学会で「今後は言語の起源に関する論文はいっさい受け付けない」という宣言がなされ、ロンドンの言語学会でも会長が似たような発言をして以来の国際的な了解が伝統になっているためだとしている。現代の言語学者たちの主要な関心は、発生や変化、発展といった歴史的な問題から離れ、言語の構造や音声学、意味論、言語類型といったことに限られるようになった。

確かに、言語が成立した事情がわからないのは日本語くらいで、世界の主要な言語の成立過程はほとんど解明し尽くされている(アメリカ・インディアンやニューギニアの諸部族の言語系統は明らかになっていない)。わが国の言語学者が、そのような世界的風潮にしたがい母国の言語の系統の探究意欲を失ったかにみえるが、むしろ門外漢である意欲的な研究者が自由な発想で、新しい発見ができる楽しみがすべてに開かれていることを幸いとすべきであろう。

## 用語解説

## Z39.50

図書館情報大学 総合情報処理センター 宇陀 則彦

Z39.50 は情報検索用の標準プロトコルの規格名で、「Z39」とは ANSI(American National Standard Institute) の構成団体である NISO(National Information Standard Organization) が定めた規格であることを示す ANSI 内の分類番号であり、「50」とは 50 番目という意味である。Z39.50 は正式には「ANSI/NISO Z39.50-1995 Information Retrieval(Z39.50): Application Service Definition and Protocol Specification (情報検索 (Z39.50): アプリケーションサービスの定義及びプロトコル仕様)」といい、1998 年には Z39.50 は ISO(International Organization for Standardization) の規格 ISO 23950-1998 となり、続いて 1999 年には JIS の規格 JIS X0806-1999 となった。

Z39.50 はクライアントとサーバの間で検索および返戻を行う際の約束事を定めた規格である。Z39.50 で決めたのは通信の際の手順とデータ構造だけであり、クライアントとサーバ自身については決めていない。したがって、クライアントやサーバは自由に設計でき、それでいて共通のプロトコルに従っているので相互に接続できる。

Z39.50 の基本機能は初期化、検索、返戻の 3 つであり、それぞれクライアントから要求を出し、サーバがそれに応えるという形で行われる。初期化ではバージョン、ユーザ認証、メッセージサイズなどクライアントとサーバの基本情報について交渉し、検索セッションをはる。検索ではクライアントからデータベースやアクセスポイントを指定した検索要求が出され、サーバは要求に従ってデータベースを検索し、検索結果集合を作成し、次の返戻要求に備えて結果集合を維持する。返戻では

クライアントから検索結果集合や返戻形式を指定した返戻要求が出され、サーバは指定されたレコードを指定された形式で返戻する。Z39.50 に従って検索システムを構築することにより、ユーザは使い慣れた一つのクライアントから複数のサーバに接続できる。同時に複数のサーバに接続すれば横断検索も可能である。また、セッションを維持するので、履歴検索も可能である。Z39.50 はあくまで論理的なデータベース構造や検索構造に基づく検索に特化したプロトコルであり、WWW で使われている HTTP(Hyper Text Transfer Protocol) を通した検索エンジンとは性格が異なる。

現在、アメリカやヨーロッパなどの図書館システムを中心に実装が進んでおり、200 以上の Z39.50 サーバが構築されている。日本でも最近、急速に実装が進んでおり、学術情報センター、東京工業大学、神戸大学、早稲田大学、図書館情報大学で電子図書館の機能として実装された。Z39.50 は現在アメリカ議会図書館が規格の管理維持機関として指定され、ZIG(Z39.50 Implementors Group) が実装に関わる作業を行っている。Z39.50 の仕様書をはじめとして、Z39.50 に関する情報は <http://lcweb.loc.gov/z3950/agency/> から得られる。日本では <http://z3950.isl.intec.co.jp/> から情報が得られる。

お知らせ

## 「情報知識学会第4回月例研究会」のご案内

平田周 IT経営研究所(情報知識学会理事)

2月22日、第4回月例研究会を開催します。今回は、東京商船大学名誉教授で航海学をご専門とされ、英王立航海研究所名誉会員でもある茂在寅男先生をお招きしてお話をうかがうことにしました。航海をめぐり古代の歴史に詳しく、とりわけ日本語の成立に関連して古ポリネシアの言語が重要な影響を与えたことを実証するなど、きわめて博識、独創的な見解を持っておられます。現在86歳の高齢ながらもことにご壮健で、研究と講演などでご活躍中です。めったに聞けない話だと思いますので、ぜひふるってご参加下さい。今回お話しいただくテーマは日本語の成立にまつわるさまざまなエピソードで、情報知識学会のメインテーマから少し離れているかもしれませんが、日本語がどのようにして出来たのかということは、言語が情報の根底をなすものだと考えれば関心を寄せてもよい問題かと考えます。そうではなくても遠く縄文の時代の昔に、わが国でどのような言葉が話されていたのか、そのルーツはどこかなどの謎はロマンをかき立てるのではないのでしょうか。昨年12月に上掲された澤田洋太郎著「日本語形成の謎に迫る」(新泉社)にも、先生のご研究の成果が紹介されており、この本の内容紹介を添付しておきますので、ご一読下されば幸いです。

いつもどおり出席の連絡は不要ですが、もし事前にメールかファクスでご連絡いただければ助かります。いつも出席者が少なく、講師の先生にも気を使いますので、ぜひご協力下さい。

テーマ： 日本語成立をめぐるとのエピソード

講師： 茂在寅男先生

1914年生まれ。1938年東京高等商船学校航海科卒業。

東京商船大学教授、東京大学水産学科講師を経て、1977年

より東海大学海洋学部教授、現在東京商船大学名誉教授。

工学博士(1961年東大電気)、日本水中考古学会副会長、

英王立航海研究所名誉会員

著書に「日本語大漂流」(光文社)のほか、古代日本の航海術(小学館)など多数。NHKドラマの考証などにも参画。

開催日時： 2000年2月22日(火曜日)

午後6時30分から午後9時頃まで

場所： 世界貿易センタークラブルーム

JR浜松町駅 世界貿易センタービル 38階

TEL 03-3435-5680

会費： 2,000円(ソフトドリンク代金を含む)

【お問い合わせ】情報知識学会事務局 五所吉哉

TEL 03-3835-5692 FAX 03-3837-0368

E-MAIL: LDE01013@nifty.ne.jp

平田周 TEL 03-3456-2665 FAX 03-5444-3302

E-mail: LDG01371@nifty.ne.jp

お知らせ [論文募集]

## 情報知識学会第8回(2000年度)研究報告会 発表論文募集について

実行委員長 田畑孝一 (図書館情報大学)

情報知識学会では平成12年5月27日(土)に、総会とともに研究報告会を開催する予定です。この研究報告会の発表論文を下記要領で募集いたしますので、どうぞ奮ってご応募ください。

### 1. 募集分野

- (1) 情報知識の構造解析、モデル化、意味理解、自己組織化、可視化
- (2) 情報・知識の表現、生産、組織化、検索、提供
- (3) 電子出版、電子図書館、フルテキストデータベース
- (4) マルチメディア、電子ミュージアム
- (5) 用語、シソーラス、電子化辞書、機械翻訳
- (6) デジタル・コンテンツの流通技術・体制と著作権処理
- (7) 専門分野における情報管理、レコードマネージメント
- (8) その他情報知識学、図書館・情報学に関連する諸研究・開発

### 2. 発表申し込み方法

発表論文題目、著者名(連名の場合、登壇発表者名に○印)、所属、該当する募集分野、連絡代表者の氏名、連絡先の住所、電話/FAX番号、電子メールアドレス、予想される論文ページ数を明記の上、下記宛て、電子メールまたは郵便、FAXにてお申し込みください。

〒305-8550 つくば市春日1-2 図書館情報大学 田畑孝一

E-mail: tabata@ulis.ac.jp、tel: 0298-59-1010、fax: 0298-59-1062

申し込み期限: 平成12年3月31日、採択可否通知: 平成12年4月7日

### 2. 3. 論文執筆・発表等について

(1) 採択されたものについては論文執筆要領の詳細を別途お知らせします。発表論文は4ページ(A-4判)にカメラレディの形で仕上げ、平成12年4月28日までに、上記2.の郵便アドレスまでお送り頂くことになります。なお、カメラレディの形とは、そのまま複製して印刷物にしてもよいものを指します。4ページまでは無料ですが、それを越えると有料(1頁1,000円)になります。

(2) 発表論文の提出がないと研究報告会での発表ができません。発表時間は、質疑応答を含めて30分程度を予定しています。

(3) 研究報告会は平成12年5月27日(土)、東京都内の会場にて開催予定。

(4) 登壇発表者は情報知識学会の会員に限ります。(当日入会も可)

(5) 発表の際、OHPまたは液晶プロジェクターを使用希望の方は、その旨、事前にお知らせ下さい。

## 投稿規程

第1版(暫定板)1990年3月

第2版(暫定板)1995年7月

第3版(暫定板)1998年3月

情報知識学会では、SGML(Standard Generalized Markup Language)に基づく学術情報の編集・蓄積の実現を意図しています。SGMLはISO規格8879またJIS規格X4151で、文章データベース作成・保守・交換の技術として、電子出版だけでなく、マニュアルなど技術文書管理にも使われてきました。最近では、さらにネットワーク上でのコンカレント・エンジニアリングから電子商取引まで、文書情報の編集・蓄積の基盤技術として国際的に普及しつつあります。

### 1. 原稿の投稿資格

本学会誌に投稿原稿を投稿するには、原則として本学会会員(個人会員)に限る。ただし、投稿者が2名以上の連記の場合には、少なくともそのうち1名は本学会会員(個人会員)であることが必要です。

### 2. 原稿の種類

論文(原著論文)、総説、解説などがあり、原稿には投稿原稿と依頼原稿とがあります。原稿は日本語または英語で書いてください。主題は本学会の対象とすることであれば特に限定しません。

### 3. 査読と採否

論文の採否は編集委員会で決定し、著者に通知します。また、内容の改善を求めることもあります。論文以外の原稿については査読の結果、表現の分かりやすさなどの点から修正をお願いすることがあります。

### 4. 原稿の構成

表題(和文と英文)、著者名(和文表記と英文表記)、著者所属(和文表記と英文表記)、要約(和文と英文)、本文(和文または英文)、文献リスト、注、表、図など。

### 5. 原稿の書式

原稿のうち、テキストおよび図・表は3.5インチフロッピーディスク(1.44MB 2HDのみ)とその印字出力(2部、うち1部は査読用)で、図や表は紙(2部)でお出し下さい。

テキスト部分の文書形式は、下記の形式に限ります。いずれの形式においても文字コードはSJISに限ります。

- (a) ASCII L<sup>A</sup>T<sub>E</sub>X のコマンドが挿入されている Plain Text
- (b) Rich Text Format(MS Word、WordPerfect で使用されている書式)
- (c) Plain Text format(いわゆる単純なテキストファイル形式)

使用したソフトウェアがフロッピーディスク表面にわかるように記述してください。

Vol.7, No.2 から L<sup>A</sup>T<sub>E</sub>X で編集・印刷しております。これに伴い、本投稿規程の変更を予定しておりますので、最新の投稿規程については下記の当学会のホームページをご覧ください。

情報知識学会 HP <http://angelos.ed6.info.kanagawa-u.ac.jp/jsik/main.html>



## 6. 原稿の送付

### 6.1 送付時に必要な書類の確認

送付時に必要なものを改めて記すので確認してください。

- ・ 標題, 著者名, 所属機関, 概要 (これらはいずれも和文と英文の両者)
- ・ 本文 (和文または英文), 参考文献, 図・表
- ・ 以上の本原稿に加えてそのコピー1部
- ・ 著者1名の連絡先 (郵便宛先, 電話番号, FAX 番号, e-mail アドレス)

### 6.2 原稿の送付

原稿は学会事務局ではなく、下記宛てにお送りください。封筒に「情報知識学会誌 投稿原稿 在中」と明記してください。

〒259-1293 平塚市土屋 2946  
神奈川大学 理学部 情報科学科  
後藤 智範  
TEL: 0463-59-4111, FAX: 0463-58-9684  
E-mail: gotoh@info.kanagawa-u.ac.jp

さらに、下記の内容を含むテキストファイルを e-mail で上記のアドレス (学会誌編集委員長) 宛てに宛てにお送りください。

標題 (和文、英文), 概要 (和文、英文)  
著者 (1名) の氏名, E-mail アドレス  
所属機関 / 部課名, 所属機関の住所 / 電話番号 / FAX 番号

## 7. 著作権

著作権は原則として本学会に帰属するものとします。著作権が他の機関にある場合などで、本学会に帰属することが困難な場合には協議することもあります。本学会誌に掲載された執筆内容が第三者の著作権を侵害するなどの指摘がなされ、損害を与えた場合には、執筆者がその責任を負うこととなります。

情報知識学会役員 (2000年1月現在)

会長 藤原鎮男 東京大学 名誉教授  
副会長 藤原譲 神奈川大学 教授  
細野公男 慶応義塾大学 教授  
根岸正光 学術情報センター 教授

理事 浅田雄司 金属材料技術研究所  
有川節夫 九州大学 教授  
岩淵幸雄 日本創造学会 理事  
国沢隆 東京理科大学 講師  
高橋仁一 大日本印刷(株) 開発室長  
菅原秀明 国立遺伝学研究所 教授  
田畑孝一 図書館情報大学 副学長  
次田皓 つくばプロテオミクス  
研究所 所長

長瀬真理 静岡大学 教授  
名和小太郎 関西大学 教授  
平田周 IT経営研究所 所長  
松田芳郎 東京国際大学 教授  
山本昌弘 法政大学 教授

監事 宍戸駿太郎 環日本海経済研究所 研究顧問  
月見里禮次郎 えんがわ 代表

芦沢実 日立製作所(株) 主任研究員  
石塚英弘 図書館情報大学 教授  
菊田昌弘 シナジー・インキュベート(株) 代表取締役  
後藤智範 神奈川大学 教授  
小山照夫 学術情報センター 教授  
田隅三生 埼玉大学 教授  
田村喜代子 市川房枝記念会  
中川優 和歌山大学 教授

永村真 日本女子大学 教授  
西脇二一 奈良大学 教授  
深見拓史 凸版印刷(株) 技術企画部長  
安永尚志 国文学研究資料館 教授

(五十音順)

情報知識学会 賛助会員 (2000年1月現在)

(株) シナジー・インキュベート  
大日本印刷(株)  
凸版印刷(株)  
中根速記学校  
(社) 日本印刷技術協会  
日本ツーリスト開発(株)  
ブル(株)  
(株) 横河総合研究所

医歯薬出版(株)  
科学技術振興事業団  
駿河台大学図書館  
(株) トータルメディア開発研究所  
(株) トップラン・グループ総研  
日商岩井インフォコム(株)  
(株) PFU ソフトウェアラボラトリ  
プロセス資材(株)

(五十音順)

=====

《 編集後記 》

当号は初めての「特集号」です。当号の刊行は、巻頭言を執筆いただいた石川先生の多大な努力により可能となりました。ここに謝意を表します。

若手の会員で当学会誌の L<sup>A</sup>T<sub>E</sub>X 編集に協力いただける方がおりましたら大歓迎です。下記宛てまでご連絡ください。

gotoh@info.kanagawa-u.ac.jp 編集委員長 後藤 智範

=====

=====

情報知識学会誌 編集委員会

編集委員長	後藤 智範	神奈川県 理学部 情報科学科 (e-mail: gotoh@info.kanagawa-u.ac.jp)
編集委員	藤原 譲	神奈川県 理学部
	石川 徹也	図書館情報大学 図書館情報学部
	石塚 英弘	図書館情報大学 図書館情報学部
	根岸 正光	学術情報センター 研究開発部
	西脇 二一	奈良大学 社会学部
	菅原 英明	国立遺伝学研究所

=====

■複写をされる方に

本誌に掲載された著作物を複写したい方は、(社)日本複写権センターと包括複写許諾契約を締結されている企業の従業員以外は、著作権者から複写権等の行使の委託を受けている次の団体から許諾を受けて下さい。著作物の転載・翻訳のような複写以外の許諾は、直接本会へご連絡ください。

〒107 東京都港区赤坂 9-6-41 乃木坂ビル 学術著作権協会

TEL:03-3474-4621, FAX:03-3403-1738 E-mail: kammori@msh.biglobe.ne.jp

アメリカ合衆国における複写については、次に連絡してください。

Copyright Clearance Center, Inc.

222 Rosewood Drive, Danvers, MA. 01923, USA

TEL: 508-750-8400, FAX: 508-750-4744 URL: <http://www.copyright.com>

情報知識学会誌 Vol.9, No.4 2000年2月19日発行 編集・発行 情報知識学会  
頒布価格 3000円

=====

情報知識学会 (JSIK: Japan Society of Information and Knowledge)

会長 藤原 鎮男

事務局

〒110-8560 東京都台東区台東 1-5-1 凸版印刷(株)内

TEL: 03(3835)5692

FAX: 03(3837)0368

e-mail: LDE01013@nifty.ne.jp

URL: <http://angelos.info.kanagawa-u.ac.jp/jsik/main.html>

=====

# *Journal of Japan Society of Information and Knowledge*

## ~~~~~ Contents ~~~~~

### Special Issues: Thesauri – Systematization of Domain Knowledge

#### Invited Paper

Thesaurus - Past and Present.....Yoshio Terasawa..... 3

#### Invited Paper

"Bunrui Goiho" as a resource of information processing  
Wakako Kashiwano, Takashi Nakano, Masahiko Ishii..... 12

#### Technical Report

Thesauri and Bibliographic databases...Kenkichi Nagai, Ikuko Harada 29

#### Research Paper

Structurelization of Concepts by Integration of Semantic Relationships  
Chinatsu Ito, Norihiko Uda, Hidehiro Ishizuka, Yuzuru Fujiwara... 38

Abstracts ..... 49

Book Reviews ..... 50

Terminology ..... 52

Notices ..... 53

Information for Authors ..... 55

~~~~~

情報知識学会誌 第9巻4号 2000年2月19日発行

編集兼発行人 情報知識学会 〒110-8560 東京都台東区台東1-5-1 凸版印刷(株)内

TEL: 03(3835)5692 FAX: 03(3837)0368 (振替: 00150-8-706543)

学術刊行物 ISSN 0917-1436