

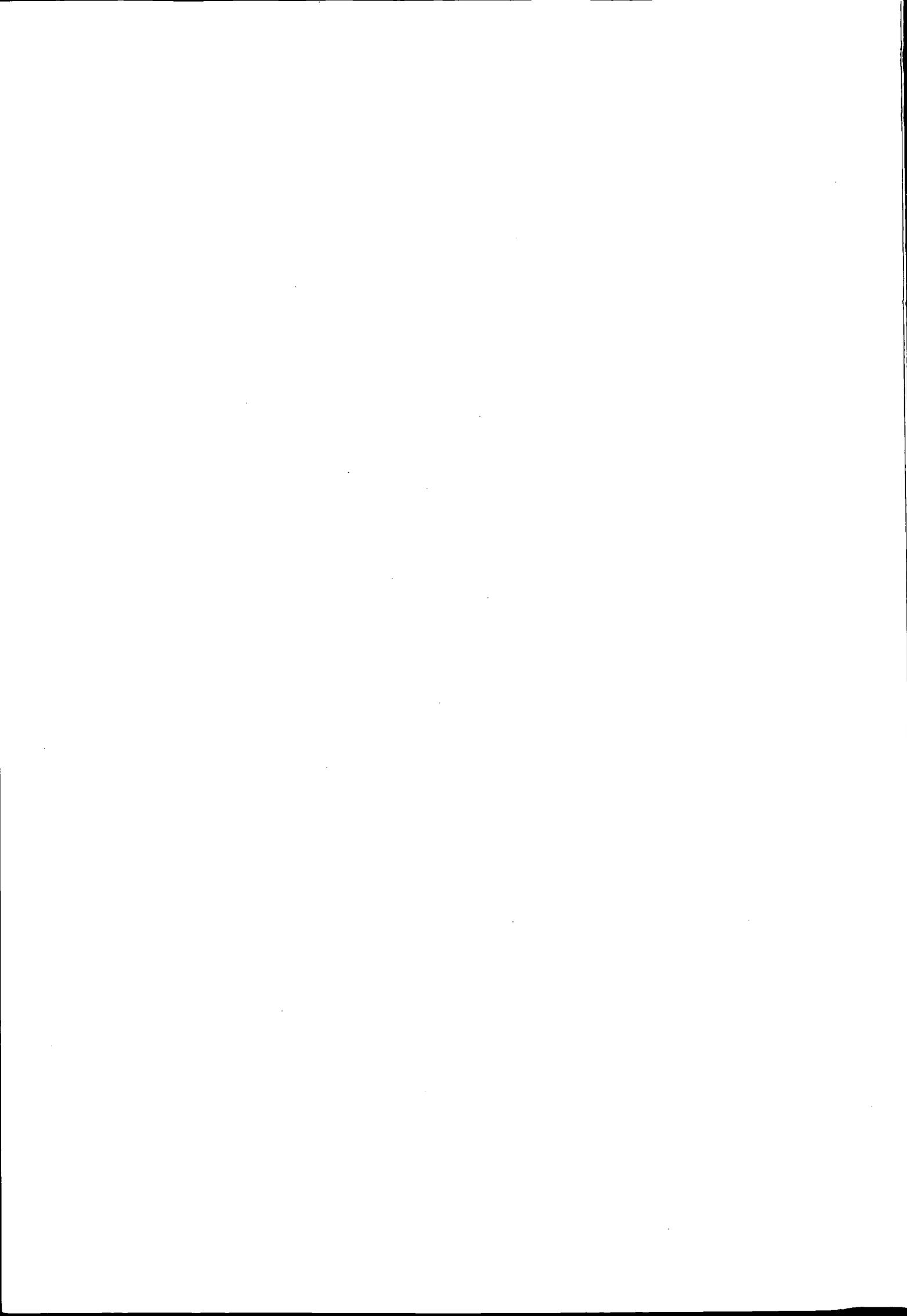
データベース構築促進及び技術開発に関する報告書

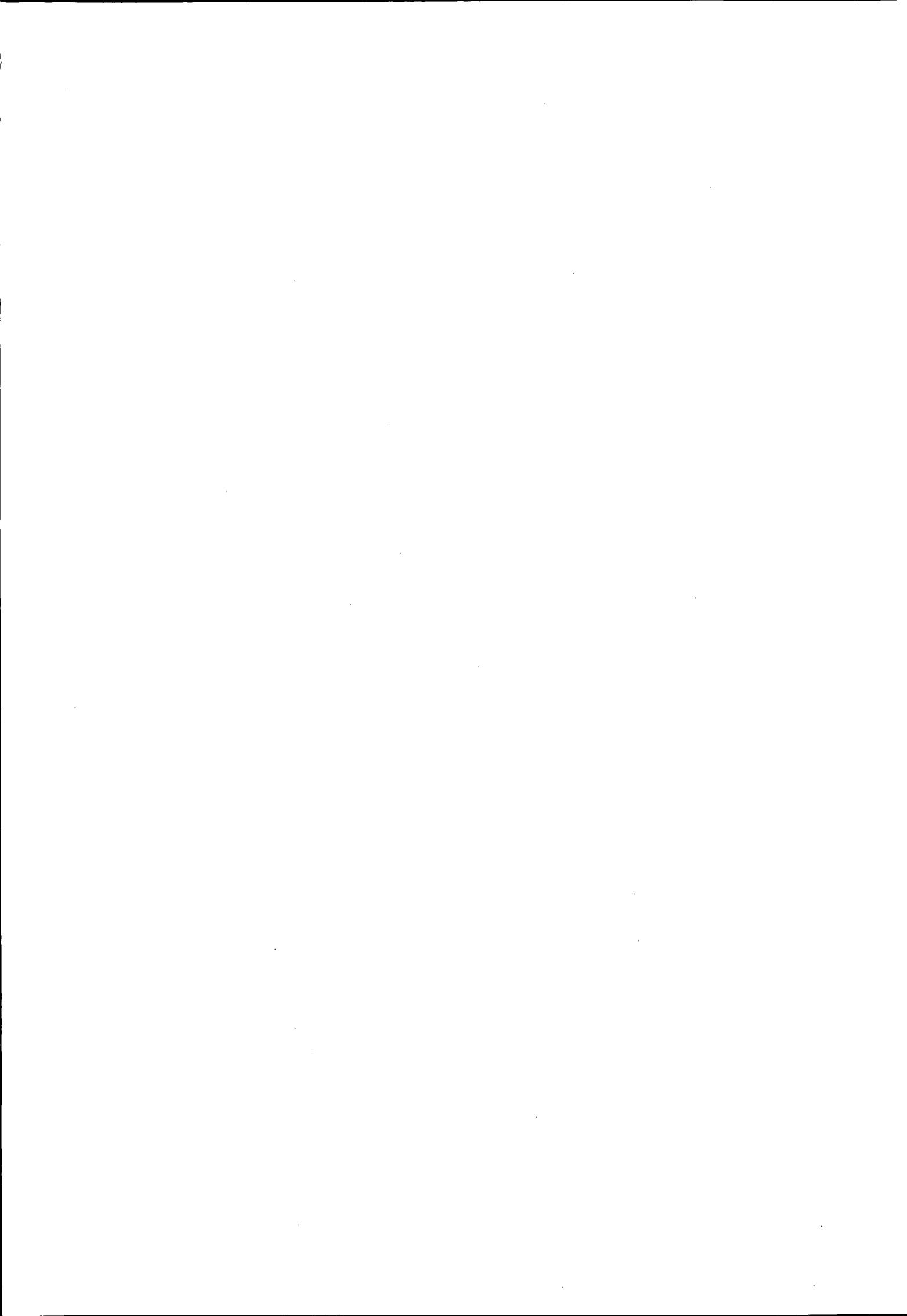
先端産業分野における専門用語の
電子辞書データベース化の調査研究

平成5年3月

財団法人 データベース振興センター

委託先 科学技術情報研究所株式会社





本事業は、日本自転車振興会から競輪収益の一部である機械工業振興資金の補助を受けて作成したものである。

序

データベースは、わが国の情報化の進展上、重要な役割を果たすもの期待されている。今後、データベースの普及により、わが国において健全な高度情報化社会の形成が期待される。さらに海外に対して提供可能なデータベースの整備は、国際的な情報化への貢献および自由な情報流通の確保の観点からも必要である。しかしながら、現在わが国で流通しているデータベースの中でわが国独自のものは1/3にすぎないのが現状であり、わが国データベースサービスひいてはバランスある情報産業の健全な発展を図るためには、わが国独自のデータベースの構築およびデータベース関連技術の研究開発を強力に促進し、データベースの拡充を図る必要がある。

このような要請に応えるため、(財)データベース振興センターでは日本自転車振興会から機械工業振興資金の交付を受けて、データベースの構築および技術開発について民間企業、団体等に対して委託事業を実施している。委託事業の内容は、社会的、経済的、国際的に重要で、また地域および産業の発展の促進に寄与すると考えられているデータベースの構築とデータベース作成の効率化、流通の促進、利用の円滑化・容易化などに関係したソフトウェア技術・ハードウェア技術である。

本事業の推進に当って、当財団に学識経験者の方々に構成されるデータベース構築・技術開発促進委員会(委員長 山梨学院大学教授 蓼沼良一氏)を設置している。

この「先端産業分野における専門用語の電子辞書データベース化の調査研究」は平成4年度のデータベースの構築促進および技術開発促進事業として、当財団が科学技術情報研究所株式会社に対して委託実施した課題の一つである。この成果が、データベースに興味をお持ちの方々や諸分野の皆様方のお役に立てれば幸いである。

なお、平成4年度データベースの構築促進および技術開発促進事業で実施した課題は次表のとおりである。

平成5年3月

財団法人データベース振興センター

平成4年度 データベース構築・技術開発促進委託課題一覧

分野	課題名	委託先
社 会	1 変異タンパク質配列データベースの構築 2 新聞縮刷版見出しデータベースの構築 3 ファジィに関する文献データベースの構築 4 医療用医薬品抗生物質データベースの構築 5 交通事故調査データベースの構築 6 楽器データベースの構築 7 人体計測データベースの構築 8 大学におけるデータベース利用教育システムのプロトタイプ作成 9 先進複合材料データベースの構築 10 博物館所蔵地図資料所在情報データベースの構築調査	日本電子計算(株) (株)朝日新聞社 (財)日本情報処理開発協会 (株)小田島 (財)日本自動車研究所 (株)ダイソメディアサービス (社)人間生活工学研究センター 日外アソシエーツ(株) (財)次世代金属・複合材料研究開発協会 (財)地図情報センター
中小企業振興 地域活性化	11 地域流通最適化データベースのプロトタイプ作成 12 異分野研究のための知的オリエンテーション・データベースシステムのプロトタイプ作成 13 在宅勤務者サポート・データベースの構築調査	(社)日本ボランティア・チェーン協会 (株)けいはんな (株)志木サテライトオフィス・ビジネスセンター
海 外	14 銅基複合材料日本特許英文データベースの構築 15 技術協力供与機材データベースのプロトタイプ作成 16 先端産業分野における専門用語の電子辞書データベース化の調査研究 17 マーケティングコードの英文データベースの構築	神鋼リサーチ(株) (財)日本国際協力システム 科学技術情報研究所(株) (株)帝国データバンク
技 術	18 安全研究における多重ソース・システム構築のための基本安全用語データベースの開発 19 3次元マッピングデータベースの技術開発 20 データベース検索サポートシステムの調査研究 21 グループウェアにおけるデータベースシステムに関する調査研究 22 パーソナルコンピュータとLANの利用による非定形データベースのプロトタイプ作成 23 知的資源型データベースの調査研究	(株)紀伊國屋書店 (株)日本総合技術研究所 セントラル開発(株)情報図書館 RUKIT (株)イフ・アドバタイジング (株)メイテック (株)ジャパンコミュニケーションズインスティテュート

「先端産業分野における専門用語の電子辞書データベース化の調査研究」
平成4年度報告書

目 次

図表目次	3
先端産業用語評価委員会名簿	5
1. はじめに	6
2. 本調査研究の概要	9
2. 1 調査研究の目的	9
2. 2 調査研究の方法	10
2. 3 調査研究結果の概要	12
3. 照合用英和・和英両引き辞書の生成	16
3. 1 両引き辞書の設計	16
3. 2 両引き辞書の生成方式	19
3. 3 両引き辞書の検索機能	23
4. 両引き辞書による和英専門語彙の特性分析	26
4. 1 両引き辞書の統計的特性	26
4. 2 和英訳語の分野別分化の状況	39
4. 3 訳語の連鎖によるクラスター分析	45
5. 和文専門用語の抽出方式	50
5. 1 実験用電子化文献の概要	50
5. 2 簡易分かち書きの方法と結果	50
5. 3 両引き辞書との突き合わせ処理	53
5. 4 未登録語の検出と検定	53
6. 和文専門用語の頻度分析と分野分類	57
6. 1 分野分類の設定	57
6. 2 出現頻度統計の方式	60
6. 3 統計結果	60

7. 専門辞書更新のシステム化	92
7. 1 出現頻度統計と連鎖検索による辞書更新の方式	92
7. 2 機械翻訳システムとの連動性	95
8. 専門用語の和英・英和訳語の特質と問題点	97
8. 1 和訳語形成の歴史的経過と訳し分け	97
8. 2 適用分野と訳し分けの関連	98
8. 3 日本起源の英語の新生術語	99
8. 4 新出術語と専門辞書更新システム	101
9. 今後の展望	103
付属資料	
1. 統合分野分類・原辞書分野分類対照表	104
2. 大学の学科・講座の分野事例	110
3. 訳語の連鎖による用語展開（日本語起点）	117
4. 訳語の連鎖による用語展開（英語起点）	118

< 図表目次 >

図 3-1-1	両引き辞書の項目と内容例	18
図 3-2-1	両引き辞書の作成手順	20
図 3-2-2	両引き辞書の作成	21
図 3-2-3	両引き辞書の作成例	22
図 3-3-1-1	完全一致検索例 (和)	24
図 3-3-1-2	完全一致検索例 (英)	25
図 3-3-2-1	部分一致検索例 (和)	24
図 3-3-2-2	部分一致検索例 (英)	25
表 4-1-1-0	両引き辞書の基本指標	27
表 4-1-1-1	適用分野数の統計	27
表 4-1-2	両引き辞書の分野別・原辞書別構成比	29
表 4-1-3	訳し分けの分野別状況 (英→和)	31
図 4-1-4	訳し分けの分野別状況 (英→和)	32
表 4-1-5	訳し分けの分野別状況 (和→英)	33
図 4-1-6	訳し分けの分野別状況 (和→英)	34
表 4-1-7	英語単語数別・適用分野別用語数	35
表 4-1-8	日本語用語の字種構成パターンと文字数	37
表 4-1-9	日本語字種構成別・適用分野別用語数	38
表 4-2-1	英訳語の適用分野別分化の状況	41
表 4-2-2	和訳語の適用分野別分化の状況	42
表 4-2-3	適用分野数と訳し分けの特異事例 (和→英)	43
表 4-2-4	適用分野数と訳し分けの特異事例 (英→和)	44
図 4-3-1	訳語の連鎖による用語展開の事例 (日本語起点)	46
図 4-3-2	訳語の連鎖による用語展開の事例 (英語起点)	47
図 4-3-1-1	訳語の連鎖による用語展開の事例 (循環型, 日本語起点)	48
図 4-3-2-1	訳語の連鎖による用語展開の事例 (循環型, 英語起点)	48
図 5-1-1-1	実験用電子化文献の例 (一般記事)	51
図 5-1-1-2	実験用電子化文献の例 (情報分野論文要旨)	51
図 5-2-1-1	分かち書き処理の例 (一般記事)	52
図 5-2-1-2	分かち書き処理の例 (情報分野論文要旨)	52
表 5-4-1	辞書未登録の専門用語候補の例 (情報分野論文要旨)	54-56
表 6-1-1	分野分類アンケート調査結果	58-59
表 6-3-1	高頻度の日本語用語 (一般記事)	61-63
表 6-3-2	高頻度の日本語用語 (情報分野論文要旨)	64-66

表 6 - 3 - 3	完全一致・部分一致による高頻度頻度の日本語用語（一般記事）	68-70
表 6 - 3 - 4	完全一致・部分一致による高頻度頻度の日本語用語(情報分野論文)	71-73
表 3 - 3 - 5	完全／部分一致係数順用語表（高係数用語、一般記事）	74-76
表 3 - 3 - 6	完全／部分一致係数順用語表（低係数用語、一般記事）	77-79
表 3 - 3 - 7	完全／部分一致係数順用語表（中係数用語、一般記事）	80-82
表 3 - 3 - 8	完全／部分一致係数順用語表（高係数用語、情報分野論文要旨）	83-85
表 3 - 3 - 9	完全／部分一致係数順用語表（低係数用語、情報分野論文要旨）	86-88
表 3 - 3 - 10	完全／部分一致係数順用語表（中係数用語、情報分野論文要旨）	89-91

先端産業用語評価委員会（執筆分担）

安達 淳	(学術情報センター研究開発部助教授)
生駒 英明	(東京理科大学理工学部教授：8)
石塚 満	(東京大学工学部教授：7.2)
坂本光子	(データベース・エンジニアリング株式会社)
佐藤清子	(データベース・エンジニアリング株式会社)
佐藤安夫	(ロゴヴィスタ株式会社常務取締役)
佐野英之	(ロゴヴィスタ株式会社常務取締役)
須田 了	((社) 日本工学会事務局長)
根岸正光	(学術情報センター研究開発部教授：1-7.1,9)
橋爪宏達	(学術情報センター研究開発部助教授)
山本恵美子	(データベース・エンジニアリング株式会社)
吉田栄治	(ロゴヴィスタ株式会社常務取締役)

事務局

清水節子	(科学技術情報研究所株式会社代表取締役)
大田康詞	(科学技術情報研究所株式会社研究員：統計図表等)
カール・ウィリアムス	(科学技術情報研究所株式会社研究員：統計図表等)

1. はじめに

機械翻訳システムは、今や実用化の段階に入り、多方面での利用が進みつつある。当初、大型計算機でのみ可能であった、翻訳ソフトウェアも、ワークステーションやパソコンの驚異的な性能の向上に伴って、現今では、むしろこうした小型機種向きのシステム開発が活発化しており、先端的翻訳手法によるシステムが、これら小型機上で稼働するようになってきている。一般にいわれる「ダウンサイジング」の傾向をここにもみることができるわけである。こうして、機械翻訳システムは、われわれにとってますます身近なものになりつつあり、幅広い適用が期待されるに至っている。

(1) 機械翻訳システムとダウンサイジング

ところで、こうした身近な存在としての機械翻訳システムを考えると、大きな問題として、翻訳用辞書の更新管理の方法があげられる。大型機による初期的な翻訳システムというのは、要するに比較的少数の、大量の翻訳需要をかかえる特定の企業・機関において、専任の担当者を配置して、組織的な運用を行うことが前提になっていると考えられる。このような場合、翻訳用の辞書ファイルを、それぞれのユーザーで実際に翻訳の対象になる文書・文献に適合するように拡充し、更新してゆくことも、こうした翻訳システム専任担当者の手によって、適切におこなわれると期待される。また、これらのユーザーでは、大量の翻訳を処理させるということから、システムをユーザーの翻訳対象資料に同調させた状態、すなわち「使い込んだ」状態をもってゆくのには要する期間も、相対的に短くなるはずである。

一方、ダウンサイズ化された翻訳システムの場合は、上記のような特定の機関ではなく、ごく普通の企業、機関、また大学の研究室、さらには個人といった、それほど大量の翻訳需要を持たないユーザーにも、手軽に使い、かつそれなりに目に見える効果を発揮するようなものでなければならない。この点からすれば、翻訳システムに対する機能・性能的要件は、従来の大型機用のシステムよりも、むしろ上述のような小型機向きのシステムの方がきびしいという見方が可能である。そして、そこにおける主要な課題のひとつが、先に指摘した翻訳用辞書の拡充、更新、同調の問題であろう。すなわち、専任者がいなくても、また、それほど大量の翻訳処理を経験しないでも、各々のユーザーの実際的な翻訳対象資料に適合した辞書が維持管理されるという状況が、実現されなければならないわけである。

翻訳の品質・精度の向上には、何はともあれ、処理の対象となる資料の具体的性格に見合った、辞書の整備、更新が基礎的な要素になることは論をまたない。そこで、こうした更新管理が、ユーザーの手をわずらわせることなく、いわば自動的に実現されるようなシステムが望まれるところである。このような機能は、翻訳システムにおける解析・学習機能を高度化させることによって、将来的に実現されうるということも、もちろん考えられないわけではない。しかし、大量の時事用語、先端的専門用語

が日々新たに発生するという現今の状況に照らせば、翻訳辞書の拡充・整備上の大きな問題は、こうした新出用語への対処という点に尽きるといってもよいであろう。こうした新出用語への対応は、翻訳アルゴリズムの高度化などとはまた別の範ちゅうの問題、すなわち新出用語を速やかに発見・検出して、その訳語を与えてゆくという、いわば運用システムの問題であると思われる。

(2) 英和翻訳辞書データベースの検討——平成3年度調査研究

本調査研究、「先端産業分野における専門用語の辞書データベース化の調査研究」は、上述のような観点に立って、平成3年度および4年度の2年間にわたって調査研究を進めてきた。すなわち、辞書の主体的な維持管理者として、ユーザーではなく、むしろ翻訳システムの開発・供給者を想定し、それらにおける効率的な先端専門用語の検出と、翻訳用辞書データベースの経常的更新を支援するシステムのありようを中心課題として検討を行ってきた。つまり、ユーザーはこうして常時更新されてゆく辞書の、経常的な配布を受けることによって、常に最新の環境において、翻訳システムを使ってゆくことができるという図式を追求するものとした。

平成3年度においては、英→和の翻訳を主題として、先端専門用語のための英日翻訳辞書の更新・維持管理の方法について検討した。先端産業分野の専門的科学技术文献において、専門的な用語が、どのように、またどの程度用いられているかを、文献を収集して実証的に調査・検討して、電子化専門用語辞書の作成・維持管理のための基礎的なデータを提供することを主眼とした。具体的には、英文の文献を収集して、用語の使用頻度を分析し、またこれら用語を既存の電子化英和辞書数種と突き合わせて、分野別の和訳語を分析した。最近の専門用語では、一般的単語を連ねて特定の専門的事項を表わす、複合語の比重が増していることに鑑み、この種の複合専門句の抽出方式についても実験に基づいた検討を行った。つぎに、専門用語の分析では、専門分野別で状況が大幅に異なることが予想される。この点について、材料、エレクトロニクス、情報の3分野について、各々の専門家による専門語彙の検討を行った。

(3) 和英・英和両用辞書データベースによる辞書更新システム化の研究

——平成4年度調査研究

上記の結果を踏まえて、平成4年度においては、和→英の機械翻訳用辞書のありようを主題のひとつとして検討した。ここで、昨年度の調査研究の結果得られた重要な知見として、翻訳用専門用語辞書における「英→和」と「和→英」の統合的な更新・維持管理という構想がある。機械翻訳システムは、英日、日英ともにすでに出回りつつある。つまり、辞書の需要としては、英和、和英の双方ともにすでに高まっている。このことは、情報流通の国際化がますます進展してゆくことの反映でもあろう。従って、海外の英語文献に発見される新出英語をいち早く英和辞書に登録すること同時に、わが国の和文文献から発見される日本語新出用語に、適切な英訳を与えて、日

本情報の海外発信をより迅速、円滑にしてゆくことも、機械翻訳システムに期待される重大な機能になっている。情報交流の濃密化に伴って、このような新訳語は急速に普及し、英→和として得られた新語は、早晩、和→英の方向に還流して出現するし、また反対の方向、つまり、日本語起源の英訳語のわが国への還流という事例も、当然ながら多発する。

つまり、ここでは、「英→和」あるいは「和→英」という一方向の翻訳に着目した辞書の更新管理は、非常な不効率を招来するであろう。従って、先端用語辞書に関しては、英和、和英の双方向からの利用を当初から折り込んだ統合的な辞書管理の方式が必要である。いわば、英和、和英の両引き辞書、兼用辞書ということと念頭においた検討が必要であり、また可能でありかつまた有効であるというのが、昨年度の「英→和」主体の調査の中で判明してきた。

この知見に即して、平成4年度の調査研究においては、和英両引きの機械翻訳用専門用語辞書の更新システムを中心課題に設定して、調査研究を進めた。すなわち、両引き辞書を設計し、既存の電子化辞書数種を併合しつつ、こうした両引き辞書を実際に生成し、これに英和、和英双方向からの高速検索機能を付加した。つぎに、この両引き辞書の諸特性について、統計的な分析を行った。すなわち、英語については、句を構成する単語数と適用専門分野の関係、日本語については、漢字だけからなる語、カタカナ語等々、字種の構成を、その適用分野別の変位をみた。また、いわゆる訳し分けの状況に関して、昨年度の英語の和訳の際の分野別の訳し分けと対照させて、新たに、日本語の英訳の際の訳し分けの状況を統計した。さらに、日本語の用語からスタートして、その英訳語を得、つぎにその和訳語を得るというような、連鎖的な翻訳をしたときに、どのような結果がえられるかを検討して、潜在的な類義語の群がどのように形成されているかを調査した。

つぎに、日本語の電子化文献を収集し、これをテスト・データとして、上記の両引き辞書と突き合わせながら、新出の日本語専門用語の検出のシステム化に関して検討し、同時に、日本語の頻度統計を行った。この際、日本語処理に固有の問題である分かち書きについても検討し、辞書登録語との完全一致と部分一致の双方を試みている。なお、本調査研究では、専門分野分類の設定が重要な点であることに鑑み、学協会へのアンケート調査等を実施して、適正な分野分類を設定し、これを両引き辞書の編成や分野別分析などに用いた。さらに、辞書更新のシステム化について、英文と和文の双方から、出現頻度統計と連鎖的検索の手法を用いて新出語を析出する方式と、機械翻訳システム自体を応用してゆく方式について検討した。また、上記の各分析を総合的に見渡したところで、専門用語における和英翻訳、英和翻訳の特質と問題点について検討している。

機械翻訳の実用化、普及にあたっては、新語、特に新出専門用語の発見とその辞書登録が、結局において決定的役割を担うものと思われる。本調査研究が、このような専門用語辞書のデータベース化、その更新整備のシステム化に対して、基礎的データと設計上の指針を与えるものとなれば幸いである。

2. 本調査研究の概要

2. 1 調査研究の目的

機械翻訳システムの実用的な翻訳精度・性能は、翻訳手法の高度化という要因もさりながら、むしろ専門用語辞書の充実の度合に依存するところが非常に大きいということが、機械翻訳システムの実用事例の積み重ねの中で、改めて確認されつつあるようである。従って、電子化された専門用語辞書をデータベース的観点に基づいて、効果的、効率的に作成し、またこれを常時更新してゆく必要がある。この場合、こうした専門辞書の拡充・維持管理を、これまでのシステムのように、個別ユーザーの問題であるとして放置してしまおうのではなく、翻訳システムの供給者における経常的ユーザー・サービスの一貫としてとらえた上で、そのための辞書データベースの整備手法を検討するのが有効である。

本調査研究では、先端産業各分野における専門用語の発生・使用状況を実態的に調査・分析することにより、専門用語辞書のためのデータベース構築・運用に関する基礎データを提供し、同時に、このような辞書管理のシステム化における有効な方法、体制について提言することを目的とする。

平成3年度には、英日機械翻訳用の辞書データベースの整備を研究主題として、エレクトロニクス、情報、材料の3専門分野を中心とする、250学協会の5,700の英文論文から400万語を抽出して、用語の出現頻度、共出現頻度、頻度の時系列変化等について調査分析を行い、電子化専門辞書整備の方法を検討した。

平成4年度においては、昨年度の英文専門用語に関する分析をさら進めるほか、新たに日英機械翻訳のための辞書の整備に関して調査研究を行い、この両者を総合して、英日・日英双方向の機械翻訳に適用できる「和英両引きの専門用語辞書データベース」の構築・更新のシステム化について、調査研究を実施した。

機械翻訳システムのための専門用語辞書の更新・維持管理においては、日々新たに大量に発生する先端的専門用語を検出し、それらを翻訳用辞書へ反映させてゆくことが、翻訳システムの実用性能の決め手ともいえることは、上述のとおりである。平成3年度の調査研究結果として、こうした先端的専門用語に関する辞書の整備にあたっては、英語文献をソースとする「英→和」翻訳用の辞書の整備と、日本語文献をもとにする「和→英」翻訳の辞書整備とは、これらを別個のものとして扱うのではなく、英和と和英を統合化して考えるのが有効であるとの結論を得た。確かに、国際的情報流通の進展が著しい現今の環境下において、先端的専門用語については、ことさら翻訳需要が高く、しかもこれは「英語→日本語」翻訳と「日本語→英語」翻訳の要求が同時に発生するはずである。

本調査研究では、このような問題認識に基づいて、英文と和文の専門用語に関する実態調査を行い、それらを比較検討しながら分析して、英日と日英の専門用語辞書デ

データベースの効率的な整備・更新の方法を明らかにする。すなわち、原文献から専門用語を能率的に検出・検定するための手法と、英日・日英辞書の相互変換の可能性に関する研究を実施する。これに基づいて、この種の辞書システムに関わる基礎的データと有効な管理手法や重要な知見を提供することにより、機械翻訳システムの精度の向上や自然語処理に関わる各種システムの開発等に幅広く貢献するものとする。

2. 2 調査研究の方法

本調査研究は、先端的な専門用語を迅速・適確に見つけだして、機械翻訳用辞書にそれを経常的に反映させてゆく方法について検討するものである。ここでは、和文と英文の電子化文献と、電子化用語辞書を応用して、専門用語の使用実態を統計的に明らかにしながら、検討を進めた。

本調査研究の実施主体である科学技術情報研究所 (J-TIES) は、これまで欧米の機関・企業と提携した調査業務を多数実施しており、先端的産業・技術の動向全般にわたって基礎的な資料、知識を集積している。本調査研究においては、こうした蓄積を元にし、さらに関係諸方面から必要な資料を収集し、これに対して分析を加えていった。一方、こうした過程全体にわたって指導を仰ぎ、また分析結果の評価を担当頂くべく、専門家により構成された「評価委員会」を組織した。すなわち、委員会における調査研究全体の方針策定、評価分析と、J-TIESにおける調査作業の実施という分担関係により、調査研究を推進した。

本調査の具体的手順は概略次のとおりである。

(1) 昨年度調査研究結果のレビューと本年度調査計画の策定 —— 先端的専門用語辞書の更新管理においては、英和・和英の統合的处理が有効であることを確認し、この認識に基づいて本年度の調査研究計画を策定した。

(2) 専門分野分類の検討と設定 —— 両引き辞書における用語の適用分野分類の設定にあたって、学協会宛のアンケート調査を実施し、これらを踏まえて、自然科学分野を対象に22の専門分野分類を設定した。

(3) 電子化両引き辞書の設計と作成および検索システムの開発 —— 既存の英日対訳電子化専門用語辞書4種類について、訳語を整理統合して、和英・英和双方から検索できる辞書を設計、作成した。また、これを高速に検索できるシステムを開発付加した。

(4) 両引き辞書の統計的特性の分析 —— 作成された両引き辞書について、英語用語の単語数、日本語用語の字種構成と適用分野の関係を統計し、分野別の専門用語の全般的傾向の把握を行った。

(5) 両引き辞書における訳し分け状況の日英比較分析 —— 英和、和英の双方について、主として専門分野別の訳し分けが行われるが、こうした訳し分けの現状について統計的に分析し、また特異事例を検出、検討した。

(6) 訳語のつながりからみた類義語群の形成状況の分析 —— 英和検索によって得られる和訳語(群)を、和英検索に投入して英訳語(群)を得るという過程を連鎖させることにより、類義語クラスターの析出を試みた。ここでの知見は、辞書更新システムや機械翻訳システムを設計する上で重要である。

(7) 調査対象文献の収集と電子化 —— わが国の代表的な学協会誌等の論文要旨(和文・英文)と通信社による配信記事(和文・英文)の相当量を収集、電子化ファイルとし、両引き辞書との突き合わせ、出現頻度統計等の実験対象データに用いた。

(8) 日本語分かち書き手法の調査と分かち書きシステムの開発 —— 和文文献から専門用語を抽出するためには、まず和文の分かち書き処理が必要である。このとき、用語辞書を用いずに、字種構成から判別して高速でなお一定の精度が得られる手法について調査の上、分かち書きのシステムを開発して、本調査研究に用いた。

(9) 日本語文献における専門用語検出システムの検討 —— 和文の文献から専門的用語を抜き出し、これを辞書と突き合わせながら、新出・未登録の専門語であると確認の上、辞書登録を行うという一連の過程について検討した。

(10) 日本語専門用語の出現頻度統計とその分析 —— 和文の上記調査対象文献について、日本語の専門用語の出現頻度の統計調査を行った。この際、分かち書きの問題があるので、辞書登録用語との全体・完全一致のほか、部分的な一致についても検索して統計し、その結果を検討した。

(11) 出現頻度と連鎖検索による日英両用辞書の更新方式の検討 —— 上記のような本年度の調査分析を、昨年度の調査研究結果と総合して、日英両用の専門用語辞書の更新・維持管理の方式について検討した。

(12) 専門語の検出と辞書更新に対する機械翻訳システムの連動性の検討 —— 本調査研究では、機械翻訳システムの運用とは一応独立の工程として、専門用語辞書整備の方法を検討してきた。そこで、辞書更新システムにおける更新結果を、機械翻訳システムの辞書に反映させてゆく際の問題を中心に、両システムの連動性に検討を加えた。

(13) 専門用語における和訳語、英訳語の特質と問題点の総合的検討 —— 上記のような調査分析を踏まえつつ、わが国と海外における先端的な専門用語発生メカニズム、またその国際間での流通過程など、現状における基本的な枠組みにさかのぼって、翻訳システムのあり方との関係を検討した。

2. 3 調査研究結果の概要

(1) 日英両用電子化辞書の作成とその諸特性

既存の電子化英和対訳専門用語辞書3種と、昨年度の調査研究の過程で得られた新出専門語を収録した独自の辞書の4者について、訳語や適用分野区分等を整理しつつ統合して、和→英、英→和の検索が可能な両引き辞書ファイルを開発した。この辞書は、訳し分けの関係を分解して、「1英文用語：1和文用語」という対応関係をもつものとし、全体で約27万エントリーからなる。異なり語数では、英語21万5千語、日本語21万2千語になる。

全体的にみると、日本語の方が英語よりも1.4%ほど語彙が小さいということであるが、ここでは、日英ほぼ同数であることが注目されよう。英→和における訳し分けと、和→英における訳し分けが、全体としてはほぼ同数という結果であるから、この点に関するかぎり、日英翻訳と英日翻訳の難易度は同等ということでもある。

つぎに、英文用語については、用語を構成する単語数、和文用語については、その字種構成（漢字・ひらがな・カタカナなど）と、適用分野の関係について統計して分析した。これは、たとえば漢語の専門語の多い分野と、カタカナ語の多用される分野といった比較検討であり、分野別専門用語辞書の整備方針に関して、有用な知見が含まれる。

(2) 和訳語・英訳語における専門分野別訳し分けの状況調査

昨年度の調査研究では、ひとつの英文専門用語対して、複数の和訳語が当てられるという、訳し分けの状況を2種の既存電子化辞書について調査した。この場合、訳し分けが専門分野別になっているのであれば、翻訳処理上の問題はかなり軽減される。すなわち、翻訳対象文献の分野区分を予め指定してやればよいからである。しかし、調査の結果をみると、訳し分けが分野毎に特定されるとはいえない例も相当あった。

そこで、本年度では、4種の辞書を統合した和英両引き辞書を用いることにより、対象範囲を拡大すると同時に、日本語からみた英訳語の訳し分けという方向についても調査を行い、英→和と和→英の比較を行った。その結果、たとえば土木や電気電子分野では、英→和に比べて和→英の際の訳し分けが多い（日本語語彙が小さい）のに対して、応用物理では逆に英語語彙の方が小さいといった、分野別の状況が読みとれる。専門分野別に機械翻訳辞書を編集・整備するにあたっては、このような分野別の特性的を見極めた上で、必要な調整を行う必要がある。

(3) 訳語検索の連鎖による類義語クラスターの析出と辞書のシソーラス化

今回の調査研究において開発した英和・和英両引き専門用語辞書の機能を応用して、英→和→英→和→・・・という訳語の連鎖的な検索を行って、翻訳語を媒介とした類義語クラスターの析出を試みた。この際、たとえば「亜鉛」を和英で引いて「ZINC」を得、つぎに「ZINC」を英和で引いて「亜鉛」だけが得られるという「単純往復」が発見されるのであれば、翻訳にとっては、これが最も好ましい状況であるといえる。しかし、現実には、このような1:1の往復関係はそれほど多くないと思われる。むしろ、同義語や類義語そして訳し分けなど結合関係を介して、多数の用語がからみ合った状況も多いのではなかろうか。こうした仮説に基づいて、いくつかの和英の用語を起点として、その関連用語をいわば手繰りよせてゆくという操作により、日英の専門用語がどのような結合関係を形成しているのかについて、その概略的感触を得るものとした。翻訳辞書の整備にあたっては、単純な対訳辞書構成では、翻訳精度にもおよそ限界があると考えられ、ここにのべたような分析を通じて、用語間・訳語間の関連性を内蔵する、まさしくデータベース的なアプローチが必要であろう。これは、従来からシソーラスと称されてきたものに近いので、翻訳辞書のシソーラス化ということもできるが、利用目的を異にしており、従って、構成や編成の手法に関して新たな研究を要するものであろう。

(4) 和文専門用語の抽出方式

和文の文献における専門的用語の出現状況を調査するについては、まず、単位語の抽出が必要である。すなわち分かち書きの工程である。現今、分かち書きの処理は、予め用意された辞書との突き合わせ処理を含む方式が有力である。しかし、辞書の構築方法自体を検討対象とする本調査研究のような場合には、辞書検索を内蔵する分かち書き処理は有効ではない。従って、ここでは、辞書を用いない分かち書きの手法について調査・検討し、国立国語研究所で開発された方法を参考にして、字種の変り目に注目した文章分割の方法を設計し、その処理プログラムを開発した。端的には、たとえば漢字の連鎖、カタカナの連鎖などを語として認定してゆく方法であり、これにより抽出された文字列を辞書と突き合わせながら、専門用語を検出してゆく手順を検討した。

すなわち、分かち書きにより取り出された「語」を、一般語からなる辞書と上述の専門用語辞書とに突き合わせて辞書未登録の語を発見し、これらについて、さらに詳細に検定するという手順、方法について、電子化された和文文献を用いて実験しつつ検討した。なお、辞書との突き合わせ処理では、文字列の部分的一致についても検出するようにして、検索精度を高めた。また、未登録語の検定では、前項にのべた連鎖検索の手法の適用可能性を検討した。

(5) 和文専門用語の出現頻度分析

前項における調査検討結果を踏まえて、和文の電子化文献における専門用語の出現頻度の統計を行った。この際、用語例が辞書登録用語と部分的に一致する場合についても統計し、各用語の独立／接続使用の実態についても明らかにするものとした。統計対象の文献データとしては、通信社の一般記事と情報関連分野の科学技術論文要旨を用い、それぞれにおける高頻度用語を調査し、また上記の部分一致と完全一致の用例数の比率による統計も行った。たとえば、他の用語との接続的使用の多い用語については、翻訳に際に注意が必要であり、むしろ接続例の方をすでに熟した専門用語と認定し、これを新たに辞書登録するといった対応が有効である。本統計調査によって、このような専門用語辞書維持管理に関する有用な知見が、その具体例とともに得られている。

(6) 専門分野分類の検討

上述における両引き辞書の開発や、専門用語の抽出、出現頻度統計等全般にわたって、専門分野別の検討が重要である。そして、その分野分類自体が、本調査研究にとっては前提的に必要とされる重要な項目であった。これは、機械翻訳のための専門用語辞書の分野別編成とも関連する問題である。そこで、本調査研究の一環として、自然科学系の学協会に対してアンケート調査を実施して、その見解を徴した。これを、大学の学科・講座編成、学術会議の研究連絡委員会構成、既存の専門用語辞書にみられる分野分類等と照合して、独自の分類体系を設定し、本調査全体についてこれを適用した。

(7) 専門用語辞書の更新管理のシステム化

情報化と技術革新の進展が急速な今日状況において、日毎に発生する専門的用語群に対して、翻訳用の用語辞書は、これに遅滞なく追従できるものでなければならない。この観点からして、翻訳用専門用語辞書のデータベース化と、その更新管理のシステム化は不可欠であろう。本調査研究では、この点について下記の二方面から検討を加えた。

その第一は、出現頻度統計と連鎖的検索による検定を活用した、日英両用専門用語辞書の更新方式の検討である。ここでは、たとえば翻訳システムの開発・供給者において、和文と英文の文献における用語の出現状況を、適切な統計を採取しながら常時監視し、十分な検定作業の上で、新出用語を認定・登録したのち、これをユーザーに経常的に配布してゆくといった体制が想定されることになる。

第二は、辞書更新システムと機械翻訳システムとの連動化という視点である。本調

査研究では、専門用語辞書の整備を、機械翻訳システムの運用とは一応独立の工程として運用することを念頭に検討してきた。そこで次に、このような辞書更新システムによる更新結果を、機械翻訳システム用の辞書に反映させてゆく際の問題点について、とくに訳語の優先度の指定方法を中心に検討を加えた。

(8) 専門用語における和訳語、英訳語の特質と問題点の総合的検討

上述のような調査分析の全体を踏まえつつ、わが国および外国における先端的な専門用語の発生メカニズムや、その国際間での流通・浸透の過程などについて、先端科学技術分野を中心として、現状の基本的な枠組みを明らかにし、こうした現状に有効に対応できるような機械翻訳システムと、そのための専門用語辞書更新システムのあり方を検討した。翻訳とは基本的に文化的な所為であって、各国の社会的環境が、その濃淡はともあれ、翻訳対象の文献に反映されている。従って、科学技術文献の翻訳においても、その研究の社会的・文化的背景に対する一定の理解が必要である。このようないみで、翻訳用辞書というものは、単なる言葉の置き換え表ではない。本項では、こうした視点もふまえて、専門用語の翻訳に関わる基本的な問題を検討した。

3. 照合用英和・和英両引き辞書の生成

3. 1 両引き辞書の設計

本調査研究では、昨年度の研究の過程で得られた知見に基づき、「英→和」と「和→英」兼用の機械化辞書を作成し、これを用いて和文専門用語の抽出方式を検討することとした。そこで、下記の4種の英和対訳辞書を整理統合して、英和、和英両引きの辞書システムを開発した。

- 1) 「文部省学術用語辞書」(インタープレス社編、科学技術用語大辞典)
- 2) 「J I S用語辞書」(インタープレス社編、科学技術用語大辞典)
- 3) 「マグローヒル科学技術用語大辞典」(日刊工業新聞社編)
- 4) J T I E S新語辞書(科学技術情報研究所作成)

1)～3)は、昨年度調査にも用いた、既成の機械化辞書であるが、これらは本来、「英→和」方向の検索利用を前提に編集されている。従って、今回の作業では、それらの「和→英」検索への逆転をした場合の状況を実験的に確かめながら、英和・和英両引き辞書に統合することを試みた。なお、4)は、昨年度研究の過程で検出された、上記3辞書に未登録の、新出専門用語を集めて独自に編集したものである。

今回開発された両引き辞書の構成は、図3-1-1に示すとおりである。ここで基本的な点は、英語と日本語の1対を辞書の記入単位(エントリー)に設定したことである。一般に、英和辞書なら、一つの英用語に対して複数の和訳語を列挙して、辞書の1エントリーが構成される。和英辞書ならその逆に1和用語とその複数英訳語が辞書の構成単位になる。両引き辞書では、こうした翻訳の方向性を解消にする必要がある。また、ソフトウェアによる高速な検索が可能でなければならない。こうした点を考慮して、上記のような1用語・複数訳語の関係を分解した、図のような1英語・1日本語の対を基本単位とする両引き辞書を設計した。このエントリーにはユニークな番号を割り当てて、機械的な管理を容易にしている。

ただし、実際には上記のような本来的な英和辞書を元にしてしているため、これらとの対応関係を保持するべく、各英用語に「エントリー番号」を割当、それに対する複数英訳語に「枝番」を付す体裁とし、「エントリー番号+枝番」によって英語・日本語の対が同定されるような構成とした。

「分野」は、本研究のために新たに設定した用語の適用分野(後述)表示であり、原辞書に表示されていた適用分野(原辞書分野)を整理統合して割り当てたものである。また「原辞書」欄は、この用語が登録されていた原辞書(上記4種の辞書)の区分を示す。

品詞区分は、翻訳システムを考えた場合、重要な情報となりうる。しかし現状では、これについて正確なデータを登録することが難しい。一方、昨年度調査の経験からみても、専門用語には、名詞の形容詞的用法も多くみられることなどから、基本的

に名詞であるとみて作業を進めることが、實際上効果的であることが分かっている。そこで、将来的な拡張性に配慮して、「品詞」欄を設けたが、現状ではすべてが名詞となっている。

これらの項目以外にも、日本語のよみ（フリガナ）、頻度情報、出典（当該用語の発見された文献情報）等が辞書項目として検討された。フリガナについては、日本語の性格上、限定できない場合も多く、この入力今回は見送った。ただし、漢字表記でもひらがな表記でも使われる用語が一部存在し、これら統合的に管理するために、フリガナは有用であると考えられるので、将来的には辞書に登録されるべきものであろう。また、頻度情報及び出典情報は、本研究の課題である専門用語辞書の更新管理に際しては、重要なデータとなりうるもので、更新管理のための「辞書データベース」には必須の項目であるが、当面の両引き辞書には含まれない。

機械翻訳システムのための電子辞書にどのような情報が必要かは、翻訳アルゴリズムに依存する面が大きい。例えば、数を数える際の単位などは、人か動物か物かによって訳し方が異なってくるから、辞書の項目として、それが人か動物か物かの区分も必要である。当面の両引き辞書システムは、このような機械翻訳用辞書の更新システムを検討するためのものであるから、性格は自ずと異なっており、上記のような構成が適当と判断された。

図3-1-1 両引き辞書の項目と内容例

エントリ- 番号	枝番	英語	日本語	品詞	分野 (複数)	原辞書分野 (複数)	原辞書 (複数)
2009492	1	barn	なや	N	建築	建築	GAKUJUTSU
2009492	2	barn	畜舎	N	建築,農業	建築,農学	GAKUJUTSU
2009492	3	barn	納屋	N	農業	農	NIKKAN
2009492	4	barn	バーン	N	応用物理	原子力	JIS
4005308	1	barnacle	ふじつほ	N	船舶海洋	船舶	GAKUJUTSU
4005308	2	barnacle	フジツボ	N	動物	無脊動	NIKKAN
2009498	1	Barnett effect	バーネット効果	N	情報	電磁	NIKKAN
2009499	1	Barnett method	バーネット法	N	情報	電磁	NIKKAN
2009500	1	barney	バニー	N	材料資源	鉱山	NIKKAN
2037682	1	friability	破碎性	N	材料資源	材料	NIKKAN
2037683	1	friagem	フリাজェム	N	地球科学	気象	NIKKAN
2037684	1	friar's cowl	アコニット	N	生物	植物	NIKKAN
4025219	1	friction	フリクション	N	機械	新語	JTIESNEW

エントリー番 英語、日本語の組み合わせに対応するナンバー

枝番 一つの英語に対して複数の訳語が存在する場合につける

分野 本委員会設定した分野

原辞書分野 原辞書に登録のあった分野名

原辞書 文部省学術用語辞書
(インタープレス社編、科学技術用語大事典)

J I S 用語辞書
(インタープレス社編、科学技術用語大事典)

科学技術用語辞書
(日刊工業新聞編、マグローヒル科学技術大事典)

J T I E S 新辞書
(科学技術情報研究所編、頻度分析・照合結果による補完辞書)

3. 2 両引き辞書の生成方式

両引き辞書の作成手順は図3-2-1に示すとおりである。各原辞書について、まず、前項にのべた辞書形式に統一するため、原辞書の形式を変換・統一する。同時に、分野区分を再編成して重複データを削除する。次に、英用語のスペルチェックを行った。既成の辞書の再編成であるから、スペルチェックは不要のようにも思われるが、昨年度の経験からみて、JTIES独自辞書のみならず、既成の辞書にもスペルミスはある程度存在するので、スペルチェックを施して、より信頼性のある辞書の作成を目指した。こうした作業の後、4種類の辞書を合体させ、重複を削除して、統合化された両引き辞書を生成した。これにより270, 772語（1英用語と1和用語の対の数、エントリー数）の両引き辞書が得られた。

両引き辞書の生成工程に関する具体的な計数を図3-2-2に示す。例えば、学術用語辞書は、元来、土木、物理、電気等々、分野別に編成されているので、これらをいったん統合したのちに、さらに4辞書の全体を統合する計数的にみると、学術用語129, 543語、JIS用語63, 654語、マグローヒル科学技術用語103, 706語、JTIES辞書56, 525語の合計353, 428語が、統合の結果、270, 772語になったということになる。

実際の統合化の過程を、FRICTION-摩擦の周辺に関して、事例的に示したものが図3-2-3である。この場合、対訳に関しては、FRICTIONが摩擦と訳されるという単純な例であるが、その適用分野区分は、原辞書それぞれにおいて独自の分類がなされている。これら分類を統合、再編成した結果が、図下段の両引き辞書の分野欄に示されている。

図3-2-1 両引き辞書の作成手順

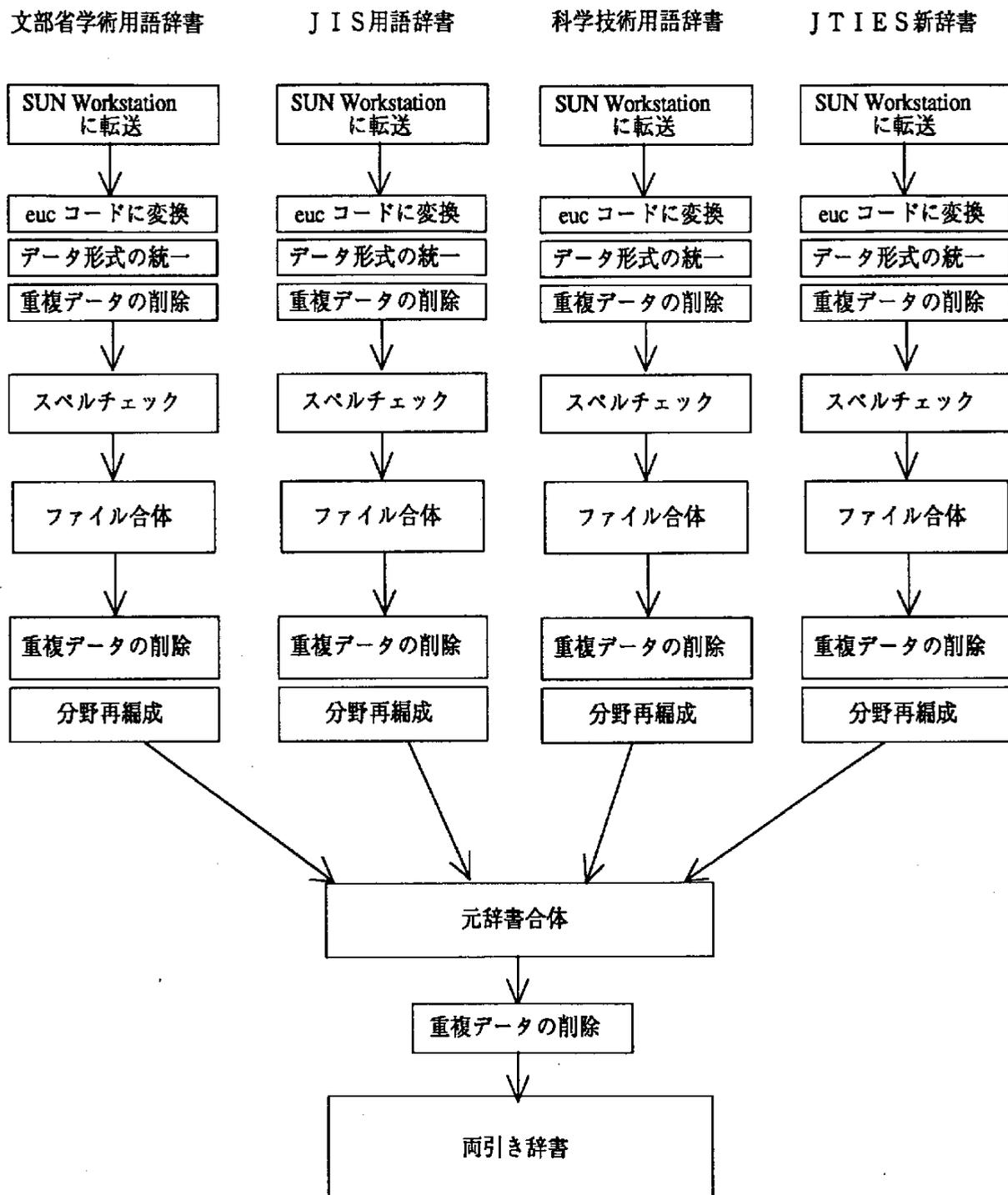


図 3 - 2 - 2 両引き辞書の作成

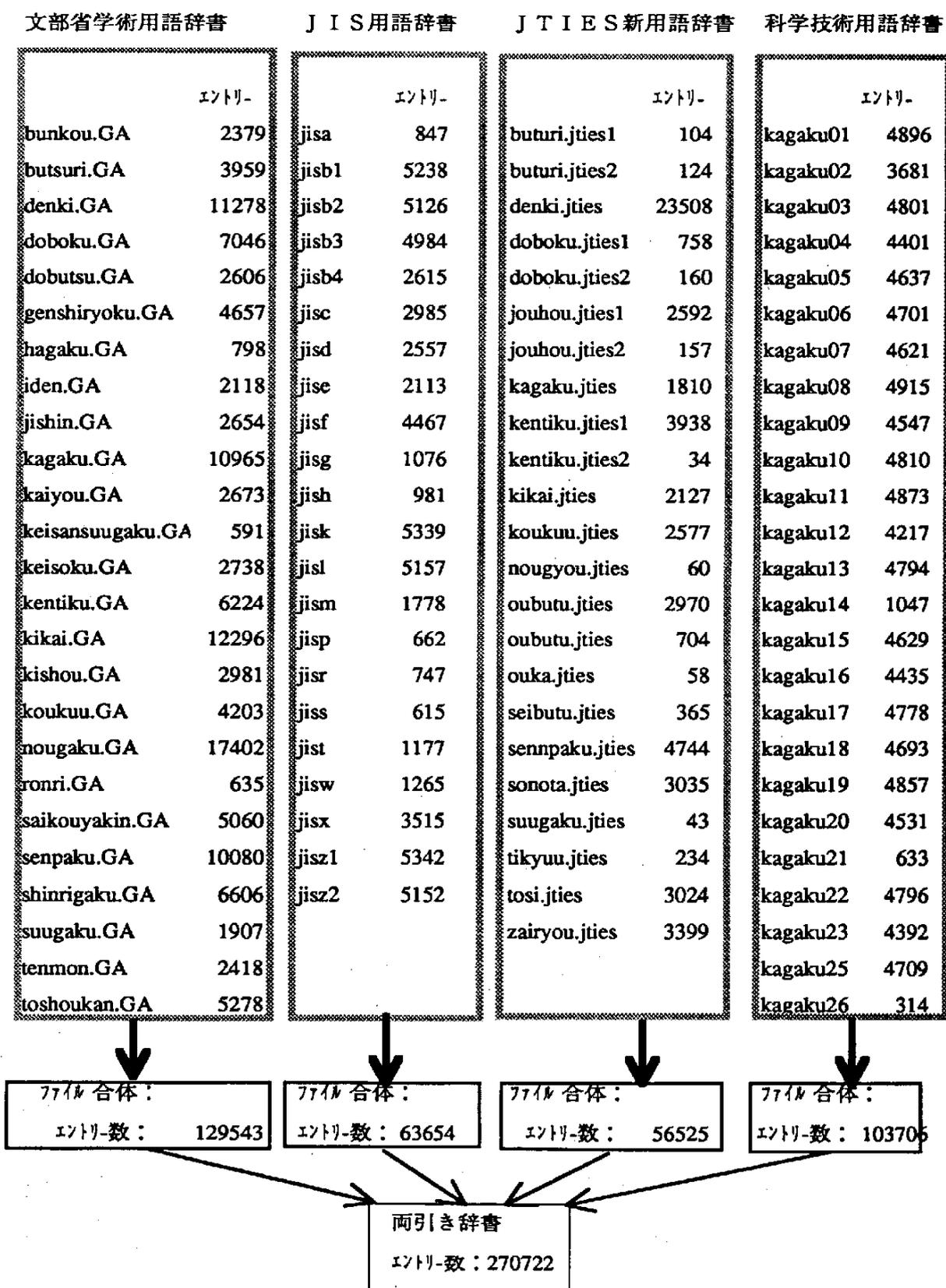


図3-2-3 両引き辞書の作成例

文部省学術用語辞書

JIS用語辞書

JTIES新辞書

科学技術用語辞書

文部省学術用語辞書					JIS用語辞書					JTIES新辞書					科学技術用語辞書				
英語	日本語	品詞	分野	原辞書	英語	日本語	品詞	分野	原辞書	英語	日本語	分野	原辞書	英語	日本語	品詞	分野	原辞書	
fretwork	透し彫	Noun	建築	GAUJUTSU	friction	摩擦	Noun	ブラ	JIS	friction	フリクション	機械	新語 JTIESNEW	Freudianism	フロイト主義	Noun	心	NIKKAN	
friction	摩擦	Noun	化学	GAUJUTSU	friction brake	摩擦ブレーキ	Noun	クラッチ	JIS	friction	摩擦	応用物理	新語 JTIESNEW	friability	破砕性	Noun	材料	NIKKAN	
friction	摩擦	Noun	機械	GAUJUTSU	friction brake	摩擦ブレーキ	Noun	自動車	JIS	friction	摩擦	化学	新語 JTIESNEW	friagem	フリアジェム	Noun	気象	NIKKAN	
friction	摩擦	Noun	計測	GAUJUTSU	friction clutch	摩擦クラッチ	Noun	クラッチ	JIS	friction	摩擦	機械	新語 JTIESNEW	frisar's cowl	アコニット	Noun	植物	NIKKAN	
friction	摩擦	Noun	建築	GAUJUTSU	friction clutch	摩擦クラッチ	Noun	工作機記	JIS	friction	摩擦	建築	新語 JTIESNEW	friction	摩擦	Noun	力	NIKKAN	
friction	摩擦	Noun	探鉱冶金	GAUJUTSU						friction	摩擦	船舶海洋	新語 JTIESNEW	friction brake	摩擦ブレーキ	Noun	機械	NIKKAN	
friction	摩擦	Noun	船舶	GAUJUTSU						friction	摩擦	電気電子通信	新語 JTIESNEW	friction clutch	摩擦クラッチ	Noun	機械	NIKKAN	
friction	摩擦	Noun	地震	GAUJUTSU						friction	摩擦	土木	新語 JTIESNEW						
friction	摩擦	Noun	土木	GAUJUTSU						friction angle	内部摩擦角	船舶海洋	新語 JTIESNEW						
friction	摩擦	Noun	物理	GAUJUTSU						friction clutch	摩擦クラッチ	船舶海洋	新語 JTIESNEW						
friction angle	摩擦角	Noun	機械	GAUJUTSU															
friction brake	摩擦ブレーキ	Noun	機械	GAUJUTSU															
friction brake	摩擦ブレーキ	Noun	船舶	GAUJUTSU															
friction brake	摩擦ブレーキ	Noun	電気	GAUJUTSU															
friction clutch	摩擦クラッチ	Noun	機械	GAUJUTSU															
friction clutch	摩擦クラッチ	Noun	船舶	GAUJUTSU															
friction clutch	摩擦クラッチ	Noun	農学	GAUJUTSU															

両引き辞書

エントリー

番号	枝番	英語	日本語	品詞	分野	原辞書分野	原辞書
1048175	YSu 1	WEw fretwork	VJw 透し彫	VHi N	VNi 建築	VOI 建築	YSO GAKUJUTSU
2037679	YSu 1	WEw Freudianism	VJw フロイト主義	VHi N	VNi その他	VOI 心	YSO NIKKAN
2037682	YSu 1	WEw friability	VJw 破砕性	VHi N	VNi 材料資源	VOI 材料	YSO NIKKAN
2037683	YSu 1	WEw friagem	VJw フリアジェム	VHi N	VNi 地球科学	VOI 気象	YSO NIKKAN
2037684	YSu 1	WEw frisar's cowl	VJw アコニット	VHi N	VNi 生物	VOI 植物	YSO NIKKAN
4025219	YSu 1	WEw friction	VJw フリクション	VHi N	VNi 機械	VOI 新語	YSO JTIESNEW
4025219	YSu 2	WEw friction	VJw 摩擦	VHi N	VNi 応用化学, 応用物理, 機械, 建築, 材料資源, 船舶海洋, 地球科学, 電気電子通信, 土木	VOI ブラ, 新語, 化学, 機械, 力, 建築, 探鉱冶金, 船舶, 地震, 計測, 土木, 物理	YSO JIS, JTIESNEW, GAKUJUTSU, NIKKAN
4025228	YSu 1	WEw friction angle	VJw 内部摩擦角	VHi N	VNi 船舶海洋	VOI 新語	YSO JTIESNEW
4025228	YSu 2	WEw friction angle	VJw 摩擦角	VHi N	VNi 機械	VOI 機械	YSO GAKUJUTSU
3023151	YSu 1	WEw friction brake	VJw 摩擦ブレーキ	VHi N	VNi 機械, 船舶海洋, 電気電子通信	VOI クラッチ, 機械, 自動車, 船舶, 電気	YSO JIS, GAKUJUTSU, NIKKAN
3023155	YSu 1	WEw friction clutch	VJw 摩擦クラッチ	VHi N	VNi 機械, 船舶海洋, 農学	VOI クラッチ, 機械, 工作機記, 新語, 船舶, 農学	YSO JIS, GAKUJUTSU, NIKKAN, JTIESNEW

3. 3 両引き辞書の検索機能

両引き辞書には、高速検索のために和・英2種のハッシュト索引を設けた。これを介することで、用語の完全一致検索は極めて高速に実行される。一方、スキャン方式によって部分一致型の検索が低速ではあるが可能になっている。図3-3-1-1~2、図3-3-2-1~2に、「亜鉛」と「STEEL」による、完全一致および部分一致検索の結果を示す。「亜鉛」は、ZINCのほか、SPELTER、ZINC SPELTERに対応する。また「亜鉛」を含む用語を検索すると、図にみられるように多くの亜鉛化合物等が検索される。「STEEL」は鋼、はがね、ハガネ、スチールという和用語に対応する。また、「STEEL」を含む英用語を検索すると、図にみるような各種の鋼製品などが検索されるという具合である。

専門用語辞書の更新においては、単に新出未登録の用語を発見するだけでなく、それがまさしく新規の登録を要するものであるか否かの「検定」が重要である。昨年の調査でも触れたように、現今では、他とは独立したまったく新しい用語が考案されることは少なく、むしろ、既存の専門用語を組み合わせて、新たな概念を表すようにするという事例が普通である。この状況を考慮すると、新出用語の「検定」のためには、部分一致型の辞書検索が有効であることが予想される。

図 3 - 3 - 1 - 1 完全一致検索例
「亜鉛」による検索例

エントリー	枝番	英語	日本語	品詞	分野	原辞書分野	原辞書
2088880	2	spelter	亜鉛	N	材料資源, 船舶海洋	採鉱冶金, 船舶	GAKUJUTSU
1129387	1	zinc spelter	亜鉛	N	材料資源	採鉱冶金	GAKUJUTSU
4056365	1	zinc	亜鉛	N	その他, 応用物理, 化学, 建築, 材料資源, 船舶海洋, 都市工学	新語, 原子力, 化, 化学, 建築, 採鉱冶金, 船舶	JTIESNEW, NIKKAN GAKUJUTSU,
4056442	1	Zn	亜鉛	N	化学, 材料資源	新語	JTIESNEW

図 3 - 3 - 2 - 1 部分一致検索例 (一部)

エントリー	枝番	英語	日本語	品詞	分野	原辞書分野	原辞書
4056386	1	zinc fume fever	亜鉛熱	N	都市工学	新語	JTIESNEW
2103506	1	zinc halide	ハロゲン化亜鉛	N	応用化学	無機	NIKKAN
4056387	1	zinc halogen battery	亜鉛ハロゲン電池	N	情報	新語	JTIESNEW
2103507	1	zinc hydroxide	水酸化亜鉛	N	応用化学	無機	NIKKAN
3063611	1	zinc immersion process	亜鉛置換法	N	材料資源	アルミ	JIS
4056388	1	zinc isotope	亜鉛同位体	N	その他	新語	JTIESNEW
2103508	1	zinc metaarsenite	メタ亜ヒ酸亜鉛	N	応用化学	無機	NIKKAN
2103508	2	zinc metaarsenite	亜ヒ酸亜鉛	N	応用化学	無機	NIKKAN
2103510	1	zinc naphthenate	ナフテン酸亜鉛	N	応用化学	有機	NIKKAN
4056389	1	zinc ore	亜鉛鉱	N	その他	新語	JTIESNEW
2103511	1	zinc orthoarsenate	オルトヒ酸亜鉛	N	応用化学	無機	NIKKAN
2103511	2	zinc orthoarsenate	ヒ酸亜鉛	N	応用化学	無機	NIKKAN
2103513	1	zinc orthophosphate	オルトリン酸亜鉛	N	応用化学	無機	NIKKAN
2103513	2	zinc orthophosphate	リン酸亜鉛	N	応用化学	無機	NIKKAN
4056390	1	zinc oxid paper	酸化亜鉛感光紙	N	材料資源	新語	JTIESNEW
3063612	1	zinc oxide	亜鉛素	N	応用化学, 応用物理	塗料, 新語	JIS, JTIESNEW
2103516	1	zinc phosphate	リン酸亜鉛	N	応用化学	無機	NIKKAN
2103517	1	zinc phosphide	リン化亜鉛	N	応用化学	無機	NIKKAN
4056395	1	zinc plating	亜鉛めっき	N	その他	新語	JTIESNEW
1129385	1	zinc point	亜鉛点	N	電気電子通信	計測	GAKUJUTSU
1129386	1	zinc protector	保護亜鉛	N	船舶海洋	船舶	GAKUJUTSU
4056397	1	zinc selenide	セレン化亜鉛	N	その他, 応用化学	新語, 無機	JTIESNEW, NIKKAN
4056398	1	zinc smelting	亜鉛精錬	N	その他	新語	JTIESNEW
2103520	1	zinc spar	菱亜鉛鉱	N	材料資源	鉱物	NIKKAN
1129387	1	zinc spelter	亜鉛	N	材料資源	採鉱冶金	GAKUJUTSU
2103521	1	zinc spinel	亜鉛尖晶石	N	材料資源	鉱物	NIKKAN
3063614	1	zinc spraying	亜鉛溶射	N	その他	溶射	JIS
3063615	1	zinc stearate	ステアリン酸亜鉛	N	応用化学	塗料	JIS

図 3 - 3 - 1 - 2 完全一致検索例
「steel」による検索例

エントリー	枝番	英語	日本語	品詞	分野	原辞書分野	原辞書
4050250	1	steel	はがね	N	機械,材料資源	新語,採鉱冶金	JTIESNEW,GAKUJUTSU
4050250	2	steel	スチール	N	機械	新語	JTIESNEW
4050250	3	steel	ハガネ	N	化学	新語	JTIESNEW
4050250	4	steel	鋼	N	化学,機械,建築,材料資源, 船舶海洋,地球科学,土木	新語,機械,建築, 採鉱冶金,船舶,地震	JTIESNEW,GAKUJUTSU, NIKKAN

図 3 - 3 - 2 - 2 部分一致検索例 (一部)

エントリー	枝番	英語	日本語	品詞	分野	原辞書分野	原辞書
1111487	2	<u>steel</u> yard	竿ばかり	N	機械,電気電子通信	工学,機械,計測	GAKUJUTSU,NIKKAN
1111487	3	<u>steel</u> yard	棒ばかり	N	機械	工学	NIKKAN
4050608	1	structural <u>steel</u> pile	鋼杭	N	建築	新語	JTIESNEW
1113085	1	structural <u>steel</u>	構造用鋼	N	建築,材料資源,土木	建築,冶金,土木	GAKUJUTSU,NIKKAN
1113085	2	structural <u>steel</u>	構造用鋼材	N	建築	新語	JTIESNEW
4050609	1	structural- <u>steel</u> shape	構造用形鋼	N	建築	新語	JTIESNEW
1115505	1	synthetic <u>steel</u>	合せ鋼	N	材料資源	採鉱冶金	GAKUJUTSU
1115623	1	T- <u>steel</u>	T形鋼	N	機械,建築,材料資源	機械,建築,採鉱冶金	GAKUJUTSU
4051575	2	telescope <u>steel</u> prop	伸縮鋼管支柱	N	建築	新語	JTIESNEW
1118068	1	Thomas <u>steel</u>	トーマス鋼	N	材料資源	採鉱冶金	GAKUJUTSU
1119320	1	tool <u>steel</u>	工具鋼	N	機械,材料資源,船舶海洋	機械,採鉱冶金,船舶	GAKUJUTSU,NIKKAN
3058407	1	torsion test <u>steel</u> wire	鋼線のねじり試験	N	材料資源	鉄鋼	JIS
1121936	1	tungsten <u>steel</u>	タングステン鋼	N	材料資源,船舶海洋	採鉱冶金,船舶	GAKUJUTSU,NIKKAN
3059835	1	twist test of <u>steel</u> wire	鋼線のねじり試験	N	材料資源	鉄鋼	JIS
1122594	1	U- <u>steel</u>	みぞ形鋼	N	建築	建築	GAKUJUTSU
4053056	1	unalloyed <u>steel</u>	炭素鋼	N	その他	新語	JTIESNEW
1123024	1	unequal angle <u>steel</u>	不等辺山形鋼	N	材料資源	採鉱冶金	GAKUJUTSU
4053355	1	untemper <u>steel</u>	非調質鋼	N	その他	新語	JTIESNEW
3060886	1	valve <u>steel</u>	バルブ鋼	N	材料資源	鉄鋼	JIS
3060886	2	valve <u>steel</u>	弁用鋼	N	材料資源	採鉱冶金	GAKUJUTSU
1124085	1	vanadium <u>steel</u>	バナジウム鋼	N	材料資源	採鉱冶金	GAKUJUTSU,NIKKAN
1124995	1	vibrac <u>steel</u>	バイブラック鋼	N	船舶海洋	船舶	GAKUJUTSU
4055148	1	welded <u>steel</u> pipe	溶接鋼管	N	その他,建築	新語,建築	JTIESNEW,GAKUJUTSU
4055400	1	wide band <u>steel</u>	帯鋼	N	その他	新語	JTIESNEW
3063322	1	wrap test of <u>steel</u> wire	鋼線の巻付試験	N	材料資源	鉄鋼	JIS
3063372	1	wrought <u>steel</u> wheels	圧延車輪	N	材料資源	鉄鋼	JIS
4055924	1	wrought <u>steel</u>	加工異鋼	N	その他	新語	JTIESNEW
4055924	2	wrought <u>steel</u>	鍛鋼	N	材料資源	採鉱冶金	GAKUJUTSU
1129234	1	Z- <u>steel</u>	Z形鋼	N	機械,材料資源	機械,採鉱冶金	GAKUJUTSU
4056378	1	zinc coated <u>steel</u>	トタン	N	その他	新語	JTIESNEW

4. 両引き辞書による和英専門語彙の特性分析

4. 1 両引き辞書の統計的特性

(1) 両引き辞書の基本指標

今回作成した両引き辞書の基本的計数は表4-1-1-0のようである。すなわち、「1英用語：1和用語」という対応関係を単位として、全体で約27万エントリーからなっている。これを和英の異なり語数でみると、英語215, 226語、日本語211, 631になる。つまり、日本語の方が英語よりも1.4%ほど語彙が小さいということになる。これら異なり語数をエントリー数270, 772で割ると、英語、日本語における訳し分けの概況を示す指標が得られる。これを仮に「訳し分け指標」と称すると、英語1.26、日本語1.28という数値になる。

日常的感覚としては、漢字熟語の多い日本語の方が語彙が大きいような感じもたれるところであるが、確立された専門科学用語という範囲で集計すると、日英ほぼ同数の用語が使われているということになる。そして、英→和における訳し分けと、和→英における訳し分けが、全体としてはほぼ同数という結果であるから、この点に関するかぎり、日英翻訳と英日翻訳の難易度は同等ということでもあろう。同時に、このことは、英日と日英の用語対訳をそれぞれに分けて考えるよりも、両引き辞書という一つのデータベースとして、本研究を進めることの妥当性を示唆しているとみることができる。

(2) 適用分野数と分野別訳し分け

表4-1-1-1は、両引き辞書の各エントリーに関して、その適用分野の数を統計したものである。これによれば、9割近くのエントリーが単一の分野に適用されることがわかる。ここでは、各エントリーが1英用語：1和用語の対からなっているという点が重要である。つまり、この両引き辞書では、「1英用語=1和用語=1適用分野」という明解な対応関係をもつエントリーが全体の90%に達する。このことは、専門用語の翻訳では、場合による訳し分けの問題が、対象文献の分野を特定することによって、大部分は解決されるということをいみする。確かに訳し分けは分野ごとに行われるのが普通であると思われるが、この点が、数字として、ある程度確認されたことになるであろう。同時に「1英用語：1和用語：適用分野」という辞書エントリーの構成方式は、機械翻訳用辞書についても有効であることが示唆されている。

なお、2分野以上に適用されるエントリーも、実数として3万程存在することがわかる。これらについて、今後詳細な検討を行って、対応訳語の適否判別と分野分類の再検討を進めることにより、分野別訳し分けの効率の高い辞書にすることができよう。

表4-1-1-0 両引き辞書の基本指標

語彙規模		訳し分け指標	
エントリー数 (A)	270772	-	
英語実数 (B)	215226	1.26	(A/B)
日本語実数 (C)	211631	1.28	(A/C)

表4-1-1-1 適用分野数の統計

適用分野	エントリー数	構成比 (%)
1	240,931	88.98
2	20,031	7.40
3	5,271	1.95
4	2,147	0.79
5	1,050	0.39
6	561	0.21
7	341	0.13
8	172	0.06
9	121	0.04
10	74	0.03
11	39	0.01
12	19	0.01
13	12	0.00
14	2	0.00
15	1	0.00
合計	270,772	100

(3) 分野別構成と原辞書との対応

両引き辞書の分野別構成を統計すると、表4-1-2左欄のようになる。複数適用分野をもつエントリーがあるため、分野別のエントリー数の単純合計は約32万になり、したがって、エントリーあたりの平均適用分野数は1.18である。

この辞書の分野別構成をみると、電気電子通信分野と機械分野がそれぞれ4万7千語(17.5%)、4万2千語(15.7%)で、他分野を大きく引き離している。これは、ここに掲げたような科学技術関係専門語彙全体を概観した結果として、興味深いものであろう。上記2分野が、語彙の大きさにおいて他の分野を圧倒しているということの、実体的意味あいも、どのように解釈するかは問題であるが、辞書構成における分野分類の立場だけからみると、これら大規模分野はいくつかに分解される方が望ましいということになる。分類の効率の一つのめやすとして、分類された結果がほぼ均等になること、つまり、等分割に近くなることがいわれる。こうした観点からして、上記2分野の再分割が検討されうることになる。もっとも、電気電子通信や機械というくくりの細部に立ち入った分類が、はたして有効に設定できるかどうかは、大いに疑問のあるところである。この場合、例えば電気と電子と通信に分解するとして、これら3者に重複的に適用される用語数がどの程度に抑えられるかという点、逆にいえば、各3分野独自の用語が相当の比率を占めるようになるかどうかということである。しかし、電気電子通信分野、機械分野ともに、応用、適用範囲の広い融合化された分野であることを考えると、こうした再分割は有効ではないように思われる。すなわち、ここに示された語彙の大きさは、分類設定上の問題に起因するのではなく、当該分野自身の性格を表しているとみるのが当面妥当であろう。ともあれ、用語の分野間での重複的適用の状況を詳細に統計して、分類設定を検討してゆくことが、専門辞書の効率的な編成にとって必要であろう。

つぎに、同表の右欄において、原辞書との対比をみると、ここに原辞書の性格が現れている。語彙自体は、学術用語辞書15万語、科学技術用語辞書13万、JIS用語8万という程度であるが、分野によって各辞書の強弱がある。まず、機械分野用語は多くをJIS用語に負っていることがわかる。応用化学も同様の傾向にある。農業分野用語は大部分学術用語からのものである。同様に学術用語の比重の大きい分野としては、化学、応用物理、土木、建築などある。科学技術用語辞書からは、地球科学、情報、生物、動物分野の用語を多く採録することができた。結果として、これら3種の辞書を併合することにより、科学技術分野全般をよくカバーする辞書が得られたといえるであろう。

表4-1-2 両引き辞書の分野別・原辞書別構成比

分野	全体		原辞書別							
	(A)	A/C(%)	文部省学術用語辞書		科学技術用語辞書		J I S用語辞書		J T I E S辞書	
建築	11,338	4.19	7,114	4.76	2,170	1.69	1,925	2.46	4,776	6.12
航空宇宙	11,083	4.09	6,882	4.60	5,245	4.08	1,839	2.35	3,296	4.23
機械	42,415	15.66	14,852	9.93	14,354	11.16	23,369	29.89	4,781	6.13
材料資源	21,825	8.06	7,006	4.69	11,969	9.30	5,491	7.02	4,936	6.33
化学	10,867	4.01	9,092	6.08	3,348	2.60	1,538	1.97	2,315	2.97
応用化学	16,313	6.02	3,091	2.07	8,443	6.56	9,096	11.63	1,095	1.40
応用物理	14,506	5.36	9,046	6.05	6,238	4.85	3,844	4.92	4,738	6.07
生物	8,219	3.04	3,084	2.06	5,860	4.55	652	0.83	776	0.99
数学	5,226	1.93	2,623	1.75	3,389	2.63	272	0.35	537	0.69
船舶海洋	20,089	7.42	13,286	8.89	3,812	2.96	5,826	7.45	5,862	7.51
情報	14,011	5.17	1,954	1.31	8,842	6.87	4,131	5.28	3,781	4.85
地球科学	15,224	5.62	6,055	4.05	11,372	8.84	756	0.97	1,519	1.95
農業	17,578	6.49	17,404	11.64	2,555	1.99	898	1.15	1,210	1.55
物理	4,571	1.69	2,701	1.81	2,539	1.97	376	0.48	638	0.82
電気電子通信	47,289	17.46	14,468	9.68	12,185	9.47	7,226	9.24	25,494	32.68
都市工学	4,094	1.51	2,139	1.43	2,109	1.64	848	1.08	3,058	3.92
土木	9,317	3.44	7,766	5.19	2,425	1.88	1,420	1.82	1,911	2.45
動物	9,610	3.55	2,780	1.86	7,879	6.12	90	0.12	275	0.35
林業	380	0.14	29	0.02	380	0.30	1	0.00	2	0.00
その他	36,503	13.48	18,134	12.13	13,553	10.53	8,584	10.98	7,007	8.98
合計 (B)	320,458	(B/C)								
エトリ-数(C)	270,772	1.18	149,506		128,667		78,182		78,007	

(4) 訳し分けの分野別の状況

図表4-1-3~6には、英和、和英の翻訳における訳し分けの状況を、分野に統計して掲げた。表4-1-3は、英用語に対する和訳語という観点から、ひとつの英用語について、いくつの和訳語（エントリー）が登録されているかを、分野別にみたものである。表の右端には各分野の英用語数に対する和訳語数の倍率を、「訳し分け指数」として掲げた。この指数は1.05から1.33の範囲であり、おおむね1.15前後のものが多い。和訳における訳し分けの多い分野は、都市工学、機械、情報である。図4-1-4は、この状況を分野別の相対比率でグラフ化したものである。これによれば、1対1の翻訳関係にある用語が少ない（訳し分けを要する）分野というのが明瞭になっており、都市工学、応用物理、情報、機械がそれにあたる。これら分野の文献は、その分野を特定しても、なお和訳時の訳し分けが比較的多いという分野であり、機械翻訳上の要注意分野ということになる。

同様の統計を、和用語を起点とし、それに対する英訳語の訳し分け状況について行った結果が、表4-1-5と図4-1-6である。「訳し分け指数」は、1.11~1.23の範囲にあり、1.20近辺のものが多い。前出の英→和に比べると、和→英の方が総体的に訳し分けが多いという状況である。1対1の対訳関係の比率は、一般に70%前後であるが、化学、生物、物理、都市工学、林業においては、その比率が80%近辺で高いことが注目される。

(5) 複合英用語の比率

英文の専門用語では、単語ではなく、複数単語からなるもの（複合語）も多い。これらの比率を分野別に統計したものが、表4-1-7である。全体的には、2語からなるものが55%を占め、1語（単語）のものが20%、3語のものが18%といった状況である。つまり、専門用語では、単語はむしろ少なく、多くは2単語からなる用語になっている。つぎに、分野別の変位をみると、生物、動物、都市工学では単語用語の比率が高いのに対し、機械と電気電子通信分野では、単語用語が少なく、3語以上の用語も30%を超えるような状況にあることがわかる。複合語の存在は、英日翻訳の際の注意点のひとつであるが、上記の機械、電気電子通信の両分野は、この点で特段の注意が必要であろう。

図4-1-3 訳し分けの分野別状況 (英→和)

訳数 分野	日本語の訳語数 (訳し分け数)												英語 数	和訳語 数	訳し 分け 指数
	1	2	3	4	5	6	7	8	9	10	11	12			
建築	9,092	842	123	29	6	3	0	0	2	0	1	0	10,098	11,338	1.12
航空宇宙	8,762	1,002	73	16	5	0	0	0	1	0	0	0	9,859	11,083	1.12
機械	29,831	4,075	832	249	98	42	13	6	3	0	2	1	35,152	42,415	1.21
材料資源	16,845	1,975	244	53	16	1	0	0	0	0	0	0	19,134	21,825	1.14
化学	8,845	735	109	36	10	2	1	0	0	0	0	1	9,739	10,867	1.12
応用化学	12,743	1,543	121	19	5	2	0	1	0	0	0	0	14,434	16,313	1.13
応用物理	9,851	1,612	267	78	35	14	5	3	0	0	0	0	11,865	14,506	1.22
生物	6,592	679	79	8	0	0	0	0	0	0	0	0	7,358	8,219	1.12
数学	3,758	583	74	15	4	0	0	0	0	0	0	0	4,434	5,226	1.18
船舶海洋	15,231	1,895	241	63	16	1	1	0	0	0	0	0	17,448	20,089	1.15
情報	9,622	1,558	268	72	21	5	4	1	0	1	0	0	11,552	14,011	1.21
地球科学	11,252	1,724	133	22	6	0	1	0	0	0	0	0	13,138	15,224	1.16
農業	13,756	1,368	231	69	16	5	1	0	0	0	0	0	15,446	17,578	1.14
物理	3,743	380	15	2	3	0	0	0	0	0	0	0	4,143	4,571	1.10
電気電子通信	39,000	3,240	411	89	26	8	6	0	0	0	0	0	42,780	47,289	1.11
都市工学	2,373	498	118	43	18	12	3	2	0	0	0	0	3,067	4,094	1.33
土木	7,641	655	77	21	7	0	1	0	1	0	0	0	8,403	9,317	1.11
動物	7,596	833	85	20	1	0	0	1	0	0	0	0	8,536	9,610	1.13
林業	343	17	1	0	0	0	0	0	0	0	0	0	361	380	1.05
その他	27,907	3,220	406	156	34	12	2	5	2	0	0	0	31,744	36,503	1.15

図 4 - 1 - 4 訳し分けの分野別状況 (英→和)

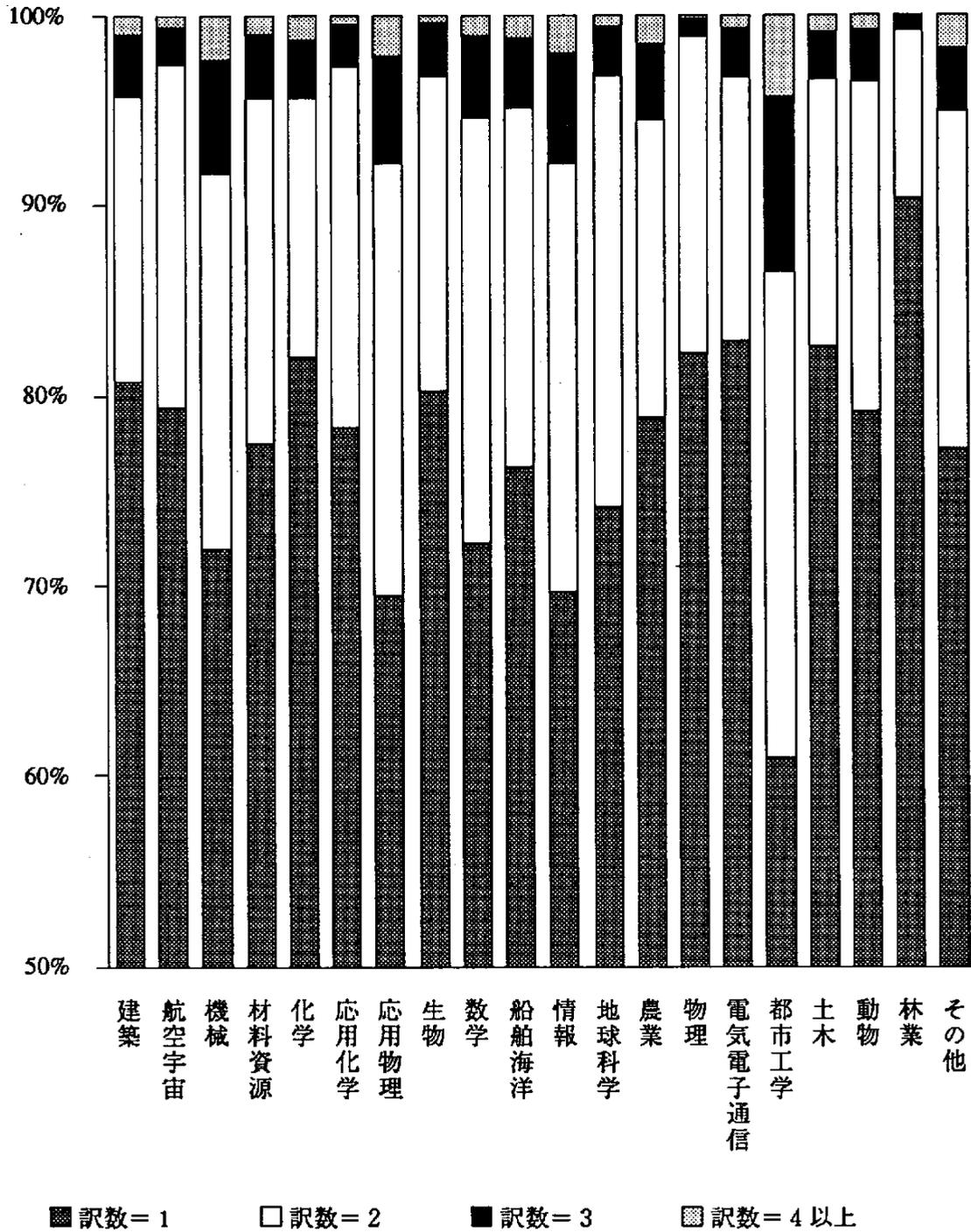


図4-1-5 訳し分けの分野別状況 (和→英)

訳数 分野	英語の訳語数 (訳し分け数)														和語 数	英訳語 数	訳し 分け 指数
	1	2	3	4	5	6	7	8	9	10	11	12	13	14			
建築	7,659	1,122	256	81	37	13	6	3	0	0	0	0	0	1	9,178	11,338	1.24
航空宇宙	7,437	1,358	220	50	10	2	0	1	0	0	0	0	0	0	9,078	11,083	1.22
機械	28,963	4,258	985	283	84	42	16	7	1	0	0	0	0	0	34,639	42,415	1.22
材料資源	15,470	1,954	444	139	56	21	11	6	2	1	0	0	0	0	18,104	21,825	1.21
化学	8,865	802	93	22	5	1	0	0	0	0	0	0	0	0	9,788	10,867	1.11
応用化学	11,356	1,734	338	85	12	10	1	1	0	0	0	0	0	0	13,537	16,313	1.21
応用物理	10,506	1,506	220	52	16	3	2	1	0	0	0	0	0	0	12,306	14,506	1.18
生物	6,657	596	81	24	5	1	0	0	0	0	0	0	0	0	7,364	8,219	1.12
数学	3,721	562	98	15	3	2	0	0	0	0	0	0	0	0	4,401	5,226	1.19
船舶海洋	13,613	2,351	360	110	38	7	2	1	0	0	0	0	0	0	16,482	20,089	1.22
情報	9,500	1,399	316	94	47	13	6	3	0	1	0	0	0	0	11,379	14,011	1.23
地球科学	10,820	1,271	328	113	36	14	8	6	3	2	1	0	0	0	12,602	15,224	1.21
農業	12,761	1,738	292	76	22	5	3	0	0	0	0	0	0	0	14,897	17,578	1.18
物理	3,692	337	44	7	4	0	0	2	1	0	0	0	0	0	4,087	4,571	1.12
電気電子通信	35,456	4,017	752	240	70	24	8	3	1	0	0	0	0	0	40,571	47,289	1.17
都市工学	3,365	259	38	13	1	4	0	2	0	0	0	0	0	0	3,682	4,094	1.11
土木	6,219	1,043	215	63	14	4	3	0	0	0	0	0	0	0	7,561	9,317	1.23
動物	7,162	926	143	29	9	1	0	0	0	0	0	0	0	0	8,270	9,610	1.16
林業	298	31	5	0	1	0	0	0	0	0	0	0	0	0	335	380	1.13
その他	25,116	3,639	768	215	91	35	19	9	6	1	1	0	0	0	29,900	36,503	1.22

図4-1-6

訳し分けの分野別状況 (和→英)

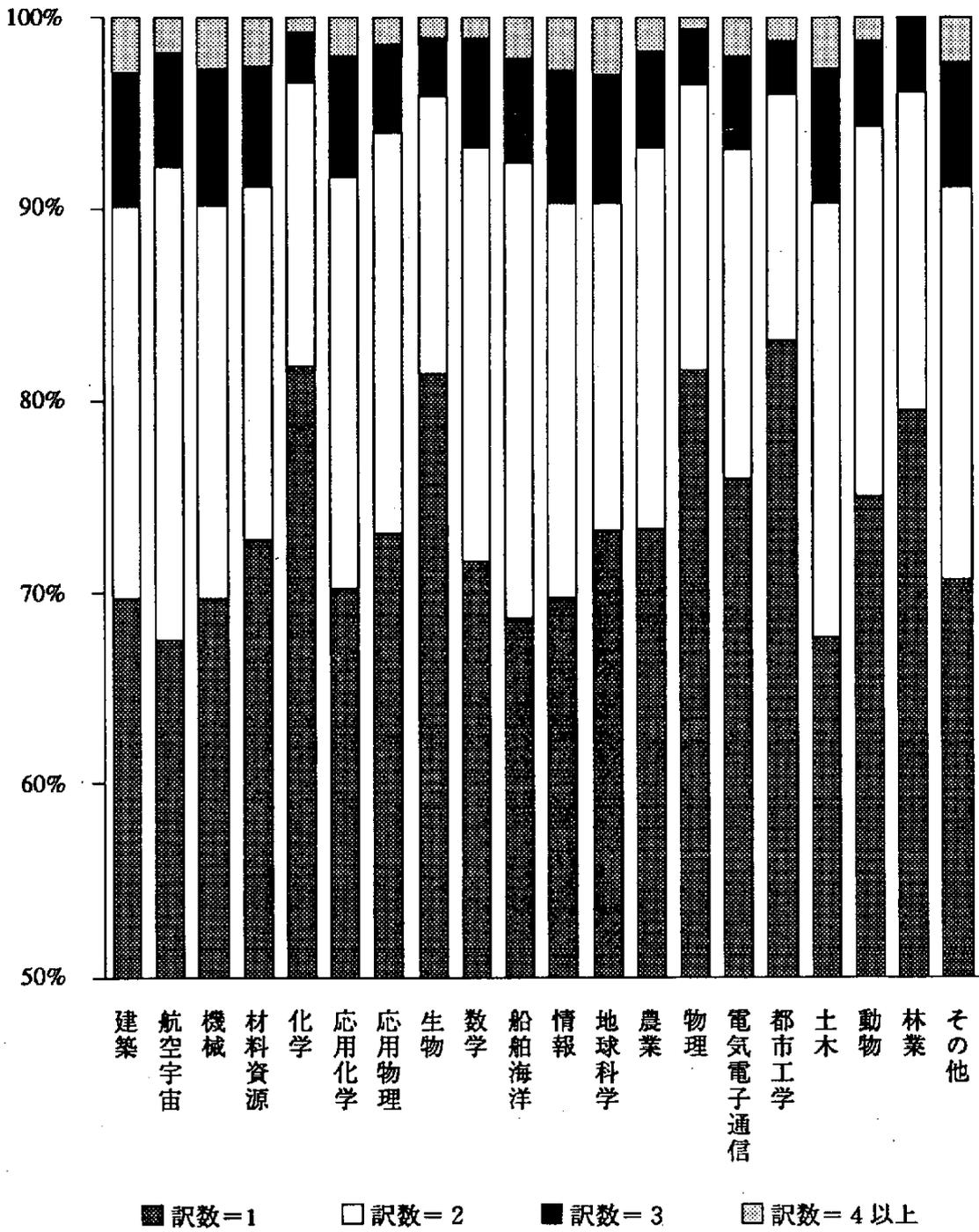


表4-1-7 英語語数別適用分野別用語数 (%)

適用分野	単語数															合計	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
建築	25	56	14	3	1	1	0	0	0	0	0	0	0	0	0	100	10,098
航空宇宙	22	57	16	4	1	0	0	0	0	0	0	0	0	0	0	100	9,859
機械	15	53	21	7	3	1	0	0	0	0	0	0	0	0	0	100	35,152
材料資源	30	51	13	4	1	0	0	0	0	0	0	0	0	0	0	100	19,134
化学	34	54	9	2	0	0	0	0	0	0	0	0	0	0	0	100	9,739
応用化学	25	55	14	4	1	1	0	0	0	0	0	0	0	0	0	100	14,434
応用物理	29	52	14	3	1	0	0	0	0	0	0	0	0	0	0	100	11,865
生物	63	32	4	1	0	0	0	0	0	0	0	0	0	0	0	100	7,358
数学	23	59	12	4	1	0	0	0	0	0	0	0	0	0	0	100	4,434
船舶海洋	19	58	16	5	1	1	0	0	0	0	0	0	0	0	0	100	17,448
情報	20	56	19	4	1	0	0	0	0	0	0	0	0	0	0	100	11,552
地球科学	32	56	9	2	0	0	0	0	0	0	0	0	0	0	0	100	13,138
農業	29	56	11	3	1	0	0	0	0	0	0	0	0	0	0	100	15,446
物理	20	61	14	3	1	0	0	0	0	0	0	0	0	0	0	100	4,143
電気電子通信	13	54	22	9	1	0	0	0	0	0	0	0	0	0	0	100	42,780
都市工学	55	40	4	1	0	0	0	0	0	0	0	0	0	0	0	100	3,067
土木	28	60	9	2	1	0	0	0	0	0	0	0	0	0	0	100	8,403
動物	71	27	2	0	0	0	0	0	0	0	0	0	0	0	0	100	8,536
林業	22	69	9	0	0	0	0	0	0	0	0	0	0	0	0	100	361
その他	30	53	12	3	1	0	0	0	0	0	0	0	0	0	0	100	31,744
全体	20	55	18	5	1	1	0	0	0	0	0	0	0	0	0	100	215,226

(6) 和用語の字種構成

英用語における複合語の統計に対応するものとして、和用語における字数を統計し、同時に、字種構成パターンの分析を行った。以前は、欧米起源の専門用語の翻訳にあたって、漢字の造語力を活かした漢字用語の考案が多く行われたが、昨今はむしろ単なるカタカナ表記で翻訳に替えることが多く行われるようになってきている。こうした状況を分野別に統計してみようというわけである。

表4-1-8がその結果である。和用語全体の45%程が漢字だけからなる用語であり、専門用語においては、漢字語優位の状態にある。もっとも、カタカナだけの用語も16%程に達している。これにカタカナを含む用語(カタカナ+漢字など)を加えると、36%程度になるから、今やこうしたカタカナ語も相当の比重をもっている。

用語の文字数についてみると、漢字語では、4文字語が32%、3文字語が22%、5文字語が17%という結果である。カタカナだけの語の場合は、6文字、7文字の語がそれぞれ17%、8文字語が13%である。カタカナ+漢字の語では6文字語が20%、漢字+カタカナの語でも6文字語が21%で、最頻値になっている。

和用語の字種構成を分野別にみたものが、表4-1-9である。漢字語優位の分野は、農業、都市工学、数学であり、逆に応用化学分野では漢字語の比率は低い。カタカナだけの語の比率に注目すると、数学、農業、物理では、その比率が低く、一方、応用化学分野において、これが32%というとび抜けて高い数字になっている。昨今におけるカタカナ語多用の傾向からすると、先端的分野でこうした傾向を示す数字が現れるものと予想されるが、こうした先端的分野は、表の分類のそれぞれに内包される形となるため、表の上に直接現れてはいないと思われる。

表 4-1-8 日本語用語の字種構成パターンと文字数

字種構成 パターン	例	用語数	文字数別構成比														
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15+
漢字のみ	連続的非単調電圧制御特性	94,729	0.58	11.94	21.55	31.80	17.27	9.50	4.10	1.96	0.77	0.30	0.14	0.04	0.01	0.01	0.01
ひらがなのみ	るつばばさみ	1,208	0.66	22.10	24.09	30.13	14.90	5.79	2.07	0.17	0.08	0.00	0.00	0.00	0.00	0.00	0.00
カタカナのみ	ワンショットアイテム	32,655	0.02	1.07	4.85	8.24	11.30	17.24	17.08	13.28	9.45	6.26	4.04	2.65	1.85	1.02	1.65
カタ・漢字	ヴラソフ方程式	23,035	0.00	0.06	1.65	7.60	16.75	20.09	17.87	13.99	9.07	5.24	3.06	1.78	1.22	0.66	0.96
漢字・カタ	鷺口瘡カンジダ	14,947	0.00	0.02	0.90	5.43	20.53	21.30	15.72	13.70	9.29	5.69	3.16	1.88	1.20	0.51	0.67
漢字・ひら・漢字	頸ふん団	9,598	0.00	0.00	6.64	20.14	26.55	19.23	12.36	6.65	4.23	2.04	1.07	0.52	0.31	0.10	0.16
漢字・ひら	鉤状砂し	6,067	0.00	3.20	31.27	28.93	21.13	9.31	4.01	1.32	0.43	0.28	0.07	0.03	0.03	0.00	0.00
ひら・漢字	わん状変形	3,883	0.00	2.55	21.79	26.32	24.77	14.40	5.82	2.73	0.90	0.44	0.21	0.05	0.03	0.00	0.00
漢字・カタ・漢字	瀝青質ケツ岩	3,771	0.00	0.00	0.19	1.83	6.50	13.71	19.20	18.80	13.95	9.63	6.89	4.03	2.41	1.54	1.33
記号・漢字	ψ 関数	2,224	0.00	7.01	19.29	19.11	22.75	11.92	6.56	4.90	2.74	1.30	1.30	0.72	0.72	0.63	1.03
カタ・ひら・漢字	ヴィヤット・マルガレスの方程式	1,883	0.00	0.00	0.00	0.37	4.20	11.47	17.42	18.59	14.82	9.56	6.74	5.63	4.04	2.87	4.30
カタ・漢字・カタ	ワンスルー燃料サイクル	1,508	0.00	0.00	0.00	0.13	0.86	2.32	6.76	11.94	17.11	15.78	13.06	10.48	6.76	4.71	10.08
漢字・ひら・カタ	携みホース	1,407	0.00	0.00	0.00	1.21	8.81	24.59	19.97	13.79	12.15	7.18	4.12	2.56	1.92	1.71	1.99
記号	π	1,270	1.42	10.63	47.87	21.50	7.64	3.86	1.73	0.94	0.55	0.79	0.47	0.31	0.39	0.16	1.73
記号・カタ	π モード	1,066	0.00	0.00	0.38	7.69	10.88	13.51	12.01	8.91	8.72	9.76	5.25	6.10	3.85	3.00	9.94
漢字・ひら・漢字・ひら	咬合の高さ	1,005	0.00	0.00	0.00	21.39	27.26	21.19	15.72	6.57	4.48	1.69	1.00	0.20	0.20	0.20	0.10
カタ・ひら	ワイヤカットれんが	491	0.00	0.20	0.41	5.09	18.74	26.48	20.16	11.00	8.15	3.87	2.44	1.63	0.20	0.61	1.02
ひら・カタ	わらバルブ	299	0.00	0.33	1.00	3.68	20.07	29.77	18.73	14.05	6.35	4.01	1.34	0.33	0.00	0.00	0.33
その他	シリコンMOSFET中の移動度	10,585	0.00	0.04	0.60	4.02	7.32	10.37	12.06	12.64	11.54	9.75	7.70	6.13	4.52	3.28	10.01

表4-1-9 日本語字種構成別適用分野別用語数 (%)

適用分野	字種構成								合計	
	漢字のみ	カタのみ	ひらのみ	漢・カ	漢・ひ	カ・ひ	漢・カ・ひ	その他		
建築	46	17	3	11	18	1	3	1	100	9,178
航空宇宙	56	13	0	16	10	0	2	3	100	9,078
機械	38	20	1	20	13	1	5	2	100	34,639
材料資源	40	23	1	22	10	1	2	2	100	18,104
化学	49	17	1	19	10	1	2	1	100	9,788
応用化学	26	32	1	23	8	0	2	7	100	13,537
応用物理	55	14	1	20	6	0	2	3	100	12,306
生物	55	13	0	25	4	0	0	2	100	7,364
数学	65	4	1	12	9	0	8	2	100	4,401
船舶海洋	51	15	1	18	11	1	2	1	100	16,482
情報	37	21	0	27	6	0	4	5	100	11,379
地球科学	55	21	0	14	7	0	2	1	100	12,602
農業	70	7	1	7	13	0	0	1	100	14,897
物理	52	8	1	18	11	1	5	5	100	4,087
電気電子通信	41	11	0	23	10	0	6	8	100	40,571
都市工学	66	18	1	7	5	0	0	3	100	3,682
土木	52	14	2	12	17	1	3	1	100	7,561
動物	54	10	1	28	7	0	0	0	100	8,270
林業	50	11	0	19	18	0	2	0	100	335
その他	61	11	1	13	9	0	2	3	100	29,900
全体	45	15	1	21	11	0	4	4	100	211,631

4. 2 和英訳語の分野別分化の状況

昨年度の調査においては、英用語を元にして、それらに対する和訳語の分野別の訳し分けの状況について調査した。本年度には、前記の両引き辞書により、和用語に対する英訳語の分化状況を調査し、英→和、和→英の双方における分野別の訳し分けの状況を比較調査することにした。

(1) 英訳語の適用分野別分化の状況

表4-2-1は、和用語からみて、その英訳語の訳し分けの数と適用分野数の関係を統計したものである。和用語21万語のうち17万語(81%)は、単一の英訳語に対応しており、これらは翻訳上問題を生じないという点で、扱い易い用語ということになる。また、訳し分けのある用語についても、その訳し分けが分野に対応している場合も、翻訳対象文献の分野を特定することで、それなりに処理が可能である。問題は、一つの分野に限っても、文脈的な訳し分けが行われるという種類の用語であり、こうしたものは、表の主対角線より下部の3角形の領域に計上されることになる。これらは、21万語中の13%程度になっている。とくに、1適用分野、2英訳語のものが1万8千語強あり、翻訳上、これらの処理の効率化が望まれるところである。

これらについては、例えば、より細分化された訳語の適用分野を設定するというのが、ひとつの方法であろう。ただし、この場合、単に概念的に分野区分全般を細かくするというのは有効とは思われない。むしろ、これら用語を個別に点検しながら、訳し分けを特定できるような分野区分を、各用語個別に考えるというゆき方が必要となるであろう。

(2) 和訳語の適用分野別分化の状況

表4-2-2は、前表と同様の統計を、英用語に対する和訳語に関して行ったものである。これを前表と比較すると、1和訳語の英用語は21万5千語中の17万8千語(83%)を占め、英訳語の場合より多少とも比率が高い。従って、和→英より英→和の方がその分簡単であろうということになるが、實際上、有意な差を生ずる程のものかどうかは疑問である。むしろ、和→英に比べて、英→和の方が、訳し分け数の上方への分散が全般的に大きいことの方が問題であろう。もっとも、主対角線下方の用語数の比率は9%程度であるから、分野を特定した場合の訳語の確定可能性は、和→英よりも、英→和翻訳の方が高くなろう。漢字の造語力を活かした分野別の翻訳語創造という先人の努力が、この辺に現れているのかも知れない。

ともあれ、1適用分野、2訳語の英用語約1万5千語に関しては、前項と同様に、個別的な検討を踏まえた、適用区分の設定が有効と思われる。

(3) 訳語分化の特異事例

表4-2-1~2において検出された、多訳語、多適用分野をもつ和英用語を具体的に示すと、表4-2-3(和→英)、表4-2-4(英→和)のようである。多分野・多訳語の用語とは、表の右下付近の用語であり、多分野・共通訳語のものは表右上に、少数分野・多訳語のものは表左下に、それぞれ計上された用語の具体例である。英→和については、カタカナ長音の有無など、英語のカナ表記にゆれに起因する和訳語数の増大例もみられる。これらは、実際には、英→和翻訳の際には問題にならないと思われるが、和→英翻訳時には、原文献に表記のゆれがあるので、和英辞書への登録は必要なものである。

表4-2-1 英訳語の適用分野別分化の状況

分野数 訳数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	合計	訳数
1	157,552	10,538	2,311	828	367	172	102	49	25	20	6	2	2	0	0	0	171,974	171,974
2	18,165	6,592	1,966	762	360	207	85	52	40	15	10	10	0	1	0	0	28,265	56,530
3	3,169	1,863	968	438	231	143	88	57	35	24	10	8	4	1	0	0	7,039	21,117
4	885	569	381	234	132	94	63	32	23	16	9	6	7	1	0	0	2,452	9,808
5	284	203	144	119	86	52	41	29	18	18	12	5	4	4	2	0	1,021	5,105
6	102	57	71	61	37	26	25	17	11	8	8	2	3	0	1	0	429	2,574
7	42	30	21	28	26	17	15	7	11	7	3	4	2	1	0	1	215	1,505
8	26	7	8	14	11	7	11	10	5	4	8	5	1	0	0	0	117	936
9	5	5	4	5	4	5	5	7	7	2	0	1	0	1	0	0	51	459
10	3	3	5	4	5	1	3	1	4	0	2	1	1	1	0	0	34	340
11	2	1	0	3	1	2	2	0	1	2	0	1	0	0	0	0	15	165
12	0	0	1	0	0	0	1	3	2	3	0	0	0	0	1	0	11	132
13	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	13
14	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	14
15	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	30
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	3	51
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	19
合計	180,235	19,869	5,880	2,496	1,260	727	443	265	182	119	68	47	25	10	4	1	211,631	
訳数	209,560	33,311	12,126	5,989	3,327	2,078	1,477	950	713	481	305	236	129	55	28	7		270,772

表4-2-2 和訳語の適用分野分化の状況

分野数 訳数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	合計	エントリ-数
1	165,344	10,138	2,090	667	248	119	69	18	15	11	6	4	1	0	0	0	0	178,730	178,730
2	14,720	8,915	2,123	741	352	182	84	41	26	17	11	2	1	0	0	0	0	27,215	54,430
3	1,137	1,885	1,248	539	274	174	105	60	26	17	8	5	4	0	0	1	1	5,484	16,452
4	182	339	418	360	190	110	81	45	20	13	14	11	2	1	0	1	0	1,787	7,148
5	30	92	134	151	119	75	54	51	24	21	11	10	2	2	0	0	0	776	3,880
6	7	16	37	58	78	56	39	33	27	23	12	2	5	0	1	0	0	394	2,364
7	3	3	11	19	34	49	44	34	22	18	20	8	4	4	1	0	0	274	1,918
8	4	2	8	13	22	16	23	16	20	20	16	5	4	2	1	2	0	174	1,392
9	0	1	2	8	6	14	23	20	12	13	8	6	3	4	3	1	0	124	1,116
10	0	0	4	0	7	5	11	12	6	10	8	4	3	2	0	1	0	73	730
11	0	0	0	1	2	3	5	7	11	7	10	8	3	2	3	0	0	62	682
12	0	0	0	0	2	3	3	2	11	1	6	5	4	4	1	0	0	42	504
13	0	0	0	1	0	1	2	2	1	5	3	3	4	0	0	0	0	22	286
14	0	0	0	1	0	0	0	0	2	3	6	4	1	1	0	0	0	18	252
15	0	0	0	0	0	0	0	2	1	0	2	4	0	5	0	2	3	19	285
16	0	0	0	0	0	0	0	0	1	0	0	3	1	0	0	0	0	5	80
17	0	0	0	0	0	0	0	0	0	0	1	0	1	1	3	1	0	7	119
18	0	0	0	0	0	0	0	0	0	0	0	2	3	1	1	1	0	8	144
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	19
20	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1	0	5	100
21	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2	42
22	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	22
23	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	23
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	24
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	30
合計	181,427	21,391	6,075	2,560	1,334	807	545	343	223	183	149	83	50	28	17	10	1	215,226	
エントリ-	199,168	35,581	12,843	6,676	4,181	2,885	2,336	1,737	1,283	1,157	1,127	671	478	307	232	107	3		270,772

表 4 - 2 - 3 適用分野数と訳し分けの特異事例 (和→英)

多分野・多訳語の例		
層	分野数	15 地球科学, 農業, 土木, 都市工学, 応用物理, 化学, 機械, 建築, 材料, 資源, 情報, 船舶海洋, 電気電子通信, 動物, その他, 数学
	訳数	12 bed, course, facies, formation, lamella, layer, layers, panniculus, ply, seam, sheaf, stratum
支柱	分野数	13 その他, 化学, 機械, 材料資源, 応用物理, 農業, 電気電子通信, 建築, 航空宇宙, 船舶海洋, 土木, 地球科学, 物理
	訳数	17 column, mast, mine timbering, pillar, pole, pole brace, post, prop, push brace, shore, stake, standards, stay, strut, support, timbering, upright

多分野・共通訳語の例		
安定性	分野数	14 生物, 農業, その他, 応用物理, 化学, 機械, 航空宇宙, 材料資源, 情報, 船舶海洋, 地球科学, 電気電子通信, 都市工学, 土木
	訳数	2 homeostasis, stability
応答	分野数	15 電気電子通信, 情報, その他, 応用物理, 化学, 機械, 建築, 航空宇宙, 材料資源, 数学, 地球科学, 都市工学, 土木, 動物, 農業
	訳数	5 answer, answering, reply, response, system response
ねじれ	分野数	13 建築, 化学, 機械, 数学, 船舶海洋, 地球科学, 電気電子通信, 物理, 応用化学, 材料資源, 土木, 農業, その他
	訳数	5 spiral grain, torsion, torsion group, twist, twisting
誤差	分野数	13 その他, 応用化学, 応用物理, 機械, 建築, 航空宇宙, 材料資源, 情報, 数学, 電気電子通信, 都市工学, 土木, 農業
	訳数	1 error

少数分野・多訳語の例		
らせん階段	分野数	2 建築, 船舶海洋
	訳数	15 VIS, caracole, circular stair, cockle stair, corkscrew stair, helical stair, newel stair, screw stair, solid-newel stair, spiral stair, spiral staircase, vice stair, vise, winding stair, winding stairs
耳折れ表紙	分野数	1 その他
	訳数	11 binding with circuit edges, binding with divinity circuit, binding with divinity edges, binding with yapp edges, box edge, circuit edge, divinity circuit, divinity edges, divinity style, turn-in, yapp

表 4 - 2 - 4 適用分野数と訳し分けの特異事例 (英→和)

多分野・多訳語の例			
core	分野数	15	その他, 機械, 電気電子通信, 応用化学, 応用物理, 建築, 航空宇宙, 材料資源, 情報, 船舶海洋, 地球科学, 都市工学, 土木, 農業, 化学
	訳数	30	しん金, ウェブ, ケーブル心線, コア, コアー, スプール, テープスプール, 果心, 核, 巻しん, 巻心, 岩心, 岩芯, 合板の心板, 磁心, 心, 心材, 心線, 心板, 芯材, 錐芯, 中子, 中心, 中心核, 中心部, 鉄心, 胴, 縄しん, 練り心, 炉心
scale	分野数	16	応用物理, 都市工学, 動物, 化学, 機械, 電気電子通信, 農業, その他, 材料資源, 船舶海洋, 土木, 建築, 地球科学, 応用化学, 数学, 生物
	訳数	20	うろこ, はかり, ものさし, りん片, りん毛, スケール, スケールきず, 音階, 垢, 尺度, 縮尺, 直尺, 湯あか, 湯垢, 物差, 物差し, 目盛, 目盛尺, 鱗, 鱗片

多分野・共通訳語の例			
brightness	分野数	16	材料資源, 都市工学, 物理, その他, 建築, 航空宇宙, 土木, 応用化学, 機械, 応用物理, 化学, 情報, 船舶海洋, 地球科学, 電気電子通信, 農業
	訳数	4	かがやき, 輝度, 白色度, 明るさ
density	分野数	17	機械, 応用物理, 化学, その他, 応用化学, 建築, 航空宇宙, 材料資源, 情報, 数学, 船舶海洋, 地球科学, 電気電子通信, 都市工学, 土木, 農業, 物理
	訳数	3	デンシテイ, 濃度, 密度

少数分野・多訳語の例			
equalizer	分野数	5	機械, 情報, 電気電子通信, 材料資源, その他
	訳数	12	イクオライザ, イクオライザー, イコライザ, スライド平衡装置, 均圧管, 均圧結線, 均圧母線, 均熱装置, 釣合いばり, 等化器, 平衡装置, 補償要素
poly	分野数	4	農業, 応用化学, 化学, 電気電子通信
	訳数	14	ポリ, ポリ-, ポリアクリル酸エステル, ポリビニルアセタール, ポリビニルアルコール, ポリビニルカルバゾール, ポリビニルピロリドン, ポリビニルプチラール, ポリビニルホルマール, ポリメタクリル酸メチル, ポリメチルビニールエーテル, ポリ塩化ビニリアン, ポリ塩化ビニル

4. 3 訳語の連鎖によるクラスター分析

(1) 連鎖検索の方法と事例

つぎに、和英の訳語の相互関連について、その概況をつかむために、和英の訳語を連鎖的に検索して、整理してみるという実験を行った。英作文の際に、まず和英辞書をひき、それだけでは要領を得ないので、そこに現れた英単語を、さらに英和辞書でひいてみるといったことは、日頃経験するところである。ここでの実験はこれと同等のことである。もっとも、ここでは、両引き辞書システムの検索機能を用いて、和→英→和→英→・・・という連鎖を、6段階まで網羅的にたどっている。

ここには、日本語を起点とした事例として、「宇宙」からスタートした場合と、英語を起点にする例で、「cycle」からスタートした場合を掲げた。実際の連鎖検索結果はかなりの紙面を要し、それほど見やすくはないので、その要約版を本文中に挿入し、検索結果自体は、付属資料に収容した。すなわち、「宇宙」の要約版は、図4-3-1であり、その詳細が<付属資料3>である。同様に「cycle」の要約版は図4-3-2で、詳細版が<付属資料4>になっている。

まず、「宇宙」をみてみると、これに対する英訳として、cosmos、cosmic、space、universeが得られる。そこで、これら英訳語に対する和訳語を検索すると、cosmosとcosmicについては、宇宙だけが得られて、この方向は完結する。一方spaceに関しては、宇宙だけにはとどまらず、「宇宙空間」、「すきま」、「間隔」、「間隔文字」、「込め物」、「空間」、「歯みぞ」、「あき領域」、「スペース」、「SP」など、実に多様な和訳語が現れる。この場合、適用分野は、物理学、情報処理、建築、機械などであって、spaceに対する訳し分けの状況が示される。これらのspace関連の和訳語について、ふたたび英訳を検索してみると、clearance、gap、opening、interstice、stemming等々のあらたな言葉が現れるので、これらに対する和訳語をつぎに検索することになる。また、universeに対する和訳語には、宇宙のほかに、「母集団」、「祖集団」という統計用語が現れ、その英訳であるpopulationから、「人口」、「個体数」、「原子滞在数」などが検索されるという具合である。

「cycle」から始めた場合は、その和訳に「周波」、「周期」、「循環」、「輪体」、「サイクル」が現れ、それらの英訳を検索すると、period、circulation、percolation、circuit等々が得られる。これにより、以下、電気関係、図書館学、医学、地学等へ用語の展開が行われる。

図 4 - 3 - 1 訳語の連鎖による用語展開の事例（日本語起点）

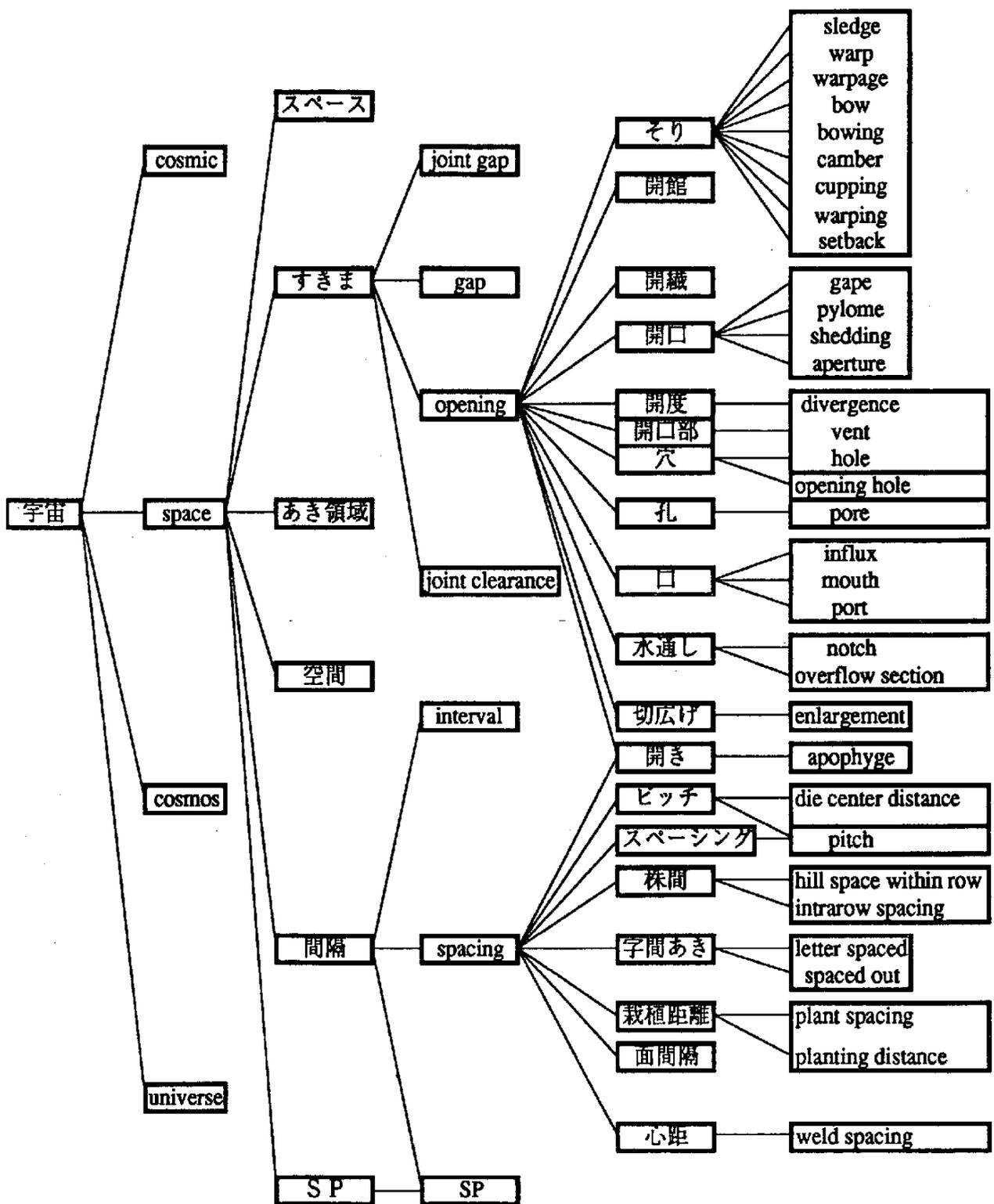
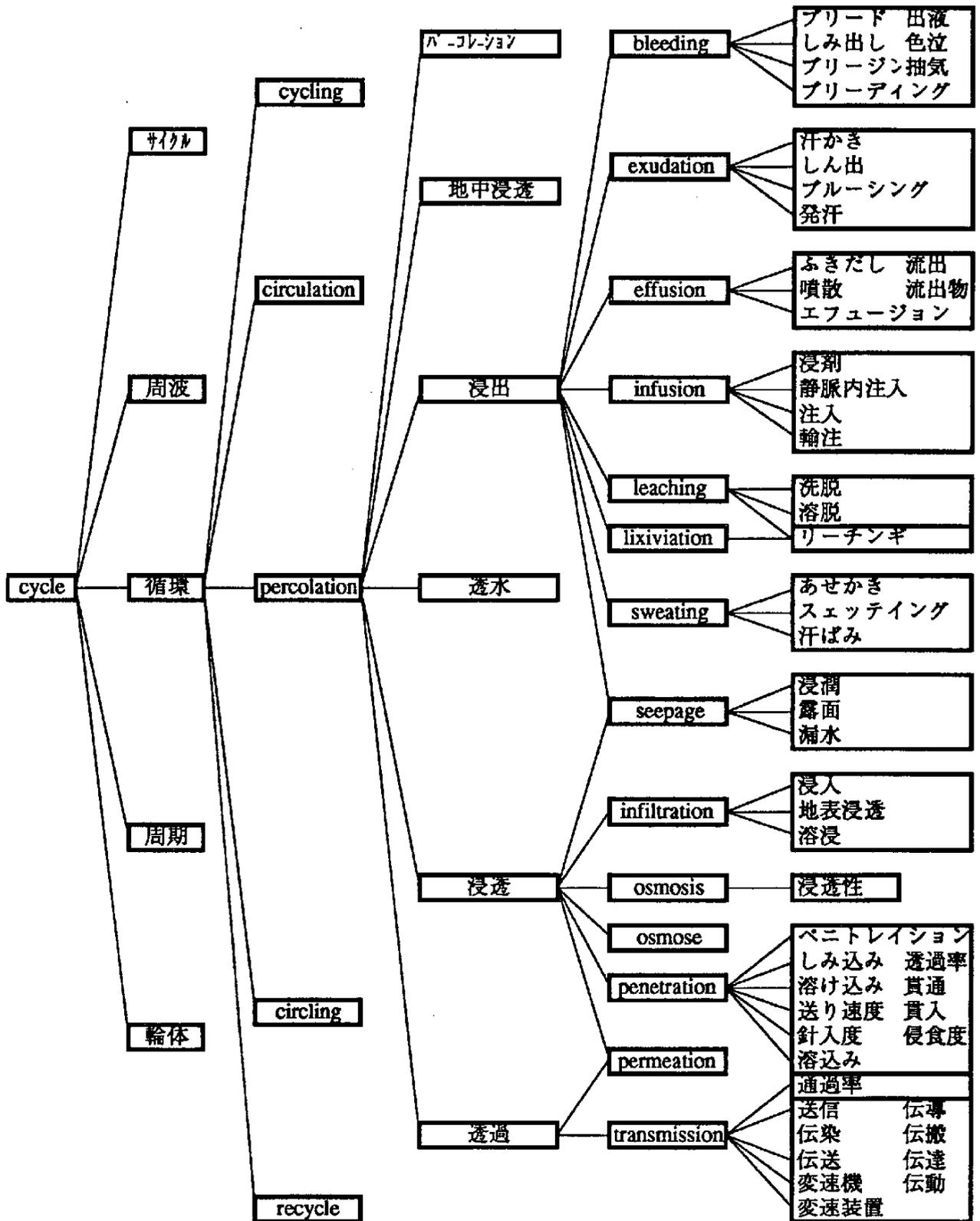


図4-3-2 訳語の連鎖による用語展開の事例（英語起点）



訳語の連鎖による用語展開の事例（循環型）

図4-3-1-1

日本語起点

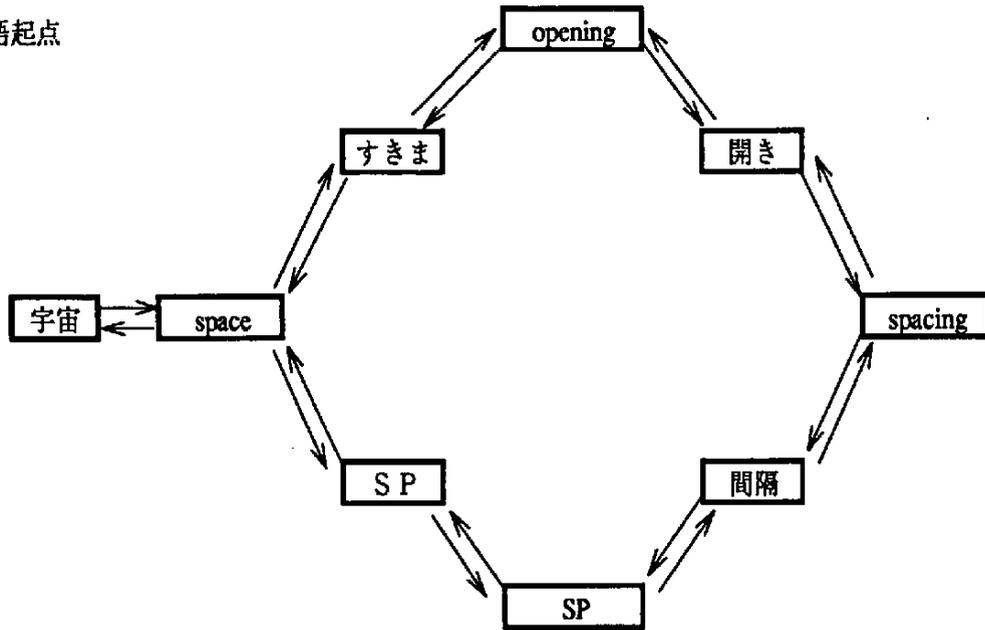
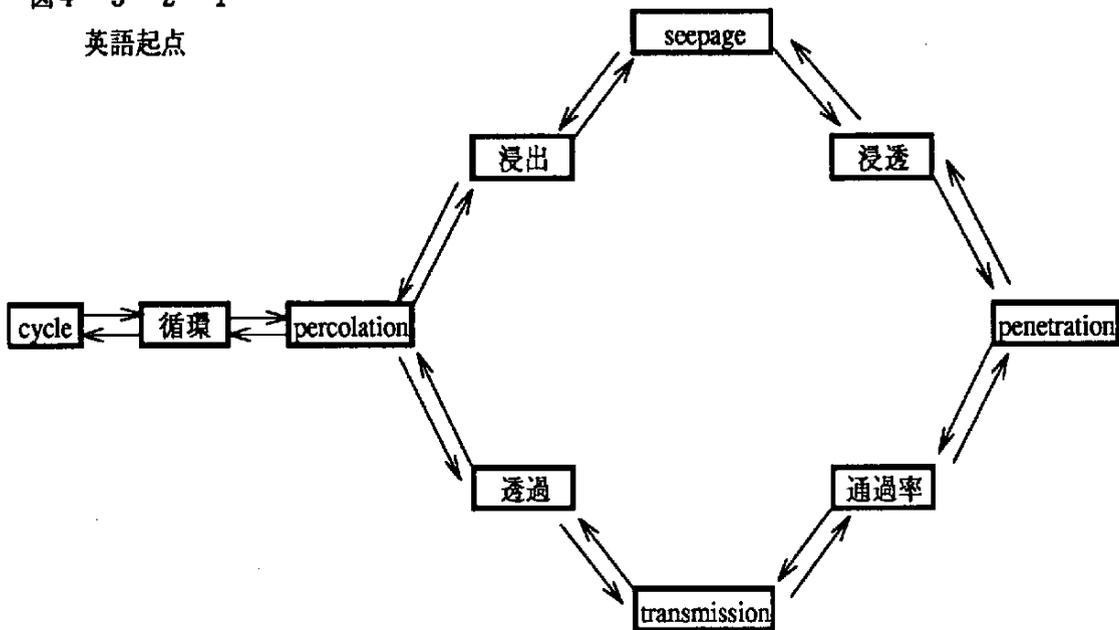


図4-3-2-1

英語起点



2) 用語のクラスター化

ここで、訳語連鎖の関係の要素をモデル化してみると、次のようなものがある(あ、い、う等は日本語、A、B、C等は英語。逆でも可)。

1) 往復：あ \leftrightarrow A

→A

2) 分岐：あ →B

→C

3) 回帰：あ→A→い→B→う→C→あ

連鎖検索の方法による分析を企画するに際しての基本的アイデアは、訳語を連鎖的に検索してゆくことによって、関連用語のクラスターが形成されるであろうというものであった。最も単純な群は、上記の1)の関係だけからなるもの、すなわち、英和、和英ともに訳し分けがなく、1英用語と1和用語が対応しているだけという場合である。これに属する用語が語数的に多いことは、図表4-1-3~6からも推定されるのであるが、その大部分は使用頻度の低い用語のようである。2)は、訳し分けが行われる部分のことである。これによって分岐したルートは、その先で1)の形の終端に到達するか、また、3)のような、複数の用語が介在する回路を経て、もとの用語に回帰する。これによって連鎖は打ち切られ、用語のクラスターが形成されることになる。

(3) 「発散語」と「用語星図」

ところで、先の事例にも現れたとおり、連鎖の途中で、多数の訳語と多数の適用分野をもつ用語に逢着すると、連鎖のルートは一気に展開する。シソーラス分析の分野でexplosion termなる概念が用いられることがあるが、訳語の連鎖検索においても、これに似た役割をはたす用語があることがわかる。これを仮に「発散語」と称することにする。

発散語は英用語、和用語ともに存在する。前掲の「宇宙」の例では、「宇宙」の訳語のひとつにspaceがあり、これが発散語として機能することによって、多くの分野の専門用語が取り込まれるようになっている。用語クラスターの「星図」は、このような発散語を中心部分に据えて、描かれることになるであろう。

新出の専門用語を辞書に登録するにあたっては、これと既登録語との関連について、綿密な調査が必要である。ここにのべた、両引き辞書による訳語の連鎖検索と、これの星図的表示は、専門用語辞書の更新管理システムの重要な要素になると思われる。この点、後に再説することにする。

5. 和文専門用語の抽出方式

5. 1 実験用電子化文献の概要

今回の調査に当たって、和文の電子化された文献を収集し、これから和文の専門用語を抽出する実験を行った。実験対象文献には、1) 共同通信社の配信記事と、2) 情報技術関係の学会論文等の要約の2種について行った。それぞれの事例を図5-1-1-1~2に示す。

5. 2 簡易分かち書きの方法と結果

和文からの用語抽出にあたっては、まず、原文を分かち書きして、用語の候補としての文字列を切り出すことが必要である。この場合、現今では、辞書を用いる方式が一般であろう。しかし、本研究のように、用語の検出方式自体を問題にするような場合には、初めから辞書を前提とするような方法は、循環論法的な誤りを犯すことになる。むしろ、なるべく多くの用語候補が見いだされるような方法が望ましい。

昨年度の調査研究において、英文複合語（複数単語）の専門用語の抽出方法を検討した際、英単語を日本語の各1文字と読み替えて処理すれば、ほぼ同様の考え方で、和文の分かち書き、すなわち用語候補の抽出ができるであろうとの予想が得られている。一方、国立国語研究所・中野洋言語体系研究部長のもとでは、辞書を用いずに、漢字個々の音訓ヨミの表と、若干の文法的接尾辞の表を用いることで、和文を分かち書きした上で、ヨミをふるというシステムを、多年にわたり開発し、語彙調査に実用している。このシステムは、基本的に字種の変化点（例えば、漢字からひらがなへの移行など）に着目し、一定の判定基準のもとに、そこに語句の切れ目を挿入する（分かち書きする）というものである。

こうした事例を参考に、本調査研究でも、字種変化に着目した語句の切り出しアルゴリズムを開発し、その結果得られた語句を用語の候補として、既述の両引き辞書と突き合わせて、用語の認定および頻度統計を行うことにした。この分かち書き処理の結果例を図5-2-1-1~2に示す。

図5-1-1-1 実験用電子化文献の例 (一般記事)

しかし硬直した大学のアンチテーゼとしてこうした遊びの達人の講義を持ってくるといった二項対立的な考え方は少し古いのではないだろうか。硬直したアカデミズムを軟派の論理で突き崩すといえは聞こえはいいが結局それはお祭り騒ぎにすぎず、最後はしょせん遊びですから式の開き直りで終わる心配もある。講義の中身は面白いんだからいそのこと軟派風の売り方はやめて、これこそが真の教養だというような大上段の構えがあてでもいいんじゃないだろうか。

産経新聞社編「大学を問う」(新潮社・一二〇〇円)には日本の大学の現状がカタログのように網羅されている副題には荒廃する現場からの報告とあり大学=荒廃というマスコミの固定観念の強さにちとあきれてしまうが仮に大学の現状を荒廃と呼ぶにしてもその中身は一九七〇年代的な荒廃とはずいぶん違っているものだと考えさせられる。例えばかつてあれほど騒がれた産学合同の研究体制などは今や問題にすらなり得ないようだそうした中から教養課程の廃止といった声も起こってくるのだろう。

図5-1-1-2 実験用電子化文献の例 (情報分野論文要旨)

この論文は、従来のクロック・システムよりもかなり有利な点を提供でき、通常非同期法に関連した欠点のない、高性能非同期アーキテクチャへの新しいアプローチについて述べている。集積化のレベルが増大すると、この技法を用いて設計した非同期波面アレイウエアフロント)は、これと等価な同期型シストリック・アレイにまさる3つの重要な利点を有する、すなわちそれはより速いスループット(全システムを通してデータがクロックされる速度)、より少ない設計の複雑さ及びより高い信頼性である。後者の利点は良く知られているが、新しいアプローチの結果である最初の2つは、自己タイミング・システムを、高性能応用に関して更に魅力的なものとする。我々の提案した非同期アーキテクチャのコストはシリコンの面積であり、それは高々、等価な同期系設計の2倍であろうと思われる。我々の非同期手法を用いる利益と欠点は三波面アレイ: 2つの一次元マルチプライヤと1つの2次元ソータを用いて強調されている。この3つは全て1つの基本的なビルディング・ブロックを用いて作ることができ、また2ミクロンCMOSを用いた時のシュミレーションは、それらが静的な論理インプリメンテーションに対して250メガヘルツ及び動的な論理インプリメンテーションに対しては4メガヘルツ以上のスループットが可能であることを示している。

図5-2-1-1 分ち書きの処理の例 (一般記事)

しかし硬直した大学のアナチテーゼとしてこうした遊びの達人の講義を
 持ってくるといった二項対立的な考え方は少し古いのではないだろうか。
 硬直したアカデミズムを軟派の論理で突き崩すといえば聞こえはいいが
 結局それはお祭り騒ぎにすぎず、最後はしょせん遊びですから式の開
 き直りで終わる心配もある。講義の中身は面白いんだからいそのこと軟
 派風の売り方はやめて、これこそが真の教養だというような大上段の構
 えがあってもいいんじゃないだろうか。

産経新聞社編「大学を問う」(新潮社・一〇〇〇円)には日本の大学の
 現状がカタログのように網羅されている副題には荒廃する現場からの報
 告とあり大学＝荒廃というマスコミの固定観念の強さにちとあきれてしまうが
 仮に大学の現状を荒廃と呼ぶにしてもその中身は一九七〇年代的な荒
 廃とはずいぶん違っているものと考へさせられる。例えばかつてあれほど騒
 がれた産学合同の研究体制などは今や問題にすらなり得ないようだそうした中
 から教養課程の廃止といった声も起こってくるのだろう。

図5-2-1-2 分ち書きの処理の例 (情報分野論文要旨)

この論文は、従来クロック・システムよりもかなり有利な点を提供
 でき、通常非同期法に関連した欠点のない、高性能非同期アーキテク
 チャへの新しいアプローチについて述べている。集積化のレベルが増大す
 ると、この技法を用いて設計した非同期波面アレイウエアフロントは、
 これと等価な同期型シストリック・アレイにまさる3つの重要な利点
 を有する、すなわちそれはより速いスループット(全システムを通して
 データがクロックされる速度)、より少ない設計の複雑さ及びより
 高い信頼性である。後者の利点は良く知られているが、新しいアプロ
 ーチの結果である最初の2つは、自己タイミング・システムを、高性能
 応用に関して更に魅力的なものとする。我々の提案した非同期アーキ
 テクチャのコストはシリコンの面積であり、それは高々、等価な同期系設
 計の2倍であろうと思われる。我々の非同期手法を用いる利益と欠
 点は三波面アレイ：2つの一次元マルチプライヤと1つの2次元ソータ
 を用いて強調されている。この3つは全て1つの基本的なビルデ
 イング・ブロックを用いて作ることができ、また2ミクロンCMOSを用
 いた時のシュミレーションは、それらが静的な論理インプリメンテーション
 に対して250メガヘルツ及び動的な論理インプリメンテーションに対し
 ては4メガヘルツ以上のスループットが可能であることを示している。

5. 3 両引き辞書との突き合わせ処理

前項で分かち書きされた各用語候補について、両引き辞書の和→英検索機能により、突き合わせを行う。このとき、単純な一致検索のみでは、有効な結果が得られないので、用語候補の文字列を分割して、すべての部分文字列を切り出し、それぞれを両引き辞書と照合した。例えば、「不等辺山形鋼」が、分かち書き処理の結果の用語候補として得られたとき、これから下記の21種の文字列を切り出し、これらについて辞書との照合を行う。

〈不〉 〈等〉 〈辺〉 〈山〉 〈形〉 〈鋼〉
〈不等〉 〈等辺〉 〈辺山〉 〈山形〉 〈形鋼〉
〈不等辺〉 〈等辺山〉 〈辺山形〉 〈山形鋼〉
〈不等辺山〉 〈等辺山形〉 〈辺山形鋼〉
〈不等辺山形〉 〈等辺山形鋼〉
〈不等辺山形鋼〉

こうした部分文字列照合方式により、原文中で「用語」の可能性のある文字のつらなりは、両引き辞書と照合され、辞書への登録の有無が点検され、また次記の頻度統計が行われる。

5. 4 未登録語の検出と検定

上記の処理の結果、辞書未登録の用語候補が出力される。これを出現頻度順にならべてみると、部分文字列の照合から派生する、一文字の漢字やカナなどが先頭部分にくるほか、専門辞書には入れられない、いわば本来の一般用語が現れる。そこで、文字数の制約と機械化常用語辞書との照合により、これらを除去すると、新出専門用語の第一次候補のリストが得られる。これを専門家がざっと一覽して、専門用語でありそうなものをだけを選んでみた例が、表5-4-1である（情報分野の論文要旨データを対象にしたもの）。

ところで、これらのうちのどれが真に専門的な用語であり、その英訳として、独特の英語専門用語を当てる必要のあるような種類の用語であるかという視点で、選別作業を進めるためには、それなりの参照資料が必要である。すなわち、その第一は、当該用語が現れた原文自体であり、またその第二として、前章の訳語連鎖分析にみたような、辞書中での関連用語を網羅的リストがあげられる。前者は具体的な用例を点検するものであるし、後者は、辞書に既登録の用語と、この未登録語との関係を点検するためのものである。これら資料のいずれについても、その調製のためのコンピュータ処理システムを用意することで、見易い形式のものをオンライン的に表示させるなど、作業の効率化を図ることができるであろう。ここに、辞書更新のためのデータベース・システムの構築が示唆される。そこで、この点について、7章に稿を改めて検討することにする。

表5-4-1 辞書未登録語の専門用語候補の例 (情報分野要旨) (1)

抽出用語	頻度	抽出用語	頻度	抽出用語	頻度
ファジィ	92	分光計算	16	認識対象	12
楕円曲線	50	物体表面	16	定量的	12
ハイパーテキスト	49	設計対象	16	段階的詳細化	12
再帰的下向	35	順序保存符号	16	対話型	12
連想形記憶	33	言語仕様	16	多数桁整数	12
属性文法	32	拡散反射成分	16	生体信号	12
並列実行	31	演算子順位	16	正当性	12
平均符号長	30	ルックアップテーブル	16	垂直方向	12
画像情報	28	非終端記号	15	次元物体	12
推論規則	26	定理証明器	15	協調動作	12
言語処理系	26	知識処理	15	逆向	12
文字枠図形	25	制約伝播	15	起因	12
汎用	25	制約充足機能	15	割合	12
次元の	25	推論法	15	解候補	12
輝度計算	25	色性反射	15	連続音声	11
一方向性関数	25	詳細化	15	並列度	11
並列性	24	計算量	15	排他制御	11
分散型	24	形記述子	15	端点値探索法	11
交点計算	24	帰納推論	15	端点区間	11
原画像	24	基本制約	15	端点解	11
マルチメディアデータベース	24	マルチコンテキスト	15	制限効果	11
離散対数問題	22	ポインタ	15	受動化条件	11
黒画素密度	22	パワースベクトル	15	写像関係	11
検索時間	22	ディレクトリ	15	実行可能	11
不均質物体	21	文章理解	14	自由変数	11
標準漢字	21	秘密鍵	14	左辺	11
学術情報	21	聴覚心理実験	14	黒画素	11
物体探索	20	知識表現言語	14	現象	11
符号語	20	大型計算機	14	研究実施計画	11
ユーザインタフェース	20	稼働	14	近似的	11
筆点運動波形	19	メタオブジェクト	14	開始点間	11
対応関係	19	マクロバイプライン	14	解析的	11
生成規則	19	様相論理	13	音素認識	11
情報較差	19	中間解	13	意味表現	11
自動生成	19	逐次的	13	意味構造	11
仕様記述	19	大容量	13	メモリシステム	11
候補漢字数	19	多言語	13	マルチメディア	11
レイトレーシング	19	線図形	13	ポロノイ	11
認識実験	18	診断対象	13	データモデル	11
二階層	18	結合型	13	ストリーム	11

表5-4-1 辞書見登録語の専門用語候補の例 (情報分野要旨) (2)

抽出用語	頻度	抽出用語	頻度	抽出用語	頻度
必要十分条件	10	署名方式	9	擬似乱数生成器	8
反射成分	10	処理結果	9	輝度値	8
特徴量	10	熟練者	9	顔画像	8
代数的仕様	10	次元表示	9	漢字認識	8
対処	10	次元空間	9	拡散相互反射	8
素因数分解	10	言語表現	9	画素単位	8
生起確率	10	空間分割数	9	メタプログラミング	8
実現方法	10	鏡面反射成分	9	ボトルネック	8
実現方式	10	幾何学的	9	プロトタイピング	8
自動形成	10	階層間	9	ファジイデータ	8
自己反映計算	10	レンダリング	9	トランスレータ	8
仕様記述言語	10	ヒューマンインタフェース	9	キャッシュミス	8
算術符号	10	アイコンックプログラミング	9	連続領域	7
高速処理	10	論理体系	8	利用形態	7
公開鍵暗号	10	連続音声認識	8	明確化	7
交通分野	10	類推理論	8	文法的	7
故障診断	10	立位姿勢時	8	復号操作	7
研究代表者	10	立位姿勢	8	復号時間	7
検索機能	10	変換行列	8	認識法	7
技法	10	平均処理時間	8	認識漢字	7
幾何	10	文字組	8	日本語文章	7
関数型	10	文字情報	8	通信計算量	7
画像検索	10	文字候補	8	逐次実行用	7
音声材料	10	文解析	8	多項式時間	7
英日機械翻訳	10	分光電力分布	8	即時表示	7
位置関係	10	表面反射率	8	臓器形状	7
暗号方式	10	筆記条件	8	数値処理	7
スーパーコンピュータ	10	熱視線	8	実行速度	7
シェイピング	10	認識精度	8	質問言語	7
並列協調問題解決	9	入力文字列	8	視覚的階層的質問言語	7
復号化	9	統語構造	8	冊子体	7
部分問題	9	抽出成功率	8	座標値	7
符号語長	9	知識工学的手法	8	高次	7
不可分命令	9	組文脈自由文法	8	高階項	7
非単調論理	9	進木構造	8	語義	7
非接触式	9	証明系	8	原画像中	7
反対語	9	昇順	8	形式的体系	7
波画像	9	充足不能	8	局所性	7
日常手書	9	辞書項目	8	協調問題解決	7
迅速	9	客観的輪郭	8	話者適応	6

表5-4-1 辞書見登録語の専門用語候補の例 (情報分野要旨) (3)

抽出用語	頻度	抽出用語	頻度	抽出用語	頻度
例外処理	6	構文解析手続	6	認識処理	5
理論的性質	6	固有物体色	6	認識手法	5
有限体上	6	研究発表欄	6	日本語訳語	5
優先度	6	研究実績	6	日本科学技術情報	5
補正用	6	検索能力	6	同時提示	5
便宜的	6	計算機実験	6	当初計画	5
変形性	6	極小点集合	6	展開係数	5
並列型	6	教育漢字	6	低域部分	5
並列化可能	6	急激	6	知的分散処理	5
並行性	6	逆対応	6	値論理回路	5
文字候補数	6	技術的	6	談話構造	5
分探索	6	輝度上	6	単語音声	5
部分問題間	6	記憶量	6	代数的手法	5
不等式系	6	機械辞書	6	対応操作文	5
評価実験	6	基礎理論	6	体表面電位図	5
表示可能	6	簡約可能	6	多値関数	5
標準辞書	6	感覚障害	6	送信遅延	5
筆点座標	6	格納用	6	総合的数式処理	5
筆者識別法	6	拡散成分	6	相互反射色	5
範囲指定形式	6	回路実現言語	6	相互拘束	5
発色特性	6	画像検索用端末	6	設計開発	5
配列構造	6	下部組織	6	声道断面形状	5
入力音声	6	音素系列	6	制約付変数	5
日本語文章推敲支援	6	一方向性置換	6	制御結果	5
特徴抽出法	6	データベースワークベンチ	6	推論方式	5
導出順序	6	センタアルゴリズム	6	推論機構	5
点字機械辞書	6	列挙	5	新聞記事	5
転送経路	6	類型	5	条件分岐	5
典拠	6	累積距離	5	処理部分	5
直接実装	6	立体保持過程	5	出力過程	5
端点値集合	6	立体知覚後	5	受動態	5
多数桁整数乗	6	理論的成果	5	実装方法	5
選択点	6	要求仕様記述	5	実行効率	5
節集合	6	方式設計	5	視点制御	5
設計解	6	母音認識	5	視点主導型知識	5
接頭辞	6	弁別索性	5	雑誌論文	5
性能比	6	変数宣言	5	最適化問題	5
図学教育	6	変換機能	5	最小生成木	5
人工内耳	6	並列処理言語	5	最終目的	5
参照関係解析	6	認識知識	5	形態素解析法	5

6. 和文専門用語の頻度分析と分野分類

6. 1 分野分類の設定

本調査研究では、専門用語辞書の整備方式を検討課題としているが、そこでは、専門分野分類の設定自体も大きな課題になる。当面は、科学技術分野を対象に考えて、これをどの程度に細分化して、両引き辞書における用語の適用分野区分や、実験用原文献の対象分野区分に用いるのが適当であるかについて検討した。

この検討には、学術用語辞書、J I S用語辞書、マグローヒル科学技術用語辞書における分野区分が、一応の参考になる。もっとも、本調査での分野分類の目的は、学問分野に関する一般的な分類を設定しようというわけではなく、用語の訳し分けと密接に関連した分野分類の設定が必要である。すなわち、この場合、語彙の分化、独自性というのが分野認定の基準になるはずである。もっとも、こうした観点からの検討に即応するような、計量的データはこれまで存在せず、むしろ、本調査研究の延長上において、集積されうるものである。従って、当面は、上記の既成辞書での分野区分に加えて、独自の調査を実施し、これらを専門家の委員会で総合的に検討してもらって、本調査のためにもちいるべき分野分類を確定した。

すなわち、分野分類のための調査として、わが国の大学における学科・講座の設置状況と、日本学術会議の研究連絡委員会構成を調査した。また、日本工学会の協力を仰いで、傘下学会を中心にアンケート調査を実施し、各学会が自己の担当専門分野をどのように考えているかの意見を徴した(表6-1-1)。これは、学会誌を翻訳対象文献とする場合に、その掲載論文にどの分野分類を与えるべきかという観点でも応用できるものである。これら調査を総合的に検討した結果、すでに示したような21分野の設定に至った。

さらに、両引き辞書の編成にあたっては、学術用語辞書等、既成の3辞書それぞれにおける分野区分を、今回の分類に統合する必要がある。このために定義された、各辞書の原分類と今回の対応関係は、<付属資料1>に示すとおりである。また、わが国の大学における学科・講座設置状況調査の結果を<付属資料2>に示す。

表6-1-1 分野分類アンケート調査結果(1)

分野	主専門分野となる学会	関係分野となる学会
機械	日本機械学会, 自動車技術学会, 精密工学会, 日本冷凍協会, 日本塑性加工学会, 日本潤滑学会, 日本船用機関学会, 可視化情報学会, 日本混相流学会, 粉体粉末冶金協会, 日本伝熱学会	空気調和・衛生工学会, プリント回路学会, 日本航空宇宙学会, 日本鋳物協会, 軽金属学会, 山陽技術振興会, 日本原子力協会, 表面技術協会, 日本シミュレーション学会, 計測自動制御学会
建築	日本建築学会, 空気調和・衛生工学会, 日本コンクリート工学協会	土木学会, 土質工学会, 日本冷凍協会, 照明学会, 骨材資源工学会, 腐食防食協会
応用物理	応用物理学会, 日本応用磁気学会	日本物理学会, 日本分析化学会, 照明学会, 日本混相流学会, 計測自動制御学会, 日本セラミックス協会, 日本航空宇宙学会, 日本原子力協会, 表面技術協会, 日本写真学会
化学	日本化学会, 高分子学会, 日本分析化学会, 化学工学協会, 日本セラミックス協会, 石油学会, 日本ゴム協会, 電気化学協会, 繊維学会, 触媒学会, 腐食防食協会, 日本エネルギー学会	日本コンクリート工学協会, 表面技術協会, 日本潤滑学会, 資源・素材学会, 資源処理技術研究会(浮選研究会), 粉体粉末冶金協会, プリント回路学会, 日本鋳物協会, 日本写真学会
航空宇宙	日本航空技術協会, 日本航空宇宙学会	日本機械学会
材料資源	日本鉄鋼協会, 日本金属学会, 溶接学会, 日本溶接協会, 日本鋳物協会, 表面技術協会, 資源・素材学会, 軽金属学会, 日本写真学会, 骨材資源工学会, 資源処理技術研究会(浮選研究会)	日本セラミックス協会, 日本塑性加工学会, 日本ゴム協会, 電気化学協会, 腐食防食協会, 日本エネルギー学会, 日本潤滑学会, 日本建築学会, 日本原子力協会, 日本混相流学会, 山陽技術振興会

表6-1-1 分野分類アンケート調査結果(2)

分野	主専門分野となる学会	関係分野となる学会
船舶海洋	日本造船学会	日本船用機関学会
情報	情報処理学会, 日本シミュレーション学会, 山陽技術振興会	電子情報通信学会, 電気学会, 計測自動制御学会, 精密工学会, テレビジョン学会, 電気通信協会, 画像電子学会, 日本写真学会, プリント回路学会
電気電子通信	電子情報通信学会, 電気学会, プリント回路学会, 日本鉄道電気技術協会, 計測自動制御学会, テレビジョン学会, 照明学会, 電気通信協会, 電気設備学会, 画像電子学会	自動車技術学会, 日本応用磁気学会, 日本シミュレーション学会, 応用物理学会, 山陽技術振興会, 粉体粉末冶金協会
土木	土木学会, 土質工学会, 日本コンクリート工学協会	日本建築学会, 骨材資源工学会, 日本ゴム協会, 日本混相流学会
物理	日本物理学会	応用物理学会, 日本原子力協会, 日本応用磁気学会, プリント回路学会
地球科学		日本分析化学会
都市工学		日本水環境学会, 土木学会, 日本建築学会
農業		日本冷凍協会
生物		腐食防食協会
数学		日本物理学会, 日本シミュレーション学会
林業		
水産		日本冷凍協会
医工学		日本機械学会, 腐食防食協会

6. 2 和文専門用語の出現頻度統計の方式

和文専門用語の出現頻度の統計は、「専門用語の抽出方式」において既述のとおりである。すなわち、通信社配信の一般記事と情報関係の論文抄録について、分かち書き処理、部分文字列への分解、両引き辞書との照合という過程を通じて、出現頻度を統計した。辞書との照合の際、分かち書きされた用語全体と一致する用語が辞書に登録されていた場合（完全一致カウント）と、部分文字列が辞書に発見された場合（部分一致カウント）とを別々に集計している。たとえば、「不等辺山形鋼」の場合、《不等辺山形鋼》は完全一致として集計され、また、《鋼》、《形鋼》、《山形鋼》、《等辺山形》、《等辺山形鋼》等々のうちで、辞書中に発見される用語は部分一致として集計される。この統計結果をもとに、完全一致カウントと部分一致カウントとの関係に着目して整理した集計表を、後掲のとおり作成し、検討した。

6. 3 統計結果

(1) 高頻度用語

表6-3-1~2は、一般記事と情報関係論文抄録における、完全一致、部分一致カウントを通じた頻度の高い用語のリストである。当然ながら、接頭辞、接尾辞的用語が、大きな部分一致カウントを獲得して、多くリストされることになっている。一方、完全一致カウントの大きいものは、日、行などである。これらは、大部分、助詞などに囲まれて文章中に独立して現れることが多い日常語である。しかし、専門用語辞書にも登録されていることから、分野によって専門的な用法のあるものである。機械翻訳にあたっては、これらの、日常的用法か専門的用法かの判別が、最も判定の難しいところと思われる。

表 6 - 3 - 1 高頻度の日本語用語 (一般記事) (1)

E = 完全一致, P = 部分一致 (二重下線部による照合)

頻度	日本語	照合	英語	分野
5296	行	E	horizontal line	情報
3896	日	E	day	応用物理, 航空宇宙, 電気電子通信, 都市工学
3270	時	E	hour	電気電子通信
1957	輪	E	areola	動物
1689	差	E	inequality	その他, 応用物理, 航空宇宙
1654	出	E	rise	その他
1366	西	E	west	応用物理, 航空宇宙, 都市工学
1353	日本	E	Japan	その他
1262	上	E	raise	その他
1201	大	E	greater	数学
1189	人	E	person	その他
1151	分	E	arcmin	数学
1100	入	E	on	電気電子通信, 物理
1027	対	E	pair	応用物理, 機械, 電気電子通信
1019	運動	E	manoeuvre	航空宇宙
1016	<u>日</u>	P	day	応用物理, 航空宇宙, 電気電子通信, 都市工学
902	<u>編注</u>	P	remark	情報
902	<u>編注</u>	P	note	情報
902	<u>編注</u>	P	computer program annotation	情報
902	<u>編注</u>	P	comment	情報
896	初日	E	first date	地球科学, 農業
883	<u>タイトル</u>	P	Torr	応用物理, 機械, 電気電子通信
883	<u>タイトル</u>	P	tight	機械
883	<u>タイトル</u>	P	tie	機械, 建築, 土木
882	表	E	table	応用物理, 航空宇宙, 情報, 数学, 都市工学
874	東	E	east	地球科学
856	開	E	off	電気電子通信
792	リグ	E	rig	機械
744	<u>野球</u>	P	spherulite	地球科学
744	<u>野球</u>	P	sphere	応用物理, 機械, 数学

表 6 - 3 - 1 高頻度の日本語用語（一般記事）（2）

E = 完全一致, P = 部分一致（二重下線部による照合）

頻度	日本語	照合	英語	分野
722	口	E	ass	農業
679	合	E	conjunction	応用物理, 航空宇宙
629	カ	E	mosquito	動物
619	高	E	High	その他
599	年	E	annual volume	その他
590	野	E	field	数学
571	時半	P	semi-	数学
569	以上	E	greater or equal	数学
546	雨	E	rain	応用物理, 建築, 地球科学, 都市工学, 土木
533	次	E	order	応用物理, 航空宇宙, 都市工学
517	点	E	lambda point	機械
502	切	E	off	電気電子通信
480	前節	E	propodite	動物
478	目	E	order	応用物理, 生物, 船舶海洋, 都市工学, 動物, 農業
475	月	E	month	応用物理, 航空宇宙, 都市工学
473	写真	E	photograph	化学
453	秒	E	s	機械
449	ロシア	P	sheer	地球科学
448	問題	E	problem	応用物理, 情報, 電気電子通信
445	北東	P	east	地球科学
443	曇	E	overcast	航空宇宙, 地球科学
430	天気	E	weather	航空宇宙, 地球科学
425	下	E	Lower	その他
414	度	E	frequency	数学
413	後	E	back	応用物理, 電気電子通信
401	頭	E	head	応用物理, 都市工学, 動物, 農業
390	米	E	rice	農業
390	ラ	E	mule	農業
367	竹下派	P	bundle	数学
367	号	E	issue	その他

表 6 - 3 - 1 高頻度の日本語用語（一般記事）（3）

E = 完全一致, P = 部分一致 (二重下線部による照合)

頻度	日本語	照合	英語	分野
345	必要	E	necessary	数学
345	ドイツ	E	Germany	電気電子通信
341	関連	E	association	応用物理, 数学, 生物, 都市工学, 動物, 農業
334	増	E	Increase	応用物理, 情報
333	方	E	shift	応用物理, 化学, 材料資源, 都市工学
333	語	E	word	応用物理, 情報, 電気電子通信, 都市工学
326	前	E	Forward	情報
319	テレビ	E	television radar air navigation	航空宇宙
318	今回	P	turn	電気電子通信
318	今回	P	gyrus	動物
318	位	E	place	数学
311	雪	E	snow	応用物理, 建築, 地球科学, 土木
309	中心	E	center	機械, 数学, 電気電子通信
308	試合	P	conjunction	応用物理, 航空宇宙
300	回	E	gyrus	動物
295	確率	P	factor	電気電子通信, 物理
295	確率	E	probability	応用物理, 機械, 建築, 航空宇宙, 情報, 数学,
294	確率予報	P	prediction	応用物理, 航空宇宙, 都市工学
294	確率予報	P	forecast	航空宇宙, 地球科学
294	確率予報	P	FCST	航空宇宙
294	確率予報	P	probability	応用物理, 機械, 建築, 航空宇宙, 情報, 数学,
294	確率予報	E	probability forecast	地球科学
294	降水	E	precipitation	化学, 建築, 船舶海洋, 地球科学, 土木, 農業
291	父	E	sire	農業
288	今後	P	back	応用物理, 電気電子通信
288	今後	P	Backward	情報
287	手	E	hand	機械, 電気電子通信, 動物
287	金	E	gold	応用物理, 化学, 材料資源, 都市工学
284	特賞	P	reward	その他
283	馬	E	horse	建築

表6-3-2 高頻度の日本語用語（情報分野論文要旨）（1）

E=完全一致, P=部分一致（二重下線部による照合）

頻度	日本語	照合	英語	分野
1349	行	E	horizontal line	情報
1172	システム	E	system	応用物理,機械,情報,電気電子通信
728	データ	E	data	応用物理,機械,情報,数学,地球科学,電気電子通信
668	場合	P	conjunction	応用物理,航空宇宙
551	対	E	pair	応用物理,機械,電気電子通信
540	開発	E	development	その他
523	上	E	raise	その他
500	研究	E	investigation	電気電子通信
483	モデル	E	model	応用物理,航空宇宙,情報,電気電子通信,都市工学
448	方法	E	method	応用物理,機械,情報,電気電子通信,土木
433	プログラム	E	program	機械,航空宇宙,情報,地球科学,電気電子通信,土木
415	必要	E	necessary	数学
410	図	E	MAP	応用物理,航空宇宙,都市工学
403	次	E	order	応用物理,航空宇宙,都市工学
392	基	E	group	化学
386	結果	E	fruiting	農業
348	処理	E	information processing	情報
314	アルゴリズム	E	algorithm	応用物理,情報,数学,電気電子通信,都市工学
299	利用	E	usage	その他
292	問題	E	problem	応用物理,情報,電気電子通信
290	データベース	E	data base	情報,電気電子通信
286	設計	E	design	応用物理,機械,建築,情報,電気電子通信,土木
284	値	E	value	応用物理,情報,都市工学
275	本研究	P	research	応用物理,情報
275	本研究	P	investigation	電気電子通信
275	本研究	P	examination	電気電子通信
269	可能	E	possible	応用物理,情報
266	生成	E	generate	電気電子通信
265	大	E	greater	数学
255	記述	E	describe	機械

表 6-3-2 高頻度の日本語用語（情報分野論文要旨）（2）

E = 完全一致, P = 部分一致 (二重下線部による照合)

頻度	日本語	照合	英語	分野
248	方式	E	manner	電気電子通信
235	構成	E	configuration	情報, 電気電子通信
233	表	E	table	応用物理, 航空宇宙, 情報, 数学, 都市工学
231	以下	E	less or equal	数学
229	出	E	rise	その他
229	言語	E	language	情報, 電気電子通信
225	解析	E	analysis	応用物理, 機械, 数学, 地球科学, 都市工学, 土木
224	定義	E	define	情報
221	評価	E	evaluation	応用物理, 情報
220	対応	E	correspondence	数学
218	文字	E	letter	電気電子通信
218	表現	E	representation	応用物理, 情報, 数学, 都市工学
212	集合	E	assemblage	船舶海洋
210	計算	E	calculation	応用物理, 電気電子通信
205	プロセッサ	E	language processor	情報
203	ソフトウェア	E	software	機械, 情報, 電気電子通信
202	情報	E	information	応用物理, 航空宇宙, 情報, 地球科学, 電気電子通信,
199	合	E	conjunction	応用物理, 航空宇宙
199	呼	E	call	電気電子通信
197	関数	E	function	応用物理, 航空宇宙, 情報, 数学, 電気電子通信, 都市工学
192	法	E	modulo	数学
186	機能	E	function	電気電子通信, 農業
184	数	E	number	応用物理, 情報, 数学
183	組	E	set	その他
183	漢字	E	kanji	情報
181	構造	E	gamma structure	応用物理
181	メモリ	E	memory	電気電子通信
177	操作	E	manipulation	電気電子通信
173	高	E	High	その他
173	関係	E	relation	情報, 数学

表6-3-2 高頻度の日本語用語（情報分野論文要旨）（3）

E = 完全一致, P = 部分一致（二重下線部による照合）

頻度	日本語	照合	英語	分野
170	画像	E	image	電気電子通信
169	構築	E	construction	情報
169	ベクトル	E	becquerel	電気電子通信
166	抽出	E	percolation leaching	材料資源
165	知識	E	knowledge	電気電子通信
164	目的	E	purpose	機械, 情報
164	対象	E	object	その他
163	符号	E	sign	情報, 数学, 地球科学
162	手法	P	slope	農業
162	手法	P	modulus	数学
162	手法	P	modulo	数学
162	手法	P	divisor	数学
159	特徴	E	characteristic feature	電気電子通信
154	検索	E	retrieval	応用物理, 情報
153	点	E	lambda point	機械
151	認識	E	recognition	応用物理, 情報, 電気電子通信
148	形	E	shape	応用物理, 機械
147	拡張	E	dilation	その他
147	応用	E	application	応用物理, 情報, 電気電子通信
145	式	E	mood of syllogism	その他
144	種類	E	kind	電気電子通信
140	実験	E	experiment	応用物理, 電気電子通信
137	確認	E	identification	応用化学, 機械, 船舶海洋, 農業
135	実行	E	execution	情報, 電気電子通信
132	存在	E	being	応用物理
130	導入	E	installation	応用物理, 都市工学
129	並列	E	parallel in	機械
128	変換	E	nuclear transformation	物理
124	意味	E	meaning	電気電子通信
124	パターン	E	antenna radiation pattern	電気電子通信

(2) 高頻度用語 (完全一致・部分一致合計値による)

表6-3-3~4は、完全一致カウントと部分一致カウントの合計値を用いて、高頻度の用語をリストしたものである。ここでも、接頭・接尾辞類と日常語が混在して多くリストされる結果となっており、専門用語らしいものは少ない。一般記事を対象にした場合、これはやむを得ないものと思われる。もっとも、情報分野の論文抄録に関する場合は、この表でも、専門用語が意外に多く含まれることが注目されよう。

ここで、同一用語に関する完全一致と部分一致カウントの関係に着目して、下記のような係数を計算してみた。

$$\text{完全・部分一致係数} = \frac{\text{完全一致カウント}}{\text{(完全一致カウント+部分一致カウント)}}$$

これは、字種変化を基準とする分かち書き処理において、独立して切り出されることの多い用語、少ない用語というような判別の指標になる。すなわち、この係数が1に近いものは、独立して切り出される用語、0のものは接続して抽出される用語ということになる。

(3) 完全・部分一致係数による用語表

前記の完全・部分一致係数に注目して、これの高(1)、中(0.5)、低(0)の領域にどのような用語が現れるかをみた。すなわち、表6-3-5~7が、一般記事における高、中、低係数の用語例であり、表6-3-8~10が情報分野の論文抄録におけるそれである。

論文抄録における高係数の用語リストには、情報分野の専門用語が効率よくリストされる結果になっている(表6-3-8)。一般記事の場合でも、専門的用語が多くリストされる結果になっており、これらは、おおむね専門辞書によって、そのまま翻訳してよいもののものである(表6-3-5)。もっとも、バブル等は大抵「バブル経済」からきたものであるので、経済専門用語ということにはなるが、ここで用いた両引き辞書は科学技術用語からなっているから、これによる翻訳は、原理的には正しく行われなければならないはずである。分野別の辞書編成の重要性が看取されるであろう。

中係数の用語表(表6-3-6、表6-3-9)には、それ自体一定のいみをもつ専門用語であるが、同時に複合的専門用語の造語要素にもなるような用語が現れる。このあたりが、翻訳にあたって最も難しい用語群であると思われる。

低係数の用語表(表6-3-7、表6-3-10)には、接頭・接尾辞的造語要素が現れる。このとき、短いカタカナの高頻度語は、大部分もっと長いカタカナ用語の切れ端であって、有意なカウントではない。特定の専門分野の用語として、このような短いカタカナ語が登録されているために、こうした結果になっているわけである。ここに、カタカナの接続については、部分一致による辞書照合は無効であることが示唆される。

表6-3-3

完全一致・部分一致による高頻度の日本語用語（一般記事）

(1)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
10778	日	3896	6882	0.36
7261	行	5296	1965	0.73
6674	表	882	5792	0.13
4853	時	3270	1583	0.67
4254	輪	1957	2297	0.46
4157	人	1189	2968	0.29
3141	日本	1353	1788	0.43
3051	上	1262	1789	0.41
2740	分	1151	1589	0.42
2499	口	722	1777	0.29
2451	部	263	2188	0.11
2265	場	278	1987	0.12
2261	年	599	1662	0.26
2219	差	1689	530	0.76
2204	西	1366	838	0.62
2192	東	874	1318	0.40
2173	目	478	1695	0.22
1973	合	679	1294	0.34
1941	後	413	1528	0.21
1849	入	1100	749	0.59
1825	島	256	1569	0.14
1713	月	475	1238	0.28
1673	タイ	160	1513	0.10
1626	共同	137	1489	0.08
1590	野	590	1000	0.37
1564	相	49	1515	0.03
1540	円	270	1270	0.18
1533	運動	1019	514	0.66
1474	度	414	1060	0.28
1469	球	207	1262	0.14
1421	節	15	1406	0.01
1420	ラン	36	1384	0.03
1393	手	287	1106	0.21
1388	前	326	1062	0.23
1359	回	300	1059	0.22
1357	派	257	1100	0.19
1354	高	619	735	0.46
1339	イド	3	1336	0.00
1323	対	1027	296	0.78
1309	子	163	1146	0.12

表6-3-3

完全一致・部分一致による高頻度の日本語用語（一般記事）

(2)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
1309	トル	2	1307	0.00
1290	下	425	865	0.33
1281	初日	896	385	0.70
1280	注	180	1100	0.14
1273	カ	629	644	0.49
1255	点	517	738	0.41
1241	問題	448	793	0.36
1239	山	256	983	0.21
1229	馬	283	946	0.23
1218	方	333	885	0.27
1205	写真	473	732	0.39
1166	力	157	1009	0.13
1141	タイト	7	1134	0.01
1122	地	114	1008	0.10
1112	金	287	825	0.26
1091	半	29	1062	0.03
1083	時間	133	950	0.12
1081	京都	193	888	0.18
1067	性	28	1039	0.03
1060	成績	161	899	0.15
1039	終日	4	1035	0.00
1036	体	75	961	0.07
988	以上	569	419	0.58
941	補	40	901	0.04
939	会議	147	792	0.16
914	リン	62	852	0.07
872	真	73	799	0.08
803	賞	79	724	0.10
794	心	94	700	0.12
792	秒	453	339	0.57
789	次	533	256	0.68
787	曇	443	344	0.56
771	雨	546	225	0.71
764	率	130	634	0.17
760	正	176	584	0.23
745	関係	184	561	0.25
728	シア	38	690	0.05
715	市場	152	563	0.21
712	側	165	547	0.23
705	文	29	676	0.04

表6-3-3

完全一致・部分一致による高頻度の日本語用語（一般記事）

(3)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
697	米	390	307	0.56
694	代	83	611	0.12
689	調査	226	463	0.33
678	数	175	503	0.26
663	トン	30	633	0.05
658	木	184	474	0.28
656	中心	309	347	0.47
653	水	127	526	0.19
649	晴	267	382	0.41
647	号	367	280	0.57
643	予報	23	620	0.04
634	会社	125	509	0.20
632	情報	134	498	0.21
620	法	76	544	0.12
616	日出	0	616	0.00
609	大学	151	458	0.25
607	頭	401	206	0.66
603	新聞	68	535	0.11
602	名	142	460	0.24
600	崎	103	497	0.17
594	南	264	330	0.44
591	世界	243	348	0.41
590	確率	295	295	0.50
585	注意	66	519	0.11
581	比	214	367	0.37
578	界	33	545	0.06
576	道	103	473	0.18
573	監督	131	442	0.23
571	天気	430	141	0.75
570	車	160	410	0.28
565	候補	245	320	0.43
552	関連	341	211	0.62
551	送信	152	399	0.28
551	州	191	360	0.35
549	議会	47	502	0.09
545	企業	161	384	0.30
539	期	37	502	0.07
537	訂正	164	373	0.31
534	口	161	373	0.30
526	中央	26	500	0.05

表6-3-4

完全一致・部分一致による高頻度の日本語用語（情報分野論文要旨） (1)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
1430	システム	1172	258	0.82
1354	数	184	1170	0.14
1350	研究	500	850	0.37
1274	データ	728	546	0.57
1231	処理	348	883	0.28
1078	式	145	933	0.13
982	度	18	964	0.02
981	語	78	903	0.08
959	上	523	436	0.55
845	性	35	810	0.04
776	計算	210	566	0.27
679	言語	229	450	0.34
675	情報	202	473	0.30
662	開発	540	122	0.82
606	方法	448	158	0.74
604	解析	225	379	0.37
591	対	551	40	0.93
566	問題	292	274	0.52
562	分	121	441	0.22
557	点	153	404	0.27
549	モデル	483	66	0.88
547	結果	386	161	0.71
530	文字	218	312	0.41
526	方式	248	278	0.47
514	グラム	14	500	0.03
512	ベース	80	432	0.16
504	認識	151	353	0.30
502	可能	269	233	0.54
501	画像	170	331	0.34
493	文	118	375	0.24
483	図	410	73	0.85
478	値	284	194	0.59
464	プログラム	433	31	0.93
462	必要	415	47	0.90
456	記述	255	201	0.56
451	利用	299	152	0.66
445	号	15	430	0.03
437	符号	163	274	0.37
436	次	403	33	0.92
430	設計	286	144	0.67

表6-3-4

完全一致・部分一致による高頻度の日本語用語（情報分野論文要旨） (2)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
428	像	23	405	0.05
424	並列	129	295	0.30
424	反射	49	375	0.12
423	力	37	386	0.09
416	形	148	268	0.36
409	生成	266	143	0.65
402	集合	212	190	0.53
400	リズム	19	381	0.05
398	関数	197	201	0.50
393	構造	181	212	0.46
388	データベース	290	98	0.75
385	表現	218	167	0.57
380	表	233	147	0.61
374	漢字	183	191	0.49
367	構成	235	132	0.64
366	評価	221	145	0.60
365	関係	173	192	0.47
352	実験	140	212	0.40
346	機能	186	160	0.54
337	アルゴリズム	314	23	0.93
336	時間	72	264	0.21
331	型	101	230	0.31
325	下	34	291	0.10
323	系	30	293	0.09
322	物体	103	219	0.32
322	ター	0	322	0.00
321	知識	165	156	0.51
317	価	8	309	0.03
313	係	16	297	0.05
305	実行	135	170	0.44
304	部	44	260	0.14
298	定義	224	74	0.75
290	検索	154	136	0.53
287	論理	51	236	0.18
284	対応	220	64	0.77
283	意味	124	159	0.44
283	プロセッサ	205	78	0.72
281	後	53	228	0.19
276	類	12	264	0.04
273	解	86	187	0.32

表 6-3-4

完全一致・部分一致による高頻度の日本語用語 (情報分野論文要旨) (3)

頻度数	用語	完全一致カント	部分一致カント	完全/部分一致係数
272	操作	177	95	0.65
261	抽出	166	95	0.64
260	組	183	77	0.70
260	トル	3	257	0.01
258	辞書	112	146	0.43
257	音声	75	182	0.29
251	以下	231	20	0.92
241	変換	128	113	0.53
241	線	10	231	0.04
240	タイ	30	210	0.13
238	色	67	171	0.28
237	高速	87	150	0.37
231	子	12	219	0.05
231	メモリ	181	50	0.78
230	リン	5	225	0.02
228	部分	87	141	0.38
227	効率	102	125	0.45
224	ソフトウェア	203	21	0.91
222	推論	85	137	0.38
222	計算機	83	139	0.37
221	約	92	129	0.42
220	特徴	159	61	0.72
220	種	110	110	0.50
215	演算	69	146	0.32
213	入力	93	120	0.44
212	日本	32	180	0.15
208	変数	110	98	0.53
207	構文	22	185	0.11
205	仕様	62	143	0.30
203	目的	164	39	0.81
203	方向	85	118	0.42
203	制約	72	131	0.35
201	自動	17	184	0.08
199	形式	63	136	0.32
199	ベクトル	169	30	0.85
197	文法	93	104	0.47
196	有効	122	74	0.62
196	条件	84	112	0.43
194	要素	103	91	0.53
194	通信	86	108	0.44

表6-3-5

完全/部分一致係数順用語表 (高係数用語, 一般記事)

(1)

頻度数	用語	完全一致カント	部分一致カント	完全/部分一致係数
480	前節	480	0	1.00
294	降水	294	0	1.00
133	湿度	133	0	1.00
133	最高気温	133	0	1.00
76	天災	76	0	1.00
66	ウルグアイ	66	0	1.00
53	最大風速	53	0	1.00
48	進行方向	48	0	1.00
45	プルトニウム	45	0	1.00
37	胸	37	0	1.00
34	メカニズム	34	0	1.00
29	アンゴラ	29	0	1.00
27	雰囲気	27	0	1.00
23	ヘリコプタ	23	0	1.00
20	プライバシ	20	0	1.00
20	タイプ	20	0	1.00
19	枠組	19	0	1.00
19	義足	19	0	1.00
19	ノウハウ	19	0	1.00
18	空気	18	0	1.00
18	トンネル	18	0	1.00
18	シナリオ	18	0	1.00
17	無回答	17	0	1.00
17	芋	17	0	1.00
16	缶	16	0	1.00
15	識別	15	0	1.00
15	殻	15	0	1.00
15	ドラフト	15	0	1.00
14	詳細	14	0	1.00
14	根拠	14	0	1.00
14	カレンダー	14	0	1.00
13	ユネスコ	13	0	1.00
13	ヒビ	13	0	1.00
12	複写機	12	0	1.00
12	不純物	12	0	1.00
12	仕切	12	0	1.00
12	温帯低気圧	12	0	1.00
12	ホルモン	12	0	1.00
12	ビタミン	12	0	1.00
11	口火	11	0	1.00

表6-3-5

完全/部分一致係数順用語表 (高係数用語, 一般記事)

(2)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
11	確信	11	0	1.00
11	プロセス	11	0	1.00
11	プログラム	11	0	1.00
11	トランク	11	0	1.00
11	スカウト	11	0	1.00
11	アトリエ	11	0	1.00
10	有感地震	10	0	1.00
10	農作物	10	0	1.00
10	静電気	10	0	1.00
10	机	10	0	1.00
10	ピリオド	10	0	1.00
10	アクセサリ	10	0	1.00
9	理念	9	0	1.00
9	毛玉	9	0	1.00
9	頂点	9	0	1.00
9	前面	9	0	1.00
9	責務	9	0	1.00
9	信念	9	0	1.00
9	儀式	9	0	1.00
9	寄与	9	0	1.00
9	移送	9	0	1.00
9	バンパ	9	0	1.00
9	ストレス	9	0	1.00
9	サルモネラ	9	0	1.00
8	労働生産性	8	0	1.00
8	利点	8	0	1.00
8	廃棄処分	8	0	1.00
8	同居	8	0	1.00
8	塗料	8	0	1.00
8	電波望遠鏡	8	0	1.00
8	酢	8	0	1.00
8	神経質	8	0	1.00
8	講義	8	0	1.00
8	活力	8	0	1.00
8	角膜	8	0	1.00
8	レモン	8	0	1.00
8	バルブ	8	0	1.00
8	トカゲ	8	0	1.00
7	埋蔵量	7	0	1.00
7	標的	7	0	1.00

表6-3-5

完全/部分一致係数順用語表 (高係数用語, 一般記事)

(3)

類度数	用語	完全一致カント	部分一致カント	完全/部分一致係数
7	赤外線	7	0	1.00
7	推定位置	7	0	1.00
7	求心力	7	0	1.00
7	機長	7	0	1.00
7	悪夢	7	0	1.00
7	ライトバン	7	0	1.00
7	ハンドル	7	0	1.00
7	サテライト	7	0	1.00
7	イリオモテヤマネコ	7	0	1.00
6	有機農業	6	0	1.00
6	帽子	6	0	1.00
6	風向計	6	0	1.00
6	肺	6	0	1.00
6	凍死	6	0	1.00
6	接続	6	0	1.00
6	消化器系	6	0	1.00
6	散乱	6	0	1.00
6	原動力	6	0	1.00
6	見当	6	0	1.00
6	横行	6	0	1.00
6	一時停止	6	0	1.00
6	リモコン	6	0	1.00
6	ライフル	6	0	1.00
6	ズボン	6	0	1.00
6	カジノ	6	0	1.00
6	アスベスト	6	0	1.00
5	力量	5	0	1.00
5	法定伝染病	5	0	1.00
5	判例	5	0	1.00
5	判断力	5	0	1.00
5	培地	5	0	1.00
5	動力炉	5	0	1.00
5	都市構造	5	0	1.00
5	調整作業	5	0	1.00
5	調査方法	5	0	1.00
5	第一段階	5	0	1.00
5	生命維持装置	5	0	1.00
5	整列	5	0	1.00
5	制約	5	0	1.00
5	柔軟性	5	0	1.00

表6-3-6

完全/部分一致係数順用語表 (低係数用語, 一般記事)

(1)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
616	日出	0	616	0.00
132	暴風	0	132	0.00
108	ボジ	0	108	0.00
99	タケ	0	99	0.00
93	ピン	0	93	0.00
72	沿岸	0	72	0.00
68	脱硫	0	68	0.00
59	シカ	0	59	0.00
59	カメ	0	59	0.00
57	風速	0	57	0.00
53	注出	0	53	0.00
51	マスト	0	51	0.00
51	ビデ	0	51	0.00
48	ベント	0	48	0.00
42	深層	0	42	0.00
42	ルナ	0	42	0.00
42	タイガ	0	42	0.00
39	衛生	0	39	0.00
39	ウシ	0	39	0.00
38	山地	0	38	0.00
35	遺伝	0	35	0.00
32	生理学	0	32	0.00
31	ミリ	0	31	0.00
29	人道	0	29	0.00
28	物理学	0	28	0.00
28	自転	0	28	0.00
27	糖	0	27	0.00
27	イネ	0	27	0.00
26	曜日	0	26	0.00
26	胞	0	26	0.00
25	一次	0	25	0.00
25	ヘルツ	0	25	0.00
24	図書	0	24	0.00
24	ウマ	0	24	0.00
23	原本	0	23	0.00
23	家主	0	23	0.00
22	中期	0	22	0.00
22	致死	0	22	0.00
22	査察	0	22	0.00
21	予備	0	21	0.00

表6-3-6

完全/部分一致係数順用語表 (低係数用語, 一般記事)

(2)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
21	人工呼吸	0	21	0.00
21	ログ	0	21	0.00
21	レア	0	21	0.00
20	工学	0	20	0.00
20	ラバ	0	20	0.00
20	マイクロ	0	20	0.00
20	ツバ	0	20	0.00
20	クマ	0	20	0.00
19	棟	0	19	0.00
19	鳥居	0	19	0.00
19	運動公園	0	19	0.00
19	スキン	0	19	0.00
19	カー	0	19	0.00
18	素子	0	18	0.00
18	十字	0	18	0.00
18	光子	0	18	0.00
18	ロゴ	0	18	0.00
18	ウジ	0	18	0.00
17	里子	0	17	0.00
17	総生産	0	17	0.00
17	精製	0	17	0.00
17	水素化	0	17	0.00
17	植木	0	17	0.00
17	气象台	0	17	0.00
17	ブイ	0	17	0.00
17	ナブ	0	17	0.00
17	トキ	0	17	0.00
17	シバ	0	17	0.00
17	オクラ	0	17	0.00
16	標準	0	16	0.00
16	全長	0	16	0.00
16	水素化脱硫	0	16	0.00
16	タラ	0	16	0.00
15	予知	0	15	0.00
15	調教	0	15	0.00
15	小石	0	15	0.00
15	広域公園	0	15	0.00
15	沿海	0	15	0.00
14	防護	0	14	0.00
14	天球	0	14	0.00

表6-3-6

完全/部分一致係数順用語表 (低係数用語, 一般記事)

(3)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
14	注水	0	14	0.00
14	弾頭	0	14	0.00
14	大潮	0	14	0.00
14	石油精製	0	14	0.00
14	石器	0	14	0.00
14	滑走	0	14	0.00
14	マイル	0	14	0.00
13	内甲	0	13	0.00
13	素粒子	0	13	0.00
13	潜水	0	13	0.00
13	上米	0	13	0.00
13	交差	0	13	0.00
13	割引	0	13	0.00
13	ラメ	0	13	0.00
12	表作	0	12	0.00
12	総合計画	0	12	0.00
12	真理	0	12	0.00
12	市松	0	12	0.00
12	号外	0	12	0.00
12	近海	0	12	0.00
12	間接	0	12	0.00
12	肝炎	0	12	0.00
12	外線	0	12	0.00
12	河川敷	0	12	0.00
12	下水	0	12	0.00
12	トルエン	0	12	0.00
12	サクラ	0	12	0.00
12	グラム	0	12	0.00
12	グラス	0	12	0.00
12	オア	0	12	0.00
12	イヌ	0	12	0.00
11	膜	0	11	0.00
11	保育	0	11	0.00
11	病原	0	11	0.00
11	内陸	0	11	0.00
11	地動	0	11	0.00
11	地震予知	0	11	0.00
11	請書	0	11	0.00
11	生糸	0	11	0.00
11	三日月	0	11	0.00

表6-3-7

完全/部分一致係数順用語表 (中係数用語, 一般記事)

(1)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
3141	日本	1353	1788	0.43
1354	高	619	735	0.46
1273	カ	629	644	0.49
988	以上	569	419	0.58
792	秒	453	339	0.57
787	曇	443	344	0.56
697	米	390	307	0.56
656	中心	309	347	0.47
647	号	367	280	0.57
594	南	264	330	0.44
590	確率	295	295	0.50
565	候補	245	320	0.43
552	関連	341	211	0.62
493	組	223	270	0.45
479	女性	276	203	0.58
460	減	218	242	0.47
447	小	254	193	0.57
426	重	209	217	0.49
390	北	239	151	0.61
367	科学	193	174	0.53
348	先	155	193	0.45
344	顔	193	151	0.56
344	開発	154	190	0.45
339	形	173	166	0.51
324	反対	190	134	0.59
320	首	174	146	0.54
316	ドル	165	151	0.52
308	要求	157	151	0.51
283	国内	129	154	0.46
273	内容	166	107	0.61
273	決定	129	144	0.47
255	現場	129	126	0.51
239	ビル	135	104	0.56
234	状況	124	110	0.53
227	強化	120	107	0.53
221	導入	100	121	0.45
218	対立	121	97	0.56
212	白	123	89	0.58
199	右	91	108	0.46
195	左	97	98	0.50

表6-3-7

完全/部分一致係数順用語表 (中係数用語, 一般記事) (2)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
189	状態	103	86	0.55
188	発生	90	98	0.48
188	出発	106	82	0.56
171	患者	96	75	0.56
171	会合	89	82	0.52
170	秋	87	83	0.51
169	挿入	102	67	0.60
167	観測	75	92	0.45
165	不安	94	71	0.57
164	一致	87	77	0.53
163	芝	91	72	0.56
163	案内	91	72	0.56
159	拡大	75	84	0.47
157	合併	87	70	0.55
147	積	67	80	0.46
147	正午	76	71	0.52
142	方向	75	67	0.53
142	部分	78	64	0.55
135	高知	76	59	0.56
135	クラブ	72	63	0.53
132	日程	61	71	0.46
132	材料	62	70	0.47
127	星	66	61	0.52
127	上昇	72	55	0.57
126	安定	57	69	0.45
121	改善	57	64	0.47
116	経験	68	48	0.59
114	方法	61	53	0.54
114	相談	50	64	0.44
112	焼	55	57	0.49
111	分析	63	48	0.57
107	鼻	63	44	0.59
106	舞台	59	47	0.56
105	信頼	53	52	0.50
104	輸送	49	55	0.47
104	泉	58	46	0.56
102	世	48	54	0.47
100	反応	50	50	0.50
98	文書	44	54	0.45
98	中止	48	50	0.49

表6-3-7

完全/部分一致係数順用語表 (中係数用語, 一般記事)

(3)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
98	竹	43	55	0.44
94	削除	49	45	0.52
94	ベスト	51	43	0.54
91	崩壊	46	45	0.51
90	談合	44	46	0.49
90	搜索	54	36	0.60
90	肝臓	47	43	0.52
88	機能	46	42	0.52
88	火災	52	36	0.59
86	促進	46	40	0.53
86	衝突	53	33	0.62
85	得点	44	41	0.52
84	意識	49	35	0.58
82	攻撃	41	41	0.50
81	効果	44	37	0.54
81	回答	45	36	0.56
79	柳	36	43	0.46
77	禁止	35	42	0.45
76	西日本	40	36	0.53
75	柱	41	34	0.55
72	腎臓	34	38	0.47
72	集中	32	40	0.44
71	指導者	31	40	0.44
71	ランク	34	37	0.48
70	現実	36	34	0.51
67	壁	31	36	0.46
67	基盤	29	38	0.43
67	アメリカ	29	38	0.43
66	負担	31	35	0.47
66	転換	30	36	0.45
66	脱硫装置	33	33	0.50
66	沖	34	32	0.52
63	復帰	38	25	0.60
63	完全	28	35	0.44
62	大規模	28	34	0.45
62	正方形	36	26	0.58
61	有効	37	24	0.61
61	報告書	37	24	0.61
61	破壊	31	30	0.51
61	設定	34	27	0.56

表6-3-8

完全/部分一致係数順用語表 (高係数用語, 情報分野論文要旨) (1)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
119	コーディング	119	0	1.00
89	スロット	89	0	1.00
65	ストローク	65	0	1.00
62	カラーモニタ	62	0	1.00
52	ステージ	52	0	1.00
46	フォーリエ	46	0	1.00
44	非常	44	0	1.00
40	枝	40	0	1.00
39	フローグラフ	39	0	1.00
39	ステップ	39	0	1.00
36	定式化	36	0	1.00
36	期待	36	0	1.00
34	自動的	34	0	1.00
32	部分集合	32	0	1.00
31	前者	31	0	1.00
27	インプリメント	27	0	1.00
27	アプローチ	27	0	1.00
26	オペレーティングシステム	26	0	1.00
25	サービス	25	0	1.00
24	オンライン	24	0	1.00
22	入射角	22	0	1.00
21	ラベル	21	0	1.00
21	フェーズ	21	0	1.00
21	ソフトウェア	21	0	1.00
19	利点	19	0	1.00
19	欠点	19	0	1.00
18	主語	18	0	1.00
17	マスク	17	0	1.00
17	プログラマ	17	0	1.00
17	フレネル	17	0	1.00
16	特別	16	0	1.00
16	データフロー	16	0	1.00
16	ガンマ	16	0	1.00
15	推進	15	0	1.00
15	安全性	15	0	1.00
15	スケジューリング	15	0	1.00
14	メカニズム	14	0	1.00
14	サンプル	14	0	1.00
13	理由	13	0	1.00
13	代入文	13	0	1.00

表6-3-8

完全/部分一致係数順用語表 (高係数用語, 情報分野論文要旨) (2)

度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
13	主観的輪郭	13	0	1.00
13	左	13	0	1.00
13	マイクロプロセッサ	13	0	1.00
13	デバッグ	13	0	1.00
12	短縮	12	0	1.00
12	情報圧縮	12	0	1.00
12	三刺激値	12	0	1.00
12	花	12	0	1.00
12	ワードプロセッサ	12	0	1.00
12	ビデオテープ	12	0	1.00
12	アッドロック	12	0	1.00
12	スカラ	12	0	1.00
12	エントロピー	12	0	1.00
11	分光分布	11	0	1.00
11	単体	11	0	1.00
11	色座標変換	11	0	1.00
11	情報源	11	0	1.00
11	上位	11	0	1.00
11	除去	11	0	1.00
11	基本設計	11	0	1.00
11	ロボット	11	0	1.00
11	プラスチック	11	0	1.00
11	アナログ	11	0	1.00
10	類似度	10	0	1.00
10	相互接続	10	0	1.00
10	色知覚	10	0	1.00
10	規定	10	0	1.00
10	基準点	10	0	1.00
10	回避	10	0	1.00
10	一連	10	0	1.00
10	ユニット	10	0	1.00
10	ドキュメント	10	0	1.00
10	コンピュータグラフィックス	10	0	1.00
10	アプリケーション	10	0	1.00
9	定義域	9	0	1.00
9	最大値	9	0	1.00
9	具体化	9	0	1.00
9	応答時間	9	0	1.00
9	スケッチ	9	0	1.00
9	コンフリクト	9	0	1.00

表6-3-8

完全/部分一致係数順用語表 (高係数用語, 情報分野論文要旨) (3)

度数	用語	完全一致カント	部分一致カント	完全/部分一致係数
9	インプリメンテーション	9	0	1.00
9	アレイ	9	0	1.00
8	物理量	8	0	1.00
8	比例	8	0	1.00
8	相互作用	8	0	1.00
8	情報量	8	0	1.00
8	識別子	8	0	1.00
8	吸収係数	8	0	1.00
8	位数	8	0	1.00
8	暗号文	8	0	1.00
8	マニュアル	8	0	1.00
8	ピッチ	8	0	1.00
8	ピクセル	8	0	1.00
8	ヒストグラム	8	0	1.00
8	パーサ	8	0	1.00
8	ニューロン	8	0	1.00
8	テレビジョン	8	0	1.00
8	クロック	8	0	1.00
8	エネルギー	8	0	1.00
7	陽	7	0	1.00
7	非	7	0	1.00
7	銅	7	0	1.00
7	通信方式	7	0	1.00
7	値域	7	0	1.00
7	垂直入射	7	0	1.00
7	順番	7	0	1.00
7	修飾	7	0	1.00
7	周波数領域	7	0	1.00
7	試作機	7	0	1.00
7	再構築	7	0	1.00
7	互換性	7	0	1.00
7	意味分析	7	0	1.00
7	プロダクションシステム	7	0	1.00
7	プリプロセッサ	7	0	1.00
7	ファントム	7	0	1.00
7	ドイツ	7	0	1.00
7	トレース	7	0	1.00
7	トラフィック	7	0	1.00
7	データベース	7	0	1.00
7	スレッド	7	0	1.00

表6-3-9

完全/部分一致係数順用語表 (低係数用語, 情報分野論文要旨) (1)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
667	ログ	0	667	0.00
72	ラメ	0	72	0.00
65	ストロー	0	65	0.00
57	レート	0	57	0.00
49	トレー	0	49	0.00
45	ジュール	0	45	0.00
41	エクス	0	41	0.00
37	輝度計	0	37	0.00
37	ダイ	0	37	0.00
36	針	0	36	0.00
35	仮名	0	35	0.00
31	帰納	0	31	0.00
30	フェイス	0	30	0.00
30	ビデオ	0	30	0.00
27	脈	0	27	0.00
25	波数	0	25	0.00
25	対数	0	25	0.00
25	周波	0	25	0.00
25	パー	0	25	0.00
23	不均質	0	23	0.00
22	板	0	22	0.00
22	ボジ	0	22	0.00
21	微分	0	21	0.00
20	分光計	0	20	0.00
20	マット	0	20	0.00
20	タック	0	20	0.00
19	リーム	0	19	0.00
19	スプレイ	0	19	0.00
18	排他	0	18	0.00
18	二階	0	18	0.00
18	導体	0	18	0.00
18	電気	0	18	0.00
18	点集合	0	18	0.00
18	線図	0	18	0.00
17	関節	0	17	0.00
17	ガン	0	17	0.00
16	算術	0	16	0.00
16	感覚	0	16	0.00
16	ダム	0	16	0.00
15	星	0	15	0.00

表6-3-9

完全/部分一致係数順用語表 (低係数用語, 情報分野論文要旨) (2)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
14	二次	0	14	0.00
14	電位	0	14	0.00
14	素因数	0	14	0.00
14	語音	0	14	0.00
14	可分	0	14	0.00
14	センシング	0	14	0.00
14	カニ	0	14	0.00
13	跡	0	13	0.00
13	赤外	0	13	0.00
13	情報通信	0	13	0.00
13	書体	0	13	0.00
13	九州	0	13	0.00
13	ベンチ	0	13	0.00
12	断面	0	12	0.00
12	谷	0	12	0.00
12	処理装置	0	12	0.00
12	科	0	12	0.00
12	ビジョン	0	12	0.00
12	ジン	0	12	0.00
12	エントロピ	0	12	0.00
11	相対	0	11	0.00
11	研究所	0	11	0.00
11	教授	0	11	0.00
11	プラス	0	11	0.00
10	不安	0	10	0.00
10	底	0	10	0.00
10	通分	0	10	0.00
10	剰余	0	10	0.00
10	小数	0	10	0.00
10	系統	0	10	0.00
10	プロダクション	0	10	0.00
10	バラ	0	10	0.00
10	ネック	0	10	0.00
10	トン	0	10	0.00
10	セント	0	10	0.00
9	柄	0	9	0.00
9	表色	0	9	0.00
9	大域	0	9	0.00
9	多目	0	9	0.00
9	小数点	0	9	0.00

表6-3-9

完全/部分一致係数順用語表 (低係数用語, 情報分野論文要旨) (3)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
9	謝辞	0	9	0.00
9	時間計	0	9	0.00
9	ロール	0	9	0.00
9	レース	0	9	0.00
9	フラン	0	9	0.00
9	テレビ	0	9	0.00
9	スティック	0	9	0.00
9	エネルギー	0	9	0.00
9	アルファ	0	9	0.00
8	物理学	0	8	0.00
8	二進木	0	8	0.00
8	動性	0	8	0.00
8	絶対	0	8	0.00
8	推理	0	8	0.00
8	互換	0	8	0.00
8	一階	0	8	0.00
8	ボトル	0	8	0.00
8	ファン	0	8	0.00
8	トキ	0	8	0.00
8	シカ	0	8	0.00
8	ゲット	0	8	0.00
8	アリ	0	8	0.00
7	腹	0	7	0.00
7	表面電位	0	7	0.00
7	備品	0	7	0.00
7	認証	0	7	0.00
7	肉	0	7	0.00
7	内耳	0	7	0.00
7	代数方程式	0	7	0.00
7	赤外面像	0	7	0.00
7	実質	0	7	0.00
7	喉頭	0	7	0.00
7	ワーブ	0	7	0.00
7	リカー	0	7	0.00
7	ラジオ	0	7	0.00
7	ムサ	0	7	0.00
7	プリ	0	7	0.00
7	ジオ	0	7	0.00
7	ウシ	0	7	0.00
6	量子	0	6	0.00

表6-3-10

完全/部分一致係数順用語表 (中係数用語, 情報分野論文要旨) (1)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
1274	データ	728	546	0.57
604	解析	225	379	0.37
566	問題	292	274	0.52
530	文字	218	312	0.41
526	方式	248	278	0.47
502	可能	269	233	0.54
478	値	284	194	0.59
456	記述	255	201	0.56
437	符号	163	274	0.37
402	集合	212	190	0.53
398	関数	197	201	0.50
393	構造	181	212	0.46
385	表現	218	167	0.57
380	表	233	147	0.61
374	漢字	183	191	0.49
366	評価	221	145	0.60
365	関係	173	192	0.47
352	実験	140	212	0.40
346	機能	186	160	0.54
321	知識	165	156	0.51
305	実行	135	170	0.44
290	検索	154	136	0.53
283	意味	124	159	0.44
271	対象	164	107	0.61
258	辞書	112	146	0.43
241	変換	128	113	0.53
228	部分	87	141	0.38
227	効率	102	125	0.45
222	推論	85	137	0.38
222	計算機	83	139	0.37
220	種	110	110	0.50
213	入力	93	120	0.44
208	変数	110	98	0.53
203	方向	85	118	0.42
197	文法	93	104	0.47
196	有効	122	74	0.62
196	条件	84	112	0.43
194	要素	103	91	0.53
194	通信	86	108	0.44
176	性能	93	83	0.53

表 6-3-10

完全/部分一致係数順用語表 (中係数用語, 情報分野論文要旨) (2)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
174	章	86	88	0.49
168	キー	100	68	0.60
166	図形	76	90	0.46
165	決定	81	84	0.49
164	表示	75	89	0.46
157	制御	71	86	0.45
155	分析	75	80	0.48
154	タイプ	65	89	0.42
147	正	66	81	0.45
147	診断	58	89	0.39
146	日本語	74	72	0.51
145	動詞	78	67	0.54
144	領域	66	78	0.46
144	基礎	59	85	0.41
142	最適	58	84	0.41
141	翻訳	65	76	0.46
141	単語	55	86	0.39
139	動的	66	73	0.47
135	表面	56	79	0.41
133	環境	62	71	0.47
123	概念	71	52	0.58
121	証明	73	48	0.60
120	接続	67	53	0.56
116	分類	63	53	0.54
115	過程	47	68	0.41
114	中心	59	55	0.52
114	効果	52	62	0.46
110	合成	47	63	0.43
106	位置	62	44	0.58
105	精度	40	65	0.38
105	出力	55	50	0.52
104	ユーザ	57	47	0.55
103	指向	50	53	0.49
99	作業	52	47	0.53
99	形状	37	62	0.37
99	学習	48	51	0.48
96	理解	49	47	0.51
95	要求	41	54	0.43
95	ロック	45	50	0.47
93	認識率	51	42	0.55

表 6-3-10

完全/部分一致係数順用語表 (中係数用語, 情報分野論文要旨) (3)

頻度数	用語	完全一致カウント	部分一致カウント	完全/部分一致係数
93	セル	50	43	0.54
92	相互反射	43	49	0.47
91	十分	41	50	0.45
87	詳細	53	34	0.61
84	段階	39	45	0.46
81	命令	49	32	0.60
81	定理	46	35	0.57
80	マシン	42	38	0.53
79	項	48	31	0.61
78	圧縮	43	35	0.55
76	分割	36	40	0.47
75	識別	39	36	0.52
75	完全	44	31	0.59
74	目録	37	37	0.50
74	特性	44	30	0.59
73	光線	33	40	0.45
73	改善	43	30	0.59
72	単位	36	36	0.50
72	視覚	32	40	0.44
71	状態	38	33	0.54
71	計測	36	35	0.51
67	区間	35	32	0.52
66	配列	34	32	0.52
63	状況	31	32	0.49
63	インタフェース	31	32	0.49
62	故障	29	33	0.47
61	方程式	26	35	0.43
61	多数	29	32	0.48
60	範囲	32	28	0.53
60	係数	28	32	0.47
59	共通	32	27	0.54
56	負荷	27	29	0.48
54	行列	25	29	0.46
52	照合	20	32	0.38
51	補正	22	29	0.43
51	頻度	30	21	0.59
51	解釈	31	20	0.61
50	同期	30	20	0.60
50	キャッシュ	24	26	0.48
48	変動	26	22	0.54

7. 専門辞書更新のシステム化

翻訳用の用語辞書は、情報化と技術革新の進展が急速な今日状況において、日々新たに発生する専門用語群に、遅滞なく追従してゆかなければならない。このためには、翻訳用専門用語辞書のデータベース化と、その更新管理のシステム化が不可欠である。これに向けては、すでにのべた辞書の連鎖検索と出現頻度統計を活用することで、効果的なシステムを構成できる可能性がある。そこで本章では、これまでに得られた知見を踏まえて、用語の出現頻度統計と連鎖検索を組み合わせた、新出専門用語の検定システムを中心に検討する。また、このような新出専門語の検出と検定に対して、機械翻訳システムを、どのように関連させ、活用することができるのかという、将来的な展望についても考察する。

7. 1 出現頻度統計と連鎖検索による辞書更新の方式

(1) 日本語と英語からの双方向接近

現今、情報流通の国際化には著しいものがあり、本調査研究のテーマとなっている機械翻訳システムも、この背景において、一層必要性が高まっているわけである。このような国際化状況の中で、文献、記事等の中から、新たな用語を早期に適確に発見してゆくためには、英日翻訳のためにもつばら英文資料を点検する、あるいはその逆に、日英翻訳用に日本語資料だけを通覧するといった、一方通行的アプローチでは、効率が悪いことは明きらかである。

すでに、本調査研究では、和英両引きの機械化専門用語辞書の構成を試みて、これに統計的分析を加え、その大要を明らかにしてきた。この和英双方向からのアプローチは、単に辞書構成上の問題にとどまらず、新語の発見機構のためにも有効である。今日、新たな用語は、英語文献にも日本語文献にも同時多発的に登場するとみられる。従って、例えば英日翻訳を主体に考えて、英文資料だけから新語を見つけ出して、それらに和訳語を与えてゆくといった、一方向の接近では、新出専門用語への追従性に限界がある。日英翻訳と英日翻訳を均等にとらえて、和文と英文の資料双方から、新語を採取してゆく体制が、ここに提案されるところである。

(2) 新出用語の検出と検定

新出の専門用語とは、単純には、辞書への未登録語ということになると考えられる。しかし、實際上、事態はそれほど簡単ではない。現在、新規に考案される専門用語の大部分は、形態的にみると、複合語である。すなわち、英用語では、複数の単語を連ねた語句の形態であり、和用語の場合は、既成の用語を接続させたものである。

従って、これらの構成要素である、それぞれの英単語、あるいは和文の要素語は、すでに辞書に登録済みである。しかし、これら各要素語に対する翻訳を連ねたからといって、専門用語としての正しい翻訳にはならないという点が問題である。

従って、何らかの方法により、こうした新出の複合的専門用語を早期に発見し、その訳語とともに、これらを辞書に登録してゆく必要がある。英文の複合専門用語の抽出方法については、すでに昨年度の調査研究において、実験的システムを開発し、その効果を検討した。和文の複合専門用語については、既述の、分かち書き結果の部分文字列への分解・照合処理において、基本的な方式が実験されている。これらの方法によれば、複合的な専門用語についても、辞書との突き合わせは可能であり、従って、未登録語も自動的に検出される。そこで、問題は、そうした未登録語が、はたして本当に新出の専門用語であるか否かの判別という点にあることになる。

この種の判定を、機械処理により全自動化するのは本来的に困難である。従って、ここでは、専門家による最終的な判別作業、すなわち新語の「検定」作業を能率化するために、効果的な新出専門語候補リストが、機械処理によっていかに提出できるかということが関心の焦点になる。

(3) 出現頻度の時系列統計

英文の場合は一定数までの単語列を、和文については一定文字数までの文字列を、辞書と突き合わせれば、新出の複合専門語のほかに、実際には複合関係にない語句も、未登録語として無数に検出されることになる。従って、これらのうちから、複合語である可能性の高いものだけに絞り込んだ、新語候補リストを生成する必要がある。こうした絞り込み処理のための指標としては、語句の出現頻度データが有効であると考えられる。

昨年度の調査研究の一環として、用語の出現頻度の時系列統計を試みたが、新語候補の絞り込み処理においては、出現頻度の時系列的な変化を検出するというのが、有効な方法として考えられる。すなわち、その出現頻度がにわかに立ち上がってきた語句を、新出の用語である可能性が高いと判定し、これを新語の候補として提示するという方式である。これによれば、相当の確度を以て、新語候補の提示が行えると考えられる。もっとも、こうした処理には、頻度データの不断の蓄積を要するので、後に提案するように、それなりの専門的機関による取り組みが必要と思われる。

(4) 連鎖検索ツール

つぎに、上記のような新語候補リストに基づいて、新出の専門用語を個別に検定してゆく際の道具として、訳語連鎖検索のシステム化が考えられる。両引き辞書を用いて、英→和→英→和→・・・のように、訳語の連鎖をたどって検索してゆくと、どの

ような結果になるかについては、すでにのべたところであるが、このような方法は、新語を新語として確認する作業において、有効に機能すると思われる。実際、専門用語を新たに辞書登録するには、慎重であるべきである。新語の登録は、それが複合語である場合、その要素語それぞれはすでに登録されていることが多いので、翻訳上、意外な「副作用」を生じさせる恐れがあるからである。従って、個別の要素語や、それらを組み合わせた語句について、その登録状況や訳語を一通り点検しておく必要がある。つまり、新語の登録の際には、その語の「周辺」についても一定の調査を要する。このことは、新語に対する訳語の選定にあたっても重要である。

ここに、連鎖検索を効率的に実行し、結果を見やすく表示するようなシステムが提案される。こうしたシステムを用意しておけば、新語の検定の精度が向上し、誤りの混入も未然に防止されるなど、作業の全般的効率化に資するところ大である考えられる。

(5) 辞書更新のセンター体制

新出専門用語の検出、検定には、上述のような辞書データベース・システムの道具立てと、実際に作業にあたる専門家を要する。従って、これを本格的に実施してゆくということは、個々の機械翻訳システム・ユーザーにおける、ユーザー辞書の整備というような枠組みには、到底おさまりきらないであろう。一方、機械翻訳システムの今後の普及を見通すと、最新の専門用語辞書の重要性に対する認識はますます高まると思われる。従って、機械翻訳システムの供給者にとっては、常に最新の専門用語辞書を提供できるという体制の整備が、今後における重大な課題になるであろう。

これらを勘案すれば、上述のような辞書データベース・システムと専門要員を配した、辞書更新センター的なものが構想される。ここでは、和文と英文の文献における用語の出現状況を、適切な統計を採取しながら常時監視し、十分な検定作業の上で、新出用語を認定・登録したのち、これをユーザーに経常的に配布してゆくといった体制が想定される。機械翻訳システムの今後の普及・発展に向けては、翻訳システム開発・供給者において、このような辞書更新システムの体制整備が図られることが望まれよう。

7. 2 機械翻訳システムとの連動性

新出専門語の検出と辞書の更新に対しては、機械翻訳システムそのものを、こうした作業に応用するという視点もある。本調査研究では、専門用語辞書の整備を、機械翻訳システムの運用とは一応独立の工程として運用することを念頭に検討してきた。そこで次に、このような辞書更新システムによる更新結果を、機械翻訳システム用の辞書に反映させてゆく際の問題点について検討を加えておく。

(1) 訳語の優先度

本調査研究において開発されたような、英和・和英両引き辞書は、専門分野辞書として、新熟語の登録、更新が容易になるということで利点が大である。反面、これを機械翻訳用辞書に持ち込むことを考えると、複数訳語が存在する用語については、訳語の適用優先度の指定に関して問題を生じる。

たとえば、情報分野で、和英では、

経路 — route, path
算法 — algorithm, calculus
競合 — conflict, race

であるが、英和では、

channel — 通信路、伝送路、チャンネル
circuit — 回路、回線
normal form — 正規形、標準形

などとなる。さらに英和では、

node — 節点、ノード
tree — 木、トリー
pattern — 図形、パターン

など、英語をそのままカタカナ書きした和訳も同列に扱えば、膨大な数の複数訳語をもつ専門用語が存在することになる。

この場合、新たに複数訳の方の用語が出現した場合も、両引き辞書では独立に辞書登録されることになる。これが、そのまま英和翻訳用、和英翻訳用辞書に編成されるとすると、2個以上の訳語が、特別な優先度をもたずに記載され、翻訳システムでは、多くの場合、先頭の訳語から順に使用されることになる。これで問題のない場合もあるが、本来なら優先度が低いはずの訳語が優先されて、不適切な訳語選択となる場合も多くなる。

このような、両引き辞書による専門用語の管理に伴う問題を解決するには、例えば、次のような方法があろう。すなわち、ある新用語（たとえば、calculus）を登録するとき、その訳語（たとえば、算法）に対する元言語の他の訳語（たとえば、

algorithm) が既に登録されているとき、その事実を登録者に知らせて、訳語としての優先順位を付してもらうようにする(たとえば、1. algorithm、2. calculusのように)。この場合、登録者に広い知識が必要とされることや、現在登録しようとする用語の方が身近であることが多く、幾分これを優先する傾向になる等の課題は残るであろうが、両引き辞書による専門用語管理に伴う問題を軽減することになると思われる。

(2) 専門用語辞書と一般用語辞書の優先度

訳語の優先度は、これを敷衍すると、専門用語辞書と一般用語辞書の使い分けの時にも生じる重要な問題である。通常、専門用語辞書の方が優先して使用されるので、いわゆる「訳し過ぎ」といった問題が生じる。情報分野を例にとると、「意味」は、専門的には「semantics」であることが多いのであるが、すべての「意味」をこのように訳したのでは「訳し過ぎ」が生じる。一般用語の「meaning」とすべき箇所もあるからである。また、「term」は、専門的に「項」とすべき場合もあるが、一般用語の「用語」とすべき箇所も多いであろう。

このような訳語の使い分けは、一般には、文脈を理解し、意味上の制約を考慮して、より適切な訳語の選択を行う必要がある。しかしながら、文脈理解や意味理解は、自然言語理解技術において未解決の困難な問題であり、実用的に利用できるレベルには到達していない。当面の対処法としては、かな漢字変換における、漢字の優先度の自動的変更方式のように、対話的機械翻訳、あるいは機械翻訳の後編集時において、ある箇所でユーザが「意味」の訳語として「meaning」を選択したら、以後はこの部分の専門用語辞書の優先度を、一般用語辞書より低くするといった方法が考えられている。

8. 専門用語における和訳語、英訳語の特質と問題点

情報流通の国際化に呼応し、また新生術語に迅速に対応するなどの面から、「英→和」または「和→英」という、一方向の翻訳にのみ着目した辞書の管理や更新は、きわめて非効率的であって、当初から、この双方向からの利用を考慮した、いわゆる「両引き辞書」を生成してゆくのが有効であることは、すでに検討したところである。また、専門用語辞書の恒常的な更新・拡張にあたっては、技術の進展にともなっていて続々と生み出される、新生術語に対する適切な対応が、きわめて重要であることも、すでに述べた通りである。そこで本章では、以上において行った調査分析の結果を踏まえつつ、先端的な専門用語の発生メカニズムと、その国際間での流通・浸透の仕組みなどについて分析・解明を試みた上で、そうした状況に有効に対応できる機械翻訳システムと、そのための専門用語辞書更新システムのあり方について、検討を加えることにする。

8. 1 和訳語形成の歴史的経過と訳し分け

現在すでに術語となっている用語も、かつて、それが誕生した時点では、新生術語であったという事実を踏まえると、既存の術語そのものの発生過程や、現在の状況について把握することが、今後発生するであろう新生術語の誕生のメカニズムを探る上で、きわめて有効であることはいうを待たないであろう。この観点から、今回行った調査・分析結果を詳細に検討してみる。

今回行った調査は、きわめて広範囲に及ぶものであるが、その中には、新生術語発生のメカニズムを探る上で、参考になる結果も多い。まず、今回行った【「英→和」および「和→英」訳し分けの分野別状況の調査】（表4-1-3、図4-1-4、表4-1-5及び図4-1-6）においては、各専門分野とも、ある1つの英語術語に対する和訳語の数、および1つの日本語術語に対する英訳語の数、すなわち訳し分け指数は、ほぼ1に近い値となっている。これは一見すると、英語術語と日本語術語との対等性を示しているかのように思われる。

しかし厳密に分析してみると、かならずしもそういう結論には至らない。すなわち、自然科学や、それを基礎とした近代技術は、基本的には、まず初期において欧米で発展し、それが、後代になって、日本に輸入されてきているという経緯を、ほとんど例外なく（どの分野でも）辿っているという事実は明白である。もちろん、その後日本における科学技術の進展はまことに目覚ましく、すでにいくつかの分野においては、わが国が世界の最先端を進んでいるといえる。

しかし、上記のような科学技術の発展過程からすると、現在すでに術語として定着し、かつ長期に亘って使用されている用語の大部分は、元来、英語術語として導入され、それに、わが国において適当な日本語の訳がつけられたものが、圧倒的に多いと

思われる。この点が、新生術語の発生機構を考える上での、第1の重要なポイントであると考えられる。

そこで、きわめて単純な考察すれば、もしそうであるなら、1つの英語術語に対して、もっと沢山の和訳語が対応しているはずである。これは、その術語の導入のされ方、分野及び翻訳者の多様性に依じて、1個の英語術語に対応して、多様な和訳語が付けられると考えられるからである。しかし、現実には、上の調査のような結果が得られている。これは、次のようなことを示しているものと思われる。

すなわち、欧米からの科学技術が輸入されて以来、きわめて多くの年月がたち、多くの分野では、日本も同等に、あるいは日本の方がより発展を遂げているという状況になっている。その間に、これらの術語は十分に活用・習熟され尽くして、その日本語訳語の意味、またはその使用分野が拡大し、それが、また国際流通の活発化に伴って、異なった英語訳語として輸出されたか、または、異なった英語術語と新たに対応付けられていった（すなわち、異なった英語術語の訳語として普及した）ものと考えられる。

なお、上にあげた2種の経過では、後者のほうが圧倒的に多いものと思われる。これはまた、4.3の訳語検索の連鎖による類義語クラスタの析出と、辞書のシソーラス化のところでもみたように、同義語や類義語及び訳し分けなどの結合関係に基づいて、多数の用語が絡み合っている状況と同義と考えてもよい。結局、それらのために、英訳語、和訳語の訳し分け指数が、たまたま1に近い値になっているのではあるまいか。したがって、これらに対しては、同じく4.3でも触れたように、いくつかの和英の用語を起点として、その関連用語を手繰り寄せていくという方法が考えられる。

8.2 適用分野と訳し分けの関連

上記のことは、また「適用分野数と訳し分けの特異事例（和→英）」（表4-2-3）の調査例からも、ある程度うかがえる。ここでは、1つの日本語術語に対して、多分野にわたって、多数の英訳語がある例として、「層」、「支柱」などの単語があがっている。統計上、特異事例ということになっているが、いずれも、きわめて広い専門分野にわたって、似たような構造・形態を表わす基本的な術語として用いられるものであり、そういう意味では特異な用語ではない。

これらは、その意味からして、最初は英語術語から出発し、その訳語として用いられているうちに、広い分野で導入された、多くの異なった（ただし、似たような構造や形態を表わす）英語術語の訳語として用いられるようになってきたと考えられる。ちなみに、これらの日本語術語に対する英訳語の数は多いが、それらは、基本的に同じ意味を持つか、または似た構造・形態を持つものを表現する言葉がほとんどである。

これはまた、英→和における多分野・多訳語の特異事例と対照してみると興味深い。この事例には、carrier、columnなどがあるが、これらの術語は、必ずしも、日本語の実例として上にのべた「層」や「支柱」ほど一般的・基本的な言葉ではない。また、その和訳語も、carrierの場合、「電荷担体」、「通信路」、「航空母艦」、「保菌者」など、きわめて異なった広い範囲に亘る意味を持った言葉になっていることがわかる。これは、最初の術語の発生が、英語（欧米語）としてであったという事実を、明確に示すものと思われる。

この考察はまた、今回の調査における「英訳語の適用分野分化の状況」（表4-2-1）および「和訳語の適用分野分化の状況」（表4-2-2）の比較・対照によっても支持されると考えられる。これらの表は、いくつかの興味ある事実を示している。まず際立った特徴は、1つの分野で1つの訳語を持つ術語が、英語、日本語とも圧倒的に多いということであり、これは、有効な専門辞書電子化の可能性と、その比較的容易さを示すものである。しかし、ここではそれらの議論には立ち入らず、術語の発生由来という観点から、これらの表を検討することにする。

これらの表は、いずれも、マトリックスの右・上半分にのみ数値が分布しているのが大きな特徴であり、これは英語術語、日本語術語の場合とも共通である。すなわち、多くの分野にわたっていくつかの訳語を持つ場合のほうが、少数の分野で数多くの訳語を持つ場合よりも多いことを示している。しかし細かく見てみると、マトリックスの対角線の右上の部分、とくに中央よりやや右よりの部分で、和訳語の訳語数が、英訳語のそれより大きくなっていることがわかる。これは、おそらく上で述べたように、いくつかの分野に共通な、基本的な英語術語が導入され、それに、それぞれ分野別に（または同一分野内で）異なった和訳語がつけられた事情を示しているのではあるまいか。すなわち、発生的には、英語術語が先行していることの反映と考えられる。以上、従来から存在する既成の術語を、その発生経過という見地から分析すると、上記のようなことがいえる。

8. 3 日本起源の英語の新生術語

しかし、最近では、わが国における科学技術の著しい発展によって、日本においても、新たな科学技術用語が生まれてきている。その発生状況を考察してみると、次のようなことがいえる。

まずその第1点は、日本で新たに発生した新生術語は、日本語とは限らず、むしろ、英語術語として誕生するケースが多いといっても、過言ではないということである。それには次のような背景がある。

現在、わが国にはきわめて多くの学会、協会が組織されていて、活発に活動しているが、そこで発行されている学術雑誌（オリジナル論文が掲載されているもの）には、日本語論文のみを掲載するもの（例えば、情報処理学会、建築学会、その他多

数)、英語論文のみを掲載するもの(代表的なものは、応用物理学会)、およびその両者を掲載するものがある。それぞれ専門分野によって、国際性の強さは非常に異なるのであるが、上記のような英文学術雑誌の状況は、これに密接に関連している。

例えば、エレクトロニクスの分野は、いうまでもなく日本は、きわめて高いレベルをキープしており、また、非常に国際的な分野として知られている。とくにその中でも、マイクロエレクトロニクス、オプトエレクトロニクスおよびそれらの基礎となる材料分野においては、日本国内だけでの論文発表・研究発表では、あまり高く評価されないという事情がある。上に述べたように、これら分野の論文が多く発表される応用物理学会の学会誌が、英文であるのも、それらの理由が大きい。

こういう分野においては、新たに誕生する技術用語も、それが日本で発生したにも拘らず、ほとんど英語であることは、容易に理解できるであろう。マイクロエレクトロニクスの分野において、最近新たに誕生して、世界的にも用いられるようになった術語としては、例えば、

high electron mobility transistor (略語: H E M T)、
trench capacitor、
reactive ion etching (略称: R I E)、
separation by implanted oxygen (略語: S I M O X)

などが上げられる。

これらは、いずれも、わが国において発明・開発された技術であり、同時に、その際の英語による命名が、ほぼそのまま新生術語として海外にも輸出され、いまや、国際的な術語として定着している。このほかにも、まだ上記の例ほどには定着していない和製の英語術語は多く、半導体分野においては、新たに開発した技術や素子の名前を、英語で命名する場合は圧倒的に多いことが知られている。これは、そうしないと、その研究成果が国際的に通用しないからであるし、また、欧米諸国も、日本の高い技術に注目しており、そういう方面からの要求も、背景として存在するためであろう。

当然ながら、これら和製の英語術語に対する和訳語も、日本においてつけられる。例えば上掲の用語に対しては、それぞれ「高電子移動度トランジスタ」、「溝型キャパシタ」、「反応性イオン・エッチング」、「サイモックス」などの和訳語が当てられている。

このように、国内で英語形で発生し、同時に、それに対応する和訳語もできるような用語の場合には、訳語の統一化が比較的容易で、かつ意味も明確である。すなわち、1つの英語術語に対して、1つの和訳語が対応するという、自動翻訳にとっては有利な状況になりやすいと思われる。分野にもよるであろうが、このような事例は、今後さらに増大するのではなからうか。

一方、日本語術語として新たに発生する用語もあろうが、後に、これに英訳語が与えられて、国際的に普及、定着したという例は少ないようである。元来、日本語で表

されるわが国独特の術語には、伝統的技術を背景にしたものが多いと考えられる。例えば、建築・土木などの分野をみると、欧米と日本では生活様式、建築様式、周囲の環境条件や立地条件などの相違があり、そうした伝統、文化の差異に対応して、学問上の基本的な考え方も相違し、従って、用語にも差異が生じることになるのは当然である。そのため、欧米独特の、の、また日本独特の用語が存在することになる。

このような事情に鑑みて、日本語術語がまず誕生し、それに後から英訳語が付けられたという事例は、最初から英語として発生した術語に比較して、ごく少数ではないかと考えられる。これらについては、さらに詳細に調査する必要がある。

8. 4 新出術語と専門辞書更新システム

(1) 英語新生術語の重要性

上記の分析より、応用物理、電気電子（エレクトロニクス関係）分野などのように、国際性が強く、国内での論文発表も英語で行われているような分野においては、専門辞書を常時更新・拡張しつつ維持・管理していくために、まず、英語の新生術語に重点的に注目して、その早期抽出につとめ、さらに、その日本語訳語を早急に準備していくことが、最も重要かつ効率的であることが分かる。英→日、和→英の両引き辞書は、これに基づいて、整備・拡張してゆくのがよいと思われる。

一方、論文が主として日本語で出版される、情報処理、建築・土木などの分野においては、日本語の新生術語についても注意を払う必要がある。この場合、各分野の実情に応じて、例えば過去数年における、そのような日本語の新生術語の発生頻度を調査した上で、専門辞書の更新・拡張のための重点的監視領域を設定してゆくのも、ひとつの方法であると思われる。

(2) 複合語、略語の重要性

新生術語は、多くの場合、新しい技術の考案・開発に伴って生まれてくると考えられる。新しく開発される技術のうちには、従来には無かった、まったく新しい原理やアイデアに基づくものもあるであろう。しかし、実際には、従来技術、あるいは少なくともこれまでに知られていた技術や、既知の原理の新たな組み合わせによって生まれるものが、圧倒的に多いと考えるべきであろう。

このような状況を、専門用語の側面からみれば、上の例もそうであったように、ほとんどの新生術語は、複合語として誕生するものと思われる。新原理や、従来まったく無かった基本的アイデアに基づいて生まれた技術または素子などは、単一の用語（例えばトランジスタ、レーザなど）で表わされる場合が多いのであるが、このような例はきわめて限定されるということからして、新生術語としては、複合語に重点的

に注意すべきであるという結論になろう。同時に、これらの複合語は、いずれも比較的複雑な表現となっているので、多くの場合、それを略語（上例のHEMT、RIE、SIMOXなど）で用いる場合が多くなっている。したがって、略語で表わされる新生術語にも、注意してゆくべきであろう。

（3） 専門用語の動向

翻訳とは結局、基本的に1つの文化的所為であるから、翻訳用辞書は、単なる言葉の書き換え表では済まない。したがって、これを生きた辞書とするには、各国の文化的背景への理解も必要であることは論を待たない。例えば、ヨーロッパ諸国においては、ギリシャ哲学、さらには後代のドイツ観念論を代表するカント哲学、あるいはデカルトの哲学などの伝統に基づき、古来より、基礎的でアカデミックな学問体系が重んじられ、その応用である技術分野は、どちらかといえば、ややレベルの低い、極端に表現すれば、卑しいものとされる風潮があった。これは、ヨーロッパから強い影響を受けた日本においても、ある時期には見られた傾向であった。そのため、重厚なあるいは重厚な印象を伴う術語が、中心的な術語群を形成するという面が見られる。

これに対して、比較的歴史の新しいアメリカにおいては、その国家の歴史的発展の形態からして、基礎的、アカデミックな学問体系よりも、どちらかといえば、その応用分野である技術分野の方を重視するという傾向が、明らかに存在する。最近の日本においても、とくに終戦後において、ヨーロッパよりもアメリカの影響をより強く受けた結果、学問的な原理よりも、それを応用する技術分野を重視する傾向が顕著に現れている。

一方、技術の著しい発展に伴って、一部の分野では、従来のように、アカデミックな学問体系がまず形成され、ついで、その応用である技術分野が発展していくという傾向とは異なって、応用面からの研究開発が先行し、その基礎的な体系付けが、後から行われてゆくという場合が、しばしば見られるようになってきている。これからの専門用語の動向についても、こうした背景を抜きにしては考えられないであろう。このように、自動翻訳のための専門辞書の有効な更新・拡充・維持管理システムの開発と拡充に向けては、今後とも、多方面からの分析・調査・考察が継続される必要があると考えられる。

9. 今後の展望

機械翻訳システムは、いよいよ本格的な実用・普及期を迎えたとみられる。従来、多くの関心は、どちらかといえば、翻訳システムのいわば本体部分である、翻訳手法と一般用語辞書に向かっており、専門用語辞書には、付加的なものとしての位置付けしかあたえられていなかったといえよう。しかし、翻訳システムを、一旦実用に供するとすると、その翻訳の精度は、専門辞書の充実度合いに大いに依存することが分かってくる。

専門的用語は、基本的に名詞句であり、その翻訳は、文法的処理を要する本来的な「翻訳」ではなく、単純な「置換」の問題に過ぎないと、これまでは考えられがちであった。従って、専門辞書の充実といっても、それは、単純な用語の登録作業としてしか認識されていなかったといえよう。幸いにして、翻訳システムの実用化の進展に伴って、こうした観念的な発想は払拭されつつあるように見える。

本調査研究は、このような背景において、機械翻訳システムのための専門用語辞書の更新方式を中心的な研究課題として発足したものである。平成3年度においては、英日翻訳を対象として、英文専門用語の出現状況を主体に調査研究を実施した。平成4年度においては、前年度研究の過程で得られた知見に立脚し、英日辞書と日英辞書を、同時平行的なものにとらえて、両引き辞書を編成して、和文専門用語の語彙統計調査や、英文用語との比較等を実施した。さらに、これらの結果に基づいて、専門用語辞書の効果的な更新に向けて、辞書のデータベース化と、その更新のシステム化について検討を加えた。

機械化された両引き専門用語辞書による分析というのは、これまでに例のないものと思われる。そこには、日英の専門語における訳し分け状況の比較、日英の訳語を連鎖的に検索して得られる、用語の群生状況調査などが含まれる。また、これらに基づいて、わが国と外国における、専門用語の発生や、翻訳をとおした用語の国際流通の機序に関わる考察も行われた。こうした検討の結果を一口に要約すれば、それは、技術進歩の激しい今日、専門用語辞書の整備や更新については、組織的、体制的、専門的な取り組みが必要であるということになるであろう。

専門用語の編成・更新事業は、今後、翻訳システム開発・供給者等において本格化が期待される。ここに報告された各種の統計的データやその分析結果が、そうした事業の構成、態様に示唆を与え、機械翻訳システムの実用化の促進に資するものとなれば幸いである。ところで、本調査研究の過程において、さらに解明すべき点も多く発見されている。これらを含めて、専門用語辞書に関する調査研究は、機械翻訳システムの実用化の進展に従って、今後各方面においてさらに活発に展開される必要があるであろう。

付属資料1 統合分野分類・原辞書分野分類

(1)

統合分野	原辞書分野名					
	J I S	語数	日刊	語数	学術	語数
機械	機械振動・衝撃用語	388	機械工学	3,309	機械	12,307
	火力発電用語	478	極低温学	101		
	往復動内燃機関用語	1,602	工学	5,757		
	円すい用語	41	時計学	127		
	研磨材、研削といし及び研磨	40	設計工学	1,379		
	布紙の用語及び記号		熱力学	614		
	巻上機用語	350	兵器	2,161		
	圧力容器の構造共通用語	298	力学	1,458		
	加工方法記号	380				
	家庭用ミシン用語	379				
	家庭用縫機用語	87				
	家庭用燃焼機器用語	584				
	金銭登録機用語	126				
	公害防止装置用語	102				
	ねじ加工工具用語	470				
	ねじ用語	544				
	C A D用語	55				
	工業用へり縫いミシン頭部品用語	807				
	工業用燃焼装置用語	340				
	工作機械の操作表示記号	190				
	工作機械の名称に関する用語	687				
	ばね用語	157				
	リーマ用語	146				
	トラックの普通荷台に関する用語	130				
	エンジン駆動発電セット用語	146				
	ドリル用語	145				
	歯車用語	293				
	歯切工具用語	377				
	数値制御工作機械用語	270				
	クレーン用語	226				
	クラッチ及びブレーキ用語	87				
	事務機械の名称に関する用語	155				
	水車及びポンプ水車用語	512				
	バルブ用語	156				
	バイト用語	203				
	送風機・圧縮機用語	760				
バッキン及びガスケット用語	227					
コンベヤ用語	220					
フライス用語	309					
フォークリフトトラック用語	188					

付属資料 1 統合分野分類・原辞書分野分類

(2)

統合分野	原辞書分野名					
	J I S	語数	日刊	語数	学術	語数
機械	ブローチ	180				
	プレス機械用語	272				
	プラスチック加工機械用語	82				
	プロッタ用語	33				
	切削工具用語	124				
	ターボポンプ用語	651				
	真空用語	249				
	自走クレーン用語	126				
	自動車の種類に関する用語	1,532				
	自動車排出物質の公害防止関連用語	261				
	自転車用語	219				
	鉄鋼製管継手用語	72				
	鉄道車両用語	1,274				
	福祉関連機器用語	700				
	転がり軸受用語	384				
	複写機用語	121				
	木材加工機械の名称に関する用語	217				
	無人搬送車類の用語	155				
	油圧及び空気圧用語	469				
	溶接用語	893				
流体素子用語	176					
48,058		20,845		14,906		12,307
応用化学	火薬用語	530	化学工学	1,100		
	化学繊維機械用語	172	繊維	808		
	繊維原料用語	236	食品工学	201		
	繊維織物用語	332	石油工学	2,086		
	繊維試験用語	371	生化学	438		
	繊維雑品用語	169	分析化学	611		
	繊維糸用語	147	分子生物学	98		
	繊維製織用語	108	無機化学	2,139		
	繊維製品用語	131	有機化学	8,658		
	繊維染色用語	310				
	レース用語	175				
	メリヤス用語	138				
	石炭利用技術用語	407				
	ゴム用語	604				
	紙・バルブ用語	665				
	プラスチック用語	810				
	製織機械用語	423				

付属資料1 統合分野分類・原辞書分野分類

(3)

統合分野	原辞書分野名					
	J I S	語数	日刊	語数	学術	語数
応用化学	染色加工機械用語	275				
	編組機械	227				
	塗料用語	757				
	分析化学用語	977				
	紡績用語,紡績機械用語	722				
	防せい防食用語	170				
24,995		8,856		16,139		
建築	建築モジュール用語	21	建築	357	建築	6,227
	建築用内外装材料用語	115	建築構造	673		
	せっこう, 石炭及びマグネシアセメント用語	180				
	コンクリート用語	190				
	ショベル系掘削機用語	524				
	8,287		1,030		1,030	6,227
応用物理	原子力用語	1,167	結晶学	454	原子力	4,672
	光学用語	792	固体物理学	483		
	フェライト磁心通則	84	光学	1,764		
	分析化学用語	520		1,726		
			統計力学	160		
			分光学	400		
		ニウレトラス (核工学)	1,092			
13,314		2,563		6,079	4,672	
化学	界面活性剤用語	304	化学	1,301	化学	10,981
	芳香族製品及びタール製品用語	196	物理化学	767	分光	2,409
15,958		500		2,068	13,390	
航空宇宙	航空用語	313	宇宙物理学	166	航空	4,244
		270	航空宇宙工学	1,275		
		269	航行学 (術)	2,126		
		200	流体力学	853		
		131				
		83				
9,930		1,266		4,420	4,244	
船舶海洋	船舶救命及び消火設備の図記号	147	海洋学	829	海洋	2,675
	船舶通風系統図記号	31	造船工学	835	船舶	10,088
	船用電気照明図記号	155				
	船用電気通信図記号	269				
	船用電気動力図記号	203				
	舟艇用語	64				
	造船用語	3,690				
18,986		4,559		1,664	12,763	

付属資料1 統合分野分類・原辞書分野分類

(4)

統合分野	原辞書分野名					
	J I S	語数	日刊	語数	學術	語数
材料資源	形状記憶合金用語	61	鉱山学 (採鉱学)	2,135	採鉱冶金	5,070
	鉱山用語	1,377	鉱物学	2,545		
	アルミニウム表面処理用語	315	材料学	3,175		
	ガラス繊維用語	220	組織学	440		
	耐火物	314	宝石学	56		
	水素吸蔵合金用語	96	冶金学	2,596		
	ボーリング用機械・器具用語	159				
	制振材料用語	125				
	鍛造加工用語	277				
	電気めっき用語	185				
	鉄鋼用語	1,090				
	粉末冶金用語	236				
	20,472	4,455		10,947		5,070
生物	バイオテクノロジー用語	545	古生植物学	69	遺伝	2,119
			古生物学	693	植物	10,673
			菌類学	343		
			遺伝学	377		
			細胞学	381		
			植物学	2,569		
			生態学	782		
			生物学	2,218		
			生物物理学	12		
			生理学	911		
			分類学	78		
			発生学	469		
			微生物物理学	986		
	ワイル学	102				
23,327	545		9,990		12,792	
数学			数学	3,038	数学	1,912
			統計学	356	統計数学	592
5,898			3,394		2,504	
動物			解剖学	1,882	動物	2,611
			脊椎動物学	2,619		2,611
			獣医学	186		
			進化	52		
			人類学	196		
			動物学	379		
	無脊椎動物学	3,712				
11,637			9,026			

付属資料1 統合分野分類・原辞書分野分類

(5)

統合分野	原辞書分野名					
	J I S	語数	日刊	語数	学術	語数
情報	情報処理用語 電子通信用語	3,163 407	情報処理 通信工学 電磁気学 システム工学	4,391 2,054 1,968 55		
12,038		3,570		8,468		
電気電子通信	計測用語 音響機器用語 音響用語 医用電気機器用語 工業プロセス計測制御用語 がいし及びプッシング用語 レーザ安全用語 パルス用語 ヒューズ用語 プリント配線版用語 集積回路用語 産業用ロボット用語 自動制御用語 照明用語 電子管用語 電子機器用受動部品用語 電子測定器用語 鉄道信号保安用語 日本語ワードプロセッサ用語	237 251 211 313 685 136 89 348 97 179 101 247 64 717 642 485 213 289 120	音響学 制御システム 電気工学 エレクトロニクス(電子工学)	409 2,020 3,148 5,366	計測 電気	2,744 11,358
30,469		5,424		10,943		14,102
地球科学			気候学 気象学 岩石学 測地学 地球科学 地球物理学 地質学	192 194 1,089 145 82 688 6,387	気象 地震	12,303 2,656
23,736				8,777		14,959
都市工学	公害防止装置用語 コンタミネーションコントロール用 自動車排出物質の公害防止関連用語	102 115 261	建築 水文学	357 986	建築	6,227
8,048		478		1,343		6,227

付属資料1 統合分野分類・原辞書分野分類

(6)

統合分野	原辞書分野名					
	J I S	語数	日刊	語数	学術	語数
土木	せっこう, 石炭及びマグネシアセメ	180	土木工学	1,385	土木	7,050
	コンクリート用語	190				
	ショベル系掘削機用語	524				
	電車路線用金具用語	90				
	鉄道線路用語	235				
	鉄道分岐器類用語	229				
9,883	1,448		1,385		7,050	
農業					農学	17,414
17,414						17,414
物理			核物理学	513	物理	3,986
			原子物理学	253		
			プラズマ物理学	102		
			相対性理論	123		
			素粒子物理学	252		
			物理学	1,563		
			量子力学	354		
7,146			3,160		3,986	
林業			植物病理学	334		
			農学	171		
			林業	67		
572				572		

付属資料2 大学の学科・講座の分野事例(1)

- 〈機械〉
- ・機械工学科 機械材料学, 機械実験工学・機械検査法, 機械設計学, 機械熱学, 機械力学, 往復機械, 加工生産学, 個体力学, 回転機械, 材料力学, 潤滑油圧工学, プラント工学, 切削加工, 塑性工学, 動力工学, 熱力学, 熱流動工学, 燃焼工学, 伝熱工学, 噴射機械, 非切削加工, 流学, 流体工学
 - ・機械情報工学科
 - ・産業機械工学科 高分子工学, 構造安全工学, システム設計工学, 生産システム工学, 設備装置工学, 輸送現象学
 - ・精密機械工学科
 - ・精密工学科 科学機器, 機械要素, 計算機制御, 振動工学, システム工学, 制御工学, 生産工学, 精密機械材料学, 精密加工学, 精密工作, 精密測定学, 特殊加工
 - ・船用機械工学科
- 〈建築〉
- ・建築学科 建築意匠学, 建築計画, 建築環境学, 建築構造学, 建築史, 建築施工, 建築設備建築環境調整学, 建築基礎工学, 建築施設計画, 地域生活空間計画, 鉄筋コンクリート構造学, 鉄骨構造学
 - ・建築工学科 建築計画学, 建築環境工学, 建築構造学, 建築人間工学, 建築施工学, 建築力学
 - ・衛生工学科 衛生設備学, 環境衛生学, 水質工学, 水道工学, 産業衛生工学, 放射線衛生工学
- 〈応用化学〉
- ・応用化学科 工業物理化学, 石炭化学・石炭工業有機工業化学・油脂化学, 電気化学工業物理化学, 天然高分子化学, 無機工業化学, 有機工業化学・油脂化学
 - ・応用精密化学科 機能高分子化学, 精密合成化学, 精密資源化学, 接触反応化学, 分子設計化学, 無機精密化学

付属資料2 大学の学科・講座の分野事例(2)

- 〈応用化学〉
- ・化学工学科 機械系単位操作, 拡散系単位操作, 化学工学基礎理論, 化学工学熱力学, 化学プロセス工学, 移動現象論, 触媒工学・表面工学, 装置工学, 装置制御工学, 生物化学工学, 分子工学, 分離工学, 反応工学, 反応システム工学, 輸送現象論
 - ・合成化学科 原料製造化学, 合成高分子化学, 合成反応化学・触媒化学, 高压化学, 重合化学, 生合成化学, 物理有機化学, 有機金属化学, 有機合成化学, 有機材質化学, 有機接触化学, 遊離基合成化学
 - ・高分子化学科 基礎高分子化学, 高分子合成, 高分子構造, 高分子材料化学, 高分子物性, 高分子分子論, 高分子力学, 放射線高分子化学
 - ・工業化学科 応用固体化学, 工業生化学, 工業電気化学, 工業物理化学, 工業分析化学, 天然物有機化学, 無機構造化学, 有機反応化学
 - ・食品工学科 栄養化学, 酵素化学, 食品化学, 微生物生産学
 - ・石油化学科 基礎炭化水素化学, 触媒化学, 触媒工学, 触媒物理学, 石油化学加工学, 石油変換工学, 炭化水素物理化学
- 〈応用物理〉
- ・応用物理学科 応用光学・応用分光学, 応用物性学・放射線工学, 計測学, 分析機器
 - ・原子核工学科 核燃料工学, 原子核機器学, 原子核反応工学, 原子炉工学, 原子炉材料学, 同位体工学
 - ・原子力工学科 核燃料工学, 原子核機器学, 原子核化学工学, 原子炉工学, 原子炉材料学, 原子炉物理学
 - ・物性物理工学科 結晶物理学, 固体理論, 金属物性, 光物性, 高压物理, 磁性物理, 統計物理学, 低温物理
 - ・物理工学科 材料強度学, 材料物性学, 熱流体物性学, 物性分光學, 量子物性学

付属資料2 大学の学科・講座の分野事例(3)

〈化学〉

- ・化学科 金相学, 金属物性学, 構造化学, 構造有機化学, 集合有機分子機能, 生物化学, 天然物有機化学, 物性物理化学, 物理化学, 分光化学, 分析化学, 反応物理化学, 放射化学, 放射線化学, 無機化学, 量子化学, 有機化学, 有機合成化学, 有機生物化学, 有機反応構造論
- ・高分子学科 高分子化学構造論, 高分子固体構造論, 高分子合成化学, 高分子物理学, 高分子溶液論

〈航空宇宙〉

- ・宇宙・地球科学科 基礎宇宙学
- ・宇宙物理学科 宇宙物理学, 銀河物理学
- ・航空学科
- ・航空工学科 空気力学, 航空器構造学, 構造強度学, 推進工学, 振動学, 流体力学
- ・天文学科

〈材料資源〉

- ・金属加工学科 結晶塑性学, 金属繊維学, 金属物理学, 鋳造加工学, 特殊鋼学, 溶液工学
- ・金属工学科
- ・材料開発工学科 界面制御工学, 環境材料工学, 細晶成長工学, 材料精製工学, 材料物理化学, 反応制御工学
- ・材料学科
- ・材料物性工学科 格子欠陥物理学, 構成材料学, 材料組織学, 材料塑性工学, 材料物性学, 粉体物性工学
- ・資源開発工学科
- ・資源工学科 開発工学, 応用計測学, 応用地質学, 加工設備学, 精製工学, 探査工学
- ・生産加工工学科 加工基礎学, 加工設計学, 固相加工工学, 材料構造強度学, 信頼性評価工学, 生産加工プロセス・機器工学, 生産加工システム工学, 融体加工学
- ・冶金学科 金属材料学, 電気冶金学, 鋳造冶金学, 鉄冶金学, 非鉄冶金学, 冶金反応及び操作

付属資料2 大学の学科・講座の分野事例(4)

- 〈船舶海洋〉 ・ 船舶海洋工学科，海洋利用計画学，運動制御学，船舶海洋構造力学，船舶海洋材料学，船舶海洋流体力学
- 〈数学〉 ・ 応用物理学科 応用数学・計算機械，計算数学
 ・ 数学科 位相数学，確立統計数学，応用数学，計画数学，関数解析学，幾何学，幾何学，整数論，数学解析学，数学解析，代数学，情報数学，統計数学，複素解析学，微分方程式論
 ・ 数理工学科 応用数学，応用力学，計画工学，制御理論，論性システム
- 〈水産〉 ・ 水産学科 水産化学，水産生物学，水産物理学，水産微生物学
- 〈生物〉 ・ 応用生物工学科 酵素工学，細胞工学，生態工学，生物化学工学，生物材料化学，生物資源工学，生物情報工学，醗酵システム工学
 ・ 植物学科 一般細胞学，植物環境応答機構解析学，植物生理・生態学，植物分類学，分子植物学
 ・ 生物学 遺伝学，細胞生理学，生物物理化学，比較生理学，微生物学，放射生物学
 ・ 生物学科
 ・ 生物化学科
 ・ 生物工学科 高次情報制御学，刺激生理学，生体計測工学，生体エネルギー学，生体物性学，分子生物学
 ・ 生物物理学科 形質発現学，原形質物性学，生態高分子構造学，生態高分子反応学，分子生物学，量子生物学，理論生物物理学

付属資料2 大学の学科・講座の分野事例(5)

〈情報〉

- ・情報科学科
- ・情報工学科

言語処理工学, 計算機言語学, 計算機援用工学, 計算機構学, 計算機システム, 計算機ソフトウェア, ソフトウェア構成学, 情報基礎論, 情報解析学, 情報工学基礎, 情報処理, 情報ネットワーク学, 情報システム工学, 知識ベース工学, 知識情報処理学, 論理回路
- ・情報システム工学科

計画情報システム工学, 集積システム工学, 情報システム基礎論, 情報システム構成学, 知識システム工学

〈動物〉

- ・獣医学科
- ・動物学科

人類進化論, 自然人類学, 動物系統・遺伝学, 動物生理・生態学, 発生生物学, 放射線生物学

〈電気電子通信〉

- ・制御工学科

計測機器, 計測学, 制御機器, 制御系構成学, 制御工学基礎理論
- ・通信工学科

通信工学基礎論, 通信システム工学, 通信方式論, 通信網工学, 電磁波工学
- ・電気工学科

計測制御工学, 組織工学, 制御工学, 電気機器, 電気応用工学, 電気回路学, 電気工学基礎理論, 電気工学基礎論, 電気材料工学, 電気磁気学, 電気物性工学, 電子工学基礎理論, 電力工学, 発送配電工学, 放電工学, エネルギー変換機器, 自動制御工学, 電気回路網学, 電力系統工学, 無線通信工学, 有線通信工学
- ・電子工学科

個体電子工学, 光波電子工学, 高周波工学, イオン工学, 制御システム工学, 電子回路工学, 電子工学基礎論, 電子装置, 電子ビーム工学, 電子物理学, 半導体工学, 量子エレクトロニクス
- ・電子制御機械工学科

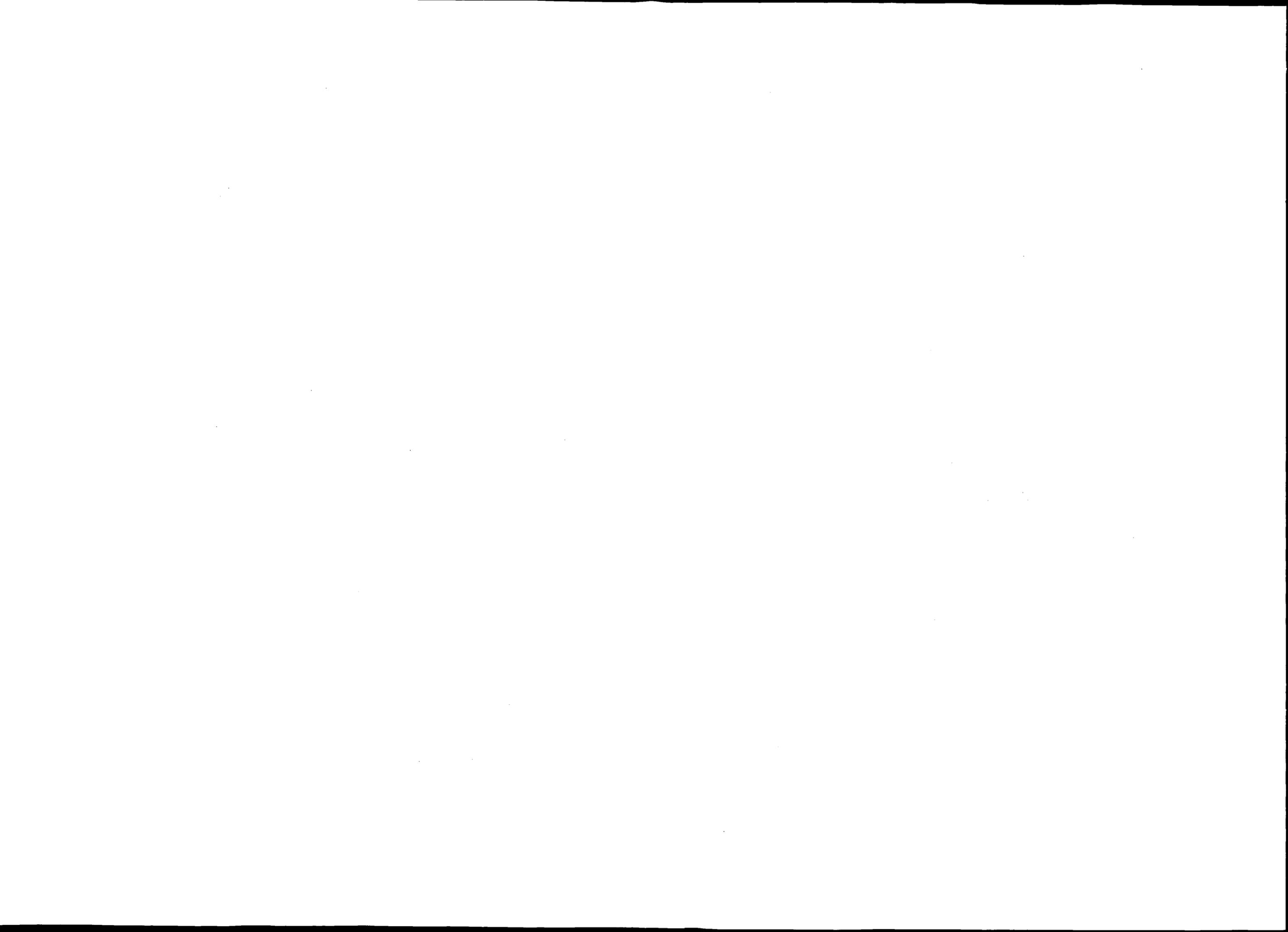
機械現象解析学, 計算機援用生産工学, ロボット工学, 制御工学, 情報変換処理工学, 電子機械基礎
- ・電子情報工学科

付属資料2 大学の学科・講座の分野事例(6)

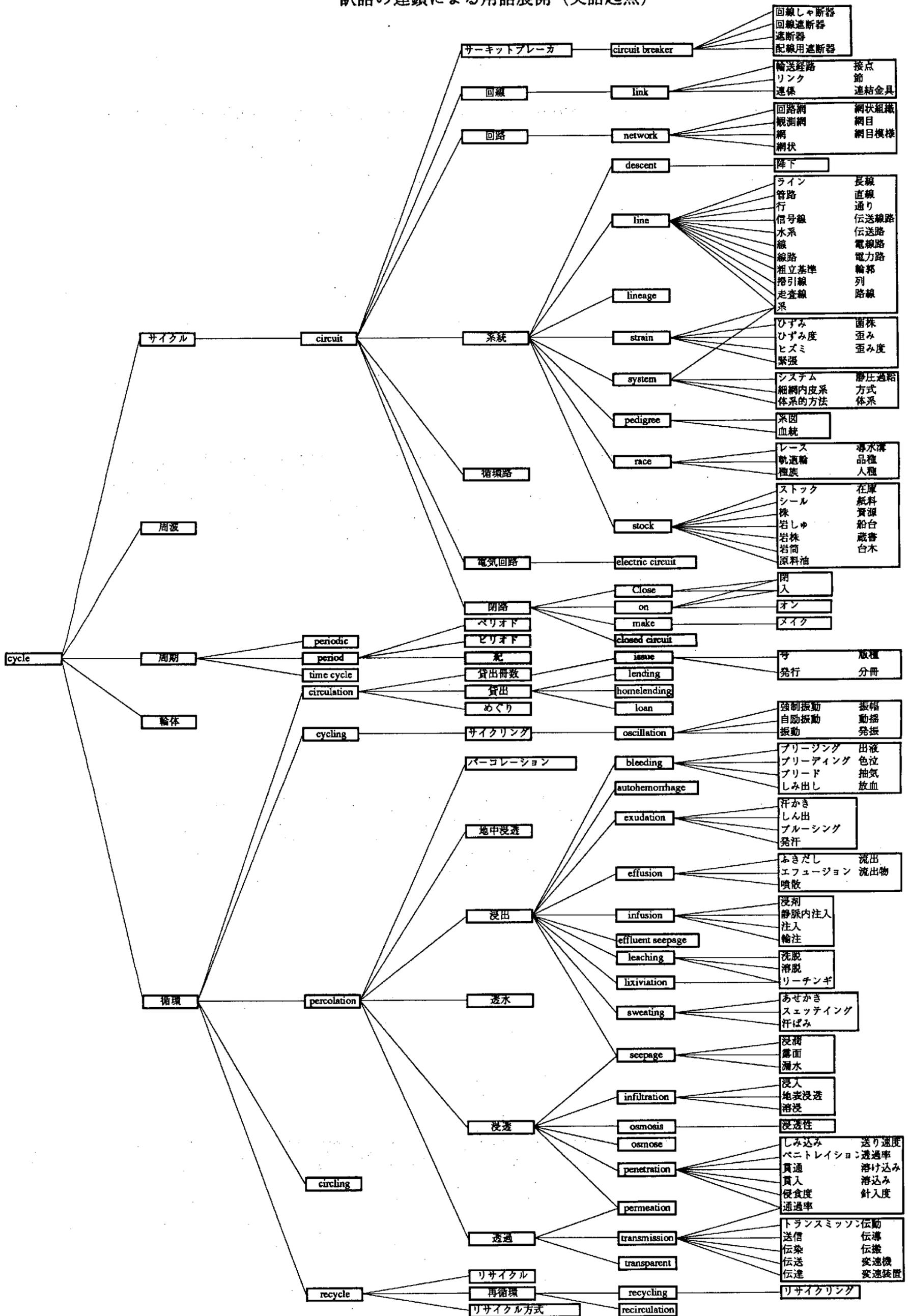
- 〈地球科学〉
- ・宇宙・地球科学科 自然物質学, 地球構造学, 地球物性学
 - ・地球物理学科 気象学, 海洋物理学, 応用地球物理学, 地殻物理学, 地球電磁気学, 地震学, 物理気候学
 - ・地球惑星物理学科
 - ・地学科
 - ・地質学鉱物学科 岩石学, 鉱物学, 地史学, 地層学, 物理地質学
- 〈都市工学〉
- ・環境工学科 温湿度調整工学, 環境計画学, 環境設計工学, 空気浄化工学, 水資源工学, 水質管理工学
 - ・都市工学科
- 〈土木〉
- ・交通土木工学科 起終点施設学, 交通施設計画, 都市交通工学, 路線施設学, 路盤基礎工学
 - ・土木工学科 海岸工学, 海岸工学・港湾工学, 河川工学, 応用構造学, 橋梁工学, 基本構造学・構造論, 交通工学, 構造力学, 水工計画学, 耐震工学, 水理学・河川工学, 土質基礎工学, 土質力学, 土木計画学, 土木材料学, 土木施工学
- 〈農業〉
- ・食品工学科 農業分析学, 農産製造学
 - ・農学科 果樹園芸学, 薬花卉園芸学, 育種学, 作物学, 雑草学
 - ・農芸化学科 細胞物理化学, 細胞有機化学, 植物栄養学, 植物分子生物学, 制御発酵学, 生物化学, 生物調節化学, 天然高分子化学, 土壌学, 分子細胞育種学, 発酵生理及び醸造学
 - ・農業経済学科
 - ・農業工学科
 - ・農業生物学科
 - ・農業工学科 かんがい排水学, 水利工学, 農業施設工学, 農産加工機械学, 農地計画学, 農用原動機学, 農用作業機械学
 - ・農林経済学科 農学言論, 農業経営学, 農業計算学, 農史, 農政学
 - ・農林生物学科 応用植物学, 昆虫学植物病理学, 植物病理学, 実験遺伝学
 - ・畜産学科 家畜栄養学, 家畜育種学, 家畜生体機構学, 家畜繁殖学

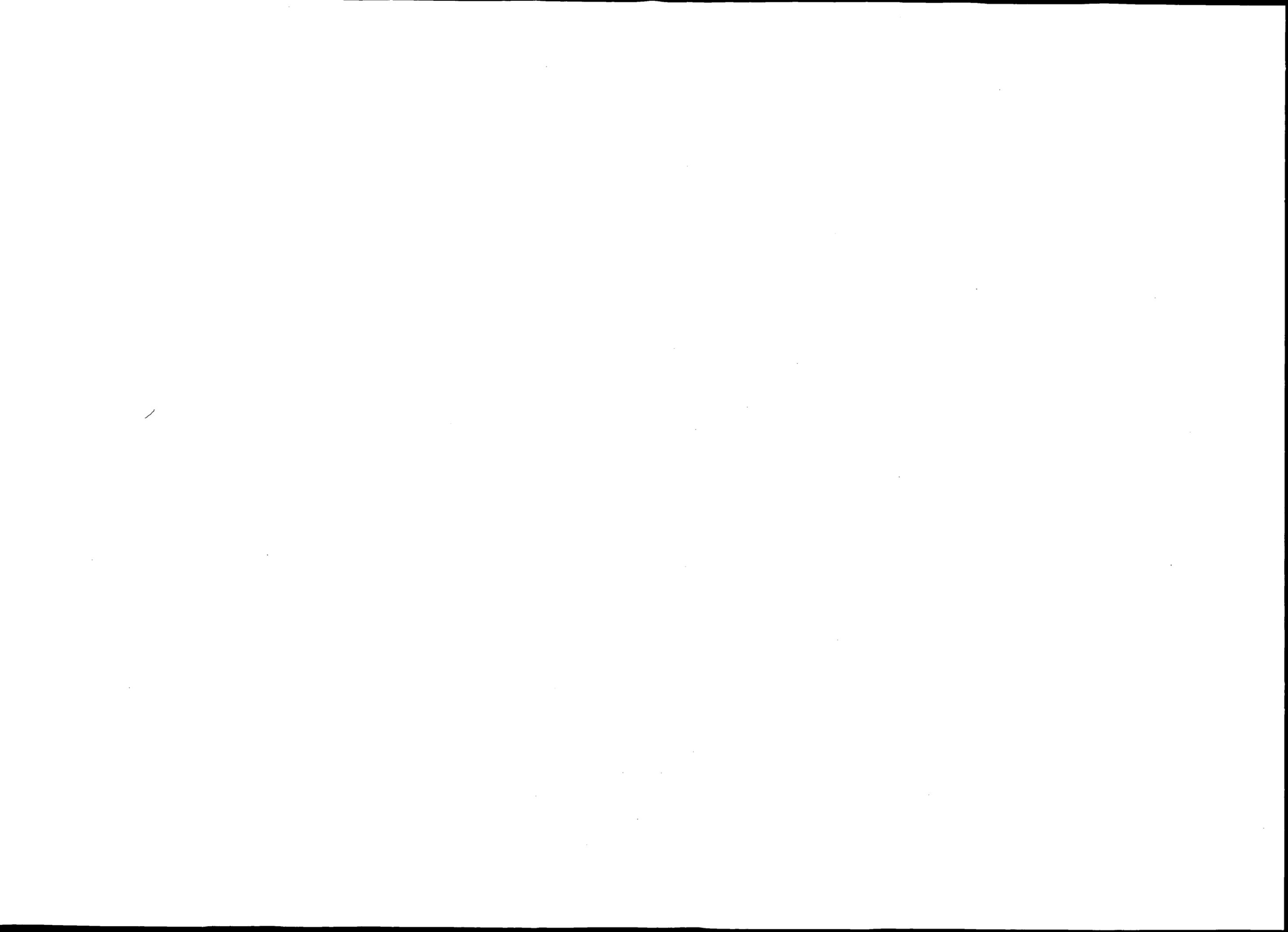
付属資料2 大学の学科・講座の分野事例(7)

- 〈物理〉
- ・物理学科 結晶物理学, 宇宙線物理学, 核エネルギー学高エネルギー物理学, 核分光学, 核反応物理学, 原子核論, 原子核物理学第一, 原子核物理学第二, 原子物理学, 固体分光学, 極低温物理学, 基礎物理学, 高エネルギー物理学, 高分子物理学, 素粒子論, プラズマ物理学, 電波分光学, 中性子物理学, 物性基礎論, 物性物理学第一, 物性物理学第二, 分子物理学, 量子物理学第一, 量子物理学第二, 流体物理学, 輻射物理学
- 〈林業〉
- ・農林経済学科 林政学
 - ・林学科 砂防学, 森林経理学, 森林生態学, 造園学, 林業工学
 - ・林産学科
 - ・林産工学科 木材化学, 木材加工材料学, 木材工学, 木材構造学, 林産機械学, 林産化学



付属資料-4
訳語の連鎖による用語展開 (英語起点)





—— 禁無断転載 ——

平成 5年 3月発行

発行 財団法人 データベース振興センター
東京都港区浜松町二丁目4番1号
世界貿易センタービル7階
TEL 03-3459-8581

委託先 科学技術情報研究所株式会社
東京都港区赤坂八丁目13番13号
ポーン赤坂503
TEL 03-3423-2400

印刷所 株式会社J S K
東京都台東区千束2丁目5番3号
TEL 03-3876-6105



