

英語科学論文中の能動態および受動態の  
他動詞＋名詞コロケーションリスト清水 眞<sup>\*1</sup> 村田真樹<sup>\*2</sup>東京理科大学<sup>\*1</sup> 鳥取大学大学院<sup>\*2</sup>

(makoto@rs.kagu.tus.ac.jp, murata@ike.tottori-u.ac.jp)

## 1 はじめに

著者達は一連の研究で、編纂した英語科学論文誌コーパスより、他動詞と名詞のパターンを自動抽出、分析してきた。本論文では、これらの分析結果を示し、関連する他の研究との比較対照を行い、専門用語のリストについて論じる。

## 2 研究の背景と目的

特殊目的のための英語 (English for Specific Purposes, ESP) の隆盛とあいまって、コーパス言語学では最近、分野別コーパスが編纂され、研究が進められている。Coxhead(2000) は 31 分野の学術論文、Connor & Upton (2004) は学術・法律、Biber (2006) は大学で用いられる英語、石川、小濱(2006) は工学、田地野他(2008) は 35 分野の論文コーパスを編纂している。しかしながら、その研究対象は単語のみであることが多く、コロケーションを取り上げたものは少ない。

著者達は理工系の学術論文コーパスを編纂し、コロケーションを分析することとした。コロケーションには様々なタイプが存在するが、この研究では能動態の他動詞とその目的語、および受動態の他動詞とその主語に焦点をあてる。

コロケーション情報に関しては、*LDOCE*(CD版)、『英辞郎 on the Web Pro』などがあるが、その情報は主に一般的表現に関するもので、個々の分野の専門用語の情報が充分ではない。様々な専門用語の辞典は、コロケーション情報がない。

理工系学術論文コーパスの編纂、コロケーションリストの作成、そしてそれらの英語学術論文のリーディングおよびライティングの指導への応用というのが最終的な目標である。

## 3 研究方法

データは英語物理化学論文誌 (*Journal of Physical Chemistry*, 以降 JPC と略)、英語化学論文誌 (*Journal of American Chemical Society*, 以降 JACS と略) よりそれぞれ論文を 100 本選び、コーパスを構築した。詳細については、清水(2011) 参照。述べ語数は、JPC が 524, 010 語、JACS が 620, 545 語となった。ふたつのコーパスから、能動態の他動詞＋目的語である名詞パターンおよび主語である名詞＋受動態の他動詞を自動抽出した。用いたソフトウェアは、Charniak Reranking Parser

(Charniak(2000) 参照。)である。抽出の基準は、

(1) a. (VB?? ???) (NP ???)

b. (NP ???) (AUX be) (VBN ???)

のパターンにマッチする文である。

得られたパターンのうち、トークンが 2 以上、2 本以上の異なる論文に生起するものを人手で分析した。名詞を 12 タイプに分類し、その後、チェックした。次に、頻度の高い動詞上位 10 語およびその動詞の共起語である名詞について調べた。共起する名詞は、「実験」、「議論」、「装置」、「式」、「数値」、「図表」、「方法」、「物質」、「現象」、「状態」、「場所」、「その他」、という 12 の範疇に分類した。

## 4 分析結果

## 4.1 能動態

JPC では、1250 トークンが抽出された。トークンの生起数上位 10 語は、*use*(462)、*show*(200)、*provide*(139)、*contain*(123)、*form*(91)、*determine*(90)、*give*(89)、*represent*(88)、*include*(85)、*produce*(85) である。( ) 内の数字は生起数である。目的語として *use* と共起する頻度が高いのは、「装置」、「式」であった。*show* は「現象」、「状態」、「数値」、「図表」、*give* は「議論」、「現象」、「実験」、「数値」、*provide* は「議論」、「方法」、「実験」、「数値」、*obtain* は「現象」、「実験」、「状態」、「数値」、*contain* は「物質」、*see* は「議論」、「実験」、「図表」、「数値」、*thank* は「その他」との共起が顕著であった。

JACS では、1599 トークンが抽出された。トークンの生起数上位 10 語は、*use*(852)、*show*(366)、*give*(274)、*provide*(269)、*contain*(257)、*form*(187)、*represent*(137)、*involve*(134)、*yield*(121)、*see*(116) である。目的語として *use* と共起する頻度が高いのは、「装置」、「方法」であった。*show* は「現象」、「状態」、「数値」、「図表」、*give* は「議論」、「現象」、「実験」、「数値」、*provide* は「議論」、「方法」、「実験」、「数値」、*obtain* は「現象」、「実験」、「状態」、「数値」、*contain* は「物質」、*see* は「議論」、「実験」、「図表」、「数値」、*thank* は「その他」との共起が顕著であった。

JPC と JACS に共通する動詞は 7 タイプである。順位

はかなり異なる。それぞれの能動態の他動詞の JPC と JACS における頻度の差を計量するためカイ二乗検定を行った。*use, show, provide, contain, form, represent* は、すべて 5%水準で有意差あり、*give* は 0.1%水準で有意差ありという結果となった。

#### 4.2 受動態

JPC では、7279 トークンが抽出された。トークンの生起数上位 10 語は、*used*(387)、*shown*(337)、*observed*(234)、*calculated*(205)、*obtained*(161)、*found*(152)、*performed*(145)、*determined*(140)、*measured*(108)、*given*(108)である。主語として *used* と共起する頻度が高いのは、「装置」、「方法」であった。*shown* は「現象」、「状態」、「数値」、「図表」、*observed* は「議論」、「現象」、「実験」、「数値」、*calculated* は「議論」、「方法」、「実験」、「数値」、*obtained* は「現象」、「実験」、「状態」、「数値」、*found* は「物質」、*performed* は「議論」、「実験」、「図表」、「数値」、*determined* は「その他」、*measured* は「その他」、*given* は「その他」との共起が顕著であった。

JACS では、7348 トークンが抽出された。トークンの生起数上位 10 語は、*used* と共起する頻度が高いのは「議論」、「方法」であった。*shown*(256)、*observed*(248)、*performed*(184)、*obtained*(175)、*found*(127)、*determined*(108)、*added*(102)、*measured*(93)、*removed*(90)である。*used*(335)、*shown*(256)、*observed*(248)、*performed*(184)、*obtained*(175)、*found*(127)、*determined*(108)、*added*(102)、*measured*(93)、*removed*(90)である。

JPC と JACS に共通する動詞 9 タイプである。順位はかなり異なる。それぞれの受動態の他動詞の JPC と JACS における頻度の差を計量するため、カイ二乗検定を行った。*used, shown, found, calculated, determined* は 0.1%水準で有意差あり、*measured* は 5%水準で有意差あり、*observed, obtained, performed* は有意差なしという結果となった。

#### 4.3 BNC との比較

JPC において頻出する他動詞上位 10 語と、一般的な英語の状況を反映していると考えられるサンプリングコーパスである British National Corpus (BNC) について、頻度の差を計量するためカイ二乗検定を行った。ただし、能動態、受動態を区別しないレンマの比較である。*use, show, observe, calculate, obtain, find, perform, determine, measure* はすべて 0.1%水準で有意差があり、JPC の方が頻度が高く、*give* は 0.1%水準

で有意差があり、JPC の方が頻度が低いという結果になった。

JACS において頻出する他動詞上位 10 語と BNC についても、頻度の差を計量するためカイ二乗検定を行った。能動態、受動態を含めたレンマの比較である。*use, show, observe, perform, obtain, find, determine, add, measure, remove* はすべて 0.1%水準で有意差があり、JACS の方が頻度が高いという結果になった。

#### 4.4 物理化学論文(JPC)と化学論文(JACS)の違い

他動詞生起度の順位だけを見ると JPC と JACS は似ているという印象を受けるが、他動詞と共起する名詞の生起数には相当な違いがある。代表的な例が、「スペクトル」を意味する *spectrum, spectra* である。JPC ではそれぞれ 20 回、5 回、合計 25 回生起しているのに対し、JACS ではそれぞれ 10 回、4 回、合計 14 回である。総語数は、JPC が 524,010 語、JACS が 620,545 語であることを考えると、JPC における頻度は相当に高い。カイ二乗検定を行うと 5%水準で有意差ありという結果になった。物理化学の専門家にコメントを求めたところ、まさに物理化学と化学の研究特徴の違いを反映しているという答えであった。JPC、JACS で生起数が多い名詞について、動詞との共起を調べた。

石川・小浜(2007)、田地野他(2008)も指摘しているとおり、同じ語でも一般語彙と学術語彙では、意味が異なることが多い。例えば、*effect* の一般的な意味は、「結果、影響」であるが、物理学における専門的な意味は、「((物理)効果)である。LDOCE に記載されている共起語は、6 タイプである。これに対し、JPC で観察された共起語は 22 タイプである。LDOCE と JPC に共通する動詞は 1 タイプ、*show* のみに過ぎない。JACS で観察されたコロケーションは 15 タイプである。LDOCE と JACS に共通する動詞は 1 タイプ、*show* のみに過ぎない。一方、JPC と JACS に共通する動詞は、10 タイプ、JACS の 15 タイプの 2/3 が JPC と共通する。

*energy* の一般的な意味は「勢力、活気」、専門的な意味は「((物理)エネルギー)、LDOCE に記載されている共起語は 10 タイプ、JPC で観察された共起語は 30 タイプである。LDOCE と JPC に共通する動詞タイプはない。JACS で観察された共起語は 9 タイプである。LDOCE と JACS と共通する動詞タイプはない。JPC と JACS に共通する動詞は、8 タイプで、実に JACS の 9 タイプの 8/9 が JPC と共通する。

*group* の一般的な意味は「集団、グループ」、専門的な意味は「((化学)基、元素団)、LDOCE に記載されている共起語は 3 タイプ、JPC で観察された共起語

は3タイプである。LDOCE と JPC に 共通する動詞タイプは1タイプである。JACS で観察された共起語は10タイプである。LDOCE と JACS 共通する動詞タイプはない。JPC と JACS に 共通する動詞は、1タイプのみである。

*increase*の一般的な意味は「増加、上昇」、専門的な意味は「増大量」、LDOCE に記載されている共起語は2タイプ、JPC で観察された共起語は4タイプである。LDOCE と JPC に 共通する動詞タイプはない。JACS で観察された共起語は7タイプである。LDOCE と JACS 共通する動詞タイプはない。JPC と JACS に 共通する動詞は、2タイプである。

*reaction*の一般的な意味は「反作用、反動」、専門的な意味は「(化学)反応」、LDOCE に記載されている共起語は7タイプである。JPC で観察された共起語は9タイプである。LDOCE と JPC に 共通する動詞タイプはない。JACS で観察された共起語は18タイプである。LDOCE と JACS に 共通する動詞タイプはない。JPC と JACS に 共通する動詞は2タイプである。

## 5 先行研究との比較

### 5.1 石川・小浜(2007)

石川・小浜(2007:309-310)は、論文テキストに含まれる語を、基本語彙、準専門語彙、専門語彙に分類し、それぞれ次のよう定義する。

- (1) 基本語彙：論文テキストだけでなく、どのテキストにも繰り返し現れる高頻度語
- (2) 準専門語彙：専門分野に関わらず、さまざまな分野の論文テキストに共通に含まれている語彙で、基本語彙以外の語。すなわち、論文テキストには高頻度で表れるが、他のテキストでは使用頻度が低く、論文テキストの特徴語となる語彙
- (3) 専門語彙：特定の専門分野でのみ使用される語彙

石川・小浜は、先行研究においては準専門語彙と専門語彙の境界線が明確ではないと指摘し、さらに、準専門語彙をCoxhead(2000)の提唱するAcademic Word List (以下AWL)と同一視することを批判している。石川・小浜によれば、AWL のデータとなったコーパス(約350万語)には、芸術・法学・商学・科学の4分野しかないため、科学テキストに使用される語彙は4分の1しか反映されておらず、AWLが文系の分野に偏っているためであるとする。

石川・小浜は工学系学術論文テキストのみからなる学術論文コーパス(約52万語)を構築し、語彙表を作成する。そして、3分野以上の工学系専門分野にまたがって学術論文テキストに特徴的に表れる語彙群を、「準

専門語彙」とする。また、工学系学術論文の中でも、機械・生物・電気など、工学の特定分野でのみ使用される語を「専門語彙」とする。したがって、人文・法学などの分野で学術論文に使用される頻度が低い場合でも、工学系の多分野で高頻度で使用されている場合は、「準専門語彙」と定義される。

この石川・小浜の主張にいくつかの疑問がわく。まず、専門語彙は特定の分野でのみ使用されるという点である。電子工学には電子工学の専門分野が存在するということが、極めて自然であるように感じられるかもしれない。しかし、同じ語が複数の分野で専門用語とされていることがある。しかも、その際、日本語訳が異なることがある。ただ単にそれぞれの分野においてその語を最初に訳した研究者が異なる訳語を使用したという場合もあるだろうが、分野によりかなり概念が異なる場合もある。このような際、これを専門語彙と見なすのか、それとも準専門語彙と見なすのか。

石川・小浜(2007:313-4)自身、*blade*が電気分野、機械分野で多く生起し、一般の意味とは異なることから、「特定の1ジャンル、または、近い分野のみにまたがった2ジャンルに限定されているので、ここでは専門用語とみなす」と主張する。これは、先にあげた石川・小浜の定義といささか齟齬をきたす。

同じ語でも、概念が異なる場合、異なる専門語彙とみなすべきであるかもしれない。ただ、異なり方が極めて小さいものから非常に大きいものがあるのが現実で、どこまでが専門語彙で、どこからが準専門語彙であると具体的に線引きするのは困難ではある。

また、準専門用語、専門用語には、一般：学術>文系：理系>工学系：理学系>物理：化学>物理化学：化学>理論系：実験系>基礎系：応用系

というような階層性が存在する可能性を指摘したい。>という記号は、その左の后者、例えば、「一般：学術」では「学術」の下位区分が、「文系：理系」であることを示す。便宜的にこのように図示したが、もちろん、「文系：理系」のいずれにも属さない、あるいは、どちらにも属するという分野もあるであろう。また、「物理：化学」のみでなく、「生物、薬学、医学」などの分野があることは言うまでもないが、ここでは省略した。「理論系：実験系>基礎系：応用系」は、「基礎系：応用系>理論系：実験系」の可能性もあるし、「物理：化学」の上にすることも考えられる。

極めて興味深く、有益である石川・小浜の準専門語彙という概念であるが、理論の精密化、具体的な分析、分析の検証が求められる。

## 5.2 田地野他(2007)、(2008)

田地野他(2007:113-124)は、学術目的の英語(EAP)を、あらゆる専門分野に共通する一般的な学術言語技能を対象とする、一般学術目的の英語(EGAP)と、ある特定の専門分野に特化した学術言語技能を対象とする特定学術目的の英語(ESAP)に区分する。前者が、石川・小浜の準専門用語、後者が専門用語に相当するであろう。5.1で行った批判がここでもあてはまる。EGAPとESAPの区別は容易であるのか、複数の分野にまたがり、しかも訳が異なり、概念も互いにずれる語の取り扱いはどのように行うのかである。

## 5.3 他のワードリストとの比較

この研究における頻出コロケーションリストと先行研究から派生したワードリストを比較した。

一例としてJPC中の能動態上位30語をあげるが、AWLには5語、名工大EGSP1304には6語、京大術語彙データベースには3語、東工大英単には10語が共通した語が存在している。

共通した語が少ない理由はいくつか考えられる。まず、本研究は頻出するコロケーションを取り上げたが、先行研究は頻出する単語を取り上げた。よって、本研究ではuseのような一般的な単語もリストアップされているが、先行研究では専門用語しか取り上げられていない。分野も関係するかもしれない。本研究では物理化学、化学であるが、先行研究ではそうではない。

## 6 おわりに

英語物理化学論文、化学論文コーパスから他動詞+名詞パターンを自動抽出、分析し、分析結果を先行研究と比較した。他動詞と名詞の共起を論じ、分析により明らかになった問題を論じた。

今後の課題は、まず、動詞の分類である。名詞は12範疇に分類したが、他動詞の分類は行っていない。動詞の分類により、ある範疇の名詞がどの範疇の動詞と共起するかを調べることが可能になる。

物理学のコーパスは既に編纂済みである。工学諸分野のコーパスを編纂中である。数学のコーパスの編纂も予定している。編纂後、分析を行わなければならない。コロケーションリストをどのように用いるかも検討し、計画を具体的に進めていく必要がある。

## 謝辞

東京理科大学理学部第一部化学科築山光一先生、宮村一夫先生、荒木光典先生から貴重なアドバイスを頂

いた。厚くお礼を申し上げます。ここに感謝したい。本研究は平成24年度科学研究費助成事業(学術研究助成金)(基盤研究(C))(課題番号:24520655)を受けた。

## 辞書・ワードリスト

Academic Word List (AWL) in Coxhead(2000)

Longman Dictionary of Contemporary English (LDOCE, Longman)

英辞郎 on the Web (アルク)

京大術語彙データベース基本単語1110 (研究社)

東工大英単: 科学・技術例文集 (研究社)

名工大EGSP1304 <http://language.sakura.ne.jp/y/>

## 参考文献

Biber, D. (2006). *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Charniak, E. (2000). A Maximum-Entropy-Inspired Parser. in *Proceedings of NAACL*, 132-139.

Connor, U. & Upton, T. A. (Eds.) (2004). *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly* 34:2, 213-238.

石川有香・小山由紀江(2007). 「学術論文リーディングを目的とした指導語彙の選定」『中部地区英語教育学会紀要第36巻』、309-316.

Schmid, H. (1997). Probabilistic part-of-speech tagging using decision trees. in D. Jones and H. Somers (eds.) *New Methods in Language Processing*, London: UCL Press.

清水真(2011). 「学術論文コーパスの設計、構築に関する問題点について」、『東京理科大学紀要(教養篇)第43号』、255-267.

Shimizu, M., Murata, M., Nakatani, Y. and Hijikata, Y. (2010). Collocations in physical chemistry papers, in *Proceedings of the 6th ICT in Analysis, Teaching and Learning of Language (CD-ROM)*, 211-219.

田地野彰・寺内一・笹尾洋介・マスワナ紗矢子(2007). 「総合研究大学における英語学術語彙リスト開発の意義—EAPカリキュラムデザインの観点から—」『京都大学高等教育研究第13号』、121-131.

田地野彰・寺内一・金丸敏幸・マスワナ紗矢子・山田浩(2008). 「英語学術論文執筆のための教材開発に向けて—一論文コーパスの構築と応用」『京都大学高等教育研究第14号』、111-121.