

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 1 日現在

機関番号：32660

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24520655

研究課題名(和文) 英語科学論文読解作成支援のためのコーパス作成およびコロケーションの研究

研究課題名(英文) The construction of science and technology paper corpora and the compilation of collocation lists

研究代表者

清水 眞 (SHIMIZU, Makoto)

東京理科大学・理学部・教授

研究者番号：50187465

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：英語科学技術論文誌30誌より各100ずつ論文を選び、コーパスを構築した。そのコーパスを用いて、n-gramおよび能動態、受動態を自動的に抽出し、分析を行った。
n-gramに関しては、4-gram以上では、あまり興味深い表現は抽出できなかった。能動態、受動態に関しては、頻度の高い動詞を認定し、どのような名詞と共起するかを調査し、コロケーションリストを作成した。

研究成果の概要(英文)：We constructed corpora from 100 papers selected from 30 science and technology academic journals. N-gram expressions and active and passive constructions were automatic extracted and analysed.
As for n-gram expressions, interesting expressions were not found among 4-gram and above. As for active and passive constructions, we compiled collocateion lists of verbs + nouns, and analysed the features of verbs and nouns.

研究分野：外国語教育

キーワード：コーパス コロケーション 英語科学技術論文 自動抽出 能動態 受動態 n-gram 名詞

1. 研究開始当初の背景

Brown コーパス、LOB コーパスなど、汎用コーパスから始まったコーパス言語学であるが、最近では、分野別コーパスの編纂が盛んである。Coxhead(2000) は 31 分野の学術論文、Connor & Upton (2004) は学術・法律などの分野、Biber (2006) は大学で用いられる多様な英語、石川、小濱(2006) は理学・工学論文、田地野他(2008) は 35 分野の論文のコーパスを編纂し、研究を行っている。

Shimizu et al. (2004)は、英語の母語話者が通常 *dream* と共に *achieve* あるいは *fulfill* という動詞を共起語として用いるのに対し、日本人の学生は日本語の「夢に到達する」という日本語の表現に影響され、*reach* という動詞を共起語として用いたことを指摘した。このような共起語情報へのアクセスは、*LDOCE* の CD 版、『英辞郎』、特に『英辞郎 on the Web Pro』など、ここ数年めざましく改善されてきた。しかしながら、その情報は主に一般的表現に関するもので、個々の分野、例えば、物理、化学などの諸分野の専門用語に関しては、情報が充分であるとは言い難い。研究社の『理化学辞典』は、各分野の専門用語が詳しい解説と共に記載されている有用な辞書であるが、出版から年月がたっているため *nanoparticle* など最新の用語が収録されていないという問題のほか、コロケーション情報がないという欠点がある。

2. 研究の目的

こうした背景を踏まえ、著者およびその共同研究者は、理工系の学術論文コーパスを編纂し、各種の分析を行うことを計画した。その分析のひとつが、n-gram、もうひとつが他動詞 + 名詞などのコロケーションである。特に、英語の非母語話者である日本人の学生にとって、英語のリーディングおよびライティングの際、困難と感じる項目の

ひとつが、他動詞 + 名詞パターンだからである。「+ 名詞」というのは、他動詞の目的語ということの意味している。動詞と名詞のコロケーションは、動詞と主語など他のタイプも存在するが、この研究では、他動詞とその目的語のみを扱う。

理工系の学術論文コーパスを編纂し、コロケーションリストの作成を行えば、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に極めて有益であると考えられる。

3. 研究の方法

英語科学論文誌および英語技術論文誌 30 誌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。)論文の選択にあたっては、以下の基準を設けた。

- 1 ファーストオーサーのファーストネームが英語のネイティブスピーカーであると思われる
- 2 ファーストオーサーの所属先がいわゆる英語圏の国にある

その結果、30 のコーパスが編纂された。

編纂されたコーパスより、4-gram 以上の表現を自動抽出した。また、能動態とその目的語、受動態とその主語を自動抽出した。

自動抽出した表現のうち、トークンが 2 以上のものを人出で分析した。

4. 研究成果

研究実施計画において 10 本コーパスの編纂を行うという計画を立てていたが、計画通り 10 本のコーパスを編纂できた。最終目的の 30 本という目的が達せられた。

n-gram に関しては、謝辞に用いられる表現を除いては、あまり有用な表現は見出せなかった。

物理化学の能動態の他動詞のうち頻出するものは、化学の能動態の他動詞のうち頻出するものとはかなり一致しており、目的語として他動詞と共起する名詞の頻度が高いのは、物理化学と化学の間でかなり一致している。受動態の場合も、物理化学の学術誌に頻出するものは、化学の学術誌に頻

出するものとかかなり一致しており、主語として他動詞と共起する名詞の頻度が高いのは、物理化学と化学の間でかなり一致している。

材料科学の能動態の他動詞のうち頻出するものは、物理化学、化学の能動態の他動詞のうち頻出するものとかかなり一致している。材料科学の上位 10 語のトークン数と物理化学、化学のトークン数を 二乗検定にかけたところ、10 語中 7 語が有意であった。目的語として他動詞と共起する名詞の頻度が高いのは、材料科学、物理化学、化学の間でかなり一致している。受動態の頻出動詞は、物理化学とはあまり一致しない。材料科学の上位 10 語と物理化学のトークン数を 二乗検定にかけたところ、10 語中 3 語しか有意ではなかった。化学とはかなり一致する。材料科学の上位 10 語と化学のトークン数を 二乗検定にかけたところ、10 語中 7 語が有意であった。

生命科学の能動態の他動詞のうち頻出するものは、物理化学、化学の能動態の他動詞のうち頻出するものとかかなり異なる。生命科学の上位10語のトークン数と物理化学、化学のトークン数を 二乗検定にかけた。生命科学と物理化学では、10語すべてに0.1%の水準で有意差があった。生命科学と化学では、10語中5語は0.1%の水準で、2語は1%の、2語は5%の水準で有意差があった。1語には有意差がなかった。目的語として他動詞と共起する名詞の頻度が高いのは、材料科学、物理化学、化学の間でかなり異なる。

受動態の頻出動詞は、生命科学と物理化学ではかなり異なる。生命科学の上位10語と物理化学のトークン数を 二乗検定にかけたところ、10語中9語が0.1%の水準で有意差があった。1語には有意差がなかった。生命科学と化学ともはかなり異なる。生命科学の上位10語と化学のトークン数を 二乗検定にかけたところ、が0.1%の水準で有意差があった。1語は5%の水準で有意差があった。

10 誌における能動態の他動詞とその目的語、受動態の他動詞とその主語のコロケー

ションリストを作成した。下記の研究代表者のウェブページ上で公開している。

5 . 主な発表論文等
(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 件)

〔学会発表〕(計 4 件)

Makoto Shimizu and Masaki Murata (2012) 'Passive and active voices in scientific papers', AsiaTEFL 2012, New Delhi, India, October 5

清水 眞、村田 真樹 (2013)「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」言語処理学会、名古屋大学(愛知県名古屋市) 3月14日

Makoto Shimizu and Masaki Murata (2013) 'Passive and active voices in engineering papers', GloCALL 2013, Da Nang, Viet Nam, November 9

Makoto Shimizu and Masaki Murata (2015) 'Active and passive voices in biology', CamTESOL 2015, Phnom Penh, Cambodia, March 1

〔図書〕(計 件)

〔産業財産権〕
出願状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

<http://www.rs.kagu.tus.ac.jp/makoto/>

6. 研究組織

(1) 研究代表者

清水 眞 (Makoto SHIMIZU)
東京理科大学・理学部第一部・教授
研究者番号：50187465

(2) 研究分担者

村田 真樹 (Masaki MURATA)
鳥取大学・工学研究科・教授
研究者番号：50358884

(3) 連携研究者

()

研究者番号：