

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 27 日現在

機関番号：10101

研究種目：基盤研究(C)

研究期間：2011～2014

課題番号：23520655

研究課題名(和文)ブレ・アカデミック英語コーパスの構築と「学習英語」の有用動詞コロケーションの特定

研究課題名(英文)A compilation of a pre-academic English corpus and identification of its useful verbal collocations

研究代表者

高見 敏子 (Takami, Satoko)

北海道大学・大学院メディア・コミュニケーション研究院・准教授

研究者番号：60333639

交付決定額(研究期間全体)：(直接経費) 3,900,000円

研究成果の概要(和文)：英語圏の小学校・中学校レベルの科学の教科書・教材図書・ビデオの本文・スクリプト・字幕および中学校レベルの歴史・地理の教科書の本文を電子テキスト化して約168万語の「ブレ・アカデミック英語コーパス」を構築し、同コーパスにおける高頻度動詞やその共起語を調べ、日本の大学生の学習に有益と思われる表現を収集した。また、英語表現の習得を支援するためのウェブ学習システム「Smile」を開発した。Smileは一文ずつ例文を提示し、学生に記憶させた後、提示文の表示を消して学生に例文を再現させる「英文再現力養成システム」である。当方は登録人数を限定してシステムを運用し、次第に利用者を拡大していく予定である。

研究成果の概要(英文)：As one of the major products of this project, a “Pre-academic English Corpus” has been compiled for research purposes. It consists of about 1.7 million words of texts, scripts and subtitles taken from textbooks, books and videos for science, geography and history education targeted at English-speaking children ranging from the elementary to the middle school levels. Verbal collocations have been collected in the compiled corpus, which seem to be good to learn for Japanese university students. In order to help students learn these useful expressions, a new e-learning system named “Smile” has been developed. This system, which works by displaying sentences one by one, allows students time to memorise each. It then hides the sentence and lets the students reproduce the same sentence from memory. The system will be open to a limited number of students in the first year, but it is planned to accept more students later if it works well.

研究分野：コーパス言語学

キーワード：英語 動詞 コロケーション Webシステム 英文再現力

1. 研究開始当初の背景

高校英語と学術英語の間には大きなギャップがあり、多くの大学生にとってはこの移行が容易ではないのが現状である。高校までは「英語を学ぶ」のが中心であり、大学で一足飛びに「英語で学術的な内容を学ぶ」ことに困難を感じるのは自然なことである。このギャップの橋渡しとして、まず「英語で平易な内容を学ぶ」ことから始めることが必要である。

上記の現状の改善に有効と考えられるのが、英語圏の学校教育で使用されている図書や字幕付ビデオ教材の利用である。そこで、こうした教材で使われている「学習のための英語」を「プレ・アカデミック英語」と呼ぶことにし、コーパスを構築してその特徴を分析し、さらに、その知見を学術英語の準備段階としての「学習英語」の指導のために役立てられるようにしたい、というのが本研究開始時の背景であった。「読めば意味が理解できる」というだけでなく、学習者が実用的に英語で話したり、英文を書いたりすることができるようになることが必要であるという認識から計画した研究である。

2. 研究の目的

(1) 英語圏の小学校高学年～高校で用いられている学習教材を電子テキスト化し、「プレ・アカデミック英語コーパス」を構築する。書き言葉としては学習図書の本文を、話し言葉としては学習ビデオの音声文字化したものを対象としてコーパスを作成する。

(2) 構築したコーパスを分析して、「プレ・アカデミック英語」の特徴を調査し、特に英文を組み立てる際に重要な要素となる動詞に注目して、有用な動詞のコロケーションを収集する。

(3) 特定した動詞のコロケーションに関する情報を公表するとともに、成果の応用として、その学習を支援するウェブシステムや教材を作成する。

3. 研究の方法

(1) 英語圏の小学校高学年～高校で用いられている科学・歴史・地理の学習教材（図書またはビデオ）を購入し、本文（ビデオの場合はスクリプト）が PDF ファイルで提供されている場合はそのファイルから、紙媒体で入手したものはスキャナーで画像化したのち、OCR ソフトを用いて電子テキスト化を行い、スペルチェッカーのみに頼らず本文に目を通して読み取り結果を確認し、誤変換を修正した。

(2) (1)で作成した英文テキストの品詞タグ付けには品詞タグの GoTagger（後藤一章氏作成）を使用した。次に高機能なコンコーダンサである AntConc (Laurence Anthony 氏

作成)を用いて、(1)で作成した「プレ・アカデミック英語コーパス」の品詞タグ付語彙頻度表を作成し、頻度の高い動詞項目からそれぞれを含む文を検索し、動詞を含む n-gram や頻度の高い動詞の共起語、またその使われ方などを調べた。

(3) (2)のデータをもとに、日本の大学生にとって有用と思われる動詞のコロケーションを収集した。コロケーションの学習を支援するウェブシステムの機能を検討した結果、コロケーションを文脈の中で学ぶことが必要であるという結論に至り、文単位で学習するウェブシステムを考案した。作成したコーパスの本文は著作権上の問題からそのまま公表することはできないため、使用可能な例文を準備する必要がある。しかしある出版社に使用料を支払うことで本文や朗読音声を教材用に利用できないか打診してみたところ難しいとの回答であり、市販の教材を従来と異なる形式で利用することは予想以上に困難であることが判明した。

4. 研究成果

(1) 小学校レベルの科学ビデオ教材（ナレーション形式・対話形式）、小学校高学年レベルの科学図書・科学コミック、小学校高学年～中学校レベルの科学教材、中学校レベルの科学・地理・歴史の教材図書、中学校レベルの科学ビデオ（ナレーション形式）、中学校レベルの世界史・米国史・地理の教科書、一般向けの科学・歴史ビデオについて、合計約 168 万語の「プレ・アカデミック英語コーパス」を構築した。

(2) プレ・アカデミック英語コーパスにおける高頻度動詞は以下のとおりである。（s 以外は原形で[レマ化して]集計したデータである。's のほとんどは is の縮約形であった。）

1. be	11. take
2. have	12. form
3. do	13. get
4. make	14. find
5. use	15. know
6. call	16. go
7. see	17. become
8. live	18. produce
9. move	19. change
10. 's	20. come

本研究で構築したコーパスは科学分野のテキストが多かったため、さまざまなサブコーパス（あるいはさまざまなジャンルのテキスト）を含む Bank of English や British National Corpus といった大規模コーパスとは異なる特徴が高頻度動詞の一部にも現れている。Leech *et al.* (2001: 282)によれば、British National Corpus における高頻度動詞は(法助動詞を除いて)上から順に *be, have,*

do, say, get, make, go, see, know, take, think, come, give, look, use, find, want, tell, put, mean, become であり、call, live, move, form, produce, change などの動詞の使用が多いことがプレ・アカデミック英語の特徴と言える。以下、一般動詞の中でもっとも頻度が高かった5つの動詞の共起語を含む表現について記述する。

make

全体としては能動態で多く用いられているが、受動態も比較的よく用いられている。特に made of (例 *made of ice*) や make/made up (例 *Two hydrogen atoms and one oxygen atom make up a water molecule. / Oxygen makes up about 20% of the air.*)、あるいはこれらを組み合わせた made up of, made out of などの表現がよく用いられる。プレアカデミック英語コーパスにおける make の後に来る特徴のある名詞としては food が挙げられ、*Plants make their own food.* のように主に植物が養分を作るという意味で用いられている。植物に関連した表現として他に make seeds (種ができる/種をつける) や make sugar(s) (訳語としては「糖」に相当するが、日本語の説明文では「デンプンを作る」と書かれていることが多い) という表現が挙げられる。また、make + 目的語 + 形容詞 (または動詞の原形) の構文となる表現がよく見られる。make things move, make work easier, make you sick などがその例である。他に make sure, make sense, make use of などの定型表現もあるが、これらは一般によく知られている表現であると言える。ただし、読めば理解できる学生であっても英文を作る際に自由に使えるとは限らないので、日本の英語教育でよく取り上げられているこのような表現についても使用できるようにするための一層の学習支援が必要である。一方で先に挙げた *Oxygen makes up about 20% of the air.* における句動詞 make up (割合などを「占める」) はこれまでの日本の英語教育では比較的なじみがなかった表現と思われる、今後積極的に取り上げていくべき有用表現であると言える。

use

プレ・アカデミック英語コーパスにおいては use は約 40% が過去分詞として用いられ、直後に来るもっとも多い語は目的を示す to で、*be used to generate electricity* の例のようにほとんどのケースは不定詞を伴っている。過去分詞の直前に用いられる副詞との組み合わせとしては also used, often used, commonly used, widely used, still used (+today), later used, probably used, even used, once used, sometimes used, not used, never used, first used, already used などが挙げられる。また used in, used for, used by, used on など、前置詞との組み合わせでもよ

く用いられるが、これらの前置詞は一般的な意味で用いられており、使用にもさほど困難はないと思われる。一方、use を含む句動詞としては use up (例 *use up fossil fuels*) の頻度が特に高く、この用法を習得していない学生は多いと推測されるため、有用性があり、教育効果が期待できる表現であると言える。

call

プレ・アカデミック英語コーパスでは能動態でも用いられるが、半数以上は過去分詞で用いられている。過去分詞の直前に用いられる副詞との組み合わせとしては also called, sometimes called, often called, commonly called, now called, generally called などが多く、上記の use のケースとある程度の共通点が見られる。また、*a process called photosynthesis* のような表現における先行詞 (本例では process) としては、上記の process のほか、structure(s), particle(s), unit(s), group(s), rock, substance(s), something, animals, instrument, material, cells, tube(s), system, organisms, area(s), land, energy, region, molecule(s), force, device, tissue, things, organ, machine, liquid, plants, tool, vessels, muscle などが頻度の高い名詞で、これらの名詞から、「NOUN called X」が科学・地理などの説明文における典型的なフレーズの1つであることがわかる。

see

プレ・アカデミック英語コーパスにおける see の大きな特徴は、原形の用法が約 52% を占めることであり、書籍における読者への指示文に用いられる命令形 (例 *See the map [chart] on page 10.*) のほか、to 不定詞や、can, can't, cannot, could, couldn't, will (縮約形を含む)、would, may, might などの助動詞とともに用いられる割合が高い。また、機能語とともに用いられて定型的な表現となっているのも特徴である。そのような例として、as you can see, see ... as, never (been) seen before, you can see from the data [diagram / map / table], (as) seen here, can see [be seen] in ..., saw[see/seeing/seen] how[that], to see if, see what [when / where / whether / which / who / why], see what happens/happened などがある。see through という表現も多いが、句動詞としての用法ではなく、*see through glass, cannot see through opaque objects* のように文字通り「通して見る」という意味で用いられている。科学に関する特徴的な表現として see[as seen] under a[the] microscope, see through a telescope[microscope], be seen with[without] a microscope, cannot [too small to be] seen with the naked eye なども挙げられる。また good [glad] to see you という口語的表現も含まれていたが、これは科

学コミックにおける登場人物の挨拶の台詞としてシリーズの中で何度か使われていたものである。

live

プレ・アカデミック英語コーパスにおける「動詞」すなわち「使用した品詞タグで動詞タグが付されたレマ live」の用例の約4割は living という語形であり、living thing(s), living organism(s), living cell(s) あるいは living space, living conditions など、主に生物の教材に見られる表現で用いられている。これらの living の用法は英和辞典では一般的に「形容詞」と分類されているものであるが、本稿では動詞タグにしたがって動詞の一用法に含める。「住んでいる」という意味の live は中学校で学習し、その際、日本語の影響から誤りやすい用法として「この意味では通常現在形は使わない」という指導もなされる。このことと混同しないようにという教育的配慮からと推測するが、living という表現は少し遅れて指導され、読めば理解できるといふ点では問題ないが、「生物」を英語で表したいときに living things という表現は思いつけず学生が実は少ないと思われる。その他、多くの用法としては live in や live on など、一般に習得度が高い(あるいは理解が容易である)と思われる、後ろに場所を表す語句を伴う表現が多いが、live in extreme environments [conditions], live in groups, live in harmony, live[lived] in poverty など特徴的な表現であると言える。また live を含む句動詞としては *This slime mold live off dead plants.* の live off が挙げられる。

以上、高頻度動詞5つの用法や共起語について述べたが、より低頻度な動詞も含め、一般的な大学生が習得していないと考えられる表現は、学生の習得状況から以下のようにいくつかのパターンに分類することができる。すなわち、(A) *The northern hemisphere is tilted toward the sun.* の tilted のように単語自体にあまりなじみがないもの、(B) *Plants give off oxygen.* の give off のように個々の語は知っているが組み合わせの意味(本例では句動詞)が未習のもの、(C) *Rocks come in various shapes and sizes.* という文の come in のように表現自体は学んでいるが用法が異なるもの、(D) *Some trains are powered by electricity.* という文の中の powered のように、動詞以外の品詞(本例では名詞)の用法のみ習得しており他品詞の用法があるという認識が希薄なもの、(E) 組み合わせから比較的容易に意味が理解できると思われるためとりたてて指導される機会が少ないと思われるもの、などである。(A)の類例としては inclined plane や revolve around など、(B)の類例としては blend in や stay put など、(C)の類例としては、既に make の項の記述で言及した *Oxygen makes*

up about 20% of the air. の make up など、(D)の類例としては *The water is warmed by the sun.* の warmed や“The sea water gets heated by hot rock.”の heated など、(E)の類例としては take in, break down, fall to などが挙げられる。

本研究で構築したコーパスはそのほとんどが英語圏の学校教育用の資料であり、したがって特定された表現はいずれも英語圏にける一般常識であると考えられるが、日本の英語教育ではなじみがないものも少なくない。中高6年間および大学で英語を学習しても英語で自由に表現できないのは、根本的にはやはり学習項目が少なすぎるためであると考えられる。

(3)(2)に述べたように日本の大学生が学ぶ価値があると思われる英語表現を特定してきたが、このような動詞のコロケーションを習得するには該当する部分の語句のみを覚えるだけでは不十分であり、文脈の中で一文が組み立てられるようになる必要がある。しかし長期記憶として例文を完全に暗記するのは非常に苦痛を感じる学習法でもある。そこで例文を画面に提示して学習者にその場でいったん記憶させ(図1)、直後に例文を隠して再現させる(図2)という動作を繰り返す英文再現力養成システム Smile を開発した。

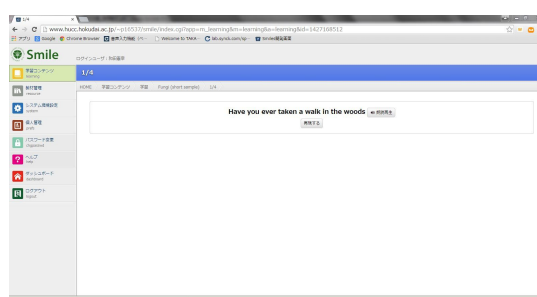


図1: 例文提示画面

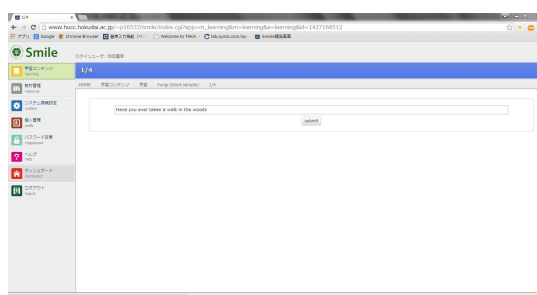


図2: 解答画面

本システムの開発にあたり工夫した点は、対応音声ファイルがない場合も合成音声で例文を読み上げさせることができる機能をつけたこと、キーボードだけでなく発声によっても入力することができるようにしたこと、例文のコピーペーストができないようにしてあること、学習者が達成感を得られるように、再現した語数を累積し、再現語数のランキング表示がされるようになってい

と、などがある。

本システムは平成 27 年度から、まず利用者を少人数に限定して試験運用を開始し、学習者からフィードバックを得つつ様子を見て利用者数を拡大し、今後の英語学習支援に活用していく予定である。

本プロジェクトでは動詞を含む表現を中心に研究を進めてきたが、今後は構築したコーパスで他品詞（形容詞・名詞など）の有用表現等についても調査し、発信型の英語力養成の支援となるデータの入手・作成や教材の開発を続けていくとともに、教育での活用も進めて、論文やホームページ等で研究や教育の成果を公表し、社会に還元していきたいと考えている。

[使用ソフトウェア]

AntConc (Laurence Anthony 氏作成)

<http://www.laurenceanthony.net/software.html>

GoTagger (後藤一章氏作成)

<http://web4u.setsunan.ac.jp/Website/GoTagger.htm>

(Eric Brill 氏作成の Brill Tagger を Windows の GUI で利用可能にしたソフトウェア)

言語研究に非常に有益なソフトウェアを作成し、自由に利用できるよう提供されている両氏に感謝の意を表する。

[参考文献]

Leech, G., P. Rayson and A. Wilson (2001) *Word Frequencies in Written and Spoken English*. Pearson Education.

5 . 主な発表論文等

採択期間中には論文や口頭発表という形で成果の公開に至らなかったが、平成 27 年度中に論文の執筆を予定している。

〔その他〕

ホームページ等

本プロジェクトのホームページ

<http://www.hucc.hokudai.ac.jp/~p16537/kaiken2011/>

英文再現力養成システム Smile 設置 URL

<http://www.hucc.hokudai.ac.jp/~p16537/smile/>

6 . 研究組織

(1)研究代表者

高見 敏子 (TAKAMI, Satoko)

北海道大学大学院メディア・コミュニケーション研究院・准教授

研究者番号：60333639