

平成22年3月18日現在

研究種目：若手研究(B)  
 研究期間：2006～2008  
 課題番号：18720144  
 研究課題名（和文） 基本語彙で豊かな英語表現を可能にするための高頻度句動詞の特定と  
 WELL 教材の開発  
 研究課題名（英文） Finding the Most Commonly-Used Phrasal Verbs in English to  
 Develop Web-Enhanced Language Learning Materials  
 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

研究成果の概要：日本人の英語習得上の問題点の一つは、句動詞のように基本語を組み合わせた自然な表現の知識が乏しいことである。本研究では英語圏の出版社から刊行されている英語学習者用の読み物のシリーズと英語圏の初等教育用図書のシリーズからなる約1,036万語のコーパスを作成した。このコーパスは当初の狙いであった英語圏で日常的に使用される句動詞の分析に役立つだけでなく、その他の基本語彙からなる実用的な表現の研究や教材の開発にも活用することができる。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	1,400,000	0	1,400,000
2007年度	1,200,000	0	1,200,000
2008年度	800,000	240,000	1,004,000
年度			
年度			
総計	3,400,000	240,000	3,640,000

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：英語教育

## 1. 研究開始当初の背景

## (1) 句動詞の重要性と学習の難しさ

多くの日本人に共通する英語習得上の問題点の一つは、主としてやや硬い表現に知識が偏り、英語圏で日常的に使われているような基本語を組み合わせた自然な表現、中でも特に phrasal verb (句動詞) と呼ばれる表現の知識が乏しいということである。

これまでの英語教育において句動詞の指導が比較的手薄になってきた背景には、無数

にある句動詞の中で、どれを優先的に教えるべきかという目安になる情報が不足していたことがあると推測される。

句動詞を指導する上でもう1つ問題になる点として、同じ句動詞が実にさまざまな意味で用いられるということがある。例えば come in という表現をプログレッシブ英和辞典(第4版)で調べてみると「〈人が〉(家などに)はいる；〈光・音などが〉中にはいる」「〈電話が〉かかってくる；〈ニュースが〉(放送局などに)はいつてくる」「〈給料・金などが〉(…から)はいる」「(競走で) (順位が…

で) ゴールする」「《スポーツ》(特に打者が)試合に出る;《クリケット》打手となる」など、句動詞として13もの用法が載っている。このように一般によく用いられる句動詞に多数の用法があることは珍しいことではなく、これが学習者にとって句動詞の習得が難しい要因になっている。

さらに辞書のスペースの都合上、一般に句動詞の用例の掲載は少なくなっているのが普通である。このため、通常の辞典だけでは句動詞の用法の理解を深めにくい状況にある。

以上のことから、英語圏の日常表現としてよく用いられているにもかかわらず日本の英語教育では比較的注意を向けられてこなかった高頻度句動詞を特定し、頻度・意味・用法のデータを集めて、これらの句動詞を学習者が学べるような教材を作成して公開することができれば、英語学習者の発信型の英語力の向上に役立てることができるであろうと考えた。

## (2) 「英語学習者に有用な句動詞データ」を得るために必要なコーパス

実際の英語でどのような表現が多く使われているかを知るために言語研究・言語教育で利用が進んでいるのが British National Corpus や Bank of English のような大型コーパスである。「よく用いられる句動詞」を調べるためにこれらの巨大コーパスを用いることがまず選択肢として考えられる。

しかし一方、これらのコーパスに含まれているテキストは英語圏の新聞や書籍、官公庁の文書など、英語を母語とする成人の言語能力相応の英文であり、当然ながら日本の大多数の英語学習者の英語力とは釣り合わないという問題がある。

そこで考えたのが、英語を母語とするライターの書いた学習用図書のシリーズを多数コーパス化するという方法である。英語を母語とするライターが英語学習者や英語圏の児童を対象として比較的易しい英語でわかりやすく書いたテキストを大量に集めることで、英語圏では子供でも知っているような日常的な表現を、ほぼ基本語で書かれた理解しやすい文脈の中で見ていくことができる。また、学習用図書であれば学習者にとって不適切なスラングやレベルの高すぎる表現を避けることができる点も教育用には好都合である。

## 2. 研究の目的

本プロジェクトの最終的な目的は、句動詞の学習に役立つ、ウェブ上で利用可能な教材

( Web-Enhanced Language Learning [WELL]教材)を開発し、公開することにある。そのためには、日本の英語学習者にとって有用な句動詞を特定し、それらの句動詞の意味・用法を分析することが必要である。そこで、そのような句動詞の調査に役立つコーパスを作成することが第一に必要ということになる。つまり、作成したコーパスから高頻度句動詞のコンコーダンスを作成し、意味・用法を分析して教育・学習に役立つ句動詞データを作成するとともに、このデータに基づいて公開可能な例文を作成し、教材開発に役立てられるようにすることが本研究の目的である。

## 3. 研究の方法

まず、英語圏の出版社 (Cambridge University Press, Houghton Mifflin, Macmillan, Oxford University Press, Pearson Longman) から出版されている主要な graded readers と leveled readers のシリーズを中心にコーパスの作成を開始した。(詳細は次節に記載。)

具体的には、頁あたりの語数が多い図書については、1冊ずつ頁ごとに切り離し、両面が読み取れるスキャナーにかけて画像ファイル (PDF ファイル) を作ったのち、OCR ソフトで文字を読み取らせてテキストデータに変換するという作業を行った。

頁あたりの語数の少ない図書については上記の作業を行うよりも各頁のテキストをそのまま入力する方が効率が良いため、直接人手で入力するという方法を採用した。

前者の場合には OCR ソフトの読み取り・変換ミス、後者の場合には入力者のタイプミスによるデータの誤りが発生する。このような誤りの修正についてはワープロソフトのスペルチェッカーでは見逃されてしまうものが少なからずあり、できるだけ正確さを期すためには全文を人間が読んで確認していく必要があり、それでも少数のエラーが残る。このため当初予定していたよりもはるかに多くの作業時間がかかることになった。

書籍をコーパス化する際に問題になることの1つが行末のハイフネーションである。OCR で読み取った結果表れる行末のハイフンには、「文中であっても必要なもの」・「単語の途中で次の行に移るために現れた、本来は単語の一部ではないハイフン」・「ダッシュ記号が変換されたもの」の3種類があり、それぞれ異なる対処をする必要があるため、これも1つ1つ人間が読んで確認し、適切な処理を選択する必要があった。

ここまでの作業でもっとも基本的な形のコーパスを構築することができたことにな

るが、さらに POS (part of speech) タグを用いたより複雑な検索を可能にするために、POS タグを付与する作業を行う必要がある。この作業はタグプログラムを利用して行うが、その前段階としてテキストを文単位に改行で区切る作業が必要になる。この作業についても、正確を期すためには人が確認しながら行う必要があり、コーパスの規模が大きいため、一部のシリーズを除いて現在も作業を継続中である。

以上のようにして作成したコーパスに対して句動詞の検索やコンコーダンスの作成を行い、主要な graded readers や leveled readers における句動詞の使われ方やその意味などを検討した。

#### 4. 研究成果

本研究の成果として、合計約 1036.3 万語のコーパスを作成することができた。その内訳は表 1 のとおりである。このコーパスは大別すると 3 つのタイプのテキストからなっている。1 つは英語圏の出版社から刊行されている graded readers と呼ばれる使用語彙の段階別になっている英語学習者用の読み物のシリーズのコーパスであり、作成したコーパスの大部分を占める (約 971 万語)。もう 1 つは leveled readers と呼ばれる英語圏の初等教育用段階別読み物のコーパス (約 61.1 万語) である。残りの 1 つは leveled readers と同様に英語圏の児童向けのものだが、シリーズ内のレベル区分がなされていないものである。これは graded readers の大部分が物語や小説であるため、タイプの異なる factual な文章を補うために追加したものであるが、全体に占める割合はわずかである (約 4.2 万語)。

表 1 : 作成したコーパスの内訳

シリーズ名	レベル数	冊数	総語数
Graded readers			
Cambridge English Readers	7	68 冊	100 万語
Macmillan Readers	6	143 冊	181 万語
Oxford Bookworms Library*	7	175 冊	210 万語
Penguin Readers	7	359 冊	480 万語
Leveled readers			
Houghton Mifflin Leveled Readers**	6	354 冊	31 万語

Longman Literacy Land Info Trail	5	102 冊	12 万語
Longman Literacy Land Story Street	12	116 冊	8 万語
Oxford Reading Tree Fireflies	8	48 冊	3.3 万語
Oxford Reading Tree Storybooks	9	160 冊	4.5 万語
Macmillan Springboard	16	128 冊	2.3 万語
Other readers			
Little Tiger Press Nature Storybooks	1	8 冊	0.7 万語
Scholastic Rookie Read-About Geography	1	51 冊	1.6 万語
Scholastic Rookie Read-About Holidays	1	33 冊	1 万語
Scholastic Time-to-Discover	1	60 冊	0.9 万語

\*Oxford Bookworms Library は 2008 年にシリーズ構成の変更があったが、本研究は 2006 年のテキスト入手当時の分類に基づく。  
\*\*Houghton Mifflin Leveled Readers の一部についてはテキスト修正を継続作業中。

コーパスの大部分を占める graded readers のレベル別の内訳は表 2 のとおりである。

表 2 : Graded readers のレベル別テキスト数

Level (Stage)***	C	M	O	P
Starter/Easystarts	8 冊	/	21 冊	33 冊
Level 1	11 冊	13 冊	25 冊	34 冊
Level 2	10 冊	32 冊	32 冊	83 冊
Level 3	11 冊	33 冊	31 冊	82 冊
Level 4	10 冊	11 冊	30 冊	49 冊
Level 5	10 冊	41 冊	19 冊	48 冊
Level 6	8 冊	13 冊	17 冊	30 冊

C: Cambridge English Readers  
M: Macmillan Readers

O: Oxford Bookworms Library

P: Penguin Readers

\*\*\*Oxford Bookworms Library ではレベル区分の表示に Stage という語を用いている。(他の3シリーズは Level 表記。)

主要な graded readers および leveled readers のコーパスを作成したことにより、これらのシリーズにおける任意の表現を検索することが可能になった。例えば、Oxford Bookworms Library では fall off の使用例が屈折形を含めて 70 例あることが容易に調べられる。いずれも「落ちる」という基本的な意味で用いられた例であり、このうち自動詞の用法と「～から落ちる」という他動詞の用法とがちょうど同数の 35 例ずつみられた。

まず自動詞の用例の一部を挙げると以下のようなものがある。注目すべき箇所には下線を施している。(本節の用例はすべて Oxford Bookworms Library からの引用。出典は本節末尾を参照。)

- ① Dorothy's left shoe *fell off*.
- ② At the kitchen door one of his slippers *fell off*.
- ③ His crown nearly *fell off* because he was shaking so much.
- ④ [...] sometimes pieces of wood *fell off* into the sea.
- ⑤ When the horse went on again, he *fell off* behind.
- ⑥ Sometimes he *fell off* sideways as well, [...]

これらの例から、fall off の主語は人でも物でも良いことが示される。ここでは特に靴やスリッパと言った履物も fall off の主語になることがわかる。この場合、日本語では「脱げる」と表現することになるが、「(靴を)脱ぐ」は take off と習っていても「脱げる」を fall off で表現できることを習ったことのない学習者は多いものと思われる。また、⑤⑥の例からは「後ろに落ちる」「横に落ちる」など、落ちる方向を示す表現ができることも知ることができる。

次に、他動詞の用例の一部を挙げると以下のような例が見られる。

- ⑦ Did you *fall off* the bed?
- ⑧ I've *fallen off* a river boat.
- ⑨ A picture *fell off* a shelf, [...]
- ⑩ A man called Burke *fell off* a ladder outside his home.
- ⑪ The surprised officer nearly *fell off* his chair.
- ⑫ [...]the water was still *falling off* the leaves of the trees.

- ⑬ You've never *fallen off* a high building in your life.
- ⑭ The red shoes *fell off* her feet, [...]

このようにいくつもの実例から、fall off X で「Xから落ちる」と表現できることを無理なく示すことができる。また、⑭の例からは自動詞用法と同様に、他動詞用法でも履物を主語として表現できることがわかる。

英和辞典では紙面の厳しい制約から掲載される例文の数はわずかであるが、ここで示したように、コーパスから得たわかりやすい実例をいくつも読むことで学習者が句動詞の用法の理解度を高めることができるようになるという効果がある。

ここで強調したいことは、このコーパスから得られる例文のほとんどどれを選んでも一般的な学生にも比較的容易に意味が理解できる英文になっているということである。これは学習者に示す例文として非常に重要な点であり、既存の巨大コーパスでは得られない、本研究で構築してきたコーパスならではの利点であると言える。

POS タグを付与した部分については POS タグを利用して、例えば「動詞+前置詞」を検索することで句動詞を集中的に集めることができるようになることを期待していたが、実際に試してみると、「動詞+前置詞」の組み合わせではあるが句動詞ではないものや、タガーで付与された POS 自体が間違っていて句動詞に該当しないものなども多数ヒットしてしまうという問題があることがわかった。これらについてはやはり人間が 1 例ずつ判断して除外していく必要があるため、正確な情報を得るには事前に予想していたように機械的に高頻度句動詞を網羅的に特定できるというわけにはいかなかった。そのため、高頻度句動詞の特定やその結果に基づく WELL 教材の開発については今後の進展を待つことになったが、それまでの間は上述の fall off の例のように、POS タグのないコーパスで有用な例文が多数得られる句動詞について調査を進めていきたい。

著作権の問題があるため、コーパスから得た例文をそのまま WELL 教材に使用することはできないが、著作権の侵害を行わないように注意を払いながら、作成したコーパスから得られた知見を教育に利用することができるよう、ウェブなどで公開していく予定である。

一般に多くの日本人が英語で自分の言いたいことを十分に表現できないのは語彙力がないためと考えられる傾向がある。しかし、英語学習者用読み物という、限られた数の基本語で書かれている本が多数あり、その本で多くの人が楽しんだり感動したりしているという事実からもわかるように、基本語彙の

用法の理解を深めることで英語の表現力を相当な程度まで向上させることができる。その意味でも **graded readers** や **leveled readers** の存在は英語学習者に大きな希望を与える教材であると言える。

近年、特に多読・多聴学習の教材として graded readers や leveled readers の利用が全国的に一層の広がりを見せており、学習や教育に役立てるため、これらのテキストにおける言語使用についてのより詳細なデータを求めるニーズも生まれてきているようである。本プロジェクトで得られる知見はそのような要請にも応えられるものになるだろう。今後も作成したコーパスに手を加えながら、多角的な活用を進めていきたい。

[用例の出典]

●すべて Oxford Bookworms Library シリーズ (Oxford University Press) のタイトルである。

- ①⑭ *The Wizard of Oz* (Stage 1) written by L. Frank Baum, retold by Rosemary Border
- ② *The Monkey's Paw* (Stage 1) written by W. W. Jacobs, retold by Diane Mowat
- ③⑤⑥ *Through the Looking-Glass* (Stage 3) written by Lewis Carroll, retold by Jennifer Bassett
- ④ *Grace Darling* (Stage 2) written by Tim Vicary
- ⑦ *The Withered Arm* (Stage 1) written by Thomas Hardy, retold by Jennifer Bassett
- ⑧ *Huckleberry Finn* (Stage 2) written by Mark Twain, retold by Diane Mowat
- ⑨ *Meteor and Other Stories* (Stage 6) written by John Wyndham, retold by Patrick Nobes
- ⑩ *Deadheads* (Stage 6) written by Reginald Hill, retold by Rosalie Kerr
- ⑪ *The Silver Sword* (Stage 4) written by Ian Serraillier, retold by John Escott
- ⑫ *The Year of Sharing* (Stage 2) written by Harry Gilbert
- ⑬ *Love Story* (Stage 3) written by Erich Segal, retold by Rosemary Border

## 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)

[その他] (計2件)

(研究発表)

高見 敏子、「Graded readers に見られる語彙分布の特徴」、統計数理研究所言語系共同研究グループ研究発表会「言語と統計 2009」、2009年3月15日、統計数理研究所

(公開講座)

高見 敏子・原田 真見、「英語圏の児童向け学習図書で学ぶ英語表現」、北海道大学大学院メディア・コミュニケーション研究院公開講座、2009年5月11日・5月18日・6月1日・6月29日、北海道大学

## 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