

英語句動詞用例サイトの開発

佐野 洋

東京外国語大学 外国語学部

〒183-8534 東京都府中市朝日町 3-11-1

e-Mail : sano@tufs.ac.jp

概要

外国語学習は言語知識の学習に加えて、技能科目であることから学習した言語知識を、その知識が含まれる用例を使った運用練習が不可欠で且つ重要である。いわゆる語法の習得や語感の獲得に関係するので練習用例数は多いほどよい。文法学習書や辞書では、学習項目毎に、その用例が、紙媒体という記録制約もあって1例しか掲載されていないことが多い。自然言語処理技術を用いた、コーパスからの英文抽出によって多数の用例を得ることができる。またウェブインタフェース技術を用いることで、インターネットを通じた効果的な運用練習が実現する。筆者は、学習項目用例を英語コーパス(British National Corpus)から抽出した。提示システムは、運用練習で効率的に学習できるよう語法上の気づきを促進する仕組みを、Ajax フレームワークを用いて開発している。

1. はじめに

1.1 研究目的

筆者は、現代的なメディア技術を使った学習効果を向上させる教材開発の方法を研究開発している[1,2]。その一つとして、コーパスや自然言語処理技術とウェブインタフェース技術を活用した教材作成に取り組んでいる。2005年には、著者の研究室と(株)小学館コーパスネットワークの共同研究により英語文型教授のための教育用英語用例サイト(1.2節を参照)を開発した。

本稿は、英語句動詞用例の英語コーパスからの抽出とそれら用例を提供する仕組みについて述べる。既に公開している英語文型教授のためのサイトが、目次方式のインタフェースを提供しているのに対し、英語句動詞用例の提供では、索引方式のインタフェースを提供することが特徴である。

初学者にとって、一つの用例の中に複数の学習項目が含まれる場合、習得の負担になる。一方、中級・上級の学習者には、複数の学習項目を含む用例が利用できることは、一度に複数の語法が学べるから効率的な学習教材となり、学習効率を向上させることができる。

句動詞用例について、用例提示の際に複数の言語特徴を明示的にインタフェースで示すことができるように、(1) 用例検索の手続きを工夫し、用例抽出を行っている、(2) 同時に Ajax フレームワークを使

ったインタフェース実装を行っている。

なお、句動詞用例集は、2006年9月に(株)小学館コーパスネットワークから公開予定である。

1.2 英語文型教授のための用例サイト

2005年度、著者の研究室と(株)小学館コーパスネットワークの共同研究により英語文型教授のための教育用英語用例サイト¹を開発した。本用例サイトには、BNC(British National Corpus, 1億語)から、コンピュータを使って自動抽出した1,320項目の英語文型用例(およそ80万例文)がアップされている(2005年9月に公開済)。



図 1 用例サイトの様子

¹ <http://www.corpora.jp/>

文法項目名(144 種類)と文の種類(14 種類)をメニューで指定すると、対応する用例がブラウザ画面に表示される(図 1)。

1.3 句動詞用例

句動詞(Phrasal Verbs)は、動詞と前置詞もしくは副詞の連鎖で構成される動詞相当語句である。句動詞は、簡単な動詞のあとに前置詞や副詞を付けて、さまざまな意味を表すことができるので、口語表現を中心に使用頻度が高いといわれている。例えば、set apart, set aside, set at, set before, set by, set in, set off, set over のような語句である。

冠詞や仮定法などと同様に、日本語文法の品詞枠組みに、前置詞や句動詞に対応する文法カテゴリがないので、一般に学習や習得が難しい項目であるともいわれている。句動詞の学習のための説明や用例は、前置詞の項目にもあり、動詞の項目でも現れわれ、しばしば副詞の説明の中にもある。

句動詞別の用例集は、使用頻度の高い表現を学習するための教材として重要である。

1.4 用例集と用例提示の方法

1.2節で示した用例サイトは、学習文型が目次としてウェブ画面に提示され、目的の用例を探す仕組みである。これは、目次方式の検索インタフェースである。

目次方式のインタフェースに対して、一方、索引方式もある。目次方式の代表は、よく知られているように YAHOO!™で、人手で整理した分野別一覧から目的のページを探す仕組みになっている²。

一方、検索方式の代表は Google™である。2002年に Google が、キーワードとそれに関連する広告を表示する仕組みを提供して以来、ネット広告を通じた購買確率が上昇した。学習心理的に考えると、検索連動型の広告は、検索結果に複数の概念を提示することに相当し、連想記憶を活性化させる。語学学習上の事柄に言い換えると、語法上の気づきを促進していると考えられる。

目次方式の検索は、初心者にも使いやすいが細かい検索が難しい。図 1 で示す用例提示では、文型(144 種類)が基本的な用例選択の手段である。

句動詞用例の提示を考える。目次方式を想定する

と、アルファベット順に句動詞(語形)が配置されることになり、単純目次で且つ、語数が多くなるほど、選択が面倒となり使いづらい。用例以外には、基本的に語形が提示されるだけで、工夫しても句動詞の訳を表示できるくらいである。したがって、句動詞用例の提示では、句動詞の特徴を他の情報から得て、言語的な特徴を使ったガイドを利用者へ提示することで用例を絞り込めるようにすることが必要になる。

以下では、BNC からの句動詞用例抽出の方法と、それら抽出された句動詞用例を提供するインタフェースについて説明する。

2. 句動詞用例抽出

筆者は、句動詞見出しについて文型記述を基に CQL³式への展開を行った[2,3]。CQL 式に展開することで BNC から用例を自動的に抽出することができる。

2.1 補語文型

補語文型は、目的語や補語の種類を規定するものである。『英語句動詞文例辞典』[4]を用いて句動詞データを作成した。[4]は、句動詞を単に見出し語リストとして収集したのではなく、以下の考え方で分類している。

- 前置詞と副詞を区別しない。前置詞と副詞はパーティクル(particle : 不変化詞)と呼ぶ
- 動詞の目的語やいわゆる前置詞の目的語に下位分類を設定する(名詞、代名詞、it など)
- 37 文型を句動詞に設定している

[4]によると、補助部を構成する要素は 7 種類である。なお、明示的に指定される 7 要素以外の要素という指定、各要素が義務的に必要な場合と、任意の場合があること、さらに語順の違いがあることから 37 の文型があるとする。なお、文型の CQL 式への展開と用例抽出については、[2]に詳しい。

2.2 前置詞/副詞の意味分類

[4]では前置詞/副詞を、「テーマ」という概念で分類している。テーマの例を挙げる(表 1)。テーマの種類は 33 で、さらに幾つかのサブテーマがある。テーマが前置詞/副詞の基本的な意味であって、その

³ CQL は、(株)小学館コーパスネットワークが提供するコーパス検索言語(Corpus Query Language)である。この言語様式を使って文型を記述し、コーパス検索エンジンを通じて BNC から用例抽出を行う。

² 現在は、索引方式も併用されている。
<http://www.yahoo.co.jp/>

意味分類の中に句動詞が配置されている。これを元に個々の句動詞エントリーとテーマの対応データを作成した。

表 1. テーマの例([4]から引用して一部を掲載)

	テーマ
1	動き(狭い範囲: 小さい/一時的な)
2	動き(広い範囲: 大きい/連続的な/目的を持った)
3	接近・到着・訪問
4	出発・去ること・発送
5	位置・(非)存在
6	接触・発見・近隣
7	攻撃(言葉による/物理的な)
...	...

2.3 動詞部文型

動詞部文型とは、動詞述部の定形・非定形パターン。主語を有する動詞述部の類型を展開したものである。複数の文法書[5,6,7,8]に従って動詞部文型の調査を行い、動詞の定形・非定形パターン記述を行った。その記述を基にして検索式(CQL式)を作成した。

定形動詞が現在形で肯定の場合、およそ 30 の定形パターンがある。英語の場合、主語をとることのできる定形の種類が現在形、過去形の 2 種類あり、さらに否定があるので、約 4 倍のパターン数になる⁴。例えば、“[have, has] to have been Ving” は 5 語から構成される動詞部文型である。ちなみに CQL 式に展開すると以下ようになる。なお、文型の CQL 式への展開については[2]に詳しい。

```
{P="N.*|PN.|CRD"}{W="have|has"}
P="VHB|VHZ"}{W="to"} P="TOO"}{L="have"}
P="VHI"}{W="been"} P="VBN"}{L="@Var1"} P="V.G"}
```

2.4 用例検索と蓄積・配信

句動詞ごとに BNC から用例を検索するための CQL 式を作成した。文型の違いや定形部分の文型の違いによって変化するが、一つの句動詞用例を検索するには、平均的に 50~100 行程度の CQL 式が必要になる。この CQL 式は、CQL スクリプトとしてコーパス検索エンジンで実行され、英語コーパス(BNC)から用例が抽出される。

図 3 には、作業工程の概念図を示す。図中の 1,2 の工程を終え、3 の抽出作業を行っている。4 は、データベースである。用例データを XML データで蓄積管理することが目的でインタフェースからの表

示要求を受けて、用例を配信する。5 のインタフェースについては、現在、設計を実施している。一部プロトタイプを作成を行っている。

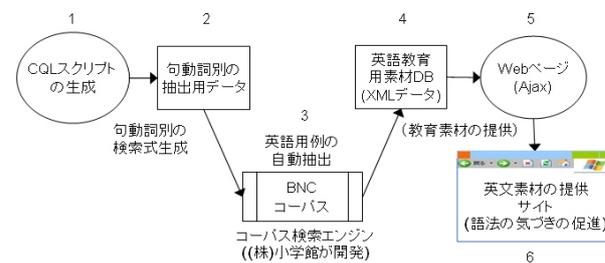


図 2. 作業工程の概念図

3. 用例提供のインタフェース

3.1 電子化データとその種類

[6]をもとに、句動詞の見出し語と対応する文型情報を電子化し、CQL 式を作成した。前置詞/副詞(不変化詞)と対応するテーマ、サブテーマ名も電子化し、用例検索結果と対応がとれるようにした。動詞部文型も整理し電子化したので、最終的に、抽出結果としての各用例には、表 2に示す電子データが抽出用例に対応する。

表 2. 電子化データ

句動詞	数	言語上の制約
見出し語数	約 4300	語形、訳
補助部文型	37	文型、語彙分布(目的語、補語)
動詞部文型	約 100	文型、語彙分布(主語)
前置詞/副詞	46	語形、テーマ、サブテーマ

句動詞用例の提示に目次方式を想定すると、アルファベット順に句動詞(語形)が配置され、語形と文型で制限された用例が表示される。先にも述べたように、見出し語数が 4000 以上もあるので使いづらいし、例えば”set up”の訳を表示しても他の語法の気づきに結びつかない。例えば、“up”の意味分類や他の”~up”の句動詞との共通性や違いなどである。

不変化詞”up”の意味分類(「テーマ」)や”set up”の意味分類(「サブテーマ」)、”set up”の動詞部文型ごとの用例表示など、表 2で示すさまざまな関連データを活用した索引方式のインタフェースが必要になる。

3.2 Ajax によるインタフェース実装

現在、Ajax でのインタフェース実装を目指し、デザインとプロトタイプの開発を行っている。

Ajax フレームワークは、サーバ(本稿の場合、用例サイト)との通信結果に応じて XML データとやり

⁴ さらに受動形もあるので、動詞部文型を網羅すると、150 以上の文型パターンがある。

とりを行い DHTML で動的にページの一部を書き換えるインタフェースを実現するというアプローチを取る。用例は、Ajax に対応するようデータベースから配信する。

33 項目あるテーマは図形化することで、一画面での操作が可能になる。同画面に、索引入力の入力ボックスを設ける予定である。基本的な検索インタフェースは、Google™サジェスト風に、(1) 句動詞見出し語の入力支援、(2) 前置詞/副詞の入力支援を行う予定である。選択肢の入力支援及び選択肢の候補の自動リストアップを行いながら、利用者が求める用例にアクセスできるようにしたい。

3.3 実装予定

句動詞用例集は、2006 年 9 月に(株)小学館コーパスネットワークからの公開を予定している。

用例を格納するデータベース設計とプロトタイプの開発は終わっている。テーママップ画面の基本設計は終了し、テーマのイメージを作成中である。また、選択肢の入力支援及び選択肢の候補の自動リストアップ部分を設計している。

図 3 と図 4 はプロトタイプとして作成中のトップページとテーママップである。テーマに対応する図式化は未だである。イラストを作成し、デザイナーに発注することになっている。

4. おわりに

本稿は、句動詞用例の提供について述べた。

用例集を、語学学習に活かすには、利用者が、現時点で(学習項目上)どの視点で検索をしているかが意識でき、且つ、用例を絞り込むための他の視点をナビゲートできるようなインタフェースが理想である。



図 3. 開発途中の用例検索サイト(プロトタイプ)

中級あるいは上級者には、一度に複数の語法が学べるから効率的な学習教材となり、学習効率を向上させることができるだろう。



図 4. 開発途中のテーママップ(プロトタイプ)

参考文献

- 1) 佐野洋:「ESP 適合の教材コンテンツを実現する語学教育支援システム」,『最新外国語 CALL の研究と実践』, コンピュータ利用教育協議会 (CIEC)・外国語教育研究部会 (34~44,10 頁),2003 年 3 月.
- 2) 佐野洋:「BNC を利用した教育用英語句動詞用例集」, コンピュータと教育研究会, 2006-CE-84, 1pp-6pp, 2006 年
- 3) Nakamura, T., Tateno, J. and Tono, Y. (2004) Introducing the Shogakukan Corpus Query System and the Shogakukan Language Toolbox. Williams, G. and Vessier, S. (eds) EURALEX 2004 Proceedings . The Eleventh EURALEX International Congress, July 6-10, 2004, Lorient, France, pp. 147-152.
- 4) クリストファーバーナード,『英語句動詞文例辞典』, 研究社, 2002.
- 5) 綿貫陽他,『ロイヤル英文法』, 旺文社, 2004.
- 6) 安井稔,『英文法総覧』, 開拓社, 2003.
- 7) 鈴木英一,『統語論』, 開拓社, 1995.
- 8) AJ トムソン/AV マーティネット, 江川泰一郎訳注,『実例英文法 第 3 版』, オックスフォード大学出版局, 1983.