

## BNC を利用した教育用英語句動詞用例集

佐野 洋

東京外国語大学 外国語学部

sano@tufs.ac.jp

### 概要

(株)小学館マルチメディア局と共同で、BNC(British National Corpus)を用いた、教育用の英語句動詞の英文用例集を作成している。筆者は、およそ 4300 に及ぶ句動詞見出しについて、文型記述を行い、CQL 式への展開を行った。文型記述の特徴は、動詞見出しを単純な動詞句としないで定形(finite form)パターンに展開したことである。この特徴によって、BNC を通じた用例抽出によって、主語と句動詞の関係分析(主語の性質とモーダル表現)が可能になった。本稿では、文型記述の詳細と CQL 式への展開行程について報告する。

## Retrieval of phrasal verbs from the BNC

### A new approach to quantitative study for English education

SANO, Hiroshi

Faculty of Foreign Studies, Tokyo University of Foreign Studies

[sano@tufs.ac.jp](mailto:sano@tufs.ac.jp)

### Abstract

This paper presents a new grammatical approach on corpus analysis and the application process to the CQL or corpus query language. The aim of this study is to improve the current status of qualitative English language learning.

With this approach, a relational analysis between subjective complements and phrasal verbs has been realized. To get authentic example sentences, retrieval styles with the CQL for 4300 phrasal verbs are made and executed.

The feature of this analysis is to focus on finite forms, whereas most of the existing sentence pattern studies concentrate on object complements. These finite forms include not only tense and aspect but also possible forms of modality, such as "might have been Ving".

The result provides a clue to quantitative English language learning. The extracted example sentences will be released shortly through Shogakkan Corpus Network. This study is conducted by a collaboration of Shogakkan multimedia department and Sano laboratory.

# 1. まえがき

## 1.1 研究目的

一般に教材作成は、教授者の専門分野の知識、教授経験や教授手法とその表現の力量に依存する部分が多い。仮に学習項目や学習対象者が同じであっても、複数の教授者が作成した教材の内容と品質は、作成者ごとに異なっている。もちろん教授設計法に従い、記述のひな型やガイドラインを用意することで、こうした事態を避けようとする教育現場も多いだろう。

筆者は、教材作成の問題点を『教材の作成が、教授者の専門分野の知識、教師としての経験、教授手法に依存するため、教材内容と教材品質を一定に維持することが難しい。』と捉えている。そして、このことから、教材作成のプロセスが、一般に労働集約的であって、教材開発にかかるコストが高くなっていると考えている。

理屈上は、一つの教材で教授する学習者を増やすことや教材内容を再利用することで、教育事業全体としての収支の均衡を図ることができる。しかし、教材作成の分野を英語教育に限定すると次の点が問題として挙がる。

- 日本人教授者として、英語母語話者のようには語感が働き難く<sup>1</sup>、教材品質向上には、言語の運用実態を広範に捉える手立てが必要である。
- 医療分野、金融ビジネス分野や IT 分野など多様な教育ニーズがあり、専門分野の英語に対応する教材作成を支援する仕組みが必要である。

筆者は、現代的で効率の良い教材開発の方法を研究開発している[1]。その一つとして、言語運用データと自然言語処理技術を活用した教材作成に取り組んでいる。

## 1.2 英語文型教授のための用例サイト

2005 年度、筆者の研究室と(株)小学館コーパスネットワーク<sup>2</sup>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を開発した。本用例サイトには、BNC(British National Corpus)から、コンピュータを使って自動抽出した 1320 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)。

メニューで示される文法項目名を指定すると該当の文型が含まれる用例がブラウザ画面に表示される。それら用例をダウンロードする

ことも可能である(CSV 形式で保存される)。図 1 に、公開されている用例サイトの様子を示す。



図 1. 用例サイトの様子

## 1.3 句動詞用例

筆者はさらに、学習文型用例だけでなく、多様な学習視点に基づく用例集の提供を進めている。現在、句動詞用例を収集している。

簡単な(且つよく使われる)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句動詞は、「動詞 + 前置詞」あるいは「動詞 + 副詞」を一つのまとまった表現として考えられる。句動詞は、口語表現を中心に使用頻度が高いといわれている。

一方、一般に文法書では、句動詞についての説明は、比較的后方に位置する<sup>3</sup>。単独で文法説明項目にならずに前置詞の説明項目に組み入れられていることもある。さらに、前置詞や句動詞は、日本語には、対応する文法カテゴリがないので、学習が難しい項目であると予想される。最近では、イメージ解釈やメタファー解釈によって前置詞を理解しようとする試み[2]もある。

こうしたことから、句動詞別の用例は、使用頻度の高い表現を学習するための補助教材として重要であると考えられる。

## 1.4 句動詞用例サイトの開発

筆者は、およそ 4300 におよぶ句動詞見出しについて、文型記述を基に CQL 式[3]への展開を行った。CLQ 式に展開することで BNC から用例を抽出することができる。文型記述の特徴は、動詞見出しを単純な動詞句としないで定形(finite form)パターンに展開したことである。この特徴によって、用例抽出の結果から、主語と

<sup>1</sup>社会的文化的な背景知識もあって、言語体系全体に対する認識や実際の運用についての認知が難しい。

<sup>2</sup> <http://www.corpora.jp/>

<sup>3</sup> 『ロイヤル英文法』(改訂新版, 旺文社, 2004)[4]でも、わずか 3 頁(406~408)の記述しかない。

句動詞の関係分析(主語の性質とモーダル表現)が可能になった。

本稿では、文型記述の詳細と CQL 式への展開行程について報告する。以下、2 章で句動詞文型とその CQL 式記述について説明する。3 章では、定形パターン展開について説明し、句動詞表現の最終的な CQL 式記述の方法について述べる。4 章は、まとめと句動詞用例サイトのインタフェース・イメージについて構想する。

## 2. 句動詞文型

### 2.1 句動詞の種類

句動詞一覧をデータ化するために参照した『英語句動詞文例辞典』(クリストファバーナード著, 文献[5])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、且つシンボルグラマーと呼ばれる独自の方法で 37 の文型を認めている<sup>4</sup>。

- 前置詞(preposition)と副詞を区別せず、パーティクル(particle: 不変化詞)とする
- 補部(動詞の目的語やいわゆる前置詞の目的語)の下位分類を設ける
- 補部を義務的にとる文型と任意にとる文型に分ける
- 上記の方針に沿って文型分類し、文型が異なれば、インデックスが同じでも異なる句動詞とする

表 1 には、パーティクルの種類と、そのパーティクルを構成要素とする句動詞の数(種類)を示す。なお、it は”take it out”である。また、例えば、”take up with”は、”up”にカウントされている。

表 1. パーティクルと句動詞の数

不変化詞	数	不変化詞	数
about	12	forward	14
above	4	from	29
across	11	in	296
after	17	into	179
against	40	it	1
ahead	18	of	42
along	33	off	396
among	4	on	305
apart	16	onto	10
around	146	out	626
as	8	over	145
aside	23	overboard	1
at	58	past	3
away	202	round	6
back	133	through	89

<sup>4</sup> 一番のポイントは、句動詞が意味的な分類基準に基づいてまとめられていることである。

before	6	to	76
behind	15	together	38
below	1	towards	9
beneath	1	under	14
between	3	up	764
beyond	1	upon	4
by	30	with	73
down	325	without	3
for	78		

[5]では、動詞に後続する要素(いわゆる述語要素)を句動詞の特性によって 37 の類型を分類している。[5]によると、補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。

表 2. 動詞に後続する要素の種類

後続要素	シンボル
名詞(名詞句も含む)	
代名詞	
名詞または代名詞	
再帰代名詞	self
仮目的語の it	it
動詞 ing 形	+ Ving
to 不定詞形	+ to do
(上記以外の)他の要素が後続	+

表 2 のシンボルは、シンボルグラマー[5]記述に使われるシンボルである。をパーティクルで表現し、”(...)”は、任意要素であることを表現する。例えば、表 3 に示すような表現方法で句動詞の文型が表現される。なお、参考のために英語文法書にみられる表現を対応させている。

表 3. シンボルグラマー表現(一部)

シンボル グラマー表現	英語文法書にみられる表現
	他動詞+名詞/代名詞+副詞
	他動詞+代名詞+副詞
self	他動詞+再帰代名詞+副詞
( )	自動詞+前置詞(+名詞)
self	自動詞+前置詞+再起代名詞
+ Ving	自動詞+前置詞+動名詞

## 2.2 CQL 式への展開

### 2.2.1 CQL 式

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。教育用英語用例サイト開

発では、1320 項目の英語文型すべてを CQL 式で記述し、用例抽出を行った。

例えば「how を使った感嘆文」であれば次のような記述になる。

```
^(W="how") {P="AJ0|AV0"} [0,10] {L="!"}$
```

この CQL 式は、『"How" という単語で文が始まり、形容詞もしくは副詞が続き、0 個以上 10 個以下の単語を間に挟んで"! "で終了する文を検索』することを意味する。

## 2.2.2 XML による検索式記述

シンボルグラマーにおいて、” ”記号は、名詞または代名詞を表現する。少なくとも検索式が 2 つ必要になる。そして、それぞれ抽出した用例の和集合が必要である。また、再起代名詞が分類として存在していることから、名詞として抽出した用例から再起代名詞を含む用例を除く必要もある。

CQL には、集合演算も用意されている。用例抽出結果に対して、ファイル単位での集合演算(和、積、差)を行う。例えば、文頭付近の特徴で抽出した結果ファイルと、文末付近の特徴で抽出した結果ファイルの和集合ととることで、文頭と文末に特徴がある文型に対応する用例を取り出すことができる。

文型記述の観点から考えると、用例抽出結果のファイルを意識して、検索式記述を行うのは、記述式が複雑になったり、記述数が増えたりするので面倒である。

我々は、教育用英語用例サイト開発の経験から、文型パターンの CQL 記述作業の効率化を目指し、集合演算も組み入れた形式で、CQL 式を記述できる枠組みを、XML を利用して作成した(XCQL フォーマット)。現在、この記述形式を使って、句動詞の文型記述を行っている。

例えば、” ”のシンボルグラマー表現(自動詞+前置詞+名詞/代名詞)は、リスト 1 に示す記述になる。

```
<item number="3" name=" " ptn=" "
isa="Phrasal Verb Construction">
<or>
<cql><finite>{L="@Var1"
P="VV.*"}</finite><prp>{L="@Var2"
P="AVP|PRP"}
{P="N.*|PN.|CRD"}</prp></cql>
<cql><finite>{L="@Var1"
P="VV.*"}</finite><prp>{L="@Var2"
P="AVP|PRP"}
{P="NN.|ATO|AJ.|DPS|DT0|CRD|ORD|AV0"}
{P="N.*|PN.|CRD"}</prp></cql>
<cql><finite>{L="@Var1"
P="VV.*"}</finite><prp>{L="@Var2"
P="AVP|PRP"}
{P="PNP|PNI|DT0"}</prp></cql>
</or>
</item>
```

リスト 1. XML を利用した検索式記述

XML のタグ表現を用いて、<cql>CQL 式</cql>だけでなく、和集合演算<or> ~ </or>、積集合演算<and> ~ </and>や差集合演算<diff> ~ </diff>を記述できるようにした。なお、<finite> ~ </finite>タグについては後述する。

変数記述の@Var1,@Var2 に例えば、句動詞エントリーの”account for”を代入して、式展開を行うと、リスト 2 で示す CQL スクリプトとなる。この CQL スクリプトをコーパス検索エンジン上で実行すると、スクリプトの各 CQL 式が解釈されて用例が抽出される。

fcql はコーパス検索エンジンに対して、CQL 式の解釈・実行を指示するコマンドである。cql-or は用例抽出結果ファイルを対象に和集合演算を指示するコマンドである。

```
# ITEM NUMBER 3 Copyright (C) 2005-2006 by
SANO-LAB
fcql '{P="N.*|PN.|CRD"} {L="account" P="V.B|V.Z"}
{L="for" P="AVP|PRP"} {P="N.*|PN.|CRD"}' > naha3.1
fcql '{P="N.*|PN.|CRD"} {L="account" P="V.D"} {L="for"
P="AVP|PRP"} {P="N.*|PN.|CRD"}' > naha3.2
fcql '{P="N.*|PN.|CRD"} {L="account" P="V.B|V.Z"}
{L="for" P="AVP|PRP"}
{P="NN.|ATO|AJ.|DPS|DT0|CRD|ORD|AV0"}
{P="N.*|PN.|CRD"}' > naha3.3
fcql '{P="N.*|PN.|CRD"} {L="account" P="V.D"} {L="for"
P="AVP|PRP"}
{P="NN.|ATO|AJ.|DPS|DT0|CRD|ORD|AV0"}
{P="N.*|PN.|CRD"}' > naha3.4
fcql '{P="N.*|PN.|CRD"} {L="account" P="V.B|V.Z"}
{L="for" P="AVP|PRP"} {P="PNP|PNI|DT0"}' > naha3.5
fcql '{P="N.*|PN.|CRD"} {L="account" P="V.D"} {L="for"
P="AVP|PRP"} {P="PNP|PNI|DT0"}' > naha3.6
cql-or naha3.1 naha3.2 naha3.3 naha3.4 naha3.5
naha3.6 > naha3.7
# ITEM NUMBER 3, See the file naha3.7 (max file
suffix -> 7)
cp naha3.7 result/naha3
# End of Script
```

リスト 2. 展開された CQL 検索式(CQL スクリプト)

従来、抽出結果のファイルを明示的に意識すると同時に、最終の抽出結果を得るためにファイル間での集合演算を行っていたが、この記述形式を採用することによって、煩雑な手間を省くことができている。

## 3. 定形パターンと用例抽出

### 3.1 定形・非定形

必要に応じて主語との一致を示す動詞の形を定形(finite form)という。動詞の形が、その主語によって定まるかどうかを基準とする分類である[6,7]。英語では、一般動詞は、現在形と過去形だけが主語と結びつく形である。過去分詞形、-ing 形、不定詞は非定形(non-finite form)と呼ばれている。

一般動詞の非定形は、助動詞(be, have, do, will 等)や助動詞相当句(be going to など)と結合し、その結合形の用法は複雑である。結合形を構成したとき、本動詞は、結合形の最後に位置し、非定形をとり、実質的な意味を担う。主語と結びつく定形の助動詞は、時間関係や話し手の心的態度(モーダル表現)を担っている。

図 2には、述語の構成イメージを示す。シンボルグラマーで表現される 37 の文型は、主動詞とパーティクルを含む(複数の)補語の構造を記述したものである。なお、"may have been Ving"のように、非定形の助動詞が連続することもある。

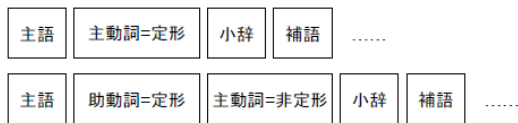


図 2. 述語の構成概念図

### 3.2 定形パターンの展開

例えば、"account for"は、"account"が定形で使われることが多く、しかも現在形の場合が多い。それに対して、"acquaint with"は、しばしば辞書などで"be acquainted with"として語法説明がされるように、"acquaint"は非定形(過去分詞形)をとり、実質的な意味を担うけれども、主語と結びつくのは、be-動詞である。be 動詞は定形をとり、BNC から抽出した用例を見ると、過去形(was/were)が使われていることが分かる。

述語の定形パターン展開を行うことで、抽出用例の細かな分類ができる。そして、句動詞ごとの語法観察の視野がより広く、精緻になると考えられる。そこで、定形パターンの展開を行い、CQL 式で記述した。

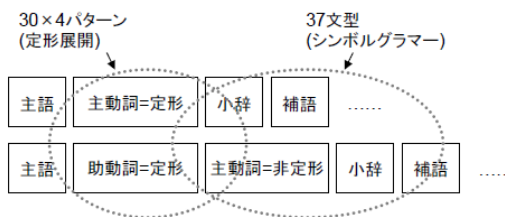


図 3. 定形パターン展開の概念図

図 3には、述部の定形パターンの展開の概念図を示す。30(定形動詞が現在形で且つ肯定の場合)の定形パターンがある。ただし、助動詞相当語句の認め方などでパターン数(30)は増減する。定形の種類が現在形、過去形の 2 種類あり、

述部には、否定辞を含めることができるので、4 倍のパターン数になる<sup>5</sup>。

```
<pattern number="1" name="Finite Verb Construction"
version="1(2006/03/06)">
<!-- 主語 + 時制動詞(定形動詞句) -->
<item number="1" name="A1" ptn="主語 + 現在形"
isa="Finite Verb Construction">
<cql>{P="N.*|PN.|CRD"} {L="@Var1"
P="V.B|V.Z"}</cql>
</item>
<item number="2" name="A2" ptn="主語 + 現在形"
isa="Finite Verb Construction">
<cql>{P="N.*|PN.|CRD"} {W="am|is|are" P="VBB|VBZ"}
{L="@Var1" P="V.G"}</cql>
</item>
<item number="3" name="A3" ptn="主語 + 現在形"
isa="Finite Verb Construction">
<cql>{P="N.*|PN.|CRD"} {W="have|has"
P="VHB|VHZ"} {L="@Var1" P="V.N"}</cql>
</item>
.....
```

リスト 3. 定形パターンの記述(一部)

リスト 3には、定形パターンの記述(一部)を示す。なお、定形パターンも XML のタグ表現を用いて記述している。

### 3.3 検索式生成

#### 3.3.1 シンボルグラマーの CQL 化

句動詞の見出し語データを電子化し、見出し語毎に句動詞を構成する主動詞、不変化詞、対応するシンボルグラマーのタイプを記録した。次に、XCQL フォーマットを用いて、シンボルグラマーの CQL 化を行った(リスト 1を参照)。変数記号で主動詞、不変化詞の対応を行い、シンボルグラマーのタイプで相互参照ができるようにした。本動詞部分は、定形パターン展開することを示す<finite> ~</finite>タグでマークしている。

#### 3.3.2 定形パターンの CQL 化

現在 30(×4)の定形パターンがある。すべて XCQL フォーマットを用いて、CQL 化を行った(リスト 3参照)。この情報を使って、<finite> ~</finite>タグでマークされた部分を定形パターンで展開する。シンボルグラマーのタイプによって CQL 式数に違いがあるが、肯定文だけで、おおよそ 100 行程度の CQL 式に展開される。

#### 3.3.3 CQL スクリプトの生成

句動詞の見出し語データと用いて、個々の句動詞見出しから CQL スクリプトを生成する。句動詞見出しレベルで、2 つのシンボルグラマ

<sup>5</sup> 受動形も含めるとさらにパターン数が増える。受動形パターンの展開作業も進めている。

ーがある場合があるので、全展開すると 5772 の CQL スクリプトが生成される。

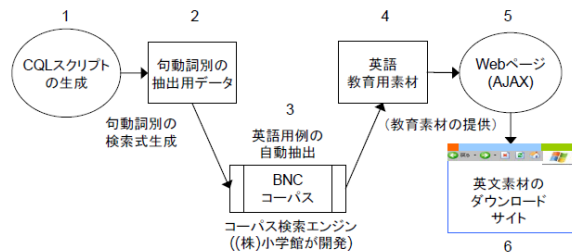


図 4. 作業工程の概念図

図 4には、作業工程の概念図を示す。1,2 の工程を終え、3 の抽出を行っている。5 のインタフェースについては、現在、設計を実施しており、イメージについては、次章で簡単に触れる。6 の BNC を利用した教育用英語句動詞用例集サイトは、今年の 9 月頃を目処に公開を目指したい。

### 3.4 用例抽出

## 4. おわりに

### 4.1 まとめ

およそ 4300 に及び句動詞見出しについて、文型記述を行い、CQL 式への展開を行った。文型記述の特徴は、動詞見出しを単純な動詞句としないで定形パターンに展開したことである。この特徴によって、BNC を通じた用例抽出によって、主語と句動詞の関係分析(主語の性質とモーダル表現)が可能になった。本稿では、文型記述の詳細と CQL 式への展開行程について報告した。

語法を確認することに役立つ用例集として、さらに、that 節による定形節をとる動詞用例や to 不定詞などの非定形節をとる動詞用例の抽出を実施する予定である。形容詞補語用例集や慣用句用例集なども計画をしている。

### 4.2 インタフェース・イメージ

抽出予定の句動詞用例は、句動詞の見出しが 4300 以上もあること、語法確認ができるように定形パターン展開していることなどから工夫が必要である。

英語文型教授のための教育用英語用例サイトのインタフェースは、アクセス経路が”文型”と下位分類(肯定、否定、疑問など)である。2 項目のクロスするポイントに用例が存在している。

句動詞用例サイトの提供にあたっては、(1) 句動詞の見出し語の入力支援、(2) 句動詞の不変変化詞の入力支援、(3) シンボルパターンの入力支援、(4) 定形パターンの入力支援を考慮す

る必要がある。4 つの項目について、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集にアクセスできるようにしたい。もっか、AJAX でのインタフェース実装を想定している。

### 謝辞

本研究は、平成 16-17 年度文部科学省科学研究費(基盤研究(C)(2))「ESP 教材の提供を目指した英語 e-Learning の研究開発」(研究代表者：高橋作太郎)と、平成 16-17 年度文部科学省科学研究費(基盤研究(C)(2))「教材制御の枠組みに基づく英語 e-Learning の研究開発」(研究代表者：馬場彰)支援を受けた。

### 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] アンドレア・タイラー、ビビアンエバンス、『英語前置詞の意味論』、研究社、2005。
- [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] 綿貫陽他、『ロイヤル英文法』、旺文社、2004。
- [5] クリストファバーナード、『英語句動詞文例辞典』、研究社、2002。
- [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。