

9月8日(木) 研究発表1 第3室(341)

語彙制限と文型制限が可能なコーパスからの英語教材抽出システム

A system to extract English educational material from a corpus
that allows vocabulary and grammar restrictionミン・ダニー 佐野 洋[†] 中村隆宏[‡]

北九州市立大学 外国語学部 〒802-8577 北九州市小倉南区北方 4-2-1

[†] 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[‡] 小学館コミュニケーション編集局デジタル編集開発室 〒101-0051 東京都千代田区神田神保町 2-30 昭和ビル 4FE-mail: danminn@kitakyu-u.ac.jp, [†] sano@tufs.ac.jp, [‡] takahiro@shogakukan.co.jp

1. はじめに

1.1. 研究目的

本稿は、BNC コーパスを利用した英語教育用教材の自動作成の方法と、作成した英語教材を広く提供するための教材素材ウェブサイトについて述べる。

筆者等は、効率的な教材開発の方法を研究開発している。まず、中学校・高等学校で利用される全種類の英語教科書を調査し、英語教育課程で学習されている英語文型を調査した。そして、(株)小学館マルチメディア局との共同研究により、英語文型をコンピュータで利用できる検索式に書き換え、BNCを使って文型パターン検索を実施した。抽出した英文用例に文型パターンに対応する文法項目情報等を付加してHTML化し、英語教育素材を提供するウェブサイトを構築した。本サイトは、小学館コーパスネットワーク(SCN)のサービスとして提供される。

学習教材作成の効率化に加えて、この取り組みの背景には、教材作成が労働集約的な作業であって、短時間に多様で且つ大量の教材が作成できないこと、作成された教材内容の品質を、教授者の経験や学習量といった属人的力量に依存せずに決めたいこと、などの問題点の解消に向けた取り組みもある。

1.2. 教材素材提供ウェブサイト

筆者等はこれまで、多様な学習要求に適合する語学教育方法論(N-Cube)を研究開発してきた[1,2]。図1は、N-Cubeを中心にした研究全体の作業工程を示している。本稿の内容がカバーする事項は、図中の(3)、(5)、(6)、(7)である。(4)のLTB(Language Tool Box)は、(株)小学館が開発したコーパス利用のためのソフトウェアである。詳細については、[3,4]を参照されたい。

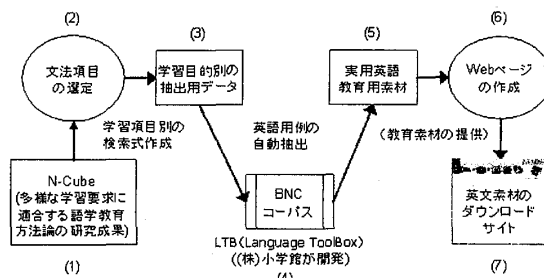


図 1. 全体構成図

我々は、教科書から選び出した学習項目を文型パターンとして検索式に書き換え(3)、LTBを利用して、BNCから文型パターンを含む用例を抽出した(5)。抽出した英文用例に文法項目情報等を付け加えてHTML化し、英語教育素材を提供するウェブサイトを構築した(7)。

2. 英文用例の自動抽出の方法

2.1. 文法項目の選定

日本国内で販売される英語教科書に関する市場調査(売り上げ高)を基準に選択した31種の英語の教科書及び日本の英語教育で広く使われている参考書4種を対象に文法項目を調査し[1,2], 144の文法項目を整理した。この144項目は、"There + is + 名詞"といった単純な文法項目から"条件法"などの難しい文法項目まで含む。

2.2. 文型パターンと検索式作成

144の文法項目は肯定文を基本とする。そして各文法項目に対し、否定文や疑問文など14の下位項目を設けた。144×14(2016)項目から、英語構文として存在しない634のパターンを除き、総計で1,382文型を整理した。そして、1,382の英語文型をコンピュータで利用できる検索式に書き換え

BNC を使って文型パターン検索を実施した。例えば「how を使った感嘆文」の場合、CQL 式は次の検索式になる。

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

2.3. 語彙レベルフィルタリング

文型パターンで用例抽出を行った結果に対し、語彙レベルフィルタリングすることで、抽出用例の語彙制御を行っている。語彙データには JACET8000、小学館プログレッシブ英和中辞典の重要語(1800語/4600語/8100語)¹、及び SVL12000²を用いた。語彙制御をしない用例に加えて、少なくとも JACET8000、小学館プログレッシブ英和中辞典(4600語/8100語)の語彙レベルでフィルタリングした用例の提供を行う予定である。

3. ウェブサイトの開発

3.1. ウェブデザイン

筆者等は、英語教育素材を提供するウェブサイトの利用者として現在、中学生、高校生や大学生に英語を教授する立場にある人を想定している。文型限定され、語彙制限された英文用例は、恐らく利用者の使い方として、補助教材やテストなどを作成する際の英文参照やサンプル利用が考えられる。

紙媒体の教科書では、一般的に、文法項目名が目次として強調され、文法項目が教授される順序に配列されている。利用者が用例を見つけるために用いるキーワードは文法項目名であると仮定し、ウェブデザインは、文法項目名(展開項目を含む)からウェブページをナビゲートし、文型説明が確認でき、そして英文用例が参照できたり、ダウンロードできたりするデザインとした(図 2)。

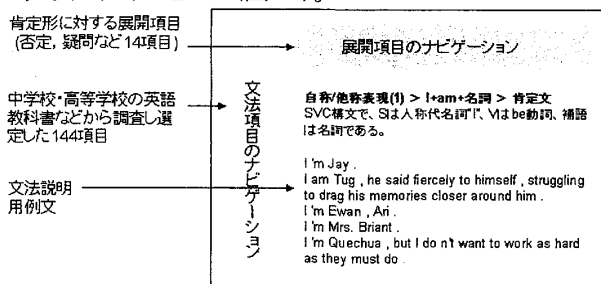


図 2. ウェブデザイン

¹ 小学館プログレッシブ英和中辞典 Copyright © (株)小学館 All rights Reserved.

² SVL12000 Copyright © (株)アルク All rights Reserved.

3.2. 構築したウェブサイト

図 3に開発したウェブサイトの様子を示す。画面左には、文法項目を選択するリストボックスがある。画面右上には、文法項目(肯定形)の下位項目の選択をするボタンが並んでいる。

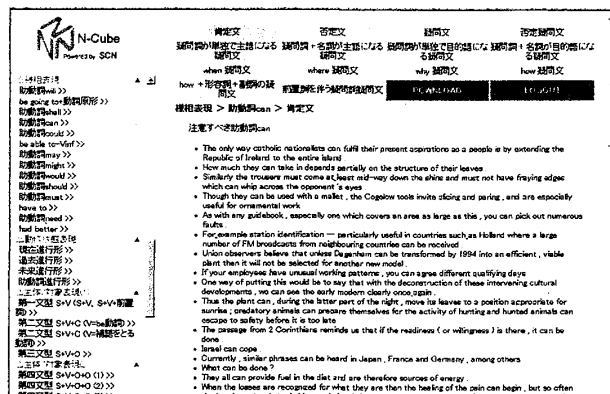


図 3. 開発したウェブサイト

文法項目と下位項目が選択されると、簡単な文法説明文と共に、画面で示すように BNC コーパスから抽出された用例が表示され、同時に画面右上にダウンロードボタンが表示されて、抽出用例をダウンロードすることができる。

謝辞

本研究は、平成 14-16 年度文部科学省科学研究費(基盤研究(B)(2))「全電子化検定済み教科書データの解析と大規模日本語コーパスの構築」(研究代表者:佐野洋)と、平成 15 年度(株)小学館・マルチメディア局委託研究の支援を受けた。

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] 佐野洋, 猪野真理枝, 宇野陽一郎:「多様性適合の学習環境を実現する語学教育支援システム」, 情報処理学会, 情報学シンポジウム講演論文集(55~62,8 頁),2002 年 1 月。
- [3] Nakamura, T. and Tono, Y. (2003) Lexical profiling using the Shogakukan Language Toolbox. In Murata, Yamada & Tono (eds.) ASIALEX 2003 Proceedings. Dictionaries and Language Learning: How can Dictionaries Help Human & Machine Learning?, pp. 170-176.
- [4] 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.