

文法項目別英語例文表示システム GPPS with SCoRE の構築

中條 清美* 西垣 知佳子** 赤瀬川 史朗† 内山 将夫‡

*日本大学 〒275-8576 千葉県習志野市新栄 2-11-1

**千葉大学 〒263-8522 千葉県稲毛区弥生町 1-33

†Lago 言語研究所 〒520-0863 大津市千町 2-7-4

‡情報通信研究機構〒610-0289 京都府相楽郡精華町光台 2-2-2

E-mail: *chuuju.kiyomi@nihon-u.ac.jp, **gaki@faculty.chiba-u.jp, †lagoinst@gmail.com, ‡mutiyama@nict.go.jp

あらまし GPPS (Grammatical Pattern Profiling System) with SCoRE (Sentence Corpus of Remedial English) は、日本人の英語学習者と教師のデータ駆動型学習 (Data-driven Learning: DDL) の支援を主な目的として開発された文法項目別英語例文表示システムである。調べたい文法項目を選択して英語例文を3段階の難易度別に表示する。教育用例文コーパス SCoRE を構成する例文は、日本人学習者にとって必要性の高い文法項目に対応する英語例文を英語母語話者が作例し、日本語対訳を人手で付けたものである。暫定版 GPPS with SCoRE はウェブ上で公開されており (<http://score.lagoinst.info/>), 著作権フリーの対訳付例文を教育用に利用することができる。本稿では、GPPS with SCoRE 構築の背景、概要、教育用コーパス SCoRE の作成手順、今後の予定について報告した。

キーワード 教育用コーパス, 文法項目別, 例文コーパス, 3レベル, 著作権フリー

Development of the Grammatical Pattern Profiling System with Sentence Corpus of Remedial English

Kiyomi CHUJO* Chikako NISHIGAKI** Shiro AKASEGAWA† and Masao UTIYAMA‡

* Nihon University 2-11-1 Shin'ei, Narashino-shi, Chiba, 275-8576 Japan

**Chiba University 1-33 Yayoi-cho, Inage-ku, Chiba, 263-8522 Japan

†Lago Institute of Language 2-7-4 Sencho, Otsu, Shiga 520-0863, Japan

‡ National Institute of Information and Communications Technology 2-2-2 Hikoridai, Seika-cho, Soraku-gun, Kyoto 610-0289, Japan

E-mail: *chuuju.kiyomi@nihon-u.ac.jp, **gaki@faculty.chiba-u.jp, †lagoinst@gmail.com, ‡mutiyama@nict.go.jp

Abstract A free, user-friendly Grammatical Pattern Profiling System (GPPS) with Sentence Corpus of Remedial English (SCoRE) has been created, using a corpus created especially for remedial or lower proficiency EFL learners and providing simple example sentences of grammar or lexical items. Teachers, students and material writers can use the GPPS to view and download beginner, intermediate or advanced language example sentences, with Japanese translations, of particular targeted grammar structures or lexical terms for classroom use. In this paper the creation of GPPS with SCoRE is explained, along with the rationale and procedure for creating the SCoRE sentences.

Keywords educational corpus, grammatically-categorized, sentence corpus, three proficiency levels, copyright-free

1. GPPS with SCORE 構築の背景

コーパス言語学の急速な進展とともに、コーパスを英語教育に利用するデータ駆動型学習 (Data-driven Learning: DDL) が世界的に広まっている。DDL では、英語学習者がコーパス(「生」の英語の集積データ)から検索ツールを使ってターゲット語を検索し、豊富な

言語使用例を見てことばの規則を発見して学んでいく。しかし、研究者向けに作られた検索ツールと「生」の英語のコーパス例文を利用する DDL は難易度が高く (Allan, 2009)^[1], 初級・中級学習者を対象とした DDL の実践例は、世界的に見ても少数にとどまっている (Boulton, 2009; Cobb & Boulton 2015; Mizumoto & Chujo, 2015)^{[2] [3] [4]}。本研究の目的は、主に初級学習者を対

象とした DDL の推進を目指して、①適切なレベルのコーパスと、②使い易い検索ツールの供給という DDL の 2 つの課題への対応を図ることであった。

具体的には、① 初級レベルの学習者が利用できるように、適切な難易度の英語例文とその日本語対訳を集積した教育用例文コーパス Sentence Corpus of Remedial English (SCoRE)を作成し、② SCoRE コーパスの例文をウェブ上で文法項目別に日本語訳とともに自由に閲覧・ダウンロード可能にする文法項目別英語例文表示システム Grammatical Pattern Profiling System (GPPS)を構築することであった。なお、GPPS with SCoRE の開発は 2012 年より開始された [5] [6]。

2. GPPS with SCoRE の概要

本節では GPPS with SCoRE 例文表示システムの概要を説明する。GPPS with SCoRE は、現在、暫定版がウェブ上で公開されており、SCoRE コーパスの例文を自由に閲覧することができる。<http://score.lagoinst.info/> にアクセスすると、図 1 のスタート画面が現れる。英語版(English Interface)あるいは日本語版(Japanese Interface)を選択すると、図 2 に示す画面から文法項目別に例文を検索できる。

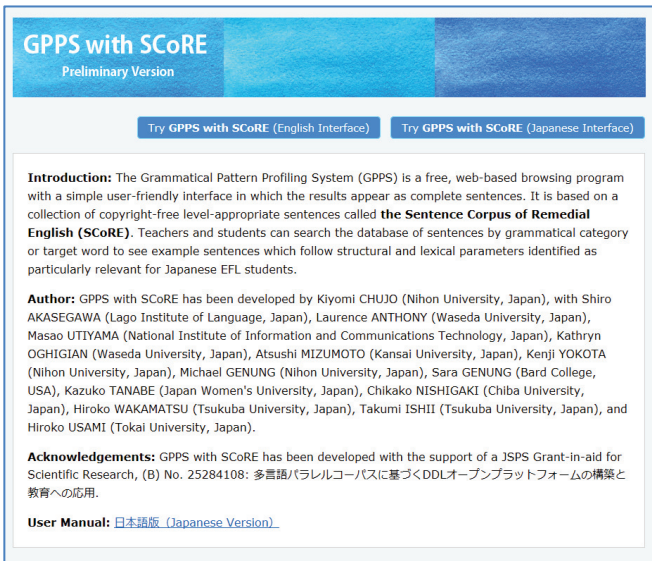


図 1 GPPS with SCoRE のスタート画面

図 2 に示す画面から、閲覧したい文法項目(図 2 の左側の文法パターンパネル)と、見たいキーワード(中列のキーワードパネル)を選んでクリックすると、著作権フリーの英語例文とその日本語対訳文(右側の例文パネル)を閲覧およびクリップボードにコピーすることが可能である。図中の文法パターンパネルに現われた数字は現在公開されている文法項目別の英語例文数、中列のキーワードパネルの数字はキーワード/キーワード別の例文数を示す。

文法パターンパネルの下にある 4 つのボタンを利用すると、表示する英語例文の難易度レベルを変更することができる。レベルには、初級、中級、上級があり、図 2 のように「すべて」を選ぶと、すべてのレベルの例文が表示される。初級の例文は先頭の赤のブロックで、中級は緑のブロックで、上級は紫のブロックで区別できるようになっている。

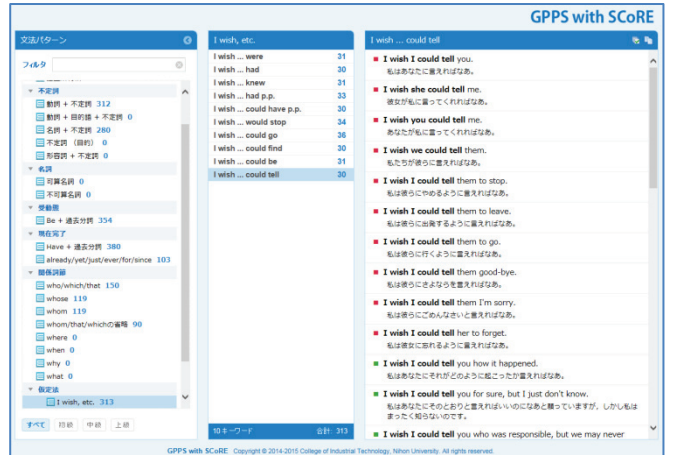


図 2 GPPS with SCoRE の I wish . . . could tell の画面

文法パターンパネルにはフィルタ機能があり、パネルの上にある入力ボックスに検索したい文字列を入力すると、文字列に関連した文法項目のみが表示される。

例文パネルには、クリップボードコピー機能がある。例文をクリックしてから、パネル右上の「コピー」ボタンをクリックすると選択した例文がコピーできる。表示されているすべての例文をコピーしたい時には、コピーボタンの左の「すべてを選択」ボタンをクリックしてから「コピー」ボタンをクリックすると全文をダウンロードすることができる。

3. 教育用コーパス SCoRE の作成手順

本節では、図 2 に示した文法項目別英語例文表示システム GPPS with SCoRE で検索可能な教育用コーパス SCoRE の作成手順について述べる。

検索ツールに加えて、コーパス教育利用の普及の柱は、学習者が「実際に使われている英文」と感じられる「簡潔で自然な」例文コーパスの開発である。本研究では、日本人学習者にとって必要性の高い文法項目に対応する英語例文を英語母語話者が作例し、日本語対訳を人手で付け、例文コーパスを作成した。その手順の概略は以下のとおりである。

(1) 学習対象者レベルの設定

教育実践に適したコーパスを作成するために、まず、対象となる学習者のレベル設定が必要であった。本研

究では、「全国の大学在籍者の半数以上は英検 3 級または 4 級程度の英語力」であることを明らかにした小野(2005)^[7]の大規模な英語習熟度調査などに基づいて、SCoRE の主要な対象者を英検 3 級または 4 級程度のリメディアルあるいは初級レベルの大学学部生(以下、対象学習者)に設定した。詳細は、Chujo, et al. (2012, 2015)^[8]^[9]、中條他(2015)^[10]を参照されたい。

(2) ソース・コーパスの作成

著者らがこれまでに行ってきた英語初級者向けコーパスデータの適性に関する基礎研究^[11]^[12]に基づいて、日本の高校英語教科書レベルの範囲内であって、対象学習者に適切なレベルと推定された 3,000 万語の英文テキスト(多読教材、各国の英語教科書、子供向け Web ニュースなど)で構成されるソース・コーパスを作成した。

(3) 対象学習者に必要な文法項目

対象学習者が苦手とする文法項目の調査を行った^[13]。その結果、中学文法項目においては所有代名詞、名詞複数形、現在完了形、受動態など、高校文法項目においては仮定法、関係詞、前置詞、などが不得意項目であることを特定することができた。

(4) 文法項目に対応するキーワードの選定

不得意文法項目に対応する高頻度動詞などのキーワードをソース・コーパスから抽出した。たとえば、受動態に用いられる高頻度動詞は got, been, seen, gone, done, come, made, given, lost, changed などであり、wish を含む仮定法の高頻度語は could, were, had, would, knew などであった。

(5) 英語例文を作成する理由

ソース・コーパスから抽出された英文は、文法項目に対応したキーワードを含んでおり、学習者に適切なレベルの英文になっていた。しかしながら、一文が長すぎるものや、内容が教育的に適切でないものなどが多く、さらに、致命的な欠陥として、著作権が付与されているため、それらの抽出文を公開して教育実践で使用することは不可能であった^[14]。

上述した理由により、SCoRE を構成する英語例文は、英語教育経験のある英語母語話者が、ソース・コーパスからの抽出英文などを参考にして、簡潔で自然な英語例文をオリジナルに作成したものである。学習者にとって利用しやすいことと、例文の authenticity を可能な限り確保するためにこのような作成手順を踏んだ。教育的配慮を重視したために DDL の例文の authenticity が低下したのではないかという議論につい

ては、Chujo, et al. (2015)^[15]の pedagogical criteria (appropriateness and usability)を参照されたい。

(6) 3 レベルの英語例文の作成

図 2 の左下の選択ボタンに示されたように、SCoRE の英語例文は、リメディアル・レベル/初級(以下、初級と表記)、中級、上級の 3 つの難易度レベルに分けて作成されている。前述のように、主要な対象学習者は初級レベルであるが、教育用例文コーパスとして、ある程度対象レベルの幅を広げて多様なレベルに対応できるように、中級と上級の教育用例文も作成した。難易度レベルを選択できることによって、たとえば、教室内での学習者の英語習熟度に個人差がある場合の学習支援や、また、教師や教材作成者が少し手ごたえのある例文を教材として使いたい場合にも役立つよう設計されている。

習熟度レベルの区分には、Chujo, et al.(2007)^[16]において、テキストの難易度の判断に有効であるとされた指標のうち、「文長」と「語彙習得学年」を用いて作成された。すなわち、初級は 8 語以内の文で、米国の語彙習得学年の 1 年生から 2 年生の語彙を用いて作成され、中級は 5 語から 11 語以内の文で、米国の語彙習得学年の 1 年生から 3 年生の語彙を用い、上級は 9 語以上の文で、米国の語彙習得学年の 4 年生以上の語彙を含むことを基準としている。各キーワードについて、レベルごとに 10 文ずつ例文を作成した^[17]。

(7) 日本語対訳の付与

英語例文に日本語対訳を付与した。続いて、日本語対訳の各文を複数の日本人研究者が個別に検討した。この段階は日本語対訳文の検討だけでなく、当該英語例文が SCoRE の教育用例文として妥当であるかどうか最も最終的に点検する段階となった(若松・石井・中條, 2015)^[18]。

現在、ウェブ上で公開されている文法項目は、属格表現、to 不定詞、受動態、現在完了、関係詞節、仮定法のそれぞれの文法項目について初級・中級・上級の英文 3,142 文とその日本語対訳である。取得した例文は、学習者の実情に合わせて、適宜、英文や日本語対訳文を自由に修正したりすることは容認されている。2015 年度中に「操作マニュアル」(日本語版と英語版)を作成し、例文の使用条件について明記する予定である。

4. 今後の予定

以上、文法項目別英語例文表示システム GPPS with SCoRE 構築の背景、概要、教育用コーパス SCoRE の作成手順について報告した。GPPS with SCoRE は現在

進行中の DDL オープン・プラットフォーム構築プロジェクト(中條, アントニ, 赤瀬川, 西垣, 水本, 内山, 2013)^[19]の主要な部分として, 新規に作成されたものである。著作権フリーの簡潔で自然な英語例文とその日本語対訳文を, 文法項目別, レベル別に表示することが可能であり, 英語学習者の DDL や, 教師の教材作成時のコーパス利用推進に役立つと期待されている。

今後の GPPS with SCoRE の改良項目として, まず, 新規例文の拡充があげられる。Minn, et al. (2005)^[20]が指摘するように, コーパスの構築は time-intensive & labor-intensive, すなわち, 非常に人手と時間のかかる作業である。GPPS with SCoRE では, 例文と対訳を人手で入念に作成しているため, 現在のところ公開されている例文数は限られている。次回の GPPS with SCoRE (Ver.2)の改訂版では 7 つの文法項目, 3,000 例文の追加を予定している。

より多くの学習者や教師が利用しやすいようにインターフェースを改良する予定である。具体的には, 練習問題出題インターフェースの追加, KWIC 検索インターフェースの追加などを行うことによって, DDL 教育実践で使用するタスクの作成が容易になる。また教師向けに授業用に例文を印刷して利用しやすいように PDF 版例文の提供が予定されている。

すでに GPPS with SCoRE は教師の教材作成の補助ツールとして試用され^[21], 改良に向けてフィードバックの収集を行っている。今後は, GPPS with SCoRE を DDL 授業実践において試用し, 活用方法の探求とともに, 教育効果を検証していきたいと考えている。

謝辞: 本研究は平成 25-28 年度科学研究費補助金基盤研究(B)(25284108)を受けて行われました。

文 献

- [1] Allan, R. "Can a graded reader corpus provide 'authentic' input?" *ELT Journal*, 63(1), 2009, 23-32.
- [2] Boulton, A. "Testing the limits of data-driven learning: language proficiency and training," *ReCALL*, 21(1), 2009, 37-54.
- [3] Cobb, T. and Boulton, A. "Classroom applications of corpus analysis," In D. Biber & R. Reppen (eds.), *The Cambridge Handbook of Corpus Linguistics*. Cambridge: Cambridge University Press, 2015, 478-497.
- [4] Mizumoto, A. and Chujo, K. "A meta-analysis of data-driven learning approach in the Japanese EFL classroom," *English Corpus Studies*, vol. 22, 2015.
- [5] Chujo, K., Oghigian, K. and Akasegawa, S., "Developing a level-appropriate, grammatically-categorized browsing system of EFL example sentences for teachers and students," The 10th International Conference on Teaching and Language Corpora (TaLC10), Institute of Applied Linguistics,

University of Warsaw, Poland, July 2012.

- [6] Chujo, K., Oghigian, K. and Akasegawa, S. "A Corpus and grammatical browsing system for remedial EFL learners," In A. Leńko-Szymańska and A. Boulton (eds.), *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam: John Benjamins, 2015, 109-128.
- [7] 小野博, 村木英治, 林規生, 杉森直樹, 野崎浩成, 西森年寿, 馬場真知子, 田中佳子, 國吉丈夫, 酒井志延, "日本の大学生の基礎学力構造とリメディアル教育," NIME 研究報告, 6, 2005, 1-147.
- [8] See [5].
- [9] See [6].
- [10] 中條清美, 若松弘子, 石井卓巳, 宇佐美裕子, 横田賢司, キャサリン・オヒガン, 西垣知佳子, "教育用例文コーパス SCoRE の作成," 日本大学生産工学部研究報告 B, 第 48 巻, 2015.
- [11] 中條清美, 西垣知佳子, 山保太力, 天野孝太郎, "英語初級者向けコーパスデータとしての教科書テキストの適性に関する研究," 日本大学生産工学部研究報告 B, 第 44 巻, 2011, 13-23.
- [12] 中條清美, 西垣知佳子, 山保太力, 落合太一, "米国 Reading 教科書と英語 Graded Readers の英語初級者向けコーパスデータとしての適性に関する研究," 日本大学生産工学部研究報告 B, 第 45 巻, 2012, 29-42.
- [13] 中條清美, 横田賢司, 長谷川修治, 西垣知佳子, "リメディアル学習者の英語習熟度と英語文法熟達度調査," 日本大学生産工学部研究報告 B, 第 45 巻, 2012, 43-54.
- [14] 中條清美, 赤瀬川史朗, 西垣知佳子, 横田賢司, 長谷川修治, "LagoWordProfiler による英語 Graded Reader Corpus の Collocation/Colligation 頻度分析," 日本大学生産工学部研究報告 B, 第 45 巻, 2012, 55-71.
- [15] See [6].
- [16] Chujo, K., Utiyama, M. and Nishigaki, C., "Towards building a usable corpus collection for the ELT classroom," in E. Hidalgo, L. Querreda, and J. Santana (eds.), *Corpora in the Foreign Language Classroom*, Amsterdam, Rodopi, 2007, 47-69.
- [17] See [10].
- [18] 若松弘子, 石井卓巳, 中條清美, "学習支援用日英例文パラレルコーパス SCoRE の構築における課題: 日本語対訳例文の訳出に焦点を当てて," 英語コーパス研究, 第 22 号, 2015.
- [19] 中條清美, アントニ・ローレンス, 赤瀬川史朗, 西垣知佳子, 水本篤, 内山将夫, "教育利用可能なパラレルコーパス検索プラットフォームの構築に向けて," 英語コーパス学会第 39 回大会, 東北大学, 2013 年 10 月 6 日.
- [20] Minn, D., Sano, H., Ino, M. and Nakamura, T. "Using the BNC to create and develop educational materials and a website for learners of English," *ICAME Journal*, 29, 2005, 99-113.
- [21] 西垣知佳子, 小山義徳, 神谷昇, 中條清美, "中学校におけるコーパスを利用したデータ駆動型英語学習の実践: ペーパー版 DDL からタブレット端末 DDL まで," 英語コーパス学会第 40 回大会, 熊本学園大学, 2014 年 10 月 4 日.