

TOEIC 公式問題集に基づく新コーパス作成と それを用いた品詞分け TOEIC 英単語学習用アプリケーションの開発

C0119340 リュウ カテイ

1. はじめに

同じ英語力でもテスト時の時間配分の工夫は得点が大きく変動するので、テスト時の時間配分を前提とする勉強法が有効である。また、一般に単語力強化のため、単語帳が広く利用されている。TOEIC テスト (以下、TOEIC) の場合、特に使用語彙が極めて安定しているため [3]、公式問題集での高頻出単語を学ぶことが高得点への鍵となる。しかしながら現状では、TOEIC 単語帳教材が多数存在するにもかかわらず、品詞ごとに単語を分ける単語アプリケーションはまだない。このような観点から、単語を品詞ごとに分別する英語単語帳アプリケーションを開発する。

2. 関連研究

2.1 初級者向け TOEIC 語彙選定とその効果に関する研究 [1]

中篠は、公開されている TOEIC 問題と市販の TOEIC 練習問題から TOEIC テストデータベースより、補習用学習語彙を選定しており、「TOEIC 語彙 1, 2」の学習により、TOEIC スコアで 140 点の向上が期待できると報告している。

2.2 英語教科書コーパスに基づく TOEIC 語彙の研究 [2]

卯城らは、新課程の中高英語教科書と TOEIC 公式問題集をコーパスとして、新課程教科書の特徴分析と、それらの教科書による語彙学習効果の推定を行い、語彙サイズは TOEIC スコア変動の高重要な説明要因と報告している。

2.3 TOEIC と語彙学習教材との関係性分析研究 [3]

長橋は、TOEIC と語彙学習教材との関係性に着目したコーパス研究を行った。大学生が入学前後で入手可能な学習教材を対象に、TOEIC で用いられる語彙のカバー率、対象教材の語彙知識の豊富さ、語彙の複雑さを分析し、TOEIC での使用語彙が極めて安定していると報告している。

以上のことより、本研究では TOEIC テストの品詞問題で高点数が取れるよう、TOEIC 問題集のコーパスを作成し、それに基づく品詞分け単語帳アプリを開発する。

3. 本研究の概要

新 TOEIC 公式問題集の電子版テキストコーパスを作成するとともに、自然言語処理を施すなどして、品詞分けされた TOEIC 単語帳アプリケーションを構築する。具体的には、ス

マートフォンで問題集の写真を撮影、スマートフォンアプリケーションで TOEIC 公式問題集 8 冊を電子版に変更、データ整理後コーパス作成をこの順に行う。電子版データは Python の NLTK ライブラリで下処理を行い、品詞ごとに単語の出現回数によって単語リストを作成する。また、Python を利用してスクレイピングを活かして web 上の「Webio 英和辞典」から英単語意味の取得も行う。その後、Django を利用して、ウェブアプリを作成する。



図1. 本研究で構築するシステムの概要図

4. 研究計画

表1 研究のスケジュール表

	8月	9月	10月	11月	12月	1月	2月
アルゴリズムを書く							
アプリケーション化							
TOEIC公式問題集電子版データの追加							
プログラミングの改善							
最終レポートを書く							

5. 進捗状況

- TOEIC 公式問題集 1 のデータを電子版に変更した。
- 電子版データの下処理を完成した。下処理はストップワード削除、名詞単数化、動詞原型変更、品詞ごとの単語出現回数の測定調査という四つの段階を含めている。

6. おわりに

TOEIC テストの勉強支援を目的として、品詞分け単語アプリケーションの開発について述べた。

参考文献

- [1] 中篠青美, 英語初級者向け「TOEIC 語彙 1, 2」の選定とその効果, 日本大学生産工学部研究報告 B, No.36, pp. 27-42(2003).
- [2] 卯城祐司, 浜田章, “新課程の英語教科書コーパスに基づく TOEIC® 語彙の特徴と難易度の推定, TOEIC Research Report 6, pp. 1-33 (2017).
- [3] 長橋雅俊, TOEIC と語彙学習教材との関係性におけるコーパス分析: 語彙の多様性と複雑さの比較に焦点を当てて, 聖徳大学研究紀要, No.30, pp. 103-111- (2019).