

学習者の読み書き頻度に基づいた 英語スピーキング学習支援システム

青木開生, 鷹野 孝典

神奈川工科大学 情報学部 情報工学科

An English Learning System for Speaking Practice based on Frequency of Lerner's Reading and Writing

Kaisei Aoki, Kosuke Takano

Department of Information and Computer Sciences, Faculty of Information Technology,
Kanagawa Institute of Technology

This study presents an English-speaking learning system for non-native English people who can read and write English but have an awareness of being not good at speaking and listening to. The feature of our proposed system is that our system can provide English sentences for the speaking practice to a learner based on frequencies of English words appeared in documents that the learner has read and written so far. By incorporating a method focusing on English speaking practice into usual English reading and writing experiences and learnings, a learner can primarily practice pronunciation of English words that are likely to be spoken in actual English conversation. In the experiment using a prototype system, we confirm the feasibility of our proposed system

キーワード: 発音練習, 英語学習, 英単語, 出現頻度, 英語音声認識, e-Learning システム

1. はじめに

グローバル化に伴う英語を用いたコミュニケーション機会の増加とともに、英語で聞く、話すといったコミュニケーションスキルの修得が益々重要になりつつある。これまでに効果的な英語学習環境の実現を目的として、e-Learning システムを活用したものも含めて、優れた学習教材や有用な教授方法が提案・開発されているものの、依然として英語で聞くこと、話すことに苦手意識を持つ日本人が多いと言われている⁽¹⁾。

英語での読み書きがある程度できるものの、聞く話すことが苦手な理由としては、適切な指導者や練習機会の不足、継続性の欠如といった学習環境・姿勢による問題点に加えて、日本人特有のカタカナ英語読みになってしまう、日本語の発音にない音素を聞き取れない、英語の発音における正しい舌の動きが理解できていない、日本語発音で黙読する習慣がついてしまっているために英語発音に転換することが困難である等、

様々な要因が挙げられる。これらの要因は、適切な発音練習の繰り返しで克服可能であると考えられるが、大きな労力と時間がかかる場合が多い。本研究では、英文の読み書きはできるが、会話中に簡単な単語が思いつかない、または、単語を知っていたとしても発音が通じないといった経験を持つ等、話すことに苦手意識のある学習者を対象として、音声認識機能を活用した英語スピーキング学習システムを提案する。本システムの特徴は、学習者がこれまでに読み書きした英文中に含まれる英単語の出現頻度に基づいて、発音練習する対象となる英文を提示する点にある。普段の英語の読み書き経験・学習において、正しい英語発音を意識した学習を導入することにより、学習者は、英会話において実際に発話する可能性が高いと考えられる英単語の発音を優先的に練習することができる。

2. 関連研究

日本人の多くが英語で話す能力の不足を感じているとされ、その学習支援を目的とした e-Learning システムが数多く開発・提案されている。野本らは従来の英語発音指導の問題点として、指導者の立場から、授業時間数が少ないこと、コミュニケーション重視の英語教育の中において発音指導を行っていくこと、とりたてて発音を指導する必要はないと考えている教員がいること、発音の指導法がわからない教員がいること、の4点を挙げている⁽¹⁾。また、野村らは英語教育の専門家ではない工学専門教員であっても容易に英語教育を実現することを目的として、意識を通してより自然な日本語への翻訳を目指す機能、および発音記号に準拠した的確なスピーキングを目指す機能を実装した e-Learning システムを開発している⁽²⁾。また、竹野らは e-Learning システムを使用しての学習の際に日本語と英語の音声の弁別の仕方、特に母音の発音の違いをスピーキングとリスニングを通して認識させる英文の学習の重要性を述べている⁽³⁾。

これらに対して提案する英語スピーキング学習支援システムでは、学習者が英語でコミュニケーションを行う状況において、その場で発話する可能性の高い英単語を正しく発音できるようになることを目指して、既に読み書きした英文中の英単語の出現頻度に基づいて、発音練習用の英文を提示する機能を組み込んでいる点に特徴がある。

3. 提案システム

提案システムの概要を図1に示す。本システムは、(1) 学習者の英文読み書き履歴において出現した英単語頻度算出機能、(2) 英語例文出題機能、(3) 発音正誤判定機能より構成される。提案システムでは、学習者がこれまでに読み書きした Web 記事、論文原稿、配布資料、プレゼンテーション資料等を対象として英文を抽出し、英文中に含まれる英単語の出現頻度に基づいて、例文データベースから発音練習用の例文を学習者に提示する。学習者が発音した内容は、音声認識機能を用いてテキスト化され、正解文章と照らし合わせることで、正誤判定される。

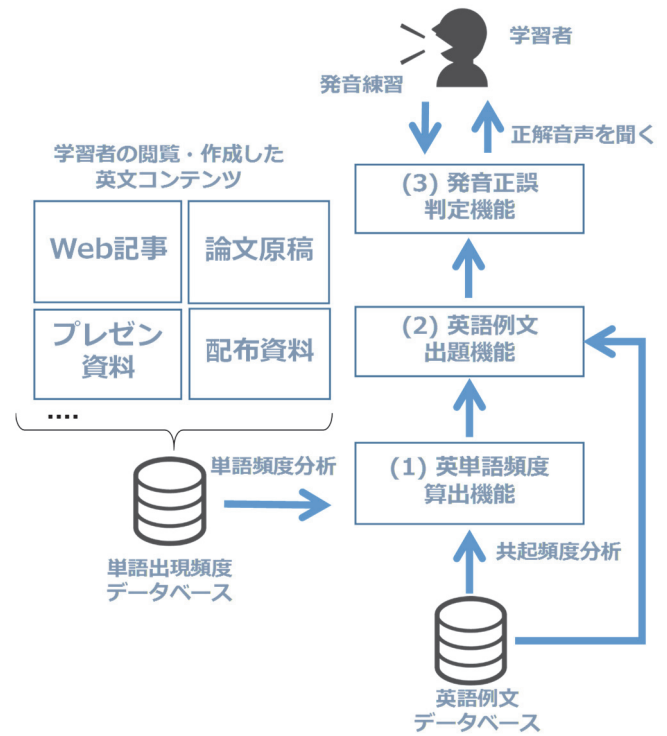


図1 提案システムの概要図

3.1 英単語頻度算出機能

学習者の英文読み書き履歴において出現した英単語頻度算出機能は、下記の手順により実現される。

- Step-1: 学習者がこれまでに読み書きした英語文書 $d_1, d_2, d_3, \dots, d_n$ を選択し、英文を抽出する。
- Step-2: Step-1 で英文から英単語群を抽出する。その際、各英単語はステミング処理を行い、単数形や原形に変換される。また、冠詞や代名詞等の不要語は、ストップワード（表1）として削除される。
- Step-3: 各文書に出現した英単語 t の出現回数の和に基づいて、英単語 t のスコア $Score(t)$ を算出する。

$$Score(t) = \frac{\sum_{i=1}^n TF(t, d_i) \times W_1^{t, di}(c) \times W_2^{t, di}(s)}{N}$$

ここで、 N は学習者が読み書きした英文コンテンツ中に出現する英単語の総数である。

また上式では、英単語 t が出現した文書形式 c に応じた重み $W_1^t(c)$ 、および、各文書を学習者自身が作成した/他者が作成したといった状態 s に応じた重み $W_2^t(s)$ をかけている。例えば、Web 文書の英文よりも、プレゼンテーション資料の英文に重点を

置きたい場合、 W_1' (Web 文書)=1, W_1' (プレゼンテーション資料)=2 のように設定する。また、学習者自身が作成した英文に重点を置きたい場合は、 W_2' (自身が作成)=2, W_2' (他者が作成)=1 のように設定する。

表 1 ストップワードの例

'a', 'an', 'the', 'i', 'my', 'me', 'mine', 'you', 'your', 'yours', 'we', 'our', 'us', 'he', 'his', 'him', 'she', 'her', 'hers', 'they', 'their', 'theirs', 'it', 'its', 'this'

3.2 英語例文出題機能

英語例文出題機能は、下記の手順により実現される。

- Step-1: 例文データベースから英文 e_i を選択する。
- Step-2: 英文 e_i に含まれる全ての単語 t について、英単語頻度算出機能によりある学習者 u に対して算出した $Score(t)$ を抽出する。
- Step-3: 英文 e_i について、英文スコア $SentenceScore(e_i)$ を、英単語スコア $Score(t)$ の和として算出する。

$$SentenceScore(e_i) = \sum_{t \in e_i} Score(t)$$

- Step-3: 英文スコア $SentenceScore(e_i)$ に応じて英文をランキングし、英文スコアの上位である英文から発音練習用例文として、学習者 u に出題する。

3.3 発音正誤判定機能

問題英文 E に対して、学習者が発話した内容を音声認識機能が返した英文 F とする。英文 E と英文 F の照合を行い、英文発音の正答率 r を算出する。英文 E に含まれる単語数を $\#E$ 、正解した単語数を $\#correct$ とすると、下記式のように計算される。

$$r = \frac{\#correct}{\#E}$$

4. プロトタイプシステム

図 2 にプロトタイプシステムの概要図を示す。本プロトタイプシステムは、Web ブラウザから利用可能な Web アプリケーションとして構築した。ユーザインタ

ーフェースは、PC からの利用だけではなくスマートフォンやタブレット端末からの利用も想定して設計した (図 3, 図 4)。

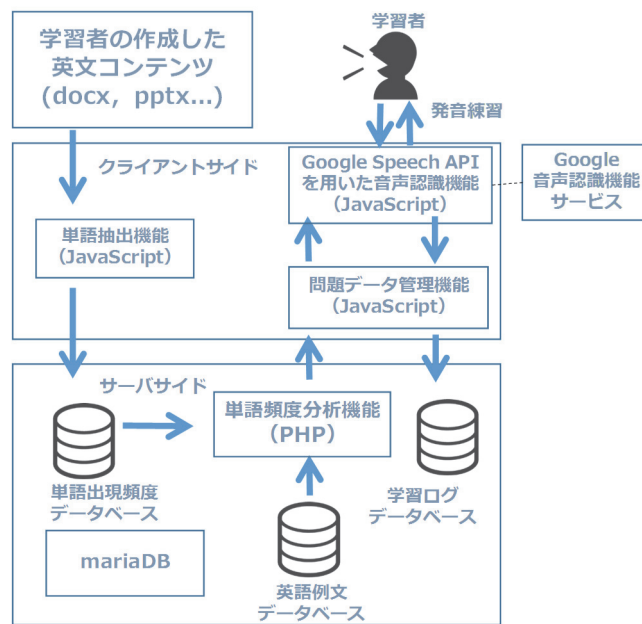


図 2 プロトタイプシステムの概要図

学習者がこれまでに読み書きした英文中に含まれる英単語の出現頻度を計算するために、学習者の作成したコンテンツ分析機能を JavaScript で実現した。本機能は、学習者がアップロードしたマイクロソフト Office 形式のワードファイル (docx) やパワーポイントファイル (pptx) から、自動で英文データを抽出し、英単語の出現頻度を計算することができる。また、発音練習用の例文データベースには、学术论文によく用いられる英文がまとめられたテンプレート集である Academic Phrasebank⁽⁴⁾ に掲載されている 560 個の英語例文、およびビジネスや一般的なマナーを考慮した英語例文集サイトである英借文ドットコムに掲載されている 261 個の英語例文を用いて構築を行った。なお、データベース管理システムとして MariaDB を用いた。

学習者による英文発話の正誤判定を行うために、音声認識 API である Google Speech API を適用した。学習者の発話から認識された英文文字列と正解データを照合することで正確判定を行っている。また、学習者の発話内容や正誤判定結果等の学習履歴は、学習履歴データベース上に記録される。

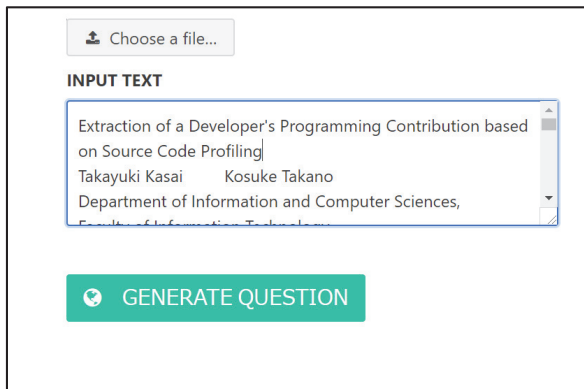


図 3 英文コンテンツ登録画面

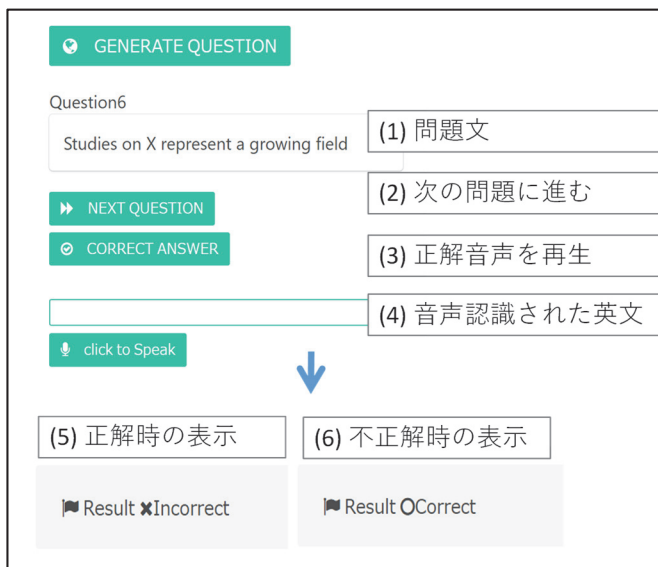


図 4 英語スピーキング練習画面

表 2 Academic Phrasebank に収録されている英文例

	英語例文
1	Approximately half of those surveyed did not comment on ...
2	The incidence of X has been estimated as 10% ...
3	The average of 12 observations in the X, Y and Z is 19.2 mgs/m ...
4	The number of Xs reached a peak during ...

表 3 英借文ドットコムに収録されている英文例

	英語例文
1	The company also plans to address a number of key issues including a focus on overseas production and procurement capabilities.
2	Investigation committee has released a draft of preventive actions to address the problem of the "sick house" syndrome.
3	Licensing agreement will need an assignment clause to address this possibility.
4	Sorry for the delay in my response, however my contact in Sony is travelling and will not be able to respond until late next week.

5. 実験

4 章で示した提案システムのプロトタイプを用いた実験により、提案システムの実現可能性を検証する。

5.1 実験環境

5 名の学習者を対象とした実験を行う。各学習者がこれまでに読み書きした英文コンテンツとして、Microsoft ワード文書で作成されたものを利用した。各学習者の英文コンテンツの内容を表 4 に示す。今回の実験では、英文コンテンツは全て学習者自身が作成した Microsoft ワード文書であるため、文書形式に応じた重み $W_1'(c)$ 、および、各文書を学習者自身が作成した/他者が作成したといった状態に応じた重み $W_2'(s)$ はともに 1 を設定した。

表 4 学習者データ

学習者	英文の種類	単語数
A	学術論文	1719
B	学術論文	1753
C	学術論文	1373
D	日本茶と紅茶の解説	1391
E	自己紹介と学校施設の紹介	556

5.2 実験方法

表 4 に示す学習者を対象として、3 章で示した方式に従い、学習者の英文読み書き履歴において出現した英単語頻度に基づいてランキングされた英語例文の出題結果について考察する。

5.3 実験結果

各学習者の英文コンテンツから算出した英単語出現頻度の結果(上位 5 単語)を表 5~表 9 に示す。また、また、英単語出現頻度に基づいた、発音練習用の英文ランキング結果(上位 4 英文)を表 10~表 14 に示す。

表 5 の結果より、学習者 A の英文コンテンツからは、「デジタル」"かんばん"を利用したプロジェクト管理システムに関する「task」「Kanban」「digital」といった英単語のスコアが高く、頻出単語であることがわかる。さらに、表 10 の結果より、この英単語スコアに基づいて、学習者 A に対する発音練習用の英文として、「task」「digital」「applications」等の英単語を含んだ英文が上位にランキングされていることが確認できる。

学習者 B～学習者 E の英単語出現頻度の結果, および英文ランキング結果についても同様の考察ができる.

これらの結果は, 学習者が提案システムを利用して, これまで読み書きした経験から, 今後の英会話において実際に発話する可能性が高いと考えられる英単語の発音を優先的に練習することができる可能性を示している.

表 5 学習者 A の頻出単語

順位	単語	英単語スコア
1	task	0.0506
2	Kanban	0.0482
3	digital	0.0197
4	application	0.0139
5	smartphone	0.0139

表 6 学習者 B の頻出単語

順位	単語	英単語スコア
1	image	0.0650
2	sketch	0.0422
3	color	0.0359
4	search	0.0171
5	sample	0.0245

表 7 学習者 C の頻出単語

順位	単語	英単語スコア
1	developers	0.0407
2	software	0.0407
3	performance	0.0174
4	code	0.0174
5	source	0.0160

表 8 学習者 D の頻出単語

順位	単語	英単語スコア
1	tea	0.0791
2	from	0.0179
3	leaves	0.0143
4	that	0.0122
5	green	0.0122

表 9 学習 E の頻出単語

順位	単語	英単語スコア
1	food	0.0233
2	from	0.0179
3	snow	0.0179
4	japanese	0.0161
5	Niigata	0.0143

表 10 学習者 A の英文ランキング結果

順位	英文
1	Women are slower than men at certain precision manual <u>tasks</u> , such as ...
2	to complete two <u>tasks</u> .
3	The data were recorded on a <u>digital</u> audio recorder and transcribed using a ...
4	has several practical <u>applications</u> . Firstly, it points to ...

表 11 学習者 B の英文ランキング結果

順位	英文
1	A <u>search</u> of the literature revealed few studies which ...
2	Notwithstanding the relatively limited <u>sample</u> , this work offers valuable insights into ...
3	Although the current study is based on a small <u>sample</u> of participants, the findings suggest ...
4	The lack of X in the <u>sample</u> adds further caution regarding the generalizability of these findings.

表 12 学習者 C の英文ランキング結果

順位	英文
1	Part of the aim of this project is to <u>develop</u> software that is compatible with ...
2	Part of the aim of this project is to develop <u>software</u> that is compatible with ...
3	The experiments were run using custom <u>software</u> written in...
4	Semi-automated genotyping was carried out using X <u>software</u> and ...

表 13 学習者 D の英文ランキング結果

順位	英文
1	In 1972, a perennial herb "stevia" was introduced to Japan from Paraguay to manufacture a natural sweetener.
2	I am writing to inform you that I shall be leaving New York at the end of this month to assume a new assignment for International Planning Division, Gebew Corporation in Tokyo.
3	We hope to hear from you shortly.
4	I am glad (pleased, happy) to learn (know) from your letter (fax, email) of May 1 that you have been promoted to Manager in your department.

表 14 学習者 E の英文ランキング結果

順位	英文
1	This week in the Red Coach Restaurant, we are featuring a wide selection of fresh oyster and sea food .
2	Food processors should place food security above profit.
3	In 1972, a perennial herb "stevia" was introduced to Japan from Paraguay to manufacture a natural sweetener.
4	We hope to hear from you shortly.

おわりに

本研究では、学習者がこれまでに読み書きした英文中に含まれる英単語の出現頻度に基づいて、発音練習する対象となる英文を出題する音声認識機能を活用した英語スピーキング学習システムを提案した。また、提案システムのプロトタイプを用いた実験により、提案システムの実現可能性を確認することができた。

今後の予定として、単語の出現頻度のみではなく、単語の共起頻度に基づく関連度により、発音練習用の英語例文を抽出する機能、および、学習者が読み書きした英文コンテンツを網羅的に抽出し単語出現頻度データベースに格納する機能等についてのシステム拡張を行っていくことを検討している。さらに、提案システムを用いた英語スピーキング学習の効果を測定するために、中長期的な授業計画に沿った実証実験を継続的に実施していくことも今後の課題として挙げられる。

参考文献

- (1) 野本尚美, 平塚紘一郎: “e-Learning を用いた英語発音指導システム”, 仁愛女子短期大学研究紀要, No. 47, pp. 17-21 (2015)
- (2) 野村 由司彦, 坂本 良太: “発音, 逐語訳, 意識を重視した英語教育をサポートする e-Learning システム”, 工学教育, Vol.56, No.6, pp. 96-101 (2008)
- (3) 竹野純一郎, 学松田雅子, 学南津佳広, Judith Mikami: “音声訓練とオリジナル・スピーキングテストサイトの開発”, 中国地区英語教育学会研究紀要, Vol.46, pp.131-140 (2016)
- (4) John Morley: “Academic Phrasebank”, The University of Manchester (2017)
- (5) 英借文ドットコム: <http://www.eishakubun.com/>, (2017.11.28 訪問)