

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

学籍番号：1321084 氏名：青木開生

指導教員：鷹野孝典

背景

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話すというスキルが重要に益々なりつつある.

→e-Learningシステムを用いた英語学習も増加している.

一方で, 日本人は英語での読み書きがある程度できるものの
発話、聞き取りという分野においては

カタカナ英語発音になってしまう, 練習継続性の欠如や適切な指導者の不足などの問題を抱えている.

関連研究

[1]e-ラーニングを用いた英語発音指導システム[野本2015]

既存の英語教育の発音問題を指摘し，解決策としてe-ラーニングを用いた英語学習について提案している

[2]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[竹野2016]

e-ラーニング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、特に母音の発音に違いをスピーキングとリスニングに重点に置いて学習させる。

[3]発音, 逐語訳, 意識を重視した英語教育をサポートするeラーニングシステム[野村2016]

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ため、意識を通してより自然な日本語への翻訳を目指す機能，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-ラーニングシステム

研究動機

提案システムでは

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提示する。

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。

研究目的

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
- 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話すことに苦手意識のある

上記のような問題を抱えた学習者を対象として
読み書き頻度に基づいた出題が可能な、
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。

読み書き頻度による出題が有効と思われる例

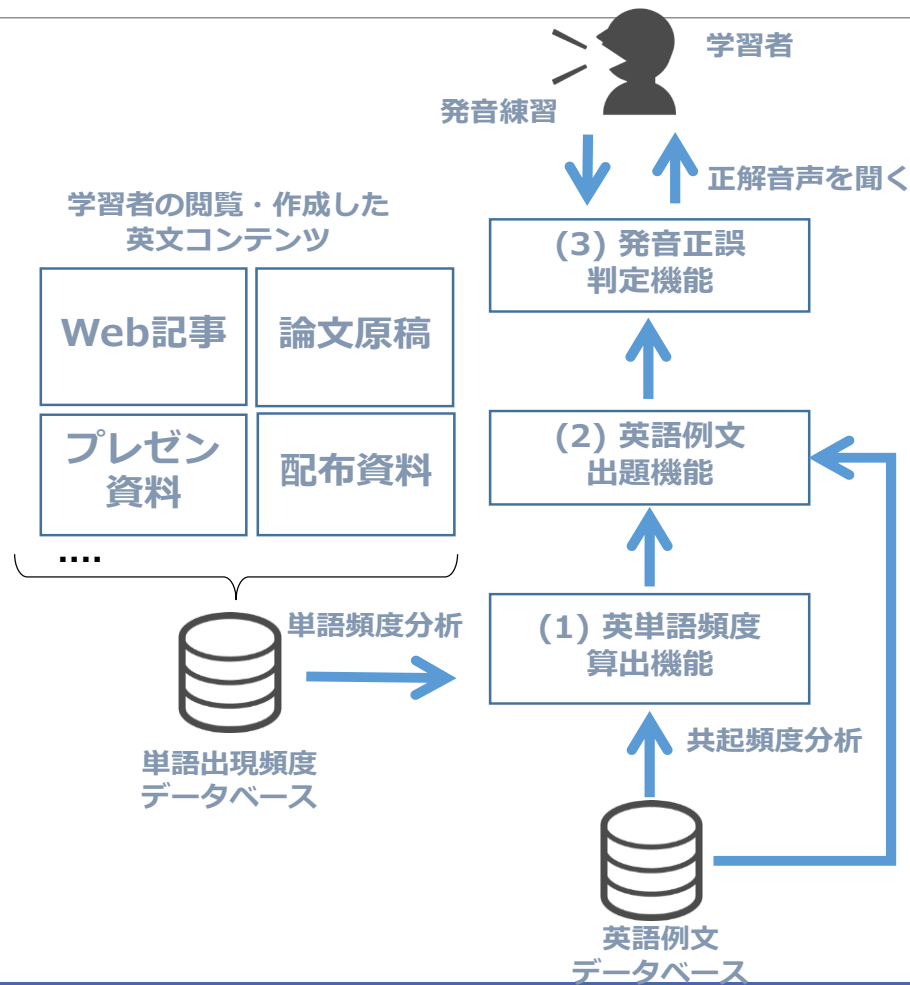
同じ意味あいの英文であっても、

学習者の読み書き頻度の高い英文に応じて使用する英文が分かれる

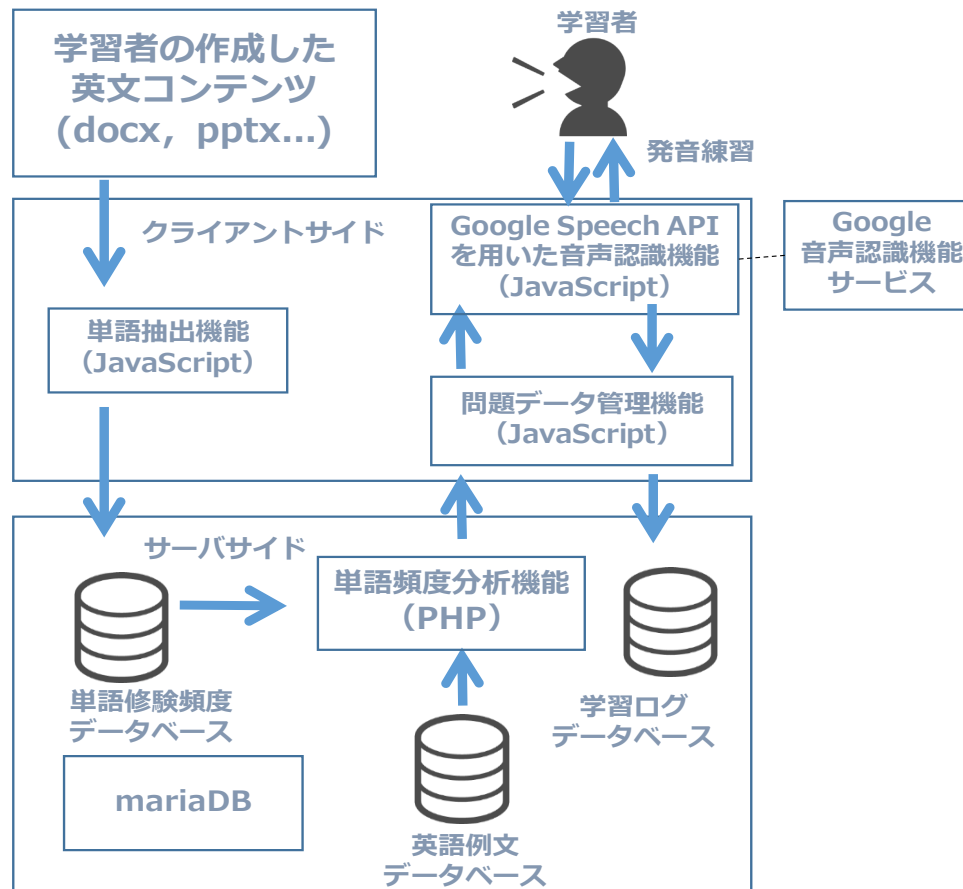
「開会式はどこで行われましたか？」という文章を英語で発話したい

学習者ごとの 読み書き頻度が高い英文	「開会式はどこで行われ ましたか？」	備考
<u>ceremony</u>	Where was the Opening <u>ceremony</u> ?	単語レベルで一致した表現
<u>Where did X take</u>	<u>Where did ceremony</u> <u>take</u> place?	句レベルで一致した表現

提案方式



実装システム図



主な実装

実験システム内で出題する問題用のデータを以下の2つから取得したい

1. チャットサービスでの英文コミュニケーション

学習者が参加してる英語での会話が行われているため、

学習者の英会話に用いる英文が直接取得できる。

slackであればログファイルを取得できる



研究室内で利用されているチャットサービス
Slack

2. 学習者が英語執筆したドキュメント

執筆を通して用いた、英文は学習者自身が選択したものであり、

学習者との関連度が高い。

英単語頻度を算出1

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 $\text{Score}(t)$ を算出する.

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'

英単語頻度を算出2

ドキュメントの形式ごとに単語に重みを設定する.

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^t (Web文書)=1, W_1^t (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^t (自身が作成) = 2, W_2^t (他者が作成) = 1のように設定する.

英語例文出題機能

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

実装

学習者がアップロードしたマイクロソフト Office (docx, 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。

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 を使用した。

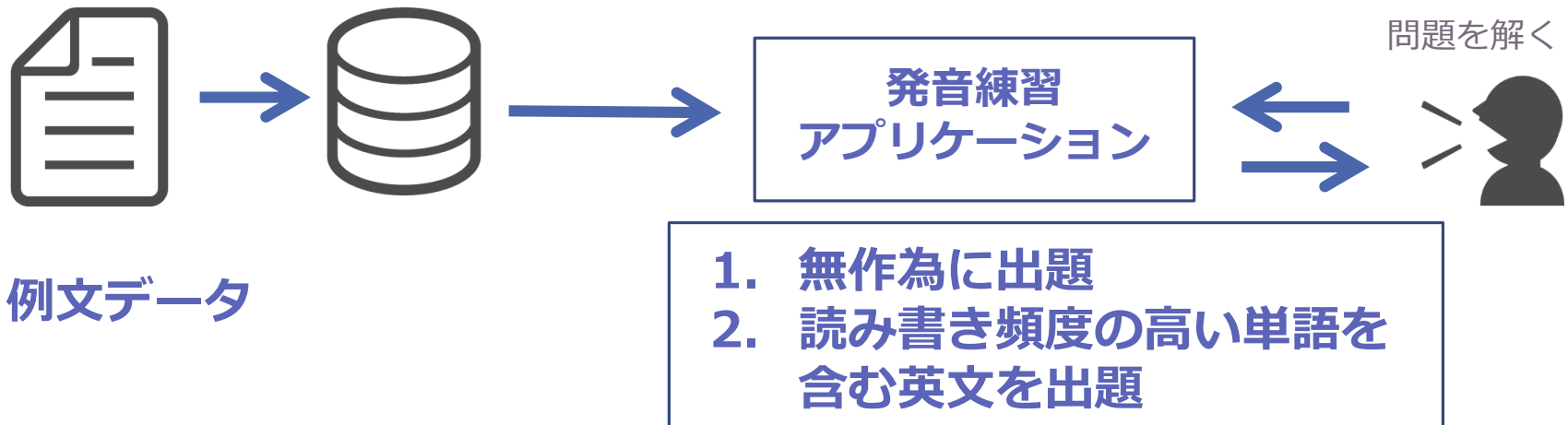
発話練習用の例文データベースとして、Academic Phrasebank と英借文ドットコム of 英文を利用した。

The screenshot shows a web application interface for English practice. At the top, there is a button labeled "Choose a file...". Below it, a section titled "INPUT TEXT" contains a text area with the following content: "Extraction of a Developer's Programming Contribution based on Source Code Profiling", "Takayuki Kasai Kosuke Takano", and "Department of Information and Computer Sciences, Faculty of Information Technology". Below the text area, there are three buttons: "GENERATE QUESTION", "NEXT QUESTION", and "CORRECT ANSWER". Below these buttons, there is a text input field containing the text "Studies on X represent a growing field". Below the input field, there is a button labeled "click to Speak". Below the "click to Speak" button, there is a section labeled "Result" which is currently empty. At the bottom, there is a button labeled "Finish".

実験

実験目的

提案システムを用いた学習方法(読み書き頻度に基づいて学習する例文を選択する)とベースライン(無作為に例文を抽出する)とで、学習意欲と会話成功率の面で比較することで提案システムの有用性を検証する



Academic Phrasebank

学術論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

総例文数560

	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 ...

英借文ドットコム

ビジネスや一般的なマナーを考慮した英語例文集

総例文数261

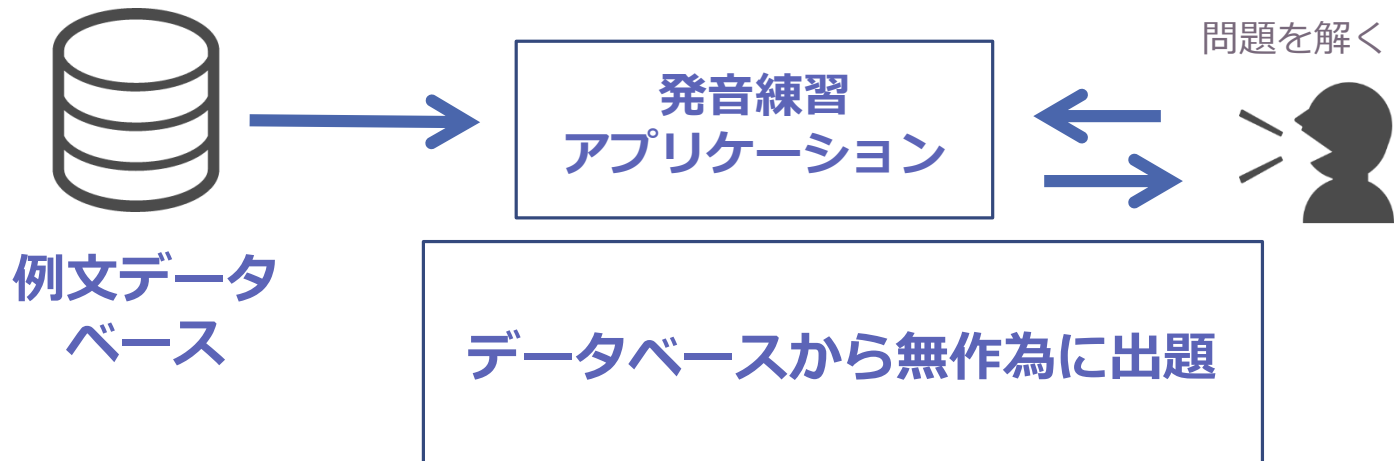
	英語例文
1	The company also plans to address a number of key issues including a focus on overseas production and procurement capabilities.
2	Investigation committee has released a draft of preventive actions to address the problem of the "sick house" syndrome.
3	Licensing agreement will need an assignment 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.

実験手順

1. 被験者1に英文で記述されたドキュメントをアップロードさせる
2. 被験者1に実験システムを用いて,アップロードされたドキュメントの頻出スコアをもとにした英文を出題してたものを学習させる
3. 被験者2にランダムに例文データから出題した問題を用いて学習させる
4. それぞれの被験者に出題された英文についてアンケートをとる
5. 被験者1, 被験者2にランダム単語と提案システムが抽出した単語を合わせた問題を出題し正答率を記録する

ランダム出題システム

- 実験システムと同一の例文データベースを用いる
- 一回の学習での出題数は実験システムと同一
- 英文をランダムに出題する



発音テストシステム

- ランダム単語と提案システムを用いて抽出した例文を1:1の割合で主題する.
- 実験の学習フェーズで出題したものと同一の題材から主題する.