

あなた専用の速習ロードマップ！

「〇〇語で、語学以外の専門分野(例:科学)を勉強する語学学習者」を支援したい→**専門分野の語彙の学習が特に重要**

- ・学習者ごとに学習すべき語彙が異なる→「あなた専用」(個人化)
- ・学習者の専門分野と語学の両方を熟知した人材は少ない
- 人手で教材を作ることは困難。自動化したい。

「覚えるべき語」がわかったとして、語のどの意味を覚えるのか？

・語と学習ゴールとの関係は？

専門分野の細かい分類は難しい。そこで、学習者に、自分が読めるようにしたい本を挙げてもらう。このとき、語彙学習のためには、次が知りたい:

- どの語を覚えると、どの本が読める？
- どの本が読めると、どの語が覚えられる？

どうやって「あなた専用」にする？

- × システムを一か月使ってもらえれば...
- × TOEIC等の民間試験を受けてもらえれば...
- ✓ 数十分の単語テストを受けてもらえれば...

研究テーマ1:

既存手法: 学習者が覚えた方がよい単語の推薦

問題点: 単語には複数意味がある。どれを覚えれば？

研究目標: 単語の用例の違いを自動認識して、学習者が覚えた方がよい用例を教える

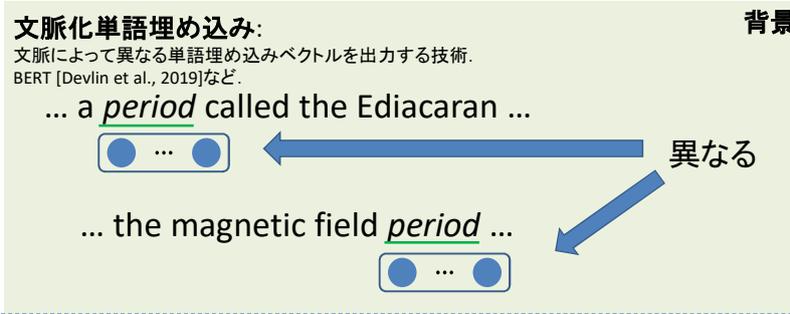
1年半の達成状況の凡例: 大枠はできた部分, 今後の課題

研究テーマ2:

既存手法: 学習者が本を読むために必要な語の判定

問題点: 学習に役立つのは、ギリギリ読めるぐらいの本ある語を知っていると、本が簡単すぎるのか、ギリギリ読めるのか、度合いがうまく測れない

研究目標: 「本が読める度合い」の定量化と、学習すべき本の推薦



背景

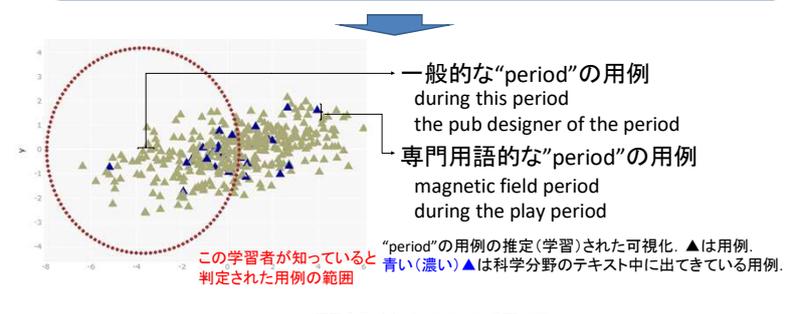
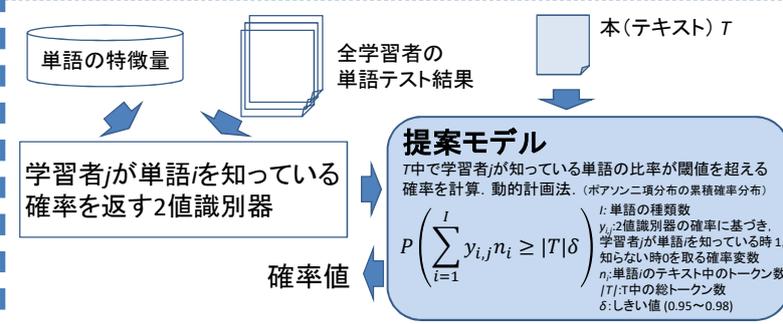
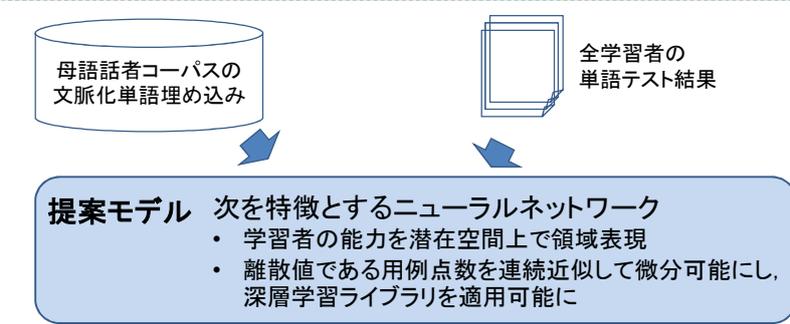
a): どの語を覚えると、どの本が読める？

テキスト中の語の95%以上~98%以上(述べ語数)を知っている → テキストが読める [Laufer 1989, Nation 1989, Nation 2006, Laufer 2010] (応用言語学分野で複数の実験報告)

b): どの本が読めると、どの語が覚えられる？

- 全単語を知っている本, 出てくる単語が全く分からない本 → 学習に役立たない

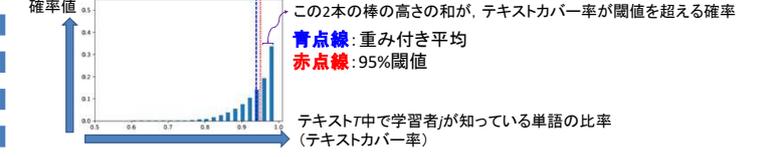
ギリギリ読めるぐらいの本が学習に役立つ! [Krashen, 1985]



評価: 単語テストと読解テストの両方を被験者100人に解かせたデータセットを新たに作成。単語テスト結果だけから、各学習者が読解テストの各設問に正答/誤答するかを予測。表中の数値はMean Average Precision多数の(読解課題テキスト, 設問文)ペアの中から、各学習者が正答するペアの検索精度に相当。



	Methods	Laufer	Short	
Conventional	VST	0.4880	0.5437	2値化
	H-LR	0.5797	0.5304	
	H-NN	0.5810	0.5393	
	H-LR+GLOVE	0.5113	0.5613	
	H-NN+GLOVE	0.5250	0.5631	
W. Avg.	A-LR	0.4880	0.4885	確率×頻度の重み付き平均 (低い方に引きずられてしまい、既存研究の閾値が使えない)
	A-NN	0.4880	0.4885	
	A-LR+GLOVE	0.4880	0.4885	
	A-NN+GLOVE	0.4880	0.4885	
Proposed	UA-LR	0.6314	0.6533	提案手法
	UA-NN	0.6172	0.6533	
	UA-LR+GLOVE	0.6305	0.6743	
	UA-NN+GLOVE	0.6159	0.6524	



今後の進展:

- ・学習者が知っている用例のデータセット作成 & 半教師あり学習
- ・多義語(多クラス)に対応可能な主要な用例の検知・可視化
- 主要な用例 = 外れ値(異常)ではない用例の発見
- 教師なし深層異常検知を提案モデルに組み込む [江原, 言語処理学会2020, 人工知能学会2020]

関連成果: NLP若手の会第14回シンポジウム奨励賞(7/93件), PACLING poster, ICDM MLCS workshop, ICMLA 2019 (採択率28.5%), NL4XAI workshop, Learning Analytics and Knowledge (LAK 2020) poster

今後の進展:

- ・単語テスト結果 & 学習者が読めるテキストの両方を記録したデータセット作成 & 半教師あり学習
- ・あるテキストを読むことにより学習者が覚えられ単語数を利得としたとき、利得の期待値と分散(リスク)を両方考慮した適切な学習用テキストの推薦 → 推薦のための強化学習

関連成果: 第44回教育システム情報学会全国大会大会奨励賞(論文+発表審査で4/79件) ICMLA 2019 (採択率28.5%), 人工知能「私のブックマーク 語彙学習支援システム」(解説記事)