

機械可読辞書データの活用による語学支援サービスに関する研究

A study of language learning and teaching service by utilizing vocabulary data in machine-readable dictionaries

小野 真嗣 (Masatsugu ONO)

The author aims to propose highly effective, efficient and satisfactory service of language teaching and learning by means of educational technology from the viewpoint of usability engineering. In the field of computational/quantitative linguistics based on corpora data, many linguists have been numerically analyzing lexical characteristics for each English vocabulary and compiled them as a vocabulary list which is machine-readable since 1960s. On the other hand, a lexical database of semantic relations called WordNet has also been developed in the field of computing science.

This study consists of three components. One is to enrich, refine and integrate already-existing vocabulary information in terms of “frequency”, “difficulty”, and “relativeness”. Another is to develop the system of language learning and teaching service with the integrated vocabulary information. And the other is to evaluate the system from usability engineering point of view. So far, the author reached and conducted the first two kinds of studies, succeeded in deriving the newly-integrated vocabulary list and visualizing the word relations in each experiment.

1 目的と背景

本研究の目的は、教育工学の観点から外国語の学習や指導に対して効果的かつ効率的な手法を求め、指導者や学習者に満足感、充実感、達成感を満たす手法を提案することである。教育にかかる一連の活動をサービスとして捉え直し、指導者や学習者をそのサービスのユーザーとして再定義し、教育工学に基づいて開発された指導法や教材利用を、ユーザーエクスペリエンス(UX)の尺度に基づいて評価することを目指すものである。

外国語学習には語彙の習得に多大な労力を要する。本研究では、既に公開されているものの活用度合いが低い機械可読の語彙情報を収集・集約してデータベース化を図り、その活用を進めながら、学習や指導の労力軽減を図りながら、新たな学習支援サービスを構築するというものである。言語教育にかかる作業効率を高め、学力や運用力といった実務上の効果を引き上げながら、ユーザーとなる学習者や指導者に対して高い満足度を提供する一連の学習や指導にかかる提供サービスの開発が本研究の最終目標となる。

2 既存語彙データの再評価と統合

言語学および情報学の分野において、これまでに多数の語彙分析が行われており、多くは機械可読辞書の形でまとめられている。1960年代に始まる頻度分析を皮切りに、90年代の難易度分析、また上位語や下位語など語彙の階層や同義性などの関係を網羅した概念辞書が開発されており、表1に一部を列举した。これらそれぞれの研究で得られた分析結果としての語彙情報を教育へ応用し、学習者向けに提供できる頻度および難易度情報として一元化し、約15000語からなる統合語彙情報コンテンツを作成し、学習サービスの基盤情報として集約した。

3 統合情報に基づく語彙関係性の可視化

統合語彙を基に数値で示される各語彙のデータについて、学習用途向けに難易度と関係性を可視化する実験に着手した。図1は可視化実験結果の一例である。具体的には統合語彙

を検索キーに、WordNet の synset 毎に上位語と下位語を抽出し、検索語を中心に図の中位に同意語を配列しつつ、下部にさらに具体性を持つ語群となる下位語を、上部に抽象性が高まる上位語を配置し、各語の難易度について赤、緑、青の順に難易度が上がるよう色付けを施して表示した。今後はこの関係性の図を踏まえつつ、語義、頻度、難易度、例文といった総合的な語法情報への拡充化も視野に、洗練させる予定である。

4 今後の計画

今後は、語彙学習に向けた情報提供サービスのためのシステム開発と、ユーザーエクスペリエンス(UX)による言語学習評価モデルの検討を進めていく予定である。

参考文献

1. 小野真嗣, 曾我聡起: ユーザーエクスペリエンスによる言語学習評価モデルの検討 – VR 英会話活動の試行実践 –, コンピュータ利用教育学会北海道支部(CIEC-Hokkaido) PC カンファレンス北海道 2021, PC カンファレンス北海道論文集, pp.30-31, (2021).
2. 小野真嗣, 曾我聡起, 菊地真人, 田邊鉄: 既存学習語彙表の再評価の試み – 学習サービスの視点から –, コンピュータ利用教育学会(CIEC) PC カンファレンス 2022, PC カンファレンス論文集, pp.252-255, (2022).
3. 細川大和, 西山幹泰, 小野真嗣, 按田将吾, 菊地真人, 大園忠親, 曾我聡起, 田邊鉄: 英語学習語彙表に基づく WordNet 収録語彙の抽出と可視化, 電子情報通信学会 人工知能と知識処理研究会(AI) SMASH22 Summer Symposium, 電子情報通信学会技術研究報告(IEICE Technical Report), pp.64-69, (2022).
4. 小野真嗣, 曾我聡起: 学習語彙情報の収集と語学教育への応用, コンピュータ利用教育学会北海道支部(CIEC-Hokkaido) PC カンファレンス北海道 2022, PC カンファレンス北海道論文集, pp.27-28, (2022).

表 1 語彙表の種別と例

語彙表	例	語彙数
語彙表	Kucera & Francis (1967)	-
頻度表	Kilgariff(1995)	6318
	杉浦リスト 1 (2002)	5776
	杉浦リスト 2 (2002)	7286
学習語彙表	北大語彙表 (1996)	7453
	JACET 8000 (2003)	8000
	COCET 3300 (2007)	3300
	ALC SVL 12000	12000
概念辞書	WordNet	150000

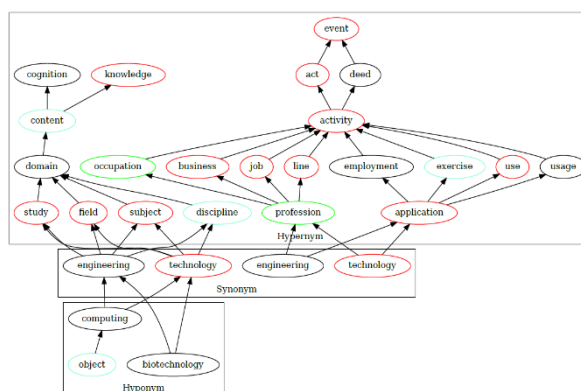


図 1 WordNet に基づく語彙情報の可視化結果例