

博士学位論文要旨等の公表

学位規則（昭和28年4月1日文部省令第9号）第8条に基づき、当該博士の学位の授与に係る論文の内容の要旨及び論文審査の結果の要旨を公表する。

氏名 小野 真嗣

学位の種類 博士（理工学）

報告番号 甲第34号

学位授与の要件 学位規程第4条第2項該当

学位授与年月日 令和8年3月21日

学位論文題目

語義間距離モデルによる語彙ネットワークの構造分析

論文審査委員 主査 教授 曾我 聡起

委員 教授 田邊 鉄

委員 教授 小林 大二

学 位 論 文 要 旨

理工学研究科 理工学専攻

学籍番号： D2210040

氏 名： 小野 真嗣

(論文題名)

語義間距離モデルによる語彙ネットワークの構造分析 Structural Analysis of Lexical Networks Based on Sense Distance

本研究は、語彙表が本来的に抱える「語彙間の意味的關係を扱えない」という構造的限界に対し、語義間距離モデルを導入することで語彙体系を構造として再記述し、語彙研究および語彙学習支援に新たな分析軸を提供することを目的としたものである。従来の語彙表は頻度や難易度といった単一指標に基づき語を並列的に提示する形式をとるため、語彙体系の中心性・階層性・周辺性といった構造的特徴を十分に反映できなかった。本研究では、WordNet の gloss を文埋め込みモデル (Universal Sentence Encoder) によりベクトル化し、語義間のコサイン類似度を距離として扱う語義間距離モデルを構築することで、語彙表に意味的構造を付与する工学的基盤を整備した。

構築した語義間距離モデルを JACET8000、北海道大学英語語彙表、BNC 語彙表の三語彙表に適用し、語義間距離の分布特性を比較した結果、同義語・反意語・上位語／下位語の三種の語義関係は明確に異なる距離帯を示し、語彙ネットワークには中心領域・境界領域・階層的連結領域といった複数の意味的領域が存在することが確認された。さらに、Dunn 検定および Welch の t 検定により、語彙表間の距離分布には統計的に有意な差が認められ、語彙表の設計思想 (難易度・教育課程・頻度) が語彙体系の構造に反映されることが示唆された。具体的には、JACET8000 は基礎語彙が中心領域を形成する密集型構造、北海道大学英語語彙表は教育課程に沿う階層型構造、BNC 語彙表は頻度語彙の多義性と語域の広さを反映した拡散型構造を示した。

これらの特徴を統合的に捉えることで、語彙ネットワークは多義語を核とする中心語彙、中間語彙、周辺語彙からなる三層構造を持つことが明らかとなり、語義間距離が語彙体系の中心性・階層性・分散性を捉える指標として有効であることが示された。また、語義間距離を可視化に組み込んだ学習支援システムの試作と評価を通じて、語彙ネットワークの構造を視覚的に提示することが学習者の語彙理解、自律性、有能感の向上に寄与する可能性が示された。特に、距離情報を反映した動的可視化は、語彙の意味的まとまりや語彙体系の広がりを直観的に把握できる学習環境として有効性が示唆された。

さらに、語彙表分析に基づく教育的示唆として、基礎語彙の中心性は段階的語彙指導の妥当性を支持し、高難度語彙の孤立性は文脈依存型支援の必要性を示した。また、頻度上位語であっても意味的中心性が必ずしも高くないことから、頻度情報と語義間距離を組み合わせた語彙指導の重要性が示唆された。加えて、本研究の枠組みはCOCAやOxford 3000/5000といった国際的語彙表、さらにはCEFR (A1～C2) との接続可能性を持ち、語義レベルでの国際的難易度推定モデルの構築に向けた基盤としても意義を有する。

一方、本研究には語義粒度の辞書依存性、語彙表の一般化可能性、語彙レベル再定義の実証的検証といった課題が残されている。今後は、大規模言語モデルを活用した語義表現の精緻化、多言語語彙表への適用、CEFR との統合的分析、および教育現場での語彙レベル再編成の実証研究が求められる。

本研究は、語義間距離という共通の枠組みにより、語彙体系の構造分析、学習支援システム的设计、語彙指導および語彙レベル再定義に関する知見を統合的に提示した点に新規性を有し、語彙研究と語彙教育の双方に対して理論的・実践的貢献を果たすものである。

論文審査の結果の要旨

本論文「語義間距離モデルによる語彙ネットワークの構造分析」は、従来の語彙表研究が抱えてきた「語彙間の意味的關係を扱えない」という構造的限界に対し、情報工学的アプローチを導入することで新たな分析枠組みを提示したものである。語彙表は頻度・難易度・教育課程といった基準に基づき語を並列的に提示する形式をとるため、語彙体系の中心性・階層性・周辺性といった構造的特徴を十分に反映できなかった。本研究はこの課題に対し、WordNet の gloss を文埋め込みモデル (Universal Sentence Encoder) によりベクトル化し、語義間のコサイン類似度を距離として扱う「語義間距離モデル」を構築することで、語彙表に意味的構造を付与する工学的基盤を整備した点に大きな新規性がある。なお、WordNet の gloss とは、各語義に付与された短い定義文であり、辞書における語義説明に相当するテキストである。

本研究の情報工学的貢献は多岐にわたる。第一に、語義定義文を高次元ベクトルとして扱うための大規模処理パイプラインを構築し、WordNet 全117,659語義を対象とした gloss ベクトル化を実現した点である。特に、文埋め込みモデルの選定、前処理、分散処理による高速化など、実装面での工夫は本研究の技術的基盤を支えている。第二に、語義ベクトル間の距離を統一的な尺度として扱うことで、語彙表の内部構造を数理的に比較可能にした点である。従来の語彙表研究では、語彙の重複率やレベル分布といった表層的比較に留まっていたが、本研究は語彙表の設計思想（難易度・教育課程・頻度）が語彙体系の構造にどのように反映されるかを統計的に検証し、語彙表研究に新たな分析軸を提供した。

構築した語義間距離モデルを JACET8000、北海道大学英語語彙表、BNC 語彙表に適用した結果、同義語・反意語・上位語／下位語の三種の語義関係は明確に異なる距離帯を示し、語彙ネットワークには中心領域・境界領域・階層的連結領域といった複数の意味的領域が存在することが確認された。さらに、Dunn 検定および Welch の t 検定により、語彙表間の距離分布には統計的に有意な差が認められ、語彙表の設計思想が語彙体系の構造に反映されることが示された。具体的には、JACET8000 は基礎語彙が中心領域を形成する密集型構造、北大語彙表は教育課程に沿う階層型構造、BNC 語彙表は頻度語彙の多義性と語域の広さを反映した拡散型構造を示した。これらの特徴を統合的に捉えることで、語彙ネットワークは多義語を核とする中心語彙・中間語彙・周辺語彙からなる三層構造を持つことが明らかとなり、語義間距離が語彙体系の中心性・階層性・分散性を捉える指標として有効であることが示された。

さらに本研究は、語義間距離を可視化に組み込んだ学習支援システムの試作と評価を行い、語彙ネットワークの構造を視覚的に提示することが学習者の語彙理解、自律性、有能感の向上に寄与する可能性を示した。距離情報を反映した動的可視化は、語彙の意味的まとめりや語彙体系の広がりを直観的に把握できる学習環境として有効性

が示唆されており、情報工学的成果が教育的応用へと接続されている点も高く評価できる。

加えて、語彙表分析に基づく教育的示唆として、基礎語彙の中心性は段階的語彙指導の妥当性を支持し、高難度語彙の孤立性は文脈依存型支援の必要性を示した。また、頻度上位語であっても意味的中心性が必ずしも高くないことから、頻度情報と語義間距離を組み合わせた語彙指導の重要性が示唆された。さらに、本研究の枠組みは COCA や Oxford 3000/5000、さらには CEFR (A1~C2) との接続可能性を持ち、語義レベルでの国際的難易度推定モデルの構築に向けた基盤としても意義を有する。総じて、本論文は語彙研究に情報工学的アプローチを導入し、語彙体系の構造理解に新たな視座を提供した点で高く評価できる。語義間距離モデルの構築、語彙表の構造分析、可視化システムの試作と評価、教育的示唆の提示という一連の成果は博士後期課程にふさわしい学術的貢献を備えており、今後の語彙研究・語彙教育の発展に寄与するものと認められる。

最終試験（口頭試問）は、博士論文公聴会に引き続き実施され、提出論文の内容に基づき、研究目的の妥当性、方法論の理解度、結果の解釈の適切性、ならびに研究の独自性と発展可能性について詳細な質疑が行われた。本試験では、研究の基盤となる語義間距離モデルの構築意図、語彙表比較の方法論的妥当性、語彙ネットワーク構造の解釈、そして教育的応用の可能性など、多岐にわたる観点から議論が深められた。

副査Aからは、語彙表における「難易度」概念の扱いについて、理工学的視点からの重要な指摘があった。従来の語彙研究では、難易度は教育学的・語彙学的知見に基づき、頻度・習得順・教育課程などの指標から定義されてきた。しかし、これらは一定の恣意性を含む可能性があり、工学的研究としては、語義間距離モデルによって算出される中心性・周辺性といった構造的特徴から、難易度を再定義する方向性もあり得るのではないか、という趣旨である。

これに対し受験者は、本研究の主目的が「既存語彙表情報の拡張」にあり、既存の難易度・頻度情報に語義間距離に基づく構造情報を付加することを目指したものであると説明した。すなわち、難易度そのものを再定義することは本研究の射程外であるが、語義間距離モデルが難易度の新たな指標となり得る可能性は認識しており、今後の発展的課題として位置づけられることを述べた。この点については、研究の方向性と工学的応用の可能性を踏まえた適切な回答であったと評価できる。

副査Bからは、JACET8000・北海道大学英語語彙表・BNC 語彙表の三つを研究対象とした理由について質問があった。受験者は、これら三語彙表がいずれも6000~8000語規模で比較可能であり、かつ設計思想が「難易度」「教育課程」「頻度」と大きく異なる点に着目したと説明した。異なる設計思想を持つ語彙表を語義間距離モデルにより比較することで、語彙表の構造的特徴がどのように現れるかを検証する意図があったことを明確に述べた。

この説明は、研究デザインの妥当性を裏付けるものであり、語彙表比較研究としての意義を十分に示すものであった。

さらに副査 B からは、語彙ネットワーク構造が当初から予期されていたものなのか、それとも分析の過程で新たに明らかになったものなのかについて質問があった。受験者は、当初は語彙表のリスト情報しかなく、語義ネットワークの構造的特徴は完全に未知であったと説明した。その上で、WordNet の gloss を USE によりベクトル化し、コサイン類似度に基づく語義間距離を算出し、統計分析を行うことで、結果として「密集型」「階層型」「拡散型」という三類型が導かれたことを述べた。この回答は、研究が探索的であり、未知の構造を明らかにした点が本研究の成果であることを示しており、博士研究としての独自性と価値を十分に説明するものであった。

以上の質疑応答を通じて、受験者は研究内容を深く理解しており、論理的かつ誠実に回答した。研究目的・方法・結果の整合性は明確であり、語義間距離モデルの構築と語彙表の構造分析という工学的アプローチは新規性と妥当性を備えている。また、語彙ネットワーク構造の発見は語彙研究・教育工学の双方に貢献する成果であると評価された。

主査および二名の副査による審査の結果、本研究内容は博士（理工学）に値する内容であるとの評価で一致し、最終試験の判定は「合格」とする。