博士論文審査及び最終試験の結果

審査委員（主査）　吉冨　朝子　　印

学位申請者　　小谷　尚子

論　文　名　　Extraction of English Domain-Specific Words from Specialized Corpora in Economics and Finance

【審査結果】

審査委員会は吉冨朝子を主査とし、主任指導教員の投野由紀夫教授、副指導教員の工藤洋路教授、外部委員として佐野洋本学名誉教授、特定非営利活動法人 CLIL 教員研修研究所(元東洋英和女学院大学教授)の笹島茂理事長の 5 名で構成された。上記論文の公開審査ならびに最終試験は 2026 年 8 月 3 日に行われ、審査委員会は全員一致で、申請者に対し博士（学術）の学位を授与するのに相応しいとの結論に達した。

【論文概要】

本論文は、経済・金融分野における専門英語語彙（technical words）が、その制度的目的によって、どのような意味特徴を持ち、いかに体系化されているかを定量的なコーパス分析と専門家視点に基づいた定性的な意味分類によって明らかにし、その教育的含意を考察するものである。

特定の目的のための英語（English for Specific Purposes：以下 ESP）教育における専門用語の知識は、経済・金融分野の実務者だけでなく大学で経済や金融について学ぶ学生にとっても主だった経済指標や金融市場での一次情報に触れ理解を深める際に重要である。既に当該分野でも分野固有の専門用語を収録した経済学辞典や金融用語辞典は存在するものの、近年の急速な産業構造の変化や金融制度の改正など、現在ニーズに沿った専門語彙を十分に反映しているとは言えない。本研究では、専門分野の文書と既存の教材や辞書との間のギャップを埋めるため、経済・金融分野の 4 種類の専門分野コーパスを構築し、特徴語の抽出を行い、それら特徴語を専門家視点に基づき意味分類を施した。専門用語が文書タイプの制度的目的によって体系化されることを明らかにし、さらに専門分野コーパスから教育的な配慮の上で特徴語を含む実用例を選び出し用例集も作成した。

先行研究は 3 つの観点からまとめられている。第一に語彙研究を概観し、既存の専門語彙の判定基準が研究者の判断に依存するところが多いことから、本研究では、統計的指標である単語頻度-逆文書頻度（Term Frequency-Inverse

Document Frequency：以下 TF-IDF 値）を用い、その分野で特徴的な語を抽出することとした。第二にジャンル・レジスター研究では、ジャンル分析やレジスター分析によって、語彙・文法の体系的な違いが示されていることを概観した上で、制度的目的が語彙選択に与える影響を、複数の実務文書で比較した研究は限られていることに着眼した。第三に、コーパスを用いた計量的手法を用いた研究を概観し、TF-IDF 値による特徴語抽出をしたものは数多くあるものの、異なるジャンルを比較し、第二言語習得理論も考慮して教育への応用を検討した研究は少ないことを指摘した。

以上を踏まえ、本研究では 3 つの研究設問を掲げた。(1) 専門分野コーパス（4 種）から抽出された語彙項目は経済・金融分野に固有の専門語を構成するか[対比的 TF-IDF による抽出語彙の妥当性の検証]。(2) コーパスのジャンル特性および抽出語彙のコーパス相互の偏り具合の定性分析を通じて特定される経済･金融文書における専門語の特性はどのようなものか[各コーパスの語彙群の比較分析を通じた意味カテゴリーの体系化]。(3) 専門用語リストに加え、経済・金融分野の ESP 教育において、とくに実務者（であり中級水準の学習者）を対象とした効果的実例文とはどのようなものか[具体的には専門用語が主語や目的語となる用例の教育学的な検討]。

分析は、大きく 5 つのステップで進められた。1）4 種類の専門分野コーパスを構築した。比較対象となる参照コーパスとして現代アメリカ英語のコーパス（The Corpus of Contemporary American English、以下 COCA）を利用した。2） 各コーパスについて TF-IDF 値を算出し、特徴語リストを作成した上で、対比的 TF-IDF により抽出語彙の妥当性を検証した。3） 抽出した特徴語について、専門家視点に基づきコーパス間の語群の偏り具合を分析し、経済・金融分野の専門用語の意味カテゴリー分類を行った。4） 専門分野コーパスを依存構造解析し（自然言語処理ライブラリ spaCy）、抽出した特徴語が、主語並びに目的語として現れる用例を自動抽出した。5） 抽出された用例を教育的観点（実務を担う、中級水準の英語力を有するなど）から評価し、経済・金融分野 ESP の教材として利用可能な例文集を作成した。

分析対象となったのは、業務上必要となる開示文書（Form 10-k 年次報告書と Prospectus 目論見書）と、マーケット情報を得るための経済専門メディア文書（*The Economist* 誌と *The Financial Times* 誌）という、二種類の文書から構成される、4 つの専門コーパスである。Form 10-k 年次報告書コーパスは、米国証券取引委員会に提出された、ダウ工業株 30 種平均株価指数に採用された企業の年次報告書から作られ、約 210 万語から成る。企業の法的義務としての開示文書であり、規制・コンプライアンス語彙が多く含まれている。Prospectus 目論見書コーパスは、JP Morgan、Morgan Stanley、UBS が発行した投資信託目論見書から構築したもので、約 8 万 8 千語から成る、投資家への説明という明確な目的を持つ文書群である。*The Economist* コーパスは約 12 万語で、経済政策や国際情勢を分析的に論じる記事が多い。*The Financial Times* コーパスは約 177 万語で、マーケット動向や企業活動を即時的に報道するメディアである。

参照コーパスとしては、具体的には一般学術英語（COCA の Academic ジャンルのうち 5 文書から無作為抽出した 5 万語、以下 COCA Academic）を採用し、コーパス構築時にバーストとして現れた COVID 関連のキーワードは、分析結果への影響を避けるため除外した。

分析に際し前処理として、Form 10-k や Prospectus は PDF 形式からテキスト形式に変換し、正規表現を利用して分析に不要な文字列（図表部分のテキストやヘッダーやフッター部分)を削除した。一方、*The Economist* と *The Financial Times* は ProQuest というデータベースから取得したテキストを用いた。抽出したテキストには、ページ番号や記号、数値だけの行など、分析に不要な情報が含まれているため、正規表現を用いてノイズを除去し、すべてのコーパスに同一のクリーニング基準を適用した。続いて、RNNTagger を用いて品詞タグ付けとレマ化することで、語形変化による影響を抑えた。さらに、NLTK のストップワードリストを用いて一般的な機能語を除去した。これら一連の前処理を統一することで、コーパス間の比較可能性と分析の再現性を確保した。

抽出指標として採用した TF-IDF 値は、ある語が特定の文書では頻繁に出現する一方で、他の文書ではあまり出現しない場合に高い値を与える指標で、「ある文書の中で重要でありながら、他の文書では一般的ではない語を抽出する」ことが可能である。4 種類のコーパスの各文書タイプの特徴を比較できるよう、文書単位で重み付けを行う TF-IDF 値を選択した。文書間の類似度計算はコサイン類似度を用いた。専門分野コーパスどうし、並びに参照コーパスである COCA Academic との比較をそれぞれ行った。

以下、結果である。研究設問(1)については、分析の結果、金融文書同士では語彙分布の類似度が高い一方で、一般学術英語の語彙とは異なる語彙分布を示していることがわかった。つまり高い TF-IDF 値の語彙が一般語ではなく、金融分野に特徴的な語彙を反映していることが示された。また、Form 10-k コーパスは法令に基づいて作成される文書であり、構成や表現が一定程度標準化されているため、コサイン類似度が比較的高くなった一方で、COCA Academic は多様な学問分野を含むため、金融分野に特有の語彙は相対的に少なく、コサイン類似度も低くなった。すなわち、「TF-IDF 値で抽出された高い TF-IDF 値の語彙は、経済・金融分野を特徴づける領域固有の専門用語である」という結論が導き出された。

研究設問(1)については、さらに辞書との照合による補完的な検証も行った。Oxford Dictionary シリーズの専門用語事典『A Dictionary of Finance and Banking』(2014)および日本で最大規模の見出し語を掲載する金融専門辞書『経済金融ビジネス英和大辞典』（2012, 日外アソシエーツ）を用い、両辞書に共通する見出し語を基準に、抽出語と辞書との照合を行ったところ、金融辞書に掲載されている基本的な専門用語の多くが抽出語と一致した。一方で、コーパスごとに、辞書には収録されていない語として、10-k では cloud、AI といった新しい金融技術に関する語、*The Economist* や *The Financial Times* では

fintech、Bitcoinといった暗号資産関連の語が見受けられた。これにより、今回のコーパス分析によって新しい専門用語を実際の使用状況に基づいて抽出できる利点が明らかになった。

研究設問(2)については、4 種類のコーパスにおいてそれぞれ TF-IDF 値の高い特徴語を確認したところ、*The Economist* では、経済全体についてその背景を含めて伝え、政策や社会的影響を論じるという経済誌の特徴を反映して、マクロ経済や経済政策を中心とした語彙が上位を占めた。*The Financial Times* では、即時性の高いマーケット報道というメディアの特徴を反映して、市場動向や金融イノベーションに関する語彙が上位を占めた。一方、Form 10-k では、企業が投資家や規制当局に対して財務状況を正確に報告するという目的を反映して、企業業績や会計処理に関する語彙が中心となった。そして Prospectus では、投資家が商品内容を理解し、適切な投資判断を行うことを支援するという目論見書の目的を反映して、投資商品や資産運用に関する語彙が上位を占めた。以上のことから、4 種類のコーパスがいずれも金融分野に属しているものの、それぞれ異なる金融文書を出す組織の制度的目的を持つため、TF-IDF 値の高い特徴語の構成が体系的に異なっていることが示された。これは、その文書を作成・公表する組織が、それぞれ異なる制度的役割や目的を持っていることと関係しており、特徴語の違いは各組織が読者や利用者に対して「何を伝えることを目的としているのか」という制度的な役割の違いを反映していると解釈された。

さらに金融会社のキャリアを持つ著者は、その専門家の観点で抽出した特徴語を意味カテゴリーごとに整理した結果、マクロ経済・金融市場と投資・リスク管理・ガバナンス・金融革新の 5 つの意味領域に分類できた。Form 10-k コーパスでは、企業の収益性を表す語、すなわち投資家が企業の経営状況を評価する語彙や、企業の財政状態を示す会計用語、さらに、特に法的責任や規制遵守という Form 10-k の文書を作成する組織の制度的役割を反映した語彙が目立った。Prospectus コーパスでは、投資商品の内容を説明する語彙や、投資家保護や契約条件に関わる語彙、環境に配慮した ESG 投資に関する語彙が確認された。つまり、同じ『開示文書』であっても、Form 10-k は企業の説明責任と法令遵守を目的としているのに対し、Prospectus は投資家の意思決定をサポートする目的、また善管注意義務を意識した語彙選択を反映していることが判明した。対してメディア・コーパスについては、*The Economist* では、景気動向やマクロ経済の変化を捉える語彙や、政府の政策や制度との関係から分析するという特徴を反映した語彙、さらに新しい金融テクノロジーや金融商品についても積極的に取り上げていることが反映される語彙が見られた。つまり、*The Economist* は経済動向や金融市場の動きの背景を踏まえながら、政策や社会への影響を分析することも重視している一方、*The Financial Times* は市場参加者を意識し、金融市場や資産価格にどのように影響がでているのかをより直接的に捉えるメディアであることが推察できた。どちらも金融メディアだが、メディアの異なる特性から、文書を作成する組織の制度的目的が異なるため、特徴語にも違いが生じていることが明らかとなった。

また、4 種類のコーパスを横断的に比較したところ、同じ意味カテゴリーの語彙であっても、各コーパスでは制度的目的に応じて異なる意味で語彙が使用されることを明らかにした。

次に作成した専門用語リストが、実際の金融実務文書に対してどの程度の妥当性を持つかを検証するため、本研究で作成した金融用語語彙リストを用いて、金融実務でよく使われている Bloomberg の記事 10 本を対象にカバー率を調査した。その結果、不動産、サステナブルファイナンス、政治・経済、信用市場など、中核的な金融分野では約 80%以上のカバー率が得られたことから、作成された語彙リストが実際の金融記事を読む際にも十分な妥当性を持つ可能性が示された。

以上に基づき、研究設問(2)については、TF-IDF 値によって抽出された専門用語が、実際の金融実務文書においても高いカバー率を示し、経済・金融分野に固有の専門語彙であることが実務的にも裏づけられた。メディアは解釈的な枠組みの提示を、Form 10-k はコンプライアンスを、そして Prospectus はリスク開示と将来志向の情報提供をそれぞれ重視していることが示された。

研究設問(3)に関する教育用用例は、全文を、自然言語処理ライブラリ spaCy を用いて依存構造解析し、各文の文法構造を分析した上で、コーパス全体で高い出現頻度を持つ特徴語が主語または直接目的語として現れる文を抽出するという方法をとった。次に、抽出した文の中から、専門用語と高い頻度で共起する動詞を優先して抽出することで、より実践的な使用パターンを示すことを試みた。さらに、文の長さ、明確さ、領域との関連性などを基準としてマニュアルで精選し、教育利用に適した例文を選定した。これらの用例は、5 つの意味カテゴリーごとに、構文パターンとともに示した。経済・金融分野の構文は、例えば、マクロ経済のカテゴリーでは、中央銀行が能動的な行為者として描かれる構文、金融イノベーションのカテゴリーでは、助動詞の could を用いて、まだ不確実な事象を推測的に語る構文が特徴的であった。このような構文パターンも特徴語と同様に、それぞれの文書が担う制度的な目的を反映していることが確認できた。

本研究の理論的意義は、経済・金融分野では各専門文書を作成する組織の制度的目的が語彙選択を影響する主要な要因であることを指摘した点である。方法論的な意義は、TF-IDF 値に基づく語彙抽出、ジャンル横断的な比較分析、そして文レベルでの構造分析という、3 つの分析手法を 1 つの枠組みの中で統合した点である。そして教育的な意義としては、実際の金融実務で使われる文章における語彙を反映しつつ、中級レベルの学習者にも扱いやすい、データに基づいた専門用語の選定ができた点である。

本研究は、コーパス言語学の分析手法を用いて、専門用語研究と ESP 教育をつなぐアプローチを提示できた。今後の課題としては、今回対象としなかった複合語の分析を行うこと、金融商品や規制の枠組みは常に変化しているため、コーパスを定期的に更新し、新しい金融技術や商品の語彙にも対応していくこ

と、FRB のベージュブックや FOMC 議事録、公式の政策声明といった、政策関連文書を分析対象に含めていくこと、本研究で採用した TF-IDF 値に基づく抽出と用例分析の枠組みを他の ESP 分野に展開していくこと、などが挙げられた。

【講評】

本論文は以下の点において非常に優れているとの評価がなされた。

1. 本研究は、専門的な金融文書における語彙選択と使用を規定する主要因として制度的目的を位置づけることにより、コーパス基盤の ESP 研究に貢献した。

2. 実際の業務文書から抽出した語彙リストおよび用例は、従来の辞書や高頻度語リストに依拠した教材では可視化しにくかった語彙のコミュニケーション機能を学習者に明示できたという点で、既存の ESP 教材とは異なる教育的価値を有している。

3. 本研究で生成された専門語リストおよび用例は、制度的文脈とカリキュラム目標に応じた ESP 教材開発のための基盤を提供し、学習者が学術的および職業的文脈において金融ディスコースに適切に関与するための教育資源の開発に寄与するものである。

4. 小谷氏が経済・金融分野を専門としてきたにも関わらず、博士後期課程でコーパス言語学や分析手法・技術（プログラミングを含む）について一から学び、従来からの関心であった経済・金融分野の英語教育資材開発に貢献できるような研究を遂行できた。

5. オリジナルコーパスを構築した上で、定量的分析・定性的分析を行った労作であり、制度的特性を抽出できた点は評価に値する。今後他の分野の同様な分析への応用や、生成 AI を活用した研究にもつながりうるなど、発展性が大いに期待できる研究である。

一方で、本論文について以下の質疑や指摘がなされた。

1. 全体的に経済・金融分野を専門とする者の視点で書かれているため、英語教育の観点からは説明が足りない部分があった。論文構成上、見出しをトピックセンテンス化するなど、もう少し読みやすい書きぶりにできたのではないか。また、経済・金融分野が専門である強みを生かして、英語教育との融合として、内容言語統合型学習（CLIL）の可能性について、もっと議論が展開できたのではないか。

2. 経済は活動であるから身体的アナロジーや基礎語彙との共起関係、金融は制度であるから抽象的概念や指標的・抽象的記号との結びつきがある、といったさらに抽象化された意味カテゴリーの違いや、分野別の構文パターンの違

いがあるのではないか。その違いが示せれば当該の専門分野だけでなく他分野の ESP 教育にも活かせるのではないか。

3. 今回作成できたリストをどのように教材開発につなげるのか。英語教育への示唆について、より具体的な提案ができたのではないか。抽出語が主語・目的語にあたる構文に用例を絞ったのはなぜか。頻度の高い複合語やコロケーションも扱ったほうが、教育的意義がより高かったのではないか。

4. 方法論について、通例参照コーパスは無色の大規模なものを用いるが、なぜ 5 万語の COCA Academic としたのか。分析方法として、特徴語抽出では対数尤度比などを用いた keyness が一般的だが、なぜ TF-IDF 値を採用したのか。TF-IDF 値を用いた場合の考慮すべき問題点は何か。RNNTagger では単数形しか抽出できていないため、抽出法の改善がありうるのではないか。今後は生成 AI を活用した特徴語抽出が可能になるのではないか。

5. ESP 研究である割には、ESP 研究に関する先行研究がやや不十分である。また、第二言語習得理論についても記述に偏りが見られ、理論の本質的な点が捉えられていない論述が見受けられる。

これらの指摘に対する小谷氏の応答は以下のとおりであった。指摘 1 および 2 については、指摘のとおりであり、そのような視点から今後の研究を行いたいとのことだった。指摘 3 に対しては、リーディング教材やプレゼンテーション活動への応用といった具体的な教材・指導案について回答した後、例文抽出ガイドラインに沿った用例リストを作成したことや、今回対象としている初・中級学習者に対しては基礎構文に必須の主語・目的語位置の抽出語が最もわかりやすいと判断したが、今後は複合語やコロケーションの抽出についても視野にあることを述べた。指摘 4 の方法論についての様々な質疑に対しては、TF-IDF 値の有効性およびその限界（偏在する中核語は特徴語で重みが下がること）・参照コーパスの活用法の改善・生成 AI を活用した分析の検討、などに関して、本論文の限界を認識しつつ、改善点に関して前向きな回答があった。さらに指摘 5 については、第二言語習得理論における着目点にやや偏りがあったことを認めつつ、理論に基づく英語教育の重要性を認識していることを強調した。

審査委員から出された指摘や質疑は、いずれも今後の研究の発展を期待しての提言であり、本論文がコーパス言語学に基づく ESP 研究として、大いに発展性が期待できる貢献であることを否定するものではなかった。いずれの指摘や質疑に対しても小谷氏は誠意ある回答をし、自らの研究の課題についても十分に自覚していた。また今後の研究に向けた展望についても、具体的な構想を述べることができた。

【総合評価】

最終試験は、2026 年 8 月 3 日 17:00-19:00 に実施され、小谷氏から約 40 分間にわたって論文要旨の発表があったのち、80 分間の質疑応答を行った。博士論文審査および最終試験での質疑応答の内容から、本論文は学術的意義および教育実践的意義の高い実証研究であると評価され、学位申請者が研究者としての資質を十分に有していると判断された。よって審査委員会は全員一致で、学位申請者が博士（学術）の学位を授与するにふさわしいとの結論に達した。