

## 論文審査の結果の要旨

論文提出者氏名 ロージャズ アンナ

ロージャズ アンナ氏の博士論文 *Multilingual computational lexicography: frame semantics meets distributional semantics* (多言語計算辞書学：フレーム意味論と分布意味論の接点) の審査結果について以下に報告する。本論文は、多言語辞書記述における理論的・技術的課題に挑んだ斬新な研究である。論文は全8章からなり、巻末には分析に使用したデータセットが収録されている

第1章 “Introduction” におけるコンピュータ辞書学についての概観と研究目的の提示に続き、第2章 “Frame approach to semantic analysis” では、認知言語学において辞書構築のために最もよく整備された意味理論である、フレーム意味論の概観を行い、同時にその課題を指摘している。第2章ではまた、コーパスに基づいた意味関係抽出手法として近年注目を集めている分布意味論を取り上げ、二つの分析手法による意味表示の特徴を論じている。

第3章 “Event structure for a multilingual FrameNet” は本論文の研究対象である posture verbs (姿勢動詞) に焦点を当て、英語についてこれまでなされてきたフレーム意味論に準拠の語彙記述の問題点を指摘している。特に、動詞のアスペクト属性について、英語に見られるような比較的限られたアスペクト形式と比べて、ロシア語等のスラブ系言語では多様なアスペクトが形態論の一部となっているケースを扱うための上位スキーマが提案される。

第4章 “Discovering verb classes in untagged corpora” は本論文の方法論の妥当性をイベント構造の領域で検証するべく、タグ無しコーパスに対して語の共起関係をもとに意味・文法的属性を導出している。ロシア語コーパスに基づいた動詞のアスペクト属性、および日本語コーパスに基づいた動詞の自他属性の機械学習による予測を試み、成功をおさめている。

第5章 “Selectional preferences across languages” は第2-3章の議論をふまえ、言語ごとの異なり、すなわち語彙化における意味属性の選好をどのようにモデル化するかという課題が取り上げられている。ここでも姿勢動詞 (英語ならば lie, sit, stand など) をサンプルとして、英語、日本語、ロシア語のデータにもとづいた分析が行われる。姿勢動詞は典型的な場面から非典型的な場面への適用範囲が言語ごとに大きく異なり、多言語比較には好適な分野となっている。本章ではアンケート調査に基づいて、三つの言語の共通点と相違点が示されている。

第6章 “Classifying selectional preferences” は前章で取り組んだ結合パターンをどのようにモデル化するかが考察されている。統計的学習モデルを修正すべく、いくつかの提案と検証がされる。文法情報を強化したデータセットでは大幅な精度の向上が見られた。一方、負例 (否定的データ) すなわち通常の言語使用では見られない結合を加えたデータセットでは、有意な精度の向上が見られなかった。

第7章 “Discussion: toward a unified semantics” は本論文全体の研究成果の上に立ち、語彙記述におけるいくつかの理論的問題の検討にあてられている。第8章 “Conclusion” は本研究の示唆するところ、今後の展望について述べられている。

本論文の学術的意義については、以下の通りである。

第一に、既存の **FrameNet** は本質的に個別言語指向であり、多言語を比較しつつ共通の土台で記述する設計にはなっていない。この問題を克服し、アスペクト属性をきめ細かく捉えるために、より一般化されたイベントのデータ構造としてのスキーマが提案されている。この点は多言語辞書学における重要な貢献である。

第二に、語彙分析の一環として、三つの言語における姿勢動詞の適合場面を共通のフォーマットによって分類し、違いを明示すると同時に、意味上の制約を体系的な形で示した。これは精度の高い記述の試みとして高く評価できる。言語間で基本的な共通性が存在しつつ、拡張例における相違が見られるという事態は、きわめて多くの語彙分野で見られる。さらなる分析のためのモデルケースを提示したという点は大きな意義をもつ。

第三に、分析手法については、フレーム意味論に基づく **FrameNet** と、語の共起関係に基づく分布意味論に着目しつつ、安易な折衷に終わらずそれぞれの方法論をより高度なものとするための掘り下げた議論が行われた。分布意味論において、負例の投入が解析精度の向上につながらなかったという事実から、より狭い空間を分割するための方法が必要であることが明らかとなった。フレーム意味論についても、有効な多言語比較を行うためにはフレームの同一性ではなく、状況内での等価性に注目すべきことが主張される。これらは重要な理論的創見である。

審査においては、幾つかの問題提起と、生産的な議論が交わされた。第一に、本研究ではアナロジーに基づいて単語間の関係を測定するが、複雑な認知過程をとるであろうアナロジーを分類プロセスの出発点にすることの妥当性について掘り下げた議論が行われた。第二に、日本語のテイル形の扱いについて、意味上の主動詞との結合関係をより明確に表示すべきであるとの指摘がなされた。第三に、分布意味論のベクトル表現によって捉えるものはどのような現実か、検定システムとの整合性についてどう考えるべきかという問題提起がなされた。これらはいずれも、今後の研究のための発展的な課題の提示であり、本論文の学術的価値をそこねるものではない。加えて、審査では、本研究で取り上げた三つの言語において、被験者間の変異の度合いが異なるが、それは社会学的要因によるか、それとも個々の言語の類型論的相違が何らかの制約を与えているのかという点についても、活発な議論がかわされた。

以上、本論文は辞書学と計算言語学の最新の知見を取り入れ、動詞の意味解析の課題に挑み、理論・実証両面において多くの貴重な知見を提示した。学術的価値がきわめて高く、この分野における優れた研究成果として高く評価すべきものと判定する。したがって、本審査委員会は博士（学術）の学位を授与するにふさわしいものと認定する。