研究報告書

申請者：宮下博幸（関西学院大学）
助成対象者：Julian Michael Stawecki, PhD（デュッセルドルフ大学）
研究課題「ドイツ語の項構造同定を目的とする解析システム（パーサ）の開発」

現在、ドイツ語に関しては大規模な電子データが存在している。しかしそのデータを言語学的に重要な句の単位に分け、またその句の単位から構成される文の項構造を自動的に同定するようなツールがないのが状況である。生成 AI は有用なツールとなりうるが、この作業に特化したものでないため現段階ではなお実用的ではない。このような背景から、本共同研究はドイツ語の大規模コーパスにおいて名詞句、前置詞句などの句を認定し、さらにそれらと動詞との組み合わせを自動的に表示することを可能とする解析システム（パーサ）の開発することを目的とした。パーサの開発によりこのような解析の自動化が可能になれば、ある動詞がどのような句要素とともに出現するのかを大規模コーパスのデータをもとに実証的に示すことが可能となり、さらには動詞と句要素との共起パターンを頻度として把握できることで、今後の辞書記述にも多大な貢献となることが期待される。

以上の目的を達成すべく、助成対象者の Stawecki 氏とともに研究を進めた。その際、必要な解析内容の検討を含む解析システムの立案は主に申請者が行い、Stawecki 氏はその立案に基づいてプログラミングを行う形で進めていった。解析システムは実際に解析データ上で走らせながらさまざまな改善を積み重ねていく必要があったが、日独間のメールならびにオンライン会議による検討を重ね、昨年末までに試行版のパーサが完成した。続けて試行版パーサの精度の検証を行った。検証を行うにあたっては新聞、フィクション、ノンフィクション、学術のそれぞれの比較的短いテクストを用意し、それらのテクストをパーサに分析させ、手作業でその結果を分析した。その結果、新聞では 0.91，ノンフィクションでは 0.92 の精度を得ることができた。一方、フィクション、学術テクストに関してはこれより精度が低く、それぞれ 0.81，0.84 であった。これは新聞、ノンフィクションのテクストに比べてフィクションや学術テクストの複雑性が高いことに起因すると考えられる。4 つのテクストの平均の精度は 0.87 であった。

以上の結果は 2025 年 3 月に長崎の出島メッセで開催された第 31 回言語処理学会（NLP2025，2025 年 3 月 10 日～14 日）で報告した。Stawecki 氏はこの共同の報告のために来日した。滞在中は学会報告で得られたフィードバックについて氏と対面で話し合うことができ、大変有意義な機会となった。氏の帰国後も、引き続きパーサのさらなる改良のための議論を継続している。

今後の課題としては、開発したパーサにとって苦手なテクスト種であることが判明したフィクションや学術テクストに関し、さらに精度を上げるための工夫を行っていくこと、またこれまでは精度検証のために非常に小規模なサンプルテクストを用いていたが、今後例えば平均で 0.95 程度の精度に達した時点で、本来の目的である大規模のデータの処理へと移行し、またそこから冒頭に目的として挙げたドイツ語の項構造の量的分析につなげていくことが求められる。Stawecki 氏とともに、これらの課題を引き続き進展させてきたい。

なお本奨学金の費用明細は以下の通りとなっている。

往復渡航費（フランクフルト～日本）	20 万円
日本滞在費（宿泊費、交通費等）	26 万円
言語処理学会会費・参加費	2 万円
研究関連の書籍購入	2 万円

本奨学金を与えていただいたことに改めて感謝申し上げたい。

2025 年 11 月 4 日