

## 令和6年度 奈良女子大学研究推進プロジェクト経費研究報告書

奈良女子大学長 殿

研究代表者

所属・職 研究院 工学系 工学領域・教授  
氏 名 吉田哲也

本年度の交付を受けた研究推進プロジェクト経費について、下記のとおり報告いたします。

申請区分	<input type="checkbox"/> (A) 多様な人材の協働による研究
	<input checked="" type="checkbox"/> (B) 一般研究
研究課題	文字列の編集操作に対する組合せ最適化を用いた同根語推定手法の開発

1. プロジェクトの成果（研究推進プロジェクト経費が交付された期間に、この研究によって得られた新たな知見等の成果を、申請書の「研究目的、研究計画・方法」に対応させて記載すること。また、支出内訳との関係についても記載すること。）  
2 ページ以内で記載してください。

同根語とは、イタリア語の **errore** とポルトガル語の **erro** のように共通の起源と意味を持つ異なる言語の単語対である。しかし、言語に対する豊富な領域知識を持つ専門家にとっても同根語の推定は非常に困難な問題であり、膨大な手作業と数十年の年月を要することも多い。これまで歴史言語学などの分野で研究されてきたが、近年の計算資源の発展に伴い、情報技術を積極的に活用するアプローチが求められている。たとえば、規則性の高い暗号の解読で用いられる置換暗号に基づく手法や、ニューラル機械翻訳を適用する手法なども提案されてきた。しかし、現代語と異なり一般に対訳コーパスを準備することが難しく、学習に必要なデータ量が少ないという課題がある。

本研究では、言語に対する領域知識を必要としない汎用的な同根語推定の実現に向けて、文字列の編集操作に対する組合せ最適化を用いた同根語推定の実現に取り組んだ。具体的には、下記の3項目に取り組んだ。

- (1) 文字列の編集操作に基づく同根語推定の定式化
- (2) 組合せ最適化アルゴリズムの開発
- (3) システム実装と実データでの評価・検証

同根語とは同じ言語族の言語間での単語の対であるため、それぞれの単語を文字列とみなし、文字列に対する編集距離に基づいて同根語の推定問題を定式化した。その際、編集距離に対する文字の対応の影響を、編集操作の頻度を要素とする行列として表現し、この行列を用いて、文字の対応づけを単語の対応のもとで編集距離の和を最小化する制約付き最適化問題として定式化した。この定式化のもとで、編集操作に基づく行列を用いて線形計画法を繰り返し適用して同根語を推定するアルゴリズムを開発した。さらに、開発したアルゴリズムを計算機上に実装し、既往研究との比較を行った。アルゴリズムの開発や計算機実装をするための計算機とその周辺機器の購入、および、研究成果を発表するための旅費や参加費などにプロジェクト経費を支出した。なお、アルゴリズム開発用計算機は他の予算とあわせて購入し、購入した計算機に必要な周辺機器を消耗費としてプロジェクト経費で購入するとともに、発表旅費や参加費なども本プロジェクト経費から支出した。

## 2. プロジェクト成果の発表

(論文名, 発表者, 発表雑誌等, 巻・号, 発表年等)

線形計画法の重み付き適用に基づく一対多同根語推定の実現, 梶 春華, 吉田 哲也, 人文科学とコンピュータシンポジウム (じんもんこん2024), A-1-1, 2024

3. 支 出 内 訳						
備 品 費		消耗品費	旅 費	謝金等	その他	計
品 名	金 額					
アルゴリズム 開発用計算機	47,600 円	782,040円	67,060 円	0 円	3,300 円	900,000 円