

京都大学	博士（文学）	氏名	SEO MINCHEOL
論文題目	言語研究のための言語資源の収集と活用に関する研究		
<p>（論文内容の要旨）</p> <p>本論文は、言語研究をするために必要な言語資源の収集方法と活用方法を取り扱うものである。本論文で提起される技法を利用して、事例研究として日本語と韓国語の分析をも行い、論者が開発した方法が実際の言語研究にどのように適用できるのかを示すとともに、日韓それぞれの述語反復構文に関する記述を前進させ、対照言語学的な考察をも加えている。</p> <p>第I部では、主に検索エンジンと検索演算子を利用して、研究に必要な用例をどのように収集すればいいのか、また、その研究方法を用いるとどのような研究が可能になるのかについて述べられる。まず、第2章では、検索エンジンGoogleと検索演算子を利用して、言語研究に必要な用例を収集する方法が紹介される。とりわけ、Googleでは「site・inurl・AROUND」演算子を組み合わせることができるので、研究者が用例を大量に、そして、効率よく検索し、収集することを可能にしている。そのため、他の検索エンジンに比べ、Googleは破格的な構文や周辺的な構文の分析に有用であると言える。</p> <p>第3章では、第2章で示した検索方法を利用し、事例研究を行っている。事例研究では、Googleと検索演算子を利用し、お好み焼きを食べるときに用いる「コテ」や「ヘラ」という語形が地域・年齢・性別によって、どのように用いられているのかを調査・分析している。その結果、「コテ」と「ヘラ」の使用には、地域差だけでなく、年齢や性別による差もあることを明らかにしている。</p> <p>第II部では、日本語の述語反復構文「PことはP」を分析するために、どのような言語資源を収集して分析に活用できるのか、また、既存のコーパスをどのように活用できるのかについて論じている。最初に第4章では、OTTサービスの一つであるNetflixと、ウェブブラウザの拡張機能であるLanguage Reactorを利用して日韓のドラマや映画の（翻訳）字幕を収集し、日韓・韓日パラレルコーパスを構築する方法について提案する。ここで、OTTサービスの字幕における「重複」「修正可能性」「重訳」「複数翻訳者」などの問題点や、方言・外国語の翻訳字幕を使用する際の注意点などについても論じられている。</p> <p>第5章では、第4章で構築した日韓・韓日パラレルコーパスを利用して、日本語の述語反復構文「PことはP」と、それとの対応関係が見られている韓国語の述語反復構文「P-kinun P」「P-kinun ha-」との対照研究を試みる。韓日パラレルコーパスからは、かなり多くの用例を収集することができる。それに対し、日韓パラレルコーパス</p>			

を対象として「PことはP」を検索した結果は、3例のみであった。韓日パラレルコーパスを利用した用例収集を経て、韓国語の述語反復構文の日本語訳を分析した結果では、「P-kinun {P/ha-}」と「PことはP」が一对一の対応関係にあるのではなく、一対多の対応関係を成している可能性があることが明らかにしている。

第6章では、第5章で利用した日韓パラレルコーパスにおいて「PことはP」の用例が非常に少なかった点を踏まえ、日本のドラマの字幕（台詞）を大量に収集し、『ドラマコーパス』を構築している。『ドラマコーパス』を対象として「PことはP」の用例を収集し、分析した結果では、次の点が明らかにしている。まず、「PことはP」の名詞化辞としては「こと」以外にも、「の」や「 \emptyset 」が用いられており、とりわけ、「 \emptyset 」の使用が最も多い。そして、「PことはP」に続く逆接の表現としては「が」よりも「けど」の使用が多い。また、「PことはP」における二つの述語Pは、同型で現れる場合が多いことも明らかにしている。

第7章では、実際の日常会話を収録した『CEJC』を利用し、「PことはP」の用例を検索・収集する方法について説明し、実際に分析を行った。その結果には、『ドラマコーパス』から収集した「PことはP」を分析した結果と、多くの共通点が見られた。どちらのコーパスにおいても、「PことはP」の名詞化辞としては「 \emptyset 」が用いられる場合が多く、逆接の表現の使用も「が」よりは「けど」の方が多い。そして、「PことはP」の先行述語と後行述語が同じ時制形式で出現している例が多くを占めており、「PことはP」に敬体が使われる際には「ます」ではなく、「です」の使用が著しい。

第8章では、第II部において、複数の言語資源を利用して「PことはP」を分析した結果を踏まえ、「PことはP」を「完全反復構文」「部分反復構文」「代用反復構文」に分けることを提案している。そのような分類を行うことが可能な理由として、「PことはP」と、「Vはする」「ADJくはある」「{N/ADJN}ではある」が意味的に類似している点、また、これらの構文の述語の階層構造における振る舞いが類似している点を示している。そして、「PことはP」における後行述語が先行述語をコントロールしているため、後行述語の階層構造よりも先行述語の階層構造を複雑にすることができないことも確認している。最後に、述語反復構文に見られるいくつかの特徴をまとめて一般化を行い、それが日本語だけでなく、韓国語の述語反復構文にも当てはまることについて論じている。

検索エンジンは検索機能を向上させるために改善を続けている。その改善が、言語研究をする研究者に有用な改善となるのかどうかについては、引き続き目を向け、多様な検索方法を試み、その結果をまとめていく必要があると考えられる。そして、第2章で取り上げた検索演算子と、その他の検索演算子を組み合わせることによって、どのような検索が可能になるのか、その検索方法が言語研究をする際、どのように活用可能なのかについても、今後もさらなる研究が続けられる必要がある。

本論文では主に日本語に見られる言語現象を取り上げて分析を行っている。しかし、第2章や第4章を通して示した言語資源の利活用の方法は、他の言語においても利用可能な研究方法である。さらに、第4章や第6章において構築した日韓・韓日パラレルコーパスや『ドラマコーパス』を利用すると、述語反復構文以外の言語現象の用例を収集・分析することも可能であり、本研究は今後、拡張可能性が高い研究だと言えることができる。

(論文審査の結果の要旨)

本論文は、言語研究をするために必要な言語資源の収集方法と活用方法を取り扱うものである。論者は自身が提案する技法を利用して、事例研究として日本語と韓国語の分析をも行う。論者の技法が実際の言語研究にどのように適用できるのかを示す中で、日韓それぞれの言語記述を前進させるとともに、対照言語学的な考察をも加えている。実際に言語研究を進めることで調査技法の有効性を示す試みであると言える。

第I部では、主に検索エンジンと検索演算子を利用して、研究に必要な用例をどのように収集すればいいのか、また、その研究方法を用いるとどのような研究が可能になるのかについて述べている。第1章は序論であり、既存の言語資源の限界について論じられるとともに、その克服のための二つのアイデアが示される。一つはウェブを検索可能な大規模言語資料、すなわちコーパスとみなす観点であり、もう一つは話し言葉のコーパスを研究者自身が新たに構築する方策である。第2章では、検索エンジン Google と検索演算子を利用して、言語研究に必要な用例を収集する方法が論じられる。Google では既存の研究では取り上げられていない演算子の組み合わせも利用することができ、用例を大量に効率よく、多様な観点から検索し、収集することを可能にする。ウェブ検索は破格的・周辺の構文や新たに生まれる表現分析に有用であるだけではない。論者の手法によれば方言形式を対象とする検索も可能になる。第3章では、第2章で示した検索方法を利用した、事例研究が行われる。具体的には Google と検索演算子を利用し、特定の同義語彙がどう用いられているのかを調査・分析し、その使用に、地域差だけでなく、年齢や性別による差もあることを明らかにしている。

第II部では、日本語の述語反復構文を分析するために、どのような言語資源を収集して分析に活用できるのか、また、既存のコーパスをどのように活用できるのかについて論じている。第4章では、日韓のドラマや映画の原語字幕と翻訳字幕を収集し、日韓・韓日パラレルコーパスを構築する方法について論じる。これらの字幕に見られる問題点や、方言・外国語の翻訳字幕を使用する際の諸問題にも注意が払われている。第5章では、第4章で構築した日韓・韓日パラレルコーパスを「準口語」すなわち「あらかじめ用意された台本に基づいて発せられた発話」のコーパスと位置付け、これを利用して、日本語の述語反復構文、および形式的・意味的な対応関係を示す韓国語の述語反復構文との対照研究が試みられる。韓日パラレルコーパスからの用例を分析した結果では、韓日の述語反復構文が一对一の対応関係にあるのではなく、一对多の対応関係を成している可能性があることが明らかになった。第6章では、第5章で利用した日韓パラレルコーパスの問題を踏まえ、日本のドラマの字幕(台詞)を大量に収集し、もう一つのコーパスを構築した。このコーパスを対象として述語反復構文の用例を収集・分析し、既存の記述で指摘されていない複数の言語特徴を明らかにしている。第7章では、論者の構築した準口語コーパスを評価する観点から、日本語

の日常会話を収録した口語コーパスを利用し、述語反復構文の用例を検索・収集して分析を行っている。その結果を論者の構築したコーパスによる分析結果と比べると、相違点もあるが、多くの共通点が確認できる。言語研究のための言語資源として準口語がどのような貢献をすることが可能なのか、一定程度示すことに成功している。

第8章では、第II部において、複数の言語資源を利用して述語反復構文を分析した結果を踏まえ、日本語の当該の構文を「完全反復構文」「部分反復構文」「代用反復構文」に分けることを提案している。意味的な類似のほか、これらの構文の述語の階層構造における振る舞いの類似を示し、述語反復構文における後行述語が先行述語をコントロールしているため、後行述語の階層構造よりも先行述語の階層構造を複雑にすることができないと分析している。最後に論者は、述語反復構文に見られる特徴について一般化を行い、その一般化が日本語だけでなく、韓国語の述語反復構文にも当てはまることについて論じている。

本論文で示された、ウェブ上の言語資源を Google 検索によって利用する方法は、安易な検索とヒット数確認とは異なるものである。論者は極めて慎重な調査を重ねつつ、社会言語学的な分析が可能な水準の情報を見事に引き出してゆく。これまで見逃されてきたウェブの活用法を案出し、ウェブをコーパスとみなす観点が有用であることを示した意義は大きい。しかし、ウェブ検索による言語研究の限界を示していない点は惜まれる部分である。例えば第I部で示された検索方法は、第II部では活用されていないが、このことが資料体の性格の違いに帰せられるべきものなのか、検索技法に関わる問題なのか、必ずしも明らかにされていない。とはいえ、第I部で、ウェブ上の情報から話者の属性や発話に関連づけられる場所を特定する発想は斬新であり、ウェブ検索による言語研究のさらなる可能性を提示した点は高く評価できる。

第II部で事例として扱われる構文は、論者も認める通り、周辺的な構文であり、先行研究の記述の抜けを見つけ出した成果と日韓対照分析まで行った成果が認められるが、言語研究自体のスケールはさほど大きくない。また、口語と準口語の違いについては興味深い示唆を得ながら十分な議論が尽くされたとは言い難い。しかしながら、自身の構築した二つのコーパスを含む大規模資料体同士の比較のために、より規模の小さい資料体では見出されにくい構文を事例に選ぶことは適切でもあり、何より本論文の主眼は、調査技法の新たな可能性を提示することにある。その議論の副産物として記述的な成果を得、資料体の性格を比較する中で言語研究の新たな課題を示したものと見れば、これら諸点は本論文の価値を高めこそすれ低めるものではない。構築されたコーパスは無論のことながら、論者の技法は全体として、応用可能性が高い。

以上、審査したところにより、本論文は博士（文学）の学位論文として価値あるものと認められる。令和5年2月15日、調査委員四名が論文内容とそれに関連した事柄について口頭試問を行った結果、全員一致で合格と認めた。