

所属・資格 国文学科・教授

申請者氏名 荻野 綱男

研究課題		ウェブをコーパスとして活用する日本語研究
報告の概要	研究目的 および 研究概要	<p>ウェブは世界最大のコーパスである。さまざまな書かれた文書が記録されている。それを日本語のコーパスとして活用する方法をさまざま試してみることを研究目的とする。</p> <p>たとえば、特定の表現がウェブでどのように使われているかを検索すれば、たくさんの用例を瞬時に入手することができる。それを用いて研究を進めることで、膨大な量の日本語コーパスを使うことが容易に行える。</p> <p>本研究では、そのようなことを念頭に、具体的な日本語表現をネット検索にかけ、どんなふうに使われているのかを把握したい。</p> <p>用例を整理するために、人手によるチェック・分類が必須である。</p>
	研究の結果	<p>平成30年度は、いくつかのテーマについてウェブを検索して研究を進めた。その一つとして、以下では日本語の語順の結果を示そう。</p> <p>「一郎が台所で酒を飲む」のように、動詞の前に三つの名詞句（名詞＋助詞）がつながる場合、それぞれの名詞句が現れる順序はどれが多いか。これを明らかにするために、「国会会議録」から「飲む」の用例を検索し、整理してみると、「～が～で～を」が16例で最も多く、次が「～で～が～を」14例となる。それ以外の語順は2例以下の少数だった。この結果から、日本語の基本語順の傾向がわかった。</p>
	研究の考察・反省	<p>当初予想していたよりも、得られた用例数が少なかった。国会会議録以外の資料ではもっと少なくなってしまう。特に WWW の検索エンジンでは、三つの名詞句が揃って現れる例が極端に少なく、手間をかけるだけムダになった。これは、日本語の場合、いわなくてもわかるものは省略してしまう傾向があることが影響していると思われる。</p> <p>国会会議録のような（純粹の話し言葉でない）読み言葉が多用されるような場合にある程度の用例数が確保できることがわかった。</p> <p>今後は、「飲む」以外の動詞についても同様の調査を行い、どんな動詞にはどんな基本語順があるのかを調べて行きたい。</p>
研究発表 学会名 発表テーマ 年月日/場所	研究発表	<p>2018.9.6「日本語研究に大規模ウェブテキストデータを扱うためには？【初級編】ウェブの検索結果を利用する」コーパスとしてのウェブテキスト活用シンポジウム（国立国語研究所）</p> <p>2018.9.29 「特別講義『データで学ぶ日本語学入門』……私ならこう読む」第62回計量国語学会大会</p> <p>2018.10.13 「第三者敬語の待遇レベルの数量化（完成版）」日本語学会 2018年度秋季大会予稿集,pp.139-146</p>
研究成果物 テーマ 誌名 巻・号 発行年月日 発行所・者	研究成果物	<p>2018.10.1 「言語研究のデータと調査方法論」『日中言語教育と日本語教育』第11号, pp.1-21</p> <p>2018.10.20 「計量言語学」日本語学大辞典、東京堂 pp.293-295</p> <p>2018.12.20 「解説・多人数質問調査法の現在（5）回答の質の問題」計量国語学 Vol.31 No.7, pp.507-516</p> <p>2019.2.28（荻野綱男・保阪靖人・板倉歌・塚本聡・田中ゆかり・韓東力）「コーパス言語学の学際的研究」日本大学文理学部人文科学研究所研究紀要第97号 pp.257-305</p> <p>2019.3.25（伝康晴・荻野綱男）（編）『講座 日本語コーパス 7 コーパスと辞書』朝倉書店, 211p.</p>