2019（令和元）年 10 月 23 日
「統語・意味解析コーパス NPCMJ を用いた言語研究: 二重主語構文と並列構文の事例」
発表者：大久保 弥（東京外国語大学大学院博士後期課程）

本発表では、国立国語研究所の共同研究プロジェクト「統語・意味解析コーパスの開発と言語研究」で開発されている NPCMJ (NINJAL Parsed Corpus of Modern Japanese)の概要と、発表者による NPCMJ を用いた過去の研究事例の紹介をした。

NPCMJ は、ペン通時コーパスを基にしたアノテーション方針で統語解析された現代日本語コーパスである。特徴として、例文はツリー構造で解析され、形態的情報だけではなく、統語的な文法情報がアノテートされているため、テキスト情報に頼らない検索が可能である。そのため、日本語の文法記述で認められてきた言語学的に重要な概念を基にコーパス調査を行うことができる。例えば、述語では動詞、イ形容詞、ナ形容詞、名詞述語をはじめ、態や法に関わる助動詞の区別、節では主節、疑問節、補部節などの区別、さらに、音形を持たない関係化の痕跡やゼロ代名詞の情報も付与されており、これらの情報を例文検索に用いることができる。NPCMJ はプロジェクトウェブサイト(http://npcmj.ninjal.ac.jp/)で公開されている NPCMJ Explorer と NPCMJ Search からウェブブラウザ上でも利用可能であり、発表では、実際にそれらを用いたいくつかの検索例を紹介した。

最後に、NPCMJ を利用した研究事例として、発表者の過去の研究発表を簡潔に紹介した。二重主語構文における項の名詞句の意味的性質の違いの傾向を検証した研究では、テキスト情報を頼りにした検索では収集することが困難であるような構文もコーパス研究の対象にできること、構文中の特定の統語的位置（つまり、それぞれの名詞句）に現れる語彙項目の抽出が可能であることを示した。一方、二つの節が特別な接続形式なしに一文中に並ぶ並列構文での後置詞（が、は、も）の使用と文の解釈の関係性について検証した研究では、複雑な統語構造を持つ文について、特定の統語的位置に現れる語彙項目（つまり、それぞれの節の主語句の後置詞）を指定した上で、例文が収集可能であることを示した。この研究は、文の並列的な読みと従属的な読みとの曖昧性について、「も」のような特定の助詞の使用によるその読みの影響を指摘するものであった。