

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*

大久保 弥 (東京外国語大学大学院 / 国立国語研究所非常勤研究員)
okubo.wataru.m0@tufs.ac.jp / wataru.okubo@ninjal.ac.jp

Outline

- | | |
|---------------|---------|
| 1. NPCMJ とは？ | 3. 実践 |
| 2. 特徴・何ができるのか | 4. 研究事例 |

1 NPCMJ (NINJAL Parsed Corpus of Modern Japanese)

- 国立国語研究所プロジェクト「統語・意味解析コーパスの開発と言語研究」で開発されているコーパス。
 - ツリーバンクコーパス (ペン通時コーパス (Penn Historical Corpus) に従ったアノテーション方針)
 - 2019年3月現在、約3万文 (50万語)。2022年までに6万文を公開予定。
- プロジェクトの詳細
 - <http://npcmj.ninjal.ac.jp/>

2 特徴・何ができるのか

- 特徴
 - ツリー構造 (理論に中立的) で統語解析されている。
 - テキスト (文字) や形態的情報 (品詞・動詞の活用形など) だけではなく、統語解析情報を基にした階層的な文の検索が可能。
 - 言語学的に重要な概念を基にコーパス調査ができる。
- 統語解析情報
 - 文法関係
 - * 主語、第一目的語、第二目的語、論理的な主語など
 - 述語
 - * 動詞、イ形容詞、ナ形容詞、受動、自発・可能・尊敬、助動詞

*本発表は、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」の研究結果を報告したものである。

- * 軽動詞、モーダル、否定、形式名詞など
- 節
 - * 主節、疑問節、補部節、条件節、関係節など
- 空要素
 - * 関係化における痕跡
 - * ゼロ代名詞 (話し手、聞き手、虚辞など)
- 詳しくは「アノテーションマニュアル」を参照。

3 実践 (ウェブブラウザ)

- ウェブブラウザで利用可能。
- NPCMJ Explorer
 - 益岡・田窪著『基礎日本語文法 改訂版』(くろしお出版) の文法項目に該当する用例を検索できる。
- NPCMJ Search
 - 文字列検索
 - ツリー検索
 - indexed モード (指示対象の可視化)

4 研究事例

- 統語解析情報を利用して、どんな研究ができるのか？
 - 二重主語構文
 - * どんな名詞が使われやすい？
 - * 関係化？
 - 並列構文
 - * 先行節-後続節での後置詞 (が、は、も) の組み合わせと文の解釈にはどんな関係があるか？