

国立国語研究所学術情報リポジトリ

現代日本語統語・意味解析コーパス（NPCMJ）の概要と検索

メタデータ	言語: Japanese 出版者: 公開日: 2021-06-25 キーワード (Ja): キーワード (En): 作成者: パルデシ, プラシャント, 長崎, 郁, 鈴木, 彩香 メールアドレス: 所属:
URL	https://doi.org/10.15084/00003347

NPCMJとは

▶ NINJAL Parsed Corpus of Modern Japanese (NPCMJ)

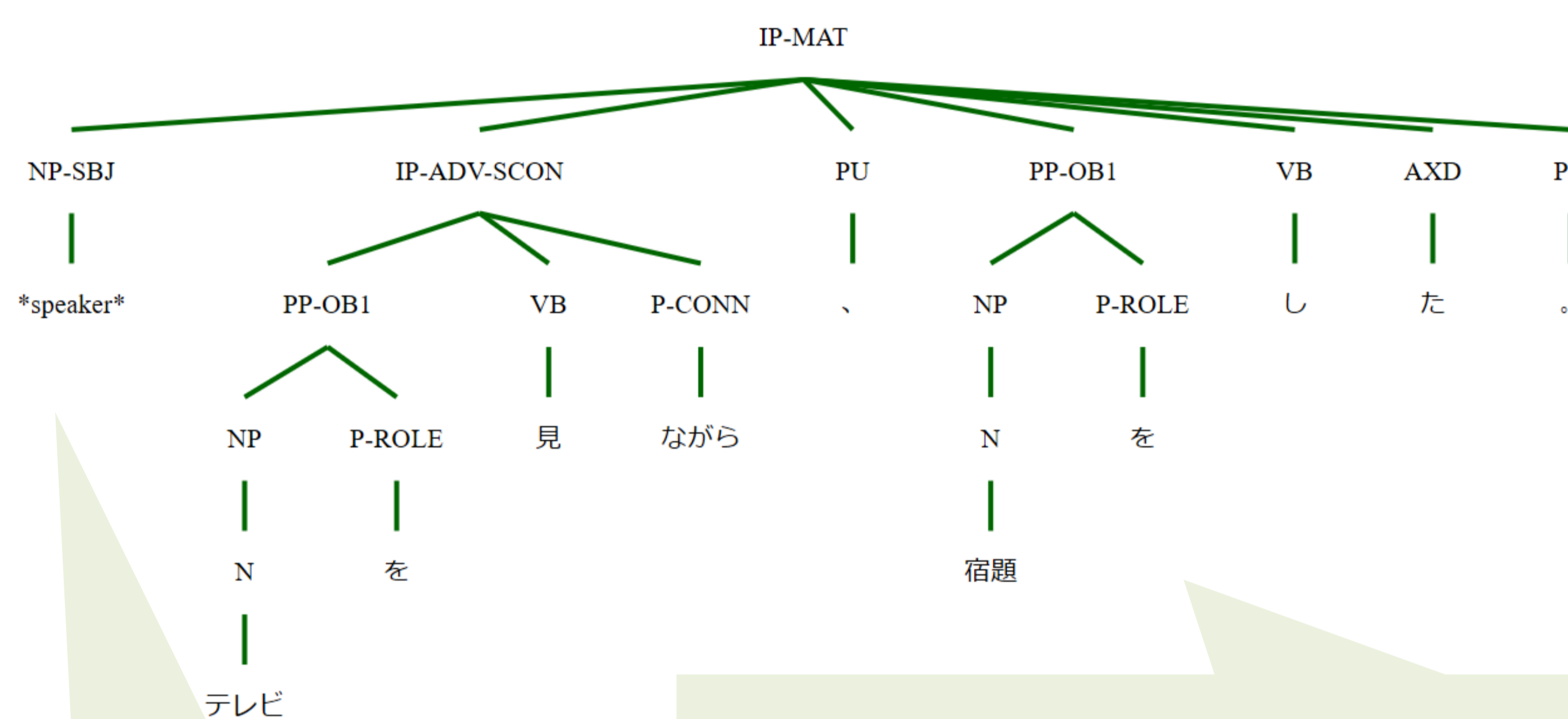
- ▶ **文の構造や意味に関する詳細な情報**が付与された言語資料
 - ▶ 日本語の**様々なタイプの文**を検索し、言語研究に役立てることができる
- ▶ 国立国語研究所で**2016年4月**から開発開始
 - ▶ <http://npcmj.ninjal.ac.jp/>
- ▶ 上記ウェブサイトから**全文データのダウンロード・各検索インターフェース**が利用できる
 - ▶ 2019年3月現在、**3万文**を無償公開中
 - ▶ 毎年1万文追加、プロジェクト終了（2022年3月）までに**6万文**公開予定

データの出典（2019年3月時点）

出典	文数
青空文庫	4,646
聖書	1,664
その他の小説の抜粋	923
ノンフィクション	223
書籍の一部	552
ウィキペディア	2,746
新聞記事	4,666
法律文	337
国会議事録	1,698
テッドトーク	1,453
教科書	6,048
辞書	3,419
その他	2,085
合計	30,460

統語分析

speaker テレビを見ながら、宿題をした。

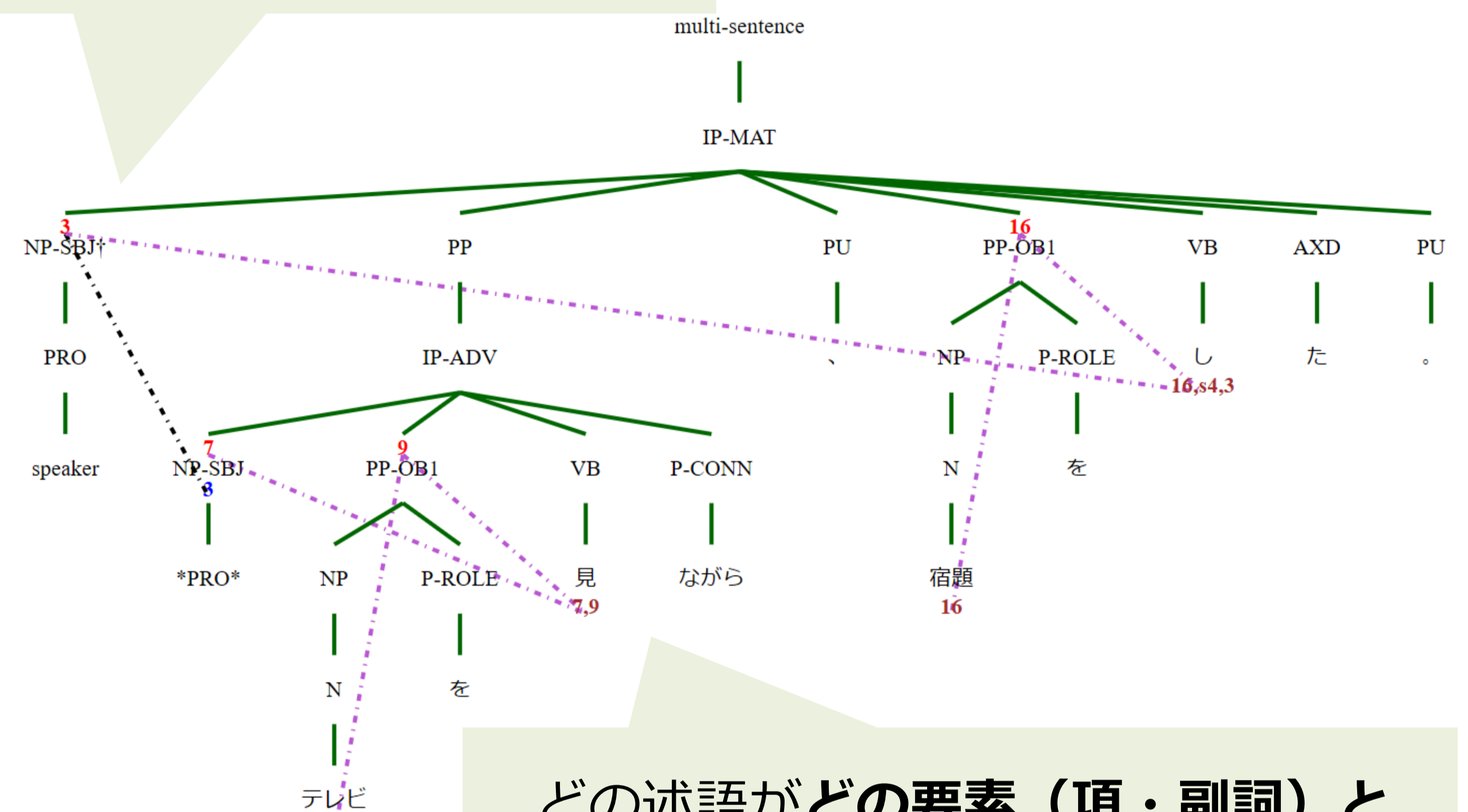


省略された主語や目的語もわかるようになっている

品詞、句、それらの修飾関係や、**主語・目的語**といった句のはたらきがわかる

意味解析

名詞が持つ**指示**がどのように受けつがれるかがわかる



どの述語がどの**要素（項・副詞）**と関係を持つかがわかる

検索インターフェース

初中級者向け

▶ NPCMJ Explorer Pattern Browse

- ▶ **日本語のいくつかの主要な文法項目**に対して、それを調べるための検索式が用意されている
- ▶ 項目名をクリックするだけでコーパス中の用例を見ることができる

用意されている検索式を確認し、**ツリー検索にジャンプ**することもできる

益岡・田窪（1992）の節項目

中上級者向け

▶ NPCMJ Search ツリー検索

- ▶ 自分で検索式を書いて、コーパス中の用例を調べることができる
 - ▶ たとえば…
「の」でマークされた主語

TGrep-lite 検索結果

[PP-SBJ] < ([P-ROLE] < の) >
Submit all data reveal

- それは全く **前例** のないことであつたが、*T* 医学に *T* 深い理解をいたは、これを許した。 [book_excerpt-44_7](#)
- *pro* あなたもよくご存じのように、*T* ほとんど一年じゅう店の外にしのセールスマンは、かげ口や偶然や *T* **いわれ** のない苦情の犠牲になり、そうしたものを防ぐことはまったくできないんです。 [aozora_Harada-1960_374](#)
- 地** のある限り、*speaker* 種まきの時も、刈入れの時も、暑さ寒さも、**昼** も**夜** もやむことはないであろう」。 [bible_old_242](#)

『**主語となる助詞句（PP-SBJ）**が、「**の**」を支配する**格助詞（P-ROLE）**を支配している』

- ▶ **文字列検索**や、**検索式の作成を補助するインターフェース**などもある