

2022.03.04

統語・意味解析コーパス (NPCMJ) チュートリアル

初心者向けインタフェース

鈴木彩香（国立国語研究所）

はじめに

- NPCMJの初心者向けインターフェースである**NPCMJ Explorer**の使い方を紹介する
- NPCMJでどのような文法現象を検索できるのか、イメージをつかんでもらう
 - <http://npcmj.ninjal.ac.jp/explorer/>

NPCMJ Explorerとは？

NPCMJ Explorerの想定するユーザー

- NPCMJを使って、「雨が降った時（に／は）、傘をさした」のような例をとりたい（トキ節の複文のテンスについて調べたい）場合に検索式を書くと...
 - `(NP-TMP < ((/IP/ < AXD) $ (N < とき|時))) $ AXD`
 - `(/PP/ < ((NP < ((/IP/ < AXD) $ (N < とき|時))) $ (/P-/ < は|
に))) $ AXD`

NPCMJ Explorerとは？

NPCMJ Explorerの想定するユーザー

- NPCMJを使ってみたいが、**自分で検索式を書くのはちょっと難しそうだな**、と感じた人にこそ使ってほしい
 - 自分の興味のある文法現象がNPCMJで検索できるのか、気軽に試してみたい
 - NPCMJでどのような文法項目を調べることができるのか、大まかにつかみたい
 - NPCMJのタグ付けがどのように行われているのか、手を動かしながら覚えたい

NPCMJ Explorerとは？

NPCMJ Explorerの特徴

- NPCMJ Explorerでは、**自分で検索式を書くことなく** NPCMJの用例文を見ることができる
- もちろん自身の研究に本格的にNPCMJを使うことになれば、既成の検索式でなくオーダーメイドの検索式が必要になる
 - その場合の検索式の書き方、ツールの使い方についてはこの後詳しく説明があります
 - その意味で、NPCMJ Explorerは「**初心者向け**」

NPCMJ Explorerとは？

NPCMJ Explorerの特徴

- NPCMJ Explorerでは、**自分で検索式を書くことなくNPCMJの用例文を見ることが出来る**
 - しかし、初心者には役に立たないツールではない
 - 中級者、上級者へ**橋渡し**をするツール
 - 文法現象に対応した**検索式の索引**として使えるツール
 - **予備的調査**や**研究のネタ**を探すのに使えるツール

2 種類のツール

- ユーザーは用途に合わせて2つのツールを使い分けることができる
 - **Pattern Browse**
 - **String Search**

2 種類のツール

Pattern Browse

- 用途
 - 参照したい文法現象が決まっている
 - NPCMJで検索することのできる文法現象をある程度網羅的に参照したい
- 具体的な文法現象に応じて、あらかじめ検索式が用意されている
- 益岡隆志・田窪行則（1992）『基礎日本語文法』（くろしお出版）の節項目をもとに、検索項目を選定

2 種類のツール

Pattern Browse

- NPCMJのタグと文法現象の対応
 - NPCMJのタグは、特定の現象に対応していることもある
 - 内の関係／外の関係の連体修飾節
- (1) a. [IP-REL 融資した] お金
b. [IP-EMB 柑橘類がよく育つ] こと

2 種類のツール

Pattern Browse

- NPCMJのタグと文法現象の対応

- しかし、直接対応していないことも多い

- **IP-SMC (小節)**

- (2) a. ぜひ、天田くんにも **[IP-SMC-OB1 出席して]**もらいたい
んだけどもー **(受益)**
- b. 心の問題は3年目に **[IP-SMC-OB1 多く]** なった。
(変化)
- c. 被害額を **[IP-SMC-OB1 実際より低く]** 見積もった。
(目的語繰り上げ)

2 種類のツール

Pattern Browse

- **文法現象ベース**で用意された検索式の一覧があると便利
 - 益岡・田窪 (1992)の**136**の節項目のうち、**69**項目をカバー
 - 「序論」や「基本的性格」などの項目を除き、特定の文法現象を扱っている項目のみ対象
 - 特定の節の中で記述されている内容の一部のみを満たす場合は、項目をさらに細分化

2 種類のツール

Pattern Browse

- **注意点！**

- Pattern Browseで得られる検索結果は、益岡・田窪 (1992)の記述に完全に沿っていることを保証するものではない
- 文法書の記述を完全には満たさない結果を返す項目については注釈をつけている
 - 例：主題の八と対比の八の区別ができない等
- 注釈がついているもの以外にも、意図しない誤りが含まれているケースがある

2 種類のツール

Pattern Browse

- **注意点！**

- また逆に、Pattern Browseでリストアップされていない現象はNPCMJで検索することができない、というわけでもない
- ここでの節項目の立て方は、益岡・田窪 (1992)の記述の方法による制限がある

2 種類のツール

Pattern Browse

- 中・上級者への橋渡しの機能
 - 検索結果や用意されている検索式を参照して、**どのようなタグで現象を検索しているのか**を確認することができる
 - 自分で検索式を作る場合にも、用意されている検索式を参考にすることができる
 - 例：特定の環境に生起する副詞の特徴が見たい等

2 種類のツール

String Search

- 用途
 - 特定の語彙項目を持つ用例を参照したい場合
- **語の区切り**に基づいた検索のほかに、**線状的な配列**に基づく検索ができる
 - 語の区切りがわからない要素であっても検索可能
- NPCMJ Development Interfacesの「文字列 (KWICK) 検索」に相当

インターフェースの使い方

- NPCMJ Explorer トップページ(<http://npcmj.ninjal.ac.jp/explorer/>)から **ツールを開く**
- インターフェースの切り替え
- パターンパネルと用例パネル
 - **Table View** と **Tree View**
 - 検索結果の一覧と構造の表示
 - **Pattern Info**
 - 注釈、および検索式の表示はここ

インターフェースの使い方

- パターンパネルの見方
 - 節項目の表示
 - 矢印のアイコンを押して検索
- ジャンルの変更
- パターンの絞り込み

インターフェースの使い方

- 用例パネルの見方
 - **Table View**の使い方
 - 用例とジャンル
 - ※用例表示にはゼロの要素が含まれることに注意
 - ツリーのアイコン (**Tree View**)
 - 用例数、表示件数、ページ切り替え

インターフェースの使い方

- 検索式や注釈の確認

- **Pattern Info**

- 検索式

- Tregex

- XPath

- ※lemmaや品詞、活用形などの情報を参照して検索しているものはXPath式のみ

- 注釈

インターフェースの使い方

- String Search

- 検索方式

- Search as String

- 例：「ことにな」

- Search as Word

- 例：格助詞の「と」（一文字からなる語）

期待される使い方

- NPCMJの検索方法に慣れる
 - NPCMJのアノテーション方法を具体的な文法現象に基づいて把握する
 - あらかじめ用意された検索式に手を加えるという形で、検索式の書き方に慣れる
 - 自分で検索式を書く時に楽をする

期待される使い方

- 言語研究につなげる
 - 予備的調査や、研究のネタ探し
 - 共起する要素や環境の偏りはないか？
 - 先行研究が見過ごしていた観点はないか？
 - ジャンルの偏りはないか？
 - 普段、コーパスを使った研究をしていない人だからこそNPCMJを使ってほしい