

# 4. TGrep-lite 検索式 1

統語・意味解析コーパス (NPCMJ) チュートリアル@弘前大学

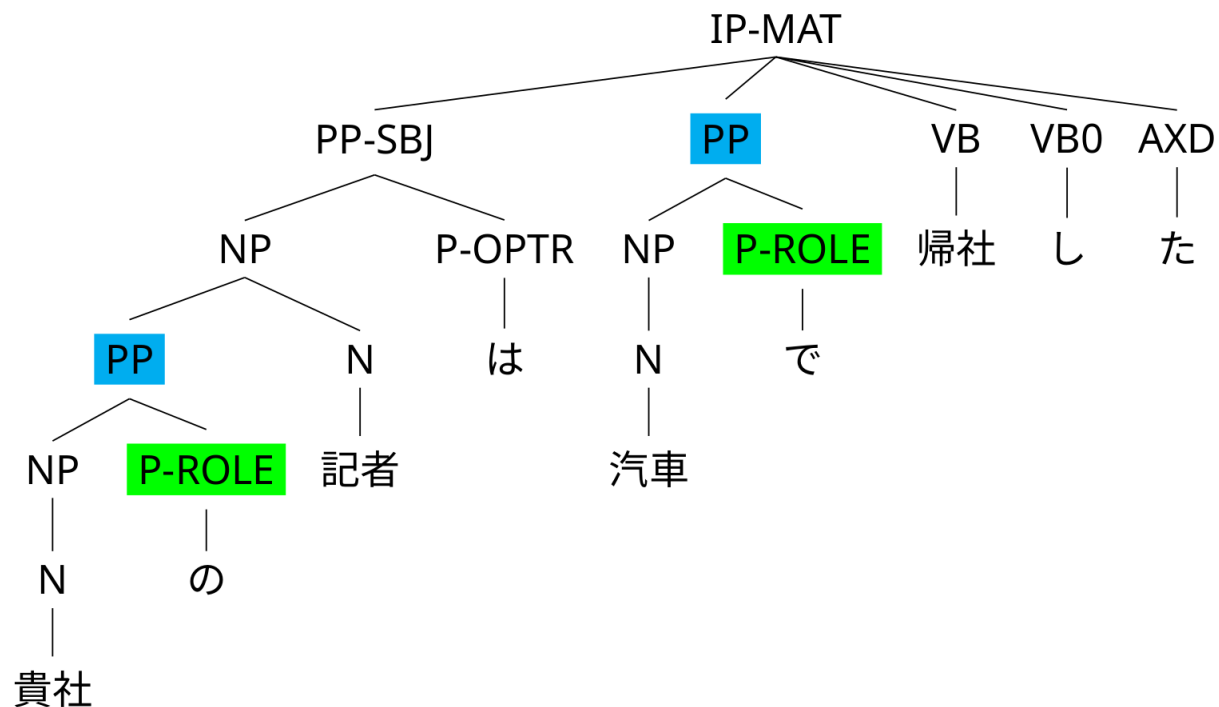
鈴木彩香・長崎郁\*

2019.5.11

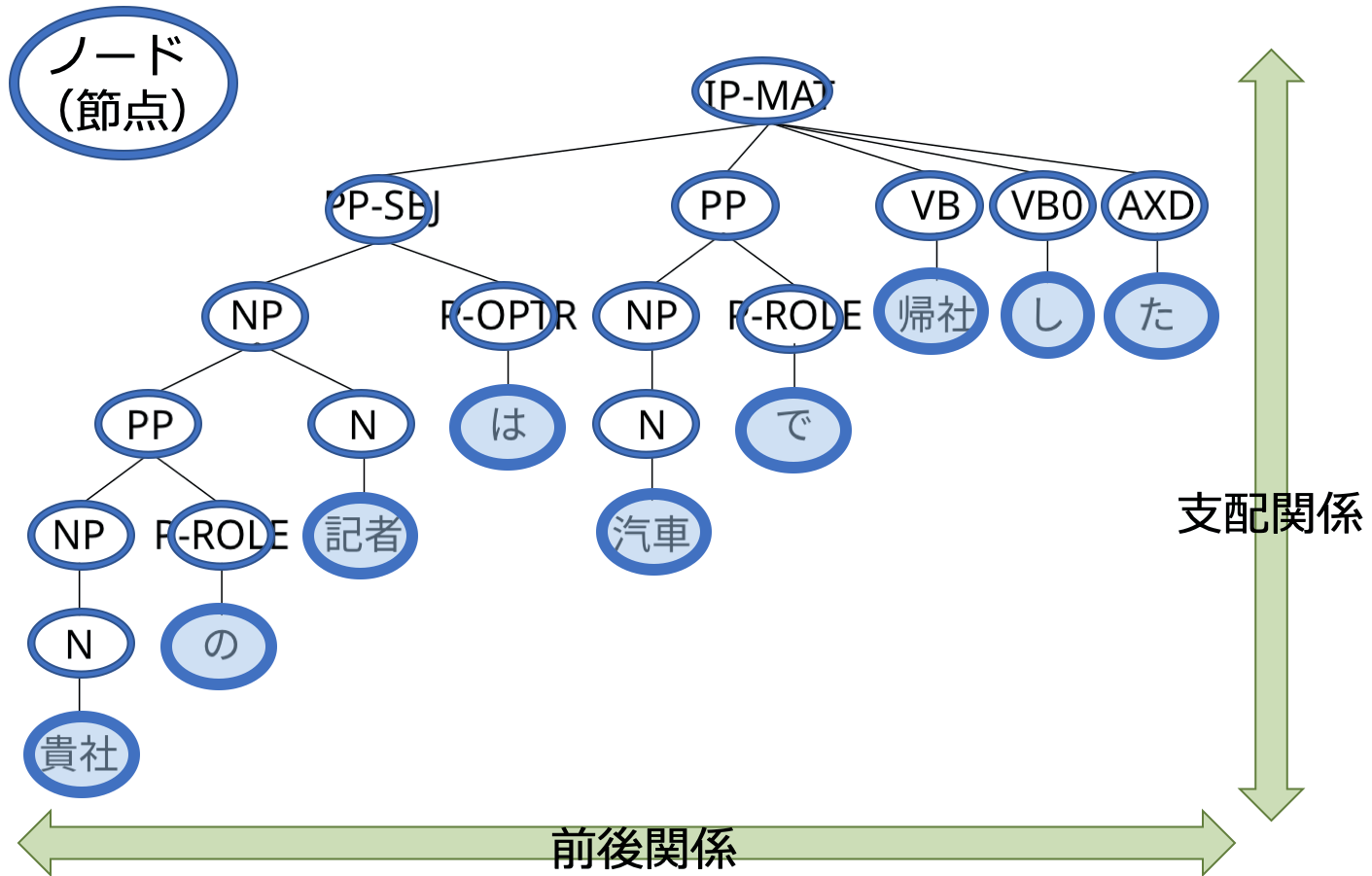
\*本資料は、長崎郁「TGrep-lite 検索式 1」（統語・意味解析コーパス (NPCMJ) チュートリアル@東北大学 2019.01.26）による。

# はじめに

- TGrep-liteとは  
NPCMJ のデータ(Penn 方式のデータ)を、**ノードや、ノードとノードの関係**を指定して検索するためのツール
- 単純な例:  
[<sup>^</sup>PP] < [<sup>^</sup>P-ROLE] (P-ROLE を直接支配する PP)



# TGrep-lite 検索式



## TGrep-lite 検索式

- 検索式は次のように分類することができる
  - ノードを記述しただけのもの
  - 単純な関係表現：ノード記述と、それらの間の関係を1つ記述したもの
  - 複雑な関係表現：ノード記述と、それらの間の関係を2つ以上記述したもの
- この時間は、**ノードの記述**について扱う

## ツリー検索画面へ

- 表示されている画面上部の ツリー検索 をクリック



# ツリー検索画面

The screenshot shows the 'ツリー検索' (Tree Search) interface. At the top, there is a navigation bar with buttons for 'コーパス概要', 'タグ', '語の依存関係', '文字列検索', and 'ツリー検索'. Below this, there are four informational buttons: 'ツリー検索について' (circled 4), 'TGrep-lite について', 'XPath について', and 'テキストを解析してクエリを作成'. The main search area is a large text input box (circled 1) with a 'Submit' button (circled 3) and a dropdown menu for source types (circled 2) currently set to 'all data'.

- 1 – 検索式入力ボックス
- 2 – ソースの種類を選択（デフォルトは **all data**）
- 3 – 実行ボタン
- 4 – 各項目についての解説を表示
  - 検索式入力ボックスに何らかの文字列（例えば「食べ」）を入力し、検索を実行してみましょう

# 検索結果の表示

## TGrep-lite 検索結果

/食べる/

Submit all data  reveal

1. 日本人も欧米人並みに肉や乳製品を **食べる** ようになった。 [textbook\\_djg\\_advanced\\_28](#) **5**
2. もう **食べる** 時間 \* ? [textbook\\_purple\\_intermediate\\_183](#)
3. / **食べる** [news\\_KAHOKU\\_457\\_2](#)
4. 彼はよく **食べる** なあ。 [textbook\\_djg\\_intermediate\\_188](#)
5. 私はピザを **食べる** のが好きです。 [textbook\\_purple\\_basic\\_23](#)
6. 父はいつも新聞を読みながら朝食を **食べる**。 [textbook\\_kisonihongo\\_1098](#)
7. \*speaker\* そんな \*T\* まずいものを **食べる** くらいなら、\*speaker\* 何も食べないほうがましだ。 [textbook\\_particles\\_401](#)
8. 毎日ピザを **食べる** のは体に悪いです。 [textbook\\_purple\\_basic\\_24](#)
9. \*pro\* インドでは、左手で **食べる** ことはない。 [textbook\\_purple\\_intermediate\\_570](#)
10. しかし、\*T\* \*T\* 天から下ってきたパンを **食べる** 人は、決して \*pro\* 死ぬことはない。 [bible\\_new\\_477](#)
11. 京都の和菓子は、\*T\* \*arb\* 宮中や公家、寺社、茶家におさめたり、\*T\* 特別なお祝いのためにあつらえる「上菓子」、\*T\* \*arb\* ふだんに **食べる** 「おまん（饅頭の略）」や「だんご」「餅菓子」 [wikipedia\\_KYOTO\\_16\\_27](#)
12. \*pro\* \*arb\* \*pro\* 茶道に於ける薄茶（うすちゃ、\*pro\* \*pro\* お薄（おうす）とも）や濃茶（こいちゃ）とともに **食べる** こともあり、\*T\* 味覚は元より\*美的鑑賞にも堪えることを期待されて [wikipedia\\_KYOTO\\_16\\_6](#)
13. \*speaker\* \*pro\* もうちょっと **食べる**。 [textbook\\_purple\\_basic\\_248](#)
14. \*T\* それを **食べる** 者は、いつまでも生きるであろう。 [bible\\_new\\_479](#)
15. \*T\* \*pro\* **食べる** 物もなければ、\*T\* \*pro\* 住む所もない。 [textbook\\_particles\\_12](#)
16. クマは、植物の根を **食べる** ために土を掘り返す。 [dict\\_vv-lexicon\\_2907](#)
17. \*speaker\* あと二つ **食べる**。 [textbook\\_purple\\_basic\\_249](#)
18. 例えば僕と妹は子供の頃のリスの脳みそを **食べる** 競争をしていました [ted\\_talk\\_10\\_8](#)
19. \*speaker\* クッキーをもう二つ **食べる**。 [textbook\\_purple\\_basic\\_247](#)
20. \*arb\* 刺身は生で **食べる**。 [textbook\\_kisonihongo\\_777](#)
21. 私も **食べる** けど。 [textbook\\_purple\\_basic\\_273](#)
22. \*speaker\* \*pro\* もっとたくさん **食べる**。 [textbook\\_purple\\_basic\\_245](#)
23. \*speaker\* 食事を作り終わって、もう **食べる** ばかりのところ、電話がかかってきました。 [textbook\\_particles\\_33](#)
24. 芋煮会は、\*pro\* 河原などの野外において、家族や友人、学校や職場の仲間が集まって、大鍋に汁をつくって **食べる** 季節行事である。 [wikipedia\\_Sendai\\_City\\_277](#)
25. デザートは \*pro\* \*pro\* **食べる** ばかりになっています。 [textbook\\_djg\\_basic\\_13](#)

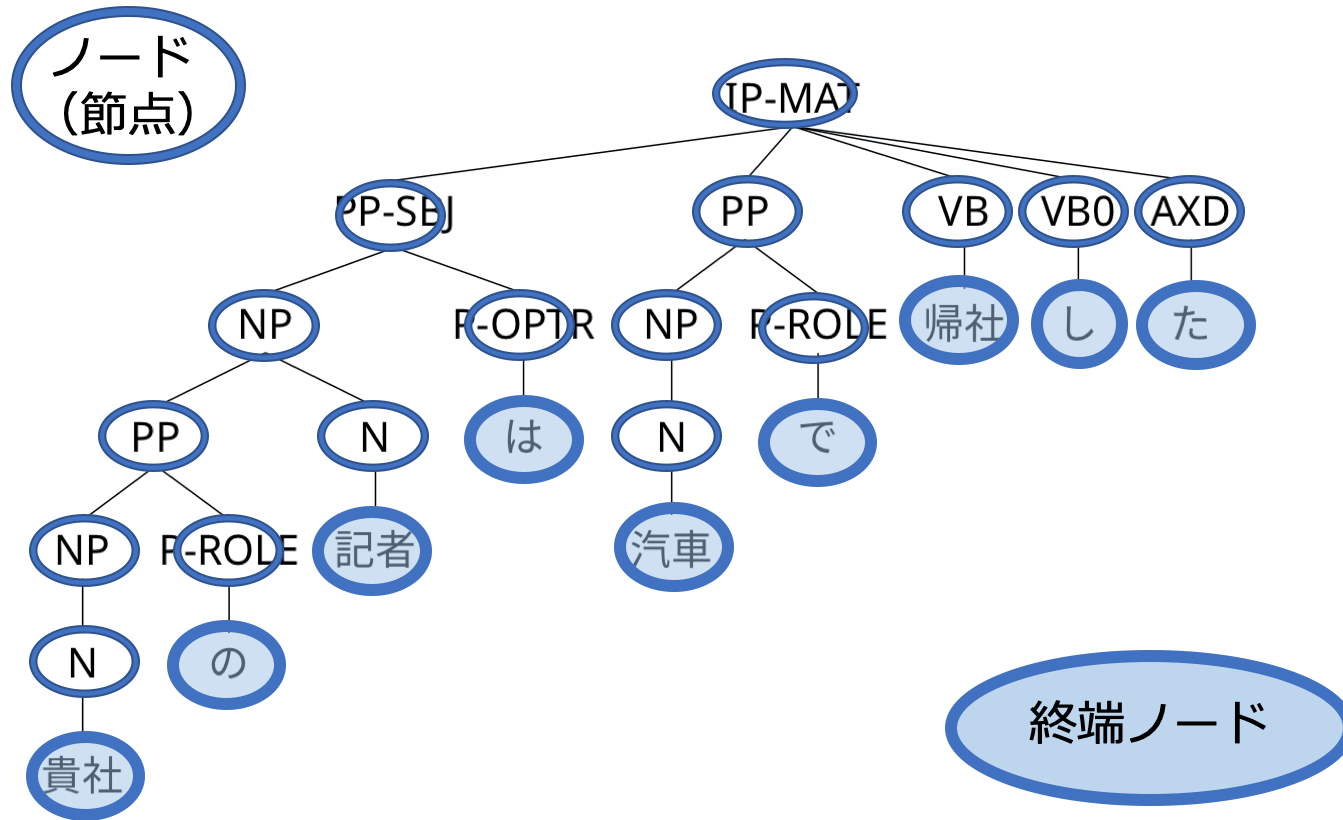
↓ **2**

Goto Page 1 of 3 (currently showing results 1 to 25 of 56) **1**

Download all results comma-separated values **4**

- 1 – マッチ数（25 件表示）
- 2 – 次の 25 件を表示
- 3 – 検索結果をダウンロード
- 4 – ダウンロードファイルの形式を選択
- 5 – クリックでツリーを表示

# ノードの記述 (終端ノードと非終端ノード)





# ノードの記述（終端ノードと非終端ノード）

- 終端ノード

- 単なる文字列を入力すると、  
食べ ⇒ 「食べ」に**完全一致**  
(「食べる」「食べ慣れる」「食べ物」などは排除される)
- 文字列を /.../ (半角スラッシュ) で囲むと**正規表現**となる  
/食べ/ ⇒ 「食べ」に**部分一致**  
(「食べ」だけでなく、「食べる」「食べ慣れる」「食べ物」などにもマッチ)

- 非終端ノード

- **必ず [...]** (半角角括弧) で囲む 常に**正規表現**となる  
[SBJ] ⇒ "SBJ" に**部分一致**  
("PP-SBJ"、"PP-SBJ2"、"NP-SBJ"、"NP-SBJ2"、"NP;\*SBJ\*"  
などにマッチ)

## ノードの記述（正規表現）

- 正規表現とは?  
文字列の集合を**パターンとして表現**するための記法
- 単にスラッシュや角括弧で囲むと「**部分一致**」の意味になる
- **語頭**（ノード先頭）、**語末**（ノード末尾）、**選言**、**繰り返し**などを表現することができる

## ノードの記述（正規表現）

/.../	部分一致（終端ノード）	/食べ/ 「食べ」を含む
[...]	部分一致（非終端ノード）	[SBJ] “SBJ”を含む
^	ノード先頭	/^食/ 「食」で始まる [^PP] “PP”で始まる
\$	ノード末尾	/する\$/ 「する」で終わる [ADV\$] “ADV”で終わる /^する\$/ 「する」に完全一致
\b	単語の区切り	[^NP\b] “NP”、“NP-SBJ”、 “NP;{person}”などに一致 (ただし“NPR”(固有名詞) には一致しない)

## 練習問題 1

- 次の検索表現はそれぞれどう違うか?

(1) と

(2) /と/

(3) /^と/

(4) /と\$/

(5) /^と\$/

まず予想を述べ、次にツリー検索画面で確かめなさい。

## ノードの記述（正規表現）

	選言 (A または B)	/好き 嫌い/ 「好き」または「嫌い」を含む [SBJ OB1] “SBJ” または “OB1” を含む
(...)	グループ化	/^(好き 嫌い)\$/ 「好き」または「嫌い」に完全一致 [^(WPRO WADV WD)\$] “WPRO” (疑問代名詞)、 “WADV”(疑問副詞)、“WD” (疑問限定詞)に完全一致
—	(アンダースコア 2 つ) ワイルドカード	— すべてのノード ワイルドカードはスラッシュや角括弧で囲まずに使うことに注意

## 練習問題 2

- 動詞「思う」が使われている用例を検索するためのノードの記述の方法を考えなさい。

動詞の活用形の扱いについては、「データの概要とタグの検索」の補足資料、第 2 節を参照

# ノードの記述（正規表現）

.	(ピリオド)任意の 1 文字	<code>/^..\$/</code> 二文字の終端ノード <code>/^あ.\$/</code> 「あ」で始まる 二文字の終端ノード
*	(アスタリスク)直前の文字 の 0 回以上の繰り返し	<code>/^あ.*/</code> 「あ」で始まる (" <code>/^あ/"</code> と同じ)
<code>\1, \2, ...</code> <code>\9</code>	検索式の中の1~9番目の (...) の中身にマッチ	<code>/^(..)\1/</code> 始まりの二文字が もう一度繰り返される 終端ノード
<code>\</code>	(逆スラッシュ)直後の文字 を特殊記号ではなく通常 の文字として扱う	<code>/^\*/</code> *で始まるもの (空要素にマッチ)

## 練習問題 3

- 次の検索表現は何を表すか?

/^ビ.\*ル\$/

まず予想を述べ、次にツリー検索画面で確かめなさい。



# 非終端ノードの記述に関する補足

- 非終端ノードの記述では、“\$”を使うか、“\b”を使うか、何も使わないかでマッチするものが変わるので注意が必要
  - [^NP\$]  
NPに**完全一致**
  - [^NP]  
NPで**始まるすべてのノード**にマッチ  
("NP"、"NP-SBJ"、"NP;\*SBJ\*"、"NPR"(固有名詞)など)
  - [^NP\b]  
NPに**完全一致**するほか、NPの後に**境界記号 (ハイフンやセミコロン) のあるもの**にもマッチ  
("NP"、"NP-SBJ"、"NP;\*SBJ\*")
- ハイフンは**拡張タグとの境界**に、セミコロンは**照応情報**や quantification のアノテーションに用いられる