



統語・意味解析情報付き
日本語コーパスの構築に向けて
デモンストレーション

アラスデア・バトラー (NINJAL)

窪田 愛 (NINJAL)

窪田 悠介 (筑波大学)

目的:

インターフェイスの設計と機能の紹介

進め方:

- ▶ 事例 1: ヨリ句が形容詞と共起する比較構文を探す
- ▶ 事例 2: CSC 違反の例を探す
- ▶ 事例 3: 連用節とモダリティ形式の共起関係を調べる
- ▶ インターフェイスの各種機能の紹介

いろいろな「より」

- (1) 2007年1月28日より高速鉄道の営業運転を開始した。
- (2) 新1号での単独路線で番組を継続することになった（「僕らのスーパーヒーロー伝説」より）。
- (3) 通常の大雨によるものより動きが速かった
- (4) 企業との共同研究では自分の技術をそのまま使うより形を変えて応用することが多い。
- (5) 一部の子どもの問題がより深刻化していることがうかがえる。

探したい構造:

[S ... [PP [NP ...] より] ... 形容詞]

いろいろな「より」

- (1) 2007年1月28日より高速鉄道の営業運転を開始した。
- (2) 新1号での単独路線で番組を継続することになった（「僕らのスーパーヒーロー伝説」より）。
- (3) 通常の大雨によるものより動きが速かった
- (4) 企業との共同研究では自分の技術をそのまま使うより形を変えて応用することが多い。
- (5) 一部の子どもの問題がより深刻化していることがうかがえる。

探したい構造:

[S ... [PP [NP ...] より] ... 形容詞]

「より」をGoogle検索

より - Wikipedia

<https://ja.wikipedia.org/wiki/より> ▼

よ (**より**、U+309F) は、平仮名「よ」と平仮名「り」を組み合わせた (合字) 平仮名 (合略仮名) のひとつ。「より」と発音される。近世から近代の日本で使用されており、主に手紙の差出人の末尾に使用される。英語ではfromの意味にあたる。

[使用例](#) · [よ](#)に関する諸事項 · [符号位置](#) · [脚注](#)

より - Wiktionary

<https://en.wiktionary.org/wiki/より> ▼ このページを訳す

Particle[edit]. より (romaji yori) than; rather than; more so than. 上田 (うえだ) さんは僕 (ぼく) よりせか高 (たか) い。 Ueda-san wa boku yori se ga takai. Mr. Ueda is taller than I am. オレンジ**より** 林檎 (りんご) が好 (す) き。 Orenji yori ringo ga ...

[Etymology](#) · [Pronunciation](#) · [Particle](#) · [Usage notes](#)

「空飛ぶ車」は想像よりも早くやってくるかもしれない - GIGAZINE

gigazine.net/news/20161130-flying-cars-closer-think/ ▼

9時間前 - 安全上の理由から、「人類が車で空を飛ぶことを望まない」と語るアンドリーセン氏ですが、地上で**自動運転カーを実現するより**も自動で空を飛ぶ車を開発する方が容易であることを明かしています。その理由はさまざまなトラブルに対して**より** ...

いよりこどもクリニック

www.shujii.com/iyori/ ▼

日本脳炎ワクチンの納入量がますます制限されて、あまり入荷されなくなっております。そのため、ただ今 (2016/10/17 pm7) **より**一般診療時間内の日本脳炎の接種は受け付けませんこといたしました。日本脳炎をご希望の方は、月火木金の14:00-15:00の時間 ...

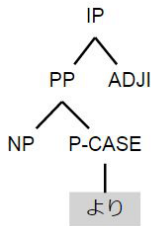
NPCMJで「より」+形容詞を検索

1. 「文字列検索」で検索ボックスに「より」と入れて Submit を押す。
2. 85 例出る [よ][り]_P-CASE_PP_IP が「より」が後置詞としてタグ付けされているケースなので、このリンクをクリック。
3. 上から 10 番目位に「平日昼間より土休日昼間の方が利用者が多い」という例がある。これが探している構造に適合する比較的構造が単純な文なので、これを基に検索パターンを作る。
4. この文の「より」のタグである P-CASE をクリックして、この文の構造木を別の窓に表示させる。
5. 「Build query」を押して表示されるグリッドで、以下のように選択するノードを指定。(IP-MAT が gen になっていることに注意。)

IP-MAT									
gen									
PP tag		PP-TMP				PP-SBJ		ADJI	PU
								tag	
NP tag	P-CASE	NP			P-CASE	NP	P-CASE	多い	。
	tag								
N	より	PP		N	が	N	が		
	word								
平日昼間		NP	P-CASE	方		利用者			
		N	の						
		土休日昼間							

NPCMJで「より」+形容詞を検索 (続き)

6. Construct ボタンを押すと、以下のような検索パターンが生成されるので、そのまま Submit を押す。



7. すると欲しいパターンの例文を抽出できる。

Coordinate Structure Constraint (等位接続構造制約)

Coordinate Structure Constraint (CSC; 等位接続構造制約)

In a coordinate structure, no conjunct may be moved nor may any element contained in a conjunct be moved out of that conjunct.

(Ross 1967: 89)

- (6) a. *This is **the magazine** that John [[bought ___] and [read the newspaper]].
- b. This is **the magazine** that John [[bought ___] and [didn't read ___]].

- (7) Here's **the whiskey** which I [[went to the store] and [bought
—]]. (Ross 1967)
- (8) That's **the stuff** that the guys in the Caucasus [[drink —]
and [live to be a hundred]]. (Schmerling 1972)

日本語の「テ形」とCSC

日本語ではどうか? (Kubota & Lee 2015 など)

- (9) a. *太郎が [[雑誌を買って]、 [__ 読んだ]] 本
b. 太郎が [[__ 買って]、 [__ 読んだ]] 本

例外はあるか?

- (10) 太郎が [[紀伊国屋に行って]、 [__ 買った]] 本

→ こういう例外を Google から拾ってくるのは大変!

探したい構造:

[RC ... [... て] [... た]] 名詞



どちらかの節からのみ抜き出しが起こっている

NPCMJでテ形のCSCの例外を検索

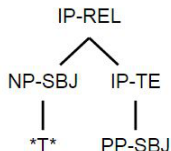
1. 「パターン・ブラウザ」の「拡張タグ」から「て-clause」を選択し、Submit を押す。
2. 5 文目「暗くて、たくましくて、ちょっともの悲しいリアス式海岸は…」がテ形を中に含む関係節の例になっているので、このリンクをクリックし、別窓に木構造を表示する。
3. 「Build query」をクリックして表示されたグリッドで、IP-REL とその直下の IP-TE に tag 指定をして、検索パターンを作成。
4. 作成された、IP-REL の直下に IP-TE があるパターンで検索をする。(Submit を押す。)
5. 例文を順に見ていくと、44 文目に「資金が尽きて活動を断念した団体もある」という文がある。これが、求めているパターンに当てはまるケースなので、これを基に検索を絞り込む。

NPCMJでテ形のCSCの例外を検索 (続き)

6. ツリー表示、「Build query」のグリッドで以下のように選択。

IP-REL				
tag				
NP-SBJ	IP-TE			
tag	tag			
ZERO	PP-SBJ		VB	P-CONJ
	tag			
T	NP	P-CASE	尽き	て
word				
	N	が		
	資金			

7. Construct ボタンを押すと、以下のような検索パターンが生成されるので、そのまま Submit を押す。



8. すると欲しいパターンの例文を抽出できる。

