

前提投射のNPCMJコーパスでの検索

窪田悠介 (国語研)、峯島宏次 (お茶大)

概要

- NPCMJのwebインターフェースで前提投射の実例を収集する

概要

- NPCMJのwebインターフェースで前提投射の実例を収集する

アウトライン

- 背景
- 「前提投射」の説明
- タスク・手法の概要
- 検索操作のデモ
- 結果の考察
- 今後の展望、課題

背景

(0) ベテラン農家も加われば、長年の経験で培われた技術を共有できる。(河北新報, 38_news_KAHOKU_1106)

背景

- (0) ベテラン農家も加われば、長年の経験で培われた技術を共有できる。(河北新報, 38_news_KAHOKU_1106)
- ⇒ ベテランでない農家がすでに「加わ」っている
(= 農業をしている)

背景

(0) ベテラン農家も加われば、長年の経験で培われた技術を共有できる。(河北新報, 38_news_KAHOKU_1106)

⇒ ベテランでない農家がすでに「加わ」っている
(= 農業をしている)

- 理論言語学研究でのコーパスの利用
 - 実証性の担保 (Harris & Potts 2009; CommitmentBank (de Marneffe et al. in *SuB23*))
- 「深い」自然言語処理研究のための言語資源の構築
 - ccg2lambda (Mineshima et al. 2015) などの意味解析システムの検証のための実データの利用

前提投射

(1) ジョンも来た。

真理条件: ジョンが来た

前提: ジョン以外の誰かが来た

前提投射

(1) ジョンも来た。

真理条件: ジョンが来た

前提: ジョン以外の誰かが来た

(2)

- a. ジョンも来たの? (疑問文)
- b. ジョンも来たらしい。 (モーダル)
- c. ジョンも来たなら、さぞ賑やかだっただろう。(条件節)
- d. ジョンも来たという噂が街に流れた。 (埋め込み)

前提投射の実例の検索

探したい構造:

[疑問文 ... Xも ...]

[... Xも ... モーダル]

[[... Xも ... たら/なら] ...]

[... [... Xも ...] と/ことを/のを V]

→ ツリーバンクを使うと構造を指定して検索できる。

手法

- 利用するコーパス: NPCMJ web公開版
(実例1万文; 調査を行った2018年9月時点)
- 調査対象:
 - 前提トリガー: 「も」「まで」「再び」「さらに」
(累加的意味を表す取り立て詞・副詞)
 - 埋め込み環境: 埋め込み節、条件節、文末モーダル
- 調査方法:
 - 窪田と峯島でまず別々に検索結果を分類
 - 結果を付き合わあせて調整

検索式

% 埋込み

```
[IP-EMB] << ([P] < も)  
[IP-EMB] << ([P] < まで)  
[IP-EMB] << さらに  
[IP-EMB] << 再び
```

% 条件節

```
[IP-ADV-CND] << ([P] < も)  
[IP-ADV-CND] << ([P] < まで)  
[IP-ADV-CND] << さらに  
[IP-ADV-CND] << 再び
```

% モーダル

```
[PP] < ([P] < も) $ [MD]  
[PP] < ([P] < まで) $ [MD]  
__ < (__ < さらに) $ [MD]  
__ < (__ < 再び) $ [MD]
```

検索画面



Tree search - Mozilla Firefox

Tree search × +

npcmj.ninjal.ac.jp/interfaces/cgi-bin/tree_search.sh?lang=jp&db=np 133%

コーパス概要 タグ 語の依存関係 文字列検索 ツリー検索 ABC English クレジット

ツリー検索について TGrep-liteについて XPathについて テキストを解析してクエリを作成

ツリー検索

[IP-ADV-CND] << ([P] < /も/)

Submit news

Copyright © 2018 NINJAL Parsed Corpus of Modern Japanese. All rights reserved.

検索画面



Tree search results - Mozilla Firefox

Tree search results × +

npcmj.ninjal.ac.jp/interfaces/cgi-bin/search.sh?pat=[IP-EMB]<<+([P] 133%

コーパス概要 タグ 語の依存関係 文字列検索 ツリー検索 English クレジット

検索結果について TGrep-lite について

TGrep-lite 検索結果

[IP-EMB] << ([P] < /も/)

Submit news reveal

1. *speaker+pro* 授業中や登下校中など*T*さまざまな場面の地震発生を想定し、**児童がいつでも*T*適切な避難行動を取れる**ように備えています。 [news KAHOKU 36 104](#)
2. 環境省動物愛護管理室によると、震災では*T* *pro* 被災者とペットの生活空間を分けた避難所があった一方で、*pro* 同行避難しても*pro* 鳴き声やアレルギーなどを理由に受け入れが認められなかったケースもあった。 [news KAHOKU 35 12](#)
3. 結局、県連は*pro* *arb* **どちらも推薦しない**自主投票とした。 [news KAHOKU 28 91](#)
4. *speaker+pro* **子どもらも楽しんで参加できる**よう、*pro* 訓練に併せてバーベキューをしたり、参加者に景品を配ったりします。 [news KAHOKU 34 126](#)
5. 検討会議は、*pro* **他県で広域災害が発生した際にも支援拠点の役割を果たす**よう提案。 [news KAHOKU 78 67](#)
6. *pro* **記事の表示や配置の仕方を工夫し、更新頻度も高める**ことで、*T*より使いやすいサイトへの成長を図る。 [news KAHOKU 97 19](#)
7. *pro* ***T* *T* デジタル紙面の公開期間である過去1週間分の記事も、手軽に読み直す**ことが

結果

タイプ	文数	説明
yes	60	前提投射
local	6	ローカル環境で前提が満足
factive	23	叙実術語に埋め込み
factive, local	7	叙実術語、ローカル
ambig	11	統語的、語彙的曖昧性
n/a	281	該当しない例
total	388	

データ公開URL:

<https://www2.ninjal.ac.jp/kubota/ljsj-presup-data.csv>

論文 (言語学会予稿集):

<https://www2.ninjal.ac.jp/kubota/papers/ljsj-presup-proc0.pdf> 14

結果 (内訳)

	yes	n/a
も	44	201
まで	1	74
再び	7	1
さらに	8	5
合計	60	281

結果: yesの例

埋め込み節

パターン:[... [... さらに ...] **COMP V**]

(3) 14年は誘致実現へ向け、土台をさらに固める大事な年になる見通しだ。(yes)

(8_news_KAHOKU_89;K201401040A0F70XX00001;JP)

条件節

パターン:[[...**Xも**...たら/なら]...]

(4) ベテラン農家も加われれば、長年の経験で培われた技術を共有できる。(yes)

(38_news_KAHOKU_1106;K201401010A0M106B00004;JP)

結果: 時間軸上の変化を表す埋め込み節

(5) 臨済宗が再び活気に満ち溢れるようになるのは、北宋代であり、[...] (yes)

(9_wikipedia_KYOTO_2;BDS_00002;9;JP)

(6) 当初は平日のみだったが、2006年3月18日のダイヤ改正から土曜・休日にも運転されるようになった。(yes)

(46_wikipedia_KYOTO_22;RLW_00004;60;JP)

(7) 沿岸部は入院ベッドが今も減ったままだ。(yes)

(14_news_KAHOKU_93;K201403050A0T10XX00201;JP)

結果:「も」のゆるい例

(8) 仙台市は、大企業の支店の存在に経済を支えられている側面が強く、また工業も発展していないため、経済関係者の間から「支店経済都市」「支社都市」と呼ばれている。(local + factive)

(217_wikipedia_Sendai_City;JP)

(9) 記事の表示や配置の仕方を工夫し、更新頻度も高めることで、より使いやすいサイトへの成長を図る。(local)

(19_news_KAHOKU_97;K201402230A0T20XX00001;JP)

(10) 近畿日本鉄道にも近鉄奈良線があるため、「JR奈良線」と案内されることもある。(yes)

(13_wikipedia_KYOTO_22;RLW_00004)

結果：曖昧性のため判定が困難な例

統語的曖昧性

(11) 同時に、警報を国から自治体へ、その先の市民へ、地震で寸断されることなく伝達する通信システムの開発と整備も進める必要がある

(81_news_KAHOKU_63;K201401110A0T10XX00001;JP)

も > 必要がある:

前提: 他のものを進める必要がある

必要がある > も:

前提: 他のものを進める (ことがすでに決まっている)

結果：曖昧性のため判定が困難な例

語彙的曖昧性

(12) しかし、それでも、国語の苦手な子には、静かに本と向きあう時間が苦痛になることも考えておかなくてはなりません。

(5_book_excerpt-28;JP)

結果 (まとめ)

- NPCMJ1万文のデータから前提投射の実例を取ることが出来た。

言語学的に興味深い点:

- 「よくなる」(時間的变化) のようなものが前提投射を引き起こす。
- 累加の「も」といわゆる詠嘆の「も」の間の境界は明確か? (cf. 沼田 (1995, 2009))

(10) 近畿日本鉄道にも近鉄奈良線があるため、「JR奈良線」と案内されることもある。

(13_wikipedia_KYOTO_22;RLW_00004)

今後の課題

- 他の環境や他のトリガー、他の「投射的意味」(Potts (2005) のCIなど)を調べる
 - **環境:** 疑問文、否定、引用節
 - **トリガー:** 叙実動詞、分裂文、所有格名詞
 - **「投射的意味」:** 非制限的關係節、評価副詞、程度副詞、尊敬語
- **生起頻度が低いトリガーの効率的な検索方法の検討**
 - BCCWJを機械的に全文パーズして検索する?
 - データ量と精度のトレードオフ

Thank you!

謝辞

- 本研究には以下のコーパスを利用しました。
国立国語研究所『NINJAL Parsed Corpus of Modern Japanese』 <http://npcmj.ninjal.ac.jp/interfaces/> (2018年9月30日確認)
- 本研究はJSPS科研費 18K00523、15H03210、および、国立国語研究所共同研究プロジェクト「統語・意味解析コーパスと言語研究」「対照言語学的観点から見た日本語の音声と文法」の研究成果の一部である。