

国立国語研究所学術情報リポジトリ

統語コーパスを用いた言語研究の試み：
日本語の等位接続構造制約をめぐって

メタデータ	言語: jpn 出版者: 公開日: 2021-06-25 キーワード (Ja): キーワード (En): 作成者: 窪田, 悠介 メールアドレス: 所属:
URL	https://doi.org/10.15084/00003358

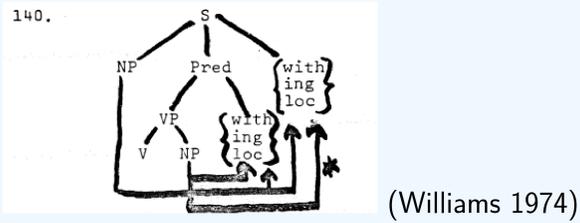


窪田悠介
国立国語研究所

研究の目的

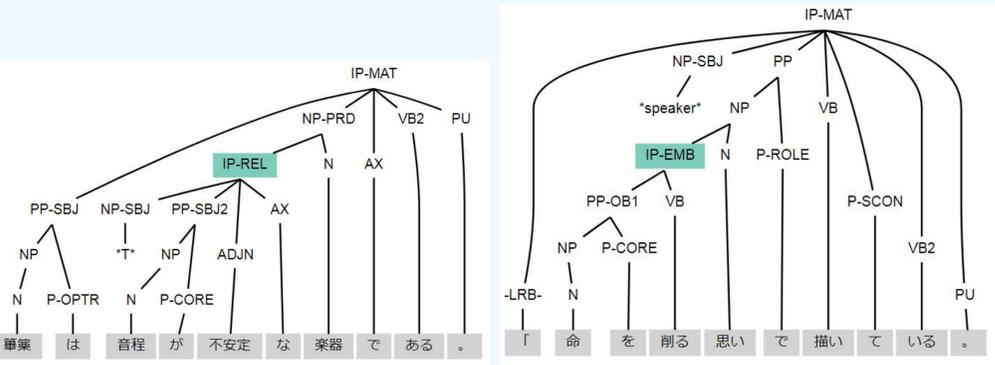
1. 統語コーパスを用いた言語研究の可能性を模索する
2. 1を通して統語コーパスのコーパスとしての特徴を考える

「紙とエンピツ」の言語学の時代



NPCMJコーパスの特徴

- ▶ 国語研で2016年から開発している統語コーパス (ポスター①を参照)
- ▶ 言語学研究での利用を念頭に置いたタグ付け (文法関係、ゼロ代名詞など)



▶ NPCMJ検索インターフェイス: <http://npcmj.ninjal.ac.jp/>

統語コーパス (ツリーバンク) のコーパスとしての特徴

	BCCWJ	NPCMJ
サイズ	大 (100M words)	小 (0.5M words)
粒度	形態論情報のみ ¹	統語構造
主な用途	量的分析	質的分析

統語コーパス (ツリーバンク) の向き不向き

- ▶ 構造的な要因が鍵になっているような現象の分析に有効
- ▶ 特に、節境界の判定などがある程度精確にする必要がある場合に役に立つただし、
- ▶ 大量のデータが必要な研究には向かない
- ▶ アノテーション基準を正しく理解して使う必要がある

ケーススタディ: 日本語の等位接続構造制約

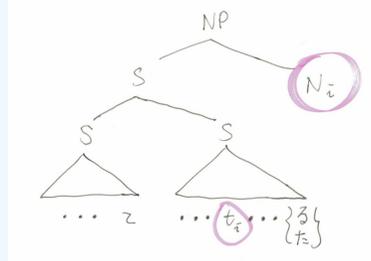
NPCMJコーパスを用いた言語研究の一例として、日本語の等位接続構造制約に関して考察を行った Kubota & Kubota 2019の研究を紹介する。

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

- ▶ 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)
- (1) a. *This is the magazine that John [[bought __] and [read the newspaper]].
b. This is the magazine that John [[bought __] and [didn't read __]].
- ▶ 等位接続構造制約の例外:
- (2) Here's the whiskey which I [[went to the store] and [bought __]]. (Ross 1967)
- (3) That's the stuff that the guys in the Caucasus [[drink __] and [live to be a hundred]]. (Schmerling 1972)

日本語の「等位接続構造制約」

- (4) 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
 - (5) [[針を刺して] [__ 出した]]血液
 - (6) [[紀伊国屋に行って] [__ 買った]]本
- ▶ 課題: (5)や(6)のような文の実例を採取することはできるか?



Google検索の限界

クエリ: *し(て) *した *は
問題点:

- ▶ 検索対象がサ変動詞に限られてしまう
- ▶ 検索結果にゴミが多い

NPCMJによる検索

クエリ:
IP-REL < (IP-ADV < (NP-SBJ < *T*)
< VB
< (P < て|で))
< (NP-SBJ !< *T*)

得られた実例

- (7) 日本人がロシア文学を読んで感じる遠さ
- (8) 株主総会が終了し、会場を後にする三洋電機の株主たち
- (9) ダム建設で原野が水没して姿を消した種類

検索結果の分類

	true hit	ambig.	false hit	error, etc.	total
ren'yoo/1st/sbj	1	1	0	0	2
-te/1st/sbj	0	0	0	0	0
ren'yoo/1st/obj	0	0	1	0	1
-te/1st/obj	0	0	3	0	3
ren'yoo/2nd/sbj	1	15	5	8	29
-te/2nd/sbj	7	7	25	6	45
ren'yoo/2nd/obj	0	7	1	2	10
-te/2nd/obj	12	8	15	1	36
total	21	38	50	17	126

結果の考察

- ▶ CSC違反は連用形よりテ形でより多く見つかった。(19例対2例)
 - ▶ テ形のほうが「従属節的」な意味関係を表しやすいという直観と合う
 - ▶ ただし、これは傾向性に過ぎない。(先行研究 Tokashiki (1989) の、連用形は常にCSCに従うという主張への反例)
- ▶ 右側 (定形節) からの抜き出しのほうが多い。(20例対1例)
 - ▶ 主名詞に近いから?

謝辞: 本研究は国立国語研究所共同研究プロジェクト「統語・意味解析コーパスの開発と言語研究」、JSPS科研費JP15H03210の研究成果の一部である。