

目次

第1章	イントロダクション	1
1.1	ツリーバンクの意義	1
1.2	ツリーバンクの特徴	6
1.3	日本語ツリーバンク・グループ	10
1.4	この本の構成	11
第2章	アノテーションの概要	13
2.1	統語情報の表記	13
2.2	セグメンテーション	15
2.3	品詞	16
2.4	句構造のスキーマ	17
2.5	節の構造	21
2.6	主要文法役割	23
2.7	任意文法役割	26
2.8	インデックスを使用しない空要素	27
2.9	インデックスを使用する空要素	30
2.10	照応関係	31
第3章	単純な構文	33
3.1	助詞	33
3.1.1	はじめに	33
3.1.2	格助詞 (P-ROLE)	34
3.1.3	接続助詞 (P-CONN)	48
3.1.4	補文助詞 (P-COMP)	50
3.1.5	終助詞 (P-FINAL)	51
3.1.6	間投助詞 (P-INTJ)	52
3.1.7	とりたて助詞 (P-OPTR)	52

3.2	句	54
3.2.1	名詞句 (NP)	54
3.2.2	量化名詞句	63
3.2.3	副詞句 (ADVP)	68
3.2.4	副詞節以外の構成素の並列構造	71
3.3	述語	74
3.3.1	はじめに	74
3.3.2	動詞 (VB)	75
3.3.3	軽動詞 (VB0)	75
3.3.4	イ形容詞 (ADJI)	76
3.3.5	ナ形容詞 (ADJN)	78
3.3.6	句や節が述語となる場合 (-PRD)	79
3.3.7	コピュラ (AX)	80
3.3.8	テンス標識 (AXD)	83
3.3.9	否定辞 (NEG)	84
3.3.10	助動詞 (AX)	84
3.3.11	モーダル要素 (MD, ADJI-MD, ADJN-MD)	86
3.3.12	形式名詞 (FN)	88
3.3.13	補助動詞 (VB2)	89
3.3.14	受動補助動詞 (PASS, PASS2)	92
3.4	曖昧な語形 (1) 「に」「と」「の」「で」	92
3.4.1	に	92
3.4.2	と	94
3.4.3	の	95
3.4.4	で	96
第4章	複雑な構文	97
4.1	節	97
4.1.1	はじめに	97
4.1.2	主節 (IP-MAT) と準主節 (IP-SUB)	98
4.1.3	関係節 (IP-REL)	98
4.1.4	空所なし名詞修飾節 (IP-EMB)	102

4.1.5	名詞化節 (IP-NMZ)	106
4.1.6	小節 (IP-SMC)	107
4.1.7	副詞節 (IP-ADV)	110
4.1.8	終助詞節 (CP-FINAL)	113
4.1.9	命令節 (CP-IMP)	114
4.1.10	感嘆節 (CP-EXL)	118
4.1.11	疑問節 (CP-QUE)	119
4.1.12	補部節 (CP-THT)	122
4.1.13	コントロール	126
4.1.14	等位節と ATB 抽出	131
4.2	様々な構文	133
4.2.1	はじめに	133
4.2.2	二重主語文	133
4.2.3	受動文	135
4.2.4	使役文	137
4.2.5	テアル構文	138
4.2.6	テアゲル／テクレル構文	140
4.2.7	目的語繰り上げ構文	141
4.2.8	主語繰り上げ構文	142
4.2.9	N-bar 削除	142
4.2.10	右方節点繰り上げ	143
4.2.11	動詞なしの付帯状況構文	143
4.2.12	移動の目的を表す節	145
4.2.13	括弧挿入句 (PRN)	145
4.2.14	主要部内在型関係節	147
4.2.15	代名詞残留型関係節	147
4.3	話し言葉特有の表現	148
4.3.1	はじめに	148
4.3.2	間投詞 (INTJ) および間投詞句 (INTJP)	148
4.3.3	言い誤り (FS)	149
4.3.4	縮約	150

4.4	曖昧な語形 (2) 「れる／られる」	151
4.4.1	直接受動	151
4.4.2	間接受動	152
4.4.3	自発	153
4.4.4	可能	154
4.4.5	尊敬	154
4.5	曖昧な語形 (3) 「よう」	154
4.5.1	様態	155
4.5.2	直喩	156
第5章	検索の方法	157
5.1	文字列検索	158
5.2	ツリー検索	160
5.2.1	ツリー検索画面	162
5.2.2	検索対象ファイルの指定	162
5.2.3	検索結果の表示	163
5.2.4	検索結果のダウンロード	166
5.3	検索パターン	169
5.3.1	ノードの表現	169
5.3.2	正規表現の記述	170
5.3.3	ノード間の関係	172
第6章	検索の実例	189
6.1	語順	189
6.2	主語と主題	190
6.3	「(人) に会う」と「(人) と会う」	192
6.4	受動文	193
6.5	テアル構文	194
6.6	疑問文	195
6.7	名詞句	196
6.8	名詞節	197
6.9	関係節	198

6.10 副詞節	199
付録 逸脱的な「のが」文の成立要因 三好伸芳	201
1.1 はじめに	201
1.2 先行研究	202
1.3 調査方法	206
1.4 調査結果と分析	208
1.5 おわりに	213
第6章課題の解答	215
謝辞	221
参考文献	225

第 1 章

イントロダクション

1.1 ツリーバンクの意義

この本は、高度な文法解析情報付き日本語コーパスである「かいのきツリーバンク」(Kainoki 2022) を日本語の研究や教育・学習に利用することを希望する人々のために、そのアノテーション (構文解析情報) および例文検索用のオンライン・ツールの解説書として書かれた。かいのきツリーバンクのアノテーションは日本語の構文すべてをカバーするので詳細であり、また検索パターンも初めて目にする人には手ごわく感じられるかもしれない。しかし、本書では、文解析の原理から説き起こし、アノテーションについても検索ツールについても、実例を交えてできるだけ平易な説明を心掛けた。

かいのきツリーバンクを日本語文法の研究に応用して有用なデータを得られることの例として、語順の問題を考えてみよう。ここでは、動詞が伴う 2 つの目的語の間の語順を考察することにする。これらの目的語は、かいのきツリーバンクでは第一目的語 (OB1) および 第二目的語 (OB2) と呼ばれ、それぞれ伝統文法の直接目的語および間接目的語にほぼ相当する。

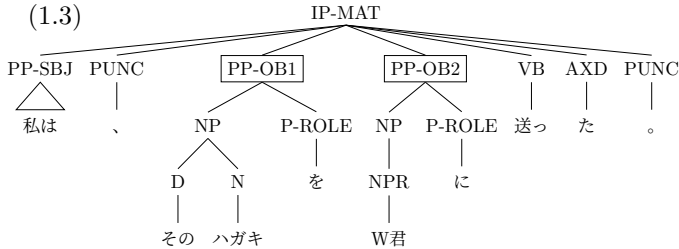
以下の (1.1) および (1.2) に、第一目的語と第二目的語の順番の異なる例文を挙げる。これらは共に、太宰治『酒ぎらい』(1940) の中に近接して出現している。

(1.1) 私は、そのハガキを W君に 送った。

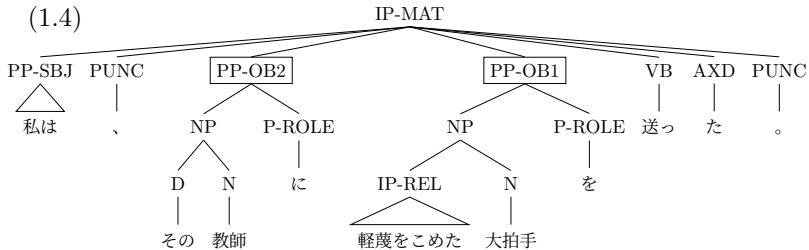
(1.2) 私は、その教師に 軽蔑をこめた大拍手を 送った。

かいのきツリーバンクにおけるこれらの解析ツリー (文に対する統語アノテーションのツリー表示) を (1.3) および (1.4) に示す (一部、省略表示)。両方の文とも、助詞「は」で表示された主語、動詞「送る」、さらに第一目的語 (OB1) と第二目的語 (OB2) を有することが見て取れる。他方、(1.1)/(1.3)

では <第一目的語-第二目的語> の語順なのに対し、(1.2)/(1.4) では <第二目的語-第一目的語> の語順となっているという違いがある。



(40_aozora_essay_dazai1940a)



(55_aozora_essay_dazai1940a)

本書の第5章で説明するウェブ上のインターフェースに検索パターンを入力することによって、特定の統語論的条件を充たす例文データを収集することができる。表1.1はそのようにして集めた、<第一目的語 (表では OB1)-第二目的語 (OB2)-動詞> の語順 (「目1-目2」と略記) と <第二目的語 (OB2)-第一目的語 (OB1)-動詞> の語順 (「目2-目1」と略記) の文の用例数を比較したものである。それぞれの検索パターンも合わせて示す (検索パターンの基本については5.3節で、また語順に関するパターンの書き方については6.1節で解説する)。ただし、コーパス・データから教科書例文等の作例や翻訳文を除いた57,504文 (1,149,637語) について調査を行った。このような検索対象の絞り込みは、オンライン・インターフェースの“Files”のボックス (5.2.2節を参照のこと) に“/x_\|trans/!”と記入することにより可能となる。集計によると、<OB1-OB2> の語順を持つ文は233例で、2つの目的語を持つ例文全体の37.9%、これに対して

第 2 章

アノテーションの概要

本章および続く第 3 章と第 4 章では、読者がかいのきツリーバンクを使うために必要なアノテーションに関する知識を提供することを目的としている。ここでの叙述は概要の理解を目指したものであるため、個々の問題の詳細については「かいのきツリーバンク解析ガイド」(長崎他 2021 以下, 「解析ガイド」)を参照していただきたい。

2.1 節では、アノテーションとして与えられる統語情報をどのように表記しているかについて述べる。アノテーションの最も基礎をなすのはテキストの語への分解であるが、その大まかな方針について 2.2 節において、それぞれの語に与えられる品詞について 2.3 節で説明する。続いて、語がどのように組み合わせられて句を構成するか、また句に対してどのようなタグを与えるかについて、2.4 節で解説を行う。さらに 2.5 節で、文構造の最上位のレベルである節がどのように構成されるか、また節にはどのような種類のものがあるかについて述べる。2.6 節では、節の重要な構成素である、主要文法役割を表す句について、また続く 2.7 節では任意文法役割を持つ句について解説を行う。さらに、主要文法役割を表す句を中心とする空要素のアノテーションに関し、2.8 節ではインデックスを使用しない場合、2.9 節ではインデックスを使用する場合について説明する。最後に 2.10 節で、代名詞やゼロ代名詞を同定するためのソート情報について述べる。

2.1 統語情報の表記

かいのきツリーバンクでは、データ中のすべての文に統語構造(統語木、ツリー)がアノテーションされている。例えば、「太郎は手を上げた」という文の統語構造をツリーとして表示すると、(2.1) のようになる。

第 3 章

単純な構文

本章では、1 個の述語を中心として構成される比較的単純な文が、そのバリエーションに応じてどのようなアノテーションを受けるかについて解説する。

まず [3.1](#) 節で助詞について、次に [3.2](#) 節で句の様々な種類とそのアノテーションについて、続いて [3.3](#) 節で文の中核となる述語がどのように分析されるかを説明する。最後に [3.4](#) 節で曖昧な語形「に」「と」「の」「で」を取り上げ、その用法の区別のための手掛かりについて述べる。

3.1 助詞

3.1.1 はじめに

助詞 (P) は句や節に付加されて、その句や節の持つ様々な文法的機能や意味を表示する。助詞は機能ごとに、格助詞 (P-ROLE)、接続助詞 (P-CONN)、補文助詞 (P-COMP)、終助詞 (P-FINAL)、間投助詞 (P-INTJ)、とりたて助詞 (P-OPTR) の 6 つのグループに分けられ、必ず固有の拡張タグを付与される。いくつかの語が連語となって助詞としての役割を果たす、いわゆる複合助詞も、単独の助詞として 6 つのグループのいずれかに分類される。また、同一の助詞が同一グループの中で複数の機能を果たすことがある。例えば、「に」は、主語、第一目的語、第二目的語、論理的主語、被使役者を示す他、到達点や目的等を示すことができる。

助詞は通常、助詞句 (PP) を投射する。ただし、補文助詞 (P-COMP) や終助詞 (P-FINAL) は節に相当する構成素 (CP) を投射すると分析する ([4.1.12](#) 節および [4.1.8](#)、[4.1.10](#)、[4.1.11](#) 節を参照)。また、助詞の中には述語の拡張要素に付随してあらわれるものがあり、この場合は助詞を節 (IP) の下に置く。このような助詞については、接続助詞 (P-CONN) について扱った [3.1.3](#) 節および、述語拡張要素について扱った [3.3.7](#) および [3.3.11](#) 節の説

第 4 章

複雑な構文

前章では基本的な文のアノテーションについて解説を行った。本章ではより複雑な構文のアノテーションを取り上げる。まず、[4.1](#) 節において節 (clause) の分類と各々の節について述べる。次に、[4.2](#) 節において [4.1](#) 節で扱いきれない多様な構文について述べる。さらに [4.3](#) 節で、話し言葉に特有の様々な表現の取り扱いについて説明する。最後に [4.4](#) 節において曖昧な動詞「れる／られる」形の、また [4.5](#) 節において「よう」の多様な用法の区別について解説する。

4.1 節

4.1.1 はじめに

本コーパスでは、節を大きく IP 節と CP 節に分類する。IP 節は述語によって投射され、項、様々な助詞句、副詞句、および従属節を支配する節である。以下のように下位分類され、必ず固有の拡張タグを付与されるが、さらにその後に機能的分類に即した拡張タグが付与されることがある。

主節 (IP-MAT), 準主節 (IP-SUB), 関係節 (IP-REL), 空所なし名詞修飾節 (IP-EMB), 名詞化節 (IP-NMZ), 小節 (IP-SMC), 副詞節 (IP-ADV)

CP 節は、終助詞を伴う文、および様々な発話行為を表す文や、伝達・認識動詞の内容を表すために設定された節である。以下のように下位分類され、IP 節と同様に固有の拡張タグを付与される。また、その後にさらに機能的分類に即した拡張タグが付与されることがある。

終助詞節 (CP-FINAL), 命令節 (CP-IMP), 感嘆節 (CP-EXL), 疑問節 (CP-QUE), 補部節 (CP-THT)

第 5 章

検索の方法

本章ではかいのきツリーバンクのオンライン検索インターフェースと、検索方法の 1 つであるツリー検索を行うための検索パターンの作り方を説明する。コーパスを利用するに当たって最も基本的と考えられる事項、すなわち、コーパスの検索と検索結果の閲覧、および検索結果のダウンロードを扱う。

かいのきツリーバンクのオンライン検索インターフェースは、誰でも特別な手続きなしで利用することができる。かいのきツリーバンクの URL (<https://kainoki.github.io>) にアクセスし、メニューの「検索インターフェース」をクリックすると、検索インターフェースのトップページ (図 5.1) があらわれる。

The screenshot shows the search interface for the Kainoki Treebank. At the top, there are links for 'Toggle', 'tree | source', '[Front page/フロントページ]', and 'Help'. The main area is titled 'The Kainoki Treebank search interface' and contains a search form with fields for 'Files:', 'TGrep2 search expression:', and a search button. Below the search form, there are statistics: '495 files; 102,875 trees; 1,537,367 words'. A table displays search results with columns for 'Filename', 'tree count', 'word count', and 'content description'.

	Filename	tree count	word count	content description
1	academic_akaike2018	45	1177	essay : Academic essay : Akaike, Takaaki (2018). <i>Honyuurui ni okeru Manabinomori</i> 83
2	academic_aoki2009	65	1291	essay : Academic essay : Aoki, Junken (2009). <i>Seitaimaku no shishitsi Manabinomori</i> 50
3	academic_arai2012	50	1165	essay : Academic essay : Arai, Hiroyuki (2012). <i>Choojushakai to nin Toohokudaigaku koohooshi Manabinomori</i> 60

図 5.1 検索インターフェースのトップページ

第 6 章

検索の実例

ここでは、日本語文法研究でよく取り上げられる 10 種類の構文について、その文例データ収集のための検索パターンの書き方を具体的に検討する。読者には実際に課題を解きながら、検索パターンの書き方に習熟してほしい。巻末に解答を掲載するが、これは自分で得た答えの確認のために使用すること。

6.1 語順

日本語の格成分（主要文法役割や任意文法役割を担う助詞句、名詞句、および副詞句）の間の順序は比較的自由であることがよく指摘される（[佐伯 1975](#)、[佐伯 1998](#)、[益岡・田窪 1992](#) 等）。他方、文章を書いている、文の落ち着きを良くするために語順を変えてみることはよく経験する。格成分相互の語順が実際のテキストにおいてどうなっているかを調べる手始めとして、〈主語 – 第一目的語〉の語順で出現する文、および逆の語順で出現する文の用例をそれぞれツリーバンクから得ることを考えてみよう。

(6.1) /SBJ\b/ \$. /OB1\b/

このような検索パターンをまず思い浮かべるかもしれない。確かにこれは、助詞句であれ名詞句であれ、主語としての拡張タグを与えられた句が第一目的語のタグを持つ句に先行している文を探してくれる。しかし、実はこれでは主語や第一目的語が空要素であるような用例も検索結果に含まれることになる（[2.8](#) 節を参照のこと）。これらがゼロ代名詞や関係節のトレースに相当して表層の表現にあらわれない場合に、その語順を議論するのは無意味である。[2.8](#) 節で説明したように、ゼロ代名詞やトレースのような名詞句相当の空要素が認められる場合、*pro* 等のゼロ代名詞やトレース *T* を終端ノードとして設定し、これに NP-SBJ や NP-OB1 のよう

付録

逸脱的な「のが」文の成立要因—助詞「が」に関連する構造の分布—

三好伸芳

1.1 はじめに

「のが」(準体助詞「の」+格助詞「が」)を含む文(以下、「のが」文)には、「のが」句を受ける述語との関係において接続助詞的、あるいは逸脱的な性質を持つものがあることが知られている。[天野 \(2014\)](#) は、以下のような「のが」文にいくつかの逸脱的な特徴が見られるとする(以下、先行研究から引用した例文は、下線などを追加・削除する場合がある)。

- (1) a. 入院中は、毎日のように彼女が見舞ってくれた。(中略) それでも、心細さは消えなかった。都内の実家を離れ、1人暮らしを始めて3年目。ろくに実家に帰ることはなかったのが毎日、病院の公衆電話から家族の声を聞いた。

(天野 [2014](#): 25, 『朝日新聞』「患者を生きる」2011年7月13日)

- b. きょうはみなさん、遠いところを息子のためにおいでくださって、のあたりまでは、ごくふつうのあいさつだったのが、友人は急回転して、こんなことをいいはじめた。

(天野 [2014](#): 25, 須賀敦子「本に読まれて」)