

現代日本語における節の分類体系について

丸山 岳彦[†]

佐藤 理史[‡]

夏目 和子[‡]

[†] 国立国語研究所 言語資源研究系

[‡] 名古屋大学大学院工学研究科

1 はじめに

本稿では、現代日本語（話し言葉・書き言葉を含む）における「節」の分類体系について論じ、それをテキスト中にアノテーションする際の問題点を検討する。

節は、文節より大きく文より小さな統語的単位である。1つの節が1つの文（単文）を構成することもある。複数の節が1つの文（複文）を構成することもある。日本語における節の体系的分類および統語的・機能的な分析についてはすでに多くの記述的研究があるが、本稿では日本語コーパスに対して「節境界ラベル」をアノテーションするという観点から、節の分類体系を整理する。また、実際の処理で困難になる点について検討する。なお本稿では、益岡・田窪(1992)に準じて、節を「複文を構成するところの、述語を中心とした各まとまり」と定義しておくことにする。

2 言語の階層構造における節

言語表現は一般に、レベルの異なる複数の単位によって、以下のような階層構造を成す。

文字・音素 < 語 < 文節 < 節 < 文 < 文章・談話

このうち、節を除く5つについては、1つ以上の要素の連結によって上位要素が構成されることから、基本的には線条的な連結（チャンキング）に基づく上下関係として捉えられる。これに対して節の場合は、節と節を連結して上位の要素（文）が得られる場合（＝(1a)）だけでなく、節が語に連結して文の一部となる場合（＝(1b)）や、準体助詞を伴う節が文の項となる場合（＝(1c)）など、同じレベルの要素間の単純な連結関係として捉えられない場合がある。

- (1) a. 雨が降って + 遠足は中止になった。
 b. 雨が降った + 日は月曜日だ。
 c. 雨が降った + のは月曜日だ。

ここで、現代日本語コーパスの中に現れた節を網羅的に検出するというタスクを想定すると、まず、節の形態・統語的な特性に基づいて、節の種類を体系的に分類しておく必要がある。そこで以下では、日本語文法研究における従属節の分類について検討する。

3 日本語文法研究における従属節の分類

節は、文末に位置する「主節」と、それ以外の「従属節」に分けられる。従属節は、その形態的・統語的・機能的な特性から、複数の下位クラスに分類することができる。伝統的な一般言語学においては、主節に対する機能に着目した分類として、従属節を「補文・関係節・副詞節・等位節」という4種類に分けることが多い。一方、日本語文法研究における従属節の分類には、例えば以下のように、複数の見方がある。

- 補足節・副詞節・連体節・並列節 (益岡・田窪, 1992)
- 連用節・名詞節・連体節 (野田, 2002)
- 連用節・連体節 (村木, 2007)

益岡・田窪(1992)に挙げられているそれぞれの従属節の例文を、以下に示す。(下線は筆者)

- (2) a. 鈴木さんは車の調子が悪いことに気づいた。
 b. 時間があれば、出席します。
 c. この小説を書いた作家
 d. 太郎は休んだが、花子は休まなかった。

(2a)の下線部は主節の述語を補う働きをする補足節である。野田(2002)では名詞節、村木(2007)では連体節の下位にある擬似連体節として分類される。(2b)は主節の述語を修飾する副詞節である。野田(2002)・村木(2007)ではともに、連用節とされる。(2c)は名詞を修飾する連体節で、野田(2002)・村木(2007)でも連体節と呼ばれる。(2d)は主節に対して対等に並ぶ並列節である。このような例は野田(2002)では連用節として一括され、村木(2007)では連用節の下位にある並列節として分類されている。これらの対応関係を整理すると、表1のようにまとめられる。

表 1: 従属節の分類の比較

益岡	補足節	連体節	副詞節	並列節
野田	名詞節	連体節	連用節	
村木	連体節		連用節	
	擬似連体節	真性連体節	状況節	並列節

益岡・田窪 (1992) が副詞節と並列節を区分しているのに対して、野田 (2002) と村木 (2007) では両者が連用節として一括されている。英語などの言語では、副詞節と並列節 (等位節) は形態的に大きく異なるため、両者を区分する必然性があるが、日本語の並列節の場合、述語句の末尾形式を変化させて後方に連鎖させるという点では副詞節と同様であり、形態的には連用接続形式の一種と見なしてよい。野田 (2002)・村木 (2007) はその点を重視しているものと考えられる。

また、益岡・田窪 (1992) と野田 (2002) でそれぞれ補足節・名詞節として連体節と区別されている節が、村木 (2007) では連体節の中に含まれている。村木は連体節の下位クラスとして「真性連体節 (後続の自立的な名詞に接続する節)」と「擬似連体節 (後続の自立的な名詞以外の形式に接続する節)」を設けている。このうち、擬似連体節が形式名詞や準体助詞に接続した「名詞相当節」が、補足節・名詞節に相当する。形態的に見れば、形式名詞に対する連体修飾構造が生じている点で、補足節・名詞節を連体節の一種と見なすことも不自然ではない。一方で、被修飾部の自立性の違いに着目して、補足節・名詞節と連体節を明確に区分することもまた自然である。ある節の構造が複数の見方で捉えられる問題については、6 節で後述する。

さて、日本語コーパスを対象として節の情報を網羅的にアノテーションするためには、上記のような分類に対して、形態的・機能的な観点から複数の下位クラスを設定し、そこに含まれる節がどのような形式によって実現されるかを体系的に整理しておく必要がある¹。以下では、現在我々が進めている「節境界解析」の設計について示す。

4 日本語テキストに対する節境界解析

テキストに含まれる節の終端境界の位置を特定し、当該の節の形態的・統語的なタイプを表す「節境界ラベル」を付与する処理のことを、節境界解析と呼ぶことにする (丸山他, 2004, 2006; 佐藤他, 2016)。テキスト中に節境界ラベルを付与することにより、そのテキストに含まれる節境界の分布に関する統計的指標を得たり、テキスト・発話の分割処理に応用したりするなど、幅広い応用可能性が考えられる。

テキスト中に節境界ラベルを付与した結果の例を、以下に示す。(3) は『現代日本語書き言葉均衡コーパス』(BCCWJ)、(4) は『日本語話し言葉コーパス』(CSJ) からの例である。ここでは便宜的に、節境界ラベルを / / で囲んで示す。

¹池原 (2007) には、「節間意味分類体系」としてかなり網羅的な節の表現リストが記載されている。

(3) パソコンの画面や本などに集中しながら、
/ナガラ節/ 自分の入れた /連体節/ 飲み物に手を
伸ばし、 /連用節/ 飲み物にはまったく目を遣ら
ないまま /ママ節/ 飲む、というのは /補足節/ だ
れでもやる /連体節/ ことだろう。 /文末/

(BCCWJ: PB12.00001)

(4) 带状疱疹というのは水ぼうそうのウイルスによ
って /テ節/ 起こる /連体節/ 病気で /連用中止節/ 大
概小さい頃に水ぼうそうをやった /連体節/ 人は
必ず体の中に体の中のどこかに水ぼうそうのウ
イルスというのが残っていて /テ節/ でストレス
だとか /トカ節/ 後凄く極度の疲れとかによって
/テ節/ 突然水ぼうそうのウイルスがまた暴れ出し
て /テ節/ 発病するという /トイウ節/ ものなんで
すが /並列節ガ/ ...

(CSJ:S00F0210)

このようなアノテーションを自動的に実現するためには、(1) どのような位置に、(2) どのようなラベルを付与するか、という 2 点をまず設計しなければならない。このためには、節の具体的な出現形式を網羅的に収集し、それらを体系的に分類した上で、適切な節境界ラベルとその付与規則を準備することになる。

このうち、(1) 節境界の検出アルゴリズムについては、佐藤他 (2016) で論じている。以下では、(2) 付与する節境界ラベルの分類について示す。

5 節境界ラベルの分類と付与

我々は現在、BCCWJ に対して節境界解析を実施し、節境界ラベルを自動的に付与する実験を行なっている。これは、従来 CSJ に付与されていた節境界ラベルを、BCCWJ にも拡張することを目指すものである。

既存の節境界解析システムとしては、丸山他 (2004) による CBAP がある。しかし、入力となる形態素列が IPADic 体系に限定されていること、ルールが複雑で拡張可能性に難があること、などの問題があった。これに対して今回、新たに Rainbow のシステムを利用し、UniDic 体系で形態素解析された BCCWJ のテキストに対して、節境界ラベルを自動的に付与するシステムを作成した (加納・佐藤, 2014; 佐藤他, 2016)。

我々が新たに作成したシステムにおいて採用した節の分類体系を、図 1 に示す。全体の枠組みとしては、基本的に益岡・田窪 (1992) を踏襲しているが、従属節を大きく 3 つに分けて、並列節を連用節の下位に位置づけている点では、野田 (2002) の分類に近い。

さらに、実際に出現する節末形式に応じて、節境界ラベルを準備した。その一覧を、簡略化した形で図 2 に示す。これは、CBAP が検出対象としていた節境界

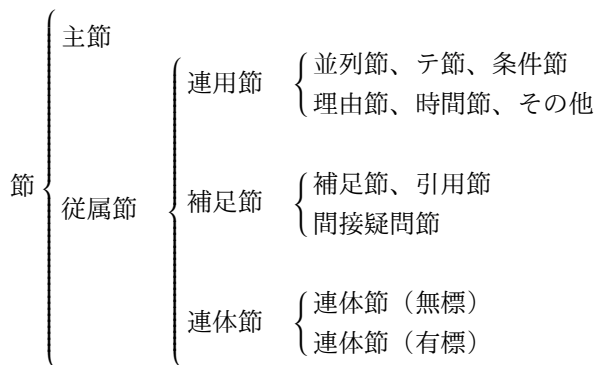


図 1: 本システムにおける節の分類体系

ラベルを拡張したものである。なお、*がついているものは、さらに助詞が後続する可能性がある場合を表す。

1. 連用節

- a. 並列節: ガ、ケド、ケレド、ケレドモ、シ*、タリ*、デハ、トカ、ヤラ
- b. テ節: テ、テカラ*、テハ、テモ
- c. 条件節: カギリ*、ケッカ、タラ、ト、トコロデ、ナラ、ナラバ、バ、バアイ*、モノノ
- d. 理由節: カラ*、タメ*、タメニ*、ノデ
- e. 時間節: アト*、アトデ、アトニ*、イマ*、イライ、トキ*、トキデ、トキニ、トキニハ、マエ*、マエニ*、その他*
- f. その他: 連用節、シニ、ダケ*、ダケニ、ツテ、ツツ*、トカ*、ドコロカ、ナガラ*、ナガラモ、ナド*、ニセヨ、ニハ、ニモ、ノニ、バカリ*、ホカ*、ホカニ、ホド*、ホドニ、マデ*、マデニ、マデニハ、マデハ、マデモ、ママ*、ママデ*、ヨウ*、ヨリ*、その他

2. 補足節

- g. 補足節: 補足節
- h. 引用節: 引用節*、ツテ、トイウ、ナンテ
- i. 間接疑問節: 間接疑問節*、カドウカ*

3. 連体節

- j. 連体節（無標）: 連体節
- k. 連体節（有標）: アトノ、カギリノ、タメノ、ダケノ、テカラノ、テノ、トキノ、トノ、ナドノ、バアイノ、ホドノ、マエノ、マデノ、ヨウナ

図 2: 節境界ラベルの分類体系

図 2 で、「1. 連用節」の下にある「f. その他」に割り当てられた「連用節」とは、動詞または判定詞の連用形による従属節に相当する。また、「f. その他」には多様な形式の節末形式が割り当てられているが、さらに未整理の接続形式が「その他」に含まれている。これには、「～する+代わりに・反面・上・以上・際・一方」などの形を含む。これらをより具体的な節境界ラベルに反映させる点は、近い将来の課題である。

2015 年末の時点で、BCCWJ のコアデータ（書籍）の長単位データベースに対して節境界解析を実施し、その付与結果をチェックした上で、規則や節境界ラベルの微調整を行なっている。2015 年度内には、国立国語研究所コーパス開発センターから、節境界ラベルの付与結果（第 1 版）を公開する予定である。

6 節境界解析で問題になる表現

最後に、節境界の認定において注意を要する表現や、問題になる点について示す。

連用節と連体節の曖昧性 まず、連用節と連体節の曖昧性という問題がある。野田 (2002) が「どの部分を節と考えるかによって、連用節になったり連体節になったりすることがある (p.11)」と指摘する問題である。

(5) 唐辛子を食べた時、涙が出た。

(5) の下線部は、時間節として認定されるべきものであるが、形態的に見れば、名詞「時」を連体節が修飾していると解釈することもできる。この場合、連体節よりも時間節を優先的に認定する必要がある。

これは一見、単純な問題のようにも見えるが、連体節の形を取って連用節を形成する要素（我々は「節末機能文節」と呼んでいる）の範囲を網羅的にリストアップすることは、かなり難しい。上記の「時」の位置には、「瞬間、途端、後、際」など多くの名詞が入り得るし、他にも「前、間、頃、うち、たび、折」などの名詞や、さらにそれらに格助詞・係助詞・副助詞などが後続する場合など、時間節を構成する節末機能文節だけでも実に多様な表現がある。また、時間節の表現をどこまで認めるかという問題もある。

- (6) a. 娘が小学生になった年、息子が生まれた。
- b. 娘が小学生になった昨年、息子が生まれた。

さらに、連用節と連体節の連続性は、時間節以外にも多く見られる。次の例では、後続する助詞の有無によって、a. は条件節、b. は連体節として区別している。

- (7) a. 協議を重ねた結果、ついに合意に達した。
- b. 協議を重ねた結果に対して、文句が出た。

形容詞による連体節の認定 形容詞が名詞を修飾する場合に、連体節として認定するか、という問題がある。

- (8) a. 高いウィスキーをいただいた。
b. 値段が高いウィスキーをいただいた。

(8)のb.では、「値段が高い」という連体節が形成されている。一方、形容詞は単独で述語になれる以上、a.の「高い」も、b.と同様に連体節と認定される可能性がある。ここで、「高い」の直後に節境界（連体節）を認めるかどうか、という点が問題となる。

寺村(1981)は、動詞や形容詞が名詞を修飾する構造について、「それが節、つまりそれだけ切りはなせば文として成り立つような内容をもつものかどうかは、外形からは判断できない(p.89)」とした上で、「お茶がほしい人」のように主体が文脈から分かること、その述語にテンスの意識があること、という2点を満たす場合にのみ「節」を認める、という立場を取っている。逆に、「やせた人」「嬉しい知らせ」のように、上記の条件を満たさない場合は、節ではなく「句」とする。

この立場に立つと、「飲むヨーグルト」「やせている男」「曲がった道」のように、対象の属性を規定する名詞修飾表現は（タ形を取っていても）連体節とは認められないことになる。しかし、このような意味的な違いを表層の単語列から判定するのは極めて難しい。

同様の問題は、(9)のb.において、「子どもの」の直後に連体節を認めるかどうか、という場合にも生じる(b.の場合、「の」は判定詞の連体形である)。

- (9) a. 子どもの時に会ったことがある。
b. 彼が子どもの時に会ったことがある。

文脈によって解釈が分かれる例 文脈によって、認定すべき節境界の解釈が分かれる場合がある。

- (10) a. 名前を呼ばれたのに気がついた。
b. 名前を呼ばれたのに気がつかなかった。

(10)のa.は、格助詞を伴った補足節が「気がついた」という述語の項になっていると解釈できる。この場合、「名前を呼ばれたことに気がついた」と言い換えられる。一方、b.の下線部は、補足節としても、逆接を表す連用節としても解釈することができる。補足節の解釈であれば「名前を呼ばれたことに気がつかなかった」と言い換えることができるが、逆接を表す連用節としてはそのような言い換えはできず、「名前を呼ばれたにも関わらず気がつかなかった」などの言い換えになる。つまり、文脈（発話意図）によって補足節と連用節の解釈が完全に曖昧になっている。

同じような曖昧性は、次の例にも見られる。

- (11) a. 太郎が帰ったところで、花子が来た。
b. 太郎が帰ったところで、花子は来たところだ。
c. 太郎が帰ったところで、花子には関係ない。

いずれも「～入ったところで」という従属節を含んでいるが、a.は補足節として、b.は並列的な連用節として、c.は譲歩を表す連用節として、それぞれ解釈される。これは、従属節と主節との関係的な意味を計算しなければ決められず、やはり文脈（この場合は主節との関係）によって解釈が分かれるケースと言える²。

形態的な指標がない場合 節境界として認定すべき箇所、形態的な指標がない場合がある。(12)のa.は部分並列、b.は「して」が省略された構文であるが、表層の単語列からこれらの境界を認定することは難しい。

- (12) a. 太郎は次郎と、花子は陽子と帰った。
b. 地図を頼りに、知らない町を歩く。

7 おわりに

本稿では、現代日本語における節の分類体系を整理した。節境界解析によってテキスト中に節境界ラベルを付与する立場から、節境界の分類案を示し、BCCWJに付与するラベルの設計について報告した。さらに、節境界解析において問題となる表現について検討した。

BCCWJに対する節境界のアノテーションについて、最終的な付与結果やその数量的な分布、解析精度などについては、稿を改めて論じることとする。

参考文献

- 池原悟(2007).「意味類型パターン記述言語仕様書」.
加納隼人・佐藤理史(2014).「日本語節境界検出プログラム Rainbow の作成と評価」.『第13回情報科学技術フォーラム(FIT2014),E-005,第2分冊』,pp.215-216.
丸山岳彦,柏岡秀紀,熊野正,田中英輝(2004).「日本語節境界検出プログラムCBAPの開発と評価」.『自然言語処理』,11(3),39-68.
丸山岳彦,高梨克也,内元清貴(2006).「第5章 節単位情報」.『日本語話し言葉コーパスの構築法』,国立国語研究所報告書124,pp.255-322.国立国語研究所.
益岡隆志・田窪行則(1992).『基礎日本語文法—改訂版—』.くろしお出版.
村木新次郎(2007).「日本語の節の類型」.『同志社女子大学 学術研究年報』,58.
野田尚史(2002).「単文・複文とテキスト」.『日本語の文法4 複文と談話』,pp.3-61.岩波書店.
佐藤理史,丸山岳彦,夏目和子(2016).「現代日本語書き言葉均衡コーパスに対する節境界付与」.『言語処理学会第21回年次大会予稿集』.言語処理学会.
寺村秀夫(1981).『日本語の文法(下)』.日本語教育指導参考書(5).国立国語研究所.

²a.とc.の「で」は格助詞、b.の「で」は判定詞の連用形だが、形態素解析の段階でこれらを正確に区別することもまた難しい。