

第5章

節単位情報

丸山岳彦・高梨克也^{*1}・内元清貴^{*2}

本章では、CSJに付与された「節単位 (Clause Unit)」という研究用情報について、その設計方針と認定手順について述べる。節単位とは、基本的に「節 (clause)」の終端境界で発話を分割することによって得られる、文法的・意味的なまとまりを備えた言語的単位である。節単位とはそもそも、「係り受け構造情報」、「重要文情報」、「談話境界情報」など談話レベルにかかわる研究用情報の付与に共通して用いる基本単位として、独自に設計された単位である。すなわち、係り受け関係を付与する範囲、重要文として抜き出す範囲、談話境界情報を付与する範囲を一元的に取りまとめるために認定された単位である。

以下、この5章では、節単位という独自の単位を設計するに至った経緯と、その設計方針について述べる。さらに、節単位情報を実際に認定するまでの手順や規則の詳細について、具体例を挙げながら示す。

5.1 話し言葉の「文」をどう捉えるか

5.1.1 書き言葉の「文」と話し言葉の「文」

「文 (sentence)」の定義には古来より諸説あるが、伝統的な見方に従うと、例えば、「書き手あるいは話し手の完結した思想の表現であり、1つ以上の単語の規則的な連結により構成される言語的単位」と規定することができる。また文法論的な見方からすると、文とは、その下位の言語的単位である「語 (word)」が文法規則に従って連結されることにより、文法的に適格な構造を持ち、かつ意味的な完結性を備えた単位として具現化したものである、ということもできる。さらに文は、その上位にある「文章・談話 (discourse)」を構成する最小の言語的単位として捉えることも可能である。

文は、語とともに、言語表現を構成する最も基本的な単位として捉えられてきた。従来の言語研究—特に20世紀初頭以降の統語論的な研究—では、文を客観的な分析対象の上限とした上で、その統語規則 (syntactic rule: 語が文に合成されるための規則や制約) や構成的意味 (compositional meaning: 構文要素と統語構造から合成的に得られる意味解釈) に関して研究が進められてきた。特に Chomsky に始まる生成文法の流れを汲む理論言語学においては、その理論が備える反証可能性 (falsifiability) を保障しなければならないという必要性から、研究対象を内省によって得られる文に限定して研究が進められてきた。母語話者の直観に基づく文法

*1 京都大学学術情報メディアセンター, 東京大学大学院情報学環

*2 情報通信研究機構

性・非文法性を問うための実験材料として用いられる「文」は、この場合、一種の抽象的・記号的な構築物として扱われていると言ってよい。言語表現を構成する文法的単位には、形態素、語、文節、節、文、段落、文章・談話などさまざまな種類や長さの単位が存在するが、文はその中でも中核的な位置を占める言語的単位として位置づけられる。

以下、この5.1.1節では、本章全体の前置きとして、「話し言葉の中から文の範囲を認定することは可能か」という問題について考えてみたい。言語における最も基本的な単位として見なされる文が、実際の話し言葉の中ではどのように存在し得るのか（存在し得ないのか）という問題について、少しばかり述べてみたい。

まず、書き言葉の文について見てみよう。大半の書き言葉では、書き手自身によって文末位置に句点が付与されるため、文の範囲を取り出すことは比較的容易であると言える。例えば(1)のような新聞記事では、文の終端が句点によって明示的に示されている。そこで、句点（あるいはテキスト冒頭）から句点までの範囲を取り出せば、文の範囲を認定することができるということになる。

- (1) 被災地からは、壊れた住宅の再建支援制度を求める声が強い。だが、「公的資金による援助は個人財産の形成につながる」との意見も強く、論議は足踏み状態だ。その中で鳥取県は昨年秋、被災者住宅再建支援基金を設け、国に先行して一步を踏み出した。政府、国会も論議を加速すべきだろう。

(毎日新聞 2002年1月17日 朝刊)

一方、話し言葉では、話し手自身が文末位置を明示的に示すということはない。もとより、句点のように文末位置を明示するために特化した装置が話し言葉には存在しない。とすると、話し言葉の中から文の範囲を取り出すためには、何らかの手がかりをもとに文の範囲を認定しなければならないということになる。例えば次のような発話の書き起こし^{*3}から文の範囲を取り出すためには、何を手がかりとしたらよいだろうか。

- (2) 今から八年前の |_p 十二月に |_p 私は |_p 友人と二人で |_p オランダのディズニーワールドに行ってきました |_p 八泊十日とちょっと長めだったのですがお金を頑張って溜めて |_p 行きました |_p ここは昔から |_p テレビで何度も見て一度でもいいから行ってみたいと強く思っていたところです |_p 東京ディズニーワールドも |_p 大好きで何度も行きました |_p...

(S01F0050)

(2)の中から文を取り出すことを考えた場合、手がかりとしてまず目に付くのは、「行ってきました」「行きました」「ところです」などの**文末表現**である。文末表現は、文の完結点を判定するための形態的な指標として考えることができる。そこで、その直後を文境界と見なし、文末表現の直後で発話を分割することにより、以下のような4つの文を取り出すことができる。

- (3) 1. 今から八年前の十二月に私は友人と二人でオランダのディズニーワールドに行ってきました
2. 八泊十日とちょっと長めだったのですがお金を頑張って溜めて行きました
3. ここは昔からテレビで何度も見て一度でもいいから行ってみたいと強く思っていたところです
4. 東京ディズニーワールドも大好きで何度も行きました

(S01F0050)

ところが話し言葉の場合、文末表現が常に安定的に得られるとは限らない。特に自発的な話し言葉では、「～

^{*3} ここでは、転記テキストに含まれる情報のうち、基本形として書き表された言語情報のみを抜き出して示す。|_pは転記基本単位の境界、つまり200ミリ秒以上のポーズが挿入されている位置を表す。「(F えー)」はフィラー、「(D ご)」は語断片を表す。

が「～けれども」などの接続助詞で発話が延々と続くことにより、最初に現れる文末表現までが非常に長くなることがある。例えば次のような例である。

- (4) 带状疱疹というのは |_p(F あの一)|_p 水ぼうそうのウイルスによって起こる病気で |_p(F え一)|_p 大概小さい頃に水ぼうそうをやった人は必ず (F あの一) 体の中に |_p 体の中のどこかに |_p 水ぼうそうのウイルスというのが残っていて |_p で (F え) ストレスだとか後凄い極度の疲れとかによって |_p 突然 (F その一) 水ぼうそうのウイルスがまた暴れ出して |_p(F あの一) |_p 発病するというものなのですが |_p 水ぼうそうの時にできる (F あの一) 水疱 |_p を伴ったぶつぶつ |_p じんま疹みたいなぶつぶつが |_p(F え) 突然できるんですけれども |_p(F え一) |_p 私の場合はですね |_p(F え一)|_p(F あの一) |_p 首の後ろの脊椎 |_p に |_p(F え) ウイルスが残っていて |_p ちょうど首の |_p この |_p 部分に |_p ここから |_p(F え一)(D な) 首胸の上から顔の顔までの間に出る |_p ということで |_p(F え一)|_p 体のですね大体片側 |_p の限られた部分に |_p(F え一) |_p その発疹が |_p 出る |_p 病気です |_p で (F え一)(F その) 神経に沿って (F え一)(D は)|_p 大体 (D は) 発疹が出るので ... (S00F0210)

(4) に現れる文末表現は、下線部の「病気です」という部分のみである。先ほどと同じように、文末表現の直後で発話を分割して文を取り出すという手続きを考えると、(4) の冒頭から「病気です」の直後までが 1 文として認定されるということになる。しかしながら、この取り出された範囲は 1 つの文としてはあまりに長く、冒頭で示した「書き手あるいは話し手の完結した思想の表現」という文の規定からはあまりに逸脱してしまっている。文末表現という形態的な指標は、文末位置を知る手がかりとしては確かに有用であるが、文末表現が現れるまでの範囲を 1 文とする文の認定方法は、話し言葉の場合、不適切な結果を生じさせることがある。

さらに話し言葉の中には、「フィラー」「語断片」「言いさし」「言い直し」「挿入」「倒置」など、通常の手書き言葉には見られないさまざまな「非流暢性 (disfluency)」に関する現象が現れることがある。そのような非流暢現象が原因となり、文の境界が不明瞭になってしまうことがある。例えば、以下の発話を見てみよう。

- (5) その |_p 泊まった部屋っていうのが |_p(D な一)|_p 何畳ぐらいかな (D けつ)|_p 何畳ぐらいになるのかな |_p(F ん)(?) もう |_p(F ん一)|_p 八畳 |_p 八畳は狭いですかね (D なん) 何だろうとにかく凄く広くて |_p(F えと一)|_p 船旅なのに |_p 普通船旅と違って |_p 二等とかでしか行ったことないからあたしにとっては衝撃だったんですが |_p もうとにかく部屋にベッドが二つあってソファもあって |_p でクロゼットとかも全部付いててバルコニーが付いていて |_p でしかも何か (F その) |_p(F ま) お風呂とかバストイレも (D べ) しっかり (D ん)|_p 部屋の中にあるみたいな |_p 感じ |_p でした ... (S01F0183)

(5) には、「フィラー」「語断片」「言いさし」「言い直し」「挿入」など、多くの非流暢現象が現れている。ここで発話中における話者の心的な過程を考えてみると、以下のようなになるだろう。話者は、「泊まった部屋」について説明をしようと発話を開始したが、途中で行き詰まり、自ら発話した内容やこれから発話する内容に対して逡巡したり、発話開始時には考えていなかった新しい内容をその場で付け加えたりしながら、最終的には「～みたいな感じでした」という文末表現で一連の説明を終えている。

具体的な非流暢現象として現れているのは、次に話す内容や既に話した内容について逡巡する発話の「挿入」(「何畳ぐらいかな」「八畳は狭いですかね」「何だろう」)、既に話した内容を新しい表現で置き換える「言い直し」(「八畳～とにかく凄く広くて」)、途中まで言いかけた内容(「船旅なのに」)を破棄して新しい発話(「普通

船旅と違って)を開始する「言いさし」などで、この結果、発話全体に構造的な乱れが生じている。

意味的には発話全体がゆるやかなまとまりを成しているように見えるものの、(5)全体が適格な構造を持つ1文であるとは言いがたい。構造的な適格性にさまざまな破綻をきたしながらも、逡巡しつつ発話を継続・展開し、自らの発話を文末表現によって何とか完結させようとした結果、最終的に(5)のような発話が形成された、と見るべきであろう。

話し言葉に見られるこのような現象について、大石(1971:255-256)には、次のような記述がある。

話しことばに不整表現が多いということも、つねに指摘されるところである。不整表現の中には、語順の混乱、首尾の不照応、話線の混交等、文の構造に関するものが少なくない。これらは要するに、話し手の表現管理に破綻が生じ、表現の統一の崩れたところに生まれた現象である。書きことばにおいては、言語行動の条件から、概してそうした破綻が少なく、また、破綻のつくろわれる余裕がある。話しことばでは、事情がこれに反して、破綻が多く、また、その結果生じた不整をおおうべきがない。

「不整をおおうべきがない」のは、話し言葉の実体が音声であるということに起因する。さらに言えば、話し言葉の本質が、実時間に沿って行われる「**発話生成** (utterance production)」という行動の結果の現れそのものであるということに起因する。そもそも発話生成とは、発話開始時から時間とともに動的に展開される一回的・個別的な行動であり、発話開始時において、発話終了時までの発話形式が厳密に決められているわけではない。話し手は、大まかな発話プランを策定した上で発話を開始し、文法的に適格な構造と意味的に完結した内容とを備えた言語形式を、実時間内に構築し、かつ線状的 (linear) に話さなければならない (Levelt 1989)。特に自発的な話し言葉では、時間をかけて編集することが可能な書き言葉と異なり、状況に応じて発話プランをその場で変更・更新したり、当初の発話プランを破棄して新しいプランに沿った発話を新たに開始したりするという、即興的な性格が強くなる。その結果、「話し手の表現管理に破綻が生じ」、(5)のような発話が形成されることになる。

このような結果生じた一まとまりの発話の中から、「文法的に適格な構造を持ち、かつ意味的な完結性を備えた単位」である文の範囲を認定するということは、本質的に矛盾した考え方であると言わざるを得ない。発話生成という一回的・個別的な行動の結果である「発話」と、一種の抽象的・記号的な構築物である「文」とは、そもそも概念として異質なものであるからである。

このことに関連して、話し言葉を収録してその「総合文型」を網羅的に記述することを試みた国立国語研究所(1960:57)には、次のような記述がある。

もし、「話しことば」を“発話”の意味に解し、“文”をその抽象的形式の意味に解するならば、“話しことばの文”ということば自体、2つの相いれない概念を結ぶという点で、矛盾を含む。“話しことば”を“発話”の意味に解することに異論はないが、“話しことばの文”とは、“発話”と“文”との、具体と抽象の間であって、“文”により近く位置する単位としての、やはり抽象的形式であり、ここでの“話しことばの”という形容の語は、“話し言葉を資料とし、ここから抽象の段階を踏むところの”の意味である。

以上、「話し言葉の中から文の範囲を認定することは可能か」という問題について考えてきた。言語表現における基本的な単位として見なされてきた文が、ある種の抽象化を経たものでなければ存在し得ないものである以上、話し言葉の中から文の範囲を認定するという考え方自体に矛盾が含まれる、ということになる。文はあくまでも抽象的・概念的な構築物であり、一方、発話は外界に具現した実在的な構築物である。話し言葉を

研究対象として扱う場合、文と発話という両者が本質的に乖離した存在であるということについて、常に自覚的でなければならないであろう。

5.1.2 談話レベルにおける基本単位の必要性

前節で述べた内容を前置きとして、ここでは、CSJの構築過程において「節単位」という情報がなぜ必要となったのか、その経緯について述べておく。

CSJには、本報告書で解説されている種々の研究用情報以外にも、次のような情報が付与されている。構文や談話のレベルに関わるこれらの情報は、主に情報通信研究機構（当時は通信総合研究所）において設計・付与されたものである。

- **係り受け構造情報** — 文節間の係り受け関係に関する情報
- **要約・重要文情報** — 講演内容の重要な部分を選択した情報
- **談話境界情報** — 話し手の意図を推測して談話構造を区分化した情報

これらの情報を実際に付与する段階においてまず問題となったのは、ある講演のうち、どこからどこまでの範囲を対象として情報を付与すればよいか、ということであった。係り受け構造情報を例にとると、どこからどこまでの範囲に対して係り受けの修飾関係を付与すべきかを最初に考える必要があった、ということである。この段階においては、転記テキストで1行ごとに配置されている**文節**が最大の文法的な単位であり、それより大きい文法的単位は存在しなかった*4。そこで、作業者間で「意味的に係り受け関係を結んでいると考えられる範囲に対して係り受け関係を付与する」と申し合わせ、作業の試行を行った。その結果、係り受け関係を付与する範囲が作業者により大きく異なってしまうという結果が生じた。ある作業者はごく狭い範囲に対してのみ係り受け関係を付与していたのに対し、ある作業者は談話的なまとまりを持つ広い範囲に対して係り受け関係を付与していた。これと同様の問題、すなわち、講演のどの範囲に対して情報を付与すればよいかという問題は、重要文情報・談話境界情報の付与作業においても生じると考えられた。

ここから、構文レベル・談話レベルに関わる情報の付与作業に共通する分析用の基本単位の確定することが先決である、という結論に達した。すなわち、係り受け関係を付与する範囲、重要文として抜き出す範囲、談話構造を構成する範囲を一元的に取り扱うための共通単位をあらかじめ認定しておき、その単位をもとに各作業を進める、という方針を立てた。問題は、そのような基本単位のどのようにして認定するかということになった。

この場合、基本単位として最も一般的に考えられるのは、「文」という単位である。しかしながら、先に述べたように、自発的な話し言葉の中から文という単位を取り出すことには本質的な矛盾が含まれる。連綿と続く話し言葉の中から、文法的・意味的なまとまりを備えた範囲を取り出し、言語的な単位として認定することを考えた場合、どのような単位が想定されるであろうか。

ちょうど同じ時期に、本章の第一筆者が所属していたATR音声言語コミュニケーション研究所第4研究室（当時）において、「**節境界検出プログラム CBAP**（Clause Boundary Annotation Program）」というプログラムが開発されていた（丸山他 2004）。このプログラムは、形態素解析された日本語のテキストから「**節境界**（clause boundary）」の位置と種類を自動的に検出するという、節境界解析の機能を持っていた。例えば、

*4 200 ミリ秒以上のポーズで区切られる転記基本単位は文節よりも大きな単位を作ることが多いが、それらは必ずしも文法的な単位を構成するものではない。

(6) のようなテキストを入力すると、節境界の位置と種類を特定し、(7) のように「節境界ラベル」を付与した結果を出力する。/連体節/や/文末/などが節境界ラベルである*5。

(6) 三十一日のニューヨーク外国為替市場は、ロンドン市場で円高が進んだ流れを引き継いで円を買う勢いが強まり、円は一時一ドル百二円五銭まで円高ドル安が進みました。円が一ドル百二円台で取引されるのは、今年初め以来ほぼ三ヶ月ぶりのことです。

(7) 三十一日のニューヨーク外国為替市場は、/主題ハ/ ロンドン市場で円高が進んだ/連体節/ 流れを引き継いで/テ節/ 円を買う/連体節/ 勢いが強まり、/連用節/ 円は一時一ドル百二円五銭まで円高ドル安が進みました。/文末/ 円が一ドル百二円台で取引されるのは、/補足節/ 今年初め以来ほぼ三ヶ月ぶりのことです。/文末/

同研究室では、ニュース解説番組や新聞記事など長文を多く含むテキストを対象に、あらかじめ節境界解析を行った上で、ある特定の節境界の直後で文分割（発話分割）を行うことにより、長文が効果的に分割されることが確かめられていた（丸山他 2004）。そこで、この CBAP を用いて CSJ に含まれる発話を分割し、基本単位の自動的な抽出を試みることにした。ただし CBAP は、日本語形態素解析システム「茶釜*6」で解析された結果を入力形式としていたため、異なる仕様の形態論情報を持つ CSJ に対して、そのままの形では適用できなかった。そこで、CSJ の形態論情報の仕様に沿って CBAP を改編し、新たに「節境界検出プログラム CBAP-csj」を開発することにした。

5.1.3 節単位が認定されるまでの流れ

新たに開発された CBAP-csj は、CSJ に付与されている形態論情報（主に短単位）を手がかりとして節境界の位置と種類を検出し、その文法的属性を表示する機能を持っている。例えば先の (2) に挙げた例に対しては、(8) のような結果を出力する*7。

(8) 1. 今から八年前の十二月に私は友人と二人でオランダのディズニーワールドに行ってきました
[文末]
2. 八泊十日とちょっと長めだったのですが/並列節ガ/
3. お金を頑張って<テ節>溜めて<テ節>行きました [文末]
4. ここは昔からテレビで何度も見て<テ節>一度でもいいから<理由節カラ>行ってみたいと
<引用節>強く思っていたところです [文末]
5. 東京ディズニーワールドも大好きで<並列節デ>何度も行きました [文末] (S01F0050)

CBAP-csj によって検出された節境界の直後には、CBAP と同様、その節の種類に応じて「節境界ラベル」が自動的に付与される。[文末]、/並列節ガ/、<テ節> などが節境界ラベルである。オリジナルの CBAP と異

*5 CBAP が検出する対象は、節境界だけでなく、主題要素なども含まれる。丸山他 (2004) 参照。

*6 <http://chasen.aist-nara.ac.jp/>

*7 以降、CBAP-csj の出力結果を示す場合、フィラーや語断片などの情報は省略して示す。

なるのは、発話分割処理に利用するという目的を第一に考えて、全ての節境界を検出するのではなく、発話分割処理に関連する節境界のみを検出対象としたこと*8、そして、その節境界直後の発話の切れ目の大きさという観点から、節境界の種類に3段階のレベルを設けたこと、の2点である。これらの理論的背景や具体的な仕様については、5.2節以降で詳しく述べる。

CBAP-csjは、節境界の位置と種類を検出すると同時に、[]および/ /で囲まれた節境界の直後で発話を分割する。上記の例で言えば、[文末] および/並列節ガ/の直後が分割位置となり、発話が5つの部分に分割されている。

特定の節境界の直後で発話を分割するという方針を採ったことにより、接続助詞などで発話が延々と続くような場合も、文末の出現を待たずに発話を分割することが可能になった。先に挙げた(4)は、CBAP-csjによって次のように分割される。

- (9)
1. 帯状疱疹というのは水ぼうそうのウイルスによって<テ節>起こる病気で<並列節デ>大概小さい頃に水ぼうそうをやった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/
 2. で<接続詞>ストレスだとか<ト力節>後凄い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/並列節ガ/
 3. 水ぼうそうの時にできる水疱を伴ったぶつぶつじんま疹みたいなぶつぶつが突然できるんですけれども/並列節ケレドモ/
 4. 私の場合はですね首の後ろの脊椎にウイルスが残っていて<テ節>ちょうど首のこの部分にここから首胸の上から顔の顔までの間に出るという<トイウ節>ことで<並列節デ>体のですね大体片側の限られた部分にその発疹が出る病気です [文末]
 5. で<接続詞>神経に沿って<テ節>大体発疹が出るので<理由節ノデ>... (S00F0210)

しかしながら、特定の節境界の直後で発話を分割する処理だけでは、結果として不適切・不自然な分割結果しか得られないことがある。特に、非流暢性の影響によって、不適切な分割結果が生じてしまうことがある。例えば、(10)のような発話をCBAP-csjで分割すると、(11)のような形が出力される。

- (10) テーマ |_p 今まで人生を振り返って |_p 印象に残ったこと |_p タイトルは長男の嫁として |_p 長男に嫁いだのが |_p 二十四歳の時でした |_p 夫も二十四歳 |_p 母 (F あの一) おしゅうとめさんが四十八歳二回り上の |_p で |_p (F (?うーん)) |_p (D おし) (D う) しゅうとが |_p 五十一歳です |_p それでとても封建的な |_p 時代 (D (?つ)) |_p だったのでとても苦労しました |_p (F あ一の一) |_p |_p 母 (D (?お)) (F あの一) |_p おしゅうとめさんはとても血圧が高い人で |_p 朝から |_p ずっと寝てました |_p |_p 食事も全部嫁 |_p 洗濯も嫁 |_p (F あの一) おしゅうとさんが |_p (F あの一) 会社に行くのも |_p 嫁の私が全部食事の支度をして |_p ... (S05F1517)

*8 CBAPが検出対象とする節境界は144種類、CBAP-csjが検出対象とする節境界は49種類である。

- (11) 1. テーマ今まで人生を振り返って<テ節>印象に残ったことタイトルは長男の嫁として<テ節>長男に嫁いだのが二十四歳の時でした [文末]
 2. 夫も二十四歳母おしゅうとめさんが四十八歳二回り上ので<接続詞>しゅうとが五十一歳です [文末]
 3. それで<接続詞 L>とても封建的な時代だったので<理由節ノデ>とても苦労しました [文末]
 4. 母おしゅうとめさんはとても血圧が高い人で<並列節デ>朝からずっと寝てました [文末]
 5. 食事も全部嫁洗濯も嫁おしゅうとさんが会社に行くのも嫁の私が全部食事の支度をして<テ節>... (S05F1517)

この例の音声を聞くと、「テーマ」「今まで人生を振り返って印象に残ったこと」「タイトルは長男の嫁として」はそれぞれ独立した発話として成り立っていることが分かる。意味としてのまとまりという点からすると、それぞれの直後で発話が分割されることが望ましい。また、「夫も二十四歳」「食事も全部嫁」「洗濯も嫁」などはいわゆる「体言止」の構造を取る表現であり、やはりその直後で分割されることが望ましい。さらに、「母おしゅうとめさんが四十八歳二回り上の」「二回り上の」は、直前の「四十八歳」を修飾する要素であり、倒置の構造となっている。しかしながら、CBAP-csj がこれらの点を発話の分割位置として適切に検出することは非常に困難である。

そこで、作業者が CBAP-csj の自動分割結果をチェックし、不適切な分割結果が生じている部分に対しては人手で修正を施すことにした。例えば (11) の例は、人手でチェックされた後、以下のような形に修正される。

- (12) 1. テーマ - ; 体言止
 2. 今まで人生を振り返って<テ節>印象に残ったこと - ; 体言止
 3. タイトルは長男の嫁として<テ節> - ; 体言止
 4. 長男に嫁いだのが二十四歳の時でした [文末]
 5. 夫も二十四歳 - ; 体言止
 6. 母おしゅうとめさんが四十八歳 <<二回り上の>> - ; 倒置-後切り
 7. で<接続詞>しゅうとが五十一歳です [文末]
 8. それで<接続詞 L>とても封建的な時代だったので<理由節ノデ>とても苦労しました [文末]
 9. 母おしゅうとめさんはとても血圧が高い人で<並列節デ>朝からずっと寝てました [文末]
 10. 食事も全部嫁 - ; 体言止
 11. 洗濯も嫁 - ; 体言止
 12. おしゅうとさんが会社に行くのも嫁の私が全部食事の支度をして<テ節>... (S05F1517)

CBAP-csj の出力では分割されていないが、文法的・意味的なまとまりを備えた言語的単位としては分割されるべきであると判断された点には、「-」記号を人手で付与し、その直後を切断した。また、倒置要素であると判断された要素は「<< >>」で囲むことにした。さらに、どのような理由でそのような操作を行ったのかを、「; 体言止」のようにコメントとして記述することにした。その他、係り受け構造情報、重要文情報、談話境界情報などの付与作業における必要性などを考慮して修正すべき項目を定義し、「人手修正規則」を定めた。この規則に従って人手による修正作業を行い、最終的な分割結果を得た。

以上のような手順に従って、CBAP-csj で自動発話分割を行い、その結果を必要に応じて人手で修正した最終的な分割結果を「節単位 (Clause Unit)」と呼び、各情報付与作業に共通して用いる単位とすることにした。節単位認定までの流れを示すと、図 5.1 のようになる。

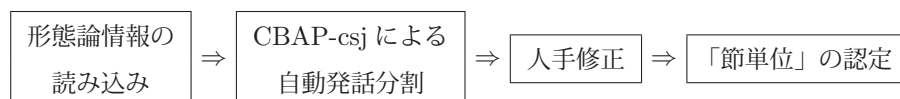


図 5.1 節単位が認定されるまでの流れ

これによって認定される節単位とは、以下のように定義される。

節単位とは、話し言葉における「文」に相当する単位として、係り受け構造情報、重要文情報、談話境界情報など構文・談話レベルの情報を付与するための基本単位として想定された、文法的・意味的なまとまりを備えた言語的単位を指す。基本的には特定の節境界の直後で発話を分割することによって得られる節 (述語を中心としたまとまり) が節単位を構成するが、非流暢現象などが原因となり、単独の名詞句や言いさされた部分、発話の断片などが節単位を構成することもある。

節単位は、ある特定の節境界を分割点とするという点で、基本的には文法的な性格を帯びた単位である。CSJ に付与された文法的な単位のうち、節単位は、短単位、長単位、文節に続いて大きく、かつ、1つの講演を構成する最大の文法的単位である。CSJ に含まれる文法的単位の階層を示すと、図 5.2 のようになる。

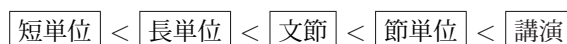


図 5.2 CSJ に含まれる文法的単位の階層

以上、「節単位」という独自の単位を認定することになった経緯と、節単位を認定するまでの一連の手順について述べた。以降、5.2 節と 5.3 節では CBAP-csj による自動発話分割処理のアルゴリズムについて詳細に述べる。続く 5.4 節と 5.5 節では、CBAP-csj の自動発話分割結果を人手修正するための規則について述べる。

5.2 CBAP-csj による自動発話分割処理

5.2.1 理論的背景

話し言葉の中から文という抽象的構築物を取り出そうとすることに本質的な矛盾が含まれるということは、冒頭に述べた。その一方で、話し言葉を分析対象として扱う場合、発話をあらかじめ何らかの分析用の単位に分割しておくことが必要であることもまた事実である。

これと同様の問題意識から、話し言葉を対象とした従来の研究においては、文に代わるさまざまな言語的単位が提唱されてきた。例えば、以下のようなものである。

- “C-Unit (Communication Unit)” (Loban 1966)
- “TCU (Turn Constructional Unit)” (Sacks et.al. 1974)
- “Intonation Unit” (Chafe 1987)
- “Utterance” (Crookes 1990)
- “AS-Unit (The Analysis of Speech Unit)” (Foster et.al. 2000)

このうち Utterance と Intonation Unit は、基本的にポーズやイントネーションなどの音声的な情報を手がかりとして発話を分割し、得られる単位である。ただし、自発的な発話ではポーズがさまざまな位置に生起し得るため、独話の分析単位として適切な分割結果が得られる可能性は不明である。TCU は、「ターン」を構成する単位を指す。「ターン（発話順番）」は、メイナード泉子（1993:56）によると「会話において一人の話者が話す権利を行使するその会話中の単位で、会話の当事者によりその何らかの意味または機能を持っていると認められたもの」と定義され、談話分析・会話分析の研究においてよく用いられるが、独話の発話単位としてどれだけ利用できるかは定かではない。C-Unit とそれを改良した AS-Unit は、表層に現れている統語的な特徴、特に「節（clause）」の境界を分割位置として得られる文法的な単位（syntactic unit）である。CSJ における節単位は、設計方針という点では C-Unit や AS-Unit に近い。

さて、節境界を利用して発話のある種の単位に分割することを考えた場合、その構造的な特性から、日本語は英語などと比べて比較的有利な立場にあると言ってよい。まず、日本語では述語句の部分が形態的によく発達しており、表層の形態素列から豊富な文法情報を獲得することができる。さらに日本語の述語句は常に節の終端に配置されるため、述語の活用形や接続助詞の種類などを手がかりとして、節の終端境界の位置およびその文法的な特性をかなり正確に把握することができる。発話の中から節境界の位置と種類を検出し、その形態的特徴や文法的特性に応じて分割位置を定めることにより、節境界を基本とする柔軟な分割処理を実現することができる。

そもそも、ある種の節境界が統語的に大きな切れ目として文の分割位置になり得るということは、日本語の記述的文法研究の中でも早い時期から指摘されてきた。例えば、従属節の種類とその従属度との対応関係を論じた三上（1953,1959,1963）や南（1974,1993）などは、その代表である。三上（1953:182）は、「代表的な活用形の平均的陳述度を分数値で表し」たもの、すなわち従属節の種類と文中の切れ目の大きさとの関係として、表 5.1 に示すような対応を示している。

表 5.1 従属節の種類と切れ目の大きさの対応

中止連用形	何々シ（テ）	1/4
連体形	何々シタ	1/2
仮定形	何々スレバ	3/4
終止形	何々シタ、セヨ	1

さらに、「連用補語を食止めるか否か」「連体として収まるか否か」「普通体を丁寧体に変更することがふさわしいか否か」という統語的テストによって、従属節（三上の言う「活用形」）の各種類を「単式」「軟式」「硬式」という3つのクラスに割り当てている。これらのクラスは「句切りの力」の違いに対応づけられており、後に述べる南（1974,1993）による従属節の分類に引き継がれていくことになる。

次に、南（1974,1993）は、従属節の内部に現れ得る要素の範囲、および従属節相互の包含可能性が従属節の種類によって異なるという統語的な事実に着目し、従属節が備える「文らしさ」という観点から、従属節を「A類」「B類」「C類」という3つのクラスに分類した。

A類: つつ、て（様態）、ながら（継続）、連用形反復…

B類: て（理由、原因）、と、ながら（逆接）、ので、連用形…

C類: が、から、けれど、し、て、連用形…

例えば、「*食べたつつ」は非文法的だが「食べたので」「食べたが」は文法的である、また「*食べただろうので」は非文法的だが「食べただろうが」は文法的である、というように、A 類から C 類に進むほど従属節内部に現れ得る要素の範囲は広くなり、それに応じて主節への従属度が低く（主節からの独立度が高く）なる。つまり、C 類に近づくほど従属節に「文らしさ」が備わっていくということであり、そこに属する節境界を文の分割位置として利用できる可能性が高くなることを示している。

三上や南によるこれらの記述は、従属節の形態ごとに切れ目の大きさ（従属度）が対応付けられている点で、発話分割処理にそのまま応用することができる。節境界の位置と種類を特定できれば、その直後の切れ目の大きさを推測し、発話分割位置と見なせるかどうかを決定することができるわけである。

節境界検出プログラム CBAP、CBAP-csj の開発^{*9}は、以上のような理論的考察が背景となっている。以降では、CBAP-csj が節境界を検出するアルゴリズム、および動作の仕様を具体的に解説していくことにする。

5.2.2 節境界の検出アルゴリズム

まず、CBAP-csj が節境界を検出するアルゴリズムについて述べる。CBAP-csj は、節境界をパターンマッチによって検出するための規則を記述した「節境界検出ルール」と、それを実行するためのプログラム本体とから構成されている。「節境界検出ルール」は、節境界の位置と種類を特定するための「形態素列パターン」と、節境界の名称を示す「節境界ラベル」を含む、292 個の規則の集合（テキストファイル）である。プログラム本体の実装は Perl で行った。

CBAP-csj は形態論情報を読み込み、パターンマッチによって節境界の位置を特定し、節境界の直後の位置にその種類を表す節境界ラベルを挿入する。CBAP-csj への入力には、「MOR 形式」と呼ばれる独自の形式を用いた。MOR 形式とは、形態論情報として付与されている短単位から「基本形」「品詞」「その他の情報 1」「活用形」「活用の種類」「その他の情報 2」という 6 つの情報を取り出し、1 つの形態素を以下のような形式で表現したものである^{*10}。

基本形_品詞-その他の情報 1_活用形_活用の種類-その他の情報 2

例えば、「被ったけれども」という表現は、MOR 形式では次のように表現される。

被っ_動詞_ラ行五段_連用形_促音便 た_助動詞_終止形 けれど_助詞-接続助詞_ も_助詞-副助詞_

図 5.3 MOR 形式による「被ったけれども」

(2) で示した例の冒頭部分を MOR 形式で示すと、次のようになる。

^{*9} なお、「CBAP」および「CBAP-csj」の著作権は、株式会社国際電気通信基礎技術研究所 ATR 音声言語コミュニケーション研究所が有しており、研究目的に限り無償で提供されている。

^{*10} 例えば助詞類の活用形のように、情報が付与されていない場合は、値なしとして表現される。なお、以降では MOR 形式で表される 1 つの単位を便宜的に「形態素」と呼ぶことにする。

(13) 今_名詞_ から_助詞-格助詞_ 八_名詞-数詞_ 年_接尾辞_ 前_名詞_ の_助詞-格助詞_ 十_名詞-数詞_ 二_名詞-数詞_ 月_接尾辞_ に_助詞-格助詞_ 私_代名詞_ は_助詞-係助詞_ 友人_名詞_ と_助詞-格助詞_ 二人_名詞_ で_助詞-格助詞_ オーランド_名詞-固有名詞_ の_助詞-格助詞_ ディズニー_名詞-固有名詞_ ワールド_名詞_ に_助詞-格助詞_ 行っ_動詞-カ行五段_連用形_ て_助詞-接続助詞_ き_動詞-カ行変格_連用形_ まし_助動詞_連用形_ た_助動詞_終止形_ 八_名詞-数詞_ 泊_接尾辞_ 十_名詞-数詞_ 日_接尾辞_ と_助詞-格助詞_ ちょっと_副詞_ 長_形容詞-形容詞型_語幹_ め_接尾辞_ だっ_助動詞_連用形_ た_助動詞_連体形_ の_助詞-準体助詞_ です_助動詞_終止形_ が_助詞-接続助詞_ ...

入力された MOR 形式は、節境界検出ルール内の形態素列パターンと照合される。もし入力中に形態素列パターンにマッチする表現があると、節境界検出ルールが定める当該の箇所に節境界ラベルが挿入される。1 個の節境界検出ルールは IN 行と OUT 行という 2 行の組から構成されており、IN 行ではパターンマッチを行うための形態素列パターンが、OUT 行ではマッチした形態素列に対する操作が指定されている。例として、**/並列節ガ/**、**<条件節タラ>** という節境界ラベルを付与するための節境界検出ルールを、図 5.4 に示す。

IN: (が_助詞-接続助詞_);
 OUT: \$1 /並列節ガ/;
 IN: ((たら|だら)_助動詞_仮定形);
 OUT: \$1 <条件節タラ>;

図 5.4 /並列節ガ/、<条件節タラ> を付与するための節境界検出ルール

図 5.4 の上段のルールでは、入力された MOR 形式中に「が_助詞-接続助詞_」という 1 形態素からなる形態素列パターンを発見したらその直後に **/並列節ガ/** という節境界ラベルを挿入することが指定されている。下段では、「たら_助動詞_仮定形」もしくは「だら_助動詞_仮定形」という 1 形態素からなる形態素列パターンを発見したらその直後に **<条件節タラ>** という節境界ラベルを挿入することが指定されている。

(13) から節境界の検出を行った結果を以下に示す。下線部は、形態素列パターンにマッチした部分を示す。

(14) 今_名詞_ から_助詞-格助詞_ 八_名詞-数詞_ 年_接尾辞_ 前_名詞_ の_助詞-格助詞_ 十_名詞-数詞_ 二_名詞-数詞_ 月_接尾辞_ に_助詞-格助詞_ 私_代名詞_ は_助詞-係助詞_ 友人_名詞_ と_助詞-格助詞_ 二人_名詞_ で_助詞-格助詞_ オーランド_名詞-固有名詞_ の_助詞-格助詞_ ディズニー_名詞-固有名詞_ ワールド_名詞_ に_助詞-格助詞_ 行っ_動詞-カ行五段_連用形_ て_助詞-接続助詞_ き_動詞-カ行変格_連用形_ まし_助動詞_連用形_ た_助動詞_終止形 [文末] 八_名詞-数詞_ 泊_接尾辞_ 十_名詞-数詞_ 日_接尾辞_ と_助詞-格助詞_ ちょっと_副詞_ 長_形容詞-形容詞型_語幹_ め_接尾辞_ だっ_助動詞_連用形_ た_助動詞_連体形_ の_助詞-準体助詞_ です_助動詞_終止形_ が_助詞-接続助詞_ = /並列節ガ/ ...

5.2.3 CBAP-csj が検出する節境界の種類

次に、CBAP-csj が検出する節境界の種類について述べる。CBAP-csj の設計段階において、寺村 (1981) や益岡・田窪 (1992) などの記述文法書で記述されている従属節の種類を参考にして、49 種類の節境界を準備した。また、付与された節境界ラベルを発話分割処理に利用することを念頭に置き、その節境界直後の構造的な切れ目の大きさという観点から、節境界に次の 3 段階のレベルを設定した。

- 絶対境界 (Absolute Boundary)
- 強境界 (Strong Boundary)
- 弱境界 (Weak Boundary)

絶対境界とはいわゆる文末表現に相当する境界であり、その直後には発話の完全な切れ目があると見なされる。節境界ラベルは [文末] のように表示される。強境界とは南 (1974) の言う C 類の従属節にほぼ相当するもので、いわゆる文末表現ではないが、構造的に大きな切れ目を成すと考えられる節境界である。「並列節ガ」や「並列節ケレドモ」などがこれに相当する。節境界ラベルは /並列節ガ/ のように表示される。弱境界とは、通常は発話の切れ目になることはないが、稀に発話の切れ目になり得る従属節の境界である。「理由節ノデ」「条件節タラ」などがこれに相当する。節境界ラベルは <理由節ノデ> のように表示される。これら 3 段階のレベルは、先の三上や南らによる従属節の従属度に関する知見を参考にして、49 種類の節境界すべてに対して個別に割り当てた。

これに加えて、検出された節境界ラベルのうち、絶対境界および強境界の直後を発話の分割位置として採用することにした。発話の完全な切れ目である絶対境界と、従属度が相対的に低い強境界の 2 種類を分割位置とすることにより、発話の構造的な切れ目の大きさという文法的特性に基づく発話分割が可能になると判断した。

CBAP-csj で検出される 49 種類の節境界ラベルを、図 5.5 に挙げる。

- 絶対境界： [文末], [文末候補], [と文末]
- 強境界： /並列節ガ/, /並列節ケレドモ/, /並列節ケレド/, /並列節ケドモ/, /並列節ケド/, /並列節シ/, /ヨウニ節/
- 弱境界： <条件節タラ>, <条件節タラバ>, <条件節ト>, <条件節ナラ>, <条件節ナラバ>, <条件節レバ>, <理由節カラ>, <理由節カラニハ>, <理由節カラ-助詞>, <理由節ノデ>, <タリ節>, <タリ節-助詞>, <テ節>, <テハ節>, <テモ節>, <テカラ節>, <テカラ節-助詞>, <テ節-助詞>, <トカ節>, <トカ節-助詞>, <ノニ節>, <連用節>, <引用節>, <引用節-助詞>, <引用節トノ>, <トイウ節>, <間接疑問節>, <間接疑問節-助詞>, <連体節テノ>, <並列節ダノ>, <並列節デ>, <並列節ナリ>, <フィラー文>, <感動詞>, <接続詞>, <接続詞 C>, <接続詞 L>, <接続詞 CL>, <接続詞 M>

図 5.5 CBAP-csj で検出される 49 種類の節境界ラベル

節境界のラベル名は、その形態的特徴、および節の前後で結ばれる関係的意味の違いを考慮して、「並列節ケレドモ」「理由節カラ」「トカ節」「引用節」のように表現した*11。それぞれの節境界ラベルごとの形式的・意味的な特徴や具体例については、5.3節で一覧として示す。

さらに、これら49種類の節境界ラベルを付与する節境界検出ルールとは別に、出現文脈に応じて節境界ラベルに付与されたレベルを強制的に書き換える「メタ規則」を導入した。例えば、弱境界の直後に接続詞が出現した場合、その境界を強境界に書き換える、というメタ規則がある。図5.6は、一度弱境界のラベルが付与された後、その直後に接続詞が出現している場合に、ラベルを強境界に書き換えるメタ規則である。

```
IN: <(.*?)> ((.*?)_接続詞_);
OUT: /$1/ $2 ;
```

図 5.6 直後に接続詞が出現した場合に、弱境界を強境界に書き換えるメタ規則

例えば「検査をして」という表現はその直後に弱境界の<テ節>という節境界ラベルが挿入されるのが原則であるが、次の例のようにその直後に接続詞が出現する場合には、メタ規則によりラベルが強境界の/テ節/に書き換えられ、そこが発話分割位置となる。

- (15) 1. で<接続詞>耳鼻科の有名な病院を紹介してもらいました [文末]
 2. で<接続詞>そこに診察に行つて<テ節>子供の頃からの状況とか後聴力の検査をして/テ節/
 3. で<接続詞>手術をすれば<条件節レバ>回復の見込みがありますよという<トイウ節>ことを伝えられました [文末] (S00M0053)

また、弱境界の<テ節>と<条件節ト>に限り、その節を構成する述語句がデスマス体・ゴザイマス体になっている場合は、その境界を強境界に変更するというメタ規則を設けた。

```
IN: ((て | で)_助詞-接続助詞_);
OUT: $1 <テ節> ;

IN: ((でし | まし)_助動詞_連用形 て_助詞-接続助詞_) <テ節> ;
OUT: $1 /テ節/ ;

IN: (ませ_助動詞_未然形 ん_助動詞_終止形 で_助詞-接続助詞_) <テ節> ;
OUT: $1 /テ節/ ;
```

図 5.7 <テ節>を付与するための節境界検出ルールと、デスマス体の場合に/テ節/に書き換えるメタ規則

*11 なお、<感動詞>から<接続詞 M>までの6種は、実質的には節境界ではないが、すぐ後に述べる「メタ規則」の適用などに関わるものであるため、他の節境界ラベルと同様に検出対象としている。接続詞の各種類については、5.3.3.26節を参照。

図 5.7 の上段のルールにより、全ての「て_助詞-接続助詞_」または「で_助詞-接続助詞_」の直後に<テ節>というラベルが挿入される。その上で、もしその述語句が「でし_助動詞_連用形 て_助詞-接続助詞_」「まし_助動詞_連用形 て_助詞-接続助詞_」「ませ_助動詞_未然形 ん_助動詞_終止形 で_助詞-接続助詞_」というデスマス体・ゴザイマス体の表現であった場合、中下段のメタ規則により、<テ節>が強境界の/テ節/に書き換えられることになる。

<条件節ト>を付与する節境界検出ルール、およびデスマス体・ゴザイマス体の場合に強境界に書き換えるメタ規則を示すと、以下のようになる。

```
IN:   (と_助詞-接続助詞_) ;
OUT:  $1 <条件節ト> ;
IN:   ((です|ます)_助動詞_終止形 と_助詞-接続助詞_) <条件節ト> ;
OUT:  $1 /条件節ト/ ;
```

図 5.8 <条件節ト>を付与するための節境界検出ルールと、デスマス体の場合に/条件節ト/に書き換えるメタ規則

次の例では「言い伝えますと」の直後が強境界の/条件節ト/に書き換えられ、発話分割位置となっている。

- (16) 1. 父親が動くと<条件節ト>とてもお腹が苦しいので<理由節ノデ>お父さんが動くと<条件節ト>
苦しいよと<引用節>言い伝えますと/条件節ト/
2. 父は極力動かないようになって<テ節>なってきました [文末] (S06M0895)

5.2.4 コアに含まれる学会講演・模擬講演の節境界ラベル検出結果

コアに含まれる学会講演 70 講演および模擬講演 107 講演を CBAP-csj で解析した結果として、絶対境界・強境界・弱境界の出現頻度を表 5.2 に示す。また、出現した節境界ラベルを頻度順に集計した結果を表 5.3 に示す。

表 5.2 177 講演に現れた絶対境界・強境界・弱境界

	学会講演 (70 講演)		模擬講演 (107 講演)	
絶対境界	5,744	(28.15%)	4,839	(18.93%)
強境界	2,829	(13.86%)	4,644	(18.16%)
弱境界	11,835	(57.99%)	16,086	(62.91%)
合計	20,408	(100%)	25,569	(100%)

表 5.2, 表 5.3 の節境界ラベルの分布を見ると、絶対境界の割合が学会講演では 28.15%, 模擬講演では 18.93% と、大きな開きがあることが分かる。このうち、[文末] の割合は学会講演で 25.91%, 模擬講演で 15.86% となっており、文末表現が学会講演では現れやすく、模擬講演ではなかなか現れないという傾向がうかがえる。その他、検出された節境界ラベルを用いることにより、さまざまな分析が可能になると考えられる。

表 5.3 学会講演 70 講演および模擬講演 107 講演に対して付与された節境界ラベル (頻度順)

表 5.3.1: 学会講演 (70 講演)

[文末]	5,288 (25.91%)
<テ節>	2,684 (13.15%)
<接続詞>	2,653 (13.00%)
<トイウ節>	1,568 (7.68%)
<引用節>	1,148 (5.63%)
/並列節ガ/	1,148 (5.63%)
<接続詞 L>	750 (3.68%)
<連用節>	560 (2.74%)
/テ節/	512 (2.51%)
<並列節デ>	447 (2.19%)
/並列節ケレドモ/	433 (2.12%)
<理由節ノデ>	360 (1.76%)
/条件節ト/	334 (1.64%)
<条件節ト>	327 (1.60%)
[文末候補]	322 (1.58%)
<テハ節>	229 (1.12%)
<条件節レバ>	183 (0.90%)
<トカ節>	158 (0.77%)
[と文末]	134 (0.66%)
<テモ節>	129 (0.63%)
/並列節ケドモ/	127 (0.62%)
<理由節カラ>	94 (0.46%)
<間接疑問節>	83 (0.41%)
<連体節テノ>	75 (0.37%)
/並列節シ/	70 (0.34%)
<間接疑問節-助詞>	65 (0.32%)
<条件節タラ>	64 (0.31%)
<感動詞>	59 (0.29%)
<タリ節>	58 (0.28%)
/並列節ケド/	51 (0.25%)
<文末候補>	49 (0.24%)
/ヨウニ節/	48 (0.24%)
<ノニ節>	36 (0.18%)
/間接疑問節/	28 (0.14%)
/連用節/	20 (0.10%)
/並列節デ/	13 (0.06%)
<フィラー文>	12 (0.06%)
/並列節ケレド/	12 (0.06%)
<テ節-助詞>	10 (0.05%)
/感動詞/	10 (0.05%)
/理由節ノデ/	9 (0.04%)
<引用節-助詞>	8 (0.04%)

<条件節タラバ>	7 (0.03%)
<テカラ節>	7 (0.03%)
/タリ節/	6 (0.03%)
<接続詞 C>	5 (0.02%)
/理由節カラ/	4 (0.02%)
<条件節ナラバ>	2 (0.01%)
<条件節ナラ>	2 (0.01%)
<理由節カラニハ>	1 (0.01%)
<並列節ダノ>	1 (0.01%)
<トカ節-助詞>	1 (0.01%)
/文末候補/	1 (0.01%)
/フィラー文/	1 (0.01%)
/テモ節/	1 (0.01%)
/テハ節/	1 (0.01%)

学会講演 合計 20,408(100%)

表 5.3.2: 模擬講演 (107 講演)

[文末]	4,054 (15.86%)
<テ節>	3,491 (13.65%)
<接続詞>	2,705 (10.58%)
<引用節>	1,830 (7.16%)
<トイウ節>	1,320 (5.16%)
/テ節/	1,145 (4.48%)
<接続詞 L>	958 (3.75%)
<並列節デ>	938 (3.67%)
/並列節ケレドモ/	884 (3.46%)
<理由節ノデ>	834 (3.26%)
/並列節ガ/	732 (2.86%)
/並列節ケド/	707 (2.77%)
[文末候補]	677 (2.65%)
<条件節ト>	637 (2.49%)
<トカ節>	412 (1.61%)
/並列節ケドモ/	395 (1.54%)
<条件節タラ>	379 (1.48%)
<理由節カラ>	354 (1.38%)
<タリ節>	322 (1.26%)
<テモ節>	290 (1.13%)
<文末候補>	270 (1.06%)
<連用節>	266 (1.04%)
<条件節レバ>	253 (0.99%)
/並列節シ/	248 (0.97%)
/条件節ト/	187 (0.73%)

<テハ節>	132 (0.52%)
<感動詞>	130 (0.51%)
<間接疑問節>	126 (0.49%)
[と文末]	108 (0.42%)
<フィラー文>	102 (0.40%)
/並列節デ/	92 (0.36%)
<テカラ節>	86 (0.34%)
<ノニ節>	77 (0.30%)
/並列節ケレド/	51 (0.20%)
/理由節ノデ/	46 (0.18%)
/感動詞/	45 (0.18%)
<間接疑問節-助詞>	39 (0.15%)
<連体節テノ>	35 (0.14%)
/連用節/	21 (0.08%)
/理由節カラ/	20 (0.08%)
<タリ節-助詞>	19 (0.07%)
<テカラ節-助詞>	16 (0.06%)
/引用節/	13 (0.05%)
/タリ節/	13 (0.05%)
<テ節-助詞>	12 (0.05%)
/ヨウニ節/	10 (0.04%)
<引用節-助詞>	9 (0.04%)
/フィラー文/	9 (0.04%)
<条件節ナラ>	8 (0.03%)
/条件節タラ/	8 (0.03%)
<接続詞 C>	7 (0.03%)
<トカ節-助詞>	7 (0.03%)
/文末候補/	7 (0.03%)
<理由節-助詞>	6 (0.02%)
<条件節タラバ>	5 (0.02%)
<条件節ナラバ>	3 (0.01%)
<引用節トノ>	3 (0.01%)
/トイウ節/	3 (0.01%)
<接続詞 M>	2 (0.01%)
/条件節レバ/	2 (0.01%)
/間接疑問節/	2 (0.01%)
<理由節カラニハ>	1 (0.00%)
<並列節ダノ>	1 (0.00%)
<接続詞 CL>	1 (0.00%)
/条件節ナラ/	1 (0.00%)
/ノニ節/	1 (0.00%)
/テモ節/	1 (0.00%)
/テカラ節/	1 (0.00%)

模擬講演 合計 25,569(100%)

5.3 節境界ラベルの一覧

以下、5.3節では、CBAP-csjが付与する節境界ラベルについて、代表的な形態素列パターン、形式的・意味的な特徴、具体例などを図5.9のような形式で記述していく。なお、ここで記述する形態素列パターンだけで全ての節境界ラベルの付与が制御されているわけではないことに注意されたい。

<p>節境界ラベル名</p> <p>パターン： 節境界検出ルールに含まれる代表的な形態素列パターン。</p> <p>形式： その節境界が備える形式的な特徴。</p> <p>意味： その節境界が備える意味的な特徴。</p> <p>注意点： CBAP-csjの解析結果に対して注意すべき点。</p> <p>例文</p>
--

図 5.9 節境界ラベル一覧の記述形式

5.3.1 絶対境界

5.3.1.1 [文末]

パターン：

IN: (終止形|終止形-(省略|撥音便A|促音便A)|命令形) ;

OUT: \$1 [文末] ;

形式： 用言の終止形・命令形などによって発話が形式的に完結される点。

意味： 叙述の完結点であり、話し手の心的態度が文法的な手段により表示される。

注意点： 「雨！」のようないわゆる一語文の文末や、「ちょっと変。」のような形式的に明示的でない（有標な形態的特徴を有しない）文末は、CBAP-csj 検出対象とならないため、[文末] のラベルは付与されない。人手修正の際にそのような箇所が発見された場合、修正の対象となる。「行くんですね」のようにノダ形に終助詞が後接した場合は [文末] となるが、「行くね」のようにノダ形以外の用言に終助詞が後接した場合は [文末候補] となる。

- | | |
|--|------------|
| (17) 簡単に最初に復習をしておきたいと<引用節>思います [文末] | (A01M0074) |
| (18) で<接続詞>それぞれ十六名の有効データを得ました [文末] | (A01F0055) |
| (19) あたしニワトリ歩いてるのって見たことが今までなかったんですね [文末] | (S03F0062) |
| (20) まず左側の方を御覧ください [文末] | (A01M0048) |

以下の例において、「十三人」「大変」の直後は [文末] に相当する箇所として分割されるべき位置であるが、

形式的に明示的でないために CBAP-csj の検出対象とはなり得ない。このような例は、人手修正の対象となる。(参照 → 5.5.1.1 節)

- (21) 被験者は全部で十三人で<接続詞>システムが何か聞いては<テハ節>答えるっていうのの繰り返し
ですね [文末候補] (A01M0021)
- (22) 金曜日の夜の終電中央林間行きとか乗っちゃうと<条件節ト>もう凄い大変もうそれぐらいたくさん
の人が住んで<テ節>大手町とか都心に働きに行ってるんですね [文末] (S03F0062)

5.3.1.2 [文末候補]

パターン：

IN: (終止形|終止形-(省略|撥音便A|促音便A)|命令形)(. *?)_助詞-終助詞_);
OUT: \$1 [文末候補];

形式： いわゆる文末表現に相当するものの、終助詞が後接していることによって、文脈によっては文末にならない可能性がある場合。典型的には、用言の終止形に終助詞「か」「ね」などが後接する場合。

意味： [文末] と同様、叙述の完結点であり、話し手の心的態度が文法的な手段により表示される。

注意点： CBA-csj による節境界解析では絶対境界として発話分割位置となるが、人手修正の際に文末ではないと判断された場合、修正の対象となる。典型的には、付与された終助詞が間投助詞として機能している場合（「昨日ですね」）や、発話の一時的な中断点になっている場合（「何だったかな」）など。

- (23) 一番左が最初千文書を取得した段階での初期検索の精度ですね [文末候補] (A03M0045)
- (24) 犬を飼っている方は多分犬を全般的に好きな方が多いですね [文末候補] (S00F0031)
- (25) 1. それにしても<テモ節>私が本で読んだあの情報は一体何だったのでしょうか [文末候補]
2. 保守的な生活を好むはずの人々がそんなに急進な手段を取るのでしょうか [文末候補]
(S00F0173)

以下の例において、「那覇に到着した後ですね」の「ですね」は文末表現ではなく間投助詞として働いており、文末表現とは認められないため、不適切な分割結果として人手修正の対象となる。(参照 → 5.5.5.1 節)

- (26) 1. 那覇に到着した後ですね [文末候補]
2. 船酔いに苦しんだ多くの乗客は帰りのチケットをキャンセルしていました [文末] (S02M0011)

また、以下の例において、「うちの主人も」は「離れ難かった」に係っており、「何て言うんですか」は発話中に挿入された部分であると認められるため、不適切な分割結果として人手修正の対象となる。

- (27) 1. で<接続詞>結局うちの主人も何て言うんですか [文末候補]
2. 離れ難かったのかもかもしれませんけどね/並列節ケド/... (S03F0314)

5.3.1.3 [と文末]

パターン：

IN: (と_助詞-格助詞_) <引用節> ((.*?)(感動詞|接続詞)_);
 OUT: \$1 [と文末] \$2 ;

形式： いわゆる文末表現に格助詞「と」が後接して発話が完結している場合。

意味： 「～と」という形を取る，叙述の完結点。演説調の発話スタイルの中でよく見られる。

注意点： 用言の終止形などに格助詞「と」が後接すると<引用節>のラベルが付与される（参照 → 5.3.3.17 節）が，その直後に感動詞または接続詞が出現した場合に，ラベルが[と文末]に書き換えられる。

- | | |
|------|--|
| (28) | 1. まず音声データベースからメルケプストラム分析行ないまして/テ節/
2. メルケプストラムとその動的特徴量デルタとデルタデルタを求めたいと [と文末]
3. で<接続詞>後は音素ごとにエイチエムエムを学習しまして/テ節/... (A01M0007) |
| (29) | 1. で<接続詞>総論では賛成なんだけれども/並列節ケレドモ/
2. 少し煮詰める必要があるんじゃないだろうか [と文末]
3. で<接続詞>私もそういう風に思います [文末] (S06M0894) |

5.3.2 強境界

5.3.2.1 /並列節ガ/

パターン：

IN: (が_助詞-接続助詞_);
 OUT: \$1 /並列節ガ/ ;

形式： 接続助詞「が」により導かれる節。

意味： 前件と後件が並列的關係（逆接または順接）にあることを表す。話題の導入や前置きの表現として用いられることも多い。

注意点： いったん発話した内容について補足するために，発話の途中に「～が」という節が挿入される場合がある。この場合，不適切な分割結果が生じるため，人手修正の対象となる。

- (30) 1. 預けるの連用形預けは通常それだけでは名詞になりませんが/並列節ガ/
2. お預けと語頭におが付くことによって<テ節>犬の仕付けの一つや転じて<テ節>実行が保留されている状態を表わすようになります [文末] (A02F0038)
- (31) 1. で<接続詞>結果ですが/並列節ガ/
2. まず絶対音感群の結果から見ていききたいと<引用節>思います [文末] (A01F0067)
- (32) 1. まず一つ目の検証方法ですが/並列節ガ/
2. 単純に音響尤度のみによる検証を行いません [文末] (A01M0103)

以下の例において、「色々なパターンを」は「集めてみました」に、「小学校一年生になる時に」は「引っ越しまして」にそれぞれ係っており、「ここに書いてある数字は頻度ですが」「昔は僕は足立区に住んでいたんですが」は発話中に挿入された部分であると認められるため、不適切な分割結果として人手修正の対象となる。(参照 → 5.5.2.2 節)

- (33) 1. 色々なパターンをここに書いてある数字は頻度ですが/並列節ガ/
2. たくさん集めてみました [文末] (A01M0021)
- (34) 1. 僕は小学校一年生になる時に昔は僕は足立区に住んでいたんですが/並列節ガ/
2. 今住んでる文京区に引っ越しまして/テ節/... (S02M0161)

5.3.2.2 /並列節ケレドモ/ /並列節ケレド/ /並列節ケドモ/ /並列節ケド/

パターン：

IN:	(けれど_助詞-接続助詞__ も_助詞-副助詞__);
OUT:	\$1 /並列節ケレドモ/;
IN:	(けれど_助詞-接続助詞__);
OUT:	\$1 /並列節ケレド/;
IN:	(けど_助詞-接続助詞__ も_助詞-副助詞__);
OUT:	\$1 /並列節ケドモ/;
IN:	(けど_助詞-接続助詞__);
OUT:	\$1 /並列節ケド/;

形式： 接続助詞「けれど」＋副助詞「も」により導かれる節。または、接続助詞「けど」＋副助詞「も」、接続助詞「けれど」、接続助詞「けど」により導かれる節。

意味： 前件と後件が並列的關係（逆接または順接）にあることを表す。話題の導入や前置きの表現として用いられることも多い。

注意点： いったん発話した内容について補足するために、発話の途中に「～けれども」などの節が挿入される場合がある。この場合、不適切な分割結果が生じるため、人手修正の対象となる。

- (35) 1. 女性の場合でもやはり同様な傾向だったんですけれども/並列節ケレドモ/
2. 男性の場合の時に見られたような逆転現象が有意な程には見られていません [文末] (A01M0096)
- (36) 1. まずその教室の様子ですけれども/並列節ケレドモ/
2. 教室の名前は俳句文法教室と言います [文末] (S04F1495)
- (37) 1. 今からグラフお見せしますけれども/並列節ケレドモ/
2. 縦軸と横軸の説明だけ先にしときます [文末] (A01M0096)

以下の例において、「合計四個の」は「電流双極子」に、「早速」は「チェックしました」にそれぞれ係っており、「この緑の部分なんですけれども」「夜着いたんですけれども」は発話中に挿入された部分であると認められるため、不適切な分割結果として人手修正の対象となる。(参照 → 5.5.2.2 節)

- (38) 1. この例では左右に二個ずつの合計四個のこの緑の部分なんですけれども/並列節ケレドモ/
2. 電流双極子を必要としました [文末] (A01M0025)
- (39) 1. ホテルの部屋の中も早速夜着いたんですけれども/並列節ケドモ/
2. チェックしました [文末] (S01F0050)

5.3.2.3 /並列節シ/

パターン：

IN: (し_助詞-接続助詞_) ;

OUT: \$1 /並列節シ/ ;

形式： 接続助詞「し」により導かれる節。

意味： 前件と後件が並列的關係にあることを表す。關係の意味としては、並列・累加・理由などを表す。

- (40) 1. 御覧のようにエフゼロにも非常に大きな違いが観察されますし/並列節シ/
2. から<接続詞>デュレーションにも非常に大きな違いが観察されます [文末] (A01M0074)
- (41) 1. だから<接続詞 I>飼い主なりあるいは<接続詞>隣近所の人が日本語で言えば<条件節レバ>
ちゃんと座るし/並列節シ/
2. 来いと<引用節>言えば<条件節レバ>来る [文末] (S02M1698)
- (42) 1. で<接続詞>その豪華な船に乗れたっていう<トイウ節>体験もできたし/並列節シ/
2. 楽しかったと<引用節>思っています [文末] (S01F0183)

5.3.2.4 /ヨウニ節/

パターン：

IN: (ます_助動詞__連体形 よう_形状詞__ に_助動詞__連用形) ;
 OUT: \$1 /ヨウニ節/ ;
 IN: (まし_助動詞__連用形 た_助動詞__連体形 よう_形状詞__ に_助動詞__連用形) ;
 OUT: \$1 /ヨウニ節/ ;

形式： 助動詞の連体形「ます」+形状詞「よう」+助動詞の連用形「に」で導かれる節。

意味： 例示や引用の意味を表す。特に「先程申し上げましたように」「一番下に書いてありますように」のような形で、講演中に注釈を添えるような場面で用いられることが多い。

注意点： /ヨウニ節/という節境界ラベルは、「ますように」「ましたように」という接続にしか付与されず、強境界しか存在しない。

- | | | |
|------|--|------------|
| (43) | 1. 先程申し上げましたように/ヨウニ節/
2. 条件付き確率の式はこのように異なっています [文末] | (A03M0010) |
| (44) | 1. よく知られますように/ヨウニ節/
2. 線形予測法というのがございまして/テ節/... | (A01M0147) |

5.3.3 弱境界

5.3.3.1 <条件節タラ> <条件節タラバ>

パターン：

IN: ((たら|だら)_助動詞__仮定形) ;
 OUT: \$1 <条件節タラ> ;
 IN: <条件節タラ> (ば_助詞-接続助詞__) ;
 OUT: \$1 <条件節タラバ> ;

形式： 助動詞の仮定形「たら」「だら」により導かれる節。または「たら」「だら」+接続助詞「ば」により導かれる節。

意味： 前件で示される内容が、後件で示される事態の仮定条件や確定条件（時系列関係）になっていることを表す。

注意点： 「そうしたら」「ひょっとしたら」「どうしたら」のような固定的な表現に対しても<条件節タラ>が付与される。

- (45) 学部の成績で選ばれたら<条件節タラ>多分落ちていたと<引用節>思うんですけども
/並列節ケレドモ/... (S01M0051)
- (46) 初めに言わば素人段階でどのぐらいできるかと<引用節>調査をしたら<条件節タラ>六十七パー
セントが鼻濁音できないんですね [文末] (A05M0040)
- (47) で<接続詞>乗り移って<テ節>ビーのやつも乗り移ったらば<条件節タラバ>まず最初にやることは
ですね見学履歴と興味を共有相違部分を検出する [文末] (A04M0047)
- (48) で<接続詞>その時にちょうど先生がもしかしたら<条件節タラ>私は東京都指定の特殊な難病かも
しれないと<引用節>言われました [文末] (S02F0100)

5.3.3.2 <条件節ト>

パターン：

IN: (と_助詞-接続助詞_);

OUT: \$1 <条件節ト> ;

IN: ((です|ます)_助動詞_終止形 と_助詞-接続助詞_) <条件節ト> ;

OUT: \$1 /条件節ト/ ;

形式： 接続助詞「と」により導かれる節。

意味： 前件で示される内容が、後件で示される事態の仮定条件や確定条件（時系列関係）になっていることを表す。

注意点： 用言がデスマス体の場合、メタ規則により弱境界が強境界に書き換えられる。「まとめますと」「そ
うしますと」のように、談話の展開を表す表現として用いられることも多い。

- (49) 言語カテゴリーとして<テ節>どういうものを持ってくると<条件節ト>比較的うまく行くのかとい
う<トイウ節>ことを考えたという<トイウ節>ことです [文末] (A01M0157)
- (50) 翌日朝食を済ませると<条件節ト>私達は旅行代理店へ向かいました [文末] (S00F0173)
- (51) 1. 以上をまとめますと/条件節ト/
2. 我々は本研究におきまして<テ節>周波数伸縮を用いた話者正規化法の種々の年齢層話者に対す
る効果これを検証いたしました [文末] (A01M0115)

5.3.3.3 <条件節ナラ> <条件節ナラバ>

パターン：

IN:	((終止形 終止形-(省略 撥音便A 促音便A)) なら_助動詞__仮定形) ;
OUT:	\$1 <条件節ナラ> ;
IN:	<条件節ナラ> (ば_助詞-接続助詞__) <条件節レバ> ;
OUT:	\$1 <条件節ナラバ> ;
IN:	(連体形 (の ん)_助詞-準体助詞__ なら_助動詞__仮定形) ;
OUT:	\$1 <条件節ナラ> ;

形式： 用言の終止形 + 助動詞の仮定形「なら」により導かれる節。または「なら」 + 接続助詞「ば」により導かれる節。

意味： 前件で示される内容が、後件で示される事態の仮定条件や推論の前提条件になっていることを表す。

注意点： 「できるなら」のような固定的な表現に対しても<条件節ナラ>が付与される。用言の連体形 + 準体助詞「の」 + 「なら」で<ナラ節>が構成される場合もある。

- (52) 日本のテレビと英国のテレビ一番違う面というのは英国のテレビはいかにも見るなら<条件節ナラ>何かを学べっていう<トイウ節>ものが多かったのをよく覚えてます [文末] (S00M0213)
- (53) もし形式的にこの会話の中で起きた他の相づちに習うならば<条件節ナラバ>ここではそうかなそうだよといったものになるはずなのではないでしょうか [文末候補] (A06F0049)
- (54) できるならば<条件節ナラバ>遭難でなくてですね<テ節>無人島に行って<テ節>そういう経験したいなど<引用節>文章作りながら思いました [文末] (S07M0833)

5.3.3.4 <条件節レバ>

パターン：

IN:	(ば_助詞-接続助詞__) ;
OUT:	\$1 <条件節レバ> ;

形式： 接続助詞「ば」で導かれる節。

意味： 前件で示される内容が、後件で示される事態の仮定条件や推論の前提条件になっていることを表す。

注意点： 「～すればいい」のような固定的な表現に対しても<条件節レバ>が付与される。

- (55) でも<接続詞>一生懸命頑張っていれば<条件節レバ>いつか爽快な気持ちでゴールテープを切る日が来るのかしらという<トイウ節>風に思うのです [文末] (S00F0131)
- (56) で<接続詞>動的に音素ツリーを生成するのであれば<条件節レバ>各状態で残すパスは一つで良いのですが/並列節ガ/... (A01M0030)
- (57) で<接続詞>ついでにこれもいい日本語に替えてくれれば<条件節レバ>いいなど<引用節>そんな気がしてます [文末] (S06M0373)

5.3.3.5 <理由節カラ>

パターン：

IN: (から_助詞-接続助詞__);

OUT: \$1 <理由節カラ>;

形式： 接続助詞「から」で導かれる節。

意味： 前件で示される内容が、後件で示される事態の原因・理由や、推論の根拠となっていることを表す。

(58) どうしてもこの遊びの方が先に入ってたから<理由節カラ>そっちを優先しちゃったっていう
<トイウ節>ことでした [文末] (S02F0012)

(59) 全体としては精度は約八割で<接続詞>ベースラインが大体三割ですから<理由節カラ>有意な正しい有効な結果が出ていると<引用節>考えられます [文末] (A03M0059)

5.3.3.6 <理由節カラニハ>

パターン：

IN: <理由節カラ> (に_助詞-格助詞__ は_助詞-係助詞__);

OUT: \$1 <理由節カラニハ>;

形式： <理由節カラ>に格助詞「に」+ 係助詞「は」が後接したもの。

意味： 前件で示される内容が、後件で示される判断の前提となっていることを表す。

(60) ここで意味的な繋がりと言うからには<理由節カラニハ>何らかの基準が必要となってきます [文末]
(A03M0138)

(61) で<接続詞>逆に言うと<条件節ト>常任理事国になるからには<理由節カラニハ>我が国が我が国の
軍事組織が外国で行動することいわゆる海外派兵これが問題にならないように憲法を改めざるを得
ない訳です [文末] (S06M0373)

5.3.3.7 <理由節カラ-助詞>

パターン：

IN: <理由節カラ> ((.*?)_助詞-(係|副)助詞__);

OUT: \$1 <理由節カラ-助詞>;

形式： 「からこそ」などのように、<理由節カラ>に係助詞または副助詞が後接したもの。

意味： 「からこそ」の場合、前件で示される内容が唯一の理由になって後件が引き起こされたことを表す。

- (62) 寝泊まりを一緒にしたからこそ<理由節カラ-助詞>長く続く友達ができたんじゃないかと<引用節> 思います [文末] (S00F0083)
- (63) ただ<接続詞>私は地元なので<理由節ノデ>逆に地元だからこそ<理由節カラ-助詞>お店にあまり 入らなくて<テ節>どこのお店がおいしいだとか<ト力節>よく行き付けのお店がっていうのは実は ないんですね [文末] (S03F0224)

5.3.3.8 <理由節ノデ>

パターン:

IN: (連体形 (の|ん)_助詞-準体助詞__ で_助動詞__連用形) ;
 OUT: \$1 <理由節ノデ> ;

形式: 用言の連体形 + 準体助詞「の」「ん」 + 助動詞の連用形「で」で導かれる節。

意味: 前件で示される内容が、後件で示される事態の原因・理由や、推論の根拠となっていることを表す。

- (64) 縮むことがないので<理由節ノデ>安心して<テ節>何度でも洗います [文末] (S04F0013)
- (65) 御飯一に対して<テ節>イクラの割合で三杯ぐらい食べちゃったんで<理由節ノデ>多分それが原因かなと<引用節>思います [文末] (S00M0025)

5.3.3.9 <タリ節> <タリ節-助詞>

パターン:

IN: (連用形 (たり|だり)_助詞-副助詞__) ;
 OUT: \$1 <タリ節> ;
 IN: <タリ節> ((は|も)_助詞-(係|副)助詞__) ;
 OUT: \$1 <タリ節-助詞> ;

形式: 用言の連用形 + 副助詞「たり」「だり」で導かれる節。

意味: 「～たり～たり」という形で、並列する動作や事態を表す。例示の意味で用いられることもある。

注意点: <タリ節>に係助詞「は」「も」が後接する場合は、<タリ節-助詞>となる。

- (66) これに対してですね<テ節>抽象化をこちら側は抽象化によって<テ節>ノードが小さくなったり <タリ節>大きくなったり<タリ節>する訳ですけども/並列節ケドモ/... (A03M0005)
- (67) 多義を判定する場合というのは非常に主観が入ってしまったり<タリ節>するので<理由節ノデ>あまり無理やり分けないという<トイウ節>こと方針にいたしました [文末] (A03F0109)
- (68) いわゆる制作進行助監督みたいなこともやったりも<タリ節-助詞>しました [文末] (S01M0227)

5.3.3.10 <テ節>

パターン：

IN:	((て で)_助詞-接続助詞__)	;
OUT:	\$1 <テ節>	;
IN:	((でし まし)_助動詞__連用形 て_助詞-接続助詞__)	<テ節> ;
OUT:	\$1 /テ節/	;
IN:	(ませ_助動詞__未然形 ん_助動詞__終止形 で_助詞-接続助詞__)	<テ節> ;
OUT:	\$1 /テ節/	;

形式： 接続助詞「て」「で」で導かれる節。

意味： 様態，事態間の継起関係，因果関係，並列関係などを表す。

注意点： 用言がデスマス体の場合，メタ規則により弱境界が強境界に書き換えられる。「について」「によって」「おいて」「に関して」のような固定的な表現に対しても<テ節>が付与される。テ節の直後に「いる」「ある」「しまう」などが後接してアスペクト的意味を表す場合，<テ節>は付与されない。

- | | |
|--|------------|
| (69) で<接続詞>自分達でレンタカー車を飛ばして<テ節>行きました [文末] | (S00F0014) |
| (70) 1. で<接続詞>当日は四時に起きまして/テ節/
2. で<接続詞>最少のスタッフで現場に直行いたしました [文末] | (S01M0227) |
| (71) ある時は失恋して<テ節>泣いている私に寄り添ってもくれました [文末] | (S00F0031) |
| (72) これに対し<連用節>日本語話者は日本語の長短母音のカテゴリー判断では物理的な長さのみで決まるカテゴリーを持っていて<テ節>それに基づき<連用節>長さの判断をしていると<引用節>考えられます [文末] | (A01F0122) |
| (73) では<接続詞>実際に我々がこれまで構築してきた検出システムの評価結果について<テ節>述べていきます [文末] | (A01M0015) |
| (74) 脳磁図測定によって<テ節>得られた代表的な結果を示します [文末] | (A01M0025) |

5.3.3.11 <テハ節>

パターン：

IN:	(< /) テ節 (> /) (は_助詞-係助詞__)	;
OUT:	\$3 <テハ節>	;
IN:	((連用形 連用形-促音便) ちゃ_助詞-接続助詞__融合)	;
OUT:	\$1 <テハ節>	;

形式： <テ節> /テ節/ に係助詞「は」が後接したもの。

意味： 繰り返す動作を表す。「に関しては」「については」「によっては」「に対しては」などの形は，ある種

の主題を提示する。

注意点： 「してはいけない」のような固定的な表現に対しても<テハ節>が付与される。「ては」の縮約形として「ちゃ」がある。強境界の/テ節/の直後に「は」が現れた場合でも、弱境界の<テハ節>に書き換えられる。

- (75) もう初日から彼らを探しては<テハ節>きゃあきゃあ騒ぎ過ぎて<テ節>もう声枯れしちゃって
<テ節>大変でした [文末] (S01F0050)
- (76) 明らかな言い淀みに関しては<テハ節>このようにシステムの出力が何らかの形で重なってれば
<条件節レバ>それを正解と見なしました [文末] (A01M0015)
- (77) 周波数伸縮を用いたさまざま伸縮法につきましては<テハ節>これまでもさまざまな手法関数等が
提案されてきております [文末] (A01M0115)
- (78) 十一月頃イラン革命が始まって<テ節>まず戒厳令と言って<テ節>夜の最初は八時以降は外出しちゃ
<テハ節>いけない [文末] (S03F0072)

5.3.3.12 <テモ節>

パターン：

IN: (<|/) テ節 (>|/) (も_助詞-係助詞_) ;

OUT: \$3 <テモ節> ;

形式： <テ節> /テ節/ に係助詞「も」が後接したもの。

意味： 前件が成立したとしても、そこから予期される後件が成立しないという、譲歩関係を表す。「に関して」「につきましても」などの形は、ある種の主題を提示する。

注意点： 「してもいい」のような固定的な表現に対しても<テモ節>が付与される。

- (79) で<接続詞>今回ちょっと運が悪かったみたいですね<並列節デ>ヨーロッパに行きましても
<テモ節>ヴィトンの商品自体がないんですよ [文末] (S02M0245)
- (80) そういう機会を設けてくれた上司に対しても<テモ節>凄くこう感謝の気持ちでいっぱいだったんで
すね [文末] (S01F0074)
- (81) リコールと呼んでも<テモ節>いいんですけども/並列節ケドモ/... (A03M0005)

5.3.3.13 <テカラ節> <テカラ節-助詞> <テ節-助詞>

パターン：

IN: (<|/) テ節 (>|/) (から_助詞-格助詞__);
 OUT: \$3 <テカラ節>;
 IN: <(テ|テカラ) 節> ((.*?)_助詞-(.*?)__);
 OUT: \$2 <\$1 節-助詞>;

形式： <テカラ節>は、<テ節>に格助詞「から」などの助詞が後接したもの。<テカラ節-助詞>は、<テカラ節>に助詞が後接したもの。<テ節-助詞>は、<テ節>に格助詞「から」以外の助詞が後接したもの。

意味： <テカラ節>は、前件の内容が完了した後に後件の内容が生起することを表す。

- | | |
|--|------------|
| (82) そして<接続詞>最後に八月に入って<テ節>大学の授業が終わってから<テカラ節>修道院を訪ねました [文末] | (S01F0157) |
| (83) 寝たきりになってからは<テカラ節-助詞>どんどんと衰弱していきました [文末] | (S02F0180) |
| (84) この群においてのみ<テ節-助詞>原型への選好が有意に認められました [文末] | (A01F0055) |

5.3.3.14 <トカ節> <トカ節-助詞>

パターン：

IN: (終止形 と_助詞-格助詞__ か_助詞-副助詞__);
 OUT: \$1 <トカ節>;
 IN: <トカ節> (.*?)_助詞-(.*?) 助詞__);
 OUT: \$1 <トカ節-助詞>;

形式： 用言の終止形 + 格助詞「と」 + 副助詞「か」で導かれる節。

意味： 「～とか～とか」という形で、並列する動作や状態を表す。例示の意味でも用いられる。

注意点： <トカ節>に助詞が後接する場合は、<トカ節-助詞>となる。

- | | |
|--|------------|
| (85) 洗濯機の音ですとか<トカ節>テレビの音ですとか<トカ節>色々あります [文末] | (A01M0137) |
| (86) で<接続詞>その時点でも何の疑いもなく<連用節>例えば御家族じゃないですかとか<トカ節>色々なこと向こうは言うてくるんですね [文末] | (S02F0121) |
| (87) で<接続詞>そうすると<条件節ト>大体遅く食うとか<トカ節>あるいは<接続詞>残すとかに<トカ節-助詞>なる訳ですけど/並列節ケド/... | (S02M0161) |

5.3.3.15 <ノ二節>

パターン：

IN: (連体形 の_助詞-準体助詞__ に_助詞-格助詞__);

OUT: \$1 <ノ二節>;

形式： 用言の連体形に準体助詞「の」+格助詞「に」が後接したもの。

意味： 典型的には、前件と後件の逆接関係を表す。内容を表す節（名詞節）として用いられることもある。その場合、「ことに」「ものに」などに言い換えることができる。

- (88) 1. ですね<接続詞>食べ物がろくに出ないのに<ノ二節>吐き気だけを催すというですね
<トイウ節>一番つらい状況になった訳です [文末] (S02M0011)
- (89) 基本的に人間って変わらないんだなっていうのに<ノ二節>気が付いた時自分が外国に対して<テ節>
持っていた情熱は何だったんだろうって<引用節>考えたんです [文末] (S00M0213)
- (90) 何か自分が死ぬのが悲しいんじゃなくて<テ節>親にそういう思いをさせている自分というのに
<ノ二節>凄く悲しみを感しました [文末] (S02F0100)

5.3.3.16 <連用節>

パターン：

IN: ((.*?)(動詞|助動詞)(.*?)(連用形));

OUT: \$1 <連用節>;

IN: (で_助動詞__連用形) <連用節> (ある_動詞_ラ行五段_終止形);

OUT: \$1 \$2;

IN: <連用節> ((始め|始める)_動詞_マ行下一段_(.*?));

OUT: \$1;

IN: (思い切り_動詞_ラ行五段_連用形) <連用節>;

OUT: \$1;

形式： 動詞、助動詞の連用形で導かれる節。

意味： 前件と後件が並列的關係にあることを表す。關係的意味としては、事態間の繼起關係、因果關係、並列關係などを表す。

注意点： 形容詞の連用形は、連用修飾語として働く場合と述語として働く場合とがある。後者は<連用節>のラベルが付与されるべき箇所であるが、CBAP-esjでは検出することができない。人手修正の際にそのような箇所が発見された場合、修正の対象となる。また、助動詞の連用形「で」に「ある」「はない」などが後接する場合や、動詞の連用形の直後に動詞「始める」「続ける」、接続助詞の「て」などが後接する場合には、<連用節>は付与されない。「思い切り」「何となく」「繰り返し」「思わず」などは述語として働く場合よりも連用修飾語として働く場合が多いので、<連用節>のラベルは付与されない。

- (91) 花火が上がり<連用節>紙吹雪が舞い<連用節>みんなもう酔っ払ったアメリカ人が踊りながらハッピーニューイヤーと叫び合います [文末] (S01F0050)
- (92) その一本の足でも地面に付くことができなくなると<条件節ト>体重が支えられなくなり<連用節>やがて体中の毛細血管が収縮して<テ節>命取りになってしまうのです [文末] (S00F0131)
- (93) 次に刺激提示条件は聴覚提示のみ視覚提示のみ視聴覚提示の三条件とし<連用節>被験者は比較刺激の一つを数えるという<トイウ節>課題としました [文末] (A01M0025)
- (94) ちなみに<接続詞>この時に使ったシャツやパンツゼッケン運動靴はその後使わずに<連用節>ずっと取ってあります [文末] (S01M0091)
- (95) 聞くところによると<条件節ト>この地方を治めていたスペイン人はとても温厚な人柄だったらしく<連用節>先住民と争うようなことはあまりなかったのだそうです [文末] (S00F0173)

以下の例において、形容詞の連用形「多く」「安く」は述語として機能しており、その直後が連用節の境界として判断できるため、人手修正の対象となる。

- (96) 単純語の五百九十七語に入っていない動詞でもこのように合成語としてなら<条件節ナラ>名詞になり得る語は非常に多く分類語彙表中の合成語で<並列節デ>後項が連用形名詞のものを取り出して<テ節>見ると<条件節ト>それだけでも二千百九十五語ありました [文末] (A02F0038)
- (97) 非常にお野菜物とかの物価が非常に安く今は若者達の集まるイタ飯屋さんとか中華のお店とかちよっと行列のできる店になってるくらい少し周りの情景が変わってます [文末] (S03F1477)

5.3.3.17 <引用節> <引用節-助詞> <引用節トノ>

パターン：

IN:	((終止形 命令形 未然形 助詞-終助詞__)と_助詞-格助詞__);
OUT:	\$1 <引用節>;
IN:	((終止形 命令形 未然形 助詞-終助詞__) (つて て と なんて)_助詞-副助詞__);
OUT:	\$1 <引用節>;
IN:	((終止形 命令形 未然形 助詞-終助詞__) など_助詞-副助詞__ と_助詞-格助詞__);
OUT:	\$1 <引用節>;
IN:	<引用節> ((は も)_助詞-係助詞__);
OUT:	\$1 <引用節>;
IN:	<引用節> ((さえ すら しか まで も)_助詞-副助詞__);
OUT:	\$1 <引用節-助詞>;
IN:	<引用節> (の_助詞-格助詞__);
OUT:	\$1 <引用節トノ>;

形式： 用言の終止形、命令形や終助詞など、文末表現に相当する述語句の直後に、格助詞「と」や副助詞「つて」など引用を表す助詞類が後接した形式。

意味： 発話や思考などの内容をまとめて引用し、後方に現れる発話・思考系の述語句を補足する。

注意点： 格助詞「と」の位置には、副助詞「つて」「なんて」や、副助詞「など」+格助詞「と」などが現れ得る。＜引用節＞の直後に係助詞「は」「も」が後接した場合、その直後までが＜引用節＞となる。副助詞「さえ」「まで」「も」が後接した場合、その直後は＜引用節-助詞＞に、また格助詞「の」が後接した場合、その直後は＜引用節トノ＞になる。

- | | |
|---|------------|
| (98) 何となく雰囲気は分かったと＜引用節＞思います [文末] | (A05F0039) |
| (99) で＜接続詞＞結局みんなそのメンバーが凄くいい恋愛をしてたから＜理由節カラ＞全部引っくるめて＜テ節＞トータルで凄く楽しい時期だったなって＜引用節＞思ってます [文末] | (S01F0166) |
| (100) 先住系の人々は皆貧しくて＜テ節＞とても柄が悪いとか＜ト力節＞先住民は反政府的な活動しているなどと＜引用節＞言い＜連用節＞皆あまり勧めてはくれませんでした [文末] | (S00F0173) |
| (101) だから＜接続詞＞凄く自立が遅くなってるなどは＜引用節＞思うんですね [文末] | (S06F0167) |
| (102) もう私としては裏切られたとしか＜引用節-助詞＞言いようがありません [文末] | (S02F0012) |
| (103) スペイン語ではないので＜理由節ノデ＞分からないとの＜引用節トノ＞ことでした [文末] | (S00F0173) |

5.3.3.18 <トイウ節>

パタン：

IN: <引用節> (いう_動詞_ワア行五段_連体形) ;

OUT: \$1 <トイウ節> ;

IN: ((終止形|命令形|未然形|助詞-終助詞__)) っちゅう_助動詞__連体形) ;

OUT: \$1 <トイウ節> ;

形式： <引用節>に動詞の連体形「いう」が後接した形式。

意味： 発話や思考などの内容をまとめて引用し、後方に現れる体言を修飾する。「こと」を修飾する場合も多い。

注意点： 引用の形式が助動詞の連体形「っちゅう」のみで表される場合がある。

- (104) そして<接続詞>正しい答え正しいシラブルの数が入力されるまでそのトライアルを繰り返すという
<トイウ節>手法です [文末] (A01M0048)
- (105) 人類言語学者のシルバースタイン一九七六は言語の機能は大きく分けて<テ節>二つの機能があるという
<トイウ節>ことを述べています [文末] (A06F0049)
- (106) 何か汗がすぐ塩になっちゃう感じで<並列節デ>あんまりだらだらと汗をかいたっていう<トイウ節>
イメージがないです [文末] (S03F0072)
- (107) そこで<接続詞>やっぱどうしても働くっちゃう<トイウ節>意欲がやっぱり人間には備わってくる
んじゃないかと<引用節>一つ考えてみたいと<引用節>思います [文末] (S05M1505)

5.3.3.19 <間接疑問節> <間接疑問節-助詞>

パターン：

-
- IN: ((終止形|終止形-(省略|撥音便A|促音便A))か_助詞-副助詞_);
- OUT: \$1 <間接疑問節>;
- IN: (連体形の_助詞-準体助詞_か_助詞-(終|副)助詞_);
- OUT: \$1 <間接疑問節>;
- IN: <間接疑問節>(どう_副詞_か_助詞-副助詞_);
- OUT: \$1 <間接疑問節>;
- IN: (何_代名詞_だ_助動詞_終止形か_助詞-副助詞_)<間接疑問節>;
- OUT: \$1;
- IN: <間接疑問節>((.*?)_助詞-(格|係|副)助詞_);
- OUT: \$1 <間接疑問節-助詞>;
-

形式： 用言の終止形 + 副助詞「か」、または、用言の連体形 + 準体助詞「の」 + 副助詞（または終助詞）「か」で導かれる節。

意味： 動作や事態を不定化した上でまとめ、後方に現れる述語句を補足する。

注意点： 「～かどうか」という形で用いられることもある。「何だか」という表現は連用修飾語として働く場合が多いので、<間接疑問節>のラベルは付与されない。<間接疑問節>の直後に格助詞、係助詞、副助詞が後続した場合は、<間接疑問節-助詞>となる。

- (108) そういったモデルがまだどうしたら<条件節タラ>良いか<間接疑問節>よく分からないというのが
実態だと<引用節>思います [文末] (A11M0469)
- (109) とにかく夢のような八日間で<並列節デ>今でも本当にそこに行ったのか<間接疑問節>時々分から
なくなるくらいです [文末] (S01F0050)
- (110) で<接続詞>キリマンジャロは五千メートルも高さがあるので<理由節ノデ>それまでに体力を蓄え
られるかどうか<間接疑問節>分からないんですけれども/並列節ケレドモ/ldots (S01F0151)
- (111) どのモデルが良いかは<間接疑問節-助詞>講演によって<テ節>異なります [文末] (A11M0369)
- (112) この大成誠太のせいの部分に短縮が起きているかどうかを<間接疑問節-助詞>先程のインタビュー
の場合と同様に二人で聞いて<テ節>判断しました [文末] (A05M0068)

5.3.3.20 <連体節テノ>

パターン:

IN: (<テ節> の_助詞-格助詞_) ;
OUT: \$1 <連体節テノ> ;

形式: <テ節>に格助詞「の」が後接した形式。

意味: <テ節>に「の」が後接することにより、後方に現れる体言を修飾する。「としての」「についての」「に
対しての」のような固定的な表現に対しても<連体節テノ>が付与される。

- (113) 要するに<接続詞>片仮名で書く外来語にしか使わないからという<トイウ節>安心感があっての
<連体節テノ>ことだとは<引用節>思うんですが/並列節ガ/... (A05M0040)
- (114) 日本語のネイティブスピーカーとしての<連体節テノ>発言も欲しいという<トイウ節>ことで
<並列節デ>そういう立場から少し発言させていただきます [文末] (A05F0154)
- (115) 研究の内容についての<連体節テノ>発表はここまでで<並列節デ>実際に実装したツールのデモン
ストレーションを行ないたいと<引用節>思います [文末] (A03M0138)

5.3.3.21 <並列節ダノ>

パターン:

IN: (終止形 だ_助動詞__終止形 の_助詞-格助詞_) ;
OUT: \$1 <並列節ダノ> ;

形式: 用言の終止形に助動詞「だ」+格助詞「の」が後接した形式。

意味: 「～だの～だの」という形で、並列する動作や事態を表す。例示の意味で単独でも用いられる。

(116) で<接続詞>そこに行くのにはやはり日本人で凄く幼く見られるので<理由節ノデ>もう入り口からパスポート見せるだの<並列節ダノ>身分証を見せるとか<ト力節>凄くうるさかったんですけども/並列節ケレドモ/... (S01F0050)

5.3.3.22 <並列節デ>

パタン：

IN: (で_助動詞__連用形) ;
 OUT: \$1 <並列節デ> ;
 IN: <並列節デ> ((あつ|あら|あり|ある|あれ|ある)_動詞_ラ行五段_(.*?)) ;
 OUT: \$1 ;
 IN: <並列節デ> (ない_助動詞__終止形) ;
 OUT: \$1 ;
 IN: <並列節デ> (は_助詞-係助詞__ ない_形容詞_形容詞型_終止形) ;
 OUT: \$1 ;

形式： 助動詞の連用形「で」で導かれる節。

意味： 前件と後件が並列的關係にあることを表す。關係の意味としては、事態間の並列關係、因果關係などを表す。

注意点： 「で」に「ある」「ない」「はない」などが後接する場合には、<並列節デ>のラベルは付与されない。

(117) 横軸が時間で<並列節デ>縦軸が周波数になっています [文末] (A01M0133)

(118) それで<接続詞>その為に大型のチェーン店が次々と入ってくるようで<並列節デ>益々治安が悪くなるんじゃないかと<引用節>みんなでもとても心配しています [文末] (S03F0133)

(119) んで<接続詞>元の親分の宮沢さんが大体加藤さんの行動に反対しましたもんで<並列節デ>あれも随分響いてるんじゃないかなと<引用節>思いましたよね [文末候補] (S06F1034)

5.3.3.23 <並列節ナリ>

パタン：

IN: (終止形 なり_助詞-副助詞__) ;
 OUT: \$1 <並列節ナリ> ;

形式： 用言の終止形に副助詞「なり」が後接した形式。

意味： 「～なり～なり」などの形で、並列する動作や事態を表す。例示の意味でも用いられる。

- (120) 結局そっから自分で達で道を聞くなり<並列節ナリ>何なりして<テ節>うまく帰ったり<タリ節>したんですけども/並列節ケレドモ/... (S00F0033)

5.3.3.24 <フィラー文>

パターン：

IN: (そう_副詞__ です_助動詞__終止形_ね_助詞-終助詞__) [文末候補] ;
 OUT: \$1 <フィラー文> ;

形式： 「そうですね」という形式の、それだけで独立した発話。

意味： 自発的な話しことば（特に独話）の場合、発話中に逡巡する際「そうですね」という形が多く見られる。形式としては [文末候補] と同等であるが、「そうですね」という形に限ってはフィラーのような働きをしているため、<フィラー文>という節境界ラベルを付与することにした。

- (121) 歩いて<テ節>そうですね<フィラー文>十分ぐらいでしょうか [文末候補] (S03M0317)
- (122) それから<接続詞>そうですね<フィラー文>これは<感動詞>このことはそうですね<フィラー文>一般的なサンプリング誤差の範囲は超えていると<引用節>と思いますが/並列節ガ/... (A07F0844)

5.3.3.25 <感動詞>

パターン：

IN: ((.*?)-感動詞__);
 OUT: \$1 <感動詞> ;

形式： 「ええ」「うん」「はい」など、単独で発話される感動詞。

意味： 基本的には、「はい」「ええ」「うん」「いいえ」「いや」など応答に用いられる要素。発話冒頭や発話中に現れてフィラーのような働きをする場合も多い。

注意点： 短単位の品詞が感動詞であり、かつ基本形にタグ (F) やタグ (M) が付いていないものに対して付与される。厳密には節境界ではないが、[と文末] のラベル付与に関わるため、検出対象となっている。人手修正の際にも、修正する必要性の有無を判断する手がかりとして用いられることがある。なお、感動詞の直後に「と」「って」など引用の形式が後接した場合、<感動詞> のラベルは付与されない。

- (123) 補償のことはって言いましたら<条件節タラ>いや<感動詞>それは考えてないっていう<トイウ節>
風にあの方も正直な方っていう風に新聞には出てますけど/<並列節ケド/>... (S06F1566)
- (124) 1. はい/<感動詞>/
2. それでは<接続詞 L>始めさせていただきます [文末] (A06F0075)
- (125) なぜかうん<感動詞>面白かった気がします [文末] (S03M0089)
- (126) よいしょ<感動詞>続いて<テ節>ワンパストライグラムデコーダーにおける探索パスのマージにつ
いて<テ節>説明します [文末] (A01M0030)

5.3.3.26 <接続詞> <接続詞 C> <接続詞 L> <接続詞 CL> <接続詞 M>

パタン：

IN: ((.*?)_接続詞__);
 OUT: \$1 <接続詞>;
 IN: ((.*?)_接続詞__) <接続詞> ((は|も|こそ)_助詞-(係|副)助詞__);
 OUT: \$1 \$3 <接続詞 C>;
 IN: ((.*?)_接続詞 L__);
 OUT: \$1 <接続詞 L>;
 IN: ((.*?)_接続詞 L__) <接続詞 L> ((は|も|こそ)_助詞-(係|副)助詞__);
 OUT: \$1 \$3 <接続詞 CL>;
 IN: (/(*.*/)(/ (って_助詞-副助詞__ 言う_動詞_ワア行五段_終止形 か_助詞-終助詞__));
 OUT: \$1 \$2 <接続詞 M>;

形式： <接続詞>は接続詞「で」のように一つの短単位から接続詞になるもの、<接続詞 C>は接続詞「けれど」+副助詞「も」のように短単位の接続から1語相当の接続詞と見なすことができるもの、<接続詞 L>は代名詞「それ」+格助詞「で」のように短単位では2単位だが長単位では1単位の接続詞「それで」になるもの、<接続詞 CL>は接続詞 L「ですけれど」+副助詞「も」のように長単位と短単位の接続から1語相当の接続詞と見なすことができるもの、<接続詞 M>は強境界直後に現れる副助詞「って」+動詞の終止形「言う」+終助詞「か」を1語相当の接続詞と見なしたものである。

意味： 語と語、発話と発話をさまざまな関係的意味によって連結する。

注意点： 短単位では複数の単位だが長単位では1単位になる「接続詞 L」は、CBAP-csjに入力される MOR 形式の段階ですでに1単位になっているものとする。

- (127) ただし<接続詞>関東では四都県内での県境を越える移動また<接続詞>関西では同一府県内での移動は構わないと<引用節>しました [文末] (A06F0128)
- (128) けれども<接続詞 C>実際の会話の中で起こる相づちは常に同じ形式のものが義務的に使われてる訳ではありません [文末] (A06F0049)
- (129) だから<接続詞 L>そういう意味では歴史のある町なんですね [文末] (S03M0046)
- (130) だからこそ<接続詞 CL>今そういう状況なってるのかもしれないけれども/並列節ケレドモ/... (S03M0046)
- (131) 1. ちょっとこれ汚い川なんですけど/並列節ガ/
2. って言うか<接続詞 M>ちょっとだいたい汚い川なんですけども/並列節ケドモ/ (S03M0098)

5.4 人手による修正と節単位の認定

5.4.1 人手修正作業の概要

以下では、前節までで述べてきた CBAP-csj による発話分割処理について、その分割結果を人手で修正することの必要性について、また、発話分割結果を人手で修正の際の操作の概要と手順について述べる。

人手で修正を行うことの最大の理由は、5.1.3 節でも述べたように、CBAP-csj の発話分割処理によって生じた不適切な分割結果を修正することである。先の (10) から (12) の例では、節境界以外の位置で発話を分割する必要がある場合（「体言止」や「倒置」など）に、人手で修正する事例を取り上げた。その他にもさまざまな理由により、人手修正が必要となる事例が存在する。また、構文レベル・談話レベルの情報付与作業（係り受け構造情報、重要文情報、談話境界情報）からの要請によって修正が必要となる事例も、多く存在する。以下では、例として、引用節構造を含む発話の過分割という事例を取り上げて説明する。

5.2 節で示したように、CBAP-csj は、局所的な形態素列のみを手がかりにして節境界の位置と種類を検出し、発話を分割するプログラムである。形態素の接続関係をパタンマッチで捉えることによって節境界の位置と種類を同定するこの手法は、非常に軽い処理で精度の高い結果を出力することが可能である。が、その一方、局所的な処理であるがゆえに、その節境界を含む大局的な構造までを把握することはできないという側面を持つ。

発話の大局的な構造を把握できないという特性により、場合によって発話が過度に分割されてしまうことがある。例えば、以下の例を見てみよう。

- (132) 1. 国際政治家の多くの方々は今現在国際社会は国家という枠組みを基礎にして<テ節>出来上がっているが/並列節ガ/
2. この国家という壁はいずれなくなり<連用節>越えられるであろう [文末]
3. しかし<接続詞>どうしても越えられないものが二つある [文末]
4. それは宗教と民族であると<引用節>言っていますが/並列節ガ/... (S02M0068)

(132) は、2つの [文末] と 2つの /並列節ガ/ によって 4つの部分に分割されている。この節境界の検出結果そのものは正しいのであるが、発話全体の分割結果としては適切ではない。発話の大局的な構造を考えると、1行目の「国際政治家の多くの方々は」は4行目の「言っています」に係る要素であり、その間にある部分は全て引用された要素、すなわち4行目の<引用節>に含まれる要素であることが分かる。しかしながら、上記の分割結果では、引用節に含まれる [文末] /並列節ガ/ の直後で発話が分割されてしまったことにより、係り受け関係を付与すべき冒頭と末尾の要素が分断されてしまっている。このような過分割の問題は、局所的なパタンマッチによる節境界の検出とそれに基づく発話分割処理では解決することができない問題である*12。

このような過分割の事例に対しては、CBAP-csj による発話分割結果を人手でチェックし、修正を加えることが必要となる。(132) の不適切な分割結果は、以下のように修正される。

(133) 国際政治家の多くの方々は { 今現在国際社会は国家という枠組みを基礎にして<テ節>出来上がっているが/並列節ガ/ : ; 引用節構造 この国家という壁はいずれなくなり<連用節>越えられるであろう [文末] : ; 引用節構造 しかし<接続詞>どうしても越えられないものが二つある [文末] : ; 引用節構造 それは宗教と民族である } と<引用節>言っています/並列節ガ/... (S02M0068)

引用節に含まれる絶対境界・強境界の直後で分割されていた箇所には「:」をマークし、分割された箇所を結合する。また、そのように操作した理由を「;」以下に記述する。さらに、引用されている範囲を{ }で囲んでマークする。これにより、CBAP-csj によって4つの単位に分割されていた(132)は1つにまとめられ、1つの節単位として認定されることになる。

人手修正作業は、基本的には、CBAP-csj による分割結果を作業者が目で確認した上で、さまざまな要因からその分割が不適切だと判断された場合に、「切断する」「結合する」「囲む」といういずれかの操作を適用して、適切な形に修正することである。修正操作の種類に応じて、「切断記号」「結合記号」「範囲記号」という人手修正操作記号が付与される。人手修正操作記号の一覧を、表5.4に示す。

表 5.4 人手修正操作記号の一覧

切断記号	-	分割されていなかった箇所を切断したことを表す記号
結合記号	+	分割されていた箇所を結合したことを表す記号
	:	引用節・連体節内で分割されていた箇所を結合したことを表す記号
範囲記号	{ }	絶対境界・強境界を含む引用節・連体節の範囲を表す記号
	« »	倒置要素の範囲を表す記号
	()	挿入節、挿入文、フィラー文の範囲を表す記号

5.4.2 人手修正作業で扱う項目の分類

次に、人手修正作業で扱う項目の分類について述べる。人手修正作業の種類は、その修正がどのような原因によって必要とされるのかという点から、図5.10に示すように、大きく5つに分類できる。

*12 この問題を解決するためには、発話分割処理に係り受け解析を組み込み、引用節の開始点・終了点のような大局的構造を把握する必要がある。係り受け解析を利用した話し言葉の発話分割処理については、下岡他(2005)を参照。

- | | |
|---------------------|---------------------------|
| 1. 大局的な文法構造に関するもの | → 大局的な文法構造に起因する不適切な分割結果 |
| 2. 非流暢性に関するもの | → 非流暢性に起因する不適切な分割結果 |
| 3. 係り受け構造に関するもの | → 係り受け構造情報の付与にとって不適切な分割結果 |
| 4. 談話構造上の問題に関するもの | → 談話構造に関わる不適切な分割結果 |
| 5. 不適切な節境界ラベルに関するもの | → 検出された節境界ラベルに問題がある場合 |

図 5.10 人手修正作業が必要となる項目の分類

これらの分類に対して、26 種類の具体的な修正項目を割り当てた。その上で、各項目に対してどのような修正を施すべきかの基準を作り、「**人手修正規則**」としてまとめた。図 5.11 に、修正対象となる項目の一覧を示す。数字は、コアに含まれる学会講演・模擬講演の計 177 講演にその修正項目が現れた総数を表す。括弧内の数字は、学会講演 70 講演・模擬講演 107 講演に各修正項目が出現した数を表す。それぞれの修正項目ごとの具体的な事例や人手修正規則の詳細については、5.5 節で一覧として示す。

1. 大局的な文法構造に関するもの			4. 談話構造上の問題に関するもの		
1-1 体言止	674	(359 : 315)	4-1 話題導入表現	536	(247 : 289)
1-2 引用節構造	513	(255 : 258)	4-2 直後がまとめ表現	341	(103 : 238)
1-3 連体節構造	45	(28 : 17)	4-3 話題の転換点	92	(23 : 69)
2. 非流暢性に関するもの			4-4 大きい切れ目	483	(76 : 407)
2-1 倒置	223	(32 : 191)	5. 不適切な節境界ラベルに関するもの		
2-2 挿入節	606	(222 : 384)	5-1 と文末	203	(92 : 111)
2-3 挿入文	82	(39 : 43)	5-2 間投助詞	135	(27 : 108)
2-4 フィラー文	94	(14 : 80)	5-3 例文など	34	(32 : 2)
2-5 言いさし	391	(155 : 236)	5-4 格助詞相当表現	10	(10 : 0)
2-6 言い直しマーカー	128	(56 : 72)	5-5 強→弱	206	(133 : 73)
2-7 述語の言い直し	131	(69 : 62)	5-6 文末候補	165	(73 : 92)
3. 係り受け構造に関するもの			5-7 非文末	63	(28 : 35)
3-1 主題の共有	471	(192 : 279)	5-8 タグミス	97	(40 : 57)
3-2 主題の飛び越し	121	(56 : 65)	総数 (学会講演 : 模擬講演)		
3-3 格要素の飛び越し	48	(19 : 29)	5,929 (2,403 : 3,526)		
5-1 連体形	37	(23 : 14)			

図 5.11 人手修正の対象となる項目の一覧と、コア 177 講演における出現数

5.4.3 人手修正作業の流れ

実際の人手修正作業は、各講演について次の手順で行った。

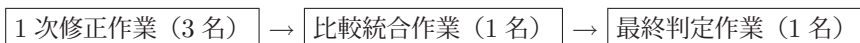


図 5.12 人手修正作業の手順

3名の1次修正作業者がCBAP-csjによる分割結果をチェックし、5.5節で示す人手修正規則に従って修正を行った。その結果を、1次作業とは異なる1名の作業者が比較統合し、3者の作業結果の間に生じた不一致の箇所とその内容を記録した。さらに、不一致があった箇所について1名の作業者が判定を行い、最終的な修正結果を一意に定めた。また、1次作業、比較統合作業、最終判定作業のいずれにおいても、音声を参照した。

さらに、人手修正作業において修正を行った場合、どのような理由でそのような修正操作を行ったのかを記録するために、その修正項目名を「義務的コメント」として記録しておくことにした。図 5.11 で定義されている修正項目のうち、どの項目に該当する修正を行ったのかを、当該の位置に「 ; 体言止」のような形で記録した^{*13}。

5.5 人手修正規則の一覧

以下では、人手修正規則の詳細と具体的な修正例について、修正項目ごとに示す。

5.5.1 大局的な文法構造に関するもの

5.5.1.1 体言止

話し言葉の中には、単独の名詞によって発話が完了している場合や、名詞述語に後続すべき助動詞（だ、です等）が省略される場合（いわゆる体言止）など、形態的に有標でないさまざまな発話の完了点が存在する。以下の例では、「十三人」「大変」の直後に絶対境界に相当する大きな発話の切れ目があるが、CBAP-csjで検出することは不可能である。

- (134) 被験者は全部で十三人で<接続詞>システムが何か聞いては<テハ節>答えるっていうのの繰り返し
ですね [文末候補] (A01M0021)
- (135) 金曜日の夜の終電中央林間行きとか乗っちゃうと<条件節ト>もう凄い大変もうそれぐらいたくさん
の人が住んで<テ節>大手町とか都心に働きに行ってるんですね [文末] (S03F0062)

そこで、単独の名詞によって発話が完了していると判断された箇所には、「体言止」を認定する。当該箇所の直後に「-」をマークして発話を切断した上で、「 ; 体言止」を義務的コメントとして残す。

^{*13} ここでは見やすさを考慮して、義務的コメントを「 ; 体言止」のような形で記述するが、実際のXML文書中には「 ; 」の記号は付与されていない。

- (136) 1. 被験者は全部で十三人 - ; 体言止
 2. で<接続詞>システムが何か聞いては<テ八節>答えるっていうのの繰り返しですね [文末候補]
 (A01M0021)
- (137) 1. 金曜日の夜の終電中央林間行きとか乗っちゃうと<条件節ト>もう凄い大変 - ; 体言止
 2. もうそれぐらいたくさんの人が住んで<テ節>大手町とか都心に働きに行ってるんですね [文末]
 (S03F0062)
- (138) 1. 赤い線が音響尤度のみによる検証 - ; 体言止
 2. それに加えて<テ節>その青いラインが母音列を用いた検証 - ; 体言止
 3. 緑色のラインが誤認識傾向に基づく検証です [文末] (A01M0103)

また、発話冒頭に「はい」「ええ」のような<感動詞>が現れている場合、そこだけが単独の発話を構成する部分と見なす。この場合にも「体言止」を認定し、当該箇所を切断する。

- (139) 1. はい<感動詞> - ; 体言止
 2. タイトルは不連続きで<並列節デ>忍耐の一年だったっていう<トイウ節>ことで話したいと
 <引用節>思います [文末] (S02F0121)

さらに、タイトルなどが単独で発話されている場合や、箇条書きの記号や例文番号を読み上げている場合にも、「体言止」を認定して発話を切断する。

- (140) 1. テーマ - ; 体言止
 2. 今まで人生を振り返って<テ節>印象に残ったこと - ; 体言止
 3. タイトルは長男の嫁として<テ節> - ; 体言止
 4. 長男に嫁いだのが二十四歳の時でした [文末] (S05F1517)
- (141) 1. 四 - ; 体言止
 2. 太郎がゆっくり歩いている [文末]
 3. 五 - ; 体言止
 4. 太郎が枝をゆっくり折っている [文末] (A02M0107)

5.5.1.2 引用節構造

引用節やトイウ節で導かれる引用構造の内部に絶対境界・強境界が含まれる場合、引用の途中で発話が分割されるという不適切な結果が生じる。以下の例では、「寒い」「狭い」「寂しい」のいずれもがトイウ節に含まれる引用部分であるが、「寒い」「狭い」の直後に [文末] が付与されているため、そこが分割位置となっており、不適切な分割結果となっている。

- (142) 1. うちに遊びに来る友人達にはスリーエスと言われていて<テ節>寒い [文末]
 2. 狭い [文末]
 3. 寂しいという<トイウ節>風に言われています [文末] (S00F0177)

そこで、引用されている範囲全体を「{ }」で囲み、不適切に分割された箇所には「:」をマークして発話を結合する。また、「; 引用節構造」を義務的コメントとして記述する。なお、「{ }」で囲む引用の範囲は、音声および文脈に基づいて判断する。

- (143) うちに遊びに来る友人達にはスリーエスと言われていて<テ節> { 寒い [文末] : ; 引用節構造 狭い [文末] : ; 引用節構造 寂しい } という<トイウ節>風に言われています [文末] (S00F0177)
- (144) 母音の場合には { 強母音が弱化して<テ節>弱母音になって/テ節/ : ; 引用節構造 で<接続詞>更に<接続詞>それが落ちる } という<トイウ節>ことはかなり頻繁にあるんですけども /並列節ケレドモ/... (A05F0154)
- (145) 看護婦さんで<並列節デ>優しい声で { おめでとうございます [文末] : ; 引用節構造 女の子です [文末] : ; 引用節構造 お母さんも元気ですよ } って<引用節>言って/テ節/... (S01M0706)

絶対境界・強境界以外の位置で分割されるべき点が引用節の内部に入り込んでいる場合、これらの直後に「-」をマークして一旦発話を切断し、さらに「:」をマークして発話を再度結合する。この場合、「; 引用節構造-内部切断」を義務的コメントとして記述する。

- (146) 私最初これを見た時に { 一体どういうことなのかな<文末候補> - : ; 引用節構造-内部切断 何かアルバイトでもしてほしいという<トイウ節>ことなのかな } という<トイウ節>風に思いました [文末] (S06M0894)
- (147) イントネーションについて { 日本語のイントネーションどうあるべきなのか<間接疑問節> - : ; 引用節構造-内部切断 特に外国人教える場合どうなのか/間接疑問節/ : ; 引用節構造 あるいは<接続詞>日本人同士の正しい喋り方はどうなのか } と<引用節>いった場合のイントネーション論については殆ど言及がないんですね [文末] (A05M0002)

引用節中に「体言止」に相当する切れ目がある場合は「; 引用節構造-内部切断-体言止」、また、「と文末」に相当する切れ目がある場合は「; 引用節構造-内部切断-と文末」を、それぞれ義務的コメントとして記述する。

- (148) で<接続詞>このように分かりにくさを記述する為の情報として<テ節> { 分かる為にどういふ情報が必要かという<トイウ節>観点 - : ; 引用節構造-内部切断-体言止 それから<接続詞 L>問題の所在がどこにあるか [文末候補] : ; 引用節構造 それから<接続詞 L>分かりにくさの程度がどういふものがあるか } と<引用節>いったような幾つか観点が必要になってくるという<トイウ節>ことが分かってきました [文末] (A03F0108)
- (149) で<接続詞> { この部分がこういった形で出血すると<条件節ト>右半身不随と - : ; 引用節構造-内部切断-と文末 後左脳だったもので<並列節デ> (左左脳には言語中枢があるんですよ [文末]) + ; 挿入文-その他 で<接続詞>言語中枢のところで失語症になる可能性がある [文末] : ; 引用節構造 で<接続詞>言葉も結局うまく喋れないようになる [文末] : ; 引用節構造 で<接続詞>最悪車椅子の生活を強いられる } っていう<トイウ節>ことを聞いたんですね [文末] (S02F0189)

引用された部分の末尾に絶対境界があり、発話が不適切に分割されている場合、引用部分を「{ }」で囲む。ただし、引用内部の要素同士を結合するわけではないので、「:」ではなく「+」で結合する。この場合には、「; 引用節構造-末尾境界」を義務的コメントとして残す。

- (150) それで<接続詞 L> { もうこのエーさんとは終わったんだな [文末候補] } + ; 引用節構造-末尾境界 つつって/引用節/で<接続詞>振り返って見たら<条件節タラ>今言ってるように妄想とか単なる憧れ恋愛に対する憧れみたいの高一まで持っていた男の子みたいな感じなんですけれども/並列節ケレドモ/... (S01M0101)

なお、人手修正作業の目的は節単位の境界を確定することであるため、引用された範囲を「{ }」で囲むのは CBAP-csj の分割結果を修正する必要がある場合に限る。逆に言えば、出現したすべての引用節の範囲をマークするわけではない。以下の例における引用された範囲（「一度でもいいから行ってみたい」）については、CBAP-csj の分割結果を修正する必要がないため、「{ }」で囲む必要はない。

- (151) ここは昔からテレビで何度も見て<テ節>一度でもいいから<理由節カラ>行ってみたい と<引用節>強く思っていたところです [文末] (S01F0050)

5.5.1.3 連体節構造

連体節（連体修飾節）の内部に絶対境界・強境界が含まれる場合、連体節の途中で発話が分割されるという不適切な結果が生じる。以下の例では、1行目の「文の日本語らしさ」から2行目の「残した」までの範囲が連体節として「要約文」に係る構造になっているが、途中に/連用節/が含まれているため、発話が不適切に分割されている。

- (152) 1. つまり 文の日本語らしさを維持しており/連用節/
2. かつ<接続詞>意味的重要部分を残した 要約文を生成することができます [文末] (A03M0004)

そこで、連体節の範囲を「{ }」で囲み、不適切に分割された箇所「:」をマークして発話を結合する。ま

た、義務的コメントとして「 ; 連体節構造」を記述する。なお、「{ }」で囲む連体節の範囲は、音声および文脈に基づいて判定する。

- (153) つまり { 文の日本語らしさを維持しており/連用節/ : ; 連体節構造 かつ<接続詞>意味的重要部分を残した } 要約文を生成することができます [文末] (A03M0004)
- (154) で<接続詞> { いいえとか違いますしか答えない [文末] : ; 連体節構造 それ以外何も言わない } 入っているのは大体四分の一しかいませんでした/テ節/... (A01M0021)

絶対境界・強境界以外の位置で分割されるべき点が連体節の内部に入り込んでいる場合、これらの直後に「-」をマークして一旦発話を切断し、さらに「:」をマークして発話を再度結合する。「 ; 連体節構造-内部切断」を義務的コメントとして残す。

- (155) 私が行った先というのは { お父さんお母さん - : ; 連体節構造-内部切断 お父さんは先生なんです
が/並列節ガ/ : ; 連体節構造 お母さんは専業主婦で/並列節デ/ : ; 連体節構造 で<接続詞>子供が
女七人上の七人が女で<接続詞>下二人が男 } の合計十一人家族でした [文末] (S00F0083)

5.5.2 非流暢性に関するもの

5.5.2.1 倒置

絶対境界・強境界で分割された点の直後の要素が、直前の節内の用言に係る関係にあると判断される場合、「倒置」の構造が発生していると見なす。以下の例の3行目の「待ってる人と」は、その直前の節の述語「目が合うんですよ」に係る要素であるため、倒置された要素であると判断される。

- (156) 1. んで<接続詞>落ち着いて<テ節>したんですけど/並列節ケド/
2. そしたらですね<条件節タラ>目が合うんですよ [文末]
3. 待ってる人と で<接続詞>凄いプレッシャー掛けてくるような気がするんですよ [文末] (S05F1600)

倒置の構造が発生している場合、係り受け関係を結ぶはずの要素間が分断されるという不適切な結果が生じる。そこで、倒置された要素を「<< >>」で囲んだ上で、その直後を「-」で切断し、直前の絶対境界・強境界の切れ目を「+」で結合する。切断した点と結合した点のそれぞれに、「 ; 倒置-つなぎ切り」を義務的コメントとして記述する。なお、「<< >>」で囲む倒置要素の範囲は、音声および文脈に基づいて判定する。

- (157) 1. んで<接続詞>落ち着いて<テ節>したんですけど/並列節ケド/
 2. そしたらですね<条件節タラ>目が合うんですよ [文末] + ; 倒置-つなぎ切り <<待ってる人と>>
 - ; 倒置-つなぎ切り
 3. で<接続詞>凄いプレッシャー掛けてくるような気がするんですよ [文末] (S05F1600)
- (158) 1. 一軒家の借家を借りてくれまして/テ節/ + ; 倒置-つなぎ切り <<あたしの父親が>> - ; 倒置-
 つなぎ切り
 2. で<接続詞>そこで半年ぐらい住んでまして/テ節/... (S05F1517)

弱境界の直後の要素が直前の節の用言に係り、かつその要素の直後で発話が分割されるのが自然であると判断される場合、倒置された要素を「<< >>」で囲み、直後を「-」で切断する。義務的コメントとして、「; 倒置-後切り」を記述する。

- (159) 1. 外車ですからね<理由節カラ> <<何せ>> - ; 倒置-後切り
 2. ですんで<接続詞 L>もう何せよく分かんないことになってるんですよ [文末] (S01M0227)
- (160) 1. 後はもう一度カード会社に連絡をして<テ節>伝票を取り寄せて<テ節> <<一切公開しないとい
 う<トイウ節>約束で>> - ; 倒置-後切り
 2. んで<接続詞>その伝票と筆跡を合わせていったんですよ [文末] (S02F0121)

節境界ではないが発話の切れ目となる箇所の直後の要素が直前の節の用言に係り、かつその要素の直後で発話が分割されるのが自然であると判断される場合にも、倒置された要素を「<< >>」で囲み、直後を「-」で切断する。「; 倒置-後切り」を義務的コメントとして記述する。

- (161) 1. 母おしゅうとめさんが四十八歳 <<二回り上の>> - ; 倒置-後切り
 2. で<接続詞>しゅうとが五十一歳です [文末] (S05F1517)

5.5.2.2 挿入節

発話中に発話プランが変更されることにより、発話の途中で「～ですけども」「～ですが」のような形の「但し書き」が急遽挿入されることがある。この但し書きとして挿入された部分が強境界の形を取り、かつ挿入部分の前後に係り受け関係を持つ要素がある場合、不適切な分割結果が生じる。以下の例の「ここに書いてある数字は頻度ですが」は発話の途中に急遽挿入された部分であり、かつそれが強境界の形を取っているために、1行目の「色んなパターンを」とその係り先である2行目の「集めてみました」が分断されてしまっている。

- (162) 1. 色んなパターンをここに書いてある数字は頻度ですが/並列節ガ/
 2. たくさん集めてみました [文末] (A01M0021)

このような挿入によって係り受け関係が分断されてしまっている場合に限り、「挿入節」を認定する。挿入節として認めるのは、原則として、/並列節ガ/, /並列節ケレドモ/, /並列節ケレド/, /並列節ケドモ/,

/並列節ケド/のいずれかの形を取っているものに限る。挿入された部分を「()」で囲み、不適切に分割された箇所「+」をマークして発話を結合する。義務的コメントとしては、「 ; 挿入節」を記述する。「()」で囲む挿入節の範囲は、音声および文脈に基づいて判断する。

- (163) 色んなパターンを (ここに書いてある数字は頻度ですが/並列節ガ/) + ; 挿入節 たくさん集めて
みました [文末] (A01M0021)
- (164) 成城やら久我山の方から多くの買い物客 (普通にスーパーにお買い物するお客さんなんですけれど
も/並列節ケレドモ/) + ; 挿入節 そういった方が凄く集まって<テ節>集客する場所らしいんです
ね [文末] (S03F0224)
- (165) 高校に入りましたら<条件節タラ> (中学でもやっておったんですけれども/並列節ケレドモ/) + ;
挿入節 野球をそのまま続けまして/テ節/... (S02M0092)
- (166) 上海って (多分皆さん御存じだと<引用節>思うんですけど/並列節ケド/) + ; 挿入節 凄くもう
近代的な奇麗なところなんですよ [文末] (S05F1600)

5.5.2.3 挿入文

発話中に発話プランが変更されることにより、発話の途中で「～です」「～ですね」のような形の「但し書き」が急遽挿入されることがある。この但し書きとして挿入された部分が絶対境界の形を取り、かつ挿入部分の前後に係り受け関係を持つ要素がある場合、不適切な分割結果が生じる。以下の例では、「中森明菜」を発話した直後にその補足として「当時の松田聖子さんと双肩結構双壁をなすアイドルですね」という発話が急遽挿入され、その直後が発話分割位置となっている。この分割によって、1行目の「彼はですね」とその係り先「やっておりました」が分断されてしまっている。

- (167) 1. 彼はですね中森明菜当時の松田聖子さんと双肩結構双壁をなすアイドルですね [文末候補]
2. で<接続詞>こちらの追っ掛けをやっておりました [文末] (S02M0092)

このような挿入によって係り受け関係が分断されてしまっている場合、「挿入文」を認定する。挿入文として認めるのは、原則として、絶対境界の形を取っているものに限る。挿入された部分を「()」で囲み、不適切に分割された箇所「+」をマークして発話を結合する。義務的コメントとしては、挿入部分が「～です」のようなコピュラ文の文末表現を取っている場合は「 ; 挿入文-判定詞」を記述する。「()」で囲む挿入節の範囲は、音声および文脈に基づいて判断する。

- (168) 彼はですね中森明菜（当時の松田聖子さんと双肩結構双壁をなすアイドルですね [文末候補] + ; 挿入文-判定詞）で<接続詞>こちらの追っ掛けをやっておりました [文末] (S02M0092)
- (169) で<接続詞>こちら訓練の一回目（こちらですね [文末候補]）+ ; 挿入文-判定詞 では五十パーセントから八十三パーセントぐらいの弁別率の範囲だったんですが/並列節ガ/...
- (170) そいで<接続詞 L>やはり妻との対話も（うちの母親ですね [文末候補]）+ ; 挿入文-判定詞 対話も少なく我が家庭内も暗かったという<トイウ節>印象があります [文末] (S05M0412)

また、挿入部分がコンピュータ文の文末表現以外の形を取っている場合は、「; 挿入文-その他」を義務的コメントとして記述する。

- (171) 1. 空間音響特性依存エイチエムエムの評価方法ですが/並列節ガ/
2. まず音源とマイクロホンとの距離が音響モデルの学習時とその音声の認識時で同じ場合（これをマッチドエイチエムエムと呼びます [文末]）+ ; 挿入文-その他 そして<接続詞>最大尤度基準に基づき<連用節>空間音響特性依存エイチエムエムを選択して<テ節>認識した場合（これをライクリーフッドセレクションと呼びます [文末]）+ ; 挿入文-その他 そして<接続詞>まず音響特性を学習していないエイチエムエムで認識した場合アイピーイーエイチエムエムの三種類の認識率で比較します [文末] (A01M0151)

さらに、発話の途中で逡巡している場合に挿入される発話に対しても、「挿入文」を認定する。

- (172) テニスコートも（何面ですかね [文末候補]）+ ; 挿入文-判定詞 かなりの面はあります [文末] (S03M0003)
- (173) ところがですね<接続詞 L>まもなく（あれはどこの辺だったでしょうか [文末候補]）+ ; 挿入文-判定詞（静岡沖辺りを通過した辺りからでしょうか [文末候補]）+ ; 挿入文-判定詞 船が揺れ始めました [文末] (S02M0011)

5.5.2.4 フィラー文

CBAP-csjによって<フィラー文>が付与されている箇所以外にも、「何て言うんですかね」「何て言うんでしょう」「そうですね」「すみません」のように、実質的な意味をほとんど持たず、それ単独でフィラーのような働きをする発話が挿入される場合がある。このような挿入部分の前後に係り受け関係を持つ要素がある場合、不適切な分割結果が生じる。以下の例では、「何て言うんですか」というフィラーのような働きを持つ発話が挿入されることにより、1行目の「うちの主人も」とその係り先「離れ難かった」が分断されてしまっている。

- (174) 1. で<接続詞>結局うちの主人も何て言うんですか [文末候補]
2. 離れ難かったのかもしれませんがね/並列節ケド/ (S03F0314)

このような分割結果によって係り受け関係が分断される場合に限り、挿入された部分を「（ ）」で囲み、不

適切に分割された箇所「+」をマークして発話を結合する。義務的コメントとしては「; フィラー文」を記述する。

- (175) で<接続詞>結局うちの主人も (何て言うんですか [文末候補]) + ; フィラー文 離れ難かったの
 かもしれませんけどね/並列節ケド/... (S03F0314)
- (176) その先にボンアートというちょっとややいかがわしいですね<文末候補> (何でしょうか [文末候補])
) + ; フィラー文 画材屋さんのようながあります [文末] (S03M0141)
- (177) 自分はどれだけ悪い回収をしてたんだという<トイウ節>ことを (そうっすね [文末候補]) + ;
 フィラー文 初めて気が付きました [文末] (S00M0221)
- (178) で<接続詞>これら語頭の子音クラスター内のアールエルの調音についての<連体節テノ> (すいま
 せん [文末]) + ; フィラー文 全話者の調音地図がこちらですが/並列節ガ/... (A05M0031)

5.5.2.5 言いさし

発話中に発話プランが変更されたり、発話が途中で言いやめられたりすることで、係り先のない要素ができる場合がある。以下の例の音声を見ると、「被験者群による」と言いかけてその後別の発話を始めてしまったため、「被験者群による」の係り先が消失してしまっていることが分かる。

- (179) 1. 強さ条件でもピッチパターンの効果は有意でした [文末]
 2. 被験者群による 同様に強さパターンの効果が見られました [文末]
 3. また<接続詞>実験三の非言語音でもああの言語音とほぼ同様の結果が得られました [文末]
 (A01F0122)

そこで、発話プランの変更によって係り先が消失してしまった要素を「言いさし」として認定する。言いさされた要素の直後を「-」で切断し、「; 言いさし-係り先なし」を義務的コメントとして記述する。

- (180) 1. 強さ条件でもピッチパターンの効果は有意でした [文末]
 2. 被験者群による - ; 言いさし-係り先なし
 3. 同様に強さパターンの効果が見られました [文末]
 4. また<接続詞>実験三の非言語音でもああの言語音とほぼ同様の結果が得られました [文末]
 (A01F0122)

言いさされた要素は、後になって同じ形が言い直されるものも多い。言いさされた要素と後方の言い直された要素との間に絶対境界・強境界がある場合、言いさされた要素には「; 言いさし-言い直しあり」を義務的コメントとして記述する。

- (181) 1. こちらは論文の方にも掲載させていただいておりますけれども/並列節ケレドモ/
 2. 集団エーというのが - ; 言いさし-言い直しあり
 3. これ一番下になりますけれども/並列節ケレドモ/
 4. 集団エーというのがですね本システムを利用した集団です [文末] (A04M0026)
- (182) 1. 五か月児 - ; 言いさし-言い直しあり
 2. それぞれこちらが原型旋律の平均聴取時間平均それから<接続詞 L>こちらの白い方が変形旋律の聴取時間平均です [文末]
 3. で<接続詞>この有意差を見ましたところ五か月児八か月児のいずれにおいても<テモ節>有意差は認められませんでした [文末] (A01F0055)
- (183) 1. ここの目玉はシャークリーフと言いまして/テ節/
 2. シュノーケルで - ; 言いさし-言い直しあり
 3. 海の中とそっくりに作られた大きいプールがありまして/テ節/
 4. そこにシュノーケルで入るんですけども/並列節ケレドモ/... (S01F0050)

なお、言いさされた要素と後方の言い直された要素との間に絶対境界・強境界がない場合は、人手修正の対象とはしない。以下の例では、発話冒頭で言いさされた要素「正解率」は後方で言い直されており、かつそこまでの間に絶対境界・強境界が存在しないため、人手修正として操作すべき項目はない。

- (184) 正解率今回の提案手法であります音素で見た時の正解率がこのようになっております [文末] (A01M0096)

5.5.2.6 言い直しマーカー

「～と言いますか」「～と言うんでしょうか」などの形で発話が一旦中断し、その後言い直しの表現を含む形で発話が継続される場合がある。以下の例では、1行目の「パート・オブ・スピーチって言いますか」は2行目の「品詞」に言い直すための表現であり、[文末候補]として発話の切れ目と見なすのは適切ではない。

- (185) 1. で<接続詞>前回の発表で単純に パート・オブ・スピーチって言いますか[文末候補]
 2. 品詞的な情報を使うよりはもう少しこう用法的な情報を使うと<条件節ト>うまく行くという
 <トイウ節>話をしてたんですが/並列節ガ/... (A01M0157)

このような言い直し表現の前後で発話が不自然に分割されてしまっている場合、当該箇所を「+」で結合し、義務的コメント「 ; 言い直しマーカー」を記述する。

- (186) で<接続詞>前回の発表で単純にパート・オブ・スピーチって言いますか [文末候補] + ; 言い直し
 マーカー 品詞的な情報を使うよりはもう少しこう用法的な情報を使うと<条件節ト>うまく行くとい
 う<トイウ節>話をしてたんですが/並列節ガ/... (A01M0157)
- (187) ある何かのセグメントって言うんですかね [文末候補] + ; 言い直しマーカー まとまりに対して
 <テ節>以下のステップを行ないます [文末] (A01M0070)
- (188) で<接続詞>これをビデオと言うんでしょうか [文末候補] + ; 言い直しマーカー 映像と音声の形で将
 来的には公開いたしますので<理由節ノデ>是非御研究の何かの参考にさせていただければと<引用節>
 思っております [文末] (A05M0002)

5.5.2.7 述語の言い直し

絶対境界や強境界の形を持つ述語句部分が言い直されている場合、言い直しの途中で発話が分割されるとい
 う不適切な結果が生じる。以下の例の音声を見ると、述語句の「外れます」が「飛びます」に言い直されてい
 ることが分かるが、「外れます」の直後に [文末] が付与されているため、そこが分割位置となり、不適切な分
 割結果が生じている。

- (189) 1. 後は話がちょっと外れます [文末]
 2. 飛びますが/並列節ガ/... (S03M0003)

そこで、述語句の言い直しによって発話が不適切に分割されている場合、「述語の言い直し」を認定する。該
 当箇所には「+」をマークして発話を結合し、義務的コメントして「; 述語の言い直し」を記述する。当該の表
 現が述語句の言い直しになっているかどうかは、音声および文脈に基づいて判定する。

- (190) 後は話がちょっと外れます [文末] + ; 述語の言い直し 飛びますが/並列節ガ/... (S03M0003)
- (191) この結果を次にします [文末] + ; 述語の言い直し 示します [文末] (A01F0145)
- (192) 調査調査アンケート調査の概要については資料集の方を御覧ください [文末] + ; 述語の言い直し 予
 稿集の方を御覧ください [文末] (A06F0120)

5.5.3 係り受け構造に関するもの

5.5.3.1 主題の共有

係助詞「は」や「も」でマークされる主題要素は、強境界の前後において複数の係り先を持つという、広い
 スコープ構造を取る場合がある。次の例において、主題要素「聴取者は」は、1行目の「聞き取り」だけでなく
 2行目の「理解しなければなりません」にも係っていると解釈できる。これは、「聞き取り」と「理解し」とい
 う2つの述語句が並列の関係にあるためである。係り受け構造情報の付与作業を考えると、これらの述語句は
 1つの節単位の内部に納まっている方が望ましい。

- (193) 1. このような場合聴取者は注目した一方の音声を正確に聞き取り/連用節/
2. また<接続詞>確実に内容を理解しなければ<条件節レバ>なりません [文末] (A01M0110)

そこで、分割された前後において主題要素が共有されていると判断された場合、強境界で分割された箇所「+」をマークして発話を結合する。義務的コメントとして「; 主題の共有」を記述する。

- (194) このような場合聴取者は注目した一方の音声を正確に聞き取り/連用節/ + ; 主題の共有 また
<接続詞>確実に内容を理解しなければ<条件節レバ>なりません [文末] (A01M0110)
- (195) 実音の確定というのは本当にさまざまな解釈がございまして/テ節/ + ; 主題の共有 大変難しゅうござ
います [文末] (A02F0116)
- (196) で<接続詞>この尤度の上がりも大きいところもあるし/並列節シ/ + ; 主題の共有 こう小さいところ
もあると [と文末] (A01M0070)
- (197) 富士山を見ながらの生活というのはこれまでも経験ありませんし/並列節シ/ + ; 主題の共有 想像
もしてこなかったことでして/テ節/ + ; 主題の共有 大変気に入ってます [文末] (S03M0098)

5.5.3.2 主題の飛び越し

係助詞「は」や「も」でマークされる主題要素は、強境界で分割された後方の節内にもみ係り先を持つ場合がある。次の例において、主題要素「こちらの町は」は2行目の「東京都東村山市秋津町です」に対して係る要素であり、1行目の「所在地で言いますと」には係り先を持っていない。

- (198) 1. こちらの町は所在地で言いますと/条件節ト/
2. 東京都東村山市秋津町です [文末] (S03M0201)

このように、主題要素が強境界を飛び越して係り先を持っていると判断された場合、強境界で分割された箇所に「+」をマークして発話を結合する。義務的コメントとして「; 主題の飛び越し」を記述する。

- (199) こちらの町は所在地で言いますと/条件節ト/ + ; 主題の飛び越し 東京都東村山市秋津町です [文末]
(S03M0201)
- (200) で<接続詞>セット内の刺激提示の順序は順序効果を考慮しまして/テ節/ + ; 主題の飛び越し ラン
ダムといたしました [文末] (A05M0031)
- (201) 今日は簡単ですが/並列節ガ/ + ; 主題の飛び越し コソボ問題について<テ節>述べさせていただきました
ました [文末] (S00M0199)
- (202) で<接続詞>ここは毎晩十二時になりますと/条件節ト/ + ; 主題の飛び越し カウントダウンショー
が行なわれます [文末] (S01F0050)

5.5.3.3 格要素の飛び越し

格助詞でマークされる格要素が、強境界で分割された後方の節内にのみ係り先を持つ場合がある。次の例において、格要素「音声が」および「ヘッドホンに」は2行目の「出力される」に対して係る要素であり、1行目の内部には係り先を持っていない。

- (203) 1. で<接続詞>システム側の音声が被験者のヘッドホンに音声合成を使いまして/テ節/
2. 出力されるという<トイウ節>ことになっています [文末] (A11M0846)

このように、格要素が強境界を飛び越して係り先を持っていると判断された場合、強境界で分割された箇所に「+」をマークして発話を結合する。義務的コメントとして「; 格要素の飛び越し」を記述する。また、格助詞「の」でマークされる連体句や、副詞のような格助詞を伴わない連用修飾句が強境界を飛び越して係り先を持つ場合にも、同様の修正を行う。

- (204) で<接続詞>システム側の音声が被験者のヘッドホンに音声合成を使いまして/テ節/ + ; 格要素の飛び越し 出力されるという<トイウ節>ことになっています [文末] (A11M0846)
- (205) この確率付き項構造を利用した曖昧性解消の小規模ですが/並列節ガ/ + ; 格要素の飛び越し 予備実験を行ないましたので<理由節ノデ>その結果を報告したいと<引用節>思います [文末] (A03M0061)
- (206) やっぱりとても坂道もないし/並列節シ/ + ; 格要素の飛び越し 住み易いところなんだと<引用節>思います [文末] (S03F0133)

5.5.3.4 連体形

5.2.2 節で述べたように、CBAP-csj への入力には、短単位から「基本形」「品詞」「その他の情報 1」「活用形」「活用の種類」「その他の情報 2」という6つの情報を取り出した「MOR形式」が入力される。このうち、活用形が「終止形」と認定された述語について、大域的な談話レベルの構造や係り受け構造情報の付与を考えると、「連体形」と見なす方がよいと思われるものがある。以下の例において、1行目の「飛行機が通る」の「通る」は「終止形」と認定されており、その直後に [文末] が認定されている。

- (207) 1. で<接続詞>わりとその拝島駅の北側の真上を飛行機が通る [文末]
2. 何て言うか<文末候補>道道空の道が飛行機が通る道になってるんで<理由節ノデ>もう飛行機の爆音が凄いうるさいです [文末] (S03F0184)

しかしながら、2行目には「道道空の道が飛行機が通る道に」という言い直された表現があることから、そもそも1行目の「飛行機が通る」は、2行目の「道」を修飾する表現として発話されていたと推測される。局所的に見ると1行目の「通る」は終止形として認定されるが、係り受け構造情報の付与にとっては、1行目と2行目の発話はつながっていた方が都合がよい。

そこで、述語の活用形が「終止形」となっていることによって発話が分割され、かつ大域的な談話レベルの

構造を考えた時にその分割が望ましくないと判断された場合に限り、人手で「連体形」を認定する。終止形の直後で分割された箇所に「+」をマークして発話を結合し、義務的コメントとして「; 連体形」を記述する。

- (208) で<接続詞>わりとその拜島駅の北側の真上を飛行機が通る [文末] + ; 連体形 何て言うか
 <文末候補>道道空の道が飛行機が通る道になってるんで<理由節ノデ>もう飛行機の爆音が凄
 うるさいです [文末] (S03F0184)
- (209) で<接続詞>曲の提示順先程あった長調六曲短調六曲の提示順それから<接続詞 L>音源の左右右か
 ら出たり<タリ節>左から出たりと<引用節>いう [文末] + ; 連体形 (半分ずつ出るんですけど
 も/並列節ケレドモ/) + ; 挿入節 その左右の順は全て被験時間それから<接続詞 L>月齢間でカウ
 ンターバランスを取りました [文末] (A01F0055)

5.5.4 談話構造上の問題に関するもの

5.5.4.1 話題導入表現

局所的な節境界の検出としては正しいものの、大域的な談話レベルの構造を考えた場合に、CBAP-csjによる分割処理がふさわしくない結果を生じさせている場合がある。次の例における談話構造を考えてみると、2行目の「ワイ君はどういう人かと言うと」という表現は、それ以降の話題（ワイ君に関する説明）を導入する機能を果たしており、それに対して「凄く頭がよくて」「凄くおとなしいんですけど」「超面白い感じで」などが並列的に叙述される、という格好になっている。

- (210) 1. 当時小学校六年生の時に一学期に隣の席のワイ君のことが好きになったんですが/並列節ガ/
 2. で<接続詞>ワイ君はどういう人かと<引用節>言うとか<条件節ト>凄く頭が良くて/テ節/
 3. で<接続詞>凄くおとなしいんですけど/並列節ケド/
 4. 超面白い感じで/並列節デ/
 5. で<接続詞>女子にも隠れて<テ節>人気があったけど/並列節ケド/
 6. どっちかって<引用節>言うとか<条件節ト>男子に人気があって<テ節>慕われてるとい
 う<トイウ節>感じで/並列節デ/... (S02F0094)

談話境界情報の付与作業においては、「ワイ君はどういう人かと言うと」という部分を話題の導入表現として認定するために、その部分が独立した単位として切り離されていた方が都合がよい。そこで、大域的な談話レベルの構造において話題導入を行っているとして解釈できる部分を「話題導入表現」として認定する。話題導入表現の直後を「-」で切断し、「; 話題導入表現」を義務的コメントとして記述する。

- (211) 1. 当時小学校六年生の時に一学期に隣の席のワイ君のことが好きになったんですが/並列節ガ/
 2. で<接続詞>ワイ君はどういう人かと<引用節>言うと<条件節ト> - ; 話題導入表現
 3. 凄く頭が良くて/テ節/
 4. で<接続詞>凄いいおとなしいんですけど/並列節ケド/
 5. 超面白い感じで/並列節デ/
 6. で<接続詞>女子にも隠れて<テ節>人気があったけど/並列節ケド/
 7. どっちかって<引用節>言うと<条件節ト>男子に人気があって<テ節>慕われてるとい
 う
 <トイウ節>感じで/並列節デ/... (S02F0094)

話題導入表現として認定される箇所は、上記のような<条件節ト>だけでなく、<並列節デ>、<テ節>、主題要素などの場合もある。

- (212) 1. ビーグループとシーグループの比較も同じで<並列節デ> - ; 話題導入表現
 2. ビーグループの方が全体の正答率が高くオーエヌエーエヌの区別が多少できてますが/並列節ガ/
 3. シーグループの方はアイエヌの正答率が高くなります [文末]
 4. つまりオーエヌエーエヌの区別ができると<条件節ト>アイエヌの正答率が下がるという
 <トイウ節>結果が出ました [文末] (A05F0043)
- (213) 1. まとめといたしまして<テ節> - ; 話題導入表現
 2. 音声の組み合わせによって<テ節>その聞き分け易さが大きく異なることが分かりました [文末]
 3. また<接続詞>被験者と再生用音源の設置位置関係が聞き分け易さに影響を及ぼすことも確認い
 たしました [文末]
 4. なお<接続詞>日本語と英語の組み合わせによるメッセージの聞き分け易さの評価が良好であっ
 たことは目的の意味から考えて<テ節>大変意義深いと<引用節>... (A01M0110)
- (214) 1. うちに来た経緯というのは - ; 話題導入表現
 2. その獣医さんの紹介なんですけれども/並列節ケレドモ/
 3. 捨て犬を保護してくれというのがその獣医さんのところに来まして/テ節/
 4. その獣医さんが (R × ×) さんちでどうですか<文末候補>みたいな話をしたんですけれども
 /並列節ケレドモ/
 5. それで<接続詞 L>それで<接続詞 L>その犬を見に行っただけなんですけれども... (S02M0198)

5.5.4.2 直後がまとめ表現

話題導入表現とは逆のパターンとして、ある話題に関する叙述が続いた上で、最後にそこまでの叙述を受けて話題をまとめるための表現が現れる場合がある。次の例における談話構造を考えると、ある話題（町の住みやすさ）に関する叙述として、1行目の「わざわざ～買えますし」、2行目の「魚屋さんも～ありましたから」、3行目の「安くありましたから」などが続き、3行目の「もうとても便利でした」によって一連の話題がまとめられている。

- (215) 1. わざわざ新宿の方まで買いに行かなくても<テモ節>そこで十分色々な新鮮な野菜も買えますし
/並列節シ/
2. 魚屋さんもお肉屋さんもしっかりとありましたから/理由節カラ/
3. で<接続詞>安くありましたから<理由節カラ>もうとても便利でした [文末] (S03F1577)

話題のまとめ表現として働いている「もうとても便利でした」は、それ自体が独立した単位として切り離されていた方が、談話境界情報の付与作業上、都合がよい。そこで、大域的な談話レベルの構造において、話題のまとめを行っている部分を「まとめ表現」として認定する。まとめ表現の直前を「-」で切断し、「;直後がまとめ表現」を義務的コメントとして記述する。

- (216) 1. わざわざ新宿の方まで買いに行かなくても<テモ節>そこで十分色々な新鮮な野菜も買えますし
/並列節シ/
2. 魚屋さんもお肉屋さんもしっかりとありましたから/理由節カラ/
3. で<接続詞>安くありましたから<理由節カラ> - ;直後がまとめ表現
4. もうとても便利でした [文末] (S03F1577)

まとめ表現として認定されるのは、<テ節>、<並列節デ>、<理由節ノデ>などの直後が多い。

- (217) 1. 大きい丸い石が転がってきて<テ節>逃げ惑うスタントの人 - ;体言止
2. そして<接続詞>プロペラ機が出てきて<テ節>プロペラ機が大爆発して<テ節>炎に包まれます
[文末]
3. で<接続詞>その炎がもうこっちに移りそうで<並列節デ>熱くて<テ節> - ;直後がまとめ表現
4. こんなショー日本では不可能だと<引用節>思いました [文末] (S01F0050)
- (218) 1. で<接続詞>私達はとにかく一番安いツアーに行ってるものですから<理由節カラ>部屋もそんなに良くはなかったんだと<引用節>思うんですけども/並列節ケレドモ/
2. それでも<接続詞 L>もう普段とても狭いところで暮らしてる私達にしてみれば<条件節レバ>もう何か本当にリゾートと言うか<文末候補>バカンスっていう感じで<並列節デ> - ;直後がまとめ表現
3. とてもいいところだったなという<トイウ節>印象でした [文末] (S00F0014)
- (219) 1. やはり中には産地直送なんつって<テ節>八百屋さんなんかありますので/理由節ノデ/
2. で<接続詞>凄くもう新鮮だし/並列節シ/
3. 食べるものも野菜とか果物も新鮮で<並列節デ>安いので<理由節ノデ> - ;直後がまとめ表現
4. やはり買い物にはもう事欠かなかったです [文末] (S03F1577)

5.5.4.3 話題の転換点

ある話題に関する叙述が続いた後、絶対境界・強境界以外の位置で、話題が転換されていると認められる場合がある。次の例では、「遊んだ内容」という話題に関して「隠れん坊」「缶蹴り」「おままごと」に関する叙述

が並列的に並べられている。その談話的な構造を考えてみると、「缶蹴りですとか」という表現の直後で別の叙述内容（「おままごと」）に移っていることが分かる。

- (220) 1. 青山墓地の中は本当絶好の隠れん坊をする場所として<テ節>凄くみんながよく隠れん坊をいたしましたし/並列節シ/
 2. 後は缶蹴りですとか<ト力節>それから<接続詞 L>今もちょうど桜の時期ですけれども/並列節ケレドモ/
 3. お花の木の下でおままごとをしたり<タリ節>ですとか<ト力節>... (S03F0108)

形式的には発話の分割点にはなり得ないものの、話題が移っていると判断される箇所が発話が分割されている方が、談話境界情報の付与作業上、都合がよい。そこで、大域的な談話レベルの構造において、話題の転換点となっている箇所を「話題の転換点」として認定する。話題の転換点となっている箇所を「-」で切断し、「; 話題の転換点」を義務的コメントとして記述する。

- (221) 1. 青山墓地の中は本当絶好の隠れん坊をする場所として<テ節>凄くみんながよく隠れん坊をいたしましたし/並列節シ/
 2. 後は缶蹴りですとか<ト力節> - ; 話題の転換点
 3. それから<接続詞 L>今もちょうど桜の時期ですけれども/並列節ケレドモ/
 4. お花の木の下でおままごとをしたり<タリ節>ですとか<ト力節>... (S03F0108)
- (222) 1. 告発して<テ節>暫くたってから<テカラ節>事件は全て一応解決したんだけど/並列節ケド/
 2. やっぱりお財布とかに結構気を付けるようになったっていうのと - ; 話題の転換点
 3. 後何かこうそれまでは人をあんまり疑うこととかはなかったのに<ノ二節>この人ってってやっぱり一緒に仕事してても<テモ節>そういうことがある訳だから<理由節カラ>やっぱり考えるようになっちゃった [文末] (S02F0121)

5.5.4.4 大きい切れ目

談話の構造を大域的に見たとき、<テ節>、<並列節デ>、<理由節ノデ>などの弱境界の直後に、発話の進行上、大きい切れ目が存在する場合がある。次の例では、2行目で「坂が多い」という話題を出し、3行目の前半（<並列節デ>までの部分）で区役所が坂の本を出していることについて説明した後、後半で「そんな本出されても困る」という主観的なコメントを述べ、4行目以降ではまた本の説明に戻っている。

- (223) 1. そんなような感じで<並列節デ>非常に落ち着いた町なんですけど/並列節ケド/
 2. で<接続詞>坂が多いって<引用節>さっき言ったんですけど/並列節ケド/
 3. それは本当にもう坂の町って何か区役所の方が文京の坂道とかいう本を出してるぐら
 いで<並列節デ>そんな本出されても<テモ節>困るんだけどって<引用節>思うんですけど
 /並列節ケド/
 4. なかなかそれなりの厚さのある小冊子で/並列節デ/
 5. で<接続詞>坂の名前とか場所がずっと書いてあるっていうので/理由節ノデ/... (S03M0046)

ここで、3行目の前半「それは～本を出してるぐらいで<並列節デ>」から、3行目の後半「そんな本出されても～思うんですけど/並列節ケド/」に対する係り受け関係は存在しない。発話の大域的な流れを考えると、両者は別々の部分を構成しており、談話の構造上、そこには大きい切れ目があると考えられる。

そこで、大域的な談話レベルの構造において、弱境界の直後が発話の進行上大きい切れ目となっており、かつその前後に係り受け関係が存在しない場合、その弱境界の直後を「大きい切れ目」として認定する。当該箇所直前を「-」で切断し、「; 大きい切れ目-係り先なし」を義務的コメントとして記述する。

- (224) 1. そんなような感じで<並列節デ>非常に落ち着いた町なんですけど/並列節ケド/
 2. で<接続詞>坂が多いって<引用節>さっき言ったんですけど/並列節ケド/
 3. それは本当にもう坂の町って何か区役所の方が文京の坂道とかいう本を出してるぐら
 いで
 <並列節デ> - ; 大きい切れ目-係り先なし
 4. そんな本出されても<テモ節>困るんだけどって<引用節>思うんですけど/並列節ケド/
 5. なかなかそれなりの厚さのある小冊子で/並列節デ/
 6. で<接続詞>坂の名前とか場所がずっと書いてあるっていうので/理由節ノデ/... (S03M0046)
- (225) 1. また<接続詞>買っただけの同じ割合で両替える時も戻ってくるっていう/トイウ節/
 2. んで<接続詞>そこの常連さんなんか一月五六十万擦るなんていうのはもうざらって感じで
 <並列節デ> - ; 大きい切れ目-係り先なし
 3. 僕自身も一か月で正直な話百五十万ぐらい擦ったことあんですけど/並列節ケド/
 4. 勝ちはいせいで一日最高十五六万ぐらいですかね [文末候補] (S00M0071)
- (226) 1. そんなことで入院がどんどんどん長引いていったので<理由節ノデ>私の母の兄弟姉妹での
 交代での付き添いが始まって<テ節> - ; 大きい切れ目-係り先なし
 2. 私もその時にお腹が大きかったので<理由節ノデ>実家に殆ど帰ってるが多かったんですが
 /並列節ガ/
 3. 母もうちにいなく<連用節>付き添いで<並列節デ>一晩交代で付いていることが始まったので
 <理由節ノデ>何か私も自身も不安で<並列節デ> - ; 大きい切れ目-係り先なし
 4. 何かだんだんだんだんこれから楽しみなことがあるのに<ノ二節>何かどうしたんだろうなっ
 ていう<トイウ節>何となく空気が漂い出しました [文末] (S02F0113)

また、ある節全体が直前の節に対する倒置要素になっているような場合にも、「大きい切れ目」を認定する。

次の例の3行目「ちゃんと耳が出るようにして<テ節>」という部分は、直前の「帽子を作ってあげたことがあるんですね」に係る部分であり、逆に直後の「そしたら～」には係り先がない。そこで、<テ節>の直後に「大きい切れ目」を認定し、発話を分割する。

- (227) 1. で<接続詞>後うち猫がいるんですが/並列節ガ/
 2. 冬は寒いかなと<引用節>思って<テ節>猫は寒いのは駄目なので<理由節ノデ>一度帽子を作
 てあげたことがあるんですね [文末]
 3. ちゃんと耳が出るようにして<テ節> - ; 大きい切れ目-係り先なし
 4. そしたら<条件節タラ>思い切り嫌われて<テ節>頭を振って<テ節>取られてしまいました
 [文末]
 5. で<接続詞>それから暫く寄ってこなくなり<連用節>... (S04F0013)

さらに、弱境界以外の要素の直後において談話構造上の大きい切れ目となっている場合にも、「大きい切れ目」を認定する。

- (228) 1. 私が生まれた場所はですね神奈川県の川崎市でして/テ節/
 2. 父が製鉄所ですかに<間接疑問節-助詞>勤めていた関係で - ; 大きい切れ目-係り先なし
 3. 父父の出身は山形の酒田なんですけども/並列節ケドモ/
 4. で<接続詞>酒田から - ; 言いさし-言い直しあり
 5. 二十代前半ぐらいですか [文末候補]
 6. ちょうど戦争が終わって<テ節>まだそんなに程たっていない頃だと<引用節>思うんですけども
 /並列節ケドモ/
 7. 酒田から川崎まで出てきまして/テ節/
 8. それで<接続詞 L>縁があって<テ節>製鉄所へ職工として<テ節>勤めてまして/テ節/...
 (S03M0098)

さらに、上記の分類に該当しない場合でも、発話の進行上大きい切れ目となっていると判断された箇所には、その直後を「大きい切れ目」として認定する。当該箇所の直前を「-」で切断し、「; 大きい切れ目-その他」を義務的コメントとして記述する。

- (229) 1. 本当は今家族と同居してるんで<理由節ノデ>いずれはいずれは出ていかなくは<テハ節>い
 けないんですが/並列節ガ/
 2. 烏山というところがとてもアクセスもいいし/並列節シ/
 3. 町自体が凄くこう雰囲気もいいので<理由節ノデ> - ; 大きい切れ目-その他
 4. できれば<条件節レバ>その近辺に住みたいなっていう<トイウ節>風に今思ってます [文末]
 5. で<接続詞>ただ<接続詞>住みたいなどと<引用節>思っても<テモ節>色々... (S03F0224)

5.5.5 不適切な節境界ラベルに関するもの

5.5.5.1 間投助詞

「～ですね」という形に [文末候補] が付与されている場合、それがいわゆる間投助詞として機能しているために、実際には発話の完結点になっていないことがある。次の例の「ですね」は間投助詞として働いており、1行目の「到着した後」は2行目の「キャンセルしてしまいました」に係る要素であるが、「ですね」の直後に絶対境界のラベルが付与されているため、両者が分断されてしまっている。

- (230) 1. 那覇に到着した後ですね [文末候補]
2. 船酔いに苦しんだ多くの乗客は帰りのチケットをキャンセルしてしまいました [文末] (S02M0011)

そこで、間投助詞として働いている要素に対して [文末候補] のラベルが付与され、それによって不適切な分割結果が生じている場合、「間投助詞」を認定する。間投助詞の直後で不適切に分断された箇所を「+」で結合し、「; 間投助詞」を義務的コメントとして記述する。

- (231) 那覇に到着した後ですね [文末候補] + ; 間投助詞 船酔いに苦しんだ多くの乗客は帰りのチケットをキャンセルしてしまいました [文末] (S02M0011)
- (232) エヌがこのエヌベストのエヌが大きくなってしまえば<条件節レバ>殆ど同じですね [文末候補] + ; 間投助詞 認識率になってしまうんですが/並列節ガ/... (A01M0030)
- (233) 一日を仕事をハードに終えて<テ節>床に就く時娘の寝顔を見る時ですね [文末候補] + ; 間投助詞 何物にも代えない凄い幸せってか幸福感を感じます [文末] (S01M0706)

「ですね」以外の形に対しても、「間投助詞」が認定される場合がある。

- (234) 恐らくですねタイタニック号ですか [文末候補] + ; 間投助詞 それ並みにですね揺れたんだと<引用節>思います [文末] (S02M0011)
- (235) これは日本の道百選て言うんでしょうかね [文末候補] + ; 間投助詞 そんなのにも確か載ってると<引用節>思いますけれども/並列節ケレドモ/... (S03M0317)

5.5.5.2 と文末

CBAP-csjによって [と文末] が付与されている箇所以外にも、<引用節>のラベルが付与された節境界の直後が [と文末] に相当する発話の切れ目になっている場合がある。以下の例では、「治るんじゃないかと<引用節>」は、直前の「できていないと [と文末]」と同様、[と文末] に相当する発話の切れ目になっている。

- (236) 1. きちんと音を伝えるようにできてないと [と文末]
 2. で<接続詞>手術をすれば<条件節レバ>治るんじゃないかと<引用節>そういうことで
 <並列節デ>去年の十一月に手術をした訳です [文末] (S00M0053)

このような場合、当該箇所「-」をマークして発話を切断し、「;と文末」を義務的コメントとして残す。なお、人手で「と文末」を認定する際には、音声および文脈に基づいて判定する。

- (237) 1. きちんと音を伝えるようにできてないと [と文末]
 2. で<接続詞>手術をすれば<条件節レバ>治るんじゃないかと<引用節> - ;と文末
 3. そういうことで<並列節デ>去年の十一月に手術をした訳です [文末] (S00M0053)
- (238) 1. 閉じれば<条件節レバ>低くなるし/並列節シ/
 2. 開けば<条件節レバ>上がると<引用節> - ;と文末
 3. そういう関係が予測される訳であります [文末] (A01M0074)

5.5.5.3 例文など

学会講演（特に人文系）において、例文が読み上げられることがある。例文中に絶対境界・強境界が含まれる場合、例文の途中で発話が分割されるという不適切な結果が生じる。以下の例では、「荒らす」「たたる」「向く」が例文として読み上げられているが、それぞれの直後に [文末] が付与されているため、そこが分割位置となっており、不適切な分割結果となっている。

- (239) 1. 特殊なアクセントになっているものは殆どが起伏式の語で<並列節デ>平板式の語は荒らす
 [文末]
 2. たたる [文末]
 3. 向く [文末]
 4. 痩せるの四語だけでした [文末] (A02F0038)

そこで、例文として読み上げられている途中で不適切に分割された箇所には「例文など」を認定し、「+」をマークして発話を結合する。また、「;例文など」を義務的コメントとして記述する。

- (240) 特殊なアクセントになっているものは殆どが起伏式の語で<並列節デ>平板式の語は荒らす [文末]
 + ;例文など たたる [文末] + ;例文など 向く [文末] + ;例文など 痩せるの四語だけでした [文末]
 (A02F0038)
- (241) 五十パーセントのセンテンスが文末思考動詞と呼ばれている思うとか<トカ節>気がする [文末] + ;
 例文など 感じるという<トイウ節>言葉で終わっております [文末] (A06F0075)
- (242) ここでもし形式的にこの会話の中での他の相づちに習えば<条件節レバ>酷いね [文末候補] + ;例
 文など 程度の言い方も十分考えられますけれども/並列節ケレドモ/... (A06F0049)

5.5.5.4 格助詞相当表現

弱境界の節境界ラベル<テ節>は、直後に接続詞が後接した場合、メタ規則の適用により、強境界の/テ節/に書き換えられる。しかしながら、「に対して」「において」「によって」のように複合的な格助詞に相当すると思われる表現の直後は、接続詞が後続したとしても、やはり弱境界として認定されるべきである。このように、いわばメタ規則の過適用によって、不適切な分割結果が生じることがある。例えば以下の例では、「対して」の直後が強境界となり、発話分割位置となっている。本来は、1行目の「正しい確認に対して」と2行目の「間違っただ確認に対して」は並列された構造であるため、この分割結果は適切ではない。

- (243) 1. で<接続詞>正しい確認に対して/テ節/
 2. あるいは<接続詞>間違っただ確認に対して<テ節>それぞれどんな特徴があったかと<引用節>言う
 くと<条件節ト>... (A01M0021)

そこで、格助詞相当表現の直後が強境界となり、不適切な分割結果が生じている場合に限り、「格助詞相当表現」を認定し、当該箇所「+」をマークして発話を結合する。また、「; 格助詞相当表現」を義務的コメントとして記述する。

- (244) で<接続詞>正しい確認に対して/テ節/ + ; 格助詞相当表現 あるいは<接続詞>間違っただ確認に対して<テ節>それぞれどんな特徴があったかと<引用節>言うくと<条件節ト>... (A01M0021)
- (245) こういったものを元にいたしまして/テ節/ + ; 格助詞相当表現 本発表で考える文段の接続のあり方としては大きく三種類を考えました [文末] (A02F0082)

5.5.5.5 強→弱

<テ節>および<条件節ト>は、述語句がデスマス体・ゴザイマス体の場合、または直後に接続詞が後接した場合、メタ規則の適用により、強境界に書き換えられる。しかしながら、発話の構造的な切れ目としては認めにくく、弱境界へ戻ることがふさわしいと判断される場合がある。例えば以下の例の「御覧いただきまして」には強境界の/テ節/が付与されているが、このテ節は後方の「分かるように」にのみ係る要素であり、その直後に構造的な切れ目があるわけではない。

- (246) 1. 御覧いただきまして/テ節/
 2. 分かるように幾つかのパターンが示されると<引用節>思います [文末] (A01F0067)

このようなメタ規則の過適用によって不適切な分割結果が生じている場合に限り、節境界ラベルのレベルを強境界から弱境界に戻す操作を行う。当該箇所「+」をマークして発話を結合し、「; 強→弱」を義務的コメントとして記述する。

- (247) 御覽いただきまして/テ節/ + ; 強→弱 分かるように幾つかのパターンが示されると<引用節>思います [文末] (A01F0067)
- (248) そうしますと/条件節ト/ + ; 強→弱 今度は文字が多過ぎて<テ節>読み切れないという<トイウ節>ことがあります [文末] (A11M0469)
- (249) 青山と言いますと/条件節ト/ + ; 強→弱 御存じの方はやっぱりお洒落な町都会的な都内でもお洒落な町ということで知られてるところかもしれないんですけども/並列節ケレドモ/... (S03F0108)
- (250) もうきちんとノートに書くようになってから<テカラ節>もう三年くらいたってるんで<理由節ノデ>随分当初に比べますと/条件節ト/ + ; 強→弱 乱暴な書き方になってきてしまってます [文末] (S00F0031)

5.5.5.6 文末候補

節境界ラベルは付与されていないが、[文末候補]に相当する発話の明らかな切れ目となっている箇所が見つかる場合がある。次の例は、音声を聞く限り、「入れてくださいみたいな」の直後に発話の切れ目が存在する。しかしながら、[文末候補]のラベルを付与する定義には該当しない表現であるため、節境界ラベルは付与されていない。

- (251) 1. 畑がまだたくさんあるんで<理由節ノデ>その日取れた野菜とかをこう山積みにして/テ節/
2. で<接続詞>お金は缶の中に入れてください<文末候補>みたいなだから<接続詞 L>入れてる人もいれば<条件節レバ>入れてない人もいるから<理由節カラ>まだそういうところがこうちょっとほのぼのして<テ節>... (S03F0062)

また次の例では、音声を聞く限り「なぜか」の直後に明らかな切れ目が存在するが、やはり節境界ラベルは付与されていない。

- (252) 1. ところが<接続詞 L>これまでのところ具体的自主的な検討は進んでおりません [文末]
2. なぜか主として<テ節>常任理事国が特権を手放したり/タリ節/
3. あるいは<接続詞>他の国に自分達と同様の特権を持たせるっていう<トイウ節>ことをこれ嫌ってる訳ですね [文末] (S06M0373)

そこで、[文末候補]に相当する発話の明らかな切れ目となっている箇所の直後を「-」で切断し、「 ; 文末候補」を義務的コメントとして記述する。

- (253) 1. 畑がまだたくさんあるんで<理由節ノデ>その日取れた野菜とかをこう山積みにして/テ節/
 2. で<接続詞>お金は缶の中に入れてください<文末候補>みたいな - ; 文末候補
 3. だから<接続詞 L>入れてる人もいれば<条件節レバ>入れてない人もいるから<理由節カラ>まだそういうところがこうちょっとほのぼのして<テ節>... (S03F0062)
- (254) 1. ところが<接続詞 L>これまでのところ具体的自主的な検討は進んでおりません [文末]
 2. なぜか - ; 文末候補
 3. 主として<テ節>常任理事国が { 特権を手放したり/タリ節/ : ; 引用節構造 あるいは<接続詞>他の国に自分達と同様の特権を持たせる } っていう<トイウ節>ことをこれ嫌ってる訳ですね [文末] (S06M0373)
- (255) 1. 基本的に狭いところなんで<理由節ノデ>ちょっと変わったものを買おうと<引用節>思うと<条件節ト>まずないと<引用節> - ; と文末
 2. 巣鴨とか池袋に出ないと<条件節ト>いけないっていう<トイウ節> - ; 文末候補
 3. そんなようなのでもやっぱり営業できてるところがこの町の凄いところかもしれませんね [文末候補] (S03M0046)
- (256) 1. これから私の意見に入ります [文末]
 2. 常任理事国入りと現行憲法の平和主義これをどう整合させるか<間接疑問節> - ; 文末候補
 3. 二十世紀初頭の日本にとって<テ節>極めて悩ましい問題です [文末]
 4. で<接続詞>勿論私も軽々に結論を出す訳には行きません [文末] (S06M0373)

5.5.5.7 非文末

CBAP-csj によって認定された絶対境界が、これまでに述べてきたこと以外の理由によって、不適切な位置での分割を生じさせていることがある。例えば以下の例では、「～にしる」の直後に絶対境界 [文末] が付与されているが、ここで発話が分割されるのは適切ではない。

- (257) 1. そのような訳で<並列節デ>殆どの行政にしる [文末]
 2. 色んな催し物にしる [文末]
 3. 八潮市っていうところは草加市の陰に甘んじていると言うか<文末候補>... (S03M0996)

そこで、CBAP-csj によって付与された絶対境界のラベルが不適切な分割結果が生じさせている場合、「非文末」を認定する。当該箇所「+」をマークして発話を結合し、「 ; 非文末」を義務的コメントとして記述する。

- (258) そのような訳で<並列節デ>殆どの行政にしろ [文末] + ; 非文末 色んな催し物にしろ [文末] + ; 非文末 八潮市ってところは草加市の陰に甘んじていると言うか<文末候補>... (S03M0996)
- (259) うちの犬がそういう音が嫌いで<並列節デ>大騒ぎして<テ節>植木は倒すわ [文末候補] + ; 非文末 網戸は破くわで<文末候補>大変です [文末] (S03F0119)
- (260) スピッツとかチンとかコリーって見掛けないんで<理由節ノデ>どこ行ってしまったんだろうと [と文末] + ; 非文末 ええ<感動詞>いうこと感じますね [文末候補] (S02M0191)
- (261) 活用語尾と語幹また<接続詞>名詞動詞形容詞などでは変化に遅い [文末] + ; 非文末 速いがあるんだろうかとか<ト力節>地の文と会話の文また<接続詞>位相差ってというようなもので何か表記にかかわりはないんだろうかとか<ト力節>... (A02F0116)

5.5.5.8 タグミス-その他

その他, どの項目にも当てはまらないが, CBAP-csj による発話分割の結果が不適切であると判断された部分には, 「タグミス-その他」を認定し, 該当箇所の修正 (結合, 切断) を行う。結合を行った例としては, 以下のようなものがある。

- (262) 1. 行き当たりばったりで<並列節デ>作るっていう<トイウ節>ことだと<条件節ト>実際の開発が難しいし/並列節シ/ + ; タグミス-その他 メンテナンスも難しいんで<理由節ノデ> - ; 直後
がまとめ表現
2. 我々は方法論として<テ節>こういう方法論を今回提案します [文末] (A04M0047)
- (263) ただ<接続詞>図の軸が違いますので<理由節ノデ>先程の (M あ) に比べますと/条件節ト/ + ; 強
→弱 非常に狭い範囲で/並列節デ/ + ; タグミス-その他 しかし<接続詞>コンシステントに差が出て
るといふ<トイウ節>ことでもあります [文末] (A01M0074)
- (264) そうすると<条件節ト>それを最も徹底させ/連用節/ + ; タグミス-その他 かつ<接続詞>恐らくは
意識的に行なっていたのが定家であるといふ<トイウ節>解釈が可能かと<引用節>思います [文末]
(A02M0076)

また, 切断を行った例としては, 以下のようなものがある。特に会話が引用されている場合には, 「 ; 体言止」や「 ; タグミス-その他」によって修正されることが多い。

- (265) 1. ここでは虫よけ対策のスプレーについての<連体節テノ>話をしています [文末]
 2. この前ねあたしのいつも行き付けのねお花ねお花屋さんでね - ; タグミス-その他
 3. おお<感動詞> - ; 体言止
 4. レモンの香り - ; タグミス-その他
 5. ああ<感動詞> - ; 体言止
 6. レモンの香りのするスプレー売ってて<テ節> - ; タグミス-その他
 7. ああ<感動詞> - ; 体言止
 8. それは虫よけになりますって<引用節>言ったから<理由節カラ>何か - ; タグミス-その他
 9. うん<感動詞> - ; 体言止
 10. おんなじようなものなんじゃない [文末] + ; 倒置-つなぎ切り <<柑橘類のにおいって>> - ; 倒置-つなぎ切り
 11. そうかな - ; 文末候補
 12. うん - ; 体言止
 13. うんうん<感動詞>らしいよ [文末候補]
 14. 柑橘類のにおいが効くのかな/文末候補/
 15. でも<接続詞 L>レモンはそうだった [文末]
 16. そううん<感動詞>それは初耳だね [文末候補]
 17. うん<感動詞> - ; 体言止
 18. 例の三ではビーの相づちに注目してください [文末]
 19. この会話の中ではビーはやや男性的なあるいは<接続詞>ぞんざいな... (A06F0049)

5.6 節単位情報の管理と XML 文書への格納

最後に、本章で示してきた節単位に関する情報が、最終的にどのように管理され、また、XML 文書の中でどのように格納されているかについて、簡単に示しておく。

人手修正を経て認定された全ての節単位に対しては、当該の講演内における ID が一意に付与される。これを「節単位 ID」と呼ぶ。節単位 ID は、講演の冒頭に現れる節単位に対して「0」を付し、以降、講演の終わりまで、すべての節単位に対して連番で振られる通し番号である。講演 ID と節単位 ID によって、CSJ に含まれる全ての節単位は、一意に定められることになる。

次に、節単位に関する情報が XML 文書の中でどのように格納されているかについて示す。節単位に関する情報は、すべて短単位 (SUW 要素) の属性として格納されている^{*14}。SUW 要素に格納されている節単位情報の属性名について、表 5.5 に示す。

以下、節単位に関する情報について解説する。節単位 ID (ClauseUnitID) は、その節単位の開始点となる SUW 要素の属性として付与される。節境界ラベル (ClauseBoundaryLabel) は、その節境界の末尾となる SUW 要素の属性として付与される。人手修正操作記号 (CU_OperationSign, CU_PreBracket, CU_PostBracket) は、切断、結合、範囲確定などの人手修正操作を行った当該の位置にある SUW 要素の属性として付与される。

*14 節単位 ID については、稀に、SUW 要素でなく Noise 要素の属性として格納されている場合がある。

表 5.5 XML 文書内における節単位情報の格納

属性名	説明	値
ClauseUnitID	節単位 ID	全ての節単位に対して与えられる番号
ClauseBoundaryLabel	節境界ラベル	CBAP-csj により付与される節境界ラベル
CU_OperationSign	人手修正操作記号 (切断・結合)	切断記号 (+), 結合記号 (: -) の 3 種類
CU_PreBracket	人手修正操作記号 (範囲-開始点)	範囲記号の開始点 (({ <<) の 3 種類
CU_PostBracket	人手修正操作記号 (範囲-終了点)	範囲記号の終了点 () } >>) の 3 種類
CU_ObligateComment	義務的コメント	人手修正操作を行った理由

義務的コメント (CU_ObligateComment) もまた、実際に修正を行った箇所に当たる SUW 要素の属性として付与される。

その他、XML 文書に付与された各要素の内容については、8.3 節の表 8.1 を参照されたい。

5.7 節単位が持つ意味と今後の展望

本章では、話し言葉における分析用の基本単位として、「節単位」という独自の単位を設計するに至った経緯と、その認定手順の詳細について述べてきた。CSJ における構文レベル・談話レベルの研究用情報—係り受け構造情報、重要文情報、談話境界情報—を付与するために用いる共通単位を確定することを目的として、節境界を主たる基準とした統語的な単位を取り出すことを試みた。実際には、人手修正のための基準作成に比較的長い時間が費やされたものの、最終的には話し言葉の分析用単位としてその仕様を確定し、節単位の認定作業を終えることができた。また、係り受け構造情報、要約・重要文情報、談話境界情報については、本章で示した基準によって認定された節単位をもとに、情報付与作業が実際に行われた。これらの詳細については、『日本語話し言葉コーパス』付属マニュアルのうち、以下のものを参照されたい。

- 『日本語話し言葉コーパス』における係り受け構造付与
- 『日本語話し言葉コーパス』における自由要約・重要文抽出データについて
- 『日本語話し言葉コーパス』の談話境界情報について

話し手と聞き手相互のインタラクションから談話が構築されていく対話とは異なり、一人の話し手が話し続ける独話という発話様式は、最初に想定された発話プランに従って実時間上に展開されていく一回的な行動 (発話生成) の結果である。その結果だけを取り出して見ると、独話とは、連綿と続くいわば「一本の紐」のような状態であると言える。この発話生成の結果からある種の言語的な単位を取り出すことを考えようとする、従来言語の中心的な単位であった「文」という概念は、途端に通用しなくなることが分かる。これはそもそも、従来の言語研究—特に文法研究—が、実際の話し言葉に現れる種々の現象をノイズとして切り捨て、特に母語話者による内省を主たる立脚点として、言語の客観的な分析という作業に対して洗練を重ねてきた結果であるとも言える。

これに対して、話し言葉の大規模なコーパスが完成した現在、従来とは全く異なるアプローチによって、言語におけるある種の単位を捉え直すことができるかもしれない。一例を挙げると、本章で示した節単位という

独自の単位は、独話における話し手の心理的な単位を反映したものである可能性が挙げられる。動的な発話生成の過程において、話し手が発話のために管理できる概念や情報の量はさほど多いわけではない。話し手は、多くの情報と複雑な構造を持つ長大な「文」をあらかじめ想定して発話を開始するわけではなく、むしろ、比較的短く、かつ統語的・意味的なまとまりを持つような単位を次々と産出し、それらに関連を持たせながら動的につながり合わせていくことによって、独話を構成していくと考えられる。節単位は、そのような発話生成時における基本的な処理単位、あるいは話し手の心理的なチャンクを反映する単位として捉えられる可能性がある。

本章で示した節単位という概念に関する一般言語学的な妥当性は、さまざまな側面から話し言葉を分析し、記述が進められることによって、改めて問われるものとする。そのような視点は、同時に、話し言葉の言語学的研究—特に文法研究—に大きく寄与するものとする。最後に、大石(1971:34)による次のような言葉を引用しておきたい。

話しことばの浮動性、不安定性をさしひいてもなお、書きことばの上に現れない特殊な言語事実が、かずかず話しことばの上に発見される。たとえば文法に関しても、書きことばの上に見られない構造や傾向が、話しことばの上に、いろいろ見いだされる。これをとらえなければ、やはり、日本語の文法を全体的に明らかにすることができないのではなからうか。話しことば独自の文法的事実から抽出されるものを組み込んで、日本語の文法体系は完成されるのではなからうか。