

文節の仕様について

Version 1.0

西川 賢哉[†] 小椋 秀樹[†] 相馬さつき[†] 小磯 花絵[†]
間淵 洋子^{†,‡} 土屋 菜穂子^{††} 斉藤 美紀^{†,‡‡}

[†] 国立国語研究所 [‡] 東京都立大学大学院 ^{††} 青山学院大学大学院 ^{‡‡} 東京大学大学院

目次

0. 初めに
1. 本則 [A] 助詞・助動詞連続
2. 本則 [B] 助詞・助動詞を伴わない自立語
3. 本則 [C] 体言連続
4. 細則 [A] 例外規則
5. 細則 [B] 話し言葉に特有の現象に対する処理
6. 細則 [C] タグの処理
7. 文節分割例

0 初めに

- 本マニュアルでは、転記テキストにおける文節の認定規則の概略を示す。
- 本マニュアルで定義する文節は、転記テキストにおける二つの表記法 — 基本形と発音形 — の対応を取り易くすることを目的として導入されたものである（「基本形」「発音形」についての詳細は、transcription.pdf 参照）。転記テキストにおいては、文節単位境界は改行によって示される。以下のサンプルにおいて、一行に記されたものが文節に相当する。

0001 00000.113-00006.570 L:		
それでは	&	ソレデワ
最初に	&	サイショニ
融合研究の	&	ユーゴーケンキューノ
話し言葉コーパスの	&	ハナシコトバコーパスノ
設計と	&	セツケート
実現への	&	ジツゲンエノ
方策という	&	ホーサクトユー
ことで	&	コトデ
簡単に	&	カンタンニ
話を	&	ハナシオ
させていただきます	&	サセテイタダキマス

- 本マニュアルで定義する文節は、係り受け解析における基本単位としても利用される。
- 本マニュアルで定義する文節は、形態論情報（短単位・長単位）の上位に位置する単位である。つまり、短単位・長単位が複数の文節に跨がることはない。ただし、短単位・長単位の定義に利用される「文節」という概念と本単位は必ずしも一致しない¹。
- 転記テキストにおいては、文節の他に、転記基本単位によっても発話が分割される。多くの場合転記基本単位の境界と文節の境界は一致するが、これらが一致しないこともある。その為、以下の例に示すように、一つの文節であっても複数の行に跨がることもある。

	そういう	&	ソーユー
	文化活動として	&	ブンカカツドートシテ
	0367 01314.557-01315.379 L:		
	古代ギリシャ	&	コダイギリシャ
一文節	0368 01315.650-01317.468 L:		
	劇を	&	ゲキオ
	(F お)	&	(F オ)
	復活させようと	&	フッカツサセヨート
一文節	0369 01317.919-01321.891 L:		
	というような	&	(W ユ; ユウ)(W ヨ; ヨー)ナ
	形で	&	カタチデ
	オペラというのが	&	オペラトユーノガ
	(F あー)	&	(F アー)

- 本マニュアルでは、転記基本単位や発音形に特に言及する必要のない場合、基本形のみを改行を用いずに示す。その際、以下に挙げる記号類を用いる。

凡例

	文節の切れ目を示す。
	同上。特に当該規則で着目している箇所。
-	切らずに繋げることを示す。
=	同上。特に当該規則で着目している箇所。
下線	特に着目している箇所 (分かりにくい場合)。

¹短単位・長単位の定義に利用される「文節」は、あくまで短単位・長単位を認定するために導入された概念であり、形態論情報の一部として具体的に文節単位に分割している訳ではない。pos.pdf における「文節 (2.1.1)」参照。

1 本則 [A] 助詞・助動詞連続

助詞・助動詞連続の後で切る．

【例】 | 私共では || (F あの一) | (A エヌエイチケー; NHK) の || ニュースを || 音声を || データベース化すると-いう || 仕事を || やっております |

1.1 助詞相当句・助動詞相当句中の助詞

助詞相当句・助動詞相当句については，構成要素の間に副助詞などが挿入されることがあるが，その場合も全体で一つの複合辞と見なす．

【例】 | 凄く | 食べて=ばっか=いましたね |

2 本則 [B] 助詞・助動詞を伴わない自立語

助詞・助動詞を伴わない自立語については，以下の各項に該当する箇所では切る．

2.1 主語・主題

主語・主題の後で切る．

【例】 | 山も | あり | 緑 || 溢れる | とても | いい | 遊園地です |

2.2 連用修飾成分

連用修飾成分²の後で切る．

【例】 | 私の | お腹の | 中の | 子供が | あんまり || 育たなく || なりまして |
| そこで | ビール || 飲みながら | お弁当 || 食べながら | 待っていると |
| 先日 || 整体に | 行ったんですね | | 地中 || 深く | 根は | 根付くと |

2.2.1 「消滅する」「紛失する」「死去する」の意の「なくなる」「亡くなる」は切らない．

【例】 | 雑木林は | なく=なってしまうんだけど | | 祖父が | 亡く=なった | 瞬間に |

2.3 連体修飾成分

連体修飾成分の後で切る．

【例】 | 基本的には | あまり | 高い || 建物が | ない || ことを | | この || 国家と-いう || 壁は |

2.4 中止法・終止法・命令法

用言の中止法・終止法・命令法の後で切る．

【例】 | 川も | あり || 森も | あり || とても | 大自然の | 中に |
| 裏切られたとしか | 言いようが | ありません || あんなに |
| で | 何か | (F あの一) | 頑張り || 池田高校ナイン |

2.5 接続詞

接続詞の前後で切る．

【例】 | 夏は | プール || そして || 夜空を | 彩る | 花火大会 |

2.6 感動詞

感動詞の後で切る．

【例】 | はい || そうです | 月曜日ですっていう |
| おい || おい || こんなに | 取ってしまってもいいのかというぐらい |

感動詞のうち，タグ (F) で囲まれた要素については，規則 6.7 に従って処理する．

2.7 独立格

体言の独立格の後で切る．

【例】 | 犬の | 方から | (F あ一) | お父さん || 起きてよというような | ことで |

² 以下のいずれかに該当するものを連用修飾成分と見なす：

- (1) 体言又はそれに相当するものに格助詞が付いたもので格機能を持つもの。ただし格助詞が省略されることがある。
- (2) 形容詞・形容動詞連用形又は副詞(名詞の副詞的用法も含む)で格機能を持たず，述語の実質概念を修飾限定するもの。

3 本則 [C] 体言連続

以上の規則に該当しないものについて、次の各項を適用する。

3.1 一部が連体修飾成分を受けている体言連続

体言連続の一部が連体修飾成分を受けている場合、その部分の後に切る。

【例】 | 小泉内閣の | 支持率 || 上昇 | | 声道の | 狭め || 形成運動とか |

ただし、次に挙げる語が付いた場合には切り離さない。

当たり 以下 以後 以降 以上 以前 以内 以来 外 間 後 ごと 視 自身 自体
上 前 前 達 的 同士

【例】 | インディーズの | 教祖=的な | 存在に | なってしまった |
| 発展の | 方は | 文章の | 途中=以降で |

3.2 同格・言い換え

同格・言い換えの関係にある体言連続は、互いに切り離す。

【例】 | 計量国語学会が | 設立され | (F えー) | 機関誌 || 計量国語学が | 発刊され |
| この | 差を | 取った | 値が | 閾値 || (A シータ;) よりも | 大きい | 場合に |

以下のようなものにも当規則を適用する。

【例】 | で | (F えっとー) | 被験者 || (A ビー; B) | こちらは | (M うん) という | 言葉を |
| 文 || (A ビー; B) と | (F え) | 文 || (A シー; C) に関して |

3.2.1 同格・言い換えの関係にある体言連続を受ける、もしくはそれに係る体言・接辞がある場合は切らない。

【例】 | 機関誌=計量国語学-発刊の | 経緯 | | 九月十五日=敬老の日-生まれの |

3.3 並立

並立された語は、互いに切り離す。

【例】 | 検出結果の | 比較 || 検討を | 行なっています |
| 先行研究に | (F え) | 定延 || 田窪 や | 田窪 || 金水 などが | あります |

人名が英語接続詞「アンド」によって並立された場合にも、当規則を適用する。

【例】 | トンプソン || アンド || ミュラックでは | cf. | (A ビー-アンド-ビー; B & B) |

3.3.1 並立の関係にある語連続のうち、並立されたもの全体を受ける、もしくはそれに係る体言・接辞・形式的な意味の「する」「できる」「なさる」「いたす」がある場合は切らない。

【例】 | 上位=下位-関係を | cf. | 上位 | 下位の | 関係を |
| 益岡・=田窪-氏の | cf. | 益岡 | 田窪には |
| 益岡・=田窪-一九九二には |
| 比較=検討-してみましたので | cf. | 比較 | 検討を | してみましたので |

3.4 二文節以上を受ける / に係る形式

二文節以上からなる形式全体を受ける，もしくはそれに係る接辞および体言的な形式は，その前後で切る．

【例】 | 基本周波数パターンが | (F えー) | への | 字 || 型と | なる | 区間を | ³
| 森首相の | 神の | 国 || 発言と | その | 波紋 |
| 全然 | (F あの) | (D み) | 寝耳に | 水 || 状態で |

それが結び付ける要素の両方またはいずれかが二文節以上である「対」も同様に処理する．

【例】 | 夏の | 北海道 || 対 || 冬の | 九州 | 沖縄 |

3.5 数量表現を含む体言連続

数量表現を含む体言連続が以下のいずれかに該当する場合，数量表現の前または後ろで切る．

- 前の数量表現が，数量で表わされるものの範囲を規定している場合

【例】 | パチスロの | 場合だったら | 一箱 || 三万ぐらいなんですけど |

- 数量表現が後の要素の表わす位置・時間を限定している場合

【例】 | そこから | 一歩 || 先に | 踏み込んだ | 政策議論が |

- 数量表現が前の要素の数量を表わしている場合

【例】 | 平均値 || (A 三 . ゼロ六 ; 3 . 0 6) というような | 値に | なっております |
| 短単位は | 原則として | 最小単位 || 二つが | 一次結合したのから | なります |

- 前の要素が計量の単位，もしくは計量の手段・方法を表わしている場合

【例】 | ベーキングパウダー | 重曹 | それぞれ | 小匙 || 一杯 |

- 数量表現が前で列挙された要素の個数を表わしている場合

【例】 | 日米韓 || 三国の | 対応 |

- 前の要素が，数量を表わす要素に連用的に係る場合

【例】 | で | 知床には | 熊がですね | 推定 || 三百頭 | いると | 言われています |
| 不動産の | その | 物件自体で | 月々 | 平均 || 二十五万ぐらい | 掛かるんですよ |
| 当時 || 四十歳の | その | 友人は |

以上に該当しないものは，切り離さない．

【例】 | 昭和=十三年=八月=八日の | 荒木文部大臣の | 発言や |
九十八年=秋には	関西にも	進出し	
朝=八時半に	家を	出て	
これは	築島=(A 千九百六十六 ; 1 9 6 6) 年による	点検	
野村=一九九四で	提唱された	概念ですが	
まずは	例文=三十三を	御覧ください	
予稿集=八十七ページ	(F あの)	訂正を	お願い申し上げます

3.5.1 規則 3.5 により切り離されるもののうち，数量表現を含む体言連続全体を受ける，もしくはそれに係る体言がある場合は切り離さない．

【例】 | 当時 | オドル=三百六十円-換算で | 千八百億円の | | 日独伊=三国=軍事同盟 |

³ この規則の適用により，接辞からなる文節が発生することがある．

4 細則 [A] 例外規則

規則 1 から 3.5 にかかわらず，以下に挙げるものは各規則に定めたように処理する．

4.1 内部で切らないもの

以下に挙げるものは，その内部が他の規則で切ることになっていても切らない．

• 以下に挙げる固有名

人名	【例】	^{みなもとの} 源=頼朝 舞の=海 明石の=尼君
国名	【例】	クレートブリテン=及び=北部アイルランド 連合王国
行政区画名	【例】	自由が=丘 お茶の=水 西が=丘
駅名	【例】	青山=一丁目の 駅を 通ってきたんですけど
地名形	【例】	場所は 丹沢の 塔の=岳が 使われます
場所名	【例】	丸の=内 虎の=門 交差点
建物名	【例】	五重の=塔 万里の=長城
組織名	【例】	東京の=教育 (A 二十一; 2 1) 研究開発委員会
グループ名	【例】	ピンキーと=キラー ジュディー・=アンド・=マリー
競争馬の登録馬名	【例】	マチ=カネ=ワラウ=カド オジサン=オジサン
歴史的出来事の名称	【例】	関が原の=戦い 本能寺の=変
祝日・記念日	【例】	こどもの=日 防災の=日 耳の=日
商品名	【例】	きのこの=山 電車で=ゴー
題名 (副題は除く)		
書物	【例】	走れ=メロス 複合動詞の=構造と=意味用法
テレビ・ラジオの番組名	【例】	母を=訪ねて=三千里 夜の=ヒットスタジオ
映画・舞台・演劇の題名	【例】	火の=鳥 ベルサイユの=ばら
楽曲名	【例】	白鳥の=湖 恋は=レインボー
美術作品名	【例】	レースを=編む=女 ミロの=ビーナス

• 動植物名

【例】 | タツノ=オトシゴ | | ユキノ=シタ | | ワレモ=コウ |

• 固定化した表現⁴

【例】 | 茶の=間 | | 万が=一 | | 負けず=嫌い |

4.2 数式・公式等の読み上げ

4.2.1 四則演算子(「足す」「引く」「掛ける」「割る」等)，比較演算子(「イコール」等)の前後で切る．

【例】 | (A ディー; D) || イコール || ゼロの | 時 | | 縦 || 掛ける || 横が |
| (A エックス; X) || プラス || (A ワイ; Y) という | 領域は |
cf. | どの | 時間にも | プラス=一を | 付けました |

4.2.2 分数「 x 分の y 」・累乗「 x の y 乗」は全体を一続きのものとする．

【例】 | 後続単語種類数分の=先行単語頻度 (D n な) の | 関数に | なっています |
| それは | 五分の=一から | 六分の=一に | なるらしいんですけど |

ただし， x, y に当たる要素の両方またはいずれかが二文節以上である場合「の」の後で切る．

【例】 | 標準化周波数分の || 帯域幅 || (A パイ;) |

4.2.3 切る位置が不明な場合は，全体で一文節とする

【例】 | 初期値 | (0 シグマハットマイナスイェルバーマイナスイェル) には |

⁴ 用語リストに登録されているもののみ．

4.3 「何とか」を含む自立語的形式

一つの自立語となるべき要素の一部を「何とか」で置き換えたものは、それを含む全体を一続きのものとする。

【例】 | 電話受け付けとか | お客様 (D す) 何とか=センター とか |

4.4 助動詞連用形を含む体言的形式

体言の一部を構成する「助動詞連用形 + { 名詞/接辞 }」については、助動詞連用形の後で切らない。

【例】 | いじめられ=っ子 | | 憎まれ=役 | | 言わせ=方 |

4.5 述語

次に挙げる形式が述語 (の一部) を構成する時、その内部で切らない。

- 「形容詞 + 助詞 + { ある/ない/ございます }」

【例】 | あんまり | 大きくは=ないんですね |
| そう | 大きくは=ありません |

- 「形容詞ウ音便 + ございます」

【例】 | 大変 | 難しゅう=ございます |
| おはよう=ございます |

- 「助動詞 + (係助詞/副助詞) + { 動詞/形容詞/接尾辞 }」

【例】 | 分散和音型を | ずっと | 聞かせ=続けて |
| 逆に | 敬遠され=がちかと | 思いますけれども |
| あってほしく=は=ないんですけれども |

ただし、「連用形 + { 係助詞/副助詞 } + する」については、助詞の後で切る。

【例】 | 殴られは | する | cf. | 泳ぎは | する |

- 「{ お/御 } + { 動詞連用形/名詞 } + { する/できる/くださる/いただく/なさる/いたす/願う/申し上げる/あそばす }」

【例】 | いちいちの | 用例について | 御説明=申し上げられませんが |

- 「{ お/御 } + { 動詞連用形/名詞 } + になる」

【例】 | 数字を | 見ると | お分かりに=なりますように |

- 伝聞を表わす「...だ+そう+{ だ/です }」

【例】 | あの | 人は | 九十幾つとかだ=そうです |

4.6 連体修飾成分と切り離さない語

次に挙げる語が直前に連体修飾成分を伴って出現した場合、その語の前で切らない。

ところ (とこ), もの (もん), はず, 訳^{わけ}, よう, みたい, そう

【例】 | 住んでいる | 私達としては | 非常に | (F えー) | 便利な=ところでもあります |
ただ	時間は	もう	無制限に	ある=訳ですから
今	言った=ような=ものを	作って		
結局	スキーに	来た=みたいで		
ーから	勉強し直すべきだと	王位継承者を	批判した=そうです	

4.7 同一の要素，類似の要素の連続

同一の要素および類似の要素の連続は以下のように処理する。

言い直しにより生じた同一要素，類似要素の連続は，規則 5.1 に従って処理する。

4.7.1 同じ代名詞・副詞・接続詞・感動詞の連続のうち，次に挙げるものは切り離さない。

| いつ=いつ | | ごく=ごく | | さて=さて | | ただ=ただ | | なお=なお | | 何=々 | | まず=まず |
| バイ=バイ | | また=また | | まだ=まだ | | もし=もし | | よく=よく |

【例】 | ごく=ごく | (F えー) | 普通の | 女の子が |
| 期待も | 含めて | まず=まずの | スタートと | なりました |

4.7.2 指示代名詞の連続・指示副詞の連続は，切り離さず全体を一続きのものとする。

【例】 | どこ=どこ | | あち=こち | | あれ=これ | | どう=こう |

4.7.3 オノマトペの連続は，全体を一続きのものとする。

【例】 | 腰のところ | が | くる=くる=くるって | こう | 巻かれて |
みんなで	わいわい=わいわい	(F えー)	やりながら
みんなで	わいわい=がやがや	言いながら	
どこか	ときどき=わくわく-する	ような	気分を

4.7.4 同じ名詞 (数量表現含む) の連続は互いに切り離す。

【例】 | その | 合間 || 合間に | 夫の | 単身赴任という | 期間も |
| 各異音 | 一つ || 一つの | 異音ごとに |

4.7.5 同じ動詞連用形の連続は互いに切り離す。

【例】 | 繰り返し || 繰り返し | | 組み替え || 組み替え |

4.8 体言+「ない」

体言に「ない」が直接続く場合，体言と「ない」を切り離さない。

【例】 | あるいは | 素っ気=ない | 口調の | | うちに | 来て | 間違い=なく | その | 犬だと |

4.8.1 前の体言が連体修飾を受けている場合は，用言部分を切り離す。

【例】 | もう | 彼との | 関係 || ないよ | cf. | もう | 関係=ないよ |

4.9 体言+「する」「できる」「なさる」「いたす」

体言に形式的な意味の「する」「できる」「なさる」「いたす」が直接続く場合，体言と用言を切り離さない。

【例】 | 外来音について | 若干 | 許容=してきて | おります |
| 有益な | 出会いを | 演出=できる | ような | ことが |

国語辞典でさ変動詞語幹としての用法が記述されていないものについても，形式的な意味の「する」「できる」「なさる」「いたす」が直接続く場合は，体言と動詞とを切り離さない。

【例】 | 青空に | 桜の | 花が | 満開=して | てる | 様子は | | 更に | 三鼻母音=した | 結果 |

以下のようなものにも当規則を適用する。

【例】 | もう | みんな=して | 集団悔し泣き状態に | なって | しまって |
| それから | 二三日=すると | | 十万円=する |

4.9.1 前の体言が連体修飾を受けている場合は、用言部分を切り離す。

【例】 | 日本語の | 勉強 || するんです | cf. | 日本語を | 勉強-するんです |

4.9.2 指示代名詞「あれ」「これ」「それ」「どれ」およびこれらの複合(規則 4.7.2)に「する」「できる」「なさる」「いたす」が直接続く場合、代名詞と用言を切り離す。

【例】 | 寄宿舍生活ですね | それ || してた | 為に | | あれ-これ || してみたのですが |

4.10 副詞+「する」「できる」「なさる」「いたす」

副詞のうち、形式的な意味の「する」「できる」「なさる」「いたす」に係るものの後では切らない。

【例】 | 目を | きらきら=させながら | 熱い | 視線で |

以下のようなものにも当規則を適用する。

【例】 | それから | 暫く=すると |
| こう=したら | いいんじゃない | ああ=したら | いいんじゃないという |
| 慇懃無礼というのが | どう=して | (F ま) | 不愉快なのかと |

副詞と形式的な意味の「する」「できる」「なさる」「いたす」との間に助詞が挿入された場合は切り離す。

【例】 | エビが | ふっくらと || するそうです | cf. | エビが | ふっくら=するそうです |

4.11 「もしか」+「する」

「もしか」に「する」が後続する場合、「もしか」と「する」とを切り離さない。

【例】 | もしか=すると | | もしか=して | | もしか=したら |

4.12 「{ こう/そう/ああ/どう }+{ いう/いった }+{ 体言/準体助詞 }」

「{ こう/そう/ああ/どう }+{ いう/いった }+{ 体言/準体助詞 }」については、指示副詞と動詞とを切り離さない。

【例】 | 彼らは | どう=いう | 風に | 思うだろうか |
| また | こう=いった | 施設が | 近くに | ある | ことも |

4.13 引用成分+「みたい」

引用成分を受ける「みたい」(形状詞)の前では切らない。

【例】 | 東京から | 来たのかい=みたいな | 乗りで |

4.14 連濁

連濁している複合語は切らない。

【例】 | 思い出=^{ぶか}深い | | 印象=^づ付けられた |

ただし、規則 3.4 が適用される場合は、そちらを優先する。

【例】 | 最終的には | 耳 | 人間の | 耳 || ^{たよ}頼り | なんですけど |

4.15 口語表現

以下に挙げる口語表現については、全体を一続きのものとする。

(体言)n=ち	<	「のうち(家)」の口語形
こん={時/中}	<	「この{時/中}」の口語形
そん={時/中}	<	「その{時/中}」の口語形
あん={時/中}	<	「あの{時/中}」の口語形

【例】 | 僕n=ちは | | こん=中で | | そん=時は |

5 細則 [B] 話し言葉に特有の現象に対する処理

5.1 言い直し

言い直しについては、以下のように処理する。

1. 語の一部のみを述べた直後に語全体を言い直している場合 切り離す

【例】 | 益岡・田窪氏の | 基本日本語 || 基礎日本語文法 (D2 の) での |
| 昭和十六年 | (F えー) | 太平洋開戦 || 太平洋戦争開戦の | 年に | 発表した |

2. 前に述べた語の一部のみを直後で言い直している場合 切り離す

【例】 | 先程 | 申し上げました | (F あー) | 阪倉篤義さん || 篤義先生の |
| 語幹と | 活用 (D ろ) 語尾とに | 切って (D い) おく || おく | 方が |
| 群れで | 生活する || してる | 動物ですから |

3. 前に述べた語全体を言い直している場合 切り離す

【例】 | 長野県の | (F えー) | 高原農家 || 高原野菜農家で | 働いている |
| 向こうで | 教育機関 || 教育事業 | 始めたいという | ことで |
| この | 場所 | における || においては |⁵

4. 一つの語の内部に言い直しがある場合 切り離さない

【例】 | 国立=日本語=国語研究所 |

言い直しのうち、タグ (D2) が付与された要素については、規則 6.6 に従って処理する。

5.2 言い差し

言い差された部分の前後で切る。

【例】 | 国立教育 || 今度は | 国立国語研究所の | 話を | します |

言い差しのうち、タグ (D) が付与された要素については、規則 6.7 に従って処理する。

5.3 別の文節の割り込み

一つの文節となるべきものの内部に別の文節が割り込んだことが明らかでない場合、割り込んだ文節の前後で切る⁶。

【例】 | 車で | 約 || 出入口まで || 三分ぐらいの | 場所に | あります |
| 車で | 約三分ぐらい || 出入口まで || の | 場所に | あります |
cf. | 車で | 出入口まで | 約-三分ぐらい-の | 場所に | あります |

5.4 文末形式・節末形式の介在

体言的要素とそれを受ける助詞・助動詞の間に文末形式・節末形式が介在する場合、その形式の後で切る⁷。

【例】 | 縦軸が | (F え) | 弁別率 ですね || を | パーセンテージで | 示して |

⁵ この規則の適用により、助詞から始まる文節が発生することがある。

⁶ この規則の適用により、接辞・助詞・助動詞からなる文節が発生することがある。

⁷ この規則の適用により、助詞・助動詞からなる文節が発生する。

6 細則 [C] タグの処理

タグの扱いについて、以下のように定める。タグの仕様については、transcription.pdf を参照のこと。

6.1 タグ (R),(L),(X),(笑),(咳),(泣),(あくび)

これらは、タグの有無にかかわらず、上記までの規則に従って文節を認定する。

ただし、タグ (L),(X),(笑),(咳),(泣),(あくび) が転記基本単位を越える場合には、括り直す。

0000 00000.000-00000.000 L:	
さてと	& (L サテト)
0000 00000.000-00000.000 L:	
次は	& (L ツギワ
何だったっけ	& ナンダッタッケ)

6.2 タグ (O)

6.2.1 外国語が節以上の形式で出現した場合には、分かる範囲で統語上の切れ目 (節・文など) や音声上の切れ目 (ポーズ・イントネーション) と一致させる。

6.2.2 古語や方言が句以上の形式で出現した場合には、分かる範囲で上記までの規則に従って文節を認定する。

歌 (短歌, 和歌, 俳句) の場合には「五, 七, 五 (, 七, 七)」の単位で分割してもよい。

6.2.3 数式に関しては、規則 4.2 に従う。

6.2.4 一文節があまりにも長くなりすぎる場合、上記の規則にかかわらず適当な箇所です切る。

6.3 タグ (M)

タグ (M) については、以下の手順で処理する。

- (M) 要素全体を一旦一つの体言と見なした上で、(M) 要素の前後の文節を認定する。

【例】 | 助動詞 | (M れる) | (M られる) |

- (M) の値に対して上記までの規則を適用し、内部の文節を認定する。

【例】 | こういった | (M 五回 | 中日は | 二死 | 一・二塁から) という | 文章を |

(M) の内部に別の (M) が含まれる場合、以上の規定にかかわらず内側の (M) の前後で文節を切る⁸。

【例】 | 中括弧で | (M 国会 || (M スラッシュ)) || 平仮名で | (M こっかい || (M 中括弧閉じる) || では) |

6.4 タグ (W),(?) [値あり]

タグ (W),(?) の内部では、他の規則で切ることになっていても切らない。

(F んー)(F あの)	&	(F ンー)(F アノ)
空気	&	クーキ
入れれば	&	イレレバ
鳴るようになる	&	ナルヨー (W ナツ; ニナル)
(F ま)	&	(F マ)
世界の	&	セカイノ
半分	&	ハンブン
占める っていうんですね	&	シメル (? ッテユー) ンデスネ
よく	&	ヨク
(? 主に付いて, 思い付いて)	&	(? オモニツイテ, (W オモニツイ; オモイツイ) テ)
出てくるのに	&	デテクルノニ

⁸ この規則の適用により、接辞からなる文節が発生することがある。

6.5 タグ (A)

タグ (A) の内部では切らない。

複数の文節に跨がる要素は、一つの (A) にまとめず、括り直す。

- × | (A ビーシー五百六十六; B C 5 6 6) 年から | (A 四百八十六; 4 8 6) 年まで |
 - × | (A ビーシー | 五百六十六; B C 5 6 6) 年から | (A 四百八十六; 4 8 6) 年まで |
| (A ビーシー; B C) | (A 五百六十六; 5 6 6) 年から | (A 四百八十六; 4 8 6) 年まで |
- cf. 規則 3.4

6.6 タグ (D2)

6.6.1 一文節となるべきものの中に割り込んだ (D2) の連続は、その文節に含める。

【例】 | 住まい=(D2 を)=は | どこに | あるかって |
| 有名に | 限らず | 私立=(D2 小)=(D がっ) 中学校に | 入りたいという |

6.6.2 上記の規則に該当しない (D2) の連続については、各々の前後で切る。

【例】 | そのまま || (D2 お) || お見せしますけども |
| その || (D2 長) || (D2 長) || 長母音 | (F えー) | 促音などの | 特殊拍の |

6.7 タグ (F) (D)

6.7.1 一文節となるべきものの中に割り込んだ (F) (D) の連続は、その文節に含める。

【例】 | この | 比較=(F うー)=語彙研究序説というような | 本を |
| 地下鉄=(D ん)=に | 乗り=(F えー)=ながら |
| この | 三者の | 関係-(D2 に)=(D つい)=について |

6.7.2 文節の一部を構成していることが明らかな語の言い淀み・感情表出系感動詞については、その文節に含める。

【例】 | 予稿集では | (D しゃしん)=を | 斜線を | 引いていますけれども |
これは	言語音でも	(D2 非)=(D げごん)	非言語音でも	同じでした
言ったもんですから	(F へー)=って	驚かれまして		
村の	住人の	(D ディーエヌ)=鑑定	(A ディーエヌエー; DNA) 鑑定を	

6.7.3 上記の規則に該当しない (F) (D) の連続については、同種のものの連続の前後で切る。

【例】 | そういう | ことが || (F えーっと) || (D 都み) || 都民なら | 誰でも |
| いつも || (F あのー) || (D う)=(D つ) || 凄く | たくさんの | 自転車が |
| でも || (F まー)=(F そのー) || (D おおが) || (F あのー) || 大きい | 立体駐車場も |

6.8 タグ (K)

タグ (K) の内部では切らない。

6.9 タグ <H>, <Q>, <P>

タグ <H>, <Q>, <P> は、直前の要素が属する文節に含める。

6.10 タグ <FV>, <息>, <笑>, <咳>, (?) [値なし]

これらのタグは、直前の要素が属する文節に含める。ただし、転記基本単位の先頭に出現する場合は、後続発話の先頭に付与する。

これまで<FV> & コレマデ<FV>
0000 00000.000-00000.000 L:
少しの & <息>スコシノ
0000 00000.000-00000.000 L:
(?) & (?)
0000 00000.000-00000.000 L:

7 文節分割例

	0065 00182.393-00185.239 L: (F えー) そもそも (A エイチアールディー; HRD) あるいは 企業内教育っていうのは	& & & & &	(F エー) ソモソモ エイチアールディー アルイフ キギョーナイキョーイクツテユーノワ
規則 4.12	0066 00185.487-00186.993 L: どういう ことを	& & &	ドーユー コトオ コーリヨ(W シュ; シ) ナイト
規則 4.9	考慮しないと	&	
規則 6.7.2	0067 00187.225-00189.142 L: (A アイティー; IT)(D 技)	&	アイティー (D ギ)
規則 6.7.3	(D アイ)	&	(D アイ)
規則 4.6	(A アイティー; IT)っていう <u>もの</u> が	&	アイティー (? ッテ)(W ユ; ユウ) モノガ
	0068 00189.360-00191.213 L: (F あー) 浸透していかないのかと	& &	(F アー) シントーシテイカナイノカト
(一文節)	0069 00191.419-00195.179 L: いう ことを	& & &	ユー コトオ (F マー)(F アー)
規則 6.7.3	(F まー)(F あー) 初めに	& &	ハジメニ
	0070 00244.094-00245.296 L: 聴覚研究会で	&	チョーカクケンキューカイデ
規則 4.5	0071 00245.837-00247.832 L: 御報告さ せていただいてるんですけども	&	ゴホーコクサセテイタダイテルンデスケレドモ
	0072 00248.203-00249.419 L: 簡単に 概要だけ	& & &	カンタンニ ガイヨードケ
規則 4.5	0073 00249.597-00249.749 L:<雑音> 0074 00249.815-00250.982 L: 御説明いたしますと	& &	ゴセツメーイタシマスト
	0075 00251.855-00258.888 L: 一言で 言うつと	& & &	(W ヒトゴト; ヒトコト) デ ユート エージーシーニ
規則 6.7.3	(A エージーシー; AGC) に (F えー)(F ま)	& & &	(F エー)(F マ) ダイナミックレンジノ
規則 3.1	ダイナミックレンジの 調整 機能が 入っているという 形に なっています	& & & & & &	チョーセー キノーガ ハイツテイルトユー カタチニ ナツテイマス
規則 4.6	0076 00259.440-00264.279 L: (F (? ま)) 補聴器の ように	& &	(F (W (? ワ; (? マ))) ホチョーキ<H>ノヨーニ
	(F その) 短時間で 瞬時に	& & &	(F ソノ) タンジカンデ シュンジン
規則 4.9, 4.6	圧縮する ような 方法というのは	& &	(W アッシュク; アッシュク) スルヨーナ ホーホートユーノワ
規則 4.10	どうしても (F その)	& &	ドーシテモ (F ソノ)
	0077 00264.893-00269.786 L:		

	0447 01127.253-01130.327 L:		
	コーパスを	&	コーパスオ
	良く	&	ヨク
	するという	&	スルトユー
	方向に	&	ホーコーニ
規則 6.7.1	進め (D たす) たいと	&	ススメ (D タス) タイト
	思っております	&	オモツテオリマス
	0448 01132.332-01137.614 L:		
	(F エー)	&	(F エー)
	で	&	デ
	以上	&	(W ジョー; イジョー)
	コーパスですが	&	コーパスデスガ<H>
	(F エ)	&	(F エ)
	それを	&	ソレオ
	使った	&	ツカッタ
規則 3.3	今度は	&	コンドウ
	音声認識	&	オンセーニンシキ
	音声理解の	&	オンセーリカイ<H>ノ
	研究課題としては	&	ケンキューカダイトシテワ
	0449 01138.115-01140.363 L:		
規則 3.3.1, 4.9	認識理解する	&	ニンシ (? キ) リカイスル
	為の	&	タメノ
規則 3.3	音響モデル	&	オンキョーモデル
	言語モデル	&	ゲンゴ (W モデル; モデル)
	0450 01140.602-01141.014 L:		
	それを	&	ソレオ
	0451 01141.453-01142.548 L:		
	(F エー)	&	(F エー)
規則 4.9	要約する	&	ヨーヤクスル
	モデル	&	モデル
	0452 01143.142-01144.724 L:		
	(F エー)	&	(F エー)
規則 4.9	話題を	&	ワダイオ
	抽出する	&	チューシュツスル
	方法	&	ホーホー
	0453 01145.149-01145.677 L:		
	それから	&	ソレカラ
	0454 01146.133-01147.905 L:		
	(F エー)	&	(F エー)
規則 4.9	精度を	&	セードオ
	向上させる	&	コージョーサセル
	為の	&	タメノ
	0455 01148.159-01150.258 L:		
規則 3.3	話者適応	&	ワシャテキオー
	雑音適応	&	ザツオンテキオー
	タスク適応	&	タスクテキオー
(一文節)	0456 01150.590-01152.345 L:		
	などの	&	(W ナト; ナド) ノ
	(D アルオ)	&	(D アルオ)
	アルゴリズムの	&	アルゴリズムノ
	研究が	&	ケンキューガ
	0457 01152.803-01153.580 L:		
	(F エー)	&	(F エー)
	必要です	&	ヒツヨーデス
	0458 01154.642-01156.271 L:		