

『現代日本語書き言葉均衡コーパス』  
かかり受け・並列構造アノテーション作業メモ  
Version 0.7.0

浅原正幸

平成 25 年 5 月



# 目次

第 1 章	はじめに	1
1.1	作業概要	1
1.2	本文書におけるかかり受け・並列構造アノテーションの表現法	1
1.2.1	かかり受け関係の表現法	1
	通常のかかり受け関係の表現法	1
	倒置の表現法	2
	交差の表現法	2
	かかり先のない文節の表現法	3
	“D” ラベル	4
	文節境界修正 “B” / “B+” ラベル	4
1.2.2	並列構造・同格構造の表現法	4
	「セグメント」と「グループ」の導入	4
	並列構造の表現法	4
	同格構造の表現法	5
	その他の「セグメント」と「グループ」	5
第 2 章	BCCWJ かかり受けアノテーション基準：概論	7
2.1	通常のかかり受け関係	7
2.1.1	格要素と述語	7
2.1.2	連体修飾（体言と体言）	7
2.1.3	連体修飾（連体修飾節の述語と被修飾語）	7
2.1.4	連用修飾（副詞や活用語連用形）	7
2.1.5	連用修飾（従属節の述語と主節の述語）	7
2.2	かかり先なしを認めるもの	7
2.2.1	感動詞	7
2.2.2	記号・補助記号	8
2.2.3	URL	9
2.2.4	空白	9
2.2.5	接続詞	9
2.2.6	その他	10
2.3	英単語・ローマ字文・漢文（・古文）	10
2.4	言い直し・言いよどみ	11
2.5	並列構造	11
2.5.1	並列構造範囲アノテーション	11
2.5.2	名詞句の並列	11
2.5.3	述語並列	12
2.5.4	部分並列内の関係	12
2.5.5	並列構造と接続詞	12
2.5.6	3 つ以上の並列構造	12
2.5.7	並列構造の複数の要素に左からかかる場合	13
2.6	同格関係	13
2.6.1	同格構造範囲アノテーション	13
2.6.2	同格関係	13
2.6.3	具体例と総称の同格関係・具体例と数詞の同格関係	14
2.7	文節の修正	14
2.8	文境界の修正	14
2.9	「格表示誤り」などの取り扱い	15
2.10	前処理（自動処理）一覧	15
第 3 章	BCCWJ におけるアノテーション基準：各論	17
3.1	従属節とかかり受け	17
3.2	連用修飾	18
3.2.1	副詞と呼応	18
第 4 章	かかり受け関係ラベルの比較～ CSJ のマニュアルに基づく比較 (1)	19
4.1	かかり受け関係ラベルのまとめ	19
4.2	かかり受け構造、並列構造、同格構造	19
4.3	文節境界を示すラベル	21
4.4	節境界を示すラベル	22
4.5	その他のラベル	22
4.6	コメントの記述	24
4.7	複数ラベルの併記	24

第 5 章	CSJ のマニュアルに基づく比較 (2)	27
第 6 章	KC のマニュアルに基づく比較	29
6.1	通常のかかり受け関係 . . . . .	29
6.1.1	格要素と複数の述語の関係 . . . . .	29
6.1.2	かかり先が非常に曖昧な場合 . . . . .	31
6.2	並列構造アノテーション . . . . .	32
6.2.1	並列構造一般 . . . . .	32
6.2.2	部分並列 . . . . .	32
6.2.3	括弧内の複数文 . . . . .	33
6.2.4	テ形 . . . . .	34
6.2.5	連体形 . . . . .	34
6.2.6	からまで . . . . .	34
6.3	同格構造アノテーション . . . . .	35
6.3.1	住所・職業・続柄と人名 . . . . .	35
6.3.2	範囲を表す「からまで」 . . . . .	36
6.3.3	「体言+ら   たち   その他   など   と   すなわち   つまり   とりわけ   特に」 . . . . .	37
6.3.4	「体言 (+、) + 「～」」 . . . . .	39
6.3.5	「用言+など」 . . . . .	40
6.3.6	節とそれをまとめる名詞 . . . . .	40
6.3.7	同格と交差 . . . . .	41
6.3.8	A の BC . . . . .	41
付録 A	Q/A	43
A.1	Version 0.2.0 から Version 0.3.0 への変更における Q/A . . . . .	43
A.2	Version 0.1 から Version 0.2.0 への変更における Q/A . . . . .	43
付録 B	変更履歴	45
B.1	Version 0.6.0 から Version 0.7.0 への変更履歴 . . . . .	45
B.2	Version 0.5.0 から Version 0.6.0 への変更履歴 . . . . .	45
B.3	Version 0.4.0 から Version 0.5.0 への変更履歴 . . . . .	45
B.4	Version 0.3.0 から Version 0.4.0 への変更履歴 . . . . .	45
B.5	Version 0.2.0 から Version 0.3.0 への変更履歴 . . . . .	45
B.6	Version 0.1 から Version 0.2.0 への変更履歴 . . . . .	45
参考文献		47
索引		48

## 第 1 章

# はじめに

### 1.1 作業概要

本文書では「現代日本語書き言葉均衡コーパス」(以下 BCCWJ) に対するかかり受け・並列構造アノテーションにおける、2 次チェックアノテータの作業基準を示す。

アノテーション作業は 2011 年 12 月に頒布されている BCCWJ DVD データを底本とする。

作業において、かかり受け構造と並列・同格構造の両方を確認し修正する。かかり受け構造修正作業は、文節単位のかかり先を判別する。あらかじめ、自動解析システムにより解析されたものがかかり受け構造 1 次チェックアノテータにより修正されている。今回、新たにラベルを導入したり、かかり受け関係の定義が変更した点を中心に修正する。

並列・同格構造は、国語研短単位をもっとも短い単位として、対応する範囲を入れ子を許して同定する。あらかじめ、並列・同格構造 1 次チェックアノテータによって付与されたものが、かかり受け構造 1 次チェックアノテータによりかかり受け関係との整合性を見ながら修正されている。今回、再度かかり受け関係を修正するとともに、新しい同格 (Generic) を導入する部分を中心に修正する。

DVD に収録されている形態論情報(国語研短単位)については改変しないが、明らかに不備がある部分については、メモを残すこと。

DVD に収録されている文節境界、文境界については、元データを尊重したうえで、かかり受け関係、B ラベルや Z ラベルを用いて、修正を行う。

作業は、各レジスタ毎に国語研で定義しているファイルの優先順位に基づいて行う。標準的な作業者の作業速度は 48 文/時間である。

### 1.2 本文書におけるかかり受け・並列構造アノテーションの表現法

本節では、本文書におけるかかり受け・並列構造アノテーションの表現法を示す。いくつかアノテーション基準を示しているが、本節は表現法とアノテーションツール ChaKi.NET 上での表示の説明を目的としており、より詳細な記述は 2 節以降にゆずる。

BCCWJ のかかり受け構造、並列・同格構造アノテーション基準は、先行研究である日本語話し言葉コーパス [?] のアノテーション基準や京都大学テキストコーパスのアノテーション基準 [?] と異なっている。

BCCWJ の基準については BCCWJ と記し、日本語話し言葉コーパスのかかり受けアノテーション基準については CSJ と記し、京都大学テキストコーパスのかかり受けアノテーション基準については KC と記すことで区別する。

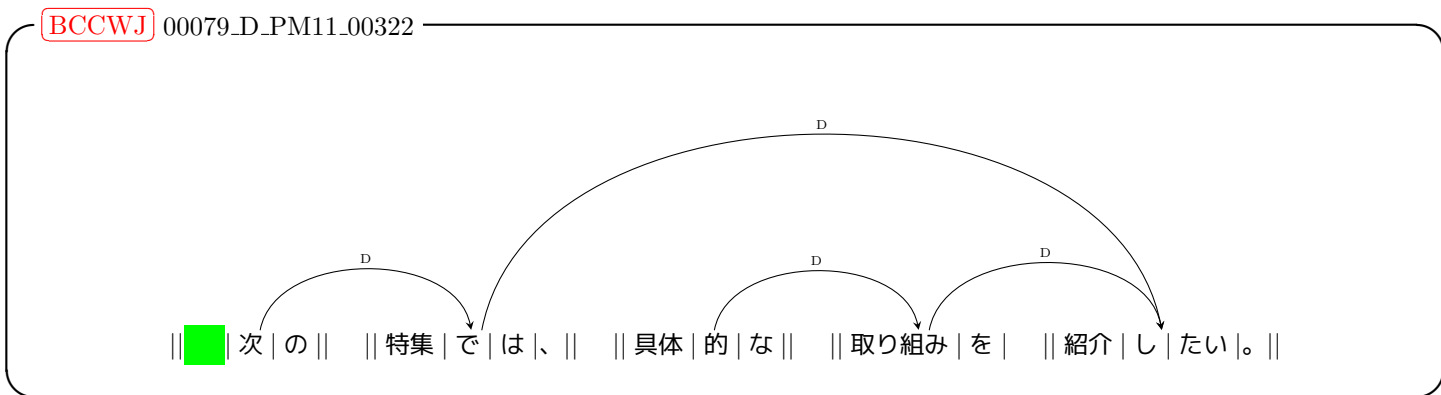
BCCWJ の右に付与されている XXXXX.Y\_ZZZZ.ZZZZZ は用例が出現したファイルを示す。XXXXX が作業の優先順位、Y が優先順位を表す部分集合名、ZZZZ.ZZZZZ が DVD に収録されている元データのファイル名である。

また BCCWJ 2 次チェックで新たに導入された規則を BCCWJ 2 次チェック と記す。かかり受け修正作業においては BCCWJ 2 次チェック と記されている箇所のみならず、全てをチェックする。

#### 1.2.1 かかり受け関係の表現法

通常のかかり受け関係の表現法

かかり受け関係は以下のような図で示す。



| は国語研短単位形態素境界 (BCCWJCSJ) もしくは JUMAN 品詞体系形態素境界 (KC) を表す。かかり受け関係の説明に形態素境界が問題にならない場合には、省略する。

|| は国語研文節境界 (BCCWJCSJ) もしくは京都大学テキストコーパス文節境界 (KC) を表す。あいまい性がない場合には空白をあけるのみで表現する。

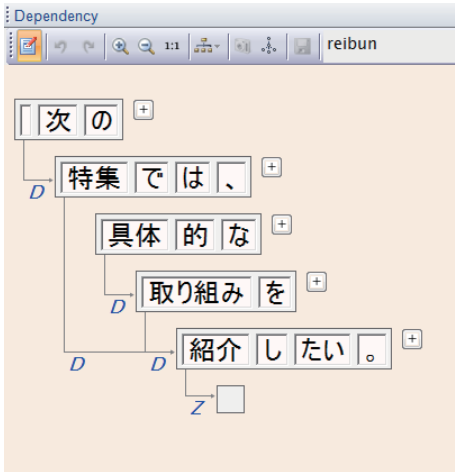
全角空白を明示的に示すために、  を用いる。

かかり受け関係は文を表す文節列上に矢印で表現する。格要素 → 述語、修飾語 → 被修飾語などの方向に付与する。

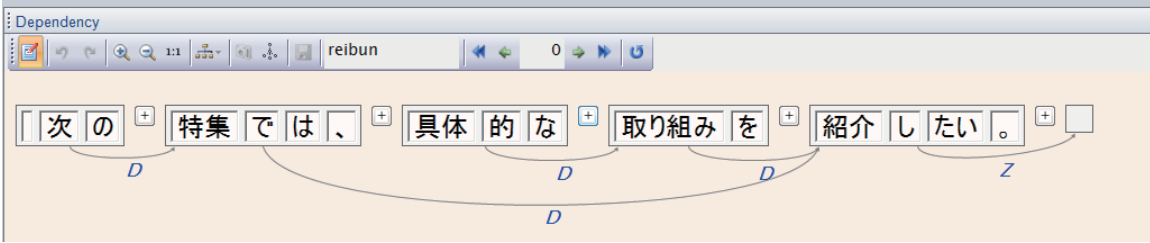
アノテーションの際に利用するアノテーションツール ChaKi.NET の Dependency Panel 上では以下のような 2 種類の表示方法がある。

BCCWJ

- Diagonal Mode



- Horizontal Mode



どちらの Mode で作業しても構わないが、本文書では Diagonal Mode で以下表示する。尚、各モードの末尾の□は文の外の DUMMY ノードで、最右要素などのかかり先がない要素をかける \*1 ためのノードである。

DUMMY ノード

倒置の表現法

[KC]の基準においては、Strictly Head Final の原則から常に左から右にかかる。[BCCWJ][CSJ]の基準においては、右から左にかかることを許す。

D ラベル

F ラベル

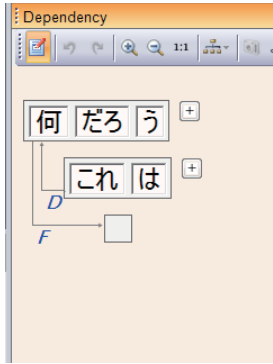
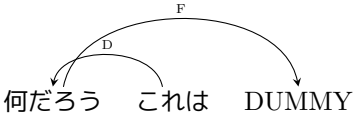
倒置

[CSJ] では右から左にかかることをラベル “R” を用いて明示するが、[BCCWJ] においては特に明示しない。

[BCCWJ] において、最初の「何だろう」はかかり先なしの根ノードになるが、アノテーションツール上では末尾の DUMMY ノードにかけることにより表現する。倒置箇所は通常のかかり受けとして “D” を用いる。倒置などによりかかり先が不定になる箇所には DUMMY ノードにかけて “F” を用いる。

奈良先端大でかかり受け関係の区別を不要とする場合には “F” を “D” に一括返還すること。

BCCWJ



交差の表現法

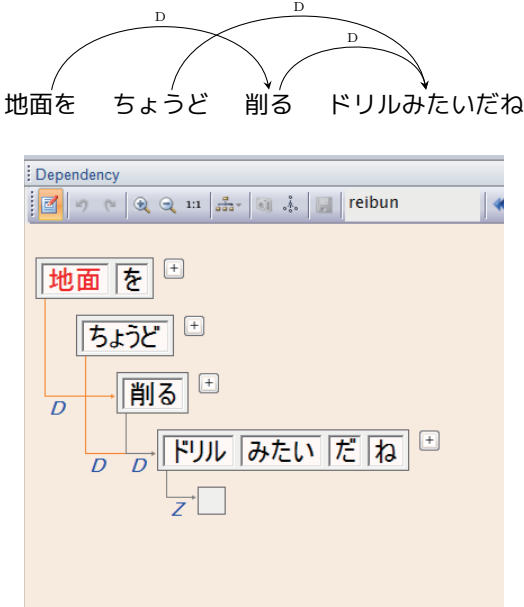
[KC]の基準においては、非交差制約の原則からかかり受け関係が同格表現以外においては交差することを許さない。

交差

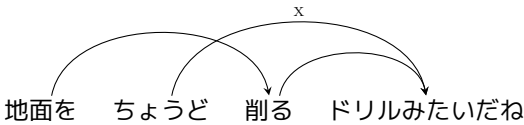
[BCCWJ][CSJ]の基準においては、かかり受け関係が交差することを許す。[CSJ] ではかかり受け関係が交差することをラベル “X” を用いて明示するが、[BCCWJ] においては特に明示しない。ChaKi.NET の GUI 上では、交差があった場合にはかかり受け関係の色が自動的にオレンジに変更される。

\*1 かかり受けを付与することを便宜上「かける」と表現する。

BCCWJ



CSJ



かかり先のない文節の表現法

[KC]の基準においては、かかり先のない文節を許さない。[BCCWJ][CSJ]の基準においては、かかり先のない文節を許す。

[BCCWJ]においてはかかり先のない文節を DUMMY ノードにかけることによって表現する。作業に利用するアノテーションツール ChaKi.NET においては、かかり先を文の末尾の空要素にかけることによってかかり先がないことを表現するため、本文書でも、文の末尾に DUMMY ノードをおき、そこにかけることでかかり先のないことを示す。かかり先のない文節については、その分類を後述するかかり受け関係ラベル分類に基づいて記述する。

尚、[CSJ]においてはかかり先のない文節を ラベル “N” により表現する。

以下の例では「中学校を」が DUMMY ノードにかかる。文末の「いたんですね」もかかり先がなく DUMMY ノードにかかる。

[BCCWJ] において、文末以外のかかり先がないノードは ラベル “F”、文末はラベル “Z” として区別する。DUMMY ノードを図に記述しない場合にはこのかかり受け関係は省略する（がアノテーションする）。

奈良先端大ではかかり受け関係ラベルの区別を不要とするが、本規定ではそれを採用しない。かかり受け関係ラベルの区別を不要とする場合には “F”、“Z” を “D” に一括変換すること。

繰り返しになるが[CSJ] において、かかり先がないノードは “N” を用いる。

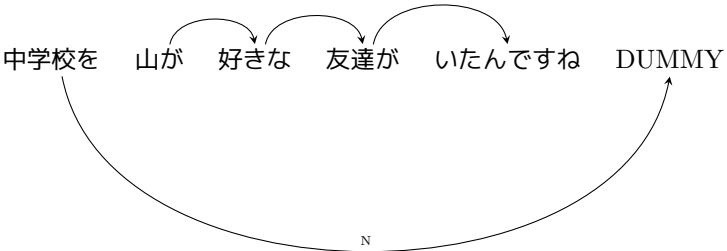
かかり先なし  
F ラベル  
DUMMY ノード

F ラベル

BCCWJ



CSJ



“D” ラベル

`BCCWJ``KC`においてラベル “D” は通常のかかり受け関係を示し、`CSJ`においてはラベル “D” は言い直しを示す（通常のかかり受け関係はラベルなし矢印）。一つのかかり受け木で `CSJ` を含む複数のかかり受け基準が一致することを表現する場合、あいまい性がない場合において、`BCCWJ``KC`におけるラベル “D” 相当のかかり受け関係を、ラベルなし矢印で表現することとする。

`BCCWJ``CSJ``KC`の 3 つの基準が一致する場合の “D” ラベルの省略

`CSJ`言い直し (“D” ラベル) の例

D ラベル

文節境界修正 “B” / “B+” ラベル

文節境界を修正する記述をかかり受け関係ラベル `BCCWJ`において “B” ラベル、`CSJ`において “B+” を用いて表現することがある。その場合、かかり受け関係は文節列の下に付与する。アノテーションツール上でも、ラベルを `BCCWJ`において “B”、`CSJ`において “B+” にすることにより付与する。

尚、形態論に基づく国語研文節単位はそのままとする。ChaKi の Dependency Panel 上で「はさみツール」や「+」を用いて、文節単位を修正することは行わない。

かかり受けアノテーションにそぐわない文節単位のみを、上のようなかかり受け関係を用いて表現する。

`BCCWJ`

`CSJ`

文節境界修正

B ラベル

B+ ラベル

1.2.2 並列構造・同格構造の表現法

「セグメント」と「グループ」の導入

`BCCWJ`において、かかり受け関係とは別に、文に含まれる並列構造・同格構造をアノテーションする。並列構造もしくは同格構造の構成要素に対しては「セグメント」と呼ばれる範囲を国語研短単位形態素境界に基づいてアノテーションし、複数の「セグメント」を「グループ」化することによりアノテーションを行う。

アノテーションツール上では下のように表示されるが、本文書では下図のように角丸四角により「セグメント」を示し、隣接する角丸四角間を無向リンク（矢印なしのリンク）で結ぶことにより「グループ」化を表す。角丸四角の範囲が示す文節が「セグメント」を表現するのではなく、角丸四角の中の単語列が「セグメント」を表現することに注意すること。

並列構造の表現法

並列構造を表現するために実線の角丸四角およびリンクを利用する。並列構造の「セグメント」と「グループ」を “Parallel” と呼ぶ。

セグメント

グループ

並列

Parallel セグメント

Parallel グループ



BCCWJ00079\_D\_PB46\_00066

|| 売れる || 時 | に || 売れる || 商品 | 構成 | と || 存在 | 感 | を || 持つ ||

Parallel

Dependency

売れる

D

時に

D

売れる

D

商品構成と

D

存在感を

D

持つ

Z

同格構造の表現法

同格構造を表現するために点線の角丸四角およびリンクを利用する。同格構造の「セグメント」と「グループ」を Apposition と呼ぶ。

BCCWJ

米国大統領    ジョン・F・ケネディ が    暗殺された

Apposition

Dependency

1: Apposition

米国大統領

D

1: Apposition

ジョン・F・ケネディが

D

暗殺された

Z

同格  
Apposition セグメント  
Apposition グループ

その他の「セグメント」と「グループ」

[BCCWJ]においては、Parallel と Apposition の他に Generic という「セグメント」と「グループ」を規定する。これらは破線の角丸四角およびリンクで表示する。

BCCWJ

泥棒は    指輪など、    多数の    高級品    を    盗んだ

Generic

Dependency

泥棒は

D

1: Generic

指輪など、

D

1: Generic

多数の

D

高級品を

D

盗んだ

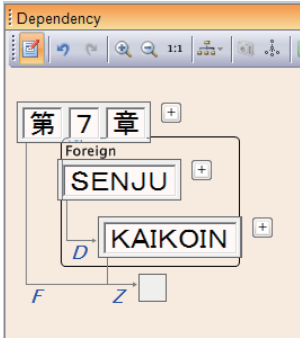
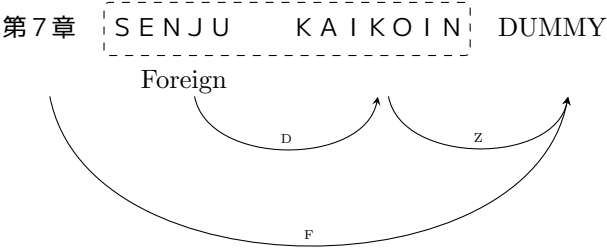
Z

Generic セグメント  
Generic グループ

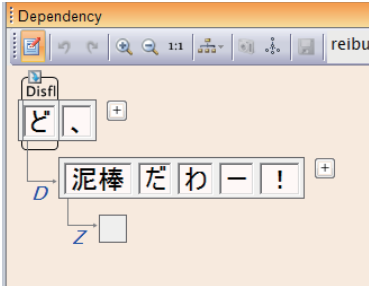
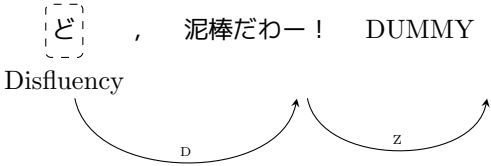
さらに、[BCCWJ]においては、Foreign, Disfluency などの「セグメント」を規定する。これらは破線の角丸四角で表示する。

Foreign セグメント  
Disfluency セグメント

BCCWJ 00004\_A\_PB22\_00002



BCCWJ 00329\_A\_OC03\_00651



第2章

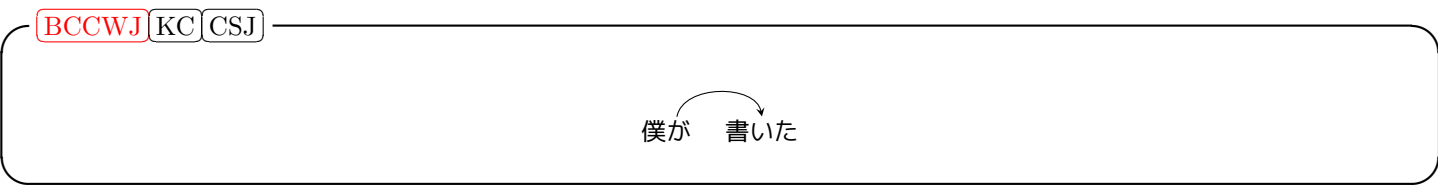
BCCWJ かかり受けアノテーション基準：概論

2.1 通常のかかり受け関係

かかり受け関係は、国語研文節単位に対して、主語-述語の関係、目的語（その他格要素）-述語の関係、修飾-被修飾の関係、並列・同格の関係、接続-被接続の関係などを付与する。

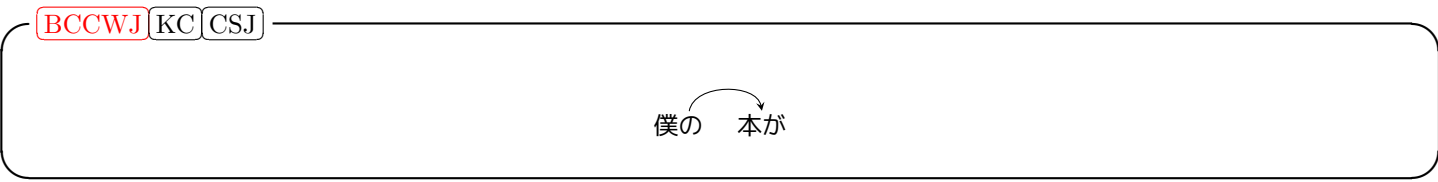
2.1.1 格要素と述語

格要素



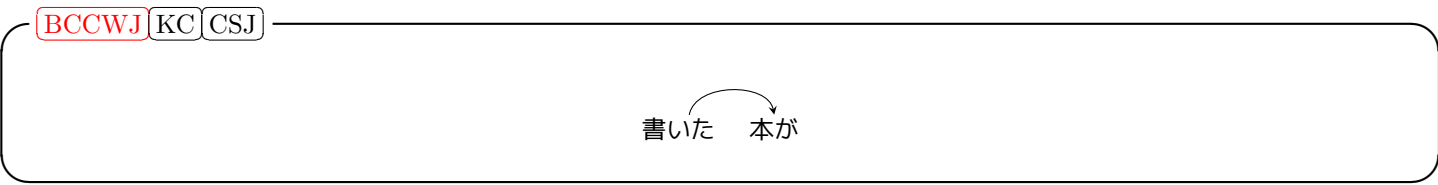
2.1.2 連体修飾（体言と体言）

連体修飾



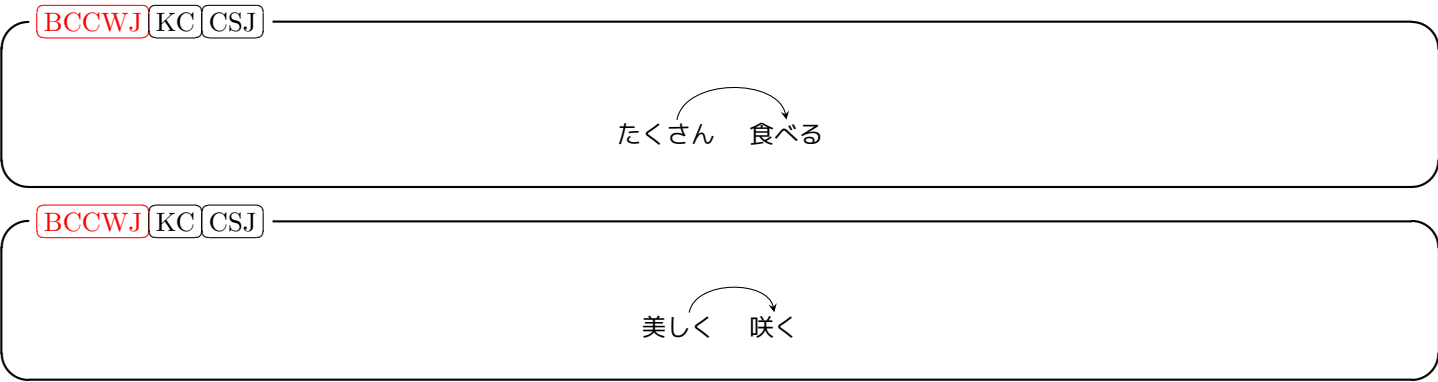
2.1.3 連体修飾（連体修飾節の述語と被修飾語）

連体修飾

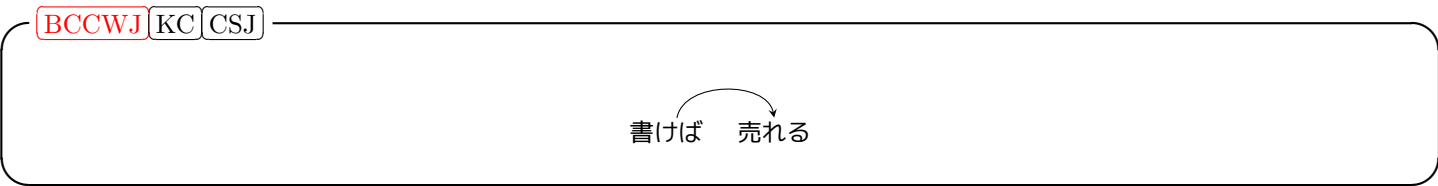


2.1.4 連用修飾（副詞や活用語連用形）

連用修飾



2.1.5 連用修飾（従属節の述語と主節の述語）



2.2 かかり先なしを認めるもの

[BCCWJ 2 次チェック](#)  
[BCCWJ]では[CSJ]にならい、かかり先なしの要素を許す。BCCWJ アノテーションにおいては、ラベル “D” のかかり受け関係が 2.1 節に示すような統語的に重要なものを表現するが、ラベル “F” のかかり受け関係は統語的に重要でない不定の表現を表現する。後述する文末ラベル “Z” とラベル “F” との両方が考えられる場合は “F” を優先する。  
「感動詞」「補助記号」「URL」などはラベル “F” でかかり先なしとしてアノテーションする。  
尚、このような事例について、かかり先を DUMMY ノードではなく、文の最右要素（もしくは、ラベル “Z” のかかり元となる要素のうちで最右要素）として認定したいという要望があるが、本規定ではそれを採用しないが、復元できるよう、ラベル “F” 相当は残す。

F ラベル

Z ラベル

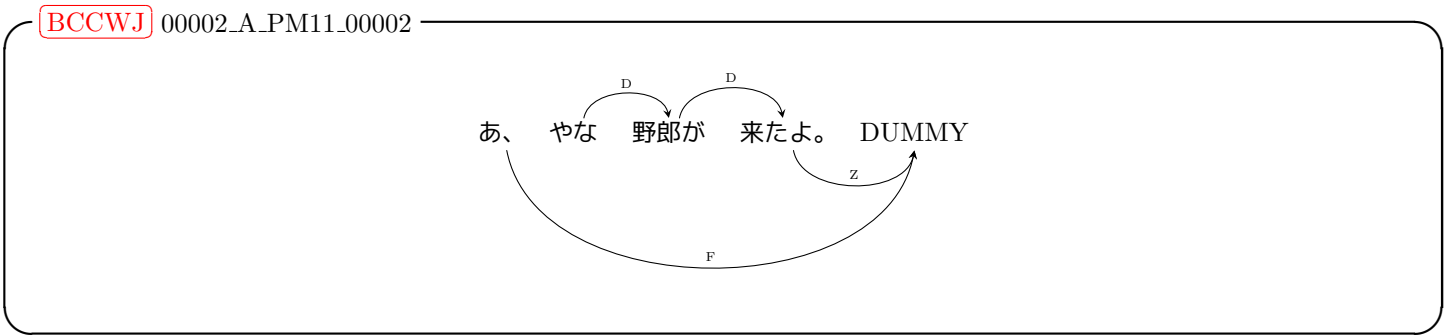
2.2.1 感動詞

感動詞

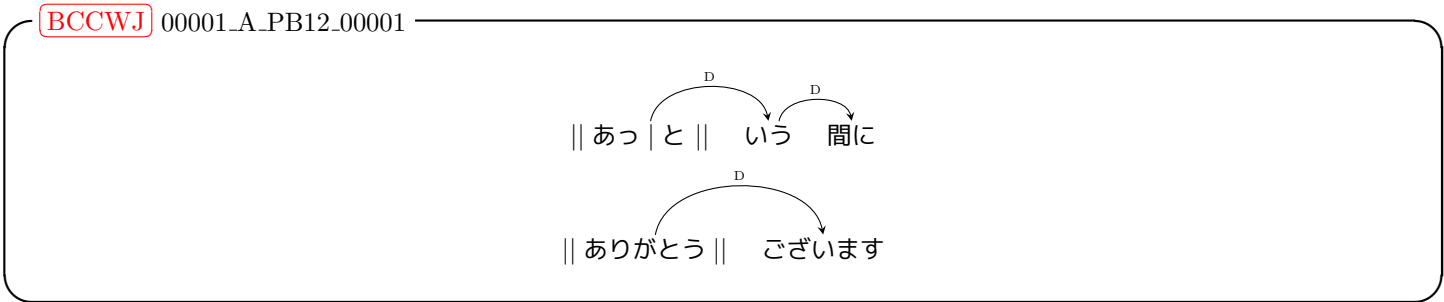
[BCCWJ 2 次チェック](#)

感動詞などが（記号などを除いて）単体で出現する文節のかかり先が決定できない場合にはラベル “F” でかかり先なしとしてアノテーションする。

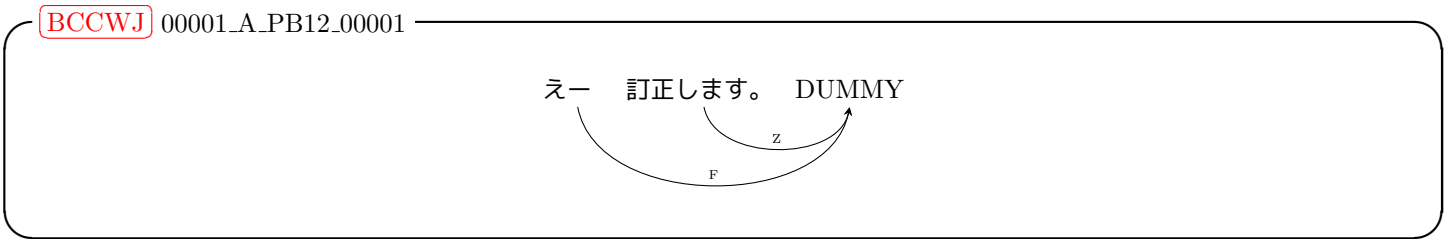
尚、自動変換プログラムにより、感動詞を含む文節のかかり受けラベルは “FX” が付与される (878 事例)。これらについては人手で確認すること。  
さらに文節内に感動詞のみ含まれるもので「ありがとう」以外のものはかかり受けラベルは “F” が付与されかかり先を DUMMY に変更される (90 事例)。



以下のように引用の「と」など付属語と一緒に出現し、かかり先が明示的にわかる場合には、通常のかかり受け関かかのアノテーションを行う。



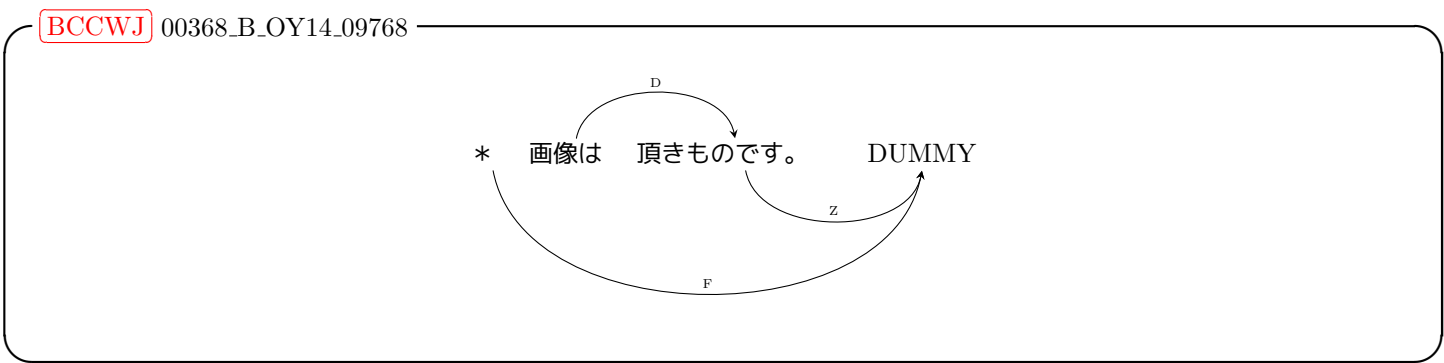
また、BCCWJ の品詞体系ではフィラーは感動詞の下に分類されているが、ラベル “F” でかかり先なしとしてアノテーションする。



2.2.2 記号・補助記号

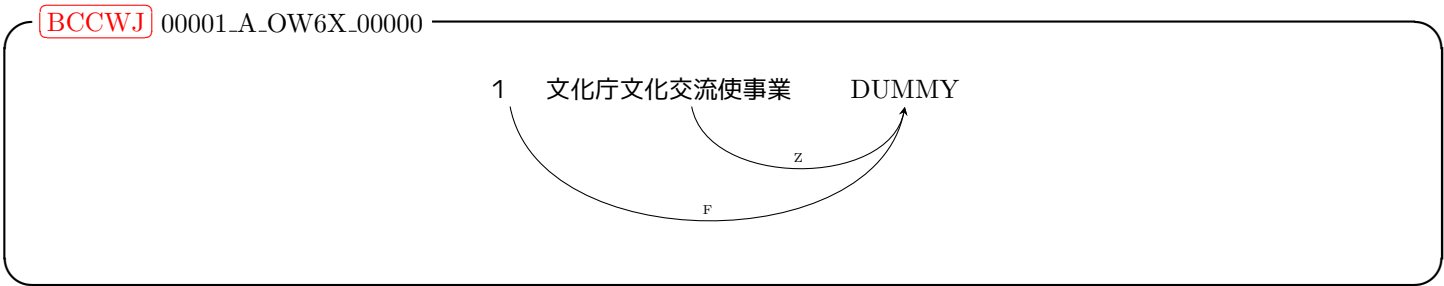
BCCWJ 2 次チェック

記号・補助記号が単体で文節に出現し、かかり先が不明な場合、ラベル “F” でかかり先なしとしてアノテーションする。



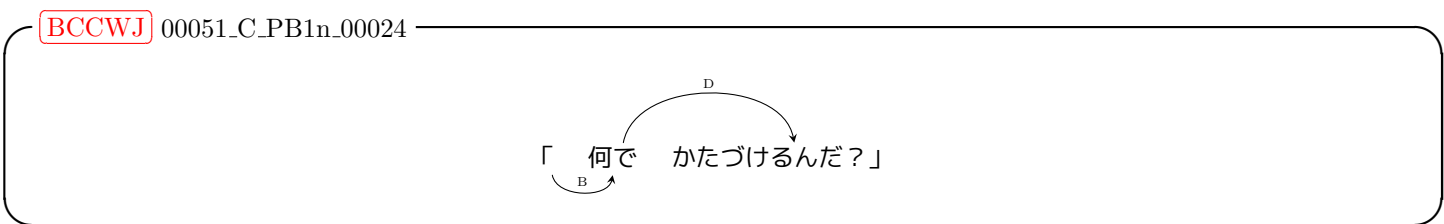
拡大解釈し、リストマーカー相当句についても、かかり先が不明な場合、ラベル “F” でかかり先なしとしてアノテーションする。  
記号のみを含む文節のかかり受けラベルは、リストマーカー相当かどうかを確認して、人手で、かかり先を DUMMY に変更すること (記号-一般 53 事例、記号-文字 12 事例)。

尚、自動変換プログラムにより、補助記号-一般のみを含む文節のかかり受けラベルは “F” が付与され、かかり先を DUMMY に変更する (補助記号-一般 1677 事例)。



1 文節で独立した括弧開は、右隣接する文節に “B” ラベルで結合する。結合された文節は、括弧以外の要素（主辞相当）がかかるものにかかる。

尚、自動変換プログラムにより、補助記号-括弧開のみを含む文節のかかり受けラベルは “B” が付与され、かかり先を右隣接要素に変更する (補助記号-括弧開 1090 事例)。



これに対して、文節内の括弧は、括弧以外の要素（主辞相当）がかかるものにかかる。

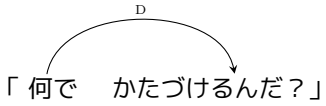
記号

補助記号

リストマーカー

括弧

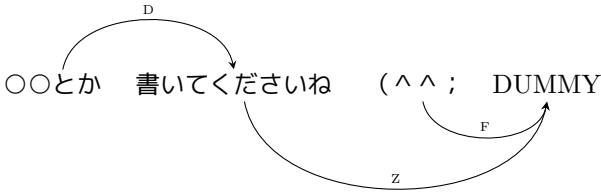
**BCCWJ** 00051\_C\_PB1n\_00024 の文節が分かれていない場合



1 文節で独立した括弧閉が 10 件出現するが、いずれも文頭に出現するものである。  
これらは文境界にまつわる修正において人手で対処する（前処理で作業済みだが作業がまだなものを見つけたら報告する）。当該文節のかかり先は文境界修正時に不定になるために必ず作業時に確認すること。  
また、**BCCWJ**の品詞体系では顔文字などのアスキーアートは補助記号の下に分類されているが、ラベル“F”でかかり先なしとしてアノテーションする。  
尚、自動変換プログラムにより、補助記号-AA-顔文字のみを含む文節のかかり受けラベルは“F”が付与され、かかり先を DUMMY に変更する（補助記号-AA-顔文字 33 事例）。  
奈良先端大においては顔文字はパンクチュエーションと拡大解釈し B で隣接する文節に縮退させるが、本規定ではそれを採用しない。採用する場合は 33 事例のかかり受けを自動で修正すること。

顔文字

**BCCWJ** 00004\_A\_OY15\_00014



2.2.3 URL

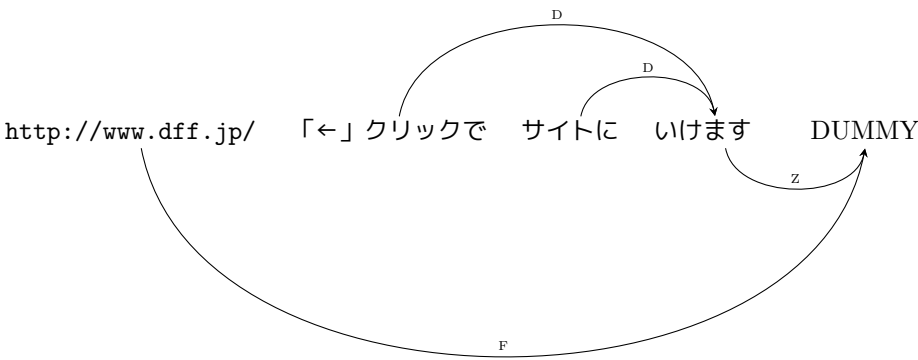
**BCCWJ 2 次チェック**

URL や、電話番号、店の営業時間など、かかり受け関係として構造しがたい表現については、ラベル“F”でかかり先なしとしてアノテーションする。  
自動変換プログラムにより、URL を含む文節のかかり受けラベルは“F”が付与され、かかり先を DUMMY に変更する（URL 223 事例）が、電話番号などは人手で確認すること。

URL

電話番号

**BCCWJ** 00382\_B\_OY14\_11567



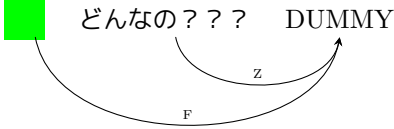
2.2.4 空白

**BCCWJ 2 次チェック**

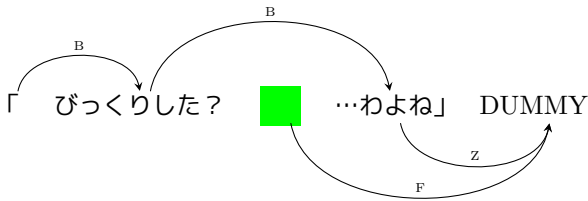
独立して文節をなす空白は自動でかかり受けラベル“F”とし、かかり先なしとしている。空白を排除した結果、後ろの文節が付属語のみであるなど前後の文節が 1 つの文節をなす場合がある。人手で確認すること。  
奈良先端大においては空白を B で隣接する文節に縮退させるが、多くは雑誌や Web テキスト上のレイアウトのための空白であり本規定ではそれを採用しない。採用する場合は 267 事例のかかり受けを自動で修正すること。

空白

**BCCWJ** 00075\_A\_OY07\_00548



**BCCWJ** 00030\_B\_PM11\_00055



2.2.5 接続詞

**BCCWJ 2 次チェック**

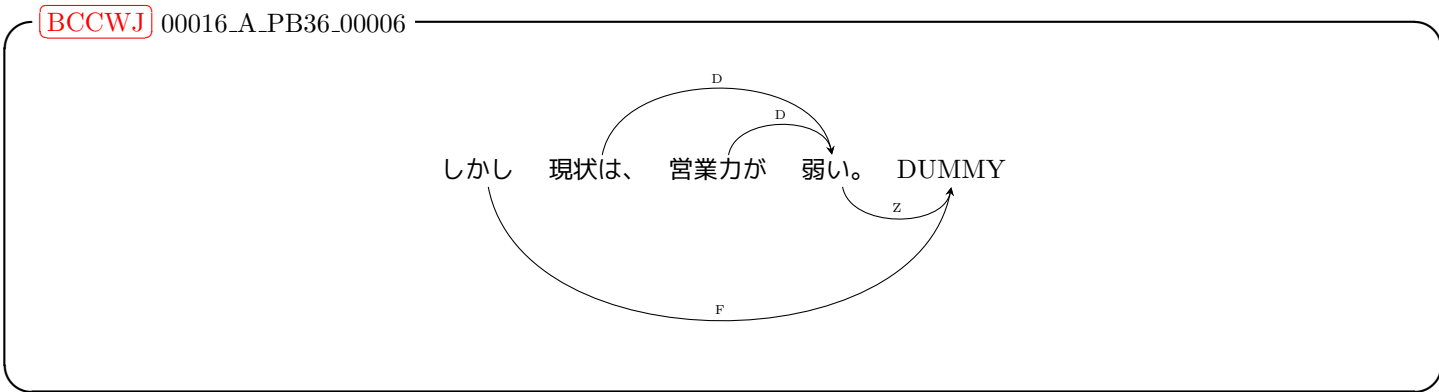
**CSJ** においては、接続詞相当句はかかり先なしとしてラベル“C”を定義している。**CSJ** のかかり受けアノテーション基準に準じて、文頭にある文接続詞についてはかかり先なしとしラベル“F”を付与する。

接続詞

尚、自動変換プログラムにより、文頭に出現する接続詞のみを含む文節のかかり受けラベルは“F”が付与され、かかり先を DUMMY に変更する (484 事例)。それ以外の接続詞のみを含む文節のかかり受けラベルは“FX”が付与される (1227 事例)。

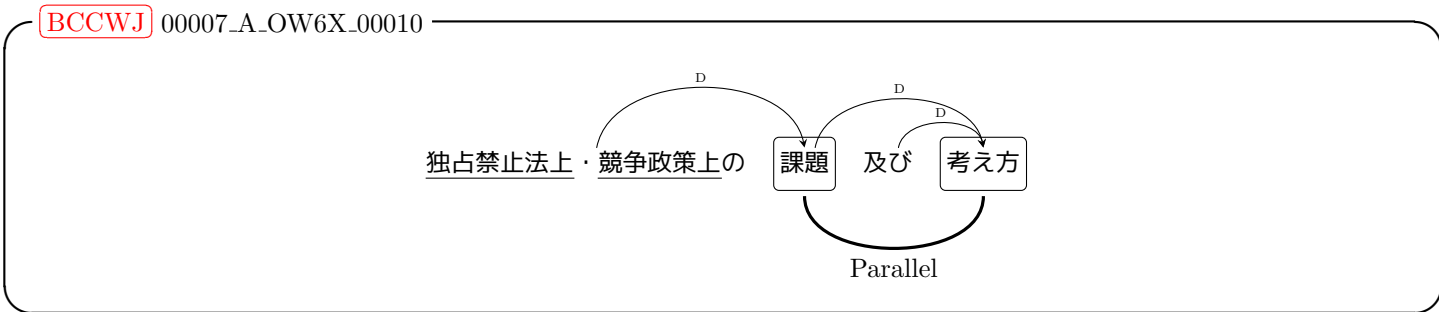
2 形態素に分かれている「だ—から」「例え—ば」なども接続詞に準じて扱う。

奈良先端大においては文接続詞を文末要素にかけると、現在のところ文もしくは節の単位が不定であるために本規定では現時点ではそれを採用しない。今後、節相当のスコープ（左端を含む）を定義する作業において、精緻化を行う予定である。

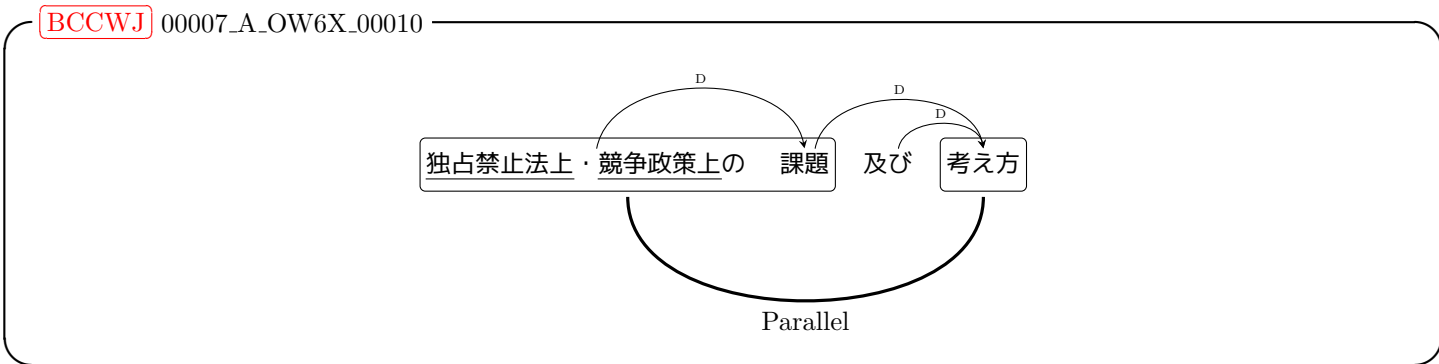


但し、並列句の間に出現する接続詞など、明示的に接続詞のスコープ（有効範囲）がわかる場合には、右隣接する並列句の最右文節にかける（2.5.5 参照）。

以下の例では、接続詞「及び」は右隣接する「考え方」にかかる。尚、この例では第 1 文節中に、「独占禁止法上」と「競争政策上」の並列構造を含むとともに、この文節が、「課題」と「考え方」の両方にかかると解釈する。



仮に「独占禁止法上・競争政策上の」の文節が「課題」にしかかからないとするならば、次のようにアノテーションを行う。

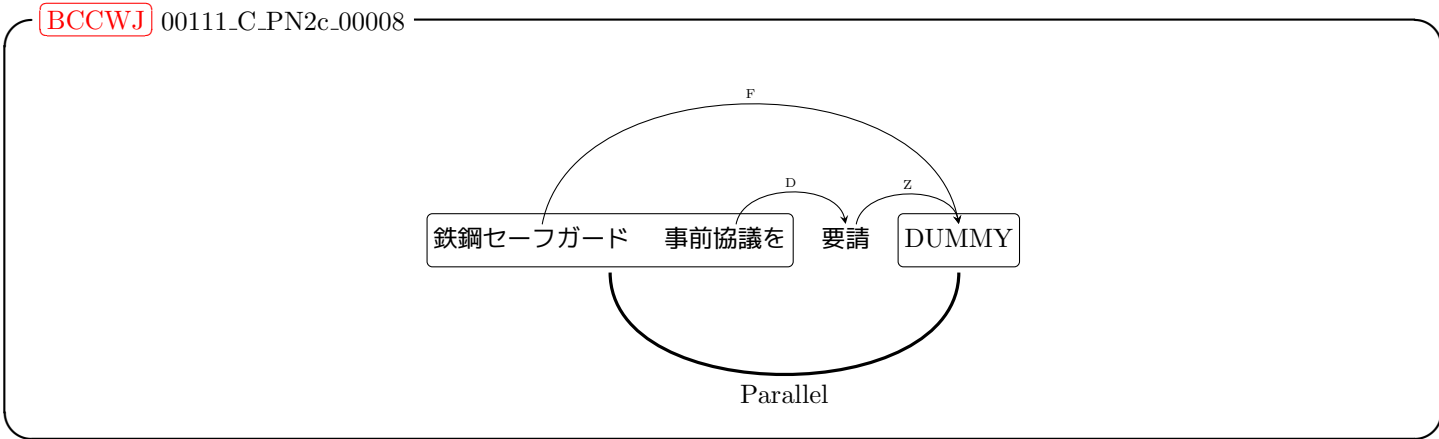


2.2.6 その他

BCCWJ 2 次チェック

BCCWJ では、冒頭のリーダ、ラジオ番組の報道時間、新聞のスポーツ種別など、「簡条書きの記号」に相当するものは“F”でかかり先なしとする。

簡条書き



CSJ では、「呼びかけ」がかかり先なしとして認定されているが、BCCWJ においても相当する表現があった場合にはラベル“F”でかかり先なしとしてアノテーションする。

2.3 英単語・ローマ字文・漢文（・古文）

BCCWJ 2 次チェック

英単語などは範囲をセグメントで指定する。「セグメント」名は“Foreign”とする。最右要素を除いた範囲内部の要素はラベル“D”で右隣接要素にかける。

尚、この処理は自動で行う (英単語 61 事例、ローマ字文 7 事例、漢文 2 事例)。最右要素を除くセグメント範囲内の要素のかかり受け関係の精緻化は、本作業の範囲外とするが、最右要素のセグメント範囲の外との関係は確認すること。

この英単語・ローマ字文・漢文については文節を越えるもののみをセグメントを付与すること。

英単語

ローマ字文

漢文

古文

BCCWJ00004\_A\_PB22\_00002

第7章

SENJU KA IKOIN

DUMMY

Foreign

D

Z

F

BCCWJ00329\_A\_OC03\_00651

jibundesagase / / / / / / / /

Foreign

注意：上の例は自動で付与しています。  
以下の漢文の事例ではト格の要素になるので通常のかかり先として認定する。この場合、セグメントでの範囲指定は不要。

BCCWJ00034\_B\_PB39\_00013

それが「山如釜形故名、其下即釜山浦也」とうたわれ釜山になつたという。

D

D

D

D

D

2.4 言い直し・言いよどみ

BCCWJ 2次チェック

BCCWJにおいて「言いよどみ」については、言い直し要素にかけてラベル“D”を付与するとともに、範囲をセグメントで指定する。セグメント名は“Disfluency”とする

BCCWJ00329\_A\_OC03\_00651

ど、泥棒だわー！DUMMY

Disfluency

D

Z

言い直し  
言いよどみ

2.5 並列構造

並列構造・同格構造については、国語研短単位に対して対応する範囲を付与する。以下、さまざまな並列構造についてのアノテーション基準を示す。

2.5.1 並列構造範囲アノテーション

BCCWJ のアノテーション基準の特色として、並列構造の範囲と対応する並列句を、かかり受け木とは独立に範囲を付与する点がある。以下の例では、かかり受け関係ラベルを全て“D”としたうえで、「科学技術の向上」と「国民経済の発展」が対応する並列構造として、セグメント Parallel で切り出され、グループ化される。文節単位ではなく短単位に付与するが、並列構造範囲が文節内で閉じている場合には対象としない。

並列構造

BCCWJ00006\_A\_OW6X\_00009

科学技術の向上と国民経済の発展に資することを

D

D

D

D

D

Parallel

2.5.2 名詞句の並列

名詞句については、対応する名詞句をセグメント Parallel で切り出し、グループ化する。かかり受け関係は通常のかかり受けと同じラベル“D”を付与する。



BCCWJ

Parallel

尚、CSJKCにおいては、ラベル “P” によりアノテーションを行う。

CSJKC

P

2.5.3 述語並列

CSJKCでは一部の述語並列について、並列構造を認定しラベル “P” を付与しているが、BCCWJにおいては、全ての述語並列を並列とみなさず、通常のかかり受けとして定義する。

並列構造：述語並列

BCCWJ

Parallel

CSJKC

P

また、判定詞を伴う名詞述語や形状詞述語が、名詞などと並列構造をなし、判定詞を共有する場合には並列構造範囲を付与する。

並列構造：述語並列の例外：判定詞

2.5.4 部分並列内の関係

CSJKCでは以下のような構造について、非交差制約を順守するためにラベル “T” を付与し、真のかかり先でないものにかけている。BCCWJにおいては、範囲を規定したうえで、通常のかかり受け関係として真のかかり先にかける。

並列構造：部分並列

BCCWJ

Parallel

KCCSJ

P

2.5.5 並列構造と接続詞

BCCWJ 2次チェック

並列句の間に接続詞が出現する場合には、右隣接する並列句の最右文節にかける。

並列構造：接続詞

BCCWJ 00144\_C.PN2c.00010

Parallel

接続詞

2.5.6 3 つ以上の並列構造

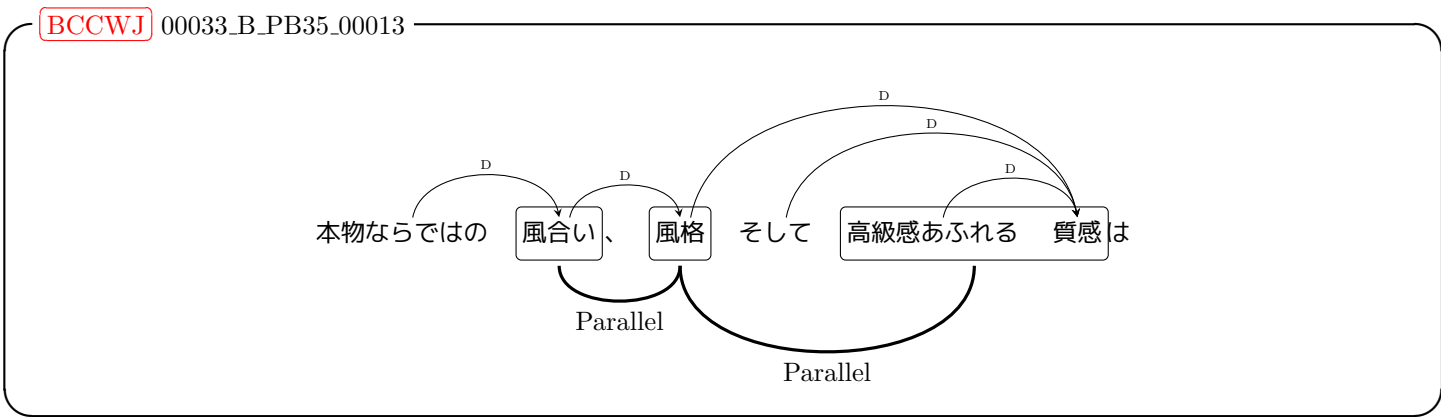
3 つ以上の並列構造の場合、隣接する並列句にかける。

並列構造：3 並列

接続詞

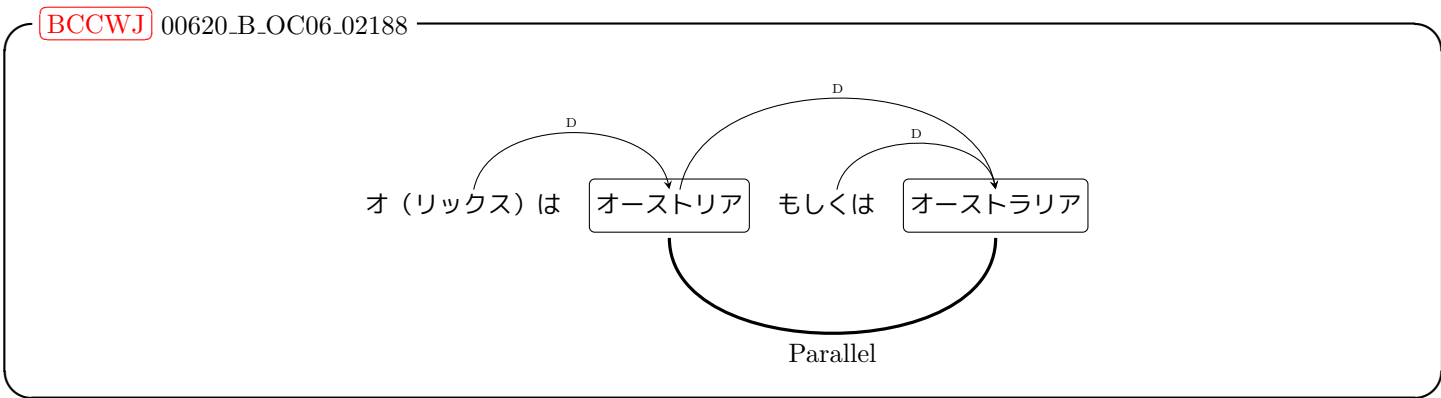


3つのセグメントによって並列句のグループ化を行う。並列句の間に接続詞が出現する場合には、右隣接する並列句の最右文節にかける。以下の事例では、「風合い」「風格」「高級感あふれる質感」が並列構造をなす。

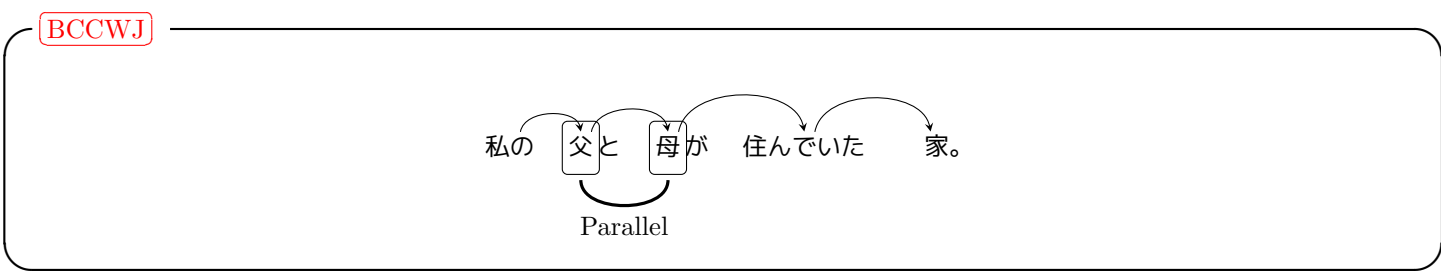


2.5.7 並列構造の複数の要素に左からかかる場合

以下のように「オ（リックス）は」は「オーストリア」と「オーストラリア」の両方にかかる場合には、並列構造範囲が付与されている。最左要素である「オーストリア」にかけることにより、両方にかかっていることを表現する。項を共有するなど、部分文字列を共有するようなかかり受け関係を対象とする。



並列構造：複数の要素が左からかかる場合



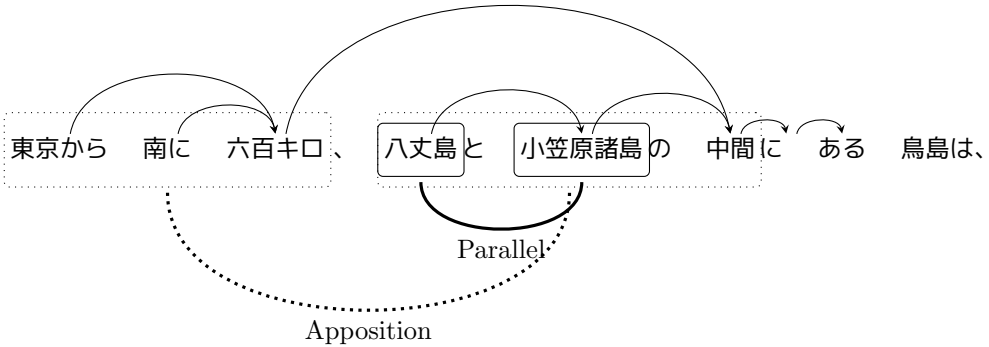
BCCWJ 2次チェック

2.6 同格関係

同格構造については、国語研短単位に対して対応する範囲を付与する。

2.6.1 同格構造範囲アノテーション

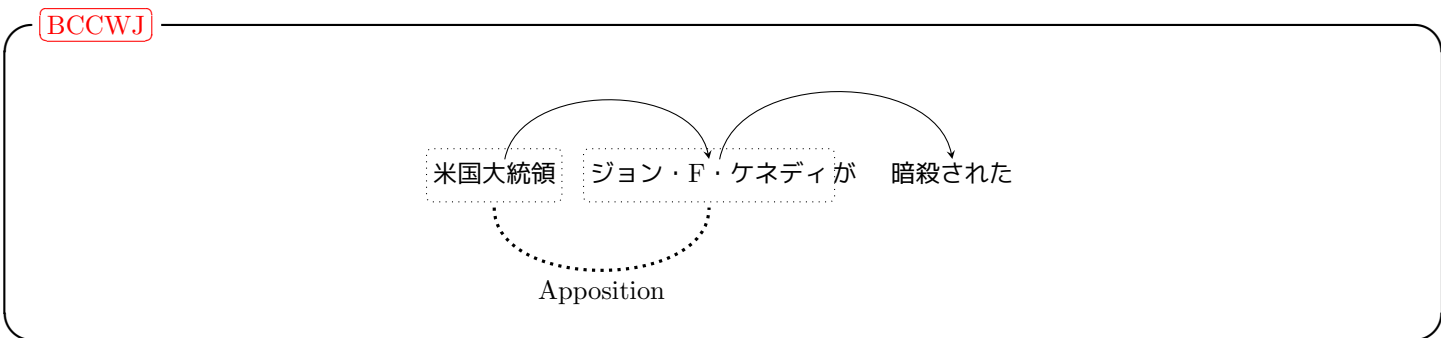
BCCWJ のアノテーション基準の特色として、同格構造の範囲と対応する並列句を、かかり受け木とは独立に範囲を付与する点がある。

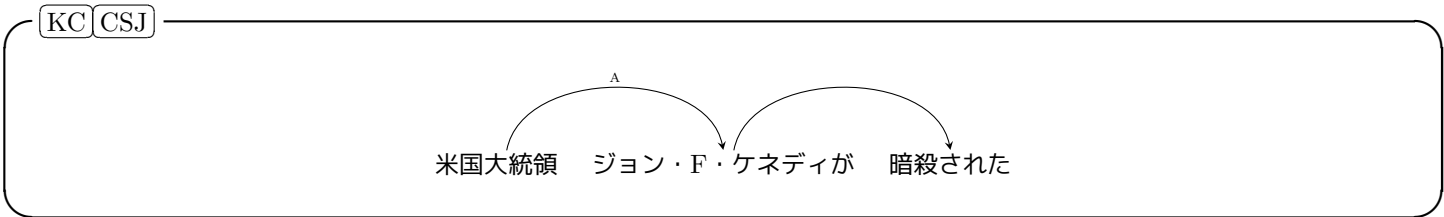


2.6.2 同格関係

通常同格関係は、対応する名詞句をセグメント Apposition で切り出し、グループ化する。かかり受け関係は通常のかかり受けと同じラベル “D” を付与する。

同格





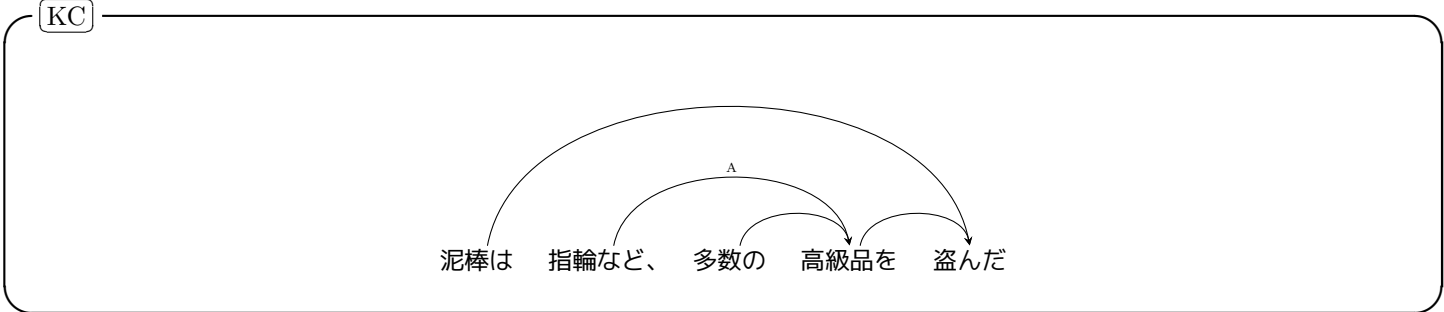
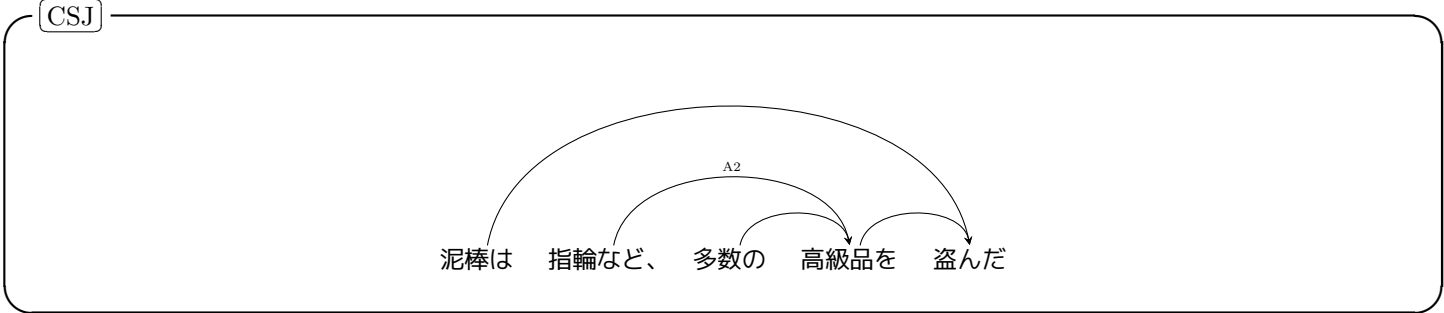
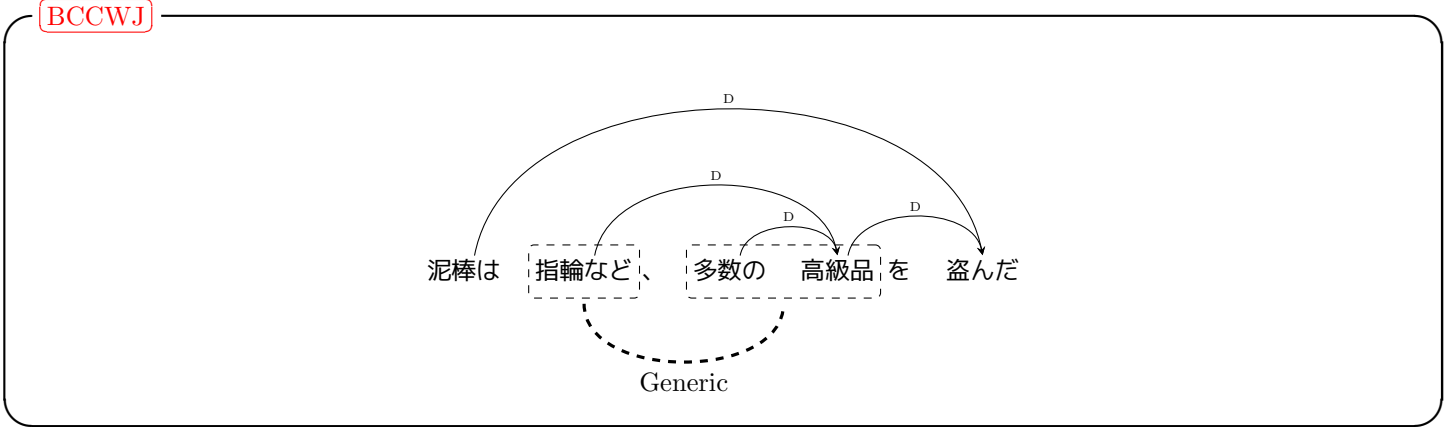
2.6.3 具体例と総称の同格関係・具体例と数詞の同格関係

[BCCWJ 2 次チェック](#)

具体例と総称の同格関係、具体例と数詞の同格関係については、対応する名詞句をセグメント Generic で切り出し、グループ化する。かかり受け関係は通常のかかり受けと同じラベル “D” を付与する。

同格:具体例と総称

同格:具体例と数詞

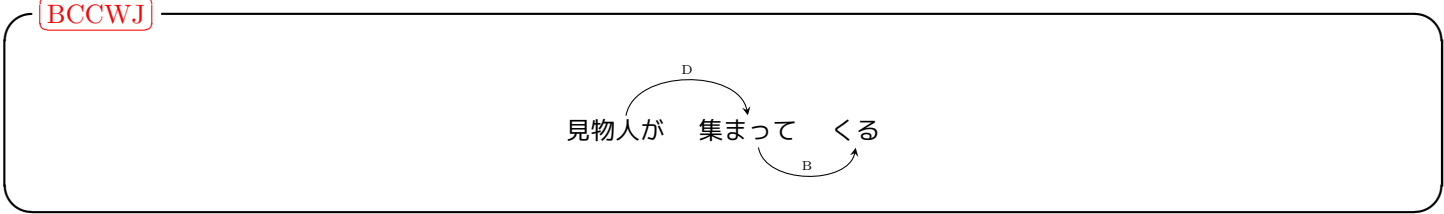


2.7 文節の修正

[BCCWJ 2 次チェック](#)

国語研文節単位は形態論情報に基づいて規定されており、かかり受けとしては 1 つの単位として扱いたい語彙的な複合動詞が 2 文節に分割している場合がある。そのような場合、ラベル “B” を用いて、左要素を右要素に連結することを許す。

文節境界修正



- 文節は連結する方法にのみ修正する。現在のところ以下のものを連結する。
- 右要素が「動詞-非自立可能」相当のもので左要素と連結して、語彙的な複合動詞をなすもの。
  - 「どう || いう」は連結する。「そう || いう」は連結する。
  - 「そう || いえば」は連結し、接続詞に準じてアノテーションする。
  - 「ため」は、事前の処理の長単位認定で、単体で 1 文節になる場合とそうでない場合が弁別されているために、その定義を尊重する。
  - 「と || し」「と || する」は、事前の処理の長単位認定で、単体で 1 文節になる場合とそうでない場合が弁別されているために、その定義を尊重する。
- 例) 「中心と || する」など

詳しくは Web 上のメモに示す接続事例を参照すること。

また小説などのタイトルで「かかり受けの判定が困難な場合」には文節結合すること。タイトルでもかかり受けが明示的に定義できるものについては、かかり受けを付与すること。

2.8 文境界の修正

[BCCWJ 2 次チェック](#)

BCCWJ DVD の元データは、文単位の定義として、文の入れ子を許している。文書構造（レイアウト）に基づいて、一番外側の文について superSentence タグが付与されている。本来文の構造としては superSentence タグが付与されるべきものであって、文書構造中改行がある場合など superSentence タグが付与されていない場合、かかり先のない文節が、となりの文に出現する場合があります。このようなことのないように、BCCWJ かかり受けアノテーションにおいては、前処理で、文書構造（レイアウト）

文境界修正

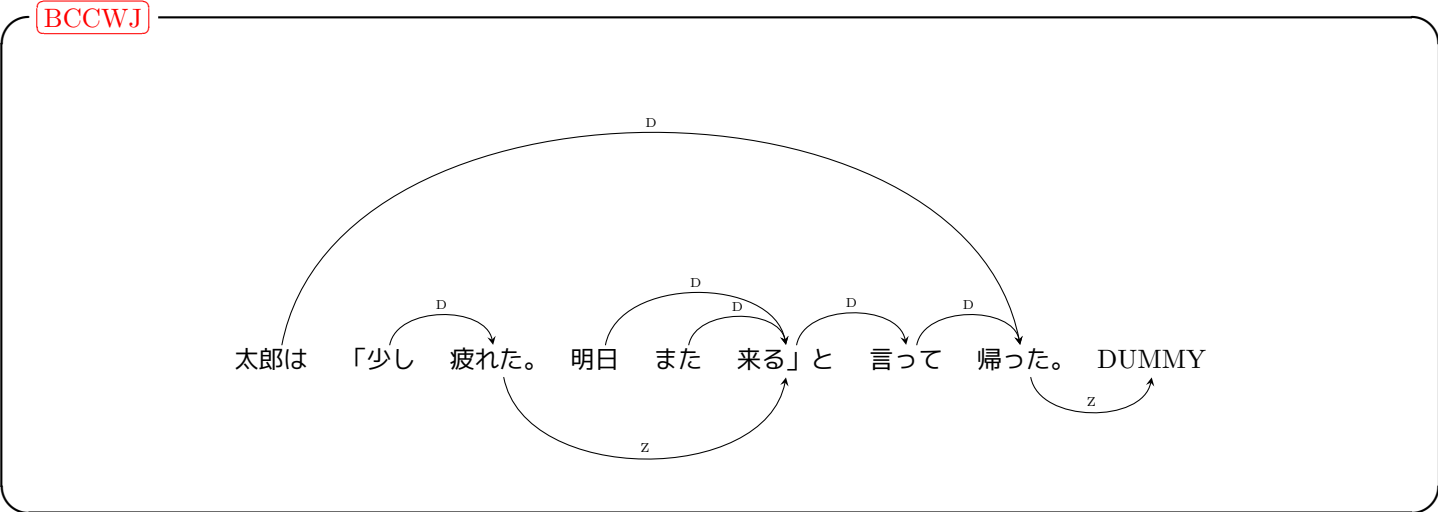
を考慮せずに、新たに superSentence 相当情報を付与する。

このようにすることで、文内に文境界相当の文節端が出現する場合がある。入れ子でない文境界もありうる。例えば、「呼びかけ」が文境界に相当する。そのような場合には、かかり先なしとし、ラベル “Z” を付与する。かかり先が不定な場合は DUMMY にかける。

この文境界の修正は、前作業として文献 [?] に記述されている文の最大範囲が規定されていることを想定している。しかし、既存の文境界がかかり受けに影響を与える場合には連絡して、文連結作業を実施する。例えば、箇条書きが文境界をまたいでいるなどの場合には文連結作業を行う。

奈良先端大における作業では、入れ子の文や引用のカギ括弧や注釈の丸括弧を Nest セグメントで処理しているが、BCCWJ のアノテーションにおいては、従前の Nest セグメントを排除したうえで、元のレイアウトに基づく、superSentence タグ、sentence タグ (type “quasi” を含む) を尊重したうえで、かかり受け関係に影響を及ぼす部分のみ「連結」作業を行った。

本作業においては、「連結」したもののについての「分割」作業のみを行う。「分割」作業により、1 行中にラベル “Z” が複数出現することもありうる。左範囲規定は、節のスコープ認定作業とともに行う予定である。



2.9 「格表示誤り」などの取り扱い

格表示誤り

以下の作業は不要

アノテーションツール ChaKi.NET のかかり受けに対するアノテーション機能を用いる。Dependency Panel でかかり受け関係を選択すると Attribute Panel 上にかかり受け関係の属性が表示される（未実装：今後対応される予定）。

2.10 前処理（自動処理）一覧

前処理

自動処理

- 今回の作業にあたり、従前基準から半自動的に変換できる部分についてはある程度自動処理による前処理を行った。
- 文末（元ファイルにおいて </sentence> 相当が後置するもの）相当の位置にある文節はかかり受けラベル “Z” を付与し、DUMMY ノードにかけている。文が入れ子になっている場合には「と」などで節（文）を受ける文節にかける。かかり受け構造を確認すること。
  - 感動詞を含む文節
    - － 感動詞のみを含む文節はかかり受けラベル “F” を付与し、かかり先を DUMMY ノードにかけている。
    - － しかし、「ありがとう || ございます」の「ありがとう」のような事例については通常の “D” で隣接要素「ございます」にかけている。これに相当する事例があった場合には修正すること。
    - － 上記以外の感動詞を 1 つ以上含む文節についてはかかり受けラベル “FX” を付与している。かかり先が存在する場合（助詞などを伴い格要素になるなど）には修正すること。
  - 補助記号-一般のみを含む文節はかかり受けラベル “F” を付与し、かかり先を DUMMY ノードにかけている。
  - 補助記号-括弧開のみを含む文節は、文末に出現しなければ、かかり受けラベル “B” を付与し、かかり先を右隣接要素にかけている。文末に括弧開が出現した事例があった場合には報告すること。
  - 補助記号-括弧開のみを含む文節は、文末に出現しなければ、かかり受けラベル “DX” を付与している。10 事例出現するが、判断が確定していないために無視してよい。
  - URL のみを含む文節はかかり受けラベル “F” を付与し、かかり先を DUMMY ノードにかけている。
  - 空白のみを含む文節はかかり受けラベル “F” を付与し、かかり先を DUMMY ノードにかけている。空白を削除すると一つの文節をなすような場合には、隣接要素同士を “B” で結合すること。
  - 接続詞を含む文節
    - － 接続詞のみを含む文節はかかり受けラベル “FX” を付与している。文接続詞相当は “F” ラベルに変更し、DUMMY にかける。明示的にかかり先がわかる場合には “D” に変更して正しいかかり先にかける。
    - － 接続詞のみを含む文節のうち文頭に出現するものは文接続詞とみなし、自動で “F” ラベルに変更し、DUMMY にかけている。明示的にかかり先がわかる場合には “D” に変更して正しいかかり先にかける。
  - 英単語・ローマ字文・漢文は “Foreign” セグメントで切り出している。内部のかかり受け構造は最右要素以外は右隣接要素に “D” で自動でかけている。
  - 言いよどみは “Disfluency” セグメントで切り出しているが、言い直し相当句に “D” でかけること。



第3章

BCCWJにおけるアノテーション基準：各論

本節では[BCCWJ]アノテーション基準において質問があった場合に、どのようにアノテーションするのかを規定する。

3.1 従属節とかかり受け

6.1.1 節で言及している格要素と複数の述語の関係について、従属節の観点から、より詳細に規定する。尚、等位接続は 6.1.1 節冒頭の定義にならうこと。

南の従属節分類と内部に現れる要素（文献 [?] 第14表 p.128-129）3.1 にならい、述語的部分以外の要素が構成要素になりうるか否かを判断する。基本的にこの表は「は」「が」が内側に来ないかどうかの判断に用いる こと。この表において内側に来ることができる要素であっても、外側（より右側）の要素にかかることが正しいことがありうる。

従属節の分類でA類に分類されるものについては主語のガ格、提示のハ（モ）格を従属節ではない主節側にかける。それ以外の格要素（ヲ格、二格）は従属節側にかける。A類の従属節の内側に認定するのは、状態副詞・程度副詞・A類従属節にとどめ、それ以外は従属節の外側にかける。

従属節の分類でB類に分類されるものについては提示のハ（モ）格を従属節ではない主節側にかける。主語のガ格を含む格要素は従属節側にかける。B類の従属節の内側に認定するのは、状態副詞・程度副詞・A類従属節・時の修飾語・場所の修飾語・ジツニの類・評価的意味の修飾語・B類修飾節にとどめ、それ以外は従属節の外側にかける。

従属節の分類でC類に分類されるものについては、提示のハ（モ）格、オソラクの類も含めて従属節側にかける。

連用形については複数の分類に含まれているので注意すること。A類連用形反復は「酒を飲み飲み考えた」「後ろを振り返り振り返り立ち去った」のような例である。A類連用形（形容詞・形容動詞）は「足音も高く出て行った」「意志堅固にやりとおした」のような例である。B類連用形とC類連用形は表中の識別では困難であるが、継起を表現する（～ト、～タラなどに近い）ものについてはB類とし、並列を表現する（～シ、～ケレドなどに近い）ものについてはC類とする。今後適切な言語テストを設定する。

テ形については複数の分類に含まれているので注意すること。A類テ形は従属節内の述語が打消の形および丁寧の形を取りうるかどうかの言語テストを行い、取りえない場合にA類テ形として判定する。B類テ形（2種）とC類テ形は表中の識別では困難であるが、継起を表現する（～ト、～タラなどに近い）ものについてはB類とし、並列を表現する（～シ、～ケレドなどに近い）ものについてはC類とする。理由を表現するテ形（～タメに近い）についてはB類とする。今後適切な言語テストを設定する。

そのほか表 3.2 野田 [?] の分類も参考にすること。これらの分類に基づき、主節と従属節の主格が同じか否かを判定しながらアノテーションを行うことが必要である。

表 3.1 南 (1974) の従属節分類と内部に現れる要素 (p128-129)

従属節の種類		A 類				B 類											C 類								
構成要素		～ナカラ (継続・付帯状況)	～ツツ (付帯状況)	～ナニテ (継続句・付帯状況)	ナニテ連用形反復	ナニテ連用形 (形容詞・形容動詞)	～ナニテ	～ト (継続節・仮定節)	～ナカニテ (逆接)	～ナニデ (理由: 焦点)	～ニ (理由: 焦点)	～バ (仮定)	～タラ (仮定)	～ナラ (仮定)	～テモ (仮定)	ナニテ (継続節)	ナニテ連用形 (継続節)	～ス (ズニ)	～ナイテ	～ガ	～カニテ (理由)	～ケレド (並列)	～シ (並列)	ナニテ (引用)	ナニテ連用形 (並列)
述語の部分以外の節内要素	名詞 + 格助詞 (ガ以外)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	状態副詞	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	程度副詞	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	A 類従属節	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	主語 (～ガなど)	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	時の修飾語	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	場所の修飾語	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	ジツニ、トニカク、ヤハリの類	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	評価的意味の修飾語	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	B 類従属節	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
述語の部分の要素	提示のことば (～ハなど)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
	オソラク、タブン、マサカの類	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
	C 類従属節	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
	用言	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	使役形	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	受身形	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	受給の形	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	尊敬の形	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	丁寧の形	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	打消の形	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
過去形	-	-	-	-	-	(-)	-	-	+	+	+	+	+	+	+	(-)	-	(+)	(+)	+	+	+	+	+	
用言 + 形式名詞	-	-	-	-	-	-	-	-	+	+	-	-	+	-	-	-	-	-	-	+	+	+	+	-	
意志形	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	
推量形	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	

+ はある要素がその節の中に存在可能であることを示す  
- はある要素がその節の中に存在不可能であることを示す  
( ) 内のものは、そう認めることに問題があるものを示す  
† テ形は述語的部分の要素に対して言語テストを行い 4 つに分類すること  
‡ 連用形は述語的部分の要素に対して言語テストを行い 4 つに分類すること

従属節  
連用修飾  
主題  
主語  
主題・主語以外の格要素  
テ形



第 4 章

かかり受け関係ラベルの比較～ CSJ のマニュアルに基づく比較 (1)

本節では、かかり受け関係ラベルを比較することで、文献 [?] に出現する事例に基づいて、BCCWJCSJKC の基準を比較する。

4.1 かかり受け関係ラベルのまとめ

次の表は各コーパスのかかり受け関係ラベルの違いを示す。

かかり受け関係のラベル	<span>BCCWJ</span>	(グループ)	<span>CSJ</span>	<span>KC</span>
通常のかかり受け	D	-	ラベルなし	D
並列	D	(Parallel)	P	P
部分並列	D	(Parallel)	I	I
同格	D	(Apposition)	A	A
同格（総称、数詞）	D	(Generic)	A2	A
言い直し・言いよどみ	D	-	D	未定義
倒置	D	-	R	未定義
倒置によりかかり先が不定になるもの	F	-	未定義	未定義
文節境界に関するラベル	<span>BCCWJ</span>	-	<span>CSJ</span>	<span>KC</span>
後続文節と接続	B	-	B+	未定義
その他	<span>BCCWJ</span>	(セグメント)	<span>CSJ</span>	<span>KC</span>
フィラー	F	-	F	未定義
顔文字	F	-	未定義	未定義
接続詞	F or D		C	D
感動詞	F or D	-	E	D
呼びかけ	Z	-	Y	未定義
非言語音	F	-	ラベルなし	未定義
かかり先のない文節	F	-	N	未定義
記号・補助記号	F	-	未定義	未定義
URL・空白	F	-	未定義	未定義
かかり受け関係の交差	D	-	X	未定義 (A のみ)
英単語・ローマ字文・漢文	D	(Foreign)	未定義	未定義
古文	D	(Foreign)	K	未定義
古文	D	(Foreign)	K:S1	未定義
古文	D	(Foreign)	K:E1	未定義
文境界相当	Z	-	未定義	未定義
コメント	<span>BCCWJ</span>	(セグメント)	<span>CSJ</span>	<span>KC</span>
	未定義	-	S:格表示誤り「が」	未定義
	未定義	-	S:格表示誤り「を」	未定義
	未定義	-	S:格表示誤り「に」	未定義
	F	(Disfluency)	S:複数文節の言い直し	未定義
	F	(Disfluency)	S:複数文節の言い直し:S1	未定義
	F	(Disfluency)	S:複数文節の言い直し:E1	未定義

BCCWJ において “D”(通常のかかり受け関係), “B”(文節接続), “F”(かかり先なし), “Z”(文境界) の 4 つのラベルを定義する。その他作業用のラベルとして “DX”, “BX”, “FX”, “ZX” の 4 つのラベルを定義する。“X” が付与されているラベルは作業者から管理者に注意を促す用途に用いる。

CSJ KC と双方で、違う意味でラベル “D” が用いられている。かかり受けアノテーションの事例の中でCSJ KCを一度に表現する場合、本来 KC では ラベル “D” を使うかかり受けについては ラベル “D” を省略する

CSJ において、「S:複数文節の言い直し」「S:複数文節の言い直し:S1」「S:複数文節の言い直し:E1」については、ChaKi のセグメント機能を使うべきだと考える。

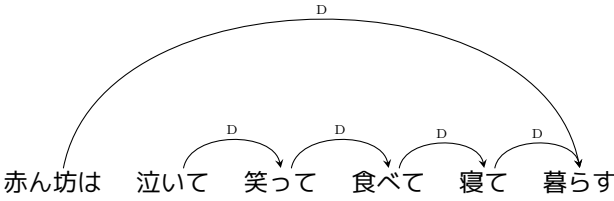
4.2 かかり受け構造、並列構造、同格構造

以下では文献 [?] の用例に基づいて、BCCWJCSJKC の差異について示す。

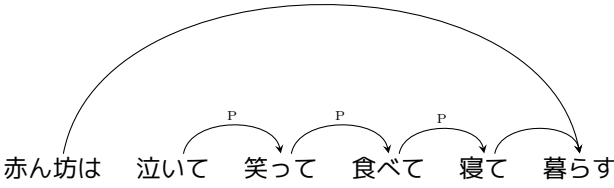
- 並列関係 (BCCWJ D (Parallel) CSJ KC P)

BCCWJ においては、並列関係はかかり受けラベル “D” と 並列構造セグメント “Parallel” で表現するが、以下のテ形の例では並列構造を認めない。

BCCWJ

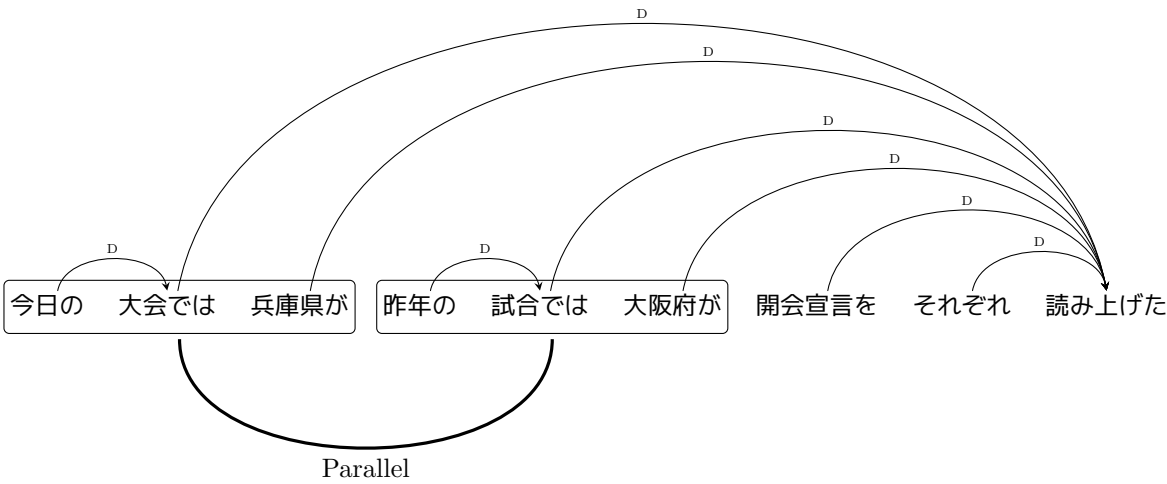


CSJ KC

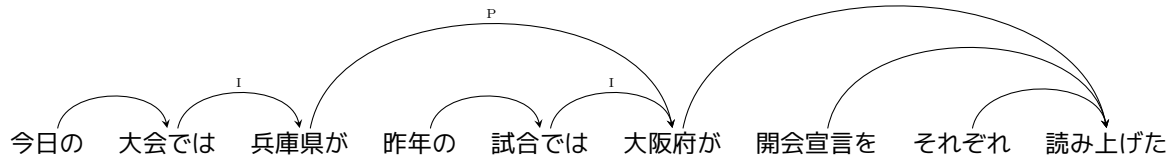


- 部分並列 (BCCWJ D (Parallel) CSJ KC I)

BCCWJ



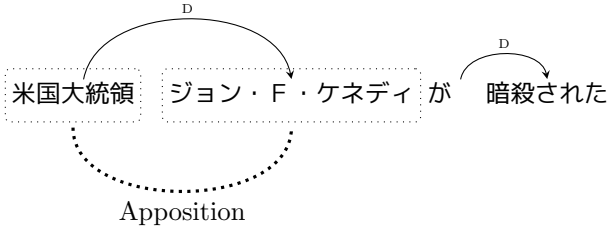
CSJ KC



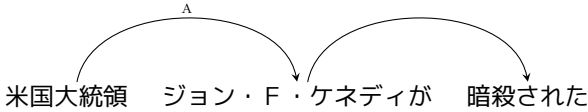
- 同格関係 (BCCWJ D (Apposition) CSJ KC A)

同格

BCCWJ



CSJ KC

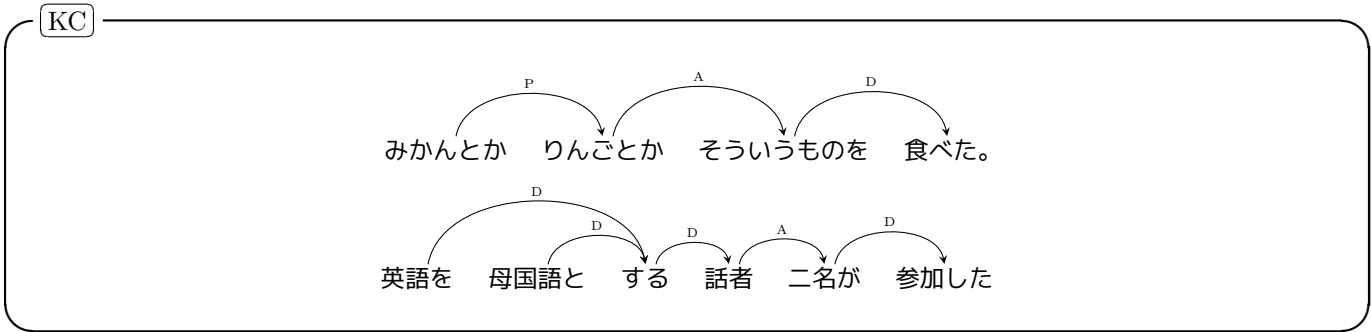
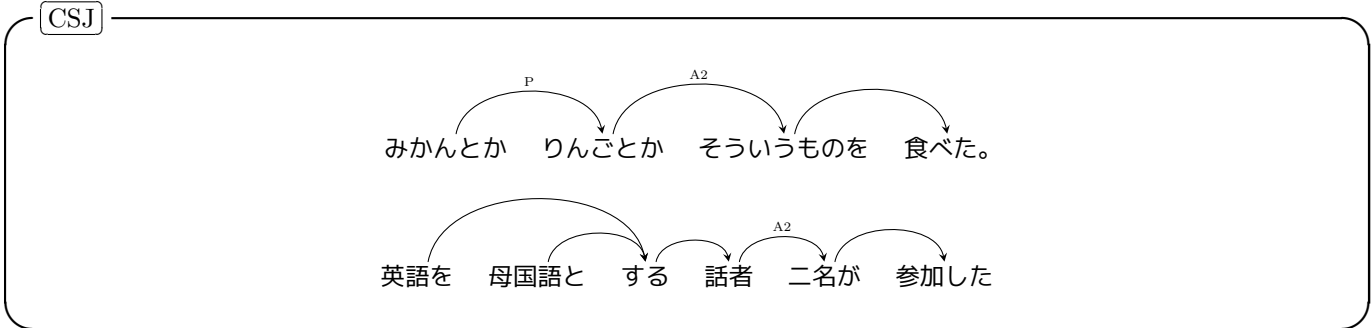
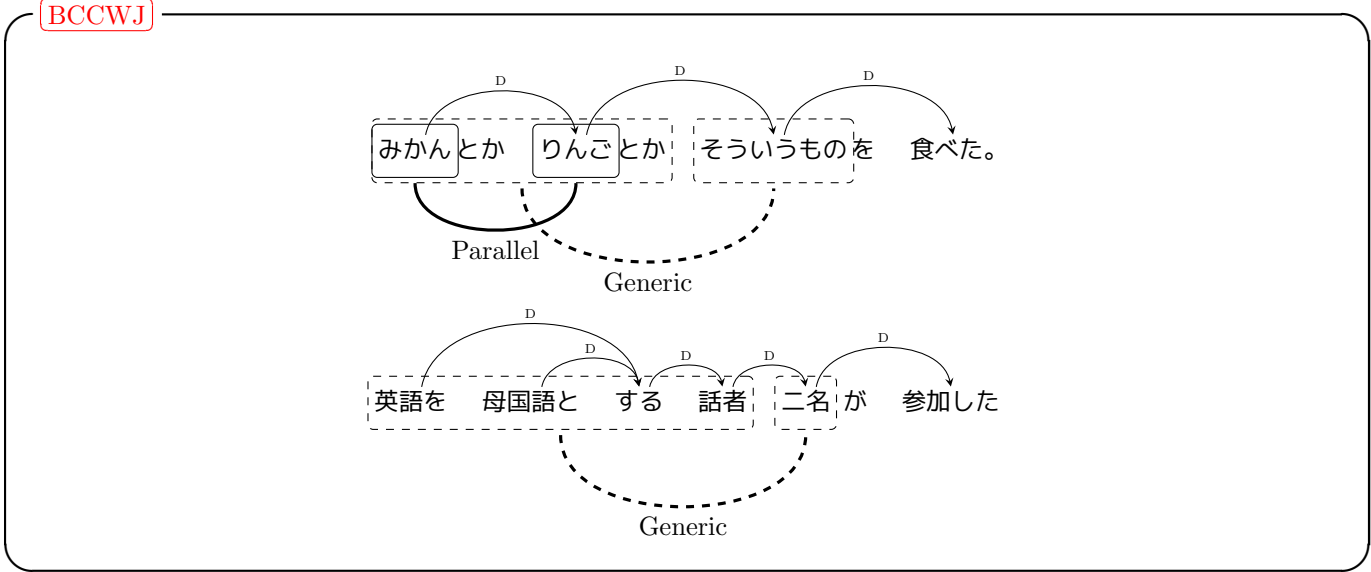


- 具体例と総称の同格関係・具体例と数詞の同格関係 (BCCWJ D (Generic) CSJ KC A)

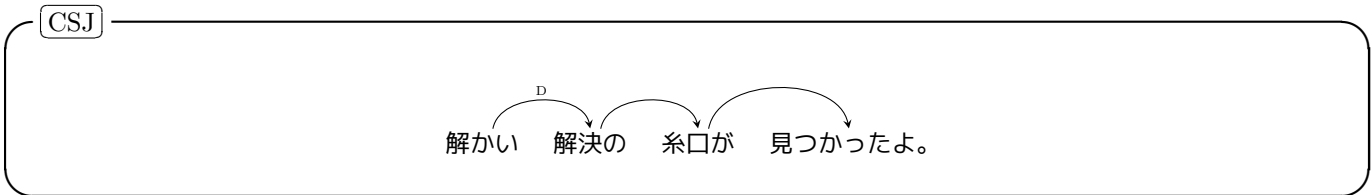
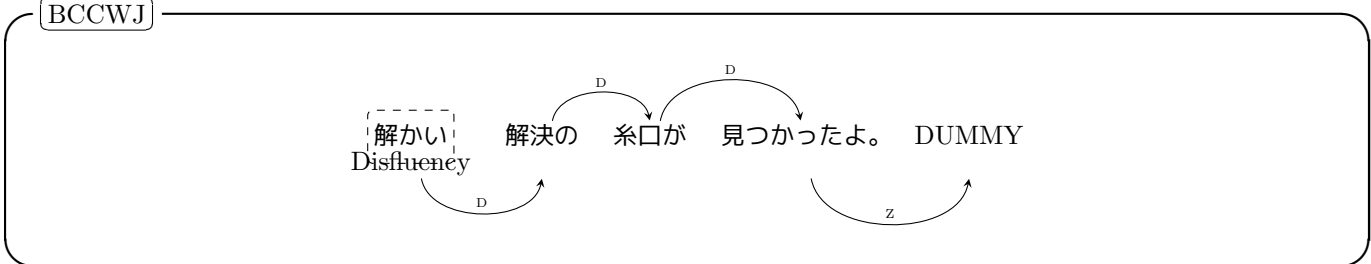
同格:具体例と総称

同格:具体例と数詞



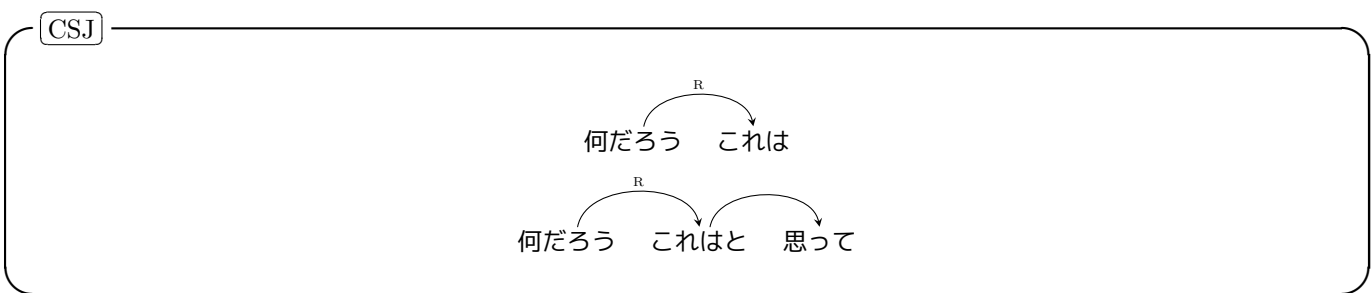
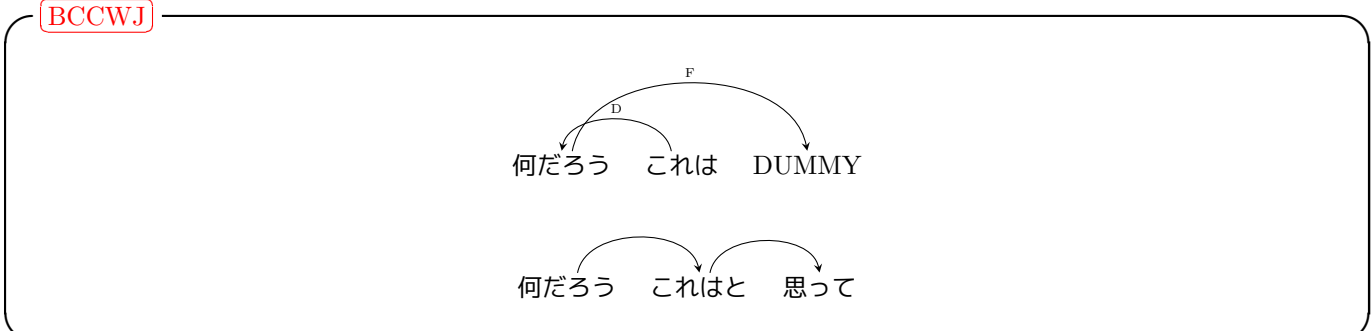


- 言い直し (BCCWJ) Disfluency (CSJ) D (KC) 未定義  
(CSJ) 特有の現象である言い直しは、(CSJ)においては“D”ラベルを付与する。(BCCWJ)においては Disfluency セグメントを付与し、かかり受けラベル“D”として、かかり先は言い直し要素とする。



- 倒置 (BCCWJ) D (CSJ) R (KC) 未定義  
(BCCWJ) において、後ろから前にかかるかかり受け関係を許す。倒置のラベルは“D”とするが、倒置によりかかり先が不定になった文節は“F”とする。  
(CSJ) において、倒置はラベル“R”により示す（が、かかり受け関係の矢印の向きで表現する方法に変更中）。

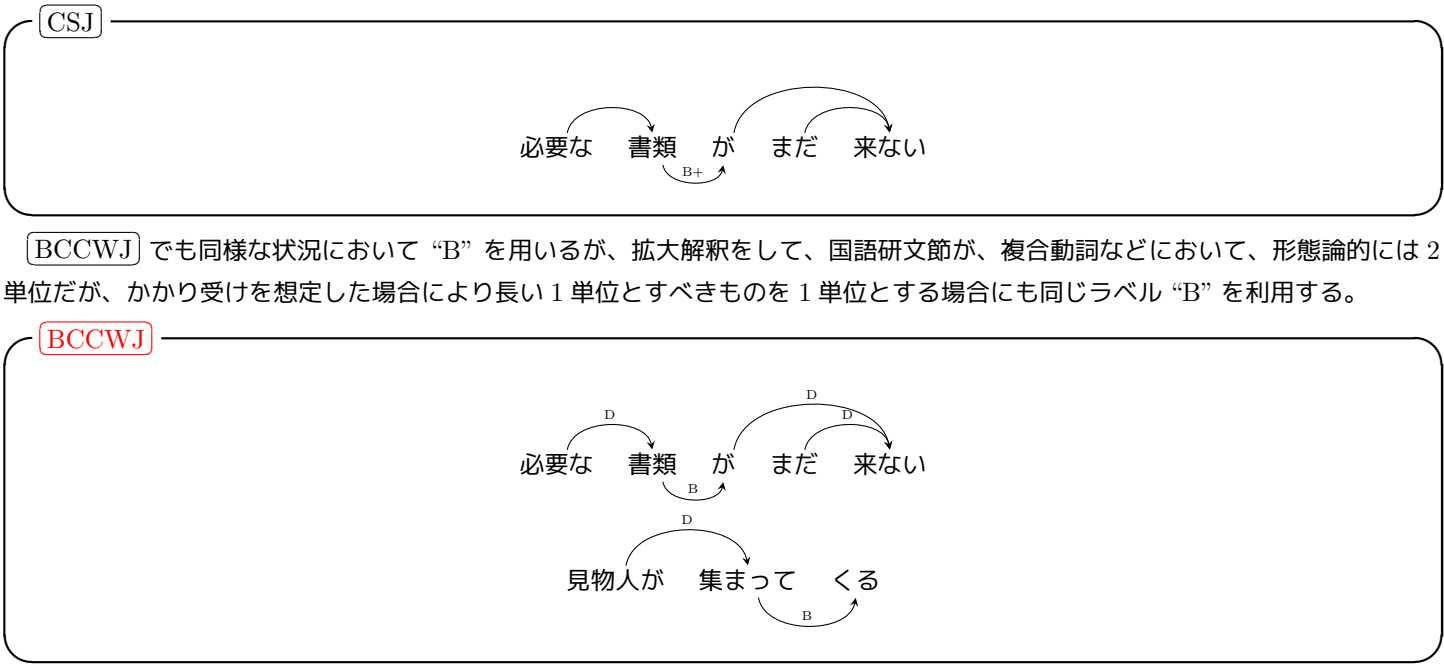
倒置



4.3 文節境界を示すラベル

文節境界修正

(CSJ) においては、転記の際に 1 文節中に転記基本単位の境界が存在することにより、1 つの文節が複数の単位に分割されることがある。それをつなぐ場合にラベル“B+”を用いる。



4.4 節境界を示すラベル

CSJ においては、49 種類の節境界ラベルが付与されている。BCCWJ においては、節境界ラベルはかかり受け・並列構造アノテーションとは別に付与するため、本文書では規定しない。

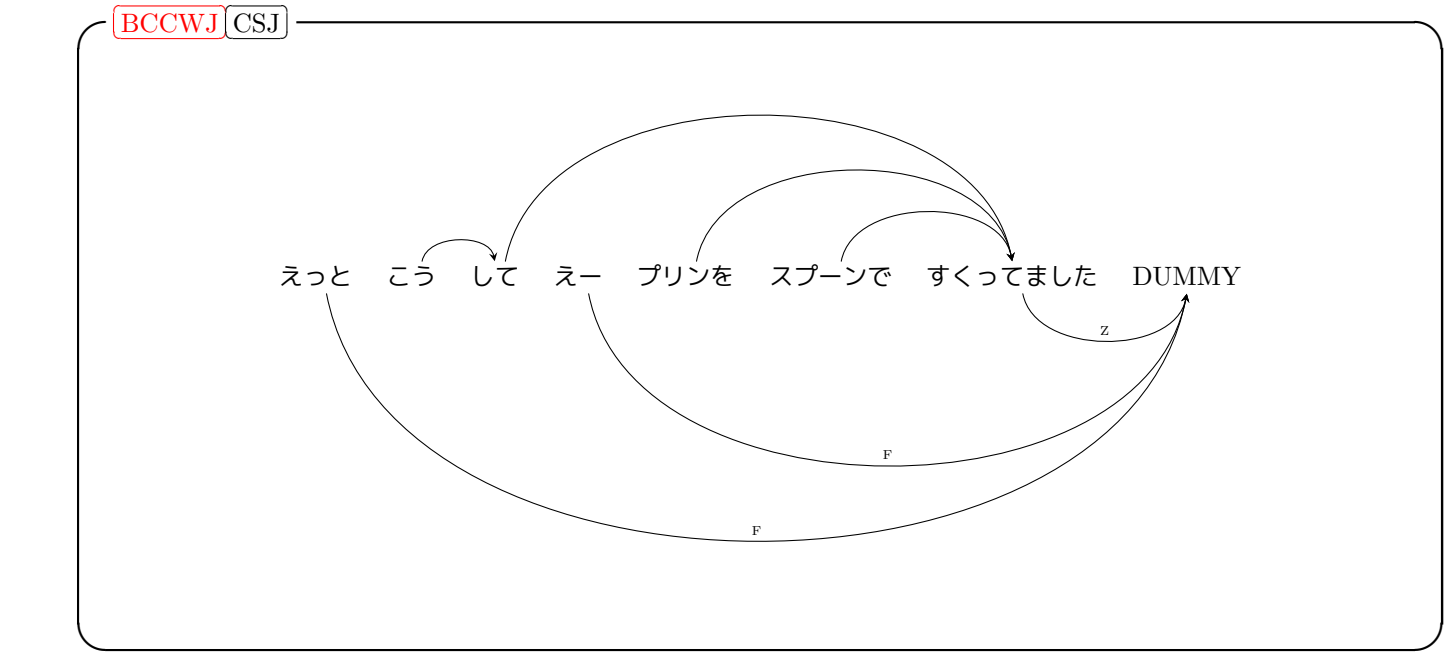
節境界

4.5 その他のラベル

1. フィラー

CSJ は、ラベル “F” を用い、フィラーのかかり先は定義しない。DUMMY にかけることによってかかり先なしを示す。BCCWJ では、同様に、ラベル “F” を用い、フィラーのかかり先は定義しない。DUMMY にかけることによってかかり先なしを示す。

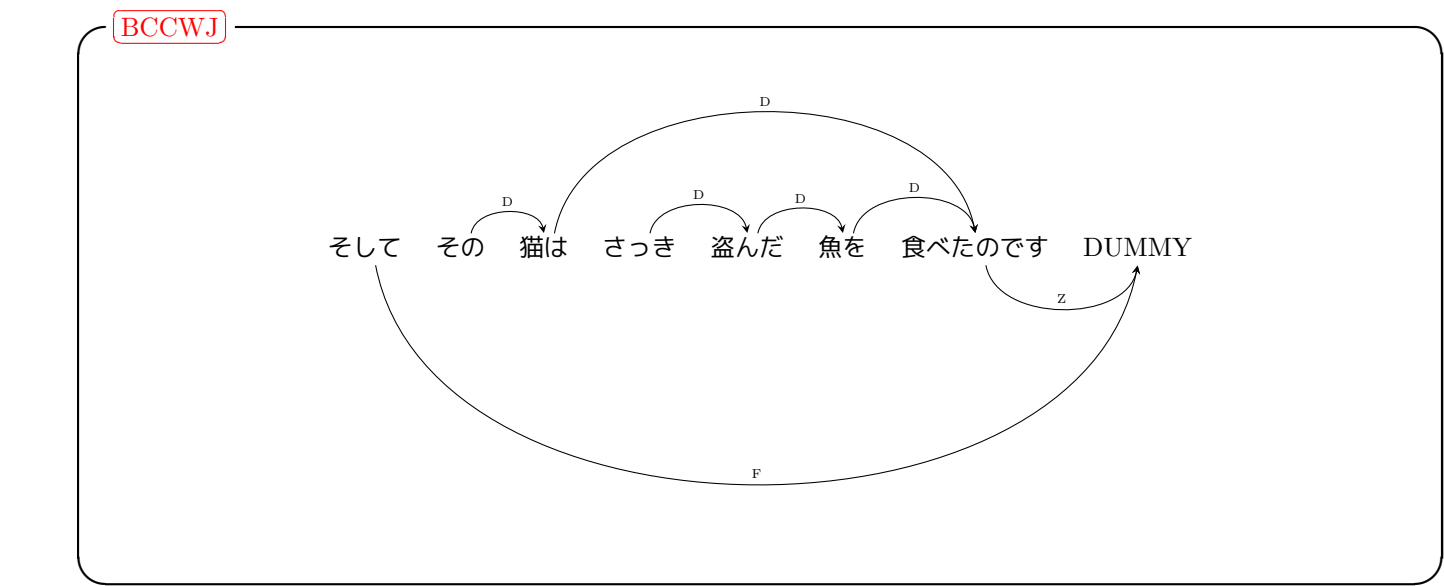
フィラー

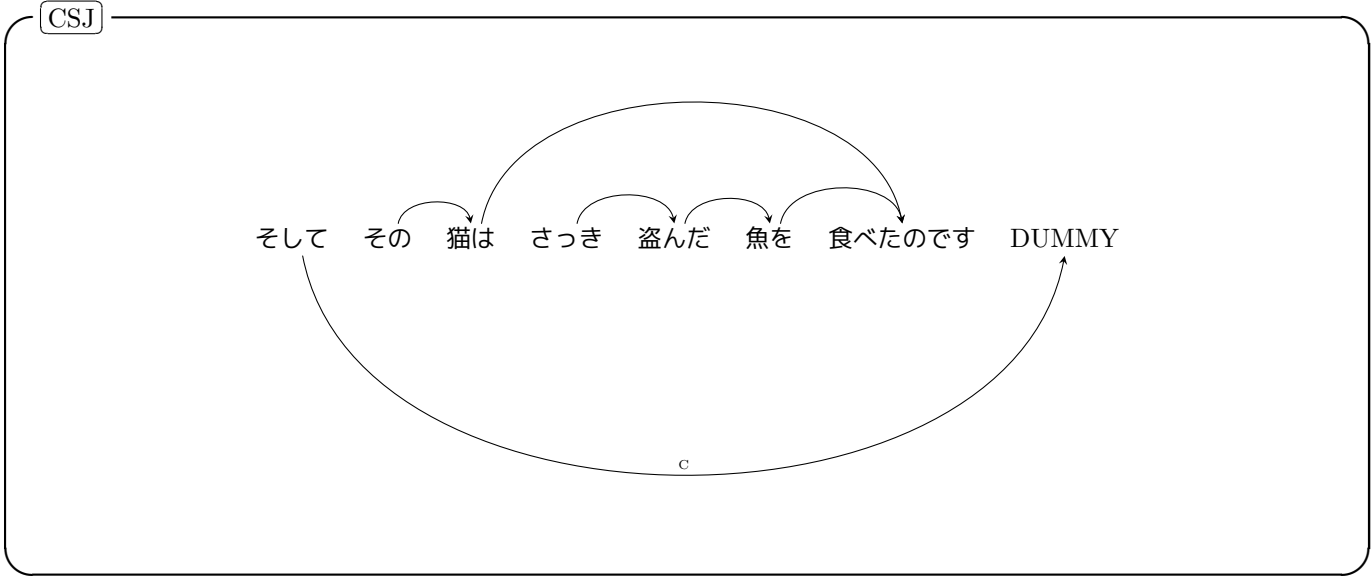


2. 接続詞

CSJ は、ラベル “C” を用い、接続詞のかかり先は定義しない。-1 位置の DUMMY にかけることによってかかり先なしを示す。BCCWJ では、文頭の接続詞でかかり先判定が難しい際にラベル “F” を用い、接続詞のかかり先は定義しない。-1 位置の DUMMY にかけることによってかかり先なしを示す。

接続詞



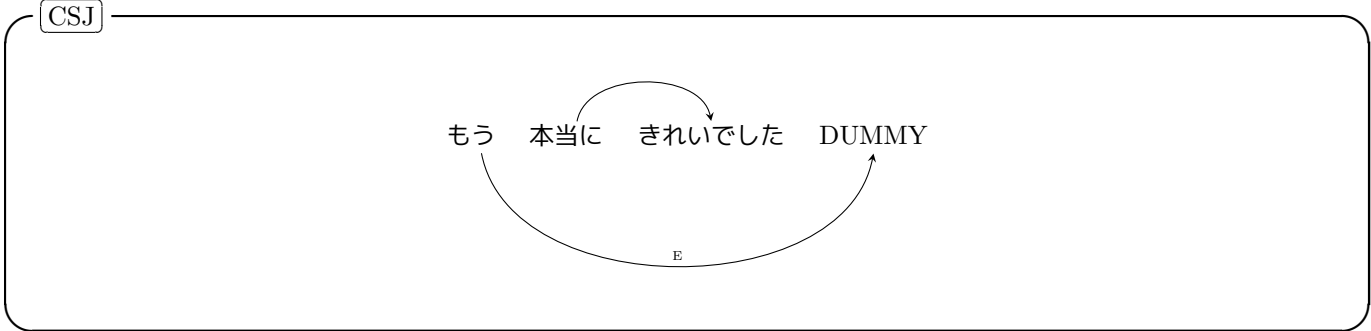
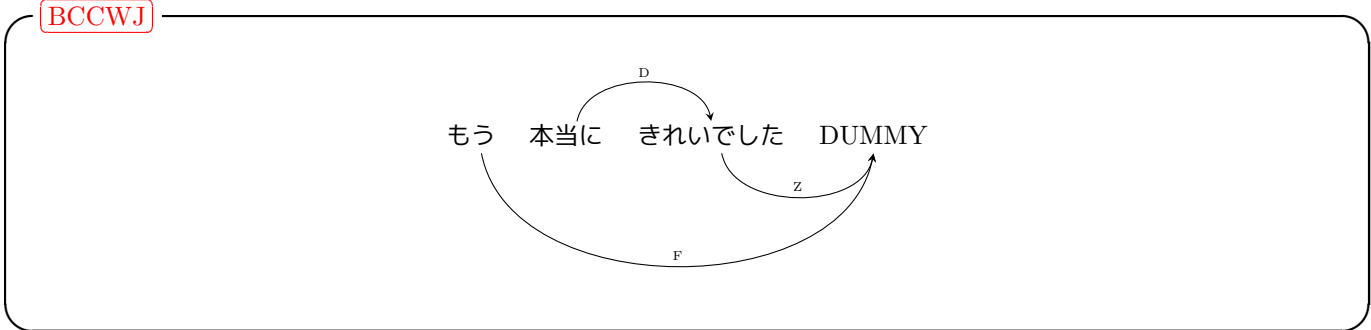


※並列句内の接続詞

3. 感動詞

CSJは、ラベル “E” を用い、感動詞のかかり先は定義しない。-1 位置の DUMMY にかけることによってかかり先なしを示す。BCCWJでは、ラベル “F” を用い、感動詞のかかり先は定義しない。-1 位置の DUMMY にかけることによってかかり先なしを示す。

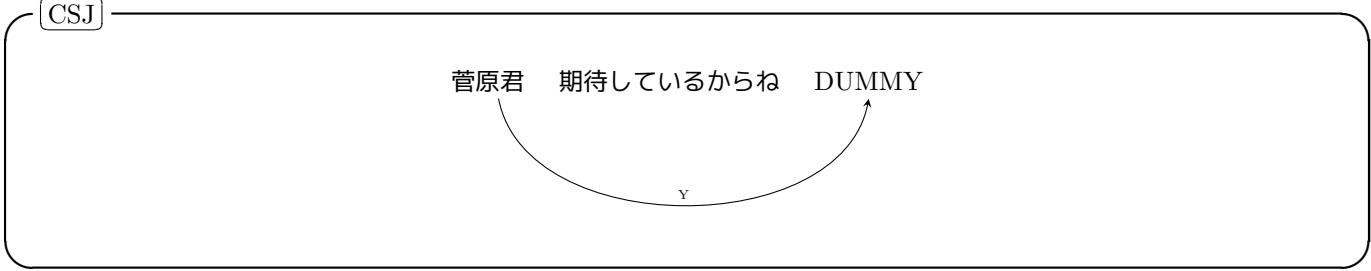
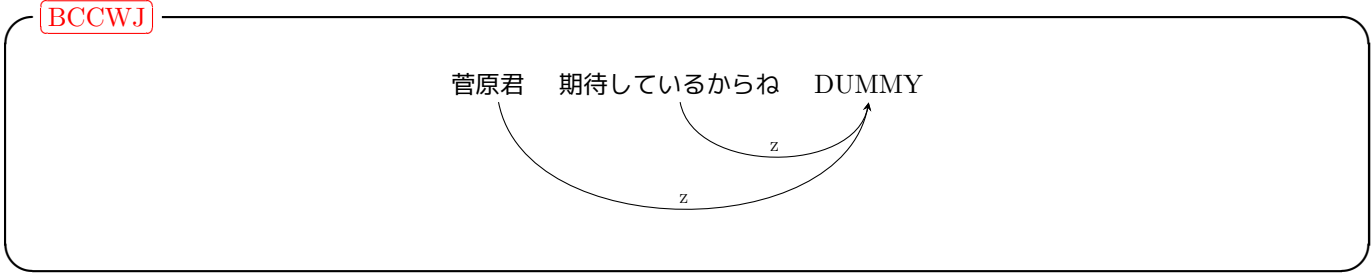
感動詞



4. 呼びかけ

CSJは、ラベル “Y” を用い、呼びかけのかかり先は定義しない。-1 位置の DUMMY にかけることによってかかり先なしを示す。BCCWJでは、ラベル “Z” を用い、呼びかけのあとに文境界相当の区切りを規定する。-1 位置の DUMMY にかけることによってかかり先なしを示す。

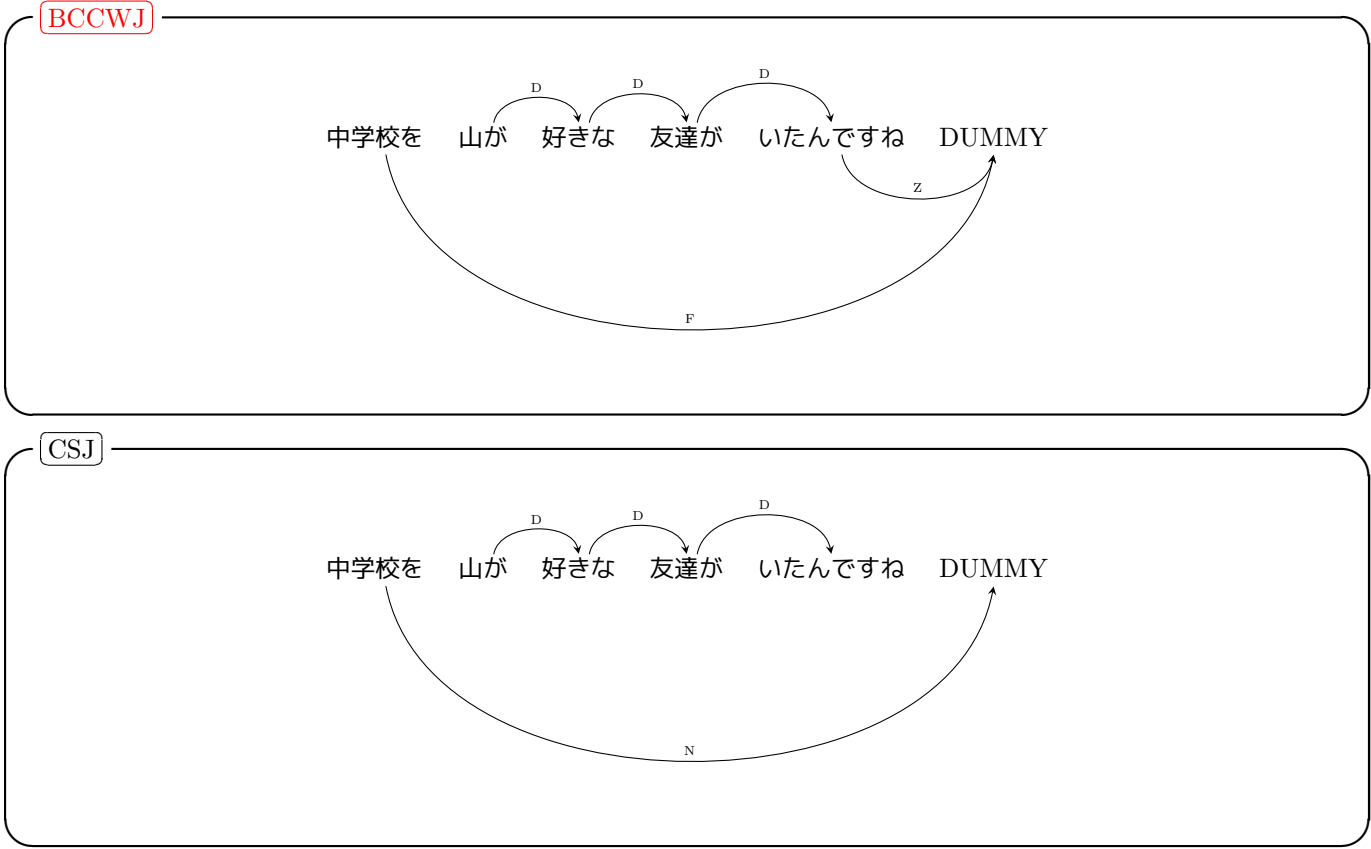
呼びかけ



5. かかり先が消失している場合に付与するラベル

CSJは、ラベル “N” を用い、-1 位置の DUMMY にかけることによってかかり先なしを示す。BCCWJは、ラベル “F” を用い、-1 位置の DUMMY にかけることによってかかり先なしを示す。

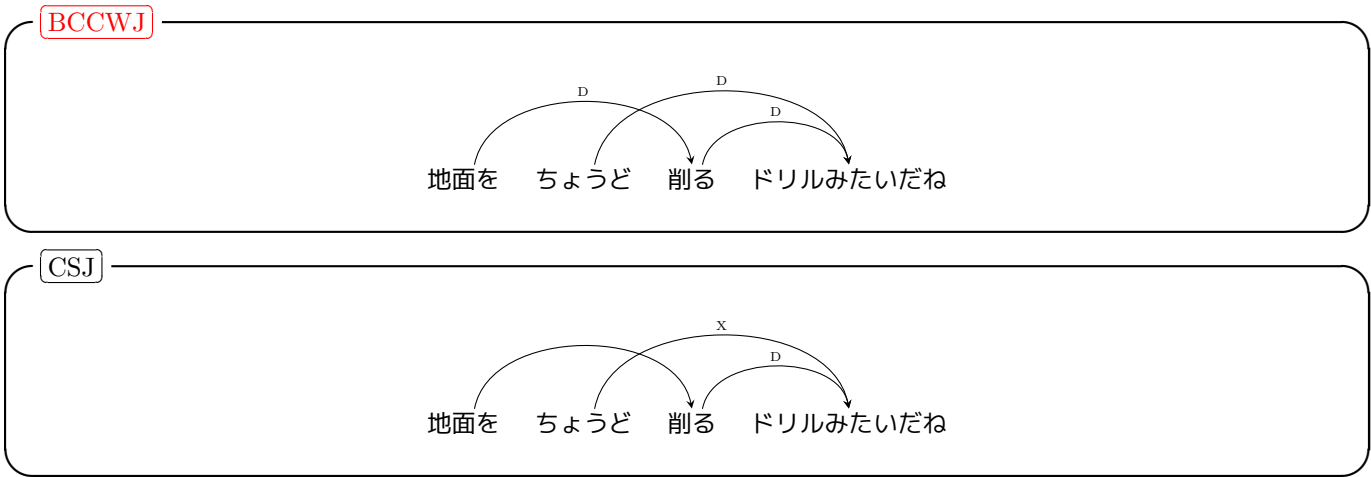
かかり先なし



6. かかり受け関係が交差する場合に付与するラベル

交差

**CSJ**は、ラベル “X” を用い、かかり受け関係が交差することを明示する。  
**BCCWJ**では、かかり受け関係の交差を許しており、単純にかかり受け関係を付与すればよい。



7. 古文・外国語に付与するラベル

古文

**CSJ**は、ラベル “K:S1” と “K:E1” を用い、それぞれ古文の開始点と古文の終了点をかかり受け木上に付与する。  
**BCCWJ**では、かかり受け木とは独立にセグメント “Foreign” として古文の範囲を指定する。同じラベルを用いて外国語についてもラベルづけする。

外国語

4.6 コメントの記述

**CSJ** は以下のコメントラベル “S” が定義されている：

- S:格表示誤り「修正候補の助詞」
- S:複数文節言い直し
- S:複数文節言い直し:S1
- S:複数文節言い直し:E1

**BCCWJ** では ChaKi の機能によりかかり受け関係にコメントが付与できる。かかり受け関係の Comment として「格表示誤り「修正候補の助詞」」を付与する予定だが、かかり受け関係に対してコメント機能が動作しないために記述しない。

**BCCWJ** では複数文節言い直しの範囲をセグメント “Disfluency” により定義する。

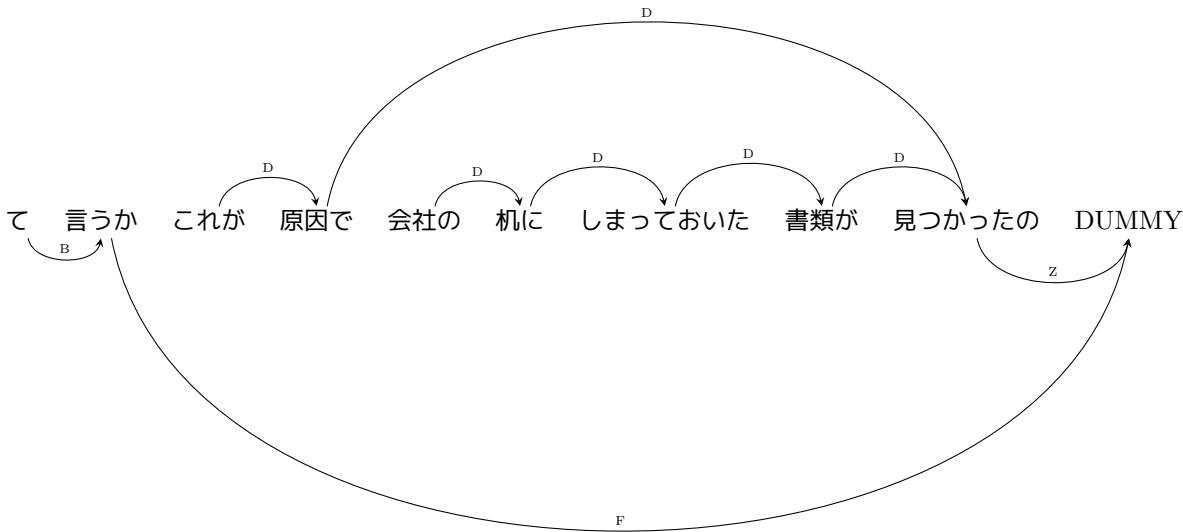
4.7 複数ラベルの併記

**CSJ** は付与するラベルが複数個になる場合は、ラベルとラベルの間に半角アンダーバーを付与する。

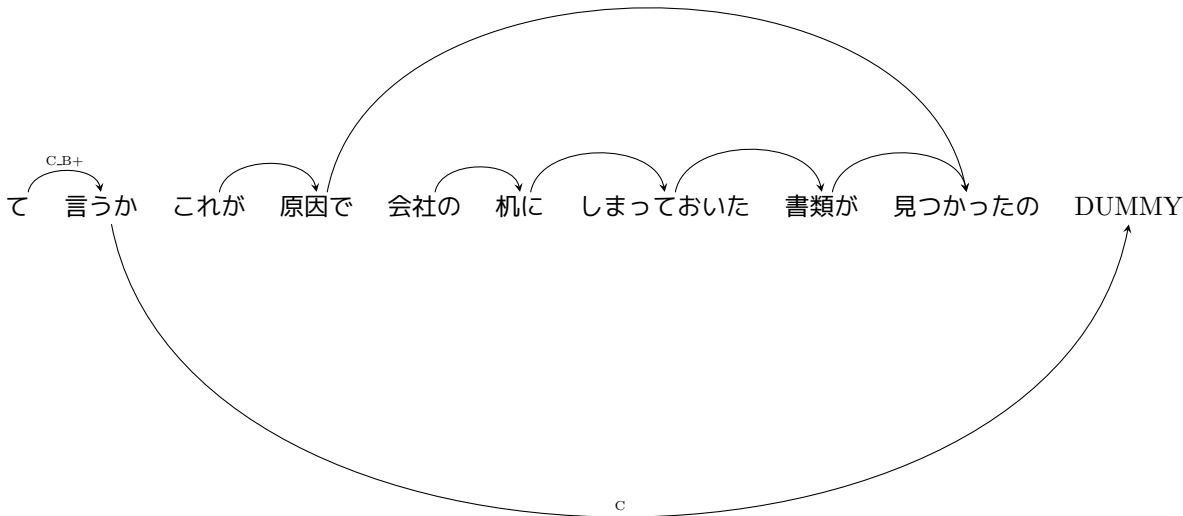
**BCCWJ** は付与するラベルが複数個にならないように、以下のように取り扱う

- ラベル “B” は文節が連結する範囲で付与し、その新しく連結されてできた文節のかかり先は最右要素のかかり先と同じものとする。
- **CSJ** におけるラベル “S” 相当の情報は **BCCWJ** において、コメント機能やセグメント “Disfluency” によって表現し、かかり受け関係のラベルに含めない。**BCCWJ** はラベル “D” でよい。**BCCWJ** はラベル “D” でよい。
- **CSJ** におけるラベル “K” 相当の情報は、**BCCWJ** において、セグメント “Foreign” によって表現し、かかり受け関係のラベルに含めない。
- **CSJ** におけるラベル “X” 相当の交差の情報は、**BCCWJ** において明示しない。

BCCWJ



CSJ





## 第 5 章

### CSJ のマニュアルに基づく比較 (2)





第 6 章

KC のマニュアルに基づく比較

本節では文献 [?] に出現する事例に基づいて、BCCWJCSJKC の基準を比較する。

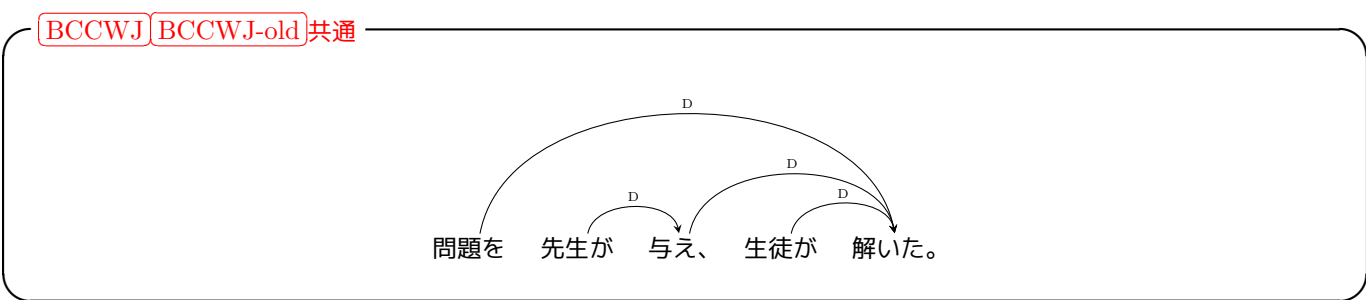
6.1 通常のかかり受け関係

6.1.1 格要素と複数の述語の関係

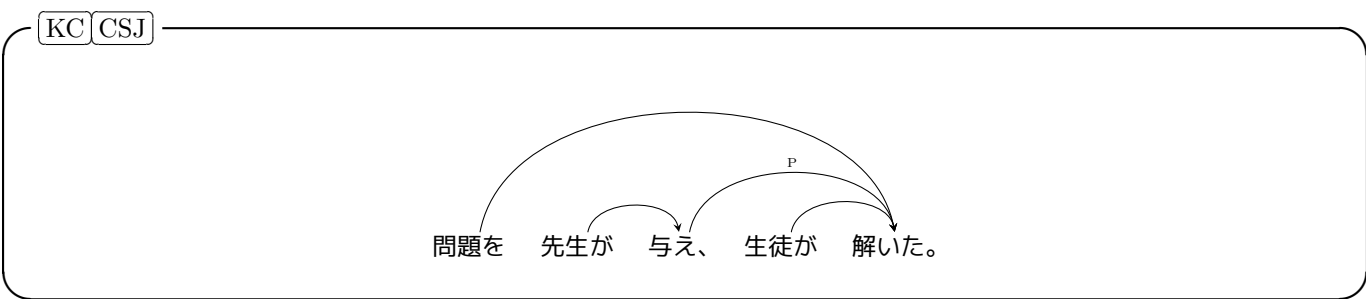
BCCWJ の 1 次アノテーションにおいて「格要素についてはかかりうる最も遠いものにかける」という基準を決めていたが、基本的に KC に合わせることにした。1 次アノテーションの基準と、今回修正する作業基準とを区別するために、前者を BCCWJ-old とする。この点については重点的に見直すこと。

1. 並列する複数の述語との関係（等位接続相当）BCCWJ 2 次チェック

BCCWJ では、述語並列を並列とみなさない。

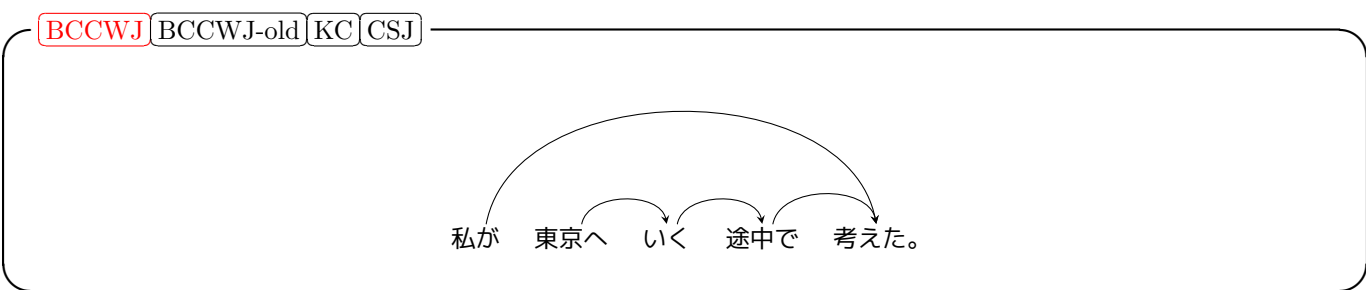


KCCSJ では、並列ラベル “P” を認定し、格要素については、並列する述語の最後の述語にかかる形で扱う。



2. 主題・主語と、従属節・主節の述語との関係

BCCWJKCCSJ とともに、従属節・主節の述語の両方にかかる場合には、主節の述語にかける。

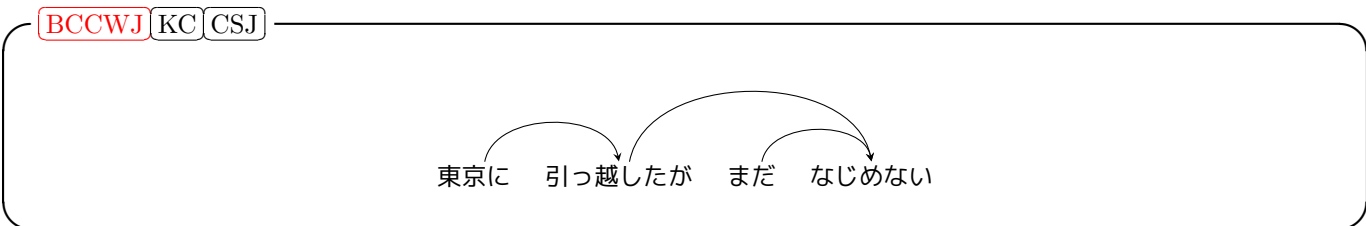


主題

主語

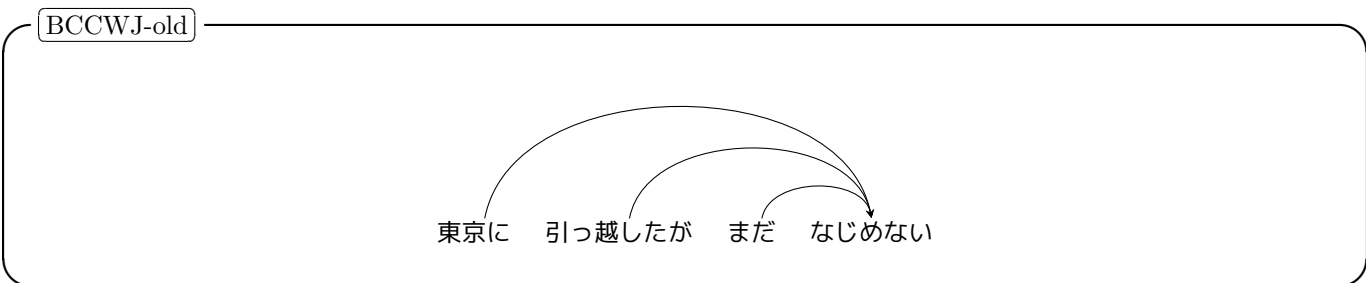
3. 主題・主語以外の格要素と、従属節・主節の述語との関係

BCCWJKCCSJ は従属節・主節の述語の両方にかかる場合には、従属節の述語にかける。偶然同一の表層格となっただけで、実際には主節の格要素が省略されているとみなす。



主題・主語以外の格要素

BCCWJ-old は、従属節・主節の述語の両方にかかる場合には、主節の述語にかける。尚、主節の主語は（文脈中に）別に存在し、従属節の主語でしかない場合はもちろん従属節にかける。

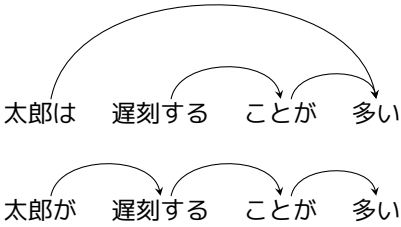


4. 文末に補助的な述語がある場合

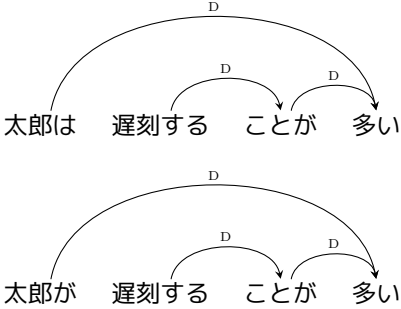
・「ことが多い」「経緯がある」「する予定」など、名詞/形式名詞相当があるものについて「は」「も」と「が」の扱いを以下のように変更すること。

何が「補助的な述語」であるのかは KC には定義されていない。BCCWJ においては別途補助的な述語リストを作成する。

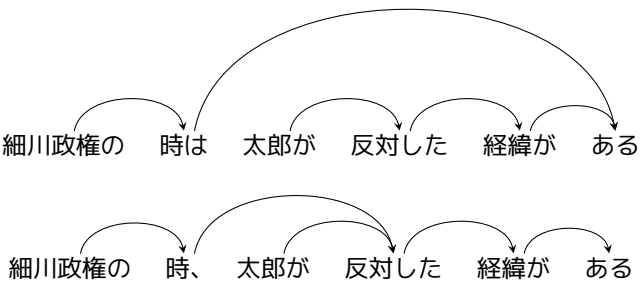
BCCWJ KC CSJ



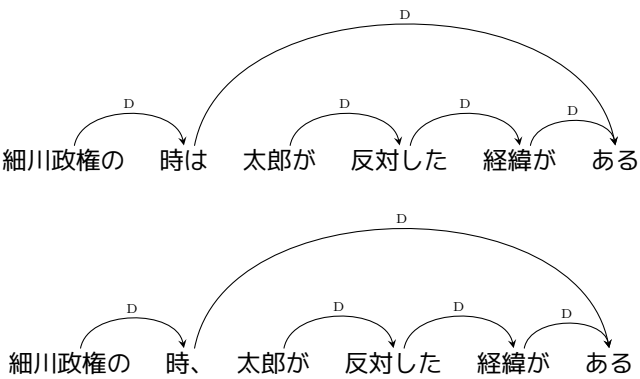
BCCWJ-old



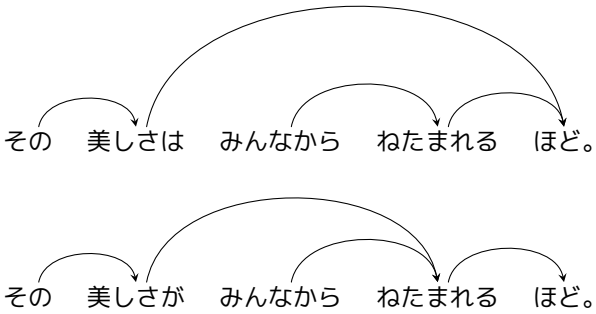
BCCWJ KC CSJ



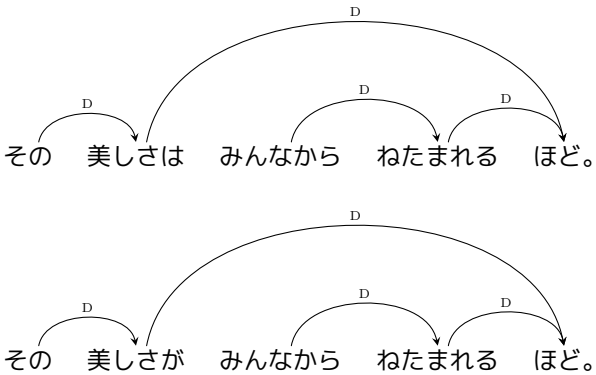
BCCWJ-old



BCCWJ KC CSJ

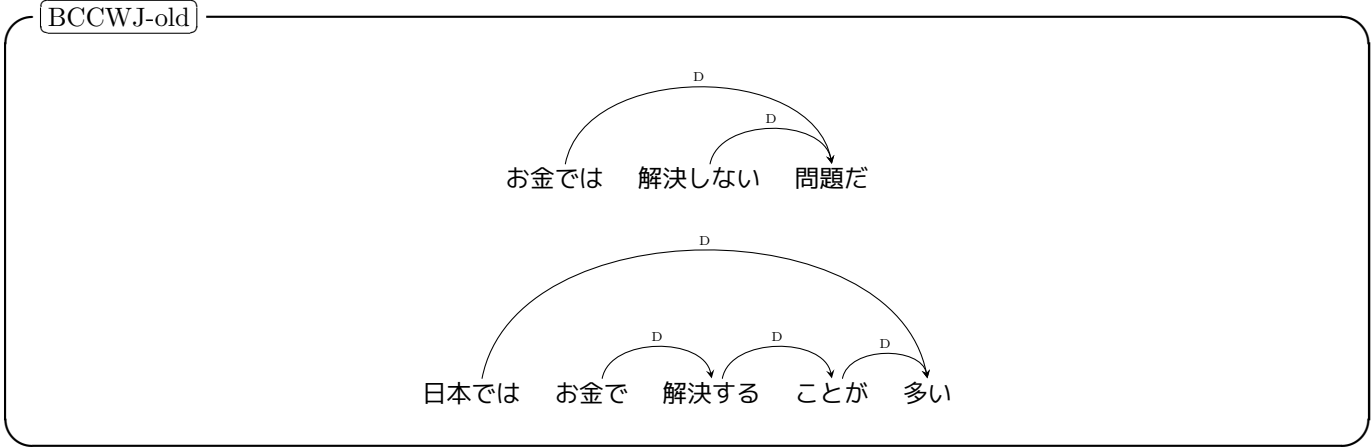
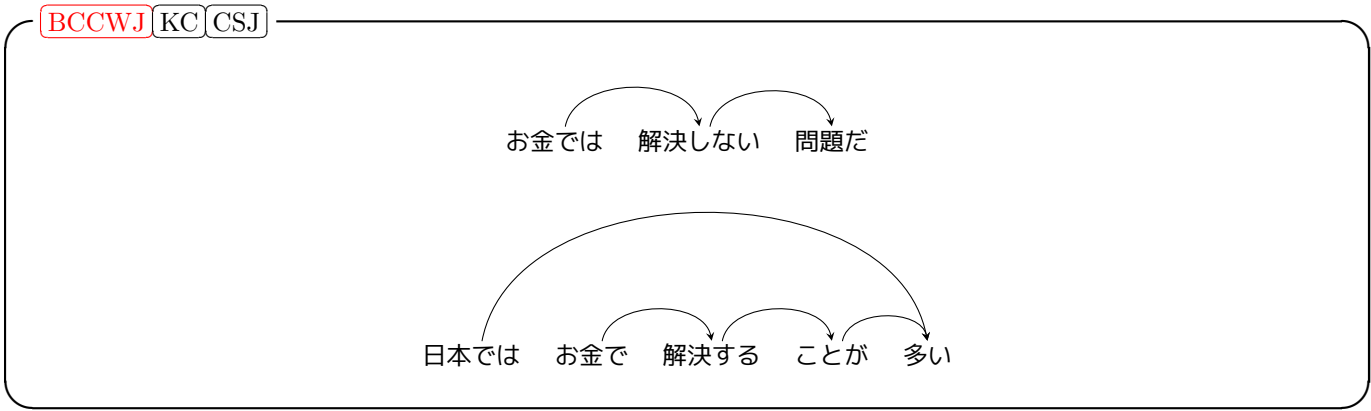


BCCWJ-old

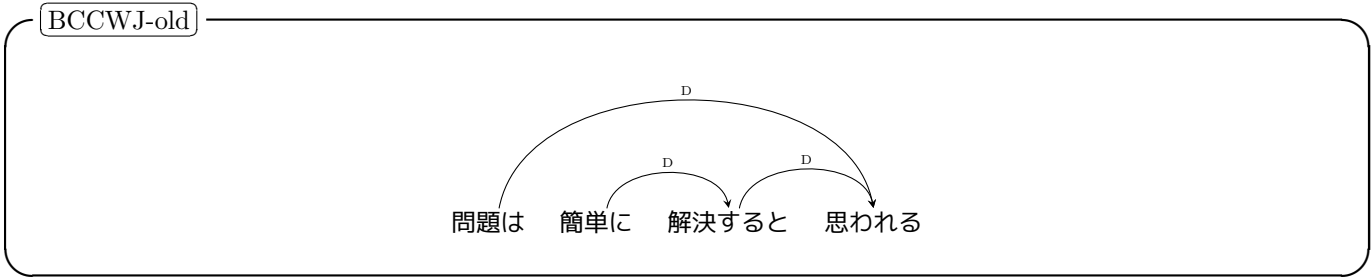
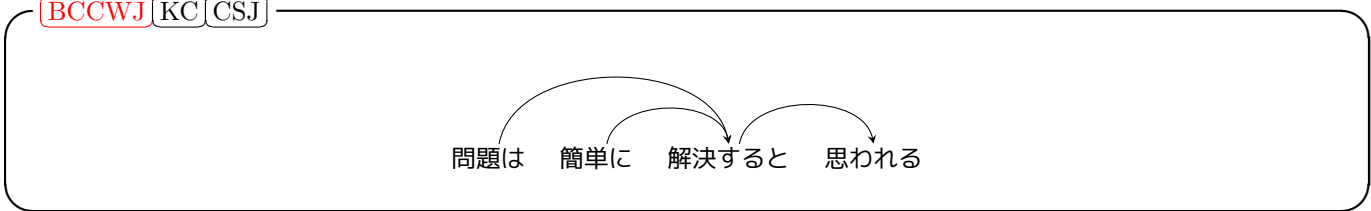


・「では」「には」で判別が難しい場合  
以下の例では意味にまで踏み込んで考慮する。「お金では」「問題だ」にはかからない。「日本では」は「多い」にはかかる。

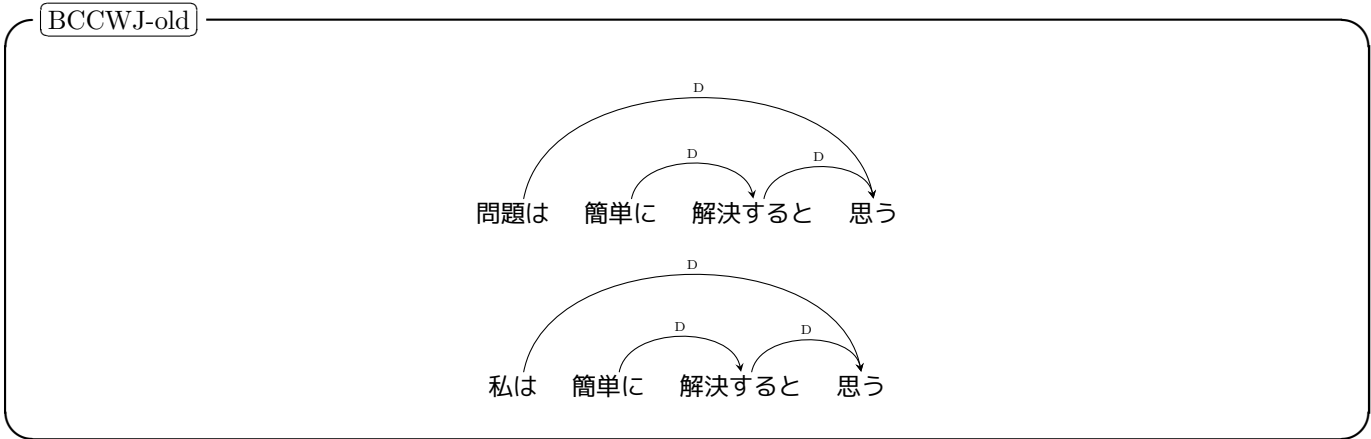
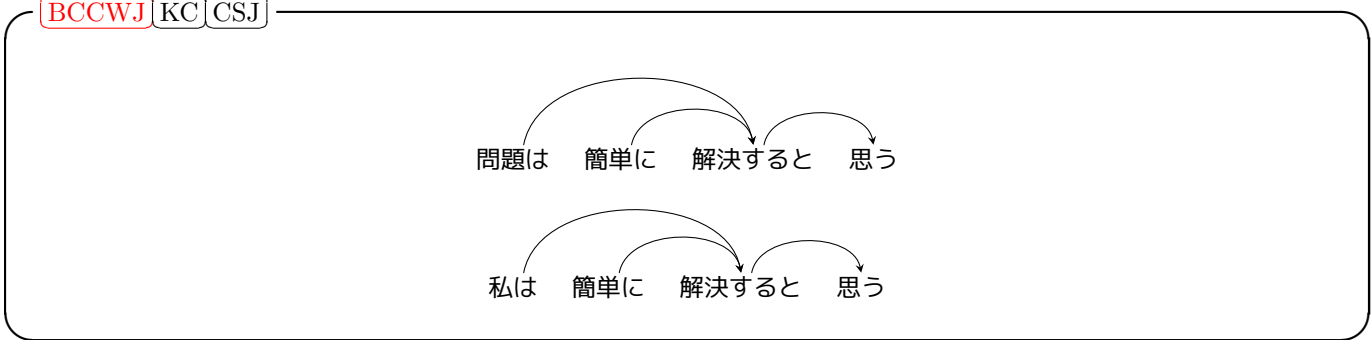
では  
には



・「 — は — すると（ように）思われる」（いわれる、される、みられる、考えられる）  
以下の例では意味にまで踏み込んで考慮する。「問題は」は「思われる」にはかからない。



・「 — は — すると（ように）思う」（思える、見える、聞く、いう、いえる、いってよい、いわざるをえない）

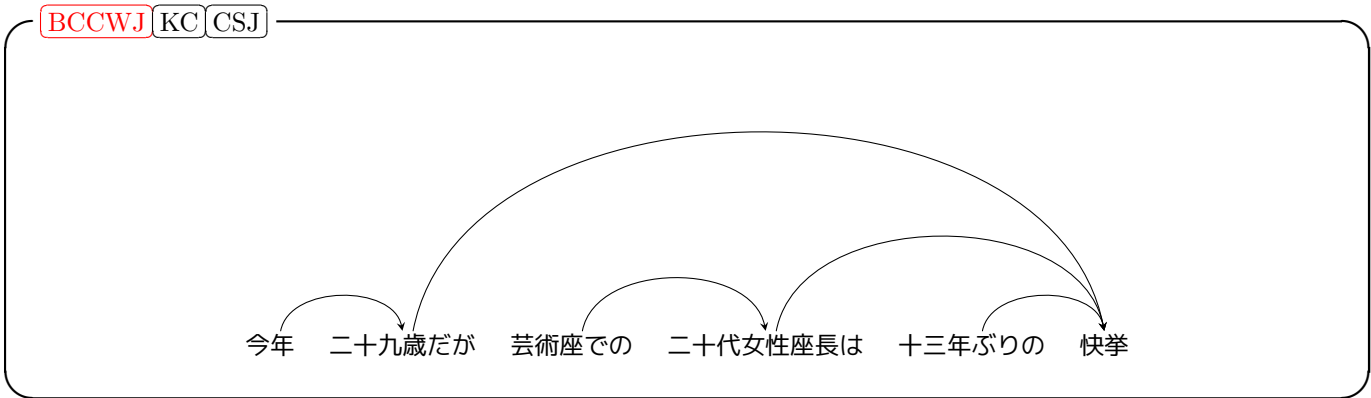


・「しかない」「ほかない」  
KC 一文節  
CSJBCCWJ 一文節にするために B+ タグもしくは B タグで連結

6.1.2 かかり先が非常に曖昧な場合

1. 接続詞

2. 従属節
- 接続詞
- 従属節
- 従属節などについてかかり先が非常に曖昧な場合は、文末（最後の述語を含む文節）にかける。



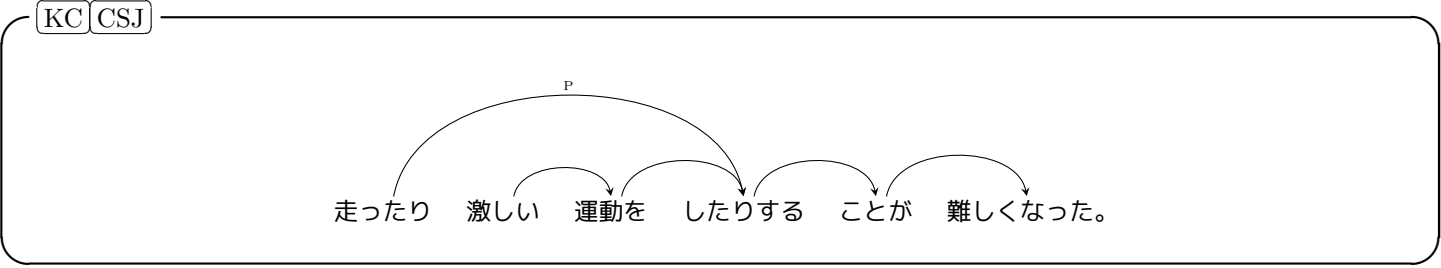
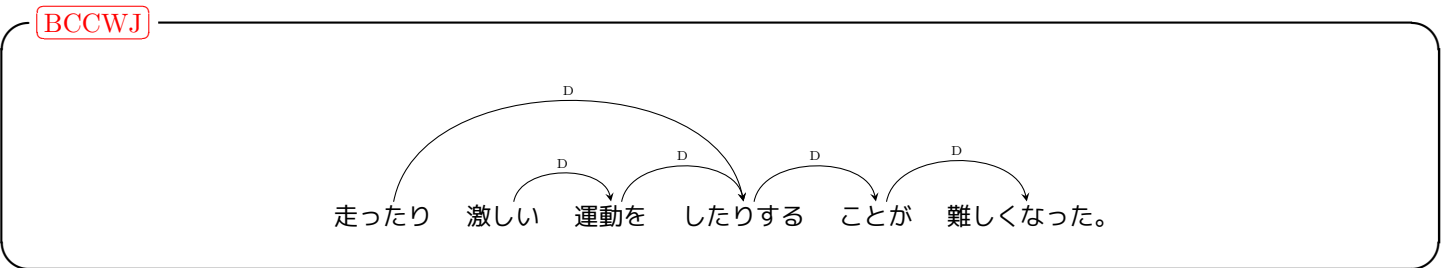
3. AのBのC

KC CSJ では、「AのBのC」や「AしたBのC」などで、かかり先が非常に曖昧な場合、近いかかり受けを優先する。  
BCCWJ でも連体修飾については近いかかり受けを優先する。

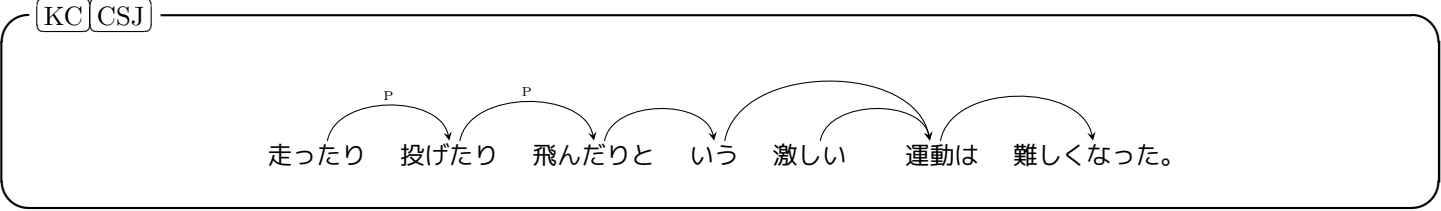
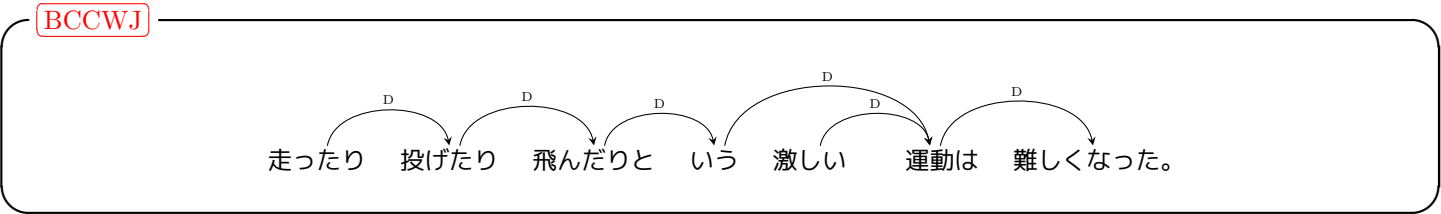
6.2 並列構造アノテーション

6.2.1 並列構造一般

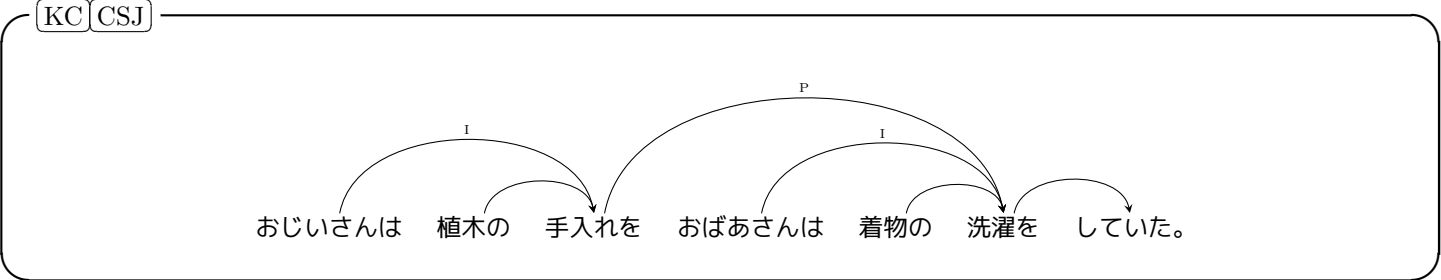
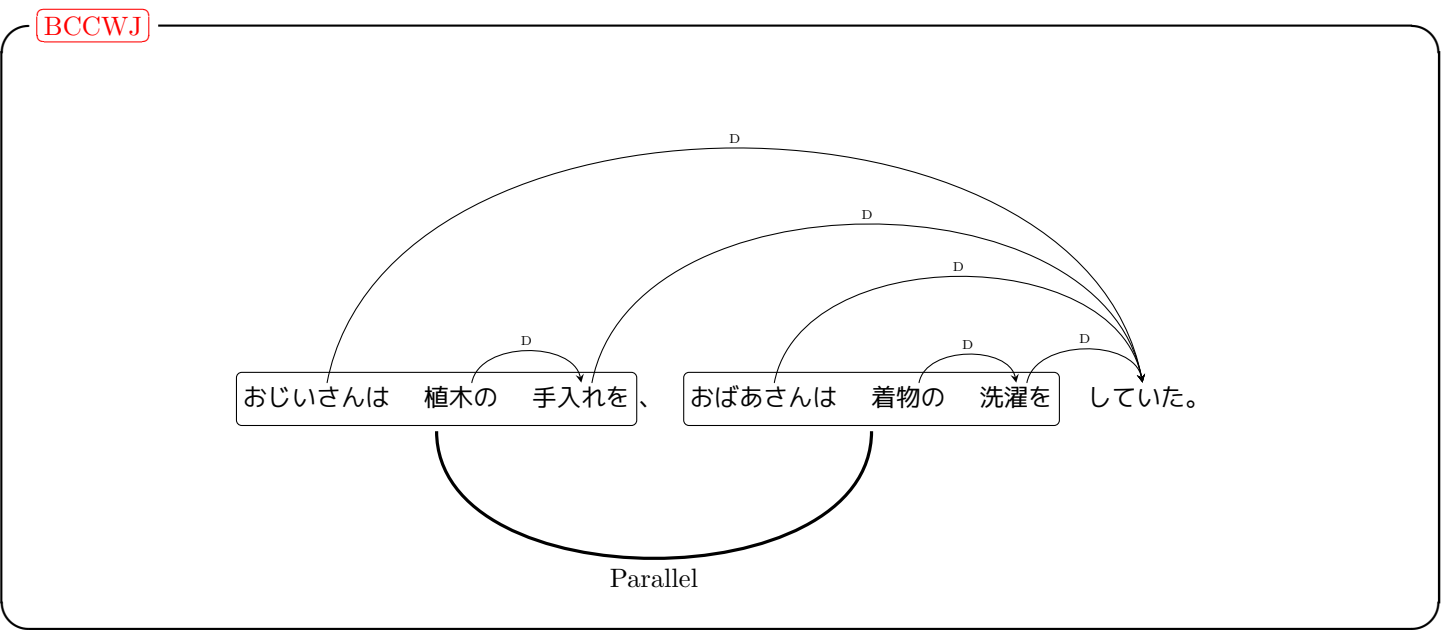
BCCWJにおいて、述語並列は並列としない。

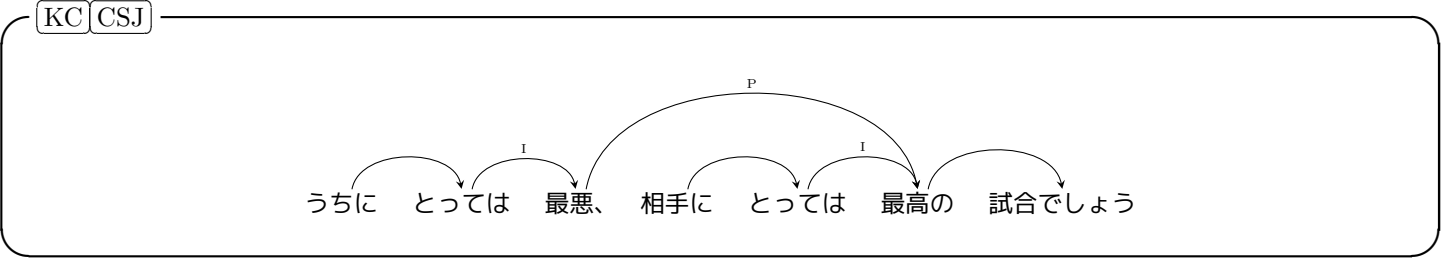
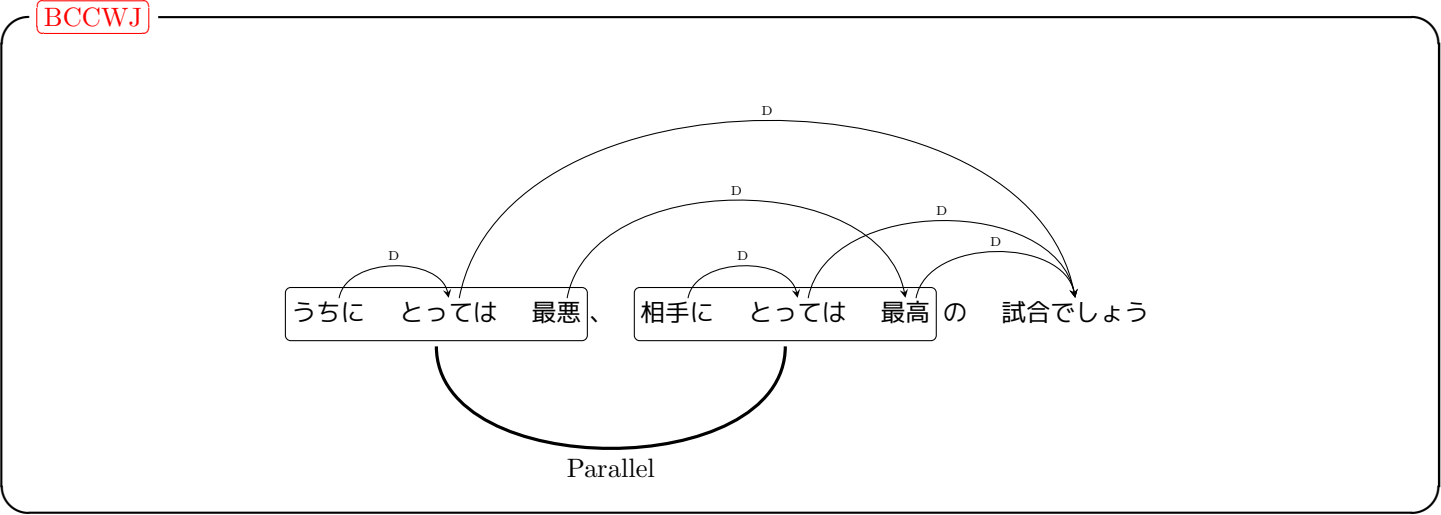
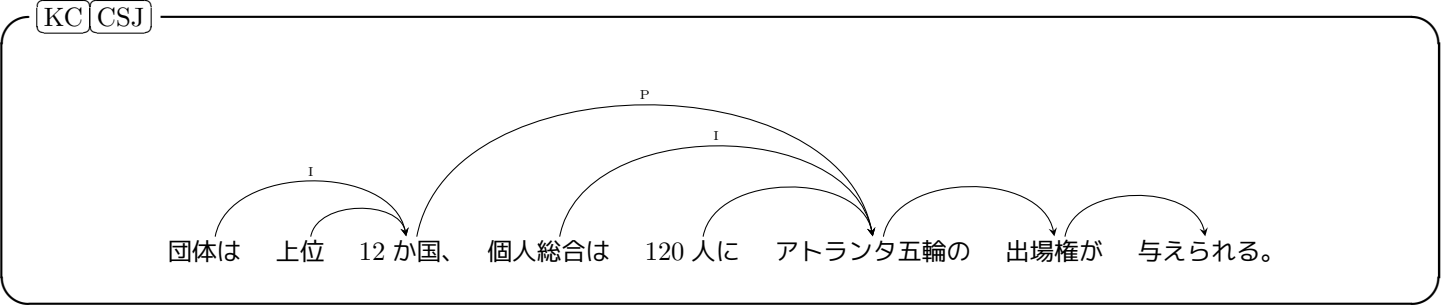
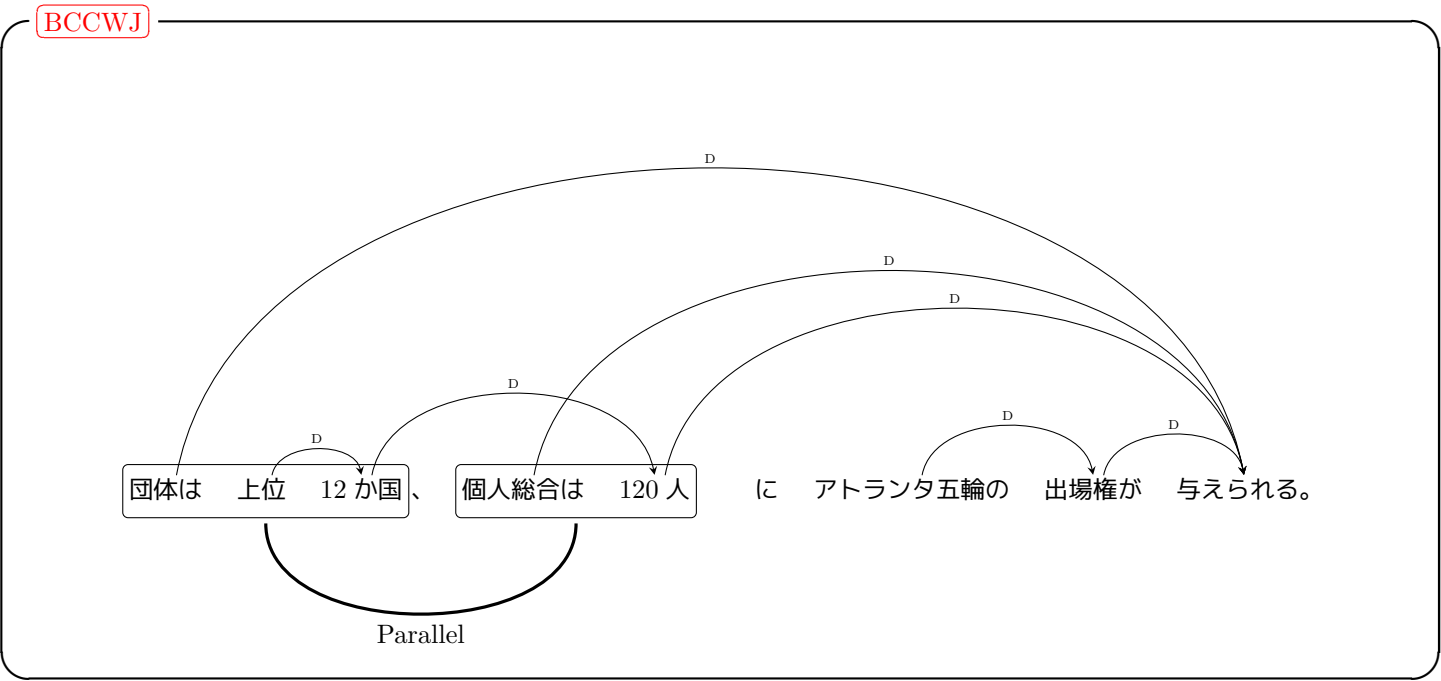


BCCWJにおいて、述語並列は並列としない。



6.2.2 部分並列

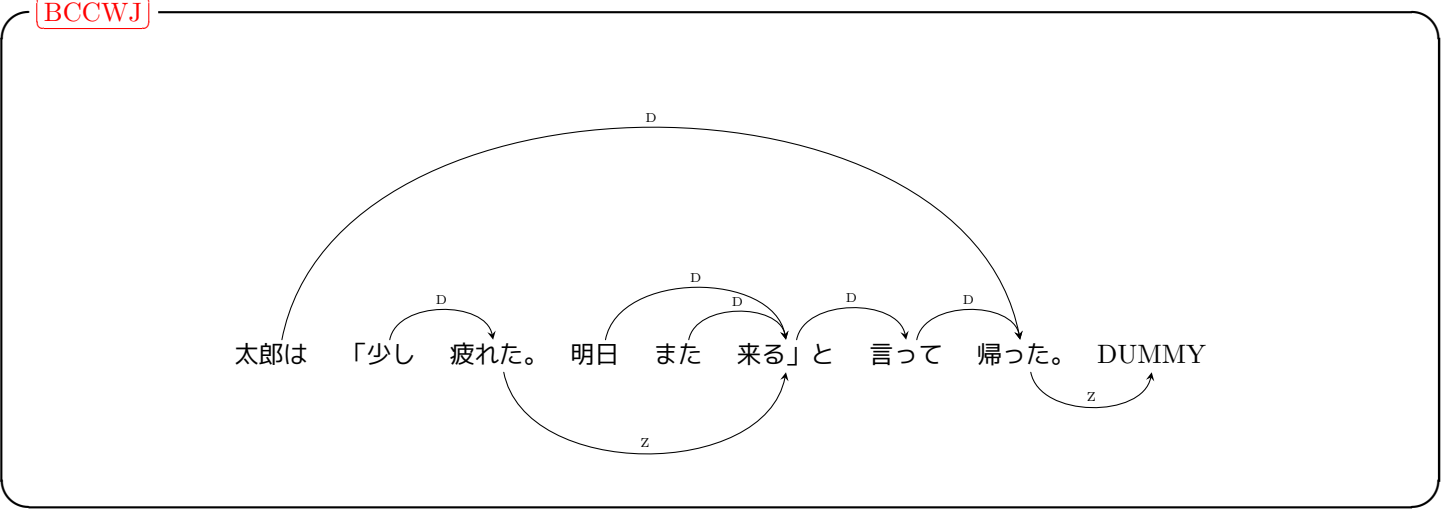




6.2.3 括弧内の複数文

文境界修正

BCCWJでは、括弧の内部に句点によって区切られている文がある場合は、ラベル“Z”を付与する。かかり先は括弧の内部の最終要素を含む文節とする。



KCでは、括弧の内部に句点によって区切られている文がある場合は、各文の末尾の述語を並列関係として扱う。

KC

CSJ は音声データを対象としており、上のような表現は想定していない。

6.2.4 テ形

テ形

BCCWJ では基本的にテ形を並列として扱わない。  
KC CSJ では以下の取り扱いを行っている。

1. 単なる修飾の場合→BCCWJ KC CSJ並列としない

BCCWJ KC CSJ

2. (弱い) 時間経過の場合→BCCWJ KC CSJ並列としない

BCCWJ KC CSJ

3. 非常に類似性が高く、等位接続とみなせる場合→BCCWJ並列としないKC CSJ並列とする

BCCWJ

KC CSJ

6.2.5 連体形

連体形が続く場合は並列とはせず、それぞれが体言にかかっているとみなす。

BCCWJ KC CSJ

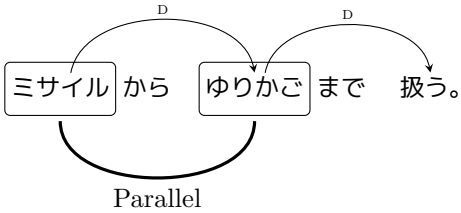
6.2.6 からまで

からまで

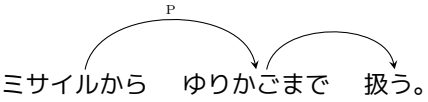
「からまで」という表現は次のように扱う

1. 範囲を表す「A から B まで扱う」のような場合、「A」と「B」を並列、「B まで」と「扱う」を通常のかかり受けとして扱う

BCCWJ

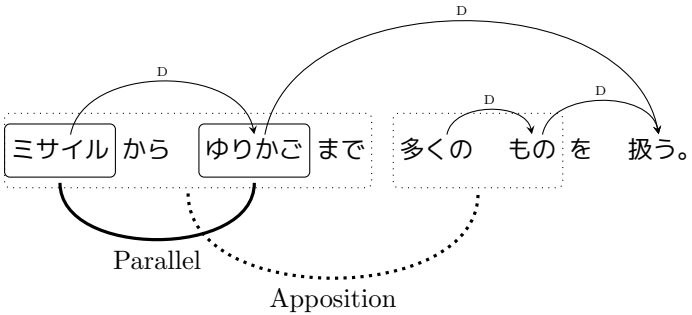


KC CSJ



2. 範囲を表す「A から B まで多くのものを扱う」のような場合、「A」と「B」を並列、「B まで」と「扱う」を通常のかかり受けとし、「A から B まで」と「多くの商品」の間を同格とする。

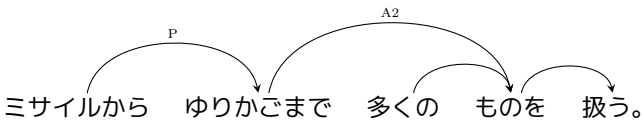
BCCWJ



KC

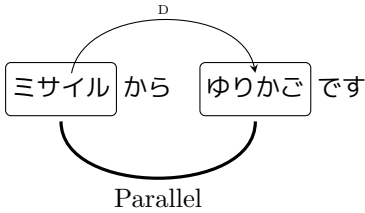


CSJ



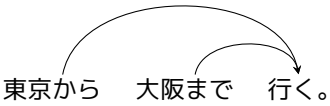
3. 範囲を表す「A から B です」のような場合、「A」と「B」を並列、「B として扱う

BCCWJ



4. 経路を表す「A から B まで行く」のような場合、「A から」と「B まで」は「行く」にかける

BCCWJ KC CSJ

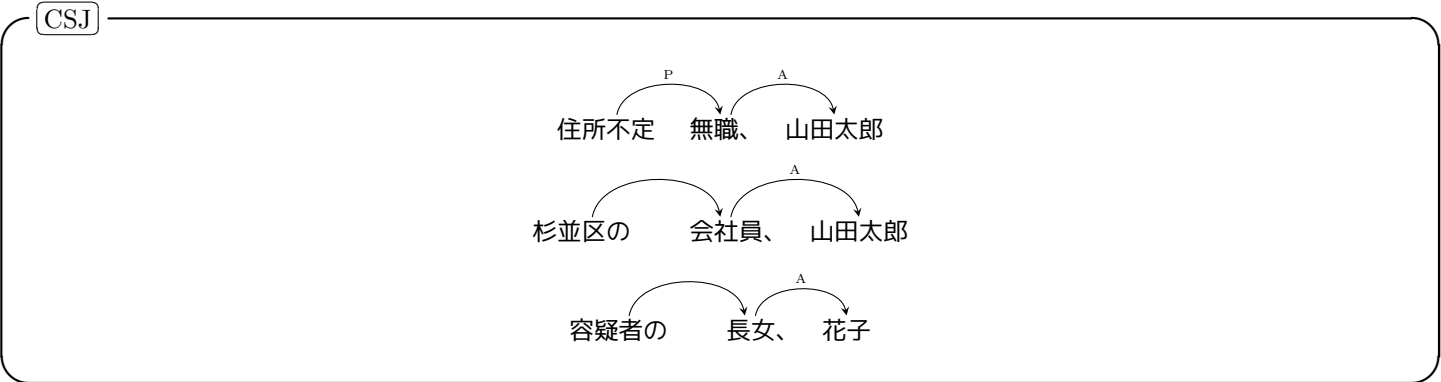
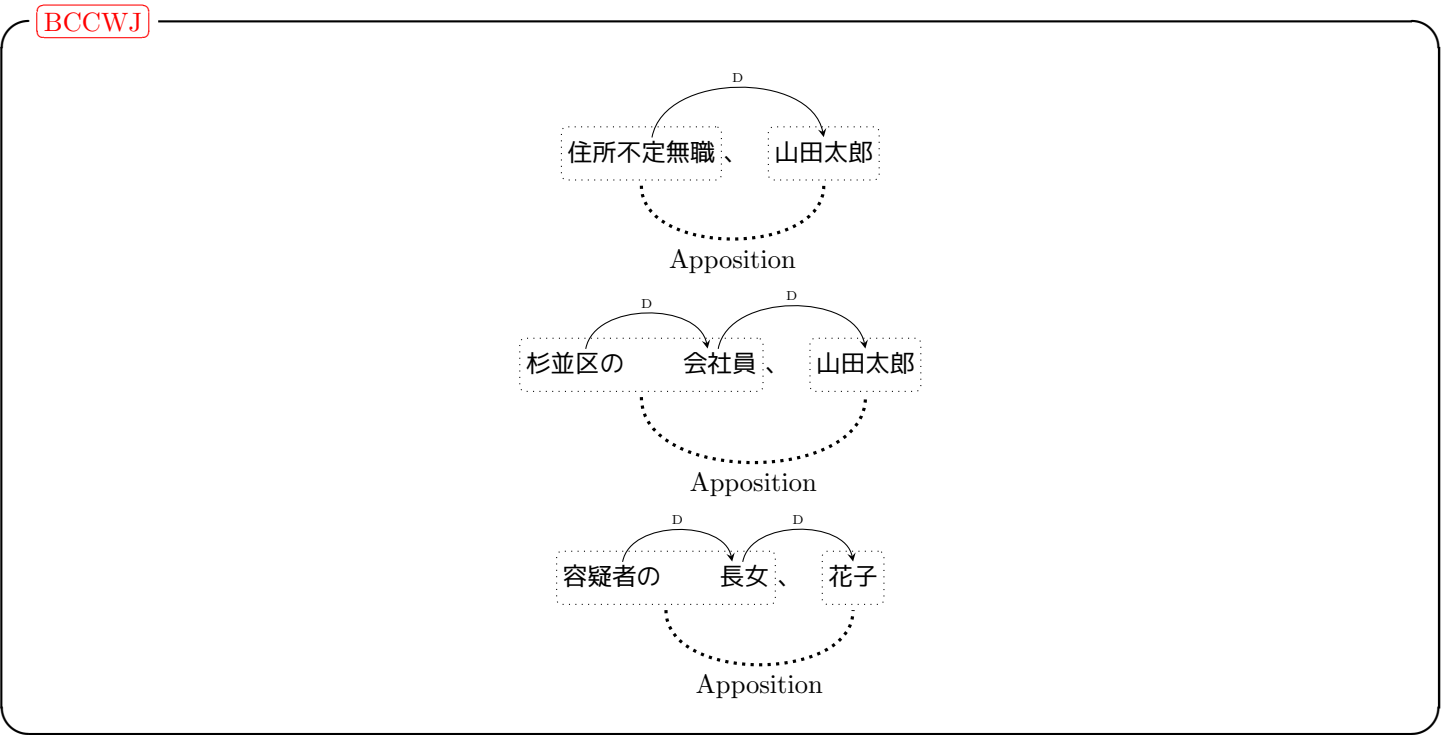


- 5. 記号の場合はつけなくてよい
- 6. 文節内で並列構造が閉じている場合にはつけなくてよい

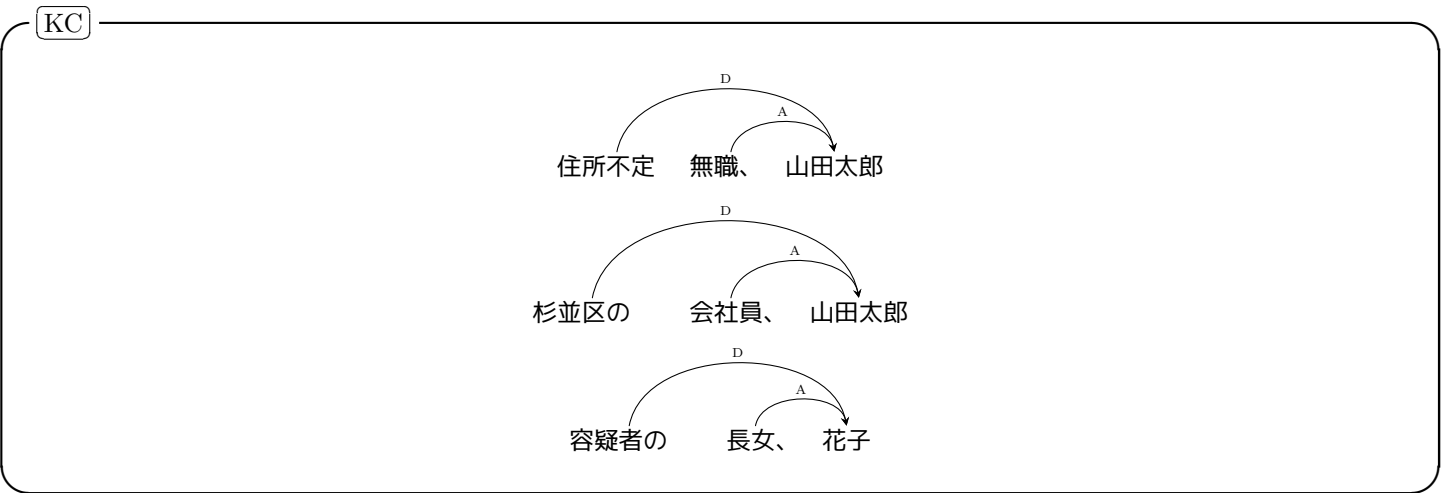
### 6.3 同格構造アノテーション

#### 6.3.1 住所・職業・続柄と人名

[BCCWJ][CSJ]においては、新聞記事に見られる「住所、職業、人名」「住所の職業、人名」などの表現では、きちんと範囲を認定したうえで同格として扱う。敬称や容疑者も同格に含める。



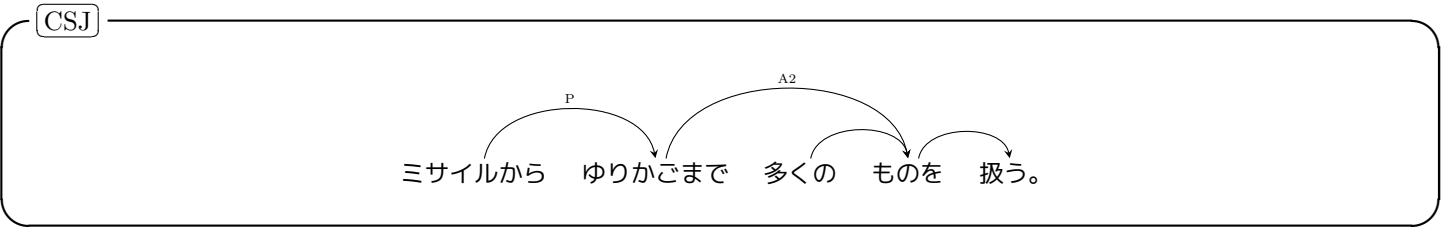
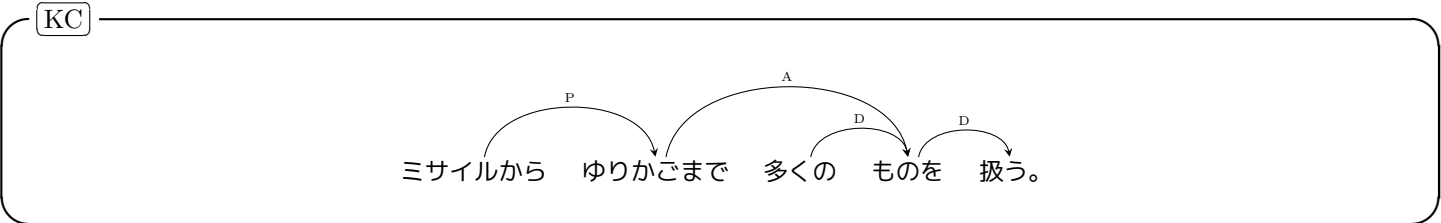
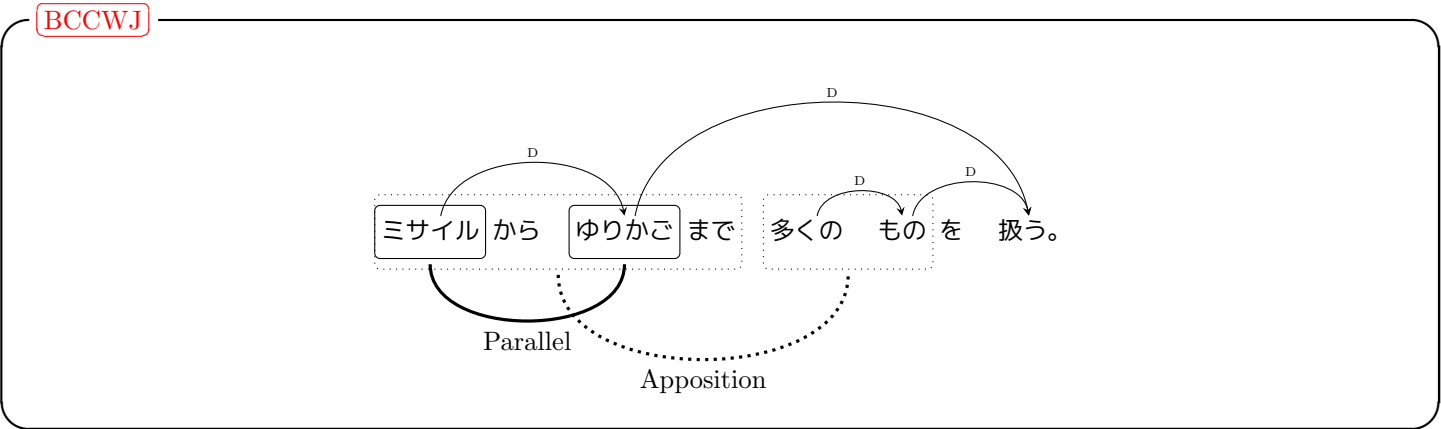
[KC]においては、新聞記事に見られる「住所、職業、人名」「住所の職業、人名」などの表現では、住所は人名にかかり、職業は人名と同格として扱う。また、続柄と人名も同格とみなす。



6.3.2 範囲を表す「からまで」

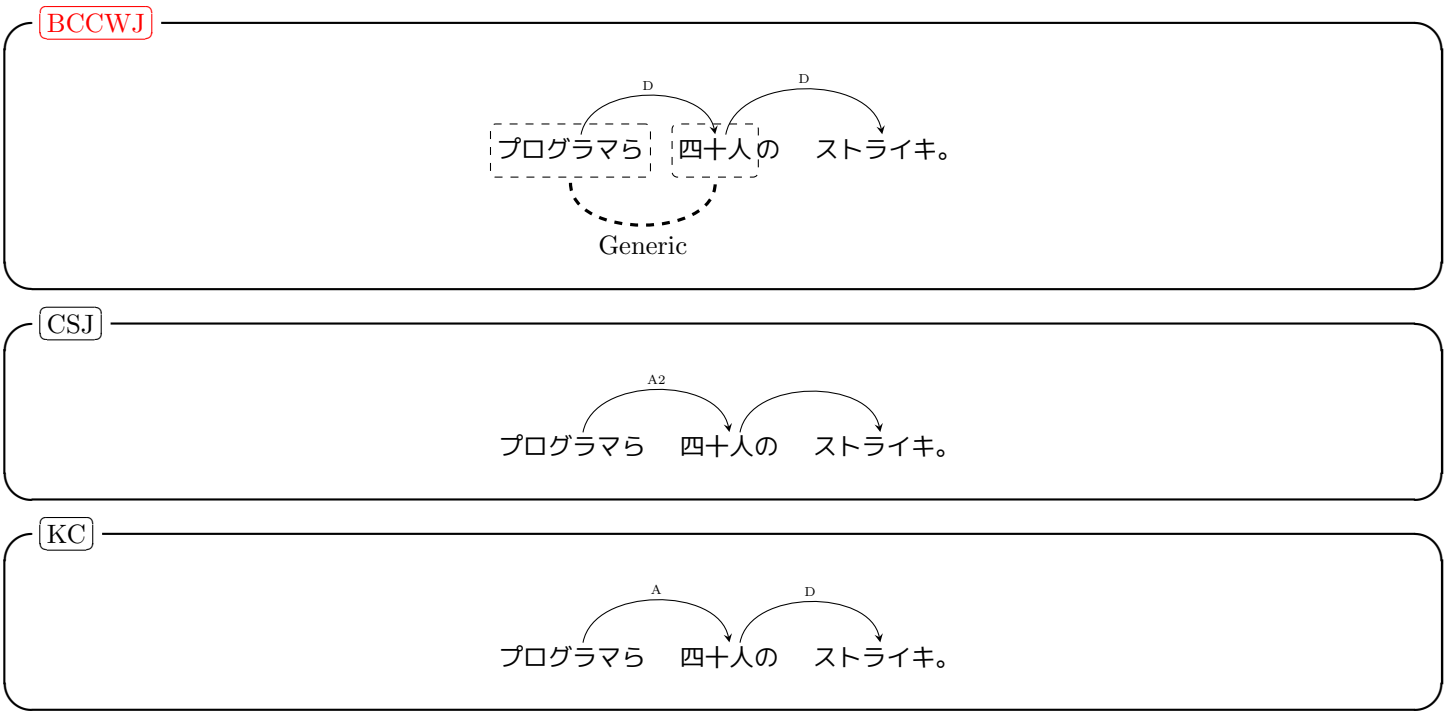
からまで

「A から B まで多くのものを扱う」のような場合、「A から」と「B まで」を並列、「B まで」と「扱う」を通常のかかり受けとし、「A から B まで」と「多くの商品」の間を同格とする。

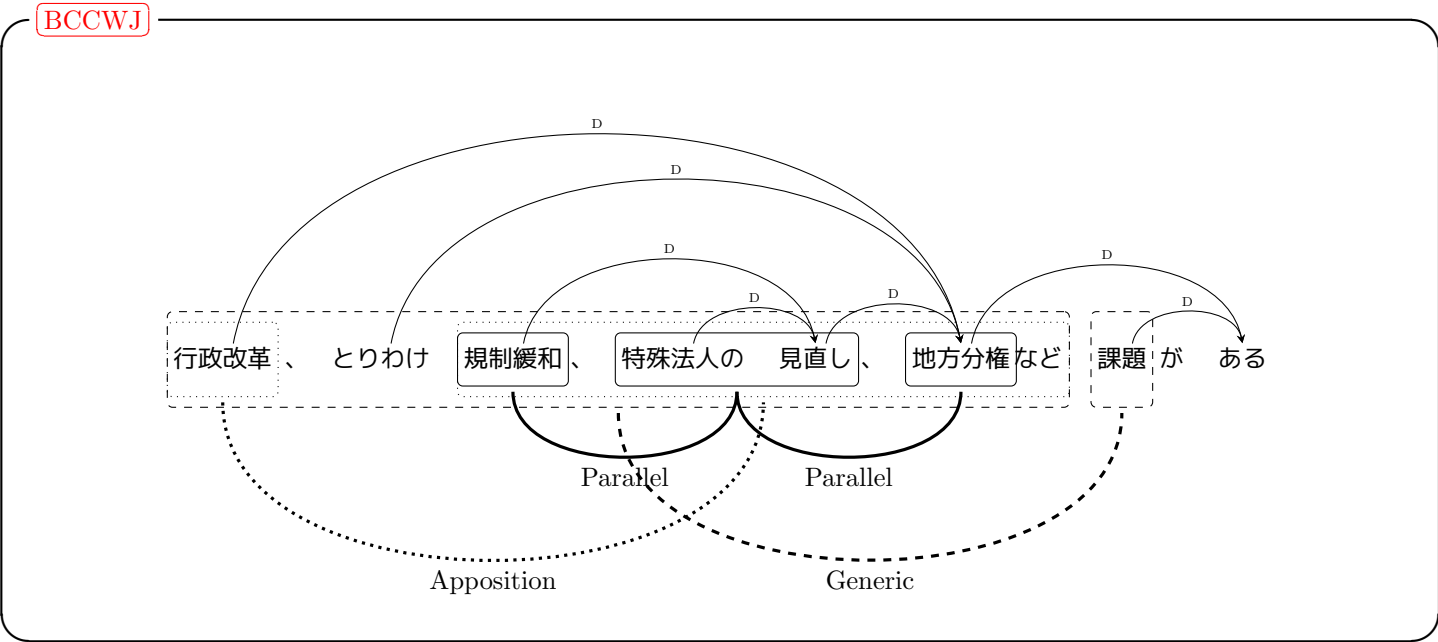




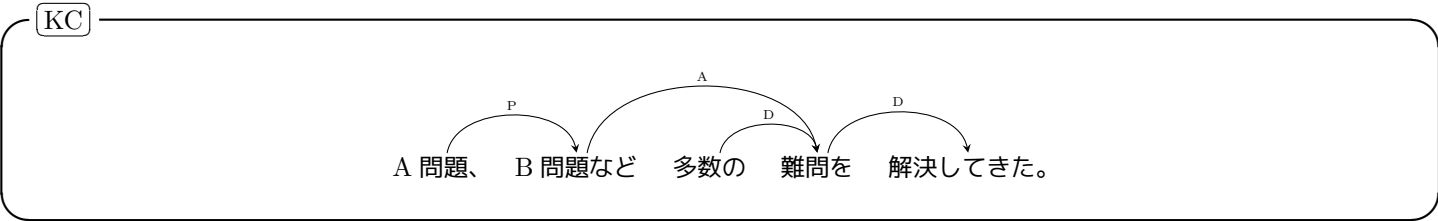
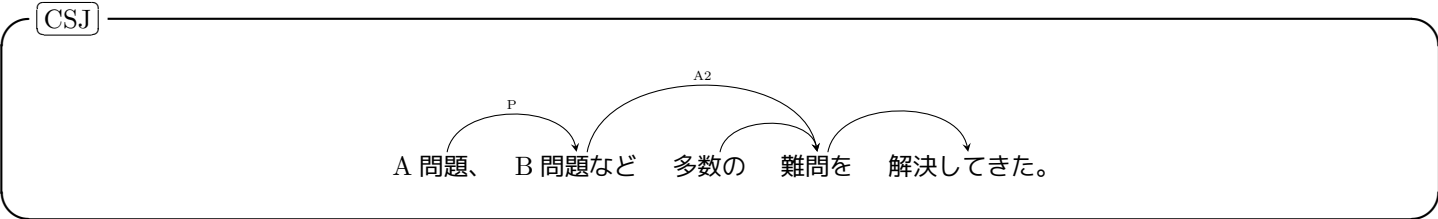
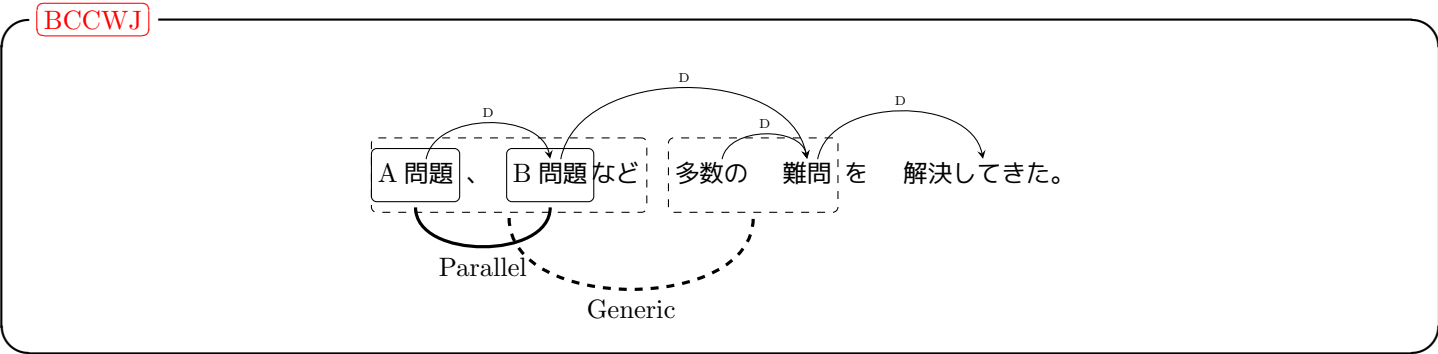
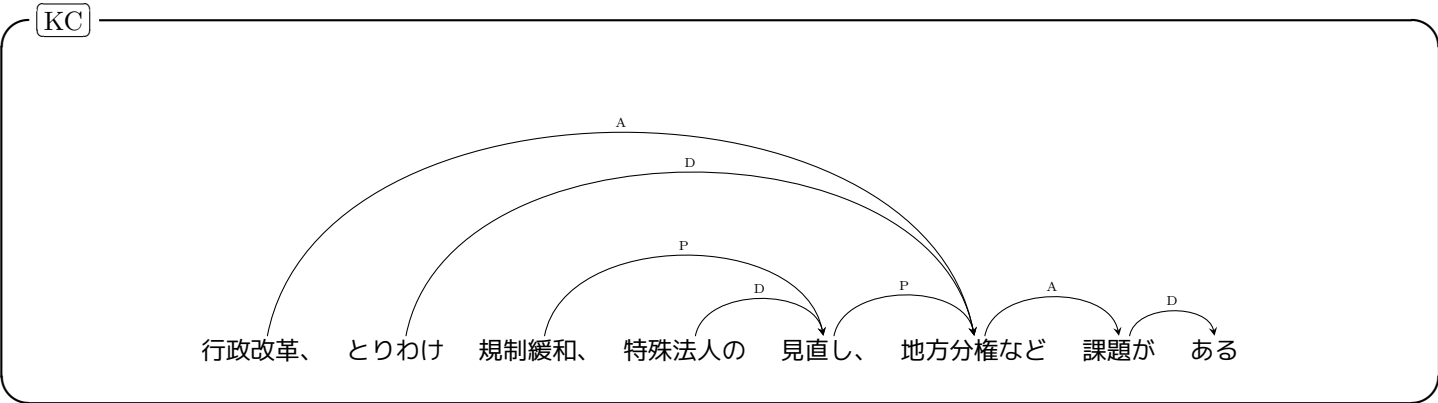
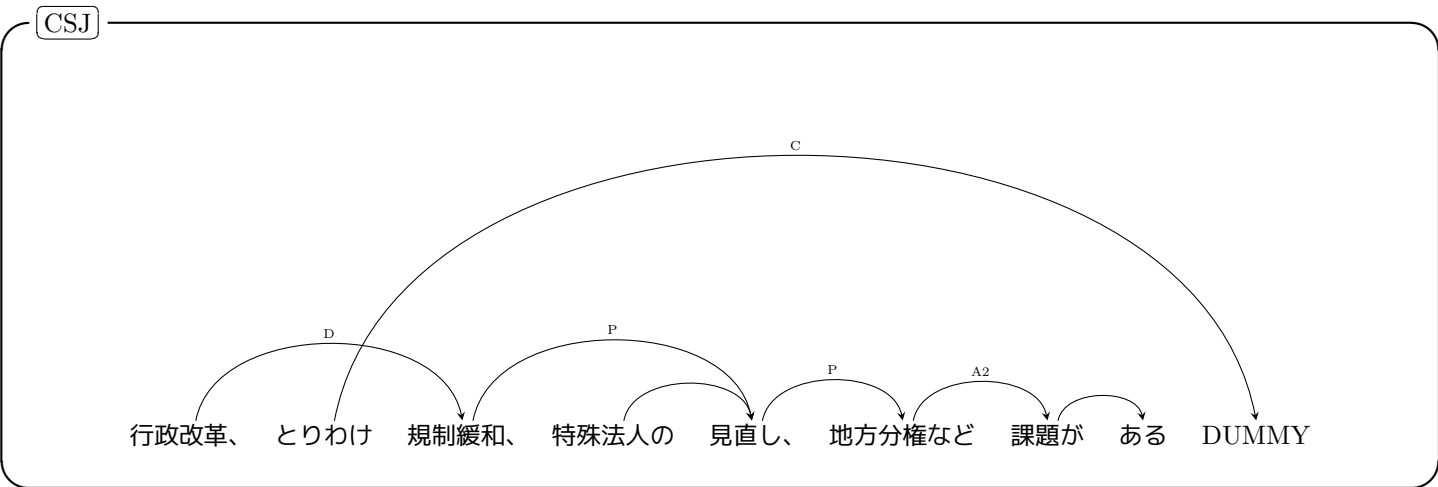
6.3.3 「体言+ら | たち | その他 | など | と | すなわち | つまり | とりわけ | 特に」



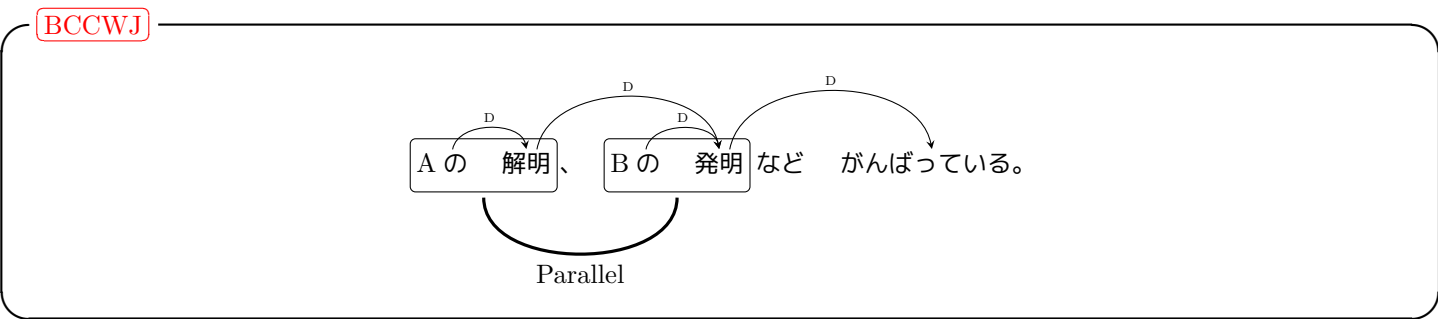
・後ろに体言の並列がある場合は全体を同格とみなす



CSJ で「すなわち」「とりわけ」は「言い直し」と認定している。



・同格関係の体言がない場合



CSJ KC

A の 解明、 B の 発明など がんばっている。

BCCWJ

市立船橋は 4 試合で 15 得点と 高い 攻撃力 を 誇る。

Apposition

※以下の例は CSJ で同格になるかどうか不明。

CSJ KC

市立船橋は 4 試合で 15 得点と 高い 攻撃力を 誇る。

BCCWJ

地球規模の 環境、 人口、 食糧 など 広範に

Parallel Parallel

Generic

CSJ

地球規模の 環境、 人口、 食糧など、 広範に

KC

地球規模の 環境、 人口、 食糧など、 広範に

6.3.4 「体言（+、）＋「～」」

体言とカギカッコが隣接する場合には同格の可能性を考慮する。

BCCWJ

アニメビデオ 「蓮如さま」を 発売した。

CSJ KC

アニメビデオ 「蓮如さま」を 発売した。

・BCCWJでは、以下のような事例も同格とみなさない。

BCCWJ

Diagram illustrating the BCCWJ dependency parse for the sentence: 5日タ刊 1面 「パソコンTVに 受信料」の 記事で、. The root node is D, which branches into D and D. The left D branches into D and D, which then branch into 5日タ刊 and 1面 respectively. The right D branches into D and D, which then branch into 「パソコンTVに 受信料」 and の respectively. The final D branches into 記事 and で、.

CSJ で以下の例は類似性が低いために同格にしない

CSJ

Diagram illustrating the CSJ dependency parse for the sentence: 5日タ刊 1面 「パソコンTVに 受信料」の 記事で、. The root node is D, which branches into D and D. The left D branches into D and D, which then branch into 5日タ刊 and 1面 respectively. The right D branches into D and D, which then branch into 「パソコンTVに 受信料」 and の respectively. The final D branches into 記事 and で、.

KC

Diagram illustrating the KC dependency parse for the sentence: 5日タ刊 1面 「パソコンTVに 受信料」の 記事で、. The root node is A, which branches into D and D. The left D branches into D and D, which then branch into 5日タ刊 and 1面 respectively. The right D branches into D and D, which then branch into 「パソコンTVに 受信料」 and の respectively. The final D branches into 記事 and で、.

6.3.5 「用言+など」

BCCWJ では同格を認定しない。

BCCWJ

Diagram illustrating the BCCWJ dependency parse for the sentence: 世界各地で 作品を 展示するなど 名が 知られた 人たちばかり。 The root node is D, which branches into D and D. The left D branches into D and D, which then branch into 世界各地で and 作品を respectively. The right D branches into D and D, which then branch into 展示するなど and 名が respectively. The final D branches into 知られた and 人たちばかり。

CSJ では、名詞にかける。

CSJ

Diagram illustrating the CSJ dependency parse for the sentence: 世界各地で 作品を 展示するなど 名が 知られた 人たちばかり。 The root node is A2, which branches into D and D. The left D branches into D and D, which then branch into 世界各地で and 作品を respectively. The right D branches into D and D, which then branch into 展示するなど and 名が respectively. The final D branches into 知られた and 人たちばかり。

KC

Diagram illustrating the KC dependency parse for the sentence: 世界各地で 作品を 展示するなど 名が 知られた 人たちばかり。 The root node is A, which branches into D and D. The left D branches into D and D, which then branch into 世界各地で and 作品を respectively. The right D branches into D and D, which then branch into 展示するなど and 名が respectively. The final D branches into 知られた and 人たちばかり。

・「～など+する」は通常のかかり受けとする。

BCCWJ

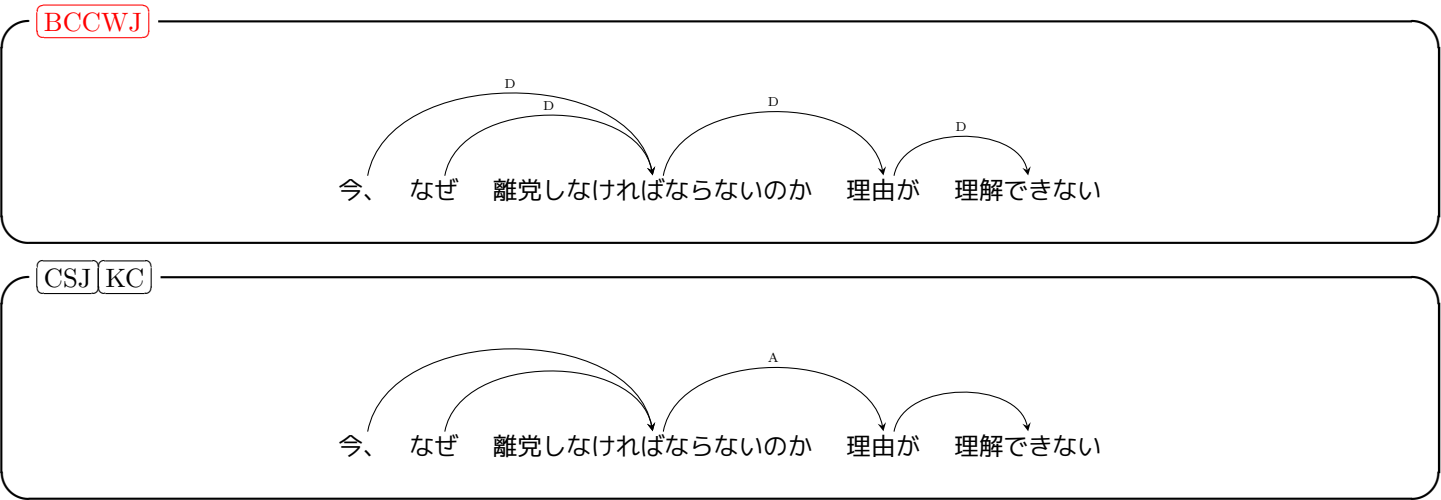
Diagram illustrating the BCCWJ dependency parse for the sentence: ドナーと 患者の DNA が 結び付くなど した ものを 電気泳動装置に かける. The root node is D, which branches into D and D. The left D branches into D and D, which then branch into ドナーと and 患者の respectively. The right D branches into D and D, which then branch into DNA が and 結び付くなど respectively. The final D branches into した ものを and 電気泳動装置に かける. A Parallel label is shown below the first two nodes.

CSJ KC

Diagram illustrating the CSJ KC dependency parse for the sentence: ドナーと 患者の DNA が 結び付くなど した ものを 電気泳動装置に かける. The root node is P, which branches into D and D. The left D branches into D and D, which then branch into ドナーと and 患者の respectively. The right D branches into D and D, which then branch into DNA が and 結び付くなど respectively. The final D branches into した ものを and 電気泳動装置に かける.

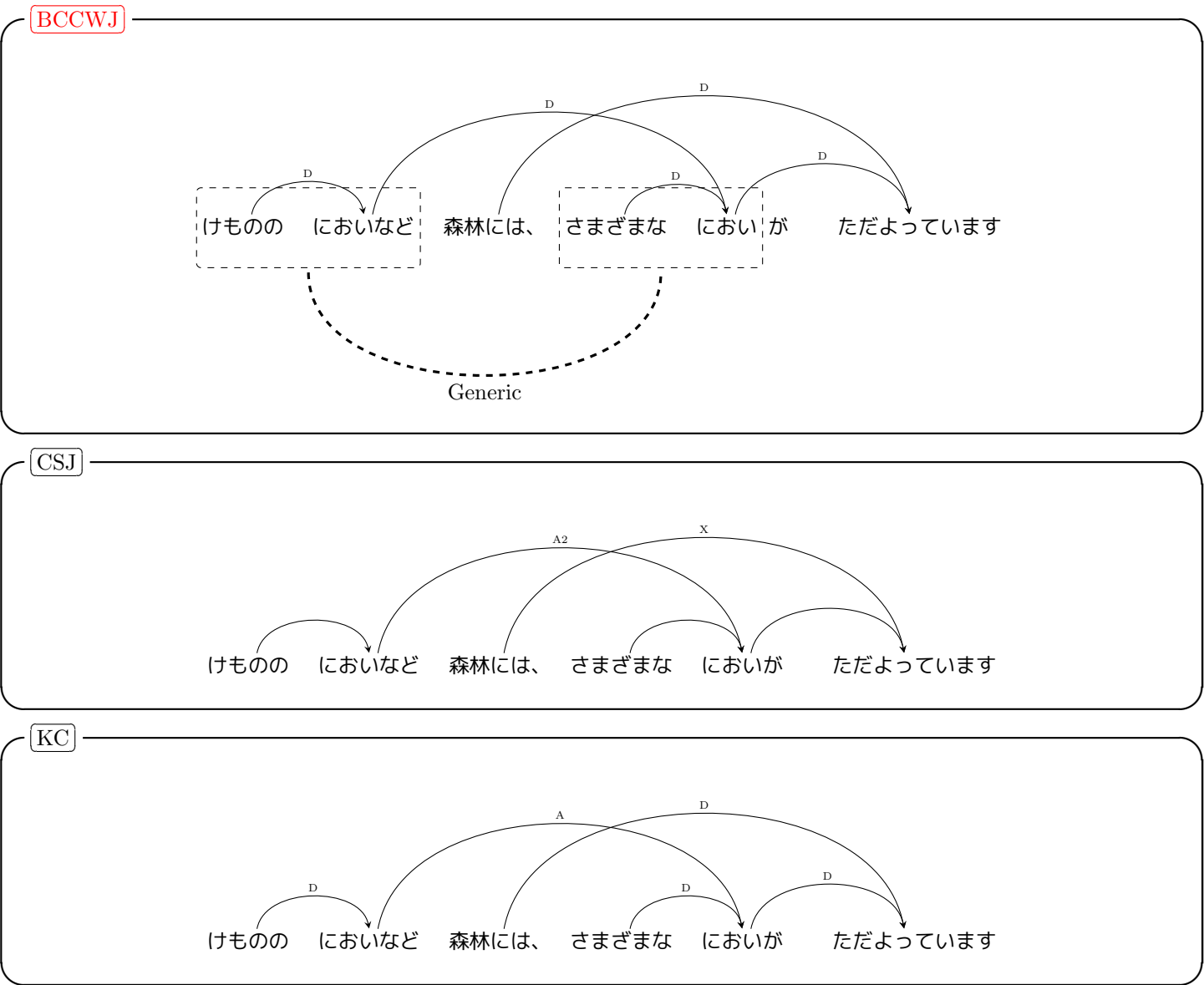
6.3.6 節とそれをまとめる名詞

BCCWJ では同格を認定しない。



6.3.7 同格と交差

- [BCCWJ] においては非交差条件を前提とせずにアノテーションする。
- [CSJ] においては非交差条件を破る場合には X タグを付与する。
- [KC]においては、同格関係の以下のような事例については、非交差条件を破るものをみとめるとする。



6.3.8 A の BC

[KC]において、「A の B（同格）C」のような表現で「A」のかかり先があいまいな場合、一般の名詞句「A の B の C」と同様に基本的には、「B」にかかる構造を優先する。

1. 近くにかかる例

- || 米国内の || 電波望遠鏡 || 十基 || (CSJ A2, BCCWJ Generic)
- || 村山内閣の || 防衛費 || ○. 八五五%増 || (CSJ A2, BCCWJ では同格としない)
- || 米国第二位の || ビールメーカー || 「～社」 ||
- || 未完の || オペラ || 「～」 ||
- || 自民党の || 総裁 || 「～」 ||
- || アルバムの || 中の || 一曲 || 「～」 || (BCCWJ では同格としない)
- || 香港の || 中立系紙 || 「～」 ||
- || 遊眠社の || 若手人気俳優、 || 羽場裕一ら ||
- || 健三郎さんの || 長男 || 光さん ||

2. 遠くにかかる例

- || 子供 || 二人の || 家族 || 四人 || (CSJ A2, BCCWJ では同格としない)
- || 自民、 || 社会 || さきがけの || 与党 || 三党 || (CSJ A2, BCCWJ Generic)
- || 同社の || 創立 || 30周年記念 || (BCCWJ では同格としない)
- || 一日付の || 機関紙 || 「赤旗」 ||
- || 俳優座劇場で || 公演の || 俳優座 || 「～」 || (BCCWJ では同格としない)
- || 同乗の || 妻 || 晴子さん || (BCCWJ では「同乗の」は「妻」にかける)



## 付録 A

## Q/A

### A.1 Version 0.2.0 から Version 0.3.0 への変更における Q/A

- Q: 「中学校を」だけで終わっているような中途半端な文の場合、「中学校を」は DUMMY にかけるしかないのですが、その場合は “F” ですか？

A: 国語研文単位相当タグ `<sentence>` タグには `type` として “quasi” 属性が定義されており、この事例に `<sentence>` が “quasi” 属性つきでラベルづけして文境界認定されていればラベル “Z” です。しかしながら、次の Q / A に書いてあるように、基本的には `<supersentence>` タグ相当で文境界を認定しなおすという作業を行っており、`<sentence>` タグ相当の文を超えてかかり先がどこかにあるので、かかり先があるところまで文を連結して対処します。そこまでしてもかかり先が真にならない場合はラベル “F” になります。

- Q: かかり受けラベルは、全て “D” でよいのではないか

A: まず、文境界相当のラベル “Z” について：既存の BCCWJ の文境界タグ `<sentence>` タグが底本となっているもとのデータ（紙や HTML タグなど）のレイアウトに基づき自動についており、文の入れ子などで問題が起きています。これを修正するために別のレベルでつけているもので、この情報を尊重して残すために “Z” で表記します。そのうえで、この作業は単文の連結（`<supersentence>` 相当）に重きを置いており、単文の内側の文末については粒度がそろっていない問題があり、これについては今回の作業内容に含めたいと思います。詳しくは 2013 年 3 月のコーパス日本語学ワークショップで発表いたしました。

次にラベル “F” ですが、感動詞など、ある程度自動で DUMMY ノードにかけることを行う際に、「自動で修正した」ことを明示するために導入しています。また、「文末は DUMMY ノードではなく、文の最右要素である」と主張する方がラベル “F” 相当のかかり受けを文の最右要素にかけるという変換を行うことを想定して、事後も残します。

最後にラベル “B” ですが、「かかり受けに用いる文節単位」と「国語研文節単位」が異なるということを明示するために導入します。基本的に国語研文節単位を分割することではなく、連結する方向にかかり受けに用いる文節単位を規定しますので、かかり受けのラベルとして表現しますが、実際のかかり受け解析器を構成する際には連結した 1 文節として扱うよう変換すればよいと考えます。尚、「国語研文節単位」の規定は尊重したうえで、変更させるという働きかけは行わないつもりです。

### A.2 Version 0.1 から Version 0.2.0 への変更における Q/A

- Q: セグメント・グループの表示がわかりにくい

A: メモ作成者の latex の技術の限界。表記方法がわかり次第修正する。尚、下線のみだと、下線どうしの境界がわからない。

- Q: 文節境界を正しく直さないのか。直すとして DVD 頒布版に反映しないのか。

A: 国語研文節単位は規程集にのっとしてわかち書き基準が決められており、その基準は尊重する。尊重したうえで、かかり受け関係付与に不都合があるものをラベル “B” により、「かかり受けアノテーション版文節境界」として別のものとして扱う。実際に解析器を構成する場合には、ラベル “B” 相当箇所は文節を連結した形で扱ってよい。

- Q: 文境界を正しく直さないのか。直すとして DVD 頒布版に反映しないのか。

A: 国語研文単位の基準は尊重する。尊重したうえで、かかり受け関係付与に不都合があるものをラベル “Z” により、「かかり受けアノテーション版文境界」として別のものとして扱う。実際に解析器を構成する場合には、ラベル “Z” 相当箇所を文境界と扱ってよいが、文が入れ子になっている点に注意すること。

- Q: 「顔文字」はパンクチュエーションの一種として直前の文節と一緒にするという取扱は検討したのか。

A: 検討していないが、パンクチュエーションが独立して出現した場合には顔文字と同様にかかり先なしとして扱うことは検討した。

- Q: 「英単語」などもかかり受けをつけるべきではないのか。

A: 英単語が格要素になっている場合には日本語として扱いかかり受け構造を付与するが、英文として出現した場合には “Foreign” セグメントとして切り出し、かかり受けアノテーションを行わない。必要と考えるのであれば、“Foreign” セグメントのみ抜き出して、別途かかり受けアノテーションを適切な基準に基づいて行えばよいと考える。

- Q: 同じ話が複数回出てくるのだが

A: それぞれデータの出自が異なっており、基準が安定したら適切に配置する。





## 付録 B

# 変更履歴

### B.1 Version 0.6.0 から Version 0.7.0 への変更履歴

- ChaKi の画面ダンプの差し替え
- 2012 年度作業疑問点の総括
- 2013 年度作業初期教示時の疑問点に対する回答

### B.2 Version 0.5.0 から Version 0.6.0 への変更履歴

- 並列構造・同格構造の表現から下線を削除（四角の範囲が元の下線相当になる）
- 2.7 節に文節の連結基準について追記。今後文節修正基準を追加する。
- 3.2.1 節に副詞のかかり先について、呼応を考慮することを追加した。
- 3.1 に従属節に対する判断基準を追記。今後節分類の言語テストを追加する。

### B.3 Version 0.4.0 から Version 0.5.0 への変更履歴

- 3.1 節に従属節に対する判断基準を追記。今後節分類の言語テストを追加する。
- 2.2.5 節に 2 形態素以上に分かれている接続詞相当句について言及した。
- 6.1.1 節に述語が 2 つ以上出現する場合の説明を追記。
- 2.7 節に文節の連結基準について追記。今後文節修正基準を追加する。

### B.4 Version 0.3.0 から Version 0.4.0 への変更履歴

- 2.10 節に前処理（自動処理）の概要について示した。

### B.5 Version 0.2.0 から Version 0.3.0 への変更履歴

- ラベル “F” が自動変換プログラムで修正したものであることを明示した。修正後もラベル “F” と区別することとする。
- 感動詞に対して自動変換プログラムでラベル”F” を付与し、1 単語で文節をなす場合には「ありがとう」をのぞいて DUMMY に自動でかけていることを記述した。
- 感動詞「ありがとう」について、後ろに「ございます」が後置する場合には、「ありがとう」をラベル “D” で「ございます」にかけることを明記。
- 記号・補助記号について、品詞細分類ごとに規定を精緻化した。
- 記号-一般、記号-文字、補助記号-一般のいずれかの品詞で 1 単語で文節をなす場合には、自動変換プログラムでラベル “F” を付与し DUMMY に自動でかけていることを記述した。
- 補助記号-括弧開で 1 単語で文節をなす場合には、自動変換プログラムでラベル “B” を付与し右隣接要素に自動でかけていることを記述した。
- 補助記号-括弧閉で 1 単語で文節をなす場合には、自動変換プログラムで左隣接要素ラベル”B” を付与し当該文節に自動でかけ、当該文節のラベル “DX” を付与し当該文節のかかり先が不定であることを明示することを記述した。
- Foreign セグメント内のかかり受け関係ラベルを “D” でかけるように変更した。これにより「かかり先なし」ではなくなるために「かかり先なし」の subsection の外へ移動した。
- Disfluency セグメントの要素で、かかり受け関係ラベルを “D” でかけるように変更し、かかり先は言い直したものに変更した。
- 格表示誤りについては、ChaKi.NET のかかり受けに対するコメント機能が動作しないために、通常のかかり受けとして扱うように変更した。

### B.6 Version 0.1 から Version 0.2.0 への変更履歴

- 文末の要素が DUMMY 要素に ラベル “F” でかかることを明示した。
- 文節内の部分要素にかけることを意味するラベル “S” の廃止。4.1 節ほか。
- 2.2.2 節において、独立した括弧をかかり受け関係ではなく、文節の結合（ラベル “B”）で対処することとした。
- 2.2.4 節において、独立した空白は、文頭でも、文中でも、かかり先なし（ラベル “F”）とした。文中の空白で、前後の文節が 1 つの文節をなす場合には、文節の結合（ラベル “B”）で対処することとした。
- かかり先なし（ラベル “F”） であり DUMMY ノードにかける矢印を文の下に移動。
- 2.5.6 節において、ラベル “S” を削除。
- その他誤記の修正。



# 参考文献

[1] 小西光, 小山田由紀, 浅原正幸, 柏野和佳子, 前川喜久雄. BCCWJ 係り受け関係アノテーション付与のための文境界再認定. 第3回コーパス日本語学ワークショップ発表論文集, 2013.

[2] 黒橋禎夫, 居倉由衣子, 坂口昌子. 形態素・構文タグ付きコーパス作成の作業基準 (version 1.8). Technical report, 京都大学, 2000.

[3] 野田尚史. はとが. 日本語文法セルフ・マスターシリーズ (1). くろしお出版, 1985.

[4] 内元清貴, 丸山岳彦, 高梨克也, 井佐原均. 『日本語話し言葉コーパス』における係り受け構造付与 (version 1.0). Technical report, 『日本語話し言葉コーパス』の解説文書.

[5] 南不二男. 現代日本語の構造. 大修館書店, 1974.

# 索引

Apposition グループ, 5  
Apposition セグメント, 5  
言い直し, 11  
言いよどみ, 11  
英単語, 10  
F ラベル, 2, 3, 7  
顔文字, 9  
かかり先なし, 3, 23  
格表示誤り, 15  
格要素, 7  
箇条書き, 10  
括弧, 8  
からまで, 34, 36  
感動詞, 7, 23  
漢文, 10  
外国語, 24  
記号, 8  
空白, 9  
グループ, 4  
交差, 2, 24  
呼应, 18  
古文, 10, 24  
主語, 17  
主語, 29  
主題, 17, 29  
主題・主語以外の格要素, 17  
主題・主語以外の格要素, 29  
Generic グループ, 5  
Generic セグメント, 5  
自動処理, 15  
従属節, 31  
従属節, 17  
セグメント, 4  
節境界, 22  
接続詞, 9, 12, 22, 31  
Z ラベル, 7  
DUMMY ノード, 2, 3  
テ形, 17, 34  
D ラベル, 2, 4  
Disfluency セグメント, 5  
では, 30  
電話番号, 9  
, 2, 21  
同格, 5, 13, 20  
同格:具体例と数詞, 14, 20  
同格:具体例と総称, 14, 20  
には, 30  
Parallel グループ, 4  
Parallel セグメント, 4  
B+ ラベル, 4  
B ラベル, 4  
ファイラー, 22  
Foreign セグメント, 5  
副詞, 18  
文境界修正, 14, 33  
文節境界修正, 4, 14  
文節境界修正, 21  
並列, 4  
並列構造, 11  
    3 並列, 12  
    述語並列, 12  
    述語並列の例外：判定詞, 12  
    接続詞, 12  
    部分並列, 12  
    複数の要素が左からかかる場合, 13  
補助記号, 8  
前処理, 15  
URL, 9  
呼びかけ, 23  
連体修飾, 7  
連用修飾, 7  
連用修飾, 18  
連用修飾, 17