# 基礎インターフェースガイド

この文書はコーパスのインターフェースを使用するための手引きである。

## 1 概要(Overview)ページ

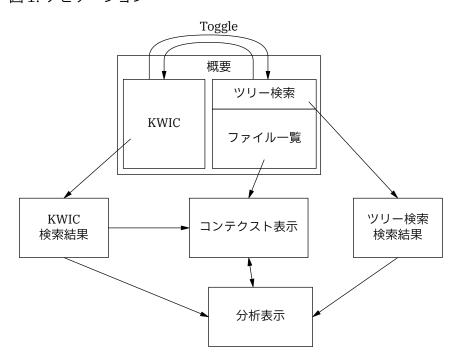
概要(Overview)ページには、2つの種類がある:

- ツリー検索・ファイル一覧(Tree search/file list)(デフォルト)
- KWIC 検索

これら 2 つを切り替えるには、 $\underline{\text{Toggle}}$ (ページ左上)をクリックする(図 2 と図 8 を参照)。

インターフェースの様々なページから概要(Overview)ページに戻るには、 $\underline{\text{Top}}$  を クリックする。 $\underline{\text{Top}}$  をクリックして表示されるのは、ツリー検索・ファイル一覧と  $\underline{\text{KWIC}}$  検索のうちの最後に選ばれていた方である。図 1 に見るように、すべてのページから コーパスデータの分析表示(Analysis view)ページおよびコンテクスト表示(Context view)ページへと飛べるようになっている。

図 1: ナビゲーション



# 2 ツリー検索・ファイル一覧(Tree search/file list)ページ

ツリー検索・ファイル一覧(Tree search/file list)ページは図2のように構成される。 コーパス中のすべてのファイル数、ツリー数、語数も示される。

#### 図 2: ツリー検索・ファイル一覧(Tree search/file list)ページ

tree word

Toggle	<u>tree</u>   <u>source</u>   .	[ <u>F</u> 1	ront page/ <u>フロントページ</u> ] <u>He</u>	<u>lp</u>
Tree search	n interface			
Files:				
TGrep2 sear	rch expression:			
Search	basic 🗸 🗘	Toggle tag set	TGrep2 <u>search guide/検索ガイ l</u>	ピ

40 files; 20,083 trees; 80,772 words

	Filename		count	content description
1	diet_kaigiroku-1	75	1081	parliamentary language : Kokkai kaigiroku
2	diet_kaigiroku-2	61	873	parliamentary language : Kokkai kaigiroku
3	diet_kaigiroku-3	159	3930	parliamentary language : Kokkai kaigiroku
4	diet_kaigiroku-4	37	722	parliamentary language : Kokkai kaigiroku
<u>5</u>	diet_kaigiroku-5	111	2869	parliamentary language : Kokkai kaigiroku

このページから辿ることのできるページは次のとおりである:

- ファイル番号をクリックすると、そのファイルのコンテクスト表示(Context view)ページに切り替わる。
- Toggle をクリックすると、ツリー検索・ファイル一覧(Tree search/file list)ペ ージと KWIC 検索ページが切り替わる。
- tree や source など、「」」で区切られたオプションのいずれかをクリックすると、 分析表示(Analysis view)ページで表示される分析方法が変わる。現在選択さ れているオプションは背景がグレーのものである(分析方法の変更は、分析表示 (Analysis view) ページでも行うことができる)。
- Front page/フロントページ は、ウェブサイトのトップページへのリンクである。
- Help は利用可能なドキュメント置き場へのリンクである。
- search guide/検索ガイド は、検索式の作り方の説明へのリンクである。
- Toggle tag set をクリックすると、アノテーションに用いられているタグの表示・ 非表示が切り替わる。

## 2.1 ファイルの選択

「Files:」の横のテキストボックスにファイル番号を入力することで、検索対象のファイ ルを選択することができる。例えば、3pは一覧の3番目のファイルを選択する。6,12p は6番から12番までのファイルを選択する。3p;6,12pは3番目のファイルと6番か ら12番までのファイルを選択する。このとき、最後の pを省略してもよい。例えば、 3p;6,12pを 3p;6,12と入力することもできる。

図3に見るように、ファイル名をスラッシュ(/)で囲んでもよい。複数のファイルを 指定する場合は、'\|'でファイル名を区切ることに注意されたい。

#### 図 3: ファイル名による検索対象ファイルの指定

<u>Toggle</u>	<u>tree</u>	sourc	<u>ce</u>	[ <u>Front page</u> / <u>フロントページ</u> ] <u>Help</u>	
Tree search interface					
Files: /di	Files: /diet_kaigiroku-12\ diet_kaigiroku-17/				
Tregex search expression:					
Search basic 〉  う Toggle tag set Tregex search guide/検索ガイド					
383 trees; 5,600 words					
Filen	ame	tree ount	word count	content description	
12 diet_l	kaigiroku-12	87	1228	parliamentary language : Kokkai kaigiroku	
17 diet_l	kaigiroku-17	296	4372	parliamentary language: Kokkai kaigiroku	

テキストボックスの内容に基づき検索対象となるファイルが制限されると、検索結 果で表示されるツリーと語数は、指定されたファイルのもののみになる。ファイル選択 の情報は、テキストボックスの入力内容が変更または削除されるまで持続される。

## 2.2 TGrep2/Tregex 検索式

ツリー検索のための検索式には、TGrep2(Rohde 2005) または Tregex(Levy and Andrew 2006)を使う。これらは異なる検索言語ではあるが、密接に関連している。 TGrep2 と Tregex のどちらを使うかは、検索対象ファイルを選択するか否かによって 決まる。

- 検索対象ファイルが選択されていない場合(つまり、図2のように、ファイル選 択のためのテキストボックスが空白である場合)は、TGrep2 を使い、コーパス**全** 体を検索する。
- 検索対象ファイルが選択されている場合(つまり、図3のように、ファイル選択 のためのテキストボックスに入力がなされている場合)は、Tregex を使い、選択 されたファイルのみを検索する。

TGrep2 と Tregex はどちらも、Penn Treebank と共に配布された検索言語、TGrep (Pito 1994) とよく似ている。

TGrep2 あるいは Tregex を使った検索式は、主としてツリーにおけるノードにマッ チする表現と、ノードとノードの関係または関係の否定の表現から成る。ノード表現に は、単純な文字列、選言された文字列、あるいは正規表現を使うことができる。複雑な 検索式は、ノード表現(マスターノード)とそれに続く関係表現から成る。

検索式の作り方の詳細については、図2または図3の search guide/検索ガイドを クリックされたい。

### 2.3 Search ボタンを押す

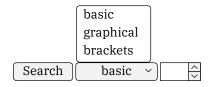
検索式の入力後、Search ボタンをクリックすると検索が開始される。検索対象ファイルが選択されていない状態で、検索式を入力せずに Search ボタンをクリックしても何も起こらない。検索対象ファイルを選択した状態で、検索式を入力せずに Search ボタンをクリックすると、画面に表示されるファイルのリストが選択されたもののみになり、ツリー数と語数もアップデートされる。

### 2.4 ツリー検索の結果表示の選択

デフォルトのツリー検索の結果表示の形式は、tree yield である。tree yield とは、ツリーの終端ノードを1行(文字列)に抽出したデータのことである。

画面の basic と表示されたボタンをクリックすると、図 4 のような選択メニューが現れる。メニューから graphical を選択すると、検索結果がグラフィカルツリーとして表示される。メニューから brackets を選択すると、検索結果が括弧付きツリーとして表示される。

図 4: 表示形式の選択



ツリー検索の結果は、検索式に一致したツリーをコーパス順に並べ、任意の箇所を出発点とし、そこから 50 ツリー(表示方法を graphical にした場合)、あるいは 500 ツリー(表示方法を basic または bracketed にした場合)を取り出して表示される。通常、出発点となる箇所は自動的に選ばれる(検索結果が 50 あるいは 500 以下であれば、すべての例が表示される)。

検索式入力ボックスの下にあるボックス(図 4)に数字を入力すると、出発点となる箇所を制御することができる。例えば、70 と入力すると、70 番目から 50 あるいは500 のツリーが表示される。

# 3 ツリー検索の検索結果ページ

ツリー検索の検索結果は、次の3つの方法のいずれかによって表示される。

- basic を選択すると、検索式に一致したツリーのうち、最大 500 がコーパス順に tree yield として表示される。その際、検索式のマスターノードと一致した範囲の 文字列が強調表示される。
- graphical を選択すると、検索式に一致したツリーのうち、最大 50 がコーパス順にグラフィカルツリーとして表示される。その際、検索式のマスターノードと一致したノードが強調表示される。
- brackets を選択すると、検索式に一致したツリーのうち、最大 500 がコーパス順に括弧付きツリーと tree yield の両方で表示される。その際、検索式のマスターノードと一致した、括弧付きツリー中のノードの前に、2 つのアンダースコア('\_\_')が加えられ、強調表示される。

tree yield 表示では、ゼロ要素(例えば、ゼロ代名詞や関係節中のトレース)は通常は表示されない。ただし、検索式のマスターノードの下に置かれたゼロ要素は表示される。

図 5 は、検索ファイルを 3p として(つまり、diet\_kaigiroku-3 として)、検索式 を /REL/、表示形式を basic とした場合の検索結果である(検索ファイルの指定については、図 2 および図 3 を参照)。

#### 図 5: basic 表示

$\underline{\text{Top}}$					
Tregex search pattern:					
/REL/					
Search basic >					
The search returned 2 hits. 1 text was searched (2110 words [1 text]; frequency: 9.48 instances per ten thousand words).					
Download all results comma-separated values ∨					
See analysis tree Y					
□ <u>1</u> 31_diet_kaigiroku <u>この際、新たに就任されました</u> 政務次官を御紹介申し上げます。					
□ 2 34_diet_kaigiroku-3 通産的変か宣友拝命しました 山下でございます					

検索結果ページ上部の検索式入力ボックスから検索を再実行あるいは修正することができる。

検索式入力ボックスの下には、検索式にヒットした数とヒット含むファイルの数が示される。検索対象のファイルを指定した場合は、頻度情報も示される。

検索結果が 50 以上(graphical を選択した場合)あるいは 500 以上(basic または brackets を選択した場合)ある場合、検索を再実行するか、あるいは表示形式切り替え ボタン横のボックスに数字(N)を入力することで、その時点で表示された検索結果以外のツリーを表示されることができる。

basic または brackets 表示では、それぞれのツリーにチェックボックスとヒット番号が与えられる。ヒット番号はツリーの分析表示(Analysis view)へとリンクされている。graphical 表示では、それぞれのツリーの ID が分析表示(Analysis view)へとリンクされている。

#### 3.1 検索結果のダウンロード

Download all results ボタンを(図 5)クリックすると、すべての検索結果がダウンロードされる。検索式にヒットした箇所がひとつのツリー中に複数個あった場合は、それと同じ数のツリーがダウンロードファイルに含まれる。ダウンロードファイル中の検索結果はコーパス順に並ぶ。よって、検索を再実行してもダウンロードファイルの内容は変わらない。ダウンロードファイルの形式は2種類ある。

- CSV 形式 (comma-separated values)
- ・ 括弧付きツリー形式 (bracketed trees)

CSV 形式のファイルは表計算ソフトで開くことができる。ファイルは1行4列になっており、最初の列はID、2列目は検索式にマッチしたノードよりも前の内容、3列目は検索式にマッチしたノードよりも後の内容であ

る。ゼロ要素 (例えば、ゼロ代名詞や関係節中のトレース) は検索式にマッチした内容 である場合にのみ表示される。

括弧付きツリー形式のファイルは CorpusSearch プログラム (Randall 2009) と互換性のある形式である。この形式では、それぞれのツリーが"wrapper"をもつ。"wrapper"とは、ツリーの内容と ID ノードを囲む、ラベルなしの括弧のペアのことである。各ツリーの ID には、@ と数字が加えられる。この数字は検索式にマッチしたノードの位置を深さ優先順に数えた数字である。さらに、検索式にマッチしたノードは、2 つのアンダースコア('\_\_')によって示される。

#### 3.2 分析モードの選択

図 5 の tree ボタンをクリックすると、図 6 のような選択メニューが現れる。

図 6: 分析モードの選択

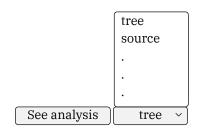


図7はsourceを選択したところである。

図 7: Source モードの選択



分析モードを選択した後に、See analysis ボタンをクリックすると、チェックしたすべてのツリーが選択した分析モードで表示される。

# 4 KWIC 検索ページ

KWIC 検索ページは、図 8 のような構成になっている。テキスト入力ボックスに文字列を入力し、Search ボタンをクリックすると、検索が開始される。この検索方法では、検索対象ファイルを選択することはできない。Toggle をクリックするとツリー検索・ファイル一覧(Tree search/file list)ページに切り替わる。

図 8: KWIC 検索ページ

Toggle	<u>tree</u>   <u>source</u>	[Front page/ <u>フロントページ</u> ] <u>Help</u>
KWIC inte	rface	
Search	corpus order v	

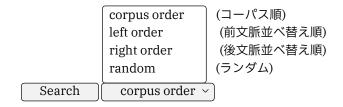
### 4.1 文字列検索表現

KWIC 検索ではテキストボックスに単純な語/文字を入力する。検索対象となるのは tree yields、つまりツリーの終端ノードを 1 行に抽出したデータから句点を排除したものである。また、ひとつのツリーは通常、ひとつの文に相当する。

## 4.2 検索結果の表示順

図8の corpus order をクリックすると、図9のように、検索結果の表示順を変更するための選択メニューが現れる。

#### 図 9: 検索結果の表示順の選択



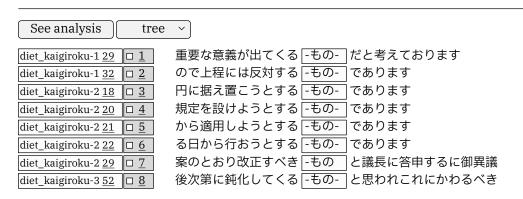
## 5 KWIC 検索の結果表示

図 10 は、デフォルトのコーパス順( $corpus\ order$ )のままで、「もの」という語を検索した結果である。

図 10: コーパス順(corpus order)の KWIC 検索結果

Top

Search pattern:  $\bullet o$ 



検索結果は伝統的な KWIC(Key Word In Context)として示される。つまり、検索表現が中央に一列に並び、左右に検索表現の前後の表現が示される。これにより、検索表現が使われる文脈を見ることができる。

図 10 ではコーパスの中で発生する順番で結果が表示されている。コーパスの中で発生する順番とは、ファイル名およびファイル内のツリーの通し番号を昇順にソートした結果に基づく順番のことである。

### 5.1 KWIC の検索結果における行の内容

図 10 では、検索結果が KWIC で表示されているほか、各行に次のような情報が示されている:

- 当該の例を含むファイル名
- ファイル中における例の通し番号。これはさらに、当該の例をハイライトしたコンテクスト表示(Context view)ページへとリンクされている(6 節を参照)。
- コーパス順の番号。これはさらに、当該の例の分析表示(Analysis view)ページ へとリンクされている(7 節を参照)。
- See analysis ボタンをクリックして表示させる例を選択するためのチェックボックス

各行に示される情報の配置は、図11のようになる。

図 11: KWIC 検索の結果表示における行の内容

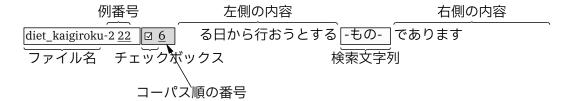


図 11 では、チェックボックスにチェックが入っている。この状態でページ下部の See analysis ボタンをクリックすると(その前に分析モードを選択しておく必要がある (3.2 節を参照)、チェックした例の分析が表示される。複数の例をチェックし、それらすべての分析を表示させることもできる。

#### 5.2 前文脈並べ替え順(Left order)

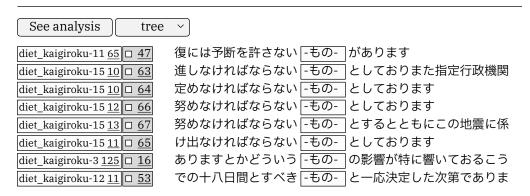
結果表示の表示順にはコーパス順(corpus order)以外の選択肢もある(4.2 節を参照)。 検索を一度行った後に表示順を変更するには、ブラウザの機能を使って KWIC 検索ペー ジへ戻り、同じ検索文字列で他の表示順を選択し、検索を再実行する。

前文脈並べ替え順(left order)は、図 12 に見るように、検索文字列の前に現れた文字のコード順に検索結果を並べ替えたものである。

#### 図 12: 前文脈並べ替え順(left order) の結果表示

Top

Search pattern: もの



前文脈並べ替え順(left order)の結果表示は、検索文字列とその前の表現を含むより大きなコロケーションを捉えるために有効である。

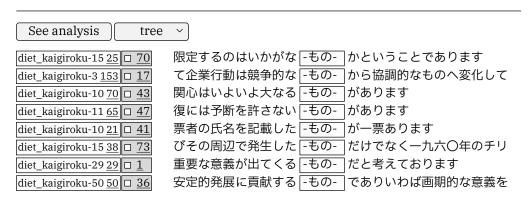
## 5.3 後文脈並べ替え順(right order)の結果表示

後文脈並べ替え順(right order)は、図 13 に見るように、検索文字列の後に現れた文字のコード順に検索結果を並べ替えたものである。

図 13: 後文脈並べ替え順(right order)の結果表示

Top

Search pattern: もの



後文脈並べ替え順(right order)の結果表示は、検索表現とその後の表現を含むより大きなコロケーションを捉えるために有効である。

### 5.4 ランダム順 (random order) の結果表示

ランダム順(random order)の結果表示は、検索結果をランダムな順に並べたものである。検索結果が多く、データセット全体からの印象を必要とする場合に役立つであろう。

## 6 コンテクスト表示 (Context view)

コンテクスト表示(Context view)は3つの方法で見ることができる:

- ツリー検索・ファイル一覧(Tree search/file list)ページで、ファイル番号をクリックする(図2を参照)
- KWIC 検索の結果表示において例番号をクリックする(図 11 を参照)
- 分析表示(Analysis view)ページ上部の Context をクリックする(図 16 および図 17 を参照)

2番目と3番目の方法でコンテクスト表示(Context view)を表示させると、当該の例がハイライトされ、その例のチェックボックスにチェックが入った状態になる(チェックを外すこともできる)。このとき、ページの一番上ではなく、当該の例がすぐに見えるように表示される。

図14に見るように、コンテクスト表示(Context view)では、ページ上部にファイルのメタデータ(タイトル、日付、ソースなど)が示され、また、ファイル中のすべての文が表示される。それぞれの文には番号とチェックボックスが付いており、文番号をクリックすると、その文の分析表示(Analysis view)に切り替わる。分析表示(Analysis view)のモードは、直前に見た形式になる(もし選択が変更されていなければ、tree (ツリー)モードになる)。

#### 図 14: コンテクスト表示(Context view)ページ



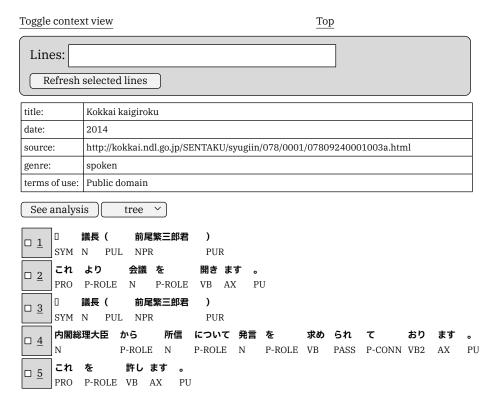
複数のチェックボックスにチェックを入れ、複数の文を選択することができる。選択解除も自由に行うことができる。複数の文を選択する際に、それらが隣接している必要はない。文を選択し、ページの下部の See analysis ボタンをクリックすると、選択された文の分析表示(Analysis view)に移動する。また、分析モードはプルダウンメニュー(図 6 を参照)から選択することもできる。

「Lines:」の横のテキストボックスに文番号を入力し、Refresh selected lines ボタンをクリックすると、その文をハイライトすることができる。

文番号の入力方法は、ツリー検索における検索対象ファイルの入力方法(2.1 節)と類似する。3p はそのファイルの中の3 番目の文を、6,12p は、6 番目から12 番目までの文を、3p;6,12p は、6 番目から12 番目の文および3 番目の文を表す。p は省略することができる。

図14の <u>Top</u>をクリックすると、概要(Overview)ページへ戻る。また、<u>Toggle word analysis</u>をクリックすると、語の品詞情報の表示/非表示が切り替わる。図 15 は品詞情報を表示したところである。

図 15: 品詞情報付きのコンテクスト表示(Context view)



# 7 分析表示 (Analysis view)

分析表示(Analysis view)を見るにはいくつかの方法がある:

- 1. グラフィカルツリーとして表示された検索結果で ID をクリックする
- 2. 括弧付きツリーとして表示された検索結果で、ヒット番号をクリックする(図 5を参照)
- 3. 括弧付きツリーとして表示された検索結果で、チェックボックスから例を選択し、 See analysis ボタンをクリックする(図 5 を参照)
- 4. KWIC 検索の結果表示において、コーパス順の番号をクリックする(図 11 を参照)
- 5. KWIC 検索の結果表示で、チェックボックスから例を選択し、See analysis ボタンをクリックする(図 10、図 12、図 13 を参照)
- 6. コンテクスト表示(Context view)で文番号をクリックする(図 14 および図 15を参照)

7. コンテクスト表示(Context view)で、チェックボックスから例を選択し、See analysis ボタンをクリックする(図 14 と図 15 を参照)

1、2、4、6の方法は、1つの例に対する分析を表示する。そのときのモードは、直前に見たモードと同じものである(もし何もしなければ、tree(ツリー)モードになる)。3、5、7の方法では、複数の例に対する分析表示を一度に表示する(図 6 を参照)。

## 7.1 Tree (ツリー) モード:ツリーを1つ表示させた場合

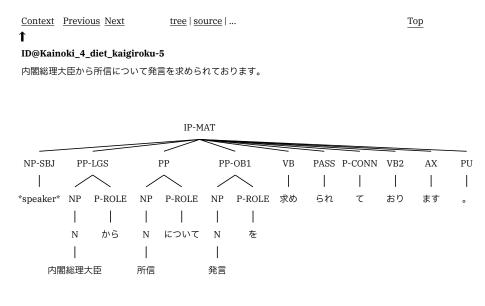
デフォルトの分析表示(Analysis view)では、その解析情報がグラフィカルツリーとして表示される。一例として、(1) の文の分析を考えてみよう。これは、コーパスのdiet\_kaigiroku-5 というファイルの 4 番目の文であると仮定する。

(1) 内閣総理大臣から所信について発言を求められております。

図 16 は (1) の解析情報をグラフィカルツリーとして表示したところである。

図 16: 単一のツリーの tree (ツリー) モードでの表示

Previous Next



前述したように、ページの最下部から分析モードを変更することができる(source は図 16 での利用可能なもうひとつの分析モードである。閲覧しているコーパスによっては、他にも選択肢がある)。

図 16 おけるグラフィカルツリーの上には、そのツリーの tree yield(終端ノードを抽出したもの)が表示されている。その上には、そのツリーの識別情報、つまり、Kainoki コーパスの中の、diet\_kaigiroku-5 というファイルの中の、4 番目の文であることが表示されている。

識別情報のすぐ上には、大きな上向き矢印がある。この上向きの矢印をクリックすると、直前のツリー(つまり、diet\_kaigiroku-5の中の3番目のツリー)の分析が現れる(7.2節を参照)。また、グラフィカルツリーのすぐ下には大きな下向き矢印がある。この下向き矢印をクリックすると、直後のツリー(つまり、diet\_kaigiroku-5の中の5番目のツリー)の分析が現れる。

上向きの矢印の上には、いくつかのリンクがある。<u>Context</u> はファイル(ここでは diet\_kaigiroku-5)のコンテクスト表示(Context view)へのリンクであり、グラフ

ィカルツリーとして表示されていたツリー(4番目のツリー)の yield が全体のコンテクストと共にハイライトされる。

Previous をクリックすると、現在表示されているツリーの前のツリーが表示される。Next をクリックすると、現在表示されているツリーの次のツリーが表示される。Previous と Next は、ファイルの最初あるいは最後まで繰り返すことができる。これらのリンクは表示されるツリーを切り替える。したがって、上向き矢印、下向き矢印とは使い方が異なる。

Top をクリックすると、概要(Overview)ページに移動する。

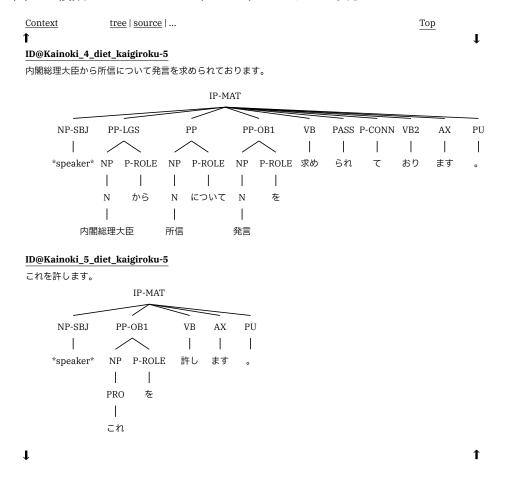
## 7.2 Tree (ツリー) モード:複数のツリーを表示させた場合

前節ではツリーを 1 つ表示させた場合の Tree(ツリー)モードについて見てきた。次に、 $diet_kaigiroku-5$  の 4 番目と 5 番目の文である、(2)について考えてみる。

(2) 内閣総理大臣から所信について発言を求められております。これを許します。

図 17 では、(2) の 2 つの文の両方の解析情報がグラフィカルツリーとして表示されている。

図 17: 複数のツリーの tree (ツリー) モードでの表示



複数のツリーを表示させたときも、図 17 のように、直前あるいは直後のツリーを、 上向き矢印あるいは下向き矢印をクリックして表示させることができる。複数のツリー を表示させている場合は、最初のツリーの右上にある下向き矢印をクリックするとその ツリーが表示されなくなる。また、最後のツリーの右下にある上向き矢印をクリックするとそのツリーが表示されなくなる。

左上の上向き矢印と右上の下向き矢印の上には、概要(Overview)ページに戻るためのリンク Top と、表示されているツリーをハイライトさせたコンテクスト表示 (Context view) を開くためのリンク Context がある。

表示されているツリーの上には、そのツリーの識別情報が表示される。この識別情報は、Kainoki コーパスの中の、diet\_kaigiroku-5 というファイルの中の、4番目と5番目の文であることを意味する。識別情報をクリックすると、そのツリーのみの分析表示に切り替わる。