Corpus Interface Guide

This document is an introduction to using the corpus interface.

1 Overview page

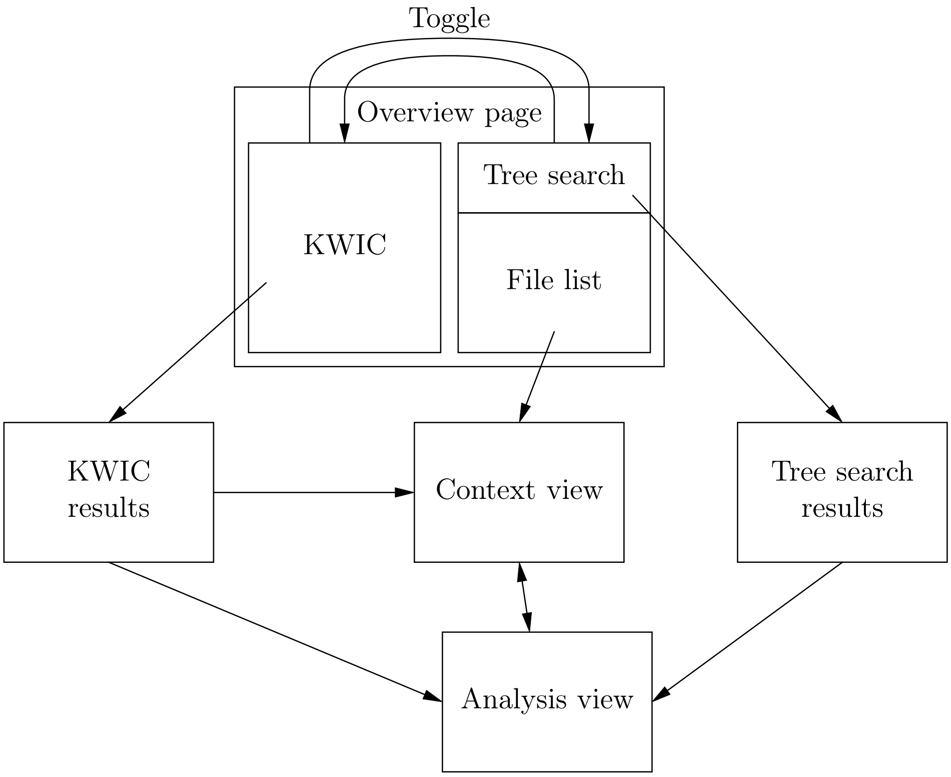
The overview page has two different states:

- a tree search/file list state, which is the default
- a keyword in context (KWIC) state

It is possible to toggle between these two states by clicking the Toggle link (see Figures 2 and 8 below).

It is always possible to return to the overview page by clicking the Top link of any of the other pages. The state returned to will correspond to how the overview page was last seen. Other possible navigations through the interface are illustrated in Figure 1. Notably, all pages lead towards reaching analysis and context views of the corpus data.

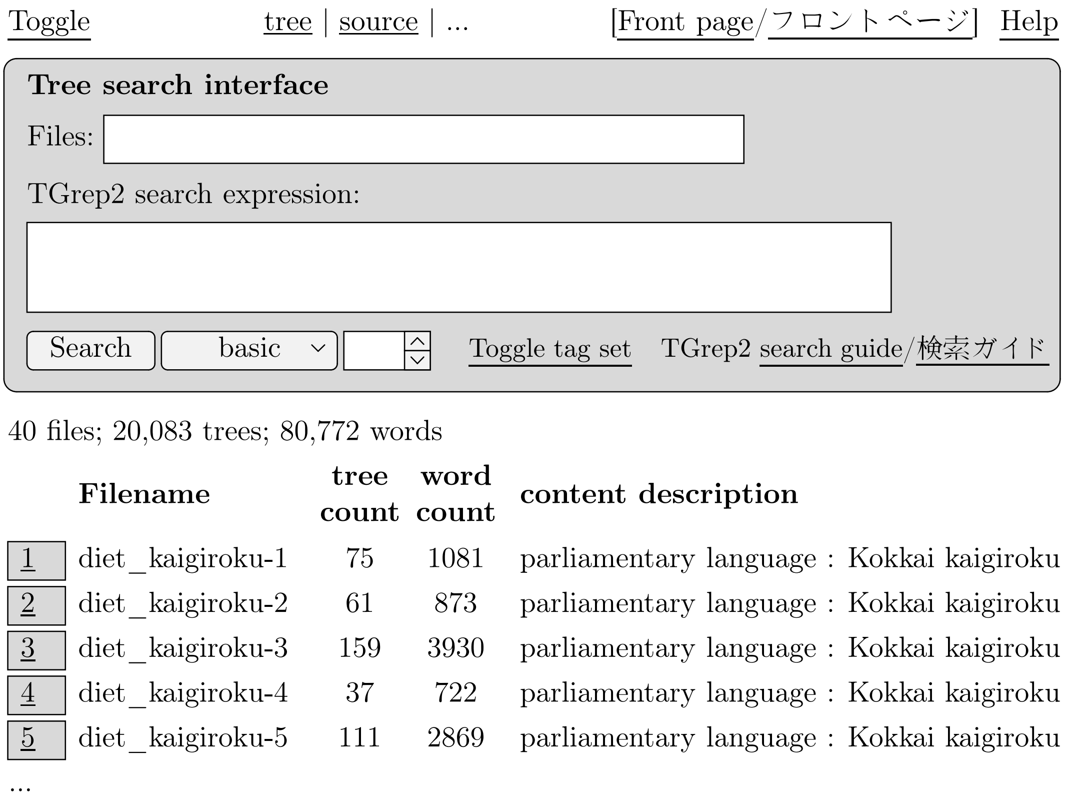
Figure 1: Navigation



2 Tree search/file list overview page

Figure 2 shows the tree search/file list state of the overview page.

Figure 2: Tree search/file list overview page


Toggle
tree | source | ...
[Front page/フロントページ] Help
Tree search interface
Files:
TGrep2 search expression:
Search
basic
Toggle tag set
TGrep2 search guide/検索ガイド

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

	Filename	tree count	word 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
5	diet_kaigiroku-5	111	2869	parliamentary language : Kokkai kaigiroku

...

This includes information about the overall numbers of files, trees and words for all the listed data.

Here is information about the various available links seen in Figure 2:

- The instances of 1, 2, etc. are links to go to a context view page for the referenced file.

- Clicking Toggle changes to a different state of the overview page.

- Clicking on tree, or source, or any of the other options that follow separated by '|' changes how analysis is shown when an analysis page is reached. The currently selected analysis option is indicated with a grey text background. (Note: The analysis view option can also be changed from an analysis view page.)

- Clicking Front page navigates to the online front page for the corpus giving overview details of the corpus.

- Clicking Help changes the page to show the available documentation.

- Clicking search guide opens an explanation for the TGrep2 query language used to search the corpus.

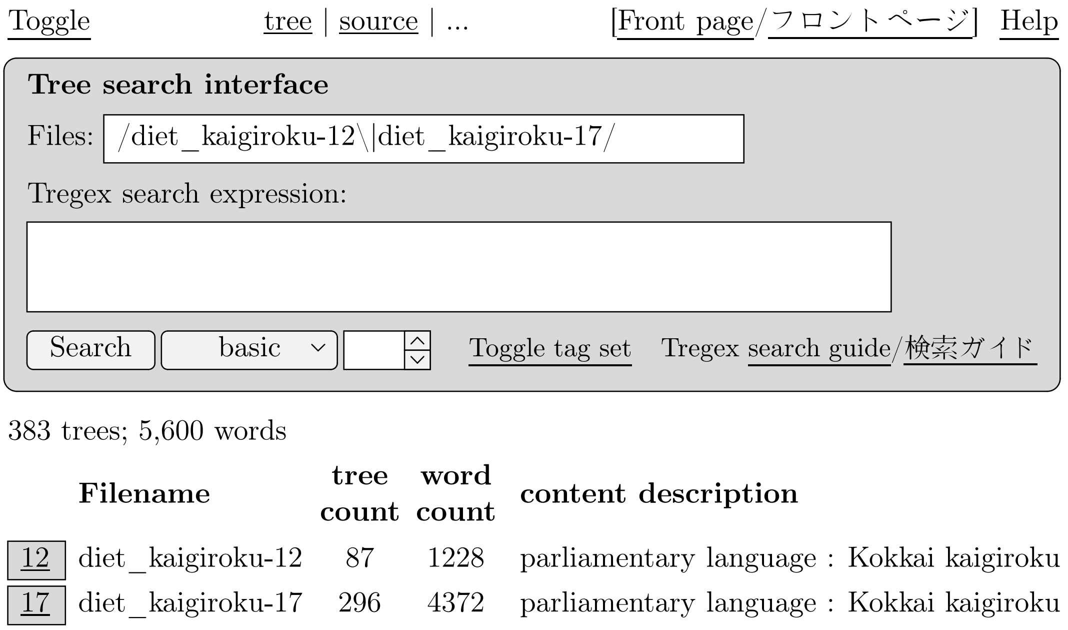
- Clicking Toggle tag set reveals/hides the tag set for the annotation when clicked.

2.1 Selecting files

Files can be selected with line addressing entered into the Files input text box: `3p` selects the third file, while `6,12p` selects the range of files from the sixth to the twelfth file (inclusively), and `3p;6,12p` selects both the third file and the range of files from the sixth to the twelfth file. The final `p` of a selection made with line addressing can be dropped, while other `p` instances are required. For example, `3p;6,12p` can be entered as `3p;6,12`.

Files can also be selected with regular expressions indicated by surrounding slashes (`/`) and used to match the names of corpus files, as Figure 3 demonstrates. Note: The '`\|`' character combination signals disjunction within the regular expression.

Figure 3: Tree search/file list overview page with file selection


Toggle
tree | source | ...
[Front page/フロントページ]
Help
Tree search interface
Tregex search expression:
Search
Toggle tag set
Tregex search guide/検索ガイド

Files: /diet_kaigiroku-12\|diet_kaigiroku-17/

basic

383 trees; 5,600 words

	Filename	tree count	word count	content description
12	diet_kaigiroku-12	87	1228	parliamentary language : Kokkai kaigiroku
17	diet_kaigiroku-17	296	4372	parliamentary language : Kokkai kaigiroku

In Figure 3, the listing of files is restricted to files selected by the content of the Files input text box. Also, the tree and word counts are for the selected files. Furthermore, should a search be made, the search applies only to the selected files. The information about selected files persists until the content of the Files field is either changed or deleted.

2.2 Making TGrep2 or Tregex search queries

Search queries over trees are made with two different but closely related search tools: TGrep2 (Rohde 2005) and Tregex (Levy and Andrew 2006). The search tool that is used for a given search depends on what is being searched:

- If there has been no selection of files (that is, the Files field of the overview page is empty, like in Figure 2), then TGrep2 is used to search **all** of the corpus.

- If files have been selected (that is, the Files field of the overview page has content, like in Figure 3), then Tregex is used to search only the **selected files**.

The query languages of TGrep2 and Tregex closely resemble the query language TGrep (Pito 1994), which was the original tree-matching program distributed with the Penn Treebank.

TGrep2 and Tregex queries are expressed as patterns that mainly consist of expressions to match nodes and relationships defining links or negated links to other nodes. Nodes of searched trees are matched either with simple character strings, or OR'd character strings, or regular expressions. A complex node expression consists of a node expression (the **master node**) which is followed by relationships.

For a full explanation of these query languages, click the search guide link seen in Figures 2 and 3. Note that Figures 2 and 3 also have a Toggle tag set link, which reveals/hides the tag set for the annotation when clicked.

2.3 Pressing the 'Search' button

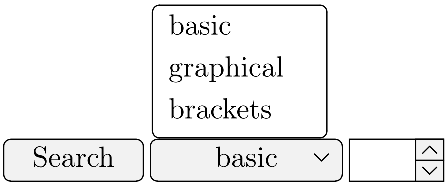
Once a search expression is entered, clicking the 'Search' button triggers the search. If the 'Search' button is pressed without there being a search expression, there is either (a) no change to the displayed page, or (b) if the file selection information of the Files input text box has changed then the page will show the new selection of files and also update the counts for trees and words.

2.4 Tree search type selection

A tree yield is the extraction of the terminal nodes of trees into single lines (character strings) of data. By default, tree search results are displayed as tree yields.

A click on the 'basic' button seen in Figures 2 and 3 brings up the selection menu seen in Figure 4. With the selection of 'graphical', tree search results are displayed as graphical trees. With the selection of 'brackets', tree search results are shown as bracketed trees.

Figure 4: Tree search type selection



As section 3 will describe, the results of a tree search are shown in corpus order, with up to 500 results displayed with the basic yield format or as bracketed trees, or with up to 50 results displayed as graphical trees. Each shown result is associated with a number that reflects its place in an overall ordering following corpus order of all results (shown and not shown).

Where in the overall corpus order the slice of 500 or 50 consecutive results comes from depends on the starting point adopted for the returned results. If the starting point is not specified, it is decided randomly when the number of results exceeds the number that can be displayed.

You can specify the starting point by entering a number in the number field seen as the blank field in Figure 4 next to the search type selection button. For example, if you enter the number 70, then 500 or 50 search results will be shown starting from the 70th result in corpus order onwards. This way, when repeating a search, you are able to go to a predicable selection from the overall results of the search.

3 Tree search results page

Tree search results are displayed in either of three ways:

- as a listing of the yield from **up to 500** matched trees, in corpus order, each with a highlighted span corresponding to what was matched by the master node of the query

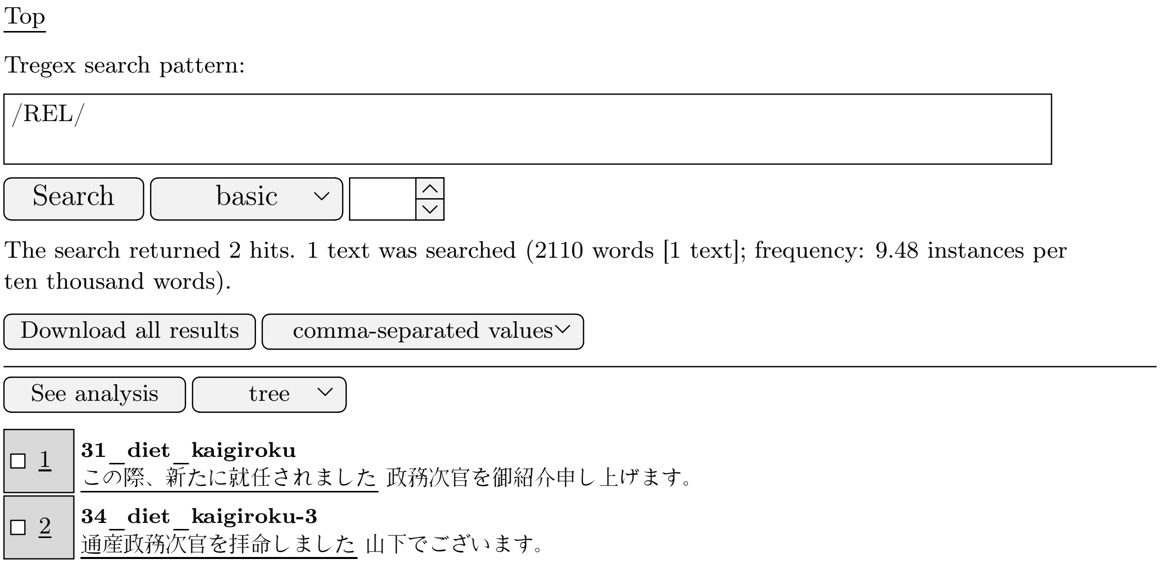
- as a listing of **up to 50** graphical trees, in corpus order, each with a highlighted node corresponding to what was matched by the master node of the query

- as a listing of **up to 500** bracketed trees together with tree yields, in corpus order, each with a highlighted node preceded by two underscore characters (‘__’) corresponding to what was matched by the master node of the query

With the display of results as tree yields, zero elements (e.g., zero pronouns and relative clause traces) are typically not shown, but will appear highlighted if they happen to be all that is matched as the master node.

Supposing the contents of the Files input text box (seen in Figures 2 and 3 above) is `3p` restricting the searched files to only the third file of the corpus, `diet_kaigiroku-3`, and supposing the contents of the search expression entry box is `/REL/` to match all trees with a node that contains the `REL` substring, then the returned basic tree search results page will look like Figure 5.

Figure 5: Basic tree search results page



Tregex search pattern:

/REL/

The search returned 2 hits. 1 text was searched (2110 words [1 text]; frequency: 9.48 instances per ten thousand words).

□ 1 **31_diet_kaigiroku**
この際、新たに就任されました 政務次官を御紹介申し上げます。

□ 2 **34_diet_kaigiroku-3**
通産政務次官を拝命しました 山下でございます。

There is a search box at the top of the results page which makes it possible to re-run/revise the search expression.

The search reports the number of hits and the number of files to which those hits belong. If the search is made with a limited range of selected files, then you will also see word frequency information.

If there are more than 500 (for basic/bracketed trees) or 50 (for graphical trees) results, you can see different results by re-running the search after changing the number (`N`) in the number field next to the search type selection button. After the search is re-run, results for up to 500 or 50 trees from the `N`-th tree onwards in corpus order will be be shown.

With basic or brackets search results, each returned entry is given a check box and a hit number. The hit number serves as a clickable link to the analysis view page for the given example. With graphical search results, each graphical display of a tree is shown with its ID in the corpus. The ID serves as a clickable link for reaching an analysis view for the example.

3.1 Download all search results

You can download all search results by clicking the ‘Download all results’ button seen in Figure 5. The download will give as many entries as there are hits reported, with a single hit per entry. Results are listed in corpus order, so a re-run of a download will give you the same download. There is a pull-down-selector for choosing between two methods for obtaining results:

- as comma-separated values
- as bracketed trees

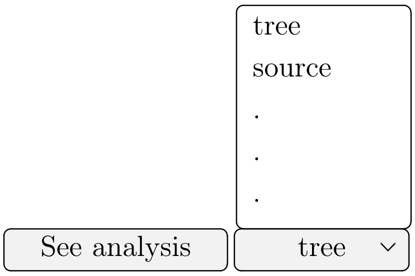
The comma-separated values format can be opened by a spreadsheet program. Each entry consists of a row with four columns. The first column is the ID for the example. The second column is the yield of content before the master node hit. The third column is the yield of the master node hit. The fourth column is the yield of content after the master node hit. Zero elements (e.g., zero pronouns and relative clause traces) only appear if they happen to be all that is matched as the master node.

The trees of a bracketed download are formatted to be compatible with the CorpusSearch program (Randall 2009). Under CorpusSearch format, every tree has a ‘wrapper’. A wrapper is a pair of unlabelled parentheses surrounding the tree content together with an ID node. An extension consisting of the ‘@’ character and a number is added to the ID node of each tree. Following depth-first order, this number picks out the node that matched as the master node. In addition, the label of the picked out node is also prefixed with two underscore characters (‘__’).

3.2 Analysis mode selection

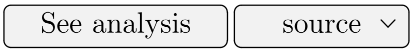
A click on the ‘tree’ button seen in Figure 5 brings up the selection menu seen in Figure 6.

Figure 6: Analysis mode selection



Moving the mouse to highlight ‘source’ and left clicking selects ‘source’ mode, resulting in Figure 7:

Figure 7: Source analysis selection

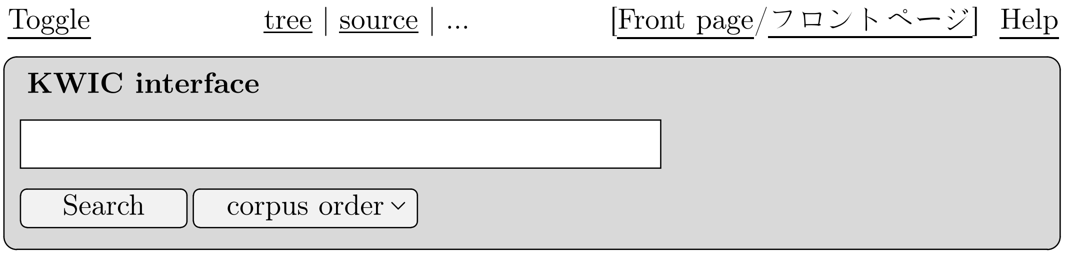


The choice of ‘source’ has consequences for how the analysis page will be shown. Pressing the ‘See analysis’ button will open the analysis view page with displayed analysis for all the selected search results. Search results are selected as a consequence of having their check box marked. Thus, with ‘tree’ selected, the analysis page reached will show the tree analysis for the selected search results.

4 KWIC page

Figure 8 shows the keyword in context (KWIC) state of the overview page. This gives a text entry box into which string search expressions can be entered. Clicking the ‘Search’ button performs the search over the entire corpus. With string search it is not possible to place restrictions on the files that are searched. There is also a Toggle link, which will continue a cycle through the available states of the overview page.

Figure 8: The KWIC page



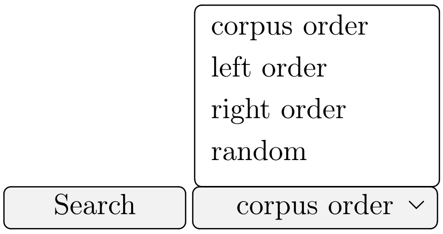
4.1 String search expressions

Expressions entered into the text box of Figure 8 need to be simple word/character based searches, including the space character for separating words to be matched. String searches are case-insensitive. The underlying searched data consists of **tree yields**, that is, extractions of the terminal nodes of trees into single lines (character strings) of data. A tree typically corresponds to the content of a single sentence. All punctuation is removed from the searched data.

4.2 Ordering of search results

Clicking ‘corpus order’ of Figure 8 brings up the selection menu in Figure 9 to change the ordering in which search results will be shown. Consequences of the differing orderings will be seen in section 5.

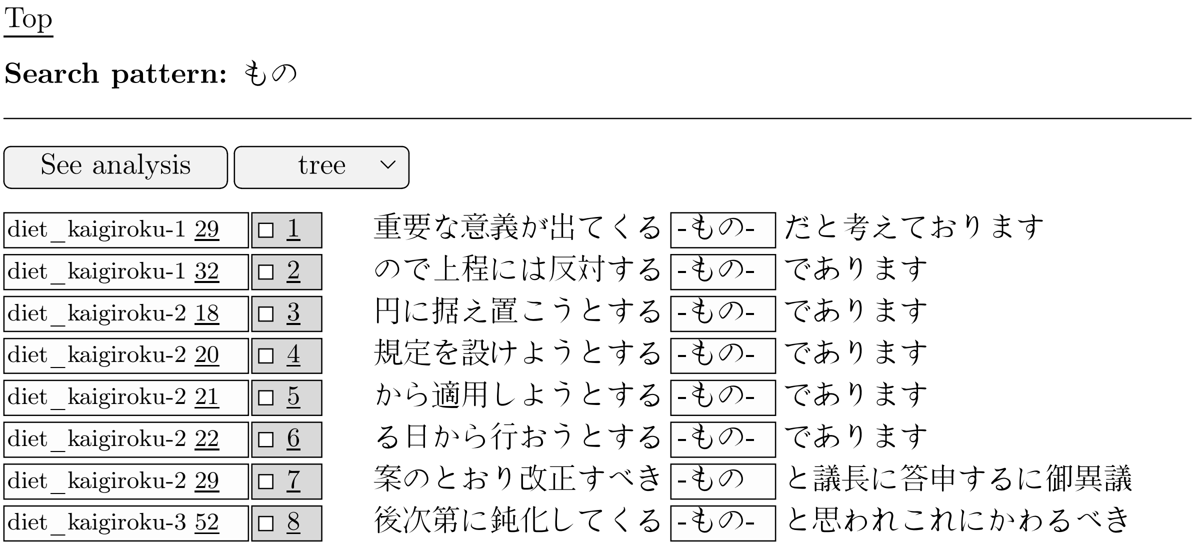
Figure 9: Ordering of search results selection



5 KWIC results page

This section describes the KWIC results page. This page is reached after a search string is entered into the KWIC page and the ‘Search’ button is clicked. For example, Figure 10 shows a KWIC results page returned from a search expression that comprises the word ‘もの’ with the default ‘corpus order’ unchanged.

Figure 10: The KWIC results page with corpus order



Search pattern: もの

diet_kaigiroku-1 29	□ 1	重要な意義が出てくる	-もの-	だと考えております
diet_kaigiroku-1 32	□ 2	ので上程には反対する	-もの-	であります
diet_kaigiroku-2 18	□ 3	円に据え置こうとする	-もの-	であります
diet_kaigiroku-2 20	□ 4	規定を設けようとする	-もの-	であります
diet_kaigiroku-2 21	□ 5	から適用しようとする	-もの-	であります
diet_kaigiroku-2 22	□ 6	る日から行おうとする	-もの-	であります
diet_kaigiroku-2 29	□ 7	案のとおり改正すべき	-もの	と議長に答申するに御異議
diet_kaigiroku-3 52	□ 8	後次第に鈍化してくる	-もの-	と思われこれにかわるべき

Figure 10 presents a traditional KWIC (Key Word In Context) concordance result. This lines up matches for the search expression in the middle of the display. The words that occur to the left and right of the matched content are also displayed to provide the context for the match.

The results of Figure 10 are said to be in ‘corpus order’ because results are shown following the order in which they occur within the corpus. This is reflected in the rising numbers seen in the names of corpus files, e.g., starting with `kaigiroku-1` and ending with `kaigiroku-3`. There is also the rising of numbers for examples that are matched from the same corpus file. For example, two examples are sourced from `kaigiroku-1` with numbers 29 and 32.

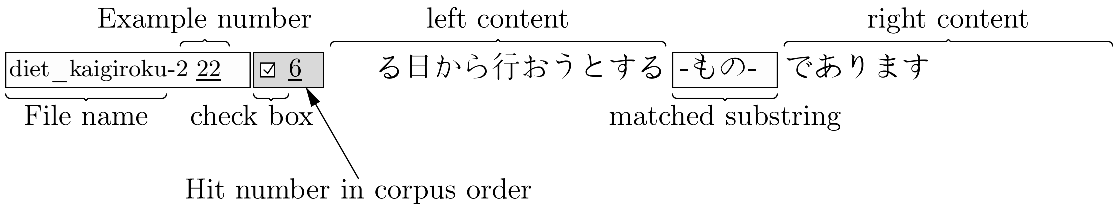
5.1 Row content of the KWIC page

We noted the KWIC presentation of Figure 10, where each row is a returned search result with centered match and context to the left and right. Additionally, each row provides:

- the name of the corpus file that contains the given example
- the number of the given example in the corpus file, which is also a link to the context view page centered on the given example that will be highlighted (see section 6)
- the hit number in corpus order, which is also a link to the analysis view page for the given example (see section 7)
- a check box for selection of the given example, for when the 'See analysis' button is clicked

Figure 11 indicates the location for the above functions within the row content.

Figure 11: Row content of the string results page



diet_kaigiroku-2 22 ☑ 6 る日から行おうとする -もの- であります

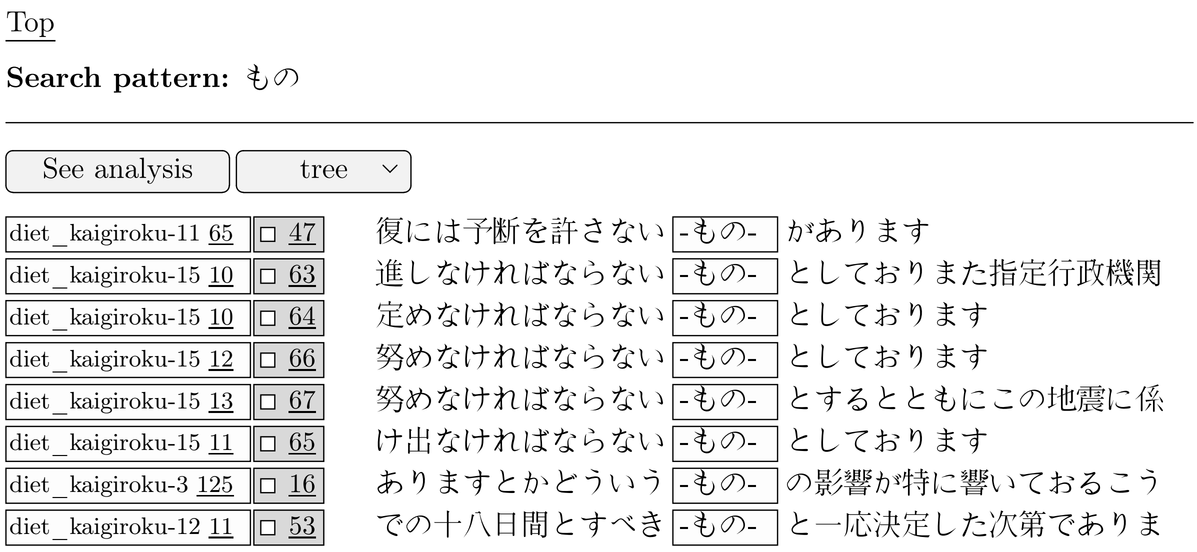
Note that Figure 11 also shows how row results are selected as a consequence of having their check box marked. Pressing the 'See analysis' button at the bottom of the page (possibly after selecting an analysis mode; see section 3.2) will open the analysis page with all the hit numbers of the selected input rows taken as the source for the analysis.

5.2 Left order

So far we have seen the string results page with corpus order, but the ordering depends on how the string search results page is accessed (see section 4.2). In this regard, your browser's back page functionality allows going back to the string search state of the overview page to then select an alternative ordering **while keeping the search expression**.

One alternative ordering is 'left order'. This sorts results alphabetically by the words that occur immediately to the left of the match, as demonstrated by Figure 12. Note that for a Japanese Kanji/Kana based corpus this option sorts results in character code order by the characters that occur immediately to the left of the match.

Figure 12: The string search results page with left order



Top

Search pattern: もの

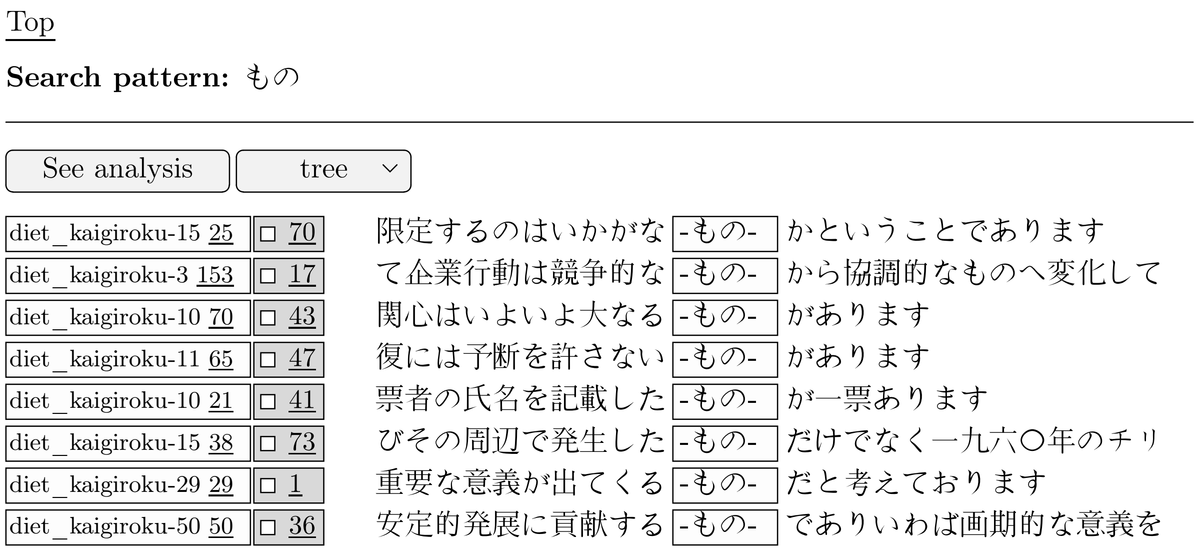
diet_kaigiroku-11 65	□ 47	復には予断を許さない	-もの-	があります
diet_kaigiroku-15 10	□ 63	進しなければならない	-もの-	としておりまた指定行政機関
diet_kaigiroku-15 10	□ 64	定めなければならない	-もの-	としております
diet_kaigiroku-15 12	□ 66	努めなければならない	-もの-	としております
diet_kaigiroku-15 13	□ 67	努めなければならない	-もの-	とするとともにこの地震に係
diet_kaigiroku-15 11	□ 65	け出なければならない	-もの-	としております
diet_kaigiroku-3 125	□ 16	ありますとかどういう	-もの-	の影響が特に響いておるこう
diet_kaigiroku-12 11	□ 53	での十八日間とすべき	-もの-	と一応決定した次第でありま

This ordering of results is useful for identifying larger collocations that involve the match and prior words.

5.3 Right order

Another alternative ordering is ‘right order’. This sorts alphabetically by words that occur immediately to the right of the match, as demonstrated by Figure 13. Note that for a Japanese Kanji/Kana based corpus this option sorts results in character code order by the characters that occur immediately to the right of the match.

Figure 13: The string search results page with right order



Top

Search pattern: もの

diet_kaigiroku-15 25	□ 70	限定するのはいかがな	-もの-	かということであります
diet_kaigiroku-3 153	□ 17	て企業行動は競争的な	-もの-	から協調的なものへ変化して
diet_kaigiroku-10 70	□ 43	関心はいよいよ大なる	-もの-	があります
diet_kaigiroku-11 65	□ 47	復には予断を許さない	-もの-	があります
diet_kaigiroku-10 21	□ 41	票者の氏名を記載した	-もの-	が一票あります
diet_kaigiroku-15 38	□ 73	びその周辺で発生した	-もの-	だけでなく一九六〇年のチリ
diet_kaigiroku-29 29	□ 1	重要な意義が出てくる	-もの-	だと考えております
diet_kaigiroku-50 50	□ 36	安定的発展に貢献する	-もの-	でありいわば画期的な意義を

This ordering of results is useful for identifying larger collocations that involve the match and following words.

5.4 Random order

Yet another alternative ordering is 'random order'. This ordering creates random permutations of the search results. This can be useful in cases where there are lots of results and you would like an impression from the overall dataset of the corpus.

6 Context view page

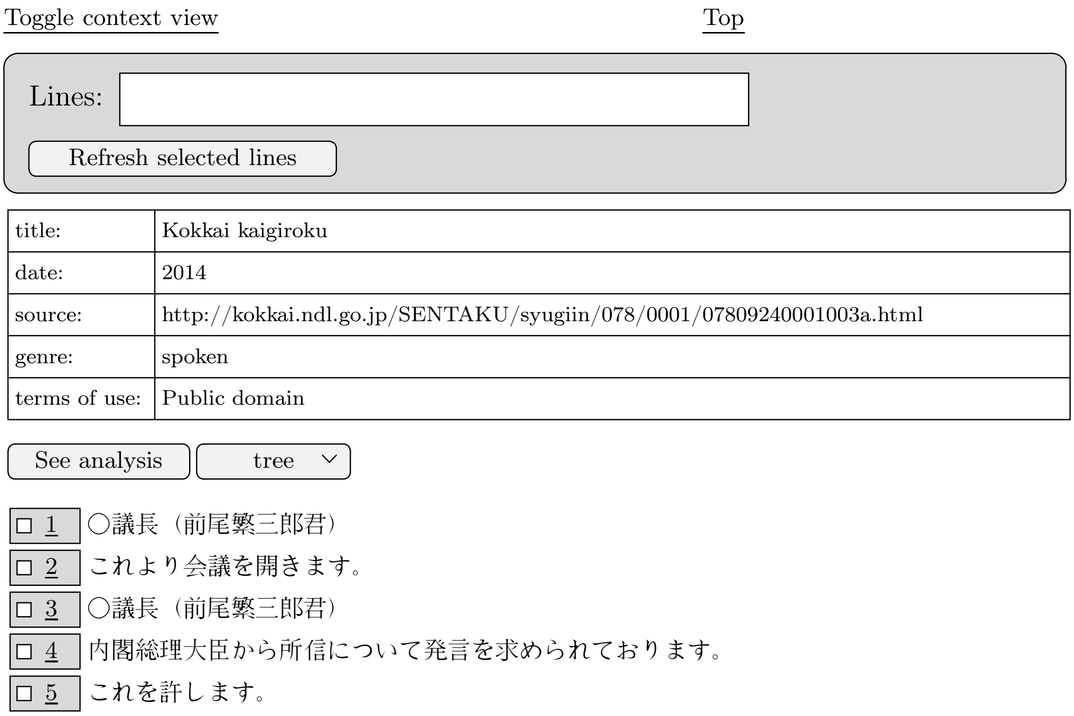
The context view page can be entered in three different ways:

- from clicking the number for a file of the tree search/file list overview page (see Figure 2 above)
- from clicking an example number of the string search results page (see Figure 11 above)
- from the Context link of an analysis page (see Figures 16 and 17 below)

With the second and third methods, the context view page is entered from links that originate from particular examples in the overall file of examples shown by the context view page. These particular examples are highlighted, and their associated check box is marked (but this can be unmarked with a selection click). Information about the highlighted items is also entered into the 'Lines:' text box as line addressing information. Furthermore, the context view page is entered from a position that can be other than the top of the page, so that the highlighted examples are visible.

The context view page seen in Figure 14 consists of a page that first gives metadata with entries to a table (title, date, source, etc.) for the file in question, and then lists — in file order — the yield for all the trees/sentences of the file, with each sentence preceded by a number and a check box. Clicking the number of a sentence takes you to an analysis view for the sentence. The 'mode' of the analysis view entered will be the same as how the analysis view was last seen. (If not changed, this will be the 'tree' mode).

Figure 14: Context view page



title:	Kokkai kaigiroku
date:	2014
source:	http://kokkai.ndl.go.jp/SENTAKU/syugiin/078/0001/07809240001003a.html
genre:	spoken
terms of use:	Public domain

□ 1 ○議長（前尾繁三郎君）
□ 2 これより会議を開きます。
□ 3 ○議長（前尾繁三郎君）
□ 4 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。
□ 5 これを許します。

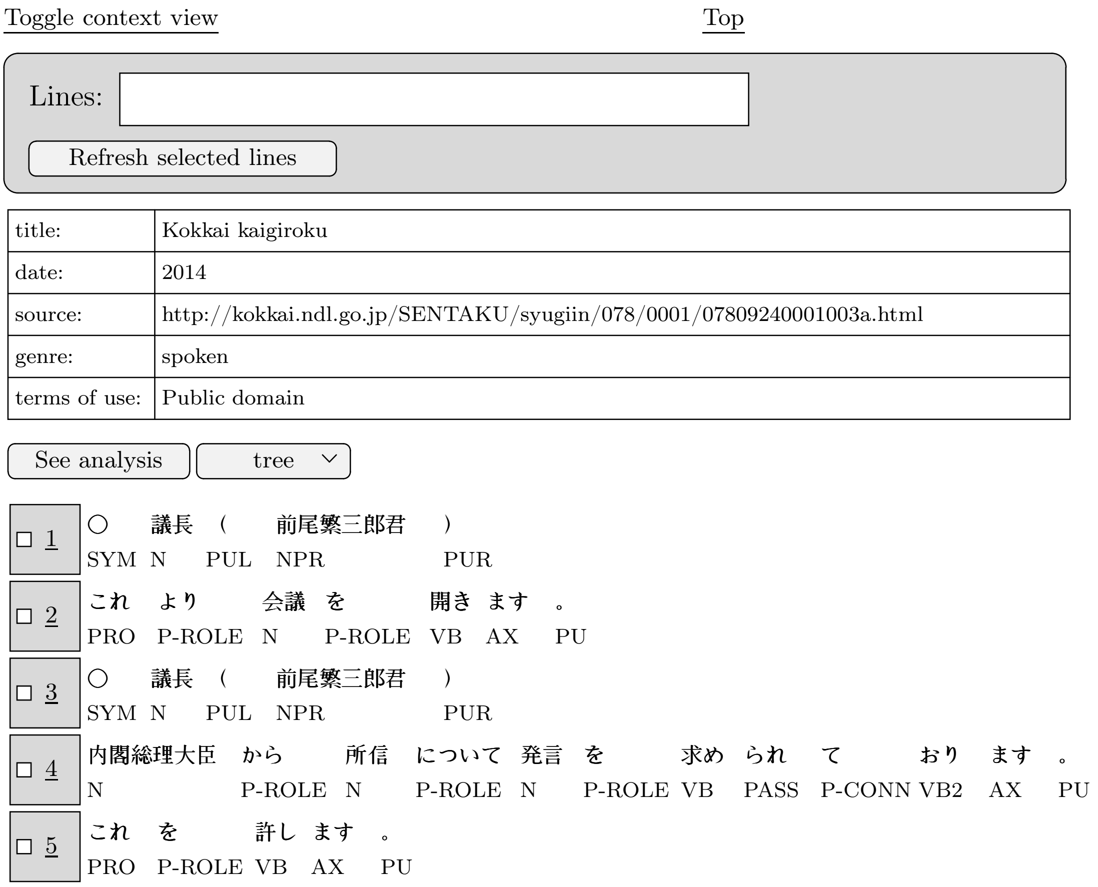
You can also select with the check boxes multiple sentences. Sentences can be freely selected (and de-selected), and need not be adjacent. Then, you can go to the analysis page for the selected sentences by clicking the 'See analysis' button. There is also a pull down menu (the same menu as shown in Figure 6 above) to select the 'mode' of analysis for the selected items.

You can also change the selection of sentences by editing the line addressing information of the 'Lines:' input text box and pressing the 'Refresh selected lines' button, which will reload the page with the selected lines of the line addressing freshly highlighted.

Line addressing for the selection of sentences/trees works like the line addressing used for selecting files described in section 2.1. Thus, `3p` selects the third sentence, while `6,12p` selects the range of sentences from the sixth to the twelfth sentence (inclusively), and `3p;6,12p` selects both the third sentence and the range of sentences from the sixth to the twelfth sentence. All instances of `p` can be dropped.

Note that Figure 14 also shows a Top link for returning to the overview page as it was last left. There is also a Toggle context view link, which when clicked will change the presentation of the context view page, for example, to reveal or hide word class information. Thus, Figure 15 shows the result of clicking Toggle context view in Figure 14, while Figure 14 is returned to with a click of Toggle context view in Figure 15.

Figure 15: Context view page with word class toggled



Toggle context view Top

Lines:

Refresh selected lines

title:	Kokkai kaigiroku
date:	2014
source:	http://kokkai.ndl.go.jp/SENTAKU/syugiin/078/0001/07809240001003a.html
genre:	spoken
terms of use:	Public domain

See analysis tree

□ 1 ○ 議長 （ 前尾繁三郎君 ）
SYM N PUL NPR PUR

□ 2 これ より 会議 を 開き ます 。
PRO P-ROLE N P-ROLE VB AX PU

□ 3 ○ 議長 （ 前尾繁三郎君 ）
SYM N PUL NPR PUR

□ 4 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 。
N P-ROLE N P-ROLE N P-ROLE VB PASS P-CONN VB2 AX PU

□ 5 これ を 許し ます 。
PRO P-ROLE VB AX PU

7 Analysis view page

The analysis view page can be entered in a number of different ways, notably:

1. from clicking an ID link of the graphical tree search results page
2. from clicking a hit number of a tree search results page (see Figure 5 above)
3. from the check box selection of examples of a tree search results page and subsequent clicking of the ‘See analysis’ button (see Figure 5 above)
4. from clicking a hit number of the string search results page (see Figure 11 above)
5. from the check box selection of examples of the string search results page and subsequent clicking of the ‘See analysis’ button (see Figures 10, 12, and 13 above)
6. from clicking an example number of the context view page (see Figures 14 and 15 above)
7. from the check box selection of examples of the context view page and subsequent clicking of the ‘See analysis’ button (see Figures 14 and 15 above)

Methods 1, 2, 4, and 6 open an analysis view for a single example, with the ‘mode’ being the same as how the analysis view was last seen. (If not changed, this will be the ‘tree’ mode). By contrast, with methods 3, 5, and 7, it is possible to open a view for multiple examples and to choose the 'mode' via the selection menu in Figure 6 above.

Once you have reached a particular analysis view, you can change to an alternative 'mode' by clicking one of the available links at the page bottom, for example, tree or source of Figure 16 below.

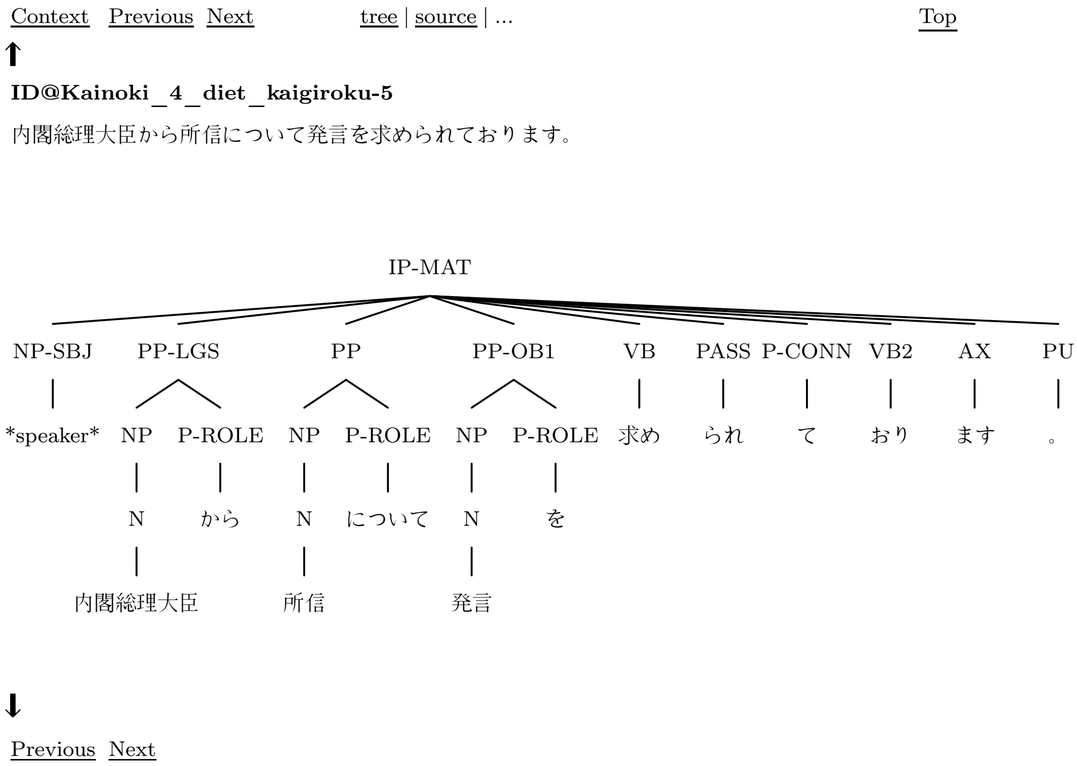
7.1 Tree mode, single tree case

The default of the analysis view page is to present the parse information as a graphical tree. As an example, let's consider analysis for the sentence in (1), which we will suppose occurs as the 4th example in a file named `diet_kaigiroku-5`.

(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。

Figure 16 illustrates the tree mode with its graphical presentation of the parse tree for (1).

Figure 16: Tree mode analysis view with single tree



Context Previous Next tree | source | ... Top

↑

ID@Kainoki_4_diet_kaigiroku-5

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。

↓

Previous Next

As with the overview page, in the centre at the page top, there is the means to change the analysis 'mode'. (Specifically, source is the other available option in Figure 16, but there may be further options depending on the corpus you are viewing.)

Above the graphical tree in Figure 16, there is a basic yield for the tree. There is identification information immediately above the yield that tells us we are seeing analysis for the 4th example of the corpus file named `diet_kaigiroku-5`.

Immediately above the identification information on the left edge of the page, there is a large up arrow. Clicking this up arrow brings the immediately prior example (so, example 3 of `diet_kaigiroku-5`) into the analysis view, to thereby have a view with the analysis of multiple trees (see section 7.2). Related to this ability to widen the analysis, there is a large down arrow immediately under the graphical tree on the left edge of the page. Clicking this down arrow brings the immediately following example (so, example 5 of `diet_kaigiroku-5`) into the analysis view.

Above the large leftside up arrow, there is a series of links. The Context link opens the context view page for the corpus file (here `diet_kaigiroku-5`) with the display moved to show the highlighted yield for the current tree (the 4th tree) and its surrounding context.

A Previous link appears when there is a previous tree, which when clicked changes the page content to an analysis view of the immediately preceding tree. Similarly, a Next link appears when there is a following tree, which when clicked changes the page content to an analysis view of the immediately following tree. When present, these Previous and Next links are also repeated at the very top or bottom of the page. Use of these links differs from the use of the large up and down arrows, since for the Previous and Next links the current tree is removed from the resulting analysis view.

Finally, there is a Top link for going to or returning to the overview page.

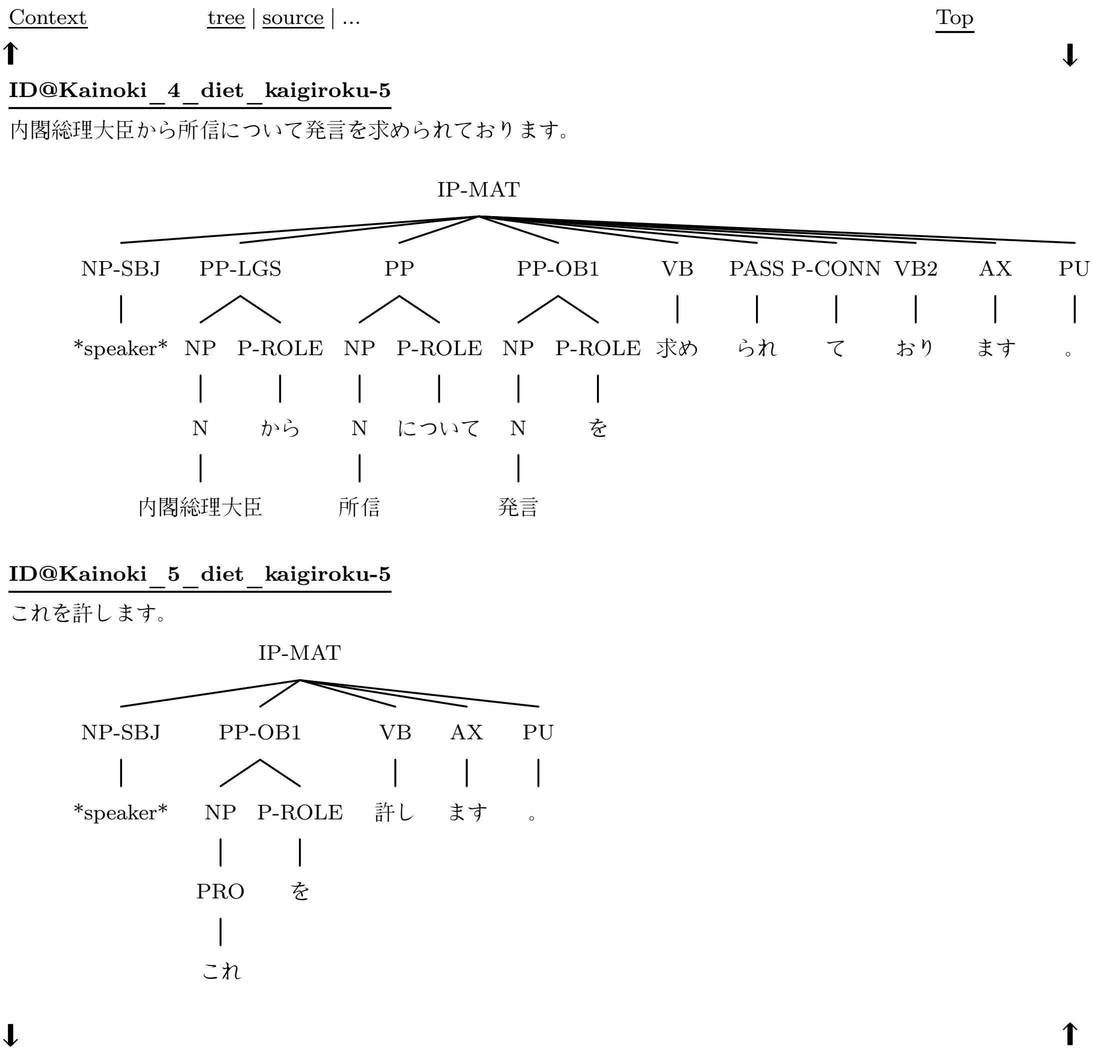
7.2 Tree mode, multiple tree case

So far we have seen the tree mode with a single tree. Let's now consider the sentences of (2), which we will suppose occur as the 4th and 5th examples in `diet_kaigiroku-5`.

(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。

Figure 17 illustrates the tree mode with its graphical presentation of parse trees for both sentences of (2).

Figure 17: Tree mode analysis view with multiple trees



Context tree | source | ... Top

ID@Kainoki_4_diet_kaigiroku-5

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。

ID@Kainoki_5_diet_kaigiroku-5

これを許します。

As with the single tree case in Figure 16, the multiple tree case in Figure 17 allows for the extension of the analysis presentation to include immediately preceding or immediately following trees by clicking respectively (when available) the leftside up arrow, or the leftside down arrow. For the case of Figure 17, where more than one tree is shown, it is possible to remove the first tree from view with a rightside down arrow, or remove the last tree from view with a rightside up arrow.

Above the large leftside up arrow and the large rightside down arrow, there is a Top link for returning to the overview page, and a Context link for opening the context view page with the display moved to show the highlighted yield for the current trees and their surrounding contexts.

Finally, above each tree yield there is identification information for the shown tree. Here, the identification information tells us that we are seeing analysis for the 4th and 5th examples of the corpus file named `diet_kaigiroku-5`. Furthermore this identification information comes underlined, with each underlined ID serving as a link for changing the analysis view to that isolated example.