

統語・意味解析コーパス（NPCMJ）チュートリアル

セッション3

NPCMJ Development Interfaces の基本操作

アラステア・バトラー、長崎郁

2021.3.13

NPCMJ Development Interfaces

- 新たに公開されたインターフェース
- データの閲覧と検索ができる。
 - 文字列検索
 - Tregex (Levy and Andrew 2006) を使ったツリー検索
- 正式公開前のデータも収録（現在、437 ファイル、896146 語、74776 ツリー）
 - 正式公開データは 338 ファイル、764266 語、67018 ツリー
 - 正式公開データについても、解析ミスの修正を行っている
 - 不定期にデータが更新されるので注意

1

セッション3の内容

1. ファイル一覧
2. 文字列検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果の保存
6. インターフェースに関するドキュメント

2

セッション3の内容

1. ファイル一覧
2. 文字列検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果の保存
6. インターフェースに関するドキュメント

3

ファイル一覧：入口

NPCMJ Development Interfaces を開く



Kainoki

Links to interfaces for viewing and searching under-development annotation

Contemporary Japanese

- [Kainoki](#) — the under-development NINJAL Paroed Corpus of Modern Japanese (NPCMJ)
- [PLENPCMJ](#) — Predicate-Argument Structure Thesaurus (PAT) meets NPCMJ
- [Kainoki](#) — the under-development NPCMJ with changed morphological base

Child language Japanese

- [Sawano](#) — the under-development child language corpus of the NPCMJ project with [WAKACHIKO](#) v8.1 morphological b

Tuguru Narratives

- [Matsumi](#) — the under-development Tuguru Narratives corpus (for more information about the [Tuguru dialect](#))

Old Japanese

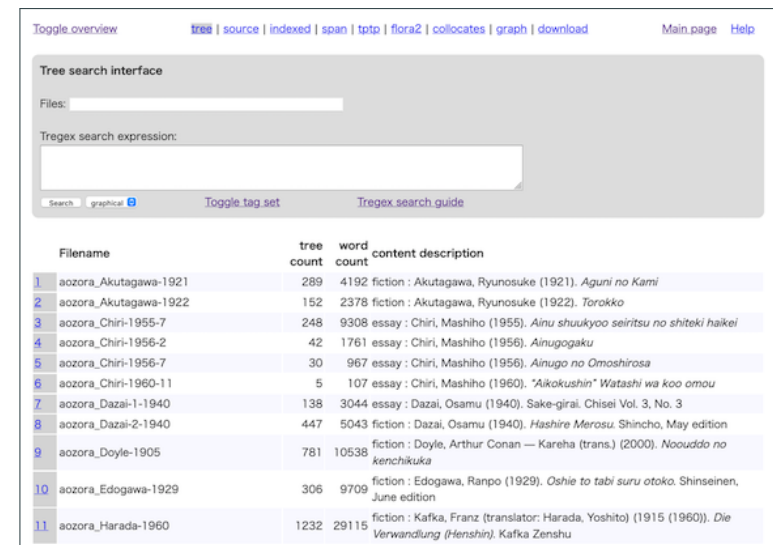
- [Ono](#) — the under-development Oxford-NINJAL Corpus of Old Japanese (ONCJO)
- [myST](#) (also [myST2](#)) — tree annotations for the first 97 poems of the Manyōshū (MY37)

Contemporary English

- [TSPC](#) — the under-development [Treebank Semantics](#) Paroed Corpus (a testing ground for [Treebank Semantics](#))

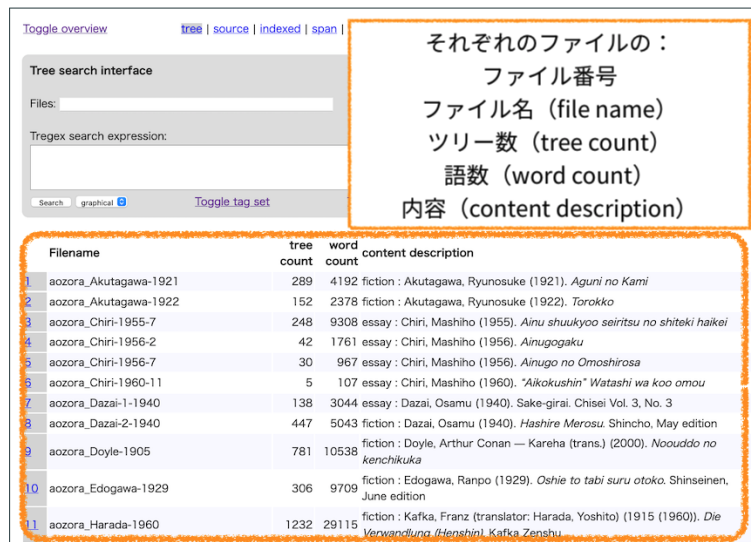
4

ファイル一覧：Overview（概要）ページ



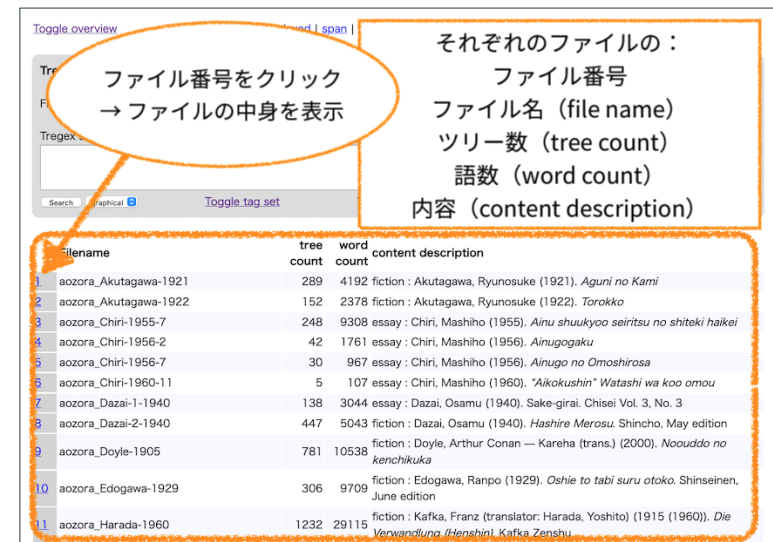
5

ファイル一覧：Overview（概要）ページ



6

ファイル一覧：Overview（概要）ページ



7

ファイルの中身：Context view（コンテキスト表示） ページ

Toggle context view Go to Kusunoki Corpus overview

title	Torokko
author	Akutagawa, Ryunosuke
date	1922
source	Aozora Bunko (https://www.aozora.gr.jp/cards/000879/card43016.html)
genre	fiction
subcorpus	Aozora Bunko
terms of use	Creative Commons license

メタデータ

- 1 トロッコ
- 2 芥川龍之介
- 3 小田原 熱海（あたま）間に、軽便鉄道敷設（ふせつ）の工事が始まったのは、良平（りょうへい）の八つの年だった。
- 4 良平は毎日村外（はず）れへ、その工事を見物に行った。
- 5 工事を——といったところが、唯（ただ）トロッコで土を運搬する——それが面白さに見に行ったのである。
- 6 トロッコの上には土工が二人、土を積んだ後（うしろ）に佇（たたず）んでいる。
- 7 トロッコは山を下（くだ）るのだから、人手を借りずに走って来る。
- 8 備（あお）るように車台が動いたり、土工の神天（はんでん）の裾（すそ）がひらついたり、細い線路がしなったり——良平はそいと思う事がある。
- 9 せめては一度でも土工と一しょに、トロッコへ乗りたいと思う事もある。
- 10 トロッコは村外れの平地へ来ると、自然と其処（そこ）に止まってしまう。
- 11 と同時に土工たちは、身軽にトロッコを飛び降りるが早いか、その線路の終点へ車の土をぶちまける。
- 12 それから今度はトロッコを押し押し、もと来た山の方へ登り始める。
- 13 良平はその時乗れないまでも、押す事さえ出来たらと思うのである。
- 14 或（ある）夕方、——それは二月の初旬だった。
- 15 良平は二つ下の弟や、弟と同じ年の隣の子供と、トロッコの置いてある村外れへ行った。

8

ファイルの中身：Context view（コンテキスト表示） ページ

Toggle context view Go to Kusunoki Corpus overview

title	Torokko
author	Akutagawa, Ryunosuke
date	1922
source	Aozora Bunko (https://www.aozora.gr.jp/cards/000879/card43016.html)
genre	fiction
subcorpus	Aozora Bunko
terms of use	Creative Commons license

メタデータ

クリックで overview へ

- 1 トロッコ
- 2 芥川龍之介
- 3 小田原 熱海（あたま）間に、軽便鉄道敷設（ふせつ）の工事が始まったのは、良平（りょうへい）の八つの年だった。
- 4 良平は毎日村外（はず）れへ、その工事を見物に行った。
- 5 工事を——といったところが、唯（ただ）トロッコで土を運搬する——それが面白さに見に行ったのである。
- 6 トロッコの上には土工が二人、土を積んだ後（うしろ）に佇（たたず）んでいる。
- 7 トロッコは山を下（くだ）るのだから、人手を借りずに走って来る。
- 8 備（あお）るように車台が動いたり、土工の神天（はんでん）の裾（すそ）がひらついたり、細い線路がしなったり——良平はそいと思う事がある。
- 9 せめては一度でも土工と一しょに、トロッコへ乗りたいと思う事もある。
- 10 トロッコは村外れの平地へ来ると、自然と其処（そこ）に止まってしまう。
- 11 と同時に土工たちは、身軽にトロッコを飛び降りるが早いか、その線路の終点へ車の土をぶちまける。
- 12 それから今度はトロッコを押し押し、もと来た山の方へ登り始める。
- 13 良平はその時乗れないまでも、押す事さえ出来たらと思うのである。
- 14 或（ある）夕方、——それは二月の初旬だった。
- 15 良平は二つ下の弟や、弟と同じ年の隣の子供と、トロッコの置いてある村外れへ行った。

1行 = 1ツリー（文）

9

セッション3の内容

1. ファイル一覧
2. 文字列検索
3. （ツリー検索）ファイルの指定
4. （ツリー検索）検索結果の表示方法
5. （ツリー検索）検索結果の保存
6. インターフェースに関するドキュメント

10

Overview から String search（文字列検索）画面へ

Toggle overview tree | source | indexed | span | tptp | flora2 | collocates | graph | download Main page Help

Tree search interface

Files:

Tregex search expression:

Search graphical Toggle tag set Tregex search guide

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4192	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9308	essay : Chiri, Mashiho (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1761	essay : Chiri, Mashiho (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	967	essay : Chiri, Mashiho (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashiho (1960). <i>"Aikokushin" Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3044	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3
8 aozora_Dazai-2-1940	447	5043	fiction : Dazai, Osamu (1940). <i>Hashire Merosu Shincho</i> , May edition
9 aozora_Doyle-1905	781	10538	fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Noouddo no kenchikuka</i>
10 aozora_Edogawa-1929	306	9709	fiction : Edogawa, Ranpo (1929). <i>Oshie to tabi suru otoko</i> . Shinseinen, June edition
11 aozora_Harada-1960	1232	29115	fiction : Kafka, Franz (translator: Harada, Yoshito) (1915 (1960)). <i>Die Verwandlung (Henshin)</i> . Kafka Zenshu

11

Overview から String search (文字列検索) 画面へ

Toggle overview クリック:
以下の3つの画面を循環
1. Overview (概要)
2. String Search (文字列検索)
3. Grammar analytics (文法分析)

Filename			
1 aozora_Akutagawa-1921	289	4192	fiction : Akutagawa, Ryunosuke (1921). <i>Aguri no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9308	essay : Chiri, Mashiho (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1761	essay : Chiri, Mashiho (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	967	essay : Chiri, Mashiho (1956). <i>Ainugo no Omoshiro</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashiho (1960). <i>"Aikokushin" Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3044	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisel Vol. 3, No. 3
8 aozora_Dazai-2-1940	447	5043	fiction : Dazai, Osamu (1940). <i>Hashire Merosu</i> . Shincho, May edition
9 aozora_Doyle-1905	781	10538	fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Noouddo no kenchikuka</i>
10 aozora_Edogawa-1929	306	9709	fiction : Edogawa, Ranpo (1929). <i>Oshie to tabi suru otoko</i> . Shinseinen, June edition
11 aozora_Harada-1960	1232	29115	fiction : Kafka, Franz (translator: Harada, Yoshito) (1915 (1960)). <i>Die Verwandlung (Henshin)</i> . Kafka Zenshu

12

String search (文字列検索) 画面

- 文字列検索：テキストボックスに入力された文字列に機械的に一致するものを抽出する。
- 読点 (。) はテキストデータから除外されているので、検索文字列に使うことはできない。
- 以下は、テキストボックスに検索文字列として「ところを」を入力したところ

String search interface

ところを

Search corpus order

13

String search (文字列検索) 画面

- 文字列検索：テキストボックスに入力された文字列に機械的に一致するものを抽出する。
- 読点はテキストデータから除外されているので、検索文字列に使うことはできない。
- 以下は、テキストボックスに検索文字列として「ところを」を入力したところ

String search interface

ところを

Search corpus order

検索文字列を入力したら
① 表示順 (後述) を選択し
② [Search] をクリック

14

String search (文字列検索) 表示順の種類

- corpus order (コーパス順 – デフォルト)
ファイル名のソート順と、ファイル中のツリー順に従った順番
- sort left (前文脈順)
検索文字列に先行する文字の順番
- sort right (後文脈順)
検索文字列に後続する文字の順番
- random
ランダムな順番

15

String search (文字列検索) 検索結果画面

Context pattern	Context pattern
Search pattern	Search pattern
1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	34
35	36
37	38
39	40
41	42
43	44
45	46
47	48
49	50
51	52
53	54
55	56
57	58
59	60
61	62
63	64
65	66
67	68
69	70
71	72
73	74
75	76
77	78
79	80
81	82
83	84
85	86
87	88
89	90
91	92
93	94
95	96
97	98
99	100

すべての検索結果が KWIC 形式で表示される

16

String search (文字列検索) 検索結果の保存

- ブラウザのメニューから (あるいは右クリックして) 「すべてを選択」「コピー」
→ 表計算ソフトやテキストエディタにペーストして保存
- 以下は、検索結果を Excel にペーストしたところ

ホーム	挿入	ページレイアウト	数式	データ	校閲	表示	操作アシスト
送ゴシック Regular [本文] 12 A A							
B I U L 背景色 塗りつぶし 罫線 表 表のデザイン 表のスタイル 表のテーマ 表のデザイン 表のスタイル 表のテーマ							
68 × f6 まわってきき五人ではいつ来たどでの切れた所へ来たのです							
13	essay_Nakaya 20	25	身体を覚えていないで覚える場合に覚えていないところを	明示すればオクターもとのurlと版権と転載自由でってのさえ明記すれば一人も文			
14	azora_Nakaya 1900 44	17	或る数値表のうちの抜けたところを	埋めるような種類の研究になりやうい			
15	azora_Miyazawa 1902 23	16	てる様子がとかで普通の人で観望したところを	あらわれていた			
16	wehmedia_MYTOT 14 4	42	に一口大に切った豆腐をいれ置いたところを	引き揚げつてダレで食べる			
17	azora_Miyazawa 1934 275	11	やをちゃんと机の上へもって置いたところを	息もつかずじっと読んでいました			
18	azora_Haas 1939 225	5	キャラコさは正確なところを	いう			
19	azora_Natsume 1908 91	19	そういつまでも悩めるところを	もって見るやと前には待てはるまいと言った			
20	textbook_Koonhong 89	35	そのところを	もう少し詳しく書いていただけませんか			
21	azora_Miyazawa 1934 428	13	ざら一部はびきのくつについたところを	しっくり押えました			
22	textbook_dig_haas 348	33	お仕事のこと	すみません			
23	azora_Miyazawa 1934 765	14	そのとき向こうの河原のわの木のところを	大が四人眠れぬがなったり眠れぬがなったりしてこっへ来るのでした			
24	textbook_particles 1129	41	学生は遊ぶところを	たくさん知ってるからなという学生はあまり勉強していないよに聞こえるけど			
25	azora_Harada 1940 813	3	というのは父親が歩で進むところを	彼は数限りない戦で進んでいかなければならないのだ			
26	azora_Dayle 1905 199	7	間ことうとアタマを叩きつけているところを	思い出しました			
27	textbook_Koonhong 81	34	その足跡は雪から逃げようとしているところを	捕まえた			
28	azora_Haas 1939 747	6	食糧に困ってつらさを覚えているところを	見たとこれがよばよばした昨日までの老人だとはどうも思われない			
29	textbook_Koonhong 1059	37	高津さんは花子と太郎と話しているところを	見かけた			
30	azora_Nakaya 1900 174	18	その分布状態などについて議論は尽きるところを	知らず物理哲学の蔵きがありありと物理学の上に現れていた			
31	textbook_Koonhong 1000	38	警察は犯人が逃げから出てくるところを	捕まえた			
32	azora_Parsons 1902 141	20	ダグとオズの間に足をかけようとするところを	脚中 (どうなの) をひとつきまきしてころしてしまいました			
33	azora_Dazai 1940 177	1	うは足音ともある王人の怪の存在するところを	見せてやろう			
34	azora_Kaguchi 1948 77	21	その食べ物と性格について己の供するところを	ハッキリ云えればそれでいい			
35	azora_Kurida 1925 272	10	そう急ぎ立てるところを	見たときではもう手が届いたか			

17

String search (文字列検索) 個々の用例の前後の文脈や解析情報を調べる

textbook_dig_haas 348	33	お仕事のつてのこと	すみません
azora_Miyazawa 1934 275	11	やをちゃんと机の上へもって置いたところを	息もつかずじっと読んでいました
textbook_particles 1129	41	学生は遊ぶところを	たくさん知ってるからなという学生はあまり勉強していないよに聞こえるけど
azora_Harada 1940 813	3	というのは父親が歩で進むところを	彼は数限りない戦で進んでいかなければならないのだ
azora_Dayle 1905 199	7	間ことうとアタマを叩きつけているところを	思い出しました
textbook_Koonhong 81	34	その足跡は雪から逃げようとしているところを	捕まえた
azora_Haas 1939 747	6	食糧に困ってつらさを覚えているところを	見たとこれがよばよばした昨日までの老人だとはどうも思われない
textbook_Koonhong 1059	37	高津さんは花子と太郎と話しているところを	見かけた
azora_Nakaya 1900 174	18	その分布状態などについて議論は尽きるところを	知らず物理哲学の蔵きがありありと物理学の上に現れていた
textbook_Koonhong 1000	38	警察は犯人が逃げから出てくるところを	捕まえた
azora_Parsons 1902 141	20	ダグとオズの間に足をかけようとするところを	脚中 (どうなの) をひとつきまきしてころしてしまいました
azora_Dazai 1940 177	1	うは足音ともある王人の怪の存在するところを	見せてやろう
azora_Kaguchi 1948 77	21	その食べ物と性格について己の供するところを	ハッキリ云えればそれでいい
azora_Kurida 1925 272	10	そう急ぎ立てるところを	見たときではもう手が届いたか
textbook_particles 1129	41	学生は遊ぶところを	たくさん知ってるからなという学生はあまり勉強していないよに聞こえるけど
azora_Harada 1940 813	3	というのは父親が歩で進むところを	彼は数限りない戦で進んでいかなければならないのだ
azora_Dayle 1905 199	7	間ことうとアタマを叩きつけているところを	思い出しました
textbook_Koonhong 81	34	その足跡は雪から逃げようとしているところを	捕まえた
azora_Haas 1939 747	6	食糧に困ってつらさを覚えているところを	見たとこれがよばよばした昨日までの老人だとはどうも思われない
textbook_Koonhong 1059	37	高津さんは花子と太郎と話しているところを	見かけた
azora_Nakaya 1900 174	18	その分布状態などについて議論は尽きるところを	知らず物理哲学の蔵きがありありと物理学の上に現れていた
textbook_Koonhong 1000	38	警察は犯人が逃げから出てくるところを	捕まえた
azora_Parsons 1902 141	20	ダグとオズの間に足をかけようとするところを	脚中 (どうなの) をひとつきまきしてころしてしまいました
azora_Dazai 1940 177	1	うは足音ともある王人の怪の存在するところを	見せてやろう
azora_Kaguchi 1948 77	21	その食べ物と性格について己の供するところを	ハッキリ云えればそれでいい
azora_Kurida 1925 272	10	そう急ぎ立てるところを	見たときではもう手が届いたか

ファイル名の右側の番号 → Context viewへ

チェックボックスの右側の番号 → Analysis viewへ

18

String search (文字列検索) 個々の用例の前後の文脈や解析情報を調べる

Context view (コンテキスト表示) では、検索文字列を含む文がハイライトされるので、その前後の文脈を確認することができる

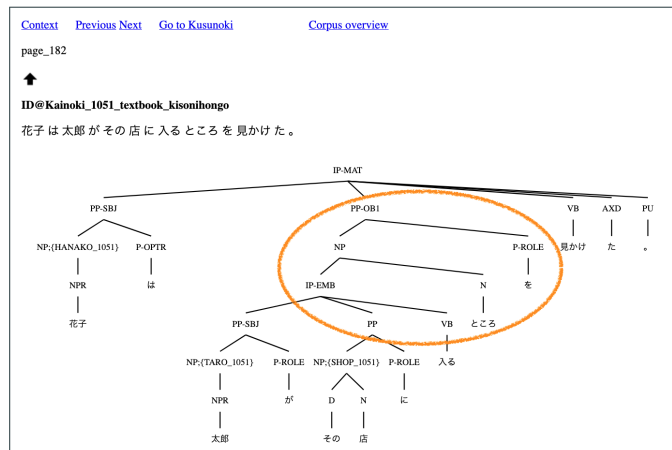
198	「ブラインドは降りていましたか？」とホームズが聞いた。
199	「自信はありませんが、半分だけ降るに当たったような気がします。」
200	いや、間違いない。
201	あの 方が 窓を 押し開こうと、ブラインドを 引き上げて いるところを 思い出しました。
202	ステッキが見つかりませんが、『気にすることはいい、きみ。きつと、また 何度も 会うことになるだろうから、今度 会うときまで 預かっ
203	そのとき、金庫は 開けて 放し、 書棚は まとめて テーブルの 上に 置いて ありました。
204	ブラインドに 隠れる 時刻は ありませんでしたから、アタリ・アームズで 一泊しました。
205	今朝の 恐ろしい 事件の ことは、 朝刊を 読むまで 何も 知り ませんでした。
206	レストレイドは、この 注目すべき 解説を 聞きながら、2、3度、 扉を 釣りあげたが、ここでも やく口を 開いた。
207	「他に 何か 聞いて おきたい ことは ありますか、 ミスター・ホームズ？」
208	「ブラインドに 行って みる までは 何と とも」
209	「ノーウッド じゃないですか」と レストレイド。
210	「ああ、 そうか。」
211	確かに それで ですよ」と ホームズが 聞いた 微笑を 浮かべて 言った。
212	ホームズの 剣刀のような 頭脳は、 レストレイドの 理解の 及ばないところまで、 事実を 鋭く 切り取る ことができる。
213	レストレイドは それを、口で 言う以上、 経験から 学んで いる。
214	レストレイドは ホームズを 興味深げに 眺めた。
215	「少し お話を 伺って おく ほうが よさそう ですね、 ミスター・シャーロック・ホームズ」と レストレイド。
216	「さて、 ミスター・マクファレン、 戸口に 警官を 2人、 外に 西輪を 待た せて ある」

19

String search (文字列検索)

個々の用例の前後の文脈や解析情報を調べる

Analysis view (分析表示) では、検索文字列を含む文のツリーが図示されるので、解析情報を確認することができる



20

練習問題 1 (文字列検索)

- ① 文字列「てくれる」を、sort left (前文脈順) で検索しなさい。
- ② いくつかの例を analysis view (分析表示) で見て、「て」と「くれる」の品詞タグ、および品詞タグの上の句タグを確認しなさい。
- ③ 検索結果を自分のコンピュータに保存しなさい。

21

セッション3の内容

1. ファイル一覧
2. 文字列検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果の保存
6. インターフェースに関するドキュメント

22

ツリー検索

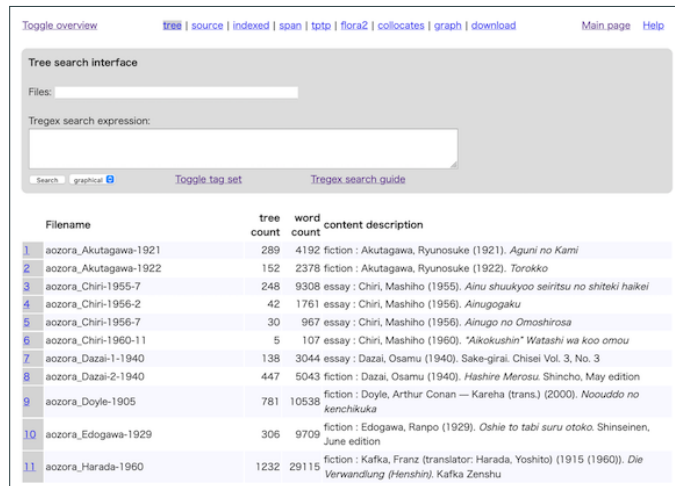
- ツリー内のノード (語・形態素の文字列、品詞タグ、句や節のタグ) やノード間の関係 (前後関係や支配・非支配関係) を検索式で指定し、検索する。
- 本インターフェースは、Tregex (Levy and Andrews 2006¹) を使った検索式に対応
- 検索式の作り方については session 4 と session 5 で
- このセッションでは、ファイルの指定、検索結果の表示方法、検索結果の保存について説明する。

¹<https://nlp.stanford.edu/software/tregex.shtml>

23

(ツリー検索) 検索ファイルの指定

ツリー検索は、Overview ページから行う。



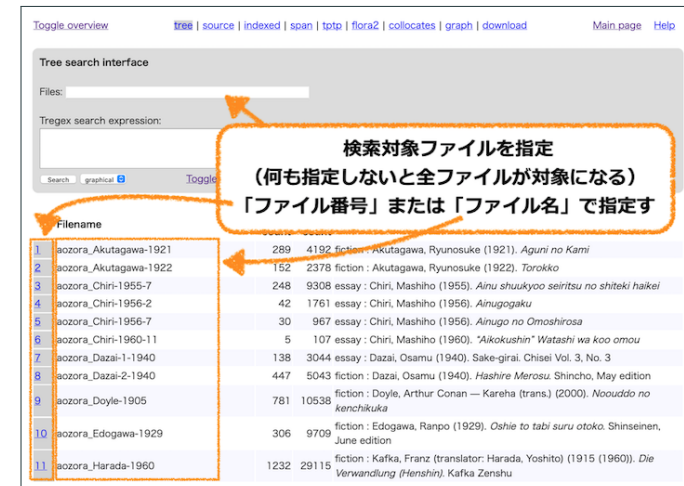
The screenshot shows the 'Toggle overview' page with a 'Tree search interface' section containing 'Files:' and 'Tregex search expression:' input fields. Below is a table of files with columns: Filename, tree count, word count, and content description.

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4192	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9308	essay : Chiri, Mashiho (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1761	essay : Chiri, Mashiho (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	967	essay : Chiri, Mashiho (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashiho (1960). <i>"Aikokushin" Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3044	essay : Dazai, Osamu (1940). <i>Sake-girai. Chisei Vol. 3, No. 3</i>
8 aozora_Dazai-2-1940	447	5043	fiction : Dazai, Osamu (1940). <i>Hashire Merosu. Shincho, May edition</i>
9 aozora_Doyle-1905	781	10538	fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Noouddo no kenchikuka</i>
10 aozora_Edogawa-1929	306	9709	fiction : Edogawa, Ranpo (1929). <i>Oshie to tabi suru otoko. Shinseinen, June edition</i>
11 aozora_Harada-1960	1232	29115	fiction : Kafka, Franz (translator: Harada, Yoshito) (1915 (1960)). <i>Die Verwandlung (Henshin). Kafka Zenshu</i>

24

(ツリー検索) 検索ファイルの指定

ツリー検索は、Overview ページから行う。



The screenshot is similar to the one on page 24, but with orange annotations. A box highlights the 'Files:' input field with the text: '検索対象ファイルを指定 (何も指定しないと全ファイルが対象になる) 「ファイル番号」または「ファイル名」で指定す'. Arrows point from this box to the first and second rows of the file list table.

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4192	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9308	essay : Chiri, Mashiho (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1761	essay : Chiri, Mashiho (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	967	essay : Chiri, Mashiho (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashiho (1960). <i>"Aikokushin" Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3044	essay : Dazai, Osamu (1940). <i>Sake-girai. Chisei Vol. 3, No. 3</i>
8 aozora_Dazai-2-1940	447	5043	fiction : Dazai, Osamu (1940). <i>Hashire Merosu. Shincho, May edition</i>
9 aozora_Doyle-1905	781	10538	fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Noouddo no kenchikuka</i>
10 aozora_Edogawa-1929	306	9709	fiction : Edogawa, Ranpo (1929). <i>Oshie to tabi suru otoko. Shinseinen, June edition</i>
11 aozora_Harada-1960	1232	29115	fiction : Kafka, Franz (translator: Harada, Yoshito) (1915 (1960)). <i>Die Verwandlung (Henshin). Kafka Zenshu</i>

25

(ツリー検索) 検索ファイルの指定

ファイル番号で指定

- 1 番目のファイルだけ

Files: 1p

- 1 番目と 10 番目のファイル

Files: 1p;10p

- 1 番目から 10 番目までのファイル

Files: 1,10p

- 1 から 10 番目までと、15 から 20 番目までのファイル

Files: 1,10p;15,20p

26

(ツリー検索) 検索ファイルの指定

ファイル名で指定

- wikipedia_Kamen_Rider だけ
(ファイル名を半角スラッシュで囲む)

Files: /wikipedia_Kamen_Rider/

- aozora_Akutagawa-1921 と aozora_Akutagawa-1922
(ファイル名の間に \ | を入れ、全体を半角スラッシュで囲む)

Files: /aozora_Akutagawa-1921\|aozora_Akutagawa-1922/

27

(ツリー検索) 検索ファイルの指定

検索ファイルを指定し、検索式を入力しない状態で、[Search]をクリックすると、検索対象ファイル全体のツリー数と語数が表示される。



28

(ツリー検索) 検索ファイルの指定

ファイル番号とファイル名、どちらで指定するのがいい？

- ファイル番号で指定の方が簡単。ただし、コーパスに新たなファイルが追加されるとファイル番号が変わる可能性がある。
- 何度もアクセスして、特定のファイル群を検索したい場合は、ファイル名のリストを事前に作成して保存しておくといよい（以下は、会話データをすべて指定する場合）。

```
/spoken_JF1\|spoken_JF10\|spoken_JF13\|spoken_JF14\|spoken_JF15\|spoken_JF2  
\spoken_JF3\|spoken_JF4\|spoken_JF5\|spoken_JF6\|spoken_JF7\|spoken_JF8\|spoken_JF9  
\spoken_JM1\|spoken_JM10\|spoken_JM15\|spoken_JM2\|spoken_JM3\|spoken_JM4\|spoken_JM5  
\spoken_JM6\|spoken_JM7\|spoken_JM8\|spoken_JM9/
```

- 正式公開されたデータだけを検索したい場合も同様。
- 正式公開されたデータのファイル名は以下を参照
<http://npcmj.ninjal.ac.jp/interfaces/cgi-bin/index.sh?db=npcmj&lang=jp>
- 参考資料ファイル (03_file_list_sample.txt)
正式公開されたデータのファイル名のリスト

29

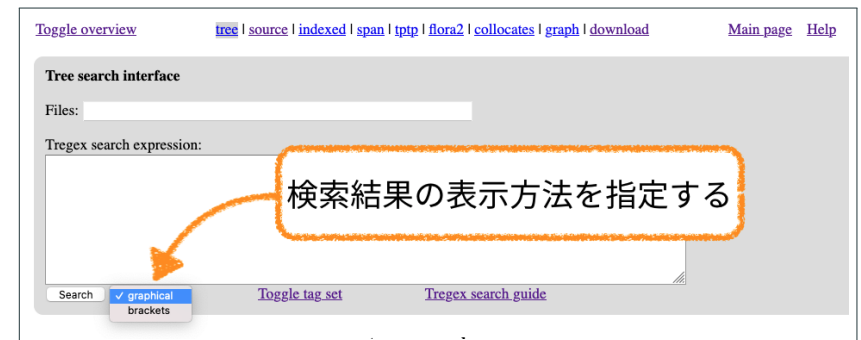
セッション3の内容

1. ファイル一覧
2. 文字列検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果の保存
6. インターフェースに関するドキュメント

30

(ツリー検索) 検索結果の表示方法

2つの表示方法：graphical と brackets

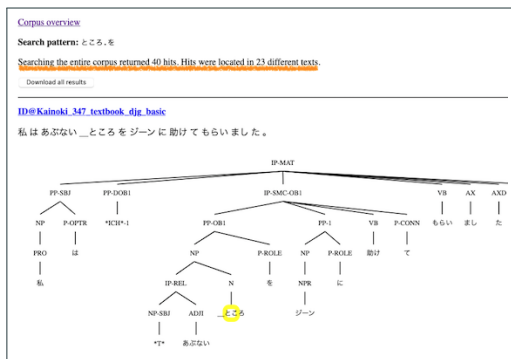


31

(ツリー検索) 検索結果の表示方法

graphical (検索ファイルを指定しない場合)

- 検索式のヒット数、ヒットを含むファイル数
- グラフィカルなツリー (100 の検索結果まで、表示内容は検索の実行毎にランダムに選ばれる)

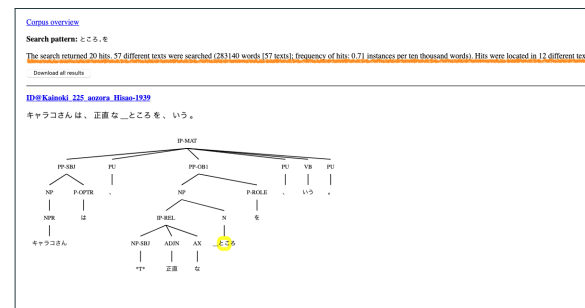


32

(ツリー検索) 検索結果の表示方法

graphical (検索ファイルを指定した場合)

- 検索式のヒット数、指定したファイル数、指定したファイル全体の語数、頻度 (10,000 語あたりのヒット数)、ヒットを含むファイル数
- グラフィカルなツリー (100 の検索結果まで、表示内容は検索の実行毎にランダムに選ばれる)

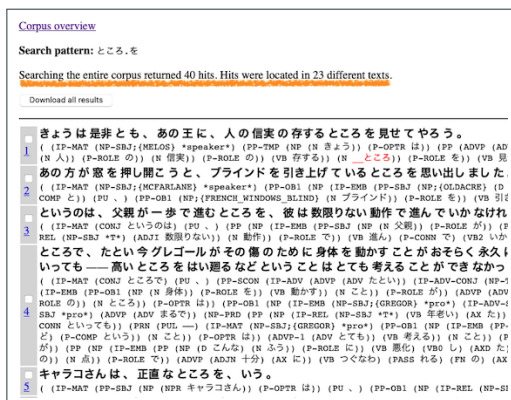


33

(ツリー検索) 検索結果の表示方法

brackets (検索ファイルを指定しない場合)

- 検索式のヒット数、ヒットを含むファイル数
- 文と括弧付きツリー (すべての検索結果)

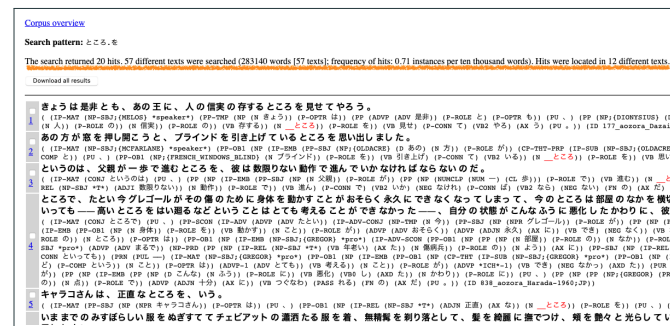


34

(ツリー検索) 検索結果の表示方法

brackets (検索ファイルを指定した場合)

- 検索式のヒット数、指定したファイル数、指定したファイル全体の語数、頻度 (10,000 語あたりのヒット数)、ヒットを含むファイル数
- 文と括弧付きツリー (すべての検索結果)



35

(ツリー検索) 検索結果の表示方法

brackets 表示では、ひとつのツリーの中に複数のヒットがあっても、そのツリーを1度しか表示しない。

- 例えば、検索式「ところ」で検索したときに、以下のツリー(文)のようにヒットが2つあっても、画面に同じツリーが2度表示されることはない。

というのは、どう考えても、彼自身が何かしら憑きものがあるような、よく町裏の小暗いところを歩いたりしている様子が、どこか普通の人間離れたところをあらわしていた。

- ヒット数と表示されたツリーの数が違うときがありうる。
- ダウンロードファイルでは、複数のヒットを含むツリーの扱いが異なる(後述)。

36

セッション3の内容

1. ファイル一覧
2. 文字列検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
- 5. (ツリー検索) 検索結果の保存**
6. インターフェースに関するドキュメント

37

(ツリー検索) 検索結果の保存

検索結果ファイルをダウンロード

- graphical 表示でも、brackets 表示でも、結果表示画面の上部の **Download all results** からすべての検索結果をファイルでダウンロードできる。
 - ファイル内の用例は括弧付きツリー形式になっており、ローカルにインストールした Tregex² やテキストエディタで開くことができる。
 - ダウンロードファイルでは、ひとつのツリーに複数のヒットがある場合、ヒットした数の分の同じツリーが含まれる。よって、ヒット数とファイルに収録されたツリーの数になる。

²<https://nlp.stanford.edu/software/tregex.shtml>

38

(ツリー検索) 検索結果の保存

brackets 表示の画面からコピー

- brackets 表示にすると、検索結果はすべて画面に表示される(複数ヒットを含むツリーの扱いに注意)。これをブラウザのメニューから(あるいは右クリックして)「すべてを選択」「コピー」し、表計算ソフトやテキストエディタにペーストしてもよい。

39

セッション3の内容

1. ファイル一覧
2. 文字列検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果の保存
6. インターフェースに関するドキュメント

40

インターフェースに関するドキュメント

The screenshot shows the Tregex web interface. At the top, there are navigation links: [tree](#), [source](#), [indexed](#), [span](#), [tntp](#), [flora2](#), [collocates](#), [graph](#), [download](#), [Main page](#), and [Help](#). The main section is titled "Tree search interface" and contains a "Files:" input field and a "Tregex search expression:" input field. Below these are buttons for "Search" and "graphical".

Annotations with orange boxes and arrows point to specific features:

- Help : インターフェースガイド置き場へ** (Help: Interface guide location) points to the [Help](#) link at the top right.
- Toggle tag set : タグ一覧の表示・非表示** (Toggle tag set: Display/Hide tag list) points to the [Toggle tag set](#) button.
- Tregex search guide : Tregex のガイド置き場へ** (Tregex search guide: Tregex guide location) points to the [Tregex search guide](#) button.

Below the buttons is a table showing search results:

	tree	word	content	descr
	int	count		
5	89	4192	fiction : Akutaga	
6	52	2378	fiction : Akutaga	
7	48	9308	essay : Chiri, Ma	
8	42	1761	essay : Chiri, Ma	
9	30	967	essay : Chiri, Mashiho (1956). <i>Ainugo no Omoshiro</i>	
10	5	107	essay : Chiri, Mashiho (1960). <i>"Aikokushin" Watashi wa koo omou</i>	
11	138	3044	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3	
12	447	5043	fiction : Dazai, Osamu (1940). <i>Hashire Merosu</i> . Shincho, May edition	
13	781	10538	fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Noouddo no kenchikuka</i>	
14	306	9709	fiction : Edogawa, Ranpo (1929). <i>Oshie to tabi suru otoko</i> . Shinseinen, June edition	
15	1232	29115	fiction : Kafka, Franz (translator: Harada, Yoshito) (1915 (1960)). <i>Die Verwandlung (Henshin)</i> . Kafka Zenshu	

41

練習問題2 (ツリー検索)

- ① 検索対象ファイルを1番目から57番目までのファイルとし、[Search] ボタンをクリック、検索対象ファイルに含まれるツリー数と語数を確認しなさい。
- ② Tregex search expression のボックスに P-FINAL (すべて半角) と入力し、まず、検索表示方法を tree にして検索しなさい。その後、検索表示方法を brackets にして検索しなさい。
- ③ 検索結果をダウンロードし、テキストエディタで開いて中身を確認しなさい。

42