

# Co-Chuマニュアル

Ver2.  
2021.7.

1

## 目次

1. はじめに	p. 4
2. データ作成	p. 7
2.1 データ作成 (csv)	p. 8
2.2 データ作成 (txt)	p. 9
2.3 データ作成 (メタ情報付きのcsv)	p.10
3. コーパス作成	p.11
3.1 Build (コーパスの階層構造を決定する)	p.12
3.2 Import (データを確認して、システムにアップロードする)	p.16
3.3 Edit (MeCab-UniDicによる形態素・コロケーション解析)	p.24
3.3.1 形態素・コロケーション解析の実行	p.24
3.3.2 形態素解析の確認	p.26
3.3.3 形態素解析の結果の訂正 (1) Fix Chainでの訂正	p.28
3.3.4 形態素解析の結果の訂正 (2) Fix Chainで直せない誤解析の訂正	p.32
3.3.5 形態素解析の結果の訂正 (3) その他の誤解析の訂正	p.37
3.4 ラインメタ情報のあるデータのimport	p.38

2

4. コーパス分析	p.42
4.1 Analyze(分析) 画面の構成	p.43
4.2 分析対象データの選定	p.44
4.3 検索項目の指定	p.45
4.4 メタ情報の条件指定 (メタ情報で結果を絞り込む)	p.52
4.4.1 スピーカー (speaker) メタデータ	p.52
4.4.2 ライン (line) メタデータ	p.53
4.5 アウトプット設定 (出力方法を指定して結果を表示する)	p.56
4.5.1 モード切り替え	p.57
4.5.2 アウトプット形式	p.60
4.5.3 ソート基準	p.61
5. その他	p.63
5.1 便利な機能 (①データセット情報、② Export Dataset、③ speakerのメタ情報の確認)	p.64
5.2 検索比較機能	p.70
5.3 Before/After フィルター機能	p.73

参考文献・参考URL

3

## 1. はじめに

Co-Chuは、研究者や日本語教師が自ら収集したオリジナルデータを取り込んでコーパスを作成し、分析できるウェブアプリケーションである。

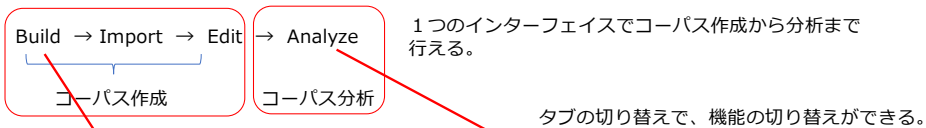
Co-Chu の入り口: <https://planning-lasso-cliff.projectxero.net/>

< 3つの特長 >

- 1) オリジナルデータを取り込んで、コーパス作成からコーパス分析まで1つのインターフェイスで使用できる。
- 2) 誤用や話しことばなど形態素解析において誤解析となりうるものにも対応できる (山本他2018)。
- 3) さまざまなメタ情報を自由に付与し、分析に用いることができる (本間他2019)。

4

# Co-Chuのインターフェイス



# Co-Chuが使用可能なブラウザ (2020.3. 時点)

Windowsの場合 : Google Chrome

Mac の場合 : Google Chrome (タブレットは使用不可)

このマニュアルでは以下の形で例やコツを紹介する。

例を示します

便利な機能、ちょっとしたコツなどを紹介しています

## 2. データ作成

• Co-Chuでは、データはcsv形式かtxt形式で取り込むことができる。

2.1 データ作成 (csv)

2.2 データ作成 (txt)

2.3 データ作成 (メタ情報付きのcsv)

### 2.1 データ作成 (csv)

#### 例 1 : 会話データ

B列に発話内容を記載。  
以下、この部分を「ライン」と呼ぶ。

A列に発話者を記載。

A	B	C	D	E	F	G	H	I	J
1	日本人	あ、お話ししたいです。							
2	アメリカ人	自分自身から話さなければいけません。							
3	日本人	日本語を、はい。							
4	アメリカ人	はい、その、えー、日本語の先生になるために、(はい) それ何という専攻なんですか。							
5	日本人	あ、えーっと、授業自体は、えーっと、日本語教授法？ (はい) のことを受けていて、(ああ) 何(なん)か、ABCDとある。							
6	アメリカ人	なるほど (はい) ですね、僕もそう思う、あの、学生が、(はい) 見学してくる授業に (はい) 参加したことはない。							
7	日本人	あの、どうですか。							
8	アメリカ人	日本の『日本語教育センター』の授業を(あ) 行らんですか？							
9	日本人	はい、はい。							
10	アメリカ人	あ、素晴らしいです。							
11	日本人	(笑)							
12	アメリカ人	スーパースターが観るんですか？							
13	日本人	あ、そうですね、スーパ、(笑) すごい勉強しました。							
14	アメリカ人	はい、そうですね、驚かしてあげてください。							
15	日本人	(笑)							
16	アメリカ人	それ、学生のほうに質問するんですか。							
17	日本人	あ、見られて。							
18	アメリカ人	見られて、その、学生はすごく、いい印象を付けて、(ああ、ああ) ああ、観るんですか、まあ、私の、まあ、							
19	日本人	はい、(笑)							
20	アメリカ人	はい、もう聞いてますね。							
21	日本人	あ、そうですね、(笑)							
22	アメリカ人	はい、はい、ありがとうございます。							
23	日本人	(笑) ありがとうございます。							
24	アメリカ人	はい。							
25	アメリカ人	えー、そうですね、(はい) ああ、君のほうですね。							
26	日本人	はい。							
27	アメリカ人	えー、(笑) とお話しして、(はい) えー、アメリカから来ました。							

**発話内容の入力上の注意**  
文字化のルールは、研究者が自由に決定できる。  
左の会話データ例には、  
同時発話 //  
あいづち (はい)  
上昇イントネーション ↑  
笑い声 {笑}  
などのメタ情報が含まれる。  
これらの記号及び( )等、記号にはさまれた文字列は形態素解析の対象にはならないので、メタ情報を付しても分析の妨げにはならない。

## 2.2 データ作成 (txt)

### 例2 : 作文データ (文章のみ)

発話者情報などが必要な文章のみのデータの場合、テキストファイル (.txt) の形で取り込み用のデータを作成してもよい。



9

## 2.3 データ作成 (メタ情報付きのcsv)

### 例3 : 作文データ

文レベルの問題で  
気になったことなどを  
ラインメタ情報として  
右のように入れておく  
こともできる。

例

A列 : 「コメント」 = 文レベルの誤用  
B列 : 「その他」 = 誤用以外の情報

C列 : speaker情報 (この場合はwriter)  
を検索で活用するために設けておく。  
\* speaker 情報はp.21 参照

	コメント	その他	speaker (writer)	D	E	F	G	H	I	J	K
1											
2											
3	おじれ文										
4											
5											
6											
7											
8											
9											
10											
11											
12											
13	おじれ文										
14	文体										
15											
16											
17	おじれ文										
18											
19											
20	文体										
21											
22											

\* ラインメタ情報も検索キーとして利用可能。例えば「おじれ文」だけを検索したい時などに利用する。10

## 3. コーパス作成

- コーパス作成は、  
Build → Import → Editの順に行う。

3.1 Build : コーパスの階層構造を決定する

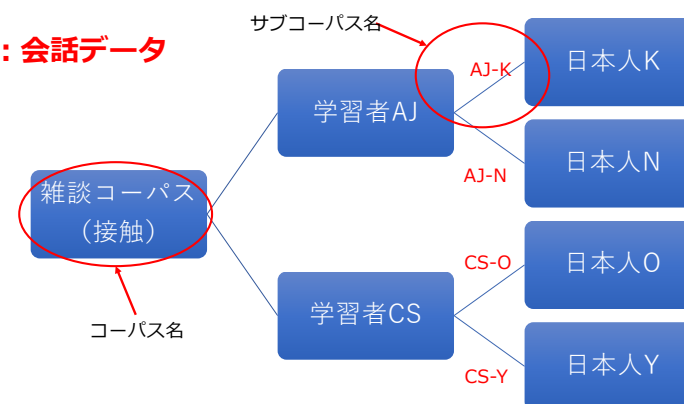
3.2 Import : データを確認して、システムにアップロードする

3.3 Edit : MeCab-UniDicによる形態素・コロケーション解析

11

## 3.1 Build : コーパスの階層構造を決定する

### 例1 : 会話データ

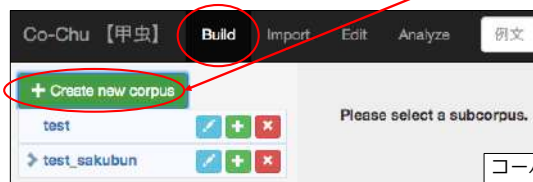


コーパス名

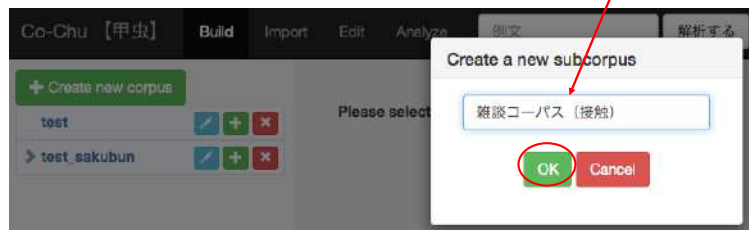
12

例1：会話データを取り込む場合

Buildを選択して出た画面の  
Create new corpusをクリック。



コーパス名を入力し、OKをクリックする。

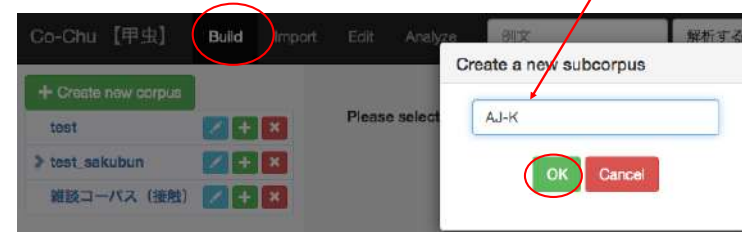


13



新しいサブコーパスを作る位置で+を押す

サブコーパス名を入力し、OKをクリックする。



14

Importの前にコーパスの階層構造が正しく入力されているかを確認する。



>をクリックすると、下位のコーパスが  
確認できる。

サブコーパスは、50音順（ABC順）に並ぶので、  
並べたい順番がある場合は、サブコーパス名に  
番号を付けておくと便利  
ex. 1.自己紹介, 2.ある物事の説明, 3.体験  
の報告, 4.2つのものの比較 (p.38参照)

15

### 3.2 Import : データを確認し、システムにアップロードする



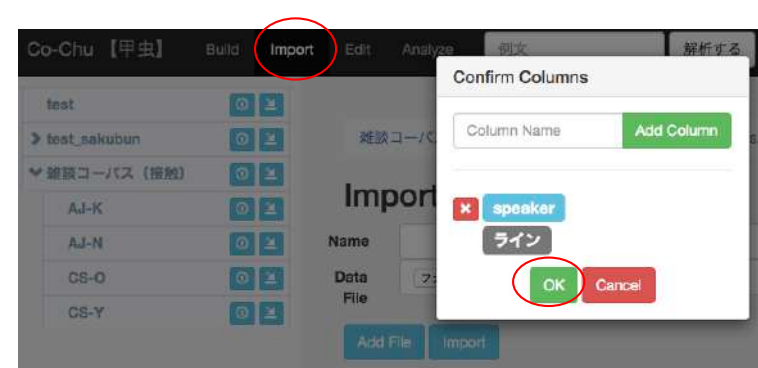
Import画面で、データを取り込むサブコーパス(AJ-K)の  
ここをクリックする。

16



ファイル選択をクリックして、取り込みたいデータを選択し、Importをクリックする。

1つのサブコーパスに同時に複数のファイルを取り込む時に使う。



ラインメタ情報を入れない場合はこのままOKをクリックする。  
\*ラインメタ情報の入れ方についてはp.38~p.41参照。



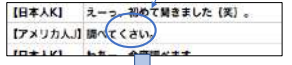
OKをクリックすると、このように取り込んだデータが表示される。スクロールして一番下まで入っているかデータを確認する。



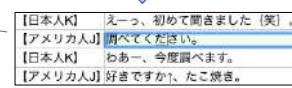
データを確認後OKをクリックする。何か問題があったらCancelをクリックしp.16に戻ってデータを修正して入れ直す。

\*少しの修正箇所(タイプミスなど)なら、この画面に直接書き込んで直すこともできる。

例えばこのようにミスを見つけたら



修正ができる





Speakersのメタ情報を入れる

- (1)Keyに属性、(2)Valueにメタ情報を入力する。複数のメタ情報を入れる場合、(3)Add Metadataをクリックするとメタ情報の入力行を増やすことができるので、必要だけAdd Metadataをクリックして、情報を入れる。Speakerごとに行う。



\*メタ情報は、検索キーとして使用できる。レベル別、母語別等検索に活用したい情報を入れておくことができる。



各Speakerについて記入が完了したらOKをクリックするとImportが開始される。  
→ 次ページへ



オレンジから緑に変わればImport完了。



### 3.3 Edit : MeCab-UniDicによる形態素・コロケーション解析

#### 3.3.1 形態素・コロケーション解析の実行

Edit画面を選択

オレンジの「形態素解析を行う」をクリックする。ボタンが緑色に変われば解析完了。

形態素解析を行う

コロケーション解析を行う

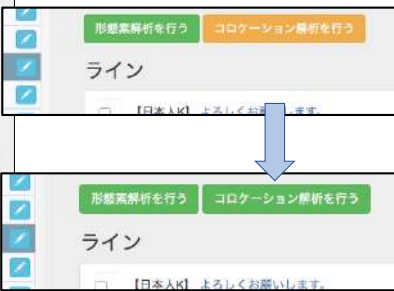
ライン

ライン

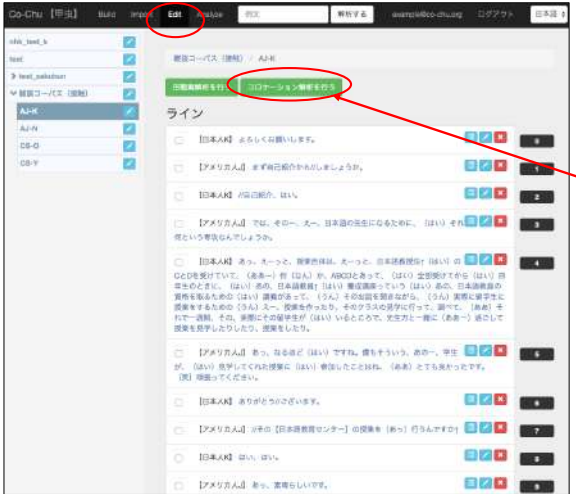


Edit画面でつづいて  
コロケーション解析を行う。

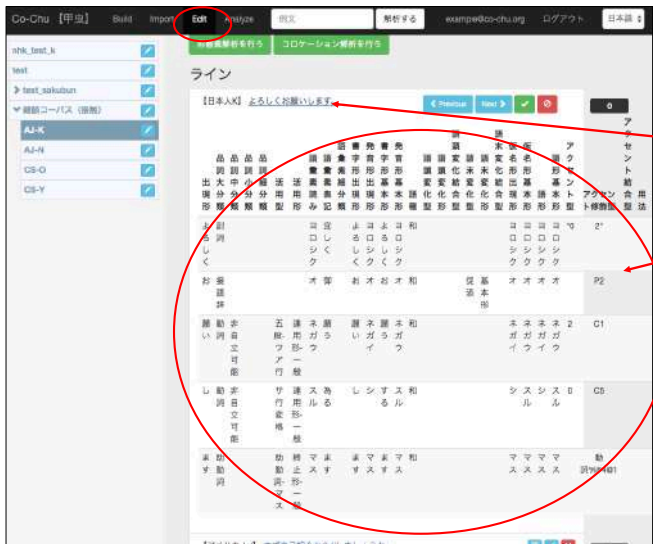
オレンジの「コロケーション解析を行う」  
をクリックする。ボタンが緑色に変われば  
解析完了。



### 3.3.2 形態素解析の確認



Importが完了した状態の画面では、  
形態素解析が行われた部分が青字で  
表示される。  
青字の部分をクリックすると、  
そのラインの形態素解析の結果が  
表示される。→次ページへ



解析結果の表示を隠すには  
クリックしたラインを再度  
クリックする。

クリックしたラインの形態素  
解析の結果が、このように表  
示される。  
正しく解析されているか確認  
できる。

### 3.3.3 形態素解析の結果の訂正 (1) Fix Chainを用いた訂正



形態素解析に誤りがあった  
場合は、まずFix Chainを  
クリックする。  
→次ページへ

ここでは、「いや」が  
「感動詞」と解析される  
べきところ、  
「形容詞いい」と「助詞  
や」に分かれて解析され  
ている。



Fix Chainをクリックして  
出た画面の、Nextを  
クリックすると、形態素  
解析の別案を検索できる。  
正しい形態素解析が  
表示されるまでNextを  
繰り返しクリックする。



正しい形態素解析が  
表示されたら、  
Accept Fixesをクリック  
する。

「いいや」の形態素  
解析が「感動詞」

Nextを繰り返しクリック  
しても正しい形態素解析  
が表示されなかった場合  
は→p.32へ



Accept Fixesをクリックすると、このような  
画面が出るので、ライン全体の形態素解析に  
誤りがないか再度確認する。

形態素解析結果の表示を出さない元の表示に戻  
しても、このラインが編集済みであることが  
「✓」のマークによって示される。



### 3.3.4 形態素解析結果の訂正 (2) Fix Chainで直せない誤解析の訂正



Nextを繰り返しクリックしても  
正しい形態素解析が表示されな  
かった場合は、誤解析の部分  
を直接書き換えて訂正する。

ここでは、「あの」が「感動詞」  
と解析されるべきところ、  
「連体詞」と解析されている。

編集したい箇所である「連体詞」  
をクリックする。→次ページへ





### 3.3.5 形態素解析結果の訂正 (3) その他の誤解析の訂正

話しことばや、誤用があるために、形態素解析がきちんとできない場合には、タグを付けて対応する。

\*タグ：話しことばや誤用など誤解析の要因となる箇所を「正しい形」(タグ 実際の発話)のようにすることで、正しく形態素解析ができるようになる。

p.35の要領で編集する

ここでは、「や、やだったことか」に「|やだっ| (淀 や、やだっ) たことか」のように、「言い定め」を示す「淀」タグを付与

タグはこのようなEditでつけるほか、元データに予めタグ付けしておいてもいい。

### 3.4 ラインメタ情報のあるデータのImport

p.10の例3のようなラインメタ情報のある作文データをImportする際には、ラインメタ情報も取り込む必要がある。まずp.16~p.18までを行った上で、ラインメタ情報をImportする。→次ページへ

コメント その他

	A	B	C	D	E	F
1		タイトル				
2						
3		ねじれ文				
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14		文体				
15						
16						
17		ねじれ文				
18						
19						
20		文体				
21						
22						
23						

ラインメタ情報名を入れる。

\*入れる順序は次ページ

(1) Column NameにデータのA列のメタ情報名(コメント)を入力し、Add Columnをクリックする。

(2) 同様に次のB列のメタ情報名(その他)を入力し、順に全ての列のメタ情報名を入力する。

(3) メタ情報名の並び順を元データでのメタ情報の並び順と同じになるようにドラッグして動かす。

(4) メタ情報名の並び順を確認したら、OKをクリックする。→次ページへ

コメント・・・A列  
その他・・・B列  
Speaker・・・C列  
ライン・・・D列

コメント	その他	speaker	ライン
	タイトル		私の故郷のお祭り リム
ねじれ文			ベトナムにとって、お正月は一年間で一番大切な祭りだ。しかし、私にとって、テットというベトナムのお正月の他に、私の故郷のお祭りが大切で大好きだ。お祭りの前はリムと呼ばれて、テットの日に旧暦の1月13日に行われる。リムの目的は耕作準備のために神様に祈ることである。祈ること以外に、リムの中にはたくさん活動が楽しめるが、一番有名でみんなに知られるのは伝統的ゲームというのは中国風棋やつなひきや武道の戦いなどである。リムには中国風棋が非常に特別だ。チェスのこまは人が全部に渡される。だから、祈祭するのは皆さんに渡される。稲穂の畑に、つなひきも人気である。ベトナムにはどこでもお祭りがあつなひきがなりたい。なぜなら、昔から、このゲームは昔の協調性や理解感を示す。また、ファンホのおかげで、リム祭が非常に有名です。ファンホの歌いは非常に特別な歴史と文化的を持っているから、UNESCOに無形文化遺産だから、ファンホが素晴らしいのを準備すると、多くの人が集まる。リムで、ファンホを演奏する場所はどこでも行うものである。その上、変化するコンテストもあるため、参加者がうれしくて、夢中させる。このように、リム祭は文化的な祭りを持つのである。本当に私はリム祭に一回だけ参加できました。だから、毎年、お正月の頃にいつも故郷へ帰ってリム祭りに参加したい。しかし、行く時間が短くなる時間になるから、参加できない。ちょっと悪いが、みんながいてくれる、必ず参加しよ。

データを確認後OKをクリックする。つづいて、Confirm Speakersの画面が出るので、p.21と同じようにspeaker(writer)の情報を入れる。

その後はp.37までと同様の操作を行う。

## 4. コーパス分析

• コーパス分析は、Analyze で行う。

- 4.1 Analyze (分析) 画面の構成
- 4.2 分析対象データの選定
- 4.3 検索項目の指定
- 4.4 メタ情報の条件指定：メタ情報で結果を絞り込む
- 4.5 アウトプット設定：出力方法を指定して結果を表示する

### 4.1 Analyze(分析)画面の構成

①分析対象データの指定

②検索項目の指定

③メタ情報の条件指定

④結果の表示方法の指定 (アウトプット)

### 4.2 分析対象データの選定

コーパス (あるいはサブコーパス) をクリックすると、分析対象として指定できる。もう一度クリックすると解除できる。  
\* 選択するとコーパスの色が (白抜き) に変わる。

一度にすべての選択を解除する「全選択解除」ボタン。

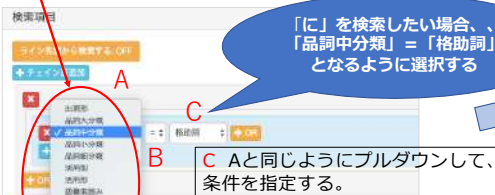
Analyzeでは、前回のAnalyzeの際に選択したコーパスが、そのまま選択されているため、毎回、分析操作をはじめる前にこのボタンをクリックしてから開始すると、間違いない。

接触場面 (アメリカ人Jと日本人K) で各話者が助詞をどのくらい、どのように使っているかを調べたい





A ×印の隣にあるボックスからプルダウンして指定したい条件を選択し、つぎにその分類の下位項目を右側のValueというボックスに入力して指定する。



「に」を検索したい場合、「品詞中分類」=「格助詞」となるように選択する

B C Aと同じようにプルダウンして、条件を指定する。

- ・「条件にあうもの(=)」
  - ・「それ以外(≠)」
  - ・「あいまい(≈)」
- の3種類の検索が指定できる。

「に」を検索したい場合、例えば「語彙表記=に」のほか、「語彙読み=ニ」などでも可能

選択肢の中からp.47の結果を参考に検索項目を具体的に指定する。

「品詞中分類」=「格助詞」という条件だけでは「に」に絞れないので、品詞中分類を指定したボックスのすぐ下の水色の「+条件追加」をクリックして条件を追加する



### 3) 検索項目を追加するには

このグレーの枠一つが形態素一つに相当するので、形態素の連鎖(チェーン)を検索したいときは、前後に条件を追加する。



例：「に」または「を」を検索する場合

例：「には」を検索する場合

条件が決まったら



次ページへ

### 4) 結果の出力

200の発話文の中に58の「に」があることがわかる。

もっと詳しく検索したいとき  
 \*例：発話者ごとの回数を知りたい → **speaker メタ情報の指定** (p.52)  
 \*例：「に」が、ラインの中のどこにあるのかわかりたい。  
 「に」の前になる要素を知りたいetc. → **アウトプット設定** (p.56~)

ライン	話者ID	ライン番号	出現	結果数: 58 / ライン数: 200
では、その一、えー、日本語の先生になるために、(はい) それ何という職業なんだろうか。	9309	3	に	
では、その一、えー、日本語の先生になるために、(はい) それ何という職業なんだろうか。	9309	3	に	
あ、なるほど(はい)です。僕もそういう、あのー、学生が、(はい) 競争してくれた授業に(はい) 参加したこと、(ああ) とても良かったです。(あ) 頑張って下さい。	9309	5	に	
あ、えーっと、授業内容は、えーっと、日本語教授法(はい) のCDを聴かせて、(ああ) 何(なん)か、ABC Dとあって、(はい) 全部聴けたら(はい) 自分定めたよ、(はい) ああ、日本語教授法(はい) 講座聴いていう(はい) ああ、日本語教授法の資格を取ったため(はい) 講座があって、(うん) その辺話を聞きながら、(うん) 実際に留学生に授業をするための(うん) えー、授業を作ったり、そのクラスの理解に行くと、調べて、(ああ) それでー講師、その、実際にその留学生が(はい) いるところで、先生方と一緒に(ああ) 講義して授業を授けたりしたり、授業をしたよ。	9304	4	に	
あ、えーっと、授業内容は、えーっと、日本語教授法(はい) のCDを聴かせて、(ああ) 何(なん)か、ABC Dとあって、(はい) 全部聴けたら(はい) 自分定めたよ、(はい) ああ、日本語教授法(はい) 講座聴いていう(はい) ああ、日本語教授法の資格を取ったため(はい) 講座があって、(うん) その辺話を聞きながら、(うん) 実際に留学生に授業をするための(うん) えー、授業を作ったり、そのクラスの理解に行くと、調べて、(ああ) それでー講師、その、実際にその留学生が(はい) いるところで、先生方と一緒に(ああ) 講義して授業を授けたりしたり、授業をしたよ。	9304	4	に	

### 4.4 メタ情報の条件指定:メタ情報で結果を絞り込む

#### 4.4.1 スピーカー(speaker)メタデータ



Importの際に入れた話者の属性を検索条件として使用する場合

母語別に検索したい

(2) ImportのAdd Metadataで入力したspeakerの属性を指定するのであれば、speakerを選択する。

作文などに付与したラインメタ情報を検索したい場合はlineを選択する(p. 53)

\*属性(メタデータ)の確認・入れ方はp.68、69参照

(3) 検索したい属性を入力する。

(1) メタデータ・フィルタをクリックすると、画面中央に選択する窓が表示される。

4) 画面左下のメタデータ・フィルタに検索したい値を入力する。

#### 4.4.2 ライン(line)メタデータ

内容に関するメタ情報を検索条件として使用する場合

1) 「メタデータ・フィルタ」をクリックし、lineを選択する。

2) Metadata Itemのボックスに検索したい内容を入れた「列のラベル名」を入力する。

3) 「メタデータ・フィルタ」に検索したい内容を入力する。

「メタデータ・フィルタ」をクリックし、lineを選択する。

Metadata Itemのボックスに「コメント」を入力する。

「メタデータ・フィルタ」に「コメント」を入力する。

A列のラベルを「コメント」として、importしたので、ここでは「コメント」と入力している (p.38-41参照)

「メタデータ・フィルタ」をクリックし、lineを選択する。

Metadata Itemのボックスに「コメント」を入力する。

「メタデータ・フィルタ」に「コメント」を入力する。

A列のラベルを「コメント」として、importしたので、ここでは「コメント」と入力している (p.38-41参照)

メタデータ・フィルタを使うときは、「検索項目」のすぐ下の「ライン先頭から検索する」の切り替えをクリックしてONにする。

「ライン先頭から検索する」をクリックしてONにすると、ボタンがオレンジから緑色にかわり、ラインの先頭から検索される。

「ライン先頭から検索する」をクリックしてONにすると、ボタンがオレンジから緑色にかわり、ラインの先頭から検索される。

#### ラインメタ情報での検索条件を設定した検索画面

「ライン先頭から検索する」を必ずONにする。(緑色で表示される)

検索項目の条件はなにも入れず、空欄のまましておく。

\* 先頭検索について：先頭検索をonにすると、ラインメタ情報の入っているラインが、1回だけ（先頭の形態素だけ）検索される。Onにしないと、ラインメタ情報の入っているラインを形態素の数だけアウトプットしてしまう。

#### 4.5 アウトプット設定：出力方法を指定して結果を表示する

##### 検索結果表示モードの種類

- 語を探す
- 語を数える
- Nグラムを見る

(選択されたものが濃い青で表示される)

指定のアウトプット形式で検索結果が下に表示される。

助詞「に」がどのように使われているか

助詞「に」が何回使われているか

助詞「に」を含んだコロケーション情報を見る

モード切り替え

検索設定

アウトプット形式: 詳細

シート標準

サンプル内の形態素を生成して出力する

出力情報

100

形態素	ライン番号	出現回数
に	0000	1
に	0000	3
に	0000	4
に	0000	4

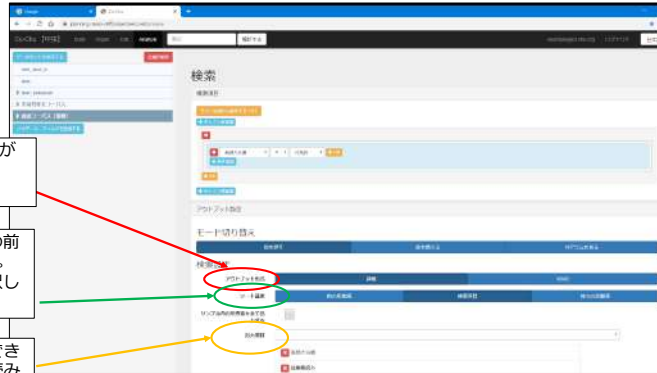
#### 4.5.1 モード切り替え

##### ・語を探す

アウトプット形式には詳細とKWICが選べる。(次ページ参照)ここでは詳細が選択されている。

ソート基準には検索語(形態素)の前後と検索語の3種類の設定ができる。ここでは、検索項目でソートを選択した。

出力項目はプルダウンで複数選択できる。ここでは品詞大分類、語彙素読みが選択されている。



57

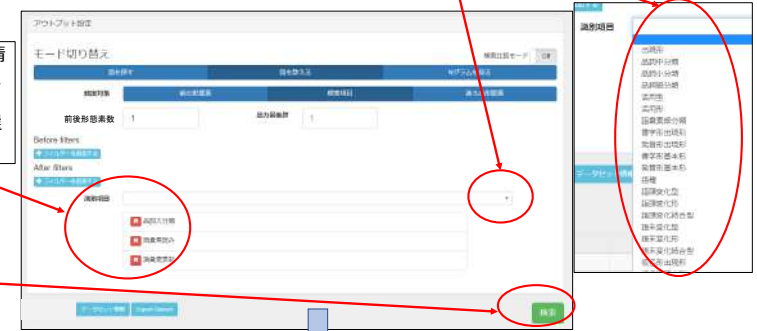
##### ・語を数える

識別項目から、表示したい情報を選ぶと、下に赤い×印とともに項目名が表示される。赤い×印をクリックすると選択をはずすことができる。

検索ボタンをクリックする。

検索結果の表示

▼をクリックすると識別項目の一覧が表示できる



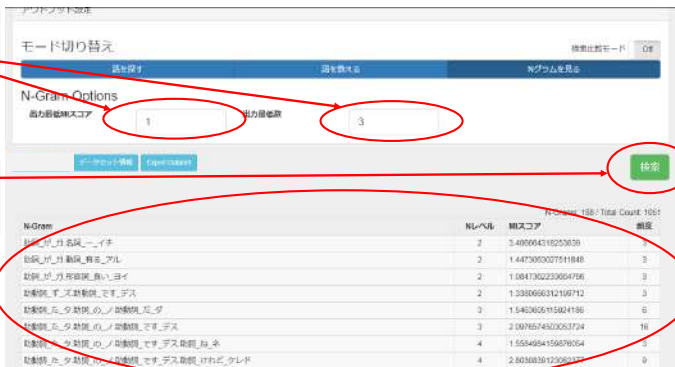
58

##### ・Nグラムを見る

ここに数値を入力して、出力最低MISコアと出力最低数を指定する。

検索ボタンをクリックすると、下にN-Gram、MISコアが表示される。

\* ここでは、MISコアはWei & Li (2013)に基づき、2から6までの形態素について計算したものの、頻度とともに表示される。



59

#### 4.5.2 アウトプット形式

アウトプット形式には、<詳細>と<KWIC>の2つの形式がある。

<詳細>

<KWIC>



・このように、同じ検索をしても、アウトプット形式を変えると結果の表示形式が異なる。

60

### 4.5.3 ソート基準

- 検索項目（形態素）、前の形態素、後ろの形態素の3種類の指定ができる。

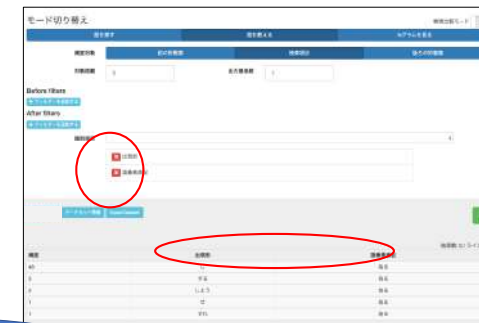
表示したい項目は選択できる  
(次のページ)



### 出力項目

プルダウンで表示させたい項目を選択できる。

語を探す



動詞「する（語彙素表記は『為る』）」が、実際はどんな形で何回使われたのか

## 5. その他

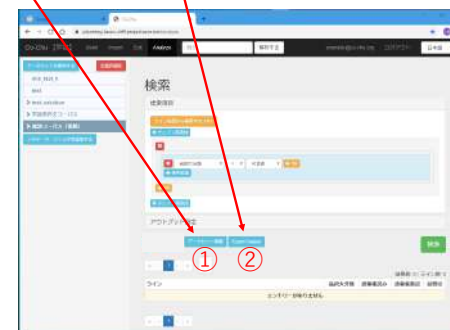
### 5.1 便利な機能

### 5.2 検索比較機能

### 5.3 Before/Afterフィルター機能

## 5.1 便利な機能

- ① 「データセット情報」
- ② 「Export Dataset」
- ③ speakerのメタ情報を確認する方法





## ① 「データセット情報」

水色の「データセット情報」をクリックするとライン数、トークン数、タイプ数、識別項目種類数を確認できる。

このOKボタンを押すと情報は消える。

項目	数
ライン数	263
短単位全語数(Token)	3600
短単位種類数(Type)	689
識別項目種類数(User Type)	790

識別項目はp.58を参照

\* Co-Chuでは、「品詞大分類」「語彙素表記」「語彙素読み」の異なり数をタイプ数として出力している。

\* 識別項目種類数は、当該検索において、ユーザーが識別項目に指定しているものの異なり数を出力する。(ここでは「出現形」を識別項目として指定)

## ② 「Export Dataset」

対象コーパスをエクセルファイルとして保存できる。

・「Download Dataset」と表示されたら「Download」という青い文字をクリックする。

・画面左下にzipファイルとしてダウンロードされたことが表示される。  
・ここをクリックすれば次のようなエクセルファイルが表示される。

## ③ speakerのメタ情報を確認する方法

\* メタデータ・フィルタで、話者や母語などで条件指定して検索する場合、どのように元データにメタ情報を付したかを確認したい時に使用する。

Build画面で対象のコーパスを指定することで、メタ情報が確認できる。

id	性別	属性
9305	男性	【アメリカ人】
9306	女性	【アメリカ人】
9307	男性	【アメリカ人】
9308	女性	【アメリカ人】
9309	男性	【アメリカ人】
9310	女性	【アメリカ人】
9311	男性	【アメリカ人】
9312	女性	【アメリカ人】
9313	男性	【アメリカ人】
9314	女性	【アメリカ人】
9315	男性	【アメリカ人】
9316	女性	【アメリカ人】
9317	男性	【アメリカ人】
9318	女性	【アメリカ人】
9319	男性	【アメリカ人】
9320	女性	【アメリカ人】
9321	男性	【アメリカ人】
9322	女性	【アメリカ人】
9323	男性	【アメリカ人】
9324	女性	【アメリカ人】
9325	男性	【アメリカ人】
9326	女性	【アメリカ人】
9327	男性	【アメリカ人】
9328	女性	【アメリカ人】
9329	男性	【アメリカ人】
9330	女性	【アメリカ人】
9331	男性	【アメリカ人】
9332	女性	【アメリカ人】
9333	男性	【アメリカ人】
9334	女性	【アメリカ人】
9335	男性	【アメリカ人】
9336	女性	【アメリカ人】
9337	男性	【アメリカ人】
9338	女性	【アメリカ人】
9339	男性	【アメリカ人】
9340	女性	【アメリカ人】
9341	男性	【アメリカ人】
9342	女性	【アメリカ人】
9343	男性	【アメリカ人】
9344	女性	【アメリカ人】
9345	男性	【アメリカ人】
9346	女性	【アメリカ人】
9347	男性	【アメリカ人】
9348	女性	【アメリカ人】
9349	男性	【アメリカ人】
9350	女性	【アメリカ人】
9351	男性	【アメリカ人】
9352	女性	【アメリカ人】
9353	男性	【アメリカ人】
9354	女性	【アメリカ人】
9355	男性	【アメリカ人】
9356	女性	【アメリカ人】
9357	男性	【アメリカ人】
9358	女性	【アメリカ人】
9359	男性	【アメリカ人】
9360	女性	【アメリカ人】
9361	男性	【アメリカ人】
9362	女性	【アメリカ人】
9363	男性	【アメリカ人】
9364	女性	【アメリカ人】
9365	男性	【アメリカ人】
9366	女性	【アメリカ人】
9367	男性	【アメリカ人】
9368	女性	【アメリカ人】
9369	男性	【アメリカ人】
9370	女性	【アメリカ人】
9371	男性	【アメリカ人】
9372	女性	【アメリカ人】
9373	男性	【アメリカ人】
9374	女性	【アメリカ人】
9375	男性	【アメリカ人】
9376	女性	【アメリカ人】
9377	男性	【アメリカ人】
9378	女性	【アメリカ人】
9379	男性	【アメリカ人】
9380	女性	【アメリカ人】
9381	男性	【アメリカ人】
9382	女性	【アメリカ人】
9383	男性	【アメリカ人】
9384	女性	【アメリカ人】
9385	男性	【アメリカ人】
9386	女性	【アメリカ人】
9387	男性	【アメリカ人】
9388	女性	【アメリカ人】
9389	男性	【アメリカ人】
9390	女性	【アメリカ人】
9391	男性	【アメリカ人】
9392	女性	【アメリカ人】
9393	男性	【アメリカ人】
9394	女性	【アメリカ人】
9395	男性	【アメリカ人】
9396	女性	【アメリカ人】
9397	男性	【アメリカ人】
9398	女性	【アメリカ人】
9399	男性	【アメリカ人】
9400	女性	【アメリカ人】
9401	男性	【アメリカ人】
9402	女性	【アメリカ人】
9403	男性	【アメリカ人】
9404	女性	【アメリカ人】
9405	男性	【アメリカ人】
9406	女性	【アメリカ人】
9407	男性	【アメリカ人】
9408	女性	【アメリカ人】
9409	男性	【アメリカ人】
9410	女性	【アメリカ人】
9411	男性	【アメリカ人】
9412	女性	【アメリカ人】
9413	男性	【アメリカ人】
9414	女性	【アメリカ人】
9415	男性	【アメリカ人】
9416	女性	【アメリカ人】
9417	男性	【アメリカ人】
9418	女性	【アメリカ人】
9419	男性	【アメリカ人】
9420	女性	【アメリカ人】
9421	男性	【アメリカ人】
9422	女性	【アメリカ人】
9423	男性	【アメリカ人】
9424	女性	【アメリカ人】
9425	男性	【アメリカ人】
9426	女性	【アメリカ人】
9427	男性	【アメリカ人】
9428	女性	【アメリカ人】
9429	男性	【アメリカ人】
9430	女性	【アメリカ人】
9431	男性	【アメリカ人】
9432	女性	【アメリカ人】
9433	男性	【アメリカ人】
9434	女性	【アメリカ人】
9435	男性	【アメリカ人】
9436	女性	【アメリカ人】
9437	男性	【アメリカ人】
9438	女性	【アメリカ人】
9439	男性	【アメリカ人】
9440	女性	【アメリカ人】
9441	男性	【アメリカ人】
9442	女性	【アメリカ人】
9443	男性	【アメリカ人】
9444	女性	【アメリカ人】
9445	男性	【アメリカ人】
9446	女性	【アメリカ人】
9447	男性	【アメリカ人】
9448	女性	【アメリカ人】
9449	男性	【アメリカ人】
9450	女性	【アメリカ人】
9451	男性	【アメリカ人】
9452	女性	【アメリカ人】
9453	男性	【アメリカ人】
9454	女性	【アメリカ人】
9455	男性	【アメリカ人】
9456	女性	【アメリカ人】
9457	男性	【アメリカ人】
9458	女性	【アメリカ人】
9459	男性	【アメリカ人】
9460	女性	【アメリカ人】
9461	男性	【アメリカ人】
9462	女性	【アメリカ人】
9463	男性	【アメリカ人】
9464	女性	【アメリカ人】
9465	男性	【アメリカ人】
9466	女性	【アメリカ人】
9467	男性	【アメリカ人】
9468	女性	【アメリカ人】
9469	男性	【アメリカ人】
9470	女性	【アメリカ人】
9471	男性	【アメリカ人】
9472	女性	【アメリカ人】
9473	男性	【アメリカ人】
9474	女性	【アメリカ人】
9475	男性	【アメリカ人】
9476	女性	【アメリカ人】
9477	男性	【アメリカ人】
9478	女性	【アメリカ人】
9479	男性	【アメリカ人】
9480	女性	【アメリカ人】
9481	男性	【アメリカ人】
9482	女性	【アメリカ人】
9483	男性	【アメリカ人】
9484	女性	【アメリカ人】
9485	男性	【アメリカ人】
9486	女性	【アメリカ人】
9487	男性	【アメリカ人】
9488	女性	【アメリカ人】
9489	男性	【アメリカ人】
9490	女性	【アメリカ人】
9491	男性	【アメリカ人】
9492	女性	【アメリカ人】
9493	男性	【アメリカ人】
9494	女性	【アメリカ人】
9495	男性	【アメリカ人】
9496	女性	【アメリカ人】
9497	男性	【アメリカ人】
9498	女性	【アメリカ人】
9499	男性	【アメリカ人】
9500	女性	【アメリカ人】

例えば、この場合、話者の一人はidが【アメリカ人】であるので、

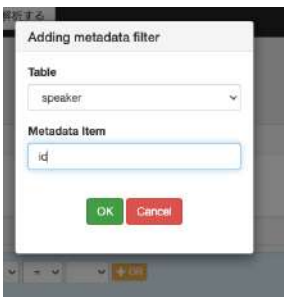
これを利用して、話者ごとに検索を実行する。



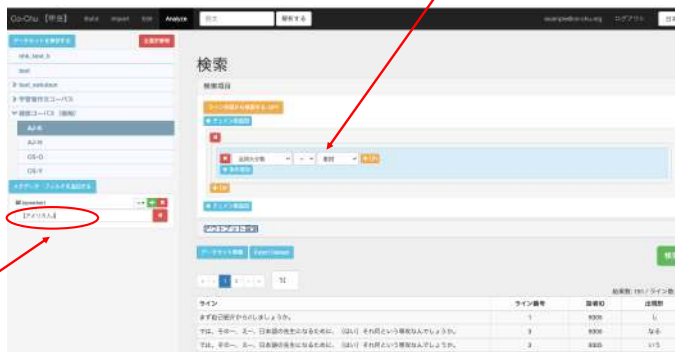
エクスポートされたエクセルファイル

編集 (= 取り込み後、今、システムにある) データ

編集前 (= 取込時の) データ



メタデータ・フィルタで Speaker情報のidを指定し、値として、【アメリカ人】を入れる



ここでは、【アメリカ人】のを使用した「動詞」を検索している。

## 5.2 検索比較機能

2つのコーパスの検索結果を並べて表示する時に使用する



アウトプット設定を「語を数える」にして、

検索比較モードを Off → On にする

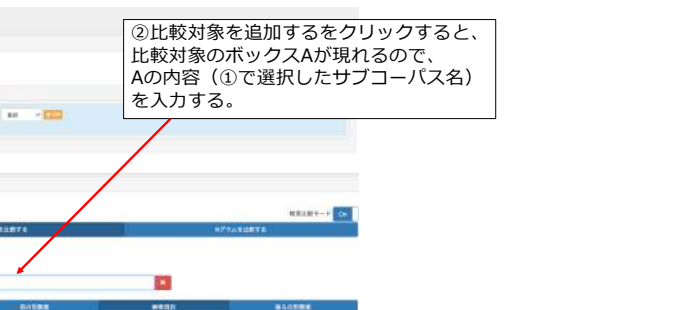


70

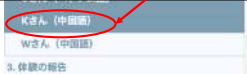
① 比較したいコーパスを選択する



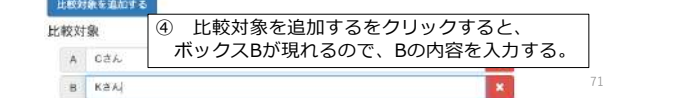
② 比較対象を追加するをクリックすると、比較対象のボックスAが現れるので、Aの内容（①で選択したサブコーパス名）を入力する。



③ 比較したいコーパスをもう一つ選択する。

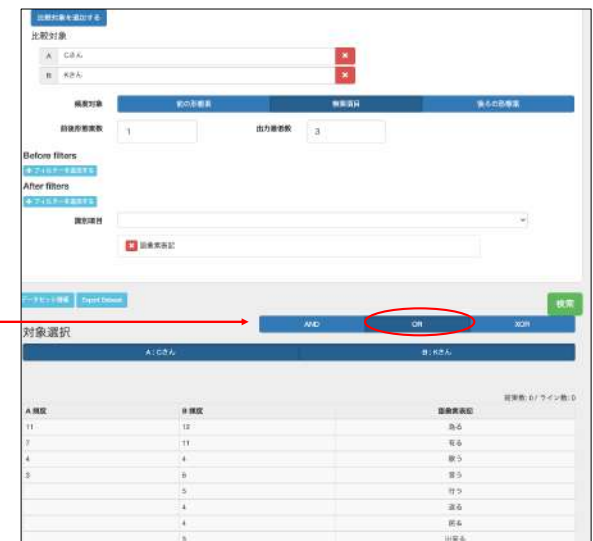


④ 比較対象を追加するをクリックすると、ボックスBが現れるので、Bの内容を入力する。



⑤ 検索 を実行する。

AND、OR、XOR で比較ができる。

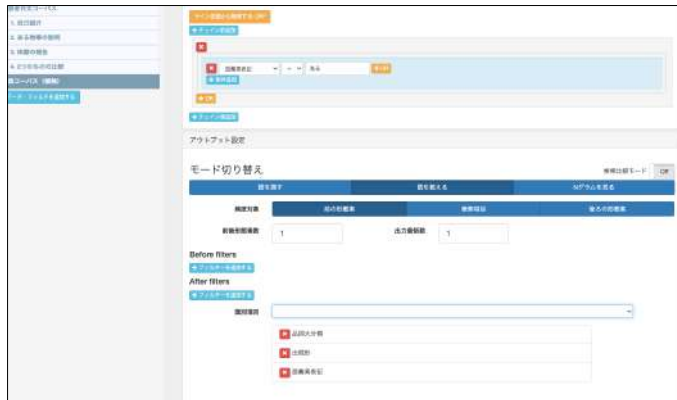


これは、OR検索の結果。  
Cさんの作文では「する」が11回  
Kさんの作文では「する」が12回  
用いられていることがわかる。

71

## 5.3 Before/After フィルター機能

検索結果をより絞って出力したい時に使用する



漢語動詞「\*\*する」において、  
どんな「\*\*」が多いのかを  
調べたい時、

「語を数える」で、「頻度対象」を  
「前の形態素」にして検索すれば、

「する」の前に使われている要素  
を数えることができる。



しかし、

73



頻度	品詞大分類	出現形	語彙表表記
6	助詞	を	を
5	名詞	緊張	緊張
4	名詞	安心	安心
2	名詞	期待	期待
2	名詞	準備	準備
2	名詞	研究	研究
2	名詞	見学	見学
2	動詞	願う	願う
2	名詞	卒業	卒業
2	助詞	と	と

このように、「する」の前になる全ての要素がカウントされる。  
漢語動詞だけを見たいので、「を」や「と」のような「助詞」は、除外したい場合には、  
Before フィルターを使用する。

74



前接する要素を、予め  
「名詞」に指定しておく。



結果には名詞のみが出力される。

頻度	品詞大分類	出現形	語彙表表記
5	名詞	緊張	緊張
4	名詞	安心	安心
2	名詞	期待	期待
2	名詞	準備	準備
2	名詞	研究	研究
2	名詞	見学	見学
2	名詞	卒業	卒業

75

## 参考文献

本間 妙・山本 裕子・川村 よし子・小森 早江子 (2019) 「コーパスシステムCo-Chuの作文指導への活用」『日本語教育方法研究会誌JLEM』26.1pp.62-63.

山本裕子・川村よし子・小森早江子・本間妙(2018)「話し言葉や誤用の含まれたテキストに対応可能なコーパス分析システムの開発」『2018年度日本語教育学会秋季大会予稿集』pp.295-300.日本語教育学会.

山本 裕子・本間妙・川村よし子 (2020) 「コーパス分析システムCo-Chuにおけるタグ検索機能とその活用 -誤用や話し言葉にどのように対応するか」『人文学部論集』43号,pp.1-24. 中部大学人文学部.

Wei, N., and Li, J. (2013) A new computing method for extracting contiguous phraseological sequences from academic text corpora. *International Journal of Corpus Linguistics* 18, 4, 506-535.

MeCab <https://taku910.github.io/mecab/> (2020年3月28日閲覧)

The UniDic Consortium. (2013) UniDic (Version 2.1.2). [Computer software]. Available from <https://osdn.net/projects/unidic/> (2020年3月28日閲覧)

本研究は科学研究費基盤研究 (C) 18K00723 (2018年~2020年度 研究代表者: 山本裕子) の助成を受けています。

76