

# 日本語コーパス研究ツールの現在

赤瀬川 史朗 (Lago 言語研究所 lagoinst@gmail.com)

## 1. コーパスから何が分かるか



## 2. コロケーション

### 2.1 コロケーションとは何か

イギリスの言語学者 John Firth (1890-1960)

You shall know a word by the company it keeps.

語というものは、その語がどんな語と一緒に使われるか見れば分かる。

→ 語というものは特定の語とよく使われる (=共起する) 傾向がある

...**collocation** is the phenomenon surrounding the fact that certain words are more likely to occur in combination with other words in certain contexts. A collocate is therefore a word which occurs within the neighbourhood of another word. (Paul Baker, et al. 2006)

**Collocates.** The Definition of regular or significant collocates was 'lexical items occurring within five words either way of the headword with a greater frequency than the law of averages would lead you to expect'. The importance of collocation was stressed and we were asked to record these items in dictionary examples. Function words such as prepositions, determiners, pronouns and auxiliaries were ignored, but delexical verbs were significant. Collocation was established only on the basis of corpus evidence. (Sinclair 1987)

### 2.2 コロケーションを調べる

▶ 動詞「流れる」を例にしてコロケーションを調べてみる

▶ 使用するコーパス

BCCWJ 領域公開データ (2009 年度版) のうち、以下のサブコーパスを使用

全体で約 6500 万語

サブコーパス	略号	語数
書籍サブコーパスの地の文	NR	45,018,820
書籍サブコーパスの会話文	SP	7,696,817
国会会議録	OM	5,041,424
Yahoo!知恵袋	OC	5,386,019
Yahoo!ブログ	OY	2,436,423
	総語数	<b>65,579,503</b>

※「書籍」は生産実態と流通実態サブコーパスの書籍を合わせたもの

※「総語数」は用例として抽出したセンテンスの総語数

▶上記のコーパス（6500万語）には「流れる」は4,724回現れる

### 2.2.1 それぞれの「流れる」の前後5語にどのような語がくるかを集計してみる

そのまんなかを川が西から東へ流れている。(吉田光男編『日韓中の交流』, 2004, 2 歴史)

- ▶単語に区切る：その|まんなか|を|川|が|西|から|東|へ|流れ|て|いる|。
- ▶前後5語を調べる：が|西|から|東|へ|**流れ**|て|いる|。
- ▶すべての用例についてこれを繰り返して、最後に集計する

共起語	合計	左	右	L5	L4	L3	L2	L1	R1	R2	R3	R4	R5
が	2802	2299	503	177	130	270	112	<b>1610</b>	11	107	106	166	113
て	2770	440	2330	65	81	70	21	203	<b>1979</b>	69	136	83	63
の	2447	1508	939	344	186	<b>802</b>	17	159	73	202	277	231	156
、	2422	1145	1277	339	239	<b>383</b>	58	126	373	181	235	258	230
に	1992	1457	535	168	176	226	74	<b>813</b>	4	167	88	158	118
。	1507	6	1501	2	2	1	0	1	127	<b>330</b>	423	420	201
た	1478	176	1302	59	44	70	3	0	<b>491</b>	132	492	114	73
を	1169	769	400	112	69	159	6	<b>423</b>	0	111	65	133	91
は	984	557	427	142	124	<b>144</b>	35	112	5	108	81	139	94
いる	669	30	639	13	11	5	1	0	0	<b>577</b>	16	18	28
と	669	417	252	82	77	88	16	<b>154</b>	38	36	55	59	64
で	615	359	256	50	52	69	28	<b>160</b>	1	55	42	96	62
から	457	376	81	58	56	89	13	<b>160</b>	4	12	26	21	18
い	424	10	414	5	4	1	0	0	0	378	10	10	16
よう	376	128	248	10	53	7	58	0	<b>144</b>	26	49	20	9
な	371	203	168	45	16	<b>136</b>	4	2	0	74	23	42	29
も	354	238	116	35	39	42	24	<b>98</b>	0	28	25	43	20
川	272	137	135	15	23	3	<b>96</b>	0	42	43	21	20	9
へ	228	212	16	11	10	23	44	<b>124</b>	0	1	3	4	8
水	224	156	68	10	28	8	<b>110</b>	0	25	9	18	8	8

## 2.2.2 コンコーダンスで調べる

▶ コンコーダンス (concordance) …キーワードを中心に並べて左右の文脈を確認できる

L5	L4	L3	L2	L1	Node	R1	R2	R3	R4	R5
			朝もや	の	流れる	中	に	、	法螺	の
尾	を	引く	よう	に	流れ	た	。			
籠り	、	村	々	に	流れ	た	の	で	ある	。
前	に	は	川	が	流れ	、	その	せせらぎ	は	都会
二	条	の	河	が	流れる	ところ	で	、	環境	的
その	娘	たる	光明皇后	に	流れ	て	いる	わけ	で	あつ
う	と	の	噂	が	流れ	た	。			
甘肅	省	南東	部	を	流れる	黄河	は	、	現在	、
n	形	を	描い	て	流れ	過ぎる	こと	に	なる	わけ
ほぼ	平行し	て	四百	キロ	流れる	が	、	両者	の	間
の	黄河	の	支流	が	流れ	て	いる	。		
の	間	を	抜け	て	流れ	て	ゆく	。		
は	真下	の	溝	に	流れる	。				
、	北	から	南	へ	流れ	て	い	た	。	
懐しん	だ	という	話	も	流れ	て	い	て	、	何だか
クーラー	の	風	が	強く	流れる	よう	な	所	に	、
			鼻	から	流れ	た	もの	が	、	やはり
水車	用	流水	溝	を	流れる	水	の	よう	に	速く
」	波止場	の	下	を	流れ	て	湾	へ	と	入っ

▶ 文前後を並べ替えることで特定の振る舞いを調べることができる

L5	L4	L3	L2	L1	Node	R1	R2	R3	R4	R5
前	に	は	川	が	流れ	、	その	せせらぎ	は	都会
二	条	の	河	が	流れる	ところ	で	、	環境	的
う	と	の	噂	が	流れ	た	。			
の	黄河	の	支流	が	流れ	て	いる	。		
から	実に	ながい	時	が	流れ	て	しまっ	て	いる	。
た	人生	の	時間	が	流れ	て	い	て	、	私
四	年	の	歳月	が	流れ	て	い	た	。	
を	横切っ	て	星	が	流れ	た	。			
			血	が	流れ	た	。			
部分	は	、	水	が	流れ	て	とかす	ので	、	まるで

▶ しかし、ガ格の名詞にどのようなものが多いかまでは簡単に分からない

L5	L4	L3	L2	L1	Node	R1	R2	R3	R4	R5
前	に	は	川	が	流れ	、	その	せせらぎ	は	都会
いう	巨大	な	川	が	流れ	て	いる	。		
走り	、	つまり	川	が	流れ	て	ゆく	の	意	で
たて	て	、	川	が	流れ	て	いる	。		
の	中央	を	川	が	流れ	て	いる	。		
に	は	大きな	川	が	流れ	て	い	まし	た	。
風景	に	は	川	が	流れ	て	い	た	。	
に	は	多治比	川	が	流れ	て	いる	。		
の	支流	中津	川	が	流れ	て	いる	。		
そこ	を	水無瀬	川	が	流れる	絶景	。	夕景	色	は

## 2.2.3 NINJAL-LWP for BCCWJ (NLB) で調べる

流れる 頻度=4,831

グループ別 パターン頻度順 基本

名詞+助詞

パターン	頻度	比率
...が流れる	2,243	
...は流れる	488	
...も流れる	119	
...の流れる	156	
...を流れる	418	
...に流れる	1,200	
...へ流れる	197	
...で流れる	368	
...と流れる	36	
...から流れる	451	
...まで流れる	45	
...より流れる	9	

名詞+複合助詞

パターン	頻度	比率
...として流れる	11	

「が流れる」 554件

コロケーション	頻度	MI	N-S
血が流れる	144	9.66	-2.31
川が流れる	97	8.79	1.18
水が流れる	92	7.52	-0.34
噂が流れる	87	10.11	-1.02
【人名】が流れる	68	1.35	0.73
涙が流れる	61	8.93	1.13
歳月が流れる	60	11.86	0.95
時間が流れる	54	5.54	0.27
空気が流れる	54	8.80	1.15
音楽が流れる	53	8.05	0.45
声が流れる	51	5.92	0.95
電流が流れる	51	11.40	1.06
情報が流れる	44	6.19	-0.40
沈黙が流れる	42	10.12	0.95
曲が流れる	36	7.87	0.52
汗が流れる	34	8.63	0.75
【一般】が流れる	28	2.52	0.36
マッシュアップが流れる	26	11.84	0.45

血が流れる

- 書籍・地の文: 2.53
- 書籍・会話文: 4.84
- 国会: 0.2
- 知恵袋: 0.37
- ブログ: 0.83

血が流れた。  
(岸見信吾『地底ドドン(男)』, 1986, 9 文学)

血が、流れていた。  
(前田珠子著『柘榴の影』, 1991, 9 文学)

血がだらだら流れてる  
(剛しいら著『別れごっこ』, 2005, 9 文学)

血が流れはじめている。  
(喜多嶋隆著『シエームス・ディーンに、似ていた』, 1988, 9 文学)

唇からも、血が流れてる。  
(松岡やよい著『いつてきます!』, 2002, 9 文学)

多くの血が流れましょう。  
(宇佐美浩然著『中国遊侠伝』, 2003, 9 文学)

カンには黒い血が流れている。  
(真木健一著『海軍座に捧ぐ百一発の砲声』, 1997, 9 文学)

運平の手から血がながれました。  
(原稿十原作/六石好文構成/山上製書文『むくはとじゅうの名犬物語』, 1997, 分類なし)

NINJAL-LWP for BCCWJ Copyright © 2012 National Institute for Japanese Language and Linguistics, Lago Institute of Language. All rights reserved.

レキシカルプロファイリング…語の振る舞いを文法パターン別に整理して表示

→より人間の感覚に近い形でコロケーションを調べることができる (パルデシ・赤瀬川 2011)

▶ ガ格と「流れる」が離れていても検索できる → コンコーダンスで調べるのはかなり難しい

川が流れる

NR: 1.87 SP: 0.71 OM: 0.6 OC: 0.55 OY: 1.22

そのまんなかを川が西から東へ流れている。  
(吉田光男編『日韓中の交流』, 2004, 2 歴史)

天神山と青光山の間には多治比川が流れている。  
(童門冬二著『小説毛利元就』, 1996, 9 文学)

それは満州の兵営で、兵営の中に川が流れていた。  
(飯塚浩二著『日本の軍隊』, 2003, 3 社会科学)

その西のふもとを信濃川の支流中津川が流れている。  
(市川健夫監修・著『雪国のくらし』, 1997, 分類なし)

エルベ河が廃墟のなか、はっきりと流れ続けている。  
(柘植久慶著『逆撃第三帝国の崩壊』, 1998, 9 文学)

砂に川が流れて込んで、水が消えたことを言うのか？  
(ジョセフィン・テイ著/壁野佐和子訳『歌う砂』, 2005, 9 文学)

川が流れてて南アルプスの山山があるあの風景です。  
(Yahoo!知恵袋, 2005, 地域、旅行、お出かけ)

Page 1 / 1 97件中 1 - 97を表示

### 3. NINJAL-LWP for BCCWJ (NLB) とは

- ▶ 国立国語研究所が開発した『現代日本語書き言葉均衡コーパス』（BCCWJ）を検索するためのツール
- ▶ Lago 言語研究所が開発した LWP (LagoWordProfiler) をベースとして、国語研と Lago 言語研が共同開発
- ▶ 使用しているコーパスは、2.2 で示したもの

### 4. NLB の特長

#### ❖ ユーザーフレンドリーなインターフェース

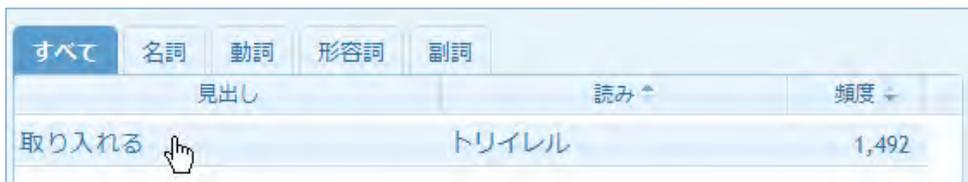
- クリックするだけでコロケーションを調べられる
- 文法パターン、コロケーション、用例、文脈などを一画面で確認できる

#### ❖ 瞬時に結果が返ってくる

- 見出し語を選んで調べたいパターンをクリックするだけ

#### ❖ 複数の表記をまとめた見出し語で検索

- 「取り入れる」という見出し語から他の表記のものも同時に検索できる



▼ 書字形		
取り入れる	1183	79%
とり入れる	125	8%
採り入れる	81	5%
とりいれる	77	5%
取りいれる	12	1%
取入れる	8	1%
摂り入れる	3	0%
採りいれる	3	0%

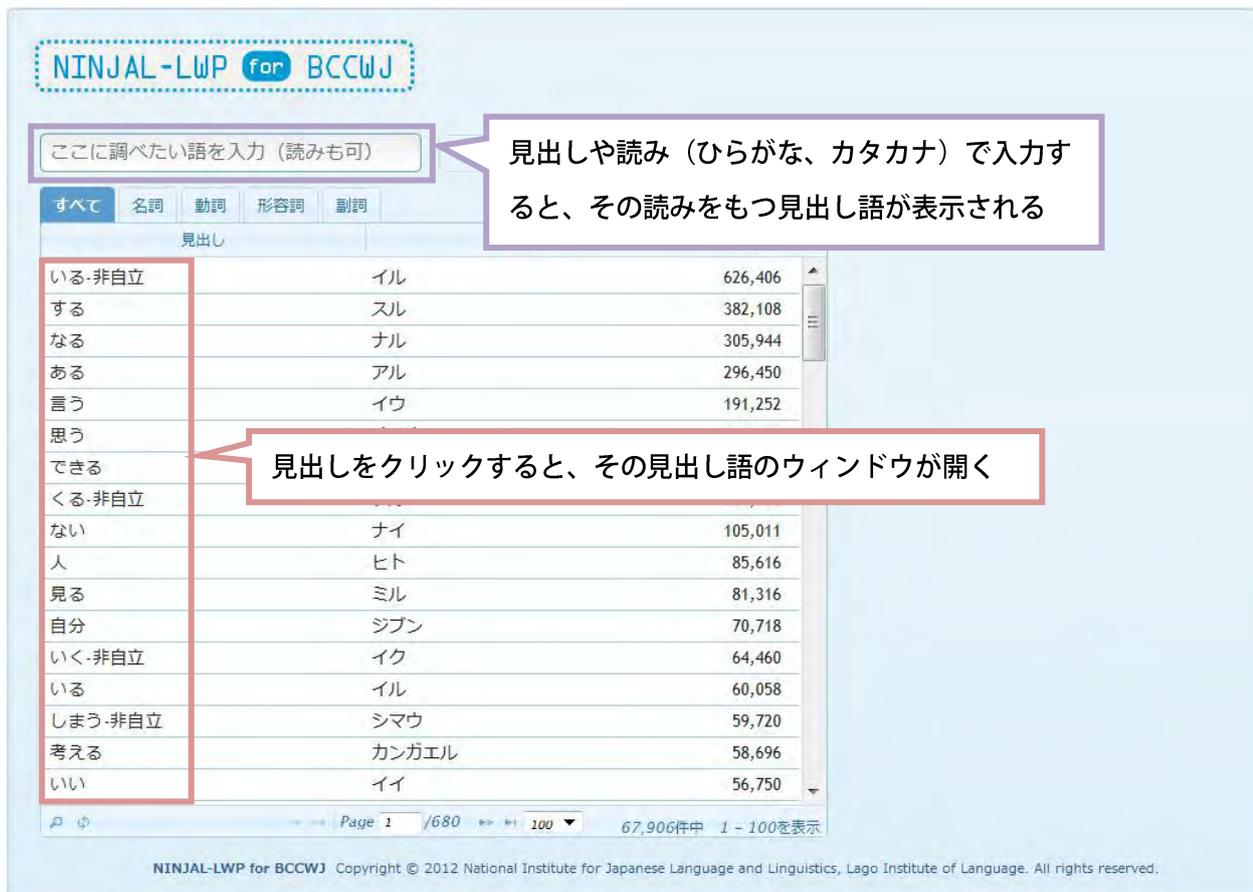
技術を取り入れる  
■ NR: 0.38 ■ SP: 0.18 ■ OM: 0 ■ OC: 0.18 ■ OY: 0

- 社史に自社製品や新技術を取り入れた例は、まだ少ない。  
(村橋勝子著 『社史の研究』, 2002, 3 社会科学)
- 明治以降、日本は死にもの狂いで外国の技術を取り入れた。  
(唐津一著 『儲かる企業の新常識』, 2003, 5 技術・工学)
- 社史の制作に自社製品や技術を採用すれば、別の意義も見出せる。  
(村橋勝子著 『社史の研究』, 2002, 3 社会科学)
- 基本的な技術は日本の技術を取り入れているから問題はないと思います。  
(Yahoo!知恵袋, 2005, スポーツ、アウトドア、車)
- 鉄砲や、大船などの新技術を取り入れたことが然り、情報や謀略を重視したことも然り。  
(遠藤周作ほか著 『たかが信長されど信長』, 1992, 2 歴史)
- 自動車においても新しいモデルや新しい技術を取り入れた車種が次々と市場に出されている。

Page 1 / 1 | 100 | 19件中 1 - 19を表示

## 5. NLB の機能

### 5.1 見出し語検索ウィンドウ



NINJAL-LWP for BCCWJ

ここに調べたい語を入力 (読みも可)

見出しや読み (ひらがな、カタカナ) で入力すると、その読みをもつ見出し語が表示される

すべて 名詞 動詞 形容詞 副詞

見出し

いる-非自立	イル	626,406
する	スル	382,108
なる	ナル	305,944
ある	アル	296,450
言う	イウ	191,252
思う		
できる		
くる-非自立		
ない	ナイ	105,011
人	ヒト	85,616
見る	ミル	81,316
自分	ジブン	70,718
いく-非自立	イク	64,460
いる	イル	60,058
しまう-非自立	シマウ	59,720
考える	カンガエル	58,696
いい	イイ	56,750

見出しをクリックすると、その見出し語のウィンドウが開く

Page 1 / 680 67,906件中 1 - 100を表示

NINJAL-LWP for BCCWJ Copyright © 2012 National Institute for Japanese Language and Linguistics, Lago Institute of Language. All rights reserved.

## 5.2 見出し語ウィンドウ

聞く 35706

グループ別    パターン頻度順    基本

名詞 + 助詞

パターン	頻度	比率
...が聞く	2,241	
...は聞く	4,104	
...も聞く	1,387	
...の聞く	126	
...を聞く	10,423	
...に聞く	4,141	
...へ聞く	41	
...で聞く	2,403	
...と聞く	486	
...から聞く	1,686	
...まで聞く	259	
...より聞く	55	

名詞 + 複合助詞

パターン	頻度	比率
...として聞く	43	
...において聞く	1	
...にかけて聞く	2	
...に沿って聞く	1	
...に対して聞く	1	
...について聞く	99	

...を聞く 1048件

コロケーション	FQ	LD	MI	N-S
話を聞く	2,099	10.80	8.83	-31.48
ごとを聞く	899	5.95	3.55	-17.91
意見を聞く	540	9.88	8.61	0.13
それを聞く	518	6.96	4.65	-4.28
声を聞く	497	8.89	6.99	-0.73
言葉を聞く	406	8.49	6.53	2.00
のを聞く	352	1.81	-0.61	-2.62
音を聞く	249	8.49	6.94	1.69
お話を聞く	186	8.81	8.48	-3.87
音楽を聞く	176	8.97	7.55	-1.42
これを知る	173	5.56	3.27	1.40
説明を聞く	122	7.77	6.56	0.11
事情を聞く	107	7.84	7.07	-2.31
名前を聞く	105	7.46	6.14	-1.47
噂を聞く	93	7.91	7.97	-1.33
理由を聞く	84	6.94	5.42	-2.02
何を聞く	72	4.69	2.44	-2.64
報告を聞く	72	7.17	6.21	-1.38
歌を聞く	69	7.23	6.50	-0.80

話を聞く

NR: 31.39    SP: 62.87    OM: 23.04    OC: 30.26    OY: 21.23

話を聞くこと  
(小畑勇二郎顕彰会編『大いなる秋田を』, 2001, 2 歴史)

話を聞いた私は  
(柴田鉄治著『新聞記者という仕事』, 2003, 0 雑記)

話を聞かせて。  
(赤川次郎著『幽霊教会』, 1997, 9 文字)

人の話を聞けー!  
(大槻はぢめ著『公衆的恋のススメ』, 2005, 9 文字)

話を聞かんかい!  
(ポール・スチュワート作; クリス・リデル絵; 岩井明子訳『嵐を追う者たち』, 2001, 分類なし)

こんな話を聞いた。  
(阿刀田高著『風の組曲』, 2004, 9 文字)

同級生に話を聞く。  
(MMプロジェクト編『モーニング娘。21人の足跡』, 2003, 7 芸術・美術)

話を聞いたらどうが  
(菊地秀行著『魔界都市ブルース』, 1992, 9 文字)

Page 1 / 21    100    2,099件中 1 - 100を表示

グループ別    **パターン頻度順**    基本

名詞 + 助詞

パターン	頻度	比率
...が聞く	2,241	
...は聞く	4,104	
...も聞く	1,387	
...の聞く	126	
...を聞く	10,423	
...に聞く	4,141	
...へ聞く	41	
...で聞く	2,403	
...と聞く	486	
...から聞く	1,686	
...まで聞く	259	
...より聞く	55	

名詞 + 複合助詞

パターン	頻度	比率
...として聞く	43	
...において聞く	1	
...にかけて聞く	2	
...に沿って聞く	1	
...に対して聞く	1	
...について聞く	99	

文法パターンをグループごとに表示

文法パターンの頻度の高い順に表示

グループ別    **パターン頻度順**    基本

パターン	頻度	比率
...を聞く	10,423	
聞いた	7,879	
聞いて + 動詞	7,465	
聞く ⇨ 動詞	7,422	
動詞 ⇨ 聞く	6,691	
副詞 ⇨ 聞く	4,982	
聞きた + 名詞	4,412	
...に聞く	4,141	
...は聞く	4,104	
聞いている[ある]	2,919	
...で聞く	2,403	
聞く + 名詞	2,370	
...が聞く	2,241	
聞きます	1,808	
聞かせる	1,710	
...から聞く	1,686	
聞かれる	1,432	
聞き + 自立動詞	1,430	
...も聞く	1,387	
(未分類)	1,375	
聞きたい	1,051	
動詞 + て聞く	944	

▼ 書字形		
問く	31170	87%
きく*(効く;問く;利く)	4499	
訊く	2427	7%
聴く	2087	6%
尋く	22	0%

▼ サブコーパスごとの書字形分布		
<b>書籍-地の文</b>		
問く	15907	82%
きく*(効く;問く;利く)	3370	
訊く	1992	10%
聴く	1375	7%
尋く	16	0%
<b>書籍-会話文</b>		
問く	4377	88%
きく*(効く;問く;利く)	775	
訊く	369	7%
聴く	216	4%
尋く	6	0%
<b>国会会議録</b>		
問く	4615	100%
きく*(効く;問く;利く)	27	
聴く	1	0%
訊く	0	0%
尋く	0	0%
<b>Yahoo!知恵袋</b>		
問く	5318	95%
きく*(効く;問く;利く)	274	
聴く	254	5%
訊く	54	1%
尋く	0	0%
<b>Yahoo!ブログ</b>		
問く	953	79%
聴く	241	20%
きく*(効く;問く;利く)	53	
訊く	12	1%
尋く	0	0%

グループ別	パターン頻度順	基本
▼ サブコーパス		
書籍-地の文	19290	428.5
	80%	
書籍-会話文	4968	882.37
	20%	
国会会議録	4616	916.91
Yahoo!知恵袋	5626	1038.18
Yahoo!ブログ	1206	492.33
<a href="#">▶ 書字形</a> <a href="#">▶ サブコーパスごとの書字形分布</a> <a href="#">▶ 活用形</a> <a href="#">▶ 後続助動詞の割合</a>		

▼ 活用形		
連用夕接続	18849	53%
連用形	6708	19%
基本形	5248	15%
未然形	4070	11%
仮定形	426	1%
未然ウ接続	316	1%
命令 e	86	0%
仮定縮約 1	3	0%

▼ 後続助動詞の割合		
れる・られる	1432	4.0%
聞かれる	1432	
せる・させる	1710	4.8%
聞かせる	1710	
ない・ぬ	964	2.7%
聞かない	725	
聞かぬ	168	
聞きません	71	

パターン	頻度	比率
...が聞く	2,241	
...は聞く	4,104	
...も聞く	1,387	
...の聞く	126	
...を聞く	10,423	
...に聞く	4,141	
...へ聞く	41	
...で聞く	2,403	
...と聞く	486	
...から聞く	1,686	
...まで聞く	259	
...より聞く	55	

文法パターン、コロケーション、用例、文脈が連動して一つの画面で確認できる

...を聞く 1048件				
コロケーション	FQ	LD	MI	N-S
話を聞く	2,099	10.80	8.83	-31.48
ことを聞く	899	5.95	3.55	-17.91
意見を聞く	540	9.88	8.61	0.13
それを聞く	518	6.96	4.65	-4.28
声を聞く	497	8.89	6.99	-0.73
言葉を聞く	406	8.49	6.53	2.00
のを聞く	352	1.81	-0.61	-2.62
音を聞く	249	8.49	6.94	1.69
お話を聞く	186	8.81	8.48	-3.87
音楽を聞く	176	8.49	7.56	-1.42
これを聞く	173	5.56	3.27	1.40
説明を聞く	122	7.77	6.56	0.11
事情を聞く	107	7.84	7.07	-2.31
名前を聞く	105	7.46	6.14	-1.47
噂を聞く	93	7.91	7.97	-1.33
理由を聞く	84	6.94	5.42	-2.02
何を聞く	72	4.69	2.44	-2.64
報告を聞く	72	7.17	6.21	-1.38
歌を聞く	69	7.23	6.50	-0.80
曲を聞く	63	7.09	6.33	-0.53
様子を聞く	50	6.60	5.57	-0.25
名を聞く	48	5.91	4.21	0.00
感想を聞く	48	7.08	7.95	-0.11
ニュースを聞く	45	6.86	6.91	-0.73

噂を聞く  
 ■ NR: 1.33 ■ SP: 2.66 ■ OM: 0.2 ■ OC: 2.95 ■ OY: 0.41

そこで噂を聞いてね  
(木谷恭介著『謀殺列島赤の殺人事件』, 1995, 9 文学)

大阪で文展の噂を聞いた。  
(連城三紀彦, 伊集院静, 澤田ふじ子著『吉川英治文学新人賞篇』, 2004, 9 文学)

開港した横浜の噂を聞いているか。  
(秋山香乃著『五稜郭を落した男』, 2004, 9 文学)

---そんなうわさを聞きました。  
(Yahoo!知恵袋, 2005, 暮らしと生活ガイド)

そういえば噂を聞いたことがありますよ。  
(別所真紀子著『残る蜚』, 2004, 9 文学)

3 噂を聞いて、多くの人々がここに集まる。  
(情報処理推進機構編著『日本の技、日本の匠』, 2005, 0 総記)

未亡人も、その噂を聞いて当然動揺しました。  
(上坂冬子著『償いは済んでいる』, 1995, 9 文学)

八兵衛のことは公家衆からも噂を聞いている。  
(濱邊勝利著『禁中忍者伝承』, 2000, 9 文学)

あっしも『石塔磨』の噂を聞いたんでござんす。  
(六道慧著『おぼろ隠密記』, 2001, 9 文学)

近頃ではおとなしくしているらしく、噂を聞かない。

Page 1 / 1 100 93件中 1 - 93

連城三紀彦, 伊集院静, 澤田ふじ子著『吉川英治文学新人賞篇』, 2004, 9 文学

一人になった部屋で、私はやはり黙ってその絵を見守り続けていた。私は感動のために何もわからなかった。我にかえると、絵には斜めに大きく赤い筋がひかれ、私の震えた手が握った鉛筆から、真紅の絵具が滴り落ちていた。それが自分の手だとは、すぐには気がつかなかった。白河の戻る前に部屋を出ると、簡単に荷物をまとめ、その夜のうちに私は東京を逃げだしたのだった。

大阪で文展の噂を聞いた。白河の名は出なかった。その後も白河が画壇に登場したという話は聞かない。理由はわかっていて。命を削るような思いまでして描いた絵を、たった一本の赤い線で葬られてしまった白河には二つの道しかなかった筈だ。死ぬか、私のように鉛筆を絶つか—事実、あの赤い筋は、私が白河の命に切りつけて出来た傷口であった。

閉じる

## 6. NLB のコーパスデータの処理

毎日毎日、家族の将来を考えています。(西出真由美著『がんばって！っていわないで』, 2003, 9 文学)

### 将来を考える

■ 書籍・地の文: 0.95 ■ 書籍・会話文: 0.72  
■ 国会: 2.22 ■ 知恵袋: 1.87 ■ ブログ: 0.42

■ 毎日毎日、家族の将来を**考え**ています。  
(西出真由美著『がんばって！っていわないで。』, 2003, 9 文学)

### 考えている

■ 書籍・地の文: 108.23 ■ 書籍・会話文: 134.38  
■ 国会: 188.77 ■ 知恵袋: 145.8 ■ ブログ: 75.86

■ 毎日毎日、家族の将来を**考え**ています。  
(西出真由美著『がんばって！っていわないで。』, 2003, 9 文学)

### 家族の将来

■ 書籍・地の文: 0.07 ■ 書籍・会話文: 0  
■ 国会: 0 ■ 知恵袋: 0 ■ ブログ: 0

■ 毎日毎日、**家族の将来**を考えています。  
(西出真由美著『がんばって！っていわないで。』, 2003, 9 文学)

- ・形態素解析辞書の拡張（代表表記に対応）
- ・センテンス（用例）データの作成
- ・形態素・係り受け解析
- ・後処理（サ変動詞・ナ形容詞）
- ・全品詞の頻度表作成
- ・内容語の頻度表作成
- ・内容語の見出し語リスト作成
- ・見出し語ごとの用例データ作成
- ・文法パターンの抽出
- ・データベース登録用ファイルの作成

## 6.1 形態素解析辞書の拡張

### オリジナルの IPA 辞書

やっぱ,1281,1281,6053,副詞,一般,\*,\*,\*,\*やっぱ,ヤツパ,ヤツパ  
 やっぱり,1281,1281,6026,副詞,一般,\*,\*,\*,\*やっぱり,ヤツパリ,ヤツパリ  
 やっぱし,1281,1281,6048,副詞,一般,\*,\*,\*,\*やっぱし,ヤツパシ,ヤツパシ  
 やはり,1281,1281,5999,副詞,一般,\*,\*,\*,\*やはり,ヤハリ,ヤハリ  
 やっぱ,1281,1281,6053,副詞,一般,\*,\*,\*,\*やっぱ,ヤツパ,ヤツパ  
 やっぱり,1281,1281,6026,副詞,一般,\*,\*,\*,\*やっぱり,ヤツパリ,ヤツパリ  
 やっぱし,1281,1281,6048,副詞,一般,\*,\*,\*,\*やっぱし,ヤツパシ,ヤツパシ  
 やはり,1281,1281,5999,副詞,一般,\*,\*,\*,\*やはり,ヤハリ,ヤハリ

### 代表表記に対応するように拡張した IPA 辞書

やっぱ,1281,1281,6053,副詞,一般,\*,\*,\*,\*やっぱ,ヤツパ,ヤツパ,やはり,ヤハリ,\*,\*  
 やっぱし,1281,1281,6048,副詞,一般,\*,\*,\*,\*やっぱし,ヤツパシ,ヤツパシ,やはり,ヤハリ,\*,\*  
 やっぱり,1281,1281,6026,副詞,一般,\*,\*,\*,\*やっぱり,ヤツパリ,ヤツパリ,やはり,ヤハリ,\*,\*  
 やつぱり,1281,1281,5999,副詞,一般,\*,\*,\*,\*やつぱり,ヤツパリ,ヤツパリ,やはり,ヤハリ,\*,\*  
 やはり,1281,1281,5999,副詞,一般,\*,\*,\*,\*やはり,ヤハリ,ヤハリ,やはり,ヤハリ,\*,\*  
 ヤツパリ,1281,1281,5999,副詞,一般,\*,\*,\*,\*ヤツパリ,ヤツパリ,ヤツパリ,やはり,ヤハリ,\*,\*  
 矢っ張,1281,1281,5999,副詞,一般,\*,\*,\*,\*矢っ張,ヤツパリ,ヤツパリ,ヤツパリ,,やはり,ヤハリ,\*,\*  
 矢っ張り,1281,1281,5999,副詞,一般,\*,\*,\*,\*矢っ張り,ヤツパリ,ヤツパリ,やはり,ヤハリ,\*,\*  
 矢つ張り,1281,1281,5999,副詞,一般,\*,\*,\*,\*矢つ張り,ヤツパリ,ヤツパリ,やはり,ヤハリ,\*,\*  
 矢ッ張り,1281,1281,5999,副詞,一般,\*,\*,\*,\*矢ッ張り,ヤツパリ,ヤツパリ,やはり,ヤハリ,\*,\*  
 矢ツぱり,1281,1281,5999,副詞,一般,\*,\*,\*,\*矢ツぱり,ヤツパリ,ヤツパリ,やはり,ヤハリ,\*,\*  
 矢張り,1281,1281,5999,副詞,一般,\*,\*,\*,\*矢張り,ヤハリ,ヤハリ,やはり,ヤハリ,\*,\*

▼ 書字形		
やはり	13918	68%
やっぱり	5750	28%
やっぱ	539	3%
矢張り	62	0%
やっぱし	30	0%
やつぱり	14	0%
ヤツパリ	9	0%
矢っ張り	5	0%
矢っ張り	4	0%
矢ッ張り	2	0%

## 6.2 センテンス（用例）データの作成

OY01\_00105|S2|OY ファイル ID|センテンス ID|サブコーパス

彼女は、2人でも大勢と表現するので、真偽の程が定かでないが、これで鼻をかむ客が気になる程度の数、出現したという事は間違いないのだろう。

## 6.3 形態素・係り受け解析（MeCab+Cabocha）

OY01\_00105|S2|OY

\* 0 6D 0/1 0.800369

彼女 名詞,代名詞,一般,\*,\*,\*,彼女,カノジョ,カノジョ,彼女,カノジョ,\*,混

は 助詞,係助詞,\*,\*,\*,は,ハ,ワ,...

、 記号,読点,\*,\*,\*,、,、,、,、,1,、

\* 1 3D 0/1 2.242968

2人 名詞,副詞可能,\*,\*,\*,2人,フタリ,フタリ,二人,フタリ,\*,和

でも 助詞,副助詞,\*,\*,\*,でも,デモ,デモ,...

\* 2 3D 0/1 5.229882

大勢 名詞,一般,\*,\*,\*,大勢,オオゼイ,オーゼイ,大勢,オオゼイ,\*,\*

と 助詞,格助詞,一般,\*,\*,\*,と,ト,ト,...

\* 3 6D 1/2 0.433439

表現 名詞,サ変接続,\*,\*,\*,表現,ヒョウゲン,ヒョーゲン,表現,ヒョウゲン,\*,\*

する 動詞,自立,\*,\*,サ変・スル,基本形,する,スル,スル,する-スル-サ変・スル,スル,する,\*,和

ので 助詞,接続助詞,\*,\*,\*,ので,ノデ,ノデ,...

、 記号,読点,\*,\*,\*,、,、,、,、,1,、

\* 4 5D 0/1 2.698557

真偽 名詞,一般,\*,\*,\*,真偽,シンギ,シンギ,真偽,シンギ,\*,\*

の 助詞,連体化,\*,\*,\*,の,ノ,ノ,...

\* 5 6D 0/1 2.595211

程 名詞,非自立,一般,\*,\*,\*,程,ホド,ホド,ほど,ホド,\*,\*

が 助詞,格助詞,一般,\*,\*,\*,が,ガ,ガ,...

\* 6 18D 0/3 0.000000

定か 名詞,形容動詞語幹,\*,\*,\*,定か,サダカ,サダカ,定か,サダカ,\*,和

で 助動詞,\*,\*,特殊・ダ,連用形,だ,デ,デ,...

ない 助動詞,\*,\*,特殊・ナイ,基本形,ない,ナイ,ナイ,...

が 助詞,接続助詞,\*,\*,\*,が,ガ,ガ,...

、 記号,読点,\*,\*,\*,、,、,、,、,1,、

## 6.4 後処理（サ変動詞・ナ形容詞）

OY01\_00105|S2|OY

\* 0 6D 0/1 0.800369

彼女 名詞,代名詞,一般,\*,\*,\*,彼女,カノジョ,カノジョ,彼女,カノジョ,\*,混

は 助詞,係助詞,\*,\*,\*,は,ハ,ワ,...

、 記号,読点,\*,\*,\*,、 、 、 ,,1,,

\* 1 3D 0/1 2.242968

2人 名詞,副詞可能,\*,\*,\*,2人,フタリ,フタリ,二人,フタリ,\*和

でも 助詞,副助詞,\*,\*,\*,でも,デモ,デモ,,,

\* 2 3D 0/1 5.229882

大勢 名詞,一般,\*,\*,\*,大勢,オオゼイ,オーゼイ,大勢,オオゼイ,\*,\*

と 助詞,格助詞,一般,\*,\*,\*,と,ト,ト,,,

\* 3 6D 1/2 0.433439

表現する 動詞,自立,\*,\*,\*,サ変・ースル,基本形,表現する,ヒョウゲンスル,ヒョーゲンスル,表現する-スル-サ変・ースル,表現する,表現する,\*,\*

ので 助詞,接続助詞,\*,\*,\*,ので,ノデ,ノデ,,,

、 記号,読点,\*,\*,\*,、 、 、 ,,1,,

\* 4 5D 0/1 2.698557

真偽 名詞,一般,\*,\*,\*,真偽,シンギ,シンギ,真偽,シンギ,\*,\*

の 助詞,連体化,\*,\*,\*,の,ノ,ノ,,,

\* 5 6D 0/1 2.595211

程 名詞,非自立,一般,\*,\*,\*,程,ホド,ホド,ほど,ホド,\*,\*

が 助詞,格助詞,一般,\*,\*,\*,が,ガ,ガ,,,

\* 6 18D 0/3 0.000000

定かで ナ形容詞,\*,\*,\*,ナ形容詞,連用形・ダ,定かだ,サダカデ,サダカデ,定かだ-サダカダ-ナ形容詞,サダカダ,定かで

ない 助動詞,\*,\*,\*,特殊・ナイ,基本形,ない,ナイ,ナイ,,,

が 助詞,接続助詞,\*,\*,\*,が,ガ,ガ,,,

、 記号,読点,\*,\*,\*,、 、 、 ,,1,,

## 6.5 全品詞の頻度表作成

、 記号	、	3352681
の 助詞	ノ	2817774
。 記号	。	2464261
に 助詞	ニ	2050492
は 助詞	ハ	1914313
を 助詞	ヲ	1814233
て 助詞	テ	1720566
た 助動詞	タ	1712662
が 助詞	ガ	1648285
だ 助動詞	ダ	1159749

## 6.6 内容語の頻度表作成

順位|見出し|読み|全頻度|累積頻度割合|サブコーパス頻度(PMW)|表記ごとの頻度|活用形頻度|れる・られる接続|せる・させる接続|ない・ぬ・ません接続

1 いる-イル-一段-非自立 イル 626481 8.49% NR: 475337 (10744.72), SP: 48723 (8730.04), OM: 33197 (6686.27), OC: 50005 (9371.44), OY: 19219 (8010.41) いる: 626481 (NR: 475337, SP: 48723, OM: 33197, OC:

50005, OY: 19219,) 基本形: 353015, 連用形: 236973, 未然形: 32602, 假定形: 3267, 未然ウ接続: 493, 命令 r o: 104, 假定縮約 1: 10, 命令 y o: 9, 体言接続特殊: 8 0.2% (いられる: 1075, 入れる: 22) 0.0% (いさせる: 3) 5.6% (いない: 31352, いません: 3608, いぬ: 87)

2 する-スル-サ変・スル スル 381786 13.67% NR: 260251 (5882.82), SP: 37084 (6644.6), OM: 33934 (6834.71), OC: 38360 (7189.05), OY: 12157 (5066.99) する: 381786 (NR: 260251, SP: 37084, OM: 33934, OC: 38360, OY: 12157,) 連用形: 222826, 基本形: 89950, 未然レル接続: 24601, 未然形: 20103, 假定形: 8488, 未然ウ接続: 6182, 未然ヌ接続: 4534, 文語基本形: 2641, 命令 y o: 2047, 体言接続特殊: 186, 假定縮約 1: 100, 命令 r o: 70, 命令 i: 58 4.8% (される: 18201, しられる: 15, さられる: 2) 1.5% (させる: 5909, ささせる: 1, しさせる: 1) 5.6% (しない: 19855, しません: 1411, しぬ: 27)

3 なる-ナル-五段・ラ行(成る;生る;鳴る;なる) ナル 305951 17.82% NR: 215882 (4879.89), SP: 27728 (4968.22), OM: 23721 (4777.69), OC: 27637 (5179.45), OY: 10983 (4577.67) なる: 305951 (NR: 215882, SP: 27728, OM: 23721, OC: 27637, OY: 10983,) 連用タ接続: 138863, 基本形: 95999, 連用形: 51603, 未然形: 12631, 假定形: 4320, 未然ウ接続: 2029, 命令 e: 188, 体言接続特殊: 146, 未然特殊: 113, 假定縮約 1: 57, 体言接続特殊 2: 2 0.2% (なられる: 588) 0.0% (ならせる: 34, ならさせる: 2) 4.1% (ならない: 8656, ならぬ: 2786, なりません: 1147)

## 6.7 内容語の見出し語リスト作成

いる-イル-一段-非自立 V.00001  
 する-スル-サ変・スル V.00002  
 なる-ナル-五段・ラ行(成る;生る;鳴る;なる) V.00003  
 ある-アル-五段・ラ行 V.00004  
 言う-イウ-五段・ワ行促音便 V.00005  
 :  
 表現する-スル-サ変・スル V.00347

## 6.8 見出し語ごとの用例データ作成

表現する (V.00347)  
 OY01\_00105 S2 OY 名詞\_彼女\_彼女\_彼女\_カノジョ\_代名詞\_\*\_彼女\_カノジョ\_1\_0:6 助詞\_は\_は\_は\_ハ\_係助詞\_\*\_0:6 記号\_、\_、\_、\_、\_読点\_\*\_1\_0:6 名詞\_二人\_二人\_二人\_フタリ\_副詞可能\_\*\_二人\_フタリ\_1\_1:3 助詞\_でも\_でも\_でも\_デモ\_副助詞\_\*\_1:3 名詞\_大勢\_大勢\_大勢\_オオゼイ\_一般\_\*\_大勢\_オオゼイ\_1\_2:3 助詞\_と\_と\_と\_ト\_格助詞\_\*\_2:3 動詞\_表現する-スル-サ変・スル\_表現する\_表現する\_ヒョウゲンスル\_自立\_基本形\_表現する-スル-サ変・スル\_表現する\_表現する\_3:6 助詞\_の\_の\_の\_ノ\_デ\_接続助詞\_\*\_3:6 記号\_、\_、\_、\_、\_読点\_\*\_1\_3:6 名詞\_真偽\_真偽\_真偽\_シンギ\_一般\_\*\_真偽\_シンギ\_1\_4:5 助詞\_の\_の\_の\_ノ\_連体化\_\*\_4:5 名詞\_ほど\_程\_程\_ホド\_非自立\_\*\_ほど\_ホド\_1\_5:6 助詞\_が\_が\_が\_ガ\_格助詞\_\*\_5:6 ナ形容詞\_定かだ\_定かだ\_定かだ\_サダカデ\_\*\_連用形\_ダ\_定かだ-サダカダ-ナ形容詞\_サダカダ\_定かだ\_6:18 助動詞\_ない\_ない\_ない\_ナイ\_\*\_基本形\_6:18 助詞\_が\_が\_ガ\_接続助詞\_\*\_6:18 記号\_、\_、\_、\_、\_読点\_\*\_1\_6:18 名詞\_これ\_これ\_これ\_コレ\_代名詞\_\*\_これ\_コレ\_1\_7:9 助詞\_で\_で\_で\_デ\_格助詞\_\*\_7:9 名詞\_鼻\_鼻\_鼻\_ハナ\_一般\_\*\_鼻\_ハナ\_1\_8:9 助詞\_を\_を\_を\_ヲ\_格助詞\_\*\_8:9 動詞\_かむ-カム-五段・マ行(噛む;搦む)\_かむ\_かむ\_カム\_自立\_基本形\_かむ-カム-五段・マ行(噛む;搦む)\_カム\_かむ\_9:10 名詞\_客\_客\_客\_キヤク\_一般\_\*\_客\_キヤク\_1\_10:12 助詞\_が\_が\_ガ\_格助詞\_\*\_10:12 名詞\_気\_気\_気\_キ\_一般\_\*\_気\_キ\_1\_11:12 助詞\_に\_に\_ニ\_格助詞\_\*\_11:12 動詞\_なる-ナル-五段・ラ行(成る;生る;鳴る;なる)\_なる\_なる\_ナル\_自立\_基本形\_なる-ナル-五段・ラ行(成る;生る;鳴る;なる)\_ナル\_なる\_12:13 名詞\_程度\_程度\_程度\_テイド\_一

般\_\*程度\_テイド\_1\_13:14 助詞\_の\_の\_の\_ノ\_連体化\_\*\_\_13:14 名詞\_数\_数\_数\_カズ\_一般\_\*\_数\_カズ\_1\_14:15 記号  
 、\_、\_、\_、\_、\_読点\_\*\_1\_14:15 動詞\_出現する-スル-サ変・ースル\_出現する\_出現し\_シュツゲンスル\_自立\_連用形\_出  
 現する-スル-サ変・ースル\_出現する\_出現し\_15:16 助動詞\_た\_た\_た\_タ\_\*\_基本形\_\_15:16 助詞\_という\_という\_とい  
 う\_トイウ\_格助詞\_\*\_という\_\_15:16 名詞\_こと\_事\_事\_コト\_非自立\_\*\_こと\_コト\_1\_16:18 助詞\_は\_は\_は\_ハ\_係助詞  
 \*\_\_\_16:18 名詞\_間違い\_間違い\_マチガイ\_ナイ\_形容詞\_語幹\_\*\_\_17:18 助動詞\_ない\_ない\_ない\_ナイ\_\*\_基本形  
 \_\_18:-1 名詞\_の\_の\_の\_ノ\_非自立\_\*\_の\_ノ\_1\_18:-1 助動詞\_だ\_だ\_だろ\_ダロ\_\*\_未然形\_\_18:-1 助動詞\_う\_う\_う\_ウ  
 \*\_\_基本形\_\_18:-1 記号\_。\_。\_。\_。\_句点\_\*\_1\_18:-1

## 6.9 文法パターンの抽出

[名詞+ト+動詞] の抽出パターン

A009 名詞+助詞 < …と%(TARGET)% 1 名詞 %(POS=名詞)%%(基本形=と;POS=助詞)%%(TARGET)%  
 A009 名詞+助詞 < …と%(TARGET)% 1 名詞 %(POS=名詞;CHECKLINK)%%(基本形=と;POS=助  
 詞)%%(…)%%(TARGET;GETLINK)%

## 6.10 データベース登録用ファイルの作成

コロケーションデータ

V.00347.A009	A009.00038	大勢と{表現する}	1	8.75	4.20	0	0.00	0.00	0.00
				0.00	0.42				

用例データ

V.00347.A009.00038 OY01\_00105.S2 彼女は、2人でも大勢と表現するので、真偽の程が定かでないが、こ  
 れで鼻をかむ客が気になる程度の数、出現したという事は間違いないだろう。 8 15 74 (Yahoo!  
 ブログ, 2008, ビジネスと経済) OY

The screenshot shows a software interface with two main panels. The left panel, titled "...と表現する 78件", displays a table of collocations. The right panel, titled "大勢と表現する", shows a detailed view of the selected collocation, including a search filter and a text snippet from a blog.

コロケーション	頻度	MI	N-S
大勢と表現する	1	8.75	0.00
夜と表現する	1	6.14	0.02
地獄と表現する	1	9.02	0.02
国と表現する	1	4.82	0.02
善人と表現する	1	12.78	0.02
反復と表現する	1	11.57	0.02
卵と表現する	1	7.95	0.02
勝と表現する	1	10.23	0.02
創造と表現する	1	8.81	0.02
元気と表現する	1	8.19	0.02
企図と表現する	1	14.12	0.02
代表と表現する	1	7.06	0.00
人形と表現する	1	8.83	0.02

大勢と表現する

書籍・地の文: 0  書籍・会話文: 0  
 国会: 0  知恵袋: 0  ブログ: 0.42

彼女は、2人でも大勢と表現するので、真偽の程が定かでないが、これで鼻  
 をかむ客が気になる程度の数、出現したという事は間違いないだろう。  
 (Yahoo!ブログ, 2008, ビジネスと経済)

## 7. NLB 演習

### 7.1 基本的な使い方

コンコーダンスでは検索語句を入力すると、その結果がコンコーダンスラインとして返ってきます。NLBでは、検索語句を入力する代わりに、まず調べたい語（NLBではこれを「見出し語」と呼びます）を選びます。見出し語として選べるのは、名詞、動詞、イ形容詞、副詞の4種類の内容語です。ナ形容詞については今後対応する予定です。



ここでは、「言葉」という名詞について調べてみます。まず画面上の入力ボックスに「言葉」または「ことば」（カタカナも可）を入力して、[絞り込み] ボタンをクリックします。下のリストに「言葉」という見出し語が現れますのでそこをクリックします。



クリックすると、「言葉」の見出し語ウィンドウが開きます。ここでは、「言葉を」の後ろにどのような動詞が来るかを調べてみることにします。左側の文法パターンパネルの[グループ別]を選ぶと、一番上に[> 助詞+動詞]グループがあります。このなかにある[言葉を...]というパターンをクリックします。



すると、中央のコロケーションパネルに「言葉を+動詞」のコロケーションが頻度順に表示されます。左側の用例パネルには、最も頻度の高い「言葉を…している」の用例が表示されます。

言葉 頻度=26,690

グループ別 パターン頻度順 基本

パターン	頻度	比率
言葉を...	3,729	
言葉は...	1,808	
言葉も...	467	
言葉の...	107	
言葉を...	8,169	
言葉に...	3,210	
言葉へ...	2	
言葉で...	1,909	
言葉と...	192	
言葉から...	153	
言葉まで...	25	
言葉より...	30	

言葉を... 846件

コロケーション	頻度	MI	N-S
言葉をいる	601	11.22	6.60
言葉を使う	473	6.94	1.71
言葉聞く	452	6.62	1.83
言葉をする	360	12.30	5.13
言葉をかける	234	6.78	2.66
言葉を交わす	175	10.11	1.38
言葉を切る	143	7.11	2.95
言葉を失う	134	7.38	2.84
言葉を思い出す	129	7.56	2.11
言葉を続ける	127	7.02	2.57
言葉を発す	119	8.75	1.82
言葉を備りる	115	8.09	1.00
言葉を言う	105	2.09	-0.42
言葉を通る	103	10.02	2.29
言葉を選ぶ	99	6.66	0.84
言葉を備じる	91	6.76	-1.78

言葉をい... 11.78 言葉を... 5.16  
 社会: 2.98 知覚負: 4.08 ブログ: 2.45

言葉を伏えているのよ。  
 (イーヴリン・E.スミス著、長野きよみ訳『ミス・メルヴィルの復讐』, 1991, 9 文学)

沢渡が言葉を探している。  
 (五十嵐真久著『1985年の奇蹟』, 2003, 9 文学)

俺は言葉をなくしていた。  
 (宮川ゆうこ著『執筆は夜に嘘をつく!』, 2005, 9 文学)

和可子は言葉を失っていた。  
 (内館秋子著『愛しすぎなくてよかった』, 2001, 9 文学)

という言葉を残しています。  
 (秋庭道博著『「この人についていこう」と思われる上司になる』, 2002, 1 哲学)

頭の隅で言葉を探している。  
 (栗根文生著『旅タオトパイ』, 2002, 2 歴史)

幸太郎は言葉を失っていた。  
 (遊道著『オルガナー』, 2005, 9 文学)

NINJAL-LWP for BCCWJ Copyright © 2012 National Institute for Japanese Language and Linguistics, Lago Institute of Language. All rights reserved.

次に、特徴的なコロケーションを表示してみます。NLB では、頻度順だけでなくMIスコアの順でも並べ替えることができます。MIスコアは統計指標の一つで、特徴的なコロケーションほど数値が高くなる傾向があります。ただし、低頻度のコロケーションの数値が過剰に高くなるため、低頻度のものを排除する必要があります。コロケーションパネルのヘッダーの「MI」をクリックしてから、パネル上で右クリックして「頻度 5 以上」を選びます。

言葉を... 846件

コロケーション	頻度	MI	N-S
言葉を自発する	1	12.93	0.02
言葉を耳語する			
言葉をそびれる			
言葉を輪唱する			
言葉を途切らせる			
言葉をする	360	12.30	5.13

頻度3以上  
 頻度5以上  
 頻度10以上  
 リセット

すると、「言葉を途切らせる」、「言葉を濁す」など、「言葉を」に続く特徴的な動詞が上位に現れます。

言葉を... 194件

コロケーション	頻度	MI	N-S
言葉を途切らせる	18	12.57	0.40
言葉をする	360	12.30	5.13
言葉をさせる	9	12.19	0.13
言葉を濁す	67	11.70	1.15
言葉をいる	601	11.22	6.60
言葉を入れる	18	11.19	0.02
言葉を反響する	19	10.41	0.42
言葉を述べ伝える	5	10.16	0.11
言葉を交わす	175	10.11	1.38
言葉を紡ぐ	16	10.08	-0.05
言葉を紡ぎ出す	6	10.08	-0.07
言葉を換える	35	10.02	0.31
言葉を通る	103	10.02	2.29
言葉をくる	48	9.67	0.82
言葉を聞き逃す	5	9.39	0.09
言葉を投げつける	29	9.28	0.62

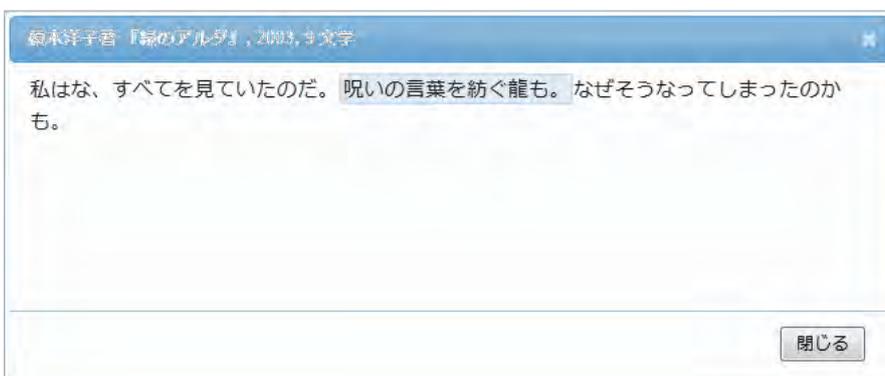
それぞれのコロケーションをクリックすると、右のパネルにその用例が表示されます。ここでは、10番目の「言葉を紡ぐ」をクリックしてみます。コーパスで使われている実際の用例を一つずつ確認することができます。



用例はセンテンス単位で表示されます。さらに前後の文脈を確認するときは、それぞれの用例の出典の部分をクリックします。



以下のような前後のセンテンスを表示するダイアログが開きます。



このように、NLB では、検索ウィンドウでまず見出し語を選んで、見出し語ウィンドウ上で、文法パターン、コロケーション、用例を行き来しながら、その見出し語の振る舞いの全体像をつかむことができます。クリックだけで簡単に操作できますので、思考を途切らせることなくさまざまな表現を思いのままに調べることができます。

## 7.2 動詞を調べる

見出し語検索ウィンドウの見出し語リストを「動詞」に切り替えます。



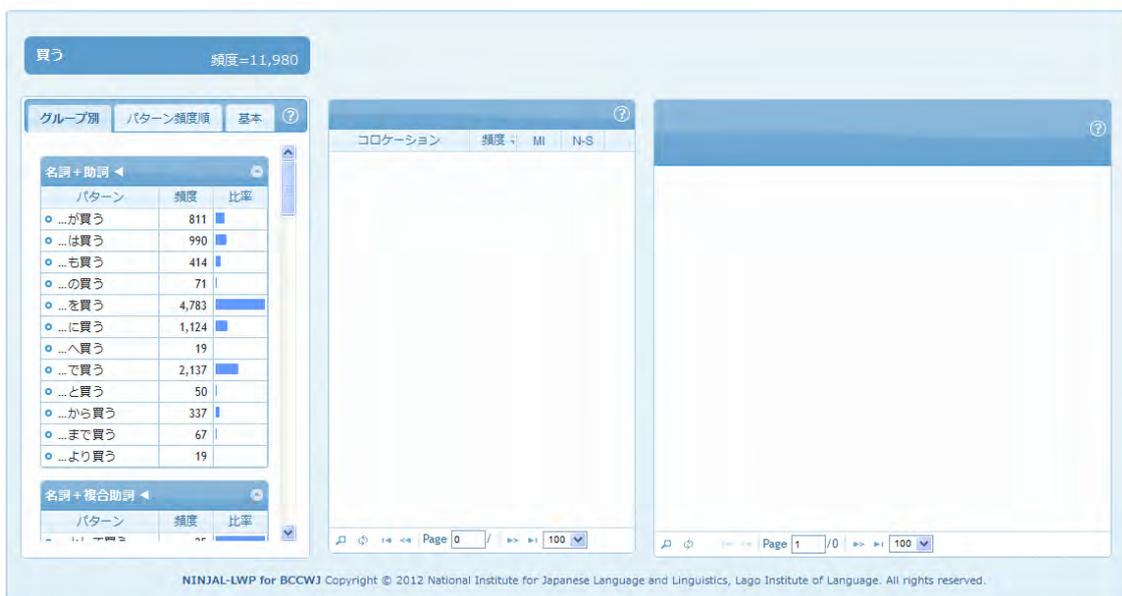
見出し	読み	頻度
いる-非自立	イル	626,406
する	スル	382,108
なる	ナル	305,944
ある	アル	296,450
言う	イウ	191,252
思う	オモウ	127,188
できる	デキル	116,370
くる-非自立	クル	109,361
見る	ミル	81,316
いく-非自立	イク	64,460

絞り込みの入力ボックスに、「かう」と入力し、「絞り込み」ボタンをクリックします。リスト中の「買う」をクリックします。



見出し	読み	頻度
買う	カウ	11,980
飼う	カウ	1,402
かう	カウ	464
支う	カウ	55
交う	カウ	49

「買う」の見出し語ウィンドウ（タブ）が開きます。



パターン	頻度	比率
...が買う	811	
...は買う	990	
...も買う	414	
...の買う	71	
...を買う	4,783	
...に買う	1,124	
...へ買う	19	
...で買う	2,137	
...と買う	50	
...から買う	337	
...まで買う	67	
...より買う	19	

[文法パネル] の [名詞+動詞 ◀] パターンの […を買う] をクリックします。

パターン	頻度	比率
○ …が買う	811	
○ …は買う	990	
○ …も買う	414	
○ …の買う	71	
○ …を買う	4,783	
○ …に買う	1,124	
○ …へ買う	19	
○ …で買う	2,137	
○ …と買う	50	
○ …から買う	337	
○ …まで買う	67	
○ …より買う	19	

文法パターンパネルの [名詞+動詞] パターンの […を買う] をクリックします。「名詞+ヲ買う」のコロケーションとその用例が表示されます。パネルをスクロールして、どのようなコロケーションがあるか確かめます。

…を買う 1330件

コロケーション	頻度	MI	N-S
ものを買う	177	3.92	-2.29
本を買う	119	5.72	-0.30
車を買う	99	6.39	0.16
物を買う	96	6.34	0.57
株を買う	87	8.42	-1.11
切符を買う	79	10.77	-0.48
家を買う	74	4.43	-0.23
のを買う	71	1.04	-1.99
土地を買う	66	6.85	-0.46
それを買う	59	2.65	-0.91
服を買う	56	7.85	-0.63
品を買う	50	6.71	0.12
商品を買う	49	6.28	0.32

Page 1 / 14 100

ものを買う

■ 書籍・地の文: 1.65 ■ 書籍・会話文: 3.94  
■ 国会: 2.42 ■ 知恵袋: 11.43 ■ ブログ: 3.75

お金でものを買う。  
(袁老孟司著『脚の冒険』, 2003, 4 自然科学)

…自分の食いたいものを買いました  
(徳野道流著『無明の闇』, 2000, 9 文学)

猫草は市販のものを買ってあります。  
(Yahoo!知恵袋, 2005, 教養と学問, サイエンス)

多少高くても良いものを買いましょう。  
(Yahoo!知恵袋, 2005, インターネット, PCと家電)

か安いものを買いなおすのかも迷っています。  
(Yahoo!知恵袋, 2005, スポーツ, アウトドア, 車)

あとで同じものを買って丁重におわびしました。  
(竹谷洋子著『帝國ホテルが教えてくれたこと』, 1999, 6 産学)

Page 1 / 2 100 177件中 1 - 100を表示

特徴的なコロケーションを見るために、MIスコアで並べ替えてみます。コロケーションパネルのヘッダーの [MI] をクリックし、パネル上で右クリックして [頻度5以上] を選びます。

…を買う 1330件

コロケーション	頻度	MI	N-S
エステマルシータを 買う	1	13.70	0.00
オトトを買う			
キューブキュービ 買う			
ザクウォーリアを			
スッキリンを買う	1	13.70	0.00
スニカーを買う	1	13.70	0.00
テレキャスターモデルを 買う	1	13.70	0.00
デニッシュベストリーを 買う	1	13.70	0.02
トロターを買う	1	13.70	0.00
ハズレタマを買う	1	13.70	0.00

Page 1 / 14 100

「鬻鬻を買う」、「歓心を買う」などの慣用的な表現が上位に来るのが分かります。

コロケーション	頻度	MI	N-S
鬻鬻を買う	46	13.01	-0.48
歓心を買う	18	13.01	-0.40
不興を買う	25	12.70	-0.04
失笑を買う	9	11.70	-0.22
不評を買う	10	11.70	-0.02
反感を買う	39	10.98	0.14
切符を買う	79	10.77	-0.48
パンプスを買う	6	10.22	0.02
馬券を買う	21	9.79	0.14
土産を買う	45	9.70	0.21
宝くじを買う	12	9.69	-0.40
おみやげを買う	9	9.53	-0.04
一役を買う	5	9.47	0.09

「鬻鬻を買う」の4番目の用例の出典部分をクリックして、前後の文脈を確認めます。

鬻鬻を買う

■ 書籍・地の文: 0.77 ■ 書籍・会話文: 1.25  
■ 国会: 0.2 ■ 知恵袋: 0.19 ■ ブログ: 1.25

■ 鬻鬻を買う 無礼な態度。  
(佐藤賢一著『双頭の鷲』, 2001, 9 文学)

■ ひんしゆくを買っているとき  
(山本清治著『白い眼』, 1989, 9 文学)

■ と業界内で鬻鬻を買ったほどである。  
(三神万里子著『メガバンク決算』, 2003, 3 社会科学)

■ これは周囲のヒンシュクを買いまくった。  
(菅沢朋著『少年たちの終わらない夜』, 1993, 9 文学)

■ または、周囲のヒンシュクを買いやすい人。  
(Yahoo!知恵袋, 2005, 教養と学問、サイエンス)

■ 王族たちの鬻鬻を買うことは、火をみるよりも明らかだったのだ。

前後の文脈を読むと、なぜ「周囲の鬻鬻を買った」のかが分かります。

菅沢朋著『少年たちの終わらない夜』, 1993, 9 文学

高校を卒業したあと、浩次はひどいバカだったためどんな大学にも行けず、その上親が金持ちなものだから私費留学という名目でアメリカに逃亡を果たした。これは周囲のヒンシュクを買いまくった。

良ちゃんは高三からなぜだかひどく熱心に勉強をはじめ、邪魔する間もなく北大に合格してしまった。

閉じる

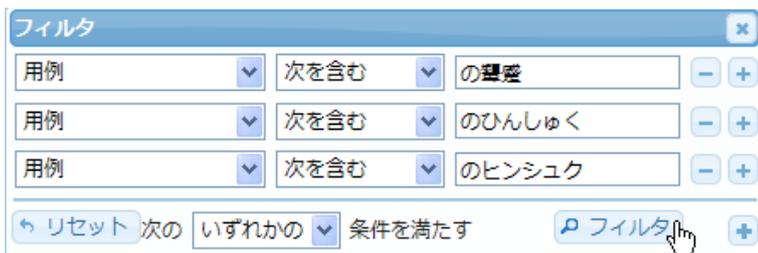
用例を読んでいくと、「…の鬻鬻を買う」という言い方が多いのに気がつきます。フィルターを使って、「…」の部分にどんな名詞が多いのかを調べてみます。まず、用例パネルの左下にある「[フィルタ]」ボタンをクリックします。



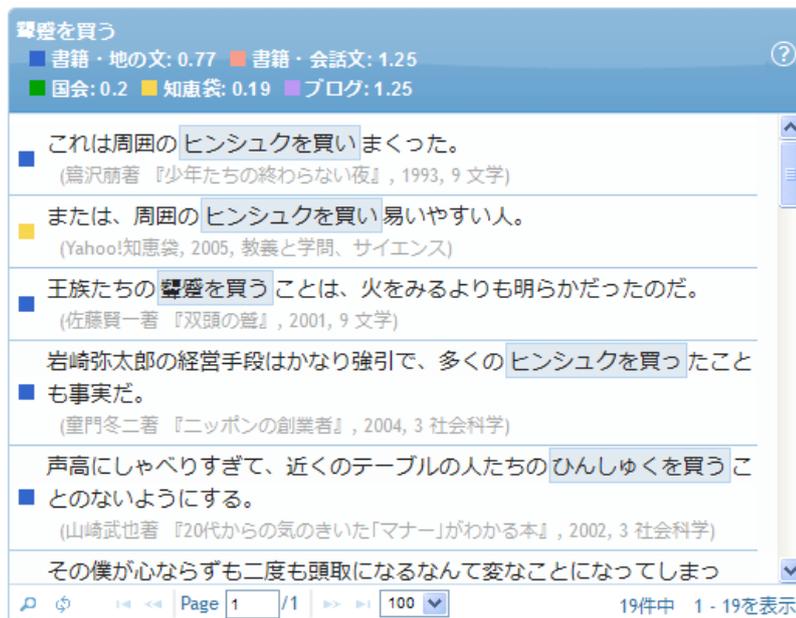
フィルタダイアログが表示されますので、順に条件を入力してきます。まず一番左のボックスから [用例] を選び、右のボックスに「の響盛」を入力します。



ひらがなやカタカナの表記もあるので、[+] ボタンをクリックして、同じように条件を追加していきます。最後に [いずれかの条件を満たす] を選び、[フィルタ] ボタンをクリックします。



先ほど文脈を確認した用例を含めて、全部で 19 件の用例が見つかりました。「響盛を買う」の 46 例のうち、半分近くの 19 例が「…の響盛を買う」という形で使われていることが分かります。

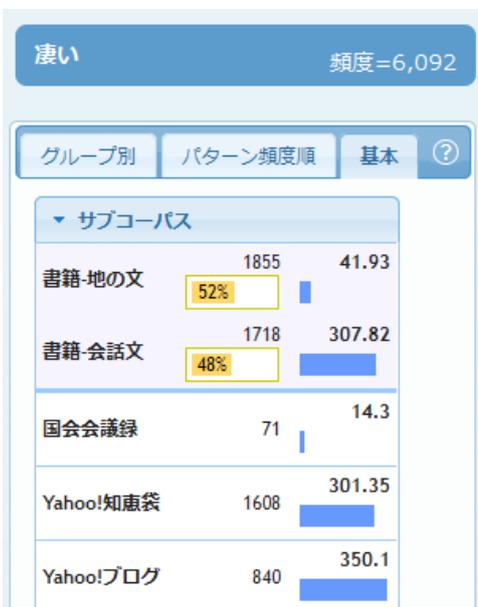


### 7.3 形容詞を調べる

見出し語検索ウィンドウの見出しリストを「形容詞」に切り替えて、「すごい」と入力して、「絞り込み」ボタンをクリックして、「凄い」をクリックして、見出し語ウィンドウを開きます。



文法パターンパネルの「基本」をクリックします。「サブコーパス」を見て、サブコーパスごとの頻度や100万語当たりの頻度（PMW）を確認します。PMWを見ると、「凄い」は口語的表現であることが分かります。



次に「書字形」をクリックします。ここでは、どの表記がどの程度の割合で使用されているかが分かります。ひらがな表記が8割を超えていることが確認できます。

書字形	頻度	割合
すごい	5021	82%
凄い	1036	17%
スゴイ	35	1%

サブコーパスごとの書字形の分布も分かります。「サブコーパスごとの書字形分布」をクリックします。知恵袋やブログといったウェブテキストでの漢字表記の割合が高いことが分かります。ウェブテキストでの漢字表記使用の傾向は、他の品詞についても見られる現象です。

サブコーパスごとの書字形分布		
書籍・地の文		
すごい	1545	83%
凄い	290	16%
スゴい	20	1%
書籍・会話文		
すごい	1533	89%
凄い	183	11%
スゴい	2	0%

国会会議録		
すごい	71	100%
凄い	0	0%
スゴい	0	0%
Yahoo!知恵袋		
すごい	1295	81%
凄い	311	19%
スゴい	2	0%
Yahoo!ブログ		
すごい	577	69%
凄い	252	30%
スゴい	11	1%

次に、文法パターンパネルを「グループ別」に切り替えて、「凄い+名詞」のコロケーションを調べてみます。

グループ別		
パターン頻度順		
基本		
名詞		
パターン	頻度	比率
● 凄い+名詞	1,374	
● 凄い+接尾名詞	116	
名詞+助詞		
パターン	頻度	比率
● ...が凄い	664	
● ...は凄い	367	
● ...も凄い	212	
● ...の凄い	153	
● ...を凄い	58	
● ...に凄い	90	
● ...で凄い	99	
● ...と凄い	10	
● ...から凄い	25	
● ...まで凄い	8	

凄い+名詞 590件			
コロケーション	頻度	MI	N-S
凄いこと	117	3.50	-5.30
凄いん	76	4.94	-4.57
凄いの	62	2.62	-1.63
凄い勢い	49	10.09	0.14
凄い人	41	3.35	-1.68
凄いもの	36	3.40	-0.96
凄いところ	35	5.00	-0.02
凄い音	33	7.86	-1.23
凄いスピード	22	8.91	-0.07
凄い話	14	4.50	-0.36
凄い雨	14	7.18	-0.09
凄い剣幕	13	12.60	-0.13
凄い美人	11	8.82	-0.58
凄い人気	10	6.93	-0.70
凄い言葉	9	3.94	-0.45
凄い数	9	5.13	-0.49

コロケーションパネルの5番目の「凄い人」をクリックして、用例を確認します。

凄い人

- 書籍・地の文: 0.29
- 書籍・会話文: 1.97
- 国会: 0
- 知恵袋: 1.31
- ブログ: 4.17

すごい人だ。  
(山藤草二著 『「笑い」の混沌』, 1991, 7 芸術・美術)

すごい人なんだぜ。  
(神楽明著 『堂の上の支配人』, 1989, 2 歴史)

実は、凄い人なの？  
(Yahoo!知恵袋, 2005, エンターテインメントと趣味)

ともかく、凄い人でした。  
(山城新伍著 『かなり好きです』, 1994, 7 芸術・美術)

(@@) すごい人・・・  
(Yahoo!ブログ, 2008, Yahoo!サービス)

すごい人だなーと思います。  
(Yahoo!ブログ, 2008, 生活と文化)

...なんて、すごい人なんだろう？  
(水上レイ著 『ジュエリーデザイナーの交友関係』, 2004, 9 文学)

41件中 1 - 41を表示

1 番目と 2 番目の用例の文脈を調べて、どのような意味で使われているかを確認します。同じ表現でも文脈によって意味が違っているのが分かります。単にコロケーションの分布を調べるだけでなく、個々のコロケーションの用例を見ることの重要性を示す好例です。

山藤章二著 『「笑い」の混沌』, 1991, 7 芸術・美術

声がすごいし、あの人が出てくるだけで場内を支配しちゃうってところありますね。  
すごい人だ。

井上

俳優さんによって自分の半径二メートルぐらいしか支配できない人と、かろうじて舞台をやっと支配できる人と、すまさんのように客席まで全部コントロールしちゃう人と、いろいろ器であります。

閉じる

神室明若 『雲の上の支配人』, 1989, 2 歴史

....、渋谷の八子公前の交差点を知ってるかい？ すごい人なんだぜ。一回の信号で上宝村の全人口ぐらいの人間が道を渡るんだぞ。

閉じる