

NLT(NINJAL-LWP for TWC) 利用マニュアル

目次

1. NLT の概要	2
2. 使い方ガイド	5
3. チュートリアル.....	6
4. コーパス分析手法 コンコーダンスとレキシカルプロファイリング	9
5. NLT の2つのウィンドウ.....	13
6. 見出し語検索ウィンドウの操作	15
7. 見出し語ウィンドウの操作	21

1. NLT の概要

1.1 1.1 NLT とは

NINJAL-LWP for TWC (ニンジャル・エルダブリュピー・フォー・ティーダブリュシー、略称 NLT)

(<http://corpus.tsukuba.ac.jp>)は、筑波大学が日本語のウェブサイトから収集して構築した約 11 億語のコーパス『筑波ウェブコーパス』(Tsukuba Web Corpus: TWC) を検索するためのツールです。検索には、国立国語研究所 (以下、国語研) と Lago 言語研究所が共同開発したコーパス検索システム NINJAL-LWP (NINJAL-LagoWordProfiler) を利用しています。同じシステムを利用したツールに、国語研が構築した 1 億語の『現代日本語書き言葉均衡コーパス』(Balanced Corpus of Contemporary Written Japanese: BCCWJ) を検索する NINJAL-LWP for BCCWJ (NLB) (<http://nlb.ninjal.ac.jp>) があります。

このツールを用いると、名詞や動詞などの内容語の共起関係や文法的振る舞いを網羅的に調査・比較することができます。

The screenshot displays the NINJAL-LWP search results for the verb '走る' (to run). The search results are organized into three main sections:

- 走る (頻度=128,836):** A table showing search results for the verb '走る'. The table has columns for 'グループ別' (Group), 'パターン頻度順' (Pattern Frequency Order), and '基本' (Basic). The table lists various patterns such as '...が走る', '...は走る', etc., along with their frequency and ratio.
- ...が走る 1874種類:** A table showing search results for the verb '走る' in the context of '...が走る'. The table has columns for 'コロケーション' (Collocation), '頻度' (Frequency), 'MI', and 'LD'. The table lists various collocations such as '車が走る', '痛みが走る', etc., along with their frequency, MI, and LD values.
- 車が走る 全889件:** A list of search results for the verb '走る' in the context of '車が走る'. The list shows various examples of the verb '走る' used in the context of '車が走る', such as 'SL機関車が走る音。', '車が飛ばいばい走っている。', etc.

1.2 使用しているコーパス

筑波ウェブコーパス (TWC) ver.1.10 では、日本語のウェブサイトから収集した 11 億 3800 万語のデータを使用しています。

1.2.1 筑波ウェブコーパスの構築

ウェブ上からのテキストの収集では、検索エンジンの API を利用して、ウェブページの URL を収集した後、その URL のデータを収集する一般的な手法に従っています。以下が具体的なコーパス構築の手順です。

- シードおよびタプルの生成

検索エンジンのクエリパラメータに与えるタプルを構成するシードには、NLBの開発過程で作成したBCCWJ(2009年の領域公開データの一部、約6千2百万語)の頻度リストを利用しました。品詞ごとに分かれた頻度リストのうち、内容語である名詞、動詞、形容詞、副詞のリストをマージして、上位500語をシードとして選びました。ただし、名詞のうち、数詞、固有名詞は排除し、また、動詞、形容詞については活用形も含めました。この500語のシードから無作為に3語を選び出し、計50万組のタプルを作成しました。以下にタプルの例を示します。

駄目 皆 構造
 条件 とても 様々
 法律 (答える OR 答え OR 答えよ OR 答えれ OR 答えろ OR 答えりゃ OR 答えん) 人々

●検索エンジン API による URL の収集

URLの収集には、Yahoo!ウェブ検索APIを利用しました。1タプル当たりで収集するURL数は10ページとし、2012年1月初旬から下旬にかけて計500万URLを収集しました。重複したURLを削除したURL総数は約3割減の約350万件になりました。

●HTML ページの収集

URLデータを5万件ごとに分割した上で、3台の端末を利用して2週間をかけてHTMLページを収集しました。

●テキストの抽出

次に収集したHTMLファイルからテキストを抽出する作業を行いました。具体的には、HTMLタグの削除、文字コードの統一(utf8)、日本語以外の言語で書かれたテキストの削除などの作業を行いました。

●不適正なページの排除

筑波ウェブコーパスの構築の目的は日本語の用例を採取することにあります。そのため、単に項目やリンクを列挙しただけのページ、広告と思われる内容の多いページ、センテンス境界の判定が難しいページは、あらかじめコーパスデータの対象から外しました。

●センテンスの抽出

レキシカルプロファイリングツールNINJAL-LWPでは、センテンス単位にした用例の中にどのようなコロケーションが含まれるかを文法パターンごとに抽出します。そのため、コーパスデータはあらかじめセンテンス単位に分割しておく必要があります。一つ前の作業でセンテンス境界の判定が難しいページを排除したのもこの理由によります。

●用例データの抽出

センテンス単位のデータのなかには、見出しに相当するものや、メニュー項目に相当するものが含まれています。センテンス中にどの程度名詞が含まれるか、センテンス中に動詞は現れるか、「クリック」や「ログイン」などのウェブページで多用される表現が用いられているかなどの複数の観点から、用例としての適正度を数値化し、用例としてふさわしいデータを抽出しました。また、同一ページで同じセンテンスが現れた場合も、最初の1件のみを用例として採取し、不要な重複を避ける工夫をしました。

●重複する用例データの削除

一つ前の作業で、同一ページでは同じ用例が複数回採取されないようにしましたが、6億語弱のパイロット版 NLT を開発して実際に運用してみたところ、同一サイトで同一の用例が頻出することが確認されました。そのため、URL の情報をもとに同一サイト での同じ用例は一度だけ採取するように改良し、最終的に語数にして 11 億 3781 万語、用例数にして 4672 万 7 千例の筑波ウェブコーパスが完成しました。

1.2.2 アノテーション

NLT では、コロケーションや文法的振る舞いの情報を抽出するために、筑波ウェブコーパス (TWC) のデータにアノテーションを付与した上で解析しています。アノテーションに使用している解析器・辞書は以下の通りです。

形態素解析 MeCab 0.98 + IPA 辞書 2.7.0

形態素解析用の IPA 辞書は、代表表記の情報を含まないため、独自に拡張して代表表記に対応させています。

係り受け解析 CaboCha 0.60

1.3 ご利用にあたって

1. 【動作環境】ブラウザは Firefox、Chrome、Safari、IE (バージョン 8 以上) に対応しています。処理速度の観点から、Firefox、Chrome、Safari のご利用をお勧めします。
2. 【クッキーの設定】使用にあたってはブラウザのクッキーをオンにしてください。オフにした状態ではデータが表示されません。クッキーをオンにする方法については、各ブラウザのマニュアルなどをご確認ください。
3. 【表示結果】NLT では、機械的に処理した結果をそのまま表示しています。形態素・係り受け解析や抽出処理の精度の限界により不適切なデータが混入しています。あらかじめご了承ください。
4. 【論文・記事を公表する場合】NLT を研究・教育に利用して論文や記事を執筆される場合は、以下のように、必ず NLT を利用した旨を明記してください。

筑波大学、国立国語研究所、Lago 言語研究所『NINJAL-LWP for TWC』 (<http://corpus.tsukuba.ac.jp>)

5. 【用例の削除依頼】このコーパスは、教育・研究目的で、ウェブ上からデータを収集したものです。すべての用例は出所先であるページのタイトルと URL を明記した上で表示しています。自らが著作権を有するウェブページから抽出された用例の削除を希望される場合は、お問い合わせのメールアドレスまでご連絡ください。申請者ご本人のページであることを確認した上で削除させていただきます。

1.4 お問い合わせ

NLT に関するお問い合わせは以下までお願いいたします。

kyoten@intersc.tsukuba.ac.jp

2. 使い方ガイド

見出し語検索ウィンドウの見出し語リストと、見出し語ウィンドウの各パネルでは、簡単な操作法を示した使い方ガイドを表示することができます。リストやパネルの右上の [?] ボタンをクリックします。

すべて	名詞	動詞	連体詞	形容詞	副詞
見出し		読み	ローマ字表記	頻度	[?]
いる-非自立	イル	iru	11,402,357		
こと	コト	koto	9,542,386		
の	ノ	no	6,953,360		
する	スル	suru	6,634,633		
ある	アル	aru	6,572,844		

下のような使い方ガイドが表示されます。

タブを切り替えると、その品詞の見出し語のみを表示します

読みとローマ字表記と頻度で並べ替えることができます

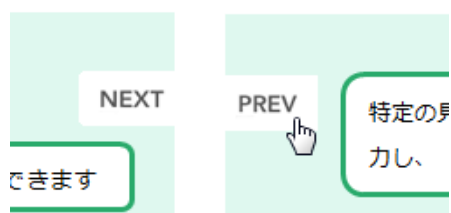
見出しをクリックすると、その見出し語のウィンドウが開きます

見出し語検索 1

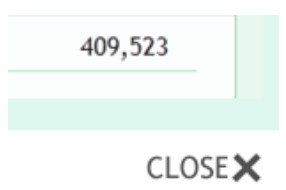
NEXT

CLOSE X

次の説明に移動するときは、右上の [NEXT] の部分をクリックします。ひとつ前の説明に戻るときは、左上の [PREV] の部分をクリックします。



使い方ガイドを閉じるときは、画面右下の [Close x] をクリックします。



3. チュートリアル

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。NLTでは、検索語句を入力する代わりに、まず調べたい語（NLTではこれを「見出し語」と呼びます。正確には、形態素解析によって認定された形態素を指します）を選びます。見出し語として選べるのは、名詞、動詞、形容詞、連体詞、副詞の5種類の内容語です。

NINJAL-LWP for TWC

ここに調べたい語を入力（読みも可）

すべて 名詞 動詞 連体詞 形容詞 副詞

見出し	読み	ローマ字表記	頻度
いる-非自立	イル	iru	11,402,357
こと	コト	koto	9,542,386
の	ノ	no	6,953,360
する	スル	suru	6,634,633
ある	アル	aru	6,572,844

ここでは、「走る」という動詞について調べてみます。まず画面上的入力ボックスに「走る」または「はしる」（カタカナも可）または「hashiru」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「はしる」という読みの見出し語が3つ表示されますので、一番上の「走る」をクリックします。

はしる

すべて 名詞 動詞 連体詞 形容詞 副詞

見出し	読み	ローマ字表記	頻度
走る	ハシル	hashiru	128,836
はしる	ハシル	hashiru	954
奔る	ハシル	hashiru	303

クリックすると、「走る」の見出し語ウィンドウが開きます。ここでは、「走る」の主語となるガ格の名詞にはどのようなものがあるの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[名詞+助詞 <]グループがあります。一番上の[…が走る]というパターンをクリックします。

走る 頻度=128,836

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
…が走る	14,782	
…は走る	7,415	
…も走る	2,979	
…の走る	1,337	
…を走る	30,443	
…に走る	23,992	
…へ走る	2,102	
…で走る	15,571	
…と走る	1,088	
…から走る	3,415	
…まで走る	2,476	
…より走る	304	

すると、中央のコロケーションパネルに「[名詞+が走る]」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「車が走る」の用例が表示されます。

走る 頻度=128,836

グループ別 パターン頻度順 基本 ?

名詞+動詞

パターン	頻度	比率
...が走る	14,782	
...は走る	7,415	
...も走る	2,979	
...の走る	1,337	
...を走る	30,443	
...に走る	23,992	
...へ走る	2,102	
...で走る	15,571	
...と走る	1,088	
...から走る	3,415	
...まで走る	2,476	
...より走る	304	

名詞+複合動詞

パターン	頻度	比率
...とあって走る	1	
...として走る	204	
...と共に走る	15	
ににおいて走る	1	

...が走る 1874種類

コロケーション	頻度	MI	LD
車が走る	889	7.85	6.55
痛みが走る	565	9.00	7.54
激痛が走る	513	13.92	9.92
電車が走る	513	9.86	8.20
バスが走る	413	8.49	7.03
【一般】が走る	356	2.31	1.08
列車が走る	343	9.72	7.93
衝撃が走る	275	9.82	7.88
人が走る	251	2.61	1.37
緊張が走る	248	9.24	7.45
線が走る	236	6.65	5.30
【人名】が走る	214	1.34	0.10
自転車が走る	181	7.98	6.40
道路が走る	178	6.98	5.56
ウマが走る	142	7.95	6.30
私が走る	137	2.47	1.23
自動車が走る	134	6.81	5.36
電気が走る	134	6.88	5.43
鉄道が走る	128	7.94	6.26
自分が走る	121	2.77	1.52

車が走る 全889件

SL機関 車が走る音。
([乗り物 交通] フリー効果音素材 [Senses Circuit])

車がいっぱい走っている。
(すーちゃんの部屋～運転免許を取るの)

なんと車が走っています。
()

車が走っていません...。
(r782, r781 (サバイバル鉄道へようこそ))

車がまっすぐ走らない。
(ホイールアライメント調整 | タイヤPIT | ピット | Super☆AUTOBACS TOKYO BAY東宝)

左手を車が走っています。
(ばきらのお散歩日記～玉川上水を歩くのだ！ 5)

発問 車が走っています。
(ふるさとの木の葉の駅)

車が走っているだけです。
([ハワイゴルフマガジン] RYOのおもしろハワイエピソード その12)

蒸気機関車が走っていた道。
(2011年10月: ちよつと歴史っばい西宮)

次に、特徴的なコロケーションを表示してみます。NLTでは、頻度順だけでなく MI スコアなどの他の統計値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 20 以上] を選びます。

...が走る 109種類

コロケーション	頻度	MI	LD
虫が走る	44	16.05	6.60
激音が走る	81	14.10	7.45
激痛が走る	513	13.92	9.92
戦慄が走る	49	12.48	6.70
悪寒が走る	59	12.42	6.95
閃光が走る	68	11.93	7.11
稲妻が走る	67	11.92	7.09
市電が走る	46	11.66	6.57
都電が走る	25	11.35	5.72
SLが走る	27	11.24	5.82
亀裂が走る	102	11.05	7.50
唾が走る	42	10.55	6.34
寒気が走る	36	10.31	6.12
汽車が走る	59	10.24	6.70
動揺が走る	48	10.19	6.45
電車が走る	513	9.86	8.20
電撃が走る	20	9.84	5.31
衝撃が走る	275	9.82	7.88

...が走る 1874種類

コロケーション	頻度	MI	LD
虫が走る	44	16.05	6.60
豊じょうが走る	1	14.65	1.15
延丁が走る	1	14.23	1.15
激音が走る	81	14.10	7.45
激痛が走る	513	13.92	9.92
怖気が走る	513	13.92	9.92
火山脈が走る	413	8.49	7.03
警邏隊が走る	275	9.82	7.88
ガソリンカーが走る	251	2.61	1.37
稲光が走る	12	12.82	4.72
戦慄が走る	49	12.48	6.70

頻度5以上
頻度10以上
頻度20以上
頻度30以上
リセット

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、6番目の「閃光が走る」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。



さらに、その右の矢印のアイコンをクリックすると、別ウィンドウ（またはタブ）にもとのウェブページに表示します。



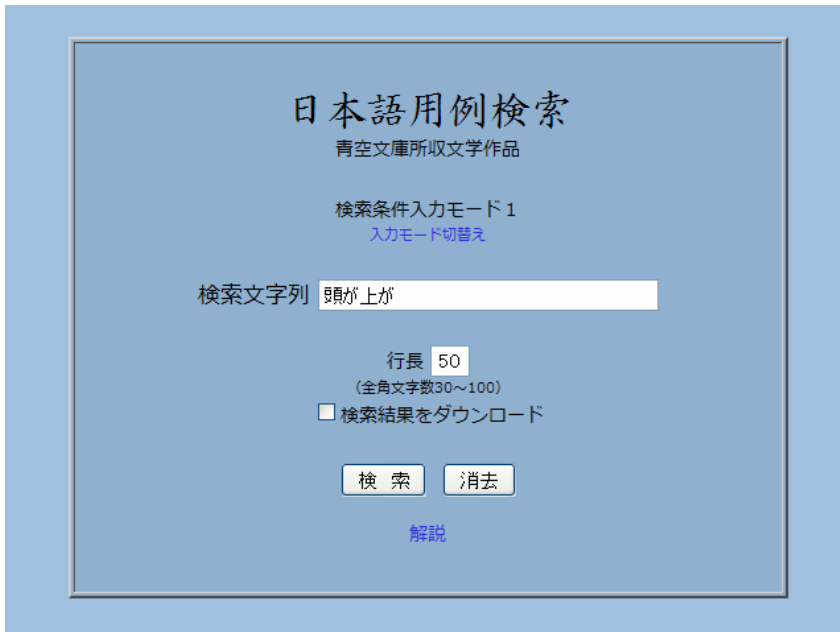
このように、NLTでは、検索ウィンドウでまず見出し語を選んで、見出し語ウィンドウ上で、文法パターン、コロケーション、用例を行き来しながら、その見出し語の振る舞いの全体像をつかむことができます。クリックだけで簡単に操作できますので、思考を途切らせることなくさまざまな表現を思いのままに調べることができます。

4. コーパス分析手法 コンコーダンスとレキシカルプロファイリング

4.1 コンコーダンス

コーパスの分析手法にはさまざまなものがありますが、最も一般的なのはコンコーダンスです。コンコーダンスは、調べたいキーワードを指定して検索します。キーワードのほか、キーワードと同時に現れる共起語を指定することもできます。大阪大学の田野村先生が公開されている日本語用例検索のサイトでは、青空文庫の検索結果をコンコーダンスで出力します。

▶ 日本語用例検索 : <http://www.let.osaka-u.ac.jp/~tanomura/kwic/aozora/>



ここでは、「頭が上がる」を調べてみます。「上がる」は動詞で活用形があるので、検索文字列は「頭が上がる」とします。「検索」をクリックすると、以下のようなコンコーダンスラインが返ってきます。コンコーダンスラインは、ノードと呼ばれるキーワードを中央に置いて、その前後の文脈が同時に確認できる形式になっています。「頭が上がる」の結果を見ると、10件中9件までが「頭が上がるらない」または「頭が上がるらぬ」という否定形で用いられ、「相手に引け目や負い目があって対等に振る舞えない」という意味であることが分かります。また、文字通りの動作を表す例は1件だけであることも確認できます。このように、コンコーダンスでは、個々の例を見ながら語句の振る舞いを判断できるため、文脈の確認が重要な言語研究にはきわめて有効なツールです。

検索結果 / 頭が上がる

れてしまった。実に面目次第もないが事実の前には**頭が上がる**らない。 ……一つは女を犯人と認めて行く道…… 夢野久作「暗黒公使」
れず、そうすりゃ、あれの代になっても、また一生**頭が上がる**らずに、貧乏たれで暮さにゃならんせに、今、ちい 黒島伝治「電報」
う公卿は、おおかた吉備彦に借りがあった。それで**頭が上がる**らなかった。恐るべきは金と女！ もう間もなくそ 国枝史郎「大鵬のゆくえ」
農程無抵抗主義なものはない。権力の前には彼等は**頭が上がる**らない。「田家衣食無厚薄、不見県門身即楽」で、 徳富健次郎「みみずのたはこと」
ると大の男である亭主も、小の女である女房の前に**頭が上がる**らない。昔の封建日本は男性横暴だつたといふが、 高田保「恋文」
りにわずかの髪をのこしている、お父親さんの禿た**頭が上がる**り、声の来た方へ向いたので、お蘭もそっちへ顔を 国枝史郎「猿ヶ京片耳伝説」
していたのでは、結局、俸給をくれるものに対して**頭が上がる**らない。上長に対しては、正しいと思ったことも言 相馬愛蔵「私の小売商道」
も私の一生の大失敗であると、いまでも主人の前に**頭が上がる**らないのであります。 よく売れるといっても知れ 相馬愛蔵「私の小売商道」
なく、ゴツゴツした持ち前だなどとなると、とても**頭が上がる**らない。とにかく商売くらい正直で正確で自由なも 相馬愛蔵「私の小売商道」
つて威張るので、あんな高慢な山木様も、家内では**頭が上がる**らないさうです、一一先生、外国人は矢ッ張り目が 木下尚江「火の柱」

10件見つかりました。

その一方で、コンコーダンスでは用例を1件ずつ見る必要があることから、検索件数が多い場合や語句の振る舞いが多様な場合には、コンコーダンスラインからだけでは全般的な傾向を読み取るのは難しくなります。特に辞書執筆・編集にコーパスを利用するケースでは、見出し語の振る舞いの全体像を短時間に効率的に把握する必要があるため、コンコーダンスのみで作業を行うことは相当な困難を伴います。このような問題の解決策として考案されたのが、レキシカルプロファイリングというコーパス検索・解析手法です。

4.2 レキシカルプロファイリング

レキシカルプロファイリングとは、語の振る舞いを網羅的に調べるために、あらかじめ共起語などの情報を集積しておき、その結果を統計的に処理した上で、重要度の高いコロケーションや文法的振る舞いなどを提示する手法のことをいいます。コーパスを全面的に採用した初のEFL用英英辞典 Colins COBUILD English Language Dictionary での編纂作業の経験を踏まえて、Patrick Hanks らが 1989 年に MI スコアなどの統計指標をコロケーション分析に取り入れる有効性を示したのが始まりです。その後、90 年代の終わりになって、レキシカルプロファイリングの手法を本格的に取り入れた検索ツール Word Sketch が辞書執筆の現場で使われるようになり、レキシカルプロファイリングの有用性が広く認知されるに至りました。

レキシカルプロファイリングの検索方法は、7.1 で示したコンコーダンスの検索方法とはまったく異なります。先ほどの「頭が上がる」で説明してみます。今度は BCCWJ 用のオンラインコンコーダンス『中納言』を利用して見ます。中納言の利用には文書による申請が必要です。

▶ 中納言 : <https://chunagon.ninjal.ac.jp/>

中納言には、短単位検索、長単位検索、文字列検索の3種類がありますが、ここでは短単位検索を利用します。まず [キー] に「(語彙素の) 頭」を指定します。[後方共起条件の追加] をクリックして、[後方共起 1] に「(語彙素の) が」を入力し、[後方共起条件の追加] をもう一度クリックして、[後方共起 2] に「(語彙素の) 上がる」を入力して、[検

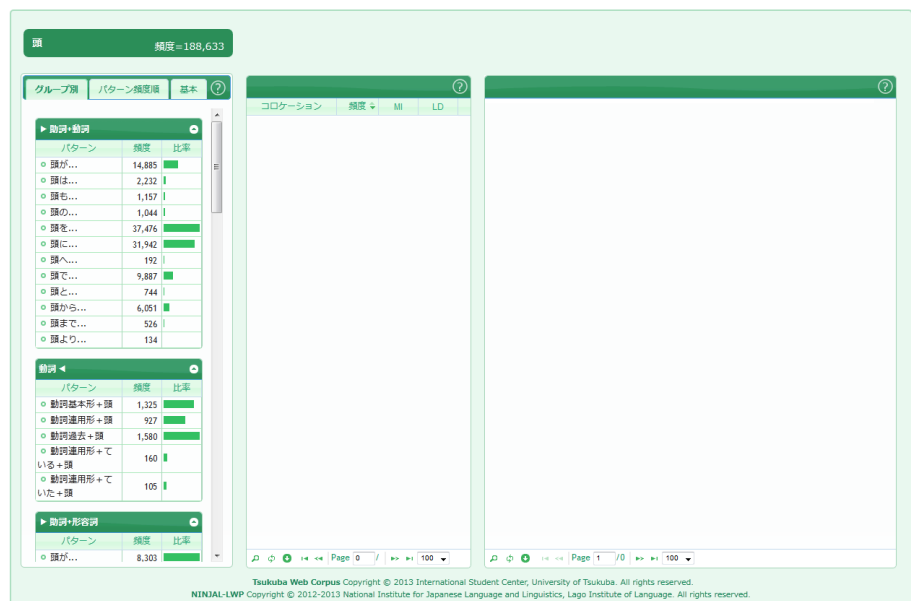
案] をクリックします。すると、以下のような検索結果が返ってきます。中納言では、ヘッダー部分をクリックすると、前文脈や後文脈などで並べ替えることができます。

48 件の結果が見つかりました。

テーブルの幅を固定 短

サンプル ID	前文脈	キー	後文脈	サブコーパス名	執筆者	書名/出典
PB49_00011	て、いやがる私たちを無理やり結婚させたの。ブルースはお父さんに	頭	(が)(上から)なかったのね。彼が家出したのだって、お父さんが	出版・書籍	サンドラ・フィールド(著)大沢 晶(訳)	ロマンスへの誘い
LBi9_00115	して睨んでるでな、やっぱし駄目だ、俺ア養子だでな、かかァに	頭	(が)(上から)ねえだ、そいで、どうでもこれだけ(三本指を立てる)	図書館・書籍	山田 洋次(著)	真二つ
OC01_10644	に火をつけて攻撃します。そのため、次期ボス候補の主人公は	頭	(が)(上から)ない。	特定目的・知恵袋		Yahoo!知恵袋
OC01_04143	女らしく、美人でおとやかに。反面、ノリスケは繊細な面もあり、ノリスケは	頭	(が)(あから)ない。 ※フジテレビのホームページで紹介されたアニメ版公式設定	特定目的・知恵袋		Yahoo!知恵袋
OY14_53827	でもあらゆるすべてのことについてあなたの気持ちに殉じてあなたには自分は	頭	(が)(上から)ないのだけどそれは永久に絶対にどんな場合も現実としても	特定目的・ブログ		Yahoo!ブログ
PB19_00061	局長、いつも言っていましたよ。うちの奴は揉まれればなし、全然	頭	(が)(上から)ん、恐ろしい奴だって」「あら、私そんなに嫌天下かしら。皆	出版・書籍	高橋 門之進(著)	浮橋
OY14_47461	でもあらゆるすべてのことについてあなたの気持ちに殉じてあなたには自分は	頭	(が)(上から)ないのだけどそれは永久に絶対にどんな場合も現実としても	特定目的・ブログ		Yahoo!ブログ
PB49_00457	ものね。小野 そればもう...表に出たら与太者連中、ハツちゃんにまる	頭	(が)(あから)ないんだ。『いえ、もう、先生にお手向い、なんてとんでもない。あつし	出版・書籍	坂口 安吾(著)大下 宇陀児(著)小林 真二(著)小野 佐世男(著)松井 翠声(著)淡谷 のり子(著)	安吾からの挑戦状

このように、コンコーダンスでは、検索語を詳細に指定して検索しますが、レキシカルプロファイリングでは、見出し語、文法パターン、コロケーションという順で調べていきます。ここでは、実際に NLT で「頭が上がる」を調べてみます。見出し語検索ウィンドウで、「あたま」と入力し、[絞り込み] をクリックします。次に見出し語リストの「頭」をクリックして、「頭」の見出し語ウィンドウを開きます。



ウィンドウが開いたら、左の文法パターンパネルの「グループ別」から文法パターンを選びます。一番上の「▶ 助詞+動詞」パターンの「頭が…」をクリックして、「頭が+動詞」のコロケーションを表示します。

The interface shows the 'グループ別' (Group) tab selected. Under '助詞+動詞', the 'パターン' (Pattern) list includes '頭が...' which is selected. The 'コロケーション' (Collocation) table on the right lists various combinations of '頭が' followed by a verb, along with their frequency, MI, and LD scores.

コロケーション	頻度	MI	LD
頭が下がる	1,782	11.26	9.69
頭がなる	1,428	4.17	2.94
頭がいる	1,251	11.59	9.80
頭がする	675	25.62	10.54
頭がくる	518	13.03	9.76
頭がしまう	483	11.74	9.27
頭が上がる	415	8.23	6.82
頭が混乱する	360	11.12	8.76
頭が回る	344	8.91	7.34
頭がある	341	1.99	0.76
頭が出る	274	4.87	3.62
頭がすつきりする	260	10.80	8.35
頭が働く	235	6.57	5.24
頭が冴える	205	12.08	8.50
頭がくらくらする	187	14.95	8.64
頭が見える	181	5.37	4.09
頭が切れる	151	8.73	6.88
頭がてる	143	4.42	3.16
頭が動く	142	6.73	5.31
頭がつく	121	4.80	3.52
頭がいく	111	5.80	4.44

中ほどより下にある「頭が上がる」をクリックします。右の用例パネルに「頭が上がる」の用例が表示されます。

The interface shows the 'コロケーション' table updated with '頭が上がる' selected. The '用例' (Examples) panel on the right displays several sentences using the pattern, such as '頭が上がりません。' and 'もう頭が上がりません。'.

このように、レキシカルプロファイリングでは、見出し語、文法パターン、コロケーションという順に調べていきます。調べたい表現だけでなく、類義表現なども同時に調べることができる点もレキシカルプロファイリングのメリットの一つです。

5. NLT の2つのウィンドウ

NLT には2種類のウィンドウがあります。見出し語を検索するための見出し語検索ウィンドウと、それぞれの見出し語の振る舞いを調べるための見出し語ウィンドウです。

5.1 見出し語検索ウィンドウ

見出し語検索ウィンドウは、見出し語を表示する見出し語リストと、見出し語を絞り込むための入力ボックスとボタン、絞り込みを解除するためのボタンから構成されます。

The screenshot shows the NINJAL-LWP search interface. At the top, there is a search input field with the text 'ここに調べたい語を入力 (読みも可)'. Below it are buttons for '絞り込み' and '元に戻す'. A navigation bar includes 'すべて', '名詞', '動詞', '連体詞', '形容詞', and '副詞'. The main content is a table with columns: '見出し', '読み', 'ローマ字表記', and '頻度'. The table lists various words and their frequencies. At the bottom, there is a footer with copyright information: 'Tsukuba Web Corpus Copyright © 2013 International Student Center, University of Tsukuba. All rights reserved. NINJAL-LWP Copyright © 2012-2013 National Institute for Japanese Language and Linguistics, Lingo Institute of Language. All rights reserved.'

見出し	読み	ローマ字表記	頻度
いる-非自立	イル	iru	11,402,357
こと	コト	koto	9,542,386
の	ノ	no	6,953,360
する	スル	suru	6,634,633
ある	アル	aru	6,572,844
なる	ナル	naru	6,062,116
よう	ヨウ	you	3,839,372
その	ソノ	sono	3,568,598
できる	デキル	dekiru	3,428,364
この	コノ	kono	2,997,119
思う	オモウ	omou	2,670,036
言う	イウ	iu	2,534,361
もの	モノ	mono	2,498,040
ため	タメ	tame	2,050,227
人	ヒト	hito	1,801,373
場合	バアイ	baai	1,745,676
くる-非自立	クル	kuru	1,654,006
とき	トキ	toki	1,500,975
自分	ジブン	jibun	1,365,966

見出し語リストの【見出し】の列をクリックすると、その見出し語のウィンドウ（タブ）が開きます。

The screenshot shows the NINJAL-LWP search interface with the search results for 'する' (frequency: 6,634,633). The search results table is partially visible, showing 'する' and 'ある'. A mouse cursor is pointing at the 'する' entry. To the right, a detailed view window for 'する' is open, showing a table of patterns and their frequencies and ratios. The detailed view window has a title bar 'する 頻度=6,634,633' and a navigation bar with 'グループ別', 'パターン頻度順', and '基本'. The table in the detailed view has columns: 'パターン', '頻度', and '比率'. At the bottom, there is a footer with copyright information: 'Tsukuba Web Corpus Copyright © 2013 International Student Center, University of Tsukuba. All rights reserved. NINJAL-LWP Copyright © 2012-2013 National Institute for Japanese Language and Linguistics, Lingo Institute of Language. All rights reserved.'

パターン	頻度	比率
○...がする	441,962	
○...はする	352,601	
○...もする	188,301	
○...のする	16,366	
○...をする	2,669,860	
○...にする	1,757,787	
○...へする	8,026	
○...でする	296,209	
○...とする	857,603	
○...からする	65,291	
○...までする	40,125	
○...よりする	9,025	

5.2 見出し語ウィンドウ

見出し語ウィンドウは、3つのパネルから構成されます。左から、文法パターン・基本情報パネル、コロケーションパネル、用例パネルです。見出し語の振る舞いを調べるときは、左のパネルから右のパネルに向かって、文法パターン、コロケーション、用例という順に見ていきます。

The screenshot shows the NINJAL-LWP interface for the word '必要' (necessary). The top bar indicates the search term and frequency: '必要 頻度=1,191,471'. Below this are three main panels:

- Left Panel (Filters):** Contains 'グループ別' (Group), 'パターン頻度順' (Pattern Frequency), and '基本' (Basic). It lists various patterns like '必要が...' with their frequencies and ratios.
- Middle Panel (Collocations):** A table titled '必要が... 209種類' (Necessary... 209 types) with columns for 'コロケーション' (Collocation), '頻度' (Frequency), 'MI', and 'LD'. It lists various collocations such as '必要がある', '必要がくる', etc.
- Right Panel (Examples):** A list of example sentences under the heading '必要がある 全224666件' (Necessary all 224,666 items). Examples include '(新入社員の心構えについて)', '(これで納得！文章問題の解き方)', etc.

At the bottom, there is copyright information: 'Tsukuba Web Corpus Copyright © 2013 International Student Center, University of Tsukuba. All rights reserved. NINJAL-LWP Copyright © 2012-2013 National Institute for Japanese Language and Linguistics, Lingo Institute of Language. All rights reserved.'

用例の前後を見るときは、文脈を表示するダイアログで確認できます。前後5センテンス（句点で終わるもの）を見ることが出来ます。

This screenshot shows the same interface as above, but with a context dialog box open over an example sentence. The dialog box is titled 'よくあるQ&A' (Common Q&A) and contains the following text:

Q 8. 玄米を食べ始めたら、アトピーが酷くなりました。 玄米が合わないのでしょうか？
 A. 先にも述べたように、玄米も、正しい炊き方、食べ合わせが肝心です。
必要があります。 まずいと思って食べているなら止めてください。
 アトピーが酷くなったように見えますが、それは排毒といって、今まで、体の中に溜め込んでいた毒素を出す反応（好転反応）が起こります。ただし、この場合は、正しい知識を持った人のアドバイスを受けながら続けることをおすすめします。
 自分で勝手に思い込まないことも大切です。

The dialog box has a '閉じる' (Close) button at the bottom right. The background interface remains visible but slightly dimmed.

6. 見出し語検索ウィンドウの操作

6.1 見出し語ウィンドウの開き方

見出し語リストの見出し列をクリックすると、その見出し語のウィンドウ（タブ）が開きます。

6.2 見出し語リストの並べ替え

見出し語リストは、ウィンドウを開いた最初の状態では頻度の高い順に表示されています。リストは、[読み] と [ローマ字表記] と [頻度] で並べ替えできます。図のように、[読み] のヘッダー部分をクリックすると見出し語の読み順（昇順）に並べ替えます。もう一度クリックすると、今度は見出し語の読み順（降順）に変わります。[ローマ字表記] も [読み] と同じです。[頻度] の場合は、最初にクリックしたときは高頻度順（降順）、もう一度クリックすると低頻度順（昇順）になります。

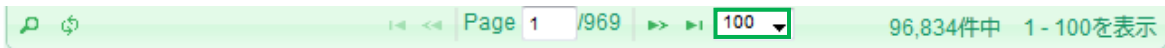
見出し	読み	ローマ字表記	頻度
ああ	アア	aa	6,134
愛	アイ	ai	55,817
藍	アイ	ai	9,065
アイアイ	アイアイ	ai ai	347
譚々	アイアイ	ai ai	5
譚譚	アイアイ	ai ai	1

6.3 見出し語リストのページ

見出し語リストは、初期設定で1ページに100語の見出し語が表示されます。リストを順に見ていくときは、リスト下の [ページ切り替えボタン] をクリックしてください。ボタンの間にある入力ボックスに直接ページを入力して、指定したページにジャンプすることもできます。



1ページに表示する見出し語数は、[ページ切り替えボタン]の右のコンボボックスで変更できます。1ページの見出し語数は、100語、200語、300語の3つから選べます。



6.4 見出し語リストの切り替え

見出し語リストには6種類のリストがあります。[名詞]、[動詞]、[形容詞]、[連体詞]、[副詞]ではそれぞれの品詞の見出し語が、[すべて]ではこれら5つの品詞のすべての見出し語が表示されます。リストの上部にあるタブをクリックすると、見出し語リストが切り替わります。ナ形容詞の扱いについては、8.7をご覧ください。



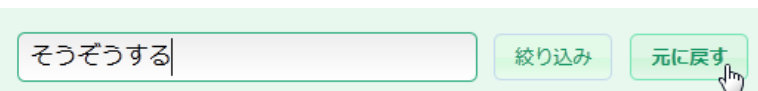
6.5 見出し語の検索 (絞り込み)

特定の見出し語を検索したいときは、見出し語リストの上にある入力ボックスに、その見出し語か、その読み（ひらがな、またはカタカナ）か、ローマ字（ヘボン式）で入力します。ローマ字の表記については、27ページのローマ字一覧をご覧ください。見出し語は代表的な表記で示されていますが、ユーザはどの表記が代表的な表記が分からないので、ふつうは読みで検索することをお勧めします。入力できたら、エンターキーを押すか、隣の [絞り込み] ボタンをクリックしてください。

NINJAL-LWP for TWC

見出し	読み	ローマ字表記	頻度 ↓
想像する	ソウゾウスル	souzousuru	28,853
創造する	ソウゾウスル	souzousuru	13,520

絞り込んだ状態のままで並べ替えすることもできます。検索する見出し語が見つかったら、その見出し語をクリックして、見出し語ウィンドウを開きます。絞り込みの状態を解除して元に戻すには、[元に戻す] をクリックします。



見出し語は、完全一致以外に、前方一致、後方一致でも検索できます。前方一致で調べたいときは、先頭に「^」（全角でも半角でも可能）を付けます。右の例では、[すべて]の見出し語リストから「思」で始まる見出し語をすべて表示します。ローマ字表記でも前方一致で検索できます。

NINJAL-LWP for TWC

^思 絞り込み 元に戻す

すべて **名詞** 動詞 連体詞 形容詞 副詞

見出し	読み	ローマ字表記	頻度
思う	オモウ	omou	2,670,036
思い	オモイ	omoi	215,786
思える	オモエル	omoeru	96,574
思い出す	オモイダス	omoidasu	56,596
思想	シソウ	shisou	51,082
思考	シコウ	shikou	50,027
思い出	オモイデ	omoidede	34,161
思わず	オモワズ	omowazu	23,576
思い切る	オモイキル	omoiikiru	16,606
思いつく	オモイツク	omoitsumu	15,301
思い込む	オモイコム	omoiikomu	14,141
思いやり	オモイヤリ	omoiyari	10,756
思い切り	オモイッキリ	omoiikkiri	10,510
思春	シシュン	shishun	8,616
思い浮かべる	オモイウカベル	omoiukaberu	8,112
思い込み	オモイコミ	omoiikomi	6,403
思いやる	オモイヤル	omoiyaru	6,259
思惑	オモワク	omowaku	5,640
思い通り	オモイドオリ	omoidoori	5,395
思い起こす	オモイオコス	omoiokosu	4,852

Page 1 / 2 100 157件中 1 - 100を表示

後方一致の場合は、末尾に「\$」（全角でも半角でも可能）を付けます。右の例では、[動詞]の見出し語リストから非自立の動詞をすべて表示します。ローマ字表記でも後方一致で検索できます。

NINJAL-LWP for TWC

非自立\$ 絞り込み 元に戻す

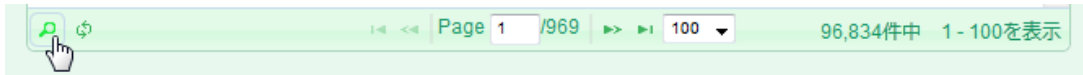
すべて **名詞** **動詞** 連体詞 形容詞 副詞

見出し	読み	ローマ字表記	頻度
いる-非自立	イル	iru	11,402,357
くる-非自立	クル	kuru	1,654,006
いく-非自立	イク	iku	1,358,504
下さる-非自立	クダサル	kudasaru	1,280,811
しまう-非自立	シマウ	shimau	1,165,628
おる-非自立	オル	oru	947,367
いただく-非自立	イタダク	itadaku	690,112
みる-非自立	ミル	miru	688,322
くれる-非自立	クレル	kureru	570,963
てる-非自立	テル	teru	509,649
なる-非自立	ナル	naru	486,777
おく-非自立	オク	oku	363,679
もらう-非自立	モラウ	morau	318,322
致す-非自立	イタス	itasu	314,263
いける-非自立	イケル	ikeru	247,997
続ける-非自立	ツツケル	tsuzukeru	189,377
過ぎる-非自立	スギル	sugiru	173,124
始める-非自立	ハジメル	hajimeru	166,195
ある-非自立	アル	aru	156,678
あげる-非自立	アゲル	ageru	95,352

Page 1 / 1 100 72件中 1 - 72を表示

6.6 見出しリストの絞り込み

前節では見出し語を絞り込む方法を示しましたが、見出し、読み、頻度の3つの項目を自由に組み合わせて絞り込むこともできます。見出し語リストの左下の「[フィルタ]」ボタンをクリックすると、「[フィルタ]」ダイアログが開きます。



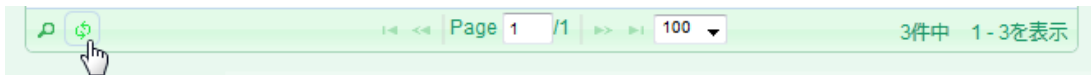
以下では、頻度が10万件以上の副詞だけを表示します。リストの右下に該当する見出しの件数(44件)が表示されます。

見出し	読み	ローマ字表記	頻度
どう	ドウ	dou	675,913
そう	ソウ	sou	467,560
さらに	サラニ	sarani	455,440
特に	トクニ	tokuni	368,974
もう	モウ	mou	358,521
少し	スコシ	sukoshi	354,194
とても	トテモ	totemo	345,798
実際	ジッサイ	jissai	324,788
やはり	ヤハリ	yahari	294,376
よく	ヨーク	yoku	284,004
まず	マズ	mazu	282,067
本当に	ホントウニ	hontouni	277,406
まったく	マッタク	mattaku	259,233
かなり	カナリ	kanari	240,389
すぐ	スグ	sugu	239,610
必ず	カナラズ	kanarazu	237,443

さらに複数の条件を指定することもできます。条件を追加するときは、「[フィルタ]」ダイアログの「+」ボタンをクリックします(逆に「-」ボタンをクリックすると、その条件が削除されます)。以下では、「と」で終わる頻度10万件以上の副詞を表示します。複数の条件を指定するときは、ダイアログの一番下にあるコンボボックスで、「すべての」条件を満たす見出しを表示するか、それとも「いずれかの」条件を満たす見出しを表示するかを指定します。

見出し	読み	ローマ字表記	頻度
ちよっと	チョット	chotto	226,749
もっと	モット	motto	192,997
ずっと	ズット	zutto	108,744

条件を解除してすべての見出し語を表示するには、[フィルタ] ダイアログの[リセット]をクリックするか、見出し語リストの左下の[リセット] ボタンをクリックしてください。



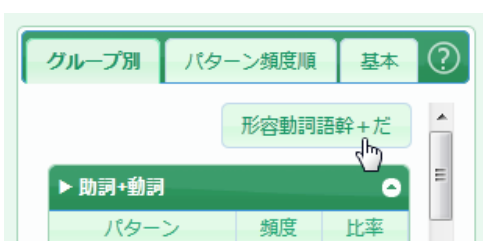
6.7 ナ形容詞の扱い

NLTでは、形態素解析にIPA辞書を利用しています。IPA辞書にはナ形容詞（あるいは形容動詞）という品詞分類はありません。そのため、NLTでは「形容動詞の語幹となる名詞」と「助動詞ダ」が連続する場合にナ形容詞と判定する便宜的な方法を用いています。検索するときは、使いやすさを考慮して、語幹となる名詞からでも、形容詞からでも検索できるようにしています。例えば、「必要な」というナ形容詞は、語幹となる「必要」という名詞からでも、「必要な」という形容詞からでも検索することができます。

まず語幹となる名詞「必要」から検索してみます。名詞の見出し語リストを選択し、入力ボックスに「ひつよう」と入力し、[絞り込み] ボタンをクリックします。見出し語リストの「必要」をクリックすると、名詞「必要」の見出し語ウィンドウが開きます。

グループ別	パターン	頻度	比率
助詞+動詞	必要が...	238,862	
	必要は...	18,676	
	必要も...	6,426	
	必要の...	2,356	
	必要を...	2,883	
	必要に...	71,191	
	必要へ...	1	
	必要で...	44	
	必要と...	147,688	
	必要から...	593	
	必要まで...	29	
必要より...	8		
動詞	動詞基本形+必要	295,737	
	動詞連用形+必要		

形容詞としての用法を調べたいときは、[文法パターン・基本情報パネル] の[グループ別] タグの右上にある[形容動詞語幹+だ] というボタンをクリックします。



以下のように、ナ形容詞（形容動詞語幹+助動詞ダ）の用法のパターンにジャンプします。

グループ別	パターン頻度順	基本	?
形容動詞語幹+な+名詞			
パターン	頻度	比率	
○ 必要な+名詞	248,391	<div style="width: 100%;"></div>	
名詞+助詞 ⇄ 形容動詞語幹+だ			
パターン	頻度	比率	
○ ...が必要だ	24,968	<div style="width: 100%;"></div>	
○ ...は必要だ	5,346	<div style="width: 10%;"></div>	
○ ...も必要だ	3,575	<div style="width: 10%;"></div>	
○ ...の必要だ	3		
○ ...を必要だ	109		
○ ...に必要だ	6,217	<div style="width: 10%;"></div>	
○ ...で必要だ	1,106	<div style="width: 10%;"></div>	
○ ...と必要だ	73		
○ ...から必要だ	236		
○ ...まで必要だ	80		
○ ...より必要だ	70		

形容詞「必要な」から検索する場合は、形容詞の見出し語リストを選択し、入力ボックスに「ひつような」と入力し、[絞り込み] ボタンをクリックします。見出し語リストの「必要な」をクリックすると、先ほどと同じ名詞「必要」の見出し語ウィンドウが開きます。これ以降の操作は、先ほどの名詞の場合と同じです。

ひつような	絞り込み	元に戻す			
すべて	名詞	動詞	連体詞	形容詞	副詞
見出し	読み	ローマ字表記	頻度	?	
必要な	ヒツヨウナ	hitsuyouna	1,191,471		

7. 見出し語ウィンドウの操作

7.1 見出し語ウィンドウの構成

本章では見出し語のウィンドウの使い方について説明します。見出し語ウィンドウは、以下の3つのパネルから構成されます。

文法パターン・基本情報パネル

パターン	頻度	比率
...が切る	5,160	
...は切る	5,184	
...も切る	1,469	
...の切る	702	
...を切る	43,425	
...に切る	13,242	
...へ切る	243	
...で切る	8,819	
...と切る	838	
...から切る	1,662	
...まで切る	753	
...より切る	114	

パターン	頻度	比率
...といえども切る	1	

コロケーションパネル

コロケーション	頻度	MI	LD
電源を切る	2,091	10.87	9.88
電話を切る	1,946	7.70	7.79
スタートを切る	1,806	10.32	9.52
縁を切る	1,318	9.67	8.97
木を切る	1,292	8.23	8.08
首を切る	1,101	8.77	8.36
シャッターを切る	917	11.44	9.17
ハンドルを切る	873	10.74	8.98
梶を切る	772	12.09	9.04
髪を切る	694	9.25	8.28
スイッチを切る	585	9.69	8.27
円を切る	560	4.84	5.05
枝を切る	495	8.90	7.84
水を切る	494	5.30	5.45
水気を切る	482	12.37	8.43
【一般】を切る	445	1.08	1.39
手を切る	439	4.99	5.15
風を切る	429	6.38	6.30

用例パネル

電源を切る 全2091件

- 火元・電源を切る。
(千代田区防災ホームページ・地震対策のソフトウェア～管理上のポイント～)
- 不要時は電源を切る。
(ラジオ(雑音障害)と災害時の機材について | 受信サービス株式会社)
- フィルタの電源を切る。
(tips)
- 電源を切ってもいいの？
(ETCカードで高速道路料金を安くする裏ワザ)
- ぶつりと、電源を切る。
(MARIA)
- 電源を切ってください。
(NV-GS150/GS250 回答 | よくあるご質問一覧 | デジタルビデオカメラ | お客様サポート | Panasonic)
- 携帯電話の電源を切る。
(障害者のための「就活」マニュアル・働きたい!と思ったら - ATARIMAEプロジェクト)

7.2 文法パターン・基本情報パネル

文法パターン・基本情報パネルは、以下の3つのタブから構成されます。[グループ別]と[パターン頻度順]は文法パターンを表示するためのタブで、中央のコロケーションパネルと連動しています。[基本]には基本的な頻度情報が集約されています。

7.2.1 基本情報タブ

基本情報タブは、[書字形]、[活用形]、[後続助動詞の割合]の3つのパネルで構成されます。ヘッダーをクリックするとそのパネルが開くようになっています。活用のない名詞や副詞や連体詞の基本情報タブには、[活用形]と[後続助動詞の割合]のパネルはありません。基本情報タブは、他の文法パターンを表示するタブとは異なり、コロケーションパネルと連動しません。以下、動詞「申す」を例にして、5つのパネルを説明します。

❖ [書字形]には、それぞれの表記の頻度と割合が表示されます。漢字表記が96%を占めていることが分かります。

書字形	頻度	割合
申す	24874	96%
もうす	943	4%
白す	49	0%

❖ [活用形] には、それぞれの活用形の頻度と割合が表示されます。このパネルは、活用形のある内容語（動詞、形容詞）のみです。「申す」は連用形での使用が8割を占めていることが分かります。

▼ 活用形		
連用形	20787	80%
基本形	2821	11%
未然形	1233	5%
仮定形	498	2%
命令e	314	1%
未然ウ接続	206	1%
仮定縮約1	7	0%

❖ [後続助動詞の割合] には、動詞または形容詞の直後に助動詞が後続する頻度と割合が表示されます。動詞の場合は、「れる・られる」、「せる・させる（使役）」、「ない・ぬ・ません（否定）」の助動詞、形容詞の場合は「ない・ぬ・ありません（否定）」が表示されます。左が動詞「申す」、右が形容詞の「素晴らしい」の例です。「素晴らしい」は10万件ほどの例がありますが、否定形が直後に後続するのはわずか32件で、否定形では用いられることはきわめてまれであることが分かります。

▼ 後続助動詞の割合		
れる・られる	757	2.9%
申される	757	
せる・させる	73	0.3%
申させる	73	
ない・ぬ	747	2.9%
申しません	528	
申さぬ	179	
申さない	40	

▼ 後続助動詞の割合		
ない・ぬ	32	0%
素晴らしい	32	

9.2.2 グループ別タブ

[グループ別] タブと、次のサブセクションで説明する [パターン頻度順] タブは、どちらも文法パターンを一覧表示します。[グループ別] では、文法パターンがグループごとに分類して表示され、[パターン頻度順] では、文法パターンが頻度の高い順に表示されます。どちらのタブも、パターンの列をクリックすると、そのパターンのコロケーションが右のコロケーションパネルに表示されます。以下では、「時間を+動詞」のコロケーションが表示されています。

グループ別	パターン頻度順	基本
▶ 動詞・動詞		
パターン	頻度	比率
○ 時間が...	106,380	
○ 時間は...	26,304	
○ 時間も...	12,111	
○ 時間の...	5,056	
○ 時間を...	130,904	
○ 時間に...	37,241	
○ 時間へ...	109	
○ 時間で...	14,693	
○ 時間と...	5,775	
○ 時間から...	1,604	
○ 時間まで...	2,464	
○ 時間より...	734	

動詞	パターン	頻度	比率
○	動詞基本形+時	54,841	

時間を..., 2255種類			
コロケーション	頻度	MI	LD
時間をかける	16,137	9.29	10.54
時間を過ごす	7,799	9.72	10.24
時間をいる	6,204	10.76	10.29
時間をする	4,836	25.33	10.24
時間を要する	3,911	23.02	9.94
時間を使う	3,642	5.44	7.12
時間をいただく	3,500	6.79	8.14
時間を費やす	2,927	11.16	9.40
時間を超える	2,891	7.67	8.54
時間を作る	2,740	5.25	6.89
時間を割く	2,538	11.67	9.24
時間をとる	2,443	6.74	7.93
時間をできる	2,163	2.46	4.32
時間を下さる	1,960	3.73	5.51
時間を取る	1,827	6.16	7.40
時間を持つ	1,559	3.83	5.56
時間をしまう	1,458	10.20	8.40
時間をおく	1,411	7.96	8.00

文法パターンのグループは、品詞ごとに設定されています。動詞の場合、以下の9のグループに分類されます。どのパターンにも属さなかった用例は未分類としてまとめています。

ID	文法パターングループ	説明
A	名詞+助詞 ◀	動詞の前に名詞+格助詞が先行するパターン
B	名詞+複合助詞 ◀	動詞の前に名詞+複合助詞が先行するパターン
C	▶ 名詞	動詞の直後に名詞が後続するパターン
E	▶ 助動詞	動詞の直後に助動詞が後続するパターン
F	複合動詞	複合動詞の前項または後項にくるパターン
G	近接動詞	当該動詞と前後3~5語以内の共起する動詞とのパターン
H	▶ 形容詞	動詞の直後に非自立の形容詞が後続するパターン
I	副詞 ◀	動詞とそれを修飾する副詞のパターン
J	形容詞連用形 ◀	動詞とそれを修飾する形容詞連用形のパターン
Z	(未分類)	上記のどのパターンにも属さない用例

7.2.3 パターン頻度順タブ

[パターン頻度順] タブは、文法パターンが頻度の高い順に表示されます。頻度と見出し語の全頻度に占める比率が示されます。比率は棒グラフで示されていますが、棒グラフの上にマウスのカーソルを置くとその割合(数字)が表示されます。このタブは、見出し語の振る舞いの全体像をつかむときに利用すると便利です。以下は、名詞の「可能」の例です。助動詞が後続するパターンが最も頻度が高く、その右のコロケーションを見ると「可能だ」「可能です」という形がほとんどを占めていることが分かります。

The screenshot shows two panels. The left panel, titled 'グループ別' (Group by), displays a list of grammar patterns sorted by frequency. The right panel, titled '可能+助動詞 10種類' (Possible + Auxiliary Verbs 10 types), shows collocation data for the top 10 patterns.

パターン	頻度	比率
可能+助動詞	260,789	■
可能+名詞・接尾	247,463	■
動詞基本形+可能	139,368	■
名詞+可能	111,441	■
可能+助詞	95,111	■
可能な+名詞	85,804	■
可能に...	50,662	■
可能に+動詞	45,402	■
可能と+動詞	42,011	■
名詞+の+可能	34,976	■
体言止め	16,711	■
動詞連用形+ている+可能	14,319	■
可能+名詞	11,996	■
接頭辞+可能	9,561	■
動詞過去+可能	8,570	■
連体詞+可能	6,399	■
...が可能だ	3,011	■

コロケーション	頻度	MI	LD
可能だ	142,559	4.97	7.86
可能です	117,996	5.58	8.46
可能らしい	92	1.17	2.78
可能や	58	2.20	2.59
可能なり	54	1.55	2.37
可能じゃ	17	2.20	1.01
可能たり	6	-0.92	-0.66
可能っす	5	1.60	-0.71
可能たい	1	-8.38	-5.71
可能やす	1	-1.79	-3.07

7.3 コロケーションパネル

コロケーションパネルは、文法パターンパネルで選んだコロケーションを表示します。以下は、「時間が+動詞」のコロケーションの例です。最初に表示したときはコロケーションは頻度の高い順に並んでいますが、MI スコア、LogDice 係数という2種類の統計値でも並べ替えることもできます。並べ替えるときは、それぞれのヘッダーをクリックします。

7.3.1 コロケーションパネルの統計値

❖MI スコアで並べ替えると、頻度にかかわらず特徴的なコロケーションが上位に現れます。

高頻度順で表示

コロケーション	頻度	MI	LD
時間がかかる	33,435	10.38	11.51
時間がある	10,613	4.11	5.70
時間が経つ	5,178	10.22	10.12
時間がある	4,880	10.72	10.18
時間がしまう	3,754	11.86	10.04
時間が過ぎる	3,286	13.36	9.94
時間が掛かる	2,898	10.33	9.51
時間になる	2,464	2.12	3.71
時間がたつ	2,238	9.13	8.94
時間ができる	1,827	2.51	4.08
時間が経過する	1,706	9.45	8.72

MI スコア順で表示

コロケーション	頻度	MI	LD
時間がする	198	21.01	5.93
時間が不足する	128	20.38	5.30
時間がはっきりする	21	16.78	2.69
時間が達する	10	16.71	1.62
時間が要する	34	16.47	3.39
時間がオーバーする	46	16.32	3.82
時間が過ぎる	3,286	13.36	9.94
時間がセットする	1	12.38	-1.70
時間がしまう	3,754	11.86	10.04
時間がくる	1,029	11.19	8.25
時間が現状する	1	10.80	-1.70

ただし、MI スコアの特性として、低頻度のコロケーションが過度に強調されるため、低頻度のものを排除したほうが人間の直感に近い結果が得られます。パネル上で右クリックすると、頻度を設定するメニューが表示されます。どの頻度を選ぶかはケースバイケースです。[頻度 10 以上] で十分絞り込めていない場合は [頻度 20 以上] にするなど、実際の結果を見ながら判断するのがよいでしょう。

コロケーション	頻度	MI	LD
時間がする	198	21.01	5.93
時間が不足する	128	20.38	5.30
時間がはっきりする	21	16.78	2.69
時間が達する	10	16.71	1.62
時間が要する	34	16.47	3.39
時間がオーバーする	46	16.32	3.82
時間が過ぎる	3,286	13.36	9.94
時間がセットする	1	12.38	-1.70
時間がしまう	3,754	11.86	10.04
時間がくる	1,029	11.19	8.25
時間がある	4,880	10.72	10.18
時間がかかる	33,435	10.38	11.51

コロケーション	頻度	MI	LD
時間がする	198	21.01	5.93
時間が不足する	128	20.38	5.30
時間がはっきりする	21	16.78	2.69
時間が達する	10	16.71	1.62
時間が要する	34	16.47	3.39
時間がオーバーする	46	16.32	3.82
時間が過ぎる	3,286	13.36	9.94
時間がしまう	3,754	11.86	10.04
時間がくる	1,029	11.19	8.25
時間がある	4,880	10.72	10.18
時間がかかる	33,435	10.38	11.51

❖LD は LogDice 係数で、コロケーションの統計処理でよく利用されるダイス係数を対数化したものです。共起頻度、見出し語の頻度、共起語頻度の3つの変数から導いた統計値で、降順に並べると、MI スコアよりは単純頻度に近い結果が出ます。

時間が... 269種類			
コロケーション	頻度	MI	LD
時間がかかる	33,435	10.38	11.51
時間がある	4,880	10.72	10.18
時間が経つ	5,178	10.22	10.12
時間がしまう	3,754	11.86	10.04
時間が過ぎる	3,286	13.36	9.94
時間が掛かる	2,898	10.33	9.51
時間がたつ	2,238	9.13	8.94
時間が経過する	1,706	9.45	8.72
時間が取れる	1,693	8.20	8.37
時間がくる	1,029	11.19	8.25
時間がとれる	1,154	8.29	8.02

7.3.2 コロケーションパネルでの固有名詞と数詞の表示

固有名詞については、以下の4つのグループにまとめて表示しています。このうち、【一般】には、形態素辞書にない固有名詞以外の未知語が含まれることがあります。また、数詞については、【数字】としてまとめて表示されます。

コロケーションパネルでの表示	該当する固有名詞
【人名】	姓や名、著名人の名前など
【地域】	一般的な地名や国名など
【組織】	企業などの組織名
【一般】	上記以外の固有名詞
【数字】	数詞

以下は、動詞「読む」の「名詞+を読む」のコロケーションに現れた「人名+を読む」の例です。

...を読む 3828種類

コロケーション	頻度	MI	LD
手紙を読む	880	7.94	7.56
テキストを読む	844	7.78	7.47
のを読む	819	0.17	1.92
空気を読む	816	6.89	7.18
雑誌を読む	768	7.26	7.24
話を読む	763	3.94	5.40
【人名】を読む	732	0.10	1.85
先を読む	726	4.17	5.56
ことを読む	724	-0.46	1.30
説明を読む	724	5.01	6.15
資料を読む	709	5.50	6.43
原作を読む	708	8.72	7.42
メールを読む	700	5.14	6.21
続きを読む	693	9.26	7.45
歌を読む	683	6.84	6.99
文献を読む	653	8.17	7.25
書籍を読む	624	7.79	7.13
コメントを読む	619	6.46	6.77

【人名】を読む 全732件

- 眠る前に **荘子** を読む。
- 谷崎を **読む** で耐寒睡。
- 創竜 **伝** を **読み** ました。
- マタイを **読み** 始めます。
- 僕も **マルサス** を **読む** だ。
- **漱石** を **読む** だのはいつ？
- ボルヘス** を **読む** だ時と同じ。
- プレヴェール** を **読む** できます。

7.3.3 特定のコロケーションの検索（フィルタ）

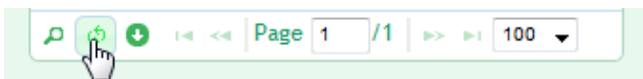
特定のコロケーションを探したいときは、画面左下の［フィルタ］ボタンをクリックして［フィルタ］ダイアログを表示します。



［フィルタ］ダイアログの使い方は、見出し語検索ウィンドウの見出し語リストの場合と同様です（8.6を参照）。以下では、「時間が+動詞」のコロケーションのリストから「時間が過ぎ去る」を検索している例です。



絞り込んだ状態を元に戻すには、［フィルタ］ダイアログの［リセット］ボタンをクリックするか、コロケーションパネルの［フィルタ］ボタンの右の［フィルタ解除］ボタンをクリックします。



7.3.4 コロケーションリストのダウンロード

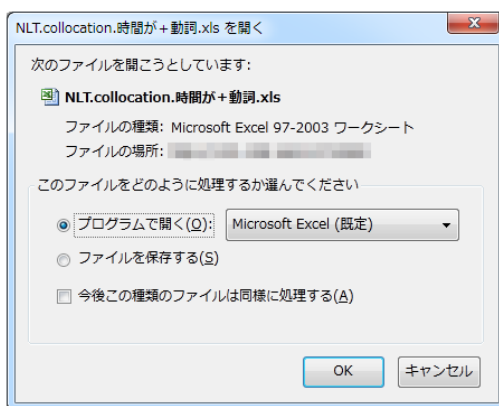
表示中のコロケーションのリストをダウンロードしたいときは、コロケーションパネルの左下の［ダウンロード］ボタンをクリックします。



クリックすると、ダウンロードするファイルの形式を選択するダイアログが表示されます。ファイル形式は、Excel ファイル（XLS ファイル）と CSV ファイル（カンマ区切りのファイル）の 2 種類です。ダウンロードできるのは表示中パターンの全コロケーションのリストです。フィルターダイアログで絞り込んだ結果を保存することはできません。



ファイル形式を選択して、[ダウンロード] ボタンをクリックすると、FireFox の場合、以下のようなダイアログが表示されますので、[プログラムで開く] か、[ファイルを保存する] かを選び、[OK] ボタンをクリックします。



以下は、Excel ファイルでダウンロードし、Excel で開いたときの画面です。出力される項目は、左から順に、コロケーション、頻度、MI スコア、ログダイスです。csv ファイルの場合も同様です。

	コロケーション	頻度	MIスコア	ログダイス
1				
2	<時間>がかかる	33435	10.38	11.51
3	<時間>がある	10613	4.11	5.7
4	<時間>が経つ	5178	10.22	10.12
5	<時間>がいる	4980	10.72	10.18
6	<時間>がしまつ	3754	11.86	10.04
7	<時間>が過ぎる	3286	13.36	9.94
8	<時間>が掛かる	2898	10.33	9.51
9	<時間>がなる	2464	2.12	3.71
10	<時間>がたつ	2238	9.13	8.94
11	<時間>ができる	1827	2.51	4.08
12	<時間>が経過する	1706	9.45	8.72
13	<時間>が取れる	1693	8.2	8.37
14	<時間>が増える	1300	6.15	7.14
15	<時間>が流れる	1222	7.53	7.82
16	<時間>がとれる	1154	8.29	8.02
17	<時間>が足りる	1149	8.16	7.98
18	<時間>がくる	1029	11.19	8.25
19	<時間>がいく	886	5.96	6.81
20	<時間>が異なる	767	5.31	6.33

[ファイル情報] のシートには、ダウンロードしたコロケーションの情報が記載されています。

	A
1	ファイル情報
2	NLTバージョン： 1.10
3	見出し語：時間 (id=N.00022)
4	パターン：時間が+動詞
5	パターン頻度：106380
6	コロケーション数：1410
7	出力日：2013-04-01 14:33

7.4 用例パネル

用例パネルは、コロケーションパネルで選択したコロケーションを含む用例を表示します。用例は短い順に表示されます。以下は、「時間がかかる」の用例です。

時間がかかる 全33435件

- 時間がかかる。
(歎異抄講読「後序について」) →
- 時間がかかる。
(みなと総合法律事務所 ウェブサイト) →
- 時間がかかる。
(ファイナルファンタジータクティクス 獅子戦争 レビュー(評価):: PSP mk2) →
- 時間がかかる。
(富士日記2) →
- 時間がかかる。
(多面評価(360度評価)ヒント | 株式会社ジイ・ディー・エル (GDL)) →
- 時間がかかる。
(震度7の朝、妻は妊娠9ヶ月だった) →
- 時間がかかるの。
() →
- 時間がかかります。
(さんさんの会 3.11 三陸気仙復興協議会【公式ブログ】 : 家がある人となない人の地元コミュニティーの問題) →

Page 1 / 335 100 33,435件中 1 - 100を表示

7.4.1 用例の前後の文脈の表示

各用例の出典（グレーの文字の部分）をクリックすると、前後の文脈を表示するダイアログが開きます。

時間がかかる 全33435件

■ 時間がかかる。
(歎異抄講読「後序について」) →

■ 時間がかかる。
(みなと総合法律事務所 ウェブサイト) →

歎異抄講読「後序について」

二十過ぎから四十五位までが本当にしっかりする時である。 僕もそうだと思う。 けれども、も一つ幾つでも手おくれはないということも力説しておかなくてはならないと思う。

我々は思知らずだとか仏法無視だとかは仲々わからない。 とくに若い時ではわからない。 時間がかかる。 その時間に耐えなければならない。 そしてついに本当に自分が何であるかがわかってこなければいけない。 これが「諸々の天下幽冥の所を炎照す」である。 そして願わくば「四十までに一応の成果を」となりたい。 二十過ぎで出発すれば大体そうなっていくであろうと思う。

http://homepage3.nifty.com/Tannisho/Gojo/6_1.html

閉じる

7.4.2 特定の表現を含む用例の検索（フィルタ）

特定の表現を含む用例を検索することもできます。以下は、「髷髻を買う」の用例です。全体で 413 件あります。



用例を読んでいくと、「周囲の髷髻を買う」という表現が 2 件見つかります。この表現が全体(413 件)のうちどのくらいあるのかを調べてみます。用例パネルの左下の [フィルタ] ボタンをクリックし、[フィルタ] ダイアログに以下のように入力し、[フィルタ] ボタンをクリックします。



以下のように、13 件の用例が見つかります。

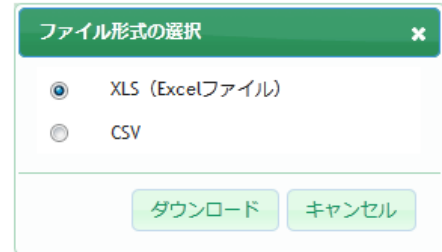


7.4.3 用例のダウンロード

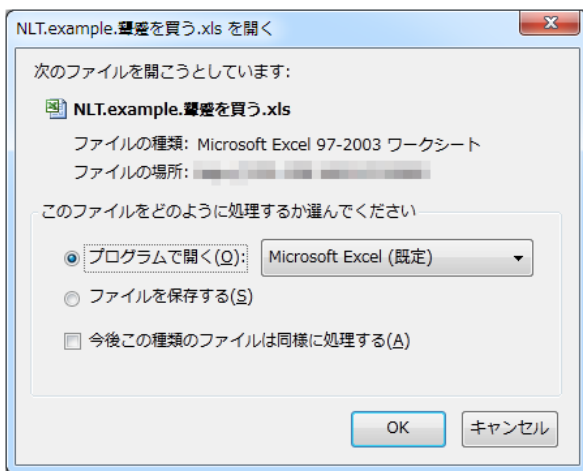
表示中の用例をダウンロードしたいときは、コロケーションパネルの左下の [ダウンロード] ボタンをクリックします。



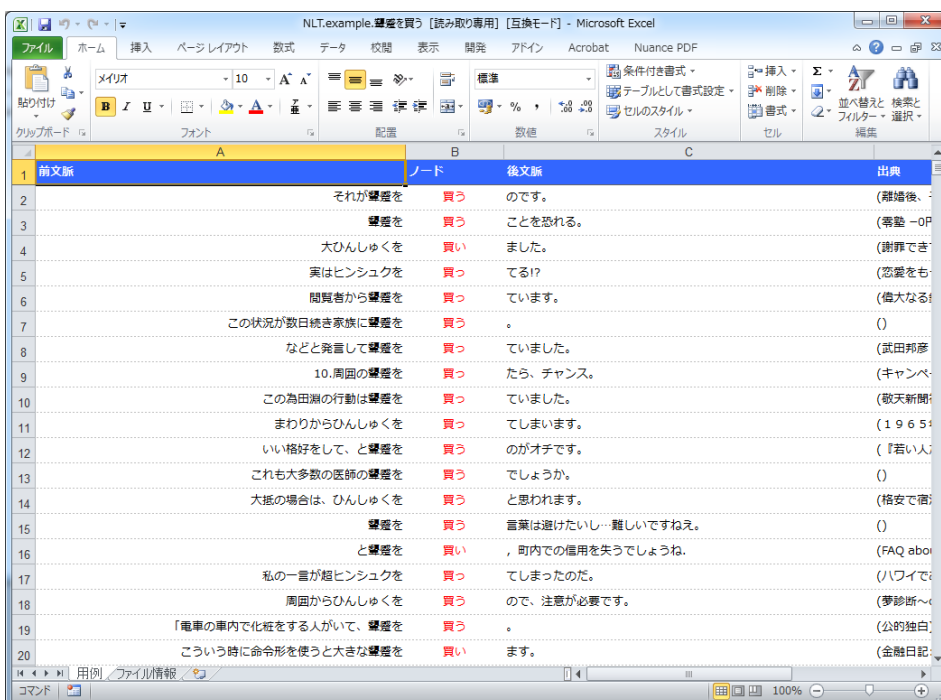
クリックすると、ダウンロードするファイルの形式を選択するダイアログが表示されます。ファイル形式は、コロケーションリストと同じ、Excel ファイル (XLS ファイル) と CSV ファイル (カンマ区切りのファイル) の 2 種類です。用例は最大で先頭から 1 万件まで抽出できます。



ファイル形式を選択して、[ダウンロード] ボタンをクリックすると、以下のようなダイアログが表示されますので、[プログラムで開く] か、[ファイルを保存する] かを選び、[OK] ボタンをクリックします。



以下は、Excel ファイルでダウンロードし、Excel で開いたときの画面です。



[ファイル情報] のシートには、ダウンロードしたコロケーションの用例の情報が記載されています。

	A
1	ファイル情報
2	NLTバージョン : 1.10
3	見出し語 : 買う (V.00079)
4	コロケーション : 寝るを{買う}
5	用例数 : 413
6	出力日 : 2013-04-01 15:00

■ ローマ字表記で使用しているローマ字一覧 ■

あ	い	う	え	お			
a	i	u	e	o			
か	き	く	け	こ	きゃ	きゅ	きょ
ka	ki	ku	ke	ko	kya	kyu	kyo
さ	し	す	せ	そ	しゃ	しゅ	しょ
sa	shi	su	se	so	sha	shu	sho
た	ち	つ	て	と	ちゃ	ちゅ	ちょ
ta	chi	tsu	te	to	cha	chu	cho
な	に	ぬ	ね	の	にゃ	にゅ	にょ
na	ni	nu	ne	no	nya	nyu	nyo
は	ひ	ふ	へ	ほ	ひゃ	ひゅ	ひょ
ha	hi	fu	he	ho	hya	hyu	hyo
ま	み	む	め	も	みゃ	みゅ	みょ
ma	mi	mu	me	mo	mya	myu	myo
や		ゆ		よ			
ya		yu		yo			
ら	り	る	れ	ろ	りゃ	りゅ	りょ
ra	ri	ru	re	ro	rya	ryu	ryo
わ				を	ん		
wa				wo	n		
が	ぎ	ぐ	げ	ご	ぎゃ	ぎゅ	ぎょ
ga	gi	gu	ge	go	gya	gyu	gyo
ざ	じ	ず	ぜ	ぞ	じゃ	じゅ	じょ
za	ji	zu	ze	zo	ja	ju	jo
だ	ぢ	づ	で	ど	ぢゃ	ぢゅ	ぢょ
da	ji	zu	de	do	ja	ju	jo
ば	び	ぶ	べ	ぼ	びゃ	びゅ	びょ
ba	bi	bu	be	bo	bya	byu	byo
ぱ	ぴ	ぷ	ぺ	ぽ	ぴゃ	ぴゅ	ぴょ
pa	pi	pu	pe	po	pya	pyu	pyo
ふぁ	ふぃ		ふぇ	ふぉ			
fa	fi		fe	fo			

このマニュアルについて

本マニュアルの著作権は、筑波大学と国立国語研究所と Lago 言語研究所が所有します。無断転載・複製を一切禁じます。

Copyright © 2013 University of Tsukuba, National Institute for Japanese Language and Linguistics, Lago Institute of Language. All rights reserved.