

## NINJAL-LWP の類義語比較機能

赤瀬川 史朗 (Lago 言語研究所)  
プラシャント・パルデシ (国立国語研究所)  
今井 新悟 (筑波大学)

### The Function to Compare Near-Synonyms in NINJAL-LWP

Shiro Akasegawa (Lago Institute of Language)  
Prashant Pardeshi (National Institute for Japanese Language and Linguistics)  
Shingo Imai (Tsukuba University)

#### 要旨

NINJAL-LWP は、レキシカルプロファイリング型のコーパス検索ツールである。現在、BCCWJ 向けの NLB (NINJAL-LWP for BCCWJ) と TWC (筑波ウェブコーパス) 向けの NLT (NINJAL-LWP for TWC) の 2 つのバージョンが公開されている。レキシカルプロファイリング型のコーパスツールの特長は、特定の語彙の振る舞いのおおよその全体像を把握できる点にある。なかでも、コロケーションを網羅的にしかも簡単に調査できることは、コンコーダンスにはない優れた特長と言える。今回、この NINJAL-LWP に、内容を対象とした類義語の調査を可能にする 2 語比較機能を実装した。これまでは難しかったコロケーションレベルの類義語の研究に活用されるものと期待される。類義語のペアは、同じ品詞の内容語のほか、形容詞と連体詞(「小さい」と「小さな」)にも対応している。形容詞では、イ形容詞とナ形容詞の比較(「美しい」と「きれいな」)も可能である。また、片方の語に見られる特徴的なコロケーションだけでなく、両方の語に現れるコロケーション(「楽しい気分」と「うれしい気分」)の差異を用例レベルで簡単に比較できる機能も搭載している。本稿では、コロケーション分析におけるレキシカルプロファイリング型のコーパスツールの利点と、NINJAL-LWP の 2 語比較機能の概要を述べた上で、いくつかの比較例を提示する。

#### 1. レキシカルプロファイリングとコロケーション分析

NINJAL-LWP は、レキシカルプロファイリング型のコーパス検索ツール(以下、レキシカルプロファイラー)である。現在、BCCWJ 向けの NLB (NINJAL-LWP for BCCWJ) と TWC (筑波ウェブコーパス) 向けの NLT (NINJAL-LWP for TWC) の 2 つのバージョンが一般公開されている(関連 URL 参照)。

レキシカルプロファイリングは、あらかじめ設定された検索式に基づいて、コーパスから様々なタイプのコロケーションの情報を抽出した結果を、文法パターンごとに整理してユーザに提示するコーパス検索手法である。特定の語彙の文法的振る舞いやコロケーションをマクロ的視点から調査できる点に大きな特長がある。一方、最も一般的なコーパスツールであるコンコーダンスでは、特定の言語事象を対象にしたミクロな視点からの観察が可能である。その意味で、レキシカルプロファイリングとコンコーダンスは、コーパス検索において相互補完的な役割を果たしていると言える。

レキシカルプロファイリングがコロケーション分析に適している理由は3つある。一つ目は、上述の通り、検索式に基づいてコロケーションを網羅的に抽出するため、コンコーダンスに比べて、一般のユーザでも簡単にコロケーション調査ができることがある。コンコーダンスを使って、コロケーションを抽出する場合、1) コンコーダンスの出力をダウンロードする、2) 共起語を集計する、3) 集計した結果を頻度などの統計値を用いて並べ替える、という3つの作業が必要になる。これに対して、レキシカルプロファイラーでは、このような作業は事前に行われるため、ユーザは調査したいパターンを選ぶだけで結果が得られる。図1は、NLBで「冷える」のガ格名詞のコロケーションを調べた画面である。左側のパネルの「…が冷える」をクリックするだけで、右側にそのコロケーションが表示される。

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
○ …が冷える	258	
○ …は冷える	117	
○ …も冷える	29	
○ …の冷える	27	
○ …を冷える	4	
○ …に冷える	88	
○ …へ冷える	2	
○ …で冷える	39	
○ …と冷える	11	
○ …から冷える	20	
○ …まで冷える	8	
○ …より冷える	6	

  

…が冷える 110種類		コーパス全体	
コロケーション	頻度	MI	LD
体が冷える	44	8.63	4.98
身体が冷える	15	9.18	5.50
足が冷える	12	8.40	4.73
【人名】が冷える	8	1.43	-2.22
お腹が冷える	8	9.81	6.07
からだが冷える	7	9.77	6.02
手足が冷える	6	11.21	7.24
下半身が冷える	6	12.09	7.90
空気が冷える	5	8.58	4.87
頭が冷える	4	6.00	2.34
足元が冷える	4	9.60	5.79
腰が冷える	4	7.95	4.25
指先が冷える	4	10.50	6.55

図1 NLBによるコロケーション分析

2つ目の理由は、コンコーダンスでコロケーションを抽出する際に必要になる語数指定がレキシカルプロファイラーでは不要なことである。例えば、自動詞とガ格名詞のコロケーションを調べるときは、ガ格名詞と自動詞との間に付加詞や副詞などの修飾要素が置かれることを考慮に入れる必要がある。通常、そのような語は5語以内に収まることが多いと思われるが、実際の用例ではその範囲を超える可能性もある。不足なく抽出するには、抽出する範囲を広げるとよさそうに思えるが、あまり広げ過ぎると今度は、主述の関係にないガ格名詞と動詞を拾ってしまう場合が出てくる。

NINJAL-LWPでは、係り受け解析の結果を利用するため、上記のような中心語と共起語の語数の問題に悩まされることはない。例えば、動詞「冷える」と共起するガ格名詞では、以下のように中心語と共起語がかなり離れた用例でも抽出される(下線は筆者)。

(1) マグマがマグマ溜りの中でゆっくり冷えるにつれて次々といろいろな鉱物ができ、マグマ溜りの下に沈んでいく。(BCCWJ教科書サブコーパス, 高等学校地学I)

3つ目の理由は、レキシカルプロファイラーでは、頻度以外の統計値があらかじめ用意されているため、それらの統計値を用いて並べ替えができる点である。NINJAL-LWPでは、単純頻度以外に、MIとlogDiceという統計値が用意されている。また、NLBでは、サブコ

コーパス間の使用分布を比較するために、PMW (100万語あたりの調整頻度) が表示される。コンコーダンスの場合、頻度以外の統計値が用意されていないときは、ユーザ側で計算する必要が出てくるが、その際、共起頻度のほかに、中心語の頻度、共起語の頻度、コーパスの総語数などが分かっているなければならない。

このように、コンコーダンスを利用してコロケーションを分析するには、かなりの労力が要求される。また、3番目に挙げた統計値の算出では、種々の頻度データを用意する必要があるが、これらは一般のユーザが簡単に準備できるものではない。レキシカルプロファイラーでは、このような一連の作業を事前に行った上で、コロケーションのリストを提示するため、ユーザは一番肝心なコロケーションの分析作業にすぐ取りかかることができる。

## 2. NINJAL-LWP の類義語の比較機能

まず、従来の単一の見出し語のプロファイリング画面を 2 つ開いて、類義語の比較を行うとどのようになるかをシミュレートしてみたい。図 2 は、NLB で「冷える」と「冷める」の見出し語ウィンドウをそれぞれ開いて、ガ格名詞のコロケーションのリストを比較したものである。

コロケーション	コーパス全体		
	頻度	MI	LD
体が冷える	44	8.63	4.98
身体が冷える	15	9.18	5.50
足が冷える	12	8.40	4.73
【人名】が冷える	8	1.43	-2.22
お腹が冷える	8	9.81	6.07
からだ冷える	7	9.77	6.02
手足が冷える	6	11.21	7.24
下半身が冷える	6	12.09	7.90
空気が冷える	5	8.58	4.87
頭が冷える	4	6.00	2.34
足元が冷える	4	9.60	5.79
腰が冷える	4	7.95	4.25
指先が冷える	4	10.50	6.55
全身が冷える	4	9.09	5.32
先が冷える	4	5.60	1.95
マグマが冷える	4	12.46	7.90
エンジンが冷える	4	8.54	4.81

  

コロケーション	コーパス全体		
	頻度	MI	LD
熱が冷める	31	11.86	7.59
ほとぼりが冷める	17	17.92	11.34
気持ちが冷める	17	9.03	4.79
【数字】が冷める	6	0.68	-3.54
お湯が冷める	6	11.46	7.06
酔いが冷める	5	11.95	7.44
目が冷める	4	5.32	1.09
料理が冷める	3	7.38	3.13
愛情が冷める	3	9.80	5.46
愛が冷める	3	8.13	3.87
心が冷める	3	5.62	1.39
紅茶が冷める	2	10.41	5.93
窯が冷める	2	10.97	6.39
空気が冷める	2	7.84	3.57
熱狂が冷める	2	11.76	6.96
熱意が冷める	2	11.14	6.52
怒りが冷める	2	8.43	4.14

図 2 単一の見出し語ウィンドウによる類義語の比較

表 1 は、図 2 をもとに、ガ格名詞を手作業でグループごとにまとめたものである。このような比較によって、2つの動詞のガ格名詞にどのようなものがあるか、おおよその判断はつけられることが分かる。だが、共起語の数が増えてくると、手作業では、見落としや見誤りが出てきたり、客観的な基準で両者を比較することが困難になったりすることが予想される。

表 1 「冷える」と「冷める」のガ格名詞

「冷える」のガ格名詞	「冷める」のガ格名詞
体/全身/お腹/指先/手足/足/足元/腰/下半身が冷える	熱/熱狂/熱意/ほとぼりが冷める
空気が冷える	酔いが冷める
マグマが冷える	愛/愛情が冷める
エンジンが冷える	お湯/紅茶/料理が冷める

今回実装した 2 語比較のプロファイリングの画面では、図 3 のように、2 つの動詞のコロケーションが同じパネルの左右に表示され、一方に顕著なコロケーションは段階的な濃淡でハイライト表示される。上のほうにくるコロケーションほど「冷える」に顕著なもの（「下半身が冷える」、「マグマが冷える」、「手足が冷える」など）、逆に、下のほうにくるコロケーションほど「冷める」に顕著なもの（「ほとぼりが冷める」、「熱が冷める」、「酔いが冷める」など）である。両方の動詞でともに現れているガ格名詞にどのようなものがあるかも、一目で確認できる（「体が冷える/冷める」、「気持ち冷える/冷める」など）。

...が冷える — ...が冷める									
冷える				冷める				LD差	
コロケーション	頻度	MI	LD	コロケーション	頻度	MI	LD		
下半身が冷える	6	12.09	7.9					7.9	
マグマが冷える	4	12.46	7.9					7.9	
手足が冷える	6	11.21	7.24					7.24	
指先が冷える	4	10.5	6.55					6.55	
お腹が冷える	8	9.81	6.07					6.07	
からだ冷える	7	9.77	6.02					6.02	
足元が冷える	4	9.6	5.79					5.79	
身体が冷える	15	9.18	5.5					5.5	
全身が冷える	4	9.09	5.32					5.32	
エンジンが冷える	4	8.54	4.81					4.81	
足が冷える	12	8.4	4.73					4.73	
体が冷える	44	8.63	4.98	体が冷める	2	4.75	0.52	4.46	
腰が冷える	4	7.95	4.25					4.25	
頭が冷える	4	6	2.34					2.34	
先が冷える	4	5.6	1.95					1.95	
空気が冷える	5	8.58	4.87	空気が冷める	2	7.84	3.57	1.3	
				目が冷める	4	5.32	1.09	-1.09	
気持ちが冷える	1	4.36	0.7	気持ちが冷める	17	9.03	4.79	-4.09	
				お湯が冷める	6	11.46	7.06	-7.06	
				酔いが冷める	5	11.95	7.44	-7.44	
				熱が冷める	31	11.86	7.59	-7.59	
				ほとぼりが冷める	17	17.92	11.34	-11.34	

図 3 類義語の比較 (2 語比較機能)

対照されるコロケーションの顕著さを測る指標として用いているのは、 $\logDice^1$ の差（画面上では「LD差」と表示）であり、Sketch EngineのSketch-Diffと同じ手法を用いている。例えば、「下半身が冷える」では、「下半身が冷める」というコロケーションが1件もないため、 $\logDice$ 差は  $7.9-0=7.9$  となる。また、「ほとぼりが冷める」の場合は、「ほとぼりが冷える」というコロケーションがないため、 $\logDice$ 差は  $0-11.34=-11.34$  となる。「気持ちが冷える/冷める」の場合は、 $\logDice$ 差は  $0.7-4.79=-4.09$  となり、「冷める」のコロケーションとして優勢だということになる。つまり、この例では、プラスの値の絶対値が大きくなればなるほど、「冷える」のコロケーションとして顕著であり、マイナスの値の絶対値が大きくなればなるほど、「冷める」のコロケーションとして顕著だということになる。

コロケーション比較の尺度においては、NINJAL-LWPは、Sketch-Diffの手法をそのまま採用しているが、インターフェースの面では、Sketch-Diffよりも自由度が高く、分析能力が強化されている。図4は、Sketch EngineのSketch-Diffの画面である（jpTenTen11で「冷える」の「冷める」のSketch-Diffをall in one blockモードで表示）。1列目が共起語、2列目が「冷める」との共起頻度、3列目が「冷える」との共起頻度、4列目が「冷める」の $\logDice$ 、5列目が「冷える」の $\logDice$ で、 $\logDice$ の差でソートされている（ $\logDice$ 差の列自体は表示されない）。初期画面では、 $\logDice$ 差の大きいものがデフォルトで最大25まで表示されている。さらに多くの共起語の表示したい場合は、いったん設定画面に戻り、指定し直して再度実行する必要がある。Sketch-Diffの画面はスタティックな画面なので、リストを並べ替えたり、同じ画面にとどまったまま、表示する共起語の数を増減させたりすることはできない。

nounが	26,049	29,117	-1.9	-1.8
足先	0	361	--	7.6
手足	0	938	--	7.2
肝	0	583	--	6.4
マグマ	0	109	--	6.0
身体	0	1,646	--	5.9
溶岩	0	94	--	5.6
背筋	0	210	--	5.6
朝晩	0	80	--	5.6
足腰	0	89	--	5.5
足元	0	548	--	5.5
手先	0	101	--	5.4
末端	0	100	--	5.4
内臓	0	151	--	5.2
御腹	0	947	--	5.0
体	0	6,339	--	5.0
指先	0	261	--	5.0
情熱	209	0	4.8	--
熱気	156	0	5.6	--
余韻	256	0	5.7	--
ホトボリ	44	0	5.8	--
熱	4,772	75	7.4	1.4
興奮	1,243	0	6.3	--
興	158	0	6.3	--
酔い	780	0	7.8	--
ほとぼり	2,658	0	11.3	--

図4 Sketch-Diff

これに対して、NINJAL-LWPの2語比較のコロケーションリストには、並べ替えやフィルターの機能が備わっているので、リストを操作しながら、さまざまな角度から分析することができる。図5は、コロケーションリストの簡易フィルタ機能である。頻度と $\logDice$ 差と出現位置の3つを自由に組み合わせて、表示する共起語の範囲を調整できる。例えば、出現位置を「両方の語」にすると、2つの動詞で共通する共起語のみが表示される。図6は、「冷える」と「冷める」の両方で共通して現れるガ格名詞である。

頻度	すべて	2以上	5以上	10以上
LD差	すべて	±2以上	±3以上	±5以上
出現位置	すべて	両方の語	左の語のみ	右の語のみ

図5 コロケーションパネルの簡易フィルタ機能

<sup>1</sup>  $\logDice$ の算出法については、Statistics used in the Sketch Engine

(<http://www.sketchengine.co.uk/documentation/raw-attachment/wiki/SkE/DocsIndex/ske-stat.pdf>)を参照してほしい。

...が冷える — ...が冷める								
冷える				冷める				LD差
コロケーション	頻度	MI	LD	コロケーション	頻度	MI	LD	
体が冷える	44	8.63	4.98	体が冷める	2	4.75	0.52	4.46
コーヒーが冷える	3	8.2	4.47	コーヒーが冷める	1	7.19	2.91	1.56
空気が冷える	5	8.58	4.87	空気が冷める	2	7.84	3.57	1.3
水が冷える	2	4.78	1.13	水が冷める	1	4.36	0.13	1
肉が冷える	1	5.98	2.28	肉が冷める	1	6.56	2.3	-0.02
酒が冷える	1	5.27	1.59	酒が冷める	1	5.85	1.61	-0.02
お茶が冷える	1	7.02	3.26	お茶が冷める	1	7.59	3.29	-0.03
愛が冷える	2	6.97	3.27	愛が冷める	3	8.13	3.87	-0.6
料理が冷える	1	5.21	1.54	料理が冷める	3	7.38	3.13	-1.59
愛情が冷える	1	7.64	3.83	愛情が冷める	3	9.8	5.46	-1.63
気持ちが冷える	1	4.36	0.7	気持ちが冷める	17	9.03	4.79	-4.09

頻度:      
LD差:      
出現位置:

図 6 両方の語に現れるコロケーションの表示

### 3. 比較例

NINJAL-LWPの2語比較機能では、同一品詞の内容語(名詞、動詞、形容詞、副詞、連体詞)の比較が可能である。形容詞では、イ形容詞とナ形容詞の比較(「美しい」と「きれいな」)も可能である。また、形容詞と連体詞の比較(「小さい」と「小さな」)も可能である。ここでは、名詞の比較と形容詞の比較を取り上げながら、2節で触れなかった機能を中心に紹介したい。

#### 3.1 「図書」と「書籍」

最初に、「図書」と「書籍」の使い方の違いを文法パターンの割合に注目してNLBで調べてみることにする。図7は、「図書」と「書籍」の文法パターンの割合を示している。ハイライト表示になっている項目は、それぞれの語に顕著なパターンを示している。「図書」の場合、接尾辞が続くパターンや、前項または後項となって複合名詞を形成する割合が、「書籍」の場合と比べ、きわだっていることが分かる。図書/書籍+接尾辞をクリックして、実際にどのような表現があるかを示したのが図8である。「図書」+接尾辞では、「図書券」と「図書室」が全体の85%以上を占めることが確認できる。

図書			書籍		
パターン	頻度	比率	パターン	頻度	比率
図書+名詞・接尾	456	36.1%	書籍+名詞・接尾	67	8.4%
図書+名詞	422	33.4%	書籍+名詞	117	14.7%
名詞+図書	412	32.6%	名詞+書籍	101	12.7%
図書+助詞	375	29.7%	書籍+助詞	488	61.3%
図書+の+名詞	110	8.7%	書籍+の+名詞	84	10.6%
名詞+の+図書	88	7.0%	名詞+の+書籍	81	10.2%
並立	71	5.6%	並立	110	13.8%
(未分類)	63	5.0%	(未分類)	51	6.4%
図書を...	62	4.9%	書籍を+動詞	90	11.3%
動詞過去+図書	40	3.2%	動詞過去+書籍	43	5.4%
動詞基本形+図書	40	3.2%	動詞基本形+書籍	38	4.8%
図書に...	24	1.9%	書籍に+動詞	25	3.1%
図書が+動詞	24	1.9%	書籍が+動詞	51	6.4%

すべて表示

図 7 文法パターンの割合の比較

「図書」+接尾辞では、「図書券」と「図書室」が全体の85%以上を占めることが確認できる。

図書+名詞・接尾 — 書籍+名詞・接尾								
図書				書籍				LD差
コロケーション	頻度	MI	LD	コロケーション	頻度	MI	LD	
図書券	73	12.04	9.07					9.07
図書室	321	12.41	9.54	書籍室	1	6.85	1.25	8.29
図書隊	9	8.41	5.48					5.48
図書係	6	8.26	5.3					5.3
図書碑	1	7.76	4.42					4.42
図書賞	3	7.33	4.36					4.36
図書展	1	6.42	3.38					3.38
書局	3	5.86	2.99					2.99
図書編	1	5.88	2.9					2.9
図書比	1	5.3	2.37					2.37
図書課	2	5.03	2.16					2.16
図書等	14	5.21	2.38	書籍等	5	6.49	0.9	1.48
図書学	1	4.31	1.43					1.43
図書社	1	4.13	1.25					1.25

図 8 「図書/書籍+接尾辞」の比較

### 3.2 「楽しい」と「うれしい」

次に、形容詞の「楽しい」と形容詞の「うれしい」に続く名詞を比較しながら、2つの形容詞の違いをNLTで調べてみることにする。まず頻度200以上<sup>2</sup>、logDice差を±3以上に絞り込むと、図9のような結果になる。

2つの形容詞に後続する名詞をまとめると、表2のようになる。両者の比較からそれぞれの形容詞の意味を考えると、「楽しい」は、ある時間、ある場所で活動することで満足した気分になることを表す。それに対して、「うれしい」は、外からの働きかけによって、期待したことが実現し、満足した気分になることを表す。さらに、「うれしい」では、

楽しい+名詞 26 嬉しい+名詞 15								
楽しい				嬉しい				LD差
コロケーション	頻度	MI	LD	コロケーション	頻度	MI	LD	
楽しい雰囲気	786	7.77	7.62					7.62
楽しい会話	430	7.32	6.99					6.99
楽しい時間	3863	6.66	7.15	嬉しい時間	40	1.14	0.6	6.55
楽しい旅	366	7.36	6.89	嬉しい旅	3	1.5	0.42	6.47
楽しい人生	449	5.99	6.15					6.15
楽しい授業	377	6.03	6.11					6.11
楽しい旅行	320	6.09	6.08					6.08
楽しい遊び	212	6.45	6.04					6.04
楽しいゲーム	249	5.6	5.63					5.63
楽しい食事	215	5.32	5.38					5.38
楽しい生活	583	4.71	5.14					5.14
楽しい作業	219	4.67	4.9					4.9
楽しい学校	329	4.45	4.82					4.82
楽しいひと時	1245	12.17	9.46	嬉しいひと時	25	7.61	4.74	4.72
楽しいイベント	579	7.25	7.13	嬉しいイベント	20	3.47	2.6	4.53
楽しい場所	246	4.19	4.53					4.53
楽しい仕事	359	4.05	4.48					4.48
楽しい思い出	1452	9.86	9.11	嬉しい思い出	47	5.99	4.71	4.4
楽しい日々	432	6.79	6.68	嬉しい日々	18	3.27	2.42	4.26
楽しい作品	252	4.67	4.94	嬉しい作品	14	1.58	0.93	4.01
楽しい毎日	437	8.4	7.48	嬉しい毎日	21	5.1	3.7	3.78
楽しい企画	231	6.24	5.99	嬉しい企画	18	3.63	2.68	3.31
楽しいサービス	29	0.89	1.26	嬉しいサービス	248	5.06	4.46	-3.2
楽しい言葉	35	0.91	1.31	嬉しい言葉	355	5.32	4.74	-3.43
楽しいニュース	13	2.63	2.12	嬉しいニュース	266	8.06	6.95	-4.83
楽しい限り	12	0.86	1	嬉しい限り	1864	9.21	8.49	-7.49
				嬉しい誤算	227	13.67	8.19	-8.19
				嬉しい悲鳴	315	11.57	8.45	-8.45

図 9 「楽しい」と「うれしい」の後続名詞の比較

<sup>2</sup> フィルターダイアログを利用すると、任意の頻度、logDice 差を設定することが可能。

意外性を表す語が後続することから分かるように、期待以上の状況が生まれることも含意している。「うれしい悲鳴」という慣用表現にもそのような意外性が含まれている。

表 2 「楽しい」と「うれしい」に後続する名詞

楽しい
時間を表す語…時間、ひと時、日々、毎日
活動を表す語…旅、レッスン、授業、食事
娯楽を表す語…会話、おしゃべり、遊び、ゲーム
その他…雰囲気、人生、生活、イベント
うれしい
知らせを表す語…ニュース、知らせ、報告
意外性を表す語…誤算、驚き、サプライズ
その他…感想、特典、配慮、再会
慣用表現…悲鳴

では、両方の形容詞に共通して出現する後続名詞を調べてみよう。少なくとも一方の共起頻度 200 以上で、出現位置を「両方の語」にした結果が図 10 である。

楽しい+名詞 28				嬉しい+名詞 28				LD差
コロケーション	頻度	MI	LD	コロケーション	頻度	MI	LD	
楽しい時間	3863	6.66	7.15	嬉しい時間	40	1.14	0.6	6.55
楽しい旅	366	7.36	6.89	嬉しい旅	3	1.5	0.42	6.47
楽しいひと時	1245	12.17	9.46	嬉しいひと時	25	7.61	4.74	4.72
楽しいイベント	579	7.25	7.13	嬉しいイベント	20	3.47	2.6	4.53
楽しい思い出	1452	9.86	9.11	嬉しい思い出	47	5.99	4.71	4.4
楽しい日々	432	6.79	6.68	嬉しい日々	18	3.27	2.42	4.26
楽しい作品	252	4.67	4.94	嬉しい作品	14	1.58	0.93	4.01
楽しい毎日	437	8.4	7.48	嬉しい毎日	21	5.1	3.7	3.78
楽しい企画	231	6.24	5.99	嬉しい企画	18	3.63	2.68	3.31
楽しいお話	317	6.46	6.31	嬉しいお話	34	4.31	3.42	2.89
楽しい経験	248	4.55	4.84	嬉しい経験	37	2.88	2.24	2.6
楽しい体験	217	5.33	5.39	嬉しい体験	43	4.07	3.3	2.09
楽しい気分	537	7.63	7.31	嬉しい気分	142	6.79	5.79	1.52
楽しい話	368	4.09	4.52	嬉しい話	157	3.93	3.37	1.15
楽しいもの	3364	4.88	5.43	嬉しいもの	1575	4.86	4.35	1.08
楽しいはず	394	5.27	5.55	嬉しいはず	185	5.26	4.61	0.94
楽しいとき	721	3.4	3.93	嬉しいとき	422	3.7	3.18	0.75
楽しいところ	374	3.46	3.94	嬉しいところ	266	4.04	3.5	0.44
楽しいん	1231	4.35	4.88	嬉しいん	933	5.03	4.5	0.38
楽しいこと	5552	3.67	4.25	嬉しいこと	6179	4.9	4.4	-0.15
楽しいの	2069	2.71	3.28	嬉しいの	2363	3.97	3.47	-0.19
楽しいよう	199	0.17	0.73	嬉しいよう	401	2.26	1.75	-1.02
楽しい気持ち	441	5.2	5.52	嬉しい気持ち	946	7.38	6.75	-1.23
楽しい出来事	88	5.49	4.93	嬉しい出来事	213	7.84	6.69	-1.76
楽しいサービス	29	0.89	1.26	嬉しいサービス	248	5.06	4.46	-3.2
楽しい言葉	35	0.91	1.31	嬉しい言葉	355	5.32	4.74	-3.43
楽しいニュース	13	2.63	2.12	嬉しいニュース	266	8.06	6.95	-4.83
楽しい限り	12	0.86	1	嬉しい限り	1864	9.21	8.49	-7.49

図 10 「楽しい」と「うれしい」の両方に後続する名詞



このリストのなかから、比較的両方の頻度の高い「楽しい気分」と「うれしい気分」を例にとって、2つの表現にどのような意味の違いがあるかを実際の用例を見ながら調べてみたい。コロケーションリストのコロケーションの列をクリックすると、文法パターンパネルが閉じて、代わりにコロケーションリストの右に用例パネルが現れて、そのコロケーションの用例が表示される。「楽しい気分」をクリックしたときの画面が図11である。

The screenshot shows a web interface for searching Japanese corpora. On the left, there is a table of collocations for '楽しい' (fun) and '嬉しい' (happy). The table has columns for '楽しい+名詞' and '嬉しい+名詞', each with sub-columns for 'コロケーション', '頻度', 'MI', and 'LD'. Below the table are filters for frequency and LD difference. On the right, a panel titled '楽しい気分 全537件' displays a list of example sentences with their source URLs.

楽しい+名詞 28				嬉しい+名詞 28			
コロケーション	頻度	MI	LD	コロケーション	頻度	MI	LD
楽しい時間	3863	6.66	7.15	嬉しい時間	40	1.14	0.6
楽しい旅	366	7.36	6.89	嬉しい旅	3	1.5	0.42
楽しいひと時	1245	12.17	9.46	嬉しいひと時	25	7.61	4.74
楽しいイベント	579	7.25	7.13	嬉しいイベント	20	3.47	2.6
楽しい思い出	1452	9.86	9.11	嬉しい思い出	47	5.99	4.71
楽しい日々	432	6.79	6.68	嬉しい日々	18	3.27	4.21
楽しい作品	252	4.67	4.94	嬉しい作品	14	1.58	0.93
楽しい毎日	437	8.4	7.48	嬉しい毎日	21	5.1	3.7
楽しい企画	231	6.24	5.99	嬉しい企画	18	3.63	2.68
楽しいお話	317	6.46	6.31	嬉しいお話	34	4.31	3.42
楽しい経験	248	4.55	4.84	嬉しい経験	37	2.88	2.24
楽しい体験	217	5.33	5.39	嬉しい体験	43	4.07	3.3
楽しい気分	537	7.63	7.31	嬉しい気分	142	6.79	5.79
楽しい話	368	4.09	4.52	嬉しい話	157	3.93	3.37
楽しいもの	3364	4.88	5.43	嬉しいもの	1575	4.86	4.35
楽しいはず	394	5.27	5.55	嬉しいはず	185	5.26	4.61
楽しいとき	721	3.4	3.93	嬉しいとき	422	3.7	3.18
楽しいところ	374	3.46	3.94	嬉しいところ	266	4.04	3.5
楽しいん	1231	4.35	4.88	嬉しいん	933	5.03	4.5
楽しいこと	5552	3.67	4.25	嬉しいこと	6179	4.9	4.4
楽しいの	2069	2.71	3.28	嬉しいの	2363	3.97	3.47
楽しいよう	199	0.17	0.73	嬉しいよう	401	2.26	1.75
楽しい気持ち	441	5.2	5.52	嬉しい気持ち	946	7.38	6.75
楽しい出来事	88	5.49	4.93	嬉しい出来事	213	7.84	6.69
楽しいサービ	29	0.89	1.26	嬉しいサービ	248	5.06	4.46
楽しい言葉	35	0.91	1.31	嬉しい言葉	355	5.32	4.74
楽しいニュース	13	2.63	2.12	嬉しいニュース	266	8.06	6.95
楽しい限り	12	0.86	1	嬉しい限り	1864	9.21	8.49

頻度: すべて, 2以上, 5以上, 10以上  
LD差: すべて, ±2以上, ±3以上, ±5以上  
出現位置: すべて, 両方の語, 左の語のみ, 右の語のみ

楽しい気分 全537件

楽しい気分が台無しになります。  
(気分屋で短気な夫(長文です) - 夫婦・家族・教えて! goo)

楽しい気分が台無しになります。  
(気分屋で短気な夫(長文です) - 質問・相談ならMSN相談箱)

明るく、楽しい気分で過ごした。  
(足立区 相談の受け方)

楽しい気分になった押絵でした。  
(アンケートの答え)

院長が明るく楽しい気分になる。  
(患者さんの声 :: 交通事故や腫瘍治療のバシモ康大井登病院::)

楽しい気分になっちゃいますよね!  
(One's Want!! 新しい家族持ってま〜す。クリスマススペシャル企画!)

リラックスして楽しい気分に乗る。  
(Marinedoor.com)

だから今もとても楽しい気分だよ。  
(RPG Data Library : アンジェリーク会話集2)

すっかり楽しい気分になったらしい。  
(-Baltimore滞在記-)

楽しい気分になせられてしまうのだ。  
(遺囑うどん)

とっても楽しい気分になれる曲です。  
(アクロ・スタント技の辞典)

楽しい気分であることはまれである。  
(障害児のきょうだい達の心の健康〜きょうだい達をどう健やかに育てるか〜)

とっぴり楽しい気分になりました。  
(マッチステック・メン評価)

客を楽しい気分になせ、笑いを誘う。  
(その3ユーモアトークの考え方)

楽しい気分になりたいと親子はいう。  
(Page 1 / 6)

図 11 用例パネルの表示(「楽しい気分」)

同様に、「うれしい気分」をクリックして、順にその用例を確認する。以下に、「楽しい気分」と「うれしい気分」の用例を一つずつ挙げる(下線は筆者)。

- (2) 賑やかな能が多く、観客は楽しい気分で家路につく。(世阿弥の生涯、<http://kajipon.sakura.ne.jp/kt/haka-topic14.html>)
- (3) 個人懇談では先生が一生懸命褒めてくださり、ちょっと嬉しい気分です。  
(SkipMamaClass すきっぷママクラス PDD 子育て支援セミナー、<http://www.jsnhc.org/smc/navis/navi2/syuuryouseikansou.html>)

(2)では、能の鑑賞(活動)に満足していることを表している。一方の(3)は、個人面談で褒めてもらえた(外からの働きかけ)ので満足していることを表している。先ほどの「楽しい」と「うれしい」の後続名詞の分析の結果と一致していることが確認できる。

#### 4. まとめ

本稿では、コロケーション分析におけるレキシカルプロファイリングの利点を挙げ、NINJAL-LWP に実装された 2 語比較機能の概要を述べ、実際の比較例を紹介した。2 語比較機能は、類義語の分析のみならず、反義語の分析にも応用できる。例えば、「強い」と「弱い」のガ格名詞を比較すると、片方の語にのみ共起する名詞の特徴を詳細に調べることができる（「イメージが強い」というが「イメージが弱い」とはあまり言わない、など）。今後、NINJAL-LWP がコーパス基盤の類義語の研究に広く活用されることを期待したい。

#### 文 献

- Curran, James (2004) *From Distributional to Semantic Similarity*. PhD thesis, Institute for Communicating and Collaborative Systems, School of Informatics, University of Edinburgh.
- Gatto, Maristella (2014) *Web As Corpus: Theory and Practice*. A&C Black.
- 今井 新悟、赤瀬川 史朗、プラシャント パルデシ (2013) 「筑波ウェブコーパス検索ツール NLT の開発」 第 3 回コーパス日本語学ワークショップ予稿集、pp.199-206.  
([http://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop\\_no3\\_papers/JCLWorkshop\\_No3\\_web.pdf](http://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no3_papers/JCLWorkshop_No3_web.pdf) よりダウンロード可能)
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrz and David Tugwell (2004) The Sketch Engine. In *Proceeding of EURALEX 2004*, pp.105-116.
- Murphy, M. Lynne (2010) *Lexical Meaning*. Cambridge University Press.
- プラシャント・パルデシ、赤瀬川史朗 (2011) 「BCCWJ を活用した基本動詞ハンドブック作成—コーパスブラウジングシステム NINJAL-LWP の特徴と機能—」 『現代日本語書き言葉均衡コーパス』 完成記念講演会予稿集、pp.205-216.
- プラシャント・パルデシ、赤瀬川史朗 (2012) 「レキシカルプロファイリング手法を用いた BCCWJ 検索ツール NINJAL-LWP とその研究事例」 日本言語学会第 144 回大会予稿集、pp.364-369.
- Rychlý, Pavel (2008) A Lexicographer-Friendly Association Score. In *Proceedings of Recent Advances In Slavonic Natural Language Processing, RASLAN 2008*, pp.6-9.

#### 関連 URL

- NLB (NINJAL-LWP for BCCWJ) <http://nlb.ninjal.ac.jp/>  
NLT (NINJAL-LWP for TWC) <http://corpus.tsukuba.ac.jp/>  
The Sketch Engine <https://the.sketchengine.co.uk/>