

# コーパスに基づいたコロケーション分析

－「素敵」「立派」「素晴らしい」を例に－

趙 聖花

ZHAO Shenghua

中国大連理工大学ソフトウェア学院

劉 玉琴

LIU Yuqin

中国大連理工大学ソフトウェア学院

**要旨：**国語研と Lago 言語研究所が共同開発した『NLB』コーパスを用いて類義語である「素敵」「立派」「素晴らしい」のコロケーションの情報を抽出して異同について調査を行った。本研究では、調査結果を基に学習者にとって必要と思われる「素敵」「立派」「素晴らしい」のコロケーション情報についての試案を述べる。

**キーワード：**コーパス、コロケーション、素敵、立派、素晴らしい

## 1. はじめに

近年、コーパスに基づいた日本語の研究が注目されている。その中にコロケーションの重要性が多く指摘され(田野村 2009、大曾 2003、中溝ら 2011)、コロケーションの情報をどのように日本語学習者に提示すべきかについて述べている。しかし、「素敵」「立派」「素晴らしい」に関する先行研究は少ない。辞書にそれぞれの意味は記述されているものの、これらはどのぐらい類似しているかについての説明は不十分である。どれも「すぐれている」「すばらしい」「いい」様子を表すと記述されていて、その用法の異同の説明が不十分である。したがって、こうした類義語については、意味的な異同の説明ばかりではなく、コロケーション情報をコーパスから抽出し、使用実態に基づいて使い方の異同を考察するものであり、典型的、または特徴的といえる共起情報を収集する。実際にどのように、どのような語と共起することが多いか、その頻度、MI スコア、LD 係数でそれぞれのコロケーションの特徴を見る必要がある。

このようなことから、本稿では『NLB』コーパスを用いて学習者がこのような類義語の異同の習得に一助となることを目的とする。

## 2. 先行研究と問題点

2.1. 『大辞林』(第二版)での記述(三省堂 <http://www.sanseido.net/>)

すてき:

- (1) 心を引き付けられるさま。すばらしいさま。

- (2) 程度のはなはだしいさま。並はずれたさま。

りっぱ:

- (1) 非常に素晴らしいさま。非常にすぐれているさま。  
(2) 堂々としているさま。  
(3) 非難する点のないさま。十分であるさま。

すばらしい:

- (1) 思わず感嘆するようなさまを表す。  
(ア) (客観的評価として) この上なくすぐれている。際立って立派だ。  
「一・いアイデア」  
(イ) (主観的評価として) きわめて好ましい。心が満たされる。  
「一・い日曜日」「一・いニュース」  
(2) 程度がはなはだしいさまをいう。  
(ア) 現代語では、多く好ましい状態について用いられる。驚くほどだ。  
(イ) 「一・く広い庭園」「一・く青い空」  
(ウ) 近世江戸語では、多く望ましくないさまをいうのに用いられる。ひどい。

## 2.2 『新明解国語辞典』(第二版)での記述

すてき:心をひきつけられるほどすぐれている様子。(pp. 587)

りっぱ:特に劣っている所が見つからないほど、出来上がりが見事な様子。「一[=見事]にやってのける。一[=十分]に生活している。どこへ出しても一な[=恥ずか

しくない]青年だ。－な[＝すばらしい]贈物。これだけでできれば－なものだ。[＝尊敬に値する]。－[＝崇高]な人格。－[＝偉大]な業績。－な[＝だれからも後ろ指を指されない]処置。－[＝公明正大]に戦う。すっかり－におなりだね[＝人間的に成長し、また前と違って社会的地位を得るようになったことへの感嘆の言葉]。(pp. 1164)

すばらしい：

- (1) 思わず感嘆するほどだ(いい)。
- (2) 程度がはなはなしく、驚くべきほどだ。ものすごい。(pp. 589)
- (3)

### 2.3 『外国人のための基本語用例辞典』での記述

すてき：とてもよい。女の人が多く使う。  
/→すばらしい。(pp. 495)

りっぱ：

- (1) 美しく堂々としているようす。
- (2) すぐれている。みごとだ。
- (3) 十分な様子。(pp. 1095)

すばらしい：

- (1) たいへんりっぱだ。非常によい。
- (2) 「すばらしく」の形で副詞的に用い、「たいへん」をもっと強めた気持ちで使う。びっくりするほど。

☆「すばらしい・すばらしく」はよい場合とか、ほめる場合にしか使わない。(pp. 497)

上記の辞典での記述を見るとどれも「すぐれている」或いは「すばらしい」「いい」様子だと記述されていて、なかなか使い分けできないのである。

### 2.4 『現代形容詞用法辞典』での記述

すてき：

心がひきつけられる魅力のある様子を表す。プラスイメージの語。日常会話中心に用いられる。また会話では女性が好んで用いたり、女性に関する事柄について用いられたりする傾向にある。

「すてき」は「すばらしい」に似ているが、

「すばらしい」のような無条件の賞賛の暗示はなく、対象にやや距離を置いて外から眺める表現になっている点が異なる。まったく同じ文脈で「すてき」と「すばらしい」が用いられると、次のようなニュアンスの違いを生ずる。

- すてきな女性に出会った。/(服装・センスなどが魅力的だ)
- すばらしい女性に出会った。/(外面・内面が賞賛すべき様子だ)

また、「すてき」の表す対象の魅力は、都会的なしゃれた雰囲気をもっている暗示があり、極端なもの、田舎じみたものについては用いられないことが多い。(pp. 308-309)

りっぱ：

- (1) 非常に優れている様子を表す。プラスイメージの語。

「りっぱ」は対象が非常に優れている様子を表す語であるが、その対象はある程度の大きさが必要で、威厳・貫禄・高価・豪壮などの暗示があり、小さいものについてはあまり用いられない。

- ? 立派なチワワ(犬)を飼っている。
- 隣の家は立派な秋田犬を飼っている。
- ? 彼女は立派な宝石を買った。

「りっぱ」は対象が優れていることについて、かなり客観的なニュアンスがあり、対象についての感動などは原則として暗示しない。この点で「すばらしい」と異なる。

- × なんて立派な眺めなんだ。
- なんてすばらしい(見事な)眺めなんだ。

- (2) 完全である様子を表す。プラスマイナスのイメージはない。(pp. 595-596)

すばらしい：

非常に優れていて感嘆すべき様子を表す。プラスイメージの語。「すばらしい」は、対象が非常に優れているので無条件に感嘆している様

子のうかがえる語であって、感動そのものを表す語といってもさしつかえなく、客観性はあまりない表現になっている。(pp. 311)

『現代形容詞用法辞典』での意味記述は同じ文脈での場合、例を挙げてニュアンスの違いを説明している点では意義がある。特に学習者にとって非常に役立つのである。しかし、取り上げた例文があまりにも少なくてやはり使い分けが難しいのである。

本研究では、『NLB』コーパスを用いて類義語である「素敵」「立派」「素晴らしい」のコロケーション情報を抽出してその異同について調査を行う。

### 3. コーパス調査

#### 3.1. 使用するコーパス

本研究で使用するコーパスは NINJAL-LWP for BCCWJ (以下、NLB) は、国立国語研究所 (以下、国語研) が構築した『現代日本語書き言葉均衡コーパス』(Balanced Corpus of Contemporary Written Japanese: BCCWJ) を検索するために、国語研と Lago 言語研究所が共同開発したオンライン検索システムである。NLB ver. 1.20 では、BCCWJ の DVD 版公開データ (2011) の文字ベース XML の可変長データを使用している。NLB 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、BCCWJ のデータにアノテーションを付与した上で解析している。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ある。形態素解析 MeCab 0.99 + IPA 辞書 2.7.0 で、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る。係り受け解析 CaboCha 0.64 で、コーパス総語数は 104,805,763 である。

#### 3.2. コロケーション研究に用いられる統計方法

##### 3.2.1. 共起頻度

頻度とはそもそもすべて特定の母集団の中で決定されるものである。たとえば漠然と the の頻度を言うことはできない。1 億語のコーパスの中での the の頻度、100 万語のコーパスの中での the の頻度というように、すべての頻度は母集団の全体に占める比率として解釈さ

れるのである。このため共起頻度・中心語頻度・共起語頻度と同時に、コーパス総語数を見る必要が生じるのである。(石川慎一郎 2008)

##### 3.2.2. MI スコア

相互情報量はある語が共起相手の語の情報をどの程度持っているかを示す指標である。その語が出れば、共起相手は自動的に決まる、というような場合に相互情報量は高くなる。相互情報量は概念的には、2 語の頻度の実測値と期待値の比を対数に誘導したもので、式は下記のようなになる。(石川慎一郎 2008)

$$I = \log_2 \frac{\text{共起頻度} \times \text{コーパス総語数}}{\text{中心語頻度} \times \text{共起語頻度}}$$

##### 3.2.2. ダイス係数

ダイス係数は情報理論やウェブサイトからの関連語自動抽出のアルゴリズムなどとしても広く使用されている。ダイス係数の計算にはコーパス総語数は必要ではなく、中心語頻度と共起語頻度の関係だけで 2 語のコロケーション強度を計測してゆく。(『NLB』ユーザマニュアル 1.20 : 22)

ダイス係数は共起頻度を中心語頻度と共起語頻度の和で割って 2 倍した値で、式は次のようになる。(石川慎一郎 2008)

$$D = 2 \times \frac{\text{共起頻度}}{\text{中心語頻度} + \text{共起語頻度}}$$

式からわかるようにダイス係数は単純ではあるが、有用性・妥当性の高い指標で、特徴語検出における 9 種類の統計値の妥当性を比較した中條・内山 (2004) によれば、ダイス係数の精度がもっとも良かったことが報告されている。(石川慎一郎 2008)

##### 3.3. 後続する名詞との共起情報

『NLB』コーパスを用いて検索した結果を表 1、表 2、表 3 に示す。後続する名詞の頻度の中で、上位 25 語をそれぞれ抽出し、MI スコアと LD 係数も考えてコロケーションの特徴を見る。

表-1 素敵な+名詞 (上位 25 語)

素敵な＋名詞 (653 種)									
序号	名詞	頻数	MI	LD	序号	名詞	頻数	MI	LD
1	人	107	4.38	3.56	14	思い出	17	8.45	6.91
2	こと	76	2.70	1.89	15	[の] <sup>注3</sup>	16	0.49	-0.33
3	[数字] <sup>注1</sup>	64	0.69	-0.13	16	名前	15	5.89	4.91
4	女性	57	6.78	5.88	17	時間	15	3.80	2.94
5	[ん] <sup>注2</sup>	42	3.62	2.79	18	曲	15	6.46	5.40
6	もの	34	3.11	2.29	19	恋	14	7.90	6.46
7	方(かた)	33	4.28	3.44	20	夜	14	5.57	4.61
8	男性	31	7.21	6.19	21	場所	14	5.15	4.22
9	出会い	27	9.42	7.75	22	クリスマス	11	8.02	6.40
10	プレゼント	27	9.40	7.75	23	作品	11	5.48	4.49
11	笑顔	23	8.54	7.12	24	音楽	11	6.07	5.00
12	ところ	22	4.14	3.29	25	ドレス	10	8.95	6.77
13	言葉	18	4.87	3.98					

表-2 立派な＋名詞 (上位 25 語)

立派な＋名詞 (1086 種)									
序号	名詞	頻数	MI	LD	序号	名詞	頻数	MI	LD
1	もの	139	4.88	4.32	14	人物	11	6.19	5.26
2	人	78	3.66	3.10	15	体格	11	10.21	7.09
3	こと	55	1.97	1.42	16	ん	10	1.28	0.72
4	方(かた)	32	3.97	3.39	17	先生	10	4.29	3.62
5	人間	32	5.35	4.72	18	態度	10	6.33	5.33
6	建物	30	7.63	6.70	19	ところ	9	2.59	2.00
7	家	26	4.24	3.64	20	教育	9	4.10	3.43
8	仕事	22	4.98	4.34	21	犯罪	9	6.18	5.18
9	大人	19	6.91	5.99	22	紳士	9	8.93	6.58
10	[数字] <sup>注4</sup>	16	-1.58	-2.13	23	人格	8	7.70	6.02
11	の	14	0.03	-0.52	24	女性	8	3.69	3.03
12	もん	14	6.23	5.36	25	施設	8	4.21	3.51
13	男	14	4.20	3.57					

表-3 素晴らしい＋名詞 (上位 25 語)

素晴らしい＋名詞 (1201 種)									
序号	名詞	頻数	MI	LD	序号	名詞	頻数	MI	LD
1	こと	262	3.59	3.67	14	仕事	24	4.47	4.41
2	もの	212	4.85	4.92	15	女性	24	4.63	4.56
3	の <sup>注5</sup>	67	1.65	1.74	16	才能	24	8.63	7.21
4	人	62	2.69	2.76	17	眺め	24	10.31	7.61

5	作品	52	6.82	6.60	18	方(かた)	22	2.79	2.83
6	ところ	40	4.10	4.13	19	景色	21	8.73	7.12
7	[地域] 注6	34	0.23	0.32	20	アイデア	20	8.15	6.87
8	[一般] 注7	28	0.35	0.44	21	音楽	18	5.88	5.52
9	ん	28	2.13	2.19	22	効果	17	4.90	4.74
10	人生	28	6.46	6.12	23	演技	17	8.63	6.87
11	世界	27	4.53	4.49	24	自然	17	4.69	4.55
12	[数字]	26	-1.52	-1.43	25	場所	16	4.43	4.32
13	選手	26	6.40	6.05					

### 3.4 後続する名詞の特徴

#### 3.4.1. 共通点

以上の表 1、表 2、表 3 から分かるように、「素敵」「立派」「素晴らしい」はどれも後ろに「人」「こと」「もの」「数字」「方(かた)」「の」

と共に起ることが分かった。その6つの共起する名詞の中で頻度が一番高い名詞は「人」「こと」「もの」であった。その特徴を以下の表 4 に示す。

表-4 後続する同じ名詞 (数字は頻度を表す)

頻度 \ 名詞	素敵な+名詞	立派な+名詞	素晴らしい+名詞
人	107	78	62
こと	76	55	262
もの	34	139	212
[数字]	64	16	26
方(かた)	33	32	22
の	16	14	67

#### 3.4.2. 相違点

「表 1 素敵な+名詞」の共起情報 MI スコアと LD 係数から分かるように、「出会い」「プレゼント」「ドレス」「笑顔」「思い出」「クリスマス」「恋」「男性」が高い。これらの名詞はほとんど女性が憧れるもの、女性の心を引きつける魅力的な感じを与えるものばかりである。この特徴は『現代形容詞用法辞典』での意味記述とだいたい合っている。

「表 2 立派な+名詞」の共起情報 MI スコアと LD 係数から分かるように、「体格」「紳士」「人格」「建物」「大人」「態度」「犯罪」が高い。これらの名詞は「犯罪」を除いて主として「大きい」「豪壮」「威厳」「完全」「崇高」の特徴を現している。『現代形容詞用法辞典』の意味記述は不十分ではあるものの、このような意味合いはちょっと触れている。「立派な犯罪」について

は「好ましくない事柄について用いられた例で、その範囲内に完全に収まっているという意味である」と記述している。(『現代形容詞用法辞典』pp. 596) 劉 (2013) の『映像コーパス』に基づいて「立派な」を検索したら、「立派な犯罪じゃん」「立派な家宅侵入じゃないですか。」「立派な殺人事件です。」等のこのましくない事柄について用いられた例がいくつも出てきた。

『現代形容詞用法辞典』を除いてほかの日本語の辞書にはこのような意味合いを持っている例は出てこない。長年日本語を教えてきたが、「立派な犯罪」のような例は『映像コーパス』を通して偶然発見して、その意味が分かった。

「表 3 素晴らしい+名詞」の共起情報 MI スコアと LD 係数から分かるように、「眺め」「景色」「才能」「演技」「アイデア」「作品」「人生」「選手」が高い。これらの名詞は褒める場合或



いは思わず感嘆すべき様子を表す場合に使われる名詞が多い。この特徴は『現代形容詞用法辞典』での意味記述と合っている。

#### 4. おわりに

以上のような調査の結果から、「素敵」「立派」「素晴らしい」それぞれのコロケーション情報としては、以下のような点を学習者に提示すべき内容と考え、試案として提案する。

1. 「素敵」「立派」「素晴らしい」はどれも後ろに「人」「こと」「もの」「数字」「方(かた)」「の」と共起するし、この六つの中で、頻度が一番高い名詞は「人」「こと」「もの」であった。
2. 「素敵な」と一番共起しやすい名詞は「出会い」「プレゼント」「ドレス」「笑顔」「思い出」「クリスマス」「恋」「男性」である。女性が憧れるもの、女性の心を引きつける魅力的な感じを与えるものが多い。
3. 「立派な」と一番共起しやすい名詞は「体格」「紳士」「人格」「建物」「大人」「態度」「犯罪」である。「犯罪」を除いて主として「大きい」「豪壮」「威厳」「完全」「崇高」の特徴を現している。「立派な犯罪」は「好ましくない事柄について用いられた例で、その範囲内に完全に収まっているという意味である」
4. 「素晴らしい」と一番共起しやすい名詞は「眺め」「景色」「才能」「演技」「アイデア」「作品」「人生」「選手」である。褒める場合或いは思わず感嘆すべき様子を表す場合に使われる名詞が多い。

#### 注

- 注1. 「素敵な+数字」の中で、「数字」は「一日」「一時間」「一年」「一週間」などの名詞を指す。
- 注2. 「素敵な+ん」の中で、「ん」は「素敵なんだ」「素敵なんだけど」「素敵なんじゃない?」「素敵なんだろう」などのような例文での使い方である。

注3. 「素敵な+の」の中で、「の」は「素敵なのがいっぱいあります。」「素敵なのは…」「素敵なのかな…」「素敵なのではないだろうか。」などのような例文での使い方である。

注4. 「立派な+数字」の中で、「数字」は「一品」「4代目」「一国」「10年選手」「四合院」「一人前」などの名詞を指す。

注5. 「素晴らしい+の」の中で、「の」は「素晴らしいのです。」「素晴らしいのは…」などのような例文での使い方である。

注6. 「素晴らしい+地域」の中で、「地域」は「素晴らしい釜石の朝でした。」「素晴らしい日本の文化」「素晴らしいフランス料理」「素晴らしい中国」などのような例文での地域を指す。

注7. 「素晴らしい+一般」の中で、「一般」は「港町」「里山」「高原」「丘」「マリア」などのような例文での名詞を指す。

#### 参考文献

- 田野村忠温 (2009) 「コーパスからのコロケーション情報抽出-分析手法の検討とコロケーション辞典項目の試作-」『阪大日本語研究』No. 21, pp. 21-41
- 大曾美恵子・滝沢直宏 (2003) 「コーパスによる日本語教育の研究-コロケーション及びその誤用を中心に-」『日本語学』No. 22, pp. 234-244
- 中溝朋子・坂井美恵子 (2011) 「漢語名詞『進歩』と『向上』のコロケーションの異同について」『大学教育』Vol. 8, pp. 88-95
- 金田一京助ら (昭和 54) 『新明解国語辞典』(第二版), 三省堂
- 文化庁 (昭和 46) 『外国人のための基本語用例辞典』, 大蔵省印刷局
- 飛田良文・浅田秀子 (1991) 『現代形容詞用法辞典』, 東京堂出版
- 石川慎一郎 (2008) 「コロケーションの強度

をどう測るか-ダイス係数、tスコア、相互情報量を中心として-」言語処理学会14回大会チュートリアル資料, pp. 40-50

中條清美・内山将夫 (2004) 「統計的指標を利用した特徴語抽出に関する研究」『関東甲信越英語教育学会紀要』No. 18, pp. 99-108

劉玉琴ら (2013) 『映像コーパス』  
<http://www.japanesecorpus.com/>