

# 現代日本語書き言葉均衡コーパスにおける 漢語名詞「影響」のコロケーションの特徴 修飾語および述語動詞との共起を中心にー

中 溝 朋 子  
坂 井 美 恵 子  
金 森 由 美

## 要旨

日本語学習者のコロケーション習得を目的とした教材を作成するために、『現代日本語書き言葉均衡コーパス』(国研,2009)を用いて漢語名詞に関するコロケーションについて調査を行った。本稿では,その一例として「影響」を取り上げ,「影響」を修飾する語と共起する述語動詞について明らかにするとともに,学習者に提供するコロケーション情報の試案を作成する。収集した用例の分析には,現在開発中の共起関係抽出プログラムを用いて共起頻度を明らかにし,ダイス係数とともにコロケーション強度の指標とした。

## キーワード

現代日本語書き言葉均衡コーパス,コロケーション,漢語名詞,「影響」,共起

## 1 はじめに

近年日本語教育においてコロケーションの重要性が多く指摘され(大曾,2005;三好,2007他),コロケーション習得に特化した教材も出版されている。しかしコーパスなどを利用したコロケーションの使用実態について,日本語教育を目的とした研究はまだ少数であり,どの範囲をコロケーションと考えることが学習者にとって有益なのか,具体的にどのようなコロケーション情報を学習者に提示すべきかについては,今後も検討が必要と考えられる。

本稿ではこのような現状を踏まえ,漢語名詞「影響」を一例として取り上げ,日本語学習者に役立つようなコロケーション情報に関

する試案を提案し,教材作成の一助となることを目的とする。

## 2 本稿で考えるコロケーション,およびコーパスの有用性

### 2.1 本稿で考えるコロケーション

日本語教育のためのコロケーションの範囲,また学習者にとって有用なコロケーション情報は何かという点は難しい課題である。従来日本語教育では,コロケーションは固定度が高い慣用句や自由結合との区別という点から議論されてきた「連語的慣用句」(宮地,1985),「連語」(国広,1997)などの概念に近い,ある程度固定的なレベルであった(三好,2007)。これらは「風呂から上がる」「暇を潰す」など

「語と語の結び付きかたは決まっているけれども、全体の意味は個々の語の意味からすぐ分かる」もので「外国人学習者には特に必要」な知識とされている(国広,1997:128-129)。しかし近年これらに加え、学習者の母語の影響や誤用の可能性などを考慮し、三好(2007)では「薬を飲む」、大曾(2005)では「風呂を沸かす」など自由結合に分類可能なものの中にも日本語教育においてはコロケーションとして取り上げるべき結び付きがあると指摘されている。本稿でも三好(2007)や大曾(2005)と同様に、学習者が自然な日本語を効率的に習得するために必要な情報を提供することが重要であると考え、コロケーションの範囲を広く考える。

また大曾(2005)では、学習者に提供すべきコロケーション情報を選定するためのコーパス利用の有用性が述べられている。例えば文法的にはともに可能な「大きい+Noun」と「大きな+Noun」の共起頻度の比較、「議論」と「論議」がスル動詞として使用可能か、「-を呼ぶ」などの動詞と共起可能かなど数値を挙げて説明し、その違いを明らかにしている。日本語学習には、こうした語の使用法は重要な情報であり、そのためには語の使用実態を明らかにすることは不可欠である。このことから本稿でもコーパスを用いて、使用実態に基づいたコロケーション情報の試案を作成する。

なお本稿では、日本語教育における有用性という観点からコロケーションを捉えるため、一般的には含まれない接辞も含めて検討する。これは例えば「悪影響」の「悪」など、接頭辞の一部は形容詞や連体詞と同様の意味を持ち、名詞を修飾するため、学習者が日本語を産出する際には、接頭辞も形容詞等と同様の選択肢として扱うであろうことを考慮しているためである。

### 3 漢語名詞の特徴と本稿で取り上げる「影響」について

漢語名詞は、中級レベル以上でその使用が増加し、特に書き言葉において多用される傾向がある。この漢語名詞の使用、特に漢語名詞と動詞との共起については、和語や外来語と比較して以下のような特徴があり、学習者、特に初級から中級への移行期にいる学習者や非漢字圏学習者には漢語名詞を使用する際の困難点となっていると考える。

一つ目は、漢語名詞の一部は複合動詞、いわゆるスル動詞として使用可能であるが、どれがスル動詞として使用可能か区別がつきにくいこと、二つ目は、スル動詞には自動詞、他動詞、両用動詞の区別がありそれらを区別しなければならないこと、その上でスル・サレル・サセルなどヴォイスを適切に選択する必要があることが挙げられる。三つ目は、これらスル動詞を構成する名詞は、主に「動作名詞=なんらかの動的な運動が名づけられている名詞(村木,1991:214)」であるが、これらはスル動詞とともに、本来、動詞が表すべき動作の内容、すなわち動詞の「実質的な意味を名詞にあずけて、みずからはもっぱら文法的な機能をはたす」機能動詞(村木,1991:203)(例えば「(連絡を)とる/(決定を)下す」など)と共起することが多いという点が挙げられる。学習者にとって、これらのどの動詞が、どの漢語名詞と具体的に共起するかは予測が難しいと考えられる。四つ目は、漢語名詞は書き言葉で多く使用される傾向があるため、例えば初級で学習する同様の意味を表す動詞では、文体や位相などにより共起不可能となる場合があるという点である。

「影響」は「影響スル」というスル動詞があり、文法的には「AガBニ影響スル」とい

う自動詞であるが、意味的には「AガBニ影響ヲ与える」という他動詞が述語の文と同じ事象を表す。そのため、学習者にとっては自他、およびヴォイスの選択には困難が伴う語と考えられる。

さらに村木(1991)は、機能動詞のひとつの役割として「動作名詞と形式(的な)動詞のくみあわさった合成述語形式が<能動><受動>のヴォイスの対立を表す」点を挙げ、これを「迂言的な受動表現」と呼び、例として「受ける<受動>」を挙げている(村木,1991:197)。例えば「攻撃を受ける」と「攻撃される」は、ともに受動の代替表現となるが格支配は異なると述べている(村木,1991:238)。例えば「敵から」は両動詞とも共起可能であるが、「敵からの/敵の」は「攻撃される」とは共起不可能である。「影響」もその動作動詞という性質から機能動詞との共起、さらに授受を表す動詞との共起が予想され、これらの使用は学習者にとって困難であることが予想される。また一方で、初級で学習し、授受表現の代表的な語である「あげる/もらう」は「影響」と共起しない<sup>1)</sup>。

このような特徴を持つ「影響」は、漢語名詞の学習上の困難点を典型的に持つ語の一つと考えられることから、本稿では漢語名詞の一つの例として「影響」を取り上げることとした。

#### 4 コーパスおよび調査方法

本稿で使用したコーパスは、『現代日本語書き言葉均衡コーパス』2009年度版モニター公開データ(国研,2009)である。同コーパスは、現代日本語書き言葉の実態の正確な縮図となることを目指し、新規出版物の刊行情報に基づく生産実態サブコーパス(1300万語)、図

書館蔵書に基づく流通実態サブコーパス(1500万語)、特定目的の日本語を収集した非母集団サブコーパスから成る。収録語数は4520万語(短単位)で、ベストセラー(230万語)、白書(480万語)、「Yahoo!知恵袋」(520万語)、国会会議録(490万語)が収録されている(石川,2010)。本研究では、コンコーダンサ「ひまわり」<sup>2)</sup>を利用し「影響」が含まれる8916例の用例を収集した。そのうちスル動詞(「影響スル」)879例と、合成語(複合語や派生語)911例は除外した。残りの7126例について共起する語の抽出と統計値を算出するために、共起関係抽出プログラム<sup>3)</sup>を利用した。このプログラムは、形態素解析器 MeCab および係り受け解析器 CaboCha を用い、「名詞+(助詞+)(主に)動詞・形容詞」、「修飾語+被修飾語(名詞)」、「副詞+被修飾語(主に動詞・形容詞)」の共起関係を抽出するものである。本稿ではコロケーション強度を示す値として共起頻度とダイス係数を採用しており<sup>4)</sup>、用例の確認、およびダイス係数の計算には Microsoft Excel を使用した。以下、調査の結果について述べる。

### 5 調査結果

#### 5.1 「影響」を修飾する語の内訳

「影響」7126例を修飾する語の用例について調査した結果を表1に示す。それぞれ「名

表1 「影響」を修飾する語の形式別内訳

語・品詞	用例数
名詞+の+影響	2601例
-い+影響(イ形容詞218例、および、「かもしれない」2例)	220例
-な+影響(形容動詞455例、連体詞637例、助動詞等6例を含む)	1098例
動詞+影響(ナイ形7例を含む)	687例
計	4606例

詞+の」による修飾が 2601 例,「～い影響」が 220 例,「～な影響」が 1098 例,「動詞+影響」が 687 例であった。

### 5.2 「～い影響」と「～な影響」

次に「～い影響」と「～な影響」の内訳をダイス係数の高い順に表 2 に示す<sup>5)</sup>。この中では「大きな」がダイス係数,共起頻度ともに圧倒的に高い値を示している。以下,同様に「影響」の程度を表す語としては「重大な」「強い」などが,プラス・マイナスの評価を表す語としては,「深刻な」「悪い」「良い」などが両指標とも高い値を示しており,これらが「影響」にかかる代表的な修飾語と言うことができる。

表 2 「～い影響」と「～な影響」の語の内訳

～い影響	ダイス	共起頻度	～な影響	ダイス	共起頻度
強い	0.670	53	大きな	4.423	512
悪い *	0.560	45	重大な ***	1.565	79
良い	0.489	42	深刻な	1.134	55
少なからぬ	0.467	21	多大な	0.924	42
著しい	0.406	20	有害な	0.375	17
大きい **	0.195	13	重要な	0.354	26
深い	0.143	9	決定的な	0.324	15
測りしれない	0.134	6	様々な	0.322	26
好ましくない	0.132	6	直接的な	0.241	11
好ましい	0.108	5	大変な	0.238	13

\*「悪しき」2 例を含む \*\*落字「大き」1 例を含む

\*\*\*「重大なる」1 例を含む

### 5.3 「影響」の評価を表す語

次に,表 3 に「影響」を修飾する語のうちプラスとマイナスの評価を表す修飾語を取り上げて分けて示す。

全体的にはマイナス評価の語が,異なり語数,共起頻度,ダイス係数すべてにおいてプ

表 3 プラス評価とマイナス評価の修飾語

プラス評価	ダイス	共起頻度	マイナス評価	ダイス	共起頻度
良い	0.489	17	悪 -	5.860	310
好 -	0.296	42	悪い	5.600	45
好ましい	0.108	5	深刻な	1.134	55
有益な	0.044	2	有害な	0.375	17
好ましい	0.020	1	甚大な	0.222	10
			好ましくない*	0.110	6
			ネガティブな	0.089	4
			良くない	0.060	3
			ひどい	0.037	2

\*「好ましからざる」1 例を含む

ラス評価の語を大きく上回っている。具体的な語としては,接頭辞「悪 -」,形容詞「悪い」がダイス係数,共起頻度ともに突出して高く,以下,「深刻な」「有害な」などが続いている。プラス評価の語の中では,「良い」「好 -」が他の修飾語に比較すると,両指標とも比較的高い。

さらに表 4 では,両評価の上位語が「影響」を修飾する場合に述語となる動詞を比較する。両評価とも「与える」「及ぼす」など「影響」を授与する動詞の頻度が高いという傾向が見られる。

表 4 評価の修飾語と共起する述語動詞

プラス評価修飾語	共起した動詞	共起頻度	マイナス評価修飾語	共起した動詞	共起頻度
好 - (17 例)	及ぼす	8	悪 - (310 例)	及ぼす	90
	与える	8		与える	71
	出る	1		ある	27
				出る	19
良い (42 例)	与える	24	悪い (45 例)	与える	21
	出る	5		及ぼす	11
	及ぼす	各 4		もたらず	6
	ある			受ける	5
好ましい (5 例)	与える	4	深刻な (55 例)	与える	16
				及ぼす	10
				ある	8
有益な (2 例)	もたらず	各 1	有害な (17 例)	与える	各 4
	及ぼす (否定形)			受ける	
					及ぼす

#### 5.4 「～的影響」か「～的な影響」か

次に「影響」の修飾語で「～的」と「～的な」のダイス係数の上位語を表5に比較する。「～的影響」は計176例、「～的な影響」は計97例あったが、前者では、「経済」「社会」など「影響」の分野を表す普通名詞が両指標とも高く、後者では「影響」の強度や評価、その伝わり方など、「影響」の特徴を表す語が両指標とも高い傾向が見られた。

なお、このような傾向は、「～的」と「～的な」が他の名詞を修飾する際にも同様に見られると考えられるため、今後他の名詞との共起についても検討が必要である。

表5 「～的影響」と「～的な影響」

「～的」	ダイス	共起頻度	「～的な」	ダイス	共起頻度
経済的	0.741	39	決定的な	0.324	15
社会的	0.484	28	直接的な	0.241	11
直接的	0.195	9	間接的な	0.111	5
遺伝的	0.156	7	社会的な	0.106	5
確率的	0.134	6	致命的な	0.089	4
思想的	0.110	5	潜在的な	0.088	4
間接的	0.109	5	圧倒的な	0.066	3
政治的	0.095	5			

#### 5.5 「影響」と共起する述語動詞

次に、「影響」と共起する述語動詞についてダイス係数および共起頻度が高い語を表6に

表6 「影響」と共起する述語動詞

述語動詞	ダイス	共起頻度
及ぼす	14.807	755
与える	13.793	1223
受ける	7.539	912
もたらす	1.202	71
出る	1.083	196
及ぶ	1.004	65
生じる	0.782	49
懸念する	0.775	36
見る	0.729	74
現われる	0.652	47

示す。述語動詞では上位2語「及ぼす」「与える」のダイス係数が圧倒的に高く、また比較的ダイス係数の値が高い上位6語のうち3語は「影響」の授与(「及ぼす」「与える」「もたらす」)を表す動詞である。

次に共起する述語動詞の中で圧倒的に両指標が高かった上位3語の「及ぼす」「与える」「受ける」と「影響」の修飾語の共起について、表7で比較する。

表7 述語動詞「与える」「及ぼす」「受ける」と「影響」の修飾語

	与える		及ぼす		受ける	
述語動詞頻度	1215		752		903	
～の影響	48(4%)		37(5%)		548(61%)	
～い影響	96(8%)		43(6%)		32(4%)	
上位語5語	良い	29	悪い	8	強い	20
	強い	16	著しい	8	深い	3
	悪い	16	強い	7	良い	2
	著しい	4	大きい	7	著しい	1
	大きい	4	良い	5	大きい	1
～な影響	456(38%)		215(29%)		87(10%)	
上位語5語	大きな	281	大きな	95	大きな	45
	深刻な	21	重大な	24	多大な	7
	重大な	19	深刻な	10	有害な	5
	多大な	16	重要な	10	深刻な	4
	重要な	13	多大な	9		

全体的には、3語の中で「影響」の授与を表す「与える」「及ぼす」と享受を表す「受ける」で異なる傾向を示している。まず「～の」による修飾が、「受ける」では頻度の61%を占めているのに対し、「与える」では5%、「及ぼす」では5%のみである。一方、「～い」「～な」による修飾については、「受ける」の頻度に占める割合は、それぞれ4%、10%であるのに対し、「与える」「及ぼす」はその約1.5倍から3倍以上の割合を占めるという逆の傾向を示している。すなわち「受ける」では名詞を用いて影響を与えている事物を限定する修飾が多いのに対し、「与える」「及ぼす」では「～い」「～な」など主に形容詞を用いて影響の規模や評価を述べる場合が多いことが大

きな特徴として挙げられる。

また「～い」「～な」の修飾語については、3語ともほぼ同様の語が共起している。

## 6 学習者に提供すべき「影響」のコロケーション情報試案

以上のような結果から、本稿では漢語名詞「影響」のコロケーション情報として、次のような点を共起する修飾語および述語動詞の基本的な情報として、学習者に提示することを提案したい。なお語の選定にあたっては、学習者が「影響」を使用する際には、様々な「影響」の種類やアスペクトが使用される可能性があることから、語の意味別に、共起頻度やダイス係数の値の高いものを提示することとする。

(1)「影響」と共起する代表的な動詞(規模や程度を強調する場合は[ ]内の修飾語を使用)

「影響」の授与：～に[大きな/重大な/深刻な/強い]影響を及ぼす/与える/もたらす

「影響」の享受：～の[大きな/強い]影響を受ける

「影響」の出現：～の影響が出る/及ぶ/現われる

「影響」の存在：影響がある

(2)「影響」の評価

プラス評価：～に良い影響を与える

マイナス評価：～に悪い影響を及ぼす、(悪い/深刻な)影響を与える

(3)その他「影響」の修飾語

(経済的/社会的/遺传的)影響

(決定的な/直接的な)影響

## 7 今後の課題

以上、コーパスを利用して漢語名詞「影響」と共起する語の実際を調査し、学習者に提供すべき情報について試案を作成した。今後も共起関係抽出プログラムの精度を高めつつ、2011年にDVDで公開された『現代日本語書き言葉均衡コーパス』など大規模コーパスを利用し、漢語名詞について広く共起関係を調査し、情報を蓄積させていきたい。これらを通じて学習者にどのような情報が必要とされるのかをさらに検討し、教材作成に反映させていきたい。

(山口大学留学生センター 准教授)

(大分大学国際研究教育センター 准教授)

(大分大学国際研究教育センター 講師)

### 【参考文献】

石川慎一郎, 2008a, 「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.

石川慎一郎, 2008b, 『英語コーパスと言語教育』大修館書店.

石川慎一郎, 2010, 「『現代日本語書き言葉均衡コーパス』(BCCWJ)における複合動詞『～出す』の量的分析」『統計数理研究所研究レポート』238, pp.15-34.

大曾美恵子, 2005, 「コーパスによるコロケーションの特定 - 日本語学習辞書の充実を目指して -」影山太郎編『レキシコンフォーラム No.1』ひつじ書房, pp.11-23.

国広哲弥, 1997, 『理想の国語辞典』大修館書店.

国立国語研究所, 2009, 『現代日本語書き言

葉均衡コーパス」モニター公開データ(2009年度版)』DVD.

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』18号, pp.99-108.

宮地裕, 1985, 「慣用語の周辺-連語・ことわざ・複合語」『日本語学』Vol.4 No.1, pp.62-75. 明治書院.

三好裕子, 2007, 「連語による語彙指導の有効性の検証」『日本語教育』134号, pp.80-89.

村木新次郎, 1991, 『日本語動詞の諸相』ひつじ書房.

#### 【注】

1) 「あげる」「もらう」が「影響」と共起しない理由は, 文体や位相の問題ではなくこれら2語が抽象名詞と共起しにくいことが主な理由と考えられる。

2) 「ひまわり」は国立国語研究所によって開発された全文検索システムで, 『現代日本語書き言葉均衡コーパスモニター公開データ(2009年度版)』に組み込まれているものである。

3) 本プログラムは, 山口大学メディア基盤センター刈谷丈治元教授により, 現在開発中である。

4) ダイス係数は中心語頻度, 共起語頻度, 共起頻度の三つのパラメータを同時に評価する指標で, 数値が高いほど共起頻度の妥当性が高い。算出方法は次のとおりである。

$D = 2 \times \text{共起頻度} / (\text{中心語頻度} + \text{共起語頻度})$ これはTスコア, MI 係数などと同様に共起の強度を測る指標の一つであるが, これら指標の中からダイス係数を採用した理由は, 中條他(2004)の9種類の指標の妥当性を比較した研究で, ダイス係数の精度が最も高いと検証されているなど, 有用性, 妥当性が高く, 統計値の単独指標として最も有効だと多

くの研究で指摘されているためである(石川2008a, 2008b)。なお, 本稿では, ダイス係数については便宜上, すべて100倍した値で示す。

5) 以下, コーパス中で漢字とひらがななど表記が異なっても同じ語と認められる場合は, 同一の語として集計し, 本稿における表記は漢字書きで統一した。

#### 【謝辞】

本研究は科研費(基盤研究(C)20520473)の助成を受けたものである。また, 本稿で言及したプログラムは, 山口大学メディア基盤センター刈谷丈治元教授により現在も開発中である。この場をお借りして心よりお礼を申し上げます。