

# 日中同形語のコロケーション

―「感染」を例として―

## The Collocation of Japanese and Chinese Homographs: For "kansens"

何 龍

He Long

### Abstract

This research is based on both the Japanese corpus and Chinese corpus. Based on the statistical frequency, t-score, MI-score and Dice coefficient, this essay confirms the significant collocation of Japanese and Chinese homograph [kansens]. In addition, this essay considers the collocation of [kansens] in Japanese and Chinese. As a result, it provides support for the idea that Chinese Speakers learning Japanese get a negative effect from their mother tongue which causes them to overuse or underuse the [kansens]. The negative effects from their mother tongue seem to come from the differences between Japanese and Chinese, the kinds of co-occurrence (direct・indirect) and the relationship between co-occurring words. At the same time, this essay discusses the positive effects of the mother tongue. Chinese speakers learning Japanese can use the knowledge of their mother tongue to help them learn Japanese and Chinese homographs [kansens].

### 1. はじめに

日本語と中国語は、共に漢字を使用することで、日本語においても、中国語においても、表記が同じである日中同形語が生まれた。日中同形語は、同じ漢字を使用する中国人日本語学習者にとって、学習しやすいことは言うまでもない。しかし、多くの先行研究によると、日本語が中国語と同じ意味を取る日中同形語を使用する際に、中国人日本語学習者は、誤用を起こすことが明らかになった。その原因は、日中同形語が中国語の意味と日本語の意味と異なるだけでなく、日中同形語のコロケーション、連語形式や品詞などの用法が異なるからである。

本稿はコーパスによる検索の研究手法を用いて、日中同形語のコロケーションの実態を調査し、日本語教育の立場から、中国人日本語学習者は、母語のコロケーションから受ける影響を明らかにすることを旨とする。そして、中国人日本語学習者に役立つ日中同形語のコロケーション情報を提供できればと願っている。

## 2. 先行研究

### 2.1. コロケーションの捉え方

本稿で問題視しているコロケーションとは何か。次に、「辞書の捉え方」、先行研究の「石川（2006）の捉え方」及び「田野村（2009）の捉え方」を紹介しながら、本稿で扱うコロケーションについて述べたいと思う。

#### 辞書の捉え方

まず、コロケーションについて、3つの辞書を調べた。その結果は以下の通りである。

『日本国語大辞典』<sup>1</sup>によると、コロケーションとは「文や句において、文法的、意味的に関連する二つ以上の単語の結合がある程度固定化している関係。また、その結合のしかたをいう」ということである。『デジタル大辞泉』によると、コロケーションとは「二つ以上の単語の慣用的なつながり。連語関係」ということである。『広辞苑（第5版）』によると、コロケーションとは「文・句における語の（慣用的な）つながり方。連語法」ということである。

辞書の捉え方から分かるように、コロケーションは、2つ以上の単語が連結することであり、その連結は、ある程度に固定化または定着している。よって、学習者にとって、より母語話者に近い自然な言語を使うために、コロケーションは、非常に重要である。

#### 石川（2006）の捉え方

石川（2006）は、コロケーションに「語と語の機械的な連結である」と定義づけ、コロケーションを「偶然的な・1回のなコロケーション」と「傾向性をもった有意なコロケーション」に分けた。さらに、「コロケーションは共起要素が個別的な表記形であるため、関係が計量的に把握しやすい」と指摘している。

石川（2006）の捉え方から分かるように、コロケーションの検出について、計量的な統計は可能である。いわゆる統計手法を用いて、有意なコロケーションを検出できる。

#### 田野村（2009）の捉え方

田野村（2009）は、コロケーションに「言語表現全般（語、語の連鎖、句、節など）の間に観察される習慣的な共起関係である」と定義づけ、「コロケーションに関する正確な情報を得るには、大量の用例を収集して分析することが必要である」と指摘している。

田野村（2009）の捉え方から分かるように、コロケーションに関する研究は大量の用例を分析することによってはじめて成り立つものである。時間とコストがかかるため、そのような研究は進まなかった。しかし、大規模な言語コーパスが相次ぎ構築されたので、大量の用例は簡単に集められるようになり、コロケーションに関する研究がしやすくなった。

#### 本稿の捉え方

本稿は、日中同形語のコロケーションについて、議論を展開するため、コロケーションを次のように取り扱う。

コロケーションは中心語となる日中同形語とその共起語との固定化された関係である。

本稿は、日本語教育の立場において、中国人日本語学習者に、日中同形語について、より正確なコロケーション情報を提供するため、統計指標に基づいて、傾向性をもった有意なコロケーションを検出し、取り上げることを目的とする。

## 2.2. コロケーションの検出に関する研究

先行研究を調べた結果、コロケーションの検出に関する研究は、石川（2006・2008）、中溝ら（2012）が挙げられる。

石川（2006・2008）によれば、コロケーションを検出する上で、使用される統計指標としては、「tスコア<sup>2)</sup>」、「MIスコア<sup>3)</sup>」や「ダイス係数<sup>4)</sup>」などがある。さらに、石川（2006・2008）は、それらの指標を使って、コロケーション検出の実験を行った。

石川（2006）は、各指標のコロケーション検出力を比較した結果、「tスコア」は、頻度に偏る指標で、生起頻度の高い一般的なコロケーション検出に強く、「MIスコア」は、頻度情報にあまり依存しない指標で、生起頻度が少ないものの、特徴性・共起傾向性を示すコロケーション検出に強い。さらに、以上の統計指標は、研究の目的に応じて使い分けしたり、複数の指標を組み合わせて使ったりする必要のあることも分かった。

中溝ら（2012）は、日本語漢語名詞のコロケーションの検出方法について、以下のように説明している「検索した文について形態素解析エンジン MeCab、および Microsoft Excel を使用し、共起する動詞、前接及び後接する語を抽出し、語数のカウントを行った」。さらに、検出方法に従って、得られたデータに基づいて、日本語漢語名詞のコロケーションについて分析した。

## 2.3. 日中同形語のコロケーションに関する先行研究

先行研究を調べた結果、日中同形語に関する研究が盛んに行われるようになったことが分かった。しかし、その研究の多くは、日中同形語の意味に注目してきた。日中同形語のコロケーションに関する先行研究が多くなかったのは、事実である。しかし、小森ら（2012）と三國ら（2015）が挙げられる。

小森ら（2012）は、日本語の意味と中国語の意味が同じである日中同形語に対し、中国人日本語学習者が日本語と中国語において、同じ共起語をとれるものととれないものを、どの程度正しく区別できるかについて調査を行った。その結果、日本語能力試験1級以上の学習者でも習得が進んでいないことが分かった。

三國ら（2015）は、同じく日本語の意味と中国語の意味が同じである日中同形語に対し、中国人日本語学習者が日本語の共起語と中国語の共起語が異なるものについて、訂正できるかどうかを明らかにするため、調査を行った。その結果、日本語能力試験1級以上の学習者でも、その訂正が容易ではないことが明らかになった。

以上の先行研究から、日本語の意味と中国語の意味が同じである日中同形語においても、異なるコロケーションをとることによって、中国人日本語学習者の学習が進んでいないことが分かった。

## 2.4. 先行研究の問題点

石川(2006・2008)は、統計の手法に基づいて、コロケーションの検出に、より正しい方法を提示したが、コロケーションに関する分析は、数値の統計に留まった。中溝ら(2012)は、日本語漢語名詞のコロケーションを分析したが、そのコロケーションの検出方法について詳しく説明していなかった。小森ら(2012)と三國ら(2015)は、日中同形語のコロケーションに関して、議論を展開したが、コロケーションの検出基準は明確ではなく、コロケーションの強度問題も問題視しなかった。

本稿は、先行研究の統計指標を受け継ぎ、日中同形語のコロケーションを検出する。そして、検出した日中同形語のコロケーションの強度情報を用いて、分析を行う。その分析をすることによって、中国人日本語学習者は、日中同形語の学習において、母語のコロケーションからどのような影響を受けているのかを明らかにする。

## 3. 本論

### 3.1. 目的

本稿の目的は、以下の2つである。

1. コーパス検索の研究手法を用いて、日中同形語の日本語コロケーションと中国語コロケーションを検出し、そのコロケーション強度を測る。
2. 日中同形語の日本語コロケーションと中国語コロケーションを比較することによって、中国人日本語学習者が日中同形語の学習をする時の問題点を分析する。

### 3.2. 分析対象

本稿は、日中同形語「感染」のコロケーションについて議論する。日中同形語の「感染」に焦点をあてる理由は、以下の2つである。

1. 胡(2012)は「日本語『感染』はマイナスのイメージを持つのに対して、中国語“感染”はプラスのイメージを持っている」と指摘している。中国人日本語学習者は、日本語「感染」を学習する際に、母語から影響を受ける可能性が高いと考えている。
2. 辞書を調べた結果、日中同形語「感染」は、日本語においても、中国語においても、常用語<sup>5</sup>であることが分かった。

### 3.3. 使用されるコーパスについて

本稿は、以下の基準に従って、コーパスを選ぶことにする。

1. 広範囲からサンプリングしなければならないため、日本語コーパスも、中国語コーパスも、均衡コーパスでなければならない。
2. 話し言葉の影響を抑えるために、日本語コーパスも、中国語コーパスも、書き言葉コーパスでなければならない。
3. 時代の影響を避けるために、日本語コーパスも、中国語コーパスも、現代語コーパスでなければならない。

以上の基準から、本稿は、国立国語研究所が2011年に公開した『現代日本語書き言葉均衡コーパス』（以下はBCCWJと称する）と中国教育部语言文字应用研究所が2009年に公開した《国家语委现代汉语通用平衡语料库》（以下は《语料库》と称する）を使用する。

BCCWJに収録された語数は、約1億語である。この中には、書籍、雑誌、新聞、白書、教科書、広報紙、Webの掲示板、ブログなど多様な日本語が含まれている。サンプルデータは、公開されている各種出版データや東京都下の公共図書館の蔵書データから、無作為に抽出されたものである。ここで断っておきたいのは、本稿が書き言葉に焦点をあてて研究するため、「yahoo 知恵袋」、「yahoo ブログ」と「国会会議録」という話し言葉を含む可能性が高いジャンルを外すことにした。よって、BCCWJの総語数は、79,357,975語となった。

《语料库》に収録された語数は、12,842,116語である。この中には、“人文与社会科学类”、“自然科学类”と“综合类”など多様な中国語が含まれている。サンプルデータは、主に教材、書籍、新聞、応用文から、無作為に抽出されたものである。

### 3.4. 分析手順

本稿は、以下の手順に従って、分析を行うことにする。

- a. BCCWJと《语料库》から、日中同形語「感染」をキーワードとして検索する。日本語「感染」と中国語“感染”の使用頻度を調べる。
- b. コーパス検索で得られた形態素解析済みのデータ<sup>6</sup>をエクセルにコピーする。
- c. 日中同形語「感染」の前後5語を範囲として、日中同形語「感染」のコロケーションを抽出する。
- d. 抽出した日中同形語「感染」のコロケーションについて、日本語と中国語のコーパスでの使用頻度を統計する。
- e. 日中同形語「感染」の使用頻度とコロケーションの使用頻度でtスコアを計算する。
- f. tスコアの計算値に従い、日中同形語「感染」の有意なコロケーションを検出する。
- g. 有意なコロケーションに対して、MIスコア、ダイス係数を計算する。
- h. 日中同形語「感染」のコロケーション強度情報を用いて、日本語「感染」のコロケーションと中国語“感染”のコロケーションについて、対照分析を行う。

### 3.5. データのまとめ方

上記の手順に従い、コロケーションについて、分析を行った。その結果を次のようにまとめる。有意な日中同形語「感染」のコロケーションを名詞、動詞及びその他の品詞の3つに分ける。そして、名詞グループをさらに前接と後接の2つのグループに分ける<sup>7</sup>。

### 3.6. 結果の分析

#### 3.6.1. 日中同形語「感染」と前接する名詞について

BCCWJから日本語「感染」の検索結果を表1にまとめ、《语料库》から中国語“感染”の検索結果を表2にまとめた。

表1 日本語「感染」と前接するコロケーション(名詞)

共起語	共起頻度	MIスコア	ダイス係数
ファイル	12	5.15	0.002
パソコン	7	4.13	0.001
ディスク	7	6.65	0.002
ワーム	7	10.53	0.003
ブート	6	10.77	0.003
ウイルス <sup>8</sup>	223	10.57	0.06

表2 中国語“感染”と前接するコロケーション(名詞)

共起語	共起頻度	MIスコア	ダイス係数
情緒	20	12.06	0.02
艺术	8	7.86	0.003
形象	7	9.07	0.005
精神	6	7.82	0.002
观众	6	9.37	0.005
病毒	57	14.92	0.09

表2から分かるように、中国語“感染”と前接する名詞で、“病毒”と“观众”以外は、すべて抽象名詞である。その中では“情緒”、“艺术”、“形象”、“精神”、“观众”の5つは、中国語“感染”と比較的強いコロケーションを持っている。一方、表1から分かるように、日本語「感染」と前接する名詞には、そのような表現がなかった。それが原因で、中国人日本語学習者は、母語のコロケーションから負の影響を受け、日本語「感染」を過剰に使用する恐れがあると考えられる。

以上の予想を検討するために、《语料庫》から中国語“感染”と最も強いコロケーションを持つ“情緒”について、調べた結果、以下の例文が得られた。

例1：海燕的矫健、勇猛、搏击风浪的飒爽英姿，启发了高尔基，使他从海燕的这些审美属性提供的**情绪感染**<sup>9</sup>和意向暗示中结合着自己的性格特点和审美要求，经过心灵的酿造，再对象化到革命者的形象上去，这样，既肯定了海燕，也肯定了自己——无产阶级革命文艺战线上的“海燕”，更肯定了革命者——俄国无产阶级革命的“海燕”。〔审美主客体 陆贵山 1989<sup>10</sup>〕

例1を見ると、中国語“感染”は“情緒”のような抽象名詞と一緒に共起することが分かった。一方、BCCWJを調べると、有意なコロケーションとして、日本語「感染」は、抽象

名詞と一緒に共起する用例文が見つからなかった。中国人日本語学習者は、母語の用法から影響を受け、日本語「感染」をふさわしくない用法で使用する恐れがあると考えられる。

一方、表1から分かるように、日本語「感染」と前接する名詞には、「ウイルス」以外には、「ファイル」、「パソコン」、「ディスク」、「ワーム」、「ブート」などの「パソコン」に関連する名詞がある。しかも、すべて日本語「感染」と比較的強いコロケーションを持っている。表2から分かるように、中国語“感染”と前接する名詞には、そのような表現がなかった。それゆえに、中国人日本語学習者は、母語のコロケーションから負の影響を受け、「パソコン」に関する名詞と共起する場合、日本語「感染」の使用を避けてしまう。すなわち、日本語「感染」を過少に使用する恐れがあると考えられる。

以上の予想を検討するために、BCCWJから日本語「感染」と強いコロケーションを持つ「ファイル」と「ディスク」を調べた結果、以下の例文が得られた。

例2：もしそのプログラムがウイルスのような動きを示し、他のファイルに感染を試みようとしたらそのプログラムを捕獲する。「出典：危ないネットの歩き方 川口友萬など 2002」

例3：ディスクアクセスを監視し、アクセスしたディスクに感染する。このウイルスは感染するだけで発病はしない。「コンピュータウイルス不正アクセス対策マニュアル 中村達 1998」

さらに、BCCWJから日本語「感染」と最も強いコロケーションを持っている「ウイルス」の例文を調べた結果、以下の例文があった。

例4：住所録に登録されたメールアドレスに対して自動的にウイルス感染したファイルを送りつけるという機能をもっていたことだ。「危ないネットの歩き方 クロサワトシオなど 2002」

例5：ウイルスの侵入を未然に防ぐことができるだけでなく、ウイルスに感染してしまったファイルから、ウイルスだけを取り除くこともできます。「朝日ビジネスPASO（ビズぱそ） 中村南 2002」

例2～例5から分かるように、「ウイルス」は「パソコン」に関連する名詞として使用されることが分かった。一方、《语料庫》から中国語“感染”と前方に共起する“病毒”を調べた結果、いずれも「パソコン」に関連する名詞ではないことが分かった。それが原因で、中国人日本語学習者は、「パソコン」に関連する名詞と前方に共起する場合、日本語「感染」の使用を避けることが考えられる。

### 3.6.2. 日中同形語「感染」と後接する名詞について

BCCWJから日本語「感染」の検索結果を表3にまとめ、《语料庫》から中国語“感染”の検索結果を表4にまとめた。

表3 日本語「感染」と後接するコロケーション（名詞）

共起語	共起頻度	MIスコア	ダイス係数
ウイルス	57	8.6	0.016
人	46	2.74	0.0006
菌	32	8.15	0.0104

表4 中国語“感染”と後接するコロケーション（名詞）

共起語	共起頻度	MIスコア	ダイス係数
病毒	281	13.41	0.031
人	30	7.08	0.0017
菌	7	11.43	0.0102

表3と表4から分かるように、日本語「感染」と後接する「ウイルス」も、中国語“感染”と後接する“病毒”も、比較的強いコロケーションを持っている。

《语料库》から中国語“感染”と後接する“病毒”を調べた結果、以下の例文を得た。

例6：基础区和母本区的种源母株应定期进行系统的鉴定、看有无重新感染病毒。「果树昆虫学 北京农业大学 1980」

例7：因感染病毒的剂量多少、毒力强弱、机体状况及免疫反应是否正常等因素不同、临床表现也不一样。「卫生防疫学 南京医学院 1980」

BCCWJから日本語「感染」と後接する「ウイルス」を調べた結果、以下の例文を得た。

例8：メールを読んだだけで感染するタイプのウイルスには、以下の手順で、プレビューウィンドウを表示させないように設定する方法が有効「朝日ビジネスP.A.S.O（ビズぼそ） 中村南 2002」

例9：ヒトに対して感染性をもつレトロウイルスだったからである。「人体改造の世紀 森健 2001」

例6～例9を見ると、中国語“感染”と後接する“病毒”は、主に“感染病毒”のように直接的に共起するのに対して、日本語「感染」と後接する「ウイルス」は、「感染するタイプのウイルス」のように間接的に共起している場合のあることが分かった。一見、日本語「感染」と後接する「ウイルス」も、中国語“感染”と後接する“病毒”も、比較的強いコロケーションを持っているため、中国人日本語学習者は、誤用が起りにくいと思いがちである。実際に、日本語「感染」と後接する「ウイルス」は、中国語と異なって、間接的に共起する場合があるので、中国人日本語学習者は、母語のコロケーションから負の

影響を受け、「ウイルス」を日本語「感染」の直後に使ってしまおうと考えられる。

以上の予想を検討するために、《语料库》から中国語“感染”と後接する“人”を調べた結果、以下の例文が得られた。

例 10：如今、转到了人民教师的岗位上、看她那强烈的幸福感、多么感染人！「老师的心 传华 1964」

例 11：如能注意声音技巧、并把感情倾注在声音之中、这样说出来的话富有音乐性、就会感染人、使人产生共鸣、收到较好的效果。「汉语口语 陈建民 1984」

実際に、BCCWJから日本語「感染」と後接する「人」を調べた結果、以下の例文を得た。

例12：C型肝炎ウイルスはセックスが主要感染経路と主張している人がいます。「知られざる肝炎 国中るみ子 1992」

例13：地域で百六十五万人に達し、感染者は千八百万人以上にのぼる。「人類サバイバルの条件 吉田康彦 1998」

例10～例13を見ると、中国語“感染”と後接する“人”は“感染人”のように直接的に共起するのに対し、日本語「感染」と後接する「人」は、「感染者は千八百万人以上」のように間接的に共起する場合のあることが分かった。それが原因で、中国人日本語学習者は、母語のコロケーションから負の影響を受け、誤用が起りやすいと考えられる。

表3と表4から分かるように、日本語「感染」と後接する名詞には、接尾辞が多いのに対し、中国語“感染”と後接する名詞は、ほとんど普通の名詞であり、接尾辞が少なかった。そのため、中国人日本語学習者は、日本語「感染」の直後に接尾辞の使用を避けることが予測できる。

以上の予想を検討するために、《语料库》から中国語“感染”と共起する“菌”を調べた結果、以下の例文が得られた。

例 14：这种菌是鳞翅目幼虫天然的病原菌、约有 130 种鳞翅目幼虫容易感染这种菌。「昆虫病理学 南开大学生物系昆虫研教室 1980」

実際に、BCCWJから日本語「感染」と共起する「菌」を調べた結果、次の例文が得られた。

例15：黄色から褐色あるいは赤色を示す感染菌によっては悪臭が強くなる。「母子保健学 海野信也 2003」

例16：体力を消耗するため初回温浴療法には危険を伴うことや感染菌の移動にも配慮する必要がある。「肺炎 渡辺彰 2003」

例 14～例 16 を見ると、日本語「感染」と後接する「菌」は、「感染菌」のように接尾辞として使用されるのに対し、中国語“感染”と後接する“菌”は、“感染这种菌”のように

普通の名詞として使用されることが分かった。それゆえに、中国人日本語学習者は、母語のコロケーションから負の影響を受け、日本語「感染」の後に、接尾辞ではなく、普通の名詞を使用する傾向があると考えられる。

### 3.6.3. 日中同形語「感染」と共起する動詞について<sup>11</sup>

BCCWJ から日本語「感染」の検索結果を表5にまとめ、《语料庫》から中国語“感染”の検索結果を表6にまとめた。

共起語	共起頻度	MIスコア	ダイス係数
受ける	11	8.58	0.005
生じる	5	3.46	0.0008

共起語	共起頻度	MIスコア	ダイス係数
教育	6	7.65	0.0021
影响	6	7.31	0.0018
产生	5	7.1	0.0015

表5と表6から分かるように、中国語“感染”と共起する動詞の中に、“教育”、“影响”などの抽象的な意味を持つ動詞がある。一方、日本語「感染」の有意なコロケーションには、そのような動詞がなかった。よって、中国人日本語学習者は、母語のコロケーションから負の影響を受け、日本語「感染」を過剰に使用することが予測できる。

《语料庫》から中国語“感染”と共起する“教育”と“影响”を調べた結果、以下の例文が得られた。

例17：文学艺术主要是用生动、具体、感人的艺术形象去反映生活、以潜移默化的力量去感染、鼓舞和教育人民。「思想政治 国家教育中学思想政治课教材编写组 1992」

例18：因此教师必须以严肃认真的教学态度、实事求是的科学精神、循循善诱的教学方法、在潜移默化中感染学生、教育学生、使学生的知识、才智和思想都得到不断的提高。「普通教育学 集体 1983」

例19：文学作品具有审美教育作用、在于文学家能把从人民当中所体验的感情传达出来、以感染、影响读者。「高校审美教育手册 郭延萍 1990」

例17～例19を見ると、“教育”と“影响”が中国語“感染”とは共起関係で、同格語でもあることが分かった。よって、中国語“感染”は、同格語の“教育”、“影响”と同様に、動詞として使わなければならない。しかし、BCCWJを調べた結果、日本語「感染」の

有意なコロケーションには、日本語「感染」と同格語の関係を持つ動詞はなかった。さらに、日本語「感染」と共起する動詞を分析した結果、日本語「感染」は、動詞と共起する場合、ほとんど名詞として使用されることが分かった。そのため、中国人日本語学習者は、共起する動詞に影響されて、もともと名詞であるべき日本語「感染」を動詞として使ってしまう傾向があると考えられる。

### 3.6.4. 日中同形語「感染」と共起するその他の品詞について<sup>12</sup>

BCCWJから日本語「感染」の検索結果を表7にまとめ、《语料库》から中国語“感染”の検索結果を表8にまとめた。

表7と表8から分かるように、中国語“感染”と共起する副詞の中には、程度を表す“极”と“最”は、比較的強いコロケーションを持つのに対し、日本語「感染」と共起する副詞には、程度を表すものなかった。そのため、中国人日本語学習者は、母語から負の影響を受け、日本語「感染」と共起する場合に、「きわめて」、「最も」などの程度を表す副詞を過剰に使用する恐れがあると予測できる。

表7 日本語「感染」と共起するコロケーション（その他の品詞）

共起語	共起頻度	MIスコア	ダイス係数
やすい	29	5.01	0.003
多い	25	3.05	0.0007
高い	21	3.53	0.0009
強い	17	3.74	0.001
非常	8	3.55	0.0009
いわゆる	5	3.54	0.0008
重大	5	5.24	0.001
十分	7	3.95	0.001

表8 中国語“感染”と共起するコロケーション（その他の品詞）

共起語	共起頻度	MIスコア	ダイス係数
容易	23	10.52	0.01
多	7	6.48	0.001
严重	10	9.34	0.006
强烈	11	10.54	0.01
极	6	8.32	0.003
高	6	7.29	0.002
最	5	6.59	0.001

以上の予想を検討するために、《语料庫》から中国語“感染”と共起する“极”と“最”を調べた結果、以下の例文が得られた。

例 20：乳状病能否迅速传播、与宿主对这种菌的感染性关系最大、也与土壤中寄主昆虫的虫口密度有关、敏感寄主虫口越稠密、病害的散布也越快。「昆虫病理学 南开大学生物系昆虫研教室 1980」

例 21：这时、梅花鹿温柔善良的眼睛里、伤心的泪水忽然不见了、闪动着勇敢的、正义的、智慧的光辉、使欣欣感染到极大的鼓舞和振奋。「会伤心的画 刘咏 1990」

中国語は、いつも物事をストレートに表現する。よって、例 20 と例 21 のように、中国語には、程度を表す副詞が使用される。それに対して、日本語において、程度を表す副詞がそれほど頻繁に使用されていない。中国人日本語学習者は、日本語と中国語の言語相違に影響され、程度を表す副詞を過剰に使用する恐れがある。

#### 4. まとめ

本稿は、日中同形語「感染」を取り上げて、そのコロケーションについて検討した。その結果、中国人日本語学習者は、母語のコロケーションから負の影響を受け、過剰または過少に使用する恐れがある。その具体的な表れは、以下の通りである。

1. 中国人日本語学習者は、その共起語の品詞に影響され、誤用を起こす可能性がある。
2. 中国人日本語学習者は、日中両言語の言語相違に影響され、誤用を起こす可能性がある。
3. 中国人日本語学習者は、日中両言語が共起種類（間接的・直接的）の違いに影響され、誤用を起こす可能性がある。
4. 中国人日本語学習者は、日中両言語が共起語との関係（同格語・非同格語）の違いに影響され、誤用を起こす可能性がある。

#### 5. 今後の課題

今後の研究は、次を課題として取り組んでいきたい。

1. 今回の研究は、日中同形語「感染」に絞って研究したが、今後、さらに研究対象を増やし、検討していきたい。
2. 今回の研究では、キーワードとなる日中同形語「感染」の前後5語を範囲として、コロケーションを検出した。これから、範囲を縮小し、日中同形語「感染」の連語形式を検討していきたい。
3. 今回の研究では、日中同形語「感染」のコロケーションに基づいて、中国人日本語学習者は、母語のコロケーションからどのような影響を受けるかを予測した。それ

は、データを読み取って、得られた結果である。それを検証するため、調査を実施するつもりである。

以上は、日中同形語に関する基礎研究にすぎない。これからは、このような基礎研究を数多く実施することによって、日中同形語に関する学習のメカニズムを明らかにしたい。

## 引用文献

- 石川 慎一郎 (2006) 「言語コーパスからのコロケーション検出の手法—基礎的統計値について—」『言語コーパス解析における共起語検出のための統計手法の比較研究』 pp. 1-14
- 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—」『言語処理学会第 14 回大会チュートリアル資料』 pp. 40-50
- 胡 春艶 (2012) 「日中同形詞的“感情色彩”探析」『语文学刊』 6号 pp. 4-5
- 小森 和子・三國 純子・徐 一平・近藤 安月子 (2012) 「中国語を第一言語とする日本語学習者の漢語連語と和語連語の習得—中国語と同じ共起語を用いる場合と用いない場合の比較—」『小出記念日本語教育研究会』 20号 pp. 49-60
- 齋藤 俊雄・中村 純作・赤野 一郎 (2005) 『英語コーパス言語学：基礎と実践 (改訂新版)』 研究社出版 p.132
- 新村 出 (1998) 『広辞苑』 第5版 岩波書店
- 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』 21号 pp. 21-31
- 中溝 朋子・坂井 美恵子・金森 由美 (2012) 「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」『大学教育』 9号 pp. 78-85
- 三國 純子・小森 和子・徐 一平 (2015) 「中国語を母語とする日本語学習者の漢語連語の習得—共起語の違いが誤文訂正に及ぼす影響—」『中国語話者のための日本語教育研究』 6号 pp. 34-49

## 参考文献

- 大河内 昭爾 (2012) 『常用語辞典』 成美堂出版
- 大曾 美恵子 (2002) 「コーパスから得られるコロケーション情報—『影響、刺激、感動』を中心に—」『言語文化論集』 2号 pp. 3-12
- 中溝 朋子・坂井 美恵子・金森 由美・大岩 幸太郎 (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』 8号 pp. 88-95
- 中條 清美・内山 将夫 (2004) 「統計的な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

李 宇明 (2008) 《現代汉语常用词表:草案》 商务印书馆

Kjellmer G. (1991). A min of phrases. In K. Ajimer & B. Altenberg (Eds.), English corpus linguistics; Studies in honour of Jan Svartvik pp. 111-127

## 関連アクセス

『現代日本語書き言葉均衡コーパス』 <<http://chunagon.ninjal.ac.jp/>> (2015年11月20日最終閲覧)

《语料库在线》 <<http://www.cncorpus.org/index.aspx>> (2015年11月20日最終閲覧)

北海道大学高見敏子研究室 <http://www.hucc.hokudai.ac.jp/~p16537/ECL/index.html> (2015年10月11日最終閲覧)

## 注

- 『日本国語大辞典』と『デジタル大辞泉』の記述は Japanknowledge によるものである。次のリンクからアクセス可能：<http://japanknowledge.com/library/>
- 石川 (2006) によると、tスコアとは、「統計学から転用された、2つの語の共起関係の統計的有意性を図る指標」であり、「共起の程度が偶然による確率を超えていると、どのぐらいの確かさで言えるかを示す指標」である。その計算式は、以下の通りである。そして、一般に、tスコアが2以上の場合に、そのコロケーションは、傾向性のある組み合わせになっている。

$$t = \left( \text{共起頻度} - \frac{\text{中心語} \times \text{共起語頻度}}{\text{コーパス総語数}} \right) \div \sqrt{\text{共起頻度}}$$

本稿はtスコアの計算は北海道大学高見敏子研究室の自動計算シートを使用する。

(MIスコアとダイス係数は同様)

- 石川 (2006) によると、MIスコアとは、任意の語を与えられたときに、どの程度、その共起語が予測できるかという指標である。そして、齋藤ら (2005) は「MIスコアが低頻度でも意味的な興味深いコロケーションを特定する上で有効である」と指摘している。ただし、MIスコアは「単語の頻度そのものが低い場合には、逆に値が高くなってしまふ」という性質がある。その計算式は、以下の通りである。

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

- 石川 (2008) によると、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そうである。そして、中條 (2004) は、コロケーション検出における9種類の統計値の妥当性を比較した結果、ダイス係数は、精度が最も高いようである。ただし、頻度が低い場合、ダイス係数は、著しく小さい値を示す可能性がある。その計算式は、以下の通りである。

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

- 5) 本稿は、中国語の《現代汉语常用词表(草案)》と日本語の「常用語辞典」に収録している語彙を参照して、その語彙を常用語と見なす。
- 6) 本稿は、BCCWJ と《语料库》が提供している形態素情報を利用する。
- 7) 日中同形語「感染」のコロケーションを検出した結果、日本語においても中国語においても名詞が最も多いことが分かった。よって、名詞グループをさらに前接と後接に分けることにする。
- 8) 本稿では、「ウィルス」と「ウイルス」のような場合によって表記が異なる語彙に対し、その使用頻度は、すべての表記の使用頻度を合計したものを取るようになる。そして、その表記は、最も使用頻度が高いものを取るようになる。
- 9) 本稿で挙げる例文は、キーワードとなる日本語「感染」と中国語“感染”を下線で、コロケーションを網掛けと下線で表記する。
- 10) コーパス検索により、得られた例文に対して、コーパスに掲出している出典情報を加えることにする。その情報は「出典 著者 出版年」という形で表記する。
- 11) 本稿は、日本語と中国語の基本的な語順の相違を考慮して、動詞グループを前接と後接のグループに分けずに、まとめて検討する。
- 12) 本稿は、日本語と中国語の言語相違を考慮して、その他の品詞グループを前接と後接のグループに分けずに、まとめて検討する。