

# 二字名詞日中同形語の共起語について

～「二字名詞漢語+的/の+二字名詞日中同形語」を中心に～

## The co-occurrence of Japanese and Chinese homographs with two-kanji:

### Focusing on two-kanji Sino-Japanese nominals + de/no + two-kanji Japanese and Chinese homographs

何 龍

He Long

#### Abstract

This paper focuses on the co-occurrence of Japanese and Chinese homographs with two-kanji. I used the data from both Japanese corpus and Chinese corpus. I picked the data as follows: two-kanji Sino-Japanese nominals + de/no +two-kanji Japanese and Chinese homographs. As a result I got that there are a lot of two-kanji Japanese and Chinese homographs that have different collocation in Chinese and Japanese. The Chinese Japanese learners should make great efforts to remember that. For the Primary stage learners, they should remember this collocation. For the Intermediate stage learners, they be able to know the reason for the particular structure.

#### 1. はじめに

近年、『現代日本語書き言葉均衡コーパス』（以下、BCCWJ と称する）、《国家語委現代汉语平衡语料库》（以下、《语料库》と称する）のような均衡コーパスが相次ぎ構築されている。従って、均衡コーパスを用いて、計量的な研究が可能となる。しかし、コーパスを用いて、日中同形語についての研究が進んでいなかったのは事実である。本稿は、『日本語教育語彙表』に収録されていた名詞の語彙を対象とし、均衡コーパスのデータを用いながら、統計指標で有意として認めた名詞の共起語を抽出する。中国人日本語学習者（以下、学習者と称する）に役立つ連語情報を論じることを目的とする。

#### 2. 先行研究

##### 2.1. 共起の捉え方

『大辞林』には、「共起」について、次の意味記述があった。「共起とは複数の言語現象が同一の発話・文・文脈などの言語的環境において生起することである」。つまり、ある言

語現象が、ある言語現象と同じ文脈で現れることによって、「共起」という概念は成り立つ。しかし、上記の定義に従えば、「専門学校」のようなものも、「学校の椅子」のようなものも、「共起」となる。本稿は、日本語の連語と中国語の連語の対照に焦点を当てるため、日本語の「名詞<sub>1</sub>+の+名詞<sub>2</sub>」と中国語の“名詞<sub>1</sub>+的+名詞<sub>2</sub>”を「共起」として扱う。つまり、「学校の椅子」、「地球的表面」のような連語のみを研究対象として扱う。

## 2. 2. 日本語と中国語における連語論に関する先行研究

鈴木康之（1994：97）は、大量の用例を集め、日本語のノ格名詞的な連語のむすびつき関係を分類した。その結果、次の3つのむすびつき関係が得られた。

1. 関係的なむすびつき。例：いすの足・友人の時計・息子の学校・蜂子の弟 など
2. 状況的なむすびつき。例：日本の雑誌・今朝の味噌汁・絶交の悲しみ など
3. 規定的なむすびつき。例：子供の前掛・戦争の武器・家政の先生・詩人の伝記 など

一方、中国語の連語のむすびつき関係について、李紹群（2011）は“定中结构”の視点から議論を行った。その結果、次のことが明らかとなった。

1. 領属関係。例：我的书包・公司的老板・隔壁的老王 など
2. 属性関係。例：铁的纪律・木头的房子・绿色的书包 など

さらに、王学群（2016：2）は、鈴木康之（1994、2010）と李紹群（2011）の研究結果に基づいて、日本語の連語と中国語の連語の対応関係について、実例を示しながら、議論していた。その結果、中国語の“領属关系”に属する連語が、日本語の「関係的な連語」に相当し、中国語の“属性关系”に属する連語が、日本語の「規定的な連語」に相当することは明らかとなった。本稿は、王学群（2016）の見解に従う。

## 2. 3. 先行研究の問題点

日中同形語に関する先行研究には、日中同形語そのものに注目したものが多かったのに対し、日中同形語の共起語に関するものはまだ多くない。さらに、何龍（2016）によれば、日中同形語を使用する際に、学習者は、その共起語からも影響を受けているそうである。しかし、何龍（2016）は、日中同形語の「感染」を例として検討した。他の日中同形語も同じ傾向があるかどうか、不明なままである。よって、日中同形語の共起語については、更なる検討が必要となる。

## 3. 本論

### 3. 1. 研究対象の出典について

李在鎬、砂川有里子（2015）の『日本語教育語彙表』に収録された語彙を研究対象として扱う。『日本語教育語彙表』は、17,921項目の見出し語から構成されている。各見出し語には、日本語教師の主観判定に基づく6段階の難易度、旧日本語能力試験の等級、品詞・語種・表記・アクセントなどの情報、語義と用例（作例とコーパスからの実例）が付載されている。

### 3. 2. 統計指標について

日中同形語とその共起語のむすびつき関係に注目するため、統計指標に基づいて、日中同形語の有意な共起語を抽出する。石川慎一郎（2008：46）は、「tスコアは、共起の有意さを判断する統計指標として、しばしば使用されている。」と指摘している。tスコア<sup>1</sup>は、コーパス総語数を考慮し、全体における個々の語の出現比率を比較する。頻繁に用いられる一般性の高いコロケーションの評価に適している。一般に、tスコアは、2以上の場合に意味のある組み合わせであるとされる。本稿は、tスコアの数値に基づいて、日中同形語の有意な共起語を抽出する。

### 3. 3. コーパスについて

BCCWJは、収録した語数が約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、ブログなどの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を母集団として、そこから無作為に抽出したものである。そして、BCCWJを検索するために、「中納言」<sup>2</sup>というコーパス検索アプリケーションを利用する。

《语料库》は、収録した語数が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”などの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、新聞などから無作為に抽出したものである。現在、“语料库在线”というサイトを通じて、日本からアクセスすることができる。

### 3. 4. 分析の手順

次の手順に従って、分析を行う。

1. 『日本語教育語彙表』により、二字の日本語漢語を抽出し、その二字の日本語漢語を前文字と後文字<sup>3</sup>に分け、それぞれを『常用漢字表』にあてはめる。前文字と後文字は共に『常用漢字表』に属するもの、計7390語に絞る。
2. 7390語の二字の日本語漢語を文化庁(1978)のS語<sup>4</sup>に属するもの、計973語に絞る。
3. 973語の二字の日本語漢語を中国語の簡体字に変換<sup>5</sup>し、前文字と後文字に分け、《現代汉语通用字表》にあてはめる、前文字と後文字は共に《現代汉语通用字表》に属するもの、計964語に絞る。
4. 964語の日中同形語を『現代日本語書き言葉均衡コーパス語彙表』<sup>6</sup>と《現代汉语语料库分词类词频表》の品詞情報に基づいて、名詞であるもの、計502語に絞る。
5. 502語の日中同形語に、tスコア<sup>7</sup>の数値に基づいて、それぞれ日本語「の」と、中国語“的”と、前方に共起することが有意かどうか判断する。日本語「の」と、中国語“的”と、前方共起が有意として認めたもの、計37語<sup>8</sup>に絞る。
6. 37語の日中同形語に、BCCWJと《语料库》で検索し、その前方共起語を集める。そして、37語の日中同形語とその前方共起語に、tスコアを計算し、有意な前方共起語をまとめ、共起頻度と使用頻度<sup>9</sup>を統計する。それぞれ共起頻度が上位3語の共起語

に絞る。その結果、日本語の82語で、中国語の20語が得られた。

### 3. 5. 結果の考察

#### 3. 5. 1. 結果の概要

日本語「の」と、中国語“的”と、有意に共起する日中同形語は、次の37語に絞った。

方法、結果、理由、報告、裁判、金融、様子、全部、表情、印象、航空、図書、態度、提案、証拠、信用、結論、通知、生理、見解、勇気、重点、中華、気象、地理、意向、体温、使命、指令、教訓、功績、情景、愛好、模範、手腕、見聞、宝庫<sup>10</sup>

上記の37語と有意として認めた前方共起語をまとめた。その結果、日本語連語が82語、中国語連語が20語<sup>11</sup>に絞った。具体的な結果を次のようにまとめた。

日本語の連語（計82語）<sup>12</sup>

唯一の方法、管理の方法、評価の方法、調査の結果、検査の結果、実験の結果、最大の理由、本当の理由、相当の理由、虚偽の報告。状況の報告、結果の報告、日本の裁判、事件の裁判、日本の金融、民間の金融、世界の金融、部屋の様子、当時の様子、生活の様子、事務の全部、業務の全部、事業の全部、困惑の表情、苦悶の表情、余裕の表情、最初の印象、全体の印象、自分の印象、日本の航空、全国の航空、学校の図書、大学の図書、地域の図書、自分の態度、政府の態度、相手の態度、最初の提案、自分の提案、日本の提案、唯一の証拠、事件の証拠、調書の証拠、企業の信用、会社の信用、顧客の信用、会議の結論、前項の通知、目的の通知、招集の通知、人間の生理、女性の生理、教授の見解、政府の見解、自分の見解、運動の重点、施策の重点、財源の重点、日本の気象、各地の気象、地域の地理、日本の地理、住民の意向、国民の意向、業者の意向、人間の体温、自分の体温、自分の使命、本来の使命、教育の使命、震災の教訓、事故の教訓、歴史の教訓、最大の功績、生前の功績、自分の功績、周囲の情景、家族の情景、監督の手腕、情報の宝庫、食材の宝庫、栄養の宝庫<sup>13</sup>

中国語の連語（計20語）

科学的方法、实验的方法、人工的方法、草案的报告、情况的报告、工作的报告、脸上的表情、所得的印象、所得的结论、错误的结论、下面的结论、人体的生理、机体的生理、植物的生理、斗争的勇气、工作的重点、历史的使命、历史的教训、眼前的情景、当时的情景、

上記の結果から、日本語と中国語において、異なる共起語を取る日中同形語は圧倒的に多いため、学習者は母語の知識を転用できなくなる。すると、初級の学習者に連語そのまま覚えさせる必要があると考えられる。

一つの特別な例があった。それは、同じ前方共起語を取る「歴史の教訓」と“历史的教訓”であった。

《语料库》で“历史的教訓”を検索した結果、次の例1が見つかった。

例1: 可见当代的教育家不管是自觉的或是不自觉的, 都要顾及到两派的长处, 这是历史的教訓<sup>14</sup>。  
「毛祖桓、从方法论看教育学的发展」<sup>15</sup>

BCCWJで「歴史の教訓」を検索した結果、次の例2が見つかった。

例2: 我々は、過去の戦争についての歴史観の相違を超え、歴史の教訓を謙虚に学び、平和な国際社会を築いていかなければならない。「文芸春秋、戦後50年日本人の発言」

例1の“历史的教訓”において、“历史”と“教訓”は、“属性关系”のむすびつき関係を持ち、例2の「歴史の教訓」において、「歴史」が「教訓」とは、「規定的な連語」のむすびつき関係を持つ。王学群(2016)の見解によると、“历史的教訓”の“属性关系”のむすびつき関係と「歴史の教訓」の「規定的な連語」のむすびつき関係が対応できる。しかし、その部分は、数が少ないので、学習者は注意しない可能性がある。すると、学習者にその部分の存在を気付かせる必要があると考えられる。

一方、有意として認めた日本語の連語の数は中国語の連語のおよそ4倍であった。なぜそのような結果が出たか。それは、日本語のノ格名詞的な連語は、「の」でつながらなければならないのに対し、中国語の連語は、“的”を使っても、使わなくてもよいからと考えられる。李紹群(2011:19)によれば、中国語には、“中国的桐油”と“中国桐油”が同時に存在するが、日本語は、中国語と違って、ノ格名詞的な連語には、「の」が必須である。それで、日本語のノ格名詞的な連語を使用する際に、学習者は、名詞と名詞を繋ぐ「の」を省略する恐れがあると考えられる。

### 3. 5. 2. 日本語の連語と中国語の連語の対照

鈴木康之(1994)と李紹群(2011)による連語のむすびつき関係の判断基準に基づいて、日本語の連語と中国語の連語のむすびつき関係について、分析を行った。その結果を次の表1と表2にまとめた。

表1 むすびつき関係<sup>16</sup>の割合(研究対象全体)

むすびつき関係	日本語連語	中国語連語
関係的な連語・領属关系	59語(72%)	13語(65%)
規定的な連語・属性关系	23語(28%)	7語(35%)
合計	82語(100%)	20語(100%)

表2 むすびつき関係の割合（日中重なる10語）<sup>17</sup>

むすびつき関係	日本語連語	中国語連語
関係的な連語・領属関係	14語（54%）	6語（32%）
規定的な連語・属性関係	12語（46%）	13語（68%）
合計	26語（100%）	19語（100%）

表1から、日本語においても、中国語においても、「関係的な連語・領属関係」は、「規定的な連語・属性関係」より多く、「関係的な連語・領属関係」が全体の7割近くを占め、「規定的な連語・属性関係」が全体のおよそ3割を占める。

《语料庫》で“属性関係”のむすびつき関係を持つ“科学的方法”を検索した結果、次の例3が見つかった。

例3：只有这样一个问题，才是一个具有科学性因而可以用科学的方法（主要是实验、数学和逻辑推理方法）来解决的问题。 「曾近义，徐天芬、自然辩证法总论」

BCCWJで「規定的な連語」のむすびつき関係を持つ「唯一の方法」を検索した結果、次の例4が見つかった。

例4：この当時は、どれだけ認識されていたか怪しいが、ここで浮上した円安・ドル高政策は、まさに日米双方のニーズに合致する唯一の方法だった可能性があったのである。 「吉田恒、さよなら円高」

例3の“科学的方法”で、“科学”と“方法”は、“属性関係”のむすびつき関係を持ち、例4の「唯一の方法」で、「唯一」と「方法」は、「規定的な連語」のむすびつき関係を持っている。“科学的方法”と「唯一の方法」において、日中同形語「方法」は、異なる共起語の“科学”、「唯一」を取るが、対応できる“属性関係”と「規定的な連語」のむすびつき関係を持つため、学習者が中国語との違いに気づきやすいと考えられる。

表2から分かるように、日本語の連語において、「関係的な連語・領属関係」も、「規定的な連語・属性関係」も、ほぼ5割を占める。一方、中国語の連語にはおいて、「関係的な連語・領属関係」が全体のおよそ3割を占めるが、「規定的な連語・属性関係」が全体の7割近くを占める。それは、日中同形語は、日中両言語において、異なるむすびつき関係でその共起語と繋がる可能性が示された。それで、中級及び上級段階の学習者に日本語の連語と中国語の連語との違いを気づかせる必要がある。さらに、そのむすびつき関係についての説明は必要であると考えられる。

《语料庫》で“領属関係”のむすびつき関係を持つ“错误的结论”を検索した結果、次の例5が見つかった。

例5:如何挑取与决定尺度,这是一个很重要的也是很关键的问题,尺度取得合适,就能得出符合于真实物理过程的结果,否则将会导致不精确以至于错误的结论。

〔伍荣生、动力气象学〕

BCCWJで「規定的な連語」のむすびつき関係を持つ「会議の結論」を検索した結果、次の例6が見つかった。

例6:単なる会議だけでなく、会議の結論を受けて、アフリカの五十二カ国の中から適当な国を選び、私の考えを實際行動に移す計画である。〔笹川良一、人類みな兄弟〕

例5と例6から、日中同形語「結論」は、異なるむすびつき関係で、異なる共起語と共起することが分かった。学習者は、母語との違いに注意しながら、日本語の連語とその共起語を学習しなければならないので、その学習が決して容易ではない。その部分の学習方法として、次のことが挙げられる。

1. 初級段階の学習者に、その部分の日本語の連語をそのまま覚えさせた方が効率よく学習できる。
2. 中級及び上級段階の学習者に、日本語の連語論を詳しく説明し、理解させた方がより自然な日本語連語の産出することに繋がる。

#### 4. おわりに

「二字名詞漢語+的/の+二字日中同形語」という表記形式を持つ二字の日中同形語とその有意な前方共起語について検討した。その結果、学習者にとって、連語の学習は、非常に重要であることが明らかとなった。より自然な日本語を身につけるために、初級段階の学習者に、日本語の連語をそのまま覚えさせる必要のあることが示唆された。さらに、中級及び上級段階の学習者に、日本語の連語論についての説明が不可欠である。具体的には、次のことが明らかとなった。

1. 二字の日中同形語は、日中両言語において、前方に異なる共起語を取るものが多い。
2. 数が多くないが、二字の日中同形語は、前方に同じ共起語と同じむすびつき関係で繋がるものがある。
3. 前方に異なる共起語を取る二字の日中同形語は、同じむすびつき関係を持てば、学習者はその違いに気づきやすい。
4. 二字の日中同形語が異なるむすびつき関係で、異なる共起語と共起する場合は、学習者は気づきにくく、その部分の学習が容易ではない。

本稿は、日中同形語に関する基礎研究のひとつにすぎない。これからは、学習者による作文の実例を示しながら、後方共起語や前方と後方の両方に共起する連語なども視野に入れて、検討するつもりである。

本稿は、国際連語論学会第4回大会で口頭発表した内容をまとめたものである。

## 参考文献

- 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:統計数理研究所共同研究レポート』 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
- 王学群 (2016) 「連語とカテゴリー化について―日中対照の観点から―」『国際連語論学会第4回大会シンポジウム資料』 p. 2
- 何龍 (2015) 「日中同形語の連語形式について―「感染」を例として―」国際連語論学会 2015年7月例会発表資料
- 何龍 (2015) 「日中同形語の品詞の違いによる誤用について―中国人の日本語学習者を対象として―」『第8回コーパス日本語学ワークショップ予稿集』 pp. 1-10
- 何龍 (2016) 「日中同形語のコロケーション―「感染」を例として―」『愛知淑徳大学論集―グローバルカルチャー・コミュニケーション研究科篇』8号 pp. 49-63
- 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
- 鈴木康之 (1994) 『現代日本語の名詞的な連語の研究』日本語文法研究会 pp. 98-104
- 鈴木康之 (2011) 『現代日本語の連語論』日本語文法研究会 pp. 93-104
- 文化庁 (1978) 『中国語と対応する漢語』 pp. 147-180
- 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
- 李绍群 (2011) 《现代汉语“名<sub>1</sub>+ (的) +名<sub>2</sub> 定中结构研究》厦门大学出版社 pp. 34-62

## 関連アクセス

- 『三省堂 大辞林』 weblio 辞書  
 <<http://www.weblio.jp/content/%E5%85%B1%E8%B5%B7>>, (2015年12月12日参照)
- 『現代日本語書き言葉均衡コーパス』 コーパス検索アプリケーション  
 <<https://chunagon.ninjal.ac.jp>>, (2015年12月12日参照)
- 《国家語委现代汉语通用平衡语料库》 语料库在线  
 <<http://www.cncorpus.org/index.aspx#P0>>, (2015年12月12日参照)



『日本語教育語彙表』

<<http://jreadability.net/jev>>, (2015年12月12日参照)

「楽訳中国語」 中国語簡体字変換ツール

<[http://www.jedic.com/chinese\\_convert](http://www.jedic.com/chinese_convert)>, (2015年12月12日参照)

『常用漢字表』(2010年改定版) (2016年4月12日参照)

<[http://kokugo.bunka.go.jp/kokugo\\_nihongo/joho/kijun/naikaku/kanji/](http://kokugo.bunka.go.jp/kokugo_nihongo/joho/kijun/naikaku/kanji/)>,

《現代汉语通用字表》

<<http://www.cncorpus.org/Resources.aspx>>, (2016年4月12日参照)

## 注

- 1) 石川慎一郎 (2008:46) によると、tスコアの計算式は次の通りである。

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

- 2) 「中納言」は、国立国語研究所で開発したコーパスを検索することができる Web アプリケーションである。短単位・長単位・文字列の3つの方法によって、コーパスに付与された形態情報を組み合わせた高度な検索ができる。
- 3) 二字の漢語の前文字と後文字を次の通りに扱う。「生活」は、前文字が「生」で、後文字が「活」である。以下、同様。
- 4) 文化庁 (1978) は、日中同形語を S 語 (意味が同じか、極めて近いもの)、O 語 (意味が一部かさなってはいるが、両者の間にずれのあるもの)、D 語 (意味が著しく異なるもの) と N 語 (日本語の漢語と同じ漢語が中国語に存在しないもの) の4つのタイプに分けた。名詞の日中同形語の共起語に焦点を当てるため、研究対象を文化庁 (1978) の S 語に限る。
- 5) 「楽訳中国語」というツールを使用し、日本語の漢字から中国語の簡体字に変換する。ただし、日本語の漢字と対応する中国語の簡体字が複数ある場合、筆者の手作業により、最も頻繁に使用される中国語の簡体字に変換する。
- 6) 『現代日本語書き言葉均衡コーパス語彙表』、《現代汉语语料库分词类词频表》は、品詞情報は精度が95%以上に達しているため、その品詞情報を利用することにする。
- 7) tスコアの計算は、高見敏子の計算シートを利用する。高見敏子の計算シートは、次のリンクからアクセスできる。  
<<http://www.hucc.hokudai.ac.jp/~p16537/ECL/index.html>>, (2016年5月1日参照)
- 8) 「二字名詞漢語+的/の+二字名詞日中同形語」という表記形式を持つ連語について検討するため、日中同形語は「的/の」と有意に共起しなければならない。
- 9) 共起頻度とは、同じコーパスにおいて、キーワードとなる語彙とその共起語が同じ文脈において同時に現れる回数である。使用頻度とは、コーパスにおいて、キーワードとなる語彙が現れる回数である。
- 10) BCCWJ においての使用頻度の降順に並び替え、日本語のみ表記する。
- 11) 日本語教育の視点から、学習者に役立つ連語情報の提供を目指すため、使用頻度が高いものに、その共起語を抽出する。有意な共起として認めたものは、それぞれ使用頻度が上位3語に絞る。
- 12) 有意として認めた連語は、単独に意味を持たなければならない。つまり、「理由の通

知」(tスコア=2.41)のようなtスコアで有意となるが、連語が独自に成立しないため、研究対象から外す。

- 13) 本稿では、『現代日本語書き言葉均衡コーパス語彙表』と《現代汉语语料库分词类词频表》に収録された品詞情報は名詞であるものを名詞と見なす。以下、同様。
- 14) コーパスの検索で、キーワードとなる語彙に下線をつける。以下、同様。
- 15) コーパスによる例文に、作者と出典の情報を加える。以下、同様。
- 16) 王学群(2016)の理論に従って、中国語の“領属关系”とそれに相当する日本語の「関係的な連語」を「関係的な連語・領属关系」と称する。“属性关系”とそれに相当する「規定的な連語」を「規定的な連語・属性关系」と称する。
- 17) 分析結果として、日中両言語において、有意として認めた前方共起語を持つものは次の10語である「方法、報告、表情、印象、結論、生理、重点、使命、教訓、情景」。以下、その10語を「日中重なる10語」と称する。