

ISSN 2186 – 3989

コロケーションの視点から見る中日
「同形同義語」の相違性

叶 栩 邑

A study on the differences of “Isomorphic Synonyms” between Chinese
and Japanese from the perspective of collocation

Xuyi Ye

北 陸 大 学 紀 要
第52号(2022年3月)抜刷

コロケーションの視点から見る中日 「同形同義語」の相違性

叶 栩邑*

A study on the differences of “Isomorphic Synonyms” between Chinese and Japanese from the perspective of collocation

Xuyi Ye*

Received December 13, 2021

Accepted January 19, 2022

Abstract

As is known to us, there are numbers of homographs in Chinese and Japanese languages, and these words can be, based on their meanings, classified into isomorphic synonyms, homonyms and isomorphic near synonyms. They are playing a special role in the Contrastive Study of Chinese and Japanese because of their similarities in form and conceptual meaning, and their advantages in quantity and proportion. Especially Isomorphic Synonyms, because of their highly similar conceptual meaning, we can not fully and objectively grasp the differences between Chinese and Japanese only by means of dictionaries and introspection. This study uses the corpus to investigate the differences of Isomorphic Synonyms between Chinese and Japanese from the perspective of collocation. At the same time, it also summarizes the investigation methods, steps and precautions in operation from three aspects:1) Constraints on syntagmatic relation and combination, 2) Analysis of co-occurrence words, 3) Combined strength and characteristic collocation. Through these practical investigations and the summary of methodology, it is expected to be helpful to the understanding and use of bilingual users and the front-line educational activities of Japanese linguistics.

Key Words : Isomorphic Synonyms, collocation, Chinese and Japanese Homographs

*北陸大学国際交流センター(天津外国語大学交流教員) International Exchange Center, Hokuriku University

はじめに

中日国交正常化に伴って、中日両言語は同じ漢字文化圏に属しているため、漢字使用という点が注目の的となっている。特に、中日同形語はその形態上の類似性により、特殊的な語彙グループとして次第に関心が寄せられ、中日対照研究の重要な分野の一つとなっている。その下位分類としては、現在同形同義語・同形類義語・同形異義語という三種類に分けることが一般的となっている。

文化庁(1978)は、中日同形語を①S類(Same)中日両国語における意味が同じか、または、極めて近いもの、②O類(Overlap)意味が一部重なってはいるが、両者の間にずれのあるもの、③D類(Different)意味が著しく異なるものという三種類¹に分け、その後の分類調査に広く影響を及ぼしている。分類調査の統計データによると、中日同形語のうち、「同形同義語」(S類)が最も多く、全体の7割~9割を占め、絶対的な優位性を保っている(文化庁1978、曾根1988、西川1991、橘1994、王2001)。一方、各分類調査のほとんどは辞書を頼りに分類作業を行い、客観的・明確な物差しに見えるものの、両語間の概念的意味の対照比較に止まり、分類結果にも多くのずれが生じている。

近年来、中日同形語をめぐる研究は伝統的な言語学の枠を超え、コーパス言語学や日本語教育学、翻訳学などの視点から多面的、多角的に考察する傾向が見られる。語義の精密な分析と記述において、単語よりもっと拡張的な単位を考察対象にするのは曖昧性が解消しやすく、もっと適切であるという主張も表れている(Sinclair1991,1996ab,2004;卫2012,2015)。そもそも形態・概念上に高度な類似性を持ち、使い分けが非常に難しい中日「同形同義語」に対して、辞書頼りや内省のみで、対象語の相違性或使用実態を網羅的に把握することは無理がある。それを踏まえて、本研究はコロケーションの視点から、対象語の使用実態に目をつけ、中日「同形同義語」の相違性をまとめている。

1. 先行研究

中日同形語は形態・音韻上、ひいては意味用法の類似性により、特徴的な語彙グループとして、中日対照研究においてよく取り上げられるテーマの一つとなっている。これまで、中日同形語をめぐる対照研究の多くは意味用法の相違点に注目し、実例を通して分析を行っている(大河内1992、曲1995、潘1995、王2004など)。しかし、中国語を第一言語とする日本語学習者にとって、日本語はともかく母国語である中国語の場合でも、内省で対象語の使用実態や特徴を網羅的に把握することは無理がある。近年来コーパス言語学の発展とともに、中日同形語の対照研究も作例や限られた例文の収集・分析からコーパス利用でのデータ処理に転換し、大規模な量的分析によって、より客観的に使用実態を把握し、それぞれの特徴や相違点などをまとめている。施(2019)は意味、品詞、文法機能、コロケーションという側面から中日同形語の相違性をまとめ、特にコロケーションという立場から実態調査を把握することの重要性を指摘している。

「コロケーション」という概念を最初に提出したのはFirth(1957)で、「慣習的に共に用いられる(2語以上の)語」と定義し、記述言語学の一つの研究領域としてコロケーション研究の重要性を唱えている。日本におけるコロケーション研究は、主に欧米の関連研究を参考に発展してきたものである。結合や連語、慣用句などの用語を使ったこともあるが、コロケーションとそれぞれ強調する部分や使用される分野が異なっている。堀(2009)は語彙間だけではなく、より広い範囲から関連性を重視して、コロケーションを「語と語の間に

おける、語彙、意味、文法などに関する習慣的な共起関係」(堀 2009 : 7) と定義している。

これまで中日同形語に関する対照研究は関心度が高まる一方、現時点ではまだ幾つかの問題点が残されている。従来の研究は主に中日同形類義語に目を付け、「同形同義語」を対象として取り扱われる研究が少ない。しかしながら、語彙グループ内に占める割合からも、使い分けの難易度や誤用・誤訳率からも、形態・概念上高度な類似性を持つ「同形同義語」の相違点に着目し、さらに詳しく分析を行う必要がある。研究の立場や視点においては、言語学における意味論の研究は、大半「意味は何を指しているか」や「意味構造の記述」などの課題解決を目的とすることで、二つの言語間での意味用法の対照という課題に対し、理論的にも方法論的にも系統性が欠けているのは事実である。特に、中日「同形同義語」の場合、内省や辞書のみではその相違性を把握することが非常に限られているため、コーパスを利用することで、コロケーションの視点からもっと全面的かつ客観的に言語事実に潜まれる特徴や相違点を捉えるのが求められている。

2. コロケーション調査の手順と問題点

中日「同形同義語」の相違点を明らかにすることを目的とするコロケーション調査は、コーパスにおける膨大なデータから有意なコロケーションを抽出し、共起ルールを生成するメカニズムを解明するのが重要である。以下においては①使用するコーパス及びソフト、②調査の手順とコーパス利用の問題点を紹介する。

2.1 使用するコーパス及びソフトについて

本研究はコーパスとコンコーダンス・ソフトを選定する際に、規模・正規性、または学習者が入手しやすい一般公開型ということから考慮して、以下の二種類に絞っている。

日本語の場合、NLB(NINJAL-LWP for BCCWJ バージョン 1.40)を利用する。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』(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。オンライン・コンコーダンスとして国語研では BCCWJ 用の中納言と少納言も公開されているが、NLB は異なるレキシカルプロファイリングという手法を使用し、内容語の共起関係や文法的振る舞いを網羅的に表示できるのがその最大の特徴とされている。以下は NLB のサブコーパスの構成状況一覧である。

表 1 NLB サブコーパスの構成状況

サブコーパス	媒体	記号	語数 (語)
出版	書籍	PB	29331243
	雑誌	PM	4626162
図書館	書籍	LB	31491513
特定目的	ベストセラー	OB	4139668
	知恵袋	OC	11329919
	法律	OL	938198
	国会会議録	OM	5086855
	広報紙	OP	4108480
	教科書	OT	1041844
	韻文	OV	232508
	白書	OW	4766280
	ブログ	OY	12018522
総語数			109102192

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うえで解析している。その際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

係り受け解析 CaboCha 0.64

具体的な検索結果としては、NLB はグループ別とパターン頻度順という二種類の並び方がある (図 1 と図 2 を参照)。

The screenshot displays the NLB search results interface. On the left, there is a sidebar with a search filter set to '頻度=382'. The main area is divided into two panels. The top panel, titled 'グループ別' (Group-wise), shows a table of patterns and their frequencies. The bottom panel, titled 'パターン頻度順' (Pattern frequency), shows a list of patterns and their frequencies. The right panel shows a detailed view of a specific pattern, including its frequency and a list of usage examples with their respective frequencies.

パターン	頻度	比率
○...が享受する	72	
○...は享受する	18	
○...も享受する	5	
○...の享受する	1	
○...を享受する	284	
○...へ享受する	1	
○...に享受する	39	
○...と享受する	1	
○...で享受する	19	
○...から享受する	3	
○...まで享受する	4	
○...において享受する	1	
○...より享受する	1	

パターン	頻度	比率
○...として享受する	8	
○...において享受する	3	
○...にわたって享受する	1	

コロケーション	頻度	Mt	LD
者が享受する	7	5.71	0.18
国民が享受する	6	8.87	3.33
【地域】が享受する	5	3.03	-2.50
我々が享受する	5	9.12	3.59
人々が享受する	4	8.40	2.86
人が享受する	4	4.30	-1.23
全てが享受する	3	7.13	1.59
ら	3	6.44	0.90
たちが享受する	3	5.45	-0.08
住民が享受する	2	8.38	2.84
子供が享受する	2	5.75	0.22
施設が享受する	1	6.54	1.00
業者が享受する	1	7.68	2.14
権が享受する	1	6.35	0.82
海軍が享受する	1	9.01	3.45
男女が享受する	1	8.22	2.67
画が享受する	1	8.05	2.50
社会が享受する	1	5.11	-0.43
自分が享受する	1	3.84	-1.70
輸送が享受する	1	8.57	3.01
達が享受する	1	7.36	1.82

右側の詳細ビューには、検索されたパターン「者が享受する」に関する文脈が示されています。例として、「参議院での附帯決議に盛り込まれました、障害者が著作物を享受する機会の確保という事です。」や「いま購入者が享受している高い利回りは、本来ならば過去の購入者が受け取れるはずだった利益。」などが挙げられています。

図 1 NLB における検索結果の表示画面 (グループ別)



図2 NLBにおける検索結果の表示画面（パターン頻度順）

現段階では中国語を対象に精度と完成度が高い大規模な解析済みデータが蓄積され、しかも一般公開されたコーパスはそれほど多くない。本研究では、NLBに相当する代表性と正規性を持つ“国家語委現代汉语通用平衡语料库”（『国家言語委員会現代中国語通用均衡コーパス』以下CNC）を利用することとしている。現在コーパスの総字数は1億字で、作成当時の構成状況は以下の通りである（表2を参照）。

表2 作成当時のCNC構成状況

種類	下位分類	字数（万字）
テキスト	大・中・小学校のテキスト	2000
人文と社会科学	政治・法律関連（政治・法律・哲学・宗教など）	3000
	歴史（民族関係も含む）	
	社会（社会学・心理・言語・教育・文芸理論・新聞学・民俗学など）	
	経済	
	芸術（音楽・美術・舞踊・演劇など）	
	文学（口語類も含む）	
	軍事類	
自然科学	農業・医学・工程と技術関連	300
新聞雑誌	新聞	900
	雑誌	400
応用文	政府文書・手紙・説明書・広告など	400
初期段階の総字数		7000（現在1億）

CNCは基本的な検索と分析機能は搭載しているが、すべてのデータが形態素解析済みのデータではなく、係り受け解析も行われていない状態である。しかし、日本語と対照し、

コロケーションの実態調査を行うためには、さらなる解析が必要とされ、本研究ではコンコーダンス・ソフトウェア AntConc(バージョン 3.2.1)をツールとして利用し、CNC の検索結果を二次分析している。

さらに、AntConc の利用で KWIC 検索 (Key Word in Context の略)、つまり対象語の共起状況を確認することができる。しかも「Collocates」の機能を利用して、スパン²(span)設定後の共起語リストの作成も可能である。

Hit	KWIC	File
1	天上/nl 的/u 火/n 愉快/v 延格/v 凡/n/n、/w 使/v 他们/r	享受_xls.txt 0 4
2	/c 人的/u 缺乏/v 民主/a 素质/n 和/c 不/d 能/vu 充分/a	corpuz_享受_xls.txt 0 5
3	物/n 的/u 相称/v、/w 是/vl 从/p 自/r 个性/n 体内/nl	corpuz_享受_xls.txt 0 6
4	落后/a 现象/n 的/u 存在/v、/w 都/d 使得/v 少数民族/n 在/p	corpuz_享受_xls.txt 0 1
5	/u、/w 在/p 当时/nt 的/u 历史/n 条件/n 下/nd、/w 能/vu	corpuz_享受_xls.txt 0 2
6	/v 众/n 甘整/n、/w 深表/v 同情/v、/w 而/c 不/d 能/vr	corpuz_享受_xls.txt 0 3
7	看/n 分之/u 十年 前	corpuz_享受_xls.txt 0 3
8	/u 补贴/n、/w 31	corpuz_享受_xls.txt 0 3
9	1 退休/v 干部/v	corpuz_享受_xls.txt 0 4
10	等教育/n 自学考试/n 专科/n 毕业生/n	corpuz_享受_xls.txt 0 3
11	/n、/w 有/v 五/m 名/q 考试/v 不/d 及格/v、/w 仍/d	corpuz_享受_xls.txt 0 8
12	/d 率/v 群/n 斗争/n、/w 对/n/d 排除/v 阻碍/n、/w 也/d	corpuz_享受_xls.txt 0 1
13	/u 借以/n 剔/c 剔除/p 其/r 本身/n 的/u 等级/n 而/c	corpuz_享受_xls.txt 0 2
14	里/nd、/w 少数族群/n 公民/n 而/c 深表/vs 公民/n 一样/a 都/d	corpuz_享受_xls.txt 0 3
15	一定/a 的/u 义务/n、/w 在/p 产品/n 知识/n 上/nd 在/r	corpuz_享受_xls.txt 0 3
16	/c 知照/n、/w 地/r 能/v 够/v 够/n、/w 恩德/d 地/u	corpuz_享受_xls.txt 0 3
17	n、/w 286 因此/c、/w 素质/n	corpuz_享受_xls.txt 0 3
18	/k 只/d 可/vu	corpuz_享受_xls.txt 0 3
19	/c 那/r 条/n 街上/nl 的/u 住户/n 根本/d 没有/v 力量/n	corpuz_享受_xls.txt 0 4
20	/v 它/r 的/u 计划/n 与/c 360 称	corpuz_享受_xls.txt 0 5
21	并挂/c	corpuz_享受_xls.txt 0 2
22	/v) /w 好/a 了/u、/w 现在/nt 我们/r 可以/vu 尽量/d 地/u	corpuz_享受_xls.txt 0 4
23	/n 欣赏/v 著作/v 是/vl 群众/n 接受/v 革命/a 教育/n 和/c	corpuz_享受_xls.txt 0 2
24	nd 本队/d 劳累/a 多/a、/w	corpuz_享受_xls.txt 0 1
25	/a、/w 有的/r 较/d 贫困/a、/w 造成/v 青少年/j 的/u 物质/n	corpuz_享受_xls.txt 0 1
26	/n 自身/n 的/u 逻辑/n 结构/n 中/nd、/w 刚强/v 发挥/v、/w	corpuz_享受_xls.txt 0 2
27	所以/c 我们/r 应该/vu 非	corpuz_享受_xls.txt 0 3
28	知/c 贡献/d	corpuz_享受_xls.txt 0 3
29	/vd 的/u 人/n、/w 甚至/d 是/v 公共汽车/n、/w 也/d 要/vu	corpuz_享受_xls.txt 0 4
30	享受 /v 文明/a 与/c 幸福/a 生活/n、/w 458 悲剧/n 本来/d 是/vl	corpuz_享受_xls.txt 0 4
31	享受 /v 民主/a 权利/n、/w 所以/c 在/p 民主/a 生活/n 中/nd 战争/n	corpuz_享受_xls.txt 0 5
32	享受 /v 平等/a 权利/n、/w 60 在/p 等级/v 差别/n 的/u 战争/n	corpuz_享受_xls.txt 0 6
33	享受 /v 平等/a 权利/n 方面/n、/w 必然/a 要/vu 受到/v 事实上/nl 的/	corpuz_享受_xls.txt 0 1
34	享受 /v 良好/a 学校/n 教育/v 的/u 只能/vu 是/vl 统治阶级/n 的/u	corpuz_享受_xls.txt 0 2
35	享受 /v 权利/n、/w 235 那/c 观念/n 首先/d 看到/v 形象/n、/	corpuz_享受_xls.txt 0 2
36	享受 /v 那/a 单位/n 在/vl 干部/n 一切/r 管理/v 福利/n 待遇/n、/	corpuz_享受_xls.txt 0 3
37	享受 /v 高等/a 学校/n 专科/n 毕业生/n 的/u 待遇/n、/w 356 可以/c	corpuz_享受_xls.txt 0 3
38	享受 /v 初级/a 工资/n 待遇/n、/w 有/v 各/q 无放/d 不/d	corpuz_享受_xls.txt 0 4
39	享受 /v 最好/a 的/u 东西/n、/w 38 可以说/v、/w 这/r 种/q 甜糖/	corpuz_享受_xls.txt 0 3
40	享受 /v 不同/a 的/u 称号/n、/w 知道/n (/w mai /ws) /w、/w 即/)	corpuz_享受_xls.txt 0 8
41	享受 /v 平等/a 的/u 选举权/n 和/c 被选举权/n、/w 97 因为/c 群众/n	corpuz_享受_xls.txt 0 1
42	享受 /v 一定/a 的/u 权利/n、/w 因而/c 对/p 双方/n 都/d 是/	corpuz_享受_xls.txt 0 1
43	享受 /v 天然/a 的/u 产物/n、/w 每逢/v 转弯/nt 之/u 季/nt、/w	corpuz_享受_xls.txt 0 2
44	享受 /v 卫生/a 的/u 设备/n、/w 所以/c 地下/nl 根本/d 没有/	corpuz_享受_xls.txt 0 3
45	享受 /v 同样/a 的/u 权利/n、/w 承担/v 同样/a 的/u 义务/a、/w 36	corpuz_享受_xls.txt 0 3
46	享受 /v 幸福/a 了/u、/w 364 想/v —/w 想/v 那些/r 可怜/a 的/	corpuz_享受_xls.txt 0 3
47	享受 /v 有趣/a 的/u 文化/n 娱乐/n 的/u 重要/a 活动/n 方式/n、/	corpuz_享受_xls.txt 0 4
48	享受 /v 少/a、/w 其/d 她/r 也/d 够/v 够/a 了/u、/w 45	corpuz_享受_xls.txt 0 5
49	享受 /v 所谓/a、/w 因此/c 导致/v 一些/r 不/v 稳定/n 享受 /v 的/	corpuz_享受_xls.txt 0 4
50	享受 /v 所谓/a /w 文本/n 的/u 快乐/a /w、/w 230 极尽/v 奢	corpuz_享受_xls.txt 0 2
51	享受 /v 快乐/a、/w 奢侈/v 它/r、/w 敬祝/v 它/r、/w 353 例如/	corpuz_享受_xls.txt 0 3
52	享受 /v 与/c 一般/a 菜肴/n 不同/a 的/u 特殊/a 的/u 一切/r、/w	corpuz_享受_xls.txt 0 2
53	享受 /v 知/c 欢乐/a ; /w 劳动者/n 所/u 失去/v 的/u 一切/r、/w	corpuz_享受_xls.txt 0 4
54	享受 /v 这/c 一个/nl 老妻/n 的/u 向题/n、/w 在/p 自己/r 的/	corpuz_享受_xls.txt 0 3
55	享受 /v 与/c 金钱/n、/w 那/r 文化/n 的/u 地位/n、/w 产生/v	corpuz_享受_xls.txt 0 1
56	享受 /v 那/c 斗争/n、/w /w /w 准备/v 为/p 取得/v 胜利/a 地/	corpuz_享受_xls.txt 0 1
57	享受 /v 知/c 精神/n 享受 /v、/w 使/v 自己/r 的/u 生活/n 过/	corpuz_享受_xls.txt 0 1
58	享受 /v 知/c 思想/n 感知/v、/w 233 十年树木/v、/w 地/u 树	corpuz_享受_xls.txt 0 2
59	享受 /v 知/c 金钱/n 的/u 获得/v、/w 在/p 表现/v 方法/n 上/	corpuz_享受_xls.txt 0 3
60	享受 /v 与/c 白人/n 平等/a 的/u 权利/n、/w 387 /w 这/r 种/q	corpuz_享受_xls.txt 0 4

図3 “享受”の後方共起一覽

2.2 調査の手順とコーパス利用の問題点

二言語間のコロケーションの対照研究は非常に複雑な作業である。そもそもコロケーションの対照は決して切りがない単純な相違点の並べあけではない。膨大な言語データから有効な情報を抽出するためには、明確な方向性と抽出基準が必要とされる。

それに、コーパスは万能な存在ではなく、田野村 (2012) はそれがあくまでも文字の集積物だと指摘し、コーパスを利用する際によくある問題点を以下の四種類に分けて挙げている。

- ① 同表記異語を区別できない
- ② 多義を区別できない
- ③ 表現の統語構造に関する情報を利用できない(「長い象の鼻」という表記において「長い」と「象」の2語が係り受けの関係にないことを統計処理に反映させたくてもむずかしい)
- ④ 表現内容の意味特性を知ることができない

(田野村 2012:219)

そこで、本研究はコロケーションの視点から中日「同形同義語」の相違性を考察し、統語的繋がりや組み合わせの制限要因、共起語の分析、結びつきの強さや特徴的なコロケー

ション側面から相違状況や特徴をまとめている。以下は調査の手順を説明し、各操作段階における共通の問題点と注意事項を整理する。

第一段階：統語的繋がりや組み合わせの制限要因

Benson (1985) は、コロケーションを「何度も繰り返し頻繁に用いられる組み合わせ」と定義し、コロケーションを語彙的コロケーション (lexical collocation) と文法的コロケーション (grammatical collocation) に分類している。前者は、「動詞+名詞」「形容詞+名詞」、「副詞+動詞」、「副詞+形容詞」など、ある語句に対して結びつきが強く、高頻度で 사용되는組み合わせを指している。一方、後者は特定の文法機能と習慣的に共起する関係を意味している。

語と語の組み合わせは、統語または語のもつ意味などの要因によって制限されている。中日「同形同義語」を考察する場合、他の語句との共起や組み合わせのパターンから、その特徴や文法機能を見ることができる。第一段階ではまず中日両語の文法的振る舞いに着目し、統語的繋がりや組み合わせの制限要因という大枠から中日「同形同義語」の相違点をまとめる。

第二段階：共起語の分析

コロケーションの文法的性質を把握してから、次は各パターンにおける共起語の性質に注目することである。対象語と常に高い頻度で共起する語の性質分析によって、対象語の意味的統語的特徴も浮き彫りになってくる。逆に言えば、中日同形語の意味や統語上の違いは、共起する語彙グループの性質や構成にも反映されている。

共起語を抽出するとき、シンタクスの面からコロケーションを観察するアプローチも必要とされる。解析済みのデータをそのまま利用することではなく、文脈において、その意味的統語的な特徴から、総合的に判断しなければならない。例えば中国語では“享受幸福”“享受幸福生活”の場合、対象語“享受”の直接後続語がいずれも“幸福”であるが、前者の目的語は“幸福”であるのに対して、後者は“生活”となっている。この場合“幸福”は直接“享受”と関連することではなく、“生活”の連体修飾語として働いていると見られる。または同じコロケーションであっても、文脈によって、品詞性が異なる場合が多いため、文脈から切り取ってコロケーションを考察することは好ましくないとと思われる。

第三段階：結びつきの強さや特徴的なコロケーション

コロケーションの結びつきの強さを計る指標としては、raw frequency (素頻度)、Log Dice 係数、MI スコア (相互情報量) などがある。その内、raw frequency (素頻度) は最も基本的な指標で、Log Dice 係数は頻度の高い一般的なコロケーションの検出に強い指標で、素頻度と似たような結果を示す。一方、MI スコア (相互情報量) は意味論的な特徴が強調される指標である。特徴的なコロケーションほど数値が高くなる傾向がある。ただし、「低頻度の項目を高く評価する傾向があり…あまり頻度の低い語の場合、コロケーションの結びつきが過大評価される」(田畑 2012: 119) という点には注意する必要がある。故に、本研究では MI スコアが高くても、使用頻度の低いコロケーションは考察の対象外にしている。

実際に考察を行う際に、また以下の問題点に注意を払う必要がある。

① スパン設定

スパンの設定は共起語の抽出やコロケーションの統計に大いに影響を及ぼしている。調査の目的や対象語の特徴に合わせて、コリゲーションの想定パターンを考慮に入れながら、適切なスパン設定にたどり着くのが非常に重要である。

② 中日「同形同義語」の対応関係

これまでのコロケーション研究は、同一言語内で行われることが多く、前方・後方という方向性でコロケーション状況を統計するのが一般的である。しかし、中日両言語は言語

種類や語順が異なっているため、逐一確認するより、意味と統語的特徴や文中機能に注目して、最も本質的な部分から対照を行う必要がある。

本研究は異なる言語システムにおけるコロケーション状況の対照であるため、相違性が確実に存在している。そこから、研究目的と合わせて、言語学的あるいは中日対照に有意義な相違点を抽出することは何より重要である。

③シンタクスからの考慮

イレナー (2013) は形容詞と名詞のコロケーションを例に、統計する時のシンタクス分析を強調している。氏の研究では形容詞の振る舞いを徹底的に分析するために、形容詞単独で名詞を修飾するもの (Ai+N) だけに注目することではなく、語連鎖のパターンも視野に入れている。

a. 連体修飾節の述語の機能を持つ形容詞で、節が名詞を修飾するもの

[N が Ai]N、 [N の Ai]N、 [N も Ai]N、 [N は Ai]N など

砂抜きが甘い店、地震危険度の高い地域など

b. 所有の「の」 + 形容詞で名詞を修飾するもの

[Poss の+Ai+N]

タレの甘辛い味、男性の太い声など

c. 名詞 + 形容詞という構造をもった複合形容詞で名詞を修飾するもの

[N+Ai]+N

香り高いコーヒー、テンション高い人など

本研究で利用するデータとして、NLB のほうは完成度が高い解析済みのデータであるのに対して、CNC の場合は AntConc の利用で二次分析が施されている。

実際に二次分析を行うとき、下記のようなパターンも表れている。

例 (1) 絶対的組織優勢

例 (2) 絶対离不开

ここで、“絶対”は直接に後接している“組織”“离不开”という共起語と関連づけることなく、後ろにある構造全体と修飾関係が成り立っているのである。

コロケーションの関連情報を統計整理するとき、シンタクスも考慮に入れ、直接的な共起語よりその実質的な修飾関係に注目し、コロケーションの非連続性などを配慮すべきである。

④精度の問題

コーパスや解析ツールの利用であくまでも機械的な分析で、100%の正確性が保証できない。

a. 分類ミス：

例 (3) このように、養殖経営体は、種類を多様化させつつ、その時代の需要を反映する形で増減してきているが、一方、はまちをはじめとする魚類養殖における赤潮問題やほたて養殖における大量へい死の問題等も経営体数の変化に影響を与えているものと思われる。

(漁業白書, 1978)

検索キーワード：大量

NLB におけるパターン分類：～+助詞+動詞

b. 分割問題：

例 (4) ここに全面的な「客観的総体性」つまり「絶対的総体性」が成立する、と言うことができよう。

(高山守著 『ヘーゲル哲学と無の論理』, 2001)

NLB においては「総体性」が分割され、「絶対的」と「総体」の組み合わせとして見ているが、実例を見れば「絶対的」と「総体性」の組み合わせとして扱ったほうが適切である。

c.非該当例

例(5)到现在为止,应用这些测验所取得的结果,主要对于心理学家具有重要意义,离开实际应用还有一定距离。

(周昌忠《创造心理学》,1983)

検索キーワード:家具

検索条件によって、“心理学家具有”という非該当例も表れている。そのために、コンコーダンス・ソフトの利用で効率化客観化を図ると同時に、手作業による確認や、該当例であるかの再確認も必要とされる。

3. 実例分析

この節では中日同形語に関する主な先行調査(文化庁1978、曾根1988、西川1991、橘1994、王2001)において、いずれも「同形同義語」(S類)と判定されている「享受」を実例として、前節の調査手順を踏まえて、コロケーションの視点から中日の相違点を明らかにする。

第一段階:統語的繋がりや組み合わせの制限要因

享受(日本語)

検索総件数:享受する382件/享受220件、合わせて602件

NLBではサ変動詞をデータ処理する際、名詞と動詞を異なる検索語として扱い、それぞれ統計することになっているため、「享受」を分析する場合、「享受」だけの入力ではなく、「享受する」を指定語として再検索する必要があることに注意しなければならない。以下はNLBにおけるコリゲーション状況を後方共起と前方共起という二種類に分けて再整理した結果である。

表3 NLBにおける「享受」の基本検索状況のまとめ

パターン	後方共起	例	前方共起	例
享受する	～+名詞 ～+助動詞 ～+形容詞	～する権利 ～しよう ～しやすい	名詞+助詞+～ 名詞+複合助詞+～ 副詞+～ 形容詞連用形+～	利益を～ 一員として～ まさに～ よく～
享受	～+助詞+動詞 ～+助詞+形容詞 ～+助詞+形容動詞 ～+助詞+名詞	～を求める ～の永い ～が可能だ ～の前提	形容動詞+～ 連帯詞+～ 名詞+の+～	怠惰な～ その～ 快樂の

享受(中国語)

検索総件数:508件

CNCでは設定する検索範囲内の結果を頻度順で配列することができるが、両言語間の対照分析を進めるために、検索データの二次分析及び手作業が必要とされる。

表 4 CNC の品詞分類一覧

表記	品詞種類	表記	品詞種類	表記	品詞種類
n	普通名詞	v	動詞	u	助詞
nt	時間名詞	vd	趋向動詞	e	嘆詞
nd	方位名詞	vl	联系動詞	o	拟声詞
nl	处所名詞	vu	能愿動詞	i	习用語
nh	人名	a	形容詞	j	縮略語
nhf	姓	f	區別詞	h	前接成分
nhs	名	m	数詞	k	后接成分
ns	地名	q	量詞	g	語素字
nn	族名	d	副詞	x	非語素字
ni	机构名	r	代詞	w	标点符号
nz	其他专名	p	介詞	ws	非汉字字符串
		c	連詞	wu	其他未知的字符

後方共起を調べた結果、主に「動詞+目的語」あるいはそのバリエーションで展開されていることが分かる。

v+n:～権利

v+a(u)+n:～民主權利，～一定的權利

v+c+v+n:～和使用權利

v+a+c+a+n:～文明与幸福的生活

v+mg+n+u+n:～一下帝王的待遇

そのほか:

n+u+n:～的待遇

v+c+v (目的語無し): (为了)～而斗争,～或欣赏

また、“在…享受中”“享受着”“享受过”などの例もある。

前方共起:

a+u+n:美的～

n (u) +n:物质～, 精神的～

a+u+v:充分地～

d+v:尽情～

d+vu+v:也不愿～

vu+v:可以～, 要～, 能～

vu+a+u+v:应该充分地～

v+n:追求～, 得到～

v+a+u+n:得到美的～

v+n+n:得到物质～

上記のパターンを整理し、中日対照考察した結果、以下の四点に注意する必要がある。

①中国語と違い、日本語の場合、名詞としての「享受」は目的語としてあまり使用されていない。NLBで調べた結果「～を目的とする」「～を求める」「～を願う」という3件に止まっている。一方、中国語では“追求～”“得到～”などv+n型が多く見られているのである。

②修飾関係において、日本語の場合「用言+享受」のパターンが非常に少なく、NLBでの分類結果から見る限り、形容詞または動詞連体形修飾の該当例は、「怠惰な～」という1件しかなかった。一方、CNCで調べた結果“美的享受”だけでも36件もあることが明らかになっている。

ただし、意味的な関連性のみにはフォーカスすると、「用言+享受」というパターンとは少し構造が異なるが、日本語では「美的(な)～」「芸術的な～」という「名詞+的(な)+～」のパターンは確かに存在していることが調査で分かった。しかし、量的には依然として少ない状態であり、その修飾関係の特徴から見ればやはり中日間にずれが存在している。

③中国語“享受”の後方共起は、主に動詞とそのバリエーションが中心として展開されている。日本語の場合「～を享受する」という前接パターンに変換されるが、文中での働きや動詞と目的語の組み合わせという点に変わりはない。

④辞書で「享受」の品詞性を確認したところ、中日いずれも名詞と動詞という種類しか記載されていない。しかし、そのコロケーション状況を調査した結果、日本語では、用例が少ないにも関わらず「享受的+名詞」という形で形容動詞に相当する働きを見せている。さらに、中国語では、“真是享受”“十分享受”の例も存在し、どちらかという形容動詞に近い用法となっている。

第二段階：共起語の分析

動詞として文中で一番基本的且つ重要な働きは特定な目的語と組み合わせて意味を表すことである。以下は中日それぞれのコロケーション調査における高頻度の「目的語+動詞」(動賓結構)の例である。

日本語：

利益を享受する(28件)

メリットを享受する(18件)

恩恵を享受する(15件)

生活を享受する(12件)

中国語：

享受権利(60件)

享受生活(38件)

享受教育(19件)

享受幸福(18件)

目的語として中日いずれも「権利」「生活」が多用されているが、NLBにおいては「幸せを享受する」が3件、「教育を享受する」が1件のみで、中国語の高頻度状態と大いに離れている。

次は他の名詞との共起状況である。

中国語では、“物質享受”“艺术享受”“精神享受”というコロケーションが多く見られる一方、日本語の場合「物質」「精神」という表現が存在しているにも関わらず、「物質(的)享受」「精神(的)享受」などのコロケーションが見当たらず、代わりに「物質的豊かさを享受する」「精神的豊かさを享受する」など動詞と目的語の組み合わせとなっている。また、

日本語では「芸術的享受」「芸術享受」のコロケーションも存在するが、使用頻度は極めて低く、中国語とかなり異なっている。このような中日「同形同義語」の使用頻度と使用傾向の違いは表面化されず、普段の辞書頼りでも発見しにくいいため、日本語学習者にとってさらに注意を払う必要があると思われる。

第三段階：結びつきの強さや特徴的なコロケーション

以下は日本語で MI スコアが 10 を超えているコロケーションのまとめである。ただし使用頻度が 3 件以下の例は今回の統計から除外した。

表 5 NLB における「享受」の特徴例

コロケーション	MI スコア	使用頻度
恵沢の享受	19.84	4
美的享受	16.55	5
享受している	14.52	87
享受してくる	14.27	10
快樂の享受	13.75	3
享受しうる	13.48	7
恩恵を享受する	13.02	15
享受しえる	12.38	7
便益を享受する	12.35	4
メリットを享受する	11.71	18
恵みを享受する	11.93	6
享受すうことを妨げる	11.89	4
繁栄を享受する	11.20	5
特権を享受する	11.06	6
充分に享受する	10.77	6
享受し続ける	10.68	3
享受していく	10.30	3
利益を享受する	10.25	28

前述のように、MI スコアは他の語彙に比べる特徴性を示す指標で、同一言語システム内の共起特徴を反映している。表 5 で示されたように、MI スコアが最も高くなっているのは「恵沢の享受」というコロケーションで、使用頻度から見れば 4 件しかないが、上記の表には収録されていない「恵沢を享受する」の 2 件と合わせて、全部で 6 件となっている。NLB において 15 件の検索結果しかない「恵沢」にしては、「享受」との共起件数はもはや少ないとは言えない状況である。

しかし、特徴的な共起語をまとめて見ると「恵沢」「快樂」「恩恵」「便益」など、文体的には硬めな漢語が多く見られている。それらの共起語の性質からも、「享受」という対象語の文体的特徴も推知することができる。

おわりに

中日同形語の対照研究は、コーパス言語学の発展につれて、今までの辞書だけを頼りにして、作例や手作業で用例集めて分析を行うことから一変して、最近の研究ではコーパスを利用して調査を行うことが多くなってきている。コーパスの利用で、大規模な言語データ処理が可能になり、共起状況についてもより全面的・客観的に捉えられるようになってきている。二つ異なる言語間のコロケーション対照は、無意味に相違点を並べるのではなく、日本語学習者の理解や使い分けに向け、言語学的あるいは中日対照に有意義な相違点を抽出することは重要である。

本研究では、形態・概念上高度な類似性を持ち、内省や辞書のみで相違性を把握し難い中日「同形同義語」に目を付け、コーパスを利用することで、コロケーションの視点から言語事実に潜まれる両語の相違点や特徴性をまとめている。さらに方法論に基づき、統語的繋がりや組み合わせの制限要因、共起語の分析、結びつきの強さや特徴的なコロケーションという三段階にわけて、実際に考察を行う際の調査手順を説明し、各操作段階における共通な問題点と注意事項を整理する。

考察を行う際、中日「同形同義語」「享受」を実例として重点的に分析を行い、相違性をまとめているが、同形類義語・同形異義語の相違性考察にも汎用でき、中日同形語全体に適応すると見込まれる。また、このような相違性が潜在的であり、誤用率が高い特殊な語彙グループを主な対象に、調査方法や手順を詳しくまとめることは、日本語学習者の理解と使い分けだけでなく、日本語教育現場での語彙指導教育までにも役立つと期待できる。

なお、実際の操作中においては、適切なスパン設定や、両言語間の対応関係及び分析対象の確定、コロケーションの非連続性などの問題も並存し、さらなる検討の余地があると思われる。今後の課題としては、調査精度と効率の両立や考察項目の全面性などを目指し、研究を進めるつもりである。

注

¹文化庁(1978)の分類においては、S類O類D類以外に、もう一種類N類(Nothing)も存在している。N類は「日本語の漢語と同じ漢字語が中国語に存在しないもの」を指しているため、普通中日同形語の範囲には入っていない。

²コロケーションの範囲を指す。スパンの設定は共起語の抽出やコロケーションの統計に大いに影響を及ぼすため、調査の目的や対象語の特徴に合わせて、コロケーションの想定パターンを考慮に入れながら、適切なスパン設定が非常に重要である。

参考文献

- 大河内康憲「日本語と中国語の同形語」『日本語と中国語の対照研究論文集(下)』,411-447(1992).
- 王蜀豫『日中語彙の対照的研究：同形語を中心に』四川文芸出版社,2001.
- 曾根博隆「日中同形語に関する基礎的考察」『明治学院論叢』424,61-96(1988).
- 橋純信「現代中国語における中日同形語の占める割合」『国際関係学部研究年報』15,99-116(1994).
- 田畑智司「日本語のコロケーション」『これからのコロケーション研究』,107-152(2012).

- 田野村忠温「日本語のコロケーション」『これからのコロケーション研究』,193-226(2012).
- 西川和男「現代漢語頻率詞典からみた日中同形語について」『関西大学文学論集』40(3),49-65(1991).
- 福島一人「定義されたコロケーションとその有用性:コロケーション辞典執筆の見地から」『情報研究』36, 253-282(2007).
- 堀正広『英語コロケーション研究入門』研究者,2009.
- 文化庁『中国語と対応する漢語』大蔵省印刷局,1978.
- 潘钧《中日同形词词义差异原因浅析》《日语学习与研究》3, 19-23(1995).
- 曲维《中日同形词的比较研究》《辽宁师范大学学报》6, 34-37(1995).
- 施建军《中日现代语言同形词汇研究》北京大学出版社, 2019.
- 王晓《关于日语教学中的中日同形词》《日语学习与研究》S1, 40-43(2004).
- 卫乃兴《共选理论与语料库驱动的短语单位研究》《解放军外国语学院学报》1, 1-6(2012).
- 卫乃兴、陆军《对比短语学探索》外语教学与研究出版社, 2015.
- Benson. Collocations and Idioms. "Dictionaries, Lexicography and Language Learning". Oxford: Pergamon Press,1985.
- Firth,J.R. "Paper in Linfuistics 1934-1951" Oxford University Press,1957.
- Sinclair,John. "Corpus Concordance Collocation" Oxford University Press,1991.
- Sinclair,John.The search for units of meaning. "Textus",9(1),75-106(1996a).
- Sinclair,John.An international project in multilingual lexicography. "International journal of lexicography",9(3),179-196(1996b).
- Sinclair,John. "Trust the Text" Routledge,2004