直訳傾向にある日中同形語の諸問題

―各種コーパスを用いた“基本”の使用実態調査から―

宮本　華瑠（大阪大学　院生）

１．はじめに

「日中間で対応するか否か」惑わされやすい類の同形語には，意味的類似度が高く，辞書記述ではほぼ同義として扱われ，日中間で直訳傾向にある所謂同形同義語が挙げられる．しかし，日中間の微妙なズレが原因となり，訳された内容が不自然になるケースをよく見かける．日中同形語の研究及び同形異義語，同形類義語の研究はその歴史も長く，研究成果も多くみられるが，同義の部分に注目し，その微妙なズレを調査対象とした研究はまだ少ない．本発表はこのような，所謂同形同義とされてきた「基本」と“基本（以降　JIBEN）”を調査対象に，各種コーパスを用いた検証を行うことで真の対応関係を明らかにすることを目的とする．直訳傾向に関しては，北京日本学研究センターの「中日対訳コーパス（以降　対訳コーパス）」を用い確認を行った結果を図１に示す．

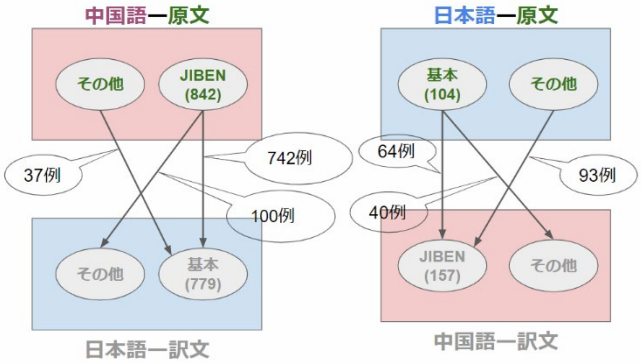

図１:「JIBEN」と「基本」の翻訳傾向

２．先行研究

日中同形語研究には日本文化庁（1978）をはじめ，荒川（1979），守屋（1979）などがあげられる．これまでは主に「どのような表記形（簡体字，繁体字の区別など）を日中同形語にすべきか」，「同形語をいかに分類（同義，類義，意義）するか」について多く論じられてきた．一方，意味分野では多くの場合，同形異義，同形類義についての考察が主流となっている．しかし，コーパスの普及と共に，所謂「同義」とされてきた同形語の研究も増えつつあり，その研究事例として，曹（2007），施・洪（2013），熊・玉岡（2014）などがあげられる．熊・玉岡（2014）は，1383 語を用い複数の辞書をもとに日中両言語品詞性の対応関係について調査を行った．その結

果，日中で品詞が一致する語は 714 語（51.63%）を占めると示している．しかし，品詞情報が掲載されている二つの中国語辞書の記述が不一致の場合には，品詞性の判断が不明であるため集計から外した語が 126 語みられると述べている（p38）．リスト（p50）を確認したところ「基本」は品詞性の判断が不明なグループに含まれていることが明らかとなった．

曹（2007）は，対訳コーパスを用い日中同形語「基本」と「JIBEN」の比較を行っている．しかし，同氏は名詞，形容詞，副詞として用いられる「JIBEN」に対し，日本語の「基本」は名詞として使用された用例のみ調査対象としており，品詞性の捉え方の側面で問題点が見られた．

３．日中対照研究にコーパスを用いる意義

Whorf（1940＝1970）は次のように述べている．

> 言語は「社会的事実」に対する指針である．（中略）事実は「現実の世界」というものは，多くの程度にまで，その集団の言語習慣の上に無意識的に形づくられているのである．二つの言語が同一の社会的現実を表すと考えてもよいくらい似ているということはありえない．住みついている社会集団が違えば世界も異なった世界となるものであり，単に同じ世界に違った標識がつけられたというものではないのである．（池上嘉彦訳『文化人類学と言語学』：2）

日中同形語を比較する際にも同様，一見同形同義として認識されがちな語であっても「二つの言語が同一の社会的事実を表すと考えてもよいくらい似ているということはありえない」し，「住みついている社会集団が違えば世界も異なった世界となるものであり，単に同じ世界に違った標識がつけられたというものではない」のである．しかし，異なった社会集団（日中）が用いる標識（言語）というのは，その言語全体から量的に調査を行わない限り「無意識的に形づくられている」「集団の言語習慣」を考察することはできない．「集団の言語習慣」の現れとして考えられるのは，日常生活での使用頻度，どのような対象物によく用いるかと言った周辺に現れる語との結びつき，付与された社会文化的意味などがあるのではないかと考える．本発表ではこれらのことを「語彙プロフィール（Stubbs：2002）」と呼ぶことにする．語彙プロフィールの考察には，コーパスを利用し実際の言語使用から推論できるとする観点が存在する．Stubbs（2002＝2006）は「意味は目に見えない（おそらくは心的な）現象であり，語の意味を観察することは不可能である．しかし，意味を確実に推論できるような証拠を観察することは十分に可能である．語の意味の証拠となるものは，主として，当該語の前後に来る他の語，とくに反復される共起パターンである．」と述べている（同上：21）．「反復される共起パターン」は共起頻度の分析から得ることができる．

昨今，共起頻度が有意かどうかの判定にはTスコアもしくはMIスコアがある．MIスコア(または相互情報量）は頻繁につかわれる一般性の高いコロケーションの評価に適しているとされ，統計解析の指標として次の式で値が定義されている．

$$MI = log_2 \frac{\text{共起頻度} \times \text{総語数}}{\text{中心語頻度} \times \text{共起語頻度}}$$

MIスコアの値は次のようにも解釈できる．

$$MI = log_2 \frac{\sqrt{\text{共起頻度}}}{\text{中心語頻度}} \times \frac{\sqrt{\text{共起頻度}}}{\text{共起語頻度}} \times \text{総語数}$$

本稿では解析指標「中心語相対頻度」と「共起語相対頻度」を設け，中心語と共起語牽引力の値を求める．

$$\text{中心語相対頻度} = \frac{\text{共起頻度}}{\text{中心語頻度}} \times 100 \qquad \text{共起語相対頻度} = \frac{\text{共起頻度}}{\text{共起語頻度}} \times 100$$

表1：「基本」の相対頻度集計表一部抜粋（BCCWJコーパス）

連番	共起語	品詞	P(A)	中心語用例数	中心語相対頻度	共起語用例数	共起語相対頻度	共起語	中心語Log	共起語Log	割合	右3合計
1	台帳	名詞	0.985	17898	1.849	578	57.266	台帳	0.267	1.758	57.266	331
2	方針	名詞	0.891	17898	8.420	5933	25.400	方針	0.925	1.405	25.400	1507
3	構想	サ変名詞	0.872	17898	2.894	2777	18.653	構想	0.462	1.271	18.653	518
4	理念	名詞	0.845	17898	2.403	2316	18.566	理念	0.381	1.269	18.566	430
5	人権	名詞	0.900	17898	2.006	4146	8.659	人権	0.302	0.937	8.659	359
6	条例	名詞	0.762	17898	1.106	2361	8.386	条例	0.044	0.924	8.386	198
7	策定	サ変名詞	0.613	17898	1.307	3302	7.087	策定	0.116	0.850	7.087	234
8	計画	サ変名詞	0.740	17898	7.023	27585	4.557	計画	0.847	0.659	4.557	1257
9	操作	サ変名詞	0.743	17898	0.983	4862	3.620	操作	-0.007	0.559	3.620	176
10	原則	名詞	0.797	17898	1.536	7694	3.574	原則	0.187	0.553	3.574	275
11	姿勢	名詞	0.843	17898	1.084	6027	3.219	姿勢	0.035	0.508	3.219	194
12	事項	名詞	0.721	17898	1.229	8594	2.560	事項	0.090	0.408	2.560	220
13	設計	サ変名詞	0.672	17898	0.665	5415	2.198	設計	-0.177	0.342	2.198	119
14	施策	サ変名詞	0.427	17898	0.508	4288	2.122	施策	-0.294	0.327	2.122	91
15	方向	名詞	0.788	17898	1.101	11552	1.705	方向	0.042	0.232	1.705	197
16	調査	サ変名詞	0.779	17898	2.693	29492	1.634	調査	0.430	0.213	1.634	482
17	定める	動詞	0.524	17898	1.145	14822	1.383	定める	0.059	0.141	1.383	205
18	基づく	動詞	0.529	17898	0.872	13839	1.127	基づく	-0.060	0.052	1.127	156
19	構造	名詞	0.415	17898	0.749	13708	0.978	構造	-0.126	-0.010	0.978	134
20	考え方	名詞	0.806	17898	3.777	87061	0.776	考え方	0.577	-0.110	0.776	676
21	政策	名詞	0.377	17898	0.559	13453	0.743	政策	-0.253	-0.129	0.743	100
22	作成	サ変名詞	0.476	17898	0.492	12525	0.703	作成	-0.308	-0.153	0.703	88
23	問題	ナイ形容	0.581	17898	1.363	71463	0.341	問題	0.135	-0.467	0.341	244
24	技術	名詞	0.381	17898	0.520	27587	0.337	技術	-0.284	-0.472	0.337	93
25	情報	名詞	0.485	17898	0.525	39386	0.239	情報	-0.280	-0.622	0.239	94

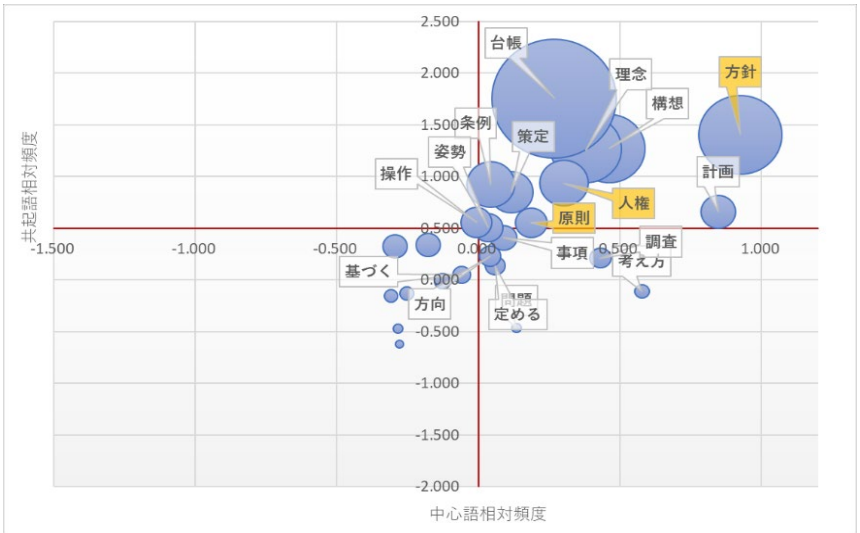

図２：【BCCWJ コーパス】「基本」の相対頻度（高頻度共起内容語）

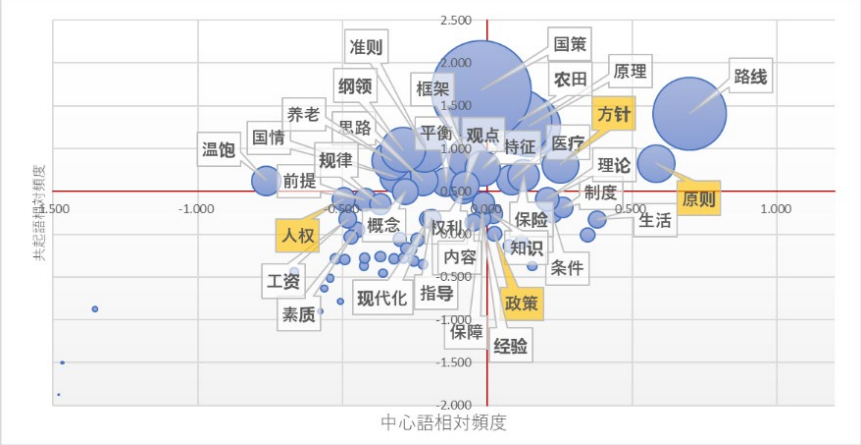

図３：【CCL コーパス】「JIBEN」の相対頻度（共起語－名詞）

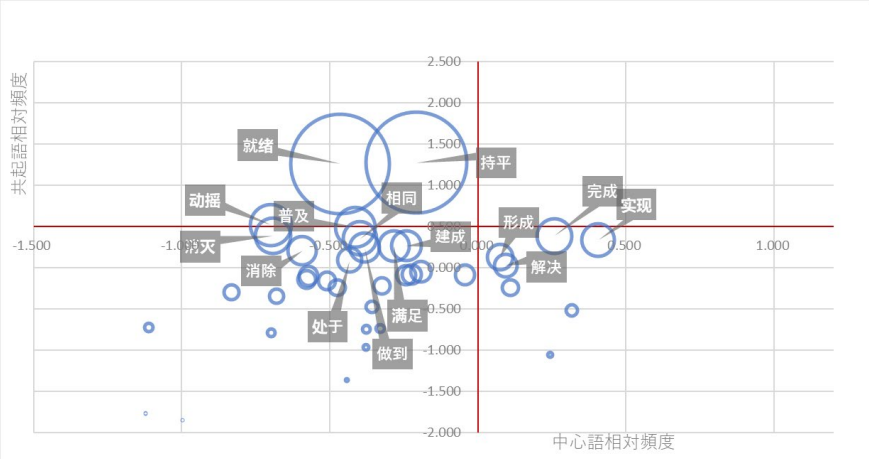

図４：【CCL コーパス】「JIBEN」の相対頻度（共起語－動詞，形容詞，形容動詞）

高頻度共起語は Dice 係数 0.014 以上のものを採用し，高頻度共起語の採択域は次の部分集合確率演算式で求める．n(A)はキーから右 3 語範囲内で共起した語の度数総和，n(Ω)はキーから左右 5 語範囲内の共起語度数総和となる．

$$P(A) = \frac{n(A)}{n(\Omega)} \qquad \text{採択域：} P(A) \geq 0.3$$

上記図 3，4，5 は中心語相対頻度（Log10 値）を横軸，共起語相対頻度（Log10 値）を縦軸，バブルのサイズは共起語度数を分母とし，中心語から右 3 語範囲内で現れた共起語度数の総和を分子にした割合の値を採用している．

４．各種コーパスを用いた分析

4.1　日中対訳コーパス

対訳コーパスの用例からは，日本語の原文中「教育の基本，しつけの基本，接し方の基本，のびのびと成長させるための基本，商売の基本」など被修飾語として「基本」が用いられる場合と，「ほめるのが基本｛だ／である／として｝」のような文末で用いる使用が多く確認できた．しかし，このような使用は中国語「JIBEN」には存在しない．中国語原文中，単純語「JIBEN」の使用は 1 例のみとなっており，見出し語を除き，単純語として用いるケースはほとんど見られなかった．以上のことから，名詞の働きをする側面で，「基本」と「JIBEN」は全く異なる統語関係を持っていることが明らかとなった．　また日中同形語が構成素となる一語化は言語によって単語の意味合いが異なるため，このケースも「異義語」として扱うべきであると考える．また，副詞の働きをする「基本」と「JIBEN」は，「意味の微妙なズレ」による曖昧性は見られない．　以上のことから直訳された場合，日中同形語「基本」と「JIBEN」の使用で棲み分けが曖昧な状態にあるのは主に漢語と結びつくケースであることが推論できる．

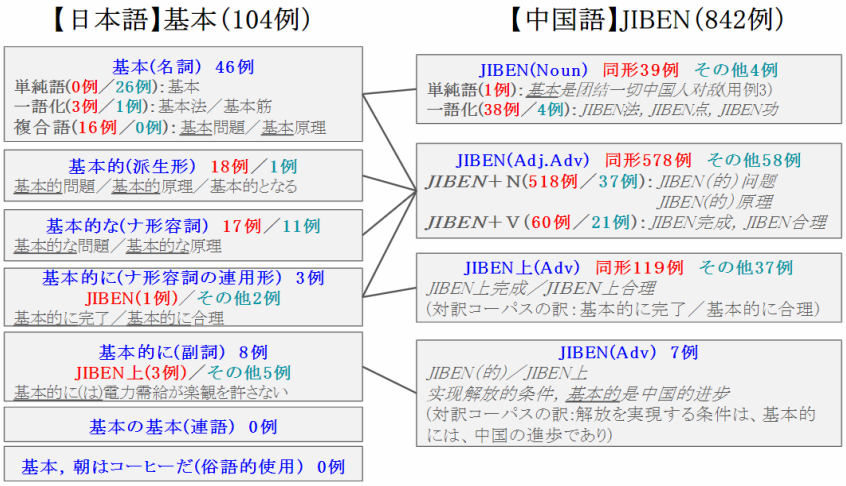

図 5：直訳された「基本」と「JIBEN」の対応関係（日中対訳コーパス）

4.2　日中対訳 EGA コーパスと BCCWJ による検証

日中対訳 EGA コーパス[1]（以降　EGA コーパス）と「現代日本語書き言葉均衡コーパス Web（BCCWJ）」から得られたデータを基に，テキストマイニングソフトウェア KH Coder を用い，調査語「JIBEN」と「基本」の 3 グラム範囲内高頻度共起語の傾向を調べた結果，日本語のみに存在する共起語（Ⅰ類）が 19 語「台帳，構想，理念，姿勢，考え方，方向，調査，条例，操作，計画，事項，設計，策定，基づく，定める，情報，作成，施策，構造」，中国語のみに存在する高頻度共起語（Ⅱ類）が 98 語“国策，平衡，思路，原理，框架，准則 act.”，日中で共通する高頻度共起語（Ⅲ類）が 6 語「原則／8 位(原則／30 位)，方針／3 位(方针-43 位)，問題/17 位(问题/91 位)，政策／25 位(政策，73 位)，人権／2 位(人权／86 位)，技術／24 位(技术／103 位)」が明らかとなった．

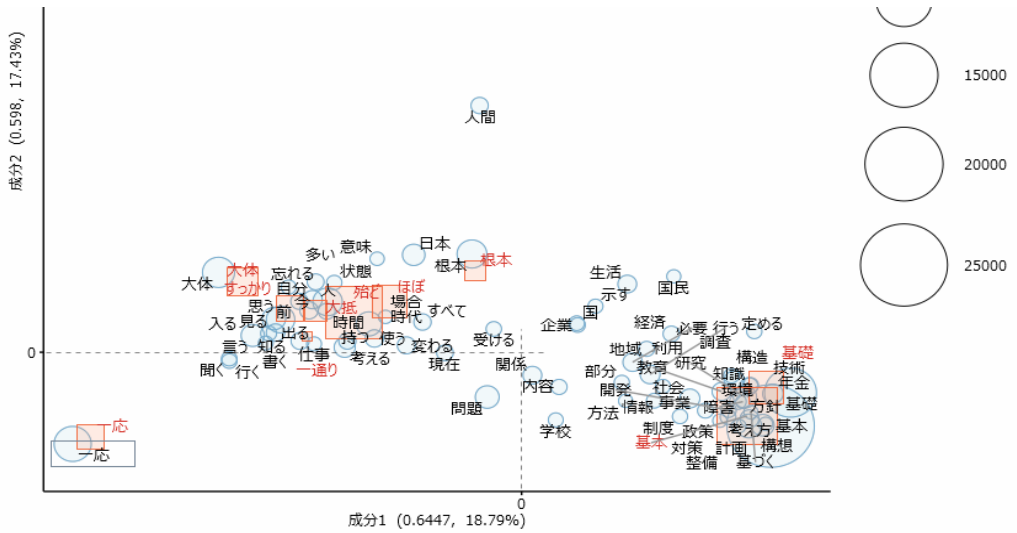

図 6：　10 グラム範囲内に現れた内容語の対応分析（BCCWJ コーパス）

図 6 から読み取れるように，日本語「基本」と「基礎」の類似度が高く，類義語群「ほぼ，殆ど，大抵，一通り，すっかり，大体」が互いに類似度が高く，孤立傾向にあるのは「一応，最低」である結果となっている．EGA コーパスを用い同じ手法で「JIBEN」の分析を行った結果，「JIBEN」は“大致（DAZHI）”と類似度が高く，「JIBEN 上」は“大概（DAGAI）”と類似度が高いことが明らかとなった．そして“根本（GENBEN）”は“总（ZONG）”“总是（ZONGSHI）”との類似度が高く，孤立しているのは“基础（JICHU）”であると見受けられる．

[1] 宮本華瑠（2023）「日中対訳コーパスの構築と公開に向けて」『言語資源ワークショップ 2023 予稿集』国立国語研究所．https://clrd.ninjal.ac.jp/lrw2023-programme.html

５．まとめと今後の課題

5.1　「基本」と「JIBEN」の語彙プロフィール

「基本」は名詞として使われる場合と，「基本的」という派生形を用い副詞的役割を担う場合がある．日本語から中国語「JIBEN」に訳される語は，同形語からではなくその他の類義語から訳されるケースが（図１の 93 例）多く見られたが，それは「基本」と類似度の高い対応語として副詞（大体，殆ど，ほぼ，大抵，根本など）が多く存在しているからである．中国語「JIBEN」は 86%の割合で同形語「基本」に訳されるが，それは「JIBEN」の癖語（集団の言語習慣），癖語以外に存在する「JIBEN」が結びつきやすい語，及び「JIBEN」との結びつきを必要とする語が中国語に（特に）多く存在しているからである．しかし，「JIBEN」には「基礎」という意味合いで用いられるケースは見られず，最も類似度の高い語として用いられるのは副詞“大致（DAZHI)”であることから，中国語「JIBEN」は，品詞面で名詞と副詞に分類されるが，実際は多くの場合副詞として使われていることが読み取れる．また，「基本」には中国語「JIBEN」が持たない独自使用が存在し，このようなケースにおいては，「基本」は「JIBEN」に訳されず，他の対応語に置き換えれらる．以上のことから直訳傾向にある日中同形語「基本」と「JIBEN」は，副詞として用いられる場合に限って「基本/基本的」と「JIBEN/JIBEN 上」は対応関係にあり，同じ意味として扱うことが可能である．名詞として用いる時，「基本」と「JIBEN」は日中とも癖語が多く，とくに「JIBEN」は中国特有表現が目立っており，翻訳の際には無暗に直訳するのではなく，日本人でも理解しやすい用語で工夫する必要があると考える．但し，日中で共通する共起語として「原則，方針，問題，政策，人権，技術」6 語が存在する．

表２　同形語「基本」と「JIBEN」の使用傾向対照表

	日本語「基本」	中国語「JIBEN」
共通する共起語	原則,方針,問題,政策,人権,技術	
癖語 （第１象限）	基本｛台帳(57.27%),方針(25.40%),構想(18.65%),理念(18.57%),人権(8.66%),条例(8.39%),策定(7.09%),計画(4.56%),原則(3.57%),姿勢(3.22%)｝	JIBEN｛路线(25.6%),农田(20.6%),原理(18.3%),原则(6.70%),方针(6.40%),保险(4.86%),医疗(4.30%)｝
中心語が結び付きやすい対象(第４象限)	設計,施策,基づく	理论,制度,条件,知识,生活,政策,建设,要求,情况,问题
結びつきやすい対象として中心語を必要とする語(第２象限)	操作	国策,准则,纲领,框架,养老,特征,国情,思路,温饱,平衡,观点,权力,概念,人权,前提,保障,规律,指导,经验,现代化,工资,内容,素质
類義語	基礎	大致

類似度の高い対応語	根本, 大体, 殆ど, ほぼ, 大抵, すっかり, 一通り	大概, 根本, 总,总是
類似度の低い対応語	最低, 一応	基础
独自使用	俗語：基本，朝はコーヒーだ。 文末表現：～のが基本(だ)。	

5.2　今後の課題

熊・玉岡(2014)で示した「品詞性の判断が不明であるため集計から外した語」 126 語を優先すべき調査対象とし，意味の類似度が高い日中同形語の分析を積み重ねていく予定である.

文献

Michael Stubbs. (2002). *Words and Phrases: Corpus Studies of Lexical Semantics.* Blackwell Pub, 南出 康世・石川 慎一郎監訳(2006).コーパス語彙意味論 ： 語から句へ.研究社.

Benjamin Lee Whorf & Edward Sapir. (1940).池上嘉彦訳（1970）『文化人類学と言語学』弘文堂.

荒川清秀（1979）「中国語と漢語―文化庁『中国語と対応する漢語』の評を兼ねて」『愛知大学文学論叢』62 号.361-388

施建軍・洪潔（2013）「汉日同形词意义用法的对比方法研究」『外语教学与研究』 45（4）531-542

熊可欣・玉岡賀津雄(2014)「日中同形二字漢字語の品詞性の対応関係に関する考察」『ことばの科学』27 号.25-51 名古屋大学言語文化研究会.

曹大峰(2007)「多言語コーパスと日本語研究−『中日対訳コーパス』の利用研究例から」 『特集 コーパス日本語学の射程　日本語科学』22 号. 59-77.国書刊行会.

文化庁（編）(1978)『中国語と対応する漢語』大蔵省印刷局.

守屋広則（1979）「資料・日中同形語―その意味用法の差違」『日本語学校論集』06.159-168.東京外国語大学外国語学部付属日本語学校.

ツール

KH coder (3.Beta.07b) https://khcoder.net　(2023 年 6 月 20)

関連 URL

北京日本学研究センター(2003)「CD-ROM 中日対訳コーパス」(2023 年 6 月 20)

北京大学「Center for Chinese Linguistics PKU（CCL 语料库)」
http://ccl.pku.edu.cn:8080/ccl_corpus/CCLCorpu s_Readme.html　(2023 年 8 月 5)

国立国語研究所「現代日本語書き言葉均衡コーパス（BCCWJ)」
https://chunagon.ninjal.ac.jp/bccwj-nt/search（2023 年 8 月 5 日).