

BCCWJ で知る東アジア漢字圏四字成語の受容と変容

砂岡和子 Sunaoka Kazuko (早稲田大学政治経済学術院)

羅鳳珠 Lo FengJu (台湾元智大学中国語文学系)

王雷 Wang, Lei (北京大学計算言語学研究所, 北京大学英語系)

姜柄圭 Kang, Byeong Kwu (韓国ソウル市立大学中文系)

1 漢字四字成語の流通と受容

四字成語は東アジア漢字圏の智慧と文化を凝縮する表現形式のひとつであり、簡潔で口調の良い言語形式とあいまって漢字圏に広く流通する。四字成語の総数はその定義と準拠するデータの性質により諸説あるが、出現頻度自体は高くなく、5万語中では8%弱、1万語では0.2%程度である[2011 姜柄圭]。このうち常用成語は頻度 100 回が 2200 条で全体の 85%を占め、1000 回以上の高頻度成語は 110 条で 2 割弱に過ぎない[同上]。現代中国語の常用成語は 2000-3000 条と見なせよう[2007 劉長征他]。出現ジャンル別では通俗・古典文学から翻訳物まで小説類での四字成語使用が 8 割近くを占め、新聞が 2 割とそれに次ぐ [2011b 砂岡和子他]。

四字成語は韓国や日本にも伝わり、当地の言語文化と接触と受容、変容を繰り返してきた。中国語は孤立語系統に属し、語と語の意味関係による結合規則が統辞法を左右する。中国語の四字成語は意合結合の縮図であり、日本語や韓国語など膠着語系言語に見られない多様な句構成があり、品詞機能も動詞、名詞を主体に豊富である【表 1】

属性	句構成	成語例	略号	品詞機能	比率
pz	修飾関係	孜孜不倦、一衣带水	IV	動詞機能成語	72.7%
zw	主述関係	草木皆兵、枯木逢春	IN	名詞機能成語	11.6%
bl	並列関係	不学无术、千锤百炼	IA	形容詞機能成語	5.3%
db	動目関係	蹉跎岁月、饱经风霜	ID	副詞機能成語	4.7%
dbu	動補関係	重于泰山、无动于衷	IB	區別詞機能成語	2.6%
dbb	動目補関係	问道于盲、入木三分	IM	修飾機能成語	1.7%
ld	連動関係	打草惊蛇、兴妖作怪	IC	補語機能成語	1.4%
ju	兼語関係	请君入瓮、唆人作恶			
fz	複指関係	半斤八两、半夜三更			

頻度	北大[A]	台湾[C]	台湾[H]	韓国[E]	日本[名大]
1	艰苦奋斗	一毛不拔	不知不觉	战战兢兢(戰戰兢兢)	老若男女
2	自力更生	掩耳盗铃	不可思议	名副其实(名實相符)	臨機應變
3	实事求是	一丘之貉	理所当然	유언비어(流言蜚語)	流言飛語
4	千方百计	囫圇吞棗	莫名其妙	자유자재(自由自在)	立身出世
5	全心全意	班門弄斧	前所未有	자초지종(自初至終)	油断大敵
6	坚定不移	朝三暮四	截然不同	자유분방(自由奔放)	優柔不斷
7	因地制宜	竭澤而漁	自然而然	이목구비(耳目口鼻)	無味乾燥
8	独立自主	天衣無縫	脫穎而出	동서고금(東西古今)	無我夢中
9	成千上万	杞人憂天	迫不及待	시시각각(時時刻刻)	三日坊主
10	坚持不懈	水落石出	不約而同	동서남북(東西南北)	滿場一致
11	战无不胜	名落孫山	層出不窮	방방곡곡(坊坊曲曲)	本末顛倒
12	轰轰烈烈	胸有成竹	無可奈何	비일비재(非一非再)	傍若無人
13	前所未有	指鹿為馬	依依不捨	의기양양(意氣揚揚)	暴飲暴食
14	欢欣鼓舞	舉一反三	自由自在	신진대사(新陳代謝)	不眠不休
15	滔天罪行	江郎才盡	總而言之	좌지우지(左之右之)	百發百中
16	丰富多彩	青出於藍	成千上萬	우유부단(優柔不斷)	和洋折衷
17	朝气蓬勃	口若懸河	不遺餘力	각양각색(各樣各色)	理路整然
18	明目张胆	驚弓之鳥	與眾不同	일사불란(一絲不亂)	理路整然
19	阴谋诡计	含沙射影	大街小巷	유명무실(有名無實)	竜頭蛇尾
20	以身作则	一敗塗地	取而代之	요지부동(搖之不動)	八方美人

【表 1】中国成語句構成と品詞機能

【図 1】中日韓トップ 20 成語

対して日本語の四字漢語は 2 字+2 字の緩い連合形式から成るものが全体の 90 パーセント以上を占め、一語としての凝縮力が弱い¹。このため単語としての統辞機能も極めて制約を受け、二字

¹ [1974 野村雅昭]の調査は約 19 万新聞語彙の四字漢語全体についてであるが、日本語四字成語にも当てはまる特徴と言える。

漢語や三字漢語と比べサ変動詞や形容動詞の語幹になることが少ない²。日本語の四字成語は中国語の窮屈な統辞制約を逃れ、母語に適した表現法へと活路を開いた。例えば中国語四字のうち基幹2字だけ残し、“自相矛盾”を“矛盾”に、“完璧歸趙”は“完璧”に、“画蛇添足”を“蛇足”に、“杞人憂天”は“杞憂”といった軽量化と脱典故化で日本語での受容力を強めた。また“三顧の礼（中国語原文は“茅廬三顧”）”“枚挙にいとまがない（中国語原文は“不勝枚舉（不可枚舉，不遑枚舉とも）”）”“多士濟々（中国語原文は“多士濟濟”）”のように日本語語順や平仮名など漢字以外の文字種が混在する。このため日本語の成語は形式上、異形や類語が極めて多く[1982 宮地裕]、定義や中国語との同定が困難である。韓国語の成語も同様の傾向を示す[2011 姜柄圭]。

2 中台英韓日マルチリンガル成語辞典の編纂

現在筆者らは国際共同プロジェクト「東アジア歴代語言知識庫」の一環として、中台英韓日マルチリンガル成語辞典を構築している。編纂作業は主に[第一段階]中国、台湾、韓国、日本の各地域を代表する現代言語コーパスからそれぞれ常用の中国（由来）の成語を抽出し、[第二段階]中台韓日で重複率の高い成語約2000語を選出、[第三段階]で成語の百科知識と各種言語情報を付与し、[第四段階]で英語、韓国語、日本語に翻訳、マルチリンガル中国成語辞典としてWebで公開する。筆者[砂岡]は日本語常用成語選定と中国成語とのマッチング作業にBCCWJを利用する。

辞典の本体は北京大学計算言語学研究所（PICL）開発の「成語知識庫」で、同所の『現代漢語語法情報辞典』の成語5000条を核に「人民日報」ならびに『成語大全』『学生成語詞典』など中国出版の代表的成語辞書5冊から3万項目余の現代中国語成語を収集し、成語の頻度情報、語法属性、品詞、異形、類語、反義語、典故、積義など16種の言語属性を記したコーパスである[A 北大三萬詞簡體成語庫][2010 王雷]。これを台湾で使われる繁体字に変換校正し[A' 北大三萬詞繁體成語庫]、台湾中央研究院のバランスコーパスと照合して台湾地域常用成語約1万語をリストアップした[B 台湾中央研究院平衡語料成語庫]³。さらに台湾教育部公布の成語データから台湾香港地域の常用成語約1万語を抽出した[C 台湾教育部成語詞條]。

3 中日韓成語マッチング作業

中国語成語の使用頻度は依拠するデータのジャンルで大きく異なる。[A] [B] [C]はともに中国成語であるがトップ20語に共通の成語はない。韓国[E]、日本「名大成語」⁴は更に異なり、5種データベース共通で比較的上位の成語は“自由自在”“流言蜚語”“傍若無人”“隨機应变”“百發百中”“虎頭蛇尾”“千篇一律”と極少数に留まる【図1】。そこで[A] [B] [C]各データベースから3者の重複回数が高い中核となる成語3000語を抽出し、各原表のIDを付し、重複回数に応じ4レベルに分けた[D 三千核心成語庫]。韓国、日本チームは[D]を参照し、それぞれ韓国国立国語研究院世宗コーパス[E 世宗成語庫]、現代日本語書き言葉均衡コーパス（BCCWJ）を基準に[F BCCWJ 成語庫]、各地域の常用成語を選別抽出する。

² 同上 pp 44 「周章狼狽（～スル）」「一知半解（～ダ）」などは例外的な存在とする。

³ 抽出には台湾研究院院士の鄭錦全教授の手を煩わせた。

⁴ 当時BCCWJを利用できなかったため、名古屋大学佐藤理史研究室「基本慣用句5種対照表」所収常用成語を利用。

ジャンル	成語数	%	ヒット数	%
書籍文学	93	47	1346	65%
書籍社会科学	35	18	359	17%
書籍歴史	24	12	140	7%
Yahoo!ブログ	11	6	28	1%
書籍言語	11	6	35	2%
書籍芸術	7	4	13	1%
書籍哲学	7	4	48	2%
Yahoo!知恵袋	6	3	14	1%
国会会議録	4	2	31	1%
書籍自然科学	3	2	7	0%
雑誌総合	3	2	9	0%
教科書	2	1	3	0%
書籍工学	2	1	30	1%
韻文	1	1	1	0%
新聞	1	1	1	0%
書籍総記	1	1	10	0%
計	211	100	2075	100
検索不能	1		1836(*500)	

【図2】中国頻用成語[少納言]検索

検索語列	不知順位	BCCWJ	出版年	BCCWJ
大所高所	5	38	2005	24
古希	10	33	2003	23
苦心慘澹	10	24	2004	21
豪放磊落	7	21	2001	19
多士濟濟	2	16	2002	18
一木一草	20	13	2008	17
螻蛄の斧	4	10	1997	7
特筆大書	3	9	1998	7
立錐の余地も無い	10	8	1984	6
衆議一決	8	7	1996	6
出処進退	20	6	1989	5
四角四面	16	5	2000	5
四分五裂	16	2	1986	4
粗製乱造	13	1	1991	4
灯火親しむべき候	1	0	1999	4
手の無い足の踏む所を知らず	5	0	1990	3
同病相哀れむ	8	0	1992	3
功成り名遂げる	13	0	1979	2
精神一到何事か成らざらん	13	0	1985,9 3,95	各2
千里の堤も蟻の穴から崩れる	16	0	1977,8 0,81,8	各1
好事門を出でず	16	0	7,88,9	
三拝九拝	20	0		

【図3】学生不知成語[少納言]検索

4 BCCWJによる日本語成語の検索

日本語成語とのマッチング作業に先立ち、日本語での頻出ジャンルを特定するため、BCCWJ 検索器「少納言」で[D]の重複回数が4-3回ある高頻度中国成語計2005条を手作業で検索を試みた。日中成語の表記の完全一致は期待できず、「少納言」検索では日中台の字体の違いを含め異形での検索も行なった。例えば「不思議（中国語原文は「不可思議）」「興味津津OR興味（中国語原文は「津津有味）」「背水の陣OR背水（中国語原文は「背水一戦）」など、部分一致は自動抽出の後、人手で選別が必要となる。結果は中国成語2005条の内、ヒットした日本成語は211個（重複回数4回が74,3回が126）で全体の約10%に過ぎないが【図2】⁵、「書籍の文学,社会科学,歴史」で9割をカバーすることが判明した。日本語成語の多出ジャンルをほぼ特定できたため、次にBCCWJ本体(DVD媒体)で半自動によるマッチング作業に移った。被検索語は北京大学8000語と後述の名古屋大学佐藤理史研究室「基本慣用句五種対照表」所収の約3000語と大量の上、慣用句の表記の揺れ、解析器の誤認などが原因でヒット率は良くない。解決策としてマッチング用テキストの短縮化、バイグラムの導入、原文テキストでの検索など方策を検討している。

5 BCCWJによる漢字慣用句認知度検証

中日同形成語のマッチング作業と並行し、2010年と2011年にかけて日本人大学生計53名を対象に日本の基本慣用句⁶の認知度調査を行なった。結果は「多士濟濟」「螻蛄の斧」「豪放磊落」「大

⁵ 「完璧」のヒット数は1836件と多いが、「少納言」は500件までしか検索結果が表示されず出現ジャンルの特定ができないため「検索不能」とした。

⁶ 名大「基本慣用句」は慣用句1918語,ことわざ344語,四字熟語265語,故事成語63語,その他不明1038語に分類される。テストはこのうち四字漢語を多く含む四字熟語と故事成語計328語について行った。

所高所」「同病相哀れむ」「衆議一決」「古希」「苦心慘澹」「立錐の余地も無い」「功成り名遂げる」「粗製乱造」「四分五裂」などの成語を 80%以上が理解できない。「灯火親しむべき候」「老いては子に従え」「蛍雪」「晴耕雨読」「粗製乱造」「薄利多売」「一利一害」は半数が知らない。調査した 328 語は「呉越同舟」「異口同音」のような中国語伝来の成語（62%）と、「一所懸命」「油断大敵」「絶体絶命」など日本製の慣用句（38%）に分類できる。大多数が知る成語は「一喜一憂」「十人十色」「春夏秋冬」「自由自在」「自暴自棄」など後者の慣用句に多い。認知度調査の結果は日本の常用成語リスト選定に反映する。

6 BCCWJ で知る漢字四字成語の受容と変容

上掲学生が知らない慣用句（成語）トップ 20 語⁷を[少納言]で検索した【図 3】。結果は「大所高所」36 件、「古希」33 件、「苦心慘澹」24 件、「豪放磊落」21 件、「多士濟濟」16 件がヒットした。明治大正期の近代小説には BCCWJ より更に豊富な漢字成語を観察できる[2009 野浪正隆]。言語コーパスは共時的な言語情報だけでなく通時的な流通状況を知ることができる言葉の履歴書である。多地域のコーパスを連携すれば言語間の接触と変容を正確なデータで把握可能となり、比較文化研究、翻訳はもちろん、語学教育の科学性と効率向上に寄与するに違いない。

謝辞：本研究は平成 22-24 年度文部科学省科学研究費補助金〔基盤（C）課題番号：22520445：研究代表者砂岡和子〕および 2010 年 7 月 1 日～2013 年 6 月 30 日台湾蔣經國國際學術交流基金會による国際プロジェクト「歴代語言知識庫の構築」(RG013-D-09, 研究代表台湾元智大學羅鳳珠)の助成で進行中の成果の一部である。

参考文献

- 1974 野村雅昭, 四字漢字の構造, 国立国語研究所報告 54, 電子計算による国語研究
- 1982 宮地裕編, 慣用句の意味と用法, 明治書院
- 2007 劉長征&秦鵬, 基于中国主流报纸动态流通过语料库的成语使用情况调查, 語言文字應用 No 3
- 2007 佐藤理史, 基本慣用句五種対照表, <http://sslslab.nuee.nagoya-u.ac.jp/index.html>
- 2009 野浪正隆, 近代小説に使われた故事成語について, 早川勝度教授・山田勝久教授退官記念論文集, 大阪教育大学国語教育講座・日本アジア言語文化講座 52 卷, PP27-56
- 2010 王雷, 俞士汶, 朱學鋒, 李芸, 面向中文資訊處理的成語知識庫的建設與應用, 第五屆文學與資訊科技國際會議, 台灣亞洲大學主辦, 2010.1/22-23.
- 2011 姜柄圭, 漢韓成語的語法結構對比分析及教學策略, 第四屆華語文教學國際研討會發表, 2011.3/11~12, 台灣銘傳大學
- 2011a 砂岡和子, 羅鳳珠, 以「歴代語言知識庫」發展的中日韓漢字成語教學模式, 第四屆華語文教學國際研討會發表, 2011.3/11~12, 台灣銘傳大學
- 2011b 砂岡和子, 羅鳳珠, 日本語常用漢字熟語の選好変化と自然言語処理, 日本語処理学会 (NLP2011) 論文集 CDROM, E3-2
- 現代日本語書き言葉均衡コーパス (BCCWJ) : 「少納言」 <http://www.kotonoha.gr.jp/shonagon/>
- 韓国国立国語研究院世宗コーパス : <http://www.korean.go.kr>; <http://kkma-sc.snu.ac.kr>

⁷ 同位があるため計 22 語。