

日本語常用漢字熟語の選好変化と自然言語処理

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)

I 口語表現の Intelligence 向上

ウェブやモバイルの発展、多言語入力拡大に伴い、母語による口語表現での発信が可能となり、SNS、ブログ等に一次情報の口語テキストが地球規模で増大している。口語による表現は日々のコミュニケーションに欠かせない言語基盤である。従来、一過性で一次消費された口語表現領域の一部に、近年「自然言語処理」が踏み込み始めた。人間が生得的に獲得する口語は、インフォーマルな主観表現に使用され、音声によるオンラインの進行、重複や言い淀み、相づちや身振りなどの非言語情報が多出する。本来、極めて「自然」な口語の特長を「従来の枠に収まらない」とするのは、自然な言語に対応しきれない情報処理側の言い訳である。動的情報を含む口語表現の場合は今後ますます拡大するであろう。自然言語処理研究は一層可処理言語領域を広げ、自然な言語への対応を急ぐ必要がある。

口語による表現の拡大は、個人間のコミュニケーションを促進し、市民社会の育成に寄与する一方、書き言葉に比べ使い慣れた平易な表現や仲間内のことばが選好される結果、ありきたりの表現しかできない現代人を産出する恐れがある。知能や知性に富んだ口語基盤を失えば、国民の言語知性は低下し、新たな創造も生まれない。言語政策的に口語表現の知識情報を向上させ、持続的に Intelligence な言語資源を確保すれば、口語の言語処理にも有利である。書き言葉同様、口語表現の一定の規範性維持は、人間同士のコミュニケーションの質を高め、言語処理の効率を向上させる。

II デジタル漢字の功罪

日本語表現の Intelligence 向上を考えるさい、漢語の使用を抜きにできない。ツイッターや携帯メールに代表されるように、現在、ネット上の口語表現の多くは平易な口語を文字で写し取った短いテキスト文である。日本人の若年層もデジタル文入力には漢字を積極的に使用することが分かっている。2003年、大学生の携帯メール表記に見る漢字使用意識調査によると、9割近くが「なるべく漢字に変換する」と答え、その理由に相手の「読みやすさ」や、「漢字を使ったほうが文意が伝わりやすい」利点を挙げている¹。日本語中の漢字表記の功罪については明治期以来、賛否両論が繰り返されて来たが、情報機器の漢字変換機能が標準装備されて以来、若者も漢字の書記の煩わしさから解放され、漢字の意味弁別機能の高さと表記イメージを好み、積極的に漢字表現を行っている。同調査では小学校6年生の約9割が日本語から漢字をなくすことに反対しており、漢字表記は今後も若年層の支持を獲得し続けよう²。

¹ 棚橋 2005

² 同上

デジタル漢字の選好に反し、若者世代の漢字語彙力が低下傾向にあるのは、漢字が多い書物を読んだり、緊縮した漢語表現を使う機会が少ないためと考えられる。漢語は今なお日本語の客観表現と造語に重要な役割を占め、口語表現が豊かな言語資源であり続けるには、書記言語から栄養を吸収することが必要である。書き言葉と話し言葉の相互交流は、文章語テキストを対象にしてきた自然言語処理にとっても歓迎すべき姿である。

Ⅲ 漢字慣用句の認知度

周知のように中国語は漢字だけで書記され、成人の常用漢字は 4000—8000 字を越す。原則一漢字を単語単位とし、統語法規則に代わり、語(漢字)の造語力が強い。なかでも四字熟語は漢字造語が定型表現に昇華した代表格といえる。中国の小学六年生の日記には、本文 80 漢字中に“阳光灿烂”“乌云密布”“闷闷不乐”“愁眉苦脸”“郁郁寡欢”“愈加小心”“一无是处”“稍不留神”“平淡无奇”“无忧无虑”“充满阳光”“喜笑颜开”“酸甜苦辣”“味味俱全”“更上一层楼”“二两棉花四把弓”など 16 種もの四字句や慣用句が使われる³。四字成語は簡潔な言語形式に民族の智慧と文化知識が凝縮され、古代から現代まで様々なジャンルで愛用される。韓国や日本語にも伝わり、現地でそれぞれ独自の発展を遂げた。

筆者はそれぞれ日本と台湾の大学で中国語と中国文化の教育に従事しており、近年、学生の漢字語彙力が低下していることに危惧を抱いている。以下、日本人学生に対して行った漢字慣用句の認知度調査を例に、日本語中の常用漢字慣用句の選好変化を観察する。

調査目的：中国伝来の漢語を含む日本語常用慣用句や成語を、日本人中国語学習者がどの程度認知しているか調査する。同じ調査を少数の社会人と在日中国留学生にも行い、異なる言語環境での認知度の違いを比較する。

被験者：日本人大学生 31 名（全員政治経済学専攻で第二外国語として中国語を履修中の学部生、内 6 名は中国語学習時間約 450 時間の 2 年生、別の 25 名は中国語学習時間約 150 時間の 1 年生クラス、両クラスとも欧米圏からの帰国子女が含まれる）。日本人社会人 4 名（20 代 1 名、30 代 1 名、60 代 2 名）、在日中国留学生 6 名（理系院生 4 名、文系院生 2 名、20 代社会人 1 名）にも同じ調査を実施した。

調査方法：下記、名古屋大学佐藤理史研究室作成「基本慣用句五種対照表」所収の計 3628 語から日本語常用慣用句 328 語を選び、アンケート形式で学生に知っている熟語を選択させた。直感テスト方式で、所要時間はパソコン使用のデジタル回答方式で 10・15 分、紙媒体では 20・30 分と、どちらも短時間で回収した。

調査語句：「五種対照表」は、1982—2006 年にかけて日本で出版された一般人から小学生向

³分数，改变了我的生活 <http://www.zww.cn/zuowen/html/21/180744.htm>

けの辞書から、慣用句を抽出・整理・分類して対照表とし、電子データとして公開する⁴。

“慣用句”はさらに「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、計 3628 の慣用的表現を登録する。今回はこのうち四字漢語を多く含む「四字熟語」と「故事成語」について調査を行った。

調査結果： [表 1]に見るように漢字を含む慣用句の認知度は、60 代社会人が 100%と一番高く、ついで文系留学生、20-30 代社会人、大学 2 年中国語履修者、在日 4 年以上中国留学生、在日 2 年未満中国留学生と続き、大学 1 年中国語履修者と理系中国留学生は 328 語中 25%から 30%弱が認識できない。

被験者	慣用句未知度
社会人 (60 代)	0%
留学生 (文系)	1.8%
社会人 (20-30 代)	4.2%
中国語履修大学生 (2 年)	4.6%
留学生 (在日 4 年以上)	13.7%
留学生 (在日 2 年未満)	14.3%
中国語履修大学生 (1 年)	24.2%
留学生 (理系)	28.0%

[表 1]

被験者	中国成語認知率
社会人 (60 代)	99.75%
留学生 (文系)	93.56%
留学生 (在日 4 年以上)	86.80%
社会人 (20-30 代)	84.65%
中国語履修大学生 (2 年)	82.34%
留学生 (在日 2 年未満)	78.38%
留学生 (理系)	77.10%
中国語履修大学生 (1 年)	69.84%

[表 2]

大学 1 年中国語履修クラスの 80%以上が、「大所高所」「特筆大書」「多士済済」「衆議一決」「手の舞い足の踏む所を知らず」「好事門を出でず」「千里の堤も蟻の穴から崩れる」「螳螂の斧」「精神一到何事か成らざらん」「立錫の余地も無い」「出处進退」「無念無想」「四分五裂」「四角四面」「苦心惨澹」「三拝九拝」「功成り名遂げる」「昼夜兼行」「得手勝手」「前門の虎、後門の狼」「豪放磊落」などを見てすぐ理解できない。「老いては子に従え」「蛍雪」「晴耕雨読」「灯火親しむべき候」を半数の学生が知らないのは時代の変化で無理ないとしても、政治経済専攻の学生が「粗製乱造」「薄利多売」「一利一害」が認知できないのは問題である。対して「言うは易く行ふは難し」「穴が有ったら入りたい」「井の中の蛙」「背水の陣」「意気消沈」「起死回生」「起承転結」「疑心暗鬼」「奇想天外」「喜怒哀楽」「極悪非道」「五分五分」などは大多数の学生が知っている。未知語の傾向は他のグループも共通する。

調査した 328 語は、「吳越同舟」「異口同音」のような中国語伝来の成語 (62%) と、「一

⁴ 依拠した五種類の辞書はそれぞれ以下の通り。『慣用句の意味と用法』(明治書院 1982 年、宮地裕編、1280 項目所収)、『小学生のまんが慣用句辞典』(学研 2005 年、金田一秀穂監修、336 句所収)、『新レインボー小学国語辞典』(学研 2005 年、金田一春彦他監修、うち慣用句・ことわざ・四字熟語・故事成語など 2585 項目所収)、『日本語慣用句辞典』(東京堂出版 2005 年、米川明彦他編、公称 1563 句所収)、『小学館学習国語新辞典』(小学館 2006 年、金田一京助編、熟語類 1841 語を収録)、詳細は佐藤理史 2007-NL-178, pp. 1-6、および「基本慣用句五種対照表」URL <http://sslab.nuee.nagoya-u.ac.jp/index.html> を参照。

所懸命」「絶体絶命」など日本製の慣用句（38%）に分類できる。留学生は中国語の成語で強みを発揮し、それぞれ順位が繰り上がるが、日本人の順位は変わらない[表 2]。日本語慣用句の認知度に漢語知識の大小が影響することが分かる。

認知度の個人差も大きい。中国語履修 1 年生クラスには 98%知っている学生がいる一方、認識率 60%以下の学生も 2 割いる⁵。2 年生クラスのトップは 90%だが、65%しか知らない帰国学生もあり、特に中国語成語の認知率は 58%である。「五種対照表」の慣用句は日本人に膾炙する語句のはずだが、一部若年層の日本語話者から漢語の知識が減退しているため、漢語を含む慣用句の認知度も低い。

IV 持続発展可能な口語言語資源の構築

日本語に比べ現代中国語の常用四字成語は 3000—8000 語と数が多いうえ、成語全体の統語機能と四字句内部の語構成が多様である。北京大学計算言語学研究所が開発を進める『中国成語デジタル大辞典』（中文名「成語知識庫」）は中国の新聞記事や代表的成語辞書から集めた現代中国語成語 3 万項目余に、使用頻度や分布密度、典故の由来、英訳、難度レベル、語法属性、品詞機能など 16 種の言語属性を付す⁶。筆者らはこれに日本・韓国の成語との対照情報を加え、マルチリンガル四字成語データベースの構築を目指す。

デジタル世代の日本語話者が型に囚われない情感表現を選好する動きは阻止できない。音声や動態情報を記録する情報センサーはますます感度が好くなり、口語コミュニケーション情報の処理負担は増大する。処理方法の対策を講じるのと同時に、発信側の言語形態を美しく整形する事業も必要だ。書き言葉は文化遺産を継承し、国際的コミュニケーションツールとしての客観表現の源泉である。中国語と日本語は、その長い借用の歴史経過から漢字用法に共通の部分があり、漢語力が落ちると日本語の造語力も衰える。口語表現が書面語と緊張関係を維持し、漢語から語彙を吸収し持続発展可能な言語資源となるよう、自然言語処理側の能動的働きかけが期待される。

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〕および羅鳳珠代表「歴代語言知識庫建置計畫」九十八年度蔣經國國際學術交流基金會の助成で進行中の成果の一部である。

参考文献

- 佐藤理史 2007、基本慣用句五種対照表の作成、2007-NL-178, pp. 1-6、情報処理学会研究報告 <http://sslslab.nuee.nagoya-u.ac.jp/index.html>
- 棚橋尚子 2005、現代社会における漢字表現、朝倉漢字講座第四巻、前田富祺、野村雅昭編、朝倉書店
- 宮地裕編 1982、慣用句の意味と用法、明治書院
- 陳力衛 2008、日本の諺・中国の諺、明治書院

⁵ 途中棄権の可能性も排除できない。

⁶ 砂岡 H P <http://www.f.waseda.jp/ksunaoka/corpus/chengyuzhishiku.pdf> の紹介文参照