

日中二字漢字語における客観的音韻類似性指標¹⁾

——主観的音韻類似性指標との比較——

早川 杏子²⁾ (関西学院大学)

于 劭贇³⁾ (名古屋大学)

初 相娟⁴⁾ (天津外国語大学)

玉岡 賀津雄⁵⁾ (名古屋大学)

これまで、日中における漢字の音韻的差異の判断に関しては、主観的な判定に基づいて行われることが多かった。しかしながら、今後言語間の音韻がもたらす影響に関する研究を、より推進していくためには、判定者の属性や判定方法の違いの影響を受けない測定指標が必要である。そうしたことから、本研究では、文字配列計算のアルゴリズムによって算出した客観的な音韻類似性指標の可能性を探るため、中国語母語話者に対して行った主観的音韻類似性の指標との相関を求め、両指標による類似度平均に基づいて、分析を行った。その結果、中程度の相関 ($r = -0.49$, $p < .01$) にとどまり、両言語による音表表記の対応付けなどに課題が残っていることがわかった。

キーワード：音韻類似性、日中同形語、日中異形語、音韻距離 (phonological distance)

1. 問題の所在と目的

日中の漢字をめぐる研究は、これまで、2言語間における意味の差異について議論されることが多かった。しかし、近年、日中の漢字語の形態 (morphology)・書字 (orthography)・音韻 (phonology) の異同とオンラインの処理 (on-line processing) に焦点を当てた研究が次第に増えつつある。とりわけ、日中両言語間で語彙的に類似した書字を持つ場合には、言語非選択的活性化 (language non-selective activation) が起

1) English Title: Objective indexes of phonological similarity between Japanese and Chinese two-kanji compound words: A comparative study on objective and subjective indexes

2) HAYAKAWA, Kyoko (Lecturer, Center for Japanese Language Education, Kwansai Gakuin University, E-mail: hayakawa@kwansai.ac.jp)

3) YU, Shao Yun (Graduate student, Graduate School of Languages and Cultures, Nagoya University, E-mail: s.yu@protonmail.com)

4) CHU, Xiang Juan (Associate Professor, Department of Japanese Studies, Tianjin Foreign Studies University, E-mail: chuxiangjuan@aliyun.com)

5) TAMAOKA, Katsuo (Professor, Graduate School of Languages and Cultures, Nagoya University, E-mail: tamaoka@nagoya-u.jp)

こることが脳波実験 (Tamaoka, Miyatani, Zhang, Shiraishi & Yoshimura, 2016) や眼球運動実験 (熊, 玉岡, マンスブリッジ, 2016) で示されていることから、両言語間での語彙的特性を記述したデータベースは今後の研究の発展に大きく寄与するであろう。

日中間における漢字の音韻の異同を調べた資料には、単漢字では茅本 (1995)、二字漢字語では当銘・費・松見 (2012) 等がある。茅本 (1995) は、日本語の読みと中国語の読みのペア 1,107 を、日本語学習歴が 2~13 年までの中国人日本語学習者を対象に、1~7 までの尺度による音韻類似度調査を行った。手法としては、日本語での読みをカタカナ、中国語での読みをピンイン (拼音, pin yin) で併記し、表記を通して類似度を判定させるというものであった。一方、当銘・費・松見 (2012) では、日本語の学習経験のない中国語母語話者に、352 語の二字漢字を聴覚呈示し、1~7 までの尺度による音韻類似度判定を求めた。当銘・費・松見 (2012) の調査では、はじめに中国語の発音、続けて日本語の発音が音声呈示された点で、茅本 (1995) の手法に比べて物理的な音声ベースで行われている。しかしながら、評定用紙には単語が漢字表記されていたと記されており、判定者には、聴覚情報だけでなく、視覚的な情報も同時に与えられたようである。

このように、オンライン研究で音韻を扱うにあたり、対象語を選定するため、これまでは目的に応じてそれぞれに日中の音韻類似度を測定し、個別にデータの蓄積を行ってきた。これまでに使用されてきた音韻類似度の指標は、単語を視覚ないし聴覚呈示して、その語がどれだけ中国語ないし日本語に似ているかということの主観的に判定するものであった。しかしながら、日中間における膨大な語を、個々に調べていくには、非常に労力がかかる上、判定者の属性や判定方法の違いによって判断の基準のブレが生じる可能性がある。また、語単位で音声呈示する方法では、二字漢字語を構成する個々の漢字の音韻類似性を調べることができない。さらに、日本語の漢字には複数の発音を有するものが多く、常用漢字 2,136 字のうち、音読み 1 つが 69 字、訓読み 1 つが 743 字であるため、2 つ以上の発音も持つ漢字は 1,324 字となり、全体の 61.99% になる (Tamaoka, Makioka, Sanders & Verdonshot, in press)。さらに、日本語の漢字では、複数の発音が同時に活性化されることが知られている (Verdonshot, La Heij, Tamaoka, Kiyama & Schiller, 2013)。そのため、個々の漢字の音韻類似性を日中で比較することは極めて難しい。とはいえ、今後、日中間の漢字および漢字語における音韻の役割について、研究成果を蓄積していくためには、こうした手続き上の問題を解消する必要がある。

そこで、本研究では、判定者の主観的な判定に依らない客観的基準で音韻類似性の指標を計算することにした。日中漢字語における音韻類似性の客観指標として、R の cba パッケージで提供される *sdists* 関数 (Buchta & Hahsler, 2016) を用いて算出し、

主観的な音韻類似性の指標との相関を調べ、問題点と課題を検討する。

2. 客観的音韻類似性指標

現在、筆者は旧・日本語能力試験出題基準〔改訂版〕(国際交流基金・日本国際協会, 2007) から、1~4 級までの二字漢字語を抽出した日中音韻データベースを構築しているところである。このために、まず2 級から4 級までの同形の二字漢字語をリスト化した朴・熊・玉岡 (2014 a, 2014 b)、熊・玉岡 (2014) をデータベースの基礎資料⁶⁾として利用した。そのデータベースに、「速達」や「予定」など、中国語には使われず、日本語にのみ存在する異形語(和製漢語にあたる)や、1 級相当語の同形語を加え、それらに中国語ならびに日本語の形態・音韻情報を追加した。1 級までの語彙表に載せられた8,009 語のうち、二字漢字で表せる語は3,852 語で、全体の48.1 %を占める⁷⁾。

本研究では、この日中音韻データベースから、2 字とも音読みであり、中国語と日本語ともに存在する同形語、日本語にのみ存在し、中国語には存在しない異形語を対象として、日本語のローマ字表記、中国語のピンイン表記を基準に、R の cba パッケージで提供される `sdists` 関数 (Buchta & Hahsler, 2016) を用いて、これら日中二字漢字語における日中音韻類似性の客観的指標を算出した。この `sdists` 関数は、アルファベットで表記された2 つの単語の形態上の一般化レーベンシュタイン距離 (generalized Levenshtein distance) を自動計算する。一般化レーベンシュタイン距離は、重み付け編集距離 (operation-weight edit distance) とも呼ばれ、一つの文字配列をもう一つの文字配列に変形するための挿入・削除・置換操作に重みを付け、編集操作のコストの合計が最小の時の値を指す (Gusfield, 1997; Boytsov, 2011)。sdists 関数のデフォルトとしては、挿入は1、削除は1、置換は2と、それぞれの操作に重みがつけられており、本研究ではこのデフォルトの重み付けを使用した。例えば、文字列 A 「omomi」を文字列 B 「hosoi」に変形するのに当たって、表1に示したように、「omomi」の先頭に「h」を挿入し、一番目の「m」を「s」に置換し、二番目の「m」を削除するという3つのステップによる合計が最もコストが低い。したがって、文字列 A と文字列 B の間の一般化レーベンシュタイン距離は4となる。

中国語では現在、ピンイン (拼音, pin yin) と呼ばれる中国式の音表規則によってその発音を表すようになったことから、日本語の発音をアルファベット表記化したローマ字と、アルファベット表記によって音韻の対応付けを行うことが可能となった。

6) 日本語、中国語、韓国語の3言語を対象とした2,058語の情報が掲載された Web 検索エンジンが構築されている (<http://kanjigodb.herokuapp.com/>)。検索方法については、于・玉岡 (2015) を参照のこと。

7) 筆者算出。読みによる語数ではなく、書字に基づく集計数である。

表1 一般化レーベンシュタイン距離の算出の例

A	B	編集操作	コスト
-	h	挿入	1
O	o	-	0
M	s	置換	2
O	o	-	0
M	-	削除	1
I	i	-	0
合計			4

sdists 関数は、もともとアルファベット表記の言語に応用化されているので、日中共通の表記である漢字は、アルファベット文字に比べ複雑な形態であり、これによって日中の形態（書字）的類似性を算出することはできない。その代わりに、両言語の発音が共通の音表表記で示されたことは、両言語の「音韻」が音素単位での配列によって、直接比較できるようになったことを意味し、多くの労力を割かなくても、膨大な漢字語の「音韻類似性」を自動的に計算することが技術的に可能となることを示唆する。

実際、Miwa, Dijkstra, Bolger, & Baayan (2014) では、アルファベットで表記された日英語の単語間の一般化レーベンシュタイン距離を両言語間の音韻距離の尺度とし、言語間の表記が異なる日本語の外来語と英語の発音の音韻距離（音韻類似度）を、sdists 関数によって算出した。この変数を含んで予測統計を行った結果、音韻類似性が認知処理の有意な予測要因となることを示した。そこで、本研究では、日中間の音韻類似性の客観的指標を算出し、日英と同様に、日中二字漢字語の音韻距離が認知処理に有意な予測要因となりうるかを検証することを前提条件として、まずはアルファベット表記化された両言語の音素配列による音韻距離の算出を R の cba パッケージで提供される sdists 関数によって行う。

2.1 音韻距離（phonological distance）の算出

はじめに、日本語の読みは、日本で広く使われているヘボン式ローマ字に変換した。そのため厳密な音素表記ではない。摩擦音の「し」は“shi”、「じ」は“ji”と表記し、破擦音の「ち」は“chi”、「つ」は“tsu”と表記した。長音は、例えば「永遠（えいえん）」は“eien”とし、“een”のように同じ音素を重複表記することはしなかった。促音では、「一致（いっち）」の場合、前項の漢字の基本的な子音を残す形で、“itchi”と表記した（その理由は後述）。

次に、中国語の読みは、同形語の場合、『現代汉语词典』第6版の記述に従った。一方、異形語の場合は、中国語において読みが一つしかないものは個々の漢字のピンイン表記に従い、読みが二つ以上あるもの場合、構成する二字の漢字が持つ意味か

ら、読みを特定した。例えば、「到着」の「着」には、“zhe（轻声）”と“zhao 2”という読みがあるが、前者は動作の継続を表す場合の読みであり、当該語の意味からすると、「接触」を意味する後者の読みのほうが適当である。本調査で扱った語に関しては、中国語母語話者にも判定をしてもらい、最終的に読みを決定した。

こうして二字漢字語の読みを特定した後、さらに「到着」は「到」“to/dao”、「着」“chaku/zhao”のように、前項、後項の読みにそれぞれ分けた。つまり、日中漢字語の音韻距離は、①二字漢字語全体の読み、②二字漢字語の前項の読み、③二字漢字語の後項の読みの三種類が算出されることになる。「一致」のような促音を含む語を、“icchi”ではなく、前項の漢字の基本的な子音を残す形で“itchi”としたのは、前項の漢字の語幹を残すためである。

これら三種の音韻距離の算出にあたっては、上述したように、日本語および中国語の両言語の音素配列をもとに、R の cba パッケージの `sdists` 関数を用いて両言語の音韻類似度を自動計算した。この操作によって算出された、客観的な音韻類似性指標である音韻類似度の値をもとに、日中音韻データベース内における各単語の音韻類似度のタグ付けを行った。音韻距離の値が小さいほど、類似していることを示すため、2以下を類似性が高い「H」とし、3~4を中程度の類似「M」、5以上を類似性が低いとみなし、「L」とした。そして、前項および後項の読みの音韻距離の値によって、「HH、HM、HL、MH、MM、ML、LH、LM、LL」の9種類に分類した。1~4級までの二字漢字語 3,628 語のうち、音読みの語は 3,266 語あり、全体の 90.0%であった。この 3,266 語を音韻類似度別に 9 分類にした個数は、以下の表 2 の通りである。

表 2 1~4 級までの音読みの二字漢字語における音韻類似度別の個数と割合

計 3,266 語	HH	HM	HL	MH	MM	ML	LH	LM	LL
抽出数	312	390	363	375	470	433	269	339	315
割合 (%)	(9.6)	(11.9)	(11.1)	(11.5)	(14.4)	(13.3)	(8.2)	(10.4)	(9.6)

以上の中から、主観的音韻類似性を行うにあたり、それぞれの分類の中から、語頭に含まれる仮名にできるだけ幅が出るようにし、中国語にもその語が存在する同形語を書字的に「一致」、中国語にはその語が存在せず、日本語でのみ使用されている異形語を書字的に「不一致」とし、それぞれが同数となるよう、最終的に表 3 のように、400 語を抽出した。

これらの二字漢字語前項漢字および後項漢字の読みの客観類似性指標による音韻類似度平均に違いがあるかどうかを、書字的の観点も含めて、2（日中間の書字的の一致性：一致、不一致）×3（前項漢字内の音韻類似度：H、M、L）×3（後項漢字内の音韻類似度：H、M、L）の三元配置分散分析で確認した。その結果、書字的の一致性の主効果 [$F(1, 382) = 0.270, ns, \eta_p^2 = .001$] は有意ではなく、前項漢字内 [$F(2,$

382) = 384.76, $p < .001$, $\eta_p^2 = .67$] および後項漢字内の音韻類似度 [$F(2, 382) = 374.93$, $p < .001$, $\eta_p^2 = .66$] のいずれの主効果も有意であった。また、前項漢字内と後項漢字内の音韻類似度の交互作用は [$F(4, 382) = 2.31$, ns , $\eta_p^2 = .02$] で、書字的な一致性、前項漢字内、後項漢字内の音韻類似度の交互作用はいずれも有意ではなく、効果量も小さかった [$F(4, 382) = 2.46$, ns , $\eta_p^2 = .02$]。

表3 客観的な音韻類似度により分類された二字漢字語の一部

分類	N	一致	不一致
HH	40	意欲、災難、手段、神秘、速度、待遇	安易、案外、案内、有無、運搬、看板
HM	36	安心、印鑑、会計、頑固、気圧、技術	医者、意味、運賃、開催、改修、記入
HL	60	印刷、外出、改良、記憶、緊急、在学	運転、快適、可決、頑丈、期日、気楽
MH	54	延期、演技、家具、家族、勧誘、関与	一番、一瞬、会社、我慢、掲載、詐欺
MM	26	圧迫、営業、豪華、座席、就任、地球	恐縮、気配、購入、出費、食事、世話
ML	72	影響、円滑、援助、演説、温泉、患者	育成、一日、一緒、一方、映像、屋上
LH	28	往復、各自、火災、合致、許可、興奮	行儀、強引、借金、邪魔、上品、相談
LM	32	永久、学歴、共存、欠席、嫉妬、習慣	栄養、格差、許容、欠如、懸命、全然
LL	32	逆転、共通、経過、傾向、軽蔑、国立	臆病、活発、窮屈、給料、拒否、見学

注：Nは抽出された語の数を表す。

書字的な一致性による音韻類似度の平均値を見ると、「一致」（同形語）の平均値は7.02で、標準偏差（以下、 SD ）は0.08であり、「不一致」（異形語）の平均値が6.96、 SD は0.08で、若干「一致」のほうが距離が大きかった。また、前項漢字内の音韻類似度は、 $H=5.08$ ($SD=0.09$)、 $M=6.84$ ($SD=0.11$)、 $L=9.06$ ($SD=0.10$)で、当然のことながら、 H が最も値が距離が小さく（すなわち、最も類似性が高い）、 L が最も距離が大きかった（すなわち、最も類似性が低い）。一方、後項漢字内の音韻類似度も、 $H=5.16$ ($SD=0.10$)、 $M=6.95$ ($SD=0.11$)、 $L=8.88$ ($SD=0.09$)と同様の傾向であった。

以上のことから、今回抽出した400語は、書字的な一致性と関係なく前項漢字ないし後項漢字の音韻類似度の平均値に違いがない上、前項漢字内および後項漢字内の音韻類似度が H 、 M 、 L と適切に異なる類似度で区分されているという前提条件を満たしていることが確認できた。

3. 主観的な音韻類似性指標

3.1 調査の実施

本調査は主観的な音韻類似性の判定であり、判定者の属性が判断に大きく影響すると考えられる。その可能性の一つは、目標言語である日本語の学習経験である。漢語

は、発音が似ているかどうかはともかく、中国語読みと音読みの音韻規則は比較的対応しているものがある。例えば、「海外」という語の場合、中国語は“hai wai”、日本語の音読みは“kai gai”で、初頭語は異なるものの、後音列はほぼ一致する⁸⁾。その他にも、日本語学習の段階が進むと、漢字の読み慣れるとともに、両言語間の漢字音の関連性を規則的に学習することが考えられる。結果として、中国語との音韻的な開きに関する直観の差に違いが生じてくる可能性がある。判定者の意識には上らないような音韻規則の学習によって、予測のストラテジーが変わる可能性を避けるためには、できれば日本語学習経験のない中国語母語話者が望ましい。しかし、本調査を行うにあたって、日本語未学習の協力者を一定数確保することが困難であったため、仮名の学習を終えたばかりで、入学して2週間の新1年生を本調査の対象者とした。

3.2 調査協力者

調査は、中国華北地域にある大学で行った。日本語の学習を始めてまだ2週間の学部1年生38名（男性5名、女性33名、平均年齢18.4歳）を対象とした。調査時は、ひらがなの学習が終了した時点で、念のため過去の学習歴をフェイスシートによって尋ねたところ、平均日本語学習歴は0.16ヶ月（標準偏差0.43ヶ月）であり、全員日本への留学歴はなかった。調査時点では、日本語でのコミュニケーションが成立しないため、調査にあたっての指示はすべて中国語で行った。

3.3 調査語および手続き

調査対象語は、2.1で抽出した400語である。調査は、中国語と日本語の音声を聞いて比較し、その類似度を主観的に判定するものであった。類似性の最も低いものを1「完全不相似（まったく似ていない）」、最も高いもの「非常相似（非常によく似ている）」を7とし、その間の尺度は「不相似（似ていない）」=2、「不太相似（やや似ていない）」=3、「无法判断（どちらとも言えない）」=4、「略微相似（やや似ている）」=5、「相似（似ている）」=6とした。

調査対象語の並び順をランダム化して、リストを作成した。さらに、単純作業の疲労による影響を考慮し、昇順ないし降順別に並んだ2つのリストを、グループAあるいはBに割り当てた。調査協力者は38名であったので、これを半分に分け、各グループ19名が、それぞれのリストの類似判定を行った。なお、音声情報によってのみ判定をしてもらうために、評定用紙には、リスト番号のみが振られ、対象語の漢字表記は一切示されなかった。

調査対象語の音韻刺激は、日本語の音声を日本語標準語話者（埼玉県出身、30

8) 漢語が借用された時代によって、漢音、唐音、呉音などの違いがあり、全てを同様にパターン化することはできないが、リストの中にはこうした語も実際に多く存在する。

代)、中国語の音声在北京語普通話（プートンファ）話者（黒竜江省出身、20代）が自然な速さで発音した音声で、中国語-日本語の順序で音声呈示した。

実際の調査の流れは次の通りである。まず、中国語で調査語のリスト番号が発音される。その1秒後に、中国語による読みで単語が音声呈示され、その1秒後に同語の日本語読みで単語が音声呈示される。判断時間には、5秒間が与えられ、1秒間のインターバルのあと、次の語が呈示された。1語あたりにかかる時間は、およそ10秒であった。対象語はA・Bグループともに4セット毎にまとめられ、セットの間には数分間の休憩を取った。

さらに、調査終了後、フェイスシートに日本語学習歴等を記入してもらった。

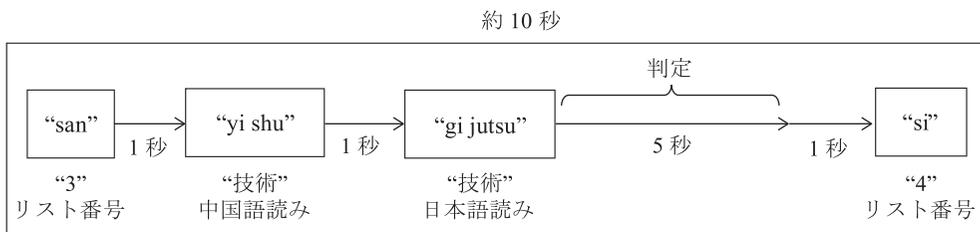


図1 1 試行あたりの音韻類似性判断の流れ

4. 分析結果および考察

4.1 主観的音韻類似性における三元配置分散分析の結果

主観尺度による音韻類似度平均は3.64 ($SD = 1.26$)で、5.0以上の似ている傾向にあった語は67語(16.8%)、3.0以下の似ていない傾向にあった語は143語(35.8%)で、似ていないと判定された語は、似ていると判定された語よりも2倍程度あり、190語(47.5%)はどちらとも言えないという範囲に入っていた。単漢字の日本語読みと中国語読みの主観的音韻類似度調査を行った茅本(1995)では、全体の平均は2.38と似ていないほうに偏っており⁹⁾、全体的に似ていなかった(茅本1995:68)と報告されている。二字漢字語で行った本研究でも、似ていないとされた語のほうが似ているとされた語の2倍近くあり、全体の3~4割を占めていたことから、茅本(1995)と概ね似た傾向が見られた。

なお、主観・客観類似指標それぞれの平均および標準偏差は、表4のようになった。

客観的音韻類似性による二字漢字語の前項漢字および後項漢字の読みの分析と同じ手続きで、2(日中間の書字の一致性; 一致, 不一致)×3(前項漢字内の音韻類似度; H, M, L)×3(後項漢字内の音韻類似度; H, M, L)の三元配置分散分析を行

9) 本研究と同様、7段階評定で、7が「よく似ている」、1が「まったく似ていない」であった。

表 4 主観・客観類似指標の平均および標準偏差

書字	音韻		語数	主観類似指標		客観類似指標	
	前項	後項		平均	標準偏差	平均	標準偏差
一致	H	H	20	4.98	0.92	3.20	0.95
		M	28	4.61	1.13	5.36	0.99
		L	30	3.75	1.14	7.10	1.03
	M	H	27	4.35	1.12	4.89	0.80
		M	13	3.64	1.01	6.69	0.85
		L	36	3.09	1.00	8.83	1.11
	L	H	14	3.20	0.94	7.64	1.01
		M	16	3.08	1.32	8.88	1.36
		L	16	2.76	1.04	10.63	0.89
不一致	H	H	20	4.93	1.09	2.60	1.19
		M	28	3.66	1.01	5.07	0.94
		L	30	3.51	1.22	7.13	1.01
	M	H	27	3.98	1.15	5.41	1.22
		M	13	3.07	1.08	6.69	0.63
		L	36	3.12	1.03	8.56	1.25
	L	H	14	2.85	0.66	7.21	1.12
		M	16	3.11	1.30	9.00	1.41
		L	16	2.72	0.94	11.00	1.46

った。その結果、書字的に一致性に主効果 [$F(1, 382) = 5.99, p < .05, \eta_p^2 = .015$] が見られ、書字的に一致する語 ($M = 3.72, SE = 0.81$) のほうが、不一致の語 ($M = 3.44, SE = 0.81$) よりも主観的音韻類似度評定の平均値が低く、類似性が高くなっていた。すなわち、客観的音韻類似性による指標と違って、主観的音韻類似度評定では、書字的に一致する語（同形語）のほうが、書字的に不一致の語（異形語）よりも類似性が高いと判断されていた。

また、書字的に一致性、前項漢字内、後項漢字内の音韻類似度の交互作用は有意ではなかったが [$F(4, 382) = .95, ns, \eta_p^2 = .01$]、前項漢字内と後項漢字内の音韻類似度の交互作用は [$F(4, 382) = 2.71, p < .05, \eta_p^2 = .03$] と有意であった。しかし、水本・竹内（2008）によれば、多元配置の分散分析の効果量は、偏イータ (η_p^2) の場合は .06 が中程度の目安とされており、前項漢字内と後項漢字内の音韻類似度の交互作用の値はそれに満たないため、効果は小さいと考えられる。

4.2 主観的および客観的類似性指標による比較

4.2.1 相関分析

はじめに、音韻距離を R の cba パッケージで提供される `sdists` 関数で計算を行った。さらに、日中の発音表記に従い算出された客観的類似性指標が、日中二字漢字語の音韻類似性の指標となり得るかを検討するため、実際に両言語による音声で判定さ

れた主観尺度による音韻類似性指標と、客観的類似性指標との相関を分析した。ピアソンの積率相関係数は、 $r = -0.49$, $p < .01$ で、中程度の相関が見られた¹⁰⁾。音韻類似性については、両尺度が同じ特性を測定することを意図しているので、ある程度高い相関が期待される。しかしながら、本研究の相関は必ずしも高い相関であるとは言えず、現段階での客観的類似性指標に、さらなる改善の余地があることを意味していると言えよう。

中程度の相関しか得られなかった理由としては、以下のことが考えられる。まず、日中語によって、アルファベットでの発音表記の方法に違いがあることである。例を挙げると、国際音声字母 (IPA) では、摩擦音となる [tɕ] は、日本語のヘボン式ローマ字表記では “ch”、中国語のピンインでは “q” が用いられる。同様に、[ɕ] は日本語の場合 “sh”、中国語では “x” となる。また、/r/ のような同じ音素表記を用いても、日中では調音法が異なり (日本語でははじき音、中国語はそり舌音)、主観的な音声的印象が大きく異なる場合もある。

このように、日本語のヘボン式ローマ字と中国語のピンインでは、それぞれの言語で音素と表記の対応関係が異なり、実際の発音とうまく対応付けられていない場合がある。しかも、これらの音素の出現頻度がある程度高いことを考えると、両言語の特性が正確に表示されていない以上、cba パッケージで計算された音韻類似性も正確な指標を算出できないであろう。今後、音声と表音法の問題は、両言語の発音をより近い音素に類似した表記で対応付けるよう自動的に修正するプログラムを実装することで、より主観尺度に近い指標を表すことができるようになるはずである。したがって、今後、音韻類似性の客観的指標を利用するためには、どのような表音法で変換すれば、実際の主観的類似指標に近いかたちになるかを検討し、現段階のものに修正を加えていく必要があるだろう。

4.2.2 各指標による比較

次に、主観的類似性指標の側面から、語の分類について検討する。表5は、書字の一致別に音韻類似度の高い上位10語を一覧にしたものである。これを見ると、客観指標では音韻距離が大きいとみなした5以上の値 (L) の単漢字も含まれており、必ずしも主観指標と一致していないものがあることがわかる。これらの語を音素レベルで観察すると、前項と後項の漢字の初頭あるいは語末に、共通の音素、同じ調音点や調音法を持つ音素が含まれていることがわかる。例えば、「修理」は日本語では “shu ri”、中国語では “xiu li” と表記し、子音は日本語は “sh”、中国語は “x” で、後項の漢字は、日本語は “r”、中国語は “l” で表されている。しかし、“sh” と “x” は、

10) 音韻距離は値が大きいほど距離が大きく、類似度が低いことを示し、主観尺度は値が小さいほど類似度が低いことを示すので、類似傾向を示す場合の値は逆相関になる。

いずれも同様の調音法である摩擦音 [ç] であり、“r” と “l” は調音法に違いはあるものの、調音点は凡そ同じである。これらの子音は、両言語で表記が異なるとはいえ、日本語話者および中国語話者が特に異なる音素として弁別していない音であると思われる。また、上位 20 語のうち、/u/ と /ou/、/u/ と /iu/、/oo/ と /ao/、/en/ と /ian/、/ai/ と /ei/、/yo/ と /iao/ などのペアが入っており、連母音、半母音のペアは、似ていると判定される傾向があるようである。そして、「点」の “ten” と “dian”、「当」の “tou” と “dang” のように、/t/ と /d/ などの子音の有声／無声の違いも類似性にはあまり関与していなかったことは、中国語母語話者ならではの特徴であると考えられる。

表 5 書字の一致別による主観的類似度指標上位 10 位の語

No.	語	書字	音韻タイプ	前項	後項	主観指標	客観指標
1	安心	一致	HM	H	M	6.70	3
2	優秀	一致	HM	H	M	6.45	5
3	年代	一致	MH	M	H	6.39	3
4	誘導	一致	HH	H	H	6.39	2
5	修理	一致	MH	M	H	6.32	6
6	災難	一致	HH	H	H	6.26	2
7	要点	一致	HL	H	L	6.18	6
8	要求	一致	HM	H	M	6.18	5
9	散歩	一致	HM	H	M	6.13	6
10	分担	一致	MH	M	H	6.08	6
1	安易	不一致	HH	H	H	6.50	1
2	俳優	不一致	HH	H	H	6.18	3
3	有料	不一致	HL	H	L	6.13	6
4	案内	不一致	HH	H	H	6.13	2
5	面倒	不一致	MH	M	H	6.13	4
6	満点	不一致	HL	H	L	6.03	5
7	一瞬	不一致	MH	M	H	5.87	2
8	弁当	不一致	ML	M	L	5.82	9
9	辛抱	不一致	MH	M	H	5.82	6
10	改修	不一致	HM	H	M	5.82	6

表 6 は、書字の一致別に音韻類似度の低い下位 10 語を一覧にしたものである。こちらは、全体的に音韻距離の大きい値が多く含まれている傾向にあるが、前項と後項の漢字で、音韻的な類似度が高いとみなしたのも、若干含まれている。しかし、類似度の高いと判定されたものに比べると、ばらつきは少ないようである。

これらの語を音素レベルで観察すると、「頭痛」や「日程」など、それぞれ “zu tsu” と “tou tong”、“nit tei” と “ri cheng” のように、共通の音素が少ないものばかりだけでなく、語末に /ku/ や /tsu/ を含む語が多かった。表 6 だけを見ても、「逆」

“gyaku”、「約」“yaku”や「血」“ketsu”、「節」“setsu”など計20語のうち9語にこれらの音節が入っている。ただこれに関しては、/ku/ や /tsu/ が特殊な音素だというわけではなく、日中の音節上の違いが関係していると思われる。中国語は、基本的に1つの漢字の読みが1音節である。しかし、日本語では、「約」“ya”と“ku”のように、1つの漢字の読みが2音節あるものも少なくない。もともと、中国語母語話者が二字漢字であれば2音節という感覚を持っているならば、日本語の「逆転」の“gya”“ku”“ten”の3音節や「適切」“te”“ki”“se”“tsu”の4音節で構成されている語は、似ていないと判定される傾向があるのかもしれない。これに関しては、さらに全体の傾向を観察した上で、はたしてそうした傾向があるかどうか、中国語母語話者への意識調査などを行い、検証していく必要があるだろう。

表6 書字の一致別による主観的類似度指標下位10位の語

No.	語	書字	音韻タイプ	前項	後項	主観指標	客観指標
1	頭痛	一致	ML	M	L	1.21	8
2	輸血	一致	ML	M	L	1.34	9
3	伝言	一致	LM	L	M	1.46	10
4	逆転	一致	LL	L	L	1.47	11
5	節約	一致	LM	L	M	1.63	9
6	学歴	一致	LM	L	M	1.68	7
7	蓄積	一致	LM	L	M	1.71	9
8	徹夜	一致	LH	L	H	1.76	8
9	証拠	一致	LM	L	M	1.76	10
10	日程	一致	ML	M	L	1.79	9
1	帰宅	不一致	ML	M	L	1.18	9
2	欠如	不一致	LM	L	M	1.50	11
3	転落	不一致	LL	L	L	1.53	11
4	適切	不一致	LL	L	L	1.55	11
5	拒否	不一致	LL	L	L	1.61	8
6	日時	不一致	LM	L	M	1.68	6
7	地味	不一致	HM	H	M	1.68	5
8	活発	不一致	LL	L	L	1.71	11
9	食事	不一致	MM	M	M	1.74	7
10	育成	不一致	ML	M	L	1.79	9

5. おわりに

本研究は、Rのcbaパッケージで提供されるsdists関数(Buchta & Hahsler, 2016)を用いて算出した音韻距離を、主観的判定に依らない客観的な日中漢字語の音韻類似性指標として利用できるかどうかを検討した。そして、主観的な音韻類似性の指標と比較分析し、問題点と課題を探った。

本研究では、日本語のローマ字と中国語のピンインを使って両言語の音韻的な距離を数値化し、客観的な音韻類似性指標を算出した。しかし、客観指標と主観指標の相関係数は中程度 ($r = -0.49$, $p < .01$) であった。さらに、旧日本語能力試験の1級までの漢字二字語を、HH、HM、HL、MH、MM、ML、LH、LM、LL の9つに分類して前項と後項の漢字の音韻距離を基準に比較した。個々の語をみても、客観的および主観的類似性指標には大きな違いのあるものが散見された。以上のように、主観的指標と客観指標には、かなりのズレがあることが分かった。本研究の比較考察の結果、音声と表記の対応付けが両言語において異なっていることが大きな原因であることが分かった。

この問題を解決するために、音声的に近い音素を両言語で同じように対応付けて表記できるような表記変換自動プログラムを作って、客観指標の精度を上げていかなければならないであろう。どのように音素変換すれば、相互に対応付けができるのかについては、今後の大きな課題である。本研究で用いたプログラムでは、アクセントや声調を反映させていないが、これらの音声的情報が主観的な類似性判断に影響することは十分に考えられる。こうした情報を含むかどうかについても、検討の余地がある。

客観的指標の開発は、濁音・半濁音化、促音化などの音韻変化を持つ漢字の読みや、複数の読みを持つ漢字など、単漢字レベルでの音韻の差異の影響を見るための指標とすることができ、さらには、日中の漢字語の形態・書字・音韻の異同が引き起こすオンライン処理の仕組みを明らかにできる可能性が広がるはずである。今後も、引き続き日中漢字語における音韻的指標の整備を進め、日中の漢字語の形態・書字・音韻の異同がもたらす処理のメカニズムを明らかにしていきたい。

謝辞： 明治大学の小森和子先生、文化学園女子大学の三國純子先生には、調査に際し、多大なご尽力をいただきました。名古屋大学院生の熊可欣さんには、日中同形語データベースをご提供いただき、中国語の翻訳に関してもご協力いただきました。また、小森先生ならびに名古屋大学の野口裕之先生には、実験計画の過程で貴重なご助言をいただきました。また、調査にご協力していただいた学生の皆様なくしては、今回の試みも実現しませんでした。本研究に関わってくださったすべての皆様に、心より感謝を申し上げます。

本研究は、JSPS 科学研究費（若手研究 B）「形態と音韻の相互作用が第2言語の語彙処理にもたらす効果」（課題番号：15 K 16781, 代表：早川杏子）および挑戦的萌芽研究「動詞と共起する名詞群の検索エンジンの構築と読解および聴解に対する共起知識の貢献」（研究課題番号：16 K 13242, 代表：玉岡賀津雄）の助成を受けたもの

である。

参考文献

- 于劭贇・玉岡賀津雄 (2015) 「日韓中同形二字漢字語の品詞性ウェブ検索エンジン」『ことばの科学』 29, 43-61.
- 茅本百合子 (1995) 「同一漢字における中国語音と日本語の音読みの類似度に関する調査」『広島大学日本語教育学科紀要』 5, 67-75.
- 国際交流基金・日本国際協会 (2007) 『日本語能力試験出題基準〔改訂版〕』 凡人社
- 当銘盛之・費曉東・松見法男 (2012) 「日本語漢字二字熟語における中国語単語との音韻類似性の調査－同形同義語・同形異義語・非同形語を対象として－」『広島大学日本語教育研究』 22, 41-48.
- 朴ソンジュ・熊可欣・玉岡賀津雄 (2014 a) 「同形二字漢字語の品詞性に関する日韓中データベース概要」『ことばの科学』 27, 3-23.
- 朴ソンジュ・熊可欣・玉岡賀津雄 (2014 b) 「同形二字漢字語の品詞性に関する日韓中データベース」『ことばの科学』 27, 53-111.
- 水本篤・竹本理 (2008) 「研究論文における効果量の報告のために－基礎的概念と注意点－」『英語教育研究』 31, 57-66.
- 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, 25-51. 名古屋大学言語文化研究会
- 熊可欣・玉岡賀津雄・マンスブリッジ パトリック マイケル (2016) 「2 言語間の非選択的活性化は統語情報の処理においても起こるか－日中同形同義漢語動詞の受動態の処理を例に－」『認知科学』 23 (4), 395-410.
- Boytsov, L. (2011). Indexing methods for approximate dictionary searching: Comparative analysis. *Journal of Experimental Algorithmics*, 16, 1-1. DOI 10.1145/1963190.1963191
- Buchta, C., & Hahsler, M. (2016). cba: Clustering for business analytics. R package version 0.2-17. <https://cran.r-project.org/package=cba>
- Gusfield, D. (1997). *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. New York: Cambridge University Press.
- Miwa, K., Dijkstra, T., Bolger, P., & Baayan, H. (2014). Reading English with Japanese in mind: Effects of frequency, phonology, and meaning in different-script bilinguals. *Bilingualism: Language and Cognition*, 17 (3), 445-463. DOI: 10.1017/S 1366728913000576
- Tamaoka, K., Miyatani, M., Zhang, C., Shiraishi, M., & Yoshimura, N. (2016). Language-non-selective lexical activation without its use for sentential interpretation: An event-related potential (ERP) study on the processing of L1 Chinese and L2 Japanese sentences. *Open Journal of Modern Linguistics*, 6, 148-159. DOI: 10.4236/ojml.2016.62015
- Tamaoka, K., Makioka, S., Sanders, S., & Verdonschot, R. G. (in press). www.kanjidatabase.com: a new interactive online database for psychological and linguistic research on Japanese kanji and their compound words. *Psychological Research*. DOI: 10.1007/s 00426-016-0764-3
- Verdonschot, R. G., La Heij, W., Tamaoka, K., Kiyama, S., You, W-P., & Schiller, N.O. (2013). The multiple pronunciations of Japanese kanji: a masked priming investigation. *Quarterly Journal of Experimental Psychology*, 66, 2023-2038. DOI: 10.1080/17470218.2013.773050