日本語とベトナム語で共有される2字漢字語の客観的な音韻類似性指標の開発[1]

ホアーン ティ ラン フォン（HOANG, Thi Lan Phuong）[2]
玉岡賀津雄[3]
于 劭贇[4]

DOI: 10.18999/stul.33.133

要約：本研究では，日越両言語間で使用される2字漢字語の音韻類似性を，音韻的距離と音素類似性の 2 つの指標で客観的に計算した。日越で共通に使用される 2 字漢字語（訓読みを含まない）1,475 語の音韻的距離は，R で提供された cba パッケージ（Buchta & Hahsler, 2016）で，音素類似性は，R の phonosim パッケージ（Yu, 2016; Version 0.1）で，自動的に計算した。その結果，日越両言語間の音韻的距離の平均は 6.06 で，標準偏差は 2.52 であった。音韻的距離が小さければ小さいほど，音韻類似性が高くなることを示す。一方，日越両言語間の音素類似性の平均は 0.5 で，標準偏差は 0.2 であった。音素類似性は 0 から 1 まで分散し，1 に近いほど，音韻類似性が高いということを示す。音韻的距離と音素類似性の相関係数は非常に高く（N=1,475, r=-0.92, p<.001），2 つの指標が類似した指標であることを示した。日越両言語間の音韻類似性データベースは，ウェブ検索エンジン（http://kanjigodb.herokuapp.com, 詳細は，于・玉岡・ランフォン, 2019 を参照）で検索できるようになっている。この検索エンジンは，ベトナム人日本語学習者の 2 字漢字語の教授・学習カリキュラムおよび教材開発に貢献することが期待される。

キーワード：2字漢字語，ベトナム語，音韻類似性，音韻的距離，音素類似性

---

[1] English Title: Developing objective indexes of phonological similarities for two-kanji compound words used in both the Japanese and Vietnamese languages

[2] HOANG, Thi Lan Phuong (Graduate Student, Graduate School of Humanities, Nagoya University, E-mail: lanp2579@gmail.com)
[3] TAMAOKA, Katsuo (Professor, Graduate School of Humanities, Nagoya University, E-mail: ktamaoka@gc4.so-net.ne.jp)
[4] YU, Shaoyun (Graduate Student, Graduate School of Humanities, Nagoya University, E-mail: s.yu@nagoya-u.jp)

1. はじめに

日本語の漢字語は，多く中国語から借用された。さらに，江戸時代（1603-1867）から日本で作成された和製漢語も多く，これらは中国に逆輸入された（陳，2001；高島，2001）。本研究では，それらを含めて借用語とする。漢字は，日本語化した「音読み」で使用されるが，「訓読み」が追加され，一つの漢字が複数の発音を持つことが多い。一方，ベトナム語では，中国の唐朝（618-907）時代から新しい概念の語彙として漢語が輸入された。13 世紀から 19 世紀にかけ，ベトナムでチュノム[5]が普及して，中国語から借用された漢語はベトナム語化して，漢越語と呼ばれ，日常生活に一般的に使用されるようになった。現代のベトナム語はローマ字で表記されるが，中国語から借用された漢語（漢越語）は多く残っており，語彙全体の約 70%になる。漢字語は，アルファベット表記であっても 1 つの音節ユニットで表記される。以上のように，日本語とベトナム語は，漢語を共有しており，表記が異なっているものの，形態素としての漢字を介して音韻的に類似している場合が多い。

長野（2017）は，2 字漢字語の音韻的側面に焦点を当てて両言語間の音韻類似性について，7 段階尺度（1「全然似ていない」～7「非常に似ている」）で評定した。旧・日本語能力試験（国際交流基金・日本語国際協会，2002）の 2～4 級の 2 字漢字語とベトナム語には「同形語」および「異形語」の 2 種類がある。同形語は，たとえば，日本語の「集中」であれば，ベトナム語で同じ意味を持つ漢語は tập trung である。この語は，tập が「集」で，trung が「中」と対応しており，漢字の形態素単位で考えると，日本語と完全に一致している。一方，異形語は，日本語の「貯金」であれば，ベトナム語で同じ意味を持つ漢語は tiết kiệm である。しかし，この語は，tiết kiệm と表記され，tiết は「節」で，kiệm は「倹」に対応しており，日本語の「貯金」はベトナム語で「節倹」という漢字表記になり，漢字で表記した場合には両言語で一致しない。日本語の「貯金」は，ベトナム語の漢語音で表記すれば，trữ kim となるが，無意味語（あるいは非単語）である。

日越両言語の同形語は，旧・日本語能力試験出題基準〔改訂版〕（国際交流基金・日本語国際協会，2002）で出題される約 4,000 語の 2 字漢字語で，約 54%を占める（松田・タンティキムテュエン・金村・中平・三上，2008）といわれている。長野（2017）は，初めて日越両言語間の音韻類似性を計算した研究である。長野（2017）の調査では，同形語および

[5] チュノムは，語や形態素を意味的に表す表記であり，中世までベトナム語の「国語」だと考えられていた。

異形語の 110 語ずつを聴覚提示して，ベトナム人で日本語の学習経験がない 25 名に両言語間の音韻類似性を判断させた。その結果，同形語の音韻類似性の平均は 2.25（標準偏差は 1.30），異形語の平均は 2.35（標準偏差は 1.24）であった。日本語を学習したことがないベトナム人の主観的な判断であるため，個人差が大きいのではないかと予想される。さらに，結果をみると，同形語と異形語の音韻類似性の平均の差は小さく，わずか 0.1 にすぎない。標準偏差も同形語が 1.30 で，異形語が 1.24 であり，分散も小さい。以上の点から，長野（2017）の主観的な測定が，ほんとうに両言語の心的な音韻類似性を判定した指標になっているかどうか，疑問が残る。

そこで，本研究では，同形語か異形語かに関わらず，2 字漢字語の日越両言語間の音韻類似性を数値化して客観的に計算することにした。さらに，日越両言語の客観的音韻類似性データベースを作成し，それをウェブ上（http://kanjigodb.herokuapp.com，詳細は，于・玉岡・ランフォン, 2019 を参照）で，無料で自由に検索できるようにした。

2. 客観的音韻類似性の計算法

本研究では，日越両言語間の音韻類似性を音韻的距離と音素類似性の 2 つの指標で算出してデータベースを作成した。

2.1 訓令式ローマ字による音素表記

音韻類似性については，両言語間で比較できるようにするために，日越両言語に共通して使用される 2 字漢字語のベトナム語と日本語の発音を，音韻論の分野でも音素表記に使用されている訓令式ローマ字表記で表記した。長母音は，母音を 2 回繰り返した（たとえば，「公園」は，/kooen/など）。ベトナム語については，2 重子音の th, tr, ch は t の変化であるため，/t/で表記し，gi は z で発音するので，/z/で表記した。たとえば，「時間」は，ベトナム語で thời gian と表記するが，ローマ字表記では/toi zan/とした。また，c はベトナム語では k と同様の発音であり，日本語の訓令式には/c/がないので，いずれも/k/で表記した。qu は/kw/で示したので，「関心」はベトナム語で quan tâm であるが，/kwan tan/となる。さらに，語末の 2 重子音-ng, -nh, -m, -n は，日本語では撥音の/N/と似ているので，音素では/n/とした。たとえば，「東南」はベトナム語で đông nam であるが，/don nan/となる。ベトナム語の v は，日本語の訓令式に存在しないため，発音が最も近いと思われる/b/にした。同様に，l

は，/r/にした。たとえば，日本語の「絶望」は，ベトナム語では tuyệt vọng と表記されるが，ローマ字では/tuet bon/となる。なお，日本語には声調がないので，ベトナム語の声調は，音韻的距離には考慮しなかった。

2.2 音韻的距離の計算

音韻的距離とは，2 つのローマ字（アルファベット）表記の文字列がどの程度異なっているかを示す指標である。音韻的距離が小さければ小さいほど，音韻類似性が高くなると判断する。音韻類似性を客観的に評価する指標の多くが発音を記号化した文字表記を基にしている。その内，一般化レーベンシュタイン距離（generalized Levenshtein distance）が最も代表的な計算方法である（Levenshtein, 1966; Gooskens & Heeringa, 2004; Schepens, Dijkstra, & Grootjen, 2011; Schepens, Dijkstra, Grootjen, & van Heuven, 2013）。本研究では，一般化レーベンシュタイン距離で日越両言語間の音韻的距離を計算し，音韻類似性の 1 つの指標とした。

音韻的距離は，R の cba パッケージで提供される sdists 関数で自動的に計算できる（Buchta & Hahsler, 2017）。音韻的距離の具体的な計算では，ローマ字表記にした 2 つの文字列で，文字列を変形するための編集コストが最小になる「最適整列（optimal alignment）」を求め，比較を行う。各比較対象の文字が一致すれば 0，置き換えが必要であれば 2，挿入または削除が必要であれば 1 とする。こうした挿入，削除，置換のコストを合計した値が音韻的距離になる。実例として，ローマ字で表記された両言語間の音韻的距離を計算する先行研究は，Miwa, Dijkstra, Bolger & Baayen（2014）と早川・于・初・玉岡（2017）などがある。Miwa et al.（2014）は，ローマ字化した日本語の外来語と英語の語彙を比較して，両言語間の音韻的距離を計算した。また，早川・于・初・玉岡（2017）も，日本語と中国語の客観的な音韻類似性は，日本語の漢字をローマ字で書き直し，中国のローマ字のピンインを利用して両言語間の音韻的距離を測定した。

2.3 音素類似性の 2 種類の計算法

音素類似性は，音韻的距離と異なり，2 つの語の文字列の相違した部分ではなく，類似した部分に注目して音韻類似性を計算するもう 1 つの指標である。音韻的距離は，2 つの語の最適整列で異なった文字に対して所要の編集コストを計算するのに対し，音素類似性は 2 つの語の最適整列で共通した文字数を合計する。その上，音素類似性は語長に

よる影響を抑えるために，共通文字数を 2 倍し，2 つの語の長さの和で数値を標準化する（式 1）。音素類似性の値は，必ず最小 0，最大 1 の範囲に入る。たとえば，2 つの語は，類似性が皆無の場合は共通文字数が 0 であるため，音素類似性も 0 になる。一方，2 つの語が完全に一致している場合は，2 つの語長が必ず同じであり，共通文字数も語長と等しい値になるため，式 1 により音素類似性は最大値の 1 になる。音素類似性の指標は，R の phonosim パッケージで自動的に計算される（Yu, 2016; Version 0.1）。

$$\text{音素類似性} = \frac{\text{文字列 A と B の最適整列における共通文字数} \times 2}{\text{文字列 A の長さ} + \text{文字列 B の長さ}} \quad \text{（式 1）}$$

ヨーロッパ言語で音韻類似性を計測する研究は，基本的には語全体を比較対象としている。しかし，漢字語は 1 つの漢字が1つの音節に対応しており，原則としては形態素とも一対一の関係にあるのが特徴的である。そのため，漢字語の音韻類似性を検討する際には，語全体で類似性を計算する方法だけではなく，漢字1文字ずつという形態素単位で類似性を計算し，その平均を取る方法も考えられる。本研究では，漢字語全体を対象とする「漢字語彙音素類似性」と，形態素単位で求めた音素類似性を平均化する「漢字平均音素類似性」という 2 種類の方法で日越両言語の音素類似性を計算し，比較した。

$$\left(\frac{\text{同一音素数}_1 \times 2}{\text{（越）音素}_1 + \text{日・音素}_1} + \frac{\text{同一音素数}_2 \times 2}{\text{（越）音素}_2 + \text{日・音素}_2} \right) \Big/ 2$$

式 2　日越両言語の漢字平均音素類似性を計算する式

漢字平均音素類似性の具体的な計算方法としては，まず，2 字漢字語を前方と後方の漢字に分け，式1を用いて両言語間の音素類似性を形態素ごとに計算し，2 つの音素類似性の値を算出する。そして，前方と後方の漢字の音素類似性を加算し，それを漢字の数，つまり2で平均した結果を計算する（式2）。この漢字平均音素類似性の結果は，語全体で計算した音素類似性と同様に，0 から 1 までの範囲で変化する。たとえば，前方と後方の漢字のどちらも両言語で類似性がない場合は，形態素ごとの類似性がいずれも 0 であり，それを平均した値も当然 0 になる。また，前方と後方の漢字が両言語ですべて一致

した場合は，形態素ごとの類似性がいずれも最大の 1 であり，その合計値を漢字数で平均すると，1 になる。

音素類似性の計算について，日本語の 2 字漢字語の「海外」（/kai gai/）は、ベトナム語で hải ngoại”（/hai goai/）を比較する。式 1 の音素類似性の計算を使用して「海外」の日越両言語間の音素類似性を計算すると，1つ目の漢字の/kai/と/hai/は，類似した音素 2 つ（/a/と/i/）であり，両言語間の文字数は 6 つ（/k/, /a/, /i/, /h/, /a/, /i/）である。つまり，音素類似性は（2×2）/（3+3）=0.667 となる。同じように，2 つ目の漢字の/gai/と/goai/の場合は，類似した文字数は 3 つ（/g/, /a/および/i/）であり，文字数は 7 つ（/g/, /a/, /i/, /g/, /o/, /a/, /i/）であり，（3×2）/（3+4）＝0.857 となる。2 字漢字語の「海外」の音素類似性は，2 つの形態素（漢字）の平均を取って，（0.667＋0.857）/2＝0.762 となる。この値は，1 にかなり近いため，音韻類似性が高いといえよう。

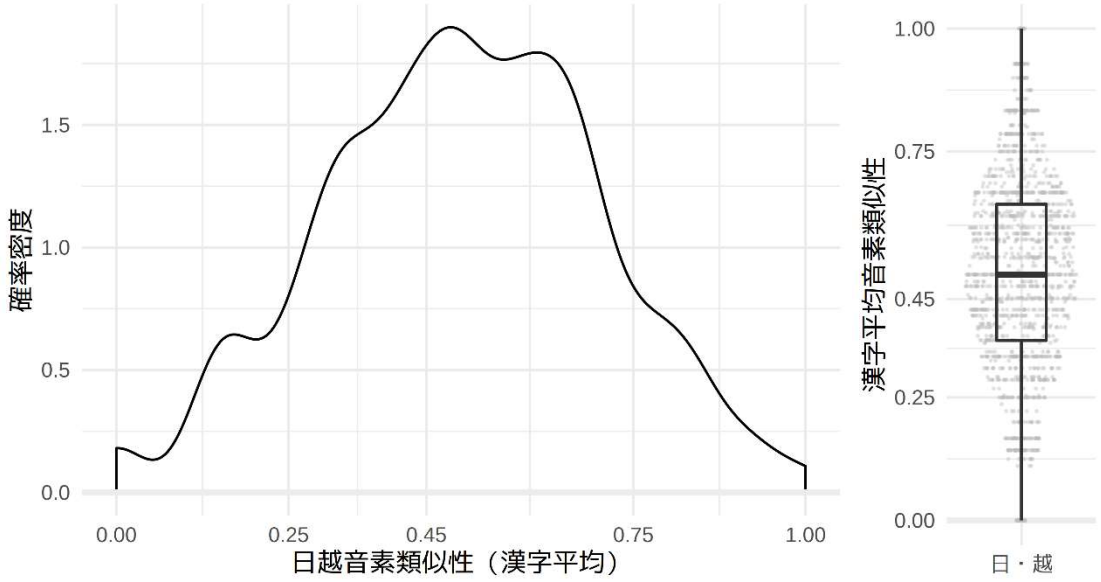

図 1. 日越両言語の 2 字漢字語の漢字平均音素類似性の分布

注:左側の図がカーネル密度推定で，右側が箱ひげ図および散布図である。

語レベルの音素類似性は，日越両言語で共通に使用した 2 字漢字語（1,475 語）の平均は 0.5 で，標準偏差は 0.2 であった。分布は図1に示したように，尖度は-0.25 で，歪度は-0.96 である。音韻類似性の指標としては細かい数値で示される。漢字平均音素類似性は，日越両言語で共通した全 2 字漢字語で，平均は 0.50 で，標準偏差は 0.20 であった。分

布は，尖度は 2.74，歪度は-0.10 で，正規曲線ではないが，対称的な分布をみせた。

2.4 音素類似性 2 種類の特性

漢字平均音素類似性と漢字語彙音素類似性の対応関係を検討するために，2 つの音素類似性の散布図を図 2 に描いた。2 種の音素類似性はほぼ直線的な関係にあり，対応が非常によいことが示された。なお，相関係数も 0.95 と極めて高かった。しかし，この 2 種の音素類似性を詳細に比較すると，漢字平均音素類似性のほうが漢字の音韻的特徴により適合していることが分かる。なぜなら，漢字語彙音素類似性は，語全体で計算されるため，個々の漢字に音韻的類似性がまったく認められなくても，最適整列を求めるアルゴリズムは漢字の境界線を無視して，語全体に共通した文字を見つけ出すからである。

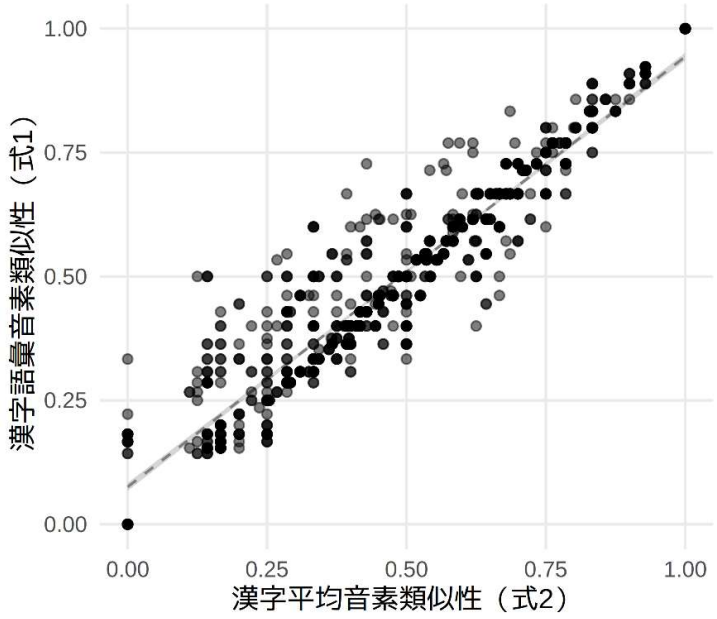

図 2.　漢字語彙音素類似性と漢字平均音素類似性との関係

注: 色が濃い点は，2 つ以上同じ数値が重なっていることを示す。つまり，複数の 2 字漢字語において，2 種類の音素類似性が同じ数値を示した場合である。

2 つの漢字を一緒にして 1 つの語として計算した場合の共通文字数は，1 つの漢字が 1 つの音節に対応するという漢字語の音韻的構造から乖離しており，真の音素類似性を評価するには適切ではない。図 2 の左下にある，漢字平均音素類似性が 0 であるのに，漢字語彙音素類似性が 0 ではないという語はこうした状況を反映している。

表 1 「図表」の語全体の最適整列の算出例（cba パッケージ）

日本語	ベトナム語	変形操作
Z	-	削除
U	D	置換
H	-	削除
Y	-	削除
O	O	一致
O	-	削除
-	B	挿入
-	I	挿入
-	E	挿入
-	U	挿入

たとえば，「図表」という漢字語は本研究のデータベースにおいて日本語での音素表記が/zu hyoo/であり，ベトナム語での音素表記が/do bieu/になっている。日本語の「図表」には，共通の音素がないため，漢字平均音素類似性は 0 と計算される。それにもかかわらず，「図表」の漢字語彙音素類似性は 0.17 でゼロではなかった。その理由は、語全体で音韻類似性を求めると，R の cba パッケージは，文字列の変形操作のコストが最小になるように，漢字単位の音韻的構造を無視した最適整列を求めてしまうためである（表 1）。日本語の後方の/hyoo/とベトナム語の前方の/do/では，/o/が共通している。一方，漢字平均音素類似性は漢字ごとに計算されるため，こうした問題点を回避できる。このように，漢字ごとの音素類似性を計算し平均化をする漢字平均音素類似性は，漢字語の音韻的構造を考慮したよりよい計測法であるといえよう。

2.5 日越両言語の音韻的距離の分布

日越両言語の音韻的距離の計算は，たとえば日本語の「家族」であればローマ字による音素表記では/ka zoku/である。同じ語は，ベトナム語では，gia tộc と表記され，音素表記では/za tok/となる。両者を比較すると，1 つ目の漢字の/ka/と/za/は，/k/を/z/に置き換えるため 2 で，/a/が一致しているので 0 である。2 つ目の漢字の/zoku /と/tok/は，/z/を/t/に置き換えるため 2 で，次の/o/と/k/は両言語で一致しているので 0，最後の/u/は挿入（または削除）であるため 1 になる。つまり，1 つ目の漢字の音韻的距離は 2+0=2 で，2 つ目の漢字の音韻的距離は 2+0+0+1=3 である。漢字に基づいた計算では，日本語の「家族」と gia tộc の

音韻的距離は，2+3=5 となる。一方，語を単位とする音韻的距離の計算では，日越両言語で共通に使用される 2 字漢字語（2,058 語の中の 1,475 語で，71.67%）で，平均は 6.05 で，標準偏差は 2.52 であった。分布の尖度は 2.78 で，歪度は 0.14 であった。分布は図 3 に示した。

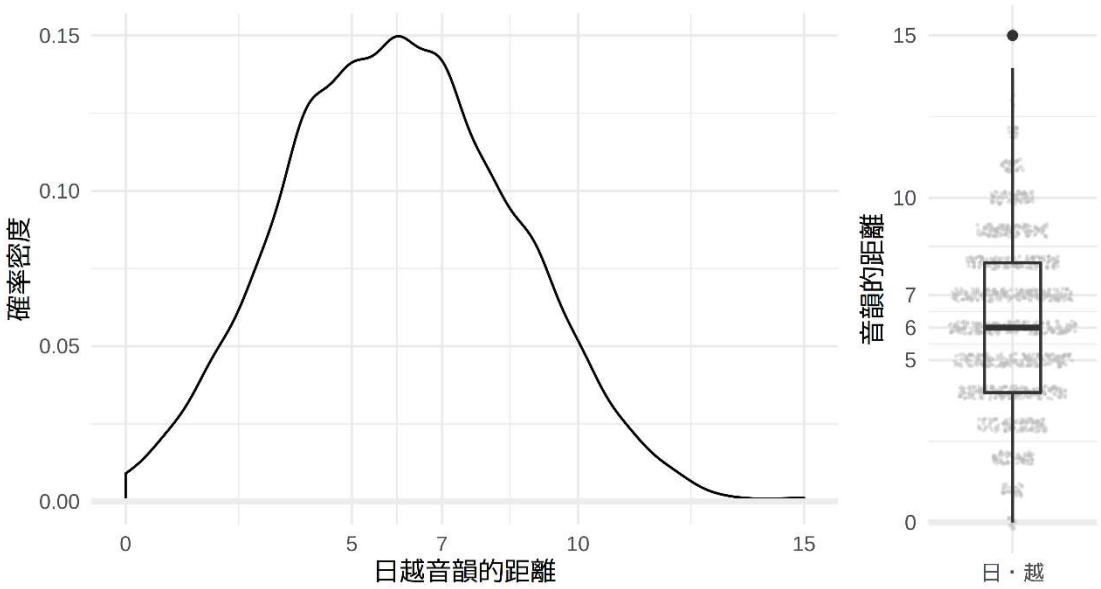

図 3. 日越両言語の 2 字漢字語の音韻的距離の分布

注:左側の図がカーネル密度推定で，右側が箱ひげ図および散布図である。

2.6. 音韻的距離と音素類似性の相関係数

日越両言語間の音韻類似性は，音韻的距離と音素類似性の 2 種類の方法で計算した。日越両言語の音韻的距離と漢字平均音素類似性のピアソン積率相関係数は，r=-0.88，p<.001（n=1,475 語）であった。相関係数 r の絶対値は，0.8 以上であったので，両指標が非常に類似していることを示唆した。

3. データベースの公開

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）の 4 級から 2 級までの 2 字漢字語のデータベース（朴・熊・玉岡，2014a, 2014b; 熊・玉岡，2014; 于・玉岡，2015）を資料として，日越両言語間の音韻類似性を計算した（日本語で音読み

のない語は除外）。データベースでは，日本語の2字漢字語の2,058語の中で，日越両言語で共通に使用される1,475語であり，全体的の71.67%を占める。

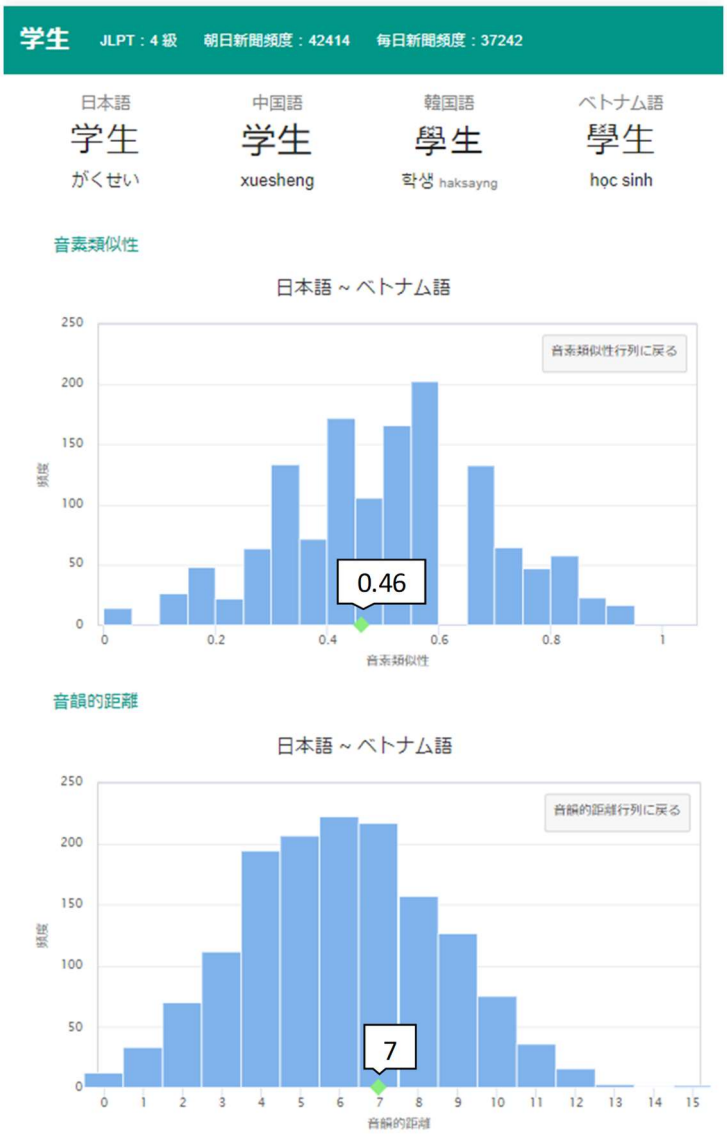

図 4. 2 字漢字語の「学生」の漢字語彙音素類似性と音韻的距離で表した音韻類似性

本研究では，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）の 4 級から 2 級までの 2 字漢字語のデータベース（朴・熊・玉岡, 2014a, 2014b; 熊・玉岡，2014； 于・玉岡，2015）を基にして，日越両言語間の音韻類似性データベースを作成した。音韻類似性の指標は音韻的距離と音素類似性から測定され，Web 上で公開された（『日韓中越同形二字漢字語データベース』http://kanjigodb.herokuapp.com）。現時

点では，音素類似性の情報は語全体を対象に計算した漢字語彙音素類似性となっている。たとえば，2 字漢字語の「学生」を調べた結果，図 4 のような画像で，日越両言語間の音素類似性と音韻的距離の情報をみることができる。図 4 をみると，日越両言語間の漢字語彙音素類似性の分布では，0.46 の値で，分布の真ん中くらいに位置している。中程度の音素類似性を持つと考えられる。日越両言語の音韻的距離の分布では 7 となり，やはり中程度に位置する。この検索は，現代のベトナム語の表記とチュノムの表記でもできる。『日韓中越同形二字漢字語データベース』では，日越ばかりでなく，日中，日韓，中越，中韓，韓越のそれぞれの言語対の音韻類似性の情報を公開している。

5. おわりに

日本語と現代のベトナム語は表記が異なっているが，両言語で共通した 2 字漢字語彙の音韻類似性を計算するために，訓令式ローマ字表記で，両言語の発音を音素表記した。そして，日越両言語間の音韻類似性を客観的に計算する方法を提案し，2 つの指標で算出した，音韻類似性データベースを作成した。またこのデータベース用に，ウェブ検索エンジン（于・玉岡・ランフォン，2019 を参照）を開発した。しかし，ベトナム語で使用される音素は，日本語に存在しない発音もあり，発音が大きく異なる両言語間で音韻類似性を測定するのは非常に難しい。本研究では，それらの微妙な違いを考慮しなかった。この点は今後の検討課題である。

［参考文献］

陳力衛（2001）「和製漢語の形成とその展開」東京：汲古書院

国際交流基金・日本国際教育協会（2007）『日本語能力試験出題基準（改訂版）』（第 4 版）東京：凡人社

高島俊男（2001）『漢字と日本人』東京：文藝春秋

早川杏子・于劭贇・初相娟・玉岡賀津雄（2017）「日中二字漢字語における客観的音韻類似性指標 ―主観的音韻類似性指標との比較―」『関西学院大学日本語教育センター紀要』6, 21–34.

松田真希子・タンティキムテュエン・ゴミントュイ・金村久美・中平勝子・三上喜貴（2008）

「ベトナム語母語話者にとって漢越語知識は日本語学習に どの程度有利に働くか―日越漢字語の一致度に基づく分析―」『世界の日本語教育』18, 21-33.

長野真澄（2017）「日本語漢字単語とベトナム語漢越音における音韻類似性調査」『広島大学日本語教育研究』27, 35-41.

朴善娟・熊可欣・玉岡賀津雄（2014a）「同形二字漢字語の品詞性に関する日韓中データベースの概要」『ことばの科学』27, 53-111.

朴善娟・熊可欣・玉岡賀津雄（2014b）「同形二字漢字語の品詞性に関する日韓中データベースの概要」『ことばの科学』27, 3-23.

玉岡賀津雄（2005）「命名課題において漢字 1 字の書字と音韻の単位は一致するか」『認知科学』12（2）, 47-73

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

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.

于劭贇・玉岡賀津雄・ホアーン ティ ラン フォン（2019）「日韓中越 4 言語における2字漢字語の音韻類似性に関するデータベースおよび検索エンジンの構築」『ことばの科学』33, 75-93.

熊可欣・玉岡（2014）「二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.

Buchta, Christian and Michael Hahsler (2017). cba: Clustering for business analytics. R package version 0.2-19. https://cran.r-project.org/package=cba .

Gooskens, Charlotte and Wilbert Heeringa (2004). Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language variation and change, 16,* 189-207.

Levenshtein, Vladimir I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady, 10,* 707-710.

Miwa, Koji, Ton Dijkstra, Patrick Bolger and R. Harald Baayen (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.

Schepens, Job Johannes, Ton Dijkstra and Franc Grootjen (2011). Distribution of cognates in

Europe as based on Levenshtein distance. *Bilingualism: Language and cognition, 15,* 157-166.

Schepens, Job Johannes, Ton Dijkstra, Franc Grootjen and Walter J. B. van Heuven (2013). Cross-Language Distributions of High Frequency and Phonetically Similar Cognates. *PLO S ONE,* http://dx.plos.org/10.1371/journal.pone.0063006

Yu, Shaoyun (2016). phonosim: An experimental R package for calculating phonological similarity. https://github.com/rongmu/phonosim

Developing objective indexes of phonological similarities for two-kanji compound words used in both the Japanese and Vietnamese languages

HOANG, Thi Lan Phuong
(Graduate Student, Graduate School of Humanities, Nagoya University, Japan)
TAMAOKA, Katsuo
(Professor, Graduate School of Humanities, Nagoya University, Japan)
YU, Shaoyun
(Graduate Student, Graduate School of Humanities, Nagoya University, Japan)

Abstract: In this study, phonological similarities of two-kanji compound words used by both the Japanese and Vietnamese languages were objectively calculated by two indexes of phonological distance and phoneme similarity. The phonological distance of 1,475 (71,67%) of two-kanji compound words used in both languages (71,67%) was automatically calculated by R cba package (Buchta & Hahsler, 2016), and by the R phonosim package (Yu, 2016; Version 0.1) for phoneme similarity. The mean of phonological distance was 6.06 with a standard deviation of 2.52. This index indicates that the smaller the phonological distance, the higher the phonological similarity. On the other hand, the mean of phoneme similarity between two languages was 0.5 with a standard deviation of 0.2. Phoneme similarity was dispersed from 0 to 1, and the closer it was to 1, it showed that phonological similarity was higher. The correlation between phonological distance and phoneme similarity was very high (N=1,475, r=-0.92, p<.001): the two indicators were very similar. In the indexes of phonological similarities between the Japanese and Vietnamese languages can be found on the web search engine (http://kanjigodb.herokuapp.com/, for more information, see Yu, Tamaoka & Hoang, 2019). This search engine is expected to contribute to developing the learning/teaching curriculum and materials of two-kanji compound words for Vietnamese learners of Japanese.

Keywords: two-kanji compound words, Vietnamese, phonological similarity, phonological distance and phoneme similarity