

漢字データベースを用いた日中韓越漢字語の学習教材の開発 ～音韻的類似性を用いたアプローチ～

Development of Learning System with Kanji Database in Japanese, Chinese, Korean, Vietnamese;
An Approach by Using Phonological Similarity

ドウ・フィ・チュオン, 山崎 恭司, 陳 守誠, 江見 圭司 (京都情報大学院大学)

Do Phi Truong, Yasushi Yamazaki, ShouCheng Chen, Keiji Emi

(The Kyoto College of Graduate Studies for Informatics)

Abstract

The similarity of the consonant Grimm's law is known between German words and English words. By using Grimm's law, English-speaking people can easily understand German words.

Correspondences such as Grimm's law have been found for the pronunciations among Chinese, Japanese, Korean, and Vietnamese Kanji. We have built a database in order to easily learn the correspondences. We will report effective learning of Chinese, Japanese, Korean, and Vietnamese Kanji words by using our database.

1. はじめに

1.1 はじめに

外国語を勉強するのは難しい。例えばベトナム人にとってはローマ字アルファベットを使っているのが漢字を勉強するのが大変である。昔から漢字に基づいてベトナムでは漢越というものが使われてきた。現在は文字がローマ字に変わり、漢越の字はなくなったが、漢越の発音はまだ残っているものをローマ字で書いている。そこで、漢越発音の勉強をすればベトナム人にとっては漢字が覚えやすくなるだろうと考えた [1]。しかし、ベトナム語漢字語の音韻的類似性に関する研究はまだ少なく、漢字語の音韻的類似性を調べる辞書やアプリケーションはまだない。

ヨーロッパでは、例えばゲルマン語系のドイツ語単語と英語単語の間にはグリムの法則などの子音の類似性が認められ、それを用いれば英単語既習者がドイツ語単語を簡単に理解することができる [2]。例を挙げる。(独)はドイツ単語、(英)は英単語を示す。

・(独) T -> (英) D, (独) G -> (英) Y

Blut -> blood

Tag -> day

Gestern -> yesterday

・(独) D -> (英) TH, (独) B -> (英) F

Broder -> brother

Ding -> thing

Halb -> half

Selbst -> self

Dieb -> thief

同様のことは、中国語の漢字音から周辺国である韓国語、ベトナム語、日本語へのそれぞれの漢字音の対応があることがある程度は分かっており、それを理解すれば既習言語の漢字音からそれぞれを理解することが可能である [3]-[5] という考えから、筆者(江見)は 2009 年にデータベースを作成した [6] [7]。ところが、このデータベースは Microsoft 社の古い技術であるデータバインドを用いている [7] ので、本稿執筆時点の環境では動作しないことが多い。そこで、本研究では、漢字語研究者及びそれに興味を持っている人のために、日中韓越の漢字語の音韻的類似性を最新のコンピュータ環境でも動作する、研究・学習するウェブシステムを開発することにした。特に開発者に中国語話者(陳守誠)とベトナム語話者(ドウ・フィ・チュオン)を加えて、漢字文化圏、日本・中国・韓国・ベトナム、各言語の漢字語彙を学習できるように開発した [8]。特に文献 [8](b) はドウ・フィ・チュオン、山崎恭司、陳守誠の修士研究をベースにしているため、詳細な手順やデータベースの使い方などはそちらを参考にし

てほしい。

1.2 漢字語と漢字

漢字はそれぞれが意味を持ち、個々が音節に対応していて、言語学用語では形態素と呼ばれる。形態素とは意味を持つ表現要素の最小単位とされ、意味を持たなくなる単位まで言語を分割した、音素としての一つ一つを指す（単語もしくは単語の構成要素でもある）。現代中国語では単語の大部分は2字以上の漢字を組み合わせたものである。

我々の考えるところ、中国において、漢字は一つの文字を単位として重んじてきたが、日本では言葉の意味を重んじてきたため、音（発音）は違っても意味が同じ言葉が存在する。まず、日本では一つの漢字に音読みと訓読みがあることがある。さらに、例として「あめふり」、「あまやどり」や「きりさめ」などの言葉がある。「あめ」、「さめ」、「あま」などは「雨」という同じ意味に分類できる形態素にあたる。中国、韓国では基本的に一字一音なので、このような点が日本語の特徴の一つとして挙げられるが、このような訓読みについては本システムでは深入りしていない。

さて、漢字音の古い形態を残しているのが韓国語漢字音や広東語漢字音である。今回は、広東語は見送るので、韓国語漢字音を核にして、現代中国語漢字音、ベトナム語漢字音の順に開発していくことにした。文献[6]にあるように、日本語漢字音は歴史的仮名遣いを使った[9]。

本論文は以下のように構成した。まず、第2章では「各国の漢字語」の特徴を韓国語、中国語、ベトナム語の順に講述する。第3章と第4章で、開発したデータベースとデータベースマネジメントシステムについてそれぞれ述べる。第3章では「日中韓越漢字データベース」のデータベース設計を詳しく述べる。第4章では「日中韓越漢字システム」として、利用したJavaScriptの技術とユーザーインターフェースのデザインについて詳しく述べる。些末な部分については割愛したので、文献[8](b)を参照して欲しい。第5章では「学習教材としてのデータベース利用と検証」を韓国語、中国語、ベトナム語の順に講述するが、学習教材として使い方の詳細は割愛するので、文献[8](b)を参照して欲しい。

2. 各国の漢字語

2.1 韓国の漢字語

2.1.1 韓国の漢字語の概要

中国語に由来するものが主となる漢字語である。それらは古来中国語に関するものが多いのと同時に、現代の各漢字語間で共通するものも多くある。文献[6]でも述べたが、韓国漢字音から日本漢字音（歴史的仮名遣い）の簡単な対応を示す。

[-m] → 「～ム」(現代仮名遣いでは「～ン」に変化)

[-n] → 「～ン」

[-ŋ] → 「～イ」「～ウ」, (唐音)「～ン」[12]

[-p] → 「～フ」(現代仮名遣いでは「～ウ」, 「～ツ」に変化, 無声音の前では「～ッ」に変化)

[-k] → 「～ク」・「～キ」・(無声音の前で)「～ッ」

[-t] → 「～ツ」・「～チ」・(無声音の前で)「～ッ」

韓国漢字語研究は文献[10]が詳しく、辞書は文献[11]などを利用した。本研究の韓国漢字語研究及びデータベースは文献[12]・[14]によるものである。

天下 (천하)	[チョンハ]	「天下」
結婚 (결혼)	[キョルホン]	「結婚」
愉快 (유쾌)	[ユクウェ]	「愉快」
學問 (학문)	[ハンムン]	「学問」
四面楚歌 (사면초가)	[サミョンチョガ]	「四面楚歌」
漁夫之利 (어부지리)	[オブヂリ]	「漁夫の利」

漢字1字の漢字語の中には、固有語と同様に浸透しているものが多い。

山 (산)	[サン]	→ 「山」
江 (강)	[カン]	→ 「川」
冊 (책)	[チェク]	→ 「本」

2.1.2 日本起源の韓国の漢字語

近代、日本が支配した時代に日本から多くの語彙が朝鮮語に漢字語として取り入れられた。日本から伝わった漢字語には数種の類型がある。

日本独自の漢語は、西欧から伝わった思想などの用語を訳した漢字語などが多くある。

個人 (개인)	[ケイン]	「個人」
主權 (주권)	[チュコン]	「主權」

義務 (의무)	[ウイム]	「義務」
鉛筆 (연필)	[ヨンピル]	「鉛筆」
電話 (전화)	[チョンホワ]	「電話」
博物館 (박물관)	[パンムルグワン]	「博物館」

上のような漢字語彙は中国でも取り入れられ中国語で使われるが、中国語で独自に語彙が作られた場合には、朝鮮語と日本語の間に同じものが存在するが、中国語との間では同じになるものはない。

飛行機 (비행기 피헝기)	飛機 フェイチ
汽車 (기차 [キチャ])	火車 ホウチャ
野球 (야구 [ヤグ])	棒球 パンチュ
会社 (회사 [ホイサ])	公司 ゴンス
写真 (사진 [サデン])	照片 チャオピエン

日本では「同音の漢字による書きかえ」が規定され、一部で日本独自の漢語において表記変更がされた。しかし、朝鮮語では旧漢字での表記での読みのままである。

弘報 (홍보 [ホンボ])	広報
慰藉料 (위자료 [ウィザリョ])	慰謝料
刺戟 (자극 [チャグク])	刺激
破毀 (파훼 [パフウエ])	破棄
手帖 (수첩 [スチョプ])	手帳
洗滌 (세척 [セチョク])	洗浄

2.1.3 日本語における訓読み熟語

訓読み熟語の中で朝鮮語になったものには、重箱読みや湯桶読みも存在する。

建物 (건물)	[コンムル]	「建物」
賣上 (매상)	[メサン]	「売り上げ」
船着場 (선착장)	[ソンチャンチャン]	「船着場」
不渡 (부도)	[プド]	「不渡り」〔重箱読み〕
待合室 (대합실)	[テハプシル]	「待合室」〔湯桶読み〕

動作的な意味を表す単語は後ろに「する」に当たる接尾辞「하다」[ハダ]を付け動詞化される。以下の例は後ろに「하다」が付けられる単語である。

貸出 (대출)	[テチュル]	「貸し出し」
上廻 (상회)	[サンフエ]	「上回ること」
組立 (조립)	[チョリプ]	「組み立て」

取扱 (취급)	[チュイグプ]	「取扱い」
割引 (할인)	[ハリン]	「割引」

意味が変わった漢字語をあげる。

生産 (생산：センサン)「生産」：「子供を生む」⇒「製品を作り出す」
放送 (방송：パンソン)「放送」：「罪人を放す」⇒「broadcasting」
自然 (자연：チャヨン)「自然」：「おのずから・ひとりでに」⇒「おのずから・ひとりでに + ‘nature’ の訳語」
發明 (발명：パルミョン)「發明」：「罪人が自ら事情を公開する」⇒「存在しなかったものを作り出す」

漢字語から朝鮮独自の語になったものもある。

가난 [カナン]	「貧乏」	艱難 (간난)
성냥 [ソンニャン]	「マッチ」	石硫黄 (석류황)
감자 [カンチャ]	「じゃがいも」	甘藷 (감저)
설매 [ソルメ]	「そり」	雪馬 (설마)
고추 [コチュ]	「唐辛子」	苦椒 (고초)
요 [ヨ]	「敷布団」	褥 (욕) - [しとね]
김치 [キムチ]	「キムチ」	沈菜 (침채)
재주 [チェヂュ]	「才能」	才調 (재조)
나인 [ナイン]	「女官」	内人 (내인)
지렁이 [チロンイ]	「ミミズ」	地龍이 (지렁이)
나중 [ナヂュン]	「のち」	乃終 (내종)
짐승 [チムスン]	「獣」	衆生 (중생)
대충 [テチュン]	「大凡」	大總 (대총)
천둥 [チャンドゥン]	「雷」	天動 (천동)
보배 [ボメ]	「宝」	寶貝 (보패)
초 [チョ]	「蠟燭」	燭 (촉) - [ともしび]
사냥 [サニャン]	「狩り」	山行 (산행)

2.1.4 韓国におけるハングル

朝鮮の人々は独自の文字を持っていなかった。朝鮮半島には日本より数百年早く漢字が中国から伝わり、それ以来文書は漢字で書き表していた。漢字を利用して十分な表記はできず、漢字を学べる者が限定的であったため、漢字が読み書きできない者には表記する手段がなかった。

世宗大王(朝鮮王朝4代目)が一般民衆のために指揮をとり、ハングルが作成された。そうして朝鮮語を書き記せるようになった。知識人や官僚たち

は漢字を重んじてハングルを排斥していたため、歴史的にハングルは必ずしも多くの民衆に使われていたわけではなかった。

韓国における漢字の廃止は1948年に大韓民国が成立し、「公の文書は例外なくハングルで記載する」という法律の制定によって始まる。漢字の廃止後、韓国では新聞・雑誌・書籍および教科書、街の看板でも、漢字は見られなくなっている[12]。

2.1.5 韓国における漢字の廃止

漢字語で表されていた概念などを示す専門用語その他、古くから漢語で記されてきた貴重なたくさんの言葉が、一般的には次第に使われなくなっていく。多くの評論・研究論文や新聞・雑誌の記事などにおいて、ほぼすべてが書き言葉であふれ、驚異的な語彙の貧困化に陥ったといえる。

現在の韓国において、文章は単純化しているといわれ、言葉の奥行きは浅くなった。例では「シャ」＝(사:サ)が同じである。

社会	=	사회:サフェ
射撃	=	사격:サギョク
写真	=	사진:サデン
斜面	=	사면:サミョン
謝罪	=	사죄:サヂョイ

韓国の漢字語で発音・意味がほぼ同じものをあげておく。

意味 (의미):	ウィミ
都市 (도시):	トシ
計算 (계산):	ケサン
酸素 (산소):	サンソ
散歩 (산책):	サンチェク
看板 (간판):	カンパン
家族 (가족):	カジョク
約束 (약속):	ヤクソク
無理 (무리):	ムリ
温度 (온도):	オンド
算数 (산수):	サンス
気分 (기분):	キブン

韓国語を解釈する場合、以下の特徴がある。

- ・同じ意味で使われる場合、漢字に変換すれば素早く意味をつかめる。

- ・使用頻度を優先して変換を試みるのが効果的である。
- ・韓国語の場合、同音異義語が多くなることに留意する。

以上のような特徴を踏まえて、現在日本では多数の韓国語漢字音を利用した学習書が存在する[12]・[16]。また、中国でも韓国語漢字音を中国語漢字音に対応させた学習書[17]がある。これらの要約はブログ記事[18]などで読むことができる。

2.2 中国の漢字語

2.2.1 簡体字の歴史

市販の中国語参考書によると、清末の1909年、陸費逵が「教育雑誌」創刊号に「普通教育当採用俗体字（普通教育に俗字を採用すべきだ）」という論文を発表したことが簡化運動の始まりとされる。五四運動時代の1920年、銭玄同は「新青年」に「減省漢字筆画的提議（漢字の筆画を減少させる提案）」を発表し、1922年には陸費逵らと連名で国語統一籌備会に常用の漢字すべての筆画を減少させる提案をしている。1934年には国語統一籌備会に簡略された字体の収集を提案し、翌1935年には2400字余りの「簡体字譜」の草案が編まれた。一方で、国民政府教育部でも324字の「第一批簡体字表」を公布したが、1936年に「不必推行」（普及の必要なし）の命令が出され実施されることはなかった。その後、簡略字体の収集が盛んになり1937年には字体研究会が1700字の「簡体字表第一表」を発表している。

中華人民共和国が建国された3年後の1952年、簡略字体の議論を受けて、漢字研究の機関として「中国文字改革研究委員会」が設立された。1954年に憲法が制定され、政務院が改組されるなど新体制への変化の中で、中国文字研究改革委員会も中国文字改革委員会に改名し、1955年に「漢字簡化方案草案」を発表した。翌年の1956年1月、この草案を基に「漢字簡化方案」が国務院より公布され1964年に「簡化字総表」にまとめられた。

1977年、中国文字改革委員会は新たに「第二次漢字簡化方案草案」を発表し、8年間の試行で廃止された。

本研究では、簡体字と正体字（繁体字）の両方をデータベースでは使用した。

2.2.2 ピンイン

ピンインは、中国語で音節を音素文字に分け、ラテン文字化して表記する発音表記体系を指す。一般的な場合この語は、漢語拼音と呼ばれる1958年に中華人民共和国が制定した漢語拼音方案という表記法、またはそれに基づく漢語拼音字母という文字を指す。なお、台湾の中国語は中国大陆と同様であるが、読み方はラテン文字（つまりピンイン）は使用せず、文字は繁体字を使っている。本研究のデータベースでは中国語漢字音はこのピンインを使用した。

2.2.3 中国語漢字音

中国語漢字音は研究成果があるので、市販の書籍に頼れば、十分な知見 [19]-[24] を得ることができる。日本語漢字音から中国語を調べることができる辞書がある [19]。特に、日本語の呉音は文献によって解釈が分かれている。現代中国語漢字音と呉音（歴史的仮名遣い）との対応がもっとも規則的になるので、我々は藤堂明保の著書 [22]-[24] を参照した。

日本語漢字音と中国語漢字音との対応をさせた文献も存在する [25]。謝国芳 [25] は日本語漢字音の慣用音についての知見がないので、日本語漢字音と中国語漢字音が対応がつかない理由が説明できないとしている。

2.3 ベトナムの漢字語

2.3.1 ベトナムの漢字語構成

ベトナムの歴史を見ると、8世紀から文章には漢字語を用いていた。それから17世紀から19世紀の間に、ベトナム語の標準語が生まれてきた。その当時、ベトナム人は多くの言語を使用していた。例えば文章には4つの文字が使われていた。チュノム字（ハノム字）を使ったり、フランス語のアルファベットを使ったり、もちろん標準のアルファベットも使用していた。

1945年に戦争が終わってから全国標準語だけを使うことになってきた。ベトナム語の標準語の中に、漢字語を借りていた言葉を使うためにアルファベットで表記するものがある。これが漢越語である。構成でいえば、漢越語約45%、元来からのベトナム語は約40%、外来語は約15%と言われており、ベトナム語の語彙の多くを占めることがわかる。

ベトナム語漢字音の日本における研究は三根谷 [27] の研究が有名である。漢字音に言及している辞書として文献 [28] を利用した。それ以外にはベト

ナム人が書いた文献 [29]-[32] も利用した。

日本の漢字でよく使われる常用漢字とほぼ同じで約2000文字ある。それぞれは漢越語で書き換えることができるが、実は単純ではない [10]-[13]。以下では左に漢字語、右に漢越語を対応させるように例をあげる。

字	Tự
学	Học
日	Nhật
本	Bản
月	Nguyệt

“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”という修士論文 [1](a) によると、漢字語対照データベースの分類基準は7種類ある。

2.3.2 AB / AB

同じ意味を表す語が同一の漢字を用いる場合、発音が似ている漢字語と漢越語もある。それぞれの漢字語を調べる場合、すぐに意味が分かると考えられる。

会談	Hội (會) Đàm (談)
注意	Chú (注) Ý (意)
態度	Thái (態) Độ (度)
成立	Thành (成) Lập (立)

2.3.3 AB / AC, CB

同じ意味を表す語が漢字語と漢越語では一字異なるが、同一の字が同じ位置にある場合

診察	Chẩn (診) Sát (察)
診断	Chẩn (診) Đoán, Đoạn (断)
無断	Vô (無) Đoạn (断)
意見	Ý (意) Kiến (見)
拝見	Bái (拝) Kiến (見)

2.3.4 AB / BA

日本語の漢字語と同じ漢字を用いるが、語順が逆転する場合。実はベトナム語の Giai (階) Đoạn (段) という言葉は日本語の漢字で段階の意味を持っている。階段の意味は別の言葉で Cầu thang という言葉が意味を表す。こういう言葉は結構少ない。

階段	Giai (階) Đoạn (段)
段階	Đoạn (段) Giai (階)

2.3.5 AB / BA

同じ意味を表す語が漢日語と漢越語では一字異なる

り、同一の字が異なる位置にある場合

学校 Học (学) Hiệu (校)
学校 Trường (場) Học (学)
空港 Không (空) Cảng (港)
空港 Phi (飛) Trường (場)
節約 Tiết (節) Ước (約)
節約 Tiết (節) Kiệm (儉)

2.3.6 AB / CD

同じ意味を表す語が漢日語と漢越語では全く異なっている場合

貯金 Trữ (貯) Kim (金)
貯金 Tiết (節) Kiệm (儉)

同じ意味を表す語が漢越語ではなくベトナムの固有語である場合。固有語と漢越語との混種語も含む。

大人 Đại (大) Nhân (人)
大人 Người Lớn
電車 Điện (電) Xa (車)
電車 Xe Điện

二語が一語、あるいは三語以上の漢越語に相当する場合／複数の漢越語が該当し、どの語もよく用いられる場合

電気 Điện (電) Khí (氣)
電気 Điện (電)

3. 日中韓越漢字データベース

3.1 漢字語データベース概要

本システムの4カ国語漢字語データベースはメイン漢字データベース1つ及びサブ語彙データベース4つにより構成している。おおよそ26000行の多言語データを蓄積している。しかしデータベースには、まだ言語的な不備（各言語の多音字への完全対応など）、不完全な項目がある。言語分野で各言語に精通する方の協力を必要とするであろう。

オリジナルのメイン漢字データベースの表計算ソフト Excel 表形式のデータは、27列の多言語漢字データを収録していて、一部のデータ列はまだ完成していない。本システムでは処理しやすくするため、いくつかの箇所を改変した。例として、オリジナルのデータベースが作成された2000年時点では、ウェブブラウザの多言語表示問題はまだ数多く存在した。筆者（江見）は文字化けを防ぐため、文字数値

参照で日本語・中国語・韓国語のデータを蓄積している。HTML5が実用化されて以来、こういう問題が発生しにくくなり数値文字参照を使う意味がなくなった。故に本システムは現行のデータベースにおいて、数値文字参照は使っていない。本システムで稼働されている漢字データベースは複数のウェブサイトのデータを使った [33]-[38]。

中国語漢字語彙データベースは漢字水平考試 (HSK) 公表による級分けの漢字語彙データベースによるものである。合計5340件の語彙データを含んでいる。これらの語彙が HKS により6つのレベルに分けられる。本システムでの仕様は表3.1.2に示している通りである。韓国語データベースAは、韓虎林、今井久美雄、梅田博之の「漢字でおぼえるしっぽをつかむ韓国語」[12]によるもので、韓国語データベースBは市吉則浩の「日本人だからできる！漢字で覚える韓国語」[13]に基づくものである。ベトナム語のデータベースは漢越日ウェブ辞典 [38]をつかった。

本システムでの仕様は以下の表で示す。

表 3.1.1 メイン漢字データベース (tableData1) の構造

#	ヘッダー	コード	説明
1	日漢	JKanji	日本語における漢字（基準とする）
2	呉音 1	JGoon1h	日本語の歴史的仮名遣いの一つ
3	-	JCoon1p	呉音 1 の音声合成に使う。表示しない。
4	呉音 2	JGoon2h	呉音の発音ケース 2, null あり
5	漢音 1	JKanon1h	日本語の歴史的仮名遣いの一つ
6	-	JKanon1p	漢音 1 の音声合成に使う。表示しない。
7	漢音 2	JKanon2h	漢音の発音ケース 2, null あり
8	-	Jkana01	現代日本語名遣いケース 1。未搭載。
9	-	Jkana02	現代日本語名遣いケース 2。未搭載。
10	-	Jkana03	現代日本語名遣いケース 3。未搭載。
11	-	Jkana04	現代日本語名遣いケース 4。未搭載。
12	唐宋音	Jtousouon2h	日本語の歴史的仮名遣いの一つ
13	慣用音	JKanyouonh	日本語の歴史的仮名遣いの一つ
14	学年	Jgakunen	日本語国語教育における習得学年
15	繁漢	CTraditional	中国語繁体字
16	簡漢	CSimplified	中国語簡体字

17	ピンイン	Cpinyin	中国語ピンイン
18	声母	Cinitial	中国語ピンインにおける声母
19	韻母	Cfinal	中国語ピンインにおける韻母
20	声調	Ctone	中国語ピンインにおける声調
21	広東	CantonPynin	中国語広東語の発音
22	レベル	level	中国語学校教育における習得学年
23	ハングル	KHangul	韓国語表音文字ハングル
24	韓音	Kromaji	韓国語発音のアルファベット表示
25	越南 1	Vietnam1	ベトナム語における読み方, ケース 1
26	越南 2	Vietnam2	ベトナム語における読み方, ケース 2

表 3.1.2 中国語漢字語彙データベース (tableData4) の構造

#	ヘッダー	コード	説明
1	中国語簡体字	Ckanji	中国語簡体字 (基準とする)
2	ピンイン	Pinyin	中国語表音文字ピンイン
3	レベル	Level	HSK 語彙級分け
4	繁体字	Chankan	中国語繁体字
5	Weblio	Address	ウェブ辞書へのリンク

表 3.1.3 韓国語漢字語彙データベース A (tableData3) の構造

#	ヘッダー	コード	説明
1	日本語漢字	JKanji	日本語における漢字 (基準とする)
2	ハングル読み方	Hanguru1	韓国語における読み方
3	Weblio	Weblio	ウェブ辞書へのリンク
4	同音語彙	Kdouonn	韓国語において、発音の同じ語彙

表 3.1.4 韓国語漢字語彙データベース B (tableData5) の構造

#	ヘッダー	コード	説明
1	한국 한자 표기	JKanji	日本語における漢字 (基準とする)
2	カタカナ	Katakana	カタカナにおける読み方
3	한글	Hanguru	韓国語における読み方
4	의미	Meaning	意味に関する注釈
5	日本語注	Jps	日本語注釈
6	中文	Cps	中国語注釈
7	Vietnhanzi	Viet1	ベトナム語における読み方
8	Vietnam	Viet2	ベトナム語における意味

表 3.1.5 ベトナム語漢字語彙データベース (tableData2) の構造

#	ヘッダー	コード	説明
1	日本語漢字	JKanji	日本語における漢字 (基準とする)
2	日本語読み方	1Jyomi	日本語における読み方, ケース 1
3	越南語読み方	2Vyomi	越南語における読み方, ケース 2
4	越南語意味	Vimi	越南語における意味

オンライン日-越辞書と「詳解ベトナム語辞典」という本から漢字語と漢越語の熟語を検索し、よく使用されている言葉をまとめ、表計算ソフト Excel でデータを入力し、データベースを作った (図 3.1)。

日本語能力試験は N5 から N1 までのレベルがある。そして N5 から N1 まで全部で 2136 字があり、その中で意味がある熟語を組み合わせ、全部で 1 万 4 千ほどの熟語の漢字語と漢越語をデータとして入力している。そこに読み方と意味詳解のデータを付けている。

	A	B	C	D
1	漢字	読み方	ベトナム語読み方	ベトナム語意味
8	黄金	くがね	HOÀNG KIM	Vàng
9	基金	ききん	CƠ KIM	Quỹ
10	金塊	きんかい	KIM KHỐI	Vàng cốm
11	金額	きんがく	KIM NGACH	Số tiền
12	金冠	きんかん	KIM QUAN, QUẢN	Vương miện vàng
13	金柑	きんかん	KIM CAM	Kumquat
14	金魚	きんぎょ	KIM NGƯ	Cá vàng
15	金銀	きんぎん	KIM NGÂN	Vàng bạc
16	金具	かなぐ	KIM CỤ	Phụ kiện kim loại
17	金型	かながた	KIM HÌNH	Khuôn
18	金庫	かねくら	KIM KHỐ	An toàn
19	金鎖	きんぐさり	KIM TỎA	Dây chuyền vàng
20	金歯	きんば	KIM ?	Răng vàng
21	金色	こんじき	KIM SẮC	Vàng
22	金星	きんせい	KIM TINH	Sao Kim
23	金銭	きんせん	KIM ?	Tiền
24	金属	きんぞく	KIM CHỨC, THUỘC, CHÚ	Kim loại
25	金槌	てつち	KIM CHUY	Búa

図 3.1 漢字語と漢越語のデータベース見本

3.2 データベースの欠陥について

筆者 (江見) がメイン漢字データベースを表計算ソフト Excel で作成する時点では、各言語の漢字音の多音字について、十分に意識できなかった。多音字への完全対応を達成するため、リレーショナルデータベースが必須である。つまり、本研究を完成するには、データベースの再設計が必要である。

筆者 (陳) の考えでは、Indexed DB API は本システムの再設計に相応しいと判断した。しかし、データの不備などを考慮に入れると、作業量はあまりにも膨大すぎる。課題として残されている。図 3.2 にシステム構造を示す。本システムは文献 [1](b) のサ

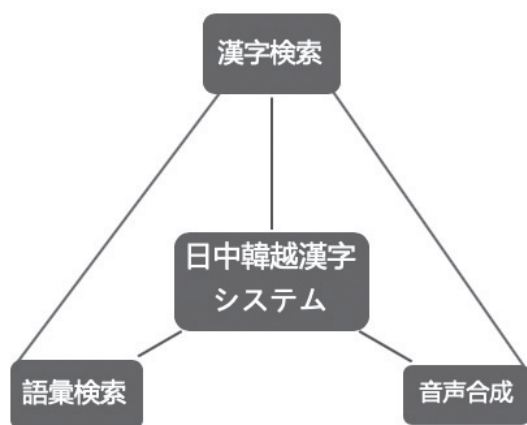


図 3.2 システム構造

イトの「(New 試作中) 日中韓越漢字」にある。

4. 日中韓越漢字システム

4.1 システム概要

本システムは、日中韓越漢字の発音およびそれに関連する情報をデータベースに蓄積し、その発音の類似性を研究・学習しやすくするためのものである。研究室や学校にとどまらず、実用性のあるアプリケーションの作成を目指している。このシステムは、日中韓越の中のどちらかを母語とする、4 言語の音声の類似性を学習・研究したい人を対象としている。

結果として、「漢字検索」「音声合成」「語彙検索」三つの機能を有するウェブアプリケーションができた。およそ半年の期間でこのような複雑なアプリケーションを完成させることは断念した。

開発環境として、本システムの開発のほとんどは「MacOS」で行われている。テキストエディターは「Brackets」を使用している。UTF-8 対応である。

開発技術として、基盤技術である「HTML5」「CSS3」「JavaScript」「PHP」、JavaScript フレームワークである「jQuery」「Bootstrap」「Tabulator」、デザイン仕様である「Material Design」「Font awesome」。パッケージ管理システム「NPM」などが使用されている。

実行環境として、「MacOS」「Windows」「iOS」「Android」で「Chrome」「Opera」「Edge」「Firefox」「Safari」の作動を確認した。結果として、「Safari」以外は思った通りに稼働した。

データベースには、2 万 6 千件以上の多言語データが保存されており、アプリケーションの豊富なフィルター機能を通じて、自由自在に検索できる。

しかし、データベースには、まだ言語的な不備（各言語の多音字への完全対応など）、不完全な項目が随所にある。言語分野で各言語に精通する方の協力が必要とされている。

漢字語の音声を即時に再生するため、Google, Baidu, FPT 社が提供している人工知能 TTS (Text-to-Speech) API が使用されている。今のところでは、日本語・中国語・韓国語・ベトナム語に対応している。言語の類似性研究に有益な広東語 TTS API も見つけたが、時間及び予算の制限によりまだ作っていなかった。後日に追加する予定がある。

4.2 システム構造

本システムは、Tabulator というフロントエンドインタラクティブテーブル JavaScript ライブラリを使ってデータ処理を行っている。解説しやすいため、Tabulator を「モデル、ビュー、コントローラー」三つの部分に分けて説明する。簡略に説明すると、モデルは実際にデータ処理・演算を行うモデル。ビューはモデルの演算結果をユーザーインターフェースで図形や文字として表示するためのモデル。コントローラーはモデルとビューの間でデータを伝送したり、処理手順を決めたりするモデルである。三つのモデルを合わせて、ようやく一つのプログラムとして動作するようになる。

Tabulator は Ajax データか JSON フォーマットのデータを読み取り、HTML インタラクティブテーブルとして表示する機能を持っている。ユーザーの操作により、リアルタイムで表示内容、表示方式を改変できる。他のフレームワークか DOM 操作と合わせれば、かなり複雑な機能が実現できる。本システムのデータ操作はほぼ全て Tabulator に関わっている。

本システムは、三つのサブシステムにより構成されている。それらは、漢字検索システム、語彙検索システム、音声合成システムである。その中で、語彙検索システムと音声合成システムは、漢字検索システムのデータに依存している。具体的に言うと、DOM 操作により、Tabulator model からユーザーが選んでいるデータを取得してから語彙検索システムおよび音声合成システムに渡す。両システムは受け取ったデータをもとに、語彙検索か音声合成処理を行う。

4.2.1 漢字検索

本システムの全ての操作は漢字検索から始まる。特定された漢字データを基として他の処理を実行する。

図 4.2.1 で示すように、ブラウザがページの読み取りを始めたら、自動的に JSON オブジェクトデータベースを Tabulator model に注入する。そして、ブラウザは Tabulator controller を通じて Tabulator view を形成する。形成された Tabulator view を User interface の一部として取り込み、これでページの初期レンダリングが終わる。

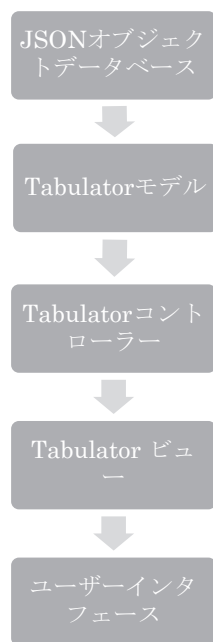


図 4.2.1 読み取りプロセス

漢字検索を行う場合、図 4.2.2 で示しているループに入る。まず、ユーザーがインタフェースで検索したい条件を、フィルターを設定して、Tabulator controller に渡す。フィルターの操作法などについては「4.3 漢字データフィルター」で詳しく説明する。Tabulator model は controller から処理された検索データを受け取る。それに従い、条件に合っているデータを取り出し、Tabulator view に渡す。Tabulator view が受け取ったデータを可視化し、User interface で更新する。これで漢字検索循環が終わる。

語彙検索及び漢字音合成の詳細については「4.7 人工知能 API による日中韓音声合成」「4.8 中韓越漢字語語彙検索」を参照してほしい。

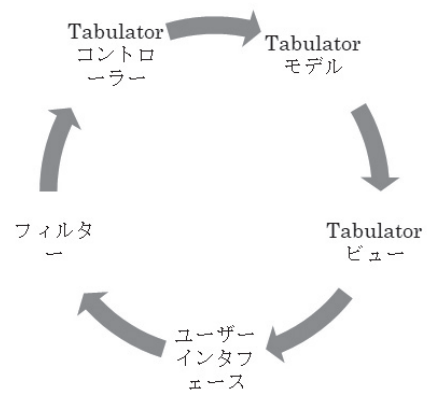


図 4.2.2 漢字検索のしくみ

4.3 漢字データフィルター

4.3.1 フィルター概要

前文で紹介した通り、本システムは日中韓越の漢字語においての音声的類似性を研究・学習するためのシステムである。故に、漢字語の音声的類似性を反映、研究するための機能が必要とされている。この機能を果たしているのは、JavaScript 及び Tabulator による漢字データフィルターである。

現時点の漢字フィルターモデルでは、全ての中国語音声反映しているピンインフィルターグループ、韓国語の語頭・語末を反映するハングルフィルターグループ、及び研究のため開発された、漢字を自由自在に検索できるカスタム検索機能がある。これからは各言語に特化するフィルターを追加する予定である（例として、日本語の漢字に特化するフィルター）。

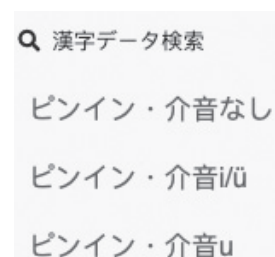


図 4.3.1 ピンイン検索ボタングループ

4.3.2 ピンインフィルターグループ

本システムのピンインフィルターは、厳密な言語的ルールによるものではない。中国語がある程度勉強できた中国語学習者の視点からできたものである。

介音の有無及び種類により、図 4.3.2 が示しているように、全ての中国漢字音を三つのサブメニュー

に分けた。「介音なし」は介音の存在しない中国語漢字音のことを指す。「介音 i/ü」は介音「i/ü」が存在する漢字音を指す。「介音 u」は介音「u」が存在する漢字音を指す。

NULL	-A	-E	-I	-ER	-AI	-EI	-AO	-OU	-AN	-EN	-ANG	-ENG
0-	A	E	I	ER	AI	EI	AO	OU	AN	EN	ANG	ENG
B-	BA	-	BI*	-	BAI	BEI	BAO	-*	BAN	BEN	BANG	BENG
P-	PA	-	PI*	-	PAI	PEI	PAO	POU*	PAN	PEN	PANG	PENG
M-	MA	ME*	MI*	-	MAI	MEI	MAO	MOU*	MAN	MEN	MANG	MENG
F-	FA	-	-	-	FAI	FEI	-	FOU	FAN	FEN	FANG	FENG
D-	DA	DE	DI*	-	DAI	DEI	DAO	DOU	DAN	DEN	DANG	DENG
T-	TA	TE	TI*	-	TAI	TEI*	TAO*	TOU	TAN	-	TANG	TENG
N-	NA	NE	NI*	-	NAI	NEI	NAO	NOU	NAN	NEN	NANG	NENG
L-	LA	LE	LI*	-	LAI	LEI	LAO	LOU	LAN	-	LANG	LENG
Z-	ZA	ZE	ZI	-	ZAI	ZEI	ZAO	ZOU	ZAN	ZEN	ZANG	ZENG
C-	CA	CE	CI	-	CAI	-*	CAO	COU	CAN	CEN	CANG	CENG
S-	SA	SE	SI	-	SAI	-*	SAO	SOU	SAN	SEN	SANG	SENG
ZH-	ZHA	ZHE	ZHI	-	ZHAI	-*	ZHAO	ZHOU	ZHAN	ZHEN	ZHANG	ZHENG
CH-	CHA	CHE	CHI	-	CHAI	-*	CHAO	CHOU	CHAN	CHEN	CHANG	CHENG
SH-	SHA	SHE	SHI	-	SHAI	SHEI	SHAO	SHOU	SHAN	SHEN	SHANG	SHENG
R-	-	RE	RI	-	-	-	RAO	ROU	РАН	REN	RANG	RENG
G-	GA	GE*	-	-	GAI	GEI	GAO	GOU	GAN	GEN	GANG	GENG
K-	KA	KE*	-	-	KAI	KEI	KAO	KOU	KAN	KEN	KANG	KENG
H-	HA	HE*	-	-	HAI	HEI	HAO	HOU	HAN	HEN	HANG	HENG
本来J-	本来Q-	本来X-	Y-	W-	K-P	K-L	K-K	J-N-	J-G-	CANTONG	NG-	

CANCEL SEARCH

図 4.3.2 ピンイン検索ボタンの例。図 4.3.1 の「介音なし」のフィルター

詳細検索メニューの一番左の列は声母のみ設定するフィルター、一番上の行は韻母のみ設定するフィルターである。「-」は対応する漢字音が存在しないことを示す。最後の 1 行において、「本来*」は歴史的な仮名遣いの特例である。大文字 K は韓国語を意味する。大文字 J は日本語を意味する。CANTONG は中国語広東語を意味する。「*-」は語

語頭

K P-	K B-	K M-	K K-	K G-	K H-	K T-	K D-	K N-	K R-	K CH-	K J-	K S-
------	------	------	------	------	------	------	------	------	------	-------	------	------

語尾

K-P	K-AP	K-EOP	K-YEOP	K-EUP	K-IP							
K-M	K-AM	K-EOM	K-YEOM	K-EUM	K-IM							
K-L	K-AL	K-EOL	K-YEOL	K-IL	K-OL	K-UL	K-YUL	K-EUL	K-WAI	K-WOL		
K-N	K-AN	K-EON	K-YEON	K-IN	K-ON	K-UN	K-YUN	K-EUN	K-WAN	K-WON		
K-K	K-AK	K-YAK	K-AEK	K-EOK	K-YEOK	K-IK	K-OK	K-YOK	K-UK	K-YUK	K-EUK	K-WAK
K-NG	K-ANG	K-YANG	K-AENG	K-EONG	K-YEONG	K-ING	K-ONG	K-YONG	K-UNG	K-YUNG	K-EUNG	K-WANG

他

J N-	J G-	CANTONG NG-
------	------	-------------

頭で、「-*」は語尾である。詳しい言語的規則は筆者（江見）の論文 [6] を参照してほしい。

4.3.3 ハングルフィルターグループ

ハングルは、朝鮮語を表記するための表音文字である。筆者（江見）は日本語、中国語、韓国語の語頭、語尾を中心に研究し、日韓両言語の音声的類似性を解明した。本システムのハングル語頭・語尾フィルターは筆者（江見）の従前のデータベース [6] の改良である（図 4.3.3. 参照）。

4.3.4 カスタム検索

前文で述べた通り、このカスタム検索機能は研究のために開発された機能である。全てのデータ列を望むままに検索できる。新しい言語的ルールの探求に役立つツールとして期待できる。しかし、このカスタム検索機能の使い方は簡単ではない。使いこなすにはそれなりの勉強や練習が必要かもしれない。現時点では、カスタム検索機能にはまだ問題点がある。例として、不完全な検索条件を入力しても警告が出ない。対処が必須と思われる。



図 4.3.4 カスタム検索

図 4.3.4 で示している通り、フィールドで検索したいデータ項目を、演算子で演算法を、検索内容で内容か内容の特徴を入力することでデータベースを検索する。つまり、演算子と検索内容のペアで内容の特徴を決めて、この特徴に合っているフィールドが属するデータ項目をデータベースから抽出する機能である。

演算子について説明する。「=」記号は、文字的な完全

一致を意味する。「!=」記号は、文字的な完全不一致を意味する。「Like」記号は、文字的に含むことを意味する。Regex 演算子はもっとも機能の強い演算子で正規表現という技術 [41] を使っている。

4.4 データの表示

4.4.1 テーブル形式のデータ表示

本システムは、「データ蓄積及び処理をフロントエンド技術に任せる」という理念を基として作られたものである。故に、データベース作成の前に、データの表示法を先に作る必要がある。于・玉岡は日韓中同形二字漢字語の品詞性ウェブ検索エンジンを作成するプロジェクトにおいて、語彙ごとにページを形成する方法を選んだ。それは于・玉岡が作成した検索エンジン [26] で、語彙ごとに詳細な多言語品詞性の表示が重視されているからである。本システムでは、語彙間の変化及び類似性の方が重要であるから、より比較しやすいテーブル表示法を採用した。図 4.4.1 が示しているのは于・玉岡が作成した日韓中同形二字漢字語の品詞性ウェブ検索エンジンで、図 4.4.2 は本システムの初期画面である。



図 4.4.1 日韓中同形二字漢字語の品詞性ウェブ検索エンジンのデータ表示 [26]

4.4.2 データ列選択

前文に書いた通り、本システムのデータ処理は全て Tabulator というフロントエンドインタラクティブテーブル JavaScript ライブラリに依存している。データ列選択機能は Tabulator データ列「visible」属性の有無によりコントロールされている。

このデータテーブルから参照できるデータは 20

日中韓越漢字										
漢字データ検索	日漢	吳音	漢音	繁漢	簡漢	ピンイン	声調	広東	ハングル	越南
ピンイン・介音なし	亜	エ	ア	亞	亚	ya	4	a3	아	Á
ピンイン・介音u	哀	オ	アイ	哀	哀	ai	1	oi1	애	AI
ピンイン・介音u	愛	オ	アイ	愛	愛	ai	4	ngoi3	애	ÁI
ハングル・語頭・語末	惡	アク	アク	惡	恶	e	4	ngok3	악	ÁC
カスタム検索	握	アク	アク	握	握	wo	4	ak1	악	ÁC
音声合成	圧	エフ	アフ	壓	压	ya	1	aat3	압	ÁP
日本語音声・吳音	扱	セウ	サウ	扱	扱	xi	1	chaap3	급	TRÁP
日本語音声・吳音2	安	アン	アン	安	安	an	1	on1	안	AN
日本語音声・漢音	暗	オム	アム	暗	暗	an	4	am3	암	ÁM
日本語音声・漢音2	案	アン	アン	案	案	an	4	on3	안	ÁN
中国語音声	以	イ	イ	以	以	yi	3	yi5	이	DĪ
韓国語音声	位	ㄷ	ㄷ	位	位	wei	4	wai2	위	VĪ
韓国語音声2	依	エ	イ	依	依	yi	1	yi1	의	Ŷ
越南語音声・越音	偉	ㄷ	ㄷ	偉	伟	wei	3	wai5	위	VĪ
越南語音声・越音2	圍	エ	ㄷ	圍	围	wei	2	tung1	위	VI
Replay / safari Play	委	ㄷ	ㄷ	委	委	wei	3	wai1	위	ŪY
漢字語彙検索	威	ㄷ	ㄷ	威	威	wei	1	wai1	위	UY
中国語・漢字語彙	尉	ㄷ	ㄷ	尉	尉	wei	4	wai3	위	UŶ
韓国語・漢字語彙A	意	イ	イ	意	意	yi	4	yi3	의	Ŷ
韓国語・漢字語彙B										
越南語・漢字語彙										

図 4.4.2 本システムのデータ表示

列ある。ブラウザを右クリックしてソースコードを表示させれば確かめることができる。ユーザーが気持ちよく使えるようにするため、一度に22列のデータを全て表示するのは賢い選択ではない。特に縦置きスマートフォンに対して、一度におおよそ4列のデータを表示するのは限界である。しかし漢字システムとして、どうしても全てのデータを参照できるようにする必要がある。

この矛盾を解決するため、データ列選択機能が開発された。この機能を通じて、ユーザーが必要とするデータ列のみ表示することが可能となった。13インチのPCを基準とし、そのモニターで最も重要な10列をスクロールなしに表示できるように工夫した。表示幅はこれより小さい端末において、自動的に横スクロールできるように切り替える。表示幅は13インチより小さいのに対して、フィールドの幅が自動的に大きくなる。

図4.4.3が示しているのはデータ列選択メニューで、デフォルトで表示されるデータ列がチェックされていることが分かる。このように表示列は簡単にカスタマイズすることが可能である。

☒ 日漢	☒ 数漢	☒ ハングル	☒ 越南
☒ 呉音	☒ 簡漢	☐ ハングル2	☐ 越南2
☐ 呉音2	☒ ピンイン	☐ 韓音	
☒ 漢音	☐ 声母	☐ 韓音2	
☐ 漢音2	☐ 韻母		
☐ 唐宋音	☒ 声調		
☐ 慣用音	☒ 広東		
☐ 学年	☐ レベル		

CANCEL SEARCH

図 4.4.3 データ列メニュー

4.5 フロントエンドデータベース

本システムのデータベースシステムには、三つの特徴が求められている。一つ目、ローカルでも動けること。そのため、完全にフロントエンドで動作できないといけない。全ての演算をユーザーのデバイスに任せる。二つ目、表計算ソフト Excel からのデータ遷移には、手数がかからない。このシステムの言語的理論はまだ研究段階である。故に、データベースも完成からまだ遠く、大きな変更が予想される。手間のかかるデータ修正は受け入れない。三つ目、UIの流暢さに支障がないくらいの処理スピード。本システムで処理されるデータの量はそれほど膨大ではないから、極めて素早いデータ処理は求めている。けれども、ユーザーエクスペリエンスへの悪影響は認めない [40]。

Web 開発において、MySQL か mongo DB のようなサーバーで稼働するデータベースを採用するのが普通だが、インターネットに接続していない場合と電波状況の悪いところでは使えないという欠点がある。本システムに関して、システムデザインを行う時点から、オフラインでも稼働できるオフライン Web アプリケーションの作成を目指していた。つまり、オンラインで一度アクセスしてから、ブラウザを閉じて Web アプリケーションに関するあらゆるデータを削除せずに、そのままキャッシュに保存することで、オフラインのアクセスを可能にすることである。故に、バックエンドで稼働する性能の強い重いデータベースより、2万件以上のデータに対処できる軽いフロントエンドデータベースの方が望ましい [39],[40]。

4.5.1 データベース技術

現時点では、W3Cにより標準化されたフロントエンドオフラインストレージ API は主に二つある。一つ目は「Web Storage」、ブラウザにデータをキーと値の組み合わせで蓄積し、利用する API である。二つ目は「Indexed Database API」、Web Storage API と同様、キーと値のペアで JavaScript のオブジェクトを蓄積し、利用する API である。一般的なりレシヨナルデータベースにおける、インデックスやトランザクションを利用できることが特徴である。けれども、今度のアプリケーションのデータベースには特別な特徴がある。まず、システムによる改ざん操作は全くない。そして、オリジナルのデータファイルは表計算ソフト Excel である。多音字対応を無視すれば、Web Storage API でも、Indexed Database API でも最善な解決策ではないことが明白である。故に、本システムは、JavaScript により、26000 行以上 JSON 化した表計算ソフト Excel データの読み取りおよび検索処理を実行することになっている。

JSON (JavaScript Object Notation) は、軽量のデータ交換フォーマットである。人間にとって読み書きが容易で、マシンにとっても簡単にパースや生成を行える形式である。JavaScript プログラミング言語 (ECMA-262 標準第 3 版 1999 年 12 月) の一部をベースに作られている。本システムに必要とされるデータをいくつかの配列化した JSON オブジェクトに蓄積している。以下はその一例である。


```
{ "JKanji" : " 亜", "JGoon1h" : " エ", "JGoon2h" : "", "JKanon1h" : " ア", "JKanonh" : "", "Jtousouon2h" : "", "JKanyouonh" : "", "Jkana01" : " ア", "Jkana02" : "", "Jkana03" : "", "Jkana04" : "", "CSimplified" : " 亞", "CTradional" : " 亞", "Cpinyin" : " ya", "Cinitial" : "", "Cfinal" : " ya", "Ctone" : " 4", "level" : " 1", "KHangul" : " 아", "Kromaji" : " a", "Vietnam1" : " Á", "Vietnam2" : "", "CantonPinyin" : " a3", "Jgakunen" : " 7" }
```

日本語における「亜」という漢字の多言語データに当てはまる。デフォルトとして、最も重要な10列のみがデータテーブルに表示されている（図4.5.1）。

日漢	呉音1	漢音1	繁漢	簡漢	ピンイン	声調	広東	ハンダ
亜	エ	ア	亞	亚	ya	4	a3	아

図 4.5.1 日本語漢字「亜」のデフォルトデータテーブル表示

4.5.2 スピード測定

図 4.5.2 と図 4.5.3 はデータ検索操作のパフォーマンス詳細である。JavaScript には、約 80ms (0.08s) の時間がかかった。本システムの仕様環境に対して、性能的な問題がないと思われる。

また、このプロジェクトはまだ完成とは言えない状態にあるので、JavaScript コードの効率化などはまだ行っていない。故に、さらなるスピードアップが期待できる。

Range: 1.70 s – 3.15 s



図 4.5.2 初期化パフォーマンス

Aggregated Time

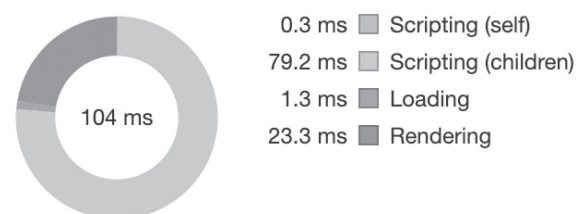


図 4.5.3 漢字検索パフォーマンス

4.6 多端末対応

4.6.1 モバイルファーストデザインの重要性

Statista の調査 [42] により、2009 年から 2018 年までの 10 年間、移動端末からのウェブサイト閲覧量は 0.9% から 52.2% まで成長してきた。2018 年にデスクトップ端末を超え、もっとも主流なウェブサイト閲覧デバイスとなっている。故に、今からの Web 開発に対して、モバイル端末の対応を考慮しないプロジェクトは半分以上のアクセス数をなくす恐れがある。本システムは、多端末対応を一番重要なポイントとして位置づけている。

MDN[39] により、モバイルファースト (Mobile First) とは、ある種のプログレッシブエンハンスメントで、ウェブ開発やウェブデザインにおいて、モバイル画面サイズのデザイン・開発をデスクトップ画面サイズのデザイン・開発より優先して注力するアプローチである。本システムの開発においては、モバイルファーストの理念が徹底されている。そのため、多ブラウザ対応、レスポンシブデザイン、Full Screen API の適用、パフォーマンス最適化などを行った。

4.6.2 多ブラウザ対応

多端末対応を実行するには、ブラウザの違いを細かいところまで考慮しないといけない。少なくとも、主流なブラウザでテストを行って修正する必要がある。文献 [42][43] が示しているのは StatCounte の調査による 2017.12 から 2018.12 までの世界的なブラウザの占有率遷移である。

本システムの目標ユーザーから考えて、結果的にデスクトップでもモバイルデバイスでも同じく Chrome, Safari, Firefox, Edge の四つのブラウザを基準とすることになった。実際の開発において、多ブラウザ対応にかなり苦労した。開発時間の半分以上は多ブラウザ対応にかかった。特に Safari の場合、安全対策は他のブラウザより明らかに厳しい。レポートを書く時点で、まだいくつかの問題が解決されていない。例えば、iOS の Safari に対して、全自動な音声合成 (自動再生) は今でも実現しておらず、どうしてもユーザーから直接に再生ボタンを押す必要がある。Full Screen 機能も Safari で稼働できない。Safari への対応も一つの課題として残っている。

4.6.3 モバイルファーストに関する技術

多種多様な端末が登場し続けている現在、昔のようにデバイスごとにページを作るのもありえない。そこで、より良いユーザーエクスペリエンスを提供するため、一つの HTML で、デバイスの特性に応じてレイアウトやデザインを変化させるレスポンシブデザイン技術が不可欠である。レイアウトやデザインの変更には様々な手段があり、本システムのレスポンシブデザインは Bootstrap などのフロントエンドフレームワークの他、CSS アニメーション及びオリジンの CSS メディアクエリも使用されている [39](c)。

ユーザーインターフェースのデザインの設計の詳細は文献 [8](b) を参照して欲しい。

4.6.4 パフォーマンス最適化

多端末に対応できる Web アプリケーションを作成するには、デスクトップデバイスとモバイルデバイスの演算力の差に配慮する必要もある。本システムにおいて、データベースを含むとおおよそ 30000 行の JavaScript コードを実行する必要があるため、JavaScript コードの実行順序及び演算効率への配慮は不可欠である。実際のところ、JavaScript コードを少し変えることで、実行効率が倍以上に上がることもあった。アニメーションの流暢さにも大きな影響を及ぼす。

前に紹介した通り、本システム全てのデータ処理は JavaScript で実行している。最初のページを読み取る段階では、30000 行以上の JavaScript コー

ドを処理する必要がある。Google Chrome Version 71.0.3578.98 (Official Build) (64-bit), 2.7 GHz Intel Core i5 を搭載する MacBook Pro late 2015 では、図 4.6.1 が示しているように、初期コード処理に約 1.5 秒の時間がかかる。また、メニューアニメーションを実行してからのデータ再読み取りにも、0.2 秒の時間がかかる。JavaScript 演算により、アニメーションの FPS が急に落ちることも明らかである。きちんと対処しないと、非常に気持ち悪い遅延が出る恐れがある。演算力の弱いモバイルデバイスではなおさらである。

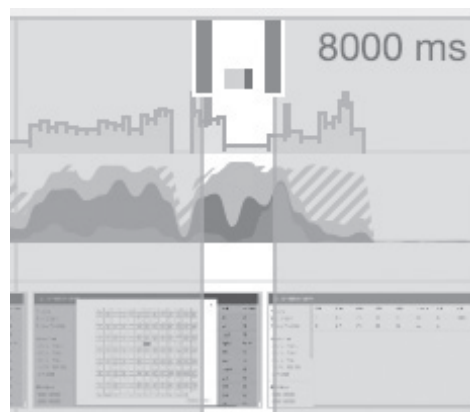


図 4.6.1 フィルターを設置してからのデータ再読み込み

4.7 人工知能 API による日中韓音声合成

4.7.1 音声合成の概要

このシステムにおいて、日本語・中国語・韓国語・ベトナム語の漢字音合成機能が搭載されている。

その中、日本語および韓国語の音声は Google

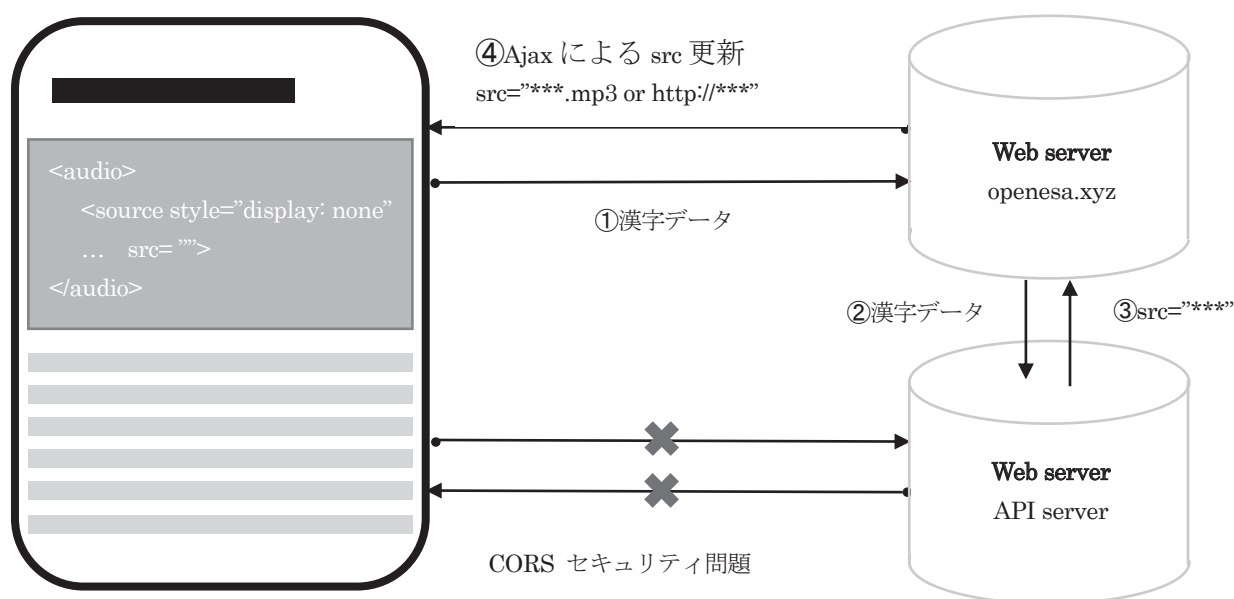


図 4.7.3 音声合成モデルにおけるデータの流れ

Cloud Text-to-Speech API によるものである。中国語の音声は Baidu の音声合成によってできている。ベトナム語の音声はベトナムの FPT が開発した OpenFPT API に依っている。同じく音声合成 API であるが、操作法はかなり違っている。本システムは、PHP で HTTP request を駆使し、API サーバーに音声合成を請求することになっている。

前文で述べたように、本システムでできる限りフロントエンド技術で機能を実現したい。この音声合成の部分は、唯一バックエンドの処理が必要とする部分である。もともとフロントエンド言語である JavaScript でも HTTP request を發送できるが、近年各ブラウザのセキュリティ対策の強化により、JavaScript による操作はかなり制限されている。

4.7.2 CORS セキュリティ問題

MDN (2019) によると、オリジン間リソース共有 Cross-Origin Resource Sharing (CORS) は、追加の HTTP ヘッダーを使用して、あるオリジン (ドメイン) で動作しているウェブアプリケーションに、異なるオリジンのサーバーにある選択されたリソースへのアクセスを許可することができる仕組みである。本システムの音声データ取得はこの CORS パターンの一例である。

異なるオリジンのサーバーから送ったデータの安全性が証明しにくいいため、多くの現代ブラウザは JavaScript による CORS 操作を禁止している。この CORS 問題を回避するため、PHP により第 2 のサーバーからデータを取得してクライアントへ送る。本システムの開発において、各 API の仕様の違いにより、サーバーから取得できるデータのフォーマットも違っている。そのため、第 2 のサーバーから取得したデータを処理する必要がある。現在のブラウザでは、JavaScript によるストレージアクセスも制限されている。この二つの理由により、本システムの作成において、PHP でのデータ処理プログラムが JavaScript より作りやすい。故に、PHP を採用することで、ブラウザのセキュリティ仕様による妨げを避けることに成功した。その代わりに、オフラインで稼働できないという欠点が出てきた。これまで JavaScript のセキュリティ問題についてたくさん書いたが、実際、JavaScript での音声合成 API 操作は不可能ではない。筆者の技術力不足及び時間制限により、レポートを書く時

点までで対処できなかったことであり、課題の一つとして残っている。

4.7.3 音声合成モデルの構造

MDN によると、Asynchronous JavaScript + XML (AJAX) はそれ自身がある種の技術というわけではないが、Jesse James Garrett によって 2005 年に作られた言語で、既存の技術同士を組み合わせた新しいアプローチを意味する。これらの技術が Ajax のモデルで組み合わせられると、ウェブアプリケーションはより速く、表示されているページ全てを再読み込みすることなく、ユーザーインタフェースを逐次更新できるようになる。本システムの音声合成モデルにおいては、ユーザーの請求により異なる音声ファイルを流す必要がある。故に、Ajax のようなユーザーの操作に応じてページの一部を更新する技術が不可欠である。

図 4.7.3 で示しているように、本システムの HTML ファイルには、CSS により不可視化された source タグが存在する。ユーザーが音声合成ボタンを押すたびに、この source タグの src 属性が Ajax によりサーバーから送られた新しいアドレスに更新される。

アドレスを取得するには、一連のデータ伝送が必要としている。

CORS セキュリティ問題により、ユーザーと API サーバー直接通信が禁止されているため、システムサーバー [7] での中継が必要とされている。ユーザーがボタンを押すと、フロントエンドで JavaScript により、ユーザーが選んでいるデータ列が処理される。処理したデータは JavaScript HTTP request によりシステムサーバーに發送する。システムサーバーで、PHP により二次処理を行い、漢字データを各音声合成 API の仕様に変更し API サーバーに転送する。API サーバーで音声合成されてから、そのアドレス (Baidu, OpenFPT) か JSON 化した MP3 (Google) をシステムサーバーに送る。システムサーバーで処理されてから、Ajax により、HTML ファイルにある不可視化された source タグの src 要素を新しい MP3 のアドレスに更新する。Google Cloud Text-to-Speech API の場合、JSON 化された MP3 データをデコードするプロセスもある。

Baidu 及び OpenFPT の音声合成 API には、おおよそ 2 秒の遅延がある。今のところ、対処する

方法がない。

4.8 中韓越漢字語彙検索

本システムの語彙検索演算は、全てフロントエンドで JavaScript により実行する。ユーザーが語彙検索ボタンを押すと、vocabulary.js にある語彙検索関数が動き始める。

まず、語彙検索関数は Tabulator コントローラーにユーザーが特定している漢字データ(selectedData)を請求する。Tabulator コントローラーから漢字データを受けとってから、ユーザーインタフェースにある語彙ウィンドウを更新しながら語彙データベースで検索条件に合っているデータをHTML テーブルテキストとして table_str に蓄積する。全ての検索が終わってから、一気にユーザーインタフェースにある語彙ウィンドウに注入する。これで語彙検索操作は完了とする。

最初は JavaScript の処理スピードを心配していたが、実際には全く問題ない。欠点として、これ以上の拡張が難しいことがある。例えば、語彙にも音声合成機能を追加したい場合には、このやり方では極めて困難である。漢字データテーブルのように Tabulator を使用する手もあるが、ユーザーの端末に過剰な負担をかける恐れがある。できるだけ避けたい。語彙検索のより良い実現法を考慮するのも、これからの課題である。

5. 学習教材としてのデータベース利用と検証

我々が作成したデータベースを利用して、実際にどのように漢字語彙を増やしていくのかの概略を示したい [8](b)。

5.1 韓国語漢字語彙

まずは、韓国語の漢字語彙を示す。
図 5.1.1 のように調べたい漢字の行を選択する。ここでは「慶」を選ぶ。



図 5.1.1 漢字の選択

図 5.1.1 の左端の“韓国語・漢字語彙 A” [13] をクリックする。図 5.1.2 のように選択した漢字を使った語彙が表示される。

慶 敬

한자	한글	Weblio	同音語
慶	강연	Weblio	
慶	강직비	Weblio	
慶	강주	Weblio	
國慶日	국강일	Weblio	

図 5.1.2 韓国語の「慶」

カスタム検索で、検索フィールド 1 > “ハングル” + “=” + “중” で検索した場合、図 5.1.3 のようになる。ハングルで “중” と書く漢字が表示される。

日漢	漢音1	漢音1	漢漢	韓漢	ピンイン	声調	広東	ハングル
衆	ス	シュウ	衆	衆	zhong	4	jung3	중
重	チュウ	チュウ	重	重	zhong	4	chung4	중
中	チュウ	チュウ	中	中	zhong	1	jung1	중
仲	チュウ	チュウ	仲	仲	zhong	4	jung6	중

図 5.1.3 ハングル文字を用いたカスタム検索結果

例えば、ブログ記事 [18] を読んで語彙を増やしたいとする。ハングル・語末の “K-ONG” で検索すると “空” の「ハングル」表示の一致が確認でき、日本語で “コウ” と読む漢字は韓国語では “kong” になるという記事の内容を確かめることができる。

また、データベースでは他の漢字の各言語における読みを比較することができ、同じ読み方をする他の漢字を同時に確認することもできる。

5.2 中国語漢字語彙

中国語の場合を示す。中国は発音として存在しう

るすべてのピンインを三つの画面に分類して操作するようにしている。ピンイン・介音なし、ピンイン・介音 i/ü、ピンイン・介音 u の 3 画面である。たとえば、ピンイン・介音 i/ü を選択して、qing を調べてみよう。まずは QING ボタンを選択して、SEARCH ボタンを押す。すると一覧が表示される。そうすると、日本語の呉音や漢音の表示を見れば、カ行またはサ行の音であることがわかる。韓国語、広東語、ベトナム語などを比較して中国語の発音を理解することが可能である。

たとえば、「慶 / 庆」を選んで、右側の発音メニューをクリックするとそれぞれの言語の発音を再生することができる。また、右メニューの「中国語語彙」をクリックすると HSK 頻出単語が図 5.2.1 のように表示される。

庆 qing 4声

中国語簡体字	ピンイン	レベル	繁体字	Weblio
国庆节	guóqīngjié	五	國慶節	Weblio
庆祝	qīngzhù	五	慶祝	Weblio

図 5.2.1 中国語の「慶」

5.3 ベトナム語漢字

ベトナム語の場合は、日本語漢字語彙を単純にベトナム語読みしたものを表示している。そしてベトナム語の意味も併記している。これは漢越語の対応が日本語、中国語、韓国語などについて研究途上であり、まだ不明な部分が多いからである。それでもベトナム人が日本語漢字を学習することはできるであろう。

たとえば、「慶 / 庆」を選ぶと図 5.3.1 のようになる。

慶 KHÁNH

×

日本語漢字	日本語読み方	越南語読み方	越南語意味
御慶	ぎょけい	NGŨ, NHA, NGŨ KHÁNH, KHUÔNG, KHANH	Kei của ban
慶事	けいじ	KHÁNH, KHUÔNG, KHANH SỰ	sự kiện tốt lành
慶弔	けいちょう	KHÁNH, KHUÔNG, KHANH ĐIẾU, ĐỊCH	Xin chúc mừng và chia buồn
慶祝	けいしゅく	KHÁNH, KHUÔNG, KHANH CHÚC, CHÚ	Celebration
慶賀	けいが	KHÁNH, KHUÔNG, KHANH HẠ	sự chúc tụng

BACK

図 5.3.1 ベトナム語の「慶」

次に漢字語と日本語を勉強するために、漢越語を知る必要があると考えられる。例えば“風各”という中国語の言葉であり、漢越語に翻訳すれば“風”は“Phong”また“各”は“Cách”，この二つの字

を組み合わせたら，“Phong Cách”という熟語が出来上がる。ベトナム人なら意味が分かるはずである。

それから“注意”という漢字であり、漢越語に翻訳すれば“注”は“Chú”また“意”は“ý”，この二つの字を組み合わせると，“Chú ý”という熟語が出来上がる。この語の意味は全く同じである。

5.4 使用者の感想・意見

5.4.1 中国語学習

筆者（江見）によると、京都情報大学院大学の 10 年前ぐらいの学生がこのデータベース [7] を用いて、日本語から中国語の漢字音をスムーズに学習することができた。日本人（京都情報大学院大学生）が使用した場合、日本語から中国語漢字音がよく分かり、大変効果的で HSK3 級にも合格した。

ところが中国人であれば、日本語を学習する上で、およそ想像がつく結果であるので、あまり役立つというレベルではない。理由は中国語の音素数が日本語よりもはるかに多いからであろう。今後は韓国人やベトナム人が中国語を学習するときのどの程度役立つかは今後の課題である。

5.4.2 韓国語学習

Ksp 理数学院／Ksp コリア学院 代表者の鍵本聡氏によると、「大変興味深い。学習に役立つでしょう」とのことである。

日本人学生には、韓国語漢字語のカタカナ表記はリエゾンがわかり役立つ。

中国人留学生は、ハングル文字の謎が解けて、韓国語への敷居が下がった。

ベトナム人は、韓国語の漢字音がベトナム語と似ていることが分かった。

5.4.3 ベトナム語学習

ベトナムの漢越語の分類が不十分であるため、ベトナム語を学習するには不十分である。（現状ではベトナム語の漢字語の研究成果が十分に存在しないため。）日本人と中国人にとっては、ベトナム語の敷居が下がった。

日本語や漢字語と中国語などを勉強すると共にベトナム人が漢越語を学ぶのが、漢字を一番早く学ぶ方法であると期待している。そうなれば、漢越語－日本語の漢字語データベースを使った漢越語－日本語の漢字語学習システムは、大変意味があるだろう。

一方で日本人や中国人に対しては、この日本語－

漢越語学習システムが役立つ場合が考えられる。なぜなら現在ベトナム語は漢字語を使っていないので、2.3の「ベトナムの漢字語」を見ると、いろいろな分類のある漢越語を理解するのが難しいからである。

5.4.4 日本語学習

中国人留学生にとっては、日本語漢字音は中国語からおおよそ推定できるのでそれほど役立つかどうかはわからない。

韓国人は、日本語との語彙の共通性がよく分かった。

ベトナム人は、日本語の漢字音がベトナム語と似ていることがよく分かり、学習の補助になる。

6. おわりに

筆者(江見)が2009年に作成したデータベース[6][7]を、この度、修了プロジェクトのテーマとして、日中韓越漢字システムのデータベースの作成を行いつつ、もう一つのテーマとして、漢字音・漢字語についての研究に取り組んだ。外国人にとって日本語を学習するのは難しいことで、ベトナムには漢越語が残っているので、その発音から漢字語が学習しやすいと考え、同様のことを中国語・韓国語でも実施できるシステムとして完成させることができた。システムにはまだまだ改善点が考えられるため、さらに実用的なものにしていきたいと考えているので、今後も努力を継続していきたいと思う。

6.1 技術的な側面からのまとめ

まずは技術的な側面から達成できたことは以下の通りである。

- ・漢字語のデータベース作成
- ・各言語の語彙データベース
- ・各列での条件検索による表示
- ・韓国語の語頭語末類似発音でのフィルタリング
- ・各言語での音声再生機能の実装
- ・音声再生機能以外のローカル化
- ・Safari 以外のブラウザのマルチデバイス対応

次に技術的な側面から達成できなかったことは以下の通りである。

- ・漢字語データベースはまだ完全なものではない
- ・ベトナム語の語彙データベース

- ・ベトナム語の類似発音におけるフィルタリング
- ・音声再生機能のローカル化
- ・Safari ブラウザーへの対応
- ・各言語の多音字への完全対応

6.2 将来

将来は以下のように改良したい。

- ユーザーのフィードバックによる改善
- データベースの完成度

ベトナム人が日本語を学ぶことはほぼ可能であろう。しかし、漢越語の分類が不十分なため、日本人、中国人、韓国人がベトナム語を学習することは難しい状態である。

また、語彙のデータベースは日本語と韓国語の類似度が高いことがわかったので、そこを中心に中国語とベトナム語(漢越語)へ拡張するのがいいという目処は立っている。

- ユーザーインターフェースや使い勝手

この件は当面の課題としておきたい。Safari ブラウザーへの対応は当面の課題としたい。特に iOS Safari の全自動音声再生、iMac・iOS Safari の Full Screen API への対応などは必要である。

- 各言語の多音字への完全対応

これは多対多のデータベースをつくることになるので、大変難しい課題である。これを達成するには、Indexed DB でフロントエンドデータベースを作り直す必要があり、作業量は非常に膨大となることが予想される。

現状では表 6.1 のように日本語漢字でも多：多の関係が発生している。このような日本語漢字に数字の添え字を振っている。簡体字と日本語漢字が同一の場合は添え字は 0 から始まるようにした。

表 6.1 多：多の関係のある漢字

日本語	繁体字	簡体字	ピンイン	ハングル
連 0	連	连	lián	연
連 1	聯	联	lián	연
画 0	畫	画	huà	화
画 1	劃	划	huà	획
弁 1	辨	辨	biàn	변
弁 2	瓣	瓣	bàn	판
弁 3	辯	辩	biàn	변
弁 4	辯	辩	biàn	변
弁 5	辦	办	bàn	판

(e) ベトナム語の音声類似性に関する研究

実際、日本語漢字音、中国語漢字音、韓国語漢字音に比べてベトナム語漢字音や漢越語に関する研究成果は十分ではないので、これを生かした学習教材を作るのは現状では難しい状態である。今後の研究成果に期待したい。

(f) オープンソース化

オープンソース化により、各地にいる研究者及び学習者がお互い協力しながらシステムを改善する。

(g) JavaScript CORS 問題へ対応して発音がどこでも再生できるようにしたい。

このプロジェクトではサーバー技術の分からない方にもデータベースを配布することにこだわったので、クライアント技術のみを使用した。ウェブサーバーに本システムを置いた場合は、発音を再生できるが、本システムをクライアント端末で実行した場合は、データ検索はできても JavaScript CORS 問題が原因で発音の再生ができない。これは今後の課題である。

(h) オフラインアプリケーション

誰でも、どのような端末からでも、電波状況の悪いところでも便利に使えるように、当初はフロントエンド技術のみでオフラインアプリケーションの開発を目指していた。筆者(陳)の技術力不足により、システム設計段階で間違いを起こし、オフラインアプリケーションの作成に失敗した。多言語対応、ユーザーインタフェースの最適化(誤操作への警告)なども完成させたい。

謝辞

この論文を作成するに当たって、多くの方から助言とご指導をいただいた。Ksp 理数学院/Ksp コリア学院代表の鍵本聡氏には韓国語の教材作成について貴重な助言をいただいた。厚く御礼申し上げます。

参考文献

- [1] (a) 松田真希子, Tuyen T.T.K., Thuy N.M., 金村久美, 中平勝子, 三上喜貴, “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, (2008). ; (b) 中澤信幸「中国語発音学習における日本漢字音活用の可能性について」
<http://dl.ndl.go.jp/info:ndljp/pid/3510638?tocOpened=1>
 - [2] 岡田朝雄, 畔上司, 「読んでおぼえるドイツ単語 3000—英語からドイツ語へ」, pp.1-283, (朝日出版社, 1987)
 - [3] 藤井友子, 「漢字音—すぐに役立つ!! 日中朝ベトナム・共通語彙 408」, pp.1-136, (朝日出版社, 1986)
 - [4] 下出憲一, 「聴いてびっくり漢字の世界—日中韓越の漢字音はこんなに似ている」, pp.1-174, (風詠社, 2018)
 - [5] (a) 増田弘, 大野敏明, 「古今各国「漢字音」対照辞典」, pp.1-269, (慧文社, 2006 年); (b) 大野敏明, “詳説世界の漢字音”, (慧文社, 2017) .
 - [6] 江見圭司, “現代日本漢字音を中国語、韓国語と歴史的仮名遣いから理解するためのデータベースと電子教材の作成”, NAIS Journal vol.8, pp.37-50, 2013. ;
 - [7] 中国語, 広東語, 日本語, 韓国語, ベトナム語, 漢字 (仮), <http://emich.world.coocan.jp/kanji/kanji.html> (2019 年 11 月 1 日閲覧, Windows 7 IE でのみ利用可能) Windows 7 は 2020 年 1 月 14 日で公式サポートを終了した。
 - [8] (a) 日中韓越漢字, <http://emich.world.coocan.jp/kanji/fp7/index.html>, (2019 年 11 月 1 日閲覧); (b) マニュアル <http://emich.world.coocan.jp/kanji/kanji-user-manual.pdf>
 - [9] (a) 府川充男, 小池和夫, 「旧字旧かな入門—日本人の手習い」, pp.1-251, (柏書房 2001); (b) 山口 謠司, 「ん—日本語最後の謎に挑む (新潮新書)」, pp.1-190, (新潮社, 2010 年)
- 韓国語関連
- [10] 伊藤 智ゆき, 「朝鮮漢字音研究 (本文篇, 資料篇)」, (汲古書院, 2007)
 - [11] 小学館 (編集), 金星出版社 (編集), 「朝鮮語辞典」, pp.1-2065 (小学館, 1992)
 - [12] 韓虎林 (原著), 今井久美雄 (訳, 著), 梅田 博之 (監修), 「漢字でおぼえるしっぽをつかむ韓国語」, pp.1-295, (インターブックス, 1999)
 - [13] 市吉則浩, “日本人だからできる! 漢字で覚える韓国語”, (河出書房新社, 2006) .
 - [14] 尹聖媛 (原著), 奉英娥 (原著), 「漢字でおぼえる韓国語単語」, pp.1-270, (国際語学社, 2007)
 - [15] 兼若逸之, “韓国語の漢字語マスター法則が分かれば語彙が増える!”, (HANA, 2017) .
 - [16] 鶴見ユミ, “韓国語単語スピードマスター 漢字語 3300”, (ジェイ・リサーチ出版, 2010) .
 - [17] 謝国芳 (謝国芳), 「破解韩国语单词的奥秘:3000 常用韩国语汉字词速记」, pp.1-313, (外语教学与研究出版社, 2008)
 - [18] 大北章二, “韓国語通訳の大北章二さんのブログ”, <http://honyakuka.jugem.jp/?pid=1> (2019 年 11 月 01 日閲覧)
- 中国語関連
- [19] 北浦藤郎, 蘇英哲, 鄭正浩, 「50 音引き中国語辞典」, pp.1-1184, (講談社, 2000)
 - [20] 相原 茂 (編集), 「講談社中日辞典 第三版」, pp.1-2224, (講談社, 2010 年)
 - [21] 相原 茂 (編集), 「講談社日中辞典」, pp.1-2214, (講談社, 2006 年)
 - [22] 藤堂明保 (編集), 加納喜光 (編集), 「学研新漢和大字典 普及版」, pp.1-2222, (学習研究社, 2005)

- [23] 藤堂明保 (編集), 「漢字源 改訂第六版」, pp.1-2281, (学習研究社, 2018)
- [24] 藤堂明保, 相原茂, 「中国語概論」, pp.1-334, (大修館書店, 1985 年)
- [25] 謝国芳, 「日語漢字读音規律揭秘」, pp.1-469, (世界图书出版有限公司, 2008 年)
- [26] 玉岡賀津雄, “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』第 29 号, pp.43-62, 2015
- ベトナム語関連
- [27] 三根谷 徹, 「中古漢語と越南漢字音」, pp.1-543, (汲古書院, 2009 年)
- [28] 川本邦, “詳解ベトナム語辞典 = Tu Dien Tuong Giai Viet Nhat. 3rd ed. To`kyo”, 大修館, 2011.
- [29] Can.N, “Thủ phân kì lịch sử 12 thế kỉ của tiếng Việt. Ngon Ngu. 6th ed”, pp.7-12, 1998.
- [30] Loc V. (翻訳と詳解), “An nam dịch ngữ = An Nam”, 1995.
- [31] Long.P, “Họngôn ngữ Nam Á – Austroasiatic Languages”, Ngonngu.net. <http://ngonngu.net/index.php?p=284> (2019 年 11 月 01 日閲覧)
- [32] Long. P, “Về việc phân định các giai đoạn lịch sử trong quá trình phát triển củatiếng Việt (phần 1)”, Ngonngu.net <http://ngonngu.net/index.php?p=292> (2019 年 11 月 01 日閲覧)
- 第 3 章で参照したウェブサイトや文献
- [33] 中国語よめ～る君の詳細情報: Vector ソフトを探す! <http://www.vector.co.jp/soft/win95/edu/se412675.html>
- [34] ハングル・漢字対照表・かじりたてのハングル <http://kajiritate-no-hangul.com/kshanja.html>
- [35] ハングルヨミ変換 <http://cattools.nii.ac.jp/hangul/>
- [36] Unicode 対応表・広東語辞書 JCanExpress - http://www.cantonese.jp/pages/list_unicode.htm
- [37](a) 漢字－漢越オンライン辞書 2001 <http://kanji.tudiennhatviet.com/>
- [37](b) 漢越日ウェブ辞典 <http://hanvietnhat.vietnam55.com/?cat=11>
- 第 4 章で参照したウェブサイトや文献
- [38] Oli Folkerd, “Tabulator - Interactive JavaScriptTables”, Tabulator.info <http://tabulator.info/> (2019 年 11 月 01 日閲覧)
- [39] (a) MDN, “AJAX”, MDN web docs <https://developer.mozilla.org/ja/docs/Web/Guide/AJAX> (2019 年 11 月 01 日閲覧); (b) MDN, “CSS アニメーションの使用”, MDN web doc https://developer.mozilla.org/ja/docs/Web/CSS/CSS_Animations/Using_CSS_animations (2019 年 11 月 01 日閲覧); (c) MDN, “Fullscreen API”, MDN web docs https://developer.mozilla.org/ja/docs/Web/API/Fullscreen_API (2019 年 11 月 01 日閲覧); (d) MDN, “Mobile First (モバイルファースト)”, MDN web docs https://developer.mozilla.org/ja/docs/Glossary/Mobile_First (2019 年 11 月 01 日閲覧); (e) MDN, “オリジン間リソース共有 (CORS)”, MDN web docs <https://developer.mozilla.org/ja/docs/Web/HTTP/CORS> (2019 年 11 月 01 日閲覧)

- [40] (a) Ken Lunde, 小松章 (訳), 逆井 克己 (訳), 「CJKV 日中韓越情報処理」, pp.1-1127, (オライリージャパン, 2002);
- (b) Ken Lunde, “CJKV Information Processing: Chinese, Japanese, Korean, and Vietnamese Computing Second Ed.”, pp.1-900, (O’Reilly Media, 2009)

■その他

- [41] Percentage of all global web pages served mobile phones from 2009 to 2018, <https://www.statista.com/statistics/241462/global-mobile-phone-website-traffic-share/>, (2019 年 11 月 01 日閲覧)
- [42] StatCounter, “Desktop Browser Market Share Worldwide”, StatCounter Global Stats <http://gs.statcounter.com/browser-market-share/desktop/worldwide> (2019 年 11 月 01 日閲覧)
- [43] Statistic, “Mobile share of website visits worldwide 2018”, Statistic.com <http://gs.statcounter.com/browser-market-share/desktop/worldwide> (2019 年 11 月 01 日閲覧)

◆著者紹介

ドウ・フィ・チュオン Do Phi Truong

京都情報大学院大学 2019 年 3 月修了。

山崎 恭司 Yasushi Yamazaki

京都情報大学院大学 2019 年 3 月修了。

陳 守誠 Shou Cheng Chen

京都情報大学院大学 2019 年 3 月修了。

江見 圭司 Keiji Emi

京都情報大学院大学准教授

2009 年, 2010 年中国・福建師範大学への派遣講師として 8 か月の中国滞在中に漢字データベースをつくったのが本研究のきっかけである。