

(注意) これは漢字データベースを使うためのレポートです。  
よくご覧ください。

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  
～音韻的類似性をもちいたアプローチ～

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

2019年2月15日

京都情報大学院大学  
応用情報技術研究科

DO PHI TRUONG  
M17W0010

山崎 恭司  
M17W0122

陳 守誠  
M17W0160

(2017年4月ウェブビジネス技術専攻入学)

題名	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 ～音韻的類似性をもちいたアプローチ～		
学籍番号 氏名	M17W0010 DO PHI TRUONG M17W0122 山崎 恭司 M17W0160 陳 守誠	指導教員	江見 圭司

### <要旨>

プロジェクトの目的：

漢字文化圏、日本・中国・韓国・ベトナム、各言語の漢字語の音韻的類似性を研究・学習するウェブシステム開発することを目的とする。システム構築は可能な限り、クライアント技術だけとした。

### 実施内容・手法

以下のような事柄を実施した。

- ・漢字語（日中韓越の各語および広東語）でのデータベースを作成
- ・各言語におけるの語彙（熟語）及び意味の表示
- ・各言語の文字表示でのソート機能
- ・各列で一意的条件検索による表示
- ・類似する発音（各言語）でのフィルタリング機能、
- ・各言語での音声再生機能の実装

### 結果・成果・結論

以下のような成果物ができた。

- ・アプリケーションのインターフェースについては、表示列の任意選択の機能、メニューの表示・非表示の機能、スマートフォン・タブレットからの閲覧・操作にも対応したものになっている。
- ・データベースを利用して各言語で漢字語の発音が同じ場合に、規則性の確認を行いつつ、各言語の語句をまとめて学習する事が可能である。
- ・データベースにおいて、音声再生は外部サーバーを利用するが、それ以外については、オフラインでの動作が可能である。

以下のウェブサイトでも閲覧できる。

<http://www.openesa.xyz>

## 目次

第1章 はじめに	1
1.1 プロジェクトの目的	1
1.2 役割分担	1
第2章 漢字語について	2
2.1 漢字について	2
2.2 韓国の漢字語	2
2.3 中国の漢字語	6
2.4 ベトナム語の歴史と漢字語の概要	7
第3章 日中韓越漢字データベース	10
3.1 4か国漢字語データベース	10
3.2 ベトナム語データベース	12
3.2 データベースの欠陥について	13
第4章 日中韓越漢字システム	14
4.1 システム概要	14
4.2 システム構造	15
4.3 漢字データフィルター	17
4.4 データの表示	19
4.5 フロントエンドデータベース	21
4.6 多端末対応	23
4.7 人工知能 API による日中韓音声合成	28
4.8 中韓越漢字語語彙検索	30
4.9 これからの課題	31
第5章 学習教材としてのデータベース利用の検証	32
5.1 韓国語漢字語彙	32
5.2 中国語漢字語彙	40
5.3 ベトナム語漢字	42
5.4 使用したときの感想	44
第6章 終わりに	45
謝辞	45
附録	48

# 第1章 はじめに

## 1.1 プロジェクトの目的

外国語を勉強するのが難しい。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で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人は漢越と言うものが使われてきた。現在は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。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。

本プロジェクトの目的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研究・学習するウェブシステムを開発することである。

開発内容は以下のとおりである。漢字文化圏、日本・中国・韓国・ベトナム、各言語の漢字語の音韻的類似性を研究・学習するウェブシステム開発する。

- ・ウェブシステムの開発によって、システム構築の実践力向上につなげる。
- ・漢字音単体のデータベースはすでに指導教員がつくったものを活用するがデータベースを表示する部分をウェブアプリケーションに対応させる
- ・漢字音を日本・中国・韓国・ベトナム、各言語から同音のものを中心に検索して漢字語の音韻的類似性を研究・学習する。
- ・各言語において単漢字の発音を合成できるようにする。
- ・さらに漢字の語彙を学習できるようにする。

## 1.2 役割分担

本システムの開発においては、ベトナム語のデータベース、韓国語のデータベース、データベースシステム(DBMS)の開発が必要であった。役割分担として、DO PHI TRUONG はベトナム語データベースの作成を担当する。山崎恭司は韓国語データベースの作成を担当する。陳守誠は漢字データベースシステム(DBMS)の作成を担当する。

本プロジェクトの指導教員である江見准教授は、2013年に論文「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」を発表した。しかし、ウェブ技術の激しい変化のため、現時点ではすでに稼働不能となった。これは本プロジェクトが実行された理由の1つである。

本プロジェクトの成果物である「日中韓越漢字システム」に搭載されているメイン漢字データベース及びピンイン、漢字、韓国語語頭語末フィルターは江見准教授の研究成果である。システム自体、各言語の語彙データベースは本プロジェクトの成果となる。

## 第2章 漢字語について

### 2.1 漢字について

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1つ1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」という同じの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。

例を理解するために、まずは母音及び子音の概念について紹介する。母音とは呼気の流れが妨げられないで発する言語音（長く伸ばせる）である。子音とは呼気の流れが妨げられて発する言語音（長く伸ばせない）である。他には、渡り音・流音・鼻音・摩擦音・閉鎖音などもある。聞こえる度を尺度とし、以下のように並べる。

聞こえる尺度（降順）

母音	a, e, o
渡り音	i(y), u(w)
流音	r, l
鼻音	m, n, ŋ
摩擦音	s, z, ʃ, ʒ
閉鎖音	p, b, t, d, k, g

### 2.2 韓国の漢字語

#### 2.2.1 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（下に例示）も多くある。本プロジェクトの韓国漢字語研究及びデータベースは市吉則浩，兼若逸之，韓虎林・今井久美雄・梅田博之，鶴見ユミの3冊の著作によるものである。

天下 (천하)	[ チョンハ ]	「天下」
結婚 (결혼)	[ キョルホン ]	「結婚」
愉快 (유쾌)	[ ユクウエ ]	「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」

### 2.2.2 日本起源の漢字語

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 [ピヘンギ])	飛機	フェイチー
汽車 (기차 [キチャ])	火車	ホウチャー
野球 (야구 [ヤグ])	棒球	パンチュー
会社 (회사 [ホイサ])	公司	ゴンス
写真 (사진 [サヂン])	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 (파괴 [パフウエ])	破棄
手帖 (수첩 [スチョプ])	手帳	洗滌 (세척 [セチョク])	洗淨

### 2.2.3 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・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] を付け動詞化がされる。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

生産 (생산 : 센サン)	「生産」	：「子供を生む」⇒ 「製品を作り出す」
放送 (방송 : 판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 : 팔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。

가난 [カナン]	「貧乏」	艱難 (간난)
성냥 [ソンニャン]	「マッチ」	石硫黄 (석류황)
감자 [カンヂャ]	「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 [チロソイ] 「ミミズ」	地龍이 (지렁이)
나중 [ナヂュン] 「のち」	乃終 (내중)
짐승 [チムスン] 「獸」	衆生 (중생)
대총 [テチュン] 「大凡」	大總 (대총)
천둥 [チャンドゥン] 「雷」	天動 (천동)
보배 [ポメ] 「宝」	寶貝 (보패)
초 [초] 「蠟燭」	燭 (촉) - [ともしび]
사냥 [サニャン] 「狩り」	山行 (산행)

#### 2.2.4 韓国におけるハングル

朝鮮の人々は独自の文字をもっていなかった。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。

しかし、漢字を利用しても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。

しかし、知識人や官僚たちは漢字を重んじてハングルの排斥されていたため、歴史的にハングルは民衆にほとんど使われてこなかった。

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。

#### 2.2.5 韓国の漢字廃止について

漢字語であらわ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た。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

社会 = 사회 : サフェ	射撃 = 사격 : サギョク
写真 = 사진 : サヂン	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

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) : サンチェック	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

- ・ 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。

## 2.3 中国の漢字語

### 2.3.1 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）」という論文を發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，錢玄同は「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）」を發表し，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年にも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」を發表している。

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」を發表した。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」を發表し，8年間の試行で廃止された。

### 2.3.2 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。

## 2.4 ベトナム語の歴史と漢字語の概要

### 2.4.1 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（ハンノム字）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。

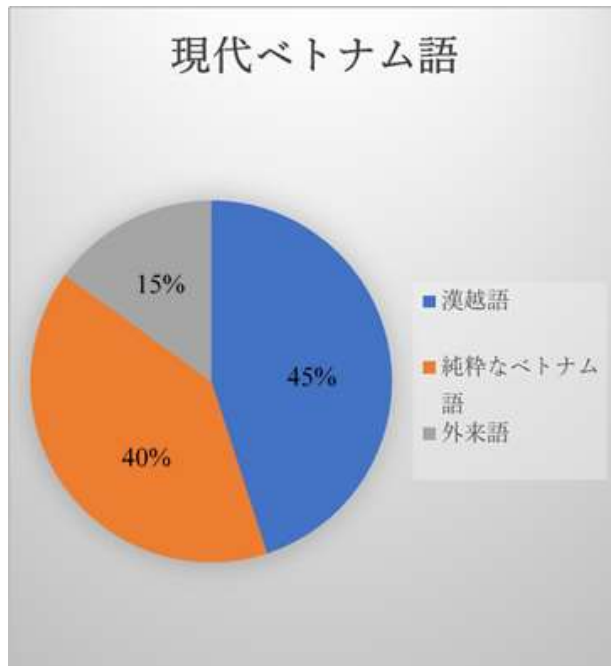


図2.1 現代ベトナム語の構成

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。

日本の漢字で良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。

漢字語

字  
学  
日  
本  
月

漢越語

Tự  
Học  
Nhật  
Bản  
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”という博士論文によると、漢字語対照データベースの分類基準は7種類ある。[10]

#### 2.4.2 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。

漢字語  
会談  
注意  
態度  
成立

漢越語  
Hội (会) Đàm(談)  
Chú (注) Ý (意)  
Thái(態) Độ (度)  
Thành (成) Lập (立)

#### 2.4.3 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

漢字語  
診察  
診断  
無断  
意見  
拝見

漢越語  
Chẩn(診) Sát(察)  
Chẩn(診) Đoán, Đoạn (断)  
Vô (無) Đoạn(断)  
Ý(意) Kiến (見)  
Bái (拝) Kiến (見)

#### 2.4.4 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。こう言う言葉は結構少ない。

漢字語  
階段  
段階

漢越語  
Giai (階) Đoạn(段)  
Đoạn (段) Giai (階)

#### 2.4.5 AB / BA

同じ意味を表す語が漢日語と漢越語では一字異なり、同一の字が異なる位置にある場合

漢字語  
学校  
学校  
空港  
空港  
節約  
節約

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

#### 2.4.6 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

漢字語	漢越語
貯金	Trù (貯) Kim (金)
貯金	Tiết (節) Kiêm (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。

漢字語	漢越語
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

漢字語	漢越語
電気	Điện (電) Khí (気)
電気	Điện (電)

## 第3章 日中韓越漢字データベース

### 3.1 4か国漢字語データベース

本システムの4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）、不完全な項目が随所であり、完成レベルにはほど遠い。言語分野で各言語に精通する方の協力を必要とする。

オリジナルのメイン漢字データベースの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、webブラウザの多言語表示問題はまだ数多く存在した。江見准教授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。表3.1.1が示しているのは、本システムで稼働されている漢字データベースの詳細である。

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、梅田博之の著作である「漢字でおぼえるしっぽをつかむ韓国語」[4]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」[2]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（「<http://hanvietnhat.vietnam55.com/?cat=11>」

「<http://kanji.tudiennhatviet.com/>」）によるものである。

本システムでの仕様は表3.1.3、表3.1.4及び表3.1.5に表示されている。

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	Jtousouon2 h	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	CTradional	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	CantonPyni n	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	Viethanzi	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 1	Jyomi	日本語における読み方, ケース 1
3	越南語読み方 2	Vyomi	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

### 3.2 ベトナム語データベース

始めて, 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, エクセルでデータを入力し, データベースを作った。

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2136字があり, その中に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
2	九	CỬU		chín
3	九九	くく	CỬU, CỬU CỬU, CỬU	bảng cửu chương
4	九月	くがつ	CỬU, CỬU NGUYẾT	tháng chín
5	九州	きゅうしゅう	CỬU, CỬU CHÂU	Kyushu
6	九十	くじゅう	CỬU, CỬU THẬP	chín mươi
7	九日	ここのか	CỬU, CỬU NHẬT, NHẬT	chín ngày
8	金	KIM		vàng
9	黄金	くがね	HOÀNG KIM	vàng
10	基金	ききん	CƠ KIM	Quỹ
11	金塊	きんかい	KIM KHỐI	vàng thỏi
12	金額	きんがく	KIM NGẠCH	Số tiền
13	金冠	きんかん	KIM QUAN, QUÂN	vương miện vàng
14	金柑	きんかん	KIM CAM	quất
15	金魚	きんぎょ	KIM NGƯ	cá vàng
16	金銀	きんぎん	KIM NGÂN	Vàng và bạc
17	金具	かなぐ	KIM CỤ	phụ kiện kim loại
18	金型	かながた	KIM HÌNH	khuôn
19	金庫	かなぐら	KIM KHỐ	an toàn
20	金鎖	きんぐさり	KIM TỎA	dây bằng vàng
21	金歯	きんば	KIM XỈ	răng vàng
22	金色	こんじき	KIM SẮC	màu vàng
23	金星	きんせい きんぼし	KIM TINH	sao Kim
24	金銭	きんせん	KIM TIỀN	tiền
25	金属	きんぞく	KIM CHẤT, THUỘC, CHỦ	kim khí
26	金権	てっぴい	KIM QUYỀN	búa
27	金髪	きんぱつ	KIM PHÁT	tóc vàng

図 3.2 漢字語と漢越語のエクセルデータベース

### 3.3 データベースの欠陥について

指導教員である江見准教授がメイン漢字データベースを 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。多音時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、本プロジェクトを完成するには、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する。しかし、データの不備などを考慮に入ると、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。



## 第4章 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の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・学習しやすいするためのものである。研究室や学校に止まらず、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言語の音声的類似性を学習・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」3の機能を有するウェブアプリケーションが出てきた。しかし、筆者（陳）ひとりの力では、お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するには無理がり、数多くの課題が残されている。例として、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が起こし、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、ユーザーインターフェースの最適化（誤操作への警告）なども完成できなかった。残された課題の解決に関して、本学の後輩の活躍が期待されている。

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。

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しており、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。

漢字語の音声即時再生のため、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 ライブラリーを使ってデータ処理を行なっている。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。簡略で不厳密に説明すると、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。3つのモデル合わせて、ようやく1つのプログラムとして動けるようになる。

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。

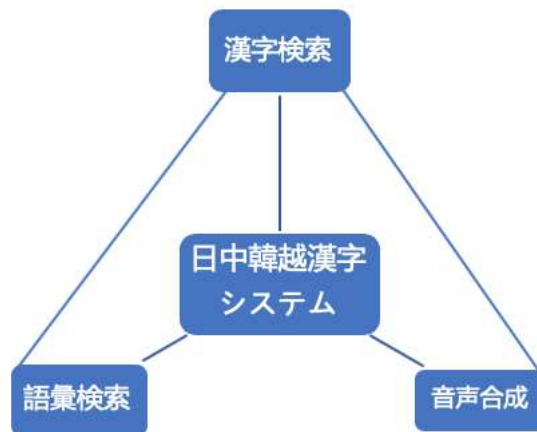


図 3.2.1 システム構造

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、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ことである。

#### 4.2.1 漢字検索

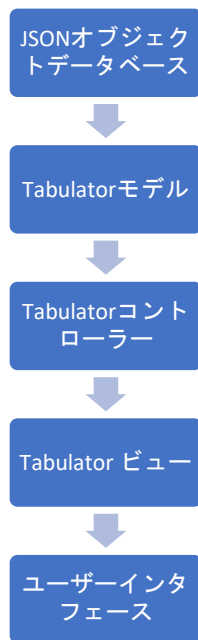


図 4.2.1.1 読み取りプロセス

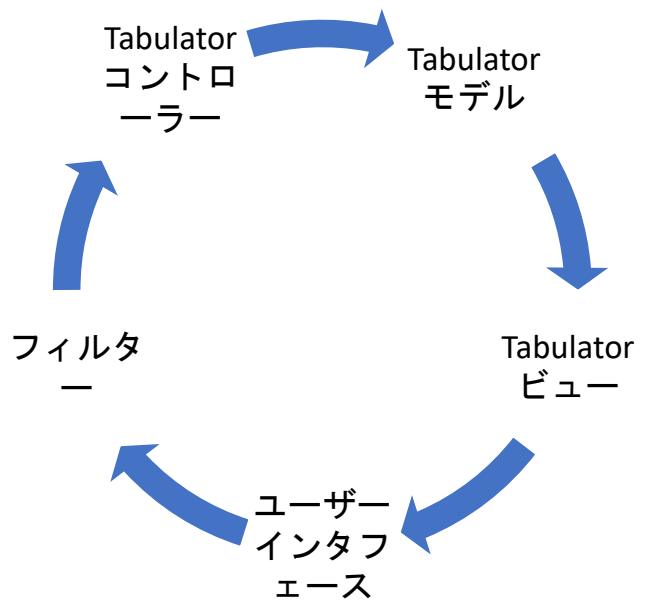


図 4.2.1.2 漢字検索ループ

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本として他の処理を実行する。

図 4.2.1.1 が示すように、ブラウザがページの読み取りを始めた瞬間、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 の一部として取り組み、これでページの初期レンダリングが終わる。

漢字検索が行う場合、図 4.2.1.2 が示しているループに入る。まず、ユーザーがインター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、「3.3 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、「3.7 人工知能 API により日中韓音声合成」「3.8 中韓越漢字語語彙検索」に参照してください。

### 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，全ての中国語音声が入力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）。

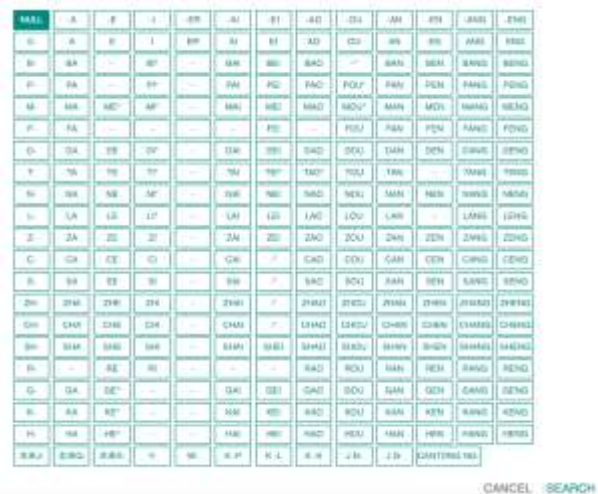
#### 4.3.2 ピンインフィルターグループ

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 4.3.2.1 が示しているように，全ての中国漢字音を 3 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」が存在する漢字を指す。

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。「-」は対応する漢字音は存在しない



韻母	A	E	I	ER	AI	EI	AO	OU	AN	EN	ANG	ENG
C	A	E	I	ER	AI	EI	AO	OU	AN	EN	ANG	ENG
B	BA	BE	BI	BER	BAI	BEI	BAO	BOU	BAN	BEAN	BANG	BENG
F	FA	FE	FI	FER	FAI	FEI	FAO	FOU	FAN	FEN	FANG	FENG
M	MA	ME	MI	MER	MAI	MEI	MAO	MOU	MAN	MEN	MANG	MENG
P	PA	PE	PI	PER	PAI	PEI	PAO	POU	PAN	PEN	PANG	PENG
T	TA	TE	TI	TER	TAI	TEI	TAO	TOU	TAN	TEN	TANG	TENG
L	LA	LE	LI	LER	LAI	LEI	LAO	LOU	LAN	LEN	LANG	LENG
Z	ZA	ZE	ZI	ZER	ZAI	ZEI	ZAO	ZOU	ZAN	ZEN	ZANG	ZENG
C	CA	CE	CI	CER	CAI	CEI	CAO	COU	CAN	CEN	CANG	CENG
S	SA	SE	SI	SER	SAI	SEI	SAO	SOU	SAN	SEN	SANG	SENG
ZH	ZHA	ZHE	ZHI	ZHER	ZHAI	ZHEI	ZHAO	ZHOU	ZHAN	ZHEN	ZHANG	ZHENG
CH	CHA	CHE	CHI	CHER	CHAI	CHEI	CHAO	CHOU	CHAN	CHEN	CHANG	CHENG
SH	SHA	SHE	SHI	SHER	SHAI	SHEI	SHAO	SHOU	SHAN	SHEN	SHANG	SHENG
R	RA	RE	RI	RER	RAI	REI	RAO	ROU	RAN	REN	RANG	RENG
G	GA	GE	GI	GER	GAI	GEI	GAO	GOU	GAN	GEN	GANG	GENG
K	KA	KE	KI	KER	KAI	KEI	KAO	KOU	KAN	KEN	KANG	KENG
H	HA	HE	HI	HHER	HAI	HEI	HAO	HOU	HAN	HEN	HANG	HENG
X	XA	XE	XI	XHER	XAI	XEI	XAO	XOU	XAN	XEN	XANG	XENG
Y	YA	YE	YI	YER	YAI	YEI	YAO	YOU	YAN	YEN	YANG	YENG
YU	YU	YU	YU	YUR	YU	YU	YU	YU	YU	YU	YU	YU

図4.3.2.1 検索ボタングループ

図4.3.2.2 ピンイン検索ボタンメニュー

ことを示す。最後の1行において，「本来\*」は歴史的な仮名使いの特例である。大文字Kは韓国語を意味する。大文字Jは日本を意味する。CANTONGは中国語広東を意味する。「\*-」は語頭で，「-\*」は語尾である。詳しくの言語的規則は江見准教授の論文に参照してください。

### 4.3.3 ハングルフィルターグループ

#### 語頭

KP-	KB-	KM-	KK-	KG-	KH-	KT-	KD-	KN-	KR-	KCH-	KJ-	KS-
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	-----	-----

#### 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

JN-	JG-	CANTONG NG-
-----	-----	-------------

図 4.3.3.1 韓国語・語頭・語末検索ボタンメニュー

ハングルは、朝鮮語を表記するための表音文字である。江見准教授は日本語、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・語尾フィルターは江見准教授の研究によるものである。

### 4.3.4 カスタム検索

前文に述べた通り、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のままに検索できる。新しい言語的ルールの探求にやくた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。使いこなせる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も出ない。対処が必須と思われる。

The screenshot shows a search interface with three rows of search criteria. Each row consists of a search field, an operator, and a search content field. A dropdown menu is open over the first operator field, showing options: '=', '!=', 'Like', 'in', and 'Regex'. The first search field is active, and the dropdown is currently showing the '=' option. At the top right, there is a 'フィルター数' (Filter Count) dropdown set to '1'. At the bottom right, there are 'CANCEL' and 'SEARCH' buttons.

図 4.3.4.1 カスタム検索メニュー

図 4.3.4.1 が示している通り，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，文字的の完全一致を意味する。

「!=」記号は，文字的の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 演算子は一番機能の強い演算子とも言われるが，使い方が結構複雑で正規表現という技術が使われている。理論的にどんな文字列でも特定できる。詳しくは [https://developer.mozilla.org/ja/docs/Web/JavaScript/Guide/Regular\\_Expressions?fbclid=IwAR3o3C2fGPXpRv3s0mYQ4dW10U\\_GQVp1850ChlgSsb4Wi9R4FLRZF4VfHw](https://developer.mozilla.org/ja/docs/Web/JavaScript/Guide/Regular_Expressions?fbclid=IwAR3o3C2fGPXpRv3s0mYQ4dW10U_GQVp1850ChlgSsb4Wi9R4FLRZF4VfHw) を参照。

## 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す」という理念をもとに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は語彙中心で，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4.4.1.1 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 4.5.1.2 は本システムである。

未来		
日本語	中国語	韓国語
1 未来	17 未来	11 未來
2 みらい	18 weilai	12 미라이 mirai
品詞性		
日本語		
3 『新明解』	4 『角川』	5 『岩波』
6 『明鏡』	7 『旺文社』	
名	名	名
中国語		
19 『漢語辞典』	20 『现代辞典』	21 両辞書の一致性
名・形	名・形	✓
韓国語		
13 『한글』	14 『한글』	15 『한글』
한글 動詞	한글 動詞	한글 動詞
X	X	X
16 미		
		X
日中語学的関係		
22 文化庁 (1978)	23 『中日対比辞典』	
◎	S	

図 4.4.1.1 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



項目	漢字	漢音	漢音	漢音	漢音	ピンイン	声母	韻母	声調	ハングル	備考
日漢	一	イチ	イツ	イツ	イツ	ichi	チ	i	1	일	
呉音	イ	イ	イ	イ	イ	i				이	
漢音	イ	イ	イ	イ	イ	i				이	
漢音2	イ	イ	イ	イ	イ	i				이	
唐宋音	イ	イ	イ	イ	イ	i				이	
慣用音	イ	イ	イ	イ	イ	i				이	
学年	イ	イ	イ	イ	イ	i				이	
繁漢	一	一	一	一	一	ichi	チ	i	1	일	
簡漢	一	一	一	一	一	ichi	チ	i	1	일	
ピンイン	一	一	一	一	一	ichi	チ	i	1	일	
声母	一	一	一	一	一	ichi	チ	i	1	일	
韻母	一	一	一	一	一	ichi	チ	i	1	일	
声調	一	一	一	一	一	ichi	チ	i	1	일	
広東	一	一	一	一	一	ichi	チ	i	1	일	
レベル	一	一	一	一	一	ichi	チ	i	1	일	
ハングル	一	一	一	一	一	ichi	チ	i	1	일	
ハングル2	一	一	一	一	一	ichi	チ	i	1	일	
韓音	一	一	一	一	一	ichi	チ	i	1	일	
韓音2	一	一	一	一	一	ichi	チ	i	1	일	
越南	一	一	一	一	一	ichi	チ	i	1	일	
越南2	一	一	一	一	一	ichi	チ	i	1	일	

図 4.4.1.2 本システムのデータ表示

#### 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 ライブラリーに依存している。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。

```
var table = new Tabulator("#example-table", {
  height: h,
  selectable:1,
  //layoutColumnsOnNewData:true,
  layout:"fitColumns",
  columnMinWidth:115,
  //responsiveLayout:"collapse",
  tooltipsHeader:true,
  //resizableColumns:false,
  columns:[
    {title:"日漢", field:"JKanji"},
    {title:"呉音", field:"JGoon1h"},
    {title:"呉音2", field:"JGoon2h", visible:false},
    {title:"漢音", field:"JKanon1h"},
    {title:"漢音2", field:"JKanonh", visible:false},
    {title:"唐宋音", field:"Jtousouon2h", visible:false},
    {title:"慣用音", field:"JKanyouonh", visible:false},
    {title:"学年", field:"Jgakunen", visible:false},
    {title:"繁漢", field:"CTradional"},
    {title:"簡漢", field:"CSimplified"},
    {title:"ピンイン", field:"Cpinyin"},
    {title:"声母", field:"Cinitial", visible:false},
    {title:"韻母", field:"Cfinal", visible:false},
    {title:"声調", field:"Ctone"},
    {title:"広東", field:"CantonPynin"},
    {title:"レベル", field:"level", visible:false},
    {title:"ハングル", field:"KHangul"},
    {title:"ハングル2", field:"KHangul2", visible:false},
    {title:"韓音", field:"Kromaji", visible:false},
    {title:"韓音2", field:"Kromaji2", visible:false},
    {title:"越南", field:"Vietnam1"},
    {title:"越南2", field:"Vietnam2", visible:false}
  ],
});
```

図 4.4.2.1 漢字データテーブルコントロールコード

図 4.4.2.1 が示しているように、このデータテーブルから参照できるデータは20列ある。ユーザ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。けれども、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、そのモニターでちょうど最も重要な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インチより小さいに対して、フィールドの幅が自動的に大きくなる。

<input checked="" type="checkbox"/> 日漢	<input checked="" type="checkbox"/> 繁漢	<input checked="" type="checkbox"/> ハングル	<input checked="" type="checkbox"/> 越南
<input checked="" type="checkbox"/> 吳音	<input checked="" type="checkbox"/> 簡漢	<input type="checkbox"/> ハングル2	<input type="checkbox"/> 越南2
<input type="checkbox"/> 吳音2	<input checked="" type="checkbox"/> ピンイン	<input type="checkbox"/> 韓音	
<input checked="" type="checkbox"/> 漢音	<input type="checkbox"/> 声母	<input type="checkbox"/> 韓音2	
<input type="checkbox"/> 漢音2	<input type="checkbox"/> 韻母		
<input type="checkbox"/> 唐宋音	<input checked="" type="checkbox"/> 声調		
<input type="checkbox"/> 慣用音	<input checked="" type="checkbox"/> 広東		
<input type="checkbox"/> 学年	<input type="checkbox"/> レベル		

CANCEL SEARCH

図 4.4.2.2 図 データ列選択メニュー

4.4.2.2 が示しているのはデータ列選択メニュー、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。

#### 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、3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す。二つ目、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い、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程膨大ではないから、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。

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、インターネットに繋いで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、ブラウザを閉じてでも 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 cache 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である。

#### 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「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。一般的なリレーショ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の特徴がある。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 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 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す。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なえる形式である。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": "亚", "CTradional": "亞", "Cpinyin": "ya", "Cinitial": "", "Cfinal": "ya", "Ctone": "4", "level": "1", "KHangul": "야", "Kromaji": "a", "Vietnam1": "Á", "Vietnam2": "", "CantonPynin": "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x. 6.1 に参照してください。

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

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

Range: 1.70 s – 3.15 s



図4.5.2.1 初期化パフォーマンス



図4.5.2.2 漢字検索パフォーマンス

図4.5.2.1及び図4.5.2.2は Google chrome performance console から取り出したデータである。4.5.2.1は、最初にページを読み込み時のパフォーマンス詳細である。データベースを読み込みを含み、全ての JavaScript が実行完了には、約 1500ms(1.5s) がかった。4.5.2.2は、データ検索操作のパフォーマンス詳細である。JavaScript には、やく 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。

#### 4.6 多端末対応

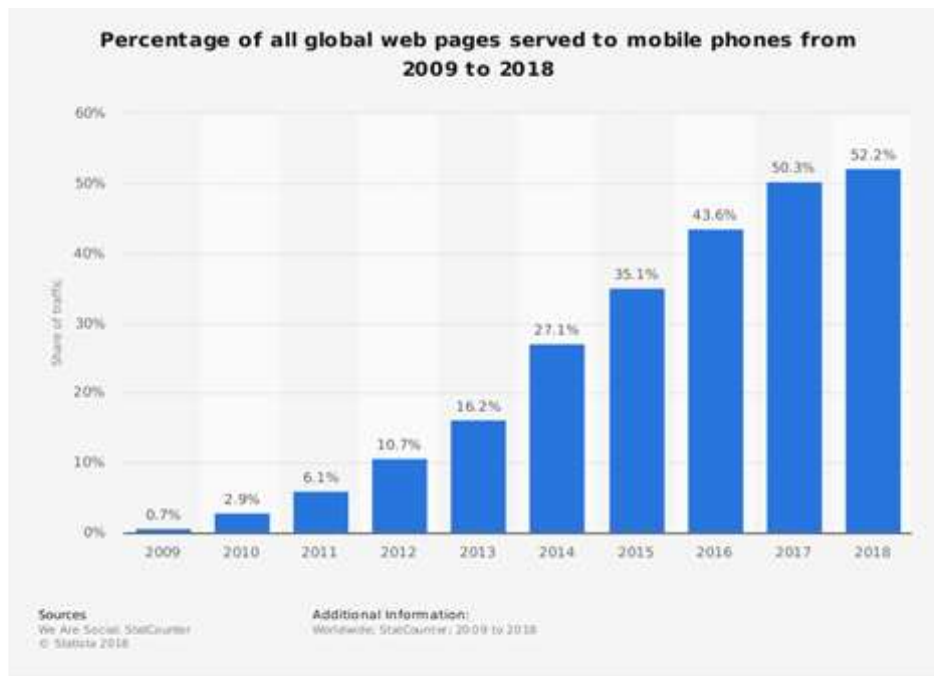


図 4.6.1 Percentage of all global web pages served mobile phones from 2009 to 2018

#### 4.6.1 モバイルファーストデザインの重要性

Statistaの調査（図 4.6.1）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，一番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，他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 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。図 4.6.2 が示しているのは StatCounter の調査による 2017.12 から 2018.12 までの世界的ブラウザの占有率遷移である。

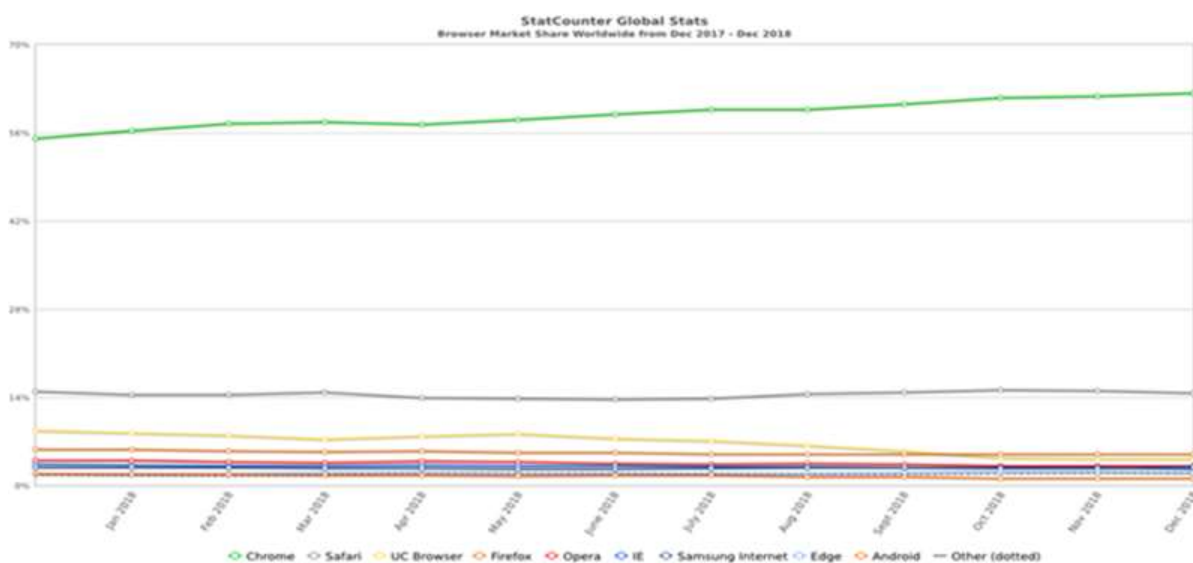


図 4.6.2 Browser Market Share Worldwide From Dec 2017 - Dec 2018

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 この4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勞した。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まで，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，全自動的な音声合成（自動再生）は今でも実現していない，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1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

多種多様の端末が登場し続けているいま、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。

MDNによると、CSS アニメーションとは、あるCSSのスタイル設定を別の設定へ遷移させることを可能にします。アニメーションは、2種類の要素で構成されています。それは、アニメーションについて記述するスタイルと、アニメーションの始めと終わりのCSSスタイルを示すキーフレームである。特徴のある機能を提供するため、本システムのメニューの開閉アニメーションはオープンソースのフレームワークによるものではなく、独自に開発されたものである。

日中韓越漢字							
	日漢	興音	漢音	繁漢	簡漢	ピンイン	声調
RESET							
データ列選択	亞	エ	ア	亞	亚	ya	4
漢字データ検索	哀	オ	アイ	哀	哀	ai	1
ピンイン・介音なし	愛	オ	アイ	愛	爱	ai	4
ピンイン・介音WU	惡	アク	アク	惡	恶	e	4
ピンイン・介音u	握	アク	アク	握	握	wo	4
ハングル・語頭・語末	圧	エフ	アフ	壓	压	ya	1
カスタム検索	扱	セウ	サウ	扱	扱	xi	1
音声合成	安	アン	アン	安	安	an	1
日本語音声・興音	暗	オム	アム	暗	暗	an	4
日本語音声・興音2	案	アン	アン	案	案	an	4
日本語音声・漢音	以	イ	イ	以	以	yi	3
日本語音声・漢音2	位	ヰ	ヰ	位	位	wei	4
中国語音声							
韓国語音声	依	エ	イ	依	依	yi	1

図4.6.3 タブレットデバイス（min-width:768px, max-width:992px）の実例

前文に書いた通り、本システムは違う端末で異なる形として表示されている。スマートフォン(max-width :768px)向けのページの横置き実例は図 3. 6. 4. 1 及び図 3. 6. 4. 2 に参照する。また、max-width :768px のデバイスに対し、デフォルトとしてメニューは表示していない。図 4. 6. 3 はタブレットデバイス ( min-width:768px, max-width:992px ) 向けの縦置きページの実例である。デスクデバイス (min-width: 992px) 向けのページ の実例は図 3. 4. 1. 2 に参照してください。

#### 4. 6. 4 Full Screen API

MDN によると、Full Screen API とは、特定の Element (およびその子孫) を全画面モードで表示したり、必要なくなったときに全画面モードを抜けたりする方法を追加します。これによって、要求されたコンテンツ - オンラインゲームなど - がユーザーの画面全体で表示され、全画面モードが終了するまで、ブラウザのユーザーインターフェイス要素や他のアプリケーションをすべて画面から排除することができる。

前文に述べたように、ブラウザの違いにより表示結果に大きな違いが生じる。表示幅の小さいモバイル端末に対して、ブラウザのアドレス欄は大きな問題になる。図 4. 6. 4. 1 が示しているように、横置きのモバイルデバイス場合、アドレス欄は3分の1以上の縦幅を占めることのあり、データ表示に大きな支障をもたらす。Full Screen APIが適用してから、図 3. 6. 4. 2 で示した通り、表現がかなり良くなってきた。今の所、本システムのFull Screen機能はまだ safari ブラウザーに対応してない。1つの課題として残っている。



図4. 6. 4. 2 モバイルchromeでの Full Screen 表現

#### 4.6.5 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、データベースを含みおおよそ 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 コードをちょっと変えることで、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 4.5.2.1 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、図3.6.5のように、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 4.6.5 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。操作法は4.2.3で説明した。

その中、日本語および韓国語の音声の Google 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 に頼っている。同じく音声合成 API であるが、操作法は結構違っている。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。

全文に述べたように、本システムできる限りフロントエンド技術で機能を実現したいのである。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 問題が回避するため、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ための方法である。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2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り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1つとして残っている。



### 4.7.3 音声合成モデルの構造

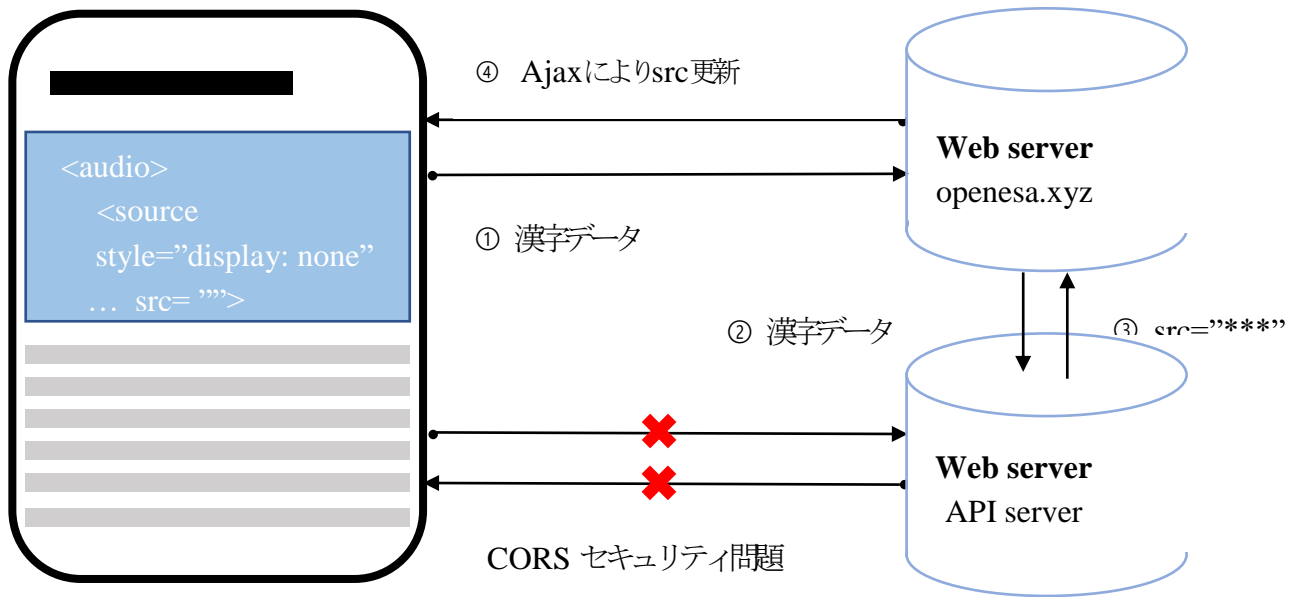


図 4.7.3 音声合成モデルにおいてデータの流れ

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ありませんが, **Jesse James Garrett** によって 2005 年に作られた言葉で, 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早く, 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更新する技術が不可欠である。

図 4.7.3 が示しているように, 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 (`http://openesa.xyz`) 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 (`http://openesa.xyz`) に発送する。システムサーバ (`http://openesa.xyz`) で, **PHP** により二次処理が行い, 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 (http://openesa. xyz) に送る。システムサーバ

(http://openesa. xyz) 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。今のところでは、対処できる方法がない。

#### 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。図 3.8 が示しているのはベトナム語語彙検索関数 viet\_voc()。

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 の処理スピードに心配していたが、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。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、これから課題である。

```
function viet_voc(){
  selectedData = table.getSelectedData();
  if(selectedData[0] == undefined){
    $("#viet_model").html("<p>Please select a character.</p>");
  }
  else{
    var table_title = "";
    var table_str = "";
    $("#viet_model_title").html("");
    table_title = "<h5 class='model-title'> " + selectedData[0].Jkanji + "&nbsp;&nbsp;&nbsp;" + selectedData[0].Vietnam1 + "&nbsp;&nbsp;&nbsp;" +
    selectedData[0].Vietnam2 + "</h5>";
    $("#viet_model_title").append(table_title);
    $("#viet_model").html("");
    table_str += "<table style='border-collapse: collapse; border: 1px solid black;'><thead><tr><th style='border: 1px solid black;'>日本語漢字</th><th style='border: 1px solid black;'>日本語読み方</th><th style='border: 1px solid black;'>越語漢字</th><th style='border: 1px solid black;'>越語読み方</th></thead><tbody>";
    for(var i = 0; i < tableData2.length; i++){
      if(tableData2[i].Jkanji.includes(selectedData[0].Jkanji)){
        table_str += "<tr><td style='border: 1px solid black;'>" + tableData2[i].Jkanji + "</td><td style='border: 1px solid black;'>" +
        tableData2[i].Jyomi + "</td><td style='border: 1px solid black;'>" + tableData2[i].Vyomi + "</td><td style='border: 1px solid black;'>" + tableData2[i].Vimj + "</td></tr>";
      }
    }
    table_str += "</tbody></table>";
    $("#viet_model").append(table_str);
  }
}
```

図4.8 ベトナム語語彙検索関数 viet\_voc()

#### 4.9 これからの課題

最後に今までで、解決できなかった技術問題をまとめる。

1. safari ブラウザーへの対応 (iOS safari の全自動音声再生, iMac ・ iOS safari の Full Screen API への対応)。
2. JavaScript CORS 問題への対応。この問題が解決できたら, システムの完全オフライン化が期待できる。
3. 語彙検索モデルの再構築。ユーザーのデバイスに過剰な負担をかけないことを前提とし, 機能が拡張できる新しいやり方での構築。
4. 漢字データベース及びサブ語彙データベースの完成度を上げること。

# 第5章 学習教材としてのデータベース 利用の検証

まずは付録1に示したようなデータベースの使い方を修得してもらおう。次に、付録2を参考にして歴史的仮名遣いをマスターしてもらおう。そのうえで、この章では、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を示したい[22]。

## 5.1 韓国語漢字語彙

まずは、韓国語の漢字語彙を示す。  
調べたい漢字の行を選択して、左端の“韓国語・漢字語彙 B” [2] をクリックする



図 5.1.1 操作画面 1

選択した漢字を使った語彙が表示される

(“ 의미 : 意味” の列での表示は、日本の漢字語をハングルに直したものが使われていない場合)



図 5.1.2 操作画面 2

カスタム検索で、検索フィールド 1 > “ハンゲル” + “=” + “중”  
で検索した場合

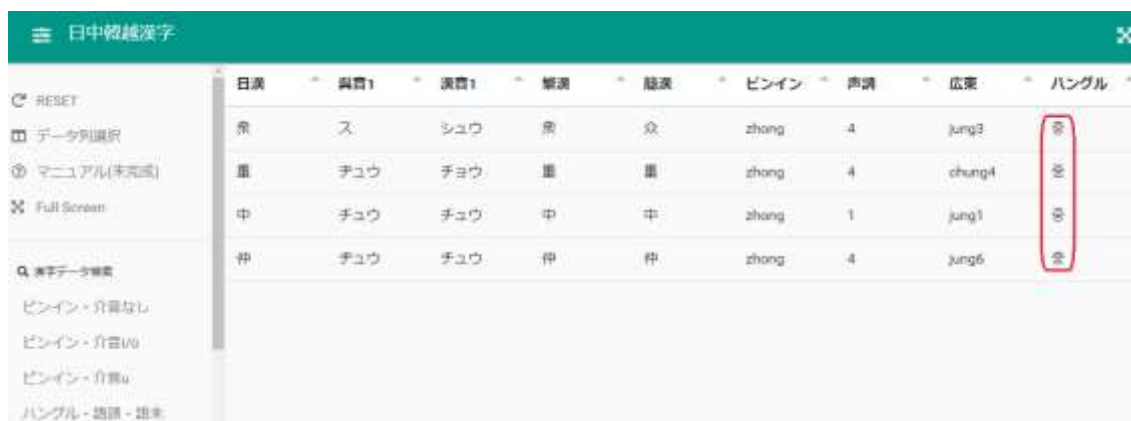


図 5.1.3 操作画面 3

ハンゲルで “중” と書く漢字が表示される

たとえば、以下のようなブログを記事を読んで語彙を増やしたいとする。

[参照：<http://honyakuka.jugem.jp/?pid=1> (韓国語通訳の大北章二さんのブ

ログ・2011.01.29) [23] の記事をもとに「語末のㅇ[ng]」を調べる

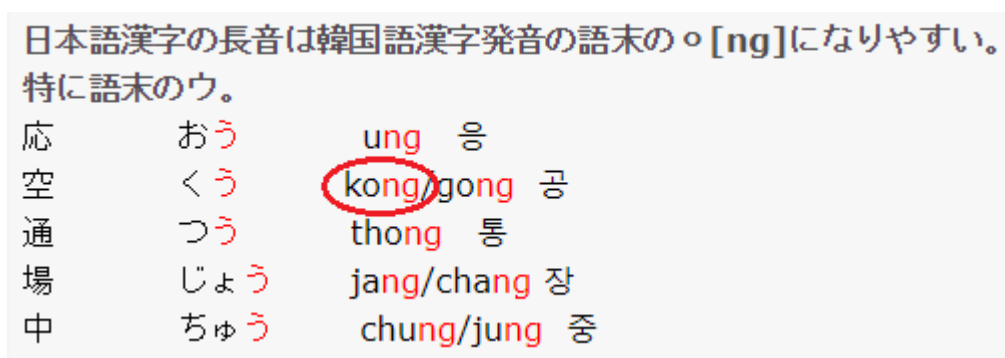


図 5.1.4 韓国語通訳の大北章二さんのブログ 1

ハンゲル・語末の “K-ONG” で検索すると

	日漢	編音1	漢音1	新漢	語漢	ピンイン	声調	広東	ハングル
RESET	葎	ウ	ゾウ	葎	葎	weng	1	yung1	ㅇ
データ別選択	供	ク	キョウ	供	供	gong	1	gung1	ㅇ
マニュアル(未完成)	共	ク	キョウ	共	共	gong	4	gung1	ㅇ
Full Screen	恐	ク	キョウ	恐	恐	kong	3	hung2	ㅇ
漢字データ検索	恭	ク	キョウ	恭	恭	gong	1	gung1	ㅇ
ピンイン・介算なし	空	クウ	コウ	空	空	kong	1	hung1	ㅇ
ピンイン・介算あり	公	ク	コウ	公	公	gong	1	gung1	ㅇ
ピンイン・介算なし	功	ク	コウ	功	功	gong	1	gung1	ㅇ
ハングル・語頭・語末									
カスタム検索									

図 5.1.5 操作画面 4

“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”になりやすいという記事の内容が検証できた。

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、また他国での読みの変化などの共通性を読みとることもできる。

日本語漢字の語末子音はそのまま韓国語漢字発音に使える場合が多い。

見、犬	けん	kyun/gyun	견
韓	かん	han	한
線、先	せん	sun	선
学	がく	hak	학
格	かく	kyuk/gyuk	격
箔、博	はく	pak/bak	박
測、側	そく	chuk	측
<b>【参考】</b>			
撃、激	げき	gyuk/kyuk	격
席	せき	suk	석
責	せき	chaek	책

図 5.1.6 韓国語通訳の大北章二さんのブログ 2

“격” = “kyuk” を調べるため、ハングル・語末の “K-YEOK” で検索した場合（検索後に「ハングル▲」をクリックしてソート）

日漢	異音1	漢音1	異漢	羅漢	ピンイン	声調	広東	ハングル
格	キヤク	カク	格	格	ge	2	gaa63	리
隔	キヤク	カク	隔	隔	ge	2	gaa63	리
撃	キヤク	グキ	撃	击	j	1	gk1	격
激	キヤク	グキ	激	激	j	1	gk1	격
壁	ヒヤク	ヘキ	壁	壁	bi	4	ba1	벽
癖	ヒヤク	ヘキ	癖	癖	pi	3	pa1	벽
域	キキ	ヨク	域	域	yu	4	wik6	역
疫	ヤク	エキ	疫	疫	yi	4	yik6	역

図 4.1.7 操作画面 5

“格”・“撃”・“激”の「ハングル」表示の一致が確認でき、記事の内容が検証できた。

日本語漢字発音「エ段+い」は韓国語漢字発音で-yung -yeになる場合が多い。

永、英、映	えい	yung 영
芸	(げい)	ye 예
迎	(げい)	yung 영
軽、慶	(けい)	kyung/gyung 경
恵	(けい)	hye 혜
系、計	(けい)	kye/gye 계
訂、定	てい	chung/jung 정
寧	ねい	nyung/yung 녀/영
制、製	せい	che/je 제
例、礼	れい	rye/ye 레/예
令、靈、零	れい	ryung/yung 령/영

図 5.1.8 韓国語通訳の大北章二さんのブログ 3

カスタム検索で、検索フィールド1 > “漢音1” + “=” + “テイ”  
で検索



図 5.1.9 操作画面 6

(検索後に「声調▲」をクリックしてソート)



図 5.1.10 操作画面 7

日本語漢字発音「エ段+い」は韓国語漢字発音で **-yung -ye**になる場合が多い。

永、英、映	えい	<b>yung 영</b>
芸	(げい)	<b>ye 예</b>
迎	(げい)	<b>yung 영</b>
軽、慶	(げい)	<b>kyung/gyung 경</b>
恵	(げい)	<b>hye 헤</b>
系、計	(げい)	<b>kye/gye 계</b>
訂、定	てい	<b>chung/jung 정</b>
寧	ねい	<b>nyung/yung 녀/영</b>
制、製	せい	<b>che/je 제</b>
例、礼	れい	<b>rye/ye 레/예</b>
令、靈、零	れい	<b>ryung/yung 령/영</b>

図 5.1.11 韓国語通訳の大北章二さんのブログ 4

この場合、日本語漢字（漢音≒音読み）が同じ2つの漢字が、ハングルおよび中国・ベトナム語の読みでも一致していることが分かる。

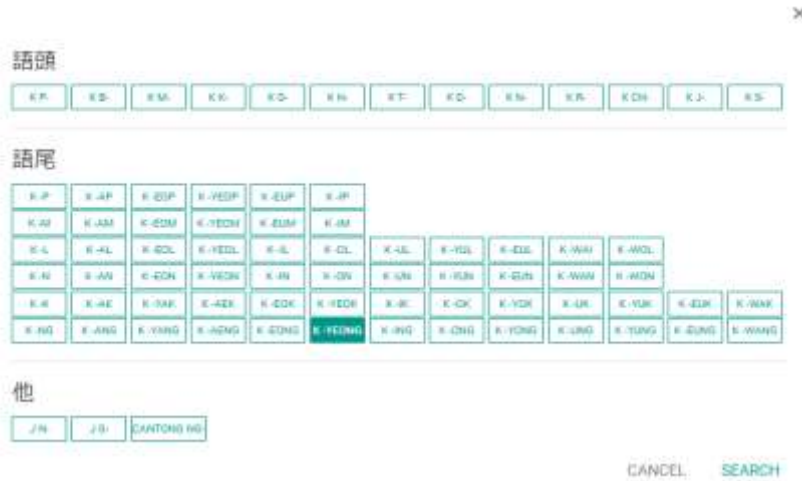


図 5.1.12 操作画面 8

ハングル・語末（語尾）の “K-YEONG” で検索



図 5.1.13 操作画面 9

記事のとおり、ハングル表記 “영” での一致が確認できる。

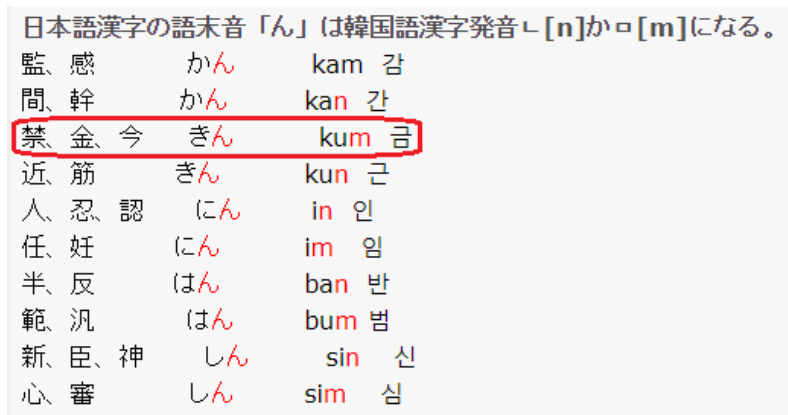


図 5.1.14 韓国語通訳の大北章二さんのブログ 5





図 5.1.15 操作画面10

カスタム検索で、検索フィールド1 > “呉音1” + “=” + “コム” で検索

日漢	呉音1	漢音1	簡漢	簡漢	ピンイン	声調	広漢	ハングル
勘	コム	カム	勘	勘	kan	1	ham3	간
堪	コム	カム	堪	堪	kan	1	ham1	간
感	コム	カム	感	感	gan	3	gam2	간
筭	コム	キム	筭	筭	jin	4	gam1	간
理	コム	キム	理	理	je	4	kam1	간
金	コム	キム	金	金	jin	1	gam1	간
欠	コム	クム	欠	欠	qian	4	him3	간
劍	コム	クム	劍	劍	jan	4	gim3	간
今	コム	キム	今	今	jin	1	gam1	간

図 5.1.16 操作画面11

記事の検証が完了

(「呉音」の“コム”は、発音が現在では「コン」になっているものである。)

ハングル・語頭の “K B-” で検索

日本語漢字発音でハ行は、韓国語漢字発音でㅂ[p/b]、ㅍ[ph]になる場合が多い。

派	は	pha	파
俳	はい	bae	배
被	ひ	phi	피
普	ふ	po/bo	보
<b>兵</b>	へい	pyung/byung	<b>병</b>
本	ほん	pon/bon	본

図 5.1.17 韓国語通訳の大北章二さんのブログ 6

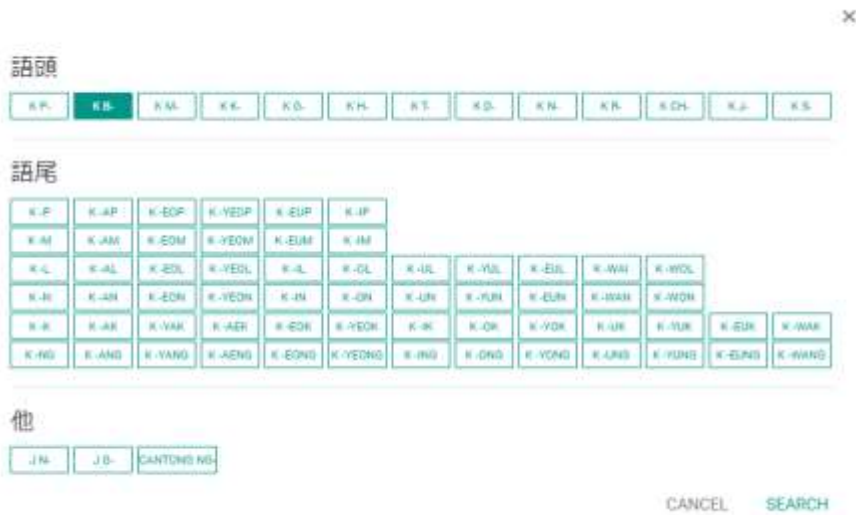


図 5.1.18 操作画面12

日漢	漢音1	漢音1	韓漢	韓漢	ピンイン	漢音	広東	ハングル
別	ベチ	ヘツ	別	別	bie	2	bin6	ㅍ
莖	ビヤウ	ヘ	莖	莖2	bing	4	bing6	ㅎ
莖	ビヤウ	ヘイ	莖	莖	bing	4	beng6	ㅎ
瓶	ビヤウ	ヘイ	瓶	瓶	ping	2	peng2	ㅎ
丙	ビヤウ	ヘイ	丙	丙	bing	3	bing2	ㅎ
併	ヒヤウ	ヘイ	併	併1	bing	4	bing3	ㅎ
<b>兵</b>	<b>ヒヤウ</b>	<b>ヘイ</b>	<b>兵</b>	<b>兵</b>	<b>bing</b>	<b>1</b>	<b>bing1</b>	<b>ㅎ</b>
柄	ヒヤウ	ヘイ	柄	柄	bing	3	beng3	ㅎ
歩	ブ	ホ	歩	歩	bu	4	bou6	ㅃ

図 5.1.19 操作画面13

検索結果より、“병”との対応が検証できる。

## 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3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みれば、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、「青」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 頻出単語が表示される。

江見准教授によると、京都情報大学院の10年前ぐらいの学生がこのデータベースを用いて、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と聞いている。ところが中国人である筆者（陳）が日本語を学習する上においては、およそ想像がつく結果であるので、とても役立つというレベルではない。今後検証したいのは韓国人やベトナム人が中国語を学習するときどの程度役立つかである。



図 4.2.1 操作画面14

✕

NULL	-I	-IA	-IE	-IAO	-IU	-IAN	-IN	-IANG	-ING	-Ü	-ÜE	-ÜAN	-ÜN	-HONG
Y-	YI	YA	YE	YAO	YOU	YAN	YIN	YANG	YING	YU	YUE	YUAN	YUN	YONG
B-	BI	-	BIE	BIAO	-	BIAN	BIN	-	BING	-	-	-	-	-
P-	PI	-	PIE	PIAO	-	PIAN	PIN	-	PING	-	-	-	-	-
M-	MI-	-	MIE	MIAO	-	MIAN	MIN	-	MING	-	-	-	-	-
F-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D-	DI	-	DIE	DIAO	-	DIAN	-*	-	DING	-	-	-	-	-
T-	TI	-	TIE	TIAO	-	TIAN	-*	-	TING	-	-	-	-	-
N-	NI	-	NIE	NIAO	NIU	NIAN	NIN	NIANG	NING	NÜ	NÜE	-	-	-
L-	LI	LIA	LIE	LIAO	LIU	LIAN	LIN	LIANG	LING	LÜ	LÜE	-	-	-
J-	Ji	JIA	JIE	JIAO	JIU	JIAN	JIN	JIANG	JING	JU	JUE	JUAN	JUN	JIONG
Q-	Qi	QIA	QIE	QIAO	QIU	QIAN	QIN	QIANG	QING	QU	QUE	QUAN	QUN	QIONG
X-	XI	XIA	XIE	XIAO	XIU	XIAN	XIN	XIANG	XING	XU	XUE	XUAN	XUN	XIONG
本来J-	本来Q-	本来X-	W-	子音なし	K-P	K-L	K-K	J-N-	J-G-	CANTONG NG-				

CANCEL    SEARCH

図 4.2.2 操作画面15

目録	異音1	異音1	異音	韻母	ピンイン	声調	広標	ハングル	越南1
ほ	キヤフ	クイ	ほ	ほ	-ing	1	hng1	호	HO-THAI
庫	キヤフ	クイ	庫	ハ	-ing	4	hng5	고	HO-SAMH
靴	キヤフ	クイ	靴	ハ	-ing	1	hng1	고	HO-SNH
構	シヤフ	セイ	構	構	-ing	2	ching4	기	THAI
構	シヤフ	セイ	構	構	-ing	2	ching4	기	THAI
構	シヤフ	セイ	構	構	-ing	1	ching1	기	THANH
構	シヤフ	セイ	構	構	-ing	3	ching3	기	THINH
高	シヤフ	セイ	高	高	-ing	1	ching1	기	THANH
部	キヤフ	クイ	部	部	-ing	1			
構	キヤフ	クイ	構	部	-ing	3			
構	シヤフ	セイ	構	構	-ing	1			
軍	ギヤフ	グイ	軍	軍	-ing	2			
軍	-	-	軍	軍	-ing	1			

図 4.2.3 操作画面16

情	シヤウ	セイ	情	情	qing
晴	シヤウ	セイ	晴	晴	qing
清	シヤウ	セイ	清	清	qing
請	シヤウ	セイ	請	請	qing
青	シヤウ	セイ	青	青	qing
擘	キヤウ	ケイ	擘	擘	qing
擘	キヤウ	ケイ	擘	擘	qing
擘	シヤウ	セイ	擘	擘	qing

図 4.2.4 操作画面17

青 qing 1声 ×

中国語簡体字	ピンイン	レベル	繁体字	Weblio
青	qīng	五	青	Weblio
青春	qīngchūn	五	青春	Weblio
青少年	qīngshàonián	五	青少年	Weblio

BACK

図 4.2.5 操作画面18

### 5.3 ベトナム語漢字

同様に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には耐えうるものになっていそうである。

擘	シヤウ	セイ	擘	擘
清	シヤウ	セイ	清	清
請	シヤウ	セイ	請	請
青	シヤウ	セイ	青	青
擘	キヤウ	ケイ	擘	擘

図 4.3.1 操作画面19

日本語漢字	日本語読み方	越南語読み方	越南語意味
紺青	こんじょう	CÁM THANH	màu xanh nước biển
緑青	りょくしょう	LỤC THANH	lên ten xanh
青地	あおじ	THANH ĐỊA	nền màu xanh
青年	せいねん	THANH NIÊN	tuổi trẻ
青春	せいしゅん	THANH XUÂN	tuổi trẻ
青木	アオキ	THANH MỘC	Aoki
青柳	あおやなぎ	THANH LIỄU	Aoyagi
青梅	あおうめ	THANH MAI	ome
青物	あおもの	THANH VẬT	rau
青田	あおた	THANH ĐIỀN	Qingtian
青畳	あおだたみ	THANH ?	mat mới
青砥	あおと	THANH CHỈ	Aoto
青磁	あおじ	THANH TỬ	Celadon
青空	あおぞら	THANH KHÔNG, KHÔNG, KHÔNG	blue sky
青竹	あおだけ	THANH TRÚC	cây tre
青筋	あおすじ	THANH CÂN	Aosuji
青色	せいしよく	THANH SẮC	màu xanh
青草	あおくさ	THANH THẢO	cỏ xanh
青菜	あおな	THANH THÁI	rau
青葉	あおば	THANH DIỆP, DIẾP	Aoba
青虫	あおむし	THANH TRÙNG	sâu bướm
青蛙	あおがえる	THANH OA	con ếch
青豆	あおまめ	THANH ĐẬU	đậu xanh
青貝	あおがい	THANH BỐI	schrenckii
青酸	せいさん	THANH TOAN	axit hydrocyanic
青銅	せいどう	THANH ĐỒNG	huy chương đồng
青雲	せいうん	THANH VÂN	Ching
青鮫	あおざめ	THANH GIAO	màu xanh cá mập
青鷺	あおさぎ	THANH LỘ	xanh diệp

BACK

図 4.3.2 操作画面20

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”又“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”又“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。

もちろん、上の“2. 漢越語構成”をみると他の分類があるが、やはり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はないかと期待している。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と考える。

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たない場合が考えられる。なぜなら現在ベトナム語は漢字語を使っていないので、漢越語を学習して、そのままの言葉を使うとベトナム人にとって変な意味や分からない言葉となるものも多いからである。

## 5.4 使用したときの感想

### 5.4.1 中国語学習

日本人（京都情報大学生）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 HSK 3 級にも合格した。  
韓国人、ベトナム人が使いやすくするのはこれからの課題。

### 5.4.2 韓国語学習

Ksp 理数学院/Ksp コリア学院 代表者の鍵本聡先生：「大変興味深い。学習に役立つでしょう。」

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、韓国語への敷居がさがった。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のため、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結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さがった。

### 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わかり、学習の補助になる。

## 第6章 おわりに

この度、修士プロジェクト報告書のテーマとして、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、ベトナム人留学生が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思っているので、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事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

まずは技術的な側面から達成できなかった事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ブラウザへの対応
- ・各言語の多音字への完全対応

### 6.2 今後の改良方針

今後の改良方針を以下のようにまとめたい。

#### (a) ユーザーのフィードバックによるフィードバックによる改善

この件は、いろいろな民族の利用を考えてユーザーのフィードバックを行いたい。

#### (b) 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、ベトナム語を学習することは難しい状態である。

あと、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。



(c) 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ブラウザへの対応も当面の課題としたい。

(d) 各言語の多音字への完全対応

これは多：多のデータベースをつくることになるので、たいへん、難しいか  
だいである。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。

現状では以下のように日本語漢字でも多：多の関係が発生しているが、これ  
を音まで行うのはかなり難しい。

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

(e) ベトナム語の音声類似性に関する研究

実際、ベトナム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  
お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、各地にある研究者及び学習者お互い協力しながらシ  
ステムを改善する。

(g) JavaScript CORS問題への対応して発音がどこでも再生できるようにする  
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。

## 謝 辞

本報報告書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きました。Ksp 理数学院／Ksp コリア学院代表の鍵本聡先生には韓国語の教材作成について貴重な助言をいただきました。厚く御礼申し上げます。

## 参考文献

- [1] 大野敏明, “詳説 世界の漢字音”, 慧文社, 2017.
- [2] 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.
- [3] 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.
- [4] 韓虎林 (著, 原著), 今井久美雄 (編著), 梅田博之 (監修), “漢字でおぼえるしっぽをつかむ韓国語”, インターブックス, 1999.
- [5] 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.
- [6] Can. N, “Thủ phân kì lịch sử 12 thế kỉ của tiếng Việt. Ngon Ngu. 6th ed”, pp. 7-12, 1998.
- [7] 府川充男, 小池和夫, 「旧字旧かな入門シリーズ日本人の手習い」, 柏書房2001, pp. 1-251
- [8] 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.
- [9] 川本邦, “詳解ベトナム語辞典= Tu Dien Tuong Giai Viet Nhat. 3<sup>rd</sup> ed.Tōkyō”, 大修館, 2011.
- [10] 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”, 世界の日本語育, 2008.
- [11] Loc V. ( 翻訳と詳解), “An nam dịch ngữ = An Nam ”, 1995.
- [12] Long.P, “Hạng ngôn ngữ Nam Á – Austroasiatic Languages”, Ngonngu.net. <http://ngonngu.net/index.php?p=284> (2019年1月20日閲覧)
- [13] Long. P, “Về việc phân định các giai đoạn lịch sử trong quá trình phát triển của tiếng Việt (phần 1)”, Ngonngu.net <http://ngonngu.net/index.php?p=292> (2019年1月20日閲覧)
- [14] Oli Folkerd, “Tabulator – Interactive JavaScript Tables”, Tabulator. info <http://tabulator.info/> (2019年1月20日閲覧)
- [15] MDN, “AJAX”, MDN web docs <https://developer.mozilla.org/ja/docs/Web/Guide/AJAX> (2019年1月20日閲覧)
- [16] 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](https://developer.mozilla.org/ja/docs/Web/CSS/CSS_Animations/Using_CSS_animations) (2019年1月20日閲覧)
- [17] MDN, “Fullscreen API”, MDN web docs [https://developer.mozilla.org/ja/docs/Web/API/Fullscreen\\_API](https://developer.mozilla.org/ja/docs/Web/API/Fullscreen_API) (2019年1月20日閲覧)

- [18] MDN, “ Mobile First ( モバイルファースト ) ” , MDN web docs  
[https://developer.mozilla.org/ja/docs/Glossary/Mobile\\_First](https://developer.mozilla.org/ja/docs/Glossary/Mobile_First) (2019 年 1 月 20 日閲覧)
- [19] 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 1 月 20 日閲覧)
- [20] StatCounter, “ Desktop Browser Market Share Worldwide ” ,  
StatCounter Global Stats  
<http://gs.statcounter.com/browser-market-share/desktop/worldwide> (2019 年 1 月 20 日閲覧)
- [21] Statistic, “ Mobile share of website visits worldwide 2018 ” , Statistic.com [http://gs.statcounter.com/browser-market-share/desktop/worldwide\\_](http://gs.statcounter.com/browser-market-share/desktop/worldwide_) (2019 年 1 月 20 日閲覧)
- [22] 于劭贇, 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.
- [23] 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 1 月 20 日閲覧)

# 付録1 漢字データベースの使い方

この章では、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を示したい。

(日中韓越漢字語の音韻的類似性を研究・学習するウェブシステム)

## 1 韓国語漢字語彙

まずは、韓国語の漢字語彙の表示を示す。

調べたい漢字の行を選択して、左端の“韓国語・漢字語彙B”をクリックする

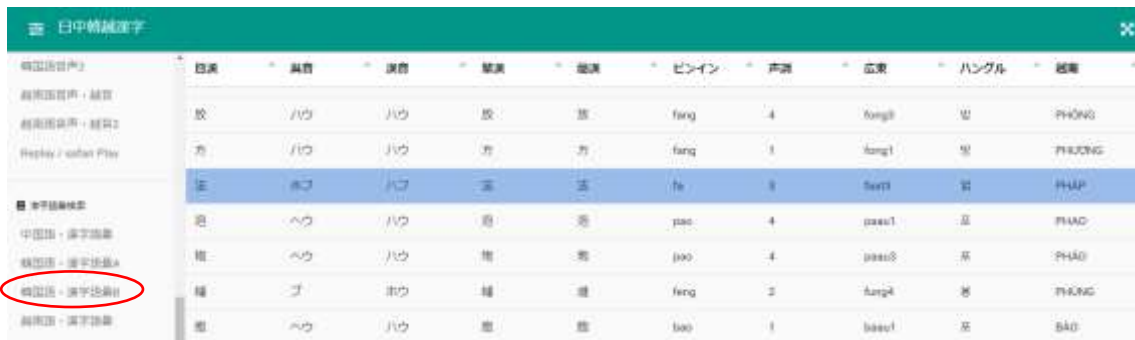


図 4.1.1 操作画面 1

選択した漢字を使った語彙が表示される

(” 의미 : 意味” の列での表示は、日本の漢字語をハングルに直したものが使われていない場合)



図 4.1.2 操作画面 2

カスタム検索で、検索フィールド1 “ハングル” + “=” + “명” で検索 (SEARCH をクリック) [ハングルの入力には、主に IME 韓国語およびソフト

ウェアキーボードを使用，またはウェブの表示・翻訳などからコピー&ペーストを利用]

The screenshot shows a search interface with three rows of search fields. The first row is pre-filled with '한글' in the first field, '=' in the second, and '명' in the third. A '필터수' (Filter count) dropdown is set to '1'. At the bottom right, there are 'CANCEL' and 'SEARCH' buttons.

図 4.1.3 操作画面 3

한글で“명”と書く漢字が表示される。

The screenshot shows a dictionary interface with a table of search results. The table has columns for Japanese, Korean, Chinese, and other linguistic information. The '한글' (Hangeul) column is highlighted with a red box, showing the character '명' for the search term '명'.

日本語	韓国語	漢語	英語	語彙	ピンイン	声調	広義	한글	韓国
名	명	名	名	名	ming	3	ming3	명	MANG
名	명	名	名	名	ming	2	ming2	명	DANG
名	명	名	名	名	ming	4	ming4	명	MANG
名	명	名	名	名	ming	2	ming4	명	MANG
名	명	名	名	名	ming	2	ming4	명	MANG

図 4.1.3 操作画面 4

再度，調べたい漢字の行を選択して，左端の“韓国語・漢字語彙 B”をクリックする。

このように，漢字語彙の表示を繰り返すことによって漢字語彙を増やしていくことが可能である。このシステムを使用する利点としては，発音のつながりに加え，漢字の繋がりからも語彙を学習できる点にあると考える。

## 2 中国語漢字語彙

中国語の漢字語彙の表示を示す。

調べたい漢字の行を選択して、左端の“中国語・漢字語彙”をクリックする

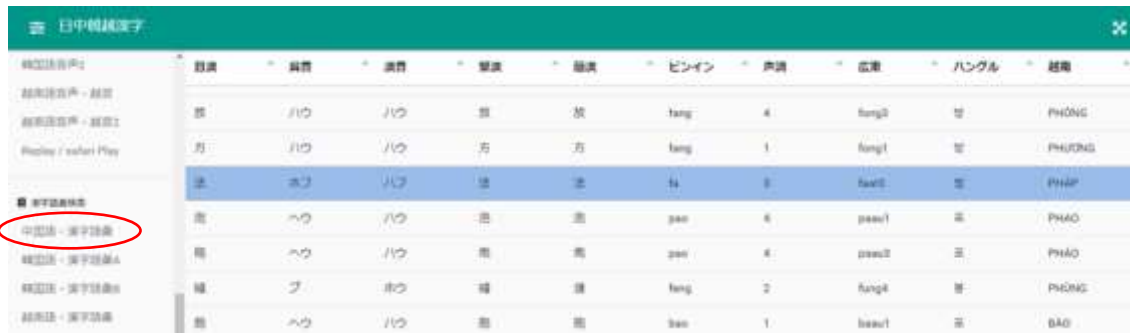


図4.1.5 操作画面 5

選択した漢字を使った語彙が表示される

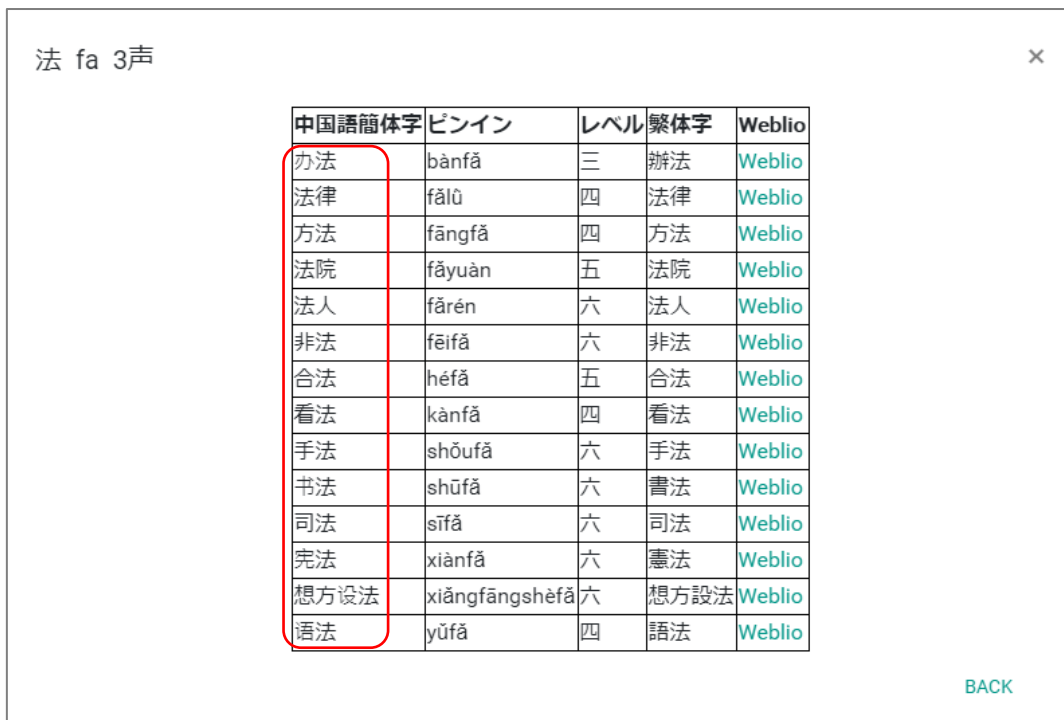


図4.1.6 操作画面 6

再度、調べたい漢字の行を選択して、左端の“中国語・漢字語彙”をクリックする。単語の意味を参照する場合はウェブ辞書を使用する。（“Weblio”をクリック）

カスタム検索で、検索フィールド1 “ピンイン” + “=” + “fa” で検索 (SEARCH をクリック)

図 4.1.7 操作画面 7

ピンインで “fa” の漢字が表示される。

日漢	韓語	漢語	蒙語	藏語	ピンイン	声調	広東	ハングル	越語
瓶	瓶子	バツ	瓶	瓶	fa	1	fa0	병	PHAT
瓶	瓶子	バツ	瓶	瓶	fa	1	fa0	병	PHAT
伐	ボチ	ハツ	伐	伐	fa	2	fa5	병	PHAT
罰	ボチ	ハツ	罰	罰	fa	2	fa5	병	PHAT
法	ボフ	ハフ	法	法	fa	2	fa5	법	PHET
法	ボフ	ハフ	法	法	fa	3	fa0	법	PHAP
法	ボフ	ハフ	法	法	fa	2	fa5	법	PHAP
法	ボチ	ハツ	法	法	fa	2		법	

図 4.1.8 操作画面 8

再度、調べたい漢字の行を選択して、左端の “中国語・漢字語彙” をクリックする。

同様に、漢字語彙の表示を繰り返すことによって漢字語彙を増やす。



### 3 越南語漢字語彙

越南語の漢字語彙の表示を示す。  
調べたい漢字の行を選択して、左端の“越南語・漢字語彙”をクリックする

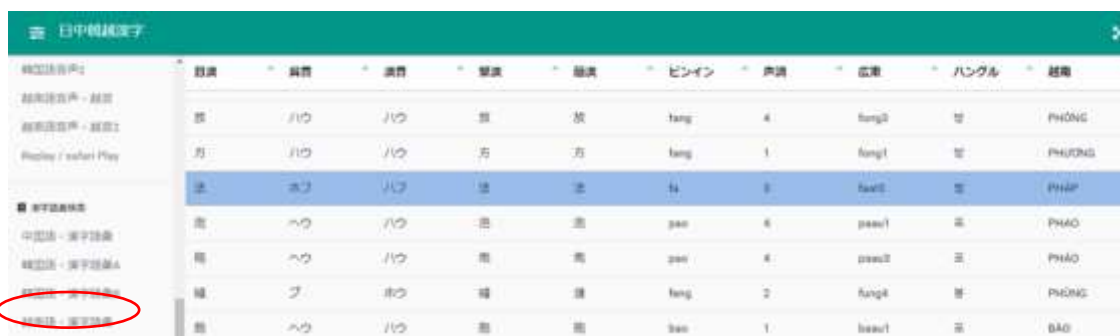


図4.1.9 操作画面 9

選択した漢字を使った語彙が表示される

法 PHÁP

×

日本語漢字	日本語読み方	越南語読み方	越南語意味
不法	ふほう	BẤT, PHẬU, PHỦ, PHI PHÁP	phóng túng
乘法	じょうほう	? PHÁP	nhân giống
仏法	ぶつぼう	? PHÁP	Phật giáo
公法	こうほう	CÔNG PHÁP	Public law
兵法	ひょうほう	BINH PHÁP	Art of war
刑法	けいほう	HÌNH PHÁP	luật hình sự
剣法	けんぼう	? PHÁP	hàng rào

筆法	ひつぼう	BÚT PHÁP	Phong cách của cậu viết chữ
算法	さんぼう	TOÁN PHÁP	toán học
解法	かいほう	GIẢI, GIỚI, GIẢI PHÁP	dung dịch
語法	ごほう	NGỮ, NGỮ PHÁP	cách diễn tả
軍法	ぐんぼう	QUÂN PHÁP	luật quân sự
違法	いほう	VI PHÁP	bất hợp pháp
適法	てきほう	THÍCH, ĐÍCH, QUÁT PHÁP	tính hợp pháp
遵法	じゅんぼう	TUÂN PHÁP	Tuân theo pháp luật
除法	じょほう	TRỪ PHÁP	phân công
魔法	まほう	MA PHÁP	ma thuật

BACK

図 4.1.10 操作画面10

今回、特に一番下の“魔法”の「越南語意味」の“ma thuật”に着目する。

“ma thuật”の“ma”は“魔法”の“魔”に対応すると考えられるが、“thuật”に対応する漢字を調べることにする。

この場合、画面から“thuật”を選択し、コピーする。次に、カスタム検索で、検索フィールド1“越南” + “Like” + “thuật”（検索内容1に“thuật”を貼り付ける）で検索する（SEARCH をクリック）

検索フィールド	演算子	検索内容
検索フィールド1	Like	thuật
検索フィールド2		
検索フィールド3		

図 4.1.11 操作画面11

今回、検索の演算子に“Like”を使うのは、越南後のアルファベット表記には発音記号が使われているため、“=”では検索条件が完全一致になってしまい、期待した結果が表示されないためである。

種類	漢字	漢音	漢字	漢音	ピンイン	声調	広東	ハングル	越南
	術	ズテ	シュツ	術	shu	4	su04	術	THUẬT
	術	ズテ	シュツ	術	shu	4	su04	術	THUẬT

図 4.1.12 操作画面12

検索結果の漢字を見てみると，“術”および“述”が表示され“thuật”の表記は“術”に対応し，“ma thuật”の表記は日本語の“魔術”に対応し，越南語では“魔法”ことを“魔術”と表現していることが推測できる。

このように，越南語は漢字に対応していることが多いので，検索によって漢字を調べて，理解を深めることに役立つと考える。越南語でも再度，調べたい漢字の行を選択して，左端の“越南語・漢字語彙”をクリックする。

同様に彙の表示を繰り返すことによって漢字を増やす。

なお，音声的な類似度をつかいはじめにチェックするような事例集もつくったので掲載する。

### 韓国語

ハングル	発音記号	呉音	漢音	漢字
법 (ボブ)	[pob](-op)	ポプ	fafu	法
칠 (チル)	[chil](-il)	シチ	シツ	七、漆
국 (クク)	[kug](-uk)	コク	コク	菊、局、国
감 (カム)	[gam](-am)	カン	カム	勘、甚、感、憾、敢、甘、監、鑑、減、紺
만 (マン)	[man](-n)	モン	バン	万、晩、蛮、慢、満、漫、灣
경 (キョン)	[kyon](-ong)	キョウ	ケイ	京、競、境、鏡、驚、傾、徑、慶、敬、景、經

### ベトナム語

日本語 ハングル	発音記号	音読み	漢音	ベトナム語														
				漢字														
法/법 (ボブ)	[pob](-op)	ボウ	fafu	PHÁP														
七/칠 (チル)	[chil](-il)	シチ	シツ	THẤT	TÁT													
菊/국 (コク)	[kug](-uk)	コク	コク	CÚC	CỤC	QUỐC												
感/감 (カム)	[gam](-am)	カン	カム	KHÂM	THÂM	CÂM	HÂM	CÂM	CAM	GIÂM	GIÂM	GIÂM	CÂM					
万/만 (マン)	[man](-n)	モン	バン	VẠN	VẢN	MIAN	MAN	MÁN	MAN	LOAN								
京/경 (キョン)	[kyon](-ong)	キョウ	ケイ	KINH	CẢNH	CẢNH	KINH	KINH	KHUYNH	KINH	KHÁNH	KINH	CẢNH					

## 付録2 歴的仮名遣い

以下において、現代＝現代かなづかい・字音＝字音仮名遣（歴史的仮名遣いあるいは旧仮名遣い）とする。

表1. 歴史的仮名遣い表

現代	字音	現代	字音	現代	字音	現代	字音	現代	字音
イ	い	カ	か	ゴウ	がう	ジョク	じょく	ヒュウ	ひう
	ゐ		くわ		がふ		ぢょく	ビュウ	びう
イキ	いき	ガ	が		ぐわう	ジン	じん	ヒュウ	ひょう
	ゐき		ぐわ		ごふ		ぢん	ビュウ	ひやう
イク	いく	カイ	かい	ジ	じ	ズ	ず	ヒョウ	へう
	ゐく		くわい		ぢ		づ	ビョウ	びやう
イチ	いち	ガイ	がい	ジキ	じき	ズイ	ずい		べう
イツ	いつ		ぐわい		ぢき	ズウ	づう	ホウ	ほう
イン	いん	カク	かく	ジク	じく	ソウ	そう		はう
	ゐん		くわく		ぢく		さう		はふ
エ	え	ガク	がく	ジツ	じつ		さふ		ほふ
	ゑ	カチ	かち		ぢつ	ゾウ	ぞう	ボウ	ぼう
エイ	えい	カツ	かつ	ジャ	じゃ		ざう		ばう
	ゑい		くわつ	ジャク	じゃく		ざふ		ぼふ
エキ	えき	ガツ	がつ		ぢやく	チュウ	ちゅう	ミョウ	みやう
エチ	ゑち		ぐわつ	ジュ	じゅ		ちう		めう
エツ	えつ	カン	かん	シュウ	しゅう	チョウ	ちょう	モウ	もう
	ゑつ		くわん		しう		ちやう		まう
エン	えん	ガン	がん		しふ		てう	ユウ	ゆう
	ゑん		ぐわん	ジュウ	じゅう		てふ		いう
		キュウ	きゅう		じう	トウ	とう		いふ
			きう		じふ		たう	ヨウ	よう
オ	お		きふ		ぢゅう		たふ		えう
	を	ギユウ	ぎゅう	ジュク	じゅく	ドウ	どう		えふ
オウ	おう	キョウ	きょう	ジュツ	じゅつ		だう		やう
	あう		きやう		ぢゅつ	ニュウ	にゅう	リュウ	りゅう
	あふ		けう	ジュン	じゅん		にう		りう
	おふ		けふ	ジョ	じょ		にふ		りふ
	わう	ギョウ	ぎょう		ぢよ	ニョウ	にょう	リョウ	りょう
	をう		ぎやう	ショウ	しょう		ねう		りやう
オク	おく		げう		しゃう	ノウ	のう		れう
	をく		げふ		せう		なう		れふ
オチ	おち	コウ	こう		せふ		なふ	ロウ	ろう
オツ	おつ		かう	ジョウ	じょう				らう
	をつ		かふ		じゃう				らふ
オン	おん		くわう		ぜう				
	をん		こふ		ぢやう				
					でう				
					でふ				

[出典：旧かな入門（シリーズ日本人の手習い）]， 柏書房2001， pp. 1-251]

## 付録3 ベトナム語の歴史の概要

ベトナム語の漢字語を理解するために、ベトナム語の歴史をまとめておく。

### 1, ベトナム語形成歴史段階

Vuong loc (1995)によると、ベトナム語形成歴史がこの下通りに分けられる：

- A) プロト (Proto) ベトナム語 10世紀以前
- B) プレ ベトナム語 (Pré Viet-Muong) 10世紀以前
- C) 10世紀： 漢越語の形成時期
- D) 古代ベトナム語15世紀
- E) 中世ベトナム語17世紀
- F) 現代ベトナム語 19世紀

表 ベトナム語の形成時期

A)	プロト (Proto) ベトナム語 10世紀以前	ベトナム語	紀元前から8世紀まで
B)	プレ ベトナム語 (Pré Viet-Muong) 10世紀以前	2つの言語：漢語とベトナム語 文章：漢字で書く	8世紀9世紀ぐらい
C)	漢越語の形成時期 (Viet-Muong commun)	2つの言語：漢語とベトナム語 1つ文章：漢字で書く。 漢字から改善し、一部のベトナム人がハンノム字を生み出したが、文章にはまだ出てこなかった。	10世紀から12世紀まで
D)	古代ベトナム語 (Viet ancien)	2つの言語：漢語とベトナム語 2つ文章：漢字で書くまたはハンノム字 (Han-Nom)	13世紀から15世紀まで
E)	中世ベトナム語 (Viet moye)	2つの言語：漢語とベトナム語 3つの文章：漢字、ハンノム字 (Han-Nom) , 国語 (現代使われている国語 - ロマー字)	17世紀から19世紀間
F)	近代ベトナム語	2つの言語：フランス語とベトナム語 文言の漢語 4つの文章：フランス語, 漢字, ハンノム字 (Han-Nom)、国語	フランスの植民地時期
G)	現代ベトナム語	1つ言語；ベトナム語 1つ文章：国語	1945年以降

Nguyễn Tài Cẩn(1998)によると、様々な異なる言語や文章などが互いに交わり、影響しているので、ベトナム語の形成時期が下図の通りに分かれています：

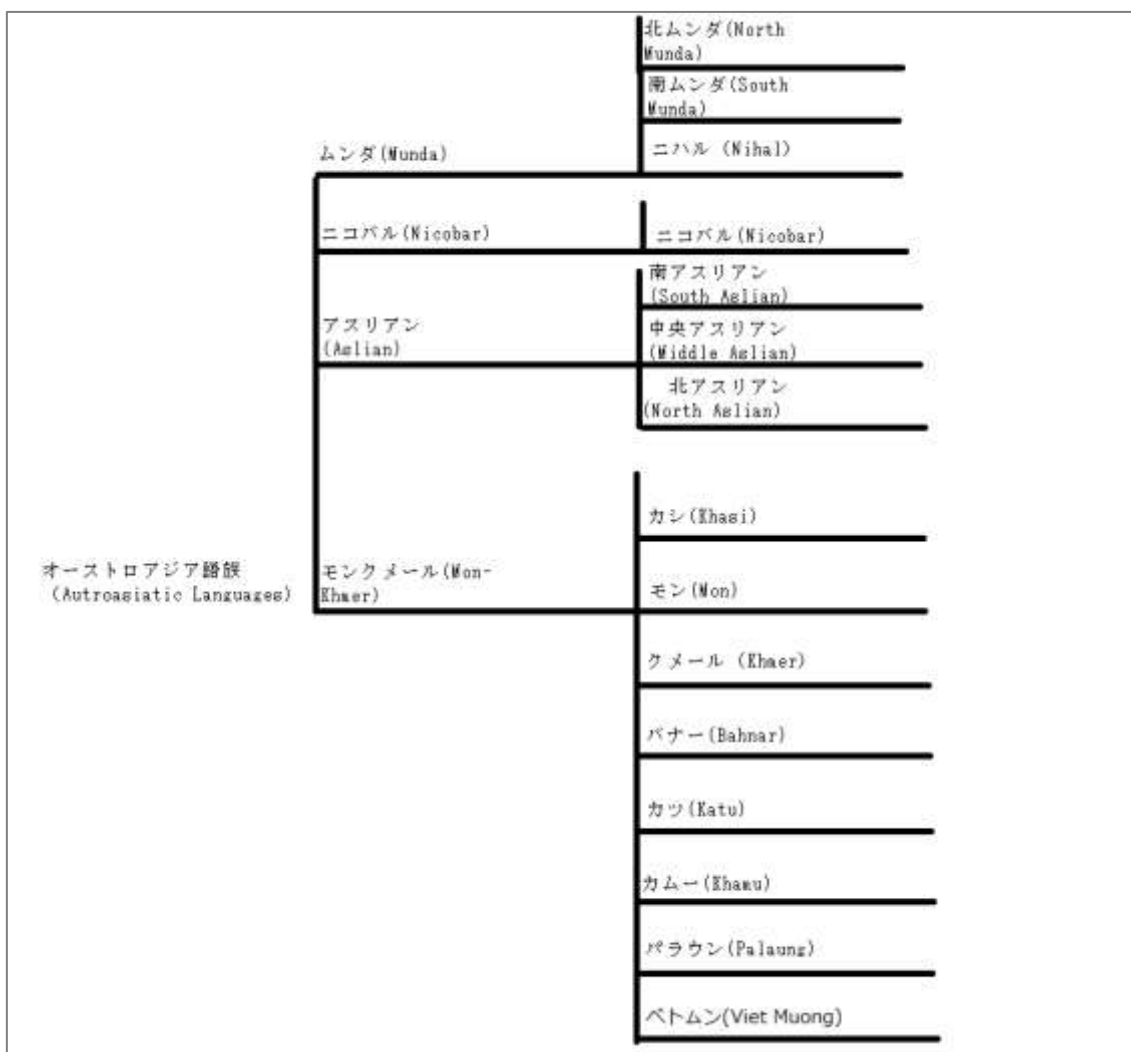


図 Austroasiatic 語族

## 1. プロト (Proto) ベトナムオン語 10 世紀以前

ベトナム語起源 Austroasiatic 語族に属する (Austroasiatic Languages) の時期は現代のベトナム語とベトナムオンに属する現代の各言語と一緒に共通語に 1 つのグループになっており、東モンクメール言語グループから分離して、他の言語が属するモンクメールグループ言語と別の歴史を持っていたようになると考えられている。

ベトナム語の歴史における発展の第一段階と見なすことができる。すなわち、それは最初のベトナム族の共通言語であり、プロト (Proto) ベトナムオンというものである。言い換えれば、プロト (Proto) ベトナムオン語時代は南アジア言語族の一つとしてベトナム語の歴史が始まる。

## 2. プレベトナム語 (Pré Viet-Muong) 10世紀以前

この時代は漢語と近接する影響でプレベトナム語が多様化して、成長していく。文章は漢字で書いていた。

## 3. 漢越語の形成時期—ベトナムコンムン語 (Viet-Muong commun)

この時代にはベトナムが独立国になっていて、ベトナムコンムン語は標準になっていったが、教育や文学などにとって漢語はまだ重要な役割をしていた。文章は漢字で書いていた。

それから一部愛国心が強いベトナム人が漢字から改善し、ハンノム字という新しい字を創造していた。残念ながら、文章にはハンノム字がまだ出てこなかった。

## 4. 古代ベトナム語 (Viet ancien)

この時代はベトナム語が分化していて、ベトナム語 (Viet 語) とムオン語 (Muong 語) に分かれていた。

元々昔から現代にかけて、ベトナムは54民族たちが一緒に暮らしてきた。各民族たちが様々な言語を使用するという原因であり、ベトナム語が豊かな言語になってきたと言う訳です。その為、ムオン語 (ムオンと言う民族がベトナムに居る) は最も近い言語と言われている。

この時代は始めて、ハンノム字がどんどん漢字の代わりに交代していく。例えば Quốc Âm Thi Tập (国音詩集) という本が書かれている。筆者は Nguyễn Trãi (阮廌) という国民的英雄 - 世界の有名人である。

## 5. 中世ベトナム語 (Viet moye)

この時代からベトナム語とムオン語がしっかり分離していく。漢語以外でヨーロッパ言語に近接するために、1651年 A. de Rhodes の Annam-Lusitan-Latin という辞書が始めて出版された。おかげで国語 (標準語) が出来てきた。

## 6. 近代ベトナム語

この時代からベトナムはフランスの植民地になっていたため、その時期ベトナム全国では漢語の代わりにフランス語とベトナム語を使っていた。文章にはまた4つの文字タイプが書かれていた: 漢字, フランス, 国語 (標準語), ハンノム。

## 7. 現代のベトナム語

1945からベトナム社会主義共和国が成立した。国語 (標準語) として発展していく。