

# 日本語文から漢字を増やし仮名を除くことによる 日本人でも中国人でも読める偽中国語の自動生成システム

温婉言<sup>†1</sup> 橋田朋子<sup>†1</sup>

日本語の文章から仮名を除去し漢字だけ残した「偽中国語」は、一見中国語のように見えて日本人も中国人も読める面白さから日中のソーシャルメディアを中心に話題になっている。本研究では、人が手動で行っていた日本語文からの偽中国語文への変換を、自動化することを目指す。特に可読性を高めるために、日本語形態素分析システム Mecab、かな漢字変換ツール Google CGI API for Japanese Input などを利用して、含まれる漢字を増やした上で日本語文を偽中国語文に変換するシステムを提案する。本稿では提案システムの実装の詳細、生成された偽中国語の評価実験の結果、偽中国語の特性をいかした応用例について述べる。

## 1. はじめに

日本語の文から平仮名とカタカナを全て取り除き、漢字のみ残した文は「偽中国語」と呼ばれる[1]。偽中国語は一見中国語や漢文に似ているが、日本語の文法を持つ文を元としているため、日本人には読める文章となる。一方で、漢字を使う中国語を第一言語とする中国人にとっても、偽中国語は、文章を構成する漢字のそれぞれの意味から文章全体が伝えたいことを何となく推測できる。中国人も日本人も読めるという特徴により、偽中国語は日本のツイッターで一時期流行り、さらに Weibo などの中国のソーシャルメディアでも話題となった[2]。実際に偽中国語を使用したコミュニケーションを取った日本人と中国人もいる[3]。

本研究では、従来人が手動で行っていた日本語の文章の偽中国語への書き換え作業を自動的に行うことを目指す。その際に単純に元の文章から仮名を除去するだけでは情報が減ってしまうため、文章に含まれる平仮名表記の語を漢字表記に書き換えることや、動詞の活用語尾により態や時制を表す漢字を動詞の前後に加えるなど文章内の漢字を増やす工夫を行う。これにより日本語文章に含まれる情報を最大限に保留し、日中どちらの人でも容易に読める偽中国語を出力する。本稿では偽中国語の自動生成手法とその可読性などに関する実験結果を報告すると共に、多様な応用可能性(図1)についても述べる。

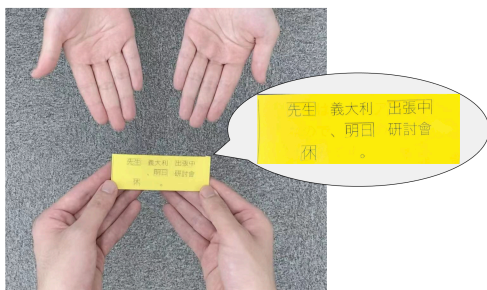


図1 提案システムで生成した偽中国語メモシール  
Figure 1 Pseudo-Chinese memo sticker generated by the proposal system.

## 2. 関連研究

### 2.1 異言語間の共通点から生じる面白い現象について

異言語間での文字や音声に見られる共通点とある種の間違いの許容から生じる面白い現象として、別の言葉に聞き取れてしまう「空耳」が挙げられる。空耳は洋楽の歌詞などを音韻が近い日本語の歌詞に書き換える遊びとしてよく楽しまれる。このような遊びでは、異なる言語間の意思の疎通はあきらめ、元の文と全く関係のない文を作り出す。しかし、外国語を聴く際にまるで日本語を聴いているような錯覚が生み出され、外国語を全く分からない人でも原文の内容を追求せず、その言語を聴くだけで楽しむことができる。空耳を研究として探求したものに、羽鹿らが考案した洋楽の歌詞を和文の空耳フレーズへ書き換える手法[4]や、島谷らが開発した Soramic という邦楽の歌詞を特定の単語群で構成された空耳日本語文への自動生成システム[5]などがある。偽中国語も異言語間での言語の共通点を積極的にいかし小さな間違いを許容する試みという点では空耳と近いが、二つの言語に共通している文字に着目する点や意思の疎通を目指している点が異なる。

### 2.2 日本語と中国語の類似している部分に着目した研究

日本語と中国語が漢字を扱うという点で類似していることを積極的に活かす研究もいくつか行われている。魯らが提案した文の部分翻訳の度合いを調整できる言語学習者向けの補助システム[6]では、スライダーでの操作により中国語へ翻訳する日本語文の割合が動的に調整され、中国語の学習者が原文と訳文の対応を比較しながら中国語への理解を深めることができる。この研究は、ある種日本語と中国語の中間言語を生み出していると言えるが、あくまでも言語の学習者が使用することを想定している。本研究は言語の学習者だけではなく、どちらかの言語にしか習得していない人でも気軽に読み取れる偽中国語に着目したシステムを実現する点が異なる。UD-Kundoku[7]は、漢文(古典中国語)に送り仮名や返り点などを自動的付加する漢文自

<sup>†1</sup> 早稲田大学  
Waseda University

動訓読システムである。このシステムは入力された白文が伝える意味の完全性及び正確さを重視し、文に含まれる情報をできるだけ保留、補足する。一方、本研究は日本語文から仮名を取り除きある意味文に含まれる情報を減らすことで、日本語としても中国語としても不完全だが中国人も日本人も同時に読める文を生成する。

### 3. システム設計

本システムは最終的に、中国人にも日本人にも読める偽中国語を出力することが目的とする。本文における「読める」を以下と定義する。

- 文章の内容を理解できること
- 文章を流暢に読めること

日本語文を偽中国語文に書き換えるプロセスは、文に含まれる中国人には理解できない仮名を取り除く行為を通して、ある意味で情報を捨てる過程とも言える。この中でも偽中国語の可読性を高めるためには、本システムで出力される偽中国語文には、日本人も中国人も同時に読める漢字を多く含める、仮名を取り除く際に削除する情報を最小限に留める、という二つの要件を満たすことが望ましい。

一般的に人が手動で偽中国語を作成する際には、まず対象となる日本語の文章の中で、漢字表記に書き換える語を漢字表記に変え、最後に文の中に含まれるひらがなやカタカナを全部消去するという一連の手順が行われる。このプロセスを参考にして、提案システムでは大まかに以下の手順で偽中国語の生成を行う。

- ① 入力された日本語文に対して形態素分析を行い、文を語ごとに分割して処理
- ② 形容詞、動詞、名詞など物事を指している語で、ひらがな表記になっている語を漢字表記へ、または漢語にあたる類義語へと変換
- ③ 外来語などカタカナで書かれる語を意味の近い和文、もしくは中国語に書き換え
- ④ 動詞の活用語尾により、過去、願望、可能、使役などの態や時制を表す漢字をその動詞の前後に補足
- ⑤ 文の中に含まれるひらがな、カタカナを全て除去

図2に、ある日本語文を入力として、偽中国語が生成されるまでの手順と実行の予想結果を示す。なお、仮名除去済みの偽中国語文は日本語と同じく語順がSOV型であり、動詞を目的語の前に移動するなどの調整をした方がSVO型の中国語に近づくが、日本人が偽中国語を読む際の流暢さが落としてしまう恐れがあるため、本研究では文の語順に対する調整を考慮しないものとする。

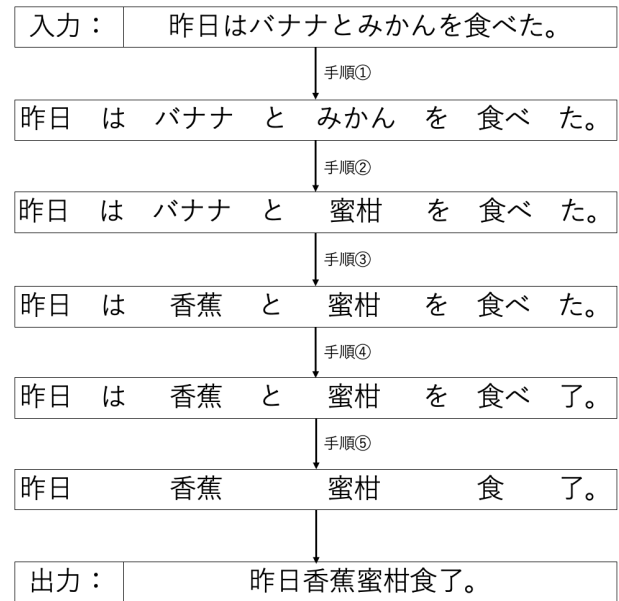


図2 日本語文を偽中国語へ書き換える過程

Figure 2 The process of translating Japanese into Pseudo-Chinese.

### 4. 実装

#### 4.1 システム概要

提案した偽中国語自動生成システムを macOS 11.4/Windows 10 環境に Python3 (Python 3.8.3) 言語で実装した。

本システムは上述したように5つのプロセスを通じて偽中国語への生成を実現する。最初に行う前処理では、日本語形態素解析エンジン MeCab を用い、入力された日本語文を形態素解析して個々の語に分割する (手順①)。次に、文に含まれる漢字を増やすために、分割されたそれぞれの語に対し、名詞、形容詞、動詞、感動詞のどちらかに属する語を品詞ごとに漢字表記へ置き換える処理を行う (手順②)。ひらがな表記である形容詞や名詞などにインターネットサービス Google CGI API for Japanese Input を用い、出現頻度が高い順に出された漢字変換候補から適切なものを選び出し、元の語と置き換える。動詞には、漢字変換候補の選出の際に日本語版 Wikipedia の全テキストデータを学習させた Word2Vec モデルにより語と語の類似度を計算し、前後の語と類似度が一番高い変換候補を選ぶ方法を使い、同音異義語による影響を解消する。一方、外来語に対して、和文へ変換できない外来語は数多くあるため、Microsoft Translator API などの翻訳サービスを使い、対象の外来語だけを中国語に翻訳する方法で変換を行う (手順③)。さらに、動詞の活用語尾に基づき、「希望」、「無」、「了」など態や時制を表す漢字を動詞の後に加える (手順④)。全ての変換を行った後、処理済みの語を元の順番に再び文として組み立て、その文から平仮名、カタカナを全て除去し

得られた文が偽中国語文として出力される（手順⑤）。

## 4.2 システムの動作

本システムを用いて、3つの日本語文例に対して出力した偽中国語の結果を表1に示す。ひらがな表記の名詞及び形容詞の漢字表記への書き換え、外来語の中国語への翻訳、動詞の後ろへの態や時制を表す漢字を付加するなど、予想した通りの動作や出力結果が得られることを確認した。

また、一般のユーザが利用しやすいように、Python用のウェブアプリケーションフレームワーク Flask を用いて、提案システムとの間に通信を行い、入力された日本語文から偽中国語の生成や表示ができる Web ページを HTML/CSS で制作した。Web ページの動作を図3に示す。

表1 提案システムを用いて生成した偽中国語例

Table 1 Sample of Pseudo-Chinese generated by using the proposed translating system.

入力文	出力文
私ほりんごとチョコを食べたい。	私林檎巧克力食希望。
お忙しい中お返事頂きありがとうございます。	忙中返事頂謝。
昨日誰と会ったかを全然覚えていない。	昨日誰会了全然覚無。

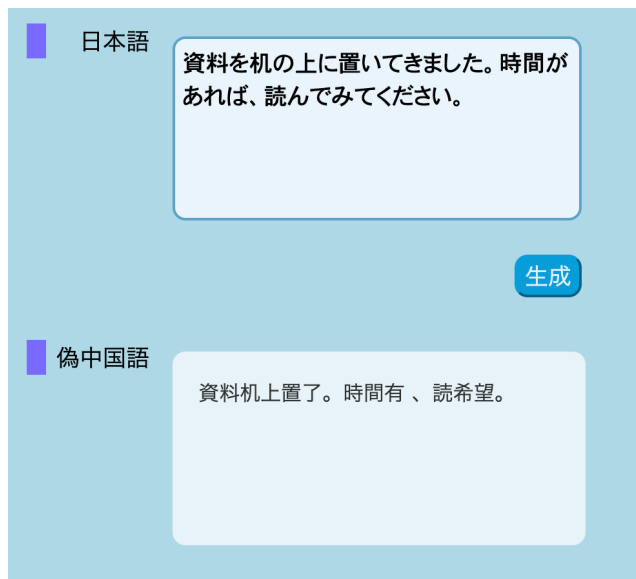


図3 偽中国語を生成できる Web ページ

Figure 3 The Web page which could create Pseudo-Chinese.

## 5. 実験・調査

### 5.1 概要

本システムを用いて生成された偽中国語文の可読性を確認するため、オンラインでアンケート形式の調査を行った。まず実験参加者のグループは4水準を設けた。日本語母

語話者（以下、日話と略す）、日本語上級者である中国語母語話者（中話-上級）、日本語初中級者である中国語母語話者（中話-初中級）、日本語学習経歴なしの中国語母語話者（中話-無）であり、それぞれ6名ずつ参加してもらった。なお予備実験により日本語学習歴ありの中国語母語話者は、それぞれの日本語能力により回答にばらつきが出る傾向があったため、今回の本実験では日本語学習歴のある中国語話者を上級と初・中級に分けることにした。本文における「日本語上級者」、「日本語初中級者」をそれぞれ「日本語能力試験（JLPT）2級と相当するまたはそれ以上の日本語能力を有する者」、「日本語能力試験（JLPT）3級と相当するまたはそれ以下の日本語能力を有する者」を指す。

実験では、日本語母語話者と中国語母語話者である実験参加者全員に対して、本システムによって生成された3つの偽中国語文を参加者自身の母語に翻訳してもらった後に、「①偽中国語文の内容を理解できる」、「②偽中国語文を流暢に読める」二つの質問項目において、それぞれの項目に「あてはまらない」を1、「あてはまる」を5にした5段階評価を行ってもらった。

また、アンケートに使用した偽中国語文と日本語問題文には「外来語」、「挨拶言葉」、「指示代名詞」、「過去、完了が含まれる表現」をそれぞれ一つ以上含まれるものとした。

### 5.2 結果と考察

5段階評価の結果を図4に示す。実験参加者のグループと質問項目を要因として、二要因分散分析を行った結果、二つの要因の主効果が有意であり、交互作用は有意ではなかった。下位検定の結果、日話、中話-上級、中話-初中級の三つの水準と中話-無との平均値の間にそれぞれ有意な差が確認され、グラフから中話-無のグループの平均値が、他のグループよりも低い。また、グラフから質問①よりも質問②の方が平均値が低い。以上から、まず日本語能力を有する人は、その日本語能力の高さや母語に関係なく偽中国語を理解することが可能であると示唆される。ただし理解ができてみずしも流暢に読めているとは限らない。一方、全く日本語能力を有していない人にとっては、偽中国語を理解することや、流暢に読むことはやや難しいことであると考えられる。

5段階評価による定性的な評価に加えて、参加者が実際に偽中国語を理解した程度を測るために、各参加者が偽中国語を自身の母語に訳した文に対して、「正しく翻訳されている語が全体に占める割合」を基準として、100%正確に翻訳されたことを5とし、0%正確に翻訳されたことを1として、筆者がそれぞれの訳文に対して理解度点数を付けた。グループごとの平均点数を図5に示す。上記の評価実験の結果と同じく、日本語能力を有する参加者の偽中国語への理解度が高いことが確認された。一方で、日本語能力

を有していない者でも偽中国語への理解度の平均点が3を超えていた。ここから偽中国語の内容を少なくとも半分程度は理解していると考えられる。つまり日本語能力を有していない者でも、各自が思っているよりも内容を理解できていた可能性が示唆される。

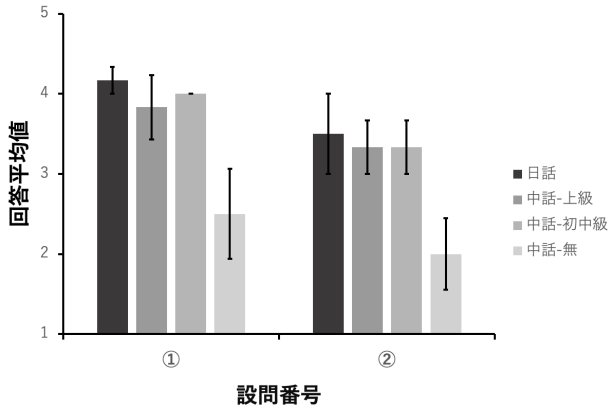


図4 五段階評価の結果

Figure 4 Result of five-grade evaluation.

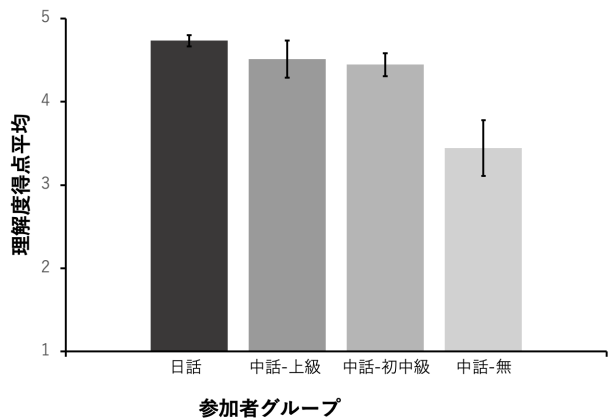


図5 参加者グループごとの偽中国語理解度の平均点数

Figure 5 The average score of the participants' comprehension of Pseudo-Chinese.

## 6. 応用例

### 6.1 概要

本システムはまずウェブアプリケーションという形で偽中国語への自動変換の仕組みを実現した。一方で提案システムを日常の場面で実際に使ってもらうためには、日本語を偽中国語へ変換する過程を手軽に楽しめたり、日常的なコミュニケーションに利用可能な形であることが望ましい。そこで、提案システムを応用した仕組みとして、剥がすことによって日本語文と偽中国語文の表示を切り替える上下2枚組のメモシールが出力されるアプリケーションを実装した。HTML, Javascript で制作される Web ページに日本語文を入力すると、Flask を通じて本システムと通信し、

Python Image Library (PIL) を使用し、2枚組のシールの画像が表示、さらにラベルライター (Brother 製) で印刷することができる (図9)。この2枚のシールは上下に組み合わせ合わせて利用するが、2種類を作り分けることができる。まず1種類目は、上部のシールには変換後の偽中国語文が印刷され、入力された日本語文と共通な漢字の部分にだけ四角が印字されている。今回の実装ではこれをユーザが自分でカッターやクラフトパンチなどで穴あけする。下部のシールには偽中国語文から取り除かれた漢字が穴から見えるように日本語入力文が印刷される。この方法で作成したシールを2枚重ねることで、図6(右)に示すように上から偽中国語に見える仕様となり、上部のシールを剥がしたら、上部のシールで覆われた日本語文が見えるようになる。また、2種類目として図7に示すように、上部に日本語文を印刷し、下部に偽中国語を印刷することで、はじめは日本語が見えているがシールを剥がすとその偽中国語が見えるような形を実現することも可能となる。

実際にラベルライター (Brother PT-P910BT) でシールが印刷される様子とシールが使用される様子を図8, 図9に示す。本アプリケーションにより、図1のような日本人が中国人にメモを残すなどの日中間の交流過程に、直感的に提案システムを使用することができると思われる。

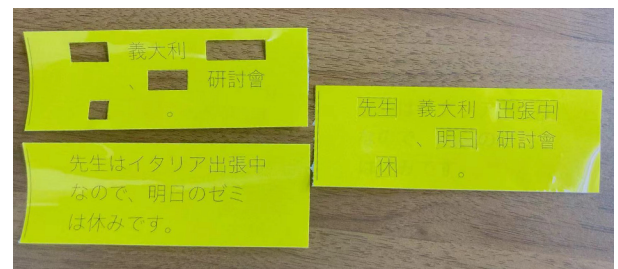


図6 上部に偽中国語を表示するシール

Figure 6 Sticker showing Pseudo-Chinese.

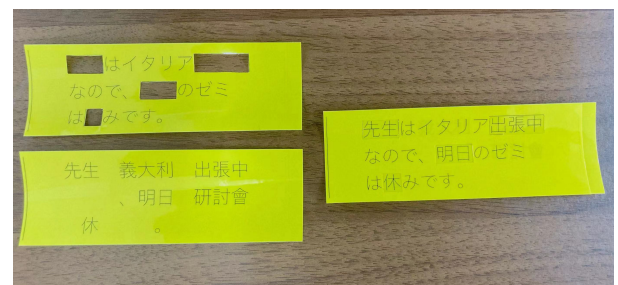


図7 上部に日本語を表示するシール

Figure 7 Sticker showing Japanese.



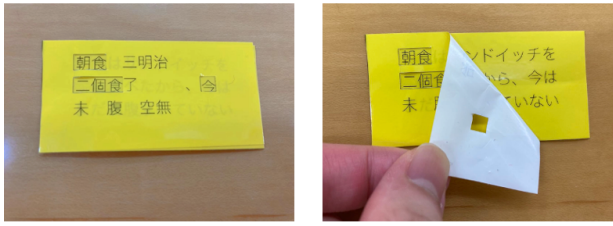


図 8 シールが使用される様子

Figure 8 The photo showing how a sticker is used.



図 9 ラベルライターでシールが印刷される様子

Figure 9 The photo showing how a sticker is printed by a label printer.

## 6.2 ユーザスタディ

5章に記載した予備実験と本実験のどちらかに1回参加した者の中で、日本語上級者である中国語母語話者3名と、日本語母語話者3名合計6名に対して、実際に本アプリケーションを使用してもらった。参加者は自由に入力した日本語文が偽中国語に変換され、自分が選んだ二種類のいずれかのタイプの偽中国語のシールとしてラベルライターで印刷されるという一連の動作をそれぞれ3回体験した。その後に参加者全員に対して、「①どのような使い方を想定して今回のシールを作成しましたか?」、「②このシールは今後どのような使い方ができると思いますか?」二つの項目を質問し、このアプリケーションに対する感想についても聞いた。

質問項目①に関しては、中国語ネイティブとコミュニケーションと取りたかった(3名)、日常のことを友人に話したかった(1名)、自分が考えたことが偽中国語で何で書くかを知りたかった(1名)、ネット上の友達と話したかった(1名)など、多様な場面や相手についての回答が得られた。

質問項目②に関しては、中国語を勉強する時に使えるかもしれない(3名)、中国語話者の方と会話する時に使える(2名)、SNSなどにリラックスした雰囲気話したい時に使える(1名)など本アプリケーションの幅広い可能性や使用可能な場面についての回答を得ることができた。

また、本アプリケーションに対して、「(シールを)剥が

して詳細を確認できる機能のは面白いと感じた」、「一瞬で分かるようになったのは面白い」、「文章を即時に偽中国語に出してくれるのは便利だった」などの感想が寄せられた。

## 7. まとめ

本研究では日本語文を対象として、ひらがな表記の語を漢字表記にするなど漢字を増やすなどの手法により、日本語文章に含まれる情報を最大限に保留し、日中どちらの人でも容易に読める偽中国語を自動生成するシステムを構築した。Webページ上で偽中国語の自動生成を楽しめるアプリケーションに加えて、日常な場面を想定した応用としての偽中国語と日本語の表示の切り替えができるシールを提案、実装した。さらなる応用可能性として、SNS上に投稿する際に日本語文章を自動的に偽中国語へ変換できる拡張機能の実現や、言語学習者に完全もしくは部分的に偽中国語変換を果たした文章を提示することで日本語から中国語への理解を深めるアプリケーションなどが挙げられる。

偽中国語変換システムを用い、日中どちらの方でも気軽に読める偽中国語を出力できたが、偽中国語への主観的な理解度が実際の理解度より低い、日本語を勉強したことがない人にとって偽中国語が読みづらい、日本人が外来語より変換された偽中国語への理解に支障が出るなどの課題がある。今後の展望として、類義語変換などの手法を導入し、より自然かつ可読性が高い偽中国語文を生成することや、さらに、今回は日本語のみからの偽中国語文の作成を考慮していたが、中国語から偽中国語へ変換するシステムの制作も試みたい。

## 参考文献

- 1) Record China, 日本でにわかにブームの「偽中国語」、中国で驚きの声, 2017-03-30, <https://www.recordchina.co.jp/b173638-s0-c30-d0046/> (参照 2022-07-26)
- 2) エキサイトニュース, 河野外相、再びツイッターの「偽中国語」で中国ネットを沸かせる, 2019-08-26, [https://www.excite.co.jp/news/article/Searchina\\_20190826016/](https://www.excite.co.jp/news/article/Searchina_20190826016/) (参照 2021-07-26)
- 3) Record China, 日本の大手漫画専門書店が偽中国語連投、中国人も偽中国語で応答「君偽中国語素晴」「運営苦労様追跡留意」, 2019-07-17, <https://www.recordchina.co.jp/b729668-s0-c30-d0054/> (参照 2022-07-26)
- 4) 羽鹿諒, 山西良典, Jeremy S. White: 洋楽歌詞からの空耳フレーズ自動生成に向けた一検討, エンタテインメントコンピューティングシンポジウム 2016 論文集, 2016, pp.114-119(2016).
- 5) 島谷二郎, 中村泰: Soramimic - 限定された単語による空耳日本語文自動生成システムの開発, 研究報告ヒューマンコンピュータインタラクション(HCI), 2018-HCI-176(5), pp.1-6(2018).
- 6) 魯薇, 寺田努, 塚本昌彦: 翻訳度合いを調節するインタフェースを備えた中国語学習支援システムの設計と実装, マルチメディア, 分散協調とモバイルシンポジウム 2021 論文集, 2021(1), pp.943-948(2021).
- 7) 安岡孝一: 漢文の依存文法解析にもとづく自動訓読システム, 日本漢字学会第3回研究大会予稿集, pp.60-73(2020).