

熟語に注目した漢字学習支援システムの提案

Kanji learning support system focusing on compound kanji words

仲宗根 英人, 小渡 悟

Hideto NAKASONE, Satoru ODO

沖縄国際大学 産業情報学部

Department of Industry and Information Science, of Okinawa International University

Email: 17DB091@okiu.ac.jp

あらまし：漢字学習に関する手法や学習サービスにおいて、収録漢字の登録や関連問題の作成を開発者側が手作業で行うものが多い。本研究では、著作権が消滅した作品や著者が許諾した作品のテキストデータに形態素解析を施し、語の出現頻度などをもとに学習サービスへの収録漢字と関連問題を収納するデータベースの自動構築を行った。また、漢字の頻出度や繋がりを可視化できるシステムを構築した。

キーワード：漢字, 漢字学習, 漢字学習支援, 二字熟語, 形態素解析, 小説, 自動構築, 漢字ゲーム

1. はじめに

漢字学習の手法として反復練習がある。これは、漢字の形と意味を何度も読み書きすることによって漢字を学習していく手法である。この手法は現在、数多くの学習ソフトで採用されている⁽¹⁾。

反復練習以外にも、漢字1文字が印字されたカードを使って熟語を作るゲームの授業内実証報告や、意味の理解に重きを置いた漢字学習ソフトの提案などがある^{(2),(3)}。特に後者の研究では、漢字学習における漢字の意味の理解と、他の単語との関係性に注目した学習法が効果的であることが報告されている。しかし、これらの研究が提案しているシステムでは、収録漢字の登録や関連問題の作成を開発者の手作業で行っていた、そのため、作業時間や労力の負担が課題とされていた。また、学ぶ漢字の優先順位について考えた時、日常生活において使用頻度の低い漢字を覚えるよりも、高い頻度で使用されている漢字を優先的に学習した方が効率的である。文化庁は常用漢字表を公開しているが、どの漢字が日常生活においてどれ程使用されているといった頻出度に関する情報は記されていない⁽⁴⁾。そのため、これらの情報を元に学習の優先順位を考えることは困難である。

そこで本研究では、漢字を学習する上で、その漢字の意味のみを覚えるのではなく、他の漢字との繋がりをも含めて理解させることを目的とした漢字学習支援システムを提案する。提案システムでは小説を解析元としたデータベース構築の自動化、ならびに漢字の頻出度を視覚的に提示する機能を実現する。

本研究では漢字同士の繋がりを、熟語を分解することで示せると考えた。例として「勇氣」をあげる。これは「勇」と「気」の漢字が結合した熟語である。ユーザーに熟語を構成する漢字(「勇」、「気」と構成漢字以外と繋がった熟語(「勇敢」、「元気」、「病気」など)を提示することで、熟語「勇氣」のみだけでなく、漢字から派生した熟語も学べる環境を提供する。

ただし、本研究における熟語とは、2文字の漢字によって構成された形式を指す。

2. 提案システム

2.1 データベースの構築

システムの実現には、日常生活における漢字の頻出度と、使われ方を分析する必要があった。データの質と信頼性を担保する為にも、多くの漢字が使用された文章を選択した。本研究では、青空文庫の著作権が消滅した作品(48作品, 約4,254,700文字)を対象に MeCab を用いて形態素解析を実施した。

形態素解析では、文章を単語別に分解し、品詞分類や正規表現を用いて熟語を特定した。最後に、除外すべき単語の選別処理を経てデータベースへ格納させた(詳細は4章で議論する)。登録情報は「漢字の文字情報」、「ID」、「英語訳」、「頻出回数」、「Google画像検索結果の画像URL」の5項目とした。

英語翻訳と画像検索には GCP の API を用いた。最終的にデータベースには、1文字漢字を1,626文字、熟語を2,405文字格納しているが、このAPI利用料金の関係で今回は頻出度の高い単語、熟語のみを格納の対象とした為、解析元のデータ量に対して、格納数が減少している(詳細は4章で議論する)。

2.2 採用した技術

システムを使用するユーザーの漢字へ対する興味関心を妨げない為には、ユーザービリティは極めて重要な要素である。今回は Ruby(Sinatra), Vue.js, GCP, PostgreSQL, Firebase の技術を採用した。SPAとして開発したことで、十分なユーザービリティを担保することが出来た。

3. 動作検証

3.1 登録されている1文字漢字の一覧表示 (Words)

登録漢字は Words ページで閲覧できる(図1)。表示されている情報は、漢字の文字情報、英語訳、Google画像検索結果の画像、星の画像である。具体的には頻出度の高い漢字では星の画像を5個表示、順次頻出度に応じて4個、3個と表示している。画面右上の入力フォームにて、絞り込み検索が可能である。



図1 動作画面の例 (Words ページ)

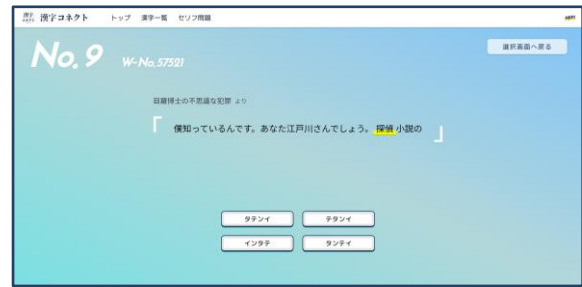


図4 動作画面の例 (Media Question ページ)

3.2 漢字の詳細ページ (MIX MAP)

Words ページで漢字を選ぶと、MIX MAP ページへ遷移する (図 2)。中心に選んだ漢字が表示され、周囲にはその漢字と熟語を形成できる漢字が最大 10 個まで表示される。また、使用頻出度に応じて中心からの距離と繋ぐ線の色が変わる。「人」を例にあげると、「主」や「間」との組み合わせが、他と比べて頻出度が高いことが視覚的に強調されている。この時、例えば「夫」という漢字をクリックすると、「人」と「夫」によって作られる熟語の詳細を画面左側にて確認することが可能である。さらに、組み合わせ対象として選んだ「夫」という漢字をクリックすることで、今度は「夫」の MIX MAP ページへ遷移するが可能である。

3.3 セリフ問題 (Media Works)

作品毎に簡単なクイズ問題を用意した。この問題も自動的に生成している。仕組みとしては、まず正規表現や形態素解析によって文章を分析し、セリフ部分のみを抽出した。次にその中で使用されている熟語の文字と読み方をデータとして取得することで実現した。この作品一覧画面で好きな作品を選ぶことにより、各々の問題選択画面へ遷移する (図 3)。



図2 動作画面の例 (MIX MAP ページ)



図3 動作画面の例 (Media Works)

3.4 問題表示画面 (Media Question)

ここでは、作中のセリフを表示させ、その中で使用されている漢字をランダムに選出し、その読み方をユーザーへ回答させるクイズ形式となっている (図 4)。問題の自動生成によって多くの問題数を確保できた。好きな作品のセリフを通じ、漢字の様々な使われ方、表現を確認することで、ユーザーの漢字に対する興味関心が高まることを期待する。

4. まとめ

形態素解析を用いてデータベースを自動構築させたが、人名や地名といった固有名詞の除外が不完全である。様々な条件を設定し、フィルターを追加したが、完全に単語を区別するには未だ至っていない。現状、除外候補の単語らを別ファイルへ分類し、設計者の最終確認にてデータの質を担保している。ただ、解析した作品数が増加するにつれ、特定の作品にのみ出現する漢字や熟語の特定が容易になる点や、除外する漢字データが蓄積されていくことによって自動化が図れると考えている。

また、API の利用料金が負担となる点も考慮しなければならない。実際、形態素解析の終了時には、1 文字漢字が 3,695 文字、熟語が 18,810 文字の分析が終了していたが、この問題によって先で述べた様に 1 文字漢字が 1,626 文字、熟語が 2,405 文字と極端に格納文字数が減少してしまっている。

今後は実際にユーザーに利用してもらい、漢字に対する印象や学習への意識の変化を取得し、評価する必要がある。

参考文献

- (1) 間辺 美樹, 並木 美太郎, 兼宗 進 et al.: “意味の理解に着目させる漢字学習ソフト「熟語マニア」の開発と評価”, 情報処理学会論文誌, 教育とコンピュータ, Vol.4, No.1, pp.16-30 (2018)
- (2) 丹羽 正之: “漢字カードを使った熟語教育”, 神戸夙川学院大学・夙川学院短期大学 教育実践研究紀要, 2015 巻, 2013-2014 号, pp.32-36 (2015)
- (3) 文化庁: “常用漢字表”, https://www.bunka.go.jp/kokugo_nihongo/sisaku/joho/joho/kakuki/14/tosin02/index.html (参照 2020.12.16)