Web 上の読解教材作成を支援するツールの開発及び活用法

北村 達也[1]　荊木 亜里沙[1]　森川 結花[2]　永須 実香[3]
川村 よし子[4]　前田 ジョイス[5]　斉木 美紀[4]　金 善子[6]
[1]甲南大学知能情報学部
[2]甲南大学国際言語文化センター
[3]上智大学国際教養学部
[4]東京国際大学言語コミュニケーション学部
[5]東京国際大学経済学部
[6]東京国際大学学習支援室

要旨

文章中の単語に訳語がリンクされた Web 教材を作成するためのツールを開発した．このツールは 2 つのプログラムから成る．第 1 のプログラムは，利用者が入力した文章を形態素解析し，各形態素の読みや訳語の情報とともに Microsoft Excel のファイルとして保存する．第 2 のプログラムは，第 1 のプログラムで得られたファイルの情報に基づいて入力文章中の単語に振り仮名と訳語を付与し，Web ブラウザに表示させる．第 1 のプログラムで得られるファイルの内容を利用者が変更することによって，形態素解析や訳語の誤りを修正することができる．作成した教材は利用者自身の Web ページやブログに掲載して公開したり，学習者にメールで送付したりすることができる．

【キーワード】形態素解析，読解教材，Web 教材，インターネット，Microsoft Excel

1. はじめに

近年，Web ブラウザ上の単語にマウスポインタを合わせたりクリックしたりするとその意味や訳語が表示されるツール（例：Google ツールバーのマウスオーバー辞書や POP 辞書.com）が利用できるようになっている．このようなツールは多くの人々を辞書引きの手間から解放している．とりわけ，母語以外で書かれた文章を読まねばならない人々は大きな恩恵を受けていると考えられる．このようなツールには，日本語教育の支援を目的としたもの（例：リーディング・チュウ太の辞書ツール（川村他，2000））もあり，教材作成等に活用されている．しかし，これらのツールには，処理結果に誤りがあっても利用者がそれを修正するのは困難もしくは不可能という問題がある．これは教育目的での利用にとっては致命的なことである．そこで，本研究では形態素解析や訳語の誤りを利用者が容易に修正できるツール e-chuta（editable-chuta）を開発した．

ここで，リーディング・チュウ太の辞書ツール（川村他，2000）を例にとり，単語へ訳

語を付与した Web ページを自動生成する仕組みを説明する．利用者により入力された文章は，まず形態素解析システムにより形態素に分割される．形態素解析では，読み，品詞，活用している単語の終止形等も得られる．次に，各単語を辞書データベースで検索する．そして，Web ページを記述するための言語（マークアップ言語）である HTML（Hyper-Text Markup Language）を用いて単語と訳語をリンクするよう指定したファイル（HTML ファイル）を自動生成する．この HTML ファイルを Web ブラウザで表示させると一連の処理は終了である．利用者が Web ブラウザ上の単語をクリックすれば訳語が表示される．

上述の処理ではいくつかの段階で誤りや不都合が生じ得る．まず，形態素解析に誤りが少なからず生じる．また，文脈との関係で訳語が正しくなかったり，利用者が教えている学習者のレベルと訳語の難易度が合っていなかったりする．さらに，単語単位で訳語を検索するので，複数の語から成る慣用句のような表現に対応することができない．処理結果の HTML ファイルを利用者自身の PC にダウンロードして編集すればこれらの問題は修正することができるが，それには HTML に関する知識が求められる．

そこで，本研究では，上述の処理において辞書検索が終わった中間段階をファイルに出力し，利用者が形態素解析や訳語の誤りを修正できるようにする．その際，中間段階を多くの PC 利用者に馴染みのある Microsoft Excel（以下 Excel と略記）の表の形式で保存することによって，修正作業を容易にする．この作業において HTML に関する知識は不要である．そして，修正後のファイルに基づいて HTML ファイルが自動生成される．これによって誤りのない読解教材が作成できる．このシステムはインターネット上にて誰でも無料で利用できる．

2. e-chuta

e-chuta は，入力文章を形態素解析して表形式のファイルに変換するプログラムとそのファイルを HTML 形式に変換するプログラムの 2 つから成る．これらのプログラムは Perl というプログラミング言語で開発されている．

2－1　入力文章の表形式への変換

文章は e-chuta の Web ページ（http://basil.is.konan-u.ac.jp/e-chuta/）上のテキスト入力エリア（図 1）に入力される．入力された文章は，形態素解析システム MeCab（工藤，2006）により単語に分割される．MeCab 用の辞書は IPA 辞書を用いている．形態素解析により，読み，品詞，単語の終止形が得られる．次に，各単語を辞書データベースで検索して訳語を得る．語形変化している状態では辞書データベースを検索することができないため，終止形で辞書データベースを検索する．

本システムで用いている辞書は，川村を中心に開発が進められている多言語辞書である（川村・金庭，2006）．この辞書には日本語能力試験の旧出題基準の 1 級から 4 級までの単

語が含まれている．現時点では利用者からの要望が多い英語と中国語の辞書に対応している．利用者は文章入力時にどちらの辞書を用いるか選択する．

最後に，形態素解析及び辞書データベース検索の結果を Excel の表の形式で保存する（図 2）．この表では，1 列目に入力文章の単語，2 列目にその終止形，3 列目に読み，4 列目に訳語（この図では英訳）が記載されている．なお，1 列目の単語が平仮名だけから成る場合には読みは記載しない．また，MeCab により助詞と判定された単語には訳語をつけない．

将来的にこのデータ形式は容易に拡張できる．例えば，5 列目に例文，画像ファイル名などを記載する仕様にすれば，単語をクリックしたときにそれらの情報を表示させるような Web 教材を作成することも可能になる[(1)]．

なお，Excel のファイル形式での保存は，Perl の Spreadsheet::WriteExcel モジュールを利用することにより実現されている．

図 1：e-chuta の Web ページ（http://basil.is.konan-u.ac.jp/e-chuta/）

2—2　表形式から HTML への変換

上述の処理が終了すると，Web ブラウザは図 3 のような表示に変わる．利用者は"Download xls-file" をクリックして生成された Excel のファイルを自分の PC に保存する．

利用者は，この Excel ファイルを開き，必要に応じて修正を加える．複数の行にまたがっている単語（例えば「口」，「が」，「軽い」）を 1 つの行にまとめて 1 つの慣用表現（この例では「口が軽い」）とし，それに訳語をつけることもできる．

修正が済んだら，このファイルを表形式から HTML 形式に変換するシステムにアップロードする．このシステムは，1 列目の単語から入力文章を再構築するとともに，3 列目のデータが存在する単語には振り仮名をふる．さらに，4 列目のデータが存在する単語には訳語をリンクする．ファイル中の全ての行に関してこの処理を行い，HTML ファイルを生成し，それを Web ブラウザに表示させる．

これら一連の処理は自動的に行われるため，利用者は誰にも気兼ねせず何度でも Excel ファイルを修正し，納得がいくまで教材を作り込むことができる．

なお，Excel のファイル形式での読み込みは，Perl の Spreadsheet::ParseExcel モジュールを利用することにより実現されている．

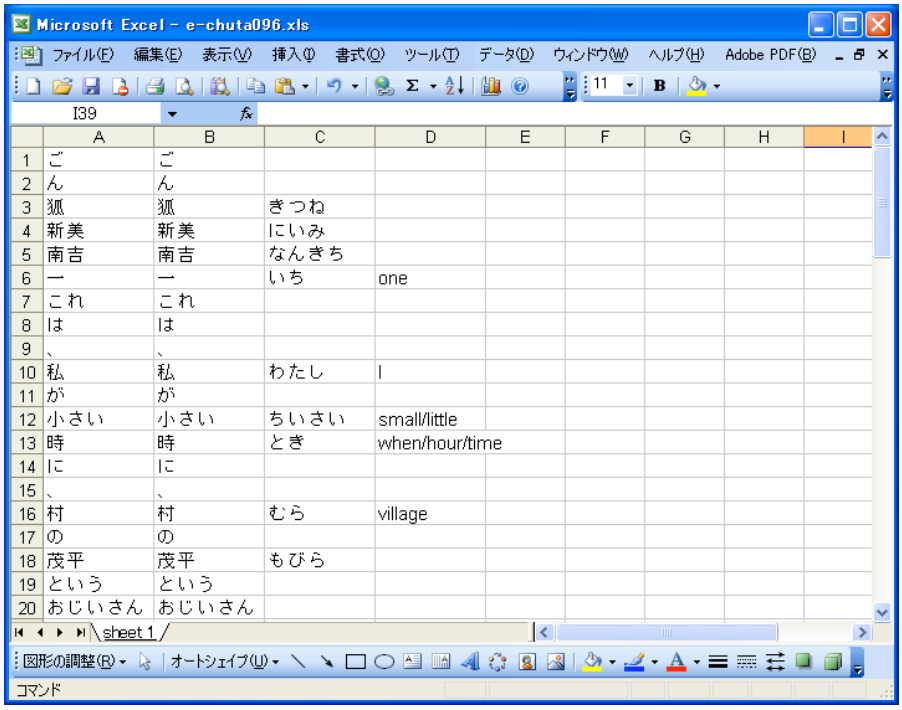

図 2：自動生成された Excel の表形式の例．「ごん狐」が 3 行に分けられているのは形態素解析の誤りによるものである．また，「茂平」の読み（3 列目）も誤っている．このような誤りを利用者が修正することによって，教材として利用できる Web ページを作成する．

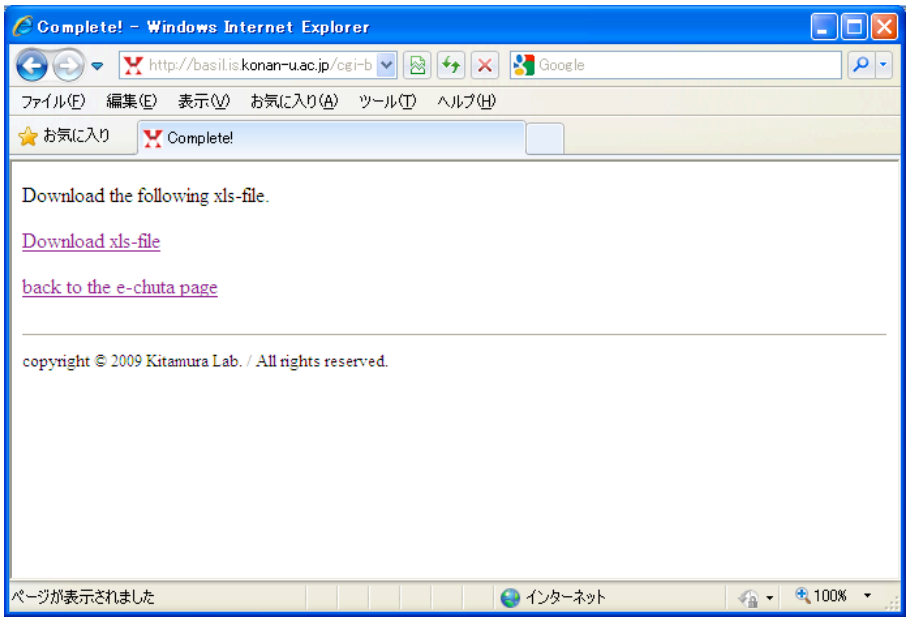

図 3：表形式への変換が終了した画面

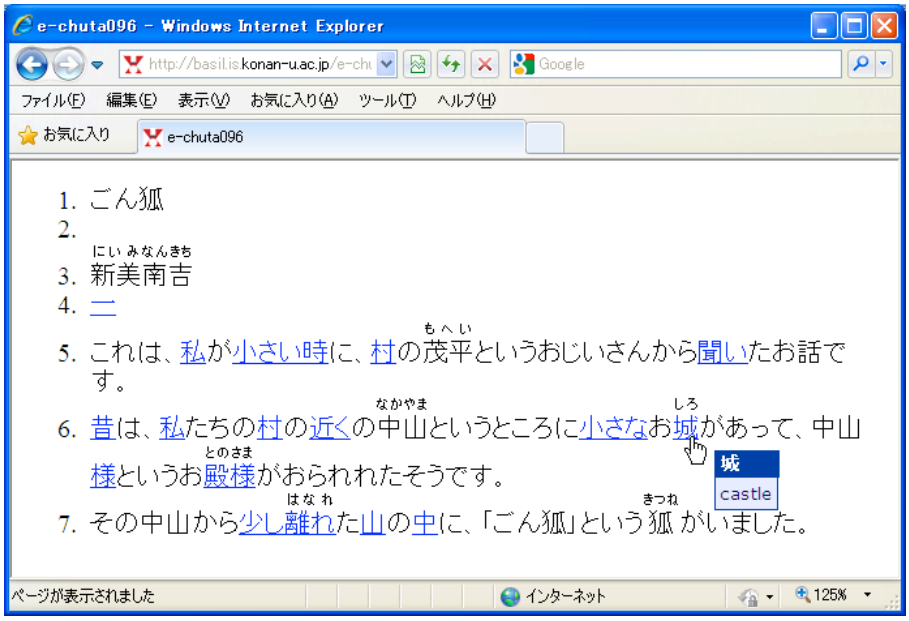

図 4：e-chuta を用いて作成した教材 Web ページの例（振り仮名は InternetExplorer でのみ表示可能）．「城」にマウスポインタを合わせることによってその英訳がツールチップに表示される．

2—3　処理結果

上述の 2 つのシステムにより処理された結果を図 4 に示す．本システムでは，文章全体の見通しがきかない初中級学習者に配慮して 1 文ごとに行番号を付与して表示するようになっている．新たな行番号をふる基準として以下の規則を採用している．

1. ″．″ の後は必ず新しい文番号をふる．このとき，″．″ の後に空行がなくてもよい．
2. ただし，″.」″ のように ″．″の後に ″」″ が続く場合には，″．″ の後ではなく ″」″ の後に新しい文番号をふる．
3. Excel の空行の後は必ず新しい文番号をふる．つまり，″．″ 以外で改行したいところに空行を入れればよい．

また，訳語へのリンクは，その単語にマウスポインタを合わせると小さなウィンドウ（ツールチップ）が表示される方式を採用している（図 4）．

3. 活用法

3—1　HTML ファイルの活用

本システムにより生成された HTML ファイルは，Web ブラウザの「ページを保存」等の機能により利用者の PC に保存することができる．保存したファイルは，利用者自身の Web ページに掲載したり，電子メールに添付して送付したりして学習者に見せることが可能である．さらに，HTML ファイルの内容をコピーすることによってブログに教材を載せることもできる(2)．

また，ホームページビルダー（IBM 社）等の Web ページ作成ソフトウェアを利用すれば，生成された HTML ファイルに図を挿入したり，レイアウトを変更したりする作業をワープロ感覚で行うことができる．

3—2　教材サイトの構築

e-chuta を用いて「日本語上級者のための日本文学珠玉の小品集」(http://basil.is.konan-u.ac.jp/tutor/bunko/) という教材サイトを開発した（森川他，2010）．この教材サイトでは，青空文庫（http://www.aozora.gr.jp/）から入手した日本文学作品に関して，英訳や挿絵を付与した本文，朗読音声，読解問題等を提供するとともに，ブログを併設して学習者との交流を測っている．

この教材サイトでは，e-chuta の機能を利用して作品中の会話の話者を学習者に知らせる工夫を加えている．すなわち，Excel の表において，会話の最初の ″「″ の行の 4 列目に話者の名前を記入している．これによって，図 5 に示すように ″「″ にマウスポインタを合わせたときに話者の名前がツールチップに現れるようなっている．

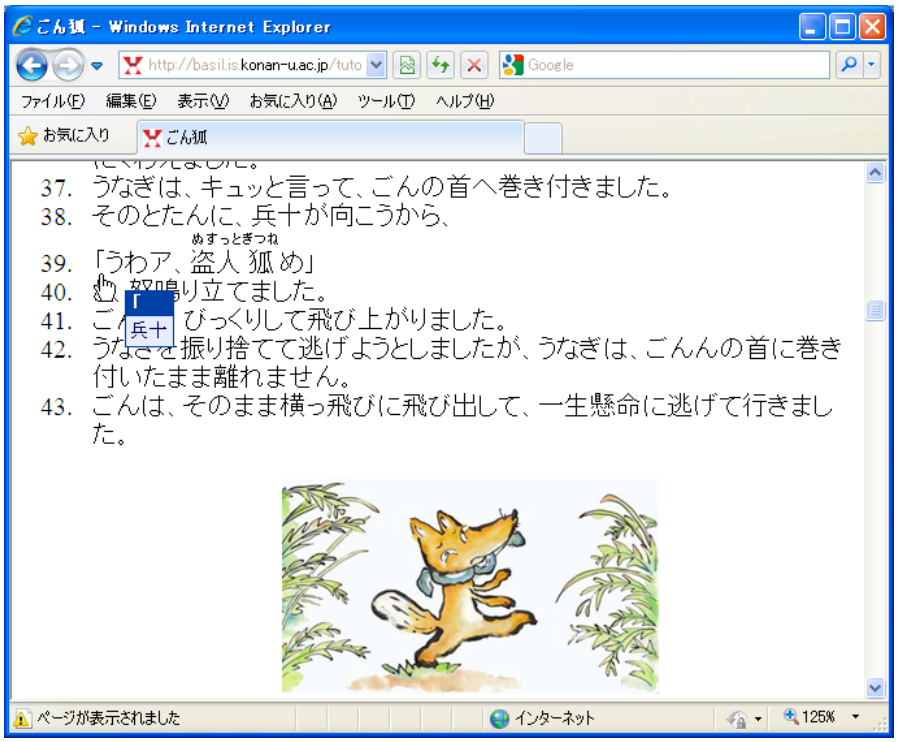

図 5：ツールチップで話者を表示した例

この他，どの単語に英訳がリンクされているのかを学習者に気付かせないために，リンクしてある単語の色を黒に設定している．これは，HTML ファイルに修正を加えることによって実現している．

e-chuta は，この教材サイトのように文学作品を教材化する場合に便利なツールである．文学作品では凝った語彙や表記が用いられるので，機械的に単語単位で訳語を示しても学習者の読解にとっての効果は薄い．e-chuta では利用者独自の修正を加えることができるので，凝った語彙や表記に対しても適切な訳語や説明を示すことができる．また，教材の作業過程で Excel の表を作成するため，そこから容易に単語リストを作成できる．この教材サイトでも，Excel の表から抽出した初中級学習者向けの単語リストを公開している．

4. おわりに

本研究では，利用者が読解教材を作り込むことができるツール e-chuta について報告した．このツールは，入力文章中から読解教材を自動生成するにあたってその中間段階を Excel のファイルとして出力する．この Excel のファイルを変更することによって，利用者が希望する読解教材を作成できる．

なお，本稿ではツールチップで英訳が表示される事例のみを示した．しかし，利用者が Excel の表形式の 4 列目にタガログ語や韓国語の訳語を記入すれば，その訳語がツールチッ

プで表示される．つまり，本ツールは様々な言語を母語とする学習者のための読解教材作成にも利用できる．また，日本語による解説を記入すれば，それがツールチップで表示される．このことを利用すれば，本ツールは日本語教育に限らず，例えば古文や漢文の教育等にも利用することができる．

本システムと同様に利用者が形態素解析や訳語の誤りを修正することができるシステムに Nagaya (2010) による読解アシスタントがある．このシステムは Web ブラウザ上で修正作業ができるという高度な機能を有している．それに対して，本システムでは中間段階を一旦自分の PC にダウンロードして処理する必要があるものの，多言語に対応している点，Excel の表形式を採用したため多くの人が操作に慣れている点，どの部分をどのように修正したのかが把握しやすい点などにメリットがあると考えている．今後さらに多言語辞書への対応を進めるとともに，多くの利用者の要望を取り入れていく予定である．

謝辞

本研究の一部は甲南大学総合研究所，平成 21 年度シーズ発掘試験(11-143)，平成 22 年度科学研究費補助金 基盤研究(B)(21320095)の支援を得て行われた．

注

(1) 2010 年 11 月現在はそのような仕様にはなっていない．

(2) 2010 年 11 月現在，Blogger でのみ動作確認済みである．この具体的方法については本システムの Web ページに解説動画を掲載している．

参考文献

(1) 川村よし子，北村達也，保原麗 (2000)「EDR 電子化辞書を活用した日本語教育用辞書ツールの開発」『日本語教育工学雑誌』Vol. 24 (Suppl.), 7-12

(2) 川村よし子，金庭久美子 (2006)「国際共同編集による日本語学習者のための多言語版 web 辞書の開発」『日本語教育学会春季大会予稿集』 61-66

(3) 工藤拓「MeCab: Yet Another Part-of-Speech and Morphological Analyzer」<http://mecab.sourceforge.net/>

(4) 森川結花，永須実香，春名宣明，北村達也 (2010)「日本語読解学習支援サイト"tutor.bunko"の構想と開発：総合的な技能養成を目指した方向性とそのコンテンツ」『甲南大学情報教育研究センター』Vol. 9

(5) Nagaya, Y. 「読解アシスタント」<http://dokkai.mit.edu/index.cgi>