読解教材の作り込みを支援するツール e-chuta の開発

Development of a tool for assisting to prepare reading materials “e-chuta”

北村達也
KITAMURA Tatsuya
甲南大学
Konan University
〒658-8501 神戸市東灘区岡本 8-9-1
E-mail: t-kitamu@konan-u.ac.jp

Abstract : We developed a support tool for preparing Japanese reading materials. This tool consists of two programs. The first program analyzes input text morphologically, and saves results in a table-form file (XLS format). The table lists morphemes in the text, their original form, pronunciation, and English translation. The second program converts the XLS file to a reading material in HTML format. By using these two programs, users who are not skilled in HTML can modify the reading material by editing the XLS file.

キーワード：読解教材開発，インターネット，表形式，Microsoft Excel，HTML

１． はじめに

リーディング・チュウ太の辞書ツール(川村，2000)は，入力文章に自動的に辞書情報をリンクして web ブラウザ上に表示するシステムである．この辞書ツールを使うことによって様々な文章を読解教材にすることができる．その反面，利用者が web ブラウザに出力された結果を編集するには，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)に関する知識が必要であり，容易とはいえない．辞書ツールの出力結果を簡単に編集できるようになれば，学習者のレベルに応じた読解教材作成や，独自の情報を追加した読解教材作成が可能になる．そこで，本研究では，誰でも利用しやすい表形式のファイル上で辞書情報等を編集できるツール e-chuta (editable-chuta)を試作した．このシステムはリーディング・チュウ太の各ツールと同様にインターネット上で自由に利用できる．そして，このシステムを利用することによって，教師は読解教材を作り込むことができる．

２． e-chuta システム

e-chuta システムは，入力文章を形態素解析して表形式に変換する部分と表形式のデータを HTML 形式に変換する部分から成る．

2.１ 表形式への変換

前者では，まず形態素解析システム MeCab (工藤，2006)を用いて入力文章を形態素に分解する．このとき読みの情報も同時に得られる．次に，川村ら(2006)が開発した辞書から各形態素の英訳を得る．この辞書には日本語能力試験出題基準の 1 級から 4 級までの単語が含まれている．最後にこれらの情報を図 1 に示す表形式(Microsoft Excel 形式)のファイルに出力する．

この表形式では，1 列目に入力文章の形態素，2 列目にその辞書形，3 列目に読み，4 列目に英訳が記載されている．なお，このデータ形式は現時点での仕様であり，利用者からの要望に応じて今後修正や拡張を行う予定である．

◇	A	B	C	D	E	F
1	日本語	日本語	にほんご	Japanese		
2	を	を	を	(particle)		
3	勉強	勉強	べんきょう	study		
4	し	する	し	to do		
5	て	て	て	post-fix		
6	いる	いる	いる	to exist/to be		
7	皆さん	皆さん	みなさん	everybody/all		
8	なら	だ	なら	copula		
9	誰	誰	だれ	who		
10	でも	でも	でも	but		

図 1：e-chuta により得られる表形式データ

2.2 表形式から HTML 形式への変換

利用者は上記の表形式ファイルをダウンロードし，必要に応じて変更を加える．例えば，読みや英訳を表示する必要のない形態素についてはそれらを削除したり，英訳をより適切なものに変更したりできる．

次に，このファイルを表形式から HTML 形式に変換するシステムで処理する．このシステムは，1 列目の形態素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形態素には振り仮名をふる(逆に，3 列目のデータが存在しない形態素には振り仮名をふらない)．さらに，4 列目のデータが存在する形態素には英訳

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．この HTML ファイルをダウンロードすることによって，利用者自身の web ページにて読解教材として公開することができる(ただし，著作権に留意が必要)．

これら一連の処理は自動的に行われるため，誰にも気兼ねせず何度でも教材を作り込むことができる．

3． e-chuta の使用例

e-chuta を利用して「日本語上級者のための日本文学珠玉の小品集(略称：tutor.bunko)」(森川ら，2010)を作成し，http://basil.is.konan-u.ac.jp/tutor/bunko/)にて公開した(図 2)．この web ページでは，青空文庫(http://www.aozora.gr.jp/)から入手した日本文学 6 作品に関して，辞書情報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図っている．さらに，朗読音声を Podcast として iTunes Store にて無料公開するなど，複合的な学習環境の構築を目指している．

教材にする文学作品として，新美南吉著「ごん狐」(1932)，芥川龍之介著「南京の基督」(1920)，有島武郎著「一房の葡萄」(1920)，太宰治著「桜桃」(1948)，宮沢賢治著「セロ弾きのゴーシュ」(1934)，寺田寅彦著「夏目漱石先生の追憶」(1932)を選択した．

e-chuta を用いた教材作成においては，日本語教師 2 名が上記文学作品のテキストデータから生成された表データに読み，英訳，会話文の話者に関する情報を付与した(森川ら，2010)．表データの編集が教材にどのように反映されるのかを容易に確認できるので，教材を作り込むことができた．

図 2：「日本語上級者のための日本文学珠玉の小品集」トップページ

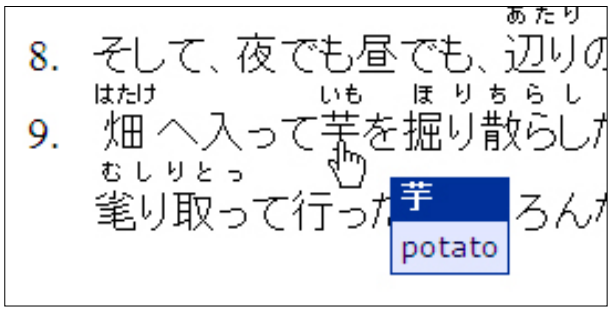

図 3：ツールチップによる英訳情報の表示(ルビ表示は InternetExplorer のみ可能)

なお，tutor.bunko においては，英訳情報はツールチップにより表示するようにしている(Zorn，2009)．利用者が英訳を知りたい単語をマウスポインタで指すだけで(クリック不要)，図 3 に示すような英訳が表示される．

4． おわりに

読解教材の作り込みを支援するシステム e-chuta を試作し，それを用いて日本語読解学習支援サイトを作成した．今後，利用者の意見を取り入れシステムの仕様を確定した上で本格的に公開する予定である．本システムが日本語読解教材の蓄積及び相互利用，また教材作成の省力化につながれば幸いである．

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 21 年度科研費基盤研究(B)(21320095)の支援を得て行われた．システム開発にご協力いただいた甲南大学知能情報学部 荊木亜里沙氏，辞書データをご提供いただいた東京国際大学言語コミュニケーション学部 川村よし子先生，システムにコメントをいただいた甲南大学国際言語文化センター 森川結花先生，上智大学国際教養学部 永須実香先生に感謝します．

参考文献

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」『日本教育工学雑誌』24(Suppl.)，7-12.

工藤拓(2006)「MeCab: Yet Another Part-of-Speech and Morphological Analyzer」http://mecab.sourceforge.net/.

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』，61-66

W. Zorn(2009)，「DHTML JavaScript Tooltips」，http://www.walterzorn.com/tooltip/tooltip_e.htm

森川結花，永須実香，春名宣明，北村達也(投稿中)「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』