

多言語版日本語辞書編集システムの開発と運用実験

川村 よし子 (東京国際大学) kawamura@tiu.ac.jp

共同研究者：植木正裕 金庭久美子 川村ヒサオ 前田ジョイス 根津誠 保原麗

Hans Coppens Jonathan Bunt Kristina.S.Hmeljak Saleh Adel

はじめに

本研究の目的は、世界各国の日本語教育者・研究者との共同作業によって、日本語学習者が必要としている各言語版の日本語辞書を作り上げるシステムの構築にある。本研究の対象である辞書編集システムは、辞書情報を XML 化して管理することによって、多言語情報を一元的に管理することを目指している。システムはインターネット上に公開し、インターネットを介して各教育者が個別に各言語版の辞書の編集作業を行う仕組みを持つ。世界各国の日本語教育者による運用実験と評価を通して、各々の言語環境における問題点の洗い出し作業を行う。その結果明らかになった問題点に対して適宜対応することによって実際に多言語環境で編集可能なシステムに仕上げ、インターネット上で自由に編集作業が行える多言語版辞書編集システムを完成させる。本システムによって作成される各言語版日本語辞書は、Web 辞書として公開するとともに、すでにインターネット上で公開している日本語読解学習支援システム「リーディング・チュウ太(Reading Tutor)」の辞書ツールに組み入れる予定である。

1. 先行研究

コンピュータを用いた読解支援の草分け的存在としてはパデュー大学と筑波大学の共同研究による「CATERS」(<http://tell.fll.purdue.edu/JapanProj/CATERS/>)、国立教育研究所の「CASTEL/J」、国際基督教大学の鈴木らによる読解システム等があるが、いずれも教材が限定された、いわば閉じられたシステムである。これに対して寺ら(1996)による読解支援システム「DL」は、入力された任意の文章の辞書引きが自動で行えるという点で教育工学的にもすぐれたシステムであった。この DL に学習履歴管理機能を持たせる(北村ほか 1999)とともに、辞書情報として日本電子化辞書研究所の EDR を用いた辞書ツール(川村ほか 2000)を組み入れたのが、本研究の基盤となる日本語読解学習支援システム「リーディング・チュウ太」(<http://language.tiu.ac.jp>)である。このシステムは、辞書ツールに加えて、文中の単語や漢字のレベルを判定する語彙チェッカー、漢字チェッカー等のツールを備え、インターネット上の情報を学習者や教師がその場で教材化できる開かれたシステムである。その意味でインターネットを活用した読解支援として先駆的な役割を果たしてきた。リーディング・チュウ太は他の言語教育分野にも影響を及ぼし、ロンドン大学の D.J.Lefevre による Reading Assistant は英語版の読解学習支援システムである。また、東京国際大学の Maeda らの FLC (Maeda et al.2000)はリーディング・チュウ太の語彙チェ

ッカーの英語版である。日本語教育の分野でも、東京工業大学の仁科ら(2002)により構文解析機能のある学習支援システム「あすなる」が開発・公開され、また、文中の漢字に振り仮名をつける「ひらがなナビ」、辞書情報・漢字情報を表示する「理解 com」等、様々な読解支援が提供されるようになった。だが、その一方で、学習者の母語で書かれた対訳辞書がほしいという切実な要望に対しては十分に応えきれていないのが現状であり、リーディング・チュウ太の辞書ツール多言語化への要望が強い。本研究による多言語版日本語辞書編集システムは辞書ツール多言語化の早期実現とともに、多言語環境での共同編集・教材の共有化を可能にする新システムである。

2. 日日辞書編集システムの概要

2003年にスタートした「辞書ツール多言語化プロジェクト」では、世界各国の編集者がインターネット上で日日辞書及び対訳辞書の編集を自由に行うことのできるシステムの開発を行ってきた(川村ほか2003)。その結果、日日辞書編集システムは2004年に完成し、現在日日辞書の編集作業中である。この日日辞書編集システムにおける主な辞書情報項目とその編集上の特徴は次の通りである。(図1)

- ・ **見出し語**：現在日本語能力試験の出題基準にある約 8000 語を見出し語として取り上げている。今後、日本語学習者にとって必要不可欠な語を補い必要度の高いものから順次編集作業を進める。語の難易度・複合語・連語に関する情報も補う。

- ・ **品詞情報**：将来リーディング・チュウ太の辞書ツールに組み入れることを目的としているため、品詞情報は「茶釜」に準拠する。複数の品詞があれば列挙し、品詞ごとに意味情報を記述する形をとる。

- ・ **意味情報**：日本語学習者のための辞書を作ることを目的としているため、意味情報は概念ブロックごとに中心語義から並べるとともに、意味説明もできるだけ平易な日本語

(可能であれば日本語能力試験 2 級レベル)で行う。初中

級者向けの情報と上級向けの情報を区別する仕組みも整える。意味情報を一語で表すことのできる類義語がある場合はこれも入力する。

- ・ **例文**：特に派生的な意味の場合には不可欠という基準で、例文情報を入力する。意味情報同様、できるだけ平易な日本語で書かれた例文を作成する。



図1 日日辞書の情報項目

3. 多言語版日本語辞書編集システムの開発

多言語版日本語辞書編集システムは、上記の日日辞書をもとに世界各国の編集者が各言語版の対訳辞書を編集するシステムである。各言語版の編集作業は概念説明の翻訳および訳語の登録、例文・用法の翻訳である。言語ごとに辞書作成責任者・編集者・閲覧者等を管理し、各言語版の辞書情報を SQL サーバーに保存する仕組みになっている。

編集にはあらかじめ編集者としての登録が必要である。言語ごとに責任者を決め、各言語の責任者が各々の編集者を登録する。

図2がこのシステムのメニュー画面である。編集者はまず画面下部の Search Word によって編集したい単語を検索し、自分の単語リスト (My Word List) に登録する。単語の検索は見出し語、編集者の担当言語、品詞、日本語能力試験の級、編集状態による多重検索が可能である。

図3が多言語版辞書編集システムの編集画面である。編集者は画面上のボックスに対訳情報を入力する。入力項目は、Sense (概念説明)、Same Word (概念説明に該当する代表的な訳語)、Translation (概念説明に該当する Same Word 以外の訳語)、Example (例文)、Note (用法等の説明) である。

入力された辞書情報は自動的に辞書編集者の担当言語の対訳情報として SQL サーバーに保存され、他の言語の辞書情報とリンクされる。

4. 多言語版日本語辞書編集システム 版による運用実験

2005年8月に完成した多言語版日本語辞書編集システム 版をもとに運用実験を開始した。第1回目の運用実験は、英語、オランダ語、スロヴェニア語版日本語辞書編集者によって2005年8月から9月にかけて行われた。

実験の結果明らかになった問題点のうち、メニュー画面および入力画面の表示方法等に



図2 多言語版辞書編集システムのメニュー画面



図3 多言語版辞書編集システムの編集画面

関しては、逐次改良を行っている。また日日辞書の概念の記述方法および例文の提示方法にも変更が必要なことが判明した。運用実験によって明らかになった主な問題点とその対応方法は次の通りである。

a. 画面表示方法の改良

問題点 版では日日辞書と各言語の編集画面を左右に並べて表示する形をとっていた。ところが、実際に作業を行うと、辞書編集者のインターネット環境によって編集画面の見え方が異なり、作業効率も悪いことが判明した。

対応方法 図3のように日本語情報と対訳情報を上下並べた形で編集できるように編集画面を修正する。

b. 例文の必要性

問題点 上記のように例文は単語の「派生的な意味の場合には不可欠」という編集方針で日日辞書の編集を進めていた。ところが運用実験の結果、学習者にとってはすべての意味について例文があるほうがわかりやすいことが判明した。

対応方法 可能な限りすべての意味について例文を作成する。編集済みのものに対しても例文作成を行う必要が生じたため、新たに辞書編集者を増員する。

c. 例文の書き方

問題点 例文作成は日本語の母語話者で主に日本語教育にたずさわっている辞書編集者に依頼している。例文はやさしい日本語で日本語として自然な文であることを原則としている。ところが短文で書いた場合、前後の文脈が明らかでないため、翻訳しにくい文があることが判明した。

対応方法 例文作成の際に、文と句を区別して作成し、句点は文の場合にのみつける。句については、訳文作成時に動詞句の形で訳すことにする。

d. 言語ごとの特殊事情

問題点 言語によっては、訳文だけではわかりにくく、文法、用法、意味等について、説明が必要な場合がある。

対応方法 当該言語による注記を可能にするために、見出し語だけでなく、意味ごとにNote欄を設ける。

e. 概念・用法の説明や例文等への疑問

問題点 翻訳時に概念、用法、あるいは例文に関して日本語辞書編集者との情報交換が必要な場合がある。

対応方法 各見出し語ごとに掲示板を設け、コメントを書き込める仕組みを作る。この場合、担当者、担当言語名を自動的に表示するとともに、当該語の日本語辞書編集者にコメントがあったことを伝えることが可能な仕組みにする。

一方、例文翻訳時における問題点とその対応方法は次の通りである。

a. 主語・目的語の省略された文の翻訳

例： [遅れる] 例文：電車が遅れて遅刻した。

問題点 誰が何に遅刻したのか明らかでない。

対応方法 可能な限り一般化して翻訳するが、難しければ主語等を適当に補う。

b. 性別・単複等が明らかでない文の翻訳

例： [紛失] 例文：学生が本を紛失した。

問題点 主語の性別や単複が明らかでないので、翻訳しにくい。

対応方法 可能な限り一般化するが、難しければ文脈から適当に判断して訳す。

c. 意識か逐語訳かの選択

例： [開く] 例文：風が吹いて、ドアが開いた。

The wind blew the door open. / The wind blew and the door opened.

問題点 見出し語を生かして翻訳することもできるが、より自然な訳では見出し語を含まない訳、あるいは品詞が異なる訳になってしまう。

対応方法 2つの訳を併記する。上記のように左側に自然な言い方（意識）をおき、スラッシュで区切って、右側に意味の通じる範囲での逐語訳をおく。

d. 翻訳がむずかしい場合

例1： [歩く] 例文：フォアボールでバッターは一塁まで歩いた。

問題点 例えばスロヴェニアでは野球に関する知識を持つ学習者は非常に少ない。こうした場合、単に例文を翻訳しただけでは意味が通じない。

対応方法 意味ごとに設けられた Note 欄に当該言語で簡単な説明を加える。

例2： [ある] 例文：我輩は猫である

問題点 「我輩」「である」が作り出す文体レベルと「猫」との間にギャップがあり、特殊な文である。

対応方法 例文作成時に十分に配慮する。万一この例文を採用する場合には、Note 欄で十分な説明を加える。

以上、一連の運用実験で明らかになった問題点は、一般の辞書の編集にも共通するものである。また、特に例文の提示方法、翻訳方法に関しては、言語教育における語彙の指導方法にも示唆を与えるものである。

5. 多言語版日本語辞書の活用

各国語版の対訳辞書編集による成果物については、電子辞書としてインターネット上に公開する予定であり、すでに Chuta Dictionary 版 (<http://marmot.chuta.jp>)として実験的に公開を始めている。この辞書は、見出し語ばかりでなく見出し語の読みでも検索できる仕組みのため、同音異義語の意味・用法の違いを知ることが可能である。辞書としては、日日辞書に加えて、英語、オランダ語、スロヴェニア語版日本語辞書が



図4 Web辞書: Chuta Dictionary 版

閲覧可能である。図4は「会う」の英語版日本語辞書の表示画面である。

見出し語の右肩にある星印は日本語能力試験出題基準による級を表している。〔 〕内は別表記を示している。その他の辞書項目に関しては3.で述べたとおりだが、それに加えて、例えば英語版ではこの例のように各項目に英語訳がついたものが表示されている。また、辞書利用者からの意見・要望があることを想定し、画面下にコメント欄も設けてある。入力されたコメントは辞書編集者に伝えられ、編集者は辞書に適宜必要な修正を行う。

さらに、一定の編集作業が完了したものについては、リーディング・チュウ太の辞書ツールに組み入れ、インターネット上で公開していく予定である。本発表を機に、さらに多くの言語による対訳辞書の編集が行われることを期待している。また運用実験への協力者も広く募っている。

一方、本研究によって完成するシステムは多言語情報を一元的に管理することができるため、日本語教育のみならず、他の教育分野においても新しい形の教材共同編集システムとして活用可能である。現在、インターネット環境によって情報の共有が可能になったとはいえ、情報が特定の言語に偏っている傾向が強い。また、単一言語環境においてのみ情報が共有されている場合が多い。多言語環境に配慮したシステムの構築によって、個々の学習者の母語に対応した教材や辞書の共同編集・相互利用が可能になる。今後こうした形で教材の共同活用がさらに進んでいくことを期待している。

謝辞 本研究は、(財)電気通信普及財団の平成17年度研究助成および東京国際大学の平成17年度特別研究助成による支援を受けた。ここに記して感謝の意を表したい。

参考文献

川村よし子・金庭久美子ほか(2003)「辞書ツール多言語化プロジェクトの基本構想」『第8回ヨーロッパ日本語教育シンポジウム論文集』89-94.

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」『日本語教育工学会論文誌』24-Suppl., 7-12.

北村達也・川村よし子ほか(1999)「学習履歴管理機能を持つ日本語読解支援システムの開発とその評価」『日本教育工学会論文誌』23-3,127-133.

国立教育研究所(1998)『CASTEL / J CD-ROM V1.2』

鈴木庸子ほか(1998)「上級日本語読解教材「新書ライブラリー」の表現調査」『人文科学とコンピュータ』38-9,87-94.

寺朱美・北村達也ほか(1996)「WWW ブラウザを利用した日本語読解支援システム」『日本語教育方法研究会誌』3-1,10-11.

仁科喜久子(2002)「多言語表示日本語読解学習支援システム「あすなる」の開発」『第3回「日本語教育とコンピュータ」国際会議』71-74.

Maeda, Kawamura, Kaneniwa(2002), Making More Effective Uses of On-Line Reading Material, Proceedings of the FLEAT IV Conference,468-474.