

## 多言語版日本語辞書編集システムの開発と運用実験

|       |          |                       |
|-------|----------|-----------------------|
| 研究代表者 | 川 村 よし子  | 東京国際大学言語コミュニケーション学部教授 |
| 共同研究者 | 金 庭 久美子  | 横浜国立大学教育人間科学部講師       |
| 〃     | 前 田 ジョイス | 東京国際大学経済学部助教授         |
| 〃     | 川 村 ヒサオ  | ハッソウセンター代表取締役         |

### 1 研究の目的

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あげるシステムの構築にある。本研究の対象である多言語版日本語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。編集システムはインターネット上で公開し、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えるという仕組みを持つ。本研究において、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ついては、適宜対応を行い、インターネット上で自由に編集作業が行える多言語版日本語辞書編集システムを完成させる。本システムによって編集される各言語版の日本語辞書は、1 単語ごとに、編集が完了すると逐次 Web 辞書として公開できる仕組みになっている。また、完成した辞書については、すでに 1999 年から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ていく予定である。本研究で開発するシステムは、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り、ICT (Information and Communications Technology) 時代に対応した教材作成・共同編集のあり方の一つを示すものである。

### 2 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」、また、これを利用した国際基督教大学の読解システム (鈴木 1998) 等があるが、いずれも教材が限定された、いわば閉じたシステムである。これに対して、寺ほか (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) を組み入れたものが、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、入力された文章に含まれている単語や漢字のレベルを個々に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。

三輪ほか (2005) では、このシステムを用いて聴解・読解試験のための日本語学習支援システムを開発している。一方、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ている。ロンドン大学の D. J. Lefevre によって Reading Assistant が開発されたが、これ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 ほか (2000) の FLC 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は、東京工業大学の仁科ほか (2002) により構文解析機能のある学習支援システム「あすなろ」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。

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、いまだに十分に答えられていないというのが現状であり、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、この辞書ツール多言語化の早期実現を可能にする。

### 3 電子辞書編集のための基礎調査

多言語版日本語辞書編集システムの開発にあたり、国内外の日本語学習者を対象に電子辞書の利用に関する基礎調査（金庭/川村 2006a）を行った。これは辞書の利用の現状を調べ、日本語辞書に対する学習者のニーズを調べることを目的にしたものである。アンケートの詳細はここでは省略するが、調査の結果、以下のことが明らかになった。

学習環境の面から見ると、漢字圏の学習者（以下漢字圏）は電子辞書を多く利用している（84%）。また、漢字圏（3%）より非漢字圏の学習者（以下非漢字圏）（16%）の方が web 辞書の利用が多い。学習レベル別にみると、電子辞書の利用はレベルが上がると上昇する。また、利用頻度に関しては、漢字圏はバイリンガル辞書と日日辞書の双方を利用しているようである。一方、非漢字圏はバイリンガル辞書を「いつも」利用する者が多い（39%）が、日日辞書については、「ぜんぜん」と答える者が多く（34%）、日日辞書の利用は少ない。以上のことから、非漢字圏学習者にとって、バイリンガル辞書の必要性は高く、本システムを開発する必要性が極めて高いことが裏付けられた。

一方、学習者が辞書の利用内容のどこに注目しているのかに関する調査では、次のようなことが明らかになった。非漢字圏は他の項目に比べ、「意味」に注目している（86%）。一方、漢字圏は非漢字圏に比べ「発音・アクセント」を見ている（24%）。次に学習レベル別に見ると、初級学習者は特に「意味」に注目し（80%）、中級・上級学習者の約4割が「例文」に、上級学習者の88%が「読み方」に、33%が「慣用句・連語」に関心を示していた。これらのことを総合して考えると、辞書の開発においては、①意味がすぐわかるような記述をする ②見出し語を明確に示す ③例文、慣用句・連語を示す、などが必要であることが判明した。

また、自由記述式の回答からも、例文や文法・用法に対するニーズが非常に高いことがわかった。日本語学習者用の辞書には、質の高い例文とわかりやすい用法説明が必須である。また、英語以外のバイリンガル辞書に対する要望が多いが、それと同程度に学習者の母語と英語という形で、多言語を同時併記する日本語辞書に対する要望もある。さらに、検索方法については「検索ボックスが1カ所で意味や例文等の検索ができるもの」、「検索の入力文字が漢字、ひらがな、ローマ字のいずれでもできるもの」が求められていることが明らかになった。以上の調査結果をもとに、辞書編集システムの開発をすすめることにした。

### 4 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、日本語辞書編集者が作成した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（川村 2003）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・閲覧者等が設定でき、各言語版の辞書情報は SQL サーバに保存される仕組みになっている。



図1 多言語版日本語辞書編集システムのメニュー画面

図1は多言語版日本語辞書編集システムのメニュー画面である。辞書編集者は、画面左の 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、

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。言語に関してはデフォルト値として担当言語が表示される仕組みになっているが、検索のみであれば担当言語以外の情報を閲覧することも可能である。単語の編集は My Word List から一単語ずつ選んで行うことになる。

画面右は、辞書の編集にかかわるすべてのプロジェクトメンバーを対象にした掲示板である。Comments on My Word には自身が担当している単語へのコメント、New Comments にはそれ以外の単語へのコメントが掲載され、日本語編集者と各言語版の編集者との間で相互に意見交換できる仕組みになっている。また、辞書の利用者からのコメントもここに掲載される。

編集作業を行うには、まず、言語ごとに編集責任者を登録する必要がある。編集責任者は、担当言語および日本語が堪能であること、十分なコンピュータスキルを持っていること、コンピュータ環境が整っていることを条件に依頼している。各言語の責任者は、担当言語の編集者・閲覧者を適宜登録することができる。

The screenshot shows a web browser window titled "WordEdit - Microsoft Internet Explorer". The address bar shows the URL "http://mamot.chuta.jp/chutaic/WordEdit.asp?worduser\_ID=94790". The page content includes a navigation bar with "MainMenu > My Word List > Word Edit(French)" and a "Logout" link. The main form is for editing a word entry with the following fields and values:

- ID: 7461
- Headword: 会う 会う / あう
- Hiragana: あう
- Class: 動詞/自立
- Sense 1: 人と人が面会する
- Same Word: 面会する[メンカイ・スル]
- Translation:
- Example: 約束の時間に先生に会う
- Note:
- Sense 2: 思いがけない人や事件と出会う
- Same Word: 遭遇する[ソウグウ・スル]

図2 多言語版日本語辞書編集システムの編集画面

図2は多言語版辞書編集システムの編集画面である。各言語の編集者は辞書項目ごとに指定された入力ボックスに対訳情報を入力していく。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 サーバに保存され、他の言語の辞書情報とリンクされる。

多言語版辞書編集システムによる運用実験は、英語版・トルコ語版・ブルガリア語版編集チームの協力を得て行われた（川村ほか 2006）。その結果、辞書項目の見直し、特に例文の提示方法に関する見直しが行われた。また、言語によって個別の説明が必要となる意味項目があることが明らかになったため、意味ごとに Note 欄を設けることにした。例えば「歩く」という語には「野球において、四死球を得て塁に出る」という意味があるが、スロヴェニア、チェコ等では、野球に対する関心がほとんどなく、野球のルールについての知識もないため、個別の説明が必要となる。こうした場合に、この Note 欄が活用可能である。また Note 欄は文法の説明等にも利用できる。

## 5 日本語学習者のための多言語版 Web 辞書

多言語版日本語辞書編集システムを利用して編集された各言語版の日本語辞書は、各単語の編集が完了すると、自動的にそのままインターネット上で公開される仕組みになっている。これが、日本語学習者のための多言語版 Web 辞書である（<http://chuta.jp/>）。

この辞書は、1) 日日辞書としても多言語対訳辞書としても利用が可能、2) インターネット上での利用が可能、3) 日本語能力試験出題基準に準拠した約 8600 語の見出し語、4) 各単語のすべての意味について学習者向けの例文付き、5) 用法の説明あり、6) 学習者からのフィードバック機能あり、7) 単語の追加編集が

随時可能、8)辞書情報の追加、修正、削除が随時可能、という特徴を持っている。特に6)7)8)は従来の辞書にない大きな特徴で、「日本語学習者のための進化する Web 辞書」であり、時代やニーズにあった最新情報を盛り込むことが可能な辞書となっている(金庭/川村 2006b)。

現在、辞書としては、日日辞書に加えて、英語、ブルガリア語、トルコ語、ドイツ語、スペイン語、ポルトガル語、マレー語、韓国語、マラティ語、チェコ語、スロヴェニア語、中国語、タガログ語版の日本語辞書の編集が始まっている。また、フランス語、ハンガリー語、イタリア語、ルーマニア語、ヒンディー語等から辞書編集の申し出を受けている。日日辞書については、日本語能力試験の出題基準に含まれるすべての単語の編集が完了している。今後、さらに日本語学習者にとって必要な語(徳弘ほか 2006、川村ほか 2007)を適宜追加していく予定である。一方、各言語版の日本語辞書は、それぞれ編集済みの単語はまだ多いとはいえない。しかし、各単語の編集作業が済むと、自動的に辞書が更新されるシステムであり、母語による対訳辞書を一日でも早く利用したいと考えている学習者に最新版の辞書を提供できるという利点がある。また、辞書編集者も自らの編集作業の成果をその場で確かめることが可能である。

図3 多言語版 Web 辞書検索画面

図3は多言語版 Web 辞書の検索画面である。検索したい語を入力ボックスに入力し、中央の言語ボックスから言語を選択し、「Search」ボタンを押すと、選択した言語版の辞書情報が閲覧可能になる。また、見出し語を、かな(ひらがな・カタカナ)あるいはローマ字で入力すると「会う」「逢う」「合う」「遭う」のような同音異義語がすべて表示される。

図4 トルコ語版の辞書画面

図4は「会う」のトルコ語版の辞書画面である。見出し語右肩の星印は、日本語能力試験の出題基準に準拠した級を示している。意味概念ごとに、概念説明、訳語、例文およびその対訳が示されている。語の用法等に説明が必要な場合には、Note の形で注がつけられている。画面下のコメント欄は、単語ごとに設けられ、辞書利用者が意見・要望等を辞書編集者に送ることができる。書き込まれたコメントは、単語情報が自動的

に付加された形で当該言語の担当者に届き、適宜対応できる仕組みである。

この多言語版 Web 辞書は、例文検索システムの機能もあわせもっている。図 3 の検索ボックスに検索語を「\* (半角アスタリスク)」で挟んで入力すると、例文検索システムとしても機能し、その語を含むすべての見出し語の辞書項目に加えて、見出し語とは無関係にその語を含むすべての例文が表示される。各々の例文には対訳情報がつけられているのみならず、その例文が属している見出し語へのリンクがはられている。そのため、リンクをたどりながら次々に新しい語を学習していくことができる。さらに、部分一致の検索も可能なため、助詞や助詞相当句を含む例文を検索したり、活用形で検索したりすることも可能である。この例文検索システムを活用した教育方法については、金庭/川村 (2007) で詳しく述べているので、ここでは省略するが、このシステムを活用した語彙学習を授業に取り入れた結果、特に語構成の学習や連語の学習に役立つことがわかった。派生語では、語の意味を類推しながら学べるので語彙数を増やす効果がある。また、「名詞+助詞+動詞」を組み合わせて学習させる連語の必要性は従来から指摘されているが、このシステムを利用すれば名詞と動詞のいろいろな結びつき方を学習者が自ら確かめながら学習することが可能になる。

## 6 今後の課題

多言語版日本語辞書編集システムの運用実験は、現在も英語版・トルコ語版・ブルガリア語・チェコ語版編集チームの協力を得て行っている。運用実験の結果、日日辞書の概念の説明方法、例文などの提示方法に関して再検討が行われた。また、各言語版と英語版の併記、ローマ字による検索、例文への振り仮名付加、モバイルへの対応等の必要性が明らかになり、対応可能なものから順次改良を行っている。すでに、ローマ字による検索およびモバイル対応は実現している。この運用実験については、今後も継続するとともに、利用者からのフィードバックも参考にしながら、学習者にとってより使いやすい辞書を作り上げていく予定である。

一方、各言語版辞書の編集作業についても継続的に行われるが、欧州のみならず世界各地の日本語教育関係者にも呼びかけ、さらに多くの言語版の編集協力者を募っていく予定である。また、多言語版日本語辞書編集システムによって完成した辞書は、「リーディング・チュウ太」の辞書ツールに組み入れる予定である。

本発表の辞書編集システムは、母語話者と非母語話者双方の意見を交換しながら辞書の編集を進めるという国際共同編集を可能にしている。編集者同士がメーリングリストやコメント欄を活用して相互に情報交換を行い、その過程で明らかになった問題点について適宜修正を行うとともに、各国の日本語学習者の視点からの辞書内容の修正も継続的に行われている。こうした新しい形の国際共同編集は、ICT 時代の学習環境構築のあり方や言語教育用教材の作成方法等にも示唆を与えるものと言えよう。

**謝辞** 本研究は、(財)電気通信普及財団の研究助成(平成 17 年度・18 年度)および科学研究費(課題番号 18320083)の研究助成(平成 18 年度・19 年度)を受けている。ここに記して感謝の意を表したい。

## 【参考文献】

金庭久美子・川村よし子

日本語学習者のための電子辞書編纂の基礎調査, *Japanese Language Education in Europe*, vol.10, 77-82, 2006a.

金庭久美子・川村よし子, 日本語学習者のための電子辞書編纂, 日本語教育方法研究会誌, vol.13 No.1, 26-27, 2006b.

金庭久美子・川村よし子

例文検索システムの構築と日本語教育への応用, 日本語教育方法研究会誌, vol.14, No.1, 6-7, 2007.

川村よし子

辞書ツール多言語化プロジェクトの基本構想, *Japanese Language Education in Europe*, vol.8, 89-94, 2003.

川村よし子・北村達也・保原麗

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, 2000.

- 川村よし子・前田ジョイス・金庭久美子・植木正裕・川村ヒサオ・根津誠・保原麗・Hans Coppens・Jonathan Bunt・Kristina S. Hmeljak・Saleh Adel,  
多言語版日本語辞書編集システムの開発と運用実験, *Japanese Language Education in Europe*, Vol.10, 146-151, 2006.
- 川村よし子・前田ジョイス・金庭久美子, 日本語学習者のための基本語選定の一試案 *Japanese Language Education in Europe*, Vol.11, 72-79, 2007.
- 北村達也・川村よし子・内山潤・寺朱美・奥村学  
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, 1999.
- 国立教育研究所  
CASTEL/J CD-ROM V1.2, 1998.
- 鈴木庸子  
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, 1998.
- 寺朱美・北村達也・落水浩一郎  
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, 1996.
- 徳弘康代・川村よし子  
漢字の造語能力に関する基礎調査, 日本語教育方法研究会誌, Vol.13 No.2, 16-17, 2006.
- 仁科喜久子・奥村学・八木豊・戸次徳久・沢谷孝志・傅亮・杉本茂樹・阿辺川武  
構文表示と多言語インターフェースを備えた日本語読解学習支援システムの開発, 言語処理学会第8回年次大会発表論文集, 228-231, 2002.
- Maeda, Kawamura, Kaneniwa, Making More Effective Uses of On-Line Reading Material, *Proceedings of the FLEAT IV Conference*, 468-474, 2002.
- 松本裕治・北内啓・山下達雄・平野善隆・松田寛・高岡一馬・浅原正幸  
日本語形態素解析システム『茶筌』 version 2.2.1 使用説明書, 2000.
- 三輪譲二・川村よし子  
Web 多言語辞書機能を活用した聴解・読解試験のための日本語学習支援システム、日本語教育学会 2005 年度第 9 回研究集会(口頭発表), 2005.

### 〈発 表 資 料〉

| 題 名                               | 掲載誌・学会名等               | 発表年月       |
|-----------------------------------|------------------------|------------|
| 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         | 2006 年 5 月 |
| 複合語検索システムの基本構想                    | 日本語教育文法研究会             | 2006 年 7 月 |
| 日本語学習者のための基本語選定の一試案               | 第 11 回ヨーロッパ日本語教育シンポジウム | 2006 年 9 月 |
| 漢字の造語能力に関する基礎調査                   | 日本語教育方法研究会誌            | 2006 年 9 月 |
| 例文検索システムの構築と日本語教育への応用             | 日本語教育方法研究会誌            | 2007 年 3 月 |