辞書ツール多言語化プロジェクトの基本構想

The basic concept for multilingualization
of the Reading Tutorial Dictionary Tool

川村よし子（東京国際大学教授）
kawamura@tiu.ac.jp

共同研究者：植木正裕 金庭久美子 川村ヒサオ 根津誠 保原麗

はじめに

1997 年から開発を進め 1999 年 10 月からインターネット上で公開してきた日本語読解学習支援システム「リーディング・チュウ太」は学習者がインターネット上の情報を用いて自由に読解学習をすすめることができる自律学習支援システムである（http://language.tiu.ac.jp）。公開以来すでに 25 万件のアクセスがあり、毎日平均 200 人の人が利用している。このシステムには辞書ツール・レベル判定ツール・リンク集・読解教材バンク・読解クイズ・文法クイズ等がモジュール化されて組み込まれ、有機的に統合されている。このうち、辞書ツールとしてはすでに日日・日英・日独の３種類を提供しているが、さらに多くの言語への対応が望まれているというのが現状である。

本研究は、この辞書ツールの多言語化をめざしたものである。辞書の登録・編集システムをインターネット上に置き、日本語辞書に対し、世界各国の編集者が各国語版の対訳をいつどこでも自由に編集できるようにして電子辞書を作成する。現在、オランダ語、スロヴェニア語、中国語などの編集を予定している。本プロジェクトの研究成果物は、電子辞書として完成させるだけでなく、読解学習支援システム「リーディング・チュウ太」に組み入れ、インターネット上で公開する予定である。

１．読解学習における辞書ツールの役割と教育効果

読解学習において辞書を引くことは不可欠の要素である。ところが、日本語学習者にとって辞書引き作業は、他の言語の学習者とは比較にならないほど多くの時間と努力を要する。個々の漢字の読みを知っていても、熟語の読みがわからなければ、一般の日本語辞典をひくことはできない。漢字から検索可能な辞典もありはするが、これを使うのは時間がかかるし、すべての単語が載っているわけではない。こうしたことから、特に非漢字圏学習者にとっては、辞書引きの負担が大きい。そのため、中上級レベルの学習者であっても語彙リストは不可欠で、語彙リストのついた教材以外の読解学習はなかなか進まないという状況が長いこと続いていた。

1999 年から公開をはじめた「リーディング・チュウ太」には、文中の辞書引きを自動で行う辞書ツールが組み込まれている。学習者が読みたい文章をコピー＆ペーストで辞書ツールに入力しさえすれば、コンピュータが文章を解析し、辞書引き作業を自動的に行う。結果の出力画面の本文の単語はすべて辞書情報とリンクされているので、学習者は読みや意味のわからない単語をクリックすれば、即座に辞書情報が得られる仕組みになっている。

この辞書ツールを利用するメリットは次のとおりである。

① 辞書情報が即座に提供されるので、学習者は辞書引き作業に時間をとられずにすみ、読解そのものに専念できる。

② すべての単語の辞書情報が提供されているので、レベルの異なる学習者に対しても同一の教材を使用することが可能である。

③ 電子化された日本語の文章であればどんな文章でもその場で教材化されるので、学習者が自らの興味やニーズにあわせて読解学習をすすめることができ、自律学習につなげることが容易である。

④ 単語の意味が語彙リストではなく、辞書情報の形で提供されているため、各単語の文中における意味を考えるという読解学習の重要な部分は学習者自らが行う。

⑤ クラス全体が同一の辞書を利用することになるので、各単語の中心的な意味や派生的な意味、また本文中での意味を考えるという作業をクラスワークとして行うことも可能である。

こうしたメリットがある一方、辞書引きがあまりにも簡便なので、学習語彙の習得が進みにくいのではないかという懸念がある。それを避けるため、辞書ツールは学習履歴を残す仕組みを備えている。学習者が単語をクリックするごとに学習履歴が残り、「あなたの単語リスト」という形で、辞書を参照した単語と参照した回数とが表示される。このリストの各々の単語は辞書情報とリンクされているので、学習者は読解終了後に復習することができる。さらに、このリストを保存したり、印刷したりすることも可能である。

図１は辞書ツールを用いた読解学習の際に学習履歴機能を利用した場合と利用しなかった場合との語彙学習の効果を比較したものである。この図で示されるように、学習履歴の活用は、語彙の習得度を高める効果がある。（北村・川村他 1999）

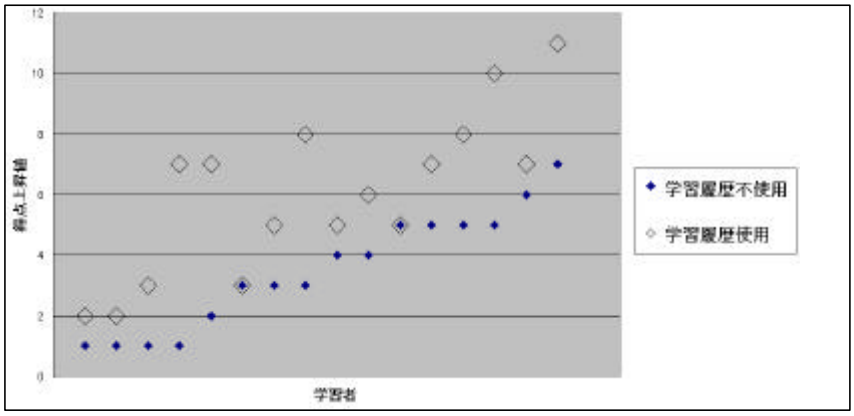

図１ 学習履歴の活用による語彙学習の効果

２．辞書ツールに必要な辞書情報

では、読解学習のための辞書ツールとしてはどのような辞書情報が必要になるのだろうか。また、辞書ツールに組み入れることを前提にした辞書としてはどのような形で編集すればいいのだろうか。

日独辞書ツール開発の際に行った評価実験（川村 2001）の結果、次のような問題点が明らかになっている。

① 見出し語の選定

辞書ツールでは「茶筌 2.02」（松本他 1999） を用いて形態素解析を行っているため、見出し語を茶筌の辞書に準拠した形にしないと、辞書情報を表示できない。複合語や慣用表現についての情報も不可欠なのだが、やはり茶筌の制約がある。これらに対して、何らかの対応方法を考える必要がある。

② 品詞情報
適切な辞書情報を提示するためには品詞情報も茶筌に準拠したものにするとともに、各見出し語に対して可能性のある品詞情報を網羅する必要もある。
③ 意味情報
意味情報が多すぎると、単語の意味が捉えにくいことがある。日本語学習者にとって必要な意味から順に並べると共に、意味情報を厳選する必要がある。
④ 例文
各々の単語がどのように使われるのかを示す例文がほしい、特に文脈によって意味が異なる語に関しては不可欠である。

――以上の点をふまえ、今回開発する辞書ツール用の辞書は次のような特徴を備えたものにする。

① 見出し語・複合語・慣用表現
日本語学習者にとって必要な語から順次、辞書の編集作業を行う。見出し語としては、基本的には茶筌の辞書に準拠した形にするが、複合語や慣用表現等、茶筌にない語句であっても見出し語として登録し、こうした情報も検索可能な仕組みを別途考える。(詳細に関しては稿を改めて述べることにする。)
② 品詞情報
品詞情報に関しても茶筌に準拠したものにするが、日本語学習者の便宜を考えて、二つ以上の品詞にわたる場合にはそれを列挙し、品詞ごとに意味情報を記述することにする。
③ 意味情報
各単語の意味情報をできるだけ絞り込むと同時に、下記のような配慮を加え、できるだけ学習者本位の辞書を作り上げる。
a. 意味情報は概念ブロックごとにまとめる
b. 中心となる語義から順に並べる
c. 初中級者向けの情報と上級以上の学習者向けの情報とを区分する
d. 派生的な意味を扱う場合には、用例を示す
④ 例文
現状での対応は難しいが、全体の仕組みとしては、将来例文の提示もできるような汎用性をもったものを考える。

３．辞書の編集システムの概要

辞書ツールの多言語化のためには、各国語版を一元的に管理できるようにしなければならない。そこで、従来は日本語、英語等異なった言語のプラットフォーム(OS やブラウザ)ごとにシステム開発を進めていたが、今回、プラットフォームに依存せずに処理が可能なシステムを開発することにした。辞書の基本構造を決定し、すべての辞書情報をこの枠組みにあわせて XML（eXtensible Markup Language）化する。これにより、それぞれの言語ごとにその使用言語にあわせたタグの挿入のみで、各国語版を統一して管理できるようになる。

XML とはテキストの構造（階層構造や順序関係）などを表すために、もととなるテキスト情報に特定の形式の文字列を埋め込んだもので、この XML 化によって、テキスト情報（例えば辞書情報）を辞書項目（例えば、「見出し語」「表記」「読み」「品詞」等）ごとに別個に扱うことが可能になる。今回の多言語化辞書の基本構造やその XML 化に関する詳細についてはここでは省略するが、

基本的な構造としては後掲の「**資料**」のような形を考えている。

辞書の編集作業の基本的な流れは次のとおりである。

① XML 化のための辞書の基本構造を決定する
② 語彙を選定し、見出し語を決定する
③ 各語の基本情報を入力する
④ 各語の日本語情報を入力する
⑤ 各語の各国語情報を入力する
⑥ 各国語版を完成させる
⑦ 各国語版の辞書をチュウ太の辞書ツールに組み入れる

辞書の編集に関しては、プラットフォームの違いの問題に可能なかぎり対応するとともに、書式を統一されたものにするために、インターネット上で利用可能な辞書登録・編集システムの開発をすすめた。完成したシステムはインターネット上に置き、まず、日本語情報を担当する編集者が③の基本情報および④の日本語情報を入力する。日本語情報の入力が完了したものは適宜インターネット上に公開し、世界各国の各国語版辞書の編集者がそれぞれ個々に⑤の各国語情報を作業できるようにする。

４．日本語情報の編集システム

現在すでに日本語情報の登録・編集システムが完成し、インターネット上に置かれている。図２はその入力画面である。辞書の編集者が各々の項目に必要事項のみ入力すれば、各項目の辞書情報が自動的に基本構造に従って XML 化できる。

辞書の編集作業は日本電子化辞書研究所の『EDR 日英対訳辞書』をもとにして行う。この辞書は現行のリーディング・チュウ太の辞書ツールが採用しているもので、約 25 万語の情報が含まれている。このうち日本語学習者にとって重要な単語（基本語彙）から逐次編集作業を進めていく。

編集に関しては、複数の編集者が並行して作業を進めていく必要があるため、単語（ブロック）ごとにあらかじめ担当者を決定するという方式を採用する

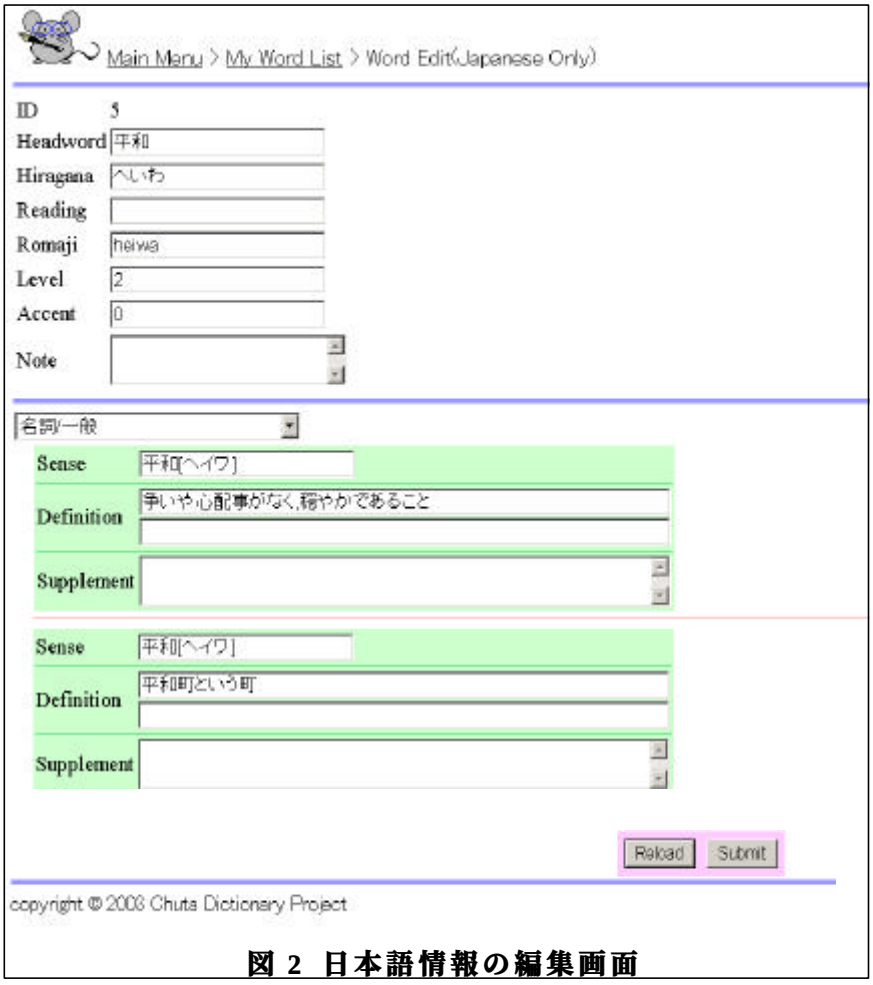

図 2 日本語情報の編集画面

ことにした。担当者の決まった単語については担当者以外は編集できない仕組みになっている。編集済みの辞書項目に関して修正が必要な場合も、修正は各々の単語の担当者が行う。他の編集者が問題点に気づいた場合には、当該の単語のコメント欄に書き込み、各単語の担当者がこれをもとに修正する。

辞書情報に関しては、各単語の見出し語、読み、表記（複数表記も含め）、発音（長音等のように平がなと発音が異なっている場合）、能力試験に準拠した難易度レベル等の基本情報に関してはすでに入力が完了している。日本語情報の編集者は、品詞、概念語、概念説明等の項目について編集作業を行う。編集にあたっては、2.で述べた編集方針をもとに、あくまでも日本語学習者向けの辞書であることに留意する。

５．各国語版の編集

各国語版の編集システムに関しては現在開発中である。基本語彙に対して日本語情報の入力が完了した時点で試行実験を行い、その後、一般に公開する。各国語版については、言語ごとにあらかじめ責任者を決め、編集者の登録や担当項目の決定等は責任者が行うことにする。各国語版の作成にあたって入力が必要な個所は、「**資料**」の「基本構造」に示した <p lang=" "> の部分である。意味情報の「概念説明」と「訳語」が主要な入力項目となる。また、言語によって補足説明が必要な場合を考慮して<extra_info>の項目を設けてある。編集の結果得られた成果物に関しては、一定量の辞書情報が集まり次第、辞書ツールに実装し、ネット上で公開していく。

辞書ツールの多言語化は単に色々な辞書が便利に使えるようになるというメリットだけではない。対訳の日本語辞書自体が存在しない、あるいは、入手が難しい国・地域も多い。辞書ツールの多言語化が実現すれば、それぞれの国・地域の日本語学習者にとって極めて有用な学習支援になるにちがいない。

このプロジェクトに関しては、リーディング・チュウ太のサイトで適宜進展状況を報告するとともに、ヨーロッパ教師会のメーリングリストにも情報を流す予定である。多くの国の方々の参加、協力によって日本語学習者のための辞書を完成させていきたい。

謝辞：本研究はプロジェクトメンバーに加えて、ベルギーのルーバン・カトリック大学のハンス・コッペンス氏から多大な協力を得た。また、プロジェクトに対して e-Japan の助成を得ている。ここに記して感謝の意を表したい。

参考文献：

川村よし子(2001)「日独辞書ツールの開発とその評価」『第 14 回日本語教育連絡会議論文集』pp.59-63.

北村達也・川村よし子・内山潤・寺朱美・奥村学(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), pp. 127-133.

日本電子化辞書研究所(1996)『EDR 電子化辞書仕様説明書』

松本裕治・北内啓・山下達雄・平野善隆・松田寛・浅原正幸(1999)「日本語形態素解析システム『茶筌』version 2.0 使用説明書第二版」NAIST-IS-TR99012.

資料：

辞書の基本構造

[< >内はタグ名 、" "内には番号や種別が入る。右の（ ）内は各タグの意味]

```
<entry index="  ">                                          (見出し語 番号="  ")
        <headword>                                          (見出し語)
        <notation number=" ">                               (別表記 番号="  ")
        <accent>                                            (アクセント)
        <hiragana>                                          (ひらがな表記)
        <reading_info>                                      (発音)
        <romaji>                                            (ローマ字表記)
        <level>                                             (語彙のレベル)
        <stylistic_category>                                (文体)
        <field_label>                                       (専門分野)
        <etymology>                                         (語源)
                <ety_lang>                                  (語源の属する言語名)
                <ety_word>                                  (語源となった語)
        <onomatopoeia type=" " >                            (擬声語 擬態語 分類="  ")
        <compound>                                          (複合語)
                <part number=" " ref=" " romaji=" ">          (要素 番号="  " 参照=" " ローマ字="  ")
                <part number=" " ref=" " romaji=" ">          (要素 番号="  " 参照=" " ローマ字="  ")

        <part_of_speech number=" " type=" ">                (品詞 番号=" " 種類="  ")
                <sense number=" ">                          (意味 number=" ")
                        <usage>                                     (用法)
                                <j></j>                                     (日本語)
                                <p lang="  ">                               (外国語 言語="  ")
                        <same>                                      (概念語)
                                <j></j>                                     (日本語)
                                <p lang="  ">                               (外国語 言語="  ")
                        <definition>                                (概念説明)
                                <j></j>                                     (日本語)
                                <p lang="  ">                               (外国語 言語="  ")
                        <translation>                               (訳語)
                                <p lang="  ">                               (外国語 言語="  ")
                        <extra_info>                                (意味補足説明)
                                <p lang="  ">                               (外国語 言語="  ")
                        <example number=" " type=" ">               (例文 番号=" " 種類="  ")
                                <j></j>                                     (日本語)
                                <r></r>                                     (読み)
                                <p lang="  ">                               (外国語 言語="  ")
                        <syn number=" " ref=" ">                    (同義語・類義語 番号=" " 参照=" ")
                        <ant number=" " ref=" ">                    (反意語・対義語 番号=" " 参照=" ")

                <idiom number=" " type=" ">                 (連語・慣用句・ことわざ 番号=" " 種類="  ")
                                <j></j>                             (日本語)
                                <r></r>                             (読み)
                                <p lang="  ">                       (外国語 言語="  ")
```