

# 学習辞書編集支援データベースの構築

砂川有里子・李在鎬・高原真理  
筑波大学

キーワード：日本語学習辞書 辞書編集支援 二言語辞書 データベース

## 1. はじめに

海外での日本語学習者は2009年に365万人を越え、過去30年間で28.7倍に増加した<sup>1</sup>。日本語能力試験の受験者数も年々増加しており、とりわけ上級レベルの増加が目立つ。2009年度には上級レベル(1・2級)の受験者数が10年前の6.4倍になり、その比率も、全受験者数の55%であった10年前に比べて76%と大きく伸びている<sup>2</sup>。

日本語を学習するには、学習者の母語の特徴を踏まえて開発され、母語で書かれた日本語学習辞書、すなわち、母語によるバイリンガル日本語学習辞書が役に立つ。特に中級以降になると、自力で本を読んだりものを書いたりする機会が多くなり、レベルが上がるにつれて受信と発信のどちらの目的も満たす日本語学習辞書がより強く求められるようになる。しかし、現状では、日本語学習者が多数を占める中国や韓国など少数の国を除き、初級者向け、あるいは旅行者向けの簡易な辞書しか手に入らない地域がほとんどといってよい。

辞書開発には莫大な費用と労力と時間が必要である。筆者が経験した母語話者向けの日本語辞書の場合、ベテランの辞書執筆・編集者と出版社編集部との強力なチームが、10年近くもの歳月をかけて試行錯誤し、ようやく完成させるという手のかけようであった。母語話者用日本語辞書とは比べようもなく貧弱な市場でしかない世界各地の日本語教育現場では、採算という点からも人材という点からも、本格的な辞書の執筆や編集に取りかかる余裕がないというのが実情である。

しかし、最近ではインターネットという強力なメディアが出現し、この状況に大きな変化をもたらしつつある。出版社を通じて書籍の辞書を刊行するには、多くの経費と時間をかけなければならないし、各国の出版と販売事情で普及に支障がある場合が少なくないが、インターネットを通じて発信するウェブ辞書なら、辞書に必要なコンテンツの作成と学習者に利用してもらうためのシステムの開発という問題さえ解決できれば、ほとんど経費をかけずに利用者に辞書を届けられる。学習辞書の普及という問題は、インターネットの利用で大きく改善され、世界中に無料で辞書を提供できる状況が整いつつある。

問題は、辞書コンテンツの作成とそれを学習者に配信し、利用してもらうためのシステム開発という点である。そこで本プロジェクトは、日本語学習辞書に必要なコンテンツをデータベースとして電子的な形で構築し、データベースの情報を世界各地にインターネットを通じて配信することを目的とする。それにより、その地域の辞書編集者がその情報を自由に取り入れてその地域にふさわしいものに加工したり、その地域に必要な情報を新たに加えたりすることによって、その地域の利用者に適したバイリンガルの日本

---

<sup>1</sup> 「海外の日本語教育の現状 日本語教育機関調査・2009年 概要」

<http://www.jpff.go.jp/japanese/survey/result/index.html> (2012年7月21日)

<sup>2</sup> 「日本語能力試験受験者の推移」<http://www.jlpt.jp/statistics/index.html> (2012年7月21日) に掲載されている統計資料に基づいて筆者が計算した数値に基づく。

語学習辞書を編集し、それをウェブ辞書として無料で発信したり安価な書籍として刊行したりすることができるようになる。

ウェブ辞書という点においては、すでに多言語版ウェブ辞書「チュウ太の Web 辞書」(<http://chuta.jp/>)がある。これは、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)」の辞書ツールとして開発されたもので、現在 32 言語に対応し、その数は今後も増える見込みである。「チュウ太の Web 辞書」は、多数の言語によるバイリンガル学習辞書の可能性を広げる意欲的な試みであるが、あらかじめ作成した日日辞書をもとにして多言語化を推し進めるものであるため、各国語の実情に即した自由な辞書作りにつながるものではないという制約がある。中・上級レベルの学習者が使いやすい辞書を開発するには、学習者の母語との対照研究を踏まえ、母語の実情に即したその地域固有の辞書を自由に編集できる環境が望ましい。本プロジェクトが構築を目指す「日本語学習辞書編集支援データベース」は辞書そのものを作るのではなく、どの母語話者の日本語学習にとっても必要と思われる汎用的な語法情報と適切な用例を提供することにより、各国語の母語話者がその情報を活用してそれぞれの母語に適した固有の辞書開発を行うことにつなげるという点で、どのような母語にも対応できるバイリンガル日本語学習辞書開発の環境を整えるところに特色があり、この点がこれまでにない新しい試みであると言える。

以上のような状況を踏まえて、私たちのプロジェクトチームは、日本語学習辞書編集に必要な情報を搭載したデータベースを構築し、辞書編集を志す世界各地の人々にバイリンガル日本語学習辞書の編集支援を行うことを目指す。このプロジェクトは、科学研究補助金基盤研究 (A) の助成を受け、『汎用的日本語学習辞書開発データベース構築とその基盤形成のための研究』(略称『学習辞書科研』) という名称のもとに 2011 年 4 月に開始され、2014 年度までの 4 年間のプロジェクトとして活動する。以下ではこのプロジェクトの概要を紹介することにしたい。

## 2. プロジェクトチームの構成

本プロジェクトは、日本語辞書編集支援データベースの構築チームと、その活動を周辺から支える研究協力チームの 2 チームからなる。

データベース構築チームは 30 名で、代表者の砂川有里子の他、研究分担者 11 名、連携研究者 18 名、非常勤研究員 1 名からなる。これらのメンバーは日本語研究グループ、コーパス研究グループなどに分かれて、日本語の語法情報をデータベースに組み込む方法や、コーパス研究を辞書記述に役立てる方法を検討するほか、データベース構築の作業にも携わる。日本語研究グループのもとには、(1)コロケーション研究班、(2)類義語研究班、(3)文法情報研究班、(4)文化情報研究班、(4)音声情報研究班、コーパス研究グループのもとには、(1)コーパス情報研究班、(2)基本語彙研究班、(3)学習者コーパス研究班、(4)言語処理班を配置し、それぞれの研究にあたる。

研究協力チームは現在の時点で 47 名で海外在住の研究協力者多数を含む。国内の研究協力者は英語辞書編集、コーパス言語学、日本語研究、英語/フランス語/ドイツ語などの外国語研究、日本語教育研究に携わる研究者で、それぞれ独自の観点から日本語学習辞書編集に資する研究を行い、口頭発表、論文発表などを通じてその成果を全メンバーと共有する。海外の研究協力者は、日本語辞書編集、各種のコーパス構築、日本語研究、日本語教育研究の研究者で、国内メンバーと同じく独自の研究成果を本プロジェクトに還元するとともに、海外での日本語学習辞書のニーズ調査や学習者コーパスの作成協力、学習者の誤用研究など、データベース構築に必要な調査・研究を行う。

### 3. 日本語辞書編集支援データベースとは

辞書の開発に必要なのは、対照研究や言語学の研究成果を踏まえた日本語の語法に関する詳細な記述である。そこで、本プロジェクトは、中級から上級レベルの学習者を対象とした日本語学習辞書の編集を支援することを目的として、以下の情報を搭載したデータベースを構築する。

- (a) 見出し語の語法情報（語義・文法情報・音声情報・類義語情報・コロケーション情報・文体情報・文化情報・コーパスに基づく頻度情報など）。
- (b) 語義ごとに典型的な用法の用例を中級・上級レベルの学習者に即した難易度で作成した例文。
- (c) 日本語学習者が犯しがちな誤用情報。

以上の情報には Creative Commons ライセンスを付与する。それにより、世界各地の辞書編集者が当データベースに自由にアクセスし、営利、非営利の別を問わず、現地の状況に即して自在に情報を加工することによって、それぞれの母語によるバイリンガル日本語学習辞書を開発することを可能にする。

以上のデータベースを構築するため、研究期間内に以下の作業を行う予定である。

- (a) 日本語学習者にとって必要な日本語教育基本語彙の選定を行う。
- (b) 既存の日本語コーパスを活用するなどして、日本語教育基本語彙の語法情報をデータベースに搭載するための研究を進める。
- (c) 既存の学習者コーパスを活用して誤用情報をデータベースに搭載するための誤用研究を行う。
- (d) 既存の日本語コーパスを活用し、各見出し語の語義に即した典型的な用例を調査し、中級から上級レベルの日本語学習者に適切な作例を行う。
- (e) 語法情報整備のためのシステムと語法情報執筆用のコーパス活用ツールを開発する。
- (f) データベース公開用のシステムと利用者向けツールの開発を行う。

本プロジェクトでは、データベース作成の第一ステップとして、どのような語彙を対象にデータベースを作成するかについて検討し、約 18,000 語からなる「日本語教育基本語彙表」（以下、「教育語彙表」）の構築を行なった（詳しくは李（2013）を参照のこと）。

### 4. 現在の進捗状況

現在、上述の「教育語彙表」をもとに、辞書開発のためのデータベース作成とシステム開発を行なっている。具体的には、データベース作成作業として、語義文をもとに用例作成作業を行なっている。語義文は、学研の『基本語データベースー語義別単語親密度』と川村よしこ氏の開発した「リーディングチュー太」の情報を参考にする。特に、「リーディングチュー太」に関しては、「チュー太の Web 辞書」に搭載された 8,000 語の単語 ID、例文 ID、見出し語、読み、Note、品詞、語義、語義ごとの例文 27,000 文を活用する。それに対して、本プロジェクトで作成した語法や用例に関する情報を「チュー太の Web 辞書」に搭載することが予定されており、両プロジェクトの緊密な相互連携のもとに作業が進められている。

用例作成は、複数の外部協力者に依頼し、一つの語義に対して作例を 3 つ、コーパスからの用例を 3 つ入れる形で、作業を進めている。具体的には、当該見出し語の難易度を超えないレベルの語だけで用例を作成する作業が行われており、その作業のための支援ソフトが開発されている。

次に、システム開発では、本データベースをウェブインタフェースで検索し、ダウンロードできるようにすることを目的としており、現在、図 1 のプロトタイプ版が完成している。



図1. 辞書検索システムのプロタイプ

図1の辞書検索システムは、「辞書検索」と「ダウンロード」の2種から成る。「辞書検索」は語義と用例その他の情報を表示するもので、検索ボックスに見出し語を入れ、検索を行うと、全体一致と部分一致で検索結果が表示される。「ダウンロード」は、以下の画面の必要な項目にチェックを入れると、それらの一覧がCSVファイルでダウンロードできる。

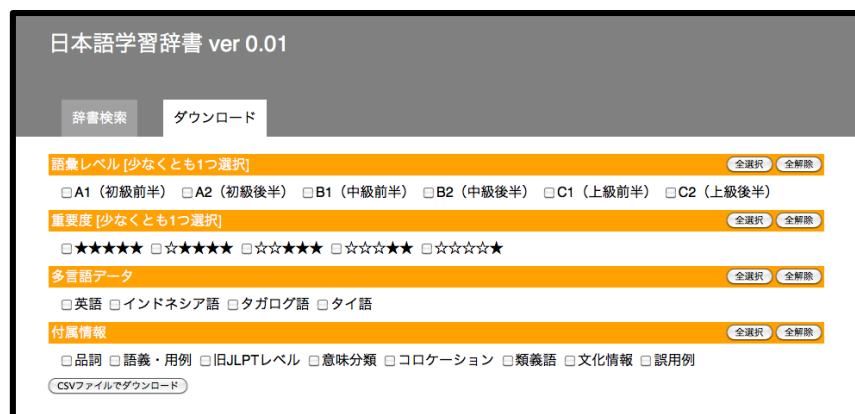


図2. 「ダウンロード」のページ

図3は、「辞書検索」を用いて「花」という名詞を検索した結果である。「花」の語義と用例が見たければ、全体一致の「語義を表示」ボタンをクリックすると、図4の画面が表示される。

日本語学習辞書 ver 0.01

辞書検索    ダウンロード

花

全体一致  
1件

部分一致  
25件

14445	花 (ハナ)	【名詞-普通名詞-一般】 A2 (初級後半) ☆★★★★
711	生け花 (イケバナ)	【名詞-普通名詞-一般】 B1 (中級前半) ☆☆☆★★
2485	開花 (カイカ)	【名詞-普通名詞-サ変可能】 B2 (中級後半) ☆☆☆★★
2770	花器 (カキ)	【名詞-普通名詞-一般】 C1 (上級前半) ☆☆☆★★
3130	花壇 (カダン)	【名詞-普通名詞-一般】 B2 (中級後半) ☆☆☆★★
3299	花瓶 (カビン)	【名詞-普通名詞-一般】 B1 (中級前半) ☆★★★★
3323	花粉 (カフン)	【名詞-普通名詞-一般】 B2 (中級後半) ☆☆☆★★
3324	花粉症 (カフンシヨウ)	【名詞-普通名詞-一般 and 接尾辞-名詞的-一般】 B2 (中級後半) ★★★★★
4107	きゃしゃ (キャシャ)	【形状詞-一般】 C1 (上級前半) ☆☆☆★★
4645	草花 (クサバナ)	【名詞-普通名詞-一般】 B1 (中級前半) ☆☆☆★★
5242	けんか (ケンカ)	【名詞-普通名詞-サ変可能】 B1 (中級前半) ★★★★★
9346	生花 (セイカ)	【名詞-普通名詞-一般】 C1 (上級前半) ☆☆☆★★
9755	線香花火 (センコウハナビ)	【名詞-普通名詞-一般 and 名詞-普通名詞-一般】

図3. 「花」の検索結果

日本語学習辞書 ver 0.01

辞書検索    ダウンロード

花

全体一致  
1件

14445 花 (ハナ) 【名詞-普通名詞-一般】  
A2 (初級後半) ☆★★★★

1. 花道という芸,生け花  
 ○ お花の先生に会った。[作例]

2. 桜の花  
 ○ 花の便り[タヨリ]がとどいた。[作例]

3. 植物の花  
 ○ 友だちの誕生日に花をあげた。[作例]

図4. 「花」の語義と用例

部分一致では、「花火、花器、花盛り、花束、花壇、花畑、花瓶」のように「花」が前に来る語と、「開花、菜の花、生花、線香花火、草花」のように「花」が後に来る語が表示される。この例で明らかなように、部分一致は「ハナ」という単語ではなく「花」という文字と部分一致するもののリストである。さらに、ページを下にスクロールするとその語彙と意味的に関連のある語彙項目が「関連語」として表示される。例えば、「りんご」で検索を行った場合、図5の結果が表示される。



図5. 「りんご」の関連語

図5では、「りんご」が持つ「果物」ないしは「植物」としての特徴から「アーモンド、アップル、アボカド、いちよう」などが関連語として表示されている。関連語は、「分類語彙表」によって定義されており、「りんご」であれば、「体-自然-植物-木本」（「分類語彙表」の情報）が抽出されるので、それと一致したものが表示される仕組みである。

検索の仕様に関して、最近では複雑な検索が可能なシステムが開発され、ユーザーが全体一致（検索キーワードと完全に一致した場合）か部分一致（検索キーワードと部分的に一致した場合）のどちらかを選んだり、部分一致の場合では前方一致（語彙項目の前半部が一致した場合）や後方一致（語彙項目の後半部が一致した場合）を選んだり、発音による検索・表記による検索など、種々の検索方法を指定したりできるものなどが登場している。しかし、ウェブ検索に不慣れた利用者にとっては、こうした細かい検索条件はかえって使いにくいと判断し、本プロジェクトでは、ユーザーが単語を入れて検索ボタンをクリックするだけの単純な仕組みで、システム側が全体一致による検索結果と部分一致による検索結果を同時に表示する仕様となっている。また、検索語の表記に関して、漢語は漢字で、外来語はカタカナで入力する必要があるが、和語の場合はひらがな表記か漢字表記かに関係なく、一回の検索で必要な結果が表示される仕様になっている。和語をひらがなで検索すると、同音異義語の場合、「着る」「切る」のように同音異義語のリストが表示される。

なお、「部分一致」は、名詞や動詞などの複合語を調べるのに便利である。例えば、動詞の一部をキーワードとして入れることで、複合動詞類を検索することができる。一例として、ひらがな入力で「かける」

と入れ、検索をすると、部分一致のところに「追い掛ける、押しかける、腰掛ける、仕掛ける」など「かける」を使った複合動詞が一度に表示される。検索結果の語義が分からない場合は、すでに述べたように「語義を表示」ボタンをクリックすると、語義や用例が表示される（図6）。



図6. 複合動詞の表示と語義・用例

一方、「関連語」は類義語や反義語を調べるのに便利である。例えば、「さわやか」で検索すると、「快い」「すがすがしい」「爽快」「心地よい」などの類義語や、「うっとうしい」「不快」などの反義語を見出すことができる。

## 参考文献

- 李在鎬(2013)「大規模コーパスに基づく語彙リストの検証」マレーシア国際研究集会発表予稿
- 李在鎬(2011)「日本語能力試験の挑戦～新しい日本語能力試験を例に」(国際交流基金事業レポート14)『日本語学』第30巻1号、pp.95-107、明治書院
- 押尾和美・秋元美晴・武田明子・阿部洋子・高梨美穂・柳澤好昭・岩元隆一・石毛順子(2008)「新しい日本語能力試験のための語彙表作成に向けて」『国際交流基金日本語教育紀要』第4号、pp71-86、国際交流基金日本語国際センター。