

学習辞書編集支援データベース作成について

- 『学習辞書科研』プロジェクトの紹介 -

砂川有里子（筑波大学）

QWU00504@nifty.com

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¹。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²。

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た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という状況である。

辞書開発には莫大な費用と労力と時間が必要である。筆者が経験した国語辞書の場合でも、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。国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助っ人が出現し、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。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、そ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当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

¹ 「2009年度海外日本語教育機関調査」結果（速報値）

<http://www.jpf.go.jp/j/japanese/survey/result/index.html>（2011年8月31日）

² 「日本語能力試験受験者の推移」<http://www.jlpt.jp/statistics/index.html>（2011年8月31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。

ようになるのである。

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³。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 28 言語に対応し、その数は今後と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げるものではないという問題がある。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、辞書編集を試みる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本年 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 29 名で、代表者の砂川の他、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。この 29 名の役割については 4 節で述べる。

研究協力チームは 39 名で、その内訳は国内の研究協力者 26 名と海外の研究協力者 13 名である。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 リーディングチュウ太の作成者である川村よし子氏は本プロジェクトの研究協力者でもある。同氏のご厚意により、「チュウ太の Web 辞書」の情報を本プロジェクトに活用する一方で、本プロジェクトの成果を「チュウ太の Web 辞書」に活用するという相互連携が行われることになった。

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、中級から中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。
- (e) 辞書執筆に必要な情報を得るための先行文献や参考書のリスト。

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、中級から中上級レベルの日本語学習者に適切な用例を行う。
- (e) 語法情報整備のためのシステムと語法情報執筆用コーパス活用ツールの開発を行う。
- (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。
- (g) 日本語の語法研究や辞書開発に役立つ参考書・研究論文・辞書の文献リストを作成する。

4. データベース構築チームの構成

日本語辞書編集支援データベース構築チームは、基本方針の策定、作業手順の大枠の決定、各研究グループの統括と各種作業の調整を行う「統括グループ」のもとに、以下の3つの研究グループを配置する。

- (1) 日本語研究グループ：日本語の語法研究、習得研究、対照研究などを行い、その成果をデータベースに取り込む方法を考える。
- (2) コーパス研究グループ：コーパスを活用した日本語研究を行い、その成果をデータベースに取り込む方法を研究する。
- (3) 文献リスト作成グループ：複合辞やコロケーションなどに関する研究文献と辞書作成に役立つ参考文献や辞書のリストを作成する。

以上のうち、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, (5)用例研究班の 5 つの班を配置する。
 また, 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)言語処理班の 4 つの班を配置する。

以上の組織とメンバーをまとめた表を以下に示す。

表 1 日本語学習辞書編集支援データベース構築チームの構成

グループ	班	メンバーおよび専門	
総括グループ (計5名)		砂川有里子	(語彙・文法)
		井上優	(対照言語学)
		アンドレイ・ベケシュ	(コーパス言語学)
		今井新悟	(教育工学)
		李在鎬	(コーパス言語学)
日本語研究グループ	コロケーション研究班 (計3名)	小野正樹	(語彙・文法)
		石田プリシラ	(慣用句)
		長谷川守寿	(計量国語学)
	類義語研究班 (計3名)	井上優	(対照言語学)
		鈴木智美	(語彙)
		砂川有里子	(語彙・文法)
	文法情報研究班 (計4名)	山内博之	(語彙・文法)
		大関浩美	(言語行為)
		小林ミナ	(文法教育)
		橋本直幸	(語彙教育)
	文化情報研究班 (計1名)	築島史恵	(文化教育)
	音声情報研究班 (計2名)	松崎寛	(音声教育)
		五十嵐陽介	(音声学)
	用例研究班 (計3名)	杉本武	(文法)
福永由佳		(コミュニケーション教育)	
パルデン・プラシヤント		(対照言語学)	
コーパス研究グループ	コーパス情報研究班 (計4名)	千葉庄寿	(言語学)
		川端一光	(計量心理学)
		近藤明日子	(日本語学)
		田中牧郎	(日本語学)
	基本語彙研究班 (計2名)	今井新悟	(教育工学)
		李在鎬	(コーパス言語学)
	学習者コーパス研究班 (計2名)	迫田久美子	(第二言語習得)
		仁科喜久子	(学習支援ツール)
	言語処理班 (計4名)	李在鎬	(コーパス言語学)
		宇津呂武仁	(自然言語処理)
		アンドレイ・ベケシュ	(コーパス言語学)
		山本和英	(自然言語処理)
文献リスト作成グループ (計3名)	アンドレイ・ベケシュ	(コーパス言語学)	
	砂川有里子	(語彙・文法)	
	高原真理	(日本語教育)	
			※下線は班長

研究代表者の砂川有里子と研究分担者のアンドレイ・ベケシュは『教師と学習者のための日本語文型辞典』（くろしお出版 1998 年）、砂川有里子は『明鏡国語辞典』（大修館 2002 年）、連携研究者の鈴木智美は『日本語表現活用辞典』（研究社 2004 年）、研究分担者の杉本武は『デイリーコンサイス国語辞典』（三省堂 1991 年）、『デイリーコンサイス漢字辞典』（三省堂 1995 年）の編集・執筆経験を持ち、研究分担者の今井新悟は自己のホームページ上に自作の **Japanese Learner's Dictionary** を公開している。また、研究分担者の井上優は、国立国語研究所 2006-2009 年度プロジェクト「日本語学習者のための日本語用例用法辞書のモデル作成」、パルデンシ・プラシヤントは、同研究所 2009-2013 年度独創・発展型プロジェクト「日本語学習者用基本動詞用法ハンドブックの作成」の研究代表者である。さらに、研究代表者と研究分担者、連携研究者の多くは科学研究費補助金特定領域 2006-2010 年度「代表性を有する書き言葉コーパスを活用した日本語教育研究」（代表者：砂川有里子）に携わり、国立国語研究所によって構築された「現代日本語書き言葉均衡コーパス」を日本語学習辞書や日本語教材に活用する方法を研究している。

辞書開発に必要な情報を蓄積するには辞書研究や日本語そのものの研究が欠かせない。辞書の編集、コーパス言語学、日本語研究、対照研究、習得研究など多方面の領域に豊富な経験を持つメンバーを備えた本プロジェクトは、データベースを構築する過程で各領域の研究をさらに発展させることをも目指している。データベース構築の副産物として辞書研究と日本語研究に新たな知見を提供し、これらの領域の研究を活性化させることも本プロジェクトの重要な目的のひとつである。

5. 年度ごとの研究計画

最後に、本プロジェクトの年度ごとの研究計画を示す。以下に挙げた計画は、当該年度に開始する研究課題を示したものであり、次年度以降にその研究が継続して行われる場合については特に明記していない。ほとんどの場合は次年度以降にも継続される予定であると理解していただきたい。

【平成 23 年度】

- ・基本方針と作業手順の決定
- ・日本語教育基本語彙の選定と語彙表の作成
- ・語法記述の基本フォーマットの検討
- ・既存の用例集の収集と整理

【平成 24 年度】

- ・データベースの基本設計検討
- ・基本語彙の語法記述開始
- ・データ処理環境の整備
- ・データの一部公開（語彙リスト）
- ・HP による広報開始
- ・用例作成の方針と手順の検討

【平成 25 年度】

- ・用例作成開始
- ・コーパス検索システムの構築
- ・データの一部公開（システムとコーパスツール）

【平成 26 年度】

- ・データベース完成
- ・用例付き最終データ公開
- ・データベースの利用普及のためのワークショップ開催

参考文献

北原保雄（2002）『明鏡国語辞典』大修館書店

グループジャマシイ（1998）『教師と学習者のための日本語文型辞典』くろしお出版

姫野昌子（2004）『日本語表現活用辞典』研究社

佐竹秀雄・三省堂編修所（1991）『デイリーコンサイス国語辞典』三省堂

佐竹秀雄・桜井隆・杉戸清樹・三省堂編修所（1995）『デイリーコンサイス漢字辞典』三省堂