

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 4 日現在

機関番号：12603

研究種目：基盤研究（B）

研究期間：2010 ～ 2012

課題番号：22320103

研究課題名（和文） 多言語会話文・語彙データベース構築と異文化交流におけるその活用に関する研究

研究課題名（英文） Study on multilingual sentences and vocabulary database construction and their utilization in cross-cultural communication

研究代表者

萬宮 健策 (MAMIYA KENSAKU)

東京外国語大学・大学院総合国際学研究院・准教授

研究者番号：00403204

研究成果の概要（和文）：言語横断的な会話文および語彙データベース作成が、本研究の最大の目的であった。言語によってデータベースの大きさにばらつきは出たものの、多言語間での語彙や短文の簡易検索や、一部の言語における映像資料との連携などが可能となった。日本では商業ベースで注目されることが少ない、いわゆる LCTL (Less Commonly Taught Languages) の外国語教育への応用も含め、これまでにあまり例を見ない使用に耐えるデータベースが構築され、当初の目的は概ね達成できたと考える。

研究成果の概要（英文）：The main focus of the research is the construction of the database in multi languages, especially in the less commonly taught languages in Japan. The database of the basic sentences and basic vocabulary in LCTL including Urdu, Thai and Arabic etc. was constructed and the basic vocabulary or sentences can be searched between any of the two languages with sounds and images, for example, from Thai to Spanish or Swahili to Urdu. etc. The database is in the practical use in the classes of the languages in the universities.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010 年度	5,200,000	1,560,000	6,760,000
2011 年度	4,400,000	1,320,000	5,720,000
2012 年度	2,600,000	780,000	3,380,000
年度			
年度			
総計	12,200,000	3,660,000	15,860,000

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：多言語資源、異文化交流

1. 研究開始当初の背景

現代社会において、言語の運用能力を必要とする場面は非常に多いが、英語などのごく一部の言語をのぞけば、必ずしも常に言語に堪能な通訳等が入手可能なわけではない。たとえば、地震や水害などの自然災害が発生した際に、現場へ赴き緊急支援を行うということを想定した場合、支援者と被災者との間で

のコミュニケーションがスムーズにいかないということも少なくない。

本研究の前段階と位置づけられる挑戦的萌芽研究（課題番号 19652058、『災害救援者教育のための多言語会話文・語彙データベース構築に関する基礎的研究』、研究代表者：堀一成）では、アラビア語、タイ語、ウルドゥー語の 3 言語で、災害救援という限定的な

場面を想定した語彙や会話文のデータベース構築とそのデータベースを活用するソフトウェア開発を行ってきた。ウルドゥー語データベースについては、2006年10月にパキスタン北部で発生した地震をきっかけとして、被災者救援に必要な基本語彙や会話文、またパキスタンに関する基本情報をあわせてウェブサイトで公開したことが発端となって上記挑戦的萌芽研究に発展した。

しかし、災害救援に限定すると、どうしても基礎語彙も会話も偏りが生じるため、また、せっきくのデータベースの使用範囲が自ずと限定されてしまうため、本研究では、それを拡充するとともに、言語の数についても拡充することを目指した。日本国内では、商業ベースに乗りにくい言語への注目度がどうしても下がることを踏まえ、本研究では、あえてそうした言語を中心に各種データベースの構築を行うこととし、同時にそのデータベースを外国語教育や異文化理解に応用することを目指すこととした。日本国内ではふだん接する機会が少ない言語であるため、可能な限り、その語彙（および仕草）が指すものを理解する一助とすべく、映像や写真などを用いる必要性もあると考えた。

2. 研究の目的

本研究は、平成19年度科学研究費補助金（挑戦的萌芽研究）で得られた成果および克服すべき課題を踏まえ、その研究を発展させることを目的としている。すなわち、各言語における基礎語彙および基本文データベースの構築および拡充である。以下で指摘するように、語彙が持つ意味は多岐にわたることが多く、複数言語間でその意味が一致することはまれであるが、それでも、似た意味を持つ語彙が複数言語間で簡易辞書的に得られれば、外国語としてその言語を学ぶ際の補助教材としても活用しうる。その際、本研究では、上述のとおり、商業ベースで開発が難しいと思われるいわゆる Less Commonly Taught Languages に焦点を当てる。また、本研究は、その成果を教育に応用することも目指している。大阪大学外国語学部や、東京外国語大学をはじめとする、外国語教育の拠点と言える場所でのデータベースを活用した教育への応用も目指す。

上記データベースを効率よく活用するためのソフトウェア開発も大きな目的の1つである。先述の挑戦的萌芽研究では、被災地で用いることを想定した簡易翻訳機能を有するソフトウェアを開発した。本研究においては、簡易翻訳機能は日本語と特定の外国語間だけでなく、任意の複数言語間での翻訳が可能なように機能の拡張を目指し、より汎用性を持たせることを目指す。また、映像や音声データとの連携も本研究の重要な目的の

1つである。構築されるデータベースは、原則として随時追加が可能なように、ウェブサイト上での公開、使用を前提とするため、その利点を活かすべく映像や写真をはじめとしてマルチメディア素材を含むデータベースを構築することとした。

今さら指摘する必要もないが、研究期間中療護も、随時データベースは更新し、任意の複数言語間での辞書として、あるいは各言語の学習事典的性格を持たせたいと考える。

3. 研究の方法

本研究の柱は、各言語の会話文、基礎語彙データベース構築・拡充と、そのデータベースを活用するアプリケーション開発である。

(1) データベース構築・拡充

まず現存しているデータベースがある言語に限定し、各言語の担当者が、本研究で補強すべき項目が何かを明確にし、その部分を重点的にデータベース構築を開始する。日本語からの訳だけでなく、多言語間の相互変換をも前提に、基本となる日本語によるデータベースを更新し、各言語担当者がそれに応じたデータベースを構築していくという作業を中心とする。

既存のデータベースが構築されていない言語については、既存のデータベースを元に、対応する語彙、会話文データベースをまず構築する。その後、その言語に特徴的な、欠かすことができない文化的な側面等も考慮に入れつつ、その言語独自のデータベース構築を目指す。ただし、任意の複数言語間での簡易辞書機能を活かすため、基礎語彙、基礎会話文については、可能な限り一対一対応で、入力を進めていく。

平成23年度以降は、構築したデータベースを実践的にさまざまな場面や授業の場で試用し、本格運用に向けた課題を探り、かつその課題克服を目的としてデータベースの完成度を高めていく。必要に応じて、ネイティブ・スピーカーからのアドバイスを得、あるいは実地での試行を繰り返しながら、実用的なデータベース構築及びそのデータベースを最大限に生かすことができるソフトウェアの改良・開発を行い、より高い完成度を目指す。また、当初の計画に含まれていなかった言語についても、可能であればデータベースに含め、データベース全体としての価値重要性を増すことを考える。

学習事典としての正確を持たせるため、日本国内では馴染みのない事物については、その語彙に対応する画像や映像を可能な限り収集し、見てわかるデータベース構築を目指す。

(2) アプリケーション開発

(1) で構築されたデータベースを、活用するアプリケーションを開発することは、本研究の最大の課題である。前出の挑戦的萌芽研究で、アプリケーションの大枠が完成していたため、本研究では、その際課題として残った、映像・画像データの活用、任意の多言語間での簡易翻訳機能、非ローマ字言語における任意のフォントによる表示機能を盛り込んで、新たなアプリケーションを開発する。

4. 研究成果

我々の研究グループが構築した災害救援者教育のための言語資源の内容について説明する。基本的なデータについては、多数の言語を横断的に検索・比較できるものとなるよう作成している。これは、言語教育の教材としてだけでなく、対照言語学・言語類型論を研究する際の基礎データとして意義もある言語資源としたい、との思いからである。多言語間の単語や文の対応で、各言語の対象とする語彙範疇は単純に一対一対応するものでないことは明らかであるが、あえて単純に一対一対応の表形式まとめることで、多言語間の対応関係調査を容易にするものを提供できるのではないかと考えている。

ウルドゥー語資源の規模は、基本的な語彙文例データが 188 例、救援に際して考えられるシチュエーション別に語彙文例を集めたデータが、洪水関係 700 例、伝染病関係 1, 125 例である。データは Excel 表データとして保存されている。多少の重複データはあるが、総計 2, 013 例の言語データが完成した。基本語彙・会話の 188 例については、大阪大学の外国人特任教員の協力を得て、発音と発話に対応するジェスチャーを確認できる MP3 音声データ、WMV 形式の動画データも作成した。

データ番号	日本語	ウルドゥー語	英語
1	私の名前を教えてください	اپنی نام بتائیے	What is your name?
2	私の名前が何ですか	میرا نام کیا ہے؟	What is my name?
3	私の名前が何ですか、わかりましたか？	میرا نام کیا ہے؟ سمجھا؟	What is my name? Did you understand?
4	私の名前が何ですか、わかりましたか？	میرا نام کیا ہے؟ سمجھا؟	What is my name? Did you understand?
5	私の名前が何ですか、わかりましたか？	میرا نام کیا ہے؟ سمجھا؟	What is my name? Did you understand?
6	私の名前が何ですか、わかりましたか？	میرا نام کیا ہے؟ سمجھا؟	What is my name? Did you understand?
7	私の名前が何ですか、わかりましたか？	میرا نام کیا ہے؟ سمجھا؟	What is my name? Did you understand?

図1) ウルドゥー語データの一部 (元データは、エクセルによる)

これらの成果については、以下5. で示したとおりの業績・成果として、発表した。

しかし、これで一連のデータ構築が完成したとは考えておらず、今後も引き続き構築されたデータベースを活用することを検討していく予定である。

以下に、本研究で構築されたデータベースおよびそれを用いた成果の一部を示しておく。



図2) 構築された映像データベースの一部。音声や、図1で示した語彙、会話文データベースと連動している



図3) 構築されたデータベースを用いての授業の一コマ。構築されたデータベースに含まれる基礎会話文を用いて応用練習をしているところ

データ番号	日本語	スペイン語	スワヒリ語	タイ語
110	109 喉が痛い	Me duele la garganta.	koo inauuma / naumwa na koo	เจ็บคอ
111	110 風邪を引いた	Estoy resfriado	nimepata mafua	เป็นหวัด
112	111 喉が出る	Tengo tos.	ninatoka vikohozo / ninakohoa	ไอ
113	112 熱がある	Tengo fiebre.	nina homa	มีไข้
114	113 だるい	Me siento flojo.	ninahisi / ninaona utepetevu	เหนื่อย
115	114 眠れない	No puedo dormir.	sipati usingizi	นอนไม่หลับ

図4) 多言語間での簡易辞書の元となる、語彙データベースの一部。

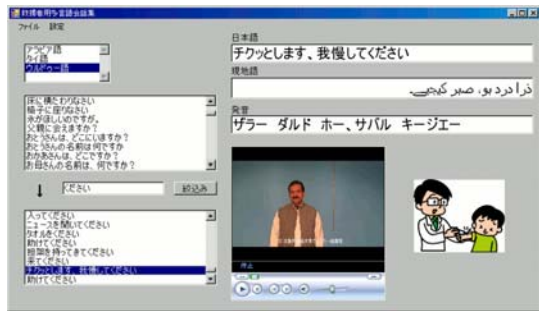


図5) 本研究で開発された、データベースを利用するためのアプリケーション。画像や映像も表示することができ、また各言語に固有のフォントも任意のものを用いることができるようになっている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 17 件)

- ① 宮本マラシー、味を表すタイ語表現における比喩、大阪大学『言語文化研究』、査読有、39 巻、2013、125-148
- ② 竹村景子、スワヒリ語変種記述調査報告 (2)ーキベニ変種およびキドティ変種基礎語彙 600 語ー、スワヒリ&アフリカ研究、査読有、24 巻、2013、50-72
- ③ 萬宮健策、ウルドゥー語におけるヴォイス、東京外国語大学語学研究所論集、査読有、17 巻、2012、184-199
- ④ 米田信子、スワヒリ語のアクセントーアクセントフレーズを中心にー、自立調和的視点から見た音韻類型のモデル、査読無、2012、139-148
- ⑤ So Yamane、Laa Musaaviyaat se Janam lene waalii Aawaazen: Unniisviin Sadii ke Aaghaaz men "Urdu Rasm al-Khat" kaa Tahqiiqii Jaa'iza, Tahqiiq (原文ウルドゥー語、邦訳は「から生まれた音：19 世紀初頭の『ウルドゥー語正書法』の研究」)、査読有、198 号、2011、1-48
- ⑥ 米田信子、Swahilization of Ethnic Languages in Tanzania: The Case of Matengo, African Study Monographs、査読有、31(3)、2010、139-148
- ⑦ 鈴木慎吾、山崎直樹、堀一成、多言語資源作成のための統語・オントロジー情報を付与するアプリケーションの開発、第 9 回情報科学フォーラム論文集第 4 分冊、査読有、119-122、2010

[学会発表] (計 20 件)

- ① 堀一成、竹原新、石島悌、片桐真子、平松初珠、萬宮健策、災害救援者教育のための言語資源構築、言語処理学会第 19 回全国大会、2013 年 3 月 13 日、名古屋大学

② 平松初珠、石島悌、萬宮健策、堀一成、災害救援教育者用 Web アプリケーションの開発、産技研研究発表会、2012 年 11 月 15 日、大阪府立産業技術総合研究所

③ 竹村景子、スワヒリ語諸方言調査報告(1)ーチャアニ変種(ザンジバル島北部県北部 A 郡)について、日本アフリカ学会第 49 回学術大会、2012 年 5 月 26 日、国立民族博物館

④ 堀一成、多言語資源活用のためのアプリケーション開発と災害環境下での音声明瞭度の改善、言語処理学会第 18 回全国大会、2012 年 3 月 16 日、広島市立大学

⑤ 堀一成、Multi-language Parallel Corpora and XML Annotation Tools, Osaka Symposium on Digital Humanities 2011、2011 年 9 月 13 日、大阪大学

[図書] (計 8 件)

- ① 萩田博、萬宮健策、白水社、ニューエクスプレス ウルドゥー語、2012、150
- ② 塩田勝彦、竹村景子、小森淳子、米田信子、品川大輔、神谷俊郎他 14 名、溪水社、アフリカ諸語文法要覧、2012、301

[その他]

ホームページ等

<http://mlab.celas.osaka-u.ac.jp/>

6. 研究組織

(1) 研究代表者

萬宮 健策 (MAMIYA KENSAKU)

東京外国語大学・大学院総合国際学研究院・准教授

研究者番号：00403204

(2) 研究分担者

堀 一成 (HORI KAZUNARI)

大阪大学・全学教育推進機構・准教授

研究者番号：80270346

松本 健二 (MATSUMOTO KENJI)

大阪大学・言語文化研究科(研究院)・准教授

研究者番号：00283838

(H22：連携研究者)

石島 悌 (ISHIJIMA DAI)

地方独立行政法人大阪府産業技術総合研究所・業務推進課・主任研究員

研究者番号：80359398

平松 初珠 (HIRAMATSU HATSUMI)

地方独立行政法人大阪府産業技術総合研究所・業務推進課・主任研究員

研究者番号：30416250

片桐 真子 (KATAGIRI MAKI)

地方独立行政法人大阪府産業技術総合研究所・製品信頼性課・主任研究員

研究者番号： 50359379
(H23：連携研究者)

(3)連携研究者

米田 信子 (YONEDA NOBUKO)
大阪大学・言語文化研究科 (研究院)・教授

研究者番号： 90352955
(H22→H23)

藤家 洋昭 (FUJIE HIROAKI)
大阪大学・言語文化研究科 (研究院)・准教授

研究者番号： 90283837

山根 聡 (YAMANE SO)
大阪大学・言語文化研究科 (研究院)・教授

研究者番号： 80283836

宮本 マラシー (MIYAMOTO MARASRI)
大阪大学・言語文化研究科 (研究院)・教授

研究者番号： 00200212

竹原 新 (TAKEHARA SHIN)
大阪大学・言語文化研究科 (研究院)・准教授

研究者番号： 20324874

竹村 景子 (TAKEMURA KEIKO)
大阪大学・言語文化研究科 (研究院)・准教授

研究者番号： 20252736

高橋 明 (TAKAHASHI AKIRA)
大阪大学・言語文化研究科 (研究院)・教授

研究者番号： 50187994

近藤 久美子 (KONDO KUMIKO)
大阪大学・言語文化研究科 (研究院)・教授

研究者番号： 90273739

長谷川 信弥 (HASEGAWA SHINYA)
大阪大学・言語文化研究科 (研究院)・教授

研究者番号： 20228448