

災害救援者教育のための言語資源構築

堀 一成, 竹原 新†, 石島 悌†, 片桐 真子†, 平松 初珠†, 萬宮 健策★

大阪大学 全学教育推進機構, 言語文化研究科†,

大阪府立産業技術総合研究所†, 東京外国語大学 大学院総合国際学研究院★

hori@celas.osaka-u.ac.jp, takehara@lang.osaka-u.ac.jp†,

{ishijima, katagiri, hiramatsu}@tri-osaka.jp†, k_mamiya@tufs.ac.jp★

1 はじめに

大阪大学と大阪府立産業技術総合研究所を中心とする多言語資源研究グループは、主に言語教育に役立つことを目的とした様々な言語資源の構築を進めてきている [3]。その中でも、近年、特に災害救援者教育のための言語資源構築に注力している。本報告では、上記の言語資源構築作業の成果を紹介する。

大規模災害が発生した際には、救援者の派遣や物資の調達など様々な面からの国際的支援が行われる。2011年3月の東日本大震災において、諸外国から支援があり、2010年1月ハイチの地震など、海外で大規模災害が発生した際には、日本も救援活動に参加している。しかし、その救援現場では、言葉の壁が大きな問題となっている。

2005年10月のパキスタン北部地震に際して、旧大阪外国語大学における学生主体の救援者用のウルドゥー語（パキスタンの国語）会話集作成がおこなわれた。この活動が、本研究の着想の源流である。パキスタンの災害現場で、救援活動を行う作業者のかかえる言葉の壁の問題を少しでも減らすため、日本国内にいるウルドゥー語の専門知識を有する者のできるボランティア活動のアイデアとして、学生の中から提案があったものである。このとき、本報告の分担執筆万宮を含むウルドゥー語専攻教員もサポートすることで、救援に際して役立つと思われる言語データを作成した。さらに、現地新聞の翻訳、地図に日本語情報を追加記入したものなども作成し、現地活動者へ送って利用してもらった (図1参照)。

この学生らの自発的な活動の成果を継承発展させるため、2007年より本研究グループが中心と



図1: 旧大阪外国語大学のパキスタン震災救援協力について伝える朝日新聞社 Web ページ

なり、災害時に役立つ、また将来災害救援の場で活躍する可能性のある日本語話者（主に大阪大学外国語学部や言語文化研究科所属の学生のうち意欲のある者）に対して教育を行う際の基礎データとなりうる言語資源を構築すべきと考え、開発を進めている。

我々の構築する多言語資源構築は、社会還元を意識した外国語教育の実践方法であると自負している。対象とする主な言語は、日本国内で教育される機会の少ない、LCTL (Less Commonly Taught Languages) であり、また得られた成果は、広く公開している。最も早くから着手していたウルドゥー語データは、早期に大学のウルドゥー語専攻 Web ウェブサイトで公開し (図2参照)、救援ボランティア活動をする団体などから高い評価を受けている [6]。また、これまでに構築済みのデータは、本研究グループの Web サイトで整理・

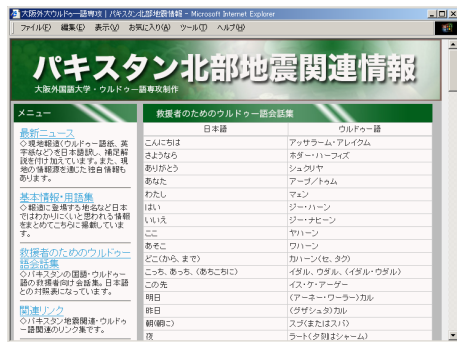


図 2: 大阪大学言語文化研究科ウルドゥー語専攻提供の救援者会話集 Web ページの一部

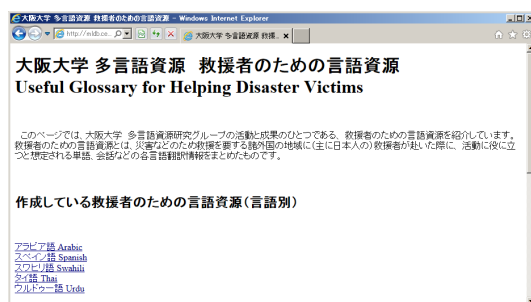


図 3: 本研究グループが言語資源を公開している Web ページの一部

公開している (図 3 参照) [5].

構築した災害救援のための言語データを現実の災害現場で活用することを想定し、過酷な環境でも利用できる堅固な Windows ノート PC に搭載するためのアプリケーションも併せて開発している [1], [2]. 専用ソフトウェア開発したのは、当該 PC の画面サイズや操作性の制約のため、単に文字データの一覧を表示するだけでは災害現場で使いづらいものとなるため、情報提示方法に工夫が必要だったからである。

また、当該 PC の内蔵スピーカーの品質が高くないことから、音声データ提供方法にも工夫する必要がある、蓄積した音声データが騒音下でどのように聴取されるかを調べる聴取実験を行った。そして、災害現場で想定される騒音をペルシア語音声データに重畳し、明瞭度を損なわないためには、どのような音声データ処理が適切なのかを検討した [4].

2 災害救援者教育のための言語資源

2.1 構築した資源の概要

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。

前述したとおり、言語資源構築は小規模のウルドゥー語単語会話集から始めた。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。

2.2 ウルドゥー語資源

ウルドゥー語データについては、研究開始時において試作されたデータを元に、データの拡充・検証作業を行い、構築を進めた。2013 年 1 月現在で、完成し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。その一部を図 4 に示す。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。図 5 に動画再生時のキャプチャ画面を示す。

また、図 6 は構築したウルドゥー語資源と後述するアプリケーションを、ウルドゥー語授業の教材として受講生に提示し、解説している様子である。このように、構築済み言語資源は順次教材として教育現場での活用を開始している。

	A	B	C	D
1		この薬を使ってください	یہ دوا استعمال کیجیے	イェ ダワー イステマール キージェー
2		私の腕をつかみなさい	میرا بازو پکڑو	メーレー バーズー バクリエー
3		私の言うことがわかりますか？	میری بات سمجھتے ہیں؟	メーリー バート サマジュデー ハエン？
4		私の言うことが、わかりましたか？	میری بات سمجھ؟	メーリー バート サマジュデー？
5		食べますか？	کیا آپ کھاتے ہیں؟	キャー アーブ カーエーン ゲー？
6		飲みますか？	پینتے ہیں؟	ピーエーン ゲー？
7		見えますか？	دیکھ سکتے ہیں؟	デーク サクデー ハエン？

図 4: ウルドゥー語データの一部



図 5: ウルドゥー語動画データのキャプチャ画面

	A	B	C	D
1	日本語	ヨミ	アラビア語	
2	頭	ラアス	الرأس	
3	顔	ワジュフ	الوجه	
4	眼	アイン	العين	
5	耳	ウズナ	الاذن	
6	鼻	アンフ	الأنف	

図 7: アラビア語データの一部



図 6: 構築したウルドゥー語データとアプリケーションを授業中に紹介している様子

2.3 その他の言語資源

2.3.1 アラビア語データ

アラビア語データについては、2008 年度に行った海外調査の結果も踏まえ、特に医療関係用語の充実を意識し作成作業を行った。現在救援医療時に必要な会話を 130 例、単語を 330 例が構築できている。ただし、カタカナ読み情報が付与されているのは 33 単語のみである。データの一部を図 7 に示す。

2.3.2 スペイン語・スワヒリ語・タイ語データ

タイ語・スペイン語・スワヒリ語は、ウルドゥー語基本語彙文例と同内容のものそれぞれ 188 例を作成した。いずれも各言語をカタカナ読みする情報も提供している。データの一部を図 8 に示す。

3 おわりに

本稿では、災害救援者の教育に役立つことを目的とした 5 言語の語彙・会話文言語資源を構築したことを報告した。言語資源の規模は言語によって異なるが、構築対象とした言語はウルドゥー語・アラビア語・スペイン語・スワヒリ語・タイ語である。

今後の研究展開について述べる。言語資源構築研究の最終目標は、大阪大学外国語学部が専攻語（日本語を除く）としている 24 言語に対し、現時点のウルドゥー語資源と同規模のデータを完成することである。これを公開し、有事の際に利用できるようにすることで、世界中のほとんどの地域に対する社会貢献ができる成果となるのではないかと考えている。また、アプリケーション開発においては、災害現場における操作性の向上やス

	A	B	D	F	J
	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	咳が出る	Tengo tos.	ninatoka vikohozi / rinakohoa	ไอ
113	112	熱がある	Tengo fiebre.	nina homa	มีไข้
114	113	だるい	Me siento flojo.	ninahisi / ninaona utepetevu	เมื่อย
115	114	眠れない	No puedo dormir.	sipati usingizi	นอนไม่หลับ

図 8: スペイン語・スワヒリ語・タイ語データの一部

スマートフォン対応などのアプリケーション開発、音声明瞭度向上法のさらなる研究を進めていきたいと考えている。

謝辞

本研究は、科学研究費補助金 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) を元に推進し、同基盤研究(B) 課題番号:22320103 『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』(研究代表者:萬宮 健策) の補助を受け研究を継続しているものである。

言語資源構築にあたって各言語担当教員や作業者の取りまとめを担当された、大阪大学言語文化研究科(スペイン語) 松本健二 准教授に特に謝意を表す。また、データ整備や関連実験実施に協力してくださった、大阪大学言語文化研究科の教員・大学院生各位、外国語学部(特に音声聴取実験に協力してくれたペルシア語専攻)の学部学生の方々に謝意を表すものである。

参考文献

- [1] 平松 初珠, 石島 悌, 萬宮 健策, 山根 聡, 堀 一成. 多言語会話文、語彙データを利用した災害救援者教育用アプリケーションの開発. 情報処理学会 第 72 回全国大会講演予稿集 第 4 分冊, pp. 469 – 470, 2010.
- [2] 平松初珠, 石島悌, 萬宮健策, 堀一成, 松本健二. 災害救援者教育用 Web アプリケーションの開発. 平成 23 年度大阪府立産業技術総合研究所研究発表会要旨集, pp. 56–57, 2011.
- [3] 堀一成, 竹原新, 上原順一, 小島一秀, 藤家洋昭, 萬宮健策. GDA に基づく統語情報付与 XML 化多言語平行資源の構築. 言語処理学会 第 17 回年次大会発表論文集, pp. 348–351, 2011.
- [4] 堀一成, 石島悌, 片桐真子, 平松初珠, 竹原新. 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 回年次大会発表論文集, pp. 911–914, 2012.
- [5] 大阪大学 多言語資源研究グループのホームページ
<http://mlldb.celas.osaka-u.ac.jp/>.
- [6] 山根聡. 災害救援者支援のための会話集等作成について. 大阪大学世界言語研究センター論集, Vol. 1, 第 1 号, pp. 217–225, 2009.