

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 24 年 6 月 19 日現在

機関番号：34412

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21500113

研究課題名（和文）医療福祉分野における日本語学習・運用支援のための対話例集構築支援の研究

研究課題名（英文）A study on constructing example sentence database in nursing-care context for Japanese language learning

研究代表者

竹内 和広（KAZUHIRO TAKEUCHI）

大阪電気通信大学・情報通信工学部・准教授

研究者番号：20440951

研究成果の概要（和文）：

医療福祉分野に従事する外国人の日本語学習・運用の困難性を軽減するため、医療福祉分野の対話例データベースを構築するための対話モデルを構築し、それに関わる知識記述およびその知識に基づいた多様な文の処理技術を開発することができた。また、それらの技術を支援として用いて、日本語教師と共に外国人の日本語学習に役立つ対話例集を試作し、その対話例を基本とした協調的な例文蓄積・学習の支援ソフトウェアの開発を行った。

研究成果の概要（英文）：

In order to facilitate the difficulties in learning and using the Japanese language that professional foreigners in medical and health welfare fields are facing, this study built software modules to support to construct cooperative example sentence database for Japanese language teachers and learners. As bases of the software, we proposed a dialogue model in nursing-care context and developed language processing techniques coping with a variety of sentences found in the context by utilizing the relevant semantic representation of the domain specific knowledge.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	900,000	270,000	1,170,000
2010 年度	900,000	270,000	1,170,000
2011 年度	800,000	240,000	1,040,000
総計	2,600,000	780,000	3,380,000

研究分野：総合領域

科研費の分科・細目：情報学、メディア情報学・データベース

キーワード：コーパス・医療福祉対話・日本語教育支援

1. 研究開始当初の背景

世界に先駆けて高度高齢化社会に向かう我が国において、外国人労働者の医療福祉分野への受け入れが求められるようになってきている。研究開始当初は外国人の医療福祉分野における就労黎明期であり、外国人看護師・介護士の日本語教育プログラムが十分に整備されていない状況にあった。具体的には、外国人労働者が勤務先でサービス受領者あ

るいは同僚との間で言語コミュニケーションを取る際の困難が予想されていたものの、どのような日本語学習が必要なのか、どのような日本語運用の支援が適切であるのか明らかとはいえなかった。また、特定機関での日本語研修の後、外国人看護師・介護士が所属する施設は日本全国に及ぶため、学習環境の保持が難しい場合もあり、継続的な学習支援のあり方についても課題があった。

異文化における言語コミュニケーションでは、日本人が相手の国の文化や言語を習得することも対応の選択肢の一つであるが、医療福祉分野のサービス受領者の多数を占めるのは高齢者であるため、そのような学習や努力を一律に求めることは難しいといえる。そこで、そのような問題を克服するため、外国人労働者の日本語学習や日本語運用をコンピュータにより支援することが期待されるが、医療福祉分野のサービス場面における外国人とサービス受領者のとのコミュニケーションに資する基盤資源整備は不十分な状態にあり、当該分野における日本語学習資源構築とその支援のための研究が必要となる背景があった。

2. 研究の目的

本研究は、外国人がコンピュータ支援により医療福祉分野の日本語を学習・運用する際に役立つ対話例集の作成・収集支援環境の整備を行うことが目的である。具体的には、登録しようとする対話例文の適切性、非文法性を計算文法、および、その他の意味処理機構に基づいて評価し、外国人の医療福祉分野従事者が日本語を学習する上で役立つ、かつ実際の医療福祉コミュニケーションを支援するシステム構築にも役立つような言語資源の構築支援環境の研究・開発を行う。

3. 研究の方法

研究開始時には、看護・介護分野のドメイン分析のため関係各所にインタビューを行い、また、関連する会合に参加することにより当該分野における日本語教育の課題調査を行う。研究初年度以降は、日本語教育を仕事あるいはボランティアで行っている複数人の日本語教師の協力を得ながら、特定分野の談話に関わる知識を日本語学習教材に応用する上での課題を適宜整理しつつ研究を進める。

また、目的志向性の高い文の処理技術を開発するために、既存の日本人向け医療福祉分野の実践テキストを調査・整理し、医療福祉分野における対話モデルの構築を行う。このことにより、外国人労働者の日本語学習・運用に十分な資源を体系的に定義付けると同時に、該当ドメインの語彙やその意味特徴を適切に把握・制限する。その結果に基づいて計算文法および意味処理を開発するとともに、それらの処理の基盤となる計算資源を構築する。

さらに、日本語学習・運用支援技術を開発するために、上記の方法により得られた知見および基盤技術に基づいて、実際に日本語学習者向けの教育教材となる対話例集を構築し、その学習支援システムを試作する。その際、医療福祉分野に従事する外国人への日本

語教育、教材作成の経験をもつ複数の日本語教師を研究協力者として迎え、彼ら彼女らの知見を基礎技術・知識記述に盛り込むとともに、試作システムに対する意見を求めて実践的に研究開発を進める。

4. 研究成果

(1) ドメイン知識体系化による意味処理機構

① 介護・看護等の医療福祉分野における日本人向け教科書を資料として、医療福祉場面における行動を名詞-動詞の組み合わせに整理して表現し、それらを介護・看護場面としてクラスタリングすることにより、名詞・動詞の言語表現から当該行動が生じうる場面を推定する知識体系に整理した。この整理により、対話が行われる場面、行動に関連して生じる問題や症状の文脈をトイレやベッドといった単語から候補限定する道具立てを得ることができ、対話例を作成する上で、関連文脈の言語表現を収集する基盤が構築できた。

② 上記の医療福祉場面における行動知識とは別に、事前調査の結果から特に重要性の高かった、心身の状態・状況に関する対話例の表現を整理する意味構造を次のような基本的な組<評価対象、評価表現>に基づいて整理した。

評価対象については、人体に関わるものだけではなく、医療福祉の場面において日常的に接するガーゼやタオルなどの機具に関しての知識も記述した。評価表現に関しては感性極性を導入し、ドメイン依存シソーラスの体系化に基準を与えた。感性極性は、特定の評価対象について、対立する評価語提示するSD法によるアンケート等によく利用される意味表現である。ドメイン依存シソーラスは様々な計算機処理において意味的な中間言語（表現）として用いることができる、言語の意味処理を支える基盤資源である。

評価表現は文脈依存性が高いため、対象ごとにどの感性極性が有効であるかを定義しなくては、多様な表現の類似性扱うことは困難である。そこで、アンケート分析に関わる書籍を網羅的に調査し、できるだけ多くの評価的意味を表現する約400軸の感性極性を整理した。その上で当該ドメインでの評価対象を評価する感性極性に関して標準的な文脈を、インターネット上の大量のデータから、肌感覚、味覚、苦痛といった様々な文脈において否定・肯定感性に関わる表現を網羅的に整理するテキストマイニング手法を開発し、当該対象の評価表現の拡張候補を抽出した。そのようにして得られた表現を適宜追加したドメイン依存シソーラスを拡張した例を図1に示す。図1は身体部位「顔」に対する症状や状況を隣接ノードとして示し、さらに

そういった症状や状況といった対象を評価する表現を隣接ノードとして拡張したもののうち直接隣接ノードに限定してグラフ表示した例である。また、症状や状況に対する評価表現の拡張は、後述の(3)で示すように、例文作成の支援となるよう、上記開発技術による表現候補抽出だけでなく、人手による検討を行っている。

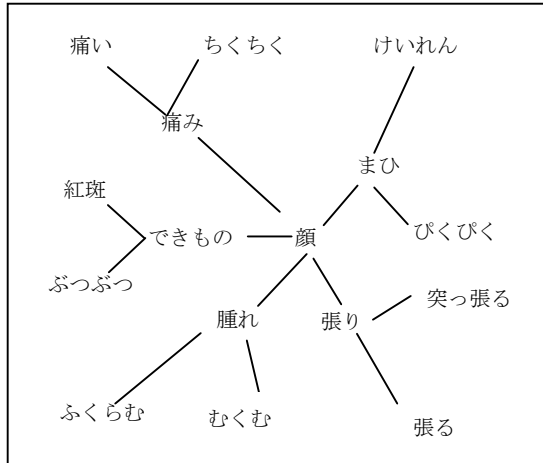


図1. 「顔」を中心とした表現拡張の例

③当該ドメインにおける対話例を知識との相関により整理する対話モデルを構築した。その際、知識・場面・行動との対応付けを考慮し、ドメインにおける特色的な対話間の発話目的を明確化するように発話機能を整理した。

例えば、①で整理したベッドで「体を起こす」という場面の行動知識に対して「体を起こしましょうね」という発話は「促し」という発話機能をラベル付与することにより整理する。このような発話機能群をドメインに特徴的な目的に基づいて整理・定義することにより、①②で構築した知識と対話対が標準的な応答や発展的な応答といった多様性を包含して体系的に整理・開発することが可能となる。この体系化に際しては、研究方法で述べたように、医療福祉分野に従事する外国人に対しての日本語教育経験をもつ複数の日本語教師が関連する様々な教材を分析した上で整理した。

整理した発話機能の例を表1に示す。表中の関連知識欄が、①②の知識とどのように関連づけられているかを示し、発話中の(+)(-)表示は対話の文脈の感性極性が肯定的か否定的かを示している。特に語彙の極性とも関連付けて整理したのは、極性が応答の選択に関わるからである。

以上①②③の成果は医療福祉ドメインの日本語学習教材を構築支援する上での計算処理上の意味処理資源を具体的に整理しただけでなく、それぞれの構築に採用・開発した手法は、今後、他分野のドメイン分析や意

味処理資源の開発にも適用・応用が可能である。

表1. 対話例の発話機能対の例

関連知識	第一発話	基本応答	発展応答
行動	提案	受諾/拒否	訴え
行動	申し出	受容/拒否	訴え
行動	促し(+)	受諾/拒否	訴え・弁明
行動	依頼	受諾/拒否	訴え・弁明
状況・状態	訴え(-)	受容/拒否	提案・助言
状況・状態	報告(+)	承認/(否認)	同調

(2) 多様な表現を扱うための計算文法の整備

本研究課題のような医療福祉分野での発話を扱うためには、既存の言語解析ツールでは解析結果が頑健ではなく、表現や構文の多様性の面でも問題がおこる。また、本研究課題の目的は学習者や学習者をサポートする人々に対して、当該の文の特徴についての補助情報を提供することにあるため、日本語の学習文法の考え方を考慮にいれなければならない。

以上の要件に対して、本研究では計算文法として文型パターンを採用し、既存の言語処理モジュールとの整合性をとる形による言語解析の方向性で研究・開発を進めた。その中で最も特色ある成果は、文の構文構造の特徴を形作る文型パターンを機能表現中心に定義した点にある。機能表現には様々な定義が存在するが、本研究では次のような文の多様性に対応させることを主眼に既存の機能表現研究や日本語教育教材を参考に機能表現とそれに基づく文型パターンの整理を行った。

「あまり食欲が {ありません/ないよ/なくて}」

その結果、上記のような「Xがある/ない」を表現する日本語文型パターンを機能表現である「が」「ありません」などに基づいて整理しただけではなく、例えば、敬意表現を含む丁寧体と普通体の発話といった機能表現に基づく文体は違うものの、構文的パターンは同じといった文の文型的特徴を解析するモジュールを開発することができた。このことにより、実際の発話場面に応じた実践的な発話例を計算機的に扱うことができる程度可能となった。

達成の客観的指標として、Conditional Random Field(CRF)による機械学習を用いて、多様な文体の文に対してどの程度対応できるかを検証した結果を表2に示す。表2は訓

練・評価用データとして、書籍(PB)、雑誌(PM)、新聞(PN)、白書(OW)、さらにインターネット上のブログ(OY)、特定目的掲示板(OC)といった明らかに文体が異なる6分野に属するコーパスを使用し、各分野に属するコーパスから約6,000文ずつ取り出して機械学習し、学習したモデルの文体とは異なる文集合に対しての文における機能表現の解析精度をパーセントで示したものである。この結果が示すように、本研究課題で整備した機能表現辞書を用いた文解析では、どのような文体にも一定以上の精度で対応が可能であることを示している。

表2: 多様な文体の機能表現分析対応

学 習	評価データ						平均
	PB	PM	PN	OW	OY	OC	
PB	-	96.9	97.9	96.2	94.3	95.6	96.2
PM	95.8	-	97.6	94.8	94.0	95.3	95.5
PN	94.6	95.3	-	96.8	91.9	92.7	94.3
OW	89.6	90.4	95.7	-	84.1	85.6	89.1
OY	94.8	95.5	96.6	94.1	-	95.7	95.3
OC	92.7	93.9	95.1	90.6	93.2	-	93.1

上記の機能表現解析モジュールは、既存の係り受け解析や形態素解析といった言語解析モジュールと並行して処理を行い、他の言語解析モジュールで文解析がうまく行えなかった場合でも、機能表現解析により文のおおまかな構造解析が同定でき、かつ当該文の特徴を多面的に検討することができるよう開発した。この技術は意味処理との連携の点からも、統語的な言語表現が多少崩れている場合も内容に関わる表現を抽出することがある程度可能となり、多様な医療福祉サービスの受容者が日常的に使用するくだけた言語表現の分析に対応する処理基盤となる。

(3) 対話例構築と支援技術

①ドメイン知識と対話例とを体系付ける対話モデルを基礎に、一般的あるいは初級の日本語教材からの連続的な発展教材となるように医療福祉分野の目的志向性を持つ日本語対話例集を構築した。構築には外国人看護師・介護福祉士候補者への日本語教育、あるいは看護介護分野の日本語教材作成の経験をもつ複数の日本語教師の協力を得ることができた。2011年度末の対話例集の文数は約800文対である。

②上述①の日本語対話例集は本研究課題で開発した資源と基盤技術を応用した支援を行いながら構築した。具体的な構築過程としては、複数の研究協力者と研究者が議論を行いながら、次の(4)に示す試作システムの基本データなることを確認しつつ漸進的に構築した。その際、作成対話例の評価を行う上で、基盤知識と高い一貫性を保つこと、教

育的効果の期待できる事例であることが重要な視点であった。すなわち、対話例の中に盛り込む候補となる表現や特徴を、標準的・代表的な対話例に様々に結びつけて議論・検討する必要があった。その過程を効率的かつ客観的に行うために、語彙的な多様性や構文の特徴付け、既に作成した例文との相関性を検討できるよう、以下の支援技術を開発した。

- 多角的な日本語特徴の類似検索
- 評価対象・評価表現の類似文脈検索

③対話例に用いる表現については、(1)のドメイン依存ソーラスに整理した評価表現をさらに重点的に拡張した。具体的にはオノマトペ表現を代表とする口語体で身体や心理的な状態を表現する上で、直感的、感覚的に用いられる表現を登録した。また、②の支援技術により効率的に収集した表現候補を教材の対話例として採用するための基準についても研究を行った。具体的には、複数の日本語教師によって当該語彙を基本的な日本語表現を用いて言い換えた場合でもその結果には多様性があることを確認した上で、教育・学習に適した共通性を持つ表現の意味的特徴を明らかにした。母語話者の場合は、個別性の高い表現であっても、その意味を文脈に応じて柔軟に判断することができるため意識することが難しいが、この手法により日本語学習者に明示的に説明する上で、特定の表現がどのような点で特徴的かを把握することができた。この検討方法は同義性をもつ表現の教材化を検討する教育語彙研究の有益な手法になりうる。

(4) 日本語学習支援システムの試作と発展性

①教材文の協調的作成と運用を目的として、協調的日本語学習支援システムを試作した。システムの前提となる、対話例集は本課題で試作した例だけではなく、将来的により個別性の高い対話例が必要になってくること、また、既に本課題とは別に医療福祉分野の日本語例文集を構築した試みもあることを考慮にいった。そのようなデータの日本語学習教材としての適切性を検討し、学習者に有益な情報を提供する、以下の要素モジュールを開発した。

- 日本語の文型特徴解析
- 類似文脈での他表現検索

これらの要素モジュールは、WebAPIとして本研究課題とは関係のないシステムでも利用できるように実装した。それらのWebAPIを利用すれば、必ずしも標準的な日本語ではない入力文に対して、日本語のあいまいな部

分、あるいは、日本語の誤りを指摘・訂正を支援するインターフェースの処理基盤とする等の多様な応用が可能である。

②本課題で構築した資源と技術を、より有効に日本語学習あるいは運用支援に生かす機能を試作システムの開発に盛り込んだ。具体的には、日本語教育の専門性が必ずしも高くないボランティアや同僚と外国人が強調的に日本語学習・運用する環境下で、対話例集を多角的な教材として協調利用できる問題集システムを構築した。このシステムはワークショップ等の複数の会合での議論を通して、多数の日本語教育実務者から有効性について確認を得た。なお、この成果は一般的に試用可能な形でWebページに公開した。

知識記述や計算文法を前提に、対話例や文例といった教育資源を共有することによって、個別的な日本語学習や運用に益するモジュール開発は大きな可能性を秘めており、今後、さらに研究開発を継続したい。

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1件)

①竹内和広、西村由美、介護分野におけるオノマトペ表現の言い換え分析に基づく意味的特徴の調査、大阪電気通信大学人間科学研究、査読有、第14号、2012年、pp.131-143.

[学会発表] (計9件)

①竹内和広、介護場面の日本語学習に向けた例文データベースの構築支援技術、第17回公開シンポジウム「人文科学とデータベース」、2012年1月7日、京都工芸繊維大学(京都)

②嶋村和樹、西村由美、竹内和広、日本語学習者のための多様な文体変換を考慮した基本例文データベースの構築、電子情報通信学会「人工知能と知識処理」研究会、2011年10月20日、京都大学(京都)

③西村由美、竹内和広、目的別日本語教育におけるオノマトペ表現の重要性、言語処理学会第17回年次大会、2011年3月10日、豊橋技術科学大学(愛知)

④Kazuki SHIMAMURA, Kazuhiro TAKEUCHI, Kiyota HASHIMOTO, An effective visualization of style inconsistencies for interactive text editing, 16th International Conference on Artificial Life and Robotics, 2011年1月29日、大分国際会議場(大分)

⑤Yumi NISHIMURA, Kazuhiro TAKEUCHI, Kiyota HASHIMOTO, A Consideration on the Support for Japanese Onomatopoeia Learning in Japanese for Specific Purposes, 2nd Asian Conference on Education, 2010年12月5日、ラマダホテル(大阪)

⑥北中佑樹、竹内和広、印象調査アンケートのためのオノマトペ表現による評価尺度の拡張、電子情報通信学会技術研究報告「思考と言語」、2010年11月27日、機械振興会館(東京)

⑦So NODA, Kazuhiro TAKEUCHI, Kiyota HASHIMOTO, Awareness Promoting Visualization of Style Inconsistencies in Texts, IEEE 2nd International Symposium on Aware Computing, 2010年11月2日、国立成功大学(台湾)

⑧北中佑樹、秋吉信吾、竹内和広、日本語教育に向けたオノマトペ表現データベースの構築、第20回ソフトサイエンスワークショップ、2010年3月14日、名城大学(愛知)

⑨安藤秀明、秋吉信吾、竹内和広、作文指導に向けた表現データベースの構築と表現評価指標の検討、第15回公開シンポジウム「人文科学とデータベース」、2009年11月28日、神戸大学(兵庫)

[その他]

ホームページ等

<http://takelab.org/iitai/>

6. 研究組織

(1) 研究代表者

竹内 和広 (KAZUHIRO TAKEUCHI)

大阪電気通信大学・情報通信工学部・准教授

研究者番号：20440951