

平成18年度
研究開発成果報告書

コミュニケーションロボットの音声対話理解
システムに対する大規模対話知識の研究開発

委託先： (株)言語理解研究所

平成19年4月

情報通信研究機構

平成18年度 研究開発成果報告書

(地域中小企業・ベンチャー重点支援型)

「コミュニケーションロボットの音声対話理解システムに対する 大規模対話知識の研究開発」

目 次

1	研究開発課題の背景	2
2	研究開発の全体計画	
2-1	研究開発課題の概要	3
2-2	研究開発の最終目標	7
2-3	研究開発の年度別計画	9
3	研究開発体制	
3-1	研究開発実施体制	10
4	研究開発実施状況	
4-1	実施概要	11
4-1-1	対話知識構築の研究開発実施概要	11
4-1-2	平成18年度10月から12月の実施概要	11
4-1-3	平成18年度1月から3月の実施概要	11
4-2	サブテーマと実施内容	11
4-2-1	研究開発に関わる全体的な管理の実施	11
4-2-2	要求意図知識ベースの研究開発	14
4-2-3	対話知識ベースの研究開発	15
4-2-4	応答知識ベースの研究開発	16
4-3	総括	17
4-3-1	研究開発の総括	17
4-3-2	研究開発に対する技術的評価	19
4-3-3	研究開発の今後の課題	21
4-3-4	研究開発による波及効果	21
5	参考文献	
5-1	研究発表・講演等一覧	

1 研究開発課題の背景

1-1 背景と目的

ロボットの研究開発に関して、日本は海外に比べて進んでいるが、現在のロボット市場は製造業分野が大半を占めている。ロボット工業会、経済産業省、総務省などの調査報告書によると、2025年にはロボット市場は約8兆円となり、生活や医療福祉分野などの市場が約7兆円と大きく拡大し、製造業分野市場の1兆円を大きく上回ると予測されている。特に、少子高齢化社会における介護や育児の対話による支援、ストレス社会による健康支援としての対話支援など、コミュニケーションロボットの活躍が期待されている。

ロボットによる物理的な作業支援環境では、メカトロニクスと安全確保技術が重要な技術となる。これに対して、人間とのコミュニケーションを支援するロボットには、音声対話により人の要求意図を理解できる技術が必要となり、さらには「心の豊かさ」を感じ取れるコミュニケーションロボットへの発展が期待される。

本研究課題は、後者のコミュニケーションロボットの実現に的を絞ったものであるが、音声対話理解研究は大きく遅れているのが現状である。音声対話理解技術は、1) 音声認識と音声合成技術、2) 人間の要求意図を理解するための大規模な対話知識データベース、3) 1) と 2) を駆動して対話システムを稼働させる制御エンジンに分割できる。研究開発現状において、1) の音声認識と音声合成エンジンは多くの実用化製品が存在し、3) 対話システムを稼働させる制御エンジンは、受託企業が製品名「音声対話理解エンジン「コンジューニアル™」」の事業化を行い、関係企業による技術連携販売を開始している。

以上より、本研究開発では、種々のシチュエーションで使われる人間の要求意図知識データベース、対話知識データベース、応答知識データベースを構築することに目標を絞る。言語知識の構築には、多くの時間と経費を必要とし、民間企業で実現するのはリスクの高い分野である。特に、大規模な対話知識の構築を目的とする本研究では、以下の点を十分に考慮する必要がある。

(1) 質の高い対話知識と大規模対話知識を同時に実現するには、人手による質の確認が長期間必要であり、人の交代や多くの異なる視点による概念化などの基準の揺れが生じる。

(2) (1) の観点より、大規模対話知識が短期間に半自動的に構築できる支援モジュールの充実が必要不可欠である。

(3) 支援モジュールには、言語知識（概念、類義語、意味共起関係知識など）と言語解析エンジンが必要不可欠であり、更に、言語知識と解析エンジン群は、製品化などの実績のある質の高い物である必要がある。

(4) (3) の質の高い言語知識と解析エンジン群を駆使することで、本研究の大規模対話知識の効率的な研究開発を実現することが目的となる。

これに対して、大学発ベンチャー企業（株）言語理解研究所では、大規模言語知識と言語理解を基盤とする多種多様な商用製品を事業化しており、音声対話に必要な音声認識と音声合成の企業連携、ロボットコミュニケーション応用の企業連携、研究成果の事業化連携の実績を確立しているため、上記対話知識データベースを研究開発できる日本の数少ない企業であり、本研究開発の意義は非常に高いといえる。

1-2 本研究で関係する技術の概要

本研究における一つの手法は、大規模テキストから要求意図を有する用例文を収集することであり、これら用例文から対話知識の候補を一次的に絞り込む研究開発がある。この意図情報に関連する研究として、意見情報抽出 (Yu et al. 2003 ; 小林他 2005)、好評不評判定 (Fuketa et al. 2005)、主観性判定 (Hatzivassiloglou et al. 1997 ;

Hatzivassiloglou et al. 2000)、そして、製品紹介の文タイプ（性能記述、価格記述などの文タイプを検出する）による評価（Tokunaga et al. 2004）があるが、これらは間接的な意図を理解する研究には発展していないのが現状である。

また、特定の要求と応答の組み合わせを質問と回答として検索できる関連研究としては、Wilensky (Wilensky et al. 1984)らがUNIX 操作要求に対応するコマンドを回答するUC というシステムを提案した。但し、このシステム専用に構築された知識ベースを必要とするなどの問題があった。また、Hammondらは (Hammond et al. 1995) 自然言語テキストの知識ベースによるFAQシステムFAQ Finderの提案を行っている。FAQの機能は、一般的に固有のFAQ知識ベースをもとに、質問を入力として回答を得るところから、機能面から考えると、本研究の要求意図と応答意図の関連付けに類似している。更に、大規模テキストを用いる質問応答システムに関する研究（清田他 2003； Ferret et al. 2001； Fukumoto et al. 2002； 渡辺他 2005）があるが、この研究も要求意図と応答意図を関連づける本研究開発に間接的に関係する。

意図理解や肯定否定という意図判定や分類検索研究では、Nasukawa (Nasukawa et al. 2003； Kadoya et al. 2005； Fuketa et al. 2002； Pang et al. 2002； Kwon et al. 2003； Kwon et al. 1999； Lam et al. 1999； Moens et al. 1997； Turney 2002)らの研究はあるが、これらは深い要求意図理解や応答意図との関係について、あるいは大規模対話知識の構築方法は議論されていない。

また、分野連想語関係の研究手法としては、Fuketa et al., 2006； Lee et al. 2002； Atlam, et al. 2006； Atlam, et al., 2003 が利用され、用例文の検索加工と意味共起関係に対する検索技術は、Aho, et al., 1975； Aoe, 1989, Aoe et al., 1996； Morita et al., 2003 が使用された。

辞書検索や文字列関係の研究手法としては、Aho et al. 1975； Aoe 1989； Aoe et al. 1996 が使用された。

2 研究開発の全体計画

2-1 研究開発課題の概要

2-1-1 研究の目的

少子高齢化社会における介護や育児の対話による支援、ストレス社会による健康支援としての対話支援など、コミュニケーションロボットの活躍が期待されているが、この研究開発課題において、音声対話理解研究は大きく遅れているのが現状である。従って、「心の豊かさ」を感じ取れる気の利いたコミュニケーションロボットを効率的に開発できる汎用的音声対話知識が必要不可欠である。この社会ニーズにより、本研究課題では、生活分野での衣料、食事、住居、健康に関する対話を実現するための大規模対話知識を研究開発し、以下の3つの知識ベースを構築することを目的とする。

1. 要求意図理解知識ベースの研究開発
2. 対話知識ベースの研究開発
3. 応答知識ベースの研究開発

2-1-2 研究課題の位置づけ

まず、本制度の研究開発と受託企業が自主的に行う研究開発を区別する。

本制度の研究開発と自主的に行う研究開発を明確に区別するために、全体システムとそれぞれの部分エンジン群を図1に示す。

(A) 上位アプリケーション層

ロボット本体の制御部分であり、アプリケーションとして音声対話理解システムを連携する部分である。

(B) 音声認識・合成エンジン層

アプリケーションに連結するためのAPI（アプリケーションプログラムインターフェース）が定義される部分であり、人間で言えば、口と耳に相当する部分である。ここでの音声認識エンジンと音声合成エンジンは市販のエンジン4種類が組み込まれており、今後自社開発で更に組み込みエンジンの種類を多くして、汎用的な音声インターフェースとする。

(C) 音声対話理解エンジン層

音声認識エンジンの結果を意図理解し、対話制御の判断により、適切な応答文を生成する。人間で言えば、脳に相当する部分である。意図理解エンジンは、音声入力に対する意図を理解する部分であり、「のどが渴いた」の要求は、飲み物不足を理解し、それを充足する意図結果を生成するエンジンである。対話制御は、意図が十分に理解できないとき、聞き返すための応答や過去の対話履歴や個人嗜好記憶により、「親しみやすい」制御を司るエンジンである。

応答文生成エンジンは、年齢や性別や、感情の状態などの立場変化により自然な応答文を生成する。音声認識辞書は要求意図知識ベースの言語モデルが対話の話題により動的に切り替えられるものであり、また音声合成のための合成辞書（読み上げアクセント辞書）は、応答文に応じて切り替えができるものであり、それぞれ自社開発ですでに事業化されている。

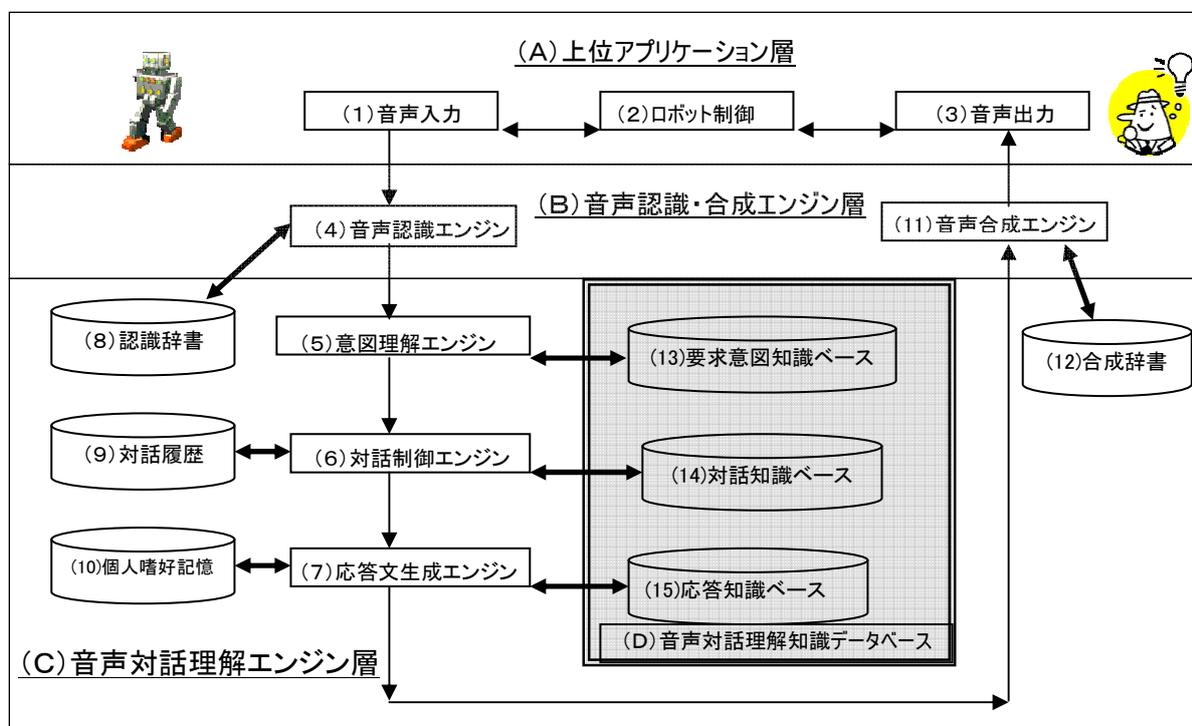


図1 コミュニケーションロボット音声対話ライブラリと研究開発内容の説明図

以上のエンジンと辞書群は、自社により評価版としてすでに事業化し、2年後から開発環境として販売するものである。しかしながら、以下の知識データベースは、食べ物、スポーツ、車関係の単純な知識データベースしか構築されておらず、コミュニケーションロ

ボットの普及を目指すには、多種多彩なシチュエーションにおける対話理解知識（汎用的な知識ベース）が必要不可欠であり、(D) 音声対話理解知識ベースの構築を本研究開発として推進するものである。

2-1-2 研究方針と方法

人間の直接的な要求意図理解は容易であるので、研究方針は以下に説明される間接的な要求意図理解に着目する（山田他 1994）。例えば、飲み物を依頼する言い方は、以下のよう疑問文、命令文、平叙文で表現できる。

疑問文：“飲み物をくれませんか？”

命令文：“飲み物をくれ”

平叙文：“飲み物が欲しいです”

これらの発話要求は直接的であり、意図理解は容易である。このような発話を直接的発話と呼ぶ。このような直接的発話でなく、話し手が聞き手に情報提供や行動を要求するための発話行為(speech act)においては、言語外の知識を利用して、発話の真の意図を理解する必要がある。このような表現を間接的発話と呼ぶ。例えば、次の発話 p を考えてみる。

“のどが渇かないですか？” 発話 p

発話意図を「飲み物が欲しい」と仮定すると、“ジュースはいかがですか？”なる聞き手の応答は話し手を満足させるものである。話し手は、質問の意図を満足させる応答を聞き手に期待しているので、その知識を以下の深層表現知識で構築する。

上記の発話 p “のどが渇かないですか？”の真の意図は「ある状況 C において、体内に水分が不足している」ことに起因すると解釈できる。これを組み合わせると、発話 p の意図は (C (状況), 体内 (場所), 水分 (対象), 不足 (充足関係)) なる深層表現で構築する。聞き手が、発話 p の意図を理解し、話し手の期待する応答 q をするためには、「水分 (対象)」を「飲み物 (対象)」に、充足関係にある「不足」を「補給」に変換して、応答意図の深層表現 (C (状況), 体内 (場所), 飲み物 (対象), 補給 (充足関係)) を得ることが必要であり、この深層表現から表層の応答候補 q = “ジュースはいかがですか？”が生成できる (図 2 参照)。以上より、要求意図理解知識と応答知識はそれぞれの深層表現で構築できる。対話知識は、この深層表現の変換規則として構築される。

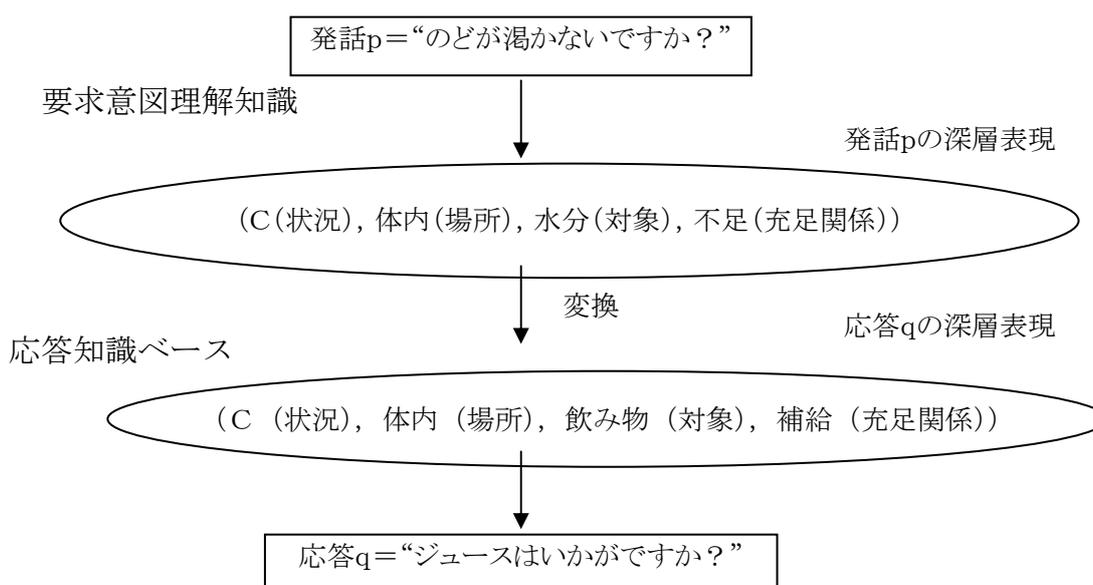


図 2 充足関係による意図理解知識と深層表現による変換例

2-1-3 研究課題内容とサブテーマ

対話理解知識ベースの研究開発内容では、人間生活にもっとも深くかかわる「衣料、食事、住居、健康」に限定した生活分野4知識、ストレス解消のための10種類個人嗜好分野の質問応答対話知識を期間1年半で研究開発する。

(1) 要求意図知識ベースの概要

1) 意味共起関係による言語理解の研究手法

対話における人の発話は、要求意図を有する場合が多い。この意図を理解するために受託企業で保有する大規模意味共起知識(約8,000万関係)を主体的に利用する。意味共起知識とは、単語「当たる」だけでは、正確な意味が理解できないが、「宝くじに+当たる」は「嬉しい」と意味理解できる。この単語と単語の2項関係知識が意味共起知識である。受託企業は、最もシンプルな2項関係知識に着眼して、各種事業化を成功させており、本研究でもこの方針を採用する。

2) 種々の意味共起関係知識

意味共起関係には、上記のような文章中の表現だけでなく、概念関係の「バスと乗り物」、場所関係の「富良野と北海道」、分野関係の「イチローと野球」、会社の階層関係「トヨタと自動車メーカー」、名寄せの共起関係「NECと日本電気」などのように、言語の意味理解には多種多彩な共起関係があり、研究課題では、これらも広義の意味共起関係と定義している。

3) 要求意図知識の定義

上記の膨大で且つ多種多彩な意味共起関係から対話に有用な要求意図知識を2重組知識(用例文, 要求意図)として分析・精選・拡充研究を経て構築するものである。例えば、2重組は、次のような知識表現で構築され、2,000万関係(組)を構築の目標とする。

- (「富良野に行きたい」, 旅行意図)
- (「富良野は北海道ですか?」, 場所確認意図)
- (「セーターはどこで買う」, 衣料購入意図)
- (「セーターは衣類ですか?」, 分類確認意図)
- (「スパゲッティが食べたい」, 食事意図)
- (「スパゲッティは麺類?」, 分野確認意図)
- (「大学に合格した」, 朗報意図)
- (「イチローと言えば野球」, 話題意図)

4) 個人嗜好を理解するための質問知識による要求意図

また、コミュニケーションの親密感を得るために、ロボット側が相手の個人嗜好を把握することは極めて有効である。例えば、

- 「イチローを知ってますか?」
- 「朝青龍って?」
- 「湯川秀樹は?」

のように、対象の分野「野球」「相撲」「科学」への興味度合いをこの質問の解答から類推する質問知識は、上記に対して以下の2重組で表現される。

- (「イチローを知ってますか?」, 野球分野興味確認意図)
- (「朝青龍って?」, 相撲分野興味確認意図)
- (「湯川秀樹は?」, 科学分野興味確認意図)

5) 大規模要求意図知識ベースの検索エンジン

これら知識ベースは、受託研究が大学からの技術移転である高速辞書検索手法「ダブル配列(Aoe, 1989, Aoe et al. 1996; Morita et al. 2003)」による意味共起辞書に融合化する研究方針で計画し、高速性を劣化させない設計と開発を目標とする。

(2) 対話知識ベースの概要

1) 対話知識ベースの定義

要求意図を理解して、脈絡のある応答文を生成することが「親しみ」のある対話の原点であるので、対話知識ベースは2重組要求意図知識に応答側の意味づけを研究することであり、3重組知識（用例文，要求意図，応答意図）で構築される。上記の例の一部に対して、以下の3重組が対話知識に対応する。

（「富良野に行きたい」，旅行意図，案内意図）

（「セーターはどこで買う」，衣料購入意図，販売店意図）

（「スパゲッティが食べたい」，食事意図，料理意図）

（「大学に合格した」，朗報意図，共感意図）

対話知識ベースは、ロボットの応答を決定する重要な知識であるので、基盤となる100万件の3重組知識を研究開発する。

2) 対話連鎖知識の定義

対話連鎖知識とは、人間から問いかけを理解して、ロボットが更に問い返すことで、対話を連続（連鎖）させる知識であり、基盤となる10万件の3重組知識を構築する。上記の例で説明すると、「富良野に行きたい」，旅行意図，交通手段問いかけ）の3重組知識となる。

3) 対話知識ベースの検索エンジン

この対話知識ベースも「ダブル配列」によるバイナリー辞書検索エンジンを開発するが、高速性を維持できる設計と開発を目標とする。

(3) 応答知識ベースの概要

1) 応答知識の定義

応答知識は上記(2)の3重組の応答意図に対応する具体的な応答文を追加した4重組（用例文，要求意図，応答意図，応答文）の知識となり、2,000万組の4重組知識を構築する。上記の例の一部では、以下の4重組に対応する。

（「富良野に行きたい」，旅行意図，案内意図，「ラベンダーがいいですね」）

（「富良野に行きたい」，旅行意図，交通手段問いかけ，「飛行機で行きますか？」）

また、個人嗜好の要求意図知識（「イチローを知っていますか？」，野球分野興味確認意図）に対しては、対話知識ベースを介さず

（「イチローを知っていますか？」，野球興味分野確認意図，肯定，「知っています」）

の4重組知識を応答知識ベースで構築する。

2) 応答知識ベースの検索エンジン

この応答知識ベースは、受託研究が大学からの技術移転である高速辞書検索手法「ダブル配列」による意味共起データ辞書に融合化するが、高速性を劣化させない設計と開発を目標とする。

2-2 研究開発の最終目標（平成20年3月末）

研究受託企業で製品化されている音声対話理解システム開発支援システムで使用できる以下の知識データベースを構築する。

(1) 要求意図知識ベース【サブテーマ】

ア) 意味共起8,000万関係への意味づけ開発

生活分野の衣料、食事、住居、健康の4知識の表層展開算出で、2,000万件の要求意図知識を構築。

イ) 個人嗜好の意図理解辞書

10種類の個人嗜好分野（スポーツ、娯楽、趣味関係）で要求意図を持つ質問知識 26,000 件以上を構築する。

- ウ) 既存の意味共起データ辞書への融合化ソフトウェア開発
100 万件格納辞書における検索速度を 1 件で最悪でも 0.1 秒以内の検索速度とする。

(2) 対話知識ベース【サブテーマ】

- ア) 対話事例の収集

要求と応答事例を Web データなどからの収集を 3,000 万文以上とする。

- イ) 対話辞書管理データベースシステムへの登録管理作業

上記 3,000 万文より、展開型表層表現で 100 万対の以上の対話変換知識と 10 万件以上の対話連鎖知識を構築する。

- ウ) 対話バイナリー辞書のソフトウェア開発

100 万件格納辞書における検索速度を 1 件で最悪でも 0.2 秒以内の検索速度とする。（要求意図と応答知識からのアクセス時間を含む）

(3) 応答知識データ【サブテーマ】

- ア) 意味共起 8,000 万関係への応答知識開発

生活分野の衣料、食事、住居、健康の 4 知識の表層展開で 2,000 万件の応答知識候補を構築。

- イ) 個人嗜好の応答理解辞書

10種類の個人嗜好分野（スポーツ、娯楽、趣味関係）で要求意図を持つ質問知識 26,000 件以上を構築する。

- ウ) 既存の意味共起データ辞書への融合化ソフトウェア開発

100 万件格納辞書における検索速度を 1 件で最悪でも 0.1 秒以内の検索速度とする。

2-3 研究開発の年度別計画

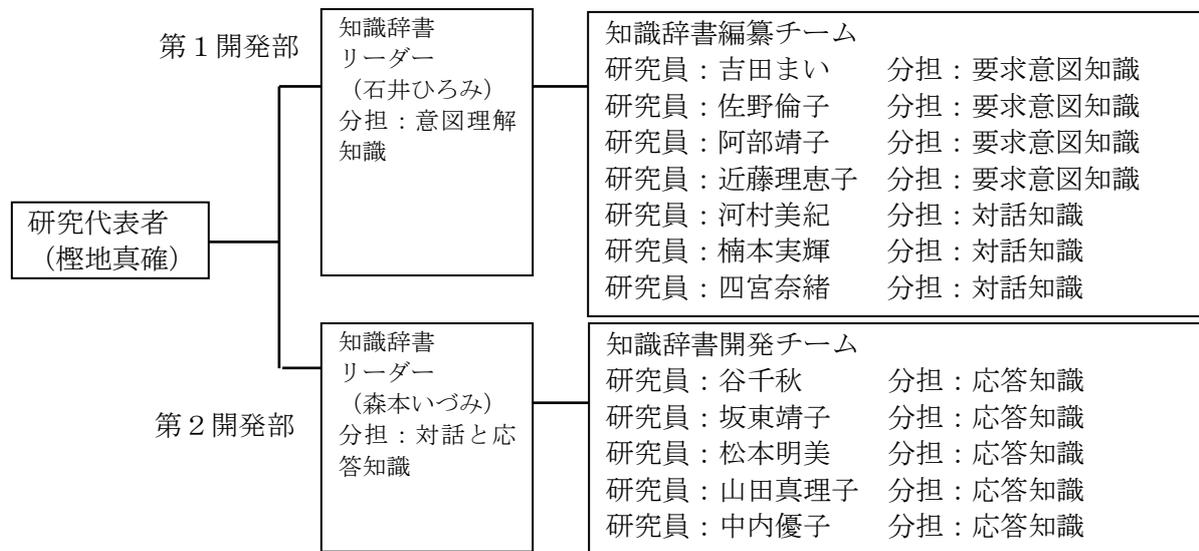
金額は非公表

研究開発項目	18年度	19年度	年度	計	備考
コミュニケーションロボットの音声対話理解システムにおける対話知識の研究開発					
1) 要求意図知識ベース 【サブテーマ】	→	→			
2) 対話知識ベース 【サブテーマ】	→	→			
3) 応答知識ベース 【サブテーマ】	→	→			
間接経費額 (税込み)					
合 計					

- 注) 1 経費は研究開発項目毎に消費税を含めた額で計上。また、間接経費は直接経費の30%を上限として計上 (消費税を含む)。
 2 備考欄に再委託先機関名を記載。
 3 年度の欄は研究開発期間の当初年度から記載。

3 研究開発体制

3-1 研究開発実施体制



共同研究先

- 1) 徳島大学大学院ソシオテクノサイエンス研究部
- 2) 筑波大学大学院ビジネス科学研究科

4 研究開発実施状況

4-1 実施概要

4-1-1 対話知識構築の研究開発実施概要

- 1) 本研究は対話知識を人手ですべて構築するのではなく、受託企業が有する既存の基盤知識や基盤エンジンなどを駆使して、対話知識を「如何に素早く、如何に精密に、如何に自動化」するかを研究することである。この効率化により、作業工数の膨大化のリスクが解消され、高度知識を利用した応用製品の事業化への道が開かれるものとなる。
- 2) この観点により、本研究開発では、スキルの異なる多くの研究者が一つの目的を持った知識を構築するので、まず判断基準を統一する基盤教育を実施した。具体的には、用例文の選別、対話知識に採用すべき話題、要求意図表現、応答知識表現などの試行錯誤的（動的）な作業を初期段階で実施した。
- 3) 平成 18 年度は統一基準や効率化のための試行錯誤的な研究分析が多く発生するので、研究計画では平成 18 年度の対話知識達成率は平成 19 年度より低く設定されている。この研究計画に沿って、サブテーマに関する研究開発概要を以下に記載する。

4-1-2 平成 18 年度 10 月から 12 月の実施概要

- 1) 研究内容の異なる研究員同士の情報共有を促進するために、異なる分野の作業も交代で行うことで、意味共起関係、言語解析、分野抽出結果、各種エンジンからの出力結果、品詞や概念など基盤となる言語知識の内容、及び基盤言語データベースの参照や操作法の相互理解を進めた。
- 2) 研究補助員に対しても、基本的知識の作業を通じて統一的判断基準を育成した。さらに、対話知識を効率的に構築するために必要とされる支援モジュール（言語解析連動の登録画面など）やその他必要とされる自動構築のために支援モジュールなどの洗い出しを行った。

4-1-3 平成 18 年度 1 月から 3 月の実施概要

- 1) 平成 18 年 10 月から 12 月の前半で試行錯誤を重ねて構築された意味共起関係や話題（分野）情報などから、対話知識の具体的体系の定義と同時に構築データの精選、割り当て、分析評価を行った。
- 2) この後半の作業において、平成 18 年度の数値目標を達成し、構築された対話知識を統一基準の参照事例として、平成 19 年度の研究開発へと進むための基盤研究が確立されたことになる。

以上の研究実施概要をふまえて、サブテーマの研究開発成果を記述する。

4-2 サブテーマと実施内容

4-2-1 研究開発に関わる全体的な管理の実施

4-2-1-1 対話知識データベース管理の実施内容

ア) データベース管理と登録支援モジュールの実施内容

委託研究開発による要求意図知識、対話知識、応答知識を新規に登録編纂できる知識データベース管理モジュールを開発し、既存の登録画面の改良と拡充を含めて実施する。

イ) ダンプファイル生成モジュールの実施内容

研究開発された知識情報を定期的にテキストファイルダンプできるモジュールを開発する。(テキストファイルによるバックアップも兼ねる)

ウ) データベース派遣研究補助員の研究指導の実施内容

人材派遣研究員のデータベース関係作業における研究指導を継続的に実施する。

4-2-1-2 対話知識データベース管理の実施状況

ア) データベース管理と登録支援モジュールの実施状況

- 1) 「昨日管理栄養士試験に見事に合格しました」のような気持ちを表す用例文を収集し、その用例文から要求意図を表す意味共起関係を抽出する研究が4-2-2からのサブテーマで実施されるので、その知識データをデータベースに効率的に格納する手法を研究し、登録の支援モジュールを開発した。
- 2) 受託研究実施前には、データベース管理画面とは独立に解析エンジンを立ち上げて、バッチ処理で形跡結果を参照しているが、大量の用例文を分析評価するためには、言語解析エンジンをデータベース管理モジュール上で稼働できれば、単語区切り、単語の概念や品詞情報を参照することで、中心情報「試験に合格しました」が分析しやすく、見分けやすくなる。
- 3) この目的により、受託企業が所有する言語解析エンジンをデータベース(マイクロソフト ACCESS など)画面上から呼び出すための支援モジュール(エンジンのCOM化)を実施した。
- 4) 平成19年1月末に試験版が完成し、平成19年3月末までにデータベース画面上での組み込み試験も完了したので、目標は達成された。

イ) ダンプファイル生成モジュールの実施状況

- 1) 知識辞書構築データベースから要求意図知識、対話知識、応答知識など多彩な知識をテキストダンプし、ファイルバックアップを実施した。
- 2) エクセル形式での作業単位ごとのダンプを行えるようにして、試行錯誤的な知識構築実験ができる環境を実現し、効率的知識構築のための支援目標は達成できた。

ウ) データベース派遣研究補助員の研究指導の実施状況

- 1) 上記ア)とイ)に関する研究補助員(開発派遣者2名)に対しては、解析エンジンAPIの説明とデータベース管理のスキーマ設計などの教育を実施した。
- 2) 用例文収集と知識構築支援を行う研究補助員(派遣者5名)に対しては、解析エンジン結果、知識の種類など非常に多くの専門知識を教育し、採用時から約3ヶ月間随時実施し、目標は達成された。

4-2-1-3 対話知識データベース管理のまとめ

ア) データベース管理と登録支援モジュールのまとめ

解析エンジンの支援モジュール(エンジンのCOM化)は上記実施状況で記載のように、研究計画に沿って達成された。今後、平成19年度ではデータベース管理画面も種類が増えるので、効率化のために随時支援モジュールの組み込みを実施する計画である。

イ) ダンプファイル生成モジュールのまとめ

知識辞書構築データベースから要求意図知識、対話知識、応答知識など多彩な知識のテキストダンプ支援モジュールを開発し、構築知識のバックアップを兼ねる背品モジュール拡充した。効率的知識構築のための支援目標は達成できた。

ウ) データベース派遣研究補助員の研究指導のまとめ

開発派遣者2名に対する解析エンジンAPIの指導に関しては研究開発の目的に沿って実施できたが、知識構築支援を行う研究補助員派遣者5名に対する知識内容と種類には多くの時間を必要としたので、研究開発の目的のみならず、言語知識に関するスキルも必要であることが分かった。人手を増やせば良いということではない、今後の参考にしたい。

4-2-1-4 大規模テキストデータの収集とテキストクレンジングの実施内容

ア) Web 情報からの用例文収集の実施内容

対話事例を収集するために、Web 情報として、ブログ、掲示板、広報情報などを主体とする大規模テキストを継続的に収集する。

イ) 文情報の切り出しの実施内容

タグ情報、文区切り情報により、文書情報のみを切り出す。

ウ) 不適切表現のフィルタリングの実施内容

対話知識に不必要な不適切表現を検査し、削除する。

エ) 話題分野への分類の実施内容

対話知識の話題分野に分類する作業を実施する。

オ) 対話知識構築の事例文1,000 万文構築の実施内容

文単位のテキストに細分化する。平成18年度の収集目標は、1,000 万文とする。
(平成19年度の最終目標は3,000 万文)

4-2-1-5 大規模テキストデータの収集とテキストクレンジングの実施状況

ア) Web 情報からの用例文収集の実施状況

- 1) Web から定期的に掲示板、広報情報、ブログテキストなどを収集し、対話事例となる用例文を収集した。
- 2) 受託企業における既存の大規模コーパスからも用例文抽出を行った。用例文抽出は目標の1,000 万文に対して、定期的な収集が実施された。

イ) 文情報の切り出しの実施状況

- 1) 大規模コーパスから独立した文（あるいは、複数文の固まり）を切り出す文切り出し支援ソフトを改良し、切り出し支援モジュールの開発を実施した。
- 2) 記号連続、同一表現文字連続などを最長部分文字列抽出法で検出し、日本語としての非文となる事例をフィルタリングした。

ウ) 不適切表現のフィルタリングの実施状況

- 1) 掲示板やブログの大規模コーパスにおいては、誹謗中傷、卑猥表現などの本研究の目的とする対話知識内容で不適切な情報が多く含まれることが分かった。
- 2) 文字列照合で不適切表現候補を高速に検索し、その部分を更に部分形態素解析できる受託企業のモジュールを利用して、この不適切表現のフィルタリングを10月から12月に集中的に実施した。

エ) 話題分野への分類の実施状況

- 1) 上述のウ) の不適切フィルタリングの後、事例文データを話題分野に分別する研究開発を実施した。
- 2) この話題分野は受託企業の有する支援モジュールで約100 分野であり、対話知識の効率的構築作業に大きな貢献があった。

オ) 対話知識構築の事例文 1,000 万文構築の実施状況

- 1) 上記の作業を経由して、大量の事例文を構築するのは、最初の数ヶ月において大きな工数が発生し、目標の 1,000 万文構築の見通しが悪かったが、受託企業の感性抽出支援モジュールを活用して有効な事例文の抽出を推進した。
- 2) 平成 18 年度 1 月から、上記 1) の効果により、収集効率は向上し、平成 18 年度末までに目標は達成できた。

4-2-1-6 大規模テキストデータの収集とテキストクレンジングのまとめ

- 1) 上述のように、作業内容と数値目標は達成できたが、対話知識を構築するための管理支援と前処理支援の研究開発としては、平成 19 年度も必要であり、受託企業の保有する種々の解析エンジンの活用、不適切表現フィルタリング支援、話題の類別支援モジュールなどを適宜改善する必要がある。
- 2) 上記 1) の改良は、平成 19 年度の研究計画のサブテーマ内の開発において、適宜実施するものとする。

4-2-2 要求意図知識ベースの研究開発

4-2-2-1 要求意図知識ベースの研究開発の内容

ア) 意味共起への意味づけ開発の実施内容

意味共起を主体とする要求意図知識を構築し、表層文による展開算出で 500 万件以上の構築を目標とする。(最終目標は 2,000 万件以上)

イ) 生活分野(衣料、食事、住居、健康)の分類の実施内容

要求意図知識を生活分野(衣料、食事、住居、健康)の分野に分別する情報を構築する。“サッカーを観戦したい”のように、名詞句“サッカー”のみで分野が特定できる場合は、名詞句の分野情報とする。

ウ) 代表 10 分野の質問知識の実施内容

代表的 10 種類(スポーツ、娯楽、趣味関係など)の個人嗜好分野における質問知識を構築し、目標は 6,000 件以上とする。(最終目標は 26,000 件以上)

4-2-2-2 要求意図知識ベースの研究開発の実施状況

ア) 意味共起への意味づけ開発の実施状況

- 1) 平成 18 年度 10 月～1 月において、受託企業が保有する 8,000 万関係と収集用例文の意味共起知識に対する要求意図知識の抽出実験と分析評価を行い、引き続き 2 月と 3 月で用例文と要求意図の意味づけと関連づけ研究を行った。切り出し文が複数の行である場合は、要求意図の意味づけは最後の文の意図を優先して研究を進めた。
- 2) また、対話知識に不足する解析上の語彙の検出や分析作業も並行して実施された。この実施においては、受託企業の有する感性理解エンジンによる事前処理を経て、感性意味が抽出されない用例文に対する評価分析を進める研究手法が実施され、評価分析工数の効率化が実現された。
- 3) 具体的実施内容は、用例文「大学に合格した」に要求意図知識の意味づけ「朗報」を関連づけることである。
- 4) 実施の進捗状況(目標構築数に対する比率)は、1 月末の段階で達成率が 60%(文数 100%であるが、精度面が 60%)であり、平成 18 年度末までに 100%が達成された。

イ) 生活分野(衣料、食事、住居、健康)の分類の実施状況

- 1) 話題に対する要求意図分類知識の構築が主体的な実施内容であり、具体例とは、商品名「ipod を買いたい」を「音楽機器」に関連づけることである。
- 2) この話題判定により、ロボットは音声認識の範囲や応答の範囲を絞り込むことができる。人間も同様な認知をしているという研究手法である。対話知識に不足する解析上の語彙の検出や分析作業もこの研究開発で実施された。
- 3) 実施の進捗状況（目標構築数に対する比率）は、生活分野（衣料、食事、住居、健康）の分野別目標に対して、1月末の段階で達成率が90%となった。
- 4) 「サッカーを観戦したい」は、「サッカー」のみで分野が特定できるが、「シュートを投げる」と「シュートを蹴る」はそれぞれ意味共起により、分野の野球とサッカーを類別する必要がある。「サッカーを観戦したい」の要求意図は、サッカーと観戦の共起関係が必要となる。しかし、分野を認定する上記の意味共起関係は単語に比べて構築が難しいので、研究実施では、単語や句の表現を利用した用例文を構築した。
- 5) 一般的な対話においては、固有表現である商品、地名など要求意図は重要であるので、重点的に意味共起の事例構築を進めた。詳細分野の研究を進め、平成18年度末には100%達成が実現できた。

ウ) 代表10分野の質問知識の実施状況

- 1) 個人嗜好を問いかける質問知識が実施内容であり、具体例としては、対象固有表現「ipod」を知っていれば、「音楽機器に興味があり」、そうでなければ、「音楽機器に興味がない」という個人嗜好情報を獲得する研究手法である。
- 2) 実施の進捗状況（目標構築数に対する比率）は、1月末の段階で達成率80%であるが、食事分野への偏りも含めて、手薄な趣味娯楽などの細分化分野への調整を実施した。
- 3) 上記2)の結果、平成18年度末までに目標は100%達成された。

4-2-2-3 要求意図知識ベースの研究開発のまとめ

- 1) 上述のように、数値目標は達成できたが、生活分野の分野数をどの程度にするかは、研究手法として判断の難しいものとなった。現状研究では、イ)の生活分野の細分化分野から10分野を決定するが、生活分野で代表的な10分野を決定するのは容易でないので、平成19年度では偏りのないように研究開発を進め、実際の対話試験を通して、代表的分野の優先順位を決定する手法をとる計画である。
- 2) 一般的対話知識として、要求意図を検討すると、商品や製品に対する話題、あるいは購入要求が汎用的であるので、平成19年度では、製品の広報情報を多く収集し、対話知識に反映していく計画である。

4-2-3 対話知識ベースの研究開発

4-2-3-1 対話知識ベースの研究開発の内容

ア) 対話知識の構築の実施内容

要求意図と応答知識の組知識として対話知識を構築する。4-2-2-1のア)の要求意図知識を抽出した500万件情報から対話知識を選別して構築する。

イ) 対話辞書データベースへの登録の実施内容

4-2-1-4のオ)の収集テキスト1,000万文からの対話知識として、本サブテーマの上記ア)の知識と融合した上で、表層展開で30万対以上（最終目標は100万対以上）の対話変換知識と3万件以上（最終目標は10万件以上）の対話連鎖知識（対話知識のつながり）を構築する。

4-2-3-2 対話知識ベースの研究開発の実施状況

ア) 対話知識の構築の実施状況

- 1) 要求意図から応答文を生成するための対話知識構築研究であり、「大学に合格した」に要求意図知識の意味づけ「朗報」に対して、応答の意味付け「感動」が定義されると、「朗報」、「感動」の組が対話知識となる。
- 2) 研究手法としては、要求意図から応答意図への変換組の定義が中心となり、収集コーパスからの対話事例より収集する手法と、人手で応答事例を分析しながら構築する手法で実施された。
- 3) 同時に、対話知識に不足する解析上の語彙の検出や分析作業も実施された。実施の進捗状況（目標構築数に対する比率）は、1月末の段階で要求意図と応答知識の組知識の達成率は50%であったが、有用な意味共起（概念知識など）の洗い出しにより、効率化が進められ、平成18年度末までに100%の目標が達成された。

イ) 対話辞書データベースへの登録の実施状況

- 1) 対話を連鎖（継続）させ、答えを要求する質問形式知識を構築する研究であり、例えば、「どこの大学に合格しましたか？」の質問であれば、大学名の応答が連鎖的に期待されることが研究手法のコンセプトであった。
- 2) 目標とする30万件の達成率は、1月末で50%であったが、10月から1月までの多くの試行錯誤により、効率的作業行程が確立できたのと、商品などの具体的な対話の有用性が判明した。
- 3) 上述2)の重点化により、平成18年度末までに100%の目標が達成された。また、3万件の達成率は1月末では40%であったが、平成18年度末まで100%が達成できた。

4-2-3-3 対話知識ベースの研究開発のまとめ

- 1) 上述のように、数値目標は達成できたが、要求意図知識構築に比べて、対話知識の構築は遅れ気味となり、課題を残した。
- 2) この理由は上述の例のように「朗報」、「感動」の組が対話知識となり、応答意図に対応する「感動」を多くの種類で発見する研究が難しいことに起因している。
- 3) 平成19年度においては、対話事例の収集と同時に、対話の場面や対話の状況を具体的に加味することで、応答意図の決定を効率化する研究手法を導入する計画である。

4-2-4 応答知識ベースの研究開発

4-2-4-1 応答知識ベースの研究開発の内容

ア) 表層応答文の構築の実施内容

- 1) 上記4-2-2-1のア)で構築した表層展開500万関係に対する応答候補知識を構築する。
- 2) 本知識は、複数の応答文候補が存在するので、応答候補文総数の構築目標を500万文とする（最終目標は2,000万文）。
- 3) 更に、4-2-3-1のイ)の対話変換知識3万件への応答知識を構築する。

イ) 質問知識の応答知識構築の実施内容

- 1) 上記4-2-2-1のウ)で構築した個人嗜好分野における質問知識に対して、その解答知識を応答知識として構築する。

- 2) 目標は 5,000 件以上とする。(最終目標は、26,000 件以上)

4-2-4-2 応答知識ベースの研究開発の実施状況

ア) 表層応答文の構築の実施状況

- 1) 要求意図から応答文を生成するための対話知識構築である。上述の例のように、「大学に合格した」に要求意図知識の意味づけ「朗報」に対して、応答の意味「感動」が定義されると、「朗報」、「感動」の組が対話知識となり、研究手法ではこの応答意図（深層表現）「感動」に対する応答文「合格良かったね。すばらしい」を生成することである。
- 2) 応答文が単調になると、対話が飽きられやすいので、表層的な文でも多く構築する研究手法が実施された。その中でも、若者表現や口語表現の言い回しを工夫する研究手法が実施された。
- 3) 上記対話知識の構築で試行錯誤による作業確立に時間がかかり、1月末までの達成率は50%であったが、10月から1月までの多くの試行錯誤により、効率的作業行程が確立できたので、平成18年度末までに目標数値の100%が達成できた。
- 4) 今後、上記4-2-3で汎用性が議論された商品や製品への問い合わせと応答についても重点的構築を進める。

イ) 質問知識の応答知識構築の実施状況

- 1) 個人嗜好に関する質問への応答知識の構築が実施された。例えば、質問文「ipodは何ですか？」に対して回答知識「音楽機器です」を構築する。
- 2) また、「ipodを知っていますか？」で「はい」と応えれば、「音楽機器」に関連する個人嗜好が理解できる。
- 3) 一般的知識として、食事分野を進めたので、多少の偏りはあるが、5,000件達成率は平成18年度末までに100%が達成できた。

4-2-4-3 応答知識ベースの研究開発のまとめ

- 1) 4-2-3の対話知識構築に比べて、限定した応答意図に対する応答文生成が構築の難易度は低いが、表層的な変化をどのように生成するかは、人手に全てを頼ることなく、半自動生成手法を確立する必要があり、この点が課題として残されている。
- 2) 平成19年度では、生成文法や格構造文法を使用した文生成の研究成果を調査し、応答知識構築に利活用を検討する計画である。

4-3 総括

4-3-1 研究開発の総括

以上の平成18年度の研究開発を総括すると表1の各サブテーマにおける研究項目と研究内容・目標に対する実施概要にとりまとめでき、以下の点を総括としてまとめる。

- 1) 全体として、目標数値は達成できた。
- 2) 知識構築を効率化するためには、データベース登録、分析評価に関するソフトウェア支援モジュールが必要であり、試行錯誤的な実験研究においては、継続して支援モジュールを拡充する工数も必要となる。
- 3) 多くの研究者が関わる知識構築においては、判断基準の統一が必要となり、異なるスキルの研究員や研究補助員（派遣研究員も含む）に対する学習期間が

- 必要不可欠であり、本研究開発では平成18年10月から12月まで必要となった。
- 4) 本研究における対話知識構築においては、要求意図知識の意図と応答知識の意図を関係づける組（要求意図から応答意図への変換定義）知識の構築が非常に難しく、大規模な用例文から如何に事例を検索するかが、今後の研究開発で重要な課題となってくる。
 - 5) 対話知識の汎用性については、研究員の意見の分かれるところであり、年齢や職業や立場により、対話の内容は変化するので、日常商品の購入要求は一般性のある分野であると思われる。
 - 6) 高齢者や身障者に対する一人暮らしの人に、癒し効果を与えるには、商品や物に関する具体的な応答を返すよりは、共感的で、精神的な応答知識が必要になる。この点については、平成19年度の研究開発を進める予定である。
 - 7) 個人嗜好を理解して、対話を進めることは重要であるが、個人嗜好の分類は非常に難しいといえる。現状のポータルなどの分類においても、固定された分類はないので、平成19年度では100分野程度の中規模分野について研究開発を進める計画である。
 - 8) 応答知識において、有限の固定された応答文では、いつかは同じ対話が出現し、システム利用において飽きてしまうおそれがある。この点に対して、方言や口語調の導入、立場や対話目的により、変化できる動的な応答文生成の研究手法が必要となるので、平成19年度の研究課題として残しておく。
 - 9) Web情報の利用としては、ブログ参加者の増加により、多彩な用例文が収集できるのは事実であるが、携帯ブログなどのように、表現自体がくだけていて、音声認識や音声合成が対応できない場合が多くなってきている。また、顔文字や絵文字などで表現する事例を参照すると、音声対話を主体とはするが、入力か出力が文字ベース（音声以外）である場合も、対話として成立させる必要がある。顔文字や絵文字の混在する文字ベースでも要求意図理解ができる知識構築も課題である。
 - 10) 対話知識において、誹謗中傷、猥褻表現など理解できないよりは、理解できた上で、応答において無視できる枠組みも必要であり、不適切表現全てを削除するのではなく、意図理解知識に組み込むことも本研究開発以外で取り組む必要がある。
 - 11) 現時点で、将来の事業化を検討するには、固定された対話知識では限界があると思われるので、本研究開発による対話知識を自動的に補完する研究開発も必要となってくる。この点は、現状のWeb知識からの情報抽出モジュールとの連携も見据えておく必要がある。

表1 各サブテーマにおける研究項目と研究内容・目標に対する実施概要

研究項目	平成18年度研究内容と目標の概要	実施の概要
「4-2-1）」研究開発管理 「4-2-1-1のア）」対話知識データベース管理	データベース管理モジュール、登録支援モジュール	開発組み込み終了 今後、適宜改良
「4-2-1）」研究開発管理 「4-2-1-1のイ）」対話知識データベース管理	ダンプファイル生成モジュール	開発完了 今後、適宜改良
「4-2-1）」研究開発管理 「4-2-1-1のウ）」対話知識データベース管理	データベース作業派遣 研究補助員の研究指導	指導完了 今後、適宜指導

「4-2-1)」研究開発管理 「4-2-1-4のア)」テキスト収集と分析	Web 情報からの用例文収集	実施完了 継続作業
「4-2-1)」研究開発管理 「4-2-1-4のイ)」テキスト収集と分析	文書情報の切り出し タグの除去	実施完了 継続作業
「4-2-1)」研究開発管理 「4-2-1-4のウ)」テキスト収集と分析	不適切表現のフィルタ リング	実施完了 今後、適宜改良
「4-2-1)」研究開発管理 「4-2-1-4のエ)」テキスト収集と分析	話題分野への分類	実施完了 話題分野の定義が課題
「4-2-1)」研究開発管理 「4-2-1-4のオ)」テキスト収集と分析	対話知識構築の事例文 1,000 万文構築	目標達成
4-2-2 要求意図知識ベースの研究開発「4-2-2-1のア)」意味共起への意味づけ開発	意味共起への意味づけ 開発 500 万件	目標達成
4-2-2 要求意図知識ベースの研究開発「4-2-2-1のイ)」生活分野の分野分別	生活分野（衣料、食事、 住居、健康）の分類	目標達成
4-2-2 要求意図知識ベースの研究開発「4-2-2-1のウ)」代表10分野の質問知識	代表10分野の質問知識 6,000 件	目標達成
4-2-3 対話知識ベースの研究開発 「4-2-3-1のア)」対話事例の収集	対話事例の収集 1,000 万文	目標達成
4-2-3 対話知識ベースの研究開発 「4-2-3-1のイ)」対話辞書データベースへの登録（変換知識対数/対話連鎖知識件数）	対話辞書データベース への登録 30 万対/3 万件	目標達成
4-2-4 応答知識ベースの研究開発 「4-2-4のア)」表層応答文の構築	表層応答文の構築 500 万件	目標達成
4-2-4 応答知識ベースの研究開発 「4-2-4のイ)」質問知識の応答知識構築	質問知識の応答知識構築 5,000 件	目標達成

4-3-2 研究開発に対する技術的評価

平成18年度末に、言語処理学会や情報処理学会を中心とする全国大会が開催され、それぞれの分野で最先端技術が発表される。本研究の技術的評価を比較するために、これら大会に参加して、関連する研究成果の比較評価を行う。

1) 意味解析技術

意味解析技術では、以下の論文と比較した。

- ・ 黒住亜紀子, 徳久雅人, 村上仁一, 池原悟. (2007). “結合価パターン辞書からの情緒を明示する用言の知識ベース化”, 言語処理学会第13回年次大会, pp. 39-42.

「情緒を明示する用言」については意図的な情報を一部もつが、要求意図には汎用化できない。また、大規模辞書構築手法については論じられていないので、本研究の手法が優れているといえる。

- ・ 長崎英紀, 古宮嘉那子, 但馬康宏, 小谷善行. (2007). “機械学習による代名詞「自分の人称判別システム”, 言語処理学会第13回年次大会, pp. 59-62.

人称判別の機械学習は、直接的に本研究とは関係ないが、興味ある研究であり、今後参考にできるものであった。

2) 対話技術

対話技術では、以下の論文と比較した。

- ・ 竹澤寿幸, 大熊英男, 葦莉豊, 清水徹. (2007). “音声認識結果と大規模コーパスに基づくユーザ意図に近い言語表現の検索”, 言語処理学会第13回年次大会

会, pp. 139-142.

本研究の要求意図知識に近い研究であり、コーパス利用についても関連性はあるが、言語表現に関する研究であり、要求意図から応答意図への変換を利用した本研究の発想は存在しなかった。

- ・ 奥村明俊, 池田崇博, 西沢俊広, 安藤真一. (2007). "ロボットとの対話によるマルチメディアブログ創作システム", 言語処理学会第 13 回年次大会, pp. 143-146.
- ・ 市川宙, 徳永健伸. (2007). "情報探索雑談における自然なトピック遷移の実現", 言語処理学会第 13 回年次大会, pp. 151-154.
- ・ 中里収. (2007). "対話中に期待される聞き手の反応について", 第 49 回人工知能学会言語・音声理解と対話処理研究会, pp. 33-38.

以上 3 件も、対話理解に関する研究発表であるが、最初の論文は話題がマルチメディアに限定され、大規模なものではない。2 番目の論文は、シナリオに関係するトピックの遷移であり、要求意図の研究成果は含まれていない。3 番目の論文は、聞き手に反応と期待に関する研究であり、手法としては最も参考になるものであるが、具体的実現方法に弱点があり、本研究の優位性は上記 3 点に対して明らかに保持されている。

3) マイニング技術

マイニング技術については、以下の論文と比較した。

- ・ 那須川哲哉, 宅間大介, 竹内広宜, 荻野紫穂. (2007). "コールセンターにおける会話マイニング", 言語処理学会 第 13 回年次大会, pp. 590-593.
- ・ 高橋大和, 廣嶋伸章, 古瀬蔵, 片岡良治. (2007). "意見性判定手法の評価と精度向上", 言語処理学会第 13 回年次大会, pp. 594-597.
- ・ 堤田恭太, 岡本潤, 内山清子, 石崎俊. (2007). "連想概念辞書とコーパスを組み合わせる曖昧性解消手法の検討", 言語処理学会第 13 回年次大会, pp. 614-617.

最初の論文は、コールセンターの意図理解に限定した技術では参考になった。2 番目の論文は、意見の判定であり、要求意図ではないが、手法としては関連性ある。3 番目の論文は、連想概念辞書に関して分野連想と関連すると思われたが、本研究の方向性とは異なっていた。言語を理解するマイニングとしては、本研究の要求意図より話題の分類の粒度が大きいので、本研究の手法の優位性は明らかである。

4) 知識獲得

知識獲得技術については、以下の論文と比較した。

- ・ 阿部修也, 乾健太郎, 松本裕治. (2007). "事態含意名詞の利用と共起パターンの学習による事態間関係知識の獲得", 言語処理学会第 13 回年次大会, pp. 883-886.
- ・ 吉永直樹, 鳥澤健太郎. (2007). "Web からの具体物の属性・属性値情報の自動獲得", 言語処理学会第 13 回年次大会, pp. 887-890.

最初の論文は、共起の学習に関する研究であり、本研究の意味共起の自動構築には参考になる。2 番目の論文は、受託企業ですでに確立している技術である。

5) 言い換え技術

言い換え技術については以下の論文と比較した。

- ・ 松吉俊, 佐藤理史. (2007). "体系的機能表現辞書に基づく日本語機能表現の言い換え", 言語処理学会第 13 回年次大会, pp. 899-902.
- ・ 加藤修平, 藤田篤, 佐藤理史. (2007). "句を対象とした構成的な言い換えの生成", 言語処理学会第 13 回年次大会, pp. 903-906.
- ・ 大田浩志, 山本和英. (2007). "理解補助を目指した動詞句の換言", 言語処理学会第 13

回年次大会, pp. 907-910.

言い換え表現は、要求意図と応答意図の変換を言い換えと考えれば関連するし、また用例文の言い換えを少し変化させると、同じ意味を持つ異なる用例文が生成できるので、用例文収集や応答文生成に関係がある。これらの論文は、本研究の対話知識の構築技術と比較する物ではないが、本研究の支援的技術として有用性は高い。ただし、受託企業で上記の基盤技術は実現しているのが現状であり、全体としては本研究手法の評価に直接影響するものではない。

4-3-3 研究開発の今後の課題

1) 情報通信技術によるストレス発生に対して

情報通信技術の発達には、物理的豊かさを提供するが、その反面、副作用が見逃されている。例えば、情報伝達の速さは、仕事の効率化を達成するが、それが原因でストレス社会を引き起こしている。本来の人間同士のコミュニケーションにおいては、情報過多になったときに、情報伝達の速さを遅らし、情報量を少なくする上に、「お疲れ様」「ご苦労様」などの「慰労」の表現が使用される。この意味でも、本研究課題では、仕事や要件のみを伝達するコミュニケーションだけでなく、日常対話に近い対話知識の構築で貢献できる可能性がある。

2) 情報通信技術による迷惑や犯罪発生について

不特定他者からの一方的な情報提供は、迷惑や悪戯メールに代表される副作用であり、さらには誹謗中傷や新しい犯罪への温床を与えることになっている。このような誹謗中傷や猥褻などの不適切な表現を含むコミュニケーションについては、対話の遮断を行える不適切表現によるフィルタリング機能を有する対話知識において貢献できるので、この観点を見据えて研究開発に取り組む計画である。

3) 少子高齢化社会による介護や育児の生活支援について

一人暮らしや引きこもりに対する支援と回復など情報通信分野の製品とサービスにおいて、貢献を果たすことは本研究の目標でもある。この支援については、「買い物」、「依頼」などの具体的な要求意図に対する応答意図を関係づける必要がある。この観点より、インターネットのポータル的な音声対話の応用が事業化として検討できるので、この点も見据えて研究開発に取り組む予定である。

4) 「心の豊かさ」を実現について

情報通信技術の一つにロボットコミュニケーションの普及があり、このサービスは、人間のもつ言葉による癒しや励まし能力を広く普及させるものである。上記1)のストレスに対する貢献と同様であるが、心の豊かさは、具体的な商品購入の要求意図に対する関連する応答のみならず、発話に対して、共感すること、相づちを打つこと、オウム返しをすることなどの極めて素朴な応答によって貢献できる場合が多いと考える。従って、相手の気持ちを認知したことを率直に答える対話知識の貢献も見据えていく計画である。

4-3-4 研究開発による波及効果

1) 医療福祉分野への波及効果

受託予定会社では、連携会社が広く普及させている医療電子カルテの音声入力分野に、電子カルテからの患者へのインフォームドコンセント、あるいは、薬投与の過誤などの医療ミスの防止支援など国民のリスクに関係する波及効果がある。

2) 身障者や高齢化社会への波及効果

音声合成技術連携会社は、声帯機能を失った患者の音声合成実用化を完成しているが、将来急激に高齢化する社会では、音声対話機能の低下した高齢者が多くなり、介護の負担増、コミュニケーション不足による事故などの問題が必ず生じるので、身障者や高齢者支援への波及効果がある。

5 参考文献

- Aho, A. V. and Corasick, M. J. (1975). "Efficient string matching: An aid to bibliographic search." *Communications of the ACM*, 18(6), pp.333-340.
- Aoe, J. (1989). "An efficient digital search algorithm by using a double-array structure." *IEEE Trans. Softw. Engr.*, SE-15(9), pp.1066-1077.
- Aoe, J., Morimoto, K., Shishibori, M., and Park, K-H.(1996). "A Trie Compaction Algorithm for a Large Set of Keys." *IEEE Transactions on Knowledge and Data Engineering*, Vol.8, No.3, pp.476-491.
- Atlam, E., Ghada, E., Morita, K., Fuketa, M., and Aoe, J. (2006). "Automatic Building of New Field Association Word Candidates Using Search Engine." *Information Processing & Management*, Vol.42, No.4, pp.951-962.
- Atlam, E., Fuketa, M., Morita, K., and Aoe, J. (2003). "Documents similarity measurement using field association terms." *Information Processing & Management*, Vol.39, pp.809-824.
- Ferret, L., Grau, B., Plantet, M. H., Illouz, G., Jacquemin, C., Masson, N., and Lecuyer, P. (2001). "QALC: the Question-Answering system of LIMSI-CNRS." In *The Ninth Text Retrieval Conference (TREC-9)*, pp.235-244.
- Fuketa, M., Lee, S., Tsuji, T., Okada, M., and Aoe, J. (2000). "A document classification method by using field association words." *An International Journal of Information Sciences*, 126(1), pp.57-70.
- Fuketa, M., Kadoya, Y., Atlam, E., Kunikata, T., Morita, K., Kashiji, S., and Aoe, J. (2005). "A Method of Extracting and Evaluating Good and Bad Reputations for Natural Language Expressions." *Information Technology & Decision Making*, Vol.4, No.2, pp.177-196.
- Fukumoto, J., Kato, T., and Masui, F. (2002). "Question Answering Challenge (QAC-1) Question answering evaluation at NTCIR Workshop 3." in *Working Notes of the Third NTCIR Workshop Meeting, Part IV: Question Answering Challenge (QAC1)*, pp.1-10.
- Haamond, K., Burke, R., Martin, C., and Lytinen, S. (1995). "FAQ Finder: A Case-Based Approach to Knowledge Navigation." in *Proceedings of the 11th Conference on Artificial Intelligence for Applications*, pp.80-86.
- Hatzivassiloglou, V. and McKeown, K. R. (1997). "Predicting the semantic orientation of adjectives." in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL-EACL)*, pp.174-181.
- Hatzivassiloglou, V. and Wiebe, J. M. (2000). "Effects of adjective orientation and tradability on sentence subjectivity." in *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, pp.299-305.
- Kadoya, Y., Morita, K., Fuketa, M., Oono, M., Atlam, E., Sumitomo, T., and Aoe, J. (2005). "A Sentence Classification Technique by Using Intention Association Expressions." *Computer Mathematics*, Vol.82, No.7, pp.777-792.
- 清田 陽司, 黒橋 禎夫, 木戸 冬子. (2003). "大規模テキスト知識ベースに基づく自動質問応答 -ダイアログナビ-. " *自然言語処理*, Vol.10, No.4, pp.145-175.

- ・小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. (2005). “意見抽出のための評価表現の収集.” 自然言語処理, Vol.12, No.3, pp.203-222.
- ・Kwon, O. and Lee, J. (2003). “Text categorization based on k-nearest neighbor approach for Web site classification.” An International Journal of Information Processing and Management, 39(1), pp.25-44.
- ・Kwon, O., Jung, S., Lee, J., and Lee, G. (1999). “Evaluation of Category Features and Text Structural Information on a Text Categorization Using Memory Based Reasoning.” Paper presented at the Proceedings of the 18th international conference on computer processing of oriental languages (ICCPOL '99), 1, pp.153-158.
- ・Lam, W., Ruiz, M., and Srinivasan, P. (1999). “Automatic Text Categorization and Its Application to Text Retrieval.” IEEE Transactions on Knowledge and Data Engineering, 11(6), pp.865-879.
- ・Lee, S., Shishibori, M., Sumitomo, T., and Aoe, J. (2002). “Extraction of field-coherent passages.” An International Journal of Information Processing and Management, 38(2), pp.173-207.
- ・Moens, M. and Uyttendaele, C. (1997). “Automatic Text Structuring and Categorization as a First Step in Summarizing Legal Cases.” An International Journal of Information Processing and Management, 33(6), pp.727-737.
- ・Morita, K., Kadoya, Y., Atlam, E., Fujita, Y., Sakakibara, A., and Aoe, J. (2003). “Word Classification and Systematization Using Co-Occurrence Word Information.” 7th World Multiconference on Systemics, Cybernetics and Informatics (SCI2003), Vol.12, pp.305-310.
- ・Nasukawa, T. and Yi, J. (2003). “Sentiment Analysis.” Capturing Favorability Using Natural Language Processing. in Proceedings of the second International Conference on Knowledge Capture (KCAP), pp.70-77.
- ・Pang, B., Lee, L., and Vaithyanathans, S. (2002). “Thumbs up? Sentiment Classification using Machine Learning Techniques.” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.79-86.
- ・Turney, P. D. (2002). “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews.” in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ALC), pp.417-424.
- ・Tokunaga, H., Atlam, E., Fuketa, M., Morita, K., Tsuda, K., and Aoe, J. (2004). “Estimating sentence types in computer related new product bulletins using a decision tree.” International Journal of Information Sciences, Vol.168, No.1-4, pp.185-200.
- ・渡辺 靖彦, 横溝 一哉, 西村 涼, 岡田 至弘. (2006). “メーリングリストを利用した質問応答システムのための知識獲得.” 自然言語処理, Vol.12, No.6, pp.25-44.
- ・Wilensky, R., Arens, Y., and Chin, D. (1984). “Talking to UNIX in English: An Overview of UC.” Communications of the ACM, Vol.27, No.6, pp.574-593.
- ・山田耕一, 溝口理一郎, 原田直樹. (1994). “質問応答システムにおけるユーザー発話モデルと協調的応答の生成.” 情報処理学会論文誌, Vol.35, No.11 pp.2265-2275.
- ・Yu, H. and Hatzivassiloglou, V. (2003). “Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Sentences.” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.129-136.

5-1 研究発表・講演等一覧
なし