

道しるべ：いまこそ話しことば処理技術の研究を



竹澤 寿幸 takezawa@slt.atr.co.jp

(株)ATR 音声言語通信研究所

まえがき

音声認識技術の性能が著しく向上し、携帯電話に音声ダイヤラーが搭載され、パソコンでも大語彙連続音声認識(ディクテーション)ソフトが動作するようになってきた¹⁾。また、日英・英日を中心に多くの翻訳ソフトが市販されている²⁾。このように、音声情報処理の観点からも、自然言語処理の観点からも、いま話しことば処理技術が花開くための下地が整いつつある。

しかしながら、音声情報処理と自然言語処理を組み合わせるだけで簡単に話しことばが処理できるわけではない。たとえば、「道しるべ：ここまできた音声認識技術」¹⁾と「道しるべ：機械翻訳技術とその適用」²⁾を組み合わせれば音声の翻訳³⁾ができるのだろうか。それだけでは十分でない。話しことばを処理する技術が必要なのである。

音声の翻訳より対象を広げて、たとえば、図-1のような音声対話システム⁴⁾を想定しよう。入力音声を問題解決器への指令(意味表現)に変換する機能を音声理解システムと呼ぶ。音声対話システム実現のためにはテキスト・音声・画像等を組み合わせた出力系も重要であるが、本稿では入力系に絞って話を進める。音声を利用した入力系の主要な要素技術として、まず音声認識技術と自然言語処理技術が挙げられる。さらに、その境界分野、もう少し前向きに言えばフロンティア領域の技術が必要であり、それを「話しことば処理技術」と呼ぶことにする。

以下では、まず、なるべく分かりやすい例を挙げながら、典型的な技術課題を紹介することで、研究分野の必要性を説明する。次に、その研究基盤であるコーパス作成を進めている大型プロジェクトを紹介する。さらに、文献、会議、研究グループ、関連URLなどの情報を紹介する。

話しことば処理技術とは

典型的な技術課題として「人間が喋る単位と機械が処理する単位」「話の流れの把握と利用」「音声認識誤り等に対して頑健な処理技術」の3つを紹介する。音声をテキストに変換するのが音声認識の主な課題である。テキストを対象に翻訳などの処理をするのが自然言語処理の主な課題である。これから紹介する3つの技術課題は、その狭間で必要な技術であり、このような技術なくしては音声認識と自然言語処理は決してつながらないのである。

人間が喋る単位と機械が処理する単位

人間はたいてい単語や文を意識して喋ったりしない。しかし、翻訳や解釈のためには処理の単位が必要となる。

仮に単文なら何でも受け付けられるようにできたとしても、たいてい単文とは子供でも分かるような簡単な文のことだろうくらいにしか解釈されない。そのため、た

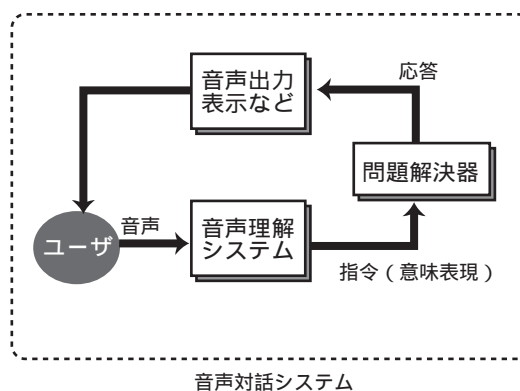


図-1 音声対話システムの基本構成

例えば「青じゃなくて赤です」のような発話がなされることになる。「青じゃなくて」が並列節を構成し、「赤です」が主節になるので、いまのは複文だという説明があっても、使えないシステムと判断されるだけであろう。

そこで、そのような形式的な制約を課すことなく、日常的な会話に現れるような言い回しを対象にすることが望ましい。そのため、たとえば、ATR音声翻訳通信研究所では、音声翻訳研究のために、日本語話者と英語話者の、通訳者を介した旅行会話等を収集し、データベース化している⁵⁾。その会話音声翻訳研究用データベース⁵⁾には、しばしば次のような言い回しが見受けられる。無音区間の長さの情報を[]内に記す。

- (1)お待たせいたしました。[440ms]シングル1泊1万円のお部屋でしたね。
 (2)お部屋を調べます。[170ms]しばらくお待ちください。

たいてい翻訳や解釈は文を単位として処理される。そのような単位を言語処理単位と呼ぶことにすれば、例(1)、(2)ともに1つの発話が2つの言語処理単位で構成されている。発話に自然に含まれる息継ぎ等の無音区間、つまりポーズ情報を利用して分割するという手段がまず考えられる。しかし、日本語には「っ」で表記される促音がある。会話文に含まれる促音で分割されないようにするには、目安として300ms以上の無音区間が必要となる。例(1)のポーズは300msより長いので、無音区間の情報だけで分割することができるが、例(2)のポーズは300msより短いので、無音区間の情報だけでは分割することができない。

そこで、たとえば文献⁶⁾では、境界位置の前2語と後1語の合計3語の範囲の品詞・活用形・活用型の情報を利用して、入力発話を言語処理単位に分割する手法を提案し、実験用音声翻訳システム³⁾で利用できる程度の性能を達成している。

話の流れの把握と利用

やりとりの多い対話の場合には省略される語句が多い。解釈や翻訳のためにはそれを補う必要がある。

ATRの会話音声翻訳研究用データベース⁵⁾を調べると、英語に翻訳する際に必須格となる話し手と聞き手に関する情報は、日本語の会話文ではほとんど常に省略されるといってよい。例を次に示す。省略されている話し手と聞き手に関する必須格を*で囲んで下線を引いた形式で補って示す。

- *こちらは*ニューワシントンホテルでございます。
- *私は*部屋の予約をしたいと思って*あなたに*電話したんですけど。

日本語の会話文ではこのような話し手や聞き手の情

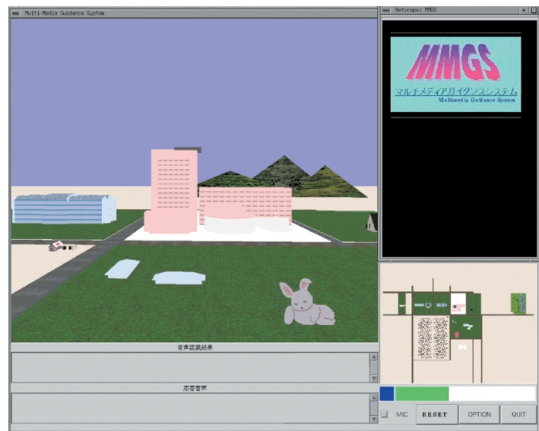


図-2 マルチモーダル入力、マルチメディア出力の案内システム: MMGS

報は明示的に言わない方が自然であり、その情報は「ございます」「したい」「思って」「電話したんです」等の述部の言い回しに含まれて、聞き手に運ばれているのである。しかし、とにかく日本語から英語に翻訳するためには話し手や聞き手に関する情報を必須格として明示的に伝える必要がある。そのため、英語に翻訳する際に必須となる主語等を日本語の述部等の情報から補う技術が研究されている⁷⁾。

図-1のような音声対話システムでデータベース検索など問題解決に必要なのは話し手や聞き手に関する情報ではなく、それ以外のガ格、ヲ格、ニ格などの情報である。ATRの会話音声翻訳研究用データベース⁵⁾から例を次に示す。書式は先ほどと同様である。

- チケット持ってるんですけども、*便を*変更したいんですけども。
- 5月3日は*チケットが*ありますか。

ATRの会話音声翻訳研究用データベース⁵⁾を調べると、発話のほぼ半分にこのようなデータベース検索に必要な語句の省略表現が含まれている⁸⁾。図-2に示すような画面に地図等を表示しながら、音声や画面上の指示動作により質問ができる観光案内システム⁸⁾において、次のような会話がなされたとしてよう。

- U1: 奈良県新公会堂はどれですか。
 S1: 奈良県新公会堂はこちらです[画面を動かす]。
 U2: 近鉄奈良駅から歩いて行けますか。

[]内は画面上の操作を示し、Uはユーザの発話、Sはシステムの発話をそれぞれ表すものとする。U2のユーザ発話は「奈良県新公会堂まで」が省略されていると考えるのが妥当である。案内システムが応答するためには、その省略を補うことが必須である。しかし、U2のユーザ発話を英語に翻訳するのであれば、聞き手が協調的な対話

用いた自然なヒューマンコンピュータインタラクションの実現を目指し、特に、適切なタイミングでの発話の理解および生成が円滑な対話の実現に不可欠であることに着目し、話し言葉を漸進的に処理する手法について研究が進められている。

さらに、関連するデータベースとして、車内音声対話データベース等が収集されている。運転操作に関する多様な情報の収集が可能な車内音声対話収集車を設計製作し、実走行中の音声対話を収録している。カーナビゲーションに関するオペレータとの対話音声や、音素バランス文の読み上げ音声等を収録しており、最終的には500名分の音声収録が予定されている。実走行中の車内音声対話データであるため、音声による車内機器操作に関する現実的な研究を進める基盤となるものと期待できる。

他にもいろいろな研究やデータベースの構築が進められている。詳しくは解説¹⁰⁾をご覧ください。

科学技術振興調整費による開放的融合研究推進制度 話し言葉の言語的・パラ言語的構造の解明に基づく「話し言葉工学」の構築

新聞記事などのあらかじめ用意されたテキストの読み上げ音声であれば、かなり高い精度で認識できるようになり、パソコン上で動作するソフトウェアも市販されている。しかし、言い淀み、言い直し、言い誤り、不正確な発音などの現象を含んだ自然な話し言葉を認識しようとする、性能が低下してしまう。このような話し言葉の音声認識・理解・要約に関する技術の確立を目的として、国立国語研究所と通信総合研究所が中心となって、大規模コーパスに基づく「話し言葉工学」の構築に関するプロジェクト²⁾が平成11年度に発足した。活動期間は平成15年度までの5年間で予定されている。

このプロジェクトでは、タスクを限定しないモノローグを主体として、延べ約800時間、日本語形態素で約700万語規模の話し言葉コーパスの作成を目標としている。大語彙連続音声認識研究のための統計的言語モデル構築のために十分な規模のデータとなることが期待できる。

モノローグ音声の収集は、データの集めやすさなどのために、国内の種々の学会での講演、大学での講義、国立国語研究所の設備を用いた模擬講演等を対象としている。模擬講演は一般の話者が簡単なメモを頼りに日常的な話題について行なったスピーチである。収録された音声は、書き起こし、日本語形態素解析等がなされる。詳しくは解説¹¹⁾をご覧ください。

ATR 音声言語通信研究所

実世界で利用可能な音声翻訳技術の実現を目指して(株)ATR 音声言語通信研究所³⁾が平成12年1月31日に

研究活動を開始した。研究活動期間は5年2カ月の予定である。

平成12年2月末に研究活動を終了したATR 音声翻訳通信研究所では、対話に代表される日常の話し言葉は、必ずしもきれいな理論的な枠組みのみによってとらえることはできないという認識に立ち、対話に関する具体的な表現を集めた大規模対訳コーパス⁵⁾を作成した。そして、音声認識、言語翻訳等の要素技術の研究を進めるとともに、実験用音声翻訳システム³⁾を構築し、種々の視点から評価実験を実施した。その結果、限定された話題であれば、コーパスに基づく手法により日常の話し言葉がかなりの精度で翻訳可能であることが実証された。

ATR 音声言語通信研究所の研究目標の1つは、限定された話題に対して実証されたコーパスに基づく対話音声翻訳技術が、実世界で利用可能であることを示すことである。もう1つの研究目標は、従来まったく未開拓であった講演等の独話の音声翻訳を可能とすることである。詳しくは解説¹²⁾をご覧ください。

文献等関連情報

パソコンに限らず、携帯電話やPDAが普及すると、話しことは処理技術の出番がますます増えるに違いない。しかも大型プロジェクトの研究活動の進展に伴い、使い切れないほどのコーパスがもうすぐ利用可能となる。従来の研究で利用されていた音声データが無響室や録音スタジオのようなノイズのない環境で収録されたものであるのに対して、現在進行中の大型プロジェクトで収集しているコーパスは、実走行中の車内での対話や学会での講演のように、いずれも実世界の音声データであり、実世界で利用可能な話しことは処理技術の研究が待ち望まれている。したがって、「いまこそ話しことは処理技術の研究を」なのである。そこで、最後に、興味を持った読者のために、文献、会議、研究グループ、関連URLなどの研究動向を調べるうえで有益な情報を紹介する。

教科書・参考書

話しことは処理技術は音声情報処理と自然言語処理の境界分野、もう少し前向きに言えばフロンティア領域に相当するため、音声情報処理から自然言語処理までの範囲を中心に、対話处理的な分野を加えて紹介する。志は高く、ダブルメジャーになるつもりでその範囲を理解すると良いだろう。書誌情報を次に列挙する。

- 古井貞熙: 音声情報処理, 森北出版(1998).
- 鹿野清宏, 中村 哲, 伊勢史郎: 音声・音情報のデジタル信号処理, 昭晃堂(1997).
- 中川聖一: 確率モデルによる音声認識, 電子情報通信学会(コロ

²⁾ <http://www.crl.go.jp/orc-speech/>

³⁾ <http://www.slt.atr.co.jp/>

ナ社 X 1988).

- 新美康永: 音声認識, 共立出版(1979).
- Lawrence Rabiner, Bing-Hwang Juang (古井貞熙監訳): 音声認識の基礎, NTTアドバンステクノロジー(1995).
- 北 研二, 中村 哲, 永田昌明: 音声言語処理, 森北出版(1996).
- 長尾 真(編): 自然言語処理, 岩波書店(1996).
- 田中穂積: 自然言語解析の基礎, 産業図書(1989).
- 野村浩郷: 自然言語処理の基礎技術, 電子情報通信学会(コロナ社 X 1988).
- 長尾 真(監修): 日本語情報処理, 電子通信学会(コロナ社)(1984).
- 長尾 真: 言語工学, 昭晃堂(1983).
- 堂下修司, 新美康永, 白井克彦, 田中穂積, 溝口理一郎(共編): 音声による人間と機械の対話, オーム社(1998).
- 益岡隆志, 田窪行則: 基礎日本語文法 - 改訂版 -, くろしお出版(1992).
- 村木新次郎: 日本語動詞の諸相, ひつじ書房(1991).
- 国立国語研究所: 話しことばの文型(1) - 対話資料による研究 -, 秀英出版(1960).
- 国立国語研究所: 話しことばの文型(2) - 独話資料による研究 -, 秀英出版(1963).

会議・研究会

国際会議や研究会も音声情報処理や自然言語処理を対象とするものをカバーするのが良いだろう。また, 人間と機械のインタラクションに関連するならば, 対話処理やヒューマンインタフェースに関する会議・研究会も適している。

音声情報処理関連の国際会議では, 隔年に交互に開催される ICSP (International Conference on Spoken Language Processing) と EUROPEECH (European Conference on Speech Communication and Technology) が大きなものである。自然言語処理関連の国際会議では, COLING (International Conference on Computational Linguistics) と ACL (Annual Meeting of the Association for Computational Linguistics) が大きなものである。たとえば「人間が喋る単位と機械が処理する単位」に関連する発表がこれらの会議で見受けられるので, 話しことば処理技術の分野は会議の話題に含まれていると考えられる。

国内の学会としては, 音声処理関連では日本音響学会の年2回の研究発表会, 言語処理関連では言語処理学会の年次大会が挙げられる。また, もちろん, 情報処理学会の全国大会も適当なものである。

研究会は関連のあるものの名称と URL を次に列挙する。

- 情報処理学会 自然言語処理研究会 (NL)
<URL: <http://cl.aist-nara.ac.jp/SIGNL/>>
- 情報処理学会 音声言語情報処理研究会 (SLP)
<URL: <http://winnie.kuis.kyoto-u.ac.jp/sig-slp/>>
- 情報処理学会 ヒューマンインタフェース研究会 (HI)
<URL: <http://www.ipsj.or.jp/sig/hi/>>
- 電子情報通信学会/日本音響学会 音声研究会 (SP)
<URL: <http://www.ieice.or.jp/iss/sp/jpn/sp-index-j.html>>
- 電子情報通信学会 言語理解とコミュニケーション研究会 (NLC)
<URL: <http://www.ieice.or.jp/iss/nlc/jpn/nlc-index-j.html>>
- 人工知能学会 言語・音声理解と対話処理研究会 (SLUD)

⁴ <http://www.ciair.coe.nagoya-u.ac.jp/db/>

⁵ <http://www ldc.upenn.edu/>

⁶ <http://www.lang.astem.or.jp/CSRC/>

<URL: <http://winnie.kuis.kyoto-u.ac.jp/sig-slp/>>

話しことば処理技術研究用コーパス

話しことば処理技術を研究するためにはコーパスが必要である。日本語の音声認識研究のための標準的なデータベースは日本音響学会音声データベース調査研究委員会で構築されている。そのウェブページ⁴には関連情報へのリンクが整備される予定である。

米国では LDC (Linguistic Data Consortium)⁵ という組織が音声やテキストのコーパスを収集・頒布している。米語が中心であるが, 最近では日本語や中国語等のコーパスも扱っている。コンソーシアム組織であり, 会員になるとその年度にリリースされたコーパスが入手できる。非会員に対しても一部のデータは非会員価格で販売することもある。

連続音声認識コンソーシアム

情報処理学会の音声言語情報処理研究会 (SLP) のもとに「連続音声認識コンソーシアム」⁶ が設立された。平成11年度まで実施されていた情報処理振興事業協会 (IPA) のプロジェクトの成果物である「日本語ディクテーション基本ソフトウェア」をベースに, オープンソースのソフトウェアとして一層の拡充・発展を図ることが目的である。

平成9年度から平成12年度まで毎年8月に会津大学において情報処理学会音声言語情報処理研究会 (SLP) の主催で開催された「音声認識セミナー 基礎からディクテーションまで」の内容をベースに近いうちに充実した教材の入った CD-ROM 付きの教科書が出版される予定である。コンソーシアムの会員になれば, その教科書を使った講習会やセミナーにも参加できる。

参考文献

- 1) 河原達也: 道しるべ「ここまでできた音声認識技術」, 情報処理, Vol.41, No.4, pp.436-439 (Apr. 2000).
- 2) 大山芳史: 道しるべ「機械翻訳技術とその適用」, 情報処理, Vol.40, No.4, pp.428-432 (Apr. 1999).
- 3) 横尾昭男, 竹澤寿幸, 山本誠一: C-STAR 音声翻訳国際共同実験, 情報処理, Vol.40, No.9, pp.908-909 (Sep. 1999).
- 4) 白井克彦, 竹澤寿幸: 音声対話処理, 人工知能学会誌, Vol.9, No.1, pp.48-56 (Jan. 1994).
- 5) 竹澤寿幸, 中村 篤, 隅田英一郎: ATR の会話音声翻訳研究用データベース, 音声研究, Vol.4, No.2, pp.16-23 (Aug. 2000).
- 6) 竹澤寿幸, 森元 暉: 発話単位の分割または接合による言語処理単位への変換手法, 自然言語処理, Vol.6, No.2, pp.83-95 (Jan. 1999).
- 7) 山本和英, 隅田英一郎: 決定木学習による日本語対話文の格要素省略補完, 自然言語処理, Vol.6, No.1, pp.3-28 (Jan. 1999).
- 8) 高橋和子, 竹澤寿幸: マルチモーダル対話システムにおけるインタラクション機構, 電子情報通信学会論文誌, Vol.J83-D-II, No.11, pp.2481-2489 (Nov. 2000).
- 9) 脇田由実, 河井 淳, 飯田 仁: 意味的類似性を用いた音声認識正解部分の特定法と正解部分のみ翻訳する音声翻訳手法, 自然言語処理, Vol.5, No.4, pp.111-125 (Oct. 1998).
- 10) 武田一哉, 板倉文忠: 文部省 COE プログラム統合音響情報研究拠点 (CIAIR) - 音声・音響情報処理の多角的研究 -, 日本音響学会誌, Vol.56, No.11, pp.748-751 (Nov. 2000).
- 11) 古井貞熙, 前川喜久雄, 井佐原均: 科学技術振興調整費開放的融合研究推進制度 - 大規模コーパスに基づく『話し言葉工学』の構築 -, 日本音響学会誌, Vol.56, No.11, pp.752-755 (Nov. 2000).
- 12) 山本誠一: エイ・ティ・アル音声言語通信研究所 (ATR-SLT) - 実世界で利用可能な音声翻訳技術の実現に向けて -, 日本音響学会誌, Vol.56, No.11, pp.756-759 (Nov. 2000).

(平成13年1月9日受付)