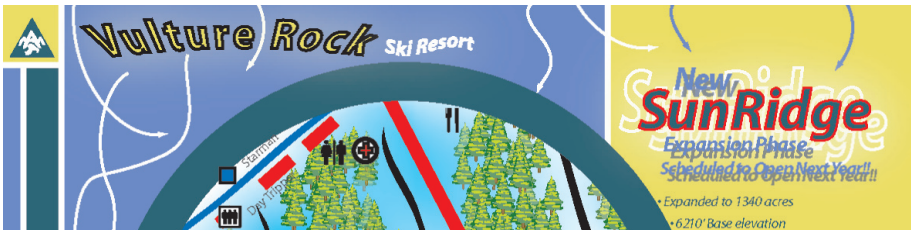
特集 5



ロボットの多言語使用の課題と現状 ―通訳ロボット―

飯田　仁
東京工科大学メディア学部
iida@media.teu.ac.jp

天野 真家
（株）東芝 研究開発センター
shinya.amano@toshiba.co.jp

多国語を操るロボットとして，最も有用と想定されるものの１つは，通訳ロボットである．通訳は多言語を処理するために，語用論といわれる文化，社会制約までも含む部門を扱う必要がある．さらに，音声通訳の場合，音声処理系と言語処理系の融合も必要になる．本稿では，これらの点に言及する．

自動通訳の歴史と現状

1983 年，NEC は TELECOM'83 において世界に先駆け自動通訳機の研究モデルを発表し，イメージ・デモンストレーションを行った．その後，ピボット方式のプロトタイプシステム作成や半音節単位に基づく音声認識手法などの実現を経て，1991 年に INTERTALKER と呼ぶ音声通訳の実験システムを開発した．英語のほかにフランス語，スペイン語の音声合成モジュールが稼働していた．当時は，限定された語彙による旅行会話が翻訳対象であった．

1987 年，東芝は，TELECOM'87 において，世界初の長距離自動通訳を介した会話の公開実験を 8 日間にわたってジュネーブの TELECOM 会場に訪れた一般見学者と，川崎にいる研究者との間で行った．この会話実験では，音声は使われず日英間のキーボード会話であったが，語彙も文法も会話対象も限定しない完全にオープンな日常的言語処理環境で行われた．

1993 年，ATR は日米独間の 3 局自動翻訳実験，さらに 1999 年，日米独韓中の一部擬似モバイルフォンを使った自動翻訳実験を行った．これに先立つ 1991 年，ATR/ITL（ATR 自動翻訳電話研究所），CMU（カーネギーメロン大学），シーメンス社ならびにドイツ UKA（カールスルーエ大学）の日米独の間で研究協力協定を結び，世界規模のコンソーシアム C-STAR I（1991 ～ 93 年：http://www.c-star.org/）を設立し，参加国語間の音声通訳共同実験を実施することを合意した．

1993 年の実験の成功を受け，第 2 期のコンソーシアム C-STAR II が立ち上がった．韓国 ETRI（電気通信研究所），イタリア ITC/IRST（科学技術研究所），フランス CLIPS がこれに参加し，日韓米独語の音声通訳実験と米独仏伊間の実験が実現した．これにより，音声通訳研究活動の基盤が構築された．音声ならびに言語のデータベースや対訳データ，翻訳用知識，対話音声収録の知見，音声通訳実験出力用の音声合成モジュールなどが整い始め，各研究機関がデータやツールを相互に共有できるようになった．さらに，その後第 3 期 C-STAR III に至り，

中国科学院CASの自動化研究所がこの活動に参加し，より広範囲の音声ならびに言語に関するソフトウェア技術が流通するようになった（C-STARの安定的なサイト：http://cstar.atr.co.jp/cstar-corpus/index.htm）．

第1期のC-STAR Iにおいては，各研究機関が日米独語の音声認識と各相手言語への翻訳を行い，テキスト・ベースの翻訳結果を相手研究機関に送り，各研究機関がそれらを自国言語に音声合成し，出力することを目指した．国際会議に参加登録する際の申込者と事務局とのやりとりを翻訳対象の対話に設定した．発話の自由度は少なく，話し言葉ではなく，書き言葉としての文法を満たしていることが前提であり，使用できる語彙も限定されていた．

一方，ドイツの大学を中心に，一部フィリップス社なども参加する大規模音声通訳研究プロジェクトVERBMOBILが1993年に始まった．DFKI（ドイツ人工知能研究所）を中心に33の研究グループが16のサブプロジェクトを構成した．このプロジェクトは第1期（1993～96年）と第2期（1997～2000年）とに分けられる．

第1期では，情報を一方的に獲得するための対話ではなく，対話者双方が交渉して最適解を求める対話を扱うことを目指した．ただし，対話の目的が明確に定まった会議日程の調整などであり，ビジネス対話における交渉などの対話を扱うにはまだ不十分であった．システム性能の点では，翻訳率だけをとると，80%以上の精度を実現した．個々の発話の訳が厳密に正しくなくても対話が成立するという点で，翻訳を介した対話を実行して目的を達成する割合で見ると，90%の対話目標の達成率であった．

1990年代前半までの自動通訳実験は，音声認識，機械翻訳，音声合成の各モジュールの後戻り処理がない線形結合により実行された．そして，そのシステム上の制約により，対話の目的を固定し，各発話が自ずと制限される状況において自動通訳を可能としたといえる．

多言語使用の問題点—自動通訳の場合

ここで例示する機械通訳を通した2カ国語会話の例は1987年，スイス—日本間で，文字ベースの自動通訳チャットとして8日にわたる公開実験から得られたものである．自動通訳を実現するために克服しなければならない多国語使用の際に生じるさまざまな言語現象を一部ではあるが紹介する．

言語を理解するということは，文字列としての記号そのものの解析だけでできるものではない．言語はこの世界の状況—自身の思考も含めて—を表現するものであるからである．言語で表現された内容は，空想も含めて，必ずこの世界の事象への何らかの写像になっている．そうでないものは，人間にとっては意味不明の表現と考えられる．ロボットと人間が言語でコミュニケーションを図る場合，言語で表現された内容は世界の事象と対応がとれている必要がある．多国語を使用できるマルチリンガルロボット実現の難しさは，単に言語理解の難しさだけではなく，複数世界の文化，社会規約などの事情に精通していなければならないという点にある．各世界の「在り方」は形態論，統語論，意味論，語用論のような言語構造にも各レベルで反映されているだろう．

語用論レベル

言語の特異性により引き起こされる通訳の難しさをいくつか例示しよう．この問題は，表面的には適切な「訳語選択の問題」としてとらえることができるが，訳語の選択基準の定式化が難しいという意味で各言語が持っている特異性に帰着する．

会話 1:

日本側：　会場の様子を教えてください．
機械通訳：please teach me the condition of meeting place.
英語側：　I don't understand.

この通訳結果が英語側の会話者に理解されなかった理由は大きく分けて2つある．「会場の様子」というような日本的曖昧さを持つ発話の翻訳を機械的に行うのはきわめて難しい．逐語訳はできても，相手には理解できるとは限らないし，誤解を招くこともあるだろう．大阪の「もうかりまっか」，「ぼちぼちですわ」というような文化にかかわる発話と類似の問題を含んでいるのである．「会場の様子」も具体的に，何を示すのかがこの発話では分からない．「大きいのか狭いのか」，「混雑しているのか，空いているのか」，「きれいなのか雑然としているのか」，「様子」の解釈は当事者の意識がどこにあるかで変わるのであるが，日本的曖昧さの下では，発話者は特にいずれかを意識しているわけではなく，何でもよいのである．さしあたり問われた側の主観で答えればよいのであるが，そうはいかない文化もあるということであろう．これは言語学的には語用論（プラグマティクス）に属する問題であり，日本人のプラグマティクス（思考・行動様式）を知らなければ通訳ができない．

さらに，"teach", "condition", "meeting place" の訳語も適切とはいえない．単独の語としての訳語を見れば，ほとんど問題がないのであるが，文全体としての外国語

との対応の難しさが如実に現れている例である．「唇」，「足」，「貢献する」のような基礎的単語でさえ，それぞれ"lip"，"foot"，"contribute"には完全には一致しない．工学などの学術用語以外は，恐らくほとんどすべての単語は異言語間で意味の被覆範囲の完全一致をみることはないだろう．これはその典型例である．「唇」は，粘膜の部分であるが，"lips"はそれを囲む一回り大きな可動部分，端的な違いとしては，口髭が生えている部分は日本語では「鼻の下」，英語では"upper lip"である．「足」は一般には"legs (+feet)"，「貢献」は，「良い」意味を含意するが，"contribute"はそうではない．「悪に加担」することも含まれる．

今の場合，"meeting place"がTelecom'87の会場であると推測できるかどうかは会話者個人の理解力に依存するが，"teach"，"condition"の理解は難しいだろう．

今の場合，

"Please tell me what the exhibition place is like."
(展示場はどんな感じですか？)

と通訳すれば，まだ多少は返事がきたかもしれない．訳語の問題は，このように語用論的な問題に帰着することも多い．

もう1つ例を挙げよう（論点になっていない部分の機械通訳は省略する）．

英語側　：Do you want something to drink?
日本語側：はい．
英語側　：What drink do you want?
日本語側：温かいコーヒーを飲みたい．
機械通訳：I want to drink warm coffee.
英語側　：warm coffee? not a hot coffee?

これは説明を要しないだろう．会話するロボットの難しさは，このようなプラグマティクスまでも獲得していないと，人間とのコミュニケーションに齟齬をきたすことになる．

意味論レベル

文書翻訳では一般に文が長いが，会話においては文は相対的に短い．長い文では係り受けの曖昧性の問題が発生するが，短い文では，この問題は軽微になる．しかし，単語の曖昧性の問題はここでも致命的な問題を惹起する．

英語側：Kondo returns from lunch now.
機械通訳：金堂は今，昼食から返ります．
(正：近藤は今，昼食から帰ります．)

このシステムの辞書には，「Kondo」に「金堂」，以下の複数のエントリが，「return」に「返す」以下の複数のエントリがあったが，正しい訳語を選択できなかった例である．

統語論レベル

会話文の解析は一般には非常に難しい．関西弁ワープロができない1つの理由は，どれほど需要があるのかという市場の論理もあるが，実際のところ大阪や京都で使われる言語の形態素解析さえできていないからである．まして，話し言葉で，かつ省略の多い会話文の統語解析は文法規則が詳しくは研究されておらず，きわめて難しい．特に省略の問題は，認識一般に存在する問題でもある．省略が存在していると分かるのは，省略前の完全な文が推定され，認識されているからである．しかし，完全な文を推定するには，元の（省略された）文を，省略されたものだと正しく認識していなければならない．この鶏―卵問題を機械で解決することは非常に難しい．実システムでは，省略であると認識する前に，文法的ではないとして解析に失敗してしまうからである．しかし，会話では，解析に従って通訳に失敗しても，会話相手の理解力に依存して会話が継続できる可能性もある．

(家族や友人の話をしている文脈が確立されている環境において)
英語側　：married, 2 children, many friends mostly female.
機械通訳：結婚された，2人の子供，多くの友達，たいていは，女性の

"married"の翻訳に失敗しているものの，この機械通訳の結果を見れば，多くの会話者は「結婚していて，2人子供がいます．友人は大勢いますが，大抵は女性です」という意味を正しく把握できるだろう．

語彙論レベル

英語側　：I love japan.（正：Japan）
機械通訳：私は漆器が好きです．

これは単純な語彙の問題である．入力を「Japan」と正しく行えば，問題なく通訳できた．しかし，この問題は音声によるロボットとの会話では大きな問題となるだろう．人間の会話に置いても文脈なしに，いきなりこの発話を聞かされた場合，「日本」か「漆器」かは判断でき

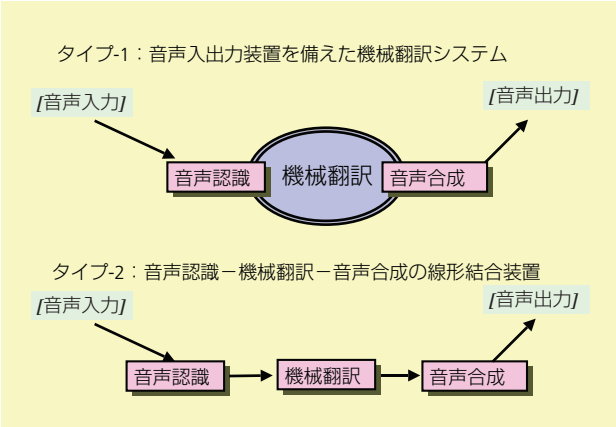


図-1　音声認識―機械翻訳―音声合成のかかわり

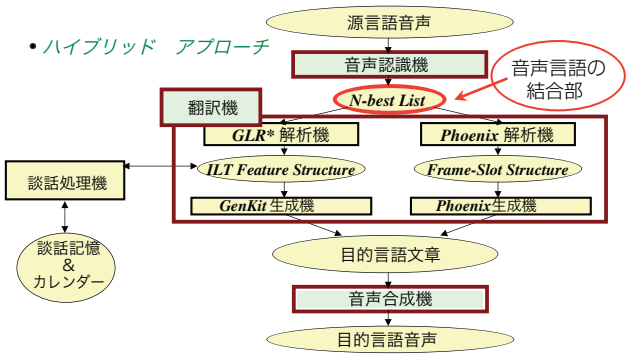


図-2　初期音声通訳システムJANUSの構成

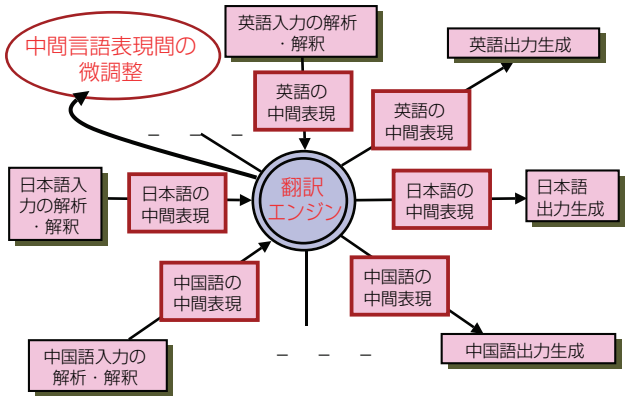


図-3　中間言語方式の考え方

ない．この語彙論的問題の解決には文脈解析が必要になる．

以下の章では，これらの問題がどのようにしてどの程度まで克服されているかを見ていく．

自動通訳システムの初期の構成

音声認識，機械翻訳，音声合成の各モジュールの線形の結合では実現できない通訳システムの課題と解決に向けた試みについて説明する．この章では，最初に説明した1990年代前半の音声通訳システムのシステム構成を見ることにより，次章で対話機能と通訳機能の同時実現の難しさについてまとめたい．

音声処理系と言語処理系をつなぐモデルとして，**図-1**に示すように，タイプ-1と，タイプ-2がある．タイプ-1は機械翻訳システムにおいてその入出力部分に音声認識と音声合成とを備えたシステムであり，音声通訳の原型と位置付ける．それに対し，C-STARの中心的な研究機関であったCMUの1990年代半ばまでの音声通訳システムJANUSは，音声認識モジュールが翻訳モジュールと直結し，さらに翻訳モジュールが音声合成モジュールと直結する音声認識・機械翻訳・音声合成の線形結合装置（タイプ-2）になっている．そのシステム構成を**図-2**に示すが，それは単なるタイプ-1の機械翻訳システムでないことが分かる．それは音声認識結果がN-best Listという複数の可能候補として翻訳モジュールへの入力になり，翻訳モジュールでは2種類の翻訳手法が試される点にある．その2種類の翻訳手法とは，文法解析を行い目的言語の文を生成する手法と，意味記述用にあらかじめ用意された意味フレームを参照して入力文を直に意味解析し，目的言語表現を作成するフレーム主導の翻訳手法である．特に，この後者においては，interchange formatという中間言語に相当する意味記述を対象領域内の語彙それぞれに知識として事前に用意することから，英語，ドイツ語，フランス語，イタリア語，韓国語，日本語から見て必要十分な記述内容を設計しておく必要がある．図-2における言語翻訳モジュールのうちの2つのパスがこの2種の翻訳手法を示す．なお，中間言語とは，いったんそこを経由して他の言語に翻訳するための人工言語である．多言語翻訳では，言語対分だけの処理系が必要になるが，中間言語を経由すれば，処理系の数は言語数に対して線形になる（**図-3**）．

一方，ATRのシステムでは，音声認識および音声合成の位置付けはCMUと変わらないが，**図-4**に示すように，翻訳モジュールはフレーズごとの入力に従って漸進的に翻訳処理を進める用例翻訳を中核として稼働する．この用例翻訳を使う多言語翻訳実行のために，各言語対ごとの用例翻訳向きの用例知識を用意することになる．音声通訳全体のシステムは，C-STARの第1期，第2期に対応してASURA，ATR-MATRIXと名付けられた統合システムとして広く公開された．

次に，VERBMOBILのシステム構成を見てみる．1期，2期全体に渡るプロジェクト活動の間，必要なモジュー

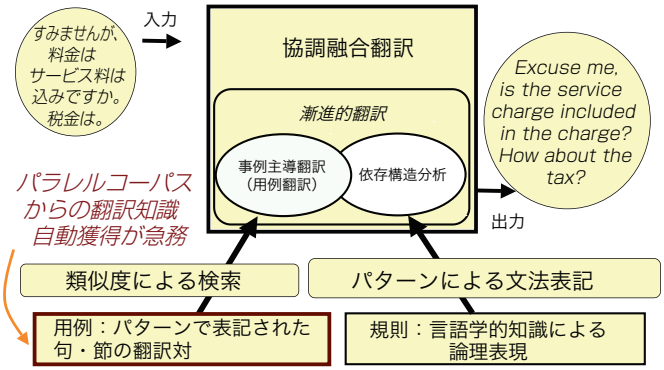


図-4　ATR-MATRIXにおける協調融合翻訳方式

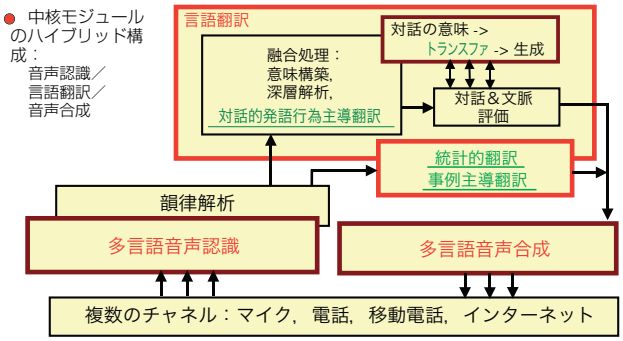


図-5　VERBMOBILにおけるシステム構成の概要

ルは適宜追加されていったと見なせる．最終のシステム構成は概略ではあるが，およそ**図-5**に示すようなモジュールが配備されている．このシステムでは，翻訳モジュールという一括りでとらえることは適当ではなく，音声認識と音声合成との間には，翻訳に関連する多くのモジュールが用意されている．対話構造と文脈の観点から翻訳結果を評価するモジュールがあり，対話構造と対話展開により生じる制約に基づく意味記述とその翻訳結果はこの評価モジュールによりその妥当性が判断される．

このように，VERBMOBILプロジェクトで目指したシステムの実現形態はタイプ-1，あるいはタイプ-2と異なる処理を目指し，音声通訳を狙ったシステム構成を当初から意識していたといえるであろう．

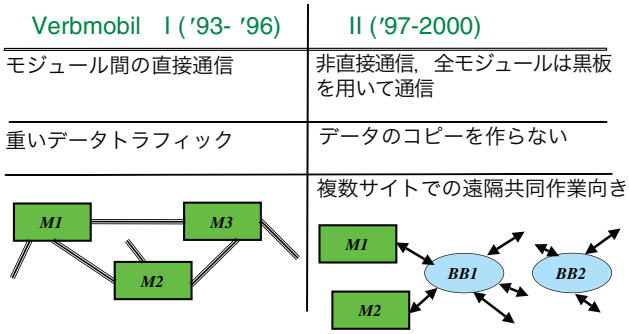
類似した試みとしては，MITのGalaxy Communicatorと呼ぶ音声対話によるボストンの町案内システム，ならびにその音声通訳への展開の試みも当初から音声対話を扱うシステム設計になっている点で，タイプ-1，タイプ-2と異なる．このシステムの制御機構を見ると，当初は音声認識モジュールを線形に翻訳モジュールとつなげてシステムを作成していたが，新たな制御機構では，その他のモジュールも含め，線形の接続関係は現れない．すべてのモジュールが1つの制御機構の下で必要なモジュールと連繋できるようになっている．

対話機能と通訳機能との同時実現の困難さ

音声対話のやりとりを扱うためには，その発話の状況依存性を解決することが大きな問題となる．対話理解という側面では，従来の文単位の言語処理の延長では文脈に影響される状況の把握が不十分である．そこには，社会活動をする上での生活文脈という状況も含まれ，状況を理解するためには，いわゆる知識処理という十把一絡げの単純な扱いでは及ばない．一方，通訳という側面からみると，タイプ-1のように1文単位の翻訳を1つの独立した文脈自由な解析ならびに翻訳により実行していたのでは，訳出された発話の連なりが文脈依存の対話を形成し得る保証がない．

そのような課題に対し，翻訳モジュールが文脈を管理し，その下で最適な訳文を作り出すためのシステム構成が望まれる．VERBMOBILでは必要性に応じモジュールの追加をしやすく設計することにより，文脈を考慮した翻訳を目指し，モジュールの強化を図っている．統計ベースの翻訳手法，事例主導の翻訳手法，対話の発話行為を翻訳対象に据えたDialog-act based 翻訳（対話的発語行為主導翻訳）手法が同時，あるいは前後して起動する．したがって，それらモジュールを制御するために，第1期においてはモジュールをエージェントと見立てて，マルチ・エージェント間の直のコミュニケーションを自立的にとるシステム構成をとった．しかし，モジュール間のデータの受け渡しなど重い処理を背負い込むことになった．そのため，第2期では，マルチ・ブラックボード制御というデータ・トラフィックの少ない機構を実現した．そこでは，モジュール間のデータ受け渡しを直に行うことはなく，常にブラック・ボードに書き込み，他のモジュールはそのブラック・ボードを参照することで情報の受け渡しを実現するようにした．その対比を**図-6**に示す．

このブラック・ボードによる制御機構に類似して，Galaxy Communicatorにおいては，大きな単位のモジュールを対象としたHubというシステム全体を制御するモジュールが用意される．たとえば，音声通訳という過程では，I/Oサーバがまず起動し，音声入力情報が音声認識モジュールで処理され，その結果が言語理解モジュールによって意味解釈され，文章生成モジュールにより言語表現化されて，文字から音声への変換モジュール



注）Mn：マルチ・エージェント
BB：ブラックボード

図-6　マルチ・エージェント対マルチ・ブラックボード

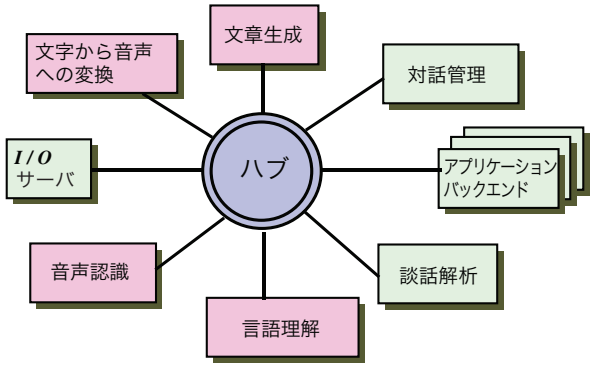


図-7　Galaxy Communicator のHub アーキテクチャ

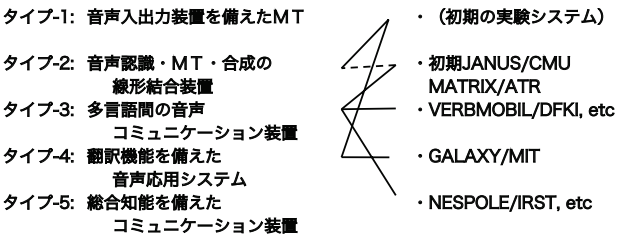


図-8　音声通訳システムの主要5タイプ

で音声に変換される．その結果はI/Oサーバが駆動することで，音声出力が実現する．Hubによるモジュール間制御の機構を**図-7**に示す．

しかし，音声対話の状況依存性をとらえるためには，まだ適切な観測情報とそれらの関係から推定できる状況について十分な知見が得られていない．そのため，現状では，対話の目的，対話のタスク，対話参加者の社会的役割などの要素を固定していくことにより，対象となる音声対話の状況ならびに文脈を自ずと制限し，対象対話における単語のperplexity[☆1]を押さえ，語義や訳語を自然に制限できるようにしているといえる．ここでは，対話の目的やタスクごとの分類ではないが，音声通訳システムのタイプとして挙げたタイプ-1，タイプ-2の分類に，VERBMOBILやGalaxy Communicatorなどのタイプを追加して，5種類のタイプ分けによる主要音声通訳システムを**図-8**にまとめる．破線はタイプ分けの根拠が弱いことを示す．また，NESPOLE（http://nespole!.itc.it）とはITC/IRST，CMU，CLIPSが中心となって新たに作った国際コンソーシアムであるが，音声・言語の多言語通訳処理を包括するマルチモーダル・インタラクションを目指している．その目的から，タイプ-5に位置付けた．e-commerce空間上の操作などを対象にした具体的な研究が進んでいる．

このように，音声通訳システムのタイプ分けをしてみると，これまでの研究はタイプ-3，ないしはタイプ-4を目指してきたといえる．現状の技術レベルから考えても自然な流れであったとみることができよう．しかし，音声ならびに言語の処理を統合して本来の音声の翻訳を実現しようという観点からそれらシステムをみると，まだ解明が不十分な研究課題が多い．また，音声を伴う日常的言語活動ではその言語表現の状況依存性が高く，話題などの変化が大きいことから，ドメインやタスクを限定すること以外には，状況依存性や話題の変化をとらえる適切な対処法をまだ見出していない．

通訳システムの能力を支える処理メカニズムのさまざまな工夫

エージェント間コミュニケーションを基本とした各種モジュールの設定とそれらの制御法，あるいはハブ構造や協調的融合機構などを音声通訳の全体のシステムの中で説明する．前章で挙げた対話機能と通訳機能の両立という課題は，音声・言語の統合の観点から見た課題，ならびに音声言語のオープンネスへの対応の観点から見た課題として論じることができる．つまり，状況を扱う問題は両課題に含まれ，特に対話機能の向上と状況依存の発話の翻訳という意味で，紋切り型の顧客対応の対話に相当する社会制度的な制約を緩和した下での自由な音声対話を扱うという課題を明確にすることにより本来の音声通訳機能実現に向かうといえる．

音声言語の統合の観点から見た課題

音声認識がうまく機能しない場合には，社会常識などの背景にある知識も総動員して，言語運用の知識などの支援を受け，利用者である人とインタラクティブに誤りの検知と修復を進める．そこで，認識誤りの修復を2つの課題から検討できる．

☆1 perplexityとは，ある言語の単語当たりのエントロピーをある単語に後続する単語数という尺度に置き換えた数値である．

1) 音声認識誤りの検出・修復

人間の判断はきわめて広域な情報によって行われる．先に述べたように語用論的なレベルまでも含むので，音声認識から言語解析に至る処理過程における判定等の決定は早い時点の限られた情報のみで判断せず，できる限り遅延し，可能性を保った扱いをすることが望ましい．これまで，複数の候補をn-best候補として選定し，それらを次のモジュールあるいは処理過程に引き継ぐことをしているが，n-best候補を確定すること自体，プロセスごとの断定的な処理と変わりなく，音声から言語へという全体の処理の中で決定されるべき情報を切り捨てる可能性を大いにはらむ．そのため，判定プロセスの遅延が重要であり，音声認識結果を，すでに 間違いを含んでいる可能性のある単語列で記述するのではなく，word-graphという単語に分節化されていない状態で記述し，言語処理側も単語以前の状態を記述単位とする処理能力の拡張が必要となる．

2) インタラクションを介した誤りの回避

従来より，認識誤りが生じても，利用者に聞き返したり，確認したりするインタラクティブな対話により，誤りを回避できるといわれてきた．しかし，聞き返されても，同じ発声や同じ句を発していたのでは，同じ誤りを繰り返す可能性は大きく，これらにかかわる認識誤り回避の対話は収束するよりも発振を招く危険性が大である．ここで，誤り部分を問いただし，聞き返すことはインタフェースとして基本の動作であることから，その前提の下でいかに対処するかが問題となる．つまり，その確認・修復行為を確実に実行できることが問題となる．そのために，オウム返しの聞き返しなどは不適当で，他の同等な内容を表す表現を使うことが有効であり，言い換えの技術を生かした対話による誤り修復の研究が重要となる．

音声言語のオープンネス対応への工夫

音声通訳のシステム能力を規定するドメインならびにタスクのポータビリティ（異なるドメインやタスクに適用できる能力が高い，あるいは差異を学習などによってすぐに吸収できる能力を表す．融通性に優れた能力ともいえる）を高めることで，随意性の大きい音声言語活動に対処することが求められる．そのための試みとして，最近のJANUS/CMU上の意味概念フレームによる発話内容の記述とそのパターンの自動学習の研究などがある．また，ATRの音声言語コミュニケーション研究所では，前身の研究所が取り組んできた翻訳手法である用例主導の翻訳をよりコーパス中心の翻訳手法に拡充して，翻訳精度のみならず翻訳の融通性とシステムの立ち上げ効率を改善している．また，EUにおいては2002年2月からシーメンス社，ドイツIBM，ノキア社などが参加して，日常社会の中の広範囲なドメインにおける音声通訳のための多言語対訳データ作りを目指すプロジェクトLC-STAR（http://www.lc-star.com）が開始された．

寡黙な通訳ロボットと，饒舌な通訳ロボット

前章では通訳時の誤りをどのように回復していくかの戦略を述べた．最も自然な方法は，人間が行っているように発話が分からない場合，発話者に聞き直すものである．ただ，すでに指摘したように聞き直すにも戦略が必要になる．この章では，実例を用いつつ，そのような戦略の検討を「寡黙」と「饒舌」の対比で見ていくことにする．ここで用いた「寡黙」と「饒舌」は，通訳ロボットが発話者に発話内容に関する問い返しを行うかどうかという意味である．行わない場合，前者，行う場合，後者であると規定しておく．

人間の同時通訳者が対話者と同席している場合，同時通訳者は，通訳中に時に話者と確認などをしているが，国際会議のブース内の同時通訳者は，そのようなことはしない．その代わりにあらかじめ講演内容に対する確認をとるような作業を行う．これは，リアルタイムでの通訳情報が得られない分，あらかじめドメインやタスクの特徴をとらえた通訳の機構を作っているのである．いずれにしても，人間の通訳は，完全に寡黙であることはない．一方で，通訳ロボットの場合，音声認識，言語理解などのあらゆる過程で，人間に問合せをしたい状況が発生する．この場合，饒舌な通訳ロボットとなる．通訳ロボットが寡黙でいられるか，いられないか，前記の実験で得られたデータを用いて考察する．

寡黙な通訳ロボット

ここでいう「寡黙」の意味をもう少し詳しく特定すると，音声処理系，言語処理系が通訳できないような言語処理上の問題に遭遇した場合，その解決を行うために発話者に問い合わせることをしないことを意味する．処理系が処理に失敗し，結果が誤りであることを認識していても，発話者に何の問い返しも行わず間違ったままの通訳を相手に行うのである．深刻な問題が発生する可能性がある政治的場面での通訳などの場合ではなく日常会話の場合には，この戦略はある程度利用可能である．この場合処理系の誤りはそのまま相手に伝わる．単語が辞書に登録されていないような語彙論的誤りでは単語が原語のまま相手に伝わる可能性が高く，この場合，相手は理解できないだろう．統語論的誤りが生じていれば，訳文は文法的文章にならない．以上の場合，聞き手が誤りの

存在に気づくことができ，何らかの対応措置を会話相手に求めることができる．

意味論的誤りの場合，少し様相が異なる．現実世界と照合して，まったく意味不明な発話内容の場合，聞き手は恐らく相手が間違えたのだろうと推測できる．しかし，処理系が意味論的間違いをした結果，意味の理解できる他の解釈が提示された場合，相手が間違いに気がつくことはかなり難しい問題になる．この発話の正当性は，文脈でしか確認できない．しかし，文脈による確認は不安定性を持つ．単に相手が突然話題を変えただけかもしれない．このことを相手に問い合わせると，それは，会話に対する会話，つまりメタ会話になり，メタ会話はさらなるメタ会話を生むかもしれず，終いには会話の破綻につながる可能性が高い．寡黙な通訳は，問題の解決を会話者に委ねてしまう結果，このような問題を持つ．

饒舌通訳ロボット

したがって，通訳ロボットは寡黙ではいられない場合が存在する．例を挙げよう；

(1) 日本語側：特にスポ-ツ番組が好きです．
(2) 機械通訳：I like スポ-ツ program especially.
(3) 英語側　：Could you translate this kanji, please?
(4) 日本語側：はい．プログラムのことです．
(5) 機械通訳：It is the thing of a program.

これは，タイプライタによるチャットであるために起きた現象である．「スポーツ」とタイプすべきところを，長音記号を誤りハイフンにしてしまい「スポ-ツ」となった．この現象は，音声会話の場合には，単に音声認識の誤りと等価であり頻繁に起きる可能性がある．寡黙通訳の結果として「スポ-ツ」はそのまま相手に提示された．相手は，「日本人は漢字を使う」という知識を持っていたので，(3) のように誤りの部分を「この漢字」として指定した．原文に含まれる漢字は，「特」，「番組」，「好」である．「スポ-ツ」は漢字ではないので，問合せを受けた日本側は，この 3 字の中から「番組」だと適当に推測して，このメタ会話に (4) で応えた．この機械通訳は (5) になる．通訳された文の意味は，まったく理解されず，この会話はこの後，破綻してしまった．

この場合，通訳ロボットには，「スポ-ツ」が辞書になく，通訳できないことは分かっているので，発話者に問い返しをすれば問題は恐らく起きなかっただろう．

それでは饒舌なロボットの方が優れているかというと，一概にそうは言えない．今度は人間とロボットとの間で，メタ会話が起きる可能性がある．人間が言い換えなどの適切な戦略を行えば問題は解決されるだろうが，これは結局は通訳ロボットの音声・言語処理系の頑健性に依存することになる．

通訳ロボットの現状と将来

最近 2，3 年をみると，MT-SUMMIT VII (Sep. 1999) においてMT and Speechという特別セッションが設けられ，MT-SUMMIT VIII (Sep. 2001) では音声通訳の招待講演が用意された．最もホットなところでは，40th ACL (July 2002) においてSpeech-to-Speech Translationと題するワークショップが併設され，最近のEUの音声通訳プロジェクトの紹介やNESPOLE! の実用面での評価法，前章で触れたATRでの最近の成果などの発表があった．1990 年代の研究から振り返ると，音声処理研究，機械翻訳研究，多言語応用研究など 1 つの研究領域にとどまることなく，音声通訳研究が関連領域にまたがる研究開発活動へさらに広がっているように見える．EUではスペイン，イタリア，ドイツによる用例翻訳による音声通訳の試みであるEU-TRANS Projectを，米国ではDARPAが戦闘状況下でのマン-マシンの音声通訳を中核とする情報収集システムを目指して，BABYLON projectを立ち上げ，新たな展開を探ってきたといえる．そのような状況下で，音声通訳は研究面でも実用化の面でもまだまだ多くの課題を抱えている．

音声入力が限定的である場合，カウンターサービスに類する場面で多言語対応が可能であろう．また，音声入力が任意の位置から受理されるようになり，話者認識も可能になれば，多言語による家電を含む一般のリモートコントロールや多言語会議などが可能になるだろう．ロボットという実在を明確にするならば，多言語に通じた秘書として，ヘッドセットマイクなどを使わなくとも適宜メッセージを話してやれば，翻訳音声メールを送信してくれるなど，異文化間コミュニケーションのためのエージェントロボット，あるいは基幹パスを提供することが可能になるだろう．

参考文献

1) Miike, S. et al.: Experiences with an On-Line Translating Dialogue System, Proc.of ACL, pp.155-162 (1988).
2) Morimoto, T. et al.: ATR'S SPEECH TRANSLATION SYSTEM: ASURA, in Proc. of EUROSPEECH'93, pp.1291-1294 (1993).
3) Iida, H.: Prospects for Advanced Speech Translation, in Proc. of MT-SUMMIT VII, pp.107-113 (1999).
4) Proc. of Speech-to-Speech Translation Workshop, 40th ACL (2002).

(平成 15 年 11 月 12 日受付)