特　集「ビッグデータとAI」

Massive AI時代の音声・言語技術

Speech and Natural Language Processing in Massive AI Era

中村　哲　奈良先端科学技術大学院大学，情報通信研究機構
Satoshi Nakamura　Nara Institute of Science and Technology. / National Institute of Information and Communications Technology.
s-nakamura@is.naist.jp, http://isw3.naist.jp/Contents/Research-ja/43_lab-ja.html

Keywords: machine translation, speech translation, spoken dialog system, web information processing.

1. インターネットと成長的フィードバック

現在，Webには，全世界の全言語の情報を加えるとゼタバイト（10の21乗）オーダのデータが存在するといわれている[IDC 11]．この中には，情報案内や情報提供のページ，多言語のページ，E-Commerceのページ，さらには，一般の利用者のブログ，ビデオ映像，マルチメディアデータ，最近ではtwitterなどの情報が含まれている．これらWebデータは多様な利用者によって日々生成，更新されており，日々変化していく実社会を射影した世界を構成している．最近では，これらの情報をも瞬時にスマートフォンで取得できるようになってきた．この背景には，多くの要素技術の進化と，大規模なデータを利用する仕組みの整備がある．スマートフォンに代表される小型デバイスの進化と高速の通信が可能な通信インフラにより，携帯端末では不可能な膨大な計算が，遠隔地に存在する多数の並列計算機で行えるようになった．それと同時に利用者が発信するデータをクラウド上に集積することが可能となり，加速度的にデータが増加した．センサによって集積できるフィジカル情報も含めると巨大なサイズであり，これら巨大集積データの処理技術として，「ビッグデータ（Big Data）」研究の重要性が語られるようになった[喜連川 11]．いわゆるビッグデータには，Webサイトデータ，マルチメディアデータ，ソーシャルメディアデータ，カスタマデータ，オフィスデータ，ログデータ，オペレーションデータ，センサデータがあり，全体では2011年で2ゼタバイト，2016年に約8ゼタバイトに拡張するという予測されている[IDC 11]．特に，科学研究（宇宙，気象など）のセンサデータはE-Scienceデータと呼ばれている．ビッグデータ研究は，これらのビッグデータを活用することによる異変の察知や，近未来の予測などを通じ，利用者個々のニーズに即したサービスの提供，業務運営の効率化や新産業の創出を可能とし，これらの活用により米国ヘルスケアで年間3,000億ドル，EU公共セクターで年間2,500億ユーロ，位置情報データの活用により年間6,000億ドルの消費者価値を創出できるという予測もある[MIC 11]．

このような膨大なデータを扱うことの重大さは世界の科学技術政策の担当機関でも認知されている．最近，米国の情報関連研究施策を取りまとめている機関であるNITRD（The Networking and Information Technology Research and Development Program）が，情報通信に関する10の重点課題を以下のとおりとりまとめている[NITRD 12]．

- ◇ Big Data（BD）
- ◇ Cyber Physical Systems（CPS）
- ◇ Cyber Security and Information Assurance（CSIA）
- ◇ Health Information Technology Research and Development（Health IT R&D）
- ◇ Human Computer Interaction and Information Management（HCI&IM）
- ◇ High Confidence Software and Systems（HCSS）
- ◇ High End Computing（HEC）
- ◇ Large Scale Networking（LSN）
- ◇ Software Design and Productivity（SDP）
- ◇ Social, Economic, and Workforce Implications of IT and IT Workforce Development（SEW）
- ◇ Wireless Spectrum Research and Development（WSRD）

この順番はアルファベット順ではあるものの，“Big Data”，“Human Computer Interaction and Information Management”が重点課題にあげられている．

一方，人工知能の研究として知識表現，知識獲得，パターン認識，言語処理，推論などの研究が行われてきた．種々の理論研究を実用に供する際に大きな課題となってきたのは，トイ課題でない実際の大きな課題にどう取り組むかであった．具体的には実世界の知識・規則の獲得と生成，パターンの獲得と生成をいかに行うかが大きな課題の一つであった．近年，大規模なデータを収集し大規模な機械学習を適用することで多くの課題が解決され始めている（DARPA Grand Challenge [Thrun 07]）．

本稿ではWeb上およびインターネットを介した大規模データによる人工知能的研究について述べる．先に述べたように，インターネット上にはゼタバイトのデータ

や種々のサービスが存在するが，利用者がそれによる反応をインターネットにフィードバックすることで，インターネット上のデータやサービスはさらに急激な速度で成長していく．例えば，E-commerce サイトの推薦機能では，多くの人が使えば使うほど，自分と同じ購買行動をする人が類型化でき，その購買履歴から的確な商品の推薦を行うことができる．後述する音声認識，機械翻訳などの技術においても，多くの人がスマートフォンでサービスを利用すれば，その発話データを用いて性能を向上させることができ，ますます，利用者が増える成長的ループを構成できる．

インターネットを介した Massive Parallel な大規模成長ループを利用して，Massive Data（Big Data）を継続的に収集し，それらを対象に進める人工知能の研究は，非常に新たな取組みといえ，本稿では，このような研究を Massive AI 研究と呼ぶことにする．あまり使われていないようだが，Webster 辞典でも“Massive”は，1.forming or consisting of a large mass: 2a : large, solid, or heavy in structure b : large in scope or degree とされているので，このような枠組みをうまく表現しているように思う．

インターネット上の Web データは，多様で膨大であり，多様な利用者によって日々生成，更新されている．この情報の大洪水の中で必要な情報を取り出すための中核的技術が言語関連技術である．Web 上の情報の検索のほか，ネット通販の商品推薦，仮名漢字変換の候補提示，トレンド分析，情報分析など多くの分野で言語処理が注目され利用されている．

音声・言語の関連では，米国国防省で最近まで進められていた，情報関係のプロジェクト GALE（Autonomous Language Exploitation）[Soltau 09] は，アラビア語と中国語のニュース，ラジオなどの放送を，音声認識，英語への機械翻訳，情報抽出を行う技術を開発するもので，それまでの人間により行っていた同種の作業の自動化を目的にしたもので，大規模データを用いた音声・言語処理の技術開発の一例といえる．

本稿では，著者がこれまでに従事してきた多言語音声自動翻訳についてフォーカスしながら研究の動向を述べる [中村 08, 中村 11]．音声・言語処理の研究は大規模データと統計モデルにより大きな発展を遂げ，世界初の旅行会話音声翻訳ネットワークサービスとして運用が続いている．ネットワーク上の利用者が使うほどデータが集まり性能が高度化する，この技術はビッグデータに基づく AI 技術と捉えることができる．ここでは，コミュニケーションのための音声・言語処理，Web 上の多言語情報抽出のため音声・言語処理の現状と今後について述べる．

2. 音声・言語処理の対象としての Web 情報

前述のように，Web には，日々変化していく実社会を射影した多種多様な情報が存在する．これらの情報の関連づけ，検索，提示などを行うためには，実社会の言語情報そのものを取り扱う大規模な処理系が必要となる．また，昨今ではビデオ動画のようなマルチメディアコンテンツが多く蓄積されており，音声処理も重要になっている．実際，インターネット上のデータ通信量では今やビデオ動画が圧倒的な量を占有しているといわれている．

言葉は，ヒトを中心にした多様な情報の相互の関連付けのために極めて重要なツールとなっている．特に“モノ”と“コト”をいかにつなぐかが重要である．従来は用途ごとにその関連付けを人手で付与するか，自動的なアノテーション技術の精度向上を待つしかなかった．しかし，インターネット上の E-commerce サイトでは，出品者が買い手により魅力的な情報を提供するために自らアノテーションを施していく．この活動により，より多くの言語アノテーションが加速度的に増加する．インターネット上の E-commerce の広がりは極めて興味深いといえる．

もう一つの観点は音声・言語処理利用がこれらの Web コンテンツの情報処理と利用者のインタラクションの高度化に不可欠ということである．一般に，利用者は検索すべき情報をあらかじめクリアにしているわけでなく，あるいは，検索方法，検索キーワードをうまく準備できないため，適切な情報を効率的に獲得することができないことが多い．検索エンジンで下手なキーワードを入力すると何十ページもの検索結果が容易に提示されるが，せいぜい最初の 1 ページしか見ずに必要な情報がとれないことが多く生じる．

このような課題のため，自然言語，音声・言語による対話インタフェースが研究されている．特に，昨今，音声認識の語彙数，性能が向上したことにより，スマートフォンで Web 音声検索（Google 音声検索など [Schalkwyk 00]），音声対話（Apple の Siri, NICT の AssisTra [水上 11]，質問応答システム一休 [鳥澤 10]）や音声翻訳（NICT の VoiceTra [中村 11]），などのサービスが登場し注目を集めている．

膨大な Web コンテンツは音声・言語処理の高度化にも利用可能である．Web 上にある膨大な音声データ，テキストデータをクローリングして利用することで，音声認識の音響モデル，言語モデルの高度化が可能となる．また，処理系をスマートフォンのような端末とサーバを接続した形態にすることで，多くの使用者からのデータを集約し集合知として利用することで，「使えば使うほど賢くなる」まさに Massive AI 的なシステムを構築することが可能となってきた．

3. インターネットと音声・言語研究

著者の関連する音声・言語処理に関する研究動向について述べる．1970年代の後半に，雑音のある通信路モデルに基づき，音声の時間的，特徴量的な揺らぎを統計的にモデル化し，モデルパラメータの推定と，認識を統一的な枠組みで計算できる隠れマルコフモデル（HMM: Hidden Markov Model）の原型が提案された [Jelinek 76]．対象言語の音声コーパス（本稿ではメタ情報付き音声・テキストデータをコーパスと呼ぶ）とテキストコーパスを大量に収集することでモデルパラメータの推定を行い入力音声の認識を行う技術である．その後，多くの改良を経て最近では不特定話者の連続音声認識が，かなり高いレベルまで到達している．この統計的なモデリングは，言語翻訳でも1990年代に入り本格化し，統計翻訳という名前で一つの大きな流れを構成するに至っている．統計翻訳では，同じ意味をもつ異なる言語の文対（対訳コーパスと呼ぶ）と，対象言語のテキストコーパスを大規模に収集し，これらから統計翻訳のモデルパラメータの推定を行う．これまでのルールベース翻訳ではルールの作成が不可欠だったが，これが不要になり，ドメイン（翻訳の対象範囲）の拡大，翻訳言語対の増加が劇的に容易になった．これらの二つの技術は，いずれもコーパスを収集し，そのコーパスからモデルパラメータを推定する点が特徴である．この際，統計モデルのパラメータ推定のためのコーパスは多いほど良いため，ⅰ）質の良い大規模なコーパスをいかに多く効率的に収集するか，ⅱ）対象となるアプリケーションのドメインに合致したモデルを構築するために対象ドメインから学習コーパス，辞書をいかに効率的に多く収集するか，ⅲ）対象ドメインだけでなく言語現象として日々変貌する言語現象をコーパス，辞書として，いかに捉えるか，が課題となっている．

音声・言語処理に関連する代表的な海外のプロジェクトの例として，前述の米国のDARPAのGALE（Global Autonomous Language Exploitation, 2006-2011）プログラム [Soltau 09] が有名である．2012年からは，さらに一般の中国語とアラビア語の各種方言の話し言葉を対象にリアルタイムで音声翻訳，情報抽出，検索処理するためのBOLT（Boundless Operational Language Translation）プログラム [BOLT 11] が進んでいる．このプロジェクトでは英語からの情報検索，対話や意味解析もターゲットに含んでいる．また，少量のデータから新たな言語の音声認識，翻訳を実現するBabelプロジェクトを開始している [BABEL 11]．

GALEプロジェクトでも取り組まれているように，情報抽出，検索処理はWeb，ニュースなどの大規模データ情報と対象とした処理で不可欠になる．このような検索，情報抽出，要約，質問応答の四つの技術はまとめて情報アクセス技術と呼ばれており，米国では1980年代から国防省やNISTにより大規模プロジェクトとして遂行されてきた（TREC [Lupu 11]，GALE [Soltau 09] など）．

4. 機械翻訳の研究

自動翻訳の歴史 [Hutchins 95] を振り返ると，第二次大戦直後1949年に米国においてWeaverによって研究の提案がなされ，1954年にGeorge Town大学とIBM社による人類初のコンピュータによる翻訳実験が行われた．このシステムは，当時最先端の計算機であったIBM709に，250単語の対訳辞書と6個の単語を参照して，目的言語側の単語・語順を制御する構文変換の規則とからなるロシア語から英語への翻訳システムであった．その後，精力的に種々の研究が行われた．1964年に，米国政府内に，機械翻訳に関してAutomatic Language Processing Advisory Committee（ALPAC）という委員会が設置され機械翻訳の研究の将来性に関する調査と議論がなされた [Koerner 95]．その結果，1966年に機械翻訳研究の将来性に対する否定的なレポートが出され，その後約10年研究資金の多くが停止，削減されたため米国における研究は大きく滞った．これが有名なALPACレポートである．ヨーロッパ，カナダなどでは研究が続けられ研究は息を吹き返した．この規則翻訳に基づいた機械翻訳方式（Rule-based Machine Translation）は，現在も商用システムの基本技術として広く活用されている．しかし，システム構築の初期コストが膨大で，多言語の翻訳には向かないという問題は残されたままであった．

2000年以降，ハードウェアの処理速度や記憶容量が格段に進歩したこと，言語に関わるデータ（大量の文章の蓄積や辞書）が計算機上に集積できるようになったこと，音声認識で言語データを統計的にモデル化する手法が大成功したことなどを受けて，機械翻訳の研究において，対訳データ（同じ意味の原文と訳文の対を集めたもの）から，翻訳システムの知識を機械学習により自動的に構築する技術が生まれた．これらはコーパスベース翻訳技術（Corpus-based Machine Translation）と呼ばれ，対訳データから検索した類似した対訳文を修正して翻訳を行う用例翻訳，対訳データから2言語間の対応関係をモデル化する翻訳モデルと表現の自然さをモデル化する言語モデルを導出し両者に基づく確率を最大化するように翻訳する統計翻訳（Statistical Machine Translation）がある．

統計翻訳は，IBMの研究者らにより提案されたモデルであり，原言語の文fが与えられたとき，条件付き確率$P(e|f)$を最大化する目的言語の文eを求める問題として定式化されている [Brown 93]．条件付き確率$P(e|f)$を最大化するeを求めるためには，ベイズ則により，$P(f|e)P(e)$を最大化するeを求める．

$$\hat{e}=\arg\max_{e} P(e|f)=\arg\max_{e} P(f|e)P(e)$$

ここで，確率 $P(e)$ は対象言語の言語モデルで，条件付き確率 $P(f|e)$ が翻訳モデルである．この定式化は，「雑音のある通信路モデル」(Noisy Channel Model) を翻訳に適用したものである．この確率 $P(f|e)$ を最大にする翻訳結果を求めることを最尤復号，探索プログラムをデコーダと呼んでいる．翻訳特有の単語の順序交換，その際にどれぐらい離れた距離まで許すか，単語ベースかフレーズベースで翻訳するかなどにより 5 段階のモデルが提案されている [Brown 93]．この方法に基づき種々のデコーダが開発されており，その中で，欧州の研究プロジェクトの中で開発されたデコーダ Moses は，最新の研究結果が試されているオープンソースとして多くの研究者に利用されている [Koehn 10].

[Goh 10] らの NICT のシステムでは，フレーズベース統計翻訳を基本として，いくつかの実用のための機能追加 (固有名詞対訳の登録機能，翻字など) がなされている．図 1 に旅行会話における学習データと日英，英日機械翻訳性能 (BLEU により測定 [Yasuda 08]) の関係を示す．対象タスクによるが 50 万文のドメインが一致したデータが必要になることがわかる．

より多くの対訳コーパスを収集するために，Web クローリング (Web データの自動収集) により収集していく方法，翻訳者を入れた枠組みにより品質の高いコーパスを収集する方法がある．前者の技術では一般に一文単位の対訳が得られないため文書単位で対応する日本語と外国語の文 (コンパラブルコーパス) を自動的に抽出し，統計翻訳モデルの学習をする研究が行われ，新聞，科学技術論文，特許などさまざまな分野で活用されている．

一方，後者についての代表的な例は，[内山 09] らが翻訳のホスティングサービス「みんなの翻訳」である．著作権フリーな Creative Commons の中で統計翻訳の出力を翻訳者が修正し，それらの文を翻訳者が利用しながら，対訳文数，質を増加させるという新たな試みを行っている．この枠組みも多くの利用者が使い，フィードバックすることで機械翻訳の語彙，性能が成長的に改善されていく Massive AI 的な枠組みといえる．

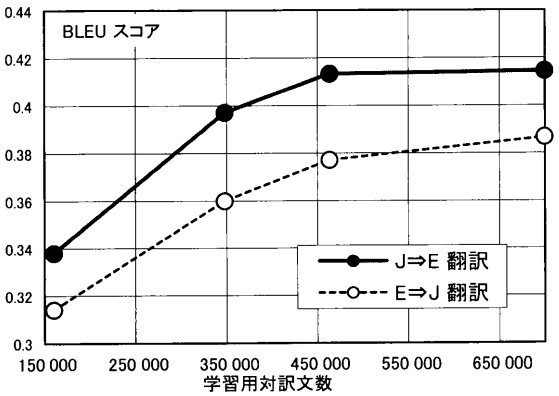

図 1　機械翻訳における学習データと性能の関係

5. コーパスベース多言語音声翻訳

音声自動通訳は原言語で発話された音声を目的言語の音声にリアルタイムで音声認識，言語翻訳，音声合成して出力する，音声処理技術と自然言語処理技術の統合技術である．本稿では，意図を理解した人間の通訳に対して，1 文ごとに音声認識，機械翻訳，音声合成処理をテキストベースで接続する技術を音声翻訳と呼ぶ．

音声翻訳がリアリティのあるシステムとして紹介されたのはテレコム'83 に NEC がラボラトリーモデルとしてデモを行った際といわれている．その後，音声翻訳実現のためには，長期的な基礎研究を行う必要があるという認識のもとに，1986 年には日本政府の主導により ATR 自動翻訳電話研究所が設立され，国内外からさまざまな研究機関の音声言語研究者が参画した．1991 年には ATR, CMU, UKA (現在 KIT) をベースに音声翻訳に関する国際コンソーシアム C-STAR が組織され，1993 年には，ATR，CMU，シーメンスによる世界 3 地点を結んだ世界初の音声翻訳実験が行われた．ATR プロジェクト開始の後，ドイツで Verbmobil [Whalstar 00]，欧州で Nespole! [Metze 02], TC-Star [Tc-Star 02]，また，米国で TransTac [Back 07], GALE [Soltau 09] の各プロジェクトが進められ，日本が世界を先導し大きな影響を与えた研究分野となっている．世界の音声翻訳のプロジェクトの研究成果としては，ATR の技術が世界初のネットワーク音声翻訳商用サービス「しゃべって翻訳」につながり，アーヘン工科大の卒業生が Google 翻訳の研究開発の中心になっている．その後，日本の携帯電話キャリアによるものなど，数々の商用サービスとして提供され始めている [中村 08, 中村 11].

現在主流の音声認識システムは，統計的音声認識手法を基礎としている．音声中の個々の音素の振舞いや，単語の並びなどを統計モデルで表現し，さまざまな仮説の中から最も高い確率が与えられる単語列を認識結果として出力する．したがって，これらの統計モデルの精度が音声認識性能に大きく影響する．[中村 11] の NICT の音声翻訳システムでは，日本語，英語，中国語，インドネシア語，ベトナム語の音声認識の音響モデルは，表 1 にあるように日本語 380 時間，英中国語 250 時間，イ

表 1　音声コーパスサイズ

言語	話者数	発話 文章数	発話 時間	タスク
日本語	4 500	226 673	387.6	旅行会話
英語	930	236 737	257.6	旅行会話
中国語	540	213 352	251.0	旅行会話
インドネシア語	400	84 000	79.5	ニュース音声
ベトナム語	30	23 424	40.5	ラジオ音声

ンドネシア語，ベトナム語は40～80時間の音声データにより学習を行い，言語モデルは，日本語，英語100万文，中国語は50万文，インドネシア語，ベトナム語は16万文の旅行会話文から学習している．しかしながら，このよう事前の大規模なモデル学習用のデータ収集はコストが極めて高く，実際の利用シーンのデータとのずれいかに小さくするかが重要となる．

2009年に総務省の委託研究で全国5地方での音声翻訳実証実験が行われ，期間中に収集された日本語約6万文，英語約17 000文，中国語約15 000文について人手による書き起しを行い，音響，言語モデルの両方について実データ学習による評価が行われた[河井 10]．

図2に音声翻訳の出力文の主観評価値（S：ネイティブ並み，A：申し分ない，B：まずまず，C：許容範囲，D：意味不明，の5段階で主観評価した際のS～Cの比率）を示す．地域に応じた固有名詞，固有表現の追加と，実際の設置場所，応用システム形態での実データが性能改善を実現していることがわかる．

2010年に音声翻訳技術の性能調査と改善を目的として，スマートフォン（iPhone）用の音声翻訳アプリケーションVoiceTraを開発し，2010年7月29日より無料公開した[中村 11]．VoiceTraは，21言語の双方向翻訳に対応しており，うち5言語については，音声による入出力が可能である．2012年10月末時点でのVoiceTraのダウンロード数およびアクセス数は70万および850万件となっている．VoiceTraは，ネットワーク型システムを採用しており，ユーザが発話した音声と翻訳結果はログとして音声翻訳サーバに蓄積される．表2に，音声ログ100件を無作為抽出し，聴取により内容を分類した

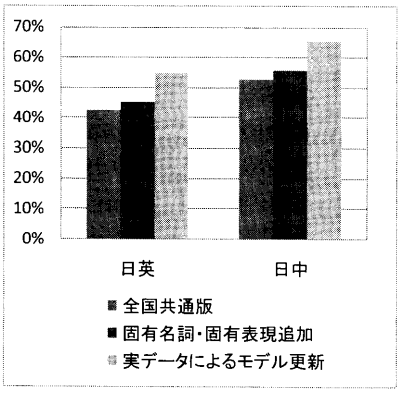

図2 音声翻訳の性能改善

表2 VoiceTraの発話の分類

分類	比率〔%〕
無音	11
無効発話（非音声など）	11
明確な旅行会話	9
旅行会話と解釈可能	42
旅行会話以外の内容	27

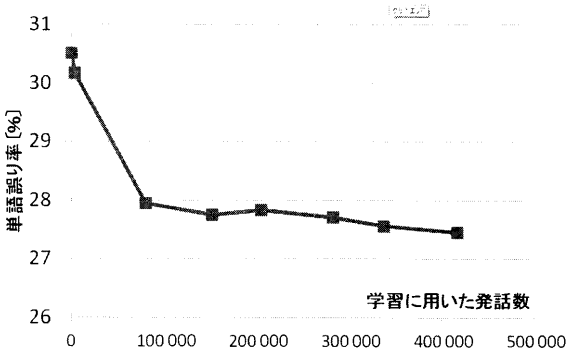

図3 VoiceTra実データ学習による性能改善

結果を示す．約半数が旅行会話的な発話の翻訳に利用されていることがわかる．

図3にVoiceTraにより収集された実データを用い，音声認識の発話単位の信頼度を用いて，教師なし学習を行った場合の単語誤り率を示す．単語誤り率は45万発話でも削減が続いており，実際の使用による実データを収集しながら教師なしで統計モデルの更新，改良を続けていくことの重要性が示唆されている．

6. 国際標準化

さらに多くの多言語の利用者に実際に音声翻訳を使ってもらい，大規模に多言語の音声を収集するためには，よりオープンで国際的な枠組みが必要となる．また，翻訳対象となる言語の知識および大規模な音声・言語資源が必要であることから，一つの組織が全言語対，全ドメインの音声翻訳を実現することは困難である．ネットワークを介して世界中に分散している音声認識，音声合成，翻訳モジュールを接続しすべての言語対を音声翻訳できるネットワーク型音声翻訳で世界的な協働の枠組みができればこの問題が解決できる．これに向けた，ネットワーク型音声翻訳の実現に必要なモジュール間通信プロトコルとデータフォーマットの国際標準化，および，この規格を用いた協働の枠組みについて説明する．

(1) アジアにおける国際標準化活動

アジアにおけるネットワーク型音声翻訳の先端研究を目的として，2006年にATR（日本），ETRI（韓国），NECTEC（タイ），BPPT（インドネシア），CASIA（中国），CDAC（インド）と共同でアジア音声翻訳先端研究コンソーシアム（A-STAR）を発足させ，2008年にはIOIT（ベトナム），I2R（シンガポール）が加盟して8か国の研究機関と共同研究を行ってきた[中村 07, Sakti 11]．

2007年にはアジア・太平洋電気通信標準化機関（ASTAP）[ASTAP 07]にて標準化活動を開始し，2009年7月，世界で初めてインターネットを介して異なるアジア言語を話す複数話者間で旅行対話を対象とした音声翻訳システムを用いて実時間音声対話に成功した．このネットワーク型音声翻訳技術をアジアにとどまらず

世界で用いられる標準化技術にすべく，標準化活動をASTAP から ITU-T に移行した．

(2) ITU-T における国際標準化活動

2009 年 10 月 ITU-T の SG16, WP2, Q21 (Multimedia architecture) /Q22 (Multimedia applications and services) において，ネットワーク型音声翻訳技術の標準化を始動した．NICT の堀 智織氏がエディタとなり，(1) ネットワーク型音声翻訳のサービス要求条件と機能，および，(2) アーキテクチャにおける要求条件の 2 件の勧告草案を作成し，2010 年 10 月 14 日に勧告 F.745 および勧告 H.625 としてわずか 1 年の期間で承認された [ITU-T 10]．図 4 にネットワークの構成図を示す．携帯電話などの端末と，音声認識，翻訳，音声合成のサーバ群がネットワークを介して接続される [ITU-T 10]．

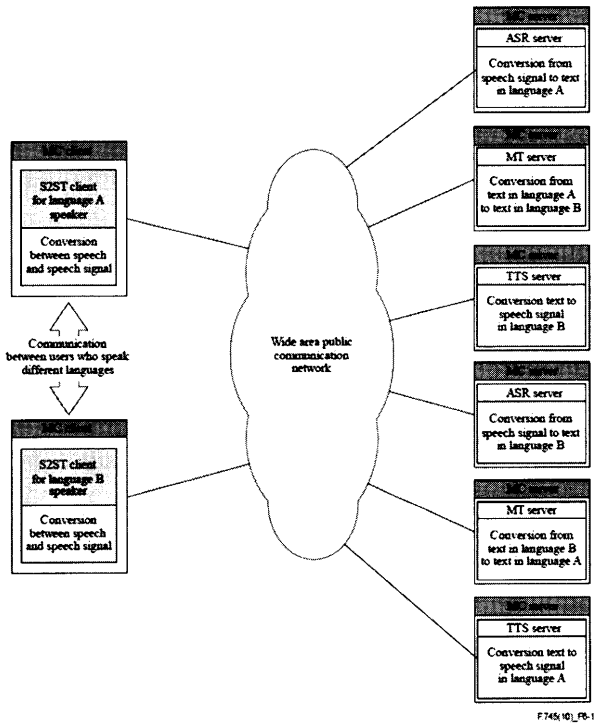

図 4　ネットワーク音声翻訳の機能モデル

(3) U-Star コンソーシアム [U-Star 10]

U-Star は，アジアの A-Star を世界に広げるべく拡張された音声翻訳コンソーシアムで，現在 23 か国・26 機関が加盟し，23 言語の音声翻訳技術の研究が行われている．ITU-T での勧告に従い，U-Star として「多言語音声翻訳サービスを提供するシステム」を共同で開発している．このシステムは，加盟機関の音声翻訳サーバを，ネットワーク型音声翻訳通信プロトコルで相互接続する．

7. Web データと音声対話研究

本章では音声対話システムについても若干触れておく．まず，1990 年代に統計モデルの基づく大語彙連続音声認識システムの研究開発が世界的に行われた．大語彙化のため，大量の音声データによる詳細な音響モデルと，大語彙辞書と言語モデルのための大規模テキストデータが必要となった．当初は新聞記事コーパスが，2000 年代頃からは Web テキストが使用された．音声認識ソフトウェア Julius のディクテーションキットに含まれている言語モデルは，2003 年頃に Web テキスト (20 億形態素) から構築されている [河原 03]．さらに最近では，検索エンジンを用いて特定のドメインに特化した言語モデルが構築されるようになった．

音声対話システムとしては，定型データに対し規則に基づく対話管理から，最近では Web 上の非定型情報を対象にした対話システムや，統計的な対話管理モデルによるシステム [Young 06] が研究開発されている．例えば，「情報コンシェルジェ」[河原 08] では，Wikipedia の文書に基づいて京都案内を行うことができる．そのほか，ピッツバーグのバス運行案内 [Raux 05]，スマートフォンを利用した京都の旅行案内 [水上 11]，スマートフォンの情報案内および Web 検索 [Schalkwyk 00] などのシステムがある．

対話システムではないが，インターネットを使ってサービスとフィードバックを行う仕組みの研究もある．Podcastle は Podcast を音声認識してテキスト化する．これにより利用者は検索などが容易になる一方で，この音声認識結果を不特定多数のユーザに修正してもらい，それによりモデルの更新を行うことが可能となる [Ogata 10]．

8. 高度言語情報融合フォーラム

平成 21 年に組織を越えて音声・言語の資源やツールを共有しつつ，言語の壁を感じさせない情報処理，コミュニケーションを実現するための技術の進歩発展・促進を図るため高度言語情報融合フォーラム (Alagin: Advanced Language Information Forum) が設立された [Nakamura 10]．現在，民間企業 (92 社)，大学・研究機関および国の関係者 (164 者) が会員となっている．辻井潤一前会長から著者が現在二代目の会長を拝命している．

このフォーラムでは，音声・言語処理技術，情報検索や信憑性判定を含めた情報分析技術，これらの技術の前提となる言語資源 (辞書，コーパスなど) のツールや言語資源を広く会員に提供すべく活動している．

現在公開中の言語資源・サービスは，

A-1.　文脈類似語データベース
A-2.　動詞含意関係データベース
A-3.　負担・トラブル表現リスト
A-4.　上位語階層データ
A-5.　単語共起頻度データベース
A-6.　日本語パターン言換えデータベース
A-7.　日本語異表記対データベース

A-8. 日本語係り受けデータベース
A-9. 基本的意味関係の事例ベース
A-10. 京都観光ブログの評価情報付与データ
B-1. 日英翻訳エンジン学習・評価用対訳コーパス
C-2. 係り受け解析システム（CNP）用中国語解析モデル
C-3. 意見（評価表現）抽出ツール用モデル
D-1. カスタム単語集合作成サポートサービス
D-2. 意味的関係抽出サービス

さらに，音声資源については，

1. 日本語高齢者音声データベース
2. ノンネイティブ英語音声データベース
3. 中国語音声データベース
4. 京都観光案内対話データベース
5. 日本語小学生音声データベース【音響モデル学習用】
6. T3デコーダ（バイナリおよびソース形式）単語数50万語彙を実時間で高精度に処理可能な，「重み付き有限状態トランスデューサ」を用いた大語彙連続音声認識ソフトウェア
7. 日本語音声データベース　音素バランス503文，雑音データベース，会話文など
8. 日英・日中バイリンガル独話音声データベース

などがあり，さらに，楽天データセットを配布している．

これらは，学習用の初期データとして貴重なもので，多様な研究に利用できるものと考えている．

9. ま と め

本稿では，Web上およびインターネットを介した大規模データによる音声・言語処理，特に，機械翻訳，音声翻訳，音声対話について述べた．先に述べたように，インターネット上にはゼタバイトのデータや種々のサービスが存在するが，多くの利用者がそれによる反応をインターネットにフィードバックすることで，インターネット上のデータやサービスはさらに急激な速度で成長していく．実フィールドで多くの利用者データを収集し，それらのデータから自動的にシステムの改善を行う成長的ループをつくるという枠組みは，種々の分野で進んでいる．

その中で言語処理分野は，インターネット，Web上のテキスト情報を利用することで最も恩恵を受け，貢献してきた分野といえる．本稿では機械翻訳と，それに続いて異なる言語を話す人々のリアルタイム異言語コミュニケーションである音声翻訳について述べた．音声翻訳の究極のゴールは人間の通訳者のように意図を解した同時通訳である．他方，GALEプロジェクトで代表される技術は，コミュニケーションというより情報分析技術といえる．情報分析技術は，経済，生活，あらゆる業種の業務に必要とされている．多言語の情報検索，情報抽出，サマライゼーション，レコメンデーションなどの情報分析技術は，今後ますます重要性を増すと思われる．これら同時通訳，高度な機械翻訳や情報分析には，意図，意味を解した処理が不可欠である．今後，Web上のテキスト情報，ネットに射影された非定型の知識，モノとコトの関係の利用とインターネットの利用による成長的な枠組みにより，この課題も少しずつ解決されていくと思われる．

謝 辞

原稿の作成にご協力いただいた情報通信研究機構ユニバーサルコミュニケーション研究所の皆様に感謝する．また，音声翻訳に関連する部分は[中村 11]の筆者の執筆部分を中心に再構成したものである．共著者の皆様に心から感謝する．

◇ 参 考 文 献 ◇

[ASTAP 07] http://www.apt.int/ASTAP-SNLP

[BABEL 11] Babel Program Broad Agency Announcement, Solicitation Number: IARPA-BAA-11-02, IARPA (April 7, 2011), https://www.fbo.gov/utils/view?id=ba991564e4d781d75fd7ed54c9933599

[Back 07] Back, N., et al.: The CMU TransTac 2007 eyes-free and hands-free two-way speech-to-speech translation system, *Proc. IWSLT 2007* (2007)

[BOLT 11] Amendment1 DARPA-BAA-11-40 BOLT - April 15.pdf (2011), https://www.fbo.gov/utils/view?id=625ffb386d2d2dd821b3fc446d28b59e https://www.fbo.gov/index?s=opportunity&mode=form&id=69bc8c59d8dfdc9907d13ac6b024ee43&tab=core&_cview=1

[Brown 93] Brown, Peter F., Della Pietra, V. J., Della Pietra, S. A. and Mercer, R. L.: The mathematics of statistical machine translation: parameter estimation, *J. Computational Linguistics - Special issue on using large corpora: II archive,* Vol. 19, Issue 2, pp. 263-311, MIT Press Cambridge, MA, USA, (1993)

[Goh 10] Goh, C. L., Watanabe, T., Paul, M., Finch, A. and Sumita, E.: The NICT translation system for IWSLT 2010, *IWSLT 2010,* pp. 139-146, Paris, France (Dec. 2010)

[Hutchins 95] Hutchins, W. J.: Machine translation: a brief history, *Concise history of the language sciences*: from the Sumerians to the cognitivists. Edited by E. F. K. Koerner and R. E. Asher, pp. 431-445, Oxford: Pergamon Press (1995)

[IDC 11] IDC 2011 Digital Universe Study: *Extracting Value from Chaos* (June 2011)

[ITU-T 10] F.745 "Functional requirements for network-based speech-to-speech translation", ITU-T (2010) http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-F.745-201010-I!!PDF-E&type=items
H.625 Architecture for network-based speech-to-speech translation services (2010) http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-H.625-201010-I!!PDF-E&type=items

[Jelinek 76] Jelinek, F.: Continuous Speech Recognition by Statistical Methods, *Proc. IEEE*, Vol. 64, No.4, pp. 532-556 (April 1976)

[河原 03] 河原達也，武田一哉，伊藤克亘，李 晃伸，鹿野清宏，山田 篤：a連続音声認識コンソーシアムの活動報告及び最終版ソフトウェアの概要，信学技報，SP2003-169, NLC2003-106, pp. 79-84 (Dec. 2003)

[河原 08] 河原達也，川嶋宏彰，平山高嗣，松山隆司：a対話を通じてユーザの意図・興味を探り情報検索・提示する情報コンシェルジェ，情報処理，Vol. 49, No. 8, pp. 912-918 (2008)

[河井 10] 河井 恒，磯谷亮輔，安田圭志，隅田英一郎，内山将夫，

松田繁樹，葦苅 豊，中村 哲：平成 21 年度全国音声翻訳実証実験の概要，2010 秋季音響論集，No. 3-9-6（2010）
[喜連川 11] 喜連川優：情報爆発のこれまでとこれから，信学誌，Vol. 94, No. 8, pp. 662-666（2011）
[Koehn 10] Koehn, P.: *Statistical Machine Translation*, Cambridge University Press（2010）http://www.statmt.org/moses/
[Koerner 95] *Concise History of the Language Sciences: From the Sumerians to the Cognitivists*, Edited by Koerner, E. F. K. and Asher, R.E., pp. 431-445, Oxford: Pergamon Press（1995）
[Lupu 11] Lupu, M., et al.: Overview of the TREC 2011 Chemical IR Track, *Proc. 20th Text REtrieval Conference,* NIST Special Publication: SP 500-295（2011）
[Metze 02] Metze, F., et al.: The NESPOLE! Speech-to-speech translation system, *2nd Int. Conf. on Human Language Technology Research, HLT'02*, pp. 378-383（2002）
[MIC 11] 情報通信審議会新事業創出戦略委員会・研究開発戦略委員会 ICT 基本戦略ボード（第 7 回）会議資料 2011 年（2011）
[水上 11] 水上悦雄，翠 輝久，河合 恒，柏岡秀紀，中村 哲：観光案内音声対話アプリケーション AssisTra，人工知能学会言語・音声理解と対話処理研究会，SIG-SLUD-63, pp.31-32（Oct. 2011）http://mastar.jp/assistra/index.html
[中村 07] 中村 哲，隅田英一郎，清水 徹，Sakriani, S.，坂井信輔，Zhang, J., Andrew, F.，木村 法幸，葦苅 豊：アジア言語音声翻訳コンソーシアム：A-STAR について，日本音響学会 2007 年秋季研究発表会講演論文集，No. 1-3-14, pp. 45-46（2007）
[中村 08] 中村 哲：音声翻訳技術の現状と今後の展開，文部科学省科学技術政策研究所科学技術動向研究センター 科学技術動向（89），pp. 8-19（Aug. 2008）http://data.nistep.go.jp/dspace/bitstream/11035/990/3/NISTEP-STT089J-1.pdf
[Nakamura 10] Nakamura, S., Torisawa, K., Kawai, H. and Sumita, E.: NICT speech and language resources and corpora, *Proc. Oriental COCOSDA 2010*（2010）
[中村 11] 中村 哲，磯谷亮輔，乾健太郎，柏岡秀紀，河井 恒，河原達也，木俵 豊，黒橋禎夫，隅田英一郎，関根 聡，鳥澤健太郎，堀 智織，松田繁樹：Web 時代の音声・言語技術，信学誌，総合報告，Vol. 94, No. 6, pp. 502-517（2011）
[NITRD 12] FY 2013 Supplement to the President's Budget, *NITRD Supplements*（2012）http://www.nitrd.gov/pubs/2013supplement/FY13NITRDSupplement.pdf, http://www.nitrd.gov/
[緒方 10] 緒方 淳，後藤真孝：a PodCastle：ポッドキャスト音声認識のための集合知を活用した言語モデル学習”，情処学音声言語情報処理研報，2010-SLP-80-10（2010）
[Raux 05] Raux, A., Langner, B., Black, A. and Eskenazi, M.: Let's go public! Taking a spoken dialog system to the real world, *Interspeech 2005*（*Eurospeech*）, Lisbon, Portugal（2005）
[Sakti 11] Sakti, S., Paul, M., Finch, A., Sakai, S., Vu, T., Kimura, N., Hori, C., Sumita, E., Nakamura, S., Park, J., Wutiwiwatchai, C., Xu, B., Riza, H., Arora, K. Luong, C. and Li, H.: A-STAR: Toward Translating Asian Spoken Languages, *Computer Speech and Language Journal*（*Elsevier*）, Special issue on Speech-to-Speech Translation（Aug. 2011）
[Schalkwyk 00] Schalkwyk, J., Beeferman, D., Beaufays, F., Byrne, B., Chelba, C., Cohen, M., Garret, M. and Strope, B.: Google search by voice: A case study, *Visions of Speech: Exploring New Voice Apps in Mobile Environments, Call Centers and Clinics*, A. Neustein, Ed., Springer（2010）http://research.google.com/pubs/archive/36340.pdf
[Soltau 09] Soltau, H., Saon, G., Kingsbury, B., Kwang, H., Kuo, J. and Mangu, L.: Advances in Arabic Speech Transcription at IBM Under the DARPA GALE Program, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 17, No. 5, pp. 884-894（July 2009）
[Tc-Star 02] http://www.tcstar.org/
[Thrun 07] Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M. and Hoffmann, G., et al.: *The Robot That Won the DARPA Grand Challenge*, Springer Tracts in Advanced Robotics, Vol. 36/2007, pp. 1-43（2007）, http://archive.darpa.mil/grandchallenge/index.asp
[鳥澤 10] 鳥澤健太郎：情報爆発と音声アプリケーションの可能性—言語処理研究者の考察—，情報処理学会研究会音声言語情報処理（SLP），2010-SLP-84（17）, pp. 1-6（2010）
[内山 09] 内山将夫，阿辺川武，隅田英一郎，影浦 峡：みんなの翻訳，言語処理学会第 15 回年次大会論文集，pp. 184-187（March 2009）http://trans-aid.jp/
[U-Star 10] http://www.ustar-consortium.com/
[Whalstar 00] Wahlster, W.（eds.）: *Verbmobil: Foundations of Speech-to-Speech Translation*, Springer（2000）
[安田 08] 安田圭志，隅田英一郎：機械翻訳の研究・開発における翻訳自動評価技術とその応用，人工知能学会誌，Vol. 23, No. 1, pp. 2-9（2008）
[Young 06] Young, S. J.: Using POMDPs for dialog management, *Proc. IEEE/ACL Workshop on Spoken Language Technology*（*SLT 2006*）, Aruba（2006）

2012 年 11 月 16 日　受理

著 者 紹 介

中村　哲（正会員）

1981 年京都工芸繊維大学工芸学部電子工学科卒業．1992 年博士（工学，京都大学）．1981 年シャープ株式会社中央研究所，情報技術研究所．1986 ～ 89 年（株）国際電気通信基礎研究所自動翻訳電話研究所出向．1994 ～ 2000 年奈良先端科学技術大学院大学情報科学研究科助教授，2000 年（株）国際電気通信基礎技術研究所音声言語コミュニケーション研究所第一研究室長，2005 年所長，取締役，2006 年（独）情報通信研究機構兼務．けいはんな研究所音声言語グループリーダ，上席研究員，MASTAR プロジェクトリーダ，知識創成コミュニケーション研究センター長，けいはんな研究所長などを経て，現在，奈良先端科学技術大学院大学情報科学研究科教授．ATR フェロー．ドイツカールスルーエ大学客員教授．音声翻訳，音声認識，自然言語処理，マルチモーダル情報処理の研究に従事．日本音響学会粟屋奨励賞，技術開発賞，情報処理学会インタラクション 2001 ベストペーパ賞，山下記念研究賞，喜安記念業績賞，本学会研究会優秀賞，AAMT 長尾賞，日本 ITU 協会国際協力賞，電気通信普及財団テレコム技術賞，ドコモモバイルサイエンス賞，総務大臣表彰，文部科学大臣表彰，ELRA Antonio Zampoli 賞など受賞．日本音響学会，電子情報通信学会，情報処理学会各会員．IEEE Senior Member, SLTC 委員，Signal Processing Magazine 編集委員，International Speech Communication Association（ISCA）理事．