社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2006－SLP－64 (41)
2006／12／22

(サーベイ) ICSLP における研究動向 －言語モデル・対話システムを中心に－

西村 竜一* 秋田 祐哉** 須藤 克仁*** 大庭 隆伸***
*和歌山大学 システム工学部 〒640-8510 和歌山県和歌山市栄谷 930
**京都大学 学術情報メディアセンター 〒606-8501 京都府左京区吉田二本松町
***NTT コミュニケーション科学基礎研究所 〒619-0237 京都府相楽郡精華町光台 2-4
E-mail: nisimura@sys.wakayama-u.ac.jp, yuya@media.kyoto-u.ac.jp, {sudoh, oba}@cslab.kecl.ntt.co.jp

あらまし 本稿では，2006 年 9 月アメリカ合衆国ピッツバーグにおいて開催された国際会議 Interspeech2006 - ICSLP における研究動向を報告する．サーベイ二部構成の後編にあたる本編では，音声認識システム，言語モデル，言語処理関連，音声対話，音声検索，要約，翻訳等のトピックを中心に報告する．
キーワード Interspeech2006 - ICSLP

A Report on ICSLP - Language Modeling and Spoken Dialog Systems -

Ryuichi NISIMURA* Yuya AKITA** Katsuhito SUDOH*** and Takanobu OBA***
*Faculty of Systems Engineering, Wakayama University 930 Sakae-dani, Wakayama-shi, Wakayama, 640-8510 Japan
** Academic Center for Computing and Media Studies, Kyoto University Yoshida Nihonmatsu-cho, Sakyo-ku, Kyoto, 606-8501 Japan
*** NTT Communication Science Laboratories 2-4 Hikaridai, Seika-cho, Shoraku-gun, Kyoto, 619-0237 Japan
E-mail: nisimura@sys.wakayama-u.ac.jp, yuya@media.kyoto-u.ac.jp, {sudoh, oba}@cslab.kecl.ntt.co.jp

Abstract This paper reports survey studies of Interspeech2006 - ICSLP held at Pittsburgh, USA in September 2006. The major topics are consisted of technologies such as ASR, Language Modeling, Language Processing, Spoken Dialog, Spoken Information Retrieval, Speech Summarization, and Speech Translation.
Keyword Interspeech2006 - ICSLP

1. はじめに

本稿は，2006 年 9 月にアメリカ合衆国ピッツバーグで開催された国際会議 Interspeech2006 - ICSLP における研究動向を報告する二部構成の後編である．取り上げるトピックおよび執筆担当者は以下のとおりである．

- 音声認識システム，言語認識（大庭）
- 言語モデル，言語処理（秋田）
- 音声対話（西村）
- 音声検索，要約，翻訳（須藤）

2. 音声認識

ここでは「System Combination」「ASR Other I」「ASR Other II」「Large Vocabulary Speech Recognition」の 4 セッションの中から特徴的な取り組みを紹介する．

2.1. 音声認識の高精度化への取り組み

高精度な音声認識は依然大きな研究課題となっている．音響・言語モデルに関する研究以外にも数多くのアプローチが試みられている．中でも複数認識器の組合せやマルチパスの音声認識器などオフライン処理を前提とした研究が多く見られた．計算機や音声認識の使用上の制約よりも認識精度を重視する傾向が見てとれる．これらの研究では共通して単体の音響・言語モデルで表現しきれない情報を利用し，音声認識精度の向上が計られている．以下，この点に着目し幾つかの研究を紹介する．

まず，Recogniser Output Voting Error Reduction (ROVER)や Confusion Network Combination (CNC)に代表される複数システムを利用した音声認識手法においては，complementary 仮説の生成が重要視されている．つまり，各認識器が異なる誤り傾向を含むことが望まれる．Zhang らは boosting 技術の一つである AnyBoosting に Minimum Classification Error を導入する手法を提案している[1]．また，Breslin らは，Minimum Bays Risk に基づいた手法で，互いに complementary なシステムを生成している[2]．一方，直接 complementary モデルを生成しないものの，Stuker らは cross-system adaptation の中で，異なる音素セットや特徴量を利用すると更に精度が向上すると報告している[3]．このほか，ROVER，CNC 関連の研究では，Chen ら[4]や Hoffmeister ら[5]が従来複数仮説の統合時に失われていた時間情報やフレーム情報を残すことで精度の改善を達成している．

マルチパス認識技術の一つに，認識誤りパターンを学習し，修正する手法がある．Zhou らは，Mandarin タスクにおいて，誤り抽出に 3 パス，修正に 6 パスをかけ 4%CER を改善している[6]．Koo らは，置換・脱落・挿入誤りごとに認識誤りの修正ルールを確率モデルで表す手法を提案している[7]．特徴的であったのは，Ma らの報告[8]で，連続数字認識において，第 1 パスで連続音声認識を使わず，各数字の抽出器をそれぞれ適用する．各数字の抽出器は独立に動作するため例えば，複数の数字の時間アライメントが重複する場合もある．抽出された各数字を有向グラフで列表現に，任意の基準で重み付けし最尤パスを連続数字認識の結果とする手法を提案している．その他のマルチパス認識手法としてリスコアリングアプローチが挙げられる．Siniscalchi らは，HMM を用いて音声を区分的に articulation 情報で 15 のクラス（摩擦音，停止，低，高など）に分類し，それから得られるスコアを知識スコアとして利用し仮説をリスコアリングする方法を提案している[9]．

2.2. 音声認識に関するその他の取り組み

GALE 関連で，アラビア語の音声認識に関する研究が 2 件報告されていた．Afify らはアラビア語の方言を認識することを目的とし，接辞のリスト等を用いた単語合成技術を利用し OOV を削減する方法を試みている[10]．Saleem らは口語体への対応のため，アラビア語特有の声門閉鎖音等の情報を利用し認識精度の向上に努めている[11]．これらのアプローチを含め，大量コーパスの収集が難しい言語に対しては，各言語固有の特徴を利用するアプローチが多く見られた．例えば，アムハラ語(エチオピアの公用語)や，ソマリ語を対象とした研究[12, 13]が報告されていた．

音声認識の利用先としては，講演音声の書き起し，Broadcast のキャプショニングのような以前から取り組まれているタスク設定が多く，画期的かつチャレンジングなタスクへの移行は見られなかった．しかしながら，その枠組みの範疇にはあるものの Lecouteux ら

は映画のシナリオやニュースの要約などを不完全な書き起し文とし，その情報と音声の同期を取ることにより音声認識精度の向上を目指すという問題設定を新たに導入していた[14]．また，Hazen は，人間手作業による講演音声の書き起しサービスを利用した場合に，コストとのトレードオフで書き起し誤りが増大してしまう点に着目し，その誤り込みの書き起しを利用し音声との時間アライメントおよび誤りの自動修正に挑戦している[15]．正確な大量コーパスの生成の低コストを目指したものである．

最後に，会議録生成において複数話者の発話が重なった場合の現象についての解析を行った興味深い報告が Cetin ら[16]により為されているので，これを紹介する．まず，複数話者の発話が重なる割合は 10%前後であり，内 90%程度が 2 話者の重複となる．つまり 3 話者以上の同時発話は殆ど存在しないことになる．また，重複部分には言語的特長も観測され，3 次以上の単語並び成分に関してはパープレキシティーが高くなり，逆に 1 次成分では低下すると報告している．このほか，会議の状況における複数話者の発話重複率の変化なども紹介している．

3. 言語認識

ここでは「Language and Dialect Recognition」セッションから数件の研究を紹介する．

Bauer らは，音声認識での音響モデルの学習に適用し成功した識別学習法である Minimum Classification Error を Language Identification (LID)の音響モデル学習に採用した．LID エラー最小化基準で学習を行うことに主な特徴を置く．最尤推定による音響モデルを利用した場合に比べ，4.7%の LID エラー率改善に成功している[17]．一方，Yang らはある言語における単語並びの妥当性を SVM を利用して評価する方法を提案している．この手法についても従来の最尤推定基準の tri-gram に対して 4 %程度の精度向上を実現[18]し，LID においても識別学習の有効性を証明している．

さて，LID の問題の一つにコーパスの収集がある．音響・言語モデルの学習のために，アノテートされた複数言語のコーパスを用意しなければならない．そのため大量に収集するには極めて多大なコストを要する．この問題に対し，LID でも unsupervised adaptation の適用が Timoshenko らにより紹介された[19]．言語への適応はできず，データベースや環境への適応となる点など LID 固有の議論あるが，基本的な適応技術は MLLR といった従来手法である．

本セッションでは 2 件韻律利用に関する研究が取り上げられていた．Lin らは従来の phonotactic ベースの LID システムに音韻ベースシステムを利用する枠組みをベイズ基準で導入する方法を提案している．これにより LID 精度は 2%程度改善された[20]．一方，Ikeno らは，US ネイティブの英語のアクセント知覚における韻律情報の役割を解析している．多角的に解析を行い最終的に，アクセント知覚の要因として音響的な特徴のみならず，発話の意味に基づくトップダウンな処理も関わっているとしている[21]．（大庭）

4. 言語モデル

本節では，Language Modeling for Spoken Dialog Systems（口頭），Language Model Adaptation, Refinement and Evaluation（口頭），Language Modeling and ASR adaptation（ポスター）の 3 セッションを扱う．

4.1. WWW の利用

近年では WWW（と検索エンジン）を利用して機械的にテキストを収集し，学習または適応に利用する手法が試みられている．これに関して今回の会議では 3 件の発表があり[22-24]，うち 2 つは学習の，1 つは適応の報告であった．WWW より収集したテキストは適切に取捨選択しなければならないが，[22]では対話的な発話スタイルの言語モデルを用いて，一方[23]では模擬データから構築した言語モデルを用いて収集テキストのパープレキシティを計算することにより，言語モデルの学習テキストを選択している．なお[23]の模擬データは文法をもとに自動生成されており，これを用いた言語モデルの学習も試みられている．一方適応手法[24]では，音声認識結果に基づいてクエリを作成して検索を行い，収集したテキストを混合して適応モデルを作成する．WWW を用いた言語モデルの構築は，今後音声認識の対象が拡大するにつれて定着していくと思われるが，ドメインごとの有効性や，テキストの選択方法についてさらなる実証が必要と考えられる．

4.2. 言語モデル適応

話題を確率的に表現できる PLSA (Probabilistic Latent Semantic Analysis)およびその発展形である LDA (Latent Dirichlet Allocation)を用いた言語モデル適応は一般的となってきており，今回の会議でも 2 件の口頭発表が行われた[25,26]．いずれも中国語のタスクにおける評価であり，文字誤り率（CER）における改善は絶対値でおよそ 0.5%[25]，1.3%[26]と報告されている．なお，[26]では比較として PLSA による適応も行われており，LDA と同等の性能であった．

このほか，適応用データにおける Unigram 確率を用いた FMA (Fast Marginal Adaptation)，すなわちバックグラウンドモデルの Trigram と Unigram の確率の比を適応データの Unigram 確率に乗じて適応を行う手法が提案されており[27]，Switchboard における評価で大幅なパープレキシティの改善が得られている．

4.3. その他の話題

言語モデルの枠組みに関するものとして，最大エントロピー法を用いて韻律情報を組み込む手法[28]，話題をベイジアンネットワークで表現して単語の確率に反映させる手法[29]，発音辞書における発音確率に代えて，変異形も言語モデル内でモデル化する手法[30]，外来語の認識のための拡張[31]，IVR システムのためのモデル自動生成[32]などが提案されていた．このほか，フランス語[33]や中国語[34]をはじめ言語特有の問題に対するアプローチ[33-36]が報告されている．

対話システム関連の発表としては，スロットフィリングに用いるモデルを，類似文のクラスタリングに基づいて自動構築する手法[37]，対話管理に音声認識器やパーザの信頼度を用いる手法[38]，音声による検索クエリ入力を想定した，発音と文字表記との対応モデルの構築手法[39]などが提案されている．

このほか，子供音声の認識における分析[40]，ディクテーションにおける訂正発話の分析[41]，コール分析における非音声情報の利用[42]などがみられた．

5. 言語処理

本節では，Language Processing Beyond and Below the Word-Level（口頭）および Spoken Language Understanding（ポスター）の 2 セッションを扱う．

5.1. セグメンテーション

口頭発表では，Below the Word-Level として，形態素・サブワードにおける処理が 3 件発表された[43-45]．まず，膠着言語であるため単語の活用・派生が多数見られるフィンランド語とトルコ語を対象に，形態素への自動セグメンテーションを行い，言語モデルの性能を競うコンテストの報告があった[43]．セグメンテーションレベルで複数の手法を統合しても効果は小さいが，認識結果を統合すると改善が得られたとのことである．他の 2 件は，これらの単位による言語モデルの構築と音声認識における評価をトルコ語[44]とフィンランド語[45]で行ったものの報告である．

Beyond the Word-Level の処理としては，CSJ 講演の統計言語モデル・SVM による文境界推定の報告[46]のほか，音声対話システムにおける言語処理を想定した，意味・概念の信頼度の提案[47]があった．

5.2. 音声言語理解

ポスターセッションでは，音声対話システムにおけ

る言語処理が多数報告された．[48]ではスロット解析に先だって話題の判別を行い，スロット解析性能の向上を図っている．[49]では，バックグラウンドの知識ベースを利用した，システムの提示する情報の最適化を提案している．[50]では木構造のメンタルモデルをもとにユーザの内的状態を判別して適切な応答を出すことで対話時間の短縮を実現している．このほかコンテキストを扱ったものとして，言語モデルに関する報告[51,52]のほか，ユーザによるコンテキストの切り替えに対応するシステムの枠組みが提案されている[53].

また，SVM による Dialog Act の判別について 2 件の発表があった．[54]では単語や話者交替などを素性とし，複数の SVM の出力を統合して多クラスを判別している．[55]では N-gram とピッチ・話速・発話長などを素性に利用して SVM を学習し，この出力を状態と見なして HMM 的なモデルを構成している．

このほか，音声対話によるオープンドメイン質問応答システム（RITEL）について，システム構成や予備的な評価の報告があった[56]．（秋田）

6. 音声対話

本節では，音声対話技術についてまとめる．Interspeech は，音声対話に関する研究が集う場として，もっとも良く機能している会議であると言っても差し支えはないだろう．セッションタイトルに Dialog の単語を含むものを数えると 6 つ．そして，音声対話への応用を前提とする技術開発は，他のセッションでもかなりの数が発表されている．本報告ではそのすべてを網羅できていないことにまずはご容赦願いたい．

6.1. 評価・分析・開発支援

音声対話システムの評価，分析を目的とした研究が目立つ傾向にある．対話システムの事例紹介では，昨年の Interspeech から引き続き報告がなされた CMU の Let's Go Bus Information System[57]が興味深い．実環境運用を通じた本格的な実態調査が始まっており，今後が期待される．同システムは電話によるバス運行問合わせシステムであり，一日に 40～60 の利用の報告がある．また，同じくバス案内タスクを扱う Tampere 大学[58]は，長期 30 ヶ月で収集したデータの分析を行っている．時間変化を追った考察が示されるとともに，閉じた対象を被験者としたユーザビリティテストと公開テストで生じた結果の比較が報告されている．Jokinen らによる PDA を用いた経路ナビゲーションタスクの評価[59]では，利用者の年齢，性別，そして音声対話に対する事前知識の違いがアンケート結果に与える影響を調べている．ポスターセッションの会場では，MIT による地図案内システム[60]のデモンストレーションが注目を集めていた．Google Map に似た地図サービスと音声対話によるレストラン検索を組み合わせたシステムである．同システムは，外部データベースからの言語モデル更新能力を持ち，異なる地域へのタスクポータビリティを意識した検討がなされている．同様に，外部データベースからの言語獲得能力を持つ対話ロボットの評価[61]では，最適な単語数の設定が問題となっている．他には，車内[62, 63]，軍事教育 [64]やスロヴァキア語対話[65]などの発表が行われた．総じて，着実な進化を歩んでいるといえるが，似たようなタスク設定が多く，目新しさには欠ける印象が残った．

また，Interspeech2006 の会期中，電話から利用できる会議案内対話システム CONQUEST[66]が稼動しており，会議参加者に利用するよう繰り返し宣伝がなされていた．例えば"When are the sessions on robust ASR?"等の質問に対して案内ができるそうである．会場から携帯電話で試してみたが，西村の英語力不足のため，残念ながら対話は成立しなかった．しかしながら，今後が楽しみな試みの一つである．

事例収集の一方で，人間の模倣を可能とするユーザシミュレーションの検討が進んでいる．時間と労力に多大なるコストを必要とする事例収集とは異なった評価アプローチとなりうる試みである．東北大の Ito らは，VoiceXML をベースにユーザシミュレータを開発している[67]．同システムは，合成音声を用いていることで，音声対話の評価に人間を必要としない．ユーザの言動系列を n-gram を用いて模倣する Georgila らの手法[68]は，n-gram エントリを構成するスロットの状態を記憶することにより，短い n-gram での有効性を確認している．Rieser らは，Wizard-of-Oz の対話事例から，EM アルゴリズムを用いて求めた cluster-based のユーザシミュレーションを提案している[69]．また，ユーザエラーシミュレーションによる評価法の提案がなされた[70].

開発支援に関連する技術としては，対話コーパスの設計を補助する LINTest[71]，状態遷移による対話管理が設計可能な対話管理マネージャを持つ tresos GUIDE[72]，オントロジーレベルの知識ベースを持つ音声対話フレームワーク PATE[73]が紹介された．

6.2. 対話戦略・談話構造

対話の主導権を論じた内容では，[74]では，車載システムでのシステム主導の利便性，前述[22]では現状のシステム主導をユーザとの混同主導型に発展させるプランが示されている．しかし，積極的にシステム主導を推進した事例は影をひそめている感がある．

鉄道案内タスクにおいて，対話事例からニューラルネットによって戦略を統計的に構築する手法が提案された[75]．Rotaru らは，談話構造の「深さ」と「推移」をコントロールすることで音声認識の誤りを制御する考えを示し，コーパス分析を行っている[76]．[77][78]は，対話管理にマルコフ決定過程(MDP)を用いている．そのうち，[77]は，open-set に対する話者同定能力を持ち，対話の過程に組み込み，利用している．一方の[78]では，sapReduction アルゴリズムによる state-action space の削減手法が提案された．Mayer らは，ピッチおよび休止時間の韻律情報を，談話構造における代名詞等の照応関係の解決に利用することを試みている[79]．状態遷移型の対話管理や，slot filling では対応が難しい over-answering 等の問題に対し，それを解決する Joint Intention theory による対話管理が提案された[80]．他にも，電話応対の call routing において，提示するプロンプトの決定に強化学習を適用した事例が AT&T から報告されている[81].

6.3. Turn-Talking, Dialog Acts

スペシャルセッション「The Prosody of Turn-Taking and Dialog Acts」が開催されるなど，発話内行為(Dialog Acts)や，話者交代/継続の検知および予測，発話の順番取り（Turn-Taking）に関連した研究が多く見られた．

対話の受け答えにおいて，韻律情報を用いることで，断片的に短縮された言葉が持つ意味の曖昧性を判別する手法が提案されている．Skantze らは，その検証を，調整された韻律を持つ合成音声を使ったスウェーデン語対話で行い，ユーザの返答時間と意味判別結果の関連を示した [82]．また，Ishi らは，日本語における句末表現を品詞情報に基づき分類し，句末のトーンの Turn-Talking や Dialog Acts における役割を調査している[83]．Schlangen は，特徴量に韻律情報と単語 n-gram を併用したモデルベースの Turn-Taking の事前予測手法を提案している[84]．Kol'a"r らは，ミーティング対話を対象に，話者の個人性に着目した検討を行っている[85]．ネイティブと非ネイティブ話者の違いにも言及している点が興味深い．

全体を通じ，Turn-Taking を知る拠り所となる特徴を探す試みが今後も増加すると思われるが，Ward[86]は，その方法論をアラビア語のコーパスを材料に展開している．Edlund ら[87]によって紹介された /nailon/ は，Prosody 研究を支援する強力なツールである．同プログラムは，各種の韻律情報をリアルタイムに近い時間で取得することが可能であるとの報告があった．

関連する話題として，チュートリアル（教育）シス

テムにおいて学生の質問する姿勢の認識手法の検討がなされ，ここでもピッチ等の韻律情報が有用であると確認された[88]．また，対話においてターンの終わりを知る移行適切場所（Transition Relevance Places）について検討があった[89]．

6.4. 音声対話に関するその他の取組み

北大の Yamada[90]らは，さまざまな状況での対話実例を集め，対話相手の音声品質（自然音声，合成音声，録音音声）が，利用者に与える影響を調べている．

[91]では，TextTiling アルゴリズムを用いて，ミーティングにおけるトピック境界の検出を行っている．

また，自然な対話を実現するには，コミュニケーションの中で生じる感情的現象を扱う必要があるだろう．[92]は怒り，[93]は皮肉，[94]は混乱や驚きの感情表現を音声対話と韻律の観点から扱っている．また，[95]では，実環境で集められた子ども発話に対する感情分析を実施している．（西村）

7. 音声検索・要約・翻訳

本節では，自然言語処理分野で扱われる検索・要約・翻訳等のタスクを音声データに対して行うための技術に関する発表についてまとめる．検索・要約・翻訳はそれぞれオーラルセッションが1つずつ，また当該分野を包含する形でポスターセッションが1つ行われた．

7.1. 音声文書検索

音声文書検索においてはサブワードモデルを用いた手法が継続的に発表されている．単語のみによって検索を行う場合は未登録語をクエリとして用いることができないため，単語よりも短い単位で検索を行うことにより，活用形の変化や発音の揺れなどによる検索漏れを抑制するための手法が重要である．本会議では，サブワード単位での一致度・類似度を定義する方法として，HMM のガウス分布間の Bhattacharya 距離に基づくサブワード間の類似度[96]，Latent Semantic Indexing(LSI)を用いたサブワードの次元圧縮による手法[97]が提案された．

また，未登録語検出を組み込んだ音声認識を用い，認識結果中に未登録語や認識誤りが多いと推定される場合にはサブワードモデルを動的に適用することで検索漏れを抑制する，という統合的なアプローチ[98]も提案された．

未登録語の問題は音声文書検索と通常の文書検索との大きな違いであり，音声分野では今後ともこうした研究が大きな位置を占めるものと思われる．しかし，単語レベルの検索を用いた場合，適合率を高く維持でき，また言語的情報を利用できる可能性もあるため，両者の適切な組み合わせについての研究が今後盛んになる可能性が高い．

7.2. 質問応答

質問応答に関してはポスターで2件の発表があった．音声データから解答を検索する場合には，言い淀みや文法的でない発話が存在し構文解析等の高次の処理が適用しづらいという問題に対し，品詞分類と固有表現認識のみを用いて解答を抽出する方法[99]が提案された．また，音声を入出力とする質問応答(データベースは Web のテキストデータ)において，質問文と Web から検索される解候補文中の semantic role を利用し，解答の正解率と明確性(余分な情報や文法的誤りが少なく，音声合成によって伝達しやすいという意味で)が向上させる手法が提案された[100]．

技術的にはテキストデータを対象とする質問応答との差分はさほど大きくなく，音声独自の着眼点はあまり見られなかった．

7.3. 音声要約

音声要約に関してはスペシャルセッションが開かれ活発な議論が行われた．冒頭のテキスト及び音声要約に関するサーベイ講演[101]では，種々の要約の自動評価手法が紹介された．

また，自然言語処理のフィールドに近いこともあってか，多数の音響的素性や音声認識結果の言語的素性をもとに，識別モデルを利用する手法が積極的に用いられ始めてきており，ロジスティック回帰や SVM を用いた重要文選択[102]，CRF を用いたニュース音声中の Soundbite(挿入されたインタビュー音声等)の検出[103]で高い精度を示すことが報告された．テキスト要約で用いられる技術と，韻律など音声独自の情報の組み合わせを行うことが重要であるという認識が広がってきている印象を受けた．

一方，音響的情報のみを利用するアプローチとして，発話のピッチ変化を維持するように単語を除去し，話し言葉に対する音声認識結果の文短縮を行う手法[104]が提案された．

また，関連技術として，音声翻訳で利用される統計的言語モデルの分野適応のために，EM アルゴリズムを用いてモデル混合比を決定する手法[105]の発表があった．

7.4. 音声翻訳

音声翻訳に関する発表は，DARPA の GALE プログラムの影響もあってかオーラル・ポスター合わせ 10 件強の発表があり，特に米国研究機関によるアラビア語から英語への翻訳タスクを統計的手法によって解くための技術の発表が多かった．

統計的機械翻訳においては，語彙数の多さや翻訳単位(単語/フレーズ)の境界誤りが翻訳モデルの学習に必要な翻訳単位の対応付けに悪影響を与えることから，それに対処するために翻訳単位の分割や語彙のまとめ上げなどの事前処理の技術が提案された[106-108]．

また，翻訳モデルにおける新しいアプローチとして，機械翻訳を単語(またはフレーズ)の分類問題と考え，最大エントロピー法および SVM を用いて解くことで生成モデルを用いた場合より高い精度が出せることが示された[109]．

GALE(および TransTac)では翻訳システムの構築が求められていることから，システム関連の発表も多く見られた．翻訳システムの設計指針として，すべてを統計的機械翻訳に任せるのではなく，代表的な翻訳事例を登録しておき，入力文を文書分類の手法を用いて分類し，該当する事例が見つかったものは事例ベースで，そうでないものは統計的機械翻訳を用いるというアプローチも紹介されていた[110][111]．また，それと異なるアプローチのものとして，中間言語への変換(言語理解)と中間言語からの文生成による翻訳システム[112]も報告された．システム実装については，音声から音声への翻訳を PDA に実装したシステムの紹介[113](英語-アラビア語)[114](英語-中国語)もあった．

音声翻訳を考える上では音声認識誤りへの対処が重要と思われるが，昨年の Interspeech で音声認識結果の N-best や単語ラティスと統計的機械翻訳の統合アプローチが多く見られたのと比較すると本年はそうした発表が見られず，機械翻訳の部分に特化した内容が多かった．

7.5. その他の検索，要約，翻訳関連技術

その他分類やタギングに関する発表があった．

文書分類でよく利用される最大エントロピーモデルの目的関数を exp(*)から{exp(*)}^K(K はハイパーパラメータ，K>>1)に変更することで，0-1 の損失関数に近づけるというアプローチ[115]が提案された．

また，チャンキングに関する発表として，HMM を用いたキーフレーズ抽出[116]，音声認識の確信度を素性とする SVM を用いた音声データからの固有表現抽出[117]，CRF と係り受け解析を組み合わせた高精度な話し言葉の文境界検出[118]があった．

これらの要素技術は情報検索・要約・翻訳などのタスクを考える上での基盤として重要な役割を占めており，自然言語処理分野の成果が積極的に取り入れられている．

7.6. 検索，要約，翻訳に関する所感

全体としては，音声データを扱っている，ということの独自性を表に出した発表は半数程度であり，自然言語処理技術の発表と見てよいものも多かった．

音声データを対象とすると，未登録語や音声認識誤り，言い淀みなどの影響で，テキストと対象とした自然言語処理よりも問題が複雑化することは間違いない．しかし，要約分野では韻律など音声に関する情報が有効に活用されていることもあり，昨今流行の識別モデルを用いた自然言語処理の方法論と，音声特有の情報・ノウハウとを組み合わせるといった研究の流れは加速していくものと考えられる．（須藤）

文　献

[1] R. Zhang, et al., Investigations of Issues for Using Multiple Acoustic Models to Improve Continuous Speech Recognition, Proc. ICSLP, pp. 529-532, 2006.

[2] C. Breslin, et al., Generating Complementary Systems for Speech Recognition, Proc. ICSLP, pp. 525-528, 2006.

[3] S. Stüker, et al., Cross-System Adaptation and Combination for Continuous Speech Recognition : The Influence of Phoneme Set and Acoustic Front-End, Proc. ICSLP, pp. 521-524, 2006.

[4] I. Chen, et al., A New Framework for System Combination Based on Integrated Hypothesis Space, Proc. ICSLP, pp. 533-536, 2006.

[5] B. Hoffmeister, et al., Frame Based System Combination and a Comparison with Weighted ROVER and CNC, Proc. ICSLP, pp. 537-540, 2006.

[6] Z. Zhou, et al., A Multi-Pass Error Detection and Correction Framework for Mandarin LVCSR, Proc. ICSLP, pp. 1646-1649, 2006.

[7] H. Koo, et al., Incremental Learning of MAP Context-Dependent Edit Operations for Spoken Phone Number Recognition in an Embedded Platform, Proc. ICSLP, pp. 2310-2313, 2006.

[8] C. Ma, et al. A Study on Detection Based Automatic Speech Recognition, Proc. ICSLP, pp. 2350-2353, 2006.

[9] S.M. Siniscalchi, et al., A Study on Lattice Rescoring with Knowledge Scores for Automatic Speech Recognition, Proc. ICSLP, pp. 517-520, 2006.

[10] M. Afify, et al., On the Use of Morphological Analysis for Dialectal Arabic Speech Recognition, Proc. ICSLP, pp. 277-280, 2006.

[11] S. Saleem, et al., Colloquial Iraqi ASR for Speech Translation, Proc. ICSLP, pp. 1634-1637, 2006.

[12] T. Pellegrini, et al., Investigating Automatic Decomposition for ASR in Less Represented Languages, Proc. ICSLP, pp. 285-288, 2006.

[13] A. Nimaan, et al., Automatic transcription of Somali language, Proc. ICSLP, pp. 289-292, 2006.

[14] Benjamin, et al., Imperfect transcript driven speech recognition, Proc. ICSLP, pp. 1626-1629, 2006.

[15] T.J. Hazen, Automatic Alignment and Error Correction of Human Generated Transcripts for Long Speech Recordings, Proc. ICSLP, pp. 1606-1609, 2006.

[16] O. Cetin, et al, Analysis of Overlaps in Meetings by Dialog Factors, Hot Spots, Speakers, and Collection Site: Insights for Automatic Speech Recognition, Proc. ICSLP, pp. 293-296, 2006.

[17] J.G. Bauer, et al., Minimum Classification Error Training of Hidden Markov Models for Acoustic Language Identification, Proc. ICSLP, pp. 405-408, 2006.

[18] X. Yang, et al., Improved language modeling using support vector machines for language modeling, Proc. ICSLP, pp. 417-420, 2006.

[19] E. Timoshenko, et al., Unsupervised Adaptation for Acoustic Language Identification, Proc. ICSLP, pp. 409-412, 2006.

[20] C. Lin, et al., Fusion of phonotactic and prosodic knowledge for language identification, Proc. ICSLP, pp. 425-428, 2006.

[21] A. Ikeno, et al., The Role of Prosody in the Perception of US Native English Accents, Proc. ICSLP, pp. 437-440, 2006.

[22] T. Misu and T. Kawahara, A Bootstrapping Approach for Developing Language Model of New Spoken Dialogue Systems by Selecting Web Texts, Proc. ICSLP, pp. 9-12, 2006.

[23] K. Wellhammer, et al., Bootstrapping Language Models for Dialogue Systems, Proc. ICSLP, pp. 17-21, 2006.

[24] M. Suzuki, et al., Unsupervised Language Model Adaptation Based on Automatic Text Collection from WWW, Proc. ICSLP, pp. 2202-2205, 2006.

[25] Y.-C. Tam and T. Schultz, Unsupervised Language Model Adaptation Using Latent Semantic Marginals, Proc. ICSLP, pp. 2206-2209, 2006.

[26] D. Mrva and P.C. Woodland, Unsupervised Language Model Adaptation for Mandarin Broadcast Conversation Transcription, Proc. ICSLP, pp. 2210-2213, 2006.

[27] D. Klakow, Language Model Adaptation for Tiny Adaptation Corpora, Proc. ICSLP, pp. 2214-2217, 2006.

[28] O. Chan and R Togneri, Prosodic Features for a Maximum Entropy Language Model, Proc. ICSLP, pp. 1858-1861, 2006.

[29] P. Wiggers and L. Rothkrantz, Topic-based Language Modeling with Dynamic Bayesian Networks, Proc. ICSLP, pp. 1866-1869, 2006.

[30] A. Ljolje, Pronunciation Dependent Language Models, Proc. ICSLP, pp. 2218-2221, 2006.

[31] H. Yamamoto, et al., Speech Recognition of Foreign Out-of-Vocabulary Words Using a Hierarchical Language Model, Proc. ICSLP, pp. 1870-1873, 2006.

[32] M. Balakrishna, et al., Automatic Generation of Statistical Language Models for Interactive Voice Response Applications, Proc. ICSLP, pp. 1898-1901, 2006.

[33] C. Lavecchia, et al., How to handle gender and number agreement in statistical language models? , Proc. ICSLP, pp. 1854-1857, 2006.

[34] X. Hu, et al., Language Modeling of Chinese Personal Names Based on Character Units for Countinuous Chinese Speech Recognition, Proc. ICSLP, pp. 1874-1877, 2006.

[35] L. A and H. Murthy, A Syllable Based Continuous Speech Recognizer for Tamil, Proc. ICSLP, pp. 1878-1881, 2006.

[36] M. Woszczyna, et al., Spontaneous Thai Speech Recognition, Proc. ICSLP, pp. 1882-1885, 2006.

[37] H. Ye and S Young, A Clustering Approach to Semantic Decoding, Proc. ICSLP, pp. 5-8, 2006.

[38] M. Purver, et al., Robust Interpretation in Dialogue by Combining Confidence Scores with Contextual Features, Proc. ICSLP, pp. 1-4, 2006.

[39] A. Horndasch, et al., Phoneme-to-Grapheme Mapping for Spoken Inquiries to the Semantic Web, Proc. ICSLP, pp. 13-16, 2006.

[40] M. Gerosa, et al., Acoustic Analysis and Automatic Recognition of Spontaneous Children's Speech, Proc. ICSLP, pp. 1886-1889, 2006.

[41] K. Vertanen, Speech and Speech Recognition during Dictation Corrections, Proc. ICSLP, pp. 1890-1893, 2006.

[42] Y.-C. Ju, et al., Call Analysis with Classification Using Speech and Non-Speech Features, Proc. ICSLP, pp. 1902-1905, 2006.

[43] M. Kurimo, et al., Unsupervised Segmentation of Words into Morphemes - Morpho Challenge 2005: Application to Automatic Speech Recognition, Proc. ICSLP, pp. 1021-1024, 2006.

[44] E. Arisoy and M. Saraclar, Lattice Extension and Rescoring Based Approaches for LVCSR of Turkish, Proc. ICSLP, pp. 1025-1028, 2006.

[45] S. Virpioja and M. Kurimo, Compact N-gram Models by Incremental Growing and Clustering of Histories, Proc. ICSLP, pp. 1037-1040, 2006.

[46] Y. Akita, et al., Sentence Boundary Detection of Spontaneous Japanese using Statistical Language Model and Support Vector Machines, Proc. ICSLP, pp. 1033-1036, 2006.

[47] C. Kobus, et al., Exploiting Semantic Relations for a Spoken Language Understanding Application, Proc. ICSLP, pp. 1029-1032, 2006.

[48] W.-L. Wu, A Spoken Language Understanding Approach Using Successive Learners, Proc. ICSLP, pp. 1906-1909, 2006.

[49] H. Pon-Barry, et al., Evaluation of Content Presentation Strategies for an In-car Spoken Dialogue System, Proc. ICSLP, pp. 1930-1933, 2006.

[50] Y. Fukubayashi, et al., Dynamic Help Generation by Estimating User's Mental Model in Spoken Dialogue Systems, Proc. ICSLP, pp. 1946-1949, 2006.

[51] V. Goel and R. Gopinath, On Designing Context Sensitive Language Models for Spoken Dialog Systems, Proc. ICSLP, pp. 1934-1937, 2006.

[52] H. Holzapfel and A. Waibel, A Multilingual Expectations Model for Contextual Utterances in Mixed-Initiative Spoken Dialogue, Proc. ICSLP, pp. 1942-1945, 2006.

[53] O.T. Stewart, et al., Conversational Help Desk: Vague Callers and Context Switch, Proc. ICSLP, pp. 1910-1913, 2006.

[54] Y. Liu, Using SVM and Error-correcting Codes for Multiclass Dialog Act Classification in Meeting Corpus, Proc. ICSLP, pp. 1938-1941, 2006.

[55]D. Surendran and G.-A. Levow, Dialog Act Tagging with Support Vector Machines and Hidden Markov Models, Proc. ICSLP, pp. 1950-1953, 2006.
[56]S. Rosset, et al., Integrating Spoken Dialog and Question Answering: the Ritel Project, Proc. ICSLP, pp. 1914, 2006.
[57]A. Raux, et al., Doing Research on a Deployed Spoken Dialogue System: One Year of Let's Go! Experience, Proc. ICSLP, pp. 65-68, 2006.
[58]M. Turunen, et al., Evaluation of a Spoken Dialogue System with Usability Tests and Long-term Pilot Studies: Similarities and Differences, Proc. ICSLP, pp. 1057-1060, 2006.
[59]K. Jokinen and T. Hurtig, User Expectations and Real Experience on a Multimodal Interactive System, Proc. ICSLP, pp. 1049-1052, 2006.
[60]A. Gruenstein, et al., Scalable and PortableWeb-Based Multimodal Dialogue Interaction, with Geographical Databases, Proc. ICSLP, pp. 453-456, 2006.
[61]P. Gieselmann and A. Waibel, Dynamic Extension of a Grammar-based Dialogue System: Constructing an All-Recipes Knowing Robot, Proc. ICSLP, pp. 449-452, 2006.
[62]S. Schulz et al., An User-Centered Development of an Intuitive Dialog Control for Speech-Controlled Music Selection in Cars, Proc. ICSLP, pp. 61-64, 2006.
[63]F. Weng et al., CHAT: A Conversational Helper for Automotive Tasks, Proc. ICSLP, pp. 1061-1064, 2006.
[64]A. Roque et al., Radiobot-CFF: A Spoken Dialogue System for Military Training, Proc. ICSLP, pp. 477-480, 2006.
[65]J. Juhar et al., Development of Slovak GALAXY / VoiceXML Based Spoken Language Dialogue System to Retrieve Information from the Internet, Proc. ICSLP, pp. 485-488, 2006.
[66]http://www.conquest-dialog.org/
[67]A. Ito, et al., A User Simulator based on VoiceXML for evaluation of spoken dialog systems, Proc. ICSLP, pp. 1045-1048, 2006.
[68]K. Georgila, et al., User Simulation for Spoken Dialogue Systems: Learning and Evaluation, Proc. ICSLP, pp. 1065-1068, 2006.
[69]V. Rieser and O. Lemon, Cluster-based User Simulations for Learning Dialogue Strategies, Proc. ICSLP, pp. 1766-1769, 2006.
[70]S. Möller et al., MeMo: Towards Automatic Usability Evaluation of Spoken Dialogue Services by User Error Simulations, Proc. ICSLP, pp. 1786-1789, 2006.
[71]L. Degerstedt and A. Jonsson, LINTest, A development tool for testing dialogue systems, Proc. ICSLP, pp. 489-492, 2006.
[72]S. Goronzy, et al., Developing Speech Dialogs For Multimodal HMIs Using Finite State Machines, Proc. ICSLP, pp. 1774-1777, 2006.
[73]N. Pfleger and J. Schehl, Development of Advanced Dialog Systems with PATE, Proc. ICSLP, pp. 1778-1781, 2006.
[74]C. Ackermann and M. Libossek, System- versus User-Initiative Dialog Strategy for Driver Information Systems, Proc. ICSLP, pp. 457-460, 2006.
[75]L.F. Hurtado, et al., A Stochastic Approach for Dialog Management based on Neural Networks, Proc. ICSLP, pp. 49-52, 2006.
[76]M. Rotaru and D.J. Litman, Discourse Structure and Speech Recognition Problems, Proc. ICSLP, pp. 53-56, 2006.
[77]F. Krsmanovic, et al., Have we met? MDP Based Speaker ID for Robot Dialogue, Proc. ICSLP, pp. 461-464, 2006.
[78]H. Cuay´ahuitl, et al., Learning Multi-Goal Dialogue Strategies Using Reinforcement Learning With Reduced State-Action Spaces, Proc. ICSLP, pp. 469-472, 2006.
[79]J. Mayer, et al., Pitch Range and Pause Duration as Markers of Discourse Hierarchy: Perception Experiments, Proc. ICSLP, pp. 473-476, 2006.
[80]R.A. Subramanian, et al., A Joint Intention-Based Dialogue Engine, Proc. ICSLP, pp. 1782-1785, 2006.
[81]C. Lewis, et al., Prompt Selection with Reinforcement Learning in an AT&T Call Routing Application, Proc. ICSLP, pp. 1770-1773, 2006.
[82]G. Skantze, et al., User Responses to Prosodic Variation in Fragmentary Grounding Utterances in Dialog, Proc. ICSLP, pp. 2002-2005, 2006.
[83]C.T. Ishi, et al., Analysis of prosodic and linguistic cues of phrase finals for turn-taking and dialog acts, Proc. ICSLP, pp. 2006-2009, 2006.
[84]D. Schlangen, From Reaction To Prediction Experiments with Computational Models of Turn-Taking, Proc. ICSLP, pp. 2010-2013, 2006.
[85]J. Kol´a˘ et al., On Speaker-Specific Prosodic Models for Automatic Dialog Act Segmentation of Multi-Party Meetings, Proc. ICSLP, pp. 2014-2017, 2006.
[86]N.G. Ward and Y.A. Bayyari, A Case Study in the Identification of Prosodic Cues to Turn-Taking: Back-Channeling in Arabic, Proc. ICSLP, pp. 2018-2021, 2006.
[87]J. Edlund and M. Heldner, /nailon/ – Software for Online Analysis of Prosody, Proc. ICSLP, pp. 2022-2025
[88]J. Liscombe, et al., Detecting Question-Bearing Turns in Spoken Tutorial Dialogues, Proc. ICSLP, pp. 69-72, 2006.
[89]R.J.J.H. van Son, et al., Prominent Words as Anchors for TRP Projection, Proc. ICSLP, pp. 465-468, 2006.
[90]S. Yamada, et al., Is Voice Quality Enough? - Study on How the Situation and User's Awareness Influence the Utterance Features, Proc. ICSLP, pp. 481-484, 2006.
[91]S. Banerjee and A.I. Rudnicky, A TextTiling Based Approach to Topic Boundary Detection in Meetings, Proc. ICSLP, pp. 57-60, 2006.
[92]F. Burkhardt, et al., Detecting Anger in Automated Voice Portal Dialogs, Proc. ICSLP, pp. 1053-1056, 2006.
[93]J. Tepperman, "Yeah Right": Sarcasm Recognition for Spoken Dialogue Systems, Proc. ICSLP, pp. 1838-1841, 2006.
[94]R. Kumar, et al., Identification of Confusion and Surprise in Spoken Dialog using Prosodic Features, Proc. ICSLP, pp. 1842-1845, 2006.
[95]R. Nisimura, et al., Analyzing Dialogue Data for Real-World Emotional Speech Classification, Proc. ICSLP, pp. 1822-1825, 2006.
[96]K. Iwata, et al., Open-Vocabulary Spoken Document Retrieval based on new Subword Models and Subword Phonetic Similarity, Proc. ICSLP, pp. 325-328, 2006.
[97]V.T. Turunen et al., Using Latent Semantic Indexing for Morph-based Spoken Document Retrieval, Proc. ICSLP, pp. 341-344, 2006.
[98]M. Akbacak et al., A Robust Fusion Method for Multilingual Spoken Document Retrieval Systems Employing Tiered Resources, Proc. ICSLP, 1177-1180, 2006.
[99]M. Surdeanu, et al., Design and Performance Analysis of a Factoid Question Answering System for Spontaneous Speech Transcriptions, Proc. ICSLP, 1165-1168, 2006.
[100] S. Stenchikova, et al., QASR: Question Answering Using Semantic Roles for Speech Interface, Proc. ICSLP, pp. 1185-1188, 2006.
[101] A. Nenkova, Summarization Evaluation for Text and Speech: Issues and Approaches, Proc. ICSLP, pp. 1527-1530, 2006.
[102] X. Zhu et al., Summarization of Spontaneous Conversations, Proc. ICSLP, pp. 1531-1534, 2006.
[103] S. Maskey et al., Soundbite Detection in Broadcast News Domain, Proc. ICSLP, pp. 1543-1546, 2006.
[104] G. Murray et al., Dialogue Act Compression Via Pitch Contour Preservation, Proc. ICSLP, pp. 1547-1550, 2006.
[105] P. Chatain, et al., Perplexity Based Linguistic Model Adaptation for Speech Summarisation, pp. 1535-1538, 2006.
[106] J. Riesa, et al., Building an English-Iraqi Arabic Machine Translation System for Spoken Utterances with Limited Resources, Proc. ICSLP, pp. 745-748, 2006.
[107] J. Lee, et al., Improving Phrase-based Korean-English Statistical Machine Translation, Proc. ICSLP, pp. 753-756, 2006.
[108] A. de Gispert et al., Linguistic Tuple Segmentation in Ngram-based Statistical Machine Translation, Proc. ICSLP, pp. 1149-1152, 2006.
[109] S. Bangalore, et al., Sequence Classification for Machine Translation, Proc. ICSLP, pp. 1157-1160, 2006.
[110] D. Stallard, et al., A Hybrid Phrase-based/Statistical Speech Translation System, Proc. ICSLP, pp. 757-760, 2006.
[111] E. Ettelaie, et al., Cross-lingual Dialog Model for Speech to Speech Translation, Proc. ICSLP, pp. 1173-1176, 2006.
[112] C. Wang, et al., High-quality Speech Translation in the Flight Domain, Proc. ICSLP, pp. 761-764, 2006.
[113] R. Hsiao, et al., Optimizing Components for Handheld Two-way Speech Translation for an English-Iraqi Arabic System, Proc. ICSLP, pp. 765-768, 2006.
[114] W. Zhu, et al., Recent Advances of IBM's Handheld Speech Translation System, Proc. ICSLP, pp. 1181-1184, 2006
[115] X. Li, et al., Improved Topic Classification over Maximum Entropy Model Using K-norm based New Objectives, Proc. ICSLP, pp. 329-332, 2006
[116] I. Alphonso, et al., Saliency Parsing for Automated Directory Assistance, Proc. ICSLP, pp. 321-324, 2006
[117] K. Sudoh, et al., Discriminative Named Entity Recognition of Speech Data using Speech Recognition Confidence, Proc. ICSLP, pp. 337-340, 2006
[118] T. Oba, et al., Sentence Boundary Detection Using Sequential Dependency Analysis Combined with CRF-based Chunking, Proc. ICSLP, pp. 1153-1156, 2006