Vol.2010-SLP-80 No.5
2010/2/12

複数の言語モデルと言語理解モデルによるラピッドプロトタイピング向け音声理解

勝丸真樹[†1] 駒谷和範[†1] 中野幹生[†2]
船越孝太郎[†2] 辻野広司[†2]
尾形哲也[†1] 奥乃博[†1]

本稿では，少量の学習データでも高精度な音声理解を実現する手法について述べる．学習データが少ない場合，単一の音声理解方式による精度は低い傾向にある．そこで本手法では，まず，複数の言語モデルと言語理解モデルを用いて複数の理解結果を得ることで，対処可能な発話を増やす．次に，得られた複数の理解結果に対して，ロジスティック回帰に基づき発話単位の信頼度を付与し，その信頼度が最も高い理解結果を選択する．ロジスティック回帰には，学習データ増加時の回帰係数の変化量に着目することで，必要最低限の学習データを割り当てる．評価実験では，学習データが少ない場合でも，単一の音声理解方式と比較して，本手法が高い音声理解精度を得られることを示す．

Speech Understanding Method for Rapid Prototyping Using Multiple Language Models and Multiple Language Understanding Models

Masaki Katsumaru,[†1] Kazunori Komatani,[†1]
Mikio Nakano,[†2] Kotaro Funakoshi,[†2]
Hiroshi Tsujino,[†2] Tetsuya Ogata[†1]
and Hiroshi G. Okuno [†1]

We aim to improve a speech understanding module with a small amount of training data. High performance is not obtained by single speech understanding methods especially when the amount of available training data is small. We utilize multiple language models (LMs) and language understanding models (LUMs) to cover various user utterances. Then, the most appropriate speech understanding result is selected from several candidates on the basis of confidence measures calculated by logistic regressions. We determine necessary amount of training data for the regressions by focusing on changes in their coefficients when the training data increases. We evaluate our method for various amounts of training data and confirm that our method outperforms every single speech understanding method even when only a small amount of training data is available.

1. はじめに

音声対話システムが一般に広く利用されるためには，システム開発初期段階で高性能なシステムを構築する技術，すなわち，**ラピッドプロトタイピング**技術が必要である．本稿では，音声理解部のラピッドプロトタイピングを扱う．ラピッドプロトタイピング技術により，開発当初からユーザの発話を高精度に理解できる．また，作成したプロトタイプを用いて発話を収集し，より質の高い学習データを得ることで，システムの性能の急速な向上ができる．

本稿では，音声理解部のラピッドプロトタイピング技術として，システム開発初期段階の学習データが少ない状況でも高精度な音声理解を実現する手法を報告する．音声理解部で用いる言語モデルと言語理解モデルには，学習データが必要な統計的なモデルと，学習データを必要としない文法ベースのモデルがある．統計的なモデルを用いた音声理解方式は，学習データが十分でない場合，性能は低い．文法ベースのモデルを用いた音声理解方式は，ラピッドプロトタイピングに向いているが，学習データが増加した場合には統計的な方式と比較して性能は低い傾向にある．そこで，本稿では，文法ベースのモデルと統計的なモデルを含む複数の言語モデルと言語理解モデルを組み合わせることで，あらゆる学習データ量において各単一手法と比較して高い性能を得る手法を開発する．この際，以下の二つの課題がある．

(1) 複数の音声理解結果から適切な結果の選択
複数の音声理解方式を用いる場合，得られる複数の結果から最終的な結果を求める必要がある．本稿では，理解結果の選択を行うモジュールを選択部と呼ぶ．従来は，ROVER 法[1] のような多数決が用いられることが多かった[2),3)]．ROVER 法では，重

†1 京都大学大学院 情報学研究科 知能情報学専攻
Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University
†2 (株) ホンダ・リサーチ・インスティチュート・ジャパン
Honda Research Institute Japan Co., Ltd.

み付き多数決を行うが，音声理解性能の高い方式と低い方式とが混在すると，高性能な方式の結果が十分に反映されなくなる場合がある．

(2) 少量の学習データの適切な分割
統計的な音声理解方式で用いる学習データと，選択部で用いる学習データは分ける必要がある．音声理解部の学習では，与えられた学習データに対して精度が最高となるように学習が行われるため，この理解結果を入力として選択部の学習を行うと，実際のデータより不当に高い理解精度に過学習した選択部が得られてしまう．

前者の課題に対し，得られる複数の音声理解結果に対して，ロジスティック回帰を用いて信頼度を付与し，適切な理解結果を選択する．後者の課題に対して，学習データ増加時のロジスティック回帰式の係数に着目し，係数の変化が小さくなった時点の発話数を，選択部に最低限必要な学習データ量とする．残りのできるだけ多くの学習データを，音声理解方式に配分し，性能を向上させる．適切に分割された学習データを用いて個々の音声理解方式と選択部を構築することで，少ない学習データでも高精度な音声理解を行う．この結果として，あらゆる学習データ量で単一手法と比較して高い性能を得る．

2. 関 連 研 究

これまで，音声対話システムにおける音声認識部・言語理解部のラピッドプロトタイピングとして，いくつかの手法が開発されてきた．音声認識部では，音声認識誤りを抑制するためにドメインに合わせた言語モデルを少ない労力で構築する手法[4),5)]が開発されている．言語理解においては，音声対話システムをドメイン・タスク依存な部分と非依存な部分に切り分け，タスク依存な部分のみを記述するだけで，システムを構築する枠組みが提案されている[6)]．以上のような，単一の言語モデルと言語理解モデルによる音声理解方式のみを用いる手法では，多様な発話に対して高精度な音声理解を実現することが難しい．複数の言語モデルと言語理解モデルを用いることで，対処可能な発話が増える．複数の言語モデルと言語理解モデルを用いて音声理解を行った例を**図 1** に記す．図 1 において，言語モデルと言語理解モデルの組み合わせを，「< 言語モデル > + < 言語理解モデル >」で表す．U1 は文法に沿った発話であるため，文法モデルを用いて音声認識を行い，Finite-State Transducer (FST) を用いて言語理解を行った結果が正解となりやすい．これに対し，U2 は文法外の発話であるため，局所的な制約である N-gram モデルを用いた方が認識精度が高くなる．さらに言語理解部で Weighted FST (WFST) を用いることで，言語理解に不要な単語や音声認識時の単語信頼度の低い語を棄却しながら，認識結果をコンセプト列に変換できる．このように複数の音声理解方式を用いることで，U1，U2 両方の発話に対して正しい音声理解結果を得ることができる．

U1 : 六月九日です．	
音声認識結果：	
- **文法**	“六月九日です．”
- **N-gram**	“六月午後のがです．”
音声理解結果：	
- **文法 + FST**	”month:6 day:9 type:refer-time”
- **N-gram + WFST**	”month:6 type:refer-time”
U2 : 二十日に お借りします. (下線部は文法外)	
音声認識結果：	
- **文法**	“二十日二時ごろです．”
- **N-gram**	“二十日に十日二時ます．”
音声理解結果：	
- **文法 + FST**	”day:20 hour:14 type:refer-time”
- **N-gram + WFST**	”day:20 type:refer-time”

図 1 複数の方式による音声理解の例
Fig. 1 Example of speech understanding results by using several speech understanding methods.

また，音声認識・言語理解が別々に研究されることが多いが，音声認識・言語理解をそれぞれ向上させたとしても，それらの組み合わせが適切でない場合は，音声理解全体としての性能向上は難しい．7) では，音声認識部と言語理解部を確率的に統合する手法の開発がなされているが，複数の言語モデルと言語理解モデルの組み合わせに関する議論はなされていない．我々は，複数の言語モデル・言語理解モデルのあらゆる組み合わせからなる音声理解方式を用い，音声理解性能の向上を目指す．

3. 発話単位信頼度に基づく音声理解結果の選択

複数の音声理解方式から出力される理解結果から，適切な結果を選択する手法について述べる．本手法による音声理解の流れを**図 2** に示す．ここで，各発話に対して，N 個の言語モデルと M 個の言語理解モデルの組み合わせによる方式から出力された音声理解結果を i ($i = 1,\ldots, n$) で表す．ただし，$n = \mathrm{N} \times \mathrm{M}$ である．

次に，一発話に対する音声理解結果 i に対し，正解である発話単位信頼度 CM_i を付与する．ここで，音声理解結果が正解とは，当該発話の理解結果が完全に正解，つまり理解結果中に誤ったコンセプトが含まれないことを意味する．次に，最も高い発話単位信頼度が付与

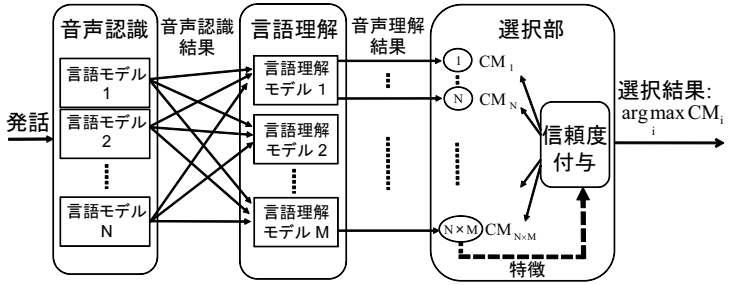

図 2 本手法における音声理解の流れ
Fig. 2 Overview of speech understanding with our method.

表 1 音声理解結果 i に関する特徴
Table 1 Features of speech understanding result i.

F_{i1} :	音声理解結果 i に関する音声認識時の音響スコア
F_{i2} :	発話検証用言語モデル使用時の音響スコアと F_{i1} の差
F_{i3} :	事後確率に基づくコンセプト信頼度の相加平均
F_{i4} :	事後確率に基づくコンセプト信頼度の 音声理解結果 i 内での最小値
F_{i5} :	音声理解結果 i に含まれるコンセプト数
F_{i6} :	音声理解結果が得られなかったか
F_{i7} :	音声理解結果が肯定・否定発話を表すものか

された結果を選択し，当該発話に対する最終的な音声理解結果を得る．つまり，選択結果は $\mathrm{argmax}_i\, CM_i$ となる．

発話単位信頼度は，本稿では音声理解時の特徴に基づくロジスティック回帰により算出する．ロジスティック回帰は，音声理解方式 i ごとに以下の式に基づき構築する．

$$CM_i = \frac{1}{1+\exp(-(a_{i1}F_{i1}+\ldots+a_{i7}F_{i7}+b_i))} \quad (1)$$

ここで，学習データを用いて係数 $a_{i1},\ldots,a_{i7}$ と 切片 b_i をフィッティングする．独立変数 $F_{i1}, F_{i2}, ..., F_{i7}$ は**表 1** に示した特徴である．特徴量は平均 0，分散 1 となるように標準化して用いる．

用いた特徴は，我々が以前用いた特徴 8) から，相関の高い特徴を除き，音声理解結果の内容に関する特徴を加えたものである．以下，詳細を述べる．特徴 F_{i1} と F_{i2} は音声認識結果から得られる特徴である．音響スコアは発話時間で正規化している．F_{i1} は使用した言語モデルに基づく音声認識時の尤度であり，F_{i2} は，音声理解時に用いたモデルとは異なる言語モデル使用時の音響スコアとの比較である．これらの特徴は音声認識結果の信頼性を表す．F_{i3} と F_{i4} は，音声理解結果の 10-best 解の事後確率に基づき算出したコンセプト単位の信頼度[9] に関する特徴である．F_{i5} は，音声理解結果のコンセプト数であり，文法外の発話は発話が長くなることがあり，そのような場合文法モデルに基づく理解結果は正解とならない可能性が高い．F_{i6} により，言語理解部において音声認識結果が受理できなかった場合を検出する．言語理解モデルによっては，受理できない音声認識結果が入力されると，音声理解結果は出力されない．そのような場合は，音声理解結果は正解にならないことが多い．F_{i7} は，肯定・否定発話に対しては比較的高精度な音声理解が可能であると考え導入した．

4. 係数変化量に基づく学習データ量の決定

本節では，学習データを分割し，音声理解方式と選択部とに配分する手法について述べる．本手法で用いるロジスティック回帰において，推定すべきパラメータは，7 つの係数と切片の計 8 つである．これは音声理解部で用いる統計的なモデル，例えば言語モデルの N-gram におけるパラメータの数と比較して非常に少ない．よって，ロジスティック回帰を含む選択部は，音声理解部と比較して少ない学習データ量で収束すると考えられる．以上を踏まえ，我々は学習データの分割を以下の方針に基づいて行う．

(1) 選択部に優先的に学習データを割り当てる．
これは，学習データが非常に少なく，統計的な音声理解方式の精度が著しく低い場合でも，学習データを必要としない文法ベースの方式による結果を適切に得る選択部を構築することで，各単一手法と比べ同等以上の性能を得るためである．

(2) 選択部への学習データの割り当ては必要最小限とする．
これは，できるだけ多くの学習データを音声理解方式に配分し，性能を向上させることで，音声理解部全体の性能向上を図るためである．

上記の方針のもと，選択部のロジスティック回帰の係数の変化量に着目することで，学習データを分割する．ここではロジスティック回帰の学習データを徐々に増やしていき，学習結果の回帰係数の変化が少なくなった時点で学習はほぼ収束したとみなす．そのときの学習データ量を，ロジスティック回帰に必要最小限のデータ量とする．具体的には，以下の操作を行う．

Step. 1 学習データとして使用可能な k_{max} のうち，k 発話，$(k+\delta k)$ 発話を用いて音声理解方式 i に対するロジスティック回帰をそれぞれ構築する．

Step. 2 得られた二つのロジスティック回帰から，係数の変化量を下記の式に基づき算出する．

$$\Delta_i(k) = \sum_j |a_{ij}(k+\delta k) - a_{ij}(k)| + |b_i(k+\delta k) - b_i(k)| \quad (2)$$

ここで $a_{ij}(k)$ と $b_i(k)$ は，音声理解方式 i に対するロジスティック回帰を k 発話の学習データを用いて構築したときの，特徴 F_j の係数と回帰式の切片を表す．

Step. 3 $\Delta_i(k) < \theta$ のとき，ロジスティック回帰の学習は収束したと見なし Step. 4 へ．そうでない場合，$k \leftarrow k + \delta k$ として，Step. 1 に戻る．

Step. 4 k 発話をロジスティック回帰，$(k_{max} - k)$ 発話を音声理解方式，つまり言語モデルと言語理解モデルの学習に割り当てる．

本稿の評価実験では，δk は，被験者の一対話分の発話数 (平均 17 発話) とした．これは，δk が小さく，特定の発話ばかりが含まれる学習データが追加された場合，ロジスティック回帰の学習が不十分であるにも関わらず，係数の変化量が小さくなってしまうためである．

また，Δ_i が小さくなる前に，ロジスティック回帰に割り当てる学習データ k が使用可能な学習データ量の上限 k_{max} に達した場合，すべての学習データを選択部に割り当て，言語モデル・言語理解モデルには，学習データを割り当てない．つまり，音声理解部で使用する音声理解方式は学習データが不要な文法ベースの方式のみとなる．

5. 実装と評価実験

5.1 用いた言語モデルと言語理解モデル

我々は MLMU の一実装として，レンタカー予約システム[10] において，2 種類の言語モデルと，4 種類の言語理解モデルを使用できるようにした．本実装時に用いる学習データや，実験時に用いる評価データは，被験者 39 名に簡単なレンタカーの予約タスクを課し，レンタカー予約システム[10] との対話により収集した．被験者 1 名あたり 8 対話行い，結果，5,900 発話を収集した．収集発話のうち，システムが検出した発話区間と，人手で付与した発話区間とが一致した 5,240 発話を実験に用いた．これは本研究の対象でない発話区間検出誤りや，タスクに関係のない発話を除くためである．5,240 発話のうち 16 名分 2,121 発話を学習データとし，23 名分 3,119 発話を評価データとした．

言語モデルは以下のモデルを用いた．

(1) 文法ベース言語モデル (文法モデル)
(2) ドメイン依存統計的言語モデル (N-gram モデル)

レンタカー予約システムにおける文法モデルは，言語理解時に用いる FST に対応させて人手で記述した．また，N-gram モデルは，学習データの書き起こしを用いてクラス 3-gram を学習し，作成した．語彙サイズは，文法モデルが 281，N-gram モデルが学習データをすべて用いた場合 420 である．それぞれを用いたときの音声認識時の単語正解精度はそれぞれ学習データでは 67.8%と 90.5%，評価データでは 66.3%と 85.0%であった．ここで，音声認識器は Julius (ver. 4.1.2) を用い，音響モデルは話者非依存 PTM トライフォンモデルを用いた[11]．また，音声認識結果を検証するための言語モデルとしてドメイン非依存大語彙統計言語モデルを用いた．これには，連続音声認識コンソーシアム配布の，Web 文章から学習した単語 N-gram モデル (語彙サイズ 60,250) を使用した[11]．

言語理解モデルは以下の 4 種類を用いた．

(1) Finite-State Transducer (FST)
(2) Weighted FST (WFST)
(3) Keyphrase-Extractor (Extractor)
(4) Conditional Random Fields (CRF)

FST による言語理解では，人手で FST を作成し，それに音声認識結果を入力することで，言語理解結果を得る．入力には音声認識結果の 10-best 候補を用い，10-best 候補の 1 位の候補から順に FST で受理可能な認識結果を探す．10-best 候補すべて受理できなかった場合，言語理解結果は出力されない．

WFST による言語理解は福林らの手法に基づく[12]．WFST の構築には，MIT Toolkit[13] を用いる．ここでは，音声認識結果の 10-best 候補それぞれを WFST によりコンセプト列に変換し，累積重みが最大となるコンセプト列を言語理解結果とする．用いる重み付けの種類は，学習データを用いて選択する[12]．WFST による言語理解では，フィラー遷移の導入により，FST では受理されない音声認識結果に対しても言語理解結果を出力できる．また，音声認識時の信頼度を重みに用いるため，音声認識誤りに頑健である．

Extractor による言語理解では，音声認識結果の 1 位の候補に対して，コンセプトに変換可能な音声認識結果の部分列を単純にコンセプトに変換する．ただし，変換されたコンセプト間に矛盾がある場合は，矛盾のないコンセプトの組み合わせを，出力コンセプト数が最大となるように選択する．ここでは，FST で受理可能な一文に同時に現れないコンセプトの組み合わせは矛盾があるとした．Extractor による言語理解は，FST では受理されない音声認識結果に対しても言語理解結果を出力できる．しかし，音声認識結果に誤りが含まれる場合もそのままコンセプト列に変換してしまう．

Vol.2010-SLP-80 No.5
2010/2/12

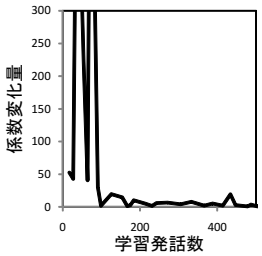

図 3 文法 + FST に対するロジスティック回帰における学習データ量と係数変化量の関係
Fig. 3 Changes of coefficients when training data increased.

CRF[14] による言語理解では，音声認識結果の 1 位の候補に対して，まず，CRF に基づき意味スロットを付与する．次に，意味スロットに当てはまる値を音声認識結果を用いて求める．CRF の構築には，CRF++ [⋆1]を用い，特徴としては，認識結果の単語に加え，単語の先頭の一文字，末尾の一文字，音声認識時の単語信頼度を用いた．パラメータは，学習データを用いて推定する．CRF による言語理解では，学習データを用いて，音声認識結果の単語列と対応するスロット列との関係をモデル化する．学習データが大量に用意できる場合は，様々なパターンの発話に対処できるが，少量の場合，言語理解精度は大きく劣化する．

5.1.1 分割の評価

5.2 音声理解精度の評価

本手法に基づく分割の有効性を評価する．本実験では，以下の二つの場合と比較する．

- 分割なし: 同じデータを用いて音声理解部と選択部の学習を行う．
- 単純な分割: 学習データを音声理解部と選択部に 1 対 1 に配分する．

学習データとして被験者一名分の 141 発話を用いた．本稿では，学習データが不要な文法ベースの音声理解方式に対するロジスティック回帰を用いて係数の変化量を求める．具体的には，文法 + FST を用いた．これは，学習データが非常に少ない段階で，文法ベースの音声理解方式による結果を適切に得るためには，その方式に対するロジスティック回帰を収束させておく必要があるからである．ここで，文法 + FST に対するロジスティック回帰における，学習データ量と係数変化量の関係を，**図 3** に示す．図 3 より，ロジスティック回帰の係数変化量は，学習データが非常に少ない段階では，大きな値であり，100 発話強の学習データで大きく減少し，収束する傾向が見られる．我々は，係数が収束したことを判定する θ は，Δ_i が大きく減少した後の値として 10 とした．

⋆1 http://crfpp.sourceforge.net/

表 2 学習データ 141 発話の分割方法ごとのコンセプト理解精度 [%]
Table 2 Concept understanding accuracy in each dividing method.

分割の方法	コンセプト理解精度	Sub	Del	Ins
本手法による分割	77.9	11.9	6.5	3.7
分割なし	74.1	17.4	5.2	3.3
単純な分割	73.5	13.6	9.9	3.0

本手法により学習データ 141 発話の分割を行ったとき，音声理解部とロジスティック回帰に割り当てる学習データはそれぞれ，30 発話，111 発話となった．各分割手法ごとのコンセプト理解精度を**表 2** に示す．表 2 より，本手法によるコンセプト理解精度は，分割なしと比較して 3.8 ポイント高い．これは，分割を行わなかった場合，ロジスティック回帰の学習のための音声理解結果に正解が非常に多くまれていたのに対し，分割を行うことで，正解に偏った結果でなくなり，それを入力とする選択部を適切に学習できたからである．また，本手法は，単純な分割と比較して 4.4 ポイント高い．これは，単純な分割の場合，選択部に配分する学習データが少なく，選択部が十分に学習されなかったからである．

5.2.1 単一の音声理解方式との比較

学習データ量を変化させた時の，本手法と単一の音声理解方式によるコンセプト理解精度を**図 4** に示す．ここで，oracle は，人手での最適な理解結果の選択を表す．本手法では，1,000 発話を越える学習データのとき，同じデータを用いて音声理解方式と選択部を学習しても，学習データに対する音声理解結果は正解ばかりに偏らず，選択部も適切に構築できると考え，学習データの分割は行わなかった．比較対象となる単一方式は使用可能な学習データをすべて用いて構築している．単一方式は，2 種類の言語モデルと 4 種類の言語理解モデルの組み合わせから 8 種類あるが，ここでは，データ量を変化させる過程で最も高精度となることがあった 3 つの理解方式の結果のみ示す．

図 4 において，oracle による精度はすべての学習データ量において単一の音声理解方式を大きく上回る．これは，複数の音声理解方式を用いることは，どのような学習データ量においても有効であることを示している．

また，本手法により，学習データが比較的少ない場合でも各単一の音声理解方式より高い精度が得られた．その結果，あらゆる学習データ量において各単一方式の性能を上回った．

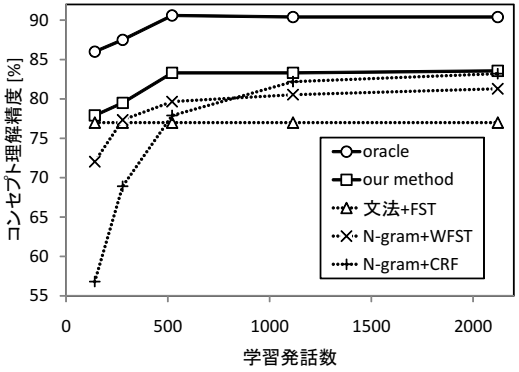

図 4 学習データ量と音声理解性能の関係
Fig. 4 Accuracy when training data increased.

6. おわりに

本研究では，音声対話システムのラピッドプロトタイピングのために，複数の音声理解手法を用いることで少ない学習データで高精度な音声理解を実現するための手法について述べた．評価実験の結果，学習データが少ない場合を含むあらゆるデータ量において，本手法は，単一方式と比較して高い音声理解精度が得られた．

本手法では，ロジスティック回帰の収束判定時には閾値 θ が必要であり，本稿では，$\Delta_i(k)$ の遷移から人手で設定したが，係数の数から自動的に設定できると考えられる．今後，閾値の設定を含めた学習データの自動分割を検討する．また，音声理解部と選択部の学習データの分割比を，学習データの中で分割比を求めるクロスバリデーションで得る手法も検討する．さらに，学習データが大量にある場合には学習データの分割を行わないことを自動的に判定することも今後の課題である．

謝辞 本研究の一部は，科研費，GCOE の支援を受けた．

参 考 文 献

1) Fiscus, J.G.: A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER), *Proc. ASRU*, pp.347–354 (1997).
2) Schwenk, H. and Gauvain, J.-L.: Combining Multiple Speech Recognizers using Voting and Language Model Information, *Proc. ICSLP*, pp.915–918 (2000).
3) Hahn, S., Lehnen, P. and Ney, H.: System Combination for Spoken Language Understanding, *Proc. Interspeech*, pp.236–239 (2008).
4) Misu, T. and Kawahara, T.: A bootstrapping approach for developing language model of new spoken dialogue systems by selecting Web texts, *Proc. Interspeech*, pp.9–13 (2006).
5) Weilhammer, K., Stuttle, M.N. and Young, S.: Bootstrapping Language Models for Dialogue Systems, *Proc. Interspeech*, pp.17–20 (2006).
6) 小暮 悟，中川聖一：データベース検索用音声対話システムにおける移植性の高い意味理解部・検索部の構築と評価，情報処理学会論文誌， Vol.49, No.8, pp.2762–2772 (2008).
7) Damnati, G., Bechet, F. and Mori, R.D.: Spoken Language Understanding Strategies on the France Telecom 3000 Voice Agency Corpus, *Proc. ICASSP*, Vol.IV, pp. 9–12 (2007).
8) Katsumaru, M., Nakano, M., Komatani, K., Funakoshi, K., Ogata, T. and Okuno, H.G.: Improving Speech Understanding Accuracy with Limited Training Data Using Multiple Language Models and Multiple Understanding Models, *Proc. Interspeech*, pp.2735–2738 (2009).
9) 駒谷和範，河原達也：音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理，情報処理学会論文誌， Vol.43, No.10, pp.3078–3086 (2002).
10) Nakano, M., Nagano, Y., Funakoshi, K., Ito, T., Araki, K., Hasegawa, Y. and Tsujino, H.: Analysis of User Reactions to Turn-Taking Failures in Spoken Dialogue Systems, *Proc. SIGdial*, pp.120–123 (2007).
11) Kawahara, T., Lee, A., Takeda, K., Itou, K. and Shikano, K.: Recent Progress of Open-Source LVCSR Engine Julius and Japanese Model Repository, *Proc. ICSLP*, pp.3069–3072 (2004).
12) 福林雄一朗，駒谷和範，中野幹生，船越孝太郎，辻野広司，尾形哲也，奥乃 博：音声対話システムにおけるラピッドプロトタイピングを指向した言語理解，情報処理学会論文誌， Vol.49, No.8, pp.2762–2772 (2008).
13) Hetherington, L.: The MIT Finite-State Transducer Toolkit for Speech and Language Processing, *Proc. ICSLP*, pp.2609–2612 (2004).
14) Lafferty, J., McCallum, A. and Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proc. ICML*, pp.282–289 (2001).