

次発話予測による音声認識結果の再順序付け\*

7N-5

巖寺 俊哲 竹澤 寿幸 石崎 雅人 森元 暉†

ATR 音声翻訳通信研究所‡

e-mail: {iwadera,takezawa,ishizaki,morimoto}@itl.atr.co.jp

1 はじめに

対話中の先行する発話から次発話を予測し、その情報を音声認識に利用することを試みている。音声認識候補中には、統語的、意味的に正しいがその文脈中では、不適切な候補が存在する場合がある。次発話予測によってこのような候補を排除し、認識精度を向上させることができる。また、次に生じ得る発話表現の可能な範囲を限定することによって、候補生成時の構文的複雑さの低減、不適切候補の生成の回避が可能となる。これによって、認識効率を向上させることができる。

次発話予測手法として2つのアプローチが試みられている。1つは、プラン [1, 2] や話題遷移モデル [3] などのあらかじめ定義された知識を用いるアプローチである。もう1つは、統計的手法を用いるアプローチである。これは、発話タイプ等の連鎖により文脈を記述し、この連鎖を統計的にモデル化し次発話タイプを予測する手法である [4, 5]。また、上記のような次発話予測手法を音声認識処理に適用されはじめている [6, 7]。

我々は、発話タイプの連鎖だけでなく対話構造を考慮した統計的アプローチをとっている [8]。この手法では、次発話の発話タイプが予測される。今回、発話タイプを利用した音声認識結果候補の再順序付けを行なった。この目的は、次発話予測手法によって予測された発話タイプ利用の効果を見積もることである。今回用いた音声認識結果候補の再順序付け手法では、まず、各認識結果候補にその発話表現に基づいて発話タイプを求める。その発話タイプと予測された発話タイプを照合し、各候補に点数付けする。次にこの点数によってソートし出力する。実験の結果、再順序付け前と比較して、認識精度が向上することを確認した。

本稿では、発話タイプを利用した音声認識結果候補の再順序付け手法について述べる。さらに、実験により本手法の評価を行なった。その結果について報告する。

2 音声認識結果候補再順序付け手法

2.1 発話タイプ

今回用いた発話タイプは、28種類である。各発話タイプは、表層表現、特に、文末表現と手がかり語と対応して定義している。この発話タイプの特徴は、ドメインに依存しない表層表現と対応していることである。表

表 1: 発話タイプの例

表層表現ボタン	発話タイプ
はい	<greet>,<acknowledg>,<yes>
*です	<inform>
すみません	<thank>,<apology>,<alert>
*たいんです*	<desire>
*どのような*か	<wh-question>
...	...

1に今回用いた発話タイプの一部と対応する表層表現ボタンの例を示す。

2.2 再順序付け手法

発話毎に、以下に示す手順で音声認識結果候補を再順序付けする。ここでは、次発話予測処理から発話タイプが、音声認識処理からは、上位 n 個の音声認識結果候補（尤度情報付き）が得られることを前提とする。

まず、音声認識結果候補のすべてに対して発話タイプを付与する。付与には、発話タイプ-表現ボタン対応テーブルを参照する。これは、ある表現ボタンとその表現ボタンを満たす発話に付与可能である発話タイプの対応関係を記述したテーブルであり、人手により作成される。

次に発話タイプを用いて音声認識結果候補をソートする。このソートは、予測された発話タイプが付与されていて、かつ、音声認識尤度の高い順になるように行なう。このソート結果を音声認識結果とする。

3 再順序付け実験・評価

次発話予測による音声認識結果の再順序付けは、(1) 次発話の発話タイプの予測、(2) 発話タイプによる再順序付け、の2つの処理を経て行なわれる。処理 (1) については、発話タイプの連鎖だけでなく対話構造を考慮した統計的アプローチの検討を進めている [8]。ここでは、処理 (2) に関して、予測された発話タイプによる音声認識精度改善の効果を評価する。したがって、今回の実験では、実際には、次発話予測は行なわず、音声認識処理の対象となる発話について、正しい発話タイプが得られることを仮定している。すなわち、次発話予測精度が100%である場合を想定している。

3.1 音声認識

実験に用いた音声認識法は、入力音声データのポーズ情報に着目し、ポーズで区切られた区間（ポーズ単位）

\*Reordering Candidates of Speech Recognition Results by Predicting Following utterance types

†Toshiaki IWADERA, Toshiyuki TAKEZAWA Masato ISHIZAKI and Tsuyoshi MORIMOTO

‡ATR Interpreting Telecommunications Research Labs.

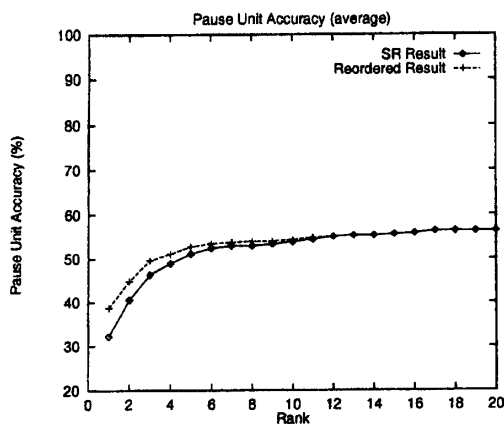


図 1: ポーズ単位認識率 (全ポーズ単位)

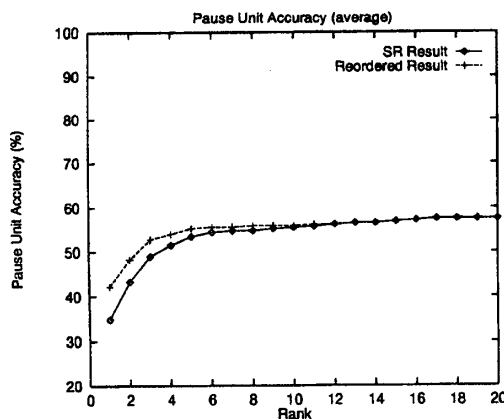


図 2: ポーズ単位認識率 (発話タイプ付与可能ポーズ単位のみ)

を音声認識の単位として認識を行なう [9]。

### 3.2 実験条件

ATRで収集中の音声言語データベース [10] から選んだ会話音声を対象に実験を行ない、音声認識結果に対して、再順序付け前後で累積認識精度を調査した。再順序付けの単位としてポーズ単位を採用し、これに対して実験を行なった。また、ポーズ単位について、(1)すべての単位、(2)音声認識結果候補に発話タイプが付与し得る単位のみ、の2つの場合について累積認識精度を測定した。実験に用いた対象データ量は、6会話、1838単語である。ポーズ単位数は、342個である。

### 3.3 実験結果

実験では、音声認識結果候補上位20位について再順序付けを行なった。実験結果を図1、2に示す。

実験結果では、累積認識率の上位の部分において認識率が向上している。発話タイプが付与可能である音声認識結果候補を含むポーズ単位の認識率は、約10%向上している。これは、発話タイプの利用が音声認識率の向上に寄与することを示している。しかし、累積認識率

の下位では、音声認識候補結果の認識率と一致してしまう。これは、再順序付けという原理上、必然的なことであり、常に、音声認識結果候補の累積認識率の最大値が上限値となってしまふ。これを解決するためには、音声認識過程において認識候補を生成する段階で発話タイプに関する情報を用いる必要がある。

## 4 おわりに

発話タイプを用いた音声認識結果候補の再順序付け手法とその効果について報告した。ここでは、発話タイプは、次発話予測手法から得られることを仮定した。実験の結果、再順序付け前と比較して、認識精度が向上した。これによって発話タイプを予測、利用することが、音声認識の精度向上に寄与することを確認した。今後は、次発話予測部との統合に関して検討を進める。また、今回は、音声認識の処理結果に対して再順序付けを行なった。今後は、音声認識過程で発話タイプに関する情報を利用することによる認識精度と処理効率の向上について検討する。

## 参考文献

- [1] 山岡 孝行、飯田 仁：“階層型プラン認識モデルを利用した次発話予測手法—話し手の意図を表す表現についての音声認識結果曖昧性の解消”、電子情報通信学会論文誌、vol.J76-DII, No.6, 1993.
- [2] 山本 哲也、太田 義一、山下 洋一、溝口 理一郎：“音声理解システム SPURT-Iのための対話管理機構:MASCOTS”、第4回人工知能学会全国大会、13-2, 1990.
- [3] 平松 敬史、吉田 英昭、野村 康雄、山下 洋一、溝口 理一郎：“音声対話理解のための話題知識の利用”、電子情報通信学会研究会、SP92-110、1992.
- [4] Nagata, T. and Morimoto, T.: “An Information-Theoretic Model of Discourse for Next Utterance Type Prediction”, 情報処理学会論文誌、Vol. 35, No.6, 1994.
- [5] Reithinger, N. and Maier, E.: “Utilizing Statistical Dialogue Act Processing in VerbMobil”, ACL-95, 1995.
- [6] 鈴木 雅実、井ノ上 直己、谷戸 文廣：“発話タイプの予測を用いた対話音声認識方式”、日本音響学会講演論文集、1995.
- [7] 鈴木 雅実、谷戸 文廣、斎藤 博昭、中西 正和：“対話音声認識における次発話予測の効果”、電子情報通信学会総合大会講演論文集、1996.
- [8] 巖寺 俊哲、石崎 雅人、森元 暉：“対話構造を考慮した次発話予測手法の検討”、情報処理学会第52回全国大会、1996.
- [9] 竹澤 寿幸：“部分木を単位とする構文規則と前終端記号のバイグラムを利用した連続音声認識”、電子情報通信学会研究会、NLC95-54、SP95-89、情報処理学会研究会、95-SLP-9-9、1995.
- [10] Morimoto, T., et al.: “A Speech and Language Database for Speech Translation Research”, Proc. of ICSLP'94, pp.1791-1794, 1994.