

小松 久美子^{*} 黒田 由香[†] 長友 健太郎[‡] 西村 竜一[‡]李 晃伸[‡] 鹿野 清宏[‡]^{*}イメージ情報科学研究所 [†]T I S 株式会社[‡]奈良先端科学技術大学院大学 情報科学研究科

1. はじめに

従来のように新聞記事などの書き言葉で書かれたテキストから作成した言語モデルを用いた音声認識システムでは、話し言葉を認識することは難しい。話し言葉に対応した言語モデルにより認識率を向上させる必要がある。そこで、話し言葉の表現を多く含むコーパスを収集し、話し言葉言語モデルの構築を試みた。本発表では、収集したコーパスの概要と作成した言語モデルの評価結果について報告を行う。

2. コーパスの収集

話し言葉が認識できる言語モデルを構築するには、話し言葉的な文章が多く含まれるテキストを大量に集める必要がある。Web 掲示板には話し言葉的な文章が多く存在する。そこで、今回は高齢者タスクとしてグルメと健康相談を想定し、グルメ、健康相談関連の Web の掲示板からテキストを収集した。収集後、一定のルール of テキストフィルタをとおして不要な記号や文字を削除、整形し、それぞれのタスクのコーパスとした。収集したコーパスの大きさを表 1 に示す。

表 1 コーパスの大きさ

コーパス	容量 (MB)	語彙数	文章数
グルメ	74	79655	805908
健康相談	11	41973	319498
毎日新聞'94	94	138663	906223

3. 言語モデルの作成

収集したコーパスを用いてグルメ・健康相談のタスクに対応した言語モデルの作成を行った。また、従来の書き言葉コーパスから作成した言語モデルと比較するため、毎日新聞'94 年度版のテキストから言語モデルを作成した。使用した語彙数は、出現頻度上位 2 万語である。

4. 言語モデルの評価

作成した言語モデルの評価基準としてカバレッジとトライグラム言語モデルのパープレキシティを使用した。

評価文には、グルメタスク文 200 文、健康相談タスク文 150 文を使用した。評価文の例を図 1、図 2 に示す。この評価文は、実際の対話例を参考にして作成した

ちょっとしたおひたしを作りたいんですけど
カニが食べたいんですけど、おすすめは何ですか？
トマトがいっぱいあるので 10 分以内でできる料理を 2、3 品教えてください

図 1 評価文の例 (グルメタスク)

もよりの接骨院はどこにありますか？
その皮膚科にはどうやって行ったらいいですか？
最近たちくらみがひどいんですが

図 2 評価文の例 (健康相談タスク)

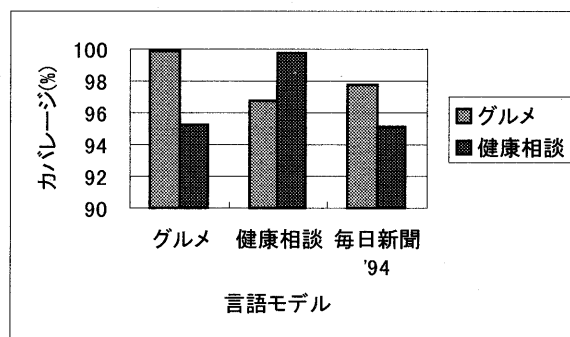


図 3 評価結果 (カバレッジ)

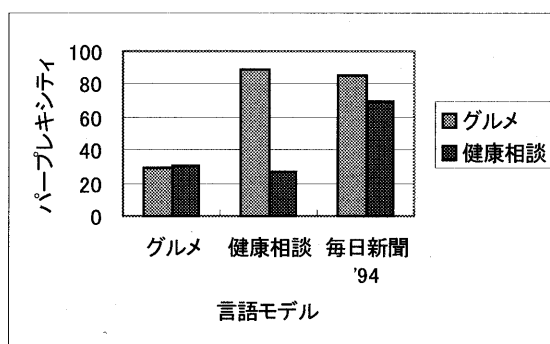


図 4 評価結果 (パープレキシティ)

比較的丁寧な話し言葉である。

評価結果を図 3、図 4 に示す。グルメタスク評価文にはグルメタスク言語モデルが、健康相談タスクには健康相談タスク言語モデルがカバレッジ、パープレキシティともに他の言語モデルに比べて最も良い値になっている。

Language Models for Spoken Language Senior Task

Kumiko Komatsu^{*}, Yuka Kuroda[†], Kentaro Nagatomo[‡],Ryuichi Nisimura[‡], Akinobu Lee[‡], Kiyohiro Shikano[‡]^{*}Laboratories of Image Information Science and Technology[†]TIS [‡]Graduation School of Information Science, NAIST

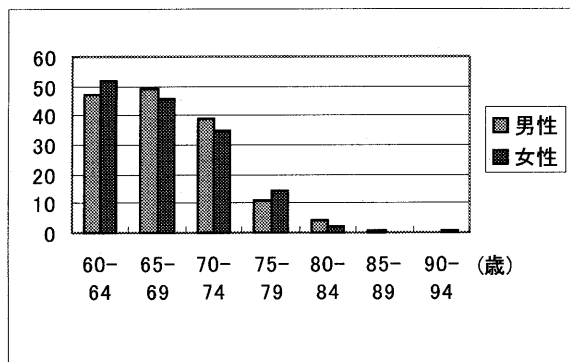


図 5 学習用音声データ話者年齢構成

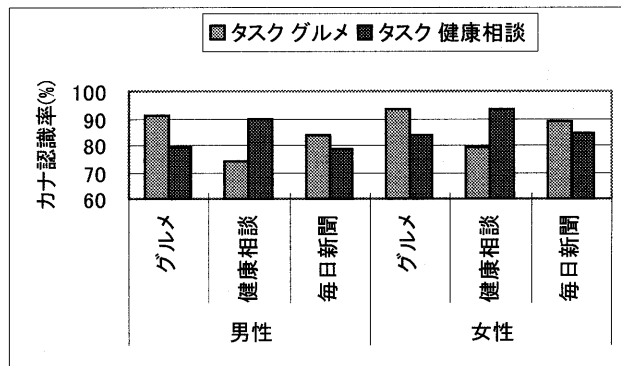


図 7 カナ単語認識率

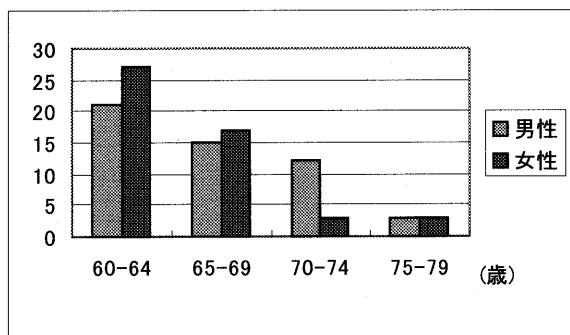


図 6 評価用音声データ話者年齢構成

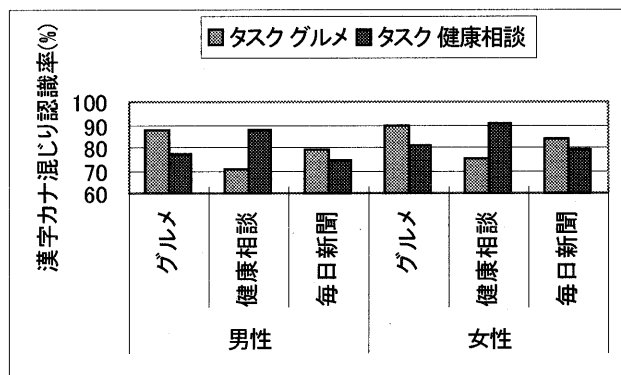


図 8 漢字カナ混じり単語認識率

つまり各タスクに対応した言語モデルの性能が良いことを示している。

5. 大語彙連続音声認識実験

大語彙連続音声認識エンジン Julius[1]を用いて認識実験を行った。言語モデルは、3節で作成したものを使用した。音響モデルは、301人×200文章の高齢者音声によって学習した高齢者向けモデル(PTM, 2000状態, 64混合, 性別依存)を使用した[2]。音響モデルの学習用音声データの話者年齢構成を図5に示す。

テストセットには、前節の評価実験で使ったテキストを読み上げた高齢者の音声データ(男性51人、女性50人)を利用した。話者の年齢構成を図6に示す。

認識実験の結果を図7および図8に示す。なお、認識率はカナ単語認識率と漢字カナ混じり単語認識率を用いた。

グルメタスクでは、新聞記事の言語モデルを使用した場合に比べ、グルメタスク言語モデルを使用した方が単語認識率が男性で8.8%、女性で6.3%高い。健康相談タスクでも同様に健康相談タスク言語モデルの方が男性で13.1%、女性で11.3%向上している。

実験結果から新聞記事の言語モデルよりタスク対応の言語モデルを使用する方が、大幅に認識率が向上することがわかった。

6. まとめ

話し言葉を認識できる言語モデルを構築するために、Web掲示板からコーパス収集を行い、収集したコーパスから作成した言語モデルの評価を行った。認識実験の結果からタスク対応言語モデルの有効性が確認できた。

今後は、他のタスクに関しても、コーパス収集、言語モデルの構築を行い、評価を行っていく予定である。

謝辞

本研究の援助をしていただいたNEDO(新エネルギー・産業技術総合開発機構)、TIS株式会社および関係者各位に感謝いたします。

参考文献

- [1] 李, 河原, 堂下: “単語トレリスインデックスを用いた段階的探索による大語彙連続音声認識”, 電子情報通信学会論誌, J83-D II, No. 12, pp. 2517-2525, 2000
- [2] 馬場, 芳澤, 山田, 李, 鹿野: “高齢者音声向け音響モデルによる大語彙連続音声認識の評価”, 情報処理学会研究会報告, 2001-SLP-35-3, 2001