音声言語情報処理 39－27
(2001. 12. 21)

# 旅行会話基本表現コーパスを用いた認識誤り訂正の検討

沖本 純幸† 山本 博史† 隅田英一郎† 菊井玄一郎†

† ATR 音声言語コミュニケーション研究所
〒 619－ 0288 京都府相楽郡精華町光台 2-2-2

E-mail: †{yoshiyuki.okimoto,hirofumi.yamamoto,eiichiro.sumita,genichiro.kikui}@atr.co.jp

**あらまし** 我々は音声認識における認識誤り箇所を検出しこれを訂正するという手法の検討を進めている．本論文では，検出された誤り箇所の訂正方法として，旅行会話基本表現コーパスによる用例を用いた訂正方法について提案する．本誤り訂正では，1) 類似用例の検索，2) 用例中の代替候補単語の抽出，3) 音響的類似度等によるリスコアリング，という方法を採る．評価実験の結果，認識誤り箇所の 20% 以上を訂正できることが確認された．本論文ではこの手法と幾つかの分析結果を報告する．

**キーワード** 誤り訂正，用例コーパス，旅行会話表現

# Correcting Mis-recognitions Using Basic Travel Expression Corpus

Yoshiyuki OKIMOTO†, Hirofumi YAMAMOTO†, Eiichiro SUMITA†, and Genichiro KIKUI†

† ATR Spoken Language Translation Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan

E-mail: †{yoshiyuki.okimoto,hirofumi.yamamoto,eiichiro.sumita,genichiro.kikui}@atr.co.jp

**Abstract** We aim at improving the performance of our speech recognition system by detecting and correcting mis-recognitions. In this paper, we propose a new error correction method based on large basic travel expression corpus. This method has three steps: 1) Searching similar examples, 2) Extracting hypothesis words from examples, 3) Rescoring hypothesis words by using phonetic distance. Our experiments show that more than 20% of mis-recognitionsare corrected. In this paper, we also report some analyses of error correction results.

**Key words** Error correction, Example Corpus, Travel expression

## 1. はじめに

近年多くの音声認識システムが開発され，その幾つかは実用レベルに近付きつつある．しかし，現状の認識システムにおいて誤認識の問題は避けられない．従来は認識誤りを，モデルと探索手法の改良によって低減するアプローチが採られてきたが，我々は認識誤り箇所の検出とこれの訂正というアプローチの検討を進めている．これは，従来の音声認識の枠組によって得られる認識結果について，より広い範囲のコンテキスト情報などを用いることで，局所的な認識誤りを訂正できるとの考えに基づく．

我々は過去において認識誤り箇所の検出について検討を行なってきた[1]．今回は認識誤り箇所の訂正方法について検討を行なう．認識誤り箇所訂正のための先行研究は，OCR の分野[2],[3]で多くなされてきた．しかし，文字認識に比べて，音声認識では入力音声の曖昧性がはるかに大きいため探索空間が広がり過ぎて，これらの手法を音声認識にそのまま適用するのは難しい．音声認識の分野では，例えば正しく認識できた部分のみ翻訳する方法[4]などが提案されているが，誤り訂正そのものを行なう検討はあまり多くなされていない．

その中で，石川ら[5]は，検出された誤り箇所に対して音韻的に類似した用例を用いて代替候補を生成し，音韻的距離と意味的距離によって候補の妥当性を判断する誤り訂正法を提案して一定の成果を収めた．しかしこの方法で用いている意味距離は，多義語の問題などのために充分な制約力を持たず，またオープンテストに対する性能も良くなかった．

我々が対象としている旅行会話のような対話における多くの発話は，会話表現集に見られるように単文を中心とした比較的単純な構造であり，大量の用例を収集することによって頻度の高い表現をカバーすることがきると考えられる．そこで我々は，用例コーパス中の用例文を用いて，用例文そのものを制約とする誤り訂正法を提案する．この提案法では，認識結果文の類似用例を用例コーパスから探索し，これを基に誤り箇所の代替候補を生成し訂正を行なう．評価実験の結果，提案方法によって 20% 以上の誤り箇所が正しく訂正できることが確認された．

## 2. 提案方式

音声認識器は通常複数の認識候補 (N-best) を出力するが，本論文ではこのうちの 1 位候補 (1-best) の系列のみを用いて，その誤り箇所を訂正する．この認識誤り箇所訂正の問題をここでは次の 3 つに分けて考える．

（1） 認識誤り箇所の検出
（2） 代替候補の生成
（3） 最適候補の選択

この中で認識誤り箇所の検出の問題は重要なテーマであるが，他稿に譲り ([8],[9] など)，本論文では後者 2 つについて検討を行なう．

我々は認識誤り箇所訂正のための情報源として，認識結果文と類似する用例文そのものを用いることを提案する．これは，同一ドメインで同じ内容を伝える文はいくつかの類似した文に大別されるであろうという仮定に基づいており，また用例文という強い言語的制約によって，誤り箇所が訂正されることを期待するものである．従って本提案手法は，あらゆる発話の認識誤りを保証するものではなく，認識対象と同一ドメインの用例コーパスによって規定された，対象領域の発話の認識誤りを訂正するものである．

### 2.1 本研究における認識誤り箇所

本論文においては，問題を明確にするため，認識誤り箇所は正確に検出できるものとして検討を行なう．すなわち，あらかじめ手作業で与えた正解の単語ラベルと，認識器の出力する 1-best の単語系列の単語ラベルを比べて，置換および挿入誤りであった部分を認識誤り箇所としている．脱落誤り箇所については，認識結果中に該当単語が表われないため検出できないので無視する．

### 2.2 代替候補の生成

検出された音声認識誤りの箇所に対する代替候補を生成する．代替候補の探索範囲として用例コーパスを用いる．すなわち認識結果文と類似する用例をコーパス中から探索し，これに表われる単語を当てはめる．類似用例の選択には，コーパス中の各文と認識結果文で DP マッチングに基づく距離計算を行ない[6]，これを類似度として順序付けを行なう．またこれと同時に得られる，認識結果文の各単語と用例文の各単語の対応関係から，認識結果の誤り箇所に対応する単語を用例文より取り出し代替候補とする．

#### 2.2.1 用例コーパス

我々は，旅行会話音声翻訳器のための音声認識部の開発を進めている[7]．したがって我々の音声認識

| | |
|---|---|
| 文数 | 92,668 |
| 平均文長 | 7.4 |
| 総単語数 | 681,525 |
| 語彙数 | 14,419 |

表1　用例コーパス

| | |
|---|---|
| 音響モデル | HMnet (コンテキスト依存, 男女別) |
| 言語モデル | 多重クラス複合 2-gram |
| 探索 | 2 パス, ビーム探索 |
| 辞書 | 語彙数 38,173 語 |

表2　音声認識システムの概要

器の認識タスクは, 旅行会話音声である. このタスクの音声の認識誤りを訂正するための用例コーパスとして, 旅行者向けのフレーズブックに表われるような旅行会話の基本表現を大量に集めたものを用いる. 集められた用例文の数は, 重なりなしでおよそ 9 万文である. 表 1にこの用例コーパスの統計値をまとめる.

**2.2.2**　用例の選択

認識誤り箇所の代替候補の集合を生成するため, 用例コーパスから認識結果文の類似文を検索する. 類似文の探索は 認識結果文と各用例文の DP マッチングに基づいて行なう. 認識結果文中の認識誤り箇所については, "誤り単語" を意味する特別な ID を与えて, 他のいかなる単語とも異なる単語としてマッチングを行なう. 認識結果文と用例文の距離を定義する式を以下に示す.

$$dist = \frac{I + D + \sum d_i}{L_{in}} \qquad (1)$$

ここで $I$ および $D$ は, それぞれ挿入誤りと脱落誤りの回数を示しており, $d_i$ は置換誤りにおける単語間の距離を示している. また $L_{in}$ は, 認識結果文の単語系列長を意味する.

単語間の距離としては, いくつかの方法を考えることができる. 最も単純には, $d_i = 1.0$ という定数値を与える方法である. 今回はこのような単語間距離に加えて, 単語間の意味距離を用いる方法についても検討した. これは, シソーラス上での意味属性の位置関係により単語間に 0 ～ 1 の意味距離を与えるものである (詳細は後述).

以上のような距離計算によって各用例ごとに得られた類似度のうち, 上位 $n$ 位の類似度を有する用例を代替候補単語を含む文として選択する.

**2.3**　最適候補の選択

認識誤り箇所に対する代替候補の集合中から, 誤り箇所に最も当てはまると考えられる単語を選択する. これには認識結果の当該箇所の音素系列と, 発音辞書に示された代替候補単語の音素系列を DP マッチングさせた結果を用いて, 次に示すスコアによって行う.

$$score = \lambda \cdot dist + (1 - \lambda) phdist \qquad (2)$$

$$phdist = \frac{I_{ph} + D_{ph} + S_{ph}}{L_{ph}} \qquad (3)$$

ここで $I_{ph}$ $D_{ph}$ $S_{ph}$ は, それぞれ認識結果の音素系列と, 代替候補単語の音素系列の DP マッチングによる挿入誤り, 脱落誤り, 置換誤りの回数である. また $L_{ph}$ は, 誤り区間の音素系列の系列長である.

この式は, 認識結果文と用例文の文としての近さと, 単語の音素系列の近さ — すなわち音響的近さ, の 2 つの要因の重み付け和によって代替候補単語を順序付けすることを表わす.

## 3.　提案方式の評価

### 3.1　実験条件

提案方式の基本性能を評価する実験を行なう. 実験に用いた音声認識システムの概要を表 2にまとめる.

認識に用いた音声は, 2.2.1節で述べたコーパスを作成する際には利用しなかった別の旅行会話文セットを, 複数の男女が発声したものを用い, 総発話数は 2,037 発話であった. 前述の音声認識システムによる評価では, この音声データに対する認識精度は単語 Accuracy で 83.9% であった. 正解の単語ラベルと認識結果を比較した結果では, 計 897 箇所に認識誤りがあった.

実験では, 2.2.1に述べた用例セットを用いて認識誤り箇所の訂正を行なう. 用例の類似度 (文間距離) の計算は式 (1) に基づいて行ない, 単語置換の距離 $d_i$ は一律に 1.0 を与えた. 代替候補の順序付け式 (2) では, 認識結果の音素系列として, 誤り区間の 1–best の単語の音素系列をそのまま用いるものとした.

### 3.2　実験結果

以上に述べたような条件の下で, 次の点に着目した評価を行なった.

- 1 つの誤り箇所に対する平均代替候補数
- 全誤り箇所に対して, 正解候補を含む代替候補が得られた誤りの箇所の割合 (可能誤り訂正率, correctable error rate)

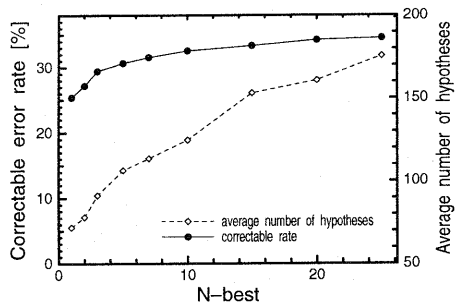


図1 Average number of hypotheses & Correctable error rate

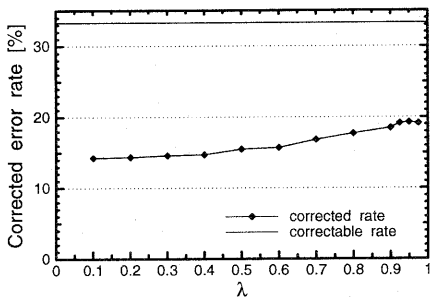


図2 Corrected error rate

- 全誤り箇所に対して, 代替候補中から正しく正解候補を選び出すことのできた誤り箇所の割合 (誤り訂正率, corrected error rate)

ただし認識誤り箇所とは, 連続した誤り単語をまとめた区間を指す. これは, 本提案方式では連続して誤っている区間全体をまとめて訂正を行なうためである.

この結果を図1, 2に示す. 図1は, 類似度 (文間距離) の上位何位までを類似用例とみなすか (上位選択順位数,N-best) に対する平均代替候補数および可能誤り訂正率を示すものである. 上位選択順位数に対する平均代替候補数は, 縦軸右側のスケールで示されており, 可能誤り訂正率は縦軸左側のスケールで示されている.

本来, 上位選択順位数と平均代替候補数は近い値となると予想されるが, ここでは平均代替候補数が非常に多い. これは, 同じ類似度を持つ用例が非常に多く表われるケースがあるためである (たとえば, 「<Error> をお願いします」という誤り文では, 非常に多くの類似用例が得られる). 図1に示された結果からは, 上位選択順位数を大きくするに伴ない, より多くの認識誤り箇所に対して正解候補を含む代替候補が得られるようになるが, 選択順位を 10 位以上に広げても, 可能誤り訂正率は伸びないことが示されている. また選択上位順位数を 15 位とした場合で, 可能誤り訂正率は 約 33% であり, 本実験で行なった方法で, 1/3 の誤り箇所が正しく訂正できる可能性があることを示している.

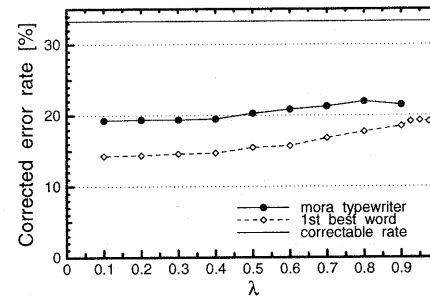


図3 Rescoring by using mora typewriter

また図2は, 代替候補を 2.3節に述べた方法によって順序付けた場合の誤り訂正率を示している. このグラフにおいて横軸は, 式 (2) における λ の値を示し, 縦軸は誤り訂正率を示している. このグラフでは, 上位選択順位数は 15 位としている. 比較のため, 図1に示した上位選択順位数 15 位における可能誤り訂正率を, 誤り訂正率の上限値として合わせて示す. この結果から, 本手法により全誤り箇所の約 20% を正しく訂正できることが示されている.

## 4. 音素距離の検討

前節の実験では, 用例により得られた代替候補を式 (2) に基づいて順序付けする際, 1 位の (誤り) 認識単語の音素系列と, 代替候補単語の音素系列の類似度を用いた. しかし音声認識の過程で, 単語候補の生成には音響モデルと言語モデルという 2 つの制約が用いられている. このため, 誤り認識単語の音素系列は言語的な制約による影響を受けていおり, しかもこれは誤った影響である可能性がある.

そこで本節では, 音響モデルだけの制約によって生成される音素系列を用いた場合についての比較実験を行なう. 具体的には, 音声認識の過程でモーラタイプライタを並行に走らせ, 認識誤り区間と判定された区間に該当する音素系列を, このモーラタイプライタ結果から取り出す.

この実験の結果を, 先の誤り認識単語の音素系列を用いた場合と合わせて図3に示す. 縦・横軸の意味は図2と同じで, 音素類似度と文類似度の重みに対する, 誤り訂正率の関係を示している. この図に

示されるように, 参照する音素系列としてモーラタイプライタによるものを用いた方が, より多くの誤り箇所を正しく訂正できることが示されている.

この結果に基づき, 以下の評価実験では最適候補選択のための音素系列として, モーラタイプライタによる音素系列を用いるものとする.

## 5. 文類似度の評価

### 5.1 文類似度

これまでの実験では, 認識結果文と用例文の距離として, 単語間の距離を定数値 1.0 とした結果について示した. 本節では, 単語間の距離を単語ごとに変えた場合について評価する.

本節で行なう実験では, 単語間の距離を意味距離に基づいて可変とする. ここで言う意味距離は, シソーラス上での単語意味属性の一致度で計算をする. これは 単語のシソーラス階層数が $N$ で, 2 つの単語の意味属性が最上層から第 $N-k$ 階層まで共通であるなら, 以下の式によって計算される.

$$d_{sem} = \frac{k}{N} \tag{4}$$

今回の実験では, シソーラスの階層数は $N=3$ で固定としているため, $d_{sem} = \{\delta, \frac{1}{3}, \frac{2}{3}, 1\}$ の 4 種類の値を取る. ($d_{sem}$ の最小値を $\delta$ としたのは, 同一語間の距離を 0.0 とし, これとの差別化を図るためである.)

しかし, 通常 1 つの単語には複数の語義が付される. このため上式のみでは, 単語間の距離を決定することができない. 本実験では, 以下の式により評価を行なった.

$$d(w_1, w_2) = \min_{i,j} d_{sem}(w_{1_i}, w_{2_j}) \tag{5}$$

ここで $w_{n_i}$ は単語 $n$ の第 $i$ 番目の語義を示す. この式は, 単語の複数語義の間の最小意味距離を当該単語間の意味距離とすることを意味する.

### 5.2 実験結果

実験結果を, 単語間距離を 1.0 に固定した場合と比較して, 図 4に示す.

この結果からは, 用例検索において単語間の意味距離を用いることによる効果は表われていない. この原因は次のように考えられる.

認識結果文と用例文を対応させて文類似度を計算する際, 認識結果中の認識誤り箇所は, "誤り単語" を意味する ID によってマスクされており, いかなる単語とも距離は 1.0 となる. このため肝心な箇所で意味距離が働かず, その効果が低減してしまっている.

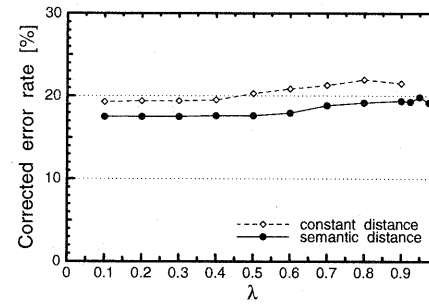


図 4 Searching examples by using semantic distance

また, 単語が持つ複数の語義を解消する方法として, 今回, 最小距離用いたが, このような単純な方法では充分な解消ができていないことも原因として挙げられる.

## 6. 誤り訂正結果の分析

### 6.1 誤り長ごとの評価

本論文中では, 誤り箇所とは連続する誤り単語の区間を指す. 本提案方式によって, どの程度の誤り長 (誤り区間における正解単語の系列長) の誤りを訂正できるのか調査した.

この結果を誤り長に対する可能誤り訂正率, および誤り訂正率の関係として図 5に示す. また同時に, 各誤り長の誤りが, 全誤りの中でどの程度の割合で出現するかを棒グラフによって重ねて示している. 本図において, 可能誤り訂正率および誤り訂正率は, 縦軸左側のスケールで示されており, 各誤り長の割合は縦軸右側のスケールで示されている.

この実験の結果から, 本提案方式により訂正可能な誤りは誤り長 2 程度までの誤りであって, これ以上の長い誤りについてはほとんど 訂正の見込みがないことが示されている. なお誤り長 2 までの誤りは, 全誤り箇所のおよそ 80% であることもグラフから読み取れる.

### 6.2 品詞ごとの評価

誤り箇所の単語に対し, その品詞ごとの可能誤り訂正率, 誤り訂正率を比較する. ここでは, 長さ 1 の誤り箇所のみについて, その正解単語の品詞ごとに分類する.

この結果を図 6に示す. 図中で棒グラフは, 誤り長 1 の全ての誤り箇所に対して, 各品詞の占める割合を縦軸右側のスケールで示している. また各品詞

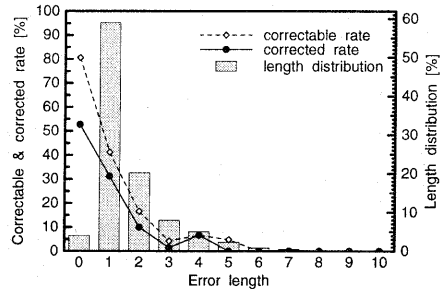


図5 Error length distribution

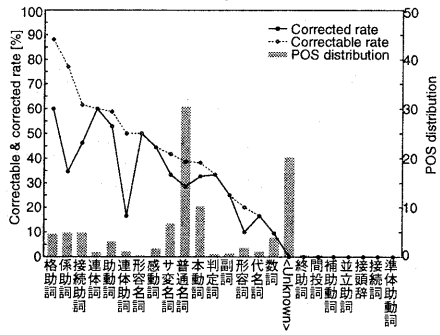


図6 Error POS distribution

ごとの可能誤り訂正率および誤り訂正率を，折れ線グラフによって縦軸左側のスケールで示している．

この結果からは，第1に，格助詞といった機能語の誤りについてはこれらの語が前後のコンテキストから比較的容易に正解語が類推できること，内容語に比べて種類が少ないこと等の理由により，比較的高い割合で代替候補中に正解語が含まれることが判る．しかし，最適候補選択においては，正解候補を正しく選択できていない．これは，これらの語が1, 2音節程度の短かい語であるため，候補選択において音素距離による差がつきにくいためであると考えられる．

また第2に，普通名詞，本動詞，形容詞といった内容語については，代替候補を選択した段階ですでに正解語が含まれていないことが多い．これは，これらの品詞は単語の種類が多いため，同様のコンテキストで交換可能な語が多く，用例文から得た代替候補のみでは候補生成が不十分であるためと考えられる．しかし，棒グラフで示された誤りに占める品詞の割合から言っても，また文の内容を正しく捉えるという意味においても，特に普通名詞，本動詞の誤りは，訂正の効果を上げる上で重要なターゲットであると考える．

## 7. ま と め

本論文では，音声認識の誤り箇所の訂正方法として，用例コーパスを用いて，認識結果文の類似用例を検索し，これによって誤り箇所を訂正する方法を提案した．

用例の検索方法，代替候補のリスコアリング方法などいくつかの比較実験を行なった結果，全認識誤り箇所のうちの 22.0% を正しく訂正できることを確認した．また，リスコアリングにより正しく正解候補を導くことのできなかった誤り箇所でも，代替候補中には正解候補が含まれているケースがあり，適切なリスコアリング方法を用いることでさらに訂正性能が向上する可能性がある．

また実際に訂正できた誤り箇所についての分析から，本訂正手法は，全誤りの 80% を占める，誤り系列長が 2 以下の誤りに対して効果的であるこが確認された．さらに，品詞ごとの訂正性能の比較から，多く誤っている普通名詞や本動詞といった品詞の誤り訂正が不十分であることが示された．これは，単語の共起関係を用いるなど，このターゲットに絞ったアプローチによってさらに訂正性能を向上させることが可能ではないかと考えている．

**謝辞** 本研究を行なうにあたり，角川書店より角川類語新辞書を提供して頂きました．ここに深く感謝いたします．

### 文 献

[1] Y.Okimoto et al.:“Evaluation of Mis-Recognition Detection Using COnfidence Measures”,In Proc. ICSP’01, pp. 685–690, 2001

[2] 竹内孔一 ら:“統計的言語モデルを用いたOCR誤り訂正システムの構築”, 情報処理学会論文誌, 40 (6), pp.2679–2689

[3] 永田昌明:“文字類似度と統計的言語モデルを用いた日本語文字認識誤り訂正手法”, 電子情報通信学会論文誌, J81-D-II (11), pp.2624–2634

[4] 脇田由美 ら:“意味的類似性を用いた音声認識正解部分の特定法と正解部分のみ翻訳する音声翻訳手法”, 自然言語処理, 5 (4), pp.111–125, 1998

[5] 石川開 ら:“テキストデータを使った音声認識誤りの訂正”, 自然言語処理, 7 (4), pp.205–227, 2000

[6] E.Sumita:“Example-based machine translation using DP-matching between word sequences”, In Proc. ACL-2001 Workshop (DDMT), pp. 1–8, 2001

[7] T. Takezawa et al.:“A Japanese-to-English Speech Translation System: ATR-MATRIX”, In Proc. ICSLP’98, pp. 957–960, 1998

[8] T. Kemp et al.:“Estimating Confidence Using Word Lattice”, In Proc. Eurospeech’97, pp. 827–830, 1997

[9] F. Wessel et al.:“Using Word Probabilities as Confidence Measures”, In Proc. ICASSP’98, pp. 225–228, 1998