

## ニュースの直前原稿を利用した音声認識誤りの自動検出法

松井 淳 加藤 直人 小林 彰夫

今井 亨 田中 英輝 安藤 彰男

NHK 放送技術研究所

NHKは平成12年3月27日より音声認識による「ニュース7」の字幕放送を試行的に開始した。「ニュース7」では、認識結果を人手により確認・修正して字幕を放送している。認識結果のみが与えられた場合、全ての認識誤りを正確に自動検出することは難しいが、ニュース音声認識では、事前に記者が作成した原稿（記者原稿）を利用することができる。本報告は、まずニュースの約7割の発話が、放送直前の記者原稿に対応することを示す。次に、記者原稿を利用した認識誤りの検出法を具体的に示し、実験により、その誤り検出の性能を評価する。

### An Automatic Error Detection Method for Broadcast News Transcription Using the Latest News Manuscripts

Atsushi Matsui, Naoto Kato, Akio Kobayashi,  
Toru Imai, Hideki Tanaka, and Akio Ando

NHK Science and Technical Research Laboratories

Although it is almost impossible to guess what will be uttered beforehand in general speech recognition task, we can do this with high possibility in news transcription task by accessing a database of news manuscripts. This is because announcer's reading scripts are usually created by editing some manuscripts in the database. In this paper, we will describe an error detection technique for a news transcription system utilizing such news manuscripts.

#### 1 まえがき

近年、聴覚障害者や高齢者を中心に、生番組とくにニュース番組の字幕サービスを求める声が高まっている。NHKはこうした要望を受けて、平成12年3月27日よりNHK総合テレビ「ニュース7」の一部で字幕放送を試行的に開始した。「ニュース7」で

は、アナウンサーの声をリアルタイムで音声認識して字幕を放送している。ニュースでは、何よりも放送内容の正確性が重要視されるが、現在の音声認識技術では、100%正確な認識結果を得ることは困難である。そのため「ニュース7」では、認識結果を人手により確認・修正した上で字幕を放送している

[1]。日本語文書における誤りの検出法としては、例えば、品詞 N-gram を用いる手法 [2] が提案されている。これは、新聞などの大規模なコーパスから抽出した品詞の連鎖に関する統計情報 (品詞 N-gram) とヒューリスティックなルールを用いて表記誤りを検出する手法である。しかし、ニュース発話で出現する単語には、形態素解析で品詞を特定できない単語 (未定義語) が多いため、品詞情報を用いる手法には限界がある。そこで本報告では、放送前に作成される記者原稿 (直前原稿) を利用した認識誤りの自動検出法を提案し、実験によりその検出性能を評価する。

## 2 発話と記者原稿

アナウンサーが放送で実際に読み上げる原稿を「読み原稿」と呼ぶ。読み原稿は、記者が事前にワープロ入力した「記者原稿」や、現場からの FAX 原稿に、手書きによる修正が加わって作成される。読み原稿は現時点では電子化されておらず、利用できない。一例として、NHK ニュース発話 (書き起こし) と、それに対応する記者原稿を示す。(1998 年 9 月 30 日放送分より。下線は一致箇所。)

(記者原稿)

台風九号の接近に伴い九州 沖縄各地で  
交通機関に影響が出ています

(書き起こし)

台風九号の接近で九州では  
交通機関に乱れが出ています

実際のニュースにおいて、放送前に入稿された記者原稿とニュースの発話がどの程度重なっているかを見るため、1998 年 9 月 30 日放送、「ニュース 7」の書き起こしと記者原稿の対応を調べた。ニュースの中には、一般話者のインタビューや、中継での掛け合いなど、記者原稿をあらかじめ用意しないもの<sup>1</sup>があるため、これらの項目を除いた計 196 文の書き起こしを作成した。そして、記者原稿を正解候補、書き起こしを出力文とみなし、音声認識で通常用いる単語正解精度 (以下、「類似度」と呼ぶ) を使って

<sup>1</sup> 記者原稿を用意しない発話には、これら以外に、ニュース冒頭の挨拶、ニュース項目の「つなぎ」がある。

書き起こしを次の 3 つに分類した。

- 「完全一致」(類似度 100%)
- 「部分一致」(類似度 50%以上 100%未満)
- 「一致なし」(類似度 50%未満)

以後、本稿ではこの分類を一貫して使う。

「完全一致」、「部分一致」および「一致なし」の書き起こしが、全書き起こしに占める割合の時間変化を、グラフにまとめ図 1 に示す。図 1 の横軸は、「ニュース 7」の開始時間、19 時から遡った原稿収集時間である。

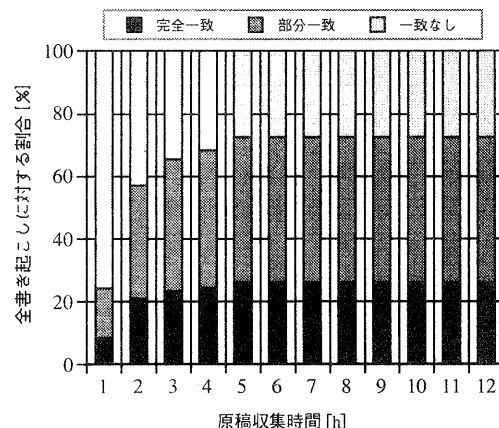


図 1 記者原稿に対応するニュース発話

図 1 のグラフで、原稿収集時間を 1 時間から 2 時間に変えた部分で、「完全一致」と「部分一致」の割合が、急激に増加している。また、原稿収集時間が 6 時間を越えたところで、「完全一致」と「部分一致」の割合が一定となった。このときの全書き起こしに占める各割合は、完全一致: 26.0%、部分一致: 46.4%、一致なし: 27.6%であった。「完全一致」と「部分一致」をあわせた書き起こしは全書き起こしの 7 割以上ある。このように、ニュース中の発話は記者原稿に類似したものが多く、記者原稿を誤りの検出に利用する手法を検討した。

### 3 提案手法

#### 3.1 誤り検出における正解文

前章で「書き起こし」と「記者原稿」の類似度（単語正解精度）を調べた結果、高い類似性を見た。一方、認識結果に含まれる誤りは、記者原稿とは特に関係がない。従って、認識結果の単語列が、記者原稿の単語列と、十分長い単語数にわたり一致するとき、その単語列に誤りが存在する可能性は低いと考えられる。そこで本稿では、認識結果と最も長い一致単語列<sup>2</sup>を持つ記者原稿の文を、認識結果の「正解文」とみなす。

記者原稿のある文を「正解文」として仮定する手法として、尾上ら [4] の方法が既に提案されている。しかし、これは文単位の比較（DP マッチング）により「正解文」を決めるため、認識結果が文末に到達するまで誤りを検出できない。そこで著者らは、単語列単位の比較による「正解文」の抽出法を検討し、早期の認識誤りの検出を試みた。

#### 3.2 単語出現位置辞書

単語列が記者原稿に出現した位置を調べるため、単語出現位置辞書 [5] を用意する。単語出現位置辞書は、記者原稿において各単語が出現した位置（単語位置）と、出現した文を指す番号（文番号）を記したものである。単語出現位置辞書の作成法の概要を図 2 に示す。

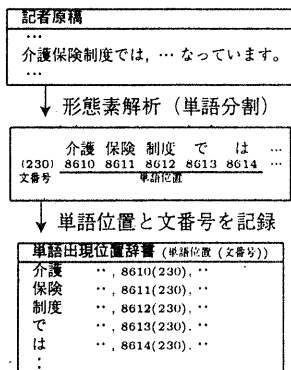


図 2 単語出現位置辞書の作成法

<sup>2</sup> 認識単語と複数の一致単語列を持つ場合には、その合計の長さで評価する。

#### 3.3 正解文の決定

与えられた認識結果から「正解文」を決定する方法を説明する。まず、認識結果の単語列を、記者原稿に出現した単語列<sup>3</sup>で分割する。そして、第 3.1 節で示した定義に従い、(合計で) 最も長い単語列をもつ文を「正解文」とする。

ここで、認識結果のある単語  $w_i$  について、複数の単語列

$$(w_j, \dots, w_i, \dots, w_k),$$

$$(w_m, \dots, w_i, \dots, w_n),$$

が記者原稿に出現する場合、単語列の長さが長い方で認識結果を分割する。ただし、単語列の長さが同じである場合は、先頭の単語がより認識結果の先頭に近いもので分割する。認識結果の単語列

介護/保険/制度/で/和解/が

を、記者原稿に出現した単語列で分割した例を図 3 に示す。図 3 の例では、認識単語「で」について、「で」を含む次の 2 つの単語列が記者原稿に出現する。

(介護/保険/制度/で)

(で/和解/が)

この場合、前者の単語列の方が 1 単語長いので、認識単語「で」は、単語列 (介護/保険/制度/で) の要素とする。

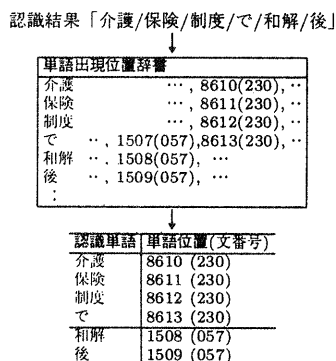


図 3 認識結果の単語位置

<sup>3</sup> ある単語列が記者原稿に出現する場合、各単語の単語位置は必ず連続する。任意の単語列が記者原稿に出現したかを調べるには、このことを利用する。

次に、逐次与えられる認識結果から、正解文を決定(仮定)する方法について、図4に示す例を用いて説明する。図4は、 $i$ 番目の単語が認識された時刻  $t_i$  における認識結果と単語位置を示す。各単語の単語位置は、各時刻における認識結果について、前述の手法により求める。

|                            |                |                      |
|----------------------------|----------------|----------------------|
| $t = t_1$                  |                |                      |
| 認識結果「介護」                   |                |                      |
| 認識単語                       | 単語位置(文番号)      | $H(230) = \dots = 1$ |
| 介護                         | 8610(230), ... | 正解文: -               |
| $t = t_2$                  |                |                      |
| 認識結果「介護/保険」                |                |                      |
| 認識単語                       | 単語位置(文番号)      | $H(230) = \dots = 2$ |
| 介護                         | 8610(230), ... | 正解文: -               |
| 保険                         | 8611(230), ... |                      |
| $t = t_3$                  |                |                      |
| 認識結果「介護/保険/制度」             |                |                      |
| 認識単語                       | 単語位置(文番号)      | $H(230) = 3$         |
| 介護                         | 8610(230)      | 正解文: 230             |
| 保険                         | 8611(230)      |                      |
| 制度                         | 8612(230)      |                      |
| $t = t_4$                  |                |                      |
| 認識結果「介護/保険/制度/で」           |                |                      |
| 認識単語                       | 単語位置(文番号)      | $H(230) = 4$         |
| 介護                         | 8610(230)      | 正解文: 230             |
| 保険                         | 8611(230)      |                      |
| 制度                         | 8612(230)      |                      |
| で                          | 8613(230)      |                      |
| :                          |                |                      |
| $t = t_8$                  |                |                      |
| 認識結果「介護/保険/制度/で/和解/後/が/必要」 |                |                      |
| 認識単語                       | 単語位置(文番号)      |                      |
| 介護                         | 8610(230)      | $H(230) = 6$         |
| 保険                         | 8611(230)      |                      |
| 制度                         | 8612(230)      |                      |
| で                          | 8613(230)      | $H(057) = 2$         |
| 和解                         | 1508(057)      |                      |
| 後                          | 1509(057)      | 正解文: 230             |
| が                          | 8616(230)      |                      |
| 必要                         | 8617(230)      |                      |

図4 正解文の決定

時刻  $t_1$  では、認識単語「介護」の出現位置は、単語出現位置辞書に記した複数の出現位置が候補となる。このとき、各単語位置: 8610(230), ... についての単語列の合計長:  $H(230), \dots$  は全て"1"となり、正解文が一意に決まらない。そのため、正解文が未定の状態で、次の単語が認識される ( $t = t_2$ ) のを待つ。

時刻  $t_2$  では、認識結果の単語列(介護/保険)が記者原稿の複数の文に出現するため、正解文が一意

に決まらない。正解文を未定として、次の認識単語を待つ。

時刻  $t_3$  では、単語列(介護/保険/制度)が記者原稿に1回出現し、それぞれの認識単語の単語位置が一意に決定する。このとき、 $H(230) = 3$  が唯一の最長単語列長となり、文番号: 230 が指す文(以下、文 230 と略す)が「正解文」となる。

時刻  $t_4$  では、時刻  $t_3$  の場合と同様に、文 230 を「正解文」とする。

さらに認識が進むと、記者原稿の別々の文に出現した単語列が、一つの認識結果に現れる場合がある。

時刻  $t_8$  では、単語列(介護/保険/制度/で)と(が/必要)は文 230 に出現するが、単語列(和解/後)は文 057 に出現する。このときの単語列長は、文 230 が合計6単語であるのに対して、文 057 は2単語である。前者の方が長いので、文 230 が「正解文」となる。このように、認識結果が部分的に複数の文と一致する場合、単語列長の合計が最大となる文を「正解文」とする。

この手法で、発話と類似する文が記者原稿に存在しない場合では、 $H(n)$  が数単語程度の短い長さで、偶然一致した文  $n$  が「正解文」として選ばれる可能性がある。そこで、 $H(n)$  をある閾値  $H_{min}$  と比較して、 $H(n) \geq H_{min}$  となる場合に限り、記者原稿の文  $n$  を正解文として仮定する。第4章で行う評価実験では、このパラメータ  $H_{min}$  を変化させて評価実験を行う。

### 3.4 認識誤りの検出

前節で示した手順で正解文を決めた後、まず、認識結果の単語列の中で、認識誤りが含まれている可能性の高い「誤り候補」を抽出する。以下、「誤り候補」の抽出の手順を模式的に示す。

```

if ( (「正解文」を仮定している) ) and (
    (正解文以外の文と対応する単語(列)が、
     正解文と対応する単語列で挟まれている) )
) then {
    挟まれた単語(列)を「誤り候補」とする
} else {
    次の認識単語を待つ
}

```

誤り候補が見つかった場合、その前後の単語の単語位置を利用して、誤り候補に対応する正解文の単語を単語出現位置辞書から取り出す。取り出した

単語を一続きの文字列に結合し、認識誤りに対する「正解文字列」を用意する。誤り候補の単語が正解文字列に存在しない場合、その単語を「誤り」とする。認識誤り検出の例を図5に示す。

認識結果が文末に到達した時点で、次単語待ちの状態にある場合には、誤り検出を行わずに処理を終了する。

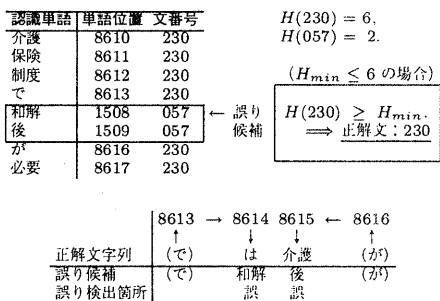


図5 認識誤りの検出

## 4 評価実験

前章で述べた手法の有効性を調べるために、誤り検出の実験を行なった。実験の対象は、「ニュース7」(1998年9月30日)の認識結果である。

### 4.1 実験条件

認識結果には、第2章の「ニュース7」の発話を、認識した結果(計196文、総合単語数6502、単語正解精度93.7)を用いた。

記者原稿は、第2章の実験結果より、類似する発話の割合が飽和する放送6時間前から放送開始までに入稿した原稿(計512文、総単語数18398)を用いた。

「ニュース7」の認識結果196文を、書き起こしと記者原稿の類似度により分類した。内訳は、下記のとおりである(図1での累積収集時間=6時間の場合と同じ)。

- 「完全一致」… 54文
- 「部分一致」… 84文
- 「一致なし」… 58文

検出対象である認識誤りの総数は、335単語である。

## 4.2 実験結果

パラメータ  $H_{min}$  を変化させた場合について、全発話(196文)について認識誤りを検出した結果の、再現率、適合率、F値を、それぞれ調べた。

ここで、F値とは再現率  $R$  と適合率  $P$  の両者を考慮した尺度であり、次式で定義される。

$$F = \frac{1}{\alpha \frac{1}{P} + (1 - \alpha) \frac{1}{R}} \quad (1)$$

なお、本実験では、特に再現率と適合率のいずれかに重みを置く必要がないので、 $\alpha = 0.5$  として両者を同等に評価した。

実験結果を表1に示す。

表1  $H_{min}$  と再現率、適合率、F値

| $H_{min}$ | 再現率   | 適合率   | F値    |
|-----------|-------|-------|-------|
| 3         | 64.2% | 41.7% | 0.505 |
| 4         | 67.8% | 41.8% | 0.517 |
| 5         | 67.5% | 43.1% | 0.526 |
| 6         | 59.9% | 46.3% | 0.522 |
| 7         | 54.2% | 47.1% | 0.504 |
| 8         | 46.7% | 48.4% | 0.476 |
| 9         | 42.5% | 46.7% | 0.445 |

表1でF値が最大となった  $H_{min} = 5$  の場合において、再現率が67.5%、適合率で43.1%となった。再現率、適合率ともに十分な値であるとは言えないが、これは、原稿と全く一致しない認識結果を評価したためであると考えられる。

そこで、認識結果を、発話と記者原稿との類似度により3つの分類(完全一致、部分一致、一致なし)に分けて、それぞれの検出結果の、再現率、適合率、平均遅れ(認識から誤り検出までの平均の遅れ単語)を評価した。また、提案手法の利点である、平均遅れの短縮効果を確かめるため、DPマッチングによる手法[4]の平均遅れ(=平均単語長/2)も調べた。評価は、表1でF値が最大となる  $H_{min} = 5$  の場合について行なった。評価の結果を表2に示す。

表2 認識誤り検出結果 ( $H_{min} = 5$ )

| 書き起こし | 再現率   | 適合率   | 平均遅れ    | 平均遅れ(DP) |
|-------|-------|-------|---------|----------|
| 完全一致  | 93.6% | 96.7% | 3.11 単語 | 16.39 単語 |
| 部分一致  | 86.7% | 57.5% | 3.59 単語 | 20.33 単語 |
| 一致なし  | 50.3% | 29.5% | 6.45 単語 | 12.93 単語 |
| 合計    | 67.5% | 43.1% | 4.10 単語 | 17.05 単語 |

「完全一致」の認識結果については、再現率(93.6%)、適合率(96.7%)ともに十分な結果が得られた。「部分一致」の認識結果については、適合率(57.5%)がやや低いものの、再現率(86.7%)はかなり高い。認識から誤り検出までの平均の遅れは、文単位で処理した場合(「平均遅れ(DP)」)に比べ、大幅に短縮された。

とくに、「完全一致」の認識結果については、平均3.11単語の短い遅れ時間で、認識誤りの93.6%を検出しており、良好な検出性能が得られたといえる。表2の「完全一致」および「部分一致」の検出結果をさらに詳しく調べるため、検出結果を次の3つに分類した。

- 成功(認識誤りを検出した箇所)
- 過検出(正解単語を誤って検出した箇所)
- 検出漏れ(認識誤りを検出できなかった箇所)

分類結果を、具体例とともに示す。なお、下線部分は検出箇所を、斜線は単語境界をそれぞれ表す。また、「原稿」とは、本手法が「正解文」として検出した記者原稿の文である。

### 成功(133箇所)

(完全一致:29箇所, 部分的な一致:104箇所)

- ・ 認識誤りを正しく検出した例(1);

(原稿) 六栄丸/は/あす/の/出港/に/備えて/  
岸壁/に/停泊/して/います  
(発話) 六栄丸/は/あす/の/出港/に/備えて/  
岸壁/に/停泊/して/います  
(認識) 六栄丸/は/つか/ず/融資/法/に/備えて/  
岸壁/に/停泊/して/います

- ・ 認識誤りを正しく検出した例(2);

(原稿) 一方/ナショナル/リーグ/で/は/パドレス/  
が/二/対/一/で/アストロズ/から/  
一/勝/を/あげ/ました  
(発話) 一方/ナショナル/リーグ/で/は/パドレス/  
が/アストロズ/から/一/勝/を/あげ/ました  
(認識) 一方/ナショナル/リーグ/で/は/パドレス/  
が/足/とも/つか/ないし/を/あげ/ました

- ・ 認識誤りを正しく検出した例(3);

(原稿) .. /続いて/いて/午後/ .. /など/西日本/  
と/東北/地方/の/ところどころ/で/ ..  
(発話) .. /続いて/いて/西日本/  
と/東北/地方/の/ところどころ/で/ ..  
(認識) .. /続いて/いて/西日本/  
や/東北/地方/の/フェリー/航路/で/ ..

### 過検出(78箇所)

(完全一致:1箇所, 部分的な一致:77箇所)

- ・ 正しい認識単語(貼り出さ)を誤って検出した例;

(原稿) .. /を/知らせる/紙/が/はり出さ/れた/ ..  
(発話) .. /を/知らせる/紙/が/貼り出さ/れた/ ..  
(認識) .. /を/知らせる/紙/が/貼り出さ/れた/ ..

### 検出漏れ(18箇所)

(完全一致:1箇所, 部分的な一致:17箇所)

- ・ 正解文を(原稿2)としたため、認識誤り(「が」)を検出できなかった例;

(原稿1) 元/役員/ら/は/東洋/ ..  
(原稿2) .. /元/役員/ら/四/人/が/東洋/ ..  
(発話) 元/役員/ら/は/東洋/ ..  
(認識) 元/役員/ら/が/東洋/ ..

成功例(1)は、記者原稿と発話と完全一致する場合であり、認識結果14単語中の4つの誤りを、全て正確に検出できている。成功例(2),(3)は、記者原稿と発話が完全には一致しない場合(部分一致)であるが、認識誤りを正しく検出できている。このように部分一致の場合で成功した検出は、合計104箇所あった。

一方、過検出の例では、正解認識と記者原稿との表記の「ゆれ」(「貼り出さ」と「はり出さ」)を、誤りとして検出していた。

また検出漏れの例では、正解(「は」)に続く単語列と、誤り(「が」)に続く単語列の両方が記者原稿に存在した場合である(原稿1, 原稿2)。このとき、認識結果に一単語長く一致する後者の原稿が「正解文」となるため、「が」を誤りとして検出しなかった。

これら過検出、検出漏れの問題は、本手法の今後の課題である。

## 5 まとめ

直前の記者原稿を利用した認識誤りの自動検法を提案した。実験によって「完全一致」の発話だけではなく、「部分一致」の発話についても良好な結果を得た。また、認識から検出までの平均遅れ時間は、4.1単語であった。これは、DP マッチングを用いた場合の17.1単語に比べてかなり速い。今後は、本手法を既存の修正システム [1] に実装する方法や、実装した場合の、修正システム全体の評価方法について検討していきたい。

## 参考文献

- [1] 後藤 淳, 今井 亨, 清山 信正, 今井 篤, 都木 徹, 安藤 彰男, 磯野 春雄, 「ニュース音声認識結果のリアルタイム修正装置」, 電子情報通信学会, A-15-15, pp.293 (2000).
- [2] 石場 正大, 竹山 哲夫, 青木 恒夫, 兵藤 安昭, 池田 尚志, 「品詞 N-gram 統計情報を用いた日本語文書における誤り検出法について」, 電子情報通信学会技術報告書, SP97-82, pp.43- 48 (1997).
- [3] 今井 亨, 安藤 彰男, 「ニュース番組自動字幕化のための音声認識システム」情報処理学会研究会, HI80-11(SLP-23-11), pp.59-64(1998).
- [4] 尾上 和穂, 今井 亨, 安藤 彰男, 「記者原稿を用いたニュース音声認識結果の修正法」, 日本音響学会 講演論文集, 1-6-6, pp.11- 12, (1998.3).
- [5] 加藤 直人, 浦谷 則好, 江原 暉将, 「記者原稿を利用したニュース音声認識のための言語モデル」言語処理学会第5 会年次大会発表論文集, pp.269-272 (1999).