

EDR コーパスを対象とした音声合成システムの誤読評価

梅村 祥之 清水 司 原田 義久

豊田中央研究所 機械認識研究室

1 はじめに

カーナビゲーションシステムを初めとする種々の情報機器が自動車に搭載され、様々な情報通信サービスが始まりつつある。提供される情報には、交通情報、電子メール、新聞記事等があり、テキスト音声合成技術の重要性が増しているが、現在まだ十分な品質とはいえない。

本稿では、音声品質を低下させている要因のうちの読み誤りを扱う。読み誤りの原因を考えると、形態素の区切り誤り、同形異音語の判定誤り、固有名詞を初めとする辞書の登録不足などがある。このような原因が複合して読み誤りが生じるため、文単位の読みの正解率は 80% 台程度となってしまう。

読み誤りを改善するための技術開発には、大量の文例による読み誤り評価と誤り例のフィードバックが必要と考えられる。しかし、漢字仮名混じり文の読みを評価するために、EDR コーパスがそのまま利用できる状態にはなっていない。すなわち、カナ表記の記載があるものの、連濁する語、例えば「会社」が連濁して「ガイシャ」と読むべき文脈で、カナ表記が、「カイシャ」と表記されている、長音化して「ガッコー」と読むべきところで、EDR コーパスのカナ表記で「ガッコウ」と表記されている、数字に関して、カナ表記がアラビア数字のままになっている、等々の問題がある。

そこで、EDR コーパスのカナ表記に、音声・音韻上の知見に基づく言語処理を行うことにより、音声合成システムが EDR コーパスの文を入力して作り出す発音情報中の読み誤りを評価できるようにする。

2 方法

2.1 処理の全体

EDR コーパスのカナ表記そのままでは、音声合成システムが出力する発音情報と異なるため、次の処理を行う。

- 1) EDR コーパスのカナ表記と音声合成システムの発音の不一致を、一義的に変換して差し支えないものに関して変換する。これに該当するもの

は、格助詞「を」「は」の発音、記号の発音などである

- 2) 一義的に変換できないものに関して、発音の候補をテーブル化し、照合段階でいずれかにマッチすればよいとして照合する。これに該当するものには、長音、連濁、英文字がある
 - 3) 数字と助数詞に関し、判定不能扱いとし、照合段階で、それを避けるよう工夫する
- 以下、各々の詳細を述べる。

2.2 一義的な変換

(1) 格助詞

格助詞「を」のカナ表記は EDR コーパスで「ヲ」となっているのに対し、市販の音声合成システムの発音出力には、「ヲ」のものと「オ」のものが存在する。「を」の発音は一義的に「オ」と考えられるため、あらかじめ、「オ」に変換してから照合する。格助詞「は」に関しては、EDR コーパスで「ハ」となっているのに対し、音声合成システムでは「ワ」となっているため、EDR コーパスのカナ表記を「ワ」に変換してから照合する。格助詞以外で、カナ表記中の「ハ」を「ワ」と発音すべきものとして、「あるいは」「または」など、いくつか存在する。そこで、品詞が助詞、接続詞、副詞、助動詞で、カナ表記に文字「ハ」を含むものを EDR コーパスから洗い出し、「初めて(△ジメテ)」などの例外を人手でチェックし、「ハ」と「ワ」の読み替えテーブルを作成し、あらかじめ、EDR コーパスのカナ表記の「ハ」を「ワ」に変換してから照合する。

(2) 記号

EDR コーパスで品詞が「記号」と定義される形態素がある。例えば、読点「、」の品詞は記号になっており、その「かな表記」は「表記」と同じ「、」になっている。そこで、カナ表記に現れる記号をどう読むべきかを調べ、読み飛ばしてもよいもの（「、」など）と、一義的に読みを定められるもの（「%」など）に分けた。

(3) 「*」と「?」

EDR コーパスのカナ表記が「*」あるいは「?」となっているものがある。その例としては、表記が「数」、品詞が「数字」、カナ表記が「*」のものがあり、例えば「数メートル」といった文脈で出現する。この例の場合には、発音は「スー」とすべきである。まら、「何メートル」の文脈に出現する「何」は、品詞が「数字」でカナ表記が「?」になっているが、この場合、発音は「ナン」とすべきである。以上のように、カナ表記が「*」か「?」のものとして、「数」、「何」、「いく」、「いくつ」、「幾」…

などがあり、それぞれに対し、読みを「スー」、「ナン」、「イク」、「イクツ」、「イク」…に変換してから照合する。

(4) 「ジ/ヂ」、「ズ/ヅ」

「ジ」と「ヂ」、「ズ」と「ヅ」に関しても、カナ表記が異なるものの、発音は同じである。音声合成システムによって、これらの発音記号に違いがあるため、これらを統一してから照合する。

2.3 発音候補テーブル

(1) 長音

例えば「行進（コウシン）」の発音が「コーシン」のように、カナ表記における母音が発音としては長母音になる現象がある[1]。しかし、EDR コーパスのカナ表記は長音化を加味した表記ではない。一方、音声合成システムは、長音化された発音を出力するため、照合にあたっての処理が必要である。ところで、カナ表記のみから長音化を推定することは不可能である。なぜなら、例えば「子牛」と「格子」のカナ表記は「コウシ」と同じであるが、「子牛」は「コウシ」で「格子」は「コーシ」というように、最終的には語で決まるからである。

以上から、本稿では、長音化の可能性のある語を計算機処理で洗い出し、音声合成システムから出力される発音と柔軟に照合する。すなわち、長音化可能部分は音声合成システムの出力が長音化していれば、EDR コーパスのカナ表記を長音化したものと照合し、長音化していなければ、EDR コーパスのカナ表記のまま長音化していない状態で照合する。なお、1形態素の中に複数箇所長音化の可能性のある部分があれば、それぞれ独立に上記の柔軟な照合を行う。例えば、「コウコウ」は「コーコウ」にも「コウコー」にもマッチする。

長音化の具体的な処理は、例えば次のように辞書引きによって行う。

"アア" "カア" "サア" "タア" "ナア" "ハア" "マア" …
それぞれについて、次のように長音化可能と見る。

"アー" "カー" "サー" "ター" "ナー" "ハー" "マー" …

(2) 連濁

連濁は、例えば「会社」が語の複合によって「〇会社」となるとき、発音が「カイシャ」から「ガイシャ」に濁音化する現象である[1, 2]。EDR コーパスでの連濁の扱いを調べると、次の例のようにカナ表記の揺れが見られる。

JC00001419 無線 電信 会社 → カイシャ

JC00000613 レコード 会社 → ガイシャ

従って、連濁する語は、EDR コーパスのカナ表記をそのまま発音として用いてはならないことがわかる。ところが、連濁現象を計算機処理するためのアルゴリズムと辞書は完全な形で公開されているわけではない。

すなわち、文献[2]で、連濁を起こす可能性のある条件は、国語学の知見から、

「力行、サ行、タ行、ハ行」で始まる語頭清音が、語の複合によって濁音化する

であることはわかっている。しかし、この規則に従って連濁可能性のあるもののうち、実際に連濁するものとししないものがあり、語毎の個別の規則に従うが、その詳細は明らかにされていない。

そこで、本稿では、上記の条件をチェックすることにより、連濁可能性のある語を判定し、実際に連濁するかどうかは、音声合成システム側を信じることにする。なお、連濁可能性の判定において、当該語と先行語が複合語になっているかどうかの判定が必要となるが、その判定に、ここでは先行語が名詞あるいは接頭語で、当該語も名詞であるという規則を用いる。

(3) 英文字

EDR コーパスの仕様説明書[3]によれば、

英語は、アルファベット読みするものとそうでないものとで、記述を分けている。

種別	表記	かな表記	説明
英字(1)	CNN	CNN	アルファベット読み
英字(2)	COBOL	コボル	非アルファベット読み
英字(3)	ppm	PPM	小文字→大文字

とある。

本法では、この仕様に従って発音を求める。すなわち、上記「英字(1)」と「英字(3)」は、カナ表記記載のアルファベットをアルファベット読みして、発音を求める。よって、「CNN」の発音は「シーエヌエヌ」となる。なお、発音の多様性として例えば「H」は「エイチ」と「エッチ」の読み方が可能といった点を考慮して、柔軟に照合を行う。上記「英字(2)」に関してはカナ表記の記載を発音として採用した。

なお、これら英文字全体については、長音化と連濁で特別扱う。

2.4 判定不能扱い

EDR コーパスの仕様説明書 [3] によれば、数字に関するカナ表記は、次のように記載されている。

種別	表記	かな表記	説明
数字(1)	3 0 0	3 0 0	対応する算用数字
数字(2)	3 0 0 万	3 0 0 0 0 0 0	対応する算用数字

これは、読みが付与されていないに等しい。そこで、数字を判定不能形態素という扱いとし、音声合成システム側の発音を信用することにする。また、数字に付随して、助数詞において、数字によって発音の変化するものに関して、発音の観点から問題のあるものが見られる。そのため、これらを洗い出して数字同様に判定不能形態素の扱いとした。次のような助数詞が該当する。

(羽 (バ バワ)) (巻 (カン マキ))
 (個 (カ コ)) (勺 (シャク セキ))
 (人 (タリ トリ ニン)) (節 (セツ フシ)) …

2.5 照合

判定不能扱いの部分が存在する可能性があるため、図1のように照合する。すなわち、初め、文頭

から形態素毎に照合して行き、判定不能形態素が現れたら、次に文末から文頭へ向かって形態素毎に照合して行く。この時点で再び判定不能形態素が現れたら、この文の処理を終了する。そのとき、前後の判定不能形態素に挟まれた判定可能形態素が存在すれば、準正解とする。前後の判定不能形態素に挟まれた判定可能形態素が存在しなければ、完全正解とする。また、以上の照合の最中にミスマッチした形態素が始めて現れた時点で、その文の処理を終了し、不正解文としてカウントする。

3 評価例

本評価法の適用例として、市販の音声合成システムを対象にした読み誤り評価結果を示す。音声合成システムから、発音と韻律の情報を持つ中間コードが出力されるので、その中から韻律情報を除いて発音の情報を得て用いる。

以下で用いる用語「完全正解数」「準正解数」「不正解数」は、前述の「照合」の節で述べた量である。

評価した文の総数 = 192,231 文
 完全正解数 = 140,097 文
 準正解数 = 12,831 文
 判定不能形態素に挟まれた判定可能形態素数 = 8.6
 不正解数 = 39,303 文
 1文当りの平均形態素数 = 24.4
 文当り正解率(準正解も正解に含める) = 79.6%

表1に、本評価システムから得られた読み誤り例のうち、上位15の高頻度形態素を表示する。このうち頻度最大の助詞の「は」は、次の例のような、形態素の区切り誤りに起因すると思われる誤りである。

… 大統領選挙はいよいよ本番 …
 … ダイトーリョーセンキョハイヨイヨホンバン …

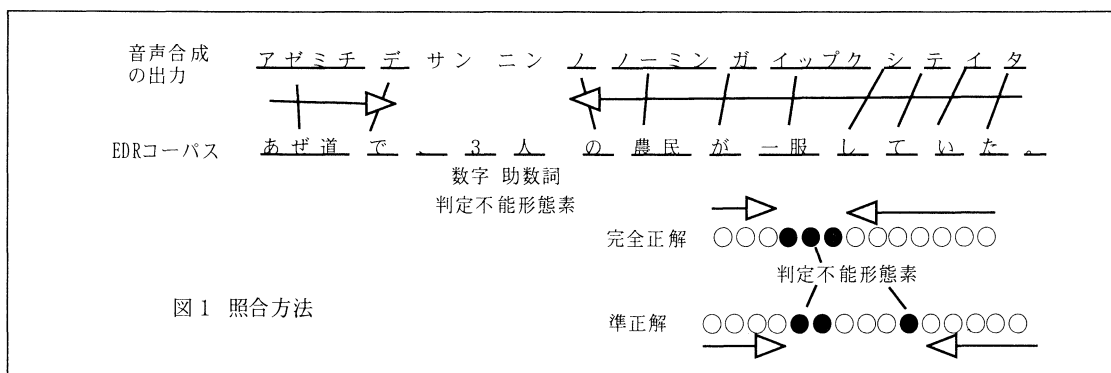


図1 照合方法

最も多く出現する誤りの種類は、「行った(イッタ／オコナッタ)」「今日(キョウ／コンニチ)」など同形異音語の読み分け[4]に関するものである。なお、誤りとして出力されたものの中に「私(ワタクシ／ワタシ)」など、本来発音に任意性があると考えられるものが含まれている。これは、本評価法でEDRのカナ表記に従って判定して、一方を正解、他方を不正解とした結果であり、本来不正解とすべきではない。

高頻度で誤る形態素を改善すれば、正解率向上への寄与は大きい。そこで、頻度順に累積して、異なり形態素数とカバー率の関係を算出し表2に示す。

表1 読み誤り形態素の例

<表記>	<品詞>	<カナ表記>	<出現頻度>
は	助詞	ワ	2013
今年	名詞	コトシ	1384
私	名詞	ワタクシ	625
行	動詞	イ	566
ソフトウェア	名詞	ソフトウェア	487
その後	名詞	ソノゴ	485
他	名詞	タ	337
他の	連体詞	タノ	324
金	名詞	カネ	317
後	名詞	アト	316
言	動詞	イ	288
分	接尾語	ブン	279
後	名詞	ゴ	205
今日	名詞	キョウ	193
末	名詞	スエ	189

表2 異なり形態素数と累積形態素数

異なり数	累積数	カバー率
1	2013	0.051
2	3397	0.086
4	4588	0.117
8	6221	0.158
16	8184	0.208
32	10146	0.258
64	12511	0.318
128	15139	0.385
256	18052	0.459
512	21121	0.537
1024	24383	0.620
2048	27930	0.711
4096	31554	0.803
8192	35650	0.907
最後まで	39303	1

4 まとめ

テキスト音声合成システムの言語処理部が出力する漢字仮名混じり文の読みに関し、EDRコーパスの文例から読み誤り率を算出し、誤り文例データを出力する読み誤り評価法を開発した。EDRコーパスに含まれるカナ表記は、音声合成システムの発音と必ずしも対応しないため、これを考慮した照合を行う方法を用いた。また、本法のないため、カナ表記と発音との関係有効性を確認するために、市販の音声合成システムを対象にして読み誤りの評価を行い、誤り率を得、誤り傾向を分析した。

本評価法は、音声合成システム相互の比較のための指標を提供し、出力される誤り文例集には、性能向上に寄与する指針が多く含まれる。

出力結果を検査して本評価法自体の精度を調べると、本評価システムが誤りと判断したものの中に、EDRコーパスおよび本評価法側の問題によって誤りと判定されているものが約30%程含まれていた。この点が今後の評価法改良の課題である。

謝辞

本研究を行うにあたり、貴重なご助言を頂いたトヨタ自動車(株) 第3電子技術部 青島滋樹氏に感謝いたします。日頃より御指導いただく(株)豊田中央研究所 杉本軍司部長に感謝いたします。

参考文献

- [1] NHK 編：日本語発音アクセント辞典，日本放送出版協会（1985）
- [2] 杉藤美代子編：講座日本語と日本語教育 第2巻 日本語の音声・音韻（上），明治書院（1989）
- [3] 日本電子化辞書研究所：EDR 電子化辞書仕様説明書（1995）
- [4] 梅村祥之，清水司：決定リストによる同形異音後の読み分け，言語処理学会第4回年次大会（1989）