

日本語文章読み上げ技術

5J-2

- AUDIOTEX -

松岡浩司 武石英二 浅野久子 市井亮美 大山芳史

NTT情報通信研究所

1. はじめに

新聞記事などのテキストを分かりやすい表現に書き換えて音声で読み上げるテキスト音声合成システムAUDIOTEXを開発した。従来の日本語文章を音声に変換するテキスト音声変換システムの問題として、①読み上げ対象の文が目読を前提に記述されているため、そのまま音声化したときに聞き取りにくい点や、②読み誤りを抑止できない点が挙げられる。このため、従来、校正支援などの限られた分野での応用に留まっていたシステムを一般ユーザを対象にした情報案内サービスに発展させることを狙いとして、テキスト解析処理の改善を行った。

2. システム概要

本システムの構成を図1に示す。①文を単語に分割する形態素解析処理、②文章を音声化に適した表現に修正する書き換え処理、③読み、アクセント、ポーズからなる仮名アクセント文を生成する韻律処理、④仮名アクセント文から音波形を生成する音声合成処理^[1]からなる。

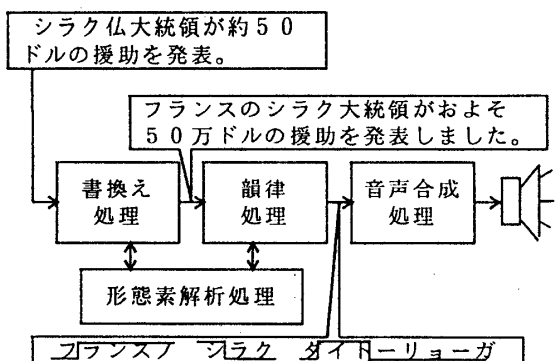


図1 システム構成

AUDIOTEX: A Text-To-Speech System for Japanese Text
 Koji MATSUOKA, Eiji TAKEISHI, Hisako ASANO,
 Ryomi ICHII, Yoshifumi OYAMA
 NTT Information and Communication Systems Laboratories

3. システムの特徴

(1)形態素解析処理

従来、利用していた局所総当たり法^[2]は、漢字列の末尾の単語を接続開始単語として、その前方に接続する漢字単語列と、その後方に接続するひらがな単語列で総当たりに単語を生成した後でそれぞれ絞り込みを行ない、両列を接続して1つの単語列を生成していた。本システムは、両列の単語列を絞り込まずに全ての組み合わせを生成した。ただし、同形語（見出しが同一である単語）を1つのグループにまとめて、そのうち代表的な単語で処理することにより、処理量を抑えた。解析結果の例を図2に示す。

(例) 今年はやる。

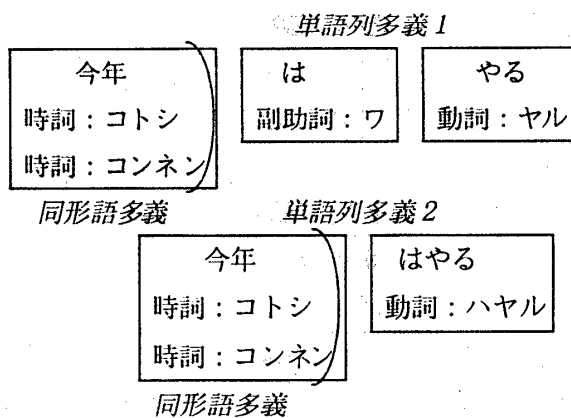


図2 形態素解析多義

生成された多義を用いて、固有名詞などの読みやアクセントの誤りを補正する手段として、①入力文中に括弧で読みを挿入して多義を指定できる機能と、②ユーザがGUIを使用して多義をメニューで選択する機能を設けた。

①の例（漢字の表記の自動説明で利用^[3]）

洋(ヒロシ)さんは太平洋の洋(ヨウ)。

補正前は前後の洋をヒロシと読む。

この他の多義生成の効果として、各単語列に対

して適用された絞り込みルールの一覧と尤度とをトレースとして出力することにより、解析結果のチューニングが容易になった。

(2) 書換え処理

書換え処理では音声化に適した表現への自動変換を行う。新聞記事の文章をターゲットとして実現した主な書換え機能を表1に示す。このとき、記号類や一部の略号など文字列情報で書き換えられる表現は形態素解析の前に書き換えて形態素解析の処理量を抑止するために、表層・形態素の2段階の処理を設けた。

表1 新聞記事文章の書換え機能

書き換え要因	分類	書き換え例	利用した情報
誤って聞こえる	接辞	約五十→およそ五十 今秋 → 今年の秋 (今週)	表層 形態素
	略語	衆院 → 衆議院	表層 形態素
聞き取りにくい	漢字 国名	対米貿易 → アメリカとの貿易	形態素
	難語	詳細は → 詳しいことは	形態素
記述文特有	節末 表現	計画によると → 計画によりますと	形態素
	体言 止め	～の予定。 → ～の予定です。	形態素
音声にできない	記号類	東京・首相官邸 → 東京の首相官邸	表層 形態素

なお、書き換え規則は利用分野に依存するので、柔軟な処理系が必要である。変形の対象とする木構造と変形結果の木構造を指定することにより、木構造変形を行うCプログラムを生成するツールを開発し、規則の記述に用いている。(図3)

(例) (defrule 人名-国名
 (入力パターン (文節(人名 接辞国名 役職名))
 (出力パターン (文節 (カナ国名 "の" 人名 役職名)))

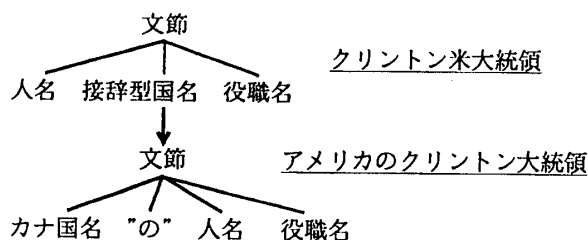


図3 木構造変形の例

(3) 韻律処理

複合語の係り受け解析による構成単語の結合関係を利用してアクセント句境界やポーズの位置を決定した。^[4]この他、形態素解析処理において、①辞書上に長単位語を登録して補助用言や格助詞相当語の認定精度を向上させる、②サ変名詞+「する」、形容動詞型名詞+「だ」を長単位語として認定してそれぞれ動詞と形容動詞と同等に扱うことで韻律精度を向上した。

4. 評価

①書換え処理：新聞記事302文に対して、人手により書換えられる表現のうち、51%をカバーでき、処理対象の95%を正しく書き換えた。

②韻律処理：新聞記事など149文に対して、自動(多義による補正なし)で、読み誤り率0.3%、アクセント誤り率4.1%、ポーズ誤り率4.6%であった。*なお、以上の処理はリアルタイムでの実行が可能である。

*全てのアクセント句について、それぞれ、読み、アクセント、ポーズの誤りを含むアクセント句の割合を示す。

5. おわりに

上記のように良好な書換え精度と韻律精度を得ることができた。将来は、口語文を含む、広範な分野の文章に対して、高品質な音声を提供できるようにしたい。なお、音声合成処理はNTTヒューマンインタフェース研究所音声情報研究部のテキスト音声合成システムFLUETを使用した。

参考文献

[1] Kazuo HAKODA, Tomohisa Hirokawa, Hajime TUKADA, Yuki YOSHIDA, and Hideyuki MIZUNO, "Japanese Text-To-Speech Software based on Wave Form Concatenation Method" AVIOS' 95
 [2] 宮崎, 大山 "日本文音声出力のための言語処理方式", 情処論文誌, Vol. 27, No. 11, pp. 1053-1061
 [3] 大山, 浅野, 松岡 "音声出力による姓名漢字表記の説明方式", 第52回情処全大4D-6, 1996
 [4] 浅野, 松岡, 市井, 大山 "テキスト音声変換における読み韻律付与の評価" 第51回情処全大3R-3, 1995