

日本語テキスト音声合成のための言語処理の検討

鈴木 和洋, 鳥原 信一, 齊藤 隆
日本アイ・ビー・エム(株)東京基礎研究所

あらまし: 日本語テキスト音声合成において, 正確な読みやアクセントを実現することは, 実際に音声合成を利用する上で重要である。本報告では, 未知語に関する処理を中心に, 正確な読みやアクセントを導出するための言語処理について概要を述べる。未知語処理は, 未知語と思われる単語の字種に応じて, ひらがな単語, 表記揺れ単語などに, 読みやアクセントの推定を行っている。本報告では, さらに未知漢字列の読みを推定する方法として, 読み付き文書データから, 漢字trigramのデータの収集と類似読み漢字のグループ化の検討した。最後に, 本言語処理の解析精度の評価を行ない, 読み正読率96.5%, アクセント正読率93.0%を得た。

A text analyzer with unknown word handlings for Japanese text-to-speech

Kazuhiro Suzuki, Shin'ichi Torihara, Takashi Saito
Tokyo Research Laboratory, IBM Japan Ltd. oratory, IBM Japne Ltd.

abstract: This paper describes a text analysis method for a Japanese text-to-speech system, focusing on unknown word handling. We classified unknown words into four categories, according to the kinds of characters and error patterns of text analysis. The unknown word handler uses four processes for the unknown word categories to estimate the representation and accentuation of each unknown word. We describe an approach for estimating the representations of unknown word consisting of Kanji characters by making single-Kanji trigram data and grouping Kanji characters according to their similarity. Finally, we report the result of test on the text analyzer and unknown word handler, and give the correction ratios for representations and accents.

1. はじめに

近年, パーソナル・コンピューターの普及に伴い, テキスト音声合成の需要が高くなってきている。テキスト音声合成では, 任意の文章を読むという観点から, その合成音声の品質の良し悪しとともに, 文章をどれだけ正確に読めるかが重要である。

一般に音声合成は, 言語処理部, 韻律制御部, 音声合成部の3つの処理から構成される。言語処理部では, 単語辞書をもとに単語の接続の規則に基づいて, 構成単語を同定し, 単語または文節の並びに対して, その依存関係の解析がなされる。次に求められた個々の単語の読み, 基本アクセント, 文法情報や, 単語・文節間の局所的な依存構造から, 韻律規則が適用され, 各音韻の継続時間長, アクセント句の位置・大きさ, 発話境界位置, および, ポーズの位置や長さなどの韻律情報が決定される。最終段の音声合成部では, 音韻情報と韻律情報をもとに, 合

成単位を選択し, つなぎ合わせることによって合成音声を得られる。

文章を正確に読む上で, 重要となるのは, 上記で述べた3つの処理のうちの言語処理部にあたる。この処理の中で, 正確な読みやアクセントを得るために, 最も重要であることは, 単語辞書の充実や形態素解析の精度である。

しかし, 任意の文章を対象としていることから, 固有名詞といった辞書に存在しない単語の出現などによって, その語が正確に読めないなどの問題が起こり, 全体の精度に大きく影響を与えることも多い。

我々は, これまで局所的な文構造をもとにした発話境界の解析^[1]や, 環境依存性を考慮した音声合成単位の生成^{[2],[3]}などについて報告を行っているが, 今回は, 正しい読みやアクセントを導出するための言語処理について, 特に未知語の処理を中心に述べる。我々は, 未知語処理において, テキストの解析過程で, まず, 未知語

区間を判別し、字種などを考慮して分類を行い、処理を分け、読みやアクセントを推定する方法を採用した。特に新たな試みとして、大量の読み付きテキストの情報から、漢字列の読みを推定する試みとして、漢字trigramの収集と類似読み漢字のグループ化の基本的な検討を行った。以下、言語処理部の概要、未知語の処理、そして、未知語処理を含めた、言語処理全体の評価について報告する。

2. テキスト音声合成言語処理の概要

本言語処理の概要を図1に示す。まず、入力されたテキストを、単語辞書を用いて単語に分解し、文節候補を構成する。求められた文節候補は、二文節最長一致を基本にして、もっとも評価値の高い候補を文節として確定する。確定した文節列に対して、局所的な文構造から、発話境界を求め、各種アクセント結合、話調の立て直し位置、ポーズ位置を決定する。ここで、図1の点線で囲まれた部分は未知語の処理を示す。通常の形態素解析の枠組みでは、未知語の出現によって、部分的に異なる単語や仮名、単漢字で文節が構成されたり、また文節境界が誤ったりする場合がある。未知語処理は、そうした形態素の並びから、未知語区間を決定し、区間内の字種やその組み合わせに応じて種々のルーチンを起動し、読みやアクセントを推定して、単語や文節候補を文節確定処理に渡す。文節確定では、それらの単語候補や文節、候補を取り入れて、処理を行うことにより、より正確な読み、アクセントを提供できる。ここでは、まず、形態素解、文節の確定、係り受け解析などの、基本となる言語処理の枠組みについて簡単に説明する。

2. 1. 文法体系

形態素解析の基本となる文法体系として、形態素候補間の接続条件を細分化し、バックトラッキングなしに効率よく解析する「接続ベクトル法」^[4]を採用した。これは仮名漢字変換用に関与された解析法であるが、音声合成のための解析用として、付属語辞書をアクセントに応じてきめ細かく拡充するなどの改良を行った。

2. 2. 形態素解析と文節の確定

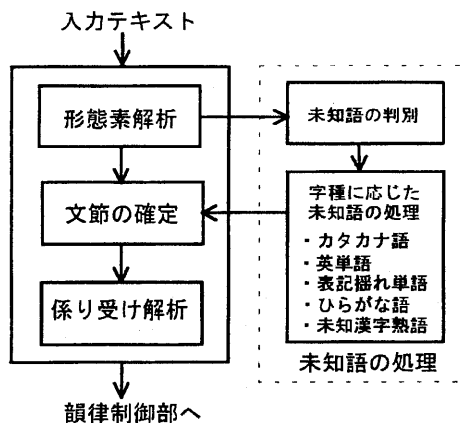


図1. 言語処理部の処理の流れ

音声合成のための形態素解析として、文節最長一致法に基づいた方法^[5]や隣接単語間の結合関係に着目した方法^[6]などが提案されている

我々は、二文節最長一致法に基づいた文節確定法を導入した。この方法では、文節を確定するために、候補文節と後続する文節候補との長さが最も長い候補を選び出す。その候補の中から、経験的に決めた以下の基準に当てはまるものを選んでいく。

- ・単語接続ペナルティーが最小のもの
- ・付属語数が最小のもの
- ・二文節の切れ目が中央に近い

ここで接続ペナルティーは、単語間の接続に伴うペナルティーで、複合語や付属語の接続の強さ、単語の頻度情報、文節前後の係り受け可能性および次節に述べる単語の付加情報を評価するものである。

2. 3. 単語辞書

本言語処理では、約77,000語の単語辞書を用いている。この辞書には、単語の見出し、読み、文法情報、アクセント、および付加情報が収められている。付加情報は、文法情報以外の、主として、読みやアクセントおよび単語間のアクセント結合に関する単語固有の性質を表現したものである。

2. 4. 係り受け解析と韻律の設定

係り受け解析は、合成音声の文節間アクセント結合、話調の切れ目、およびポーズ位置などの韻律の設定を決める上で重要である。本言語処理では、求められた文節列の品詞系列など情

報に基づいて、3文節間の局所的な係り受け解析を行っている。この係り受けのパターンと発話の制約から、上記の韻律設定を行っている^[1]。

3. 未知語処理

形態素解析での未知語処理は、音声合成をはじめとして、多くの自然言語に共通するものである。音声合成での未知語処理の目的は、解析された形態素列から未知語を抽出し、それに対して、正確な読みやアクセントをどうやって与えるかにある。

通常の形態素解析において、入力テキストの中に未知語が出現すると、それは様々な読み誤りやアクセント誤りとして、合成音声に出力される。この誤りを、どのように修正するかは、未知語が、どのような形態素列として解析されたかを観察し、どの部分の形態素列を未知語区間と判別するかを考える必要がある。そのために、まず、種々の未知語の特徴を字種の観点から分類した。

3. 1. 未知語の分類

未知語を字種の観点から分類すると次のような4つのカテゴリーに分けることができる。

- カタカナ語、英単語など未知語境界が明確な未知語。
- ひらがな語のように付属語を伴った場合に未知語境界が字種だけで判断できない未知語。
- 表記揺れなどによって漢字とひらがなの混合した未知語。
- 省略語、造語、固有名詞などのような漢字列となる未知語。

これら4種の未知語は、解析される形態素やその並びが異なるため、読みとアクセントを推定する方法もそれぞれ異なってくる。そこで、未知語処理では、解析された形態素の並び、字種に応じて、起動の仕方を変え、それぞれに対応した処理を行うことにした。

3. 2. 未知語処理の起動と未知文字列の判別

未知語処理をすべての入力文字列に適用すると、本来の形態素解析の妨げとなる恐れがある。そのため、単語解析では、未知語処理の起動のタイミングを決定するために、解析された単語候補の尤もらしさを表す単語候補信頼度を、単語候補の使用頻度、字種区分、品詞系列、単語

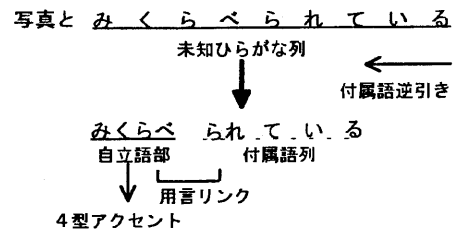


図2 未知ひらがな列の処理

長などをもとに算出する。この単語候補信頼度がある閾値を下回る場合に、その区間を未知語区間と定めて、未知語処理を起動する。現段階では、信頼度を算出する要因として、文字列内の、単一の仮名文字や単漢字、および、解析された単語の字種を考慮し、未知語区間の判別を行う。以下、その4つの処理について説明する。

(ただし、漢字列となる未知語についての処理は、現段階では検討中のため、ここでは、処理の検証について報告する。)

3. 3. 未知のカタカナ語、英単語の処理

カタカナ語は、通常外来語を表すものが多く、そのままの読みに逆3モーラルールでアクセントを与える方法や音韻の特徴的な並びに応じて外来語アクセントを推定する方法^[7]も提案されている。しかしながら、日本語においては、動植物の名前や語の強調のためにカタカナを用いることがある。また、「カッコいい」など、語の一部分がカタカナで書かれているものも存在する。こうした外来語のアクセントルールで扱えないアクセントを持つ語については、何らかの別の処理が必要である。

一方、英単語については、まず、基本的な単語については、英単語辞書をもって扱うが、辞書に登録されていないものについては、英単語の読みがルールで生成できることを利用して、読みとアクセントの推定を行っている。

3. 4. 未知のひらがな語の処理

ひらがな語の未知語は、後続する付属語と連なることが多く、そういった場合に自立語の部分特定することが難しくなる。こうしたひらがな語の未知語区間判定では、解析できなかった単一のひらがなを中心にして、周囲のひらがな単語などを含めて、未知語区間としている。

未知語区間が決まった後の処理例を図2に示

す。この場合、未知語区間を、1つの文節と考え、付属語を後方から解析、判定して、もっとも長い付属語列を採用し、それより前を自立語とする。この処理によって、付属語列については、正確なアクセントを与えることができる。自立語部のアクセントについては、

付属語の接続条件で、品詞が判別できるため、体言についてはO型アクセントを、用言についてはN型アクセントを指定しているが、この単純な規則では、必ずしも正確なアクセントは得られないため、カタカナ語と同様に何らかの別の処理が必要となる。

3. 5. 表記揺れ単語の処理^[8]

これまでに述べてきたカタカナ語やひらがな語の未知語、および仮名と漢字が入り交じった未知語などの中で、「辞書には存在するが表記揺れのために情報が得られない」場合がある。このような単語を表記揺れ単語と呼び、それらに対処するための処理を導入した。

表記揺れ単語の処理では、基本的に、入力テキストの未知語区間を、単漢字辞書を用いて仮名列に変換し、その仮名列から単語辞書を引くという処理を行っている。起動のタイミングとなる未知語区間の判定においては、複数の漢字を含む平仮名列や片仮名列が考えられるが、経験的に解析誤りの起こりやすいと考えられる「一文字以下の漢字を含む片仮名列または平仮名列」としている。

図3に処理の流れと、簡単な例を示す。この例では、「後押し」という単語が辞書に登録してある場合に、「あと押し」という単語に、読みとアクセントを与えるものである。未知文字列と判定された「あと押し」に対し、単漢字辞書を用いて、いくつかの仮名表記を生成する。これら仮名表記に対応する語が基本辞書に存在するかどうかを、仮名索引を用いて調べ、存在する場合には、表記揺れ単語候補として、文節確定時の形態素候補として追加する。この処理は、辞書に登録されている単語に対して、様々な揺らぎを吸収できるため、正読率の向上と

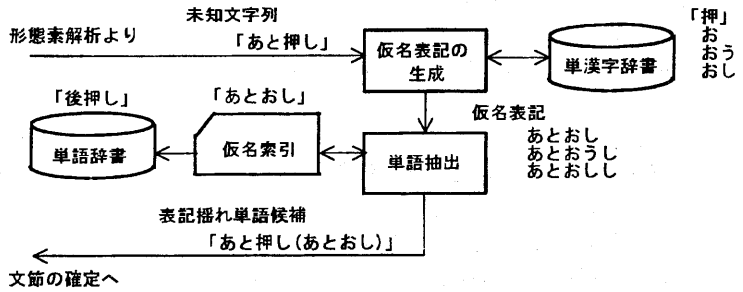
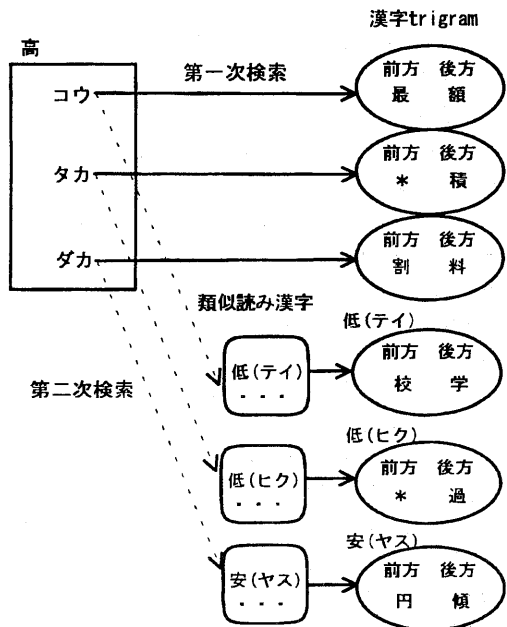


図3 表記揺れ単語の処理

もに、辞書の登録単語数を必要最小限に抑えるという利点もある。

3. 6. 未知漢字列の処理

現時点での漢字列の未知語については、漢字の音訓を考慮しながら、読みをつなぎ合わせて行く方法を用いている。そのため固有名詞や造語など正しく読めないものも多い。漢字列に対して読みを推定する方法の先行研究として漢字の比較的簡単な分類とルールを用いて、通常の日本語文に含まれる漢字を読むことが報告されている^[9]。我々は、「人が未知漢字列を読む場合、漢字の読みは、前後の漢字環境から予測している場合が多い」という仮定から、大量の読み付きテキストを利用して、読みを推定するた



めのデータベースを作成し、当該漢字の読みをその周囲の環境要因によって決定することを試みた。とりあえず、今回は、前後の漢字環境として、両端2漢字環境における単漢字の読みの情報（以下、「漢字trigram」と呼ぶ）を求め、その統計情報を用いての読みの推定を試みた。さらに、限られたデータを有効に利用するために、同環境においてある特定の読み方をする漢字（以下、「類似読み漢字」と呼ぶ）を求め、そのグループ化の検討も行った。現段階では、読み付きテキストの不足で、十分な統計情報を得られないため、今回は、基本的な有効性の検討となった。なお、アクセントの推定についても、今回は扱わず、今後の課題とした。

漢字未知語列の処理を図4に示す。未知語区間の判定では、まず構成される形態素列に音読みの単漢字が含まれる場合で、その漢字列のすべての構成漢字が、単一の読みしか持たない場合には、処理を行わない。未知語と判断された漢字列については、まず、第一次検索として、漢字列の中の複数の読みを持つ漢字について、漢字trigramのデータを用いて読み付けを行う。このとき、その漢字における前後環境が漢字trigramのデータになれば、さらに、第二次検索として、その漢字の各読みに対応する複数の類似読み漢字のグループの漢字trigramのデータを調べる。データの中に必要とされる環境が見つければ、その類似読み漢字に対応する読みを採用する。例えば、図4において、「円高傾向」の「高」の読みを推定する場合、まず、「高(コウ)」、「高(タカ)」、「高(ダカ)」の3つの漢字trigramを検索する。この3つの中に、「円-傾」という環境が見つからなければ、次の検索として、類似読み漢字のデータを検索する。このとき「高(ダカ)」の類似読み漢字「安(ヤス)」の漢字trigramに「円-傾」という環境があるため、「高」の読みは、「ダカ」と推定される。

3. 6. 1. 漢字trigramデータの収集

漢字trigramのデータを収集するために、読み付きテキストとして、EDR（電子化辞書研究所）のデータ^[10]を用いた。このデータは、主として新聞記事で構成され、その中から、207,802文のデータを使用した。これらの文に対して、漢字コード表から作成した単漢字辞書を用いて、

見出し漢字	前接漢字	後接漢字	見出し漢字の読み	頻出回数
安	で	値	アン	4
安	円	傾	ヤス	1

図5 漢字 trigram データの例

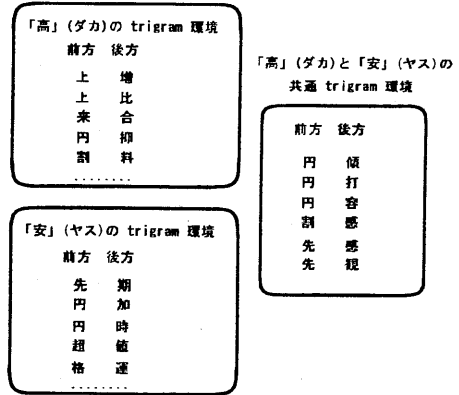


図6 類似読み漢字の例(高[ダカ]安[ヤス])

各単漢字に対応する読みを設定し、読みが単一のものは削除した。図5に、漢字trigramの例を示す。この処理によって得られた漢字trigramの数は、506,086個であった。このデータベースを用いて、文章の読み付けを行ったところ、既知語、未知語ともに読み付けできないものがあり、データ数が、まだ十分でないことがわかる。

3. 6. 2. 類似読み単漢字のグループ化

仮に、さらに大量の読み付きテキストを用いてデータベースを作成しても、すべての漢字trigramのデータを持つことは、現実には困難である。そのため、限られたデータベースを有効に用いるためには、類似読み漢字のグループ化が必要になる。今回は、そうしたグループ化が可能であるかどうかを検証するため、教育漢字約1,000個に関して、類似読み漢字を調べた。

図6に、類似読み漢字の例を示す。この例では、「高(だか)」「安(やす)」と呼ばれる2つの組みに関してみると、共通のtrigram環境が存在する。この場合に、双方のtrigram環境を入れ替えを行うと、多くの場合、入れ替えが可能である。この場合、「高(だか)」と「安(やす)」の2つの組は類似読み漢字となっており、これらの環境をグループ化して用いることが可能といえる。ただし、幾つかの例では、連濁の有無の問題が絡んでくるもの（「超高(タカ)値」と「超安(ヤス)値」）などがあり、さらに検討が

必要となる。

4. 評価

最後に、言語処理の結果が、どのように合成音声の読みやアクセントの正確さに反映されているかを調べるために評価実験を行った。ここでは、比較的大量のデータを用いての評価に焦点をおき、9日分の新聞記事について、ひらがなと漢字の境界で区切った文節データ（疑似文節データ）を用いた。データの総数は、179,079文節で、読みの評価については、全データを使用し、アクセントについては、全データから任意抽出した3156文節について、聴取による評価を行った。

また、文章の評価についても、新聞のコラム、335文（1717文節）を用いて、同様に読みとアクセントの評価を行った。結果を表1に示す。読みの誤りは、3～4%、アクセントの誤りは、4～7%であった。誤りのデータをみると読みにおいては、固有名詞の読み誤りが多くみられ、アクセントにおいては、おもに長い複合語のアクセント結合の誤りが目立っていた。今後のさらに精度を上げていくためには、固有名詞の読み、アクセントの推定や複合語の分割などの検討が必要であろう。

未知語処理（未知漢字列の処理を除く。）の効果を調べるために、未知語処理の有無に関する正読率の違いを調べた。これについても、上記の実験で用いた疑似文節および文章データを使用した。表2にその結果を示す。未知語処理において、疑似文節では、誤り全体の11.4%（1047文節）、また、文章においては、30.5%（52文節）について、読み、アクセントの改善がみられた。さらに、多くの改善を得るためには、今後、漢字未知語列の処理の導入や未知語の判別の基準の検討などが必要であろう。

5. おわりに

日本語テキスト音声合成のための言語処理について、未知語処理を中心に、処理の概要を述べ、その評価を行った。今後の課題として、
・漢字未知語に対する処理として、漢字trigramのデータ収集と類似読み単漢字のグループ化の

表1. 読みおよびアクセントの評価

	疑似文節 (179079文節)	疑似文節 (3156文節)	文 (335文, 1717文節)
読み誤り(個)	7610 (4.25%)	111 (3.52%)	47 (2.68%)
アクセント誤り(個)	-----	221 (7.00%)	71 (4.14%)

表2. 未知語処理の効果

	使用テキスト	未知語処理なし	未知語処理あり
誤り数 (個)	疑似文節 (3156文節)	375	332
	文 (335文, 1717文節)	170	118

検討、およびその導入。

・漢字未知語に対するアクセント付与ルールの検討。
などが挙げられる。

【文献】

- 鈴木, 齊藤: 日本語テキスト音声合成のためのN文節構造解析とそれに基づく韻律制御, 電子情報通信学会誌, Vol. J78-D-11, pp. 177-187 (1995).
- 橋本, 齊藤: 環境依存性を考慮した音節を合成単位とする音声合成—環境依存クラスタリングによる音素クラスターの生成—, 音響学会講演論文集, 2-1-1, pp. 245-246 (1995. 9).
- 齊藤, 橋本, 阪本: 環境依存性を考慮した音節を合成単位とする音声合成—音素クラスターを用いた音節合成単位の生成—, 音響学会講演論文集, 2-1-2, pp. 247-248 (1995. 9).
- 大河内正明: 分かち書き方式仮名漢字変換のためのバックトラックを必要としない文法解析, 情報処理学会誌, Vol. 24, No. 4, pp. 389-396 (1983).
- 小暮, 嵯峨山, 佐藤: 音声合成のための日本語テキスト解析, 音響学会春季講演論文集, 1-5-18, pp. 79-80 (1982).
- 清水, 樋口, 河井, 山本: 隣接単語間の結合関係に着目したテキスト音声変換形態素解析, 音響学会論文誌, Vol. 51, No. 1, pp. 3-13, (1995).
- 佐藤大和: 外来語アクセントの分析と規則化, 音響学会講演論文集, 1-7-7, pp. 133-134 (1989. 3).
- 鈴木, 齊藤: 日本語テキスト音韻変換における表記揺れの解消, 音響学会講演論文集, 3-5-9, pp. 295-296, (1992. 10).
- 鈴木, 中挟知, 近藤, 佐藤, 島田: 漢字シソーラスの構築と語句解析への応用, 第52回情報処理学会全国大会 4B-6 (1996).
- 日本電子化辞書研究所: EDR電子化辞書, 日本語コーパス JCO-V015 (1995).