

動画と音声認識を利用した外国人向け日本語自動詞・他動詞学習コンテンツの開発

早稲田大学日本語研究教育センター スワン彰子

早稲田大学理工学部 奥本 浩章

早稲田大学メディアネットワークセンター 藤田 真一

早稲田大学日本語研究教育センター 川本 喬

早稲田大学理工学部 成田 誠之助

swanakik@aoni.waseda.jp

概要 外国人の日本語学習者にとって、自動詞・他動詞は難易度の高い学習項目の一つである。これを学習する場合教室内で単に文字と言葉のみで説明しても実際にどのような場で使うのかを意識しづらい。発表者らはよく利用される自動詞・他動詞をピックアップし、それぞれについてシチュエーションを考え、動画として撮影して教材とした。また、単にそれを再生するだけではなく、動画内で利用されている動詞の部分を学習者にマイクで入力させ、それを音声認識技術を用いて自動採点する学習コンテンツを試作した。

1. はじめに

本教材は、早稲田大学の留学生を対象とした日本語の授業で使用されている教科書を補助する目的で開発された E-Learning システム[1]の一部である。このシステムは講義中と自宅学習での利用を想定し、教科書の内容に沿って、カタカナ語、動詞の活用、助詞など、学習者にとって習得が困難であると思われる項目について重点的に練習が行える教材である。

今回はその中で「自動詞・他動詞」について扱っている教材を開発したので報告する。日本語の自動詞・他動詞は「始まる」「始める」などといった対になる動詞を見てみると、両者の形の違いに規則性がない、音響的に似たものが多いため学習者が混乱しやすく習得が難しい文法項目の一つである。また、日常会話を考えてみると、われわれネイティブの日本人は自動詞・他動詞を無意識のうちに使い分けている。日本語学習者にとっては両者の即時的な使い分けが要求されるということであり、繰り返し学習することが必要となる。

以上の背景を踏まえ、自動詞・他動詞の即時的な使い分けを練習するための Web 教材を作成することにした。そこで、日常会話における即時的かつ感覚的な利用を考慮して、キーボード入力ではなく、音声入力による練習教材を目指した。これまでに開発した音声認識システムおよび教材の仕組みと、外国人日本語学習者による評価結果を報告する。

2. 教材システムの特徴

2.1 全体像と構成

システムは、Web サーバ (Microsoft Windows 2000 の IIS 5.0)、サーバマシン上のデータベース (Microsoft Access 2000)、クライアントマシンから成り、データベースと ASP の連携により、クライアントに動的な練習問題を提示する。問題表示用 HTML ファイルには、Java で開発した音声認識アプレットが埋め込まれているため、ブラウザを起動するだけで、音声認識機能を利用できる。このように、アプリケーションのインストールやダウンロードを行うことなく、すぐに音声入力を行える。

2.2 音声認識アプレット

音声認識に必要な処理は、すべて Java アプレットで行われる。本システムの場合、日本語の自動詞と他動詞という音響的に近い単語同士（例えば、「閉まる」と「閉める」）を正確に認識する必要があるため、一般的な音声認識エンジンのように、文法、文章の構造および単語辞書の検索による概念駆動型処理は行わず、純粹に音声から得られる特徴量のみを用いて認識を試みる。また、認識対象は、語彙数に関してはあらかじめ登録された単語に限られているが、不特定話者に対応している。認識結果としては、正解か不正解かのいずれかが直ちに表示され、確認できる。

2.3 本教材の特徴

日本語学習用教材としての本教材の特徴は音声入力を取り入れることで、学習者が自動詞・他動詞の即時的な使い分けを練習できることである。

また、音声認識を利用した教材という視点から本教材を見てみると、本音声認識システムは、インストール不要、不特定話者対応、即応性という特徴を備えており、自動詞・他動詞の即時的な使い分けを題材とした教材に最適化されているといえる。音声認識を利用する教材には発音矯正を目的とした教材[2]が多くある

が，本教材は発音の矯正よりも即時的な使い分けの練習に焦点を当て開発されている．また，一般に，音声認識処理を行うには，特別なソフトウェアのインストールが必要となる場合が多いが，本システムでは，Web上の教材にアクセスするだけで簡単に利用できる．

3. 音声認識システムの仕組

3.1 認識の流れ

音声認識システムの構成を図 1 に示す．

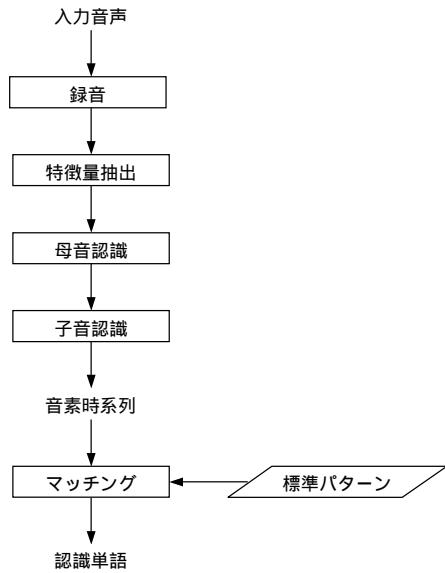


図 1. 音声認識システムの構成

入力音声は，サンプリング周波数 8[kHz]，サンプルサイズ 16bit とする．これをサンプル数 9 すなわち 64[msec]毎のフレームにわたる．

3.2 特徴量抽出

まず，入力音声の振幅値から対数音声パワーを求め，雑音区間を除外する．ここで求めた音声区間の各フレームに対して，予測次数 11 で線形予測分析(Linear Predictive Coding : LPC)および高速フーリエ変換(Fast Fourier Transform : FFT)を行い，音声のスペクトルを求める．さらに，スペクトルのピーク値よりフォルマント周波数の時系列を求める．また，スペクトル全体の平均電力と低域の電力比およびスペクトルの高域の電力と中域の電力比を求める[3]．以上の操作を行うアプリケーションを開発した．この画面を図 2 に示す．

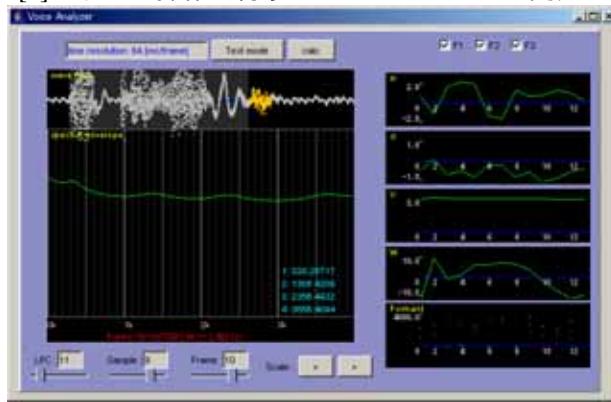


図 2. 特徴量抽出アプリケーション

3.3 音素認識

まず，フォルマント周波数の時系列から母音の種類を識別したのち，音声パワーおよびスペクトルの電力比を用いて子音を推定する．これらの推定は，あらかじめデータベースに用意してある標準パターンとのマッチングによって行う．

マッチング方法としては，現在のところ時間軸伸縮マッチング(Dynamic Time Warping : DTW)を用いている．これは，長さの異なる二つの系列データのマッチングを行う際に，フレーム間の距離尺度の累積値が最小になるように，時間軸を非線形伸縮する方法である．これにより，発声時間が不確定である音声データを扱う

ことができる。しかし、この方法は時間的な変動に対しては強力であるが、個人差などに起因するスペクトルそのものの変動に対しては弱い。そこで、これを補うために隠れマルコフモデル（Hidden Markov Model：HMM）を併用することにした。この方法では、入力音声の変動を吸収するように学習を行うことができるのと、認識精度の向上が期待できる。また、学習を行うことで、解答データにない単語にも対応できるようになる。

具体的には、まず、以下4で得られた留学生の音声データを用いて、音素単位でHMMを作成する。続いて、これらの音素HMMを連結学習することで、単語HMMを作成する。そして、生起確率が最も高い単語を認識単語とする。

4. 教材内容

4.1 自動詞・他動詞の使い分け

本Web教材では、自動詞・他動詞を多様な問題形式で練習できるようになっており、今回開発した音声を用いた教材はその中の一部である。教材全体としては、問題提示方法としてテキスト、音声、映像があり、学習者の解答方法としては、キーボード、マウス、音声入力がある。

まず、キーボードおよびマウスを使った問題形式には、文章中の空欄に正しい自動詞または他動詞を入力するもの、自動詞・他動詞のどちらを使えばよいかを選ぶもの、動詞群を自動詞と他動詞に分けるものがある。

一方、筆者が開発を行った音声認識を利用した教材については、以下に詳述する。現在開発済みの部分は、自動詞・他動詞の変換練習問題および映像を用いた総合的な自動詞・他動詞の使い分け練習教材である。

4.2 変換練習問題

これは、対になる自動詞・他動詞を発音することで、両動詞を一組で効率よく覚えることを目的とした教材である。教材にアクセスすると、自動詞または他動詞がテキストおよび音声で提示される。学習者はその動詞が自動詞であるか他動詞であるかを判断し、それが自動詞であれば対応する他動詞を、逆に他動詞であれば自動詞を発音する。例を表1に示す。

表 1. 変換練習問題例

問題	解答
つづきます（自動詞）	つづけます（他動詞）
あけます（他動詞）	あきます（自動詞）
はじまります（自動詞）	はじめます（他動詞）

この練習では、自らの発声を意識することで、テキストのみの練習と比較して学習者へのより確実な定着が期待できる。また、キーボード入力ではなく音声入力を用いるので、より即時的な練習が行える。この教材の画面を図3に示す。

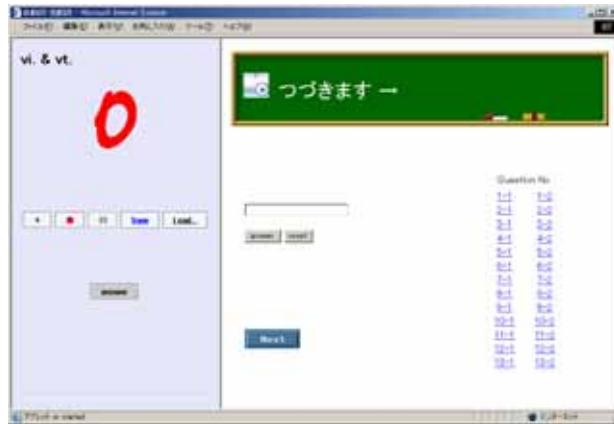


図 3. 自動詞・他動詞変換練習

4.3 映像を用いた練習問題

これは、会話中に自動詞または他動詞が出てくる映像を見て、具体的にどのような場面でどのように自動詞・他動詞が使われるのかをイメージしながら会話に参加する教材である。映像は、演劇サークルに協力を依頼し作成したもので、30秒～1分程度の短いスキットになっている。学習者はこの映像を一度見て、自動詞または他動詞が使われている部分を確認する（図4）。次に、その部分で映像が止まるので、正しい動詞を発音する（図5）。このように、映像の登場人物の台詞を発音することで、自動詞・他動詞の使い分けを学習できるようになっている。最後に、会話中に出てきた自動詞・他動詞の組が文字で表示されるので、正しい方を選択する（図6）。

以上の一連の流れで，映像，音声，文字を用いた総合的な学習が行える。また，図4～6のうち自分が練習したい部分だけを行うこともできる。

このように，映像と音声入力を併用することで，自動詞・他動詞を用いる具体的な場面を疑似体験できるようになっている。



図 4. 映像と文字による確認



図 5. 音声入力による回答



図 6. 文字による選択問題

5. 今後の展望

2004年度後期に早稲田大学の外国人日本語学習者を対象とした本教材の評価実験が予定されている。そこで本教材の有益性についてアンケート調査を実施する。また、音声認識エンジンの認識精度の向上を目指す。一方、本教材は、教材部分と音声認識部分を別々に開発しているため、音声による解答入力・自動採点を他の電子教材へ応用することは容易である。既存の電子教材での利用の可能性を模索する。

参考文献

- [1] Toshitaka TAMURA, Shinichi FUJITA, Akiko SWAN, Toshiko HOSAKA, Seinosuke NARITA. (2002). A Case Study of Web-based Supplementary Material that complements Classroom Teaching. E-Learn 2002: 2250-2253
- [2] Julie Voce. (2003). Evaluating the Use of Audio-Visual Speech Synthesis in Pronunciation Training. ED-MEDIA 2003: 491-494
- [3] 三輪譲二, 新津善弘, 牧野正三, 城戸健一. (1978). 音声スペクトルの概略形とその動特性を利用した単語音声認識システム. 日本音響学会誌: 186-193