

5B-6 「音声と画像のデータベースを活用した
日本語教育アプリケーションを制作する立場からの技術的要望」

小関基宏

日本放送出版協会

伊藤由美

(株) マルチ・エンタテインメント・ワーク・ショップ

1. 日本語文化をマルチメディアにしたい

町中のファーストフードで耳にする若者たちの言葉づかい。平板化されたアクセントに、簡略語、チョー付き言葉に、意味不明の相づち言葉……何だか外国にでも来ているような気分になる。これはこれで、時代の流れとともに変わっていく“言葉の性”ともいうべき現象なのではあるが、やはりこれが日本語と思うと少々寂しい気がしなくもない。

日本語には、お国言葉といわれる『方言』と、アナウンサーなどが用いる聞き易く正しく伝えるための『標準語』がある。前者は、全国各地の風土と歴史が生みだしたいわば伝統文化であり、後者は、あらゆる地域の世代間または日本語を用いる外国人に情報を共有させるための実用文化である。これらはともに、「意味」「発音」「高低のアクセント」などで構成されている。

言葉の意味は各種辞書により誰でもがその内容を知ることができる。しかし、膨大な語彙と非常に複雑な変化を伴う日本語の発音とアクセントについては、印刷物などの文字媒体だけでは伝授できないことが多く、専門の講師や方言語りの生の声を聞きながら学ぶことが不可欠となる。

このような状況の中、生の講師による言語学習に変わる情報処理技術を活用した画像と音声での学習コンテンツの実用化は、言語を学ぶ者にとって飛躍的な学習の機会創出となるのである。

2. 制作上で突き当たった壁

まず手がけたのは、昭和初期に研究資料のために録音された「全国方言資料」（日本放送協会編）の一部音声を使い、子供からお年寄りまで全国の方言を楽しみながら知ることができるソフトの制作である。

ここで肝心なのは、エンターテインメント性を持たせた画面進行構成とビジュアルで楽しませる演出である。ワンウェイの映像作品ならば雑作もないことであるが、双方向性を持たせて遊びながら進行させたいと考え、とどんデータ容量が増えてしまう。

一般家庭向け程度のコンピュータでもサクサク稼働できる範囲にデータの容量を押さえながら、音質・画質を落とさずに・・・となると、やはりあきらめざるを得ないところも多くなる。このソフトは、言語学者の金田一春彦氏の解説付きということであったため、できるかぎり実写動画である金田一節を使った方言研究の話を随所に入れたかったが、半分以上に減らした。そのほかにも、地元の方々の参加で新たに録音した音声を使ったコーナーでは、声だけではなく参加者の表情も入れた方言アルバムの作成を希望したが、動画から静止画への変更を余儀なくされた。

これらの壁は、圧縮技術と CPU 性能の向上などが解決してくれる問題であると予想される。一刻も早く、安価でかつ汎用性の高い技術の開発を望むものである。

次に、トライしたのが全国の方言を学習できるソフトである。全国の方言のモデル音声を聞いたり、自分の発音を各地の方言に変えて、それを聞きながら覚えるという構想であった。このソフトは、放送メディア等の発達により年々消えていく方言を復活させると共に、転勤先や転校先でなかなかご当

地の言葉になじめない人、演劇や朗読でお国言葉を使う人、地方文化を保存しようとする人々に役立てていただくというのがキッカケであった。

このもくろみは、最初から壁があった。それは、方言は一定の法則で分析することが不可能に近いため、全ての方言のサンプルをデータ化しておかなければ、合成音声を使ったとしても上記の内容での学習は実現できないということである。

例えば、既に残されている前述「全国方言資料」に納められている音声や、現在使われている各地の話し言葉を録音し、自動的に単語や文に切り出して、その言葉を文字とアクセントに変換ができれば、この「全国方言学習辞典」の制作は可能になるのではないかという素人考えはずれていたようだ。方言への思いを残しつつ、このソフトの制作はひとまずあきらめた。

元データを作ることが出来ない「方言」がダメだとしたら、元データのある「標準語」ならばできるのではないかと始めたのが「日本語発音アクセント辞典」のマルチメディア化である。

標準語の発音とアクセントについては、既に印刷媒体として「日本語発音アクセント辞典」(NHK放送文化研究所編)が出版されている。このデータを参照して構成すればと当初は簡単に考えていた。しかし、単語だけで6万9千語、ひとつの単語に複数のアクセントを持つものもあり、また、助詞がつくことでアクセントが変わるものや、慣例により変化するもの、伝統的なアクセントが現在では変化してしまっているものなど、それらを全て網羅すると8万語にもおよぶ。これらに全て音声をつけると、その容量も膨れ上がる。

データ容量を減らすために合成音声で作成する案も出されたが、文字や記号だけでは表せない微妙な発音が多いため、聞いて覚える正しい発音アクセントのモデル音声にするには熟練したアナウンサーの肉声でなければならぬ。そこで、ともかく生の声で録音することにした。

ここでひとつ壁にぶつかった。録音技術は難しいものではないが、録音された音声をひとつずつ切り出し、ファイル名を付け、文字と高低アクセント記号と音声を照合し、かつ高低アクセントをビジュアルで見せていく。この一連の作業のほとんどを手作業で処理しなければならないことである。作業内容としては極単純なものであるはずが、何故自動化できないのか。技術面に盲目である者にとって思いがけない実体に直面した。

以上3つの日本語をテーマにしたコンテンツの制作に関与し共通することは、データベースの技術は制作者のどんなわがままな希望でも実現可能な程に進んでいるが、元になる素材を作成する段階で、広い汎用性のある技術が望まれるということである。

もちろん、手作業でも経費と時間と人手をかければ可能なことも多い。しかし、これからのホームネットワーク時代では、需要する側が増えるとともに、普通の人々がソフト供給側にまわることもあるだろう。そうした中で、より安価で時間をかけずにデータベースの素材を作ることができれば、様々な分野に、データベースを活用したコンテンツの世界が広がるのではないかと考える。

A technical request from the position that Japanese educational application which the database of the voice and the image was made use of for is made.

Yumi Itoh <yumibox@mews.co.jp>

Multi Entertainment Work Shop Co.,Ltd.

2F Skycourt Shibuya 40-1 Kamiyamacho Shibuya-ku Tokyo Japan 150-0047