

全国方言談話資料データベースの作成に向けて

Database of Discourse in Japanese Dialect

井上文字子(国立国語研究所)
田原広史(大阪樟蔭女子大学)

2000/7/15

16th DB-West

1

1. 方言談話資料の現状

研究素材として見直されつつある方言談話資料

- 全国方言資料 (NHK)
- 方言談話資料 (国語研)
- 各地方言収集緊急調査 (文化庁)

2000/7/15

16th DB-West

2

2. 各地方言収集緊急調査(その1)

- 各都道府県において緊急に調査し、記録・保存することが目的
- 各県5地点 (約200地点)
- 7場面の対話に絞り、1地点10時間の録音を収集
- 老年60歳以上、若年20~30代
- 3時間分を文字化・標準語訳

2000/7/15

16th DB-West

3

2. 各地方言収集緊急調査(その2)

- 永久保存
- 公開されているのはごく一部
- 文化庁から国語研への移管を機にデータベース化を企画する
- データベース科研 (H9~)
- 文字化資料の電子テキスト化
- 音声資料のデジタル化

2000/7/15

16th DB-West

4

3. 全国方言談話資料データベース

3.1 データ量などの問題

- 2000時間の録音 (83日分)
- 600時間分の文字化 (25日分)
- どの程度が量的に妥当か?
- NHK1999との比較
2.2GB, 30時間分, 141地点
PDF形式 + 音声ファイルリンク
ただし文字化部分は画像

2000/7/15

16th DB-West

5

3.2 文字化部分の作業

- 手書き文字化・標準語訳原稿を電子テキスト化
- 表記・形式を統一
- できるだけ原本に忠実に入力
- 文字化部分と標準語訳部分を分け、エクセルで入力
- 最終的にはWordで研弼

2000/7/15

16th DB-West

6

3.3 音声部分の作業(その1)

- カセットテープをパソコンに取り込む(デジタル化)
- ソニーVAIOのギガポケット(ビデオインターフェイス)を利用
- 直接ハードディスクに談話単位(15~30分)で録画(録音)
- 画面トラックに文字化資料を同期させておく(CCDカメラで)

2000/7/15

16th DB-West

7

3.3 音声部分の作業(その2)

- パソコン内で再生しながら、音声ソフト(sp4win)を用いて再録音
- サンプリングレート、形式の調整
- 8ビット, 22.05KHz, モノラルで(NHK1999と同じ形式)
- さらに発話単位に編集して完成

2000/7/15

16th DB-West

8

3.4 データベース化の方法(その1)

二つの方法

- ワードプロ・エクセル等で作成したものをPDF化し、音声ファイルをリンクさせる(PDF方式)
- データベースソフトを用いて、文字化と音声を一体化したものを作成する(データベース方式)

2000/7/15

16th DB-West

9

3.4 データベース化の方法(その2)

- PDF方式
誰にでも使える
個々のファイルが扱いやすい
- データベース方式
検索が自由にできる
色々な角度から分析できる

2000/7/15

16th DB-West

10

3.4 データベース化の方法(その3)

現時点での方針

- PDF方式とデータベース方式の両方を作成する
- 量的に両方が無理ならば、PDF方式を優先し、データベース方式のものは、音声をユーザが後で組み込む形のものにする

2000/7/15

16th DB-West

11

4. 今後の予定および課題

- 音声談話資料の編集の難しさ
- 「編集にかかる手間・費用」と「学術的価値」との兼ねあい
- 常に利用のことを頭に置くこと
- 2004年までに完成の予定
- パソコン環境の変化への対応

2000/7/15

16th DB-West

12