# 目次

<table>
<thead>
<tr>
<th>第 1 章</th>
<th>概説</th>
<th>前川喜久雄</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>本書の目的と対象</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>CSJ の対象と特徴</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1.2.1</td>
<td>収録された音声の特徴</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1.2.2</td>
<td>学会講演と模擬講演</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1.2.3</td>
<td>その他の講演</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1.2.4</td>
<td>朗読</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1.2.5</td>
<td>再朗読</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1.2.6</td>
<td>対話</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1.2.7</td>
<td>再朗読および対話の取録についての補遺</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>1.2.8</td>
<td>話者の分布</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>研究用付加情報</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>音声の自発性</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>1.4.1</td>
<td>自然の序列</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>1.4.2</td>
<td>印象評定</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>1.4.3</td>
<td>音声収録記録票とアンケート</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>1.4.4</td>
<td>話し言葉という用語について</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>XML 文書</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>話者の個人情報</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td>CSJ の解析結果</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>1.7.1</td>
<td>語の融合の決定樹</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>1.7.2</td>
<td>意識と行動</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>1.7.3</td>
<td>語彙の変異率</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>まとめ</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>第 2 章</th>
<th>転記テキスト</th>
<th>小磯花絵・西川賢哉・間瀬洋子</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>転記テキストの概要</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>2.1.1</td>
<td>設計方針</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>2.1.2</td>
<td>転記テキストの基本構成</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
2.2 転記基本単位の認定 .......................... 32
  2.2.1 認定基準 .................................. 32
     2.2.1.1 転記基本単位の種類 ..................... 32
     2.2.1.2 転記単位A・転記単位D .................... 33
     2.2.1.3 転記単位B ............................... 34
     2.2.1.4 転記単位C ............................... 34
     2.2.1.5 転記単位間の時間的制約 ................. 35
  2.2.2 作業の流れ .................................. 36
2.3 基本形の表記法 .................................. 38
  2.3.1 表記の概要 .................................. 38
     2.3.1.1 使用する字種 ............................ 38
     2.3.1.2 字種間の書き分けの方針 ................. 38
     2.3.1.3 仮名遣い ................................. 39
     2.3.1.4 送り仮名 ................................. 39
     2.3.1.5 同音異義語 ............................... 39
     2.3.1.6 平仮名 ................................. 40
     2.3.1.7 片仮名 ................................. 41
     2.3.1.8 漢字 .................................. 42
     2.3.1.9 アルファベット・算用数字 ................. 42
     2.3.1.10 記号 .................................. 43
  2.3.2 表記の詳細 .................................. 46
     2.3.2.1 名詞 .................................. 46
     2.3.2.2 固有名詞 ................................. 48
     2.3.2.3 数詞 .................................. 50
     2.3.2.4 副詞 .................................. 51
     2.3.2.5 擬音語・擬態語 ........................... 52
     2.3.2.6 感動詞 .................................. 54
     2.3.2.7 動詞 .................................. 57
     2.3.2.8 形容詞 .................................. 59
     2.3.2.9 助詞・助動詞 ............................. 59
     2.3.2.10 接辞 .................................. 60
  2.3.3 外来語 .................................. 60
  2.3.4 口語表現 ................................. 63
  2.3.5 促音の扱い .................................. 67
  2.3.6 転記作業用の辞書の構築 ........................ 70
     2.3.6.1 用字用語辞書 ............................. 70
     2.3.6.2 仮名漢字変換用辞書 ...................... 71
     2.3.6.3 表記確認用辞書 ........................... 71
2.4 発音形の表記法 ......................................................... 73
  2.4.1 字種 ............................................................. 73
  2.4.2 発音の急激な相違・言い間違い .................................. 73
  2.4.3 非音化的母音の引き延ばし・子音の引き延ばし ......................... 73
  2.4.4 練習字における母音移動 ........................................ 74
  2.4.5 ビッチの急激な変動に伴って知覚される母音 ............................ 75
  2.4.6 暗味な発音の扱い ................................................ 75
    2.4.6.1 発音は暗味だが語自体は特定できる場合 ....................... 76
    2.4.6.2 発音が暗味でその語が存在するか否か迷う場合 .................... 77
    2.4.6.3 発音が暗味で語が特定できない場合 ............................. 77
    2.4.6.4 複数の読みを持つ語や口語表現に関連する語で迷う場合 .............. 78
2.5 転記テキストで用いるタグ .............................................. 80
  2.5.1 タガ (F) ....................................................... 82
    2.5.1.1 フィラー .................................................. 83
    2.5.1.2 感情表現系感動詞 ......................................... 85
    2.5.1.3 応答表現 .................................................. 85
  2.5.2 タガ (D) ....................................................... 88
  2.5.3 タガ (D2) ...................................................... 92
  2.5.4 タガ (?) ....................................................... 94
  2.5.5 タガ (O) ....................................................... 95
  2.5.6 タガ (M) ....................................................... 97
  2.5.7 タガ (R) ....................................................... 99
  2.5.8 タガ (X) ....................................................... 100
  2.5.9 タガ (A) ....................................................... 100
  2.5.10 タガ (K) ...................................................... 102
  2.5.11 タガ (W) ...................................................... 103
  2.5.12 タガ (B) ...................................................... 105
  2.5.13 タガ(笑), タガ(泣), タガ(咳) .................................. 107
  2.5.14 タガ (L) ....................................................... 108
  2.5.15 タガ<FV> ..................................................... 108
  2.5.16 タガ<VN> ..................................................... 108
  2.5.17 タガ<H> ..................................................... 109
  2.5.18 タガ<Q> ..................................................... 110
  2.5.19 タガ<P> ..................................................... 110
  2.5.20 タガ<笑>, タガ<咳>, タガ<息> ................................ 111
  2.5.21 タガ間の共起関係 ............................................. 111

2.6 対話 ................................................................. 114
  2.6.1 転記基本単位の認定 ............................................. 114
第3章 形態論情報

3.1 長単位・短単位 .................................................. 134
  3.1.1 単位の設計 ................................................. 134
      3.1.1.1 語彙調査の単位 .................................. 134
      3.1.1.2 CSJの単位 ....................................... 135
  3.1.2 長単位の認定基準 ........................................ 138
      3.1.2.1 文節認定基準 ................................... 138
      3.1.2.2 長単位認定基準 .................................. 139
  3.1.3 短単位の認定基準 ........................................ 141
      3.1.3.1 最小単位認定基準 ................................ 141
      3.1.3.2 短単位認定基準 .................................. 142
  3.1.4 話し言葉特有の現象の単位認定 ................................ 144
  3.1.5 タグの扱い ............................................... 150

3.2 代表形・代表表記 ............................................. 154
  3.2.1 同音異義別の必要性 .................................... 154
  3.2.2 代表形・代表表記の付与基準 ............................. 156
      3.2.2.1 代表形の付与基準 ................................ 156
      3.2.2.2 代表表記の付与基準 ............................... 160
  3.2.3 話し言葉特有の現象に対する代表形・代表表記の付与 ................. 162

3.3 品詞等の情報 .................................................. 166
  3.3.1 品詞情報の設計 ........................................... 166
  3.3.2 品詞 ....................................................... 170
  3.3.3 活用の種類・活用形 ..................................... 172
  3.3.4 その他の情報 ............................................. 173
  3.3.5 話し言葉特有の現象に対する品詞情報の付与 ........................ 174

3.4 今後の検討課題 ................................................. 178
<table>
<thead>
<tr>
<th>章節</th>
<th>項目</th>
<th>項目詳細</th>
<th>頁碼</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.1</td>
<td>単位に関する検討課題</td>
<td></td>
<td>178</td>
</tr>
<tr>
<td>3.4.1.1</td>
<td>長単位に関する検討課題</td>
<td></td>
<td>178</td>
</tr>
<tr>
<td>3.4.1.2</td>
<td>短単位に関する検討課題</td>
<td></td>
<td>178</td>
</tr>
<tr>
<td>3.4.2</td>
<td>代表形・代表表記に関する検討課題</td>
<td></td>
<td>179</td>
</tr>
<tr>
<td>3.4.3</td>
<td>品詞情報に関する検討課題</td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>3.4.1</td>
<td>単位に関する検討課題</td>
<td></td>
<td>178</td>
</tr>
<tr>
<td>3.4.1.1</td>
<td>長単位に関する検討課題</td>
<td></td>
<td>178</td>
</tr>
<tr>
<td>3.4.1.2</td>
<td>短単位に関する検討課題</td>
<td></td>
<td>178</td>
</tr>
<tr>
<td>4.1</td>
<td>短単位・長単位データベースの設計と運用</td>
<td></td>
<td>188</td>
</tr>
<tr>
<td>4.1.1</td>
<td>概要</td>
<td></td>
<td>188</td>
</tr>
<tr>
<td>4.1.2</td>
<td>単位データベースの設計</td>
<td></td>
<td>190</td>
</tr>
<tr>
<td>4.1.2.1</td>
<td>構築の背景</td>
<td></td>
<td>190</td>
</tr>
<tr>
<td>4.1.2.2</td>
<td>単位データベースの特徴</td>
<td></td>
<td>191</td>
</tr>
<tr>
<td>4.1.3</td>
<td>単位データベースの構造</td>
<td></td>
<td>193</td>
</tr>
<tr>
<td>4.1.3.1</td>
<td>単位データベースのレコード形式</td>
<td></td>
<td>193</td>
</tr>
<tr>
<td>4.1.3.2</td>
<td>レコード ID 関連</td>
<td></td>
<td>193</td>
</tr>
<tr>
<td>4.1.3.3</td>
<td>転記テキストに関する情報</td>
<td></td>
<td>195</td>
</tr>
<tr>
<td>4.1.3.4</td>
<td>短単位の情報</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>4.1.3.5</td>
<td>長単位の情報</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>4.1.3.6</td>
<td>管理情報</td>
<td></td>
<td>197</td>
</tr>
<tr>
<td>4.1.4</td>
<td>単位データベースの運用</td>
<td></td>
<td>198</td>
</tr>
<tr>
<td>4.1.4.1</td>
<td>単位データの整合性チェック</td>
<td></td>
<td>198</td>
</tr>
<tr>
<td>4.1.4.2</td>
<td>転記テキストの修正と同期</td>
<td></td>
<td>200</td>
</tr>
<tr>
<td>4.1.4.3</td>
<td>索引</td>
<td></td>
<td>201</td>
</tr>
<tr>
<td>4.2</td>
<td>人手解析による短単位・長単位データの構築</td>
<td></td>
<td>202</td>
</tr>
<tr>
<td>4.2.1</td>
<td>構築する単位データの全体像</td>
<td></td>
<td>202</td>
</tr>
<tr>
<td>4.2.2</td>
<td>構築方法の概略</td>
<td></td>
<td>202</td>
</tr>
<tr>
<td>4.2.3</td>
<td>人手短単位解析の手順</td>
<td></td>
<td>203</td>
</tr>
<tr>
<td>4.2.3.1</td>
<td>概要</td>
<td></td>
<td>203</td>
</tr>
<tr>
<td>4.2.3.2</td>
<td>人手短単位認定</td>
<td></td>
<td>203</td>
</tr>
<tr>
<td>4.2.3.3</td>
<td>「茶釜」による形態素解析</td>
<td></td>
<td>204</td>
</tr>
<tr>
<td>4.2.3.4</td>
<td>品詞情報の変換</td>
<td></td>
<td>206</td>
</tr>
<tr>
<td>4.2.3.5</td>
<td>基本形と発音形との対応づけ</td>
<td></td>
<td>207</td>
</tr>
<tr>
<td>4.2.3.6</td>
<td>単位データベースへの登録</td>
<td></td>
<td>209</td>
</tr>
<tr>
<td>4.2.4</td>
<td>人手長単位解析の手順</td>
<td></td>
<td>209</td>
</tr>
<tr>
<td>4.2.4.1</td>
<td>BUP による長単位解析</td>
<td></td>
<td>209</td>
</tr>
<tr>
<td>4.2.4.2</td>
<td>単位データベースへの登録</td>
<td></td>
<td>212</td>
</tr>
<tr>
<td>4.3</td>
<td>単位データベース修正ツール</td>
<td></td>
<td>213</td>
</tr>
<tr>
<td>4.3.1</td>
<td>概要</td>
<td></td>
<td>213</td>
</tr>
<tr>
<td>4.3.2</td>
<td>設計</td>
<td></td>
<td>213</td>
</tr>
</tbody>
</table>
4.3.3 システム構成 ................................. 214
4.3.4 修正ツールの概要 ............................... 215
4.3.5 単位データの検索と読み込み ................. 216
  4.3.5.1 方法1: GUIを用いた方法 .................. 217
  4.3.5.2 方法2: SQL文を直接記述する方法 .......... 218
  4.3.5.3 方法3: 単位データファイルを読み込む方法 .... 219
  4.3.5.4 方法4: 検索結果から再検索する方法 .......... 219
4.3.6 単位データの修正 ............................... 220
  4.3.6.1 単位分割位置の修正機能 .................. 220
  4.3.6.2 修正支援機能 ............................ 222
  4.3.6.3 単位辞書検索機能 ......................... 225
4.3.7 単位データの更新と排他制御 ................. 226
  4.3.7.1 更新処理 ............................... 226
  4.3.7.2 排他制御 ............................... 227
4.3.8 ニュースシステム ............................... 227
4.3.9 修正ツールの運用 ............................... 229
4.3.10 修正ツールによる単位データの構築結果 ......... 230
4.4 短単位辞書の構築 ............................... 231
  4.4.1 概要 ....................................... 231
  4.4.2 設計 ....................................... 231
  4.4.3 短辞書辞書の構造 ......................... 232
    4.4.3.1 辞書項目の例 .......................... 232
    4.4.3.2 DTD ................................. 234
  4.4.4 活用表 ..................................... 237
    4.4.4.1 概要 .................................. 237
    4.4.4.2 活用表の記述例 ........................ 238
    4.4.4.3 DTD ................................. 238
    4.4.4.4 解説 .................................. 240
  4.4.5 短単位辞書の構築 ......................... 250
    4.4.5.1 人手解析単位データの構築に伴うもの ......... 250
    4.4.5.2 自動解析単位データの人手修正に伴うもの .... 251
    4.4.5.3 用例の付与 ................................ 252
4.5 まとめ ........................................... 253

第5章 節単位情報 ....................................... 255
  5.1 話し言葉の「文」をどう捉えるか .................. 255
    5.1.1 書き言葉の「文」と話し言葉の「文」 .......... 255
    5.1.2 話題レベルにおける基本単位の必要性 .......... 259
    5.1.3 節単位が認定されるまでの流れ ................ 260
5.2 CBAP-csjによる自動発話分割処理 .............................. 263
5.2.1 理論的背景 ........................................... 263
5.2.2 節境界の検出アルゴリズム ............................... 265
5.2.3 CBAP-csjが検出する節境界の種類 ......................... 267
5.2.4 コアに含まれる学会講演・模擬講演の節境界ラベル検出結果 ... 269
5.3 節境界ラベルの一覧 ........................................ 271
5.3.1 絶対境界 ............................................. 271
5.3.1.1 [文末] ........................................... 271
5.3.1.2 [文末候補] ...................................... 272
5.3.1.3 [と文末] ......................................... 273
5.3.2 強境界 ................................................ 273
5.3.2.1 /並列節ヲ/ ....................................... 273
5.3.2.2 /並列節ケレドモ/ /並列節ケレド/ /並列節ケレドモ/ /並列節ケド/ ........ 274
5.3.2.3 /並列節シ/ ....................................... 275
5.3.2.4 /ヨウニ節/ ......................................... 276
5.3.3 弱境界 ................................................ 276
5.3.3.1 <条件節タラ> <条件節タラバ> .......................... 276
5.3.3.2 <条件節ト> ......................................... 277
5.3.3.3 <条件節ナラ> <条件節ナラバ> .......................... 277
5.3.3.4 <条件節レバ> ...................................... 278
5.3.3.5 <理由節カラ> ....................................... 279
5.3.3.6 <理由節カラニハ> ................................ 279
5.3.3.7 <理由節カラ-助詞> ................................ 279
5.3.3.8 <理由節ノデ> ....................................... 280
5.3.3.9 <タリ節> <タリ節-助詞> ............................... 280
5.3.3.10 <テ節> ............................................... 281
5.3.3.11 <テハ節> ........................................... 281
5.3.3.12 <テモ節> ........................................... 282
5.3.3.13 <テカラ節> <テカラ-助詞> <テ節-助詞> .......................... 283
5.3.3.14 <トカ節> <トカ節-助詞> ............................... 283
5.3.3.15 <ノニ節> ........................................... 284
5.3.3.16 <連用節> .......................................... 284
5.3.3.17 <引用節> <引用節-助詞> <引用節トノ> .................. 285
5.3.3.18 <トイウ節> ......................................... 286
5.3.3.19 <間接疑問節> <間接疑問節-助詞> ....................... 287
5.3.3.20 <連体節テノ> ....................................... 288
5.3.3.21 <並列節ダノ> ....................................... 288
5.3.3.22 <並列節デ> ......................................... 289
5.3.3.23 並列節ナリ ................................. 289
5.3.3.24 フィラー文 ................................. 290
5.3.3.25 感動詞 ................................. 290
5.3.3.26 接続詞 C 接続詞 L 接続詞 CL 接続詞 M ................................. 291
5.4 人手による修正と節単位の認定 ................................. 292
  5.4.1 人手修正作業の概要 ................................. 292
  5.4.2 人手修正作業で扱う項目の分類 ................................. 293
  5.4.3 人手修正作業の流れ ................................. 295
5.5 人手修正規則の一覧 ................................. 295
  5.5.1 大局的な文法構造に関するもの ................................. 295
    5.5.1.1 体言止 ................................. 295
    5.5.1.2 引用節構造 ................................. 296
    5.5.1.3 連体節構造 ................................. 298
  5.5.2 非流暢性に関するもの ................................. 299
    5.5.2.1 倒置 ................................. 299
    5.5.2.2 捕入節 ................................. 300
    5.5.2.3 捕入文 ................................. 301
    5.5.2.4 フィラー文 ................................. 302
    5.5.2.5 言いさし ................................. 303
    5.5.2.6 言い直しマーカー ................................. 304
    5.5.2.7 仮語の言い直し ................................. 305
  5.5.3 係り受け構造に関するもの ................................. 305
    5.5.3.1 主題の共有 ................................. 305
    5.5.3.2 主題の飛び越し ................................. 306
    5.5.3.3 格要素の飛び越し ................................. 307
    5.5.3.4 連体形 ................................. 307
  5.5.4 話構文構文上の問題に関するもの ................................. 308
    5.5.4.1 話題導入表現 ................................. 308
    5.5.4.2 直後がまとめ表現 ................................. 309
    5.5.4.3 話題の転換点 ................................. 310
    5.5.4.4 大きい切れ目 ................................. 311
  5.5.5 不適切な節境界ラベルに関するもの ................................. 314
    5.5.5.1 関係助詞 ................................. 314
    5.5.5.2 と文末 ................................. 314
    5.5.5.3 例文など ................................. 315
    5.5.5.4 格助詞相当表現 ................................. 316
    5.5.5.5 強→弱 ................................. 316
    5.5.5.6 文末滞補 ................................. 317
<table>
<thead>
<tr>
<th>章目</th>
<th>項目</th>
<th>頁数</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5.5.7</td>
<td>非文末</td>
<td>318</td>
</tr>
<tr>
<td>5.5.8</td>
<td>タグミス-その他</td>
<td>319</td>
</tr>
<tr>
<td>5.6</td>
<td>節単位情報の管理と XML 文書への格納</td>
<td>320</td>
</tr>
<tr>
<td>5.7</td>
<td>節単位が持つ意味と今後の展望</td>
<td>321</td>
</tr>
<tr>
<td>第6章</td>
<td>分節音情報</td>
<td>藤本雅子・菊池英明・前川喜久雄</td>
</tr>
<tr>
<td>6.1</td>
<td>分節音ラベリングの方針</td>
<td>323</td>
</tr>
<tr>
<td>6.2</td>
<td>分節音ラベル</td>
<td>323</td>
</tr>
<tr>
<td>6.3</td>
<td>ラベリング方式</td>
<td>325</td>
</tr>
<tr>
<td>6.3.1</td>
<td>分節音ファイルの形式と注意点</td>
<td>325</td>
</tr>
<tr>
<td>6.3.2</td>
<td>ラベリングの基準</td>
<td>327</td>
</tr>
<tr>
<td>6.4</td>
<td>利用上の注意</td>
<td>329</td>
</tr>
<tr>
<td>6.4.1</td>
<td>ラベルの融合</td>
<td>329</td>
</tr>
<tr>
<td>6.4.2</td>
<td><code>&lt;el&gt;</code>について</td>
<td>330</td>
</tr>
<tr>
<td>6.4.3</td>
<td>転記テキスト中のタグ<code>&lt;el&gt;</code>について</td>
<td>331</td>
</tr>
<tr>
<td>6.5</td>
<td>ラベリング作業の実施方法</td>
<td>331</td>
</tr>
<tr>
<td>6.6</td>
<td>XML 文書および音響モデルにおける「音素」について</td>
<td>331</td>
</tr>
<tr>
<td>6.7</td>
<td>分節音ラベリング上の問題点と CSJ での処理方法</td>
<td>332</td>
</tr>
<tr>
<td>6.7.1</td>
<td>有声破裂音/g/の異音とその表記方法</td>
<td>332</td>
</tr>
<tr>
<td>6.7.2</td>
<td>無声破裂音の摩擦音化</td>
<td>335</td>
</tr>
<tr>
<td>6.7.3</td>
<td>`/Q/+破裂音での「Q」の付与</td>
<td>336</td>
</tr>
<tr>
<td>6.7.4</td>
<td>`/N/+有声破裂音の融合</td>
<td>336</td>
</tr>
<tr>
<td>6.7.5</td>
<td>有音子音と拗音の分節</td>
<td>337</td>
</tr>
<tr>
<td>6.7.6</td>
<td><code>&lt;sv&gt;</code>, <code>&lt;uv&gt;</code>, <code>&lt;fr&gt;</code>の付与例</td>
<td>338</td>
</tr>
<tr>
<td>6.7.7</td>
<td>その他</td>
<td>340</td>
</tr>
<tr>
<td>6.7.7.1</td>
<td><code>/h/</code>の有声化</td>
<td>340</td>
</tr>
<tr>
<td>6.7.7.2</td>
<td>撥音と鼻子音の融合</td>
<td>340</td>
</tr>
<tr>
<td>6.7.7.3</td>
<td>母音無声化に関わる問題</td>
<td>342</td>
</tr>
<tr>
<td>第7章</td>
<td>韻律情報</td>
<td>五十嵐陽介・菊池英明・前川喜久雄</td>
</tr>
<tr>
<td>7.1</td>
<td>韻律情報とそのラベリング</td>
<td>347</td>
</tr>
<tr>
<td>7.1.1</td>
<td>ToBI と J_ToBI</td>
<td>347</td>
</tr>
<tr>
<td>7.1.2</td>
<td>J_ToBI の問題点と X-JToBI</td>
<td>349</td>
</tr>
<tr>
<td>7.1.3</td>
<td>トーン層</td>
<td>351</td>
</tr>
<tr>
<td>7.1.3.1</td>
<td>アクセント</td>
<td>351</td>
</tr>
<tr>
<td>7.1.3.2</td>
<td>句頭音調</td>
<td>352</td>
</tr>
<tr>
<td>7.1.3.3</td>
<td>句頭境界音調</td>
<td>353</td>
</tr>
<tr>
<td>7.1.3.4</td>
<td>句末境界音調</td>
<td>354</td>
</tr>
<tr>
<td>7.1.3.5</td>
<td>ポインター</td>
<td>355</td>
</tr>
</tbody>
</table>
7.1.3.6 エクステンダー ............................. 355
7.1.3.7 フィラーのトーンラベル ........................ 356
7.1.4 BI 層 ...................................... 356
7.1.4.1 中間値 .................................. 356
7.1.4.2 非流動性 ................................. 357
7.1.5 単語層 .................................... 359
7.1.6 プロミネンス層 ............................... 359
7.1.7 注釈層 .................................... 362
7.1.8 ラベルの融合 ............................... 364
7.1.8.1 融合ラベル ............................... 364
7.1.8.2 XML 文書中での表現 ........................ 365
7.1.9 次節以下の構成について ...................... 366
7.1.9.1 ケース分け ............................... 366
7.1.9.2 ラベリングの模式図 ...................... 366
7.2 各層の説明 ................................... 367
7.2.1 トーン層 ................................... 367
7.2.1.1 A：アクセント核 .......................... 367
7.2.1.2 H−：句頭音調 ........................... 374
7.2.1.3 %L：句頭境界音調 ....................... 379
7.2.1.4 (H%)%L：句頭境界音調「仮想的句頭境界音調」 382
7.2.1.5 L%：句末境界音調「下降調」 ............... 383
7.2.1.6 L%H%：句末境界音調「上昇調1」 ............ 386
7.2.1.7 L%L%：句末境界音調「上昇調2」 .......... 391
7.2.1.8 L%HL%：句末境界音調「上昇下降調」 ....... 392
7.2.1.9 L%HLH%：句末境界音調「上昇下降上昇調」 393
7.2.1.10 >：エクステンダー ........................ 393
7.2.1.11 L%L, H%L, LH%L, HL%L：融合ラベル（境界音調） 397
7.2.1.12 FL, FH ................................. 402
7.2.1.13 x：補助記号（F0 値が信頼できない） ....... 402
7.2.1.14 ?：補助記号（カテゴリーが不明確・存在が不明確） 405
7.2.1.15 ！：補助記号（位置が不明確） ............... 407
7.2.1.16 補助記号の融合的利用 ................... 409
7.2.2 BI 層 ................................... 409
7.2.2.1 1 ........................................ 409
7.2.2.2 2 ........................................ 410
7.2.2.3 3 ........................................ 411
7.2.2.4 1+ ....................................... 412
7.2.2.5 1+p ..................................... 413
7.2.2.6 2+ .................................................. 413
7.2.2.7 2+p ............................................. 414
7.2.2.8 2+b ............................................. 415
7.2.2.9 2+bp .......................................... 416
7.2.2.10 3+ ............................................. 417
7.2.2.11 P ............................................. 417
7.2.2.12 D , <D ........................................ 418
7.2.2.13 F , <F ........................................ 418
7.2.2.14 F2 ............................................. 418
7.2.2.15 W ............................................. 418
7.2.2.16 PB ........................................... 419
7.2.2.17 アクセント句・インドネーション句境界の認定 .......... 420
7.2.3 単語層 ........................................... 430
7.2.3.1 単語ラベル ..................................... 430
7.2.3.2 アクセントラベル ................................ 430
7.2.4 プロミネス層 ..................................... 430
7.2.4.1 PNLP .......................................... 430
7.2.4.2 FR ............................................. 431
7.2.4.3 HR ............................................. 431
7.2.4.4 EUAP .......................................... 431
7.2.5 注釈層 ........................................... 432
7.2.5.1 HBP ........................................... 432
7.2.5.2 QQ ............................................. 432
7.2.5.3 AYOR .......................................... 432
7.3 フィラード .......................................... 433
7.4 言いよどみ .......................................... 434
7.5 その他音声論上の問題 .................................. 442
7.5.1 Penult Non-Lexical Prominence ........................... 442
7.5.2 Floating Rise ....................................... 444
7.5.3 Hooked Rise ....................................... 446
7.5.4 Emphasized Unaccented Accentual Phrase ..................... 447
7.5.5 High Beginning Phrase .................................. 449
7.5.6 Quasi Question ....................................... 450
7.5.7 At Your Own Risk ..................................... 450
7.5.8 Parasistic Boundary ..................................... 451
7.5.9 言語情報を持たていないと判断される応答表現 .................. 452

第8章 XML文書 ........................................ 455
8.1 XML文書の目的 ....................................... 455
8.1.1 整合性検証 ........................................... 455
8.1.2 情報間の依存関係を利用したデータ検索 ..................... 456

8.2 XML ..................................................... 457
8.2.1 XML とは ........................................ 457
8.2.2 XML の基本 ......................................... 457

8.3 CSJ 各種情報の記述方法 .................................. 458
8.3.1 基本方針 ............................................ 458
8.3.2 各要素の内容 ........................................ 467
  8.3.2.1 Talk 要素 ........................................ 467
  8.3.2.2 IPU 要素 ......................................... 467
  8.3.2.3 TalkComment 要素 .................................. 468
  8.3.2.4 LineComment 要素 ................................. 468
  8.3.2.5 Comment 要素 ..................................... 468
  8.3.2.6 LUW 要素 .......................................... 469
  8.3.2.7 SUW 要素 .......................................... 470
  8.3.2.8 TransSUW 要素 ................................... 472
  8.3.2.9 Mora 要素 ......................................... 474
  8.3.2.10 Phoneme 要素 ..................................... 475
  8.3.2.11 Phone 要素 ....................................... 475
  8.3.2.12 Noise 要素 ....................................... 476
  8.3.2.13 NonLinguisticSound 要素 ......................... 477
  8.3.2.14 XJToBILabelTone 要素 ........................... 478
  8.3.2.15 XJToBILabelWord 要素 ............................ 479
  8.3.2.16 XJToBILabelBreak 要素 ........................... 479
  8.3.2.17 XJToBILabelPrm 要素 ................................ 480
  8.3.2.18 XJToBILabelMisc 要素 ............................ 480

8.3.3 階層構造からの逸脱に対する措置 .......................... 480
  8.3.3.1 分節音ラベルの融合 ............................... 480
  8.3.3.2 語の発音レベルでの融合 ............................ 481

8.3.4 CSJ 研究用付加情報の XML による記述例 ..................... 481

8.4 CSJ 各種情報の整合性検証 ................................ 484
  8.4.1 CSJ 各種情報の作成過程概要 ......................... 484
    8.4.1.1 転記情報 ........................................ 484
    8.4.1.2 形態論情報 ....................................... 484
    8.4.1.3 分節音情報 ....................................... 484
    8.4.1.4 語律情報 .......................................... 485
    8.4.1.5 節単位情報、係り受け構造情報、重要文情報、談話境界情報 .... 485
    8.4.1.6 話者情報 .......................................... 485
8.4.2 整合性検証の具体的内容 ................................................................. 486
8.4.2.1 転記情報と分節音情報の整合性検証 ................................................. 486
8.4.2.2 タグ<ce>箇所分節音ラベル修正 .................................................. 488
8.4.2.3 短単位 (転記) 情報・分節音ラベルの整合性検証 ............................... 489
8.4.2.4 転記タグ (?) の検証 ................................................................. 494
8.4.2.5 転記 (短単位)・分節音ラベル・顔律ラベルの整合性検証 ..................... 494
8.4.2.6 顔律情報内の整合性検証 ........................................................... 498

8.5 XML 文書の利用 ............................................................................. 499
8.5.1 XML 文書のブラウジング ............................................................. 499
8.5.1.1 XML 文書の所在 ................................................................. 499
8.5.1.2 Web ブラウザでのブラウジング .................................................. 499
8.5.2 XML 文書の整形 .................................................................... 499
8.5.2.1 XSLT と XPath ...................................................................... 500
8.5.2.2 xsltproc による XML 文書の整形 ............................................. 503
8.5.2.3 転記テキストの生成 ................................................................. 505
8.5.2.4 分節音ラベルの生成 ................................................................. 505
8.5.2.5 無声化母音周辺の情報抽出 ...................................................... 506
8.5.3 XML 文書に対する検索 ............................................................. 506
8.5.3.1 起動およびファイルの読み込み .................................................. 507
8.5.3.2 検索クエリ作成タブ ................................................................. 509
8.5.3.3 AND, OR 条件の指定 ............................................................... 511
8.5.3.4 結果 XML 表示タブ ................................................................. 516
8.5.3.5 表示整形用 XSL タブ .............................................................. 517
8.5.3.6 結果 HTML 表示タブ .............................................................. 517
8.5.3.7 IPU 可視化タブ ....................................................................... 519
8.5.3.8 各種ファイルの保存・読み込み .................................................. 519
8.5.4 XML ブラウザの利用例 ............................................................. 521
8.5.4.1 利用例 1.「ジャナイ」 .............................................................. 521
8.5.4.2 利用例 2.「タカイ」のアクセント位置 ....................................... 522
8.5.4.3 利用例 3. 言い誤り時の促音化「ッ」 ......................................... 523

第 9 章 CSJ の検索 ............................................................................. 527
9.1 GREP と正規表現 ...................................................................... 527
9.2 スクリプト言語 ........................................................................ 528
9.3 RDB と SQL ................................................................................ 532
9.4 XSLT プロセッサ ..................................................................... 537
9.5 検索に関わる注意 ....................................................................... 539
9.5.1 代表表記 ................................................................................ 539
9.5.2 短単位 ................................................................................... 540
9.5.3 活用形 ............................................. 540
9.5.4 形態論情報の誤解析 ....................................... 540
9.5.5 ピーンラベルの時間情報 ...................................... 541
9.6 まとめ ..................................................... 541

参考文献 ........................................... 543

索引 ........................................... 547