

国立国語研究所学術情報リポジトリ

目次

メタデータ	言語: Japanese 出版者: 研究社 公開日: 2026-07-24 キーワード: 作成者: メールアドレス: 所属:
URL	https://repository.ninjal.ac.jp/records/2000720

This work is licensed under a Creative Commons
Attribution-NonCommercial-ShareAlike 4.0
International License.



「シリーズ言語資源学」の刊行にあたって	iii
はじめに	v

第1章 | 会話コーパスの設計と構築 (小磯花絵) 1

1 会話コーパスの設計	1
1.1 Survey of English Usage	2
1.2 国立国語研究所における話し言葉の調査研究	3
1.3 British National Corpus	5
1.4 日本語日常会話コーパス	5
2 日本語の会話コーパス	6
2.1 現日研・職場談話コーパス	6
2.2 名大会話コーパス	8
2.3 BTSJ 1000人日本語自然会話コーパス	9
2.4 日本語話題別会話コーパス	10
3 『日本語日常会話コーパス』の構築	11
3.1 『日本語日常会話コーパス』構築の準備	11
3.2 『日本語日常会話コーパス』構築の流れ	19
4 『日本語日常会話コーパス』の量的概観	24
4.1 話者毎のデータ規模	24
4.2 話者の属性毎のデータ規模	25
4.3 話者間の関係性毎のデータ規模	26
4.4 会話の属性毎のデータ規模	30
5 おわりに	33

コラム TalkBank：国際的な会話データの宝庫 (小磯花絵)

第2章 | 会話の収録・音声動画データの加工 (臼田泰如) 37

1 音声の収録	37
1.1 録音機器の種類	37
1.2 録音機器の設定	38
1.3 実際の録音	39
2 動画の収録	42
2.1 録画機器の種類	42
2.2 録画機器の設定	43
2.3 実際の録画	44
3 メタ情報の収集	47
4 収録後の音声・動画データの加工	48
4.1 音声・動画の形式変換	49
4.2 音声・動画データの同期	50

4.3 非公開情報の処理	56
5 おわりに	61
コラム 収録の現場で起こるちょっとしたトラブル集 (小磯花絵)	

第3章 | 会話の転記 (臼田泰如) 64

1 発話の単位	64
1.1 話し言葉における単位の設定の必要性	64
1.2 既存の話し言葉の単位	65
1.3 CEJCにおける単位	67
2 表記	69
2.1 ローマ字表記	69
2.2 漢字仮名交じり表記	69
2.3 口語表現	70
3 転記用タグ	70
3.1 会話コーパスにおけるタグとは	70
3.2 CEJCの転記テキストに用いるタグ	71
4 CEJCにおける具体的な転記テキスト作成の手順	76
5 おわりに	79
コラム 埋もれた会話資料をコーパスとして蘇らせる (小磯花絵)	

第4章 | 談話行為情報 (居關友里子) 82

1 背景	82
1.1 談話行為情報の考え方	82
1.2 先行研究	83
2 『日本語日常会話コーパス』への談話行為情報付与	87
2.1 談話行為情報付与の目的と設計方針	87
2.2 談話行為情報の枠組み	87
2.3 付与作業	104
3 談話行為情報付与結果の概要	105
4 おわりに	108

第5章 | 「中納言」を活用した分析 (小磯花絵) 110

1 短単位情報・長単位情報の基本	110
1.1 単位の認定基準	110
1.2 階層的見出しの体系	111
1.3 短単位・長単位で提供する基本情報	111
1.4 品詞情報	112
2 「中納言」の基本的な機能・利用法	113
2.1 「中納言」とは	113
2.2 文字列検索	114

2.3	短単位検索・長単位検索	116
2.4	位置検索	119
3	中納言の活用事例	119
3.1	事例1	119
3.2	事例2	122
3.3	事例3	123
3.4	事例4	125
4	おわりに	127
コラム まとめて検索「KOTONOHA」の活用（小磯花絵）		

第6章 | ELANを活用した分析

（川端良子） 130

1	ELAN概要	130
1.1	マルチモーダルアノテーションツール	130
1.2	ELANのインストール	131
1.3	ELANで用いるファイルの種類	131
1.4	インタフェースの言語設定	132
2	ELANの基本操作	132
2.1	ファイルの新規作成・保存	132
2.2	画面各部の機能と用語の説明	133
2.3	動画・音声データの操作	135
2.4	注釈の作成・編集	136
2.5	注釈の検索	138
2.6	CSVファイル／TSVファイルの書き出し	139
2.7	管理語の活用	140
2.8	上位注釈層に対する下位層の設定	143
2.9	テンプレートファイルの利用	146
3	CEJCのELANファイルを用いた分析	147
3.1	ELANファイルの利用	147
3.2	コメント機能の利用	148
3.3	談話行為情報ファイルの読み込み	150
3.4	基礎統計	153
3.5	高度な検索	154
4	おわりに	160
コラム ELANの注釈を字幕ファイル形式で出力する（川端良子）		

第7章 | Praatを活用した分析

（西川賢哉） 162

1	CEJCにおける韻律情報	162
2	TextGridの活用	163
2.1	TextGridEditorの起動	165
2.2	表示と音声の再生	166
2.3	ラベルの修正・層の操作	167

2.4 F0 値の取得	168
3 Praat スクリプトの基礎	169
3.1 オブジェクトウインドウ	169
3.2 スクリプトの作成と実行の流れ	171
3.3 履歴機能	171
3.4 Praat スクリプトの仕様	173
3.5 算術演算子・算術関数	177
3.6 文字列演算子・文字列関数	178
3.7 条件分岐：if	178
3.8 ループ (1)：for	179
3.9 ループ (2)：while	180
3.10 ジャンプ：goto	180
4 Praat スクリプトの活用	181
4.1 活用事例 1：Pitch ファイルの生成	181
4.2 活用事例 2：有核語を含むアクセント句区間の F0 最大値	183
5 おわりに	193

第 8 章 | リレーショナルデータベースを活用した分析

(川端良子・小磯花絵) 195

1 リレーショナルデータベースとは	195
2 SQL の基本操作	197
2.1 SELECT 文の基本	197
2.2 列の選択 SELECT 句	197
2.3 行の抽出 WHERE 句	199
2.4 テーブルの結合 JOIN 句	201
2.5 グループ毎の集計 GROUP BY 句	204
2.6 グループ化したテーブルの条件抽出 HAVING 句	205
2.7 整列 ORDER BY 句	206
2.8 CASE 式	206
2.9 文字列関数	207
2.10 サブクエリー	209
2.11 テーブル・一時テーブルの作成	210
3 CEJC-RDB の活用	211
3.1 CEJC-RDB の基本構造	211
3.2 セグメントテーブル	213
3.3 親子関係テーブル	216
3.4 リンクテーブル	222
3.5 メタ情報テーブル	224
4 CEJC-RDB の操作環境	226
4.1 データベースの接続	227
4.2 データベース構造の確認	227
4.3 レコードの確認	228

4.4	SQLの作成と実行	228
4.5	結果の出力	229
5	おわりに	230
コラム	RからRDBに直接アクセス (小磯花絵)	
索引		233
編者・執筆者紹介		237

