

話し言葉の言語的・パラ言語的構造の解明に基づく『話し言葉工学』の構築

I 研究の全体計画

1. 目的, 意義, 必要性

本研究は自発的な「話し言葉」の情報処理技術の基盤を確立することを目的とする。

「話し言葉」情報処理技術には、インデックス付与による既存音声データの有効活用、福祉技術への応用、速記や文字起こしの自動化などの波及効果が期待される。

本研究では、通信総合研究所の有する自然言語処理技術と、国立国語研究所の有する言語学的知見を、研究総括責任者の有する音声情報処理に関する知見のもとに統合して研究を推進する。

2. 研究概要

サブテーマ1として、パラ言語情報（音声を文字化することによって欠落する情報）など、話し言葉固有の特徴を利用した「話し言葉工学」基礎技術の研究をおこなう。

サブテーマ2として、大規模な話し言葉コーパスを構築する。またその構築のために必要な付加情報の半自動付与技術の開発をおこなう。

サブテーマ3として、まとまった内容をもつ音声を認識して要約情報を出力する話し言葉要約システムのプロトタイプを開発する。

5年後の目標は、

- 「話し言葉工学」基礎技術の確立と言語学的知識の体系化
- 大規模話し言葉コーパスの構築とその効率化技術の確立
- 話し言葉要約システムの構築

である。

3. 研究総括責任者

古井 貞熙（東京工業大学 教授）

4. 融合研究機関

郵政省 通信総合研究所

文部省 文化庁 国立国語研究所

5. 研究期間

平成 11 年度～平成 15 年度

