

言語支援環境プロジェクトの概要

情報通信研究機構
自然言語グループリーダー

井佐原 均

言語支援環境プロジェクト

プロジェクトの目標

言語を主とした知識獲得、情報検索、質問応答、翻訳（支援）などの機能を持つ総合的な技術情報活用システムの開発

企業での実用化を視野に入れ、情報通信研究機構の持つ

豊富な言語資源

と

高度な言語処理技術

を利用し、企業の持つ

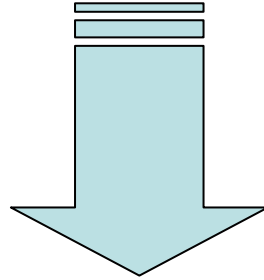
実世界のニーズ

と

システム化能力

を活用し、企業の研究者との共同で

言語支援環境の研究開発を行う。



その過程で 言語資源を中核にした

言語処理研究コミュニティ を構築することにより

実用に向けた / 広い応用が可能な
言語処理技術の開発

必要な言語資源の開発

先端的プロジェクトの共同提案

を行う。

自然言語グループ

自然言語処理の研究

文解析・生成
情報検索・抽出
要約・言い換え
機械翻訳
英語運用支援

学習に基づく処理

言語に関する基礎研究

語彙意味論
談話
感性情報処理(敬語・音楽)
意図の分析

理論基盤

ブレーク
スルー

研究ツール

学習データ
技術の実証

客観データ

言語資源

学習者コーパス
対訳コーパス
話し言葉コーパス
EDR電子化辞書

外部との協力

オープンラボ
連携大学院(神戸大学)

資源の公開

タイ自然言語ラボ

情報通信研究機構の提供する言語資源

解析技術、検索技術等の情報通信研究機構の持つ
高度自然言語処理技術およびソフトウェア

形態素解析システム

係り受け解析システム

検索システム

固有表現抽出システム

最大エントロピー法学習ソフトウェア

コーパスエディタ

言語データ

多言語情報付きコーパス(NICTコーパス)

(日中英対訳データ。形態素・係り受け情報付き)

日英の新聞記事の文対応、記事対応データ

話し言葉コーパス(700時間の日本語発話コーパス)

日本人英語学習者コーパス

(英語習得レベルの付与された300時間分の発話コーパス)

タイ語コーパスORCHID

CRL単文データ

固有表現データ

EDR電子化辞書

情報通信研究機構はEDR電子化辞書の全著作権を所有しています。

自然言語グループの応用技術

日英言語横断検索システム



(2名常駐)

沖電気

情報検索

(1名常駐)

TIS

ユーザインターフェース

言語処理技術
(学習に基づく
高精度解析技術)

言語資源
(世界最大級の
テキストデータ・辞書)

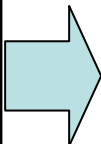


言語支援環境の構築

キーワードの自動
抽出・分類による
情報獲得支援

機械翻訳に用いる
翻訳知識の自動獲得

NICT
けいはんな



多言語データ作成

富士通

(1名常駐)

読解支援・
翻訳知識獲得

シャープ

(1名常駐、1名非常勤)

情報抽出

日本原研