

## 3.2.10 自然言語グループ

### 中期計画期間全体

#### 目 標

人間の知的活動を支援する環境の実現のため、人間のコミュニケーションの基幹をなす自然言語の処理と伝達の技術を開発する。中期目標期間終了時点では、言語情報を多言語かつ多様な形態で処理するシステムを開発する。

#### 目標を達成するための内容と方法

システム開発を視野に入れた学習に基づく自然言語処理の研究を中心に、ブレークスルーを目指す自然言語の基礎研究にも注力している。開放的融合研究を行うなど、他機関との協調、競争的資金の獲得にも努力している。

#### 特 徴

自然言語処理技術において、コンテスト等によって最高水準の成果を客観的に示している。また、第三言語翻訳などの新しい手法も提案している。自然言語の基礎研究を併せて行うことにより、他の研究機関にない幅の広い研究活動が可能となる。

### 今年度の計画及び報告

#### 今年度の計画

自然言語処理の研究開発を行う。また、その基盤となる自然言語の基礎研究を行う。これら成果の応用として、実用システムの開発と多言語処理の研究開発を行う。これらの研究の基盤となる言語資源の開発を行う。

自然言語処理の研究においては、言語解析技術の精度向上、検索等の応用技術の開発、研究用言語資源の整備を図る。自然言語の基礎研究においては、大規模データに基づく実用的な辞書の開発基盤を確立する。また、学習支援に関する研究を行う。実用システムの開発においては、オープンラボを活用し、外部機関の協力を得て、言語支援環境システムの開発を進める。このため、大規模な言語資源の開発を行う。多言語処理の研究開発においては、機械翻訳システムの研究を行う。また、中国語の研究開発を推進する。タイ自然言語ラボと協調し、アジア圏言語の処理システム及び言語資源の開発を行う。

#### 今年度の成果

自然言語処理の研究では、これまでの成果を実データに適用し、係り受け解析システムの精度を向上するとともに、応用システムとしての機械翻訳システムの開発を進めた。機械翻訳の評価法を開発し、評価コンテストを共催した。NICT コーパスの拡充を行った。自然言語の基礎研究においては、コーパス中の実例を用いて、これまでの理論を検証した。具体的には新しい距離尺度の導入により、語の相互関係に加えて、階層関係をも自動抽出している。また、学習支援システムとして、英語誤り検出システムを開発した。実用システムの開発においては、オープンラボにおいて、機械翻訳、読解支援、情報抽出支援、コーパス構築等の言語支援環境システムの開発を進めた。さらに、英語学習者コーパスを研究用に公開した。また、一般書籍としても公開した。昨年度で終了した話し言葉工学プロジェクトの成果として、国立国語研究所と共同で、日本語話し言葉コーパスを公開した。

タイ自然言語ラボとの関係を深めるため、交流会を開催した。また、日中交流会議を開催した。中国科学院に協力し、機械翻訳システムの評価コンテストを開催した。



英語学習者コーパスは一般書籍として、公開された。また、このコーパスを用いて、英語学習支援システムを開発した

