

### 3.2.10 情報通信部門 自然言語グループ

グループリーダー 井佐原均 ほか16名

#### 自然言語処理の研究開発

#### 概要

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過程を研究し、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。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、ブレイクスルーを目指す自然言語の基礎研究にも注力した。他機関との協調、競争的資金の獲得にも努力している。

#### 平成17年度の成果

従来に引き続き、自然言語処理の研究を中心に、世界最大級の研究用言語資源の開発、言語に関する基盤的研究の推進、実用的な技術開発を一元的に推進することにより、国際会議41件、誌上論文11件、特許取得10件、申請14件という研究成果を上げた。この成果に基づくシステムは情報検索コンテストにおいて、1位2件、2位2件、3位1件の成績を上げた。さらに、オープンラボを活用した企業との共同研究により、実システムに組み込み可能な技術を開発した。

自然言語処理の研究においては、基盤となる要素技術を開発・統合し、機械翻訳支援及びその他の応用システムの開発を行った。学習を用いた解析システムでは世界最高水準にある。医療用機械翻訳システムの精度向上・分野拡張(ジュネーブ大学と共同開発)、機械翻訳システム評価法の検討、話し言葉コーパスを用いた話し言葉の分割及び解析処理の研究を行った。用例翻訳・統計翻訳を融合するシステムの開発に向けて、研究用プラットフォームの構築を行った。

自然言語の基礎研究においては、大規模データに基づく客観的な語彙知識の獲得とその可視化を行った。当該分野の国際会議(GL2005)を開催した。また、敬語誤用検出システムを開発した。意図の自動抽出、英語の誤り検出の研究を行い、システム化した。

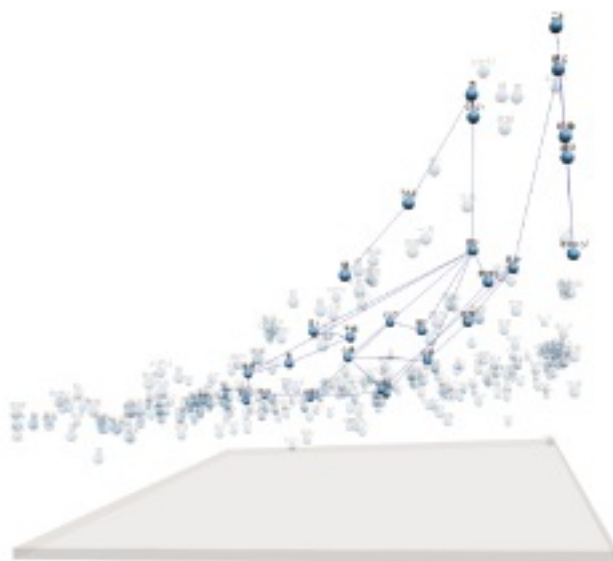
実用システムの開発においては、オープンラボを活用し、商用システムへの搭載技術を開発した。その成果を報道発表し、7紙に掲載された。

言語資源の開発においては、NICT多言語コーパスの拡充と、学習者コーパス・EDR電子化辞書の改良を進めた。日英対訳データを活用するホームページ(LinguaPark)を開発・公開した。配布している言語資源は順調に売り上げを伸ばしている。

多言語処理においては、タイ自然言語ラボへの技術移転を行った。ISOの標準化活動に参加し、日本語・中国語・韓国語の単語分割の標準化に関する検討を行った。さらに、日中自然言語処理共同研究促進会議の開催、中国語言語資源(辞書及びコーパス)の構築などを通して、中国との関係を強化した。翻訳支援ツールを開発し、京都大学と共同でそれを用いた多言語翻訳実験(日英⇒タイ、韓国、中国、マレーシア)を実施した。



LinguaPark



語彙知識の可視化