

## 3.5.1 知識創成コミュニケーション研究センター 自然言語グループ

グループリーダー 井佐原 均 ほか39名

## 言語資源と多言語情報処理の研究開発

## 概要

自然言語グループは、ナチュラルコミュニケーション技術の開発の一環として、言語情報プロジェクト、言語グリッドプロジェクト、タイ自然言語ラボラトリー（TCL: Thai Computational Linguistics Laboratory）の3プロジェクト体制で、言語障壁を起因とするデジタルデバイドの解消を目指し、今後の言語情報処理の基盤となる大規模な言語資源の作成・公開を中心に、その作成・活用に資する言語処理技術や応用システム、それらを統合しサービスとして実現する言語グリッドの開発を行っている。

## 平成19年度の成果

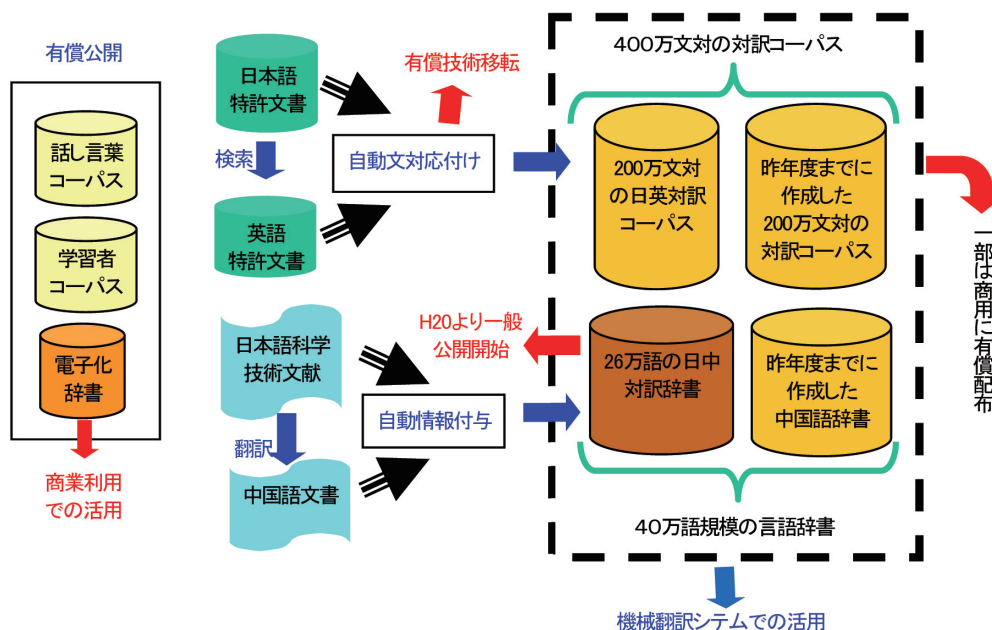
## (1) 言語情報プロジェクト

## ① 言語資源の開発

日英対訳データの収集、日中対訳データの収集及び日本語データの中国語翻訳を実施した。また、対訳データ間の文対応技術、対訳文間の構成要素の対応付与技術の開発を実施した。

新規に200万文対の対訳コーパスを開発し、既存のデータと合わせて400万文規模の対訳コーパスとした。また、EDR日英辞書の中国語への拡張を行った。これは平成20年度に一般公開される。日本語ワードネットの開発を進め、その第1版を完成させた。これも平成20年度に一般公開する。

対訳データの自動対応付けツールの性能向上を図り、有償での技術移転を行っている。



## ② 機械翻訳システムの開発

大規模な言語資源に基づく実用的な自然言語処理技術応用システムとして、日中機械翻訳システムの開発を進めている。システムは言語の構造をより深く利用する汎用的な用例翻訳手法に基づいており、開発した大規模言語資源を利用して、高性能の翻訳の実現を目指す。要素技術としては中国語の解析技術の性能向上を図った。翻訳システムの基盤となる部分は言語や対象分野に依存しないので、北京の観光情報の翻訳への展開や、以下で述べる日タイ交流授業などへも利用された。

## (2) 言語グリッドの研究開発

言語グリッドプロジェクトでは、言語の壁の克服に向けて、インターネット上の言語資源を連携させ多言語サービスとして提供する「言語グリッド」の開発及びそれを利用した異文化コラボレーションツールの研究開発を行っている。本研究開発の成果として、平成20年度の言語グリッドの実用化を目指している。平成19年度の研究開発では、言語グリッド基盤の管理ツールであるサービスマネージャを開発した。これにより、言語資源提供者による、言語グリッド上の言語資源のアクセス管理やモニタリングなどの資源管理が可能になった。このサービスマネージャを用いることで、京都大学による、非営利利用を対象とした言語グリッドの試行的な運営が2007年12月より開始され、国内外の大学、研究機関など約50団体が参加して言語グリッドの利用が始まっている。

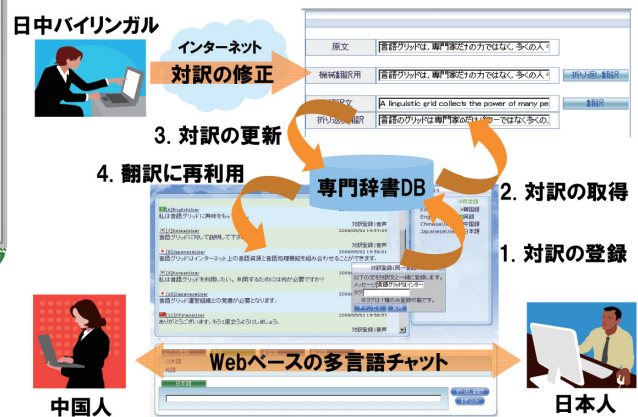
一方、異文化コラボレーションツールの研究開発では、国際交流活動を可能とするWebベースの多言語チャットツールを開発し、さらに多言語チャットのログを辞書として蓄積するスパイラル型辞書構築機能の開発も行った。これにより辞書データを増大させることで、翻訳精度の向上を可能にする。また基礎研究として、複数の機械翻訳を結合する際に生じる訳語のドリフトを防止するために、訳語選択情報を文脈とした機械翻訳連携技術の研究を行った。この技術は、国内特許に出願中であり、国際特許にも出願予定である。言語グリッドプロジェクトのメンバーが参加する論文発表は、平成19年度はジャーナル6本、国際会議12本、また報道発表は6件に上り、活動の広がりを示している。

IP Address	Data Transfer Size (Bytes)
130.54.21.141	2615
130.54.21.141	2547
130.54.21.141	2547
130.54.21.141	1475
130.54.21.141	644
130.54.130.227	1662

特定ユーザの利用ログ

## 利用状況の一覧

言語グリッドサービスマネージャの画面



スパイラル型辞書構築機能の利用例

## (3) タイ自然言語ラボラトリー

設立後5年を迎え、独自の技術開発・資源開発を進めるとともに、言語情報処理研究の成果の実証の場として活動した。また、東南アジア地区への技術移転に積極的に取り組んだ。

インターネット上での共同作業を支援するツールKUI (Knowledge Unifying Initiator)は、タイ科学技術省の公式ホームページで活用されている。また、遠隔授業システムと用例翻訳システムを組み合わせ、タイの高校と日本の東宇治高校を結んでの交流授業を実施した。自然言語処理技術に関するスクールは今年度も開催し、多くの参加者を得た。



日タイ遠隔交流授業



交流授業を伝える新聞記事