

言語運用を基盤とする言語情報学拠点

Center of Usage-Based Linguistic Informatics (UBLI)

川口 裕司

(COE 拠点リーダー)

言語情報学

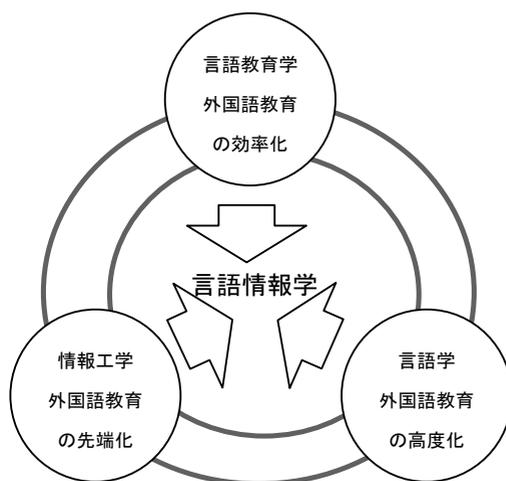
情報工学の基盤の上に言語学と言語教育学を有機的に統合し、「言語情報学」という新たな学問分野の世界的な研究拠点を創成することが、このプロジェクトの目的です。ボーダレスな多言語時代に入った現在、言語教育においても情報技術に裏づけされた多言語 e-learning システムを構築し、高度で効率的な教育を行なうことが望まれます。

研究組織

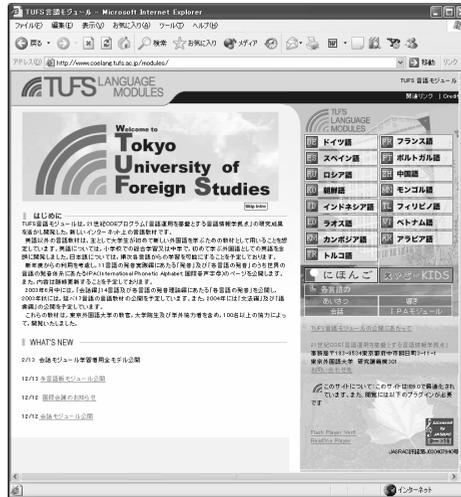
この COE 計画では、4つの研究班が組織され、各班は緊密な連携をとりつつ研究が進められています。

言語情報学班は言語情報学の研究全体を統括し、COE 計画において中心的な役割を担う研究班です。この班のもとに言語学班、言語教育学班、情報工学班の3つの班が形成され、個々の専門分野における研究のほかに、言語情報学的研究を推進するための基礎研究を行なっています。

冒頭にも述べたように、この COE 計画によって創成される新たな学問分野である言語情報学は、情報工学の基盤の上に言語学と言語教育学を統合した学問分野です。言語情報学の最も目にみえる成果は、17の言語を対象とするインターネット上の言語学習システム、「TUFUS 言語モジュール」です。



TUFS 言語モジュール (TUFS Language Modules)



<http://www.coelang.tufs.ac.jp/modules/index.html>

TUFS 言語モジュールの中で、まず最初に公開されたのは、IPA (International Phonetic Alphabet, 国際音声字母) モジュールです。続いて発音モジュールが 2003 年度に英語, ドイツ語, フランス語を始めとする 12 言語で公開されました。2004 年度には会話モジュールが, 英語, ドイツ語, フランス語, スペイン語, ポルトガル語, ロシア語, 中国語, 朝鮮語, モンゴル語, インドネシア語, フィリピン語, ラオス語, カンボジア語, ベトナム語, アラビア語, トルコ語, 日本語の 17 言語全てにおいて公開され, 一部は授業の中で利用されています。さらに 2005 年度には文法モジュールがドイツ語, フランス語, ロシア語, 中国語, フィリピン語, ベトナム語, トルコ語, 日本語の 8 言語で公開されます。

モジュール性, 通言語性, 多言語性

TUFS 言語モジュールは今までにない新しいタイプのウェブ教材です。その名のとおり, 「モジュール的発想」に基づいて作られています。言語学習は発音, 会話, 文法, 語彙の 4 つのモジュールに分解され, それぞれのモジュールはある程度まで互いに独立しながら, 全体として一つのまとまりをもった言語学習教材を構成すると考えます。もちろん言語学習が一つの統合された一体性のある営みであることは言うまでもありません。また個別言語の垣根を少しでも低くして言語一般を見つめる眼差しをもつことができるように, 「通言語的発想」に基づいています。通言語性はとくに会話モジュールの機能シラバスや 2005 年 7 月に公開された通言語文法モジュールに明確に現れていると言えます。また現在, 設計が進んでいる語彙モジュールでも通言語的発想が生かされる予定です。

ところで TUFS 言語モジュールは学習者に, より自由な学習環境を提供するわけですが, 学習の達成度や到達度を知るには, やはり何らかの統一的な物差しが必要です。この COE

計画では、TUFS 言語モジュールを用いた東京外国語大学独自の言語能力記述モデルを追求し、将来的に一つの統一したモデルを提案する予定です。17の言語にわたって、ある程度まで共通に言語能力を記述することが可能になれば、中等・高等教育における言語教育に一時代を画することになるかもしれません。

一つの同じ教材をいろいろな言語で学ぶことができたらどうでしょうか。それを実現しようというのが TUFS 言語モジュールの多言語版です。2005 年度に英語、フランス語、中国語（簡体字・繁体字）、モンゴル語、トルコ語のわかる人が、日本語の発音モジュールと会話モジュールを学習できるようになります。今後もさらにモジュールの多言語化を推し進めていきます。

言語コーパスと言語運用

言語研究には作例やインフォーマントの直観を中心とした研究、あるいは書記資料や口語資料に基づいた研究などがあります。言語教育においても、実際の会話に似せて作られた会話や特定の文型や表現を中心とした教材が用いられたり、あるいは実際の会話を利用した教材などがあります。この COE 拠点は、そうした言語研究や学習教材の中で用いられている広い意味での「言語資料（コーパス）」のあり方を問い直そうとします。言語研究者は、言語コーパスの中から言語理論や論拠立てを説明するのに都合がよい言語資料だけを引き出して提示するだけでよいのでしょうか。そうではなく言語が実際に用いられる状態、すなわち「言語運用」を知ることが真の目的であるとすれば、言語コーパスからどのようにして言語運用を引き出すことができるのでしょうか。それぞれの言語資料には個別の特徴があるはずですが、はたして巨大な言語コーパスを構築し、それを分析すれば言語運用の姿が明らかになるのでしょうか。逆に多様性や複雑な類型が立ち現れるのではないのでしょうか。こうした様々な疑問に答えるべく、この COE では既存の言語コーパスを分析したり、自ら構築した自然談話コーパスを研究し、その成果を教材にどのように応用することができるかを考えます。特に外国語学習においては、学習者が当該の言語をどのように生成しているのか、いわゆる学習者言語コーパスの分析も重要な意味をもっています。

拠点では 2002 年よりコーパス言語学、自然談話分析、学習者言語コーパスの研究を継続しており、すでに『言語情報学研究報告 3 コーパス言語学における構文分析』、『言語情報学研究報告 5 第二言語の教育・評価・習得』、『言語情報学研究報告 6 自然会話分析と会話教育』を出版しました。

第 1 回言語情報学国際会議

21 世紀 COE が採択された 2002 年末に国際会議の輪郭を決定しました。こうして 2003 年 12 月 13・14 日の両日に、東京外国語大学で第 1 回言語情報学国際会議（The First International Conference On Linguistic Informatics）が開催されました。

従来から言語理論とコンピュータ科学が言語教育や言語習得に影響を与えてきたことは

周知のことです。しかしながら、これら三つの研究領域の統合と協働は必ずしも行われてきていません。この COE 計画が目指すのはそれらの学問領域の有機的な統合です。

言語情報学により、従来の言語学と応用言語学の成果は情報工学の基盤の上に統合されます。この国際会議では言語情報学という新しい統合的学問領域の現状を認識し、将来の可能性を考えました。会議は三つのセッションに分かれました。

1. コンピュータ言語学・・・コンピュータ科学と言語学の協働の可能性
2. コーパス言語学・・・コーパス言語学の現状
3. 応用言語学・・・第二言語習得と言語理論との関連性

海外や国内の研究機関より多数の研究者を招き、2 日間で延べ 300 名の出席者があり、活発な議論が交わされました。本学からも教員と大学院生が複数の報告を行いました。言語情報学という統合的学問分野は、コンピュータ言語学、文献学、方言学、コーパス言語学、語用論、応用言語学、e-learning などの多岐にわたる研究分野と関連します。そのため会議の参加者がそれぞれの報告や議論の内容を理解できるように、事前に予稿集を出版して会議に臨みました。この会議報告集を通して、言語情報学の裾野の広さを認識するとともに、その現状と課題が明らかになりました。

第 1 回言語情報学国際会議の報告集とコーパス言語学の成果は、オランダの John Benjamins 社からシリーズ『言語運用を基盤とする言語情報学 Usage-Based Linguistic Informatics』の第 1 巻・第 2 巻として 2005 年春に出版されました。

