『言語情報学研究報告』No.14（2006）

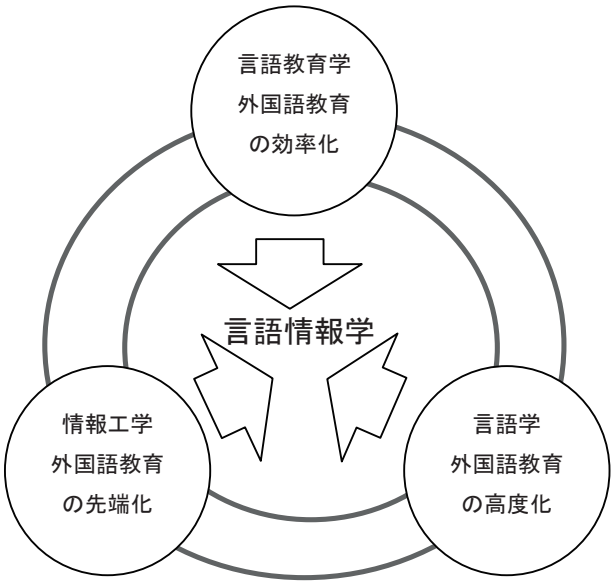
言語運用を基盤とする言語情報学拠点

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，とくにコンピュータ技術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学問分野である言語情報学は，コンピュータ技術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，「TUFS 言語モジュール」です。

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，現在では 14 言語が完成しました。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，フィリピノ語，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 言語全てで公開され，一部は授業の中で利用されています。さらに 2006 年 8 月現在において，文法モジュールはドイツ語，フランス語，スペイン語，ロシア語，中国語，モンゴル語，フィリピノ語，ベトナム語，カンボジア語，トルコ語，日本語の 11 言語で公開され，語彙モジュールも英語，ドイツ語，フランス語，スペイン語，ロシア語，中国語，朝鮮語，フィリピノ語，ベトナム語，トルコ語，日本語の 11 言語が公開されています。

TUFS 言語モジュール（TUFS Language Modules）

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

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。また語彙モジュールは日本語の基本語彙を中心にしつつ，特有な語彙を追加するやり方で，ある程度の通言語性を確保しています。

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，学習の言語能力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ます。すでにヨーロッパ共通参照枠組をモデルにした言語能力記述の妥当性に関する大規模アンケートを実施し，相対的にポジティブな結果が得られ，この枠組を用いた言語能力レベルと会話モジュールの関連づけが行われています。将来的に 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。2006 年度中に英語，フランス語，中国語（簡体字・繁体字)，朝鮮語，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。

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に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。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でしょうか。こうした様々な疑問に答えるべく，この COE 拠点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，フィールド調査,音声研究,自然談話分析，第二言語習得，自然言語処理・e-learningの研究を継続しました。これまで拠点では15冊の研究報告集を公刊しました。コーパス言語学の成果としては,『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 7　コーパス言語学における語彙と文法』,『言語情報学研究報告11　言語研究におけるコーパス分析と理論の接点』,『言語情報学研究報告12　コーパス言語学の諸相―話し言葉コーパスと書き言葉コーパス―』があります。フィルード調査と音声研究の成果には,『言語情報学研究報告8　フィールド調査による口語資料の収集及びその分析』,『言語情報学研究報告 4　通言語音声研究―音声概説・韻律分析』があります。自然談話分析の成果は『言語情報学研究報告 6　自然会話分析と会話教育』,『言語情報学研究報告 13　自然会話分析への言語社会心理学的アプローチ』があり，第二言語習得研究は『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10　教材開発・評価・第二言語習得』,『言語情報学研究報告14　第二言語習得理論に基づく言語教育と評価モデル』にまとめられています。またTUFS言語モジュールとe-learningおよび自然言語処理に関しては,『言語情報学研究報告1　TUFS言語モジュール』,『言語情報学研究報告 2　言語学・応用言語学・情報工学』,『言語情報学研究報告15　TUFS言語モジュール教材・ことばの定量分析』を出版しています。

国際会議・ワークショップ・国内会議

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。

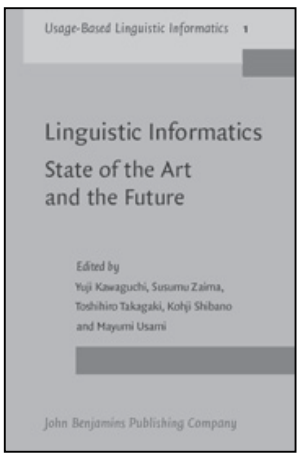
言語情報学により，従来の言語学と応用言語学の成果は情報工学の基盤の上に統合されます。この国際会議では言語情報学という新しい統合的学問領域の現状を認識し，将来の可能性を考えました。会議は三つのセッションに分かれました。

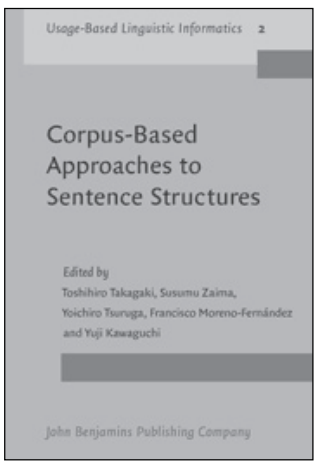
1. コンピュータ言語学・・・・コンピュータ科学と言語学の協働の可能性
2. コーパス言語学・・・・・・コーパス言語学の現状
3. 応用言語学・・・・・・・・第二言語習得と言語理論との関連性

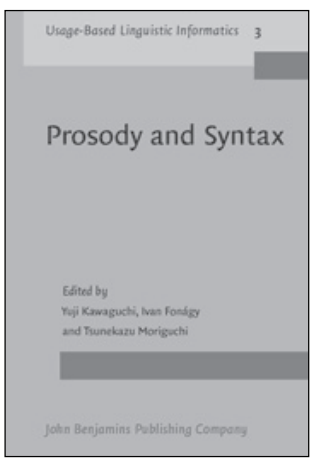
海外や国内の研究機関より多数の研究者を招き，2日間で延べ300名の出席者があり，活発な議論が交わされました。本学からも教員と大学院生が複数の報告を行いました。言語情報学という統合的学問分野は，コンピュータ言語学，文献学，方言学，コーパス言語学，語用論，応用言語学，e-learningなどの多岐にわたる研究分野と関連します。そのた

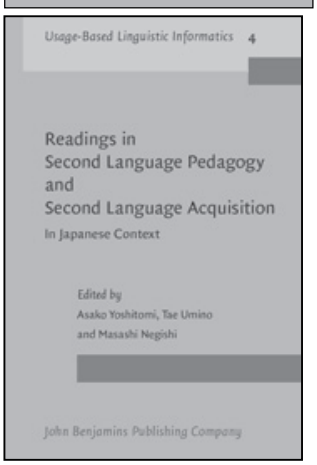
め会議の参加者がそれぞれの報告や議論の内容を理解できるように，事前に予稿集を出版して会議に臨みました。この会議報告集を通して，言語情報学の裾野の広さを認識するとともに，その現状と課題が明らかになりました。

第 1 回言語情報学国際会議の報告集 Linguistic Informatics State of the Art and the Future，コーパス言語学の論集 Corpus-Based Approaches to Sentence Structures ,プロソディーと統語論に関する論集 Prosody and Syntax，言語教育と第二言語習得の論集 Readings in Second Language Pedagogy and Second Language Acquisition はオランダの John Benjamins 社から本拠点の名を冠したシリーズ『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・第 3 巻・第 4 巻として 2005 年から 2006 年にかけて出版されました。









拠点ではその後も継続的に会議を開催しています。2005 年 1 月 27 日に東京外国語大学の 2 つの 21 世紀 COE プログラム主催による総合シンポジウム「グローバリゼーションと多文化的想像力」が開催されました。この中の「インテグラルな地域文化研究拠点の創生

に向けて」と題する個別シンポジウムのセッションAにおいて,「言語理論と言語教育の統合は可能か」と題するシンポジウムを行いました。本学から3名の教員に出席いただいて,一般言語学，応用言語学，類型論の立場から，言語理論と言語教育の有機的な連携の可能性について議論しました。このシンポジウムの内容は,『言語情報学研究報告9 シンポジウム，講演会，研究報告』に掲載されています。

21世紀COEプログラムの目的の一つに，各分野における高度な研究能力を有する若手研究者の育成があります。2005年1月28日に,「グローバル化の現実と文化の基層をみつめる人文社会研究」と題する若手研究者による研究報告会が行われました。6 名の研究発表があり，質疑応答が行われました。そのうちの5編が『言語情報学研究報告9 シンポジウム，講演会，研究報告』に掲載されました。

2005年10月4•5日の両日，東京外国語大学で「言語情報学国内会議　言語情報学とは何か」を開催しました。10月4日は，招待報告とCOE協力者の院生諸氏による研究発表を行いました。10月5日は，言語学と教育工学に関する特別講演の後に，21世紀COEの推進メンバー諸氏による,「言語情報学とは何か－言語学・応用言語学・情報工学の言語情報学への寄与－」と題する研究報告会と自由討論が行われました。

言語情報学拠点では，2002年のCOE採択当初から，日本語の話しことばコーパスの構築を推進してきました。2004年度からは，教員と院生を海外に2週間程度派遣し，現地で話しことばの録音を行い，そのデータを文字化し,「話ことばコーパス」を構築しています。これまでスペイン語，マレーシア語，ロシア語，フランス語，トルコ語，イタリア語（サレント方言),カナダにおける英語と日本語のバイリンガリズムについて自然談話の収録が行われ，すでに複数の言語で「話ことばコーパス」が構築されました。2005年度には，新たにC-ORAL-ROMのメンバーとの協力関係も確立することができました。C-ORAL-ROMはイタリアのフィレンツェ大学，スペインのマドリード自治大学，フランスのエックス・マルセイユ大学，ポルトガルのリスボン大学の4大学による,「ロマンス諸語の話ことばコーパス」に関する共同プロジェクトです。

2005年12月9日，話ことばコーパス研究に関するワークショップをC-ORAL-ROMとの共催で開催しました。翌10日には,「第2回言語情報学国際会議」を開催し，話ことばコーパス研究の意義,コーパスの談話分析,言語コーパスの言語教育への応用といったテーマを議論し，言語情報学の理論的基盤を明確にしました。ワークショップと国際会議には延べ300名以上の出席がありました。こうした2度の国際会議とワークショップを通して，コンピュータ技術を利用し，言語理論を教育実践へ応用しようとする言語情報学拠点の研究対象はさらに明確になりました。すなわち言語情報学の最も重要な学術的貢献は，おそらくコンピュータ科学とコーパス言語学の理論に基づき，書きことばや話ことばのコーパスを用いて実際の言語運用にみられる様々な言語変異や談話機能を分析することです。また同時に学習者言語コーパスを構築し，分析し，その成果を教育実践の中に組み込むことで，より効率的で高度な言語教育を実現できるようにすることが言語情報学の目標です。ここで言う教育実践とは，とりも直さず本拠点が開発したTUFS言語モジュールを利用した授業やネットワークによるTUFS言語モジュールを用いた自律的学習のことです。

2006 年度は COE プロジェクトの最終年度です。上記の目標に少しでも近づくため，9 月 14 日に「コーパス言語学―その研究領域―」と題するワークショップを開催しました。イギリス，フランス，ドイツ，ノルウェー，カナダ，アメリカ合衆国より研究者を招聘し，コーパス言語学の研究領域を考えるとともに，その様々な利用法と実践に関する研究報告を聴き，研究成果の言語教育への応用の可能性を探りました。翌 15 日には，海外から招聘した 10 名の研究者に本研究拠点の 5 年間の研究成果を報告し，できるだけ客観的な視点から本研究の評価をお願いし，様々な質疑応答がなされました。

5年間という短期間で新たな学問領域が十全なかたちで創成されるとは思えません。言語情報学とは本プロジェクトの大きな野望の別名なのであって，私たちは今，その第 1 ステージの終盤にようやくさしかかったのです。