

東京外国語大学大学院21世紀COEプログラム
言語運用を基盤とする言語情報学拠点
研究成果 (2002.10-2007.3)



東京外国語大学大学院地域文化研究科
Graduate School of Area and Culture Studies
Tokyo University of Foreign Studies

東京外国語大学大学院 21世紀COEプログラム
言語運用を基盤とする言語情報学拠点
— 研究成果 (2002.10–2007.3) —



目 次

項目またはページ番号をクリックで各ページへ移動

学長からのメッセージ	4
1. 言語運用を基盤とする言語情報学拠点	6
2. 研究組織と研究成果	8
2.1 言語学	9
2.2 言語教育学	13
2.3 情報工学	17
2.4 言語情報学	19
2.5 言語学と情報工学の連携	22
2.6 言語教育学と情報工学の連携	24
3. 若手研究者育成のための教育プログラム	26
4. COE 研究成果の外部評価	30
5. プロジェクト関係者一覧	34
言語情報学拠点の継承と発展	38

学長からのメッセージ

池端雪浦

(東京外国語大学学長)

2002年度から開始された文部科学省の「21世紀COEプログラム」は、我が国の大学に、世界最高水準の研究教育拠点（Center of Excellence）を学問分野毎に形成し、研究水準のいっそうの向上と世界をリードする創造的な人材の育成をめざしています。本学は、「人文科学」と「学際・複合・新領域」の2つの学問分野にそれぞれ1件の申請を行い、人文科学では「言語運用を基盤とする言語情報学拠点」、学際・複合・新領域では「史資料ハブ地域文化研究拠点」が採択されるというすばらしい結果をえました。本学大学院地域文化研究科の、個性ある研究教育のポテンシャルが高く評価されたことを嬉しく思います。これら二つの拠点は、言語研究と地域文化研究における世界的な教育研究拠点を目指そうとする本学の将来構想の主要な推進力・両輪であると考えられています。

この2つの拠点を全学的見地から支援するために、学長直属の「21世紀COEプログラム運営室」を設置しました。この運営室は、学長、副学長、研究科長、拠点リーダーをはじめ、大学院を支える学部ならびにアジア・アフリカ言語文化研究所の長、さらに事務局長以下事務局幹部から構成され、部局横断でかつ事務局・教員が文字通り一体となった組織です。運営室は、拠点の支援のために学内諸組織間の連携体制を構築するとともに、総計300平米におよぶスペースの提供や在外調査研究旅費などの学内予算措置をはじめとする支援を行ってきました。

「言語運用を基盤とする言語情報学拠点」は、近年めざましい発展を遂げた情報工学と言語学ならびに言語教育学の3専門分野を有機的に統合して、言語情報学という新しい学問領域を創成することです。それによって、膨大な言語運用データの集積とその分析に基づいて、言語運用の実態を解明する言語研究の新分野を切り開き、その成果を言語教育学へ応用していくことがめざされています。また大学院生の教育面では、インターネット上の言語資源情報の収集・評価・分類や、TUFSS言語モジュールの開発、多言語コーパスの解析・処理などに大学院生を参加させることによって、言語資源情報について深い理解をもち、音声学・言語学の知識やコンピュータ技術に精通した人材を育成すること、また、現地での言語フィールド調査に当たらせることで、フィールドワーカーとしての訓練を行っています。

「言語運用を基盤とする言語情報学拠点」は、本学がかつて経験したことのない、いくつかの新しい挑戦をおこなってきました。その1つは、先に言及した言語学、言語教育学そして情報工学の3専門分野が協働して言語情報学という新しい学問領域の創成に取り組んだことです。2つ

は、本プログラムは実に多くの教員、研究者、大学院生、ポストドク研究者らの参加と協力をえて、進められてきたことです。プログラム事業への協力者は学内のみならず、海外の研究者をはじめとして多くの学外研究



者を含んでいました。3つは、国際シンポジウムやワークショップ等を通して、あるいはまた大学院生やポストドク研究者の海外での学会発表等を通して、本学の研究活動の国際化が大きく前進したことです。本プログラムの研究成果がオランダのJohn Benjamins社から「言語運用を基盤とする言語情報学 (Usage-Based Linguistic Informatics)」の叢書として出版されていることも、その一環と言えるでしょう。

4つは、本プログラムの中心的活動の1つであるTUFUS言語モジュールの開発では、研究成果がただちにweb上に公開され、学外の外国語教育者のオープンな外部評価を受けて、改善されるという手法がとられてきたことです。そして5つは、大学院生の教育において、「研究活動を通じた教育」が、これまで以上に積極的にとりくまれるようになったことです。

こうした新しい挑戦によって、言語情報学拠点のプログラムは研究・教育の両面において、多くのすぐれた成果をあげてきたと私は考えております。これらの成果は、リーダーである川口裕司教授の卓越した献身的リーダーシップと、プログラムの中心的担い手である本学教員、大学院生ならびにポストドク研究者らの骨身を惜しまぬ協力によるものであり、本学はそれを誇りに思っております。

言語情報学拠点プログラムは2007年3月で、5年間のファーストステージを終了しますが、本学としては、その後もプログラムの活動を維持・発展させていくために、研究資金を含めて支援を続けたいと考えています。言語情報学拠点のこれまでの研究成果に対する率直な評価と同時に、今後の研究活動の方向性についても、ご助言・ご意見をいただければ幸いです。

2007年2月1日

1. 言語運用を基盤とする言語情報学拠点

我が国の大学が、世界のトップレベルの大学と伍していくためには、競争的環境を一層醸成し、国公私立大学を通じた大学間の競い合いがより活発に行われることが重要です。この一環として、第三者評価に基づく競争原理により、世界的な研究教育拠点の形成を重点的に支援し、国際競争力のある世界最高水準の大学づくりを推進するため、平成14年度より、「21世紀COEプログラム」を実施しています。（文部科学省の文書より）

「言語運用を基盤とする言語情報学拠点」は2002年（平成14年）度に人文科学の分野で採択された21世紀COE拠点です。

拠点の目的と必要性

言語学と言語教育学と情報工学の連関性は以前から意識されてきました。しかし三つの分野の協働作業によって新たな学問的成果があがっているとは必ずしも言えません。この拠点ではコンピュータ科学の基盤の上で言語学と言語教育学の有機的な統合をはかり、そうした研究分野を言語情報学と名づけました。

近年、アメリカ合衆国をはじめとしインターネットを利用した外国語教育が急速に台頭しつつありますが、それらの多くは英語・フランス語・ドイツ語等の教育であり、多言語を目的としたシステムは多くありません。この拠点では16の外国語と日本語による多言語教育システムを開発し、それを世界に向けて発信しました。本学大学院は、これまで数多くの外国語研究者と教育者を輩出してきており、わが国の外国語教育に対して果たす責任と任務は大きいと言わねばなりません。



<http://www.coelang.tufs.ac.jp/>



拠点の意義・波及効果

私たちがこれまで行ってきた外国語研究では単に言語理論を追求するだけでなく、本学における言語教育の実践を通して言語理論を見つめなおす眼差しの重要性が説かれてきました。こうした言語研究と言語教育の双方向的なフィードバックは、他の大学には見られない東京外国語大学の個性であり、独自の学問的特色とすることができるでしょう。

言語情報学拠点からは次の三つの研究分野における研究成果が期待されます。

①言語教育工学および外国語教育学の研究成果

この拠点では、世界的にも類例のない多言語によるウェブ学習教材を開発します。多言語をベースとするe-learningはまだまだ発展途上にあり、この拠点がその開発に着手することの意義は大きいでしょう。たとえばアジアの諸言語（ベトナム語、ラオス語、カンボジア語、フィリピン語、モンゴル語等）を対象とするe-learningは、他大学の追随を許さない東京外国語大学独自の学術的研究と言えます。また全世界的に見ても、英語以外の言語における第二言語習得研究は数が少なく、この点からも言語情報学拠点の学術的意義は極めて大きいと言えるでしょう。

②コーパス言語学および自然言語処理の研究成果

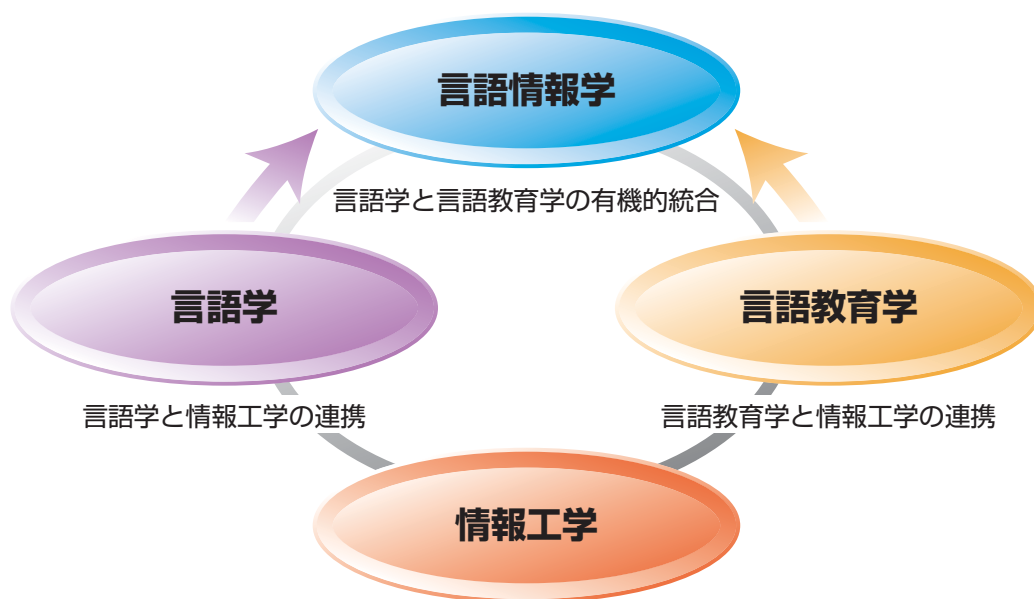
ウェブ教材の開発とともに、言語機能を重視した機能別の多言語コーパスや研究分野の目的に合わせたに目的別の多言語コーパス（言語学、音声学、文化研究等）を構築します。拠点では世界の様々な言語の膨大な言語運用データを集積し、それらにメタ言語情報を付与するなど、情報工学の知見を活用しながら、その成果を実際に言語教育に応用するよう努めます。

③言語情報学による統合的研究成果

従来の言語研究や言語教育では、言語運用の実態を科学的に分析することの必要性が必ずしも認識されてきませんでした。言語学はこれまで理論研究と内部構造の記述に終始し、実際に使用される談話やテキストを重点的に研究し始めたのはごく最近のことに過ぎません。とはいえ現実の言語運用に立脚した言語理論の構築とその言語教育への応用は、今日の言語研究の最も重要な研究課題の一つであると思います。この21世紀COE拠点では、情報工学を基盤としつつ、言語情報学の名のもとに言語学と言語教育学が統合され、その協働研究の中から「TUFUS言語モジュール」と呼ばれる多言語によるウェブ教材が開発されます。言語情報学の創成によって、この拠点からは、言語理論と情報技術の十全な知識を兼ね備えつつ、現実の言語運用に基づいた言語教育を行うことができる新しいタイプの外国語教育研究者が養成されます。これからの我が国の外国語教育の多言語化と高度化にとって、言語情報学拠点は重要な意味をもっていると考えます。

2. 研究組織と研究成果

この拠点では言語学，言語教育学，情報工学の三つの班がそれぞれの課題を研究し，研究の全体を言語情報学が統括します．言語情報学では言語学と言語教育学の有機的な統合が実現されます．さらに言語学と情報工学の連携，言語教育学と情報工学の連携もはかられ，これにより言語情報学を中心として，言語学，言語教育学，情報工学の研究分野の協働作業が実現します．また，三つの研究分野の専門知識を有する，優秀な若手研究者を養成し，わが国の言語教育の未来を担う人材育成をとくに重要視し，そのための教育プログラムを実施します．



若手研究者育成プログラム

- ・ウェブ教材開発
- ・TUFS外国語研修
- ・フィールド調査
- ・若手研究者自立支援
- ・研究連携

2.1 言語学

研究計画

世界的に見て資料価値の高い少数言語コーパスをウェブ上で公開します。またフィールド調査によって自然会話の録音を行い、データを文字化し、多言語の機能別・目的別コーパスを構築し、こうして構築されたコーパスおよび既存の言語コーパスを分析し、複雑多様な言語運用の実態を解明します。

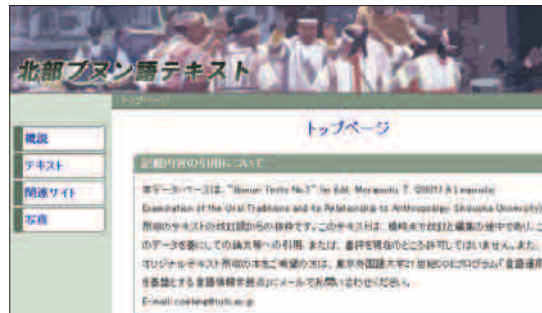
研究成果

- ・ 少数言語コーパス
- ・ 機能別・目的別言語コーパス
- ・ 言語運用に基づいたコーパス分析と音声分析

少数言語コーパス



ワヒー語基礎語彙 2005年8月公開



北部ブヌン語テキスト 2006年2月公開

ワヒー語はイラン語派東イラン語・パミール諸語に属する言語で、ブヌン語はオーストロネシア語族の西部語派に属する言語です。ワヒー語は辞書機能をそなえた基礎語彙検索ページを公開し、ブヌン語はいくつかの民話を音声付きで紹介するページを作成しました。このほかパラウン語ナムサン方言とクーイ語についてもフィールド調査を実施しました。

機能別・目的別言語コーパス

拠点では2004年度から多言語話しことばコーパスを構築するためのフィールド調査を実施してきました。フィールド調査は原則として教員1名と大学院生2・3名からなるチームで行われま

1

言語運用を基盤とする
言語情報学拠点

2

研究組織と
研究成果

3

若手研究者育成の
ための教育プログラム

4

COE研究成果の
外部評価

5

プロジェクト
関係者一覧

した。フィールド調査において、最も困難をきわめるインフォーマントの手配、録音場所の提供、さらに文字起こしのための人員確保等に関しては、現地大学の研究者の献身的な協力体制があったおかげで、各調査チームとも円滑に調査研究を行うことができました。

海外でのフィールド調査と連携機関

2004年度	2005年度	2006年度
スペイン語 (マドリード自治大学) ロシア語 (モスクワ大学) マレーシア語 (マレーシア国民大学)	フランス語 (エクス・マルセイユ大学) イタリア語 (サレント方言) トルコ語 (マルマラ大学)	フランス語 (パリ13大学) カナダのマルチリンガル話者 (バンクーバー) スペイン語 (マドリード自治大学) 中国語 (台湾) (淡江大学, 精華大学, 台湾大学, 銘傳大学) トルコ語 (マルマラ大学, イスタンブール大学)

多言語話しことばコーパスはウェブ上でいくつかの例を見ることができます。また文字データについては、利用申込の手続きをすれば、研究等に利用することが可能です。

海外でのフィールド調査のほかに、雑談場面を中心とした日本語母語話者同士の会話と日本語

母語話者と学習者の会話を収集しました。いずれも話者の年齢、性別、話題などが、統制された形で集められていますので、様々な観点から比較・対照研究ができるようになってきました。また自然会話分析の方法論的考察を『自然会話分析研究方法論ハンドブック』としてまとめました。



文字化システム (BTS-J) とコーディング (分類) の例
『BTSによる多言語話し言葉コーパス--日本語 (機能別の談話資料集) 2』より

ライン番号	発話文番号	発話文終了	話者	発話内容	発話文全体POL	文末POL
45	42-1	*	JBM03	あーそう、出身地で言わせてもらおうと、,[言わせてもらおうと、は聞き取りにくい発話]	/	/
46	43	*	YM01	はい。	NM	NM
47	42-2	/	JBM03	僕自身は生まれも育ちも東京なんで、,	/	/
48	44	*	YM01	はいはいはい。	NM	NM
49	45	*	JBM03	あーなんも面白いことがない(<笑い>)っていう…。	NM	NM
50	46	*	YM01	どのあた、どのあたりですか?。	P	P
51	47-1	/	JBM03	えーと生れたのが文京区で、(はいはい)で一幼稚園の(うん) 年長の(うん) ときに板橋区高島平に、,	/	/
52	48	*	YM01	はいはいはい。	NM	NM
53	47-2	/	JBM03	えー引っ越しをして、でずーっとそこで、暮らしてて、,	/	/
54	49	*	YM01	うんうん。	NM	NM
55	47-3	*	JBM03	でーまあ稼ぐようになってからはやっぱりひと一人暮らし始めたんですけどね。	P	P

言語運用に基づいたコーパス分析と音声分析

言語学はもともとことばの内部構造の記述をしたり、理論的に分析したりすることに専念し、ことばが現実に使われている場面での実態を科学的に分析することには必ずしも十分な配慮がなされてきたとはいえないかもしれません。実際に用いられる談話やテキストの運用に基づいた理論を構築し、それを語学学習に生かす工夫が求められています。言語学班では個別言語のコーパス分析に基づいて、これまでフランス語、ドイツ語、スペイン語、イタリア語、ロシア語、朝鮮語、中国語、日本語、マレーシア語等の言語運用に基づく研究を進めてきました。この中から14の成果を集めた論文集Corpus-Based Analyses on Sentence Structures, さらに国内外の研究者と協働して、音声学の論集 Prosody and Sentence Structuresやコーパス言語学の論集 Linguistic Informatics and Spoken Language CorporaとLinguistic Informatics and Corpus Linguisticsを刊行しました。この他にも多数の研究報告集を出版してきました。

- ・『Linguistic Informatics II Corpus-Based Analyses on Sentence Structures』2004年4月
- ・『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』2004年8月
- ・『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』2004年11月

1 言語運用を基盤とする言語情報学拠点

2 研究組織と研究成果

3 若手研究者育成のための教育プログラム

4 COE研究成果の外部評価

5 プロジェクト関係者一覧

- ・『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』 2005年10月
- ・『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』 2005年10月
- ・『Linguistic Informatics IV Prosody and Sentence Structures』 2005年12月
- ・『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』 2006年7月
- ・『言語情報学研究報告12 コーパス言語学の諸相』 2006年11月
- ・『Linguistic Informatics VI Linguistic Informatics and Spoken Language Corpora』 2006年11月
- ・『Linguistic Informatics VII Linguistic Informatics and Corpus Linguistics』 2007年3月



2.2 言語教育学

研究計画

世界に類例のない17言語による「TUFSS言語モジュール」を開発し、インターネットにより学内外で利用してもらい外部評価を受けます。また学習者言語コーパスの構築や自然談話分析等を行い、言語教育への応用を視野に入れた基礎研究、および独自の言語能力評価記述の研究を行い、言語評価モデルの構築を試みます。

研究成果

- ・「TUFSS言語モジュール」の開発と評価
- ・第二言語習得研究と学習者言語コーパスの構築
- ・独自の言語能力評価記述の研究
- ・自然談話の分析と教育への応用

「TUFSS言語モジュール」の開発と評価

TUFSS言語モジュールを開発し、外国語教育の高度化と多言語化を実現しました。TUFSS言語モジュールは発音・会話・文法・語彙の4つのモジュールからなり、それぞれが緊密にリンクしています。言語モジュールは学内外の学会やワークショップにおいて紹介され、東京外国語大学や他大学の外国語教育者から外部評価を受けました。



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

言語情報学拠点では言語を基本言語層、日常言語層、文化言語層の3つの階層としてとらえ、それぞれのモジュールと言語コーパスは以下のように各言語層に対応します。

基本言語層	日常言語層	文化言語層
発音モジュール IPAモジュール	会話モジュール, 文法モジュール, 語彙モジュール	
語彙範疇コーパス, 機能別・目的別多言語コーパス		

1

言語運用を基盤とする
言語情報学拠点

2

研究組織と
研究成果

3

若手研究者育成の
ための教育プログラム

4

COE研究成果の
外部評価

5

プロジェクト
関係者一覧

これまでの5年間で研究が行われてきたのは、主に基本言語層と日常言語層です。4つのモジュールはこれらをカバーしています。一方、語彙モジュールでは、国立国語研究所の『分類語彙表』に基づいて語彙の意味分類を行い、分類語彙表検索のページは語彙範疇コーパスの機能を果たすページになっています。また機能別・目的別の多言語コーパスとして構築された「多言語話しことばコーパス」は単に日常言語層だけでなく、音楽・歴史・教育といった文化言語層にかかわる貴重な話しことばデータを提供してくれます。



2006年9月に新サーバーを導入し、より快適なページ閲覧が可能になりました。そのためでしょうか、TUFSS言語モジュールのページは2006年10月から2007年1月までの四ヶ月間、毎月200万ページビュー（Webalizerの解析による）を越え、多くの方に利用されるようになりました。

Summary by Month										
Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Feb 2007	142696	96173	33792	1599	2619	5912165	3199	67585	192347	285392
Jan 2007	241501	166304	63764	2964	37904	158820291	91913	2038686	5161654	7486558
Dec 2006	221382	157042	64224	2644	32840	143359582	81994	1990959	4868311	6862872
Nov 2006	250156	180414	74774	2563	35386	166014433	76900	2243237	5412438	7504693
Oct 2006	293993	220187	90254	2561	35080	192284457	79416	2797890	6825807	9113789

TUFSS言語モジュールはウェブ上で広く紹介されており、2007年1月のGoogle検索エンジンの調査によると、「TUFSS言語モジュール」と「東外大言語モジュール」を含むウェブページだけでも1380ページがヒットします。検索結果の上位約35%は大学を始めとする教育機関や語学学習関連のウェブサイトです。個人のサイトでも利用評価等が紹介されています。使ったときの感想のほとんどが好意的なもので、他のサイトでは勉強できない東南アジアの諸言語も学べる点、マルチメディアを利用して学習できる点、平易でありながら網羅的に扱われている点、さらに無料で提供されている点などがとくに高く評価されています。

第二言語習得研究と学習者言語コーパスの構築

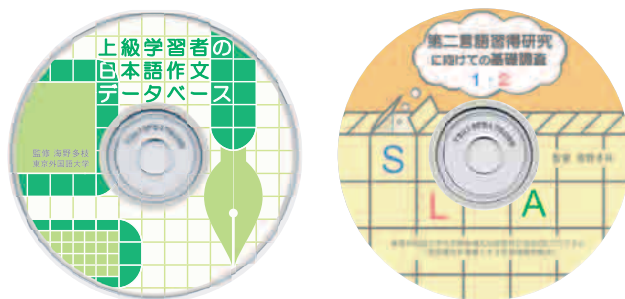
日本語学習者の作文データ、および日本人英語学習者の音声データから成る学習者言語コーパスを構築しました。この他、学習者ストラテジーの調査や、学習者言語コーパスを使用した基礎研

究を行い、言語教育・言語評価への応用を試みました。これらの成果については報告集や論文集を刊行しました。

- ・『言語情報学研究報告5 第二言語の教育・評価・習得』 2004年12月
- ・『言語情報学研究報告10 教材開発・評価・第二言語習得』 2006年3月
- ・『Linguistic Informatics V Studies in Second Language Teaching and Second Language Acquisition』 2006年3月
- ・『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』 2006年11月



- ・『第二言語習得研究に向けての基礎調査1・2』 2004年12月
- ・『上級学習者の日本語作文データベース』 2005年4月
- ・『日本人英語学習者音声データ文字化コーパス』 2007年2月



独自の言語能力評価記述の研究

言語能力記述モデルや教材評価についての論集を刊行しました。2006年にはe-learning を用いた発音学習と会話学習に関する大規模なアンケート調査を学内外で実施し、報告集にまとめました。

- ・『言語情報学研究報告2 言語学・応用言語学・情報工学』 2004年3月

1
言語運用を基盤とする
言語情報学拠点

2
研究組織と
研究成果

3
若手研究者育成の
ための教育プログラム

4
COE研究成果の
外部評価

5
プロジェクト
関係者一覽

- ・『言語情報学研究報告10 教材開発・評価・第二言語習得』 2006年3月
- ・『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』 2006年11月



自然談話の分析と教育への応用

大規模な日本語自然談話コーパスを構築し、社会心理学的な観点から談話の論理構成、会話のストラテジー、および言語機能の分析を行いました。またこうした調査を行うための方法論の研究や、言語教材への応用も試みました。

- ・『Linguistic Informatics III The First International Conference on Linguistic Informatics』 2004年10月
- ・『言語情報学研究報告6 自然会話分析と会話教育』 2004年12月
- ・『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』 2006年11月



- ・ BTSによる多言語話し言葉コーパス - 日本語会話 2005年6月公開
- ・ BTSによる多言語話し言葉コーパス - 日本語 2005年6月公開
- ・『自然会話分析研究方法論ハンドブック』 2005年6月刊行



2.3 情報工学

研究計画

音声やビデオを活用したユビキタス・マルチメディア言語教育の実現に向けた情報管理システムを構築し、多言語コーパス解析用ツールを開発します。

研究成果

- ・ e-learningによるユビキタス・マルチメディア言語教育
- ・ コーパス解析用ツールの開発と自然言語処理
- ・ 言語コーパスのデータベース化

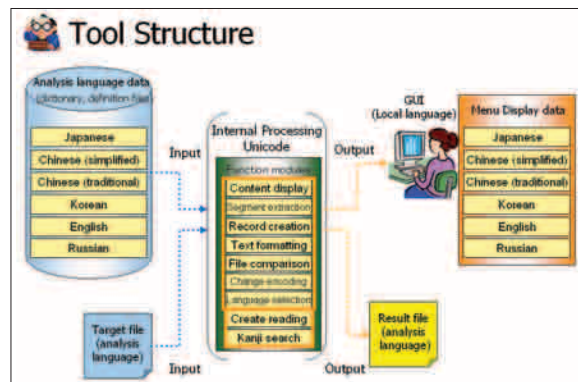
e-learningによるユビキタス・マルチメディア言語教育

2005年度にTUFUS言語モジュールの教材を利用するためのLMS (Learning Management System)として、COE-Learningシステムを構築しました。TUFUS言語モジュールの全教材が、コース単位でシステムに登録され、学習者は各コースに自由に参加できます。このシステムにより、いつでもどこでもページを閲覧だけでなく、教材の利用状況や成績管理といった学習者進捗管理が可能になり、マルチメディア言語学習環境が実現されました。



コーパス解析用ツールの開発と自然言語処理

多言語に対応した言語分析・語学教材作成支援ツール「KOTOEMON」を開発しました。このソフトウェアは、海外における日本語教育での教材作成支援を目指して、ユーザインタフェースに表示されるメニュー等のテキスト表示を多言語化したり、語学教育に携わる人の多くが、おそらくは文系出身者であること考慮し、多言語



1

言語運用を基盤とする
言語情報学拠点

2

研究組織と
研究成果

3

若手研究者育成の
ための教育プログラム

4

COE研究成果の
外部評価

5

プロジェクト
関係者一覧



での文型検索のための正規表現の入力をはじめとする処理手続きの可覚化を図ったりすることで利用しやすい工夫がされています。言語教育分野での利用の促進を狙っています。

なお『言語情報学研究報告15 TUFU言語モジュール教材・ことばの定量分析』(2006年12月)には、KOTOEMONの解説や「日本語流行歌中の外来語に関する調査」、「インターネット上に見られる慣用的表現の誤用に関する調査」など自然言語処理に関する論文を所収しています。

言語コーパスのデータベース化

言語学班は、教員1名と院生2・3名からなるフィールド調査グループを構成し、現地の大学と連携しながら自然会話の録音を行いました。録音データは現地協力者あるいは外注業者により文字起こしされ、データは情報工学班に渡されました。Transcriber等のソフトを用いて、音声データと文字データを同期させ、データをXML化し、Flashを用いて「多言語話しことばコーパス」としてウェブ上に公開しました。現在、フランス語（エックス）、フランス語（パリ）、トルコ語、マレーシア語、スペイン語、カナダ・バイリンガル話者が公開中です。また「BTSによる多言語話し言葉コーパス（日本語1・2・3、日本語会話1・2）」は、文字データのみをCD化し、オープンソースとして公開中で、利用申込が相次いでいます。さらにワヒー語、ブヌン語等の少数言語のコーパスもウェブ化しました。



http://www.coelang.tufs.ac.jp/multilingual_corpus/language_function.html

2.4 言語情報学

研究計画

言語教育に対して言語研究はその理論的な基盤を提供してきましたが、同時に、言語教育の実践を経て、言語理論へのフィードバックが求められています。拠点では、そうした言語研究と言語教育の相互交流を実現し、両者の有機的な統合を目指します。また大学院生を研究報告集や国際会議に積極的に参加させ、学術論文の執筆方法を学ばせ、世界で活躍するために必要な発表能力や討議能力を身につけさせます。

研究成果

- ・ 国際会議・ワークショップの開催
- ・ 言語学と言語教育学の有機的な統合
- ・ 多言語コーパスのTUFSS言語モジュールへの応用
- ・ 第二言語習得理論に基づく談話分析
- ・ マスコミ向け説明会の開催

国際会議・ワークショップの開催

世界的な拠点を形成するため、国際会議やワークショップを開催しました。2006年の円卓会議では海外の研究者に拠点の研究成果を評価してもらいました。

会議名（日程）	テーマ
第1回言語情報学国際会議（2004年12月13日－14日）	言語情報学－現状と未来－
国際ワークショップ（2005年12月9日）	話しことばコーパス－その重要性と応用－
第2回言語情報学国際会議（2005年12月10日）	言語学，応用言語学，情報工学の寄与
国際ワークショップ（2006年9月14日）	コーパス言語学 －その研究領域－
円卓パネル会議（2006年9月15日）	21世紀COE プロジェクト全体評価



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

1

言語運用を基盤とする
言語情報学拠点

2

研究組織と
研究成果

3

若手研究者育成の
ための教育プログラム

4

COE研究成果の
外部評価

5

プロジェクト
関係者一覧

言語学と言語教育学の有機的な統合

コンピュータ技術を用いて言語理論と言語教育の有機的な統合を目指しました。音声学の分野での重要な成果の一つはIPAモジュールの開発です。国際音声字母 (International Phonetic Alphabet) とは、世界中の言語音を同じ基準で記述できるように国際音声学会が定めた音声記号のことです。IPAモジュールでは子音、母音の具体的な音声を聞きながら学習できるようになっています。

通言語文法モジュールは世界の様々なタイプの言語の文法体系を通言語的に学習し教育するためのページです。17言語のTUFSS言語モジュールに共通する文法項目を見渡し、その共通性や特異点を理解できるように開発された教材で、性・数、時制、活用等さまざまな文法トピックを俯瞰できるよう項目ごとに解説され、各言語の文法モジュールにリンクしています。

言語文化ポータルサイトは26言語の言語・文化・歴史などを研究するための基礎資料を掲載したページです。



IPAモジュール (2004年8月公開)



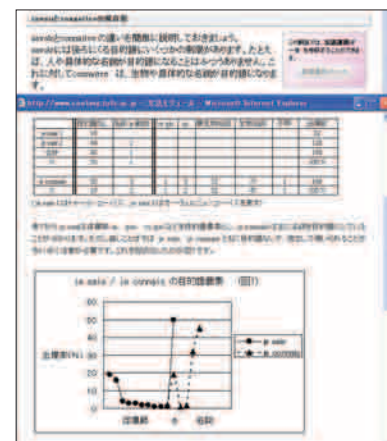
通言語文法モジュール
(2005年7月公開)



言語文化ポータルサイト
(2005年8月公開)

多言語コーパスのTUFSS言語モジュールへの応用

拠点では教員と大学院生を、2~3週間程度、現地に派遣し、臨地で話しことばを録音しました。そのデータを文字化したのが「多言語話しことばコーパス」です。このコーパスは言語運用の実態を把握する上で貴重なデータです。多言語コーパスの文法分析を行い、その成果をTUFSS言語モジュールに応用する試みが行われました。たとえば文法モジュールの解説に、話しことばのデータ分析の結果を入れて、現実の言語運用と規範文法の違いがわかるようにしました。



2.5 言語学と情報工学の連携

研究計画

機能別と目的別の多言語コーパスを構築します。これらのコーパスは言語教材の開発のみならず、異文化理解や教養教育にも資するところが極めて大きいと考えます。

研究成果

- ・機能別の多言語コーパスの構築
- ・研究目的別の多言語コーパスの構築
- ・自然言語処理

機能別の多言語コーパスの構築

既存の言語コーパスに基づく研究はいうまでもなく、言語の機能、文法、音声の研究など多目的に利用できる自然会話のコーパスの収集は拠点の重要な学術的貢献の一つです。これまでスペイン、ロシア、マレーシア、トルコ、フランス、イタリア、カナダ、台湾へ若手研究者を派遣し、現地機関の協力を得て、最大30時間程度の録音を実施し、これをもとに「多言語話しことばコーパス」を構築し、ウェブ上にいくつかの自然会話の実例を掲載しました。多言語話しことばコーパスは研究目的用のオープンソースであり、利用申請を行うことで研究に使用することができます。

トルコ語 多言語話言葉コーパス

TOP > コーパス > 父親のイメージ、長寿と日本

父親のイメージ、長寿と日本

再生 一時停止 停止 音小 音大

表示/非表示の切り替え
トルコ語 日本語 英語

トルコ語: bi(r) şeyde yazı görmüştüm # bi(r) çocuğun yüz bi(r) insanın daha doğrusu yüz yaşına gelene kadar her beş yıllık dönemler içerisinde babası hakkında düşündüklerini söylüyo(r)du <he he evet> bilmiyorum <enteresan bi(r) yazıydı okudum [ben onu]> evet gördün mü (?)

日本語: どこかで見ているのを見たんだけど、子どもは、というか、人は、より正確には、100歳になるまで5年ごとに自分の父親についての考えを述べるのだそうよ、<そう、そう> 分からないけど、<面白い内容だった、それ] 読んだよ> そう、あなたも見た(?)

研究目的別の多言語コーパスの構築

言語研究，文化研究などを行うための研究目的別の多言語コーパスを構築し，ウェブ上に公開しました．2005年度にはワヒー語基礎語彙や上級学習者の日本語作文データベースを公開し，2006年度には北部ブヌン語テキストのページが完成しました．

言語文化ポータルサイトは，東京外国語大学において専攻することができる26言語の言語・文化・歴史などを研究するための基本文献や参考資料を掲載したページです．

ワヒー語概説
検索について
ワヒー語検索
TOP ページへ

語彙一覧(26件)
※一部、再生しても無音の音声があります。今後、改善

id	音声	IPA	英訳	取得件数
1		çik wotɕn	to grow, to come out, to sprout	26件
2		çak	1) full 2) too much	
3		ças	debris (of small stones with rain etc)	
4		çak wotɕn	to increase	
5		çuk xak	to insert, insert	
6		çəzn(e) oɲnek	eyeglass	

語彙表INDEX
a b c ç d
ð dz e ə f g
y ŷ h i j
k l m n o p
q r s š t
t̥ ʈ ts u v
w x x̣ y z ẓ

絞込み検索

自然言語処理

多言語に対応した言語分析/語学教材作成支援ツール「KOTOEMON」を開発しました．このソフトウェアは，(1)用例収集の自動化と，(2)テキスト検索と整形の効率化，(3)ユーザインタフェースの向上を図ることにより，言語分析の効率化や語学教材作成を支援するものです．多言語に対応しているので海外で日本語教育に従事している外国の方が日本語教育用素材を効率的に作成するためのソフトウェア支援ツールとしても機能します．さらに，外国語教育に従事している教師が外国語教育で使用する教育素材を効率的に作成するためにも利用できます．

Tool Functions and their use

- 1 Overview
- 2 Background
- 3 Tool structure
- 4 Tool functions and their use
- 5 Demonstration
- 6 Achievements

- Display mode (six languages)
Japanese, Simplified Chinese, Traditional Chinese, Korean, Russian, and English
- The Key Word In Context (KWIC) function can search in 66 languages
- Rebooting is not necessary

2.6 言語教育学と情報工学の連携

研究計画

多言語を対象とするe-learningはいまだ発展途上の段階にあります。とりわけアジアの諸言語を対象とするe-learningは開発が遅れており、その意味で、拠点独自の研究計画と言えるでしょう。また英語・ドイツ語・フランス語等以外の言語における第二言語習得研究も蓄積が少なく、そうした研究は学問的価値が高いと思われまます。

研究成果

- ・ e-learning教材の開発
- ・ e-learning教材の利用
- ・ e-learning教材の評価

e-learning教材の開発

情報技術の支援を受けて多言語e-learning教材を実現しました。2003年度に発音モジュール（実践編）と会話モジュールをウェブ化しました。2005年度に文法モジュール、2006年度に語彙モジュールがそれぞれ公開されました。また留学生や外国の人が日本語の発音と会話を学べるようにTUFSS言語モジュール（多言語版）も開発されました。



e-learning教材の利用

英語と日本語では会話教材の教師用マニュアルや学習者ガイドが作成されました。これらの手引書は自律学習を支援したり、教室で効果的に会話モジュールを利用するためのものです。

1

言語運用を基盤とする
言語情報学拠点

- ・ えいご for KIDS: えいごではなそう! Teacher's Manual 2004年8月, 改訂版 2007年2月
- ・ TUFJ日本語会話モジュール学習者ガイド 2006年12月



2

研究組織と
研究成果

e-learning教材の評価

e-learning教材のモニタリングによる授業評価, 教材評価等を行いました。2004年度に発音モジュールの内部・外部評価を実施し, 2005年度からは学内の授業に e-learning システムが導入され, その中でTUFJ言語モジュールを使用し, 学習者の進捗状況を管理できるようになりました。2006年度にe-learningによる発音と会話の学習に関するアンケート調査を実施しました。

3

若手研究者育成の
ための教育プログラム



4

COE研究成果の
外部評価

5

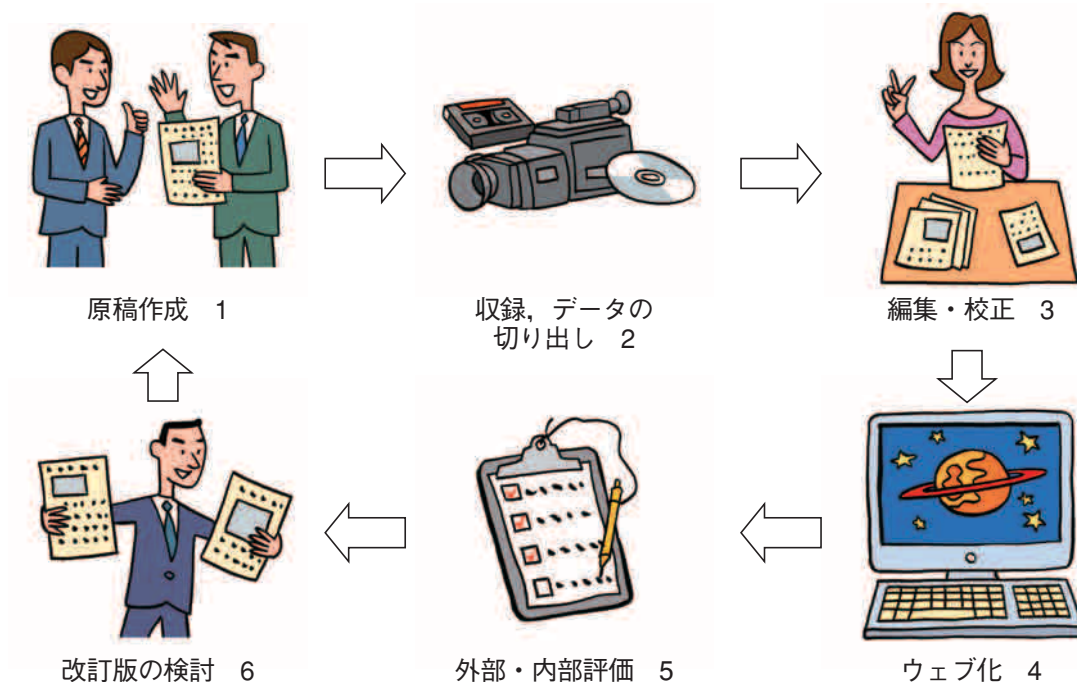
プロジェクト
関係者一覧

3. 若手研究者育成のための教育プログラム

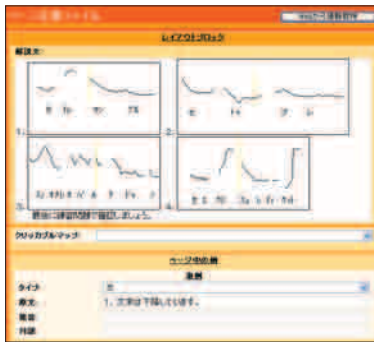
21世紀COE拠点は大学院生の教育研究プログラムを重要視してきました。言語情報学拠点においても、若手研究者を育成するための教育プログラムは、COE計画の最も重要な部分を占めてきました。言語学、言語教育学、情報工学に関する専門知識を有し、高度な言語運用能力をもち、わが国の言語教育の多言語化と高度化に貢献できるような新しいタイプの言語教育研究者を養成するために、以下の教育プログラムを実施しました。

ウェブ教材開発プログラム

TUFS言語モジュールのコンテンツ作成、ウェブ化、評価等の様々な行程に大学院生を参加させることで、彼らに音声学、言語学、言語教育学、情報工学の基礎知識とその研究方法を教育しました。

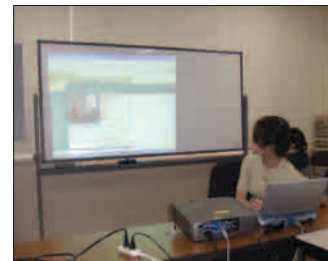


e-learning教材の開発作業に大学院生が参加することでITスキルの向上を目指しました。右ページの絵は入力フォームから発音モジュール（理論編）のXMLデータを作成する画面です。



TUFS 外国語教育研修プログラム

2003年度から大学院生の企画する夏期集中コース等の授業において、TUFS言語モジュールを利用してもらい、TUFS外国語研修プログラムを実施してきました。また研修後は実習レポートを作成し、教材評価、クラスリサーチを実施しました。『言語情報学7 言語情報学とコーパス言語学』（2007年3月）にTUFS言語モジュールの利用報告書を掲載しました。



フィールド調査プログラム

教員と大学院生が日常会話に関わる談話素材をそれぞれの国でフィールドワークを行って収集しました。彼らは2～3週間程度現地に滞在し、現地大学や研究機関の協力を得て、言語状況や運用場面の実態調査を行い、臨地録音を実施し、総合的な判断能力や問題解決能力も併せて身につけました。

フィールド調査

実施時期	対象言語・協力大学	派遣教員・協力者・院生
2004年11月18日 ～2004年12月7日	マレーシア語 マレーシア国民大学	正保勇（外国語学部）、 鵜沢洋志、野元裕樹
2004年12月15日 ～2004年12月30日	スペイン語 マドリード自治大学	高垣敏博（外国語学部）、 木越勉、結城健太郎
2004年12月16日 ～2004年12月31日	ロシア語 モスクワ大学	中澤英彦（外国語学部）、 恩田義徳、佐藤修
2005年6月22日 ～2005年6月26日	トルコ語 マルマラ大学	川口裕司（外国語学部）、 セリム・ユルマズ（マルマラ大学）

1

言語運用を基盤とする
言語情報学拠点

2

研究組織と
研究成果

3

若手研究者育成の
ための教育プログラム

4

COE研究成果の
外部評価

5

プロジェクト
関係者一覧

2005年7月1日 ～2005年7月16日	フランス語 エクス・マルセイユ大学	佐野敦至 (福島大学), 秋廣尚恵 (高等実習研究員PD), 中田俊介, 小川敬洋
2005年7月28日 ～2005年8月12日	パラウン語ナムサン方言 (ミャンマー国)	インインメイ
2005年8月18日 ～2005年8月31日	イタリア・サレント方言	山本真司 (外国語学部) ルカ・トーマ (桜美林大学), 田中慎吾
2005年9月12日 ～2005年9月21日	クワイ語 (タイ国)	太田ワランヤ
2006年2月2日 ～2006年3月1日	フランス語 パリ13大学	川口裕司 (外国語学部), 秋廣尚恵 (高等実習研究院PD), フランソワーズL. グーセンス (パリ13大学)
2006年2月19日 ～2006年3月28日	日本語, 英語, フランス語 (カナダ・バンクーバー)	時田朋子
2006年6月8日 ～2006年6月29日	トルコ語 マルマラ大学	セリム・ユルマズ (マルマラ大学), アルスンU. ユルマズ (イスタンブール大学)
2006年12月13日 ～2006年12月23日	スペイン語 マドリッド自治大学	高垣敏博 (外国語学部), 結城健太郎, 松井健吾, 臼井新
2006年12月16日 ～2007年1月6日	中国語 (台湾) 淡江大学	林俊成 (外国語学部), 須藤秀樹, 黄詩淑, 林虹瑛
2006年12月21日 ～2007年1月6日	中国語 (台湾) 台湾大学・精華大学・銘傳大学	望月圭子 (外国語学部), シア・ティエンミン・テレンス, 蔡松益



若手研究者自立支援プログラム

言語情報学の研究成果は、『言語情報学 Linguistic Informatics』と『言語情報学研究報告』を通してウェブ上に公開されました。また国際会議やワークショップを開催し、大学院生による国際会議発表を積極的に奨励し、国際学会で発表を行う院生に滞在費と渡航旅費を支給しました。彼らは国際的に評価の高いEurocall, E-Learn, ED-MEDIAをはじめ、国際語用論学会、国際歴史言語学会、国際マレー・インドネシア学会、国際韓国語教育学会等、計25の国際会議で研究発表を行ないました。

オランダのJohn Benjamins社からは、拠点の名を冠した叢書を出版し、研究成果を広く世界に向けて発信しました。優れた院生の研究論文が叢書の中に掲載されました。

シリーズ Usage-Based Linguistic Informatics, John Benjamins, Amsterdam/Philadelphia

第1巻 Linguistic Informatics State of the Art and the Future (2005)

第2巻 Corpus-Based Approaches to Sentence Structures (2005)

第3巻 Prosody and Syntax (2006)

第4巻 Readings in Second Language Pedagogy and Second Language Acquisition (2006)

第5巻 Spoken Language Corpus and Linguistic Informatics (2006)

第6巻 Corpus-Based Perspectives in Linguistics (2007)



拠点からは、情報工学の基礎を身につけ、言語学と言語教育学の理論に精通する多数の若手研究者が巣立ち、言語学・言語教育学の分野において19名が博士号を取得し、福岡大学、国学院大学、明治学院大学、清泉女子大学、東京成徳大学、東海大学、近畿大学、高麗大学（韓国）、東国大学（韓国）、和春科技大学（台湾）等の専任教員になり、2名が国立国語研究所の研究員に採用されています。



研究連携プログラム

フィールド調査プログラムでは派遣先大学の研究者と協働することで、院生たちに自然談話の調査法、転写法、分析法などを実地で学ばせました。このほか言語学班では静岡大学・東京学芸大学・清泉女子大学・東京工業大学の研究者、さらに筑波大学の院生やポストドク院生が協力して音声研究を行い、研究論文集 Prosody and Syntax を出版しました。また言語教育学班の言語能力評価プロジェクトでは、国立教育政策研究所の研究員の協力を得て、ウェブ教材の教材評価と授業評価を実施しました。拠点協力者から4名の学術振興会特別研究員（内2名はCOE枠）が出ました。こうした学外研究者との研究連携プログラムの実施を通して、大学院生は本学の大学院教育だけでは得ることのできない貴重な知見を得ることができました。

1

言語運用を基盤とする
言語情報学拠点

2

研究組織と
研究成果

3

若手研究者育成の
ための教育プログラム

4

COE研究成果の
外部評価

5

プロジェクト
関係者一覧

4. COE研究成果の外部評価

2006年9月15日に円卓パネル会議を開催し、海外から招聘した研究者に拠点プロジェクトの内容、進捗状況、研究成果を1時間半かけて説明し、その後、1時間の質疑応答を行いました。彼らは5年間の研究成果を評価し、パネル会議の内容を詳細に分析し、2006年10月に「東京外国語大学COEプログラムに関する共同報告 Joint Report on Tokyo University of Foreign Studies, Centre of Excellence」という文書を作成しました。外部評価に応じてくれた8名は、フランス、ドイツ、イギリス、カナダ、ノルウェーにおいて、それぞれの研究分野の第一線で活躍している研究者たちです。多忙なスケジュールの中、共同報告書を作成してくれた彼らの厚意に、この場を借りて感謝の意を表したいと思います。以下に共同報告書の日本語全訳を掲載します^註。



東京外国語大学COEプログラムに関する共同報告

2006年10月

1) プロジェクト計画

2002年に着手されたCOE研究プログラムの目的の詳細な内容がパネル会議には公開されていないため、私たちは十分な情報に基づいた評価を行う立場にはいない。しかし、提供された研究成果をもとに、ある程度まで当初の計画を推測することができる。研究成果を見ると、当初の計画が幾つかの非常に明確な特徴を掲げたものであったろうことは想像に難くない。第一に、音声学、形態論、統語論、語彙論だけでなく、語用論や社会言語学も含めた言語学の諸分野、第二に、特にCALL（コンピュータ補助による言語学習）教材を全体的な教育プログラムに取り入れるなどの研究方法、第三に、より効果的な言語教育を可能にするための実際的な目標、に関し賞賛に値する広い視野が見られる。

言語学研究に関しては、4回にわたる国際会議およびワークショップ、また言語情報学に関する出版物が既刊4冊、5冊目が近日出版予定、というのをみてもプログラムの質の高さと一貫した取り組みを物語っている。パネル出席者は、2006年9月に開催されたワークショップの際の見事な組織

註. 『共同報告』の英文オリジナルは『言語情報学7 言語情報学とコーパス言語学』の3.2. プロジェクト評価あるいは http://www.coelang.tufs.ac.jp/common/pdf/L17_project_evaluation_en.pdfを参照して下さい。

編成と連携ぶりに感銘を受けた。

2) 独自性

言語情報学という用語は少なくとも1980年代にまでさかのぼるが、TUFSSプロジェクトはこの用語に新しい意味合いをもたせている。言語習得、言語理論とコンピュータを統合した新しい学際的分野としての言語情報学である。当該プロジェクトには独自性をうたえない分野も含まれるが、理論言語学・(コンピュータ)応用言語学・コーパス言語学を統合し、COEプログラムで対象としている17言語もの多様な言語を扱った多言語CALLプラットフォームを構築するというCOEの目標には確かに独自性がある。理論に導かれた言語分析と実際の言語教育への具体的な応用の組み合わせは他に例を見ない。例えば、ある語彙項目が最も高い頻度で使われる文脈のコーパス分析に裏付けられた語彙項目の用法説明などである。私たちの見解では、このアプローチは教材の効果を高め、理論的研究課題の形成を鮮明にするものである。

本プロジェクトにおける多言語アプローチは、既に対象とされている言語にも、また追加される可能性のあるいかなる言語にも、包括的で有用な枠組みとなっている。これは、研究者と言語学習者の双方にとって非常に有用で、貴重な比較の道具となる。通言語文法モジュールは、母国語と第二言語の言語間の比較を提供することで学生の言語学習の補助となる。また、研究者にとっては、アジアの言語を含む広範な言語に共通の枠組みにより、広範な比較対照研究を促進するものである。特筆すべきは、この枠組みが言語間の潜在的な類似性を最大化しつつも、単一もしくは少数のテキストソースを何通りにも翻訳するとおこりがちな、対訳テキストに付随する文化的干渉を回避している点である。その結果、それぞれの言語への比較可能性と適合性がともに見られることは特に賞賛に値する。

3) 科学的性格

明らかに主要な応用であるインターネットを用いた言語学習プログラムは先に述べた通り非常に興味深い。しかし、インターフェースに使われている言語が日本語であるため、パネル出席者がインターフェースのデザインについて述べるには限界がある。英語やその他の言語によるインターフェースが加わることを待望する。

UBLI(言語運用を基盤とする言語情報学)シリーズは様々な国の学者による非常に興味深い研究成果をまとめたものである。なかには最先端に行く研究もあれば、様々な分野や下位分野の概観を示す有益な研究発表もある。しかし指定されたフォーマットのせいで個々の研究プロジェクトにおける理論的側面の発展がある程度妨げられたきらいはあるように思われる。最新の重要な言語理論、例えば、生成文法や認知言語学、構文文法、機能文法やパターン文法などへの言及が比較的乏しいことが気になる。これは、執筆者が決められた時間的制約を尊重し、コーパス分析の広く豊富な研究成果をあげることに専念しようとするあまり、時間とスペースの不足から理論的議論を要約したためと思われる。シリーズの内容は執筆依頼を引き受けた学者の専門分野によりある程度決まってしまうため、様々な下位分野の取り上げ方に少し偏りが見られた。今後のUBLI出版では、シリーズの巻や項を紹介する理論的論文を執筆または依頼することを検討してもよいだろう。

教育への応用に関しては、いくつかの分野においては最先端のコンピュータ言語学が言語学習モジュールに組み込まれるほど十分に発展していないということも認識されなければならない。これは批評というよりさらなる発展へのチャンスと受けとめていただきたい(「結果」参照)。

国際会議やワークショップの組織については、2006年9月のワークショップに招待された側の視

1

言語運用を基盤とする
言語情報学拠点

2

研究組織と
研究成果

3

若手研究者育成の
ための教育プログラム

4

COE研究成果の
外部評価

5

プロジェクト
関係者一覽

点から言うと、日本人研究者、大学院生や世界各国の研究者と知り合い、意見を交換するきわめて価値ある体験となった。大学院生にとっては、著名な学者と身近に接することのできる貴重な機会であった。

4) 国際的な貢献

東京外国語大学COEプログラムの国際的な貢献は、幾つかの観点から評価することができる。

第一に、教育的応用の対象となった言語範囲の国際性について。この点において、本プログラムは非常に国際的であり、類型学的には異なる多数の言語が、統一された方法的、体系的アプローチをもって扱われている。

第二に、国際的な研究成果の利用について。例えば、フランス語学では、本プログラムは、フランス語の口語（エックス・アン・プロバンス）コーパスを研究している最も本プログラムと関連性の深い機関の協力を得ている。以前のワークショッププログラムでは、統語的語彙（LADL）や構文解析（Intex）の領域で他機関と提携している。これら他機関との提携レベルが正確にはどの程度のものであったのか明らかではないが、本プログラムが貴重な国際的ネットワークをつくりあげていることは間違いない。

第三に、研究成果の国際的提供について。『Usage-Based Linguistic Informatics（言語運用を基盤とする言語情報学）』シリーズ（ジョン・ベンヤミンズ出版）は、国内研究者とUBLI国際ネットワークのメンバー双方の研究成果を世界中の学者に提供している。

パネル出席者の要望に対する回答によると、さらなる重要な前進が今後予定されている、あるいは既に見られるとのことである。成果は世界中の研究者に提供されることになっている。すなわち、要求に応じたコーパスの頒布、ソフトウェアのオープンソースでの公表、言語モジュールの他研究機関による利用実現が予定されている。

最後に、プログラムの国際的認知度について。この点においては、研究成果をウェブサイト上で英語でより体系的に発表しさえすれば、COEプログラムはさらなる国際貢献を達成し得るであろう（評価を下した時点では、ウェブサイトの「外殻」のみ英語で見ることができ、詳細情報のある下位レベルは日本語であった）。

5) 実現可能性（結果）

9月15日に行われたパネル会議では、出席者は本プログラムの実現可能性に関する事後評価を行うことに懸念を表明したが、今、全員が感じているところでは、例えば対照分析、類型論的記述、言語教育への応用などにおける成果は、それらが2002年に策定された計画に厳密に沿ったものであるかは我々には不明だが、東京外国語大学COEプログラムに対する肯定的評価が正当なものであることを証明するのに十分すぎるものである。

むろん、CALLプログラムを使った学習者が達成した言語学習の質を分析していくことも有益であろう。

6) 科学的業績

第2項で東京外国語大学COEプログラムの独自性を指摘し、第4項でその国際的貢献を指摘した。

9月15日の発表で、大量の口語と文語コーパスが作成され、注釈がつけられ、COEプログラムで使用できるようになったことが明かにされた。UBLIプロジェクトの報告中に、数ヶ国語についてCOE言語モジュールとこれらの「リアルデータモジュール」間の相互作用が示された。東京外国語

大学の第一の目的がCALL応用にあることを考えると、これらはすばらしい業績である。しかし、現在の成果と将来の目標がどの程度までNLP（自然言語処理）技術の発展に依存しているのかがはっきりしなかった。例えば、評価時点では文法モジュールはまだ開発段階にあり、意味の曖昧性の解消はもとより、品詞タグ付けや構文解析などの技術の統合は示されなかった。インタラクティブなCALLモジュールはこういった技術を必要とする一方で、多言語向けにそのようなツールを開発することについては、17の言語用にすでに開発されたものを発展させる明白な余地はあるものの、COEプログラムの規定目標をはるかに超えている。しかし、プログラムの業績が同じことの繰り返しに限定されないようにするには、適切な分野における有能なパートナーとの提携が次の段階では検討されるべきであろう。この点において、また他の面においても、海外で行われている同様の研究を考慮することが賢明であろう。例えば、IPA（国際発音記号）は、ピーター・ラディフォギッド著『母音と子音（Vowels and Consonants）』やオンラインのジョン・ウェルズ（John Wells on-line）などのように、それが発生する特定の言語に関連付けた説明が考えられる。さらに、市販の商品のなかには、発音の超分節的側面について高度なフィードバックが可能なものもある。

UBLIプロジェクトは多数の国の研究者を結集し、出版された研究論文集（叢書）のシリーズには最先端の研究発表が掲載されている。

全体としてみると、協力を依頼された海外の研究者を除いて、ほとんど国内の研究者だけで意欲的な言語学プログラムをつくろうとしたことは大胆な決定であった。同じような状況にある西洋の多くの大学ならば、拠点研究機関以外でのプログラムの認知度を即座に高めるため、著名な学者を迎え入れようとしたであろう。資料を見る限り、国内で開発するという方針により、明らかに遠大な目標を達成するには、恐らく5年では短すぎたということになろう。しかし、東京外国語大学が助成を受けた際の状況を知らない以上、パネル出席者の見解は限られたものではある。検証した内容を踏まえる限り、プログラムがその可能性をすべて達成できるように、助成を継続して受けられるよう心から願う次第である。

7) 若手研究者の育成

若手研究者育成について、出席者はワークショップで非常によい印象を持った。大学院生と指導教員の連携は極めて良好のようである。多くの若手の研究者が国際的なフォーラムで自主的に（9月のワークショップではそうであった）研究発表をする機会を与えられている（もしくは過去にこの機会を与えられた）ということに好感を持った。

会議参加の詳細なリストを見ると、個人であれグループであれ、国際的学者からなる観客を前にして研究成果を発表する機会を得た学生は21人を下らないことがわかる。複数回そういった機会を得ている学生もいる。こういった機会は若手研究者の育成において非常に貴重なものである。

- ペーター・ブルーメンタール（ケルン大学、ドイツ）
- ジャン＝フィリップ・ダルベラ（ニース大学、フランス）
- ジャン＝ミシェル・エロワ（ピカルディー・ジュールヴェルヌ大学、フランス）
- ピエール・クンストマン（オタワ大学、カナダ）
- シャンタル・リュック（オスロ大学、ノルウェー）
- イヴ＝シャルル・モラン（モントリオール大学、カナダ）
- ティム・プーリー（ロンドン・メトロポリタン大学、イギリス）
- アヒム・シュタイン（シュトゥットガルト大学、ドイツ）

1

言語運用を基盤とする
言語情報学拠点

2

研究組織と
研究成果

3

若手研究者育成の
ための教育プログラム

4

COE研究成果の
外部評価

5

プロジェクト
関係者一覧

5. プロジェクト関係者一覧

21世紀COE拠点事業推進担当者

	専門	役割分担
川口裕司	フランス語学・トルコ語学	拠点リーダー, 言語情報学統括
在間進	ドイツ語学	言語教育学統括, 言語情報学
富盛伸夫	理論言語学	言語学
高垣敏博	スペイン語学	言語学
敦賀陽一郎	フランス語学	言語学
亀山郁夫	ロシア文学	言語教育学
野間秀樹	朝鮮語学	言語学
芝野耕司	情報学	情報工学統括・言語情報学
峰岸真琴	言語学	情報工学・言語情報学
宇佐美まゆみ	言語社会心理学・日本語教育学	言語教育学
梶茂樹 (京都大学) (2002-2004年度)	言語学・音韻論	言語学
水林章 (上智大学) (2002-2005年度)	フランス文学・歴史学	言語教育学

研究プロジェクト2002年度～2006年度

言語情報学班

TUFS言語モジュール開発, 言語文化ポータルサイト構築, 多言語コーパス, 談話分析, 高度教養教材開発, 言語情報学研究報告・言語情報学の出版

プロジェクト責任者・協力者	研究タイトル
川口裕司, 峰岸真琴, 芝野耕司	TUFS言語モジュール開発
川口裕司, 敦賀陽一郎, 高垣敏博, 在間進	言語文化ポータルサイト構築 多言語話しことばコーパス構築
宇佐美まゆみ	談話分析
峰岸真琴, 亀山郁夫, 川口裕司	高度教養教材開発
川口裕司, 高垣敏博, 在間進, 敦賀陽一郎, 宇佐美まゆみ, 峰岸真琴, 亀山郁夫, 富盛伸夫, 芝野耕司	『言語情報学 Linguistic Informatics』, 『言語情報学研究報告』の出版

1

言語運用を基盤とする
言語情報学拠点

2

研究組織と
研究成果

3

若手研究者育成の
ための教育プログラム

4

COE研究成果の
外部評価

5

プロジェクト
関係者一覧

言語学班

コーパスに基づいた言語運用分析，通言語的観点からみたプロソディーと統語論

プロジェクト責任者・協力者	研究タイトル
敦賀陽一郎	フランス語動詞構文クラスの実例分類・頻度数調査とフランス語動詞構文研究文献調査
川口裕司	フランス語の否定構文の通時的研究， コーパスに基づいたフランス語の条件法の分析
黒澤直俊	ラテン語とポルトガル語の修飾辞と被修飾辞の語順
在間進，成田節	コーパスに基づくドイツ語の動詞構文研究
高垣敏博， 宮本正美（神戸市外大）	スペイン語コーパスの構築と関連ツール開発によるスペイン語研究の高度化
山本真司， 鈴木信吾（東京音楽大学）	イタリア語・ルーマニア語対照言語学
中澤英彦	コーパスに基づくロシア語の体の研究， ロシア語動詞研究におけるコーパスの利用
宗宮喜代子	コーパスに見る有標構文の諸相
浦田和幸	英語の接続法に関する通時的研究
三宅登之	コーパスに基づく中国語動詞構文の研究
望月圭子	原因・結果をあらわす日本語・中国語・英語の複合動詞の対照研究
風間伸次郎	口語・文語コーパスを用いた文法の記述的研究
正保勇	マレーシア語の文型調査のためのコーパス作成，マレーシア語の有標語順
吉枝聡子	ワヒー語の語彙コーパス構築
川村大	古代日本語形容詞の格表示に関するデータベース作成

プロジェクト責任者・協力者	研究タイトル
川口裕司，森口恒一（静岡大学） 斎藤純男（東京学芸大），斎藤弘子，降幡正志， 富盛伸夫，井上史雄（明海大学）， 佐藤大和（東京工業大学），鮎沢孝子（国際教養大学）	統語的曖昧性と韻律特徴，プロソディーと統語論， TUFSS会話モジュールの音声分析

言語教育学班

談話分析，第二言語習得，TUFSS言語モジュール評価

プロジェクト責任者・協力者	研究タイトル
宇佐美まゆみ 鈴木卓（フェリス女子大）	多言語話しことばコーパス構築と分析， 自然会話分析の方法論に関する基礎研究， 日本語，朝鮮語，中国語，英語の基本的文字化システムの開発

吉富朝子, 狩野晶子 (上智大)	日本人英語学習者言語コーパスの構築と分析, 英語会話モジュールの教師用マニュアル作成
海野多枝	日本語学習者言語コーパスの構築と分析, 日本語会話モジュールの学習者用マニュアル作成
根岸雅史, 高島英幸, 市川雅教, 山森光陽 (国立教育政策研究所)	言語能力記述規準の研究, TUFUS言語モジュールの評価

情報工学班
e-learning, 自然言語処理

プロジェクト責任者	研究タイトル
林俊成	TUFUS言語モジュールのe-learning化
佐野洋, 望月源	日本語教育のための教材コーパス構築, 自然言語処理

TUFUS言語モジュール監修者

IPAモジュール, 通言語文法モジュール	
IPAモジュール	斎藤純男, 中川裕, 益子幸江
通言語文法モジュール	峰岸真琴, 風間伸次郎
TUFUS言語モジュール	
P=発音, D=会話, G=文法, V=語彙	
英語	野村恵造 (G), 斎藤弘子 (P), 浦田和幸 (G,V), 吉富朝子 (D)
ドイツ語	成田節 (P,D,G), 正木晶子 (P), 在間進 (V)
フランス語	川口裕司 (P,D,G,V), 水林章 (D)
スペイン語	川上茂信 (P,D,G), 高垣敏博 (G,V)
ポルトガル語	黒澤直俊 (P,D,G,V), 武田千香 (D)
ロシア語	中澤英彦 (P,D,V,G)
中国語	平井和之 (日本大学) (P,D), 三宅登之 (G,V)
朝鮮語	趙義成 (P), 五十嵐孔一 (D), 野間秀樹 (G,V)
モンゴル語	温品廉三 (D), 岡田和行 (G,V), 斎藤純男 (東京学芸大学) (P)
インドネシア語	降幡正志 (P,D,G,V)
フィリピン語	森口恒一 (静岡大学) (P), 山下美知子 (D,G,V)
ラオス語	鈴木玲子 (P,D,G,V)
カンボジア語	上田広美 (P,D,G,V), 岡田知子 (P,D,G,V)
ベトナム語	宇根祥夫 (P,D,G,V), 田原洋樹 (立命館アジア太平洋大学) (P)
アラビア語	ロバート・ラトクリフ (P,D,G,V)
トルコ語	福盛貴弘 (大東文化大学) (P), 菅原睦 (D,G,V)
日本語	荒川洋平 (V), 川村大 (G), 佐藤ゆみ子 (P), 海野多枝 (D,G)

上記以外にも『言語情報学研究報告』に寄稿された協力者の方が多数いました。

John Benjaminsの叢書 Usage-Based Linguistic Informatics (UBLI) における海外協力者

<p>UBLI I: Linguistic Informatics State of the Art and the Future</p> <p>Piter van Reenen, Anke Jongkind (Free University Amsterdam), Christian Leclère (University of Marne-la-Vallée), Francisco Moreno-Fernández (University of Alcalá), Janet Holmes (Victoria University of Wellington), David Block (University of London)</p>
<p>UBLI II: Corpus-Based Approaches to Sentence Structures</p> <p>Francisco Moreno-Fernández (University of Alcalá)</p>
<p>UBLI III: Prosody and Syntax</p> <p>Iván Fónagy (1920-2005) (Hungarian Academy), Eva Bérard (University of Paris 3), Jacqueline Vaissière (University of Paris 3), Alexis Michaud (University of Paris 3), Piet Mertens (University of Leuven), Manfred Markus (Innsbruck University), Joaquim Brandão de Carvalho (University of Paris 8), Cédric Patin (University of Paris 3), Annie Rialland (University of Paris 3), Selim Yılmaz (Marmara University), Arsun Uras Yılmaz (Istanbul University)</p>
<p>UBLI IV: Readings in Second Language Pedagogy and Second Language Acquisition In Japanese Context</p> <p>Lynne Hansen (Brigham Young University), Joshua Rowe (Brigham Young University), Barbara Pizziconi (University of London), Amy Snyder Ohta (University of Washington), Keiko Yoshioka (Leiden University)</p>
<p>UBLI V: Spoken Language Corpus and Linguistic Informatics</p> <p>Claire Blanche-Benveniste (École Pratique des Hautes Études), Susan M. Conrad (Portland State University), Massimo Moneglia (University of Florence), Emanuella Cresti (University of Florence), José Deulofeu (University of Aix-Marseille I), Antonio Moreno-Sandoval (Autonomous University of Madrid), José Maria Guirao-Miras (University of Granada), Maria Fernanda Gorjão Bacelar do Oliveira Nascimento (Lisbon University), José Bettencourt Gonçalves (Lisbon University), Maria Amália Pereira Mendes (Lisbon University), Sandra Cristina dos Santos Antunes (Lisbon University), Selim Yılmaz (Marmara University)</p>
<p>UBLI VI: Corpus-Based Perspectives in Linguistics</p> <p>Jean-Philippe Dalbera (École Pratique des Hautes Études), Marie-José Dalbera-Stefanaggi (University of Corsica), Peter Blumenthal (University of Cologne), Pierre Kunstmann (University of Ottawa), Chantal Lyche (University of Oslo), Jean-Michel Eloy (University of Amiens), Tony McEnery (Lancaster University), Onghua Xiao (Lancaster University), Randi Reppen (Northern Arizona University), Tim Pooley (London Metropolitan University), Yves-Charles Morin (University of Montreal), Achim Stein (University of Stuttgart)</p>

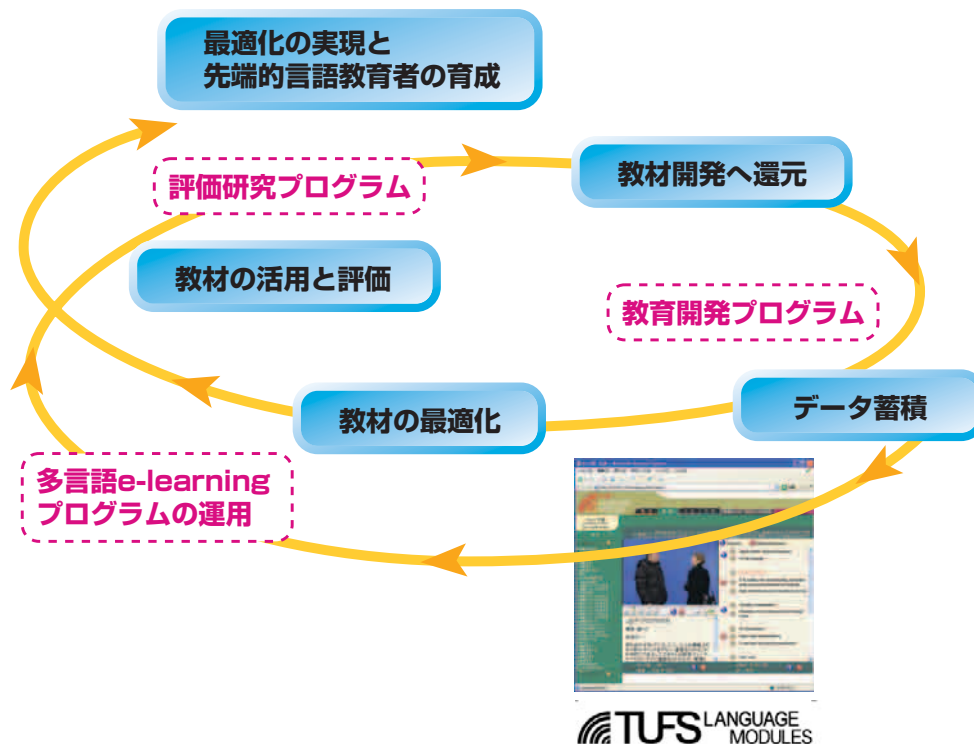
参照：http://www.benjamins.com/cgi-bin/t_seriesview.cgi?series=UBLI

言語情報学拠点の継承と発展

言語情報学拠点プログラムは2007年3月に終了しますが、この拠点の教材作成、e-learning、教材評価などの諸計画は、平成19年度特別教育研究経費（教育改革）として採択された「世界の「言語・文化・地域」理解のための最適化教育プログラム（以下、最適化プログラムと略）」によって継承されます。

最適化プログラムの事業目的は、東京外国語大学における教育研究の成果と蓄積をいかし、世界の「言語・文化・地域」に関するデジタル教材の開発と利用を推進するもので、その目的は以下の3つです。

- (1) 先端的なデジタル教材を開発し、多言語e-learning 教育プログラムのもとで活用する。IT 技術を活用することで、「言語・文化・地域」に関する本学教育の手法を高度化し、個々の学習者の個性やニーズに対応する最適化教育を実現します。
- (2) 本事業に大学院生の参加を促し、教材評価手法とデジタル教材の開発能力を身につけた言語教育者・言語研究者を育成します。
- (3) 本学で開発される世界の諸言語・文化・地域に関するデジタル教材をインターネット上で公開し、本学における教育研究の成果を広く社会に還元します。





183-8534 府中市朝日町3-11-1
東京外国語大学大学院21世紀COEプログラム
言語運用を基盤とする言語情報学拠点
☎042-330-5541 (事務局)