The 21st Century COE Program
"Usage-Based Linguistic Informatics"

C O N T E N T S

Message from the President

Setsuho IKEHATA (President, Tokyo University of Foreign Studies)

The 21st Century COE (Center Of Excellence) Program, launched by the Ministry of Education, Sports, Culture, Science and Technology in 2002, grants subsidies to distinguished universities in our country for the establishment of a center for research and education which maintains the highest academic standards in the world and covers a diverse range of fields. The Program aims to raise the level of research in our country's universities and to foster the creative academic minds of future leaders in the world. Tokyo University of Foreign Studies (TUFS) submitted applications for research projects in two of the selected programs – the Humanities and Interdisciplinary/Compound/New Sphere fields. We have obtained wonderful results; both projects were selected for funding. We are extremely pleased and encouraged by this high evaluation of the unique research projects and the educational potential of our Graduate School of Area and Culture Studies.

A strong foundation in foreign languages is vital to Area and Culture Studies. TUFS engages in education and research spanning more than 50 languages, cultures and societies in every part of the world. This contributes to cross-cultural understanding and the fostering of people who are capable of helping actualize the goal of creating a harmonious global community. In addition, a double-major system that requires students to specialize in both a language and a discipline-related course of study enables TUFS to produce graduates equipped with a high degree of language competence and a deep understanding of world cultures and societies.

Our new campus in Fuchu is proudly equipped with a state-of-the-art computing network. The most outstanding feature is the information literacy and the number of computers on campus, which ranks at the top level among liberal arts universities in our country. With such a priviledged information infrastructure, TUFS endeavors to make the best use of multimedia, the Internet and other high-tech capabilities to develop the most advanced language education system.

The University's Usage-Based Linguistic Informatics project, selected by the 21st Century COE Program, is the concrete manifestation of our plans for the future, which I have just outlined. The implementation team members are committed to this future vision and are vigorously engaged in the project. It is my avid hope that they will produce rewarding results. It is the intention of everyone at TUFS to combine our wisdom in a concerted effort to do our utmost to make the 21st Century COE Program a success.

With a view to providing full support to the program, TUFS has established the "21st Century COE Program Administration Office" which reports directly to me, as the president of the university. This office is an inter-sectional organization consisting of the president, the vice president, the deans of each division, the program leader, as well as the managers of the secretariat. It has the important role of enhancing cooperation between the various sections within TUFS and to administrate the use of space and the budget allocated for research.

Tokyo, September 1, 2004

Center of Usage-Based Linguistic Informatics (2002-2006)

Person in charge of program promotion

Name	Area of specialization	Role
Yuji KAWAGUCHI	French & Turkish Linguistics	Center leader; responsible for Linguistic Informatics
Susumu ZAIMA	German Linguistics	Responsible for Applied Linguistics & Linguistic Informatics
Nobuo TOMIMORI	Theoretical Linguistics	Linguistics
Toshihiro TAKAGAKI	Spanish Linguistics	Responsible for Linguistics
Yoichiro TSURUGA	French Linguistics	Linguistics
Ikuo KAMEYAMA	Russian Literature	Applied Linguistics
Akira MIZUBAYASHI	French Literature & History	Applied Linguistics
Hideki NOMA	Korean Linguistics	Linguistics
Kohji SHIBANO	Information Science	Responsible for Computer Sciences & Linguistic Informatics
Shigeki KAJI	Phonology	Linguistics
Makoto MINEGISHI	Linguistics	Computer Sciences & Linguistic Informatics
Mayumi USAMI	Social Psychology of Language	Applied Linguistics

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

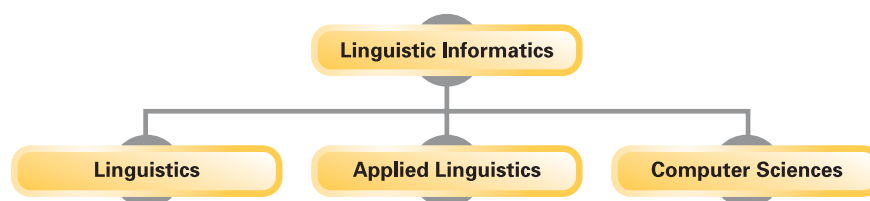# What is Linguistic Informatics?

"Linguistic Informatics" is an academic field that considers what happens when linguistics and language education are combined based on computer science. This pamphlet will provide an easy-to-understand explanation of our COE Program.

In the linguistics research conducted by Tokyo University of Foreign Studies thus far, we have reconsidered language theory and its importance through actual linguistic education practice rather than merely carry out research on language theory. It can be said that this type of interactive feedback between linguistic research and language education is representative of Tokyo University of Foreign Studies' uniqueness and its distinctive scholastic character. The COE Program capitalizes on our university's unique experience to gather linguistic usage data for a vast array of languages from around the world and then analyze that data using computer science. The results of these analyses are then put to practical use in actual linguistic education settings, thus enabling an organic integration of linguistic theory and education. This integration of linguistics and language education is highly expected to lead to the advancement and increased effectiveness of conventional foreign language education.
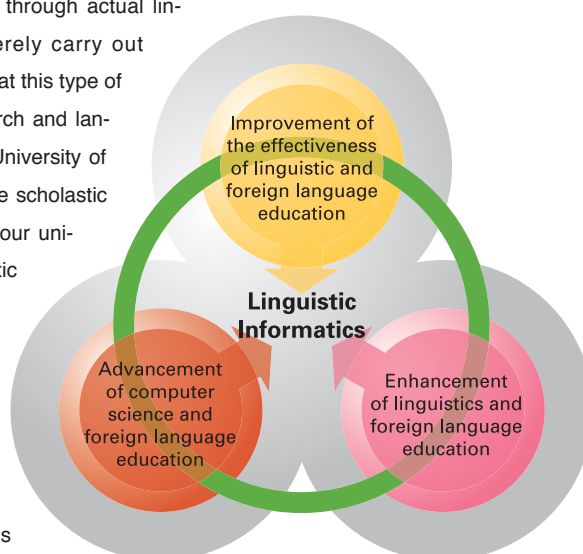
As shown in the organizational chart below, the COE program is made up of four research groups.
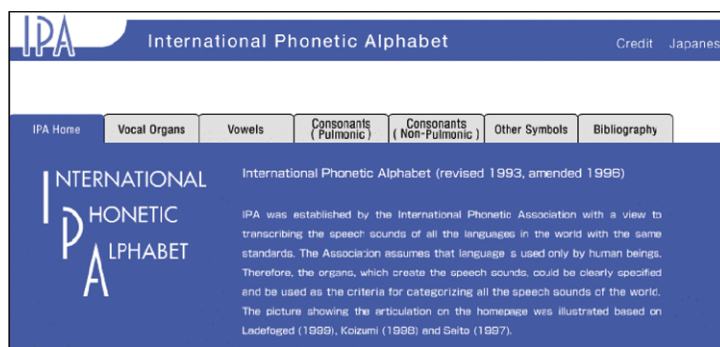
The "Linguistic Informatics" Group plays a central role in the COE program since it is the sole research group in charge of the development and application of TUFS Language Modules. There are three groups under the "Linguistic Informatics" Group: "Linguistics," "Applied Linguistics" and "Computer Sciences." These three groups conduct basic research related to the development, improvement and application of the TUFS Language Modules.

The Supervisory Group charged with the responsibility of overseeing this research organization is made up of seven of the people in charge of program promotion for the 21st Century COE Program. Important decisions related to the implementation of annual plans are made at Supervisory Group meetings. In addition, a Liaison Group has been established to forge closer ties among these groups. Liaison Group meetings are also held on a frequent basis, and information regarding the progress of the project is reported among the groups.

TUFS Language Modules

Internal disclosure of the TUFS Language Modules commenced on April 25, 2003, and the first module to be revealed to the public was the IPA (International Phonetic Alphabet) Module on June 9. IPA is comprised of phonetic symbols used to transcribe the speech sounds of all the languages in the world with the same standards. By clicking on the phonetic symbols for vowels and consonants that appear in the IPA Module and then clicking on the talking parrot at the lower-right portion of the screen, you can listen to each of the respective sounds.

http://www.coelang.tufs.ac.jp/ipa/english/index.htm

In September 2003, following the unveiling of the IPA Module, the Pronunciation Module, which covers 11 languages (German, French, Spanish, Portuguese, Russian, Chinese, Korean, Mongolian, Filipino, Vietnamese and Japanese), was released. Then, on December 12, 2003, all 17 languages of the Dialogue Module were published on the Internet in line with the holding of the "First International Conference on Linguistic Informatics." These are, however, the initial versions of these modules and are therefore subject to revision.

The TUFS Language Modules are geared towards students who understand Japanese and are studying the target language for the first time. The English module, however, is targeted at elementary and junior high school students.

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

The 17 languages of the TUFS Language Modules:

English, German, French, Spanish, Portuguese, Russian, Chinese, Korean, Mongolian, Indonesian, Filipino, Lao, Cambodian, Vietnamese, Arabic, Turkish, Japanese

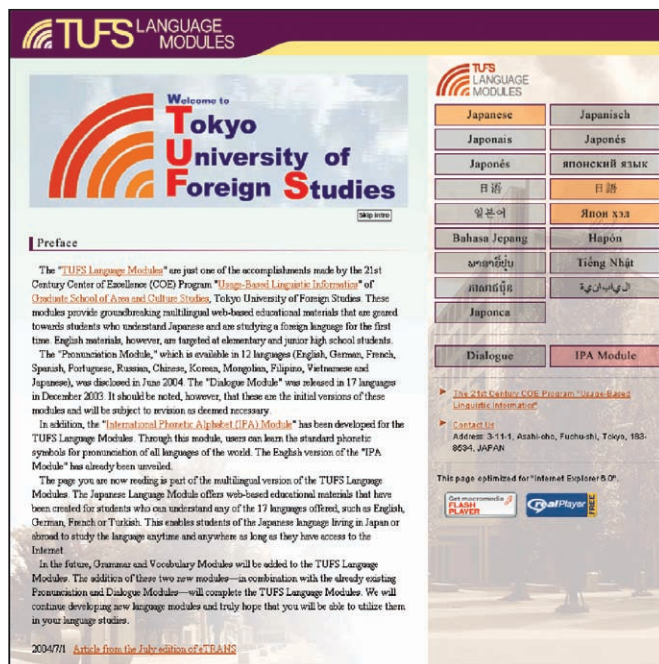## Modules and Cross-Linguistic Expressions

The TUFS Language Modules are innovative web-based educational materials that have never before been available. As the name suggests, these modules were created based on a "modular concept." Children's magazines often include supplemental projects. When you were a child, did you follow the instructions that came with projects like these to the exact letter? As long as the end result is the same, shouldn't everyone be able to do things their own way?

Language learning materials comprised of individual components that can be used on their own to a certain extent but can also be utilized together are referred to as "module educational materials." With the TUFS Language Modules, language learning can be broken down into four modules: Pronunciation, Dialogue, Grammar and Vocabulary. While learners can of course study these four modules comprehensively, they can also freely choose which of the four modules to begin their language studies with. This freedom of choice can certainly be said to be one of the advantages of the TUFS Language Modules. And unlike conventional printed educational materials, these modular educational materials can be revised or modified easily, which is another merit of this system. Moreover, by putting web links to practical use, these four modules could become more closely linked in the future.

Although the TUFS Language Modules give learners more freedom to design their own study plans, it will be necessary to develop a uniform system for measuring performance and achievement. We plan to commence research for the development of a descriptive model for language proficiency that is unique to Tokyo University of Foreign Studies and utilizes the TUFS Language Modules.

Multilingual Interactive Modules

Wouldn't it be wonderful to be able to use the same educational materials to study several different languages? The multilingual version of the TUFS Language Modules makes it possible to do just that. At present, those who understand English, Mongolian and Chinese (traditional Chinese characters) can study Japanese using the Pronunciation Module and the Dialogue Module.

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

Example page for studying Japanese dialogue in Mongolian

Example page of studying Japanese pronunciation in Chinese (traditional Chinese characters)

This multilingual version not only corresponds with the current trends toward multiculturalism and multilingualism, but can also be said to be a challenge against conventional foreign language education in Japan. By improving the multilingual version, the traditional image of Universities of Foreign Studies being universities where Japanese students study foreign languages will be shed and these universities will have the potential to undergo a transformation into universities where non-Japanese and exchange students can study Japanese or other languages. We will continue to press forward with the multilingualization of the TUFS Language Modules.

Research Projects

Linguistics

- Phonetics research

  Construction of the IPA Module

  Cross-linguistic perspectives in phonetics
- Corpus linguistics

  Sentence structure analysis based on multilingual corpus

  Analysis of grammatical function based on field research
- Practical use of the TUFS Language Modules

  Application of the results of multilingual corpus analysis to the TUFS Language Modules

Applied Linguistics

- Discourse data analysis

  Compilation of natural dialogue data based on field research

  Natural dialogue data analysis for linguistic usage research
- Second language acquisition (SLA)

  Construction of a Japanese composition (advanced-level students) database

  Research on second language acquisition strategies and beliefs

  Basic research for an English learner language corpus

  Creation of the English Dialogue Module teacher's manual
- Evaluation research

  Implementation and analysis of a questionnaire for module users

  Development of a TUFS language proficiency scale

Example page of studying Japanese dialogue in English

Computer Sciences

- E-learning

  Creation of databases and Web pages for the TUFS Language Modules
- Multilingual information processing

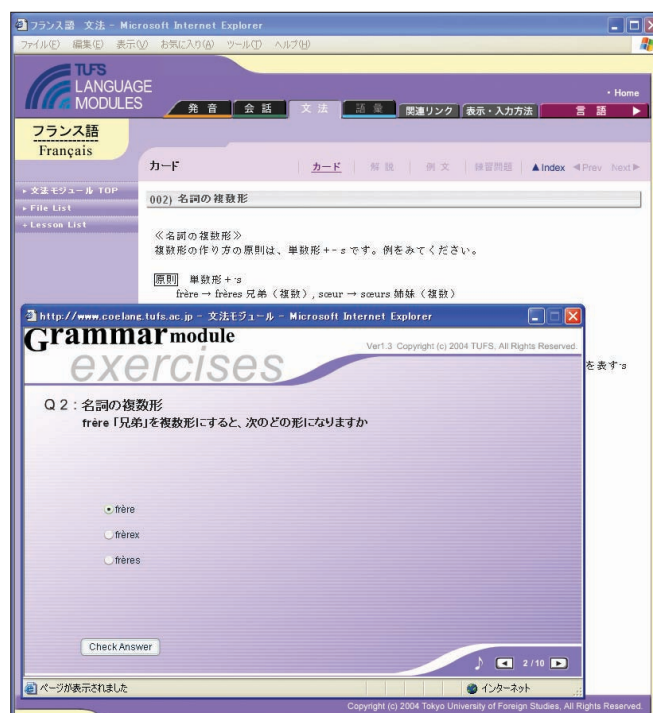 Natural language processing and analysis of English, Japanese and other languages

Research Projects in Academic Year 2004

Linguistic Informatics:

The development of the grammar module is underway, and information should be available by the beginning of 2005. The development of the vocabulary module will commence next year.

Editors of Grammar Modules

English	Kazuyuki URATA, Keizo NOMURA
German	Takashi NARITA
French	Yuji KAWAGUCHI
Spanish	Shigenobu KAWAKAMI
Portuguese	Naotoshi KUROSAWA
Russian	Hidehiko NAKAZAWA
Chinese	Takayuki MIYAKE
Korean	Hideki NOMA
Mongolian	Renzo NUKUSHINA
Indonesian	Masashi FURIHATA
Filipino	Michiko YAMASHITA
Lao	Reiko SUZUKI
Cambodian	Hiromi UEDA, Tomoko OKADA
Vietnamese	Yoshio UNE
Arabic	Robert RATCLIFFE
Turkish	Mutsumi SUGAHARA
Japanese	Tae UMINO, Futoshi KAWAMURA

Example page of studying French grammar in Japanese

Linguistics:

Corpus-based analysis of sentence structures and cross-linguistic perspectives in phonetics

Name	Topic
Yoichiro TSURUGA	Impersonal constructions in French
Yuji KAWAGUCHI	Diachronic research on negative constructions in French
Naotoshi KUROSAWA	Word order of modifier and modified constituent in Latin and Portuguese
Kiyoko SOHMIYA	Aspects of marked constructions as seen in corpora
Kazuyuki URATA	Diachronic research on the subjunctive in English
Susumu ZAIMA, Takashi NARITA	Corpus-based research on verb construction in German
Toshihiro TAKAGAKI	Construction of a Spanish corpus and the development of relevant tools to advance Spanish language research
Hidehiko NAKAZAWA	Utilization of a corpus for research on Russian verbs
Takayuki MIYAKE	Research on the syntactic characteristics of Chinese verbs based on corpus analysis
Keiko MOCHIZUKI	Comparative study of compound verbs in Japanese and Chinese that express "causal phenomena" and "resultant phenomena" and their corresponding English sentence structures
Shinjiro KAZAMA	Descriptive study of grammar using spoken and literary corpora
Isamu SHOHO	The causes and results of marked word order in the Malaysian language
Satoko YOSHIE	Construction of a Wakhi vocabulary corpus
Shinji YAMAMOTO	The Italian language in the 21st century
Yuji KAWAGUCHI, Naotoshi KUROSAWA, Shigenobu KAWAKAMI, Tsunekazu MORIGUCHI, Yoshio SAITO	Prosodic analysis of speech through the TUFS Dialogue Modules — Review of the content and the potential applications of the Pronunciation Module —

Applied Linguistics:

Discourse Analysis, Second Language Acquisition, Evaluation of TUFS Modules

Name	Topic
Mayumi USAMI	Construction and analysis of a multilingual corpus of spoken language. Basic research on methodology for natural conversation analysis. Development of a basic transcription system for Japanese, Korean, Chinese and English
Tae UMINO	Construction and analysis of Japanese learner-language corpus. Basic research aimed at the development of learner's manual for 'Japanese Dialogue Module'
Asako YOSHITOMI	Construction of an English learner language corpus. Revision of the English Dialogue Module teacher's manual
Masashi NEGISHI, Hideyuki TAKASHIMA, Masanori ICHIKAWA, Koyo YAMAMORI	Development of a Language Proficiency Scale

Computer Sciences:

E-learning, Natural Language Processing

Name	Topic
Hiroshi SANO, ChunChen LIN	Construction of an educational material corpus for Japanese language education. Construction of an English corpus

First International Conference on Linguistic Informatics

Date:	December 13 (Sat.) – 14 (Sun.), 2003
Venue:	Tokyo University of Foreign Studies, Research-Lecture Building, Multimedia Hall

The First International Conference on Linguistic Informatics was held at Tokyo University of Foreign Studies in December 2003.

Participants at the conference not only learned about the new integrated field of study known as Linguistic Informatics and its current status, but also considered its future potential. The conference was comprised of three sessions (please see below), and everything from reports to question and answer sessions was conducted in English.

1. Computers and Linguistics ····Potential for collaboration between computers and linguistics
2. Corpus Linguistics ················Status report on corpus-based linguistic research
3. Applied Linguistics ················Relationship between second language acquisition and linguistic theory

Current status of natural dialogue-based discourse analysis

Researchers from both at home and abroad conducted lectures and presented reports, and there were also a number of presentations given by graduate students. Proceedings were published prior to the holding of the conference. There were over 300 people in attendance at this two-day conference, and the event was brought to a successful close at the reception following the end of the conference.

The report collection for this conference will be published domestically in Vol. 3 of the Linguistic Informatics Series and will also be published with new content and in a different format by John Benjamins Publishing Company in the Netherlands.
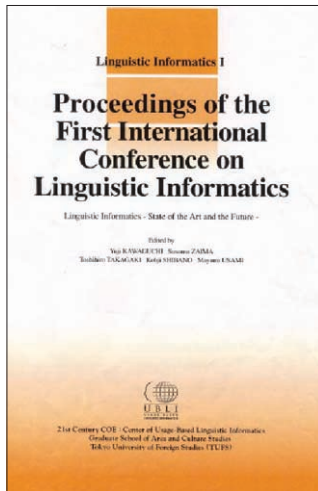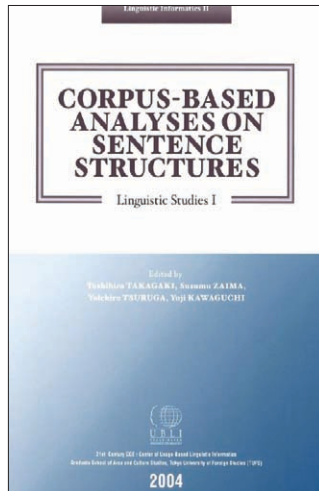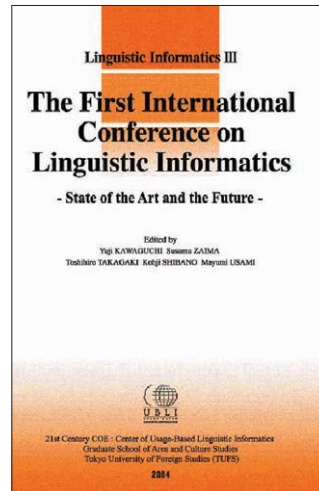

Poster
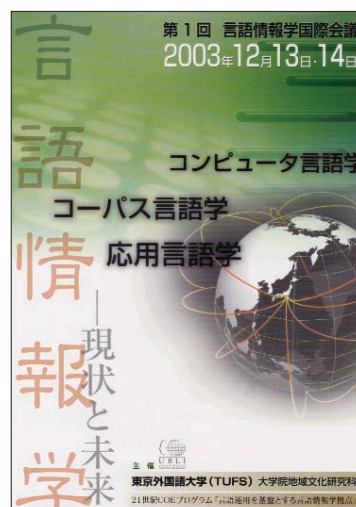
Program

Publications

Series : LINGUISTIC INFORMATICS

Vol.I 2003.10	Vol.II 2004.4	Vol.III 2004.10

To be published with new content and in a different format by

John Benjamins Publishing Company, Amsterdam, the Netherlands

Series : WORKING PAPERS IN LINGUISTIC INFORMATICS

Vol.I 2004.3	Vol.II 2004.3	Vol.III 2004.9	Vol.IV 2004.10