

1-21-2006

Institute for Language Information and Technology (ILIT)

Office of the Provost

INSTITUTE FOR
LANGUAGE INFORMATION AND TECHNOLOGY
(ILIT)

Proposal
To Create a New Institute
At Eastern Michigan University

Submitted to the
Office of the Provost

By

HELEN ARISTAR-DRY

Professor
Department of English Language and Literature

The LINGUIST List
613P Pray Harrold
Eastern Michigan University
Ypsilanti, MI 48197

Tel: 734 487-7952
Fax: 734 483-9744
hdry@linguistlist.org

January 21, 2006

Support Signatures:

Russell Larson, Head
Department of English Language and Literature

[date]

Hartmut Höft, Dean
College of Arts and Sciences

[date]

Robert Holkeboer, Associate Vice President
Graduate Studies and Research

[date]

Brian Anderson, Director
Office of Research Development

[date]

EXECUTIVE SUMMARY

This proposal suggests that an Institute for Language Information and Technology be formed at Eastern Michigan University, in order to take advantage of the presence of The LINGUIST List at EMU. The LINGUIST List is an Internet network and research facility which has become the electronic center of the discipline of linguistics, serving both as a central information resource for the field and as a laboratory for research and development of language technologies. With 21,500 subscribers, a 100,000-page website, and 4 international mirror sites, LINGUIST has provided international visibility for its two university hosts, Eastern Michigan University and Wayne State University, and has been widely recognized as a leader in the burgeoning field of language technology. LINGUIST's research into the digital archiving of language data has produced digital tools and standards which have been adopted by the International Standards Organization and by the National Science Foundation. To date, LINGUIST has been awarded 12 National Science Foundation grants valued at \$4,628,160, recovered more than \$1,031,817 in indirect costs, and generated an additional \$678,176 from development activities and direct donations.

Thus LINGUIST List is well-positioned to fill a growing need in industry, academia, and government for technology to support the production, understanding, and analysis of human language. However, LINGUIST's expansion is currently constrained by the need to share resources between two universities and by the lack of stable funding, which encourages staff turnover and the pursuit of short-term projects to the detriment of long-range development. Hence, this document proposes that LINGUIST operations be amalgamated at EMU through the establishment of The Institute for Language Information and Technology (ILIT) and the creation of a faculty position for Professor Anthony Aristar. Aristar, who currently directs The LINGUIST List operation at WSU, will become, with Helen Aristar-Dry (already faculty at EMU), a Co-Director of the new Eastern Michigan Institute.

Institute status and the guarantee of a base level of support from Eastern Michigan University will allow The Institute for Language Information and Technology to pursue a more robust research agenda, at the same time as it consolidates The LINGUIST List's prestige, financial resources, and educational potential at EMU. Building on LINGUIST's successes, ILIT may be expected to become a focus of collaboration between Linguistics and Computer Science, to continue to attract high-caliber students to both programs, and to motivate and support the development of an advanced degree in Language Technology.

PROPOSAL SUMMARY

Introduction

In 1990, Anthony Aristar founded The LINGUIST List at the University of Western Australia as an internet forum where academic linguists could discuss linguistic issues and exchange linguistic information. Within months, Helen Aristar-Dry became co-moderator of The LINGUIST List, which at that time had 69 subscribers. In 1990, Aristar-Dry moved to EMU from the University of Texas at San Antonio and Aristar left Australia for Texas A&M University. In 1992, the LINGUIST List website was launched at EMU, with an additional editing site at Texas A&M. In 1998, Aristar joined Wayne State University, which replaced Texas A&M as the second editing site of the List and became a collaborating institution on future language technology projects and the grants awarded to fund this research.

During this period, LINGUIST was evolving from a mailing list into a center for language technology and information. One of the oldest academic organizations on the Internet, LINGUIST's early utilization of web technology rapidly positioned it as the central source for electronic information in the discipline of linguistics. Acquiring more and more technical expertise as its functions expanded, LINGUIST eventually evolved into an organization whose major focus was to develop digital infrastructure for the field. Today, the LINGUIST List is a combined website, email list, and research organization dedicated to providing the general public with information on language and linguistics, and to providing the discipline of linguistics with digital tools and standards for the analysis and archiving of language data.

The LINGUIST List has grown from an initial 69 subscribers to 21,500 subscribers in 108 different countries, and from generating 200 messages a month to generating 87,000 email messages a day. Its primary website (<http://linguistlist.org>) has over 100,000 pages, and sustains over 300,000 page views a week. It has four full mirror sites (at the U. of Stockholm, U. of Edinburgh, U. of Tübingen, and Melbourne U. in Australia); and together its 5 websites receive over 3 million views per week. The LINGUIST List also hosts 129 other linguistics mailing lists and has achieved recognition for innovative applications of Internet technology to linguistic research. This recognition includes 12 grants from the National Science Foundation, an Academic Equipment Grant from Sun Microsystems, and 3 awards from the Linguistic Society of America, including the 2003 Victoria A. Fromkin Award for Extraordinary Service to the Discipline. The National Science Foundation grants alone are valued at \$4,628,160, and have generated more than \$1,031,817 in indirect cost recovery. In addition, LINGUIST has generated \$678,176 from other development activities and fund-raising efforts.

Although it retains its unassuming name, The LINGUIST List is now a major institution within the field of linguistics; and it has gained worldwide acclaim for the University. Media sources including *The Wall Street Journal*, ABC News, *The Smithsonian Magazine*, *Die Zeit* (Germany), *Corriere della Sera* (Italy), and *Science et Vie* (France) have reported on the activities of The LINGUIST List. Most recently, *Discover* magazine has requested information on its E-MELD project on endangered languages. On the Web, the National Science Foundation (NSF), Open Languages Archives Community (OLAC), and World Wide Web Consortium (W3C) websites devote space to the LINGUIST List. The National Science Foundation in its Endangered

Linguistics projects now mandates standards developed by LINGUIST, and its work on ancient languages has now become part of the new ISO-639-3 language standard.

LINGUIST has had a significant impact on EMU students and fueled the creation of new course and degree offerings within the Linguistics Program. Because of it, the Linguistics Program has grown from a handful of part time students to a program of 16 full time graduate students, most of whom go on to top-flight Ph.D. programs in linguistics. EMU continues to attract high caliber linguistics students from around the world because LINGUIST List provides both funding and high-profile work in language technology. The List's income from grants, publishers, employers, and subscriber donations, helps Linguist offer 20 fellowships to graduate students who in return serve as its editors; 13 of these fellowships are currently held at EMU.

The LINGUIST List staff, external collaborators, and technical capabilities have evolved to the point where an Institute with an increased annual revenue is a logical and feasible next step. Linguist has so far been merely a "faculty project," that is, a project undertaken without university sponsorship and only limited university support. Restricted to this status within the English Department, with income shared with Wayne State University, and without General Fund support, LINGUIST List faces severe limitations constraining its future development. The year-to-year uncertainty of present funding operations limits LINGUIST's ability to undertake long term projects and requires that a significant portion of resources and administrative effort be devoted to short term fund-raising.

However, with a director and co-director, project management and programming staff, and with all activities under the roof of one institution, the proposed Institute would have the long-term stability necessary to actively pursue funding for a more robust research and development agenda. The current ad hoc administrative structure places great pressure on the staff, constraining their ability to secure the LINGUIST List's future and expand its activities in ways already called for by the discipline.

THE NEED FOR AN INSTITUTE

The Institute for Language Information and Technology (ILIT) will provide the infrastructure necessary to allow current LINGUIST List activities to expand, and it will strengthen EMU's already prestigious Linguistics Program. At the same time, it will address pressing national and international needs within academia, industry, government, and education.

National and International Needs

Academia: The field of linguistics is newly recognizing that it requires discipline-specific computational tools to support its traditional pursuits. The discipline has long required electronic infrastructure to maintain communication among linguists, support collaborative research, disseminate new discoveries, secure positions for graduates, run conferences, maintain professional societies, and provide accurate language information to the public. And since 1990, The LINGUIST List has filled this need via an exemplary academic website that is often cited as a model for other fields. But now, even more urgently, the field requires new digital tools and standards to support documentary linguistics and the electronic archiving of language data.

Documentary linguistics is the collection and annotation of primary language data showing the full range of linguistic activities within a community. It is one of the oldest subfields of linguistics; but it has been given new prominence by a growing awareness of the need to document endangered and dying languages before they disappear. Recording, analyzing, and displaying examples of such languages in their social context is increasingly recognized as an enterprise of cultural and scientific importance, and it requires the development of specialized software to handle international character sets, multilingual texts, and multimedia displays of language data in complex configurations (e.g., interlinear glossed text). Moreover, the applications must be designed for maximum flexibility and transparency, for use by scholars who are not computer experts and who may be working in extreme weather conditions using many different types of hardware.

The need for such research and development continues to increase, as is indicated by the recent establishment of several new funding initiatives focused on language documentation, e.g., the Documentation of Endangered Languages (DEL) initiative implemented by the National Science Foundation, the National Endowment for the Humanities, and the Smithsonian; the Hans Rausing Endangered Languages Documentation Programme; and the Dokumentation Bedrohter Sprachen (DoBeS) project funded by the Volkswagen Stiftung. Its federally-funded research projects have already made The LINGUIST List a leader in this new field. The implementation of ILIT will ensure that EMU does not lose the momentum generated by LINGUIST's earlier achievements.

Industry: Software developed to meet needs within linguistics has significant potential for technology transfer to two major groups within industry: (1) software developers interested in multilingual computing and the globalization of existing applications, and (2) service industries, which are increasingly exploiting the computer's potential for processing, understanding, and generating language. Now, as never before, linguists are playing a vital role in the development of new service technologies, from automated reservation systems to talking appliances. For this reason, the field of language technology has generated a disproportionate number of the recent employment opportunities in linguistics (32% of 1351 jobs posted since 2002), even though it is just one of many linguistic subfields. ILIT will feed industry, both through the development of multilingual software and through educational activities. Ongoing collaboration between ILIT and EMU's degree programs (e.g., the M.A. in Language Technology and the Post-M.A. Certificate in Language Technology) will provide a training environment which uniquely integrates study with practical internships focused on ground-breaking applied research.

Government and education: Among world leaders, the United States lags far behind in knowledge of other languages. In spite of America's increasing emphasis on diversity and inclusion, most Americans—even those in key posts in government, business, and education—have relatively little information about the world's lesser known languages.¹ And yet, as is being increasingly recognized in both government and education, ignorance of other languages

¹ How many Americans know that there are 43 languages spoken in Afghanistan? That the Arabic spoken in Morocco cannot be understood in Iraq? Or that Indonesia, the world's largest Muslim nation, has 736 languages within its borders? Most Americans would not even know how to find such information.

hampers the United States as it attempts to respond effectively both to international threats and to opportunities for international cooperation.

Governmental and educational initiatives such as the federally-funded National Science Digital Libraries (NSDL) project have recognized the need to increase public access to authoritative linguistic information, taking full advantage of electronic methods of processing and disseminating multilingual data. With its interlinked databases developed through 15 years of information-gathering, The LINGUIST List is already recognized as an unparalleled source of language information. Funded by an NSDL grant from the Dept of Education, Linguist is already contributing language information to the National Science Digital Library. With expanded resources, ILIT will significantly heighten EMU's contributions and visibility within such important national initiatives.

Summary of national and international needs:

The Institute for Language Information and Technology will address four external needs:

- (1) the need within academia for electronic infrastructure and for digital tools to aid in the collection, analysis, and archiving of language data.
- (2) the need within industry for language technology to aid in globalization and in the development of applications which process and produce human language.
- (3) the need within government and education for accessible sources of accurate information about languages and cultures, particularly lesser-known languages.
- (4) the need within academia, industry, and government for individuals trained in both linguistics and computer science—individuals who are able to meet the challenges of an increasingly multilingual and technology-centered world.

LINGUIST List has a long and successful track record in meeting the needs of government and academia, as shown by its succession of grant-funded projects. As noted above, LINGUIST contributes to the National Science Digital Library through the ALL Language project. Moreover, the DATA project and the 5-year \$2.2 million E-MELD project have given LINGUIST authority and visibility within the international movement to preserve endangered languages. LINGUIST List moderator Aristar-Dry currently serves on the boards of the Open Language Archives Community, the ISLE (International Standards in Language Engineering) Metadata Initiative, and the Digital Endangered Languages and Musics Archive Network (DELAMAN), as well as the International Standards Organization Metadata Working Group. As an outgrowth of DATA and E-MELD, LINGUIST has begun to curate an archive of endangered languages data which has the potential to grow into a major international resource.

LINGUIST has also demonstrated its potential as a training organization. LINGUIST attracts high caliber students from around the world because it provides both funding and high-profile work in language technology. To date, LINGUIST has supported 43 students through their M.A.

degrees; and fully two-thirds of these are pursuing advanced degrees in prestigious Ph.D. programs. In most cases, the visibility of their LINGUIST List work and their language technology skills have been instrumental in distinguishing them from other applicants and gaining them admission. Moreover, LINGUIST has played a crucial role in the establishment of EMU's new Certificate in Language Technology and in the M.A. in Language Technology currently under development.

To date, LINGUIST has not had the resources to pursue technology transfer opportunities which would address needs within industry. But an Institute with stable funding and more senior staff would be able to forgo some of the short term funding pursuits that have hitherto absorbed energy and resources better devoted to long range development.

Internal Needs

With 26 employees and 5 distinct language technology projects currently underway, in addition to contract web development, LINGUIST List has outgrown its present infrastructure. But its international prominence and research opportunities continue to increase. The creation of an institute has thus become imperative for reasons of internal organization. Without an institute's stability and administrative independence, LINGUIST cannot fulfill its demonstrated potential.

Implementation of an institute, on the other hand, will address multiple internal needs and offer numerous specific benefits to LINGUIST and to the institution:

- An institute will realize the potential of The LINGUIST List and its founders to expand its various income producing activities and to pursue a more robust research, development, and training agenda, one that builds on its solid funding record. This potential is currently limited by Linguist List's status as a faculty project within the Department of English Language and Literature at EMU, by the shared/split management of research grants with Wayne State University, and by the absence of bridge funding between grant cycles.
- An institute will help to ensure that The LINGUIST List continues beyond the eventual retirement of its founders. This disciplinary concern was a major impetus for the request by The Linguistic Society of America that LINGUIST develop and host the Society website.
- An institute will provide the administrative, marketing, and clerical support necessary for The LINGUIST List to solidify and expand its leadership role as an innovator in language technology.
- An institute will carry more weight with government and private grantors, since an institute is the appropriate home for the varied service, research, and training functions performed by The LINGUIST List. These functions are not adequately reflected in its current name, organizational structure, and position within the English Department.
- An Institute will have more decision making authority, allowing it to respond quickly to requests for training and technology transfer from government agencies, the private sector, or academia.

- Institutionalization will facilitate coordination with other EMU departments, universities, the business world, and government agencies.
- The activities of the institute will secure an international reputation for EMU and its linguistics program, at the same time as they provide financial support for an expanded graduate program.

Funding requested during the first three years of the Institute operations is \$232,431 per year. An additional \$68,833 would be needed if grant funding is unavailable in a given year.

GOALS

The Institute for Language Information and Technology (ILIT) is committed to growing into a leading institute providing linguistics research, services, and education for academia, government, industry, and the general public. The goals of the ILIT are to:

- Conduct applied research in language technology.
- Provide training in language-related technologies.
- Serve as a comprehensive source on the world's languages.
- Develop and maintain an Internet infrastructure for the discipline of Linguistics
- Provide a rich learning and research environment for EMU faculty and students.

OBJECTIVES AND STRATEGIES

Objectives

- Pursue applied research in language technology to (a) develop computational tools to support the work of language documentation and analysis and (b) develop standards, procedures, and facilities to support electronic archiving of language data.
- Develop Internet infrastructure for the discipline of linguistics, including (a) enhancing our interlinked databases of information on jobs, conferences, books, journals, academic papers, linguists, and dissertations, (b) hosting mailing lists, and (c) developing and hosting websites for professional societies.
- Create a comprehensive language information center that merges existing dispersed databases, creates new content, and disseminates accessible digital resources on all world languages.

- Curate an archive of data on endangered languages.
- Provide language technology training to the field, businesses, government agencies, and the general public
- Provide information about languages and the work of ILIT to government agencies and the general public.
- Provide training and internships for EMU students.

Strategies:

- Increase proposal-writing capacity for both short and long term projects.
- Research new funding sources.
- Increase income from the LINGUIST email list and website through membership fees and more aggressive marketing of its advertising services for publishers and of its jobs service for employers in language technology.
- Develop a conversion service for legacy data on endangered languages
- Develop an electronic infrastructure model based on The LINGUIST List website to be marketed to other fields or organizations.
- Conduct ongoing “digital institutes” to formulate and review recommendations of best practice and to train a substantial core of linguists and language archivists in their implementation.
- Develop a marketing plan to advertise training and consulting to the field, to businesses, government agencies, and the general public interested in linguistics, language technology, or Internet applications.
- Provide press releases to the media to increase public awareness of the accomplishments and activities of ILIT.

Relevance to university:

The goals and objectives of ILIT mesh firmly with three of the Strategic Directives of the College and University. In particular, ILIT will promote:

- *Strategic Direction 2: Eastern Michigan University will be recognized for the synergy of theory and practice in its graduate programs.*

ILIT will provide signal opportunities to students to be involved in cutting-edge research in both Documentary Linguistics and Language Technology.

- *Strategic Direction 5: Eastern Michigan University will become a university with global and multicultural perspectives.*

The LINGUIST List is international in research orientation and collaboration, in student recruitment, and in outreach to the linguistics community worldwide. LINGUIST List student research assistants have done independent fieldwork in Ghana, Ivory Coast, and Argentina; and, under the DATA grant, LINGUIST has sent 3 students to Alaska to document Dena'ina, an Athabaskan language. The LINGUIST List also recruits its student staff through international advertisement of the "LINGUIST Assistantships"; and this year its staff contains students from Japan, People's Republic of China, Great Britain, Sweden, Russia, Taiwan, India, and the Ukraine. Finally, LINGUIST's current subscribers reside in 107 different countries and its mirror sites are maintained in 4 different countries on 3 continents. This worldwide reach gives linguistics at Eastern Michigan University extraordinary international visibility.

- *Strategic Direction 6: Eastern Michigan University will improve institutional effectiveness.*

ILIT's externally-funded projects in Language Technology will help the university to secure "state of the art facilities and the resources to maintain them," as well as providing a certain amount of "staff development, especially in the use of technology."

Additionally, through the LINGUIST List and its associated projects, students interact electronically with professional linguists across the world, collaborate as full members of Language Technology research teams composed of scientists from major American, European, and Australian universities and institutes, and report on their research at national and international conferences.

COLLABORATION WITH ACADEMIC UNITS

LINGUIST List has a long-standing history of successful collaboration with students and faculty in the Computer Science Department. A Computer Science faculty member, Prof. John Remmers, designed LINGUIST's initial suite of custom software and has regularly served as a database and training consultant. Through the years, LINGUIST has employed 10 CS graduate students as Research Assistants, in addition to providing internships to 6 students taking CS courses. In the last 4 years, LINGUIST has also employed Accounting students to manage its finances and Technical Writing students to produce organizational documentation.

INFRASTRUCTURE

ORGANIZATION

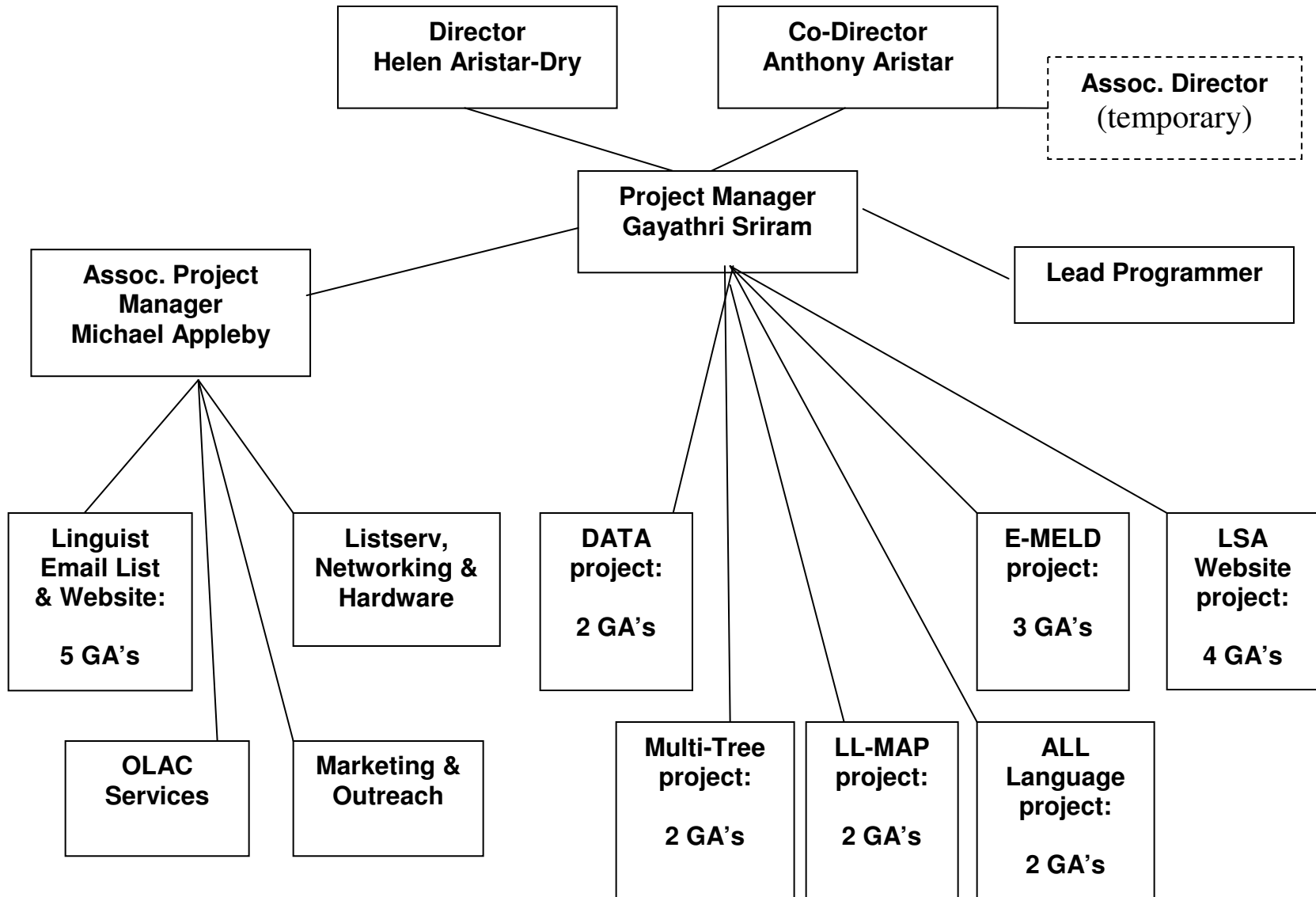
Currently Linguist List employs 26 people at 3 universities (including a book review site directed by two faculty review editors and a graduate assistant at the U. of Arizona). LINGUIST has 6 distinct externally-funded projects, in addition to its extensive core operations.

Organizational Chart of the Institute

As suggested by the following organizational chart, ILIT will provide the administrative staff and organization to effectively manage both external projects and the core infrastructure facilities for the discipline.

See Organizational Chart next page.

Figure 1: ILIT Organization



Institute Personnel

Co-Directors and Administrative Staff

Helen Aristar-Dry, Ph.D. Co-Director, Professor Linguistics Program, EMU
Anthony Aristar, Ph.D. Co-Director, Professor Linguistics Program, EMU
Associate Director (to be determined), Professor Linguistics Program, EMU
Gayathri Sriram, Project Manager
Michael Appleby, Associate Project Manager
Programmer (to be determined)
Administrative Associate (to be determined)

Faculty Associates

Daniel Seely, Ph.D., Professor, Linguistics Program, EMU
Beverley Goodman, Ph.D., Associate Professor, Linguistics Program, EMU
Veronica Grondona, Ph.D., Associate Professor, Linguistics Program, EMU
Ed Garrett, Ph.D., Assistant Professor, Linguistics Program, EMU

Faculty Associate status will be open to distinguished researchers from Eastern Michigan U., Wayne State U., and other collaborating universities and organizations.

Equipment

Currently, LINGUIST has 2 enterprise-level Sun servers and 2 Sun Ultra Sparc workstations housed in Halle Library. These were purchased by LINGUIST, using NSF grant funds and an Academic Equipment Grant from Sun Microsystems. The oldest machine, the Sun Enterprise 450, is nearing the end of its working life; and the newer machine, the Sun V880 is currently overloaded. Application has been made for another equipment grant from Sun Microsystems.

LINGUIST also has 16 desktop computers and a Dell laptop in the LINGUIST office at 613P Pray-Harrold. The laptop requires replacement, but the other equipment is adequate for the foreseeable future.

Location

LINGUIST List is currently located in 613P Pray-Harrold but seeks a larger working space. Needed is space for 18 student work stations, 3 semi-private cubicles for full-time managers and programmers, and a small meeting room with projection facilities to support the 8-10 project planning and review meetings which take place each week.

BUDGET AND SOURCES OF REVENUE

START-UP AND ONGOING RESOURCE NEEDS

- **12-month salary for Director, Helen Aristar-Dry.**
Typically 1-2 months of Aristar-Dry's summer salary is supplied by grant funds. However,

LINGUIST List is fully operational 355 days per year, requiring 12-month oversight by the Directors, whether or not grant projects are actively being pursued.

- **Faculty position for Anthony Aristar, Co-Director (12-month salary).**

Aristar is currently employed at WSU. Hiring Aristar at EMU will allow ILIT to consolidate resources and projects at EMU, with a consequent increase in revenue to offset Institute costs.

- **Temporary position for an Associate Director.**

The Associate Director should work with the current Directors for at least 2 years before assuming the position of Director, upon the Directors' retirement. This position should be established within the next 4 years.

- **Salary for the Project Manager, Gayathri Sriram.**

Retaining experienced managers between grant periods is imperative because of the highly specialized nature of the research. Sriram has functioned as technical director of the organization for 6 years; her retention is crucial to its ongoing effectiveness.

- **Salary for the Associate Project Manager, Michael Appleby**

Two full-time management positions are necessary because the 24-member LINGUIST staff consists primarily of students, who work for only 2 years, while pursuing their M.A. degrees. Undertaking professional software development with an untrained staff and constant staff turnover places and unusually heavy training and management burden on the project managers. Appleby (currently employed through WSU) has been with the organization for 4 years; and losing him would be highly detrimental to ongoing projects and to ILIT's ability to attract future funding.

- **Half of the salary for a professional programmer (bridge funding).**

Although ILIT will continue to employ linguistics and CS students and train these in programming, ILIT will require the skills of a full-time programmer experienced in software development. Half the programmer's salary will be provided through ongoing LINGUIST List funding. It is expected that grant funding will provide the other half of the salary. However, in the event that grant funding is unavailable in a given year, bridge funding will be required because of the highly specialized nature of the research. (Lack of bridge funding recently resulted in a loss of programming staff which has severely impeded current project work.)

- **Salary for a part-time administrative associate (20-hours a week, PT rate)**

Currently financial administration is being attempted by a GA. However, administration of multiple grant accounts requires skills and experience not ordinarily found in graduate students; moreover two-year turnover in this position severely impedes project efficiency.

- **Funding for 5 graduate assistants (12-month, tuition and stipend).**

Currently LINGUIST employs 20 graduate students, including 6 at WSU and 1 at U. of Arizona. However, funding is being requested only for the minimum number of students necessary to maintain core Institute operations, including The LINGUIST List website, LL email list, the central search engine for the Open Language Archives Community (OLAC), and the hosting of other linguistics mailing lists and websites (e.g., the Linguistic Society of America). It is expected that external grants and contracts will continue to fund the additional students required for ILIT projects.

- **Internet connectivity, office space, maintenance contracts for servers, and 10 hours a week sysop time.**

Requested Budget:

Appendix A shows the total annual budget requested for ILIT for the next 3 years. As explained below, the annual budget shown in Appendix A (\$379,483) includes both resources already in place and funding which will be requested only in years when grant funding is pending. When existing and bridge funding are deducted (see Table 3), the total annual amount requested is \$232,431 for Year 1, with a projected 3% annual increase for years 2 and 3.

It should be noted that Appendix A does not list the considerable funding which will be provided from LINGUIST's own resources, e.g., funding for the 15 additional GA's which compose the current 20-member student crew, and funding for half of the salary of the full-time programmer. Nor does it list the additional programmer(s) required to complete software development projects, as these will be funded from LINGUIST resources and project income.

A significant part of the budget requested in Appendix A represents resources already in place. Of the \$379,483 requested annual budget, \$78,119 represents resources already being supplied by various EMU administrative units. See Table 1 below:

Table 1: Current EMU Support	
Stipend for Graduate Students – 8 month (2* \$8500) ¹	17,000
Tuition for Graduate Students (3 * \$12384 & 1* \$6467) ²	43,619
Maintenance contracts for LINGUIST servers ³	5,000
.25 FTE for ILIT Systems Administrator (PT9) ⁴	12,500
TOTAL	78,119
<i>¹ Currently provided by English Dept</i>	
<i>² Currently provided by Graduate School</i>	
<i>³ Currently provided by the College of Arts and Sciences</i>	
<i>⁴ Currently provided by ICT</i>	

If the current College, Graduate School, and Department support continues, the annual budget requested by ILIT from the General Fund will be \$301,364.

Furthermore, as shown in Table 2, part of the requested budget may be expected to be supplied, all or in part, by grant funding. Table 2 shows that portion of the requested budget which may be considered bridge funding, to be requested only in years when grant funding is pending.

Table 2: Bridge Funding	
.5 FTE for ILIT programmer (PT9)	\$25,000
1.5 mo. Spring / Summer Salary for Aristar-Dry	13,365
1.5 mo. Spring / Summer Salary for Aristar	12,210
Fringe benefits for ILIT programmer	9,025
Fringe benefits for Aristar-Dry	4,825
Fringe benefits for Aristar	4,408
Total Bridge Funding	68,833

As shown in Table 2, an additional \$68,833.00 of the total requested budget shown in Appendix A may be considered bridge funding. In each of the preceding seven years, this amount has been provided from grant funding.

The requested annual budget for ILIT may thus vary from the total \$379,483 detailed in Appendix A to \$232,431. Table 3 summarizes this calculation.

Table 3: Summary	
Annual Requested Budget (Appendix A)	\$379,483
Resources already supplied by EMU	-\$78,119
Subtotal	\$301,264
Bridge Funding (usually supplied by grants)	-\$68,833
Expected annual request	\$232,431

Prior Funding

As detailed in the appended spreadsheet (Appendix B), in its 15 years of operation LINGUIST List has brought in \$5,311,336.00 from external sources. This total includes both grant funds and monies netted from other development and fund-raising activities.

Grants: To date, LINGUIST List has been awarded over \$4.5 million in National Science Foundation grants and recovered more than \$1,031,817 in indirect costs, which have been shared among its supporting universities. NSF grants have included:

NSF grant to customize mailing list software awarded 1993, \$4,000

NSF grant to produce an enhanced web site and search facilities, 1996, \$109,000

NSF grant to create a searchable archive of other linguistics-related mailing lists, 1999, \$169,000

NSF grant for configuration of database to support endangered language initiatives, 2000, \$50,000

NSF grant for workshop “Endangered Languages Digitization: The Need for Standards,” 2000, \$40,165

NSF Social Sciences Infrastructure Grant: “E-Meld: Electronic Metastructure for Endangered Languages Data,” 2001–2006, \$2,242,914.

NSF Dena’ina Grant, “DATA: Dena’ina Archiving, Digitization and Access,” 2003–2005, \$394,663

NSF Collaborative Project: ALL Language grant, 2003–2004, \$96,525

DHB Collaborative Research: LL-MAP, 2005–2008, \$742,669

Collaborative Research: Multi-Tree: A Digital Library of Language Relationships, 2005–2008, \$485,024

Total Grants, \$4,633,160

Total Indirect Costs, \$1,031,817

Other funding: To date, LINGUIST List has generated an additional \$678,176 from activities such as the following:

- Web Hosting for institutions, associations, organizations. *Currently the Linguistic Society of America pays \$18,000 per year for website hosting and technical support, in addition to a one-time \$65,000 web development fee.*
- Publisher services and book announcements. *To date, in 2005 revenue from this source has exceeded \$33,000.*
- Job Announcements, for the academic, government, and private sector. *To date, in 2005 revenue from job announcements has exceeded \$35,000. More marketing and outreach activities could increase this income, as would development of facilities for hosting and distribution of job-seeker resumes..*
- Fund raisers, direct donations. *In 2005 Linguist raised \$61,207 in direct donations.*

Funding Prospects

Lack of senior administrative staff has hitherto limited LINGUIST’s ability to seek additional funding, but ILIT will be able to leverage LINGUIST’s research in digital archiving into grant funding from other government agencies, e.g., NEH (specifically the Division of Preservation and Access) and The Institute for Museum and Library Services (IMLS), as well as private foundations interested in the preservation of endangered languages. Prospective sources of funding thus include:

- **National Science Foundation (NSF)**
- **National Endowment for the Humanities (NEH)**
- **The Institute of Museum and Library Services (IMLS)**

- **Private Foundations**
- **Department of Defense**
- **Department of Education**

In addition, an Institute with a full-time, experienced administrative staff will be able to develop sources of funding, through:

- **Specialized training and consulting** in language technology for business and government
- **Technology transfer** to business, other universities, and associations
- **Linguist List website and list membership.** *Currently membership is free. However, developing additional services and requesting a minimal fee for these (e.g., \$5 a year for individuals and \$15 per year for institutions) could increase yearly income by \$100,000.*

CONCLUSION

With institutional support, ILIT will become a leading institution for the creation and dissemination of digital tools for language processing. It will be an innovator in the development of standards and procedures for the digital archiving of language data. It will also serve as a comprehensive public and government information source on the world's languages, and it will manage an archive of data on endangered languages. It will function as a center for training in language-related technologies, including the development of best practices for the field; and it will continue to host the central metadata server for the discipline of linguistics. It will also continue the development of a highly sophisticated Internet infrastructure, which has the potential of technology transfer to other fields and organizations. These capabilities will allow ILIT to strengthen EMU's already prestigious Linguistics Program, at the same time as it meets identified needs within academia, government, and industry.

Appendix A: Budget for EMU Institute for Language Information and Technology			
(Does not include funding provided by LINGUIST)			
	Year 1	Year 2	Year 3
Students			
Stipend for 5 Students (12 month)	\$56,735.00	\$58,437.05	\$60,190.16
Salaries			
1.0 FTE Project manager (PT 10)	\$65,000.00	\$66,950.00	\$68,958.50
1.0 FTE Associate Project Manager (PT8)	\$45,000.00	\$46,350.00	\$47,740.50
.5 FTE Administrative Asst. (PT7)	\$18,000.00	\$18,540.00	\$19,096.20
.5 FTE Programmer (BF) (PT9)	\$25,000.00	\$25,750.00	\$26,522.50
Aristar-Dry's Summer salary (BF)	\$26,730.00	\$27,531.90	\$28,357.86
Aristar's Summer salary (BF)	\$24,420.00	\$25,152.60	\$25,907.18
.25 FTE for Systems Administrator (PT9)	\$12,500.00	\$12,875.00	\$13,261.25
Total Salaries	\$273,385.00	\$268,711.55	\$276,772.90
Fringe benefits			
Aristar-Dry's Fringe Benefits	\$9,649.53	\$9,939.02	\$10,237.19
Aristar's Fringe Benefits	\$8,815.62	\$9,080.09	\$9,352.49
Project Manager Fringe Benefits	\$23,465.00	\$24,168.95	\$24,894.02
Assoc. Project Manager Fringe Benefits	\$16,245.00	\$16,732.35	\$17,234.32
Admin. Asst. Fringe Benefits	\$6,498.00	\$6,692.94	\$6,893.73
Programmer Fringe Benefits	\$9,025.00	\$9,295.75	\$9,574.62
Total Fringe Benefits	\$73,698.15	\$75,909.09	\$78,186.37
Total Wages & Fringes	\$347,083.15	\$344,620.64	\$354,959.26
Tuition & Fees			
3 Students (out-of-state)	\$37,152.00	\$38,266.56	\$39,414.56
2 Students (in-state)	\$12,933.00	\$13,320.99	\$13,720.62
Other expenses			
Travel & Marketing	\$5,000.00	\$5,150.00	\$5,304.50
Material and Supplies	\$2,400.00	\$2,472.00	\$2,546.16
Computer hardware updates	\$10,000.00	\$10,300.00	\$10,609.00
Server maintenance contracts	\$5,000.00	\$5,150.00	\$5,304.50
Software updates & licensing fees	\$5,000.00	\$5,150.00	\$5,304.50
Total Other Expenses	\$27,400.00	\$28,222.00	\$29,068.66
Grand Total	\$379,483.15	\$377,992.64	\$389,332.42

Appendix B: LINGUIST List Income 1993-2005

LINGUIST List NSF Grants

Grant Name	Grant Number	Year Awarded	Indirect Cost Recovery	Total Awarded
NSF grant to customize mailing list software	SBR-9311748	1993		\$4,000
NSF grant to produce an enhanced web site and search facilities	SBR-9601352	1996	\$31,014	\$109,000
NSF grant to create a searchable archive of other linguistics-related mailing lists	SBR-9975299	1999	\$32,945	\$169,000
NSF Supplementary Grant: Research Experiences for Undergraduates	SBR-9975299 (Supp.)	2000		\$5,000
NSF grant for configuration of database to support endangered language initiatives	BCS-0003197	2000	\$10,877	\$50,000
NSF grant for workshop "Endangered Languages Digitization; The Need for Standards"	BCS-1097249	2000	\$2,054	\$40,165
			2002 \$139,859	
			2003 \$104,613	
			2004 \$125,089	
			2005-6 \$109,340	
			Grant Total* \$478,901	\$2,242,914
NSF Social Sciences Infrastructure Grant: "E-Meld: Electronic Metastructure for Endangered Languages Data"	SES-1099652	2001	\$25,677	
			2003 \$50,709	
			2004-5 \$76,386	\$394,663
			Grant Total* \$14,405	
NSF Polar Programs grant: "DATA: Denali: Archiving, Training and Access"	OPP-0326805	2003	\$14,620	
			2004 \$14,620	
			Grant Total* \$29,025	\$96,525
NSF National Science Digital Libraries Collaborative Project: ALL Language grant	NSDL-0333530	2003		
DHB: Collaborative Research: "L-MAP: Language and Location - A Map Annotation Project"	HSD 0527512	2005	\$228,671	\$742,669
NSF Major Research Instrumentation: Instrument Acquisition for Endangered Languages Research: Development of an Ontology-based Toolbox	CNS 0521454	2005	\$0	\$294,200
Collaborative Research: Multi-Tree: A Digital Library of Language Relationships	BCS 04040000	2005	\$141,944	\$485,024
Total Grants			\$1,031,817	\$4,633,160

Other Development

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005 (to 11/30)	Totals
Direct Donations (ORD)	\$6,178	\$5,605	\$15,830	\$18,534	\$19,527									\$59,496
Direct Donations 5-56125 (R19200)						\$16,156	\$17,979	\$16,291	\$23,468	\$43,817	\$52,198	\$58,125.01	\$61,307.00	\$289,341
Job Announcements								\$27,019	\$20,830	\$18,423	\$6,345	\$25,482.41	\$35,850.00	\$133,949
Publisher Services								\$17,067	\$17,175	\$12,745	\$14,995	\$15,945.00	\$21,545.00	\$99,472
Extended Publisher Services							\$15,000		\$14,970	\$20,000	\$20,000	\$13,450.00	\$10,629.00	\$94,049
Individual Book Announcement										\$100	\$100	\$100.00	\$1,569.00	\$1,869
Website Development													\$38,241.00	
Total Other	\$6,178	\$5,605	\$15,830	\$18,534	\$19,527	\$16,156	\$32,979	\$60,377	\$76,443	\$95,085	\$93,638	\$113,102	\$169,141	\$678,176
Grand Total														\$5,311,336