U.S. Department of Education

Washington, D.C. 20202-5335

APPLICATION FOR GRANTS UNDER THE

INTERNATIONAL RESEARCH AND STUDIES PROGRAM CFDA 84.017A. SCHEDULE 84.017A-3
CFDA # 84.017A
PR/Award # P017A090309
Grants.gov Tracking#: GRANT10262613

OMB No. 1840-0795, Expiration Date: 08/31/2010
Closing Date: APR 23, 2009

Table of Contents

Forms

Narratives

This application was generated using the PDF functionality. The PDF functionality automatically numbers the pages in this application. Some pages/sections of this application may contain 2 sets of page numbers, one set created by the applicant and the other set created by e-Application's PDF functionality. Page numbers created by the e-Application PDF functionality will be preceded by the letter e (for example, e1, e2, e3, etc.).

U.S. DEPARTMENT OF EDUCATION

BUDGET INFORMATION

NON-CONSTRUCTION PROGRAMS

OMB Control Number: 1890-0004

Expiration Date: 06/30/2005

Name of Institution/Organization:
Amherst College

Applicants requesting funding for only one year should complete the column under "Project Year 1." Applicants requesting funding for multi-year grants should complete all applicable columns. Please read all instructions before completing form.

SECTION A - BUDGET SUMMARY

U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Categories	Project Year 1(a)	Project Year 2 (b)	Project Year 3 (c)	Project Year 4 (d)	Project Year 5 (e)	Total (f)
1. Personnel	$ 18,867	$ 19,500	$ 7,681	$ 0	$ 0	$ 46,048
2. Fringe Benefits	$ 4,545	$ 4,733	$ 2,281	$ 0	$ 0	$ 11,559
3. Travel	$ 0	$ 3,000	$ 6,000	$ 0	$ 0	$ 9,000
4. Equipment	$ 6,734	$ 0	$ 0	$ 0	$ 0	$ 6,734
5. Supplies	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
6. Contractual	$ 33,809	$ 42,362	$ 38,934	$ 0	$ 0	$ 115,105
7. Construction	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
8. Other	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
9. Total Direct Costs (lines 1-8)	$ 63,955	$ 69,595	$ 54,896	$ 0	$ 0	$ 188,446
10. Indirect Costs*	$ 13,579	$ 14,055	$ 5,778	$ 0	$ 0	$ 33,412
11. Training Stipends	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
12. Total Costs (lines 9-11)	$ 77,534	$ 83,650	$ 60,674	$ 0	$ 0	$ 221,858

***Indirect Cost Information *(To Be Completed by Your Business Office)*:**

If you are requesting reimbursement for indirect costs on line 10, please answer the following questions:

(1) Do you have an Indirect Cost Rate Agreement approved by the Federal government? [X] Yes [ ] No

(2) If yes, please provide the following information:

Period Covered by the Indirect Cost Rate Agreement: From: 7/1/2006 To: 6/30/2009 (mm/dd/yyyy)

Approving Federal agency: [ ] ED [X] Other (please specify): Department of Health and Human Services

(3) For Restricted Rate Programs (check one) -- Are you using a restricted indirect cost rate that:

[ ] Is included in your approved Indirect Cost Rate Agreement? or, [ ] Complies with 34 CFR 76.564(c)(2)?

ED Form No. 524

U.S. DEPARTMENT OF EDUCATION

BUDGET INFORMATION

NON-CONSTRUCTION PROGRAMS

OMB Control Number: 1890-0004

Expiration Date: 06/30/2005

Name of Institution/Organization:
Amherst College

Applicants requesting funding for only one year should complete the column under "Project Year 1." Applicants requesting funding for multi-year grants should complete all applicable columns. Please read all instructions before completing form.

SECTION B - BUDGET SUMMARY

NON-FEDERAL FUNDS

Budget Categories	Project Year 1(a)	Project Year 2 (b)	Project Year 3 (c)	Project Year 4 (d)	Project Year 5 (e)	Total (f)
1. Personnel	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
2. Fringe Benefits	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
3. Travel	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
4. Equipment	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
5. Supplies	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
6. Contractual	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
7. Construction	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
8. Other	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
9. Total Direct Costs (lines 1-8)	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
10. Indirect Costs	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
11. Training Stipends	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
12. Total Costs (lines 9-11)	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0

Project Narrative

Abstract Narrative

Attachment 1:
Title: Pages: Uploaded File: **1235-abstract.pdf**

LangBot Project: An Intelligent Agent for Language Learning

Instant Messaging (IM) has become the communication tool of choice among secondary- and university-aged students, but language teachers have not yet succeeded in harnessing the capabilities of this tool to support foreign language learning. LangBot is a ground-breaking second language learning and research tool designed to collect detailed behavior-tracking and self-report data, generate user models, and track vocabulary development while serving as an “intelligent” language reference agent in a conversational "wrapper.” LangBot is added to a user's "buddy list" in their IM software application just like a human "buddy". Once this is done, learners can ask LangBot for help. LangBot can also initiate interactions with individual learners by asking them questions and making suggestions. Because learners are encouraged to formulate sentence-length requests for assistance in the target language (English is also possible), LangBot not only simplifies the search for language resources, it also provides an opportunity for communicative language use. LangBot can provide: 1) translations for words, phrases or sentences; 2) examples of words or phrases used in context; 3) corrective feedback on spelling, morphological and syntactic errors in sentences produced by learners; 4) automatically-generated, individualized vocabulary tests that are automatically administered and scored with the results recorded in a database; 5) automatically-scored fill-in-the-blank exercises, and 6) suggestions for readings derived from online newspapers. LangBot will be freely available via the IM software learners and teachers already have installed on their computers. It will serve as a gateway to a variety of online resources commonly used by language learners.

The project website will provide learners with access to their own assessment data and enable them to send reports of quiz results and estimations of their current vocabulary knowledge and development history to their teachers. An instructional guide will be prepared for teachers and available on the website with suggestions for how they can use LangBot to enhance their curricula.

A knowledgeable team of computational linguists, computer-assisted language learning software developers, Portuguese, Chinese, and Spanish linguists, and second language acquisition researchers will develop LangBot and all language materials. Corpora representing contemporary language use in the three languages will be developed as a resource for the LangBot system and made available as stand-alone resources for materials development and reference. All publications and presentation materials will be accessible on the project website.

Project Narrative

Project Narrative

Attachment 1:
Title: Pages: Uploaded File: **1242-LangBot_Project_2009.pdf**

Table of Contents

LangBot Project: An Intelligent Agent for Language Learning

In the past few years, the availability of online language learning resources and tools has increased significantly. Language learners and teachers alike have access to online dictionaries, grammar reference websites, machine translation tools and authentic materials ranging from news print media, Internet radio, and video. However, taking full advantage of these resources is not as straightforward as one might think. Learners and teachers must first locate resources on the Web, vet them for quality, and then determine how best to integrate them into the learning process. If language reference services are free, they often rely on banner ads or more intrusive advertisement techniques (e.g. pop-up ads) for a revenue stream making the search for language reference information cumbersome and frustrating. For learners who are successful in locating and utilizing this array of resources, each with a different interface, there is still no mechanism for tracking learners' requests for information or language reference assistance and using these data to promote long-term retention and productive use of vocabulary. The LangBot Project seeks to address this gap by providing a single interface to a wide array of language reference and resource materials together with the capability of tracking learner queries and using past learner behavior to customize assistance.

Instant Messaging (IM) has become the communication tool of choice among secondary- and university-aged students, but language teachers have not yet succeeded in harnessing the

capabilities of this tool to support foreign language learning. Instant messaging systems are designed to promote real-time communication between people, however, it is also possible for "intelligent agents" or "bots" to engage with humans in communicative interaction. A *bot* is a relatively small and focused computer application that runs continuously in the background as other programs (e.g. an instant messenger application) are being run, and responds automatically to a user's activity. LangBot is an innovative, data-driven language learning and research tool that logs learner behavior, self-report data, generates learner models, and tracks vocabulary development while serving as an “intelligent” language reference agent in a conversational "wrapper.” LangBot is added to a user's "buddy list" in their IM software application just like a human "buddy". Once this is done, learners can ask LangBot for help. LangBot can also initiate interactions with individual learners by asking them questions and making suggestions. Because learners are encouraged to formulate sentence-length requests for assistance in the the target language (English is also possible), LangBot not only simplifies the search for language resources, it also provides an opportunity for communicative language use. LangBot can provide: 1) translations for words, phrases or sentences; 2) examples of words or phrases used in context; 3) corrective feedback on spelling, morphological and syntactic errors in sentences produced by learners; 4) automatically-generated, individualized vocabulary tests that are automatically administered and scored with the results recorded in a database; 5) automatically-scored fill-in-the-blank exercises, and 6) suggestions for readings derived from online newspapers (see section 3.2 for more detail description of LangBot's capabilities). LangBot will be freely available via the IM software learners and teachers already have installed on their computers. It will serve as a gateway to a variety of online resources commonly used by language learners.

The project website will provide learners with access to their own assessment data and

enable them to send reports of quiz results and estimations of their current vocabulary knowledge and developmental history to their teachers. An instructional guide will be prepared for teachers and available on the website with suggestions for how they can use LangBot to enhance their curricula.

A knowledgeable team of computational linguists, computer-assisted language learning software developers, Portuguese, Chinese, and Spanish linguists, and second language acquisition researchers will develop LangBot and all language materials. Corpora representing contemporary language use in the three languages will be developed as a resource for the LangBot system and made available as stand-alone resources for materials development and reference. All publications and presentation materials will be accessible on the project website.

1. Need for Project

As web-based language reference resources have proliferated, the need has grown for efficient and effective means of leveraging these resources for language learning. In recent years, a sea change has occurred in the use of computer-mediated communication tools (CMC) among secondary- and university-aged students with preference shifting from the use of email to Instant Messaging (IM) for communication with peers. Clearly, IM has become the communication tool of choice among school-aged students, but researchers have made very little progress in finding ways to integrate research and data collection tools into the IM framework. Today's middle school, high school, and college students use IM and cell phones to stay in perpetual contact with their peers, a development that has prompted numerous studies examining this phenomenon from a wide range of perspectives. In the context of second/foreign language acquisition (SLA) research, IM and *chat* are often used synonymously and studies examining dyadic interaction between language learners are increasingly using IM tools for data collection by either requiring

participants to either save transcripts of their conversations or by using screen-capture software (a screen-capture utility creates a video recording of students using the IM interface) to record the online conversation (e.g. Sotillo, 2005). However, this approach to using IM for research is problematic because it places research participants in the role of being responsible for data collection. If participants either forget to record their own data and send it to the researcher or intentionally neglect to do so, this can result in potential selection bias and compromise data integrity. If the screen-capture technique is used, each screen-capture movie must be viewed and transcribed before analysis can begin. Using these approaches is very resource-intensive limiting their feasibility.

To date, very little work has been done exploring how data can be collected directly via IM tools to support instruction. This is most likely due to the fact that most IM servers are commercially owned (e.g. America Online, Microsoft) making automated data collection very difficult. Of particular interest and relevance to the present project is how IM tools can be used to promote foreign language learning beyond facilitating real-time exchange of text or audio messages between human interlocutors. Using a bot provides new opportunities for SLA researchers. It can automatically record its interactions with users, store these data in a database, and perform analyses behind the scenes to inform future interactions with individual users and assess their vocabulary development.

Learning a foreign language requires an extended period of intensive study. The fact that second language acquisition theories have been developed almost exclusively based on cross-sectional studies represents a fundamental disconnect between the empirical basis of theories and the longitudinal learning processes they attempt to explicate. There have been a number of calls from within the profession to expand our empirical focus to include investigations of student

learning outcomes that can provide more detailed accounts of language development over time (e.g. Tucker, 2000; Ortega & Iberri-Shea, 2005). To date, the majority of longitudinal investigations of second language acquisition have been descriptive in nature and consisted of small sample sizes, precluding the use of inferential statistics and negatively impacting the generalizability of results (Ortega & Iberri-Shea, 2005). One way of improving generalizability is to increase sample size; however, the feasibility of this option is limited due to the massive amount of data typically generated by rigorous longitudinal studies and the lack of techniques for automatically analyzing these data. LangBot represents a tool that addresses this need while serving the instructional needs of language learners and teachers.

Determining which vocabulary are the most important for learners to acquire first has been a focus of SLA researchers. According to Laufer and Nation (1999), the first 1000 most frequent words in a language account for around 75% of the running words in formal written texts and about 84% of informal conversational use. While these percentages may vary somewhat across languages, Laufer and Nation (ibid) argue that for the purposes of general foreign language instruction not focused on defined areas of use, knowing the first 2000 most frequent words equips learners with the vocabulary knowledge required to be a highly functional communicator in the target language. LangBot is designed to assist students in achieving productive mastery of these 2000 most frequent words as well as acquiring more specialized vocabulary.

2. Potential for the use of materials in program to others

The LangBot project will have a significant impact on language teaching, learning, and SLA research. Since LangBot is freely accessible on the three main IM networks (AOL, MSN, and Yahoo), teachers and learners anywhere can leverage LangBot's capabilities to support

language learning in a wide range of curricular contexts.

2.1 Usefulness to Teachers

The project website will include information for teachers about the capabilities of the system, the interactions and activities it supports, and the data it collects. The list of the six thousand most frequent words in each of the supported languages, separated into 500-word frequency bands, will also be available for download from the website. In addition, an instructional guide will be prepared for teachers, explaining how their curriculum can be designed to take advantage of the capabilities of the system.

2.2 Usefulness to Students

In addition to being able to use LangBot as a language reference and vocabulary training assistant, learners will have the option of viewing all of their personal data, including the number of: 1) vocabulary words that they have mastered through their interactions with LangBot; 2) vocabulary and concordance gap activities completed; and 3) reading texts requested. Learners will also have the option of sending these results and a vocabulary development report to their instructors.

2.3 Usefulness to Intelligent Computer Assisted Language Learning (ICALL) Researchers

The development of new NLP tools for Spanish, the adaptation of existing NLP tools for Chinese, and the redeployment of existing Portuguese ICALL tools in a different system will have a direct impact on the ICALL community effort to expand and popularize NLP techniques in Computer Assisted Langauge Learning. Because all developed technology will be available under the GNU public license, this project will contribute in different ways to the develpment of other instructional tools that use NLP.

First, the research on the adaptation of the existing Chinese NLP components will help us

better understand the necessary steps to transform tools that can perform error identification (finding out if an error exists) into tools that can carry out error diagnosis (describing the nature of the error). The more we learn about these necessary steps, the better we can reuse existing NLP tools that were originally designed for L1. Second, the redeployment of existing Portuguese error diagnosis modules will provide important information about portability issues in ICALL technology. As discussed in Amaral (2007), most of the NLP tools currently used in intelligent language tutors is system dependent, which means that it is very difficult to use them outside of the original system for which they were designed. One of the goals of the the LangBot project is to investigate reusability issues in ICALL and to propose new ways of developing NLP tools for multiple language tutoring systems. Finally, the development of Spanish error diagnosis modules based on the existing Portuguese ones will provide a new set of tools to the ICALL community that can serve different purposes in different systems.

3. ACCOUNT OF RELATED MATERIALS

3.1 Previous Research on Intelligent Language Tutors

The inclusion of advanced computational techniques in language learning software, or intelligent computer-assisted language learning, has been a focus of research since the early 1990s (see Holland, Kaplan, & Sama., 1995; the special Issue of the *Journal of Artificial Intelligence in Education*, 1994; Gamper & Knapp, 2002 for further discussion). Under the umbrella term of artificial intelligence, this work has focused on the application of techniques from natural language processing (e.g. Amaral and Meurers, 2008; Amaral 2007; Heift 2003; Nagata 2002) and speech recognition to help language learners improve their pronunciation (LaRocca, Morgan, & Bellinger, 1999; Rypa & Price, 1999; Dalby & Kewley-Port, 1999), assist children learning to read (Mostow & Aist, 1999), diagnose errors and provide corrective

PR/Award # P017A090309

feedback to students (see the special issue of the *CALICO Journal*, 2003), model and adapt to learner behavior (e.g., Bull, 1994), and support virtual dialogs or speech-interactive environments (Bernstein, Najmi, & Ehsani, 1999; Holland, Kaplan, & Sabol, 1999; Harless, Zier, & Duncan, 1999). While the above-mentioned intelligent language tutors have proven to be effective tools for language learning, they all serve as stand-alone systems and are intended to support language learning between one individual and a computer. Due to specialized hardware and software requirements and limited scalability, these systems have not achieved broad impact in foreign language education.

More recently, research has begun to focus on the development of conversational agents that can interact with language learners via an IM client (Zacharski, 2002; 2003). This work represents a new direction in the use of intelligent agents to support language learning by using standardized computer-mediated communication tools as opposed to specialized applications to support interaction. Zacharski has developed conversational agents that function as peers and information resources for language learners as they interact in an adventure game. The plan-based approach employed by Zacharski differs from chatterbots (e.g., Alicebot, where the bot attempts to facilitate open-ended conversational exchange) by structuring interaction around the sort of dialogs frequently found in basal foreign language textbooks and requiring students to successfully complete a number of collaborative tasks. Learners interact with multiple conversational agents and each other via IM, send and receive email messages, and view web pages in search of information to solve the required tasks (Zacharski, 2003). By customizing and integrating open-source IM (e.g., Jabber), email, and web servers together into a unified framework, the system is able to keep track of what is done, when, and by whom and respond in a manner that promotes productive collaboration on the part of the human participants (Zacharski

2003).

Building on Zacharski's prototype, the LangBot serves as an intelligent language reference agent that receives requests for assistance in the form of natural language questions from learners, retrieves and presents the requested information in a conversational "wrapper," can utilize feedback from learners to refine queries, and learns from these interactions. The probabilistic algorithm underlying the LangBot's "intelligence" learns from each interaction, thus each failure to assist learners improves its future performance.

3.2 Description of LangBot's Features and Capabilities

The first time users initiate interaction with LangBot, it introduces itself and explains briefly what it's capabilities are and what they need to do to receive help. Since it collects user data, users are asked for their consent to analyze these data for evaluation purposes. After the consent form is completed (agreement to participate isn't required to use the tool), users are presented with a series of questions regarding language history, gender, age, time spent studying abroad, country of residence, and self-assessed language proficiency that form the basis of their learner profile. These data are used by LangBot as a starting point for calibrating the level of difficulty for the readings it suggests to the user, for selecting vocabulary words, and generating learner models together with tracking and performance data. Periodic short questionnaires (2-4 questions) are administered to recalibrate user models and collect data on the use of LangBot as it relates to specific types of instructional tasks (i.e. reading or writing).

LangBot uses Naïve Bayesian classifiers to parse and classify user requests and determine the exact nature of a request. This statistical approach, referred to as "unsupervised learning", permits LangBot to "learn" from input by users. If it encounters a request it can't classify, LangBot asks for clarification from the user. Once LangBot has determined the nature

of the request, the next time a user formulates a request in a similar manner, it is “understood.” This learning process occurs entirely without human intervention on the part of a programmer, computational linguist, or language teacher. In essence, the more LangBot is used, the "smarter" or more accurate it becomes in classifying requests with greater variability in syntax and lexis. LangBot can classify fourteen different message types, ranging from translation requests to simple greetings and leave-takings.

Word of the Day: Since users will have added LangBot to their “buddy lists,” it will see when they are online and have indicated that they are “available.” Depending on an individual user’s preference (learners can indicate how frequently they wish to receive the word-of-the-day query), LangBot will select a word from the appropriate frequency band, based on the user’s reported proficiency level and tracking data, and ask for a translation equivalent. If the translation provided by the learner is incorrect, LangBot will provide a number of possible appropriate responses. Next, the learner will be asked to use the word in a sentence. The student's input will then be checked for spelling, morphological, and syntactic errors using NLP tools that are specifically designed to analyze learner language. The tools for Portuguese will be adapted from the existing TAGARELA system (Amaral and Meurers, 2008; Amaral 2007) that are available under the GNU public license, and that have been develped by the Co-PI Luiz Amaral. The tools for Chinese and Spanish will be either developed based on TAGARELA specifications, or will be existing NLP tools adapted to the needs of LangBot, such as the Stanford Chinese Word Segmenter and the Stanford Parser (both also available under the GNU public license). If the vocabulary usage is considered appropriate, the learner will be given positive feedback and provided with three additional sentences containing the specified word. For inappropriate responses, LangBot will indicate that this doesn’t appear to be correct and will

provide three sentences that the learner can use as models. Careful attention has been given to how LangBot responds to learners so that learners perceive LangBot as an "assistant" and not a "teacher" or expert. This is a very important detail that gives LangBot credibility in the eyes of students and does not foster unrealistic expectations on their part. Finally, learners who wish to receive a word-of-the-day prompt more frequently than once a day, can ask LangBot for one as often as they wish.

Periodic vocabulary tests: Repeated exposure and reinforcement is required to promote long-term retention of vocabulary. The proposed project will accomplish this by administering periodic vocabulary tests through the LangBot. These tests will consist of 5 to 10 items (optimal length of the tests will be determined through usability testing) and will have the same format as the word-of-the-day activity. Learners will be presented with a word and asked to produce its translation equivalent followed by a sentence demonstrating its use in context. Learners will receive corrective feedback on spelling and errors in morphology and syntax. The words included in each vocabulary test will consist of words from queries submitted previously, prior word-of-the-day items, and words selected from the frequency band of the frequency list that corresponds to the learners' current level.

The frequency of tests will be calculated based on the number of queries submitted together with the number of days that have passed since the last test. The default minimum period of time that can elapse between vocabulary tests for high-frequency users is currently set at one week; however, learners will have some degree of control over the time interval (the default period may be adjusted based on usability testing). Before beginning to administer a vocabulary test, LangBot asks if the learner would like to take the test at that time. If the learner declines, then LangBot will wait until the next day or the next time the learner is online, which

ever comes first. Learners will also have the capability of asking LangBot for vocabulary tests more frequently than the default period.

Concordance fill-in-the-blank: Findings from vocabulary learning research suggest that using concordances is an effective means of promoting L2 vocabulary development (e.g. Stevens, 1991; Cobb, 1997). This method, known as data-driven learning, presents learners with concordance lines instead of a dictionary definition. The learner's task is to deduce the meaning of the word based on the contextual snippets provided. One of the benefits of this approach is that learners develop a more nuanced understanding of a word's meanings. In this activity, learners are presented with 5 concordance lines, each eleven words in length, where the word of focus is replaced by a blank. The learner's task is to provide the word that belongs in the blank.

Suggested readings: The final activity supports vocabulary acquisition by suggesting readings that match learner ability based on reported proficiency level. The LangBot web crawler uses Google's news service to select articles based on topics that are currently receiving attention in the media of each target language. Each article is analyzed to determine readability by calculating the percentage of words contained in each of the twelve word frequency bands in the article. The Automated Readability Index (ARI) and the Coleman-Liau Index (CLI) were chosen over other readability indexes for their more straightforward approach. While other indexes count the number of syllables per word, the ARI and CLI consider the number of characters instead. These two indexes use a formula involving the average number of words per sentence and the average number of character per word to calculate the difficulty of an article. These readability indices were developed for Western languages. It will be necessary to modify and augment these indexes to function with non-Western languages. For instance, Lau and King (2006) adapted the Flesch (1948) and Lang (1971) formulas to gauge the readability of Chinese

web pages. LangBot asks learners if they would like to read an article with the option of choosing a topic domain (e.g. current events, sports, business, science & technology, entertainment, etc.). If learners agree, they are provided with a link to the selected article. Learners may also request an article at any time (e.g. the equivalent of "Can you suggest an article" or "I would like to read an article" in the target language).

Tracking development and generating a learner status profile: The computational architecture of LangBot will also include techniques for assessing learner vocabulary development based on a combination of translation requests submitted and subsequent vocabulary test results, including previously requested translation equivalents and words selected from the word frequency list.

3.3 LangBot Technical Specifications and Design

Pidgin (formerly known as GAIM) is a multi-platform instant messaging client with a plug-in architecture that supports the construction of bots that are compatible with common instant messaging systems (AOL, MSN, Yahoo, Jabber, etc.). LangBot's stochasitic approach to classifying requests and identifying the specifics of the request has proven quite robust. Previous research conducted on the LangBot prototype demonstrated a 95% accuracy rate for classifying message types with fairly limited training input (Payne & Lipschultz, 2007). All requests are logged with the user's screen name in a MySQL database together with the response provided. If the request is for a translation, i.e. "How do you say picture frame in Portuguese", the computation component determines that the request is for a translation, "picture frame" is to be translated, and the translation is from English to Portuguese. Next, the request will be sent to an online English-Portuguese dictionary, the response is parsed and repackaged into the conversational response: "Here are the followings ways you can say "picture frame" in Portuguese: option 1, option 2, etc. Would you like to see some examples of how [Portuguese

word or phrase] is used in a sentence." If the user responds affirmatively, three example sentences are provided from a corpus stored in the local database. Since both the translation request and result are stored in the local database, subsequent requests for the same word or phrase can be processed without needing to query the external source. If the translation request is for a full sentence, the computational component classifies this as a different message type, submits the query to the Google online translation engine, and formulates the response as, "In Portuguese you could say, "[translated sentence]".

3.3.1 Beyond Feedback on Translation

After receiving feedback on the translation of words or phrases, learners are offered the possibility to use the learned expression to create his/her own sentence. The student's input will then be checked for spelling, morphological, and syntactic errors. To perform these tasks for Portuguese we propose to incorporate the Portuguese NLP modules developed by project Co-PI Luiz Amaral for the TAGARELA system (Amaral and Meurers, 2008; Amaral 2007).

In TAGARELA, the student input is sent to a module called Analysis Manager that controls the error diagnosis process based on properties of the input, information about the learner, and characteristics of the activities. The Analysis Manager calls the NLP sub-modules that are part of the expert module of the system. The whole process starts with tokenization of the input string, and moves to spelling, syntactic and semantic analysis. The Portuguese tokenizer takes into account specific properties of Portuguese, such as cliticization, contractions, and abbreviations. After tokenization, the input is checked for non-word spelling errors using a standard spell-checker (Kuenning, 2005) with Brazilian Portuguese parameter files. Then, the system performs a full-form lexical lookup that returns multiple analyses based on the CURUPIRA lexicon (Martins et al. 2006), including detailed morphological information. In the

spirit of Constraint Grammar (Karlsson et al. 1995; Bick 2000, 2004), disambiguation rules are used to narrow down the multiple lexical analyses based on the local syntactic context. Complementing these local disambiguation rules, a CYK parser augmented with feature checking capabilities uses hand-written grammar rules to checks agreement, case relations, and some global well-formedness conditions. In TAGARELA, in addition to the form-focused processing, content assessment is performed using shallow semantic matching between the student answer and target answers provided by the teacher in the Activity Model, essentially a basic version of the approach discussed in Bailey & Meurers (2008). We would like to explore the capabilities of such an approach to content-based error diagnosis for LangBot.

The NLP tools for Portuguese are ready to be incorporated and tested in LangBot. This project proposes to use the same approach to develop error diagnosis NLP tools for Spanish and Chinese. Many of the existing tools for Portuguese can be used with minor changes for Spanish. For example, the Portuguese tokenization rules are very similar to Spanish, and only some minor changes have to be added to address small structural differences in both clitic systems, and contractions with prepositions. The parser can be used as is, but the grammar has to be addapted, and the local desanbiguation rules have to be tested. We expect to make minor changes to these rules as well.

More NLP tools will have to be incorporated in the case of Chinese. Since Chinese does not use whitespace to demarcate word boundaries, the existing LangBot architecture will have to be modified to accommodate this requirement. We will use the Stanford Chinese Word Segmenter (available under the GNU public license) to locate word boundaries and the Stanford Parser for checking the well formedness of the student input. The feedback strategies will be modified to accommodate the available NLP resources, and feedback messages will reflect the

lexical and morphosyntactic information available to the system. A Chinese computational linguist familiar with the Stanford NLP tools will work with project Co-PI Luiz Amaral to make the necessary modifications to the Stanford tools.

3.3.2 Corpora

The corpora for each of the languages targeted in this project will be representative of a broad range of contemporary language use in each respective language. The genre to be included are Wikipedia articles, blog text, online newspaper articles, feature film language tracks, and public domain literary text. Custom scripts will mine Wikipedia articles, strip all wiki mark-up and store the clean text in the local database. RSS feeds generated by blogging systems will be leveraged to build a growing corpus of contemporary language from selected blogs in each language. Feature film language tracks will be extracted from DVDs using freely available software. [Note Bene: since only disconnected sentences will be provided to learners and they won't have access to the full-length original text, the use all source texts is covered by fair use.] Frequency lists of the 6000 most frequent words for each language will be generated from each corpus and divided into 500-word frequency bands. Each 500-word band will correspond to a difficulty level with the four bands in the 2000 most frequent words corresponding to beginner to intermediate levels of language proficiency. As learners demonstrated vocabulary knowledge grows, the targeted frequency band will become more difficult.

4. Likelihood of Achieving Results

The proposed project seeks to leverage the pervasive use of IM among today’s language learners in secondary and post-secondary foreign language education to promote vocabulary acquisition in Portuguese, Chinese, and Spanish. The LangBot project is unique in that it employs Instant Messenger as a gateway to a range of language reference resources used by

language learners as a part of everyday practice, as well as specialized resources constructed as a part of this project that are not currently available for teachers and learners in less-commonly taught languages. Since many of the core computational and data infrastructure components of LangBot have already been constructed, the requested funds will be used to build corpora for Portuguese, Chinese, and Spanish, make language-specific customizations to the LangBot system (conversational language wrapper elements), and extend LangBot's capabilities to include feedback on spelling, morphology and syntax with TAGARELA technology and the Stanford Chinese Word Segmenter and Parser. The project PI and Co-PI have extensive experience designing and managing software development projects in addition to their expertise in language teaching and SLA research.

Project PI, Scott Payne, has designed and developed a specialized web-based tool for constructing and analyzing corpora to be used for the project. MOCA (Multimedia Online Corpus Analysis - http://moca.ats.amherst.edu/) enables project investigators to identify dynamic sources of language material (e.g. weblogs, online newspapers, discussion forums) where new content is frequently posted and link those sources to a corpus. MOCA automatically checks these sources and downloads any new material to the corpus, adding relevant meta-data for each resource (i.e. genre, date, source, etc.). Other static content (e.g. literary text, government documents, language tracks from feature films, etc.) can also be uploaded to a corpus. When data is added to a corpus, MOCA executes a battery of scripts that generate and update frequency lists for single-word up to six-word sequences, tag words for part-of-speech, and lemmatize verbs. Project Co-PI, Luiz Amaral, had designed and implemented TAGARELA (http://tagarela.osu.edu). TAGARELA (Teaching Aid for Grammatical Awareness, Recognition and Enhancement of Linguistic Abilities) is an intelligent web-based workbook for learners of

Portuguese (Amaral, 2007; Amaral & Meurers 2007, 2008). The system can be used as a pedagogical complement in traditional classroom settings, as well as in distance learning or individualized instruction programs. It includes six activity types: listening comprehension, reading comprehension, picture description, fill-in-the-blanks, rephrasing, and vocabulary. These activities provide opportunities for students to practice their listening, reading, and writing skills. Different from paper-based workbooks, TAGARELA offers on the spot individualized feedback on orthographic errors (non-words, spacing, capitalization, punctuation), syntactic errors (verbal and nominal agreement), and semantic errors (missing concepts, extra concepts, word choice). In contradistinction to traditional CALL exercises, specific, individualized feedback can be provided even for activities which allow a wide range of variation in the vocabulary, the morphological form, the word order, and the syntactic constructions used by the learner. For all activity types, the answers are checked by the system, i.e., the generation of feedback is completely automated.

Development of LangBot for each of the targeted languages together with the related corpora will follow an iterative cycle of development, testing, and revision. Project investigators will be intricately involved in all phases of this process. Pretesting will be carried out internally to the project team and students from language courses at Amherst College, the University of Massachusetts, and UC Davis will participate in the pilot studies.

5. Expected Contribution to Other Programs

Instructional materials and tools are limited for less-commonly taught languages in the United States, especially Portuguese. Since foreign language learners at the secondary and post-secondary levels are already accustomed to communicating with their peers via IM, the LangBot project capitalizes on this habituated behavior to provide access to a range of language resources,

assess vocabulary learning, and track changes in acquisition over time. Furthermore, very few corpora exist in Portuguese and Chinese; especially ones which consist of language beyond journalistic and literary genres. The corpora of contemporary language will also be made available, independently, via the MOCA interface to assist teachers at all levels and sectors of foreign language education in developing instructional materials in addition to serving as a reference resource for learners.

Since the language-specific versions of LangBot will be accessible via the four most common IM networks (AOL, MSN, Yahoo, Jabber), language learners anywhere in the United States and abroad can avail themselves of this resource. The LangBot system will also include a website enabling learners to track their progress and send results from vocabulary tests and their learner status profile to their teachers. This feature makes it possible for language teachers to integrate LangBot into their local curricular contexts as they see fit.

6. Plan of operation

The project PI, Scott Payne, will be responsible for overall project management and data analysis. Michael Lipschultz, a doctoral student in Computer Science at the University of Pittsburgh and developer of the LangBot prototype, will be responsible for all LangBot and NLP programming, algorithm development, and will assist in data analysis. The web application programmer (in-kind contribution) will develop the web-based tracking and reporting interfaces for learners, as well as the analysis report interfaces for the project team and external evaluator. Luiz Amaral, Co-PI, together with the Chinese computational linguist, will be responsible for NLP tools and error diagnosis specifications and evaluations, as well as the adaptation of existing tools to the needs of the project. The Co-Investigators, including Weijia Li, a Chinese linguist at Amherst College, and research assistants will be responsible for selecting language

resources, constructing corpora in Portugese, Chinese, and Spanish, and conducting analyses specific to their respective languages, in addition to translating and providing recommendations for language-specific refinements in the language LangBot uses for interaction with learners.

The first year of the project (see project time line in Appendix A) will be dedicated to constructing the corpora for the targeted languages, formulating the interactional language used by LangBot, installing and testing the Stanford Chinese NLP tools, and adapting and developing the NLP tools for Portuguese, Spanish, and Chinese. While the project team will perform some limited pilot testing during the first year, the second year of the project is where most of the testing-revision cycle will occur. During the second year, a pilot study will be conducted for all three languages without the spelling, morphology and syntax error feedback capabilities provided by the NLP tools. Year three is when the large-scale study will be conducted with the completed version of LangBot for all three languages. Year three will also have a heavy emphasis on dissemination. Project investigators will present findings and conduct workshops at national professional meetings of foreign language teachers of Portuguese, Chinese, and Spanish, as well as computer-assisted language learning and applied linguistics conferences.

6.1 Meeting the Objectives of Section 605

The LangBot project addresses many of the objectives outlined in Section 605 of the International Research and Studies Program. This project seeks to develop and validate specialized materials and techniques for promoting foreign language vocabulary acquisition in three of the less-commonly taught languages listed in Competitive Preference Priority 1. Due to the open access to LangBot, this resource can also support training overseas, as well as advancing internationalization of post-secondary programs in the United States. The LangBot project also promotes the access and dissemination of foreign language knowledge through the

use of advanced technologies.

7. Quality of key personnel

J. Scott Payne (Ph.D. Interdisciplinary: Cognitive Psychology, Second Language Acquisition, and Educational Technology, Washington State University), Director of Academic Technology Services at Amherst College. Previously, he served as the Assistant Director for Technology and Research at the Center for Language Acquisition and Senior Lecturer in the Department of Linguistics and Applied Language Studies at Penn State University. He has published in the field of computer-assisted language learning, over 17 years of experience teaching German and English as a Second Language, developed numerous online, language learning environments and courseware authoring tools. His research interests include the role that individual differences in working memory capacity play in second language acquisition processes, technology-mediated second language learning and pedagogy, and the application of emerging and advanced technologies to second language teaching and learning. He also served as the lead Instructional Technologist for Project 2001, an Andrew W. Mellon Foundation-supported initiative, based at the Center for Educational Technology at Middlebury College. Project 2001 had the mandate of providing support in the form of professional development for foreign language faculty and technology support specialists, consulting with faculty and tech support staff on software development and curricular implementation projects, and designing and building online learning environments to promote innovative teaching and learning of foreign languages at 62 of the top Liberal Arts Colleges across the United States. He will dedicate time to the project throughout the year.

Luiz Amaral (Ph.D. Linguistics) is an Assistant Professor of Hispanic Lingistics at the University of Massachusetts, Amehrst. He received his Ph.D. from The Ohio State University in

2007. His research has focused on the use of Natural Language Processing (NLP) tools for automatic error diagnosis and the development of Intelligent Language Tutoring Systems. He has developed the TAGARELA system (http://tagarela.osu.edu), an electronic workbook that uses NLP technology to provide detailed on the spot individualized feedback on orthographic, morphological, syntactic and semantic errors. For the TAGARELA process, he programmed many of the NLP tools used by the system, including the tokenizer, the parser, the grammar, the disambiguator, all shallow semantic checking modules, and all feedback modules. His research interests also include learner modeling and activity design for automatic language tutors, the use of corpora in second language instruction, and second language acquisiton theory. He will dedicate time to the project throughout the year.

Weijia Li (Ph.D. Educational Technology, University of Massachusetts, Amherst) is the Director of Chinese program in the Department of Asian Languages and Civilizations at Amherst College. She received her doctorate degree in Educational Technology from University of Massachusetts at Amherst in 2003. She also holds master degrees in applied linguistics and language teaching. Dr. Li is interested in the integration of technology in language teaching, particularly in teaching Chinese characters with multimedia. In 2005 she received a professional development grant in instructional technology for academic development from University of Massachusetts. Besides teaching Dr. Li served as a member on Teaching with Technology Committee in Smith College, and she managed the website for the Department of East Asian Languages and Literatures in Smith College. Since she came to the U. S. in the late 1990s, she has taught beginning, intermediate, and advanced Chinese at University of Massachusetts at Amherst, Smith College and Amherst College.

Before coming to the United States, Weijia Li was an English professor in Beijing

University of Science and Technology, where, for four years, she also chaired the Department of Foreign Languages, and for eight years, served as a member on the university policy-making committee. She started the multimedia program in teaching English to Chinese students in the university. As a reward to her competence and commitment, Weijia Li was honored “Excellent Young Teacher” in Beijing, and she also won the Huo, Yingdong national award as “Excellent Teacher” in the early 1990s, the highest award for teaching in China. She will work on the project throughout the year.

8. Budget and cost effectiveness

Much of the programming and design work of the LangBot framework has been completed and undergone pilot testing. The requested funding is for adding the capability of providing corrective feedback on spelling, morphology and syntax, as well as constructing the corpora and customizing LangBot's interactional language prompts to support Portuguese, Chinese, and Spanish. Building on the previous work of project PI Scott Payne and Michael Lipschultz, the LangBot framework constitutes a scalable and highly cost-effective means of conducting large-scale, longitudinal research on L2 vocabulary development. By integrating automated data collection and analysis capabilities into the system, analyses of large data sets can be significantly streamlined. All data are collected in a relational database and structured in a manner that supports direct export to statistical analysis software programs (i.e. SPSS). Furthermore, the questions and responses generated by the system will be stored externally to the program itself, making it possible to implement on-the-fly modifications to LangBot’s language prompts without disrupting the functionality of the system.

Project Co-Investigators and research assistants with expertise in each of the target languages will be responsible for translating LangBot’s language responses into Portuguese,

Chinese, and Spanish, as well as constructing the corpora in each language, from which examples of contextual language use are drawn. Custom scripts will be created for mining corpus data from the Portuguese, Chinese, and Spanish versions of Wikipedia and automatically cleaning all wiki mark-up from the text. The capability already exists in MOCA to automatically download text from weblogs and online newspapers the use RSS for syndicating content. These tools will support the rapid construction of corpora with diverse content representing a broad range of language genres.

As in-kind support, Amherst College will host the LangBot server in a climate-controlled server room with uninterrupted power supply. Academic Technology Services technical staff will secure and update server software as needed, and design the project website all at no charge to the grant.

9. Evaluation plan

The LangBot project will adopt an iterative evaluative cycle of development, testing, refinement, and full-scale implementation. Since many of the functional components of the project have already been prototyped, the focus of the project will be on developing and adapting NLP tools for implementation, data analysis and expanding the LangBot framework to accommodate Portuguese, Chinese, and Spanish.

The external evaluation of the project will be conducted by Dr. Robert Blake, Director of the University of California Language Consortium and Professor of Spanish at UC Davis. Professor Blake has agreed to conduct extensive testing of the Spanish and Chinese versions of LangBot in courses at UC Davis, as well as performing two on-site evaluations at the end of Year 2 (Fall 2009) and at the end of the Year 3 (Fall 2010). The evaluations will consist of meetings with the Co-Investigators for each language, examining the data analyses and results

for each language, and interviewing students who have had extensive experience with LangBot in focus groups. The evaluator's reports will be submitted to the U.S. Department of Education.

A central methodological question related to the evaluation of LangBot's ability to promote vocabulary acquisition pertains to how the construct of development is defined. Since conducting a series of controlled, experimental studies is impractical due to the open accessibility of the LangBot, we will employ a within-subjects design to account for developmental changes in vocabulary knowledge. The selection of vocabulary items will begin with the learner's reported proficiency level. If a learner's user profile indicates that the individual is an early beginner, items will be selected from the first frequency band (1000 most frequent words). As the learners' vocabulary knowledge increases, as measured by the percentage of correct responses in a given word frequency band, words from the next frequency band will be added to the word-of-the-day exercise and subsequently tested.

9.1 Sample Population

This research targets secondary (grades 6-12) and university-aged language learners of Portuguese, Chinese, and Spanish. The local area schools in the Pioneer Valley where both Amherst College and the University of Massachusetts are located, have strong language programs in Chinese and Spanish, including the Chinese Immersion Charter School. Due to the open accessibility of the LangBot, it is likely to attract users not currently receiving any formal language instruction, but who find LangBot useful for language maintenance. At the beginning of each large-scale study (see project timeline section 6.2), language teachers in the local are schools will be contacted, messages will be posted on listservs for Portuguese, Chinese, and Spanish language teachers announcing the availability of the tool, and announcements will be sent directly to colleagues at institutions across the United States. Due to lower enrollment

numbers nationwide in Portuguese and Chinese, it is anticipated that the number of users in these language groups will be lower than for Spanish. To bolster numbers in both of these less-commonly taught languages, we will contact the coordinators of Portuguese, Chinese, and Spanish language programs across the United States to alert them to the availability of the LangBot and encourage them to pass this information on to their students.

9.2 Research Questions

1. Do the vocabulary activities delivered via the LangBot increase vocabulary knowledge as measured by the vocabulary tests?

2. Is there a relationship between the type of vocabulary exercise and increases in vocabulary knowledge as measured by vocabulary tests?

3. Is there a relationship between reported proficiency level and vocabulary look-up queries, based on their frequency band membership?

4. What computational techniques prove most effective for selecting examples of contextualized language use for a large corpus?

5. Does the conversational interface of the LangBot promote an increase in the number of queries and request for vocabulary activities?

6. What are the user-behavior patterns exhibited by learners and how do they correlate to L2 vocabulary acquisition?

9.3 Data Analyses

The large sample size anticipated in the research will enable the use of sophisticated multivariate statistics (e.g. multiple regression) to account for the unique contributions of user-behavior patterns, exercise type, and proficiency level to L2 vocabulary development. Analyses will also be performed to determine if there are differences between groups (proficiency level,

vocabulary exercise preference, user-behavior profile) in vocabulary development.

After the initial pilots during Fall 2010 and Spring 2011 are completed, revisions will be made over the summer and the large-scale study will begin in Fall 2011 and continue through Spring 2012. Results of this study will be submitted for publication in refereed journals and presented at annual professional meetings (i.e. Chinese Language Teachers Association, American Council on the Teaching of Foreign Languages, American Association of Teachers of Spanish and Portuguese, Northeast Conference on the Teaching of Foreign Languages, Computer Assisted Language Instruction Consortium, American Association of Applied Linguistics, American Council on the Teaching of Foreign Languages).

For each additional language, the same iterative evaluation cycle will be followed: pilot test, revision and refinement, large-scale study, and write-up and publication. It is anticipated that at least four refereed publications as well as multiple presentations at national and international professional meetings will result from the project. All publications will be available for download from the LangBot project website. In addition, all source code and corpora will be released as open-source or under Creative Commons licenses and available for download from the project website.

10. Adequacy of resources

Amherst College is a Liberal Arts college that places a heavy emphasis on promoting teaching and learning. The University of Massachusetts at Amherst has strong programs in Chinese, Portuguese and Spanish with hundreds of students who will contribute to the testing and participate in the research. All project investigators have decades of language teaching experience and are fully dedicated to the project. For IT infrastructure, Amherst College is an Internet 2 institution with highly skilled IT support staff. The server room where the LangBot

server will be housed is climate-controlled and has uninterrupted power supply for all servers.

11. Description of Final Format

The LangBot project will generate an “intelligent” language reference agent in a conversational "wrapper" in the form of and IM bot that serves as a gateway to: 1) online dictionaries and machine translation engines; 2) provides examples of contextual usage of words and phrases; 3) administers automatically-generated individualized vocabulary tests and records results; and 4) selects level-appropriate readings for learners based on their vocabulary knowledge. Included is a web site where learners can track their own progress and send reports to their teachers. In addition, the corpora developed through this project will be freely accessible via MOCA, an online corpus analysis tool hosted at Amherst College. These corpora can serve as a resource for independent materials development by language teachers and as a resource for language learners. An instructional guide will be prepared for teachers to assist them in integrating LangBot into their curricula. All publications and workshop materials will be available for download from the project website.

12. Provisions for Pretesting and Revision

The iterative evaluative cycle adopted for the LangBot project will include at least three formal testing phases in addition to informal testing conducted internally by project team members. During and following the two pilots and the large-scale studies, the computational components of the IM bot will be refined as required to maximize the effectiveness of the materials. Furthermore, the project investigators will evaluate the corpora constructed for the project in their ability to inform the development of materials unrelated to the project. If they determine deficiencies in the breadth or representativeness of the corpus resources, the content will be adjusted to address these concerns.

13. Competitive Preference Priority

The instructional materials developed under the LangBot project will support language teaching and learning in three of the critical language areas highlighted by the Competitive Preference Priority: Portuguese and Chinese.

Project Narrative

Other Narrative

Attachment 1:
Title: Pages: Uploaded File: **1237-payne_cv.pdf**

Attachment 2:
Title: Pages: Uploaded File: **1238-Appendix A.pdf**

Attachment 3:
Title: Pages: Uploaded File: **1239-Appendix B.pdf**

Attachment 4:
Title: Pages: Uploaded File: **1240-cvluizamaral.pdf**

Attachment 5:
Title: Pages: Uploaded File: **1241-Weijia_CV.pdf**

CURRICULUM VITAE

J. Scott Payne

111a Seely-Mudd Building
Amherst College
Amherst, MA
Tel. (413) 542-8396
E-mail: spayne@amherst.edu

CURRENT POSITIONS

Director, Academic Technology Services, Amherst College

Affiliate, Department of Germanic and Slavic Languages and Literatures, The Pennsylvania State University

EDUCATION

Ph.D. (2000), Washington State University (Individual Interdisciplinary Doctoral Program: Second Language Acquisition, Educational Technology, and Cognitive Psychology)

Dissertation: A study of the effects of individual differences in working memory capacity and synchronous computer-mediated communication in a second language on second language oral proficiency development.

M.A. (1993), University of Arizona (German Language and Literature)

B.A. (1988), Western Washington University (German Language and Literature)

PROFESSIONAL POSITIONS

Director, Academic Technology Services, Amherst College, August 2006-present.

Supervise a staff of twelve responsible for supporting the educational and research mission of Amherst College through the development of learning and research tool and promoting new technologies across campus.

Senior Lecturer, Department of Linguistics and Applied Language Studies, The Pennsylvania State University, August 2002 - July 2006.

Taught graduate seminars on research methods and computer-assisted language learning.

Assistant Director for Technology and Research, Center for Language Acquisition, The Pennsylvania State University, August 2002 - July 2006.

Responsibilities include: proposing and leading the development of web-based tools to support research projects, consult on experimental research design and data analysis, conduct workshops training faculty and research assistants in the use of data analysis tools, serve as a resource to language departments interested in using technological tools in language instruction.

Project Co-Director, *Technologies for Advanced Foreign Language Proficiency*, Center for Advanced Language Proficiency Education and Research (National Foreign Language Resource Center), August 2003 - July 2006.

Conduct summer workshops on the use of technology for promoting advanced foreign language proficiency as well as preparing related instructional materials.

Technology Consultant, Center for Advanced Language Proficiency Education and Research (National Foreign Language Resource Center), August 2002 - July 2006.

Adjunct Lecturer, Department of Germanic and Slavic Languages and Literatures, The Pennsylvania State University, 2004-2006.

Taught a content- and project-based fifth-semester German language.

Instructional Technologist, Center for Educational Technology, Middlebury College, 2000-2002.

Responsibilities included providing instruction, and pedagogical and technical expertise as a member of a team of educator-technologists engaged in technology training and materials-development support for language faculty and instructional technologists from 62 liberal arts colleges, under the auspices of Project 2001, an Andrew W. Mellon Foundation funded program at the Center for Educational Technology, Middlebury College.

Interim Co-Director, Language Learning and Resource Center, Department of Foreign Languages and Literatures, Washington State University, Spring 2000.

Responsibilities include: 1.) managing and maintaining the departmental website, 2.) scheduling, training, and overseeing lab monitors, 3.) working with other units on campus to expand the availability of foreign language programming via satellite and make such content easily accessible to faculty and instructors, and 4.) assisting faculty with technological issues associated with instruction.

English as a Second Language Intensive Program Coordinator, American Language and Culture Program, University of Idaho, 1998-1999.

German Instructor, Department of Foreign Languages and Literatures, Washington State University, 1997-2000.

Taught second, third, and fourth semester German language courses.

English as a Second Language Instructor, Intensive American Language Center, Washington State University, 1996-1997.

Taught English to foreign students, coordinated multi-sectional courses, developed evaluation criteria and instructional materials.

Visiting Professor, Chonnam National University, Kwangju, Korea, 1995-1996.

Taught intermediate and advanced German conversation and composition courses.
Taught beginning EFL to trade, accounting, and regional development majors in the College of Business.

English as a Second Language Instructor, Language Research Center of Chonnam National University, Kwangju, Korea, 1993 - 1995.

Teaching-related activities included: high-beginner to advanced English conversation courses to university students and members of the community, instructional materials development, beginner to advanced courses in the Early English Education Program and teaching language and methodology courses in teacher-training courses held at the Language Research Center.

Graduate Teaching Assistant, University of Arizona, Tucson, AZ, 1991-1993.

Taught first and second semester German courses.

REFEREED JOURNAL ARTICLES & BOOK CHAPTERS

Payne, J. S. & Ross, B. (2005). Working memory, synchronous CMC, and L2 oral proficiency development, *Language Learning & Technology*, 9(3), 35-54.
Available online: http://llt.msu.edu/vol9num3/payne/default.html

Thorne, S. & Payne, J. S. (2005). *Evolutionary trajectories, internet-mediated expression, and language education, *CALICO Journal*, 22(3), 371-397.

* CALICO-Sony Outstanding Article Award, CALICO Journal 2004-2005

Chun, D. M. & Payne, J. S. (2004). What makes students click: working memory and look-up behavior. *System*, 32 (4), 481-503.

Payne, J.S. (2004). Making the most of synchronous and asynchronous discussion in foreign language instruction. In R. Terry, L. Lomicka, and J. Cooke-Plagwitz (Eds.), *Heinle Professional Series in Language Instruction: Teaching with Technology,* 1 (pp. 155-161). Boston: Heinle.

Payne, J.S. & Whitney, P.J. (2002). *Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal*, 20 (1), 7-32.

* CALICO-Sony Outstanding Article Award, CALICO Journal 2002-2003

Payne, J. S. & Peterson, N. S. (2000). The Civil War project: project-based collaborative learning in a virtual space, International Forum of Educational Technology & Society, July 2000.
Available online: http://ifets.ieee.org/periodical/vol_3_2000/f04.html

JOURNAL SPECIAL ISSUES EDITED

Thorne, S. & Payne, J. S. (2005). Guest co-editor of a special issue of the CALICO Journal on Computer-mediated Communication and Foreign Language Learning: Context, Research and Practice, 22 (3), 368-766.

BOOK REVIEWS

Payne, J. S. (2006). Review of New Perspectives on CALL for Second Language Classrooms (Fotos & Browne). *International Journal of Bilingual Education and Bilingualism*, 9 (4), 516-520.

NON-REFEREED PUBLICATIONS

Payne, J. S. (1995). Hypermedia and Learning Styles: Match or Mis-Match, *Language Teaching and Research*, N. XXV, 85-95.

Payne, J. S. (1995). Kids' CALL: Language Practice for Students with Limited Literacy, *Language Teaching and Research*, N. XXIV, 197-203.

RESEARCH-IN-PROGRESS

Payne, J. S. (in preparation). Using corpus analytic techniques for assessing second language development.

Payne, J. S. (in preparation). Some methodological considerations for comparing face-to-face interaction and SCMC.

Payne, J. S. (in preparation). Student documentary film: project-based learning for advanced language proficiency.

GRANTS

Title: **Working memory, language processing, and second language learning**
Role: Principal Investigator; Co-PIs: Nuria Sagarra and Judith Kroll
Funding Agency: Language Science Research Group, Penn State University
Status: Awarded ($2000)

Title: **Development of the Online Corpus Analytic Tool and a corpus-based assessment methodology**
Role: Principal Investigator; Co-PIs: James Lantolf and Steven Thorne
Funding Agency: Social Science Research Institute, Penn State University
Status: Awarded ($5,000 over 1 year)

Title: **Integrated System for Language Education and Training**
Role: Co-Investigator (PI Ed Twardy, Academic Consortium for Global Education)
Funding Agency: Office of Naval Research
Status: Pending ($1,673,461 for UMass/Amherst College subcontract over 4 years)

INVITED PRESENTATIONS

Payne, J. S. (2007). *The promise of corpus analysis: Transforming L2 pedagogy, assessment, and research*. An invtited presentation at the Center for Language Study, Yale University, New Haven, CT.

Payne, J. S. (2007). *The psycholinguistics of chat*. An invited presentation at the Center for Language Study, Yale University, New Haven, CT.

Payne, J. S. (2006). *The psycholinguistics of chat: working memory, patterns of use, and cross-modality transfer*. Paper to be presented in an Invited Colloquium at the annual meeting of the American Association of Applied Linguistics, Montreal, CA.

Thorne, S. & Payne, J. S. (2006). *Corpus Linguistics and Language Development: Research, Assessment, and Pedagogical Innovation*. Invited presentation and workshop at the Berkeley Language Center, University of California, Berkeley.

Thorne, S. & Payne, J. S. (2006). *Corpus Linguistics and Language Development: Research, Assessment, and Pedagogical Innovation*. Invited presentation and workshop at the Second Language Acquisition Institute, University of California, Davis.

Thorne, S. & Payne, J. S. (2005). *Technologies and language learning: Projects, principles, and practices.* Invited presentation at the Language Education Center, University of Pennsylvania.

Payne, J. S. (1999). *The Psycholinguistic and Pedagogical Implications of Integrating Technology into the Second Language Curriculum*. Invited Speaker, Second Language Acquisition Institute, University of California, Davis, CA.

RECENT CONFERENCE PRESENTATIONS

Dahill, B. & Payne, J. S. (2007). *Mobile language immersion: A study in iPod use in listening comprehension*. Paper to be presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), Texas State University, San Marcos, TX.

Payne, J. S & Lipschultz, M. (2007). *Using instant messenger bots and unsupervised learning techniques to promote vocabulary acquisition*. Paper to be presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), Texas State University, San Marcos, TX.

Payne, J. S. (2006). *Data-driven foreign language learning.* Paper presented at the Northeast Association for Language Learning Technology, University of Pennsylvania, PA.

Agafonova, L., Isenberg, N., Lipschultz, M., Payne, J. S., & Tasker, T. (2006). *Podcasting and foreign language education: Current practices and future possibilities.* Panel discussion at the Northeast Association for Language Learning Technology, University of Pennsylvania, PA.

Payne, J. S. & Lipschultz, M. (2006). *The BodoBot project: Pedagogical agents for foreign language learning.* Paper presented at the Penn State Symposium for Teaching and Learning with Technology, University Park, PA.

Ross, B., Payne, J. S., & Pajtek, A. C. (2006). *Data-driven Learning in ESL Writing Instruction.* Paper presented at TESOL, Tampa, FL.

Pajtek, A. C., Payne, J. S., & Ross, B. (2005). *Using data-driven learning techniques to improve academic writing*. Paper presented at the Second Language Research Forum, Teachers College, Columbia University, NY.

Thorne, S. & Payne, J. S. (2005). *Using technology to promote advanced foreign language proficiency*. Paper presented at the Northeast Conference on the Teaching of Foreign Languages, New York, NY.

Payne, J. S. (2005). *Talking with your fingertips*. Paper presented as a part of the CALICO Special Session at the Northeast Conference on the Teaching of Foreign Languages, New York, NY.

Payne, J. S., Sharma, S. & Lee, J. (2005). *A human-computer interaction study of data-driven learning*. Paper presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), East Lansing, MI.

Ross, B. & Payne, J. S. (2005). *Data-driven learning in L2 writing instruction*. Paper presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), East Lansing, MI.

Payne, J. S. (2004). *The tail wagging the dog? A critical look at the reading span test.* Paper presented at the Second Language Research Forum, University Park, PA.

Payne, J. S. (2004). *Moving towards corpus-based language assessment*. Paper presented at the annual EuroCALL Conference, Vienna, Austria.
Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/eurocall2004/

Payne, J. S. & Ross, B. (2004). *Working Memory, Synchronous CMC, and L2 Oral Proficiency Development.* Paper presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), Pittsburgh, PA.
Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/calico2004a/

Payne, J. S., Sanisk, D. & Badalamenti, J. (2004). *Using Student Documentary Film Projects to Promote Advanced Proficiency.* Paper presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), Pittsburgh, PA.
Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/calico2004b/

Payne, J. S. (2004). *Some methodological considerations for comparing face-to-face interaction and SCMC.* Paper presented at the annual meeting of the American Association of Applied Linguistics, Portland, OR.
Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/aaal2004/

Payne, J. S. (2004). *Working Memory and L2 Development: Some Implications for Future Research*. Paper presented at the Working Memory Workshop, University Park, PA.
Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/wm_workshop/

Payne, J. S., Egan, K. & Cieri, C. (2003*). Corpora and Human Language Technologies for Language Learners.* Paper presented at the annual EuroCALL Conference, Limerick, Ireland.
Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/eurocall2003/

Payne, J. S. & Chun, D. (2003*). What makes students click: working memory and look-up behavior.* Paper presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), Ottawa, Canada.

Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/calico2003a/

Payne, J. S. (2003). A (chat) window into L2 production. Paper presented as part of a panel discussion, *Conducting L2 CMC Research: A Panel Discussion of Methodologies, Models, and Techniques,* at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), Ottawa, Canada.
Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/calico2003b/

Payne, J. S., Schaumann, C., Zachau, R. & Coburn, A. (2002). *Collaboration in the Virtual Museum*. Paper presented at the annual meeting of the American Council on the Teaching of Foreign Languages (ACTFL), Salt Lake City, UT.
Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/actfl2002/

Schaumann, C., Payne, J. S. & Coburn, A. (2002). *Collaboration in the Virtual Museum.* Paper presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), Davis, CA.
Abstract available: http://www.calico.org/CALICO02/quads/Schaumann.html

Payne, J. S. & Ceglowski, M. (2001). *Using Image Databases in an Educational Setting Intelligent Searching of Media Databases*. Paper presented at the NERCOMP Image Database SIG, Smith College, MA.
Presentation hand-out available online:
http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/nercomp2001/

Payne, J. S., Williams, L., Simon, A. L. & Pease, A. (2000). *The Internet and Language Proficiency.* Paper presented at the annual meeting of the American Council on the Teaching of Foreign Languages (ACTFL), Boston, MA.

Hokanson, S, Jimenez, J. P., Lozano, M. E., Wittenberg, J., Payne, J. S. & Kogan, J. (2000). *Teaching New University Instructors to Teach Communicatively with and without Computers.* Paper presented at the annual meeting of the American Association of Teachers of Spanish and Portuguese, San Juan, Puerto Rico.

Payne, J. S. (2000). *Chatrooms as Conversation Simulators: L2 Chatroom Interaction and Oral Proficiency Development.* Paper presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), Tucson, AZ.
Available online: http://www.calico.org/00proceedings/sessions/Payne.html

Payne, J. S. & Kogan, J. (1999). *Increasing L2 Production Through Collaborative Research & Writing Online*. Paper presented at the annual meeting of the Rocky Mountain Modern Language Association, Santa Fe, NM.

Payne, J. S. (1999). *WebEdventure: a Web-based tool for collaborative research and writing in second language instruction.* Paper presented at the annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), Oxford, OH.

Peterson, N., Payne, J. S., Erickson, J. & Magee, D. (1999*). Real-Time Assessment of, and Feedback to, Student Essays.* Paper presented at the Tenth Annual Higher Education Assessment Conference, Spokane, WA.

Payne, J. S. (1998). *Demonstration of interactive WebCALL*. Paper presented at the annual meeting of the Rocky Mountain Modern Language Association, Salt Lake City, UT.

Peterson, N. & Payne, J. S. (1998). *Civil War Project: moving the dynamics of collaborative project-based learning into a virtual space*. Paper presented at the Washington State University Interdisciplinary Conference, Pullman, WA.

Payne, J. S., Peterson, N., Cressman, G., Streuli, K. & Smith, B. (1998). *Civil War Project: Student Mentoring and Writing Development in an On-line Collaborative Research and Writing Environment.* Poster presented at the 2nd Annual Distance Education Conference, Bellevue, WA.

WORKSHOPS

Payne, J. S. (2005). SALRC Pedagogy Workshop: Using Technology To Promote Learner-Centered South Asian Language Instruction, South Asia Summer Language Institute (SASLI), University of Wisconsin, Madison, June 6th - 10th.

Thorne, S. & Payne, J. S. (2005). *Using technology to promote advanced language proficiency*. Workshop for the Technology Project as a part of the CALPER professional development summer workshop series.

Thorne, S. & Payne, J. S. (2004). *CALPER technology workshop on computer-mediated communication*. Workshop for the Technology Project as a part of the CALPER professional development summer workshop series.

Payne, J. S. (2002). Workshop on Classroom Research in the Foreign Language Instruction, Williams College, Williamstown, MA.

OTHER PRESENTATIONS

Payne, J. S. (2004). *Technology in L2 Instruction*. Presentation for the Teaching Assistant Preparation Workshop, The Pennsylvania State University, University Park, PA. Available online: http://www.personal.psu.edu/faculty/j/s/jsp17/presentations/ta_prep2004/

Payne, J. S. (2003). *Automated Essay Scoring, Computers as Mentors, and Other Scary Things*. Presentation to the E-learning Special Interest Group, Department of Biology, The Pennsylvania State University, University Park, PA

Thorne, S. & Payne, J. S. (2003). *The Consumption and Production of Culture on the Web*. Invited lecture in French 581, The Pennsylvania State University, University Park, PA.

Thorne, S. & Payne, J. S. (2003). *Communication, Culture, and Internet Technologies*. Invited lecture in French 581, The Pennsylvania State University, University Park, PA.

Payne, J. S. (2001). Participated on a panel discussion addressing the topic of assessment in foreign language instruction. Vassar-Williams Foreign Language Assessment Colloquium, Vassar College, Poughkeepsie, NY.

GRADUATE COMMITTEES

Advisor, Ph.D. Committee. Noelle Isenberg, Department of Germanic and Slavic Languages and Literatures, 2004-present.

Member, Ph.D. Committee, Mathew Carlson, Department of Spanish, 2004-2006.

Member, Ph.D. Committee, Lisa Hundley, Department of Germanic and Slavic Languages and Literatures, 2006-2007.

Member Ph.D. Committee. Muhammad Salem, Department of Curriculum and Instruction, 2005-2007.

Member, Ph.D. Committee, Chi-Yen Chiu, Department of Linguistics and Applied Language Studies, 2003-2005.

Member, M.A. Committee, John Zinck, Department of Linguistics and Applied Language Studies, 2004.

Advisor, D.M.L. Committee. Ralf Borrmann, German School, Middlebury College, 2003

SOFTWARE DEVELOPMENT

Multimedia Online Corpus Analysis tool (MOCA), 2006 – present.
A web-based system for creating and analyzing corpora; in particular, longitudinal learner corpora.

Bodo Bot Project, Spring 2005 – present.
The Bodo Bot project is an SLA research tool in the form of an instant messaging bot that is designed to collect detailed behavior-tracking and self-report data, generate user models, and track language development while serving as a language reference agent in a conversational "wrapper".

Online working memory and information processing measures, 1999 - present.
Reading Span Test
Nonword Repetition Test
Spatial Span Test
Translation Recognition (Spanish)
Simon Task
Second version of working memory tests are available online:
http://www.personal.psu.edu/faculty/j/s/jsp17/tools/wm_tests/

KWICionary, Penn State University, Penn State University, release date October 2004.
An instructional tool for data-driven learning, language learners query a large corpus of German literary, journalistic, and conversational language (feature film transcripts) for

single or multi-word strings. Results are displayed in the keyword-in-context (KWIC) format, customary with concordance software.
Available online: http://kwic.la.psu.edu

Online Corpus Analytic Tool, Penn State University, expected first release Spring 2005.
A web-based system that facilitates the creation, storage, management, and searching of corpora. Capabilities include: calculating corpus statistics, automated parts-of-speech tagging, and a manual coding interface.
Available online: http://ocat.la.psu.edu

Online Research Project Management System, Penn State University, expected first release Spring 2005.
A web-based system for scheduling and managing research projects online. This environment includes: a tool for creating, administering online surveys as well as analyzing data and working memory and other psycholinguistic tests developed with either Macromedia Director or Flash.

Virtual Museum, Center for Educational Technology, Middlebury College, 2001.
A portfolio-style, multimedia writing environment that uses the museum as a metaphor for organizing collaborative groups of learners with a focus on the process of researching, synthesizing, and publishing knowledge. For each exhibit item, exhibit author(s) can receive feedback from their peers in an asynchronous discussion space and from instructors in the form of a critical review, based on learning goals that have been set for the museum. The Virtual Museum also supports the use of non-western languages.
Available online at http://language.la.psu.edu/virtualmuseum/

Chatskeller, Center for Educational Technology, Middlebury College, 2001.
A synchronous discussion tool designed to support real-time communication via text with special features that allow teachers to observe and interact with up to four chatroom groups simultaneously and facilitate research on chatroom discourse. The Chatskeller also supports non-western languages.

CET Online Survey Tool, Center for Educational Technology, Middlebury College, 2001.
Designed and developed a system for authoring and delivering surveys and course evaluations over the Web.

Online Writing Center, Center for Educational Technology, Middlebury College, 2000.
Combines a scheduling feature to facilitate students arranging appointments with writing tutors together with a writing workspace infused with asynchronous and synchronous communication tools.

Online Modules, Center for Educational Technology, Middlebury College, 2002
A system designed to facilitate the online delivery of the technology training curriculum at the Center for Educational Technology. This initiative represents three objectives: 1.) to transform onsite, classroom instruction at the CET into an individualized, project-based approach combining rich self-study resources and intensive mentoring.

Colloquium, Center for Educational Technology, Middlebury College, 2000

Designed and developed the prototype for Colloquium, a tool for web-casting educational events along with accompanying online readings and links to support pre-event preparation by participants, components for synchronous discussion and directing questions to the presenter during the live event, as well as an asynchronous discussion for continued dialog on the topic.

Special Interest Groups Website, Washington State University, 2000.
Designed and developed a website system for supporting the scholarly interaction of academic special interest groups. This system contains an annotated bibliography, threaded discussion, membership directory, information section, and a working papers component.

Learning Management System, Department of Sociology, Washington State University, 2000.
Developed a derivational version of my website system without the specific foreign language support to facilitate other instructional applications. This system includes a course management system and suite of instructional tools.

Online Quizzing Tool, Department of Psychology, Washington State University, 2000.
Developed and maintained a self-assessment tool for students taking introductory-level psychology courses at Washington State University.

Interactive graphing simulation. Department of Sociology, Washington State University, 1999.
Designed and developed an interactive graphing simulation program with Macromedia Director 7.0 together with a faculty member of the Department of Sociology at Washington State University as a vehicle for teaching this faculty member how to use Director 7.0. The program developed helps students develop an understanding of complex relationships by presenting data in graphical form and requiring students to manipulate the graph while providing corrective feedback.

Spanish Accent Tutor, 1999.
Designed and developed a CD-ROM with Macromedia Director 7.0 for a faculty member teaching Spanish socio-linguistics in the Department of Foreign Languages and Literatures at Washington State University.

Learning Management System, Department of Foreign Languages and Literatures, Washington State University, 1998 - 2001.
Developed a database-driven website design that integrates the informational, pedagogical, administrative, and alumni outreach needs of a department. All instructional tools in this system contain authoring templates and are designed to promote collaboration in learning and in teaching through collaborative curriculum building. All learning environments support the use of digital video, audio, images, and text. There are learning tools for: threaded discussion, quizzing, drill-and-practice exercises, placement/proficiency testing, collaborative research and writing, and a chatroom. The course management components enable teachers to keep track of attendance, post lesson plans with homework, maintain student information, and record grades in an online gradebook that is fully integrated with the instructional tools.

LinguaAuthor, 1996.

A digital video-based template program developed with Macromedia Director to facilitate the rapid production of language learning courseware by teachers possessing virtually no programming knowledge.

Bingo Mania, 1995.
Created three vocabulary games using Macromedia Director beginning with sound-picture recognition and progressing to word recognition and production.
Available online:
http://www.personal.psu.edu/faculty/j/s/jsp17/tools/bingomania/Food_Picts_L1.html

Messy Room, 1995.
An info-gap simulation that focuses on developing comprehension of commands and vocabulary. This program uses audio files and animation sequences provide more authentic practice and pique the interest of students. Developed with Macromedia Director.

Bingo Grid-Maker, 1994.
Teaching materials development tool that produces randomized bingo grids and saves them in a grid bank (HyperCard).

Phoneme Tutor, 1992.
Instructional tool for learning phonemic transcription based on sound-symbol relationships (HyperCard).

Bundeswehr Universität (military academy), Hamburg, Germany, 1990.
Developed computer assisted language learning materials for the academy's English language program. These materials consisted of vocabulary, grammar tutorial and reading programs, based on video materials.

Computer Lesson Programmer, Western Washington University, Bellingham, WA, 1987-1988.
Created vocabulary and reading lessons for the German department using the authoring systems for "Le Metro", "Images", "Vocab" and "Lecture".

ONLINE PROJECTS

Civil War Project, Washington State University, 1997.
Co-designed and developed an online learning environment to meet the need for online practicum and mentoring experiences for pre-service teachers at Washington State University; and to develop research, information literacy, and writing skills for eighth-grade students at Enumclaw Junior High School. In addition to being one of the mentors for a group of six of the eighth-graders, I served as the webmaster for the project.

WebEdventure Project, Washington State University, 1999.
Project designer, planner and coordinator. Using WebEdventure, second semester Spanish language students from Washington State University, Delta College in Michigan, and North Harris College in Texas, collaboratively researched eight different Spanish-speaking countries. The objective of their research was to build a database of information on accommodations, restaurants, entertainment, transportation, and other

vital information for travelers in each of the eight chosen countries. The end task for the students was to produce a story of a fictitious trip that they took through one or more of the eight countries, incorporating research data into their stories. Students produced multiple drafts of their stories, each receiving teacher and peer comments.

WebEdventure Project, Washington State University, 1999.
Facilitated a collaborative research and writing project with WebEdventure that supported upper-divisional undergraduate students researching Latin American authors and conducting literary analysis of their work.

WebEdventure Project, Washington State University, 1999.
Facilitated the German version of the virtual trip WebEdventure project.

SERVICE

Executive Board Member, Computer Assisted Language Instruction Consortium, 2006-2009.

Chair, Committee for the transformation of the CALICO Journal and web presence, 2005-present.

Editorial Board Member, *South Asia Language Pedagogy and Technology*, 2006-present.

Conference Program Co-Chair, American Association of Applied Linguistics 2004.

Member of the Web Committee for the Computer Assisted Language Instruction Consortium (CALICO).

Member of the Ad hoc Web Committee for the American Association of Applied Linguistics.

Redesign of CALICO conference and online proposal submission and adjudication system, 2005.

Developed and continue to host a database-driven online proposal submission and adjudication system for AILA 2005.

Developed and continue to host a database-driven online proposal submission and adjudication system for American Association of Applied Linguistics 2004.

Developed and continue to host a database-driven online proposal submission and adjudication system for Second Language Research Forum 2004.

Developed a database-driven online proposal submission and adjudication system for Computer Assisted Language Instruction Consortium (CALICO), 2002.

Co-chair of the Computer-Mediated Communication Special Interest Group of CALICO, 2000-present.

RELATED PROFESSIONAL EXPERIENCE

Software Design Consultant, Houghton Mifflin Co., 2000-2002.
Worked on a development team designing and building online learning environments.

Language Program Developer, Western Washington University, 1990-1996.
Developed ESL programs for the Summer English Language Program (SELP). Activities involve: establishing a marketing network in Korea and Germany, developing advertising materials, and organizing and conducting weekend excursions for SELP students, planning, organizing and leading a two-week study tour in the Western United States.

Translator, CNC-Zentrum, Hamburg, Germany, 1991.
Translated all course manuals and materials, a total of three volumes, for the machine tool courses at the center.

Translator, Vereins- und Westbank, Hamburg, Germany, 1989-1991.
Translated from German to English, company reports and the weekly stock market report which was faxed and telexed to over 200 customers in 12 different countries.

AWARDS AND SCHOLARSHIPS

CALICO Outstanding Article Award, CALICO Journal 2004-2005.

CALICO-Sony Outstanding Article Award, CALICO Journal 2002-2003.

Olsen-Rounds Fellowship ($1,500) for promoting writing in foreign language education., Washington State University, April 1999.

International Rotary Exchange Scholarship, Weinheim, Germany, 1981-82.

LANGUAGES

Speak, read and write German fluently; limited spoken fluency in French, and some reading and writing skills; very limited speaking, reading and writing abilities in Korean.

TRAVEL AND EXPERIENCE ABROAD

Resided in Kwangju, South Korea, Fall 1993 to Summer 1996.

Resided in Hamburg, Germany, 1988-1991.

Studied and resided in Montpellier, France, September 1985 to April 1986.

Studied and resided in Weinheim, Germany, April through August 1986.

Rotary Exchange Student in Weinheim, Germany, 1981-82.

Traveled to Italy, Denmark, Czechoslovakia, Yugoslavia, Austria, Switzerland and former East Germany.

Appendix A: Project Timeline

Fall 2009	• Install and configure server. • Begin constructing Portuguese corpus. • Develop Portuguese conversational prompts and responses. • Begin constructing Chinese corpus. • Develop Chinese conversational prompts and responses. • Install and test the Stanford Chinese word segmenter. • Install and test the Stanford Chinese parser. • Begin constructing the Spanish corpus. • Develop Spanish conversational prompts and responses. • Begin integrating the existing Portuguese NLP tools into LangBot. • Optimize algorithms, testing modes, and conversational prompts and responses for Portuguese, Chinese, and Spanish.
Spring 2010	• Finish the integration of the Portuguese NLP tools into LangBot. • Begin the adaptation of the Portuguese NLP tools to Spanish. • Continue developing corpora for Portuguese, Chinese, and Spanish. • Conduct internal testing of Portuguese, Chinese, and Spanish versions of LangBot without the NLP feedback modules.
Summer 2010	• Adapt TAGARELA's grammar and parser to LangBot. • Continue adapting Portuguese NLP tools to Spanish. • Begin developing NLP feedback modules for Chinese. • Complete Portuguese, Chinese, and Spanish corpora. • Construct project website with interface to user data for students.
Fall 2010	• Conduct first pilot study of the Portuguese, Chinese, and Spanish versions of LangBot without the NLP feedback modules. • Conduct unit tests for Portuguese NLP tools in LangBot. • Continue developing NLP feedback modules for Chinese. • Begin writing the Spanish grammar for LangBot parser. • Make the Portuguese, Chinese, and Spanish corpora and word frequency lists freely available for download from the project website.
Spring 2011	• Conduct second pilot study of the Portuguese, Chinese, and Spanish versions of LangBot without the NLP feedback modules. • Start implementation of LangBot NLP feedback modules for all three languages. • Conduct unit tests for Spanish NLP tools in LangBot. • Start integrating output of Chinese NLP modules into LangBot NLP feedback modules.

Summer 2011	• Complete implementation of LangBot NLP feedback modules. • System testing and pilot of Portuguese, Chinese, and Spanish versions of LangBot with full NLP error analysis capabilities.
Fall 2011	• Begin large-scale studies of Portuguese, Chinese, and Spanish versions of LangBot with full NLP error analysis capabilities. • Analyze data from large-scale studies in Portuguese, Chinese, and Spanish
Spring 2012	• Continue large-scale studies for Portuguese, Chinese, and Spanish versions of LangBot with full NLP error analysis capabilities. • Project investigators present results and give workshops at national professional meetings for language teachers. • Continue analyzing data from large-scale studies and optimizing NLP error analysis tools.
Summer 2012	• Continue large-scale studies for Portuguese, Chinese, and Spanish versions of LangBot with full NLP error analysis capabilities. • Project investigators present results and give workshops at national professional meetings for language teachers. • Continue analyzing data from large-scale studies and optimizing NLP error analysis tools. • All project source code and language resources made available under the Creative Commons license.

Appendix B: References

Amaral, Luiz (2007). *Designing Intelligent Language Tutoring Systems for Integration into Foreign Language Instruction*. Ph.D. Dissertation. The Ohio State University. Columbus-OH.

Amaral, Luiz and Detmar Meurers (2007). Conceptualizing Student Models for ICALL. In Cristina Conati and Kathleen F. McCoy (Eds.): *User Modeling 2007: Proceedings of the Eleventh International Conference, Lecture Notes in Computer Science*. Wien, New York, Berlin: Springer.

Amaral, Luiz and Detmar Meurers (2008). From Recording Linguistic Competence to Supporting Inferences about Language Acquisition in Context. In *Computer Assisted Language Learning*. 21 (5). Routledge.

Bailey, S. & D. Meurers (2008). Diagnosing meaning errors in short answers to reading comprehension questions. In J. Tetreault, J. Burstein & R. D. Felice (eds.), Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications, held at ACL 2008. Columbus, Ohio: Association for Computational Linguistics, pp. 107–115. URL http://aclweb.org/anthology-new/W/W08/W08-0913.pdf.

Bernstein, J., Najmi, A., & Ehsani, F. (1999). Subarashii: Encounters in Japanese spoken language education. CALICO Journal, 16 (3), 361-384.

Bick, E. (2000). The Parsing System “Palavras”: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus, Denmark: Aarhus University Press.

Bick, E. (2004). PaNoLa: Integrating Constraint Grammar and CALL. In H. Holmboe (ed.), Nordic Language Technology, Arbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004 (Yearbook 2003), Copenhagen: Museum Tusculanum, pp. 183–190.

Bull, S. (1994). Student Modeling for Second Language Acquisition. *Computers and Education, 23* (1-2), 13-20.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System,* 25 (3), 301-315.

Dalby, J., & Kewley-Port, D. (1999). Explicit pronunciation training using automatic speech recognition. *CALICO Journal, 16* (3), 425-446.

Flesch, R. F. (1948). A new readability yardstick. Journal of Applied Psychology, 32:221-233.

Gamper, J., & Knapp, J. (2002). A review of intelligent CALL systems. Computer

Assisted Language Learning, 15 (4), 329-342.

Harless, W., Zier, M., & Duncan, R. (1999). Virtual dialogues with native speakers: The evaluation of an interactive multimedia method. *CALICO Journal, 16* (3), 313-338.

Heift, Trude (2003), 'Multiple learner errors and meaningful feedback: A challenge for ICALL systems', CALICO Journal 20(3), 533–548.

Holland, M. V., Kaplan, J. D., & Sama, M. R. (Eds.). (1995). Intelligent language tutors: Theory shaping technology. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Holland, M., Kaplan, J. D., &Sabol, M. A. (1999). Preliminary tests of language learning in a speech-interactive graphics microworld. CALICO Journal, 16 (3), 339-360.

Karlsson, F., A. Voutilainen, J. Heikkil & A. Anttila (eds.) (1995). *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. Berlin and New York: Mouton de Gruyter.

Kuenning, G. (2005). International Ispell Web, Version 3.3.02. URL http://ficus-www.cs.ucla.edu/geoff/ispell.html.

LaRocca, S., Morgan, J., & Bellinger, S. (1999). On the path to 2X learning: Exploring the possibilities of advanced speech recognition. CALICO Journal, 16 (3), 295-310.

Lau, T.P. & King, I. (2006). Bilingual Web Page and Site Readability Assessment. ACM 1-59593-323-9/06/0005.

Laufer, B. & Nation, P. (1999). A vocabulary-size test of controlled productive ability. Language Testing, 16(1), 33-51.

Martins, R., R. Hasegawa & M. das Grac¸as Nunes (2006). Curupira: a functional parser for Brazilian Portuguese. In Computational Processing of the Portuguese Language, 6th International Workshop, PROPOR. Lecture Notes in Computer Science 2721. Faro, Portugal: Springer.

Mostow, J., & Aist, G. (1999). Giving help and praise in a reading tutor with imperfect listening—because automated speech recognition means never being able to say you're certain. CALICO Journal, 16 (3), 407-424.

Nagata, Noriko (2002), BANZAI: An application of natural language processingto web based language learning, CALICO Journal 19(3), 583–599.

Ortega, L. & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: recent trends and future directions, Annual Review of Applied Linguistics, 25, 26–45.

Payne, J.S. and Lipschultz, M. (2007). *Using Instant Messenger Bots and Unsupervised Learning Techniques to Promote Vocabulary Acquisition*. The annual meeting of the Computer Assisted Language Instruction Consortium (CALICO), San Marcos, TX.

Rypa, M., & Price, P. (1999). VILTS: A tale of two technologies. CALICO Journal, 16 (3), 385-404.

Stevens, V. (1991). Concordance-based vocabulary exercises: A viable alternative to gap-fillers. In Johns, T. & King, P. (eds.) *Classroom concordancing: English Language Research Journal, 4* (pp. 47-63). University of Birmingham: Centre for English Language Studies.

Tucker, R. G. (2000). The applied linguist, school reform, and technology: Challenges and opportunities for the coming decade. CALICO Journal, 17(2), 197-221.

Yang, S. J. (1971). A readability for Chinese language. Ph.D. Thesis for Mass Communication, University of Wisconsin.

Zacharski, (2002). *Conversational agents for language learning*. Unpublished manuscript.

Zacharski, (2003). A discourse system for conversational characters. In A. Gelbukh (Ed.), *Proceedings of the fourth international conference on intelligent text processing and computational linguistics* (pp. 492-495). Heidelberg: Spring Verlag.

Luiz Alexandre Amaral, Ph.D.

(amaral@spanport.umass.edu)

Home page: http://people.umass.edu/amaral/

Professional address:
Spanish and Portuguese
Department of Languages, Literatures and Cultures
University of Massachusetts
416 Herter Hall
Amherst-MA, USA 01003
Phone: (413) 545-4790

RESEARCH INTERESTS

- **Intelligent Computer Assisted Language Learning (ICALL):** use of innovative technology in ICALL environments; development of new ICALL tools to support individualized and classroom instruction; integration of new and existing ICALL tools into content-based approaches to foreign language instruction.
- **Intelligent Language Tutoring Systems (ILTS):** use of natural language processing technology to develop ILTS; conceptualization and development of student and instruction models for ILTS based on cognitive models of second language acquisition; activity design for ILTS.
- **Computational Linguistics:** syntactic ill-formed input processing for ICALL tools, especially the adaptation of parsing strategies to deal with learner language, and the development of shallow semantic processing mechanisms for different activity types.
- **Theoretical Linguistics:** Constraint-based syntactic theories, especially Head-driven Phrase Structure Grammar (HPSG); unbounded dependency constructions; compositionality in complex prepositions.

EDUCATION

Ph.D. Hispanic Linguistics, 2007.
The Ohio State University, USA.
Dissertation: ***Designing Intelligent Language Tutoring Systems: integrating natural language processing technology into foreign language***

teaching practice.

Advisors: Detmar Meurers (Linguistics) and John Grinstead (SPPO).

M.A. Linguistics (Summa Cum Laude), 2001.

Department of Letters, Pontifical Catholic University of Rio de Janeiro, Brazil.

Thesis: ***Seleção Semântica de Complementos Verbais em Head-driven Phrase Structure Grammar (HPSG).***

Advisor: Violeta Quental

Specialization in Teaching English as a Foreign Language – Special Didactics, 1995.

Department of Education, State University of Rio de Janeiro (UERJ), Brazil.

B.A. Letters – Language and Literature, 1997.

Department of Letters, State University of Rio de Janeiro (UERJ), Brazil.

APPOINTMENTS

Current Position:

Assistant Professor of Hispanic Linguistics
Department of Languages, Literatures and Cultures
University of Massachusetts
Amherst - MA

Previous Positions:

2007-2008 Assistant Professor of Hispanic Linguistics
Department of Hispanic and Italian Studies
University of Victoria
British Columbia - Canada

2005-2007 Graduate Research Associate
Department of Linguistics
The Ohio State University
Columbus - OH

2002-2005 Graduate Teaching Associate of Spanish and Portuguese
Department of Spanish and Portuguese

	The Ohio State University Columbus – OH
2003-2004	Adjunct Faculty of Spanish Department of Modern Languages Columbus State Community College Columbus - OH
1999-2002	Language Teacher Swiss School of Rio de Janeiro Rio de Janeiro – Brazil
1996-2002	English Teacher Sociedade Brasileira de Cultura Inglesa Rio de Janeiro – Brazil
1999-2002	Foreign Language Teacher Curso QI Rio de Janeiro – Brazil
1999	Foreign Language Teacher. Curso PH Rio de Janeiro – Brazil
1994	English Teacher Auding Idiomas Rio de Janeiro – Brazil
1993-1994	English Teacher Cursos para a Comunidade State University of Rio de Janeiro (UERJ) Rio de Janeiro – Brazil
1993-1994	English Teacher Feedback (Curso Interaction Ltda) Rio de Janeiro – Brazil

Other Professional Experience:

1999 – 2002 Foreign Language Coordinator
Swiss School of Rio de Janeiro, Brazil.
Responsibilities: Selecting and training language teachers, choosing and preparing material for classes, and coordinating the preparation of students for different English certificates by the University of Cambridge (KET, PET, FCE, and CAE).

1996 Member of the Foreign Language Examining Board
Fundação Escola do Serviço Público (FESP).
Responsibilities: Elaborating public exams carried out by the Foundation. Participation on the following exams: Exam for primary and secondary teachers of *Centro de Ensino Integrado* (CEI), February, 1996; Exam for primary and secondary teachers of the city of Macaé, May, 1996; *Supletivos* (public certificate of adult secondary education provided by the government of the state of Rio de Janeiro), July and September, 1996.

PUBLICATIONS

Papers:

To appear Amaral, Luiz and Detmar Meurers. Little Things with Big Effects: On the Identification and Interpretation of Tokens for Error Diagnosis in ICALL. In *CALICO Journal*. 27 (1).

To appear Amaral, Luiz. The Double Subcategorization of Tough-Adjectives in Brazilian Portuguese. In *Locality of Grammatical Relationships.* Edited by R.D. Levine and D.W. Meurers.

2008 Amaral, Luiz and Detmar Meurers. From Recording Linguistic Competence to Supporting Inferences about Language Acquisition in Context. In *Computer Assisted Language Learning*. 21 (5). Routledge.

2007 Amaral, Luiz and Detmar Meurers. Conceptualizing Student Models

for ICALL. In Cristina Conati and Kathleen F. McCoy (Eds.): User Modeling 2007: *Proceedings of the Eleventh International Conference*, Lecture Notes in Computer Science. Wien, New York, Berlin: Springer.

2006 Feldman, Anna, Jirka Hana, Luiz Amaral and Chris Brew. Tagging Portuguese with a Spanish Tagger. In *Proceedings of the Workshop on Cross-Language Knowledge Induction at the 11th Conference of the European Chapter of the Association for Computational Linguistics* (EACL). April 3-7. Trento, Italy.

2005 Amaral, Luiz. Desafiando a Localidade. In *Palavra* Journal, n.12, p 119 - 132. PUC-Rio. Rio de Janeiro, Brazil.

2004 Amaral, Luiz. A Forma do Objeto Direto Anafórico em Português - Uma análise motivada pela topicalidade. In *Cadernos do Congresso Nacional de Lingüística e Filologia,* Vol. VIII, n.14, p 9-22. Rio de Janeiro, Brazil.

2003 Amaral, Luiz. Um Modelo de Restrições Semântico-Selecionais para Sistemas de Processamento de Linguagem Natural. In Journal *Veredas*, v.9, p 31 - 45. Juiz de Fora, Brazil.

2003 Oliveira, Claudia, Luiz Amaral, and Milena Garrão. Recognizing Complex Prepositions Prep+N+Prep as Negative Patterns in Automatic Term Extraction from Texts. In *Proceedings of 1st Workshop em Tecnologia da Informação e da Linguagem Humana* (TIL2003). São Carlos, Brazil.

Project Proposals:

2005-2007 `Bridging the Gap between Research in Natural Language Processing and Individualized Language Instruction'. Co-authored the proposal with PI Detmar Meurers and Glaucia Silva. Arts and Humanities Grants for Innovation. Office of Research, The Ohio State University. The application was successful. $46,000, 2005-2007.

2005 `Natural Language Processing for Intelligent Computer-Aided Language Learning—Advancing Technology and Integration in Teaching'. Co-authored this proposal with PI Detmar Meurers to NSF (Information and Intelligent Systems, Human Language and Communication), 2004. The NSF panel recommended `fund if possible', but there were not enough funds to be distributed. We are planning on resubmitting a revised version.

Unpublished Manuscripts:

2001 Amaral, Luiz. *Seleção Semântica de Complementos Verbais em Head-driven Phrase Structure Grammar (HPSG).* Unpublished Master Thesis. PUC-Rio.

PRESENTATIONS

Peer-reviewed Conferences:

2008 Amaral, Luiz, and Detmar Meurers. *Tapping into the synergy between SLA, Foreign Language Instruction and Natural Language Processing in ICALL.* In AILA: 15th World Congress of Applied Linguistics. Essen, Germany. August 25.

2008 Amaral, Luiz, and Detmar Meurers. *Little Things With Big Effects: On the identification and interpretation of tokens for error diagnosis in ICALL.* In Computer-Assisted Language Instruction Consortium (CALICO 2008). University of San Francisco. March 18 – 22.

2007 Amaral, Luiz, and Detmar Meurers. *Extending Learner Models for Intelligent Computer-Assisted Language Learning Beyond Grammar.* Conference on Technology for Second Language Learning. Iowa State University. Ames, Iowa. September 21 – 22.

2007 Amaral, Luiz, and Detmar Meurers. *Putting Activity Models in the Driver's Seat: Towards a demand-driven NLP architecture for ICALL.* European Association for Computer Assisted Language Learning (EUROCALL). University of Ulster, Ireland. September 5 – 8.

2007 Amaral, Luiz. *Creating ICALL Activities for Task-Based Instruction.* Computer Assisted Language Instruction Consortium (CALICO). Texas State University. May 22 – 26.

2007 Amaral, Luiz, and Detmar Meurers. *Designing Learner Models for Intelligent Language Tutors.* Computer Assisted Language Instruction Consortium (CALICO). Texas State University. May 22 – 26.

2006 Amaral, Luiz, and Detmar Meurers. *Using a Foreign Language Tutoring System for Grammatical Feedback.* European Association for Computer Assisted Language Learning (EUROCALL). University of Granada, Spain. September 4 - 7.

2006 Amaral, Luiz, and Detmar Meurers. *Where Does ICALL Fit into Foreign Language Teaching?* Computer Assisted Language Instruction Consortium (CALICO). University of Hawaii, Honolulu. May 16 - 20.

2006 Amaral, Luiz, Vanessa Metcalf, and Detmar Meurers. *Language Awareness Through Re-use of NLP Technology*. Workshop NLP in CALL at the Computer Assisted Language Instruction Consortium (CALICO). University of Hawaii, Honolulu. May 16.

2006 Amaral, Luiz, Detmar Meurers, and Gláucia Silva. *Using Computer Assisted Language Learning Systems to Support Portuguese Instruction.* Fifth International Conference of the American Portuguese Studies Association (APSA). University of Minnesota, Minneapolis. October 5 – 7.

2006 Silva, Gláucia, and Luiz Amaral. *Communicative Abilities of Learners in an Individualized Instruction Program.* Eighth International Conference of the Brazilian Studies Association (BRASA). Vanderbuilt University. Nashville, TN. October 13 – 16.

2005 Silva, Gláucia, and Luiz Amaral. *Challenges for Establishing a Portuguese Individualized Instruction Program in the US.* South Atlantic Modern Language Association (SAMLA). Atlanta, GA. November 5.

2005 Amaral, Luiz. *The Role of the Personal Infinitive in BP Tough-Constructions.* Niagara Linguistic Society at the University at Buffalo. Buffalo, NY. September 30.

2005 Amaral, Luiz. *When Semantics Changes Syntax: the subcategorization of adjectives in Brazilian Portuguese.* V Workshop in Formal Linguistics at USP. University of São Paulo (USP), São Paulo, Brazil. August 25-26.

2005 Amaral, Luiz. *Locality in Weak Unbounded Dependency Constructions: an HPSG approach for the Portuguese tough-constructions.* 8th Ohio State University Symposium on Hispanic and Luso-Brazilian Literatures, Linguistics, and Cultures. Ohio State University. Columbus, OH April 29.

2005 Amaral, Luiz and Gláucia Silva. *Integrating CALL tools into the Portuguese Classroom*. 58th Annual Kentucky Foreign Language Conference. University of Kentucky. Lexington, KY. April 21 – 23.

2005 Amaral, Luiz. *Using Computers to Teach Foreign Language.* Ohio Latinamericanist Conference. Ohio State University. Columbus, OH. March 4.

2004 Amaral, Luiz. *Um Modelo de Restrições Semânticas de Complementos Verbais para o PB.* The Fourth International Conference of the American Portuguese Studies Association (**APSA**). University of Maryland, College Park. October 14-16.

2004 Amaral, Luiz. *A Forma do Objeto Direto Anafórico em Português - Uma análise motivada pela topicalidade.* VIII Congresso Nacional de Lingüística e Filologia. Universidade Estadual do Rio de Janeiro (UERJ). Rio de Janeiro, Brazil. August 23 – 27.

2004 Amaral, Luiz. Definindo *Locuções Prepositivas em Português: Composicionalidade de Expressões do Tipo [em+N+de]*. 57th Annual Kentucky Foreign Language Conference. University of Kentucky. Lexington, KY. April 15-17.

2003 Oliveira, Claudia, Luiz Amaral, and Milena Garrão. *How Complex are*

Complex Prepositions? 7th Hispanic Linguistics Symposium (**HLS**). University of New Mexico. October 16-18.

2003 Oliveira, Claudia, Luiz Amaral, and Milena Garrão. *Complex Prepositions and the Internal Structure of NPs.* 6th Annual Ohio State University Graduate Student Symposium on Hispanic and Luso-Brazilian Literature, Linguistics, and Cultures. March 3.

Invited Talks:

2007 *Intelligent Computer-Assisted Language Learning - An Opportunity for Interdisciplinary Research.* First OSU Spring Linguistics Symposium. Columbus – OH. April 9.

2006 *Intelligent Language Tutoring Systems: Research and Implementation.* Pontifical Catholic University of Rio de Janeiro (PUC-Rio). Rio de Janeiro, Brazil. September 1.

2005 *Creating Intelligent Exercises: Using NLP techniques in CALL*. Kenyon College. Gambier, OH. November 7.

2005 *Processamento Sintático e Representações Lingüísticas: formalização e teoria. University of São Paulo* (USP). São Paulo - Brazil. August 24.

2005 *Bridging the Gap Between Natural Language Processing and the Language Teaching Practice in ICALL Research.* Núcleo Interinstitucional de Lingüística Computacional (NILC). São Carlos, Brazil. August 19.

2004 *HPSG e Localidade.* Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil. September 10.

Presentations about Brazilian Culture, History and Politics:

2006 *Brazil in the 21st Century*. Worthington International Friends Association (WIFA). Worthington, OH. June 15.

2006 *Brazilian History and the Current Political Situation and Latin*

America. Great Decisions Series. Upper Arlington Senior Citizen Center. Upper Arlington, OH. May 24.

2006 *Brazilian Culture to Discover*. College of Education at the Ohio State University. Exchange Graduate Student Program with UFG (Universidade Federal de Goiás). Columbus, OH. April 3.

2006 *Brazil – The Unknown Giant.* Global Hotspots Workshop for Columbus high school teachers. Columbus, OH. March 8.

2006 *Youth Culture in Brazil*. Introduction to Brazilian Culture (Port 303). The Ohio State University. Columbus, OH. March 2.

2006 *Brazil – from JK to FHC.* Introduction to Brazilian Culture (Port 303). The Ohio State University. Columbus, OH. February 7.

TEACHING EXPERIENCE AT UNIVERSITY LEVEL IN NORTH AMERICA

University of Massachusetts:

SPAN 470 (Introduction to Hispanic Linguistics)
SPAN 597 (Foreign Language Teaching Methods)
SPAN 697AH (Topics in Second Language Acquisition)

University of Victoria:

SPAN 497 (Introduction to Hispanic Linguistics)
SPAN 450A (Advanced Composition, Translation and Stylistics)
SPAN 250A (Review of Grammar and Conversation)
SPAN 250B (Review of Grammar and Conversation)

The Ohio State University:

PORT 501 (Portuguese for Spanish Speakers 1)
PORT 101.01 (Elementary Portuguese 1)
PORT 102.01 (Elementary Portuguese 2)
PORT 101.51 (Elementary Portuguese Individualized Instruction 1)
PORT 102.51 (Elementary Portuguese Individualized Instruction 2)
PORT 103.51 (Intermediate Portuguese Individualized Instruction 1)

PORT 104.51 (Intermediate Portuguese Individualized Instruction 2)
SPAN 102.66 (Intensive Spanish for Review 1)
SPAN 103.01 (Intermediate Spanish 1)
SPAN 104 (Intermediate Spanish 2)
SPAN 401 (Advanced Spanish Grammar)

Columbus State Community College:
SPAN 101 (Elementary Spanish 1)
SPAN 102 (Elementary Spanish 2)
SPAN 103 (Intermediate Spanish 1)

SERVICE:

Conference Organization:

Co-chair of the Workshop on Interfaces of Intelligent Computer-Assisted Language Learning (IICALL–2006). The Ohio State University. December 14-17, 2007.

Co-chair of the 8th Ohio State University Symposium on Hispanic and Luso-Brazilian Literatures, Linguistics, and Culture. The Ohio State University, April, 2005.

Abstract Reviewing for Conference:

Member of the Program Committee of WorldCALL 2008. Fukouka, Japan. (http://www.j-let.org/~wcf/modules/tinyd8/index.php?id=2)

University Service:

University of Massachusetts:

Chair of the Language Program Committee

Director of the Master's of Arts in Teaching

Director of the Secondary Teacher Education Program

Member of the Graduate Studies Committee

Member of the Educator Licensure Advisory Council.

Member of the Dean's Ad Hoc Advisory Council on the Digital

University of Victoria:

Member of the Curriculum Committee of Hispanic and Italian Studies.

Member of the Graduate Studies Committee.

Member of the Hispanic and Italian Study Abroad Committee.

Hispanic and Italian website co-coordinator.

Member of the department's working group on teaching strategies.

Departmental representative in the Humanities Computing and Media Center (HCMC).

Member of the University's Computer Mediated Communication working group.

Member of the departmental Equity Sub-committee.

The Ohio State University:

Member of the recruiting committee of Portuguese. Department of Spanish and Portuguese.

Member of the Fulbright evaluation panel at The Ohio State University for the English Teaching Assistantship for Brazil.

FELLOWSHIPS AND GRANTS

2005	The Arts and Humanities Grant for Innovation. Office of Research at The Ohio State University. $46,000 to implement and ICALL system (TAGARELA) for the Individualized Instruction Program.
2005	Gordon P.K. Chu Memorial Fellowship. $2,000 to spend Summer 2005 in the University of São Paulo (USP) - São Carlos investigating and collecting natural language processing (NLP) resources to develop TAGARELA.
2003	Spanish and Portuguese travel grant. $300 to attend the 7th HLS.
2001 – 2002	PUC-Rio fellowship for M.A. students. Fees and Tuition for the duration of the course.

OTHER COURSES AND CERTIFICATES

Spanish – Curso de Español del Consulado General de Argentina.
Rio de Janeiro – 2001

English – Certificate of Proficiency University of Cambridge.
England – 1993.

Certificate of Proficiency University of Michigan.
USA – 1991.

Didactics – Professional Development Workshop
Teaching Spanish at the Intermediate Level
The Ohio State University, USA – 2003

Teacher Training Course (TTC)
Cultura Inglesa, Brazil – 1998.

Teacher Training Course (TTC)
Instituto Brasil - Estados Unidos, Brazil – 1996.

LANGUAGES

Portuguese:	Native speaking, reading and writing proficiency
Spanish:	Near-native speaking, reading and writing proficiency
English:	Near-native speaking, reading and writing proficiency
French:	Near-native speaking, and reading, and excellent writing proficiency
Italian:	Elementary speaking, and reading proficiency.

COMPUTING SKILL

Programming Languages:	Python, Prolog
Operating systems:	Mac OS X, Microsoft Windows, Unix, Linux
Typesetting/markup:	LaTeX, HTML

PROFESSIONAL AFILIATIONS

American Association of Applied Linguistics (AAAL)
Computer Assisted Language Instruction Consortium (CALICO)
American Portuguese Studies Association (APSA)

REFERENCES

Professor Detmar Meurers
The Ohio State University
Department of Linguistics
1712 Neil Ave
222 Oxley Hall
Columbus, OH
43210 - USA
Tel: (614) 292-0461
e-mail: dm@ling.ohio-state.edu

Professor Gláucia Silva
UMass Dartmouth
Department of Portuguese
285 Old Westport Rd
Dartmouth, MA
02747 - USA
Tel: (508) 999-8271
e-mail: gsilva@umassd.edu

Professor Donna Long
The Ohio State University
Department of Spanish and Portuguese
1775 College Rd
298 Hagerty Hall
Columbus - OH
43210 - USA
Tel: (614) 292-8864
e-mail: long.25@osu.edu

Weijia Li

Webster 109 413-542-2126
Amherst College, MA 01002 wli@amherst.edu

EDUCATION

University of Massachusetts - Amherst, Amherst MA

- Doctorate in Educational Technology 2004
- Master in Asian Languages and Literatures (expected in 2007)
- Master in Teacher Education and Curriculum Studies 2001

City University of New York, New York, NY

- M.A. program in TESL (Teaching English as a Second language) 1996

Beijing University of Science and Technology, Beijing, China

- M.A. in Applied Linguistics and Language Teaching 1987

China Broadcast and Television University, Beijing, China

- B.A. program in Chinese language and Literature 1985

Hebei Teachers' University, Shijiazhuang, China

- B.A. in English Language and Literature
  1982

WORK EXPERIENCE

Amherst College, Amherst, MA 7/06-present

- Director of the Chinese program
- Senior Lecturer in the Department of Asian Languages and Civilizations

Smith College, Northampton, MA 9/01-6/06

- Lecturer in the Department of East Asian Languages and Literatures
  Taught Chinese language courses of all levels
- Technologist: Teaching with Technology Committee
  Interviewed faculty who used technology in teaching, wrote articles on their achievement, designed websites to showcase their work
- Website Manager:
  Managed website updating for the Department of East Asian Languages and Literatures

University of Massachusetts- Amherst, Amherst MA 9/99-05/06

- Instructor/Teaching Associate: Department of Asian Languages and Literatures
  Taught Chinese language courses of all levels
- Tutorial program coordinator: United Asian Learning Resource Center
  Managed the tutorial program, hired and trained tutors, gave students academic advice
- Course facilitator: School of Education
  Facilitated Education 497 (Student Tutoring in Schools) and
  Education 597 (Leadership in Multicultural Tutoring)
- Web designer: School of Education

Designed websites for STEP (Secondary Teacher Education Program) and TEAMS (Tutoring Enrichment Assistance Models for Schools)

Beijing University of Science and Technology, Beijing, China 9/87-9/99

- Associate Professor: Foreign Languages Department
  Taught English language courses of all levels
- Chair: Foreign Languages Department
  Hired and trained teachers
  Supervised academic affairs
- Member of the university policy-making committee
  Made decisions on academic affairs and school management

United Nation and **World Bank**, Beijing, China 8/98-8/99

- Translator and interpreter:
  Translated documents and interpreted for conferences on educational and Environmental Programs in China

China Central Broadcast and Television University, Beijing, China 10/82-12/84

- Instructor: Taught English courses of all levels.
- Program designer: made radio and TV programs for distance learning

PUBLICATIONS

- **Reading Plus**, China Central Broadcast and Television University Press 2001
- **College English**, Higher Education Press, China 1999
- **Meet Your Urgent Need** (Book and CD-ROM), Beijing University Press 1998
- **A Guide to College English Test**, Chengdu Science & Technology Press 1997
- **Extensive Reading for College English**, Shanghai Jiaotong University Press 1994
- **A Guide to English Learning,** Qingdao Oceanology University Press, China 1993
- **Selected Readings for College Students,** Qingdao Publishing House, China 1993
- **Two-way Practice for English Listening and Writing,** Shanghai University Press 1991
- **The Important Role of Teachers in Class Organization**, High Education Research 1990

AWARDS

- Professional Development Grant In Instructional Technology for Academic Development
  University of Massachusetts 2005
- Award of Appreciation, United Asia Learning Resource Center
  University of Massachusetts, Amherst 2004
- Full scholarship as a visiting scholar, Ministry of Education, China 1994
- Title of Excellent Teacher, Education Bureau of Beijing 1993
- First-prize winner in Teaching Performance Competition,
  Beijing University of Science and Technology 1990
- Title of National Excellent Teacher, Ministry of Education, China 1989

Budget Narrative

Budget Narrative

Attachment 1:
Title: Pages: Uploaded File: **1234-Budget Narrative.pdf**

Budget Narrative

Personnel

Scott Payne, PI - The principal investigator is contributing his time at no charge to the budget. Dr. Payne's time includes both technical and project management functions, as well as performing all data analyses.

Weijia Li, Co-Investigator, is budgeted at 100% of two months summer salary in year 1 and 100% of one month of summer salary for year 2 of the project, using her actual salary in the calculation. Dr. Li's time includes Chinese corpus and language prompt development, and test functions, as well as dissemination at national conferences.

Michael Lipschulz, IM Bot Programmer - The programmer will be paid wages at a rate of 10 hours/week @ $40/hour for 20 weeks for years 1 and 2. In year 3 his hours will be reduced to 10 hours/week @ $40/hour for 10 weeks. He and will perform all programming required for the project and assist in data analysis.

Subcontract to the University of Massachusetts: Luiz Amaral, Co-PI, is budgeted for 100% of two months summer salary in all three years of the project, using his actual salary in the calculation. Dr. Amaral's time includes development of the NLP error feedback tools for Portuguese, Spanish, and Chinese.

TBN, Chinese Computational Linguist – the amount of $10,000 is reserved for each of the three years of the project to pay for the work on extending the Stanford Chinese word segmenter and parser to provide error feedback on spelling, morphology and syntax.

TBN, Research Assistants - The research assistant wages are calculated as wage payroll at a rate of $12/hour for 400 hours during the first two years of the project.

For project time occurring after June 30 of any given year the salaries have been adjusted for a 4.5% merit increase.

Fringe Benefits
Fringe benefits are computed using the rates of 29.7% applicable to Category I Salaries and 7.65% applicable to Category III Salaries and Wages for the current fiscal year – July 1, 2008 through June 30, 2009.

Travel
Funds are requested to cover lodging, meals and roundtrip airfare to present project findings at national and international professional meetings and for the external evaluator to visit the Amherst College campus to perform his evaluation.

External Evaluator

Robert Blake, External Evaluator – Dr. Blake will perform an external evaluation at the

end of year 2 and year 3. He will perform extensive testing at UC Davis, meet with all project investigators, examine the data analyses and results for each language, interview students who have had extensive experience with LangBot in focus groups, and write two reports during the project.

Equipment

Funds are requested for a dedicated web server, which is essential to the project. Due to the high level of usage anticipated, it is necessary to have a dedicated server for running the LangBot application.

Facilities and Administrative Costs

Facilities and Administrative Cost rates are negotiated and approved by the Department of Human and Health Services, Amherst College's cognizant federal agency. The current on-campus rate is 58% of salary, wages and fringes from July 1, 2006 to June 30, 2009.