C. Project Description

C.a. List of Participants

<table>
<thead>
<tr>
<th>Name</th>
<th>Department</th>
<th>Name</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geoffrey Aguirre</td>
<td>Neurology</td>
<td>Matthew Botvinick</td>
<td>CCN**</td>
</tr>
<tr>
<td>David Brainard</td>
<td>Psychology</td>
<td>Robin Clark</td>
<td>Linguistics</td>
</tr>
<tr>
<td>Anjan Chatterjee</td>
<td>Neurology</td>
<td>Dorothy Cheney</td>
<td>Biology</td>
</tr>
<tr>
<td>David Embick</td>
<td>Linguistics</td>
<td>Branch Coslett</td>
<td>Neurology</td>
</tr>
<tr>
<td>Steven Gross</td>
<td>Philosophy</td>
<td>Amishi Jha</td>
<td>Psychology</td>
</tr>
<tr>
<td>Aravind Joshi</td>
<td>CIS*</td>
<td>Michael Kahnana</td>
<td>Psychology</td>
</tr>
<tr>
<td>Anthony Kroch</td>
<td>Linguistics</td>
<td>William Labov</td>
<td>Linguistics</td>
</tr>
<tr>
<td>Mitch Marcus</td>
<td>CIS*</td>
<td>Christine Massey</td>
<td>IRCS</td>
</tr>
<tr>
<td>Fernando Pereira</td>
<td>CIS*</td>
<td>Gillian Sankoff</td>
<td>Linguistics</td>
</tr>
<tr>
<td>Thomas Schoenemann</td>
<td>Anthro.</td>
<td>Robert Seyfarth</td>
<td>Psychology</td>
</tr>
<tr>
<td>Daniel Swingley</td>
<td>Psychology</td>
<td>Sharon Thompson-Schill</td>
<td>Psych.</td>
</tr>
<tr>
<td>Scott Weinstein</td>
<td>Philosophy</td>
<td>David White</td>
<td>Psychology</td>
</tr>
</tbody>
</table>

*Computer and Information Science **Center for Cognitive Neuroscience

C.b Vision, Goals, and Thematic Basis

Many scientific disciplines that have little regular contact nevertheless share critical theoretical questions about the nature and processes of information transmission: communication, broadly conceived. At Penn and elsewhere, these questions are probed in graduate programs and research laboratories as diverse as Anthropology, Biology, Computer Science and Engineering, Linguistics, Neuroscience, Philosophy and Psychology. We hope to create a community of communication scientists who are capable of integrating the theoretical issues, methods, and formalisms that are currently distributed over these multiple disciplines. The cross-disciplinary environment required for such integration is technically and culturally taxing, both for graduate and postgraduate students (who must commit to extra training and research activity, some of it in unfamiliar territory) and for their mentors (who must co-advice and develop new foundational courses).

The kind of interdisciplinary cooperation in training that we envisage for the IGERT program at Penn will not start from scratch. On the contrary, cross-disciplinary education has been a hallmark of our programs for decades, though limited by the absence of a formalized administrative structure and a shortage of dedicated funding. Our successes are perhaps best illustrated by the names and present positions of past graduates of our programs (found in the advisee lists of the various bio-sketches) – some of whom have taken posts outside the discipline of their Ph.D. IGERT funding will enable these partial cross-training successes to grow and flourish. The outcome will be a cohort of young scientists who can transcend the barriers that are currently slowing scientific progress.

C.b.1. Thematic basis of the program: Past progress in several of the communication-related fields has come from building static structural theories within a single domain; for example, synchronic syntactic systems characterizing the knowledge of idealized adult speaker/hearers of English. Today many of these disciplines are moving ahead in two convergent directions. (1) Communication dynamics: New theories aim to understand not only the communicative product but the communicative process – how information-bearing signals are perceived, stored and accessed by communicating entities, and how the relevant state of these entities develops and changes over time periods ranging from milliseconds to centuries [e.g., 1, 2, 3]. (2) Context-sensitivity in communicative systems: New theories aim to understand how communicating systems respond to the various and changing environments in which messages are received and processed; for instance, how grammatical features, background knowledge, and the structure of discourse combine incrementally to allow the binding of language to the ambient reference world [e.g., 4, 15].

C.b.2. Present shortcomings in our interdisciplinary training: It seems clear that interdisciplinary training will always face difficulties in execution primarily because of the added training and research efforts involved. However, at Penn we have identified several shortcomings in our current interdisciplinary training that we believe could be rectified with IGERT funding: (1) Limited cross-disciplinary student interaction: Via outside-speaker programs and occasional ad hoc seminars, students with broad interests can keep a finger on the scientific world outside their immediate discipline. However, students in these sub-disciplines rarely
gather to communicate their own research interests, plans, methods and findings to each other. This can result in insular scientists, who rarely see broader issues and are unskilled in communicating outside their own immediate group. When smaller, regularly meeting, cross-disciplinary forums have been implemented, our experience is that student research becomes more textured, interesting, and relevant. (2) **Lack of consistent mathematic and computational training:** Only pockets of student researchers currently obtain professional-level mathematical and computational training. All students in the communication sciences need this proficiency, both as a tool of their trade and to provide a common scientific language among these researchers. (3) **Securing funding for secondary training:** Breadth that comes at the expense of advanced disciplinary training is not useful. Yet most of our participating departments can promise funding for only four years, making it difficult to get the extra cross-disciplinary training necessary without risking a year without funding. This extra training is especially crucial in the communication sciences because the methods and tools, as well as the established data sources vary so widely (including ethological field observation, ethnographic participant observation, clinical or sociolinguistic interviews, psychological tests, corpora of transcribed conversations, controlled behavioral experiments or physiological recordings, and computer simulations). (4) **Lack of gender, race and ethnicity balance:** The Cognitive Sciences are indeed making some progress toward gender balance, with the exception of Computer Science/Engineering whose student population is predominantly male. Moreover, in all fields, progress in becoming racially and ethnically representative is still far too slow. Relatedly, undergraduate students who are enrolled in smaller colleges and non-research universities (which often have higher minority representation) can go unexposed to the training and experience necessary to make them prime candidates for a scientific career in the communication sciences.

**C.b.3 Addressing these shortcomings:** Our Education and Training objectives (Section C.d) are designed to target these shortcomings. Students in this program will interact with each other on a regular basis, be taught the same mathematical foundations, and be given the time necessary to conduct research in a second related discipline. Attention and resources from IGERT will be devoted to developing a diverse applicant pool. A key tool for this is the Undergraduate Summer Workshop, an intensive 3-week program including in-depth lectures from a range of faculty, and hands-on laboratory experience. This workshop will target strong undergraduates from colleges outside the standard research university pipeline, helping to identify and motivate a talented and diverse group of students. Another key method for recruiting a more diverse population of graduate students, already in use to some extent, is to join with scientists at other local colleges, whose gender, ethnic and racial constitution is very broad, in joint research and advising of undergraduates.

To illustrate the impact of our proposed training more concretely, we sketch four hypothetical graduate student research projects. These examples are based loosely on recent research here at Penn, where partial solutions were found by much more ad hoc methods.

1. A **biology student** has developed an annotated corpus of baboon vocalizations, with special interest in a sound produced by adult males, a loud two-phased bark known as a “wahoo.” This call type appears during predator encounters, during intergroup encounters, and during aggressive interactions with other males. Its sound varies in terms of the relative length and amplitude of the two parts, the sharpness of the onset (“wahoo” vs. “bwahoo”), the pitch contour, and in other ways. The student’s questions include whether this is really a single type of call, and what relationship its acoustic variation has to its different contexts of use. To address these questions, she needs to design and apply some signal processing techniques of a kind that are familiar to speech engineers. Having taken our mathematical foundations course, and having worked in a phonetics lab as part of her secondary training in linguistics, she is easily able to try out various signal analysis and synthesis methods and to go over her results with an expert in acoustic pattern recognition. The creation of a richer dataset allows her to discover that the more subtle dynamics of the baboon’s call are indeed contingent on the contextual setting of the call. Presentation of her work at an animal communication conference sets her apart from others because she has not only noted the contextual contingencies of calls but quantified the contributing components of the signal.

2. A **computer science student** working on virtual human behavior within graphical user interfaces wants to simulate natural eye movement patterns in an avatar (a virtual human) engaged in real-time conversation. In particular, she needs to uncover the information content of eye movements in conversation, be it for expressing emotions or for reference to objects in the world. She would like to design studies to collect such data, but computer science students normally lack knowledge of the
psychological literature and experience in controlled research with human subjects. However, as a result of
the IGERT program she has taken graduate coursework in psychology, including a recently developed
graduate level course in research methods. Her psychology training included an apprenticeship in the
psycholinguistic eye movement lab at the Institute for Research in Cognitive Science (IRCS), where she
helped run a study of eye movements in face-to-face conversation. She is easily able to design and collect
a conversational eye movement database, for deriving improved avatar eye-movement rules. Her work
also benefits the psycholinguists, who having heard about her work in a research seminar, start using
avatars as controllable stimulus displays in their own research.

3. A linguistics student wants to apply mathematical models to language change (i.e., how a language
evolves over generations within a community). Using available annotated historical corpora of English
created at Penn and elsewhere, she discovers some linguistic changes that never stabilize or become
permanent. She wants to know whether successful innovations differ systematically from cases where the
change is suppressed in terms of factors such as the complexity of the change or even the social status of
the innovating group. Having learned the necessary mathematical techniques as an IGERT trainee, she is
able to model the spread and suppression of change as a dynamical system and to make predictions about
the detailed quantitative profiles of successful and unsuccessful innovations. Having secondary training in
developmental psychology, this student also realizes that similar mathematical models can be applied to
language acquisition by children, which might be modeled as an innovative process of language creation
interacting with a regulatory process that suppresses deviation from social norms.

4. A psychology student has been studying child-adult differences in sentence processing abilities by
recording subjects’ eye movements as they respond to spoken instructions. When he discovers that the
youngest children have specific difficulties in recovering from initial misinterpretations of ambiguous
phrases, he recalls something from his secondary training in neuroscience: the protracted development of
prefrontal cortex in children is associated with delayed onset of inhibitory control functions. In order to
test the hypothesis that children’s failure to use context to recover from misinterpretation could result from
reduced inhibitory control, he identifies several patients with brain damage to prefrontal cortex from the
Focal Lesion Database maintained in the Center for Cognitive Neuroscience (CCN). He is able to compare
the eye-gaze performance of these patients to that of both normal adults and children in order to test his
hypothesis about the relation between general cognitive control mechanisms and specific linguistic
abilities. Further, he is able to simulate the effects of weakened inhibition of conflicting cues on sentence
comprehension using a model of cognitive control developed by one of the computational neuroscientists
in the CCN.

C.c. Major Research Efforts
This grant is designed to promote collaboration among five existing research clusters at Penn. Here we briefly
describe these clusters, their faculty and their unique properties.

1. Neurobiological and Field Study of Animal Communication (Faculty: Cheney, Schmidt,
Schoenemann, Seyfarth & White). Researchers in this area are committed to an ethological approach that
studies animal communication in its natural social context [5]. Notable examples include Cheney’s and
Seyfarth’s current research on the communication and social behavior of wild baboons, in which field
vocalization playback techniques are used to document baboons’ response to vocalizations in a range of
environmental settings [6]. Birds’ complex communicative and social behaviors are studied by both Schmidt
and White, again crucially in interactive settings. For example, White has developed a novel technique of
using remote-controlled robotic cowbirds to produce calls and subtle social behaviors as stimuli to aviary
birds, and Schmidt now records brain responses from freely behaving birds as they listen to others’ songs [7,8].

2. Behavioral Study of Human Language Use and Acquisition (Faculty: Dahan, Gleitman, Massey,
Swingley & Trueswell). This research area is concerned with the dynamics of human communication in
natural settings and how these abilities are learned and used by children to acquire their native language [9,10].
Trueswell, known for his work on real-time spoken language comprehension, and Gleitman, known for her
research on child language acquisition, have recently teamed up to study the development of child sentence
comprehension abilities. This group was the first to develop eyetracking during listening techniques for
children, in which they hear spoken instructions to act on the world [2]. Children’s eyegaze patterns allow researchers to infer the dynamics of language interpretation in real time, as each spoken utterance unfolds.

3. The Neural Basis of Human Communication (Faculty: Botvinick, Chatterjee, Coslett, Embick, Farah, Grossman, Jha, Kahana, Thompson-Schill). A large cognitive neuroscience community at Penn is interested in human communication and its relation to other cognitive functions including learning and memory, conflict resolution, and cognitive control. A strength of this group of investigators is their reliance on convergent evidence from a large number of methodologies, including (a) analysis of cognitive performance in individuals either with chronic impairments in patients with permanent brain damage (Chatterjee, Coslett, Grossman, Thompson-Schill) or with temporary impairments in normal subjects resulting from pharmacological manipulations (Farah) or transcranial magnetic stimulation (Coslett); (b) recording of neural activity in normal volunteers and brain-damaged patients using fMRI (Botvinick, Chatterjee, Coslett, Embick, Farah, Grossman, Jha, Thompson-Schill), ERP (Jha, Kahana), and scalp and depth electrode recording (in presurgical patients; Kahana); and (c) computational simulations of normal and impaired cognition with biologically-informed models (Botvinick, Farah, Kahana). [See, e.g., 11, 12, 13]

4. Computational, Mathematical and Formal Analysis of Language Structure, Use & Change (Faculty: Clark, Embick, Joshi, Kroch, Liberman, Marcus, Palmer, Pereira, Prince, Sankoff, & Weinstein). This research can be divided into two areas. The first, corresponding to the traditional discipline of linguistics, investigates how the structures of human language map sounds to meanings, and the how these mappings differ with communicative situation and historical, geographical, and social setting [3, 14,15]. The second area, corresponding to computational linguistics, develops representations and algorithms for computational language analysis, interpretation, and use, and for automatically learning representations and processing methods from appropriate data [16,17,18].

5. Enhanced Communication Environments and Systems (Faculty: Badler, Joshi, Kearns, Palmer & Pereira). Researchers in this area work on technologies that promote new ways of communicating. Kearns and colleagues build and study speech-based systems that are adaptive, and provide multi-modal, multi-user social interaction [19]. Badler and others work on the design and application of graphical human modeling and simulation, especially for human-computer interfaces [20,21]. These groups are especially interested in how contextual factors change and dynamically influence dialogue, including the generation of machine utterances and the interpretation of utterances produced by humans.

Collaboration among these groups will be mediated by co-advising students (as in the examples in Section C.b.3) and by developing common mathematical and computational tools. For example, probabilistic sequence models such as hidden Markov models (HMMs) are broadly useful for describing and modeling the dynamics of communicative processes with a latent state, such as animal vocalizations, turn-taking in dialog, or even impulses along a neural pathway; graph-generating processes and multi-player games on graphs are powerful tools for modeling the growth and information-exchange properties of networks of communicating agents; and computationally-tractable lexicalized grammars such as tree-adjoining grammar allow researchers to model the interplay between lexical and grammatical evidence in language acquisition. Each example involves a mathematical foundation, algorithms for fitting and applying models, and practical software tools.

Section C.d. Education and Training

We want to preserve the strengths of our existing interdisciplinary graduate training in the cognitive sciences, while also remedying the weaknesses that hinder development of interdisciplinary research in the dynamics of communication. Our proposed training program will provide for its students:

(a) an intellectual community across their diverse backgrounds and core training;
(b) preservation of the excellent training that they now get within their sub-disciplines;
(c) in-depth training within a secondary discipline for each student, without financial hardship;
(d) a shared mathematical and computational foundation;
(e) recruiting and retention for an ethnically diverse set of students across sub-disciplines.

We briefly describe below the mechanisms that will achieve these goals. These mechanisms are ordered chronologically, reflecting the order in which student participants will encounter them.
<table>
<thead>
<tr>
<th>Training Mechanism</th>
<th>Purposes</th>
<th>Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate Summer Workshop</td>
<td>Recruit students to the discipline; Increase minority involvement</td>
<td>Junior or senior year of undergraduate career</td>
</tr>
<tr>
<td>Common-Ground and Common-Thread Seminar</td>
<td>Provide weekly intellectual gathering; present/discuss research</td>
<td>Every year of graduate career</td>
</tr>
<tr>
<td>Training in Core Discipline</td>
<td>Provide in-depth expertise</td>
<td>1st and 2nd years of graduate career</td>
</tr>
<tr>
<td>Mathematical Foundations Course</td>
<td>Provide training in communication theory and its mathematical basis</td>
<td>Completed by the end of 3rd year of graduate career</td>
</tr>
<tr>
<td>Graduate Training in Secondary Discipline</td>
<td>Provide training secondary area, without financial hardship</td>
<td>3rd year of graduate career</td>
</tr>
<tr>
<td>Ph.D. Dissertation Research</td>
<td>Encourage cross-disciplinary theses</td>
<td>4th and/or 5th years of graduate school</td>
</tr>
<tr>
<td>Post-Doctoral Fellowship</td>
<td>Bring in outside experts</td>
<td>Post Ph.D.</td>
</tr>
</tbody>
</table>

**C.d.1 Advanced Undergraduate Workshop in Cognitive Science and Cognitive Neuroscience.**

For graduate recruitment into this program, funding is requested for an Annual Undergraduate Summer Workshop in Cognitive Science and Cognitive Neuroscience, co-organized with the Center for Cognitive Neuroscience (CCN) and the Institute for Research in Cognitive Science (IRCS). It will provide a three-week-long intensive introduction to research topics in this area. All IGERT faculty have agreed to participate, offering a day-long session of lectures, seminars and hands-on labs within their area of expertise (lab methods include fMRI, head-mounted eyetracking, computational linguistic annotation, ERP brain-imaging techniques, animal behavior, and neuroanatomy). Each year, two research themes will be selected as foci, with one directly related to the thrust of this IGERT training proposal, and the other designed to offer breadth in training within a related area of strength at Penn.

The PI and Co-PI’s have extensive experience running similar workshops. In 1998-2003, NSF’s Science and Technology Center (STC) grant to IRCS funded such a workshop, which was wildly popular: about 100-150 students applied every year, 25-30 were offered admission, and typically 100% accepted. Many students from this workshop were recruited into Penn’s graduate programs in the Cognitive Sciences, and several are currently star students or have moved on to select post-doctoral and faculty positions. However, the STC funding ended in 2002, and after finding extra funding in 2003, Penn is not be able to offer a workshop in 2004, for the first time in six years.

**New Minority Recruitment.** The central shortcoming of our prior summer workshops was the lack of an explicit recruitment mechanism for under-represented minorities. Under IGERT, we intend to enlarge the under-represented minority component of our applicant pool by implementing the following recruitment mechanisms. First, **IGERT Summer Workshop Minority Fellowships** will be offered to a target of 10 out of 30 participants, paying all travel, housing and dining costs associated with the workshop. We will work with Penn’s Office of Minority Affairs to achieve this goal. Second, electronic and printed announcements will be sent to appropriate colleges and universities around the U.S., directed to undergraduate chairs in relevant departments and to instructors of relevant undergraduate courses. Finally, advertisements and the workshop website will encourage minority student applications and provide details regarding the minority fellowships. The goal of this program is to enlarge the minority applicant pool to graduate school not just at Penn but nationally as well, since our participants also go on to graduate programs elsewhere.

**C.d.2. Graduate School Admission.** IGERT graduate trainees will be selected from the participating core graduate programs (Anthropology, Biology, Computer and Information Sciences, Linguistics, Psychology and Neuroscience). Each year, five incoming graduate students will be offered IGERT traineeships. These trainees will be guaranteed 5 years of graduate funding (instead of the usual 4 years) achieved by using IGERT funds to pay for stipends in 2 of their 5 years. This would supplement departmental support and crucially allow each student five years to fulfill the additional IGERT training requirements (detailed below) without financial hardship. We therefore propose funding 10 student stipends per year (for 3rd and 4th year funding), yielding approximately 25 IGERT students in the program at any given time. As part of our effort to attract minority students to our program, we intend to work with well-established on-campus units (such as the McNair Scholars Program) whose task it is to attract and mentor minority...
graduate students. Continued mentoring and group meetings have been found to be important components in the success of minority students in both undergraduate and graduate training programs.

**C.d.3 Graduate training in core discipline.** IGERT trainees will be expected to satisfy all the requirements of their home department. We recognize that in-depth expertise is needed within any science, and we do not want to water-down training for the sake of interdisciplinary breadth. Most participating departments have PhD course requirements that are typically completed after the first two years.

**C.d.4 Common-Ground and Common-Thread Graduate Research Seminar** All IGERT trainees, regardless of their year, will be expected to participate throughout their career in a weekly research seminar designed to provide an intellectual gathering for those students interested in the dynamics of communication in context. Sessions will alternate biweekly between what we call From-Within and From-Without meetings. During a From-Within meeting, one student will be expected to present his/her research to the group. This has two benefits: (1) Student presenters, working with their advisor, will have to explain the broader significance of their research to non-experts in related disciplines; (2) All students will be exposed to developing research projects from related disciplines, potentially shaping these projects as they develop. During From-Without weeks, students will present and discuss selected readings by experts from outside Penn, in particular, that week’s colloquium speaker at the Institute for Research in Cognitive Science. This ‘journal club’ will expose students to outside opinions, and prepare them for the colloquium in a way that isn’t normally possible. We believe this seminar is crucial because it establishes communication across different areas of expertise (establishing common-ground) and does so throughout all stages of educational development (establishing the common-thread). IGERT faculty will be encouraged to attend. One faculty per year will be the ‘ring-leader’, obtaining graduate teaching credit for this responsibility.

**C.d.5 Mathematical Foundations.** The new Mathematical Foundations courses will form a two-semester sequence, based on practical computer exercises dealing with real problems in the associated disciplines. Initially Professor Mark Liberman, an expert in computational linguistics and digital signal processing, will teach this course. The course will be taught in a computer/media lab setting and cover relevant aspects of a wide range of topics, including information theory, game theory, formal language theory, automata theory, the logic of information flow, signal processing, machine learning, and probabilistic models. Despite the great range of material and of student backgrounds, experience in teaching a subset of this material to a similar range of students convinces us that all students can get a basic practical understanding, while even the best prepared students broaden and deepen their knowledge. This common mathematical foundation will enable students to perform sophisticated analyses of signals and to model the form and information flow of behavioral sequences, whether the object of study is a human conversation, a negotiation with a computer to book an airplane flight, or a baboon calling bout. These two semesters obviously cannot substitute entirely for the dozen or more semesters that today would be required to cover a similar range of topics. However, they can give students the ability to understand and implement algorithms from published descriptions, especially given appropriate libraries of basic functions, and to discuss alternative approaches with experts in an informed way.

**C.d.6 Graduate Training in Secondary Discipline.** During their third year of graduate school, IGERT fellows will be expected to take focused coursework in their secondary discipline, designed to support plans for their Ph.D. research. For instance, a student studying machine learning in Computer Science would be able to take a series of human and animal learning courses offered through Psychology. A psychologist planning to do her Ph.D. research on the experimental study of language learning would take linguistics courses in Syntax, Semantics and Discourse Reference. A computer scientist might become trained in cognitive neuroscientific techniques. Five years of funding will allow for this secondary training.

**C.d.7 Ph.D. Dissertation Research.** Students will be well equipped to conduct interdisciplinary research on the dynamics of communication in context. They will be funded through their fifth year, and small research funds ($1000-$2000) will be made available to students on a case-by-case basis when other funds are not available to an IGERT dissertating student.

**C.d.8 Post-Doctoral Fellowships.** Funds are requested for two half-time post-doctoral fellowships per year as part of this IGERT Training program (matching funds made available by participating faculty). Applicants will be expected to submit a three-page research proposal reflecting a collaboration between at least two IGERT faculty (crossing the core research areas). Preference will be given to applicants who bring skills and expertise not represented in the core IGERT faculty (e.g., a geneticist interested in language evolution), thereby broadening the access that IGERT graduate trainees will have to intellectual and technical expertise.
These post-doctoral fellows will be expected to carry out their proposed research and participate in the *Common-Ground Seminar* described above (presenting their work and providing advise to others).

**C.e. Management and Evaluation**

The PI together with the co-PI’s will be responsible for all aspects of the program and will make decisions about the resource allocations, overall policy, monitoring of the budget, and setting up of special committees for the development of the educational programs. We expect and will encourage active involvement by other faculty participants and will set up an advisory committee including at least one member from each subdiscipline, so as to set policy for the overall program and to review student cases.

Infrastructure and administrative resources will be provided by IRCS (Institute for Research in Cognitive Science), funded 1992-2002 by an NSF STC grant. IRCS continues to exist after the expiration of the NSF STC grant. Liberman and Kearns co-direct IRCS, whose staff include an Administrative Director and several administrative and technical staff. The Administrative Director will be responsible for the detailed monitoring of the budget, and in the preparation of relevant reports. Staff will provide administrative support for the interdisciplinary seminars and the management of the IGERT educational programs described in this proposal.

IRCS has an external advisory board which meets once a year. We will ask a subset of this committee to look specifically at the IGERT program. In addition, review of IGERT will form part of the regular internal review of IRCS, during which the opinions of participating departments and students are actively solicited.

**Minority Evaluation:** Finally, we intend to have a representative from the Penn McNair Scholars program join our advisory board each year to evaluate our progress in minority recruitment and mentoring.

**C.f Expected Resource Commitments**

There are two relevant multi-site projects that are partly based at Penn, whose activities will mesh helpfully with the proposed IGERT program: the Linguistic Data Consortium (LDC) and the TalkBank Project.

The LDC (http://www.ldc.upenn.edu) is an open consortium of universities, companies and government research laboratories. It creates, collects and publishes speech and text databases and other resources for research and development purposes. Since its foundation in 1992, the LDC has published nearly 200 digital databases, which have been delivered to researchers at nearly 1,000 institutions. Many LDC databases involve human/human or human/computer interaction, sometimes with linguistic or functional annotation as well as audio or video signals and transcripts. The LDC will provide funding to create and publish additional materials suitable for teaching and student research on communicative interaction. The fact that Penn is the LDC’s host institution, and its director is one of this proposal’s co-PIs, will facilitate this cooperation.

TalkBank (http://www.talkbank.org) is an interdisciplinary project funded by NSF, hosted by CMU and Penn. Two of TalkBank’s coPIs are among the organizers of this proposal. The goal of TalkBank is to foster fundamental research in the study of human and animal communication, by providing standards and tools for creating, searching, and publishing primary materials via networked computers. So far, efforts have focused on five disciplinary groups: animal communication, classroom discourse, linguistic exploration, gesture and sign, and text and discourse, in addition to work on generally applicable standards, libraries and programs. Students in the proposed IGERT program will participate in TalkBank working groups, attend TalkBank workshops, and have the opportunity to work with TalkBank researchers from many other institutions. TalkBank will also serve as a medium to publicize the proposed IGERT-supported integrative approach to graduate training. In addition to these two organized inter-institutional entities, participating faculty have many active research ties with companies and government laboratories, where IGERT program students will be placed for summer or term-time research positions, as many students have been in the past. These include AT&T Research, ARL, BBN, Bell Laboratories, CHI Systems, CoGenTex, Face2Face Inc., General Electric Corp. Research, IBM, Intel, Lockheed-Martin (Camden, Moorestown and King of Prussia sites), Microsoft Research, MITRE, NASA Ames, NASA Johnson Space Center, NIST, Unigraphics, and Xerox PARC.

Finally, most IGERT faculty have externally supported laboratories on campus or in the field, providing an extraordinary array of opportunities for research by graduate and post-doctoral students. Note that we do not seek IGERT funds to set up these labs – they are already available, well-funded, and ready for students to join.