Programmatic Questions

PI: Phillips, Colin
Proposal ID: 1449815
Title: NRT-DESE: Flexibility in Language Processes and Technology: Human- and Global-Scale

1. Will the Science Policy internships be optional or required? Who will coordinate the internship program? How will expected outcomes and responsibilities for the trainee and the mentor be determined and codified? What metrics will be used to evaluate the internships? Will the internships count towards trainees' degree requirements?

Required? Yes. The part-time policy internship program will be required for the NSF-funded NRT trainees. It will be encouraged but not mandatory for students who pursue the program but without NSF support. The internships will not initially count towards trainees' degree requirements: this would be impractical, given the diverse range of formal PhD requirements in the programs of participating PhD students, and it could become a barrier to participation. It could in the future become an elective credit offering through the Language Science Center.

Coordinators. The internship program will be jointly coordinated by the PI and the program coordinator. The program will also draw on the expertise of NRT faculty with experience in the specific fields where the internships will be based, e.g., an internship in the American Speech & Hearing Association (ASHA) would be brokered by faculty who already have ties to ASHA.

Outcomes and responsibilities. The science policy internship program is a very new initiative. We are developing initial plans and procedures, but we expect that these will evolve as we learn from our initial experiences. In the course of developing our NRT proposal we contacted 6 professional and government organizations, all of which expressed interest in the program (we had the space to include only 2 of these in the proposal page limit). We discussed with each of these organizations possible activities that the internships could cover (e.g., developing briefing materials for policy makers, shadowing staff, learning how government uses research). As a next step, we will work with these organizations, with our faculty, and with our students to agree on initial guidelines. These will include such topics as: (i) duration of the internship [part-time for a semester or summer]; (ii) communications skills goals [oral and written]; (iii) goals for who the interns will interact with [preferably individuals inside and outside the host organization]; (iv) "match-making" and orientation procedures for each cohort of trainees [likely using Winter Storm as convenient time]; (v) expectations for concrete products [students should pursue a primary project with a specific objective]; (vi) reporting expectations for student and host; (vii) procedures for resolving disputes, e.g., if the host is dissatisfied with the student's work, or if the student feels that s/he is being given inappropriate tasks. The primary expected outcomes relate to our NRT's broader aim to improve students' communication skills, broader understanding of the context of their science, including how science impacts society, and appreciation of diverse career opportunities.

Metrics. The following metrics will be used to evaluate the internships:

1. Was the plan jointly developed by the student and mentor effective? (non-objective)
2. Were the planned activities completed?
3. What concrete products were produced, e.g., reports or briefing materials? What audiences saw these (internal or external audiences)?
4. In what ways did the student find the internship helpful? (non-objective)
5. In what way did the mentor find the internship helpful? (non-objective)
6. What communication skills were developed? (written and oral)
7. How did students' understanding of science-policy connections change based on the internship? (via before/after focus groups: we found this very informative in IGERT for tracking students' changing perspective on interdisciplinary research)
8. Do students participate more in public-facing activities following the internship? (could be compared with students in control group)

Some of the questions about the value of an internship could usefully be asked both immediately afterwards, and after a longer delay, at which point students' perspective may have changed.

2. Describe in more detail how trainees will receive explicit communications training as part of the Winter Storm and other activities. Describe the type of the communication training, e.g. oral and/or written presentations to technical and/or general audiences, social media, and video. How will competencies, rubrics, and benchmarks be developed? How regularly will trainees be assessed? Who will provide the training, and what is their background and experience in communications training?

Specific activities. Communications training activities will consist of activities spread throughout the year, but integrated with other program activities. It will include oral and written communication with diverse audiences. We feel that in recent years we have succeeded in developing a "culture of speaking", where oral presentation skills are highly valued, but for NRT we need to make the training more systematic and explicit, and we need to foster a similar "culture of writing". Whereas our oral communication training often occurs in open settings, our writing training has remained confined to exchanges between individual students and faculty. We need to elevate the status of writing by making writing training less "private".

- Oral, interdisciplinary technical audience: students will give an annual presentation at the Language Science Lunch Talk series (Thursdays), they will receive peer feedback via written forms (pilot currently underway); specific sessions will be devoted to explicit discussion of best practices
- Oral, disciplinary technical audience: conferences and job talks provide excellent opportunities for extensive practice. We will broaden our culture of intensive talk preparation and feedback. We will seek to codify best practices in writing, with models.
- Oral, general audiences: in outreach activities students work in groups to develop presentations to children or community members. Before and after briefing sessions with faculty and peers provide training and feedback.

- Oral, general audience: students will develop short videos explaining their work (pilot version currently in preparation for Winter Storm 2015).
- Oral, general audience: students will participate in training in speaking to broad audiences, led by experts in theater and in marketing & communications.
- Oral, multiple technical audiences: students will develop skills in effective on-the-fly communication via debates and practice poster sessions (practice talks are commonplace, but we recently discovered that practice poster pitches are more difficult and more fun; this is an essential communication skill that is often overlooked).
- Oral & written, general audiences: the policy internships will provide various opportunities for communications training.
- Writing, technical audience: students' NRT proposal provides initial training in grant writing. These proposals will go through multiple drafts, including feedback from peers and from a faculty committee. The revision process should be particularly beneficial.
- Writing, diverse audiences: in the modern information environment it is essential for students to learn how to explain science in an engaging fashion online, via websites, blogs, and social media, with messages ranging from a few pages to just 140 characters. Students and student teams will be integrated into the online communications program of the Language Science Center, which provides good opportunities for writing in these genres, and for receiving expert feedback.
- Writing, technical audience: workshops for writing conference abstracts provide valuable practice in crafting a concise and compelling message.
- Writing, interdisciplinary technical audience: students will participate in explicit mentoring in writing effective articles (students ran a pilot version in 2013), and will participate in writing workshops that focus on peer feedback (with students and faculty contributing their in-progress writing for discussion).

Contributors. A number of different individuals will contribute to the communications training, reflecting the diverse needs. Some have formal experience in communications training, others have practical experience as excellent communicators.

- Leslie Felbain from UMD's School of Theater has decades of experience of working with actors, and has recently developed programs for training scientists to be more effective speakers. She has expressed enthusiasm for extending her existing program with STEM faculty to our NRT graduate students.
- Nicky Everette is Director of Communications for one of our participating colleges; she has worked closely with language science faculty on developing online materials and infrastructure. We hope to engage her help in our student training efforts.
- Rae Grad is UMD's director of federal relations, working mostly on Capitol Hill. She has offered to contribute to our student training activities, especially in the area of communicating with policy makers.
- Philip Resnik and Nan Ratner are language science faculty who have experience of working with media outlets, either through radio broadcasts or through highly successful AAAS activities.

- Norbert Hornstein and Hal Daumé (NRT co-PI) both run prominent scientific blogs in their respective fields, which requires very different writing skills than publishing papers.
- Colin Phillips has much experience of speaking about language science to diverse academic and non-academic audiences.
- The UMD group plays a key role in a current national effort to improve scientific outreach in language science, e.g., symposium in Portland, OR on Jan 9th 2015 involving Phillips, Lidz, and some of our graduate students. This effort will bring our students together with experts from other universities with expertise in communicating with diverse audiences.

Evaluation. Competencies and benchmarks will be developed via consultation with students, faculty, and our evaluation team, led by KerryAnn O'Meara (UMD) and Ann Austin (Mich St U). It is important to us that students play an active role in defining the objectives and measuring success, as this helps them to reflect more critically on their own skills.

Feedback and Assessment. Feedback on communications skills will be provided regularly, as a part of students' other training activities. For individual students, overall progress towards communications goals will be tracked each semester, as part of students' regular review process, which is built around an electronic dossier that students use to track their own progress and that forms the basis for discussion with mentor(s) and program management. For the program as a whole, progress towards achieving objectives in communications training will be integrated into the annual formative assessment cycle, which will culminate in the external advisory board meeting each spring.

3. A key feature of NRT is training of STEM graduate students for both research and research-related careers, within and outside academia. The proposal is heavily focused on training students for academic research careers. How can you broaden the professional skills training so that it is includes a focus on non-academic careers, as well as academic careers? How will students receive explicit training in transferable professional skills in addition to communication, e.g. leadership, project management, negotiation, and conflict resolution? Is the career series already in place or planned? Will professional skills training involve active learning activities? Who will provide the professional skills training?

Trainee outlook; student and program goals. Our trainees will be drawn from diverse fields, which offer rather different career options. This means that we need to be responsive to diverse needs. For most of our students who focus on human/cognitive/neuroscience work on language, their initial goal is to pursue an academic career, and they are less informed about non-academic careers. For most of our students who focus on computational/engineering approaches to language, their default goal is a non-academic career (industry positions are plentiful currently), and they may be uncertain about the feasibility of pursuing an academic career, e.g., they might have encountered the view that the cachet of their department or institution places an academic career out of reach for them. Therefore, we need to pursue multiple approaches to help students to identify and prepare appropriately for different career options.

Transferable skills (and one surprise). We want all students to know the diverse ways in which their PhD training can be put to use, and understand that they can make valuable and rewarding contributions outside academia. At the same time, we strongly believe that many of the skills that are highly valued for non-academic careers are also valuable in academic careers, though they are sometimes under-appreciated in academia. In non-academic careers it is important to have strong skills in oral and written communication for diverse audiences, leadership, teamwork, project management, mentoring, etc. These are generally overlooked and do not attract explicit training in PhD programs, but they are essential skills for success in the current academic climate, just as in non-academic careers. Therefore, we are committed to developing these skills in our PhD students, but we want students to understand that most of them are valuable for long-term success inside and outside academia.

Sometimes the differences between academic and non-academic skill sets are overemphasized because of the mismatch between (i) what it takes to get tenure/promotion in academia (narrow, sometimes discouraging teamwork), and (ii) what it takes to be more broadly successful in academia and to drive institutional and programmatic change (broader; impossible to do it alone). We emphasize to our students that they should be looking beyond the narrow questions of how they publish their next paper or get their first secure job, and should be thinking of what skills they need for life-long success. This is not always an easy sell, as students understandably are motivated by near-term pressures.

In discussing with industry colleagues the skills that they look for in the PhDs that they hire (*Google*, *Microsoft*, etc.), some things are unsurprising, e.g., they value teamwork experience and the ability to meet specific objectives. But one thing is perhaps more surprising. In industry positions it is highly valued that applicants show that they can do a “deep dive” into a topic, e.g., as evidenced by publishing multiple papers on a topic. These employers sometimes have the perception that academics are more scattered, because the ability to fill multiple roles is highly valued. This ability to focus intensively on one thing for a while is more traditionally associated with academia, but may be more needed in industry.

What is in place already? A number of individual pieces are already in existence, but a key goal for NRT is to pull together the various training needs in a more systematic fashion. For example, some skills are implicitly developed and are highly valued, but the training and the rationale are not made sufficiently clear.

- A number of Winter Storm modules have explored different aspects of professional skills development. E.g., in 2015 we will hold a session on diverse career paths. But these have not previously been part of a systematic plan.
- Professional skills development is tracked through the periodic student self-assessment process. This helps to highlight the importance of these skills. But this has not previously been linked to specific training activities.
- Students receive ample opportunities to take on leadership roles through Language Science Center activities and initiatives. But explicit training for these roles has been only sporadic.

- Some of our participating departments offer mini-courses or talk series that focus on professional development topics, though these often emphasize more "traditional" topics, such as CV preparation, interviewing, or writing journal reviews.
- Undergraduate mentoring opportunities are very common in some participating units, and much less common in others. Our cross-department efforts are already helping to bring this into fields where it did not happen previously.
- The Computer Science department has a "Future Faculty Fellows" program, which aims to support students in pursuing the option of an academic career. This is a response to the fact that many students in that field consider this as out-of-reach or less attractive. (It is modeled on this program in UMD's School of Engineering.)

Active learning. The professional skills training will certainly involve active learning, via multiple mechanisms. (i) A central aspect of students' training will be their ownership over program activities, including events, program assessment, outreach, mentoring, and planning team-based research. This provides training in leadership, project management, and negotiation. (ii) The science policy internships will provide opportunities for active development of professional skills, as well as for active experience in a non-academic career path. (iii) The communications training [see #2] offers many forms of active learning.

Contributors. The professional skills training will be provided by multiple individuals: faculty, guest speakers (including alumni), and internship mentors. In addition to our faculty, who have much experience in the different aspects of professional skills development, we can draw on a wealth of other expertise. E.g., the researchers at the Center for Advanced Study of Language (CASL) provide an unusual window into the relation between academic and government research. E.g., we have access to many experts from Washington-area organizations, including UMD graduates who have positions in such places as the State Department, the National Academies, the National Institutes of Health, and national professional organizations. E.g., one of our key advisors is David Baggett, an alumnus and successful software entrepreneur: he has participated in professional development activities that we have offered in the past.

4. How will the trainees receive vocational counseling and preparation for non-academic careers, including research-related (rather than just research-focused) careers? Who will provide that training and preparation?

Vocational counseling. Vocational counseling will be provided via professional development workshops, especially during Winter Storm, via discussions with multiple faculty mentors, via internships (NRT's science policy internships and the industry internships that are common for computational students), and via specific individual connections that we will facilitate for students, depending on their individual career goals. Students report that it is often difficult to initiate career-focused discussions with their faculty mentors. To address this, these discussions are incorporated into the progress reporting process that students will undertake each semester with their faculty mentor.

Contributors. The counseling and preparation will be guided by the same individuals who will contribute to the professional development activities in #3 above. The Language Science Center's status as a university-wide initiative with ties to university leadership, government, and industry means that our NRT students will have access to an unusually broad network of connections who can help them in career preparation.

5. Will the NRT trainees be required to mentor undergraduates in the PULSAR program? What is the nature of the mentorship? Is it focused on research, or it is broader than that? How long would the mentorship last? Will the NRT trainees receive any mentorship training? Will the mentorship count towards degree requirements? Who will manage the mentoring program?

Not required, not for credit. Undergraduate mentoring in PULSAR will not be a requirement, but it will be strongly encouraged, and there will be many opportunities for NRT trainees to mentor PULSAR students, in various settings.

Undergraduate mentoring will not count towards degree requirements in participating PhD programs. We think that would be counterproductive in the short term, and we do not see this as a barrier to participation: in our experience, graduate students find it rewarding to develop effective mentoring relations with undergraduates, and they have little difficulty recognizing that this is valuable preparation for their future career.

… but this does raise interesting questions for the future. The aim of NRT is to bring about sustainable and generalizable change in graduate training, via experimenting with new training models. We do not regard it as feasible or attractive to modify PhD requirements in as many as 10 participating PhD programs at the start of our NRT program, and we generally prefer to avoid turning best practices into degree requirements, as that feeds a "checkbox mentality" among PhD students that we discourage. Seeing beyond the checkboxes is an important transition for students as they move from undergraduate education to preparing themselves for leadership roles in science (this applies to academic and non-academic careers alike). We have had good success in creating a strong culture that helps students to adopt this perspective. But we acknowledge that this might not be scalable, and so it is important for us to consider whether some NRT best practices should ultimately be codified as credit-bearing activities or degree requirements.

Focus of the mentorship. The undergraduate mentorship activities will be broader than research, although that is certainly one component. Graduate students will help students in shaping their academic path, identifying interests, helping them to connect to faculty and students with similar interests. They can also contribute to developing specific skills, e.g., communication skills through outreach activities; leadership and teamwork through joint participation in student committees.

NRT trainees who apply and are selected as PULSAR fellows during their PhD careers will work closely with the PULSAR students as a group via the weekly PULSAR seminar and PULSAR-specific activities.

Our first two students completed this role in Fall 2014 and both found it to be a very rewarding experience.

Mentorship training. Yes. This will be a new activity for us: a number of training activities have been developed in individual departments, mostly in the context of labs where graduate students supervise large numbers of RAs, but we have not yet developed program-wide training activities. This will be incorporated into the rest of the professional development training. Additional training will be provided for students who apply for the PULSAR fellowship, covering activities specific to that role.

Contributors. The mentoring activities will be managed in a distributed fashion. The PULSAR fellowship (2 per semester) is coordinated by the PULSAR Director, currently Dr Tess Wood, Assistant Director of the Maryland Language Science Center. Other mentoring activities will be managed by the PIs, the program coordinator, and the trainees' faculty mentors.

6. Will the evaluation and surveys examine how the NRT trainees and the control group compare with respect to transferrable development skills, including communication skills? Will the information be self-reported? Will the control group include students from other universities?

Control group. Our aim is for the control group to include students in language science fields from other CIC Institutions (the academic arm of the Big 10 sports conference + U of Chicago). We view these as a promising control group because: (i) the universities are comparable in size and academic profile; (ii) many of the universities have language scientists in multiple disciplines, though without the degree of coordination across fields found at UMD; (iii) there is infrastructure in place to support cross-institution initiatives within the CIC, and it is possible to leverage this to create something that is more than an assessment control group (it would be hard to get buy-in for that alone); (iv) in May 2014 we hosted an initial workshop with representatives from all CIC institutions, in an effort to seed connections. The workshop was well received. We favor the external control group over an internal control group, as that would introduce a strong selection bias. The students at UMD who have easy access to the NRT activities but choose not to take advantage of them (for whatever reason) are not comparable to the NRT cohort.

Assessing professional skills. Yes, transferable professional skills are a key part of the training goals of the NRT program, so they will receive a lot of attention in our evaluation efforts. The head of our evaluation team, Prof KerryAnn O'Meara, is an authority on professional development for young researchers, and she also has extensive experience in professional development programs for early career researchers through her leadership of UMD's NSF-ADVANCE program for inclusive excellence. (A number of the faculty on our team have benefited from O'Meara's programs.) By a stroke of sheer luck, the ADVANCE program is housed on the same hallway as the PI's office, making effective coordination easier.

Information gathering. For the NRT and control groups, we will use a mix of self-reported data from students, plus survey reports from selected faculty. Since we cannot expect to directly observe the

students in the control group, there are limits on the data that we can collect. For the NRT group only, we will gather further information based on direct observation of student activities and research.

7. It is difficult to evaluate the plan to broaden participation because applicants to the NRT will be recruited from the current population of graduate students, and the institution's strategies to broaden participation are not described. The solicitation requires that proposers provide "quantitative data showing the recruitment and retention outcomes of participating departments over the past five years, including time-to-degree completion". Please provide those data.

Clarification. While it is true that students will formally apply to the NRT after they are already PhD students at UMD, this does not mean that NRT will not be involved in recruiting students to the university. The opportunities provided by our interdisciplinary program(s) and community feature prominently in our recruitment efforts, and these are a key reason why students choose to enroll at UMD. The part that is delayed is the development of a detailed plan of study by the individual student and formal commitments involving funding.

For our purposes, "underrepresented groups" includes ethnic minorities in all of the fields that we cover, and women specifically in computational fields. (Nationally, only around 20% of computer science PhDs are women, and the numbers at UMD are similar.)

Strategies that have not worked for us. In the past we have participated in visits to HBCUs and attended national events targeted at underrepresented groups. This strategy did not prove effective for us, though it might be more effective for other fields or institutions. The fields that we are recruiting in are often not represented at HBCUs. Students come from a wide range of institutions, and are often not coming direct from an undergraduate school. The ethnic minority students who have come to us in recent years all got their first degree from large public universities.

Strategies that have worked for us. Our successes in recruiting from underrepresented groups have come from: (i) the strong reputation of our faculty and programs: students' mentors advise them to go to very strong programs, unsurprisingly; (ii) the reputation of our highly supportive academic climate - prospective students encounter this via campus visits, via our online materials, via meeting with our students at conferences and summer schools, and via our faculty's many visits to other institutions; (iii) our various efforts as engaged scientists, i.e., existing outreach programs, proposed policy internships for NRT. For some students, these activities send a strong message about our values. For recruiting women computer scientists, the simple fact that we have strong women faculty and an unusually high proportion of women graduate students makes a big difference (most of Feldman and Daumé's students are women; this is quite unusual relative to their peers).

Aside from activities that directly contribute to recruitment to our PhD programs, we take seriously the need to expand the pipeline for underrepresented groups in language science and STEM more broadly. Fighting over the terribly small numbers that have already decided to pursue a PhD is not a satisfactory solution. That is another reason why we are committed to outreach activities that engage

with middle-school and high-school students, mostly in schools where almost all students are from groups underrepresented in STEM.

Institutional strategies. UMD's College of Behavioral and Social Sciences has a summer research program for students from underrepresented groups that seeks to attract students who may ultimately apply to one of our graduate programs. Some UMD colleges also offer funds to assist in the recruitment of students from underrepresented groups. The university as a whole has a diverse student population, and it is certainly a welcoming environment for students from diverse backgrounds. According to NSF data, UMD's rate of graduating minorities who go on to STEM PhDs is among the highest in the country.

Data from the IGERT cohort. Recruitment and retention data from our IGERT cohort is more informative than data from participating departments (also provided here - see appendix), because the IGERT students came to UMD for its language science opportunities and benefited from its supportive environment. Among the 50 PhD students who have pursued the IGERT program (the final cohort are students who began in 2012), we have the following statistics:

70% US students, 30% international students
50% received IGERT funding, 50% were funded through other sources
30% of NSF-funded students from minorities (20% AfAm + Hispanic)
60% of students were women, including majority of computational students
98% retention rate so far (includes 2 students who moved w/ PhD advisors)
92% 6-year completion rate

Relative to norms in the participating fields, these figures range from somewhat better to far better. The retention rates are substantially higher than overall rates at UMD. In terms of recruiting underrepresented minorities the sample size is so small that it is hard to draw confident quantitative conclusions. In terms of recruiting women to computational fields, we are more confident that the numbers are significantly different than in peer programs.

Data from participating programs. See appendix.

8. The proposal points out that mentoring of students from disadvantaged backgrounds is particularly critical for their success. Describe any planned mentoring approaches specific to students from underrepresented groups.

Who we are talking about: We prefer to reframe the question. "Disadvantaged backgrounds" is not the same as "underrepresented groups"; groups that are severely underrepresented in one field are overrepresented in other fields; and students' backgrounds are imperfect predictors of the risks and challenges that they face. We therefore do not ourselves plan any mentoring approaches that will be specific to one group of students. That could easily backfire, by adding to students' marginalization. However, we are certainly mindful of the steps that can be taken to improve students' chances of

success, and we are alert to the fact that some students are at greater risk than others, especially those from disadvantaged backgrounds.

Program-specific mentoring activities: We take many steps to support student success.

- In all cases is it important to build trust between students and faculty mentors, preferably with multiple faculty. Without trust, it is harder to become aware of risks and challenges.
- Addressing career concerns early and often is important for all students, but especially for students from disadvantaged backgrounds, who often lack the safety net that other students might enjoy, and who are more likely to face pressure from family to use their talents on something more lucrative more quickly.
- Fostering strong support networks is essential for student success and for combating risks. E.g., [redacted]

- Students benefit from role models who they can identify with. This is captured nicely in a recent blog post by co-PI Hal Daumé, one of our computer science faculty. Students from underrepresented groups may need more help in finding suitable role models.
- Faculty must be proactive in learning about risks and in addressing them. E.g., Naomi Feldman and Hal Daumé, computational experts from different departments, recently participated in a seminar on "Empowering underrepresented students to enhance retention". E.g., co-PI Rochelle Newman, chair of the Dept of Hearing & Speech Sciences, has been developing plans with another faculty member to introduce cultural competence training for current faculty and students. It makes a big difference when the leading scientists in a group take these issues seriously.
- The community outreach activities -- a requirement for all NRT faculty and students -- indirectly support our mentoring goals. Engagement with students from neighboring communities helps to foster greater sensitivity to the needs of students from diverse backgrounds.

University resources: Beyond the NRT team, the University of Maryland has many resources designed to support student success, including some targeted specifically at underrepresented groups.

- The University System of Maryland (UMD is the flagship campus) hosts the NSF-sponsored PROMISE AGEP program that offers multiple professional development opportunities for underrepresented STEM graduate students
- In 2014 the university launched the Maryland Center for Women in Computing, which offers numerous programs aimed at supporting women computer scientists at all levels, from middle school to the professoriate
- A graduate women mentoring series in computer science offers faculty and peer support.
- UMD sends a large contingent of women computer scientists each year to the Grace Hopper Celebration of Women in Computing (30 in 2014)

9. It is unclear if NRT trainees will include students in computational science as well as linguistics. Please clarify.

Short answer. Yes, computational research will be an integral part of our NRT program, and it will impact all NRT trainees.

Scope. Our NRT trainees will be drawn from around 10 PhD programs, covering research that spans from education to electrical engineering. Linguistics is just one among those participating programs. Faculty with computational expertise relating to language are drawn from Computer Science, the iSchool (= Information Science), Linguistics, Electrical Engineering, and the Center for Advanced Study of Language. Faculty with expertise in cognitive/psychological/neuroscience approaches to language are also drawn from many different departments/programs. The special role of the Linguistics Dept is that (i) it stands at the intersection of these two approaches, and (ii) a number of its faculty teach 'gateway' courses that serve students from diverse academic backgrounds.

Beyond 'Big Data'. An important goal of our NRT program is to move "beyond Big Data". Language scientists were early to the scene in the big data revolution, and language technologists quickly learned about the possibilities of applying statistical models to large amounts of 'found data'. The results of that have found their way into our everyday lives via Google and Apple. But these fields were also early to confront the limitations of this approach -- it's surprisingly limiting in English, despite working with vastly more data than human children learn from, and it is impractical for 99% of the world's languages. That is why our NRT program focuses on multi-scale data, in making smarter use of data sets (whether 'found' data or carefully curated data), and in learning from comparisons of human and machine learning.

(Context: The Big Data revolution in natural language processing led to a distancing of computational and human-focused research on language. One famous quote by a prominent researcher was: "Every time I fire a linguist, my system gets better!" That person revised his views many years later, and this opinion is becoming less widespread. But that climate led to a generation of young researchers who were skeptical of the value of working with experts in human language processing.)

Specific computational training. NRT trainees will engage in computational work through a number of mechanisms:

- Semester courses: Naomi Feldman's computational modeling course (LING 698 / CS828F) is specifically designed as a point-of-entry for students with limited computational background. Students with a little more experience can take the two-semester sequence in natural language processing (CMSC 723/724), machine learning (CMSC 726) or data analytics (INST 737). We hope to build on the advent of NRT to reorganize the natural language processing courses in ways that will better serve NRT trainees, covering themes such as "text as data" and deeper engagement with computational linguistics.

- Winter Storm modules: our annual 2-week intensive winter workshop provides many opportunities for computational training in smaller-than-course-sized units. We have included diverse computational modules in previous incarnations; future Winter Storms will include modules that are more directly tailored to the computational and data curation needs of NRT trainees.
- Team research projects, including Summer Camp: all NRT trainees will participate in team-based research projects, which will be developed over an extended period of time, and will include at least one intensive "Summer Camp" research workshop. Our aim is for all of these team projects to include a computational component, though the nature of this component will vary between projects.

Minimally, all NRT trainees from a non-computational background will upgrade their computational skills, learn how to work alongside computational experts, and gain a better understanding of computational research. If successful, this will already be a significant advance. But also, students with computational expertise will learn from experts in other fields how to make better use of data resources and data collection and curation skills, so that they can put their skills to better use.

Example: a collaboration between two current early-stage PhD students is a good model of what we would like to see. Sudha Rao (Computer Science) is collaborating with Allyson Ettinger (Linguistics) on computational research on automatic text processing in Chinese. A key challenge in automatic understanding of Chinese is that many words are omitted, and their meanings must be recovered from discourse context. The two students contribute skills from quite different fields to the project. Both are learning a lot about the other's area of expertise, both have been drawn outside their comfort zone, and this has already led to a paper submission. This is a model for partnerships that we would like to see in the NRT program.

Appendix: Recruitment and Completion Data for Participating Programs

We provided above information on recruiting, retention, and time-to-degree for students in our interdisciplinary language science programs. We regard those as the most informative baseline data. Data from large departments like Psychology and Computer Science are less informative about the language science programs if only a couple of faculty in each department are involved in language science. Nevertheless, we here provide further details on university and department level statistics for recent years. We do this for the 7 departments whose students have been most active in our interdisciplinary language science programs to date. We have spreadsheets with more detailed raw data that can be provided on request.

Federal categories changed around 2010, so it is difficult to combine data from before/after that time. We focus here on 2010-2014 data.

Abbreviations:

LING	Dept of Linguistics
SLLC	School of Languages, Literatures and Cultures (includes the 2nd Lang Acq. program)
HESP	Dept of Hearing & Speech Sciences
PSYC	Dept of Psychology
NACS	Program in Neuroscience and Cognitive Science (interdepartmental PhD program)
CS	Dept of Computer Science
HDQM	Dept of Human Development and Quantitative Methodology (College of Education) (this department is the result of a merger within the past 5 years)

<u>Overall UMD Profile</u>

At the undergraduate level, the University of Maryland has a fairly successful record of recruiting for diversity. The current undergraduate population includes 42% from minorities, including 15% African American students. According to institutional data (which we are unable to verify), no other US university that has such a high minority enrollment has an overall graduation rate of 70% or higher.

At the graduate level, the 2014 entering cohort is 19.6% minority, 31.2% foreign, and 43.3% US caucasian (6% unknown). The 19.6% includes 7.2% AA, 4% hispanic, 1.6% 2 or more. These numbers include PhD and Masters students. Note that at the graduate level, percentages are substantially affected by the choice of denominator. Minorities are by definition US students, but institutional data often reports percentages of all students, US and foreign alike.

Recruitment 2010 - 2014 (5 years)

Data is for 2014 alone, and for 2010-2014 combined. The data is for all enrolled doctoral students, and not only for new enrollments, so as to yield a larger sample. Summary: our departments are somewhat diverse, but not as diverse as the UMD Graduate School population statistics overall. That difference is hard to interpret, since there are so many differences across fields, including differences in the denominator, due to great variation in the proportion of foreign students.

Overall UMD Graduate School:			
	7.2% AA, 4% Hispanic, 1.6% 2 or more.	47.4% female.	2014 alone
	7.3% AA, 3.7% Hispanic, 1.6% 2 or more.	47.3% female.	2010-2014 combined
SLLC			
	1.1% AA, 5.4% Hispanic, 2.1% 2 or more.	73% female	2014 alone
	1.5% AA, 6.4% Hispanic, 1.3% 2 or more.	76% female	2010-2014 combined
HESP			
	2.3% AA, 3.4% Hispanic, 4.6% 2 or more.	88% female	2014 alone
	4.9% AA, 4.6% Hispanic, 2.7% 2 or more.	91% female	2010-2014 combined
LING			
	3.3% AA, 3.3% Hispanic.	50% female	2014 alone
	4.2% AA, 4.7% Hispanic.	49.1% female	2010-2014 combined
PSYC			
	8.8% AA, 4.7% Hispanic ,2.0% 2 or more.	76.4% female	2014 alone
	7.1% AA, 5.5% Hispanic, 1.2% 2 or more.	75.6% female	2010-2014 combined
CS			
	8.7% AA, 6.8% Hispanic, 4.1% 2 or more.	14.8% female	2014 alone
	9.1% AA, 5.5% Hispanic, 3.3% 2 or more.	13.2% female	2010-2014 combined
HDQM			
	6.3% AA, 6.3% Hispanic, 1.8% 2 or more.	81.1% female.	2014 alone
	6.5% AA, 6.9% Hispanic, 1.8% 2 or more.	81.6% female.	2010-2014 combined
NACS			
	2.2% AA, 0% Hispanic, 2.2% 2 or more.	39.1% female.	2014 alone
	4.7% AA, 0.4% Hispanic, 1.6% 2 or more.	49.1% female.	2010-2014 combined

Retention

Retention information is for students overall, and is not broken down by gender or ethnicity.

Retention and graduation rates are, of course, a lagging measure, e.g., 4-year retention rates are only available for students who enrolled in 2010 or earlier. So any changes implemented in recent years would not be reflected in graduation rates and would only impact short-term retention rates. We provide 4-year retention/graduation rates for the entering classes of 2008 , 2009, 2010 (3 years worth), and 2-year retention/graduation rates for entering classes of 2010, 2011, and 2012.

Summary: the overall numbers vary here. In the 2008-2010 data HESP and NACS are low, other language science fields either track university-wide norms, or are higher (LING and CS). In the more recent data our programs generally perform above university-wide averages. But since most students who drop out of PhD programs do so after more than 2 years, these numbers should be taken with caution. (Plus, these numbers do not distinguish students who drop out of a PhD program vs. students who follow their PhD advisors to another institution).

	2008	2009	2010		2010	2011	2012
Graduate School (all)	72.3	74.7	75.4		84.1	79.6	81.1
LING	100	100	80		80.0	83.3	80.0
SLLC	75.0	57.1	72.7		90.9	81.3	81.8
HESP	54.5	--	28.6		85.7	66.7	90.0
PSYC	83.3	75.0	76.9		76.9	90.9	100
NACS	58.3	54.5	85.7		100	100	90.0
CS	70.0	83.3	84.6		90.4	86.8	88.6
HDQM	64.7	80.0	86.7		93.9	90.9	100

* Note that the data for HESP combines PhDs and AuDs, a professional doctorate for audiologists. That program has seen a sharp change in its PhD population, driven in large part by IGERT and by the new faculty hires and the broader language science investments that were catalyzed by IGERT.

Time to Degree

Also not broken down by ethnicity. Note that this covers all doctoral degrees, not only PhDs. For HESP, we break out PhDs separately. Generally, LING has a faster time to degree than the other departments by roughly a year; SLLC has been generally improving (as has HESP, but the sample size is very small).

	FY 2010	FY 2011	FY 2012	FY2013	FY2014
LING	5.2	5	5	7.4	5.1
SLLC	7.9	8	6.2	7.5	6.7
HESP	N/A	6.9 (N/A for PhD only)	6.3 (5.5 for PhD only)	N/A	5.7 (5.7 for PhD only)
PSYC	6	6	6	5.3	6.5
NACS	6	4.7	4.9	5.7	5.7
CS	6.3	5.7	6	6	6
HDQM	6	7.3	6.5	6.3	6

Women in Computer Science

Whereas sample sizes for ethnic minorities are small enough that large fluctuations in percentages are likely, we looked in more detail at quantitative data from women in computer science, as there is a larger sample size. (Reminder: these data are for Computer Science as a whole, where women are severely underrepresented at the doctoral level. Our current enrollment of women in computational aspects of language is exceptionally strong.) It is clear that female applicant numbers are very low. Our question is how women fare once they have decided to apply and enroll.

Recruitment and enrollment. The Computer Science department receives around 1000 PhD applications per year, of which around 20% are from women. Admission rates for women are also very close to 20%. The percentage of admitted students who enroll is also around 20% for women. In the past couple of years there is a trend for a lower proportion of women to enroll than apply. This is worrying. The recent creation of the Center for Women in Computing may help to address this.

Time to degree. In our sample of 142 recent CS PhDs the time to degree is slightly shorter for women (5.9 years) than for men (6.25 years).