Report on the NSF-sponsored *Human Language Technology Workshop on Industrial Centers*

Mary Harper*, Alex Acero, Srinivas Bangalore, Jaime Carbonell, Jordan Cohen, Barbara Cuthill, Carol Espy-Wilson, Christiane Fellbaum, John Garofolo, Chin-Hui Lee, Jim Lester, Andrew McCallum, Nelson Morgan, Michael Picheney, Joe Picone, Lance Ramshaw, Jeff Reynar, Hadar Shemtov, and Clare Voss[1]

* University of Maryland
College Park, MD
mharper@casl.umd.edu

Abstract

This paper reports on the findings of the National Science Foundation (NSF) hosted Human Language Technology Workshop on Industrial Centers that was held May 3rd and 4th, 2007. Representatives from academia, industry, and government attended this meeting to discuss the feasibility of developing an NSF center-based partnership between industry and academia in the field of Human Language Technology (HLT). Currently the HLT field does not have such a center in the US. Given the considerable advances in this field with great potential for continued success and the benefits of collaborations among academic, industrial and government partners, the time is ripe to build a better understanding of how to create a center that is not only mutually beneficial to all parties, but also supports work that simply could not be done by any partner alone.

Purpose of the Meeting

On May 3rd and 4th, 2007, the National Science Foundation (NSF) in Arlington, Virginia hosted the *Human Language Technology Workshop on Industrial Centers*. Twenty-nine representatives from academia, industry, and government attended this workshop to discuss the feasibility of developing an NSF center-based partnership between industry and academia in the field of Human Language Technology (HLT).

Because the HLT field does not currently have an industry-oriented center in the US, the purpose of the workshop was to determine whether the time is ripe to begin plans for building such a center. Several factors justified convening the workshop:

- There have been considerable advances in the field, and there is great potential for continued advances in fundamental technologies ranging from speech recognition and synthesis to machine translation, text mining, and next-generation search engines.
- Planned coordination among academic, industrial, and government partners offers the potential to tackle research questions that are broader than the ones that could be addressed by any partner alone and whose solutions would be mutually beneficial.
- Such collaboration has the potential to stimulate research excellence at universities, to enhance the quality of the intellectual property of US HLT companies, and to foster university-to-industry technology transition.

Preparatory Materials for the Meeting

In preparation for the meeting, participants were asked to read the following materials related to two types of NSF centers and to focus especially on the linkages for university and industry collaboration in each.

1. **The NSF Industry/University Cooperative Research Centers (IUCRCs) program:**

   The IUCRC program seeks to develop partnerships among industry, university, and government members to stimulate cooperation for carrying out fundamental research recommended by an Industrial Advisory Board.

   - The IUCRC program web site (NSF, 2007d)
   - The IUCRC Program Evaluation Project (Gray, 2007)
   - "Managing the Industry/University Cooperative Research Center: A Guide for Directors and Other Stakeholders" (Gray and Walters, 1998), in particular, chapters 1, 2, and 5

2. **The NSF-sponsored Engineering Research Center program:**

   The ERC program seeks to develop engineering systems-focused, interdisciplinary centers at universities in close partnership with industry.

   - The ERC program web site (NSF, 2007b)
   - The Engineering Research Centers Association web site (ERC Assoc., 2007)
   - "ERC Best Practices Manual" was developed by staff of the ERCs to assist those who are planning or setting up an ERC (Absher et al., 1998). See chapter 5 on industrial relations.

Participants were also requested to consider the following issues prior to the meeting:

- Is a center a viable vehicle for collaboration between academia and industry in the area of HLT? If so, what type of center would be best?
- How can one optimize a mutually beneficial center-based partnership among academia, industry, and government with respect to the following tasks?

[1]For affiliations of authors, see Harper et al. (2007).

- Develop a long-term, strategic vision for an emerging engineered HLT system with the potential to transform a current industry or spawn something new.
- Define a research agenda that optimizes shared research interests, needs, and opportunities.
- Define partnership strategies between universities and industry and determine how to best collaborate and divide up rights and responsibilities.
- Determine strategies for protecting/sharing intellectual property while enabling timely publication of intellectual output of the center.
- Develop mechanisms for involving graduate students in industrially relevant research that also qualifies for Master's and Ph.D. level theses.

- What breadth of research should the center fund? Which areas of research are most viable for center collaboration?
- How should the center handle organizational issues?
  - Develop a strategic plan for integrating fundamental HLT-related science and engineering research. Is there a viable test bed that could be used to tie together the research threads and enable systems level evaluation?
  - Develop a strategic plan for constructing a multidisciplinary research agenda while developing a more diverse research population. Would a single site or multiple site center be more effective?
  - What is the best structure for an advisory board (i.e., balance between academic, industrial, and government oversight)?

The Meeting

The meeting was comprised of a series of presentations and breakout sessions. On the first day, there was an opening presentation by Mary Harper about the meeting's purpose and schedule, followed by four presentations on NSF center programs, two by NSF program directors: Alex Schwartzkopf (IUCRC program) and Bruce Kramer (ERC program) and two by individual center directors: Janis Terpenny (Virginia Tech IUCRC) and Adam Powell (USC ERC). These presentations were followed by two breakout sessions in the afternoon. Homework was assigned on the evening of the first day of the workshop and was discussed first thing in the morning on the second day. Discussion of the homework was followed by a third breakout session on possible next steps. In the following subsections, some of the key issues raised by the three to four focus groups in each breakout session are summarized.

Discussion Item 1

Would an HLT center be a viable vehicle for collaboration between industry and academia? What would the ideal collaboration look like?

An environment for working on large-scale problems

As centers have a fairly high management and infrastructure overhead, the participants considered what the advantages of a university-industry center would be compared to individual collaborations between one university laboratory and a single industrial partner. Some participants pointed out that an individual expert may be better suited to work on immediate well-defined problems, but a group with a diverse expertise would be needed to work on larger, less well-defined problems. A center could provide just the right environment to attract high quality students and faculty and engage industry involvement to tackle bigger problems than an individual or small group could handle. It could investigate broader efforts with multiple disciplines, while educating graduate students to work in the new emerging areas of science and technology. A center would also provide industry with more revolutionary science and engineering, produce better students for industrial partners to recruit, and produce more products and services than an individual laboratory.

Availability of shared resources

Another advantage of a center is the availability of shared infrastructure, including various types of data, tools, and computational support (e.g., the MapReduce algorithm implemented over a grid-like computational substrate to support very large-scale computation). Large data collections are essential in the light of the data-driven methodology common in HLT, but they are often quite expensive to create, extend, document, maintain, and distribute. Some data collections require human subjects' approval, while others may require the center to deal with copyrights. In addition to coordinating the development of and providing access to the right data to set the challenges for the center, it is also necessary for the center to provide shared computing environments. Members should be able to work on parts of an end-to-end system without needing to build an entire system by themselves.

Alternative models for collaborative efforts

One of the breakout groups discussed other types of models for centers or collaborative efforts that support broad multidisciplinary research in addition to IUCRCs and ERCs. These models include *Centers of Excellence* (CoE), e.g., the Johns Hopkins University CoE; *Federally funded research and development Centers* (FFRDCs), e.g., Institute for Defense Analyses (IDA), MIT Lincoln Labs, and MITRE; *University-affiliated Research Centers* (UARCs), e.g., University of Maryland Center for the Advanced Study of Language (CASL), Johns Hopkins University Applied Physics Laboratory (APL), University of Southern California Institute for Creative Technologies (ICT); *Patron-based funding* (such as Bambergers), e.g., Institute for Advanced Studies (IAS) at Princeton; University Centers, e.g., International Computer Science Institute (ICSI) at Berkeley; *DOE National Laboratories and Technology Centers*, e.g., Argonne National Laboratory, Ames Laboratory; *The MOSIS*

Service (in VLSI); *Supercomputing Centers*; *NSF Science of Learning Centers* (SLCs); *Technology Alliances* (CTAs, ITAs), e.g., Collaborative and International Technology Alliances at the Army Research Lab (ARL).

Broadening partner involvement & research portfolio

These models involve different types of partnerships between industry, university, and government (see Figure 1). They vary in the extent to which partners are involved in the initial establishment of the collaborations, in the planning of projects, the reviewing and selection of projects, the funding decisions, and the legal commitments that come with project funding (grants vs. cooperative agreements vs. contracts). For example, ARL currently manages several CTAs and ITAs, each with ***joint*** planning and cooperative agreements among industry, university, and government partners. Individual CTAs and ITAs are funded once for five years, with three-year add-on options. By contrast, UARCs and University CoEs have cycles of multi-year government funding, because they are intended to address their government stakeholders' interests over the long term. As there are a variety of organizational and funding options for tackling the grand-challenge problems for human language technologies, the HLT-focused IUCRC or ERC could partner with some of these other existing models for collaborations. This partnership would bring together researchers working within other arrangements in order to broaden the research portfolio of the partners and allow them to tackle potentially larger problems.

Attracting diverse talent pool

The advantages of a center were deemed to include the pooling of good people, ideas, and infrastructure to solve new problems, while providing a broad collection of opportunities for visiting investigators from other institutions and industry. A center would be an ideal locus for consolidating ideas and efforts from university, industry, and government researchers, each bringing different perspectives to the problems the center would tackle. The center would attract researchers that excel in their disciplines given the potential to work with other researchers with similar levels of excellence. Bringing these groups together can lead to qualitatively new research because it unifies groups that otherwise would be working from different less interdisciplinary perspectives. This consolidation of diverse, excellent researchers should also be a magnet for funding (both center-based and individual or small group awards).

Addressing industry needs

The participants considered what industry would want out of an industrially-oriented HLT center. Many companies care about recruiting students who are well-trained in emerging technologies that would be part of a successful center. Also, the companies would benefit from a center that produces solutions for difficult problems such as global communication aids, speech in real environments (e.g., sensor-based projects, cocktail party challenge), and better speech synthesis. A center would help the company partners to be more competitive (both domestically and internationally) by providing the critical mass to work on hard problems that matter to them but that they cannot afford to do themselves. The center also has potential to enable a number of new companies to be created that depend on HLT. Another potential impact of a center on research companies might be that it offers a vehicle that could potentially support broader than DARPA-focused research (DARPA has recently been engaging companies to manage research teams).

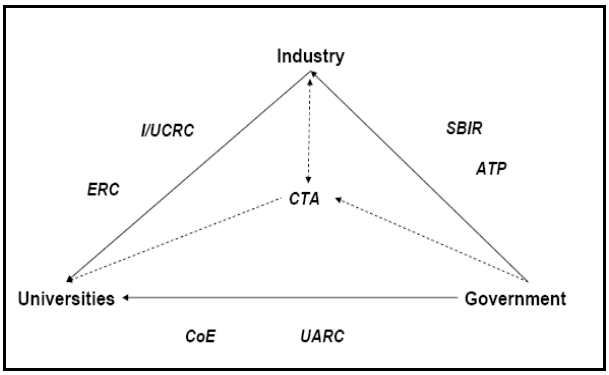

Figure 1. Center vehicles for collaboration between universities, industry, and government

Addressing needs of university researchers

The participants also considered what the university researchers would want from an industrially-oriented HLT center. Academics like to work on hard problems (e.g., deep NLP) that are not near-term. A center would provide the infrastructure and funding needed to support this type of research. Stability of funding is critical for attracting high quality students, post doctoral candidates, and faculty to the HLT center. Because obtaining center funding is challenging (especially an ERC award) and universities need steady funding to support good students (otherwise they move into other fields or leave for industry), having broad industry buy-in could help to create a stable funding base. The center would also attract visiting scholars from academia, industry, and government to help with the research agenda.

Tackling a diversified research agenda

Based on these discussions, the participants concluded that there is a good potential for a center to leverage the strengths of academic and industrial partners to tackle new human language technologies, such as virtual reality. A successful center would need to have a diversified portfolio of research problems; the research should be exciting, involve a multidisciplinary team, and result in innovations that can be used by industrial partners. If the center includes a sizable consortium of industry and government partners, it may be possible to build a massive infrastructure to support all of the partners. The center cannot simply produce core industrial products; it must also develop leading edge core technology, some of which may give rise to novel products given the guidance of the industrial partners. Some participants suggested that

the center should avoid tackling the large data processing problems, which are currently too expensive and so should be left to industry. Instead it may be better to focus on how to tackle, for example, low density languages (e.g., translation to and from rare languages with minimal parallel text, speech understanding with sparse per-language training data).

Membership cost to participate in an IUCRC or ERC

Since the preponderance of the support for an IUCRC comes from company membership fees, NSF requires a center to have at least six members with total company membership fees equalling at least $300,000 yearly. Although an ERC does not rely as heavily as an IUCRC on industrial support, NSF expects substantial financial support from industry, again typically provided through annual membership fees (usually two or three levels of membership with corresponding fees and membership benefits). Participants at the meeting believed that the cost of participating in an IUCRC or an ERC could be prohibitive for some companies, especially for smaller companies. Although it may be a challenge to obtain funding from industry, if it is clear that industrial partners have some control over how their membership fees are spent (and can leverage other funding), they will have greater interest in participating in the center. An effective IUCRC or ERC cannot take money without considering their industrial partners' needs.

Control of funds and intellectual property rights

Some industrial participants expressed the concern that in a broad-based center they would lose direct control. For example, some companies already have mechanisms for educating and recruiting students; they identify and directly support faculty who train students according to their specific needs. There was concern that being part of a center would mean that less of their funding would get to those researchers they would want to support (due to overhead and center priorities). There was also concern about losing control of intellectual property (IP). Some companies, especially small ones, keep things secret, worry about the potential risk of IP leaking, and usually do not patent.

Industrial participation mechanisms

Industrial partners would have a number of ways to influence the center. They could negotiate with the center universities (with some limitations set by the NSF programs) either when the center proposal is being developed or after it has been funded. Also by participating on the advisory board, an industrial partner can have a strong impact on the work conducted by the center (thus leveraging all of the center's funding) and recommend center affiliates. Also, industry partners who contribute more funding and effort to the center should receive greater center benefits than less engaged partners.

Identifying focus and markets for HLT products

The participants stressed the importance of identifying a multi-disciplinary focus that has an actual or potential market. If the center focus is too narrow, then it may be hard to find enough support. If the center focus is too wide, then research efforts will be less coherent and more difficult to manage. Currently there are few money-making products in speech processing or machine translation (though the opposite is true for web-search), so it is prudent not to define HLT technologies too narrowly. Additionally, projections about plausible markets are likely to need revision with potential impact on ideal partnerships. Formulating markets where language would play a role was thought to be a useful exercise even outside of the effort to define an HLT center. Several possible avenues for potential HLT products were identified:

- Social domain language-related products
- Commercial targeting of potential customers (advertising)
- Automating the creation of call center systems
- Information integration (e.g., customer relationship management, internal and external business intelligence, and brand marketing)[2]
- Construction industry language problems for foreign workers (5% of revenue is spent correcting mistakes, and there are also safety problems)
- Vertical high-accuracy translation markets, such as legal system translation
- Hospitals' need to provide medical help in a variety of languages
- Assignment of insurance categories to medical reports
- Law enforcement applications
- Service to government goals or the government organization itself
- Reducing language barriers in information access (e.g., cross-lingual search engines)
- Question answering in any language
- Translingual information mining and access across media
- Communicating with the speech impaired (text-to-speech), the manually impaired (speech-to-text), the visually impaired (speech again), or linguistic minorities (machine translation)

One thought was to look at 18-year olds to find where the markets will be in near future (e.g., instant messaging has moved into business, video gaming). Successful centers seem to involve many industrial partners, so it is not ideal to settle on just one market. Finally, it may be worth thinking about problems in two ways: what are the limiting factors in advancing language technology AND how is language technology itself a limiting factor in other applications?

Additional industrial perspectives needed

Participants raised an additional issue that should be considered more thoroughly. Since the industry representatives at this initial meeting were by and large from larger companies, some of the other important industry voices were not heard. There is a need for

[2] Perhaps companies interested in the data resources may be less competitive about the core technologies.

input from companies that are the language technology consumers, but do not have their own investments in research. It would be beneficial to involve in future meetings several representatives from technology-consuming industries that rely on HLT, but do not or cannot pay for all of the costs of HLT research and development themselves.

In summary, the meeting participants would expect a viable HLT center to be challenge-centric, with experts in the necessary disciplines, a shared vision with all partners, shared infrastructure, and ample funding to attract partners from industry and government labs, and to provide a stable base for sustained research and education of students. All participants agreed that the ideal center would have a lifetime that is longer than a standard NSF grant, with an explicit goal of becoming self-sustaining. Participants estimated this would take five to ten years, although the industry partners tended to believe that shorter durations would be possible.

Discussion Item 2

How can we best optimize the collaboration between industry and academia in a center environment?

Most participants agreed that the center should be multi-disciplinary with multiple co-PIs per center-supported project (with a mixture of perspectives). Multiple universities, government labs, and industries of a variety of sizes would contribute to building a strong center with broad impact. The center needs to be heterogeneous and inclusive, with one institution selected as the management hub for the center. Flexibility to adjust research focus was seen as an advantage, as long as expertise to meet the requirements of the challenges set by the center is maintained. Small companies were considered critical for the vibrancy of the center since they will play an important role in technology transition and product development.

IUCRC vs. ERC as initial program

Most participants felt that an ERC would be a more effective mechanism for building an HLT center than an IUCRC due to the higher levels of ERC funding, providing for the right infrastructure at the outset. Many participants believed that it would be hard to sustain a center in the long term on membership fees alone. This led others to suggest that the IUCRC should only be a first step.

Factors for building a successful center

Moving people bi-directionally between organizations was thought to be as important as money for building a successful university-industry center. It is more common for academics than researchers in industry to visit different organizations for long periods of time (e.g., sabbaticals). Industrial researchers visit other organizations, but typically only for short periods of time. Location of the center is critical for supporting this culture. Closer proximity of industries to the center may facilitate visits.

Other factors identified as critical for building a winning partnership include: an industrial liaison (master cajoler), an industry advisory board (with power), a director who reports to the board, Chief Scientist position(s), dedicated management (benign, not dictatorial, but with clear responsibilities), empowerment of PIs, encouragement for companies to place people at the center, student internships (from other institutions), and visiting faculty.

To engage students, the center should be located at one or more universities. Also, the center should focus on evolving "cool" areas of research, technology, and/or suite of potential applications. Robotics is cool for students. How about "Language/speech enabled agents," NLP–based web services, or a Universal Star Trek translator?

To engage industry, industrial partners should help define the challenges, while working with center leadership to select, filter, generalize, and modify recommendations for projects before they move forward. In some cases, industry may suggest specific applications that center efforts will generalize. It is also vital to involve industry in defining the center concept that will be proposed. Center retreats were suggested as one mechanism for obtaining industry input once the center is in place.

Although IP policies were discussed and some participants believed that they should be liberal and negotiable, much depends on the participating universities' policies. Additionally, the best practices for IUCRCs and ERCs (as defined in the center materials given at the beginning of this report) should play a role in working out IP policy. Another issue discussed is the need to develop mechanisms for pooling data resources while preserving ownership. Open versus non-open source code resources, as well as cross-licensing, should also be discussed with the industry partners.

Models for collaboration

One group drew a diagram representing one possible model of collaboration (see Figure 2). It details the flow of research prototypes and researchers, funding, special requirements, expertise for standards development, and products among government, universities, existing HLT industries, HLT consuming industries, and incubators and small companies.

Two themes were identified as candidates for organizing the center:

1. ***An HLT infrastructure and education center:*** This center might be focused on developing a component repository for HLT (essentially a reusable software version of LDC) together with an architecture and APIs for assembling components (perhaps UIMA-based). Given this framework, members could develop demonstration prototypes for research, education, and industry. To support education of students, teaching materials could be developed that are based on the components and architectures. These products would be tested among participating institutions and then shared as open source

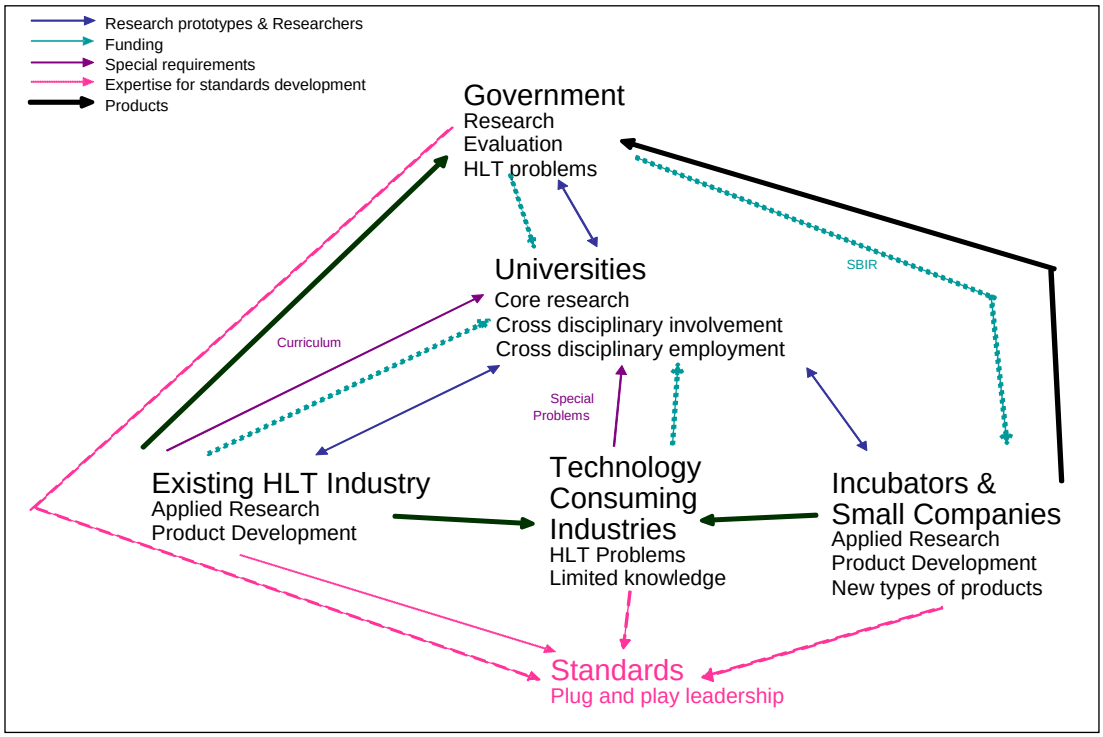

Figure 2. Possible linkages and funding options for collaboration among universities, industries, and government

(curricula, exercises, lectures, components, and data) or presented in an industry showcase for language technologies. The center would need computing and data infrastructure to build better HLT solutions. It is important to provide open access, when possible, and a firewall otherwise, for access to proprietary data. For a multi-site distributed entity, infrastructure should be accessible to all participants, including industrial partners. The CISE Computing Research Infrastructure (NSF, 2007a) and Global Environment for Networking Innovations (NSF, 2007c) programs may have a role to play in supporting this type of center.

2. ***A grand-challenge centric center:*** In this center, the challenges come from consensus among researchers and/or directly from industry, with one to three grand challenges per center. There should be spinoff technologies along the way, free cross-licensing of any and all technology among center partners should be considered, and at least some technology should be open source. Such grand challenges for the center to address could be:
   - Building a universal translator (any-language to any-language)
   - Developing personalized learning web agents that live in the web and communicate in natural language with users, read web pages, and perform a variety of useful tasks
   - Creating question answering systems for any language
   - Developing robust speech recognition with human-like capabilities to cope with cross-talk, noise, acoustic deformations (e.g., the speaker suffering from a cold, or whispering)

Homework

What breadth of research should an HLT center cover? Which areas of research are most viable for center collaboration?

Some participants focused on the possible challenges for the grand-challenge type of center:

- Robust speech recognition in cross-talk situations
- Cross-lingual (and perhaps cross-media) question answering, where answering the questions requires unifying information from more than one source (so it is not just answer retrieval), and perhaps more than one language or modality
- Rapid machine translation for resource-poor (minority or endangered) languages
- Learning from text, where the knowledge acquired is tested by performance on tasks
- "Universal" help-desk dialog system that can be rapidly configured for specific applications
- Tough problems coming from industry with 3-5 year (or longer) timeframes, where the researchers get to vet or select from longer list, focusing on the most interesting and generalizable challenges

Possible markets

Others felt that finding good science is easier than finding good markets for a center, and so focused attention on possible markets, including national security, health assistive technologies (gerontology, speech therapy, health monitoring, etc.), education, cybertrust, geospatial applications (e.g., maps), temporal applications, or alignment across media.

Center for cross-cultural communication

One comprehensive idea for a center involving both grand challenge problems and markets was proposed

that resulted in much enthusiastic discussion: A center for cross-cultural communication/collaboration technologies (in cyberspace). This center must be multidisciplinary with following necessary disciplinary areas: human language technology (automatic speech recognition, machine translation, information extraction, etc.), multimodal areas (human-computer interaction, engineering, human factors), cultural anthropology, linguistics (language experts, sociolinguistics, etc.). An international dimension was thought to be critical.

Areas that seem to be emerging that could be addressed by the center include: blogging and social network analysis, cultural specific aspects of language, mobile technologies, marketing across counties and cultures, coping with cross-language training (accent mitigation, language use, etc.), how language used by various groups changes over time (e.g., discourse analysis, rhetoric, media environment, spin, register, data, sciops (how organizations react)).

Possible markets identified for such a center include cross-cultural collaboration technologies, multicultural language-based discourse, social networking, marketing, brand monitoring, My Space, cross-border tutoring, call centers, expert finding (hiring)—e.g., email patterns, emerging market analysis, State Department, and tourism.

Many factors affect the need for the technology that the center would produce. For example, China and India have different needs and commercial interests based not only on language, but also based on societal factors; after all, good interfaces to technology take into account the diverse contexts of use and the range of individual differences among users.

Possible broader impact

One participant pointed out the findings of a recent congressional committee hearing that may affect the problems addressed by the center; they are summarized below:

- Technology is necessary, but we must evaluate its impact and invest wisely.
- Increasing the capabilities and efficiency of level one and two linguists using technology such as machine translation is critical because we will never have enough level three linguists.
- Increasing the pool of US citizens who know a second language, particularly languages of interest such as Chinese and Arabic, is a critical national priority. If technology can play a role in this, that is even better.

The needs identified by this congressional committee could help enhance the broader impact of a center's grand challenges.

Discussion Item 3

What are the next steps?

The participants agreed that the best way to move forward is to begin the process of building a center. They decided that a multifaceted approach would provide a staged, successful strategy.

- The first step would be to develop a plan for a multi-university IUCRC with a goal of leveraging this effort into a proposal for a multi-university ERC. Although the universities and their industrial partners will take over funding the center eventually, having NSF imprimatur at the start would help immensely with the development of the center. If the proposed center embraces one or more grand challenges, they should be identified and their importance and feasibility justified.
- In addition, in tandem, we should seek to develop a congressionally funded National Institute for HLT.

Developing a multi-university IUCRC then an ERC

The ERC program would provide an appropriate level of funding to create a vibrant center; however, such center funding is very challenging to win, so advanced planning is critical. Planning and coordination need to start well before the solicitation comes out, and people need time to develop the concept of the center. To begin planning for the staged HLT center, the participants suggested asking for support from deans, provosts, VPs of research, and departments at several universities. Ideally, these institutions would provide some infrastructure for developing the center concept (e.g., release time, facilities, resources for fund-raising, and co-sponsorship). Having the weight of the community behind an ERC proposal would provide the necessary base for convincing potential funders of the necessity of a center.

Leveraging the IUCRC was thought to be a good first step in developing an ERC, especially for developing the industrial component. For planning the IUCRC, the participants thought it vital to immediately begin building ties with industry (along the lines of Figure 2). This requires assembling a working group of volunteers with the time to begin the planning process. As for deciding who will lead the effort going forward, one possibility is teaming a visionary leader with someone who has great planning and execution skills. A critical mass of working group members (not too many, but not too few) would be beneficial, with one from each university. When building a list of potential partners, it is important to select some partners who are capable of making ties with industry and helping to define who the consumers of the technology products of the center (i.e., third party customers) would be. There is an issue of group dynamics that may need to be addressed; one person might end up carrying the full load (everyone is happy to play, but none willing to step up and work), reducing the overall chance of success. Members should get buy-in from their universities, and they need to contribute concretely to the action items developed by the group. Identifying which institution will lead is a priority, as well as identifying which institutions will be partners in this multi-university HLT IUCRC. Agreements between these sites cannot begin too soon.

The IUCRC working group will need to:

- Discuss possible alternative approaches, develop a high-level vision, and collect evidence to convince companies to participate in the center.
- Build ties with industry, both large and small companies. The group should develop strategies for outreach to small companies. Assembling an industry working group and running a few focus groups may help to build an industrial strategy.
- Organize a series of planning meetings. These meetings (hopefully on both coasts) should involve industry, academia (US and international universities), and others (e.g., government labs, centers such as the Hopkins CoE, LDC, and possibly professional societies). Planning meetings should involve companies of all sizes. At these meetings, the working group will present the high-level vision of the center, as well as sub-visions targeted to industry cliques. The working group will need to identify the cliques based on which companies are interested. For small companies, it may be necessary to cover some expenses to come to the meeting or possibly some of their time (although this would be somewhat challenging to do with limited planning funds provided by NSF and universities).
- Develop an international strategy. Several participants thought this was fundamental for establishing the credibility of the center and for supporting the follow-on ERC effort. The group needs to identify and court international partners in order to add new dimensions to the challenges being tackled by the center. When identifying international partners, it would be beneficial to consider value added (e.g., What expertise does a site have to offer that is not covered at the center? Does it have or is it applying for parallel funding?). The NSF Office of International Science and Engineering (OISE) can help (NSF, 2007f).
- Begin proposal planning and preparation for the IUCRC (NSF, 2007e) with the goal of a January 4, 2008 deadline for letter of intent and a March 28, 2008 deadline for a proposal for a planning grant and follow-on planning meetings that will be needed to write a successful full center proposal.

Developing a National Institute for HLT

The establishment of a National Institute for Human Language Technology (HLT) would declare HLT as a national resource. This institute would need to involve a large number of universities and companies. Some companies already take an active role in congressional actions (e.g., SAIC and Lockheed), and so involving them may contribute to our success.

There are challenges in managing an effort with a large group of companies and universities. Definition of the role of the institute is critical. Does it host meetings at conferences, have an agenda, have a goal, share information, and facilitate collaboration among PIs? Should it have an international aspect? Where should it be located? (Maybe there should be both an east and west coast arm.)

In support of the campaign for this institute, some of the attendees are working on an executive summary describing the institute and its rationale. This summary will provide talking points for members to go to the leadership of their respective institutions to obtain support to work on the institute. Many participants have indicated an interest in helping to build the institute. Moreover, highlighting success stories in the evolution of human language technology will help increase the awareness of its importance in academic, governmental, and general audiences.

Some participants agreed to discuss the prospects of the center and institute at a number of upcoming conferences, including Interspeech, ACL, and ICML. It was also suggested that we put together a mailing list to send information out to potentially interested parties and plan a future one-day workshop related to the institute to plan for its evolution. A quarterly newsletter would be useful to update interested parties. With residual funds from the workshop, we plan to set up a Wiki at the University of Maryland to support both the center and the institute efforts.

Acknowledgments

We thank Tanya Korelsky for her vision and enthusiastic support of the workshop, Bruce Kramer, Adam Powell, Alex Schwartzkopf, and Janis Terpenny for their presentations, and Shaina Castle for assistance in organizing the meeting and editorial support. We also thank Caitlin Christianson and Joe Olive for their helpful comments during the meeting. This report is based upon work supported by NSF IIS 0702910. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

Bibliographical References

M. Absher et al. (1998). ERC Best Practices Manual, www.erc-assoc.org/manual/bp_index.htm

D. O. Gray (2007). Industry-University Cooperative Research Centers (IUCRC), Program Evaluation Project, www.ncsu.edu/iucrc/ index.htm.

D. O. Gray and S. G. Walters (1998). Managing the Industry/University Cooperative Research Center: A Guide for Directors and Other Stakeholders, www.ncsu.edu/iucrc/PurpleBook.htm.

ERC Assoc. (2007). About the ERCs, www.erc-assoc.org/

Harper et al. (2007). Report on the NSF-sponsored Human Language Technology Workshop on Industrial Centers, ftp.ecn.purdue.edu/harper/Final%20HLT%20Center%20Report.pdf

NSF (2007a). CISE Computing Research Infrastructure www.nsf.gov/funding/pgm_summ.jsp?pims_id= 12810.

NSF (2007b). Engineering Research Center Program (ERC), www.nsf.gov/funding/pgm_summ.jsp?pims_id=5502&%20org=NSF&sel_org=NSF&from=fund.

NSF (2007c), Global Environment for Networking Innovations (GENI). www.nsf.gov/cise/cns /geni/.

NSF (2007d). Industry/University Cooperative Research Centers Program (I/UCRC), www.nsf.gov/eng/iip/ iucrc/.

NSF (2007e). Industry/University Cooperative Research Centers Program (I/UCRC), Program Solicitation 07-537, www.nsf.gov/pubs/2007/nsf07537/nsf07537.pdf.

NSF (2007f).Office of International Science and Engineering (OISE), www.nsf.gov/div/index.jsp?div= OISE.