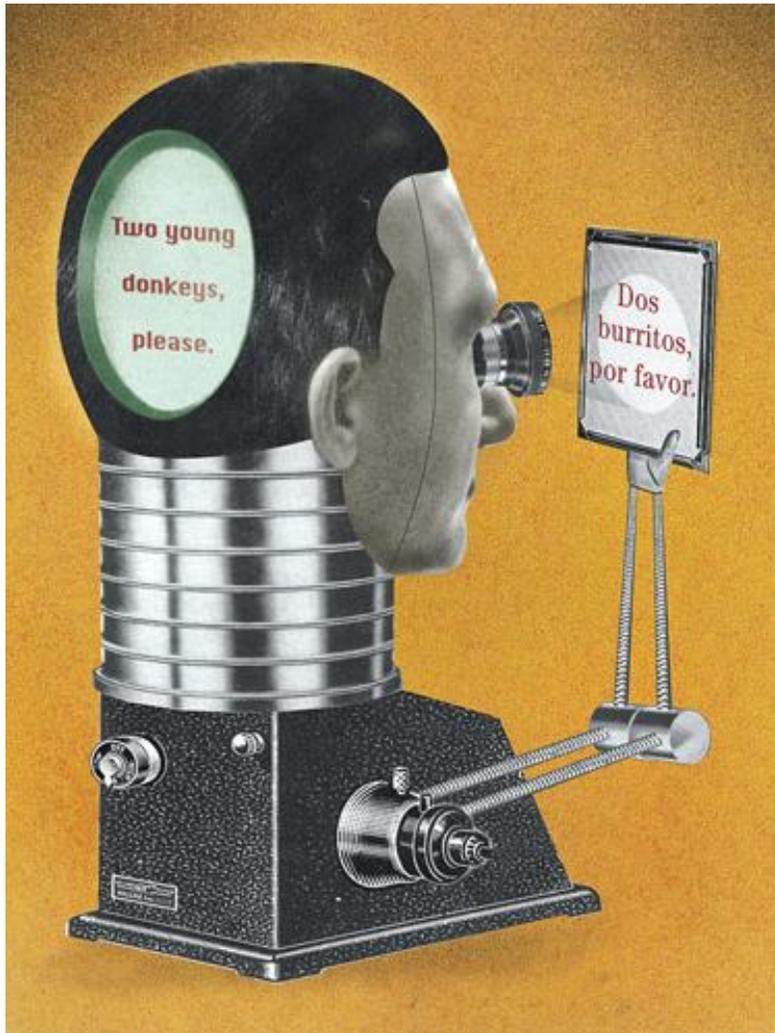# MURI Topic Proposal:

Structured Translation and Analysis of Low-Resource Languages

Chris Arney and Joe Myers
with many coordinators
Jim Harvey, Clare Voss, Barbara Broome,
Melissa Holland, Steve LaRocca (ARL),
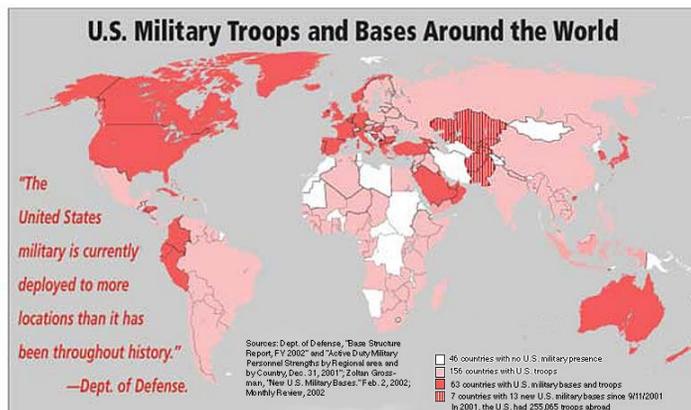Many others in the
Multilingual Computing Branch, CISD, ARL
Jerry Ball (AFRL),
Alan Schultz, Dennis Perzanowski (NRL),
Joe Olive (DARPA),
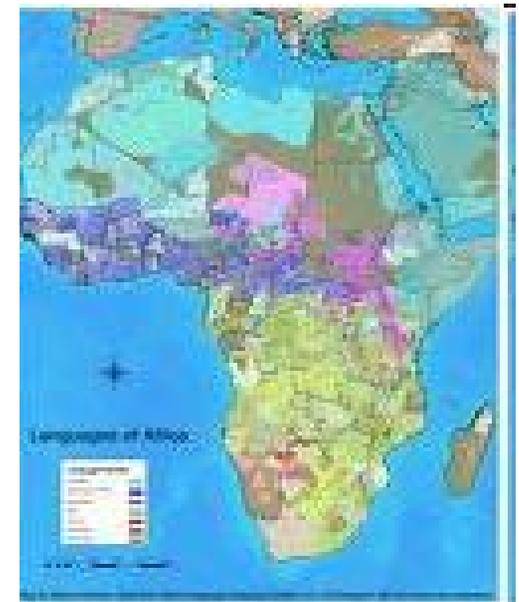Heather MacCalum-Bayliss, IARPA,
Tanya Korelsky, NSF

DoD Needs

- The need for MT breakthrough is substantial as the US currently maintains **troops in 144 countries**, many of which employ mid- and low-density languages that have no MT capabilities, and our communications and signals intelligence processes are overwhelmed in volume and numbers of languages to monitor.

- This language challenge is growing – MT is a huge problem for the newly established Africa Command. Africa is a continent with very high linguistic diversity; there are an estimated **1500-2000 African languages** with practically no MT help in any of these.

U.S. Military Troops and Bases Around the World

"The United States military is currently deployed to more locations than it has been throughout history."
—Dept. of Defense.

Sources: Dept. of Defense, "Base Structure Report, FY 2002" and "Active Duty Military Personnel Strengths by Regional area and by Country, Dec. 31, 2001"; Zoltan Grozman, "New U.S. Military Bases." Feb. 2, 2002; Monthly Review, 2002

- 46 countries with no U.S. military presence
- 155 countries with U.S. troops
- 63 countries with U.S. military bases and troops
- 7 countries with 13 new U.S. military bases since 9/11/2001

In 2001, the U.S. had 255,065 troops abroad

Need both:

- Tactical
- Strategic

Impacts

- This project will enable the **advancement of tactical (local, native) and strategic (intelligence processing/information fusion)** communication through a new framework for integrating core linguistic and adaptive learning for MT and TA.

- This research is critical to DoD's future in information and net-centric warfare. The research results will **transition to the ARL, TRADOC, operational, and intelligence communities** as the algorithms will be implemented by the intelligence community.

MURI Objective

What is needed now is a structured modeling approach to MT and natural language processing with the advantage that MT tools based on structure rather than stochastic substitutions will open the way forward to efficient MT development for hundreds of languages and many valuable TA capabilities.

Research Concentration Areas

- Improve quality of MT by **developing architectures** and (linguistic, psychological, and mathematical) **models** including semantics and pragmatics and building more effective MT and TA tools.

- Use the new MT tools, demonstrating their use for **two-way translation of low/medium-level resourced language**.

- Improve language processing by **moving closer to comprehension**.

- Use the models to **paraphrase, interpret, and extract ideas** and themes from documents.

- Use the models to **summarize, correlate, and distill the content to answers** of general or specific questions about the information in the sources.

Congratulations, CMU, USC-ISI, UT-Austin, MIT!

Major External Events

Annual reviews: Oct 2011 – Oct 2015

Annual IPRs with metrics: ditto

DDRE review presentation: Spring 2012

Program reviews: May 2012, 2014

Option decision: Spring 2013

DoD visibility/inquiries: ongoing! Please help us keep management informed by providing as available:

    -- significant developments

    -- copies of summary or significant talks and papers

Emphasis

Messages from DDRE:

- *Focus* on advancing the basic science, not on transitions or deliverables. (But do *enable* transitions and deliverables as they become appropriate).
- Every portfolio needs some risk; this is ours.

Items of interest for annual IPR data and metrics:

-- Scientific Objectives of the project, Scientific Barriers and challenges which make/have made these objectives difficult to achieve in the past and which must still be overcome, and New Scientific Opportunities that now make these objectives more tractable than in the past

-- Army Impact/Relevance -- what good will this do for the Army (I will help here)

-- What Scientific Approach is the team taking?

-- What has been done, what is planned, what is the status, how are things going?

-- What Accomplishments have been made (interim or significant)?

-- What Collaborations are going on between the team and others? Be specific: what and with whom?

-- What Leveraging is going on? What other efforts (or funding or etc) is the team taking advantage of, or has the team been able to provide to other efforts?

-- What Technology Transfer has taken place? What DoD or govt agencies (or anybody else) have you been able to give software, papers, ideas, advice, etc, that they have been able to employ in their work?

-- What are Future Plans or the next steps?

-- Various Metrics:

   -- List of all peer-reviewed publications during the previous year which list support from this project (you will want to encourage everyone to remember to acknowledge support from this grant wherever appropriate)

   -- Number of unpublished manuscripts during the previous year

   -- Number of presentations made during the last year which involve work from this project

   -- Number of patents applied for, and number received, during the last year which are related to this project

   -- Number of grad students supported by this project during the last year

   -- Number of postdocs supported by this project during the last year

   -- Number of faculty supported by this project during the last year

   -- Number of students supported by this project receiving a PhD during the last year

   -- Number of students supported by this project receiving a MS (or BS) during the last year

   -- Awards or honors won by anybody supported by/working on this project

   -- list name of honor or award (award or honor does not need to be related to the project; this is a measure of the quality of the faculty chosen to work on our projects)

   -- Anything else noteworthy, eg names of students accepting employment at Army labs, etc

Items of interest for annual IPR data and metrics:

-- Scientific Objectives of the project, Scientific Barriers and challenges which make/have made these objectives difficult to achieve in the past and which must still be overcome, and New Scientific Opportunities that now make these objectives more tractable than in the past

-- Army Impact/Relevance -- what good will this do for the Army (I will help here)

-- What Scientific Approach is the team taking?

-- What has been done, what is planned, what is the status, how are things going?

-- What Accomplishments have been made (interim or significant)?

-- What Collaborations are going on between the team and others? Be specific: what and with whom?

-- What Leveraging is going on? What other efforts (or funding or etc) is the team taking advantage of, or has the team been able to provide to other efforts?

-- What Technology Transfer has taken place? What DoD or govt agencies (or anybody else) have you been able to give software, papers, ideas, advice, etc, that they have been able to employ in their work?

-- What are Future Plans or the next steps?

-- Various Metrics:

-- List of all peer-reviewed publications during the previous year which list support from this project (you will want to encourage everyone to remember to acknowledge support from this grant wherever appropriate)

-- Number of unpublished manuscripts during the previous year

-- Number of presentations made during the last year which involve work from this project

-- Number of patents applied for, and number received, during the last year which are related to this project

-- Number of grad students supported by this project during the last year

-- Number of postdocs supported by this project during the last year

-- Number of faculty supported by this project during the last year

-- Number of students supported by this project receiving a PhD during the last year

-- Number of students supported by this project receiving a MS (or BS) during the last year

-- Awards or honors won by anybody supported by/working on this project

-- list name of honor or award (award or honor does not need to be related to the project; this is a measure of the quality of the faculty chosen to work on our projects)

-- Anything else noteworthy, eg names of students accepting employment at Army labs, etc

Government Advisory Panel

Kathy Baker (DoD)

Jerry Ball (AFRL)

Kelly Dobson (AMRDEC)

Lisa Harper (CIA)

Teri Langendoen (NSF)

John Lavery (ARO)

Marianna Martindale (DoD/CAMT)

Jeff Micher (ARL/CISD)

Rod Sturdivant (USMA)

Carol Van Ess-Dykema (Director of Technology, Office of the Director of National Intelligence/National Virtual Translation Center)

Clare Voss (ARL/CISD)

Joe Myers (ARO)

Functions:

-- Provide feedback, guidance to MURI team

-- Keep organization cognizant

-- Enable transition opportunities