

THIS PAGE INTENTIONALLY LEFT BLANK

(U) RESEARCH & TECHNOLOGY
(U) HUMAN LANGUAGE TECHNOLOGY RESEARCH

This Exhibit is SECRET//NOFORN									
	FY 2011 ¹ Actual	FY 2012 Enacted			FY 2013 Request			FY 2012 — FY 2013	
		Base	OCO	Total	Base	OCO	Total	Change	% Change
Funding (\$M)	26.4	31.0	3.0	34.0	26.0	3.4	29.4	-4.7	-14
Civilian FTE	8	8	—	8	8	—	8	—	—
Civilian Positions	8	8	—	8	8	—	8	—	—
Military Positions	—	—	—	—	—	—	—	—	—
¹ Includes enacted OCO funding.								Totals may not add due to rounding.	

(U) Project Description

(U//FOUO) The Human Language Technology (HLT) Research Project provides a coherent, concentrated focus on language analytics to exploit the volume, variety, and velocity of communications that the SIGINT system collects. HLT Research conducts research that supports the goals of the NSA/CSS' Analytic Modernization effort. This Project complements NSA/CSS initiatives to strengthen the language analyst workforce by providing the technologies that serve as force multipliers for analysts.

(U//FOUO) The HLT Research Project has an HLT Center of Excellence (COE) at Johns Hopkins University to promote academic and industry interest in intelligence challenges and attract world-class talent to work on IC HLT problems. The HLT COE focuses on critical intelligence needs that are not adequately addressed by commercial technology or other government programs. The HLT Research Project also leverages programs at the Defense Advanced Research Projects Agency (DARPA) and the Intelligence Advanced Research Projects Activity (IARPA). DARPA and IARPA programs provide foundational HLT capabilities in automatic content extraction, speech-to-text, machine translation, summarization, and question answering. The HLT Research Project conducts research and advanced development necessary to bridge research results from DARPA's and IARPA's efforts to SIGINT applications. This Project includes the Human Language Technology Research Sub-Project.

(U) Base resources in this project are used to:

- (S//SI//REL TO USA, FVEY) Research and develop voice, text, video and image analytics to enable fundamental language exploitation capabilities for all types of communication, regardless of medium.
- (S//SI//REL TO USA, FVEY) Increase the number of languages, accuracy, and speed of results for keyword search from machine-generated transformations of speech-to-text.
- (S//SI//REL TO USA, FVEY) Conduct research and advanced development on automatic document image analysis, particularly for handwritten documents, an extreme technical challenge. The primary emphasis is on core capabilities to enable triage and keyword search on the diverse kinds of documents found in intercept, including language and script identification and handwritten document detection, segmentation, and analysis.
- (U//FOUO) Research analytics that automatically analyze the linguistic content of communications. This area comprises several technologies, including content extraction and machine translation. Content analytics identifies and extracts information from language communications, turning a mass of unstructured text into usable metadata.

- (TS//SI//REL TO USA, FVEY) Research, design, and develop analytics that enable deployment of HLT capabilities nearer to the point of collection within the SIGINT system.
- (U//FOUO) Support collaborative research into human language exploitation and machine learning with commercial and academic partners.
- (U//FOUO) Develop test and training data to support scientific research and evaluation.
- (U//FOUO) Provide and maintain a computer lab to support in-house algorithm development, evaluation, and proof-of-concept demonstrations of promising solutions.
- (U//FOUO) Sustain support activities that foster cross-organizational and cross-discipline collaboration in solving hard technical problems critical to the success of NSA/CSS' SIGINT and cyber missions as well technical health of the workforce.

(U) There are no new activities in this Project in FY 2013.

(U) OCO resources in this project are used to:

- (TS//SI//REL TO USA, FVEY) Enable machine translation research and new speech processing capabilities for Afghanistan and Pakistan dialects using state-of-the-art research findings in less-common languages and by developing new language and dialect models.

(U) The CCP expects this Project to accomplish the following in FY 2013:

- (S//REL TO USA, FVEY) Develop and deploy speech-to-text models for additional languages, where the languages will be selected according to corporate NSA/CSS priorities, language analyst preparation, and scientific assessment of technology readiness. [CCP_0106]
- (S//REL TO USA, FVEY) Extend name-finding solutions to support named-entity extraction for 12 additional languages, to include at least three languages that are less-commonly taught. Create and demonstrate solutions in three to five languages for the much harder problem of extracting relations between entities. These capabilities will yield automated solutions to uncover pertinent facts within both unstructured written communications and spoken communications that have been transformed into text. [CCP_0106]
- (U//FOUO) Design techniques to reduce by 25 percent hand-annotated data required to develop models in support of speech-to-text solutions. [CCP_0106]
- (S//REL TO USA, FVEY) Research, develop, and demonstrate solutions for cross-lingual entity disambiguation to enable analysts to perform language independent retrieval of communications to, from, or about persons of interest from multi-lingual SIGINT data sets. [CCP_0106]

(U) Changes From FY 2012 to FY 2013:

(S//NF) Human Language Technology Research: -\$4.7 million (-\$5.1 Base, +\$0.4 OCO). The aggregate decrease is the result of:

- (U) Increases:
 - (S//NF) \$0.4 million in Overseas Contingency Operations (OCO) accelerates new speech processing capabilities and associated analyst applications for Afghanistan and Pakistan dialects.

- (U) Decreases:
 - (S//NF) \$5.0 million due to a FY 2012 Congressional add not sustained in FY 2013.
 - (S//NF) \$0.1 million due to a planned programmatic reduction in travel and training.

Human Language Technology Research Project Budget Chart FY 2013 Budget Request by Appropriation Account This Exhibit is SECRET//NOFORN			Funds — Dollars in Millions		
Subproject	Description	Resourcing	FY 2011	FY 2012	FY 2013
<i>Operation and Maintenance, Defense-Wide</i>		<i>Funds</i>	—	—	1.12
		<i>Positions</i>	—	—	8
Human Language Technology Research	Pay and Benefits	Base	—	—	1.12
		Positions	—	—	8
<i>Research, Development, Test, and Evaluation, Defense-Wide</i>		<i>Funds</i>	26.36	34.03	28.23
		<i>Positions</i>	8	8	—
Human Language Technology Research	Communications and Utilities	Base	0.06	0.04	0.04
	Contract Services	Base	24.35	28.07	23.36
		OCO	—	3.00	3.40
	Equipment	Base	0.57	1.76	1.36
	Pay and Benefits	Base	1.20	1.09	—
	Travel and Transportation	Base	0.17	0.07	0.07
	Positions		8	8	—
Totals may not add due to rounding.					

THIS PAGE INTENTIONALLY LEFT BLANK