

## **(S//SI//REL) How Is Human Language Technology (HLT) Progressing?**

FROM: (U//FOUO) [REDACTED]  
Language Analysis Modernization Lead (S2)  
Run Date: 09/06/2011

*(S//SI//REL) Editor's intro: At the SID town hall meeting of February 2011, [REDACTED] (pictured) briefed on Human Language Technology, i.e., tools that sort through SIGINT voice collection and automatically find the most promising nuggets, thereby saving linguists countless hours. What's happened with HLT since that time?*

(S//SI//REL) In 2011 we deployed HLT Labs to Afghanistan, NSA Georgia, Latin American SCS sites, and NSA Texas.

### **(U) Afghanistan-area targets**

(S//SI//REL) Afghan Regional Operating Cryptologic Center (AROCC) analysts started using HLT Labs to track their targets in April, and when the analytics were successfully used to find new information, the mission was expanded to include international teams.\* The Afghanistan deployment boasts some technological firsts associated with cloud computing\*\* and includes the full suite of analytics with Pashto speech-to-text (STT). Recently French analysts in the ARC were able to find target speakers on new selectors using speaker recognition.

(S//SI//REL) Our deployment to NSA Georgia enables us to partner with analysts to assess the performance of our newest STT models: Pashto and Farsi. These languages have limited training data which creates challenges for STT, and we have been focused on finding applications that are beneficial even for these low-resource languages. NSA-Georgia traffic includes noisy VHF collections which seriously degrade analytic performance; however, analysts can still find target speaker cuts on unknown frequencies.

### **(U) Spanish-speaking targets**

(S//SI//REL) Spanish is the most mature of our speech-to-text analytics, and has higher keyword-search accuracy than other deployed STT models. We've had great success searching for Spanish keywords at NSA Texas and Latin America SCS sites.

(S//SI//REL) For example, in early August a new NSA Texas user applied keyword search the morning after his training to find a previously unreported cut from a drug trafficking target. Likewise, the OIC of one of the Latin American SCS sites recently reported he was able to find foreign intelligence regarding a Cuban official in a fraction of the usual time. His comment: This same example could be used over and over by many that have to go over countless voice cuts to finally dig that gold nugget that will turn into a report.

### **(U) Development work continues**

(U//FOUO) The R6 research team is working to add new applications, improve keyword search capability, enhance analytics, add new languages, and refine the user interface. Recently the Summer Camp for Applied Language Exploration (SCALE) -- a joint NSA Johns Hopkins University exercise -- investigated new ways to use the results of HLT analytics from existing targets to find new targets. Research is also working closely with the [SPIRITFIRE](#) (voice analytics)

and [TransX](#) (translation, transcription and transliteration) efforts to ensure HLT Labs capabilities are included in the corporate solution for enterprise deployment in 2012.

(U//FOUO) More information about HLT Labs is available [here](#).

(U//FOUO) See a related *SIDtoday* article about HLT [here](#).

(U) Notes:

\* (S//REL) The international teams were from the Analysis and Research Cell (ARC), Task Force 310, and Combined Joint Special Operations Task Force (CJSOTF).

\*\* (S//SI//REL) Specifically, the Afghan deployment is the first use of [DISTILLERY](#) and [CLOUDBASE](#) on a [GHOSTMACHINE](#) platform.