(S//SI//REL) Finding Nuggets -- Quickly -- in a Heap of Voice Collection, From Mexico to Afghanistan

FROM: (U//FOUO) ██████████
Senior Technical Development Program -- Class of 2010, and Intelligence Analysis Technical Director, NSA Texas
Run Date: 05/25/2011

(U//FOUO) Recently I had a rare life-changing instance where the highly unexpected occurred, a so-called "black swan" event.* What was this event? It was my exchange with the Human Language Technology (HLT) Division in the Research Directorate and with many other HLT believers.

(U) Before submerging into HLT as my Senior Technical Development Program goal, I had no clue what HLT was. If asked, I would have defined the Human Language Technology concept as the ability of linguists to use the available technology to search for reportable intelligence from voice traffic. Speaker identification was no more than "yes, that's the guy!", I imagined. Language identification was even simpler: Spanish, not Spanish; and keyword search, well... what's a keyword?

(S//SI//REL) My definition of HLT has changed through the years. The *Human* part involves a lot more than simply understanding the target language, given the need for a language analyst to know customers' requirements, collection techniques, global networks composition, and legal rules. As for the *Language*, it is no longer enough to know what was said in speech or conveyed in writing. In order to determine intent, the analyst must know how things were said, in what tone and accent, the mood of the person, vocabulary usage, religious and political beliefs, nationality, type of device used, communication patterns, associations, and the location of the speakers to the nearest cell tower! Then there is *Technology*, which has given analysts a more comprehensive view of their target space and has also made that target space larger and impossible to navigate "a cappella."

(S//SI//REL) To successfully track those elusive and always-mobile targets, we must find them, regardless of which communication method they happen to be using. What we do not know is how many wrong-doers we might have overlooked or how many illegal operations we might have failed to uncover because of the volumes of data that we were not able to scan. *That's where Human Language Technology can help: it can find the exact traffic of interest within a mass of collection.*

(S//SI//REL) To *sell* the technology to those who would benefit from it, we need to convince a veteran linguist that a computer will actually find a target of interest using a statistically generated voice model, even if the phone number has changed! It's not an easy sell. We must also change the way we process and analyze collected intelligence. However, the HLT organization (R67, formerly R64) is helping analysts embrace the HLT concept and is enabling them to see further into their overstuffed queues.

(S//SI//REL) Thanks to the support of the SCS Director, researchers from R64 accompanied by this lone S2 language analyst, delivered HLT analytics to F6 sites in the Americas; the immediate success of the first HTL-Labs system was possible not just because the technology was mature enough, but because visionary leaders --with DIRNSA at the top-- and open-minded field analysts have accepted both the virtues and the flaws of this revolutionary technology.

(TS//SI//REL) SCS Spanish Language Voice Analysts have learned to exploit the advantages of speech-to-text keyword search, and have quickly integrated speaker, gender, and language recognition into their workflows. NSA-Texas analysts who were part of the initial testing and

validation of HLT analytics helped verified their utility, and the successes multiplied. From finding tunnels in Tijuana, identifying bomb threats in the streets of Mexico City, or shedding light on the shooting of US Customs officials in Potosi, Mexico, the technology did what it advertised: *it accelerated the process of finding relevant intelligence when time was of the essence.* (See related article.)

(S//REL) I did not expect to find myself across the world explaining the technology to military leaders and/or introducing translators at the Afghanistan Remote Operations Cryptologic Center (AROCC) to research and analysis skills, but then again, black swan events are unpredictable. DIRNSA appointed one of his military Fellows, ███████████ (*pictured below with author in Kandahar*) --a dedicated an exemplary Army officer-- to lead the deployment of HLT analytics to Afghanistan. From Kandahar to Kabul, we have traveled the country explaining NSA leaders' vision and introducing SIGINT teams to what HLT analytics can do today and to what is still needed to make this technology a game-changing success.

(S//REL) While the challenges of the Afghanistan language mission are not insurmountable, it will take many months for HLT analytics to reach the same level of performance as the Spanish systems deployed to SCS. The AROCC and SCS missions are different but the underlying problem is the same: not enough talented language analysts to process everything we collect. With a commitment from R to improve the speech-to-text functionality, with support from S, T, NSA Georgia, the AROCC and NCR leadership, and more importantly, with the keen interest of every language analyst in theater to learn about and use HLT tools, in time this technology will likely redefine the way speech is processed in the SIGINT missions of Afghanistan.

(U//FOUO) Analytic modernization is as much about technology as it is about people; understanding the needs of analysts and shaping technology to enable them to succeed is perhaps the most satisfying result, whether prosecuting narcotics traffickers in Mexico or Taliban leaders in Afghanistan.

(U) Notes:

* (U) The term comes from Nassim Nicholas Taleb's 2007 book "The Black Swan."