B14 Using Computational Forensic Linguistics to Screen Pedophilic Communications

Carole E. Chaski, PhD, Institute for Linguistic Evidence, 25100 Trinity Drive, Georgetown, DE 19947-6585; and Raye Croghan, BS, ALIAS Technology, LLC, 25100 Trinity Drive, Georgetown, DE 19947-6585*

After attending this presentation, attendees will understand a validated computational forensic linguistic method for assessing two communication types, threats and pedophilic grooming, in order to screen for pedophilic communications.

This presentation will impact the forensic and legal communities by delivering a method to discover early warning evidence of threatening and grooming behavior that results in child endangerment, abduction, and sexual assault.

Most adults experience threats and pedophilic grooming communications after the fact, when emotional or physical harm has been inflicted. Children, by virtue of their age, are even less likely to be able to accurately evaluate these covert communication types. Lack of exposure to such communications means that most of us are not able to recognize these communication types for what they are, even if we recognize that something is not right in the situation. Yet the ability to recognize and accurately classify different types of problematic texts has obvious survivability, as well as, investigative value in evaluating recidivism potential for convicted and repeat offenders. Computational Forensic Linguistics provides objective, intelligent classification for these rarely-experienced and very problematic text types.

As a branch of natural language engineering, Computational Forensic Linguistics quantifies specific linguistic features in text and dialog, and then subjects this quantification to statistical analysis for classification of documents into forensically-significant categories. ALIAS, Automated Linguistic Identification and Assessment System (Chaski 2005, 2007, 1997) is a computational forensic linguistic program with components for authorship, witness statement relatedness, and other forensically-significant questions. In this presentation, two components of ALIAS, ThreatAssess and PREText are discussed.

ALIAS ThreatAssess provides a very rapid (milliseconds) assessment of a text to determine if it is classified as a real threat or not. Using a database of real threat letters which have been involved in investigation or litigation and the Chaski Writing Sample Database of simulated threat letters, apologies, love letters, complaints, and angry letters as comparison texts, a cross-validated statistical model for classifying texts has been developed. Like the threat text type, each comparison text type has an interpersonal and emotional communicative purpose and therefore represents a good foil. Each new text fed into ThreatAssess is classified as either a real threat or a comparison type based on the statistical model whose accuracy is at least 92% with a maximum of 100%.

Built on ALIAS ThreatAssess, ALIAS PREText, or PREdator Text, provides a very rapid assessment of a text and/or a chat dialog to determine if it is classified as sexual predatory grooming or not. ALIAS PREText was developed using several different types of pedophilic communications: (1) pedophile to pedophile, (2) pedophile to victim, (3) pro-pedophile activism, (4) risky communications, and (5) defensive pedophile communications. Pedophile to pedophile data includes personal interactions between pedophiles dating back to 1996 as associated with pedophile participation in special interest pro-pedophile only membership groups where electronic communications took place through email, forums, and electronic chat. Pedophile to victim data includes grooming tactics captured between a pedophile and a child or an adult informant posing as a child where the pedophile acted on the chat by appearing physically to meet the minor victim. Many of these chats have been used in court as part of the conviction process. Pro-pedophile activism data includes known pro-pedophile activism web sites, blog articles, papers and letters promoting the pedophilia cause in defense of perceived persecution by society. Risky communication data includes electronic interaction between adults curious about pedophilic tendencies and the normalization of the perversion by the pedophile community and recruitment of new pro-pedophile member tactics. Defensive pedophile communication data includes electronic communications among pedophiles with convictions and/or admitted activities and attraction to minors despite severe penalties, including overt/covert threats against countries and persons illegalizing child pornography and persons engaged in prosecution or anti-pedophile activism.

PREText implements a statistical model of these different communication types, providing a score based on an empirically-derived threshold. ALIAS PREText's objective, quantitative, statistically- validated scoring can be used to develop techniques and training in pedophilic cybercrime investigations, to provide cloaking for investigators, and to present scientific evidence to judges and juries about communications which are, fortunately, unlikely to have been experienced firsthand by the triers of fact.

Pedophilic Communications, Forensic Linguistics, Predatory Grooming

 ** Presenting Author*