I16 Deception Leaves a Linguistic Trace: Assessment of Lying Using Computational Discourse Analysis

Robert K. Welsh, PhD, Kevin Reimer, PhD, David Arney, BA,*
Wendy Chan, BA, Lauren Stevenson, BA, Joshua Morgan, BA, and Rhea Holler, BA, Department of Graduate Psychology, Azusa Pacific University, 901 East Alosta Avenue, PO Box 7000, Azusa, CA 91702

After attending this presentation, attendees will understand the basics of computational linguistics and its application as new methodological approach for research on deception.

This presentation will impact the forensic community and/or humanity by demonstrating a new paradigm for future research on deception.

This paper presents a new methodology for the assessment of deception through computational discourse analysis. Preliminary data are presented for three studies where a computational model (Latent Semantic Analysis – LSA) was able to significantly detect group differences in college students who provided brief language samples when instructed to tell the truth or lie.

The common thread that runs through all methodologies that seek to detect deception is language. Arguably observance of behavior is a central component to one who is malingering, but the forensic examiner still ques- tions the examinee about his/her behavior and seeks an interpretation of the behavior. Language is a fundamental feature of psychometric assessment, polygraph and integrity testing and clinical interviews designed to detect deception.

Latent Semantic Analysis (LSA) is an unsupervised method of computational linguistics designed to statistically represent the contextual- usage meaning of words (Landauer, Foltz & Laham, 1998). LSA was developed at Bellcore Laboratories and refined at the University of Colorado, Boulder. Conceptually, LSA uses a matrix-driven decomposition technique, much like factor analysis, called singular value decomposition (SVD). In LSA words, sentences and paragraphs are assigned a vector as a knowledge representation estimate in high dimensional space based on cosine values ranging form -1.0 to +1.0. Cosine vectors represent the similarity and dissimilarity that are interpreted as meaning associations. High scores represent consistency and coherence of knowledge representations.

One strength of LSA that is clinically useful in the detection of deception is its ability to detect internal consistency and coherence of knowledge representation. Knowledge representation is central to the task of accurately reporting one's history. A reliable self-report requires that one has direct personal experience with the history being reported. Moreover, the author's propose that internal coherence of knowledge representation can only happen when an examinee is telling the truth. When evaluating whether or not an individual is accurately representing themselves, the forensic examiner pays close attention to the coherence, consistency and amount of convincing detail given in the examinees personal account. When inconsistency is discovered in an examinees narrative account, the forensic examiner uses the inconsistent information as data to strengthen the case that the examinee has given an unreliable self-report. However, for individuals who are adept at deception, it is often not possible, even through extended interviews, to "catch" an examinee in a lie.

The strength of LSA as an unsupervised computational linguistic model is that it mathematically approximates the same features found in human representations of meaning based on past experience. Through an individual's narrative, LSA will be able to approximate the extent to which the account reflects induction and representation based on lived experience.

LSA does not utilize word order or syntax to extract meaning repre- sentations from the narrative. Rather, it is able to use comparisons between words, sentences and paragraphs to extract the meaning from the whole at a deeper or "latent" level. Accordingly, LSA is able to determine how word choice represents meaning (Landauer, Fultz & Laham, 1998).

In this paper, the presenters use LSA to detect inconsistencies or lack of coherence of knowledge representation. Discussion of computational linguistics as a research and clinical tool will follow the presentation of the experimental data.

Deception, Malingering, Computational Linguistics