

Preliminary Report
Shane Todd Death Investigation of Alleged Suicide Notes
Carole E. Chaski PhD

I. Engagement and Executive Summary

1. I have been retained as an expert by Mary Todd in the matter of the death investigation of her son Todd Shane. In particular, I have been asked to assess documents of questioned status as suicide notes and of questioned authorship. I have been retained (1) to determine if the questioned documents are actually suicide notes and (2) to analyze the authorship of the questioned texts in relation to known writing samples of Mr. Todd. This is a preliminary report because I have completed the suicide note assessment but I have not completed the authorship analysis.

I have used a standard method for determining the status of a text as a suicide note, the ALIAS SNARE method. This method has an accuracy rate of 86% for the length of the texts I have received. The method has been developed and tested on over 1,000 documents including real suicide notes and real control texts. The method was developed independent of any litigation on ground-truth data and these results have been published in academic, peer-reviewed conference proceedings and is being included in police training in the United States by an independent police trainer.

Regarding the authorship analysis, I have completed a qualitative assessment of the native language of the questioned texts. I am preparing to use a standard method for authorship identification, the ALIAS SynAID method; this method has an accuracy rate of 94-95% when it has been tested on ground-truth data independent of any litigation and these results have been published in academic, peer-reviewed journals, conference proceedings and books.

My main findings are:

- (a) the alleged suicide note is not a real suicide note;
- (b) the fragment from the hard drive alleged to be a suicide note is not a real suicide note;
- (c) known writing of Shane Todd in the months before his death do not contain any indications of real suicide notes or suicidal language;
- (d) the language of the alleged suicide note is non-native English, so that the authorship of the alleged suicide note is not Shane Todd, a native speaker of English.

I have personal knowledge of the matters set forth in this report and, if called as a witness, I could and would testify to these same matters.

II. Credentials

2. I earned my A.B magna cum laude in Ancient Greek and English from Bryn Mawr College, my M.Ed. in Psychology of Reading from the University of Delaware, and my M.A. and Ph.D. in Linguistics from Brown University.

3. From 1987 to 1994, I taught syntax, psycholinguistics of literacy and computational linguistics at the University of South Carolina-Columbia and North Carolina State University. I taught both graduate and undergraduate students and supervised graduate research, as an assistant professor.

4. I then held a Visiting Research Fellowship at the National Institute of Justice, U.S. Department of Justice from 1995 to 1998. At the request of NIJ, I founded the Institute for Linguistic Evidence, Inc. in 1998. The institute is a non-profit research organization, of which I am Executive Director. I maintain research relationships with scholars in Ireland, Greece, Pakistan, Korea, Switzerland, Spain, Canada and Britain as well as the United States, working specifically on forensic linguistic evidence.

5. I am the president of ALIAS Technology LLC which provides consulting and software for forensic linguistic analysis and computational linguistics. I have been consulted by prosecuting, plaintiff and defense attorneys, law enforcement and private individuals on forensic linguistic evidence since 1992, and with corporations on computational linguistics projects since 2006.

6. I serve on the Advisory Board for the Linguistics Department of Montclair State University (Montclair, New Jersey) and have been an adjunct professor of linguistics at the University of Delaware (Georgetown campus). I am an adjunct professor of forensic science at George Washington University (Washington, DC) and the University of Nebraska (Lincoln, NE). I taught at the Linguistic Society of America Linguistics Institute at the University of Michigan (Ann Arbor, Michigan) in the summer of 2013.

7. I am a member of the American Academy of Forensic Sciences (fellow in engineering sciences section), Linguistic Society of America, the Association for Computing Machinery (information retrieval section), American Bar Association (criminal justice section, science and technology section, associate member), and The Association for Linguistic Evidence.

8. I have published ten peer-reviewed articles on forensic linguistic methods in academic and professional journals and contributed seven book chapters on forensic linguistics including two textbooks, a science encyclopedia and other scholarly books. I have presented my research at national and international conferences with peer-review of abstracts over 50 times. I have a patent pending on my methods for forensic authorship identification. I am the Editor of *LESLI: Linguistic Evidence in Security, Law and Intelligence*, and an associate editor of *The Linguistics Journal*, *Language and Law* and *the International Journal of Law, Language and Discourse*; I have reviewed articles for these journals as well as *the International Journal of Speech, Language and Law*, *Law and Society*, *Foundations of Information Retrieval*, *International Journal of Digital Evidence*, and *Cortex*.

9. Exhibit A is my curriculum vitae. A list of cases in which I have provided sworn testimony is included in my curriculum vitae. Currently ongoing cases are not listed in order to protect client privilege.

10. I am being compensated at the rate of \$100 per hour in this case for the authorship analysis.

III. Data and Testing Results

11. I received from Mary Todd the following data:

Alleged Suicide Notes ~775 words
Fragment from Hard Drive ~225 words
Emails from Shane Todd to family ~1700 words

12. I used ALIAS SNARE to assess each of these documents for suicidal language, i.e. the language patterns found in real suicide notes. SNARE classifies a document as a real suicide note or a control text based on linguistic and statistical analysis.

13. The Alleged Suicide Notes were not classified as a suicide note; instead these notes were classified as a control text.

14. The Fragment from Hard Drive was not classified as a suicide note; instead this text was classified as a control text.

15. Emails from Shane Todd to family were not classified as a suicide note; instead these notes were classified as control texts.

16. I analyzed the Alleged Suicide Notes as a linguist for native/non-native English. The Alleged Suicide Notes contain indications of non-native English:

non-Native English: missing determiner 'a'
about how I was loving son, grandson

native English
about how I was **a** loving son, grandson

non-Native English: missing modal
If you allow me to, I ask that my family decides

native English
If you allow me to, I ask that my family **could/can** decide

These non-Native indicators are not explainable as typos or sloppy typing because the rest of the document is well-typed. Further, culturally, the focus on family and success is an indicator of a non-American.

IV. Additional Work for Final Report

17. For a complete suicide note assessment, SNARE-Quals is used after SNARE. SNARE-Quals gathers real suicide notes from the SNARE database to be compared to the questioned suicide note.

18. For a complete authorship assessment, ALIAS SynAID is used; this method requires at least two possible suspect authors. Given the proximity of Shirley to the case, both Shane Todd and Shirley should be tested as potential authors of the alleged suicide notes. Other potential authors might include any investigators who had access to Shane Todd's computer.

ALIAS SNARE Report



ALIAS: Automated Linguistic Identification and Assessment System _SNARE Module 2008

CaseID Carole E Chaski PhD-ShaneTodd-SNARE-770

CaseQD

UserName Carole E Chaski PhD

ReportDate 3/29/2014

ReportTime 8:48:13 AM

Statistical Model Decision CONTROL TEXT

ALIAS SNARE runs one test which compares the Case Questioned Document to a database of real suicide notes against a database of control documents which have similar characteristics to suicide notes. The control document database includes apology letters, love letters, trauma narratives, essays about anger, angry letters, complaint letters, simulated and real threats to known targets, business letters related to insurance and a small number of simulated and hoax suicide notes. The current databases include hundreds of real suicide notes and hundreds of control documents, for a total of a little more than one thousand documents. A statistical model for classifying suicide notes is built on these databases.

ALIAS SNARE assesses the Case Questioned Document based on the classification obtained from applying the current statistical model to the Case Questioned Document. Important linguistic features in this statistical assessment are: afterlife, blaming, hope, features of business correspondence, emotional states, conditionality, verb temporality and pronoun ratios.

The statistical model developed from these databases using Linear Discriminant Function Analysis. ALIAS SNARE is able to distinguish real suicide notes from the control documents at 88.6% accuracy.

Only 7% of the control documents are misclassified as real suicide notes, while almost 19% of the real suicide notes are misclassified as control documents, for an total error rate of 12%.

The Suicide Note Assessment Rating Chart depicts the current statistical model with the assessment for the Case Questioned Document written as "Suicide Note" or "Control Text." In the chart below, the statistical model's error zones are depicted by the lighter gray shades, while the model's correct rate for real suicide notes is depicted by the dark gray shade, and the model's correct rate for control documents is depicted by the blue shade.

In addition to ALIAS SNARE's classification of this Case Questioned Document, investigators should take into consideration the error zones and the important linguistic features as they occur or do not occur in the Case Questioned Document. Investigators should conduct their own assessment based on reading the text for the important linguistic features and other features from their own experience in suicide note assessment.

Suicide Note Assessment Rating



ALIAS SNARE Report



ALIAS: Automated Linguistic Identification and Assessment System _SNARE Module 2008

CaseID Carole E Chaski PhD-ShaneTodd-SNARE-769

CaseQD

UserName Carole E Chaski PhD

ReportDate 3/29/2014

ReportTime 8:48:12 AM

Statistical Model Decision CONTROL TEXT

ALIAS SNARE runs one test which compares the Case Questioned Document to a database of real suicide notes against a database of control documents which have similar characteristics to suicide notes. The control document database includes apology letters, love letters, trauma narratives, essays about anger, angry letters, complaint letters, simulated and real threats to known targets, business letters related to insurance and a small number of simulated and hoax suicide notes. The current databases include hundreds of real suicide notes and hundreds of control documents, for a total of a little more than one thousand documents. A statistical model for classifying suicide notes is built on these databases.

ALIAS SNARE assesses the Case Questioned Document based on the classification obtained from applying the current statistical model to the Case Questioned Document. Important linguistic features in this statistical assessment are: afterlife, blaming, hope, features of business correspondence, emotional states, conditionality, verb temporality and pronoun ratios.

The statistical model developed from these databases using Linear Discriminant Function Analysis. ALIAS SNARE is able to distinguish real suicide notes from the control documents at 88.6% accuracy.

Only 7% of the control documents are misclassified as real suicide notes, while almost 19% of the real suicide notes are misclassified as control documents, for an total error rate of 12%.

The Suicide Note Assessment Rating Chart depicts the current statistical model with the assessment for the Case Questioned Document written as "Suicide Note" or "Control Text." In the chart below, the statistical model's error zones are depicted by the lighter gray shades, while the model's correct rate for real suicide notes is depicted by the dark gray shade, and the model's correct rate for control documents is depicted by the blue shade.

In addition to ALIAS SNARE's classification of this Case Questioned Document, investigators should take into consideration the error zones and the important linguistic features as they occur or do not occur in the Case Questioned Document. Investigators should conduct their own assessment based on reading the text for the important linguistic features and other features from their own experience in suicide note assessment.

Suicide Note Assessment Rating



ALIAS SNARE Report



ALIAS: Automated Linguistic Identification and Assessment System _SNARE Module 2008

CaseID Carole E Chaski PhD-ShaneTodd-SNARE-770

CaseQD

UserName Carole E Chaski PhD

ReportDate 3/29/2014

ReportTime 8:48:13 AM

Statistical Model Decision CONTROL TEXT

ALIAS SNARE runs one test which compares the Case Questioned Document to a database of real suicide notes against a database of control documents which have similar characteristics to suicide notes. The control document database includes apology letters, love letters, trauma narratives, essays about anger, angry letters, complaint letters, simulated and real threats to known targets, business letters related to insurance and a small number of simulated and hoax suicide notes. The current databases include hundreds of real suicide notes and hundreds of control documents, for a total of a little more than one thousand documents. A statistical model for classifying suicide notes is built on these databases.

ALIAS SNARE assesses the Case Questioned Document based on the classification obtained from applying the current statistical model to the Case Questioned Document. Important linguistic features in this statistical assessment are: afterlife, blaming, hope, features of business correspondence, emotional states, conditionality, verb temporality and pronoun ratios.

The statistical model developed from these databases using Linear Discriminant Function Analysis. ALIAS SNARE is able to distinguish real suicide notes from the control documents at 88.6% accuracy.

Only 7% of the control documents are misclassified as real suicide notes, while almost 19% of the real suicide notes are misclassified as control documents, for an total error rate of 12%.

The Suicide Note Assessment Rating Chart depicts the current statistical model with the assessment for the Case Questioned Document written as "Suicide Note" or "Control Text." In the chart below, the statistical model's error zones are depicted by the lighter gray shades, while the model's correct rate for real suicide notes is depicted by the dark gray shade, and the model's correct rate for control documents is depicted by the blue shade.

In addition to ALIAS SNARE's classification of this Case Questioned Document, investigators should take into consideration the error zones and the important linguistic features as they occur or do not occur in the Case Questioned Document. Investigators should conduct their own assessment based on reading the text for the important linguistic features and other features from their own experience in suicide note assessment.

Suicide Note Assessment Rating

