



128 Is This a Real Suicide Note? Authentication Using Statistical Classifiers and Computational Linguistics

Carole E. Chaski, PhD*, *ALIAS Technology, LLC, Institute for Linguistic Evidence, 25100 Trinity Drive, Georgetown, DE 19947*; and Denise D. Huddle, BS, *Huddle & D'Abre, 900 NE Loop 410, Suite D406, San Antonio, TX 78209*

After attending this presentation, attendees will learn a method of suicide note authentication using both a quantitative, validated software tool Suicide Note Assessment Research (SNARE) and qualitative assessment based on database extraction. Attendees will learn how this method has been used in actual cases of death investigation and how it can be used in psychological assessment. Students will also learn how validation testing is performed for behavioral and linguistic datasets.

This presentation will impact the forensic science community by showing why a suicide note assessment is essential for investigating some deaths which could be classified as a homicide or suicide. The method presented here using SNARE illustrates validation testing for behavioral and linguistic datasets, and thus also advances the methodology of forensic science.

Identifying suicide notes is an extremely difficult task. Part of this difficulty is common to many identification tasks with low accuracy rates. First, if the object to be identified is rare or experience with the object to be identified is low, identification rates are typically low because the identifiers do not have a well-founded grasp of the identifying features. Second, if the object to be identified has low internal consistency, because the object has a wide range of class characteristics, then identification accuracy is predictably low, because the object can be mistaken for so many other types of similar objects.

Suicide notes are difficult to identify accurately because, first, they are rare in most investigators' experience, and second, they are not highly consistent or stereotypical. Suicide notes are found in only the minority of actual suicides (it is estimated that notes only occur in 10 to 15% of suicides).¹ Suicide notes contain elements of other text types, such as apologies and love letters, which can make them easily confusable with other text types. Suicide notes as a whole are internally inconsistent –no one feature set is present in all. It is not surprising then that psychiatrists and psychologists are claimed to have an accuracy rate for identifying texts as suicide notes or not at ranges from 50% to 70% as even this group of professionals may not have consistent experience with real suicide notes, and the experience is informed by a group of documents which are not necessarily similar to each other.² Bennell's work on police decision-making mentions the possibility of a quantitative analysis for verifying suicide notes.³

SNARE is a computational and quantitative tool for identifying and classifying suicide notes. SNARE currently has obtained an accuracy rate of 80% on a dataset of 400+ real suicide notes and 500 control documents. When the real suicide note data is limited to brief notes (45 words or less) the accuracy rate increases to 86%. The dataset has been vetted. The statistical classifier is a linear discriminant function analysis using leave-one-out cross-validation. The reported accuracy rates are the average of true positives for the real suicide note class and the control document class.

The quantitative method enables the analyst (death investigator, psychiatrist, or psychologist) to start from an objective point in the investigation, since the software analyzes the questioned text without input from the analyst. But given an error rate from 14 to 20%, the analyst should then use SNARE to collect a pool of comparative documents from the real suicide note dataset. By comparing the questioned document to real suicide notes with similar characteristics, the analyst can determine whether SNARE has made a classification error or not.

Methodologically, the most interesting aspect of this second step is the requirement that the analyst seek to disprove the SNARE classification. The two-step analysis based on a quantitative classification followed by qualitative assessment allows for the application of analyst knowledge, preferably in line with Leenaars' work.⁴⁻⁷

References:

1. Holmes, Ronald M. and Holmes, Stephen T. (2005). *Suicide: Theory, Practice and Investigation*. New York: Sage.
2. Pestian, John P., Matykiewicz, Pawel, Grupp-Phelan, Jacqueline, Lavanier, Sarah Arszman, Combs, Jennifer, and Kowatch, Robert. (2008). "Using Natural Language Processing to Classify Suicide Notes." *BioNLP 2008: Current Trends in Biomedical Natural Language Processing*, pp.96-97. Association for Computational Linguistics.
3. Bennell, Craig. (2005). "Improving Police Decision Making: General Principles and Practical Applications of Receiver Operating Characteristic Analysis." *Applied Cognitive Psychology* 19: 1157- 1175. Published online 13 September 2005 in Wiley InterScience www.interscience.wiley.com
4. Leenaars, Anton and Balance, W. 1984. "A logical empirical approach to the study of suicide notes." *Canadian Journal of Behavioral Sciences* 16:248-56.
5. Leenaars, Anton. 1988. *Suicide Notes*. New York: Human Sciences Press.
6. Leenaars, Anton. 1996. "Suicide: a multidimensional malaise." *Suicide Life Threatening Behavior* 26:221-36.
7. Leenaars, Anton, Park, Ben, Collins, Peter I., Wenckstern, Susanne, and Leenaars, Lindsey. 2010. "Martyrs' Last Letters: Are they the same as suicide notes?" *Journal of Forensic Sciences* 55:660-668.

Suicide Note, Computational Linguistics, Validation Testing