1035-Z1-646 **Alan Safer*** (asafer@csulb.edu), Department of Mathematics and Statistics, California State University, Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840-1001, and **Eric Chuk** (echuk@csulb.edu). *Authorship and Statistical Profiling.* Preliminary report.

A writer's style or voice is something difficult to define traditionally. Descriptions are usually impressionistic, relying on adjectives that give a sense of overall tone or prominent word choices. With the advent of data and text mining methods, an inductive, quantitative approach can instead be used, turning this into a classification task. Quantitative authorship attribution is accomplished by a variety of statistical means, including chi-squared tests and decision trees. Computerized statistical analysis allows a variety of text measurements, including word counts, average sentence length, and many others. Once taken, these measurements can be tested to determine which features of a text most successfully identify its writer. Applications include plagiarism detection and other identification in documents with unknown or disputed authors. (Received September 12, 2007)