2013

Introduction to the Symposium

Lawrence M. Solan, Ph.D., J.D.

AUTHORSHIP ATTRIBUTION WORKSHOP

PREFACE

Lawrence M. Solan*

In October 2012, Brooklyn Law School's Center for the Study of Law, Language and Cognition held a two-day workshop on Forensic Authorship Attribution. This volume contains its proceedings. The workshop, sponsored by the National Science Foundation, to which Brooklyn Law School expresses its gratitude,[1] brought together leading scholars from around the world who approach the question of authorship attribution from disparate, and seemingly incompatible, perspectives. Represented among the articles that appear in this volume are works based on the algorithms of computer scientists and computational linguists (Argamon & Kopell; Chaski; Juola; Koppel, Schler & Argamon; Stamatatos; and Vogel), works by linguists who evaluate and compare stylistic regularity and nuance, often on a case-by-case basis (Coulthard; Grant), and research comparing the extent to which linguistic nuance results from regional differences between one speech community and another, and the extent to which it is based within the individual (Turell and Gavaldà). Other prominent linguists—Ronald Butters and Edward Finegan—moderated workshop sessions.

Despite these differences in approach, what emerged from the workshop and is reflected in the published articles is a recognition that those who work algorithmically can improve their models by incorporating into them some of the insights of those who work with stylistic markers, and those whose work is less computational can develop quantitative techniques to

improve the reliability of their conclusions. In my own essay in this volume, I explore the advantages and disadvantages of the various techniques employed in this volume, commenting on how the legal system tends to react to them. While it might be too ambitious to predict that the various methods will converge, there is already reason for optimism that the insights of the various approaches will influence those of the others, creating a field with a healthy combination of cooperation and competition. Thus, in this volume, we see the expanded use of linguistic features by the computational researchers, while at the same time we see efforts by stylistic researchers to introduce statistical modeling into stylistic authorship analysis.

A quick glance at the table of contents shows a lot of contributors with Ph.D.s and very few with law degrees, an unusual array for a law journal publication. Nonetheless, the legal academic community is very much present in this volume. In addition to my own essay, comments by two prominent law professors who specialize in scientific evidence—Edward Cheng and Jonathan Koehler—consider the legal community's likely response to the advances in authorship attribution described by the linguists.[2] Moreover, participating in the workshop were two statistical "consultants" (Stephen Fienberg of Carnegie Mellon University and Robert Carpenter of Columbia University).

This interaction between the scientific community and the evidence scholars was one of the workshop's main goals. Forensic identification sciences have been under severe attack as inadequately grounded in science over the past decade—largely for good reason. Just as the scientists had a lot to learn from the reactions of the legal scholars to their work, we believed that the legal scholars could benefit from seeing in action a relatively young forensic science that takes itself seriously as science. While the evidence literature decries the absence of concern over the rate of error in one forensic science after another, an annual workshop on authorship and plagiarism identification actually requires that algorithms presented be subjected to a

---

[2] Also present at the workshop were D. Michael Risinger and Michael Saks, two additional prominent scientific evidence scholars.

proficiency test before a paper is accepted for presentation at the conference.[3]

At the same time, though, some important insights that may assist a trier of fact in a case concerning authorship, such as multiple possible authors making the same kind of grammatical error in known writings (as illustrated by Malcolm Coulthard's piece in this volume), are not yet subject to mathematical analysis, creating a dilemma for the legal system. The legal commentators observed and comment on both the promising progress of the field and the challenges it still faces.

I end on a very sad note. One of the authors, María Teresa Turell, passed away on April 24, 2013, just before this volume went to press. Maite, as everyone knew her, was a Professor of English at Universitat Pompeu Fabra in Barcelona, where she directed the Forensic Linguistics Laboratory. She was an important figure in the field, devoted to bringing quantitative rigor to stylistic insight, one of the themes of the workshop. More importantly to those of us who knew Maite, her intellectual toughness was matched with a loving and generous character that will remain with us for a long, long time to come.

---

[3] The conference is known as the "PAN/CLEF" conference. *See* PAN WORKSHOP & COMPETITION, http://pan.webis.de/ (last visited May 3, 2013). As of the date of this writing (May 2, 2013), 107 teams from forty-two countries have registered for the next PAN/CLEF conference, which will be held in Valencia, Spain in September 2013.