



### E26 Facebook® and the Faceless: Authorship in an Electronic Society

Carole E. Chaski, PhD\*, Institute for Linguistic Evidence, ALIAS Technology, Limited Liability Company, 25100 Trinity Drive, Georgetown, DE 19947; and Alice B. Lustre, JD, California Department of Justice, 455 Golden Gate Avenue, Suite 11000, San Francisco, CA 94102-3658

After attending this presentation, attendees will learn about the role of authorship issues involving blogs, emails, Facebook®, and other documents in both high profile and low press cases. Attendees will attain information regarding two different methods of language-based authorship identification forensic stylistics and computational linguistics. Attendees will also gain information on validation test results for these methods and the linguistic community's reaction to these methods since the 2011 *Ceglia v. Zuberberg* Facebook-ownership case.

This presentation will impact the forensic science community by being able to accurately assess authorship identification methods by both the *Frye* and *Daubert* criteria, giving them an additional tool for handling electronic evidence.

For most of the general public, many attorneys and most crime laboratories, the term "forensic linguistics" is so unfamiliar that it evokes images of argumentative linguists or the speech of dead people. But anyone who has received an email from a Nigerian scam artist, a vixen selling herbal compounds, has read blog posts that scorch the computer screen, or has received digital applications to fill out over CAPTCHA boxes might also wonder about the reality of email addresses, screennames, and digital documents. Authorship in an electronic society provides pseudonymity, anonymity, and blithe shape-shifting.

Some recent high profile cases which have involved an authorship issue include the JonBenet Ramsey homicide, the 2004 precedent-setting case of *Cahill vs. Doe* (Shaeffer), the 2008 *Best Western International vs. John Doe* (Dial, Furber et. al.), settled for over \$2M, and now the latest Facebook® suit, *Ceglia vs. Zuckerberg*. Cases which have little or no press, but are just as pressing for societal concerns and forensic science include, from 2011, the Masters homicide and the Isaacs custody-related trials which focused, in part, on a pseudonymous Facebook post.

There are currently two main methods for author identification: forensic stylistics and computational linguistics. Forensic stylistics has practitioners in the United States, Britain, Australia, and Spain. The computational linguistics approach has practitioners and researchers in the United States, Israel, Germany, France, Greece, Spain, and Ireland. These two methods differ radically in their procedures, data requirements, implementation tools, and most importantly, in the validation testing results. The differences between authorship identification in forensic stylistics and computational linguistics have been documented by Crystal<sup>4</sup>, Chaski<sup>2-6</sup>, Koppel and Schler<sup>7</sup>, Nunberg<sup>8</sup> and now in relation to the latest Facebook® case, Zimmer<sup>9,10</sup> and Liberman<sup>11,12</sup>. In this presentation, we review these differences by references to the specific cases mentioned above will be reviewed, highlighting the core difference of validation testing. While computational linguists employ validation testing as a primary means of software and method development, forensic stylistics practitioners have never offered any test results or error rates.

The *Frye* standard of general acceptability has been subsumed by the *Daubert* criteria, not abandoned. The relevant scientific community for both authorship identification methods, forensic stylistics, and computational linguistics, is linguistics. Linguists who function independently of the forensic world and some linguists who do forensic work are not in agreement with forensic stylistics as a method or as a representation of linguistics. The forensic science community, including crime laboratories and the judicial system, should begin to pay heed to what linguists are saying about these methods. Attorneys are the key personnel for presenting the documentation reported to the judges and juries who face issues of authorship in an electronic society.

#### References:

1. Crystal, David. 1995. Review of *Forensic Stylistics*, by Gerald A McMenamin. *Language*, 71, No. 2, pp. 381-5.
2. Chaski, Carole E. 1997. "Who Wrote It? Steps Toward a Science of Authorship Identification." *National Institute of Justice Journal*. September 1997. Also available through National Criminal Justice Reference Service: <http://www.ncjrs.org> NCJ 184604.
3. Chaski, Carole E. 2001. "Empirical Evaluation of Language-Based Author Identification Techniques." *Forensic Linguistics: International Journal of Speech, Language and Law* Volume 8:1. pp. 1-64. June 2001.
4. Chaski, C. E. 2010. Empirically testing the uniqueness of aggregated stylemarkers. *International Association of Forensic Linguists/Language and Law 8<sup>th</sup> Biennial Conference*, July 12-15, 2007 University of Washington, Seattle, WA.
5. Chaski, Carole E. 2010. "Linguistics as a Forensic Science: The Case of Author Identification." In Susan Behrens and Judith A. Parker, editors. *Language in the Real World*. Routledge.
6. Chaski, Carole E. 2008. "The Computational-Linguistic Approach to Forensic Authorship Attribution." In Frances Olsen, Alexander Lorz, and Dieter Stein, editors. *Law and Language: Theory and Practice*. Düsseldorf University Press.



## Jurisprudence Section - 2012

---

- <sup>7</sup> Koppel, M. and Schler, J. 2003. Exploiting stylistic idiosyncrasies for authorship attribution. In *Proceedings of IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis*, Acapulco, Mexico, 2003. Available at: <http://citeseer.ist.psu.edu/article/koppel03exploiting.html>
- <sup>8</sup> Nunberg, G. 2005. Statement of Geoffrey Nunberg in re Hargett v Morell.
- <sup>9</sup> Zimmer, Ben. 2011a. "Decoding your E-Mail Personality." *The New York Times Sunday Review*. July 23, 2011. Also available online at: [http://www.nytimes.com/2011/07/24/opinion/sunday/24gray.html?\\_r=1](http://www.nytimes.com/2011/07/24/opinion/sunday/24gray.html?_r=1)
- <sup>10</sup> Zimmer, Ben. 2011b. "Does E-mail Have Fingerprints?" Visual Thesaurus Word Routes, July 28, 2011. Available at: <http://www.visualthesaurus.com/cm/wordroutes/2928/>
- <sup>11</sup> Liberman, Mark. 2011a. "High-stakes Forensic Linguistics." *Language Log*. July 25, 2011. <http://languagelog.ldc.upenn.edu/nll/?p=3309>
- <sup>12</sup> Liberman, Mark. 2011b. "Authors vs. Speakers: A Tale of Two Subfields." *Language Log*. July 27, 2011. <http://languagelog.ldc.upenn.edu/nll/?p=3317>

**Authorship Identification, Computational Linguistic, Frye and Daubert Criteria**