The linguistic fingerprint

(Notiz: letztes Update dieser Seite war 2007)

Referentin: Daniela Berger

The linguistic fingerprint -- Silver bullet or mere myth?

Abstract

In the wake of crime science shows like CSI and high profile criminal cases like the JonBenet Ramsey murder, the field of forensic linguistics has come to the attention of the general public. Today many laypersons know the term "linguistic fingerprint" and they have certain expectations about what it implies.

But these expectations are largely unfounded.

The lack of real knowledge about this technique are largely due to its ill-chosen "nickname". The term "linguistic *fingerprint*" puts it into the neighborhood of the "actual", i.e. dactyloscopic fingerprint and the "genetic fingerprint". But this is misleading.
Both in fingerprinting and in DNA analysis there are procedures for collecting samples, for analysing them, for comparing them to samples taken from the suspect(s) and for interpreting the results. These procedures are known for their reliability today, but it took years of research to get to this point. Still, today we are at a point where a fingerprint left at a crime scene can safely be used to to confirm the guilt of a suspect.

The use of the fingerprint metaphor in the context of forensic linguistics and authorship attribution implies that research in this field has reached the same maturity. In reality, some promising results have been found, but so far the linguistic community was not able to prove that a certain set of markers can be reliably used to confirm a person's authorship of a text. Many questions are still left to be answered.

Overview

In this talk I will give definitions of the relevant terms and concepts.

Then I will give an overview over the different fields of interest that are subsumed under "forensic linguistics". From these I chose authorship attribution as the target of a state-of-the-art report.

I will present several interesting approaches, demonstrate their application with the help of real life examples where possible, and discuss their merits and limitations. The main focus here will be

- a) on written texts such as blog entries / comments and forum articles and
- b) on the source code of software such as viruses

I will show that forensic linguistics procedures are far from having the same accuracy as fingerprinting procedures, but that - at best - they can be used to prove that the same person did or did not write a set of texts. And if that is not possible they can still be used to gather other, more general clues about the author, perhaps about his gender or his education.

For the time being this does not make the linguistic fingerprint the proverbial silver bullet, but rather it makes forensic linguistics one valuable tool in the criminological toolbox.

Revision #1
Created 2022-03-10 16:47:57 UTC by merline
Updated 2022-03-19 17:15:15 UTC by merline