

## The Grammar of the Content Plane as a Style Marker in Forensic Authorship Attribution

**Authors :** Dayane de Almeida

**Abstract :** This work aims at presenting a study that demonstrates the usability of categories of analysis from Discourse Semiotics - also known as Greimassian Semiotics in authorship cases in forensic contexts. It is necessary to know if the categories examined in semiotic analysis (the 'grammar' of the content plane) can distinguish authors. Thus, a study with 4 sets of texts from a corpus of 'not on demand' written samples (those texts differ in formality degree, purpose, addressees, themes, etc.) was performed. Each author contributed with 20 texts, separated into 2 groups of 10 (Author1A, Author1B, and so on). The hypothesis was that texts from a single author were semiotically more similar to each other than texts from different authors. The assumptions and issues that led to this idea are as follows: -The features analyzed in authorship studies mostly relate to the expression plane: they are manifested on the 'surface' of texts. If language is both expression and content, content would also have to be considered for more accurate results. Style is present in both planes. -Semiotics postulates the content plane is structured in a 'grammar' that underlies expression, and that presents different levels of abstraction. This 'grammar' would be a style marker. -Sociolinguistics demonstrates intra-speaker variation: an individual employs different linguistic uses in different situations. Then, how to determine if someone is the author of several texts, distinct in nature (as it is the case in most forensic sets), when it is known intra-speaker variation is dependent on so many factors?-The idea is that the more abstract the level in the content plane, the lower the intra-speaker variation, because there will be a greater chance for the author to choose the same thing. If two authors recurrently chose the same options, differently from one another, it means each one's option has discriminatory power. -Size is another issue for various attribution methods. Since most texts in real forensic settings are short, methods relying only on the expression plane tend to fail. The analysis of the content plane as proposed by greimassian semiotics would be less size-dependable. -The semiotic analysis was performed using the software Corpus Tool, generating tags to allow the counting of data. Then, similarities and differences were quantitatively measured, through the application of the Jaccard coefficient (a statistical measure that compares the similarities and differences between samples). The results showed the hypothesis was confirmed and, hence, the grammatical categories of the content plane may successfully be used in questioned authorship scenarios.

**Keywords :** authorship attribution, content plane, forensic linguistics, greimassian semiotics, intraspeaker variation, style

**Conference Title :** ICAL 2016 : International Conference on Applied Linguistics

**Conference Location :** Paris, France

**Conference Dates :** November 21-22, 2016