Journal of Research and Innovation for Sustainable Society (JRISS)
Volume 8, Issue 1, 2026
ISSN: 2668-0416
Thoth Publishing House

DOI: 10.33727/JRISS.2026.1.11:110-114

Intelligence Analysis: The Psycholinguistic Signature of the Author in Open-Source Investigations

Manuel Gabriel Vigariu[1] and Constantin-Alexandru Marin[2]

[1]Psychologist, independent researcher, and legal advisor, Bucharest, Romania, ORCID: 0009-0004-4470-746X
[2]Communication Studies Department, National University of Political Studies and Public Administration, Bucharest, Romania, ORCID: 0009-0007-7194-3910

E-mail: vigariu.manuel@gmail.com

Abstract. In an information-saturated environment, where open-source information is abundant, the real challenge for the security of the information ecosystem is no longer access to data, but establishing its anonymous authorship. This paper advances the thesis that, beyond the walls of digital anonymity, the human mind betrays its presence through the involuntary nuances of language. By examining the concept of "psycholinguistic signature", the article highlights a new form of identity, showing that deep cognitive patterns - shaped by lexical choice, syntactic complexity, temporal flow, and emotional markers - exhibit remarkable stability. Even when there are deliberate attempts to manipulate the communicative register or to conceal identity, these linguistic traces remain constant, offering a robust tool for identifying and profiling authors. Psycholinguistic analysis thus emerges as an essential methodology for turning the digital echo into reliable behavioral evidence within the information ecosystem.
Keywords: *open-source*, *information ecosystem, digital anonymity, psycholinguistic analysis*

1. Introduction

In today's geopolitical landscape and contemporary cross-border space, dominated by an exogenous proliferation of data, open-source intelligence (OSINT) investigations have become a critical component of the intelligence cycle. Yet the current paradox is that, although access to information is virtually unlimited, the ability to attribute that information to a real source is increasingly eroded by digital anonymity. When classical investigative techniques - based on metadata, geolocation, or technical network analysis - are bypassed through sophisticated anonymization tools, intelligence analysis, both within intelligence services and among digital investigators, often runs into an "invisible wall". At this point, the traditional toolkit of the intelligence analyst, while technically robust, often reveals a vulnerability: the lack of instruments capable of penetrating the surface layer of text to reach the author's psychological profile.

To overcome this limitation, integrating psycholinguistic perspectives[1] becomes imperative. Psycholinguistics, as the discipline that studies the dynamic relationship between cognitive processes and language use, is not limited to a simple grammatical analysis; it provides the key to understanding how an individual thinks through how they write. Every lexical choice, every syntactic structure, and

[1] Heuer, R.J. (1999). *Psychology of Intelligence Analysis. Center for the Study of Intelligence*, Central Intelligence Agency.

every variation in discourse rhythm represents an imprint of cognitive functioning, one that is difficult to consciously disguise over the long term.

However, exploiting this resource cannot be reduced to the mere use of automated algorithms. The complexity of human nature requires nuanced interpretation, suggesting a subtle reconfiguration of analysis teams. The effectiveness of this approach depends on interdisciplinary collaboration in which the technical expertise of the intelligence analyst or digital investigations expert is complemented by highly specific psychological competence. This implies the need for a psychologist within analytic structures, one trained specifically in psycholinguistic dynamics and behavioral profiling. This specialist, whose contribution is essential yet not ostentatious, acts as a decoder of subtext, transforming raw linguistic data into operational psychological profiles. We argue that only through this synthesis between OSINT techniques and specialized psychological expertise can reliable author identification be ensured, along with a richer understanding of both what can be learned and who stands behind the text.

2. Psycholinguistics: the cognitive architecture of discourse

Psycholinguistics sits at the intersection of higher cognitive mechanisms and the formal structure of language, investigating how intentions, emotional states, and personality traits are encoded in linguistic utterances. In the context of psychological analysis and profiling, it is not treated as a purely descriptive discipline, but as an explanatory one, capable of decoding the "automatic processing" of information behind linguistic content. Reference authors like David Lieberman highlight the principle that language functions as a "projector" of mental state: the more an individual tries to consciously control the message they transmit, the more cognitive leaks appear in the subconscious structure of discourse. Psycholinguistic analysis thus becomes a tool for inferring latent intentions, going beyond mere semantic decoding.

A first analytical dimension is lexical selection. Vocabulary is not random; it reflects the author's attentional focus and emotional state. For instance, the excessive use of abstract nouns and stative verbs, at the expense of concrete action verbs, may indicate a tendency toward emotional distancing or an attempt to lend an appearance of legitimacy to unverifiable claims. By contrast, a lexicon marked by intense connotations, hyperbole, or frequent use of the first-person pronoun "I" (linguistic egocentrism) may signal narcissism, a need for affirmation, or, in the case of threats, heightened emotional engagement with the act[2].

A highly predictive indicator in behavioral profiling is grammatical voice. Shifts from active to passive or reflexive/impersonal constructions are often markers of cognitive dissonance or avoidance of responsibility (reduced agentivity). Consider the following example in the context of denying guilt: the statement "I made a mistake" (active voice) implies direct ownership of agency. By contrast, resorting to "Mistakes were made" (passive voice) or "There was a misunderstanding" (impersonal construction) serves as a psychological defense mechanism, with the author effectively erasing themselves from the equation of action[3]. In intelligence analysis, identifying such shifts in voice within the same narrative is a strong indicator of attempted manipulation or truth concealment.

Likewise, repetitiveness (linguistic iterativity) offers valuable clues about the sender's cognitive architecture. Here we do not mean intentional rhetorical repetition, but repetitions that function as blockages or mental anchors. Repeating certain keywords or phrases - especially under pressure or during moments of narrative transition - can signal fixation points in an obsessive thought scheme or difficulties processing contradictory information. For example, repeatedly returning to a particular detail that is irrelevant to the broader context (cognitive perseveration) may betray a specific anxiety tied to that detail, acting as a "blind spot" the analyst should exploit.

[2] Pennebaker, J.W. (2011). *The Secret Life of Pronouns: What Our Words Say About Us*, Bloomsbury Press.
[3] Newman, M.L. et all (2003). *Lying Words: Predicting Deception from Linguistic Styles. Personality and Social Psychology Bulletin 29(5),* SAGE Publications.

3. Psycholinguistic fingerprint: the stability and uniqueness of cognitive indicators

If psycholinguistics provides the theoretical framework needed to understand the mechanisms of language, the concept of a psycholinguistic fingerprint represents the practical applicability of that framework in source identification and attribution. Defined as a complex set of linguistic and cognitive markers that are relatively stable over time, the psycholinguistic fingerprint functions analogously to a digital fingerprint, but at the level of information processing and verbal expression[4]. In OSINT investigations, this fingerprint becomes the keystone for overcoming anonymity, enabling both intelligence analysts and expert psychologists to link disparate digital identities through a behavioral constant invisible to the untrained eye. The fundamental difference between conventional content analysis and an approach based on the psycholinguistic fingerprint is that the latter does not look for what is said, but for how the message is processed and constructed, revealing the individual uniqueness behind the text[5].

A central postulate in using the psycholinguistic fingerprint as an intelligence tool is the idea of cognitive inertia. Deep patterns of thought - how an individual structures an argument, orders ideas in space and time, or reacts to emotional stimuli - are crystallized during early language development and remain relatively resistant to later change. When an author attempts to falsify a text, for example by writing in a non-native language or by imitating an ungrammatical style to appear uneducated, they can readily focus on superficial aspects such as spelling or basic syntax. Yet the cognitive effort required to alter deeper patterns - such as the complexity of idea hierarchies or the underlying causal logic - is immense; thus, beneath the layer of deliberate masking, the true psycholinguistic fingerprint persists and can be detected by analyzing rhythm fluctuations and the consistency of logical structures.

This cognitive inertia manifests through specific dimensions which, taken together, compose the unique "texture" of the fingerprint, going beyond simple stylometric analysis or word-frequency counts, which can be easily manipulated with automated tools. One dimension is the temporal profile of discourse - the narrative's so-called "internal clock": some authors tend to compress time, using many perfective past forms and moving quickly from cause to effect, while others dilate it, focusing on detailed descriptions and using imperfective past forms to create a static scene. Closely related is the structure of agentivity, where a chronic preference for certain voices becomes a personality marker: some individuals display low agentivity, frequently externalizing themselves from action through passivization, indicating an avoidant style, while others demonstrate excessive agentive control, characteristic of dominating or authoritarian thinking[6]. In addition, the way the author links sentences - cohesion and local coherence - completes the picture: fragmentary coherence with frequent logical jumps suggests an associative thinking style, whereas rigid, linear cohesion indicates systematic thinking. These traits are extremely difficult to synthesize artificially, whether by an algorithm or by a casual imitator.

In open-source investigations, extracting a psycholinguistic fingerprint must allow the construction of a reference profile (baseline profile) that serves as a standard for comparison. Once this profile is established on the basis of a text corpus - such as forum posts under a known identity - it can be overlaid onto other anonymous corpora to identify matches. Identity validation does not rely on the presence of a single keyword, but on the statistical convergence of multiple fingerprint dimensions: the same distribution of grammatical categories, the same preferences for phrasal structures, and, most importantly, the same psychological "voice". Therefore, the psycholinguistic fingerprint transforms text from a mere vehicle of information into forensic evidence, capable of supporting or refuting

[4] Coulthard, M., Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*, Routledge.
[5] Juola, P. (2006). *Authorship Attribution. Foundations and Trends in Information Retrieval*, Now Publishers.
[6] Fast, D.C. et all. (2008) Personality as Manifested in Word Use: Correlations with Self-Report, Acquaintance Report, and Observer Report in *Journal of Personality and Social Psychology 94(2)*, American Psychological Association.

hypotheses about attribution of actions in cyberspace, thus offering a robust solution to the problem of anonymity in the digital age.

4. From signature to attribution: operationalizing psycholinguistic analysis in OSINT

Translating a psycholinguistic fingerprint into actionable attribution requires a disciplined workflow that combines linguistic inference with corroboration from other open-source signals. In practice, the goal is not to "name" an author from a single text, but to establish a defensible linkage between corpora: to show, with measurable confidence, that two or more bodies of writing are likely produced by the same cognitive system.

The operational starting point is corpus construction. Investigators should collect a representative baseline set of texts from a known identity (the *reference corpus*) and a set from the target, anonymous identity (the *questioned corpus*), while controlling for topic, genre, and platform constraints as much as possible. Preprocessing matters: preserving original punctuation, paragraphing, and revisions can be essential because these features often carry stable cues. When translation is unavoidable, analyses should be performed primarily on same-language samples, since translation can flatten stylistic variance and reduce the diagnostic value of syntactic markers.

Next comes feature extraction across multiple layers of language. At the surface level, analysts can quantify function-word patterns, preferred discourse connectors, pronoun ratios, and morphological choices. At the structural level, they can evaluate sentence-length distributions, clause embedding, agentivity preferences (active vs. passive), and coherence strategies (linear argumentation versus associative jumps). At the psychological level, they can track affective loading, certainty markers, threat proximity, and repetition under pressure. The strength of the method lies in convergence: a single indicator is rarely decisive, but a consistent cluster across layers is difficult to fabricate and harder to sustain over time.

For intelligence purposes, attribution should be treated as probabilistic and auditable. Analysts should document which markers were stable, which were context-dependent, and how strongly each marker discriminated between candidate authors. Where possible, apply cross-validation: test the baseline profile against unrelated texts by the same known author and against texts from plausible alternatives. This reduces false positives and helps identify when an apparent match is driven by topic rather than by cognitive style.

Finally, psycholinguistic evidence should be triangulated with complementary OSINT signals - posting schedules, platform affordances, interaction patterns, and recurrent situational details - without allowing any one channel to dominate. Ethical and legal constraints are central: psycholinguistic profiling can be intrusive and can amplify confirmation bias if used carelessly. A multidisciplinary review process, in which the analyst's technical assessment is paired with a psychologist's behavioral interpretation, is therefore not a nice-to-have, but a safeguard that improves both accuracy and accountability.

5. Conclusion

Finally, this paper shows that, both in the dynamic context of information-field security and within the cybersecurity framework, the open information ecosystem requires a reconfiguration of classical investigative methods. Through its exploratory approach, the study underscores the need to integrate and recalibrate the way open-source investigations are conducted and beyond. Exploring the concept of the "psycholinguistic signature" has highlighted its potential to serve as an identifying anchor in the volatile context of information security.

To move from the theoretical identification of the patterns mentioned above to their operational application, it is imperative to adopt a multidisciplinary approach that directly bridges information technology and the behavioral sciences. In this context, the need to develop, within the psychology paradigm, a new generation of specialists becomes critical.

6. References

[1]. M. Coulthard, A. Johnson, *An Introduction to Forensic Linguistics: Language in Evidence*, Routledge, 2007.

[2]. T.A. Hausen, *The Psychology of Language: From Data to Theory (4th ed.)*, Psychology Press, 2014.

[3]. R.J. Heuer, *Psychology of Intelligence Analysis. Center for the Study of Intelligence*, Central Intelligence Agency, 1999.

[4]. P. Juola, *Authorship Attribution. Foundations and Trends in Information Retrieval*, Now Publishers, 2006.

[5]. D. J. Lieberman, *Mindreader: The New Science of Deciphering What People Really Think, What They Really Want, and Who They Really Are*, Rodale Press, 2022.

[6]. M.L. Newman, J.W. Pennebaker, D.S. Berry, J. M. Richards, *Lying Words: Predicting Deception from Linguistic Styles. Personality and Social Psychology Bulletin 29(5),* SAGE Publications, 2003.

[7]. J.W. Pennebaker, *The Secret Life of Pronouns: What Our Words Say About Us*, Bloomsbury Press, 2011.

[8]. L.A. Fast, D.C. Funder, Personality as Manifested in Word Use: Correlations with Self-Report, Acquaintance Report, and Observer Report in *Journal of Personality and Social Psychology 94(2)*, American Psychological Association, 2008.