
Chi ha scritto questo?



Testo originale in tedesco:
Wer schreibt denn da?
Zündlumpen #76 - 2020
zuendlappen.noblogs.org

Indice

- Introduzione
- Identificazione dell'autore presso il BKA
- Metodi di riconoscimento dell'autore e profilazione autoriale
- E adesso?

Introduzione

Una breve panoramica dei moderni metodi della linguistica forense per determinare l'autorialità.

L'articolo seguente cerca di offrire una panoramica da una prospettiva non tecnica e di proporre una valutazione corrispondente. Esistono alcune pubblicazioni accademiche su questo tema che potrebbero essere esaminate per una valutazione migliore. Tuttavia, il mio scopo principale qui è semplicemente sollevare la questione, non fornire una visione solida e conclusiva; quindi, se sapete qualcosa di più, pubblicatelo!

Evitare tracce che potrebbero ritorcersi contro di voi in futuro — magari anche dopo anni o decenni — probabilmente interessa a molte persone che occasionalmente commettono reati e si trovano in conflitto con la legge. Evitare impronte digitali, evitare tracce di DNA, evitare impronte di scarpe e fibre tessili o almeno disfarsi dei vestiti in seguito, evitare telecamere di sorveglianza, evitare tracce di strumenti, evitare registrazioni di qualsiasi tipo, riconoscere la sorveglianza, ecc. — tutto questo dovrebbe interessare chiunque commetta reati di tanto in tanto e voglia proteggersi dall'identificazione. Ma che dire di quelle tracce che spesso emergono solo dopo che un reato è stato commesso, dal desiderio di spiegare anonimamente la propria azione o persino usando uno pseudonimo ricorrente? Quando si scrive e si pubblica un comunicato?

La mia impressione è che in molti casi non venga prestata particolare attenzione a queste tracce, nonostante il rapido sviluppo tecnologico delle capacità analitiche. Questo può essere intenzionale, negligente oppure il risultato di un compromesso tra esigenze concorrenti. Senza voler qui suggerire in generale come trattare queste tracce — dopotutto ognuna deve deciderlo da sé — vorrei delineare i metodi con cui le autorità investigative in Germania e altrove lavorano attualmente (probabilmente), ciò che sembra possibile in teoria e ciò che potrebbe diventare possibile in futuro.

Forse dovrei precisare in anticipo che tutto, o almeno gran parte di ciò che presento qui, è controverso sia scientificamente sia giuridicamente. Mi interessa meno la validità legale delle analisi linguistiche — e nemmeno quella scientifica — quanto piuttosto se sembri plausibile che queste indagini possano orientare un'attività di sorveglianza, perché anche se una traccia non è utilizzabile in tribunale di per sé, potrebbe comunque condurre ad altre tracce utili.

Identificazione dell'autore presso il BKA

Secondo le proprie informazioni, l'Ufficio Federale di Polizia Criminale tedesco (BKA) mantiene un dipartimento dedicato all'identificazione delle autrici di testi. L'attenzione è rivolta a testi collegati ad atti criminali, come rivendicazioni di responsabilità, ma anche a "documenti programmatici" provenienti, tra gli altri, dallo "spettro estremista di sinistra". Tutti i testi raccolti vengono elaborati tramite studi linguistici in una cosiddetta raccolta di comunicati e possono essere confrontati e ricercati tramite il Sistema Informativo Criminale per i Testi (KISTE). Secondo il BKA, i testi vengono classificati in base alle seguenti caratteristiche biografiche delle loro (presunte) autrici: origine, età, istruzione e professione.

Tutti i testi in arrivo vengono inoltre confrontati con testi precedentemente archiviati per determinare se più testi possano essere stati scritti dalla stessa persona.

Nel contesto di indagini specifiche, i testi archiviati possono anche essere confrontati con testi di cui è nota la provenienza, per determinare se siano stati scritti dalla stessa persona o se ciò possa essere escluso.

Queste sono le informazioni ufficiali del BKA su questo dipartimento. Cosa significa in pratica?

Penso che si possa presumere che almeno tutte le rivendicazioni di responsabilità vengano registrate in questo database e analizzate per verificare se esistano altre rivendicazioni delle stesse persone. Il fatto che vengano registrati anche "documenti programmatici" consente ulteriori conclusioni: almeno sembra possibile che, oltre ai testi con rilevanza penale, vengano archiviati anche altri testi ritenuti provenire da una determinata scena politica. Per esempio, testi di giornali, dichiarazioni di gruppi/organizzazioni politiche, appelli, post di blog, ecc. Nel peggiore dei casi, supporrei che tutti i testi pubblicati su noti siti "estremisti di sinistra" (dopotutto è piuttosto facile ottenerli), così come testi di pubblicazioni cartacee ritenute interessanti dalle autorità investigative, vengano inseriti in questo database.

Questo significherebbe che per ogni rivendicazione il BKA avrebbe un insieme di testi che presume abbiano la stessa autrice. Questi possono comprendere altre rivendicazioni ma anche testi inseriti nel database. Oltre alle serie di reati, potrebbero emergere ulteriori indizi sulle autrici, come pseudonimi, nomi di gruppi — o, nel peggiore dei casi, nomi reali — con cui ho scritto una rivendicazione potrebbe aver scritto altri testi, ma anche, a seconda del contenuto, ogni tipo di altra informazione, spesso inclusi indizi sul luogo di residenza e attività di una persona, interessi tematici, caratteristiche biografiche, background educativo, ecc.

Tutte queste informazioni possono quantomeno servire a restringere il cerchio delle sospettate.

Rimane poco chiaro quali altri campioni di confronto il BKA possa ottenere. Per la maggior parte delle persone esiste certamente tutta una serie di testi ai quali le autorità investigative potrebbero avere accesso e che potrebbero essere inseriti nel database in caso di sospetto o forse anche preventivamente — ad esempio se una persona è registrata con una dicitura come “estremista violenta di sinistra”, ecc. Potrebbe trattarsi di qualsiasi cosa riporti il vostro nome, da una lettera a un'autorità a una lettera al direttore di un giornale. Nominerò intenzionalmente solo le fonti più ovvie, per non fornire involontariamente alle autorità investigative nuove idee decisive, ma sono sicuro che possiate rispondere da soli alla domanda su quali vostri testi potrebbero essere accessibili. Se i profiler del BKA riescono a restringere il campo delle persone sospettate a una caratteristica specifica che consente il confronto con grandi quantità di campioni di testo disponibili (per esempio, se si presume che un ricercatore di una certa disciplina sia responsabile di una lettera, allora tutte le pubblicazioni di quel settore potrebbero essere usate come campioni di confronto). Questo potrebbe essere, ad esempio, una possibile spiegazione (parziale) di come si sia arrivati ad Andrej Holm nel caso contro la militante gruppe (mg), almeno se si assume che il BKA non abbia semplicemente cercato su Google “gentrification”; quindi penso sia del tutto possibile che vengano effettuate anche analisi di questo tipo.

Metodi di riconoscimento dell'autrice e profilazione autoriale

Tutto ciò, però, considera soltanto ciò che il BKA sostiene di poter fare e porta queste considerazioni alle loro logiche conclusioni. Ma come funziona realmente il riconoscimento o la profilazione dell'autrice?

Chi non ha mai provato il timore che magari l'insegnante di tedesco potesse smascherarti dopo che una poesia satirica su un professore era comparsa nei bagni della scuola e tutti ridevano del fatto che solo tu avresti potuto scrivere “Leerer” (“vuoto”) invece di “Lehrer” (“insegnante”). Fortunatamente, l'intero dipartimento di tedesco cadde nella trappola, interpretando la cosa come un semplice errore ortografico e ignorando il gioco di parole fin troppo azzeccato. La linguistica forense sembra richiedere un po' di pratica, o almeno una motivazione criminologica, chissà. In ogni caso, secondo un articolo promozionale della “poliziotta linguistica” Christa Baldauf, intorno al 2002 l'analisi degli errori era uno degli strumenti più importanti del BKA insieme all'analisi stilistica. Errori ortografici, errori grammaticali, punteggiatura, ma anche refusi, ortografia nuova o vecchia, indizi sulle particolarità della tastiera, ecc.: tutto questo serve ai “poliziotti della

lingua” per raccogliere indizi sull’autrice. Per esempio, se scrivo “muß” invece di “muss”, ciò potrebbe indicare che ho perso alcune delle più recenti riforme ortografiche durante la scuola. Se invece scrivo costantemente termini che secondo le regole ortografiche richiedono “ß” e non “ss”, potrebbe significare che sulla mia tastiera non esiste il carattere “ß”. Se, per esempio, parlo di “dem Butter” invece di “die Butter”, potrebbe indicare che sono cresciuta in Baviera, ecc. Ma potrei anche fingere tutte queste cose solo per depistare i “poliziotti della lingua”. La plausibilità del mio profilo di errori fa anch’essa parte di un’analisi del genere. Analogamente, l’analisi stilistica esamina le particolarità del mio stile di scrittura. Che tipo di termini utilizzo? La struttura delle mie frasi mostra schemi specifici? Esistono combinazioni ricorrenti di termini che compaiono magari in testi diversi? Penso che chiunque osservi attentamente i propri testi riconoscerà alcune caratteristiche stilistiche personali.

Tali analisi qualitative servono principalmente a profilare autrici e autori di testi. Sebbene sia certamente possibile collegare diversi testi in questo modo, il vero valore di queste analisi sta nella possibilità di determinare elementi come età, “livello di istruzione”, “appartenenza a una scena”, origini regionali e talvolta persino indizi su professione/formazione, ecc. Si sente parlare anche di tentativi di determinare il genere dell’autrice, ma generalmente non sembrano essere così semplici.

Esistono inoltre analisi più quantitative e statistiche che esaminano tutto, dalla frequenza delle parole alle combinazioni lessicali fino alla sintassi e alla struttura delle frasi misurabili in questo modo. Questi metodi, noti come stilometria, sono talvolta molto controversi perché non è possibile dire con precisione cosa stiano effettivamente misurando, ma a volte producono risultati sorprendenti, soprattutto in combinazione con approcci di machine learning. Ritengo quindi probabile che questi approcci vengano utilizzati soprattutto per raggruppare testi differenti in base alle loro somiglianze.

Il chiaro vantaggio di tali analisi quantitative è che possono essere eseguite su larga scala. Tutti i testi disponibili digitalmente o digitalizzabili possono essere analizzati in questo modo. Dai post sui social media ai libri, i testi possono essere acquisiti tramite questi metodi. Sebbene il successo di tali tecniche sia ancora relativamente modesto e spesso si sia scoperto che testi apparentemente simili sono più simili nel genere che nell’autorialità, se si assume che gli stili di scrittura individuali possano lasciare schemi quantitativi, allora una volta conosciuti tali schemi sarà possibile attribuire in massa testi a determinati autrici.

E adesso?

Esistevano ed esistono naturalmente diversi approcci per affrontare questa consapevolezza, nessuno migliore o peggiore degli altri. Chi non scrive comunicati

evita in gran parte questo problema, ma resta comunque coinvolta nella questione della partecipazione a pubblicazioni e dell'autorialità di altri testi. Chi modifica e oscura i testi prima della pubblicazione, ad esempio facendo riscrivere e riformulare successivamente i passaggi da più persone, ecc., rischia comunque di sviluppare caratteristiche linguistiche e stilistiche sfruttabili in configurazioni ripetutamente simili oppure di non riuscire a nascondere efficacemente tali caratteristiche. Chi pensa di poter ignorare il problema perché nessuno dei propri campioni di testo è disponibile o perché ritiene che il valore legale del riconoscimento dell'autrice sia troppo fragile, rischia che in futuro quei campioni diventino disponibili in qualche modo (per esempio perché si viene condannate con successo per l'autorialità) oppure che cambi la valutazione giuridica della procedura. Chi si fida del fatto che la tecnologia non sia ancora abbastanza avanzata potrebbe essere sorpresa dagli sviluppi futuri. Chi usa soluzioni tecniche per nascondere la propria autorialità rischia di lasciare nuove caratteristiche e nuove tracce, oltre a produrre comunicati scritti male che nessuno vuole leggere. Se invece non scrivete mai alcun testo, allora semplicemente non scrivete testi.

Quindi fate ciò che vi sembra più adatto, ma fatelo d'ora in poi — se non lo avete già fatto — tenendo presenti queste tracce e quella sensazione di disagio allo stomaco che, si dice, abbia già salvato molte persone dal commettere un errore imprudente nel momento decisivo.