Istituto a Ordinamento Universitario D.M. 30/04/2018, G.U. n. 138 16/06/2018

Sede di Firenze

Corso di studi triennale in Mediazione Linguistica (classe di laurea L-12)

Indirizzo Criminologia e Cybersecurity

TESI FINALE

Il manifesto Industrial Society and Its Future: un'analisi linguistica-forense

Candidata: Beatrice Oliva

Matricola n° 787/FI

Relatore: Prof. Paolo Rondinelli

Correlatrice: Prof.ssa Valeria Semeghini

A.A. 2025/2026

Ai miei genitori, per aver reso possibile questo traguardo.

ABSTRACT

La linguistica forense è un potente strumento sul quale si basa la metodologia della profilazione autoriale. Sebbene lo studio del linguaggio, scritto e parlato, sia un ambito molto complesso e in constante evoluzione, dal 1968 ad oggi la disciplina è stata riconosciuta come metodo scientifico in grado di sviluppare prove con un peso probatorio sufficiente da poter essere presentate di fronte a una corte di tribunale. Nel corso di questa evoluzione un passaggio fondamentale è stato rappresentato dal caso di Theodore Kaczynski (Unabomber), autore del manifesto *Industrial Society and Its Future*; per tale motivo nel seguente lavoro di ricerca si è optato di analizzare come caso studio il *Manifesto* originale. L'obiettivo è dunque quello di svolgere un'analisi linguistica sul documento, in modo tale da evidenziare quali elementi riguardanti il profilo di un autore possono emergere tramite le scelte linguistiche applicate.

SUMMARY

Forensic linguistics is a powerful tool underpinning the methodology of authorship profiling. Although the study of language, both written and spoken, is a highly complex and constantly evolving field, since 1968 the discipline has been recognized as a scientific method capable of producing evidence with sufficient probative value to be presented before a court of law. Throughout its development, a pivotal milestone was represented by the case of Theodore Kaczynski (the Unabomber), author of the manifesto *Industrial Society and Its Future*; for this reason, in the present research, the case study that was chosen is the original *Manifesto*. The objective is therefore to conduct a linguistic analysis of the document, in order to highlight which aspects of an author's profile can be inferred through the linguistic choices reflected in the text.

RESUMEN

La lingüística forense es una herramienta fundamental sobre la que se sustenta la metodología de la perfilación de autoría. Aunque el estudio del lenguaje, tanto escrito como oral, constituye un ámbito altamente complejo y en constante evolución, desde 1968 esta disciplina ha sido reconocida como un método científico capaz de producir pruebas con suficiente valor probatorio para ser presentadas ante un tribunal de justicia. A lo largo de esta evolución, un hito fundamental fue el caso de Theodore Kaczynski (Unabomber), autor del manifiesto *Industrial Society and Its Future*. Por este motivo, en el presente trabajo de investigación, se ha optado por analizar como estudio de caso el Manifiesto original. El objetivo es realizar un análisis lingüístico del documento, así que se muestren los elementos del perfil de un autor pueden inferirse a partir de las elecciones lingüísticas presentes en el texto.

INDICE

INTRODUZIONE 5

CHAPTER 1 - CONTEXTUAL FRAMEWORK AND RESEARCH RATIONALE

1.1 The Ted Kaczynski case: general overview 7

1.2 The manifesto *Industrial Society and Its Future* 9

1.3 Rationale for the selection of the case study 11

CAPITOLO 2 - LA LINGUISTICA FORENSE COME STRUMENTO INTERPRETATIVO

2.1 Linguistica forense: definizione 13

2.2 Le applicazioni della linguistica forense secondo Coulthard 15

2.3 La profilazione autoriale nella linguistica forense 18

2.3.1 *Feature* lessicali 19

2.3.2 *Feature* sintattiche 20

2.3.3 *Feature* semantiche 20

2.3.4 *Feature* strutturali 20

2.3.5 *Feature* contenutistiche 20

2.4 Correlazione tra *feature* linguistiche e dati demografici 21

2.5 Metodo di *profiling*: dall'analisi del contenuto all'analisi quantitativa delle *feature* 25

CAPITOLO 3 - LO STATO DEGLI STUDI SUL MANIFESTO *INDUSTRIAL SOCIETY AND ITS FUTURE*

3.1 Studi e interpretazioni del *Manifesto* (storici, ideologici e criminologici) 27

3.2 Limiti degli studi esistenti 30

3.3 Contributo della presente ricerca 31

CAPITOLO 4 - L'IMPRONTA LINGUISTICA DI UNABOMBER: UN'ANALISI FORENSE MULTILIVELLO

4.1 Scelte lessicali e campi semantici dominanti 33

4.2 Analisi sintattica e organizzazione del discorso nel *Manifesto* 41

4.3 Le caratteristiche macro-testuali e semantiche del *Manifesto* 46

CAPITOLO 5 - I RISULTATI DELL'ANALISI LINGUISTICA: PROFILAZIONE, INTERPRETAZIONE E VERIFICA

5.1 Dall'analisi linguistica alla profilazione dell'autore 51

5.2 Dalla profilazione all'interpretazione del *Manifesto* 54

5.3 La verifica della profilazione attraverso la comparazione autoriale 57

CAPÍTULO 6 - LÍMITES, PERSPECTIVAS Y NUEVOS DE LA PERFILACIÓN DE AUTOR EN LA LINGÜÍSTICA FORENSE

6.1 Limitaciones metodológicas y aplicativas de la perfilación de autor 63

6.2 Perspectivas metodológicas para el desarrollo de la perfilación de autor 68

6.3 La inteligencia artificial como nuevo desafío para la lingüística forense 71

CONCLUSIONI 75

BIBLIOGRAFIA 79

INTRODUZIONE

Negli ultimi decenni la linguistica forense ha assunto un ruolo sempre più rilevante nell'ambito investigativo e giudiziario, dimostrando come l'analisi del linguaggio possa costituire uno strumento utile non solo per l'interpretazione dei testi giuridici, ma anche per la ricostruzione del profilo di un autore anonimo. Alla base di tale ragionamento vi è la concezione che ogni individuo possieda una propria e distintiva impronta linguistica, modellata sulla base delle esperienze e dei contesti comunicativi con cui si è entrati in contatto nel corso della vita; questo permette di articolare le proprie scelte linguistiche in *pattern* lessicali, sintattici, strutturali e semantici che con il tempo si stabilizzano, andando a formare l'idioletto di un autore. A ciò si affiancano gli studi sociolinguistici che hanno permesso di individuare possibili correlazioni probabilistiche tra determinati fenomeni linguistici e specifici dati demografici e che sono oggi alla base della profilazione autoriale.

Da questa prospettiva nasce il seguente lavoro di ricerca, con la volontà di presentare un esempio di applicazione concreta della metodologia della linguistica forense, analizzando come caso studio il manifesto *Industrial Society and Its Future*, redatto da Theodore Kaczynski, anche conosciuto come Unabomber. La vicenda, infatti, rappresenta uno dei casi più emblematici che ha contribuito significativamente a evidenziare il potenziale applicativo della materia: durante le indagini, durate quasi vent'anni, il punto di svolta è stata precisamente la pubblicazione del *Manifesto* sul giornale Washington Post, evento che ha permesso agli esperti di eseguire un'analisi linguistica del documento. I risultati di questo esame costituirono il principale materiale linguistico attraverso il quale fu possibile elaborare un profilo autoriale che orientò le investigazioni verso il responsabile. Sebbene non rappresentassero una prova diretta e definitiva dell'identità dell'autore, permisero di far progredire le indagini e ottenere ulteriori prove da poter presentare davanti alla Corte distrettuale del Montana e arrivare così a una condanna.

Nel corso dell'evoluzione e dell'affermazione della linguistica forense in ambito investigativo, quindi, la vicenda assume un ruolo centrale, poiché ha mostrato come il linguaggio possa fornire informazioni che vanno oltre il semplice contenuto di un testo.

È precisamente da questa considerazione che emerge l'interesse di fornire una ricostruzione, anche se di dimensioni ridotte, dell'analisi linguistica condotta sul *Manifesto*, senza ripercorrere l'intera vicenda giudiziaria. L'obiettivo, infatti, non è quello di proporre una lettura criminologica del caso Kaczynski, bensì quello di mostrare quali elementi emergono nel momento in cui un testo viene esaminato tramite gli strumenti della disciplina e come gli studi sociolinguistici ne inferiscono una possibile corrispondenza con specifici dati demografici. Pertanto, la domanda che guida l'intera ricerca è la seguente: quanto mostrano di ognuno di noi le scelte linguistiche che attuiamo nel momento in cui elaboriamo un testo?

Per rispondere a tale quesito, il lavoro si articola in sei capitoli.

Il primo capitolo, redatto in lingua inglese, espone brevemente la vicenda giudiziaria legata al caso Unabomber, presentando anche le tematiche affrontate all'interno del *Manifesto*.

Il secondo capitolo rappresenta l'esposizione del quadro teorico e metodologico della linguistica forense con particolare attenzione alla profilazione autoriale. A ciò segue la ricostruzione dello stato degli studi dedicati al *Manifesto*, con l'obiettivo di mostrare il contributo che la presente ricerca intende offrire.

Il quarto e il quinto capitolo rappresentano la parte centrale della tesi, poiché sono dedicati allo sviluppo dell'analisi linguistica effettuata sul documento. Essa parte dall'esame delle diverse *feature* estraibili dal testo, per inferirne successivamente una correlazione probabilistica con un dato demografico che costituisce parte del profilo dell'autore. L'analisi è arricchita da una sezione destinata alla verifica dell'analisi svolta: i risultati ottenuti riguardanti le caratteristiche linguistiche, infatti, vengono confrontati con quelli che emergono dall'esame della tesi di dottorato di Theodore Kaczynski. Attraverso questo confronto vengono individuate le ricorrenze stilistiche e linguistiche condivise tra i due scritti, mostrando come specifiche scelte si mantengano costanti sebbene presenti in contesti comunicativi differenti.

Il lavoro di ricerca si chiude con il sesto capitolo, redatto in lingua spagnolo, dedicato a una riflessione basata su due principali articoli, Suchman (2023) e Queralt (2024), riguardo i limiti metodologici e applicativi della linguistica forense e dell'attribuzione autoriale, con un particolare riguardo verso le nuove sfide derivanti dall'avvento di sistemi basati sull'intelligenza artificiale.

CHAPTER 1 - CONTEXTUAL FRAMEWORK AND RESEARCH RATIONALE

1.1 The Ted Kaczynski case: general overview

The case study examined in this research concerns the figure of Theodore Kaczynski, better known as the Unabomber, who was responsible for a series of attacks carried out between 1978 and 1995 in the United States, more specifically in the States of Illinois, California, Utah, and Tennessee.

Born in Chicago in 1942, Kaczynski showed remarkable intellectual abilities from an early age; in the early 1960s, he graduated from Harvard University and obtained a master's degree and a PhD in mathematics from the University of Michigan. In 1967, he was appointed assistant professor of mathematics at the University of California, Berkeley, but resigned after a short time (Wiehl, 2020).

The decision to abandon the academic world represents a significant moment in Kaczynski's life. During his years at Harvard, he had developed a strong interest in the social and cultural changes brought by technological progress and the rise of industrial society. After leaving the University, these reflections turned into increasingly radical ideologies that would later be widely articulated in Kaczynski's written work, the manifesto *Industrial Society and Its Future*. Over time, however, his ideology became increasingly radical[1], ultimately leading him to pursue his desire to subvert the entire social order through homemade explosive devices, constructed in his isolated cabin near the town of Lincoln, Montana.

In 1978, Kaczynski decided to initiate his criminal campaign by sending mail bombs to individuals who, in his view, embodied the technological progress that had destroyed society; therefore, his selected targets were scientists, university researchers, and airline executives.

His first attack occurred when he mailed a bomb to a professor at Northwestern University (Chicago, Illinois), an assault followed by further bombings that, at that stage, did not result in fatalities. In 1979, however, Kaczynski succeeded in placing an explosive device in the cargo hold of a commercial aircraft. Although the plane was able to land safely without the bomb detonating on board, the incident constituted a federal offense and consequently required the involvement of the FBI; a dedicated task force was thus established under the name UNABOMB (UNiversity and Airline BOMber)[2]. The attacks continued in the following years, and the first fatality occurred in 1985: Hugh Scrutton, the owner of a computer store in California, was killed by a bomb that had been placed in the parking lot of his business. A six-year hiatus followed, the reasons for which still remain unclear. When he resumed his campaign, Kaczynski sent further mail bombs that resulted in

[1] Wiehl L. (2020). *Hunting the Unabomber: The FBI, Ted Kaczynski, and the Capture of America's Most Notorious Domestic Terrorist*. Thomas Nelson.

[2] Turchie, T. D. (1996, April 2). *Affidavit in support of application for search warrant*. United States District Court for the District of Montana.

additional fatalities, including Thomas J. Mosser and Gilbert Murray. Three more years passed before his arrest on 3rd April 1996, followed by a further two years before his final conviction on 22nd January 1998 issued following a guilty plea formally entered after his admission of responsibility for the charges (Wiehl, 2020).

Having briefly retraced Theodore Kaczynski's criminal activity, it is particularly relevant to examine his distinctive modus operandi. His operational method consisted in the construction of explosive devices subsequently delivered by mail to carefully selected targets. Although the persistence of his attacks and the deliberate nature of his target selection may place Kaczynski within the broader category of serial killers, his modus operandi simultaneously distinguishes him from more conventional serial offender profiles. His peculiar characteristic lies in his use of mail bombs as the primary weapon against his victims. This choice not only reduced his exposure to identification and capture but also allowed him to remain emotionally detached from the violent act itself. This element is especially significant when compared to typical serial killer profiles, in which the offender's motivation often includes a direct confrontation with the victim and their suffering, an aspect that is notably absent in the Unabomber case (Rebecca Taylor LaBrode, 2007). Moreover, the choice itself is not merely an operational choice but also a symbolic act. By using the mail to deliver bombs that could traverse the network and reach their targets undetected, he implicitly demonstrated how the very infrastructure underpinning industrial society could be easily exploited, revealing its inherent vulnerabilities.

Consequently, it can be inferred that this distinctive modus operandi played a crucial role throughout the investigation, since it enabled Kaczynski to minimize personal risk and thus remain undetected, and even largely unsuspected, for nearly two decades (1978-1996). For this reason, the provided context frames his written production, the manifesto *Industrial Society and Its Future* which marked a decisive turning point in the case. Up to that moment, it had proven difficult to construct a reliable offender profile on the basis of the evidence recovered from the various crime scenes, and even the preliminary profile that had been developed failed to produce investigative leads or identify a concrete suspect. Nevertheless, the *Manifesto* was later analysed through an approach that was relatively unconventional at the time, treating the text as the space in which the Unabomber's ideology is structured through linguistic patterns characteristic of the author. This allowed the identification of specific elements that proved fundamental to the advancement of the investigation and, ultimately, to the resolution of the case.

1.2 The manifesto *Industrial Society and Its Future*

As previously anticipated in the foregoing paragraph, the manifesto *Industrial Society and Its Future* represented the key element that enabled the construction of an authorial profile, later connected to the Unabomber, and ultimately led to the arrest of an individual for the first time after nearly twenty years. In order to achieve this result, investigators relied on an approach that was relatively unconventional at the time, focusing primarily on the linguistic analysis of the *Manifesto* rather than on a purely content-based examination.

Since the present research aims to reconstruct the linguistic analysis that was conducted, providing a brief overview of the *Manifesto*'s content appears useful; this is intended to illustrate how the conclusions derived from a content-based approach may differ from, or fail to capture certain relevant components of the findings produced through linguistic analysis, which will constitute the focus of the following chapters.

Industrial Society and Its Future is the written work through which Theodore Kaczynski's ideological perspective can be examined and understood. It consists of an essay of approximately 35,000 words, divided into 232 numbered sections in which the critique directed at industrial society and technological progress is developed according to a relatively linear argumentative structure.

In the opening sections (sections 1-32), the author introduces the main thesis upon which the entire work is based: "The Industrial Revolution and its consequences have been a disaster for the human race" (Kaczynski, 1995). However, before illustrating his reasoning, Kaczynski integrates a series of definitions of concepts that are essential to understand his thesis, including notions such as "feelings of inferiority"[3], "oversocialization"[4].

The following sections (sections 33-68) present the theoretical cornerstone of the *Manifesto*: the "power process"[5]. This concept is presented as a biological need of human beings and is articulated through four components: the setting of a goal, the personal effort required to achieve it, the attainment of that goal, and the presence of a certain degree of autonomy necessary during the process (Kaczynski, 1995). Within Kaczynski's perspective, the "power process" has a central position, as it provides the motivation behind his criticism towards the technological progress: technology has created a social order in which individuals have progressively lost their autonomy

[3] "By 'feelings of inferiority' we mean not only inferiority feelings in the strict sense but a whole spectrum of related traits; low self-esteem, feelings of powerlessness, depressive tendencies, defeatism, guilt, self- hatred, etc." (Kaczynski, 1995).

[4] "By 'oversocialization' I mean being so strongly socialized that certain basic human instincts are repressed below the level needed for normal functioning." (Kaczynski, 1995).

[5] "Human beings have a need (probably based in biology) for something that we will call the "power process." This is closely related to the need for power (which is widely recognized) but is not quite the same thing." (Kaczynski, 1995).

due to the strengthening of institutional control. As a consequence, individuals started to engage themselves in "surrogate activities"[6], considered as artificial goals whose achievement no longer derives from genuine personal purpose. Thus, the "power process" was not eliminated by technology; rather it was modified by it, resulting in a more artificial and less authentic system.

Following the introduction of this concept, the author provides a concrete example of his argument. In sections 69-86, he compares primitive societies with modern ones, arguing that the former enables a full and autonomous exercise of personal will. This is followed by the most extensive part of the *Manifesto* (sections 87-139): the aim is to analyse technology in order to demonstrate his previous claim, i.e. the loss of autonomy is not an accidental phenomenon but it is strongly tied to the internal logic of the industrial system, which is strongly influenced by the technological progress. More specifically, Kaczynski associates technological development with an increase of social complexity, a complexity that requires the continuous proliferation of institutions and control mechanisms, which would inevitably lead to a limitation of individual freedom.

These sections lay the groundwork for the subsequent parts, which highlight the impossibility of reforming industrial society without altering its structural balance (sections 140-170) and outline a series of possible future scenarios deriving from the continued expansion of technology (sections 171-206). The *Manifesto* culminates in a justification for a possible revolution against the system (sections 207-232). The argument revolves around the idea that if the social system proves to be both oppressive and impossible to reform, then a revolution becomes inevitable and morally legitimate (Kaczynski, 1995).

Knowing the content of the work and the context in which this text was produced, it is possible to make some inferences: the author of the *Manifesto* may plausibly be an anti-establishment political extremist, targeting individuals who symbolically represent the institutions that, in his view, have irreversibly deprived individuals of their autonomy. Furthermore, given the nature of the targets and the methods employed, it might even be hypothesized that the offender was not a social outcast, but rather an individual fully integrated into the very society he criticizes. In addition, the construction of the explosive devices suggests a high degree of technical and manual competence, potentially developed through working experience (notably, the FBI initially drafted a profile identifying a working-class aircraft mechanic as a possible perpetrator[7]).

Nevertheless, the information outlined above does not allow the extraction of reliable demographic data such as gender, age, educational background, or social class. Although, in this

[6] "We use the term 'surrogate activity' to designate an activity that is directed toward an artificial goal that people set up for themselves merely in order to have some goal to work toward, or let us say, merely for the sake of the 'fulfillment' that they get from pursuing the goal." (Kaczynski, 1995).

[7] Turchie, T. D. (1996, April 2). *Affidavit in support of application for search warrant*. United States District Court for the District of Montana.

case a composite sketch was produced on the basis of eyewitness testimony (following the Salt Lake City attack), the elements that can be derived solely from content analysis remain insufficient to formulate a comprehensive offender profile. Such elements may contribute to narrowing the pool of potential suspects; however, they cannot identify the author as an individual, but rather only outline his ideological position and worldview. It is precisely at this stage that linguistic analysis becomes essential, since only through an integrated approach that combines content analysis with linguistic examination, it is possible to construct a coherent and methodologically grounded authorial profile.

1.3 Rationale for the selection of the case study

It is on the basis of these premises that the decision to focus on an emblematic case becomes justified, namely one in which linguistic analysis played a decisive role within the investigative process.

The case of Theodore Kaczynski has been selected with the intention of presenting one of the investigations that contributed to the recognition of forensic linguistics as a valid and significant evidentiary method in legal proceedings in the United States. It must be specified, however, that the final conviction was not based directly on the evidence obtained through forensic linguistic analysis. Rather, forensic linguistics played a crucial role during the investigative phase, as it made it possible to obtain the search warrant for the cabin in which Kaczynski lived; nonetheless, it was the physical evidence collected during that search that ultimately constituted the core of the prosecution's case and led to the conviction.

Despite this, the present research may be regarded as an attempt to reconstruct the linguistic process through which the specific authorial profile associated with Theodore Kaczynski was developed. For this reason, the study focuses exclusively on the analysis of the manifesto *Industrial Society and Its Future*. Thus, it is important to emphasize that this case study will not be addressed from a criminological or psychological perspective, nor will the author's biographical background be further examined, except insofar as it is strictly necessary in order to contextualize particular linguistic choices. The analysis will be directed solely toward the *Manifesto* as a linguistic product, with the aim of demonstrating the following points.

First, as previously mentioned, the present work seeks to examine one of the cases that most effectively illustrates the potentiality of forensic linguistics, documenting how the data produced through this discipline do not stem from mere intuition, but rather from established studies and validated methodologies.

Similarly, the choice of this case study is motivated by the intention to demonstrate the existence of a direct relationship between language and identity. Writing, in fact, is not limited to functioning as a communicative tool through which ideology is disseminated; rather, it becomes a space in which the author makes deliberate lexical and syntactic choices, potentially allowing a written text to be attributed to a specific authorial profile.

Finally, the decision not further address the broader case and its legal and criminological implications, but to exclusively focus on the *Manifesto*, is based on the fact that what emerges as particularly significant in this investigation is that the written text itself represented the only element that ultimately led to the resolution of the case after nearly twenty years, during which no incriminating evidence had been found and the Unabomber's attacks had continued. The chronology of the case clearly highlights this dynamic: the *Manifesto* was published by The Washington Post on 19 September 1995, constituting the first moment in which the text became publicly accessible and, therefore, available for systematic analysis, including from a linguistic perspective. From that point onward, the analysis of the *Manifesto* progressed and ultimately led, within a few months, to the arrest of Theodore Kaczynski on 3 April 1996.

In light of these considerations, the present research aims to examine the *Manifesto* as the primary source by adopting the analytical tools of forensic linguistics, the aim is to identify the author's distinctive linguistic features, which may contribute to the reconstruction of a coherent authorial profile. The following chapters will therefore outline the theoretical and methodological framework of forensic linguistics, review the existing body of scholarship concerning the *Manifesto*, and finally present a multilevel analysis of the text.

CAPITOLO 2 - LA LINGUISTICA FORENSE COME STRUMENTO INTERPRETATIVO

2.1 Linguistica forense: definizione

Come definito nel capitolo precedente, lo scopo di questo lavoro di ricerca è quello di ricostruire il profilo linguistico dell'autore del manifesto *Industrial Society and Its Future*. L'analisi parte evidentemente dallo scritto originale e, nonostante al momento della stesura della tesi si conosca già l'autore effettivo dello scritto, durante questo lavoro di ricerca il testo viene trattato come un testo anonimo. Risulta quindi necessario ricorrere ad una serie di strumenti della linguistica forense, adottando un approccio metodologico misto che unisce il metodo qualitativo e quantitativo. Tuttavia, prima di spiegare quali strumenti verranno impiegati e come, è necessario inquadrare la disciplina della linguistica forense.

La linguistica forense è un ramo della linguistica applicata che si occupa dello studio della lingua speciale dei testi legali e dell'analisi della lingua nei processi penali (Marotta e Tamponi 2021). Questo settore è relativamente recente: negli anni Sessanta Noam Chomsky guida la cosiddetta "rivoluzione chomskiana" che modifica profondamente la linguistica. Fino a quel momento tale disciplina veniva considerata come lo studio sistematico della lingua, quest'ultima intesa come un sistema organico di segni nel quale ogni elemento ha valore solo in relazione agli altri facenti parte del sistema stesso[8]; siamo, dunque nello strutturalismo di Ferdinand de Saussure che ha permesso di offrire un quadro teorico per la disciplina basato su concetti ben definiti e condivisibili da altri studiosi, alla base dei quali c'è la lingua concepita come un fenomeno sociale che esiste nella società ed esiste in relazione ad essa.

Tuttavia con la "rivoluzione chomskiana" si verifica un cambiamento radicale: la lingua viene interpretata come una facoltà innata ed interiorizzata dai parlanti, basata su regole grammaticali precise, che permettono di formulare frasi. Questo concetto portante della grammatica generativa[9] di Chomsky, che pone al centro dell'attenzione la "competenza"[10] del parlante piuttosto che l'uso effettivo della lingua ("performance"[11]), è l'elemento che porta ad una frammentazione della disciplina della linguistica ed è proprio da questa frammentazione che nasce il settore della linguistica forense, termine che appare per la prima volta nel 1968 da parte del linguista svedese Jan Lars Svartvik, come calco di *linguistic forensics*[12]; Svartvik, infatti, riuscì a dimostrare che la

[8] Saussure, F. (1916). *Cours de linguistique générale.* Payot.
[9] «La teoria generativa considera la grammatica una conoscenza rappresentata nella mente del parlante, e mette il parlante al centro della ricerca» (Benincà e Tortora, 2011).
[10] «Conoscenza della lingua da parte del parlante-ascoltatore» (Chomsky, 1965).
[11] «L'uso effettivo del linguaggio in situazioni concrete» (Chomsky, 1965).
[12] Shim A. (2022). *Forensic Linguistics: Science or Fiction?*. UC Law Journal.

confessione resa da un condannato non corrispondesse alle sue dichiarazioni precedenti, in quanto vi erano delle parti che non erano in linea con l'idioletto dell'individuo.

Nonostante il lavoro di Svartvik, la linguistica forense si affermerà e verrà riconosciuta solo negli anni Novanta dal momento che nel corso degli anni questa disciplina è stata al centro di grandi dibattiti che hanno messo in dubbio la sua scientificità, primo tra tutti quello riguardante l'ammissibilità delle prove linguistiche. Siano negli Stati Uniti, Paese che ha contribuito in maniera significativa allo sviluppo e al riconoscimento di questa materia a livello internazionale; fino agli anni Ottanta e Novanta i tribunali statunitensi applicavano il criterio giuridico conosciuto come *Frye test*[13], secondo cui la prova scientifica è ammissibile solo se generalmente accettata come affidabile dalla comunità scientifica di riferimento. Tale criterio, nel 1976, ha permesso al tribunale federale della *Northern District of California*[14] di rifiutare una prova presentata da un esperto di linguistica, ritenendo che quel tipo di analisi non fosse ancora abbastanza affidabile poiché nuova e carente di metodi consolidati.

Il punto di svolta avviene però con la sentenza *Daubert* della Corte Suprema degli Stati Uniti (1993) che ridefinisce i criteri di ammissibilità della prova scientifica introducendo il *Daubert test*[15]: questo cambiamento fece sì che le analisi linguistiche entrassero più facilmente come prova nei tribunali poiché con *Daubert* l'attenzione si spostò dall'accettazione generale dell'analisi scientifica alla verificabilità e solidità del metodo applicato. Tale criterio ha permesso l'ammissibilità delle prove linguistiche in casi quali il procedimento penale federale contro Theodore Kaczynski[16], conclusosi con un *guilty plea* nel 1998 e il procedimento civile *Kleiman v. Wright* (Florida District Court, 2020), in cui il dibattito scientifico è stato ritenuto rilevante per la valutazione della prova ma non per la sua esclusione.

Ad oggi la linguistica forense è dunque divenuta uno strumento di prova ben consolidato che aggrega in sé tutte le applicazioni della linguistica in ambito legale (Mastronardi e Trojani, 2014); le

---

[13] «La regola è che le opinioni degli esperti o dei consulenti tecnici sono ammissibili come prova nei casi in cui la questione oggetto dell'indagine sia tale che persone prive di esperienza difficilmente sarebbero in grado di formulare un giudizio corretto. Ciò avviene perché la materia trattata rientra, almeno in parte, nell'ambito di una scienza, di un'arte o di un mestiere, e richiede una precedente pratica, esperienza o studio per acquisirne un'adeguata conoscenza. Quando la questione in esame non rientra nella sfera della comune esperienza o delle conoscenze ordinarie, ma richiede competenze o conoscenze specialistiche, allora le opinioni dei testimoni esperti nella specifica scienza, arte o professione cui la questione si riferisce sono ammissibili come prova.» *Frye v. United States* (Supreme Court of the District of Columbia, 1923). Traduzione a cura della laureanda.

[14] *United States v. Hearst* (U.S. District Court for the Northern District of California, 1976).

[15] «Nulla nelle Regole nel loro complesso o nel testo e nella storia della redazione della Regola 702, che disciplina specificamente la testimonianza degli esperti, lascia intendere che la "accettazione generale" costituisca un presupposto necessario per l'ammissibilità delle prove scientifiche. Inoltre, uno standard così rigido sarebbe in contrasto con l'impostazione liberale delle Regole e con il loro approccio generale volto ad attenuare le tradizionali barriere all'ammissione delle testimonianze valutative.» Daubert v. Merrell (U.S. Supreme Court, 1993). Traduzione a cura della laureanda.

[16] *United States v. Kaczynski* (U.S. District Court for the District of Montana, 1996).

tecniche sviluppate nel corso degli anni differiscono tra loro a seconda dell'ambito di applicazione della disciplina, ma sono riconosciute e insegnate in maniera univoca in diverse parti del mondo.

2.2 Le applicazioni della linguistica forense secondo Coulthard

Come si è detto nel capitolo precedente, la linguistica forense trova applicazione in diverse aree specifiche che, come spiega Malcolm Coulthard nel suo libro *Forensic Linguistics: the application of language description in legal contexts*, si possono riassumere in tre ambiti:

- l'analisi del linguaggio di testi giuridici scritti;
- l'analisi del linguaggio parlato nel processo legale;
- il linguista come testimone esperto.

Nei testi giuridici scritti l'esperto di linguistica parte da una distinzione di ciò che è e non è un testo giuridico, inteso come un testo che interpreta o applica norme del diritto; da qui ne deriva la classificazione in testi normativi, testi interpretativi e testi applicativi, basata sulla funzione che questi testi svolgono nell'ordinamento giuridico. Il linguista poi si focalizza sul definire le caratteristiche peculiari del linguaggio tipico di questi scritti e, per fare ciò, analizza la grammatica, la costruzione dei periodi e il vocabolario, elemento che spesso risulta essere complesso o in alcuni casi addirittura arcaico. Lo scopo di questa analisi è duplice: da una parte chiarire l'ambiguità che il linguaggio legale può presentare, delineando una corretta interpretazione del testo; dall'altra verificarne la comprensibilità e suggerirne eventualmente una semplificazione. Un esempio, in tal senso, è il *California Plain Language Rewrite* del 2005, ovvero l'iniziativa da parte dello Stato della California di riscrivere le proprie *Pattern Jury Instructions* (in italiano "istruzioni standard per la giuria") in *plain english*[17]; questo ha apportato cambi significativi a una serie di istruzioni. Di seguito si riporta la definizione di falsa testimonianza prima e dopo tale iniziativa per mostrare le modifiche apportate:

> *A witness who is willfully false in one material aspect of his or her testimony is to be distrusted in others. You may reject the whole testimony of a witness who willfully has testified falsely as to a material point, unless, from all the evidence, you believe the probability of truth favors his or her testimony in other particulars* (California State Portal, 2005).[18]

[17] «Linguaggio che riflette gli interessi e le esigenze del lettore e del consumatore piuttosto che gli interessi legali, burocratici o tecnologici.» Recco R. S. (2020) *Say it in Plain English*. Università degli Studi di Enna "Kore".

[18] «Un testimone che rende volontariamente una falsa dichiarazione su un aspetto rilevante della propria testimonianza deve essere considerato inaffidabile anche per gli altri aspetti. Si può respingere integralmente la testimonianza di un testimone che abbia deliberatamente mentito su un punto sostanziale, salvo che, alla luce di tutte le prove, si ritenga che la probabilità della verità deponga a favore dell'attendibilità della sua testimonianza negli altri aspetti» (California State Portal, 2005) Traduzione a cura della laureanda.

Modificato in *plain english* nel 2005:

> *If you decide that a witness deliberately lied about something important, you should consider not believing anything that witness says. Or, if you think the witness lied about some things, but told the truth about others, you may simply accept the part that you think is true and ignore the rest* (California State Portal, 2005).[19]

Da questo processo di riscrittura si può notare come la sintassi del testo originale, più complessa poiché basata sulla subordinazione, sia stata semplificata dividendo i periodi in frasi più brevi e legandole tra loro con un rapporto di coordinazione. Nella versione del 2005 si nota inoltre un tentativo di avvicinamento al linguaggio quotidiano, tramite l'abbandono del tono informale e l'eliminazione del linguaggio tecnico-specialistico, che vede l'utilizzo di espressioni come '*willfully false*' o '*to be distrusted*', in favore di un lessico più comune ('*deliberately lied*'; '*consider not believing*').

Nonostante si possa pensare che attraverso un processo di riscrittura in *plain english* si perda precisione giuridica o ci sia il rischio di semplificare troppo un concetto, è necessario sottolineare che procedimenti come questo vengono svolti da esperti del settore che lavorano proprio con lo scopo di trovare un equilibrio tra precisione legale e comprensibilità da parte dei cittadini, poiché una norma di diritto perfetta dal punto tecnico, ma che risulta incomprensibile da parte della maggior parte della popolazione, può risultare non efficente.

Per quanto riguarda il linguaggio in un processo legale, in questo ambito il linguista è chiamato ad esaminare come si svolgono le varie interazioni giudiziarie, con lo scopo di verificare come e quanto il linguaggio verbale possa condizionare un procedimento penale e possibilmente, incidere sulla sua equità e comprensione. Nello specifico, durante un dibattimento o un interrogatorio, si osserva una chiara asimmetria di potere[20]: vi sono determinate figure, rispettivamente giudici e avvocati e pubblici ministeri, che, dato il ruolo che è a loro riconosciuto, guidano la conversazione per garantire l'ordine e ricostruire un dato avvenimento. Tuttavia in tale sistema ci sono delle aree problematiche, dettate esattamente da questa configurazione, sulle quali un linguista è chiamato a intervenire. Le questioni più rilevanti riguardano il rischio di alterare le testimonianze rese, tanto in un dibattimento quanto durante un interrogatorio, attraverso le cosiddette domande suggestive poste dall'autorità competente; esse vengono definite come domande che tendono a suggerire le risposte[21]

[19] «Se si ritiene che un testimone abbia deliberatamente mentito su un fatto importante, si dovrebbe prendere in considerazione la possibilità di non credere a nessuna delle sue dichiarazioni. Oppure, se si ritiene che il testimone abbia mentito su alcuni aspetti ma abbia detto la verità su altri, si può semplicemente accettare le parti che si ritengono veritiere e ignorare il resto» (California State Portal, 2005). Traduzione a cura della laureanda.

[20] Coulthard M. (2010). *Forensic Linguistics: the application of language description in legal contexts*. Langage et société.

[21] Tkačuková T. (2010). *The power of questioning: A case study of courtroom discourse*. DISCOURSE and INTERACTION.

e, in quanto tali, nell'ordinamento italiano sono vietate dall'articolo 499 c.p.p.[22]. Di seguito si propongono due esempi di domande suggestive: «L'auto dell'imputato era bianca o nera?»; «L'auto dell'imputato era bianca o no?»

Questi esempi mostrano come una domanda possa orientare, anche involontariamente, la risposta dell'interlocutore. In entrambi i casi, infatti, si presuppone che il testimone abbia percepito il colore della macchina a cui si sta facendo riferimento; in aggiunta, nel secondo esempio, l'alternativa proposta («bianca o no») tende a privilegiare una delle due opzioni. Tali formulazioni possono quindi influenzare il ricordo della persona a cui viene formulata la domanda, che può essere indotto a fornire una risposta conforme alle informazioni implicitamente contenute nella domanda; pertanto uno studio meticoloso delle parole e della sintassi che si sceglie di utilizzare, soprattutto nei contesti in cui è richiesta la massima neutralità, risulta essenziale.

È precisamente in questa prospettiva che si inserisce la figura del linguista forense il cui compito è quello di identificare gli elementi grammaticali e linguistici che tendono a suggerire una risposta secondo l'intento di chi conduce l'interrogatorio, con l'obiettivo di delineare delle linee guida per formulare al meglio le domande.

Un'altra problematica legata al linguaggio parlato nei processi penali riguarda la gestione dell'interrogatorio o l'ascolto di testimonianze rese da testimoni vulnerabili. Data la vulnerabilità della loro testimonianza, dovuta principalmente all'età, alle condizioni fisiche o psichiche, alla posizione rispetto a peculiari categorie di reati (Algeri, 2017), in un procedimento penale in cui la prova si costruisce sulla base di esami, controesami e domande spesso incalzanti, i rischi sono che un testimone vulnerabile non comprenda le domande che gli vengono poste, che la sua memoria venga alterata dalla pressione a cui è sottoposto, che risponda in modo compiacente poiché influenzato o che si trovi in difficoltà a raccontare gli eventi traumatici vissuti. In questo ambito, nel corso degli anni, sono stati apportati dei cambiamenti per ridurre al minimo la possibilità di alterare le dichiarazioni di questi testimoni; tuttavia, il lavoro di un esperto di linguistica rimane centrale nell'assistere la persona, per evitare che questa venga esposta a domande suggestive e per assicurarsi della piena comprensione delle domande che gli vengono poste.

Quest'ultimo punto, nonostante possa essere ritenuto di poco conto, permette invece non solo di rendere accessibile il linguaggio giuridico anche a coloro che non sono del settore, ma consente agli indagati/imputati di esercitare a pieno il proprio diritto alla difesa[23], che non si esaurisce nella assistenza di un avvocato ma concerne anche la possibilità di comprendere le contestazioni formulate e di partecipare consapevolmente al processo.

[22] «(...) Nell'esame condotto dalla parte che ha chiesto la citazione del testimone e da quella che ha un interesse comune sono vietate le domande che tendono a suggerire le risposte. (...)» Art. 499 c.p.p.

[23] «Tutti possono agire in giudizio per la tutela dei propri diritti e interessi legittimi (...). La difesa è diritto inviolabile in ogni stato e grado del procedimento. (...)» Art. 24 della Costituzione Italiana.

L’ultimo ambito di applicazione della linguistica forense riguarda i suoi impieghi investigativi che includono l’identificazione dell’autore di uno scritto anonimo o la cui paternità è incerta o contestata, l’individuazione di informazioni utili in testi redatti con finalità criminali ai fini di redigere un profilo dell'autore, la verifica della veridicità di una confessione o di una testimonianza, il riconoscimento di un plagio e l’accertamento dell’identità di un parlante. Nel seguente lavoro di ricerca il settore su cui si pone una maggiore attenzione è quello della profilazione autoriale.

2.3 La profilazione autoriale nella linguistica forense

La profilazione autoriale viene definita da Andrea Nini (2018) come il lavoro di determinare informazioni sul *background* dell'autore di un testo anonimo in base alla lingua del testo; il linguista forense dovrà dunque rilevare quelli che appaiono come elementi linguistici inusuali, per poi tentare di costruire un profilo sulla base di tali componenti.

Questa disciplina si basa sull’idea che ognuno di noi possieda la propria “impronta linguistica”: gli esseri umani, in quanto esseri abituali, tendono a comportarsi in modo ripetitivo e sviluppare abitudini che rimangono nel tempo e ciò accade anche con la scrittura[24]. Quando una persona scrive, infatti, le scelte linguistiche che compie sono spesso automatiche e seguono lo stile di scrittura che la persona ha sviluppato in base all'istruzione che ha ricevuto, all’ambiente sociale in cui è vissuto o ai contesti in cui è solito scrivere. È importante sottolineare come tali elementi, centrali per la costruzione di un possibile profilo autoriale, non emergano dall’intuizione del linguista, ma derivino da metodi di analisi che nel corso degli anni sono stati consolidati grazie ad una serie di studi condotti (van Halteren, 2004; Nini, 2015; Roemling *et al.*, 2024). Sebbene le intuizioni e la conoscenza personale del linguista possano risultare utili e rendere ancora più specifico il profilo, soprattutto quando entrano in gioco elementi culturali o dialettali, la profilazione non può basarsi esclusivamente su questo; è necessario che il lavoro svolto si basi su metodi scientifici e che le tecniche utilizzate siano applicabili anche su casi differenti, cosa che la mera intuizione del linguista difficilmente può fare.

Gli studi sopra citati, infatti, sottolineano l'importanza della stilometria nell'ambito della linguistica forense: grazie a questo approccio è possibile studiare, in maniera oggettiva, lo stile linguistico di un testo anonimo attraverso l’identificazione di tratti misurabili statisticamente. Per effettuare questo tipo di analisi la *feature engineering* risulta centrale. Con tale espressione si fa riferimento al processo che trasforma il testo anonimo, che si intende analizzare, in variabili quantificabili definite come ‘*feature*’; in altre parole si esamina lo scritto, identificando i tratti linguistici ricorrenti e distintivi che costituiscono la specifica impronta linguistica dell’autore, per

[24] Misini A., Kadriu A., Canhasi E. (2022). *A Survey on Authorship Analysis Tasks and Techniques*. SEEU Review.

poi trasformare questi dati in *feature* misurabili con lo scopo di verificarne la frequenza nel testo e ricavare il *pattern* linguistico che contraddistingue lo scritto.

L'accuratezza di tale processo dipende direttamente dalla qualità dell'analisi delle *feature*: Arta Misini, Arbana Kadriu e Ercan Canhasi (2022) dimostrano infatti come questi componenti non debbano essere esaminati esclusivamente per come si presentano, ma la loro analisi deve avere una rilevanza linguistica. Ciò implica che la presenza di determinate *feature* assume un significato nel momento in cui consente di individuare pattern linguistici relativamente stabili nel testo e dunque riconducibili all'idioletto dell'autore, tenendo in considerazione però il registro e il contesto comunicativo in cui il testo è stato redatto. Due testi formali prodotti da due magistrati diversi possono presentare le medesime caratteristiche linguistiche proprie del registro formale, quali complessità sintattica o l'uso di un lessico specialistico; in questo caso gli elementi su cui si concentrerà l'attenzione saranno quelli che riflettono delle preferenze linguistiche del singolo autore, come la modalità scelta per organizzare il discorso, l'uso di particolari e ricorrenti costruzioni sintattiche o specifiche espressioni. Il valore delle *feature* varia, dunque, in base alla tipologia di testo che si sta analizzando, poiché ricercare determinati elementi in uno scritto può risultare funzionale, ma non necessariamente quello stesso componente funziona in un caso differente.

Per svolgere questo tipo di analisi al meglio, è necessario partire da una classificazione delle *feature*, per poi soffermarsi su come la loro variazione e la loro presenza (o assenza) in un testo possano suggerire determinati tratti linguistici. Le *feature* vengono classificate in cinque categorie:

- *feature* lessicali
- *feature* sintattiche
- *feature* semantiche
- *feature* strutturali
- *feature* contenutistiche

2.3.1 *Feature* lessicali

Per *feature* lessicali si intendono quelle caratteristiche che descrivono quali caratteri e parole un autore sta utilizzando, con lo scopo di dedurre quali scelte lessicali ricorrono più frequentemente. L'analisi che viene svolta è su più livelli, poiché si parte da un indagine sui caratteri utilizzati, per poi spostarsi alle parole, osservandone la lunghezza media, i termini scelti e concludere con l'esaminare intere frasi dove si misura la frequenza con la quale determinati vocaboli appaiono nei periodi.

Un ulteriore elemento che rientra in questa categoria di *feature* sono le cosiddette "parole funzionali", dal momento che sono considerate degli importanti indicatori stilistici: con l'espressione "parole funzionali" si fa riferimento a quegli elementi linguistici che non trasmettono alcun concetto, ma costruiscono la grammatica del testo[25]. Essendo delle unità indipendenti dall'argomento trattato tendono a rimanere costanti nel tempo e in contesti anche differenti; di conseguenza il loro uso è spesso automatico ed è precisamente per questo motivo che forniscono informazioni importanti sull'impronta linguistica dell'autore, un elemento anch'esso inconscio.

2.3.2 *Feature* sintattiche

Le *feature* sintattiche racchiudono caratteristiche quali la costruzione delle frasi e le relazioni grammaticali tra le parole. In questo contesto risulta particolarmente utile il *POS tagging* [etichettatura grammaticale] che assegna una precisa categoria grammaticale sostantivo, verbo, aggettivo, articolo, etc.) a una parola.

2.3.3 *Feature* semantiche

Questa categoria non risulta ampiamente utilizzata dal momento che richiede un'analisi più complessa che non si basa esclusivamente sull'osservazione della forma del testo ma tenta di interpretare il significato delle parole e delle frasi. In altre parole, si concentra maggiormente sul comprendere il contenuto, un aspetto difficilmente quantificabile e proprio per questo, scarsamente usato.

2.3.4 *Feature* strutturali

Nello spiegare le *feature* strutturali è centrale comprendere la differenza rispetto alle *feature* sintattiche: le prime rappresentano la forma complessiva del testo piuttosto che la struttura grammaticale delle singole frasi. In questa categoria, dunque, il *focus* si sposta alla costruzione e disposizione spaziali dei paragrafi, ovvero il numero di frasi per paragrafo, la presenza di eventuali spazi bianchi o righe vuote, impaginazione del testo.

2.3.5 *Feature* contenutistiche

Questo particolare tipo di *feature* riflettono i temi e gli argomenti del testo. A differenza delle *feature* semantiche, tali informazioni non sono utilizzate con lo scopo di interpretare il significato del testo, ma consistono nell'individuazione di parole chiave, espressioni ricorrenti o legate a un

[25] Stamatatos E. (2008). *A survey of modern authorship attribution methods*. Journal of the American Society for Information Science and Technology.

tema specifico, tali da mettere in evidenza le preferenze tematiche e, talvolta, l'ideologia dell'autore.

2.4 Correlazione tra *feature* linguistiche e dati demografici

Una volta classificate le varie *feature* linguistiche, è possibile individuarle nel testo anonimo che si sta esaminando. Il passaggio successivo è analizzarle dal punto di vista quantitativo, ovvero verificare quali *feature* sono presenti, quanto spesso si presentano nel testo e quali tra queste dominano rispetto ad altre.

Questa loro variazione, direttamente legata all'impronta linguistica dell'autore, è l'elemento che, possibilmente, può far emergere dati demografici, quali genere, età, livello di educazione e classe sociale, centrali per la costruzione di un profilo. Tuttavia, è necessario specificare che le deduzioni che ne derivano non sono certe, poiché provengono da correlazioni statistiche osservate in *corpora* linguistici, ovvero raccolte di testi**,** di cui si conoscono le informazioni degli autori, analizzati per osservare l'esistenza di una correlazione tra tratti linguistici e gruppi sociali. Tale correlazione è stata successivamente confermata da numerosi studi empirici e stilometrici svolti nel corso degli anni; il profilo autoriale costruito in base a queste informazioni, quindi, non può essere considerato definitivo, ma solo un'ipotesi di profilazione, motivata da prove linguistiche.

La maggior parte degli studi condotti utilizza la *feature engineering* secondo uno schema ben preciso: per prima cosa, si identifica un *corpus* di testi adeguato e stratificato, ovvero si raccoglie un insieme di testi prodotti da autori appartenenti a gruppi sociali differenti la cui selezione non è casuale. I testi che rientrano nel *corpus* devono essere in linea con il registro individuato nel testo anonimo: possedere un insieme di testi simili in quanto a stile, permette di stabilire quale sia il comportamento che ci si aspetta dalle varie *feature* in una determinata tipologia testuale. Nel caso in cui si confrontasse uno scritto accademico, caratterizzato da uno stile più nominale e formale, con un testo narrativo, le variazioni delle *feature* che si rileverebbero tra i due testi non dipenderebbero da fattori diastratici, bensì da fattori diafasici. Al contrario, la possibilità di accostare il testo anonimo a una serie di testi simili permette di rilevare eventuali somiglianze o differenze, rispetto ai testi con i quali è stato confrontato e interpretare correttamente le variazioni e la frequenza delle *feature* come possibili indicatori di determinati gruppi sociali.

Dopo aver costruito il *corpus* di riferimento, dunque, si analizzano e misurano le *feature* presenti in ciascun testo, distinguendole nelle varie categorie (lessicali, sintattiche, semantiche, strutturali e basate sul contenuto). Successivamente, per ciascun fattore sociale, quali il genere, l'età, la classe sociale e il livello di istruzione, si osserva come i vari tratti linguistici variano in relazione ai diversi gruppi sociali e si selezionano quelli che mostrano le differenze più evidenti, ovvero quelli

maggiormente informativi rispetto a un preciso dato demografico. Infine, si confrontano i risultati ottenuti dall'analisi di ciascun testo e si costruisce un modello statistico in grado di stimare la probabilità di appartenenza di uno specifico pattern linguistico a determinate categorie sociolinguistiche. Attraverso le correlazioni tra tratti linguistici e gruppi sociali, il modello permette di formulare ipotesi sul genere, la fascia d'età, il livello di istruzione e lo *status* sociale dell'autore.

Di seguito si riportano i risultati presentati da Andrea Nini nella sua tesi *Developing forensic authorship profiling* (2018), risultati che si basano su inferenze statistiche estrapolate da diversi studi condotti su *corpus* di testi molto ampi.

Per quanto riguarda la deduzione del genere dell'autore di un testo anonimo, ci si basa sulla differenza sostanziale tra lo stile di scrittura maschile e quello femminile. Uno stile nominale in cui la scelta del nome, unita a quella dell'aggettivo e dei verbi in modi non finiti, prevale sulla scelta del verbo di modo finito[26], appare più comune negli scritti prodotti da autori maschili; ci si interfaccia dunque ad un testo le cui caratteristiche distintive sono la presenza di sostantivi derivati da verbi (nominalizzazione), frequenti enumerazioni e un lessico generalmente astratto. Sulla base di questi elementi la sintassi risulta essere maggiormente paratattica, poiché si tende a costruire le frasi intorno a un sostantivo creando così dei blocchi ricchi di informazioni spesso giustapposti tra di loro. Al contrario, gli scritti di autrici femminili mostrano uno stile tendenzialmente verbale, caratterizzato da periodi di tipo ipotattico, ovvero basati sulla subordinazione, la cui costruzione ruota attorno a un verbo coniugato.

Le *feature* che mostrano una variazione più rilevante per questo tipo di deduzione sono dunque quelle sintattiche, dal momento che l'osservazione della costruzione delle frasi permette di sostenere se lo stile adottato nel testo sia nominale o verbale. Risultano però interessanti anche le *feature* lessicali, le quali possono contribuire a inferire con maggiore precisione il genere dell'autore, rivelando se la tipologia di lessico adottata è in linea con lo stile, nominale o verbale. Vale la pena menzionare anche le *feature* basate sul contenuto, la cui variazione può rivelare informazioni sul genere; tuttavia gli studi condotti mostrano come queste disuguaglianze siano spesso piccole, ovvero osservate in pochi casi e per questo vengano prese in considerazione con cautela. Sono numerosi, invece, gli studi che confermano le differenze sintattiche e lessicali; tra questi è interessante citare Säily *et al.* (2001), che avevano già osservato questa differenza studiando un *corpus* di 660 lettere scritte tra il 1415 e il 1681.

In merito al dato demografico relativo all'età, gli studi di *profiling* linguistico evidenziano una discrepanza tra l'interesse applicativo delle forze dell'ordine e ciò che emerge concretamente

[26] Istituto della Enciclopedia Italiana. (n.d.). Stile nominale. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/stile-nominale_(Enciclopedia-dell'Italiano)/.

dall'analisi testuale: le prime sarebbero maggiormente orientate verso la determinazione dell'età biologica dell'autore, ma le variazioni linguistiche, ovvero gli unici elementi in un testo anonimo in grado di suggerire informazioni per la profilazione, risultano essere legate alla cosiddetta età sociale, ovvero l'età di una persona modellata sulla base delle esperienze vissute e contesti esplorati. Determinati eventi sociali, quali l'ingresso nel mondo del lavoro o il raggiungimento di determinate tappe anagrafiche, comportano di fatto l'esposizione a dei contesti che necessitano l'uso, e quindi l'acquisizione, di nuovi registri linguistici, talvolta maggiormente formalizzati.

In questo ambito risulta interessante ciò che Kemper, già nel 1987[27], aveva osservato: l'esistenza di una correlazione negativa tra la complessità sintattica e l'invecchiamento. Ciò implica che più una persona avanza con l'età, meno tende a utilizzare frasi complesse dal punto di vista sintattico. Kemper spiega questa tendenza dandovi due interpretazioni differenti: anzitutto menziona il ruole della cosiddetta *working memory* [memoria di lavoro] ovvero un sistema cognitivo che permette di mantenere temporaneamente attive alcune informazioni mentre se ne elaborano altre, capacità che con l'invecchiamento tende a diminuire; il fatto che la complessità sintattica si riduca parzialmente a causa della diminuzione della *working memory* comporta che i parlanti anziani producono frasi più brevi e semplici poiché la produzioni di frasi sintatticamente complesse richiede una notevole abilità di memoria. Secondo tale interpretazione, dunque, la propensione verso uno stile nominale e semplice appare parzialmente attribuibile all'età biologica; tuttavia Kemper propone una seconda ipotesi secondo la quale tale tendenza è una conseguenza diretta di esperienze pregresse attraverso le quali un individuo impara metodi più efficaci per comunicare, che gli consentono di evitare l'uso di strutture eccessivamente complesse. Il soggetto, dunque, diventa progressivamente maggiormente capace ad esprimere idee con parole precise e accuratamente scelte, alleggerendo così la sintassi. È precisamente questo fenomeno che mostra come la propensione verso l'utilizzo di uno stile più nominale e quindi sintatticamente più semplice, non sia direttamente riconducibile all'età in senso strettamente biologico, ma piuttosto derivi dall'età sociale, che dal punto di vista del *profiling* linguistico risulta più interessante da ricostruire.

Un elemento ulteriore che conferma la rilevanza dell'età sociale rispetto all'età biologica è la ricchezza lessicale, un dato che non deriva necessariamente dall'anzianità di un individuo, ma piuttosto dal livello di istruzione al quale si è o si sta avendo accesso, le abitudini di lettura e scrittura e l'attività lavorativa che si sta svolgendo o si è svolta. Questi fattori infatti permettono all'autore di entrare in contatto con contesti linguistici differenti che richiedono competenze specifiche e che favoriscono l'accrescimento del repertorio lessicale

[27] Nini A. (2018). Developing forensic authorship profiling. Language and Law / Linguagem e Direito, Vol. 5(2), p. 38-58.

Considerata la natura di queste variazioni nei vari gruppi appartenenti al dato demografico dell'età, le *feature* su cui un linguista si focalizza sono principalmente quelle di tipo sintattico e lessicale, con una particolare attenzione alla ricchezza lessicale.

Un altro dato interessante da ricavare da un testo anonimo è quello riguardante la classe sociale. A livello diastratico, gli studi mostrano una differenza significativa del livello di complessità tra testi prodotti da individui appartenenti a classi sociali differenti: uno scritto prodotto da un individuo di una classe sociale più elevata è spesso caratterizzato da una maggiore complessità linguistica, al contrario di quelli associati a individui di classi sociali più basse. Di conseguenza, l'elemento su cui si pone maggiore attenzione sono le *feature* sintattiche e in quest'ottica risulta particolarmente efficace l'analisi delle *T-unit*.

Una *T-unit* (*terminal unit*) è definita come la più breve unità grammaticale in cui può essere diviso un testo[28]; si intende quindi una proposizione principale seguita da tutte le subordinate che dipendono da essa. A differenza di una frase che dipende dalla punteggiatura, termina con l'inserimento di un punto fermo, punto esclamativo o punto interrogativo e può contenere più proposizioni, le *T-unit* contengono esclusivamente una sola proposizione principale e le sue subordinate. Di seguito si riportano degli esempi:

«Lei è andata a casa perché era stanca.». In questo caso c'è una frase ed una *T-unit* dove la principale è «Lei è andata a casa» a cui segue la subordinata.

«Lei è andata a casa e ha chiamato una sua amica.». In questo esempio si osserva sempre una frase, ma due *T-unit* poiché vi sono due proposizioni coordinate.

«Lei è andata a casa. Quando è arrivata ha chiamato una sua amica.». In quest'ultimo esempio vi sono due frasi, data la presenza di un punto fermo e due *T-unit*.

Questo concetto linguistico permette quindi di andare oltre la lunghezza delle frasi e soffermarsi invece sulla loro struttura interna, esaminando due aspetti principali: la stratificazione delle subordinate e le informazioni inserite in una singola unità. Questo implica che maggiore è il numero di subordinate e parole inserite in una singola *T-unit*, maggiore è la complessità della sintassi e gli studi mostrano come le persone appartenenti alle classi sociali più alte siano in grado di gestire meglio strutture più complesse di tipo ipotattico.

Lo stesso fenomeno si rileva negli scritti di individui che hanno ricevuto una maggiore istruzione. Tuttavia, in questi casi emerge anche l'uso di un lessico più specifico e tecnicamente preciso. Questo suggerisce che il livello di istruzione influenzi non solo la struttura della frase ma

[28] Thomas M. (). *The T-unit as a measure of 'syntactic maturity': operationalizing linguistic complexity in early generative grammar*. Simplicité et complexité des langues dans l'histoire des théories linguistiques. Pag. 428.

anche la scelta lessicale, contribuendo così a rimarcare ulteriormente la correlazione tra livello di competenza linguistica e contesto socioculturali.

Dopo aver discusso questa serie di dati si può concludere che le *feature* lessicali e sintattiche sono quelle che maggiormente indicano una variazione diafrastrica, sulla base della quale si possono elaborare delle ipotesi riguardo le caratteristiche proprie dell'autore del testo anonimo. Nonostante ciò, è bene analizzare anche le *feature* semantiche, strutturali e basate sul contenuto dal momento che queste possono fornire interessanti indizi riguardo la tipologia di testo che ha redatto l'autore, le sue intenzioni e la visione della realtà; pertanto, arricchiscono il profilo che si è elaborato, favorendo talvolta una redazione più corretta.

Nel capitolo successivo si ripercorre quindi l'intero processo che porta alla costruzione di un profilo di autore concreto, con lo scopo di riassumere tutti i passaggi seguiti durante la fase di analisi linguistica del *Manifesto*.

2.5 Metodo di profiling: dall'analisi del contenuto all'analisi quantitativa delle *feature*

Come già anticipato nel primo capitolo del seguente lavoro di ricerca, la profilazione linguistica richiede un approccio integrato che unisca l'analisi del contenuto e l'esame dal punto di vista linguistico. Pertanto, un linguista che si approccia ad un testo anonimo per la prima volta deve partire precisamente dall'analisi del contenuto

Questo tipo di analisi risulta utile per cogliere i temi centrali dello scritto, le intenzioni dell'autore e il possibile lettore ideale al quale è indirizzato il testo. Informazioni quali l'argomento trattato e il rapporto tra autore e lettore consentono infatti di ipotizzare il contesto in cui lo scritto è stato prodotto e, di conseguenza, individuare il registro adottato.

Tuttavia, il registro emerge con maggiore chiarezza attraverso l'analisi delle varie *feature*, lessicali, sintattiche, strutturali, semantiche; in questa fase successiva l'attenzione non viene posta sulle *feature* basate sul contenuto poiché rappresentano la fase iniziale dalla quale prende il via l'analisi linguistica del *Manifesto* e per questo sono già state affrontate nel paragrafo 1.2 della seguente ricerca. Durante questa fase di esame delle *feature* quindi il primo passo consiste nell'individuare per ogni tipologia di *feature* le caratteristiche linguistiche presenti nel testo. Per quanto riguarda le *feature* lessicali l'attenzione sarà rivolta verso l'individuazione del tipo di lessico che l'autore utilizza (fondamentale, di alto uso o di alta disponibilità); in questa prospettiva si farà ricorso ad alcuni indici linguistici, quali il *Type-Token Ratio (TTR)* e l'*F-score* di Heylighen & Dewaele. Altri elementi che si analizzeranno saranno i termini maggiormente ripetuti all'interno del testo e la presenza o assenza di errori ortografici.

In merito alle *feature* sintattiche l'obiettivo dell'analisi sarà quello di individuare il grado di complessità sintattica e la tendenza a uno stile maggiormente nominale o verbale; in quest'ottica risulta particolarmente efficace la misurazione delle *T-unit* con lo scopo di calcolare quali sono le relazioni gerarchiche che legano le frasi all'interno dei vari periodi del *Manifesto* maggiormente utilizzate.

L'analisi delle *feature* si concluderà con l'analisi delle *feature* strutturali e semantiche. Nonostante rappresentino la parte conclusiva di questo di esame, esse ricoprono un ruolo centrale poiché sono rispettivamente in grado di evidenziare la struttura che l'autore adotta per presentare al lettore le sue argomentazioni e possono fornire una possibile interpretazione di determinate scelte lessicali inserite nel testo. I dati che ne risultano possono rivelarsi utili per confermare e rafforzare ulteriormente le ipotesi emerse riguardo le intenzioni dell'autore dedotte già dall'analisi del contenuto.

Dopo aver estrapolato e misurato le *feature* presenti nel *Manifesto* è possibile interpretare i dati emersi sulla base di quanto Andrea Nini (2018) presenta nella sua tesi e inferire una ipotetica correlazione con dei dati demografici riconducibili all'autore.

CAPITOLO 3 - LO STATO DEGLI STUDI SUL MANIFESTO *INDUSTRIAL SOCIETY AND ITS FUTURE*

3.1 Studi e interpretazioni del *Manifesto* (storici, ideologici e criminologici)

Dopo aver delineato, in sintesi, le principali caratteristiche della linguistica forense e gli strumenti teorici e metodologici a cui si farà ricorso, in questo capitolo ci si concentrerà sull'oggetto specifico della presente ricerca, ovvero il manifesto *Industrial Society and Its Future*. Prima di procedere all'analisi linguistica, risulta interessante ricostruire lo stato degli studi esistenti, con lo scopo di comprendere quali interpretazioni siano già state prese in esame e quali aspetti risultino ancora poco approfonditi.

I principali studi condotti sul *Manifesto* si focalizzano sulla rilevanza storica e ideologica dell'opera. Lo scritto nasce a metà degli anni Novanta (1995), un periodo caratterizzato da un'accelerazione della globalizzazione e della rivoluzione digitale, già iniziata intorno agli anni Cinquanta. Questi due fenomeni hanno profondamente trasformato il mondo, riducendo le barriere geografiche per lasciare spazio a un sistema interconnesso in cui merci, persone, capitali, informazioni si muovono con maggiore libertà e rapidità. Di conseguenza, le nuove opportunità lavorative e di vita che ne derivano e la società digitale in ascesa sono generalmente accompagnate da una visione positiva verso il progresso: le innovazioni, l'accorciamento delle distanze e la diffusione di tecnologie informatiche vengono interpretate come fattori di crescita economica e di modernizzazione sociale, tutti elementi che potenzialmente apportano degli sviluppi positivi nella vita dei cittadini.

È in questo contesto che l'opera di Theodore Kaczynski si inserisce con una posizione che si può definire controcorrente, ovvero come una forte critica verso il progresso tecnologico che si interroga sui suoi reali effetti psicologici e sociali. Nonostante nel periodo in cui il *Manifesto* è stato prodotto il dibattito riguardo queste tematiche tendesse a rappresentare lo sviluppo tecnologico in maniera tendenzialmente positiva, le prospettive più negative non mancavano; tuttavia data la dominanza di una tendenza ideologica favorevole, il *Manifesto* acquisisce una certa rilevanza agli occhi degli studiosi.

Data l'ampiezza dello scritto e le vaste argomentazioni (elaborate in 232 sezioni), l'opera risulta essere un riferimento significativo per la comprensione delle ragioni intrinseche che muovono le critiche verso la società industriale, critiche che, dato il diffuso ottimismo nei confronti dello sviluppo, risultavano quasi anticonformiste. Per questo motivo sono numerosi gli studi che si focalizzano sull'interpretazione del contenuto in relazione al periodo storico in cui è stato scritto, perché un'analisi di questo tipo consente non solo di valutarne la ragionevolezza, ma permette di

verificare se tali critiche siano in qualche modo fondate. Questo elemento è ciò che muove gli studi in questo ambito, poiché riconoscere e comprendere le critiche che vengono mosse nei confronti di un sistema, se ragionevoli, possono individuare gli eventuali limiti strutturali e possibilmente gettare le basi per un suo possibile miglioramento.[29]

Un esempio concreto di tali studi lo si ritrova nel documento *From traditional society to industrial society (An anthropological analysis of the manifesto Ted Kaczynski)* redatto da Jamil e Saprijal (2025). Le riflessioni presentate dagli autori evidenziano come la visione di Kaczynski rovesci completamente la narrazione dominante riguardo la nascente società industriale percepita come positiva poiché indirizzata verso il progresso e il miglioramento delle condizioni lavorative. Al contrario nel *Manifesto* si muove una critica verso questa, percepita invece come una forma di alienazione e deumanizzazione. Gli autori quindi non condividono la sua posizione, ma la considerano utile come critica antropologica riguardo i costi possibili derivanti dalla modernità industriale.

A ciò si aggiungono le analisi prettamente ideologiche, volte a esaminare la tesi portata avanti nel *Manifesto* e i valori che l'autore intende preservare. L'interesse primario di questi studi non si esaurisce nella comprensione dei temi cardine attorno ai quali è costruito il pensiero di Kaczynski, né nell'analisi delle modalità attraverso le quali l'ideologia è articolata; ciò che risulta particolarmente rilevante è l'attualità di determinate argomentazioni presentate. Temi quali l'aumento del controllo sociale e la conseguente riduzione dell'autonomia individuale rivelano meccanismi che non appartengono esclusivamente agli anni Novanta ma sono osservabili anche oggi, se non in misura più marcata: l'avvento dei *social* ha infatti favorito la nascita di strumenti per la raccolta di dati sempre più efficienti e indiretti e l'ampia normalizzazione di queste pratiche, talvolta presentate come condizioni necessarie per l'accesso a determinati servizi, ha fatto passare in secondo piano l'effetto negativo che queste hanno nei confronti della autonomia personale. È proprio la persistenza e il rafforzamento di queste dinamiche che rende il *Manifesto* un'opera non solo dei suoi tempi, ma anche uno scritto centrale per l'interpretazione di dinamiche contemporanee, alimentando così i dibattiti su questi temi: mantenere viva la riflessione su queste questioni è ciò che permette di acquisire consapevolezza e sviluppare uno sguardo critico nei confronti di tali fenomeni ed è in quest'ottica che si inseriscono appunto gli studi sull'ideologia del *Manifesto*.[30]

Un'analisi del *Manifesto* particolarmente rilevante è quella condotta da Douglas e Youvan, *A Critical Analysis of the Unabomber's Manifesto: Industrial Society and Its Future* (2024). Il passaggio più interessante e che riassume l'obiettivo dei due autori è la frase "Comprendere le cause

[29] Jamil S. M., Saprijal. (2025). *From traditional society to industrial society (An anthropological analysis of the manifesto Ted Kaczynski)*. Journal Sulôh Nanggroe.
[30] Youvan D. C. (2024). *A Critical Analysis of the Unabomber's Manifesto: Industrial Society and Its Future*. ResearchGate.

profonde di questo malessere, come i sentimenti di alienazione, il degrado ambientale e l'erosione dei valori tradizionali, può contribuire a orientare gli sforzi volti a costruire un futuro più umano e sostenibile.”. Tale riflessione risulta centrale perché gli autori però non si limitano a esporre le tematiche trattate da Kaczynski, né condividono le sue argomentazioni, ma fanno un passo ulteriore: nel riportare le tematiche principali che muovono la critica propongono uno spunto per riflettere sulle motivazioni che muovono tali critiche che vale la pena attenzionare affinché la società in quanto tale possa migliorare.

Infine, la letteratura esistente si concentra anche sulla rilevanza del *Manifesto* nel caso giudiziario che ha coinvolto Theodore Kaczynski, autore dello scritto. Nel 1995 Unabomber comunica per la prima volta alle forze dell'ordine la sua intenzione di pubblicare il *Manifesto* sulle principali testate giornalistiche. Questo evento rappresenta una svolta decisiva nelle indagini, poiché costituisce il primo momento in cui Unabomber lascia una traccia potenzialmente riconoscibile della sua identità**;** questo elemento era qualcosa che fino a quel momento non era mai pervenuto sulle scene del crimine o dall'analisi degli ordigni esplosi ed era dunque una circostanza che aveva contribuito in modo significativo a far sì che le indagini si protrassero per diversi anni. A seguito dell'effettiva pubblicazione sul Washington Post infatti, le indagini subiscono un'accelerazione notevole che conduce nel giro di pochi mesi all'arresto di Kaczynski e alla sua successiva condanna, a seguito dell'accertamento della sua identità.

L'elemento su cui la letteratura si concentra sono quindi le motivazioni che hanno portato Kaczynski a richiedere la pubblicazione di un qualcosa che era potenzialmente riconducibile direttamente a lui, specialmente dopo essere stato così cauto a non lasciare prove di sé per diversi anni[31]. Inoltre gli studiosi hanno mostrato interesse nei confronti della successiva comparazione svolta tra il *Manifesto* e gli scritti ritrovati nella sua cabina nel bosco, poiché tale confronto è stato determinante nell'identificare Kaczynski come Unabomber in maniera ufficiale[32].

Un saggio particolarmente rilevante è quello di Roberta Melazzo, *Linguistica forense: l'emblematico caso Unabomber* (2021) nel quale si ripercorre l'analisi scientifica del linguaggio. L'autrice evidenzia come tale analisi ha permesso di restringere i sospettati a Theodore Kaczynski e identificarlo come autore del fino a quel momento anonimo *Manifesto* del quale era stata chiesta la pubblicazione dal soggetto, anch'esso anonimo, *Unabomber*.

[31] Oleson, J. C. (2023). *A requiem for the Unabomber*. Contemporary Justice Review.
[32] Melazzo R. (2021). *Linguistica forense : l'emblematico caso Unabomber.* Mimesis Edizioni.

3.2 Limiti degli studi esistenti

Ciò che emerge dalla varietà degli studi esistenti è la mancanza di un approfondimento di ciò che in questo lavoro di ricerca risulta curioso, ossia il ruolo dell'analisi linguistica del *Manifesto*.

Secondo il IV emendamento della Costituzione degli Stati Uniti[33], i cittadini sono tutelati da perquisizioni e sequestri se questi ultimi non non sono motivati da mandati emessi in presenza di una *probable cause*. La *probable cause* rappresenta lo standard legale del diritto statunitense che definisce i fatti e circostanze necessarie per ritenere che un reato sia stato commesso e che una persona ne sia responsabile; questo implica che le prove che vengono raccolte devono essere più che sospetti ragionevoli, ma non è necessario che provino l'esistenza del reato e la colpevolezza di un individuo oltre ogni ragionevole dubbio.

Nella vicenda giudiziaria che riguarda Unabomber questo principio risulta particolarmente significativo poiché è il componente che definisce la rilevanza dell'analisi linguistica svolta sul *Manifesto*. È infatti necessario precisare che il mandato di perquisizione della sua cabina non fu emesso esclusivamente sulla base delle informazioni ottenute da questo esame, soprattutto considerata l'epoca in cui si svolge la vicenda, il livello di sviluppo della disciplina della linguistica forense e i criteri di ammissibilità delle prove; al tempo le sole prove linguistiche non erano considerate elementi sufficienti per configurare la *probable cause*.

Per questo motivo il ruolo dell'analisi linguistica può essere descritto in maniera più adeguata utilizzando il termine 'catalizzatore', inteso come un elemento che velocizza il processo investigativo, orientando le indagini, fino a quel momento frammentate, verso un possibile profilo identificabile. È precisamente il ruolo da catalizzatore che le prove linguistiche hanno avuto nella vicenda l'elemento che risulta interessante, dal momento che, la convergenza di queste informazioni, sulle quali si è costruito un possibile profilo e la conferma di quest'ultimo da parte di una persona molto vicina a Kaczynski, ha portato all'emissione di un mandato.

La rilevanza dell'analisi del *Manifesto* dunque, non risiede nella capacità di identificare in modo univoco l'autore di uno scritto, poichè dare un nome e un cognome a colui che ha redatto un testo anonimo sulla base di inferenze linguistiche risulta complicato; ciò che è possibile è indirizzare le indagini verso un obiettivo più mirato, accelerando così il processo di identificazione di un individuo.

[33] «Il diritto delle persone a essere garantite nella sicurezza della propria persona, della propria abitazione, dei propri documenti e dei propri beni contro perquisizioni e sequestri irragionevoli non potrà essere violato; e nessun mandato potrà essere emesso se non in presenza di una probabile causa, comprovata mediante giuramento o dichiarazione solenne, e con l'indicazione specifica del luogo da perquisire e delle persone o delle cose da sottoporre a sequestro.». The Constitution of the United States, 1787. Traduzione a cura della laureanda.

3.3 Contributo della presente ricerca

Considerata la funzione orientativa e dunque, non conclusiva dell'analisi linguistica, la presente ricerca intende fornire un contributo nella letteratura esistente sul caso Unabomber e sull'analisi del *Manifesto*. Come evidenziato gli studi in materia si focalizzano sul caso giuridico, sull'attualità delle tematiche trattate ma soprattutto sull'attribuzione autoriale del *Manifesto*, ossia sulla comparazione tra l'elaborato e altri testi di Theodore Kaczynski ritrovati nella sua cabina per dimostrare che egli fosse autore del *Manifesto*. In questo contesto il seguente lavoro di ricerca intende fornire una prospettiva diversa: non si ricerca una corrispondenza testuale, bensì si analizza lo scritto come testo anonimo e come unico dato disponibile in grado di suggerire caratteristiche plausibili sull'autore. Il contributo che si intende fornire, dunque, non riguarda rivendicare di aver scoperto e messo per iscritto elementi che nessuno aveva mai notato, ma proporre un'analisi linguistica orientata alla profilazione autoriale finalizzata a offrire un esempio concreto di applicazione degli strumenti teorici della linguistica forense per inferire caratteristiche dell'autore.

Tuttavia, il presente lavoro non intende proporre una ricostruzione integrale del profilo elaborato dall'FBI a seguito della pubblicazione del *Manifesto* e della relativa analisi, in considerazione del fatto che un'analisi accademica si distingue evidentemente dalle pratiche di *profiling* investigativo: quest'ultime, di fatto, si fondano sull'integrazione di informazioni documentali e su conoscenze professionali specifiche e in una ricerca come la seguente risulta complicato avere accesso a documenti ufficiali derivanti dalle indagini svolte o acquisire competenze altamente specialistiche, generalmente acquisite dopo anni di studi. Inoltre, determinate conclusioni sviluppate nel processo investigativo originale derivano dalla conoscenza personale e deduzioni del *profiler* che originariamente ha sviluppato il profilo.

Pertanto, il presente lavoro si limita a riportare e analizzare gli elementi linguistici che suggeriscono determinate caratteristiche dell'autore, relative a genere, età, classe sociale e livello di istruzione, individuabili esclusivamente dal *Manifesto*; a seguito dell'analisi si fornirà una plausibile interpretazione di questi sulla base dei risultati, riportati da Andrea Nini (2018), relativi alla correlazione tra tratto linguistico e dato demografico discussi nel paragrafo 2.4. A ciò si affianca, in un secondo momento, la considerazione delle principali deduzioni attribuite al *profiler* dell'FBI con lo scopo di riportare esempi di inferenze difficilmente derivabili dai soli dati linguistici, ma che certamente, se fondati, arricchiscono in maniera significativa un profilo d'autore.

In questa prospettiva, la tesi si propone di offrire un esempio concreto di costruzione di un profilo autoriale a partire da prove linguistiche individuabili tramite strumenti propri della disciplina della linguistica forense, per mostrare come un tale approccio possa contribuire alla costruzione di

un quadro indiziario coerente, utile a orientare ulteriori attività investigative necessarie affinché si raggiunga lo standard della *probable cause.*

CAPITOLO 4 - L'IMPRONTA LINGUISTICA DI UNABOMBER: UN'ANALISI FORENSE MULTILIVELLO

4.1 Scelte lessicali e campi semantici dominanti

Per delineare un profilo autoriale plausibile del *Manifesto*, considerato come un testo anonimo nella seguente ricerca, è necessario ampliare l'analisi che si effettua sullo scritto, indirizzando l'attenzione verso la sua architettura linguistica; con il termine plausibile non si fa riferimento alla corrispondenza del profilo risultante da questa analisi con quello originale creato dall'FBI, bensì ci si riferisce alla coerenza dei risultati emersi rispetto ai dati osservabili e misurabili esclusivamente dal testo originale. Per l'esame dello stile linguistico che caratterizza lo scritto si seguono i modelli proposti dalla letteratura sulla profilazione autoriale, con lo scopo di far emergere quei parametri che possono essere indici di una correlazione tra variazioni sociolinguistiche e variazioni stilistiche proprie dell'autore. L'analisi parte dunque da una descrizione dello stile linguistico del *Manifesto* effettuata tramite l'osservazione delle varie tipologie di *feature* presenti nello scritto con lo scopo di individuarne il registro; è importante sottolineare come in questa fase iniziale le *feature* non vengano misurate o interpretate, dal momento che la reale variazione e la conseguente interpretazione demografica siano osservabili solo dopo la costruzione di un *corpus* con cui confrontare i parametri emersi.

Per individuare ed esaminare le varie *feature* presenti nel testo è necessario partire dalla classificazione di ogni parola come nome, verbo, aggettivo, avverbio, etc.; per fare ciò risulta particolarmente utile il *POS tagging*. Il *part of speech* (POS) *tagging* consiste nell'assegnazione della corretta categoria grammaticale a ciascuna parola di una frase[34] e rappresenta una delle principali applicazioni del *Natural Language Processing* [elaborazione del linguaggio naturale], una disciplina interdisciplinare che combina linguistica, informatica e intelligenza artificiale finalizzata all'elaborazione e alla comprensione del linguaggio naturale[35]. Durante l'analisi linguistica, dunque, fare affidamento su un approccio computazionale come il *POS tagging* permette non solo di velocizzare il processo di analisi, riuscendo ad esaminare grandi moli di dati testuali, scritti o digitali, ma permette anche di assegnare a ciascuna parola del testo la sua corrispondente funzione grammaticale, fornendo così una rappresentazione della struttura del linguaggio utilizzata nello scritto.

[34] Chiche A., Yitagesu B. (2022). *Part of speech tagging: a systematic review of deep learning and machine learning approaches*. Journal of Big Data 9, 10.
[35] Pio Principe M. (2021). *Elaborazione del linguaggio naturale con metodi probabilistici e reti neurali*. Università di Bologna.

Tuttavia vi è un aspetto delicato da considerare: la medesima parola può appartenere a diverse categorie grammaticali, in base alla sua posizione all'interno della frase o può avere molteplici significati, il che rappresenta un elemento che difficilmente un sistema riesce ad individuare in maniera autonoma. Da qui ne deriva il problema dell'ambiguità contestuale e lessicale, un fenomeno tipico della lingua inglese. Con l'espressione ambiguità contestuale si intende una determinata sequenza di parole che può essere assegnata a più di una struttura sintattica in base alle regole della grammatica (Javed, Jehan, & Ahmad, 2025); un esempio di ciò può essere la frase "*time flies*" che si può interpretare come struttura sostantivo-verbo oppure come verbo-sostantivo. Per quanto riguarda invece l'ambiguità lessicale, questa viene definita come una sequenza di caratteri (nella lingua scritta) che corrisponde a più di una parola e/o significato (Boland, 2023); nella lingua italiana la parola "*bat*" può essere tradotta sia come "pipistrello" sia come "mazza da baseball".

In quest'ottica il *POS tagging* moderno sta venendo sviluppato in un modo tale che il sistema operativo sia in grado di apprendere i vari modelli che gli vengono presentati così che possa attribuire le funzioni grammaticali in modo più accurato, basandosi non solo sulla parola isolata, ma anche sul suo contesto[36]. Inoltre, un'altra possibile via percorribile per ovviare a questa problematica è quella dell'annotazione manuale che vede una persona fisica contrassegnare ogni *token* con la relativa categoria, approccio a cui si è fatto affidamento per l'elaborazione di questa tesi.

Dopo aver associato ciascuna parola del *Manifesto* alla sua relativa categoria grammaticale è possibile proseguire con l'analisi delle *feature*.

Le prime *feature* prese in esame sono le *feature* lessicali. Il lessico infatti è il principale indicatore del registro, poiché le sue variazioni sono strettamente legate al grado di formalità della situazione o dell'argomento trattato; gli elementi su cui ci si focalizza sono in primo luogo la tipologia di lessico scelta e inserita nello scritto, per poi analizzare la frequenza con la quale determinati vocaboli appaiono nei periodi.

Da una prima lettura del *Manifesto* è anzitutto chiara la norma ortografica adottata: nel corso del testo, infatti, l'autore utilizza con coerenza l'ortografia dell'inglese americano, privilegiando forme quali '*behavior*', '*organization*', '*civilization*' e '*labor*' rispetto alle corrispondenti varianti britanniche ('*behaviour*', '*organisation*', '*civilisation*' e '*labour*').

Nell'ambito dell'analisi lessicale, un primo aspetto preso in esame riguarda il grado di formalità del lessico, valutato tramite il calcolo dell'*F-score* di Heylighen & Dewaele. Questo indicatore linguistico è inteso come una misura empirica della formalità lessicale, misurazione basata sulla frequenza di determinate categorie grammaticali in un testo; nel 1999 Heylighen e Dewaele, nel

[36] POS tags. (2018). Sketch Engine.

loro articolo *Formality of Language: definition, measurement and behavioral determinants*, mostrano come determinate categorie linguistiche siano più tipiche in scritti formali rispetto che in quelli informali e individuano l'*F-score* come un'ottima indicatore del grado di formalità. Nello specifico individuano sostantivi, aggettivi, articoli e preposizioni come indici di un testo più formale, ma esclusivamente se in relazione a pronomi, verbi e avverbi sono presenti con una frequenza maggiore. La frequenza con la quale si manifestano tali categorie è misurata tramite una formula, il cui risultato rappresenta l'*F-score*:

F = [(noun freq. + adjective freq. + preposition freq. + article freq.) - (pronoun freq + verb freq. + adverb freq.) + 100] / 2

È opportuno specificare che nella seguente formula le frequenze delle varie parole si esprimono in percentuali e si aggiunge 100 e divide per 2 così da ottenere un risultato compreso da 0 e 100. Da ciò si deduce quando è altro l'*F-score* così da determinare se la formalità lessicale sia tendenzialmente alta (*F-score* alto) o bassa (*F-score* basso).

Per determinare, dunque, se il lessico del *Manifesto* sia maggiormente formale o informale si è applicata la stessa formula; dato che non sono stati usati modelli che permettono di calcolare l'*F-score* automaticamente ma le frequenze delle varie categorie sono state estratta e manualmente, il risultato che ne deriva è da considerarsi come una buona stima dell'*F-score* del *Manifesto*:

F = [(25% + 8% + 14% + 8%) - (8% + 12% + 4%) + 100] / 2

F = 131 / 2

F = 65,5

Il valore approssimativo dell'*F-score* si aggira quindi intorno a 65-66.

Secondo quanto delineato da Heylighen e Dewaele, tale valore è indice di un lessico formale, privo di espressioni colloquiali ma non equiparabile al lessico della prosa accademica e tecnica. Il risultato ottenuto infatti collocherebbe il testo in maniera più adeguata vicino a testi politici e ideologici o saggi argomentativi, testi che presentano un'elevata densità nominale e numerose preposizioni che introducono strutture argomentative.

Un ulteriore elemento da attenzionare riguarda la tipologia di lessico utilizzato, il quale appare prevalentemente tecnico e astratto: l'autore infatti fa uso di una terminologia che non fa riferimento ad oggetti concreti o esperienze sensoriali di cui il lettore può fare esperienza bensì si avvale di parole che richiamano concetti astratti e precise categorie. Tuttavia analizzando in maniera più approfondita la terminologia scelta si possono notare delle variazioni lessicali che dipendono dai differenti contenuti affrontati nelle diverse sezioni; prendendo in considerazione tali variazioni il *Manifesto* si può suddividere in tre parti, ognuna con un lessico specifico e che risulta interessante analizzare per comprendere il perché di tali scelte terminologiche. Per fornire esempi concreti di

quanto si sta affermando si sono individuate le sezioni che possono funzionare come modello lessicale caratteristico in grado di mostrare in modo evidente le variazioni di lessico presenti nelle tre parti del *Manifesto* appena menzionate.

> The Industrial Revolution and its consequences have been a disaster for the human race. They have greatly increased the life-expectancy of those of us who live in "advanced" countries, but they have destabilized society, have made life unfulfilling, have subjected human beings to indignities, have led to widespread psychological suffering (in the Third World to physical suffering as well) and have inflicted severe damage on the natural world. The continued development of technology will worsen the situation. It will certainly subject human being to greater indignities and inflict greater damage on the natural world, it will probably lead to greater social disruption and psychological suffering, and it may lead to increased physical suffering even in "advanced" countries. (*Industrial Society and Its Future*, sezione 1).[37]

La prima sezione presa in esame è stata selezionata dal momento che rappresenta in modo evidente la tipologia di lessico presente nelle prime sezioni dello scritto, che costituiscono la parte maggiormente teorica e ove l'autore fornisce le definizioni di concetti che gettano le basi per la costruzione dell'intera tesi portata avanti (sezioni 1-76). In questa prima parte infatti l'autore dedica ampio spazio nell'illustrare alcuni concetti e darne le relative definizioni, che saranno centrali per la comprensione della critica portata avanti nel corso del testo. Vengono infatti introdotti concetti quali '*feeling of inferiority*', '*human drives*', '*surrogate activities*', '*oversocialization*', tutte nozioni appartenenti all'ambito psicologico. Nell'esporre tali concetti l'autore fa uso di una terminologia propria dell'ambito a cui sta facendo riferimento, di conseguenza il lessico è incentrato sulla definizione rigorosa dei concetti che menziona, eliminando termini gergali o colloquiali. Si può dunque definire che il lessico sia di alto uso[38], specialistico e ricco di termini pieni, intesi come espressioni specifiche che veicolano un significato chiaro e rimandano a concetti precisi[39].

Malgrado l'elevato grado di tecnicalità del lessico, che fornisce alle prime sezioni una sembianza di scritto scientifico, in questa prima parte del *Manifesto* è frequente l'uso di un lessico valutativo, ovvero l'uso di una terminologia volta a esprimere un giudizio o qualificare un dato elemento[40].

---

[37] «La rivoluzione industriale e le sue conseguenze sono state un disastro per la razza umana. Esse hanno incrementato a dismisura l'aspettativa di vita di coloro che vivono in paesi "sviluppati" ma hanno destabilizzato la società, reso la vita insignificante, assoggettato gli esseri umani a trattamenti indegni, diffuso sofferenze psicologiche (nel Terzo mondo anche fisiche), inflitto danni notevoli al mondo naturale. Il continuo sviluppo della tecnologia peggiorerà la situazione. Essa sicuramente sottometterà gli esseri umani a trattamenti sempre più abietti, infliggerà al mondo naturale danni sempre maggiori, porterà probabilmente a una maggiore disgregazione sociale e sofferenza psicologica e a incrementare la sofferenza fisica in paesi "sviluppati"» Traduzione a cura della laureanda.

[38] Istituto della Enciclopedia Italiana. (n.d.). Lessico. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/lessico_(Enciclopedia-dell'Italiano)/.

[39] Istituto della Enciclopedia Italiana. (n.d.). Lessico. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/lessico_(Enciclopedia-dell'Italiano)/.

[40] Cava A., M. (2011). *Lessico valutativo e scrittura accademica: un approccio corpus-based*. Studii și Cercetări Filologice. Seria Limbi Străine Aplicate.

Espressioni come '*disaster*', '*deeply-troubled*', '*craziness*', '*loser*' possiedono un forte contenuto valutativo, generalmente negativo, che, associato a specifici termini non descrivono solamente un determinato elemento ma suggeriscono la posizione che l'autore prende nei confronti di un dato concetto. La coesistenza di terminologia specialistica e lessico valutativo risulta particolarmente significativa poiché suggerisce l'intento dell'autore del *Manifesto*: descrivere scientificamente un fenomeno, suggerendo al tempo stesso una certa interpretazione del fenomeno che si sta discutendo.

Avanzando nella lettura del *Manifesto*, però, si osserva una variazione nel lessico adottato dall'autore; il repertorio lessicale passa infatti dall'essere fortemente valutativo e focalizzato sulla sfera psicologica all'essere notevolmente analitico e teorico, modificando di conseguenza l'ambito d'interesse.

> It is said that we live in a free society because we have a certain number of constitutionally guaranteed rights. But these are not as important as they seem. The degree of personal freedom that exists in a society is determined more by the economic and technological structure of the society than by its laws or its form of government. [16] Most of the Indian nations of New England were monarchies, and many of the cities of the Italian Renaissance were controlled by dictators. But in reading about these societies one gets the impression that they allowed far more personal freedom than our society does. In part this was because they lacked efficient mechanisms for enforcing the ruler's will: There were no modern, well-organized police forces, no rapid long-distance communications, no surveillance cameras, no dossiers of information about the lives of average citizens. Hence it was relatively easy to evade control. (*Industrial Society and Its Future*, sezione 95).[41]

La sezione precedentemente citata mostra chiaramente tale variazione lessicale, che sarà presente in gran parte delle successive sezioni riguardanti la critica verso la tecnologia e la società moderna (sezioni 77-170); in relazione al tema e le argomentazioni proposte, la terminologia adottata si orienta verso i tecnicismi propri dell'innovazione e della società.

Un'altra caratteristica del lessico presente in questa porzione dello scritto è la presenza di un una terminologia concettuale e astratta: termini come '*constitutionally guaranteed rights*', '*technological growth*', '*system*', '*freedom*', '*form of governments*' appartengono a quell'insieme di parole che non si riferiscono a entità fisiche ma rimandano a concetti o idee che non possono essere percepiti attraverso i cinque sensi. L'utilizzo di tale peculiare lessico suggerisce che l'attenzione dell'autore non sia rivolta all'analisi delle singole persone o singoli aspetti della

[41] «Si dice che noi viviamo in una società libera perché abbiamo un certo numero di diritti costituzionalmente garantiti. Ma questi non sono importanti come sembra. Il grado di libertà personale che esiste in una 34 società è determinato più dalla struttura economica e tecnologica che dalle sue leggi o dalla sua forma di governo. La maggior parte delle nazioni indiane del New England erano monarchie, e molte città del Rinascimento italiano erano controllate da dittatori. Ma leggendo di queste società si ha l'impressione che esse permettessero molta più libertà personale della nostra. In parte questo accadeva perché non avevano meccanismi efficienti per rinforzare la volontà del reggente: non vi era una forza di polizia moderna, bene organizzata, non vi erano comunicazioni rapide tra lunghe distanze, non vi erano camere di sorveglianza, né dossier informativi sulla vita dei cittadini comuni. Quindi era relativamente facile evadere il controllo.» Traduzione a cura della laureanda.

tecnologia ma bensì verso la tecnologia intesa come un sistema dal quale sono conseguiti delle strutture e processi sociali ormai estesi all'intera società.

Le sezioni che seguono il titolo *How Some People Adjust* (77–86) costituiscono invece un parziale distacco rispetto a questa terminologia astratta e specialistica.

> Another way in which people satisfy their need for the power process is through surrogate activities. As we explained in paragraphs 38-40, a surrogate activity is an activity that is directed toward an artificial goal that the individual pursues for the sake of the "fulfillment" that he gets from pursuing the goal, not because he needs to attain the goal itself. For instance, there is no practical motive for building enormous muscles, hitting a little ball into a hole or acquiring a complete series of postage stamps. Yet many people in our society devote themselves with passion to bodybuilding, golf or stamp-collecting. (...). (*Industrial Society and Its Future*, sezione 84).[42]

In tali paragrafi l'autore tratta temi reali, evocando situazioni pratiche e quotidiane come '*shiny new toys*', '*building enormous muscles*', '*putting in overtime*', '*taking a second job*'. A differenza dei paragrafi precedenti e quelli che seguiranno, qui la terminologia più astratta e lontana dalla quotidianità è sostituita con termini più vicini alla realtà di tutti e che richiamano immediatamente a immagini ben precise. A ciò si aggiungono i riferimenti a nomi propri di persone e luoghi reali, quali Manuel Noriega e Panama, elemento che conferma la variazione lessicale improntata verso il fornire esempi concreti e puntuali come prova della sua tesi.

L'ultima variazione lessicale si osserva nelle sezioni conclusive del *Manifesto*, ovvero quelle relative ai possibili scenari futuri e alla necessità di una rivoluzione contro il sistema nel momento in cui quest'ultimo risulta opprimente e impossibile da riformare (sezioni 171-232).

> Until the industrial system has been thoroughly wrecked, the destruction of that system must be the revolutionaries' ONLY goal. Other goals would distract attention and energy from the main goal. More importantly, if the revolutionaries permit themselves to have any other goal than the destruction of technology, they will be tempted to use technology as a tool for reaching that other goal. If they give in to that temptation, they will fall right back into the technological trap, because modern technology is a unified, tightly organized system, so that, in order to retain SOME technology, one finds oneself obliged to retain MOST technology, hence one ends up sacrificing only token amounts of technology. (*Industrial Society and Its Future*, sezione 200).[43]

---

[42] «Un altro modo con il quale si può soddisfare il proprio bisogno del processo di potere sono le attività sostitutive. Come spieghiamo nei paragrafi 38-42, un'attività sostitutiva è diretta verso un obiettivo artificiale; l'individuo la pratica per l'interesse verso il processo di "conseguimento" che ne ricava, non perché ha bisogno di soddisfare quello scopo specifico. Per esempio non vi è un motivo pratico per costruirsi enormi muscoli, inviare una piccola palla dentro una buca o acquisire una serie completa di francobolli postali. Tuttavia molte persone nella nostra società si dedicano con passione al body-building, al golf o a collezionare francobolli. (...).» Traduzione a cura della laureanda.

[43] «Fino a che il sistema industriale non sarà distrutto completamente la distruzione di quel sistema deve essere l'unico obiettivo dei rivoluzionari. Altri obiettivi distrarrebbero l'attenzione e l'energia dall'obiettivo principale. Cosa ancora più importante, se i rivoluzionari acconsentono a spostare la loro attenzione verso un obiettivo diverso dalla distruzione della tecnologia, saranno tentati di usare la tecnologia come uno strumento per raggiungere l'altro obiettivo. Se essi accettano questa tentazione ricadono giusto nella trappola tecnologia, perché la tecnologia moderna è un sistema unico,

Ciò che caratterizza questa parte dello scritto è il fatto che non si descriva più un fenomeno cercando di diagnosticarne il problema, ma si passa all'azione fornendo una progettazione su come agire; come conseguenza il lessico risulta astratto ma non teorico dal momento che non vengono presentati concetti generali o teorie ma si fa riferimento a strutture collettive ed espressioni legate a strategie di mobilitazione. Termini quali '*revolutionaries*', '*elite*', '*commitment*', '*dominant force*', '*prepare the way*', '*final push*' costituiscono esempi significativi di questa tendenza. Pur trattandosi di espressioni specifiche, queste fanno riferimento a situazioni e azioni concrete, con il preciso intento di far percepire la rivoluzione al lettore come un'ipotesi vicina e reale e di conseguenza realizzabile, un risultato a cui l'autore aspira.

Dall'analisi delle variazioni lessicali presenti nel *Manifesto* si può sostenere che, in linea generale, il testo ha una tendenza verso una terminologia maggiormente astratta e specialistica di vari settori quali psicologia, tecnologia e mobilitazione; nonostante ciò l'autore intende mantenere un contatto con la realtà inserendo talvolta delle sezioni con un lessico più concreto e che richiama a situazioni di cui il lettore può fare esperienza sensoriale. Ciò suggerisce dunque che l'intenzione dell'autore sia quella di produrre un testo accademico e rigoroso, ma che allo stesso tempo rimanga accessibile anche all'ampio pubblico, inserendo passaggi con un lessico anche più vicino alla quotidianità, come quello delle sezioni 77-86.

Un ulteriore esame che rientra nell'analisi delle *feature* lessicali e che va oltre l'individuazione del tipo di terminologia utilizzata in un testo riguarda la ricchezza lessicale. La ricchezza lessicale è un dato molto interessante da estrapolare durante un'analisi linguistica dal momento che indica l'ampiezza e la varietà lessicale che una persona possiede; evidentemente maggiore e più vario è il patrimonio di termine conosciuti, maggiore è la chiarezza con la quale un individuo è in grado di esprimersi e presentare le sue argomentazioni[44]. Dal punto di vista della profilazione autoriale però può risultare particolarmente significativo, dal momento che una conoscenza più vasta e varia di termini richiede un'esposizione a contesti comunicativi differenti e anche complessi, una formazione scolastica adeguata e possibilmente un particolare interesse alla lettura che permette certamente di ampliare il proprio vocabolario. Per questo motivo durante l'analisi del *Manifesto* si è calcolato l'indice linguistico *Type-Token Ratio* (TTR).

Il *Type-Token Ratio* (TTR) è una misura utilizzata per valutare la ricchezza lessicale di un testo e si basa sulla relazione tra parole uniche e parole totali[45]. Più specificamente, si distingue il numero

ermeticamente organizzato, così che per conservare qualche tecnologia ci si trova obbligati a conservare la maggior parte della tecnologia, quindi si finisce col sacrificare solo quantità simboliche di tecnologia.» Traduzione a cura della laureanda.

[44] Benigno V. (2013). *Ricchezza lessicale e uso delle collocazioni in produzioni scritte di italiano L2 e italiano L1*. Varietà dei contesti di apprendimento linguistico.

[45] Mitchell D. (2014). *Type-token models: a comparative study*. Journal of Quantitative Linguistics.

di *type*, ossia le parole distinte che nel corso del testo appaiono senza ripetizioni, con il numero di *token*, ovvero tutte le parole che compongono lo scritto; a seguito della misurazione delle parole totali e delle parole uniche, si pongono in relazione i *type* rispetto ai *token* con lo scopo di ottenere il TTR. Maggiore sarà il risultato, maggiore sarà la ricchezza lessicale.

Nel caso del *Manifesto* si contano approssimativamente 4.485 parole uniche rispetto alle 34.800 che compongono lo scritto; di conseguenza si ottiene un TTR di 0,128, che reso come percentuale risulta 12,9%.

Tale parametro appare un indicatore di un TTR basso. Tuttavia è necessario specificare che il *Type-Token Ratio* mostra un limite: maggiore è l'estensione del testo, minore sarà il TTR dal momento che più uno scritto si allunga e più vi è la possibilità che le parole inizino a ripetersi. Per tale ragione ad oggi l'indicatore maggiormente utilizzato nel campo della linguistica forense risulta essere il Guiraud *Index*. L'indice di Guiraud appare risolvere la problematica del TTR legata alla lunghezza di un testo**,** poiché al posto di mettere in relazione i *type* rispetto al numero di *token* totali, il numero di *type* viene diviso per la radice quadrata dei *token*. Sulla base di tale misurazione, infatti, l'indice di ricchezza lessicale risulta più elevato, circa 24,05, dato che, come indica David Mitchell (2014), rientra nella fascia identificata come indicatore di un testo con una grande varietà lessicale.

In merito alla ricchezza lessicale vi sono due dati che risultano interessanti ai fini della profilazione autoriale. Il primo riguarda la presenza di forme o parole oggi percepite come desuete o piuttosto ricercate nell'inglese moderno. Gli esempi maggiormente significativi di questa tipologia di lessico sono '*clew*' [indizio] e l'espressione idiomatica "*You can't eat your cake and have it too*" [Non si può avere la botte piena e la moglie ubriaca]. La particolarità di queste forme ortografiche è che sono state sostituite con grafie, ossia '*clue*' e "*You can't have your cake and eat it too*", che oggi risultano essere più comunemente utilizzate. Queste forme sono generalmente riconosciute come varianti ortografiche storiche delle determinate parole[46], che continuano ad esistere pur venendo percepite come forme letterali. La presenza di tali varianti quindi non modifica il significato dell'enunciato ma contribuisce a conferire al testo un tono più letterario.

Un ulteriore dato osservabile tramite l'indicatore della ricchezza lessicale sono le ricorrenze di determinate parole nel corso dello scritto. Analizzando il *Manifesto* i termini che appaiono con più frequenza sono sicuramente '*society*' (ca. 235 volte), '*people*' (ca. 220 volte), '*system*' (ca. 215 volte), '*power*' (ca. 165 volte), '*technology*' (ca. 120 volte), tutte parole che richiamano il contenuto fondante dell'intero *Manifesto*. Dal punto di vista contenutistico la ripetizione di queste precise parole appare prevedibile poiché rappresentano i nuclei centrali dell'opera; tuttavia, da un punto di

[46] Pearson Education. (n.d.). Clue. In Longman Dictionary of Contemporary English.; Zimmer B.(2006). "Forensic linguistics, the Unabomber, and the etymological fallacy". *Language Log*.

vista linguistico, questa ripetizione costante si può leggere come uno strumento linguistico pensato per aumentare la credibilità della tesi e infondere nel lettore l'idea che l'intero discorso converga inevitabilmente verso un'unica e medesima conclusione.

Questo effetto è inoltre amplificato dall'associazione costante di questi termini a concetti specifici che ogni volta che vengono ripresentati si presentano sempre nella stessa combinazione; nel *Manifesto*, ad esempio, '*system*' viene costantemente associato a controllo, alienazione e perdita di libertà, '*technology*' a degrado psicologico e dipendenza e '*modern society*' a sofferenza e disintegrazione umana. Attraverso il reiterarsi di queste connessioni, il testo costruisce una rete semantica che viene interiorizzata dal lettore con l'avanzare della lettura e semplifica in qualche misura la realtà in cui gli individui si muovono: se si presenta in maniera costante l'immagine della tecnologia o dell'intero sistema società come qualcosa che drena la libertà e l'autonomia umana, nella mente del lettore si instaura con più facilità l'idea che ciò che è opposto a questi concetti porti invece libertà e autonomia.[47] Questo tipo di costruzione, quindi, elimina zone intermedie o ambiguità presentando la realtà come divisa in due poli opposti.

Grazie all'analisi lessicale si può dunque concludere che il lessico è chiaramente accademico e tecnico, ma i risultati di questa analisi non si esauriscono a questo poichè un esame più accurato riguardo la terminologia specifica usata e la ripetizioni di questa, potrebbe mostrare il reale intento dell'autore del *Manifesto*.

4.2 Analisi sintattica e organizzazione del discorso nel *Manifesto*

Le considerazioni emerse dall'analisi lessicale evidenziano come le scelte terminologiche dell'autore contribuiscano a conferire al *Manifesto* un carattere teorico e compatto dal punto di vista ideologico. Tuttavia, la forza argomentativa del testo dipende anche dalle modalità con cui i concetti vengono presentati e collegati tra loro; in questa prospettiva, l'analisi delle *feature* sintattiche risulta fortemente significativa per comprendere la struttura che l'autore sfrutta per presentare al lettore la propria visione della realtà come il risultato di un ragionamento logico e coerente. Attraverso tale analisi è così possibile rivelare il meccanismo inconscio che si attua nel momento in cui una persona scrive un testo; tali dinamiche, ormai ampiamente assimilate, portano a effettuare scelte grammaticali in modo largamente inconsapevole e dunque difficilmente camuffabili[48]. Pertanto l'analisi della sintassi rappresenta un elemento di particolare interesse nell'attribuzione autoriale, attraverso la quale è possibile identificare l'idioletto dell'autore del testo anonimo.

---

[47] Hassan A., Barber S.J. (2021). *The effects of repetition frequency on the illusory truth effect.* Cognitive Research: Principles and Implications.

[48] Nini A. (2018). *Developing forensic authorship profiling.* Language and Law / Linguagem e Direito, Vol. 5(2), p. 38-58.

Per svolgere questo tipo di analisi è necessario seguire una precisa *pipeline* articolata in più fasi. La prima fase è quella della tokenizzazione in cui il linguista è chiamato a dividere le varie frasi di un testo in vari *token*, ovvero unità linguistiche che possono essere associate a una specifica funzione grammaticale: la frase "*The industrial society creates artificial needs*" si può dividere in sei *token*, ciascuno dei quali associabile a una categoria grammaticale mediante il processo di *POS tagging* da cui ne risulta la seguente suddivisione:

Token	***POS tagging***
The	articolo determinativo
industrial	aggettivo
society	sostantivo
creates	verbo
artificial	aggettivo
needs	sostantivo

Una volta assegnata la corretta categoria grammaticale a ciascun *token* il testo diviene una sequenza di categorie grammaticali, il che permette di estrarre la frequenza dei vari *tag* che si presentano nel testo; in altre parole, si vanno a contare quanti sostantivi, verbi o aggettivi vi sono e quali di questi si presentano con la maggior frequenza.

Un esame di questo tipo permette di individuare i *POS n-grams* ovvero sequenze contigue di etichette grammaticali utilizzate come caratteristiche stilometriche per rappresentare i modelli sintattici di un testo, riducendo l'influenza del contenuto lessicale (Kawasaki, 2023).

Se si prende come riferimento la medesima frase utilizzata in precedenza "*The industrial society creates artificial needs*" è possibile osservare due possibili *n-grams* ovvero "*industrial society*" e "*artificial needs*"; entrambe le sequenze non vanno lette dal punto di vista lessicale, ma piuttosto dal punto di vista delle funzioni grammaticali che i *token* che li formano possiedono, individuando così una ripetizione della struttura aggettivo-sostantivo. Tali sequenze sono particolarmente dimostrative della presenza di *pattern* sintattici indipendenti dal contenuto del testo, il che risulta particolarmente rilevante nell'individuazione dell'idioletto di un autore. Tuttavia è necessario specificare che i singoli *n-grams* non identificano l'idioletto poichè tali sequenze ottengono significato nel momento in cui la loro analisi viene affiancata da un esame più profondo riguardo la struttura sintattica del testo.

L’analisi sintattica, infatti, si conclude con l’osservazione delle regole grammaticali che governano le frasi e interi periodi. L'attenzione si sposta dunque verso l'individuazione delle strutture subordinate, coordinate, argomentative o l’uso della struttura passiva con l’obiettivo di individuare la logica con la quale l’autore costruisce il suo ragionamento.

Dopo aver chiarito la logica che governa l’analisi sintattica del *Manifesto* si possono esporre i risultati che ne sono derivati.

Un primo elemento che si può riscontrare sulla base dell’estrazione della frequenza dei *POS tag* è un’alta densità nominale.

> It seems that for many people, maybe the majority, these artificial forms of the power process are insufficient. A theme that appears repeatedly in the writings of the social critics of the second half of the 20th century is the sense of purposelessness that afflicts many people in modern society. (This purposelessness is often called by other names such as “anomic” or “middle-class vacuity.”) We suggest that the so-called “identity crisis” is actually a search for a sense of purpose, often for commitment to a suitable surrogate activity. It may be that existentialism is in large part a response to the purposelessness of modern life. [12] Very widespread in modern society is the search for “fulfillment.” But we think that for the majority of people an activity whose main goal is fulfillment (that is, a surrogate activity) does not bring completely satisfactory fulfillment. In other words, it does not fully satisfy the need for the power process. (See paragraph 41.) That need can be fully satisfied only through activities that have some external goal, such as physical necessities, sex, love, status, revenge, etc. (*Industrial Society and Its Future*, sezione 64).[49]

Se si associa ciascun *token*, in cui si scompone l’intera sezione, alla relativa funzione grammaticale e ci si focalizza sulla frequenza con la quale sostantivi, verbi e aggettivi compaiono ne risultano: 44 sostantivi, 16 verbi e 15 aggettivi. Se si pongono poi in relazione tra loro i dati appena menzionati si osserva che i sostantivi appaiono con una frequenza circa tre volte maggiore rispetto sia ai verbi che agli aggettivi (rispettivamente con un rapporto di 2,75: 1 e 2,93: 1).

Inoltre vi sono determinate sezioni che contengono strategie linguistiche che aumentano il numero di sostantivi presenti al loro interno; tra queste vi sono le enumerazioni nominali, ovvero un elenco di elementi costituito prevalentemente da sostantivi[50] e la nominalizzazione, ossia un

[49] «Sembra che per molte persone, forse la maggioranza, queste forme artificiali del processo di potere siano insufficienti. Un tema ricorrente negli scritti dei critici sociali della seconda meta del XX secolo è il senso di inutilità che affligge molte persone nella società moderna. (Questa mancanza di scopi è spesso chiamata con altri nomi, come "anomia" e "vacuità della classe media".) Noi sosteniamo che la cosiddetta "crisi di identità" è di fatto una ricerca di intento, spesso di impegno verso una conveniente attività sostitutiva. Può essere che 1'esistenzialismo sia in larga parte una risposta alla mancanza di scopi della vita moderna". Nella società moderna è molto diffusa la ricerca del "raggiungimento". Ma noi pensiamo che per la maggioranza delle persone un'attività il cui principale obiettivo è la realizzazione (cioè una attività sostitutiva) non sia una soddisfazione completamente soddisfacente. In altre parole, non soddisfa pienamente il bisogno per il processo del potere (vedi paragrafo 41). Questo bisogno può essere pienamente soddisfatto solo attraverso attività che hanno qualche traguardo esterno, come le necessita fisiche, il sesso, l'amore, lo status, la vendetta, ecc.» Traduzione a cura della laureanda.

[50] Istituto della Enciclopedia Italiana. (n.d.). Enumerazione. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/enumerazione_(Enciclopedia-dell'Italiano)/.

processo linguistico che consiste nel trasformare un verbo, un aggettivo o un'intera frase in un nome[51]. Un chiaro esempio di nominalizzazione vi è nella frase: "*The trouble with such theorists is that they have made the development and application of social theories their surrogate activity.*" (*Industrial Society and Its Future*, sezione 97). In questo estratto appare chiaro come nel passaggio "*they have made the development and application of social theories*" l'autore potesse semplificare la sintassi ed esprimere lo stesso concetto utilizzando la struttura soggetto-verbo-complemento "*they developed and applied the social theories*". Nonostante ciò, l'autore opta per utilizzare la strategia della nominalizzazione, strategia che contribuisce a conferire alla prosa una maggiore densità nominale.

Spostando l'attenzione sui *n-grams* che compaiono sistematicamente nel corso del *Manifesto* è interessante menzionare due sequenze di categorie grammaticali e che già da una prima lettura appaiono evidenti. La prima struttura è aggettivo-sostantivo; nel testo si incontrano, con una frequenza notevole, dei sostantivi associati a uno o più aggettivi e ciò accade specialmente nel caso dei termini che sono centrali per la tesi su cui si basa il *Manifesto*. Questo risultato, inoltre, è in linea con quanto riscontrato con l'analisi lessicale presentata nel capitolo precedente che ha messo in luce la presenza di associazioni fisse tra determinati termini e concetti specifici come accade con il termine '*technology*' costantemente associata ad aggettivi che riflettono un certo degrado psicologico quali '*oppressive*', '*destructive*'.

Un altro *n-gram* frequentemente utilizzato nel *Manifesto* è la struttura pronome-verbo-*that*; tale schema compare nei passaggi in cui l'autore intende inserire un'opinione personale. La peculiarità di questa struttura è che il pronome utilizzato non è *I* (prima persona singolare) bensì *we* (prima persona plurale), e.g. "*we argue that* (...)" (*Industrial Society and Its Future*, sezione 57); "*we contend that* (...)" (*Industrial Society and Its Future*, sezione 58); "*We don't mean that* (...)" (*Industrial Society and Its Future*, sezione 58).

L'ultimo esame che rientra nell'analisi sintattica riguarda le regole grammaticali utilizzate per legare tra loro le varie frasi, esaminata attraverso il calcolo di subordinate per *T-unit*.

Questo tipo di analisi si basa sul calcolo delle *T-unit*, definite come proposizioni principali seguite da tutte le subordinate che dipendono da esse; di conseguenza se una frase contiene più proposizioni coordinate ciascuna di esse rappresenta una *T-unit* distinta, mentre se una frase è composta da una principale e la sua relativa subordinata, questa viene considerata una *T-unit* unica. Tramite l'individuazione delle *T-unit* presenti nel *Manifesto* si è quindi in grado di esaminare il tipo di legame che unisce due o più proposizioni (subordinazione o coordinazione) e la quantità di informazioni inserite in una singola unità. Questo approccio permette di esaminare il testo in un

[51] Istituto della Enciclopedia Italiana. (n.d.). Enumerazione. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/enumerazione_(Enciclopedia-dell'Italiano)/.

modo più approfondito, andando oltre la misurazione della lunghezze delle frasi poiché questa presa come singolo dato non fornisce elementi a sufficienza per determinare se un testo presenta una sintassi complessa o no; di fatto un autore può costruire frasi abbastanza lunghe ma costruirle tramite dei legami di coordinazione, elemento che la letteratura nell'ambito della stilometria identifica come tratto di una sintassi più semplice rispetto a un elaborato le cui frasi possiedono un alto grado di subordinazione.

Pertanto attraverso questa analisi, più profonda rispetto alle precedenti, si può concludere che la sintassi del *Manifesto* presenta una tendenza verso l'ipotassi, ossia le proposizioni che si susseguono in un periodo sono legate tra loro da un rapporto di subordinazione dal quale ne deriva una struttura gerarchica in cui vi sono una o più proposizioni che dipendono gerarchicamente dalla principale. Un esempio rappresentativo di tale tendenza è il seguente:

> *We contend that the most important cause of social and psychological problems in modern society is the fact that people have insufficient opportunity to go through the power process in a normal way.* (*Industrial Society and Its Future*, sezione 58).[52]

In questo periodo si osserva una certa complessità sintattica: alla proposizione principale "*We contend*" segue una subordinata introdotta da "*that*", al cui interno compare una seconda subordinata dipendente dall'espressione "*is the fact that*" e un ulteriore subordinata che conclude la frase introdotta da "*to*". Si conta dunque una sola *T-unit* costituita da tre subordinate che dipendono gerarchicamente dalla proposizione principale "*We contend*".

La complessità del periodo è ulteriore se si considera il soggetto della prima subordinata formato dal sintagma nominale: "*the most important cause of social and psychological problems in modern society*". Una tale organizzazione gerarchica dei periodi conferisce al *Manifesto* un'apparenza di testo che segue un certo rigore logico nel quale le varie argomentazioni vengono esposte con una progressiva concatenazione causale.

A seguito dell'analisi svolta sulla sintassi del *Manifesto* si può sostenere che il testo adotti diverse strategie linguistiche tipiche degli scritti accademici. Secondo quanto afferma Balboni nel suo saggio *Il monologo e lo scritto accademici lo stile accademico* (2011), per essere considerato tale, dovrebbe presentare determinate caratteristiche linguistiche che egli articola su tre livelli: pragmatico, lessicale e sintattico. Sul piano pragmatico uno scritto accademico tende ad essere referenziale e metalinguistico; si tratta dunque di un testo incentrato sul descrivere e argomentare in modo oggettivo la realtà e che nel momento in cui compare un termine o un concetto, riserva spazio

[52] «Sosteniamo che la causa più importante dei problemi sociali e psicologici della società moderna sia il fatto che le persone dispongono di opportunità insufficienti per attraversare il processo di acquisizione del potere in modo normale.» Traduzione a cura della laureanda.

per definirlo e spiegarne il significato. Sul piano lessicale, Balboni afferma che gli scritti accademici dovrebbero fare uso di un lessico specialistico, con termini propri delle varie discipline a cui fanno riferimento. Infine, dal punto di vista della sintassi essa presenta generalmente numerose nominalizzazioni, enumerazioni, un frequente uso della forma passiva e di sintagmi nominali densi e spesso complessi.

Sulla base dei criteri individuati da Balboni si può concludere che il *Manifesto* presenti numerosi aspetti riconducibili allo stile accademico dal momento che possiede gran parte delle caratteristiche linguistiche ritenute necessarie affinché uno scritto accademico si possa definire tale. L'elemento che maggiormente lo distingue da un testo accademico in senso stretto è il ricorso frequente a un lessico valutativo. Tale scelta linguistica fa sì che l'autore non si limiti a descrivere con oggettività ma tende anche a formulare giudizi e frequenti polemiche che alimentano l'efficacia persuasiva del testo.

4.3 Le caratteristiche macro testuali e semantiche del *Manifesto*

La presenza del lessico valutativo all'interno del *Manifesto* non è la sola caratteristica che discosta il documento dai criteri delineati dal professor Balboni; Balboni infatti identifica anche precise caratteristiche macrotestuali come componenti necessarie affinché un testo si possa definire con uno stile accademico. Tra quelle da lui citate vi sono: la presenza di un indice che elenchi in maniera chiara e ordinata i capitoli che compaiono nel documento, ciascuno associato al numero di pagina corrispondente, la suddivisione in capitoli e paragrafi, la formulazione di un *abstract* introduttivo, la presenza di note e appendici e la redazione di una bibliografia. Tali elementi risultano indispensabili negli scritti accademici poiché permettono non solo di esporre il ragionamento scientifico, ma di farlo in maniera logica e fornire le fonti con le quali il lettore può verificare l'affidabilità di quanto si sta affermando; nel momento in cui uno di questi componenti non è presente, in particolar modo la bibliografia, l'elaborato diviene un testo nel quale vengono esposte delle idee di cui però non si conosce la provenienza e che non possono essere verificate. Di conseguenza, non si soddisfa l'obiettivo principale di un testo accademico che è quello di esporre concetti in maniera chiara e verificabile, così da ridurre al minimo l'ambiguità.

La presenza, dunque, di una struttura formata da specifici apparati è ciò che contraddistingue maggiormente un testo accademico da altre forme di scrittura che trattano idee di una certa rilevanza, quali saggi scientifici o manifesti politici; difatti Balboni nel suo saggio, mette in comparazione un saggio scientifico e una tesi, osservando che in un saggio spesso l'indice non è esplicitato e la struttura dell'argomentazione viene dedotta attraverso la progressione dei titoli e

sottotitoli (Balboni, 2011); al contrario, una tesi possiede generalmente le caratteristiche indicate da lui stesso, rientrando così nel genere testuale dello scritto accademico.

Se si applica tale descrizione al *Manifesto* e si analizzano le sue *feature* strutturali e dunque la costruzione e la forma complessiva del testo, si osserva la mancanza di tre componenti: l'indice, *l'abstract* e la bibliografia. Il *Manifesto*, di fatto, possiede una struttura molto rigorosa poiché diviso in 232 sezioni, ciascuna delle quali numerata in modo progressivo; inoltre include delle note, anch'esse numerate, che forniscono informazioni aggiuntive riguardo un concetto o un termine espresso in un paragrafo a cui si fa riferimento. Tuttavia, non vi è presente un indice o un abstract introduttivo e le note che l'autore inserisce non indicano le fonti a cui si sta facendo riferimento. Da queste osservazioni si può inferire che l'autore non mira a esporre al lettore le argomentazioni che ha dedotto attraverso la consultazioni di determinate fonti o gli argomenti che intende approfondire, sulla base di quanto già esistente nella letteratura a riguardo, bensì si concentra sul mostrare la forza della propria tesi che porta avanti nel *Manifesto*. Queste conclusioni non rendono il *Manifesto* uno scritto meno serio o non valido, ma lo discostano certamente da quella che è la definizione di testo accademico.

Tale affermazione trova un'ulteriore riscontro dall'analisi delle *feature* semantiche. Nonostante non siano ampiamente utilizzate per la profilazione autoriale, dal momento che richiedono un esame più complesso incentrato sull'interpretazione delle parole e costruzioni utilizzate, nel caso in esame delineano ulteriormente il disallineamento del *Manifesto* rispetto a un testo accademico.

Un primo elemento rilevante è la presenza di aggettivi qualitativi, ossia aggettivi che modificano il significato del nome, del quale specificano delle qualità[53]. Come già osservato durante l'analisi delle *feature* lessicali, l'autore non si limita a descrivere un fenomeno e presentarne le criticità al lettore in maniera neutrale, ma fornisce delle reali valutazioni; facendo riferimento a passaggi presi direttamente dal *Manifesto*, la società industriale non è presentata in quanto tale, ma è costantemente associata ad aggettivi con connotazione negativa quali *dangerous, frustrated, powerless*. Tramite l'utilizzo di aggettivi qualitativi opportunamente scelti e associati a specifici concetti, l'autore orienta il lettore verso una valutazione, positiva o negativa, nei confronti di un determinato concetto, valutazione che coincide con l'interpretazione stessa dell'autore.

A tal proposito Amelia Maria Cava, nella sua riflessione riguardo il lessico valutativo[54], riprende la definizione di Hunston (1993) che definisce il lessico valutativo come qualsiasi elemento linguistico che indichi l'atteggiamento dello scrivente nei confronti del valore di un'entità nel testo; in questa prospettiva, la presenza di numerosi aggettivi qualificativi all'interno del *Manifesto* si può

[53] Istituto della Enciclopedia Italiana. (n.d.). Aggettivi qualificativi. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/aggettivi-qualificativi_(Enciclopedia-dell'Italiano)/.

[54] Cava A., M. (2011). *Lessico valutativo e scrittura accademica: un approccio corpus-based*. Studii și Cercetări Filologice. Seria Limbi Străine Aplicate.

leggere come un indicatore della tendenza ad associare precisi concetti a specifici aggettivi, così da suggerirne l'interpretazione che l'autore intende trasmettere.

La tendenza a definire e valutare la realtà piuttosto che descriverla in maniera neutrale appare chiara anche dall'uso assiduo di verbi copulativi. All'interno del *Manifesto* infatti si incontrano raramente verbi dinamici che descrivono azioni concrete, poiché la tendenza del testo è quella di descrivere concetti astratti. Per questo motivo nel testo si riscontrano con maggiore frequenza verbi che hanno la funzione copulativa, ossia verbi la cui funzione di predicare è svolta da un aggettivo, un nome o un avverbio[55]; esempi di verbi copulativi sono diventare, sembrare, rappresentare, etc. Anche il verbo essere è considerato un verbo con una funzione copulativa, sebbene venga meglio definito come copula[56] poiché, a differenza dei verbi copulativi che contribuiscono al contenuto semantico, reso dal aggettivo, nome o verbo associato a un determinato soggetto, il verbo essere rappresenta una mera congiunzione tra le due parti[57].

Data la definizione di verbi copulativi, tra i quali rientra il verbo essere, si può sostenere che all'interno del *Manifesto* espressioni come "*Modern society is* (…)", "*Freedom means* (…)", "*The system is not (…)*" appaiono frequentemente e rappresentano effettivamente delle costruzioni con dei verbi copulativi. Sulla base della presenza di questi verbi, ciò che risulta interessante dal punto di vista semantico è che anch'essi formulano giudizi e opinioni soggettive, con la differenza però che i giudizi che vengono espressi attraverso queste espressioni tendono ad essere scambiati come verità generali, specialmente se il verbo copulativo che si sta utilizzando è il verbo essere[58]. Pertanto la tendenza a presentare concetti tramite costruzioni copulative formate dal verbo essere può essere letto come un tentativo di costruire definizioni generali in modo tale che il lettore può essere indotto a interpretarle come definizioni generali e autorevoli, nonostante non vi sia una bibliografia con le fonti di riferimento.

Un ultimo elemento da prendere in considerazione nell'analisi delle *feature* semantiche è l'uso saltuario di pronomi personali individuali. Come già indicato in precedenza, l'autore utilizza con grande frequenza il pronome '*we*', specialmente nei passaggi in cui viene esposta la posizione che assume rispetto a un dato o un fenomeno. L'utilizzo assiduo del pronome '*we*' dal punto di vista semantico si può interpretare in maniera più approfondita rispetto a limitarsi a inferire che l'autore sia più di uno o sia formato da un gruppo; nel momento in cui si svolge un'analisi linguistica per la profilazione autoriale è necessario prendere in considerazione tutte le possibili variabili. In questo

---

[55] Istituto della Enciclopedia Italiana. (n.d.). Verbi. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/verbi_(Enciclopedia-dell'Italiano)/.
[56] Istituto della Enciclopedia Italiana. (n.d.). Verbi. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/verbi_(Enciclopedia-dell'Italiano)/.
[57] Istituto della Enciclopedia Italiana. (n.d.). Verbi. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/verbi_(Enciclopedia-dell'Italiano)/.
[58] Serianni L. (1989), *Grammatica italiana*. UTET.

caso dunque, è bene specificare che l'uso del pronome '*we*' nei testi risulta più frequentemente utilizzato per la costruzione di un'identità collettiva, che raramente fa riferimento alla collettività degli autori che stanno elaborando il testo, specialmente nella scrittura accademica; al contrario, la prima persona plurale viene più largamente usata per presentare il parlante come membro di una collettività. Nell'articolo *When "We Believe that ...": The Role of Collective "we" in the Dialogic Negotiation in Hard News Discourse* (2020) redatto da Nádraská, l'autore sostiene che "nel passare da '*I*' a '*we*', la voce citata ancora esplicitamente il proprio sé al gruppo e si colloca al suo interno". In questo passaggio Nádraská non si limita a definire il pronome *we* come un pronome collettivo, ma piuttosto come uno strumento linguistico che l'autore di un testo adotta per presentarsi come membro di un gruppo, inserendovi di conseguenza le sue argomentazioni. Questo non implica che tramite l'uso della prima persona plurale il testo acquisisca maggiore credibilità o autorità, bensì fa si che le argomentazioni presentate appaiano rilevanti per un gruppo e non solo per un individuo; grazie a questa scelta linguistica infatti l'autore è in grado di favorire il coinvolgimento del lettore, facendolo sentire anch'egli parte del gruppo a cui si sta facendo riferimento.

In questa prospettiva, però, è interessante citare anche l'articolo di Ruey-Ying Liu *Constituting institutional identity in political discourse: The use of the first-person plural pronoun in China's press conferences* (2023). Sebbene l'articolo si focalizzi sull'uso del pronome '*we*' rispetto al pronome '*I*' nei contesti comunicativi, le stesse riflessioni possono essere applicate nel contesto scritto. Liu riporta infatti che la prima persona plurale è comunemente interpretata come un riferimento all'istituzione a cui chi parla è associato e, pertanto, può funzionare come indicatore dell'identità istituzionale; di conseguenza se un autore di un testo scritto fa uso del pronome '*we*' il lettore può interpretare la sua opinione come quella di un'entità più ampia e non soltanto come quella di un singolo individuo.

Sulla base delle affermazioni di Nádraská e Ruey-Ying Liu la presenza ripetuta del pronome *we* all'interno del *Manifesto* può essere quindi letta in quattro modi: l'uso della prima persona plurale può indicare l'esistenza di più autori (sebbene sia l'interpretazione meno plausibile), collocare l'autore all'interno di un gruppo con il quale condivide determinate riflessioni, rappresentare l'autore come parte di un'istituzione oppure vi è la possibilità che l'autore utilizzi consapevolmente il pronome *we* per conferire al testo un senso di maggiore credibilità e attendibilità dal momento che le argomentazioni che vengono presentate non vengono lette come la posizione di un singolo individuo, ma di un gruppo più ampio o addirittura istituzionale.

È opportuno ricordare che le interpretazioni riportate rappresentano tutte le possibili conclusioni a cui si può giungere sulla base di un elemento specifico, che in questo caso è l'uso della prima persona plurale rispetto alla prima persona singolare; tuttavia, le ipotesi che vengono formulate

vanno sempre confrontate con l'insieme di dati di cui si è in possesso per arrivare a deduzioni il più corrette e ragionevoli possibili. Per questo motivo un ulteriore analisi che viene svolta nel momento in cui si tenta di dedurre il profilo dell'autore di un testo anonimo è quella che si focalizza sul contenuto, facendo affidamento sulle *feature* basate sul contenuto. Quest'ultime risultano particolarmente utili per individuare le tematiche ricorrenti e inferire così l'ideologia dell'autore, ragion per cui all'interno del seguente lavoro di ricerca si è dedicato un paragrafo (1.2) per esporre, in maniera sintetica focalizzandosi sui passaggi più rilevanti ai fini di questa tesi, i principali nuclei tematici affrontati nel *Manifesto*.

CAPITOLO 5 - I RISULTATI DELL'ANALISI LINGUISTICA: PROFILAZIONE, INTERPRETAZIONE E VERIFICA

5.1 Dall'analisi linguistica alla profilazione dell'autore

Considerando congiuntamente i risultati emersi dall'analisi delle *feature* lessicali, sintattiche, strutturali, semantiche e basate sul contenuto è possibile misurare le variazioni delle *feature* presenti nel testo con lo scopo di inferirne una correlazione con un dato demografico e delineare così un profilo plausibile del possibile autore del *Manifesto*; questo processo rappresenta la fase conclusiva della profilazione autoriale delineata da Andrea Nini nella sua tesi *Developing forensic authorship profiling* (2018), documento già presentato all'interno del seguente lavoro di ricerca (capitolo 2.4).

Nello specifico, in merito alla deduzione del genere dell'autore si può inferire che la persona che ha redatto il *Manifesto* sia di sesso maschile. Gli elementi individuati da Nini (2018) che possono orientare verso tale conclusione sono l'elevata densità nominale e una sintassi paratattica; a seguito dell'analisi linguistica del *Manifesto*, dunque, ci si focalizza sui dati riguardanti le *feature* sintattiche e lessicali, attenzionando maggiormente i risultati emersi dal calcolo delle *T-unit* e dall'indice linguistico del'*F-score* di Heylighen & Dewaele. In particolare, il parametro relativo al numero di subordinate per *T-unit* conferma che la sintassi del *Manifesto* sia tendenzialmente ipotattica poiché costruita su un rapporto di subordinazione dalla quale ne deriva una struttura gerarchica in cui le varie proposizioni dipendono gerarchicamente dalla principale e si susseguono con una concatenazione causale. Nonostante questa conclusione non appaia in linea con quanto delineato da Nini come indicatore di un autore di sesso maschile, ma sembri piuttosto orientare verso un'autrice come redattrice del *Manifesto*, il dato relativo alla sintassi non può essere interpretato isolatamente. La profilazione autoriale, infatti, non si basa sulla valutazione indipendente dei singoli indici linguistici, bensì sulla loro interpretazione congiunta che tiene conto del quadro complessivo che si ha davanti. In questo caso la tendenza a una sintassi ipotattica deve essere letta in concomitanza con la risultanza dell'*F-score*. Il valore approssimativo calcolato relativo a tale indice linguistico si aggira intorno a 65-67, valore che secondo quanto delineato da Heylighen e Dewaele è indice di un lessico formale. Tuttavia, l'elemento che risulta maggiormente rilevante non è l'esito ottenuto, bensì gli elementi che influenzano l'*F-score* e che dunque abbassano o alzano l'indice linguistico: Heylighen e Dewaele individuano come indicatori di un *F-score* elevato e quindi di un lessico formale, i sostantivi, gli aggettivi, le preposizioni e gli articoli. Nello specifico caso del *Manifesto* l'*F-score* elevato trova riscontro anche dall'elevata densità nominale del testo derivante dalla frequenza circa tre volte maggiore con la quale i sostantivi compaiono nel testo rispetto ai verbi, dato emerso durante l'analisi lessicale. Ciò

suggerisce che nonostante la presenza di una sintassi prevalentemente ipotattica, il che orienta verso un possible autore di sesso femminile, il rapporto di subordinazione che lega le proposizioni all'interno dei periodi è spesso costruito sulla nominalizzazione, contribuendo ad aumentare la densità nominale del testo e il grado di formalità del lessico. Per le ragioni appena illustrate la deduzione relativa al genere della persona che ha redatto il *Manifesto* risulta maggiormente compatibile con quello che Nini associa a un autore di sesso maschile.

La sintassi ipotattica fornisce altresì indicazioni sull'età dell'autore. È opportuno ribadire che quando si parla di età non si fa riferimento all'età biologica, poiché non vi è una relazione diretta e necessaria tra un dato anagrafico, quale l'età biologica e l'uso della lingua; esiste piuttosto una teoria linguistica nella linguistica forense orientata alla profilazione autoriale secondo la quale la lingua riflette il comportamento linguistico di una persona (Misini *et al*., 2022). Ciò implica che è possibile inferire l'età sociale di una persona poiché questa è modellata sulla base delle esperienze vissute e dai contesti esplorati che hanno permesso di acquisire nuovi registri linguistici e farli propri, a tal punto da farli diventare scelte linguistiche automatiche nel momento in cui si scrive. In questa prospettiva Nini (2018) individua la sintassi come ottimo indicatore dell'età plausibile dell'autore dal momento che secondo la teoria di Kemper, già osservata nel 1987, esiste una correlazione negativa tra complessità sintattica e invecchiamento legata principalmente allo sviluppo, negli anni, di strategie comunicative che rendono possibile alleggerire la sintassi pur rimanendo in grado di far passare lo stesso messaggio mediante un lessico più preciso. La sintassi ipotattica e complessa identificata nel *Manifesto* non risulta dunque coerente con la tendenza alla semplificazione sintattica descritta da Kemper nei parlanti di età avanzata. Ciononostante non è possibile sostenere che l'autore rientri in una fascia di età più giovanile dal momento che vi è un indicatore che non permette di confermare ciò ovvero il *Type-Token Ratio* (TTR). Tale parametro valuta la ricchezza lessicale di un testo, che nel caso del *Manifesto* risulta essere di circa 12,9%. Nonostante il TTR possieda un limite legato all'estensione di un testo (Heylighen, Dewaele. 1999), esso fornisce comunque un'indicazione riguardo la vastità del vocabolario conosciuto dall'autore del *Manifesto*; Il valore ottenuto indica infatti che, su 34.800 parole totali, ogni 100 parole sono circa 13 le forme lessicali diverse. Questo dato permette di escludere una piena compatibilità con i profili linguistici associati ad autori più giovani; di conseguenza risulta ragionevole identificare l'autore del *Manifesto* come un individuo che abbia maturato una significativa esperienza linguistica, un ampio vocabolario e un'ottima proprietà di linguaggio piuttosto che un autore molto giovane e che non abbia ancora sviluppato un repertorio lessicale altrettanto ricco e diversificato. Tale inferenza trova ulteriore riscontro dalla presenza di un lessico specialistico, definizioni

rigorose e forme o espressioni considerate oggi letterarie quali ‘*clew*’ e "*You can't have your cake and eat it too*"[59].

I dati appena menzionati inerenti all’età appaiono in linea con la classe sociale e il livello di educazione poiché delineano un individuo che ha ricevuto un’educazione di un livello piuttosto elevato che gli ha consentito di entrare in contatto con registri e contesti linguistici più formali. Queste conclusioni possono essere avvalorate prendendo in considerazione l’analisi degli n-grams e delle *feature* strutturali. Nello specifico, tramite la tecnica del *POS tagging* è stato possibile associare a ogni parola che compone il *Manifesto* la relativa categoria grammaticale; attraverso tale processo si è potuto individuare specifici *n-grams* che si ripetevano in maniera sistematica durante il testo. La sequenza di etichette grammaticali maggiormente significativa risulta essere la struttura pronome-verbo-that nella quale il pronome che viene utilizzato è sempre quello che indica la prima persona plurale, '*we*'. Come già delineato precedentemente (capitolo 4.3) tale struttura può essere interpretata come indicatore plausibile dell’identità istituzionale e pertanto è spesso utilizzata in contesti e scritti più formali e prestigiosi[60]. Inoltre se si sposta l’attenzione sulle *feature* strutturali del *Manifesto* si osserva che l’autore ha scelto di suddividere il testo in varie sezioni numerate, inserendo talvolta titoli e sottotitoli; queste caratteristiche permettono di sviluppare il ragionamento secondo una progressione logica e argomentativa tipica di scritti formali. Pertanto, sulla base degli elementi appena menzionati, la propensione verso l’uso di un lessico formale (comprovato tramite l’*F-score*), la presenza di una sintassi ipotattica costruita principalmente sulla nominalizzazione e un’elevata ricchezza lessicale (confermata dal calcolo del TTR) è possibile inferire che l'autore possieda una familiarità e una buona competenza nella produzione di testi formali, abilità generalmente associata a un elevato livello di istruzione e a una prolungata esposizione ai registri della scrittura formale, come quelli tipici dell'ambito universitario. La convergenza di tali dati linguistici rende dunque il profilo linguistico delineato compatibile con quello che Nini (2018) osserva nei soggetti appartenenti alle classi sociali più elevate, ossia ai gruppi che hanno un accesso maggiore e più facilitato a contesti comunicativi formali.

In definitiva, gli indicatori analizzati delineano un profilo linguistico in liena con quello di un autore di sesso maschile, la cui età sociale sembra aver favorito il consolidamento di una certa familiarità con i registri formali della lingua e lo sviluppo di un ampio repertorio lessicale, conoscenze presumibilmente riconducibili a un elevato livello di istruzione.

---

[59] Pearson Education. (n.d.). Clue. In Longman Dictionary of Contemporary English.; Zimmer B.(2006). *Forensic linguistics, the Unabomber, and the etymological fallacy*. *Language Log*.

[60] Liu R. Y. (2023). *Constituting institutional identity in political discourse: The use of the first-person plural pronoun in China's press conferences.*

5.2 Dalla profilazione all'interpretazione del *Manifesto*

Il profilo del possibile autore del *Manifesto* non è l'unico elemento che si può dedurre tramite l'osservazione dei dati linguistici; è possibile, infatti, interpretare determinate *feature*, in particolar modo quelle semantiche, con l'obiettivo di inferire la tipologia di testo prodotto e le motivazioni che hanno portato all'utilizzo di determinate strategie linguistiche. Le interpretazioni che ne derivano possono costruire elementi cardine in una profilazione autoriale e soprattutto in un caso come quello di Unabomber, poiché potrebbero delineare le ragioni che muovono le azioni di colui che ha redatto il testo. Nel caso studio in esame conoscere ipotizzare le motivazioni che hanno portato Unabomber a compiere gli attacchi bomba e a richiedere la pubblicazione del *Manifesto* sulle testate giornalistiche è cruciale per le forze dell'ordine poiché permette di orientare le proprie azioni future e possibilmente restringere ulteriormente il campo dei possibili autori di reato.

Un primo elemento che è possibile estrapolare dai dati linguistici è la tipologia di testo che è stato prodotto e il registro utilizzato nel *Manifesto*.

Sebbene a primo impatto il documento possa apparire come un testo accademico dal registro formale-alto, l'analisi linguistica mostra che una conclusione tale non sia pienamente corretta, dal momento che i risultati che ne emergono delineano delle peculiarità che fanno si che il *Manifesto* non soddisfi pienamente i requisiti necessari affinché un testo sia definibile vicino al discorso accademico.

Di fatto, dal punto di vista sintattico si delineano degli elementi compatibili con un registro formale associabile a un testo di natura accademica[61]; tra questi i tratti maggiormente esemplificativi sono evidentemente l'elevata densità nominale, derivante in particolar modo dalla presenza di numerosi nominalizzazioni ed enumerazioni nominali, il forte ricorso all'ipotassi e la presenza di sintagmi nominali anche complessi. A ciò si aggiunge quanto emerge dal punto di vista lessicale, vale a dire l'assenza di colloquialismi in favore dell'uso di un lessico specialistico maggiormente orientato verso una terminologia astratta che permette all'autore di definire in maniera rigorosa determinati concetti che egli tratta nel suo elaborato.

Sulla base della classificazione di Rovere in merito alle varie tipologie di registri[62], tali elementi sono tipici di un registro alto-formale, soprattutto se si prende in considerazione la presenza di un lessico specialistico appartenente a diverse discipline (psicologia, tecnologia e sociologia); questa componente linguistica è infatti uno degli indicatori più forti di un registro alto dal momento che presuppone una certa conoscenza, da parte dell'autore, riguardo l'argomento trattato e una certa

[61] Balboni P. (2011). *Il monologo e lo scritto accademici*. Ca' Foscari.
[62] Istituto della Enciclopedia Italiana. (n.d.). Registro. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/registro_(Enciclopedia-dell'Italiano)/.

familiarità con contesti comunicativi che richiedono specifiche conoscenze disciplinari e linguistiche[63].

Tuttavia, sebbene si possa affermare che il *Manifesto* possieda un registro alto, correlabile a quello di un testo accademico, si osserva che l'uso del lessico specialistico non avvenga nella stessa maniera prevista per un elaborato in stile accademico: l'elemento che differisce in questo caso è che la terminologia inserita o i concetti di cui l'autore da le definizione non vengono collegati a una bibliografia e di conseguenza il lettore non conosce le fonti di tali argomentazioni che gli vengono presentate. Se si prende come riferimento la definizione del termine '*socialization*' che appare nel *Manifesto*, ovvero "*Psychologists use the term "socialization" to designate the process by which children are trained to think and act as society demands*" (*Industrial Society and Its Future*, sezione 24) si osservano delle strategie linguistiche precise. La struttura di tale definizione ricorda la scrittura accademica, poiché introduce un concetto di cui ne definisce il significato[64], indicando che tale descrizione non deriva dalla sua visione della realtà ma da affermazioni fatte da psicologi; eppure tale definizione non viene inserita nella letteratura esistente in merito, né è supportata da riferimenti bibliografici. Il fatto che il lettore debba accettare le definizioni presentate dall'autore è sicuramente un elemento che discosta il *Manifesto* della scrittura accademica, distanza dovuta anche dalla presenza di un lessico valutativo e strutture con verbi copulativi.

Si può sostenere quindi che il *Manifesto* possieda un registro alto e numerose strategie linguistiche tipiche di un testo accademico. Nonostante ciò, l'assenza di riferimenti bibliografici e la presenza di lessico valutativo e verbi copulativi impediscono di classificarlo come testo accademico in senso stretto.

Il *Manifesto* appare, invece, più vicino alla struttura e le intenzioni di un saggio politico-ideologico. Da un lato il testo è organizzato secondo una struttura logica con la quale l'autore presenta anzitutto una tesi centrale che sviluppa successivamente fornendo varie argomentazioni in merito. Dall'altro si configura come un discorso politico, poiché impiega strategie sintattiche e lessicali orientate alla persuasione di un pubblico ampio, in linea con la definizione proposta da Ana Albalat-Mascarell nel suo articolo *Approaches to the Analysis of Metadiscourse Features in Political Discourse* (2023). Quest'ultimo elemento risulta particolarmente interessante perché permette di aprire una discussione riguardo le strategie linguistiche orientate alla persuasione che l'autore inserisce all'interno del *Manifesto*.

Durante la valutazione del registro e della tipologia di testo, le componenti riscontrate come maggiormente rilevanti sono le variazioni lessicali verso una terminologia più concreta, l'inserimento di lessico valutativo e la mancanza di elementi strutturali, quali riferimenti

[63] Biber D., Conrad S. (2019). *Register, Genre, and Style*. ResearchGate.
[64] Balboni P. (2011). *Il monologo e lo scritto accademici*. Ca' Foscari.

bibliografici, tipici della scrittura accademica; sono esattamente questi i fattori che fanno si che il *Manifesto* si discosti da un testo accademico sebbene vi siano altre caratteristiche in linea con tale genere. La riflessione che ne consegue riguarda dunque le motivazioni che hanno indirizzato l'autore a compiere queste determinate scelte linguistiche

Una possibile interpretazione riprende il concetto di retorica presentato da Joachim Knape nel suo articolo *La teoria della retorica: questioni basilari* (2012). Secondo il professore Knape la retorica (in latino ars persuadendi quindi l'arte di persuadere) è definita come la padronanza di processi comunicativi orientati strategicamente al successo e si può dunque considerare un atto comunicativo pianificabile e intenzionale. Ciò significa che la retorica non si basa esclusivamente sul parlare bene, inteso come scegliere un lessico forbito o elegante, ma piuttosto nasce dalla capacità di scegliere e usare il linguaggio in modo consapevole e indirizzato ad ottenere un determinato effetto sul destinatario.

In questa prospettiva le scelte di utilizzare lessico valutativo, di inserire dei termini più vicini alla quotidianità, mantenendo comunque un lessico formale e autorevole e di omettere le fonti delle definizioni fornite si possono leggere come un tentativo di strutturare il linguaggio e l'intero *Manifesto* in maniera tale da costruire la retorica.

A tale interpretazione si può aggiungere un elemento ulteriore: la ripetizione sistematica di precisi *n-gram* aggettivo-sostantivo, che vedono affiancati sempre i medesimi elementi, produce un effetto che è possibile accostare al concetto di retorica espresso da Knape. La reiterazione infatti non assume una funzione meramente ridondante, ma concorre a costruire la forza persuasiva del testo rafforzando la coerenza interna dell'argomentazione, favorendo la memorizzazione delle idee principali e contribuendo a consolidarne la credibilità.

In quest'ottica il *Manifesto* appare come un testo in cui l'autore non si limita a descrivere la società cercando di farne emergere le criticità nella maniera più neutrale possibile, ma cerca di infondere nel lettore la percezione che quanto stia leggendo, poiché presentato in maniera così teorica e analitica, sia l'interpretazione corretta della realtà, fino ad arrivare al punto di prendere in considerazione la necessità di una rivoluzione per riformare il sistema. La ripetizione, in questo senso, non rende maggiormente vere le idee espresse, ma produce un effetto di inevitabilità e compattezza ideologica che può risultare estremamente persuasivo.

Nel complesso, l'analisi linguistica suggerisce che il *Manifesto* rappresenti un testo più vicino a un saggio politico-ideologico costruito attraverso l'utilizzo di precise strategie linguistiche. L'assenza di riferimenti bibliografici, il ricorso a lessico valutativo, la reiterazione di specifiche strutture lessicali e l'alternanza tra terminologia specialistica ed espressioni più accessibili contribuiscono infatti a conferire autorevolezza al discorso senza sottoporlo ai criteri di

verificabilità necessari in un testo accademico, verosimilmente finalizzato a orientare l'interpretazione del lettore.

Nel contesto della vicenda giuridica i risultati emersi dall'analisi linguistica hanno permesso di redigere un profilo più plausibile e meglio supportato dai dati linguistici rispetto a quello desumibile esclusivamente dalle scene del crimine e dal contenuto del *Manifesto*. Il profilo precedentemente stilato infatti vedeva come possibile autore un uomo contro il sistema, verosimilmente impiegato in una mansione lavorativa che richiedeva grandi capacità manuali e gli permetteva di essere a stretto contatto con gli aerei (alla luce del tentato attacco a un attacco di linea); a seguito dell'analisi dei dati linguistici emersi, però, l'attenzione si è spostata verso un uomo che aveva ricevuto un'istruzione piuttosto elevata che gli aveva permesso di riprendere diversi tratti tipici della scrittura accademica, rielaborandoli però all'interno di un testo con finalità prevalentemente persuasive. Integrando tali risultati all'analisi delle tematiche trattate e delle azioni criminose portate avanti è stato poi possibile formulare ulteriori ipotesi sul suo profilo. In particolare, si è dedotto che l'uomo potesse essere in linea con un profilo di un individuo non pienamente inserito nel contesto della società e che, presumibilmente, vivesse solo in un'area piuttosto isolata, condizione che gli permetteva di rimanere lontano dai meccanismi della società che tanto criticava e di creare gli ordigni esplosivi liberamente, senza attirare a sé sospetti.

Tramite ciò si intende sottolineare come i dati linguistici provenienti dalle analisi di linguistica forense non forniscano il nome e il cognome della persona che ha redatto il testo anonimo ma se considerate congiuntamente alle prove forensi possono orientare e restringere il campo dei possibili sospettati.

Furono proprio queste deduzioni infatti che, a seguito della pubblicazione del *Manifesto* e della segnalazione da parte di David Kaczynski (fratello di Theodore Kaczynski) che indicava come possibile autore del *Manifesto* proprio suo fratello, che le autorità ottennero il mandato di perquisizione per la capanna in cui viveva Kaczynski[65].

5.3 La verifica della profilazione attraverso la comparazione autoriale

Gli elementi che condussero Theodore Kaczynski davanti alla corte del Distretto del Montana (1996) e che contribuirono alla sua condanna, formalizzata attraverso un *guilty plea*, furono prevalentemente le prove linguistiche ricavate dal *Manifesto* che, a seguito del confronto con i vari documenti ritrovati ed esaminati nel suo capanno risultarono come coerenti. Alla luce di ciò, nel

[65] Wiehl L., Pulitzer L. (2020). *Hunting the Unabomber: The FBI, Ted Kaczynski, and the Capture of America's Most Notorious Domestic Terrorist*. Thomas Nelson

seguente lavoro di ricerca si ritiene interessante dedicare un capitolo all'analisi linguistica svolta sulla sua tesi di dottorato.

La tesi di dottorato di Theodore Kaczynski, intitolata *Boundary Functions*, è stata prodotta nel 1967, ventotto anni prima rispetto alla pubblicazione del *Manifesto*, presso l'Università del Michigan. Essa è una ricerca specialistica in ambito matematico nella quale Kaczynski sviluppa in maniera più approfondita la teoria delle *boundary functions* [funzioni al bordo]. Sebbene il contenuto di questo elaborato sia evidentemente differente rispetto a quello trattato nel *Manifesto*, la ricerca sviluppata da Kaczynski non viene analizzata con lo scopo di provare che i due testi sono stati redatti dalla medesima persona ma piuttosto per individuare quali caratteristiche della scrittura possano essere considerate espressione del suo idioletto e quali, invece, dipendono soprattutto dal tipo di testo. L'analisi, dunque, è stata condotta delineando anzitutto le convenzioni della scrittura matematica in modo tale da disporre un quadro di riferimento sulla base del quale valutare se determinate scelte linguistiche e stilistiche siano attribuibili all'autore oppure alle esigenze proprie del tipo di testo; per svolgere tale comparazione sono stati applicati i medesimi strumenti linguistici utilizzati nell'analisi del *Manifesto* così da estrapolare le *feature* lessicali, sintattiche, semantiche, strutturali e basate sul contenuto.

Per delineare le caratteristiche proprie dei testi matematici ci si è basati sull'articolo *Linguistic Conventions of Mathematical Proof Writing at the Undergraduate Level: Mathematicians' and Students' Perspectives* redatto da Kristen Lew e Juan Pablo Mejía-Ramos (2019). In questo documento Lew e Mejía-Ramos identificano un insieme di convenzioni linguistiche della scrittura matematica a partire da uno studio qualitativo basato su interviste a un gruppo di matematici e uno di studenti universitari; ad ogni gruppo sono state proposte diverse dimostrazioni matematiche, ciascuna delle quali possedeva una variazione che non rispettava una convenzione linguistica con l'obiettivo di domandare a ciascun partecipante se notavano qualcosa di scorretto nel passaggio proposto e in caso come l'avrebbero riscritto. Tramite la percezione degli esperti e degli studenti Lew e Mejía-Ramos tentano di ricostruire le consuetudini degli scritti matematici, tra le quali vi sono: l'impiego di un registro prevalentemente formale, il limitato ricorso ai pronomi personali, una rigorosa organizzazione logica del discorso e l'uso di una terminologia precisa e univoca per evitare ambiguità, sebbene vi sia il rischio di incorrere nella ripetizione degli stessi termini matematici.

Una volta delineate, sulla base della letteratura, le principali convenzioni che caratterizzano la scrittura matematica, è possibile procedere all'analisi linguistica della tesi di dottorato di Theodore Kaczynski.

Il primo aspetto preso in esame sono le *feature* lessicali, per le quali si è calcolato l'*F-score* di Heylighen & Dewaele usando la medesima formula:

F = [(noun freq. + adjective freq. + preposition freq. + article freq.) - (pronoun freq + verb freq. + adverb freq.) + 100] / 2

Attraverso l'impiego del *POS tagging* è stato possibile realizzare quantificare in maniera approssimativa la frequenza con la quale le varie categorie grammaticali compaiono nel testo, valori successivamente espressi in percentuale. Applicando la formula si ricava:

F = [(29% + 9% + 17 % + 7%) - (2% + 14% + 3%) + 100] / 2

F = 143 / 2

F = 71,5

L'*F-score* si aggira quindi intorno a 71-72, valore coerente con quello tipico della prosa accademica formale[66], caratterizzata precisamente da elevata densità nominale, ricorso limitato ai pronomi personali e alle voci verbali.

In merito alla tipologia di lessico utilizzata si può sostenere che Kaczynski prediliga l'utilizzo di un lessico astratto, derivante in parte dalla materia trattata ma in particolar modo dalla propensione verso la nominalizzazione. I termini che compaiono con maggior frequenza sono infatti theorem, lemma, proof, function, interval i quali vengono spesso affiancati da sostantivi quali *existence*, *continuity*, *convergence*.

Un ulteriore elemento da attenzionare riguarda la ricchezza lessicale, stimabile tramite l'indicatore linguistico *Type-Token Ratio*. Tenendo presente che il testo è composto da circa 12.000 parole, si contano approssimativamente 1.500–1.650 type; questi dati producono un TTR stimato di circa 13,2-13,5%. Tale valore deve tuttavia essere interpretato con cautela. È infatti opportuno ricordare che il TTR è una misura fortemente influenzata dalla lunghezza del testo che tende ad aumentare al diminuire del numero complessivo di parole[67]. Inoltre, trattandosi di una tesi di argomento matematico, il lessico risulta altamente ristretto e spesso ripetitivo dal momento che è necessario mantenere un'estrema precisione terminologica, indispensabile per garantire precisione concettuale. Di conseguenza, il valore stimato del TTR appare coerente con la natura del testo analizzato.

Un dato ulteriore osservabile dall'analisi della ricchezza lessicale è l'utilizzo di un inglese americano standard degli anni Sessanta[68]; la terminologia utilizzata, infatti, sebbene sia costituita in larga parte da termini tecnici che non presentano differenze ortografiche tra inglese britannico e americano, presenta l'ortografia americana.

---

[66] Heylighen F., Dewaele J. M. (1999). *Formality of Language: definition, measurement and behavioral determinants*. Internal Report, Center "Leo Apostel".

[67] Mitchell D. (2015). *Type-token models: a comparative study*. Journal of Quantitative Linguistics.

[68] Leech G., Hundt M., Mair C., Smith N. (2009). *Change in Contemporary English: A Grammatical Study*. Cambridge University Press.

Nel complesso, l'analisi delle *feature* lessicali evidenzia quindi un registro fortemente formale, caratterizzato da elevata densità nominale, terminologia specialistica e una marcata tendenza all'impersonalità. Tali caratteristiche individuate si riflettono anche nelle modalità con cui le informazioni vengono organizzate all'interno del periodo

In merito alle *feature* sintattiche, attraverso il calcolo delle subordinate per *T-unit* è possibile sostenere che la sintassi sia prevalentemente ipotattica, caratterizzata da un elevato grado di subordinazione. "*If f is defined in H, then the set of curvilinear convergence of f is the set of all points x∈X such that there exists some arc at x along which f approaches a limit.*"

Il periodo appena citato rappresenta un esempio significativo della complessità sintattica tipica della tesi; è infatti caratterizzato da una sintassi ipotattica formata dalla proposizione principale "*then the set of curvilinear convergence of f is the set of all points x∈X*" dalla quale dipendono gerarchicamente tre subordinate. Inoltre si osserva l'uso di costruzioni impersonali che rendono possibile presentare le informazioni senza far emergere alcun riferimento all'autore e indirizzando l'attenzione esclusivamente sull'oggetto matematico di cui si sta parlando.

Questa struttura sintattica rispecchia le caratteristiche tipiche della scrittura matematica; tuttavia, ciò che risulta particolarmente rilevante è l'elevato livello di organizzazione gerarchica con cui Kaczynski costruisce i periodi, disponendo le diverse proposizioni secondo un preciso ordine che gli permette di concentrare in un solo periodo un'elevata densità informativa.

Si può, quindi, concludere che la sintassi del testo che si sta analizzando sia tendenzialmente ipotattica e caratterizzata da un'elevata complessità derivante dalle numerose proposizioni che si susseguono nei vari periodi. A ciò si accosta un uso frequente della voce passiva e di costruzioni impersonali che favoriscono la costruzione di un testo nel quale i contenuti vengono presentati nella maniera più oggettiva e neutrale possibile. Di conseguenza dal punto di vista semantico si rileva una totale assenza di lessico o costruzioni che veicolano le emozioni o le valutazioni dell'autore.

Le caratteristiche individuate a livello lessicale, sintattico e semantico trovano ulteriore conferma dalle *feature* strutturali del documento. L'organizzazione del testo infatti segue una progressione logica scandita dalla suddivisione in capitoli e sezioni, propriamente intitolati, attraverso la quale ogni parte introduce i concetti e gli strumenti necessari per la comprensione per la sezione successiva. In aggiunta, è presente una bibliografia conclusiva, elemento tipico della scrittura accademica[69] e coerente con la natura della tesi di dottorato.

Sulla base di quanto emerso da questa analisi vi sono una serie di caratteristiche coerenti con le convenzioni proprie della scrittura matematica accademica. Tuttavia, alcuni elementi osservati

[69] Balboni P. (2011). *Il monologo e lo scritto accademici*. Ca' Foscari.

sembrano andare oltre le necessità del genere e potrebbero essere interpretati come tratti caratteristici dell'autore, indipendenti dalla tipologia di testo che sta venendo redatto.

Anzitutto emerge un forte orientamento all'astrazione concettuale. Sebbene i concetti matematici siano naturalmente nozioni astratte, la tendenza all'astrazione si osserva anche nel *Manifesto*, nel quale le argomentazioni presentate appartengono ad un dominio completamente diverso e riguardano temi quali la società, i cittadini, la tecnologia. Nonostante la differente natura degli argomenti, Kaczynski privilegia anche in questo caso la rappresentazione astratta dei fenomeni. L'elevata nominalizzazione è altresì un componente riscontrato in entrambi gli scritti attraverso il quale l'autore costruisce il discorso. A ciò si aggiunge la propensione per una sintassi ipotetica e complessa, caratterizzata da periodi spesso estesi, nei quali si incontrano sintagmi nominali, spesso anche articolati e le proposizioni che ne fanno parte sono organizzate secondo una precisa gerarchia. Infine entrambi gli scritti evidenziano un'ampia ricchezza lessicale, prevalentemente formale e specialistica, attraverso la quale l'autore costruisce definizioni precise e rigorose.

È quindi possibile ipotizzare che tali elementi costituiscano una componente relativamente stabile della scrittura di Kaczynski. Pertanto nel caso giudiziario di Unabomber le prove sulle quali si è basata l'accusa sono state precisamente i risultati emersi anzitutto dall'analisi del solo *Manifesto* e successivamente dalla sua comparazione con altri scritti di Kaczynski, tra i quali la sua tesi di dottorato.

Tale confronto rappresenta quindi il momento conclusivo del percorso metodologico seguito in questo lavoro: la comparazione costituisce infatti la verifica finale della validità del processo di profilazione linguistica sviluppato nei capitoli precedenti. In questa prospettiva, i risultati ottenuti mostrano come i dati linguistici non derivino da mere intuizioni ma bensi da un metodo scientifico basato sull'applicazione degli strumenti propri della linguistica forense. Ciò conferma, entro i limiti propri della disciplina, il valore di questa materia come strumento di supporto alla ricostruzione del profilo di un autore anonimo

CAPÍTULO 6 - LÍMITES, PERSPECTIVAS Y NUEVOS RETOS DE LA PERFILACIÓN DE AUTOR EN LA LINGÜÍSTICA FORENSE

6.1 Limitaciones metodológicas y aplicativas de la perfilación de autor

La reconstrucción del perfil lingüístico propuesta en el presente trabajo pone de manifiesto el potencial de la lingüística forense como herramienta de investigación. Sin embargo, los tribunales suelen mostrarse cautelosos a la hora de considerar las pruebas lingüísticas, ya que esta disciplina presenta limitaciones tanto metodológicas como aplicativas. Esto se debe a que la lingüística forense hereda la incertidumbre propia de la lingüística.

La lingüística no se considera de forma unánime una ciencia dura, puesto que trabaja con datos altamente variables y heterogéneos y formula hipótesis que difícilmente pueden someterse a una verificación experimental (Marotta y Tamponi, 2021). Por ello, se encuadra más adecuadamente dentro de las denominadas ciencias blandas, aunque ocupa una posición particular dentro de esta clasificación. Esta singularidad se debe a que el objeto de estudio de la lingüística son precisamente las lenguas, sobre las cuales la disciplina ha desarrollado teorías y métodos con un grado de cientificidad relativamente elevado.

En consecuencia, la lingüística comparte los mismos intereses que las ciencias humanas (clasificadas como ciencias blandas), pero aplica métodos empíricos y verificables propios de las ciencias duras. Por esta razón, ha sido definida como «la ciencia más dura entre las ciencias blandas» (Marotta y Tamponi, 2021).

Este grado de incertidumbre es heredado por la lingüística forense, ya que esta surge en un período en el que la lingüística pierde su unidad teórica. Con la aparición de la gramática generativa de Chomsky, la atención se desplaza hacia la denominada competencia lingüística, entendida como el conocimiento innato que permite a los hablantes producir y comprender un número infinito de enunciados, con el objetivo de construir una teoría general del lenguaje. Sin embargo, la realidad del lenguaje es mucho más compleja, ya que existen enormes diferencias entre los hablantes derivadas de factores como la edad, la clase social, las emociones o las situaciones comunicativas, aspectos que esta nueva orientación relegaba a un segundo plano. Como reacción a ello, diversos investigadores desarrollaron enfoques alternativos o complementarios, lo que dio lugar a una auténtica fragmentación de la disciplina, de la que surgiría posteriormente el ámbito especializado de la lingüística forense.

Como consecuencia, la lingüística forense encuentra dificultades para constituir una prueba científica, especialmente cuando el objetivo es determinar quién ha escrito un texto, ya que no es

posible asumir científicamente que un individuo posea un estilo lingüístico suficientemente estable y distintivo como para permitir su identificación inequívoca.

Por este motivo, dentro de la lingüística forense orientada a la perfilación de autor se identifican limitaciones tanto metodológicas como aplicativas. En concreto, cabe preguntarse cuáles son las dificultades que surgen durante el procedimiento mediante el cual los especialistas formulan hipótesis sobre el posible autor de un texto anónimo y cuáles son los límites dentro de los cuales la lingüística forense puede utilizarse con fines de investigación. Desde esta perspectiva, se examinan no solo las distintas fases que Nini (2018) considera necesarias para llevar a cabo una perfilación de autor, sino también los resultados obtenidos, con el objetivo de identificar posibles debilidades del método científico.

El primer aspecto en el que conviene centrar la atención se refiere a la metodología y al proceso necesarios para elaborar un perfil plausible. Al revisar las distintas fases del procedimiento, pueden identificarse tres momentos especialmente críticos, ya que, si no se interpretan correctamente, podrían desviar el proceso de perfilación: el análisis cuantitativo y cualitativo de las *features*, la identificación del estilo plausible del autor y la comparación de los resultados obtenidos con los estudios sociolingüísticos.

El análisis cuantitativo y cualitativo de los rasgos lingüísticos del texto objeto de estudio constituye la primera fase crucial para que las etapas posteriores, dedicadas a la interpretación de dichos rasgos y a la correlación entre un rasgo lingüístico y un dato demográfico, produzcan resultados coherentes con el texto analizado. Con el objetivo de elaborar una representación lo más completa posible de las características lingüísticas de un escrito, Mario Crespo Miguel, en su artículo *El análisis de autoría en Lingüística Forense: historia, concepción y revisión metodológica. Aplicación de la razón de verosimilitud a textos cortos en español* (2023), propone tres condiciones fundamentales.

En primer lugar, el especialista debe disponer de una cantidad suficiente de producción lingüística, ya que una muestra reducida puede poner de manifiesto características meramente circunstanciales. Esta problemática se debe a que, cuando un texto contiene pocas palabras, el número de rasgos lingüísticos observables disminuye considerablemente. En consecuencia, el experto dispone de menos datos sobre los que fundamentar su análisis, lo que dificulta la identificación de patrones estilísticos reales. Este aspecto constituye un elemento especialmente crítico dentro de todo el proceso, puesto que la frecuencia con la que aparecen determinados fenómenos lingüísticos determina si estos representan un dato suficientemente relevante como para ser considerado en el análisis. Así pues, en un texto breve una palabra, una expresión o un

determinado fenómeno lingüístico puede aparecer una única vez o incluso no aparecer; por ello, el análisis de los rasgos lingüísticos puede verse condicionado por la escasez de datos observables.

El segundo aspecto señalado por Crespo (2023) hace referencia a la necesidad de realizar un análisis multinivel. En efecto, no es posible delinear un perfil lingüístico mediante un examen cuantitativo y cualitativo de los rasgos centrado exclusivamente en un único elemento. Esta exigencia responde a dos motivos. Por un lado, deriva de la propia naturaleza del lenguaje, definido como un fenómeno complejo y multidimensional. Por otro lado, la convergencia de diversas elecciones lingüísticas independientes del registro o del contenido posee un mayor valor probatorio que el derivado de un único rasgo lingüístico. Por ello, el análisis debe adoptar una perspectiva multinivel que abarque todos los componentes de la escritura y del lenguaje (léxico, sintaxis, estructura, semántica y contenido).

El último aspecto identificado por Crespo se refiere a la necesidad de definir el registro y las temáticas abordadas en el texto, teniendo en cuenta que determinadas características lingüísticas pueden estar condicionadas por estos factores y no por elecciones personales del autor. Piénsese, por ejemplo, en el caso de dos textos redactados por dos jueces diferentes: es natural encontrar las mismas estructuras léxicas o un tipo de vocabulario similar. Sin embargo, estas características derivan de la naturaleza del texto, ya que los escritos jurídicos exigen el empleo de determinadas elecciones lingüísticas consideradas convencionales para este tipo de documentos. En consecuencia, al analizar ambos textos con el objetivo de extraer información sobre el estilo propio del autor, resulta necesario identificar aquellos rasgos que puedan considerarse independientes del registro y del contenido, puesto que son los que, con mayor probabilidad, constituyen elementos distintivos o recurrentes en los escritos de una determinada persona.

Por consiguiente, la fiabilidad del perfil no depende exclusivamente de un análisis correcto de los rasgos lingüísticos, sino también de la capacidad del especialista para distinguir aquellos que describen efectivamente el estilo individual del autor de los elementos condicionados por el contenido o por el contexto comunicativo. La principal dificultad de este procedimiento reside precisamente en que no existe una metodología objetiva y definitiva que permita diferenciar los rasgos derivados de las convenciones propias de la escritura de aquellos que pueden considerarse el resultado de una elección personal del autor.

La dificultad señalada anteriormente desempeña también un papel central en la deducción del estilo plausible del autor. En este sentido, Crespo (2023) sostiene que, durante este procedimiento, es esencial que los rasgos lingüísticos considerados sean discriminantes y relevantes. Esta afirmación implica que dichos rasgos, para poder considerarse relevantes desde el punto de vista del análisis lingüístico, deben ser lo suficientemente peculiares como para permitir distinguir a un

determinado autor de otros. Asimismo, es necesario que no se trate de elementos aleatorios o carentes de significado, sino que representen fenómenos lingüísticos reales, reconocidos por la lingüística como indicadores de una elección léxica, sintáctica, ortográfica o morfológica.

Aunque la definición propuesta por Crespo orienta al especialista en la identificación de los datos lingüísticos relevantes para la perfilación de autor, no existen criterios precisos y claramente definidos que permitan distinguir con un elevado grado de certeza cuáles son las características más discriminantes frente a otras. No obstante, es necesario subrayar la importancia de examinar el conjunto de los rasgos lingüísticos observados, de modo que pueda obtenerse una visión suficientemente completa. En efecto, ninguna característica lingüística puede considerarse decisiva si se analiza de forma aislada, ya que, como se ha señalado anteriormente, el estilo de un autor depende de la interacción de diferentes rasgos. De hecho, el estilo de un autor no deriva exclusivamente del empleo de determinadas palabras características o de expresiones recurrentes. Por el contrario, es el resultado de la interacción de numerosos rasgos lingüísticos que, considerados en su conjunto, tienden a mantenerse relativamente estables a lo largo del tiempo y en distintos contextos comunicativos. Además, la atribución de autor se basa en una valoración probabilística y no determinista; las inferencias formuladas derivan, en efecto, de la convergencia de múltiples evidencias.

La última fase en la que pueden observarse posibles limitaciones metodológicas es la comparación de los resultados obtenidos mediante el análisis lingüístico con los estudios sociolingüísticos presentes en la literatura de referencia. Este procedimiento permite pasar de la observación de un rasgo lingüístico a la valoración de su verdadero valor en la atribución de autor, mediante la comprobación de la distribución de dicha característica en la población de referencia. La principal dificultad reside precisamente en los estudios disponibles. Para elaborar un perfil lo más preciso posible, sería necesario disponer de un corpus de referencia perfectamente comparable con el texto objeto de análisis; sin embargo, la literatura existente se basa en corpus limitados a determinados períodos históricos, áreas geográficas, variedades lingüísticas o géneros textuales. Por consiguiente, los rasgos lingüísticos no permiten identificar directamente una categoría social, sino únicamente establecer una correlación estadística (Nini, 2018). En consecuencia, el especialista debe interpretar los datos con cautela, teniendo siempre presentes los límites de generalización de los estudios de referencia. El perfil obtenido, por tanto, no describe con certeza al autor, sino que representa la hipótesis más plausible a la luz de las evidencias lingüísticas disponibles y de la literatura existente.

A la luz de las limitaciones metodológicas expuestas, resulta oportuno centrar la atención también en las limitaciones aplicativas de la disciplina, las cuales no dependen del método en sí

mismo, sino de la aplicación de la teoría a los casos concretos y a las diferentes particularidades que estos pueden presentar.

En el ámbito de la lingüística forense orientada a la perfilación de autor también se identifican limitaciones aplicativas, es decir, aquellas relacionadas con el uso práctico de la disciplina. Desde esta perspectiva, el interés se centra en determinar cuándo y hasta qué punto la lingüística forense puede emplearse efectivamente en una investigación o en un proceso judicial. En este caso, la atención deja de dirigirse a la corrección del método para plantear, partiendo de la premisa de que la metodología es adecuada, en qué situaciones la perfilación de autor puede resultar insuficiente o incluso no ser aplicable.

La primera limitación se refiere, evidentemente, a la imposibilidad de identificar de manera inequívoca al autor de un texto. En el caso concreto del estudio de caso analizado, aunque el perfil elaborado señala como posible autor del Manifiesto a un hombre de edad madura, con un elevado nivel educativo y probablemente perteneciente a una clase social media-alta, ello no implica que el especialista pueda afirmar con certeza que el autor del documento sea Theodore Kaczynski. A la luz de estos resultados, lo que fue posible fue orientar la investigación hacia un grupo más reducido de individuos, entre los cuales podía encontrarse el verdadero autor, y solicitar una orden de registro. En consecuencia, la disciplina permite reducir el número de sospechosos y orientar con mayor precisión las investigaciones, pero no conducir directamente a una condena.

Esta limitación pone de manifiesto que la lingüística forense no puede actuar de forma autónoma, sino que debe integrarse con otros elementos probatorios obtenidos durante la investigación, como pruebas biológicas (ADN), testimonios o escuchas telefónicas. Por este motivo, en los procedimientos judiciales los jueces pueden atribuir a la lingüística forense un valor probatorio distinto al de otras pruebas científicas, precisamente porque sus conclusiones se fundamentan en valoraciones probabilísticas.

Otra limitación en la aplicación de la disciplina está relacionada con los tipos de textos sobre los que es posible realizar un análisis lingüístico. Las principales dificultades afectan a los textos breves, los documentos redactados por más de un autor y aquellos en los que se observa el fenómeno del *style masking*. En cuanto a los textos de reducida extensión, las dificultades que pueden surgir durante su análisis ya han sido expuestas anteriormente; por ello, la atención se centrará exclusivamente en las otras dos categorías.

Los documentos redactados por más de un autor, conocidos como textos colaborativos, son aquellos en cuya elaboración participan varias personas. Este tipo de escritos constituye una limitación para la aplicación de la lingüística forense, ya que esta disciplina se basa en la identificación de los hábitos lingüísticos característicos de un determinado autor (idiolecto). En

consecuencia, cuando un documento ha sido redactado o revisado por diferentes personas, las características lingüísticas observadas ya no reflejan el estilo de un único autor, sino un conjunto de rasgos pertenecientes a varios individuos. Este hecho dificulta aún más el proceso de inferencia sobre el posible perfil del autor, ya que impide determinar con suficiente fiabilidad si un determinado rasgo lingüístico refleja realmente el estilo de quien redactó el texto.

La misma dificultad se observa en aquellos textos en los que el autor modifica deliberadamente su estilo de escritura. Este fenómeno, conocido como *style masking*, consiste en que el autor altera de forma voluntaria y consciente su propio estilo de escritura (Ruiz Urbón, 2025). Esta estrategia puede adoptar distintas formas, ya que el autor puede introducir modificaciones en el plano léxico, sintáctico o del registro e, incluso, cometer errores de manera intencionada. No obstante, el objetivo suele ser alejarse lo máximo posible de su propio idiolecto, recurriendo a expresiones o formas menos características con el fin de hacer el texto lo más anónimo posible.

A pesar de ello, la literatura especializada señala que, aunque el *style masking* reduce de manera evidente la eficacia de la perfilación de autor, no consigue eliminar por completo el idiolecto de un individuo (Nini, 2018). Existen determinadas elecciones lingüísticas que se realizan de forma inconsciente y automática, como la longitud de las oraciones, la preferencia por una sintaxis hipotáctica frente a una paratáctica o la tendencia a estructurar el discurso siguiendo una determinada lógica. Estos aspectos resultan mucho más difíciles de controlar, precisamente porque suelen constituir hábitos automáticos ya interiorizados por el autor.

Son precisamente estos elementos en los que centra su atención el especialista, ya que, aunque no permiten determinar con certeza si el texto analizado es espontáneo o ha sido deliberadamente alterado, representan rasgos que, según demuestra la literatura especializada, tienden a permanecer estables y resultan difíciles de modificar.

6.2 Perspectivas metodológicas para el desarrollo de la perfilación de autor

Aunque las limitaciones expuestas anteriormente no representan una dificultad significativa en el análisis realizado en esta investigación, principalmente porque el material lingüístico disponible era lo suficientemente amplio como para permitir la identificación y cuantificación de un número adecuado de datos lingüísticos, resulta fundamental tener presente la existencia de estas limitaciones. Esta conciencia permite llevar a cabo un análisis lingüístico e interpretar los datos con la cautela necesaria, evitando extraer conclusiones excesivamente categóricas.

Al mismo tiempo, la evaluación de estas limitaciones abre la puerta a una reflexión sobre las posibles líneas de desarrollo futuro de la investigación, orientadas a reforzar el rigor metodológico de la disciplina y a perfeccionar las herramientas de análisis disponibles. Desde esta perspectiva, el

artículo de Cristina Ruiz Urbón, *Aplicación de la lingüística forense a los estudios de atribución de autoría literaria. Principios teóricos, metodológicos y conclusivos* (2025), resulta especialmente relevante, ya que ofrece una reflexión sobre las posibles líneas de evolución de la investigación en este ámbito y propone metodologías destinadas a reducir, en la medida de lo posible, el impacto de estas limitaciones.

En concreto, Ruiz Urbón insiste en la importancia de que la lingüística no presente sus resultados como afirmaciones categóricas, sino que centre la atención en la relevancia de las evidencias obtenidas en términos probabilísticos. En este sentido, la autora aboga por un mayor uso, especialmente en el ámbito de la perfilación de autor, de los modelos bayesianos, es decir, procedimientos de inferencia estadística fundamentados en el teorema de Bayes.

El teorema de Bayes describe el procedimiento mediante el cual una probabilidad inicial se actualiza a medida que se incorporan nuevas evidencias, relacionando dos elementos fundamentales: la probabilidad a priori (*prior*) y la verosimilitud (*likelihood*)[70]. La probabilidad a priori representa el grado de conocimiento sobre una hipótesis antes de observar los datos, mientras que la verosimilitud expresa hasta qué punto los datos observados son compatibles con una determinada hipótesis. La combinación de ambas permite obtener la probabilidad a posteriori (posterior), es decir, una estimación de la probabilidad de la hipótesis inicial a la luz de las nuevas evidencias.

El principal cambio metodológico derivado de la aplicación de este teorema a la lingüística reside en que los modelos bayesianos se fundamentan en una formulación matemática fija. La formulacion es: $P(\beta \mid D) = P(D \mid \beta) \cdot P(\beta) / P(D)$ en la que $P(\beta)$ representa la distribución a priori, $P(D \mid \beta)$ la verosimilitud, $P(\beta \mid D)$ la distribución a posteriori y $P(D)$ una constante de normalización.

Dada la naturaleza de este trabajo, no se considera necesario profundizar en los aspectos matemáticos de la fórmula más allá de la explicación del significado de cada uno de sus términos. Lo verdaderamente relevante es que la aplicación de este modelo a la perfilación de autor permite obtener una estimación probabilística de los parámetros analizados y cuantificar el grado de incertidumbre inherente a toda inferencia estadística.

Desde esta perspectiva, el objetivo del análisis deja de ser determinar con certeza si un individuo concreto es el autor de un texto y pasa a consistir en evaluar en qué medida la evidencia lingüística observada respalda una hipótesis frente a otra.

En el ámbito de la lingüística forense, el método que actualmente se considera más riguroso es el Likelihood Ratio (Crespo Miguel, 2023; Ruiz Urbón, 2025). Este instrumento, derivado del

[70] Imarisio L. (2025). *Modelli Bayesiani per lo studio dell'interazione tra variabili ambientali e l'attività di Mustela Nivalis*. Politecnico di Torino.

razonamiento bayesiano, expresa la fuerza con la que una determinada evidencia modifica la plausibilidad de las hipótesis planteadas. En la práctica, cuando se analiza un texto anónimo, se observan determinados fenómenos lingüísticos y, posteriormente, se establecen posibles correlaciones entre dichos fenómenos y determinadas variables demográficas a partir de la literatura científica disponible. En este contexto, el cálculo del Likelihood Ratio permite cuantificar cuántas veces una hipótesis resulta más probable que otra, proporcionando así una valoración científica de la fuerza de la evidencia obtenida.

Aunque los modelos bayesianos presentan una mayor solidez desde el punto de vista teórico, su aplicación práctica se enfrenta todavía a importantes dificultades, derivadas principalmente de la escasez de estudios poblacionales, es decir, investigaciones que analicen la distribución de determinadas elecciones lingüísticas dentro de grupos específicos de hablantes. Para aplicar estos modelos es imprescindible conocer la distribución de las características lingüísticas en una población de referencia, ya que, en ausencia de estos datos, no es posible verificar la verosimilitud de las hipótesis formuladas. En otras palabras, falta la base empírica necesaria para determinar el grado de compatibilidad o incompatibilidad entre una determinada evidencia y la hipótesis planteada.

Por este motivo, el desarrollo de los modelos bayesianos aplicados a la perfilación de autor continúa encontrándose en una fase de evolución. No obstante, Ruiz Urbón (2025) introduce en su trabajo un aspecto especialmente innovador: el concepto de estandarización. A la luz de la literatura existente y de las investigaciones realizadas hasta el momento, pretender eliminar completamente el grado de incertidumbre inherente a la lingüística forense, y en particular a la perfilación de autor, resulta una aspiración poco realista. En consecuencia, parece más razonable orientar el desarrollo de la disciplina hacia procedimientos metodológicos cada vez más rigurosos y transparentes, que permitan aprovechar el conocimiento disponible evitando interpretaciones probabilísticas divergentes.

El proceso de estandarización responde precisamente a este objetivo. Su finalidad consiste en establecer un conjunto de criterios objetivos destinados a guiar la asignación del grado de probabilidad de las inferencias obtenidas, reduciendo así la influencia de la subjetividad inherente a la interpretación del especialista.

A la luz de lo expuesto, puede afirmarse que la investigación en lingüística forense y perfilación de autor no se orienta tanto a eliminar las inevitables limitaciones que caracterizan este tipo de análisis como a desarrollar metodologías cada vez más estandarizadas. La incorporación de herramientas como los modelos bayesianos, el Likelihood Ratio y criterios objetivos para la asignación del grado de probabilidad refleja el esfuerzo por desarrollar instrumentos que permitan

fundamentar metodológicamente la interpretación de las inferencias lingüísticas y reforzar el trabajo de los especialistas. Desde esta perspectiva, el progreso de la disciplina no consiste en sustituir el juicio del experto, sino en proporcionarle herramientas que aumenten la objetividad y la coherencia interpretativa de sus conclusiones, contribuyendo así a incrementar la fiabilidad de los resultados alcanzados.

6.3 La inteligencia artificial como nuevo desafío para la lingüística forense

A pesar de las amplias perspectivas de futuro para la evolución y el perfeccionamiento de la metodología de la disciplina, el desarrollo tecnológico, especialmente en el ámbito de la inteligencia artificial, podría dar lugar a nuevas dificultades con las que los especialistas tendrán que enfrentarse.

La expresión inteligencia artificial apareció por primera vez en 1955, cuando John McCarthy y otros investigadores desarrollaron el *Dartmouth Summer Research Project on Artificial Intelligenc*e con el propósito de demostrar que cualquier aspecto del aprendizaje o de la inteligencia humana podía describirse con un grado de precisión suficiente como para permitir que una máquina lo simulara (Suchman, 2023). Este planteamiento sentó las bases para la investigación y el desarrollo de la disciplina, aunque dicha evolución no siguió un recorrido lineal. En efecto, los primeros investigadores en inteligencia artificial partían del supuesto de que la inteligencia y el conocimiento humanos constituían fenómenos objetivos y universales, independientes de factores como el cuerpo o el contexto social y cultural, y que, por tanto, podían representarse mediante un conjunto de reglas y reproducirse por una máquina.

Sin embargo, esta concepción fue objeto de numerosas críticas por parte de investigadores que cuestionaban precisamente su carácter excesivamente universalista. Según estas perspectivas, asumir que todos los individuos razonan de la misma manera y comparten los mismos conocimientos, con independencia del contexto social y cultural en el que se encuentran, supone una simplificación excesiva de la realidad. Aunque estas reflexiones introdujeron una nueva forma de entender la inteligencia y el conocimiento humanos, poniendo el énfasis en la importancia de la cultura, las tradiciones, las emociones y las experiencias individuales, los supuestos fundamentales de la inteligencia artificial apenas han cambiado. En la actualidad se ha pasado de la inteligencia artificial simbólica a enfoques basados en el machine learning. La idea de que los sistemas funcionaban mediante reglas explícitamente programadas por los desarrolladores, que posteriormente la máquina utilizaba para razonar, ha sido sustituida por un modelo según el cual el sistema aprende a partir de un gran número de ejemplos y calcula regularidades estadísticas sobre las que genera sus respuestas. No obstante, aunque ambos enfoques suelen combinarse en los

sistemas actuales, permanece inalterado el supuesto fundamental de que la inteligencia puede formalizarse y, en consecuencia, reproducirse mediante un ordenador.

Este presupuesto constituye probablemente el aspecto más interesante desde la perspectiva de la lingüística forense y, en particular, de la perfilación de autor.

La irrupción de la inteligencia artificial ha introducido, sin duda, nuevos desafíos para este ámbito, derivados de diversos factores. En primer lugar, destacan las dificultades asociadas a los modelos de inteligencia artificial generativa. Estas herramientas son capaces de producir textos lingüísticamente correctos y adaptados a las instrucciones proporcionadas por el usuario. En la mayoría de los casos, los textos generados son el resultado de una colaboración entre una persona y el sistema: el autor proporciona un prompt en el que especifica las modificaciones deseadas, las cuales pueden consistir, por ejemplo, en la imitación de un determinado registro o estilo. A partir de estas instrucciones, el sistema genera un texto que el usuario puede aceptar tal como ha sido producido o modificar mediante nuevas indicaciones.

Desde el punto de vista del análisis lingüístico, este constituye el principal desafío que la inteligencia artificial ha introducido en la disciplina, ya que en los textos elaborados o modificados mediante modelos de IA generativa desaparecen algunos de los presupuestos fundamentales para la identificación del idiolecto de un autor, dificultando así el proceso de perfilación.

Con el objetivo de identificar las modificaciones lingüísticas introducidas mediante el uso de la inteligencia artificial, en los últimos años se han desarrollado diversos programas de detección que, sin embargo, también presentan importantes limitaciones (Queralt, 2024). Estos algoritmos se basan en un procedimiento muy similar al empleado en la lingüística forense, ya que analizan determinadas características del texto y las comparan con modelos estadísticos elaborados a partir de textos generados por sistemas de inteligencia artificial. En consecuencia, estos modelos permiten identificar qué fenómenos lingüísticos aparecen con mayor frecuencia en los textos producidos por una máquina, lo que no conduce a una identificación definitiva de la inteligencia artificial, sino a una estimación de la probabilidad de que un texto presente características compatibles con las de un texto generado artificialmente. Esto implica la posibilidad de obtener falsos positivos y falsos negativos, así como la dificultad de determinar si un texto ha sido generado íntegramente por un sistema de inteligencia artificial o si ha intervenido también un autor humano. Estas limitaciones constituyen la principal debilidad de estos sistemas y explican por qué no pueden considerarse herramientas en las que la lingüística forense pueda confiar plenamente, ya que difícilmente proporcionan resultados científicamente concluyentes.

Además de la generación automática de textos, la inteligencia artificial ha hecho posible el desarrollo de los denominados *deepfakes* de voz, es decir, grabaciones que reproducen con gran

fidelidad la voz de una persona. Aunque esta tecnología pertenece al ámbito de la fonética forense, constituye igualmente un importante desafío para la autenticación de las pruebas, ya que hace cada vez más difícil distinguir entre grabaciones auténticas y contenidos generados artificialmente.

A pesar de que los problemas derivados del desarrollo de la inteligencia artificial son reales, continúan planteándose numerosos interrogantes acerca de la capacidad de estos sistemas para representar e interpretar fenómenos lingüísticos complejos. Estas dudas derivan de los supuestos generales sobre los que siguen fundamentándose los sistemas de inteligencia artificial, los cuales continúan infravalorando la importancia de determinados factores relacionados con la inteligencia, el conocimiento y la comunicación entre las personas. Como consecuencia, los modelos de inteligencia artificial son capaces de reproducir numerosos aspectos del lenguaje, como el registro, el léxico o la sintaxis, pero no pueden considerarse equivalentes a la experiencia humana, al contexto comunicativo ni a la intencionalidad del autor, elementos que continúan siendo fundamentales en el análisis lingüístico.

Por todo ello, puede afirmarse que los sistemas de inteligencia artificial no vuelven obsoletas las metodologías de la lingüística forense ni de la perfilación de autor, aunque sí modifican necesariamente sus herramientas y objetivos. En la actualidad es posible analizar textos redactados íntegramente por una persona, generados por un sistema de inteligencia artificial o elaborados mediante la colaboración entre ambos. En consecuencia, estas nuevas tecnologías no eliminan la necesidad del análisis lingüístico, sino que hacen imprescindible desarrollar nuevas técnicas destinadas a evaluar el origen de los textos.

En este sentido, Suchman (2023) señala que los enfoques interdisciplinarios, que integran conocimientos procedentes de la lingüística, la informática, la inteligencia artificial, el derecho y la ciberseguridad, constituyen una de las estrategias más eficaces. La necesidad de este tipo de aproximaciones radica en que, cuando se analiza un texto, ya no basta con preguntarse quién podría ser su autor, sino que deben considerarse numerosos factores cuya evaluación requiere competencias especializadas procedentes de distintas disciplinas. En consecuencia, la inteligencia artificial no sustituye al lingüista forense, sino que transforma su labor en una actividad necesariamente interdisciplinaria, en la que el análisis del lenguaje sigue ocupando un papel central, pero debe complementarse con otros ámbitos de conocimiento para garantizar la fiabilidad de las pruebas.

CONCLUSIONI

Come anticipato nell'introduzione, la presente ricerca nasce dall'interrogativo riguardante fino a che punto le scelte linguistiche compiute durante la scrittura mostrano di ognuno di noi. Il caso del manifesto *Industrial Society and Its Future* raffigura un contesto particolarmente significativo per affrontare questa domanda, poiché rappresenta un documento sufficientemente esteso dal quale è possibile estrapolare sufficienti indizi linguistici utili per l'analisi linguistica. A ciò si aggiunge il fatto che la vicenda ha segnato uno spartiacque nel riconoscimento della linguistica forense come metodo scientifico attendibile in grado di produrre risultati con un effettivo peso probatorio; non è un caso, infatti, che quanto risultato dall'analisi linguistica del *Manifesto* durante il processo investigativo nel caso di Unabomber, abbia soddisfatto i requisiti necessari della *probable cause* affinché potesse essere emesso un mandato di perquisizione della capanna di Kaczynski.

La seguente tesi, dunque, ha cercato di presentare un'analisi puramente linguistica del *Manifesto*, esaminato come testo anonimo, con l'obiettivo di ricavare informazioni inerenti all'autore e valutare la tipologia di inferenze che si potevano produrre. L'analisi linguistica è stata svolta su più livelli, partendo dal contenuto, focalizzandosi sugli aspetti lessicali, sintattici e strutturali e concludendo con delle considerazioni riguardo la semantica.

I risultati ottenuti evidenziano un'elevata densità nominale e un'ampia ricchezza lessicale, due elementi che alzano il valore dell'F-score, identificando così il lessico del Manifesto come prevalentemente formale. A ciò si aggiunge una tendenza all'ipotassi, dove si susseguono frasi abbastanza lunghe e complesse nelle quali talvolta si osservano dei sintagmi nominali articolati. Dal punto di vista strutturale e semantico, infine, si evidenzia un'organizzazione logica del discorso che appare imitare le convenzioni della struttura accademica; nonostante ciò il Manifesto non risulta appartenere a tale genere, bensì è stato identificato come un documento maggiormente vicino a un saggio politico-ideologico.

Tali conclusioni hanno permesso di ipotizzare il pensiero e le motivazioni che hanno portato Kaczynski non solo a produrre il Manifesto ma verosimilmente anche gli attentati. Tuttavia, le considerazioni finali alle quali si è giunti non si esauriscono a queste dal momento che si propone altresì un possibile profilo dell'autore. Tale ricostruzione identifica un individuo di sesso maschile, la cui età sociale sembra aver favorito il consolidamento di una certa familiarità con i registri formali della lingua e lo sviluppo di un ampio repertorio lessicale. Le seguenti conclusioni, sebbene non rappresentino prove definitive, si basano su dati linguistici analizzati tramite gli strumenti propri della linguistica forense in ambito di profilazione autoriale, che conferiscono ai risultati che si sono ottenuti un valore probabilistico che in un contesto forense può assumere un valore probatorio sufficiente da essere utilizzati come un elemento indiziario.

In un secondo momento tali risultati sono stati messi in correlazione con quanto evidenziato dagli studi sociolinguistici presentati da Nini (2018) nella sua tesi *Developing forensic authorship profiling* con l'obiettivo di conferire eventuali corrispondenze tra tratto linguistico e dato demografico.

L'analisi si conclude con un momento di verifica dei dati ottenuti tramite una comparazione di quest'ultimi con quanto emerso dall'analisi linguistica della tesi di dottorato di Kaczynski *Boundary Functions*. L'obiettivo era quello di mostrare la presenza delle medesime caratteristiche linguistiche in due testi che, sebbene scritti dallo stesso autore, presentano differenze in merito al registro, al contenuto e al contesto in cui sono stati scritti; grazie a ciò è stato possibile confermare che le scelte linguistiche evidenziate nel Manifesto e riscontrate altresì nella tesi di dottorato siano indipendenti dal contenuto, dal registro e dal contesto e dunque siano verosimilmente riconducibili all'autore stesso, in particolare al suo idioletto.

Il seguente lavoro di ricerca, dunque, tramite un'analisi puramente linguistica intende evidenziare le potenzialità del linguaggio di rivelare informazioni dell'autore che trascendono il suo pensiero e le sue intenzioni; così facendo è possibile giungere a una risposta alla domanda di ricerca proposta che vede la lingua come un fenomeno che, se analizzato con gli strumenti adeguati, è in grado di mostrare informazioni inerenti a un autore che trascendono le sue intenzioni o motivazioni, ma rivela altresì dati riguardanti l'idioletto di un individuo e ricavarne così informazioni sul suo contesto sociale, geografico e biologico.

Questo potenziale rappresenta un aspetto interessante che può avere anche un peso probatorio in ambito forense. Di fatto sebbene la disciplina della linguistica e in particolare della profilazione autoriale non siano in grado di fornire dati sufficienti per condurre a una condanna il loro ruolo ausiliario può essere cruciale; le prove che vengono prodotte infatti possono fornire contributi orientativi, indiziari o corroborativi, rappresentando così una disciplina con un effettivo valore probatorio.

Con l'obiettivo di fornire una quadro che sia il più completo possibile della materia e delle sue capacità, metodologiche e applicative, la tesi si conclude fornendo una riflessione riguardo la coesistenza della disciplina della profilazione autoriale e dei sistemi di intelligenza artificiale. Tale discussione prende il via dall'articolo di Sheila Queralt (2024) *Los retos de la lingüística forense en la era de la inteligencia artificial* per poi estenderla in particolare alla disciplina di profilazione. Attraverso questo approfondimento si intende concentrare l'attenzione sulla necessità di sviluppare approcci diversi in grado di far fronte alle nuove sfide derivanti dall'intelligenza artificiale, affinché la linguistica forense e la profilazione autoriale non vengano considerate obsolete in un'era in cui

l'idioletto di un autore può essere modificato e camuffato da delle macchine in grado di riprodurre il linguaggio umano.

La conclusione che si intende presentare, dunque, vede la disciplina come un ramo della linguistica che presenta enormi potenzialità nel campo forense dal momento che è in grado di formulare inferenze probabilistiche sulle motivazioni, l'ideologia e soprattutto sui dati demografici di un autore, partendo dal linguaggio. Ciononostante, è importante tenere a mente che nella società di oggi dominata dai sistemi di intelligenza artificiale la domanda sulla quale bisognerebbe maggiormente focalizzarsi non è più quanto il linguaggio rigela di un autore ma quanto di quel linguaggio è prodotto da una macchina e quanto è frutto di un essere umano.

BIBLIOGRAFIA

Albalat-Mascarell, A. (2023). *Approaches to the analysis of metadiscourse features in political discourse*. Complutense Journal of English Studies.

Algeri, L. (2017). *Il testimone vulnerabile*. Giuffrè Editore.

Art. 499 c.p.p. Gazzetta Ufficiale della Repubblica Italiana.

Balboni, P. E. (2011). *Lo stile accademico nel monologo e nella scrittura.* Il discorso accademico. Lingue e pratiche disciplinari.

Benigno V. (2013). *Ricchezza lessicale e uso delle collocazioni in produzioni scritte di italiano L2 e italiano L1*. Varietà dei contesti di apprendimento linguistico.

Beninca', P. & C. Tortora, 2011. *Grammatica generativa e variazione*. Studi italiani di linguistica teorica ed applicata.

Biber D., Conrad S. (2019). *Register, Genre, and Style*. ResearchGate.

Boland J. E. (2023). *Lexical Ambiguity*. Oxford Bibliographies in Linguistics.

California State Portal, 2005.

Cava A., M. (2011). *Lessico valutativo e scrittura accademica: un approccio corpus-based*. Studii și Cercetări Filologice. Seria Limbi Străine Aplicate.

Chiche A., Yitagesu B. (2022). *Part of speech tagging: a systematic review of deep learning and machine learning approaches*. Journal of Big Data 9, 10.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. University of Maryland.

Costituzione della Repubblica Italiana (art. 24). Gazzetta Ufficiale della Repubblica Italiana.

Coulthard, M. (2010). *Forensic linguistics: The application of language description in legal contexts*. In Linguistique légale et demande sociale: Les linguistes au tribunal.

Crespo Miguel, M. (2023). *El análisis de autoría en Lingüística forense: Historia, concepción y revisión metodológica. Aplicación de la razón de verosimilitud a textos cortos en español.* Revista Signos. Estudios de Lingüística.

Daubert v. Merrell (U.S. Supreme Court, 1993).

Frye v. United States (Supreme Court of the District of Columbia, 1923).

Hassan A., Barber S.J. (2021). *The effects of repetition frequency on the illusory truth effect.* Cognitive Research: Principles and Implications.

Heylighen, F., & Dewaele, J.-M. (1999). *Formality of language: Definition, measurement and behavioral determinants*. Center Leo Apostel, Vrije Universiteit Brussel.

Imarisio L. (2025). *Modelli Bayesiani per lo studio dell'interazione tra variabili ambientali e l'attività di Mustela Nivalis*. Politecnicno di Torino.

Istituto della Enciclopedia Italiana. (n.d.). Stile nominale. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/stile-nominale_(Enciclopedia-dell'Italiano)/.

Istituto della Enciclopedia Italiana. (n.d.). Lessico. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/lessico_(Enciclopedia-dell'Italiano)/.

Istituto della Enciclopedia Italiana. (n.d.). Enumerazione. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/enumerazione_(Enciclopedia-dell'Italiano)/.

Istituto della Enciclopedia Italiana. (n.d.). Nominalizzazione. In La grammatica italiana. https://www.treccani.it/enciclopedia/nominalizzazione_(La-grammatica-italiana)/.

Istituto della Enciclopedia Italiana. (n.d.). Aggettivi qualificativi. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/aggettivi-qualificativi_(Enciclopedia-dell'Italiano)/.

Istituto della Enciclopedia Italiana. (n.d.). Verbi. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/verbi_(Enciclopedia-dell'Italiano)/.

Istituto della Enciclopedia Italiana. (n.d.). Registro. In Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/registro_(Enciclopedia-dell'Italiano)/.

Jamil S. M., Saprijal. (2025). *From traditional society to industrial society (An anthropological analysis of the manifesto Ted Kaczynski)*. Journal Sulôh Nanggroe.

Javed, Jehan, Ahmad. (2025). *Syntactic Ambiguity in English: Cognitive Processing and Interpretation*. Advance Social Science Archive Journal.

Kaczynski, T. J. (1995, September 19). *Industrial Society and Its Future*. The Washington Post.

Kawasaki Y. (2023). *Revisiting Authorship Attribution of Tirant lo Blanc Using Parts of Speech n-grams*. ACL anthropology.

Kleiman v. Wright (Florida District Court 2020).

Kristen L., Mejía-Ramos J. P. (2019). *Linguistic Conventions of Mathematical Proof Writing at the Undergraduate Level: Mathematicians' and Students' Perspectives.* Journal for Research in Mathematics Education.

Leech G., Hundt M., Mair C., Smith N. (2009). *Change in Contemporary English: A Grammatical Study.* Cambridge University Press.

Lew, K., & Mejía-Ramos, J. P. (2019). *Linguistic conventions of mathematical proof writing at the undergraduate level: Mathematicians' and students' perspectives.* Journal for Research in Mathematics Education.

Liu, R.-Y. (2023). *Constituting institutional identity in political discourse: The use of the first-person plural pronoun in China's press conferences.* Discourse & Society.

Mariani A., Bambi F. (2013). *Lingua e diritto: scritto e parlato nelle professioni legali.* Università di Pisa (capitolo 2.2).

Marotta G., Tamponi L. (2021). *Argomenti di linguistica teorica e applicata.* Edizioni ETS.

Mastronardi, V., & Trojani, A. (2014). *Appunti di linguistica forense: Introduzione e reality monitoring*.

Melazzo R. (2021). *Linguistica forense: l'emblematico caso Unabomber.* Mimesis Edizioni.

Misini A., Kadriu A., Canhasi E. (2022). *A Survey on Authorship Analysis Tasks and Techniques.* SEEU Review.

Mitchell D. (2014). *Type-token models: a comparative study.* Journal of Quantitative Linguistics.

Nádraská, Z. (2020). *When "We believe that ...": The role of collective "we" in the dialogic negotiation in hard news discourse*. Prague Journal of English Studies.

Nini A. (2018). *Developing forensic authorship profiling*. Language and Law / Linguagem e Direito, Vol. 5(2), p. 38-58.

Oleson, J. C. (2023). *A requiem for the Unabomber.* Contemporary Justice Review.

Pearson Education. (n.d.). Clue. In Longman Dictionary of Contemporary English. https://www.ldoceonline.com/dictionary/clue.

Pio Principe M. (2021). *Elaborazione del linguaggio naturale con metodi probabilistici e reti neurali*. Università di Bologna.

Queralt, S. (2024). *Los retos de la lingüística forense en la era de la IA*. Lengua y Sociedad.

Recco, R. S. (2021). *Say it in plain English*. Università degli Studi di Enna "Kore".

Ruiz Urbón, C. (2025). *Aplicación de la lingüística forense a los estudios de atribución de autoría literaria: Principios teóricos, metodológicos y conclusivos*. Signa: Revista de la Asociación Española de Semiótica.

Saussure, F. (1916). *Cours de linguistique générale*. Payot.

Serianni L. (1989). *Grammatica italiana*. UTET.

Shim A. (2022). *Forensic Linguistics: Science or Fiction?*. UC Law Journal.

POS tags. (2018). Sketch Engine. https://www.sketchengine.eu/blog/pos-tags/.

Stamatatos E. (2008). *A survey of modern authorship attribution methods*. Journal of the American Society for Information Science and Technology.

Suchman, L. (2023). *The uncontroversial "thingness" of AI*. Big Data & Society.

Thomas M. (). *The T-unit as a measure of 'syntactic maturity': operationalizing linguistic complexity in early generative grammar.* Simplicité et complexité des langues dans l'histoire des théories linguistiques. p. 428.

Turchie, T. D. (1996, April 2). *Affidavit in support of application for search warrant*. United States District Court for the District of Montana.

Tkačuková T. (2010). *The power of questioning: A case study of courtroom discourse*. DISCOURSE and INTERACTION.

United States v. Hearst (U.S. District Court for the Northern District of California, 1976).

United States v. Kaczynski (U.S. District Court for the District of Montana, 1996).

United States. (1791). Constitution of the United States: Amendment IV.

Youvan, D. C. (2024, June 30). *A critical analysis of the Unabomber's manifesto: Industrial Society and Its Future*. ResearchGate.

Wiehl L. (2020). Hunting the Unabomber: The FBI, Ted Kaczynski, and the Capture of America's Most Notorious Domestic Terrorist. Thomas Nelson.

Zimmer B.(2006). *Forensic linguistics, the Unabomber, and the etymological fallacy*. Language Log.

DICHIARAZIONE DI CONSULTABILITÀ

*Il/La sottoscritto/a*_____Beatrice Oliva_____

*Matricola n°*__787___/___FI___

Corso di Laurea Triennale ___Mediazione Linguisitca, indirizzo in Criminologia e Cybersecurity

_____

Titolo della tesi: ____Il manifesto *Industrial Society and Its Future*: un'analisi linguistica-forense

_____

_____

Dichiara che la sua tesi:

✓ *è consultabile da subito*	☐ *potrà essere consultata a partire dal giorno ____/____/_____*	☐ *non è consultabile*

Firenze, _05_/_08_/_2026___ *Firma dello Studente* Beatrice Oliva