

ROMANE WERNER*

ANALISI QUANTITATIVA DI TRASCRIZIONI DI VERBALI DELLA POLIZIA

1. Introduzione e problematica

In Italia, gli anni Novanta sono stati caratterizzati da molteplici attentati compiuti da diversi movimenti politicizzati di estrema destra e di estrema sinistra, e da organizzazioni criminali, prima fra tutte Cosa Nostra. Gli attentati culminarono nel 1992 con le stragi di Capaci e di Via d'Amelio, che costarono la vita ai giudici Giovanni Falcone e Paolo Borsellino e alle loro scorte. Questi attentati spinsero le autorità ad aumentare i mezzi per combattere Cosa Nostra. Le testimonianze degli ex affiliati all'organizzazione criminale, divenuti collaboratori di giustizia, cominciarono a essere utilizzate in modo più efficiente per ottenere informazioni su Cosa Nostra e per sconfiggerla. Ma se da un lato, le confessioni dei collaboratori hanno permesso di affrontare uno dei principali pilastri su cui si basa l'organizzazione, ovvero l'omertà, rendendo l'organizzazione vulnerabile e smantellandola parzialmente; dall'altro, grazie alle false dichiarazioni dei collaboratori, molte persone estranee agli attentati sono state condannate all'ergastolo.

Varie ricerche hanno mostrato che queste dichiarazioni sono comunemente trascritte in verbali di polizia (KOMTER 2022), i quali mescolano il discorso orale con il linguaggio giuridico-burocratico. A questo riguardo, gli esperti considerano che la presenza sovrabbondante di termini tecnici e di espressioni arcaiche può portare a una mancanza di comprensione (FRASER 2003) e hanno denunciato la mancanza di una vera e propria norma per quanto riguarda la documentazione di queste testimonianze, il che porta all'uso di forme personali da parte delle persone che trascrivono e, pertanto, a una moltiplicazione delle forme di trascrizione (cfr. per esempio EERLAND/VAN CHARLDORP 2022). Inoltre, è stato mostrato che anche la forma data al verbale può avere grande importanza poiché il formato domanda-

* Université catholique de Louvain.

risposta sarebbe considerato come più affidabile da parte dei giudici (EERLAND/VAN CHARLDORP 2022).

In linea con quanto appena detto, abbiamo deciso di compiere un'analisi di linguistica computazionale forense che si concentrerà sul linguaggio delle due diverse forme di trascrizione dei verbali di interrogatori (i.e., le trascrizioni complete che raffigurano l'intera interazione e che includono ogni turno dei parlanti – dette “domanda-risposta” –, e le trascrizioni più brevi in cui l'interazione viene riassunta e letterarizzata, dette “narrative” o “monologo”) redatti dalla polizia italiana. Intendiamo fornire spunti di riflessione sull'uso del verbo nelle due forme di trascrizione per investigare se esse appartengono allo stesso genere linguistico e stabilire se la forma influenza la distribuzione di vari tratti linguistici. Per far ciò, ci siamo concentrati sull'identificazione e l'estrazione di caratteristiche linguistiche multilivello, che sono state estratte automaticamente da un *corpus* annotato fino alla sintassi.

Più in dettaglio, questo contributo ha come scopo di presentare i primi risultati di un lavoro quantitativo in corso volto a indagare le caratteristiche linguistiche di testi legali di collaboratori di giustizia redatti durante i processi delle stragi di Capaci e Via d'Amelio in Italia nel corso del quale condurremo un'analisi quantitativa di 24 caratteristiche linguistiche (lessicali, morfosintattiche e sintattiche) legate alle due forme di trascrizione utilizzate nell'ambito delle investigazioni in Italia.

L'articolo verrà strutturato nel modo seguente. Si comincerà spiegando l'ambito scientifico dell'articolo (ossia una combinazione tra scienze forensi, linguistica computazionale e linguistica del *corpus*). Si continuerà dando un breve stato dell'arte dell'analisi di trascrizione di interrogatori, e spiegando i dati e il metodo utilizzato, per finire con elementi conclusivi e direzioni future.

2. Perché combinare la linguistica forense (computazionale) con l'analisi del genere?

La linguistica forense ha attirato molta attenzione da quando Svartvik ha pubblicato *The Evans Statement: un caso per la linguistica forense* (SVARTVIK 1968), in cui ha dimostrato per la prima volta il reale potenziale della linguistica in contesti forensi (SOUSA-SILVA 2018: 119). La linguistica forense rappresenta quindi una sottodisciplina emergente delle scienze forensi, che «è un campo interdisciplinare della linguistica applicata/descrittiva che comprende lo studio e l'analisi della lingua nel contesto del crimine, delle procedure giudiziarie o delle controversie legali» (DANIELEWICZ-BETS 2012: 93).

In questo contesto, «l'uso della linguistica computazionale in contesti forensi è diventato indispensabile, tanto da far nascere il campo della linguistica forense computazionale» (SOUSA-SILVA 2018: 120). Quest'ultimo rappresenta un campo di studi relativamente giovane, che è una sottodisciplina della linguistica computa-

zionale che combina scienze forensi, informatica e linguistica. In altre parole, può essere considerato come «l'analisi automatizzata di tracce forensi» (GERADTS 2018: 179) in un contesto linguistico grazie all'uso di vari metodi di elaborazione del linguaggio naturale. Inoltre, la linguistica computazionale si avvale da molto tempo di *corpora* e

sebbene si faccia una distinzione tra linguistica dei *corpora* e linguistica computazionale, la prima può esistere solo come parte della seconda, non solo perché per essere disponibile in forma elettronica, un *corpus* deve essere sottoposto a un'elaborazione del linguaggio naturale, ma anche perché alcune delle procedure applicate ai *corpora* (come l'annotazione) richiedono procedure di elaborazione sofisticate e inoltre perché i *corpora* dovrebbero essere idealmente adattati per essere utilizzati nei sistemi NLP (SOUSA-SILVA 2018: 123).

Per quest'analisi, ci concentreremo sull'analisi del linguaggio come mezzo di comunicazione tra le autorità di polizia e i sospetti/testimoni e sul conflitto tra la lingua dei testi scritti e la lingua parlata del processo legale.

Quest'analisi combina anche la linguistica forense con l'analisi del genere. Studi recenti hanno mostrato che «i generi sono intesi come forme di conoscenza culturale che inquadrano concettualmente e mediano il modo in cui comprendiamo e agiamo in varie situazioni. Questa visione riconosce i generi come tipi di testi e di azioni sociali che organizzano e generano tipi di testi e azioni sociali, in relazione complessa e dinamica gli uni con gli altri» (BAWARSHI/REIFF 2010: 4). Questo campo di ricerca «invita a riconoscere come le caratteristiche formali, piuttosto che essere arbitrarie, sono collegate a scopi sociali e modi di essere e di conoscere in relazione a questi scopi» (*ibid.*). I generi sono quindi stati definiti come «produzioni linguistiche orali e scritte che, in una determinata cultura, possiedono caratteristiche comuni di tipo testuale, semantico, grammaticale stabili nel tempo» (CHARTRAND *et al.* 2015: 5).

3. Stato dell'arte

In modo generale, la trascrizione è stata definita come «il processo di cambiamento di una modalità linguistica in un'altra, cioè dal parlare allo scrivere» (LINDSAY/O'CONNELL 1994: 102) con l'agente del cambiamento rappresentato dal trascrittore, il che implicherebbe quindi che «la qualità di una trascrizione non può essere giudicata a prescindere dagli scopi e dalle capacità del trascrittore» (*ibid.*). In altre parole, la trascrizione riguarda quindi un cambiamento di formato, dal parlato al discorso scritto, la cui forma e il cui contenuto dipenderanno principalmente dalla conoscenza e dallo scopo del trascrittore. L'originaria dialogicità del discorso viene così sostituita dall'autorità monologica del trascrittore, che determina cosa è stato detto e come

rappresentarlo (RICHARDSON *et al.* 2022). Inoltre, su un piano forense e legale, la trascrizione può essere considerata sia come un processo importante nella fase investigativa di qualsiasi crimine, sia come un compito cruciale svolto dagli agenti di polizia (LINDSAY/O'CONNELL 1994). Secondo VAN POUCKE (2017), i rapporti di polizia sono da considerare come dichiarazioni ufficiali e cronologicamente scritte dagli agenti, che hanno lo scopo di fornire informazioni sia per quanto riguarda il ritrovamento di tracce e prove sia per quanto riguarda il loro autore. Pertanto, «quando gli agenti di polizia interrogano formalmente sospetti, testimoni e vittime, la loro condotta rientra in un quadro giuridico volto a garantire due scopi: l'utilità forense nell'individuazione di un reato e l'accettabilità giuridica delle prove prodotte in tribunale» (STOKOE *et al.* 2020: 21). Inoltre, «le interviste investigative della polizia sono una parte vitale del processo di giustizia penale: sono un elemento essenziale del processo di raccolta delle prove investigative, e in molte giurisdizioni andranno anche a costituire un importante elemento di prova in tribunale» (HAWORTH 2018: 428).

Tuttavia, molti esperti hanno evidenziato il fatto che le trascrizioni non rappresentano dati fedeli degli interrogatori (LINDSAY/O'CONNELL 1994: 101), a causa delle carenze metodologiche dei sistemi di trascrizione utilizzati e, pertanto, «tutte le trascrizioni risultano essere selettive e, in qualche misura, distorsive» (*ibid.*), in quanto nessuna trascrizione di un'interazione potrà mai essere completamente esatta fornendo solo una rappresentazione del discorso parlato e mai una copia esatta di quest'ultimo, e ciò comporta inevitabilmente una perdita di dati (*ibid.*). Dunque da un lato, le trascrizioni rappresentano la prima tappa della procedura penale e una delle basi del successo o meno della fase investigativa; dall'altro, si possono avere dubbi sulla loro utilità forense a causa dei problemi legati all'estrazione e all'accettabilità giuridica delle tracce forensi (*ibid.*).

Come è mostrato nella letteratura, i rapporti di polizia possono assumere due forme principali in Italia, ossia uno stile monologo e uno stile domanda-risposta, che definiremo in seguito. Dal punto di vista grammaticale, il monologo è di solito scritto dalla prospettiva dell'indagato, cioè utilizzando la prima persona singolare e il discorso diretto (EERLAND/VAN CHARLDORP 2022: 2). Le domande poste e le osservazioni fatte dall'interrogante non compaiono quindi nel verbale (*ibid.*). La domanda non è perciò documentata, mentre la risposta è fornita attraverso una narrazione in prima persona singolare. In altre parole, «l'interrogatorio comporta la produzione di una dichiarazione monologica redatta dall'intervistatore ma scritta come se fosse una narrazione in prima persona dal testimone» (HAWORTH 2018: 445). Va ricordato che nel caso dell'Italia, le trascrizioni delle interviste assumono la forma di resoconti *A.D.R.* (*A Domanda Risponde*) quando non sono state fornite registrazioni audio/video. Sembra quindi che il rapporto sia stato scritto utilizzando uno stile domanda-risposta, in quanto viene fornito l'acronimo *A.D.R.*, mentre in realtà non è stata scritta alcuna domanda e la risposta si trova sotto forma di stile narrativo. Sembra quindi che il discorso parlato sia stato trasformato in un discorso scritto attraverso un processo di *letterarizzazione* del discorso stesso (cfr. Fig. 1).

A D.R.: Ho iniziato a far parte di cosa nostra dal 1990, nel senso che in quell'anno sono stato ufficialmente "combinato", ma sin dall'infanzia ho svolto attività illecite per conto di cosa nostra, essendo io imparentato con le famiglie Galatolo, Madonna e Fontana, tutti facenti parte della mafia.

Fig. 1 - Esempio di una trascrizione con lo stile monologo

Per quanto riguarda lo stile domanda-risposta (Q&A), le domande degli agenti di polizia e le risposte dell'indagato sono scritte come tali utilizzando il discorso diretto e, talvolta, anche segnali prosodici o il discorso soprasegmentale (EERLAND/VAN CHARLDORP 2022: 2). In Italia, è stato notato che quando sono state condotte registrazioni audio/video, si tende a redigere il rapporto in forma di *A.D.R.* In tal caso, si annotano sia le domande che le risposte nel modo nel quale sono state dette spontaneamente (cfr. Fig. 2).

1. Avv. Dif.: Sì, solo due domande. Agente, il collega che inseguiva la persona che avete bloccato, era questo signore?
2. Teste F.: Guardi, mi sembra di no, mi sembra che era l'ispettore che era di servizio; perché io l'imputato l'ho visto dopo, <al momento del...>
3. Avv. Dif.: Senta, nei> momenti immediatamente precedenti alla scena che voi osservate [sic], vi trovavate, immagino, sempre negli stessi locali dell'aeroporto?
4. Teste F.: Sì, fuori però dal <posto di polizia.>
5. Avv. Dif.: Ecco, lei> in precedenza aveva notato qualcosa di anomalo, in relazione a quel cittadino che scappava?
6. Teste F.: No, no, almeno, noi, appena siamo entrati... perché, al momento che noi entravamo, già dalla porta abbiamo visto questo <che...>
7. Avv. Dif.: Grazie>, nessun'altra domanda.

Fig. 2 - Esempio di una trascrizione con lo stile domanda-risposta

Tuttavia, in Italia, la riforma Cartabia (decreto legislativo n. 150/2022) ha recentemente inciso sulle forme di documentazione del processo penale modificandone l'articolo 134 del codice di procedura penale (i.e., la norma che disciplina, appunto, le modalità di documentazione). In base a questa legge, l'atto legale deve comprendere un verbale e, in alcuni casi, l'uso di mezzi fonografici e audiovisivi. La relazione può inoltre essere redatta in forma integrale o sintetica. La forma integrale deve riportare tutte le dichiarazioni di tutte le parti, mentre la forma sintetica riporta solo le parti essenziali della deposizione. Nel caso della forma integrale, la trascrizione rappresenta in modo fedele e preciso l'interrogatorio nella sua interezza, non omettendo alcun elemento. La forma sintetica, invece, riproduce fedelmente ma sommariamente le domande e le risposte in modo essenziale. Viene anche modificato il fatto che quando il verbale è redatto in forma sintetica, l'interrogatorio deve essere documentato anche con mezzi fonografici e audiovisivi.

Nei prossimi paragrafi, verranno spiegati i dati utilizzati, e il metodo di annotazione e di analisi.

4. Dati e metodo

Per rispondere alla nostra domanda di ricerca, abbiamo deciso di selezionare un *corpus* di testi che fossero sia estratti da scenari reali sia raccolti in condizioni standard. I dati utilizzati provengono da un enorme archivio della Commissione Parlamentare Antimafia. Come menzionato sopra, quest'analisi si concentra sulle trascrizioni di due forme di rapporti di polizia, ossia lo stile monologo e lo stile domanda-risposta. Abbiamo raccolto il contenuto di 25 trascrizioni per lo stile monologo e di 25 trascrizioni per lo stile domanda-risposta. Questi testi sono stati pubblicati dal 1984 al 2018. Il nostro *corpus* contiene 50 testi e ammonta a 46.810 *token* per le trascrizioni dello stile monologo e 212.257 *token* per le trascrizioni dello stile domanda-risposta, per un totale di 259.067 *token*.

Ci siamo concentrati sull'identificazione e l'estrazione di caratteristiche linguistiche multilivello che sono state estratte da un *corpus* annotato fino alla sintassi a dipendenze per poter stabilire il profilo linguistico del *corpus*.

Il nostro *corpus* è stato analizzato automaticamente grazie a una serie di strumenti di trattamento automatico del linguaggio sviluppati congiuntamente dall'Istituto di Linguistica Computazionale Antonio Zampolli di Pisa e dall'Università di Pisa, che si sono dimostrati tra i più efficienti nel trattare testi giuridici (VENTURI 2012). Tutti i testi sono quindi stati etichettati morfo-sintatticamente dal Part-of-Speech tagger descritto in DELL'ORLETTA 2009 e sintatticamente dal parser DeSR, ovvero il parser di dipendenza sviluppato e descritto in ATTARDI 2006. Gli strumenti utilizzati si sono dimostrati in grado di «rendere evidenti le informazioni linguistiche implicite contenute nei testi annotandoli a livelli di analisi sempre più complessi» (VENTURI 2012: 141). Tutti i testi sono quindi stati suddivisi in varie frasi e ogni frase è stata poi segmentata in unità ortografiche, ovvero *token* e lemmi, con un *token* per riga in documenti .csv. inoltre, a ogni *token* sono stati aggiunti i livelli morfologici (POS, coarse_pos e fine_pos) e le relazioni di dipendenza sintattica tra i *token* (syntactic_index, syntactic_information) (VENTURI 2012) (cfr. Tab. 1 per un esempio dell'annotazione).

index	token	lemma	pos	coarse_pos	fine_pos	syntactic_inc	syntactic_info
1	Non	non	B	BN	_		4 neg
2	ho	avere	V	VA	num=s per=		4 aux
3	mai	mai	B	B	_		4 mod_temp
4	riferito	riferire	V	V	num=s mod		0 ROOT
5	le	il	R	RD	num=p gen=		6 det
6	cose	cosa	S	S	num=p gen=		4 obj
7	che	che	P	PR	num=n gen=		10 subj
8	mi	mi	P	PC	num=s per=		10 comp_ind

Tab. 1 - Esempio di *tokenization*, *POS-tagging* e *tagging* morfologico e sintattico

Abbiamo analizzato le caratteristiche linguistiche che sono state tipicamente colte nell'ambito dell'identificazione di generi (cfr. BRUNATO/DELL'ORLETTA 2017) – focalizzandoci sul verbo –, ossia 1) le caratteristiche generali del testo (i.e., lunghezza delle frasi e lunghezza delle parole); 2) le caratteristiche lessicali (es., *Type/Token Ratio* (TTR)); 3) le caratteristiche morfosintattiche (es., distribuzione delle parti del discorso, densità lessicale, distribuzione del modo, del tempo e della persona dei verbi); 4) le caratteristiche sintattiche (es., distribuzione dei tipi di dipendenza, caratteristiche di profondità degli alberi, caratteristiche dei predicati verbali, caratteristiche di subordinazione). Ci siamo anche concentrati sull'analisi di caratteristiche che sono state considerate come tipiche del linguaggio legale (cfr. DAHLMAN 2006; BUSSO 2021; VISCONTI 2016; ONDELLI 2014; DAINO 2016; PIANESI 2008), ossia 1) caratteristiche lessicali (i.e., l'uso di abbreviazioni, di termini tecnici, di aggettivi in *-(t)ivo* e *-(t)orio*, nomi deaggettivali, frasi ridotte, dialogismo) e 2) caratteristiche morfosintattiche (i.e., l'uso di verbi modali e del passivo).

Siccome si tratta di un'analisi in corso, abbiamo analizzato, per compiere una prima analisi di tratti linguistici compresi nelle due forme di trascrizione, 24 tratti linguistici, ossia (1) caratteristiche generali del testo: lunghezza delle frasi, lunghezza delle parole; (2) caratteristiche lessicali: densità lessicale, *Type/Token Ratio* dei lemmi, *Type/Token Ratio* delle forme, abbreviazioni, termini tecnici (si veda CASADEI 2014 per una risorsa dedicata ai termini tecnici del linguaggio legale in Italia), aggettivi in *-(t)ivo* e *-(t)orio*, nomi deaggettivali, dialogismo, nomi provenienti da participi passati, frasi ridotte; (3) caratteristiche morfosintattiche: distribuzione di verbi al futuro, distribuzione di verbi all'imperfetto, distribuzione di verbi al participio, distribuzione di verbi al presente, l'anteposizione dell'aggettivo al nome, l'anteposizione dell'avverbio al verbo, l'anteposizione del participio rispetto al nome, l'uso del passivo, l'uso di verbi modali; (4) caratteristiche sintattiche: proporzione di proposizioni principali, proporzione di frasi subordinate prima del verbo principale, proporzione di frasi subordinate dopo il verbo principale. Ogni tratto linguistico è anche stato statisticamente analizzato in modo da stabilire la verificabilità linguistica del suo risultato¹.

5. Analisi e risultati

Verranno ora esposti i primi risultati della nostra analisi, suddivisi secondo ogni tratto linguistico.

¹ Il test statistico utilizzato per condurre quest'analisi statistica è stato il T Test.

5.1. Caratteristiche generali del testo

Per quanto riguarda il numero di parole nelle frasi, osserviamo un numero più importante di parole nella forma domanda-risposta con una differenza, comunque, non significativa (32.81 vs. 30.93; $p = 0.4543$). Questo risultato potrebbe spiegarsi col fatto che il *POS tagger* non delimita le frasi in modo perfetto, ossia elementi che sarebbero considerati come facenti parte di una nuova frase (es., la frase «viene sentito nella veste di imputato in procedimento connesso e proprio in tale veste ha la facoltà di avvalersi... appunto ha la facoltà di non rispondere; la dobbiamo inoltre avvisare che tutto quanto lei avrà intenzione di riferire...» costituisce una sola frase, mentre dovrebbe essere rappresentata da tre frasi, ossia «(1) viene sentito nella veste di imputato in procedimento connesso e proprio in tale veste ha la facoltà di avvalersi... (2) appunto ha la facoltà di non rispondere; (3) la dobbiamo inoltre avvisare che tutto quanto lei avrà intenzione di riferire...») sono stati aggiunti come elementi linguistici della frase principale.

Osserviamo anche un numero più importante di lettere nelle parole nella forma monologo con una differenza significativa (4.74 vs. 4.42; $p = 0.0004$), come è già mostrato nella letteratura scientifica (si veda per esempio ROVERE 2005), nella quale viene indicato che il discorso legale comprende parole più lunghe che in altri generi comunicativi.

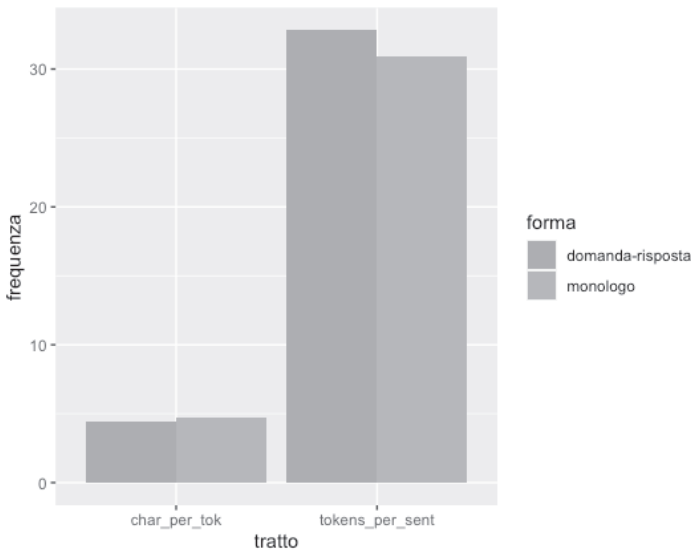


Grafico 1 - Distribuzione del numero di parole nelle frasi e del numero di lettere in ogni parola

5.2. Caratteristiche lessicali

Si osserva una densità lessicale più importante nella forma monologo con una differenza significativa (0.477 vs. 0.457; $p = 0.00005$). Nel caso dei *Type/Token Ratio*, notiamo una distribuzione più importante dei lemmi nella forma monologo con una differenza non significativa (0.625 vs. 0.608; $p = 0.5329$), mentre per quanto riguarda il *Type/Token Ratio* delle forme, si nota un numero più importante nella forma monologo con una differenza significativa (0.720 vs. 0.672; $p = 0.0007$) (si veda il Grafico 2).

Inoltre, si osservano più abbreviazioni nella forma monologo, con una differenza però non significativa (0.0024 vs. 0.0023; $p = 0.2853$) (cfr. esempi 1 e 2). Si osservano anche più termini tecnici (0.00237 vs. 0.0005; $p = 0.001$), frasi relative participiali ridotte (0.001 vs. 0.0004; $p = 0.0005$), dialogismo (0.004 vs. 0.002; $p = 0.01$), nomi deaggettivali (0.0002 vs. 0.0001; $p = 0.1$) e nomi provenienti da participi (0.0003 vs. 0.00007; $p = 0.5$) nella forma monologo, con differenze che sono significative (cfr. esempi da 3 a 8). Al contrario, emerge un numero più importante di aggettivi in *-(t)ivo* e *-(t)orio* nella forma domanda-risposta (0.0003 vs. 0.0004; $p = 0.06$) con una differenza però non significativa.

- (1) **Abbreviazione (forma monologo):** Tutte le dichiarazioni che ho progressivamente reso nei vari interrogatori di cui mi chiedono le *SSLL*² mi sono state suggerite progressivamente dai funzionari di polizia (Andriotta_17-07-2009);
- (2) **Abbreviazione (forma domanda-risposta):** assumerà ufficio di testimone, salvo incompatibilità previste dall'*art.* 197 (Scarantino_28-09-2009);
- (3) **Termini tecnici (forma monologo):** io ho appena dichiarato di non essermi mai occupato istituzionalmente del *predetto* (Narracci_27-10-2010);
- (4) **Termini tecnici (forma domanda-risposta):** Sì, intendo rispondere precisando *altresì* che trovandomi qui per altra udienza innanzi alla Corte d'Appello, ho ritenuto opportuno approfittare per venire nei vostri uffici (Ciancimino_29-10-2009);
- (5) **Frase ridotte (forma monologo):** sono venuto a conoscenza anche di reati commessi da altri appartenenti alla mia stessa organizzazione e di vicende *coinvolgenti* rappresentanti dello Stato (Scarantino_22-09-1194);
- (6) **Frase ridotte (forma domanda-risposta):** Ed è questo quello che a noi interessa diciamo in modo *preveniente* ma non è l'unico che ci interessa (Ciancimino_16-07-2009);
- (7) **Dialogismo (forma monologo):** ero stato nominato *capo-mandamento* (Mutolo_17-07-1992);
- (8) **Dialogismo (forma domanda-risposta):** io avevo chiesto.. al Comune, al Prefetto, al Sindaco, un qualunque posto di lavoro, pure *lavare scale* (Candura_10-03-2009).

² *SSLL* = *le Signorie Loro*, forma di cortesia utilizzata dagli uffici pubblici e legali per rivolgersi ai loro interlocutori con forme indirette. Rimanda pertanto in modo ovvio al linguaggio burocratico.

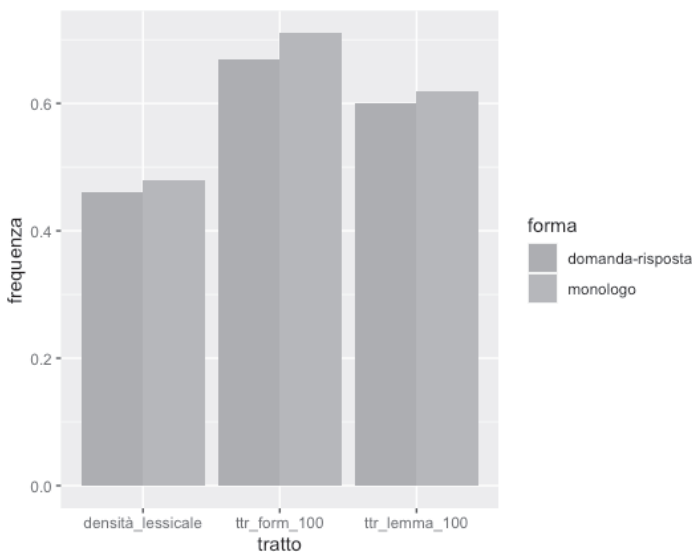


Grafico 2 - Distribuzione della densità lessicale e del *Type/Token Ratio* dei lemmi e delle forme

È stato mostrato nella letteratura che il linguaggio legale viene solitamente associato a un *Type/Token Ratio* e a una densità lessicale più alti rispetto ad altri generi (cfr. ORTMANN/DIPPER 2019; DAHLMAN 2006). Nel nostro caso, per quanto riguarda le caratteristiche lessicali, osserviamo che i tratti linguistici che sono stati evidenziati come più tipici del linguaggio legale sono associati alla forma monologo, la quale sembrerebbe pertanto essere più vicina al linguaggio legale rispetto alla forma domanda-risposta.

5.3. Caratteristiche morfosintattiche

Si osservano un numero più importante di verbi al futuro nella forma domanda-risposta con una differenza significativa (1.35 vs. 1.28; $p = 0.6477$), un numero più importante di verbi all'imperfetto nella forma narrativa con una differenza non significativa (18.4 vs. 15.2; $p = 0.07$), un numero più importante di verbi al participio passato nella forma monologo con una differenza significativa (58.2 vs. 43.6; $p = 0.0002$), un numero più importante di verbi al presente nella forma domanda-risposta con una differenza significativa (39.8 vs. 22.2; $p = 0.0002$) (cfr. Grafico 3). Per quanto riguarda l'uso dei tempi verbali, la forma monologo sembra contenere più verbi al participio passato e all'imperfetto, il che sembra coincidere con lo scopo comunicativo di quel tipo di forma (i.e., narrazione di fatti al passato), mentre la forma domanda-risposta contiene più verbi al futuro e al presente, il che coincide più con una forma orale di comunicazione con un contesto collegato alla comunicazione spontanea.

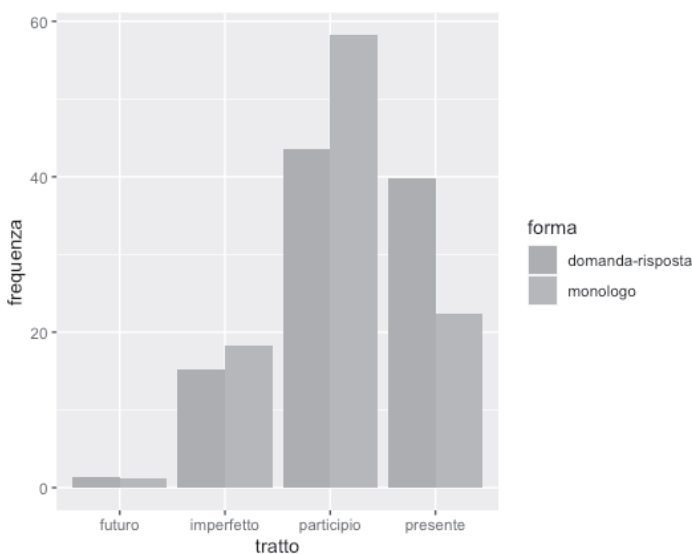


Grafico 3 - Distribuzione del numero di verbi al futuro, all'imperfetto, al participio passato e al presente

Si osservano anche una distribuzione più importante di anteposizioni dell'aggettivo al nome nella forma monologo, con una differenza significativa (0.007 vs. 0.005; $p = 0.04$), una distribuzione più importante di anteposizioni del participio al nome nel caso della forma monologo con una differenza non significativa (0.0004 vs. 0.0001; $p = 0.2729$), mentre si osserva una distribuzione più importante di anteposizioni dell'avverbio al verbo nel caso della forma domanda-risposta, con una differenza significativa (0.0135 vs. 0.0185; $p = 0.00004$). Si osservano un numero più importante di verbi modali nella forma domanda-risposta con una differenza significativa (0.0005 vs. 0.008; $p = 0.00001$) ma anche dell'uso del passivo nella forma domanda-risposta, con una differenza però non significativa (0.0023 vs. 0.0024; $p = 0.4546$) (cfr. esempi 9-18).

- (9) **Anteposizione dell'aggettivo al nome (forma monologo):** Ho deciso di collaborare, dopo essermi avvalso della facoltà di non rispondere nel corso di due *precedenti confronti* (Scarantino_28-09-2009);
- (10) **Anteposizione dell'aggettivo al nome (forma domanda-risposta):** tutti i *precedenti incontri* con il Colonnello... (Ciancimino_15-05-2008);
- (11) **Anteposizione del participio al nome (forma monologo):** e commentavamo l'*avvenuta* esecuzione (Russo_25-03-1988);
- (12) **Anteposizione del participio al nome (forma domanda-risposta):** ma di volere provare ad entrare al più presto nel materiale possesso dei *predetti documenti* per consegnarli immediatamente dopo all'Ufficio (Ciancimino_12-12-2008);

- (13) **Anteposizione dell'avverbio al verbo (forma monologo):** i cui genitori in quel tempo ed anche attualmente abitano ove abitavano i miei genitori (Tranchina_03-05-2011);
- (14) **Anteposizione dell'avverbio al verbo (forma domanda-risposta):** avevo *semplicemente detto* che per motivi familiari avrei trattenuto (Ciancimino_12-06-2009);
- (15) **Verbi modali (forma monologo):** Ribadisco ancora che a Ficuzza, dove fui portato, mi furono eseguite delle fotografie, dalle quali ben *dovrebbe evincersi* il mio stato (Mule_05-12-1984);
- (16) **Verbi modali (forma domanda-risposta):** sì perché *volevo lanciare* un magistrato (Candura_10-03-2009);
- (17) **Passivo (forma monologo):** Il Riccobono mi spiegò che, appunto, l'omicidio *era stato deciso* dalla Commissione per i motivi che ho già ricordati (Mutolo_27-11-1992);
- (18) **Passivo (forma domanda-risposta):** che io *ero stato avvicinato* dai dei mafiosi per farmi ritrattare (Andriotta_17-07-2009).

5.4. Caratteristiche sintattiche

In questa parte, osserviamo un numero più importante di proposizioni principali nella forma monologo con una differenza significativa (30.03 vs. 24.59; $p = 0.002$), un numero più importante di frasi subordinate dopo la frase principale nella forma monologo con una differenza significativa (88.74 vs. 86.01; $p = 0.01$), e un numero più importante di frasi subordinate prima della frase principale nella forma domanda-risposta con una differenza significativa (13.98 vs. 11.25; $p = 0.01$) (cfr. Grafico 4). questi risultati sono anche stati esposti in studi precedenti (cfr. per esempio SHARNDAMA 2014).

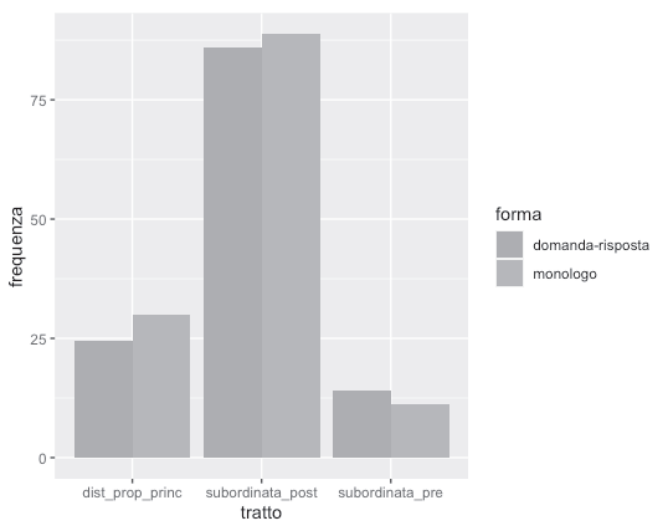


Grafico 4 - Distribuzione delle proposizioni principali, delle subordinate prima della frase principale e delle subordinate dopo la frase principale

6. Conclusione

Per concludere, abbiamo osservato che nella forma monologo le parole sembrano essere più lunghe e perciò più complesse. Anche la densità lessicale sembra essere più importante, il che ci mostra una variazione lessicale più importante e pertanto una complessità linguistica più importante. Un'altra osservazione rimanda al fatto che la forma monologo comprende più verbi al passato, mentre si usano più il futuro e il presente nella forma domanda-risposta, il che sembra coincide con lo scopo comunicativo delle due forme di trascrizione (i.e., narrazione di un fatto passato vs. domande-risposte nel corso di un intervento al presente).

Altri tratti linguistici collegati alla forma monologo (ad es. anteposizione dell'aggettivo al nome, anteposizione del participio al nome, termini tecnici, frasi relative participiali ridotte, dialogismo, nomi deaggettivali) sono tratti che sono stati considerati come facenti parte del linguaggio legale, il che enfatizza il fatto che la forma monologo sarebbe più collegata al linguaggio legale che la forma domanda-risposta, la quale potrebbe essere associata a un altro genere comunicativo, più legato alla resa orale. Queste due forme sembrano pertanto rimandare a due generi diversi e a scopi comunicativi diversi.

I dati ottenuti sono da considerare provvisori, in quanto quest'analisi rappresenta solo una prima fase di una ricerca più ampia sulle due forme di trascrizione degli interrogatori della polizia. Consideriamo comunque che siano significativi e contiamo nel proseguo della ricerca di analizzare più tratti linguistici, sia in modo quantitativo che qualitativo, e di condurre un'analisi degli scopi comunicativi che permettano di confermare la nostra ipotesi e di identificare i generi delle due forme di trascrizione.

Bibliografia

- ATTARDI 2006 = GIUSEPPE ATTARDI, *Experiments with a Multilanguage non-projective Dependency Parser*, in *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X '06)*. New York City, 2006, Stroudsburg, The Association for Computational Linguistics, 2006, pp. 166-170.
- BAWARSHI/REIFF 2010 = ANIS BAWARSHI / MARY REIFF, *Genre: An Introduction to History, Theory, Research, and Pedagogy*, Indiana, Parlor Press LLC, 2010.
- BUSSO 2021 = LUCIA BUSSO, *Lexicon and Grammar in Legal-Lay Language: A Quantitative Corpus Study on Italian*, in «Studi italiani di linguistica teorica e applicata», 2021, 51, pp. 5-32.
- BRUNATO/DELL'ORLETTA 2017 = DOMINIQUE BRUNATO / FELICE DELL'ORLETTA, *On the Order of Words in Italian: A Study on Genre vs. Complexity*, in SIMONETTA MONTEMAGNI / JOAKIM NIVRE (edited by), *Proceedings of the Fourth Interna-*

- tional Conference on Dependency Linguistics*, 2017, pp. 25-31, aclanthology.org/W17-6505.pdf.
- CASADEI 2014 = FEDERICA CASADEI, *Il lessico dell'italiano burocratico. Una ricognizione sul Grande Dizionario Italiano dell'Uso*, in FEDERICA CASADEI, ALESSANDRA SERRA / GRAZIA SOMMARIVA (a cura di), *Studi di linguistica, letteratura e filologia*, Viterbo, Sette Città, 2014, pp. 5-36.
- CHARTRAND *et al.* 2015 = SUZANNA CHARTRAND / JUDITH ÉMERY-BRUNEAU / KATHLEEN SÉNÉCHAL, *Caractéristiques de 50 genres pour développer les compétences langagières en français*, Québec, Bibliothèque et Archives nationales du Québec, 2015.
- DAINO 2016 = RICCARDO DAINO, *La lingua dei giuristi: analisi dell'italiano nei manuali di giurisprudenza*, Trieste, Università degli Studi di Trieste, 2016.
- DAHLMAN 2006 = COLONNAL DAHLMAN, *Specialità del linguaggio giuridico italiano*, Lunds, Lunds Universiteit, 2006.
- DANIELEWICZ-BETS 2012 = ANNA DANIELEWICZ-BETS, *The Role of Forensic Linguistics in Crime Investigations*, in ANDREW LITTLEJOHN / SANDHYA RAO MEHTA (edited by), *Language Studies: Stretching the Boundaries*, Cambridge, Cambridge Scholars Publishing, 2012, pp. 93-108.
- DELL'ORLETTA 2009 = FELICE DELL'ORLETTA, *Ensemble System for Part-of-Speech Tagging*, in *Proceedings of Evalita '09, Evaluation of NLP and Speech Tools for Italian*, Reggio Emilia, 2009, pp. 1-8.
- EERLAND/VAN CHARLDORP 2022 = ANITA EERLAND / TESSA VAN CHARLDORP, *The Influence of Police Reporting Styles on the Processing of Crime Related Information*, in «Frontiers in Communication», 2022, 7, www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2022.816461/pdf.
- FRASER 2003 = HELEN FRASER, *Issues in Transcription: Factors Affecting the Reliability of Transcripts as Evidence in Legal Cases*, in «Forensic Linguistics», 2003, 10(2), pp. 203-226.
- GERADTS 2018 = ZENO GERADTS, *Digital, Big Data and Computational Forensics*, in «Forensic Sciences Research», 2018, 3(3), pp. 179-182.
- HAWORTH 2018 = KATE HAWORTH, *Tapes, Transcripts and Trials: The Routine Contamination of Police Interview Evidence*, in «The International Journal of Evidence and Proof», 2018, 22(4), pp. 428-450.
- KOMTER 2022 = MARTHA KOMTER, *Institutional and Academic Transcripts of Police Interrogations*, in «Frontiers in Communication», 2022, 7, www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2022.797145/full.
- LINDSAY/O'CONNELL 1994 = JEAN LINDSAY / DANIEL O'CONNELL, *How do Transcribers Deal with Audio Recordings of Spoken Discourse?*, in «Journal of Psycholinguistic Research», 1994, 24(2), pp. 101-115.
- ONDELLI 2014 = STEFANO ONDELLI, *Ordine delle parole nell'italiano delle sentenze: alcune misurazioni su corpora elettronici*, in «Informatica e diritto», 2014, 23(1), pp. 13-39.

- ORTMANN/DIPPER 2019 = KATRIN ORMANN / STEFANIE DIPPER, *Variation between Different Discourse Types: Literate vs. Oral*, in *Proceedings of the Sixth Workshop on NLP for Similar Languages, Varieties and Dialects*, Michigan, Ann Arbor, 21019, pp. 64-79.
- PIANESE 2008 = GIOVANNI PIANESE, *Analisi linguistica comparativa di un corpus di testi del dominio giuridico. Sentenze penali italiane e francesi a confronto*, Napoli, Università degli Studi di Napoli Federico II, 2008.
- RICHARDSON *et al.* 2022 = EMMA RICHARDSON / KATE HAWORTH / FELICITY DREAMER, *For the Record: Questioning Transcription Processes in Legal Contexts*, in «Applied Linguistics», 2022, 43(4), pp. 677-697.
- ROVERE 2005 = GIOVANNI ROVERE, *Capitoli di linguistica giuridica. Ricerche su corpora elettronici*, Alessandria, Edizioni dell'Orso, 2005.
- SHARNDAMA *et al.* 2014 = EMMANUEL SHARNDAMA / YAKUBU SAMAILA / YUSUF ISHAYA TSOJON, *English for Academic Purposes: A Tool for Enhancing Students' Proficiency in English Language Skills*, in «International Journal of English Language Teaching», 2014, 1(2), pp. 14-20.
- SOUSA-SILVA 2014 = RUI SOUSA-SILVA, *Investigating Academic Plagiarism: A Forensic Linguistics Approach to Plagiarism Detection*, in «International Journal for Educational Integrity», 2014, 10(1), pp. 31-41.
- SOUSA-SILVA 2018 = RUI SOUSA-SILVA, *Computational Forensic Linguistics: An Overview of Computational Applications in Forensic Contexts*, in «Language and Law», V (2018), 2, pp. 118-143.
- STOKOE *et al.* 2020 = ELIZABETH STOKOE / CHARLES ANTAKI / EMMA RICHARDSON / SARA WILLOTT, *When Police Interview Victims of Sexual Assault: Comparing Written Guidance to Interactional Practice*, in MARIANNE MASON / FRANCES ROCK (edited by), *The Discourse of Police Interview*, Chicago, The University of Chicago Press, 2020, pp. 21-41.
- SVARTVIK 1968 = JAN SVARTVIK, *The Evans Statements: A Case for Forensic Linguistics*, Gothenburg, Acta Universitatis Gothoburgensis, 1968.
- VAN POUCKE 2017 = LIES VAN POUCKE, *Quel est le rôle de la rédaction du procès-verbal dans l'interprétation des interrogatoires de police*, Gent, Universiteit Gent, 2017.
- VENTURI 2012 = GIULIA VENTURI, *Investigating Legal Language Peculiarities Across Different Types of Italian Legal Texts: An NLP-Based Approach*, in «IAFL Porto 2012 Proceedings», 2012, pp. 138-156.
- VISCONTI 2016 = JACQUELINE VISCONTI, *La lingua del giudice*, in PAOLO CARETTI / ROBERTO ROMBOLI (a cura di), *La lingua dei giuristi. Atti dell'VIII Giornate internazionali di diritto costituzionale* (Pisa-Firenze, 24-25 settembre 2015), Pisa, Pisa University Press, 2016, pp. 32-41.

