

**A cura di**

---

**Nicola Grandi**

---

**L' ITALIANO SCRITTO  
DEGLI STUDENTI  
UNIVERSITARI**

**Quadro sociolinguistico, tendenze  
tipologiche, implicazioni didattiche**



**Materiali Linguistici  
Università di Pavia**

**FrancoAngeli** 

## 4. L'italiano scritto all'università: prime elaborazioni statistiche di dati linguistici e sociobiografici

di *Matteo Farnè, Nicola Grandi e Matteo Pascoli*<sup>1</sup>

### 1. Premessa: la domanda di ricerca

La domanda di ricerca che ha ispirato l'indagine condotta nell'ambito del progetto Univers-ITA, per la quale rinviamo all'Introduzione, è all'apparenza piuttosto semplice: gli studenti e le studentesse universitari scrivono bene o male? E la loro capacità di scrittura è davvero peggiorata con il passare del tempo?

Queste domande, che pure riscuotono, si è visto, un certo successo tra l'opinione pubblica e nella pubblicistica generalista, sono in realtà mal poste, dal momento che i concetti di “bene” e “male” si basano su percezioni del tutto soggettive che, quindi, variano da individuo a individuo e dipendono da criteri che è pressoché impossibile misurare scientificamente. Come si stabilisce, in effetti, quando un testo è scritto “bene” o “male”? Lettori diversi possono avere percezioni differenti a partire dal medesimo input. Si tratta, in fin dei conti, di una mera questione di “gusti”.

Il quesito ha però, come si è visto, di nuovo, nell'introduzione, una risonanza innegabilmente vasta e, dunque, non può essere accantonato o liquidato frettolosamente. È evidente, però, che l'unico modo per affrontare scientificamente il tema è quello di trasformare i concetti intuitivi di “bene” e “male” in parametri oggettivi e misurabili, che abbiano a che fare con la struttura e la configurazione dei testi a cui ci si riferisce e, indirettamente, con il processo di pianificazione che li ha prodotti, cioè con le scelte operate dagli e dalle scriventi. Anche in questo caso, ovviamente, non esistono strutture ineren-

1. Il lavoro, frutto di un progetto condiviso dagli autori, può essere così ripartito: Matteo Pascoli ha curato l'estrazione dei dati, Matteo Farnè ha curato le analisi e le elaborazioni statistiche e Nicola Grandi ha curato le generalizzazioni linguistiche. La redazione del testo è da attribuire a Nicola Grandi per §1, la sezione introduttiva di §3 e §5; a Matteo Farnè per §2 e §3.2; le sezioni §3.1 e §4 sono state redatte congiuntamente da Matteo Farnè e Nicola Grandi.

temente “buone” (o giuste) e strutture inerentemente “cattive” (o sbagliate). Esistono però termini di raffronto, determinati innanzitutto dalla situazione nella quale e per la quale viene prodotto un testo, rispetto ai quali è possibile in certa misura giudicare o, meglio, classificare il testo stesso. In sostanza ogni situazione comunicativa prefigura una sorta di “risultato atteso ideale”, cioè un tipo di testo perfettamente adeguato ad essa, rispetto al quale è possibile individuare eventuali scostamenti e devianze nei testi effettivamente prodotti. In questo modo, misurando cioè la vicinanza/la lontananza dal risultato ideale atteso, siamo effettivamente in grado di classificare i testi reali e di valutare, indirettamente, la competenza di chi li ha prodotti, soprattutto rispetto alla capacità di cogliere le specificità della situazione comunicativa e di operare le scelte adeguate. Per quanto concerne i testi confluiti nel corpus Univers-ITA<sup>2</sup>, la consegna trasmessa a studenti e studentesse, per la quale si rimanda al cap. 3, era quella di redigere un testo formale, scrivendo nel modo più corretto possibile e individuando in un proprio professore il destinatario ipotetico. In questo caso, il termine di confronto ideale è, dunque, un testo orientato verso il polo più alto dell’asse diafasico, prodotto in piena coerenza con quanto prevede la grammatica normativa dell’italiano. Ribadendo che, come si è già detto nell’introduzione e nel cap. 3, in questo volume si è scelto di designare con l’etichetta “neutra” di “annotazione” ogni fenomeno notevole rilevato nei testi raccolti, quindi ogni fenomeno deviante rispetto a quanto previsto dalla grammatica normativa, possiamo supporre che il numero complessivo di annotazioni rilevate in un testo sia in qualche modo direttamente proporzionale alla sua lontananza dal risultato atteso, cioè da un testo del tutto conforme a quanto prescritto dalla grammatica ufficiale della lingua (ma ciò, come vedremo più oltre, non qualifica automaticamente il testo come malformato).

Come si è detto nel cap. 3, per quantificare la distanza dal già citato risultato ideale atteso dei testi prodotti da studenti e studentesse che hanno partecipato alla rilevazione sono stati adottati due approcci, uno quantitativo e uno qualitativo.

## **2. La valutazione quantitativa dei testi del corpus Univers-ITA**

Sui 2.137 testi raccolti è stata innanzitutto effettuata una prima ricognizione puramente quantitativa, utilizzando il software READ-IT<sup>3</sup>, che calcola

2. Grandi *et al.* (2023); <https://doi.org/10.60760/unibo/univers-ita>. Cfr. anche cap. 3.

3. Cfr. Dell’Orletta *et al.* (2011). [www.italianlp.it/demo/read-it/](http://www.italianlp.it/demo/read-it/)

automaticamente un indice di complessità correlando parametri diversi. In particolare, abbiamo utilizzato il valore corrispondente a READ-IT BASE, che considera le caratteristiche tipicamente usate nelle misure tradizionali della leggibilità di un testo<sup>4</sup>: la lunghezza della frase (calcolata come numero medio di parole per frase), la lunghezza delle parole (calcolata come numero medio di caratteri per parola), il numero di frasi, il numero di parole, ecc. Come si è visto nel cap. 3 (in particolare nella tab. 1), una prima analisi quantitativa dei testi ha rivelato una loro sostanziale omogeneità, con una complessità medio-alta. Dopo questa ricognizione “grezza”, i testi sono stati classificati con granularità più fine, rilevando, sempre grazie al software READ-IT, altri parametri quantitativi utili per operare qualche generalizzazione più precisa. In particolare, risultano di un certo interesse i valori relativi al numero di frasi e di lessemi diversi presenti in ogni testo. Pur con i limiti di una ricognizione esclusivamente quantitativa (che misura, dunque, quante, ma non quali frasi e quanti, ma non quali lessemi compaiono nei testi<sup>5</sup>) e, quindi, con un’ approssimazione forse eccessiva, si può supporre che il numero di frasi sia un indicatore mediamente affidabile della complessità di un testo e che il numero dei lessemi diversi sia un indicatore mediamente affidabile della sua ricchezza lessicale.

Nel seguito di questa sezione descriveremo queste due caratteristiche dei testi prima in senso univariato, e poi rispetto ad alcune variabili sociobiografiche ricavate dal questionario descritto nel cap. 2. Lo strumento statistico principale che utilizzeremo è il test ANOVA (ANalysis Of VAriance), che consiste nel testare l’uguaglianza delle medie delle variabili quantitative oggetto di analisi rispetto alle diverse modalità di un carattere qualitativo (si rimanda a Cicchitelli, D’Urso e Minozzo, 2022 per maggiori dettagli)<sup>6</sup>.

## 2.1. Numero medio di frasi per testo

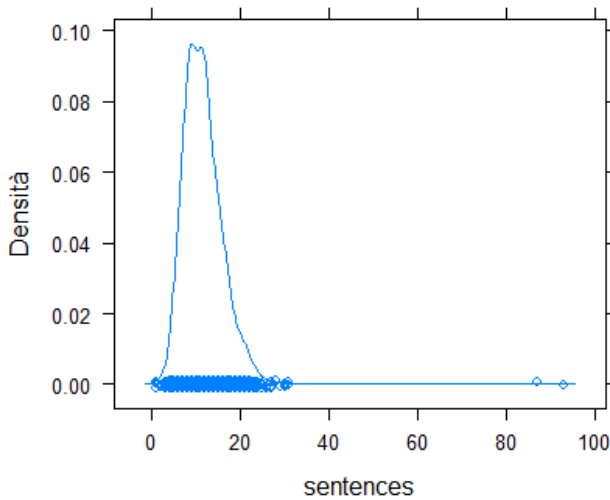
In fig. 1 osserviamo che il numero di frasi per testo ha una distribuzione approssimativamente simmetrica con una mediana di 11, una media di 11.6, un primo quartile di 8 e un terzo quartile di 14:

4. Per un riferimento generale, cfr. Collins-Thompson (2014).

5. Per altro, nel conto dei lessemi diversi il software non è in grado di riconoscere eventuali errori di battitura, quindi, ad esempio, una parola come *lbroy* verrebbe classificata come diversa da *libro* e utilizzata come indice di complessità in quanto non riconosciuta dal software stesso.

6. Nelle tabelle che seguono la significatività del p-value viene codificata come segue: \*\*\*\* tra 0 e 0.001, \*\*\* tra 0.001 e 0.01, \*\* tra 0.01 e 0.05, \* tra 0.05 e 0.1, ' ' tra 0.1 e 1. Tutte le elaborazioni statistiche presentate in questo capitolo sono basate sul dataset del progetto liberamente consultabile a questo indirizzo: <https://doi.org/10.6092/unibo/amsacta/8229>. Cfr. Grandi e Pascoli (2025).

Fig. 1 - Densità stimata del numero di frasi (sentences)



Per quanto concerne il numero delle frasi, le differenze sono dunque lievi, verosimilmente perché il limite imposto dalla finestra in cui il testo è stato digitato ha avuto, come conseguenza, una certa “compressione” del medesimo (detta più rozzamente: la gabbia imponeva un limite minimo e un limite massimo rispettivamente di 250 parole e 500 parole e, quindi, non ci si può aspettare che i testi differiscano troppo per numero di frasi; per questo, però, anche differenze lievi diventano rilevanti).

La tab. 1 riporta i dati di medie e mediane, sempre relative al numero di frasi per testo, suddivisi per aree disciplinari e aree geografiche.

Tab. 1 - Media e mediana delle frasi per testo per aree geografiche e disciplinari

		<i>Media frasi</i>	<i>Mediana frasi</i>
Area umanistica	<b>Totale</b>	<b>12,5</b>	<b>12,5</b>
	Nord	12,5	12,5
	Centro	11,0	11,0
	Sud e Isole	11,5	11,5
Area sanitaria	<b>Totale</b>	<b>11,7</b>	<b>11,0</b>
	Nord	12,8	11,0
	Centro	11,6	11,0
	Sud e Isole	10,6	10,0

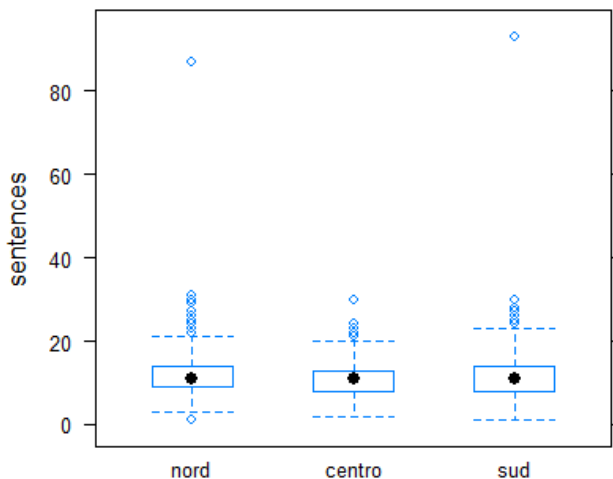
Tab. 1 - Segue

		Media frasi	Mediana frasi
Area scientifica	<b>Totale</b>	<b>11,4</b>	<b>11,0</b>
	Nord	11,8	11,0
	Centro	10,4	10,0
	Sud e Isole	12,0	11,0
Area sociale	<b>Totale</b>	<b>11,2</b>	<b>11,0</b>
	Nord	11,6	11,0
	Centro	11,2	11,0
	Sud e Isole	10,9	11,0

I testi prodotti da studenti e studentesse di area umanistica si caratterizzano per un numero di frasi per testo superiore a quello delle altre aree.

In fig. 2, si riporta il boxplot che correla il numero di frasi all'area geografica dell'ateneo. La media delle frasi è di 11.13 per il Centro, 11.47 per il Sud, 12.02 per il Nord. Il valore del Nord risulta significativamente superiore a quello del Centro, con un p-value dello 0.06%. Tra Centro e Sud invece non c'è differenza significativa:

Fig. 2 - Boxplot condizionato del numero di frasi (sentences) rispetto alla collocazione geografica dell'ateneo

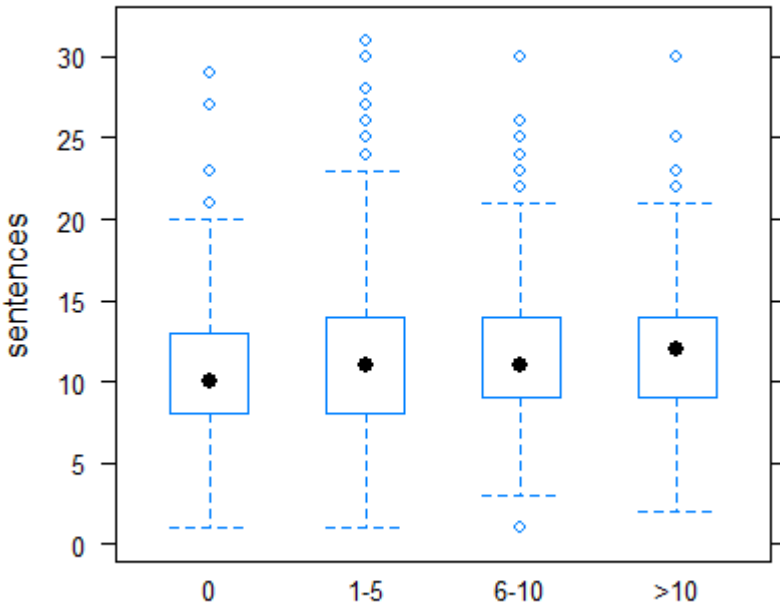


Tra le variabili sociobiografiche estratte dai questionari, risulta particolarmente esplicativa quella relativa alle abitudini di lettura (cfr. tab. 2 e fig. 3). La media condizionata delle frasi cresce infatti sistematicamente col numero di libri letti mediamente in un anno (ad esclusione di quelli necessari per la preparazione degli esami) da 10.65 per 0 libri a 11.97 per più di 10 libri:

Tab. 2 - Modello ANOVA per il numero di frasi (sentences) rispetto al numero medio di libri letti in un anno

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
numeroLibri0	10.6497	0.3694	28.826	< 2e-16 ***
numeroLibri1-5	0.8367	0.4021	2.081	0.03754 *
numeroLibri6-10	1.188	0.4184	2.841	0.00453 **
numeroLibri>10	1.3155	0.4484	2.934	0.00339 **

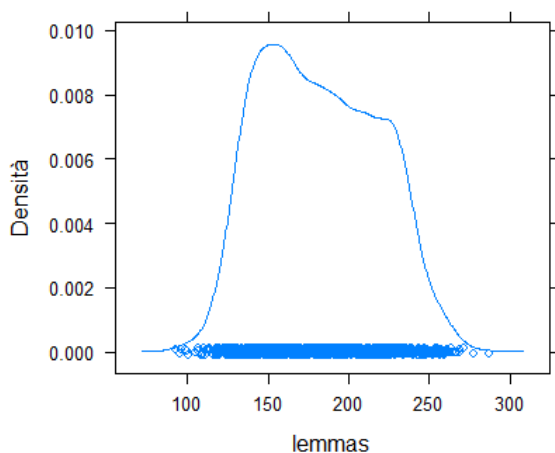
Fig. 3 - Boxplot condizionato del numero di frasi (sentences) rispetto al numero di libri



## 2.2. Numero di parole diverse per testo

Per quanto concerne il numero di parole diverse per testo, la fig. 4 rivela come esso abbia una distribuzione per nulla simmetrica: la mediana è di 179 parole, la media di 181.9, il primo quartile è 152 e il terzo quartile 211. La distribuzione appare però molto più dispersa attorno alla mediana rispetto a una variabile normale, e in modo appunto asimmetrico:

Fig. 4 - Densità stimata del numero di numero di parole (lemmas)



La tab. 3 mostra i dati di medie e mediane, sempre relativi al numero di parole diverse per testo, Suddivisi per aree disciplinari e aree geografiche:

Tab. 3 - Media e mediana delle parole diverse per testo per aree geografiche e disciplinari

		<i>Media lessemi</i>	<i>Mediana lessemi</i>
Area umanistica	<b>Totale</b>	<b>183,1</b>	<b>181,0</b>
	Nord	185,4	184,0
	Centro	183,5	184,0
	Sud e Isole	178,9	176,0
Area sanitaria	<b>Totale</b>	<b>179,4</b>	<b>179,5</b>
	Nord	184,5	186,0
	Centro	181,6	178,0
	Sud e Isole	170,0	166,0



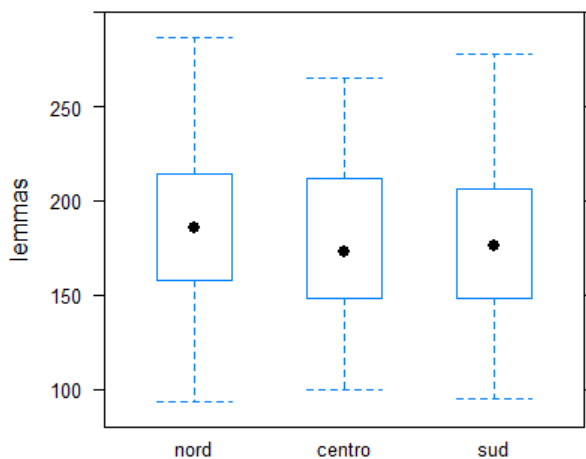
Tab. 1 - Segue

Area scientifica	<b>Totale</b>	<b>179,8</b>	<b>177,5</b>
	Nord	184,3	182,0
	Centro	172,6	163,0
	Sud e Isole	181,6	181,0
Area sociale	<b>Totale</b>	<b>184,4</b>	<b>182,0</b>
	Nord	190,4	193,0
	Centro	184,9	177,0
	Sud e Isole	176,8	166,5

In termini generali, considerando che i testi prodotti hanno in media circa 400 parole e che le parole funzionali (gli articoli, le preposizioni, ma anche gli ausiliari, ecc.) si ripetono inevitabilmente più volte in qualsiasi testo, usare in media quasi 182 parole diverse è un risultato piuttosto confortante, in quanto indica che la maggior parte delle parole non si ripete.

Per quanto riguarda le aree disciplinari, l'area sociale sopravanza quella umanistica. Per quanto concerne, invece, le aree geografiche la fig. 5 mostra che il numero mediano di parole diverse usate nei testi prodotti in atenei settentrionali è superiore rispetto a quello che caratterizza testi redatti in atenei del Centro e del Sud. Il test ANOVA mostra che il valore per il Nord (186.37) è significativamente superiore rispetto a quello del Sud, con un p-value di 0.35%, mentre tra Sud e Centro non si rilevano differenze significative:

Fig. 5 - Boxplot condizionato del numero di parole diverse (lemmas) rispetto alla collocazione geografica dell'ateneo

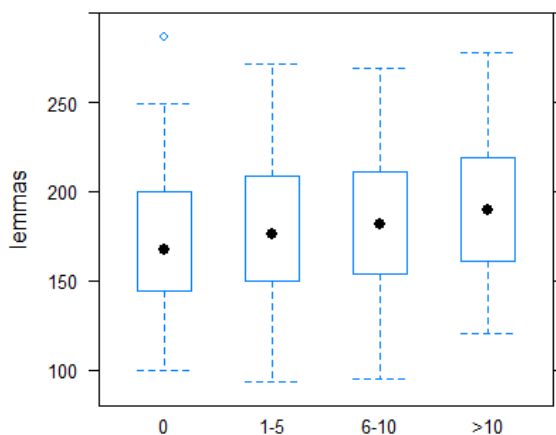


Anche in questo caso il parametro sociobiografico più esplicativo è il numero di libri letti all'anno: come si può osservare in fig. 6, il numero di parole medio cresce linearmente con esso. La tab. 4, che riporta il test ANOVA, mostra che rispetto al valore medio di parole per chi non legge libri (172.42), tutti i valori medi per le altre tre classi di *numeroLibri* sono significativi. Chi legge più 10 libri usa in media 17.48 parole in più, con un p-value inferiore allo 0.0001%:

Tab. 4 - Modello ANOVA per il numero di parole (lemmas) rispetto a numeroLibri

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
numeroLibri0	172.412	2.702	63.802	< 2e-16 ***
numeroLibri1-5	7.034	2.941	2.392	0.016845 *
numeroLibri6-10	11.271	3.061	3.683	0.000236 ***
numeroLibri>10	17.483	3.280	5.330	1.08e-07 ***

Fig. 6 - Boxplot condizionato del numero di parole (lemmas) rispetto al numero di libri letti in media all'anno



I dati, come è evidente, confermano una sensazione diffusa e cioè che la lettura abbia un effetto positivo sulle capacità di scrittura. Su questo aspetto, però, torneremo nella prossima sezione di questo contributo.

Per quanto concerne il numero di parole diverse per testo, emerge una correlazione interessante con il parametro relativo alla conoscenza di almeno

una lingua antica, che pare avere un effetto positivo sulla ricchezza del lessico: chi non conosce lingue antiche usa in media 179.19 parole diverse per testo; nei testi prodotti da chi ne conosce almeno una c'è un incremento medio di parole diverse pari a 5.9 (p-value=0.02%)<sup>7</sup>.

### 3. La valutazione qualitativa dei testi del corpus Univers-ITA

È tuttavia evidente come la natura di un testo sia troppo complessa per poter essere colta attraverso un'analisi esclusivamente automatica, visto che i fattori in gioco sono troppo articolati (si pensi, ad esempio, alla difficoltà già citata in n. 5, di rilevare automaticamente gli errori di ortografia o l'uso, proprio o improprio, della punteggiatura). Per poter tracciare un quadro davvero esaustivo delle peculiarità dei testi raccolti (allo scopo di poterli classificare rispetto al “risultato atteso” menzionato nel §1) e per poter meglio individuare quali tratti linguistici siano sistematicamente correlati ai parametri sociobiografici inclusi nel questionario (fondamentali per delineare lo sfondo sociale delle abilità di scrittura formale) si è reso necessario, come già detto nel cap. 3, abbinare all'analisi meramente quantitativa una valutazione qualitativa mediante un'accurata ricognizione manuale dei testi stessi. In questa fase sono stati annotati manualmente tutti i tratti che marcessero un qualunque scostamento dal “risultato atteso”, cioè che fossero, in qualunque modo, devianti rispetto al prototipo ideale di un testo scritto formale pienamente coerente con quanto prescritto dalla grammatica normativa dell'italiano. Come si è ricordato nell'introduzione e nel cap. 3 (e anche all'inizio di questo capitolo), non sarebbe corretto stabilire un automatismo tra fenomeni annotati ed errori. Tra i fenomeni annotati ci sono infatti forme molto eterogenee: troviamo sia errori veri e propri (cioè forme che sarebbero sanzionate in ogni tipo di testo, come, ad esempio, la resa senza *h* di *ho* prima persona singolare del presente indicativo del verbo *avere*), sia forme tollerate in testi poco sorvegliati, colloquiali, ma scarsamente tollerate in testi di media o alta formalità (come molte forme del cosiddetto *che* polivalente), sia, infine, forme non previste dalla grammatica normativa dell'italiano, ma ormai am-

7. Un altro fattore che risulta abbastanza esplicativo del numero di parole è la condizione di fuori sede e pendolare. Chi non è fuori sede ha una media di parole diverse utilizzate di 178.78; chi è fuori sede riporta in media 4.2 parole in più (p-value=2.66%), chi è pendolare 6.18 parole in più (p-value=0.11%). Questo dato può essere forse spiegato considerando la correlazione tra le condizioni di fuori sede e pendolare e il numero di libri letti in un anno, che in entrambi i casi è superiore rispetto a quello di chi studia nella città in cui risiede. Si può forse ipotizzare che il tempo del viaggio offra, a fuori sede e pendolari, più occasioni di lettura.

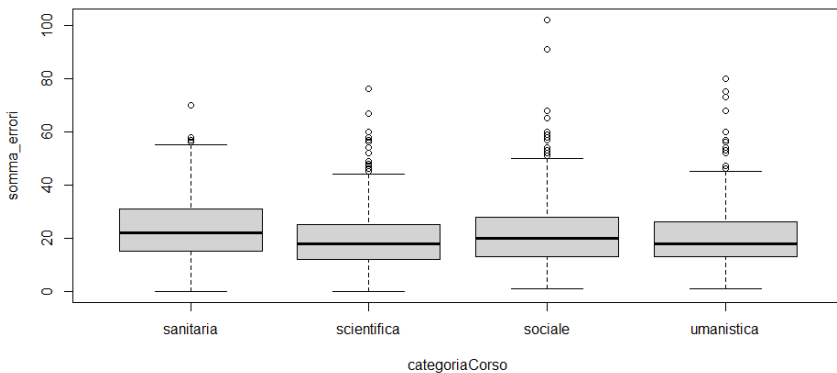
piamente tollerate anche in testi di media o alta formalità (come l'uso di *gli* pronomi dativo plurale in luogo di *loro*). Un numero elevato di annotazioni, dunque, segnala senza dubbio una certa distanza del risultato atteso ideale, ma non identifica *by default* un testo come malformato.

Nel seguito di questo paragrafo analizzeremo prima (§3.1) il numero globale di annotazioni<sup>8</sup> sui testi sia in senso univariato, sia rispetto ad alcune delle variabili sociobiografiche ricavate dal questionario (cfr. cap. 2), evidenziando quindi solo alcune tra le centinaia di correlazioni potenziali; successivamente (§3.2) presenteremo un quadro approssimativo delle correlazioni tra le singole classi di annotazione e alcune delle variabili sociobiografiche considerate.

### 3.1. Alcune correlazioni tra il numero globale di annotazioni e alcuni parametri sociobiografici

Un primo parametro significativo nella distribuzione delle annotazioni è ovviamente quello relativo all'area disciplinare del corso di laurea. Il boxplot in fig. 7 mostra che l'area sanitaria presenti il più alto numero medio di annotazioni rispetto alle altre aree, per distacco:

Fig. 7 - Numero medio di annotazioni per la categoria del corso di studi



Nella tab. 5 possiamo notare infatti che l'area sanitaria presenta un numero medio di annotazioni per testo pari a 24.15, mentre tutte le altre aree

8. Per la legenda dei codici delle annotazioni si rinvia al cap. 3.

hanno valori significativamente inferiori, in particolare l'area umanistica (p-value inferiore all'1 per 10000) e l'area scientifica (p-value inferiore all'1 per 1000000):

Tab. 5 - Modello ANOVA per il numero totale di annotazioni rispetto alla categoria del corso

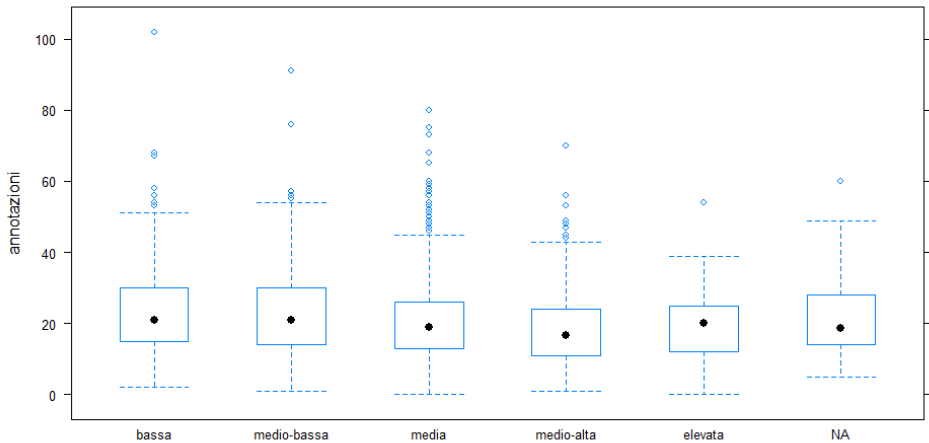
	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Sanitaria	24.1492	0.7336	32.919	< 2e-16 ***
Sociale	-2.2129	0.8763	-2.525	0.0116 *
Umanistica	-3.6919	0.8752	-4.218	2.56e-05 ***
Scientifica	-4.5245	0.8503	-5.321	1.14e-07 ***

In coerenza con quanto premesso sopra, possiamo quindi affermare come la maggior distanza tra testi effettivamente prodotti e risultato atteso ideale sia quella rilevata in area sanitaria.

Un secondo fattore esplicativo del numero totale di annotazioni risulta il genere. Le studentesse, infatti, fanno registrare un numero totale medio di annotazioni pari a 20.37, mentre negli studenti l'incremento è di 1.61 annotazioni in media (p-value pari a 0.21%).

Un fattore estremamente informativo per prevedere il numero totale di annotazioni è la classe socioeconomica degli e delle informanti. Il boxplot in fig. 8 mostra infatti un calo costante passando dalla classe bassa alla classe medio-alta e, invece, un rimbalzo per la classe alta. Il test ANOVA mostra in particolare che la media delle annotazioni per la classe media (20.72) e per la classe medio-alta (18.97) risulta significativamente inferiore rispetto alla classe bassa (24.43), con un p-value di 0.1% e 0.002%, rispettivamente:

Fig. 8 - Boxplot del numero totale di annotazioni rispetto alla classe socioeconomica

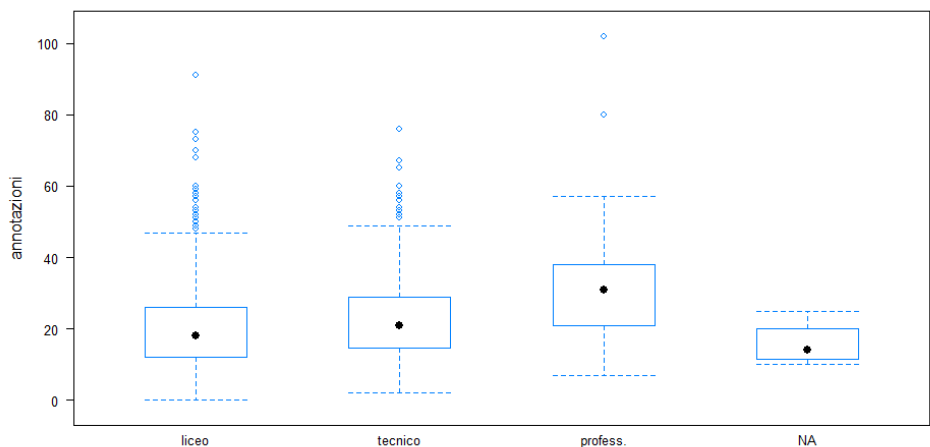


Il trend evidenziato alla fig. 8 in un certo senso conferma tendenze note e già rilevate nel cap. 2: l'accesso ad un livello superiore di istruzione e, dunque, ad una maggiore padronanza della lingua è ancora abbastanza condizionato dalle condizioni di partenza, favorendo chi vive in una situazione di maggiore agiatezza<sup>9</sup>. Occorre valutare più attentamente l'inversione della curva nel passaggio da chi dichiara di appartenere a una classe medio-alta a chi invece dichiara di provenire da un contesto familiare di grande benessere economico (in realtà solo 30 persone in tutto il campione rilevato, per metà iscritte ad atenei settentrionali, insufficienti per proporre generalizzazioni sensate).

Come prevedibile, riveste grande importanza nel prevedere il numero totale di annotazioni (e quindi la distanza del risultato atteso) anche il tipo di diploma:

9. Cfr. tra gli altri De Mauro (1963 e 2014) e Istat (2011).

Fig. 9 - Boxplot del numero totale di annotazioni rispetto al tipo di diploma



In questo caso, si osserva una differenza nettissima tra il numero di annotazioni per chi proviene dagli istituti professionali (32.10 in media) rispetto ai licei (20.08 in media). Lo scarto risulta significativo anche tra i licei e gli istituti tecnici (23.00 in media), come riportato in tab. 6:

Tab. 6 - Modello ANOVA per il numero totale di annotazioni rispetto a diploma

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Scientifico	20.0884	0.2799	71.758	< 2e-16 ***
Tecnico	2.9066	0.6331	4.591	4.66e-06 ***
Professionale	12.0160	1.4221	8.449	< 2e-16 ***
NA	-4.3384	5.7133	-0.759	0.448

Occorre ricordare, però, che in questo ambito il campione è piuttosto sbilanciato: il 77,8% dei e delle partecipanti alla rilevazione ha frequentato un liceo; il 18,9% un istituto tecnico e solo il 3,1% un istituto professionale<sup>10</sup>. La fig. 9 e la tab. 6 dunque ci dicono piuttosto che permane un gap significativo tra licei e altri istituti superiori rispetto all'accesso all'università<sup>11</sup> e, solo suc-

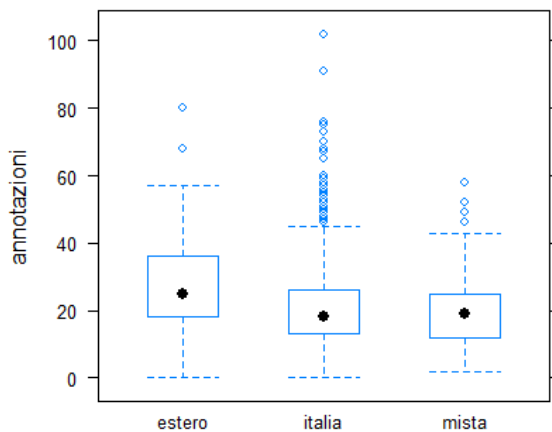
10. Cfr. anche cap. 2.

11. Si veda, di nuovo, Istat (2011) e anche il report 2024 pubblicato a questo indirizzo: <http://dati.istat.it/Index.aspx?QueryId=24985>

cessivamente, che chi non ha un'istruzione di tipo liceale fatica maggiormente nella redazione di testi formali.

Il paese di nascita e la provenienza della famiglia ovviamente impattano in modo significativo sul totale delle annotazioni. Essere nati in Italia fa diminuire il numero medio di annotazioni rilevate da 27.89 a 20.75 ( $p\text{-value}=0.00002\%$ ). Riguardo alla provenienza della famiglia, il boxplot in fig. 10 mostra un divario sostanziale tra le famiglie provenienti dall'estero e le altre, nonostante la presenza di numerose anomalie positive tra gli e le scriventi provenienti da famiglie autoctone:

Fig. 10 - Boxplot del numero totale di annotazioni rispetto alla provenienza della famiglia



La tab. 7 mostra che il numero medio di annotazioni presenti nei testi di studenti e studentesse con genitori stranieri è pari a 28.12; questo numero diminuisce di oltre 7 annotazioni per i testi redatti da studenti e studentesse con entrambi i genitori italiani e con un genitore italiano e uno straniero, con livelli di significatività estremi<sup>12</sup>:

Tab. 7 - Modello ANOVA per il numero totale di annotazioni rispetto a provenienza Famiglia

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Gen. stranieri	28.115	1.178	23.874	< 2e-16 ***
Gen. italiani	-7.424	1.207	-6.151	9.14e-10 ***
Famiglie mista	-7.719	1.555	-4.963	7.48e-07 ***

12. Su studenti e studentesse con almeno un genitore non italiano si veda anche il cap. 17.



Infine, occorre richiamare un parametro già citato in §2.2, quello relativo alla conoscenza di almeno una lingua antica, che, come si è visto, ha un effetto positivo sulla ricchezza del lessico. I dati rivelano che questo effetto positivo si estende anche alle annotazioni complessive: i testi redatti da chi non conosce almeno una lingua antica registrano 22.22 annotazioni; quelli prodotti da chi invece ha dimestichezza con almeno una lingua antica registrano in media 19.75 annotazioni ( $p\text{-value}=0.0001\%$ ). Come si è visto in §2.1, anche il numero di libri letti all'anno ha un effetto positivo sulla ricchezza del lessico. In questo caso, però, lo stesso effetto non viene registrato nel numero totale di annotazioni. La correlazione, infatti, non è significativa: il  $p\text{-value}$  ottenuto confrontando chi legge 0 libri all'anno e chi legge più di 10 libri all'anno è 7.57%. Questo dato, a nostro modo di vedere, è estremamente rilevante in quanto contribuisce a sfatare il “mito” secondo cui una notevole assiduità alla lettura determini un beneficio automatico per le abilità di scrittura: leggere molto ha senza dubbio una ricaduta positiva sulla scrittura, ma indiretta. Leggere molto, cioè, aiuta, ma non basta. Al contrario, invece, studiare e praticare almeno una lingua antica ha sulle capacità di scrittura nella lingua nativa un effetto positivo diretto. Da questo punto di vista, in sostanza, lo studio di lingue antiche pare configurarsi come una sorta di “abilità trasversale”, per utilizzare un'etichetta oggi assai in voga, dal momento che i suoi effetti si estendono ben oltre i confini della disciplina stessa<sup>13</sup>. A ben vedere, questi dati permettono di sfatare anche un secondo “mito”, quello per cui lo studio di lingue antiche sia fine a se stesso. Esso, in realtà, si configura come un'eccellente “palestra” per addestrare abilità metalinguistiche applicabili poi allo studio di qualunque lingua moderna<sup>14</sup>.

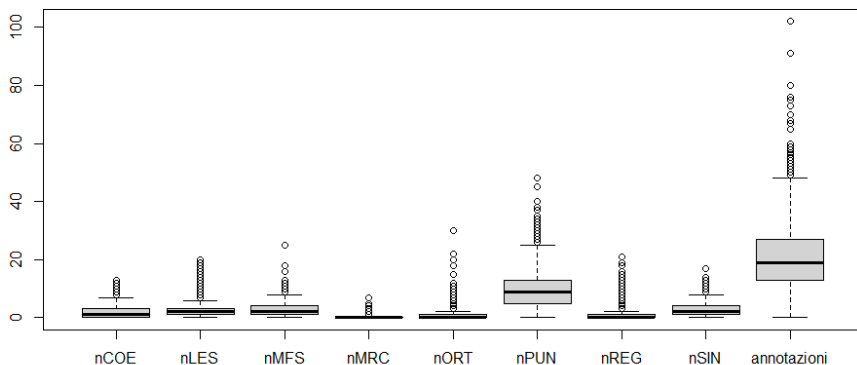
13. Sul tema vi vedano, tra gli altri, Mangiavini e Bettoni (2009), Canfora e Cardinale (2012) e Milanese (2024).

14. Ci si può chiedere se lo stesso effetto possa essere innescato da una L2 e/o dallo studio di una lingua straniera. È possibile, anzi probabile che sia così, ma al momento i nostri dati non ci permettono di dare risposte nette come quelle che, invece, riguardano le lingue antiche. Come si è visto nel cap. 2, nel questionario c'erano ovviamente anche domande relative alla conoscenza di altre lingue “vive”. Ma dal momento che lo studio di almeno una lingua straniera è di fatto obbligatorio ormai in tutti i percorsi formativi, non è possibile creare due sottocampioni nettamente distinti, uno composto da studenti e studentesse che conoscono e usano almeno una lingua diversa dalla propria e uno composto da studenti e studentesse che non conoscono e non usano almeno una lingua diversa dalla propria: questo secondo gruppo sarebbe quasi vuoto. Per verificare l'impatto della conoscenza di lingue moderne sulla competenza nella lingua nativa sarebbe necessaria una raccolta dati più granulare e mirata.

### 3.2. Alcune correlazioni tra i singoli tipi di annotazione e alcuni parametri sociobiografici

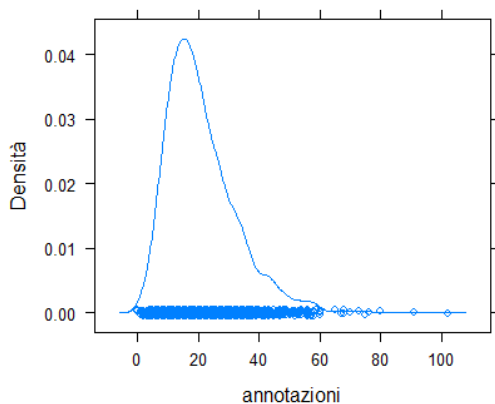
Per procedere ad un'analisi più sgranata delle annotazioni conviene prendere le mosse dalla fig. 11, che mostra il quadro dettagliato della loro distribuzione per classe:

Fig. 11 - Distribuzione delle annotazioni



In fig. 12 possiamo osservare che la distribuzione di queste annotazioni è asimmetrica a destra, a causa della presenza di parecchie anomalie positive. La mediana è pari a 19, la media è pari a 21.01, il primo quartile è 13, il terzo quartile è 27. Il massimo raggiunge addirittura 102, mentre 2 testi su 2137 non presentano alcuna annotazione:

Fig. 12 - Densità stimata del totale di annotazioni



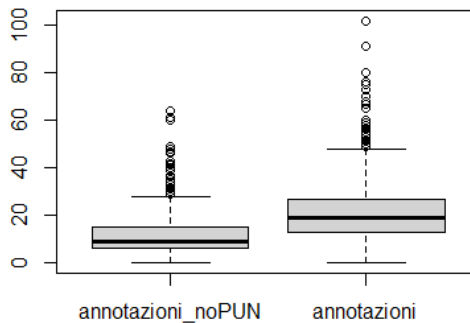
La tab. 8 riporta invece le medie osservate per ciascun tipo di annotazione. Si nota che la media totale delle annotazioni di punteggiatura per ogni testo ammonta circa a 10, quasi la metà delle 22 annotazioni totali in media per testo:

Tab. 8 - Medie per ciascuno degli otto tipi di annotazione e per il loro totale con e senza le annotazioni di punteggiatura

<i>nCOE</i>	<i>nLES</i>	<i>nMFS</i>	<i>nMRC</i>	<i>nORT</i>	<i>nPUN</i>	<i>nREG</i>	<i>nSIN</i>	Totale	noPUN
1.74	2.40	2.46	0.15	0.82	9.99	0.92	2.52	21.01	11.01

Tale squilibrio risulta ancora più evidente dalla fig. 13:

Fig. 13 - Boxplot delle annotazioni totali (a sx) e delle annotazioni senza quelle di punteggiatura (a dx)



In sostanza, la punteggiatura (nPUN), da sola, copre circa il 50% delle annotazioni rilevate in tutti i testi del corpus Univers-ITA (in media, quasi 10 annotazioni per testo) e quindi si configura, almeno a prima vista, come la vera “emergenza” nel quadro della scrittura formale all’università (cfr. cap. 14 per un’analisi approfondita dei dati e cap. 20 per possibili interventi didattici). Le categorie con meno annotazioni sono quelle relative al registro (nREG, 0,9 annotazioni a testo), all’ortografia (nORT, 0,8 annotazioni a testo) e quelle legate alla produzione di strutture sintatticamente marcate (nMRC, 0,1 annotazioni per testo in media).

Passiamo ora ad una ricognizione più dettagliata delle correlazioni tra alcuni parametri sociobiografici e le singole classi di annotazione. L’impatto

più rilevante è senza dubbio quello dell'area disciplinare del corso di studio frequentato. Prendendo come riferimento un livello di significatività dell'1%, i contrasti significativi risultano tra area scientifica e area sanitaria per annotazioni di lessico e di morfosintassi, tra area umanistica e area sanitaria per annotazioni di ortografia e tra area scientifica o umanistica e area sanitaria per annotazioni di registro e sintassi. Le tabb. 9-14 mostrano i risultati dei test ANOVA che legano queste sei variabili all'area disciplinare del corso di laurea. L'impatto più significativo in assoluto risulta quello (negativo<sup>15</sup>) dell'area scientifica sulle annotazioni errori di morfosintassi e di lessico, seguito dall'impatto (negativo) dell'area umanistica sulle annotazioni errori di ortografia, rispetto all'area sanitaria.

Tab. 9 - Modello ANOVA per le annotazioni nLES rispetto all'area disciplinare

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Sanitaria	2.9395	0.1609	18.267	< 2e-16 ***
Scientifica	-0.9243	0.1865	-4.955	7.79e-07 ***
Sociale	-0.5867	0.1922	-3.052	0.0023 **
Umanistica	-0.2313	0.1920	-1.205	0.2284

Tab. 10 - Modello ANOVA per le annotazioni nMFS rispetto all'area disciplinare

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Sanitaria	2.9637	0.1480	20.020	< 2e-16 ***
Scientifica	-0.8474	0.1716	-4.938	8.49e-07 ***
Sociale	-0.4095	0.1768	-2.316	0.0207 *
Umanistica	-0.3972	0.1766	-2.249	0.0246 *

15. *Negativo* ha un significato potenzialmente controintuitivo in questo contesto. Affermare che l'impatto più significativo è quello *negativo* dell'area scientifica su lessico e morfosintassi significa infatti che l'area scientifica fa registrare un decremento di annotazioni relative a lessico e morfosintassi rispetto alla performance peggiore, che è quella di area sanitaria. Quindi, per semplificare in modo un po' estremo, l'impatto negativo indica che la situazione dell'area scientifica è "migliore" di quella sanitaria, in quanto produce, rispetto a queste due categorie di annotazioni, testi più vicini al risultato atteso, rispetto al quale, invece, l'area sanitaria viene considerata una sorta di *worst case*.

Tab. 11 - Modello ANOVA per le annotazioni nORT rispetto all'area disciplinare

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Sanitaria	1.0847	0.1070	10.141	< 2e-16 ***
Scientifica	-0.2329	0.1240	-1.878	0.060456
Sociale	-0.2017	0.1278	-1.579	0.114516
Umanistica	-0.4720	0.1276	-3.700	0.000221 ***

Tab. 12 - Modello ANOVA per le annotazioni nPUN rispetto all'area disciplinare

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Sanitaria	11.1653	0.3965	28.161	< 2e-16 ***
Scientifica	-1.3481	0.4596	-2.934	0.00339 **
Sociale	-0.5887	0.4736	-1.243	0.2139
Umanistica	-2.0322	0.4730	-4.297	1.81e-05 ***

Tab. 13 - Modello ANOVA per le annotazioni nREG rispetto all'area disciplinare

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Sanitaria	1.2379	0.1234	10.029	< 2e-16 ***
Scientifica	-0.4346	0.1431	-3.038	0.00241 **
Sociale	-0.2310	0.1474	-1.567	0.11729
Umanistica	-0.3966	0.1473	-2.693	0.00713 **

Tab. 14 - Modello ANOVA per le annotazioni nSIN rispetto all'area disciplinare

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Sanitaria	2.7379	0.1363	20.088	< 2e-16 ***
Scientifica	-0.4664	0.1580	-2.953	0.00319 **
Sociale	-0.1045	0.1628	-0.642	0.52098
Umanistica	-0.1253	0.1626	-0.770	0.44111

La classe socioeconomica risulta di particolare impatto sulle annotazioni di coerenza testuale: le classi medio-bassa, media e medio-alta presentano un impatto significativamente negativo rispetto alla classe bassa. Si veda la tab. 15:

Tab. 15 - Modello ANOVA per le annotazioni nCOE rispetto alla classe socioeconomica

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Bassa	2.2885	0.1695	13.504	< 2e-16 ***
Medio-bassa	-0.5331	0.1892	-2.817	0.004890 **
Media	-0.1045	0.1628	-0.642	0.001010 **
Medio-alta	-0.1253	0.1626	-0.770	0.000214 ***
Elevata	0.2977	0.3629	0.820	0.412060
NA	-0.3425	0.2628	-1.303	0.192646

Per le annotazioni di morfosintassi, si rileva una differenza negativa e significativa rispetto alla classe bassa per la classe media (p-value=1.5%) e per la classe medio-alta (p-value=0.3%). La differenza risulta pure negativa e significativa rispetto alla classe bassa per la classe medio-alta sulle annotazioni di punteggiatura (p-value=0.2%) e sintassi (p-value=0.9%), e per la classe media sulle annotazioni di registro (p-value=0.8%). Un ulteriore, importante fattore esplicativo è il genere: il genere maschile infatti impatta in modo positivo<sup>16</sup> e significativo sulla media delle annotazioni di ortografia (che salgono dallo 0.70 rilevato per il genere femminile ad 1.04, p-value=0.0009%) e di punteggiatura (da 9.57 a 10.67, p-value=0.009%). In questo caso, dunque, l'analisi più dettagliata consente di circostanziare meglio quanto rilevato in §3.1 sul totale delle annotazioni.

Per quanto concerne l'area geografica, essa impatta sulle annotazioni di lessico e di sintassi, come riportato nelle tabb. 16 e 17. Nel primo caso, il Nord e il Sud presentano un impatto negativo e significativo rispetto al Centro. Nel secondo caso, vale lo stesso, ma con segno positivo.

Tab. 16 - Modello ANOVA per annotazioni nLES rispetto all'area geografica

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Centro	2.6582	0.1049	25.34	< 2e-16 ***
Nord	-0.2867	0.1353	-2.12	0.03415 *
Sud e Isole	-0.4384	0.1447	-3.03	0.00247**

16. Coerentemente con quanto affermato nella n. 15, qui *positivo* ha di fatto un valore negativo, cioè registra una performance peggiore. *Positivo* infatti indica un incremento di annotazioni, quindi, semplificando, un allontanamento del risultato atteso.

Tab. 17 - Modello ANOVA per annotazioni nSIN rispetto all'area geografica

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>p-value</i>
Centro	2.31980	0.0884	26.238	< 2e-16 ***
Nord	0.22454	0.11403	1.969	0.04906 *
Sud e Isole	0.33975	0.12194	2.786	0.00538**

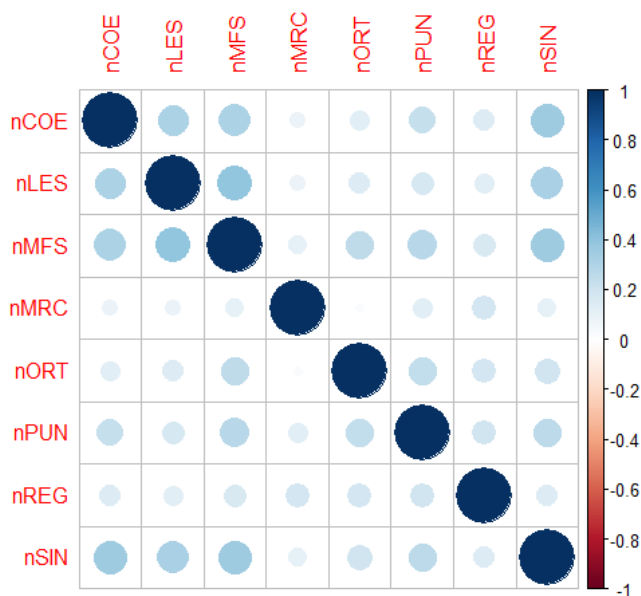
In generale, i dati presentati in questa sezione mostrano tendenze talvolta non facilmente decifrabili. La più netta è senza dubbio quella che di nuovo identifica stabilmente studenti e studentesse di area sanitaria come quelli che marcano la distanza più netta, nella redazione del testo scritto formale, dal risultato atteso per tutte le classi di annotazione considerate. L'area che, nel complesso, pare registrare il numero minore di annotazioni per le tipologie analizzate è invece quella umanistica. Gli altri dati non consentono al momento generalizzazioni di rilievo. Se infatti quanto affermato qui, con granularità più fine, sulla classe socio-economica e sul genere conferma sostanzialmente quanto rilevato nella sezione precedente sul totale delle annotazioni, occorre invece capire, con analisi più approfondite, quale sia la ragione delle correlazioni rilevate nelle tabb. 16 e 17.

#### 4. Matrice di correlazione tra le annotazioni

Uno degli obiettivi del progetto Univers-ITA era quello di affrontare le specificità della lingua scritta formale di studenti e studentesse dell'università adottando anche un approccio "tipologico", allo scopo di individuare innanzitutto pattern ricorrenti nei testi raccolti e, successivamente, di proiettarli nel quadro tanto del diasistema italiano, quanto della più ampia variazione interlinguistica. Questo secondo tema non può essere affrontato in questa sede, nella quale invece cercheremo di soffermarci sul primo, focalizzando la nostra attenzione sulle possibili correlazioni tra le diverse classi di annotazione e, quindi, sulla possibilità di individuare una sorta di tipo ricorrente nei testi analizzati.

La fig. 14 presenta una matrice di correlazione tra otto categorie di annotazioni. La grandezza e l'intensità delle sfere azzurre sono direttamente proporzionali alla significatività della correlazione:

Fig. 14 - Rappresentazione grafica della matrice di correlazione delle otto categorie di annotazione



La fig. 14 mostra che le correlazioni tra le otto classi di annotazione sono tutte positive, seppure con gradi variabili. Al tempo stesso, si nota che esiste un gruppo di variabili più fortemente correlate tra loro. L'analisi delle componenti principali ottenuta standardizzando le variabili (cfr. tab. 18) mette in luce infatti la presenza di una prima componente che spiega oltre il 30% della varianza complessiva, seguita da una seconda e una terza che spiegano tra il 10% e il 15%. Guardando ai pesi delle prime tre componenti principali, scopriamo che componente 1 in tab. 18 rappresenta una sorta di indicatore complessivo sulle annotazioni; componente 2 rappresenta il contrasto tra annotazioni di coesione, lessico, morfosintassi e sintassi ed annotazioni di frasi marcate, ortografia, punteggiatura e registro. La componente 3 è quasi identica alla 2, a parte che morfosintassi e frasi marcate si scambiano di segno:



Tab. 18 - Pesi delle prime tre componenti principali relativi alle otto variabili

	Comp.1	Comp.2	Comp.3
nCOE	0.39	0.24	0.23
nLES	0.39	0.32	0.22
nMFS	0.45	0.17	-0.02
nMRC	0.17	-0.60	0.57
nORT	0.30	-0.16	-0.70
nPUN	0.36	-0.21	-0.26
nREG	0.26	-0.58	-0.05
nSIN	0.42	0.20	0.11
%varianza spiegata	31.28	13.61	12.18

In sostanza, la co-occorrenza tra tutte le otto classi di annotazioni spiega il 31.28% della variabilità totale, cioè circa un terzo. Possiamo quindi affermare che è piuttosto probabile incappare in un testo in cui si trova almeno un esempio di annotazione appartenente a tutte le otto classi. Invece il contrasto tra annotazioni nCOE, nLES, nMFS e nSIN da una parte e le altre annotazioni dall'altra spiega il 13.61% della variabilità totale. Il contrasto tra nCOE, nLES, nMRC e nSIN da una parte e le altre annotazioni dall'altra spiega il 12.18% della variabilità totale.

Quindi, la co-occorrenza tra tutte le annotazioni, di fatto, domina in termini esplicativi ed è, di fatto, il dato che emerge in modo lampante. Ciò ci induce ad affermare che nei testi raccolti, sulla base delle annotazioni apposte, non è possibile individuare un pattern che prevale sugli altri in modo netto. I due gruppi di annotazioni in Comp.2 e Comp.3 spiegano una quota decisamente inferiore di variabilità, ma sono senza dubbio i due *cluster* più significativi in termini di varianza spiegata. Quindi, per riassumere, al netto del fatto che tutti i tipi di annotazione tendono a concorrere, si può affermare che la co-occorrenza tra nCOE, nLES, nMFS e nSIN è quella più frequente, seguita a poca distanza da quella tra nCOE, nLES, nMRC e nSIN. Comp.2 e 3 di fatto ci dicono che è possibile isolare un nucleo costante di annotazioni: nCOE, nLES e nSIN. Questo “addensamento” può sostanzialmente essere considerato come la peculiarità della maggior parte dei testi raccolti o, per ragionare in termini tipologici, come la correlazione di tratti strutturali che più li contraddistinguono.

## 5. Conclusioni

Possiamo tornare, ora, all'interrogativo che ha aperto questo contributo, cercando di operare qualche generalizzazione conclusiva. Quanto osservato nei paragrafi precedenti consente di individuare alcune linee di tendenza di ordine generale nella scrittura formale di studenti e studentesse universitari.

Rispetto alla natura dei testi raccolti, le analisi quantitative e qualitative consentono di isolare tre classi di "deviazione" rispetto al risultato atteso che tengono a co-occorrere con estrema frequenza: quelle riconducibili a coerenza (nCOE), sintassi e coesione (nSIN) e lessico (nLES). Questa correlazione costituisce in sostanza un addensamento molto evidente di fenomeni linguistici, attorno al quale si colloca poi una costellazione di altri fenomeni attestati con gradi diversi di frequenza (al netto della punteggiatura, sulla quale torneremo a breve). Il forte legame tra nCOE e nSIN, in particolare, consente senza dubbio di affermare che l'ambito in cui si registra la maggior distanza tra il risultato atteso ideale e i testi effettivamente prodotti da chi ha partecipato al progetto coincide con le attività di pianificazione, organizzazione e strutturazione dei testi stessi<sup>17</sup>. Se infatti volessimo individuare una caratteristica peculiare nella scrittura formale di studenti e studentesse universitari, alla luce di quanto emerso dai dati presentati in questo contributo, essa andrebbe senza dubbio identificata nella "difficoltà" a costruire un testo complesso, pienamente coerente e coeso, del tutto aderente alle aspettative legate al contesto nel quale è stato elaborato<sup>18</sup>. Quanto affermato sopra sulla punteggiatura, che, si è visto, costituisce l'ambito più problematico nella produzione scritta formale di studenti e studentesse, è in effetti la manifestazione di superficie più evidente di questa situazione, dal momento che la punteggiatura ha, come scopo primario, quello di scandire l'organizzazione di un testo, contrassegnandone le unità strutturali e informative<sup>19</sup>. Quanto ciò sia determinato dall'effetto delle "nuove" scritture per nulla o poco sorvegliate (tipicamente, quelle della messaggistica istantanea) dovrà essere verificato con indagini più approfondite e sistematiche.

Queste generalizzazioni, come anticipato nell'introduzione al volume, non ci consentono di trarre conclusioni nette sulla competenza di studenti e studentesse e, soprattutto, su un suo eventuale indebolimento negli ultimi decenni. Di certo, i dati analizzati in questa sede rivelano che la forbice tra quanto ci si attendeva di trovare e quanto si è effettivamente trovato non è, in media,

17. Sulla nozione di testo e sul ruolo che coerenza e coesione giocano in questo ambito cfr., tra gli altri, Andorno (2003) e Ferrari (2014).

18. Cfr. anche cap. 13.

19. Cfr. cap. 14.

significativamente larga per quanto concerne i singoli parametri analizzati o, meglio, le singole abilità linguistiche; si divarica di più, invece, quando queste singole abilità devono essere combinate tra loro nella pianificazione di strutture articolate e complesse. Per quanto, come si è detto in §1 e nel cap. 3, la produzione scritta formale di studenti e studentesse dell'università appaia piuttosto omogenea, almeno secondo le analisi quantitative, questa tendenza non è distribuita in modo uniforme dal punto di vista sociobiografico. In questo ambito, i dati raccolti confermano a grandi linee il quadro noto che vede gli atenei del Nord sistematicamente meglio posizionati di quelli del Centro e del Sud, le classi socioeconomiche basse e medio-basse caratterizzate da una competenza più incerta e un gap al momento incolmabile tra i diversi titoli di studio secondari di secondo grado. Forse meno evidente del previsto risulta la preminenza dei corsi umanistici rispetto alle capacità linguistiche, mentre appare costante, in tutti gli ambiti esplorati, il distacco dell'area sanitaria, stabilmente in ultima posizione. Viene confermato anche il beneficio che ha, sulle competenze linguistiche, la consuetudine con la lettura, tuttavia limitatamente alla ricchezza del lessico e, a quanto pare, non sulla capacità di scrittura, che trae invece giovamento dalla conoscenza di lingue antiche.

Serviranno ovviamente analisi più approfondite e una ricognizione più dettagliata dei numerosissimi dati che il progetto ha prodotto per circostanziare con più precisione queste considerazioni. Per il momento, quello che si può asserire, in conclusione, è che non possiamo certamente affermare che studenti e studentesse scrivano “peggio” rispetto al passato, ma possiamo di certo affermare che la loro produzione è diversa, in un certo senso inattesa, rispetto alle aspettative e che ciò è determinato non tanto da presunte “difficoltà” di scrittura, quanto, piuttosto, da una limitata dimestichezza con alcune specifiche tipologie testuali e con le scelte linguistiche che esse comportano.

## Riferimenti bibliografici

Andorno C. (2003), *Linguistica testuale*, Carocci, Roma.

Canfora L. e Cardinale U., a cura di (2012), *Disegnare il futuro con intelligenza antica. L'insegnamento del latino e del greco antico in Italia e nel mondo*, il Mulino, Bologna.

Collins-Thompson K. (2014), “Computational assessment of text readability: A survey of current and future research”, *ITL-International Journal of Applied Linguistics*, 165, 2: 97-135.

Cicchitelli G., D'Urso P. e Minozzo M. (2022<sup>4</sup>). *Statistica: principi e metodi*, Pearson, Milano.

De Mauro T. (1963), *Storia linguistica dell'Italia unita*, Laterza, Roma-Bari.

- De Mauro T. (2014), *Storia linguistica dell'Italia repubblicana dal 1946 ai nostri giorni*, Laterza, Roma-Bari.
- Dell'Orletta F., Montemagni S. and Venturi G. (2011), *READ-IT: assessing readability of Italian texts with a view to text simplification*, in Alm N. (ed.), *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*, Association for Computational Linguistics, Edinburgh, 73-83 , testo disponibile al sito: <https://aclanthology.org/W11-2308/>
- Ferrari A. (2014), *Linguistica del testo. Principi, fenomeni, strutture*, Carocci, Roma.
- Grandi N., Ballarè S., Chiusaroli F., Gallina F., Pascoli M. e Pistolesi E. (2023), *Corpus Univers-ITA*, doi: <https://doi.org/10.60760/unibo/univers-ita>
- Grandi N., Pascoli M. (2025), *Dataset del progetto Univers-ITA*. Università di Bologna, doi: <https://doi.org/10.6092/unibo/amsacta/8229>
- Istat (2011), *L'Italia in 150 anni. Sommario di statistiche storiche 1861-2010*, testo disponibile al sito: [www.istat.it/it/archivio/228440](http://www.istat.it/it/archivio/228440)
- Mangiavini M. e Bettoni M (2009), “Lingue classiche, complessità e competenze”, *Interaction Design and Architecture(s) Journal - IxD&A*, 7-8: 48-50, testo disponibile al sito: [https://ixdea.org/wp-content/uploads/IxDEA\\_art/8/IxDEA\\_8\\_8.pdf](https://ixdea.org/wp-content/uploads/IxDEA_art/8/IxDEA_8_8.pdf)
- Milanese G. (2024), *Le ragioni del latino*, Scholé, Brescia.

## Materiali Linguistici

---

**L'**italiano è davvero una lingua in declino? È vero che gli studenti e le studentesse oggi hanno difficoltà a scrivere e comprendere testi complessi? La tecnologia ha davvero impoverito la nostra lingua? Queste domande corrispondono a una serie di luoghi comuni molto diffusi nell'opinione pubblica, ma quasi mai confermati dai dati. Questo volume raccoglie i primi risultati di un progetto di ricerca pluriennale finalizzato a produrre la prima mappatura sistematica mai realizzata delle capacità di scrittura formale della popolazione studentesca universitaria italiana. Un gruppo di ricerca interdisciplinare (composto da linguisti, statistici, informatici e sociologi) ha analizzato sia i principali tratti linguistici e grammaticali, sia i più importanti correlati sociobiografici della produzione scritta formale di un campione di oltre 2.000 studenti e studentesse di 44 atenei italiani, rappresentativo per aree disciplinari e geografiche. Il quadro che emerge è molto complesso e articolato: un confronto con dati del passato pare mostrare, in alcune aree, un seppur lieve miglioramento nelle competenze di scrittura; tuttavia, esistono aree di oggettiva debolezza, come, ad esempio, nelle attività di pianificazione del testo.

*Contributi di:* S. Ballarè, F. Chiusaroli, F. Da Milano, E. Di Domenico, M. Farnè, F. Gallina, C. Gianollo, N. Grandi, A. Iannella, Y. Martari, E. Miola, S. Scaglione, M. Pascoli, F. Pecorari, M.L. Pierucci, E. Pistolesi, R. Pugliese, M.L. Restivo, M. Tivosanis, L. Tramutoli, S. Tusini, S. Valente.

*Nicola Grandi* è professore ordinario di Glottologia e Linguistica presso il Dipartimento di Filologia classica e Italianistica dell'Alma Mater Studiorum – Università di Bologna. Si occupa di tipologia linguistica, di sociolinguistica, con particolare riferimento alle tendenze dell'italiano contemporaneo, e di educazione linguistica. Ha fondato e dirige il sito divulgativo [www.linguisticamente.org](http://www.linguisticamente.org).