

## ARTICOLI

UNIVERS-ITA. DESCRIZIONE E PRIMI RISULTATI DI UNO STUDIO  
DELL'ITALIANO SCRITTO DI STUDENTI UNIVERSITARI

Nicola Grandi - Università degli Studi di Bologna

Silvia Ballarè - Università degli Studi di Bologna

Yahis Martari - Università degli Studi di Bologna

Emanuele Miola - Università degli Studi di Bologna<sup>1</sup>

## 1. Gli obiettivi del progetto

La discussione sullo stato di salute dell'italiano, soprattutto nelle generazioni più giovani e, in particolare, nelle studentesse e negli studenti dell'università (d'ora in poi s&s), torna periodicamente alla ribalta, soprattutto nella pubblicistica generalista. Come è noto, al centro del dibattito c'è la percezione di una regressione delle competenze di scrittura di s&s, che di fatto rappresenterebbe il sintomo più evidente di un declino generale della nostra lingua. In questo quadro, l'attenzione si è concentrata sull'università perché chi la frequenta rappresenta lo strato teoricamente più 'colto', più esposto a usi formali della lingua, sia nello scritto che nel parlato, e, dunque, al contempo più competente della varietà standard e meno ricettivo ai cambiamenti. È quindi evidente come sistematiche deviazioni rispetto allo standard da parte di s&s acquistino un significato particolare. Questa sensazione di declino è però spesso basata su intuizioni ed esperienze di carattere personale, quindi su dati certo significativi, ma non rappresentativi e generalizzabili. Il progetto *Univers-ITA. L'italiano scritto degli studenti universitari: quadro sociolinguistico, tendenze tipologiche, implicazioni didattiche*<sup>2</sup>, che qui si presenta, si è posto l'obiettivo di tracciare un profilo esaustivo dell'italiano scritto di s&s, analizzando la produzione scritta formale di un campione rappresentativo per aree disciplinari e geografiche e costruendo strumenti di ricerca utili a un'indagine

minuziosa di questo segmento del sistema linguistico italiano (nello specifico, tre corpora linguistici a libero accesso, per i quali cfr. 2.1 e 2.2). Si tratta, di fatto, del primo studio davvero sistematico condotto su questo tema: manca dunque, per il passato, un'indagine paragonabile a quella che si presenta in questa sede che possa costituire un termine di confronto tale da consentirci di stabilire se la lingua abbia davvero subito quel presunto declino che spesso viene denunciato e se, nel passaggio da una generazione all'altra, ci sia stata una reale perdita nelle capacità di scrittura. Per questa ragione, i risultati che ci accingiamo ad illustrare vanno considerati come un punto di partenza, più che come un punto di arrivo.

## 1.1. Il campione

L'indagine è stata condotta nell'anno accademico 2020/21 su s&s del secondo anno delle lauree triennali e delle lauree magistrali a ciclo unico (quindi appartenenti alla coorte 2019/20). La scelta del secondo anno è stata compiuta per distanziare la rilevazione dalla conclusione della scuola secondaria di secondo grado, al fine di mitigare l'effetto della formazione secondaria sui dati raccolti. Come si spiegherà meglio nella sezione seguente, s&s hanno redatto un testo tra le 250 e le 500 parole su una traccia comune e, successivamente, hanno compilato un questionario sociobiografico di oltre 50 domande, finalizzato a tracciare un profilo dettagliato degli scriventi<sup>3</sup>. L'impegno previsto era di circa un'ora. Per incentivare la partecipazione alla raccolta dati è stata abbinata alla medesima una sorta di lotteria che metteva in palio buoni da 25 euro per l'acquisto di libri.

1 Il presente testo è costruito sulla base del report finale del progetto PRIN 2017, *Univers-ITA*, una sintesi curata dal coordinatore del progetto Nicola Grandi e pubblicata sul sito del medesimo: <<https://site.unibo.it/univers-ita/it/divulgazione/report-finale>>. La stesura di questo contributo è così ripartita: par. 1 è attribuibile a N. Grandi, par. 2 a Y. Martari e par. 3 a E. Miola (3.1), Y. Martari (3.2) e S. Ballarè (3.3).

2 Il progetto prevedeva quattro unità di ricerca: Università degli Studi di Bologna (coordinatore N. Grandi), Università degli Studi di Macerata (coordinatrice F. Chiusaroli), Università per Stranieri di Perugia (coordinatrice E. Pistolesi) e Università degli Studi di Pisa (coordinatrice F. Gallina). Cfr. <<https://site.unibo.it/univers-ita/it/>>.

3 Il questionario conteneva domande relative al corso di studi frequentato, al percorso di studi precedente, alle abitudini di lettura e scrittura, ai modelli culturali di riferimento, alle altre lingue conosciute, all'area geografica di provenienza e ad eventuali esperienze di migrazione, alle caratteristiche socio-economiche e culturali della famiglia di provenienza, ecc. Molti di questi dati sono poi riportati sul corpus e costituiscono anche chiavi di ricerca per l'estrazione e, poi, per l'interpretazione dei dati.

Nella progettazione del campione<sup>4</sup> si è stabilito un target teorico di 3.000 partecipanti, calcolato sulla coorte precedente a quella indagata, cioè la 2018/19, che constava di 300.263 iscritti/e (in modo da avere cioè una copertura di 1/100 circa). A monte, sono stati controllati due parametri, quello relativo alla distribuzione dei partecipanti per area geografica dell'Ateneo e per area disciplinare. Per quanto concerne il primo parametro, il campione teorico è stato progettato come segue:

- Atenei del Nord: 47% [1410 informanti]
- Atenei del Centro: 23% [690 informanti]
- Atenei del Sud: 30% [900 informanti]

Nel progetto iniziale (e ideale) la rilevazione si sarebbe dovuta svolgere in presenza e a sorpresa, durante lezioni curricolari, in alcuni Atenei rappresentativi di queste aree: Milano e Bologna per il Nord, Macerata, Perugia e Pisa per il Centro e Napoli per il Sud.

Per quanto riguarda, invece, le aree disciplinari, sono stati inizialmente selezionati sette corsi che garantissero un numero sufficiente di iscritti in tutte le aree geografiche, secondo la scansione seguente:

- > Area sanitaria (classi di laurea: L-SNT/1, LM-13): 10% [300 informanti]
- > Area scientifica (classe di laurea: L-9, nello specifico Ingegneria gestionale): 36% [1080 informanti]
- > Area umanistica (classi di laurea: L-11 e LM-85bis): 19% [570 informanti]
- > Area sociale (classi di laurea: L-18, nello specifico Economia aziendale e LMG/01): 35% [1050 informanti]

Le restrizioni legate all'emergenza sanitaria, esplosa proprio al momento di avviare la raccolta dei dati, hanno imposto una profonda revisione di questo piano di lavoro. Il sistema di raccolta dei dati è stato dunque adattato per una rilevazione a distanza. Su questo aspetto si tornerà più nel dettaglio nel paragrafo successivo. In questa sezione ci limitiamo a citare gli effetti che questo cambiamento di strategia ha determinato sul campione indagato. Il progetto Univers-ITA è stato presentato nelle aule virtuali nel corso di lezioni curricolari. Nella maggior parte dei casi, s&s hanno poi svolto il compito da casa; solo in alcune circostanze ci è stato permesso di raccogliere i dati durante la lezione. La partecipazione, inizialmente pensata come 'coercitiva' e direttamente monitorabile data la presenza in aula dei partecipanti e del ricercatore, è divenuta dunque di fatto volontaria e non sorvegliata (e questo può certamente determinare un effetto leggermente distortivo sui risultati dell'in-

indagine, dal momento che si può supporre che abbiano partecipato, in media, s&s più motivati). Questo ha ovviamente comportato un decremento del tasso di partecipazione media prevista e ci ha indotto ad ampliare il numero degli Atenei coinvolti (45 alla fine del progetto). Stanti le mutate condizioni complessive, ci si è posti l'obiettivo minimo di coprire almeno il 60% del campione teorico inizialmente progettato. Nell'area tecnologica e nell'area sociale la raccolta ha avuto maggiori difficoltà (dovute in larga parte al minor interesse mostrato dalle colleghe e dai colleghi a cui abbiamo chiesto ospitalità per presentare il progetto in classe durante una loro lezione) e questo ci ha imposto di rivedere anche le classi di laurea individuate come rappresentative. Per l'area tecnologica, abbiamo quindi rilevato anche i corsi della classe L-9 con denominazione diversa da ingegneria gestionale e i corsi delle classi L-7 e L-8: questo ci ha consentito di raggiungere l'obiettivo di minima che ci eravamo posti. Per l'area sociale, abbiamo cercato di rilevare tutti i corsi della classe L-18, ma in questo caso il risultato è stato comunque inferiore all'obiettivo minimo.

Hanno complessivamente partecipato alla raccolta dei dati 2137 s&s (1/144 immatricolati nella coorte 2019/20), così ripartiti per i due parametri indicati sopra:

Area geografica:

- Atenei del Nord: 41,7% [891 informanti]
- Atenei del Centro: 30,6% [655 informanti]
- Atenei del Sud: 27,7% [591 informanti]

Aree disciplinari:

- > Area sanitaria (classi di laurea: L-SNT/1, LM-13): 11,605% [248 informanti]
- > Area scientifica (classe di laurea: L-7, L8 e L-9): 33,785% [722 informanti]
- > Area umanistica (classi di laurea: L-11 e LM-85bis): 27,421% [586 informanti]
- > Area sociale (classi di laurea: L-18 e LMG/01): 27,187% [581 informanti]

Rispetto al campione ideale, dal punto di vista geografico c'è una certa sovrarappresentazione degli Atenei del Centro rispetto a quelli del Nord e del Sud; dal punto di vista disciplinare, invece, si registra uno sbilanciamento a favore dell'area umanistica, a scapito di quella sociale.

Altri parametri, oltre ai due appena citati, sono stati controllati solo a valle della rilevazione. Tra i più significativi vi sono quelli relativi alla classificazione per genere e al titolo di studio di s&s. Per quanto riguarda il primo, il campione Univers-ITA è rappresentato da 1346 informanti di genere femminile (63% del totale) e da 779 informanti di genere maschile (il 36,4% del totale); 7 informanti non hanno seleziona-

<sup>4</sup> La costruzione del campione è stata curata dall'unità di ricerca di Perugia Stranieri.

to alcun genere. Rispetto al valore nazionale della coorte 2019/20, nella quale il rapporto tra i due generi è 55,5% (F) e 44,5% (M), c'è un significativo incremento del genere femminile. Per quanto concerne le cinque aree disciplinari, utilizzando come termine di riferimento la classificazione per genere interna alla coorte 2019/20 a livello nazionale nel campione del progetto si osserva come questo sbilanciamento sul genere femminile si concentri principalmente in area sanitaria e umanistica: in questi casi, lo scarto rispetto al valore nazionale è rilevante (prossimo al 10%). Nelle aree scientifica e sociale la distribuzione dei due generi nel campione Univers-ITA non si discosta in modo significativo da quella nazionale.

Per quanto riguarda il titolo di studio superiore, occorre innanzitutto precisare che la quasi totalità dei partecipanti al progetto (il 98,1%) ha svolto in Italia l'intero percorso di scolarizzazione. Come prevedibile, il titolo liceale è di gran lunga il più rappresentato (77,8% del totale), seguito, con distacco notevole, dal diploma di istruzione tecnica (18,9%); resta ancora sporadico l'accesso all'università da parte di chi ha un diploma di tipo professionale (3,1%). La distribuzione di s&s con titolo liceale è sostanzialmente omogenea nelle aree umanistica, scientifica e sanitaria, nelle quali esso caratterizza tra l'80 e l'85% degli immatricolati. In area sociale, invece, la percentuale scende a poco più del 60%: qui, in effetti, si concentra una quota molto rilevante di chi ha un diploma di tipo tecnico, che rappresenta poco oltre il 30% degli iscritti (il triplo rispetto alle aree sanitaria e umanistica; in area scientifica la percentuale di chi ha un diploma di tipo tecnico è di poco inferiore al 20%).

## 2. La raccolta dei dati

La risorsa Univers-ITA è costituita da tre corpora interrogabili e disponibili online, per i quali è stata necessaria una triplice raccolta di dati. Vediamo il dettaglio di questo processo, con particolare attenzione verso la raccolta che ha prodotto il corpus di testi *ad hoc*.

### 2.1. I testi *ad hoc*

Come si è accennato sopra, ogni partecipante è stato invitato a redigere, in modo anonimo, un testo secondo una traccia comune (cfr. fig. 3). La consegna conteneva indicazioni rispetto al registro (sorvegliato) e alla tipologia (argomentativa) del testo: è stato infatti chiesto di produrre un testo tra 250 e 500 parole con uno stile formale («come se scrivessi per un tuo professore»), immaginando di rispondere a un sondaggio «rivolto a tutti gli studenti sulla didattica

a distanza nei mesi di emergenza sanitaria»; con la precisazione di scrivere «in modo non schematico, i vantaggi e gli svantaggi della didattica a distanza» secondo il proprio punto di vista.

Per evitare di influenzare il campione, l'accesso al sito web per la raccolta dei dati è stato protetto da username e password forniti a s&s di volta in volta.

L'interfaccia era presentata come segue:



fig. 1. Prima schermata dell'interfaccia per la raccolta dei dati.



fig. 2. Seconda schermata dell'interfaccia per la raccolta dei dati.



fig. 3. L'interfaccia per la stesura del testo.

Dopo aver redatto il testo, s&s accedevano alla sezione dedicata al questionario sociobiografico.

I 2137 testi (per complessive 810.715 parole) sono stati poi sottoposti a una duplice valutazione: la prima automatica, prevalentemente quantitativa; la seconda critica e qualitativa. Quest'ultima valu-

tazione è stata principalmente finalizzata a correggere e bilanciare le inevitabili generalizzazioni che il processo di correzione automatizzata applica ai testi, attraverso una competenza testuale e stilistica che solo l'analisi umana può garantire. Nella prima fase, dunque, i testi raccolti sono stati lemmatizzati e pos-tagati e successivamente sono stati esaminati automaticamente utilizzando le funzionalità del software READ-IT<sup>5</sup>, che valuta la leggibilità di un testo secondo parametri come il numero di frasi,

<sup>5</sup> <<http://www.italianlp.it/demo/read-it/>>.

il numero di parole, la lunghezza delle frasi, la lunghezza delle parole, ecc. In seconda battuta, tutti i testi sono stati riletti e annotati manualmente da due lettrici con una solida competenza linguistica e metalinguistica acquisita durante un percorso di studi magistrale incentrato sulle scienze linguistiche e con una specifica formazione sulla lingua italiana. L'annotazione manuale è stata preceduta da una lunga fase di training e di allineamento delle due annotatrici ai criteri definiti in precedenza dal gruppo di ricerca per i seguenti aspetti (cfr. tab. 1):

Parametri	Tag	Fenomeni
Organizzazione del testo	PAR1, PAR2, PAR3	Mancata/scorretta suddivisione del testo in paragrafi Esposizione schematica dei contenuti
Ortografia	ORT	Assenza/impiego scorretto dell'apostrofo Uso dell'accento con forme verbali monosillabiche ( <i>fa, sa, so</i> ) e con la forma apocopata dell'avverbio poco Da non considerare: errori di battitura; uso dell'accento grave al posto di quello acuto e viceversa
Registro	REG	Lessico non adeguato al contesto scritto sorvegliato mediamente formale Uso di <i>gli sovraesteso per loro e le</i> Uso del <i>tu</i> impersonale
Frase marcate	MRC	Frase dislocate a dx/sx Frase scisse e pseudoscisse Frase a tema sospeso
Lessico	LES	Povertà/eccessiva genericità lessicale Lessico improprio Ripetizioni Platismi Violazione di collocazioni Malapropismi
Punteggiatura	PUN	Omissione dei segni interpuntivi Sostituzione di un segno interpuntivo con un altro Inserimento di segni interpuntivi in contesti incongrui
Morfosintassi	MFS	Mancato accordo per genere e numero Mancato rispetto della <i>consecutio temporum</i> Inadeguata gestione del riferimento (ad esempio, pronomi distanti dai loro antecedenti, pronomi che rimandano a referenti dotati di realtà concettuale anziché testuale) Reggenze preposizionali errate <i>che</i> polivalente
Coerenza	COE	Uso illogico dei connettivi Mancata esplicitazione delle relazioni logiche che intercorrono fra i contenuti espressi (giustapposizione) Contraddittorietà Frammentazione delle informazioni "Cortocircuiti semantici", come in espressioni del tipo <i>possibile/impossibile + potere; permettere + potere; ecc.</i>

Sintassi e coesione	SIN	Omissione della preposizione nella coordinazione di sintagmi Mancati o scorretti parallelismi Gerundi assoluti Omissioni argomentali, ad esempio: “ricominciare a recarsi in presenza” Interruzione della continuità sintagmatica, ad esempio: <i>Basterebbe pensare alle famiglie che vivono, magari anche numerose, in un monolocale.; Sono, infine, felice delle scelte fatte dai miei professori.</i>
---------------------	-----	---

tab. 1. Criteri di annotazione manuale.

Sono state annotate tutte le forme non adatte a un testo sorvegliato come quello richiesto nella consegna, quindi sia le vere e proprie deviazioni dalla norma linguistica (ad esempio ortografiche come *o comprato* senza *h*), sia le incoerenze stilistiche (ad esempio un lessico inadeguato come *il sapere del docente* trasmigra e si annida *nelle menti dei suoi alunni* in luogo di un più probabile *si sposta e si colloca*).

Tra i fenomeni annotati ci sono insomma sia elementi che nella percezione del lettore comune (cioè un lettore senza una competenza sociolinguistica e variazionista) potrebbero facilmente essere definiti “errori”, sia elementi stilistici in diversa misura già integrati nella lingua scritta comune, ma che un testo accademico sorvegliato difficilmente presenterebbe (tra questi vi sono persino piccoli scarti legati alle convenzioni grafiche su cui si è adottata una scelta inevitabilmente arbitraria: ad esempio, *Didattica a Distanza* invece di *Didattica a distanza*, *Didattica A Distanza*, *didattica a distanza*, oppure *online*, invece di *on-line* e *on line*). Si è cioè scelto di annotare tutti quei fenomeni notevoli relativamente al contesto comunicativo della prova. Per evitare equivoci nell’interpretazione dei dati e dare conto dell’eterogeneità dei fenomeni a cui rimandiamo, useremo per tutti i fenomeni segnalati nei testi del corpus l’etichetta di *annotazioni*. Va da sé che, per il linguista, alcune di queste annotazioni rivelano che in alcuni ambiti la norma, come avviene sempre in tutte le lingue vive, sta mutando in maniera significativa.

Al termine del procedimento è stato creato un dataset in cui ciascun testo (identificato tramite un codice numerico) è stato associato alle informazioni sull’università e il corso di studio dello scrivente, ai rispettivi dati del questionario sociobiografico, alle prime misurazioni quantitative ottenute tramite READ-IT e al conteggio delle annotazioni per ciascuna categoria. Questo dataset è stata la base per tutte le elaborazioni statistiche prodotte nel progetto.

I testi raccolti, annotati secondo i parametri appena descritti e accompagnati da un ampio corredo di metadati (ricavati attraverso la somministrazione

del questionario), sono confluiti nel corpus UniversITA<sup>6</sup>, interrogabile liberamente sulla piattaforma *NoSketch Engine*<sup>7</sup> utilizzando diversi filtri di ricerca, per i quali si rinvia al vademecum pubblicato su sito del progetto<sup>8</sup>.

## 2.2. I testi non *ad hoc*

Nella seconda fase del progetto è stato raccolto un grande numero di testi prodotti da s&s indipendentemente dalle finalità del progetto: articoli di giornali universitari, tesi e tesine.

Questi testi sono confluiti in due corpora. Il primo<sup>9</sup> contiene 1.630 testi di tipo giornalistico, tratti da giornali universitari, blog ecc., per un totale di 1.484.575 parole. Per tale risorsa, che non costituisce un corpus bilanciato, sono disponibili diversi metadati per contestualizzare le future indagini: argomento del testo, anno di redazione, collocazione geografica dell’ateneo. La seconda risorsa<sup>10</sup> è, come tipo testuale, decisamente la più rappresentativa della scrittura di s&s ed è costituita da un numero non elevato (773 testi) di tesi e tesine, tutte non corrette dai docenti, che equivalgono però a un considerevole numero di parole (5.484.429). È possibile consultare tale sotto-corpus sia in modalità non bilanciata sia, con alcune ovvie riduzioni, in modalità bilanciata<sup>11</sup>: in

6 Il corpus è stato realizzato da Nicola Grandi, Silvia Ballarè, Francesca Chiusaroli, Francesca Gallina, Matteo Pascoli, Elena Pistolesi (Corpus UniverS-ITA.2023).

7 <<https://corpora.ficlit.unibo.it/CUSP/crystal/index.html#open>>.

8 <<https://corpora.ficlit.unibo.it/CUSP/vademecum.pdf>>.

9 Il corpus è stato realizzato da Nicola Grandi, Silvia Ballarè, Francesca Chiusaroli, Francesca Gallina, Matteo Pascoli, Elena Pistolesi (Corpus UniverS-ITA-ProGior.2023).

10 Il corpus è stato realizzato da Nicola Grandi, Silvia Ballarè, Francesca Chiusaroli, Francesca Gallina, Matteo Pascoli, Elena Pistolesi (Corpus UniverS-ITA-ProUniv.2023).

11 In modalità non bilanciata, si ha accesso a una sottoparte della risorsa, selezionata in modo tale da essere rappresentativa della popolazione universitaria italiana in relazione ai medesimi parametri adottati per il campionamento dei testi raccolti nella prima fase (ovvero, collocazione geografica dell’ateneo e area disciplinari del corso di laurea frequentato dallo scrivente).

questo caso sono stati utilizzati gli stessi parametri adottati per il campionamento dei testi raccolti nella prima fase (par. 2.1). Anche in questo corpus sono disponibili diversi metadati: genere, regione di nascita e corso di laurea dello scrivente, anno di redazione, collocazione geografica dell'ateneo.

Anche questi due corpora sono consultabili a libero accesso sulla piattaforma *NoSketch Engine* (cfr. n. 7). Per i dettagli si rinvia di nuovo al vademecum pubblicato sul sito del progetto (cfr. n. 8).

### 3. Primi risultati dell'analisi

Data la natura delle annotazioni che sono state operate sul corpus *Univers-ITA* (cfr. par. 2.1), i loro numeri assoluti non possono in alcun modo comprovare (o, se è per questo, confutare) l'ipotesi, talvolta ventilata in passato pur senza alcuno studio scientifico a sostegno, che s&s di oggi scrivano peggio rispetto al passato o che vi sia un declino nelle abilità d'uso dell'italiano. Solo analisi più minute condotte sul corpus *Univers-ITA* e analisi diacroniche che usino i dati di *Univers-ITA* comparandoli con altri lavori potranno, se mai, dire qualcosa a questo riguardo.

È tuttavia possibile elencare alcuni risultati generali, validi in sincronia, prendendo in considerazione alcune delle variabili indagabili tramite i dati statistici raccolti dal progetto<sup>12</sup>.

#### 3.1. Ricchezza del vocabolario e quantità e qualità delle annotazioni

Iniziamo dal numero di parole diverse usate nei testi prodotti ad hoc, che potrebbe essere un buon indicatore del vocabolario di s&s delle nostre università. Chi studia in un Ateneo del Centro-Sud usa in media ( $\mu$ ) meno parole diverse di chi studia in Atenei del Nord. Lo scarto non è enorme ma tuttavia sostanzioso: s&s del Nord hanno scritto 186,4 parole diverse, mentre s&s del Centro e del Sud si attestano rispettivamente a 179,5 e 178, per una media totale di 181,9. Questo dato resta vero in ogni area disciplinare. Ha invece carattere di forte tendenza l'osservazione che coloro che studiano al Sud impiegano un vocabolario tendenzialmente meno ricco di chi studia nel Centro Italia (unica eccezione: le parole usate da s&s di area scientifica del Sud sono 9 in più rispetto a quelle di s&s della stessa area del Centro). L'area disciplinare che esibisce minore varietà di parole impiegate nei testi è quella sanitaria,

<sup>12</sup> Le elaborazioni statistiche a cui si fa riferimento in questa sezione e il grafico in fig. 4 sono frutto del lavoro di Matteo Farnè (Alma Mater Studiorum – Università di Bologna).

anche se lo scarto con le altre aree non è enorme.

È appena il caso di ricordare che questi dati si riferiscono a testi che da consegna non potevano superare le 500 parole totali (e che ne contengono mediamente circa 380). Il dato appena mostrato dunque non significa che chi studia all'università conosce meno di 200 parole, come pure si è letto su qualche articolo della stampa generalista online (v. ad es. Soncini 2023). Il dato offerto va poi interpretato considerando che le parole vuote, come gli articoli definiti e indefiniti e le preposizioni, e i verbi a più alta frequenza ricorrono, si direbbe necessariamente, più di una volta in qualsiasi testo di media lunghezza: l'impiego di 178-186 parole diverse è perciò un risultato buono, perché indica che la maggior parte delle parole non si ripete.

Il numero dei libri letti in un anno correla positivamente con il numero di parole diverse impiegate: chi dichiara di leggere in un anno dieci libri o più (oltre a quelli richiesti per gli esami universitari) ha usato nel testo prodotto per il progetto  $\mu$  17,48 parole differenti in più rispetto a coloro che hanno dichiarato di non leggerne più di cinque. Quindi, come peraltro almeno in certa misura era da attendersi, le abitudini di lettura di s&s hanno un'influenza positiva sulle loro capacità di scrittura e in particolare sulla ricchezza del lessico. Anche la conoscenza di almeno una lingua antica, ovvero di greco antico e/o latino, corrisponde a una maggiore varietà di parole adoperate, anche se questa variabile è meno potente rispetto a quella del numero di libri letti: chi conosce una lingua antica, infatti, impiega mediamente quasi 6 parole diverse in più di chi non ne conosce (circa 185 contro circa 179 lemmi diversi).

Il grafico in fig. 4 illustra la distribuzione e il numero medio di annotazioni per ciascuna categoria presa in esame:

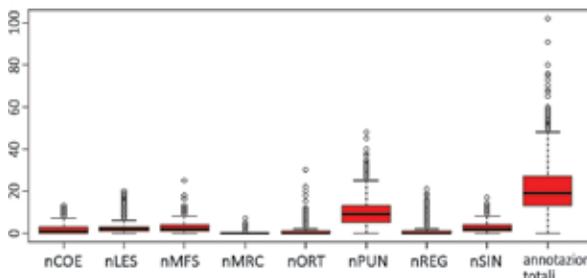


fig. 4. Quadro riassuntivo delle annotazioni.

I testi di area sanitaria sono quelli caratterizzati dal maggior numero di annotazioni ( $\mu$  24,15), mentre il numero minore si riscontra per i testi di area scientifica ( $\mu$  19,63). I testi di area umanistica hanno ricevuto  $\mu$  20,46 annotazioni, quelli di area sociale invece  $\mu$  21,94.

Se si bada alla classe socio-economica di s&s il numero di annotazioni cala costantemente passando dai testi prodotti da chi si dichiara di classe bassa ( $\mu$  24,43 annotazioni) a medio-bassa ( $\mu$  22,65) a medio-alta ( $\mu$  18,96), ma si rialza poi per chi dice di appartenere alla classe alta ( $\mu$  20,62). Si tratta di un dato che desta interesse e che probabilmente meriterebbe discussione e approfondimenti maggiori, che non ci sono possibili in questa sede.

La variabile *diploma di scuola secondaria* sembra creare una netta divisione tra il numero di annotazioni presenti sui testi di chi proviene da un liceo ( $\mu$  20) o da un istituto tecnico ( $\mu$  23) rispetto a quelle presenti sui testi prodotti da chi ha frequentato un istituto professionale ( $\mu$  32,11)<sup>13</sup>. La variabile genere, invece, non incide molto: i testi di studentesse hanno  $\mu$  20,4 annotazioni contro le 22 dei testi di studenti.

Considerando l'eterogeneità, specialmente dal punto di vista sociolinguistico, dei fenomeni annotati, a un primo sguardo sembrerebbe ragionevole ipotizzare che i dati appena discussi siano dovuti principalmente alla concorrenza di due cause: da un lato, la mancata padronanza dello standard (orto grafico e linguistico; dall'altro, una maggiore disinvoltura nell'impiegare forme espressive (o 'devianti'), forse indotta dalla (percepita) maggiore dimestichezza con la lingua.

Come anticipato all'inizio di questo paragrafo, dunque, non è possibile affermare che ci sia stato un peggioramento degli indicatori, soprattutto perché manca un termine di confronto paragonabile a questo progetto. Primi sondaggi relativi ad alcuni ambiti paiono mostrare tuttavia un leggero miglioramento: per quanto riguarda l'ortografia, ad esempio, ambito per il quale possediamo studi pregressi fatti, a partire dagli anni Novanta, su campioni abbastanza ampi (anche se non comparabili con quello di Univers-ITA), il numero di errori nella scrittura delle parole monosillabiche (come *po', sì, dà* dal verbo *dare, qua, sa, sta*, ecc.) resta costante o è in leggera diminuzione rispetto al passato.

È poi evidente che la situazione sociale attuale non è paragonabile a quella dei decenni precedenti in virtù del costante incremento di iscrizioni alla scuola secondaria di secondo grado e all'università, talché è ragionevole pensare che l'università oggi intercetti anche s&s che difficilmente avrebbe intercettato in passato. Di fatto, dunque, il campione

(e conseguentemente il campionamento) scelto per Univers-ITA non è comparabile con nessuno dei campioni selezionati per gli studi precedenti.

### 3.2. Studenti con retroterra familiare plurilingue

Focalizziamo ora l'attenzione sul sotto-corpus di testi prodotti da s&s con retroterra familiare plurilingue, selezionando informanti con almeno un genitore nato all'estero. Non si tratta di s&s stranieri, ma ascrivibili piuttosto alla categoria dei "nuovi italiani" (cfr. Vedovelli 2017; Martari 2024), poiché si è scelto di analizzare unicamente testi prodotti da s&s con almeno un titolo di studio di accesso all'università conseguito in Italia o in una scuola italiana all'estero: 211 informanti (9,9% del collettivo Univers-ITA), di cui 193 (91,5% del totale) scolarizzati interamente in Italia o in scuole italiane all'estero (2 casi). Il retroterra linguistico e familiare è però sensibilmente difforme: infatti 127 di loro provengono da famiglie di origine mista (un genitore nato all'estero) e 84 hanno entrambi i genitori nati all'estero (in 78 casi sono sposati tra connazionali).

Sulla base dei testi prodotti da questo contingente, si è rilevato che i risultati dell'analisi del sotto-corpus sono sempre raffrontabili e talora proporzionalmente simmetrici con quelli del corpus totale, se non altro rispetto alle strategie sintattiche e testuali di connessione e alle caratteristiche del lessico. Ad esempio, la media del rapporto tra numero di parole per testo e fenomeni sintattici annotati (per testo) negli elaborati dell'intero corpus è 0,0061, e, in modo quasi analogo, per i testi del sotto-corpus la media è 0,0069. La frequenza dei principali connettivi (*e, che, o, ma*, ecc.) sembra essere del tutto proporzionale (circa un decimo dei valori di frequenza del corpus complessivo). Per ciò che concerne il lessico, la media del rapporto tra numero di parole e fenomeni annotati nei testi dell'intero corpus è 0,0058, e, con un risultato decisamente comparabile, nei testi del sotto-corpus la media è 0,0067.

Nella prospettiva educativo-linguistica, occorre rilevare, in conclusione, come Univers-ITA contribuisca in ogni caso alla riflessione – di grande interesse per insegnanti e ricercatori – sull'uniformità e la difformità tra le produzioni linguistiche scritte di s&s con retroterra familiare plurilingue.

### 3.3. Scrittura e contemporaneità

Al fine di inquadrare i risultati appena discussi entro lo scenario contemporaneo, è importante fare quantomeno cenno al fatto che cambiamenti

13 Va comunque ricordato che, come si è detto in 1.1, il campione su questo è molto sbilanciato, vista la differenza nettissima tra informanti con diploma liceale e quelli con diploma tecnico. Quasi irrilevante è la percentuale di chi ha un diploma professionale.

extralinguistici (soprattutto tecnologici) avvenuti negli scorsi decenni hanno avuto certamente delle conseguenze sul piano sociolinguistico; essi, forse, possono aiutarci a meglio comprendere la presenza di tratti linguistici normalmente relegati a produzioni meno controllate in contesti formali scritti.

Nella contemporaneità, le occasioni di scrittura quotidiane sono molto frequenti e, certamente, sono aumentate rispetto a un passato non troppo distante (Palermo 2017, Cerruti *et al.* 2011, Tavosanis 2011). Per questo motivo, considerando la dimensione diafasica e diamesica, possiamo dire che il quadro sociolinguistico è certamente più ricco e più complesso rispetto al passato, poiché si è affermata una scrittura informale, attraverso cui si conducono relazioni tra pari. Il rapporto tra gli interlocutori e il loro grado di vicinanza sociale e comunicativa, come è noto (cfr. Halliday 1978), hanno un ruolo cruciale per il realizzarsi di scambi informali, poco controllati, colloquiali. L'italiano colloquiale che, fino a non molto tempo fa, era una varietà linguistica sostanzialmente esclusivamente parlata è, oggi, anche scritta. La scrittura, fino a pochi anni or sono, era praticata quasi esclusivamente in contesti sorvegliati e controllati e non costituiva affatto una pratica quotidiana attraverso la quale mantenere i rapporti personali della vicinanza. Le relazioni tra pari erano dominio incontrastato dell'oralità e la scrittura informale, di fatto, aveva un ruolo fortemente marginale.

Oggi, s&s (e non solo), scrivono decisamente più che in passato, ma lo fanno, soprattutto, in ambito informale e nella comunicazione tra pari (il tema è molto trattato: cfr. ad esempio Martari 2019, Lubello 2017). La lingua utilizzata in questi scambi, almeno in alcuni aspetti, mima l'oralità. Lo scambio è meno pianificato e decisamente più immediato rispetto ai testi scritti prototipici; questo, assieme al rapporto tra parlanti, ha conseguenze su tutti i livelli di strutturazione linguistica di questi testi (si pensi, ad esempio, alla sintassi che risulta essere tipicamente piuttosto frammentata). Negli ultimi decenni, di fatto, è andata costituendosi una sorta di grammatica dello scritto informale e trascurato che fino a non molto tempo fa (quasi) non esisteva.

È importante sottolineare che l'italiano impiegato in questi scambi non è da stigmatizzarsi a priori poiché esso è *adatto* al contesto in cui è realizzato. Come tutte le lingue parlate da ampie comunità, anche l'italiano ha al proprio interno più varietà che presentano caratteristiche strutturali diverse e sono adatte a contesti situazionali diversi. Il parlato informale differisce da quello formale e uno scenario del tutto analogo si ha nello scritto. Non esistono varietà

“giuste” o “sbagliate”, ma parte della competenza linguistica sta nel selezionare la varietà di lingua più adatta al contesto in cui avviene lo scambio comunicativo. Inserire in una tesi di laurea tratti linguistici tipici della messaggistica istantanea sarebbe, infatti, *non adatto* tanto quanto inserire in un messaggio a un amico tratti linguistici esclusivi dell'italiano aulico e formale. Ciò che assume centralità, dunque, è la selezione opportuna della varietà in base al contesto concreto in cui viene prodotto un certo testo. Sarebbe auspicabile che la scuola e l'università dessero a s&s l'opportunità di allenare abilità di scrittura diverse rispetto a quelle che, legittimamente e correttamente, usano nella comunicazione informale. È di massimo rilievo, dunque, educare alla variazione e creare occasioni perché s&s imparino a muoversi tra varietà diverse, al fine di acquisire gli strumenti necessari alla produzione di testi che siano corretti *poiché* adatti al contesto.

#### Riferimenti bibliografici

- Cerruti M., Corino E., Onesti C. (a cura di), 2011, *Formale e informale. La variazione di registro nella comunicazione elettronica*, Roma, Carocci.
- Halliday M.A.K., 1978, *Language as Social Semiotic: The Social Interpretation of Language and Meaning*, Londra, Edward Arnold.
- Lubello S., 2017, *Lo scritto factotum dei nativi digitali*, «Lingue e culture dei media», 1, pp. 143-146.
- Martari Y., 2014, *L'italiano di nuovi italiani*. In Ballarè S., Fiorentini I., Miola E. (a cura di), *Le varietà dell'italiano contemporaneo*, Roma, Carocci, pp. 145-160.
- Martari Y., 2019, “Italiano L2 di italiani? Varietà di apprendimento di scriventi in italiano lingua madre e lingua seconda”. In Carbonara V., Cosenza L., Masillo P., Salvati L., Scibetta A. (a cura di), 2019, *Il parlato e lo scritto: aspetti teorici e didattici*, Pisa, Pacini, pp. 137-150.
- Palermo M., 2017, *Italiano scritto 2.0. - Testi e ipertesti*, Roma, Carocci.
- Soncini G., 2023, “«TL;DR». I lettori annoiati alla terza riga e le quattro lettere simbolo del tempo in cui viviamo”, *Linkiesta*, 2 dicembre 2023. <<https://www.linkiesta.it/2023/12/lettori-pubblico-vocabolario-testo/>> (ultimo accesso: 11.03.2024).
- Tavosanis M., 2011, *L'italiano del web*, Roma, Carocci.
- Vedovelli M. (a cura di), 2017, *L'italiano dei nuovi italiani*, Atti del XIX convegno nazionale GISCEL, Università per stranieri di Siena, 7-9 aprile, 2016, Roma, Aracne.