IL PROGETTO PRIN 2022 PNRR "DIZIONARIO DELL'ITALIANO ACCADEMICO: FORME E FUNZIONI TESTUALI" (DIA): PRIME ACQUISIZIONI E PROSPETTIVE[1]

Davide Mastrantonio[2], *Abdelmagid Sakr*[3], *Michela Dota*[4], *Serena Nardella*[5]

1. INTRODUZIONE

In questo contributo diamo conto dell'esistenza e dei primissimi risultati del progetto PRIN 2022 PNRR "Dizionario dell'italiano accademico: forme e funzioni testuali" (DIA). Il progetto ha il duplice obiettivo di avviare una riflessione sistematica sulla natura del cosiddetto "lessico accademico" e di realizzare un dizionario di parole "accademiche" consultabile in rete in libero accesso, ricercabile sia per forme che per funzioni. Rientrano nel progetto anche finalità didattiche e di terza missione, come formazioni a docenti e attività di divulgazione dei risultati; queste serviranno a far conoscere il Dizionario e a valorizzarne le potenzialità, quale strumento pensato per aumentare la consapevolezza della strutturazione dei testi scritti formali sia in fase ricettiva (lettura, studio) sia in fase produttiva (scrittura).

Nelle pagine che seguono illustriamo le linee portanti del progetto nella loro evoluzione fino a oggi; non diamo conto di tutte le ricerche finora in corso o compiute da tutti i componenti del progetto: molte troveranno spazio in future pubblicazioni.

Il primo nodo da cui discendono gli altri è la formulazione di ipotesi sulla natura del lessico accademico: si tratta naturalmente di ipotesi di lavoro che si affinano col progredire della ricerca (§ 2).

La seconda questione (che solo per ragioni di organizzazione del testo si è deciso di posticipare al § 5) riguarda i parametri usati nella costruzione di un corpus testuale che consenta di ottenere risultati significativi per la ricerca. Come si vedrà meglio, la scelta di limitarsi fin qui ai tre generi dell'articolo di ricerca, del manuale universitario e della voce di enciclopedia dipende dal fatto che due importanti tratti costitutivi del lessico accademico sono il registro neutro o formale e il dominio cognitivo dell'astrazione, tratti ben rappresentati in questi tre generi testuali. Sarà bene notare fin d'ora che il corpus DIA non è l'unico strumento di indagine ma è comunque un riferimento indispensabile per

[1] Il contributo, elaborato nell'ambito del progetto «Dizionario dell'italiano accademico: forme e funzioni testuali (DIA)», Prot. P20222CFEE, finanziato dall'Unione Europea – Next-Generation EU – PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.1 Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN) – CUP N. H53D23009570001(Unità di ricerca dell'Università Ca' Foscari Venezia), CUP N. G53D23007160001 (Unità di ricerca dell'Università degli Studi di Milano), è frutto della discussione e stretta collaborazione degli autori, e si è giovato del confronto con gli altri membri del progetto. A fini di assegnazione formale si devono attribuire a Davide Mastrantonio i §§ 1, 2, 3 e 6; ad Abdelmagid Sakr il § 4; a Michela Dota i §§ 5, 5.1 parr. a e b; a Serena Nardella il par. c del § 5.1. Gli autori sono grati ai revisori per i suggerimenti e i preziosi spunti di riflessione offerti.

[2] Università Ca' Foscari Venezia.
[3] Università Ca' Foscari Venezia.
[4] Università degli Studi di Milano.
[5] Università degli Studi di Milano.

vari scopi: a) leggere in modo estensivo i testi per poter formulare ipotesi sui rapporti tra forme lessicali e funzioni comunicative; b) attingere i passi su cui condurre l'analisi qualitativa dei singoli rapporti forma-funzione e da cui selezionare gli esempi da citare nelle future voci del Dizionario; c) compiere indagini lessicometriche ed estrarre liste di parole in modo automatizzato, tutte operazioni che consentono di raffinare e migliorare il lemmario.

Il nocciolo della ricerca è rappresentato dal lemmario, o per meglio dire dalla struttura concettuale che è alle sue spalle. Nel § 3 presentiamo una struttura provvisoria, cioè un elenco e una possibile gerarchizzazione delle funzioni comunicativo-testuali svolte dal lessico accademico. Una specifica riflessione è dedicata al contributo offerto dalle indagini automatizzate e lessicometriche ricavabili dal corpus DIA (§ 4). Nelle conclusioni si riassumono brevemente le acquisizioni fin qui raggiunte e si tratteggiano alcune previsioni di ricerca nel breve e lungo periodo (§ 6). In fondo al contributo c'è un'appendice con il riferimento a tutti i testi inclusi nel corpus.

Qualche parola sugli aspetti amministrativi, sulla composizione del gruppo di ricerca e sul cronoprogramma.

L'avvio formale del progetto risale alla fine del novembre 2023 ma, dati i tempi richiesti dai reclutamenti, il primo gruppo di ricerca completo si è formato solo nella primavera del 2024. Al progetto collaborano tre unità di ricerca: Ca' Foscari (coordinata da Davide Mastrantonio, che è anche PI), Università Statale di Milano (coordinata da Michela Dota, vice PI) e Università per Stranieri di Siena (coordinata da Eugenio Salvatore). Prendono inoltre parte al progetto Daniele Baglioni, Doriana Cimmino, Alessio Cotugno, Giovanni Piantanida, Jessica Puliero e Abdelmagid Sakr per Venezia; Alessandro Canazza, Alice Martignoni (la quale ha lasciato il progetto nel settembre 2024), Serena Nardella, Giuseppe Polimeni e Massimo Prada per Milano; Sara Bottana e Massimo Palermo per Siena. Nel settembre del 2024 si è tenuto un primo seminario intermedio che ha coinvolto i componenti del progetto e alcuni *discussant*[6]; un secondo seminario è programmato a Siena per il gennaio del 2025. Il convegno conclusivo si terrà a Venezia a settembre del 2025 e anticiperà di poco la fine del progetto, prevista per il novembre successivo.

2. CHE COS'È IL LESSICO ACCADEMICO?

Lo studio del lessico accademico presuppone una delimitazione del campo e la formulazione di ipotesi sulla consistenza dell'oggetto di ricerca, ipotesi strettamente collegate con la costruzione del corpus e con la messa a punto di un lemmario[7].

Sulla scia di studi relativi all'inglese, anche per l'italiano ci si è accorti che nei testi orientati alla produzione e trasmissione della conoscenza (come tipicamente i generi testuali della comunicazione specialistica e didattica) esiste uno strato lessicale formato da parole trasversali rispetto ai settori[8] (dunque non tecnicismi) che servono a organizzare il discorso e le relazioni concettuali. Questo approccio ha portato alla messa a punto di liste

[6] Per la precisione i *discussant* erano quattro, due interni al progetto (Massimo Palermo e Massimo Prada) e due esterni (Emilia M. Calaresu e Carla Marello); a tutti siamo molto grati per le utilissime osservazioni.

[7] Non c'è lo spazio per discutere qui alcune questioni molto importanti, e cioè il rapporto che il "lessico accademico" intrattiene da un lato con la nozione di "discorso accademico" (cfr. ad esempio Fiorentino, 2015; Zaleska, 2016), dall'altro con quella di "lessico intellettuale" (cfr. Serianni, 2019; Gregory, 1985). Per altri aspetti che qui sono invece solo rapidamente accennati, cfr. Mastrantonio (2021; 2022; 2023a).

[8] Sarebbe più prudente dire "tendenzialmente trasversali" perché bisogna tener conto di tradizioni discorsive più o meno ristrette legate alle comunità discorsive, le quali possono coincidere con gli ambiti disciplinari o anche con una singola scuola all'interno di una disciplina. La nozione di "settore" sarà declinata in modo più specifico nel § 5: aree CUN, macro-aree disciplinari, settori scientifico-disciplinari.

di parole che ricorrono con regolarità nei testi formali indipendentemente dal settore, e che hanno un ruolo determinante nei processi di ricezione e produzione dei testi, la cui conoscenza influisce dunque sul successo o insuccesso scolastico[9]. Per l'italiano abbiamo le liste di Ferreri (2005), Spina (2010) e D'Aguanno (2019).

Va in realtà detto che la percezione di uno strato lessicale dotato di una sua omogeneità era già presente, benché non valorizzata, nella tradizione italiana di studi sui linguaggi settoriali; se prendiamo il principale lavoro di riferimento in questo campo, Gualdo, Telve (2011), vediamo citate liste di verbi ricorrenti nei testi specialistici (*avvenire*, *comportare*, *confermare*, *costituire*, *consistere*, *dipendere*, *diventare*, *esistere*, *essere*, *illuminare*, *indicare*, *rappresentare*, *realizzare*, *riferirsi*, *significare*, *stare*, *suggerire*, *svolgere*, *trovarsi*) ma, sulla scia di una tradizione che risale almeno a Maria Luisa Altieri Biagi, rubricati come fenomeno «di perdita d'importanza, o depotenziamento, del verbo» (ivi: 211).

Il lessico "accademico" (Spina, 2010; D'Aguanno, 2019) o "della conoscenza" (Ferreri, 2005) è stato definito come l'insieme di parole «di registro formale, appartenenti al lessico astratto, che permettono di realizzare diverse funzioni comunicative nella scrittura complessa relativa ad argomenti diversi» (D'Aguanno, 2019, Appendice: 45). Da questa definizione ricaviamo come pertinenti i seguenti parametri: la componente sociolinguistica («registro formale» e «argomenti diversi», dunque sono pertinenti solo le parole trasversali, non marcate dal punto di vista dei settori); il tipo di contenuto proposizionale e il suo modo di organizzazione («lessico astratto», di cui diremo meglio alla fine di questo paragrafo); la componente pragmatico-testuale («realizzare diverse funzioni comunicative»).

Rispetto agli approcci fin qui adottati, possiamo mettere a fuoco varie novità che stanno emergendo dal progetto DIA e dagli studi che lo hanno preparato, relative ai seguenti punti: (a) problema del rapporto tra forme lessicali e funzioni comunicative; (b) ruolo della sintassi; (c) esplicitezza e ricchezza di codifica; (d) frame e campi semantici; (e) problemi di tipologia testuale.

(a) *rapporto tra forme lessicali e funzioni comunicative*

Fino ad oggi il problema del lessico accademico è stato posto soprattutto nei termini di liste di parole da conoscere per frequentare la scuola e l'università[10]. Le liste sono senz'altro utili, ma non sono fruibili fintanto che non vengano chiarite le funzioni comunicative che le varie espressioni svolgono, attraverso analisi che partano dai testi e che adottino un'ottica funzionalista (cfr. § 4). Il presupposto del DIA è proprio che le parole del lessico accademico siano ordinabili e aggregabili in modo sistematico su base funzionale, al netto ovviamente dei problemi legati alla polisemia e alla polifunzionalità (per cui cfr. § 3)[11].

[9] L'idea di una varietà accademica da rendere oggetto di insegnamento risale alla distinzione tra BICS (*Basic Interpersonal Communicative Skills*) e CALP (*Cognitive/Academic Language Proficiency*), elaborata da Cummins (1979) per spiegare le cause dell'insuccesso scolastico che affliggeva studenti apparentemente solidi nei domini della comunicazione per scopi generali. Si noti peraltro che "accademico" è usato proprio come calco sull'ingl. *academic*, «relating to schools, colleges, and universities, or connected with studying and thinking, not with practical skills» (*Cambridge Dictionary* online), che è un significato diverso e più ampio rispetto a *accademico* 'universitario'.

[10] D'Aguanno (2019a; 2019b) propone un raggruppamento delle parole accademiche per funzioni, non derivate queste ultime da un'analisi testuale bensì riprese principalmente dai lavori sull'inglese accademico (ad esempio McCarthy, O'Dell, 2008).

[11] Ferreri (2005: 135) propone invece, ma forse a fini meramente esemplificativi e didattici, «aggregazioni di parole» accademiche da intendere come «abbozzi di organizzazioni, parziali e incomplete, la cui unica ragion d'essere è quella di dare il senso della trasversalità di questa porzione di lessico [cioè il lessico accademico] e attivare l'ideazione di possibili percorsi inter- e pluridisciplinari».

In relazione alle funzioni comunicative, è fondamentale osservare che esse non sono esclusive dei testi formali, ma compaiono anche nella comunicazione per scopi generali e riferita a operazioni quotidiane. Questa idea è verificabile empiricamente grazie a esempi come i seguenti[12], nei quali si riportano varie espressioni che codificano le relazioni logico-argomentative di "aggiunta" (1) e di "esemplificazione" (2)[13]:

1)

a) […] l'esposizione [...] agli oppiacei provoca assuefazione alla sostanza e quindi suscettibilità alla sindrome d'astinenza. Gli oppiacei possono, **inoltre**, indurre, con probabilità variabile a seconda dell'oppiaceo, una condizione di tossicodipendenza (*Fondamenti di tossicodipendenza da oppiacei*).

b) Gli effetti [della cocaina] **non si limitano a**lla depressione […] **ma** provocano anche problemi fisici non da poco (*Io donna*, 8 ottobre 2019).

c) [l'epilessia] limita molto la vita dei pazienti, perché **a**gli aspetti sanitari **si associa** ancora un forte pregiudizio sociale (*Io donna*, 8 ottobre 2019).

d) [Le donne] Diventano tossicodipendenti più tardi degli uomini, ma poi consumano quantità maggiori, sono più restie a chiedere aiuto e "resistono" agli interventi di recupero, faticando di più per uscire dal tunnel. **Come se non bastasse**, la società è meno disposta a "perdonare" una tossicodipendenza alle donne rispetto agli uomini. (*Io donna*, 8 ottobre 2019).

2)

a) Il deficit sensoriale può causare un deficit della stimolazione cognitiva. **Si pensi a**gli anziani che non possono più leggere il giornale o guardare la televisione per un deficit visivo (Arcuri, Arcuri, *Manuale di sociologia*, 2010).

b) Al cafè Marrakech trovi i piatti tipici della cucina marocchina, **come** il cous cous (Bettinelli, Della Putta, Visigalli, *Buona idea. Corso di lingua e cultura italiana*, livello A1, 2020).

c) cioe' gli ho sempre detto a guido tutto a posto si chiacchiera // pero' lui sparisce non si fa sentire // e poi lo ribecchi magari una sera sbronzo che ti fa il molesto // e non sai che cazzo succede // mh // **tipo** una volta // cena a casa di da e di lodo // tu non c' eri c' era ada pero' // era una bella serata avevamo organizzato io lodo ada guido c' erano anche delle altre tipe cosi' // eh cosa // nel senso no lui e' il mio tipo // em // e niente e e lui e' arrivato tutto a posto mentre noi cucinavamo ha iniziato a bere ma ha bevuto neanche tanto // una birra pero' mi sa che aveva gia' bevuto a casa // e poi e' crollato // ad addormentare non ha mangiato un cazzo (conversazione, ricavata dal corpus KIParla).

I passi in (1-2) non provengono unicamente dal corpus DIA ma sono ricavati anche da altri testi, vari per canale, genere, registro: un report scientifico (1a), un manuale universitario (2a), un racconto orale informale (2c), una recensione presente in un manuale di italiano L2 di livello A1 (2b), un articolo di un settimanale (1b-d). Come anticipato, ciò che si nota è che gli esempi citati usano risorse linguistiche diverse per esprimere rispettivamente gli stessi valori semantici, cioè la relazione di aggiunta (1) e la relazione di esemplificazione (2). Tra le differenze presenti, spiccano quelle di registro: *tipo* 'per

[12] Peraltro, l'idea che nella vita di tutti i giorni si compiono operazioni logiche del tutto analoghe a quelle della ricerca scientifica è confortata dai filosofi della scienza, cfr. ad esempio Kosso (2011), cit. in Mastrantonio (2023a: 106).

[13] Sulle relazioni logico-argomentative cfr. Ferrari (2014: 131-177).

esempio' e *come se non bastasse* 'inoltre' (con connotazione negativa) non sarebbero pensabili in un testo formale; *come* 'per esempio' e *inoltre* appaiono neutre dal punto di vista del registro; *si pensi a* 'come esempio prendiamo' e *si associa a* ('ci sono gli aspetti sanitari e in aggiunta c'è un forte pregiudizio') difficilmente vengono impiegati nel registro colloquiale.

Simili casi di variazioni mostrano da un lato che le funzioni comunicative non sono appannaggio delle pratiche discorsive formali ma sono trasversali a generi e registri; ma al tempo stesso mostrano che il registro è un parametro pertinente e molto utile per l'individuazione e l'inquadramento del nostro lessico accademico, il quale sembrerebbe essere costituito dalle realizzazioni diafasicamente formali o neutre delle funzioni comunicative (mentre restano escluse quelle basse). Questa circostanza, si noti, è alla base della decisione di limitare il corpus solamente ad alcuni generi di registro formale, almeno per il momento. Ma le analisi in corso di Eugenio Salvatore sui rapporti fra lessico accademico e lingua dei giornali potranno offrire una prospettiva aggiornata su questo frangente.

(b) *ruolo della sintassi*

Tra gli aspetti pertinenti al lessico accademico finora non o scarsamente evidenziati c'è la sintassi. Già in (1c) abbiamo visto che la relazione di aggiunta non occupa una posizione inter-frasale (che è tipica dei connettivi, come *inoltre*) ma si realizza nel sintagma verbale, dunque nel cuore della frase (*si associa a*). Lo stesso fenomeno si manifesta anche nella codifica delle relazioni di causa-effetto (*perché* > *essere determinato da*)[14] o di conclusione logica (*quindi* > *suggerire*), per cui abbiamo coppie come le seguenti:

3)
a) Il treno è arrivato tardi **perché** c'era un guasto alla linea elettrica.
b) Il ritardo del treno **è stato determinato da** un guasto alla linea elettrica.

4)
a) le due tavole hanno le stesse dimensioni, **quindi** dovevano far parte di un unico progetto.
b) le analoghe dimensioni delle due tavole **suggeriscono** che fossero parte di un unico progetto.

Sia *determinare* sia *suggerire* sono parole ben rappresentate nei testi del nostro corpus (*suggerire* è presente anche nell'elenco di Gualdo, Telve, 2011 citato sopra). Il fenomeno qui osservabile, al netto di una variazione di registro, è connesso con le nominalizzazioni: i processi vengono codificati a livello del sintagma nominale (o di una proposizione completiva) e le relazioni logiche prendono il posto del verbo. Si tratta di un meccanismo che Halliday (2004) classifica come "metafora grammaticale"[15].

(c) *esplicitezza e ricchezza di codifica*

Un ulteriore aspetto pertinente sembra essere l'esplicitazione e la ricchezza di codifica, che si associa alla nozione di "vincolo comunicativo" (Sabatini, 1999); ci riferiamo al fatto che i testi concettuali hanno la necessità di esplicitare maggiormente le informazioni, trattandosi di testi settoriali e che parlano di astrazioni; si può aggiungere il fatto che l'emittente e il ricevente non condividono il contesto della comunicazione (ma questo vale

[14] Dei rapporti di causa-effetto si sta occupando Sara Bottana dell'unità di Siena.
[15] Dal punto di vista della profondità storica, questo tipo di compattamento frasale sembra essere disponibile da sempre in italiano, ma si è reso via via più accessibile con il maggior peso assunto dai linguaggi scientifici, cfr. Mastrantonio (2021b: 235-259). Per le categorie di disponibilità/accessibilità cfr. Palermo (2021: 34-35).

solo per il canale scritto). Una coppia di esempi che aiuta a cogliere questa dimensione è la seguente (il secondo passo è ripreso da una voce dell'*Enciclopedia Treccani* online con un leggero adattamento; il primo è una riformulazione):

5)
a) I bonobo **hanno** le gambe più lunghe di altre scimmie.
b) La specie dei bonobo **si caratterizza** rispetto alle altre **per** la maggiore lunghezza degli arti inferiori.

I due passi, sinonimi, realizzano le funzioni di "indicare proprietà" e "confrontare" due entità del mondo (bonobo vs. altre scimmie). La funzione dell'indicare proprietà è codificata rispettivamente da *hanno* e da *si caratterizza per*, dove, registro a parte, l'opposizione fra il primo e il secondo termine risiede nel rango di frequenza (il primo, marcato FO, è molto più frequente del secondo, CO)[16] e nella ricchezza di codifica (*avere* è più povero, essendo generico e polisemico, mentre *caratterizzarsi per* è molto specifico).

(d) *frame e campi semantici*

Gli aspetti discussi fin qui, registro, sintassi, esplicitezza, sono indipendenti dalla semantica lessicale; ma per indagare la natura del lessico accademico è utile tener conto anche del significato delle parole. Non sembra un caso che nei testi concettuali alcune aree semantiche appaiono molto sollecitate; è il caso del campo della vista, rappresentato da parole come *aspetto*, *assistere a*, *dal punto di vista*, *emergere*, *notare*, *osservare*, *riguardare*, *vedere* e molte altre (gli esempi seguenti sono tutti ricavati dal corpus DIA):

6)
a) Affrontiamo ora un **aspetto** insondato nella storia dell'inquisizione.
b) Nel tardo Jomon (3500-3000 anni fa) **si assiste ad** un drastico aumento degli insediamenti costieri.
c) **Dal punto di vista** formale, chiarezza, completezza e congruità sono i requisiti irrinunciabili di una buona documentazione.
d) Dalla lettura dei contributi che seguono **emerge un quadro** alquanto composito e frammentato dei processi partecipativi.
e) È da **notare** come il testo non faccia cenno a quella parte di tradizione aristotelica dedicata all'etica.
f) come **è stato** ormai **osservato** e dimostrato da più parti, la lingua comune si nutre principalmente di linguaggi specialistici.
g) Le discipline che **riguardano** l'analisi dei dati e le applicazioni della statistica alla psicologia sono in Italia raggruppate nel gruppo disciplinare M.PSI/03.
h) **Vediamo** adesso il primo passo da compiere nella fase di controllo: il Process Management Chart

[16] Con FO sono marcate le parole "fondamentali", cioè le circa 2.000 parole più frequenti dell'italiano, presenti in una larghissima fetta di testi e discorsi e note a tutti i parlanti madrelingua senza distinzioni socioculturali; CO indica invece il vocabolario "comune", cioè le ca. 45.000 parole che hanno un rango di frequenza più basso, non appartengono ai lessici tecnici (in tal senso "comuni", cioè indipendenti dalla professione o dalla provenienza regionale) e sono note solo ai parlanti che hanno un livello di istruzione medio o superiore (cfr. De Mauro, 1999).

Per spiegare la presenza di questi termini dobbiamo pensare che la produzione e la trasmissione della conoscenza implicano un continuo ricorso alle procedure di osservazione. Gli studi di filosofia della scienza fanno molta attenzione a questo aspetto; ad esempio Feyerabend (1979) sottolinea la necessità di vigilare sul «linguaggio in cui descriviamo le nostre osservazioni» (ivi: 67) e «[su]i termini nei quali [le] esprimiamo abitualmente» (ivi: 57).

Alcune di queste forme sono ordinabili in *frame* o paradigmi coerenti. Ad esempio la coppia *guardare* vs. *vedere* articola il processo della percezione visiva (da intendere anche in senso metaforico) secondo il polo [± intenzionale]; sotto questo profilo, come *guardare* si comportano anche *analizzare*, *osservare*, *prendere in considerazione*, *prestare attenzione a*, *studiare*; come *vedere* si comportano *emergere*, *individuare*, *notare*; si tratta delle due funzioni che, come si vedrà nel § 3, chiamiamo "portare l'attenzione su qualcosa" e "notare un elemento nuovo" (su questo punto cfr. già Mastrantonio, 2023a: 111-112).

In altri casi il lessico della percezione visiva svolge altre funzioni: ad esempio *vediamo* in (6h) svolge una funzione demarcativa e lavora in regime di deissi testuale (vale cioè 'nella parte seguente di testo'). Bisogna mettere in conto anche le risemantizzazioni; la semantica del "vedere" appare decisamente indebolita in una parola come *aspetto* (6a) nel significato astratto di 'lato di una questione' (o 'questione' *tout court*); questa accezione costituisce un'evoluzione semantica del significato visivo «ciò che appare, che si presenta agli occhi, e il modo con cui si presenta» (ad esempio *l'aspetto del mare*, Vocabolario Treccani online, s.v. *aspetto*, 2.); ma sincronicamente i due significati sono ormai ben separati, e questa separazione andrebbe meglio sottolineata dai dizionari stessi.

(e) *problemi di tipologia testuale*.

C'è infine un'altra questione molto importante da considerare, fin qui data per scontata, relativa al dominio cognitivo dell'astrazione. Un gruppo di funzioni comunicative molto ben rappresentate nei testi del corpus DIA riguarda le operazioni del confrontare e classificare, che implicano il notare analogie e differenze. Nella Tabella 1 si presentano le funzioni comunicative in questione, seguite da formule che descrivono le proposizioni semantiche collegate alle funzioni stesse e dalle corrispettive forme lessicali (per come elaborate nel lemmario provvisorio, su cui cfr. § 3):

Tabella 1. *Alcune funzioni comunicative e il rispettivo lessico*

funzione comunicativa	**proposizione semantica**	**forme lessicali**
indicare analogie	x è analogo a y; z si trova sia in x sia in y	*accomunare*; *analogia*; *analogo*; *condiviso*; *comune*; *simile*; *(lo) stesso*
indicare differenze	x è diverso da y	*al contrario (di)*; *differenza*; *differenziarsi da*; *distintivo*; *diversamente da*; *diverso*; *invece*; *mentre*
indicare proprietà e classificare	x è qualità di y	*aspetto*; *carattere*; *caratteristica*; *contraddistinguere*; *criterio*; *distinguere/-si (da, per, tra)*; *elemento*; *essere caratterizzato da*; *parametro*; *presentare* ('avere'); *proprietà*; *qualità*; *tratto* ('proprietà caratteristica'); *suddividere*; *tipo*

Questo gruppo di funzioni rimane fuori dalle concettualizzazioni prodotte dalla linguistica del testo o dalla pragmatica, filoni invece estremamente utili per evidenziare altri tipi di funzioni (sopra abbiamo commentato relazioni logico-argomentative quali quelle di aggiunta, causa-effetto, conclusione logica, esemplificazione). Per poter lavorare su categorie comparabili bisogna spostarsi nella linguistica cognitiva; ad esempio Langaker (2013: 4) scrive che «at the core of grammatical meaning are mental operations inherent in these elemental components of moment-to-moment living», e offre un inventario di concetti basilari che mettono in rapporto abilità generali e cognizione: *focus and shift attention*; *track a moving object* (questi due sono utili a spiegare la distinzione vista sopra tra *guardare* e *vedere*); *form and manipulate images*; *compare two experiences*; *establish correspondences* (queste ultime due sono pertinenti al confrontare e classificare); *combine single elements into complex structures* ecc.

Per quel che riguarda la tipologia testuale, un possibile ma finora tenue aggancio è rappresentato da una proposta di classificazione dei testi che si deve a Hatim (1984) poi ripresa da Mazzoleni (2004), rimasta finora al margine dell'attenzione degli studiosi. Mazzoleni formula una tipizzazione di testi in cui le nozioni di "scopo comunicativo" e di "forma testuale" vengono tenute distinte. Rispetto alle ben note categorie di Werlich (1975), secondo Mazzoleni solo esporre, argomentare e prescrivere vanno intesi come scopi comunicativi; invece narrare e descrivere sarebbero forme testuali; a queste due ultime forme ne viene aggiunta una terza, rappresentata dai testi concettuali. Se ne ricava pertanto il seguente schema (Tabella 2):

Tabella 2. *Schema basato sulla tipologia testuale di Mazzoleni (2004)*

Scopo	**Forma testuale**
Espositivo	Narrativo Descrittivo Concettuale
Argomentativo	Esplicito Implicito
Prescrittivo	Nomativo Istruzionale

Concentriamoci su quella che Mazzoleni chiama «forma testuale» (che si potrebbe forse chiamare "tipo di contenuto proposizionale e suo modo di organizzazione sulla base del dominio cognitivo")[17] e in particolare sul «sottotipo concettuale»:

[17] Per "dominio cognitivo" si intendono i gruppi di abilità cognitive coinvolte nella comprensione di diversi tipi di rappresentazioni linguistiche e concettuali. La distinzione tra modo narrativo, descrittivo e concettuale interessa infatti i tipi di referenti, i tipi di processi e i tipi di relazioni logiche che sono codificate nel testo. Nel modo narrativo le componenti cognitive maggiormente sollecitate sono quelle di tempo, azione, motivo; nel modo descrittivo prevale lo spazio; alla base del modo concettuale ci sono il confronto, la classificazione, la definizione delle entità e altre componenti, ma si tratta di un terreno ancora da esplorare. Per una tipologia incentrata sull'elaborazione dell'esperienza alla base dell'ideazione cfr. Palermo (2013: 242-243) che si rifà a sua volta a Della Casa (1994).

> in un testo di sottotipo descrittivo viene presentata la configurazione spaziale, statica, di elementi/oggetti; un testo di sottotipo narrativo prevede personaggi coinvolti in eventi che si svolgono in una dinamica temporale; la caratteristica fondamentale di un testo di sottotipo concettuale sta invece nello statuto ontologico di ciò di cui si parla – non oggetti o elementi concreti, individuali, ma astrazioni, categorie, concetti appunto, connessi da relazioni logico-semantiche generali (Mazzoleni, 2004: 404)

Delle caratteristiche del testo concettuale (come del resto degli altri sottotipi: i quali però sono molto meglio descritti) non si dice molto: «astrazioni, categorie, concetti [...] connessi da relazioni logico-semantiche generali»; ma il riconoscere un (sotto)tipo testuale autonomo sembra essere di non poco aiuto nel dar conto almeno di una parte del lessico accademico, quella che si riferisce sistematicamente alle operazioni del confrontare e classificare. Si tratta di un lavoro da affrontare pazientemente; per ora possiamo limitarci a ipotizzare che una parte del lessico accademico sia spia di un determinato tipo di contenuto e del suo modo di organizzazione (tipo concettuale), oltre che di determinati macro-atti pragmatico-testuali (scopo argomentativo ed espositivo).

2.1. *Qualche riflessione conclusiva*

Le considerazioni fin qui svolte mostrano la complessità e la modularità del nostro oggetto di ricerca, al punto che ci si domanda se in futuro l'etichetta stessa di "accademico" dovrà essere superata, almeno per quel che riguarda l'inquadramento teorico del lessico da noi studiato. Per ciò che concerne invece l'obiettivo pratico, cioè la creazione del Dizionario, l'etichetta "accademico" appare pienamente legittimata nel suo significato educativo, dal momento che il Dizionario è anzitutto pensato per essere impiegato a scuola e all'università. Del resto, sembrano altrettanto (anzi più) problematiche possibili etichette alternative come quella di "lessico colto" o di "lessico intellettuale", che contengono al loro interno sezioni di lessico molto più ampie, rispetto a cui il lessico accademico come da noi inteso costituirebbe solo una piccola fetta (sul "lessico intellettuale europeo" cfr. Gregory, 1985).

Al netto di tutto ciò, appare per il momento utile provare a elencare, sia pure in via provvisoria, le caratteristiche delle parole accademiche:

- quanto al campo, sono parole trasversali ai settori;
- quanto al registro, sono neutre o formali;
- quanto al tipo di testo o discorso, e precisamente in riferimento al tipo di contenuto e alla sua organizzazione in base al dominio cognitivo, sono usate per organizzare discorsi concettuali, cioè discorsi astratti, orientati alla funzione referenziale e dotati di un alto grado di esplicitezza;
- quanto al contenuto semantico, sono in sé parole non referenziali che esprimono funzioni comunicative presenti anche nella comunicazione per scopi generali;
- quanto alla sintassi, in certi casi mettono a disposizione dei parlanti/scriventi risorse sintattiche particolari e alternative alla comunicazione ordinaria.

Il concetto di "neutralità dal punto di vista del registro" serve a dare conto dei molti casi in cui il lessico accademico si serve di forme del vocabolario di base e proprie anche delle pratiche comunicative ordinarie; esempi lampanti sono il verbo *essere* (quando è sinonimo di *consistere in* o *rappresentare*, nella funzione di "stabilire identità e definire")[18] o

[18] Questo aspetto è trattato in Mastrantonio (2021a: 358).

la congiunzione *ma* (molto sfruttata nei testi concettuali sia per le relazioni logiche di sostituzione e contrasto controaspettativo, sia per agevolare il cambio di tema). Se il lessico accademico può essere inteso come una varietà di lingua, è una varietà che presenta forti punti di tangenza con la lingua per scopi generali.

Questa linea è alternativa a quella di Spina (2010), che nell'estrazione automatizzata della sua lista elimina meccanicamente tutti i lessemi che fanno parte del vocabolario di base. Si tratta di un approccio potenzialmente problematico, perché le proprietà sociolinguistiche non sono sempre associate in modo esclusivo a un lessema, ma possono dipendere dall'uso del lessema stesso[19]; un caso esemplare in questo senso sembra essere il verbo *suggerire* (4), parola del vocabolario fondamentale (in impieghi come *suggerire la risposta*; *ti suggerisco di andare* ecc.) che però assume una fisionomia accademica nell'accezione 'far dedurre, portare a concludere logicamente'[20]. Ma persino la flessione morfologica o la posizione occupata dalla parola nella frase e nel testo possono creare una marca di accademicità e segnalare un uso formale che difficilmente si avrebbe nella comunicazione per scopi generali: esempi di questa fattispecie sono il congiuntivo *si pensi a* (2), spesso usato per indicare la relazione di esemplificazione, e il congiuntivo *si osservi* (7a), sfruttato per richiamare l'attenzione del lettore.

Naturalmente ciò non significa che i dati lessicometrici siano privi di valore, anzi. A proposito della questione se includere nel nostro lemmario parole del vocabolario di base, appare molto interessante l'indagine lessicometrica contrastiva condotta da Abdelmagid Sakr (§ 4): su un campione di 8 parole del lemmario provvisorio appartenenti al vocabolario di base, ben 7 sono nettamente più frequenti nel corpus DIA rispetto ad altri corpora non accademici dell'italiano. Tale scarto quantitativo rafforza l'idea che, se usate nei testi concettuali, le parole di tutti i giorni sono sottoposte a sfruttamenti particolari che possono interessare tanto gli aspetti morfosintattici quanto gli aspetti semantici (tutti aspetti che ovviamente solo un'analisi di dettaglio può rivelare).

3. LA STRUTTURA DEL LEMMARIO, OVVERO LA RETE DELLE FUNZIONI COMUNICATIVE

In questo paragrafo presentiamo una prima ipotesi di struttura del lemmario, messa a punto da Davide Mastrantonio. Definire il lemmario e la sua struttura sono operazioni che procedono di pari passo e pongono vari problemi; non si tratta solo di decidere se includere o escludere volta per volta una determinata parola, ma di stabilire con ragionevole sicurezza quali funzioni attribuirle, tenuto conto di due fatti:

a) allo stato attuale delle conoscenze, l'individuazione delle funzioni non può sempre poggiare su categorie concettuali consolidate (abbiamo evidenziato il caso delle funzioni del confrontare, indicare analogie ecc. che costituiscono un campo ancora da esplorare);
b) le parole accademiche – e si noti qui la divergenza rispetto ai tecnicismi – sono spesso polisemiche, cioè hanno più funzioni comunicative che si escludono nel medesimo enunciato, o polifunzionali, ovvero svolgono cumulativamente nello stesso enunciato più funzioni.

Ecco un esempio di polisemia relativa al verbo *osservare*:

[19] A noi interessano le proprietà di registro, ma il discorso vale anche per i sottocodici: pensiamo alle rideterminazioni di parole della fisica come *forza* e *momento*; si vedano al riguardo le considerazioni di De Mauro (1988, § 4, punto [a]).

[20] Accezione peraltro non registrata nei vocabolari dell'uso, ad esempio il Treccani online, il Nuovo De Mauro online, il Sabatini, Coletti (2018) online, lo Zingarelli (ed. 2008).

7)

a) La reazione iniziale [...] consiste nella cattura di un nucleo di idrogeno da parte di un altro nucleo dello stesso tipo, con la produzione di un nucleo di deuterio, di un positrone [...] e di un neutrino. **Si osservi** che esistono tre tipi di neutrini: l'elettronico, il muonico e il tauonico, associati a tre particelle cariche, rispettivamente l'elettrone, il muone e il tauone.

b) Analizzando invece le riserve aggregate delle banche centrali (Tabella 3), **si osserva** che cinque dei primi dieci paesi appartengono alla comunità BRICS+.

c) Abbiamo invece considerato come oggetti in atto i singoli numeri irrazionali e, **come osservato**, ciascuno di essi richiede, per essere espresso compiutamente, un insieme infinito di cifre.

d) È un tema posto sin dalla prima pagina del volume di Ricciardi, Fiorentino e Picone. Gli Autori **osservano** che il racconto del terremoto pare essere consegnato soltanto al canone giornalistico e, in particolare, alla declinazione dell'inchiesta giudiziaria famosa come Irpiniagate

In (7a), *si osservi* ha la funzione di "richiamare l'attenzione" istituendo una blanda discontinuità testuale ('attenzione: dato il tema del discorso, cioè i neutrini, dovete sapere che ne esistono tre tipi')[21]. Tale funzione è invece assente in (7b), dove *si osserva* si limita a esplicitare e lessicalizzare il procedimento dell'osservazione, in un luogo del testo dove peraltro c'è compresenza di più codici (l'enunciato contiene infatti un riferimento logodeittico a una tabella). In questo secondo caso, la funzione svolta è probabilmente quella che chiamiamo "presentare un fenomeno", funzione spesso lessicalizzata da espressioni come *registrarsi* (*in tutti i paesi si registrano aspri conflitti sociali interni*) o *verificarsi* (*I principali miglioramenti nella produttività si verificarono nel settore agricolo*). In (7c) e (7d) si osservano ancora altre funzioni: in (7c) *come osservato* serve a "rinviare a una parte di testo precedente"; in (7d) *osservano* opera nel piano polifonico-enunciativo e serve a "riferirsi al discorso altrui". Ciò che si osserva è dunque una varietà di funzioni che sembrerebbero escludersi a vicenda, varietà che a sua volta poggia su elementi formali (indicativo vs. congiuntivo, posizione a inizio enunciato o interna) e semantici (solo in [7c-d] *osservare* può essere coerentemente sostituito da un verbo del "dire").

Dal momento che il nostro Dizionario non si accontenta di dare definizioni *ad hoc* o parafrastiche, ma intende dar conto della rete di funzioni sottostante al lessico accademico, appare chiaro che la creazione del lemmario è un'operazione delicata poiché richiede un lavoro qualitativo parallelo sulle forme e sulle funzioni. Per quel che riguarda le funzioni, gli ambiti cui appoggiarsi con più sicurezza sono la pragmatica e la linguistica del testo, ma con un'avvertenza di non poco conto: la linguistica del testo si focalizza su ciò che accade oltre il confine d'enunciato o di unità informativa, dunque fatica a cogliere fenomeni di compattamento frasale come quelli descritti sopra in relazione ai verbi *associarsi a* (1), *determinare* (3), *suggerire* (4). Un filone di studi promettente è quello della *genre analysis* (cfr. Swales, 2004, e per l'italiano Klimová, 2016), tenuto però conto del fatto che questo indirizzo si concentra sulle caratteristiche peculiari dei singoli generi (*moves* e *steps*), mentre le funzioni del lessico accademico sono – almeno nei voti – trasversali ai generi.

Data questa complessità, la prima procedura richiesta è lo spoglio manuale, cioè la lettura estensiva di campioni del corpus DIA e di altri testi dello stesso tenore finalizzata

[21] In un'analisi testuale classica, la relazione che vige tra il primo e il secondo enunciato di (7a) è di "specificazione" (cfr. Ferrari, 2014: 157). Ma tale contenuto semantico non è veicolato dal verbo *si osservi*, che si limita a operare sul piano interazionale (richiamare l'attenzione). Si profila, in casi come questo, un problema più generale, ovvero lo sfruttamento metatestuale di dispositivi che operano primariamente sul piano interazionale.

a individuare lessico pertinente facendo ipotesi sulle funzioni comunicative coinvolte. Come illustrato al § 2, le parole non settoriali e al tempo stesso di registro formale sono spesso dei rivelatori della presenza di una funzione comunicativa che merita di essere indagata. Parallelamente, la creazione del lemmario deve tenere conto del confronto con le liste di parole accademiche, sia quelle preesistenti (cfr. § 2), sia quelle estratte dal corpus DIA, e delle indagini lessicometriche (§ 4). Rispetto alla griglia che si illustrerà tra poco, alcune conferme significative circa i rapporti forma-funzione si avranno solo via via che verranno condotte indagini approfondite su gruppi di lessemi o gruppi di funzioni; attualmente, ad esempio, Doriana Cimmino sta studiando la lessicalizzazione delle funzioni di dominio e di progressione del topic.

Nel mettere a punto la tassonomia seguente si è tenuto conto degli approcci funzionalisti; una nozione utile per pensare il livello gerarchico superiore (i piani del testo) è la tripartizione funzionale di Halliday (*ideational*, *metatextual* e *interpersonal*; cfr. Halliday, 1994), che però non è ripresa pedissequamente. Molte categorie del livello inferiore (le funzioni comunicative) sono desunte dalla linguistica del testo (ad esempio la griglia delle relazioni logico-argomentative, le relazioni di *dispositio* e le relazioni referenziali nel caso in cui siano lessicalizzate), dalla pragmatica e dagli studi sull'enunciazione (la funzione interazionale di richiesta di attenzione, i dispositivi di modulazione della forza illocutiva, gli affioramenti del punto di vista e della prima persona ecc.). Delle categorie della linguistica cognitiva si è detto qualcosa in fondo al § 2. A partire da questi impulsi, ne è finora emersa la seguente macro-strutturazione:

a) il piano dei fenomeni di cui si parla (o del contenuto proposizionale), i quali possono essere menzionati con rilievo maggiore dato all'oggetto osservato (es. *si verifica x*; *si registra x*) o al soggetto osservante (*osserviamo x*; *notiamo x*);
b) il piano del punto di vista, cioè come i diversi osservatori (vale a dire i membri di una comunità scientifica e discorsiva) interpretano i fenomeni in esame;
c) il piano dell'organizzazione del testo, cioè come l'enunciatore progetta il testo e guida il lettore nella sua interpretazione, con attenzione sia alle strategie orientate al testo (o metatestuali), sia alle strategie orientate al destinatario (o interazionali).

I livelli appena illustrati si prestano a essere inseriti discorsivamente all'interno di un macro-frame che è sostanzialmente quello dell'attività scientifica: si delimita un campo dello scibile e lo si osserva cercando di capire come funziona; si offre il corrispettivo linguistico degli oggetti osservati, entità, processi e rapporti logici (fin qui siamo nel piano fenomenico); si sottolinea la propria posizione e ci si confronta con altri studiosi che hanno compiuto analoghe osservazioni o che comunque sono competenti in materia, col potenziale esito di un contrasto o di una convergenza di opinioni (piano del punto di vista); si organizzano le conoscenze in questione in un testo che sia comprensibile al lettore (piano testuale e metatestuale), all'interno di una comunicazione simmetrica (come accade nel testo argomentativo) o asimmetrica (testo espositivo).

A queste macro-categorie corrispondono poi delle categorie di livello inferiore, che nella nostra ipotesi di lavoro coincidono con il concetto di funzione comunicativa; tra parentesi si riportano alcuni dei lessemi (una settantina) che allo stato attuale delle ricerche consideriamo rappresentativi delle varie funzioni (ma il lemmario provvisorio è più ampio):

a) <u>piano dei fenomeni</u>:

- portare l'attenzione su qualcosa (*analizzare*, *prendere in considerazione*);
- notare un elemento nuovo (*emergere*, *individuare*, *notare*);
- presentare un fenomeno (*assistere a*, *osservare*, *registrarsi*, *verificarsi*);
- stabilire identità e definire (*consistere in*, *costituire*, *rappresentare*);

- indicare analogie (*analogo*, *condiviso*);
- indicare differenze (*al contrario*, *differenziarsi da*, *distintivo*);
- indicare proprietà e classificare (*aspetto*, *criterio*, *parametro*, *suddividere*, *tratto*);
- stabilire rapporti generale-particolare (*è il caso di*, *gruppo*, *nello specifico*, *per esempio*, *sottogruppo*, *si pensi a*);
- stabilire cause e conseguenze (*alla luce di*, *causa*, *comportare*, *dal momento che*, *determinare*, *fattore*, *effetto*);
- ricavare conclusioni logiche (*dedurre*, *dunque*, *ricavare*, *suggerire*);
- indicare cambiamenti e sviluppi (*evoluzione*, *progressivo*);
- misurare e quantificare (*aumento*, *diminuzione*, *incremento*).

b) piano del punto di vista:

- riferirsi a discorsi altrui o ad altri testi (*affermare*, *osservare*, *secondo*, *sostenere*, *sottolineare*);
- parlare di contrasti di opinioni (*discutere*, *discussione*, *dibattuto*, *errato*);
- indicare il grado di certezza di un'affermazione (*indubitabile*, *presumibilmente*, *naturalmente*);
- dare una valutazione positiva o negativa di qualcosa (*condivisibile*, *problematico*);
- enfatizzare e mitigare un'affermazione (*credo*, *mi pare*, *notevole*, *significativo*, *se non vado errato*).

c) piano dell'organizzazione del testo:

- mantenere o cambiare il tema del discorso (*a tale riguardo*, *a tale proposito*, *in quest'ottica*, *in relazione a*, *invece*, *trattarsi*);
- rinviare a parti di testo precedenti o successive (*accennare*, *infra*, *menzionare*, *osservare*, *supra*, *vedere*);
- indicare disposizione delle parti (*anzitutto*, *in primo luogo*, *da un lato… dall'altro*, *in aggiunta*);
- richiamare l'attenzione del lettore (*notare*, *osservare*, *considerare*).

4. Le indagini automatizzate e lessicometriche condotte sul corpus DIA

In questa sezione si illustrano due tipi di analisi, a cura di Abdelmagid Sakr, condotte sul corpus DIA per mezzo del software Python, e precisamente: a) l'estrazione automatizzata di una lista di parole trasversali ai settori dunque potenzialmente accademiche; b) l'analisi lessicometrica di un piccolo campione di parole incluse nel lemmario provvisorio. La prima operazione ha come obiettivo immediato quello di raffinare e migliorare il lemmario provvisorio. Quanto alle analisi lessicometriche, esse sono condotte in modo contrastivo con altri corpora di italiano generale o in ogni caso non accademico, e offrono dati utili per meglio ragionare sulla definizione stessa del lessico accademico (cfr. § 2).

L'analisi è condotta su un corpus di 950.003 token, formato dalle tre sezioni del corpus DIA (778.068 token) cui si somma il corpus DIA pilota (171.935 token: su tutto, cfr. § 5). La base dell'analisi può dunque essere suddivisa in quattro componenti: i tre subcorpora DIA specifici relativi alle macro-aree umanistica, scientifica e giuridico-economica e una quarta parte, rappresentata dal corpus DIA pilota, contenente testi da tutte le macro-aree. Come base lessicale di partenza dell'analisi è stata utilizzata una lista formata da 5.165 lessemi precedentemente identificati come potenzialmente trasversali, ricavata selezionando le parole presenti in almeno 3 dei 4 componenti appena ricordati.

Per il pre-processamento e l'analisi dei dati, sono stati impiegati gli strumenti del software Python, in particolare la libreria spaCy per la lemmatizzazione avanzata del testo. In questo modo è stato facile gestire le varianti morfologiche delle parole (tipicamente i verbi), dato che le forme flesse sono ricondotte a un unico lessema in modo automatico. I due criteri principali per estrarre la lista di parole sono la frequenza e la dispersione delle parole nel corpus. Per ogni lessema nella lista predefinita, sono state pertanto calcolate:

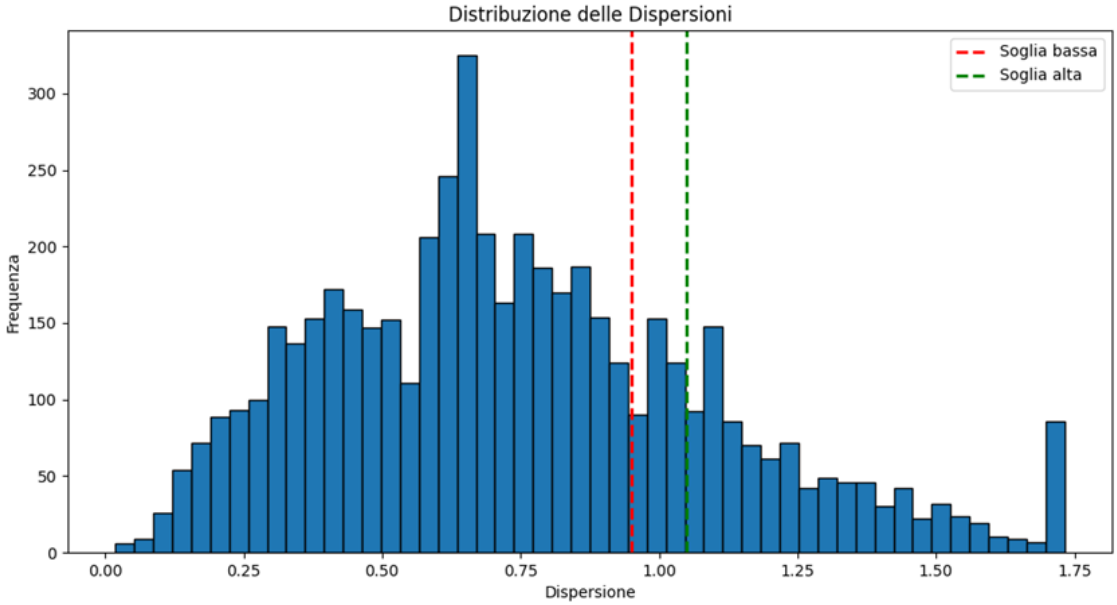
a) la frequenza in ciascuno dei tre subcorpora e nel corpus pilota;

b) la dispersione tra le varie macroaree.

La frequenza è stata calcolata come il rapporto tra il numero di occorrenze di un lessema e il numero totale di parole in ciascuna macro-area. La dispersione, invece, è stata misurata utilizzando un indice che quantifica quanto uniformemente una parola è distribuita tra le diverse macroaree. L'indice di dispersione è calcolato come il coefficiente di variazione delle frequenze nelle diverse macroaree, coi valori più bassi che indicano una distribuzione più uniforme.

Per la classificazione delle parole sono state stabilite due soglie di dispersione, una inferiore (corrispondente al valore di 0.95) e una superiore (del valore di 1.05). Queste soglie, sebbene più alte rispetto a quelle comunemente usate in studi simili (tipicamente 0.6-0.7), sono state scelte basandosi sull'analisi della distribuzione specifica del nostro corpus, come visualizzato nel grafico di dispersione (Figura 1). La scelta di queste soglie più permissive è dovuta alla natura interdisciplinare del lessico accademico e permette di catturare un *range* più ampio di parole potenzialmente rilevanti per la comunicazione accademica trasversale.

Figura 1. *Grafico della distribuzione delle dispersioni delle parole nel corpus*



La Figura 1 illustra la distribuzione delle dispersioni delle parole nel corpus. Le colonne blu rappresentano la frequenza dei lessemi per ciascun valore di dispersione, mostrando come si distribuiscono i lessemi lungo lo spettro di dispersione. La linea rossa tratteggiata rappresenta la soglia inferiore (0,95) e la linea verde tratteggiata rappresenta la soglia superiore (1,05). Come si può osservare, la maggior parte delle parole si concentra nella parte sinistra del grafico, indicando una bassa dispersione e quindi una maggiore trasversalità.

L'analisi è stata realizzata interamente attraverso il software Python, sfruttando diverse librerie specializzate[22]. In particolare, sono state fondamentali le librerie spaCy per l'analisi linguistica della lemmatizzazione e Matplotlib per la visualizzazione dei dati[23]. Il codice principale include le seguenti funzionalità chiave:

- lettura e pre-processamento dei file di testo del corpus;
- lemmatizzazione del testo utilizzando spaCy;
- calcolo delle frequenze e dell'indice di dispersione per ogni lessema;
- classificazione delle parole in base alle soglie di dispersione;
- generazione di output dettagliati per ciascuna categoria di parole;
- creazione di un grafico della distribuzione delle dispersioni con l'utilizzo di Matplotlib.

Per gestire l'analisi di grandi volumi di testo si è usata una strategia di elaborazione a blocchi, suddividendo il corpus in segmenti di dimensioni gestibili (1.000.000 di caratteri[24]) per la lemmatizzazione. I risultati dell'analisi hanno prodotto le seguenti fasce di lessemi[25], basate sui loro valori di dispersione:

- 3.830 parole trasversali (74,2% del totale): questi lessemi hanno un indice di dispersione ≤ 0,95, ciò che indica una distribuzione relativamente uniforme tra le diverse macroaree del corpus. La loro bassa dispersione suggerisce un uso consistente in vari contesti accademici;
- 361 parole in zona grigia (7,0% del totale): questi lessemi hanno un indice di dispersione compreso tra 0,95 e 1,05. Sono parole che non sono chiaramente trasversali, ma nemmeno altamente specifiche di una singola macroarea; richiedono perciò un'analisi più approfondita per determinare la loro potenziale natura accademica;
- 974 parole escluse per alta dispersione (18,8%): questi lessemi hanno un indice di dispersione > 1,05, dunque una distribuzione meno uniforme tra le diverse macro-aree; potrebbe trattarsi di termini più specifici di certe aree disciplinari o utilizzati in modi particolari in determinati contesti accademici.

L'elevata percentuale di lessemi classificati come trasversali (74,2%) suggerisce l'esistenza di un significativo nucleo lessicale trasversale potenzialmente accademico.

La scelta delle soglie di dispersione (0.95 e 1.05) rimane, comunque, un punto critico; infatti, se da un lato tali valori si basano sull'analisi della distribuzione specifica del nostro corpus, dall'altro lato sono significativamente più alti rispetto agli standard comunemente usati in letteratura. La scelta di valori più alti favorisce l'inclusività nella selezione delle parole trasversali, ma potrebbe portare a una sovrastima della trasversalità; e il bilanciamento tra inclusività e specificità nella selezione delle parole trasversali

[22] Le librerie in Python sono collezioni di codice pre-scritto che estendono le funzionalità base del linguaggio di programmazione, permettendo di eseguire operazioni complesse senza dover scrivere tutto il codice da zero. Funzionano come "scatole degli attrezzi" specializzate che possono essere facilmente integrate nei programmi per aggiungere funzionalità specifiche.

[23] *spaCy* è una libreria specializzata per l'elaborazione del linguaggio naturale, che fornisce strumenti avanzati per l'analisi linguistica, inclusa la lemmatizzazione (riduzione delle parole alla loro forma base). *Matplotlib*, invece, è una libreria per la creazione di grafici e visualizzazioni di dati, che nel nostro caso è stata utilizzata per generare l'istogramma della distribuzione delle dispersioni, permettendo una rappresentazione visiva immediata dei risultati dell'analisi.

[24] La scelta di dividere il testo in blocchi di 1 milione di caratteri è stata una soluzione tecnica per evitare che spaCy, dovendo processare grandi volumi di testo, esaurisse la memoria del computer causando rallentamenti o crash.

[25] Le fasce di lessemi ricavate dall'analisi corrispondono direttamente ai valori di dispersione calcolati. La classificazione è stata effettuata utilizzando specifiche soglie di dispersione: valori ≤ 0.95 per le parole trasversali, tra 0.95 e 1.05 per la zona grigia, e > 1.05 per quelle escluse per alta dispersione.

rappresenta una sfida continua, che richiede una valutazione critica dei risultati e possibili aggiustamenti metodologici.

L'approccio computazionale sta rivelando diverse opportunità di arricchimento e sta portando al raffinamento del lemmario provvisorio. La lista estratta ha attirato la nostra attenzione su parole finora non notate negli spogli manuali, come *focalizzare*, *concentrarsi su*, *approfondire* e *indagare*, che fanno tutte capo al "piano dei fenomeni" e che offrono sfumature più precise per descrivere il modo in cui gli estensori dei testi delimitano il loro oggetto di studio. Altri verbi che rappresentano processi centrali nel ragionamento scientifico sono *elaborare*, *interpretare*, *dedurre*, *valutare* e *stimare*. Al piano dell'"organizzazione del testo" si potrebbero ricondurre parole quali *organizzare*, *articolare*, *suddividere*. Nella lista di parole potenzialmente accademiche compaiono anche connettivi logici o formule equivalenti dotate di vari significati, alcuni dei quali di registro alto (tipicamente *nondimeno*): *nondimeno*, *ciononostante*, *eppure*, *ma*, *di conseguenza*, *in sintesi*, *riassumendo*, *vale a dire*, *supponendo che* e altri.

Come anticipato, un secondo tipo di indagine è quella di carattere lessicometrico, e precisamente il confronto della frequenza che alcune parole del vocabolario di base hanno nel corpus DIA da un lato (preso come campione rappresentativo dei testi concettuali e che si servono di lessico accademico) e in altri corpora di italiano non accademico dall'altro (corpus di parlato, corpus di italiano generale). Portare l'attenzione sul vocabolario di base è importante, dato che tra le parole della lista accademica sopra descritta è emersa una sovrapposizione considerevole tra il lessico accademico e il vocabolario di base (ca. il 63% delle 3.830 parole sicuramente trasversali della lista sono incluse nel vocabolario di base)[26]; questo conforta l'ipotesi di una forte tangenza tra vocabolario accademico e vocabolario di base espressa al § 2; ciò che la misurazione contrastiva delle frequenze può offrire è un appiglio quantitativo per valutare l'effettiva accademicità dei lessemi identificati. Per fare ciò, è stata condotta un'analisi comparativa tra il corpus DIA esteso (cioè costituito dal corpus pilota più il corpus DIA vero e proprio) e tre corpora di italiano generale (PAISÀ, Italiano del Web 2020 e KIParla)[27]. L'analisi, concentrata su un campione di 8 lessemi, ha rivelato dei *pattern* interessanti.

Tabella 3. *Confronto delle frequenze dei lessemi tra corpora (i valori sono presentati come numero di occorrenze [frequenza per milione di token])*[28]

Lessema	**DIA + DIA pilota (950,003 token)**	**PAISÀ (267,573,660 token)**	**Italian Web 2020 (14,514,566,714 token)**	**KIParla (2,326,171 token)**
Caratteristica	367 (386.31/mil)	51,616 (192.90/mil)	1,918,652 (132.19/mil)	78 (33.53/mil)
Comportare	197 (207.37/mil)	17,839 (66.67/mil)	1,212,168 (83.51/mil)	73 (31.38/mil)
Consistere	156 (164.21/mil)	18,289 (68.35/mil)	724,838 (49.94/mil)	52 (22.35/mil)
Determinare	448 (471.58/mil)	21,434 (80.11/mil)	1,488,144 (102.53/mil)	229 (98.45/mil)

[26] La percentuale è stata calcolata confrontando la lista dei 3830 lemmi trasversali con il Vocabolario di Base.

[27] *PAISÀ* (Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati) è un corpus di italiano scritto contenente testi estratti dal web, per un totale di circa 267.573.660 token. *Italian Web 2020* è il più ampio corpus di italiano disponibile, parte della famiglia TenTen di corpora web, con oltre 14.5 miliardi di token. *KIParla* è un corpus di italiano parlato che raccoglie conversazioni in diversi contesti, per un totale di circa 2.326.171 token.

[28] Per "tratto", i valori si riferiscono al senso di 'proprietà distintiva'.

Emergere	196 (206.31/mil)	12,164 (45.46/mil)	1,155,979 (79.64/mil)	43 (18.49/mil)
Individuare	254 (267.37/mil)	14,691 (54.90/mil)	1,411,711 (97.26/mil)	74 (31.81/mil)
Registrare	72 (75.79/mil)	52,641 (196.73/mil)	1,827,693 (125.92/mil)	166 (71.36/mil)
*tratto**	70 (73.68/mil)	4,649 (17.37/mil)	377,841 (26.03/mil)	6 (2.58/mil)

Come si può vedere nella Tabella 3, quasi tutti i lessemi analizzati presentano una frequenza significativamente più alta nel corpus DIA rispetto agli altri corpora; ad esempio, *emergere* possiede una frequenza di 206.31 per milione nel corpus DIA, significativamente più alta rispetto agli altri corpora, dove la frequenza varia da 18.49 a 79.64 per milione.

Per *tratto* abbiamo provato a svolgere un'analisi che tenesse conto della semantica, prendendo in considerazione l'accezione di 'proprietà distintiva'. Per i corpora con poche occorrenze complessive (21 in KIParla, 154 nel corpus DIA) è stata possibile un'analisi completa; negli altri casi si è proceduto per campioni statisticamente significativi: ad esempio in PAISÀ, su un campione di 200 occorrenze, 39 (19.5%) presentavano il significato di 'proprietà distintiva'; questa proporzione è stata poi proiettata sul totale delle occorrenze. Il risultato è che il nostro lessema, nell'accezione considerata, ha una frequenza di 73.68/milione nel corpus DIA (70 occorrenze su 154 totali), notevolmente più alta rispetto agli altri corpora, rispettivamente 2.58/milione per KIParla (6 occorrenze su 21 totali), 17.37 per Italian Web 2020 (56 occorrenze su 200 campione), 26.03/milione per PAISÀ (39 occorrenze su 200 campione).

L'unico tra i lessemi analizzati che mostra una distribuzione differente è *registrare*, rispetto al quale la frequenza in altri corpora è simile (71.36/milione in KIParla) o superiore (196.73/milione in PAISÀ, 125.92/milione in Italian Web 2020) al corpus DIA (75.79/milione). Ma se si tenesse conto della semantica, è possibile che anche in questo caso i valori si riequilibrerebbero. Per tutti gli altri casi, il fatto che la frequenza penda a favore del corpus DIA permette anche solo su base quantitativa di concludere che nei testi concettuali queste parole godono di uno statuto diverso, essendo sottoposte a un maggiore sfruttamento.

5. IL CORPUS: FISIONOMIA E CRITERI DI ALLESTIMENTO

L'allestimento del corpus DIA[29] è stato preceduto dalla costruzione di un piccolo corpus pilota, a cura di Michela Dota, per ragionare in primo luogo sulla rappresentatività e sul bilanciamento[30]. Con rappresentatività di un corpus intendiamo, con Biber (1993: 243), «the extent to which a sample includes the full range of variability in a population». Ci siamo quindi interrogati su quali generi testuali includere e 2) quale grado di diversificazione disciplinare perseguire, insieme a questioni di carattere prettamente esecutivo, ossia 3) come trattare il paratesto e 4) stimare i tempi necessari a realizzare il corpus.

Per affrontare i primi due punti, ci è parso opportuno dialogare, per dir così, con un corpus già disponibile di italiano accademico scritto, ossia quello di Spina (2010), dal quale

[29] Per ragioni di copyright, specie per quel che riguarda i manuali universitari (raramente disponibili ad accesso aperto), tanto il corpus DIA pilota quanto il corpus DIA non potranno essere resi pubblicamente consultabili. L'elenco completo dei testi contenuti nel corpus è riprodotto *infra* nell'Appendice.

[30] Sui due aspetti, cruciali nel design di un corpus, si rinvia alla sintesi in Cresti, Moneglia (2016: 586 e ss.).

la studiosa ha ricavato la sua lista di parole accademiche (cfr. § 2). Il corpus di Spina include testi dalle tre macro-aree umanistica, giuridico-economica e scientifica e risulta composto come da Tabella 4.

Tabella 4. *Corpus dello scritto accademico, confluito nel Perugia Corpus (riproduzione della tabella contenuta in Spina, 2014: 355*; *le cifre si riferiscono al numero di parole).*

	tesi	dispense	manuali	articoli	TOT.
umanistica	55.311	170.501	36.077	114.581	*376.470*
giur-eco	58.087	176.206	64.020	75.814	*374.127*
scientifica	54.460	203.197	34.464	70.872	*362.993*
TOT.	*167.858*	*549.904*	*134.561*	*261.267*	*1.113590*

Poiché anche il nostro obiettivo è individuare forme trasversali rispetto ai settori, cioè indipendenti dalle specificità disciplinari, la tripartizione di Spina, confortata da lavori consimili di area anglosassone, ci è sembrata una soluzione idonea ai nostri scopi[31].

I generi rappresentati nel corpus DIA definitivo sono tre: articolo di ricerca, manuale universitario e voce di enciclopedia. Rispetto a Spina (2010; 2014), abbiamo deciso di escludere le tesi di laurea perché possono riprodurre in modo imperfetto, e finanche errato, le forme e le locuzioni accademiche che stiamo cercando[32]. Abbiamo escluso anche le dispense, poiché riteniamo che equivalgano tipologicamente ai manuali; nel corpus di Spina, infatti, la bassa rappresentatività dei manuali appare controbilanciata dalla numerosità delle dispense.

Sulla base di queste prime considerazioni, abbiamo costruito il corpus pilota in formato docx (Tabella 5). Per il calcolo delle parole token abbiamo utilizzato AntConc, convertendo previamente i file di lavoro al formato txt compatibile col software. Abbiamo quindi personalizzato i criteri di tokenizzazione: oltre alle sequenze di lettere isolate, sono considerate equivalenti alle parole le cifre e i simboli, quali gli operatori matematici; sono invece esclusi dal computo i segni di interpunzione.

Tabella 5. *Fisionomia del corpus pilota (calcolo dei token realizzato in AntConc)*

	Manuali	**Articoli**	**Parole token totali**
Area umanistica	40.837	5.031	45.868
Area scientifica 1 (biologia)	39.239	5.484	44.723
Area scientifica 2 (matematica e informatica)	37.486	7.970	45.456

[31] Ad esempio persegue il criterio della trasversalità l'*Oxford Corpus of Academic English* (Rees, 2022: 394). Egualmente trasversale alle discipline (benché basato su una suddivisione parzialmente diversa) è il corpus ASIC (Academic Spoken Italian Corpus), su cui si veda il contributo di Peppoloni nel presente fascicolo di *Italiano LinguaDue*. Con diversa impostazione, per fornire un quadro rappresentativo dell'italiano accademico scritto avremmo potuto considerare l'incidenza, in termini di quantità di prodotti della ricerca, che ciascuna macroarea ha sul complesso delle pubblicazioni accademiche in lingua italiana, così come rilevato dall'ANVUR (Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca). Rispettare la proporzionalità del contributo effettivo delle diverse macroaree avrebbe però prodotto un corpus fortemente sbilanciato sulla macroarea umanistica e sull'area giuridica, a discapito di un'equa trasversalità tra i saperi, compresi quelli veicolati quasi esclusivamente da pubblicazioni in lingua inglese.

[32] Due casi di uso improprio di una parola accademica quale *individuare* nell'interlingua degli scriventi di scuola superiore sono commentati in Mastrantonio (2023a: 112-113). Sulle molteplici incertezze scrittorie degli studenti è ormai disponibile una vasta letteratura, in continuo (e sconfortato) aggiornamento; si veda, da ultimo, Restivo (2022).

Area economica	31.354	4.534	35.888
Parole token totali	148.916	23.019	**171.935**

Il corpus pilota consta di 8 pubblicazioni (8 articoli e 8 manuali), di 8 autori diversi, distribuite sulle 3 macroaree e riprodotte integralmente. Il formato docx è servito a formattare i testi con gli *heading styles*, in modo da visualizzare a colpo d'occhio, attraverso la localizzazione automatica delle occorrenze, la presenza e la distribuzione delle prime ipotetiche voci da includere nel lemmario (vd. Figura 2).

Nel corpus pilota la macroarea scientifica è più rappresentata poiché è la più utile per ragionare sul trattamento del paratesto e delle formule, non sempre pienamente inserite nella sintassi verbale. Per evitare di appesantire eccessivamente il file di lavoro e al contempo preservare le trame logodeittiche fra testo e paratesto su cui poggiano i testi accademici, abbiamo omesso tabelle, apparati iconografici e formule complesse, lasciando un rimando al luogo del testo originale (vd. Figura 3).

Figura 2. *Localizzazione della voce esemplificativa 'suggerisce' nel corpus pilota*

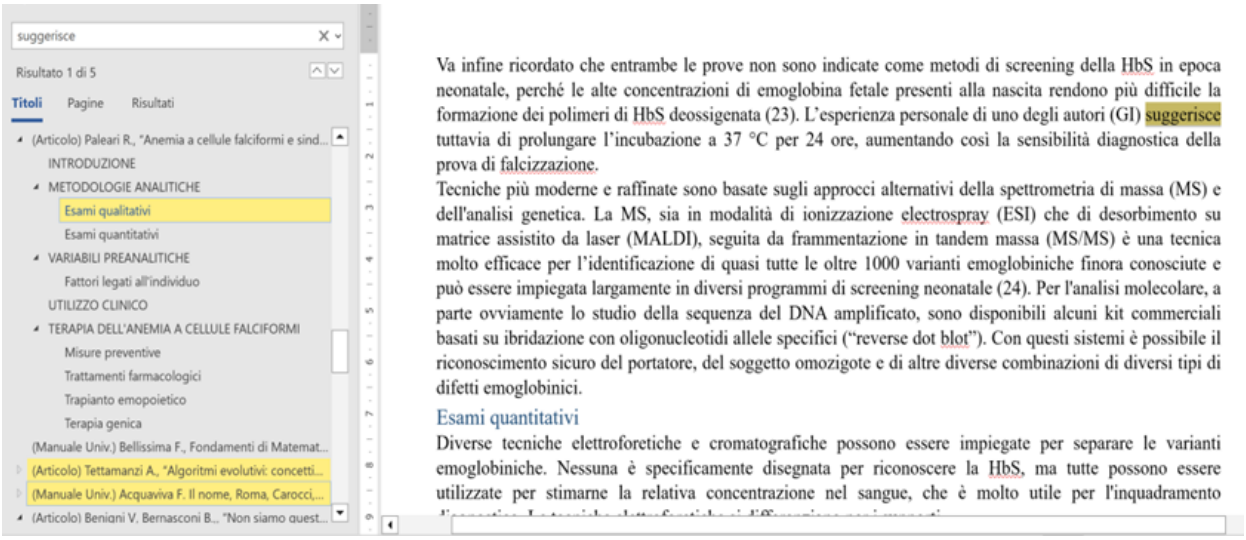


Figura 3. *Esempio di trattamento del paratesto*

chimiche specifiche (Fig. 4.4). Poiché le proteine hanno una sequenza e una frequenza di aminoacidi diversa, l'organizzazione dei gruppi R conferisce alle diverse proteine strutture e proprietà funzionali peculiari.
In natura sono presenti 20 aminoacidi che vengono suddivisi in sottogruppi, a seconda che il gruppo R sia acido (per esempio l'acido aspartico), basico (lisina), neutro polare (leucina) o neutro non polare (serina) (Tabella 4.3 e Fig. 4.5).
Il carbonio in posizione α, fatta eccezione per la glicina, è asimmetrico in quanto lega quattro gruppi diversi; quindi per ogni aminoacido possono esistere due isomeri che si designano come forme De L-, destrogira e levogira.

Tabella 4.3[p.36] Sottogruppi degli aminoacidi e loro codici a singola e tripla lettera

Nelle proteine naturali gli aminoacidi sono sempre presenti nella forma L, mentre le forme D sono presenti in alcuni piccoli polipeptidi elaborati da microrganismi.
Gli aminoacidi sono sostanze anfotere, in quanto contengono almeno un gruppo funzionale acido −COOH e

La realizzazione del corpus pilota ci ha permesso di riflettere meglio su alcuni problemi: da un lato sondare la disponibilità di testi già digitalizzati o da digitalizzare tramite OCR (ciò che comporta un dispendio di tempo significativo), dall'altro la questione dell'opportunità di diversificare al massimo gli autori e le sedi editoriali, per scongiurare il rischio di incorrere negli idioletti degli autori o in forme di omologazione dovute a interventi redazionali. Il rischio della sovrarappresentazione di un idioletto si corre anche inserendo integralmente i manuali universitari nel corpus, data la loro maggiore estensione materiale rispetto agli articoli di ricerca, molto più brevi.

Nel corpus definitivo, ovvero il corpus DIA, per equilibrare l'incidenza quantitativa dei diversi generi e degli idioletti autoriali, abbiamo preso come riferimento l'estensione tendenziale di un articolo di area umanistica, ossia circa 5.000 parole token ovvero 35.000/40.000 caratteri, bibliografia esclusa[33]. Ci è parso, inoltre, che questa dimensione fosse un buon compromesso tra il disporre di testi con un'estensione sufficiente per garantirci un input di forme e funzioni abbastanza ricco e la necessità di incrementare velocemente la quantità di testi e di autori: il nostro obiettivo iniziale prevedeva di annoverarne almeno un centinaio, badando a non includere più pubblicazioni di uno stesso autore. Gli articoli di ricerca sono stati quindi selezionati in modo che la loro estensione corrispondesse a quella prefissata e sono sempre stati trascritti integralmente, mentre i manuali sono stati campionati in modo statisticamente rilevante, prelevando escerti dall'introduzione, dal capitolo centrale e dall'ultimo capitolo o conclusione, in modo che la loro somma ammontasse a 5.000 parole token – con un certo margine di tolleranza, in eccesso o in difetto, comunque non superiore alle cinquecento parole. Gli estratti, idealmente omogenei tra loro quanto a lunghezza (circa 1700 parole token), sono stati confezionati in modo flessibile, per preservare la coesione e la coerenza interna delle singole porzioni. Come si è già detto, oltre agli articoli e ai manuali, il corpus definitivo include le voci enciclopediche, uniformi per registro (neutro o alto) agli altri due generi. Redatte sempre da accademici, le voci sono state scelte, o ritagliate, in modo che l'estensione si attestasse attorno alle 5.000 parole.

Venendo ai criteri operativi generali, abbiamo privilegiato le pubblicazioni edite dal 2000 in poi, in modo da avere una fotografia sincronica dell'italiano accademico attuale; come si dirà in § 5.1., non è stato sempre possibile rispettare questo principio a causa delle contingenze che si illustreranno.

Un ulteriore criterio, attivato soprattutto per i subcorpora giuridico-economico e scientifico, corrisponde alla prototipicità disciplinare delle pubblicazioni: per il subcorpus scientifico, ad esempio, abbiamo escluso le pubblicazioni relative alle didattiche disciplinari (quali didattica della matematica, didattica della fisica e simili) poiché, in quanto tangenti ai saperi pedagogico-didattici della macroarea umanistica, non incarnano la tipica pubblicazione di area scientifica. L'applicazione di questo criterio, specie in relazione al subcorpus scientifico, non è stata indolore, poiché attingere ai Settori delle didattiche, che dispongono di molte pubblicazioni in lingua italiana, avrebbe consentito di ovviare con maggiore agio alla forte incidenza – se non proprio egemonia – dell'inglese, specie per alcuni settori scientifico-disciplinari. L'ultimo criterio generale è, infatti, la preferenza accordata alle pubblicazioni redatte in lingua italiana, evitando, per quanto possibile, le traduzioni dall'inglese. D'altra parte, l'inevitabile presenza di qualche traduzione (vd. *infra* § 3.1., b) offre l'opportunità di isolare un subcorpus utile a indagare l'eventuale condizionamento che la traduzione ha su forme, funzioni e loro sequenze

[33] Lavorando primariamente su file in formato docx, abbiamo sempre realizzato un primo computo orientativo delle parole di ciascun testo sfruttando la funzione di conteggio nativa di Microsoft Word; il risultato è stato poi raffrontato con il calcolo generato da AntConc.

preferenziali, che possono tradursi in cascate di collocazioni (*collocational cascades*) ricorsive (Gledhill, 1995: 24).

Ognuna delle tre macroaree del corpus finale consta quindi di 48 testi di (almeno) altrettanti autori: la co-autoria, molto diffusa per certe discipline, aumenta significativamente il numero degli scriventi, che supera senz'altro 144. I 48 testi per macroarea sono ripartiti sui 3 generi: sono stati quindi inclusi 16 articoli, 16 voci enciclopediche e 16 estratti di manuali. Avendo scelto l'estensione di 5.000 parole circa per testo, ciascun genere, all'interno di ciascuna macroarea, contribuisce equamente al corpus, con 80.000/84.000 parole di Microsoft Word, equivalenti a 85.000/88.000 parole token in AntConc; ogni macroarea si attesta in media sulle 250.000 parole di Microsoft Word, equivalenti a circa 260.000 parole token di AntConc. La Tabella 6 riepiloga l'assetto quantitativo del corpus, secondo il computo di AntConc. Come per il corpus provvisorio o pilota, il computo include i metadati riferiti a ciascuna pubblicazione, ossia una breve aggiunta testuale a cura nostra con l'indicazione delle seguenti informazioni: il genere testuale di appartenenza, l'autore, il titolo, il luogo di edizione e l'editore, l'anno di pubblicazione, l'estensione in parole e la sua afferenza disciplinare (stabilita secondo criteri illustrati oltre).

Tabella 6. *Dimensione del corpus e sua consistenza per generi e macroaree*

	Manuali	**Articoli di ricerca**	**Voci enciclopediche**	**Totale**
Macroarea giuridico-economica	86.198	88.131	88.352	262.681
Macroarea scientifica	83.744	88.483	82.177	254.404
Macroarea umanistica	88.004	86.646	86.333	260.983
Totale	257.946	263.260	256.862	**778.068**

Il corpus DIA sfiora così le 780.000 parole token, una dimensione che potremmo reputare modesta se il corpus intendesse rappresentare un ampio spettro di varietà[34], ma che invece possiamo ritenere idonea agli scopi del progetto, sia teorici sia glottodidattici, nonché compatibile con i tempi a disposizione per la sua realizzazione, in relazione alla durata complessiva del progetto. Dal punto di vista teorico, infatti,

> [t]here is no ideal size for a corpus; it all depends on what the corpus contains and what is being investigated. [...] While specialised corpora may vary in size, an important point is that such corpora do not need to be as large as more general corpora to yield reliable results. The reason for this is that as specialised corpora are carefully targeted, they are more likely to reliably represent a register or genre than general corpora. Even with relatively small amounts of data, 'specialised lexis and structures are likely to occur with more regular patterning and distribution' than in a large, general corpus. (Koester, 2022: 50-51 e riferimenti; cfr. Hunston, 2012).

Considerata quindi la tipologia del nostro corpus (specializzato e funzionale alla redazione di un dizionario, con uno spiccato scopo pedagogico; cfr. Koester, 2022; Rees, 2022), un'estensione molto ampia non sarebbe necessariamente l'opzione migliore:

[34] Non si può, però, ritenerlo un corpus piccolo, generalmente identificato dalla letteratura con l'estensione di duecentocinquantamila parole (Koester, 2022: 49). Del resto, più importanti della dimensione sono la rappresentatività (Ballance, Coxhead, 2022: 410) e lo scopo (Szudarski, 2018: 137) del corpus.

> large corpora do not meet the needs of teachers and learners in ESP/EAP, as they provide 'either too much data across too large a spectrum, or too little focused data, to be directly helpful to learners with specific learning purposes' […] Smaller, more focused corpora, which have been set up for a specific research or pedagogical purpose, are much more likely to yield insights that are directly relevant for teaching and learning for specific purposes (Koester, 2022: 49 e riferimenti).

Quanto ai tempi a disposizione, essi rappresentano inevitabilmente un limite alla possibilità di confezionare un corpus che sia davvero un campione scientificamente affidabile di un determinato dominio linguistico; ma si tratta di limiti temporali comuni a progetti consimili al nostro (cui vanno aggiunti le risorse umane e finanziarie limitate), che comunque non impediscono di approdare a descrizioni e analisi utili (Ballance, Coxhead, 2022: 411; sui vincoli determinati dalle contingenze del lavoro vd. anche Hunston, 2008: 165).

In caso di necessità, il corpus si potrà eventualmente ampliare in modo mirato, ad esempio in presenza di dubbi sulla trasversalità di una forma poco rappresentata o addirittura assente in un'area. Dato l'obiettivo dichiarato nel progetto di redigere un numero predefinito e limitato di voci per il Dizionario, al momento non pare necessario saturare il corpus, ossia aggiungere altri testi (e quindi parole token) finché cessano di emergere nuove unità lessicali pertinenti (cfr. Koester, 2022: 51-52 e riferimenti). Ma dalle indagini senesi sulla lingua dei giornali è possibile che si troveranno conferme sull'opportunità di considerare anche altri registri (come si mostra in § 2 agli ess. 1-2), che potrebbero essere utili sia per consolidare le ipotesi sulle funzioni comunicative, sia per fornire l'equivalente informale o colloquiale di quelle funzioni come via di accesso didattico (procedere dal noto al nuovo).

Per individuare pubblicazioni davvero eterogenee, abbiamo deciso di affidarci al sistema CUN di classificazione dei saperi: benché non immune da criticità, esso è senz'altro uno strumento tassonomico dotato di una sua coerenza interna, perciò più oggettivo dell'attribuzione intuitiva di una pubblicazione a un certo ambito; inoltre la sua organizzazione interna a grana progressivamente più fine (area disciplinare [anche AD] > macrosettore concorsuale > settore concorsuale > raggruppamento disciplinare > settore scientifico-disciplinare [anche SSD]) lo rende adeguato a perseguire il criterio della massima diversificazione. Il sistema di classificazione delle pubblicazioni per SSD, peraltro, è certamente più visibile nei cataloghi universitari della ricerca[35] rispetto a sistemi di classificazione concorrenti, come quello europeo (ERC). I cataloghi dei prodotti della ricerca degli atenei, infatti, sono stati sistematicamente consultati per accertarci della corretta affiliazione di ciascuna pubblicazione, prodotta da autori e autrici strutturati o emeriti, o da esperti affiliati ad atenei o ad enti assimilabili, dunque, in generale, da professionisti nell'uso dell'italiano accademico. Va detto che il lavoro è stato impostato, condotto e ultimato prima dell'entrata in vigore dei nuovi SSD e Gruppi Scientifico-Disciplinari (DM n. 639 del 2 maggio 2024)[36]; tuttavia, crediamo che l'intempestiva obsolescenza del criterio eletto non abbia ripercussioni significative sull'affidabilità del lavoro. Abbiamo quindi fatto confluire le Aree CUN nella nostra tripartizione, come illustrato in Tabella 7.

[35] Sono gli archivi istituzionali di cui ciascuna università dispone per raccogliere, documentare e, talvolta, rendere liberamente consultabili le pubblicazioni del personale docente e ricercatore.
[36] https://www.mur.gov.it/it/atti-e-normativa/decreto-ministeriale-n-639-del-02-05-2024.

Tabella 7. *Confluenza delle Aree CUN nelle tre macroaree del corpus*

Aree CUN	**MACROAREA DEL CORPUS**
01 – Scienze matematiche e informatiche	SCIENTIFICA
02 – Scienze fisiche	
03 – Scienze chimiche	
04 – Scienze della Terra	
05 – Scienze biologiche	
06 – Scienze mediche	
07 – Scienze agrarie e veterinarie	
08 – Ingegneria civile e Architettura	
09 – Ingegneria industriale e dell'informazione	
10 – Scienze dell'antichità, filologico-letterarie e storico-Artistiche	UMANISTICA
11 – Scienze storiche, filosofiche, pedagogiche e Psicologiche	
12 – Scienze giuridiche	GIURIDICO – ECONOMICA
13 – Scienze economiche e statistiche	
14 – Scienze politiche e sociali	UMANISTICA

Come si vede, restano escluse le aree CUN ingegneristiche (8 e 9), perché a rigore non assimilabili alla macroarea scientifica e dunque non assorbibili nella tripartizione che abbiamo scelto di adottare. Tuttavia crediamo che non sia una lacuna grave: da un canto, buona parte dei settori disciplinari dell'area 9 ha molteplici tangenze con i settori delle Aree 1, 2 e 3, e dunque questa assenza non incide molto sul grado di rappresentatività del corpus; dall'altro canto, parte dei settori dell'area 8 potrebbero essere coperti, in caso di necessità, attingendo al corpus di italiano accademico scritto del progetto SEHA, proprio incentrato sull'architettura e l'ingegneria delle costruzioni (cfr. Piccioni, D'Angelo, Ferro, 2021).

5.1 *I tre subcorpora: fisionomia, criteri particolari di allestimento e criticità riscontrate*

a) *Subcorpus giuridico-economico*

Il subcorpus giuridico economico, assemblato da Michela Dota, si è rivelato il meno problematico da allestire, sia in termini di reperibilità ed estensione dei testi, sia in termini di equilibrio tra le due componenti, giuridica ed economica, giacché le rispettive aree CUN (12 e 13) contengono un numero simile di SSD; ciascuno dei 3 generi testuali conta quindi 8 pubblicazioni di area giuridica e 8 di area economica (per un totale di 48 testi per macroarea, lo ricordiamo). D'altro canto, nonostante l'assetto quantitativo delle due Aree sia simile, per alcuni SSD, specie di economia e statistica, le pubblicazioni in italiano (o non frutto di traduzioni) scarseggiano; in altri casi, alcuni Settori non sono stati privilegiati perché sono stati ritenuti non prototipici dell'area: le rispettive discipline, infatti, intersecano campi e oggetti di ricerca delle altre due macroaree (si pensi alle diverse statistiche, congeneri alla macroarea scientifica, o a Storia del pensiero economico, pertinente anche all'area umanistica). Si tratta, tuttavia, di un criterio di massima, cui

talvolta abbiamo derogato, a causa dell'iniqua reperibilità delle pubblicazioni tra i Settori; abbiamo dunque attinto al SSD di Diritto romano e diritti dell'antichità e di Filosofia del diritto, sebbene i rispettivi temi di ricerca collimino con quelli degli studiosi di storia antica e di filosofia. Nella Tabella 8 si fornisce il prospetto delle discipline incluse nel corpus (in neretto gli SSD rappresentati da almeno una pubblicazione).

Tabella 8. *Tavola degli SSD inclusi nel subcorpus giuridico-economico*

Area 12 - Scienze giuridiche	Area 13 - Scienze economiche e statistiche
IUS/01 DIRITTO PRIVATO	**SECS-P/01 ECONOMIA POLITICA**
IUS/02 DIRITTO PRIVATO COMPARATO	**SECS-P/02 POLITICA ECONOMICA**
IUS/03 DIRITTO AGRARIO	**SECS-P/03 SCIENZA DELLE FINANZE**
IUS/04 DIRITTO COMMERCIALE	SECS-P/04 STORIA DEL PENSIERO ECONOMICO
IUS/05 DIRITTO DELL'ECONOMIA	SECS-P/05 ECONOMETRIA
IUS/06 DIRITTO DELLA NAVIGAZIONE	**SECS-P/06 ECONOMIA APPLICATA**
IUS/07 DIRITTO DEL LAVORO	**SECS-P/07 ECONOMIA AZIENDALE**
IUS/08 DIRITTO COSTITUZIONALE	**SECS-P/08 ECONOMIA E GESTIONE DELLE IMPRESE**
IUS/09 ISTITUZIONI DI DIRITTO PUBBLICO	**SECS-P/09 FINANZA AZIENDALE**
IUS/10 DIRITTO AMMINISTRATIVO	**SECS-P/10 ORGANIZZAZIONE AZIENDALE**
IUS/11 DIRITTO CANONICO E DIRITTO ECCLESIASTICO	SECS-P/11 ECONOMIA DEGLI INTERMEDIARI FINANZIARI
IUS/12 DIRITTO TRIBUTARIO	**SECS-P/12 STORIA ECONOMICA**
IUS/13 DIRITTO INTERNAZIONALE	SECS-P/13 SCIENZE MERCEOLOGICHE
IUS/14 DIRITTO DELL'UNIONE EUROPEA	SECS-S/01 STATISTICA
IUS/15 DIRITTO PROCESSUALE CIVILE	SECS-S/02 STATISTICA PER LA RICERCA
IUS/16 DIRITTO PROCESSUALE PENALE	SPERIMENTALE E TECNOLOGICA
IUS/17 DIRITTO PENALE	SECS-S/03 STATISTICA ECONOMICA
IUS/18 DIRITTO ROMANO E DIRITTI DELL'ANTICHITÀ	**SECS-S/04 DEMOGRAFIA**
IUS/19 STORIA DEL DIRITTO MEDIEVALE E MODERNO	SECS-S/05 STATISTICA SOCIALE
IUS/20 FILOSOFIA DEL DIRITTO	SECS-S/06 METODI MATEMATICI
IUS/21 DIRITTO PUBBLICO COMPARATO	DELL'ECONOMIA E DELLE SCIENZE ATTUARIALI E FINANZIARIE

b) *Subcorpus scientifico*

Molto più variegata è invece la macroarea scientifica: conta 7 Aree, ciascuna costituita da un numero molto variabile di SSD (da un minimo di 8 dell'Area 2 ai 50 dell'Area 6). Per formare il subcorpus, curato da Alice Martignoni, i 16 testi di ogni genere sono stati distribuiti nelle 7 Aree, in modo che per ogni genere testuale vi fossero 2 o 3 pubblicazioni di ciascuna Area CUN, il più possibile diversificate quanto a SSD. Nella Tabella 9 si forniscono i prospetti riassuntivi della composizione del subcorpus per ciascun genere testuale.

Tabella 9. *Aree CUN e SSD delle pubblicazioni incluse nel subcorpus scientifico per genere testuale*

Articoli di ricerca	**Manuali**	**Voci enciclopediche**
Area CUN – SSD	Area CUN – SSD	Area CUN – SSD
1 MAT/02; MAT/03	1 MAT/03	1 MAT/03
1 MAT/02	1 MAT/05	1 MAT/05
2 FIS/01	2 FIS/01	2 FIS/02
2 FIS/04	2 FIS/02	2 FIS/04
3 CHIM/02	3 CHIM/01	3 CHIM/06
3 CHIM/02	3 CHIM/06	3 CHIM/04
4 GEO/02	3 CHIM/10	4 GEO/12
4 GEO/04	4 GEO/06	4 GEO/11
5 BIO/03	4 GEO/O8	5 BIO/07
5 BIO/12	5 BIO/14	5 BIO/14
5 BIO/07	5 BIO/19	5 BIO/18
6 MED/11	6 MED/09	6 MED/09
6 MED/09	6 MED/17	6 MED/10
6 MED/13	6 MED/25	7 VET/03; VET/08
7 AGR/04	7 AGR/02	7 VET/03; VET/08
7 AGR/19	7 VET/08	7 AGR/11

Come per gli altri subcorpora, abbiamo dovuto contemperare il principio guida della massima diversificazione con altri criteri o variabili contingenti, che hanno condizionato la selezione degli SSD. È facile immaginare come la difficoltà principale sia consistita nella scarsità o nella scarsa reperibilità per alcune discipline di pubblicazioni in italiano: l'Area 2 (Scienze Fisiche), ad esempio, ne registra poche; viceversa, le Aree di scienze mediche e di scienze chimiche, disponendo – apparentemente – di un maggior numero di pubblicazioni in italiano, sembrano non ancora monopolizzate dall'inglese. La prevedibile egemonia dell'inglese ci ha obbligati a escludere alcuni settori o ci ha costretti a selezionare dei manuali tradotti[37]. In un caso, relativo alla già menzionata Area 2, abbiamo aggirato la scarsità di testi in italiano selezionando una pubblicazione leggermente antecedente rispetto al limite temporale fissato, e cioè risalente agli anni Novanta[38].

Inattesa, invece, è stata la criticità connessa alla lunghezza media degli articoli e delle voci enciclopediche di area scientifica, che si è rivelata tendenzialmente inferiore alle cinquemila parole, almeno secondo il computo automatico di Microsoft Word. Per ovviare alla brevità di alcuni contributi, coerentemente con quanto fatto occasionalmente per gli altri subcorpora, li abbiamo bilanciati con altri articoli eccedenti il limite stabilito di 5.000 parole, in modo che la somma di due contributi raggiungesse ugualmente le 10.000 parole attese. Talvolta questo bilanciamento è stato applicato all'ordine superiore, cioè al numero di token per Area, avendo cura, per quanto possibile, di non far prevalere l'apporto di una stessa Area: a titolo esemplificativo, per le voci enciclopediche l'Area 5 e

[37] Per la precisione, sono 3 i manuali tradotti, relativi alle Aree 3 e 4: cfr. il repertorio dei testi in Appendice, dove si vede che anche il subcorpus giuridico-economico contiene 4 manuali di economia tradotti. Tale scelta potrebbe naturalmente avere ripercussioni sulla testualità e sulle forme emergenti; e del resto, l'influsso dell'inglese non è detto che si limiti alle traduzioni: nel caso di autori molto esposti all'inglese scientifico, le stesse produzioni originali italiane potrebbero risentire dei modelli discorsivi inglesi.

[38] Analogamente per il subcorpus umanistico abbiamo sporadicamente attinto a testi antecedenti agli anni Duemila, per conseguenza della ricerca della massima diversificazione disciplinare.

l'Area 7 sopperiscono al sottodimensionamento dell'Area 4, mentre per gli articoli, ancora l'Area 5, insieme all'Area 6, bilanciano le dimensioni inferiori delle Aree 1, 2 e 3.

c) *Subcorpus umanistico*

La macroarea umanistica, di cui si è occupata Serena Nardella, abbraccia tre aree CUN (10, 11 e 14): l'area 10 (Scienze dell'antichità, filologico-letterarie e storico-artistiche), che è la più corposa, è costituita da 5 aree disciplinari (AD) e 77 SSD; l'area 11 (Scienze storiche, filosofiche, pedagogiche e psicologiche) da 7 AD e 34 SSD; l'area 14 (Scienze politiche e sociali) da 1 AD e 14 SSD[39].

La difformità quantitativa tra le compartizioni degli ambiti risultava, in fase di avvio del lavoro di allestimento del corpus, un impaccio da superare attraverso la ricerca di un criterio il più possibile equo, con cui sarebbero stati trascelti i testi da accogliere. Era necessario, infatti, individuare un metodo selettivo che tenesse coerentemente conto di tutte le discipline, affinché per ogni Area potessero essere presenti contributi attinenti almeno a tutti gli ambiti disciplinari, a più settori possibili e questi ultimi in numero bilanciato rispetto all'incidenza nelle rispettive Aree, data dalla quantità di settori contenuti, di ciascun ambito.

In Tabella 10 è riprodotto un primo tentativo di campionamento basato sulla selezione di un numero di testi variabile (da 1 a 5 per AD). Come è possibile osservare, per ogni area la quantità prevista di lavori da coinvolgere sarebbe stata determinata dalla produttività disciplinare degli ambiti compresi. Dunque, se per l'area 14 l'unico AD avrebbe fornito 4 testi, l'oscillazione d'uso nelle restanti due Aree (3-5 testi per l'Area 10, 1-3 testi per l'Area 11) sarebbe stata orientata dalla proporzione in ognuna tra AD e Settori, questi ultimi più numerosi nella prima, che è meno densa di AD (pertanto, per L-ANT, con 10 SSD, la selezione avrebbe riguardato 3 testi, per L-OR, con 23 SSD, 5 testi, per L-FIL-LET, con 15 SSD, 4 testi), più ridotti nella seconda, che ospita, invece, più ambiti (con lo stesso criterio, per M-PSI, con 8 SSD, sarebbero stati isolati 3 testi e per M-EDF, con 2 SSD, un solo testo).

Tabella 10. *Prospetto del primo tentativo di selezione dei testi (t)*

Area 10 = 5 AD = 77 SSD	**Area 11 = 7 AD = 34 SSD**	**Area 14 = 1 AD = 14 SSD**
L-FIL-LET (15 ssd) = 4 t	M-DEA (1 ssd) = 1 t	SPS (14 ssd) = 4 t
L-ANT (10 ssd) = 3 t	M-STO (9 ssd) = 3 t	Area 14 = 4 t
L-ART (8 ssd) = 3 t	M-GGR (2 ssd) = 1 t	
L-LIN (21 ssd) = 5 t	M-FIL (8 ssd) = 3 t	
L-OR (23 ssd) = 5 t	M-PED (4 ssd) = 1 t	
Area 10 = 20 t	M-PSI (8 ssd) = 3 t	
	M-EDF (2 ssd) = 1 t	
	Area 11= 13 t	

Il criterio appena illustrato, plasmato sulla conformazione del campo umanistico, sarebbe stato difficilmente applicabile alle altre due macroaree; ma ha posto le basi per l'elaborazione di un sistema di selezione testuale regolato dai medesimi principi di

[39] Per il dettaglio delle tre Aree, vd. https://www.cun.it/uploads/storico/settori_scientifico_disciplinari_english.pdf.

diversificazione e bilanciamento per raggiungere un'alta rappresentatività disciplinare nell'intero corpus.

Pertanto, abbandonata questa prima ipotesi di lavoro e dopo aver stabilito che per ogni macroarea sarebbero stati coinvolti 48 testi totali, si è ritenuto opportuno ripartirne il numero, soppesando la quantità di settori generati da ognuna delle tre aree, in base allo schema che segue:

area 10 = 8 manuali universitari, 8 articoli di ricerca e 8 voci enciclopediche;

area 11 = 6 manuali universitari, 6 articoli di ricerca e 6 voci enciclopediche;

area 14 = 2 manuali universitari, 2 articoli di ricerca e 2 voci enciclopediche.

Nel nuovo processo di reperimento testuale, la varietà disciplinare non è stata l'unico criterio seguito: come per gli altri subcorpora, si è tenuto conto anche dell'estensione dei testi, della loro differenziazione autoriale ed editoriale e del contenuto linguistico, nel modo che si specifica di seguito.

1) *Varietà disciplinare.*

Per ogni Area, affinché i settori scientifico-disciplinari fossero massimamente rappresentati e per evitare che vi fossero sovrapposizioni tra questi, è stata ricercata la massima varietà, attuando anche delle sostituzioni in corso d'opera. Sono stati inseriti testi provenienti da tutti gli ambiti disciplinari inclusi nelle tre Aree e da un numero sostanziale di Settori. Come osservato, il volume degli ambiti compresi da ciascuna Area risulta variabile. Pertanto, nella selezione, è stato bilanciato (anche in base alla quantità di SSD generati dallo stesso AD) il numero di testi per AD, affinché tutti gli ambiti figurassero equamente nel lavoro finale e vi fosse un considerevole assortimento disciplinare[40]. Per l'Area 11, ad esempio, il raggruppamento disciplinare M-FIL, che contiene 8 SSD, ha fornito più testi al corpus (5, afferenti a M-FIL/02, M-FIL/03, M-FIL/04, M-FIL/06, M-FIL/08) rispetto a M-GGR che include solamente 2 SSD: per quest'ultimo è stato selezionato un solo lavoro (attinente al Settore M-GGR/01). In Tabella 11 si riportano gli SSD rappresentati nel subcorpus.

Tabella 11. *Aree CUN, AD e SSD dei testi inclusi nel subcorpus umanistico*

Area 10	
AD	**SSD**
L-FIL-LET	L-FIL-LET/09 FILOLOGIA E LINGUISTICA ROMANZA
	L-FIL-LET/10 LETTERATURA ITALIANA
	L-FIL-LET/11 LETTERATURA ITALIANA CONTEMPORANEA
	L-FIL-LET/12 LINGUISTICA ITALIANA
L-ANT	L-ANT/04 NUMISMATICA
	L-ANT/06 ETRUSCOLOGIA E ANTICHITÀ ITALICHE
	L-ANT/07 ARCHEOLOGIA CLASSICA
	L-ANT/09 TOPOGRAFIA ANTICA
	L-ANT/10 METODOLOGIE DELLA RICERCA ARCHEOLOGICA

[40] Ai fini di un possibile ampliamento del corpus, un ulteriore parametro di cui tenere eventualmente conto in relazione alla provenienza disciplinare dei testi coinvolti all'interno di ciascuna area è la quantità di docenti che vi afferisce (dati reperibili al link https://cercauniversita.mur.gov.it/php5/docenti/cerca.php corrispondente alla sezione 'Docenti' del sito Cineca): in tal caso, i testi relativi ad aree che vantano un cospicuo numero di studiosi con inquadramento accademico, come L-FIL-LET, subirebbero un significativo incremento rispetto ad altri. D'altra parte, questo approccio condizionerebbe sensibilmente il contenuto disciplinare del corpus in riferimento al problema della trasversalità.

L-LIN	L-LIN/01 GLOTTOLOGIA E LINGUISTICA
	L-LIN/08 LETTERATURA PORTOGHESE E BRASILIANA
	L-LIN/10 LETTERATURA INGLESE
	L-LIN/11 LINGUE E LETTERATURE ANGLO-AMERICANE
	L-LIN/14 LINGUA E TRADUZIONE – LINGUA TEDESCA
L-ART	L-ART/01 STORIA DELL'ARTE MEDIEVALE
	L-ART/02 STORIA DELL'ARTE MODERNA
	L-ART/06 CINEMA, FOTOGRAFIA E TELEVISIONE
	L-ART/07 MUSICOLOGIA E STORIA DELLA MUSICA
L-OR	L-OR/01 STORIA DEL VICINO ORIENTE ANTICO
	L-OR/02 EGITTOLOGIA E CIVILTÀ COPTA
	L-OR/03 ASSIRIOLOGIA
	L-OR/05 ARCHEOLOGIA E STORIA DELL'ARTE DEL VICINO ORIENTE ANTICO
	L-OR/21 LINGUE E LETTERATURE DELLA CINA E DELL'ASIA SUD-ORIENTALE
Area 11	
AD	**SSD**
M-PSI	M-PSI/03 PSICOMETRIA
	M-PSI/04 PSICOLOGIA DELLO SVILUPPO E PSICOLOGIA DELL'EDUCAZIONE
	M-PSI/07 PSICOLOGIA DINAMICA
M-FIL	M-FIL/02 LOGICA E FILOSOFIA DELLA SCIENZA
	M-FIL/03 FILOSOFIA MORALE
	M-FIL/04 ESTETICA
	M-FIL/06 STORIA DELLA FILOSOFIA
	M-FIL/08 STORIA DELLA FILOSOFIA MEDIEVALE
M-STO	M-STO/04 STORIA CONTEMPORANEA
	M-STO/05 STORIA DELLA SCIENZA E DELLE TECNICHE
	M-STO/06 STORIA DELLE RELIGIONI
	M-STO/07 STORIA DEL CRISTIANESIMO E DELLE CHIESE
	M-STO/08 ARCHIVISTICA, BIBLIOGRAFIA E BIBLIOTECONOMIA
M-GGR	M-GGR/01 GEOGRAFIA
M-PED	M-PED/01 PEDAGOGIA GENERALE E SOCIALE
M-DEA	M-DEA/01 DISCIPLINE DEMOETNOANTROPOLOGICHE
M-EDF	M-EDF/01 METODI E DIDATTICHE DELLE ATTIVITA MOTORIE
Area 14	
AD	**SSD**
SPS	SPS/04 SCIENZA POLITICA
	SPS/07 SOCIOLOGIA GENERALE DINAMICA
	SPS/08 SOCIOLOGIA DEI PROCESSI CULTURALI E COMUNICATIVI
	SPS/09 SOCIOLOGIA DEI PROCESSI ECONOMICI E DEL LAVORO
	SPS/11 SOCIOLOGIA DEI FENOMENI POLITICI

2) *Differenziazione autoriale ed editoriale*.

Riguardo alle voci enciclopediche, per non attingere unicamente dalle fonti allestite dall'*Enciclopedia Treccani* on line, ne sono state campionate alcune tratte da enciclopedie cartacee, poi acquisite con OCR. Purtroppo la scrittura in colonna sugli originali ha rallentato di molto la trascrizione e, considerata la validità del materiale già digitalizzato, molto più agevole e sorvegliato, si è preferito limitare il processo di riproduzione da stampa, sia per non superare i tempi di allestimento prestabiliti, sia perché l'operazione

manuale avrebbe incrementato consistentemente la probabilità di riversare involontariamente eventuali refusi nel corpus. Talvolta, tuttavia, si è ricorsi a risorse diverse anche per coprire settori difficilmente rintracciabili nell'*Enciclopedia Treccani* on line, almeno con i caratteri prefissati e attesi. È il caso di L-ANT/04 (Numismatica) e M-STO/06 (Storia delle religioni), che era necessario inserire per assicurare l'assortimento disciplinare di cui si è detto, nonostante non fossero rappresentati in rete in contenuti che vantassero il rispetto dei restanti criteri enunciati (estensione moderata, varietà autoriale, inequivocabile affiliazione dello scrivente al settore ricercato, contenuto linguistico idoneo alla ricerca).

3) *Contenuto linguistico.*

Potendo disporre per le tre aree trattate di un cospicuo materiale in lingua italiana, nel subcorpus umanistico non sono state impiegate traduzioni.

6. CONCLUSIONI

Nel contributo abbiamo cercato di illustrare i progressi fin qui compiuti dal progetto DIA nella descrizione del lessico accademico e la ramificazione dei percorsi di ricerca.

Definire questo insieme di parole richiede anzitutto di andare al di là della creazione di liste, cercando invece di adottare un approccio sistematicamente funzionale. Questo tipo di approccio è a sua volta limitato, ad oggi, dal fatto che non tutti i tipi di funzioni comunicative e testuali sono state studiate e descritte con la stessa accuratezza (è il concetto di «regolarità nascoste» di cui si parla in Mastrantonio, 2023b). Un aiuto in questo senso sembra dato dalla possibilità di isolare una nuova tipologia basata sul contenuto e sulla sua organizzazione (il modo "concettuale"), che permette di dare conto di molte manifestazioni linguistiche che prescindono dallo scopo comunicativo (tipicamente, tutto il lessico del classificare).

Anche dal punto di vista sociolinguistico, la ricerca ci sembra che ponga problemi teorici non di poco conto; se da un lato il lessico accademico sembra configurarsi come un registro, dall'altro lato si discosta nettamente dai sottocodici, benché ne costituisca il supporto indispensabile nelle pratiche discorsive formali (cfr. Mastrantonio, 2021: 351): infatti la polisemia delle parole accademiche e la loro non referenzialità (§ 3) sono tratti nettamente divergenti rispetto ai linguaggi settoriali.

Abbiamo cercato di mostrare come la nostra ricerca si giovi in modo equilibrato di approcci qualitativi e quantitativi: i primi sono basati sulla lettura estensiva dei testi e sulle analisi semantiche e formali di dettaglio, mentre i secondi sfruttano l'apporto dei dati lessicometrici resi disponibili dai procedimenti della linguistica computazionale (§ 4). Tra le analisi semantiche, andrà menzionata l'emersione di potenziali metafore concettuali come "Comprensione è Visione", "Ricerca è Esplorazione", "Teoria è Costruzione"; l'identificazione e lo studio di queste metafore può avere importanti ricadute, dal momento che queste strutture cognitive permettono di concettualizzare e comunicare idee complesse. Appare chiaro che tutti questi scopi possono essere raggiunti solo basandosi su un corpus che sia sufficientemente ampio e rappresentativo, come quello messo a punto e illustrato nel § 5.

Al momento di licenziare questo contributo, l'unità di Venezia sta lavorando a un prototipo di database e di sito web (progettato col supporto di Lorenzo Sartori e Andrea Boscolo Zemello dell'Area Servizi Informatici e Telecomunicazioni di Ca' Foscari) che ospiti il Dizionario e che consenta accessi multipli ai suoi contenuti, cioè per forme, per funzioni comunicative, per piani del testo e altro ancora. Quando il Dizionario sarà

popolato di dati e reso consultabile, crediamo che ne potranno risultare vantaggi in più direzioni: non solo per quel che riguarda le applicazioni didattiche (potenziamento della comprensione e della produzione nei segmenti dell'istruzione scolastica e universitaria), ma anche per la possibilità di intraprendere con più sicurezza studi sulla struttura e sulla natura dei generi testuali della comunicazione colta e specialistica, non solo in sincronia.

RIFERIMENTI BIBLIOGRAFICI

Antonelli G., Motolese M., Tomasin L. (2021) (a cura di), *Storia dell'italiano scritto*, vol. V: *Testualità*, Carocci, Roma.

Biber D. (1993), "Representativeness in corpus design", in *Literary and Linguistic Computing*, 8, pp. 243-257.

Cresti E., Moneglia M. (2016), "La linguistica italiana dei corpora", in Lubello S. (a cura di), *Manuale di Linguistica italiana*, de Gruyter, Berlin, pp. 581-611.

Cummins J. (1979), "Cognitive/Academic Language Proficiency, Linguistic Interdependence, the Optimum Age Question and Some Other Matters", in *Working Papers on Bilingualism*, 19, pp. 121-129.

D'Aguanno D. (2019a), "Il lessico accademico per l'insegnamento della scrittura nelle scuole superiori", in Palermo M., Salvatore E. (a cura di), *Scrivere nella scuola oggi. Obiettivi, metodi, esperienze*, Atti del II Convegno ASLI Scuola (Siena, Università per Stranieri, 12-14 ottobre 2017), Franco Cesati Editore, Firenze, pp. 93-106.

D'Aguanno D. (2019b), *Insegnare l'italiano scritto. Idee e modelli per la didattica nelle scuole superiori*, Carocci, Roma.

Della Casa M. (1994), *Scrivere testi*, La Nuova Italia, Firenze.

De Mauro T. (1988), "Linguaggi scientifici e lingue storiche", in Guerriero A. R. (a cura di), *L'educazione linguistica e i linguaggi delle scienze*, La Nuova Italia, Firenze, pp. 9-19.

De Mauro T. (1999), "Introduzione", in Id. (diretto da), *Grande dizionario italiano dell'uso*, con la collaborazione di G. Lepschy ed E. Sanguineti, 6 voll. più l'appendice *Nuove parole italiane dell'uso*, UTET, Torino, 1999-2003.

Ferrari A. (2014), *Linguistica del testo*, Carocci, Roma.

Ferreri S. (2005), *L'alfabetizzazione lessicale: studi di linguistica educativa*, Aracne, Roma.

Feyerabend P. (1979), *Contro il metodo: abbozzo di una teoria anarchica della conoscenza*, Feltrinelli, Milano (prima ed. it. 1973).

Fiorentino G. (2015), "Aspetti problematici del discorso accademico", in *Cuadernos de Filología Italiana*, 22, pp. 264-84.

Gallina F. (2018), "Studenti internazionali in mobilità. La questione del lessico della conoscenza in italiano L2", in Coonan C.M. *et al.* (a cura di), *La didattica delle lingue nel nuovo millennio. Le sfide dell'internazionalizzazione*, Edizioni Ca' Foscari - Digital Publishing, Venezia, pp. 323-339.

Gledhill C. (1995), "Collocation and genre analysis. The discourse function of collocation in cancer research abstracts and articles", in *Zeitschrift fu Èr Anglistik und Amerikanistik*, 43, 1/1, pp. 11-36: https://u-paris.hal.science/hal-01220327v1.

Gregory T. (1985), "Il Lessico Intellettuale Europeo", in Cicala M. C. (a cura di), *Lo storico e il suo lessico*, Atti del Convegno di Prato, 1-3 aprile 1982, La Grafica, Messina, pp. 3-13.

Gualdo R., Telve S. (2011), *Linguaggi specialistici dell'italiano*, Carocci, Roma.

Halliday M.A.K (1994[2]), *Introduction to functional grammar*, Edward Arnold, London.

Halliday M.A.K (2004), *The Language of Science*, Continuum, London-New York.

Hatim B. (1984), "A Text-Typological Approach to Syllabus Design in Translator Training", in *The Incorporated Linguist*, 23, 3, pp. 146-149.

Hunston S. (2008), "Collection strategies and design decisions", in Lüdeling A., Kytö M. (eds.), *Corpus linguistics. An international handbook*, vol. 1, De Gruyter Mouton, Berlin-New York, pp. 154-167.

Hunston S. (2012), "Corpora and language teaching: Specific applications", in Ead., *Corpora in Applied Linguistic*s, Cambridge University Press, Cambridge, pp. 198-212.

Jakubíček M., Kilgarriff A., Kovář V., Rychlý P., Suchomel V. (2020), *Italian Web 2020 (itTenTen20). Italian Corpus*, Lexical Computing CZ s.r.o.: https://www.sketchengine.eu/ittenten-italian-corpus/.

Klimová K. (2016), "Acquisizione dell'italiano accademico in contesto LS e in prospettiva comparativa", in *Études romanes de Brno*, 37, pp. 245-57.

Koester A. (2022), "Building small specialized corpus", in O'Keeffe A., McCarthy M. J. (eds.), *TheRoutledge handbook of corpus linguistics*. Second edition, Routledge, London; New York, pp. 48-61.

Kosso P. (2011), *A Summary of Scientific Method*, Springer, Dordrecht-Heidelberg-London-New York.

Langacker R.W. (2013), *Essential of Cognitive Grammar*, Oxford University Press, Oxford.

Lyding V., Stemle E., Borghetti C., Brunello M., Castagnoli S., Dell'Orletta F., Dittmann H., Lenci A., Pirrelli V. (2014), *Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati (PAISÀ). Corpus of Italian Web Texts*, Eurac Research CLARIN Centre: http://hdl.handle.net/20.500.12124/7.

Mastrantonio D. (2021a), "L'italiano scritto accademico: problemi descrittivi e proposte didattiche", in *Italiano LinguaDue*, 13, 1, 348-368: https://riviste.unimi.it/index.php/promoitals/article/view/15871.

Mastrantonio D. (2021b), "Connettivi", in Antonelli *et al.* (2021), pp. 221-258.

Mastrantonio D. (2023a), "Sul cloze mirato e semplificato nella didattica del registro accademico", in Id., Salvatore E. (a cura di), *Forme, strutture e didattica dell'italiano. Studi per i 60 anni di Massimo Palermo*, Edizioni Università per Stranieri, Siena, pp. 103-114: https://edizioni.unistrasi.it/1277/1720/Sul_cloze_mirato_e_semplificato_nella_didattica_del_registro_accademico_-_p._103.htm.

Mastrantonio D. (2023b), "Le regolarità nascoste: aspetti comunicativi sottorappresentati nella didattica L2/LS", in *Italiano LinguaDue*, 15, 1, pp. 343-358: https://riviste.unimi.it/index.php/promoitals/article/view/20397.

Mauri C., Ballarè S., Goria E., Cerruti M., Suriano F. (2019), *KIParla Corpus*, University of Bologna: http://kiparla.it.

Mazzoleni M. (2004), "Dai tipi ai generi: una tipologia testuale in chiave di didattica della traduzione", in D'Achille P. (a cura di), *Generi, architetture e forme testuali*, Atti del 7. Convegno SILFI, Società internazionale di linguistica e filologia Italiana (Roma, 1-5 ottobre 2002), Franco Cesati Editore, Firenze, pp. 401-413.

McCarthy M., O'Dell F. (2008), *Academic vocabulary in use*, Cambridge University Press, Cambridge.

McEnery T., Brookes G. (2022), "Building a written corpus: what are the basics?", in O'Keeffe A., McCarthy M.J. (eds.), *The Routledge handbook of corpus linguistics*. Second edition, Routledge, London; New York, pp. 35-47.

Nuovo De Mauro = *Nuovo Dizionario De Mauro* (disponibile on line all'indirizzo https://dizionario.internazionale.it/).

Palermo M. (2013), *Linguistica testuale dell'italiano*, il Mulino, Bologna.

Palermo M. (2021), "La dimensione testuale", in Antonelli *et al.* (2021), pp. 17-56.

Piccioni S., D'Angelo M., Ferro M. C. (2021), "I Corpora SEAH di comunicazione specializzata nel settore dell'Architettura e delle Costruzioni. Struttura, compilazione e usi", in *Linguistica*, 61, 2, pp. 99-125.

Rees G. (2022), "Using corpora to write dictionaries", in O'Keeffe A., McCarthy M. J. (eds.), *The Routledge handbook of corpus linguistics*. Second edition, Routledge, London; New York, pp. 387-404.

Restivo L. (2022), "L'italiano scritto degli studenti universitari: prime osservazioni sul corpus UniverS-Ita", in *Italiano LinguaDue*, 14 (2022), 2, pp. 797-818: https://riviste.unimi.it/index.php/promoitals/article/view/18328.

Sabatini F. (1999), "«Rigidità-esplicitezza» vs «elasticità-implicitezza»: possibili parametri massimi per una tipologia dei testi", in Skytte G., Sabatini F. (a cura di), *Linguistica testuale comparativa. In memoriam Maria Elisabeth Conte*, Atti del Convegno interannuale della Società di Linguistica Italiana (Copenaghen, 5-7 febbraio 1998), Copenaghen, Museum Tusculanum Press, pp. 141-172; rist. in Id., *L'italiano nel mondo moderno. Saggi scelti dal 1968 al 2009*, a cura di Coletti V., Coluccia R., D'Achille P., De Blasi N., Proietti D., Liguori, Napoli, 2011, vol. II, pp. 183-216.

Sabatini F., Coletti V. (2018), *Il Sabatini Coletti - Dizionario della lingua italiana* (disponibile on line all'indirizzo https://dizionari.corriere.it/dizionario_italiano/).

Serianni L. (2019), "Scrivere per imparare a leggere. Riflessioni sulla scrittura nella scuola", in Palermo M., Salvatore E. (a cura di), *Scrivere nella scuola oggi. Obiettivi, metodi, esperienze*, Atti del secondo Convegno nazionale ASLI Scuola, Siena, Università per Stranieri, 12-14 ottobre, Franco Cesati Editore, Firenze, pp. 21-36.

Spina S. (2010), "AIWL: una lista di frequenza dell'italiano accademico", in Bolasco S., Chiari I., Giuliano L. (a cura di), *Statistical Analysis of Textual Data*, Proceedings of the 10th International Conference "Journées d'Analyse statistique des Données Textuelles" (9-11 June 2010 - Sapienza University of Rome, LED, Milano, pp. 1317-1325.

Spina S. (2014), "Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione", in Basili R., Lenci A., Magnini B. (a cura di), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014*. vol. 1, Pisa University Press, Pisa, pp. 354-359.

Swales J. (2004), *Research genres. Explorations and applications*, Cambridge University Press, Cambridge.

Szudarski P. (2018), *Corpus linguistics for vocabulary. A Guide for Research*, Routledge, New York.

Treccani = *Vocabolario Treccani* on line: https://www.treccani.it/vocabolario/.

Werlich E. (1975), *Typologie der Texte. Entwurf eines Textlinguistischen Modells zur Grundlegung einer Textgrammatik*, Quelle & Meyer, Heidelberg.

Załęska M. (2016) (a cura di), *Il discorso accademico italiano: temi, domande, prospettive*, Lang, Frankfurt am Main.

Zingarelli (2008) = *Lo Zingarelli 2008. Vocabolario della lingua italiana*, Zanichelli, Bologna.

APPENDICE

Testi nel corpus DIA pilota

1. (Manuale Univ.) Acquaviva F. (2013), *Il nome*, Carocci, Roma.
2. (Articolo) Benigni V., Bernasconi B. (2023), "Non siamo queste esperte". Usi valutativi dei deittici prossimali e distali, in *Italiano LinguaDue*, 15, 2, pp. 38-50: https://riviste.unimi.it/index.php/promoitals/article/view/22014.
3. (Manuale Univ.) Bellissima F. (2008), *Fondamenti di Matematica*, Carocci, Roma.
4. (Manuale Univ.) Bianchi P.(2018), *4.0 La nuova rivoluzione industriale*, il Mulino, Bologna.
5. (Articolo) Ferrario C. (2021), "L'imposta di soggiorno: vincolo o opportunità per il settore turistico?", in *Il diritto degli affari*, XI, 2, pp. 176-186.
6. (Manuale Univ.) Luzi L. (2009), *Biologia cellulare nell'esercizio fisico*, Cortina, Milano.
7. (Articolo) Paleari R. (2010), "Anemia a cellule falciformi e sindromi correlate: aggiornamenti e prospettive", in *Biochimica clinica*, 34, 4, pp. 250-258.
8. (Articolo) Tettamanzi A. (2005), "Algoritmi evolutivi: concetti ed applicazioni", in *Mondo digitale*, 1, pp. 3-17.

Testi nel corpus DIA

Macroarea giuridico-economica (262.681 parole token[41])

Area 12 - Scienze giuridiche (131.548 parole token)

Manuali universitari:

1. Bellantuono G. (2009), *Contratti e regolazione nei mercati dell'energia*, il Mulino, Bologna (SSD IUS/02 DIRITTO PRIVATO COMPARATO).
2. Bosi G. (2017), *L'impresa culturale Diritto ed economia delle attività creative*, il Mulino, Bologna (SSD IUS/04) DIRITTO COMMERCIALE).
3. Cafaggi F. (a cura di) (2004), *Reti di imprese tra regolazione e norme sociali*, il Mulino, Bologna, Bologna (SSD IUS/01 DIRITTO PRIVATO).
4. Iudica G., Zatti P. (2004[5]), *Linguaggio e regole del diritto privato*, CEDAM, Padova (SSD IUS/01 DIRITTO PRIVATO).
5. Marchesi D. (2010), *Litiganti, avvocati e magistrati. Diritto ed economia del processo civil*e, il Mulino, Bologna (SSD IUS/14 DIRITTO CIVILE - SECS-P/01 ECONOMIA POLITICA).
6. Moriconi L. (2023), *Elementi e materiali di diritto doganale*, Milano University Press, Milano (SSD IUS/14 DIRITTO DELL'UNIONE EUROPEA).
7. Stea G. (2014), I *Principi di diritto penale nella giurisdizione europea*, Pisa University Press, Pisa (SSD IUS/17 DIRITTO PENALE).
8. Zaccaria R. (2004[5]), *Diritto dell'informazione e della comunicazione*. Con la collaborazione di Valastro A., CEDAM, Padova (IUS/08 DIRITTO COSTITUZIONALE).

Voci enciclopediche:

1. Ciucciovino S. (2016), "Prescrizione [dir. lav.]", in *Diritto on line*: https://www.treccani.it/enciclopedia/prescrizione-dir-lav_(Diritto-on-line)/ (SSD IUS/07 DIRITTO DEL LAVORO).
2. Galgano F.(2001), "Lex mercatoria", in *Enciclopedia del Diritto* - Aggiornamento V, Giuffrè Francis Lefebvre, Milano, pp. 721-729 (SSD IUS/05 DIRITTO DELL'ECONOMIA).
3. Gamba C. (2016), "Presunzioni [dir. proc. civ.]", in *Diritto on line*, 2016: https://www.treccani.it/enciclopedia/presunzioni-dir-proc-civ_(Diritto-on-line)/ (SSD IUS/15 DIRITTO PROCESSUALE CIVILE).
4. Manacorda S. (2014), "Diritto penale europeo": https://www.treccani.it/enciclopedia/diritto-penale-europeo_(Diritto-on-line. (SSD IUS/17 DIRITTO PENALE)

[41] La presente estensione e le successive sono misurate con AntConc.

5. Marrella F. (2018), "Investimenti [dir. Inter.]", in *Diritto on line* (SSD IUS/13 DIRITTO INTERNAZIONALE: https://www.treccani.it/enciclopedia/investimenti-dir-int_(Diritto-on-line)/.
6. Nappi P. (2008), "Processo agrario", in *Enciclopedia del Diritto* - Annali II, t. 1, Giuffrè Francis Lefebvre, Milano, pp. 617-647 (SSD IUS/03 DIRITTO AGRARIO).
7. Punzi A. (2002), "Filosofia del diritto", in *Enciclopedia del Diritto* - Aggiornamento VI, Giuffrè Francis Lefebvre, Milano, pp. 1163-1176 (SSD IUS/20 FILOSOFIA DEL DIRITTO).
8. Ruggieri F. (2008), "Pubblico ministero (diritto processuale penale)", in *Enciclopedia del Diritto* - Annali II, t. 1 – Giuffrè Francis Lefebvre, Milano, pp. 998-1015 (SSD IUS/16 DIRITTO PROCESSUALE PENALE).

Articoli:

1. Antonioli M. L. (2023), "Buona fede, affidamento e responsabilità civile della P.A. nel nuovo codice dei contratti pubblici", in *Ceridap*, 4, pp. 1-23 (SSD IUS/ 10 DIRITTO AMMINISTRATIVO).
2. Carullo G. (2010), "Appalti: il termine di ricorso fra incertezze interpretative e riforme legislative", in *Rivista italiana di diritto pubblico comunitario*, 2, pp. 553-565 (SSD IUS/10 DIRITTO AMMINISTRATIVO; SSD IUS/14 DIRITTO DELL'UNIONE EUROPEA).
3. Galetta D.U. (2016), "Accesso civico e trasparenza della Pubblica Amministrazione alla luce delle (previste) modifiche alle disposizioni del D. Lgs. n. 33/2013", in *Federalismi.it*, 5, pp. 1-19. (SSD IUS/10 DIRITTO AMMINISTRATIVO).
4. Masera L., Viganò F. (2013), "Abolire la Bossi-Fini?", in *Diritto penale contemporaneo*, (14 ottobre), pp. 1-10 (SSD IUS/17 DIRITTO PENALE).
5. Ramajoli M. (2016), "A proposito di codificazione e modernizzazione del diritto amministrativo", in *Rivista trimestrale di diritto pubblico*, 66, 2, pp. 347-366 (SSD IUS 10 DIRITTO AMMINISTRATIVO).
6. Vanoni L.P. (2012), "Discriminazione sul luogo di lavoro e autonomia delle organizzazioni religiose in USA: il caso Hosanna Tabor Evangelical Lutheran Church and School V. EEOC", in *Rivista aic*., 2, pp. 1-10 (SSD IUS/08 DIRITTO COSTITUZIONALE e SSD IUS/21 DIRITTO PUBBLICO COMPARATO).
7. Zanon N. (2017), "Corte Costituzionale, evoluzione della "coscienza sociale", interpretazione della Costituzione e diritti fondamentali: questioni e interrogativi a partire da un caso paradigmatico", in *Rivista aic*, 4, pp. 1-16 (SSD IUS/08 - DIRITTO COSTITUZIONALE.
8. Zendri C. (2015), "Diritto feudale – diritto canonico – diritto pubblico. Studi recenti e prospettive di ricerca", in *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte: Kanonistische Abteilung*, 101, pp. 389-398 (SSD IUS/19 STORIA DEL DIRITTO MEDIEVALE E MODERNO).

Area 13 - Scienze economiche e statistiche (131.133 parole token)

Manuali universitari:

1. Bowles S., Edwards R., Roosevelt F. (2011), *Introduzione all'economia politica. Le dinamiche del capitalismo*, Springer (traduzione di Vatiero M., Di Trocchio S., Niglia G.), Springer, Milano, (SSD SECS-P/01 ECONOMIA POLITICA).
2. Bruni L., Zamagni S. (2004), *Economia civile. Efficienza, equità, felicità pubblica*, il Mulino, Bologna (SSD SECS-P/01 ECONOMIA POLITICA).
3. Ciappei C., Citti P., Bacci N., Campatelli G. (2006), *La metodologia Sei Sigma nei servizi: un'applicazione ai modelli di gestione finanziaria*, University Press, Firenze (SSD SECS-P/07 ECONOMIA AZIENDALE; SSD SECS-P/08 ECONOMIA E GESTIONE DELLE IMPRESE).
4. De Bonis R., Vangelisti M. I. (2019), *La moneta. Dai buoi di Omero ai Bitcoin*, il Mulino, Bologna (SECS-P/12 STORIA ECONOMICA).
5. Dominick S. (2016), *Economia internazionale. Il Commercio internazionale*, Zanichelli, Bologna (SSD SECS-P/01 ECONOMIA POLITICA).
6. Frank R. H., Cartwright E. (2021), *Microeconomia*; edizione italiana a cura di Piras R., McGraw-Hill, Milano (SSD SECS-P/01 ECONOMIA POLITICA).

7. List D.A., Laibson D., John A. (2021[2]), *Principi di economia politica: Teoria ed evidenza empirica* (a cura di Macchi M. e Ricchiuti G.), Pearson Education Resources Italia S.R.L, Milano, (SSD SECS-P/01 ECONOMIA POLITICA).
8. Mankiw G. (2004[3]), *Principi di economia*, Zanichelli, Bologna (SSD SECS-P/01 ECONOMIA POLITICA).

Voci enciclopediche:

1. Ciciretti R., Carbonari G., Solinas M., Falsetti E. (2011), "Commissione nazionale per le società e la borsa (CONSOB)", in *ASSONEBB. Associazione Nazionale Enciclopedia della Banca e della Borsa*: https://www.bankpedia.org/termine.php?c_id=19250 (SSD SECS-P/02 POLITICA ECONOMICA).
2. De Cecco M. (2004), "Mercati finanziari", in *Enciclopedia del Novecento*, III Supplemento: https://www.treccani.it/enciclopedia/mercati-finanziari_(Enciclopedia-del-Novecento)/ SSD SECS-P/03 SCIENZA DELLE FINANZE).
3. Di Donato L. (2012), "Impact Assessment Board", in *ASSONEBB*: https://www.bankpedia.org/ricerca.php (SSD SECS-P/02 POLITICA ECONOMICA).
4. Petretto A. (2004), "Privatizzazione e regolamentazione", in *Enciclopedia del Novecento*, III Supplemento: https://www.treccani.it/enciclopedia/privatizzazione-e-regolamentazione_(Enciclopedia-del-Novecento)/ (SSD SECS-P/03 SCIENZA DELLE FINANZE).
5. S.a. (2010), "Denaro (enciclopedia)", in *ASSONEB* https://www.bankpedia.org/ricerca.php (SECS-P/12 STORIA ECONOMICA).
6. Sacco P. L. (2004), "Economia", in *Enciclopedia del Novecento*, III Supplemento: https://www.treccani.it/enciclopedia/economia_res-a73dd417-87f0-11dc-8e9d-0016357eee51_(Enciclopedia-del-Novecento)/ (SSD SECS-P/02 POLITICA ECONOMICA).
7. Vittorioso G., Anchino S. (2011), "Impairment test (enciclopedia)", in *ASSONEBB:* https://www.bankpedia.org/termine.php?c_id=20487 (SSD SECS-P/09 FINANZA AZIENDALE).
8. Zenezini M. (2004), "Decentramento produttivo", in *Enciclopedia del Novecento*: https://www.treccani.it/enciclopedia/decentramento-produttivo_(Enciclopedia-del-Novecento) (SSD SECS-P/0 2 POLITICA ECONOMICA).

Articoli:

1. Coveri A., Cozza C., Guarascio D. (2022), "Il capitale monopolistico ai tempi delle grandi piattaforme digitali", in *Economia & Lavoro*, LVI, 1, pp. 49-63 (SSD SECS-P/02 POLITICA ECONOMICA; SSD SECS-P/06 ECONOMIA APPLICATA).
2. Gobbi L. (2023), "De-dollarizzazione: la sfida dei paesi BRICS", in *Moneta e Credito*, 76 (304), pp. 357-372 (SSD SECS-P/01 ECONOMIA POLITICA).
3. Marzana D., Poy S., Rosina A., Sironi, E. (2018), "Alternanza scuola-lavoro e sviluppo delle soft skills: un'indagine sulle attese dei giovani", in *Ricercazione*, 10, 2, pp. 187-2017 (SSD SECS-S/04 DEMOGRAFIA).
4. Pavione E., Gazzola P., Sciarini L. (2023), "Le certificazioni B Corp nelle imprese agroalimentari: analisi di due casi studio", in *Economia Aziendale Online – Business and Management Sciences International Quarterly Review*, 14, 3, pp. 721-736 (SSD SECS-P/07 ECONOMIA AZIENDALE).
5. Ruffini R., Ingaggiati M., Barbato G. (2022), "Processi di assunzione del personale nella pubblica amministrazione: analisi strategica dei comportamenti burocratici", in *RU. Risorse umane nella pubblica amministrazione*, 2, pp. 63-78. (SSD SECS-P/10 ORGANIZZAZIONE AZIENDALE - SSD SECS-P/07 ECONOMIA AZIENDALE).
6. Sbrana F. (2020), "La grande scossa. Il terremoto dell'Irpinia tra economia, politica e territorio", in *Rivista economica del Mezzogiorno*, 34, 4, pp. 839-850 (SSD SECS-P12_STORIA ECONOMICA).

7. Shaba E., Lazazzara A., Solari L., Delle Fave A. (2023), "Industria 4.0: come cambia il lavoro con l'introduzione dei cobot nei processi produttivi manifatturieri", in *Prospettive in organizzazione*, 23, pp. 1-12 (SSD SECS-P/10 ORGANIZZAZIONE AZIENDALE).
8. Zanderighi L., Orsi L. (2020), "Commercio di vicinato e resilienza", in *Micro & Macro marketing*, 29, 2, pp. 377-394 (SSD SECS-P/06 ECONOMIA APPLICATA; SSD SECS-P/08 ECONOMIA E GESTIONE DELLE IMPRESE).

Macroarea scientifica (254.404 parole token)

Area 01 - Scienze matematiche e informatiche (33.443 parole token)

Manuali universitari:

1. Abate M., Tovena F. (2011), *Geometria differenziale*, Springer, Milano (SSD MAT/03 – GEOMETRIA).
2. Canuto C., Tabacco A. (2014), *Analisi Matematica I. Teoria ed esercizi*, Springer, Milano (SSD MAT/05 ANALISI MATEMATICA).

Voci enciclopediche:

1. Dieudonné J. A. (2007), "Analisi matematica", in *Enciclopedia della scienza e della tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/analisi-matematica_(Enciclopedia-della-Scienza-e-della-Tecnica)/ (SSD MAT/05 ANALISI MATEMATICA)
2. Kobayashi, S., Salamon S. (2007), "Geometria differenziale", in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/geometria-differenziale_(Enciclopedia-della-Scienza-e-della-Tecnica)/ 2007 (SSD MAT/03 GEOMETRIA).

Articoli:

1. Andreatta F., "Forme modulari e valori speciali di funzioni L" (2021), in *Bollettino della unione matematica italiana*, pp. 1-13 (SSD MAT/02 ALGEBRA; SSD MAT/03 GEOMETRIA).
2. Zappa G. (2002), "Sui gruppi finiti i cui sottogruppi non normali hanno tutti lo stesso ordine", in *Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali*. Rendiconti Lincei. Matematica e Applicazioni, 9, 13, pp. 5-16 (SSD MAT/02 ALGEBRA).

Area 02 - Scienze fisiche (31.743 parole token)

Manuali universitari:

1. Silvestrini V., Mencuccini C. (2017), *Fisica: elettromagnetismo e ottica: con esempi ed esercizi*, Casa Editrice Ambrosiana, Rozzano (SSD FIS/01 FISICA SPERIMENTALE).
2. Cicogna G. (2008), *Metodi matematici della Fisica*, Springer, Milano (SSD FIS/02 FISICA TEORICA, MODELLI E METODI MATEMATICI).

Voci enciclopediche:

1. Fiorini E. (2007), "Neutrini solari", in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/neutrini-solari_(Enciclopedia-della-Scienza-e-della-Tecnica)/ (SSD FIS/04 FISICA NUCLEARE E SUBNUCLEARE).
2. Peruzzi G., Regge T. (2003), "Gravitazione", in *Storia della scienza*, Treccani, Roma: https://www.treccani.it/enciclopedia/la-grande-scienza-gravitazione_(Storia-della-Scienza)/ (SSD FIS/02 FISICA TEORICA, MODELLI E METODI MATEMATICI).

Articoli:

1. Bellini G., Miramonti L. Ranucci, G. (2008), "Borexino: un rivelatore unico per lo studio dell'oscillazione dei neutrini di bassissima energia", in *Il nuovo saggiatore*, 24, 3, pp. 46-58 (SSD FIS/04 FISICA NUCLEARE E SUBNUCLEARE).

2. Modena I. (1997), "La ricerca delle onde gravitazionali", in *Il nuovo cimento,* 20, 5, pp. 629-636 (SSD FIS/01 FISICA SPERIMENTALE).

Area 03 - Scienze chimiche (36.245 parole token)

Manuali universitari:

1. Hart H., Craine L. E., Hart D. J., Hadad C. M. (2012[7]), *Chimica organica*, Zanichelli, Bologna (SSD CHIM/06 CHIMICA ORGANICA).
2. Moret S., Purcaro G., Conte L. S. (2015), *Il campione per l'analisi chimica. Tecniche innovative e applicazioni nei settori agroalimentare e ambientale*, Springer, Milano (SSD CHIM/10 CHIMICA DEGLI ALIMENTI).
3. Salvatore F., Filippelli F., Salvatore M. M. (2018), *Fondamenti di Chimica Analitica*, Aracne, Roma (SSD CHIM/01 CHIMICA ANALITICA).

Voci enciclopediche:

1. Corradini P. (2007), "Polimeri", in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/polimeri_(Enciclopedia-della-Scienza-e-della-Tecnica)/ (SSD CHIM/06 CHIMICA ORGANICA).
2. S.a. "Catalisi", in *Enciclopedia Treccani*, Treccani Roma: https://www.treccani.it/enciclopedia/catalisi/ (SSD CHIM/04 CHIMICA INDUSTRIALE).

Articoli:

1. Barone G., Varriale L. (2013), "Modellizzazione dell'inquinamento atmosferico", in *La chimica e l'industria*, 1, pp. 134-140 (SSD CHIM/02 CHIMICA FISICA).
2. Scrocco E. (2004), "Un metodo elettrochimico per lo studio dei flussi liquidi nello xilema collegati al funzionamento degli stomi", in *Rendiconti Lincei. Scienze Fisiche e Naturali*, 9,15, pp. 133-149 (SSD CHIM/02 CHIMICA FISICA).

Area 04 - Scienze della Terra (27.645 parole token)

Manuali universitari:

1. Dongarrà G., Varrica D. (2004), *Geochimica e ambiente*, EdiSES, Napoli (SSD GEO/08 GEOCHIMICA E VULCANOLOGIA).
2. Klein C. (2004), *Mineralogia*, Zanichelli, Bologna (SSD GEO/06 MINERALOGIA).

Voci enciclopediche:

1. Malanotte-Rizzoli P. (2007), "Oceanologia", in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/oceanologia_(Enciclopedia-della-Scienza-e-della-Tecnica)/ (SSD GEO/12 OCEANOGRAFIA E FISICA DELL'ATMOSFERA)
2. Panza G. F. (2007), "Onde sismiche", in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/onde-sismiche_(Enciclopedia-della-Scienza-e-della-Tecnica)/ (SSD GEO/11 GEOFISICA APPLICATA)

Articoli:

1. De Casa G., Lombardi G. (2007), "Caratteri fisico-meccanici del Tufo Giallo della Via Tiberina (Roma)", in *Rendiconti Lincei. Scienze Fisiche e Naturali*, Serie 9, 18, pp. 5-25 (SSD GEO/04 GEOGRAFIA FISICA E GEOMORFOLOGIA.
2. Leoni L., Lezzerini M., Saitta M. (2008), "Calcolo computerizzato dell'analisi mineralogica quantitativa di rocce e sedimenti argillosi basato sulla combinazione dei dati chimici e diffrattometrici", in *Atti della Società Toscana di Scienze Naturali, Memorie*, Serie A, 113, pp. 63-69 (SSD GEO/02 GEOLOGIA STRATIGRAFICA E SEDIMENTOLOGICA).

Area 05 - Scienze biologiche (37.876 parole token)

Manuali universitari:

1. Capasso F. (2011), *Farmacognosia, Botanica, chimica e farmacologia delle piante medicinali*, Springer, Milano (SSD BIO/14 FARMACOLOGIA).
2. Dehò G., Galli E. (2019), *Biologia dei microrganismi*, Casa Editrice Ambrosiana, Rozzano (SSD BIO/19 MICROBIOLOGIA GENERALE).

Voci enciclopediche:

1. Ingegnoli V. (2007), "Ecologia del paesaggio", in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/ecologia-del-paesaggio_(Enciclopedia-della-Scienza-e-della-Tecnica)/ (SSD BIO/07 ECOLOGIA).
2. Meldolesi J. (2007), "Cellula. Fisiologia e patologia della cellula", in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/cellula-fisiologia-e-patologia-della-cellula_(Enciclopedia-della-Scienza-e-della-Tecnica)/ (SSD BIO/14 FARMACOLOGIA).
3. Chiurazzi M. (2007), "Genetica molecolare", in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/genetica-molecolare_(Enciclopedia-della-Scienza-e-della-Tecnica)/ (SSD BIO/18 GENETICA).

Articoli:

1. Bussotti F., Bettini D., Carrari D., Selvi F., Pollastrini M. (2022), "Cambiamenti climatici in atto: osservazioni sugli impatti degli eventi siccitosi sulle foreste toscane", in *Forest@. Rivista di selvicoltura ed ecologia forestale*, 20, pp. 1-9 (SSD BIO/03 BOTANICA AMBIENTALE E APPLICATA).
2. Gatti R., Codemo C., De Palo E. F. (2010), "Il cortisolo nelle differenti matrici biologiche: dalla biochimica all'analisi di laboratorio", *Biochimica clinica*, 34, 6, pp. 591-599 (SSD BIO/12 BIOCHIMICA CLINICA E BIOLOGIA MOLECOLARE CLINICA).
3. Salmaso F., Quadroni S., Romanò A., Compare S., Gentili G., Crosa G. (2014), "Definizione dello stato ecologico secondo il D.M. 260/2010 in due fiumi di pianura (Adda e Ticino) interessati dal Deflusso Minimo Vitale", in *Biologia ambientale*, 28, 1, pp. 25-37 (SSD BIO/07 ECOLOGIA).

Area 06 - Scienze mediche (48.025 parole token)

Manuali universitari:

1. Moroni M, Esposito R., Antinori S. (2014), *Malattie infettive. Ottava edizione*, Edra, Milano (SSD MED/17 MALATTIE INFETTIVE).
2. Rossi A., Amore M., Carpiniello B., Fagiolini A., Maina G., Vita A. (2019), *Manuale di psichiatria*, Edra, Milano (SSD MED/25 PSICHIATRIA).
3. Rugarli C. (2015), *Medicina interna sistematica. Parti 4 e 5: Malattie dell'apparato digerente, del fegato e del pancreas*, Edra, Milano (SSD MED/09 – MEDICINA INTERNA).

Voci enciclopediche:

1. Matucci A., Romagnani S. (2007), "Allergia", in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma: https://www.treccani.it/enciclopedia/allergia_(Enciclopedia-della-Scienza-e-della-Tecnica)/ (SSD MED/09 MEDICINA INTERNA).
2. S.a. (1993), "Insufficienza respiratoria", in *Enciclopedia Medica Italiana*, Aggiornamento della Seconda edizione, USES Scientifiche, Firenze, pp. 6355-6372 (SSD MED/10 MALATTIE DELL'APPARATO RESPIRATORIO).

Articoli:

1. Averna M. (2019), "Ipercolesterolemia familiare omozigote: nuove evidenze e raccomandazioni ai medici per migliorare la diagnosi e la gestione clinica", in *HoFH Today*, 1, 2019, 12-31 (SSD MED/09 MEDICINA INTERNA).
2. Caudarella R. (2007), "Ipercalciuria idiopatica", in *L'endocrinologo*, 8, 3, pp. 209-220 (SSD MED/13 ENDOCRINOLOGIA).
3. Tarugi P., Noto D. (2019), "Livelli molto bassi di LDL e rischio cardiovascolare: dalle evidenze genetiche a quelle dei trial clinici", in *Giornale Italiano dell'Arteriosclerosi*, 10, 2, pp. 33-47 (SSD MED/11 MALATTIE DELL'APPARATO CARDIOVASCOLARE).

Area 07 - Scienze agrarie e veterinarie (39.357 parole token)

Manuali universitari:

1. Ceccon P., Fagnano M., Grignani C., Monti M., Orlandini S. (2017), *Agronomia*, Napoli, Edises (SSD AGR/02 AGRONOMIA E COLTIVAZIONI ERBACEE).
2. Fracassi F. (2023), *Algoritmi diagnostico terapeutici in medicina interna nel cane e nel gatto*, Edra, Milano (SSD VET08 CLINICA MEDICA VETERINARIA)

Voci enciclopediche:

1. Carpaneto G. M., Vigna Taglianti A. (2007), "Entomologia", in *Enciclopedia della Scienza e della Tecnica* (SSD AGR/11 ENTOMOLOGIA GENERALE E APPLICATA).
2. S.a. (2010), "Cardiomiopatia dilatativa del cane", in *Vetpedia* (SSD VET/03 PATOLOGIA GENERALE E ANATOMIA PATOLOGICA VETERINARIA oppure VET/08 CLINICA MEDICA VETERINARIA).
3. S.a. (2010), "Sindrome vestibolare", in *Vetpedia* (SSD VET/03 PATOLOGIA GENERALE E ANATOMIA PATOLOGICA VETERINARIA oppure VET/08 CLINICA MEDICA VETERINARIA).

Articoli:

1. Renna M., Di Gioia F., Leoni B., Santamaria P. (2016), "Due espressioni dell'agrobiodiversità in orticoltura: germogli e microortaggi", in *Italus Hortus*, 23, 1, pp. 31-44 (SSD AGR/04 ORTICOLTURA E FLORICOLTURA).
2. Riva J., Bondiolotti G., Trotti A., Verga M. (2007), "Possibili relazioni tra disturbi comportamentali su base ansiosa nel cane e neurotrasmettitori", in *Praxis veterinaria* 28, 3, pp. 3-10 (SSD AGR/19 ZOOTECNICA SPECIALE).

Macroarea umanistica (260.983 parole token)

Area 10 - Scienze dell'antichità, filologico-letterarie e storico-artistiche (128.056 parole token)

Manuali universitari:

1. Brambilla M. (2007), *Il discorso politico nei paesi di lingua tedesca. Metodi e modelli di analisi linguistica*, Aracne, Roma (SSD L-LIN/14 LINGUA E TRADUZIONE – LINGUA TEDESCA).
2. Brazzelli N. (2015), *L'Antartide nell'immaginario inglese: spazio geografico e rappresentazione letteraria*, Ledizioni, Milano (SSD L-LIN/10 LETTERATURA INGLESE).
3. Cadioli A. (2021), *«La sana critica». Pubblicare i classici italiani nella Milano di primo Ottocento*, Firenze University Press, Firenze, (SSD L-FIL-LET/11 LETTERATURA ITALIANA CONTEMPORANEA).
4. Fanciullo F. (2013), *Introduzione alla linguistica storica*, il Mulino, Bologna (SSD L-LIN/01 GLOTTOLOGIA E LINGUISTICA)
5. Quilici L., Quilici Gigli S. (2004), *Introduzione allo studio della topografia antica*, il Mulino, Bologna (SSD L-ANT/09 TOPOGRAFIA ANTICA).

6. Riccobono M.G., *Dante poeta-profeta, pellegrino, autore: strutturazione espressiva della Commedia e visione escatologica dantesca*, Aracne, Roma, 2013 (SSD L-FIL-LET/10 LETTERATURA ITALIANA).
7. Sciolla G. C. (2010), *Studiare l'arte. Metodo, analisi e interpretazione delle opere e degli artisti*, UTET, Torino (SSD L-ART/02 STORIA DELL'ARTE MODERNA).
8. Uberti M. L. (2005), *Introduzione* alla storia del Vicino Oriente antico, il Mulino, Bologna (SSD L-OR/01 STORIA DEL VICINO ORIENTE ANTICO).

Voci enciclopediche:

1. Barsanti G. (1993), "Candelabro", in *Enciclopedia dell'Arte Medievale*, Treccani, Roma: https://www.treccani.it/enciclopedia/candelabro_(Enciclopedia-dell'-Arte-Medievale) (SSD L-ART/01 STORIA DELL'ARTE MEDIEVALE).
2. Breglia L. (1957), "La numismatica", in *Enciclopedia classica*, a cura di. Pighi G. B. , Del Grande C., Arias P. E., Società Editrice Internazionale, Torino, sez. I, vol. III, pp. 491-531 (SSD L-ANT/04 NUMISMATICA).
3. Canciani F., Tamburello A. (1966), "Trono", in *Enciclopedia dell'Arte Antica*, Treccani, Roma: https://www.treccani.it/enciclopedia/trono_(Enciclopedia-dell'-Arte-Antica)/ (SSD L-ANT/07 ARCHEOLOGIA CLASSICA).
4. Corno D. (2011), "Retorica", in Simone R. (dir.), *Enciclopedia dell'Italiano*, Treccani, Roma, pp. 1242-1247: (https://www.treccani.it/enciclopedia/retorica_(Enciclopedia-dell'Italiano)/ (SSD L-FIL-LET/12 LINGUISTICA ITALIANA).
5. Donadoni S. (1994), "Arte egiziana", in *Enciclopedia dell'Arte Antica*, Treccani, Roma: https://www.treccani.it/enciclopedia/arte-egiziana_res-aa909ab4-66b6-11e1-b491-d5ce3506d72e_(Enciclopedia-dell'-Arte-Antica)/ (SSD L-OR/02 EGITTOLOGIA E CIVILTÀ COPTA).
6. Marocco P. (2003), "Cinema digitale", in *Enciclopedia del cinema*, Treccani, Roma: https://www.treccani.it/enciclopedia/cinema-digitale_(Enciclopedia-del-Cinema)/(SSD L-ART/06 CINEMA, FOTOGRAFIA E TELEVISIONE)
7. Medri M. (2002), "Ricerca archeologica. La documentazione dello scavo", in *Il mondo dell'archeologia*, Treccani, Roma: https://www.treccani.it/enciclopedia/ricerca-archeologica-la-documentazione-dello-scavo_(Il-Mondo-dell'Archeologia)/ (SSD L-ANT/10 METODOLOGIE DELLA RICERCA ARCHEOLOGICA).
8. Salviati F. (2002), "Dai primi insediamenti al fenomeno urbano. Estremo Oriente", in *Il Mondo dell'Archeologia,* Treccani, Roma: https://www.treccani.it/enciclopedia/dai-primi-insediamenti-al-fenomeno-urbano-estremo-oriente_(Il-Mondo-dell'Archeologia)/ (SSD L-OR/05 ARCHEOLOGIA E STORIA DELL'ARTE DEL VICINO ORIENTE ANTICO).

Articoli:

1. Bagnasco G. (2023), "Note preliminari sull'Eracle del Museo d'Arte della Fondazione Luigi Rovati", in *Aristonothos*, XIX, pp. 17-50 (SSD L-ANT/06 ETRUSCOLOGIA E ANTICHITÀ ITALICHE).
2. Bulfoni C., Feng L.S. (2023), "Gli italianismi nella lingua cinese", in *Italiano LinguaDue*, 15, 1, pp. 558-568: https://riviste.unimi.it/index.php/promoitals/article/view/20418 (SSD L-OR/21 LINGUE E LETTERATURE DELLA CINA E DELL'ASIA SUD-ORIENTALE).
3. Cabrini A. M. (2017), "'Qui in Milano': aspetti e strategie del narrare bandelliano", in Prada M., Sergio G. (a cura di), *Italiani di Milano: studi in onore di Silvia Morgana*, Ledizioni, Milano, pp. 213-224 (SSD L-FIL-LET/10 LETTERATURA ITALIANA).
4. Chiesa P. (2015), "Autografi medievali e filologia. In margine a un convegno paleografico", in *Carte Romanze*, 3, 1, pp. 359-374 (SSD L-FIL-LET/09 FILOLOGIA E LINGUISTICA ROMANZA).
5. Coiro L. (2021), "Stili vocali e coreutici nella ricerca di Alan Lomax: Cantometrics, Choreometrics ed equità culturale", in *De Musica*, 24, 2, pp. 5-21 (SSD L-ART/07 MUSICOLOGIA E STORIA DELLA MUSICA).

6. Pomponio F. (2010), "Assiriologia e letteratura poliziesca: rapporti tra due nobili avventure intellettuali", in *Quaderni di Vicino Oriente*, 5, pp. 293-307 (SSD L-OR/03 ASSIRIOLOGIA).
7. Scaramucci M. (2016), "Ch'ixi, Qhipnayra, Nkali: modi di dire l'emergenza decoloniale", in *Altre modernità*, 26, pp. 148-160: https://riviste.unimi.it/index.php/AMonline/article/view/7847/7519 (SSD L-LIN/08 LETTERATURA PORTOGHESE E BRASILIANA).
8. Scarpino C. (2021), "In transfer. Prospettive sulla letteratura americana tradotta in Italia", in *Ácoma*, 26, pp. 7-21 (SSD L-LIN/11 LINGUE E LETTERATURE ANGLO-AMERICANE).

Area 11 - Scienze storiche, filosofiche, pedagogiche e psicologiche (99. 764 parole token)

Manuali universitari:

1. Bachmann C., Luccio R., Salvadori E. (2005), *La verifica della significatività dell'ipotesi nulla in psicologia*, Firenze University Press, Firenze (SSD M-PSI/03 PSICOMETRIA).
2. Concato G. (2006), *Manuale di psicologia dinamica*, AlefBet, Firenze (SSD M-PSI/07 PSICOLOGIA DINAMICA).
3. Oliverio Ferraris A. (2013), *Il significato del disegno infantile*, Bollati Boringhieri, Torino (SSD M-PSI/04 PSICOLOGIA DELLO SVILUPPO E PSICOLOGIA DELL'EDUCAZIONE).
4. Sabbatucci G., Vidotto V. (2019), *Storia contemporanea Dalla Grande Guerra a oggi*, Laterza, Bari-Roma (SSD M-STO/04 STORIA CONTEMPORANEA).
5. Turbanti S. (2017), *Bibliometria e scienze del libro: internazionalizzazione e vitalità degli studi italiani*, Firenze University Press, Firenze, (SSD M-STO/08 ARCHIVISTICA, BIBLIOGRAFIA E BIBLIOTECONOMIA).
6. Zhok A. (2018), *Identità della persona e senso dell'esistenza*, Meltemi, Milano (SSD M-FIL/03 FILOSOFIA MORALE).

Voci enciclopediche:

1. Ciancio L. (2013), "I segni del tempo: teorie e storie della Terra", in *Il Contributo italiano alla storia del Pensiero: Scienze*, Treccani, Roma: https://www.treccani.it/enciclopedia/i-segni-del-tempo-teorie-e-storie-della-terra_(Il-Contributo-italiano-alla-storia-del-Pensiero:-Scienze)/ (SSD M-STO/05 STORIA DELLA SCIENZA E DELLE TECNICHE).
2. Ciliberto M. (2012), "Rinascimento: oltre la costruzione dei 'moderni'", in *Il Contributo italiano alla storia del Pensiero: Filosofia*, Treccani, Roma: https://www.treccani.it/enciclopedia/rinascimento-oltre-la-costruzione-dei-moderni_(Il-Contributo-italiano-alla-storia-del-Pensiero:-Filosofia)/ (SSD M-FIL/06 STORIA DELLA FILOSOFIA).
3. Grottanelli C. (2002), "Regalità, nell'antico mondo mediterraneo", in *Enciclopedia delle religioni*, a cura di Eliade M., vol. XI (Religioni del Mediterraneo e del Vicino Oriente), Città Nuova, Roma, pp. 470- 475 (SSD M-STO/06 STORIA DELLE RELIGIONI).
4. Meniconi A. (2011), "La magistratura", in *L'Unificazione*, Treccani, Roma: https://www.treccani.it/enciclopedia/la-magistratura_(L%27Unificazione)/ (SSD M-STO/04 STORIA CONTEMPORANEA).
5. Mugnai M. (2012), "Scienza e filosofia: Geymonat e Preti", in *Il Contributo italiano alla storia del Pensiero: Filosofia*, Treccani, Roma: https://www.treccani.it/enciclopedia/scienza-e-filosofia-geymonat-e-preti_(Il-Contributo-italiano-alla-storia-del-Pensiero:-Filosofia)/ (SSD M-FIL/02 LOGICA E FILOSOFIA DELLA SCIENZA).
6. Vasoli C. (2012), "Dante Alighieri", in *Il Contributo italiano alla storia del Pensiero: Filosofia*, Treccani, Roma: https://www.treccani.it/enciclopedia/dante-alighieri_(Il-Contributo-italiano-alla-storia-del-Pensiero:-Filosofia)/ (SSD M-FIL/08 STORIA DELLA FILOSOFIA MEDIEVALE).

Articoli:

1. Banini T., Picone M. (2018), "Verso una geografia per la partecipazione", in *Geotema*, 56, gennaio-aprile, pp. 3-10:

https://www.ageiweb.it/geotema/wp-content/uploads/2018/12/56_1_Banini-Picone.pdf (SSD M-GGR/01 GEOGRAFIA).
2. Benedetti M. (2016), "Frati minori e Inquisizione. Alcuni casi nell'Italia medievale", in *Territórios e Fronteiras*, 9, 1, pp. 83-96: https://periodicoscientificos.ufmt.br/territoriosefronteiras/index.php/v03n02/article/view/563 (SSD M-STO/07 STORIA DEL CRISTIANESIMO E DELLE CHIESE).
3. Ceciliani A. (2018), "Didattica integrata quali-quantitativa, in educazione motoria-sportiva, e benessere in età evolutiva", in *Formazione & insegnamento*, 16, 1, pp. 183-193: https://cris.unibo.it/retrieve/e1dcb330-e9ab-7715-e053-1705fe0a6cc9/Didattica%20Integrata%20Ceciliani.pdf (SSD M-EDF/01 METODI E DIDATTICHE DELLE ATTIVITÀ MOTORIE).
4. Ciabarri L. (2005), "Appropriazioni debite: reti sociali e gerarchie nel consumo locale dell'aiuto umanitario: il campo rifugiati di Darwanaje – Somaliland", in *Antropologia*, 5, pp. 155-169: https://www.ledijournals.com/ojs/index.php/antropologia/article/view/128/120 (SSD M-DEA/01 DISCIPLINE DEMOETNOANTROPOLOGICHE).
5. Conte P. J. A., "'Una trasgressione delittuosa': Manichini di cera e teoria della percezione", in *Itinera*, pp. 173-189: https://riviste.unimi.it/index.php/itinera/article/view/2929 (SSD M-FIL/04 ESTETICA).
6. Daniele K., Zannini L. (2013), "I criteri COREQ per valutare la validità degli studi qualitativi: un percorso a ritroso per illuminarne alcuni presupposti taciti", in *Studium Educationis* , 24, 1, pp. 84-95: https://ojs.pensamultimedia.it/index.php/studium/article/view/6485/5610 (SSD M-PED/01 PEDAGOGIA GENERALE E SOCIALE).

Area 14 - Scienze politiche e sociali (33.163 parole token)

Manuali universitari:

1. Arcuri F. P., Arcuri F. (2010), *Manuale di sociologia. Teorie e strumenti per la ricerca sociale*, Springer, Milano (SSD SPS/07 SOCIOLOGIA GENERALE).
2. Ferrera M., V. Fargion V., Jessoula M. (2012), *Alle radici del welfare all'italiana: origini e futuro di un modello sociale squilibrato*, Marsilio, Venezia (SSD SPS/04 SCIENZA POLITICA).

Voci enciclopediche:

1. Bezza B., Dosi G. (1991), "Automazione", in *Enciclopedia delle scienze sociali*, Treccani, Roma: https://www.treccani.it/enciclopedia/automazione_(Enciclopedia-delle-scienze-sociali)/ (SSD SPS/09 SOCIOLOGIA DEI PROCESSI ECONOMICI E DEL LAVORO).
2. Schizzerotto A. (1997), "Scuola", in *Enciclopedia delle scienze sociali*, Treccani, Roma: https://www.treccani.it/enciclopedia/scuola_(Enciclopedia-delle-scienze-sociali)/ (SSD SPS/08 SOCIOLOGIA DEI PROCESSI CULTURALI E COMUNICATIVI).

Articoli:

1. Dalla Chiesa N. (2016), "A proposito di "mafia capitale". Alcuni problemi teorici", in *Rivista di studi e ricerche sulla criminalità organizzata*, 1, 2 pp. 1-15: https://riviste.unimi.it/index.php/cross/article/view/6634 (SSD SPS/11 SOCIOLOGIA DEI FENOMENI POLITICI).
2. Miceli S. (2022), "Alle origini della sociologia. Una contronarrazione", in *Quaderni di teoria sociale*, 2, pp. 247-262: http://www.riviste.morlacchilibri.com/index.php/quaderniditeoriasociale/article/view/186/207 (SSD SPS/07 SOCIOLOGIA GENERALE).

CC BY SA