

IL LESSICO DELLA LINGUISTICA LETTO DA LONTANO. RILEVAZIONI QUANTITATIVE SU UN CORPUS DI ARTICOLI SCIENTIFICI (1876-2014)

*Giovanni Urraci*¹

1. INTRODUZIONE

Il presente contributo intende ricostruire e analizzare le tendenze che hanno governato l'evoluzione della terminologia linguistica attestata nell'«Archivio Glottologico Italiano» (AGI), della quale si propone una “lettura da lontano” (cfr. Moretti, 2005) basata sull'interpretazione di misure lessicometriche relative al ciclo vitale dei tecnicismi e alle dinamiche di accrescimento del lessico specialistico. Nello specifico, adottando una prospettiva prettamente quantitativa, verranno sottoposte ad analisi la dimensione del vocabolario tecnico e le oscillazioni nella frequenza relativa dei tecnicismi, quest'ultima valido indicatore del grado di specializzazione del discorso; saranno inoltre estrapolati ed illustrati dati pertinenti alla stratificazione della terminologia, alla ricchezza lessicale e al tasso di obsolescenza delle parole. In definitiva, l'obiettivo è cogliere, descrivere e, quando possibile, motivare le principali tendenze che hanno governato lo sviluppo della terminologia; ciò senza mettere a fuoco singoli termini e delineando storie di parole, bensì osservando l'oggetto indagato nel suo complesso².

Sebbene le osservazioni proposte si rivolgeranno esclusivamente alle specifiche vicende storico-lessicali dell'AGI, riteniamo che esse potranno suggerire considerazioni di più ampia portata in quanto il periodico oggetto della ricerca ben si presta a rappresentare lo sviluppo complessivo degli studi linguistici italiani. L'AGI, infatti, è la più longeva rivista del settore pubblicata nel nostro paese: fondata nel 1873 da Graziadio Isaia Ascoli, comunemente considerato il padre dello studio scientifico del linguaggio in Italia³, ha attraversato da protagonista, e spesso condizionato, l'intero arco evolutivo della Linguistica italiana; inoltre, è stata diretta da alcuni tra gli studiosi più importanti e influenti⁴ ed è testimone delle principali correnti che si sono susseguite all'interno della disciplina: pur non accogliendone tutte le branche, può essere considerata almeno rappresentativa per quanto concerne dialettologia, linguistica generale e linguistica storica. La notevole varietà tematica e l'estensione cronologica ne fanno un documento prezioso nella ricostruzione storiografica e determinano la sua importanza quale bacino terminologico.

I risultati qui illustrati non hanno la pretesa di rappresentare minutamente l'effettiva evoluzione della terminologia linguistica e andranno letti nella consapevolezza che,

¹ Università degli Studi di Padova.

² Uno sguardo più ravvicinato sugli specifici tecnicismi adoperati nell'AGI e sulla loro distribuzione cronologica è proposto in Urraci (2023).

³ Sul profilo biografico di Ascoli, indispensabile per coglierne l'importanza, cfr. almeno Radoni Zucco (1973) e Brambilla (2009).

⁴ Nell'ordine, si sono succeduti ad Ascoli: Carlo Salvioni, Pier Gabriele Goidanich, Matteo Bartoli, Vittore Pisani, Benvenuto Terracini, Giacomo Devoto, Carlo Alberto Mastrelli, Romano Lazzeroni e Paolo Ramat; la rivista è attualmente diretta da Francesca M. Dovetto, Marco Mancini e Alberto Nocentini.

sebbene inevitabilmente condizionati dalle trasformazioni della materia stessa, i mutamenti osservati scaturiscono da accadimenti interni alla rivista, compresi gli avvicendamenti alla direzione e il conseguente imporsi di diverse linee editoriali (Urraci, 2021b: 235-236). Nonostante ciò, siamo convinti che la ricerca possa configurarsi come uno studio di fattibilità attraverso cui accertare la possibilità di esaminare i tecnicismi attraverso un approccio quantitativo e, al tempo stesso, effettuare una messa a punto di strumenti e metodologie; una messa a punto efficace perché interagisce con un corpus ampio ma circoscritto, eccezionalmente vario ma internamente coerente e uniforme, testimone di un ricambio lessicale profondo ma non soverchio⁵. In definitiva, auspichiamo che il lavoro svolto possa costituire un primo passo nella direzione di una indagine più sistematica ed esaustiva della fisionomia complessiva della terminologia linguistica.

2. IL CORPUS

Il corpus su cui è stata condotta l'analisi raccoglie il testo degli articoli in italiano⁶ pubblicati sull'AGI dal secondo numero, uscito nel 1876, al volume 99 del 2014. Non è stata presa in considerazione la prima annata della rivista (1873) in quanto essa accoglie esclusivamente i *Saggi ladini* di G. I. Ascoli⁷, assimilabili a una monografia⁸, e non ci si è spinti oltre il 2014 al fine di garantire un decennio di distanza tra il momento dell'analisi e la produzione dell'oggetto studiato, una separazione che è a nostro avviso indispensabile per evitare di trasformare la storiografia in cronaca⁹. Inoltre, sono stati esclusi i volumi 5 e 6, in quanto racchiudono l'edizione critica di un antico codice, il cosiddetto *Codice irlandese dell'Antoniana*, e sono state sistematicamente omesse dall'analisi alcune sezioni tra cui sommari, indici, liste di pubblicazioni, necrologi, note di cronaca e recensioni. Infine, in risposta all'esigenza di garantire la massima uniformità del corpus e la minima dispersione del lessico tecnico, si è intervenuti anche sui singoli contributi espungendo correzioni e postille, bibliografie e *abstract*, edizioni di testi e glossari quando privi di una significativa componente di analisi linguistica.

⁵ Gli strumenti di analisi impiegati sono in corso di perfezionamento, ma comunque sostenuti da precedenti riflessioni e applicazioni. Il modello metodologico è stabilito nei contributi ospitati in Tuzzi (2018) e, per lo specifico ambito della storia della Linguistica, alcuni esperimenti pertinenti sono discussi in Urraci e Cortelazzo (2018).

⁶ Non si sono dunque considerati gli articoli scritti in lingue diverse dall'italiano, numerosi in particolare negli anni più recenti: su un totale di 55 articoli in lingue altre, ben 33 sono stati pubblicati dopo il 1992. Come prevedibile, la lingua meglio attestata è l'inglese: 35 articoli; non mancano però contributi interamente in francese (8), spagnolo (6), tedesco (2), portoghese (1) e rumeno (1).

⁷ Per una esaustiva sintesi della struttura e dei contenuti dei *Saggi ladini* rimandiamo a Lucchini (2009: 48-53). Importante, e di notevole impatto, è anche il *Proemio* che introduce il primo numero dell'AGI; al riguardo si consultino, tra gli altri, Morgana (2001) e il più recente Morgana (2009).

⁸ A differenza delle altre annate si tratta di un testo essenzialmente monotematico scritto da una mano unica, presenta quindi una eccezionale compattezza che contrasta con la natura composita del corpus e, pertanto, la sua inclusione comporterebbe il rischio di anomalie statistiche. Non di meno, l'importanza dei *Saggi ladini* nella fondazione della terminologia linguistica è ineludibile; pertanto, essi sono stati consultati e presi in considerazione in fase di compilazione della lista rappresentativa del lessico della Linguistica (cfr. §. 3).

⁹ Nello stabilire i limiti cronologici si è dovuto tenere conto anche di alcune esigenze operative; in particolare, dei vincoli imposti dalla crescente propensione dell'AGI ad accogliere articoli in lingua inglese, ormai prevalenti in alcuni fascicoli, che non possono essere integrati nel corpus.

Tabella 1. *Misure lessicometriche essenziali*

N – Token	5.089.527
V – Type	268.920
(V/N)*100 – Type/Token Ratio	5,3
(V _I /V)*100 – Percentuale di hapax	61,8

Nel complesso il corpus è costituito da 5.089.527 *token*¹⁰ distribuiti su 95 volumi¹¹ e si compone di 831 articoli opera di 358 autori, tra cui per produttività spiccano direttori e condirettori della rivista (cfr. Tabella 2). Le misure lessicometriche fondamentali utili a descriverlo sono riassunte in Tabella 1.

Tabella 2. *Autori maggiormente prolifici*

Autore	Articoli firmati	Autore	Articoli firmati
Graziadio Isaia Ascoli	55	Pier Gabriele Goidanich	13
Vittore Pisani	37	Nunzio Maccarrone	11
Giuliano Bonfante	30	Silvio Pieri	10
Carlo Alberto Mastrelli	24	Aldo Luigi Prosdocimi	10
Carlo Salvioni	20	Ruggero Stefanini	9
Matteo Bartoli	17	Domenica Romagno	8
Benvenuto A. Terracini	17	Giovanni Flechia	8
Alberto Nocentini	16	Roberto Gusmani	8
Romano Lazzeroni	16	Giuseppe Morosi	7

3. ESTRAZIONE DELLA TERMINOLOGIA SPECIALISTICA

La notevole ampiezza del corpus genera un forte rumore di fondo. Si è pertanto imposto come propedeutico a un efficace studio dei dati testuali lo sviluppo di strumenti atti a individuare e isolare, in maniera semi-automatica, i lessemi rilevanti, ossia i tecnicismi

¹⁰ Per ulteriori informazioni circa la composizione del corpus e per una dettagliata illustrazione dei criteri seguiti nelle fasi di digitalizzazione, normalizzazione e *tokenizzazione* rimandiamo a Urraci e Cortelazzo (2018: 88-89). Ci limitiamo qui a segnalare interventi quali la riduzione al minuscolo di tutte le lettere, l'inversione dell'apocope (*vocal* > *vocale*) e l'eliminazione delle vocali prostetiche (*istrumentale* > *strumentale*); si è inoltre agito sui composti, regolarizzandone la grafia, e sulla rimozione delle varianti formali, ricondotte all'uso prevalente (*indeuropeo* > *indoeuropeo*; *napolitano* > *napoletano*).

¹¹ Le annate 22 e 23 sono accorpate in un unico volume: ripropongono la silloge composta per il centenario della nascita di Ascoli.

della Linguistica, sui quali poter poi concentrare le analisi quantitative e le osservazioni qualitative.

L'estrazione automatica della terminologia tecnica da corpora rientra in un ambito di ricerca ben affermato le cui prime espressioni, sorte nel campo del *Natural Language Processing*, risalgono alla fine degli anni '90 (Heylen, De Hertog, 2015: 204); le sue declinazioni sono molteplici e le metodologie spaziano dagli approcci linguistici a quelli statistici, non di rado assumendo forme ibride¹². A nostro avviso, però, queste procedure non garantiscono risultati pienamente soddisfacenti rispetto alle esigenze sollevate dalla ricerca¹³: non possono infatti risolvere pienamente il complesso problema della sovrapposizione e dei continui interscambi tra lessico settoriale e lessico comune (Cabr , 1999: 65 e ss.), e ci    ovviamente assai problematico per uno studio che si prefigge di tracciare la distribuzione della terminologia tecnica. Si consideri inoltre che i testi studiati, pur non presentando una variabilit  verticale (Rovere, 1989; Cortelazzo, 1994), sono complessi a causa della forte interdisciplinariet  e del loro essere intrisi di lingua comune, portata con s  dalla necessaria attenzione per i referenti extralinguistici; a tali difficolt , proprie di molte lingue speciali, e che nella Linguistica si manifestano con particolare forza, si sommano ulteriori problematiche specifiche del corpus di riferimento, primariamente determinate dalla invasiva presenza della variazione diacronica, la quale si concretizza in risemantizzazioni e sostituzioni oltre che nei frequenti innesti di neologismi e nella scomparsa di non pochi vocaboli. Tutti fenomeni difficili da cogliere nella loro pienezza persino attraverso metodologie d'indagine qualitativa (De Luca, 2014: 50).

Pur riconoscendo valore e utilit  delle metodologie di estrazione automatica, alla luce di quanto sin qui affermato si   optato per una procedura che prevede un'ampia, preponderante componente manuale:   stata compilata una lista lessicale di riferimento che, confrontata con il prodotto della *tokenizzazione* della rivista, ha consentito di riconoscere e *taggare* tramite software i tecnicismi¹⁴. Tale lista   stata prodotta innanzi tutto attingendo alle voci contenute in nove vocabolari terminologici¹⁵, primariamente selezionati per la loro capacit  di garantire una buona copertura cronologica: sono stati consultati Severino (1937), De Felice (1954), Gentile (1963), Cardona (1969), Ducrot e Todorov (1972), Dubois *et al.* (1979), Cardona (1988)¹⁶, Beccaria (2004) e Casadei (2011)¹⁷. Ogni vocabolario   inevitabilmente frutto di una selezione e non si prefigge di cogliere la terminologia nella sua interezza; pertanto, al fine di catturare una pi  ampia porzione dei tecnicismi specifici del campione studiato, il lessico ripreso dalle fonti lessicografiche  

¹² Per una panoramica sui principali approcci all'estrazione della terminologia tecnica si possono consultare Pazienza *et al.* (2005) e Heylen, De Hertog (2015).

¹³ I dati presentati in Bonin *et al.* (2012) consentono di stimare l'affidabilit  dei processi di estrazione automatica tra il 61% e il 79% (65-73% per i poliformi). Bisogna inoltre considerare che, mentre possono essere conteggiate le forme erroneamente etichettate come "tecnicismo",   difficile valutare il numero di usi specialistici che sfuggono al riconoscimento degli strumenti automatizzati.

¹⁴ La lista compilata ha permesso di associare tre variabili categoriali al vocabolario del corpus attraverso le quali ogni unit  lessicale   stata *taggata* come tecnicismo (si/no), glottonimo (si/no) e tecnicismo di una disciplina contigua (si/no).

¹⁵ I lemmi riportati nei vocabolari considerati non sono stati ripresi nella loro interezza, ma si   proceduto a una scrematura finalizzata ad espungere quelli problematicamente polisemici e quelli che, nelle pubblicazioni considerate, sono attestati solo occasionalmente nelle loro accezioni specialistiche; entrambe le categorie sono state individuate mediante una sistematica verifica delle concordanze delle voci ambigue. La ripresa non si   limitata alle voci messe a lemma, ma si   attinto anche alle forme citate soltanto in sede di definizione e discussione dei concetti.

¹⁶ Si   scelto di analizzare due edizioni del dizionario di G. R. Cardona in quanto le differenze tra di esse sono cos  numerose e profonde da poterle considerare opere a s  stanti (cfr. Cardona, 1988: 5-8).

¹⁷ Per un interessante inquadramento storico di alcuni dei vocabolari consultati cfr. Iacobini (1997).

stato integrato con forme rilevanti estrapolate da AGI e *Lingua nostra*¹⁸ attraverso uno spoglio eseguito con l'ausilio di filtri basati sul calcolo delle specificità¹⁹: l'applicazione di tali filtri ha evidenziato le variazioni nella distribuzione delle occorrenze rispetto ai valori attesi ed ha favorito l'individuazione di unità lessicali associate a periodi circoscritti.

In un secondo momento sono state individuate anche le *multimword expression*, rispetto alle quali i vocabolari sono risultati poco informativi ma la cui inclusione nell'analisi è imprescindibile. Ai lessemi polirematici ripresi dai vocabolari consultati si sono aggiunti quelli suggeriti dall'analisi dei segmenti ripetuti (cfr. Giuliano, La Rocca, 2008: 221-223), una procedura che consente di riconoscere le combinazioni di parole ricorrenti in un certo testo²⁰; esse sono state successivamente ordinate secondo il loro indice di significatività e vagliate manualmente così da isolare i segmenti rilevanti²¹. Il corpus è stato quindi lessicalizzato sulla base dei poliformi individuati, attuando però una scrematura motivata dalla necessità di evitare la dispersione delle parole in un numero eccessivo di unità di analisi: sono stati dunque esclusi dal processo i poliformi con meno di 15 occorrenze quando in concorrenza con forme semplici.

Nel complesso, le procedure descritte hanno permesso l'individuazione di 7.939 termini settoriali: 6.807 tecnicismi propri della Linguistica, 684 riconducibili a discipline contigue e 448 glottonimi (cfr. Tabella 3)²². Nei calcoli rientrano anche 576 poliformi.

Tabella 3: *Composizione della lista di riferimento impiegata per l'individuazione del lessico settoriale*

Voci totali 7.939	
Glottonimi	448
Tecnicismi della linguistica	6.807
Tecnicismi di discipline contigue	684

¹⁸ La considerazione di una ulteriore rivista scientifica si è rivelata preziosa per ampliare il bacino lessicale di riferimento. *Lingua nostra*, fondata da Bruno Migliorini e Giacomo Devoto nel 1939, è la prima rivista interamente e programmaticamente dedicata alla lingua italiana, sul modello della tedesca *Muttersprache* e, soprattutto, della francese *Le français moderne* diretta da Albert Dauzat. Il 2024 ha visto la pubblicazione della sua ottantacinquesima annata.

¹⁹ Rimandiamo a Lebart *et al.* (1998) per una dettagliata disamina del calcolo delle specificità. Ci limitiamo qui a esplicitare che, nell'ambito della ricerca condotta, tale elaborazione è servita a misurare, per ogni unità lessicale, lo scarto della frequenza relativa nei vari sub corpora rispetto al valore registrato nella totalità del corpus; in altri termini, il dato ha descritto, per ciascuna annata, le variazioni rispetto alla normale distribuzione delle parole. Ciò è stato assai utile al fine di individuare i tecnicismi caratterizzanti un arco temporale ristretto in quanto ha permesso di eseguire uno spoglio delle sole forme con specificità positiva ($p\text{-value} \geq 0,25$), ossia quelle le cui occorrenze si concentrano, in maniera statisticamente significativa, in un numero limitato di volumi.

²⁰ L'operazione è stata eseguita per mezzo del software TaLTaC2 (cfr. Bolasco, 2010).

²¹ Nello specifico, ci si è serviti dell'IS relativo. Pur senza entrare nello specifico della formula applicata, per la quale rimandiamo a Morrone (1993), segnaliamo che tale misura si basa essenzialmente sul rapporto tra le occorrenze del segmento testuale e quelle di ciascuna delle forme grafiche che lo compongono; pertanto, se un segmento è costituito da parole che occorrono in prevalenza all'interno di quella specifica sequenza, e dunque se lo scarto tra nominatore e denominatore è ridotto, esso farà registrare un indice di significatività molto elevato, ed è probabile si tratti di un lessema polirematico e non di una combinazione casuale di parole. A livello operativo, l'ordinamento dei segmenti ripetuti secondo il loro IS relativo ha facilitato il lavoro di scrematura consentendo l'immediata messa in evidenza di quelli potenzialmente più rilevanti.

²² Al riguardo si vedano anche Urraci e Cortelazzo (2018: 90) e Urraci (2019: 458-459). Segnaliamo inoltre che l'efficacia della lista lessicale di riferimento è stata messa alla prova in Urraci (2021b), sebbene con obiettivi diversi e nel contesto di un confronto tra i profili lessicali di AGI e «Lingua Nostra» finalizzato a tracciare un parallelo tra Linguistica generale e Storia della lingua italiana.

Provenienza delle voci	
Lemmi dei vocabolari terminologici	4.867
Definizioni dei vocabolari terminologici	1.528
Corpus	1.135
Altre fonti	409 (solo glottonimi)

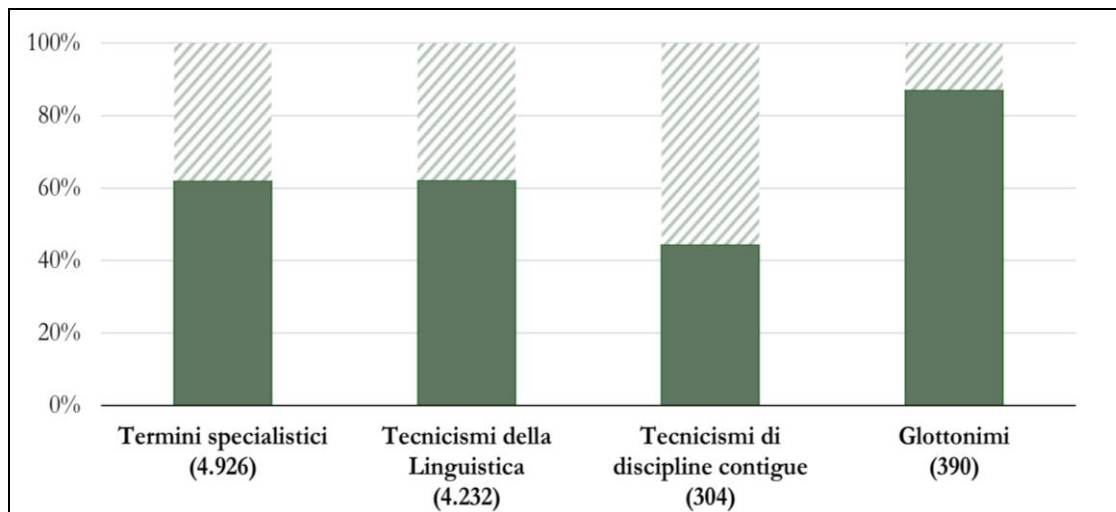
A margine della descrizione delle procedure semi-automatiche impiegate per estrapolare la terminologia tecnica non possiamo esimerci dall'indicare alcune debolezze del sistema adottato, inevitabilmente soggetto a soluzioni di compromesso; in particolare, sopravvive alle operazioni manuali di verifica e selezione un insieme di termini ambigui, etichettati come specialistici ma che in realtà non realizzano la loro accezione tecnica in tutti i contesti. Tuttavia, riteniamo che queste sporadiche discrepanze non inficino i risultati dell'analisi poiché il loro peso sul complesso delle attestazioni è marginale e, quindi, incapace di alterare le tendenze che emergono dai dati.

4. AMPIEZZA DEL VOCABOLARIO SPECIALISTICO

Nell'AGI sono presenti 4.926 dei complessivi 7.939 termini specialistici che compongono la lista di riferimento: il 62% del totale. Di questi, 4.232 sono tecnicismi propri della Linguistica, a conferma di come la rivista in esame offra uno scorcio rappresentativo del lessico normalmente impiegato nella disciplina. La copertura dei tecnicismi provenienti da settori contigui è invece meno estesa (304 su 684), mentre numerosi sono i glottonimi: 390, l'87% di quelli raccolti, un dato coerente con l'apertura allo studio di un numero significativo di lingue differenti, o quanto meno con il frequente riferimento ad esse attraverso esempi e rimandi.

Nel prosieguo della trattazione saranno presi in considerazione unicamente i tecnicismi propri della Linguistica: lo studio dei glottonimi e del lessico attinto ad altre discipline non è pertinente rispetto agli obiettivi del contributo.

Grafico 1. *Copertura del lessico settoriale*



Come già dichiarato nei passaggi introduttivi, la ricerca si prefigge una descrizione della fisionomia complessiva del vocabolario specialistico finalizzata alla individuazione dei tratti che lo caratterizzano e delle tendenze evolutive che lo hanno segnato; in virtù di tale approccio alla materia le prossime pagine non sottoporranno ad analisi singole forme e non si addenteranno in considerazioni a carattere storico e semantico. Nonostante ciò, si ritiene di potenziale interesse per il lettore uno scorcio sui materiali di lavoro, che scegliamo di presentare riportando due liste di parole, pur senza commentarle: la prima include i tecnicismi non ambigui più frequenti nel corpus (Tabella 4), la seconda una selezione di hapax (Tabella 5).

Tabella 4. *Forme semanticamente non ambigue con il maggiore numero di occorrenze*

Tabella 5. *Selezione di hapax*

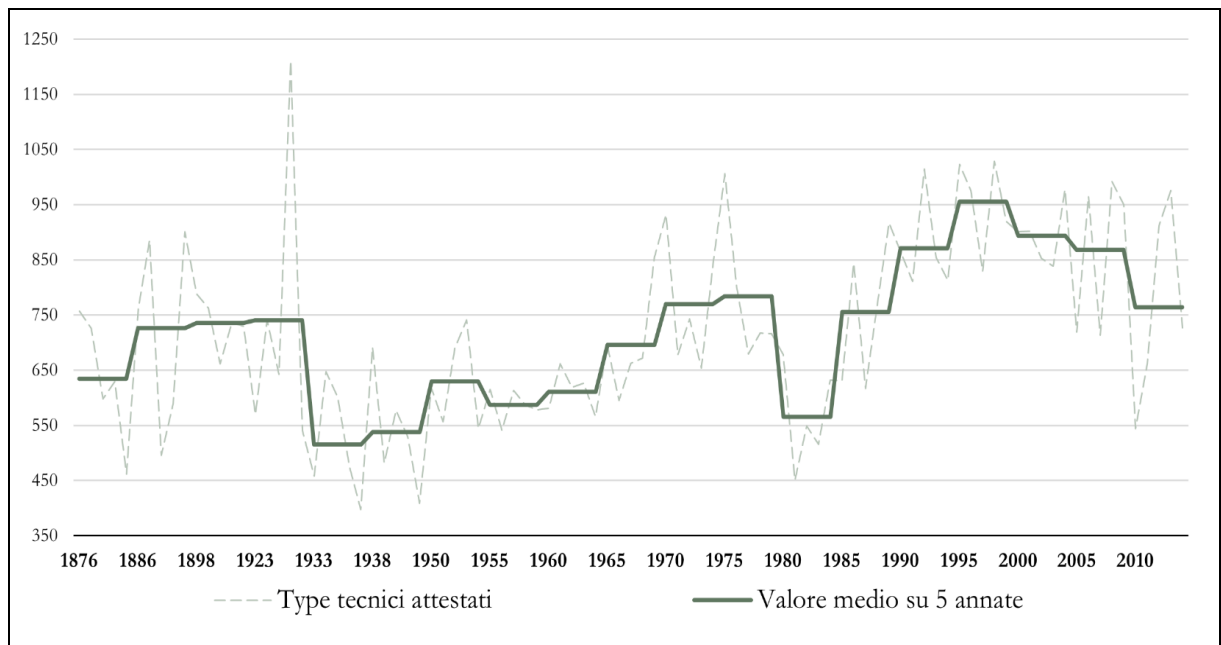
Forma grafica	Occorrenze	Forma grafica	Occorrenze
latino (g)	8.504	antesonantico	1
italiano (g)	7.605	antifonema	1
lingua	5.107	antimetamorfotica	1
greco (g)	4.997	arico	1
significato	4.004	bislungamento	1
verbo	3.978	catacresi	1
francese (g)	3.393	elemento cosale	1
plurale	3.300	enantiomertia	1
parola	3.278	espandimento	1
suffisso	2.518	fonopatema	1
testo	2.112	idiomaticity	1
vocale	2.070	incrassamento	1
fenomeno	2.059	inguria	1
dialetto	1.978	labdacistico	1
tedesco (g)	1.918	metallage	1
articolo	1.865	mouillement	1
frase	1.792	ostruentizzazione	1
linguistica	1.776	parola imitativa	1
toscano (g)	1.694	pasitelegrafia	1
soggetto	1.688	politematia	1
radice	1.581	semelfattività	1
piemontese (g)	1.573	sostratologia	1
indoeuropeo (g)	1.501	tautofonia	1
consonante	1.459	vocoide	1
linguaggio	1.419	welleristico	1

5. IL LESSICO TECNICO IN DIACRONIA

Dopo avere brevemente presentato i dati relativi ad ampiezza e composizione del vocabolario specialistico, spostiamo ora l'attenzione sul numero di tecnicismi effettivamente impiegati in ciascuna annata, valutando anche la variazione del dato in diacronia così da evidenziare le dinamiche di espansione e contrazione del vocabolario attivo. Quantifichiamo e rappresentiamo queste informazioni attraverso il Grafico 2, all'interno del quale il numero di *type* per anno è proiettato sull'asse del tempo²³.

Si è scelto di riportare i valori assoluti, e non il tasso di tecnicismi nel vocabolario, perché essi offrono una più efficace rappresentazione della dimensione del lessico settoriale e costituiscono un buon termine di confronto tra le annate: la lunghezza dei sub corpora, infatti, incide solo marginalmente sulla quantità di tecnicismi impiegati in quanto il numero di quelli necessari a trattare un certo argomento non cresce proporzionalmente alla lunghezza del testo²⁴.

Grafico 2. *Type tecnici attestati in ciascuna annata*



La figura precedente suggerisce che il numero di tecnicismi impiegati in ciascuna annata è soggetto a variazioni ridotte e, pur esibendo oscillazioni più o meno marcate, il dato non evidenzia una tendenza complessiva che possa definirsi forte e significativa: si può piuttosto intuire una sorta di moto ondulatorio caratterizzato da un affossamento tra gli anni Trenta e Sessanta. Inoltre, negli ultimi venticinque anni emerge una sensibile

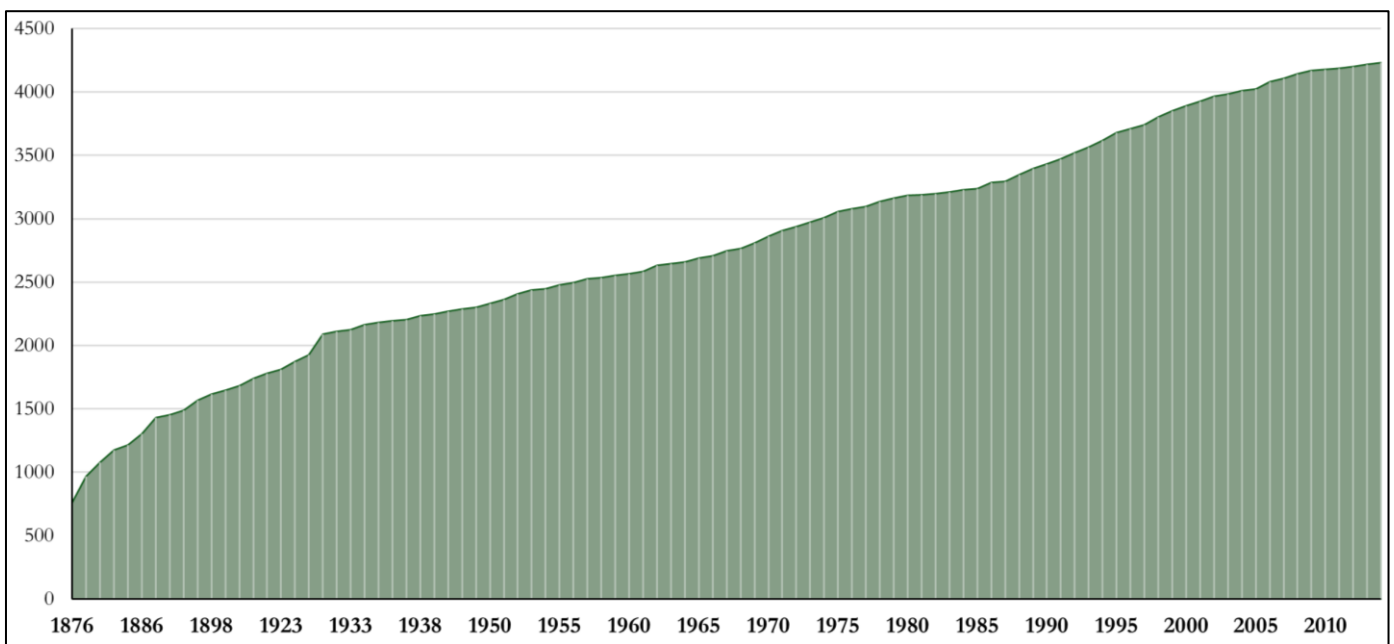
²³ In relazione ai primi numeri è problematico offrire delle datazioni esatte in quanto, agli esordi, i numeri venivano stampati nel corso di più anni e comprendevano fascicoli la cui effettiva data di pubblicazione non sempre coincide con quella indicata nel volume; al riguardo si vedano almeno Lucchini (2009: 54-56) e Santamaria (2018). Si superano tali discrepanze attribuendo a ciascun numero dell'AGI l'anno di pubblicazione del suo ultimo fascicolo.

²⁴ Un effetto rilevabile, ma comunque non in grado di alterare in maniera significativa il complesso dei dati, ha invece il numero di articoli pubblicati in un determinato volume, in quanto si ripercuote direttamente sulla quantità di argomenti trattati: il dato abnorme fatto registrare dall'annata 1929 scaturisce appunto dal fatto che essa ospita quasi il triplo dei contributi rispetto alla media.

espansione del vocabolario specialistico attestato nei singoli saggi, verosimilmente attribuibile all'iniezione di una sostanziosa terminologia associata a nuovi approcci teorici e relativa a morfologia e sintassi, ambiti sino a quel momento marginali negli interessi della rivista (cfr. Urraci, 2019: 461-462); eppure, nonostante ciò, l'incremento della complessità terminologica è minore di quanto ci si sarebbe potuti attendere: nel lungo arco di tempo preso in esame, partendo da una disciplina ancora giovane, la dimensione del vocabolario tecnico impiegato annualmente aumenta di appena il 25%, con valori medi che, comunque, non sono lontani da quelli registrati in alcuni numeri precedenti. Le tematiche affrontate sono molteplici e varie, spaziano dalla dialettologia alla linguistica indoeuropea, dalla fonetica alla sintassi, dalla etimologia alla lessicografia, dalla geografia linguistica alle escursioni nella filologia, tutte branche il cui successo si lega a periodi cronologici ben delimitati²⁵; nonostante ciò, non è possibile individuare un chiaro parallelo tra gli argomenti affrontati e il numero di tecnicismi necessario a trattarli.

L'estensione del lessico specialistico adoperato è, in definitiva, relativamente stabile. Ciò trova conferma anche nel Grafico 3, che riporta sull'asse temporale la dimensione complessiva del vocabolario tecnico, intesa come somma dei tecnicismi diversi (*type*) sino a quel momento documentati: come si può osservare l'andamento è lineare, non esponenziale.

Grafico 3: *Dimensione complessiva del vocabolario tecnico (type)*



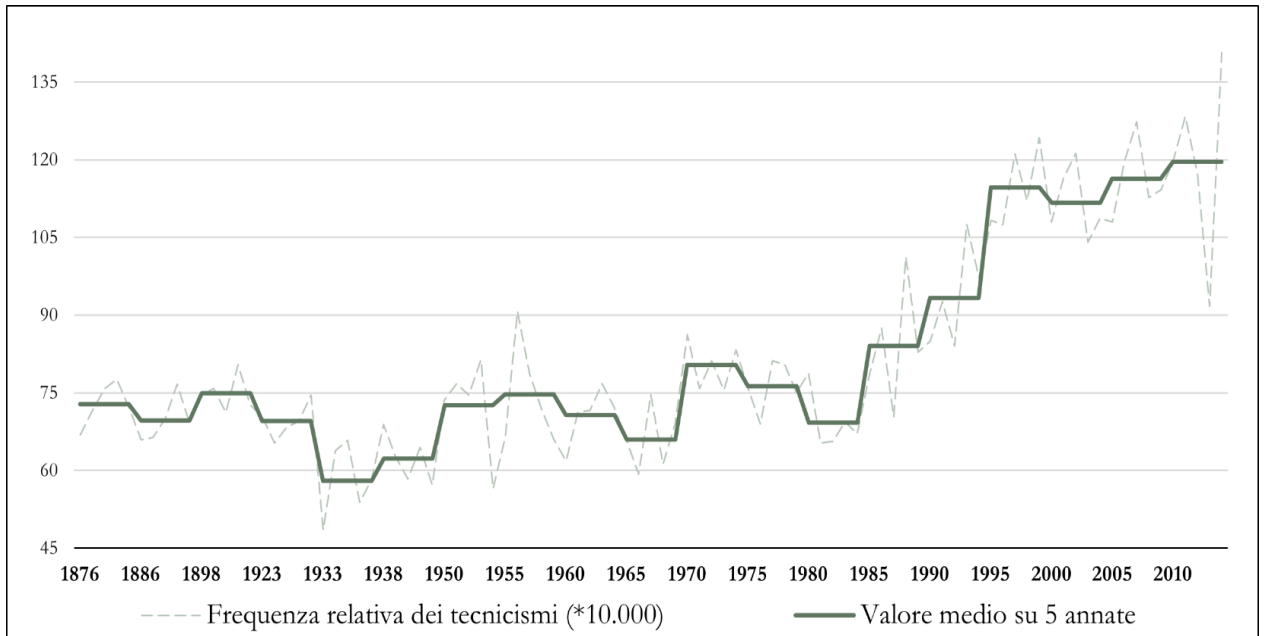
I dati sin qui considerati suggeriscono che il lessico non si sviluppa in maniera incrementale, per accumulo, bensì attraverso cicli di sostituzioni; sostituzioni che, in misura ridotta, avvengono attraverso l'avvicendamento di parole con sinonimi più recenti e coerenti con l'evoluzione dei punti di vista della disciplina²⁶, mentre più spesso consistono nella radicale scomparsa di alcuni interessi di ricerca e relative terminologie, a cui fa seguito l'introduzione di nuovi paradigmi che portano con sé un consistente numero

²⁵ Per un dettagliato quadro tematico della rivista, che ne riassume i principali interessi e segnala la marcatezza diacronica di alcuni campi d'indagine, rimandiamo a Urraci (2021a).

²⁶ Emblematico di questi processi è il tecnicismo della fonetica ottocentesca *colore*, completamente sostituito da *articolazione*; ma si prendano ad esempio anche il subentro di *dileguo* ad *ettlissi* e di *prestito* ad *accatto*.

di neologismi e, talvolta, la specializzazione di parole sino a quel momento prive di un'accezione specialistica: al riguardo, si possono considerare le vicende che hanno interessato *strategia* e *tema*, mentre inverso è il processo conosciuto da *chiaro*, *scuro* e *oscuro*.

Grafico 4. *Andamento nel tempo della frequenza relativa dei tecnicismi*



Come si è più volte ribadito, il quadro delineato dall'esame dei *type* mostra una sostanziale uniformità di fondo. Differenze tutt'altro che trascurabili emergono invece nel grado di specializzazione del discorso scientifico, che può essere quantificato attraverso la frequenza relativa dei tecnicismi: tale misura consente infatti di calcolare le occorrenze di lessico tecnico ogni 10.000 parole, in tal modo sintetizzandone la densità in un testo. Nel Grafico 4 il dato è mostrato nella sua evoluzione cronologica, con riferimento sia alle singole annate sia ai valori medi calcolati su cinque annate. La figura esibisce un andamento tendenzialmente crescente e caratterizzato da un balzo particolarmente marcato in corrispondenza dei primi anni Novanta: il divario con le fasi precedenti della rivista è forte ed evidente, la frequenza con la quale nel testo degli articoli si incontrano tecnicismi aumenta infatti del 60% – un incremento percentuale più che doppio rispetto a quello fatto registrare dai *type* (Grafico 2). Nei primi numeri è possibile infatti riconoscere un termine tecnico ogni 16 *token*, un valore che rimane fondamentalmente invariato sino agli anni Ottanta, quando il rapporto tra parole comuni e specialistiche è di 14 a 1; solo negli ultimi venticinque anni il discorso si fa sensibilmente più carico di voci specialistiche: se ne incontra una ogni 9 parole.

L'improvvisa necessità di un più frequente ricorso ai tecnicismi deriva certamente dall'introduzione di nuove tematiche di ricerca, legate alla morfosintassi e dotate di un alto grado di tecnicità, «studiate secondo una prospettiva che richiede la costante denominazione di oggetti e concetti altamente specifici» (Urraci, 2021b: 245). A ciò si accompagna il restringimento dei riferimenti alla realtà extralinguistica, per loro natura non tecnici, riferimenti che, con alterne fortune, erano sempre stati presenti tanto nella illustrazione dei significati quanto nell'attenzione per il parlante e i contesti d'uso, e che vengono ora fagocitati dall'analisi delle strutture interne alla lingua (cfr. Urraci, 2023: 49).

6. STRATIFICAZIONE CRONOLOGICA DEL VOCABOLARIO TECNICO

Tra i diversi possibili criteri di classificazione del lessico, nell'ottica di uno studio diacronico risulta di particolare interesse la sua categorizzazione basata sulla data di prima attestazione, poiché attraverso di essa è possibile ricostruire la stratificazione del vocabolario e, in tal modo, maturare considerazioni circa la persistenza delle parole e il loro ciclo vitale. Inoltre, i mutamenti nel rapporto tra termini acclimatati e neoformazioni possono offrire indicazioni eloquenti non solo circa lo sviluppo della terminologia ma anche in relazione alla disciplina stessa, della quale indirettamente testimoniano la continuità con il passato, le tensioni interne e la velocità di rinnovamento.

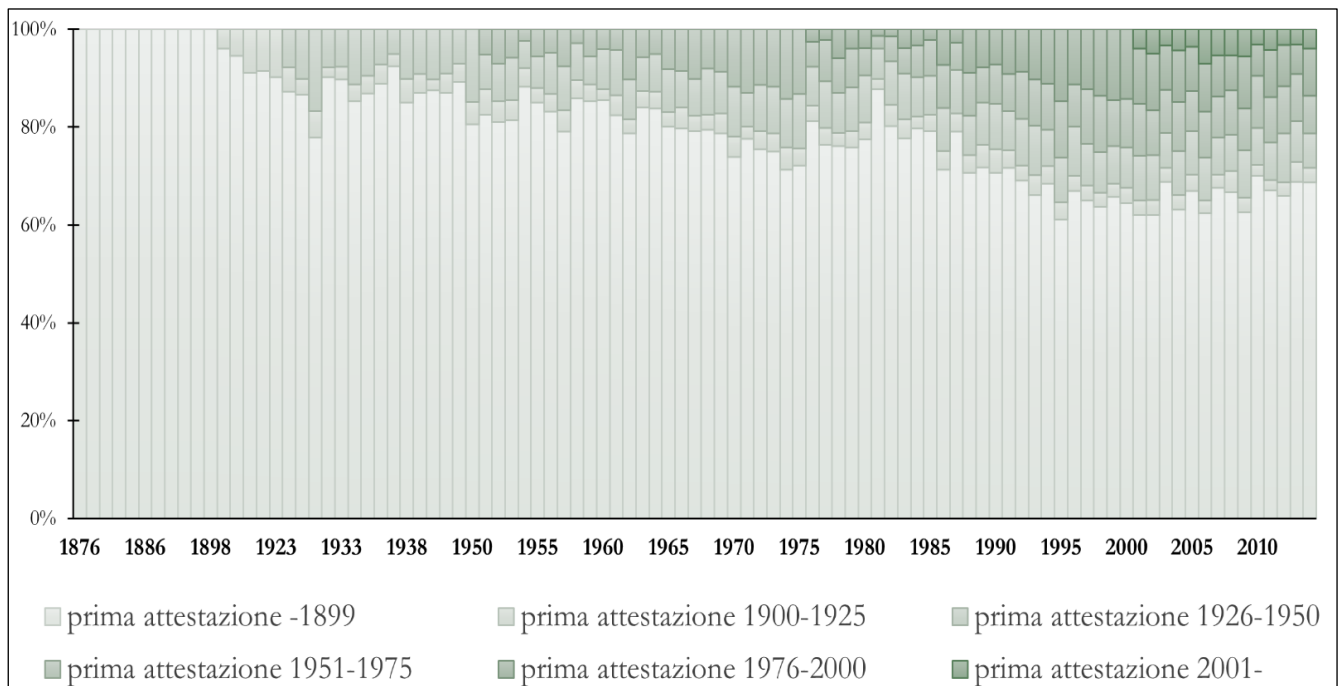
Per ricostruire la stratificazione del lessico estrapolato dall'AGI si è scelto di ricondurre la data di prima attestazione nel corpus alle seguenti fasce temporali: 1876-1899, 1900-1925, 1926-1950, 1951-1975, 1976-2000 e, infine, 2001-2014²⁷. Al riguardo, si tenga presente che la medesima estensione temporale non ospita necessariamente lo stesso numero di fascicoli, perché sino al secondo Dopoguerra la pubblicazione è fortemente irregolare: tra il 1876 e il 1900 escono undici volumi, tra il 1900 e il 1925 appena cinque e tra il 1926 e il 1950 quindici; queste oscillazioni si ripercuotono inevitabilmente su alcune misure lessicometriche, ma con effetti meno marcati rispetto a quelli che a nostro avviso avrebbe una segmentazione basata sul conteggio delle annate, perché essa porterebbe a isolare periodi segnati da una eccessiva escursione cronologica.

Il risultato dell'applicazione dei criteri sin qui presentati è il Grafico 5.

Nel Grafico le barre corrispondono al numero di *type* tecnici attestati nei singoli volumi e illustrano la composizione del lessico specialistico con riferimento all'origine cronologica dei tecnicismi. Nonostante la innegabile complessa articolazione del vocabolario, appare immediatamente evidente la preponderanza della componente pre-novecentesca che, anche nelle annate più recenti, rappresenta oltre il 70% dei tecnicismi attestati: si tratta di un nucleo costante e persistente, mentre il contributo degli altri periodi è nettamente minoritario oltre che decisamente meno stabile.

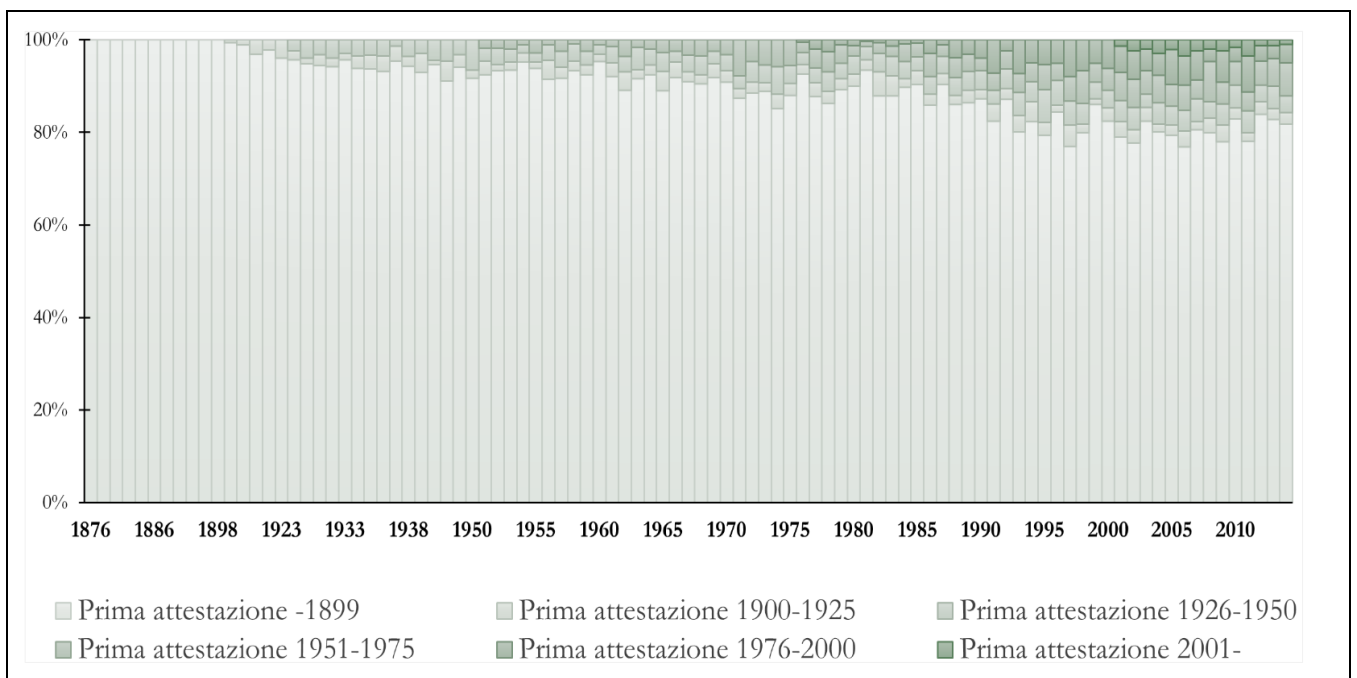
²⁷ In studi precedenti, sempre basati sull'analisi quantitativa del medesimo corpus, si sono adottate periodizzazioni differenti in risposta a diverse domande di ricerca, che vertevano sulla storia della disciplina e consideravano il lessico solo quale indicatore di altri processi. In particolare, in Urraci (2021a) e Urraci (2023) l'AGI è stato periodizzato secondo la segmentazione suggerita dall'analisi delle corrispondenze (Greenacre, 2007), una tecnica di analisi statica multivariata che converte i dati di frequenza delle parole in coordinate su un piano cartesiano e che si basa, dunque, sulle similarità tra i profili lessicali dei sub corpora; i segmenti temporali individuati sono i seguenti: 1876-1905, 1910-1923, 1926-1942, 1950-1988, 1989-1993 e 1994-2014. Queste fasi, riconosciute esclusivamente attraverso l'esame della distribuzione del lessico, si sovrappongono a momenti significativi nella storia della rivista: la prima ripartizione corrisponde alle direzioni di G. I. Ascoli e C. Salvioni, tra loro fortemente omogenee, mentre la seconda si lega alla figura di P. G. Goidanich; il terzo periodo si apre con l'affermazione della Neolinguistica determinata dall'azione e dall'influenza di M. Bartoli e B. Terracini, e si interrompe con la lunga pausa imposta alle pubblicazioni dalla Seconda guerra mondiale; la successiva cesura corrisponde al rinnovamento comportato dall'apertura, nel 1989, a Sig (Società Italiana di Glottologia) e Sli (Società di Linguistica Italiana); l'ultima fase, invece, è segnata dalla nomina a direttore di R. Lazzeroni (1994), artefice di un'autentica rivoluzione nell'impostazione scientifica della rivista.

Grafico 5: *Composizione del vocabolario specialistico (type) con riferimento alla data di prima attestazione dei tecnicismi*



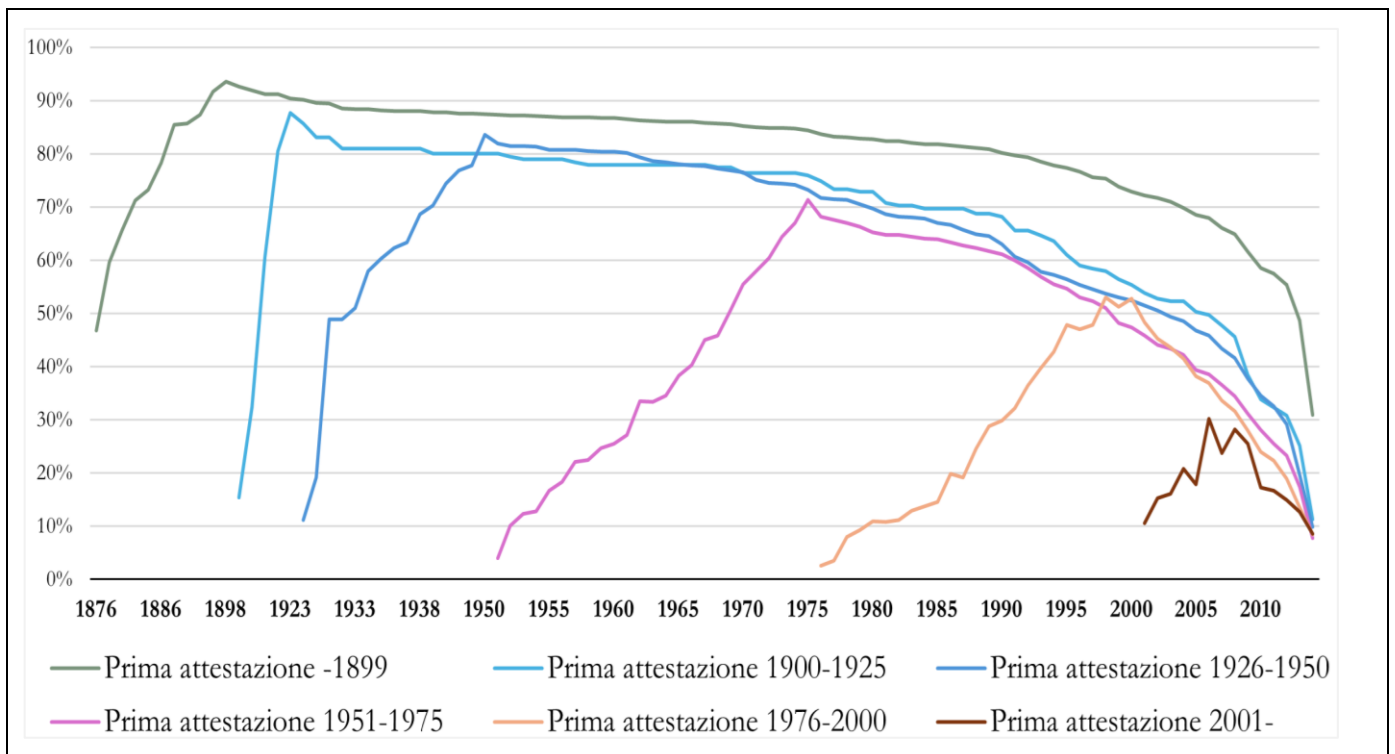
L'evidenza del dato è persino maggiore in relazione ai *token* specialistici (Grafico 6): ancora negli ultimi numeri, la percentuale di tecnicismi pre-novecenteschi è capace di coprire l'85% delle occorrenze del lessico settoriale.

Grafico 6: *Composizione del vocabolario specialistico (token) con riferimento alla data di prima attestazione dei tecnicismi*



Per meglio cogliere le dinamiche richiamate poco sopra riteniamo utile arricchire l'indagine esaminando il tasso di obsolescenza delle singole componenti cronologiche del vocabolario (Grafico 7), in quanto il dato offre una efficace rappresentazione del ciclo vitale dei tecnicismi²⁸.

Grafico 7. Percentuale di tecnicismi vitali, calcolata per ogni categoria cronologica sul totale dei type ad essa associati e proiettata sull'asse del tempo



1. 1876-1899. Vengono impiegati 1.617 tecnicismi differenti, di cui ben 499 sono ancora attestati nell'ultimo numero della rivista. Il tasso di abbandono a dieci anni è dell'8,8%, a trenta dell'11,5% e a cinquanta del 12,6%: questa componente del vocabolario è, in definitiva, eccezionalmente stabile. Non manca comunque un significativo insieme di tecnicismi che non supera la barriera del 1900, centoventi per l'esattezza, tra cui possiamo segnalare *ascitiz̃io* (13)²⁹, *tralignamento* (9), *prefissione* (7), *parassitico* (6) e *patema* (1).
2. 1900-1925. Si registrano 195 prime attestazioni, tra cui *semantica* (1221), *toponimo* (284), *fonema* (209), *occlusiva* (140), *sandhi* (122) e *semiconsonante* (19); di queste neoformazioni ventidue sono ancora presenti nell'ultima annata. Il tasso di obsolescenza a dieci anni è del 19%, a trenta del 21% e a cinquanta del 24,1%: il calo è assai più repentino rispetto al segmento precedente, ma il dato si stabilizza abbastanza rapidamente.
3. 1926-1950. Le aggiunte al vocabolario tecnico sono 519, e comprendono parole come *isoglossa* (83), *agrammaticale* (34), *spirantizzazione* (28), *ipotassi* (17) e *ablaut* (14); solo cinquantuno sopravvivono sino all'ultimo numero considerato. Il tasso di

²⁸ I valori registrati perdono di precisione avvicinandosi al presente, perché il riconoscimento delle ultime attestazioni, per sua natura precario, tende a farsi sempre più aleatorio. Pertanto, i dati non possono essere considerati in termini assoluti ma hanno validità solo in termini comparativi.

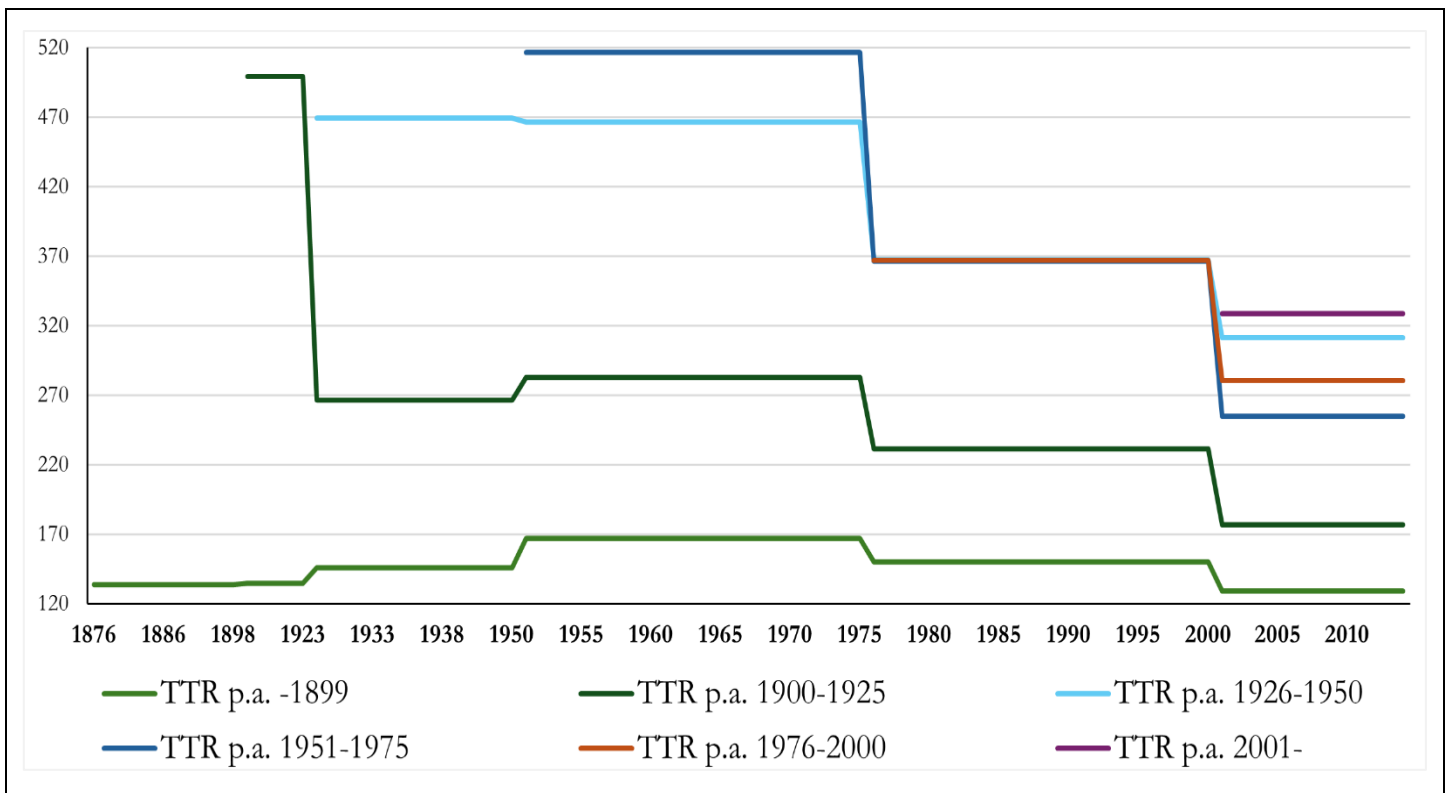
²⁹ Qui e oltre indichiamo tra parentesi tonde la frequenza registrata nel corpus.

- obsolescenza a dieci anni è del 19,7%, a trenta del 30,3% e a cinquanta del 47,6%: una porzione significativa dei neologismi di questi anni cade rapidamente in disuso e, più in generale, all'interno di questo strato lessicale la velocità di abbandono delle parole è molto rapida.
4. 1951-1975. Il numero di neologismi tecnici è decisamente elevato, ben 726, tra cui ci limitiamo qui a segnalare *lessema* (156), *sociolinguistica* (65), *polisemia* (64), *grafema* (42), *morfosintassi* (36), *ipercorrettismo* (12) e *semema* (2); di questi, però, appena cinquantasei sono attestati nell'annata 2014. Il tasso di obsolescenza a dieci anni è del 36% mentre quello a trenta raggiunge il 60,6%: entrambi i valori sono molto elevati.
 5. 1976-2000. Gli ingressi sono numericamente superiori persino rispetto al periodo precedente: 834, di cui settanta sono ancora attestati nell'ultimo numero considerato; tra i nuovi tecnicismi segnaliamo *telico* (140), *inaccusativo* (81), *topic* (69), *frasale* (63), *interfisso* (29), *attante* (28), *coreferente* (20) e *vocoide* (1). Il tasso di obsolescenza è del 76,1%, un dato estremamente elevato anche se potenzialmente influenzato dalla vicinanza all'estremo cronologico del corpus.
 6. 2001-2014. 341 tecnicismi si sommano al vocabolario, un valore più basso rispetto al quarto e al quinto segmento cronologico anche se messo in relazione alla maggiore brevità di questo periodo.

I neologismi accolti tra il 1926 e il 1950 declinano stabilmente con un tasso di abbandono medio annuo dell'1,4%, primariamente associato alla lenta ma costante marginalizzazione della geografia linguistica, a cui buona parte di essi deve essere ricondotta; minor fortuna hanno i numerosi termini introdotti tra il 1951 e il 1975, la metà dei quali non è già più documentata alla fine del secolo: questo elevato tasso di obsolescenza, che si assesta sul 2,4% annuo, è conseguenza del restringimento tematico della rivista, che si chiude a molti degli interessi che aveva accolto nel secondo Dopoguerra, alcuni di essi collaterali rispetto al nucleo centrale della disciplina (cfr. Urraci, 2023: 40-42). Ben diversa è invece la caratterizzazione dei termini di più antica attestazione, e in particolare di quelli risalenti alla fine del 1800: esibiscono una notevole resilienza dimostrata dal ridottissimo tasso di abbandono, pari allo 0,6% annuo. La terminologia ottocentesca e pre-ottocentesca costituisce quindi un nucleo cospicuo e persistente, mentre, come si è visto, le successive innovazioni faticano ad affermarsi stabilmente, spesso perché associate a campi ristretti o a specifiche teorie; possiamo dunque concludere che molti dei concetti fondamentali e basilari della disciplina, ancora oggi imprescindibili in qualsiasi tipologia di analisi, erano già stati elaborati oltre un secolo e mezzo fa: si tratta di una base terminologica che non è stata messa in discussione dalla trasformazione degli studi, dei loro approcci e metodologie.

La fisionomia delle differenti componenti del lessico tecnico, intese come stratificazioni cronologiche associate alle date di prima attestazione, può essere ulteriormente approfondita analizzando il rapporto tra la dimensione del vocabolario specialistico (*type*) e il numero di occorrenze dei tecnicismi (*token*), ossia il *Type-Token Ratio* (TTR), che costituisce un indice di ricchezza lessicale (cfr. Tuzzi, 2003: 127): più il valore è alto più il lessico impiegato è variegato, mentre valori bassi testimoniano la costante ripetizione di un insieme limitato di lessemi; in altre parole, con specifico riferimento alla realtà analizzata, il dato esprime con quale frequenza si ripresentano nel testo gli stessi tecnicismi e, dunque, quanto è varia la terminologia impiegata. L'andamento del TTR nel tempo è riportato nel Grafico 8, all'interno del quale, per una migliore leggibilità, si è scelto di non mostrare i dati relativi alle singole annate ma di riportare i valori medi per ciascuno dei periodi di riferimento.

Grafico 8. TTR della terminologia tecnica nel tempo. I dati sono accorpati in base alla prima attestazione nel corpus e mostrano i valori medi per ciascun periodo



Come si evince dalla figura precedente, la componente ottocentesca del vocabolario specialistico presenta, sin dalle fasi più antiche, una bassissima ricchezza lessicale, dimostrazione del fatto che essa designa principalmente un insieme di concetti che si ripropongono costantemente quasi in ogni articolo e volume, mentre comparativamente pochi sono i termini a bassa frequenza, dunque quelli associati a trattazioni estremamente specifiche: ciò può essere una delle chiavi di lettura della sua resistenza nel tempo. Opposta è la caratterizzazione del lessico introdotto a partire dall'ultimo quarto del Novecento, assai ricco in quanto in esso hanno un maggior peso gli *hapax*³⁰ e tecnicismi che talvolta compaiono in non più di un articolo: una simile fisionomia può segnalare una forte specializzazione della terminologia e dunque il frequente ricorso a tecnicismi che designano campi semantici ristretti o altamente contestualizzati. Certamente elevata è anche la ricchezza lessicale dei termini entrati nel vocabolario tra il 1951 e il 1975; tuttavia, in questo caso, come dimostrano le osservazioni qualitative, il dato riflette una effettiva ricchezza tematica (cfr. Urraci, 2021a: 559) e non è pertanto conseguenza della sola specializzazione del lessico: vi è una grande varietà nelle occorrenze dei tecnicismi perché numerosi sono gli argomenti trattati.

Si può infine osservare che il TTR di tutte le categorie tende a decrescere nel corso del tempo, a dimostrazione del fatto che decadono progressivamente quelle voci che, sin dalle loro prime attestazioni, avevano fatto registrare poche occorrenze, mentre si consolidano quelle caratterizzate da una maggiore versatilità.

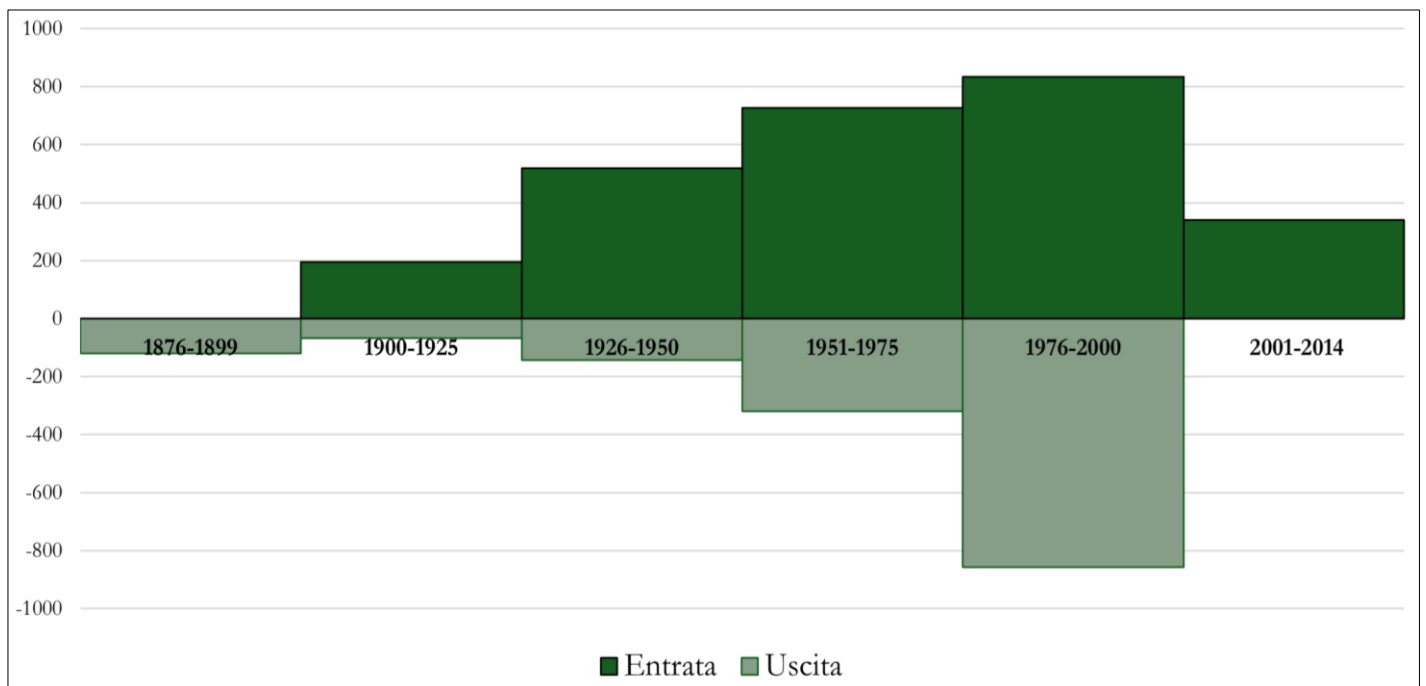
³⁰ Si può considerare, a titolo di esempio, che su 669 tecnicismi introdotti tra il 1989 e il 2014 ben 260 sono *hapax*: il 39% del totale.

7. PRIME ATTESTAZIONI, ULTIME ATTESTAZIONI

Completiamo la descrizione della terminologia scientifica impiegata nell'AGI proponendo, attraverso il Grafico 9, un sintetico bilancio dei *type* entrati e usciti dal vocabolario nel corso di ciascuno dei periodi isolati nella cronologia della rivista³¹. I dati illustrati nella figura confermano la stabile crescita del numero di tecnicismi vitali; fa eccezione solo l'ultimo quarto del XX secolo, fase in cui si registra un numero di prime attestazioni inferiore rispetto a quello delle forme non più documentate. In aggiunta, si può cogliere una progressiva accelerazione nel ricambio dei tecnicismi, i cicli di sostituzione dei quali si fanno sempre più rapidi e agiscono in maggiore profondità: il numero di tecnicismi che entra ed esce annualmente dal dizionario tende a un moderato ma costante aumento.

Tra il 1900 e il 1925 hanno la loro ultima attestazione 67 tecnicismi, tra cui *palatile* (75), *etlissi* (59), *disparizione* (10) e *omeotropia* (7). Negli anni compresi tra il 1926 e il 1950 la stessa sorte riguarda 143 voci; segnaliamo, a titolo di esempio, *faucale* (30), *anfizona* (10), *internamento* (7) e *fema* (1). Nel periodo 1951-1975 le ultime attestazioni sono 320, tra esse si possono ricordare quelle di *esplosiva* (148), *incolume* (92), *degradamento* (47), *livellazione* (30), *cimelio* (24) e *ragguagliare* (10). Infine, tra il 1976 e il 2000 sono ben 857, tra cui *furbesco* (56), *sucedaneo* (52), *glossematica* (13), *spirito della lingua* (11) e *isolessi* (2).

Grafico 9: Prime e ultime attestazioni accorpate per periodi



³¹ Individuiamo come “voci uscite dal vocabolario” i tecnicismi che, a partire dal periodo preso in esame, non sono più attestati nel corpus. Inevitabilmente si tratta di una informazione effimera: non si può escludere il futuro recupero di termini caduti in disuso e, inoltre, in alcuni casi la mancanza di documentazione sarà inevitabilmente pura casualità e non testimonianza di un effettivo, più generale declino o abbandono della parola.

8. CONCLUSIONI

L'AGI, e più in generale la Linguistica italiana, hanno conosciuto, nel corso del tempo, una evoluzione sistematica e sostanziale: i contenuti della rivista, intuibili anche attraverso un semplice spoglio dei titoli degli articoli, dimostrano un significativo slittamento negli interessi e nelle tematiche trattate. Nonostante tali trasformazioni, numerose e profonde, la terminologia non conosce alcuna rivoluzione: la solida e ampia base costituita dal lessico ottocentesco viene infatti intaccata solo superficialmente dall'innesto di neologismi che faticano a radicarsi in quanto prevalentemente circostanziali o legati a specifiche correnti; una caratterizzazione opposta rispetto a quella dei tecnicismi pre-novecenteschi i quali, come indicano chiaramente il ridotto tasso di obsolescenza e il basso TIR, sono estremamente resilienti e richiamano un nucleo concettuale essenziale e imprescindibile.

Tra le trasformazioni osservate, il mutamento principale consiste nella evidente progressiva specializzazione del discorso, segnalata sia dalla maggiore densità di lessico settoriale sia dall'incremento nel numero di tecnicismi vitali, anche se quest'ultimo dato non comporta un aumento sostanziale dei vocaboli specialistici impiegati in ciascuna annata.

In conclusione, crediamo che la ricerca qui presentata evidenzia le potenzialità associate a una lettura da lontano della terminologia linguistica, un approccio che può evidentemente offrire spunti preziosi discernendo le tendenze celate nei testi e permettendo di operare su materiali che, a causa della loro estensione, sarebbero difficilmente gestibili mediante metodologie tradizionali. In definitiva, riteniamo che si sia dimostrata la validità dell'approccio e degli strumenti adottati, i quali potranno pertanto essere efficacemente applicati a eventuali futuri studi aventi per oggetto la stessa materia; in quella occasione sarà però certamente opportuno un rafforzamento delle pratiche di ricerca attraverso un ampliamento del corpus, in termini sia di grandezza sia di varietà, e una più precisa individuazione dell'origine cronologica dei tecnicismi, che dovrà essere determinata in termini assoluti.

RIFERIMENTI BIBLIOGRAFICI

- Beccaria G. L. (2004), *Dizionario di linguistica e di filologia, metrica, retorica*, Einaudi, Torino.
- Bolasco S. (2010), *TaLTaC2.10. Sviluppi, esperienze ed elementi essenziali di analisi automatica dei testi*, Led, Milano.
- Bonin F. et al. (2012), "Lessico settoriale e lessico comune nell'estrazione di terminologia specialistica da corpora di dominio", in Ferreri S. (a cura di), *Lessico e lessicologia*. Atti del XLIV Congresso Internazionale di Studi della Società di Linguistica Italiana (Viterbo, 27-29 settembre 2010), Bulzoni, Roma, pp. 217-230.
- Brambilla A. (2009), "Graziadio Isaia Ascoli e l'accademia scientifico-letteraria. Appunti per un bilancio", in Morgana S., Bianchi Robbiati A. (a cura di), *Graziadio Isaia Ascoli "milanese"*, Led, Milano, pp. 85-107.
- Cabré M. T. (1999), *Terminology. Theory, methods and applications*, John Benjamin, Amsterdam-Philadelphia.
- Cardona G. R. (1969), *Linguistica generale*, Armando editore, Roma.
- Cardona G. R. (1988), *Dizionario di linguistica*, Armando editore, Roma.
- Casadei F. (2011), *Breve dizionario di linguistica*, Carocci, Roma.
- Cortelazzo M. A. (1994), *Lingue speciali. La dimensione verticale*, Unipress, Padova.

- De Felice E. (1954), *La terminologia linguistica di G.I. Ascoli e della sua scuola*, Spectrum, Utrecht-Anvers.
- De Luca M. T. (2014), *Il lessico della linguistica in Lingua nostra (1939-1978)*, Logos Verlag, Berlino.
- Dubois J. et al. (1979), *Dizionario di linguistica*, Zanichelli, Bologna.
- Ducrot O., Todorov T. (1972), *Dizionario enciclopedico delle scienze del linguaggio*, Isedi, Milano.
- Gentile A. (1963), *Lessico di terminologia linguistica*, Liguori, Napoli.
- Giuliano L., La Rocca G. (2008), *L'analisi automatica e semi-automatica dei dati testuali*, vol. I, *Software e istruzioni per l'uso*, Led, Milano.
- Greenacre M. J. (2007), *Correspondence Analysis in Practice*, Chapman & Hall, London.
- Heylen K., De Hertog D. (2015), "Automatic Term Extraction", in Kockafert H. J., Steurs F. (eds.), *Handbook of Terminology*, vol. I, John Benjamins, Amsterdam-Philadelphia, pp. 203-221.
- Iacobini C. (1997), "Le voci di linguistica", in *Lingua e stile*, XXXII, 1, pp. 138-145.
- Lebart L. et al. (1998), *Exploring Textual Data*, Kluwer Ac. Pub, Dordrecht.
- Lucchini G. (2009), "Ascoli e l'«Archivio Glottologico Italiano»", in Meneghetti M. L., Tagliani R. (a cura di), *Tra eadotica e comparatistica: le riviste e la fondazione della filologia romanza*, Edizioni del galluzzo, Firenze, pp. 33-65.
- Moretti F. (2005), *La letteratura vista da lontano*, Einaudi, Torino.
- Morgana S. (2001), "Fasi dell'elaborazione del Proemio ascoliano. Dall'aula dell'Accademica scientifico-letteraria alle pagine dell'Archivio Glottologico Italiano", in Barbarisi G., Decleva E., Morgana S. (a cura di), *Milano e l'Accademia scientifico-letteraria. Studi in onore di Maurizio Vitale*, vol. I, Cisalpino, Milano, pp. 261-376.
- Morgana S. (2009), "Il «Proemio» all'Archivio Glottologico Italiano", in Meneghetti M. L., Tagliani R. (a cura di), *Tra eadotica e comparatistica: le riviste e la fondazione della filologia romanza*, Edizioni del galluzzo, Firenze, pp. 175-206.
- Morrone A. (1993), "Alcuni criteri di valutazione della significatività dei segmenti ripetuti", in Anastex S. J. (ed.), *Secondes Journées internationales d'analyse statistiques des Données Textuelles*, Télécom, Paris, pp. 445-453.
- Pazienza M. T., Pennaccioti M, Zanzotto F. M. (2005), "Terminology extraction: an analysis of linguistic and statistical approaches", in Sirmakessis S. (ed.), *Knowledge Mining. Studies in Fuzziness and Soft Computing*, Springer, Berlin, pp. 255-279.
- Radoni Zucco M. (1973), "Profilo biografico di Graziadio Isaia Ascoli (1829-1907)", in Cortelazzo M. (a cura di), *Graziadio Isaia Ascoli e l'Archivio Glottologico Italiano (1873-1973)*, Società filologica friulana, Udine, pp. 13-51.
- Rovere G. (1989), "Sottocodici e registri in testi tecnici", in *Rivista italiana di dialettologia*, XIII, pp. 135-160.
- Santamaria D. (2018), *Graziadio Isaia Ascoli. Percorsi bibliografici*, Edizioni dell'Orso, Alessandria.
- Severino A. (1937), *Manuale di nomenclatura linguistica*, Le lingue estere, Milano.
- Tuzzi A. (2003), *L'analisi del contenuto: introduzione ai metodi e alle tecniche di ricerca*, Carocci, Roma.
- Tuzzi A. (2018) (a cura di), *Tracing the Life Cycle of Ideas in the Humanities and Social Sciences*, Springer, Berlin.
- Urraci G. (2019), "La parola alle parole. 150 anni di Linguistica in Italia raccontati attraverso l'evoluzione del suo lessico tecnico", in Aldinucci B., Nadal C., Salvatore E., Caruso G., La Grassa M., Carbonara V. (a cura di), *Parola. Una nozione unica per una ricerca multidisciplinare*, Edizioni Università per Stranieri di Siena, Siena, pp. 457-465.

- Urraci G. (2021a), “Parole e storia. La Linguistica in Italia raccontata attraverso il lessico dell’Archivio Glottologico Italiano”, in Diodato F. (a cura di), *Il linguaggio e le lingue: tra teoria e storia*. Atti del I convegno CISPELS, Aracne, Roma, pp. 553-568.
- Urraci G. (2021b), “La Linguistica italiana nelle riviste scientifiche: un confronto diacronico”, in Fresu R., Gualdo R. (a cura di), *In fieri, 3. Ricerche di linguistica italiana*. Atti della III Giornata dell’ASLI per i dottorandi, Franco Cesati Editore, Firenze, pp. 233-246.
- Urraci G. (2023), “L’Archivio glottologico italiano: 150 anni di storia”, in *Archivio glottologico italiano*, CVIII, 1, pp. 19-58.
- Urraci G., Cortelazzo M. A. (2018), “First Steps in Shaping the History of Linguistics in Italy: The Archivio Glottologico Italiano”, in Tuzzi A. (a cura di), *Tracing the Life Cycle of Ideas in the Humanities and Social Sciences*, Springer, Berlin, pp. 87-103.

