

Per studiare il vocabolario del passato

La posizione delle parole in epoca storica¹

Cosimo Burgassi[°], Elisa Guadagnini[°]

[°]Istituto di Linguistica Computazionale «Antonio Zampolli» (CNR – Pisa)

This paper aims to propose a new method for describing the lexicon of a language in a specific period of its history. The first paragraph outlines the two main ideas to be found in the studies concerning both synchronic and diachronic lexicology. In the second paragraph our method for lexical inquiry is presented along with its core concepts, such as textual Corpus Representativeness, Connotation, Connotation Rate (*Quoziente Connotativo*, QC) and word Position in the Center-Periphery Vocabulary Model. The third paragraph sketches two possible research lines, the first one regarding the lexicon of a given historical period (Old Italian), the second dealing with the comparison between two different linguistic historical phases (Old Italian vs. Contemporary Italian).

Keywords: Historical Lexicology, Corpus Linguistics, Word Connotation, Word Position, Center-Periphery Vocabulary Model.

1. Un'idea di vocabolario

Il vocabolario è «l'insieme dei vocaboli appartenenti a una determinata lingua» (cfr. *GradiIt* s.v.). Possiamo intendere questo insieme come una collezione di individui (le parole) isolati o irrelati, comunque considerati singolarmente, oppure concepirlo come un organismo dotato di caratteristiche proprie e descrivibile, quindi, come un'*architettura* risultante dal complesso dei suoi elementi costitutivi.

¹ Questo saggio prosegue e approfondisce la linea di studi iniziata dagli autori nel 2012 e presentata in maniera organica in Burgassi & Guadagnini 2017a; cfr. anche Burgassi & Guadagnini 2014; Guadagnini 2016; Burgassi & Guadagnini 2017b; Burgassi 2020; Guadagnini 2020 e ics. Esprimiamo un sentito ringraziamento ai revisori anonimi della rivista per la loro generosa lettura e per i preziosi suggerimenti e commenti. Nel quadro di un'elaborazione comune, si devono a Elisa Guadagnini i §§ 1-2 e a Cosimo Burgassi i §§ 3-4.

Ricordiamo, per questa seconda idea di vocabolario, l'immagine spaziale della città desunta da Wittgenstein, che De Mauro cita per esprimere il concetto della stratificazione linguistica rispetto alla lingua contemporanea osservata in sincronia (De Mauro 2005: 193-194). Secondo questa metafora, il vocabolario rappresenta lo spazio entro cui si posizionano, gli uni rispetto agli altri, i suoi elementi costitutivi.

La lessicologia storica, da parte sua, tende tradizionalmente a riferirsi alla prima concezione di vocabolario che abbiamo ricordato, intendendo cioè il lessico come un insieme di parole considerate individualmente. In questo senso, la lessicologia storica tende a sovrapporsi alla semantica storica e a coincidere con la lessicografia: la storia delle parole consiste per lo più nella descrizione dei significati assunti o perduti, nel corso dei secoli, dai vocaboli esaminati. Questo tipo di analisi si fonda, come è ovvio, sullo spoglio dei testi conservati (o su una loro selezione), rispetto alle diverse fasi storiche della lingua: che tenda a connotarsi come *corpus-based* (volta quindi alla verifica che un significato già noto del vocabolo sia testimoniato all'interno della documentazione) o *corpus-driven* (derivante dall'interpretazione semantica *a posteriori* delle attestazioni reperite del lessico)², essa mantiene un punto di vista e un obiettivo strettamente semasiologici. I contesti di occorrenza del vocabolo di interesse, insomma, confortano o determinano la descrizione lessicologica, che consiste nella definizione del significato o dell'insieme dei significati veicolati dal vocabolo, posta la sua occorrenza nella documentazione di una specifica epoca.

A nostro avviso, è oggi possibile affiancare a questo tipo di indagine un ulteriore piano di analisi: consentito, come vedremo, dall'esistenza dei grandi corpus testuali, esso mira alla ricostruzione del vocabolario come architettura, non considerandolo un insieme di elementi isolati. L'architettura del vocabolario di una certa fase storica, da questo punto di vista, è data dall'insieme delle parole documentate, che occupano diverse posizioni possibili. La concettualizzazione più efficace di questa idea adotta una metafora spaziale: come si dirà, proponiamo un modello di tipo centro/periferia.

Per ricostruire il vocabolario di una fase storica così concepito proponiamo di ripensare e adattare i metodi e la teoria linguistica che descrivono il lessico in sincronia, con riferimento alle lingue contemporanee.

² Si applicano qui le due categorie dell'indagine *corpus-based* o *corpus-driven*, così come isolate e definite da Tognini-Bonelli 2001: 65, 89.

2. Dalla linguistica statistica alla filologia supra-testuale

La linguistica statistica fornisce strumenti concettuali e metodologici estremamente efficaci per lo studio – sincronico e complessivo – del vocabolario di una lingua contemporanea. Il presupposto di questo tipo di analisi è la presenza di un corpus *rappresentativo* in senso statistico: esso deve cioè comporsi di un insieme di testi che costituisca un *campione* della *popolazione madre*, vale a dire che presenti le stesse caratteristiche qualitative e quantitative della lingua da descrivere. L'analisi del corpus permette la valutazione dei singoli vocaboli attestati, segnatamente grazie ai concetti di frequenza, rango e dispersione: la *frequenza* è il numero delle occorrenze del lessema di interesse, il *rango* è la posizione che esso occupa nella lista dei vocaboli ordinati per frequenza decrescente (la parola di rango 1 è cioè quella in assoluto più ricorrente nel corpus), la *dispersione* valuta come varia la distribuzione del lessema all'interno del corpus (osserva, quindi, se esso ricorra omogeneamente o se tenda invece a localizzarsi entro specifici sotto-corpus – un termine specialistico, per es., tenderà a occorrere in testi di natura tecnica, un vocabolo triviale tipicamente nel parlato informale, ecc.). I parametri appena ricordati permettono di posizionare i lessemi in un modello di vocabolario *a bersaglio*, che vede al centro il *vocabolario di base*, in una fascia intermedia i vocaboli di rango *comune* e nell'anello periferico i vocaboli di basso uso, antiquati od obsoleti, i termini tecnici e specialistici, i regionalismi, ecc. (Figura 1, per il lessico italiano contemporaneo)³.

³ La Figura 1 dettaglia anche le tre sottofasce che costituiscono il vocabolario di base dell'italiano contemporaneo (cfr. De Mauro 2005: 60). Sono *fondamentali* «2.049 vocaboli di altissima frequenza, le cui occorrenze costituiscono circa il 90% delle occorrenze lessicali nell'insieme di tutti i testi scritti o discorsi parlati»; sono di *alto uso* «2.576 vocaboli di alta frequenza, le cui occorrenze costituiscono un altro 6% circa delle occorrenze nell'insieme di tutti i testi scritti o discorsi parlati»; sono di *alta disponibilità* «1.897 vocaboli, relativamente rari nel parlare o scrivere, ma tutti ben noti perché legati ad atti e oggetti di grande rilevanza nella vita quotidiana (*alluce, batuffolo, carrozzeria, dogana* ecc.); «i vocaboli fondamentali, di alto uso e di alta disponibilità (quest'ultimo è il gruppo più esposto al variare della cultura materiale e richiede aggiornamenti relativamente frequenti) costituiscono nell'insieme il *vocabolario di base*».

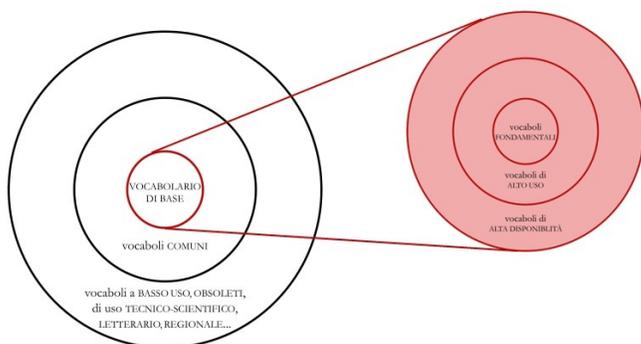


Figura 1. Modello di vocabolario a bersaglio; vocabolario di base (cfr. *GradIt*)

A parer nostro, è possibile mutuare questi concetti chiave e renderli applicabili allo studio del vocabolario di una fase storica (Tabella 1).

Tabella 1. Linguistica statistica e filologia supra-testuale

vocabolario di una lingua contemporanea	vocabolario di una lingua rispetto a una fase storica
linguistica statistica	filologia supra-testuale
corpus rappresentativo	corpus rappresentativo
dati di occorrenza del lessema	dati di attestazione del lessema
frequenza, rango, dispersione	connotazione e quoziente connotativo
modello di vocabolario a bersaglio	modello di voc. di tipo centro/periferia
vocabolario di base	nucleo lessicale

In particolare, riteniamo di poter far corrispondere alla linguistica statistica una *filologia supra-testuale*. La qualifica di *supra-testuale* esprime il livello di analisi che non interessa una singola opera o una specifica tradizione testuale, ma insiste su un insieme di testi (considerabile – come vedremo – rappresentativo di uno stato di lingua). Questa analisi è definita *filologia* perché punta alla piena comprensione del testo: come la linguistica statistica, la filologia supra-testuale si fonda su un corpus; se, però, per la prima i dati quantitativi estrapolati dal corpus sono tendenzialmente conclusivi, per la seconda essi non si accolgono immediatamente ma sono sempre oggetto di valutazione critica.

Rispetto all'analisi lessicologica del vocabolario di una fase storica, la filologia supra-testuale è applicabile secondo un metodo che può essere rappresentato come una sequenza di passaggi successivi (Figura 2).

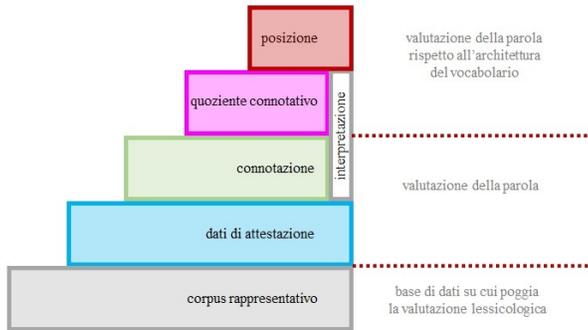


Figura 2. Filologia supra-testuale e metodo di analisi del lessico di una fase storica

Presupposto necessario del metodo che proponiamo è la presenza di un *corpus rappresentativo* (il primo scalino della Figura 2): dato che stiamo analizzando una fase storica, la rappresentatività del corpus non può configurarsi come una rappresentatività di tipo statistico, poiché è impossibile costruire un campione di una popolazione madre inattuabile. Adottiamo quindi un valore debole di rappresentatività, definibile in questo modo: è *rappresentativo* un insieme di attestazioni che permetta la ricostruzione dello stato di lingua di cui raccoglie le testimonianze fino al massimo grado possibile di veridicità, entro i limiti di quanto consente la documentazione (esistente, nota e disponibile)⁴. Assumendo in partenza, quindi,

⁴ La questione della *rappresentatività* dei corpus testuali impiegati in linguistica storica è complessa e assai dibattuta. Con riferimento ai casi che tratteremo nella parte esemplificativa di questo contributo, rispetto quindi all'italiano medievale (cfr. *infra*, § 3), consideriamo *rappresentativo* il *Corpus OVI dell'Italiano antico (Corpus OVI)*, disponibile alla libera consultazione on line all'indirizzo <<http://gattoweb.ovi.cnr.it>>, che raccoglie in via tendenzialmente esaustiva la documentazione italo-romanza edita, dalle Origini alla fine del Trecento. Per le ragioni che possono portare a considerare il *Corpus OVI* rappresentativo dell'italiano antico (e per i limiti entro cui è predicabile tale rappresentatività) cfr. Guadagnini 2016; Burgassi & Guadagnini 2017a: 7-11 e cfr. in proposito le osservazioni di Ernst 2018: 942-943. Corpus rappresentativi di altre epoche storiche, o di altre lingue, richiederebbero naturalmente considerazioni del tutto diverse: pur riconoscendo la centralità di questo aspetto, riteniamo di non poter affrontarlo in questa sede. Ricordiamo le osservazioni incipitarie di Kabatek 2013: 9: «Un corpus

i limiti del caso e della storia, si rinuncia programmaticamente alla nozione di *competenza*, all'ambizione di colmare gli inevitabili vuoti della documentazione, a risalire dal documentato al presumibile: a seconda della fase storica presa in esame, il quadro migliore possibile del corrispondente stato di lingua potrà essere dato dalla documentazione del tempo raccolta in modo idealmente esaustivo, oppure da una selezione della documentazione del tempo operata in base a criteri storico-culturali e storico-linguistici da valutare. Come in tutte le metodologie basate su corpus, la costituzione del corpus stesso è il primo gradino, che funge da fondamenta per tutto il sistema: il metodo che proponiamo in queste pagine rimane sempre entro i confini del corpus scelto come corpus di riferimento, considerato come rappresentativo dello stato di lingua che si vuole indagare.

A partire da tale corpus, la prima fase dell'analisi lessicologica consiste nella raccolta dei *dati di attestazione* del vocabolo di interesse (lo scalino azzurro della Figura 2). Le occorrenze del lessema nel corpus sono osservate secondo due parametri: quello *quantitativo* (quante sono le occorrenze del vocabolo nel corpus? entro quanti testi?) e quello *qualitativo*. Su questo secondo ordine di indagine, i dati di attestazione sono raccolti osservando la distribuzione del lessema sul piano *diacronico* (entro quale intervallo temporale si distribuiscono le occorrenze?), *diatopico* (a quali varietà diatopiche sono riconducibili i testi entro cui si distribuiscono le occorrenze?) e *diatestuale* (o *diagenerico*: entro quali generi testuali o tradizioni discorsive risulta attestato il lessema?).

A partire dai dati di occorrenza estrapolati dal corpus di riferimento, è possibile dare una prima valutazione lessicologica del lessema grazie alla *connotazione*; essa risulta da un atto interpretativo e riguarda la parola considerata individualmente, non rispetto all'architettura del vocabolario (gradino verde della Figura 2). La connotazione valuta i dati di attestazione di un lessema secondo quanto esso risulta diffuso e disperso nel corpus di riferimento: essa esprime in forma sintetica le caratteristiche di occorrenza del vocabolo. Come è noto, in linguistica il termine *connotazione* è stato impiegato più volte e in modi diversi: in opposizione alla *denotazione*, esso tende a designare i tratti di ordine soggettivo che contribuiscono al significato di una parola, come il valore affettivo, allusivo e simili. Questi tratti sono riconoscibili prevalentemente grazie alla competenza linguistica dei parlanti; posto che tale competenza non è attingibile per un'epoca

representativo para la historia de una lengua es una *construcción teóricamente imposible* ya que la lengua, aunque solo se manifieste en textos, no es la suma de los textos sino algo distinto. Un corpus representativo es, además, una *construcción empíricamente imposible* ya que la producción de textos que se ha archivado no corresponde a más que una mínima parte de la producción lingüística total...» (corsivi nostri).

storica, proprio rispetto a un'epoca storica riteniamo opportuno servirci del termine *connotazione* per denominare l'insieme dei tratti di un vocabolo che non pertengono al suo contenuto semantico strettamente referenziale (o descrittivo). Come è ovvio, questi tratti dipendono esclusivamente dalle caratteristiche di occorrenza del lessema nel corpus di riferimento poiché, come abbiamo detto, il metodo che proponiamo rinuncia programmaticamente a tenere in conto qualunque piano dell'indagine che non sia fondato sul documento⁵.

A partire dalla connotazione, è possibile stimare il *quoziente connotativo* (QC) per fornire un'ulteriore valutazione lessicologica (scalino rosa della Figura 2). Il QC proietta la connotazione del lessema nell'architettura del vocabolario: a questo livello, quindi, la parola non è più considerata da sola ma è rapportata all'insieme dei vocaboli presenti nel corpus, ciascuno con la propria connotazione. Il QC esprime, quindi, il rapporto tra i possibili valori della connotazione delle parole nel corpus di riferimento: beninteso, non si tratta di un rapporto quantitativo, poiché tali valori non sono numeri (estrapolati statisticamente) ma riflettono un giudizio interpretativo di qualità. Definiamo il QC un *quoziente* per porre l'accento sul fatto che esso esprime una correlazione, benché questa – come detto – non sia quantitativa.

Per analogia con l'articolazione del lessico contemporaneo in VdB, parole comuni e altri vocaboli (effettuata dalla linguistica statistica sulla base del rango delle parole), distinguiamo tre valori del QC: *basso*; *medio*; *alto*. Hanno un QC *basso* i vocaboli che risultano altamente diffusi e altamente dispersi nel corpus, cioè che ricorrono con molte attestazioni (relativamente alla quantità delle occorrenze totali raccolte nel corpus stesso), distribuite entro l'intero arco cronologico considerato, in tutte le varietà diatopiche documentate, in tutte le tipologie testuali e tradizioni discorsive presenti. Il QC *medio* è attribuibile ai vocaboli che occorrono nel corpus con frequenza assoluta relativamente bassa (pur potendo essere altamente dispersi), oppure con frequenza non bassa ma altamente localizzata (cioè con bassa dispersione). Hanno un QC *alto* i vocaboli bassamente diffusi e bassamente dispersi nel corpus o, per dirla positivamente, i vocaboli che

⁵ Negli studi linguistici i contenuti semantici non descrittivi che tradizionalmente individuano la connotazione dipendono dalla competenza dei parlanti (cfr. Blank 1997: 63-64) e contribuiscono a determinare i diversi piani della variazione intralinguistica nel 'diasistema'. Per esempio, si è sottolineato come vocaboli con lo stesso contenuto referenziale (o descrittivo) possono opporsi nel diasistema soltanto da questo punto di vista, per 'diversità di architettura' (Eugenio Coseriu) o 'marcatezza diasistemica'. Questi concetti di connotazione e di architettura si fondano sulla competenza dei parlanti: a questo riguardo sono pertanto affatto diversi da quelli da noi proposti. Il contenuto semantico prettamente referenziale, d'altra parte, è l'oggetto di studio tradizionale della lessicologia/lessicografia storica.

occorrono assai di rado e soltanto in modo molto localizzato sul piano diacronico, diatopico, diatessuale, in combinazione di una coppia di questi piani, o in tutti e tre simultaneamente. Come detto, i valori della connotazione interpretano i dati di attestazione del lessema (sul piano quantitativo del numero delle occorrenze, sul piano qualitativo della loro distribuzione e dispersione): il QC è, a sua volta, il risultato di un atto interpretativo teso a porre il vocabolo in correlazione al vocabolario come insieme, e non è calcolabile mediante indici numerici.

Il valore del QC determina la *posizione* del lessema nell'architettura del vocabolario: i vocaboli con QC *basso* occupano la posizione nucleare, quelli con QC *medio* l'area media, quelli con QC *alto* la periferia. Il modello di vocabolario presupposto, quindi, è di tipo *centro/periferia* (e non a bersaglio come per il lessico contemporaneo): in questo modello, la parte centrale – che rispetto a una lingua contemporanea è occupata dal VdB – individua il *nucleo lessicale* del vocabolario descritto (Figura 3).

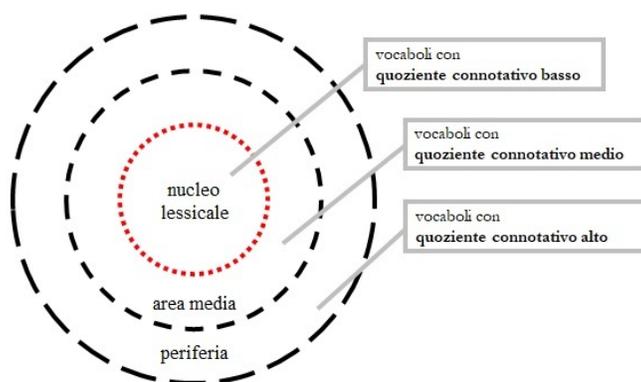


Figura 3. Architettura del vocabolario e posizioni relative dei lessemi

La differenza grafica che oppone le linee continue del bersaglio (Figura 1) a quelle tratteggiate del tipo centro/periferia (Figura 3) esprime iconicamente la differenza concettuale che passa tra il rango e il QC. Come si diceva, quest'ultimo riflette un atto interpretativo e le tre fasce entro cui si distribuiscono i suoi valori possibili rappresentano aree tendenziali dai confini non discreti (non numerici). Ciò avviene anche, ad esempio, per la classificazione delle zone climatiche, che non seguono linee di confine nette. Tale paragone è estendibile anche ai criteri che definiscono le suddivisioni areali: se, per valutare la posizione della parola nel vocabolario, il criterio puramente quantitativo (la frequenza nel corpus) è combinato con quello qualitativo (la dispersione), analogamente il clima di una zona dipende

dalla latitudine (fattore assimilabile al criterio quantitativo) e da alcune variabili come l'altitudine, la vicinanza dal mare, ecc. (fattori che, nell'insieme, sono assimilabili al criterio qualitativo).

Nell'impossibilità di affidarsi a indici numerici, alcuni lessemi ricorrono nel corpus con caratteristiche di attestazione tali da rendere certa l'attribuzione di un valore del QC e della relativa posizione nel vocabolario: nel paragrafo seguente presenteremo tre casi di questo tipo.

Mentre i valori *basso* e *alto* del QC riflettono un orientamento complessivamente omogeneo dei dati di attestazione (alta frequenza e alta dispersione vs bassa frequenza e bassa dispersione), e sono quindi attribuibili con relativa immediatezza, il riconoscimento del valore *medio* può presentare un grado di opinabilità. Infatti, in presenza di dati di attestazione di segno discordante – per esempio nel caso di lessemi di alta frequenza ma con bassa dispersione nel corpus, o all'inverso di bassa frequenza ma con alta dispersione –, l'assegnazione di uno o di un altro valore del QC non è immediata. Mentre la linguistica statistica fa spesso iniziare convenzionalmente i valori medi del rango a partire dalla prima coppia di lessemi con identica frequenza, nel caso del QC non è possibile dare indicazioni assolute: si dovrà valutare caso per caso a quale parametro attribuire maggior peso relativo, per giungere alla valutazione sintetica complessiva. Per esempio, una frequenza relativamente alta correlata a una bassa dispersione può far propendere per l'attribuzione al lessema di un QC basso (e dunque di una posizione periferica), se si ritiene di poter riconoscere in quel lessema un termine, testimoniato da una specifica tradizione discorsiva.

3. Il metodo in pratica

In questa sezione intendiamo tracciare alcune linee di studio sulla base dei principi generali presentati fin qui: esse mostrano impieghi e declinazioni possibili del metodo proposto e hanno valore puramente esemplificativo. In particolare, applicheremo il metodo a uno stato di lingua storico e al confronto tra due diversi stati di lingua: proporremo dapprima l'analisi lessicologica di alcuni vocaboli documentati in italiano antico (*cibo*, *crollare*, *gensore*; § 3.1), confronteremo poi italiano antico e italiano contemporaneo (rispetto a *cibo*, *menare*, *ambiguo*; § 3.2).

3.1 Rispetto a uno stato di lingua: *cibo*, *crollare* e *gensore* in italiano antico

Adottando la periodizzazione tradizionale, consideriamo come *italiano antico* la fase storica che va dai primi documenti italo-romanzi alla fine del Trecento⁶: assumiamo come corpus di riferimento il *Corpus OVI*, ritenendo che esso possa essere considerato rappresentativo per questa fase dell'italiano⁷.

Volendo fornire un'analisi lessicologica dei lessemi *cibo*, *crollare* e *gensore* ('più gentile') in italiano antico, raccogliamo per prima cosa i dati di attestazione di questi vocaboli nel *Corpus OVI* (Tabella 2).

Tabella 2. Dati di attestazione di *cibo*, *crollare* e *gensore* nel *Corpus OVI*

	<i>cibo</i>	<i>crollare</i>	<i>gensore</i>
ordine quantitativo	3.782 occorrenze in 338 testi	302 occorrenze in 104 testi	5 occorrenze in 5 testi ⁸
ordine qualitativo			
piano diacronico	1243 <i>ca</i> – XV <i>in</i> .	1275 – XV <i>in</i> .	1239/48 – XIII <i>ex</i> .
piano diatopico	tutte le aree geografiche ⁹	tutte le aree geografiche ¹⁰	area settentrionale e toscana
piano diatեսtuale	tutte le tipologie testuali	tutte le tipologie testuali	lirica

Sulla base dei dati di attestazione, è possibile attribuire ai tre lessemi analizzati le connotazioni descritte nella Tabella 3.

⁶ Per una sintesi delle questioni relative al concetto, in verità piuttosto problematico, di *italiano antico* cfr. per es. Tomasin 2019: 140-149; Burgassi & Guadagnini (2017a: 19-25) e la bibliografia ivi citata.

⁷ Cfr. *supra*, nota 4.

⁸ Riprendendo le sigle in uso nel *Corpus OVI* (per cui cfr. la bibliografia associata), questi sono i testi di occorrenza di *gensore*: Guido Faba, *Gemma*, 1239/48 (bologn.); Re Giovanni (ed. Panvini), XIII pm. (tos.); Guittone, *Lettere in versi*, a. 1294 (tos.); *Disputatio roxe et viole*, XIII (lomb.); *Poes. an. mant.*, XIII/XIV; *TLIO* s.v. *gensore* reperisce un'ulteriore occorrenza in Folco di Calavra (ed. Fratta), XIII (tos.).

⁹ Riprendendo le etichette di localizzazione diatopica del *Corpus OVI*, si osserva ricorrere il lessema in area toscana (fior., aret., pis., sen., prat., pist., lucch.), settentrionale (bologn., mil., venez., mant., pav., lig.), mediana e meridionale (urbin., napol., tod., abruzz., aquil., perug., rom., cass., castell., sab.) e siciliana.

¹⁰ Riprendendo le etichette di localizzazione diatopica del *Corpus OVI*, si osserva ricorrere il lessema in area toscana, settentrionale, mediana e meridionale e siciliana.

Tabella 3. Connotazione di *cibo*, *crollare* e *gensore* in italiano antico

	<i>cibo</i>	<i>crollare</i>	<i>gensore</i>
connotazione	lessema altamente diffuso e altamente disperso nel corpus	lessema bassamente diffuso e altamente disperso nel corpus	lessema bassamente diffuso e bassamente disperso nel corpus

Queste connotazioni determinano i quozienti connotativi (QC) espressi nella Tabella 4.

Tabella 4. Quoziente connotativo di *cibo*, *crollare* e *gensore* in italiano antico

	<i>cibo</i>	<i>crollare</i>	<i>gensore</i>
quoziente connotativo	QC basso	QC medio	QC alto

I rispettivi QC permettono di posizionare i vocaboli nel vocabolario secondo le rispettive posizioni illustrate nella Figura 4.

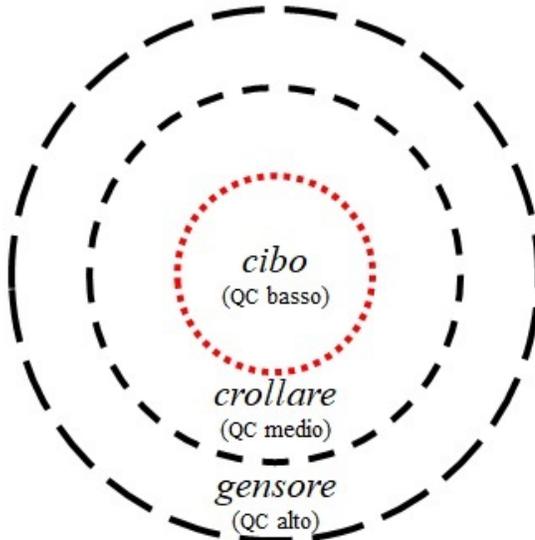


Figura 4. La posizione di *cibo*, *crollare*, *gensore* nell'architettura del vocabolario italiano antico

È importante sottolineare che il valore lessicologico attribuito a questi lessemi non cambierebbe, ove si adottasse un diverso corpus di riferimento o esso – come accade – cambiasse (per l'adozione di nuove edizioni dei medesimi testi, o per l'immissione di testi prima assenti, ignoti o inediti). *Cibo* ha valori di diffusione e dispersione talmente alti che non basterebbero decine, ma neppure centinaia di occorrenze in più o in meno a mutarne sensibilmente la connotazione (e quindi il QC e la posizione). *Gensore* si caratterizza come un francesismo della tradizione lirica pre-stilnovistica: potrebbero aggiungersi nuove occorrenze di questo vocabolo – una esterna al corpus di riferimento, come si è detto nella nota 8, è in effetti registrata nel *TLIO* –, ma non cambierebbe la natura peculiare di questo vocabolo, tipico della tradizione discorsiva lirica gallicizzante, attiva non oltre il Duecento al Nord e lungo la linea Bologna-Firenze. *Crollare*, infine, resterà sempre assai distante dalle caratteristiche di connotazione di vocaboli come *cibo*, se anche fossero immessi nel corpus centinaia di nuovi testi¹¹.

3.2 Confronto tra due stati di lingua: *cibo*, *menare* e *ambiguo* in italiano antico e italiano contemporaneo

La seconda linea esemplificativa che tratteremo riguarda il confronto tra due diversi stati di lingua. Questo tipo di confronto può essere stabilito, di per sé, tra qualsivoglia coppia (o serie) di stati di lingua isolati sulla base della periodizzazione che si accoglie, sia essa orientata – per esempio – in senso strettamente geometrico (scansione cronologica per secoli), genericamente storico-culturale (epoche intese in modo tradizionale: Medioevo, Umanesimo, Rinascimento, ecc.) o, ancora, specificamente storico-linguistico (secondo fatti e avvenimenti caratterizzanti della storia della lingua¹²).

Rispetto ad altre scelte possibili, riprendendo e approfondendo una prospettiva di analisi da noi già adottata, consideriamo un arco cronologico esteso e confrontiamo il lessico italiano medievale e quello contemporaneo; eventualmente, quale ricerca supplementare, entro queste coordinate si potrà porre l'analisi degli stati intermedi. In questa sede, proporremo quindi un raffronto tra la valutazione lessicologica dei dati restituiti dal corpus di riferimento per l'italiano antico (*Corpus OVI*; cfr. § 3.1) e i risultati dell'analisi statistica che definiscono il vocabolario italiano contemporaneo, così come quest'ultimo è ricostruito dal *GraDIt*. Questo

¹¹ Va specificato che il metodo è maggiormente efficace per la valutazione dei morfemi lessicali: va tenuto presente che su singoli lessemi affissati o derivati possono influire i tratti connotativi (nel senso di connotazione su definito) del morfema derivazionale.

¹² Per esempio, per l'italiano si può rimandare alla periodizzazione del corpus *MIDIA*, dedicato principalmente allo studio della morfologia (D'Achille & Grossmann 2017).

strumento – lo ricordiamo – assegna due ordini di marche: il primo riguarda il lessema e il secondo ogni sua accezione, combinando in questo modo informazioni lessicali e onomasiologiche. In questo contributo ci interessa il primo ordine di informazione lessicologica, relativa al lessema e non specificamente a una sua accezione¹³.

Com'è ovvio, l'analisi dei dati per questi due stati di lingua poggia su tipi di fonti non completamente sovrapponibili, in ragione della loro differente (e fisiologica) attingibilità. Se per una fase storica, infatti, la valutazione lessicale si fonda *soltanto* sui testi scritti superstiti (le uniche testimonianze di cui disponiamo), l'esame statistico per la fase contemporanea si basa *anche* sui discorsi dei parlanti, rappresentando questi – evidentemente – una fonte attingibile, e opera su un corpus campionato¹⁴. Il raffronto tra la lingua contemporanea e una fase linguistica storica risulta quindi inattuabile là dove la natura delle fonti non è in alcun modo confrontabile: è il caso di quelle parole che tendono a non essere registrate né dalla scrittura né dai discorsi orali, ma che sono ben conosciute ai parlanti e ben presenti nell'uso endofasico del lessico mentale. Questa categoria lessicale, costituita dalle parole di 'alta disponibilità' (AD), è identificata esclusivamente dalla competenza linguistica dei parlanti, che può essere attinta per la lingua contemporanea ma non può esserlo per una fase storica. Il raffronto tra il lessico di una fase storica e quello contemporaneo, pertanto, non è proponibile per la categoria delle parole ad alta disponibilità: in generale, nella documentazione di una fase storica, tendono a essere sottorappresentate le denominazioni di oggetti del quotidiano come gli utensili, i cibi e simili, la cui attestazione dipende spesso da circostanze di citazione e fattori di occorrenza casuali e desultori (questi vocaboli compaiono, per esempio, nelle liste degli strumenti degli artigiani, nelle ricette, ecc.)¹⁵.

Con queste premesse, passiamo ora a un esempio di confronto possibile tra i due stati di lingua suddetti. In parallelismo con § 3.1, consideriamo una triade di

¹³ Ciò detto, il metodo che proponiamo è applicabile anche a questa indagine di tipo onomasiologico: cfr. la discussione sul concetto di *marcatezza* in Burgassi & Guadagnini 2017a: 25-28.

¹⁴ Cfr. § 2. Per l'impiego di fonti anche orali ricordiamo, per esempio, che i vocaboli marcati FO (fondamentali) nel *GrADIt* rappresentano le parole in assoluto più ricorrenti nell'«insieme di tutti i testi scritti o discorsi parlanti» (cfr. *supra* nota 3).

¹⁵ Cfr. *supra* nota 3. Questa categoria di lessico, quindi, non è individuata con metodi statistici: «Sono le parole che diciamo o scriviamo raramente, ma che pensiamo con grande frequenza. Queste parole, come si è capito da circa vent'anni, rischiano di restare fuori dalle liste di frequenza e di uso. Per trovarle, per trovare le più importanti occorre servirsi di altre vie. [...] dobbiamo pazientemente interrogare gruppi diversi di parlanti per isolare un po' alla volta il vocabolario di più alta disponibilità» (De Mauro 1980: 150-151).

parole di diversa categoria grammaticale: riprendiamo il caso del sostantivo *cibo*, a cui affianchiamo il verbo *menare* e l'aggettivo *ambiguo*.

Partiamo valutando la posizione di questi lessemi nel vocabolario antico: percorriamo quindi in successione gli scalini della Figura 2, come si è fatto in § 3.1. Per prima cosa registriamo i dati di attestazione nel corpus di riferimento (Tabella 5), con alcune precisazioni (per praticità, qui sotto e in seguito replichiamo quanto si è detto su *cibo* nelle Tabelle 2-4).

Tabella 5. Dati di attestazione di *cibo*, *menare* e *ambiguo* nel *Corpus OVI*

	<i>cibo</i>	<i>menare</i>	<i>ambiguo</i>
ordine quantitativo	3.782 occorrenze in 338 testi	oltre 16.300 occor- renze in oltre 869 testi	7 occorrenze in 6 testi
ordine qualitativo			
piano diacronico	1243 <i>ca</i> – XV <i>in</i> .	XIII <i>in</i> . – XV <i>in</i> .	XIV sec.
piano diatopico	tutte le aree geografiche	tutte le aree geografiche	sei testi fiorentini, un testo veneto ¹⁶
piano diatestuale	tutte le tipologie testuali	tutte le tipologie testuali	testi di traduzione diretta dal latino o in stretta relazione con fonti latine

Rispetto ai dati forniti, precisiamo che, per quanto riguarda l'*ordine quantitativo* delle occorrenze di *menare*, il numero è sicuramente sottostimato: non si sono prese in considerazione, infatti, le forme *men*, *men'*, *mene*, *meno*, *minali*, *minu*, per il rumore semantico causato dall'omografia con altri vocaboli o sintagmi (*meno* aggettivo /avverbio; il composto pronominale per *me ne*; *minale* sostantivo da *mina* 'unità di peso', risalente al latino *hemina* 'unità di misura pari a metà sextarius', cfr. Stussi 1992: 252). A proposito del *piano diatestuale* di *ambiguo*, poi, si precisa che alla tipologia delle traduzioni dirette dal latino appartengono la Deca terza e quarta di Tito Livio in volgare (che restituiscono ciascuna un'occorrenza del lessema); si possono ascrivere ai testi in stretta relazione con fonti latine il *Trattato della scienza* di Jacopo Passavanti (due occorrenze) e il tardo (*ante*

¹⁶ Riprendendo le sigle in uso nel *Corpus OVI* (per cui cfr. la bibliografia associata), questi sono i testi di occorrenza di *ambiguo*: Boccaccio, *Filostrato*, 1335-36 (?); Deca quarta di Tito Livio, a. 1346 (fior.); Deca terza di Tito Livio (B, L. I-II), XIV m. (fior.); Jacopo Passavanti, *Tratt. scienza*, c. 1355 (fior.); Boccaccio, *Esposizioni*, 1373-74; Comm. *Arte Am.* (D), a. 1388 (ven.).

1388) commento veneto all'*Ars amandi* di Ovidio (unico testo non fiorentino, per un'occorrenza)¹⁷.

In base ai dati di attestazione, possiamo valutare la connotazione dei tre lessemi in oggetto (Tabella 6).

Tabella 6. Connotazione di *cibo*, *menare* e *ambiguo* in italiano antico

	<i>cibo</i>	<i>menare</i>	<i>ambiguo</i>
connotazione	lessema altamente diffuso e altamente disperso nel corpus	lessema altamente diffuso e altamente disperso nel corpus	lessema bassamente diffuso e bassamente disperso nel corpus

Proiettando queste connotazioni sul piano dell'architettura del vocabolario, ricaviamo i valori del QC (Tabella 7):

Tabella 7. Quoziente connotativo di *cibo*, *menare* e *ambiguo* in italiano antico

	<i>cibo</i>	<i>menare</i>	<i>ambiguo</i>
quoziente connotativo	QC basso	QC basso	QC alto

Rispetto all'architettura del vocabolario italiano antico, dunque, possiamo predicare una posizione nucleare per *cibo* e *menare*, di contro a una posizione periferica per *ambiguo*.

Valutata la posizione di queste parole nel vocabolario medievale, passiamo come annunciato al confronto con i ranghi contemporanei, attribuiti dal *GrADIt*. Il sostantivo *cibo* ha, in questo strumento, la marca FO (vocabolo fondamentale, appartenente al VdB); oltre al significato 'ciò che si mangia', nella voce sono registrate anche le accezioni '(fig.) nutrimento spirituale', marcato CO, e 'momento del pasto', marcato OB. Notiamo che questa articolazione semantica coincide con quella documentata per il lessema nella fase antica (cfr. *TLIO* s.v. *cibo*). Il verbo *menare* presenta la marca CO (vocabolo comune, non appartenente al VdB), in quanto sono CO le accezioni 'aspettare con violenza' (in espressioni del tipo *menare colpi a destra e a manca*) e 'picchiare, malmenare' (in espressioni

¹⁷ *Ambiguo* occorre poi in due opere boccacciane (*Filostrato* ed *Esposizioni sopra la Comedia di Dante*; una sola occorrenza in ciascun testo): in questo caso, si tratta verosimilmente di un'opzione lessicale da parte di Boccaccio in direzione del latinismo (di natura, qui, prettamente stilistica).

popolari del tipo *è stato menato da un gruppo di teppisti*). È marcato BU (basso uso) il primo significato, ‘portare, condurre’, che corrisponde al significato proprio del lessema nella fase antica (cfr. *TLIO* s.v. *menare*): esso dipende dall’evoluzione del latino *minari* ‘minacciare’ a ‘condurre minacciando (specialmente il bestiame)’ documentata già in epoca tardo-imperiale¹⁸. *GraDIt* assegna all’aggettivo *ambiguo* la marca AU (vocabolo di alto uso, appartenente al VdB). Il vocabolo ha, oggi come in antico, il significato principale ‘di esito o valore incerto’¹⁹.

Alla luce di questi elementi, possiamo dunque raffrontare le posizioni dei lessemi nel vocabolario italiano antico e nel vocabolario italiano contemporaneo (Figura 5).

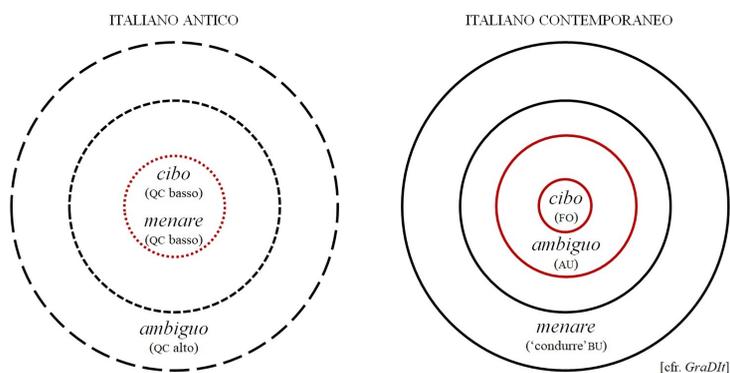


Figura 5. La posizione di *cibo*, *menare* e *ambiguo* in italiano antico e contemporaneo

Come si vede, *cibo* risulta occupare un’analoga posizione centrale nel medioevo e nell’italiano contemporaneo; *menare* risulta occupare una posizione centrale in antico e periferica oggi; *ambiguo* risulta occupare una posizione periferica in antico e nucleare oggi. I tre lessemi selezionati individuano, quindi, tre possibili fattispecie:

- uguaglianza della posizione;
- mutamento della posizione dal centro alla periferia;

¹⁸ Cfr. *FEW* s.v. *minare*. Il significato ‘percuotere’, che deriva anch’esso dalla pratica di spingere avanti il bestiame minacciandolo e pungolandolo, è invece originariamente romanesco (cfr. *DELLn* s.v. *menare* e bibliografia ivi citata).

¹⁹ Cfr. *TLIO* e *GraDIt* s.v. *ambiguo*. *GraDIt* marca come OB (obsoleto) il sottosignificato ‘dubbioso, irresoluto, perplessa’ e individua un’accezione TS (linguaggio tecnico specialistico) nell’ambito filosofico della logica formale.

- mutamento della posizione dalla periferia al centro²⁰.

Rispetto a quanto emerge tradizionalmente dagli studi di lessicologia storica, che registrano il momento dell'ingresso di un lessema nella documentazione (data di prima attestazione) o la sua uscita dalla lingua dell'uso (per le parole oggi obsolete o antiquate), l'adozione della prospettiva qui descritta porta un complemento di informazione: la linea di ricerca proposta permette, infatti, di tracciare l'eventuale mutamento di posizione di una parola nel tempo e, in generale, di ricostruire i diversi assetti del vocabolario in diacronia.

4. Conclusioni

Per concludere, ci preme ancora una volta sottolineare che il metodo di analisi lessicologica che proponiamo può essere applicato a qualsiasi dominio linguistico e a qualunque fase storica, posta la presenza di un corpus considerabile come rappresentativo. Il punto di partenza è l'idea di lessico considerato come un'*architettura*, vale a dire un insieme organico di elementi che può essere descritto nel suo complesso. D'altra parte, ogni elemento costitutivo di questa architettura (la singola parola) occupa una posizione valutabile sia in assoluto sia rispetto agli altri elementi e all'insieme: secondo la nostra proposta, il primo di questi valori è espresso dalla *connotazione*, il secondo dal *quoziente connotativo* (QC).

References

- Blank, A., 1997. *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*. Tübingen: Niemeyer.
- Burgassi, C. 2020. «Parole di conforto». Tappe storiche di una concorrenza lessicale. In *Studi di filologia offerti dagli allievi a Claudio Ciociola*. Pisa: ETS: 61-78.
- Burgassi, C., Guadagnini, E. 2014. Prima dell'*indole*. Latinismi latenti dell'italiano. *Studi di lessicografia italiana* 31: 5-43.
- Burgassi, C., Guadagnini, E. 2017a. *La tradizione delle parole: Sondaggi di lessicologia storica*. Strasbourg: ÉLiPhi.
- Burgassi, C., Guadagnini, E. 2017b. L'integrazione lessicale di *facile* nel vocabolario italiano. In E. Guadagnini, G. Vaccaro (eds.), *Rem tene, verba sequentur. Latinità e medioevo romanzo: testi e lingue in contatto*. Alessandria: Edizioni dell'Orso: 157-177.

²⁰ Avere corretta consapevolezza della posizione delle parole in fase storica, al di là della (e talvolta in opposizione alla) competenza dei parlanti odierni, costituisce un elemento importante anche sul versante filologico: per alcuni esempi cfr. Burgassi & Guadagnini 2017a: 89ss., 158-159.

- Corpus OVI* = Larson, P., et al. (eds.). *Corpus OVI dell'Italiano antico*. <<http://gattoweb.ovi.cnr.it>>.
- D'Achille, P., Grossmann, M. 2017. *Per la storia della formazione delle parole in italiano. Un nuovo corpus in rete (MIDIA) e nuove prospettive di studio*. Firenze: Cesati.
- De Mauro, T. 1980. *Guida all'uso delle parole*. Roma: Editori Riuniti.
- De Mauro, T. 2005. *La fabbrica delle parole. Il lessico e problemi di lessicologia*. Torino: UTET.
- DELIn* = Cortelazzo, M., Zolli, P. (eds.) 1999. *Il nuovo etimologico. Dizionario etimologico della lingua italiana*. Bologna: Zanichelli.
- Ernst, G. 2018. Recensione di Burgassi & Guadagnini 2017a. *Zeitschrift für romanische Philologie* 134(3): 941-950.
- FEW* = von Wartburg, W., et al. (eds.) 1922-. *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes*. <<https://apps.atilf.fr/lecteurFEW/>>.
- GraDIt* = De Mauro, T. (ed.) 1999. *Grande dizionario italiano dell'uso*. Torino: UTET.
- Guadagnini, E. 2016. Lessicografia, filologia e corpora digitali: qualche considerazione dalla parte dell'OVI. *Zeitschrift für romanische Philologie* 132(3): 755-792.
- Guadagnini, E. 2020. Alessandro, la *cautela* e altri latinismi: un esercizio traduttologico su un estratto della tradizione latino-romanza del *Secretum secretorum* (SS/B). *Francigena* 6: 243-278.
- Guadagnini, E. ics. Una breve storia del 'cadavere': caduti latini, corpi morti romanzi e una postilla dantesca. In C. Burgassi, E. Guadagnini, G. Vaccaro (eds.), *Storie di idee nell'Europa mediterranea*. CNR-ISEM Edizioni.
- Kabatek, J. 2013. ¿Es posible una lingüística histórica basada en un corpus representativo? *Iberoromania* 77: 8-28.
- Stussi, A. 1992. Testi in volgare veronese del Duecento. *Italianistica* 21(2/3): 247-267.
- TLIO* = Squillaciotti, P. (ed.). *Tesoro della lingua italiana delle origini*. <<http://tlio.ovi.cnr.it/TLIO/>>.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam/Philadelphia: Benjamins.
- Tomasin, L. 2019. *Il caos e l'ordine. Le lingue romanze nella storia della cultura europea*. Torino: Einaudi.