

NOMI FEMMINILI NEI *CORPORA ONLINE* DI ANTICHI DOCUMENTI VOLGARI ITALIANI DEL MEDIOEVO

Parole chiave: TLIO, DATINI, corpora antroponimici, nomi medievali, varianti

Esistono ormai in rete molti strumenti che consentono l'accesso a documentazione volgare di epoca medievale e la ricerca si orienta soprattutto verso aspetti di tipo lessicale e linguistico. In particolare mi sono occupata di verificare quale può essere l'utilizzazione di queste fonti per una ricerca antroponimica, esaminando l'accessibilità dei diversi strumenti, il tipo di risultati ottenibili e inoltre la possibilità di estrarre dati e materiali mirati distinguendo le tipologie onomastiche: toponimi e antroponimi (maschili e femminili, primo nome, secondo nome, soprannome). Le mie ricognizioni hanno per il momento preso in considerazione due grandi *corpora* di italiano medievale reperibili in rete: TLIO (Tesoro della Lingua Italiana delle Origini) e DATINI (fondo di lettere di Francesco Datini). In particolare la mia ricerca si è focalizzata sui nomi femminili.

TLIO

Il Tesoro della Lingua Italiana delle Origini¹ è un vocabolario storico dell'italiano antico, elaborato dai ricercatori e collaboratori dell'Opera del Vocabolario Italiano (OVI), Istituto del Consiglio Nazionale delle Ricerche con sede a Firenze, presso l'Accademia della Crusca.

Il *corpus* TLIO, la cui pubblicazione è cominciata nell'ottobre 1997², contiene (al 14 luglio 2016, data dell'ultimo aggiornamento, sempre riportato sulla maschera iniziale) 2.117 testi per complessive 21.920.635 occorrenze e 445.750 forme

¹ Consultabile all'indirizzo online <http://www.oivi.cnr.it> (oppure <http://www.vocabolario.org>, ultimo accesso il 30/10/2016), il TLIO copre l'arco temporale compreso tra le origini (il testo più antico è l'*Indovinello veronese*) e il 1375 (anno della morte di Boccaccio); di fatto sono compresi nel *corpus* anche testi che superano tale limite cronologico.

² Mosti 2010: 1–2.

grafiche distinte. È reso disponibile in rete per le ricerche linguistiche e consente di scaricare brevi citazioni per uso di ricerca.

Le notizie contenute nell'introduzione al TLIO³ ci informano che la lemmatizzazione è sostanzialmente completa per quanto riguarda le forme grafiche, con un'esigua rimanenza di forme rimaste ancora non interpretate. Perciò la ricerca di un lemma consente di acquisire l'elenco completo delle forme del lemma stesso presenti nel *corpus*.

Ad oggi sono lemmatizzate 3.674.650 occorrenze⁴ (dati riferiti all'aggiornamento di luglio 2016), rinvenibili direttamente grazie alla ricerca per lemmi.

Le altre occorrenze delle forme trovate con la ricerca per lemmi sono reperibili selezionando anche l'opzione non lemmatizzati relativa al tipo di contesti. L'introduzione ribadisce, inoltre, che la lemmatizzazione del *corpus* dell'italiano antico è resa complessa dalla difficoltà che richiede l'interpretazione di numerosissimi contesti in molte varietà linguistiche diverse, in linguaggi diversi (poetico, giuridico, botanico, medico ecc.), e in testi editi con i più diversi criteri filologici, solo per una parte dei quali esistono strumenti di interpretazione (commenti, glossari ecc.) di qualità e affidabilità molto varia. Questo lavoro è stato iniziato fin dalle origini dell'Opera del Vocabolario, ed è passato attraverso metodi linguistici e strumenti informatici diversi e attraverso l'opera di molte persone. Per queste ragioni è possibile ancora incontrare nelle lemmatizzazioni presenti nel *corpus* incoerenze e anche veri e propri errori⁵, alla correzione dei quali si sta lavorando mentre si procede nel contempo ad ampliare il numero delle occorrenze lemmatizzate.

Il *software* lessicografico utilizzato dall'Istituto del CNR per elaborare e gestire TLIO è lo stesso usato per gli altri *corpora* di italiano antico: si chiama GATTO⁶, acronimo di Gestione degli Archivi Testuali del Tesoro delle Origini, ed è stato ideato da Domenico Iorio Fili presso l'Opera del Vocabolario Italiano. La versione web, denominata GATTOWEB, è interrogabile dall'ottobre 2005⁷.

Per estrarre dal *software* GATTO i lemmi interessanti per la ricerca antroponica occorre impostare la categoria grammaticale *n.p.* (nomi propri) e il disambiguatore *antr.* (antroponimi, presenti in numero di 37810⁸ record, inclusi i casi

³ Reperibili navigando oltre la maschera introduttiva del TLIO (online: <http://tlioweb.ovi.cnr.it/HelpCorpora/Infotlio.html>; ultimo accesso il 30/10/2016). L'attuale responsabile della lemmatizzazione è Elena Artale.

⁴ Ibidem.

⁵ Ibidem.

⁶ Mosti 2010: 20.

⁷ Indirizzo diretto (online: <http://gattoweb.ovi.cnr.it>, ultimo accesso il 30/10/2016), ma vi si accede anche dall'*home page* dell'OVI, cliccando il link "Interroga le banche dati" e poi l'icona relativa alla banca dati TLIO). Tale programma può essere usato utilizzando i principali programmi per la navigazione in rete.

⁸ I numeri naturalmente sono in continuo aumento, dato il costante aggiornamento del *corpus*.

dubbi⁹), mentre per estrarre i toponimi è necessario fare uso del disambiguatore *n.g.* (nomi geografici, in numero di 21023¹⁰).

Vi sono poi casi in cui la colonna del disambiguatore non riporta alcuna indicazione, ad esempio in corrispondenza di nomi quali *Durindana*, *Fulberta*, *Gioiosa*, *Attaclera*, nomi propri delle spade dei cavalieri protagonisti dei poemi cavallereschi, cui, dunque, non è stata assegnata un'etichetta.

Per quanto riguarda gli antroponimi, non è possibile interrogare il sistema a livello più profondo, arrivando a distinguere primi, secondi nomi e nomi aggiunti. Per giungere a una disambiguazione certa in tal senso occorre cercare, tramite l'opzione "mostra contesti", uno per uno tutti i contesti testuali da cui sono estratti i nomi in questione.

Inoltre, provando ad eseguire tale operazione, appare subito evidente l'impossibilità di distinguere, oltre che tra primi, secondi nomi e nomi aggiunti, anche tra individui reali e nomi letterari, a meno di non vagliare ogni nome in ciascuna sua occorrenza tramite l'opzione "mostra contesti". Prendendo ad esempio un nome che può indifferentemente riferirsi a individui reali come a personaggi letterari, si può analizzare il caso di *Elisabetta* (Fig.1):

The screenshot shows the search results for 'Elisabetta' in the Corpus TLIO. The interface includes a search bar at the top with the text 'Corpus TLIO: risultati della ricerca -'. Below the search bar, there are several tabs: 'Contesti kwic', 'Ordinamento..', 'Selezione', 'Annulla selezione', 'Salva', 'Grafica', 'Vai a..', 'Riavvia GattoWeb', and 'Guide..'. The main content area displays a list of search results, each with a checkbox, a document reference, and a snippet of text. The results are numbered 1 through 9. The snippets show various occurrences of the name 'Elisabetta' in medieval Italian texts, such as '1 # Doc. sen., 1277-82 534.1 TS' and '2 Quad. F. de' Cavalcanti, 1290-1324 (fior.) [1300] 15.1 TS'. The interface also features a sidebar on the left with various filters and options, including 'occorrenze', 'trovate 69', 'pag. corrente 1', 'selez. 0', 'ricerca completata', 'scegli un'altra pagina', 'raffinamento', 'parole nei contesti', 'presenti', 'assenti', 'entro 10 parole', 'ricerca espansa', 'sintagmi', 'lunghezza', 'interpunzione forte', 'limita', 'posizione nel sintagma', 'prima', 'ultima', 'mostra', and 'in posizione'.

Fig. 1

⁹ Vi sono un certo numero di antroponimi contrassegnati dal sistema con il disambiguatore *antr.?*, casi ritenuti dubbi per diverse motivazioni (essenzialmente ambiguità o inconsuetudine), ad esempio *Acquaviva* e *Scolpuzio*.

¹⁰ Vd. nota precedente.

Per le occorrenze dalla 3 alla 9 il contesto di riferimento aiuta a ricostruire che il personaggio citato è la cugina della Vergine Maria e quindi ininfluenza dal punto di vista di uno studio onomastico; per le prime due occorrenze, è necessario cliccare sulla rubrica (in blu) per aprire la scheda relativa all'opera di riferimento ed ipotizzare verosimilmente la presenza di individui reali:

autore	Torrigiano di Firenze (Mastro)
titolo	Rime /due sonetti/
titolo abbreviato	Torrigiano, XIII sm. (fior.)
edizione	Poeti del Duecento, a cura di Gianfranco Contini, Milano-Napoli, Ricciardi, 1960, t. I, pp. 439-40.
curatore	Contini, G.
area generica	tosc.
area specifica	fior.
indice di qualità	
genere	lir.
tipo	O
forma	V
data descrittiva	sec. XIII s.m.
collocazione interna	E/FFN
data inserimento	1998-04-03 00:00:00
sigla	ffn
note	Anche in CLPIO, V ToFi 489 e 490
occorrenze	199
forme diverse	133
lemmi diversi	13
occorrenze lemmatizzate	16

Fig. 2

autore	=
titolo	Libro dell'entrata e dell'uscita di una Compagnia mercantile senese del secolo XIII
titolo abbreviato	Doc. sen., 1277-82
edizione	Libro dell'entrata e dell'uscita di una Compagnia mercantile senese del secolo XIII, a cura di Guido Astuti, Torino, Lattes, 1934.
curatore	Astuti, G.
area generica	tosc.
area specifica	sen.
indice di qualità	TS
genere	doc. merc.
tipo	O
forma	P
data descrittiva	1277-82
collocazione interna	Cit/1. Libro.3
data inserimento	1997-04-21 00:00:00
sigla	mk
note	
occorrenze	170921
forme diverse	4261
lemmi diversi	2382
occorrenze lemmatizzate	43086

Fig. 3

Solo nel secondo caso (Fig. 3), giacché ci troviamo di fronte a una fonte documentaria (*Il libro dell'entrata e dell'uscita di una compagnia mercantile senese del secolo XIII*), abbiamo qualche elemento in più per ricostruire che possa trattarsi di un personaggio reale.

Ne consegue, dunque, che anche ricostruire il numero degli individui è difficile.

È possibile in alcuni casi uno studio di varianti e tuttavia non vi è una lemmatizzazione efficace che consenta una visione d'insieme di tali varianti: prendendo ad esempio il caso di un nome evidentemente solo letterario, *Pentesilea*, per trovare tutte le occorrenze relative al nome è necessario cercare separatamente le due forme *Pantasilea* (Fig. 4) e *Pentesilea* (Fig. 5):

The screenshot shows the search results for 'Pantasilea' in the Corpus TLIO. The interface includes a search bar, a menu with options like 'Contesti kwic', 'Ordinamento..', 'Selezione', 'Annulla selezione', 'Salva', and 'Grafica', and a sidebar on the left with statistics: 'trovate 99 in 1 pag. pag. corrente 1 selez. 0 ricerca completata'. The main area displays three results:

Result ID	Text Snippet	Count	TS	Other
1	St. de Troia e de Roma Amb., 1252/58 (rom.>tosc.) sse guardi forte la citade de Troia e repositose pro fi a ttanto ke regina Pantasilea pertanto Priamus per aliquanti dii	50.23	TS	@
2	St. de Troia e de Roma Amb., 1252/58 (rom.>tosc.) patre fece pianto grandissimo sopra la soa sepultura. E da poi ke venne la regina Pi tucti li greci. Ma Nectolinus filio de	51.21	TS	@
3	St. de Troia e de Roma Amb., 1252/58 (rom.>tosc.) c. anni fi ad Iulium Cesarem. A la storia retornemo de Troia. Morta Pantasilea li tro nullo homo non	55.25	TS	@

Fig. 4

The screenshot shows the search results for 'Pentesilea' in the Corpus TLIO. The interface is similar to Fig. 4, but with a 'raffinemento' button at the bottom left. The sidebar statistics are: 'trovate 18 in 1 pag. pag. corrente 1 selez. 0 ricerca completata'. The main area displays four results:

Result ID	Text Snippet	Count	TS	Other
1	St. de Troia e de Roma Laur, 1252/58 (rom.>tosc.) li Troiani e Priamus commannaio ke guardassero forte la citate, et repositete fi tanto ke Pentesilea , hoste 'n aiuto ad noi. Ma pertanto Priamus per aliquanti	50.23	TS	@
2	St. de Troia e de Roma Laur, 1252/58 (rom.>tosc.) patre, gio e fece granne pianto sopra lo monumento de Acille. Ma pertanto la regina Pentesilea ven li Greci in fuga. Ma Pirrus, filio	51.22	TS	@
3	St. de Troia e de Roma Laur, 1252/58 (rom.>tosc.) mise tutti li Greci in fuga. Ma Pirrus, filio de Acilles, occise molte femine e Pentesilea lo trovao e fer non remase de commattere, fi ke	51.25	TS	@
4	St. de Troia e de Roma Laur, 1252/58 (rom.>tosc.) e ferliolo forte. Et Pirrus, sentenose feruto, non remase de commattere, fi ke non occise Pentesilea	51.28	TS	@

Fig. 5

A questo punto si può arrivare a dire che le 99 attestazioni di *Pantasilea* fanno di questa forma la maggioritaria rispetto a *Pentesilea* (18), anche se naturalmente, trattandosi di un personaggio letterario, non è possibile uno studio di tipo socio-onomastico.

In conclusione, il *corpus* TLIO non presenta ancora le caratteristiche per essere una fonte di utilizzo agevole per lo studio antroponomico, soprattutto per l'assenza di mappe distintive e di lemmatizzazioni efficaci.

DATINI

DATINI è il *corpus* lemmatizzato delle lettere e dei documenti editi dell'Archivio DATINI, fondo conservato nell'Archivio di Stato di Prato, fonte di grandissimo interesse per la ricostruzione della storia del mondo mercantile europeo nella seconda metà del Trecento. L'archivio DATINI contiene oltre 150.000 documenti sciolti e circa 600 registri¹¹. Nel 2003 l'Archivio di Stato di Prato affidava all'Opera del Vocabolario Italiano l'allestimento di una banca dati informatizzata contenente le lettere e i documenti editi dell'archivio DATINI¹², con la richiesta di indicizzare i nomi di persona e di luogo e di lemmatizzare i termini appartenenti ad alcuni settori di lessico: quello storico-economico, commerciale (incluso il commercio marittimo), militare e storico-giuridico, nonché i tecnicismi e le date rilevanti. Una seconda richiesta verteva sulla possibilità di raggruppare i lemmi prodotti in base ai rispettivi ambiti tematici di afferenza, e ottenere un livello di ricerca che guidasse l'utente nell'interrogazione del *corpus*. È nato così il *corpus lemmatizzato del carteggio Datini*, allestito con lo stesso programma che gestisce il *corpus* TLIO dell'OVI, il Gatto, in una versione appositamente dedicata¹³.

Il *corpus* consta di 2.511 testi e 45.259 forme, per un totale di 977.034 occorrenze, di cui 126.663 lemmatizzate, con 6.510 lemmi e 22 iperlemmi (l'iperlemma è ciascuna delle 22 categorie entro le quali sono stati distribuiti i lemmi del *corpus* ad esempio abbigliamento, arredi, alimenti, etc., ma sono solo lemmatizzati e non iperlemmatizzati antroponomi, toponimi, termini e verbi afferenti al campo religioso o all'agricoltura, le parti del corpo, i nomi della settimana e i termini generici).

L'intera documentazione del *corpus* abbraccia un intervallo temporale che va dal 1367 al 1432¹⁴ e contiene, oltre alle lettere e ai documenti commerciali delle varie filiali della compagnia Datini sparse per la penisola e fuori dall'Italia (ad Avignone e nell'allora Regno di Aragona), il testamento di Francesco Datini e gli scambi epistolari in cui convivono la dimensione pubblica e i rapporti privati tra gli scriventi: il carteggio di Francesco con la moglie Margherita Bandini, quello con l'amico e notaio Lapo Mazzei e quello con il medico Naddino.

Significative sono la vastità geografica rappresentata dal *corpus* datiniano, nonché il notevole assortimento linguistico, ricco di varietà italo-romanze e non: 2408 testi

¹¹ Informazioni tratte da <http://datini.archiviodistato.prato.it/> (ultimo accesso il 30/10/2016).

¹² Artale 2006.

¹³ Ibidem.

¹⁴ Ibidem: 2.

rappresentano il toscano (nelle sue varianti e nelle peculiari realizzazioni di scriventi non toscani o da lungo tempo residenti fuori dalla Toscana), 38 il marchigiano, 14 il siciliano, 8 il campano, 2 il veneto, un testo rispettivamente il calabrese, il laziale e il ligure; a questi vanno aggiunti 28 testi in latino, 9 in catalano e uno in tedesco¹⁵.

La marcatura dei lemmi è stata effettuata utilizzando le seguenti categorie¹⁶:

- agg. — aggettivo
- antr. — antroponimo
- avv. — avverbio
- n.g. — nome geografico (nome di luogo)
- n.p. — nome proprio
- s.f. (eventualmente s.f.pl.) — sostantivo femminile (sostantivo femminile plurale)
- s.i. — sostantivo di genere incerto
- s.m. (eventualmente s.m.pl.) — sostantivo maschile (sostantivo maschile plurale)
- v. — verbo

In DATINI, una volta selezionata la categoria *n.p.*, è possibile estrarre tutte le voci che sono state lemmatizzate con categoria lessicale *antr.*: tutti i nomi di persona, eventuali soprannomi, e cariche specifiche nel caso in cui l'indicazione rimandi ad un preciso personaggio storico¹⁷.

Data la grande variazione grafica nei nomi propri, sono state effettuate alcune normalizzazioni grafiche, di cui l'introduzione al *corpus* fornisce qualche esempio:

1) le forme *niccolaio*, *niccolò*, *nicolò*, sono state ricondotte al lemma *niccolò*; le forme *niccola*, *nicola*, sono state ricondotte al lemma *nicola*;

2) l'alternanza *buo-*/*bonsignore* è stata ricondotta al lemma senza dittongo *bonsignore*;

3) le forme del tipo *gabriello* sono state ricondotte al lemma *gabriele*;

4) le forme del tipo *ghirig-* / *greg-* / *grigoro* (*-orio*) sono state ricondotte al lemma *gregorio*;

5) le forme del tipo *guasparre(-o)* sono state ricondotte a *gaspare*;

6) le forme del tipo *ludovic(-a)* sono state ricondotte a *lodovico(-a)*;

7) nei lemmi è stata sempre adoperata la grafia *iacopo* (e non *jacopo*).

L'introduzione al *corpus* fornisce anche interessanti informazioni in merito al trattamento degli antroponimi femminili, che, si dice, sono stati trattati, quando

¹⁵ Ibidem.

¹⁶ Le informazioni che seguono sono tratte dalle istruzioni *infocorpus*, reperibili navigando oltre la maschera introduttiva di DATINI, all'indirizzo ([http://aspweb.ovi.cnr.it/\(S\(mwqipyaavxrpipynkblbatfzj\)\)/HelpCorpora/Infoasp.html](http://aspweb.ovi.cnr.it/(S(mwqipyaavxrpipynkblbatfzj))/HelpCorpora/Infoasp.html); ultimo accesso il 30/10/2016).

¹⁷ L'iniziale minuscola anche nei nomi propri è dovuta al *software* GATTO (che non distingue tra maiuscole e minuscole). Per aderenza al *software*, nelle citazioni dirette dal *corpus* ho mantenuto l'iniziale minuscola.

possibile, come i maschili (es. *lisa di ugototto degli agli*); solo in mancanza del cognome e/o del patronimico sono state date indicazioni del tipo donna di..., vedova di..., madre di..., serve di... o il semplice monna (o madonna) — riportando il titolo dal documento — posposti al nome, tra parentesi (es. *diana, donna di iacopo zarini*). Se è noto il nome della madre si è utilizzato un lemma del tipo *sandra (figlia di leonarda panciaticchi)*.

Nel caso di *ginevra di francesco datini*, si è preferito lemmatizzare *ginevra (figlia naturale di francesco datini)* piuttosto che *ginevra di francesco datini*.

La categoria lessicale *antr.* è stata usata anche per i nomi di famiglia, segnalati come tali nel disambiguatore.

Sono stati lemmatizzati come *n.g. (nome geografico)* i nomi di città, paesi, contrade, località, vie, piazze, porte; sono stati lemmatizzati come *n.g.* anche specifici edifici, come chiese, monasteri, palazzi, ospedali.

Vi sono, inoltre, specifici disambiguatori appositamente studiati per marcare i nomi di nave (*nome di nave*) e i nomi di ordini religiosi (*ordine religioso*).

Le forme antroponimiche estratte selezionando la categoria grammaticale *n.p.* e filtrate dalla categoria lessicale *antr.* sono in tutto 6.139; i toponimi sono 60, i nomi di nave 34 e i nomi di ordini religiosi 14.

Gli antroponimi sono stati raccolti per stringa onomastica dell'individuo (*diana donna di iacopo zarini*) e si intuisce, dunque, la necessità di lemmatizzare sulla forma normalizzata, dopo averla estratta dalla stringa onomastica.

NOMI FEMMINILI IN DATINI

I primi nomi femminili, estratti da DATINI vagliando gli antroponimi registrati, sono in tutto 245, inclusi i casi dubbi. Questi 245 antroponimi sono riconducibili a 79 forme normalizzate (il numero può essere soggetto, come vedremo, a lievi oscillazioni per via dei criteri di lemmatizzazione utilizzati), a cui sono collegate le numerose varianti; poiché ciascun nome può essere riferito più volte alla stessa persona va precisato anche il numero degli individui, che ammonta a 149, inclusi i casi dubbi.

Tra i vantaggi vi è senz'altro il fatto che il *corpus* sia reale: gli individui citati sono persone realmente esistite e quindi databili. Inoltre, sono significative la visibilità sugli identificatori sociali (*madonna* o *serva*, ad esempio) e la ricostruzione dell'identità dei personaggi, nonché l'ampia variabilità grafico-fonetica, che permette di studiare il *corpus* dal punto di vista linguistico.

Tra le criticità si evidenzia il fatto che il sistema per ricavare i nomi non sia immediato e numerose operazioni siano necessarie per rendere il *corpus* utilizzabile dal punto di vista onomastico: prevede innanzitutto delle marcature di genere e lo scorporo dei dati della stringa onomastica, che risulta aggregata (Fig. 6):

lemma	cat less	forma	disambiguatore
francesca (vedova di ticcio)	antr.	franciescha	
francesca di domenico bandini	antr.	fancescha	
francesca di domenico bandini	antr.	fraciescha	
francesca di domenico bandini	antr.	francesca	
francesca di domenico bandini	antr.	francescha	
francesca di domenico bandini	antr.	franciescha	
francesca di francesco	antr.	franciescha di francescho	
francesca di giusto (monna)	antr.	franciescha di gusto	
gaia (donna di piero di giunta del rosso)	antr.	gaia	
gaia (donna di piero di giunta del rosso)	antr.	ghaia	
gaia (monna)	antr.	ghaia	donna di piero di giunta del r
gemma	antr.	gemma	monna gemma?
gemma (monna)	antr.	giema	
ghita (donna di matteo ghetto)	antr.	ghita	
ghita (donna di matteo ghetto)	antr.	gita	
ghita di geri	antr.	gita di gieri	
gianna (madre di baldello d'a. bovattieri)	antr.	giana	
ginevra (figlia naturale di francesco datini)	antr.	g. di francesco	
ginevra (figlia naturale di francesco datini)	antr.	ginevera	
ginevra (figlia naturale di francesco datini)	antr.	ginevra	
ginevra (figlia naturale di francesco datini)	antr.	zinevra	
giovanna (monna)	antr.	giovana	
giovanna (monna)	antr.	giovanna	
giovanna (serva di margherita di domenico bandini)	antr.	giovana	

Fig. 6

Solo la ricostruzione di un database mirato può, dunque, rendere funzionale e agevole la consultazione del *corpus*.

Vi sono forme che si presentano in un'unica occorrenza: è indicata la variante attestata, qualora non coincida con quella base di lemmatizzazione:

Tabella 1

Nomi a una occorrenza		
ID	Nome	Forma attestata
1.	<i>Anna</i>	
2.	<i>Bellozza</i>	[<i>Beloza</i>] ¹⁸
3.	<i>Benedetta</i>	
4.	<i>Benvenuta</i>	
5.	<i>Bionda</i>	
6.	<i>Boneca</i>	
7.	<i>Colomba</i>	<i>Cholomba</i>
8.	<i>Dialta</i>	
9.	<i>Dolce</i>	
10.	<i>Dora</i>	

¹⁸ Nome non ricavabile dall'elenco.

Nomi a una occorrenza		
ID	Nome	Forma attestata
11.	<i>Eva</i>	
12.	<i>Farillera</i>	
13.	<i>Fattorina</i>	
14.	<i>Filippa</i>	<i>Filipa</i>
15.	<i>Gianna</i>	<i>Giana</i>
16.	<i>Gostanza</i>	<i>Gostança</i>
17.	<i>Gramigna</i>	<i>Ghramignia</i>
18.	<i>Jacopa</i>	
19.	<i>Lodola</i>	
20.	<i>Lorenza</i>	
21.	<i>Lorita</i>	
22.	<i>Luca</i>	<i>Lucha</i>
23.	<i>Maddalena</i>	<i>Madalena</i>
24.	<i>Masa</i>	
25.	<i>Mattea</i>	
26.	<i>Mea</i>	
27.	<i>Mina</i>	
28.	<i>Orsetta</i>	<i>Orseta</i>
29.	<i>Orta</i>	
30.	<i>Papera</i>	
31.	<i>Petrazza</i>	<i>Petraza</i>
32.	<i>Picchina</i>	
33.	<i>Sandra</i>	
34.	<i>Severina</i>	
35.	<i>Sibilla</i> ¹⁹	<i>Reina</i>
36.	<i>Spina</i>	
37.	<i>Tessa</i>	
38.	<i>Tonina</i>	
39.	<i>Vivola</i>	
40.	<i>Zita</i>	

¹⁹ Nome presente in elenco ma non attestato.

Vi sono due nomi che mettono in luce ulteriori criticità del sistema di lemmatizzazione: infatti, uno dei nomi presenti nell'elenco estratto dal *corpus*, Sibilla, non è realmente attestato, poiché la forma presente nel testo non è *Sibilla*, ma *Reina*: l'appellativo è riferito a Sibilla de Fortià, regina d'Aragona, e ciò ha spinto i compilatori a lemmatizzare (arbitrariamente) sul primo nome del personaggio, la cui identità è ricostruita con certezza, pur, dunque, in assenza del nome. Opposto il caso di Bellozza (presente nella variante *Beloza*), nome assente nell'elenco estraibile dal *corpus* e da me aggiunto: questo nome è stato ricavato dalle varianti di Bartolomea (lemmatizzata, infatti, come *Bartolomea detta Bellozza*); anche qui si ravvisa, quindi, un'arbitrarietà nel lemmatizzare sulla base di un nome in realtà assente dal testo. Casi analoghi si vedranno anche tra i nomi che contano più di un'attestazione.

Propongo ora un elenco dei nomi più attestati per frequenza decrescente, con riferimento al tipo onomastico; accanto è riportato il numero di individui nelle diverse forme e le relative attestazioni:

Tabella 2

ID	Nome	Individui per <i>tipo</i> onomastico	Numero individui per <i>forma</i>	Numero attestazioni per <i>forma</i>	Note
1.	<i>Caterina, Tina</i>	13	11; 2	16; 4	
2.	<i>Giovanna, Nanna, Vanna</i>	12	6, 5, 1	14; 9; 2	
3.	<i>Margherita Ghita, Chita</i>	8	5; 2; 1	19; 3; 3	Lemmatizzati sotto Margherita anche Madama e <i>Reina</i>
4.	<i>Bartolomea</i>	8	7; 1	10; 3	Lemmatizzato sotto <i>Bartolomea</i> anche <i>Bellozza</i>
5.	<i>Francesca, Checca</i>	7	6; 1?	12; 2	Non ci sono specifiche su <i>Checca</i>
6.	<i>Antonia</i>	5	5	8	
7.	<i>Beatrice, Bice</i>	4	2; 2	2; 2	
8.	<i>Piera</i>	4	4	5	
9.	<i>Agnola</i>	3	3	5	
10.	<i>Lisa</i>	3	3	4	
11.	<i>Andrea</i>	3	3	3	
12.	<i>Lapa</i>	3	3	3	
13.	<i>Chiara</i>	2	2	3	Lemmatizzata sotto <i>Chiara</i> anche <i>Monaca</i>
14.	<i>Leonarda</i>	2	2	3	

ID	Nome	Individui per <i>tipo</i> onomastico	Numero individui per <i>forma</i>	Numero attestazioni per <i>forma</i>	Note
15.	<i>Lucia</i>	2	2	3	
16.	<i>Niccolosa</i>	2	2	3	
17.	<i>Diana</i>	2	2	2	
18.	<i>Domenica, Domenichina</i>	2	4; 1	5	
19.	<i>Fiore</i>	2	2	2	
20.	<i>Lena</i>	2	2	2	
21.	<i>Rosa</i>	2	2	2	
22.	<i>Giuliva</i>	1	1	5	
23.	<i>Chiaromonda</i>	1	1	4	
24.	<i>Cilia</i>	1	1	4	
25.	<i>Ginevra</i>	1	1	4	
26.	<i>Gregoria</i>	1	1	4	
27.	<i>Dianora</i>	1	1	3	
28.	<i>Gaia</i>	1	1	3	
29.	<i>Isabetta</i>	1	1	3	
30.	<i>Regina</i>	1	1	3	Lemmatizzata sotto <i>Regina</i> anche <i>Madonna</i>
31.	<i>Tinga</i>	1	1?	3	
32.	<i>Beldì</i>	1	1	2	
33.	<i>Gemma</i>	1	1	2	
34.	<i>Paola</i>	1	1	2	Lemmatizzata sotto <i>Paola</i> anche <i>Coverciana</i>
35.	<i>Parta</i>	1	1	2	
36.	<i>Salvestra</i>	1	1	2	
37.	<i>Simona</i>	1	1	2	
38.	<i>Taddea</i>	1	1	2	
39.	<i>Villana</i>	1	1	2	

Un caso interessante riguarda *Margherita*: come nel caso precedentemente citato di *Sibilla*, anche qui troviamo due appellativi (*Reina* e *Madama*) lemmatizzati come *Margherita* in quanto riferiti a Margherita di Durazzo (attestata con il suo nome una sola volta).

Simile il caso relativo a *Chiara*: due delle tre attestazioni presenti sono ricondotte al personaggio di Chiara Gambacorti, ma in uno solo dei casi la donna è citata come Chiara, infatti la seconda attestazione è in realtà l'appellativo *Monaca*, chiaro riferimento alla sua condizione di monacazione; anche qui *Monaca* è stato lemmatizzato come *Chiara*.

Degno di nota anche il caso di *Regina*, presente in tre attestazioni (tutte riferite a Regina della Scala) lemmatizzate sotto Regina sebbene una delle occorrenze, ancorché riferita alla sua persona, sia l'appellativo *Madonna*.

Si può citare ancora il caso di *Paola*, in due attestazioni riferite al medesimo personaggio (citata in elenco come *paola detta la coverciana*), è attestata una volta come *Paola* (variante *Paghola*) e un'altra come *Coverciana*, elemento di chiaro carattere soprannominale. Occorre dunque vagliare una a una le forme effettivamente attestate per indagarne la significatività onomastica, poiché, come si è dimostrato, spesso l'etichettatura utilizzata in sede di lemmatizzazione rischia di celare forme che non sono semplici varianti grafiche o fonetiche e ciò può essere fuorviante per il linguista.

Tina è ipocoristico di *Caterina*, come conferma la documentazione storica; interessante per gli alterati *Tinuca* e *Tinuccia*, correttamente lemmatizzati sotto *Tina*.

Tabella 3

Elemento 1	Lemma	Cat less	Forma
<i>Tina</i>	Tina	antr.	<i>tina</i>
<i>Tina</i>	tina di niccolò dell'ammannato tecchini	antr.	<i>tina</i>
<i>Tina</i>	tina di niccolò dell'ammannato tecchini	antr.	<i>tinuca</i>
<i>Tina</i>	tina di niccolò dell'ammannato tecchini	antr.	<i>tinuccia</i>

Un procedimento evidentemente non omogeneo, quello di lemmatizzare gli alterati sotto la forma base: infatti nel caso di *Domenica*, l'alterato *Domenichina* (nella variante *Domeneghina*) è lemmatizzato a parte (e nel mio elenco è invece contata tra le varianti di *Domenica*, per omogeneità).

Tabella 4

Elemento 1	Lemma	Cat less	Forma
<i>Domenica</i>	domenica (donna di meo di simone)	antr.	<i>domenica</i>
<i>Domenica</i>	domenica (donna di meo di simone)	antr.	<i>domenicha</i>
<i>Domenica</i>	domenica (donna di meo di simone)	antr.	<i>domeniche</i>
<i>Domenica</i>	domenica (donna di meo di simone)	antr.	<i>domina</i>
<i>Domenichina</i>	domenichina (donna di piero di giusto benintendi)	antr.	<i>domeneghina</i>

A parte andrebbe considerata la forma *Domina* per *Domenica*: potrebbe trattarsi di un soprannome, bisognerebbe indagare consultando il testo originale nel suo contesto.

Complessa è la valutazione di *Tigna*, che potrebbe essere una forma soprannominale anche maschile²⁰, come testimonia Brattö, che cita *Tigna* come forma da confrontare con *Tignosus*, indicato come soprannome e presente in quattro attestazioni tra i nomi meno frequenti del *Libro di Montaperti*; è confrontabile, inoltre, con un personaggio maschile denominato *il Tigna*, presente nella commedia seicentesca *Il forca* di Niccolò Amenta²¹.

Tabella 5

Elemento 1	Lemma	Cat less	Forma
<i>Tigna</i>	tigna	antr.	<i>tingna</i>

Certamente femminile è la forma *Tinga* (come precisato dalla fonte). Non devono stupire le oscillazioni grafiche *-ng-*, *-ngn-* e *-gn-*, comuni nella tradizione manoscritta per rappresentare la *-n-* palatale:

Tabella 6

Elemento 1	Lemma	Cat less	Forma
<i>Tinga</i>	tinga (monna)	antr.	<i>tinga</i>
<i>Tinga</i>	tinga (monna)	antr.	<i>tingha</i>

I casi dubbi sono tali a causa dell'assenza di ulteriori informazioni circa gli individui citati: ad esempio il due volte attestato *Checca* non ha riferimenti, per cui le due attestazioni potrebbero riferirsi alla stessa persona come a due differenti; tuttavia è maggiormente presumibile che si tratti di uno solo, giacché in presenza di tratti di disambiguazione i compilatori del *corpus*, come hanno fatto altrove, avrebbero specificato la presenza di due personaggi.

Propongo ora un grafico relativo alle attestazioni dei nomi pluriattestati, per numero di individui (Fig. 7).

²⁰ Brattö 1955: 208.

²¹ Amenta 1753.

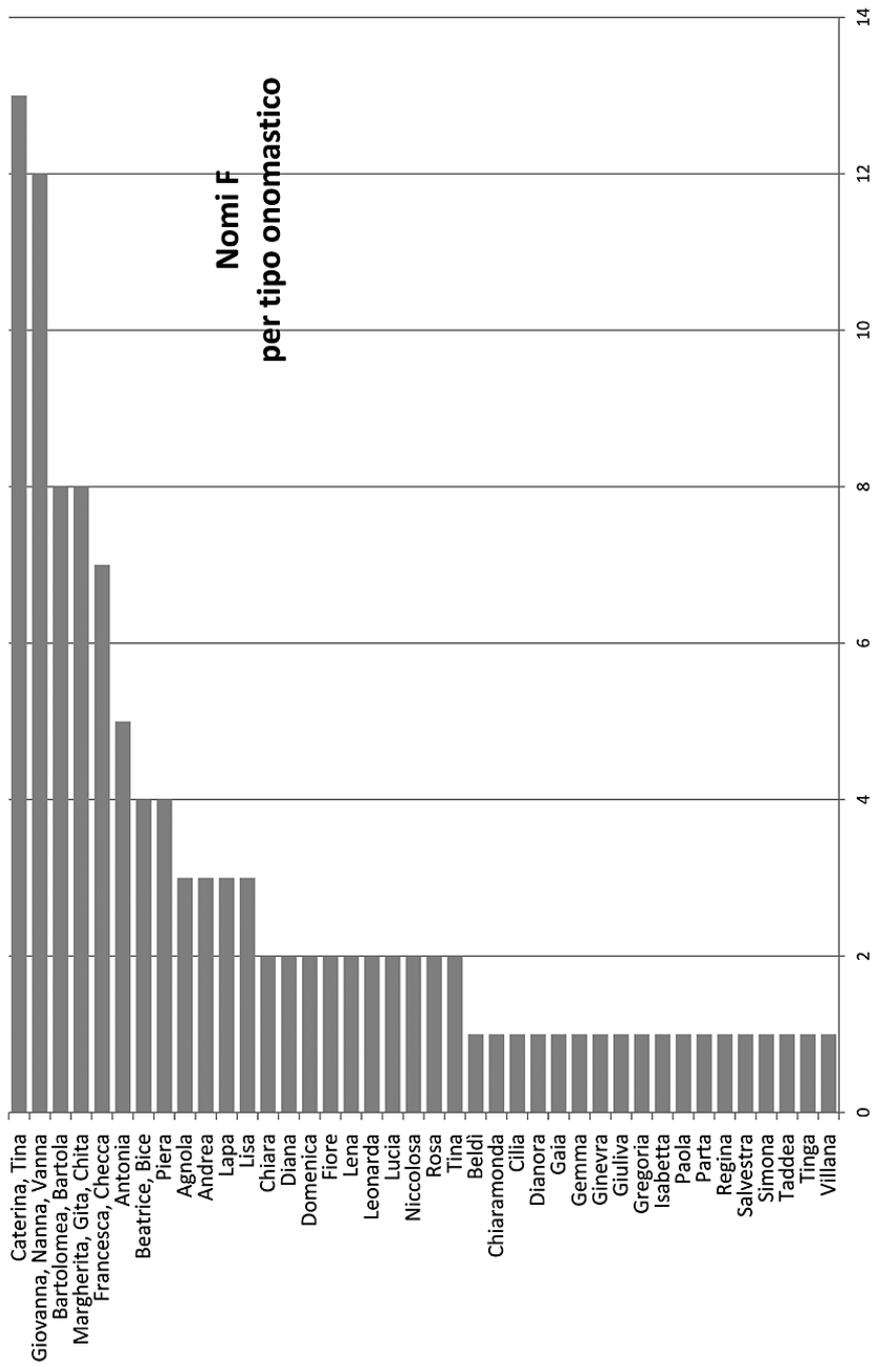


Fig. 7

I tipi onomastici più rappresentati sono:

Tabella 7

<i>Caterina</i> (13 individui totali, di cui 11 nella forma a lemma), sostenuto da <i>Tina</i> , ipocoristico ²²	13	11; 2
<i>Giovanna</i> (12 individui totali, di cui 6 nella forma a lemma), sostenuto dagli ipocoristici <i>Nanna</i> (5) e <i>Vanna</i> (1)	12	6; 5; 1

a seguire

<i>Margherita</i> (8 individui totali, di cui 5 nella forma a lemma), sostenuto dagli ipocoristici <i>Ghita</i> (2) e <i>Chita</i> (1)	8	5; 2; 1
<i>Bartolomea</i> , <i>Bartola</i>	8	7; 1
<i>Francesca</i> ; <i>Checca</i>	7	6; 1?
<i>Antonia</i>	5	5
<i>Beatrice</i> , <i>Bice</i>	4	2; 2

Questi dati permettono di ricostruire il quadro onomastico dell'epoca, associandolo a personaggi concreti, esponenti della borghesia mercantile, eventualmente utilizzabili per confronti con altri repertori femminili dello stesso periodo.

Osservazioni linguistiche

Propongo ora alcuni esempi di confronti e considerazioni linguistiche che è possibile effettuare a partire dalla grande ricchezza di varianti grafiche e fonetiche fornite dal *corpus*.

Possiamo confrontare, ad esempio, le diverse rese grafiche del fonema nasale palatale /ɲ/, ritrovando la consueta varietà che caratterizza la *scripta* toscana²³: a fianco a *gn* troviamo anche *gni* e *ngn*:

Tabella 8

Forma	Conteggio di forma
<i>agnola</i>	1
<i>agnola di lotto</i>	1
<i>angniola</i>	1
<i>angnola</i>	2
<i>ghramignia</i>	1
<i>tingna</i>	1

²² Cf. *Tecchini Caterina (Tina) di Niccolò dell'Ammannato*.

²³ Castellani 1991: 27–29.

Un'altra particolarità della *scripta* toscana è ravvisabile nella resa grafica dell'occlusiva velare sorda /k/, al cui proposito Castellani afferma, portando ad esempio il *Libro del dare e dell'avere* di Lapo Riccomanni (1281–1297), che l'autore tende a separare le occlusive velari dalle affricate palatali usando per le prime, di preferenza ma non unicamente, il digramma *ch*, sia dinanzi a *i* ed *e*, sia dinanzi ad *a*, *o*, *u*²⁴, come accade nei nomi del *corpus*:

Tabella 9

Forma	Conteggio di forma
<i>chara</i>	1
<i>chatarina</i>	3
<i>chaterina</i>	5
<i>chaterina d'andrea</i>	1
<i>chaterina d'aringho</i>	1
<i>chaterina di barnabò</i>	1
<i>chaterina di bernabò</i>	1
<i>chaterina di nicholozo</i>	1
<i>heccha</i>	1
<i>domenicha</i>	1
<i>fancescha</i>	1
<i>fraciescha</i>	1
<i>francescha</i>	3
<i>franciescha</i>	3
<i>franciescha di franciescho</i>	1
<i>franciescha di gusto</i>	1
<i>Lucha</i>	1

CONCLUSIONI

I *corpora* di testi medievali in italiano reperibili in rete possono effettivamente fornire materiale interessante per la ricerca antroponimica, ma non tutti sono adatti ad una ricerca puntuale: il *corpus* TLIO non permette di distinguere efficacemente individui reali da personaggi letterari, a causa dell'assenza di una lemmatizzazione omogenea e di mappe distintive.

DATINI, invece, grazie alla presenza di solo materiale documentario, permette l'estrazione dei nomi di soli individui reali, rendendo così efficace la ricerca antro-

²⁴ Ibidem: 30.

ponimica, pur con tutti i limiti del caso e le numerose operazioni di aggiustamento che lo studioso di onomastica deve mettere in atto per poter lavorare adeguatamente sui dati estratti.

BIBLIOGRAFIA

- Amenta N. 1753: *“Il forca”, commedia di Niccolò Amenta, avvocato napoletano*, Stamperia de Muzj, Napoli.
- Artale E. 2006: *Mercanti medievali in Internet: le lettere dell'archivio Datini in GattoWeb*, [in:] *Prospettive nello studio del lessico italiano. IX Congresso SILFI, Firenze, 14–17 giugno 2006*, Firenze University Press, Firenze.
- Brattö O. 1953: *Studi di antroponomia fiorentina. Il Libro di Montaperti (An. MCCLX)*, Elanders Boktryckeri Aktiebolag, Göteborg.
- Brattö O. 1955: *Nuovi studi di antroponomia fiorentina. I nomi meno frequenti del Libro di Montaperti (An. MCCLX)*, Almqvist & Wiksell, Stockholm.
- Castellani A. 1991: *Problèmes concernant la notion de scripta et problèmes concernant le rapport entre graphèmes et phonèmes*, [in:] *Actes Trèves 1991*, pp. 695–699 [rist. col titolo *Grafemi e fonemi: esempi medievali*, [in:] Id., *Nuovi saggi di linguistica e filologia italiana e romanza (1976–2004)*, a cura di V. Della Valle et al., Roma, Salerno Editrice, 2010].
- Mosti R. 2010: *Italiano antico e italiano moderno: notizie dal TLIO*, Giornata di studi del Consiglio Nazionale delle Ricerche “Lo stato della lingua. Il CNR e l'italiano nel terzo millennio”, Roma.

SITOGRAFIA

- Archivio di Stato di Prato, <http://datini.archiviodistato.prato.it/>
- Corpus TLIO, <http://tlioweb.ovi.cnr.it/HelpCorpora/Infotlio.html>
- Corpus DATINI, [http://aspweb.ovi.cnr.it/\(S\(mwqipavxrpiypynbklbatfzj\)\)/HelpCorpora/Infoasp.html](http://aspweb.ovi.cnr.it/(S(mwqipavxrpiypynbklbatfzj))/HelpCorpora/Infoasp.html)
- GATTOWEB, <http://gattoweb.ovi.cnr.it>
- OVI, <http://www.ovi.cnr.it>

FEMININE NAMES IN *ONLINE CORPORA* OF ITALIAN MEDIEVAL DOCUMENTS

SUMMARY

Corpora and network tools which provide access to Italian medieval documentation are nowadays accessible online. The current research is now oriented mainly towards lexical and linguistic aspects. My purpose is to investigate the use of these resources for an anthroponomical study, to verify the accessibility of the different network tools and, finally, to consider which kind of results can be obtained and also the possibility to extract data and materials, in order to distinguish onomastic types: placenames and anthroponyms (male and female names, first name, second name, nickname, etc.). My research is particularly focused on feminine names found in Corpus TLIO and Corpus DATINI.

Keywords: TLIO, DATINI, anthroponymic corpora, medieval names, variants