

Il trattamento automatico del linguaggio applicato all'italiano volgare. La redazione di un *formario* tratto dalle prime dieci *Lettere* di Alessandra M. Strozzi.

Ottavia Bersano
Università degli Studi di Firenze
Universität Bonn
ottavia.bersano@gmail.com

Nadezda Okinina
Eurac Research
Bolzano
nadezda.okinina@eurac.edu

Abstract

English. This paper aims to describe the strategies that have been experimented in the use of Natural Language Processing (NLP) on a Florentine text of the 15th century, that is the Epistolary of Alessandra Macinghi Strozzi. With the help of NLP, and in particular of POS tagging, it was possible not only to facilitate the linguistic analysis of the text by quickly obtaining the most interesting graphic, phonetic and morphological traits, but also to create, automatically, a collection of word forms, alphabetically ordered with the number of occurrences for each item.

Italiano. Questo contributo è teso a illustrare le strategie sperimentate nell'applicazione del Trattamento automatico del Linguaggio (TAL) a un testo di origine fiorentina risalente al sec. XV, l'Epistolario di Alessandra Macinghi Strozzi. Con l'ausilio del TAL, e in particolare con il processo di POS tagging, è stato possibile non solo agevolare l'analisi linguistica del testo ricavandone celermente i tratti grafici, fonetici e morfologici più interessanti, ma realizzarne anche, in modo automatico, un *formario*, ordinato su base alfabetica e comprensivo del numero di occorrenze per ciascun item.

1 Introduzione

Il presente contributo si propone di coniugare le tecniche offerte dal Trattamento automatico del Linguaggio (TAL) allo studio dell'italiano volgare, e nella fattispecie a testi di origine fiorentina del sec. XV. Esso nasce in seno ad alcune difficoltà emerse durante la selezione dei dati testuali di una tesi di dottorato – ancora in corso di stesura – tesa a fornire una nuova edizione dell'Epistolario di Alessandra Macinghi Strozzi (1406-1471): un documento assai conosciuto e apprezzato da chiunque si occupi della civiltà italiana del Quattrocento e che siamo ancora costretti a leggere nell'edizione curata da Cesare Guasti nel 1877 (Guasti, 1877). Una nuova edizione e una nuova analisi linguistico-interpretativa, condotta con gli strumenti più aggiornati, porterà indiscutibili vantaggi anche agli studi di ambito storico e letterario.

2 Le *Lettere* di Alessandra Macinghi Strozzi

Le *Lettere* – in tutto settantatré – furono scritte in un arco temporale compreso tra il 1447 e il 1470 e indirizzate, da Alessandra Macinghi – vedova di Matteo Strozzi – ai figli Filippo, Lorenzo e Matteo. Questi, raggiunta la maggiore età, ereditarono l'esilio paterno e furono costretti a lasciare Firenze: un provvedimento legislativo, infatti, stabiliva che tutti i figli maschi degli esuli, al compimento del diciottesimo anno, ne dovessero ereditare la condizione; lasciata dunque Firenze ed esercitando il mestiere della mercatura, Filippo, Lorenzo e Matteo viaggiarono per tutta Europa, trovandosi a soggiornare nei maggiori centri politico-commerciali del tempo, in particolare a Palermo, Napoli, Barcellona, Madrid, Londra e Bruges.

Il presente Epistolario – che nasce da una naturale quanto necessaria esigenza comunicativa, quella che intercorre tra una madre e i suoi figli – costituisce un raro esempio di scrittura femminile privata del sec. XV e rappresenta non solo una delle prime testimonianze in lingua volgare, ma anche una delle prime testimonianze scritte da una donna laica, le cui possibilità di scolarizzazione, com'è noto, erano al tempo limitatissime.

2.1 La rilevanza linguistica dell'Epistolario

La figura di Alessandra Macinghi Strozzi desta interesse non solo perché dimostra una certa dimestichezza nel padroneggiare la penna, ma anche perché riesce con disinvoltura a cimentarsi in partite e ragioni: un ambito lessicalmente e storicamente maschile. La prosa di Alessandra si contraddistingue per autenticità, schiettezza e onestà di valori pedagogici, tanto che Contini (1970) la definì la "paradigmatica" tra le prose domestiche del Quattrocento; la sua rilevanza linguistica è data anzitutto dal genere letterario cui va ascritta, ovverosia il genere epistolografico: «Le settantatré *Lettere* pervenuteci rispondono a un bisogno immediato, urgente di comunicazione, non hanno né preoccupazione né destinazione letteraria [...]. Sono scritti assolutamente privati, in cui la stessa grammatica è quella parlata; avvisi, massime, ricordi, notizie, resoconti [...] di fatti, avvenimenti, azioni, proposte relative al nucleo familiare, all'intimità della casa, segreti che non devono andare oltre il mittente e il destinatario [...]» (Doglio, 1984, p. 487).

Per quanto concerne il lessico delle *Lettere*, si intende redigere – in appendice e a completamento della nuova edizione – un *Glossario*, che riceverà e metterà prontamente a disposizione del LEI – *Lessico etimologico italiano* (Pfister, 1979-) – numerosi termini specialistici, fra cui primeggiano quelli relativi all'attività commerciale e finanziaria, che ebbe nell'Italia del tardo Medio Evo e del Rinascimento uno straordinario sviluppo. E poiché taluni di questi termini si sono trasmessi anche alle altre lingue europee (si pensi alla fortuna di termini come *banco*, *banchiere*, *capitale*, *credito*, *debito*, *polizza*, *assicurazione*, *lettera di cambio*, ecc.), ne deriverà un contributo molto utile alla storia della terminologia economica sovranazionale tuttora in uso.

Come messo in evidenza più volte dallo Stussi (2000), i documenti mercantili sono una fonte di inestimabile valore per la conoscenza della storia della lingua e offrono talvolta la possibilità di retrodatare parole o, addirittura, intere espressioni.

Una nuova edizione delle *Lettere* della Macinghi Strozzi metterà finalmente a disposizione degli studiosi un testo filologicamente corretto e linguisticamente affidabile, che potrà sostituire l'edizione curata da Cesare Guasti nel 1877, di cui più volte ne sono stati segnalati i limiti (Trifone, 1989). Verrà così per la prima volta alla luce la grafia realmente utilizzata dall'autrice, sistematicamente modernizzata dal Guasti, e si potranno recuperare i diversi caratteri relativi ai suoni e alle forme, che non risultano nella precedente edizione. Le *Lettere* si prestano inoltre a interessanti rilievi di ordine sociolinguistico, essendo la scrivente una donna di ceto mercantile, assimilabile alla classe dei cosiddetti "semicolti".

3 L'applicazione del TAL per la redazione del *formario*

Per il presente contributo è stato realizzato un primo modello di *formario* tratto dalle prime dieci *Lettere* di Alessandra M. Strozzi, teso a mettere in luce i tratti grafici, fonetici e morfologici caratterizzanti la lingua della scrivente, autentica espressione del fiorentino *argenteo* (cfr. Manni, 1979); trattandosi di un archetipo, lo studio si è limitato all'analisi delle prime dieci lettere dell'*Epistolario*: ciò ha consentito di eseguire un lavoro accurato, ma soprattutto di riflettere sui benefici e sulle criticità date dall'utilizzo degli strumenti offerti dal TAL.

Alcune rese grafiche dell'originale (del tipo *x* per *s*, *y/j* per *i*, *cha* per *ca*, *cho* per *co*, *chu* per *cu*) sono state anzitutto normalizzate conformemente alla grafia moderna; è stata successivamente realizzata una tokenizzazione attraverso l'impiego del modello TreeTagger per l'italiano contemporaneo elaborato da Achim Stein (Schmid, 1994). Si è quindi proceduto con l'assegnazione di un POS per ogni forma delle prime dieci *Lettere*, il cui *corpus* consta, in tutto, di 13.782 occorrenze. Il tagset impiegato, elaborato sulla base del tagset ELRA, è stato semplificato e adattato alle esigenze dell'analisi linguistica eseguita, della quale vengono esposti gli esiti al § 4; il tagset adottato è illustrato nella tabella sottostante, la quale pone inoltre in evidenza le corrispondenze tra le etichette impiegate da chi scrive e quelle proprie degli analizzatori adoperati.

Come si evince dalla tabella, è stata operata una distinzione tra articolo determinativo e indeterminativo; per quel che riguarda le congiunzioni, invece, si è preferito non introdurre alcuna distinzione tra congiunzione subordinante e coordinante, dato che nella lingua del tempo la congiunzione assume una funzione che sovente non è possibile classificare con certezza, giacché polivalente. Le preposizioni, ripartite in semplici e articolate, constano di una terza etichetta, "prep.", sotto la quale sono state fatte confluire le preposizioni improprie.

SPIEGAZIONE	ETICHETTA	CORRISPONDENZE ELRA	CORRISPONDENZE TREETAGGER	CORRISPONDENZE TINT
aggettivo	agg.	AS, AP, AN, DP, DN, DS	ADJ	A, AP, DI, PI
antroponimo	antrop.	SPN	NPR	SP
articolo determinativo femminile	art.det.f.	RS, RP	DET:def	RD
articolo determinativo maschile	art.det.m.	RS, RP	DET:def	RD
articolo indeterminativo femminile	art.indet.f.	RS, RP	DET:indef	RI
articolo indeterminativo maschile	art.indet.m.	RS, RP	DET:indef	RI
avverbio	avv.	B	ADV	B
congiunzione	cong.	C*	CON	CC, CS
interiezione	int.	I	INT	I
non verbale	x	X*	SENT, SYM, PON, FW, LS	FB, FF, FS
numerale	num.	N	NUM	NO
preposizione	prep.	E	PRE	E
preposizione articolata	prep.art.	ES, EP	PRE:det	E+RD
preposizione semplice	prep.sempl.	E	PRE	E
pronome	pron.	P*, Q*	PRO*	PC, PD, PE, DQ
pronome femminile	pron.f.	P*, Q*	PRO*	PC, PD, PE, DQ
pronome maschile	pron.m.	P*, Q*	PRO*	PC, PD, PE, DQ
sostantivo femminile	s.f.	SS, SP, SN	NOM	S
sostantivo maschile	s.m.	SS, SP, SN	NOM	S
toponimo	top.	SPN	NPR	SP
verbo	v.	V*	VER*	V*

Tabella 1. Corrispondenze delle etichette grammaticali.

Si segnalano mediante l'asterisco (*) le etichette trascritte nella forma base, senza i dettagli offerti dai tools. L'etichetta TreeTagger "PRO:demo", per esempio, è stata riportata "PRO*".

Dal momento che nessun POS tagger dispone di un modello per l'italiano volgare, il processo di POS tagging è risultato piuttosto articolato e, per ovviare all'assenza del modello, sono state sperimentate due diverse strategie: la prima è consistita nell'applicare il modello POS tagger addestrato sul *corpus* D(h)ante (Basile e Sangati, 2016), la seconda nel normalizzare il *corpus* delle *Lettere*, così da poter impiegare un POS tagger predisposto per l'italiano contemporaneo. Al fine di eseguire un valido confronto fra le due strategie, sono state manualmente etichettate le prime 1.000 parole del *corpus* delle *Lettere*, le cui assegnazioni sono state comparate con gli esiti dati dai due processi sopra descritti. Impiegando dunque TreeTagger e il parser Stanford CoreNLP – addestrati da A. Basile e F. Sangati sul *corpus* D(h)ante – sono state assegnate le parti del discorso. TreeTagger ha dato un risultato migliore rispetto al parser Stanford: il primo ha dato infatti il 59% dei tag corretti contro il 54% del secondo. Prima di approdare alla seconda strategia, il testo delle *Lettere* è stato normalizzato utilizzando il correttore ortografico GNU Aspell (<http://aspell.net>) e quindi nuovamente etichettato mediante il modello TreeTagger elaborato da Achim Stein e addestrato sull'italiano contemporaneo, il cui impiego ha consentito di approdare a una percentuale di tag corretti pari al 69%. I migliori risultati di POS tagging, tuttavia, sono stati ottenuti utilizzando il tagger per l'italiano contemporaneo Tint (Aprosio e Moretti, 2018) che, una volta applicato al testo precedentemente normalizzato, ha restituito il 72% delle etichette corrette.

Programma	Stanford D(h)ante	TreeTagger D(h)ante	Aspell+TreeTagger (Stein)	Aspell + Tint
Accuratezza	54%	59%	69%	72%

Tabella 2. Descrizione dell'accuratezza dei programmi impiegati nell'assegnazione dei tag alle prime 1000 parole del *corpus* delle *Lettere*.

Nell'attribuzione delle etichette le parti del discorso che hanno presentato maggiori difficoltà sono state aggettivi e pronomi, indipendentemente dal programma impiegato. L'etichettatore Stanford è inoltre risultato particolarmente debole e impreciso nel riconoscimento dei verbi, attribuendo erroneamente tale etichetta a molte altre parti del discorso.

Al fine di incrementare la percentuale di tag corretti, il sistema è stato perfezionato attraverso l'impiego dei dati derivanti dal dizionario TLIO (*Tesoro della Lingua italiana delle Origini*, <http://tlio.ovc.cnr.it/TLIO/>), che ha consentito di attribuire le etichette grammaticali a 5.194 forme e grazie al quale è stato possibile ricavare, inoltre, una puntuale distinzione di genere per 606 sostantivi. Per i sostantivi restanti, che non hanno trovato riscontro all'interno del dizionario TLIO, sono state elaborate alcune regole, basate su una serie chiusa di articoli e aggettivi, a seconda che questi accompagnino sostantivi femminili o maschili; tali regole – per l'elaborazione delle quali sono state manualmente redatte delle liste, comprensive di serie di articoli e aggettivi – hanno consentito di assegnare ad altri 120 sostantivi la distinzione di genere, precedentemente mancante.

Prossimamente si ritiene opportuno impiegare un analizzatore morfologico per l'identificazione automatica del genere delle parole, così da verificarne il grado di correttezza; si intende inoltre sperimentare una terza strategia – simile a quella applicata per il POS tagging del *corpus* MIDIA (Iacobini et al., 2014) –, finalizzata a perfezionare il lessico di TreeTagger per l'italiano contemporaneo attraverso le voci provenienti dal dizionario TLIO.

Per favorire il riconoscimento di antroponomi e toponimi, inoltre, sono stati impiegati i dati ricavati dal *Glossario del Libro dei debitori, creditori e ricordi* che Alessandra Macinghi Strozzi tenne tra il 1453 e il 1473, un testo dunque coevo all'Epistolario oggetto del presente studio e di mano della stessa Alessandra Macinghi Strozzi (Bersano, 2015-16, pp. 271-294).

Dal *corpus* annotato è stato tratto automaticamente un *formario*, strutturato in ordine alfabetico; sulla base di queste due fonti è stata eseguita una breve quanto esaustiva analisi linguistica, di cui si riportano gli esiti più significativi nel paragrafo successivo.

4 Esiti

Per garantire la completa affidabilità dei risultati, è stato necessario confrontarsi costantemente con la trascrizione originale – specie nella fase iniziale del TAL – onde evitare errori nel processo iniziale di trasmissione dei dati – *input* –, e poter essere certi dell'attendibilità dei dati in uscita, *output*.

Il *formario* tratto dal testo delle *Lettere* e realizzato grazie al processo sinora descritto, ha consentito a colpo d'occhio di cogliere i fenomeni e i tratti grafici, fonetici e morfologici tipici del fiorentino *argenteo*, così come di porre in luce quelli che ne esulano, presentandosi inaspettatamente 'controcorrente'. Qui di seguito una breve illustrazione degli esiti linguistici grafici e morfologici più interessanti ricavati dal *corpus* annotato e dal *formario*.

GRAFIA: si rileva un'oscillazione costante per la resa della *l* palatale, con il primeggiare del tipo *gl*, anche dinanzi a *i* (seppure risultino consistenti le rese grafiche *lgl*, rare invece quelle in *li*); meno incerta sembrerebbe la resa grafica della *n* palatale, per la quale primeggia il tipo *ngn* (38 occorrenze), più sparute le rese *ngni*, *gn*; rarissima *gni* (con 3 sole occorrenze); dinanzi a *i* risulta nuovamente schiacciante il tipo *ngn* (16 occorrenze) rispetto a *gn* (2 sole occorrenze).

Ancora, è da evidenziare l'uso della grafia *k* per l'occlusiva velare sorda dinanzi ad *a* nella voce *Karisimo*; tale resa grafica per l'occlusiva velare è fatto notevole: essa risulta infatti del tutto assente nel *Libro dei debitori, creditori e ricordi* di Alessandra Macinghi Strozzi (Bersano, 2015-16, p. 166). Scorrendo gli item presenti alla lettera *h*, spicca la scrizione etimologica *homo*, impiegata due volte in queste prime dieci *Lettere* (8 attestazioni in tutto, invece, per *uomo*, conforme alla grafia moderna).

MORFOLOGIA: notevoli sono i plurali in *-(l)gli < -li*; alcuni esempi: *begli (-lgli)* con 3 occorrenze e *fanciugli (-lgli)* con 2 occorrenze, senza esiti contrari.

Per l'art. det. masch. sing. prevale la forma *il*, sebbene risulti considerevole anche la presenza della variante *el* (si contano in tutto, rispettivamente, 88 e 13 occorrenze), penetrata nel fiorentino intorno alla seconda metà del sec. XIV per influsso dei dialetti occidentali e meridionali (Manni, 1979). Per il plurale dell'art. det. è attestata la forma *e* in luogo di *i*, il sistema ha tuttavia etichettato tutte le *e* presenti (550 occorrenze) come *coniunzioni*; occorrerà senz'altro ovviare all'errore, insegnando alla macchina che *e* può essere anche articolo se posta dinanzi a un sostantivo masch. pl., così che il sistema offra la possibilità di scegliere fra le due etichette: *cong.* oppure *art.det.m.*

Per l'uso di *mie, tuo, suo* invariabili (del tipo, *mie bisogni*) e *mia, tua, sua* pl. masch. e femm. (del tipo, *mia figliuoli*), è risultato maggiormente utile consultare il file contenente il testo delle *Lettere* etichettato automaticamente dal sistema piuttosto che il *formario*, poiché quest'ultimo è privo dei contesti, essenziali al fine di verificare a quale sostantivo pronomi e aggettivi si accordino e dunque comprendere, come nel caso presente, l'uso dei possessivi sopraccitati.

Molto agevole è stata la ricerca delle occorrenze per i numerali *duo* (prevalente) e *dua* in luogo di *due*: le 12 occorrenze di *duo* sono state erroneamente etichettate come *sostantivi*; le 3 occorrenze di *dua* sono state correttamente etichettate come *num.* (numerales) secondo il tagset elaborato da chi scrive; le 7 occorrenze di *due* sono state invece etichettate come *aggettivi*.

VERBI: è attestata una volta soltanto la forma *sete* per *siete*; anche in questo caso, è stato essenziale consultare il file contenente il testo delle *Lettere* etichettato automaticamente dal sistema piuttosto che il *formario*, così da verificare il contesto e sciogliere ogni perplessità rispetto al valore semantico della parola (*sete* sostantivo vs. *sete* verbo).

Schiacciante è la presenza dei tipi *arò, arei* per *avrò, avrei*, che non presentano esempi contrari; tutte le attestazioni, inoltre, sono state correttamente etichettate come *verbi*, a eccezione dei tipi *aresti* (4 occorrenze in tutto) e *arebe* (un'occorrenza) classificati erroneamente come *aggettivi*.

Non sono presenti esempi contrari ai tipi *dia* e *stia*; il primo, che consta di 18 occorrenze in tutto, risulta due sole volte erroneamente etichettato come *sostantivo*; il tipo *stia* occorre una volta soltanto ed etichettato regolarmente.

Infine, a riprova della conformità della lingua di Alessandra al fiorentino *argenteo*, compaiono senza esempi contrari i tipi *fussi* per *fossi* e *fusti* per *fasti*.

Per quel che concerne le desinenze verbali, non è possibile in questa sede offrirne una panoramica esaustiva; basti dire, tuttavia, che grazie al supporto informatico, l'analisi linguistica tesa a individuare le forme nonché il numero di occorrenze per ciascuna desinenza verbale è stata di facile esecuzione, oltre che rapida e accurata. Si segnala una desinenza atipica per la lingua di Alessandra, riscontrata nel *formario* e verificata nell'originale: il tipo *preghiamo* per la 1° pers. pl. del pres. ind.; tale occorrenza è un *unicum* in tutto il testo delle *Lettere*, poiché la scrivente adotta uniformemente la desinenza *-no* per la 1° pers. pl. (del tipo *noi laviano*), tanto nelle *Lettere* quanto nel *Libro dei debitori, creditori e ricordi di Alessandra Macinghi Strozzi* (cfr. Bersano, 2015-16, p. 233).

Si riscontrano 3 attestazioni per la forma metatetica *drento* (Manni, 1979) – di cui una con raddoppiamento dell'occlusiva postconsonantica: *drentto* – in luogo di *dentro* e un'attestazione per la forma metatetica *grillanda*, tipiche del fiorentino *argenteo*. Sono inoltre attestate le forme metatetiche *adrieto*, *adrietro* (con mancata dissimilazione *r-r* in *r-ø*), *indrieto* e, più rara per questo periodo, *dirieto*, derivanti dall'influsso esercitato da altri dialetti toscani (Manni, 1979, pp. 167-168).

Significativa è ancora l'attestazione di *sun* in luogo di *su* nel tipo *in sun un*, con l'inserzione della *n* eufonica in *sun*.¹

Per quel che concerne antroponomi e toponimi, ne sono stati riconosciuti in tutto 346, grazie ai dati ricavati dal *Glossario* del *Libro di debitori, creditori e ricordi* di Alessandra (Bersano, 2015-16); 127, invece, non sono stati riconosciuti secondo le etichette prestabilite ('antrop.' e 'top.');

di questi 127, tuttavia, 43 sono stati riconosciuti ed etichettati come *nomi propri*; 112 sono stati invece etichettati più genericamente come *sostantivi* e solo 4, erroneamente, come *verbi*.

¹ Non si hanno esempi, in queste prime dieci *Lettere*, per il tipo *sur*, che trae origine da *sun* per dissimilazione (*in sun un > in sur un*).

5 Conclusioni

Il *corpus* annotato, da cui è stato tratto il *formario* impiegato per l'analisi linguistica, costituisce non solo una proficua base di lavoro per la redazione del futuro *Glossario* delle *Lettere* di Alessandra Macinghi Strozzi, ma anche un primo quanto fondamentale strumento di analisi; esso ha infatti consentito di selezionare celermente i dati più significativi che andranno opportunamente registrati nel *Glossario* finale, che sarà posto in appendice alla nuova edizione delle *Lettere* di Alessandra Macinghi Strozzi.

Il lavoro di revisione effettuato a mano è stato senz'altro ingente; si ritiene tuttavia che il supporto informatico sia stato essenziale al fine di porre in evidenza taluni tratti, così come la quantità di occorrenze per ogni forma riscontrata: fattore indispensabile, quest'ultimo, per chiarire quanto un fenomeno, o un tratto, fosse frequente nella lingua dell'epoca.

Bibliografia

- Alessio Palmero Aprosio, Giovanni Moretti. 2018. *Tint 2.0: an All-inclusive Suite for NLP in Italian*, in «Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)».
- Angelo Basile, Federico Sangati. 2016. *D(h)ante: A New Set of Tools for XIII Century Italian*. Editato online sul sito della Fondazione Bruno Kessler e accessibile al seguente indirizzo: <https://dh.fbk.eu/D%28h%29ante>
- Ottavia Bersano. Tesi di laurea, a.a. 2015-16. *Il libro dei debitori, creditori e ricordi di Alessandra Macinghi Strozzi (1453-1473). Analisi linguistica*. Tesi di laurea, Università degli Studi di Firenze.
- Gianfranco Contini. 1970. *Letteratura italiana del Quattrocento*. Firenze, Sansoni.
- Maria Luisa Doglio. 1984. *Scrivere come donna: fenomenologia delle 'Lettere' familiari di A. Macinghi Strozzi*, in «Lettere italiane», XXXVI, pp. 484-97.
- Progetto GNU (a cura di). 2000-. *GNU Aspell* (<http://aspell.net>).
- Cesare Guasti (a cura di). 1877. *Lettere di una gentildonna fiorentina del sec. xv ai figliuoli esuli*, Firenze, Sansoni.
- Claudio Iacobini, Aurelio De Rosa, Giovanna Schirato. 2014. *Part-of-Speech Tagging Strategy for MIDIA: a Diachronic Corpus of the Italian Language*, in «Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it 2014)», pp. 213-218.
- Paola Manni. 1979. *Ricerche sui tratti fonetici e morfologici del fiorentino quattrocentesco*, in «Studi di Grammatica italiana», VIII, pp. 115-171.
- Max Pfister (a cura di). 1979-. *LEI - Lessico etimologico italiano*, Wiesbaden, L. Reichert, edito per incarico della Commissione per la Filologia romanza da Max Pfister.
- Helmut Schmid. 1994. *Probabilistic Part-of-Speech Tagging Using Decision Trees*, in «Proceedings of International Conference on New Methods in Language Processing», Manchester, UK.
- Alfredo Stussi. 2000. *Studi di filologia e letteratura italiana in onore di Gianvito Resta*, 2 voll., Salerno Editrice, Roma.
- Opera del Vocabolario italiano – OVI (a cura di). 1997-. *Tesoro della Lingua italiana delle Origini (TLIO)*, pubblicazione periodica online consultabile al seguente link: <http://tlio.ovi.cnr.it/TLIO/>
- Pietro Trifone. 1989. *Sul testo e sulla lingua delle lettere di Alessandra Macinghi Strozzi*, in «Studi linguistici italiani», xv, pp. 65-99.