

Valentina Pollidori
Analisi, trattamento e codifica dei dati testuali
*per la base di dati del Tesoro della Lingua Italiana delle Origini**

SOMMARIO: Nota introduttiva – I. La scheda *BTV*: dati “anagrafici” e dati bibliografici del testo – II. L’edizione e la preparazione del testo – II.1. I riferimenti – II.1.1. Riferimenti topografici: casistica e codifica – II.1.1.1. Riga della pagina e riga del testo – II.1.2. Riferimenti organici e speciali: A. Testi in prosa, B. Testi in poesia – II.2. Segni e caratteri speciali con funzione diacritica – II.2.1. Supplemento sulla rappresentazione di alcuni fenomeni grafico-fonetici.

LEGENDA:

<i>BTV</i>	base di dati della <i>Bibliografia dei testi volgari</i>
EdR	edizione di riferimento
<i>GATTO</i>	programma di <i>Gestione Archivi Testuali Tesoro delle Origini</i> di D. Iorio-Fili
<i>ItalNet</i>	base di dati testuale interrogabile in Internet tramite il consorzio <i>ItalNet</i>
OVI	Centro di Studi «Opera del Vocabolario Italiano»
P	pre-edizione / pre-editore
T	il testo
<i>TLIO</i>	base di dati del <i>Tesoro della Lingua Italiana delle Origini</i>
{abc}	forme per l’interrogazione del Corpus
← → ↑ ↓	freccie orizzontali per i rinvii fra i diversi paragrafi, verticali per quelli fra sezioni di una stessa tabella.

NOTA INTRODUTTIVA

La visibilità del quasi trentennale lavoro del centro studi «Opera del Vocabolario Italiano», raggiunta con la diffusione in Internet della banca dati del *Tesoro della*

* Queste pagine costituiscono l’adattamento di alcune sezioni di un più ampio e specialistico *Manuale* compilato ad uso interno e frutto dell’esperienza collettiva e settoriale dei colleghi e dei collaboratori dell’OVI. Anche in questa occasione non posso non ringraziare tutti coloro che a vario titolo hanno contribuito a porre, a discutere e a risolvere i problemi presentati dalla complessa gestione dei dati del Corpus. Un ringraziamento particolare va alla collaboratrice tecnica Franca Bertini che con l’esperienza acquisita nella battitura, codifica e correzione dei testi ha permesso più di una volta di individuare e correggere molte delle incongruenze sfuggite alla precedenti fasi di elaborazione dei testi.

Lingua Italiana delle Origini e con l'uscita a stampa delle prime voci del *Vocabolario*, rappresenta l'occasione per fornire alcune informazioni supplementari sul lavoro di documentazione, interpretazione e codifica del patrimonio testuale raccolto e organizzato in funzione dei due obiettivi citati.

Tale lavoro si rivela molto poco meccanico quando si valuti la peculiarità di questa base di dati nella quale confluiscono e si fondono tutte le varietà linguistiche dell'antico "italiano": da quelle pluri-documentate dei numerosi testi toscani e veneti, a quelle più rare e meno indagate offerte da testimonianze, spesso isolate, di aree linguistiche marginali e fortemente mescolate¹. Ma nella costituzione di un corpus testuale che si fonda sulle edizioni disponibili piuttosto che sulla riedizione degli originali manoscritti², l'endogena varietà dei fatti di lingua attestata dal complesso dei documenti viene sovraesposta e complicata dall'eterogeneità delle prassi editoriali, che su quella varietà intervengono spesso con finalità e modalità del tutto opposte. Accanto a edizioni rigorosamente scientifiche e filologiche, destinate a un pubblico di specialisti e di studiosi attento al complesso rapporto che lega grafia, fonetica e morfologia in una fase così antica della nostra lingua, ci sono altrettante edizioni che

¹ Si pensi al caso dei *Sermoni subalpini*, il più antico e isolato testimone di un'area piemontese ad alto grado di mescolanza linguistica con elementi francesi e latinismi (l'edizione di questo testo, curata da W. Babilas [München, Hueber, 1968] ha poi finalità storico-documentarie piuttosto che filologiche-linguistiche) a fronte dei numerosi documenti toscani e veneti pubblicati in esemplari edizioni critiche quali, per fermarsi alle più note, quelle curate da Alfredo Schiaffini (*Testi fiorentini del Duecento e dei primi del Trecento*, Firenze, Sansoni, 1926), Arigo Castellani (*Nuovi testi fiorentini del Duecento*, Firenze, Sansoni, 1952; e *La prosa italiana delle origini*: I, *Testi toscani di carattere pratico*, Bologna, Patron, 1982), Alfredo Stussi (*Testi veneziani del Duecento e dei primi del Trecento*, Pisa, Nistri-Lischi, 1965).

² Sulle motivazioni di questa scelta fondativa per l'impresa del *Tesoro* e per l'avvio della realizzazione del *Vocabolario* si rimanda allo storico intervento di Giorgio Pasquali, *Per un Tesoro della lingua italiana*, in «Atti della R. Accademia d'Italia. Rendiconti della Classe di scienze morali e storiche», serie VII, II, pp. 490-521 (poi pubblicato in AA. VV., *Per un grande vocabolario storico della lingua italiana*, Firenze, Sansoni, 1957, pp. 39-107). L'opzione a favore di una decisa preminenza dell'attività lessicografica non si è però disgiunta nel passato da un intenso lavoro filologico di promozione di nuove edizioni e di revisione di quelle esistenti, come puntualmente ricordato da Domenico De Robertis, direttore dell'Ufficio Filologico nel periodo 1965-82 (Id., *L'ufficio filologico dell'Opera del vocabolario, il suo impianto, il suo lavoro*, in *La Crusca nella tradizione letteraria e linguistica italiana*, Atti del Congresso Internazionale per il IV Centenario dell'Accademia della Crusca, Firenze, 29 Settembre - 2 Ottobre 1983, Firenze 1985, pp. 443-51). La forte riduzione del personale dell'OVI, compensata però all'esterno dalla notevole produzione di edizioni critiche, ha consigliato per l'oggi un ulteriore ridimensionamento della pratica filologica del Centro per puntare decisamente verso l'impegno lessicografico; e cfr. Pietro G. Beltrami, *Prima del Vocabolario: finalità lessicali di un archivio medievale* (Tesoro della lingua italiana delle Origini), in AA. VV., *Testi, manoscritti. ipertesti. Compatibilità informatica e letteratura medievale*, Atti del Convegno Internazionale, Firenze, SISMEL - Edizioni del Galluzzo, 1998, pp. 79-95 (in particolare 81-82). A questo stesso volume si rinvia per un quadro d'insieme delle diverse esperienze e metodologie che si sono sviluppate per imprese analoghe.

non esitano a livellare e ammodernare gli elementi grafici, fonetici e morfologici di un testo col dichiarato scopo divulgativo di agevolarne la fruizione.

Le combinazioni e le oscillazioni comprese fra questi due modelli estremi risultano poi pressoché irriducibili ad un sistema univoco, situazione che acutizza i normali problemi posti dalla codifica e che impone, caso per caso, una continua assunzione di responsabilità per lo meno pari alla discrezionalità di giudizio messa in atto nelle scelte interpretative dei diversi fenomeni testuali e di lingua e delle loro numerose rappresentazioni.

Di tutto questo lavoro condotto nel tempo, si descrivono qui gli aspetti più patenti, e sotto molti aspetti “pratici”, ma non per questo meno significativi della problematicità, empirica e teorica, posta dalla costituzione di una banca testuale che si proponga di rendere il più possibile coerenti tra loro dati e informazioni tanto eterogenei.

Nella prima parte di questo articolo si illustrano le modalità e i criteri di registrazione dei testi presenti nella banca dati o destinati ad entrarvi. Ogni testo del Corpus viene infatti schedato in un apposito data-base e corredato di una serie predefinita di informazioni (edizione adottata, data di composizione, area linguistica, ecc.) che costituiscono la “carta d’ingresso” del singolo individuo testuale. Queste stesse informazioni, collegate al programma di gestione del Corpus, vengono poi utilizzate sia per documentare in uscita la fonte dell’esemplificazione, sia per consentire in entrata ricerche più specialistiche e raffinate attraverso restrizioni relative a particolari aree geografiche, tipologie di testi, intervalli cronologici, ecc.

Nella seconda parte si descrive l’analisi che viene condotta sulle singole edizioni per individuare le specifiche caratteristiche editoriali di un testo e per fornire quindi le istruzioni necessarie affinché queste risultino coerenti con il sistema complessivo del Corpus. Infatti, i presupposti in base ai quali le richieste degli utenti di una banca dati testuale possano ricevere risposte affidabili ed esaustive – nei termini di una completezza dell’informazione rispetto alla documentazione offerta –, consistono notoriamente nell’uniformità e univocità della formalizzazione dei dati immessi e nella flessibilità e duttilità dei programmi di interrogazione. Solo questi due requisiti sono infatti in grado di sostituire l’efficienza altrimenti garantita da una lemmatizzazione integrale del corpus testuale³.

Se al secondo requisito si è già efficacemente risposto con l’utilizzazione – ormai prevista da tutti i programmi informatici – di caratteri jolly (asterischi, punti,

³ La lemmatizzazione del *TLIO* è parziale e selettiva in quanto è condotta testo per testo ed è principalmente mirata ad associare ogni forma attestata – indipendentemente dal numero delle sue occorrenze e con l’esclusione di una serie di “parole grammaticali” – al lemma corrispondente. Ma su questo tema si rinvia ad un prossimo specifico intervento su questo stesso *Bollettino*, mentre è il caso di ricordare che la lemmatizzazione, interrogabile nella base dati gestita in sede dal programma *GATTO*, non compare invece nel Corpus che si consulta via Internet.

cappelletti, parentesi, ecc.) che consentono di raccogliere sotto un'unica stringa le numerose e non sempre prevedibili grafie delle forme offerte da testi delle più diverse provenienze geografiche, cronologiche e culturali; il primo patisce ancora di una certa approssimazione dovuta tanto all'eterogeneità dei testi, quanto alla difformità, da edizione a edizione, dei particolari sistemi di interpretazione e resa di identici fenomeni testuali (compendî, errori d'autore o di copista, lacune, guasti, cancellazioni e quindi scioglimenti, correzioni, integrazioni, ricostruzioni, ecc.) e grafico-fonetici (rafforzamento consonantico, assimilazione, scempiamento, aferesi, ecc.).

I singoli sistemi – in sé per lo più perfettamente coerenti, in quanto rispondenti a precise tradizioni scritte o a dichiarate soluzioni scientifiche – risultano problematici solo in rapporto al sovra-sistema che si costituisce con la fusione dei testi nel corpus. Alla stregua di tutti gli altri dati che confluiscono nel Corpus, anche questo sovra-sistema di “segni” interpretativi deve quindi essere reso il più possibile univoco, senza però che il processo di uniformazione alteri la coerenza interna di ogni singola edizione⁴.

Un'avvertenza preliminare: i due principali programmi di gestione della banca dati del *TLIO*, quello su Internet e quello di uso interno, differiscono nel sistema di citazione dei riferimenti bibliografici (→ I e II.1). In questo intervento si fa sempre riferimento alla presentazione e alla visualizzazione del secondo, il programma *GATTO*, che fornisce anche l'uscita redazionale per la compilazione delle voci del *Vocabolario*, mentre si segnalano in nota le differenze più notevoli della versione in Internet.

I. LA SCHEDA BTV: DATI “ANAGRAFICI” E DATI BIBLIOGRAFICI DEL TESTO

Sulla rete interna del Centro è da tempo operativo un data-base di Access strutturato in una serie di campi interrogabili singolarmente o in modo combinato. Qui vengono raccolte, opera per opera, tutte le informazioni utili a corredare gli esempi del Corpus *TLIO* dei riferimenti bibliografici essenziali per il reperimento dei contesti nelle edizioni citate, e quelle necessarie invece ai diversi programmi di gestione del corpus per ricerche più selettive e mirate del materiale testuale⁵. Quando esce l'edizione di un nuovo testo oppure la nuova edizione di un testo già presente nel

⁴ Rispettare fino in fondo i sistemi di rappresentazione adottati per i singoli testi potrebbe infatti creare ambiguità nell'interpretazione di identici segni utilizzati con funzioni diverse e costringerebbe gli utenti a ricorrere continuamente alle edizioni di riferimento per comprendere il loro effettivo valore.

⁵ Non è il caso di ripetere in questa sede quali sono le possibilità di ricerca consentite dai diversi programmi sui quali gira il Corpus *TLIO* (*ItalNet* e *GATTO*), per questo si rimanda alle informazioni contenute alle relative pagine del citato sito web.

TLIO, questa viene registrata nella *BTV* secondo le modalità qui sotto indicate e già parzialmente diffuse nelle pagine web dell'OVI (<http://www.csovi.fi.cnr.it>).

I.1. ABBR[EVIASIONE] TIT[OLO]

Il campo ABBRTIT contiene l'abbreviazione con la quale il testo è citato nel Corpus e nel Vocabolario. Questa può avere una delle seguenti forme:

Autore, titolo, data (provenienza linguistica):	Bono Giamboni, Vizi e Virtudi, a. 1292 (fior.)
Autore, data (provenienza linguistica) [per autori citati con una sola opera]:	Alberto da Massa, XIII s.m. (tosco.)
Titolo, data (provenienza linguistica) [per opere anonime]:	Amaistramenti de Sallamon, 1310/30 (venez.)
Doc. provenienza linguistica, data [per tutti i testi documentari]:	Doc. amiat., 1348
Lett. provenienza linguistica, data [per le lettere]:	Lett. aret., 1370
Poes. an. provenienza linguistica, data [per tutte le rime anonime]:	Poes. an. bergam., p. 1340
Stat. provenienza linguistica, data [per tutti i testi di tipo legislativo]:	Stat. agrig., 1328

Quando si hanno identiche abbreviazioni per schede diverse, l'abbreviazione è seguita dal numero progressivo in parentesi tonde, numero progressivo che si riferisce esclusivamente all'ordine di ingresso di quel testo rispetto al precedente:

ABBRTIT: Stat. mant., 1374 Stat. mant., 1374 (2)

Quando l'abbreviazione si riferisce ad un testo di rilevanza lessicografica ma presente unicamente in edizioni scarsamente affidabili, come è il caso delle edizioni ancora settecentesche dei due testi cavalchiani della *Disciplina degli Spirituali* e del *Trattato delle trenta stolizie* (entrambi pubblicati a cura di G. Bottari, Roma, Pagliarini, 1757), l'ABBRTIT viene racchiusa fra parentesi angolari:

<Cavalca, Disc. Spir., a 1324 (pis.)> <Cavalca, Trenta stolt., a 1342 (pis.)>
--

In questo modo redattori e utenti sono immediatamente avvertiti della "criticità" degli esempi estratti da quei testi.

Quando l'abbreviazione si riferisce ad un testo presente nel Corpus in due diverse edizioni delle quali la più recente è raccomandata per una maggiore affidabilità critica, l'ABBRTIT della più antica viene racchiusa fra \$:

\$Bestiario d'Amore di R. Fornival, XIII (tosco.)\$
\$Stat. pist., 1313\$

Nei due esempi riportati la marcatura indica che i testi presenti nelle edizioni di Giusto Grion (*Il Mare Amorosio, poemetto in endecasillabi sciolti di Brunetto Latini, «Il Propugnatore»* II [1869], p. II, pp. 147-79 e 273-89) e di Sebastiano Ciampi (Pisa, Prospero, 1814) sono presenti nella base di dati anche secondo una diversa edizione giudicata più affidabile (nello specifico si tratta de *Lo Diritano bando, conforto et rimedio delli veraci e leali amadori*, edizione critica a cura di Rosa Casapullo, Firenze, presso l'Accademia della Crusca, 1997; e de *L'Opera di S. Jacopo in Pistoia e il suo primo statuto in volgare (1313)*, a cura di Lucia Gai e Giancarlo Savino, Pisa, Pacini, 1994).

Infatti, in presenza di nuove e più affidabili edizioni di testi già inclusi nel Corpus e già lemmatizzati, si è preferito integrare i nuovi prodotti senza eliminare dalla consultazione quelli preesistenti. La marcatura del titolo inoltre consente di escludere facilmente i testi così connotati dall'interrogazione del Corpus qualora si volessero estrarre dall'insieme dati statistici.

I.2. AUTORE.

Il nome dell'autore è dato nella sequenza "cognome, nome" quando il cognome è propriamente tale, altrimenti nella sequenza "nome cognome" (l'uso è fondamentalmente quello dell'*Enciclopedia Italiana*, compresi casi particolari come "Dante Alighieri", non "Alighieri, Dante").

– La dizione "Anonimo" è riservata ai testi letterari o comunque "d'autore". Per i testi documentari, giuridici, legislativi ecc. si usa invece il simbolo "=".

– I nomi dei mittenti di lettere che non siano da considerare come opere letterarie sono indicati in parentesi graffe { }.

– Le rime di corrispondenti incluse in raccolte d'autore richiedono una scheda autonoma. Per quest'ultimo punto, in considerazione della complessa questione attributiva che investe la tradizione della lirica delle Origini, prima di creare una nuova scheda viene sempre preventivamente verificata l'assenza del testo da immettere attraverso l'interrogazione del Corpus. Questo potrebbe infatti già essere incluso nella banca dati da una diversa edizione (cosa che comunque risulterebbe già nella *BTV*), sotto il nome di un diverso autore o come testo anonimo.

I.3. TITOLO

Per consentire un ordinamento alfabetico fruibile, l'articolo iniziale è posposto, per es. «Costituzioni Egidiane del 1357 (Le)»; «Denunzia volterrana del 1327 (Una)». Ugualmente posposti sono i numerali, per es. «Cedole volgari... (Due)» e brevi sintagmi come nei casi seguenti: «Documento mercantile riguardante Francesco Datini (Frammento di un)»; «Grammatica latino-italiana del sec. XIII (Una nuova)»; «Bestiario in volgare (Tre capitoli di un)».

Il titolo preceduto dal \$ sta ad indicare che quel testo compare nella *BTV* e quindi nel *Corpus* anche in una edizione diversa e più affidabile (← I.1)

I.4. EDIZIONE

Questo campo contiene i dati bibliografici dell'edizione effettivamente utilizzata per l'immissione del testo. Le riviste sono citate tramite abbreviazioni raccolte in una lista consultabile alle citate pagine web.

[TITOLO]	Capitolare degli Ufficiali sopra Rialto
[EDIZIONE]	a c. di Alessandra Princivalli (con introduzione di Gh. Ortalli), Milano, Editrice La Storia, 1993.

Al campo EDIZIONE saranno presto collegati due nuovi campi: uno di tipo logico, che servirà ad indicare la presenza nell'EdR di glossari o di indici lessicali; e un secondo di tipo testo, dove verranno raccolte le principali informazioni bibliografiche su quel determinato testo e sull'edizione citata (precedenti edizioni⁶, recensioni, articoli, note, ecc.) utilizzati o da utilizzare per la correzione del testo e per la sua interpretazione.

⁶ Informazione particolarmente utile per i testi documentari spesso riediti con titoli differenti.

I.5. CURATORE

Il nome del curatore dell'edizione, che già compare nel campo EDIZIONE, è ripetuto in questo campo distinto nell'ordine "cognome, iniziale del nome" per consentire ricerche di testi di cui sia noto solo questo dato, ma anche per ricostruire quanto è stato estratto da edizioni antologiche.

Accanto ai dati bibliografici descritti, la scheda *BTV* contiene una serie di altri sette campi dove vanno inseriti i dati più specificatamente "anagrafici" del testo immesso. Siccome il contenuto di questi campi è sempre espresso in forma abbreviata, anche se facilmente intuitiva, per lo scioglimento delle abbreviazioni adottate si rinvia alle citate pagine web.

I.6 e 7. AREAGEN[ERICA] e AREASPEC[IFICA]

Questi due campi identificano l'area linguistica del testo: il primo contiene la localizzazione geografica più generica, generalmente regionale ("toscano", "veneto" ecc., ma anche "italiano settentrionale", "mediano", ecc.); il secondo quella invece più circoscritta ("fiorentino", "aretino", "pisano-lucchese", "veneziano", ecc.). Se la provenienza non è specificabile al di là dell'area regionale, il campo AREASPEC ripete il dato del campo AREAGEN.

I.8 e 9. DATA DESC[RITTIVA] e DATACOD[IFICATA]

La datazione attribuita al testo è contenuta nel campo DATADESC "in chiaro", con abbreviazioni intuitive e secondo una casistica consultabile. La stessa data compare in forma codificata nel campo DATACOD di modo che l'ordinamento alfanumerico di questo campo dà come risultato l'ordine cronologico dei testi.

I.10. GENERE

Contiene l'indicazione del genere letterario entro il quale è stato classificato il testo: comm., doc., doc. merc., stat., lir., ecc.⁷.

⁷ Se ne veda l'elenco nella *Bibliografia dei testi in volgare fino al 1375 preparati per lo spoglio lessicale*, Firenze, C.N.R. Opera del Vocabolario, 1992, p. 497.

I.11. FORMA

Indica la forma del testo da immettere: ossia se è in versi (V), in prosa (P) o misto (M); o se una forma prevale sull'altra (P/v; V/p)⁸.

I.12. TIPO

In questo campo il testo viene classificato come originale (O), oppure come volgarizzamento (V) se l'ascendente è un testo latino, come traduzione (T) se invece dipende da un'altra lingua romanza o, ancora, come parafrasi (P) o misto (M) quando contenga parti originali e parti volgarizzate.

Naturalmente non tutti i campi vengono riempiti simultaneamente, anzi la prassi più diffusa è quella di aprire una scheda con i dati principali (autore, titolo, edizione di riferimento e quant'altro appare chiaro ad un primo rapido esame) al momento dell'uscita o del reperimento di un nuovo testo che rientri nei limiti cronologici del *TLIO*, e di riservare alla pre-edizione il compito di introdurre le ulteriori informazioni (datazione, area linguistica, apparato documentario, ecc.) attraverso una lettura più approfondita del testo e del suo corredo critico e bibliografico.

II. L'EDIZIONE E LA PREPARAZIONE DEL TESTO.

Ricordando ancora una volta che le scelte di natura più propriamente "filologica" sono demandate alla responsabilità dei singoli editori, il compito principale della pre-edizione sta nell'interpretare i particolari fatti testuali e editoriali e nel tradurli in un linguaggio informatico che rispetti e mantenga il valore e la funzione delle specifiche soluzioni adottate.

La pre-edizione o preparazione di un'opera consiste quindi in un esame delle sue caratteristiche testuali ed editoriali, esame attraverso il quale si accerta la coerenza dei criteri di trascrizione usati dall'editore (ma non sempre – va detto – esplicitamente dichiarati) e si individuano tutti gli elementi extra-testuali destinati ad accompagnare l'esemplificazione a video e a stampa.

Per quanto concerne l'accertamento dei criteri editoriali relativi alla rappresentazione e all'interpretazione del testo, una volta che è stato ricostruito il "diasistema" di segni introdotti dall'editore e che è stata verificata la coerenza della sua applicazione, il pre-editore deve porsi il problema della coerenza di quel particolare sistema col sovra-sistema costituito dall'insieme dei testi già presenti nel Corpus e

⁸ Questa sigla vale solo per quei testi nei quali la presenza di una forma rispetto all'altra è di natura assolutamente episodica e non strutturata.

provvedere di conseguenza ad armonizzare il primo col secondo, come si spiegherà più nel dettaglio al punto II.2.

Per quanto riguarda invece il secondo tipo di dati (II.1), ossia i dati extra-testuali come la disposizione del testo nell'EdR e l'articolazione originale o ricostruita del documento, la pre-edizione ha il compito di istruire un trattamento di codifica tale che nella presentazione dei contesti, a video e a stampa, il sistema di riferimenti risulti fedele al testo dell'EdR e funzionale al generale sistema di citazione.

Le risultanze di questa duplice indagine vengono registrate in un "Modulo" che accompagna il testo alla fase di battitura o allo scanner e che viene poi conservato in un apposito archivio affinché, in caso di dubbio, le soluzioni adottate dal P possano sempre essere ricostruite e verificate.

II.1. I RIFERIMENTI

Una delle prime indicazioni che devono essere fornite dal P riguarda appunto i riferimenti della citazione, ossia quei dati che a corredo degli esempi, a stampa e a video, ne consentono la localizzazione all'interno dell'edizione citata o anche, più raramente, in altra edizione⁹.

I riferimenti si distinguono in "topografici" (RT), relativi all'assetto fisico del testo nell'EdR (volume, pagina, colonna e riga della pagina), e "organici" (RO), inerenti alla struttura interna del testo (libri, capitoli, paragrafi, forme metriche, versi, ecc....). In base alle caratteristiche del testo e dell'EdR la pre-edizione deve stabilire quali di questi riferimenti devono essere utilizzati e come vanno trattati e codificati.

Come criterio generale per la presentazione degli esempi si è inoltre stabilito di non riportare nella citazione degli esempi di testi in prosa il numero di riga di riferimenti organici o speciali (del tipo: riga del capitolo, riga del paragrafo, riga dell'Introduzione, ecc.), giudicato sovrabbondante se non addirittura inutilizzabile e bastando, a supporto del RO, l'informazione topografica di pagina e riga della pagina. Al contrario, si omette l'indicazione della riga della pagina nella citazione degli esempi di opere in versi, ritenendo più che sufficiente per il reperimento dell'esempio la sola indicazione della pagina, naturalmente sempre accompagnata dall'eventuale numero o titolo del componimento e dal numero dei versi¹⁰.

II.1.1. Riferimenti topografici (RT): casistica e codifica

⁹ A scanso di equivoci è opportuno ricordare che il sistema di riferimenti di cui si tratta è rispecchiato nella presentazione dei dati gestiti dal programma *GATTO*, mentre la versione *Ital-Net*, consultabile in rete esterna tramite Internet, ha un suo proprio sistema di citazione che utilizza solo una parte di questi dati "traducendoli" in un modo diverso.

¹⁰ Questa norma è stata stabilita in tempi recenti per cui le voci di Vocabolario già pubblicate contengono ancora il riferimento relativo al numero di riga della pagina.

Con RT si intendono appunto le “coordinate” relative alla disposizione fisica del testo nell’EdR (volumi, pagine, colonne, righe, ecc.).

Nella prima colonna dello specchio che segue si presentano le più diffuse tipologie di *mise en page* seguite dalle indicazioni che caso per caso vengono fornite dal P affinché il riferimento della citazione rispecchi l’assetto dell’EdR; nella seconda colonna si danno le codifiche di elaborazione elettronica corrispondenti alle istruzioni date.

Se il T occupa un solo volume «si assumono come RT la pagina e la riga della pagina»	Il tecnico inserisce nel documento il numero di pagina – costituito dal codice \$0000\$ – ogni volta che nell’EdR si dà il cambio pagina; in questo modo gli esempi saranno seguiti dal numero di pagina e di riga dell’EdR.
Se il T occupa due o più volumi «si assumono come RT il volume, la pagina e la riga della pagina»	Il tecnico provvede a inserire la lettera maiuscola corrispondente al volume (A = vol. I, B = vol. II, ecc.) all’interno del codice di pagina prima del numero, \$A001\$ \$B001\$, di modo che gli esempi a video risultino accompagnati dall’indicazione di volume e pagina del volume.
Se il T è disposto su una o più colonne all’interno di una stessa pagina «si assumono come RT la pagina, la colonna e la riga della pagina»	Il tecnico provvede a inserire la lettera maiuscola corrispondente alla colonna (A = col. I, B = col. 2, ecc.) all’interno del codice di pagina di séguito al numero, \$001A\$ \$001B\$, di modo che gli esempi a video risultino accompagnati dall’indicazione di pagina e colonna.
Se il T si trova sul <i>recto</i> o sul <i>verso</i> di un foglio «si assumono come RT la pagina e la riga della pagina, specificando nell’indicazione della pagina il <i>recto</i> e il <i>verso</i> » ¹¹ .	Il tecnico provvede a inserire le lettere R e V maiuscole, corrispondenti al <i>recto</i> e al <i>verso</i> del foglio, all’interno del codice di pagina di séguito al numero: \$001R\$ \$001V\$.
Se le pagine su cui si trova una parte del T sono in numeri romani «si assumono come RT la pagina e la riga della pagina; specificando nell’indicazione della pagina che si tratta di numeri romani» ¹² .	Il tecnico provvede a inserire la lettera r, che indica il numero romano, all’interno del codice di pagina prima del numero della pagina, reso, come di consueto, con i numeri arabi: \$r001\$.

¹¹ Si tratta di un’eventualità molto rara e che riguarda esclusivamente l’adozione di stampe antiche nelle quali sia numerato solo il *recto* del foglio.

¹² Questa distinzione viene registrata e codificata solo nel caso in cui l’EdR contenga anche parti con numerazione araba; se invece tutte le pagine del volume portano numeri romani la specificazione risulta ridondante e quindi non viene registrata.

II.1.1.1. *Riga della pagina e riga del testo.*

Il conteggio delle righe è fatto automaticamente dai programmi. Questi attribuiscono il n. 1 alla riga di testo che compare subito dopo il codice di pagina, o sulla stessa linea o subito sotto, e incrementano il conteggio di un'unità ad ogni cambio di riga fino a quando l'incontro col successivo codice di pagina non riavverte il conteggio. A causa di questo automatismo, nei casi in cui l'opera contenga parti di testo da non trascrivere, il P è sempre tenuto a segnalarle sulla copia di lavoro indicando il numero di riga dal quale deve ripartire la trascrizione.

II.1.1.1.a. *Righe dell'edizione da eliminare.*

Di norma non vanno conteggiate le righe che non fanno parte del testo, e cioè: titoli correnti, numerazioni di parti del testo, riferimenti all'assetto dell'originale ms. (indicazione di carte; date estratte dal curatore, ecc.), apparati critici, note introduttive o a piè di pagina, ecc. Di tutto questo il P dà specifica avvertenza con la formula che segue, provvedendo a cancellare nella copia dell'EdR le righe che non devono essere conteggiate:

«Nel conteggio delle righe del testo non vanno considerati i titoli correnti (e/o la numerazione dei capitoli, e/o gli apparati ecc. ...)»

II.1.1.1.b. *Righe o parti di testo da eliminare.*

Anche all'interno dei testi si danno casi di parole, sintagmi, frasi o passi di varia estensione che devono essere esclusi dalla memorizzazione perché – ed è il caso più frequente – in altre lingue (quasi prevalentemente in latino, ma anche in antico francese o in provenzale) o perché, più raramente, introdotti dagli stessi editori da testimoni molto tardi o addirittura da altre edizioni con lo scopo di dare completezza al contesto (il caso più noto è quello del testo dantesco che accompagna alcuni commenti alla *Commedia*, e che spesso riproduce la Vulgata)¹³. In presenza di questi casi il pre-editore, valutando l'incidenza del passo in questione sull'intelligibilità dei contesti contigui, dovrà indicare come procedere. Si danno tre casi principali:

– b.1. Se l'opera contiene parti di testo la cui eliminazione non incide sulla comprensione dei contesti adiacenti, il P si limita a cancellare sulla copia di lavoro i pas-

Eliminato: è

Eliminato: di alcuni commenti alla

Eliminato: nei quali il testo dantesco

Eliminato: ¶

¹³ Sia inteso che quello che viene eliminato dall'indicizzazione è il testo dantesco della *Commedia* e non l'eventuale citazione di questo che può comparire nel commento: quest'ultima viene infatti considerata alla stessa stregua del contesto che la contiene.

si che non devono essere riprodotti indicando la pagina e la riga da dove la trascrizione deve riprendere. Sulla base di queste istruzioni il passo eliminato viene sostituito nella visualizzazione degli esempi dalla formula [... OMISSIS ...].

EdR	Visualizzazione
Item demmo a Iohannina due istaia di grano di Quaresima, al priorato di frate Sostengno, per s. xxxij. It dedimus fratri Iohanni de Tosca predicta Iohannina de dicto debito, s. xx (...) Salda la ragione con la detta Iohannina la villa di (...)	Item demmo a Iohannina due istaia di grano di Quaresima, al priorato di frate Sostengno, per s. xxxij. [... OMISSIS ...] Salda la ragione con la detta Iohannina la villa di (...)

– b.2. Se l’opera contiene parole o parti di testo che non devono essere memorizzate ma che devono però comparire nella visualizzazione perché essenziali alla comprensione dei contesti adiacenti, il P le evidenzia sulla copia di lavoro dando istruzione di includerle nell’apposito codice di “esclusione di lingua”: **&L-... &L+**¹⁴. In questo modo il passo sarà visibile nella sua interezza, ma le parole che lo compongono non compariranno fra le forme interrogabili del corpus.

EdR e visualizzazione	Codifica
(...) senza determinazione del fattore della legge. Unde apostulus ad Rom. 13: “E per questo medesimo voi prestì” (...).	(...) senza determinazione del fattore della legge. &L- Unde apostulus ad Rom. 13 &L+ : “E per questo medesimo voi prestì” (...).

– b.3. Se l’opera contiene parti di T solo parzialmente funzionali alla comprensione dei contesti contigui, il P indica sia la quantità di testo da trascrivere entro il codice di esclusione di lingua (ad esempio la prima e l’ultima riga del passo in questione, oppure le prime e le ultime tre forme, o come altro decida) sia quella da sostituire con la formula di omissione: **&L-** [...OMISSIS...] ...**&L+**. Il testo così trattato comparirà nell’esempio fin dove indicato dal P ma le sue forme non saranno interrogabili.

EdR	Codifica	Visualizzazione
(...) dicasi tutto: «Deus miseratur nostri», col «gloria»,	(...) dicasi tutto: «&L-Deus miseratur nostri&L+», col «glo-	(...) dicasi tutto: «Deus miseratur nostri», col «gloria»,

¹⁴ Poiché questa codifica è attiva da non molto tempo, può ancora capitare che forme latine o di altre lingue compaiano nell’indicizzazione: su queste si interverrà in correzione via via che verranno individuate.

responso et verso, etc., «Pater noster et ne nos inducas in temptationem set libera nos a malo <...testo da omettere...> miseratio tue pietatis absolvat» lij. Ad ricevere confrate. Quando alcuno adimanderà (...)	ria», responso et verso, etc., «&L-Pater noster et ne nos inducas [...OMISSIS...] miseratio tue pietatis absolvat &L+» . #lij@. Ad ricevere confrate. Quando alcuno adimanderà (...)	responso et verso, etc., «Pater noster et ne nos inducas in temptationem set libera nos a malo [...OMISSIS...] miseratio tue pietatis absolvat». lij. Ad recevoir confrate. Quando alcuno adimanderà (...)
--	---	---

II.1.2. Riferimenti organici e riferimenti speciali.

Più articolata e differenziata è la casistica dei riferimenti organici (RO), cioè quelli relativi all'articolazione interna di un'opera: per le opere in prosa si tratta prevalentemente di divisioni in libri, capitoli, parti, articoli, ecc.; per le opere in poesia di numerazione di rime (nei canzonieri o nelle raccolte d'autore), di ottave, di canti.

Particolari tipi di riferimento sono poi quelli detti "speciali" (RS) e cioè titoli, prologhi, proemi, introduzioni, rubriche, ecc. che si distinguono dagli altri perché asistematici.

Poiché il riferimento, di tipo organico o speciale, è uno dei parametri usati dai programmi per la definizione del taglio dei contesti, si è evitato di rispecchiare quei sistemi (del tipo: distinzione in terzine; paragrafatura molto fitta) che frantumano eccessivamente il testo rendendo spesso difficoltosa una veloce comprensione degli esempi. In presenza di questi casi si è intervenuti raccogliendo insieme più riferimenti, ma mantenendo però sempre il sistema dell'EdR. Così un testo diviso in paragrafi nell'EdR presenterà nel *TLIO* riferimenti del tipo: *L. I, cap. 3, parr. 1-5*.

Qui di séguito si danno alcune norme generali distinte fra testi in prosa (A) e testi in poesia (B). Negli esempi il segno % – codice di elaborazione elettronica che marca appunto i riferimenti ma non è visibile in chiaro – indica le parti in cui un testo viene diviso.

II.1.2.A. Testi in prosa

1) I testi editi singolarmente e privi di partizioni interne non portano altro riferimento organico che quello costituito dal codice di elaborazione elettronica "%", il che significa che gli esempi estratti da quei testi sono accompagnati dai soli RT, ossia dall'indicazione di pagina e riga.

2) I testi privi di partizioni interne, estratti da raccolte più o meno organiche ma elaborati singolarmente perché distinti tra loro anche da un solo dato "anagrafico" (di-

versa area linguistica, diversa datazione, diverso genere letterario, ecc.), portano come RO il numero progressivo attribuito a quel testo nell'EdR, come nei tre casi qui sotto riportati estratti dal volume de *La prosa italiana delle origini: I, Testi toscani di carattere pratico*, a cura di Arrigo Castellani, Bologna, Pàtron, 1982:

Titoli dell'EdR	ABBRTit	Rif. organici
1. Conto navale pisano	Doc. pis., XII pm.	% 01
2. Testimonianze di Travale	Doc. trav., 1158	% 02
3. Declaratoria pistoiese	Doc. pist., XII ex.	% 03

3) I testi suddivisi al loro interno in sezioni distinte da date proprie, del tipo Statuti, Libri privati o di conti, Memoriali, raccolte di lettere, ecc., portano queste come riferimenti organici. La data (da intendersi solo come anno) viene utilizzata come riferimento organico solo nel caso in cui individua una specificazione ulteriore rispetto all'indicazione cronologica che compare nel campo ABBRTit della *BTV*, come nell'es. che segue:

Doc. pis., 1288-1374	% (1288) % (1309) % (1318) % ...
----------------------	---

e viene racchiusa fra parentesi (tonde o quadre) quando non compare in chiaro ma è stata integrata o ricostruita dal pre-editore sulla scorta di note, introduzioni, ecc.. Questa precede qualsiasi altro riferimento come nell'es. che segue:

Stat. sen., 1320-26	% (1320) Esordio % (1320) 01 % (1320) 02 % % (1326) Esordio
---------------------	---

4) I testi che presentano un contenuto articolato per sezioni non gerarchizzate, come raccolte di novelle, trattati, lettere, ecc. portano come RO il numero progressivo delle singole componenti. In questi casi si è sempre cercato di evitare di ripetere nei RO informazioni relative al genere o al tipo di testo che siano già espresse nell'ABBRTit o anche nel TITOLO, come qui esemplificato per le *Trecentonovelle* del Sacchetti:

ABBRTit	Errata	Corrige
Sacchetti, <i>Trecentonovelle</i> , XIV s.m. (fior.)	% proemio % novella 002 rubr. % novella 002 %	% proemio % 002 rubr. % 002 %

5) I testi organizzati al loro interno in parti gerarchicamente ordinate (Libro, parte, capitolo, ecc.) riproducono nei riferimenti la medesima struttura con formule abbreviate, come nei due esempi che seguono:

Giovanni Campulu, 1302/37 (mess.)	% Intr. % L. I, prologo % L. I, cap. 01 % L. I, cap. 02 %
Ugo Panziera, <i>Trattati</i> , a. 1330 (tosca. occ.)	% 01, cap. 01 rubr. % 01, cap. 01 % 01, cap. 02 rubr. %

La nomenclatura che compare nei RO è abbastanza ristretta e le abbreviazioni impiegate nelle codifiche sono di comprensione immediata:

addizioni	add.	libro	L.
capitolo, -i	cap., capp.	paragrafo, -i	par..., parr.
distinzione	dist.	parte	pt. / Pt.
explicit	expl.	rubrica, -che	rubr.
incipit	inc.	...ecc.	

Più diversificata e ampia è invece la nomenclatura dei riferimenti speciali, che viene di norma riprodotta per esteso proprio perché desultoria e asistemica. Si riportano le denominazioni che si incontrano con maggior frequenza:

Accessus	Incipit	Proemio
Antiprologo	Indice	Prologo / Prol.
Appendice / App.	Intitulatio	Rubricario
Dedica / Ded.	Introduzione / Intr.	Sommario / Somm.
Esordio	Prefazione / Pref.	Tavola / Tav.
Explicit	Premessa / Prem.	Titolo / Tit.

Naturalmente qualsiasi riferimento non compreso nelle precedenti liste potrà sempre essere adottato in una forma coerente con questo sistema.

II.1.2.B. Testi in poesia

1) I testi in versi, dal sonetto al cantare, editi singolarmente hanno come RO la sola indicazione del genere metrico o quella generica di “comp[onimento]”:

Astore da Faenza, <i>La vostra</i> , a. 1396 (fior.), % son.
Baldo fiorentino, XIII/XIV (fior.), % canz.

2) I testi estratti da Raccolte – frequente il caso delle rime di corrispondenza – mantengono nel RO il numero d’ordine che hanno nell’EdR da cui sono riprodotti, preceduto dall’indicazione del genere metrico e seguito dalla lettera che di solito indica l’ordine dei testi poetici all’interno di tenzoni e corrispondenze, come nel caso del sonetto dubitativamente ascritto al Bellondi estratto dall’edizione Bettarini delle *Rime* di Dante da Maiano (Firenze, Le Monnier, 1969):

Puccio Bellondi (ed. Bettarini), XIII ex. (fior.), % son. D. 5a.
--

3) I testi inclusi in raccolte d’autore o di altro genere hanno come RO l’abbreviazione del genere metrico (espressa in caratteri minuscoli) seguita dal numero d’ordine che il singolo testo ha nell’EdR. Gli esempi tratti dalle *Rime* di Chiaro Davanzati (a cura di Aldo Menichetti, Bologna, Commissione per i testi di Lingua, 1965) saranno quindi presentati nella forma che segue:

Chiario Davanzati. XIII m. (fior.)	% canz. 01 – % canz. 02 – % % son. 01 – % son. 02 – % % son. D 01 – % son. D 02 – % ...
------------------------------------	---

dove il numero d'ordine è sempre reso in cifre arabe: principalmente per motivi di economia di spazio nella citazione, ma anche perché basta l'indicazione della forma metrica (canz. / ball. / son. 01) a risolvere il problema di identiche serie numeriche spesso distinte nell'EdR con cifre arabe e romane. Nei casi in cui la numerazione dei testi nell'EdR sia continua, cioè senza distinzioni per metri, l'abbreviazione della forma metrica precederà fra parentesi quadre il numero d'ordine¹⁵.

Francesco di Vannozzo, Rime, XIV s.m. (tosco.-ven.)	% [1388-89, canz.] 001 % [canz.] 002 % [1374, canz.] 003 % [son.] 004
---	---

4) I testi d'autore raccolti sotto un unico titolo complessivo avranno come riferimenti interni i titoli specifici di ogni singolo testo, come nei due casi qui sotto riportati.

Bonvesin, Volgari, XIII tu. d. (mil.)	% Disputatio mensium % De Sathana cum Virgine % De peccatore cum Virgine
Petrarca, Trionfi, 1351(?) - 74.	% T. Cupidinis I % T. Cupidinis II % T. Cupidinis III % T. Pudicitiae

5) Le opere in versi strutturate al loro interno in partizioni sistematiche gerarchicamente organizzate o contenenti riferimenti speciali vengono trattate alla stessa stregua delle opere in prosa (cfr. sopra):

¹⁵ Questo evita che nelle citazioni del Vocabolario si incontrino due parentesi tonde a contatto e indica anche che il numero del componimento, fuori di parentesi, si riferisce al numero complessivo delle rime, non al numero dei rappresentanti di una specifica forma metrica.

Bestiario moralizz., XIII (tosca./aret.-castell.)	% 01 rubr. % 01 % 02 rubr. % 02 %
Boccaccio, Filostrato, 1335-36 (?)	% incipit % proemio % pt. 01, ott. 001 rubr. % pt. 01, ott. 002 % pt. 01, ott. 002 rubr. %
A. Pucci, Reina, a. 1388 (fior.)	% cant. I, ott. 01 % cant. I, ott. 02 % cant. I, ott. 03 %

Come per la prosa, così anche per la poesia si dà un elenco delle abbreviazioni più frequenti utilizzate nei RO:

ballata	ball.	lauda, laude	l. o lauda
cantica, canto	c.	madrigale	madr.
cantare	cant.	ottava	ott.
canzone	canz.	serventese	serv.
canzonetta	canz.tta	sestina	sest.
componimento	l. o lauda	sonetto	son.
discordo	disc.	stanza	st.
frottola	frott.	strofa, strofe	str.

Un'avvertenza particolare riguarda la presenza di date specifiche per singole rime comprese in raccolte. In questi casi le date precedono la numerazione del componimento entro parentesi quadre:

Buccio di Ranallo, Cronaca, c. 1362 (aquil.)	% [1338] son. 01 % [1338] son. 02 % [c.1340] son. 03
--	--

questo anche nel caso in cui anche l'indicazione della forma metrica sia stata integrata dal P (cfr. sopra l'es. di Francesco di Vannozzo).

* * *

Le indicazioni fin qui fornite hanno naturalmente un carattere prescrittivo generale poiché, come si è già detto, neppure sulla scorta di quanto è già passato nel Corpus è possibile normare integralmente la tipologia dei riferimenti, la loro nomenclatura, la loro applicazione. Ogni edizione e ogni testo – fatte salve alcune generiche affinità come quelle che abbiamo esposto – offrono spesso problemi specifici ai quali solo la discrezionalità di giudizio del pre-editore può far fronte individuando, caso per caso, il sistema di citazione più congruo all'assetto testuale consegnato dall'editore.

II.2. Segni e caratteri speciali con funzione diacritica

Di ogni segno o carattere introdotto nel testo dall'editore il P ricostruisce ed esplicita la particolare funzione critica o diacritica, e ne indica il trattamento più appropriato rispetto alla coerenza interna del sistema specifico e di quello complessivo. Sulla base di queste indicazioni nella fase di battitura vengono quindi adottate le codifiche corrispondenti ai valori specifici.

Nella prima colonna della tabella che segue si danno i principali segni e caratteri che si incontrano nelle edizioni dei nostri testi. Di ognuno di questi sono indicate le diverse applicazioni attestate, seguite, fra virgolette basse, dalle istruzioni del P relative al loro trattamento. Nella colonna di destra si illustra invece il trattamento più propriamente tecnico che i segni e i caratteri dati richiedono affinché le forme interessate da questi vengano correttamente indicizzate e visualizzate. Sono in grassetto le "voci" dei segni nella prima colonna e i codici di elaborazione elettronica nella seconda¹⁶. Si ricorda che le forme investite da questi segni non richiedono speciali modalità d'interrogazione, in quanto sono comunque riconosciute dal programma in base alla richiesta della forma stessa o nella presentazione dei contesti¹⁷.

Usi dell'EdR e istruzioni del P	Trattamento e visualizzazione
l' apostrofo che indica elisione o apocope «si copia come si trova».	Come si trova, e cioè senza inserire uno spazio fra la forma elisa e quella che segue perché è lo stesso programma che provvede a distinguere accorpando l'apostrofo alla prima forma: <i>ll'omo</i> viene correttamente indicizzato

¹⁶ Restano fuori da questa trattazione i segni di interpunzione che di norma vengono riprodotti come si trovano nell'EdR, con l'unica eccezione degli apici, per i quali è necessaria una particolare codifica – &" – affinché siano tenuti distinti dall'analogo segno di valore diacritico.

¹⁷ Alla richiesta della forma {pe} il programma d'interrogazione, adeguatamente impostato, risponde con le seguenti nove forme attestate nel corpus e distinte da specifici diacritici: *pe* | *pè* | *pê* | *pê* | *pe·* | *pe'* | *pè'* | *ppe'* | *ppe·*. Se poi una di queste forme, in uno o più testi, dovesse essere integralmente o parzialmente interessata da parentesi quadre o da qualsiasi altro segno speciale, questo risulterà nella visualizzazione dei singoli contesti.

	in <i>ll'</i> e <i>omo</i> .
l' apostrofo che indica aferesi «va unito alla forma che segue e separato da quella che precede con uno spazio».	Lo spazio fra la parola investita da aferesi e quella che precede consente al programma di indicizzare correttamente la forma aferetica a partire dal segno ' : <i>'l</i> , <i>'nferno</i> , <i>'ntorno</i> , ecc.
l' apostrofo che indica elisione di vocale interna a parola, del tipo <i>anda'me</i> , <i>ballà'le</i> «va tenuto unito alle sillabe contigue e la parola che lo contiene va inserita entro il codice speciale previsto».	Le parole che contengono apostrofi interne vengono copiate come si trovano e racchiuse entro i codici &A...&I (&A <i>anda'me</i> &I, &A <i>ballà'le</i> &I) così che nell'indicizzazione la forme risultino nella loro interezza, e cioè <i>anda'me e ballà'le</i> .
l' apostrofo isolato o suspirium che indica la scomparsa di forme costituite da una sola vocale, del tipo <i>tra ' consoli</i> , «va copiato come si trova sul testo e cioè separato dalle forme contigue».	L'apostrofo isolato nell'indicizzazione risulta come forma a sé, indipendente dalle due forme contigue: '.
Il punto di abbreviazione , presente in forme del tipo <i>dr.</i> , <i>kl.</i> , <i>lib.</i> , <i>sol.</i> , ecc., «deve sempre essere convertito nel segno speciale previsto 8 ».	L'adozione di questo codice al posto del punto (<i>den8</i> , <i>li8</i> , <i>sol8</i> , ecc.) consente al programma di distinguerlo dall'identico segno di interpunzione e di attribuirlo correttamente alla forma cui appartiene: <i>dr.</i> , <i>kl.</i> , <i>lib.</i> , <i>sol.</i> , ecc.
Il punto alto che segnala il raddoppiamento o rafforzamento consonantico in fonosintassi, del tipo riconducibile ad un processo di assimilazione in sandhi esterno di consonanti finali latine (<i>a ffare</i>) «deve essere eliminato e sostituito da spazio».	Seguendo questa indicazione sia nell'indicizzazione (<i>a ffare</i>) che nella visualizzazione (<i>a ffare</i>) entrambe le forme vengono rappresentate senza punto alto (→ II.2.1).
Il punto alto usato nell'EdR per indicare la caduta di una consonante finale (<i>-n</i> , <i>-l</i> , <i>-m</i> , <i>-r</i>) in séguito ad assimilazione (ed eventuale successivo scempiamento) con la consonante iniziale della parola che segue (<i>i-rria</i> o <i>i-ria</i> ; <i>de-re</i> ; ecc.) «va battuto a carico della forma che precede e separato da spazio da quella che segue».	Seguendo questa indicazione, nell'indicizzazione la prima forma compare seguita dal punto in alto (<i>de· i·</i>), mentre la seconda, con o senza raddoppiamento consonantico, ne è priva (<i>re / rria</i>). Nella visualizzazione gli esempi compaiono nel modo seguente: <i>de· re</i> , <i>i· rria</i> , ecc. (→ II.2.1).
Il punto alto interno a parola, adottato nell'EdR per segnalare fenomeni di assimilazione consonantica (seguita o meno da scempiamento) interni a parola, del tipo <i>co·lla</i> (prep. art.), <i>e·llo</i> per 'nello', <i>prendo·lo</i> per 'prendonlo', ecc. «deve essere battuto col codice speciale previsto perché la forma resti unita».	Con l'applicazione del codice &· nei casi segnalati dal P (<i>co&·lla</i> , <i>e&·llo</i> , ...) la forma viene indicizzata nella sua interezza e viene visualizzata sia nell'elenco delle forme che nella presentazione degli esempi col punto alto interno: <i>co·lla</i> , <i>e·llo</i> , <i>prendo·lo</i> , ecc. (→ II.2.1).
Il punto alto isolato, preceduto o meno da un altro segno diacritico, che segnala il completo	Il punto alto, isolato o unito al carattere diacritico che lo accompagna, costituisce una

dileguo di parola ridotta ad una sola consonante (<i>che·(r)re</i> , <i>che · (r)re</i> , <i>che ·· (r)re</i>) «va copiato come si trova, ossia unito al segno diacritico che lo accompagna ma separato dalle parole contigue».	forma a sé sia nell'interrogazione sia nella visualizzazione: · ·· '· (→ II.2.1).
Il punto alto usato nell'EdR per rappresentare l'unione grafica con la parola che precede delle forme enclitiche di pronomi o articoli ridotte al solo elemento consonantico (<i>che·l</i> , <i>che·n</i> , <i>E·l</i> , <i>no·m</i> , ecc.) «deve essere unito alla forma che segue e separato da spazio da quella che precede».	In questo modo il punto alto compare unito solo alla particella enclitica tanto nell'indicizzazione (<i>che e ·l ·m ·n no ecc.</i>) quanto nella visualizzazione: <i>che ·l</i> , <i>che ·n</i> , <i>E ·l</i> ; <i>no ·m</i> ; ecc. (→ II.2.1)
Il puntino espuntorio sottoscritto usato nell'edizione di opere in versi per indicare l'espunzione di lettere o sillabe soprannumerarie (come nel caso della <i>-e</i> di <i>'mperadore</i> nel verso <i>Secondo fu lo 'mperadore Ruberto</i>) «va eliminato e sostituito con parentesi angolari che racchiudono le lettere espunte» ¹⁸ .	I puntini espuntori sottoscritti vengono eliminati e le lettere espunte vengono racchiuse fra parentesi angolari precedute codice & (<i>'mperador&<e></i>). In questo modo la parola viene indicizzata nella sua interezza (<i>'mperadore</i>), e visualizzata con la lettera finale espunta: <i>'mperador<e></i> .
Il trattino usato nell'EdR per il rafforzamento prevocalico della consonante finale dei monosillabi proclitici, del tipo <i>nonn-è</i> , «va eliminato e sostituito da uno spazio che separa le due forme contigue».	Secondo tale indicazione il trattino viene eliminato e le due parole risultano separate sia nell'indicizzazione (<i>è nonn</i>) sia nella visualizzazione: <i>nonn è</i> .
Il trattino usato per la <i>l</i> enclitica, articolo o pronome, del tipo <i>E·l</i> , <i>ke·l</i> , <i>sì·l</i> , <i>sopra·l</i> ecc., «va sostituito col codice previsto unito alla <i>l</i> e separato con spazio dalla parola che precede».	Secondo tale istruzione il trattino, sostituito dal codice 7 unito alla <i>l</i> (<i>E 7l</i> , <i>ke 7l</i> , <i>sì 7l</i> , <i>sopra 7l</i>), compare sempre a carico della <i>l</i> enclitica nella visualizzazione delle forme (<i>e ke ·l sì sopra ecc.</i>) e degli esempi (<i>E ·l</i> , <i>ke ·l</i> , <i>sì ·l</i> , <i>sopra ·l</i> , ecc.) (→ II.2.1)
Il trattino di unione che nell'EdR indica l'interpretazione unita di particolari sintagmi (<i>Cò-de-faar</i>) o l'enclisia di pronomi atoni in forme verbali (<i>'gravedà-sse</i> , ecc.) «va perforato con il codice previsto, unito alle parole contigue».	Seguendo l'indicazione il trattino viene sostituito dal codice 7 e lasciato unito alle forme contigue (<i>Cò7de7faar</i> ; <i>'gravedà7sse</i>); in questo modo sia nell'indicizzazione sia nella visualizzazione risultano come componenti di un'unica forma: <i>Cò-de-faar</i> ; <i>'gravedà-sse</i> .
Il trattino di rima al mezzo che nei testi in versi segnala la presenza di parole in rima interna (<i>dolce meo sire – nonn è lo mio core</i>) «va integrato con il codice previsto unito alla parola rima».	Il codice &T unito alla parola in rima interna e seguito da trattino (<i>sire&T-</i>) non compare nella visualizzazione dei contesti ma serve ad una particolare funzione dei programmi di interrogazione che permette di estrarre da un testo o dall'intero Corpus tutte le parole-rima,

¹⁸ Questa codifica era destinata in origine alle vere e proprie espunzioni interne a parola, ma ha mutato funzione per motivi di natura tecnica che non è qui il caso di ripercorrere.

	in fine o all'interno di verso.
Le parentesi tonde che segnalano lo scioglimento dei compendî, del tipo <i>lana (et) lib(ras)</i> , «vanno copiate come si trovano o precedute dal codice previsto quando si tratta di scioglimenti interni alle forme» ¹⁹ .	Queste vengono riprodotte come si trovano e fatte precedere dal codice & quando interne: <i>lana (et) lib&(ras&)</i> . In questo modo la forma viene indicizzata nella sua interezza (<i>et libras</i>) e visualizzata come compare nell'EdR.
Le parentesi quadre con valore di integrazione o di ricostruzione testuale, del tipo [<i>n</i>] o <i>eb[b]ero</i> , «vanno copiate come si trovano o precedute dal codice previsto quando si tratta di integrazioni interne alla forma» ²⁰ .	Queste vengono copiate come si trovano [<i>n</i>], e fatte precedere dal codice & quando l'integrazione è interna a parola, <i>eb&[b&]ero</i> . In questo modo le forme integrate vengono indicizzate senza segni diacritici (<i>'n ebbero</i>) ma compaiono nella presentazione dei contesti nella forma dell'EdR.
Le parentesi angolari usate nell'EdR con valore di integrazione o di ricostruzione «devono essere sostituite con le parentesi quadre (↑) e trattate di conseguenza».	Queste vengono sostituite dalle parentesi quadre, precedute da & quando l'integrazione è interna a parola.
Le parentesi angolari esterne che racchiudono parole ripetute (<i>San <Paulo> Paulo ...</i>) o parti di parole prive di valore lessicale (<i>Andrucciolo <p> quali ...</i>) espunte dallo scrivente o dall'editore, «vanno eliminate con quanto contengono».	Come indicato dal P, le parti espunte e racchiuse fra parentesi aguzze vengono omesse nella trascrizione del testo che, quindi, nei due casi citati appare nella forma: <i>San Paulo ... e Andrucciolo quali ...</i>
Le parentesi angolari esterne che racchiudono parti espunte dallo scrivente o dall'editore ma che contengono parole di senso compiuto, del tipo <i>Se la carta <venisse meno> s'ismarrisse</i> , «vanno copiate come si trovano».	Queste vengono riprodotte come si trovano e le forme racchiuse al loro interno vengono regolarmente indicizzate come forme del testo.
Le parentesi angolari interne con valore espuntivo, del tipo <i>s<c>anta</i> , «vanno eliminate con quanto contengono e la parola interessata da espunzione va copiata nella sua interezza».	Nella trascrizione delle parole interessate da espunzione interna la parte espunta viene eliminata insieme alle parentesi che la racchiudono senza speciali indicazioni: <i>s<c>anta</i> dell'EdR sarà quindi battuto, indicizzato e visualizzato come <i>santa</i> .

¹⁹ Per non appesantire l'operazione di battitura dei tecnici, nelle più recenti trascrizioni di testi pesantemente farciti di parentesi tonde per lo scioglimento dei compendî più usuali si è stabilito di non trascrivere le parentesi tonde fermo restando che questo tipo di intervento viene sempre e comunque registrato nel citato modulo di accompagnamento.

²⁰ Un uso particolare delle parentesi quadre è quello adottato per la ricostruzione delle rime di versi assenti nei testimoni per lacuna o omissione, del tipo: ... [*-ello*]. Poiché queste desinenze rimiche non hanno rilevanza lessicale, esse vengono copiate fra i codici di campo-formula #...@, in modo tale che siano presenti negli esempi, ma non vengono indicizzate fra le forme del Corpus.

<p>Le barre oblique scempie e doppie adottate in alcune più recenti edizioni per distinguere gli errori dello scrivente dalle cancellature dello stesso, del tipo /co/ co(n)fini, «devono essere convertite in parentesi angolari» sempre che – come si è detto sopra per le angolari (↑) – il loro contenuto abbia consistenza lessicale.</p>	<p>A seconda della consistenza lessicale delle parti delimitate da barre queste verranno trascritte fra parentesi angolari oppure non verranno riprodotte, come è stato fatto per il caso citato di /co/.</p>
<p>I punti di lacuna che indicano una o più lettere illeggibili all'interno di una forma precisamente delimitata, del tipo co..o o .rre.icio «devono essere riprodotti col codice previsto».</p>	<p>In questi casi il punto (o i punti) di lacuna corrispondenti alle lettere illeggibili vengono sostituiti dal codice 8 (co88o, 8rre8icio) che consentirà l'indicizzazione e la visualizzazione della parole lacunose nella forma dell'EdR: co..o., .rre.icio.</p>
<p>I punti di lacuna usati nell'EdR per guasti o parti illeggibili dell'originale di estensione non definibile «vanno riprodotti fino a un massimo di 10 e racchiusi fra i codici di campo-formula»²¹.</p>	<p>Questi vengono copiati fino a un massimo di 10 e inseriti entro i codici di campo-formula #...@²²: questo per distinguerli dai punti di sospensione che in quanto segni di punteggiatura possono essere utilizzati dal programma per la delimitazione dei contesti.</p>
<p>punti di lacuna che segnalano guasti di estensione testuale non precisamente definibile, del tipo et ac o l sole, «devono essere battuti come si trovano fino ad un massimo di 10, e vanno inclusi nel campo-formula insieme ai residui frammenti di parole che non sono riconducibili a forme di estensione definita».</p>	<p>La parte di testo indicata dal P viene racchiusa nel campo-formula (#ac o l@) in modo che sia visibile nei contesti ma che non venga indicizzata. Nell'esempio citato non sono ovviamente incluse nel campo-formula le due forme estreme et e sole entro le quali si verifica il guasto poiché esse hanno senso compiuto.</p>
<p>Il carattere corsivo usato nell'EdR per la ricostruzione di parti di parole illeggibili o erronee nell'originale, del tipo “una” dove si distinguono solo le lettere estreme, «va riprodotto come si trova».</p>	<p>In base alla segnalazione le parti in corsivo vengono battute entro il codice &K ...&k (u&Kn&ka): in questo modo la forma verrà visualizzata come appare nell'EdR, e cioè «una».</p>
<p>Il corsivo usato nell'EdR per lo scioglimento dei compendi «va convertito in tondo e racchiuso fra parentesi tonde (↑)».</p>	<p>Seguendo l'indicazione del P, la parte in corsivo verrà sostituita dal tondo e messa in parentesi tonde precedute dal codice &, come sopra indicato.</p>

²¹ Il programma *GATTO* permette, fra le altre possibilità, di delimitare la visualizzazione dei contesti sulla base della punteggiatura forte o debole, compresi quindi i punti di sospensione. Per questo è necessario che i punti di lacuna siano tenuti distinti da quelli di sospensione attraverso la loro inclusione nell'apposito codice del campo-formula.

²² L'applicazione dei delimitatori di campo-formula, costituiti dai simboli # e @, vale, oltre che al caso illustrato, per tutti quegli elementi del testo che, privi di valore linguistico, devono comunque comparire nella visualizzazione degli esempi (cfr. oltre).

Il corsivo usato nell'EdR per titoli di opere («nel libro <i>Di Ponto</i> ») o per singole parole con specifico rilievo «va riprodotto come si trova. Se le parole in corsivo sono in altra lingua, queste devono comunque essere copiate in corsivo e inserite nel codice di esclusione lingua (← II.1.1. 1.b.2)».	Le parole in corsivo vengono riprodotte e racchiuse entro il codice &C ...&c (<i>nel libro &Cdi Ponto&c</i>) in modo che il corsivo venga mantenuto nella visualizzazione.
I numeri romani , espressi in lettere maiuscole o minuscole (<i>iiij</i> , <i>MCCC</i>) «devono essere riprodotti come si trovano e inclusi in campo-formula» ²³ .	I numeri arabi o romani vengono racchiusi fra i codici di campo-formula #...@ (#iiij@ , #MCCC@), in modo che risultino nella visualizzazione dei contesti ma non siano indicizzati.
La nota tironiana 7 e il compendio & , «vanno sciolti con <i>e</i> , <i>et</i> o <i>ed</i> (a seconda dell'uso dei singoli testi) e racchiuse fra parentesi tonde».	I compendî citati sono convertiti nelle forme indicate dal P ed inseriti fra parentesi tonde: (<i>et</i>).

Altri elementi di cui la P dovrà indicare il trattamento sono:

- la **crux critica** o **crux desperationis** †, usata per delimitare zone del testo (parole o frasi) di senso non decifrabile viene riprodotta dove si trova entro i segni previsti e in campo-formula in modo da escluderla dall'indicizzazione: **#}crux{@**;
- il **sic** tondo o corsivo, usato nell'EdR fra in parentesi tonde o quadre per commentare forme "critiche", viene riprodotto in tondo fra parentesi quadre entro il campo-formula in modo da escluderlo dall'indicizzazione: **#[sic]@**;
- figure o simboli, usati nell'originale al posto di alcune specifiche parole e riprodotti nell'EdR (abbastanza frequente nei testi documentari l'uso del simbolo † al posto della parola *croce* in determinati sintagmi), vengono "tradotti" con la parola che simboleggiano. Questa viene battuta nel testo racchiusa entro il codice **#} { @** per escluderla dall'indicizzazione:

EdR	Codifica	Visualizzazione
... pregò la santa † ché ...	pregò la santa #}croce{@ ché ...	pregò la santa }croce{ ché ...

²³ In presenza di numeri romani che potrebbero ingenerare ambiguità nella lettura dei contesti perché perfettamente corrispondenti a lettere (*v*, *i*, *l*) o a sillabe (*li*), il P darà indicazione di inserire nel campo-formula, prima e dopo il numero romano, un punto fermo: ad esempio il numero romano *v* verrà codificato **#.v.@** e quindi visualizzato *.v*.

Nel caso non fosse possibile ricondurre l'immagine ad una parola che la identifichi, basterà usare l'indicazione *figura* codificata allo stesso modo: #}figura{@.

- I nomi dei personaggi premessi nei testi dialogici alle rispettive battute, originali o integrati che siano, «vanno racchiusi entro il codice speciale previsto [e in campo-formula se integrati dall'editore]»: quindi \Messere\; \Madonna\ se originali, ma #\Messere \@; #\Madonna\@ se integrati dall'editore.

* * *

Qualsiasi altro elemento del testo o dell'EdR non compreso in questo elenco che sia stato o sarà giudicato rilevante ai fini dell'intelligibilità del contesto o della sua reperibilità, viene riprodotto nella battitura con l'istruzione di includerlo nei codici di campo-formula. Ad esempio, nel caso della *Vita di San Petronio* curata da Maria Corti²⁴, è stato recuperato a testo un passo che la curatrice aveva relegato in nota. Per facilitare il reperimento di quella particolare sezione nel testo informatizzato la si è fatta precedere dall'indicazione di rinvio alla specifica nota, indicazione racchiusa fra parentesi quadre e campo-formula: #[n.38]@.

Lo stesso trattamento è stato adottato, in questi ultimissimi tempi, nei casi in cui il P ha ritenuto di dover modificare o integrare la lezione dell'EdR mantenendo però quest'ultima per confronto. Così nell'edizione diplomatica di *Habito laudevele in cui persona posa* di Orso Orsini²⁵, la lezione riportata dall'editore *perdere* è stata sostituita in fase di pre-edizione dalla forma integrata *p[on]dere* con rinvio alla lezione del ms. inclusa fra parentesi quadre e campo-formula: *p[on]dere* #[ms.: *pde-re*]@.

²⁴ *Vita di San Petronio, con un'Appendice di testi inediti dei secoli XIII e XIV*, a cura di M. Corti, Bologna, Commissione per testi di lingua, 1962 («Scelta di curiosità letterarie», 260).

²⁵ In G. Bertoni, *Un nuovo poeta italiano delle origini*, «Archivum Romanicum», II (1918), pp. 365-67.

II.2.1. *Supplemento sulla rappresentazione di alcuni fenomeni grafico-fonetici.*

Nella resa di alcuni fenomeni grafico-fonetici si riscontra nel Corpus una notevole varietà di rappresentazioni. Questa varietà, parzialmente endogena fin quando dipendente dalla grande eterogeneità linguistica dei testi immessi, risulta fortemente accentuata dai diversi sistemi di trascrizione adottati nelle singole edizioni. Accanto alle edizioni più antiche (sette-, ottocentesche) che spesso scontano una scarsa attenzione alla veste grafematica dei testimoni riprodotti o adottati (solo episodicamente rivista integralmente dal personale dell'OVI) privilegiando nella trascrizione gli usi consolidati dalla tradizione, si affiancano molte edizioni moderne che pur adottando sistemi di formalizzazione senz'altro più raffinati e analitici non per questo risultano sempre coerenti e univoche fra loro.

Questo naturalmente comporta che le procedure elettroniche, le quali nella loro notoria rigidità distinguono le unità linguistiche solo in base all'aspetto grafico, esaltino tali difformità di trascrizione di identici fenomeni rendendo assai dispersiva l'interrogazione di particolari zone della lingua²⁶.

Escludendo integrali interventi di revisione che richiederebbero in numerosi casi di tornare al controllo degli originali manoscritti, o normalizzazioni che potrebbero provocare un effetto domino anche sulle rappresentazioni più consolidate dall'uso, si è stabilito per il presente di intervenire sulle difformità più critiche riducendole ad un sistema il più possibile univoco ma salvaguardando contemporaneamente il principio di fedeltà all'edizione di riferimento e quindi al suo sistema interno: soluzione forse eccessivamente pragmatica ma che non esclude futuri ripensamenti anche radicali in considerazione del fatto che la flessibilità degli attuali programmi informatici consente di intervenire all'infinito nella correzione sui testi²⁷.

Per illustrare il problema nella sua portata e per cercare di agevolare le ricerche su queste zone della lingua²⁸ sono stati estratti i vari tipi di rappresentazione grafica delle forme coinvolte dai più diffusi fenomeni grafico-fonetici (rafforzamento consonantico, assimilazione e scempiamento, aferesi, elisione, ecc.) così come si pre-

²⁶ Su queste problematiche vertono alcuni degli interventi pubblicati ora citato volume *Testi, manoscritti. ipertesti. Compatibilità informatica e letteratura medievale*; in particolare si segnala l'intervento di Lino Leonardi, *Tradizione poetica e dinamica testuale nella lirica italiana del Duecento: funzioni di un ipertesto*, proprio perché fondato su un'opera ad alto grado di formalizzazione quale è quella delle *Concordanze della Lingua Poetica Italiana delle Origini* di d'Arco Silvio Avalle.

²⁷ È doveroso ricordare che alcune regole di normalizzazione grafica erano già state indicate nei primi anni dell'informatizzazione dell'Opera, ma restando la principale destinazione degli esempi quella del Vocabolario a stampa al momento della creazione della banca dati queste si sono rivelate insufficienti anche per le innovazioni nel frattempo intervenute nei criteri di trascrizione.

²⁸ A maggior ragione sfuggenti in quanto si tratta per lo più di parole grammaticali che al momento non vengono sistematicamente lemmatizzate.

sentavano nei 1348 testi che costituivano il Corpus al momento di questa indagine. Di ogni tipo si indica il trattamento che è stato applicato nella prospettiva di mantenere un certo grado di coerenza pur nei limiti imposti dalle necessità di formalizzazione richieste dall'elaborazione elettronica.

1) Il rafforzamento consonantico in fonosintassi, del tipo *a ffare*, prodotto da congiunzioni e pronomi (*che, se, ma; chi* ecc. in tutte le loro varianti grafiche), da forme ossitone e da monosillabi tonici (*né, è, sì, vè*, ecc.), da preposizioni e da pronomi tonici (*a, da, di / de, tra / intra, oltre; tu*), nelle diverse edizioni viene rappresentato:

- (a) col punto alto unito alle due forme contigue: *e·lle, né·lle, tra·lloro, tu·tti, ...*;
- (b) col punto alto unito alla prima forma e separato da quella che segue: *e· lle, né· lle, tra· lloro, tu· tti, ...*;
- (c) col punto alto a carico della seconda forma: *e·lle, né·lle, tra·lloro, tu·tti, ...*;
- (d) senza nessun segno diacritico fra le due forme: *e lle, né lle, tra lloro, tu tti, ...*;

Poiché nell'incontro fra le due forme non c'è perdita di alcun elemento, ed anzi il fenomeno fonosintattico è già evidente nel raddoppiamento della consonante iniziale, si estenderà la soluzione (d) a tutti gli altri casi: *e lle, né lle, a ffare, tra lloro, tu tti*, ecc.

2) La scomparsa della consonante finale di una parola (*-n, -l, -r*, più raramente *-m*) assimilata nel raddoppiamento della consonante iniziale della parola seguente, con eventuale successivo scempiamento (del tipo *i lluogo* o *i ria, co lloro* o *co loro, de rre* o *de re, pe mmare* o *pe ragione*, ecc.), nelle edizioni viene rappresentata:

- (a) con la ricostruzione fra quadre della consonante scomparsa per assimilazione: *i[n] (r)ria; pe[r] (r)ragione, ...*;
- (b) col punto alto unito alle due forme contigue: *de·(r)re, i·ria, ...*;
- (c) col punto alto a carico della forma priva della consonante: *co· (l)loro, de· (r)re, ...*;
- (d) coll'apostrofo: *co' lloro, i' lloro, ...*;
- (e) col trattino unito alle due forme contigue²⁹: *i-lloro*.
- (f) senza alcun segno diacritico fra le due forma contigue: *cho lloro, i lloro*.

Nel trattamento per il Corpus si mantengono inalterate le soluzioni (a), (c) ed (f), mentre nei tipi (b), (d) e (e) si conserva il segno diacritico dell'EdR e lo si trascrive a carico della forma che ha perduto la consonante finale: *i· ria, co' lloro, i- lloro*.

²⁹ Si tratta in realtà di un caso isolato, quello del *Tristano Riccardiano*, a cura di Ernesto Giacomo Parodi, Bologna, Romagnoli-Dall'Acqua, 1896.

2.1) Il completo dileguo di alcune forme monosillabiche (*il / el / lo, in / en, ne, no*) investite da scempiamento consonantico in fonosintassi e, prima, da aferesi o apocope vocalica, del tipo *ò (m)me* per *ò 'n me*; *che (r)romito* per *che (el / lo) romito*, si trova rappresentato:

- (a) con la ricostruzione fra quadre dell'elemento consonantico dileguatosi accompagnato dal segno di aferesi o di elisione: *gli è ['n] piacere*;
- (b) col punto alto isolato preceduto dal segno di aferesi: *che '· romito*;
- (c) col punto alto isolato preceduto da trattino, ma solo nel caso di dileguo della *l*, articolo o pronome, per la quale il curatore ritenga di non poter stabilire la forma originaria (*el / il / lo*) e di conseguenza il fenomeno vocalico (aferesi o apocope) che l'ha interessata: *che -· rey*;
- (d) col punto alto unito alle due forme contigue: *che-ricievate (che 'l ricevate), fa-mantenente (per fa 'n mantenente)*;

Si rispettano le soluzioni (a), (b) e (c), mentre si interviene sul tipo (d) isolando il punto alto dalle due forme contigue e integrando il segno diacritico appropriato per la scomparsa dell'elemento vocalico : *che-ricievate > che '· ricevate, fa-mantenente > fa '· mantenente*.

3) Un caso particolare è quello dell'incontro della forma *no* dell'avverbio con parola iniziante per consonante. In assenza di specifiche indicazioni dell'editore e in considerazione dell'impossibilità di distinguere il fenomeno fonosintattico rappresentato – rafforzamento da *no*; assimilazione e scempiamento da *non* – si riproduce sempre l'edizione attribuendo il segno diacritico, qualunque esso sia, alla forma *no*. Pertanto in presenza delle rappresentazioni che seguono (si assume come esempio il caso di maggior frequenza, ossia quello in cui l'avverbio è seguito da alcune forme pronominali) si mantengono i tipi (a) e (b) e a quelli si adeguano i tipi (c) (d) ed (e), attribuendo cioè il segno diacritico a *no*:

- (a) senza segni diacritici: *no (m)me; no (t)te, no (l)l', no l(l)a, no l(l)e, ...*;
- (b) col punto alto a carico dell'avverbio: *no· (m)me; ..., no· (l)l', ...*;
- (c) col punto alto unito alle due parole contigue: *no·(m)me, ..., no·(l)l', no·(l)la, no·(l)le, ...*;
- (d) coll'apostrofo a carico della negazione: *no' ll'; no' li, ...*;
- (e) col trattino unito alle due parole contigue³⁰: *no-mmi, ..., no-ll', no-(l)lo, ...*

³⁰ Anche questo limitato al citato *Tristano Riccardiano* (cfr. sopra).

Nello specifico dell'incontro fra negazione e forme pronominali valgono però perlomeno due eccezioni. La prima riguarda la scrizione unita delle forme di *no / no(n) + i pron. di 3^a e 6^a pers. (noll', nolla, nolle, ecc.)* che viene riprodotta come si trova; la seconda interessa invece l'incontro con le forme elise dei pronomi clitici del tipo *m(-e,-i), n(e)*, ecc. che si trova rappresentato:

- (a) con il diacritico (punto alto o apostrofo) a carico del pronome: *no 'm, no 'n, no 'nd(e), ...*;
- (b) con l'apostrofo unito alle due forme: *e no'm; e no'm ...*;
- (c) col punto alto unito alle due forme contigue: *no·m, no·n, no·nd(e)...*

Si rispetta la soluzione maggioritaria (a), *no 'm*, e si interviene sui tipi (b) e (c) separando la negazione dal pronome e lasciando a carico di quest'ultimo il segno diacritico: *no 'm, no ·m, no ·n, no ·nd(e)*, ecc.

3.1) Ancora più resistente ad una soluzione univoca e unitaria è la diversa resa grafica della forma *nol* per la quale è decisamente consolidato l'uso della scrizione unita ma che nelle edizioni si trova:

- (a) in grafia unita: *nol*;
- (b) in grafia separata col diacritico (apostrofo o punto alto) unito alla *l* enclitica: *no 'l, no ·l*
- (c) col diacritico unito alle due forme contigue: *no'l, no·l, no-l*.

Ferma restando la rappresentazione (a) che ha più di duemila esempi nel Corpus, il tipo (c) viene trattato alla stregua del (b) ossia col segno diacritico a carico del pronome: *no 'l; no ·l; no -l*. Nell'interrogazione del Corpus si dovranno pertanto prevedere entrambe le soluzioni, quella "sintetica", {*nol*}, e quella "analitica", {*no*} + {*l/-l*}.

4) Analoga varietà di rappresentazioni si ha nell'incontro fra le preposizioni *con, en / in e per* con le forme dell'articolo inizianti per *l-*. Per queste si trovano le soluzioni:

- (a) in grafia unita: *colla, cola, col, pel, ...*; con accento diacritico nei casi di *èllo, èlle* (per 'nello', 'nelle'), ...;
- (b) col punto alto unito alle due forme contigue: *co·lla, co·le, e·llo, pe·l ...*;
- (c) con l'apostrofo al posto della finale consonantica della prep.: *co'la, pe'li, ...*;

- (d) in grafia separata senza alcun segno diacritico: *co la, co lla, pe lo, pe lla, ...*; ma col segno di aferesi in *co 'l, pe 'l*;
- (e) col punto alto a carico della prep.: *co· lla, co· le, pe· lle, ...*;
- (f) col trattino, ma limitatamente all'unione di *en / in* con la forma non ridotta dell'articolo: *en-ella*.

Diversamente dal caso dell'avverbio di negazione, per la preposizione seguita da articolo si è scelto di interpretare i tipi (b) ed (f) come scrizioni unite – e quindi col segno diacritico codificato come segno interno – per analogia col tipo (a) che è maggioritario nel Corpus; e di mantenere separate le forme dei tipi (c), (d) ed (e). Di conseguenza nell'interrogazione le grafie *colla* e *co·lla* si raccoglieranno con la richiesta {*colla*}, mentre negli altri casi, *co la* e *co'la*, sarà necessaria un'interrogazione per cooccorrenze {*co*} + {*la*}.

5) Un altro fenomeno variamente rappresentato è quello dell'agglutinazione fra forme verbali e pronomi atoni o particelle avverbiali, integre o ridotte, in posizione enclitica:

- (a) in grafia unita con eventuale accento sulla vocale tonica: *portallo, portâllo, pòrtallo; mandoli o mandòli, mandolli, ...*;
- (b) col punto alto unito alle due forme contigue: *prendo·lo, disse·i, conve·me, fa·mene, ...*;
- (c) col punto alto a carico del clitico: *tene·me*;
- (d) col trattino: *di'-mello, disé-mello, conté-lli, aver-lli, da'-lli, è-llo, ...*

Si uniformano i tipi (b) (c) e (d) al tipo (a), eliminando lo spazio nel caso (c) e codificando i segni diacritici come segni interni. In questo modo qualunque sia la rappresentazione grafica adottata dall'EdR (*pòrtallo, disse·i, disé-mello*) le forme verranno interpretate e ricercate dai programmi di interrogazione come un'unica catena grafica, alla pari della forma unita: {*portallo*} raccoglierà *portallo, portâllo, pòrtallo, porta-lo*; ecc.

6) L'ultimo caso di una certa consistenza riguarda la grafia unita o divisa di gruppi di pronomi atoni composti con *lo* e *ne* apocopati. Questi si trovano:

- a) in grafia unita: *gliel, men, mil, sen (= se ne), tel, ...*;
- b) con l'apostrofo fra i due pronomi contigui: *se'n*;
- c) col diacritico (punto alto o apostrofo) fra i due pronomi contigui: *glie-l, li-l, me·n, se·n, te·l*, ecc.

Si conserva la rappresentazione di tipo (a), codificata dall'uso, mentre si tengono separate col segno diacritico a carico della forma ridotta quelle di tipo (b) e (c). Anche in questo caso quindi l'interrogazione del Corpus sarà duplice: {*gliel*} raccoglierà *gliel* e *gliel'*, {*se*} + {*n*} raccoglierà invece *se 'n* e *se ·n*.

* * *

In chiusura di questa panoramica generale, e necessariamente generica, sui principali criteri di analisi, trattamento e codifica dei dati per la costituzione del *TLIO*, non possiamo non dichiararci consapevoli dell'opinabilità di ogni singola scelta fatta e del lavoro che ancora resta da svolgere perché si possa offrire un sistema di informazioni più scientificamente rigoroso; ma vogliamo ancora ricordare – a parziale giustificazione del lavoro svolto – la notevole massa dei dati (con tutto quello che si è già detto sulla loro eterogenea provenienza) da gestire, e soprattutto il diverso grado di elaborazione nel quale molti di questi sono arrivati al Corpus, dopo essere stati investiti da processi di trattamento che miravano ad una destinazione del tutto diversa da quella di una base di dati.

Resta che quello stesso strumento informatico che impone continue scelte di formalizzazione, dà anche la possibilità di correggere e integrare in ogni momento qualsiasi tipo di dati. Fatto di rilievo sostanziale per un'impresa come quella del *Vocabolario* che intende documentare l'evoluzione della lingua – nelle sue varianti diacroniche, diastratiche, diatopiche – e insieme rispecchiare l'evoluzione delle nostre conoscenze che lo stesso *Vocabolario* contribuisce a promuovere.