

# Metodi di allineamento testuale bilingue per un'edizione genetica digitale dei *Mémoires* di Carlo Goldoni

Matteo Zibardi<sup>1</sup>

<sup>1</sup> Sorbonne Université, CERES, Francia – matteo.zibardi@sorbonne-universite.fr

## ABSTRACT (ITALIANO)

Il presente contributo propone i primi risultati ottenuti nell'ambito della realizzazione di un'edizione genetica digitale dei *Mémoires* di Carlo Goldoni, prodotta grazie al ricorso agli strumenti di Trattamento Automatico della Lingua (TAL). Pubblicata a Parigi nel 1787, l'autobiografia è il frutto del rifacimento del testo di diciassette prefazioni, scritte in lingua toscana, poste in testa ai tomi dell'edizione Pasquali delle opere teatrali dell'autore editi tra il 1761 e il 1778. La genesi bilingue dell'opera è stata però ignorata dalla critica goldoniana. Il presente contributo intende colmare tale aporia proponendo delle modalità operative per la realizzazione di un'edizione genetica digitale dei *Mémoires* di Carlo Goldoni. Dal momento che il testo francese è frutto di un profondo rifacimento di quello italiano, un semplice calcolo di similarità coseno tra le frasi delle due versioni non è stato sufficiente. Dunque, sono state estratte ed allineate le frasi contenenti entità nominate condivise dai due stadi testuali. L'allineamento testuale bilingue a partire dalle entità nominate si è rivelato di gran lunga più efficace e ha permesso di individuare molte porzioni comuni tra i due testi, nonostante la diversità linguistica. È stata così creata la prima versione di un'edizione sinottica che sarà presto realizzata nella sua interezza.

**Parole chiave:** edizione scientifica digitale; Named Entity Recognition; allineamento testuale bilingue; critica genetica; Carlo Goldoni.

## ABSTRACT (ENGLISH)

*Methods of bilingual text alignment for a genetic digital edition of Goldoni's Mémoires.* This contribution proposes preliminary results obtained while realizing a digital genetic edition of Carlo Goldoni's *Mémoires*. This edition has been produced through the use of Natural Language Processing (NLP) tools. Goldoni's work represents a significant case study in the field of genetic criticism: published in Paris in 1787, it is the result of the rewriting of seventeen prefaces, written in Tuscan, placed at the beginning of the Pasquali edition of the author's plays published between 1761 and 1778. However, the bilingual genesis of this work has been neglected in Goldoni's critic bibliography. This contribution aims to describe the methodology employed to realize a digital genetic edition of the *Mémoires*. Assuming that the French texts has not only been translated from the Italian one, but also modified in some of its parts, the similarity score between the sentences of the two version was not a reliable result. To find a better solution, the sentences in both source and target text containing equivalent named entities have been aligned. The bilingual text alignment based on the recognition of the same named entities has proved itself to be more efficient and permitted to detect a lot of common blocks between the two versions, despite linguistic and textual differences. So, the first draft of a synoptic genetic edition of the *Mémoires* has thus been created.

**Keywords:** digital scholarly editing; Named Entity Recognition; bilingual textual alignment; genetic criticism; Carlo Goldoni

## 1. INTRODUZIONE

I *Mémoires de M. Goldoni pour servir à l'histoire de sa vie et à celle de son théâtre* sono un caso di studio altamente significativo per la ricostruzione della vita teatrale europea del XVIII secolo e per l'analisi delle pratiche di scrittura autobiografica. L'opera, pubblicata in lingua francese a Parigi nel 1787, contiene un racconto retrospettivo ed in gran parte apologetico della vita del drammaturgo Carlo Goldoni, trasferitosi nella capitale francese nel 1762. Oltre all'indiscutibile valore documentario, in ottica genetica i *Mémoires* si configurano come il risultato di un «macrotesto autobiografico» (Anglani, 1995: 325), dal momento che Goldoni aveva già iniziato a raccontare della sua vita, in lingua toscana, all'interno del paratesto dell'edizione delle sue opere teatrali: dapprima nella prefazione all'edizione Bettinelli stampata a Venezia nel 1750, e poi nelle diciassette prefazioni in testa ai rispettivi tomi dell'edizione Pasquali, stampata sempre a Venezia dal 1761 al 1778.

Per la stesura dei *Mémoires*, Goldoni riprende gran parte del materiale precedentemente scritto in toscano traducendolo in francese, ma cambia i particolari di alcuni passaggi, elimina delle sequenze e ne aggiunge altre. Ne deriva dunque un'opera che possiede una propria autonomia strutturale e contenutistica, ma che risulta essere allo stesso tempo inscindibile dal proprio avantesto in lingua toscana. Per tale configurazione composita, i *Mémoires* sono stati definiti un'«autobiografia linguistica» (Oppici, 2016: 70), mettendone in luce la genesi plurilingue che richiede di essere analizzata attraverso le categorie della critica genetica. Tuttavia, una ricognizione della bibliografia critica goldoniana mostra una lacuna per quanto riguarda lo studio del plurilinguismo dell'autore, che si può trovare solo in contributi ormai datati (Folena, 1954; Folena, 1984), oppure di breve ampiezza, come la voce «Langue» redatta da D'Onghia (D'Onghia, 2019) e recentemente pubblicata all'interno del *Dictionnaire Goldoni*. Tali riferimenti critici non approfondiscono il rapporto tra il racconto autobiografico e l'impiego di una lingua seconda. Inoltre, gli studi più importanti dedicati ai *Mémoires* (Vianello, 1991; Alonge, 2009; Ferrone, 2011) citano marginalmente le prefazioni Bettinelli e Pasquali, senza considerare l'opera nella totalità della sua genesi testuale plurilingue. Solamente l'edizione Mondadori dei *Mémoires* (Bosisio, 1993) ha colto l'importanza del «macrotesto autobiografico» goldoniano, pubblicando in un unico volume il testo francese accanto alle prefazioni Bettinelli e Pasquali. Tuttavia, la mancanza di un approfondimento critico relativo al rapporto tra le differenti versioni e l'importanza gerarchica affidata ai *Mémoires*, pubblicati per primi nonostante siano l'ultima redazione del testo, intaccano l'eccellente intuizione editoriale e filologica. A partire da tale aporia critica, il presente contributo intende suggerire delle modalità operative per la realizzazione di un'edizione genetica dei *Mémoires*, supportata dagli strumenti offerti dal Trattamento Automatico della Lingua (TAL). Nello specifico, si dimostrerà come sia possibile allineare i blocchi di testo comuni tra le prefazioni Pasquali e il testo dei *Mémoires*, nonostante la diversità linguistica. Un'edizione sinottica così prodotta permetterà di ottenere uno strumento che favorirà lo studio del rapporto tra le due redazioni, colmando così la mancanza presente negli studi goldoniani. Da questo punto di vista, la critica genetica assume un'importanza fondamentale (Sciarrino & Anokhina, 2018), dal momento che la vicinanza al testo consentirà di analizzare in maniera approfondita il plurilinguismo di Goldoni e, in generale, degli autori bilingui o plurilingui. Nel seguito del contributo verrà descritta la metodologia impiegata che ha permesso di muovere i primi passi nella realizzazione di una vera e propria edizione genetica digitale che vedrà prossimamente la luce. Nonostante i risultati siano ancora parziali, si ritiene comunque opportuno di condividerli con la comunità scientifica per stimolare un dibattito fecondo sulla creazione di edizioni genetiche digitali, supportate dal ricorso agli strumenti di TAL.

## 2. REALIZZARE UN'EDIZIONE GENETICA DIGITALE BILINGUE

Il testo dell'edizione francese dei *Mémoires* è stato ricavato seguendo scrupolosamente l'*editio princeps* in tre tomi pubblicata a Parigi nel 1787. Per ragioni di compatibilità con i modelli linguistici presenti attualmente, il testo è stato normalizzato in francese contemporaneo attraverso alcuni cambiamenti relativi alla morfologia e all'accentazione di alcune parole. Diverso è invece il caso dell'avantesto. Mancando dei modelli allenati su dati relativi alla varietà diacronica dell'italiano del XVIII secolo, il testo delle prefazioni Pasquali non è stato modificato ed è stato ricavato a partire dall'edizione critica realizzata da Paolo Bosisio e precedentemente citata. Il procedimento prevedeva dunque l'utilizzo di due file .txt: il primo contenente in modo continuo il testo delle diciassette prefazioni Pasquali, mentre il secondo il testo normalizzato dei *Mémoires*.

Per quanto riguarda l'allineamento testuale plurilingue, si è pensato in un primo momento calcolare la similarità tra le frasi del testo italiano con quelle del testo francese utilizzando dei modelli di *text similarity*,<sup>1</sup> che hanno tuttavia offerto dei risultati poco soddisfacenti. Il procedimento è stato semplice: per ogni frase del testo italiano sono stati individuati i tre migliori possibili candidati per l'allineamento che presentassero una similarità coseno maggiore rispetto ad una determinata soglia fissata a 0.4. Il valore basso della soglia, scelto per garantire la possibilità di trovare blocchi di testo comuni anche in presenza di traduzione e riformulazioni, ha però creato un numero molto alto di falsi positivi che avrebbe richiesto uno spoglio manuale estremamente lungo per verificare la correttezza degli allineamenti. In particolare, la

---

<sup>1</sup> Quello che ha fornito le prestazioni migliori e che è stato poi tenuto in considerazione per i successivi calcoli di similarità è LaBSE, la cui documentazione è fornita all'indirizzo seguente: <https://huggingface.co/sentence-transformers/LaBSE> (cons. 10/01/2025)

fascia critica è rappresentata dall'intervallo di frasi che hanno una similarità tra 0.50 e 0.65, per cui vengono individuati nel testo francese sia falsi positivi, che frasi soggette ad una parziale riscrittura ma che condividono comunque l'informazione del testo di partenza. In alcuni casi, invece, la frase corrispondente, che pure esiste, non viene per nulla riconosciuta, come nel caso seguente:

```
"l'imer, che pensava a sostenere gl'intermezzi, e temea dell'incontro della passalacqua, fatto avea un altro acquisto.": [
  {
    "target": "mais imer croyait que l'intermède soutenait la tragi-comédie, et c'était celle-ci qui soutenait l'intermède.",
    "score": 0.5087771415710449
  },
  {
    "target": "le patron avait là des affaires, il s'y arrêta et mit pied-à-terre, je crus le moment favorable pour m'en aller;",
    "score": 0.44995665550231934
  },
  {
    "target": "je redoutais si fort ce moment terrible, qu'en sortant d'une faute j'en méditais une autre qui pouvait achever ma perte.",
    "score": 0.43072694540023804
  }
],
```

Tralasciando i limiti dovuti alla mancanza di un modello linguistico performante per i dati relativi all'italiano del XVIII secolo, l'imprecisione negli allineamenti suggerisce anche la particolarità del corpus preso in analisi. Essendo infatti presenti moltissime libere rielaborazioni di concetti e situazioni presenti nel testo italiano, i *Mémoires* non possono essere considerati alla stregua di una semplice traduzione e ampliamento dell'avantesto. Il corpus solleva di conseguenza anche delle problematiche a livello filologico, essendo molto difficile trovare una definizione di blocco testuale comune laddove a modificare il testo di partenza intervengono sia una traduzione che una riformulazione. Inoltre, è poco frequente la corrispondenza tra singole frasi e gli stessi concetti possono essere espressi da più periodi nel testo italiano, mentre da una sola frase nel testo francese, e viceversa. Come indicazione di metodo, è dunque necessario andare alla ricerca di punti di ancoraggio all'interno delle due versioni, piuttosto che perseguire uno scrupolo filologico volto al riconoscimento di varianti relative a singoli sintagmi o espressioni.

Da questo punto di vista, la bibliografia specifica relativa al tema dell'allineamento testuale non offre particolari spunti risolutivi, dal momento che la maggior parte dei test condotti dalle autrici e dagli autori ha come obiettivo l'allineamento di corpora paralleli, vale a dire costruiti a partire da un testo sorgente e dalla sua rispettiva traduzione (Zweigenbaum & al, 2017; Repar & al, 2021; Malandri & al, 2024; Molfese & al, 2024; Reboul, 2024,<sup>2</sup> Xu & al, 2024). È stato dunque necessario elaborare un altro tipo di metodologia per approcciare lo studio sinottico dei due testi. Una possibile soluzione è stata suggerita dalle caratteristiche stesse del corpus. Visto lo spiccato carattere odepórico di entrambe le versioni e constatata la presenza di numerose informazioni sugli ambienti culturali e politici dell'epoca trasmesse da Goldoni, è parso opportuno considerare tali aspetti per imbastire l'operazione di allineamento bilingue. L'intuizione si basa sull'assunto che, per quanto diversi nella forma, le due versioni condividono in ogni caso gran parte della materia autobiografica che tocca necessariamente il racconto delle persone incontrate e dei luoghi visitati. Sarà così altamente significativo per il confronto tra i due testi verificare l'aggiunta, l'eliminazione o la conservazione delle informazioni nel passaggio da un'edizione all'altra.

Si è dunque deciso di ricavare automaticamente in entrambi i testi tutte le entità nominate, vale a dire i nomi di luoghi, persone, organizzazioni etc. per utilizzarle come punti di ancoraggio nell'allineamento

---

<sup>2</sup> Questo interessante contributo affronta la questione dell'allineamento bilingue utilizzando anch'esso le entità nominate come base per il riconoscimento di sequenze comuni tra il testo dell'Odissea e alcune sue traduzioni francesi. Tuttavia, per quanto possa presentare tagli e aggiunte circoscritte, il corpus scelto rientra sempre nell'ordine della traduzione. Il metodo, dunque, non può essere integralmente replicato per il confronto dei *Mémoires* con il proprio avantesto toscano, viste le numerose inversioni dell'ordine degli elementi citati e le peculiarità specifiche del corpus goldoniano.

bilingue.<sup>3</sup> Dopo la loro estrazione, è stato necessario filtrare i risultati dall'analisi automatica dal momento che, come indicato dalla documentazione ufficiale (Zaratiana & al, 2024) i modelli di riconoscimento di entità nominate utilizzati non possiedono una precisione del 100%, raggiungendo comunque ottimi livelli di approssimazione. Inoltre, le diverse denominazioni di una medesima entità nominata all'interno dello stesso testo sono state regolarizzate assegnando a ciascun elemento un id unico. A questo punto, per ogni entità nominata riconosciuta nel testo italiano è stata individuata l'entità nominata corrispondente nel testo francese sulla base di un calcolo di similarità coseno utilizzando LaBSE. In questa fase, tutte le lettere maiuscole sono state rese minuscole, dal momento che Goldoni, secondo una prassi tipica dell'epoca, utilizza le maiuscole per designare anche alcuni nomi comuni. Un processo di normalizzazione di questo tipo permette infatti al modello di evitare delle possibili ambiguità o allucinazioni. Tale metodologia risulta essere estremamente funzionale essendo capace di ridurre al minimo i limiti derivanti dalla mancanza di modelli specifici allenati per il trattamento di dati testuali provenienti da testi italiani del XVIII secolo. Infatti, le entità nominate sono molto più facili da accoppiare tra di loro, dal momento che nel testo dei *Mémoires* non si trovano delle vere e proprie traduzioni, quanto più degli adattamenti al sistema linguistico francese che, in ogni caso, non si discosta troppo dall'originale e permette di ottenere dei risultati affidabili.

È stato così possibile accoppiare delle liste di frasi "source" per il testo italiano e "target" per il testo francese contenenti la medesima entità nominata. Per affinare ancora di più la precisione dell'allineamento, i cluster di frasi appena creati sono stati ulteriormente filtrati con la creazione di altri cluster contenenti al proprio interno due o più entità nominate. Il procedimento è molto semplice, dal momento che se una frase del testo italiano o francese è presente in più cluster, significa necessariamente che contiene al proprio interno più entità nominate. Tale procedimento si è reso necessario per filtrare le frasi che contengono al proprio interno nomi di luogo o di persona citati frequentemente, e che costituivano dunque cluster numericamente importanti. Considerando poi solo i cluster che presentavano almeno una frase nella lista "source" e una nella lista "target", sono stati calcolati gli allineamenti migliori sulla base della similarità coseno tra gli embeddings delle frasi. Ciò ha permesso di risolvere degli errori provenienti dal semplice calcolo della similarità condotto in precedenza e minimizzare le tempistiche del controllo umano per verificare la validità degli allineamenti:

```
"passalacqua (passalacqua), imer (imer)": [
  {
    "source": [
      "e la terza un'operetta in musica in sei personaggi, intitolata la fondazione di venezia, in cui cantavano l'imer, l'agnese, la passalacqua, il gandini brighella, il campagnani arlecchino, ed il mio casali cantovvi anch'egli, e si fece onore.",
      "l'imer, che pensava a sostenere gli'intermezzi, e temea dell'incontro della passalacqua, fatto avea un altro acquisto."
    ],
    "target": [
      "imer la trouva à son gré, me pria d'en prendre soin, et je m'en chargeai avec plaisir, la trouvant très jolie et très docile.\n madame passalacqua en devint jalouse;",
      "mais il n'était pas trop content de madame passalacqua: sa voix était fausse, sa manière était monotone, et sa physionomie grimacière: imer voulait soutenir les intermèdes, et un musicien de l'orchestre lui en proposa le moyen."
    ]
  }
],
```

Calcolando così la similarità all'interno dei cluster creati e confrontando i risultati con quelli derivati dall'analisi della semplice similarità tra le frasi del testo italiano e francese, è possibile notare un notevole incremento prestazionale nel momento in cui l'allineamento viene effettuato a partire dall'estrazione di frasi contenenti entità nominate. Infatti, a fronte di 54 blocchi di testo riconosciuti da entrambi i metodi, il calcolo di similarità a partire da entità nominate ha permesso di individuare ulteriori 86 porzioni testuali

---

<sup>3</sup> Per l'estrazione delle entità nominate è stato utilizzato il modello GLiNER che rappresenta in questo momento lo stato dell'arte per il compito richiesto. In particolare, `urchade/gliner_multi-v2.1` per il francese: [https://huggingface.co/urchade/gliner\\_multi-v2.1](https://huggingface.co/urchade/gliner_multi-v2.1) (cons. 10/01/2025) e `DeepMount00/GLiNER_ITA_LARGE` per l'italiano: [https://huggingface.co/DeepMount00/GLiNER\\_ITA\\_LARGE](https://huggingface.co/DeepMount00/GLiNER_ITA_LARGE) (cons. 10/01/2025).

comuni, a fronte delle 14 del semplice calcolo di similarità coseno. Dunque, all'interno dell'allineamento testuale bilingue, l'estrazione di blocchi testuali comuni a partire da entità nominate equivalenti si rivela più efficiente del semplice calcolo di similarità coseno in oltre il doppio dei casi. Per visualizzare i risultati, è stato creato un file .html che presenta una finestra bipartita contenente a sinistra il testo integrale delle prefazioni Pasquali e a destra quello dei *Mémoires* (Fig. 1), rispettando così l'ordine cronologico di redazione. Sono stati evidenziati in giallo le frasi individuate da entrambi i metodi, in rosso quelle esclusive del calcolo similarità e in verde quelle presenti nei cluster creati. Ciò permette di meglio comprendere il funzionamento del procedimento e, attraverso un'attenta analisi dei risultati, di migliorare in seguito il procedimento. Inoltre, cliccando su un blocco comune nella versione francese, oppure in quella italiana, è possibile allineare i due testi e studiare i blocchi comuni all'interno del loro contesto originale. Tale visualizzazione bipartita rispetta l'autonomia espressiva delle singole redazioni, dal momento che il testo italiano e quello francese possono essere letti autonomamente, mettendo in luce allo stesso tempo il legame che tra di esse intercorre.

merito, aveva sacrificato il buon senso al cattivo uso de' comici, e m'invogliai sempre più a rinnovar la griselda. la scrissi in verso, seguitai in gran parte la traccia del primo autore, cangiai qualche scena, e ne aggiunsi a mio capriccio, e la ridussi in istato di ricomparir come nuova. fra gli altri cambiamenti ne feci uno, che diede il maggior merito alla novità. premevami il mio casali. immaginai d'introdurre il padre di griselda, ch'era nata fra boschi: un buon vecchio, tenero, prudente, discreto, che non insuperbisce veggendo la figliuola sul trono, e non si rattrista veggendola ricadere nell'antica sua povertà, e prende parte soltanto all'offesa dell'onore e dell'innocenza. questo vecchio piacque infinitamente, e tutta la tragedia ha piaciuto, ed il pubblico rese a me questi onori, che dovevansi in parte all'autor primiero. per contentare gli attori degl'intermezzi, ne ho composto uno in due parti ed uno in tre, e terminata la piazza di padova, cioè le recite della primavera, la compagnia passò a udine per trattenervisi tutto l'estate. io mi vi resi egualmente, attirato da più motivi, di cui non era l'ultimo la ferramonti. desiderava altresì di riveder quel paese, dove vissuto aveva parecchi mesi, dove avea molti amici e dove mi lusingava di rivedere (per semplice curiosità) qualche oggetto delle prime mie

plus aise pour moi. je venais de m'occuper de ce même sujet à Venise, et la griselda de pariat n'était pas autre chose que l'opéra qu'il avait composé lui-même, en société avec apostolo zeno. j'entrepris, avec plaisir, de contenter la romana; mais je ne suivis pas exactement les auteurs du drame; je fis beaucoup de changements; j'y ajoutai le père de griselda: un père vertueux qui avait vu sans orgueil monter sa fille au trône, et la voyait descendre sans regret. j'avais imaginé ce nouveau personnage pour donner un rôle à mon ami casali: cette épisode donna un air de nouveauté à la tragédie, la rendit plus intéressante, et me fit passer pour auteur de la pièce. dans l'édition de mes œuvres faite à turin en 1777, par guibert et orgeas, cette griselda se trouve imprimée comme une pièce à moi appartenante: je déteste les plagiat, et je déclare que je n'en suis pas l'inventeur. mes comédiens avoient rempli, à padoue, le nombre des représentations convenues, et ils faisaient leurs paquets pour aller à udine, dans le frioul vénitien. imer me proposa de m'y emmener avec lui, je n'avais plus rien à craindre du côté de la limonadière, qui s'était mariée; je consentis de suivre la compagnie; mais ce ne fut pas avec le directeur que je voyageai. je lui fis mes excuses, et je partis dans une bonne voiture, avec madame ferramonti et le bonhomme son mari. à udine, mes ouvrages furent très applaudis: j'avais, dans cette

**Figura1. Esempio di blocchi testuali comuni nell'edizione sinottica dei *Mémoires*.**

La codifica non stabilisce una gerarchia tra il testo dei *Mémoires* e quello delle prefazioni Pasquali. Le porzioni di testo sono inserite all'interno di tag <span> con un id alfanumerico contenente la sigla della lingua e un numero, una classe per mettere distinguere i vari colori e l'evento onclick che grazie ad una semplice funzione chiamata alignText permette di allineare i due blocchi testuali. Di seguito due esempi completi presenti all'interno del file:

```
<span id="it-29" class="highlight-green" onclick="alignText('fr-29', 'it-29')">predicava dall'insigne pulpito del duomo di detta città il padre jacopo cataneo, agostiniano scalzo milanese.</span>
```

```
<span id="fr-29" class="highlight-green" onclick="alignText('fr-29', 'it-29')">je vais le jour des cendres à la cathédrale pour entendre le père cattaneo, augustin réformé, et je trouve son sermon admirable;</span>
```

Per questo motivo, un'eventuale codifica in formato XML in cui siano confrontati i due testi non dovrà seguire le linee guida della TEI per quanto riguarda l'apparato critico e la rappresentazione delle varianti,<sup>4</sup> dal momento che quest'ultima segue un principio verticale di subordinazione rispetto ad un testo principale. Piuttosto, risulterebbe filologicamente più adeguato l'utilizzo dell'elemento <seg><sup>5</sup> per circondare il blocco testuale scelto, unitamente all'attributo @corresp,<sup>6</sup> che presenterà un valore grazie a cui sarà possibile associarlo alla propria sequenza parallela. La natura arbitraria dell'elemento è consona rispetto al carattere eterogeneo dei blocchi paralleli, che sono stati riscontrati spesso al livello della frase; non corrispondono in nessun caso a dei paragrafi comuni e nemmeno a delle porzioni più ampie. Per questo, è da escludere l'assegnazione dell'attributo @corresp a elementi come <p><sup>7</sup> e <div><sup>8</sup> che individuano al contrario lunghe sequenze di testo. Per quanto riguarda la visualizzazione dei risultati,

<sup>4</sup> [https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#index-body.1\\_div.13\\_div.4](https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#index-body.1_div.13_div.4) (cons. 03/04/2025).

<sup>5</sup> <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-seg.html> (cons. 04/04/2025).

<sup>6</sup> <https://tei-c.org/release/doc/tei-p5-doc/it/html/ref-att.global.linking.html> (cons. 04/04/2025).

<sup>7</sup> <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-p.html> (cons. 04/04/2025).

<sup>8</sup> <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html> (cons. 04/04/2025).

potrebbe inoltre essere esplorata la possibilità di utilizzare EVT 3.0.<sup>9</sup> Sarebbe tuttavia necessario implementare manualmente dei parametri specifici e personalizzati per l'allineamento dei segmenti e la messa in evidenza dei blocchi testuali comuni, dal momento che un numero molto elevato di edizioni genetiche digitali che utilizzano il software EVT si basano sulla logica dell'apparato critico e dei diversi testimoni. Al contrario, il corpus del «macrotesto autobiografico» goldoniano solleva delle problematiche filologiche che richiedono un altro tipo di codifica, più neutra e meno gerarchicamente definita.

### 3. CONCLUSIONI

È stata presentata in questo contributo la prima versione di un'edizione genetica digitale dei *Mémoires* di Carlo Goldoni, realizzata operando un allineamento testuale bilingue dei blocchi di testo comuni ottenuti grazie all'estrazione e al filtraggio delle entità nominate nei due testi componenti il corpus. La visualizzazione sinottica ottenuta permette già di iniziare a formulare alcune considerazioni critiche circa il rapporto stilistico ed espressivo che intercorre tra le due versioni. Resta ancora da verificare la possibilità di allineare le porzioni di testo comuni che non contengono entità nominate, in modo tale da creare uno strumento che possa poi essere riutilizzato per lo studio della genesi plurilingue di alcune opere e non solo. In ogni caso, si è dimostrato che attraverso un procedimento semplice, facilmente replicabile e che non richiede il *fine-tuning* di uno o più modelli specifici sia possibile allineare blocchi di testo anche molto diversi tra di loro, ma che condividono le stesse informazioni relative alle entità nominali. Infine, in ottica contrastiva, un'ulteriore implementazione potrebbe prevedere una visualizzazione differente per i blocchi testuali esclusivi dell'una o dell'altra redazione, in modo tale che l'utente possa agevolmente comprendere lo sviluppo diacronico del testo e il rapporto stilistico e contenutistico tra le due versioni. Una volta ottenuta un'edizione genetica digitale completa, verrà messo a libera disposizione della comunità scientifica uno strumento indispensabile per colmare una lacuna presente negli studi goldoniani. Più in generale, la condivisione della metodologia impiegata per la realizzazione tecnica dell'allineamento testuale bilingue permetterà anche ad altri specialisti di realizzare edizioni genetiche per lo studio di scrittori bilingui o plurilingui. Sono già attualmente disponibili in rete delle ottime edizioni genetiche digitali, che si limitano però all'analisi di opere la cui genesi interessa una sola lingua.<sup>10</sup> Nel solco di questi progetti già avviati, la futura edizione dei *Mémoires* goldoniani intende essere un punto di riferimento per la realizzazione di edizioni genetiche bilingui o plurilingui ricorrendo agli strumenti di trattamento automatico della lingua.

### BIBLIOGRAFIA

- Alonge, R. (2009). Goldoni Mémoires: A Parigi si vive meglio. *Problemi di Critica Goldoniana*. XVI: numero speciale: Terzo centenario della nascita di Carlo Goldoni e Secondo centenario della morte di Carlo Gozzi: tomo terzo, 2009, 1000–1012. <https://doi.org/10.1400/130551>
- Anglani, B. (1995). I «Mémoires»: Bilanci e prospettive. Alberti C. & Pizzamiglio G. (Eds.). *Carlo Goldoni 1793-1993: Atti del Convegno del bicentenario*, Venezia, 11-13 aprile 1994, 325–339.
- D'Onghia L. (2016) *Langue. Dictionnaire Goldoni*. Comparini L. & Fabiano A. (Eds) Parigi: Garnier. 109–114.
- Ferrone, S. (2011). *La vita e il teatro di Carlo Goldoni*. Venezia: Marsilio.
- Folena, G. (1958). L'esperienza linguistica di Carlo Goldoni. *Lettere Italiane*, X (1), 21–54.
- Folena, G. (1984). *L'italiano in Europa. Esperienze linguistiche del Settecento*. Torino: Einaudi.
- Goldoni, C. (1993). *Memorie*. Bosisio, P. & Ranzini, P. (Eds.). Milano: Mondadori.
- Malandri, L., Mercorio, F., Mezzanzanica, M., & Pallucchini, F. (2024). SeNSE: Embedding alignment via semantic anchors selection. *International Journal of Data Science and Analytics*. <https://doi.org/10.1007/s41060-024-00522-z>
- Molfese, F., Bejgu, A., Tedeschi, S., Conia, S., & Navigli, R. (2024). CroCoAlign: A Cross-Lingual, Context-Aware and Fully-Neural Sentence Alignment System for Long Texts. *Proceedings of the 18th*

<sup>9</sup> <https://iris.unito.it/handle/2318/1896873> (cons. 04/04/2025).

<sup>10</sup> Si pensi al lavoro sulle varianti de *Il nome della rosa*: <https://www.variantidellarosa.it> (cons. 10/01/2025) oppure alla sezione "Génétique" del portale *eBalzac*: <https://www.ebalzac.com/genetique> (cons. 10/01/2025).

- Conference of the European Chapter of the Association for Computational Linguistics. Graham, Y. & Purver, M. (Eds.). San Giuliano: Association for Computational Linguistics, 2209-2220.  
<https://aclanthology.org/2024.eacl-long.135/>
- Oppici, P. (2016). I Mémoires di Goldoni come autobiografia linguistica. Regards croisés France-Italie: Langues, écritures et cultures: Atti della giornata di studi organizzata dalle Università di Macerata e Clermont-Ferrand. Bisconti D. & Fabiani, D. (Eds.). Macerata: EUM, 69-80.
- Reboul, M. (2023). Étude semi-automatique des traductions d'Homère avec "Odysseus" in Notre Homère. Deloince-Louette, C. & Salha, A. (Eds.). Grenoble: UGA Éditions, 163-182.  
<https://doi.org/10.4000/books.ugaeditions.32951>
- Repar, A., Martinc, M., Ulčar, M., & Pollak, S. (2021). Word-embedding based bilingual terminology alignment. Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference. 5-7 July 2021. Kosem, I. et al. (Eds.). Brno: Lexical Computing CZ.  
<https://doi.org/10.5281/zenodo.5547982>
- Sciarrino, E., & Anokhina, O. (2018). Plurilinguisme littéraire: De la théorie à la genèse. Genesis, 46, 11-34. <https://doi.org/10.4000/genesis.2554>
- Vianello M. (1991). L'avvocato in commedia, Goldoni e l'autobiografia. Studi Veneziani, 22, 325-358.
- Xu, Y., Hu, L., Zhao, J., Qiu, Z., XU, K., Ye, Y., & Gu, H. (2024). A Survey on Multilingual Large Language Models: Corpora, Alignment, and Bias. Frontiers of Computer Science.  
<https://doi.org/10.1007/s11704-024-40579-4>
- Zaratiana, U., Tomeh, N., Holat, P., & Charnois, T. (2024). GLiNER: Generalist Model for Named Entity Recognition using Bidirectional Transformer. Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Duh, K., Gomez, H., & Bethard, S. (Eds.). Città del Messico: Association for Computational Linguistics, 5364-5376 <https://doi.org/10.18653/v1/2024.naacl-long.300>
- Zweigenbaum, P., Sharoff, S., & Rapp, R. (2017). Overview of the Second BUCC Shared Task: Spotting Parallel Sentences in Comparable Corpora. Proceedings of the 10th Workshop on Building and Using Comparable Corpora. Sharof, S., Zweigenbaum, P., Rapp, P. (Eds.) Vancouver: Association for Computational Linguistic, 60-67. <https://doi.org/10.18653/v1/W17-2512>