Il progetto Corr(si)Ca: edizione digitale della corrispondenza Canioni

Giaufret Anna, Beatrice Dal Bo, Elena Margherita Vercelli, Laura Bonanno

Il progetto Corr<si>Ca: edizione digitale della corrispondenza Canioni

Anna Giaufret[1], Beatrice Dal Bo[2], Elena Margherita Vercelli[3], Laura Bonanno[4]

[1] Dipartimento di Lingue e Culture Moderne, Dottorato in Digital Humanities, Università di Genova, Italia - anna.giaufret@unige.it

[2] CLESTHIA (EA 7345), Sorbonne Nouvelle, France - beatrice.dal-bo@sorbonne-nouvelle.fr

[3] Dipartimento di Lingue e Culture Moderne, Dottorato in Digital Humanities, Università di Genova, Italia in co-tutela con CY Cergy Paris Université, France - elena.margherita.vercelli@edu.unige.it

[4] Dipartimento di Lingue e Letterature Straniere e Culture Moderne, Dottorato in Digital Humanities, Università di Torino, Italia in co-tutela con l'Université Jean Moulin Lyon 3, France - laura.bonanno@unito.it

ABSTRACT[1]

Il progetto Corr<si>Ca, portato avanti da docenti e dottorande/i del dottorato in Digital Humanities dell'Università di Genova, consiste nella digitalizzazione di una corrispondenza di circa 300 lettere, quella della famiglia Canioni, originaria di Olmi-Cappella, un villaggio dell'entroterra nel nord dell'isola. La corrispondenza si sviluppa nell'arco temporale 1881-1918. Gli scriventi, uomini e donne, presentano diversi gradi di alfabetizzazione nelle due lingue del carteggio, l'italiano e il francese. Il presente contributo presenterà il progetto, non ancora concluso, e i suoi obiettivi, concentrandosi sulla trascrizione delle lettere e sul protocollo XML-TEI utilizzato.

PAROLE CHIAVE

Edizione diplomatica digitale; XML-TEI; corrispondenza; scriventi semicolti; Corsica.

1. INTRODUZIONE: IL CARTEGGIO CANIONI

Il carteggio della famiglia Canioni, l'accesso al quale ci è stato permesso dai discendenti della famiglia stessa, rappresenta una corrispondenza familiare a cavallo tra Ottocento e Novecento, più precisamente tra il 1881 e il 1918. Seppure la corrispondenza si componga di un numero più elevato di lettere, il gruppo di ricerca ha deciso di stabilire una frontiera cronologica – almeno per la prima fase del lavoro – che coincide con un momento chiave della storia europea: la fine della Prima guerra mondiale.

La corrispondenza oggetto di questo studio si compone quindi di 270 lettere, scambiate dai membri e dall'entourage della famiglia Canioni, il cui nucleo centrale si trova in un villaggio della *Haute Corse*, Olmi-Cappella, situato a circa 800 metri di altitudine nella valle del Ghjunsani. Il principale destinatario delle lettere, il figlio maggiore (che è anche colui che ne ha conservato il maggior numero), si trova invece sul continente, vicino a Marsiglia. Il carteggio riflette infatti la lingua parlata sulle due sponde del Mediterraneo da tre generazioni di Canioni, uomini, ma anche donne, appartenenti alla categoria degli scriventi semicolti, con diversi gradi di alfabetizzazione. Inoltre, le lettere sono scritte nel momento in cui la popolazione còrsa passa dall'uso dell'italiano a quello del francese come lingua della scrittura: questo passaggio è visibile nel carteggio.

L'oggetto di studio presenta dunque un interesse linguistico di rilievo, perché ci permette di accedere a dati importanti che riguardano le competenze linguistiche scritte dei semicolti e delle donne al tempo della corrispondenza e la scrittura in zona di diglossia o triglossia. Inoltre, il carteggio costituisce un'importante testimonianza storica: le lettere ci forniscono infatti preziose informazioni sulla vita quotidiana, sulla cultura materiale, sull'economia e la politica locale e su molto altro. Infine, 8 lettere sono inviate dal fronte durante il periodo bellico dal nipote Canioni, il giovane Léon, a cui si aggiungono quelle dei suoi cugini Christophe e Xavier (una quindicina circa): questo gruppo di missive fornisce importanti informazioni sullo stato d'animo dei combattenti e sul rapporto dei Còrsi con la politica nazionale [12].

[1] Le autrici hanno contribuito rispettivamente alle seguenti sezioni: A. Giaufret (1. Introduzione, 2. Il gruppo e il progetto di ricerca, 3. Metodologia e trascrizione, 4. Particolarità e interesse linguistico del corpus), B. Dal Bo (3. Metodologia e trascrizione, 4. Particolarità e interesse linguistico del corpus), E. M. Vercelli (3. Metodologia e trascrizione, 5. Stato di avanzamento e risultati attesi, Bibliografia), L. Bonanno (2. Il gruppo e il progetto di ricerca, 3. Metodologia e trascrizione, 5. Stato di avanzamento e risultati attesi).

La geolocalizzazione di mittenti e destinatari delle lettere permetterebbe di acquisire informazioni sull'ampiezza della rete dei corrispondenti, che potrebbero a loro volta supportare riflessioni sulla diaspora còrsa e sul contatto con parlanti che dispongono di repertori linguistici diversi.
Di seguito viene riportata una rappresentazione grafica dell'albero genealogico della famiglia Canioni (Figura 1. In rosso i principali scriventi).

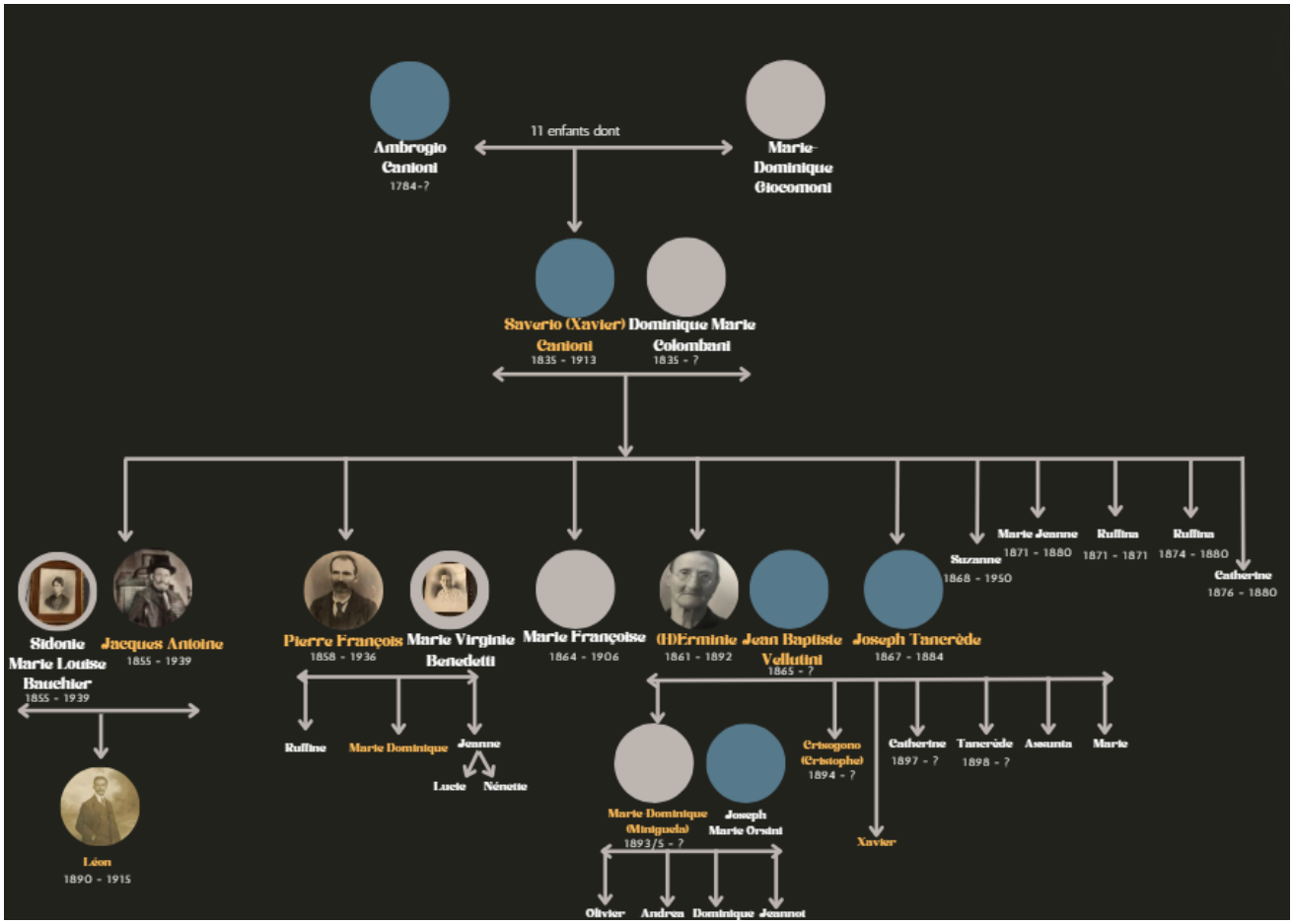

Figura 1: Albero genealogico della famiglia Canioni (in arancione i principali scriventi)

2. IL GRUPPO E IL PROGETTO DI RICERCA

Il progetto ha iniziato a svilupparsi su impulso di Anna Giaufret e Beatrice Dal Bo, prendendo le mosse dal lavoro di ricerca svolto nell'ambito del progetto Corpus 14[2] e nella tesi di quest'ultima, in cui si analizza una corrispondenza di scriventi semicolti durante la Prima guerra mondiale [6, 16]. Sulla base di questa esperienza ed essendo entrate in possesso degli originali del carteggio Canioni, Giaufret e Dal Bo hanno dato inizio al progetto Corr<si>Ca, ovvero Correspondance <numérique> Canioni, e hanno coinvolto le dottorande e i dottorandi in Digital Humanities delle Università di Genova e Torino, ognuna/o delle/dei quali mette al servizio del progetto le proprie competenze. Mentre altri sottogruppi si dedicano all'elaborazione di un'ontologia e alla realizzazione della parte divulgativa del progetto, in una prospettiva di Open Science (si veda a questo proposito la proposta Pasciuto *et al.*, "The Corr<si>Ca Project: enhancing and "querying" the Canioni family correspondence"), il gruppo composto da Giaufret, Dal Bo, Bonanno, Vercelli, Mantegazza, Bergaglio[3] si concentra sulla digitalizzazione delle lettere e la trascrizione in XML-TEI.
Il gruppo di ricerca è inoltre in contatto con il *laboratoire M3C* dell'Università di Corsica Pasquale Paoli[4], al fine di rendere fruibile la corrispondenza e i risultati del progetto alla comunità dei Còrsi.

---

[2] Gruppo di ricerca che ha realizzato la digitalizzazione di un grande corpus di lettere di "Poilus ordinaires" francesi [16, 20] (https://hdl.handle.net/11403/corpus14). (Ultimo accesso: 18 gennaio 2024).
[3] La trascrizione delle prime 100 lettere è stata realizzata da Giulia Balestrello e Alessia Rebecchi nell'ambito della loro tesi di laurea magistrale.
[4] M3C - Médiathèque Culturelle de la Corse et des Corses: https://m3c.universita.corsica/s/fr/page/home. (Ultimo accesso: 18 gennaio 2024).

3. METODOLOGIA E TRASCRIZIONE

Non è possibile tracciare nell'ambito di questa presentazione uno stato dell'arte esaustivo dell'edizione digitale (particolarmente fiorente in ambito medievista, di cui è un esempio l'esperienza del codice Pelavicino [18]). In questo intervento, ci limiteremo a una rapida panoramica degli studi sulle corrispondenze [9, 10, 14] e sulla digitalizzazione delle stesse secondo lo standard XML-TEI nell'ambito di progetti di ricerca preesistenti, come il progetto Bellini [7].

Secondo Walter et al. [24: 4-5][5], la corrispondenza è un "genere proteiforme, reticolare ed ellittico", caratteristiche che lo rendono particolarmente interessante, ma anche complesso da trattare. Per quanto riguarda la costituzione del corpus del progetto Corr<si>ca, sono state scartate dal carteggio donato dai discendenti della famiglia soltanto le lettere illeggibili (all'interno dell'arco cronologico prescelto). Il carteggio è inoltre accompagnato da materiale fotografico e da altri documenti, quale il libretto militare di Léon Canioni e alcuni taccuini del padre di quest'ultimo, Jacques Antoine, che saranno l'oggetto di una fase successiva del lavoro.

Come osserva Tomasi [23: 131], la "trascrizione delle fonti primarie è una forma di intervento interpretativo" in cui si selezionano e si identificano gli elementi più adatti per rappresentare al meglio la fonte in oggetto. Lo standard XML si rivela particolarmente adeguato a questo compito, in quanto permette ai ricercatori di elaborare documenti *well-formed*, la cui codifica riflette le intenzioni soggettive dell'interprete. Per quanto riguarda gli obiettivi del progetto Corr<si>Ca, la volontà è quella di mettere a disposizione della comunità dei ricercatori e della società in generale un corpus di lettere consultabile per ricerche di tipo linguistico innanzitutto, ma anche storico, filologico, ecc. Per svolgere questo lavoro in modo conforme allo standard di codifica di testi digitali, utilizziamo lo schema di validazione della TEI[6], che permette la validazione dei documenti, che sono dunque non soltanto "ben formati", ma anche "validi" (*valid*), ossia rispondenti agli standard previsti dal consorzio [3]. Partendo da tali obiettivi e rispettando i principi FAIR (*Findability, Accessibility, Interoperability, and Reusability*) [25], la prima fase del lavoro ha permesso di definire i principi alla base della trascrizione e gli elementi con relativi attributi e valori.

Per la trascrizione è stata utilizzato un protocollo di codifica XML-TEI, creato da Dal Bo a partire dalle informazioni accessibili di altri progetti di ricerca che lavorano su tali oggetti di studio e arricchito degli elementi necessari alla codifica delle particolarità linguistiche di questo corpus, al fine di facilitare e uniformare il più possibile il lavoro dei vari collaboratori al progetto. Questi hanno ugualmente accesso alla versione 2.0 del protocollo redatta in francese e aggiornata da Dal Bo e Giaufret nel novembre 2023. Esso si compone di tre sezioni: A) *TEI Header*, B) Organizzazione testuale della lettera, C) Corpo del testo. La prima (A) presenta alcune delle etichette da utilizzare per l'annotazione dei metadati, fra cui alcuni elementi di <correspDesc> [19]. La seconda sezione (B) presenta i tag che identificano le varie parti che compongono la lettera, come <opener> e <closer> per le formule epistolari e <postscript> per eventuali *post-scriptum*. Infine, la terza (C) contiene i tag necessari alla riproduzione dei documenti originali, sia per quanto riguarda la struttura delle sezioni del testo (ad esempio, <p>, <pb/>, <lb/>, ecc.), sia per le modifiche apportate dagli scriventi nel processo di scrittura (<del>, <add>, etc.) con i relativi attributi. In questa sezione sono anche indicate le etichette che possono essere utilizzate per annotare elementi specifici, che sono stati precedentemente identificati come oggetti di analisi linguistiche pertinenti (per esempio, <unit> per le unità di misura, <foreign> e in caso di uso di utilizzo di lingue diverse da quella principale del documento) e per l'analisi di contenuti specifici, quali i nomi di persona e di luogo (<persName>, <placeName>) per le entità nominate corrispondenti.

È stato inoltre creato un database che viene costantemente aggiornato man mano che le trascrizioni sono completate e caricate nell'archivio. Tale database contiene colonne utili a identificare per ogni lettera elementi quali la lingua principale in cui è scritta, mittente e destinatario, ma anche luogo di partenza, luogo di destinazione e luoghi menzionati. È per questo una fonte importante di dati linguistici e geografici, che saranno utili per rapide ricerche e per la creazione delle cartine geografiche che accompagneranno l'edizione.

Alla luce delle problematiche emerse durante la fase di trascrizione e annotazione del corpus, un'ulteriore versione del protocollo di codifica XML-TEI sarà definita nell'autunno 2024. Le difficoltà riscontrate riguardano principalmente la natura stessa del linguaggio di markup, che, essendo costruito su una struttura ad albero rigidamente gerarchica [3], non sempre permette di trascrivere con la fedeltà imporrebbe il progetto le particolarità testuali di alcune lettere (ad esempio la gestione di più elementi <closer> o l'inserimento di questi nel margine del foglio). La versione aggiornata del protocollo sarà quindi applicata all'insieme del corpus nella fase di lavoro successiva, quella della revisione e armonizzazione delle

[5] L'Édition numérique de correspondances https://cahier.hypotheses.org/guide-correspondance. (Ultimo accesso: 18 gennaio 2024).

[6] TEI Guidelines: https://tei-c.org/guidelines/. (Ultimo accesso: 18 gennaio 2024).

trascrizioni, prevista per l'estate 2025. La versione definitiva del protocollo sarà caricata nell'archivio digitale Zenodo. In questa versione aggiornata è prevista l'integrazione esplicita con l'ontologia di cui sopra, specialmente per quanto riguarda le classi create, che si ispirano direttamente alle etichette utilizzate nella fase di trascrizione. Lo scopo di tale lavoro è di rendere più trasparenti le scelte operate dai due gruppi che lavorano sulla codifica e sull'ontologia e il legame tra i due protocolli.

Riprendendo la distinzione di Pierazzo [15] tra edizioni *haute couture*, che prevedono l'uso di un software creato ad hoc per gli interventi a valle dell'edizione, e edizioni *prêt-à-porter*, che implicano, invece, l'adattamento di un software esistente, il progetto Corr<si>Ca si colloca nel secondo gruppo. Infatti, si prevede di utilizzare uno strumento come EVT (*Edition Visualization Technology*) per la visualizzazione e TXM[7] [13] o un altro strumento equivalente per le interrogazioni complesse di tipo linguistico nel corpus. In parallelo, lo sviluppo di un blog Wordpress permetterà di valorizzare conoscenze e di inserire strumenti di accompagnamento (per la geolocalizzazione, la visualizzazione della rete di scriventi, ecc.).
La scelta di EVT (ancora in fase di valutazione) come software per la creazione, navigazione e visualizzazione dell'edizione digitale della corrispondenza sarebbe motivata in primis dalla compatibilità con gli standard XML, HTML e Java, che permettono flessibilità e configurabilità [8, 17] e, in secondo luogo, dalla possibilità di sfruttare funzionalità di geolocalizzazione già presenti in EVT 2, ma che verranno implementate con migliori collegamenti tra testo, risorse LOD e mappe in EVT 3, la cui versione alpha è stata lanciata nel dicembre 2022 e che probabilmente verrà rilasciato nel 2024.
Precisiamo che l'edizione ha escluso per il momento di interessarsi agli aspetti materiali delle lettere (grana della carta, penna, timbri, ecc.), salvo quando questi elementi possono costituire indizi utili al collocamento temporale o geografico della missiva.

Nella prima fase del progetto, ci siamo interrogate sulla granularità della codifica, su un eventuale protocollo di anonimizzazione (per i dati sensibili) e sul tipo di edizione. Abbiamo infine optato per una trascrizione diplomatica delle lettere che rispetti criteri di fedeltà rigorosa all'originale (segmentazione, punteggiatura, uso delle maiuscole, ortografia, ecc.).

4. PARTICOLARITÀ E INTERESSE LINGUISTICO DEL CORPUS

Oltre alle caratteristiche della lingua utilizzata dagli scriventi semicolti in queste missive, la fase di trascrizione e codifica ha messo in evidenza altre particolarità del corpus, frequenti nei corpora di corrispondenze di persone "comuni", quali l'alto numero di scriventi non presenti nelle basi di dati del web semantico (come VIAF[8]) e i nomi di luogo assenti nei database di geolocalizzazione come DBpedia[9] e GeoNames[10], perché poco noti, locali e/o dialettali. Inoltre, poiché non esiste un'edizione critica precedente delle lettere, sono emersi elementi problematici, alcuni dei quali tipici dell'edizione di corrispondenze (elementi non decifrabili o non presenti, numerose grafie da decifrare e/o identificare, presenza di lettere non autografe, ecc.), alcuni più specifici alle corrispondenze di scriventi semicolti [20][11] (usi linguistici lontani dalla norma dell'epoca), altri ancora specifici alla regione di provenienza, poiché si tratta di scriventi bilingue o trilingue, le cui lettere costituiscono una preziosa testimonianza del momento di transizione linguistica dall'italiano al francese, con l'emergenza, talvolta, del substrato còrso. Infatti, la Corsica attraversa tra l'Ottocento e il Novecento una fase di francesizzazione, scandita da momenti più o meno intensi, e che agisce soprattutto sulla popolazione scolastica. Secondo alcuni studi sull'argomento [1, 22], il reale arretramento dell'italiano lingua scritta di fronte all'avanzata del francese avviene solo intorno alla metà dell'Ottocento, mentre la trasmissione intergenerazionale della lingua còrsa si mantiene fino agli anni Cinquanta del Novecento. La situazione linguistica della popolazione può quindi essere definita di diglossia, con una lingua parlata, il còrso, e una lingua scritta che è dapprima l'italiano e poi il francese, con una fase di transizione in cui il repertorio linguistico si compone di tutte e tre, come nel caso di Pierre François. Il nostro corpus contiene 8 lettere scritte interamente in italiano, tutte da Xavier Canioni (nato nel 1835) e due *post scripta*, uno di Pierre François (nato nel 1858) e uno di Xavier. Dall'analisi delle grafie abbiamo ipotizzato che le lettere in francese di Xavier (dal 1888 in poi, con qualche

[7] TXM: https://txm.gitpages.huma-num.fr/textometrie/index.html. (Ultimo accesso: 18 gennaio 2024).
[8] VIAF: https://viaf.org/. (Ultimo accesso: 10 aprile 2024).
[9] DBpedia Fr: https://fr.dbpedia.org/. (Ultimo accesso: 15 aprile 2024).
[10] GeoNames: https://www.geonames.org/. (Ultimo accesso: 15 aprile 2024).
[11] Inoltre, per studi sulla lingua dei semicolti, rimandiamo a [2] per il francese, [5, 11, 21] per l'italiano e [4] per l'italiano in Corsica.

eccezione) siano in realtà dettate al figlio Pierre François, che le traduce dal còrso o dall'italiano, ma che è tuttavia in grado di scrivere anche in quest'ultima lingua.
Gli esempi che seguono mostrano non solo come la lingua usata, sia essa italiano o francese, si discosti dalla norma dello standard, ma anche come si manifesti la presenza di lemmi della lingua còrsa (con *code-mixing*), soprattutto laddove gli argomenti affrontati sono relativi alla cultura materiale. Peraltro, l'uso del còrso non si limita ai nomi degli alimenti (in questo esempio specifico), ma si diffonde anche agli articoli e ai numerali:

"Nous vous avons expédier un petit colis postal de cinque kilos. contenant una salticca, deux fiadelli una pulpetta, una fetta di mezina e dui furmagli. Voilà le tout, nous avons coupé un petit morceau de pulpetta qui passait le poid". (02/02/1914, Pierre François; si noti anche l'anno, che si colloca alla fine dell'arco cronologico del corpus).

Ecco, quindi, la prima proposta di codifica adottata dal progetto di ricerca:

```
<p>Nous vous avons expédier<lb/>
un petit colis postal<lb/>
de cinque <unit type="weight" unit="kg"> kilos</unit>. contenant<lb/>
<foreign xml:lang="co">una salticca</foreign>, deux<lb/>
<span xml:lang="co">fiadelli una pulpetta,<lb/>
una fetta di mezina<lb/>
e dui furmagli</span>.<lb/>
Voilà le tout, nous avons<lb/>
coupé un petit morceau<lb/>
de <foreign xml:lang="co"> pulpetta</foreign> qui passait<lb/>
le poid.</p>
```

Altro esempio interessante delle specificità del corpus è quello che riguarda le unità di misura, per le quali gli scriventi continuano a usare quelle tradizionali (in particolare *rubo* e *cantaro*) anche quando il sistema decimale (usato anch'esso) si è diffuso in Francia, e quindi in Corsica:

"Mi dimandi se le uve sono bone. Si, sono bone, e molto a marchato Ecco il prezzo, 30 soldi il rubo. Cioe 6, franchi il cantaro dunque sapia che ci vole, 13 rubi duva per fare, 10020 litri di mosto. Viene dunque a, 3 soldi il litro. Il prezzo e bono". (30/01/1881, Xavier)

Questa la proposta di codifica della citazione precedente:

```
<p>Mi di mandi sele uve sono bone<lb/>
si sono bone e molto à marchato<lb/>
E cco il prezzo. 30 <unit type="currency">soldi</unit> il <unit type="weight">rubo</unit><lb/>
cioe. 6, <unit type="currency">franchi</unit> il <unit type="weight">Cantaro</unit> dunque<lb/>
sapia che ci vole, 13, <unit type="weight">rubi</unit> duvaper<lb/>
fare, 100 20 <unit type="volume" unit="L">li tri</unit> di mosto.</p>
<p>viene dunque a, 3. <unit type="currency">soldi</unit> il <unit type="volume" unit="L">litro</unit>.<lb/>
il prezzo e bono.</p>
```

Esistono infine esempi di lettere scritte nelle due lingue, ovvero in cui il corpo della lettera è in francese e il *post scriptum* in italiano. L'elemento interessante è che queste due lettere sono entrambe dello stesso scrivente, Pierre François, ma nel primo esempio è lo stesso Pierre François a esserne l'autore (20/09/1887), mentre nel secondo la lettera è firmata da Xavier, sebbene la grafia sia attribuibile allo stesso Pierre François (30/06/1893, sei anni dopo la precedente).
Ecco la proposta di codifica di quest'ultimo esempio (lettera di Xavier, scritta da Pierre François, 30/06/1893):

```
<postscript>
<p><span xml:lang="it">Caro figlio mi vene per<lb/>
iscontro un cabriole a unpr<lb/>
ezzo tutto affatto à mercato<lb/>
```

per di meglio per 90 <unit type="currency">franchi</unit><lb/>
che avendolo da comprare in<lb/>
fabrigo costarebe almeno 150 <unit type="currency">fra<add place="above">n</add>chi</unit>. ma non trovandomi<lb/>
mica instato di denaro per<lb/>
podere fare questa compra<lb/>
mi adirizzo a te si nulla ti<lb/>
pergudigueca et se ti le trovi<lb/>
a la mano fendodi il tuo<lb/>
bougletto <del rend="overstrike">et et</del> e il piu presto<lb/>
sarai pagato frutti et fondo<lb/>
Se tu decili : per la soma di 100 <unit type="currency">fr</unit> fr.<lb/>
rispondi il piu presto possibile</p>
</postscript>

Dagli esempi precedenti si evincono le difficoltà di codifica a causa di problematiche linguistiche relative alla segmentazione dei lemmi, alla distanza dalla norma generalizzata e alle interferenze continue tra le tre lingue in compresenza. Tuttavia, è proprio da queste difficoltà che potranno nascere le riflessioni più innovative sul piano scientifico. Per esempio, si prevede di utilizzare il tag <w> per indicare la segmentazione standard delle parole che si trovano ipo- o ipersegmentate nel testo originale e <choice> per segnalare le forme standard. Seguendo le raccomandazioni della TEI, l'obiettivo è di indicare, all'interno di <choice>, la forma originale attraverso il tag <orig>, e la forma standard utilizzando l'elemento <reg> e l'attributo @type per precisare il livello linguistico interessato dalla normalizzazione (per esempio, ortografia, morfologia, sintassi).

Mi <w>di mandi</w> <w>se</w><w>le</w> uve sono bone<lb/>

Nous vous avons <choice><orig>expédier</orig><reg type="morphosynt">expédié</reg></choice><lb/>

Ulteriori difficoltà di codifica sono presentate dall'organizzazione grafica del testo, le cui sezioni non sono sempre delimitate chiaramente, dall'occupazione da parte di questo di molti spazi a margine, dall'ordine testuale che non corrisponde all'ordine grafico, ecc.

5. STATO DI AVANZAMENTO E RISULTATI ATTESI

Le fasi del progetto già realizzate e previste sono quindi:

1. Elaborazione e firma di un protocollo d'intesa con i discendenti della famiglia Canioni (conclusa)
2. Scansione delle lettere in HD (terminata)
3. Elaborazione del protocollo di codifica (prima versione realizzata)
4. Trascrizione delle lettere e codifica in XML-TEI (60% circa)
5. Revisione delle trascrizioni (da effettuare alla fine della fase precedente e della fase 6.1)
6. Codifica strutturale e semantica
   6.1. Fase preliminare: metadati, testo, unità di misura, nomi di luogo e di persona (in corso)
   6.2 Fase avanzata: inserimento di etichette per standardizzare le forme e annotazione semantica delle entità nominate (attesa di finanziamenti supplementari).

Parallelamente, si sta lavorando alla firma di un accordo con l'Università di Corsica Pasquale Paoli, affinché il progetto possa essere ospitato sul sito del Laboratoire M3C ed essere reso disponibile alla comunità dei ricercatori e degli appassionati in Corsica.
I risultati attesi alla conclusione del progetto saranno quindi un portale web da cui saranno accessibili: la piattaforma di visualizzazione delle lettere, accompagnata dall'edizione diplomatica; il corpus su TXM o strumento simile per effettuare ricerche linguistiche complesse tramite etichette XML-TEI; un blog par la valorizzazione della ricerca che conterrà un apparato critico e divulgativo, compreso quello iconografico e documentale.

Siamo convinti che questo materiale potrà essere utile alla comunità degli storici, dei linguisti, degli etnografi, ecc. per approfondire la conoscenza di una microregione còrsa, il Ghjunsani, dei suoi rapporti con altre regioni dell'isola e del continente e di un periodo storico fondamentale per la storia del repertorio linguistico dei Còrsi.

6. RINGRAZIAMENTI

Ringraziamo i discendenti della famiglia Canioni, che ci hanno generosamente prestato la corrispondenza originale oggetto di questo studio, e Santu Massiani, prezioso custode della storia del Ghjunsani.

BIBLIOGRAFIA

[1] Branca, Marina, e Nicolas Sorba. "Un siècle d'évolution de la transmission intergénérationnelle du corse". Glottopol, no. 38, 2023. DOI: https://doi.org/10.4000/glottopol.3179.

[2] Branca-Rosoff, Sonia, e Nathalie Schneider. L'Écriture des citoyens. Une analyse de l'écriture des peu-lettrés pendant la période révolutionnaire. Paris: Klincksieck, 1994.

[3] Ciotti, Fabio. Digital Humanities. Metodi, strumenti, saperi. Roma: Carocci editore Spa, 2023.

[4] Colombani Giaufret, Hélène, e Anna Giaufret. "Il manoscritto dei verbali del comune di Pioggiola (Corsica), 1788-1797: analisi testuale e linguistica". Una piccola comunità corsa negli anni della Rivoluzione. Pioggiola attraverso il manoscritto delle delibere 1787-1797. A cura di Francesca Ferrando, Palermo: New Digital Press, 2022, pp. 35-59.

[5] D'Achille, Paolo. "L'italiano dei semicolti". Storia della lingua. A cura di Luca Serianni e Pietro Trifone, vol. 2, 1994, vol. 2º, Einaudi, 1994, pp. 41-79.

[6] Dal Bo, Beatrice, Francesca Frontini, e Giancarlo Luxardo. "Annotazione semantica e visualizzazione di un corpus di corrispondenze di guerra". Atti del IX Convegno Annuale AIUCDA cura di Marras, Cristina et al., Milano: Università Cattolica del Sacro Cuore, 2020. ISBN 978-88-942535-4-2. DOI: 10.6092/unibo/amsacta/6316.

[7] Del Grosso, Angelo Mario, Erica Capizzi, Salvatore Cristofaro, Maria Rosa De Luca, Emiliano Giovannetti, Simone Marchi, Graziella Seminara, e Daria Spampinato. "Bellini's Correspondence: A Digital Scholarly Edition for a Multimedia Museum". Umanistica Digitale, 18 dicembre 2019, no. 7, 2019, pp. 23-47.

[8] Di Pietro, Chiara, e Roberto Rosselli Del Turco. "Between innovation and conservation: the narrow path of user interface design for digital scholarly editions". Digital Scholarly Editions as Interfaces. A cura di Roman Bleier et al., Norderstedt: BoD, 2018, pp. 133-163.

[9] Donato, Maria Pia. "Lettere, corrispondenze, reti epistolari". Mélanges de l'École française de Rome - Italie et Méditerranée modernes et contemporaines, vol. 132-2, 2020, pp. 249-255. DOI: https://doi.org/10.4000/mefrim.9995

[10] Duménieu, Bertrand, Danièle Poublan, Jean-Damien Généro, Francine Filoche e Patricia Bleton. "Un wiki sémantique pour l'édition scientifique d'une correspondance du XIXe siècle". Humanités numériques, 1 dicembre 2022, no. 6, 2022. DOI: https://doi.org/10.4000/revuehn.3203.

[11] Fresu, Rita. "14. L'italiano dei semicolti". Manuale di linguistica italiana. Edited by Sergio Lubello, Berlin, Boston: De Gruyter, 2016, pp. 328-350. DOI: https://doi.org/10.1515/9783110360851-016

[12] Géa, Jean-Michel. "Entre identité locale et sentiment national : la posture énonciative de deux soldats corses durant la Première Guerre mondiale". Études corses, dicembre 2004, no. 59, 2004, pp. 129-143.

[13] Heiden, Serge, Jean-Philippe Magué, e Bénédicte Pincemin. "TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. " JADT 2010: 10th International Conference on the Statistical Analysis of Textual Data. giugno 2010, Rome, pp. 1021-1032.

[14] Martineau, France, e Sandrine Tailleur. "Correspondance familiale acadienne au tournant du XXe siècle : fenêtre sur l'évolution d'un dialecte". Congrès Mondial de Linguistique Française. CMLF, Paris: Institut de Linguistique Française, 2010. DOI: https://doi.org/10.1051/cmlf/2010118.

[15] Pierazzo, Elena. "Quale futuro per le edizioni digitali? Dall'*haute couture* al *prêt-à-porter*". Atti del V Convegno Annuale AIUCD. A cura di Boschetti, Federico, AIUCD, 2017. DOI: 10.6092/unibo/amsacta/5559.

[16] Praxiling - UMR 5267 (2019). Corpus 14 [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage).

[17] Rosselli Del Turco, Roberto. "Designing an advanced software tool for Digital Scholarly Editions." Textual cultures vol. 12, no. 2, 2019, pp. 91-111. DOI: https://doi.org/10.14434/textual.v12i2.27690

[18] Salvatori, Enrica, Roberto Rosselli Del Turco, Chiara Alzetta, Chiara Di Pietro, Chiara Mannari, Alessio Miaschi. “Il Codice Pelavicino tra edizione digitale e Public History”, Umanistica Digitale, 10 gennaio 2017, no. 1, 2017. DOI: https://doi.org/10.6092/issn.2532-8816/7232.

[19] Stadler, Peter, Marcel Illetschko, e Sabine Seifert. “Towards a Model for Encoding Correspondence in the TEI: Developing and Implementing <correspDesc>”. Journal of the Text Encoding Initiative, settembre 2016 – dicembre 2017, no. 9, 2016. DOI: https://doi.org/10.4000/jtei.1433

[20] Steuckardt, Agnès (a cura di). Entre village et tranchées. L’écriture de Poilus ordinaires. Uzès: Inclinaison, 2015.

[21] Testa, Enrico. L’italiano nascosto. Einaudi, 2014.

[22] Thiers, Ghjacumu. "Aspects de la francisation au XIXème siècle, en Corse". Études corses, no. 9, 1978, pp. 5-39.

[23] Tomasi, Francesca. “XML/TEI per la trascrizione delle fonti primarie e la codifica dell'apparato critico”. Journal of Latin Linguistics, 1 dicembre 2007, vol. 9, no. 3, 2007, pp. 129-148. DOI: https://doi.org/10.1515/joll.2007.9.3.129

[24] Walter, Richard, Claire Bustarret, Marie Dupond, Alexandre Guilbaud, Giancarlo Luxardo, Yvan Leclerc, Jean-Sébastien Macke, Irène Passeron, Nicolas Rieucau, e Fabienne Vial-Bonacci. L’Édition numérique de correspondances. Version 1.2, gennaio 2018. Licence CC BY-NC-SA.

[25] Wilkinson, Mark, Michel Dumontier, M., IJsbrand Jan Aalbersberg, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data, 15 marzo 2016, no. 3, 2016. DOI: https://doi.org/10.1038/sdata.2016.18