

Verso un nuovo volto dell'Atlante Linguistico Italiano (ALI). Processi di valorizzazione in ambiente ArcGIS

Cristiano Pesaresi*, Davide Pavia*, Filiberto Ciaglia*, Lorenzo Ferrarotti**, Stefano Fiori***

Parole chiave: *Atlante Linguistico Italiano (ALI), approccio geografico, dialetti, GIS, patrimonio linguistico*

Keywords: *Atlante Linguistico Italiano (ALI), geographical approach, dialects, GIS, linguistic heritage*

Mots-clés : *Atlas Linguistique Italien (ALI), approche géographique, dialectes, SIG, patrimoine linguistique*

1. L'impresa di digitalizzazione dell'Atlante Linguistico Italiano

Nonostante la sua straordinaria ricchezza di dati (si veda Rivoira, Cerruti, Ghia, Turchetta in questo volume), l'*Atlante Linguistico Italiano* (ALI) è una fonte poco utilizzata nella ricerca linguistica, sia perché la sua pubblicazione è ancora incompleta, sia perché attualmente è consultabile solo su volumi cartacei di grande formato e diffusi in numero limitato. Inoltre, la struttura dell'ALI, sebbene più densa di punti rispetto all'*Atlante Italo-Svizzero* (AIS), è meno omogenea e meno orientata agli aspetti morfosintattici (Loporcaro, 2013, p. 30). In anni recenti, l'esigenza di rendere accessibili e sostenibili digitalmente i grandi archivi linguistici ha spinto numerosi progetti atlantistici europei verso la digitalizzazione (si vedano, ad esempio, Krefeld, Lücke, 2018; Thibault *et alii*, 2021), ma l'ALI non ha ancora partecipato pienamente a questo processo, rimanendo di fatto difficilmente accessibile in quanto a *human readability* e del tutto non *machine readable*, e dunque non conforme ai principi FAIR (Krefeld, Lücke, 2023). Nel 2023, però, l'ALI ha intrapreso la strada della digitalizzazione grazie alla sua inclusione nel partenariato esteso PNRR «CHANGES» (*Cultural Heritage Active Innovation*

* Roma, Università Sapienza, Italia.

Il lavoro è frutto della collaborazione degli Autori, ma nello specifico: C. Pesaresi ha scritto il paragrafo 6; D. Pavia il paragrafo 7; F. Ciaglia il paragrafo 8; L. Ferrarotti i paragrafi 2 e 3; S. Fiori i paragrafi 4 e 5. Il paragrafo 1 è stato scritto congiuntamente da L. Ferrarotti e S. Fiori. Il paragrafo 9 è stato scritto congiuntamente da C. Pesaresi, D. Pavia e F. Ciaglia.

** Bergamo, Università degli Studi, Italia.

*** Torino, Università degli Studi, Italia.

for Next-Gen Sustainable Society), all'interno dello Spoke 3 (*Digital libraries, archives and philology*), e più specificamente nel Work Package 5 (*Languages and their legacies in oral digital archives*).

In questo quadro nasce il progetto DigitALI, che ha l'obiettivo di digitalizzare i dieci volumi dell'ALI finora pubblicati (prevedendo l'applicazione del processo anche su quelli che saranno pubblicati in futuro), creando un'interfaccia cartografica interattiva online che consenta la consultazione e l'interrogazione dei dati tramite filtri combinabili, secondo un modello *corpus-based*. Il progetto non si limita, dunque, a una semplice riproduzione delle carte a stampa, ma mira ad esempio a valorizzare anche le parti meno accessibili dell'ALI, già pubblicate altrove, come il questionario (Bartoli, Pellis, 1971) e i verbali d'inchiesta (Massobrio *et alii*, 1995), che contengono informazioni geo-demografiche e sociolinguistiche preziosissime per l'interpretazione dei dati linguistici, ai quali si aggiungono due collezioni fotografiche: il fondo Ugo Pellis, già consultabile sul sito della Società Filologica Friulana, di enorme interesse demologico ed etnografico, e quello prodotto dai raccoglitori successivi (coordinati dal direttore dell'*Atlante* dell'epoca, Benvenuto Terracini), ancora inedito, conservato nella sede dell'Istituto dell'Atlante Linguistico Italiano a Torino (presso l'Università degli Studi, in via Sant'Ottavio 20). Questi contenuti saranno integrati nella piattaforma digitale per offrire ai ricercatori una rappresentazione più completa del materiale raccolto. Inoltre, l'ALI è l'elemento cardine del PRIN PNRR 2022 DigitALi coordinato da Matteo Rivoira, che si propone di digitalizzare diversi atlanti regionali italiani, tra cui l'*Atlante Storico Linguistico ed Etnografico del Friuli*, l'*Atlante Linguistico ed Etnografico del Piemonte Occidentale*, l'*Atlas des Patois Valdôtains* e l'*Atlante Linguistico ed Etnografico Informatico della Conca Aquilana*, al fine di favorire una lettura integrata dei dati raccolti da questi progetti.

2. *Laborazione dei materiali digitali*

A differenza di altri progetti simili di digitalizzazione, come l'AIS Reloaded (Loporcaro *et alii*, 2021), che ha dovuto impiegare complessi strumenti di riconoscimento ottico dei caratteri (OCR) per la digitalizzazione dei suoi materiali, l'ALI ha potuto affrontare direttamente la fase di retro-standardizzazione (Geisler *et alii*, 2021), poiché la redazione dei volumi ha fatto affidamento, sin dalla prima pubblicazione nel 1995, su un processo informatizzato, che si basa sulla digitalizzazione delle schede, grazie a programmi appositamente creati da Claudio Tosti dell'Istituto Poligrafico e Zecca dello Stato (che fino al 2011 ha curato la pubblicazione dei volumi): la Redazione digita nel software dedicato (ALIdigi) i contenuti dell'archivio cartaceo, conservati sotto forma di schede, ciascuna delle quali riporta la risposta fornita a una particolare domanda del questionario in una data località (fig. 1). Per ogni domanda, o «voce», del questionario viene prodotto un file di testo che contiene tante righe quante sono le località nelle quali la voce ha una risposta disponibile: su ognuna viene riportato il numero del punto d'inchiesta, il suo codice alfanumerico (basato su una ripartizione di tipo geografico e utilizzato

per identificare le inchieste) e infine la risposta dialettale, seguita da eventuali commenti. Possono essere presenti anche note di testo e riferimenti fotografici associati alle risposte dialettali, che vengono digitati in file separati. In seguito, questo materiale è elaborato da altri programmi, che svolgono operazioni di normalizzazione dei caratteri, volte ad armonizzare le trascrizioni eseguite da diversi raccoglitori e, in certi casi, di unione di voci diverse, con l'obiettivo di generare l'output grafico delle carte. Bisogna però osservare che il file di testo sottostante al processo è scritto in un codice composto unicamente da caratteri ASCII (d'ora in avanti chiamato, per brevità, «ALICode») che viene trasformato in elementi grafici puramente vettoriali (non quindi caratteri digitali conformi a Unicode) per essere plottato sulle carte. In fig. 1 si vede il rapporto che intercorre tra una scheda dell'*Atlante* (a sinistra) e la stringa di testo presente nell'archivio digitale (sotto); inoltre la fig. 1 mostra come ciò che viene digitato nel programma di inserimento (al centro) venga poi reso sulle carte a stampa (a destra).

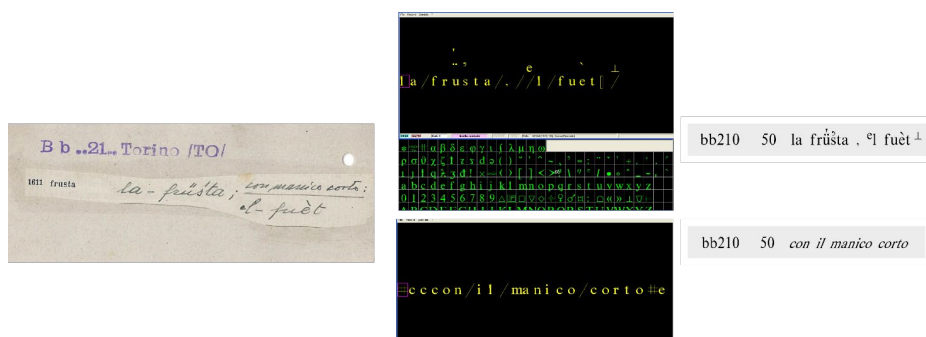


Fig. 1 – A sinistra, la schedina delle risposte raccolte a Torino per la voce 1611 «frusta»; al centro, l'interfaccia dei programmi di digitazione impiegati dalla redazione; a destra, il risultato della digitazione del testo delle risposte (in alto) e delle note (in basso).

Fonte: elaborazione su schedina dell'archivio dell'ALI, voce 1611 «frusta» punto Bb21 Torino; fermo immagine e output dal software ALIDigi.

La concezione di questo sistema è analoga a quella del cosiddetto *Beta Code* sviluppato negli anni '70 del secolo scorso per il *Thesaurus Linguae Graecae* e sfrutta la combinazione di caratteri ASCII per codificare senza perdita d'informazione il complesso insieme di segni che compone l'alfabeto fonetico dell'ALI. La facilità, l'universalità e l'indipendenza da software proprietari di questo sistema hanno consentito di preservare i dati dell'ALI dalle innovazioni informatiche successive, ma ha anche richiesto una complessa procedura di ricodifica e conversione delle stringhe di testo ALICode verso lo standard Unicode (Moran, Cysouw, 2018). Per rimediare alle numerose discrepanze che differenziano la concezione dell'ALICode rispetto a Unicode, si è fatto ricorso a una fase di pre-trattamento, nella quale uno script Python applica alle stringhe ALICode circa venti espressioni regolari in sequenza: ciò porta a una versione «normalizzata» (chiamata «NormALI») sulla quale si può quindi operare la traslitterazione, implementata sempre in Python tramite la

funzione *tokenize* della libreria *segments* (Forkel, Moran, 2018), appositamente creata per gestire dati linguistici in forma scritta, che si basa su una tabella di corrispondenze (detta «profilo ortografico») tra le codifiche ALIcode, Unicode e HTML. Lo stesso tipo di tabella sarà alla base della conversione dei simboli fonetici dell'ALI in caratteri IPA.

La fase successiva è consistita nella riconfigurazione dei file di testo contenenti i dati linguistici in un formato tabellare «lungo», in cui ciascuna riga dei file traslitterati corrisponde a un punto d'inchiesta. Tuttavia, ogni stringa può contenere più di una risposta dialettale e le sigle identificative degli informatori che forniscono le risposte, come <*ai> (= 'altro informatore'); si trovano anche informazioni aggiuntive fornite dagli informatori stessi o dalla Redazione, come etichette relative a fatti variazionali (ad es. <rust> 'rustico') o a indicazioni relative all'uso, nonché i simboli utilizzati per rimandare alle note, alla legenda, ai riferimenti fotografici e ai tipi etnografici.

Per creare un database strutturato, nel quale ciascuna informazione fosse contenuta in colonne dedicate e dunque facilmente estraibile, è stata messa a punto un'altra procedura, basata questa volta su uno script in ambiente R (v4.1.1; R Core Team, 2021), che sfrutta i segni di interpunzione già presenti nelle stringhe di testo originariamente adibiti alla separazione delle risposte, per collocare ciascuna di esse su una riga diversa per incorporare ogni informazione tra quelle menzionate in una nuova colonna dedicata. Ogni nuova riga della tabella mantiene l'informazione del numero di punto e del codice alfanumerico e una colonna con la stringa originale, precedente a questa manipolazione, in formato HTML, che serve per la riproduzione digitale delle carte a stampa. Lo script fa sì che questo processo venga ripetuto automaticamente per tutti i file relativi alle singole voci di un volume, che alla fine sono riuniti in un unico file al quale vengono aggiunte le informazioni provenienti dai file che contengono le note e i riferimenti fotografici, sottoposti in precedenza allo stesso processo di normalizzazione e riconfigurazione. In questo modo sono stati riconvertiti interamente i voll. IX e X, più recenti e affidabili, e, in un secondo momento, i voll. V, VI e VII, mentre i voll. II, III e VIII sono sotto revisione. I voll. I e IV verranno invece ridigitati per intero, non essendo stato possibile recuperare i file originali. Il vol. XI, in corso di redazione, sarà elaborato non appena sarà terminato dalla Redazione, dopo essere stato pubblicato anche in formato cartaceo.

Come si è già accennato, tuttavia, la versione digitale dell'ALI non sarà una semplice riproduzione dei materiali dell'*Atlante* cartaceo (che tra l'altro verrà implementata grazie ad apposite funzionalità geotecnologiche in termini di interrogazione e restituzione dinamica di informazioni). Infatti, sin dall'inizio del progetto, è stata portata avanti anche la digitalizzazione dei verbali d'inchiesta, già pubblicati in volume (Massobrio *et alii*, 1995), le cui informazioni sulle località e i relativi informatori rappresentano i «metadati» da collegare ai dati linguistici veri e propri: la valorizzazione di questi è un tratto caratterizzante del progetto DigitALI. Le informazioni sono state digitate manualmente in un file Excel che conta circa 1800 record: in questa fase del lavoro è stata accordata priorità ad alcune caratteristiche fondamentali degli informatori (nome, sesso, età, anno di nascita, gradi istruzione, professione) e delle inchieste (data di svol-

gimento, raccoglitore, parti del questionario somministrate), mentre altre sono state momentaneamente tralasciate (per esempio, riguardo agli informatori, il luogo di nascita dei genitori ed eventuali periodi passati lontano dal luogo d'origine, e riguardo alle località indagate, la diocesi d'appartenenza e la presenza di trasporti ferroviari o di altro genere). Tutte queste informazioni potranno comunque essere facilmente integrate nel database in un secondo momento.

3. La geolocalizzazione dei punti d'inchiesta

Naturalmente, un passo fondamentale nell'ottica di digitalizzazione dell'*Atlante* è stata la geolocalizzazione dei punti d'inchiesta. Poiché nei materiali digitali della redazione erano presenti soltanto le coordinate geometriche utili al plottaggio delle carte a stampa, le coordinate geografiche delle località indagate hanno dovuto essere riassegnate *ex novo*. Non si è trattato, tuttavia, di un processo puramente meccanico, ma di un'occasione per ritornare criticamente su alcuni assunti della cartografia «tradizionale» dell'ALI: è tutt'altro che semplice, infatti, definire che cosa esattamente rappresentino i punti d'inchiesta che si trovano sulle carte. In linea di principio, un punto dovrebbe corrispondere a un'inchiesta, ma questa non è necessariamente un evento unitario sul piano temporale, poiché spesso, specialmente durante la prima campagna di inchieste (1925-1946), il raccoglitore è tornato più volte nello stesso punto per completare il questionario o per porre domande integrative. Ma sovente non lo è nemmeno sul piano spaziale: un nutrito numero di inchieste è di fatto costituito da interviste svolte con informatori diversi in luoghi geograficamente distinti, per quanto vicini tra loro. Talvolta questa distinzione consente di far intravedere dinamiche di espansione e contrazione di tratti linguistici, come nel caso del punto 29, in cui le due inchieste rivelano una differenza tra l'eponimo centro di fondovalle (Locana, TO), più esposta all'influenza linguistica della pianura torinese, e la borgata montana di San Lorenzo, più conservativa. Quello dell'unità del punto linguistico e dei suoi rapporti con l'area circostante è un tema centrale della riflessione della scuola geolinguistica torinese e specialmente di quella di Benvenuto Terracini (1960), direttore dell'ALI dal 1953 al 1968. Oggi queste dinamiche possono essere rese più trasparenti grazie al livello di granularità che la cartografia digitale consente di raggiungere: tramite Google Earth, dunque, si è proceduto alla geolocalizzazione e al posizionamento dei punti d'inchiesta degli informatori tramite l'aggiunta di altrettanti *placemark*, basandosi nuovamente sulle informazioni contenute nei verbali. Si riporta a titolo di esempio, in fig. 2, la geolocalizzazione delle tre frazioni che compongono l'inchiesta di San Severino Marche (Punto 483).

È stata così rovesciata la prospettiva originaria, ponendo al centro della rappresentazione cartografica non tanto il punto d'inchiesta considerato astrattamente (le cui informazioni identificative originali sono comunque reperibili nel database) ma i singoli parlanti, coerentemente con l'obiettivo di dare risalto a informazioni che possono sfuggire a una consultazione

statica delle sole carte linguistiche, slegata dai materiali integrativi che ne arricchiscono l'interpretazione.

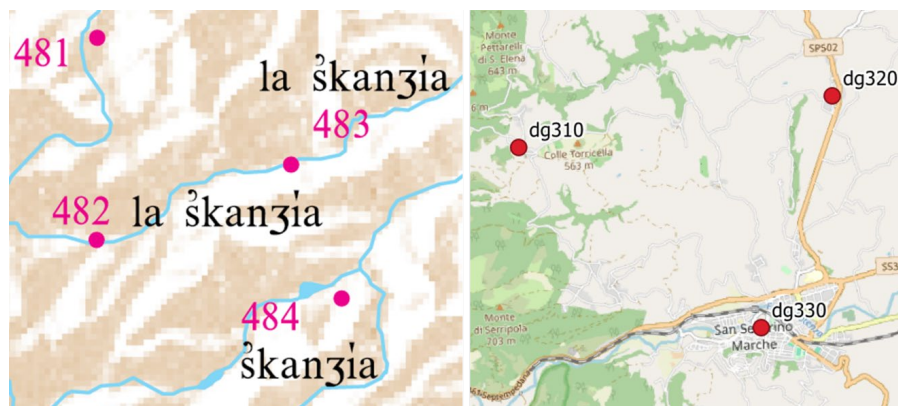


Fig. 2 – A sinistra: ingrandimento sui punti 481, 482, 484 e in particolare 483, riferito a San Severino Marche della carta «scansia» (X 954); a destra: geolocalizzazione delle tre frazioni relative al punto 483 in cui è stata svolta l'inchiesta.

Fonte: elaborazione su dati carta ALI 954 X «scansia» e Google Earth.

4. Costituzione del database

La congiunzione degli elementi descritti nei paragrafi precedenti (dati linguistici, verbali d'inchiesta e punti geolocalizzati) costituisce, assieme al questionario, il primo nucleo del database DigitALI, che è stato impostato secondo un modello relazionale classico con un'implementazione in PostgreSQL e con un'estensione PostGIS per gestire i dati spaziali relativi ai punti d'inchiesta. Il database viene gestito con l'applicazione *pgadmin* e per il momento è ospitato localmente, ma si prevede a breve la possibilità di caricarlo su uno dei server acquistati appositamente da CHANGES per l'archiviazione dei dati, mantenuti dall'Università di Bergamo. Questa architettura consente flussi di lavoro che si integrano con diversi applicativi come QGIS (il quale dispone di plugin appositi per interfacciarsi con Postgres) e ArcGIS, che sarà alla base della risorsa cartografica, come verrà illustrato nei paragrafi seguenti.

Il modello relazionale non è l'unica scelta possibile, ma è sicuramente adeguata alle caratteristiche dell'ALI, il cui archivio presenta già un grado molto alto di strutturazione; per questo si è riflettuto molto su come le diverse entità che compongono l'*Atlante* entrano in relazione e in base a quali informazioni. I punti d'inchiesta e gli informatori sono legati dall'associazione combinata del numero di punto d'inchiesta e di un codice alfanumerico, derivante dalla centuriazione della penisola, uno schema poi divenuto inoperante ma tuttora presente nei dati; le stesse informazioni legano le risposte dialettali ai punti linguistici, mentre un numero identificativo le collega alla relativa domanda nel questionario. La questione più delicata riguarda

il legame tra informatori e dati linguistici: poiché molti punti d'inchiesta hanno più di un informatore, e in alcuni di questi casi le stesse domande vengono poste a persone diverse, si tratta di una relazione «multi-a-molti» (M:N), che non può essere formalizzata direttamente nel modello relazionale classico. La versione del linguaggio SQL (*Structured Query Language*) di Postgres consente comunque di effettuare operazioni di *join* su questo tipo di dati, ma in primo luogo la *query* risulta piuttosto complessa, e in un'ottica di migrazione in ambiente ArcGIS la soluzione più efficace e ortodossa sarebbe la definizione di una tabella pivot che metta in relazione i codici identificativi di ciascun informatore con quelli delle domande del questionario a cui hanno risposto. Questo punto è tuttora oggetto di riflessione nel gruppo di lavoro.

5. Alcune ulteriori prospettive del lavoro di digitalizzazione

In generale, il processo di digitalizzazione sta senz'altro contribuendo ad avvicinare i dati dell'ALI al paradigma FAIR (Wilkinson *et alii*, 2006), convertendoli in un formato che ne favorisce la condivisione, il riutilizzo e l'interoperabilità con altre risorse. In quest'ottica, uno degli sforzi principali per il futuro sarà la conversione dei dati linguistici in IPA, oggi imprescindibile per gli atlanti digitali (Loporcaro *et alii*, 2021, Ghia 2023). Tecnicamente, è possibile impiegare strumenti già utilizzati per traslitterare ALIcode in Unicode e HTML (Geisler *et alii*, 2021). Tuttavia, come sottolineano Colcuc e Zacherl (2022), a differenza della traslitterazione, il passaggio all'IPA comporta una semplificazione, dovuta all'assenza di corrispondenze biunivoche tra ALI e IPA, soprattutto per la ricchezza combinatoria dei simboli ALI (Abete, Rivoira, 2021). Si potrà quindi prevedere una doppia trascrizione: una più dettagliata, l'altra più accessibile. La trascrizione IPA permetterà anche analisi dialettometriche basate sull'algoritmo di Levenshtein, già applicato con Gabmap all'ALT (Wieling *et alii*, 2016; Rubehn *et alii*, 2024). Inoltre, un database digitale con trascrizione semplificata faciliterà l'indicizzazione e la tipizzazione automatica, rimuovendo i diacritici dalle forme fonetiche (Genre, Massobrio, 1976). Resta però necessaria una riflessione teorica sulla tipizzazione (Cerruti, Regis, 2008; Regis, 2019), soprattutto per l'ALI, in cui – diversamente dall'*Atlante Lessicale Toscano* (ALT) – le varietà sono troppo eterogenee per applicare regole uniformi (Picchi *et alii*, 2001). Alla tipizzazione si lega inoltre la possibilità, esplicitamente prevista anche dal progetto «gemello» DigitALiA, di un approfondimento dell'aspetto semantico dei materiali dell'ALI: ciò prevede, per esempio, che ogni voce del questionario (ma la cosa potrà valere anche per i singoli tipi delle risposte) venga legata a uno o più concetti tratti da un vocabolario condiviso con altre banche di dati linguistici. Sulle strategie da adottare la riflessione è ancora aperta: tra i modelli adottabili vi è quello di Verba Alpina che fa riferimento a voci compilate su Wikidata (corredate da un codice univoco) o quello adottato dal progetto di digitalizzazione dei *Tableaux Phonétiques* che ricorre al *Concepticon* (List *et alii*, 2025).

6. La migrazione del DigitALI in ambiente ArcGIS e possibili sviluppi

Per poter effettuare una migrazione in ambiente ArcGIS finalizzata a un processo multivariato di valorizzazione geotecnologica è stata, dapprima, necessaria un'ampia attività di analisi condivisa delle tabelle precedentemente strutturate e a disposizione. Una serie di riunioni tecnico-operative in via telematica tra le unità operative costituite dalle Università di Bergamo, Roma Sapienza e Torino ha permesso di definire procedure e campi di raccordo che consentissero, poi, di lavorare nella prospettiva di concatenazioni tabellari propedeutiche all'organizzazione di un geodatabase multistrato che potesse fungere da base di riferimento, continuamente aggiornabile, e in grado di supportare le elaborazioni pilota. È stato, quindi, possibile eseguire i primi test da cui ricavare apposite informazioni sulla validità delle procedure in atto e sulla possibilità di avviare, mediante geocoding, operazioni di geolocalizzazione puntuale, in grado sia di restituire la geovisualizzazione su mappa delle voci selezionate, sia di adottare ulteriori misure *ad hoc* volte ad accrescere l'accuratezza delle procedure qualora le operazioni di geocoding avessero evidenziato situazioni «anomale». In particolare, scelte convenzionalmente alcune regioni come prime aree di studio esemplificative e selezionate specifiche voci per validare le procedure e vagliare i primi prodotti di output, la geolocalizzazione è stata eseguita su base satellitare, dopo aver effettuato una sovrapposizione con le basi territoriali dell'Istat a scala comunale e provinciale, così da ricondurre ogni elemento all'interno dei confini amministrativi di pertinenza. Ogni elemento geolocalizzato e rappresentato come primitiva punto porta con sé tutti i metadati registrati nelle tabelle degli attributi che fungono da input e questo consente una consultazione dinamica, che restituisce informazioni, ad esempio, mediante pop-up di sintesi.

Le procedure adottate, ricorrendo anche alla programmazione in Python per superare alcune problematiche emerse, e l'architettura creata sono state concepite in modo da poter replicare i passaggi e l'ambiente di lavoro per ogni regione d'Italia. La vestizione grafica, che ha cercato di mantenere quella originale dell'*Atlante* in un connubio con i canoni della cartografia digitale e con la simbolizzazione messa a disposizione da ArcGIS Pro, è stata definita in maniera da permettere un'efficace resa restitutiva a livello di geovisualizzazione e consultazione interattiva.

Per successivi sviluppi, il geodatabase organizzato in strati multipli georeferenziati e armonizzati consente di operare verso l'integrazione di fonti plurime, coniugando in un unico ambiente documenti testuali, cartografie di diverso periodo, fotografie, testimonianze espresse mediante file audio e video, così da concepire soluzioni geotecnologiche interdisciplinari che non siano la somma di prodotti più o meno «forzatamente» amalgamati, bensì il risultato di un'euritmica condivisione avvenuta a monte e portata avanti di comune accordo, sposando stessi obiettivi e step operativi comunemente avallati, conoscendo ciascun gruppo le esigenze degli altri e quindi anticipando o evitando situazioni di «cortocircuito».

In questo contributo vengono presentate alcune tavole esemplificative che riproducono in chiave moderna e in formato digitale, dinamicamente inter-

rogabile, quelle contenute nella versione cartacea dell'*Atlante Linguistico Italiano* e che possono essere messe a confronto utilizzando effetti specifici, quali ad esempio lo swipe che permette di scorrere una sorta di tendina in senso ovest-est, mostrando gradualmente due strati informativi georeferenziati e sovrapposti. Vengono quindi mostrate alcune elaborazioni, a livello regionale e nazionale, che sono il risultato di un percorso che ha portato a testare soluzioni operative in grado di conferire all'*Atlante Linguistico Italiano* un volto nuovo fondato su quello originale.

In un'ottica di progettualità in rete e veicolazione dei risultati e dei prodotti di output, il passaggio da ArcGIS Pro ad ArcGIS online o ad altre piattaforme di rete – appositamente pensato e in fase di discussione – renderebbe possibile un'opera di diffusa valorizzazione e messa a conoscenza dell'opera, che sino a ora – complice anche le poderose dimensioni di ciascun volume – risiede soltanto in alcune sedi.

Nella prospettiva di un sempre più saldo legame tra studi geografici e linguistici – che nel caso in oggetto già trova numerosi punti di incontro nei rilievi sul terreno, nel coinvolgimento attivo della popolazione mediante inchieste, nella condivisione di soluzioni geoinformatiche – la sovrapposizione e l'analisi combinata delle tavole dell'ALI tradotte in digitale e della cartografia storica georeferenziata possono supportare studi dettagliati e «inesplorati» dei territori, alla ricerca di chiavi di lettura che consentano di spiegare, con approccio intersettoriale e diacronico, aspetti concernenti:

- i diversi tipi di paesaggio;
- l'identità locale;
- i flussi e i processi di contaminazione succedutisi, talvolta favoriti o ostacolati da elementi fisico-morfologici (corsi d'acqua, rilievi montuosi, zone paludose ecc.);
- la toponomastica e le minoranze linguistiche, dando tra l'altro risalto a rilevanti elementi della varietà linguistica in Italia.

Nel 2008, Paolo Di Giovine (p. 17), Direttore dell'allora Dipartimento delle Scienze dei Segni, degli Spazi e delle Culture della Sapienza Università di Roma (poi confluito nell'odierno Dipartimento di Lettere e Culture moderne), così si era espresso durante il convegno «Flussi, Reti e Sistemi Informativi Geografici, tra presente e futuro», organizzato dalla Sezione di Geografia: «In generale, i Sistemi Informativi Geografici permettono di mettere in relazione le componenti fisico-territoriali, antropiche, economiche, culturali, ecc., per analizzare in modo più approfondito fenomeni di grande importanza». E ancora: «Questi strumenti, infatti, costituiscono e conservano, per un linguista qual è chi vi parla, la memoria storica degli eventi e consentono di svolgere indagini rigorose. Inoltre, nelle ricerche di linguistica, le rappresentazioni cartografiche di tale tipo possono offrire le chiavi per individuare e comprendere una serie di fenomeni in funzione della loro distribuzione diatopica». Rilette a distanza di parecchi anni, tali considerazioni sembrano aver aperto la strada alla migrazione in ambiente GIS dell'*Atlante Linguistico Italiano* e preannunciato i possibili sviluppi.

Del resto, nella stessa occasione, Franco Salvatori (2008, pp. 21-22), allora Presidente della Società Geografica Italiana, aveva sottolineato l'importanza di saper raccogliere sfide contrassegnate da profondi significati, all'interno delle quali approccio geografico e sistemi geotecnologici possono favorire «prospettive gigantesche di progresso» e innovazione scientifica, facendo «emergere aspetti profondi, dettagli e particolari connessioni».

In effetti, nata come sfida, questa progettualità condivisa, avente quale soggetto l'*Atlante Linguistico Italiano*, è stata da subito vista come un'occasione per amalgamare settori scientifico-disciplinari diversi in una ricerca applicata: una ricerca ispirata dal linguaggio della geo-graficità, con le sue metodologie e i suoi strumenti, così da dischiudere «frontiere di forte impatto [...], in grado di suscitare curiosità e desiderio di esplorare» (De Vecchis, 2008, p. 20) e dove risulta evidente «l'importanza di banche dati ben strutturate, con informazioni quantitative e qualitative, in grado di interfacciarsi con i principali *software* cartografici, per produrre elaborati di fondamentale supporto durante i vari passi delle ricerche» (Palagiano, 2008, p. 26). Proprio da qui si è, infatti, partiti, rendendo il geodatabase il «cuore pulsante» atto a guidare in modo armonico «la fase delle sovrapposizioni automatiche e delle comparazioni-correlazioni tra variabili» (Pesaresi, Pavia, 2025, p. 10).

Vari lavori che hanno indagato sul possibile valore aggiunto attribuibile ai GIS nelle ricerche di linguistica, mediante adozione di rigorose metodologie, hanno ad esempio sottolineato come questi possano rivelarsi strumenti di cruciale rilevanza per:

- studiare le relazioni tra dati linguistici e posizioni spaziali, attraverso funzionalità e tool relativi a geocodifica e georeferenziazione, che portano a una efficace e accurata geovisualizzazione dei dati e favoriscono interrogazioni dinamiche e analisi computazionali (Kretzschmar, Petrulovich, 2020);
- integrare approccio linguistico e tecniche geospaziali, contribuendo in modo significativo a indagini di dettaglio e all'analisi dei cambiamenti dialettali nel corso del tempo, in relazione con altri fattori connessi, come quelli sociali e culturali, così da far progredire lo stato dell'arte secondo nuove prospettive (Teerarojanarat, Tingsabath, 2011);
- realizzare atlanti digitali interattivi che possano non solo interessare ricercatori e studiosi, bensì coinvolgere un'ampia gamma di utenti, attraverso la diffusione in rete (webGIS) e con la possibilità di consultare materiali integrativi quali registrazioni audio e file video che rendano il prodotto accattivante e che facciano immaginare di trovarsi in una sorta di museo virtuale immersivo (Lowry, 2022).

Le attività sin qui condotte per la digitalizzazione dell'ALI e la sua migrazione in ambiente GIS si sono susseguite perseguendo tali prospettive e, attraverso apposite metodologie e procedure, hanno innanzi tutto portato alla:

- organizzazione di un flusso di lavoro fatto di step sequenziali e fasi di validazione;
- creazione di un geodatabase di riferimento continuamente implementato;
- realizzazione di una serie di prodotti di output intesi come mappature digitali collegate al geodatabase;

- discussione di specifici parametri per la valorizzazione in rete;
- possibile costruzione di story map che possano supportare la fase di disseminazione dei risultati e coinvolgimento su vasto raggio.

7. Strumenti e metodologia

La scelta della piattaforma ArcGIS come strumento per la migrazione dell'*Atlante* è stata mossa da molteplici fattori, riconducibili alle potenzialità dello strumento geoinformatico e alla relativa dotazione, in seno alla Sapienza, di uno *stock* di licenze tale da permetterne il pieno utilizzo negli ambienti di lavoro web e desktop. Si fa riferimento all'ampia gamma di prodotti abilitati dal *license agreement* finanziato dal progetto «InDAGIS-MODE&APP» (Pesaresi, 2022), a cominciare dall'applicazione che funge da perno dei sistemi informativi sviluppati sull'architettura di Esri Inc., ovvero da ArcGIS Pro.

Il software GIS di punta della compagnia statunitense – oggetto di specifici approfondimenti dedicati alla gestione dei *coordinate system* (Pavia, 2023a) e allo sviluppo di algoritmi in Python (Pavia, 2023b) – eredita le funzionalità del suo predecessore ArcMap con importanti innovazioni che riguardano, tra l'altro, l'integrazione tra l'applicazione e le restanti componenti della piattaforma ArcGIS, agevolando la fruizione dei geodati tramite i motori di ricerca o l'utilizzo di dispositivi mobili come gli *smartphone*.

La possibilità di pubblicare contenuti in un ambiente di destinazione di facile accesso come ArcGIS Online, basato sulle piattaforme *cloud* di Amazon (Web Services) e Microsoft (Azure) (Fu, 2022, p. 4), semplifica la configurazione di webGIS volti a visualizzare e interagire con i *layer* predisposti dall'utente tramite ArcGIS Pro, senza necessità di uscire dall'ambiente di lavoro principale della piattaforma.

Per gli obiettivi del progetto di ricerca DigitALI, tesi alla possibilità di pubblicare le voci d'inchiesta e i relativi informatori in una mappa digitale e interattiva, il potenziale della piattaforma ArcGIS è apparso il più adeguato allo sviluppo di un sistema informativo multiscala e multistrato, capace di cambiare forma e contenuto in relazione alle caratteristiche dell'*audience* di riferimento. Concretamente, questo significa riuscire a predisporre una o più mappe web condivisibili all'interno di web app come le già citate story map, adatte a divulgare i contenuti dell'*Atlante* in modo più accessibile e multimediale.

La prima fase del lavoro è consistita quindi nella predisposizione, in ambiente ArcGIS Pro, dei dati relativi al sottoinsieme utilizzato per la sperimentazione della migrazione dell'*Atlante*, estrapolati dalla precedente base dati e successivamente convertiti nei formati compatibili con i geodatabase di Esri.

Nel caso dei punti d'inchiesta, corrispondenti ai luoghi in cui si sono svolti i questionari, si è proceduto alla georeferenziazione delle coordinate delle righe appartenenti alla tabella originale, espresse nel sistema di riferimento «WGS 84» (EPSG: 4326) e declinate in gradi decimali. L'operazione ha generato una classe di punti vettoriali contenente 1.004 geometrie, per la stra-

grande maggioranza (96%) comprese all'interno dei confini del Paese¹. L'interdistanza media tra i suddetti punti è pari a circa 12,16 chilometri², mentre la somma dei punti d'inchiesta a scala regionale è riportata nella tab. 1.

Tab. 1 – Punti d'inchiesta dell'ALI per regione italiana.

1	Sardegna	109
2	Piemonte	71
3	Sicilia	69
4	Toscana	67
5	Emilia-Romagna	63
6	Lombardia	59
7	Veneto	58
8	Friuli-Venezia Giulia	53
9	Abruzzo	52
10	Calabria	52
11	Marche	51
12	Puglia	48
13	Lazio	42
14	Campania	39
15	Trentino-Alto Adige	32
16	Umbria	27
17	Basilicata	25
18	Liguria	18
19	Molise	14
20	Valle d'Aosta	11

Fonte: elaborazione su dati: Istat, ALI.

¹ 44 dei 1.004 punti si trovano fuori dai confini nazionali scaricati dalla pagina «World Bank Official Administrative Boundaries» (<https://datacatalog.worldbank.org/search/dataset/0038272/World-Bank-Official-Boundaries>, ultimo accesso: 27/06/2025). Essi risultano distribuiti nel modo seguente: Croazia (32); Slovenia (10); Francia (1); San Marino (1).

² L'interdistanza media è stata calcolata con il geoprocesso «Generate Near Table» (<https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/generate-near-table.htm>, ultimo accesso: 30/06/2025) che, per i singoli punti d'inchiesta, ricerca il punto più vicino tra le rimanenti 1.003 geometrie, per poi effettuare il calcolo della distanza tra la coppia utilizzando il metodo geodetico, dunque rispetto all'ellissoide di riferimento del *coordinate system* del dato di input («WGS 84»).

La classe puntuale è stata poi inserita in una mappa di ArcGIS Pro, al fine di permettere una prima associazione tra il suo *layer* vettoriale e le tabelle *standalone* dove si trovano le voci dell'*Atlante* e le anagrafiche dei loro informatori. Diversamente dalle associazioni che si svolgono all'interno dei RDBMS (*Relational DataBase Management System*), la cui presenza è ritenuta imprescindibile per la definizione stessa di qualunque base dati (Favretto, 2018, p. 485), si è proceduto utilizzando i geoprocessi «Add Join» e «Add Relate»³, tramite cui riuscire ad associare le tabelle «virtualmente» rimanendo nell'ambiente di lavoro della mappa, prendendo in considerazione anche quelle salvate fuori dai DB relazionali come, ad esempio, i DBF che costituiscono l'*attribute table* degli *shapefile*. Ciò ha consentito di condurre una verifica preliminare del funzionamento delle chiavi necessarie al compimento dell'associazione, utilizzando i due *tool* di *geoprocessing* sulle tabelle originali salvate all'interno di specifiche cartelle di lavoro Excel.

L'associazione tra i punti d'inchiesta, gli informatori e le voci dell'ALI ha tuttavia richiesto la creazione di una serie di colonne, dove memorizzare, tramite l'uso di algoritmi nei linguaggi di programmazione Arcade e Python, degli attributi alfanumerici formati dalla concatenazione tra le informazioni disponibili nei campi originali come «punto», «codice» e «str», il cui significato è stato già chiarito nel primo paragrafo del contributo. Al termine dell'elaborazione, i campi chiave popolati hanno permesso di associare il *layer* dei punti d'inchiesta sia con la tabella degli informatori (1:M)⁴, sia con le voci dell'*Atlante* (1:1).

La prima associazione ha avuto un obiettivo di carattere squisitamente cartografico, volto a integrare gli attributi dei punti d'inchiesta con i dati relativi alla pronuncia delle voci dell'*Atlante*, nell'ottica di farle comparire dentro apposite etichette da mostrare a fianco delle geometrie. Il *tool* utilizzato è stato quindi «Add Join», in considerazione della sua capacità di aggiungere allo schema originale dei punti d'inchiesta le colonne relative alla tabella delle voci dell'*Atlante*. Come ricorda Corbin (2015, p. 101), l'effetto di una associazione che si svolge con il *tool* «Add Join», così come con «Add Relate», viene salvato unicamente nella mappa di ArcGIS Pro, senza lasciare traccia nella base dati di riferimento del progetto GIS.

Agendo come se fossero parte dello stesso schema, i campi delle due tabelle vengono così messi a disposizione dell'utente per lo svolgimento di svariate operazioni: dalla simbologia alle *query* tabellari, dai geoprocessi all'etichettatura delle geometrie. Come anticipato, lo scopo dell'associazione tra le geometrie puntuali e le voci dell'ALI è stato quello di mostrare, a fianco

³ Per approfondimenti sulle funzionalità dei geoprocessi succitati si rimanda alla consultazione delle pagine seguenti: <https://pro.arcgis.com/en/pro-app/3.4/tool-reference/data-management/add-join.htm>; <https://pro.arcgis.com/en/pro-app/3.4/tool-reference/data-management/add-relate.htm> (ultimo accesso: 01/07/2025).

⁴ La parentesi annotata contiene il riferimento alla tipologia di cardinalità tra le tabelle oggetto dell'associazione. Nel caso in questione, la cardinalità 1:M implica l'associazione tra un punto d'inchiesta e *n* informatori dell'*Atlante*.

dei punti d'inchiesta, un'etichetta relativa alla pronuncia di ciascuna voce derivandola dalla colonna «html». In considerazione della variabilità fonetica delle voci mappate, il contenuto è stato quindi formattato con i marcatori HTML grazie a una espressione in Python che, concatenando i verbi e i sostantivi di ciascuna voce (i.e. «cullare», «culla»), ha consentito di visualizzare la formattazione originale con l'aggiunta di determinati *tag*⁵.

Con «Add Relate» si è invece proseguito alla creazione di un'associazione tra l'*attribute table* dei punti d'inchiesta e quella degli informatori, nell'ottica di rendere possibile ai lettori della mappa il riconoscimento dei soggetti delle inchieste svoltesi sul territorio, mostrandone il nome e cognome, la professione e gli attributi di carattere anagrafico raccolti in fase di stesura dell'*Atlante*. Diversamente dall'«Add Join», l'associazione che il processo crea tra le tabelle si palesa solo nel momento in cui l'utente, utilizzando i *tool* di selezione disponibili («Select By Attributes», «Select By Location»), richiede ad ArcGIS Pro la visualizzazione delle righe degli informatori che, per la presenza di attributi uguali nei due campi chiave, risultano associate a quelle dell'*attribute table* dei punti d'inchiesta oggetto della selezione.

Oltre al procedimento sopramenzionato, esiste una modalità di accesso alternativa che si svolge con «Explore», strumento di ArcGIS Pro tramite il quale navigare nelle mappe (o scene) dell'applicazione e consultare gli attributi della cella o geometria identificata con un click del mouse. Cliccando sui punti d'inchiesta, l'utente visualizzerà un pop-up dove compariranno tanto gli attributi delle geometrie puntuali quanto quelli relativi ai dati agli informatori che risultano associati al punto. In considerazione della rilevanza dei pop-up per la lettura e l'interpretazione delle mappe digitali, l'applicazione di Esri Inc. presenta un'ampia gamma di parametri di customizzazione che consentono l'arricchimento dei dati di base con l'ausilio di specifici formati testo e file multimediali.

In questo senso, il prossimo paragrafo si focalizzerà sull'importanza della configurazione dei pop-up per il progetto di ricerca DigitALI, mostrando alcune esemplificazioni derivanti dalla sperimentazione delle funzionalità del software per la configurazione dei pop-up di livello avanzato.

8. *Dal lemma al parlante. Configurare il pop-up per valorizzare i protagonisti dell'inchiesta*

Un elemento innovativo nel processo di digitalizzazione dell'ALI risiede nella capacità di conferire valore al contributo dei parlanti, non solo rispetto al repertorio di risposte fornite ma anche in merito a significative informa-

⁵ Si fa riferimento al marcatore , utilizzato per modificare lo stile del testo delle viste (e.g. pop-up, legenda) delle applicazioni Esri compatibili con il linguaggio di programmazione HTML (cfr. <https://community.esri.com/t5/classic-esri-story-maps-questions/questions/within-hosted-arcgis-online-how-do-i-label-my/td-p/134273>, ultimo accesso: 01/07/2025).

zioni biografiche. Lungi dal costituire una mera appendice nella conduzione dell'inchiesta, l'annotazione dei «dati anagrafici e socioculturali» ad opera di Ugo Pellis e dei successivi raccoglitori ha rappresentato un passaggio essenziale nel raffinare «l'interpretazione dei dati linguistici» (Cugno *et alii*, 2020, p. 195). Da questo punto di vista, infatti, si può affermare che il prototipo di cartografia digitale che deriva da tale approccio al dato geolinguistico riesca a coniugare la distribuzione dei lemmi – già valorizzata dalla versione a stampa – con le preziose notazioni custodite nei verbali redatti dai raccoglitori a margine dei rilievi sul campo (Massobrio *et alii*, 1995), caratterizzati da una sezione ideata appositamente per il parlante e intitolata «Informatore».

L'associazione tra le *attribute table* relative ai punti, alle singole voci e ai parlanti consente, infatti, di mettere in luce per ognuno dei punti un pop-up che presenta tutte le informazioni descrittive connesse al lemma e al profilo dell'informatore o degli informatori coinvolti. Valorizzare la metadattazione dei parlanti costituisce un'attività inedita nel panorama degli atlanti linguistici digitali, poiché permette di salvaguardare quella dimensione quali-quantitativa che coniuga da una parte la cristallizzazione fonetica del lemma e dall'altra le peculiarità dello «spirito del parlante» come principio ispiratore degli studi dialettologici di matrice terraciniana (Covino, 2014, p. 4).

Venendo alla configurazione del pop-up, nell'ottica di predisporre un atlante digitale fruibile non solo dagli addetti ai lavori ma anche da tutto il pubblico a vario titolo interessato allo studio dei dialetti si è proceduto a modificare la veste predefinita nella visualizzazione degli attributi. A partire dalla semplice visualizzazione degli attributi tabellari, si è avviata l'elaborazione di una veste discorsiva attingendo alle varie opzioni offerte dalla funzionalità «Configure pop-ups» del software ArcGIS Pro. Cliccando sulle icone disponibili all'interno della maschera di configurazione, si aprono una serie di finestre contenenti varie scelte di formattazione e di impaginazione per il pop-up, tra le quali si è optato per una disposizione orizzontale delle due tabelle selezionando l'opzione «split into two horizontally stacked elements».

Relativamente alla formattazione del testo ci si è avvalsi della funzionalità «Text», tenendo conto di una duplice finalità. Innanzi tutto l'operazione ha permesso di selezionare solo i campi effettivamente utili per condurre la ricerca geolinguistica: per la prima tabella – formata dall'associazione tra i campi dei punti d'inchiesta e quelli relativi alle voci dell'ALI – il pop-up contiene la trascrizione del lemma con caratteri IPA (campo «html»), il numero dell'inchiesta (campo «inchiesta»), il periodo (campo «anno») e il raccoglitore (campo «raccoglitore»); per la seconda – formata dall'associazione tra i campi delle voci dell'ALI e quelli degli informatori – il contenuto del pop-up richiama il nome dell'informatore (campo «nome»), l'anno di nascita (campo «dob»), l'età (campo «eta»), lo status lavorativo (campo «professione») e il livello d'istruzione (campo «istruzione»). In secondo luogo, cliccata l'opzione di formattazione, dall'interno della finestra «Text options» si è provveduto a selezionare ogni campo funzionale alla ricerca attraverso un menù a tendina contenente i «Fields», per immetterli all'interno di una maschera testuale statica in qualità di espressioni dinamiche che cambiano a seconda del punto d'inchiesta selezionato dall'utente. Tra-

mite le opzioni di formattazione offerte dalla funzionalità si è optato per una visualizzazione in grassetto degli attributi, al fine di agevolare la consultazione dei dati. Avere contezza istantaneamente di età, professione e livello di istruzione rappresenta un fondamentale passaggio nella consultazione dell'ALI, giacché ciò consente di mettere in relazione alle voci i contenuti tabellari fino a ora confinati nei volumi a stampa dei *Verbali d'inchiesta*, inscrivendo in un ambiente unico informazioni funzionali allo studio delle varie tipologie di variazione dialettale. Si pensi, ad esempio, alla variazione diastratica che è strettamente connessa con l'inquadramento sociale del parlante all'interno della collettività.

Per meglio comprendere l'allestimento della visualizzazione discorsiva dei risultati d'inchiesta, si illustra a titolo esemplificativo (tab. 2) il processo di configurazione del pop-up relativo alla carta «culla» nel comune di San Michele di Ganzaria (punto 1049), in provincia di Catania.

Tab. 2 – Configurazione del pop-up con dati relativi a risposta e informatore.

<i>Visualizzazione del pop-up in fase di compilazione tramite funzionalità «Text»</i>	<i>Risposta</i> Lemma {html} relativo all'inchiesta {inchiesta} del {anno} condotta da {raccoglitore}	<i>Informatore</i> L'informatore è {nome}, nato nel {dob}, di {eta} anni. Svolge/ha svolto la professione di {professione} e rispetto al livello di istruzione si definisce/ha frequentato {istruzione}
<i>Visualizzazione del pop-up a seguito del salvataggio della maschera testuale</i>	<i>Risposta</i> Lemma □ nàka* relativo all'inchiesta #0672 del (1939) condotta da Pellis	<i>Informatore</i> L'informatore è Francesco Ugo Judicelli , nato nel 1913 , di 26 anni. Svolge/ha svolto la professione di insegnante elementare privato e rispetto al livello di istruzione si definisce/ha frequentato ginnasio e abilitazione privata magistrale

Fonte: elaborazione su dati ALI.

In questa fase di sperimentazione dell'applicativo digitale, i lavori di concatenazione delle tabelle dei dati linguistici nel processo di migrazione in ambiente ArcGIS – attività tutt'ora in corso – stanno proseguendo nel tentativo di calibrare la mole di dati frutto dell'informatizzazione delle inchieste linguistiche con le esigenze proprie dei sistemi informativi geografici.

I primi fruttuosi esiti di questo progetto di ricerca di carattere interdisciplinare sono espressi in maniera significativa dalle cartografie digitali elaborate, che saranno al centro dell'analisi del successivo paragrafo. Sino ad oggi presentate al pubblico solo in occasione di diversi convegni, le rappresentazioni risultano particolarmente immediate in termini di consultazione e coniugano tale caratteristica alla capacità acclarata del software di organizzare in maniera armonica una mole imponente di informazioni proveniente da molteplici fonti di diversa tipologia (Pesaresi, Pavia, 2025, p. 10).

9. Tavole esemplificative, a livello regionale e nazionale: verso un nuovo volto dell'Atlante Linguistico Italiano

La fig. 3A, incentrata sulla Sicilia, è la prima tavola prodotta a seguito dei vari step metodologico-operativi che hanno portato alla restituzione cartografica dei dati contenuti nell'*Atlante Linguistico Italiano*. La scelta della Sicilia è stata indotta da una serie di motivazioni legate, ad esempio, alla numerosità dei punti d'inchiesta (69) e all'insularità, che agevola la consultazione tecnica delle voci, senza rischio di mescolanze grafiche.

Utilizzando la *basemap* «World Terrain Base», appartenente alla cartografia di base disponibile agli utenti della piattaforma ArcGIS, lo strato informativo della voce «culla» appare su uno sfondo che, nei toni del grigio, restituisce una contestualizzazione dei caratteri morfologici dell'area in esame. A fianco a ogni punto del livello appaiono diverse classi di etichette, ciascuna delle quali riferite a un attributo differente della geometria. In primo luogo, viene visualizzato il codice del punto nel medesimo colore della primitiva, utilizzando una formattazione che risulterà corsiva per i fondi di «Tropea» e «Piccitto» e normale (tonda) per quanto concerne il «Pellis». All'etichetta concernente il codice si aggiunge poi la classe relativa ai sostantivi e ai verbi della voce «culla».

Dalla tavola si evince una certa regolarità e fittezza dei punti d'inchiesta – ad eccezione delle zone fortemente caratterizzate da presenza di rilievi (ed è per questo importante la sovrapposizione di un template che restituisce una visualizzazione degli elementi fisico-morfologici) – e, sulla scia delle trascrizioni e dello stile grafico dell'ALI cartaceo, si può osservare la distribuzione della voce «culla». Due *inset map*, una per la zona nord-occidentale dell'isola, l'altra per la zona orientale (a sud-ovest dell'edificio etneo), forniscono appositi zoom su aree esemplificative dove tale voce acquisisce connotati particolari.

La fig. 3B mostra come le forme dialettali siano spesso accompagnate da lettere maiuscole che costituiscono i riferimenti a diversi tipi di oggetti ed elementi (nel caso specifico la «culla») restituiti nell'ALI cartaceo tramite un apposito corredo iconografico di supporto.

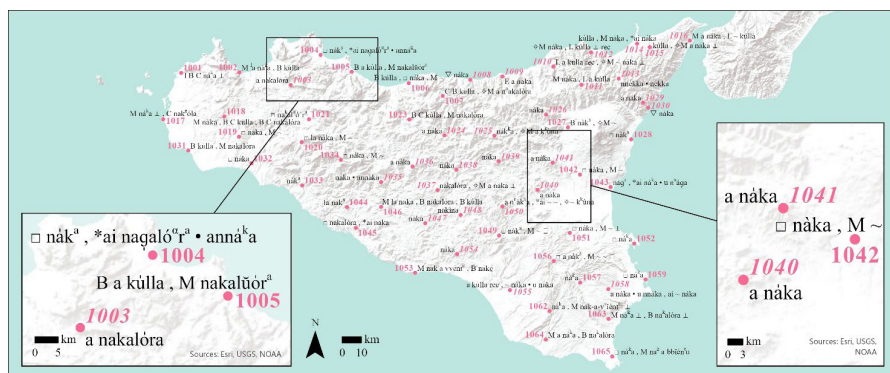


Fig. 3A – La mappa della voce «culla» relativa alla Sicilia con i simboli e le forme dialettali ricavati dall'ALI cartaceo.

Fonte: elaborazione su dati ALI, Esri.



Fig. 3B – Corredo iconografico di supporto che mostra diversi tipi di «culla» utilizzati come riferimenti, durante le campagne d'inchiesta, nell'ALI cartaceo.

Fonte: ALI.

La fig. 4 rappresenta, in sovrapposizione ai punti della voce «culla», i layer relativi ai limiti amministrativi comunali e provinciali estratti dalla base dati cartografica dell'Istat, così da favorire l'inquadramento delle singole voci d'inchiesta. Con un ridotto livello di zoom, e quindi a scala geografica minore, si visualizzano i confini provinciali con le relative sigle poste nel centroide dei relativi poligoni; con un maggiore livello di zoom, e quindi a scala più grande, compaiono le etichette dei comuni siciliani e le occorrenze della voce «culla» (intese come diverse forme dialettali).

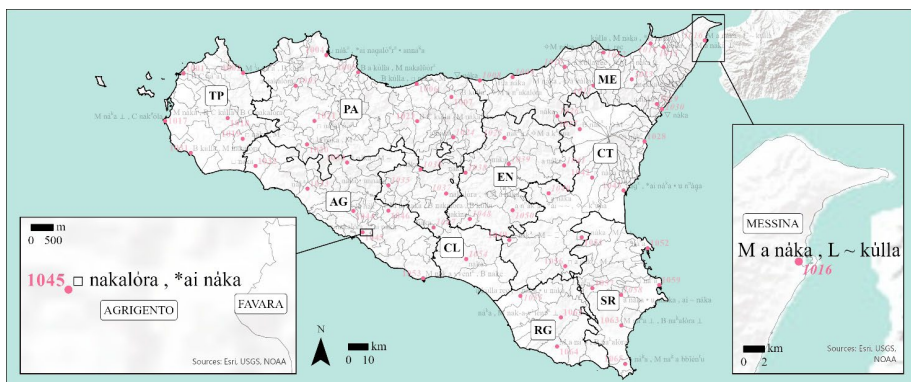


Fig. 4 – La mappa della voce «culla» relativa alla Sicilia con sovrapposti i limiti amministrativi comunali e provinciali.

Fonte: elaborazione su dati ALI, Esri, Istat.

La fig. 5 offre, dal suo canto, una rappresentazione dei punti d'inchiesta della Basilicata, in un *layout* di mappa che, ai lati dell'elaborazione principale, mostra due coppie d'*inset map* focalizzate sui comuni di Potenza-Tricarico e Pomarico-Moliterno. Al centro della mappa appaiono i punti d'inchiesta compresi all'interno dei confini regionali, dotati di attributi relativi alla pronuncia della voce «ombrello» in HTML. Questa tavola era stata prodotta in occasione dell'VIII Convegno Internazionale di Dialettologia – Progetto A.L.Ba, organizzato presso l'Università degli Studi della Basilicata (Potenza, 8-10 maggio 2025) e mostrata durante la presentazione *La digitalizzazione dell'Atlante Linguistico Italiano (DigitALI). Problemi e prospettive* (di Cerruti M., Ciaglia F., Di Giovine P., Ferrarotti L., Fiori S., Ghia A., Pavia D., Pesaresi C., Rivoira M., Turchetta B.).

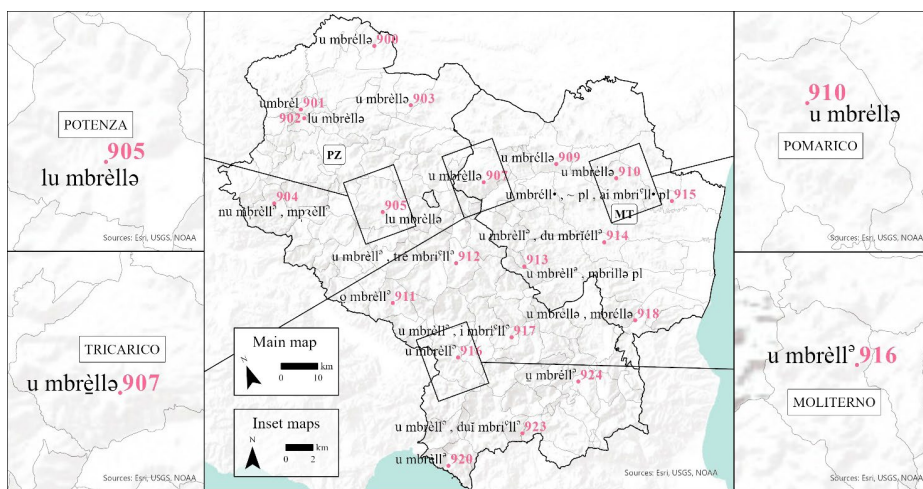


Fig. 5 – La mappa della voce «ombrello» relativa alla Basilicata.

Fonte: elaborazione su dati ALI, Esri, Istat.

La fig. 6 mostra, invece, le potenzialità dell'effetto *swipe*, utilizzato nelle applicazioni GIS per supportare l'analitica comparazione visiva tra diversi strati informativi. Tale effetto prevede che l'utente, selezionati i *layer* di proprio interesse, muova il cursore sullo schermo in senso orizzontale o verticale, così da vederli in sovrapposizione in un processo di graduale scoperta. Nel caso in questione, lo *swipe* è stato utilizzato sulla pagina dell'ALI e sulla pagina dell'A.L.Ba. (*Atlante Linguistico della Basilicata*) relativa alla pronuncia della voce «forbice», precedentemente scansionata e georeferenziata tramite l'aggiunta di una quantità di punti di controllo sufficiente ad applicare una trasformazione affine di prim'ordine polinomiale. L'effetto ha come risultato la comparsa dei punti d'inchiesta ALI riferiti alla medesima voce dell'A.L.Ba., agevolando la comparazione tra i due Atlanti in una sorta di *continuum* cartografico reso possibile dalla presenza dei confini comunali e regionali, visualizzabili a cavallo di ambedue le rappresentazioni.



Fig. 6 – L'effetto *swipe* visualizzato sulla mappa della voce «forbice» dell'ALI (a sinistra) e dell'A.L.Ba. (a destra).

Fonte: elaborazione su dati ALI, Esri, Istat, A.L.Ba.

Le figg. 7 e 8 estendono l'area mappata a tutto il territorio nazionale, visualizzando i 1.004 punti dell'*Atlante* che si trovano all'interno dei confini statuali estratti dalla banca dati della World Bank. La trama spaziale delle primitive e la scala di riduzione della mappa non hanno permesso in questo caso l'etichettatura dei punti d'inchiesta dell'*Atlante*, la quale avrebbe reso pressoché illeggibile la rappresentazione saturandola di oggetti grafici e geografici. La presenza delle «mappe insetto» è andata nuovamente incontro alla risoluzione del possibile problema di natura cartografica concernente l'eccessiva sovrapposizione di simboli e *label*; sfruttando i corpi idrici presenti nel *layout* di mappa per l'aggiunta di riquadri a scala medio-grande, le «mappe insetto» consentono di visualizzare i singoli punti d'inchiesta con le *label* relative al codice del punto e a una particolare voce dell'*Atlante*. Quelle che compaiono all'interno delle due figure a scala nazionale sono rispettivamente «meridiana» (fig. 7) e «(Condisci) con buon olio» (fig. 8), scelte per via dell'interesse che rivestono rispetto allo sviluppo di studi e ricerche sui temi cartografici e su quelli geografici inerenti all'ambiente, all'agricoltura e alla salute della popolazione.

La fig. 9 rappresenta il territorio del comune di Francavilla in Sinni (PZ), visualizzato sullo sfondo della *basemap* «World Imagery» a colori naturali.

Sovrapponendo ai contenuti del *layout* di mappa la schermata del pop-up, l'immagine restituisce una dimostrazione, anche se statica e parziale rispetto alle possibili interazioni con il sistema, dell'efficacia di questa finestra interattiva, tramite cui l'utente accede al retroterra biografico dei due informatori per lo svolgimento di ricerche geolinguistiche.

Per quel che riguarda i parlanti associati alla località lucana, con specifico riferimento a Orlando De Salvo e Ferrando De Benedictis, la fig. 10 mostra due *screenshot* delle corrispondenti espressioni testuali, dove è possibile visualizzare il contenuto di alcuni campi tabellari dinamici riportati in grassetto, concatenati con parti di testo statico in formato normale (tondo) volte a rendere più fluida e discorsiva la lettura.

La varietà toponomastica, le particolarità dei dialetti e la relativa distribuzione territoriale, le testimonianze delle minoranze linguistiche, i rapporti che si istaurano tra lingua e dialetti costituiscono un grande patrimonio da tutelare e valorizzare (Cantile, 2004; Arca, 2004), anche mediante forme di cartografia digitale che riescano a contenere, esplicitare e veicolare una simile ricchezza.

«I toponimi che si trovano sullo stesso piano l'uno accanto all'altro hanno in molti casi origini diverse per profondità cronologica e appartenenza culturale: vanno quindi interpretati secondo una lettura stratigrafica che individui l'epoca storica, la società e l'etnia che li ha fissati. [...] e spesso i toponimi rappresentano l'unica testimonianza ancora visibile di etnie e culture ormai cancellate dal tempo» (Nocentini, 2004, p. 698). Alcuni toponimi, specialmente quelli concernenti le forme del suolo e alcuni caratteri distintivi del passato sono andati progressivamente perdendosi (Arena, 2004, p. 702), a causa: dell'intensa crescita delle superfici artificiali; delle opere condotte per sanare territori e recuperare spazi da destinare ad attività economico-commerciali; delle infrastrutture di traffico e collegamento che contribuiscono a radicali cambiamenti; dei sempre maggiori flussi e scambi che coinvolgono e mettono in stretta relazione genti e luoghi diversi (cancellando vecchie tracce e creandone di nuove). Più ampio è l'arco temporale di riferimento maggiori sono le possibilità che si siano verificate tangibili modificazioni. Le dinamiche di ridistribuzione della popolazione (ad esempio tra aree interne e costiere), le configurazioni territoriali connesse con il passaggio verso il sistema industriale, prima, e la diffusione del settore terziario e quaternario, poi, fungono da motori di profondo cambiamento e incidono «sull'apparato denominativo del territorio, che non poteva rimanere estraneo ai suddetti mutamenti di personalità funzionale: una serie di denominazioni sono scomparse, in particolare microtoponimi legati agli usi tradizionali del mondo rurale, altre sono nate, in parte sostituendo designazioni precedenti e in parte configurandosi come coniazioni *ex nihilo*» (Cassi, 2004, pp. 725-724).

Quanto fin qui svolto e in parte presentato con questo lavoro, volto a mostrare i primi lineamenti del nuovo volto dell'*Atlante Linguistico Italiano*, mediante processi di raccordo e valorizzazione in ambiente ArcGIS, è stato animato dal corale desiderio di conservare e far conoscere elementi peculiari e distintivi, «fissandoli» sul territorio e «restituendoli» su un sistema interattivo di geovisualizzazione e consultazione dinamica.

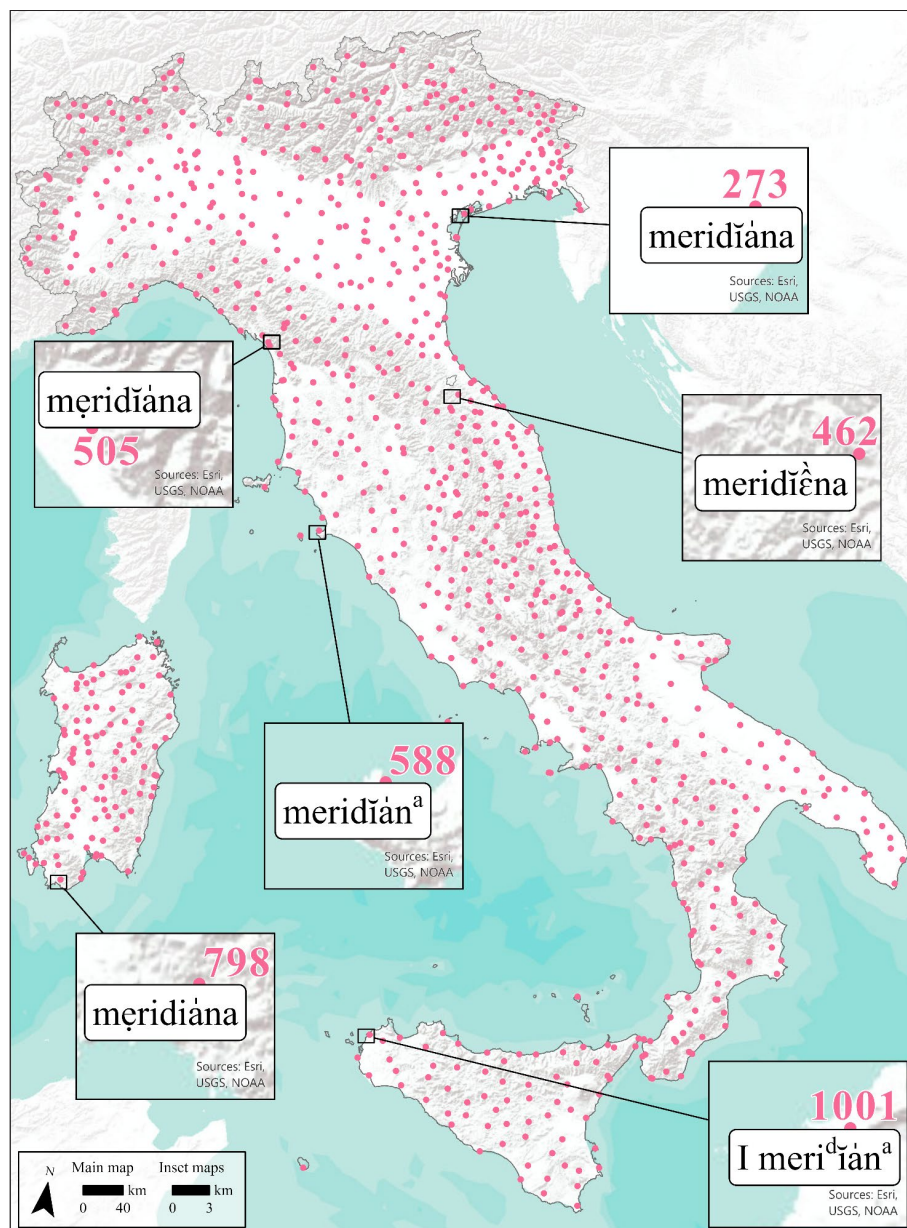


Fig. 7 – I punti d'inchiesta dell'ALI con zoomate esemplificative della voce «meridiana».

Fonte: elaborazione su dati ALI, Esri, Istat.

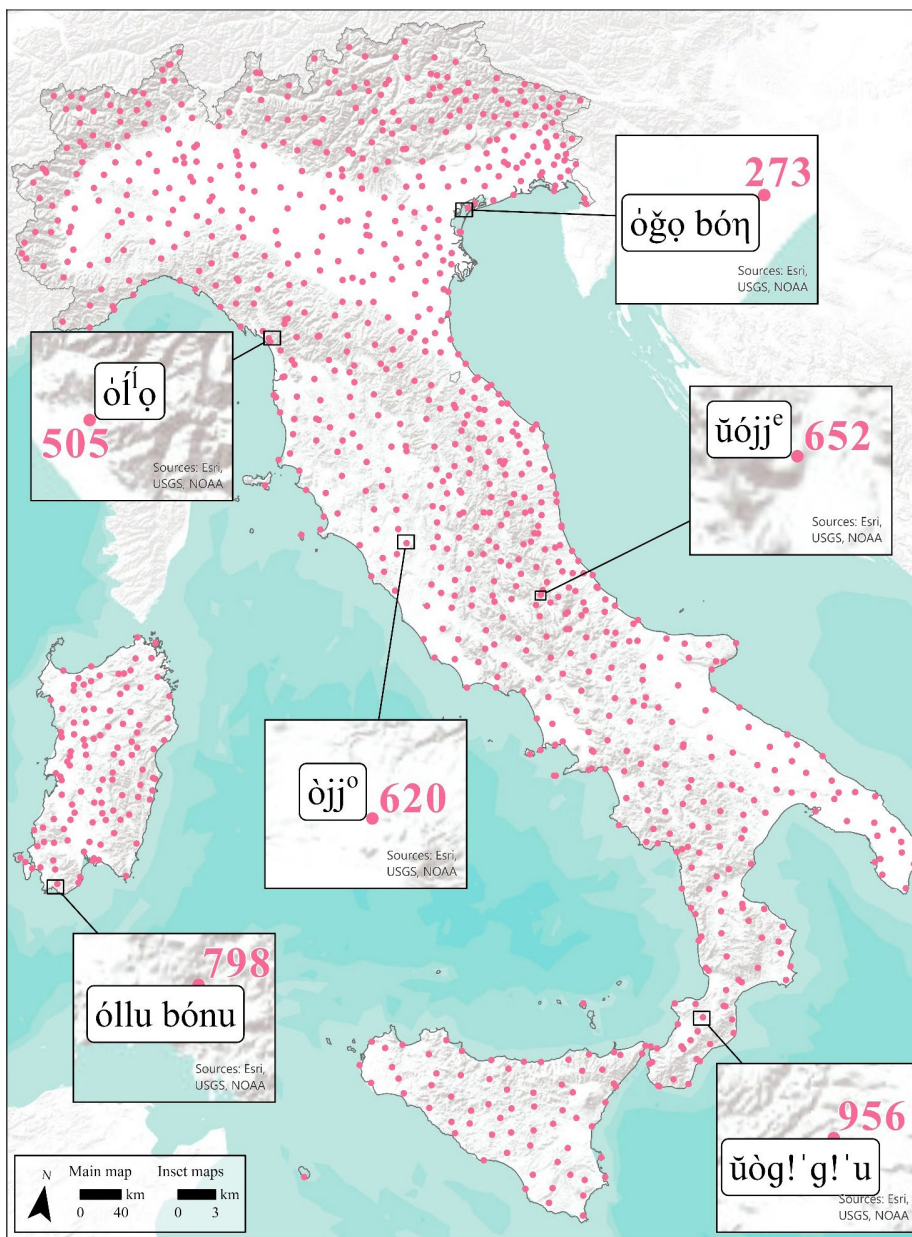


Fig. 8 – I punti d'inchiesta dell'ALI con zoomate esemplificative della voce «(Condisci) con buon olio».

Fonte: elaborazione su dati ALI, Esri, Istat.

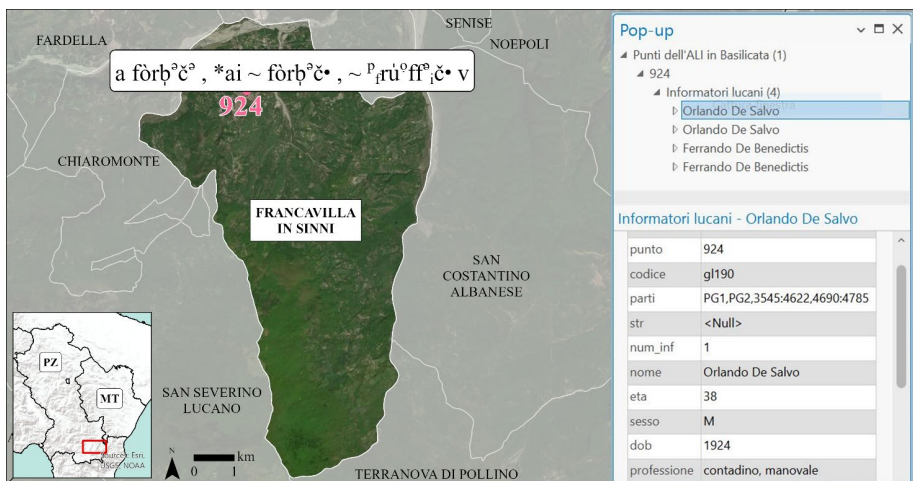


Fig. 9 – La mappa della voce «forbice» del comune di Francavilla in Sinni (PZ) con il suo pop-up. In linea generale, la duplicazione dei nomi degli informatori può derivare da (eventuali) plurime risposte da loro fornite e da aspetti tecnici riguardanti la schedatura elettronica delle risposte.

Fonte: elaborazione su dati ALI, Esri, Istat.

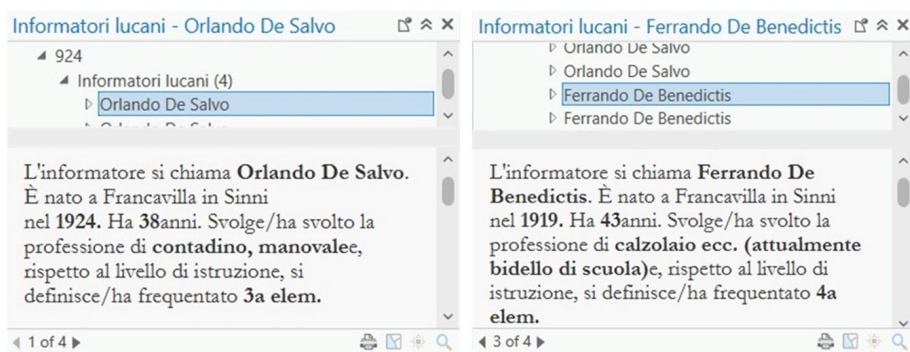


Fig. 10 – I dati relativi agli informatori Orlando De Salvo e Ferrando De Benedictis, entrambi associati al comune di Francavilla in Sinni (PZ), restituiti dall'espressione testuale elaborata nel pop-up. La duplicazione dei nomi degli informatori deriva da plurime risposte fornite e da aspetti tecnici riguardanti la schedatura elettronica delle risposte.

Fonte: elaborazione su dati ALI.

Bibliografia

- ABETE G., RIVOIRA M., «Il sistema di trascrizione dell'Atlante Linguistico Mediterraneo tra principi teorico-metodologici e prassi trascrittoria», in *Bollettino dell'Atlante Linguistico Mediterraneo* 1 (Nuova serie), 2021, pp. 143-161.
- ARCA S., «Normative internazionali di nomi geografici», in IGM (a cura di), *Italia. Atlante dei tipi geografici*, Firenze, Istituto Geografico Militare, 2004, pp. 80-82.

- ARENA G., «Denominazioni varie attribuite alle forme del terreno», in IGM (a cura di), *Italia. Atlante dei tipi geografici*, Firenze, Istituto Geografico Militare, 2004, pp. 702-709.
- BARTOLI M., PELLIS U., «Atlante Linguistico Italiano. Questionario», in GENRE A., CAMPAGNA S., MASSOBRIO L. (a cura di), *Bollettino dell'Atlante Linguistico Italiano. Supplementi*, 1971.
- CANTILE A., «Norme toponomastiche nazionali», in IGM (a cura di), *Italia. Atlante dei tipi geografici*, Firenze, Istituto Geografico Militare, 2004, pp. 83-89.
- CASSI L., «Nuovi toponimi», in IGM (a cura di), *Italia. Atlante dei tipi geografici*, Firenze, Istituto Geografico Militare, 2004, pp. 722-729.
- CERRUTI M., REGIS R., «La tipizzazione lessicale: problemi e metodi», in ALEPO (Atlante Linguistico ed Etnografico del Piemonte Occidentale) I, Scarmagno, Priuli & Verlucca, 2008 pp. 15-59.
- COLCUC B., ZACHERL F., «Tools for data processing and visualization in the project VerbaAlpina», in PALLIWODA N. (a cura di), *Data Processing and Visualization in Variational Linguistics/Dialectology* (Working Papers in Corpus Linguistics and Digital Technologies: Analyses and Methodology 7), Szeged-Hamburg, University of Szeged, 2022 pp. 55-70.
- CORBIN T., *Learning ArcGIS Pro. Create, analyze, maintain, and share 2D and 3D maps with the powerful tools of ArcGIS Pro*, Birmingham, Packt Publishing, 2015.
- COVINO S., «Benvenuto Terracini, Bruno Migliorini e la linguistica europea del Novecento», in *Vox Romanica*, 73, 2014, pp. 1-16.
- CUGNO F., RIVOIRA M., RONCO G., «L'Atlante Linguistico Italiano (ALI)», in *Romance Philology*, 74, 2020, pp. 191-215.
- DE VECCHIS G., «Indirizzo di saluto», in PESARESI C. (a cura di), *Flussi, Reti e Sistemi Informativi Geografici, tra presente e futuro, Semestrale di Studi e Ricerche di Geografia*, 2, 2008, pp. 19-20.
- DI GIOVINE P., «Indirizzo di saluto», in PESARESI C. (a cura di), *Flussi, Reti e Sistemi Informativi Geografici, tra presente e futuro, Semestrale di Studi e Ricerche di Geografia*, 2, 2008, p. 17.
- FAVRETTO A., «Cartografia per il turismo: uso di immagini remote», in *Bollettino dell'Associazione Italiana di Cartografia*, 126-127-128, 2006, pp. 59-70.
- FAVRETTO A., «GIS e Database Geo-spaziali. Un contributo per evitare di confonderli», in Atti Conferenza ASITA 2018 (Bolzano, 27-29 novembre 2018), 2018, pp. 485-488.
- FAVRETTO A., *Cartografia in evoluzione*, Bologna, Pàtron Editore, 2025.
- FORKEL R., MORAN S., *cldf/segments: Unicode Standard tokenization (version v1.2.2)*, Zenodo, 2018, <<https://github.com/cldf/segments/tree/v1.2.2>>.
- FU P., *Getting to Know Web GIS*, Redlands, Esri Press, 2022.
- GEISLER H., FORKEL R., LIST J.M., «A digital, retro-standardized edition of the *Tableaux Phonétiques des Patois Suisses Romands* (TPPSR)», in THIBAUT, A., AVANZI M., LO VECCHIO M., MILLOUR A. (a cura di), *Nouveaux regards sur la variation dialectale – New Ways of Analyzing Dialectal Variation*, Strasbourg, Éditions de Linguistique et de Philologie, 2021, pp. 13-36.

- GENRE A., MASSOBRIO L., «Progetto di pubblicazione dei materiali dell'Atlante Linguistico Italiano con particolare riguardo all'Indice», in *Bollettino dell'Atlante Linguistico Italiano*, 1, 1976, pp. 78-92.
- GHIA A., «Verso il DigitALEIC: riflessioni sulla trasformazione dei materiali fonetici della grafia fonetica dell'ALEIC all'IPA», in *Bollettino dell'Atlante Linguistico Italiano*, 47, 2023, pp. 109-142.
- KREFELD T., LÜCKE S. (a cura di), *Berichte aus der digitalen Geolinguistik* (Korpus im Text 6), 2018, <kit.gwi.uni-muenchen.de/?p=4498&v=2>.
- KREFELD T., LÜCKE S., «FAIRNESS – Medien im methodologischen Zentrum der Geolinguistik, Version 3 (22.06.2023, 16:20)», in KREFELD T., LÜCKE S., MUTTER C. (a cura di), *Berichte aus der digitalen Geolinguistik (II): Vernetzung und Nachhaltigkeit* (Korpus im Text 9), 2023, <<https://www.kit.gwi.uni-muenchen.de/?p=45335&v=3>>.
- KRETZSCHMAR W.A., PETRULEVICH A., «GIS for Language Study», in SCHUSTER K., DUNN S. (a cura di), *Routledge International Handbook of Research Methods in Digital Humanities*, Routledge, 2020, pp. 218-236.
- LIST J.M., TJUKA A., BLUM F., KUČEROVÁ A., UGARTE C.B., RZYMSKI C., GREENHILL S., FORKEL R. (a cura di), *CLLD Concepticon 3.4.0* [Data set], Zenodo, 2025.
- LOPORCARO M., *Profilo linguistico dei dialetti italiani*, Roma-Bari, Laterza, 2013.
- LOPORCARO M., SCHMIDT S., ZANINI C., PESCARINI D., DONZELLI G., NEGRINELLI S., TISATO G., «AIS, reloaded: a digital atlas of Italy and southern Switzerland», in THIBAUT A., AVANZI M., LO VECCHIO M., MILLOUR A. (a cura di), *Nouveaux regards sur la variation dialectale / New Ways of Analyzing Dialectal Variation*, Strasbourg, Éditions de linguistique et de philologie, 2021, pp. 111-136.
- LOWRY J.H., «Potential Applications of GIS for Linguistic Data», in *Senri Ethnological Studies*, 108, 2022, pp. 59-74.
- MASSOBRIO L., RONCO G., NOSENGO M.C., TUNINETTI G. (a cura di), *Atlante Linguistico Italiano. Verbali delle inchieste*, 2 voll., Roma, Istituto Poligrafico e Zecca dello Stato, 1995.
- MORAN S., CYSOUW M., *The Unicode cookbook for linguists: Managing writing systems using orthography profiles* (Translation and Multilingual Natural Language Processing 10), Berlin, Language Science Press, 2018.
- NOCENTINI A., «Toponimi italiani: origine ed evoluzione», in IGM (a cura di), *Italia. Atlante dei tipi geografici*, Firenze, Istituto Geografico Militare, 2004, pp. 698-701.
- PALAGIANO C., «I Sistemi Informativi Geografici per studi finalizzati all'operatività», in PESARESI C. (a cura di), *Flussi, Reti e Sistemi Informativi Geografici, tra presente e futuro, Semestrale di Studi e Ricerche di Geografia*, 2, 2008, pp. 25-27.
- PAVIA D., «Sull'importanza dei sistemi di riferimento per l'ambiente GIS: un focus su ArcGIS Pro e sulle trasformazioni grid-based», in *Semestrale di Studi e Ricerche di Geografia*, 2, 2023a, pp. 113-127.
- PAVIA D., «Un patrimonio di pietre da salvaguardare: applicazioni GIS per una mappatura dei muretti a secco dell'Alta Murgia», in *Bollettino dell'Associazione Italiana di Cartografia*, 179, 2023b, pp. 108-122.

- PESARESI C., *Applicazioni GIS. Principi metodologici e linee di ricerca. Esercitazioni ed esemplificazioni guida*, Novara, UTET – De Agostini, 2017.
- PESARESI C., «InDAGIS-MODE&APP. Modelli Operativi Dinamici con Elaborazioni GIS Interattive Avanzate a n Dimensioni e Applicazioni per la Preparedness e la Patrimonializzazione», in *Semestrale di Studi e Ricerche di Geografia*, 1, 2022, pp. 69-89.
- PESARESI C., PAVIA D., *Applicazioni GIS per progettualità interdisciplinari. Pensare geograficamente in chiave applicata (con ArcGIS Pro e ArcGIS Online)*, Milano, UTET Università - D Scuola SpA, 2025.
- PICCHI E., MONTEMAGNI S., BIAGINI L., «DBT-ALT: A System for Storing and Querying the Data of the Atlante Lessicale Toscano (ALT)», in *Dialectologia*, 9, 2001, pp. 85-103.
- R CORE TEAM, *R: A language and environment for statistical computing*, Vienna, R Foundation for Statistical Computing, 2021, <<https://www.R-project.org/>>.
- REGIS R., «Sul tipo lessicale», in *Géolinguistique*, 19, 2019.
- RIVOIRA M., CERRUTI M., GHIA A., TURCHETTA B., «L'Atlante Linguistico Italiano: storia, metodi e sviluppi», in questo volume.
- RUBEHN A., MONTEMAGNI S., NERBONNE J., «Extracting Tuscan phonetic correspondences from dialect pronunciations automatically», in *Language Dynamics and Change*, 14, 2024, pp. 1-33.
- SALVATORI F., «Indirizzo di saluto», in PESARESI C. (a cura di), *Flussi, Reti e Sistemi Informativi Geografici, tra presente e futuro*, *Semestrale di Studi e Ricerche di Geografia*, 2, 2008, pp. 21-22.
- TEERAROJANARAT S., TINGSABADH K., «Using GIS for Linguistic Study: a Case of Dialect Change in the Northeastern Region of Thailand», in *Procedia Social and Behavioral Sciences*, 21, 2011, pp. 362-371.
- TERRACINI B., «Il concetto di lingua comune e il problema dell'unità di un punto linguistico minimo», in *Bollettino dell'Atlante Linguistico Italiano*, 5-6, 1960, pp. 12-24.
- THIBAUT A., AVANZI M., LO VECCHIO M., MILLOUR A. (a cura di), *Nouveaux regards sur la variation dialectale – New Ways of Analyzing Dialectal Variation*, Strasbourg, Éditions de Linguistique et de Philologie, 2021.
- WIELING M., SASSOLINI E., CUCURULLO S., MONTEMAGNI S., «ALT Explored: Integrating an Online Dialectometric Tool and an Online Dialect Atlas», in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia, European Language Resources Association (ELRA), 2016, pp. 3265-3272, <<https://aclanthology.org/L16-1520/>>.
- WILKINSON M., DUMONTIER M., AALBERSBERG I. *et alii*, «The FAIR Guiding Principles for scientific data management and stewardship», in *Scientific Data*, 3, 160018, 2016.

*Verso un nuovo volto dell'Atlante Linguistico Italiano (ALI).**Processi di valorizzazione in ambiente ArcGIS*

Il progetto DigitALI, avviato nel contesto del PNRR «CHANGES», ha l'obiettivo di digitalizzare e valorizzare l'*Atlante Linguistico Italiano* (ALI) attraverso strumenti avanzati e specifiche funzionalità in ambiente GIS. L'iniziativa ha dovuto fare i conti con i limiti di accessibilità dell'ALI cartaceo, integrando e raccordando dati linguistici e socio-demografici, immagini fotografiche e coordinate geografiche, il tutto organizzato in un geodatabase relazionale implementato in PostgreSQL/PostGIS. Un aspetto centrale è la conversione dell'ALICODE – una codifica ASCII usata per rappresentare i dati fonetici – in Unicode e successivamente in IPA, per favorire la compatibilità con standard internazionali e permettere ulteriori sviluppi. Il progetto ha inoltre previsto l'accurata geolocalizzazione degli informatori, il trattamento come metadati delle informazioni presenti nei verbali di inchiesta, la ristrutturazione dei dati con script Python e R. Grazie a questi interventi e una serie di concatenazioni tabellari e specifici step metodologici atti a garantire la qualità e sovrapposibilità dei dati migrati in ambiente GIS e la fluidità del flusso di lavoro, è in corso di realizzazione una versione digitale dell'*Atlante* mediante ArcGIS Pro. In questo lavoro vengono presentate diverse carte digitali che fungono da prodotti di output prototipali, su alcune regioni e a scala nazionale, in grado di supportare interrogazioni interattive e di restituire sottoforma di pop-up i metadati collegati. A livello prospettico, DigitALI rappresenta un esempio di collaborazione interdisciplinare che apre una serie di strade per studi linguistici, geografici e storici, contribuendo a favorire in modo significativo la fruibilità e l'analisi del patrimonio linguistico italiano.

*Towards a New Face of the Atlante Linguistico Italiano (ALI).**Processes of Enhancement in ArcGIS Environment*

The DigitALI project, launched within the framework of the PNRR «CHANGES», aims to digitize and enhance the *Atlante Linguistico Italiano* (ALI) by means of advanced digital tools and specific functionalities in a GIS environment. The initiative had to address the accessibility limitations of printed ALI, integrating and connecting linguistic and socio-demographic data, photographic images and geographic coordinates, all organized in a relational geodatabase implemented in PostgreSQL/PostGIS. A key aspect of this is the conversion of the ALICODE – an ASCII-based encoding used to represent phonetic data – into Unicode and subsequently into IPA, to facilitate compatibility with international standards and enable further developments. The project also foresaw accurate geolocalization of informants, the processing of the information present in the fieldwork reports as metadata, and the restructuring of data through Python and R scripts. Thanks to these interventions and a series of tabular concatenations and specific methodological steps aimed at guaranteeing the quality and overlap of the migrated data in a GIS environment and the workflow fluidity, a digital version of the

Atlante is being developed in ArcGIS Pro. In this paper various digital maps are presented which act as prototype output products, on some regions and at national level, able to support interactive queries and give back the connected metadata in the form of pop-ups. Prospectively, DigitALI is an example of interdisciplinary collaboration, paving the way for linguistic, geographical and historical studies, contributing to significantly fostering the usability and analysis of Italy's linguistic heritage.

Vers un nouveau visage de l'Atlas linguistique italien (ALI).

Processus de valorisation dans un environnement ArcGIS

Développé dans le cadre du PNRR « CHANGES », le projet DigitALI vise à numériser et valoriser l'Atlas Linguistique Italien (ALI) grâce à des outils avancés et à des fonctionnalités spécifiques dans un environnement SIG. Cette initiative a dû faire face aux limitations d'accessibilité de l'ALI dans sa version en papier ; elle a dû harmoniser et relier des données aussi bien linguistiques que sociodémographiques, des images photographiques et des coordonnées géographiques, le tout organisé dans une géo-base de données relationnelle implémentée dans PostgreSQL/PostGIS. Un aspect important a été celui de la conversion de l'ALICode – un codage ASCII utilisé pour représenter les données phonétiques – en Unicode puis en API, afin de promouvoir la compatibilité avec les normes internationales et de permettre tout développement futur. Le projet comprend également la géolocalisation précise des informateurs, le traitement des informations des rapports d'enquête sous forme de métadonnées et la restructuration des données à l'aide de scripts Python et R. On est en train de créer une version numérique de l'Atlas avec ArcGIS P qui exploite toutes ces interventions, une série de concaténations tabulaires et des étapes méthodologiques spécifiques visant à garantir la qualité des données, leur superposition dans un environnement SIG ainsi que la fluidité du flux de travail. Ce travail présente plusieurs cartes numériques servant de prototype ; celles-ci couvrent plusieurs régions à l'échelle nationale ; elles sont capables de prendre en charge des requêtes interactives et d'afficher les métadonnées associées sous forme de fenêtres contextuelles. Finalement, DigitALI représente un exemple de collaboration interdisciplinaire ouvrant de nombreuses perspectives aux études linguistiques, géographiques et historiques ; il améliore aussi considérablement l'accès au patrimoine linguistique italien ainsi que son analyse.

