

Il progetto DigitALI: problemi e prospettive della digitalizzazione dell'*Atlante Linguistico Italiano* (ALI)*

Massimo Cerruti, Lorenzo Ferrarotti, Stefano Fiori,
Matteo Rivoira e Barbara Turchetta**

Abstract

The paper aims to describe the main methodological issues in the digitization of the *Atlante Linguistico Italiano* (ALI), a fundamental resource for dialectological research on the dialects of Italy. Despite its scientific relevance, ALI remains difficult to access due to its printed format and heterogeneous data structure. The project seeks to overcome these limitations by developing a structured, queryable database integrated into a digital cartographic interface. This paper presents the main methodological challenges encountered in the digitization process, including data standardization, phonetic transcription encoding, and georeferencing of linguistic points. The first results highlight the potential of DigitALI for socio-historical research, enabling large-scale analyses of dialectal variation.

1 Introduzione

1.1 L'Atlante Linguistico Italiano

L'*Atlante Linguistico Italiano* (ALI) è la seconda impresa atlantistica condotta sul territorio italiano dopo l'AIS (*Atlante Italo-Svizzero* o *Sach- und Sprachatlas Italiens und der Südschweiz*, Jaberg/Jud 1928–1940). Con una rete di punti d'inchiesta più che doppia rispetto a quella del primo e un questionario composto da svariate migliaia di domande, rappresenta un progetto di raccolta e di pubblicazione di dati più ampio e ambizioso del suo predecessore. Nato nel 1924 su iniziativa di Matteo Bartoli all'Università di Torino, con la collaborazione di Ugo Pellis e il supporto della Società Filologica Friulana, l'ALI iniziò le inchieste di raccolta di dati sul campo nel 1925 e le terminò nel 1965. Fino al 1942 le inchieste furono condotte da Ugo Pellis; dopo la guerra ripresero nel 1953, sotto la direzione di Benvenuto A. Terracini, con più raccoglitori, attivi principalmente nel centro-sud. La pubblicazione dell'atlante è iniziata nel

* Riconoscimenti: Questo lavoro è stato sostenuto da finanziamenti del Ministero dell'Università e della Ricerca italiano e dell'Unione Europea, PE 0000020 CHANGES, – CUP F53C22000700006, D53C22002530006 PNRR Missione 4 Componente 2 Investimento 1.3, finanziato dall'Unione europea – NextGenerationEU.

** L'articolo è frutto di lavoro e riflessioni comuni a tutti gli autori. A fini accademici, i capitoli 1 e 4 sono da attribuire a tutti gli autori, il capitolo 2 a Stefano Fiori, il capitolo 3 a Lorenzo Ferrarotti.

1995 sotto la guida di Lorenzo Massobrio ed è in corso ancora oggi. La consistenza attuale dell'ALI può essere descritta con alcune cifre.¹

- 1009 inchieste: 727 svolte tra il 1925 e il 1942 (da Ugo Pellis), 282 tra il 1953 e il 1965 (da 6 raccoglitori: Corrado Grassi, Temistocle Franceschi, Raffaele Giacomelli, Michele Melillo, Giorgio Piccitto, Giovanni Tropea)
- 947 punti cartografati (ma sono molte di più le località effettive)
- ~1.700 informatori
- 7659 voci di questionario (8309 domande), articolate in diverse parti
- ~2-3 milioni di risposte stimate 8.826 fotografie, 2500 cartellini illustrativi
- 10 volumi pubblicati su 20 (1995–2023)

L'archivio originale cartaceo su cui sono registrati i dati delle inchieste è conservato a Torino, presso la sede dell'Istituto dell'ALI. Il nucleo fondamentale di tale archivio è composto dai fascicoli delle singole inchieste svolte in ciascun punto (Figura 1, a sinistra e al centro) che conservano la copia carbone di fogli originali ritagliati e incollati su schede (Figura 1, a destra), ciascuna delle quali riporta le risposte raccolte in un singolo punto per una certa voce del questionario; tutte le schedine relative a una singola voce (ad es. “161 bocca”, ecc.) sono raccolte anche in una “mazzetta”, che costituisce, in un certo senso, il precursore di una carta linguistica. L'archivio dell'Atlante contiene molti altri materiali linguistici (inchieste gergali, traduzioni della *Parabola del figliol prodigo*) ed iconografici, come le immagini utilizzate dai raccoglitori per porre alcune domande, e soprattutto, una vasta collezione di oltre 8000 fotografie, per la maggior parte conservate presso la Società Filologica Friulana di Udine, di grande valore storico ed etnografico.²

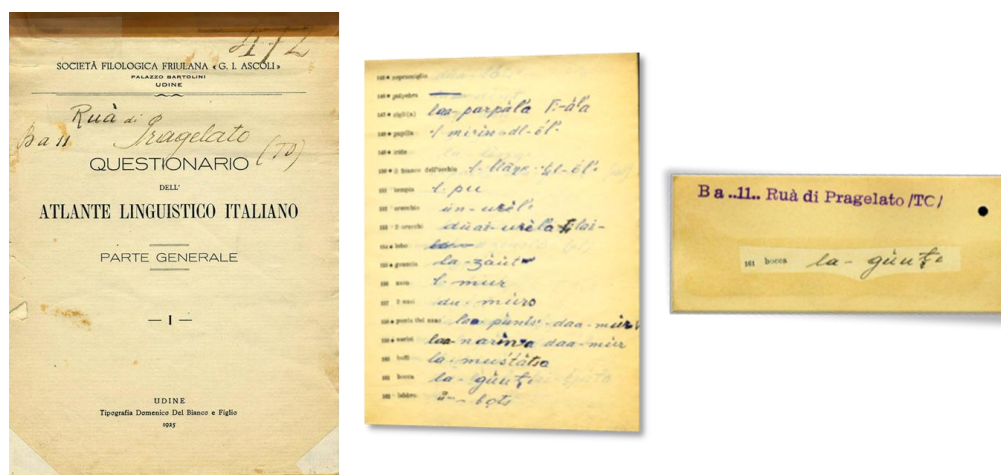


Figura 1: A sinistra, il frontespizio del fascicolo dell'inchiesta svolta a Ruà di Pragelato (TO), relativo alla Parte Generale del questionario; al centro, una pagina interna del fascicolo; a destra, la schedina contenente la risposta alla voce *bocca* (161)

¹ Per cui si vedano Cugno/Massobrio (2010: 58–73), Rivoira (2016) e Cugno/Rivoira/Ronco (2020). Molti materiali sono reperibili sul sito web dell'ALI.

² Il fondo fotografico udinese è già stato reso disponibile online sul sito della Società Filologica Friulana Fondo Fotografico Ugo Pellis e se ne prevede una futura integrazione con DigitALI (v. oltre).

Uno strumento fondamentale, anche se spesso trascurato, per la consultazione dell'Atlante sono i verbali d'inchiesta, estratti direttamente dai quaderni compilati dai raccoglitori e pubblicati in due volumi (cf. Massobrio et al. 1995). Oltre a informazioni generali circa le caratteristiche della località e le condizioni del rilievo, essi contengono un profilo a volte molto dettagliato degli informatori intervistati, non soltanto sul piano anagrafico, ma anche per quanto riguarda il carattere, gli atteggiamenti e le abitudini linguistiche. Il verbale è quasi sempre chiuso da una relazione, di lunghezza variabile, nella quale il raccoglitore annota particolari suoni riscontrati durante il rilievo ma anche ulteriori osservazioni, talvolta di lunghezza considerevole, specialmente nelle inchieste condotte dopo il 1953, che a volte arrivano ad avere quasi le caratteristiche di monografie dialettali in miniatura.

1.2 DigitALI nell'ambito di Changes

L'ALI è stato, e per certi aspetti è ancora, di difficile accesso per i ricercatori. In una prima fase, la pubblicazione tardiva, a partire dal 1995, ha fatto sì che i suoi dati fossero consultabili solo presso l'archivio di Torino; successivamente, la diffusione in formato cartaceo su più volumi non ha sempre ricevuto l'attenzione adeguata, anche a causa del grande formato. Inoltre, l'ALI si distingue dall'AIS, considerato lo standard degli atlanti linguistici in Italia, per una maggiore densità di punti ma una costruzione meno uniforme³, con inchieste talvolta molto ridotte e minore attenzione agli aspetti morfosintattici (cf. Loporcaro 2013: 30). Negli ultimi anni, il mondo accademico ha posto sempre più attenzione all'accessibilità e alla sostenibilità digitale degli archivi linguistici, promuovendone la digitalizzazione; limitandosi solo all'ambito europeo, ci si può fare un'idea della consistenza di questo movimento scorrendo i contributi presenti nella serie “Berichte aus der digitalen Geolinguistik” (cf. Krefeld/Lücke 2018), o quelli nel volume a cura di Thibault et al. (2021). Tuttavia, l'ALI, ancora in corso di realizzazione e caratterizzato da una mole considerevole di dati, non ha finora potuto essere integrato in tali processi, restando di fatto inaccessibile sia agli utenti umani sia alle macchine (cf. Krefeld/Lücke 2023) e non conforme ai principi FAIR, pur essendo dotato di un processo di redazione informatizzato.

Il progetto DigitALI ha dunque come obiettivo finale lo sviluppo di un'interfaccia cartografica digitale per l'ALI, collegata a un database strutturato e interrogabile, che possa configurarsi come un vero e proprio corpus linguistico avanzato. L'iniziativa rientra nelle attività del Partenariato PNRR CHANGES (*Cultural Heritage Active Innovation for Next-Gen Sustainable Society*) e, più in particolare, dello Spoke 3 (*Digital libraries, archives and philology*), che si propone tre obiettivi principali. Il primo è la documentazione, archiviazione digitale e analisi di patrimoni culturali scritti e orali, considerati sia in prospettiva diacronica che sincronica. Il secondo è lo sviluppo di metodologie e strumenti tecnologici per la documentazione dei patrimoni linguistici e culturali. Il terzo riguarda infine la diffusione della cultura scientifica legata

³ Il questionario originale – suddiviso in una *Parte generale* (suddivisa in due fascicoli) e in una *Parte Speciale* relativa al lessico agricolo e dei mestieri, con approfondimenti specifici per le inchieste da svolgersi in montagna, in pianura o al mare e con un'ulteriore appendice di carattere morfologico che fu sottoposta solo in alcune inchieste – fu ridotto di circa un terzo dopo il primo decennio di inchieste (nel 1935) e ulteriormente ridotto nell'ultima fase delle inchieste. In alcuni punti fu inoltre condotta un'*Inchiesta minima di confronto* limitata a 500 voci.

ai valori della diversità linguistica e culturale, con un'attenzione particolare alla divulgazione verso un pubblico ampio e diversificato. CHANGES ha un approccio interdisciplinare che coinvolge diverse aree del sapere, tra cui linguistica storica, sociolinguistica, dialettologia, etnolinguistica, etnografia, antropologia, storia, filologia, letteratura e ingegneria informatica. Gli ambiti di ricerca si articolano attorno alla linguistica del contatto, alla sociolinguistica variazionista e storica e alla filologia.

All'interno dello Spoke 3 di Changes, il progetto DigitALI si colloca nell'ambito del Work Package 5 (*Languages and their legacies in oral digital archives*), dedicato alla digitalizzazione e valorizzazione degli archivi orali e scritti, attraverso la creazione di almeno 15.000 record digitali con l'obiettivo di verificare l'interoperabilità dei contenuti e sviluppare un software in grado di consentire la lettura trasversale dei dati di diversa tipologia e la geolocalizzazione delle fonti. Per questo, sono coinvolti, oltre all'Atlante Linguistico Italiano, altri archivi distribuiti su diverse regioni italiane e all'estero, tra cui l'Opera del Vocabolario Siciliano, la Discoteca di Stato, il Museo delle Arti e Tradizioni Popolari di Roma e l'Istituto Nazionale per il Patrimonio Immateriale. La realizzazione del progetto si basa sull'adozione di tecnologie avanzate per la gestione e analisi dei dati linguistici ed etnografici, come applicazioni GIS per la geolocalizzazione e la lettura trasversale delle informazioni, software dedicati alla raccolta e all'elaborazione di dati linguistici ed etnografici e strumenti per la connessione tra tipologie diverse di formato digitale.⁴

L'impatto del progetto Changes si articolerà su due livelli. Per la comunità scientifica, consentirà la digitalizzazione e archiviazione di patrimoni linguistici e culturali attualmente conservati su supporti fragili e la messa in opera di tecnologie informatiche applicabili alla ricerca. Per la società, offrirà strumenti di divulgazione in grado di sensibilizzare il pubblico sull'importanza della diversità linguistica e culturale, promuovendo una maggiore consapevolezza del valore del patrimonio immateriale. In questo modo, Changes rappresenta un punto d'incontro tra linguistica, tecnologia e valorizzazione culturale, favorendo nuove prospettive per la ricerca e la diffusione del sapere.

In questo contributo intendiamo riferire sullo stato dei lavori e illustrare le principali tappe seguite nel corso del primo anno e mezzo di vita del progetto DigitALI, nonché gli strumenti tecnologici e metodologici adottati per far fronte alle sfide poste dalla digitalizzazione; verranno inoltre riportati i primi risultati concreti di alcune ricerche effettuate sul database e si darà una panoramica delle prospettive di sviluppo alla cui realizzazione si sta attualmente lavorando.

2 Il processo di digitalizzazione

Bisogna innanzitutto sottolineare che la realizzazione delle carte a stampa si avvale di strumenti informatici sin dalla pubblicazione del primo volume nel 1995, ed è basata essenzialmente sulla digitalizzazione delle schede (v. Figura 1, capitolo 1) da parte della redazione, la quale, tramite programmi appositamente creati dall'Istituto Poligrafico della Zecca di Stato, produce per

⁴ Un progetto simile, nell'ambito di CHANGES (Spoke 2, *Creativity and intangible cultural heritage*), è la digitalizzazione e la georeferenziazione della grande quantità di dati morfosintattici relativi ai dialetti italiani contenuti nei tre volumi di Manzini/Savoia (2005), che per molti versi rispondono a esigenze simili a quelle a cui vorrebbe rispondere il DigitALI (cf. Mazzaggio et al. 2023/2024; l'interfaccia è già consultabile).

ciascuna voce del questionario un file di testo in cui ciascuna riga riporta il contenuto di una schedina, cioè la risposta dialettale preceduta dal numero e dal codice alfanumerico del relativo punto linguistico, mentre i testi delle note e dei riferimenti fotografici, che fanno parte integrante delle carte, vengono salvati in due file separati (Figura 2). Nei decenni si è quindi accumulata una grande quantità di materiale già digitalizzato e, trattandosi soprattutto di semplici file di testo, ancora leggibile coi moderni strumenti informatici; ciò ha rappresentato un enorme vantaggio per la digitalizzazione, che ha potuto affidarsi a un processo già consolidato, concepito per la stampa ma in grado di produrre documenti digitali resistenti all’obsolescenza, risparmiando così mesi se non anni di digitazione manuale dei dati o la necessità di addestrare modelli OCR specifici, come è stato il caso per AISReloaded (cf. Loporcaro et al. 2021). Almeno per quanto riguarda i dati linguistici, dunque, il progetto DigitALI ha potuto affrontare direttamente la fase di retro-standardizzazione (cf. Geisler/Forkel/List 2021), ovvero l’adeguamento dei file già presenti in archivio ai moderni formati comunemente impiegati per la conservazione e lo scambio di dati digitali; per fare un esempio, i file delle voci sono stati riconvertiti da semplici elenchi di righe a un più consueto formato tabellare con delimitatori (come il CSV). I prossimi paragrafi approfondiranno le metodologie adottate per la riconversione dei dati linguistici e per il trattamento degli altri componenti principali del futuro Atlante digitale.



Figura 2: 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)

2.1 Standardizzazione e ricodifica dei dati linguistici

I file prodotti dai programmi di digitazione (cf. capitolo 1.2) non sono immediatamente pronti per la cartografazione, poiché il loro contenuto è scritto con una codifica speciale, che chiameremo per brevità ALIcode e che per rappresentare le trascrizioni fonetiche, spesso molto complicate, sfrutta unicamente combinazioni di caratteri ASCII (*American Standard Code for Information Interchange*), accessibili dalle tastiere di qualsiasi computer. Si tratta di un sistema concettualmente analogo al “Beta Code” sviluppato negli anni ’70 per il *Thesaurus Linguae Graecae* e adottato da molti progetti di geolinguistica digitale, per esempio *Verba Alpina* (cf. Colcuc/Zacherl 2020) e l’*Atlante Lessicale Toscano* (cf. Picchi/Montemagni/Biagini 2001), che consente di rappresentare senza perdita di informazioni, o quasi, sistemi di trascrizioni fonetica

complessi come quello dell'ALI, composto da un elevato numero di caratteri di base e diacritici che si possono combinare in maniera non completamente prevedibile, e di conservare i dati in un formato di basso livello⁵, molto resistente all'avvicinarsi delle tecnologie informatiche. Le stringhe ALIcode vengono poi ricomposte da un apposito programma nei caratteri dell'alfabeto fonetico dell'ALI, che sono tuttavia rappresentati da un font disegnato appositamente per l'Atlante e non conforme allo standard Unicode: per consentire la portabilità dei caratteri tra sistemi operativi è stato perciò necessario definire un processo di ricodifica da ALIcode a Unicode.

Fase	Stringa	
ALIcode	g%4&av %4_%3i%2~%1\	
NormALI	g%4&av##a%3i%4_%2~%1#a	
Traslitterazione	HTML	Unicode
	ǧav^{ĩ̱̀}	g\u030Cavi\u0331\u0303\u0300\u0350
Resa grafica	ğav ^î	ğav ^î

Tabella 1: Le fasi della normalizzazione delle stringhe ALIcode e le traslitterazioni in HTML e Unicode.

Blu = diacritico, rosso = lettera alzata, verde = delimitatore

Sarebbe troppo lungo descrivere nel dettaglio il funzionamento di ALIcode, ma la sua logica fondamentale è ben riassunta ad esempio dalla risposta al punto 22 (Cogne, AO) nella carta “sassolino” (IX 844)^{6, 7} per cui si veda la Tabella 1. Le codifiche sono state progettate per imitare la resa tipografica: oltre al rigo di base, i caratteri possono occupare altri sette livelli, quattro superiori (1–4) e tre inferiori (6–8), numerati a partire da quello superiore più alto^{8,9}; questa struttura impilata è ben visibile nella schermata di digitazione in Figura 3. Se i caratteri si trovano sul rigo di base occorrono da soli, per esempio <g> all'inizio della stringa, altrimenti sono preceduti da <%> seguito dal numero del livello corrispondente: visivamente, perciò, al carattere <ğ> corrisponde la sequenza <g%4&>, poiché il diacritico occorre al livello immediatamente superiore alla lettera di base.

Tuttavia, l'alfabeto fonetico dell'ALI prevede che non solo i diacritici, ma anche le lettere possano trovarsi in posizione rialzata o abbassata, nel qual caso sono anch'esse precedute da <%[1-8]>¹⁰; inoltre, poiché l'ordine delle codifiche procede linearmente dal rigo di base verso l'alto o il basso, questo fa sì che i diacritici sottoscritti precedano la lettera rialzata alla quale si

⁵ Questa espressione non deve essere interpretata in senso deteriore, ma come la proprietà del sistema di mantenersi a un livello, appunto, relativamente basso di astrazione rispetto al funzionamento della macchina, e dunque indipendente da tecnologie proprietarie di natura necessariamente più transitoria.

⁶ Il numero romano indica il volume e quello in cifre arabe la carta linguistica.

⁷ Il numero romano indica il volume e quello in cifre arabe la carta linguistica.

⁸ La cifra 5 corrisponderebbe al rigo di base, ma non è mai esplicitata in ALIcode.

⁹ La cifra 5 corrisponderebbe al rigo di base, ma non è mai esplicitata in ALIcode.

¹⁰ Notare lo spazio iniziale, che si riferisce al fatto tipografico che in quella posizione il rigo fondamentale è libero. La sequenza [1–8] è un'espressione regolare che indica una sola occorrenza di un numero compreso tra 1 e 8.

appoggiano, cosicché <̂> è rappresentato da <%4_%3i%2~%1\>¹¹. Questa caratteristica confligge con le convenzioni Unicode, secondo cui tutti i diacritici devono seguire i caratteri a cui si appoggiano, e la loro posizione è definita da un attributo che li assegna a una delle tre *Canonical Combining Classes* (0 = caratteri sul rigo, 230 = caratteri sovrascritti, 220 = caratteri sottoscritti). Prima della ricodifica, dunque, si è reso necessario un passaggio intermedio, implementato tramite uno script in Python che applica in sequenza una ventina di espressioni regolari alle stringhe originali, convertendole in un formato detto “NormALI”, che rispetta le convenzioni Unicode; nella riga corrispondente in Tabella 1, si nota come il gruppo che comprende la lettera rialzata e i relativi diacritici sia stato delimitato dalle sequenze <##a> e <#a>, rispettivamente in apertura e in chiusura, mantenendo così l’informazione originale.

	Grapheme	Class	Left	Right	HTML	UNICODE
491	O	BASE			ω	\u03C9
492	%6O	SUB			ω	\u03C9
493	%7O	SUB			ω	\u03C9
494	%8O	SUB			ω	\u03C9
495	%1O	SUP			ω	\u03C9
496	%2O	SUP			ω	\u03C9
497	%3O	SUP			ω	\u03C9
498	%4O	SUP			ω	\u03C9
499	%1P	BASE			᛫	\u16EB
500	%2P	BASE			᛫	\u16EB
501	%3P	BASE			᛫	\u16EB
502	%4P	BASE			᛫	\u16EB
503	%6@	DIA			̓	\u0313
504	%7@	DIA			̓	\u0313
505	%8@	DIA			̓	\u0313
506	%1@	DIA			̓	\u0313
507	%2@	DIA			̓	\u0313
508	%3@	DIA			̓	\u0313
509	%4@	DIA			̓	\u0313

Figura 3: Stralcio del profilo ortografico per la conversione dalle codifiche ALIcode (colonna “Grapheme”) in HTML e Unicode

A questo punto si può procedere alla traslitterazione, realizzata sempre in Python con la libreria *segments* (Forkel/Moran 2018), appositamente creata per la manipolazione di dati linguistici, seguendo le indicazioni fornite da Moran/Cysouw (2018) e l’esperienza di digitalizzazione dei *Tableaux Phonétiques des Patois Suisses Romands* (TPPSR) (cf. Gauchat/Jeanjaquet/ Tappolet 1925), descritta in Geisler/Forkel/List (2021), in cui si adottano le stesse tecnologie. Per prima cosa, è stato preparato manualmente un profilo ortografico, cioè una tabella di equivalenze tra grafemi appartenenti a diversi sistemi ortografici, di cui viene riportato uno stralcio in Figura 4, che include le codifiche per <ω> (U+03C9 = GREEK SMALL LETTER OMEGA), <•> (U+16EB = Runic single punctuation) e <’> (U+0313 = COMBINING COMMA ABOVE): la colonna “Grapheme”, obbligatoria, contiene tutte le singole codifiche ALIcode da

¹¹ Viceversa, quando una lettera è abbassata, è preceduta dai diacritici sovrascritti.

traslitterare con le sequenze corrispondenti nelle colonne “HTML” e “UNICODE”¹²; l’aggiunta delle codifiche HTML si deve al fatto che queste consentono una resa grafica dei simboli ALI sul web molto più fedele all’originale, soprattutto per quanto riguarda i caratteri alzati e abbassati. Si noter , infatti, che a tutte le varianti ALIcode che specificano la posizione di un simbolo rispetto al rigo corrisponde sempre la medesima codifica nell’ortografia: per comodit , infatti, si   deciso di introdurre questa informazione gi  in fase di normalizzazione, con un gruppo di espressioni regolari che individuano i gruppi alzati o abbassati di lettere pi  i loro diacritici e li delimitano con stringhe apposite, che vengono poi convertite rispettivamente nei tag HTML `` e ``. D’altro canto, Unicode non ha la possibilit  di rappresentare questo tipo di formattazione del testo, e il repertorio di lettere sotto- e sovrascritte che offre   troppo esiguo per soddisfare le esigenze dell’ALI; si   deciso per il momento, per mantenere l’informazione relativa alla posizione dei caratteri, di aggiungere il diacritico `<’>` (U+0350 = COMBINING RIGHT ARROWHEAD ABOVE) alle lettere alzate e `<,>` (U+0355 = COMBINING RIGHT ARROWHEAD BELOW) a quelle abbassate nelle stringhe Unicode.

Tutto il processo si pu  dunque riassumere come in Tabella 1, che tramite i colori traccia le trasformazioni subite dalle codifiche; lo script applica automaticamente la conversione a tutti i file delle voci di un volume dell’Atlante, producendo in pochi secondi una versione standardizzata e ricodificata, nella quale tuttavia le nuove codifiche non sostituiscono quelle originali, ma vengono salvate in colonne aggiuntive. Sono stati traslitterati allo stesso modo anche i file delle note e dei riferimenti fotografici.

2.2 Trattamento dei metadati

Parallelamente al lavoro appena descritto,   stata portata avanti sin dall’inizio la digitalizzazione dei verbali d’inchiesta, gi  pubblicati in formato cartaceo (Massobrio et al. 1995), che contengono preziose informazioni sui punti indagati e sulle caratteristiche biografiche e socio-demografiche degli informatori, che costituiscono i “metadati” da collegare ai dati linguistici veri e propri, la cui valorizzazione   un tratto caratterizzante del progetto DigitALI. Il trattamento dei metadati contenuti negli oltre 1000 verbali pubblicati   stato ripartito tra i membri del progetto, i quali hanno digitato un sottoinsieme di informazioni salienti¹³ in una tabella Excel che, al termine del lavoro, conta 1846 record. Questo numero   maggiore di quello degli informatori effettivi dell’ALI, che un calcolo provvisorio pone invece intorno a 1719, a causa del frequente utilizzo di informatori di controllo, una pratica adottata talvolta anche da Pellis, ma divenuta assai pi  sistematica sotto la direzione di Terracini (cf. capitolo 1.1): spesso, nei punti dove   stata interpellata pi  di una persona (poco meno della met  del totale), le sezioni del questionario gi  svolte con la fonte principale sono state ripetute, in tutto o in parte, con un altro informatore, le cui risposte sono contrassegnate nelle carte linguistiche (e dunque nei file

¹² Le colonne “Class”, “Left” e “Right” contengono, rispettivamente la tipologia del carattere e l’eventuale contesto a sinistra e a destra, sono tutte opzionali e al momento non rientrano nel processo di traslitterazione, ma non   escluso che ci  accada in futuro.

¹³ Per quanto riguarda le caratteristiche degli informatori, sono state inserite nella tabella le informazioni relative al nome, et , genere, professione e grado di cultura; sono state tralasciate, per il momento, informazioni riferite alla localit  come la diocesi, i collegamenti, le scuole presenti e cos  via, ma tutto questo pu  facilmente essere integrato nella struttura gi  esistente.

ALlcode, cf. capitolo 2.1) da sigle come <*ai> o <ai> (=altro informatore)¹⁴ e simili. Può anche accadere, tuttavia, che gli informatori secondari siano informatori principali di altre voci del questionario che non sono state chieste ad altri, nel qual caso le risposte non sono contrassegnate da nessuna sigla, e l'identità della fonte può essere appurata solo consultando i verbali d'inchiesta.

	A	B	C	D	E	F	G	H	I	J
1	loc	punto	codice	parti	str	num_inf	nome	eta	sess	dob
2	Malfa	1000	ii010	1:1496,1497:3544,3545:4785,5421:€		1	Salvatore Giuffrè	61	M	1903
3	Malfa	1000	ii010	5071:5420ter		2	Giulio Taranto	38	M	1926
4	Malfa	1000	ii010		*ai	3	Franco Carrà	17	M	1947
5	Trapani	1001	ig010	1:504,5071:5420ter	I	1	Antonino Sotera	37	M	1903
6	Trapani	1001	ig010	505:1496,5834:5965	I	2	Pietro Manno	38	M	1902
7	Trapani	1001	ig010	91bis,91ter,93bis,94bis,97bis,101bis I		3	Alberto Accardi	46	M	1894
8	Trapani	1001	ig010		*ai	3	Alberto Accardi	46	M	1894
9	Ballata	1001	ig020		II	1	Giuseppe Maranzano	35	M	1905

Figura 4: Esempio della tabella nella quale sono stati digitati i verbali d'inchiesta. Notare lo “sdoppiamento” dell'informatore 3 del punto 1001

Di tutto ciò si è dovuto tenere conto durante la compilazione dei metadati, poiché lo scioglimento di questi riferimenti è essenziale per poter collegare con precisione ciascuna risposta dialettale alla persona che l'ha fornita; così, gli informatori che sono contemporaneamente informatori principali e secondari di parti diverse del questionario sono stati “sdoppiati” su più righe della tabella, contenenti gli stessi metadati salvo quelli nella colonna “parti” e “str” (= “stringa identificativa”). Un esempio è offerto in Figura 4, dove l'informatore di Trapani Alberto Accardi occupa due righe, una per le voci del questionario chieste a lui solo e un'altra, contrassegnata da *ai, riservata alle voci che si sovrappongono con altri informatori. Le sigle <I> e <II>, invece, si riferiscono al fatto che il punto di Trapani è in realtà composto da due rilievi complementari, il primo svoltosi proprio a Trapani e il secondo a Ballata; questa circostanza verrà approfondita nel prossimo paragrafo.

2.3 Georeferenziazione

All'inizio del lavoro, esisteva in archivio un file contenente l'elenco di tutti i punti d'inchiesta dell'ALI, corredati dall'anno di svolgimento delle inchieste e dal nome del raccoglitore, ma le coordinate delle località erano disponibili solo in formato geometrico, finalizzato alla realizzazione delle carte a stampa; si è perciò provveduto a una nuova georeferenziazione manuale dei punti d'inchiesta, creando un file .kml (*Keyhole Markup Language*) con GoogleEarth, che può essere facilmente esportato con l'aiuto di un'applicazione GIS in altri formati di condivisione di dati spaziali, come GeoJSON o shapefile.

Quella che poteva apparire una procedura quasi meccanica ha però fornito l'occasione per un cambio di prospettiva, che influenza profondamente la concezione stessa dell'atlante: come si accennava nel caso di Trapani alla fine del paragrafo precedente, alcuni punti d'inchiesta (circa

¹⁴ La corrispondenza però non è sempre biunivoca, poiché si danno anche casi, come quello di Parma (punto 406, Pellis), dove dietro alla sigla *ai si nascondono in realtà tre diverse persone, due uomini di 35 e 47 anni e un terzo la cui età è ignota, o Salerno (punto 855, Giacomelli), in cui ai sta per la moglie e la figlia di uno degli informatori. Si tratta però di casi eccezionali.

una trentina) comprendono in realtà più rilievi svolti in località diverse: per esempio, il verbale dell'inchiesta di Locana (TO, punto 29, sigla bb14), paese in Valle Orco posto sul fondovalle, riferisce che l'informatore per l'inchiesta minima di controllo, siglata bb14*, è della frazione montana di San Lorenzo Piantonetto; è un tipico caso in cui la differente provenienza degli informatori è rilevante per l'osservazione di dinamiche di livellamento linguistico, dal momento che la frazione montana conserva il tipo dialettale francoprovenzale, mentre il capoluogo si può considerare in larga misura piemontesizzato. Un altro esempio è San Severino Marche (PU, punto 483, sigla dg330), al quale sono ricondotte le inchieste svolte nelle frazioni Cèsolo (dg32) e Stigliano (dg31). In tutti questi casi, alle inchieste secondarie sono state assegnate coordinate distinte da quelle del punto principale (Figura 5), cercando di legare la geolocalizzazione il più strettamente possibile, piuttosto che al singolo punto cartografato, alla posizione informatori stessi al momento dell'inchiesta, per quanto è possibile ricavarla dai verbali.

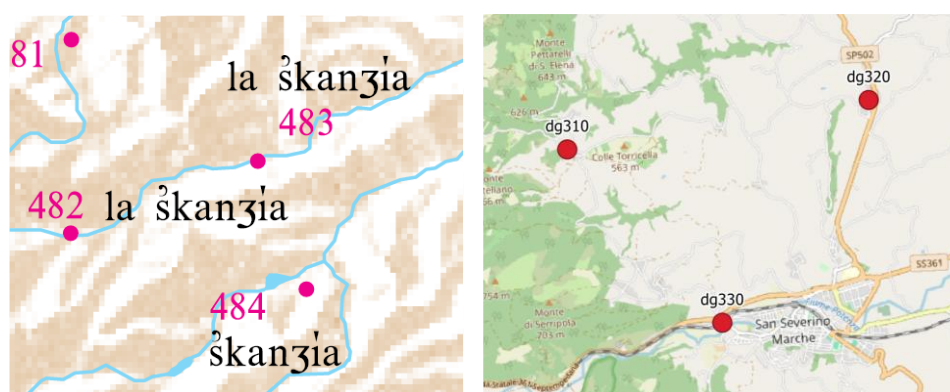


Figura 5: A sinistra: ingrandimento sul punto 483 (San Severino Marche) della carta 'scansia' (X 954); a destra: georeferenziazione delle tre frazioni in cui è stata svolta l'inchiesta

Questa procedura ha consentito di scoprire anche altri casi, meno perspicui dal momento che, contrariamente agli esempi riportati finora, le nuove località individuate non erano dotate di un codice identificativo; è il caso di Corigliano Calabro (CS, punto 932, sigla hm2), il cui verbale d'inchiesta segnala che un rilievo di controllo, privo di una sigla distinta da quella del punto principale, è stata svolta nella vicina località di Piana di Caruso, che è stata quindi identificata, georeferenziata e contrassegnata da un nuovo codice (hm2*), come se si trattasse di una nuova inchiesta corrispondente al punto 932.

La possibilità di ottenere un tale grado di precisione è uno dei vantaggi offerti da una risorsa cartografica digitale, che non deve preoccuparsi troppo della vicinanza dei punti poiché può permettersi di modificare a piacere l'ingrandimento della mappa, ed è anche in accordo con l'impostazione di tutto il progetto DigitALI, che cerca di mettere i parlanti al centro della scena, anche dal punto di vista geografico.

2.4 Creazione di un database strutturato

La congiunzione degli elementi descritti nei paragrafi precedenti (dati linguistici, verbali d'inchiesta e punti georeferenziati), sarebbe già sufficiente per una riproduzione digitale delle carte dell'ALI, alla stregua di NavigAIS (cf. capitolo 2); tuttavia, le stringhe di testo nei file delle

voci contengono non soltanto le forme dialettali, ma molte altre informazioni aggiuntive, come nei due esempi seguenti tratti ancora dalla carta “sassolino” (cf. capitolo 2.1):

1. prädüčča , *ai na ūrićča (Sora, punto 668)
2. u^m brēcč'o u , ~ brēcč'e d (Ancona, punto 471)

Il testo in corsivo che accompagna le diverse risposte legate al punto reca informazioni preziose: a Sora la sigla <*ai>, già incontrata nella tabella dei metadati (cf. capitolo 2.2), identifica la risposta fornita dall'informatore secondario, mentre ad Ancona entrambe le risposte sono della medesima informatrice, la quale però specifica che la prima forma è usata dagli uomini (<u>) e la seconda dalle donne (<d>). Esistono molte altre informazioni aggiuntive che si incontrano nel testo delle risposte dell'Atlante, le più importanti delle quali sono il rimando alle note <⊥>, il riferimento fotografico <F> e i vari simboli e lettere che rinviano alla legenda o all'apparato iconografico, che consente di disambiguare il referente di una risposta dialettale, particolarmente importante per carte di interesse etnografico come *casa* (IX 1506) e *culla* (VII 679). Oltre a simboli e sigle, si possono trovare osservazioni più o meno estese aggiunte dalla redazione o dal raccoglitore stesso.

punto	codice	str	norm	tag	legenda	simbolo	nota	rifo
666	fg070		sassòtto					
667	fg040		sassitto					
668	fh040		prädüčča					
668	fh040	*ai	na ūrićča					
669	fh010		prätélla					
670	fh050		pärtélla					

Figura 6: Estratto del file ristrutturato della voce *sassolino* (IX 844)

Per facilitare l'accesso a queste informazioni, esse sono state estratte dalle stringhe normalizzate e assegnate a colonne distinte a seconda della loro tipologia; inoltre, le risposte multiple relative ad un unico punto sono state ricollocate su righe diverse, in modo che ciascuna di esse corrisponda a un solo *record* nella tabella. Anche in questo caso, le colonne già esistenti (NormALI, HTML, Unicode) non vengono soppresse, cosicché è possibile risalire al contesto originale di ciascuna risposta. Questa ristrutturazione dei dati è stata operata in ambiente R (versione 4.1.1, R Core Team 2021) utilizzando i pacchetti *tidyr* (cf. Wickham et al. 2023) e *dplyr* (cf. Wickham/Vaughan/Grich 2023), ed è stata enormemente facilitata dal fatto che tutte le informazioni metalinguistiche sono riportate in corsivo nelle carte, che in ALIcode corrisponde a porzioni di testo aperte e chiuse, rispettivamente, dalle sequenze <#cc> e <#e>, e possono così essere individuate con le espressioni regolari¹⁵. Il risultato dell'operazione è mostrato in Figura 6, dove le risposte di Sora, convertite in formato Unicode nella colonna “norm”, compaiono su righe separate e compaiono colonne dedicate per le marche d'uso (“tag”), la legenda, i simboli, le note, i riferimenti fotografici (“rifo”) e finalmente la sigla identificativa nella colonna “str”. Si noti che proprio la presenza di quest'ultimo attributo, in combinazione con il

¹⁵ Al netto degli inevitabili errori umani nella digitazione e di alcune piccole eccezioni, che hanno dovuto essere prima riconosciute e corrette prima della normalizzazione del testo, si tratta di un altro esempio dei vantaggi portati dal rigore formale con cui i dati in archivio vengono trattati in fase di redazione.

numero di punto, il codice e il numero della voce, consente di individuare l'informatore corrispondente a ciascuna risposta nella tabella dei metadati.

Ai file delle note e dei riferimenti fotografici, invece, sono stati aggiunti solo i campi “nota” e “rifo”, rispettivamente, i quali, in combinazione con i campi comuni “punto” e “codice” e un nuovo campo “q”, aggiunto anche ai file delle risposte, contenente il numero della voce corrispondente, hanno consentito di combinare tutti e tre i file in uno solo per ciascuna voce; infine, conclude la procedura di ristrutturazione l'aggiunta del campo “vol”, che determina l'appartenenza delle voci al volume nel quale sono state pubblicate, dopodiché tutte le voci vengono riunite in un solo file che contiene tutto il materiale digitalizzato. Nella prospettiva di rendere questa struttura interoperabile e, soprattutto, di interfacciarsi con altre imprese consimili, come esplicitamente previsto dal progetto PRIN 2022 *DigitALLa. Digitalisation of Italian Linguistic Atlases*¹⁶ il campo “voce” sarà a sua volta legato a un campo “significato”, che permetterà di collegare i concetti dell'atlante con un vocabolario condiviso con altre banche dati. Sulle strategie da adottare la riflessione è ancora aperta, tra i modelli adottabili vi è quello di VerbaAlpina 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 al. 2025).

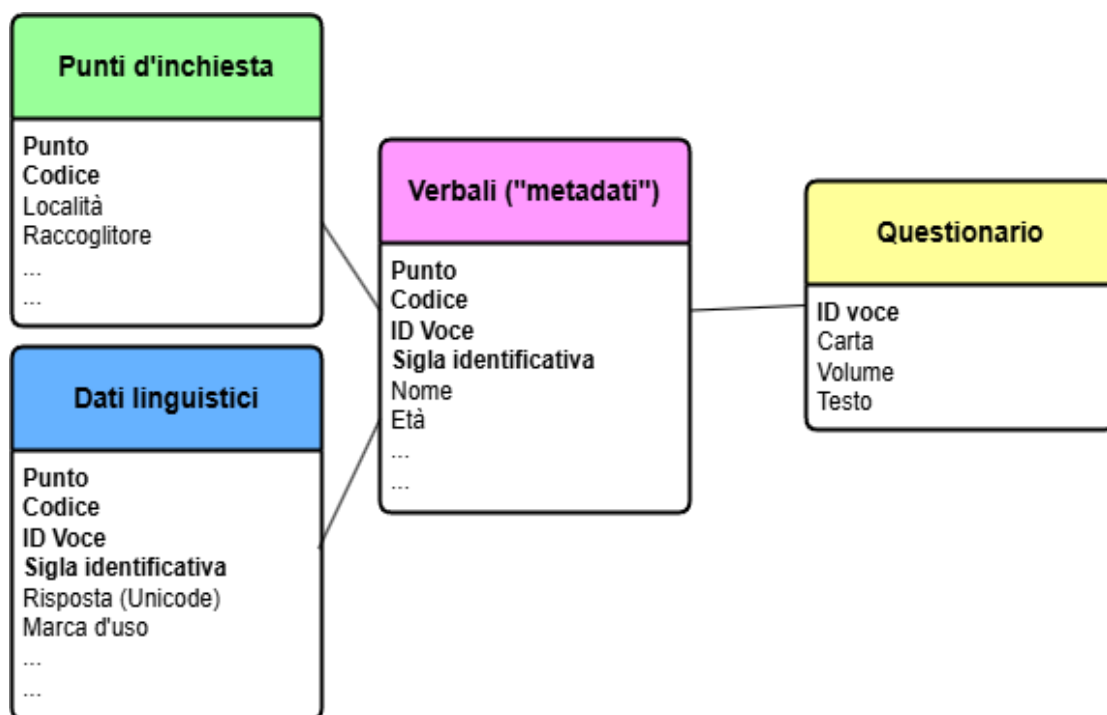


Figura 7: Rappresentazione schematica dei componenti del database DigitALI

¹⁶ Il progetto biennale è stato avviato nel febbraio del 2025 coinvolge tre unità (Torino, L'Aquila e Udine) ed è diretto da Matteo Rivoira.

2.5 Struttura e interrogazione del database

Tutti gli elementi descritti nei paragrafi precedenti sono attualmente immagazzinati in un database MongoDB, un DBMS (*DataBase Management System*) di tipo NOSQL, ovvero “not (only) SQL”. Ciò significa che non c’è una struttura relazionale nella quale le tabelle fanno riferimento l’una all’altra tramite chiavi primarie e secondarie, ma i dati sono conservati come “collezioni”, concettualmente analoghe alle tabelle, che raccolgono “documenti”, assimilabili ai singoli record, ciascuno dotato di “proprietà” che assumono certi valori. La struttura sottostante non è tabellare ma in formato BSON, una variante del popolare JSON, che è un formato standard per lo scambio di dati; tuttavia esistono interfacce MongoDB, come l’applicazione Compass, che consentono una più consueta visualizzazione dei dati in formato tabellare e soprattutto l’importazione e l’esportazione di file CSV. Uno dei vantaggi di MongoDB è che non obbliga a definire uno schema per i dati, che possono dunque avere caratteristiche molto diverse; ad esempio i documenti di una stessa collezione possono avere un numero diverso di campi, o avere solo qualche campo in comune, cosa impensabile in un database relazionale; questa caratteristica lo rende particolarmente adatto ad accogliere dati di natura anche molto eterogenea, e anche per questo è stato scelto come unica piattaforma per i dati di tutti i progetti coordinati dei Work Packages 4 e 5. Inoltre, MongoDB consente gratuitamente di ospitare un database localmente sulla propria macchina, oppure in un cloud, per database di dimensioni contenute.

In Figura 7 sono mostrate le collezioni che al momento formano il database DigitALI, visualizzate come tabelle in relazione per comodità espositiva; oltre agli elementi già discussi, si vede la collezione “ALI_questionario”, creata da un file già presente tra i materiali digitali della redazione. Come si può vedere, di fatto lo schema concettuale dei dati ALI non si discosta da quello relazionale classico: al centro delle relazioni sta la collezione “Informatori”, ovvero la tabella dei metadati, che fa da perno tra i dati linguistici, la georeferenziazione e il questionario. Attualmente, sono stati caricati nel database solo i volumi IX e X, poiché i relativi file digitali sono quelli più recentemente prodotti dalla sola redazione dell’Atlante e possono essere considerati fedeli rappresentanti delle carte a stampa, mentre i materiali precedenti al volume VIII risalgono ancora alla collaborazione con l’Istituto Poligrafico, ora interrotta, e hanno bisogno di esami approfonditi per verificare che si tratti di bozze definitive prive di errori o difformità dai volumi pubblicati. Anche così, il numero di documenti, ovvero di singole risposte dialettali nel database ammonta a 221.756, una cifra notevole, soprattutto considerando che si tratta di soli due volumi, che induce a tentare qualche primo approccio di ricerca.

A questo proposito, il database può essere interrogato tramite la già citata applicazione Compass, usando una sintassi molto intuitiva simile a JavaScript (si possono in realtà scegliere stili diversi basati su vari linguaggi di programmazione). Alcuni esempi di stringhe di ricerca (*queries*) sono riportati sotto:

- A. { vol: { \$in: [9, 10] } , tag: { \$ne: "" } }
- B. { punto: 668, str: "*ai" }

La query A si può verbalizzare come “trova tutti i record nei volumi IX e X il cui campo tag contiene un’informazione (= non è vuoto)”. La query B, invece, sta per “trova tutti i record relativi all’informatore secondario (sigla *ai) del punto 668”. I risultati di queste interrogazioni

possono essere facilmente esportati in formato CSV o JSON e analizzati secondo i propri interessi. Nel prossimo capitolo mostreremo alcuni percorsi di ricerca che si sono resi possibili, o perlomeno sono stati estremamente facilitati, dalla digitalizzazione dei dati ALI.

3 Primi esempi di ricerche con il database DigitALI

3.1 La rappresentazione della variazione negli atlanti

La raccolta dei dati dell'ALI fu concepita per fornire un quadro più o meno sincronico della variazione dialettale nello spazio linguistico italiano della prima metà del XX sec., con l'attenzione rivolta a fatti fonetici e soprattutto lessicali. Si può senz'altro affermare che, cent'anni dopo, i suoi dati sono importanti anche (e specialmente) in qualità di documenti storici, che assumono una rilevanza ancora maggiore se si considera che l'ALI rappresenta la situazione di un'epoca in cui le lingue locali erano ancora in maniera predominante acquisite e trasmesse con la socializzazione primaria e si trovavano in un rapporto di sostanziale diglossia con l'italiano. Oltretutto, bisogna considerare che al tempo erano attive molte dinamiche di contatto dialettale (come ad es. l'impiego di *koinè* regionali) oggi non più attive o comunque mutate in favore di una tendenza comune all'italianizzazione.

A tal proposito, di dati di tipo variazionistico degli atlanti linguistici della prima generazione non solo sono poco studiati, ma anche generalmente poco valorizzati nella presentazione dei loro dati: pertanto, tale limite deve essere tenuto presente, e possibilmente superato, in un progetto di digitalizzazione di un atlante tradizionale.

Nell'ALI la variazione sociolinguistica si manifesta sostanzialmente in due casi. Il primo di questi è la variazione *interspeaker*, che si può cogliere quando è presente più di un informatore in un certo punto d'inchiesta e vi è una differenza tra le risposte date alla stessa domanda del questionario (a volte anche la descrizione degli informatori nei verbali d'inchiesta fornisce dati utili in questo senso v. oltre). È particolarmente rilevante che le ultime inchieste dell'atlante, svolte da raccoglitori diversi sotto la direzione di Benvenuto Terracini¹⁷ (dal 1953 in poi, v. capitolo 1.1.) prevedono sistematicamente la presenza di un informatore di controllo per tutta l'inchiesta, pratica già seguita in parte da Ugo Pellis, ma in maniera non sistematica. Come affermava Grassi (1957: 18):

[...] il materiale del nostro Atlante è in grado di darci, attraverso le oscillazioni segnalate e documentate dalle due diverse fonti, un quadro assai vicino al vero delle varie correnti culturali e linguistiche che si intrecciano e si contrastano o assommano in un'area determinata.

(Grassi 1957: 18)

¹⁷ Questa innovazione è descritta e giustificata in Grassi (1957); alcuni limiti del criterio dell'informatore singolo per inchiesta sono evidenti nell'analisi di Genre (1970) a proposito dell'inchiesta di Ghigo di Prati. Del resto, la sensibilità "proto-sociolinguistica" di Benvenuto Terracini è del resto ben nota e l'uso di diversi informatori per punto linguistico risale già ai suoi primi studi come *Il parlare di Usseglio* (ora in Terracini 2024) su influenza dello studio di Gauchat (1905).

In secondo luogo, si può cogliere anche una buona quantità di variazione percepita dagli informatori stessi, che spesso contrastano forme “normali” con altre “marcate”, contrassegnate nell’atlante da apposite marche d’uso, che rappresentano diverse dimensioni della variazione (v. oltre).

Come si è detto, gli atlanti in formato cartaceo spesso non valorizzano dati di questo tipo, perché li trattano come alternative alle risposte principali e gli strumenti per interpretarli non sono del tutto accessibili: pertanto, è necessario un processo lungo e laborioso per integrarli efficacemente nella lettura delle carte. In questo senso, la digitalizzazione offre la possibilità di sfruttarli in maniera efficiente per la ricerca scientifica, semplificandone l’accesso e la visualizzazione. Nei paragrafi che seguono mostreremo come l’estrazione automatica di questo tipo di dati da un campione (voll. IX–X) dell’archivio dell’ALI digitalizzato possa velocizzare molto questo processo.

3.2 Informatori multipli (Sora, punto 668)

Un caso piuttosto interessante di variazione *interspeaker* tra due informatori si ha a Sora (punto 668), un centro in provincia di Frosinone, nel Lazio meridionale, in cui è parlata una varietà di tipo italiano meridionale. Il raccoglitore, Raffaele Giacomelli, che condusse due interviste nel 1954, ha descritto alcuni pattern di cambiamento da lui osservati:

Fra le mutazioni più notevoli subite in questi ultimi anni dal dialetto di Sora c’è quella relativa alla palatale intervocalica e in *s*, la mutazione dell’articolo maschile *l’l’ə* in *i* (al singolare), ed a volte, il ripristino della *d* e della *v* per influenza letteraria. Altra mutazione è la scomparsa della *u* propaginata, dopo l’occlusiva velare sorda che precedeva immediatamente, ad esempio in *kūəċinə* ‘cugino’, *kūəlōnnə* ‘colonna’, *məl’l’ikwərə* ‘ombelico’, *ċəkūərə* ‘cicciolo’ ecc. Infine, ora si sente una maggiore lenizione delle occlusive sorde intervocaliche e dopo nasale. Per quanto riguarda l’articolo, ho interrogato uno dei figli dell’informatore Evangelista, Giuseppe di anni 30, sulla pronuncia dell’articolo determinativo maschile singolare e plurale che nell’antico dialetto era *l’l’ə* e nel moderno è *i*: il più delle volte pronunciava la forma antica ma con un suono di *l’* molto attenuato, rassomigliante ad una *iə*, come scrivono attualmente molti poeti di Sora. Ad esempio nel poemetto ‘La pigna e Macciocca’ (‘Il pino di villa Macciocchi’) di Francesco Biancale (Sora, 1936), l’articolo determinativo maschile singolare e plurale compare scritto ora sotto la forma moderna *i*, ora sotto quella *ie*.

(Massobrio et al. 1995: 522)

La descrizione di Giacomelli mostra che a Sora erano in atto mutamenti significativi, interpretabili come la convergenza verso una fonetica di tipo romanesco (ad es. con la deaffricazione di [tʃ] > [ʃ], la reintroduzione di [d] e [v], e una maggior sonorizzazione parziale delle occlusive intervocaliche) e la cancellazione di tratti locali, come la sostituzione dell’articolo determinativo maschile locale [ʎ:ə] con [i] e l’eliminazione di un tratto molto marcato come [w] epentetico dopo occlusiva velare (ad es. in [kwə’tʃinə] ‘cugino’, [kwə’lɔn:ə] ‘colonna’, [mə’ʎ:ikwərə] ‘ombelico’, [’tʃəkʷərə] ‘cicciolo’).

Nel rendere conto dell'inchiesta, Giacomelli caratterizza precisamente i due informatori: il primo, nato nel 1901 circa, è innovativo (come si può intuire dal suo profilo sociolinguistico, poiché svolgeva varie professioni, anche slegate dal contesto rurale) e il suo idioletto si discosta parecchio da quello tradizionale descritto da Merlo (1920), mentre il secondo, nato due decenni prima, nel 1882, è più conservativo e appartiene a una generazione precedente.

[Informatore 1:] Giovanni Evangelista; età: 53; professione: contadino, carraio, norcino, infermiere; dimora più lunga: Sora; dimora breve: 1 anno in Brasile; dimora della madre: Sora; dimora del padre: Sora; grado di cultura: 4a elementare; mezzi articolatori: buoni.

La fonte è padre di numerosa famiglia che regge con autorità e bontà e si adatta ai più vari mestieri, mentre, specie in passato, era ricercato dalle liete brigate per le sue recitazioni poetiche in dialetto e le sue facezie. Parla il dialetto moderno, vale a dire alquanto diverso da quello esposto da C. Merlo nella *Fonologia del dialetto di Sora* (Merlo 1920), pubblicata a Pisa nel 1920 ma con materiale raccolto fra il 1898 e il 1908. Tale antico dialetto si parla tuttavia ancora a poca distanza da Sora, nella campagna alquanto fuori della strada fra gli anziani e anche fra i giovani che non frequentano la città. Quanto alle qualità dell'informatore debbo notare che egli non dette prova di una grande scrupolosità, giacché spesso buttava giù la risposta senza troppo riflettere se essa fosse più o meno consentanea allo spirito del dialetto e quanto fosse più o meno italianizzata.

[Informatore 2:] Vincenzo Mattacchione; età 72; professione: contadino; dimora più lunga: Sora; dimora breve: Roma (ivi trasferitosi dal 1951, presso i suoi figli che esercitano la professione di rivenditori di frutta e verdura).

Dopo l'inchiesta a Sora con la fonte principale, ebbi occasione a Roma di avvicinare un suo parente, di 19 anni più anziano, vissuto fino al 1951 vicino al podere in cui abita la prima fonte, uomo che parla il dialetto antico, molto intelligente e coscienzioso, col quale in Roma stessa ho rivisto quasi totalmente il materiale raccolto.

(da Massobrio et al. 1995: 521, con modifiche).

Abbiamo dunque verificato nei dati digitali a nostra disposizione quanto sia rintracciabile delle osservazioni del raccoglitore. Abbiamo trovato 55 risposte marcate con il tag **ai*, che rappresentano risposte del secondo informatore, quello descritto come più conservativo: sono state riportate nella Tabella 2. In queste troviamo la conservazione dei fenomeni descritti dal raccoglitore Giacomelli, ad es. la conservazione di [w] prima di occlusiva velare (ad es. 10, 37, 38), la conservazione di [tʃ] (ad es. 6, 34), una maggior sonorizzazione delle occlusive intervocaliche (v. 3, 33, 36), che in questa varietà erano generalmente sempre desonorizzate. Si riscontrano però anche altri tratti di convergenza fonetica (e in generale nella forma di parola) verso l'italiano o il romanesco, in accordo con la dinamica descritta da Giacomelli, da parte del primo informatore: si notino in 17 e 18 le forme locali [komən'ʔsatə] e [tra'sərə] dell'informatore più anziano sostituite da forme più italianizzanti [komin'ʔfatə] e [tə'sərə], o il ripristino di [d] in 24, dove si ha [b:an'djera] presso l'informatore più giovane e [b:a'n:era] presso quello più anziano.

Informatore 1	Informatore 2	Voce	
1. sò ssombátə	*ai ~ zzəmbátə , ~ ĩəntátə , ~ ĩəntátə	391	Saltai oltre la siepe
2. sə n ab-brüónná	*ai abbrəónná	482	Pietro si vergogna di parlare
3. aččidə	*ai aččitə	605	Vuol uccidere tutti
4. assəttətə *ai conf	*ai aššitətə	941	e siediti qua!
5. paésə	*ai paiesə	1499	villaggio
6. čittà	*ai čətà	1500	città
7. lampiónə	*ai lambəiónə	1507	fanale della strada [lampione]
8. allə šittà	*ai čətà	1584	Nelle città
9. ténɡə a uni	*ai ~ ~ ũəni	1632	Quando ho da venire da te?
10. a ppəkürünə	*ai ~ ppəkürünə	1674	Luigi camminava carponi per non essere veduto
11. ũərətà	*ai ũərdà	1715	ti dirò io la verità
12. ũinčə sèmprə	*ai sèmpə	1748	e vinci sempre al gioco
13. pèrdə la pašénza	*ai pačénzĩa	1749	Insomma, non farci perdere
14. fūitə	*ai fəiūtə	1762	sono fuggiti
15. ulərriə	*ai ũəlarĩa	1885	Vorrei vendere il grano
16. è kkominčátə	*ai ~ kkəmənzátə	1894	Quest'altro poi non è pieno: è già avviato
17. koməncátə	*ai è kkəmənzátə	1896	chi l'ha incignata?
18. təsòrə	*ai trasòrə	1907	La faccenda del tesoro nascosto
19. špərašə	*ai špərgə	2006	<i>Asparagus officinalis</i> – sparagio, asparago
20. sèllənə	*ai sələnə	2007	<i>Apium graveolens</i> – sedano
21. bbuskà	*ai abbəškà	2071	volete guadagnare troppo
22. ĩlə aũndórə	*ai ~ aũndórə	2114	Se no, perdo gli avventori
23. tənīmə aĩutà	*ai ĩlə tənīmə aĩətà	2269	Tutti dobbiamo soccorrerli
24. bbandiera	*ai bbannèra	2548	bandiera
25. soldatə	*ai solətətə	2552	soldato
26. škoppétta	*ai skəppétta	2553	fucile
27. kommünə	*ai la kəmmüna	2559	Nel nostro comune
28. è pproĩbbitə	*ai è pprəĩbbitə	2560	È vietato ai forestieri
29. gũaiə a tté	*ai ũaiə ~ ~	2608	Guai a te se fai la spia!
30. puínálə	*ai pəínálə	2623	pugnale
31. priğğónə	*ai prəšónə	2624	prigione [carcere]
32. [t aččidə]	*ai [t aččitə]	2629	un giorno o l'altro tenterà di darti una coltellata
33. ġğüdišə	*ai ġğütəčə	2706	Al giudice disse tutto
34. ũilínə	*ai ũilínə	2733	violino
35. è ušínə	*ai ũəčínə	2767	È qui vicino <?>
36. sa mbriágənə	*ai sə mbriākənə	2806	Uomini e donne si ubriacano
37. kańńòmə	*ai kũəńńòmə	2815	Né il suo cognome
38. końtəntə	*ai kũəťəntə	2825	Perciò è contento
39. du sũmpə	*ai űkə ddu zũmpə, ĩũntə	3212	Ora con due salti raggiungiamo quell'albero

Tabella 2: Risposte date da entrambi gli informatori di Sora (p. 668) nei volumi IX e X con divergenze di tipo fonetico

Il carattere innovativo del primo informatore è evidente anche da tratti non descritti direttamente da Giacomelli, di carattere lessicale, come si può vedere (Tabella 3) nella sostituzione dei tipi locali *statevi buono* (4), *pastinaca* (7) e *trasire* (13) con *arrivederci*, *carota gialla* ed *entrare*. Notevole anche la sostituzione di localismi come [rəťə'ʎ:om:əɾə] (probabilmente imparentato con il napoletano ['ʎom:erə] 'gomitolo') con l'italianismo *capriola* (16).

Informatore 1	Informatore 2	Voce	
1. bbéüə	*ai ũó la mbúmba in- terr	1223	bere <i>infant</i>
2. fátə súbbətə	*ai aĩtətəüə	1667	vi prego, affrettatevi!
3. mbiččát'ə	*ai ntəppəkátə	1670	Tonio inciampò in una pietra
4. arruudérčə	*ai štətəüə bbònə	1770	Dunque, arriverci
5. i kũadrinə	*ai ílə sòldə	1843	Hai abbastanza denaro?
6. na spiga	*ai ~ fèlla	1974	spicchio d'arancia
7. karòta ġġàlla	*ai paštənáka , paštənáta	2004	Daucus carota – carota
8. məlaòra	*ai ~ , pəmmaòra	2005	<i>Lycopersicum esculentum</i> – pomodoro
9. kannəllinə rűššə	*ai fašəriílə	2012	fagiolini
10. poərèlla	*ai pòüərammé	2046	Povera me, quante spese!
11. appəččátə	*ai abbrəšátə	2510	È bruciato tutto
12. p arrubbáílə	*ai pə ttòlłəčə	2681	per rubargli i soldi
13. ĩéndra	*ai trəsə	2745	Avanti!
14. č a pĩəsérə	*ai ĩe kkűəntəntə	2797	Vedi, lei viene volentieri al ballo
15. prədűčča	*ai na ũrićča	3161	sassolino
16. kapriòla	*ai rəčəłlómmerə	3183	Sta' attento come cammini, se no fai un capi- tombolo

Tabella 3: Risposte date da entrambi gli informatori di Sora (p. 668) nei volumi IX e X con divergenze di tipo lessicale

3.3 Etichette variazionali

Come si è visto (capitolo 3.1), vi sono casi in cui è possibile rilevare variazione sociolinguistica anche nei dati forniti da informatori singoli, vale a dire quando questi sono linguisticamente consapevoli della coesistenza di forme diverse e forniscono varianti “marcate”, che nell’archivio digitale sono etichettate con marche d’uso specifiche. Queste, che sono già in parte presenti in atlanti precedenti come l’ALF e soprattutto l’AIS, rappresentano, appunto, tipi diversi di variazione sociolinguistica. Anche se l’elenco delle marche è stato definito e sistematizzato nella fase redazionale dell’Atlante, non ne è mai stata valutata la consistenza in termini quantitativi. A questo fine abbiamo estratto tutte le risposte con marche sociolinguisticamente rilevanti nei voll. IX e X: su 222.841 entrate abbiamo individuato 1049 risposte con etichette di questo tipo, cioè circa lo 0,47%. Le marche individuate sono le seguenti, raggruppate per tipo di variazione rappresentata (Tabella 4) e in base alla loro consistenza numerica tra quelle estratte (Figura 8).

Marca	Significato	Tipo di variazione
<i>borgh</i>	<i>borghese</i>	Diastratica
<i>civ</i>	<i>civile</i>	
<i>popol</i>	<i>popolare</i>	
<i>volg</i>	<i>volgare</i>	Diafasica/diastratica
<i>gerg</i>	<i>gergale</i>	Diafasica
<i>infant</i>	<i>infantile</i>	
<i>cult</i>	<i>dotto</i>	
<i>rec</i>	<i>recente</i>	Diacronica
<i>v</i>	<i>vecchio</i>	
<i>rust</i>	<i>rustico</i>	Diatopica/diastratica
<i>d</i>	<i>donne</i>	Di genere
<i>u</i>	<i>uomini</i>	
<i>corr</i>	<i>l'informatore si corregge</i>	Relativa all'uso (sociolinguisticamente rilevante?)
<i>freq</i>	<i>frequente</i>	
<i>nega</i>	<i>contrasto tra informatori</i>	
<i>rar</i>	<i>raro</i>	
<i>sic!</i>	<i>l'informatore insiste</i>	

Tabella 4: Marche d'uso di interesse sociolinguistico e variazionale

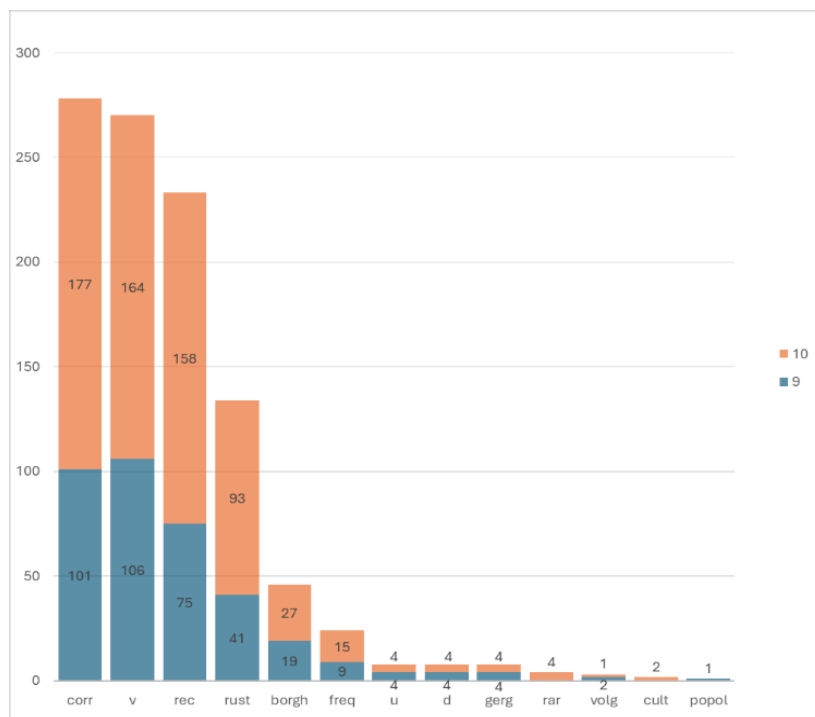


Figura 8: Distribuzione quantitativa delle marche d'uso sociolinguisticamente rilevanti nei voll. IX e X dell'ALI

In riferimento alla frequenza, si può vedere che la marca d'uso più frequente è *corr* (correzione), con 278 occorrenze, cioè quando l'informatore corregge sé stesso fornendo un'altra risposta. Pur non essendo pertinente direttamente a nessun ambito della variazione sociolinguistica, presenta comunque un interesse indiretto in questo senso, per cui saranno necessari studi specifici: gli informatori dell'ALI, infatti, non sembrano solo correggere risposte chiaramente errate o percepite come non fedeli alla propria parlata (ad es. italianismi, fatto comune nelle inchieste dialettali), ma anche la dinamica inversa, cioè la correzione di forme molto locali in favore di altre sovralocali. Un esempio della prima dinamica (correzione di una forma italianizzante o sovralocale verso una forma più locale) si può osservare ad es. a Rionero in Vulture (Potenza), punto 902, alla voce 1822 *ti prendi delle sborne*, la prima risposta *pèllə* ['pɛl:ə] è corretta dall'informatore in *pèddə* ['pɛd:ə], in cui è presente l'esito locale in luogo di quello italianizzante della prima. La dinamica opposta (correzione verso una forma non locale), più inconsueta, si ha ad es. nel punto 802 Carpino (Foggia) per la voce 1992 *ciliegia*: l'informatore Russi, peraltro descritto dal raccoglitore Melillo come "[...] tentato o meglio [...] portato ad italianizzare o ad usare una certa 'koinè' pugliese" (Massobrio et al. 1995: 642), risponde inizialmente con la forma indigena *čərasə* [tʃə'rasə] correggendola poi in *čərəsə* [tʃə'resə]; la stessa dinamica si osserva, nei dati digitali a noi disponibili, al punto 1058 Canicattini Bagni (Siracusa) per la voce 2840 *quella gente si diverte a leticare in tribunale*: l'informatore corregge il locale *ššařriarisi* [ʃ:a'ɹ:jarisi] con l'italianizzante *a llitikari* [a l:iti'kari].

La seconda è *v* (*vecchio*), che indica forme conosciute ma in disuso, utilizzate dalle generazioni precedenti, con 270 occorrenze; insieme a *rec* (*recente*), con 233 occorrenze, che indica invece forme nuove, tipiche delle generazioni giovani, rappresenta variazione di tipo diacronico in tempo apparente. La marca *rust* (*rustico*) non indica solo probabilmente un fatto di variazione geografica, ma anche una variazione di classe, essendo usata, molto spesso, in opposizione a *borgh* (*borghese*), come si potrà vedere (v. oltre): la prima ha 134 occorrenze, mentre la seconda 46. Le altre marche sono molto meno numerose nell'archivio; si segnala però la variazione di genere tra *u* (*uomo*) e *d* (*donna*), presente soprattutto al punto 471 Ancona.

Considerando la ricchezza di questi dati, seppur parziali, abbiamo provato a individuare in quali punti dell'atlante si concentrasse la maggior quantità di etichette variazionali e se, sulla base di queste, si potessero ricostruire dei pattern di variazione, anche combinando i dati estratti con le informazioni descrittive presenti nei verbali di inchiesta.

Un primo esempio viene dall'informatore di Sale Marasino (Brescia), un centro sulla sponda orientale del Lago d'Iseo. Qui, un messo comunale in pensione di 78 anni (nato nel 1859) intervistato da Ugo Pellis nel 1937, è descritto come "dalla tendenza borghese, cioè livellata di fronte al forestiero", in grado però di passare "alla schietta parlata popolana che gli è familiare", rimanendo però "talvolta [...] a mezza strada fra l'abitudine livellata e quella rustica", da cui "risulta una fluttuazione articolativa dai suoni caratteristici non ricorrenti nella lingua nazionale" (Massobrio et al. 1995: 75). I suoni a cui fa riferimento Pellis sono un tratto caratteristico di molte varietà rustiche tra Bergamo e Brescia (Sanga 1987: 19; Bonfadini 1990: 58s., 66s.; Bonfadini 2015: 281–285), cioè la perdita dell'articolazione dorsale dei fonemi /s/ e /z/ che risulta nella debuccalizzazione di /s/ in /h/ (attraverso una fase intermedia /θ/ < /ts/) e nell'avanzamento di /z/ in /ð/ (< /dz/). Bonfadini (1990: 58s.) descrive la possibilità dei parlanti dell'area bresciana passare dal sistema rustico a quello urbano a seconda del contesto comunicativo, nel

senso che /h/ e /z/ possono essere soppressi dai parlanti dei centri rustici in favore di /s/ e /z/ tipici invece delle varietà urbane: questo tratto, per molti versi, si configura come una variabile sociolinguistica. Pellis, nei verbali, riporta una descrizione molto vivida di questo uso alterante:

A proposito dell'*h* per *s* è necessario notare che col forestiero borghese si cerca di evitare l'*h*. Dico si cerca, perché non sempre ci si accorge di articolarlo, e quindi di evitarlo. L'informatore distingue, di solito scientemente, fra uso rustico (*h*) e borghese (*s*). In complesso, di questa abitudine articolativa (*h*) quasi ci si vergogna. Un uomo e una donna, fermatisi davanti all'albergo, vedendo alla finestra mia moglie, evitavano il suono *h*. Ma accaloratasi la discussione fino al litigio, ambedue perdettero ogni ritegno, e allora piovvero gli *h* senza fine e gli *s* furono sommersi completamente.

(Massobrio et al. 1995: 76).

La descrizione di Pellis rivela piuttosto chiaramente che il tratto è interpretabile come un *marcatore* (cf. Labov 2001: 196s.), nel senso che si tratta di una variabile sociolinguistica soggetta a variazione diastratica (di classe) ma anche diafasica (di stile), poiché era associata alle classi inferiori e rurali, di cui i parlanti potevano variare l'uso a seconda della situazione comunicativa (e la "perdita di controllo" descritta da Pellis è un indice chiarissimo di questo fenomeno). Allo stesso tempo, tuttavia, la sua salienza e, evidentemente, lo stigma sempre più esplicito che era associato ad esso fanno sì che la si possa interpretare anche come uno *stereotipo*, vale a dire una variabile di cui i parlanti sono particolarmente consapevoli e su cui possono intervenire (e che, solitamente, è preliminare all'eliminazione di quel tratto, perché troppo marcato).

I dati estratti per i voll. IX e X dell'ALI consentono di chiarire maggiormente il rapporto dell'informatore con questa dinamica (Tabella 5).

Risposta	tag	Voce
1. paèś , paé ^h <i>rust</i>	<i>rust</i>	1499 villaggio
2. pīāha , pīāsa <i>borgh</i>	<i>borgh</i>	1501 piazza
3. šortia , hortia <i>rust</i>	<i>rust</i>	1677 sorgente
4. poś , poḥ <i>rust</i>	<i>rust</i>	1679 Come dite la conca naturale dove l'acqua si raccoglie e la si attinge? vasca naturale
5. poś , poḥ <i>rust</i>	<i>rust</i>	1685 pozzo
6. ohtaria , ošteria <i>borgh</i>	<i>borgh</i>	1780 osteria
7. šigole , hiḡole <i>rust</i>	<i>rust</i>	1968 <i>Allium cepa</i> – cipolla
8. kōšta , kōhta <i>rust</i>	<i>rust</i>	1974 spicchio d'arancia
9. nehōla , nešōla <i>borgh</i> , nešolāna <i>gr</i>	<i>borgh</i>	1987 nocciola
10. šaréĵa , haré ^ś a <i>rust</i>	<i>rust</i>	1992 (una) ciliegia
11. špāreś , hpāreh <i>rust</i>	<i>rust</i>	2006 <i>Asparagus officinalis</i> – sparagio, asparago
12. šèleno , hèleno <i>rust</i>	<i>rust</i>	2007 <i>Apium graveolens</i> – sedano

Risposta	tag	Voce	
13. faǰól , faðól <i>rust</i>	<i>rust</i>	2011	<i>Phaseolus vulgaris</i> – fagiolo
14. kornaǰěĩ , korněč , kornaðěĩ <i>rust</i>	<i>rust</i>	2012	fagiolini
15. balánsa , balánha <i>rust</i>	<i>rust</i>	2020	bilancia
16. šoldát , holdát <i>rust</i>	<i>rust</i>	2552	soldato
17. preǰũ , preðũ <i>rust</i>	<i>rust</i>	2624	prigione [carcere]
18. šonadũr , šunadũr , honadũr <i>rust</i>	<i>rust</i>	2731	suonatori
19. ma ^h , maš <i>borgh</i>	<i>borgh</i>	5619	romano (peso) [piombino] della stadera

Tabella 5: Uso delle marche *rust* e *borgh* da parte dell'informatore di Sale Marasino nei volumi IX e X

Nei dati dell'ALI è rappresentato il fenomeno sia per la sorda [s] > [h] sia per la sonora [z] > [ð] (10, 14, 17); soprattutto, è chiaro come l'informatore oscilli in maniera costante tra le due forme, etichettando in maniera complementare come *rustiche* le forme con [h] e [ð] e *borghesi* quelle con [s] e [z]. Emerge quindi il fatto che l'informatore è sostanzialmente oscillante e non sembra propendere per le une o per le altre in maniera decisiva. Ciò che è importante è che l'ALI ci consente di collocare abbastanza bene nel tempo la dinamica di riduzione di queste forme rurali, cioè un mutamento che trova la sua causa in motivazioni di ordine sociolinguistico, sostenuto dalla varietà urbana: un parlante di uno strato sociale mobile (messo comunale) di un piccolo centro, nato a metà del XIX ne era soggetto, agendo da veicolo per la sua diffusione.

Un altro caso piuttosto rilevante è quello di Grosseto (punto 570). Anche qui l'informatore, descritto come parlante urbano, conosce le varianti “rustiche” (Massobrio et al. 1995: 451), etichetta con cui, come si è visto (Tabella 4), si può intendere una variazione sia diastratica sia diatopica. Osservando i dati per Grosseto (Tabella 6), si vede che la marca *rustico* è attribuita a forme locali, spesso con caratteristiche più genericamente mediane o italiane centrali che toscane, mentre le forme non rustiche o urbane sono appunto toscane comuni; particolarmente notevole è il caso in 5., in cui è presente la spirantizzazione dell'occlusiva velare (tipica della “gorgia” toscana) su /k/ nella forma urbana non marcata e la sonorizzazione centromeridionale in quella marcata come *rustica*. Con la marca *borghese* sono etichettate forme convergenti verso l'italiano (soprattutto la conservazione di -r(e)/ nell'infinito, ad es. 3, 8, 13 e la forma per *ciliegia* in 10), con una variazione di tipo diastratico. Siamo dunque in presenza di una convergenza doppia, verso una “koiné regionale” toscana di tipo fiorentino mediata da Siena (cf. Giannelli 2000: 88s.), ma anche verso l'italiano (quest'ultima si manifesta con la cancellazione dei tratti di tipo locale e di quelli della koiné regionale). Si noti che, comunque, spesso le due tendenze non sono del tutto separabili a causa della somiglianza strutturale di italiano e fiorentino. In ogni caso, questi dati sono piuttosto interessanti perché sono interpretabili come un processo attivo di diffusione del fiorentino all'inizio del XX secolo nella città di Grosseto, e la sua assenza negli immediati dintorni, che conservavano ancora tratti linguistici non “toscani comuni”, di tipo italiano centrale. L'evidenza fornita dall'ALI è considerevole, perché per

quest'area sono disponibili pochi dati storici e in generale oggi è livellata linguisticamente, come del resto molte zone della Toscana (cf. Marzini 2006).

Risposta	tag	Voce
1. nde lẹ ċittà <i>rust</i> , nẹl'ẹ ċittà	<i>rust</i>	1584 Nelle città
2. šfērza <i>rust</i> , frūsta	<i>rust</i>	1611 frusta
3. dẹvọ vẹni, ~ vẹnirẹ <i>borgh</i>	<i>borgh</i>	1632 Quando ho da venire da te?
4. ʃdruččolāi, ʃdrušolāi <i>rust</i>	<i>rust</i>	1672 Io sdrucchiolai su una buccia
5. fati ^h a, fadiga <i>rust</i>	<i>rust</i>	1703 la fatica le pesa
6. ęntriāmo, entrāmo <i>rust</i>	<i>rust</i>	1790 entriamo
7. mẹ lọ ripiēni <i>rust</i> , ~ ~ riēmpi	<i>rust</i>	1803 favorisci riempirlo
8. a k ^k omprāllo <i>rust</i> , ~ k ^k omprārlo	<i>rust</i>	1873 Che vada a comperarne
9. tẹsòrọ <i>rust</i> , tẹfòrọ	<i>rust</i>	1907 La faccenda del tesoro nascosto
10. sarāš'a <i>rust</i> , čilēš'a <i>borgh</i>	<i>rust/borgh</i>	1992 (una) ciliegia
11. špàrāš'o, špàrāš'o <i>rust</i>	<i>rust</i>	2006 <i>Asparagus officinalis</i> – sparagio, asparago
12. grātis, sènza nènτε <i>rust</i>	<i>rust</i>	2102 I campioni si danno gratis
13. si dẹve aiūtālli <i>rust</i> , ~ ~ aiūtārli <i>borgh</i>	<i>rust/borgh</i>	2269 Tutti dobbiamo soccorrerli
14. ẹ rré <i>rust</i> , i ~	<i>rust</i>	2549 re
15. ndel [...] ^h omùne <i>rust</i> , nel [...] ~	<i>rust</i>	2559 Nel nostro comune
16. ččerkà la-lemòš'ina <i>rust</i> , ččerkà l'elēmòš'ina	<i>rust</i>	2561 di andare a mendicare
17. barattiāmo, barattāmo <i>rust</i>	<i>rust</i>	2595 Cambiamo!
18. sta l'ontāno, šta ~ <i>rust</i>	<i>rust</i>	2769 Allora abita lontano da qui
19. affāre, interèsse <i>rust</i>	<i>rust</i>	2824 ha fatto un buon affare

Tabella 6: Uso delle marche *rust* e *borgh* da parte dell'informatore di Grosseto nei volumi IX e X

Meno complessa, ma piuttosto chiara, è la dinamica che si può registrare a Petritoli (Fermo), punto 494, nelle Marche. Pellis descrive in questi termini lo stato della parlata locale (Massobrio et al. 1995: 373):

La parlata di Petritoli sta attraversando un periodo di evoluzione più sensibile che in altre località. I vecchi e i contadini usano comunemente articolare suoni come *ö* e *α* ecc., mentre i borghesi tendono a livellare pronunciando, invece di *ö*, *é* oppure *è*; invece di *α*, un *e* più o meno aperto. Ho assistito a colloqui spontanei per la strada e all'osteria, senza che la mia presenza portasse turbamento, e ho constatato un fluttuare di articolazione di suoni fortissimo. Ognuno aveva qualche particolarità propria, qui più accentuata che altrove. In tutte c'era una tendenza a livellare. La spinta più forte viene naturalmente dalla borghesia, sebbene in forte minoranza.

(Massobrio et al. 1995: 373)

Queste tendenze sono piuttosto evidenti nei dati forniti dall'informatore (Tabella 7), piuttosto giovane all'epoca dell'inchiesta (34 anni nel 1931), che tende a valutare come “recenti” forme che perdono tratti locali marcati. Si ritrovano, nei dati estratti, non solo i fenomeni descritti da Pellis come [ø] > [e] (ad es. 3, 4, 8, 14 ecc.) [Λ] > [ε] (ad es. 9, 20 ecc.), ma anche la perdita della palatalizzazione in [ε] di [a] tonica, tipica dei dialetti tra le Marche, la Romagna e

l'Umbria (ad es. 1, 2, 7 ecc.) e la soppressione di tratti “mediani”, come il vocalismo in -[u] finale e il ripristino dell'occlusiva nel gruppo [nd] (v. ad es. entrambi i fenomeni in 1., 5). Qui si osserva dunque una convergenza verso l'italiano, come del resto avviene nella vicina Ancona; non sembrano essere presenti particolari dinamiche regionali o locali di convergenza.

Risposta	tag	Voce	Risposta
1. so zumbètu , ~ zumbàt ^o <i>rec</i>	<i>rec</i>	391	so zumbètu , ~ zumbàt ^o <i>rec</i>
2. mmazzè , mmazzà <i>rec</i>	<i>rec</i>	605	mmazzè , mmazzà <i>rec</i>
3. pójja , pèjja <i>rec</i>	<i>rec</i>	940	pójja , pèjja <i>rec</i>
4. móttetè a-ssedè , mètterè ~ ~ <i>rec</i>	<i>rec</i>	941	móttetè a-ssedè , mètterè ~ ~ <i>rec</i>
5. lu šinn'ku , sindakò <i>rec</i>	<i>rec</i>	1403	lu šinn'ku , sindakò <i>rec</i>
6. la pièzza , piàzza <i>rec</i>	<i>rec</i>	1501	la pièzza , piàzza <i>rec</i>
7. lu sarġètu , ~ sarġàtu <i>rec</i>	<i>rec</i>	1504	lu sarġètu , ~ sarġàtu <i>rec</i>
8. lu vòtturó , vòtturé <i>rec</i>	<i>rec</i>	1615	lu vòtturó , vòtturé <i>rec</i>
9. vòna , vèna <i>rec</i>	<i>rec</i>	1677	vòna , vèna <i>rec</i>
10. la kanđóna , kanđéna <i>rec</i>	<i>rec</i>	1780	la kanđóna , kanđéna <i>rec</i>
11. lu palló , ~ pallé <i>rec</i>	<i>rec</i>	1787	lu palló , ~ pallé <i>rec</i>
12. vuttètu , vuttàtu <i>rec</i>	<i>rec</i>	1809	vuttètu , vuttàtu <i>rec</i>
13. mbriè ^k u , mbrià ^k ò <i>rec</i>	<i>rec</i>	1819	mbriè ^k u , mbrià ^k ò <i>rec</i>
14. špóčči , spécči <i>rec</i>	<i>rec</i>	1844	špóčči , spécči <i>rec</i>
15. sènu , sànu <i>rec</i>	<i>rec</i>	1893	sènu , sànu <i>rec</i>
16. kumin ^š ètu , kumin ^š à'u <i>rec</i>	<i>rec</i>	1894	kumin ^š ètu , kumin ^š à'u <i>rec</i>
17. próta , préta <i>rec</i>	<i>rec</i>	1895	próta , préta <i>rec</i>
18. □ spezzèta , □ spezzà'a <i>rec</i>	<i>rec</i>	1896	□ spezzèta , □ spezzà'a <i>rec</i>
19. štàva dešperètu , dešperàtu <i>rec</i>	<i>rec</i>	1934	štàva dešperètu , dešperàtu <i>rec</i>
20. l erbátte , erbétta <i>rec</i>	<i>rec</i>	1969	l erbátte , erbétta <i>rec</i>
21. la špóka , ~ špéka <i>rec</i>	<i>rec</i>	1974	la špóka , ~ špéka <i>rec</i>
22. la pàra , pèra <i>rec</i>	<i>rec</i>	1979	la pàra , pèra <i>rec</i>
23. la mǎla , mèla <i>rec</i>	<i>rec</i>	1981	la mǎla , mèla <i>rec</i>
24. la nǝalèta , ~ nǝalàta <i>rec</i>	<i>rec</i>	1998	la nǝalèta , ~ nǝalàta <i>rec</i>
25. li kèvu <u>li</u> , kàvu <u>li</u> <i>rec</i>	<i>rec</i>	2000	li kèvu <u>li</u> , kàvu <u>li</u> <i>rec</i>
26. la patèta , patàta <i>rec</i>	<i>rec</i>	2009	la patèta , patàta <i>rec</i>
27. piècè , pià <i>rec</i>	<i>rec</i>	2048	piècè , pià <i>rec</i>
28. sǎkki , sékki <i>rec</i>	<i>rec</i>	2049	sǎkki , sékki <i>rec</i>
29. fraški , fréski <i>rec</i>	<i>rec</i>	2050	fraški , fréski <i>rec</i>
30. rrózza , rrézza <i>rec</i>	<i>rec</i>	2505	rrózza , rrézza <i>rec</i>
31. čóka , čéka <i>rec</i>	<i>rec</i>	2545	čóka , čéka <i>rec</i>
32. lu rrá , rē <i>rec</i>	<i>rec</i>	2549	lu rrá , rē <i>rec</i>
33. lu surdè , surdà <i>rec</i>	<i>rec</i>	2552	lu surdè , surdà <i>rec</i>
34. lu fučólu , fučéle <i>rec</i>	<i>rec</i>	2553	lu fučólu , fučéle <i>rec</i>
35. lu štóllu , štéllu <i>rec</i>	<i>rec</i>	2623	lu štóllu , štéllu <i>rec</i>
36. la fèrrèta , fèrràta <i>rec</i>	<i>rec</i>	2625	la fèrrèta , fèrràta <i>rec</i>

Risposta	tag	Voce	Risposta
37. lu viuló , ~ viulé <i>rec</i>	<i>rec</i>	2733	lu viuló , ~ viulé <i>rec</i>
38. lu ruvendó , ~ ruvendé <i>rec</i>	<i>rec</i>	3159	lu ruvendó , ~ ruvendé <i>rec</i>
39. sèssu , sàssu <i>rec</i>	<i>rec</i>	3161	sèssu , sàssu <i>rec</i>
40. sèssu , sàssu <i>rec</i>	<i>rec</i>	3162	sèssu , sàssu <i>rec</i>
41. du sèssi , sàssi <i>rec</i>	<i>rec</i>	3163	du sèssi , sàssi <i>rec</i>
42. rrózzete , rézzete <i>rec</i>	<i>rec</i>	3196	rrózzete , rézzete <i>rec</i>
43. lu mèrku , màrku <i>rec</i>	<i>rec</i>	5619	lu mèrku , màrku <i>rec</i>
44. pìezza dell-èrba , pìazza ~ ~ <i>rec</i>	<i>rec</i>	6417	pìezza dell-èrba , pìazza ~ ~ <i>rec</i>
45. vètróna , vètréna <i>rec</i>	<i>rec</i>	6420	vètróna , vètréna <i>rec</i>

Tabella 7: Uso della marca *rec* da parte dell'informatore di Petritoli nei volumi IX e X

3.4 Le “note”

L'estrazione dei dati ci ha permesso anche di osservare in maniera trasversale dinamiche del tutto particolari, anche non legate alle marche d'uso. Un caso singolare e per molti versi eccezionale è Gilba Superiore, frazione di Brossasco, un comune della Val Varaita in Piemonte. Qui l'informatore (nato nel 1903, messo comunale) mostra una competenza diglossica sia della varietà locale provenzale alpina (o occitana alpina) sia nel dialetto piemontese del centro principale di Brossasco (nella parte bassa della valle), che si avvicina alla varietà di *koiné* a base torinese della vicina Saluzzo. La forma locale è sempre la risposta principale, mentre quella piemontese è riportata sistematicamente nelle note (Tabella 8). È una dinamica ben nota per cui il piemontese era generalmente conosciuto tra i parlanti di occitano come lingua veicolare, anche perché nessun grande centro urbano del Piemonte era di lingua occitana. I dati dell'informatore di Gilba sono preziosi anche in virtù della loro abbondanza: si tratta di molte decine di entrate doppie di questo tipo in cui sono rappresentate le due varietà del suo repertorio. La disponibilità di questi dati in massa, e non separati in centinaia di carte linguistiche, consentirà di studiare con maggior efficacia questo caso e altri di questo tipo¹⁸.

Risposta	note	Domanda	
p ^r fümá	<i>a Brossasco dicono: per fümé</i>	461	A che cosa serve la pipa? per fumare
bóġga	<i>a Brossasco dicono: bùġa</i>	944	La tavola tentenna
na vil ^{la} , rüà	<i>è la frazione, detta a Brossasco burgà</i>	1499	villaggio
la séul ^a	<i>a Brossasco dicono: šùl'a</i>	1968	<i>Allium cepa</i> – cipolla
čaričēġa	<i>a Brossasco dicono: čarēġa</i>	1992	(una) ciliegia
majòla	<i>a Brossasco dicono: friòla</i>	1995	(una) fragola
triful ^a	<i>a Brossasco dicono: patàta</i>	2009	<i>Solanum tuberosum</i> – patata
faġòl	<i>a Brossasco dicono: faġòl</i>	2011	<i>Phaseolus vulgaris</i> – fagiolo
móŭra	<i>a Brossasco dicono: múra</i>	2037	Quest'uva è matura?
pōī	<i>a Brossasco dicono: pōu</i>	3366	Gesummaria, che paura!

Tabella 8: Note da cui traspare la competenza bilingue dell'informatore di Gilba (volumi IX e X)

¹⁸ Si pensi ad es. al punto 61 Spinetta Marengo, frazione di Alessandria a est del fiume Bormida, per cui sistematicamente sono segnalate varianti diverse tra la parte est e quella ovest del centro, che rappresentano una transizione tra il tipo dialettale alto-monferrino di Alessandria e quello più orientato verso il pavese di Tortona.

3.5 La variazione linguistica nell'ALI: una valutazione complessiva

Dobbiamo osservare che questi risultati si basano su un piccolo sottoinsieme del corpus totale di dati dell'ALI e sono stati eseguiti su una prima versione del software/database. Nonostante questo, alcune tendenze variazionali sono abbastanza forti da poter essere rilevate. La presenza di più informatori spesso è indice della presenza di una certa quantità di variazione linguistica: sia diastatica, sia diacronica in tempo apparente. Inoltre, alcuni informatori sembrano essere attenti alle forme più antiche della generazione precedente e alle forme socialmente marcate. Particolarmente rilevanti sembrano essere anche le autocorrezioni degli informatori, che possono rivelare dati di interesse sociolinguistico.

Anche da un altro punto di vista la quantità di variazione linguistica rappresentata nell'ALI è significativa, se si considera che è a tutti gli effetti uno dei pochi atlanti linguistici plurilingui esistenti¹⁹, in quanto contiene dati da 97 tra punti bilingui/trilingui e alloglotti, cioè da lingue slave (sloveno, croato), lingue germaniche, istroromeno, romani (in Abruzzo), albanese, greco, catalano (Alghero). È poi degno di nota che l'ALI contenga anche dati sull'italiano: in Alto-Adige, Istria e Venezia Giulia Pellis raccolse dati sulla competenza individui bilingui, motivata dall'ideologia nazionalista del tempo, con l'intento di verificare la diffusione e il grado di espansione della lingua nazionale presso queste comunità; nondimeno, si tratta di un'importante testimonianza per la storia del plurilinguismo²⁰.

In sostanza, si può senz'altro affermare che i dati dell'ALI permettono di individuare e collocare nel tempo alcuni pattern fondamentali del cambiamento dei dialetti italiani: la convergenza verso varietà urbane locali e/o *koinai* sovrالocali, così come la convergenza verso l'italiano e la persistenza di dinamiche locali caratteristiche. La versione digitale dell'Atlante sarà dunque concepita anche per consentire lo studio di fenomeni di questo tipo.

4 Conclusioni

La rilevanza dei dati ALI oggi è diversa rispetto a quella che si immaginava in origine: se l'obiettivo era ottenere una fotografia latamente “sincronica” della situazione linguistica dell'Italia della prima metà del XX secolo, oggi questi dati assumono una valenza sociolinguistica storica²¹. In questa prospettiva, il concetto di “mutamento interno” risulta problematico, poiché non tiene conto dell'influenza di contesti socioculturali, dei contatti linguistici e della diffusione di varietà diatopiche di altra origine. La dialettologia variazionista, in particolare, consente di interpretare fenomeni sincronicamente attestati come espressione di una variazione

¹⁹ Un esempio è lo *Sprachatlas des Deutschen Reichs* di Georg Wenker, che conteneva anche dati sulle varietà non germaniche dell'Impero Tedesco. Una versione digitale di questo atlante, il *Digitaler Wenker-Atlas* (DiWA), è consultabile sul sito del progetto *Regionalsprache.de* (REDE).

²⁰ Per l'Alto Adige cf. Ronco 1980: 138s.

²¹ La sociolinguistica storica si configura come una disciplina volta a indagare la variazione linguistica in prospettiva diacronica, superando la tradizionale dicotomia tra mutamento “interno” e “esterno”. In particolare, si mette in discussione l'assunto neogrammatico secondo cui il cambiamento linguistico deriverebbe esclusivamente da fattori interni al sistema. Come evidenziato da Labov (1994), la variazione sincronica, ossia le differenze osservabili in un determinato momento storico, è spesso motivata da fattori extralinguistici, quali le dinamiche sociali, e costituisce il motore stesso del mutamento linguistico.

socialmente motivata e storicamente significativa. I dati archiviati dell'ALI assumono un ruolo fondamentale in tal senso, poiché offrono evidenze empiriche che sfuggono a una concezione univoca e strutturale del dato linguistico, rivelando invece l'intreccio tra variazione diatopica e diastratica e fenomeni di contatto²², imponendo una riflessione su come si debba considerare il mutamento linguistico come il risultato di una complessa interazione tra fattori interni ed esterni, in cui la dimensione sociale della lingua svolge un ruolo determinante. La variazione sincronica, opportunamente interpretata, si configura dunque non come una semplice fotografia del presente, ma come una finestra privilegiata sulla dinamica evolutiva della lingua nel tempo.

Il progetto DigitALI, che renderà possibile lo studio dei dati dell'Atlante in tale prospettiva, ha raggiunto uno stadio per certi versi avanzato, ma restano aperte questioni metodologiche e tecniche cruciali. Una delle priorità è lo sviluppo di un'interfaccia cartografica interattiva per l'interrogazione del database, realizzata in collaborazione con un'équipe di geografi dell'Università di Roma Sapienza nell'ambito dello Spoke 3 del Partenariato PNRR Changes, coordinata da Cristiano Pesaresi. I primi prototipi sono già stati elaborati con il software ArcGIS e permetteranno una rappresentazione dinamica dei dati linguistici georeferenziati.

Dal punto di vista meramente grafico la rappresentazione cartografica riprodurrà in modo relativamente fedele le mappe cartacee, salvo ovviamente la possibilità di un'interrogazione dinamica del corpus di dati e la rinuncia ai "complementi", vale a dire alle voci pubblicate in elenco negli spazi bianchi del volume. L'essenzialità della raffigurazione del dato è infatti un elemento imprescindibile per la lettura e l'analisi delle carte: diffidiamo infatti di quelle rappresentazioni in cui i dati si affastellano rendendo faticosa e dispersiva la consultazione dell'atlante. Saranno i singoli studiosi ad attivare i collegamenti che ritengono necessari tra dati di natura diversa, in parte disponibili dall'ecosistema digitale dell'ALI in parte no.

Un altro nodo centrale riguarda la conversione delle trascrizioni fonetiche dell'ALI in IPA, un passo necessario per garantire interoperabilità con altri atlanti digitali e strumenti di analisi linguistica. Tuttavia, la grafia ALI è caratterizzata da una combinazione complessa di lettere e diacritici, che non trova una corrispondenza biunivoca nell'IPA. Questo implica un processo di traslitterazione che, sebbene tecnicamente realizzabile con metodologie già applicate alla retro-standardizzazione di altri atlanti (Geisler/Forkel/List 2021), comporterà una semplificazione inevitabile, in particolare nella notazione dell'apertura vocalica. La sfida sarà ridurre al minimo la perdita di informazioni, garantendo al contempo la compatibilità con database linguistici internazionali. Ulteriori livelli di semplificazione della trascrizione potranno svolgere funzioni di servizio o precedere una fase di tipizzazione formale: si tratta tuttavia di un passaggio irto di difficoltà se gestito con leggerezza.

²² Cf. ad es. Mancini (2019) sulla variazione nei repertori grafici latini tra la sequenza digrafica <xs> e il grafema unico <x>, influenzata da usi grafici greci. Questo esempio dimostra come anche in epoca antica la variazione potesse essere frutto di contatti linguistici e culturali, e non semplicemente il risultato di un'evoluzione interna al sistema. Sul piano teorico, Milroy (1998) argomenta che il mutamento linguistico non nasce necessariamente nel luogo o nel tempo in cui viene osservato, ma può essere il risultato della diffusione di innovazioni localizzate altrove, spesso introdotte da gruppi socialmente marcati. Tali innovazioni si generalizzano attraverso le reti sociali fino a diventare parte integrante del diasistema (cf. ad es. Milroy/Milroy 1985, che hanno analizzato la diffusione del cambiamento linguistico tra i diversi gruppi sociali parlanti inglese a Belfast, mostrando come la variazione sincronica sia in grado di anticipare e spiegare il mutamento diacronico).

La digitalizzazione dei dati dell'ALI apre inoltre alla possibilità di analisi dialettometriche avanzate, grazie all'impiego dell'algoritmo di Levenshtein per la comparazione fonetica, già applicato con successo all'Atlante Lessicale Toscano (ALT) e ad altre risorse geolinguistiche (cf. Wieling et al. 2016; Rubehn/Montemagni/Nerbonne 2024). Inoltre, il database potrà permettere di estrarre automaticamente forme “normalizzate”, semplificando l'indicizzazione e la tipizzazione lessicale delle risposte dialettali. A differenza di atlanti regionali più omogenei, tuttavia, l'ALI copre una vasta gamma di varietà linguistiche, rendendo inapplicabili alcune metodologie di categorizzazione automatica basata su corrispondenze fonetiche, come è stato fatto per l'ALT (cf. Picchi/Montemagni/Biagini 2001).

Nel lungo termine, il progetto mira a consolidare la sua adesione ai principi FAIR, favorendo la condivisione e il riutilizzo dei dati con altri progetti geolinguistici, tra cui quelli del PRIN DigitALIa, che prevede l'integrazione digitale con diversi atlanti regionali italiani. Parallelamente, si sta lavorando all'integrazione del database con il Fondo Fotografico Ugo Pellis, già digitalizzato dalla Società Filologica Friulana. Questa sinergia tra dati linguistici e documentazione visiva rappresenterebbe un valore aggiunto per la ricerca dialettologica ed etnografica, offrendo una visione più completa delle realtà linguistiche documentate.

In definitiva, DigitALI non è solo un'iniziativa di archiviazione digitale, ma un progetto che ridefinisce il modo in cui i dati dialettali vengono trattati e resi accessibili, con implicazioni significative per la dialettologia, la linguistica storica e le *digital humanities*.

Bibliografia

- AIS: Jaberg, Karl/Jud, Jakob (eds.) (1928–1940): *Sprach- und Sachatlas Italiens und der Südschweiz*. 7 voll. Zoltingen:Verlagsanstalt Ringier & Co.
- ALF: Gilliéron, Jules/Edmont, Edmond (eds.) (1902–1910) : *Atlas Linguistique de la France*. Paris : Champion.
- Atlante Linguistico Italiano* (ALI): Matteo et al. (eds.) (1995–2011): *Atlante Linguistico Italiano*. Voll. I–VIII. Roma, Istituto Poligrafico e Zecca dello Stato. Vol. IX–X. Torino, Istituto dell'Atlante Linguistico Italiano, 2018–. atlantelinguistico.it/ [29.06.2025].
- Bonfadini, Giovanni (1990): “Il dialetto bresciano: modello cittadino e varietà periferiche”. *Rivista Italiana di Dialettologia* 14: 41–92.
- Bonfadini, Giovanni (2015): “Il dialetto bresciano alla luce delle ricerche più recenti”. In: Piotti, Mario (ed.): *Dalla scripta all'italiano: Aspetti, momenti, figure di storia linguistica bresciana*, *Annali di Storia Bresciana* 3. Brescia, Editrice Morcelliana: 263–294.
- Colcuc, Beatrice/Zacherl, Florian (2022): “Tools for data processing and visualization in the project VerbaAlpina”. In: Palliwoda, Nicole (ed.): *Data Processing and Visualization in Variational Linguistics/Dialectology*. Szeged/Hamburg, University of Szeged: 55–70. DOI: 10.14232/wpcl.2022.7.3.
- Cugno, Federica/Massobrio, Lorenzo (2010): *Gli atlanti linguistici della Romania*. Alessandria: Edizioni dell'Orso.
- Cugno, Federica/Rivoira, Matteo/Ronco, Giovanni (2020): “L'Atlante Linguistico Italiano (ALI)”. *Romance Philology* 74: 191–216
- Cysouw, Michael (2018): *cysouw/qlcData: Zenodo release* (Version v0.2.1.0). Zenodo. DOI: 10.5281/zenodo.1137278.

- DiWA: *Digitaler Wenker-Atlas*. uni-marburg.de/de/fb09/dsa/projekte/abgeschlossene-projekte/digitaler-wenkeratlas-diwa? [07.07.2026].
- Forkel, Robert/Moran, Steven (2018): *cldf/segments: Unicode Standard tokenization* (Version v1.2.2). Zenodo. github.com/cldf/segments/tree/v1.2.2. DOI: org/10.5281/zenodo.1296127 [28.05.2025].
- Gauchat, Louis (1905): “L’unité phonétique dans le patois d’une commune”. In: Bovet, Ernest et al. (eds): *Aus Romanischen Sprachen und Literaturen: Festschrift Heinrich Morf*. Halle, Niemeyer: 175–232.
- Gauchat, Louis/Jeanjaquet, Jules/Tappolet, Ernest (eds.) (1925): *Tableaux phonétiques des patois suisses romands. Relevés comparatifs d’environ 500 mots dans 62 patois-types. Publiés avec introduction, notes, carte et répertoires*. Neuchâtel: Attinger.
- Geisler, Hans/Forkel, Robert/List, Johann-Mattis (2021): “A digital, retro-standardized edition of the *Tableaux Phonétiques des Patois Suisses Romands* (TPPSR)”. In: Thibault, André et al. (eds.): *Nouveaux regards sur la variation dialectale – New Ways of Analyzing Dialectal Variation*. Strasbourg, Éditions de Linguistique et de Philologie: 13–36.
- Genre, Arturo (1970): “L’inchiesta di Ugo Pellis per l’ALI a Prali (Torino)”. *Bollettino dell’Atlante Linguistico Italiano* 17–18: 18–32.
- Giannelli, Luciano (2000): *Toscana*. Pisa: Pacini.
- Grassi, Corrado (1957): “La valutazione della fonte nei nuovi rilievi dell’ALI”. *Bollettino dell’Atlante Linguistico Italiano* 2: 8–20.
- Krefeld, Thomas/Lücke, Stephan (eds.) (2018): *Berichte aus der digitalen Geolinguistik* (Korpus im Text, 6). Version 2. kit.gwi.uni-muenchen.de/?p=4498&v=2 [29.06.2025]
- Krefeld, Thomas/Lücke, Stephan (2023): *FAIRNESS – Medien im methodologischen Zentrum der Geolinguistik*. Version 3 (22.06.2023, 16:20). In: Krefeld, Thomas/Lücke, Stephan/Mutter, Christina (eds): *Berichte aus der digitalen Geolinguistik (II): Vernetzung und Nachhaltigkeit*. (Korpus im Text 9), Version 30. kit.gwi.uni-muenchen.de/?p=45335&v=3 [29.06.2025].
- Labov, William (1994): *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.
- Labov, William (2001): *Principles of Linguistic Change. Volume II. Social Factors*. Oxford: Blackwell.
- List, Johann Mattis et al. (eds.) (2025). *CLLD Concepticon 3.4.0* [Data set]. Zenodo. doi.org/10.5281/zenodo.14923561.
- Loporcaro, Michele (2013): *Profilo linguistico dei dialetti italiani*. Roma-Bari: Laterza.
- Loporcaro, Michele et al. (2021): “AIS, reloaded: A digital dialect atlas of Italy and southern Switzerland”. In: Thibault, André et al. (eds.): *Nouveaux regards sur la variation dialectale – New Ways of Analyzing Dialectal Variation*. Strasbourg, Éditions de Linguistique et de Philologie: 111–136.
- Mancini, Marco (2019): “Repertori grafici e regole d’uso: il caso latino <xs>”. In: Agostiniani, Lucio/Marchese, Maria Pia (eds.): *Lingua testi, storia. Atti della giornata di studi in ricordo di Aldo Luigi Prosdocimi*. Roma, Giorgio Bretschneider Editore: 13–54.
- Manzini, Maria Rita/Savoia, Leonardo Maria (2005): *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell’Orso. manzinisavoia.changes.unifi.it/ [29.06.2025].

- Marzini, Monica (2006): *Lessicografia in aree complesse: il caso di Grosseto*. In: Corino, Elisa/Marello, Carla/Onesti, Cristina (eds.): *Atti del XII Congresso Internazionale di Lessicografia – Euralex (Torino, 6–9 settembre 2006)*. Alessandria, Edizioni dell’Orso: 689–699.
- Massobrio, Lorenzo et al. (1995): *ALI. Verbalì delle inchieste*. 2 voll. Roma: Istituto Poligrafico e Zecca dello Stato.
- Mazzaggio, Greta et al. (2023/2024): “Morphosyntax of Italian and Romance Varieties: Presentation of the Manzini and Savoia (2005) Corpus and Its Digitalization”. *Bollettino dell’Atlante Linguistico Italiano* 47: 185–210.
- Merlo, Clemente (1920): “Fonologia del dialetto di Sora (Caserta)”. *Annali delle Università Toscane* 4: 117–283.
- Milroy, James (1998): “Toward a speaker-based account of language change”. In: Jahr, Håkon Ernst (ed.): *Advances in Historical Sociolinguistics*. Berlin, Mouton de Gruyter: 21–36.
- Milroy, James/Milroy, Lesley (1985): “Linguistic change, social network and speaker innovation”. *Journal of Linguistics* 21/2: 339–384.
- Moran, Steven/Cysouw, Michael (2018): *The Unicode Cookbook for Linguists: Managing Writing Systems Using Orthography Profiles*. Berlin: Language Science Press. (= *Translation and Multilingual Language Processing* 10). DOI: 10.5281/zenodo.1296780.
- Picchi, Eugenio/Montemagni, Simonetta/Biagini, Lisa (2001): “DBT-ALT: a System for Storing and Querying the Data of the *Atlante Lessicale Toscano* (ALT)”. *Dialectologia* 9: 85–103.
- R Core Team (2021): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. R-project.org/ [29.06.2025].
- Rivoira, Matteo (2016): “L’archivio dell’ALI: informatizzazione dei dati e prospettive di sviluppo”. In: Avolio, Francesco/Cigno, Antonino (eds.): *Archivi Etnolinguistici Multimediali dalla formazione alla gestione e al dialogo col territorio, Pescara, 6 ottobre 2012, Atti del Convegno*. Pescara, Museo delle genti d’Abruzzo: 49–61.
- Ronco, Giovanni (1980): “Confronto fra le inchieste linguistiche ed etnografiche compiute dai due atlanti linguistici nazionali (ALI-AIS)* nell’Alto Adige”. *Archivio per l’Alto Adige* LXXIV: 137–148.
- Rubehn, Arne/Montemagni, Simonetta/Nerbonne, John (2024): “Extracting Tuscan phonetic correspondences from dialect pronunciations automatically”. *Language Dynamics and Change* 14/1: 1–33.
- Sanga, Glauco (ed.) (1987): *Lingua e dialetti di Bergamo e delle valli*. 3 voll. Bergamo: Lubrina.
- Società Filologica Friulana. <https://www.filologicafriulana.it/biblioteca-e-archivi/fondo-fotografico-ugo-pellis/> [29.06.2025].
- Terracini, Benvenuto A. (2024): *Il parlare di Usseglio* (ed. Paolo Benedetto Mas). Lanzo: Società Storica delle Valli di Lanzo.
- Thibault, André et al. (eds.) (2021): *Nouveaux regards sur la variation dialectale – New Ways of Analyzing Dialectal Variation*. Strasbourg, Éditions de Linguistique et de Philologie: 13–36.
- Wickham, Hadley et al. (2023): *dplyr: A Grammar of Data Manipulation*. R package version 1.1.1. <https://CRAN.R-project.org/package=dplyr> [29.06.2025].

- Wickham, Hadley/Vaughan, Davis/Grich, Maximilian (2023): *tidyr: Tidy Messy Data*. R package version 1.3.0. <https://CRAN.R-project.org/package=tidyr> [28.05.2025].
- Wieling, Martijn et al. (2016): “ALT Explored: Integrating an Online Dialectometric Tool and an Online Dialect Atlas”. *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (LREC’16). Portorož, European Language Resources Association: 3265–3272. aclanthology.org/L16-1520/ [29.06.2025].
- Wikidata. [wikidata.org/wiki/Wikidata:Main_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page) [04.07.2025].