

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

I progetti DiWaC e ArchiWals

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1670241> since 2022-09-22T08:24:13Z

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

ISTITUTO
DELL'ATLANTE
LINGUISTICO
ITALIANO



UNIVERSITÀ DI TORINO

*Opera promossa dalla
Società Filologica Friulana
«G.I. Ascoli»*

BOLLETTINO

DELL'ATLANTE LINGUISTICO ITALIANO

*III Serie - Dispensa N. 41
2017*

I contributi pubblicati nella Rivista sono sottoposti a un processo
di *blind peer review* che ne attesta la validità scientifica

Questo periodico è pubblicato con il contributo del Ministero dei Beni Culturali

©
Istituto dell'Atlante Linguistico Italiano
Via Sant'Ottavio, 20 - 10124 Torino (Italia)

È vietata la riproduzione, anche parziale, non autorizzata, con qualsiasi mezzo effettuata, compresa la fotocopia, anche a uso interno o didattico. L'illecito sarà penalmente perseguibile ai sensi della Legge n. 159 del 22 maggio 1993.

ISSN 1122-1836
ISBN 978-88-98051-24-3

SOMMARIO

Anna Cornagliotti, <i>Ricordo di Max Pfister</i>	IX
Carla Marcato, <i>Due scuole di studi dialettologici in Italia: Torino e Padova</i> ..	1
Alessandro Zuliani, <i>Non solo proverbi: modi di dire, locuzioni, credenze popolari raccontati da Iuliu A. Zanne</i>	19
Alberto Borghini, <i>La fontana con “una faccia di stoffa”. A proposito di una tradizione canavesana</i>	55

LE ISOLE LINGUISTICHE TEDESCOFONE IN ITALIA: SITUAZIONE ATTUALE E PROSPETTIVE FUTURE

Workshop

(Torino, 24 febbraio 2017)

<i>Presentazione</i> , di Livio Gaeta	81
Marco Angster, Marco Bellante, Raffaele Cioffi, Livio Gaeta, <i>I progetti DiWaC e ArchiWals</i>	83
Ermenegildo Bidese, <i>The correlation between unbalanced bilingualism and language decay in small language minorities: the current status of research and future perspectives</i>	95
Pier Paolo Viazzo, <i>Quanti erano — e quanti sono — i Walser a sud delle Alpi? Dal XVI secolo alla legge 482/99</i>	109
Anna Giacalone Ramat, <i>Associazioni walser nella Valle del Lys: storia e prodotti culturali</i>	121
Peter Zürrer, <i>L'inchiesta linguistica di Coquebert de Montbret nel Primo Impero (1806-1812). Progetto di un'edizione dei documenti relativi alle isole linguistiche dell'Italia settentrionale</i>	129
Costanza Cigni, <i>Il Vocabolario comparativo dei dialetti walser in Italia: stato dell'arte e prospettive future</i>	135
Riccardo Regis, <i>Di minoranza in minoranza: ipotesi di tipizzazione lessicale</i>	147
Federica Cusan, Matteo Rivoira, <i>La ricerca toponomastica nelle comunità walser meridionali. Qualche appunto relativo all'ambito piemontese</i>	161

SOMMARIO

NOTIZIE

Luca Bellone, <i>Su un nuovo database del linguaggio giovanile torinese contemporaneo: CheSignificaTorino</i>	177
---	-----

RECENSIONI E SEGNALAZIONI

M. Musso, *L'espressione linguistica dello spazio in un'area plurilingue: il paesaggio toponomastico della media Valle del Lys* (F. Cusan), pag. 201; B. Cinausero Hofer, E. Dentese, *Oronimi del Friuli* (F. Cusan), pag. 203; A. Castiglione, *Le parole del cibo. Lingua e cultura dell'alimentazione a Troina* (A. Ghia), pag. 207; *Le lingue del Salento*, a cura di A. Romano (P. Mariano), pag. 210; T. Urgese, *Il dialetto del XVII secolo di Nniccu Furcedda, opera salentina del francavillese Girolamo Bax* (A. Romano), pag. 216; A. Biondi, *Vocabolario. Il dialetto di San Severino Marche confrontato con altri dialetti marchigiani arcaici e contemporanei* (A. Romano), pag. 218; J. Dorta (ed.), *Estudios comparativo preliminar de la entonación de Canarias, Cuba y Venezuela* (A. Romano), pag. 219; A. De Marco, E. Paone, *Dalle emozioni alla voce. Nuovi orizzonti della comunicazione in italiano L2* (V. De Iacovo, A. Romano), pag. 222; F. Olariu, *Variație și varietăți în limba română. Studii de dialectologie și sociolingvistică* (F. Cugno), pag. 225.

NOTIZIARIO DELL'ATLANTE LINGUISTICO ITALIANO

Lorenzo Massobrio, Matteo Rivoira e Giovanni Ronco, <i>Relazione tecnica sull'attività svolta dalla Redazione dell'ALI nel periodo 1° gennaio - 31 dicembre 2017</i>	231
Lorenzo Massobrio, <i>Bilancio consuntivo dell'Istituto dal 1° gennaio al 31 dicembre 2017</i>	234
<i>Composizione degli Organi dell'Istituto dell'ALI</i>	236

NOTIZIARIO DELL'ATLANTE TOPONOMASTICO
DEL PIEMONTE MONTANO

Federica Cugno e Gabriella Chiapusso, <i>Relazione sull'attività svolta dalla Redazione nel periodo 1° gennaio - 31 dicembre 2017</i>	241
---	-----

SOMMARIO

NOTIZIARIO DELL'ATLANTE LINGUISTICO ED ETNOGRAFICO
DEL PIEMONTE OCCIDENTALE

Sabina Canobbio, Monica Cini, Riccardo Regis e Tullio Telmon, <i>Relazione tecnico-scientifica sullo stato dei lavori dell'ALEPO: situazione al 31.12.2017</i>	247
---	-----

NOTIZIARIO DEL LABORATORIO DI FONETICA SPERIMENTALE
«ARTURO GENRE»

Antonio Romano, <i>Relazione tecnica sull'attività svolta nel periodo 1° genna- io - 31 dicembre 2017</i>	251
---	-----

<i>Elenco dei periodici e delle pubblicazioni ricevuti in scambio</i>	269
<i>Elenco delle pubblicazioni ricevute in omaggio</i>	271
<i>Sommari delle annate precedenti</i>	275

I progetti *DiWaC* e *ArchiWals*

MARCO ANGSTER *, MARCO BELLANTE **,
RAFFAELE CIOFFI **, LIVIO GAETA **

* Università di Zara, ** Università degli Studi di Torino
mangster@unizd.hr, marco.bellante@unito.it,
raffaele.cioffi@unito.it, livio.gaeta@unito.it

Abstract

In the context of a dramatic language decay, the project *DiWaC* and its continuation *ArchiWals* aim at safeguarding the cultural and linguistic heritage of the Walser communities of Gressoney and Issime in the Aosta Valley and of Formazza and Rimella in Piedmont. The projects consist in the construction of a digital archive which will allow the free accessibility online of the linguistic materials acquired in the data-base, conveniently elaborated for storing and archiving the specific text types produced by the communities.

Keywords

Cultural Heritage; Endangered Language; Minority Language; Corpus Linguistics.

1. *Introduzione*

Com'è noto, la presenza delle comunità walser dell'area piemontese e valdostana ha le sue radici nel periodo Basso Medievale: risultato dello spostamento di interi nuclei familiari di coloni provenienti dall'alto Vallese, esse sono divenute con il passare dei secoli isole linguistiche alemanne superiori in area gallo-romanza. Nel corso degli ultimi centocinquanta'anni, tali comunità hanno vissuto, da un lato, l'isolamento linguistico rispetto all'area tedescofona, dall'altro hanno visto progressivamente prendere piede oltre alle varietà dialettali romanze presenti anche l'uso dell'italiano a scapito dell'originario dialetto alemanno (specie nella seconda metà del secolo scorso). Tale situazione ha dunque comportato un inesorabile restringimento dei campi d'uso del dialetto stesso, fenomeno che ha dunque condotto alla progressiva decadenza della lingua e, come nei casi di alcune singole comunità, alla sua scomparsa. Gli ultimi decenni del secolo scorso hanno, però, anche visto la

nascita di numerose iniziative locali e sovra-regionali, mirate al recupero e alla valorizzazione della lingua e della cultura walser: una piccola rinascita che ha avuto come effetto principale la fondazione di associazioni e di Centri Culturali Walser a carattere sovra-regionale e, cosa di non secondaria importanza, ha stimolato la pubblicazione di opere spesso bilingui mirate alla conservazione del patrimonio culturale e linguistico delle popolazioni walser.

Un patrimonio linguistico e testuale che, vista la condizione in cui versano i dialetti walser dell'area italiana, costituisce materiale di primaria importanza dal punto di vista non solo linguistico, ma anche culturale ed etnografico. Proprio dalla volontà di conservare, sistematizzare e studiare tale patrimonio nascono i due progetti *Di.Wa.C.* e *ArchiWals*¹.

2. *Uno strumento per la conservazione dell'eredità culturale walser*

Eredità linguistica e culturale di varietà dialettali a forte rischio di estinzione, l'insieme di testimonianze scritte disponibile per le comunità walser si rivela certamente di notevole interesse per i diversi ambiti della ricerca linguistica, storica e antropologica.

Un primo elemento di interesse è rappresentato dall'accresciuto numero di testimonianze scritte prodotte negli ultimi decenni in ciascuna delle varietà prese in esame dai due progetti (rispettivamente, le varietà di Gressoney e di Issime per l'area valdostana, e quelle di Formazza e Rimella per il Piemonte). Esse costituiscono una documentazione la cui complessità e ricchezza linguistica potrà essere meglio studiata una volta operata la loro sistematizzazione in una banca dati specificatamente progettata per la loro gestione e per lo studio di tali varietà dialettali.

Non è di secondaria importanza, a questo proposito, il consolidamento del prestigio raggiunto da tali varietà nel corso della seconda metà del Novecento. Proprio per accrescere ulteriormente tale prestigio, le singole banche dati e i *corpora* testuali a esse legati, saranno di libera consultazione: fattore quest'ultimo che permetterà l'accesso da parte dei parlanti a una risorsa linguistica digitalizzata che per la prima volta raccoglie la parte più consistente della produzione letteraria e lessicografica delle singole varietà walser, e che potrebbe idealmente fungere da stimolo per un'ulteriore produzione scritta.

Una potenzialità decisiva legata alla creazione di questo tipo di raccolta — cioè un *corpus* testuale — è poi rappresentata dalla possibilità di intraprendere una descrizione sistematica delle singole varietà in oggetto attraverso un approccio “*corpus-based*”, possibilità non comune nel caso dello studio di

¹ Il portale dei progetti è accessibile alla URL: www.diwac.it.

lingue di minor diffusione, le cosiddette “*lesser-used languages*”, o in decadenza (“*decaying languages*”) ².

Le ulteriori future indagini sul campo che verranno condotte sulle quattro varietà prese in considerazione non potranno infine che trarre indubbio beneficio dall’accesso a tali *corpora* linguistici e testuali. Essi potranno rappresentare non solo un efficace esempio metodologico di trattamento di dati di lingue di minor diffusione, ma anche uno strumento di lavoro per futuri approfondimenti degli studi su singoli fenomeni linguistici o morfosintattici ³.

3. Alcune questioni riguardanti il trattamento di dati di lingue di minor diffusione

La linguistica dei *corpora* ha destato, specie negli ultimi anni, un interesse via via maggiore dal punto di vista teorico: i *corpora* costituiscono infatti una sempre più diffusa, durevole e flessibile risorsa dalla quale è possibile trarre dati utili alla ricerca linguistica. Tale campo di studio e di ricerca ha potuto avvantaggiarsi nel recente passato di due importanti innovazioni legate alla diffusione di Internet e del *World Wide Web*, lo sviluppo cioè di ampi *corpora* basati sul *Web* e l’utilizzo dei *social media* come fonte di dati linguistici. L’accresciuta attenzione nei confronti degli approcci “*corpus-*” o “*Web-based*” ha, fino ad ora, forse comprensibilmente, toccato in maniera solo marginale lo studio e l’analisi delle lingue di minor diffusione o a rischio di scomparsa. In tal senso, del tutto paradigmatica ed esemplificativa è l’esperienza dell’archivio del progetto *DoBeS* (*Dokumentierung von bedrohten Sprachen*) ⁴, strumento che permette all’utente di accedere ai materiali di varie indagini sul campo condotte su varietà pressoché prive fino a quel momento di qualsiasi descrizione. Tuttavia, alcune lingue, seppur minacciate, non mancano di documentazione e in alcuni casi sono persino provviste di un’abbondante produzione letteraria. Un patrimonio letterario e linguistico che, benché in molti casi limitato per estensione e varietà, potrebbe di contro essere valorizzato, indagato e descritto nei suoi diversi aspetti attraverso un approccio classico basato su *corpora*.

Al di là di tali prospettive di studio, certamente stimolanti, l’applicazione della metodologia legata alla costruzione di *corpora* concernenti lingue di minor diffusione comporta però una serie di problemi dei quali occorre tenere conto al momento della loro progettazione. Un primo problema è legato al sistema di scrittura delle varietà la cui ortografia spesso differisce in maniera

² A tal proposito, si vedano fra gli altri: Giacalone Ramat (1979) e Dal Negro (2003).

³ Per qualche risultato iniziale, si vedano Angster, Gaeta (2018), Gaeta (in stampa, in preparazione), Gaeta, Angster (in preparazione).

⁴ <<http://dobes.mpi.nl/>>.

molto evidente da quella delle eventuali varietà affini o delle relative varietà standard di riferimento, e ciò indipendentemente dalla distanza fonologica che le separa. In aggiunta a ciò, non è trascurabile la mancanza, per molte lingue di minor diffusione, di una norma ortografica stabile e condivisa, mancanza da cui consegue un'evidente variabilità interna non solo in testi del medesimo periodo o appartenenti al medesimo genere letterario, ma spesso anche in testimonianze scritte riconducibili allo stesso autore (o addirittura, in casi estremi, all'interno di uno stesso testo). L'assenza di standardizzazione ortografica, inoltre, non interessa solamente il modo in cui i fonemi (e conseguentemente i lessemi) vengono resi, ma anche il livello morfologico e morfosintattico della lingua (ad es. il trattamento dei clitici o, più generale, delle sequenze di grafemi morfologicamente segmentabili e dei nessi consonantici).

Un'ulteriore questione legata all'uso dei *corpora* nell'ambito dello studio di lingue di minor diffusione riguarda poi l'esistenza di strumenti di *POS-tagging* (annotazione delle parti del discorso), procedimento che si rivela spesso del tutto impossibile proprio a causa del fatto che gli strumenti di annotazione esistenti sono creati per essere utilizzati con varietà standardizzate. Tali strumenti non sono in grado di fornire risultati soddisfacenti se applicati ai dialetti in quanto questi ultimi non solo sono molto distanti dalle rispettive lingue di riferimento, ma sono anche caratterizzati, come già detto, da una grafia priva di standardizzazione.

Non marginale all'atto della progettazione di un *corpus* è infine il problema legato a due degli aspetti metodologici fondanti per la realizzazione di *corpora* linguistici: il bilanciamento e la rappresentatività. Tali aspetti appaiono di difficile applicazione nell'ambito delle lingue di minor diffusione, che si presentano spesso come codici legati all'uso orale e la cui produzione letteraria è spesso limitata a quella di singoli autori, concentrata in alcuni decenni e spesso riconducibile a pochi generi letterari. Tuttavia, proprio in ragione della condizione di pericolo in cui versano le varietà considerate, appare assolutamente necessario operare una strategia inclusiva al momento della progettazione del *corpus* e della raccolta dei testi, in modo tale da ottenere il massimo grado di conservazione e tutela di tale patrimonio testuale ed etnografico. I problemi del bilanciamento e della rappresentatività, necessariamente lasciati da parte in fase di raccolta dei dati, saranno poi presi in considerazione successivamente nel momento dell'analisi e dell'interpretazione dei dati raccolti.

4. *Cenni sulla produzione linguistica e letteraria walser delle quattro varietà di Issime, Gressoney, Formazza e Rimella*

I progetti *DiWaC* e *ArchiWals*, come già detto, si propongono di raccogliere, catalogare, digitalizzare e, infine, analizzare dal punto di vista linguistico, lessicale e morfosintattico quanto fino a noi giunto del patrimonio tradizionale (scritto e parlato) di quattro varietà walser dell'area subalpina, quelle di Gressoney e Issime per la Valle d'Aosta, e di Rimella e Formazza per il Piemonte. Esse costituiscono casi esemplari delle vicende storiche e linguistiche delle varietà walser presenti in Italia, così come del loro differente grado di contatto con le varietà di area gallo-romanza e alemanna. Poste rispettivamente alle pendici del massiccio del Monte Rosa e in prossimità del Griespass, le comunità di Gressoney (Greschèney) e di Formazza (Pomatt) per lungo tempo si sono trovate a essere più direttamente in contatto con l'area tedescofona, una continuità di rapporti spentasi, per quanto concerne Gressoney, nell'ultimo secolo, ma che in parte risulta ancora percepibile per quanto concerne l'area di Formazza. Separata da quella di Gressoney dalla presenza dell'abitato di Gaby, la comunità walser di Issime appare, al contrario, un esempio di varietà del tutto isolata, circondata da ogni lato da varietà gallo-romanze o francofone e per tale ragione da esse direttamente influenzata⁵. Così come Issime (Éische)me), infine, la piccola comunità di Rimella (Rémmalju) presenta una forte connessione con la circostante area romanza, e la varietà ivi parlata ha subito fortemente l'influenza delle varietà romanze. Le quattro comunità, in ragione di una serie di fattori geografici e storici, risultano dunque esempi tangibili della condizione di progressivo stato di attrito e slittamento linguistico che caratterizza una lingua di minor diffusione, fenomeni che bene possono risultare percepibili nella produzione scritta degli ultimi decenni.

La recente produzione scritta appare di primario interesse linguistico in virtù delle caratteristiche linguistiche, lessicali e morfosintattiche di queste varietà e delle soluzioni grammaticali da esse adottate nel tempo, spesso in maniera del tutto indipendente l'una dall'altra. La produzione testuale delle comunità walser appare in questo senso relativamente ampia, anche se del tutto disomogenea per genere e periodo di produzione: ai testi arrivati a noi come testimonianze delle ricerche svoltesi nel corso del XIX secolo, e al ricco materiale desumibile dalla letteratura atlantistica dei primi decenni del Novecento, si è infatti andata ad affiancare la letteratura del periodo successivo alla seconda metà del secolo scorso, risultato della moderna crescita letteraria

⁵ Per quanto concerne un quadro d'insieme dei casi di Issime e Gressoney, si vedano: Zürer (1999) e Zürer (2009), mentre su Formazza e Rimella si vedano rispettivamente Dal Negro (2003) e Bauen (1978).

delle varietà walser. Accanto ai dizionari delle singole varietà⁶, frutto delle iniziative delle Associazioni Walser, e ad alcune grammatiche⁷, la recente produzione letteraria è costituita *in primis* da pubblicazioni a carattere locale curate dalle diverse comunità — ad es.: bollettini parrocchiali, riviste dalla cadenza periodica, brevi pubblicazioni (legate a occasioni specifiche), da volumi bilingui di cultura tradizionale o di letteratura per bambini, da testi di carattere religioso o, ancora, da raccolte di poesie.

Analizzata dal punto di vista diacronico, la produzione a stampa nei vari dialetti Walser mostra una ricchezza e una vitalità molto discontinua nei vari decenni, del tutto dipendente sia dallo status dialettale che li caratterizza, sia dalle dinamiche sociali che nel tempo hanno comportato la progressiva riduzione del loro uso nelle comunità di riferimento. La presenza di una letteratura che nella sostanza raccoglie testi appartenenti a generi letterari molto distanti fra di loro per caratteristiche sintattiche e stilistiche (ad es.: poesia e ricette di cucina, o ancora letteratura tradizionale e letteratura per bambini) appare un aspetto fondamentale per la valutazione delle caratteristiche e dell'evoluzione di queste varietà. L'insieme della produzione testuale di ciascuna delle varietà prese in considerazione costituisce una testimonianza diretta non solo di alcuni fenomeni di cambiamento linguistico tipici delle lingue di minoranza, ma anche della loro evidente variabilità sintattica e ortografica interna, così come del recente tentativo di pervenire a una loro standardizzazione più o meno condivisa dal punto di vista ortografico.

Proprio al fine di mettere in atto un recupero per quanto possibile completo ed esteso dei materiali a disposizione, si è deciso di procedere alla raccolta esaustiva della produzione scritta disponibile e, quando possibile, delle trascrizioni dei materiali orali disponibili in registrazione: tale scelta ha comportato, come già menzionato in precedenza, il mancato rispetto del criterio di rappresentatività in favore di una più estesa e puntuale raccolta (e conservazione) di *corpora* testuali altrimenti a rischio di dispersione, o, nel caso di documenti audio già conservati su supporto digitale, di obsolescenza. Sulla base di tale volontà di conservazione e di analisi, si è proceduto alla raccolta sul campo di tali materiali a stampa che, acquisiti in digitale e successivamente riordinati per tipologia di opera e località, sono stati successivamente catalogati secondo un numero identificativo univoco, di una metadattazione che tenesse conto di criteri cronologici e di genere testuale — elementi fon-

⁶ Per quanto concerne le varietà di Gressoney e di Issime, si fa riferimento in questa sede a: WKZ-GT1 = Walser Kulturzentrum (1988); WKZ-GT2 = Walser Kulturzentrum (1998); WKZ-IT1 = Walser Kulturzentrum (1988); WKZ-IT2 = Walser Kulturzentrum (1998). Per quanto riguarda la comunità di Rimella, si rimanda a: Scilligo (1993).

⁷ A tal proposito, si veda: Bauen (1978).

damentali, questi ultimi, per il riordino e la sistematizzazione del *corpus*: tali criteri di catalogazione costituiscono fattori di non secondaria importanza non solo per la realizzazione, il riordino e la gestione del *corpus*, ma anche per l'analisi linguistica e morfosintattica dei testi in esso contenuti, da un punto di vista sincronico e diacronico, cioè secondo una prospettiva di evoluzione nel tempo delle singole varietà, oltre che per ovviare ai problemi di bilanciamento e rappresentatività sopra menzionati.

Ciascuno dei documenti acquisiti nelle varie campagne di raccolta condotte presso le biblioteche e gli archivi dei Centri Culturali Walser è stato riversato in formato digitale attraverso un applicativo di riconoscimento caratteri (OCR) e successivamente caricato sulla piattaforma che ospita i diversi *corpora*. Preliminarmente al caricamento sulla piattaforma, ciascun file di testo generato dall'OCR è stato sottoposto a un controllo di conformità con l'originale, finalizzato alla correzione degli errori di interpretazione generati dal riconoscimento caratteri: data l'assenza di un sistema di riconoscimento caratteri compatibile in maniera estensiva con la grammatica e l'ortografia dei dialetti walser, per necessità si è scelto di utilizzare un OCR addestrato al riconoscimento della grafia del tedesco standard.

Tale operazione di correzione e successivo riversamento ha comportato alcune scelte riguardanti il grado di intervento e di correzione da apportare ai testi e alla documentazione a nostra disposizione, in particolare nel caso dell'insieme di documenti prodotti nel periodo precedente agli anni Novanta del secolo scorso: tali testi appaiono infatti caratterizzati da una significativa assenza di uniformità grafica e ortografica, fattore quest'ultimo introdotto progressivamente (e non del tutto univocamente) solamente a partire dagli ultimi decenni del Novecento. Tale difformità grafica appare piuttosto evidente e diffusa all'interno della produzione testuale nelle singole varietà walser, non esclusivamente nella produzione anteriore al recente tentativo di normalizzazione grafica. L'alternanza di differenti grafie si rileva in modo particolare nell'ortografia di vocali, consonanti e nessi consonantici o, più in generale, nella resa dei prestiti dall'italiano o dalle varietà romanze (si veda la coesistenza delle forme *santimeter* e *sentimeter* per l'italiano *centimetro*). Per quanto riguarda la resa delle vocali, le oscillazioni possono essere esemplificate prendendo in analisi il caso di Gressoney dalla lenta introduzione, soprattutto in seguito alla pubblicazione dei vocabolari a fine anni '80, di una chiara distinzione ortografica fra la vocale anteriore medio-alta ([e] <e>) e quella anteriore semi-chiusa ([ɪ] <é>) e le corrispondenti vocali posteriori arrotondate ([o] <o> vs. [ʊ] <ò>). Dal punto di vista prettamente ortografico, poi, risulta evidente come alcune consonanti nel tempo (ma anche in una stessa opera) vengano scritte seguendo norme diverse, così come ben dimostrato dall'oscillazione nella resa grafica dell'occlusiva velare sorda [k], realizzato come <c>

secondo la tradizione italiana, <k> secondo quella tedesca o ancora come <gg> seguendo la grafia proposta da Eugen Dieth per lo svizzero tedesco⁸.

Questa mancanza di uniformità grafica è stata per scelta conservata nella realizzazione del *corpus* in quanto evidente fattore di variabilità intrinseca alla lingua, e punto nodale per lo studio delle dinamiche portanti di tali varietà di uso minoritario: a ciascuna delle varianti grafiche riscontrate — specie nei testi più datati — è stato infatti dedicato uno specifico spazio all'interno dei rispettivi lemmi registrati nel database e in prima istanza desunti dai vocabolari delle rispettive varietà, in modo che la difformità ortografica non infici la possibilità di operare ricerche efficaci nel *corpus* e che la lemmatizzazione del *corpus* non necessiti al tempo stesso una normalizzazione ortografica dei testi.

5. Un database multistrato

La necessità di trattare e sistematizzare questi *corpora* testuali e la mole di dati linguistici e morfologici in essi contenuti è stata risolta con la creazione di una piattaforma consistente in un database multistrato, specificatamente progettato per trattare efficacemente gli aspetti linguistici interni ed esterni delle lingue considerate. L'idea di partenza è stata quella di applicare la logica di un database relazionale su strutture minime basate su lemmi: un funzionamento di questo tipo presenta infatti il vantaggio pratico di minimizzare la capacità di calcolo richiesta per le differenti tipologie di ricerca solitamente necessarie con dizionari e *corpora*, proprio in ragione della specifica architettura usata. Il motore di ricerca del database da noi progettato, infatti, processa solamente i dati per i quali viene interrogato, costruendo di volta in volta i necessari indici, ed evitando così l'uso dell'ingente quantità di memoria necessaria per la creazione di indici ottenuti attraverso una compilazione esaustiva (per esempio come accade per quanto concerne linguaggi di marcatura come XML). La piattaforma è stata creata in un ambiente di sviluppo *ScriptCase* per *PHP*, *JavaScript* e database *MySQL*, in modo da essere accessibile da qualunque tipo di macchina.

Per rendere più immediato e lineare il funzionamento del database, per ciascun lemma è prevista la creazione di un numero identificativo (ID) univoco, al quale tutte le informazioni legate alla singola occorrenza nel *corpus* (*token*) sono connesse. Nella progettazione del database e del dizionario è stata creata una maschera specifica dedicata alla descrizione di ciascun lemma: ideata in modo da essere piuttosto flessibile, essa consente di volta in volta di compilare per ciascun lemma diversi campi e valori (ad es. lemmi derivati,

⁸ Si veda in particolare Dieth (1938).

composti o affini, informazioni etimologiche, ecc.), secondo le diverse necessità e la granularità dell'informazione linguistica di cui ciascun lemma sarà corredato. All'interno della piattaforma è anche prevista la presenza di database specifici per ciascuno dei *corpora* delle lingue analizzate e per le istanze di parole fonologiche che in essi sono riscontrabili, entrambi gestiti in maniera autonoma dal dizionario.

La navigazione e l'indagine su dizionario e su *corpora* è affidata al sistema di ricerca interno della piattaforma. Per ciascun ID univoco di un lemma la ricerca può essere filtrata attraverso diverse discriminanti quali traduzioni, varianti ortografiche o forme che appaiono più o meno frequentemente (ad es., la forma del participio preterito del verbo *goa* 'andare', attestata nel *corpus* con la forma *kanget* così come nella variante *canget* o *cannet* o *gganged*, e così via), parti di discorso, lemmi collegati, fonte testuale e posizione delle occorrenze nel *corpus*. Le ricerche all'interno del *corpus* possono essere semplici — la ricerca di una singola parola nel dizionario — o complesse — la ricerca di una singola parola all'interno dell'intero *corpus*, in modo da verificare il numero di occorrenze all'interno dei singoli testi della forma base, dei suoi allomorfi e delle sue varianti. Questo permette diversi gradi di analisi del *corpus*, o un'approfondita comparazione fra le attestazioni di un lemma all'interno del patrimonio letterario e le accezioni presenti nel dizionario.

I *corpora*, le istanze di parole fonologiche e i dizionari interagiscono nel processo di acquisizione di nuovi lemmi provenienti dai testi e non presenti nei dizionari a stampa delle diverse varietà. Le attestazioni nei dizionari fungono da riferimento per la definizione della forma di citazione di ciascun lemma, mentre nuovi inserimenti desunti dai testi verranno di volta in volta dotati di una forma di citazione eventualmente adattata alle convenzioni grafiche più diffuse. La piattaforma è progettata in modo da facilitare l'espansione del numero delle entrate a mano a mano che nuovi testi e testimonianze verranno acquisite, auspicabilmente ampliando il ventaglio di generi testuali anche con le trascrizioni di interviste e di altre interazioni orali.

6. Conclusioni

Nel momento in cui scriviamo, il lavoro di raccolta e digitalizzazione risulta avviato solo per due delle quattro varietà prese in considerazione (Gresoney e Issime), ma le operazioni sui materiali linguistici e letterari fino ad ora svolte ci hanno consentito di riflettere sulla struttura del database, così come sui diversi passaggi necessari alla realizzazione dei *corpora*. La raccolta dei testi e dei dati lessicali ci ha dato materiale utile a sviluppare e testare la versione "beta" sia della banca dati, sia del dizionario, e ci ha inoltre consen-

tito di presentarne aspetto e funzionamento ad alcuni membri delle quattro comunità: la nostra speranza è quella di coinvolgere nella sperimentazione della piattaforma un numero sempre più ampio dei possibili utenti futuri in modo da avviare un meccanismo di confronto volto al miglioramento della sua utilizzabilità.

Appena il database sarà reso liberamente accessibile a parlanti e studiosi, la nostra speranza sarà quella di veder aumentare l'interesse nei confronti dei dialetti walser a diversi livelli, e di stimolare così la produzione di nuovo materiale scritto che possa arricchire ulteriormente la banca dati e i dizionari.

Il patrimonio testuale, espanso e ampliato in questa maniera, sarà di conseguenza sempre più utile a tutti quelli che vorranno avvicinarsi a queste varietà, e, costituendosi quale esempio del loro possibile utilizzo, potrebbe divenire una risorsa fondamentale nella descrizione delle quattro varietà trattate — magari stimolando iniziative simili presso altre comunità (walser e non) — e un riferimento per chi volesse recuperare l'uso di queste varietà, sia nello scritto, sia nell'oralità. I progetti *DiWaC* e *ArchiWals*, che qui abbiamo presentato, mirano in conclusione a contribuire alla conservazione di un patrimonio linguistico e culturale in pericolo e alla sua maggiore esposizione e fruibilità da parte, certamente, degli specialisti di dialettologia e linguistica tedesca oltre che delle lingue di minor diffusione e in pericolo, ma soprattutto da parte delle comunità stesse che in essi potranno trovare un monumento duraturo al loro patrimonio culturale che può fungere da volano per il prestigio presente e futuro di queste varietà.

BIBLIOGRAFIA

- ANGSTER M., GAETA L. (2018), *Wie kurz sind die Kurzverben? Morphologische Merkmale in Gressoney und Issime*, in RABANUS S., *Deutsch als Minderheitensprache in Italien – Theorie und Empirie kontaktinduzierten Sprachwandels*, Germanistische Linguistik 239-240, pp. 211-237, Hildesheim, Olms.
- BAUEN M. (1978), *Sprachgemischter Mundartausdruck in Rimella (Valsesia, Piemont). Zur Syntax eines südwalserischen Dialekts im Spannungsfeld der italienischen Landes- und Kultursprache*, Bern, Haupt.
- CSW-RI1 = Centro Studi Walser – Rimella (1995), *Ts Remmaljertittschu. Vocabolario Italiano – Tittschu*, Borgosesia, Tipografia di Borgosesia.
- CSW-RI1 = Centro Studi Walser – Rimella (2005), *Ts Remmaljertittschu. Vocabolario Tittschu – Italiano*, Borgosesia, Tipografia di Borgosesia.
- DAL NEGRO S. (2003), *The Decay of a Language. The Case of a German Dialect in the Italian Alps*, Bern, Peter Lang.
- DIETH E. [1938] (1986), *Schnyzerdütschi Dialäktschrift. Leitfaden einer einheitlichen Schreibweise für alle Dialekte*, 2^a edizione rivista da Christian Schmid-Cadalbert, Aarau, Sauerländer.
- GAETA L. (in stampa), *Im Passiv sprechen in den Alpen*, Sprachwissenschaft, Heidelberg, Winter.
- GAETA L. (in preparazione), *Remotivating inflectional classes: an unexpected effect of grammaticalization*, in DRINKA B., *Proceedings of 23rd International Conference on Historical Linguistics, University of Texas at San Antonio, 30.7.-4.8.2017*, Amsterdam / Philadelphia, John Benjamins.
- GAETA L., ANGSTER M. (in preparazione), *Loan word-formation in minority languages: lexical strata in Titsch and Töitschu*, in TEN HACKEN P. e PANOCOVÁ R., *Interaction of borrowing and word-formation*, Edinburgh, Edinburgh University Press.
- GIACALONE RAMAT A. (1979), *Lingua, dialetto e comportamento linguistico. La situazione di Gressoney*, Aosta, Musumeci.
- SCILLIGO P. (1993), *Pumattertietsch Werterbeuch. Vocabolario Formazzino-Tedesco-Italiano e Italiano-Formazzino-Tedesco / Pumattertietsch-Waeltsch-Tietsch und Waeltsch-Pumattertietsch-Tietsch*, IFREP, Roma.
- WKZ-GT1 = Walser Kulturzentrum (1988), *Greschòneytitsch. Vocabolario Italiano – Titsch*, Quart (Aosta).
- WKZ-GT2 = Walser Kulturzentrum (1998), *Greschòneytitsch. Vocabolario Titsch – Italiano*, Quart (Aosta).

WKZ-IT1 = Walser Kulturzentrum (1988), *D'Éischemetöitschu. Vocabolario Italiano – Töitschu*, Quart (Aosta).

WKZ-IT2 = Walser Kulturzentrum (1998), *D'Éischemetöitschu. Vocabolario Töitschu – Italiano*, Quart (Aosta).

ZÜRRER P. (1999), *Sprachinseldialekte. Walserdeutsch im Aosta-Tal (Italien)*, Aarau / Frankfurt a.M. / Salzburg, Sauerländer.

ZÜRRER P. (2009) *Sprachkontakt in Walser Dialekten. Gressoney und Issime im Aostatal (Italien)*, Stuttgart, Steiner.

Risorse web

DoBeS (*Documentation of Endangered Languages*): www.dobes.mpi.nl

Di.Wa.C.: www.diwac.it