

**ARCHIVIAZIONE E ANNOTAZIONE DEL PARLATO
NELL'ATLANTE LINGUISTICO DELLA SICILIA**

Vito Matranga
Università di Palermo
vito.matranga@unipa.it

1. Il pluridecennale interesse dell'*Atlante Linguistico della Sicilia* (ALS) nei riguardi dei dialetti dell'Isola e, più generalmente, nei riguardi di tutti gli aspetti relativi al suo repertorio linguistico, ha consentito la raccolta di ormai diverse migliaia di documenti orali in grado di testimoniare l'esperienza linguistica e buona parte di ciò che rimane della cultura tradizionale nella Sicilia contemporanea.

Ai documenti sonori prodotti in questi ultimi 20 anni, in funzione dei diversi ambiti di ricerca dell'*Atlante Linguistico della Sicilia*, si aggiungono quelli di più antica data, e pertanto di un certo interesse ormai storico, che riguardano le inchieste effettuate tra la fine degli anni '60 e l'inizio degli anni '70 da Giuseppe Piccillo e da Vincenzo Orioles per la *Carta dei Dialetti Italiani* (CDI); un altro piccolo fondo relativo alle inchieste fonetiche e lessicali effettuate da Giovanni Ruffino nel 1973; le numerose registrazioni che, nella seconda metà anni '80, erano state effettuate per l'*Osservatorio Linguistico Siciliano* (OLS). Questo patrimonio costituisce l'*Archivio delle Parlate Siciliane* (APS)¹, il quale possiede, ad oggi, circa 5000 ore di registrazione relative a oltre 200 località siciliane². L'80% di questi documenti – giunti su bobina, su nastro analogico e poi su DAT – è oggi ordinato in una Banca Dati³, che ne consente un'agevole consultazione, grazie alla sistematica digitalizzazione/acquisizione in file sonori gestibili con i più comuni mezzi informatici.

Una prima struttura informatica per l'archiviazione di questi documenti era stata costruita nei primi anni '90 – man mano che i nastri venivano raccolti nella sede del Centro di studi filologici e linguistici siciliani – utilizzando le possibilità allora concesse dalle prime versioni del software FileMaker. Si è pensato, pertanto, di mantenere questo software di base, e il nucleo originario dell'Archivio, per l'impianto della nuova Banca Dati, che ne ha esteso e migliorato le funzioni per l'immissione e la ricerca dei documenti, reimpostandone la struttura originaria.

¹ Cfr. Matranga (2011°).

² Per il patrimonio documentario archiviato nella Banca Dati dell'APS fino alle metà del 2011, cfr. Galati & Luppina (2001).

³ Per informazioni più dettagliate sull'organizzazione e il funzionamento della Banca Dati dell'APS, cfr. Matranga (2011b).

Archivio delle Parlate Siciliane

n° archivio **2769**
data di accesso 29/03/2011
immissore Galati Silvia

località <input type="text" value="Casteltermini"/> (AG) <input type="text" value="329"/> <small>provincia codice della località</small>	quartiere <input type="text"/> mese <input type="text" value="07"/> anno <input type="text" value="2007"/>	durata del file sonoro (in minuti) <input type="text" value="90"/>
corpus <input type="text" value="Tesi di laurea"/> argomento generale <input type="text" value="Cultura alimentare"/>	argomento specifico <input type="text" value="concetti di base (alimentare)"/> <input type="text" value="panificazione"/> <input type="text" value="pasta"/>	<input type="button" value="SE materiale edito"/> <input type="button" value="ALM"/> <small>colore corpus e sigla argomento</small>
nome del raccoglitore <input type="text" value="Consiglio Serafina"/>	codice <input type="text" value="125"/> <small>codice del file sonoro</small> <input type="text" value="329ALM125A2769"/>	

INFORMATORI
DATI TECNICI
CORREDO INTERVISTA
NOTE GENERALI
COLLEGA FILE SONORO
STATO DELLA SCHEDATURA
STAMPA ETICHETTE

SCHEDA BASE

Figura. 1: scheda per di immissione dei dati di base dell'APS.

2. Il caricamento dei dati è strutturato su una serie di schede di interfaccia che raccolgono le informazioni (ordinate in campi) sullo stesso documento sonoro.

La prima scheda (v. fig.1) esige alcune informazioni indispensabili relative ad ogni singolo documento e consente di raggiungere, attraverso una serie di pulsanti, tutte le altre schede a esso correlate.

Le informazioni basilari riguardano: la località nella quale è stata prodotta la registrazione; la data in cui è avvenuta la registrazione; la durata della singola registrazione; il *corpus*, ossia il progetto o l'ambito di ricerca per il quale è stata effettuata la registrazione.

Ciascun *corpus* costituisce, in sostanza, un fondo documentario, omogeneo relativamente alla tipologia escussiva rispondente allo specifico obiettivo della ricerca: dalle inchieste sociovariazionali a quelle etnodialettali; dalla Carta dei Dialetti Italiani alle Etnostorie e i racconti di vita e altri fondi elencati in fig. 2.

Di ogni documento sonoro è sistematicamente indicato, in relazione allo specifico *corpus*, l'argomento trattato, che è definito primariamente in categorie generiche e sovraordinate (nel campo «argomento generale») e, contemporaneamente, in quelle più specifiche e subordinate (nel campo «argomento specifico») secondo le relazioni indicate in fig. 2.

Dalla scheda di base si accede ad altri moduli correlati, che pretendono informazioni più specifiche relative agli «informativi», agli eventuali «materiali di corredo», ai «dati tecnici» delle registrazioni etc.

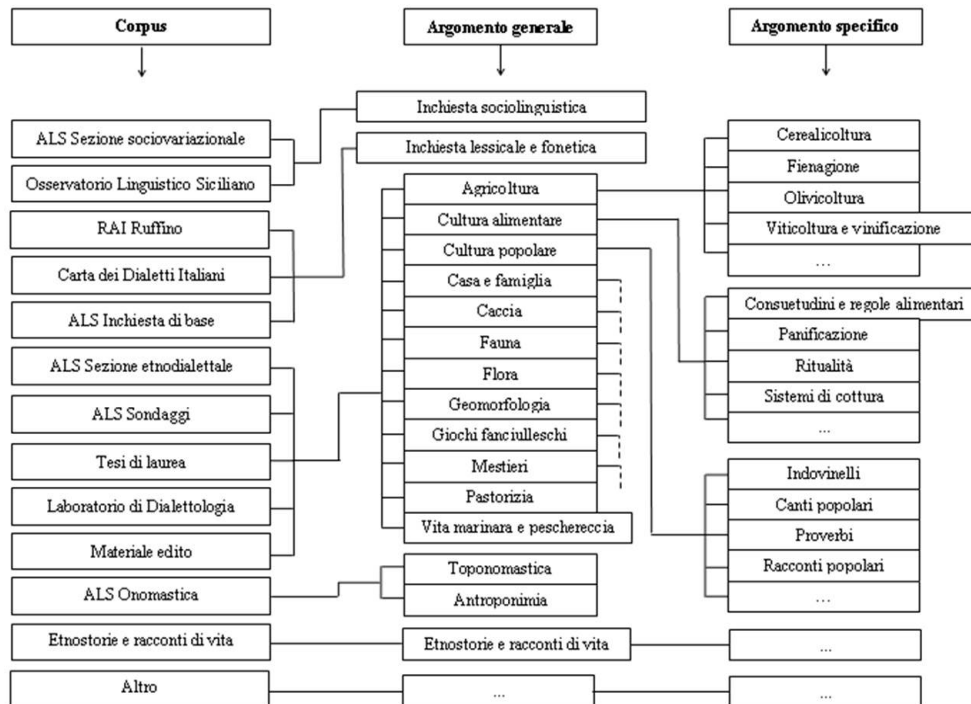


Figura 2: struttura dei corpora dell'APS.

The screenshot shows the search interface for the APS, titled "Archivio delle Parlate Siciliane". It features two main search sections:

- Cerca**: A search form with fields for "località", "corpus", "argomento generale", "argomento specifico", and "anno della registrazione". A "Cerca informatori" button is located below these fields.
- Cerca Informatori**: A search form with fields for "NOME", "SESSO", "ETÀ" (with a range selector "compreso tra ... e ..."), "CATEGORIA PROFESSIONALE", "PROFESSIONE", and "ISTRUZIONE". A "MOSTRA RISULTATI" button is located below these fields.

A large "MOSTRA RISULTATI" button is also present at the bottom of the interface.

Figura 3: scheda per la ricerca dei documenti sonori dell'APS.

I file sonori – memorizzati su *server di storage* – sono collegati alle corrispondenti unità d’archivio, affinché possano essere prontamente disponibili all’ascolto una volta effettuata la ricerca attraverso una maschera di interfaccia semplificata (v. fig. 3), che prevede alcune fondamentali chiavi di ricerca: la località, il corpus, l’argomento, o le caratteristiche del testimone.

I documenti trovati vengono restituiti simultaneamente alle altre informazioni basilari, indipendentemente dalla chiave di ricerca, e accanto a ciascuno di essi è posto (come si vede in fig. 4), un pulsante attraverso il quale è possibile l’ascolto.

Località	Prov.	Corpus	Argomento	Informatori	Anno	Minuti	Raccoglitore	n° archivio	Ascolta
Alcamo	(TP)	ALS Sezione Etnodialettale	Cultura alimentare	2	2006	15	Bono Ludana	2294	
			pasta salse, condimenti, intingoli lunade, rane, grandi						
Alcamo	(TP)	ALS Sezione Etnodialettale	Cultura alimentare	1	2006	90	Bono Ludana	2295	
			carne						
Alcamo	(TP)	ALS Sezione Etnodialettale	Cultura alimentare	1	2006	16	Bono Ludana	2296	
			carne						
Alcamo	(TP)	ALS Sezione Etnodialettale	Cultura alimentare	1	2006	15	Bono Ludana	2297	
			preliminari dollo ritualità (alimentare)						
Alcamo	(TP)	ALS Sezione Etnodialettale	Cultura alimentare	1	2006	90	Bono Ludana	2298	
			consuetudini e regole alimentari sistema di cottura conservazione degli alimenti caffè e infusi						
Alcamo	(TP)	ALS Sezione Etnodialettale	Cultura alimentare	1	2006	90	Bono Ludana	2299	
			pasta panificazione						
Alcamo	(TP)	ALS Sezione Etnodialettale	Cultura alimentare	1	2006	90	Bono Ludana	2300	
			preliminari legumi pesce, molluschi, crostacei uova e frittate frittate						
Alcamo	(TP)	ALS Sezione Etnodialettale	Cultura alimentare	1	2006	90	Bono Ludana	2301	
			arnesi e vasellame vino, liquori e bevande						
Alcamo	(TP)	ALS Sezione Etnodialettale	Cultura alimentare	1	2006	90	Bono Ludana	2302	
			ortaggi e verdure funghi carni						

esegui un'altra ricerca

Figura 4: esempio di tabulato di risposta a una ricerca dei documenti dell’APS.

L’*Archivio delle Parlate Siciliane* costituisce, dunque, la “banca” del progetto geo-socio-linguistico dell’ALS: una “banca” intesa propriamente come contenitore ed erogatore delle risorse empiriche necessarie alle successive fasi della ricerca.

Sia pur costruito, sul piano informatico, in modo tale da funzionare autonomamente, questa Banca delle parlate costituisce soltanto il primo basilare modulo di un complesso sistema informativo, del quale riporto lo schema (v. fig. 5), ma sul quale non indugero se non sulla parte che riguarda più propriamente il trattamento dei file sonori per l’analisi delle informazioni desunte dal parlato, e precisamente sulla fase della trascrizione e dell’annotazione/markup delle trascrizioni stesse.

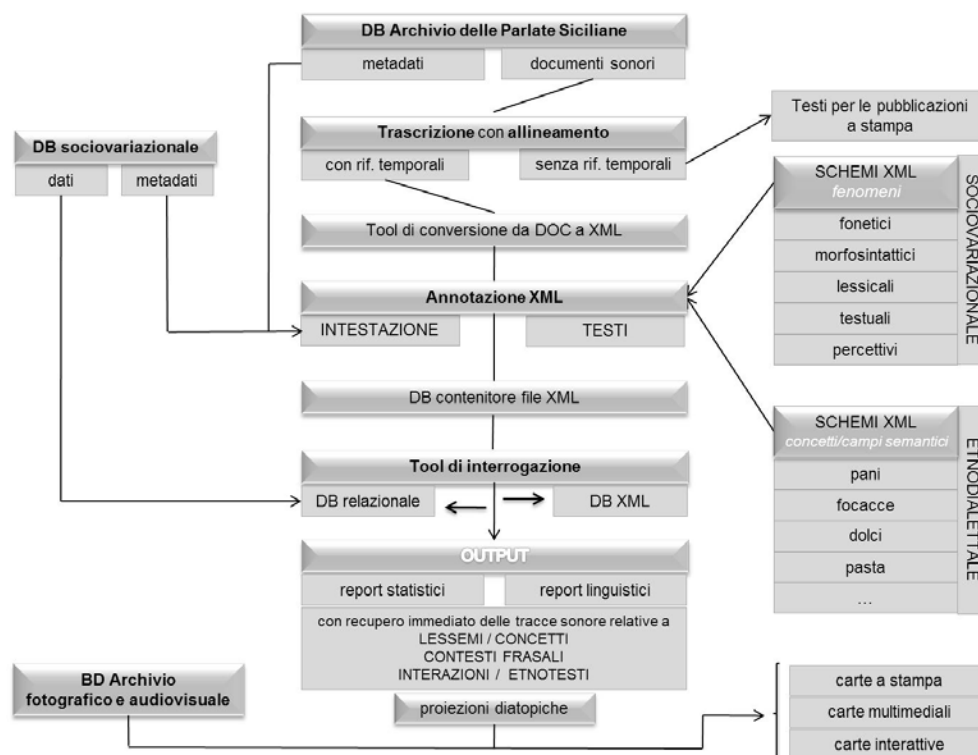


Figura 5: struttura del sistema informativo dell'Atlante Linguistico della Sicilia.

3. Prima di illustrare queste procedure, consentitemi una breve considerazione introduttiva.

In una ricerca linguistica fondata sui materiali raccolti sul campo, riteniamo che il nucleo informativo basilare debba essere l'intera intervista. Riteniamo che sul parlato, e non soltanto sulla parola, vada costruito un sistema in grado di rispondere alle interrogazioni sui vari livelli informativi potenzialmente compresi nelle interazioni tra intervistatore e intervistato. Ci siamo preoccupati, cioè, di creare la possibilità di consultare agevolmente non soltanto quelle porzioni del documento individuate come singole "risposte" a uno specifico quesito, ma anche di risalire ai rispettivi contesti linguistici e interlocutori. In altre parole, si è cercato di mettere a punto procedure informatiche in grado di evitare la frammentazione del testo/documento, la decontestualizzazione del tratto linguistico di volta in volta sottoposto ad analisi e la destrutturazione delle dinamiche interattive che l'hanno prodotto. Non ultimo, ci siamo preoccupati di evitare il definitivo accantonamento del documento sonoro nel quale le informazioni sono testimoniate. Vorremmo, insomma, evitare che le «informazioni» – una volta interpretate, incorniciate e divenute «dati» – perdano ogni rapporto con il documento sonoro, che testimonia anche il contesto linguistico e pragmatico nel quale sono occorse e nel quale continuano, in certo qual modo, a rivivere.

Queste condizioni non sono garantite, almeno nella nostra esperienza, da un'unica Banca Dati, bensì da un sistema modulare di gestione delle informazioni rilevate sul campo,

che sfrutta metodi, modalità e strumenti informatici di diverso tipo⁴, che ho provato a schematizzare in fig. 5.

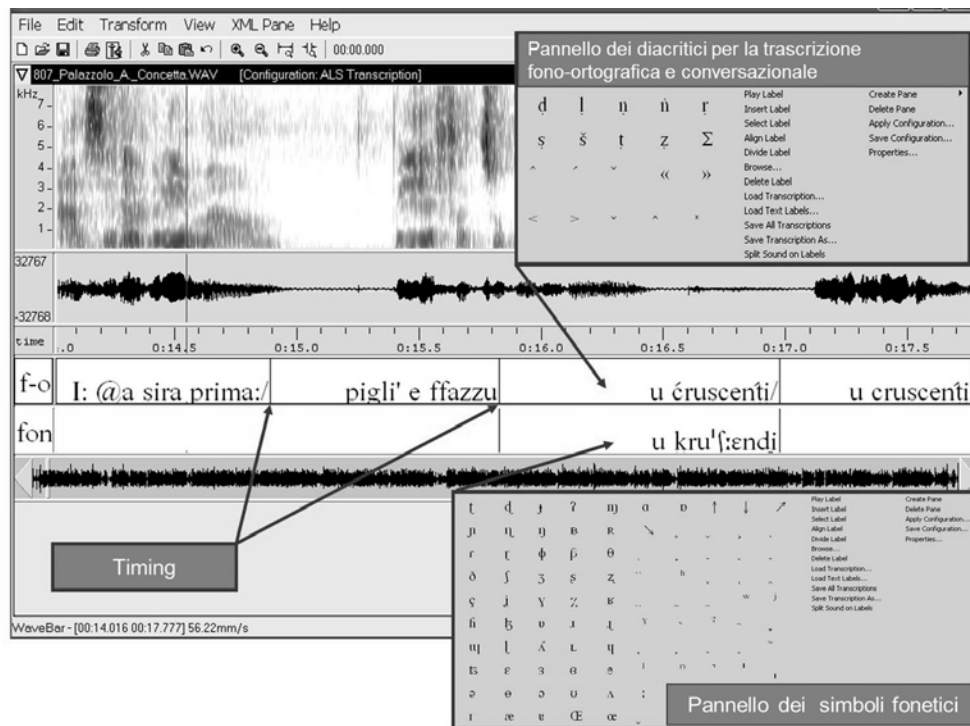


Figura 6: esempio di allineamento della trascrizione al segnale acustico.

4. Tornando alle procedure, i documenti sonori selezionati per una specifica analisi vengono innanzitutto sottoposti a trascrizione – parziale o integrale, a seconda del caso⁵ – attraverso l'uso di software e procedure che ne consentono l'allineamento temporale con il rispettivo segnale acustico: operazione, questa, che ci consente – come si mostrerà – oltre al reperimento anche l'ascolto immediato delle porzioni dei documenti sonori che testimoniano le informazioni di carattere linguistico e/o etnografico di volta in volta ricercate. A questo scopo, abbiamo adattato alle nostre specifiche esigenze⁶ il software Wavesurfer aggiungendo, parallelamente all'oscillogramma e al sonogramma, due aree in grado di ricevere la relativa trascrizione: nella superiore, quella fono-ortografica (cfr. Matranga, 2007) e conversazionale (cfr. Paternostro, 2007), nella inferiore quella fonetica (v. fig. 6).

⁴ Per la presentazione del sistema informativo dell'ALS (cfr. Matranga, *in stampa*).

⁵ Sulle scelte e i criteri della trascrizione dei documenti cfr. Matranga (2007).

⁶ Per gli aspetti informatici collabora con l'ALS, sin dal 2004, un gruppo di ricercatori, coordinato da Antonio Gentile, dell'ex Dipartimento di Ingegneria Chimica Gestionale Informatica Meccanica (DICGIM) dell'Università di Palermo. Cfr. *infra*, nota 9.

Il testo è segmentato attraverso l'inserimento di barre verticali posizionate di volta in volta in corrispondenza al segnale acustico sovrastante.

La trascrizione del documento viene salvata dal software con tutti i riferimenti temporali, di inizio e fine, di ogni segmento testuale. Questi riferimenti saranno automaticamente convertiti da un apposito *tool* in valori temporali (*timing*. V. fig. 7) delle etichette che costituiscono il sistema di annotazione basato sull'eXtensible Markup Language (XML).

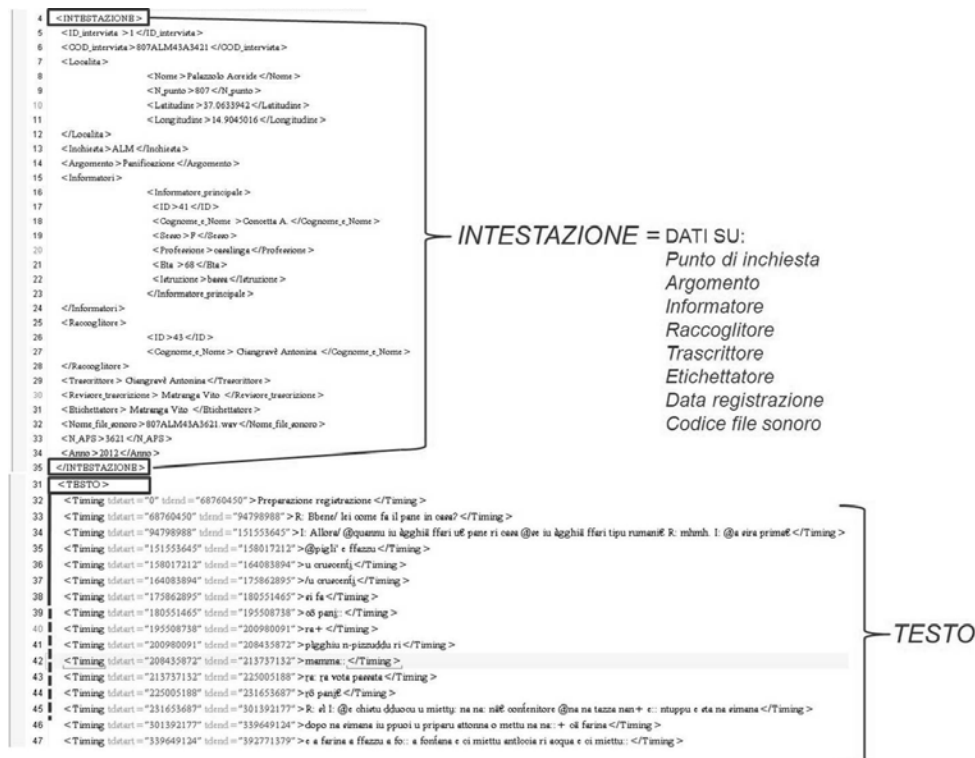


Figura 7: frammento iniziale di un testo pronto per l'annotazione.

In fig. 7 è riportato, come esempio, il frammento iniziale della trascrizione di un documento sonoro pronto per l'annotazione dopo l'allineamento con il segnale acustico. Il TESTO è anticipato dall'INTESTAZIONE contenente i dati essenziali relativi all'intervista in questione. Ciascun TESTO viene dunque sottoposto a markup utilizzando uno schema logico/interpretativo già predisposto per l'individuazione dei tratti interessati alla specifica ricerca. Sono già stati predisposti, infatti, diversi schemi logici formalizzati in .xsd/XML per l'annotazione dei tratti informativi di pertinenza fonetica, morfosintattica, del contatto lessicale e dell'analisi testuali (per quanto riguarda le inchieste sociovariazionali), e – come si vedrà – anche uno schema di base per l'annotazione dei testi di interesse etnodialeale⁷.

⁷ I modelli logici per l'annotazione dei fenomeni fonetici – indagati attraverso le alcune domande della terza parte del sezione sociovariazionale (cfr. Matranga *et alii*, 2006) – e dei

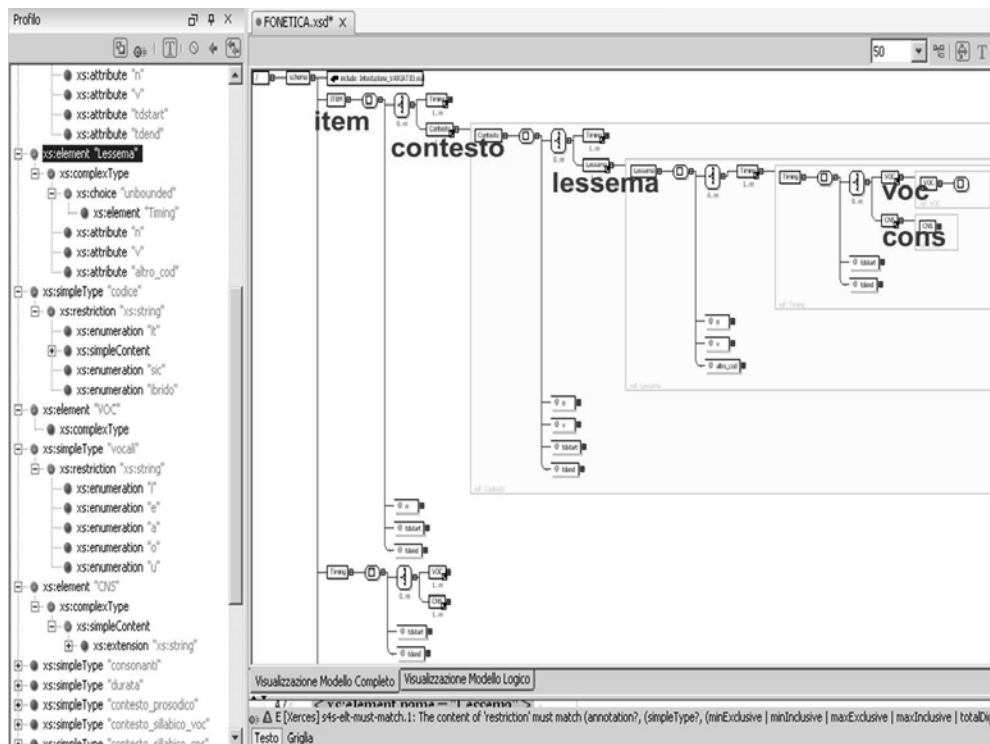


Figura 8: frammenti dello schema logico per l'annotazione dei fenomeni fonetici.

In fig. 8 sono riportati due frammenti esemplificativi dello schema logico per l'annotazione dei fenomeni di tipo fonetico investigati attraverso le apposite domande del questionario sociovariazionale dell'ALS⁸.

I testi annotati sono in seguito inseriti in un apposito «Contenitore» informatico, pronti per essere interrogati sulla base delle diverse etichette previste da ciascuno schema interpretativo precedentemente formalizzato.

La parte più complessa, dal punto di vista informatico, riguarda proprio lo strumento di interrogazione, predisposto a correlare dati e metadati relativi al parlante, e gestiti da una Banca dati di tipo relazionale, con dati linguistici desunti dai testi annotati in XML. Si è molto lavorato per la definizione di uno strumento di interrogazione che avesse

testi etnodialettali sono stati messi a punto e formalizzati da chi scrive. Con le indicazioni di Roberto Sottile (cfr. Sottile & Paternostro, 2012), Luisa Amenta (cfr. Amenta e Castiglione, 2006), Vincenzo Pinello (cfr. Pinello, *in stampa*) e Giuseppe Paternostro (cfr. Paternostro 2004 e 2013) sono già stati formalizzati anche gli schemi per l'annotazione dei fenomeni di interesse, rispettivamente, lessicale, morfosintattico, percettivo e pragmatico-testuale rilevati attraverso il questionario sociovariazionale.

⁸ Cfr. Ruffino & D'Agostino (2001).

un'interfaccia quanto più amichevolmente versatile, adattata alle esigenze di analisti non particolarmente esperti nella formalizzazione immediata di *queries* di una certa complessità⁹. In fig. 9 è riportata la finestra di interrogazione dei testi annotati per l'analisi di fenomeni fonetici (in questo esempio, la deaffricazione di /tS/).

ALSWEB - Finestra di ricerca III parte

Home Nuovo DB Xml Gestisci DB Xml Query I&II parte Query I,II&III parte

III parte del questionario
Nome database: Catania1

Tag ammissibili: CNS

Attributi:
 Cns == IS
 Cnx_fon == Deaftrz
 Durata == -- Qualsiasi --
 Fenom == -- Qualsiasi --
 Fenom2 == -- Qualsiasi --
 Ctx_pros == -- Qualsiasi --
 Ctx_sillb == -- Qualsiasi --

Proprietà tag Lessema: altro_cod == -- Qualsiasi --

Quesito	ITEM	COD. intervista	Località	ID Informatore	Codice richiesto	Prima lingua
Tutti		Tutti	Tutti	Tutti	Tutti	Tutti
I		701-54-01	Catania	01	italiano	Entrambi
II		701-54-02		02	siciliano	Italiano
III		701-54-03		03		Siciliano

Sesso: Tutti, Femmina, Maschio
 Età: Tutti, 15, 16, 18
 Fam. Sog.: Tutti, Figlio-a, Genitore, Nonno-a
 Fam. Tipo: Tutti, Tipo 0, Tipo 1, Tipo 2
 Istruzione: Tutti, Diploma di scuc, Elementare sen, Laurea, Licenza di scuc
 Professione: Tutti, insegnanti operai (o assiml)

Inserisci Cancella

Elementi criteri in OR: cerca il tag "CNS" con proprietà "Cns" ugua

Inserisci in AND ==>
 <== Modifica condizione

Esegui Query

Figura 9: finestra di interrogazione (es. deaffricazione di /tS/).

Il sistema restituisce, insieme ai dati statistici che soddisfano la richiesta formulata, anche le singole occorrenze testuali, così come sono state etichettate, e la possibilità di ascoltare la porzione del corrispettivo documento sonoro: possibilità concessa proprio grazie all'allineamento del segnale acustico alla trascrizione (v. fig. 10).

⁹ Per gli aspetti più strettamente informatici, cfr. Gentile *et alii* (2010); Russo *et alii* (2010); Pirrone *et alii* (2011); Pirrone *et alii* (in stampa).

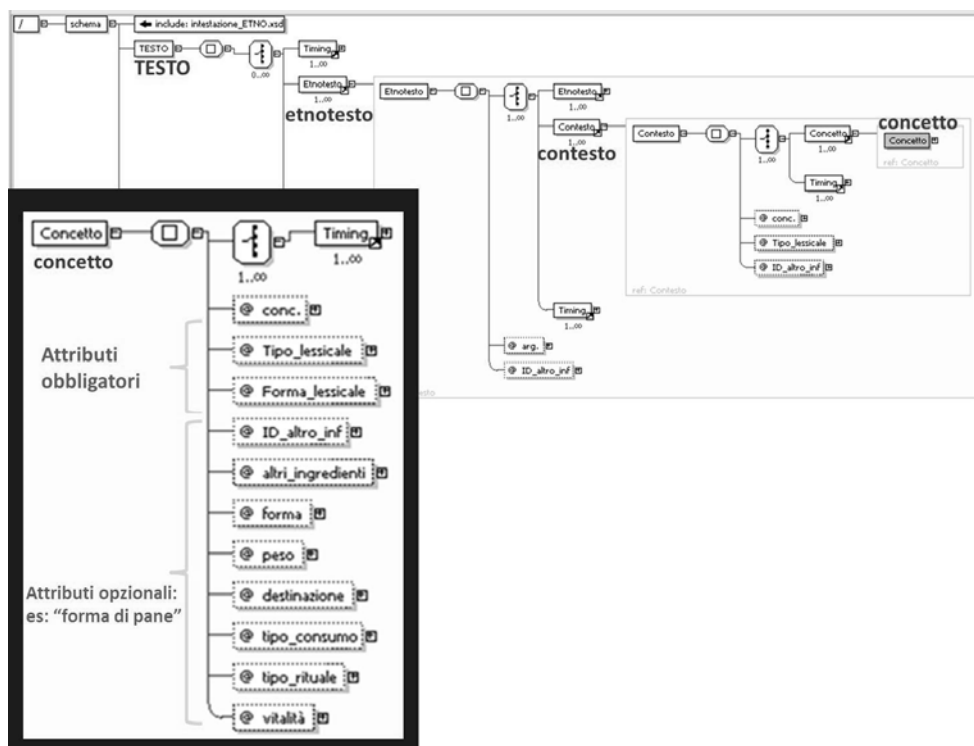


Figura 11: frammento dello schema logico per l'annotazione dei testi di interesse etnodialezionale.

L'analista seleziona e etichetta – sulla base di un ventaglio di «argomenti» preordinati – ciascun ETNOTESTO in ragione delle informazioni contenute in una determinata porzione dell'intera intervista (TESTO). Il CONTESTO rappresenta una porzione testuale minima (che generalmente coincide la frase), nella quale occorra una forma dialettale che denomi un CONCETTO interessato all'indagine su una specifica area semantica o «argomento».

In fig. 12 è riportato un esempio di annotazione di un'intervista sulla cultura alimentare, e specificamente sulla «panificazione tradizionale». In questo TESTO è stato individuato, per esempio, un ETNOTESTO sulla «lievitazione del pane», all'interno del quale occorrono, ancora per esempio, forme dialettali che esprimono CONCETTI come quello di «lievito madre», di «contenitore del lievito» etc.: forme lessicali che sono ovviamente inserite nei rispettivi CONTESTI linguistici/frasali i quali vengono anch'essi delimitati e annotati.

```

<TESTO>
  <Timing idstart="69157336" idend="94798988"> R: Bene! lei come fa il pane in casa? </Timing>
  <Blnoteto arg="Levitazione">
    <Timing idstart="94798988" idend="151553645"> I: Allora </Timing>
    <Contesto conc="PANB" Tipo_lesivale="pani,di,case">
      <Timing idstart="94798988" idend="151553645"> quannu iu agghi'a fferi </Timing>
      <Concetto conc="PANB" Tipo_lesivale="pani,di,case" Forma_lesivale="pani, di, casa">
        <Timing idstart="151553645" idend="158017210"> u panj di casa </Timing> </Concetto> </Contesto>
        <Timing idstart="158017210" idend="158017211"> ee iu agghi' a fferi tipu rumeni R: mhnh. </Timing>
      <Contesto conc="LIBVITO_MADRE" Tipo_lesivale="criscienti">
        <Timing idstart="158017211" idend="158017212"> I: a vira prima pigli' e ffazzu </Timing>
        <Concetto conc="LIBVITO_MADRE" Tipo_lesivale="criscienti" Forma_lesivale="cruscenij">
          <Timing idstart="158017212" idend="164083894"> u cruscenij </Timing> </Concetto> </Contesto>
        <Contesto conc="LIBVITO_MADRE" Tipo_lesivale="criscienti">
          <Concetto conc="LIBVITO_MADRE" Tipo_lesivale="criscienti" Forma_lesivale="cruscenij">
            <Timing idstart="164083894" idend="175862895"> u cruscenij </Timing> </Concetto>
            <Timing idstart="175862895" idend="195508738"> si fa oè panj: ra+ </Timing> </Contesto>
          <Contesto conc="LIBVITO_MADRE" Tipo_lesivale="mamma">
            <Timing idstart="200980091" idend="208435872"> pigghiu n-pizzaddu </Timing>
            <Concetto conc="LIBVITO_MADRE" Tipo_lesivale="mamma" Forma_lesivale="mamma">
              <Timing idstart="208435872" idend="213737132"> ri mamma: </Timing> </Concetto>
              <Timing idstart="213737132" idend="231653687"> rã... rã vota pavesa rã panj </Timing> </Contesto>
              <Timing idstart="231653687" idend="301392177"> R: el l' e ohiù ddocu y mettu: na na... na... confenitore na na tazza nan+ c:: ntuppu e sta na s'umana </Timing>
              <Timing idstart="301392177" idend="339649124"> dopo na s'umana iu ppoi u preparo attorna y mettu na na: + oè farina </Timing>
              <Timing idstart="339649124" idend="392771379"> e a farina a ffazzu a fo:: a foniana e oi mettu antocia ri soqua </Timing>
            <Contesto conc="LIBVITO_MADRE" Tipo_lesivale="criscienti">
              <Timing idstart="339649124" idend="392771379"> e oi mettu </Timing>
              <Concetto conc="LIBVITO_MADRE" Tipo_lesivale="criscienti" Forma_lesivale="cruscenij">
                <Timing idstart="392771379" idend="401265632"> stu cruscenij </Timing> </Concetto> </Contesto>
                <Timing idstart="401265632" idend="499106584"> chiù mpaùtu tutt'evieme e:: come oè mpaùtu attornj </Timing>
              <Contesto conc="CONTENITORE_LIBVITO" Tipo_lesivale="bainicoddu">
                <Timing idstart="401265632" idend="499106584"> a mëtterne nã+ nã+ nã+ </Timing>
                <Concetto conc="CONTENITORE_LIBVITO" Tipo_lesivale="bainicoddu" Forma_lesivale="bbainicoddu">
                  <Timing idstart="499106584" idend="502864596"> nã bbainicoddu </Timing> </Concetto> </Contesto>
                  <Timing idstart="499106584" idend="502864596"> e ppoi u ntuppu e l'indomeni matina è pprenta pi fferi: </Timing>
                  <Timing idstart="502864596" idend="506811691"> pi mpaùti </Timing>
                  <Timing idstart="506811691" idend="506811691"> u panj </Timing>
                </Blnoteto>

```

Figura 12: frammento di un testo annotato.

In fig. 13 si riportano gli stessi frammenti testuali mostrati in fig. 12, ma adattati in colonne (con una disposizione, pertanto, diversa da quella che effettivamente si presenta in fase di etichettatura XML) per agevolare una più immediata visualizzazione delle etichette.

Ciascuna occorrenza lessicale (*token*) relativa a un determinato CONCETTO previsto nello schema interpretativo è annotata, dunque, attraverso tre attributi obbligatori:

1. il «concetto», espresso in italiano, selezionato attraverso un menu tra quelli previsti dallo schema interpretativo;
2. la «forma lessicale», che riguarda la specifica esecuzione fonetica e morfologica desunta dalla trascrizione fon-ortografica;
3. il «tipo lessicale», determinato dall'etichettatore attraverso omologazioni/normalizzazioni di carattere morfologico, fonetico-fonologico e grafemico.

Le procedure relative alla tipizzazione sono, com'è noto, non prive di problemi e di insidie, sui quali non è qui il caso di soffermarsi¹⁰. Tuttavia, esse sono indispensabili al processo di markup, giacché è propriamente sul contenuto del «tipo lessicale» che si potrà pervenire all'agevole recupero delle rispettive varianti fonetiche o delle occorrenze morfologiche (ossia gli specifici 'valori' della «forma lessicale») riscontrate in tutti i documenti annotati.

¹⁰ Per un approccio ai criteri di tipizzazione adottati cfr. Matranga (2014).

Lo specifico «concetto» («conc.» in figg. 12 e 13), la rispettiva «forma lessicale» e il relativo «tipo lessicale» sono dunque i tre attributi obbligatori previsti per ogni CONCETTO. Altre annotazioni – sull'ordine linguistico e etnografico – possono essere sovrapposte, anche in momenti diversi, a uno stesso TESTO (già annotato) attraverso l'aggiunta di altri 'attributi' opzionali, i quali si possono prevedere al momento della formalizzazione dello stesso schema interpretativo, o – se sono particolarmente complessi – anche in una successiva fase di annotazione sovrapponendo un ulteriore e specifico schema, a condizione che questo mantenga la stessa struttura gerarchica dei nodi/elementi primari (ossia: ETNOTESTO, CONTESTO e CONCETTO) con i rispettivi attributi non opzionali (ossia: «conc.», «tipo lessicale» e «forma lessicale»).

In fig. 13 sono riportati, come esempio, gli attributi aggiuntivi relativi all'annotazione del concetto di 'forma di pane': ossia, «forma», «peso», «tipo di consumo», «tipo di ritualità», «destinazione», «vitalità».

ELEMENTO/TAG	TAG temporale	TESTO	chiusura
	<Timing tdstart="69157336" tdend="94798988">	R: <u>Bbene/</u> lei come fa il pane in casa?	</Timing>
	<Timing tdstart="94798988" tdend="151553645">	I: Allora	</Timing>
	<Timing tdstart="158017210" tdend="158017211">	se <u>iu agghi'</u> a <u>ffari tipu</u> <u>rumani</u>	</Timing>
	<Timing tdstart="158017211" tdend="158017212">	R: <u>mhmh</u> I: a sira prima pigli' e <u>ffazzu</u>	</Timing>
	<Timing tdstart="158017212" tdend="164083894">	<u>u cruscenij</u> /	</Timing> </Concetto> </Contesto>
	<Timing tdstart="401265632" tdend="499106584">	chištu <u>mpaštu</u> tutt'assieme e:: come cò <u>mpaštu attonnu</u>	</Timing>
	<Timing tdstart="401265632" tdend="499106584">	a mètterne nà + nò + nà +	</Timing>
	<Timing tdstart="401265632" tdend="499106584">	nù <u>bbainicieddu</u>	</Timing> </Concetto> </Contesto>
	<Timing tdstart="401265632" tdend="499106584">	e <u>ppoi</u> u <u>nùppu</u> e l'indomani <u>maùna</u> è <u>ppronfu</u> pi <u>ffari</u> ::	</Timing>
	<Timing tdstart="499106584" tdend="507864596">	pi <u>mpaštari</u> u <u>panj</u>	</Timing>

Figura 13: frammento di un testo annotato (adattato in colonne).

Ovviamente, l'interrogazione dei TESTI annotati può essere condotta sulla chiave di ogni attributo delle rispettive etichette: l'interrogazione del «concetto» restituisce ogni «tipo lessicale» riscontrato; l'interrogazione del «tipo lessicale» restituisce ogni «forma lessicale» a esso correlata.

Il sistema è in grado, dunque, di recuperare tutte le varianti lessicali e fonetiche correlate a uno specifico «concetto», corredate di report sulla frequenza delle occorrenze e contestualizzate nei rispettivi contesti frasali e etnotestuali, anche questi con aggancio immediato

delle porzioni del corrispettivo documento sonoro. È inoltre in grado di recuperare informazioni di ordine semantico ed etnografico quali – nell’esempio dei ‘pani’ – *forma, peso, tipo di consumo, ritualità, destinazione, vitalità*.

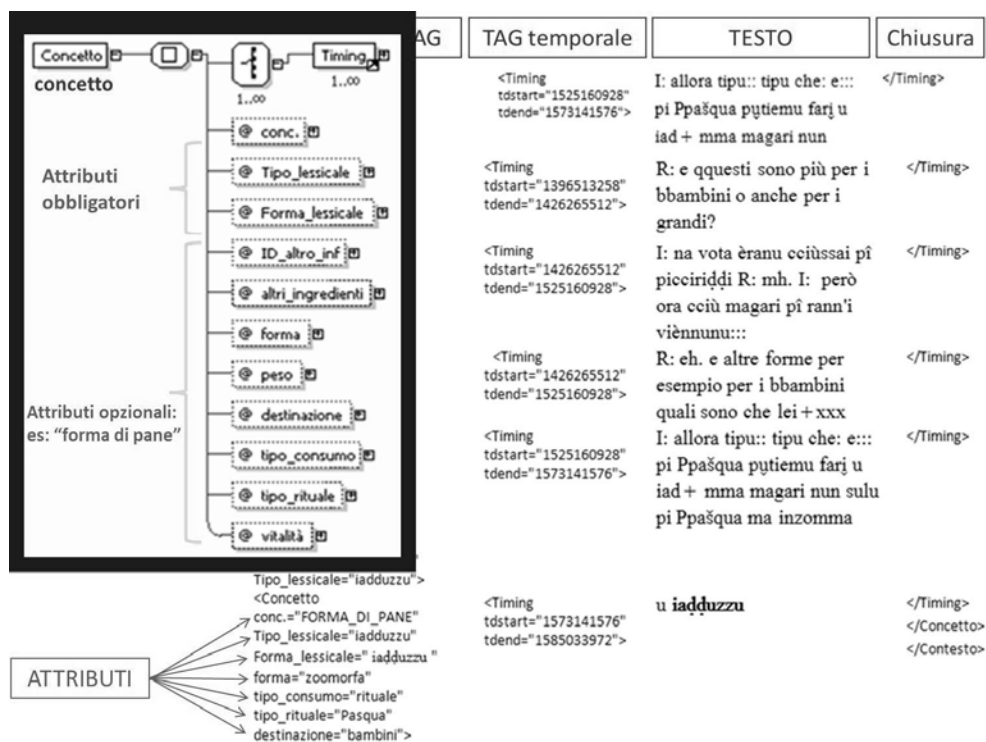


Figura 14: frammento di un testo annotato (adattato in colonne) con attributi opzionali.

6. Nonostante l’estrema sintesi, spero sia risultato evidente almeno che nel momento in cui si decide di mettere a punto un sistema inevitabilmente complesso di archiviazione, di trascrizione e, soprattutto, di annotazione delle informazioni contenute nelle interviste, si è obbligati a operare delle scelte che nessun informatico di mestiere può fare autonomamente, giacché esse impegnano un’indispensabile preventiva conoscenza dei problemi di ordine geolinguistico, sociolinguistico e – nel caso del trattamento delle interviste etnodialettali – anche di quello etnografico.

Soprattutto nella predisposizione/formalizzazione degli schemi logici interpretativi per l’annotazione (markup) dei tratti informativi, lo stesso linguista è obbligato a fare delle scelte precise, e più o meno definitive, su quali informazioni voglia ottenere dai materiali raccolti sul campo e quali altre è disposto a rinunciare.

Nel caso, poi, della tipologizzazione – ossia della definizione del rapporto tra la variante effettivamente occorsa nella nominalizzazione di un concetto (e annotata nell’attributo «forma lessicale») e il rispettivo «tipo lessicale» (preteso dal sistema di annotazione) – il linguista è poi obbligato a definire un sistema complesso di puntuali relazioni

tra le diverse “forme” e il “tipo” corrispondente, e dunque a riflettere – operando delle scelte definitive – anche sui criteri di normalizzazione della variazione diatopica.

Anche nell’ambito della geografia linguistica l’approccio informatico si è, dunque, di fatto intrecciato con alcuni aspetti di ordine teorico-metodologico, sui quali varrebbe la pena di riflettere. Ma non qui e non ora!

BIBLIOGRAFIA

Amenta, L., & Castiglione, M. (2006), Tradurre, correggere, correggersi: dati morfosintattici e processo internazionale, in D’Agostino e Paternostro, pp. 115-134.

D’Agostino, M., & Paternostro, G. (2006), a cura di, Costruendo i dati. Metodi di raccolta, revisione e organizzazione della Banca dati nella sezione socio variazionale, Centro di studi filologici e linguistici siciliani - Dipartimento di Scienze filologiche e linguistiche - Facoltà di Lettere e Filosofia dell’Università di Palermo, Palermo.

D’Agostino, M., & Ruffino, G. (2002), a cura di, Questionario dell’Atlante Linguistico della Sicilia. Sezione sociovariazionale, Dipartimento di Scienze filologiche e linguistiche dell’Università di Palermo - Centro di studi filologici e linguistici siciliani, Palermo.

D’Agostino, M., & Ruffino, G. (2005), I rilevamenti socio variazionali. Linee progettuali, Centro di studi filologici e linguistici siciliani -Dipartimento di Scienze filologiche e linguistiche - Facoltà di Lettere e Filosofia dell’Università di Palermo, Palermo.

Galati, S., & Luppina, A., a cura di, Archivio della Parlate Siciliane (APS). I documenti sonori, in Matranga 2011a, pp. 29-79.

Gentile, A., Pirrone, R., & Russo, G. (2010), Knowledge Discovery for the Linguistic Atlas of Sicily Project, International Journal of Computer Systems Science & Engineering, Vol. 25, no. 2.

Matranga, V., (2007), Trascrivere. La rappresentazione del parlato nell’esperienza dell’Atlante Linguistico della Sicilia, Centro di studi filologici e linguistici siciliani - Dipartimento di Scienze filologiche e linguistiche - Facoltà di Lettere e Filosofia dell’Università di Palermo, Palermo.

Matranga, V., (2011a), a cura di, Bollettino dell’Archivio delle Parlate Siciliane, no.1, Centro di studi filologici e linguistici siciliani - Dipartimento di Scienze filologiche e linguistiche dell’Università di Palermo, Palermo.

Matranga, V., (2001b), Archivio delle Parlate Siciliane (APS). La Banca Dati, in Id. 2011a, 13-27.

Matranga, V., Serio, S., Soriani, G., & Sottile, R., (2006), Dalla lettura al parlato: I dati fonetici, in D’Agostino & Paternostro, 135-154.

Matranga, V., (2014), Dal parlato alla parola. Procedure informatiche per il trattamento delle interviste etnodialettali nell’Atlante Linguistico della Sicilia (ALS), in Cugno F., Mantovani L., Rivoira M., Specchia S., a cura di, Studi linguistici in onore di Lorenzo Massobrio, Istituto dell’Atlante Linguistico, Torino.

Matranga, V., (in stampa), L’Archivio delle Parlate Siciliane e il sistema informativo

dell'Atlante linguistico della Sicilia, in Atti del Convegno Internazionale di Studi su Archivi Etnolinguistici Multimediali. Dalla formazione alla gestione e al dialogo col territorio (Pescara, 6 ottobre 2012).

Paternostro, G. (2004), Un parlante, tanti parlanti. Riflessioni sulle produzioni delle fonti nella geolinguistica variazionale, *Rivista Italiana di Dialettologia*, no. 28, 45-86.

Paternostro, G. (2007), La trascrizione conversazionale, in *Matranga 2007*, 103-136.

Paternostro, G. (2013), Discorso, interazione, identità. Studiare il parlato attraverso i parlanti, Peter Lang, Frankfurt.

Pinello, V. (in stampa), Modelli spaziali e cognitivi per l'interpretazione dei dati metalinguistici. L'esperienza dell'Atlante Linguistico della Sicilia, Peter Lang, Frankfurt.

Pirrone, D., Russo, G., Gentile A., Pirrone R., & Gaglio, S. (2011), The ALSWEB Framework: a Web-based Framework for the Linguistic Atlas of Sicily Project, *International Conference on Advanced Information Networking and Applications (AINA)*, 571-577.

Pirrone, D., Russo, G., Gentile A., & Pirrone R. (2013), *Collective reasoning over shared concepts for the Linguistic Atlas of Sicily project*. Submitted as chapter in *Inter-cooperative collective intelligence: techniques and applications*.

Russo, G., Gentile A., Pirrone R. & Mollame, A. (2010), Information Organization and Visualization for the ALS Project, In *Proceedings of International Conference on Complex, Intelligent and Software Intensive Systems*, (Atti del Convegno, Krakow, Poland, 15-18 Febbraio 2010), 724-728.

Sottile, R., & Paternostro, G. (2012), La variazione lessicale nella Sicilia contemporanea. Italiano e siciliano nei quesiti onomasiologici dell'Atlante Linguistico della Sicilia (ALS), in Bianchi, P., De Blasi, N., De Caprio, C., & Montuori, F., a cura di, *La variazione dell'italiano e nella sua storia. Varietà e varianti linguistiche e testuali* (Atti dell'XI Congresso Internazionale della Società di Linguistica e Filologia italiana, Napoli 5-7 ottobre 2010), F. Cessati Ed., Firenze, 811-823.