

Francesco Scaglione

“Lavorare” il dato linguistico: prospettive e limiti. *Alcune considerazioni dall’esperienza dell’Atlante Linguistico della Sicilia (ALS)*

Abstract

Il presente contributo si propone di riflettere sul rapporto sempre più stretto e interdipendente tra ricerca linguistica e nuove tecnologie per la processazione e analisi del dato. Partendo dall’esperienza dell’*Atlante Linguistico della Sicilia*, verrà presentato un sistema di annotazione in XML, formalizzato secondo il modello della *trasferenza* elaborato da Regis (2013), per il recupero e l’analisi dei dati tratti dal quesito onomasiologico (domanda I) del questionario linguistico della sezione sociovariazionale dell’ALS, quesito volto a indagare specifiche conoscenze lessicali e il contatto tra italiano e dialetto in Sicilia. Ciò permetterà, inoltre, di riflettere sui vantaggi e i limiti dell’ausilio di strumenti informatici con cui la ricerca linguistica deve sempre più confrontarsi.

The present work focuses on the relationship between linguistic research and the use of new technologies for linguistic data processing and analysis. Starting from the experience of the *Atlante Linguistico della Sicilia* (ALS), this paper describes a XML schema, based on the theory of *trasferenza* (*transference*) by Regis (2013), for the annotation and analysis of the data from the onomasiological questions of the ALS sociovariational questionnaire. Moreover, this modest case study tries to make clear the pros and cons of technological devices in linguistic research.

1. Introduzione: vecchie sfide, nuovi strumenti

Uno degli aspetti che sembra aver segnato l’evoluzione e, in parte, le direzioni della ricerca linguistica degli ultimi decenni è di certo la fortissima influenza dei recenti sviluppi tecnologici (soprattutto di ambito informatico) che hanno permesso non solo di considerare e analizzare il dato linguistico in una prospettiva del tutto nuova (con la possibilità di ottenere risultati prima impensabili), ma anche di raccogliere le sfide (ancor più complesse per le scienze “molle”, come la linguistica) che il progresso e la “contemporaneità” hanno via via lanciato.

In tal senso, l’apporto più importante riguarda principalmente la creazione e la gestione di vaste banche dati (molto spesso pensate per dialogare tra loro), fenomeno che ha accelerato o, piuttosto, rinvigorito la riflessione epistemologica e soprattutto metodologica che più che mai caratterizza la linguistica dei corpora, disciplina ormai autonoma, definibile oggi come il versante “tecnico-informatico”

della linguistica¹. Le principali motivazioni che hanno da sempre accompagnato l'allestimento di banche dati informatizzate sono legate soprattutto a una necessità connessa al “fattore tempo” che, in una dimensione bivalente, si coniuga ora nell'immediatezza del recupero del dato, ora nella velocizzazione del processo di analisi, rappresentazione e divulgazione dei risultati. Quest'ultimo aspetto rappresenta un elemento determinante soprattutto per quella ricerca sociolinguistica che si propone di restituire un quadro di fenomeni in uso o in atto, i cui lunghi tempi di analisi creano a volte un incolmabile iato, uno scollamento tra dato e realtà esaminata a causa della divulgazione di risultati ormai “vecchi”, superati e non più corrispondenti alla situazione nel tempo reale.

Ma al di là e accanto all'immagazzinamento dei dati all'interno di un grande “contenitore” informatico, il momento dell'analisi e la sua celerità pongono un'ulteriore questione che riguarda la creazione e lo sviluppo di software, frutto di specifici processi di *language engineering* (cf. Cunningham 1999), destinati all'esecuzione di *language-related tasks* che aiutino il linguista al superamento di una dimensione esclusivamente descrittiva, in favore di un'analisi caratterizzata anche da aspetti connessi alla distribuzione, variazione, occorrenza e rappresentatività del dato.

Di certo, l'aspetto “diacronico” nella sua doppia valenza (recupero e analisi del dato) e il bisogno di strumenti e sistemi informatici utili a esaminare e rappresentare particolari fenomeni accompagnano da sempre la ricerca geo-socio-linguistica, “sopraffatta” molto spesso, per dirla sinteticamente con D'Agostino e Paternostro (2006), dall'“esplosione del dato”, dalla sua complessità e “fluttuazione” e dalla necessità di chiare coordinate di analisi. I nuovi programmi e software, come si vedrà più nello specifico, rappresentano, pertanto, un importante ausilio attraverso cui “orientarsi” tra i molteplici percorsi interpretativi in cui il linguista sembra a volte “perdersi”, ma soprattutto attraverso cui gestire, leggere ed esaminare una considerevole mole di dati secondo precisi e comuni modelli analitici.

Ma all'interno di questi processi di informatizzazione, qual è il ruolo del linguista? Per prima cosa, è utile precisare che la messa a punto di qualsiasi

¹ È opportuno specificare che la linguistica dei corpora, nella sua qualità di disciplina *text-based*, ha, in realtà, una storia assai antica. Infatti, ricerche basate sull'analisi di vasti corpora (in particolar modo di ambito letterario) hanno accompagnato e sostanziato da sempre importanti imprese lessicografiche. Basti pensare a *A Dictionary of the English Language* del Doctor Johnson, al *Diccionario de Autoridades* della *Real Academia Española*, o restando in ambito italiano, alle antiche opere lessicografiche dell'*Accademia della Crusca*. Tuttavia, a partire dalla seconda metà del XX secolo, l'avvento della tecnologia digitale ha provocato un radicale cambiamento nella ricerca, che risulta oggi indissolubilmente legata all'uso di particolari software informatici (KENNEDY 1998, 5).

software o strumento informatico di archiviazione e analisi non rappresenta semplicemente un momento preliminare della ricerca, ma è essa stessa ricerca, presupponendo riflessioni di ordine teorico e metodologico affidate anche alla “sensibilità”, all’intuizione e ai bisogni del linguista. Quest’ultimo, quindi, non si limita ad acquisire passivamente una certa dimestichezza nell’utilizzo di specifici sistemi e programmi (in molti casi abbastanza intuitivi), ma assume un ruolo centrale nel dialogo per l’elaborazione di soluzioni informatiche che meglio si adattino e si “pieghino” a particolari schemi di “lavorazione” e analisi del dato, strettamente connessi a una precisa domanda conoscitiva.

2. Uno schema di analisi

Come in parte accennato, la ricerca geo-sociolinguistica ha raccolto le sfide del mondo contemporaneo, trovando nei supporti informatici importanti strumenti di ricerca per nuovi approcci di analisi e per una più veloce rappresentazione e divulgazione del dato: dalle carte geolinguistiche un tempo raccolte in enormi volumi, si è passati a più agevoli carte digitali interrogabili per concetti o per punto(-i) d’inchiesta (basti pensare al progetto *NavigAIS*, curato da Graziano Tisato, che ripropone on-line le carte dell’*Atlante Italo-Svizzero*, www3.pd.ist.cnr.it/navigais-web/, cf. Tisato 2015; o alla carta parlante relativa ai nomi della ‘trottola’ in Sicilia a cura di Ruffino 1997), fino ad arrivare a progetti di geolinguistica digitale sovraregionale (come il recentissimo *VerbAlpina*, diretto da Krefeld e Lücke in cui, tra le tante possibilità, l’utente può persino creare carte interattive *ad hoc* tramite il supporto di *Google Maps* a partire da diversi criteri, quali concetto, tipo lessicale, fonetico, morfo-lessicale, etc., cf. tra gli altri Krefeld–Lücke 2014²) e ai sempre più utilizzati sistemi XML (*eXtensible Markup Language*) per la gestione e l’interrogazione di vasti corpora riguardanti l’indagine e il *computing* di specifici tipi e usi linguistici da cui astrarre teorie e correlazioni in base a variabili geo-etnografiche, sociolinguistiche o testuali (cf. Carletta *et al.* 2005)³.

² *VerbAlpina* prevede, grazie a una specifica *App*, la fruizione dei dati tramite i più recenti supporti digitali (*smartphone*, *tablet*, etc...). Inoltre, il sistema, secondo le modalità del *crowdsourcing*, permette agli utenti la possibilità di creare un proprio “ambiente” di ricerca per la raccolta di dati (di interesse soprattutto lessicale), eventualmente condivisibili sulla rete.

³ Riguardo al rapporto tra trattamento dei dati geo-etno-sociolinguistici e nuove tecnologie, è utile inoltre ricordare le recenti imprese di ambito regionale, interregionale e microregionale, tra le quali l’*Atlante Multimediale dei Dialetti Veneti* (AMDV), il *Piccolo Atlante Linguistico dei Walser Meridionali* (PALWaM) e il progetto *Culture e Lingue delle Alpi del Piemonte* (CLAPie). Per una

Soffermandoci su quest'ultimo aspetto, ci proponiamo di presentare in questa sede un sistema di annotazione in XML messo a punto all'interno del cantiere della sezione sociovariazionale dell'*Atlante Linguistico della Sicilia* (ALS), destinato alla creazione di un *database* per il recupero e l'analisi dei dati relativi al quesito onomasiologico (Domanda I) della parte linguistica (III parte) del questionario, volto a indagare conoscenze lessicali e a "intercettare" possibili forme di contatto tra italiano e dialetto. In questa parte dell'intervista, attraverso una tecnica escussiva tradizionale (ovvero, l'ostensione di immagini), viene chiesto agli informatori di nominare prima in italiano e poi in dialetto una serie di oggetti e azioni della vita quotidiana, legati al lessico tradizionale e arcaico siciliano (Castiglione *et al.* 2006, 64-66). Di seguito (Tabella 1) vengono riportati i 36 referenti della domanda onomasiologica, secondo l'ordine e la successione previsti dal questionario (D'Agostino – Ruffino 2002, 9):

TERZA PARTE

I. Dire le forme italiane e siciliane (*utilizzare fotografie o disegni*). Nei casi in cui viene dato il tipo prettamente dialettale, chiedere: «lei usa questa parola?»

	<u>USA</u>			<u>USA</u>	
	Si	No		Si	No
1. culla	<input type="checkbox"/>	<input type="checkbox"/>	19. lumaca	<input type="checkbox"/>	<input type="checkbox"/>
2. bambola	<input type="checkbox"/>	<input type="checkbox"/>	20. pesche	<input type="checkbox"/>	<input type="checkbox"/>
3. spillo	<input type="checkbox"/>	<input type="checkbox"/>	21. terrazzo	<input type="checkbox"/>	<input type="checkbox"/>
4. rana	<input type="checkbox"/>	<input type="checkbox"/>	22. trottole	<input type="checkbox"/>	<input type="checkbox"/>
5. basilico	<input type="checkbox"/>	<input type="checkbox"/>	23. spugna	<input type="checkbox"/>	<input type="checkbox"/>
6. gruccia	<input type="checkbox"/>	<input type="checkbox"/>	24. foglia	<input type="checkbox"/>	<input type="checkbox"/>
7. albicocche	<input type="checkbox"/>	<input type="checkbox"/>	25. salvadanaio	<input type="checkbox"/>	<input type="checkbox"/>
8. prezzemolo	<input type="checkbox"/>	<input type="checkbox"/>	26. coperta	<input type="checkbox"/>	<input type="checkbox"/>
9. pipistrello	<input type="checkbox"/>	<input type="checkbox"/>	27. ciliegie	<input type="checkbox"/>	<input type="checkbox"/>
10. ago	<input type="checkbox"/>	<input type="checkbox"/>	28. pantaloni	<input type="checkbox"/>	<input type="checkbox"/>
11. tacchino	<input type="checkbox"/>	<input type="checkbox"/>	29. uva	<input type="checkbox"/>	<input type="checkbox"/>
12. sbucciare	<input type="checkbox"/>	<input type="checkbox"/>	30. carciofo	<input type="checkbox"/>	<input type="checkbox"/>
13. mela	<input type="checkbox"/>	<input type="checkbox"/>	31. setaccio	<input type="checkbox"/>	<input type="checkbox"/>
14. grembiule	<input type="checkbox"/>	<input type="checkbox"/>	32. spazzola	<input type="checkbox"/>	<input type="checkbox"/>
15. comodino	<input type="checkbox"/>	<input type="checkbox"/>	33. agave	<input type="checkbox"/>	<input type="checkbox"/>
16. ragnatela	<input type="checkbox"/>	<input type="checkbox"/>	34. mestolo	<input type="checkbox"/>	<input type="checkbox"/>
17. cencio/straccio	<input type="checkbox"/>	<input type="checkbox"/>	35. sedano	<input type="checkbox"/>	<input type="checkbox"/>
18. mandorle	<input type="checkbox"/>	<input type="checkbox"/>	36. frittata	<input type="checkbox"/>	<input type="checkbox"/>

Tabella 1. Referenti della Domanda I del questionario linguistico dell'ALS

panoramica aggiornata sullo stato di avanzamento di tali progetti si rimanda interamente ai contributi contenuti in BALI 39.

L’obiettivo principale della Domanda I è, quindi, quello di “misurare”, grazie al sistema di campionamento degli informatori dell’ALS e alla rete di punti rilevati⁴, la vitalità del sostrato lessicale arcaico in prospettiva diastratica, diatopica e diagenazionale, ma soprattutto di esaminare il contatto tra codici in una realtà non più di rigida diglossia, ma di forte compenetrazione tra varietà. Infatti, tale *naming test* non mira semplicemente a registrare il passaggio da un sistema a un altro, secondo una netta opposizione tra conservazione ~ innovazione (che si concretizza molto spesso nella perdita della parola dialettale sostituita dalla forma italiana più o meno “dialettalizzata”), ma anche a evidenziare spazi linguistici intermedi che mostrano, invece, uno scenario *in fieri* in cui è possibile osservare una crescente variabilità interna al repertorio e al (dia)sistema (Sottile – Paternostro 2012, 813; cf. anche Sottile – Capitummino 2010). Non a caso, l’analisi si concentra in ciascuna delle serie (italiana e siciliana) soprattutto sulle “forme non attese”⁵, vale a dire forme di “rottura” che non rientrano pienamente in nessuno dei due poli estremi del *continuum* e che mettono a fuoco, quindi, importanti zone di contatto e di scambio tra i codici.

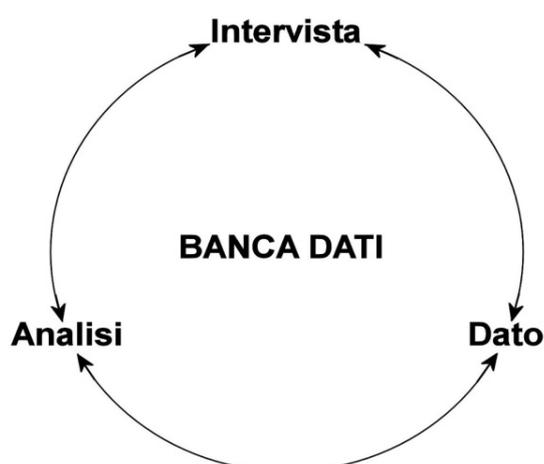
⁴ Gli informatori sono divisi in cinque famiglie costituite da Nonno, Genitore e Figlio e selezionati secondo criteri diastratici, quali prima lingua di apprendimento (italiano o siciliano) e livello d’istruzione, permettendo, quindi, di tracciare un *gradatum* (o forse, un *continuum*) che da un polo di maggiore dialettologia passa via via a una condizione di prevalente italoфония (D’AGOSTINO – RUFFINO 2005, 90).

La rete d’inchiesta, a partire da criteri socio-spaziali, raccoglie punti dinamici, recessivi, microaree (generalmente costituite da due o tre punti geograficamente contigui e socio-demograficamente e linguisticamente omogenei), punti medio-piccoli e grandi, città e aree periurbane, in una prospettiva in cui lo spazio non è uno scenario asettico, ma è un luogo vissuto, agito e riflesso dai parlanti (cf. D’AGOSTINO – PENNISI 1995; RUFFINO 1995; D’AGOSTINO – RUFFINO 2005; D’AGOSTINO 2006; per una prima ricognizione sulle dinamiche linguistiche tra centri urbani e centri rurali relative della Domanda I, cf. D’AGOSTINO 2002).

⁵ L’espressione riassume, più o meno esplicitamente, la parabola analitico-interpretativa della Domanda I all’interno del questionario linguistico dell’ALS. Infatti, nella sua fase iniziale, sulla scorta, in parte, della ricerca condotta da MOCCIARO (1989) a Mandanici (nel messinese) e sulla scia degli studi sui processi di italianizzazione del lessico dialettale (cf., sempre per la Sicilia, anche TROPEA 1991), il quesito onomasiologico si proponeva di indagare esclusivamente la conoscenza e la vitalità di una serie di arcaismi del dialetto siciliano (RUFFINO 1991, 26). Tuttavia, la variabilità riscontrata nei dati raccolti, da una parte, ha mutato la domanda conoscitiva che si è via via ricollocata verso la direzione di cogliere ed esplorare le diverse soluzioni che si discostano dalla forma attesa (per esempio, dialetto *naca*, italiano *culla*, etc.) e in cui si apprezzano in molti casi interessanti fenomeni “intermedi”, di contatto e di compenetrazione tra i due codici; dall’altra, il cambiamento di prospettiva ha determinato la necessità di classificare tali esiti che, non rientrando pienamente né nell’italiano né nel dialetto, sembrano invece disporsi (più o meno perfettamente) a metà strada tra i due sistemi.

2.1. Riflessioni e procedimenti preliminari

Il progetto dell'ALS (sia nella sezione sociovariazionale che etnodialettale) ha da sempre considerato l'intervista la base su cui porre le sue fondamenta empiriche e attraverso cui cogliere ed esaminare la variazione in ogni suo aspetto (linguistico, sociale, diatopico ed etnografico). Tale prospettiva non può che riflettersi anche nell'archiviazione e analisi dei dati tramite un sistema in cui, a partire dal parlato (e non soltanto dalla singola parola), sia possibile operare interrogazioni sui vari livelli informativi che via via si (co)costruiscono nell'interazione tra intervistatore e intervistato (Matranga 2014, 639; Matranga – Sottile 2007, 202-203). Ciò determina, quindi, un percorso di processazione e consultazione circolare (in un verso e nell'altro) che dall'intervista passa al dato e, quindi, all'analisi per poi eventualmente ritornare all'intervista (vedi Schema 1).



Schema 1. Modello banca dati ALS

I presupposti teorici che accompagnano il sistema di informatizzazione dell'ALS si riscontrano, inoltre, nell'analisi della Domanda I, in cui si è ritenuto altrettanto opportuno non recidere il legame tra intervista e dato. Al di là di un sistema di archiviazione del materiale audio che, a partire dall'attribuzione di una serie codici, permette di ricondurre l'intervista (o una sua parte) al parlante e al suo spazio linguistico, tale aspetto è stato realizzato soprattutto attraverso un processo preliminare di allineamento, tramite *Wavesurfer*, tra parlato e (tra)scritto:

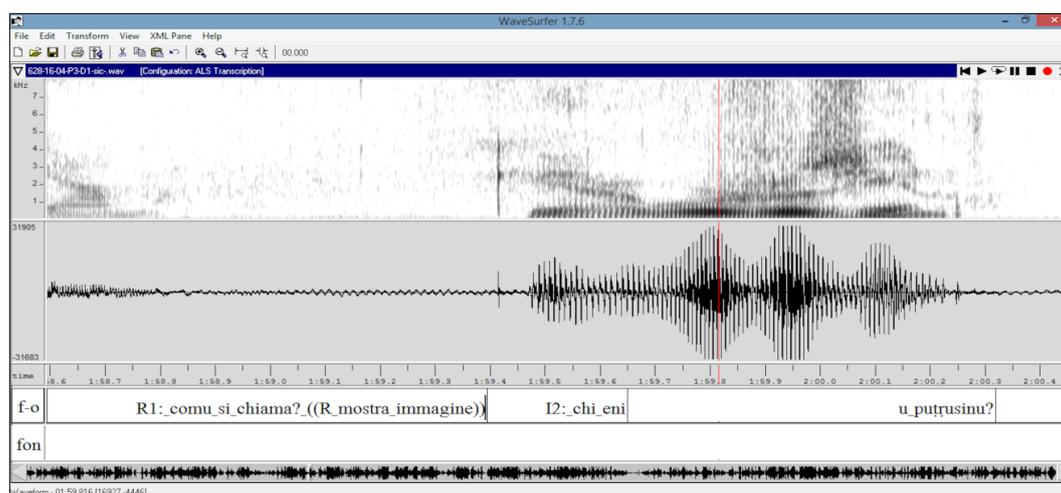


Figura 1. Processo di allineamento parlato-trascritto

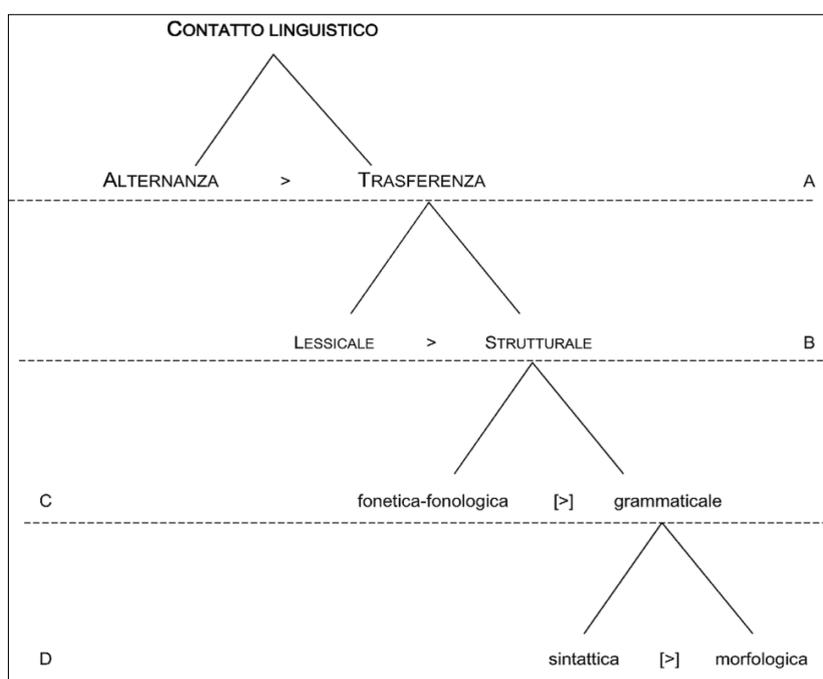
Come illustrato in Figura 1, il sistema permette di suddividere il file audio dell'intervista, isolando una precisa frazione temporale (riascoltabile in ogni momento) in cui è possibile inserire la trascrizione del contenuto⁶. Nel caso proposto, vengono riportate le varie parti del processo elicittivo in serie siciliana dell'item PREZZEMOLO, in cui l'informatore (nonno di II famiglia di Capo d'Orlando – ME) restituisce in modo diretto la forma dialettale attesa (*putrusinu*). In piena sintonia con gli approcci più tradizionali, la trascrizione continua ad essere, quindi, il punto di partenza fondamentale del processo di analisi, e ciò risulta ancor più evidente nel nostro caso in cui il dato (trascritto) mantiene il proprio legame con il contesto interazionale che lo ha generato.

Completato l'allineamento, il programma genera automaticamente un file in cui sono annotate le varie parti trascritte, precedute dalle coordinate temporali rispetto al file audio di riferimento, che verrà successivamente convertito e inserito in ambiente *Oxigen* per l'annotazione (§ 2.2).

Parallelamente a tale processo preliminare, un ulteriore momento di riflessione, questa volta però di ordine particolare e legato alla “lavorazione” della nostra base empirica di riferimento, ha riguardato l'elaborazione e la ricerca di un modello epistemologico che potesse costituire la griglia di analisi delle forme elicitate nella Domanda I. Nello specifico, si è fatto ricorso alla teoria della trasferenza elaborata da Regis (2013), il quale, ripercorrendo il solco tracciato dallo studio pionieristico di Weinreich (1953), ma con uno sguardo rivolto agli aspetti

⁶ Le due fasce presenti nella parte bassa della Figura 1 prevedono la possibilità di trascrivere sia secondo un sistema fono-ortografico (*f-o*) che in caratteri IPA (*fon*). Nella nostra ricerca viene adoperato esclusivamente il più agevole sistema fono-ortografico elaborato e formalizzato in MATRANGA (2007), a cui si rimanda per i criteri di trascrizione adottati nel presente contributo.

gerarchici implicati nel contatto linguistico (cf., ad esempio Thomason – Kaufman 1988, 74-76), propone un modello descrittivo-induttivo per la classificazione dei fenomeni di contatto, nato anche dalla necessità di “domare” l’eterogeneità dei dati lessicali emersi dalle inchieste sociovariazionali dell’ALS. Il fulcro della teoria è il concetto, per l’appunto, di trasferenza, intesa come «il passaggio di una o più forme, un sintagma, un tratto o una costruzione da un codice all’altro» (Regis 2013, 14). Nel particolare, tale nozione comprende la trasferenza lessicale (TL) e la trasferenza strutturale (TS): la prima si realizza nel momento in cui una parola o un sintagma esogeno è “trasferito” all’interno della cornice morfo-sintattica della lingua ospitante tramite prestito o *code-mixing*; la seconda, invece, si snoda in trasferenza fonetica-fonologica (relativa al passaggio di foni e proprietà fonologiche) e grammaticale (concernente aspetti sintattici e morfologici). Le categorie e sottocategorie descritte si collocano, inoltre, all’interno di un modello gerarchico-implicazionale che disciplina il rapporto tra i diversi fenomeni:



Schema 2. Modello gerarchico-implicazionale della trasferenza (REGIS 2013, 32)

In base allo Schema 2, l’alternanza⁷ implica anche la trasferenza (livello A), mentre la TL diventa “veicolo” primario di elementi strutturali (livello B). Scendendo nella subcategoria della TS (livello C), si suppone che la trasferenza fonetico-fonologica preceda quella grammaticale e che, infine, quest’ultima implichi prima un passaggio di proprietà sintattiche e poi di tratti morfologici (livello D) (*Id.*, 33).

Tale quadro epistemologico può essere letto (soprattutto per i criteri gerarchico-implicazionali appena evidenziati) sullo sfondo del paradigma della fenomenologia del contatto linguistico elaborato da Berruto (2009, 7) a partire dalla dicotomia tra *lingue a contatto* (in cui non è previsto bilinguismo individuale o comunitario) e *lingue in contatto* (dove, invece, si presuppone un certo grado di bilinguismo)⁸. Allo stesso tempo, però, il modello di Regis supera l’opposizione berrutiana, riuscendo a cogliere e descrivere indifferentemente sia forme di contatto ormai stabilizzatesi nel sistema (vale a dire nella *langue*, come nel caso di prestiti per influsso adstratico e, quindi, di lingue *a* contatto), che fenomeni relativi al discorso (o meglio, all’uso e alla *parole*, ad esempio *code-switching*, *code-mixing* e fenomeni di trasferenza, manifestazioni prototipiche di un contesto di lingue *in* contatto). Inoltre, il modello permette di analizzare esiti derivanti non soltanto da atti linguistici reali, ma anche, come nel nostro caso, da output ricavati tramite specifiche pratiche elicitative.

La teoria della trasferenza rappresenta la cornice all’interno della quale è stato elaborato uno schema interpretativo formalizzato in XSD (*XML Schema Definition*)⁹, creato *ad hoc* per l’etichettatura dei dati della Domanda I:

⁷ È utile precisare che lo studioso mostra una visione piuttosto “larga” dell’alternanza, considerata come l’uso separato di due o più codici con conseguenti fenomeni di commutazione interfrasale (tra frasi e/o proposizioni) o plurisintagmatica.

⁸ Cf. anche BERRUTO – BURGER 1985.

⁹ Il modello XSD di etichettatura, a cui ho personalmente apportato dei piccoli aggiustamenti in base a specifiche esigenze di ricerca e analisi, è stato realizzato sotto la guida di Vito Matranga.

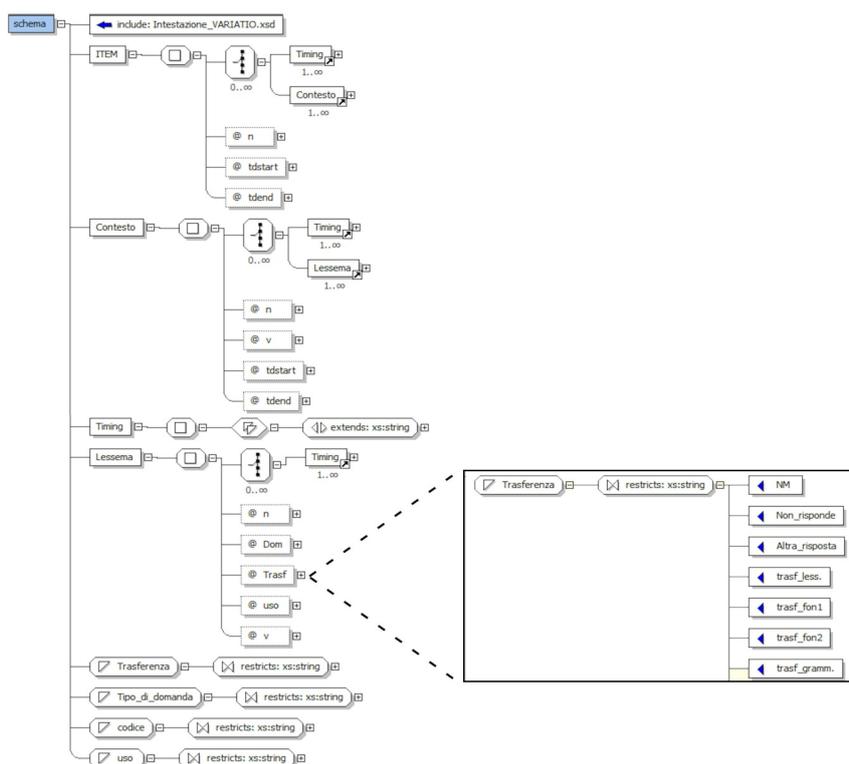


Figura 2. Modello XSD per l’etichettatura del quesito onomasiologico

Il modello di etichettatura (Figura 2), seguendo un ordine gerarchico, isola l’intero processo elicittivo che contiene il dato interessato dall’analisi (livello ITEM a cui, come illustrato in § 2, Tabella 1, si associa un numero in base all’ordine previsto nel questionario che si ripete anche nei livelli successivi), per poi separare il contesto frasale in cui è inserito l’item richiesto (livello *Contesto*). Infine, si passa al *Lessema*, con un ventaglio più eterogeneo di possibilità di etichettatura: tipo di domanda (diretta o indiretta, in base a possibili suggerimenti o interventi da parte del raccogliatore)¹⁰, *trasferenza*¹¹ (in cui, oltre alle varie tipologie di contatto, è

¹⁰ In molti casi, i raccoglitori impiegano strategie di contestualizzazione di varia natura (mimando l’azione, descrivendo il referente o suggerendolo in italiano in serie siciliana o viceversa in serie dialettale) allo scopo di facilitare e velocizzare il processo cognitivo di riconoscimento dell’immagine, col rischio, però, di trasformare a volte la prova onomasiologica in un quesito traduttivo.

¹¹ Tra le possibilità di annotazione del tipo di *trasferenza* è stata apportato un piccolo cambiamento rispetto alle categorie formalizzate da REGIS (2013), con l’intento di considerare, inoltre, i tratti fonetico-fonologici tipici dell’isola o delle aree analizzate, riconducibili a una pronuncia micro- o macro-regionale (ad esempio, fenomeni di raddoppiamento intrinseco come nel caso di *bbambola*, *bbasilico*, *rragnatela*; deaffricazione dell’affricata palatale in area occidentale e orientale come in *çiliegie*; sonorizzazione delle consonanti sorde in posizione post-nasale o intervocalica diffusa in diverse aree dell’isola, etc.). Tali fenomeni (evidenziati in serie italiana) sono stati definiti come

possibile annotare se l’informatore non risponde o dà una risposta non pertinente), uso (dichiarazione d’uso in serie siciliana, ma solo nel caso in cui venga elicitata la forma attesa)¹² ed eventuali cambiamenti nella morfologia rispetto al *target* richiesto (livello *v.*, ad esempio, forma singolare al posto di quella plurale, verbi al gerundio anziché all’infinito, etc.).

I frammenti dell’intervista (dall’ITEM al *Lessema*) presentano il proprio *timing* di inizio e di fine (*tdstart* e *tdend*), ovvero la propria cronologia rispetto al file audio da cui sono estrapolati e a cui possono essere così ricondotti¹³.

2.2. Tagging e interrogazione

Come già specificato, una volta conclusa la fase di allineamento, il file generato dal programma viene convertito per essere lavorato secondo i criteri inseriti nello schema XSD di riferimento. Proponiamo di seguito due esempi (uno per ciascuna serie), per illustrare il processo di *mark-up*:

Serie italiana

- (1) <ITEM n="31" tdstart="2424220138" tdend="2610849905">
<Timing tdstart="2424220138" tdend="2453408083">R1: /// questo di qua signora lo sa che cos'è? ((R mostra immagine))</Timing>
<Timing tdstart="2453408083" tdend="2471046728">I2: sì </Timing>
<Timing tdstart="2471046728" tdend="2495057943">R3: questo l'ha presente | [per:] I4: [sì]</Timing>
<Timing tdstart="2495057943" tdend="2523396213">I4: sì: però questo in it+ it+ R5: sì sì</Timing>
<Timing tdstart="2523396213" tdend="2557147233">R5: prima in italiano eh::

trasferenza fonetica di tipo 2 (**trasf_fon2**) per distinguerla, per l’appunto, dalla trasferimento fonetica propriamente detta, in cui il passaggio di tratti interessa fenomeni più chiaramente legati al contatto con il dialetto, con conseguenti risvolti di ordine sociolinguistico.

¹² In questa circostanza, ci muoviamo su un terreno poco agevole, basato su dati dipendenti dall’intuizione dell’informatore che possono anche non corrispondere con quanto dichiarato. In tal senso, risulta paradigmatico l’aneddoto di LABOV (1975, 106-107) riguardo a un informatore di una certa varietà di inglese americano che dichiarava di non aver mai sentito un particolare tratto bandiera della sua parlata. Tuttavia, i ricercatori, ascoltando di nascosto l’intervistato, si accorsero che quest’ultimo, nonostante le sue dichiarazioni, mostrava il fenomeno indagato.

¹³ Occorre precisare che alcuni campi della presente scheda di annotazione sono stati ripresi e rielaborati dal vecchio protocollo di processazione dei dati proposto e descritto in CASTIGLIONE *et al.* (2006, 68).

```
// in italiano?</Timing>
<Timing tdstart="2557147233" tdend="2571245561">I6: // cerniera?</Timing>
<Timing tdstart="2571245561" tdend="2576422291">no</Timing>
<Timing tdstart="2576422291" tdend="2596122187">R7: // lei come lo chiama
questo in italiano?</Timing>
<Timing tdstart="2596122187" tdend="2602903861">[italiano?] I8:
[crivu]</Timing>
<Contesto n="31" tdstart="2602903861" tdend="2610849905">
  <Lessema n="31" Dom="indir" Trsf="trasf_less.">
<Timing tdstart="2602903861" tdend="2610849905">crivo</Timing>
  </Lessema>
</Contesto>
</ITEM>
```

(Antonia Z., 65 anni, nonna III famiglia, istruzione media/bassa, Palermo)

Nell'esempio (1) viene richiesto in serie italiana l'item SETACCIO (referente n. 31 in base alla numerazione prevista dal questionario; § 2, Tabella 1) in cui l'informatrice tenta di "costruire" la forma italiana, in un primo momento, dal verbo siciliano che designa l'azione di *setacciare* (*cèrniri*, da cui la forma *cerniera*), per poi restituire l'unico termine conosciuto, corrispondente alla forma dialettale (*crivu*) morfologicamente italianizzata (con -o finale)¹⁴. Considerati gli interventi del raccogliitore nel guidare l'informatrice nella sua formulazione ultima, nel tag del livello *Lessema* viene segnalato che si tratta di una domanda indiretta (**Dom="indir"**). La forma elicitata rappresenta, quindi, un chiaro esempio di trasferimento lessicale dal dialetto all'italiano¹⁵, così come viene indicato nella specifica annotazione (**Trsf="trasf_less."**).

¹⁴ Invece, forme come *trottola* o *carciofa*, che in serie italiana mostrano l'influenza di tratti fonetici e *patter* grammaticali del dialetto (la retroflessione del nesso /tr-/ , nel primo caso, e il metaplasmo di genere sul modello del siciliano *cacòcciula*, nel secondo), verrebbero etichettate rispettivamente come fenomeni di TF1 e TG. Lo stesso vale in serie siciliana per esiti come *taddarita* per *tađdarita* (pipistrello) e *cirase* al posto di *cirasi* (ciliegie), con perdita rispettivamente dell'occlusiva retroflessa e del morfema grammaticale -i, etichettati ugualmente come TF1 e TG. Nel caso invece della presenza di più fenomeni di trasferimento all'interno dello stesso lessema, il modello gerarchico-implicazionale di Regis (§ 2.1, Schema 2) indica chiaramente quale tipo di contatto "preceda" l'altro. Ad esempio, esiti come *štracciu* (in luogo di 'straccio') in serie italiana e *àstraco* (per *àštracu*, terrazzo) in serie siciliana, in cui si osserva il trasferimento di materiale sia fonetico che morfologico dall'altro codice, verranno classificati entrambi come fenomeni di TF1 perché il passaggio di tratti fonetici anticipa quello morfologico.

¹⁵ Come è possibile notare, nel momento in cui l'informatore fornisce più risposte, viene selezionata quella più vicina al *target* richiesto rispetto alla serie. Invece, nel caso di più varianti "trasferite" dello stesso item, viene scelta generalmente l'ultima forma restituita.

Come desumibile dal processo di etichettatura appena considerato, nell’economia di un’analisi che permetta la comparabilità di dati fortemente eterogeni e difficilmente riconducibili a categorie comuni, il modello di Regis privilegia una prospettiva basata principalmente sul passaggio di materiale linguistico da un codice all’altro, “sacrificando” la rilevanza degli eventuali adattamenti al codice d’arrivo del materiale trasferito. Nel caso preso ad esempio importa che, nella serie italiana, il siciliano *crivu* (come tipo lessicale) passi all’italiano, indipendentemente dal fatto che la forma si conservi integralmente nella sua veste dialettale o che subisca un adattamento in morfologia finale (*criv+o*), giacché si tratta di un tipo lessicale non congruente con la serie.

Serie siciliana

- (2) <ITEM n="3" tstart="242582218" tend="359280793">
<Timing tstart="242582218" tend="284056776">R1: chista? /| chistu veramente
(R mostra immagine)</Timing>
<Contesto n="3" tstart="284056776" tend="293270194" v="9">
<Lessema n="3" Dom="NM" Trasf="NM" uso="sì" v="9">
<Timing tstart="284056776" tend="293270194">I2: i spìnguli </Timing>
</Lessema>
</Contesto>
<Timing tstart="293270194" tend="337423730">R3: sì. / e lei la usa sta parola
[spìngula?] I4: [sì sì]</Timing>
<Timing tstart="337423730" tend="359280793">riçemu noaṛi “pìgghiami ru
spìnguli ri ddocu”</Timing>
</ITEM>

(Gaetana C., 65 anni, nonna III famiglia, istruzione media/bassa, Valderice, Custonaci,
S. Vito Lo Capo - TP)

L’esempio (2) riporta invece il processo elicittivo in serie siciliana dell’item SPILLO (referente n. 3 del quesito onomasiologico), in cui viene restituita la forma attesa attraverso una domanda diretta e, quindi, non marcata (**Dom="NM"**), ma in una veste morfologica diversa rispetto a quella richiesta (l’informatrice, infatti, al posto del singolare *spìngula* fornisce la forma plurale *spìnguli*). Da qui il tag **v="9"** a livello *Contesto* e *Lessema*, al fine di rendere tale informazione recuperabile e, perfino, isolabile per una eventuale analisi di tale aspetto. In questo caso, viene anche chiesto l’uso dell’arcaismo (etichettato sotto il tag

uso="si"), momento in cui l'informatrice non solo fornisce una risposta positiva, ma inserisce anche l'item in un contesto pragmatico-esperenziale fittizio che "ricostruisce" in modo verosimile la circostanza della richiesta del referente (*riçemu noaṛi "pìgghiami ru spìnguli ri ddocu"*).

Concluso, quindi, il momento della "lavorazione" del dato, si passa alla creazione di una banca dati in cui poter effettuare una serie di interrogazioni a partire dai vari criteri e livelli di analisi formalizzati nello schema XSD:

Figura 3. Pannello di interrogazione della banca dati del quesito onomasiologico

Il sistema permette di interrogare la banca dati per più livelli di etichettatura (per *multiple tasks*), isolando o raggruppando particolari fenomeni linguistici e

correlandoli alle diverse caratteristiche socio-culturali del campione (Figura 3)¹⁶. Ad esempio, in questo caso viene chiesto di recuperare le forme attese del nonno di I famiglia di Palermo in serie siciliana. Una volta eseguita la *query*, il sistema recupera i dati richiesti:

The screenshot shows the 'ALSWeb - Risultati della ricerca III parte' interface. At the top left is the logo of the 'Istituto Nazionale per lo Studio e la Cura delle Alterazioni Linguistiche'. The main title is 'ALSWeb - Risultati della ricerca III parte'. Below the title, it indicates the database name: 'Nome database: Scaglione'. A section titled 'La query:' contains the search criteria: 'Cerca le occorrenze del tag Lessema in presenza dei seguenti vincoli: (tag "Lessema" in informatori con codintervista: 201-10-01 in informatori di localita: Palermo con idinformatore: 01 con codicerichiesto: siciliano) AND (tag "Lessema" con proprietà "Trasf" uguale a "NM" in informatori con codintervista: 201-10-01 in informatori di localita: Palermo con idinformatore: 01 con codicerichiesto: siciliano)'. Below this, an 'Informazioni' section provides summary statistics: 'Tags Lessema trovati: 27 - Informatori trovati: 1' and 'Campioni parziali: Campione totale: 36 -- Occorrenze del tag Lessema con vincolo: tag "Lessema" in informatori con codintervista: 201-10-01 in informatori di localita: Palermo con idinformatore: 01 con codicerichiesto: siciliano' and '27 di 36 (75,0% PARZ. - 75,0% TOT.) -- Occorrenze del tag Lessema con vincolo: tag "Lessema" con proprietà "Trasf" uguale a "NM" in informatori con codintervista: 201-10-01 in informatori di localita: Palermo con idinformatore: 01 con codicerichiesto: siciliano'. A detailed record is shown with fields: 'Codice intervista: 201-10-01 | Id intervista: 693 | Località: Palermo | Fam. sog.: Nonno-a | Fam. tipo.: Tipo 1 | Sesso: Maschio | Etá: 72 | Livello istrusione: Elementare senza licenza | Professione: - | Tag Lessema trovati: 27'. Below this are four XML snippets for 'Lessema' instances, each with a 'Timing' tag indicating start and end times and a label like 'a pupa', 'a spingula', 'a ggiurana', and 'u bbasilicò'. At the bottom, there are five buttons: 'Salva risultato in pdf con tag', 'Salva risultato in pdf senza tag', 'Salva risultato in word con tag', 'Salva risultato in word senza tag', and 'Salva come concetto'.

Figura 4. Esito della *query*

¹⁶ Le caratteristiche diastratiche degli informatori insieme alle altre informazioni connesse al tipo di campionamento ALS (famiglia e tipologia generazionale; § 2, n. 4) vengono recuperate direttamente dal sistema in base al codice attribuito all'intervista, lo stesso presente nel file XML etichettato (§ 2.1).

In Figura 4 viene riportato l'esito della *query* che computa il numero e la percentuale delle forme attese restituite dall'informatore rispetto al totale, con la possibilità di ascoltare, grazie al processo di allineamento preliminare, anche il singolo item per un ulteriore controllo. Inoltre, al di là del grado più basso di interrogazione, come nel caso del *Lessema*, cambiando i criteri della *query*, è possibile soffermarsi a un livello "macro" o "intermedio", recuperando e riascoltando l'intero processo elicittivo (livello ITEM), oppure considerare, laddove disponibile¹⁷, il contesto frasale in cui è inserito il dato. In ogni caso, qualunque siano i parametri inseriti, ritroviamo quel processo circolare evidenziato in § 2.1 in cui intervista, dato e analisi risultano sempre strettamente legati da un rapporto interdipendente. Inoltre, attraverso un più fitto e puntuale popolamento della banca dati, sarà possibile analizzare i vari fenomeni indagati secondo una prospettiva geolinguistica, comparando i dati e la loro variazione in funzione dei punti della rete d'inchiesta, e leggerli (e/o incrociarli) anche in base ai parametri diastratici e diagenetici previsti dal campionamento (§ 2, n. 4).

3. *Ciò che resta: quando i tag non bastano*

Il sistema di informatizzazione, analisi e consultazione dei dati qui presentato tenta di rispondere all'esigenza della creazione di un *database* il cui perno è (e resta) l'intervista, attorno alla quale, come un giro di compasso, si costruisce il sistema di "lavorazione" e interrogazione dei dati. Ma partendo proprio da questa esperienza per spostarci verso considerazioni di ambito più generale, sembra giusto chiedersi: i nuovi software e sistemi di immagazzinamento e processazione (XML e non solo) riescono a cogliere fino in fondo tutti gli aspetti che compongono e accompagnano il dato o qualcosa viene inevitabilmente perso, restando in una posizione marginale o addirittura fuori da qualsiasi *tag*? Come in parte evidenziato nel sistema di etichettatura del quesito onomasiologico, dinnanzi alla possibilità di utilizzo di particolari ausili informatici, il linguista si trova più che mai a dover effettuare a monte delle scelte ben precise, focalizzandosi ancor di più su aspetti specifici (a scapito inevitabilmente di altri), in una visione del dato molto spesso atomistica e frammentaria¹⁸. Pertanto, nonostante gli

¹⁷ In molti casi, infatti, contesto frasale e *lessema* coincidono, secondo una successione di "botta e risposta", in cui l'informatore si limita a fornire semplicemente l'item richiesto.

¹⁸ Ciò non significa che i software e gli schemi di analisi XML non siano flessibili e, quindi, essere integrati con l'inserimento di ulteriori variabili in base a nuove esigenze di ricerca. Tuttavia, tale evenienza comporta molto spesso la necessità di passare in rassegna l'intero corpus per includere

innumerevoli strumenti informatici a disposizione, questi non riusciranno mai a “fissare”, come lo spillo dell’entomologo, il dato in modo definitivo: la sua imprevedibilità – riflesso diretto dell’imprevedibilità del campione e/o degli informatori e delle circostanze elicitative – resta quell’incognita che accompagna (e accompagnerà) la ricerca linguistica, cambiando o molto spesso stravolgendo in corso d’opera prospettive teoriche e di analisi.

Quanto appena detto risulta ancora più vero all’interno di un progetto come l’ALS, in cui ogni intervista rappresenta un “microcosmo” ricorsivo nella struttura, ma, allo stesso tempo, reso unico dall’interazione tra intervistatore e intervistato, dalla(-e) personalità (spesso in conflitto) e dal vissuto degli informatori, e ancora dal desiderio di voler andar oltre le domande per raccontare una frammento di storia di vita e di “lingua”. Infatti, nelle varie parti (biografica, percettivo-metalinguistica e linguistica) che caratterizzano il questionario sociovariazionale dell’ALS, aspetti (apparentemente) secondari collegabili a ciò che Jaberg e Jud definiscono come “atmosfera dell’intervista”, corredano in molti casi il dato di quel plusvalore informativo-interpretativo di supporto per la motivazione e spiegazione di particolari esiti linguistici. In un’altra sede, partendo sempre dal quesito onomasiologico e addentrandoci nell’“intercapedine” tra input e output linguistico, abbiamo avuto l’occasione di mettere a fuoco il rapporto tra dialetto ed emozioni, fattore che in molti casi giustifica la dichiarazione d’uso dell’item elicitato in serie siciliana (cf. Scaglione 2016). Ma al di là e accanto ad aspetti emotivi, è possibile cogliere interessanti processi di recupero e ipotesi di dialettalità o “italianità” (§ 2.2, esempio 1) e ulteriori importanti informazioni difficilmente riconducibili a specifici *tag* o schemi prefissati:

(3) (R mostra immagine)

I1: pantaloni

R2: in siciliano.

I3: i | i pantaloni: in siciliano, *pìgghia i cà’usi*.

R4: ok, perfetto! / *e sta parola a usa? a parola càusi?*

I5: // non tanto *cchiù*. no, non sono mai || mi ada+ || *comu si diçi?*

R6: si adatta all’italiano.

quel particolare aspetto all’interno dello studio, con un conseguente prolungamento in termini di tempo dell’analisi del dato.

I7: appunto più che | più che mai. *sempri idda parra: tutta:* in italiano peccìò (sic) mi sono adattata a loro *picchi*

R: certo

I7: sono stata quasi dodici anni con mia nuora peccìò (sic) più o meno so quasi || *sugnu mmettà cu mmettà*

(*Carmela C., 80 anni, nonna I famiglia, istruzione bassa, Catania*)

(4) (R mostra immagine)

I1: eh:: *u stracciu*

R2: solo così?

I3: sì, così // *u pagghiazzu*

R4: ah:: / *e sta parola a usi?*

I5: eh:: // sì, per insultare.

(*Roberto B., 22 anni, figlio I famiglia, istruzione bassa, Catania*)

Nell'esempio (3) la nonna catanese non motiva l'abbandono dell'uso dell'arcaismo siciliano (e del dialetto in genere) per una mancata competenza o una dimenticanza del termine, ma a partire dal nuovo nucleo familiare in cui vive, caratterizzato da un prevalente uso del codice lingua da parte dei membri che lo compongono (in modo particolare, dalla nuora). Pertanto, vivere in un contesto pressoché italofono ha comportato un (ri)orientamento nelle scelte linguistiche che si riflette inevitabilmente anche a livello identitario: l'informatrice mostra ed è consapevole di possedere un'identità ormai scissa, a metà strada tra dialetto e italiano (*sugnu mmettà cu mmettà*), diretta espressione di eventi che lasciano un segno anche nella sua biografia linguistica. Nell'esempio (4) il giovane del capoluogo etneo, specifica il contesto d'uso del termine dialettale che sembra aver perso il suo valore denotativo, in favore di un significato figurato che si adatta perfettamente alla sfera dell'insulto. Pertanto, l'utilizzo dell'item non rientra del tutto in una prassi prototipicamente quotidiana, ma pare aver subito un processo di specializzazione che non sembra rassicurarci fino in fondo sulla piena vitalità del termine¹⁹.

Sebbene i fenomeni appena evidenziati non sembrano entrare a pieno nell'interesse della linguistica dei corpora e soprattutto di una ricerca su basi quantitative, tali fattori (nel nostro caso recuperabili, seppur in modo non sistematico) potrebbero tracciare e/o confermare possibili vie interpretative e di analisi, contestualizzando e, soprattutto, precisando (come nel caso dell'esempio

¹⁹ Ovviamente al momento dell'etichettatura l'intero frammento viene isolato e, quindi, anche queste parti dell'intervista possono essere recuperate ed eventualmente riascoltate. Sta di fatto, però, che questi, come tanti altri momenti del processo elicittivo, proprio per la loro ineffabilità, risultano difficilmente riconducibili a chiari schemi analitico-interpretativi.

4) il valore del dato. Quest’ultimo, infatti, all’interno dei sistemi di processazione e *computing* tende molto spesso a perdere, in condizioni e misura diverse, la sua “concretezza”, diventando una semplice unità quantitativa avulsa da quell’aspetto qualitativo che lo accompagna e determina.

Pertanto, come dimostrato, nonostante gli inestimabili vantaggi, i supporti informatici non rappresentano la soluzione definitiva in grado di poter canalizzare quella intrinseca eterogeneità e imprevedibilità del dato, che nessun *tag* riuscirebbe mai a “imbrigliare” sotto categorie predeterminate. Al linguista, quindi, il compito di non perdere mai il contatto diretto col dato, guardando dentro e al di là dell’“etichetta” che lo circonda.

Riferimenti bibliografici

BALI 39 2015

«Bollettino dell'Atlante Linguistico italiano», 39.

BERRUTO 2009

G. Berruto, *Confini tra sistemi, fenomenologia del contatto linguistico e modelli del code switching*, in G. Iannaccaro, V. Matera (a cura di), *La lingua come cultura*, Torino, 3-34.

BERRUTO – BURGER 1985

G. Berruto, H. Burger, *Aspetti del contatto fra italiano e tedesco in Ticino*, «Archivio Storico Ticinese» 101, 29-76.

CARLETTA *et al.* 2005

J. Carletta, D. McKelvie, A. Isard, A. Mengel, M. Klein, M. B. Møller, *A generic approach to software support for linguistic annotation using XML*, in G. Sampson, D. McCarthy (eds.), *Corpus Linguistics: Readings in a Widening Discipline*, London, 449-59.

CASTIGLIONE *et al.* 2006

M. Castiglione *et al.*, *Tecniche vecchie, obiettivi nuovi: i quesiti lessicali*, in M. D'Agostino, G. Paternostro (a cura di), *Costruendo i dati. Metodi di raccolta, revisione e organizzazione della banca dati nella sezione sociovariazionale*, Palermo, Centro di studi filologici e linguistici siciliani, 63-86.

CUNNINGHAM 1999

H. Cunningham, *A definition and short history of Language Engineering*, «Natural Language Engineering» V, 1-16.

D'AGOSTINO 2002

M. D'Agostino, *Reti familiari in Sicilia: dinamiche urbane e rurali*, in T. Krefeld (ed.), *Vissuto e dinamiche linguistiche. Varietà meridionali in Italia e in situazione di extraterritorialità*, Berlin, 47-60.

D'AGOSTINO 2006

M. D'Agostino, *Fra ricerca empirica e storiografia. Modelli di spazio in linguistica*, in T. Krefeld (a cura di), *Modellando lo spazio in prospettiva linguistica*, Frankfurt a.M., 35-71.

D'AGOSTINO – PATERNOSTRO 2006

M. D'Agostino, G. Paternostro, *Questionario, parlanti, spazio linguistico: l'esplosione del dato*, in M. D'Agostino, G. Paternostro (a cura di), *Costruendo i dati. Metodi di raccolta, revisione e organizzazione della banca dati nella sezione sociovariazionale*, Palermo, Centro di studi filologici e linguistici siciliani, 25-44.

D'AGOSTINO – PENNISI 1995

M. D'Agostino, A. Pennisi, *Per una linguistica spaziale. Modelli e rappresentazione della variabilità linguistica nell'esperienza dell'ALS*, Palermo, Centro di studi filologici e linguistici siciliani.

D’AGOSTINO – RUFFINO 2002

M. D’Agostino, G. Ruffino, *Questionario. Atlante Linguistico della Sicilia, Sezione Sociovariazionale*, Palermo, Centro di studi filologici e linguistici siciliani.

D’AGOSTINO – RUFFINO 2005

M. D’Agostino, G. Ruffino, *I rilevamenti sociovariazionali. Linee progettuali*, Palermo, Centro di studi filologici e linguistici siciliani.

KENNEDY 1998

G. Kennedy Graeme, *An Introduction to Corpus Linguistics*, London-New York.

KREFELD – LÜCKE 2014

T. Krefeld, S. Lücke, *Quando l’etnolinguistica alpina va online: alcune esigenze metodologiche*, in *La géolinguistique dans les Alpes au XXIe siècle. Méthodes, défis et perspectives, Actes de la conférence annuelle sur l’activité scientifique du centre d’études francoprovençales*, Aoste, Région Autonome de la Vallée d’Aoste, 39-51.

LABOV 1975

W. Labov, *Empirical foundations of linguistic theory*, in R. Austerlitz (ed.), *The Scope of American Linguistics*, Lisse, 77-132.

MATRANGA 2007

V. Matranga, *Trascrivere. La rappresentazione del parlato nell’esperienza dell’Atlante Linguistico della Sicilia*, Palermo, Centro di studi filologici e linguistici siciliani.

MATRANGA 2014

V. Matranga, *Dal parlato alla parola. Procedura informatiche per il trattamento delle interviste etnodialettali nell’Atlante Linguistico della Sicilia (ALS)*, in F. Cugno, L. Mantovani, M. Rivoira, S. Specchia (a cura di), *Studi linguistici in onore di Lorenzo Massobrio*, Torino, Istituto dell’Atlante Linguistico Italiano, 639-50.

MATRANGA – SOTTILE 2007

V. Matranga, R. Sottile, *Atlante Linguistico della Sicilia. Aggiornamenti*, V. Matranga e R. Sottile (a cura di), *Esperienze geolinguistiche. Percorsi di ricerca italiani e europei*, Palermo, Centro di studi filologici e linguistici siciliani, 199-204.

MOCCIARO 1989

A.G. Mocciano, *Due inchieste a confronto a Mandanici (Punto 819 dell’AIS): considerazioni sull’innovazione linguistica*, «RID» XII, 161-78.

REGIS 2013

R. Regis, *Contatto linguistico, linguistica del contatto: aspetti di modellizzazione*, «Studi Italiani di Linguistica Teorica e Applicata» I, 7-40.

ROMAINE 2008

S. Romaine, *Corpus linguistics and sociolinguistics*, in A. Lüdeling, M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, Berlin, 96-111.

RUFFINO 1991

G. Ruffino, *Questionario. Atlante Linguistico della Sicilia, Sezione Sociovariazionale. Inchieste di prova 1991-'92*, Palermo, Centro di studi filologici e linguistici siciliani.

RUFFINO 1995

G. Ruffino (a cura di), *Percorsi di geografia linguistica. Idee per un atlante siciliano della cultura dialettale e dell'italiano regionale*, Palermo, Centro di studi filologici e linguistici siciliani.

RUFFINO 1997

G. Ruffino, *I nomi della trottola e la memoria del gioco*, Palermo, Centro di studi filologici e linguistici siciliani.

SCAGLIONE 2016

F. Scaglione, *Dialetto ed "emozioni": alcuni aspetti descrittivi*, in Gruppo di ricerca dell'*Atlante Linguistico della Sicilia* (a cura di), *La linguistica in campo. Scritti per Mari D'Agostino*, Alessandria, 183-93.

SOTTILE – CAPITUMMINO 2010

R. Sottile, E. Capitummino, *Il contatto lingua-dialetto nei dati lessicali dell'ALS*, in G. Marcato (a cura di), *Le nuove forme del dialetto*, Padova, 277-82.

SOTTILE – PATERNOSTRO 2012

R. Sottile, G. Paternostro, *La variazione lessicale nella Sicilia contemporanea. Italiano e siciliano nei quesiti onomasiologici dell'Atlante Linguistico della Sicilia (ALS)*, in P. Bianchi, N. De Blasi, C. De Caprio, F. Montuori (a cura di), *La variazione nell'italiano e nella sua storia. Varietà e varianti linguistiche e testuali*, Atti dell'XI Congresso SILFI, Napoli 5-7 ottobre 2010, Firenze, 811-23.

TISATO 2015

G. Tisato, *Documenti etnolinguistici navigabili e parlanti: l'approccio di NavigAIS e dell'AMDV*, «Bollettino dell'Atlante Linguistico Italiano» 39, 59-82.

THOMASON – KAUFMAN 1988

S.G. Thomason, T. Kaufman, *Language Contact, Creolization, and Genetic Linguistics*, Berkeley-Los Angeles-London.

TROPEA 1991

G. Tropea, *Su alcuni aspetti dell'italianizzazione lessicale in Sicilia*, in V. Orioles (a cura di), *Innovazione e conservazione nelle lingue*, Atti del Convegno della Società Italiana di Glottologia, Messina, 9-11 novembre 1989, Agnano Pisano-Pisa, 171-99.

WEINREICH 1953

U. Weinreich, *Languages in contact. Findings and problems*, New York.

Sitografia

www.verba-alpina.gwi.uni-muenchen.de/it/ (data ultima consultazione 10/04/2017)

www3.pd.ist.cnr.it/navigais-web/ (data ultima consultazione 10/04/2017)