

# Carmelo Lupini

## Raccolta dei materiali, trascrizione fonetica ed elaborazione dei dati acquisiti

La registrazione e la trascrizione fonetica particolareggiata costituiscono la prima ed indispensabile tappa di ogni analisi linguistica, tanto più se la lingua studiata è priva di tradizione scritta.

Pur essendo l'orecchio lo strumento naturale che lo studioso ha a disposizione in questo dominio, le tecniche di registrazione e di analisi del segnale verbale, sia analogiche che digitali, gli permettono di controllare le sue impressioni e si dimostrano indispensabili nell'identificazione dei tratti pertinenti i vari suoni che compongono una determinata lingua, siano essi descritti in termini acustici o percettivi.

La fonetica occupa un ruolo particolare nello studio del linguaggio, poiché è ormai acquisito che qualsiasi descrizione linguistica debba accordare la priorità alla lingua parlata; per questo motivo l'unico riferimento sicuro, ad di là dei vari sistemi di trascrizione, resta sempre la registrazione della viva voce del parlante.

La *Fonetica Acustica* ha cominciato a svilupparsi, pur con mezzi strumentali di scarsa precisione e di difficile utilizzazione, nella seconda metà del XIX secolo.

Nel 1863, il fisico H. von Helmholtz pone le basi dell'acustica e in particolare della descrizione dei suoni vocalici, seguito su questa strada, da fonetisti quali P. Rousselot, H. Pipping e L. Herman. La Fonetica Acustica progredisce tuttavia molto lentamente, anche dopo l'invenzione da parte di T. Edison (1878) del fonografo a cilindro.

Finalmente l'uso di dischi, registratori magnetici, nastri e cassette permetterà non solo di acquisire il linguaggio con maggiore precisione e rapidità, ma anche di archiviarlo e di classificarlo più agevolmente; tutto questo per dare una soluzione efficace alla necessità di disporre di "testi acustici" (con relativa trascrizione) allo scopo di documentare *in perpetuum* le fasi di evoluzione fonetica, morfosintattica e lessicale della "parlata esplorata".

Oggi la possibilità dell'acquisizione computerizzata del prodotto vocale e della sua analisi ha permesso eccezionali passi in avanti per ciò che concerne l'elaborazione, l'archiviazione e la manipolazione del materiale acquisito.

L'utilizzazione del calcolatore come strumento potente di memorizzazione, reperimento e manipolazione dell'informazione, permette di compiere, con maggiore velocità ed efficienza, quelle operazioni che, comunque, il linguista, più specificamente il lessicografo, compiva manualmente. I lessicografi, fin dal tempo del primo Vocabolario della Crusca, hanno fondato il loro lavoro sulla lettura di testi, la schedatura dei passi più rilevanti, l'estrazione delle parole significative, l'identificazione dei contesti e degli esempi utili alla costruzione di un lessico. Queste operazioni possono essere compiute molto più efficientemente utilizzando un calcolatore, che ha capacità vastissime di memoria e infallibile regolarità nelle operazioni ripetitive. I programmi, le tecniche messe a punto e quelle che ancora vengono studiate trovano primaria applicazione nella lessicografia o, più generalmente, nel trattamento automatico dei testi.

Una delle operazioni automatiche per il trattamento di un testo acquisito è il *parsing*; il programma che segue tali operazioni è detto *parser*, la sua funzione è quella di individuare e separare i documenti, nonché di estrarre dall'interno di essi il titolo, le parole e altri componenti per poterne ricavare, ad esempio, un lessico ragionato ed evidenziare automaticamente i rapporti tra le varie parti del testo. Si tratta, infatti, di un'applicazione che legge una frase (o un testo, una frase alla volta) e, consultando una grammatica opportunamente definita e scritta in un formalismo specifico, restituisce in *output* la struttura sintattica della frase. Comunque, non è questa la sede per affrontare problemi specifici di linguistica computazionale e mi limiterò a suggerire semplici tecniche d'utilizzo del calcolatore alla portata di chi ne abbia una minima conoscenza.

A questo punto è d'obbligo una precisazione: applicare l'uso del calcolatore al trattamento del linguaggio, come a qualunque altro problema, non significa conoscere alla perfezione questo o quel calcolatore, o una decina di linguaggi di programmazione, ma, prima di tutto e soprattutto, saper "pensare come un calcolatore", cioè imparare il metodo di analisi e scomposizione di un problema in termini tali che sia possibile, poi, costruire una sequenza di operazioni semplici (istruzioni), che offrono una soluzione al problema stesso. Solo in una fase successiva, quando occorre tradurre questa sequenza di istruzioni in un programma, occorrerà la conoscenza pratica di un calcolatore e di un linguaggio di programmazione. Questo, però, è un passo tecnico marginale rispetto a quello precedente molto più importante e qualificante.

Disporre di un computer portatile con microfono per la raccolta dei materiali linguistici è oggi un grande vantaggio alla portata di tutti per svariati motivi: in primo luogo è possibile digitalizzare fedelmente e archiviare subito grandi quantità di dati, in secondo luogo i dati acquisiti sono facilmente classificabili e possono essere già organizzati in un piccolo archivio elettronico; se si desidera invece conservare i dati acquisiti col computer in supporti tradizionali, è possibile riversare i dati su nastro magnetico attraverso un comune cavo di collegamento tra il magnetofono e la scheda audio del computer.

L'altro grande vantaggio dell'acquisizione computerizzata consiste nella possibilità di intervenire subito sul materiale acustico per ottimizzarlo e soprattutto per eliminarne gli eventuali disturbi di sottofondo o i "rumori" che possono pregiudicarne la qualità.

I programmi disponibili, inoltre, permettono di visualizzare lo spettrogramma del dato acustico, quindi di osservare subito graficamente la frequenza (in verticale), la durata (in orizzontale) e l'ampiezza (annerimento, maggiore o minore, di alcune parti dello spettrogramma).

Ci sono due tipi di spettrogrammi ottenibili: quello a banda larga, più adatto per l'analisi dei singoli elementi fonici, e quello a banda stretta, più utile per analizzare certe caratteristiche, quali l'intonazione, il tono e, ovviamente, le caratteristiche individuali della voce.

Esistono applicazioni (Fig. 1) che permettono di trascrivere facilmente la voce dell'informatore e di associare la trascrizione di una singola parola alla sezione corrispondente sullo spettrogramma.



Fig. 1

Tutti i programmi di scrittura attualmente disponibili, peraltro in continua evoluzione, hanno un numero di funzioni di gran lunga maggiore del necessario e alcuni di questi si rivelano utilissimi per la manipolazione di un testo; lo strumento, che, ad esempio, permette di mettere in ordine alfabetico ogni singola parola, è un ottimo aiuto per ricavare una lista di parole da cui verranno eliminati facilmente i doppi; una volta che tutte le parole del testo in oggetto sono state analizzate e classificate a seconda della propria categoria morfologica, si ottiene in pratica il lessico di quel testo, con l'indicazione, per ogni lessema, della sua collocazione nel testo e del contesto nel quale

ricorre; in pratica si può ottenere un *index locorum* (o concordanza), sia delle forme che dei lemmi del testo.

Materialmente un testo scritto consiste di vari caratteri grafici, alfabetici, numerici e di punteggiatura. Poiché intendiamo ricercare informazioni sulle sequenze di caratteri alfabetici, ossia sulle “parole”, dobbiamo per prima cosa definire cosa è una parola; tra le varie definizioni proposte dagli studiosi ve ne è una che fa benissimo al caso nostro ed è quella di considerare parola “quello che è scritto tra due spazi o segni di interpunzione”. Qualcuno sorriderà a questa trovata, ma in questa sede bisogna accettare questa definizione, perchè è funzionale al modo di lavorare dei calcolatori, i quali basano il loro funzionamento proprio su questa capacità: riuscire a riconoscere segni delimitati da spazi bianchi come “parole”.

Questa definizione parte da un preconcetto di fondo: è formulata partendo da un testo in forma scritta. Va bene, quindi, quando si tratta di insegnare ad un programma come riconoscere una “parola”; non è che un modo per individuare le parti che ci interessano appunto in quanto parole, ossia corrispondenze (secondo alcuni arbitrarie) tra segno e significato. Questa distinzione non funzionerebbe difatti nella lingua parlata; le pause di un enunciato espresso oralmente non sempre corrispondono agli spazi bianchi dello stesso enunciato messo per iscritto. Difatti, nel fluire del discorso le parole, dal punto di vista fonico, risultano modificate e condizionate nella pronuncia e nell'accentazione (vedi per es. i fenomeni di agglutinazione dell'articolo: *lastrico* da *astracum*). Questa definizione si riferisce, quindi, alla parola in termini di unità grafologica, intesa come sequenza di sillabe e infine di lettere o fonemi.

Prendiamo le parole *amo* e *amavo*; benché ci rendiamo conto che si tratta di forme diverse, tuttavia le consideriamo come varianti della stessa parola *amare*, così come le forme *buoni* e *buone* sono varianti della stessa parola *buono*. Per tradizione in italiano, quando vogliamo indicare queste parole, usiamo la convenzione di indicare per i nomi il maschile o femminile singolare, per gli aggettivi il maschile singolare, per i verbi l'infinito, etc.; per altre lingue, invece, come ad esempio per il greco e il latino, usiamo la prima persona dell'indicativo presente singolare per indicare i verbi; per l'arabo, i dizionari sono organizzati secondo temi o radici. Chiameremo questo tipo di parole “lessemi”; il lessema è dunque l'unità fondamentale del lessico di una lingua. Le forme costituiscono le varianti dei lessemi e sono in genere parole autonome; vi sono anche casi di lessemi costituiti da più forme, per esempio le espressioni idiomatiche, che sono sequenze di parole fissate nell'uso e le cui parti non sono sostituibili.

Il lessema si realizza nel discorso come forma. Un lessema è già una forma; è la forma alla quale si fanno risalire le possibili varianti che si realizzano nel discorso; per analizzare le forme dal punto di vista strettamente morfologico, abbiamo bisogno del concetto di “base”, che è la parte della forma che ha il contenuto semantico. Le forme hanno una base, la base di un lessema nominale prenderà i morfemi che si possono usare coi nominali, la base di un lessema verbale prenderà i morfemi verbali, etc.

Il termine “lemma” appartiene invece alla lessicografia ed indica l'unità grammaticale, flessionale, semantica che figura come voce di un dizionario (per esempio la forma sing. per i nomi, la base verbale per i verbi).

Il lessico di un parlante è, dunque, la lista di lessemi dalla quale egli attinge per realizzare gli elementi della struttura enunciativa e accedere così al discorso. Il lessico di un testo o di un autore è l'elenco dei lessemi usati in quel testo o da quell'autore e l'operazione più elementare da compiersi con l'ausilio del computer è quella di ottenere automaticamente il lessico di un testo e, cosa molto interessante, di metterlo a confronto col lessico di un altro testo di uguale argomento, ma che fa capo ad un autore diverso, ad esempio le traduzioni della *Parabola del figliuol prodigo* da parte di due informatori diversi.

Una conoscenza seppur minima del linguaggio HTML, linguaggio diffuso in Internet ed estremamente compatibile coi tutti i moderni computers, ci permette di costruire con estrema facilità, e senza dover ricorrere a complicati programmi, degli utilissimi archivi linguistici, scritti ed acustici ad un tempo, nel senso che possiamo disporre, ad esempio, di un testo dialettale in formato

elettronico trascritto foneticamente, le cui singole parole (o gruppi di parole pronunciate senza soluzione di continuità) siano selezionabili per ottenerne immediatamente ogni informazione di natura morfologica, etimologica ed anche acustica.

Passo ora a trattare in breve alcuni dei problemi connessi con la trascrizione fonetica. Il problema della trascrizione è stato affrontato scientificamente verso la seconda metà dell'800; i sistemi di trascrizione che hanno avuto maggiore eco ed importanza sono quello del LEPSIUS<sup>9</sup> (1854) e quello dell'International Phonetic Association<sup>10</sup>. Essi segnarono un notevole passo avanti e si distaccarono dall'uso di far coincidere, nel caso di una lingua scritta, un grafema non latino ad un grafema latino.

Questo genere di trascrizione molto semplicistica non corrispondeva necessariamente alla vera pronuncia; infatti facendo corrispondere ogni carattere alfabetico non latino ad uno latino, potremmo trascrivere la parola neogreca ἀρνί 'agnello' *arní*, mentre, in effetti, la pronuncia è /ʎʎʎʎʎ/, o il russo xopoшo trascriverlo *horošo* mentre la pronuncia è /ʎʎʎʎʎʎʎ/.

Al sistema di trascrizione elaborato da LEPSIUS, seppur con diverse varianti, si ispirano i sistemi degli studiosi di lingue romanze e quelli in uso in dialettologia, messi a punto da ASCOLI e MERLO. Nella seguente tabella sono messi a confronto alcuni segni dell'Alfabeto Fonetico Internazionale (A), con alcuni di quelli di cui sopra e precisamente con:

- (B), quello usato da Gerhard ROHLFS nella sua *Grammatica storica della lingua italiana e dei suoi dialetti*;
- (C), quello usato da Gerhard ROHLFS nella sua grammatica storica dei dialetti italo-greci (Calabria, Salento);
- (D), quello usato da Giorgio PICCITTO in *Vocabolario Siciliano* (Catania-Palermo 1977);
- (E), quello dell'Istituto Siciliano di Studi Bizantini e Neoellenici «B. Lavagnini»;
- (F), quello usato da Pavao TEKAVČIĆ nella sua *Grammatica storica dell'Italiano*;
- (G), quello usato da Francesco AVOLIO in *Bommèsprə, Profilo linguistico dell'Italia centro-meridionale*;
- (H), quello generalmente usato per la trascrizione dei dialetti galloitalici di Sicilia;
- (I), quello usato da Fiorenzo TOSO in *Storia linguistica della Liguria*;
- (J), quello del Centro per gli Studi Dialettali Italiani (CDI).

	A	B	C	D	E	F	G	H	I	J
e, o chiusi	e, o	ɛ, ɔ				ɛ, ɔ			ɛ, ɔ	ɛ, ɔ
e, o aperti	ɛ, ɔ	ɛ, ɔ	é, ó		e, o	ɛ, ɔ			ɛ, ɔ	ɛ, ɔ
i, u semivocaliche	j, w	i, u	j		y, w	j, w		i, u	i, u	i, u
e indistinta	ʎʎ	ə		ə		ə	ə	ə	ë	ə
fricativa bilabiale sonora	β	β				β	β			ɸ
affricata mediopalatale sorda	ʎʎ	č	č		č	č	č	č	č	č
occlusiva postpalatale sorda	ʎʎʎ	č	č		č	č	č	č		č
fricativa dentale sonora	ð	ð	ð		ð	ð	ð		ð	ɸ
t, d invertite	ʎʎʎ	t, d	t, d	t, d	t, d		t, d	t, d		t, d
l, r invertite	ʎʎʎ	l, r	l, r		l, r		l, r	l, r		l, r
affricata mediopalat. sonora	ʎʎʎ	ǰ	ǰ		ǰ	ǰ	ǰ	ʎʎ	ǰ	ǰ
occlusiva postpalatale sonora	gʎ	ǰ	ǰ		ǰ	ǰ	ǰ	ǰ		ǰ
fricativa velare sonora	ʎʎ	ɣ	ɣ	ɣ			ɣ	ɣ		ɣ
h aspirata	h	h								
fricativa velare sorda	ʎʎ	ħ			χ	χ	χ			h

<sup>9</sup> Questo sistema si trova codificato nell'opera *Standard Alphabet for reducing unwritten Languages and foreign graphic Systems to a uniform Orthography in European Letters*, Williams & Nogate.

<sup>10</sup> International Phonetic Alphabet (I.P.A.).

l (pre)palatale	ʎʎ	l̥	l̥		l̥	l̥	l̥	ʎ	l̥	l̥
n (pre)palatale	ʎʎ	ñ̥	ñ̥		ñ̥	ñ̥	ñ̥	ñ̥	ñ̥	ñ̥
n velare	ʎʎ	ɲ̥		ɲ̥		ʎ			ɲ̥	ɲ̥
sibilante mediopalatale sorda	ʎʎ	š̥	š̥		š̥	š̥	š̥	š̥	š̥	š̥
fricativa (pre)palatale sorda	ʎʎ			ç̥i	ç̥	ç̥			ħ̥	é̥
sibilante dentale sorda	s	s̥	s̥	s̥	s̥	s̥			s̥	s̥
sibilante dentale sonora	z	z̥	z̥	z̥	z̥	z̥		f̥	z̥	f̥
affricata dentale sorda	ts	z̥	z̥		ts̥	ts̥			z̥	z̥
affricata dentale sonora	dz	z̥	z̥	z̥	dz̥	dz̥	z̥	z̥	z̥	z̥
sibilante mediopalat. sonora	ʎʎ	ž̥						f̥	ž̥	f̥
fricativa (pre)palatale sonora	ʎʎ			ʂ̥ʎ						ʂ̥ʎ
fricativa bilabiale sorda	ɸ	ɸ̥				ɸ̥	ɸ̥			ɸ̥
fricativa mediopalatale sorda	ʎʎ	χ̥	χ̥	hi						ħ̥
fricativa dentale sorda	θ	θ̥	θ̥		θ̥	θ̥	θ̥			θ̥

Questi ultimi sistemi di trascrizione hanno in comune alcuni principi: evitano il più possibile l'uso di lettere estranee all'alfabeto latino, tuttavia sono state introdotte alcune lettere greche, e precisamente  $\chi$ ,  $\gamma$ ,  $\theta$ ,  $\delta$ ,  $\varphi$ ,  $\beta$  che corrispondono rispettivamente ai caratteri IPA  $\chi$  /,  $\gamma$  /,  $\theta$  /,  $\delta$  /,  $\varphi$  /,  $\beta$  /. In dialettologia troviamo i segni  $f$  e  $z$  corrispondenti a quelli IPA /z/ e /dz/. Gli stessi sistemi si servono poi di diversi segni diacritici per modificare il valore di base rappresentato da un carattere singolo; ad esempio: la cosiddetta "pipa" (  $\sim$  ) generalmente indica un'articolazione palatale, il punto sottoscritto (  $\cdot$  ) ne indica una cacuminale, mentre il punto sovrascritto (  $\cdot$  ) una velare; inoltre spesso si usano tagliare orizzontalmente i caratteri delle occlusive sonore  $b$ ,  $d$ ,  $g$  con una barra (  $-$  ) per indicare le corrispondenti fricative in luogo di  $\beta$ ,  $\delta$  e  $\gamma$ .

Purtroppo questi sistemi grafici comportano alcune difficoltà, alcune di ordine pratico e altre di natura scientifica: i segni diacritici tendono a rallentare il ritmo della trascrizione con il rischio di essere tralasciati pregiudicando così il valore scientifico di un testo; per certi foni, poi, è spesso necessario l'uso di più di un segno diacritico, il che talvolta produce un accumulo di segni grottesco. Gli alfabeti usati dai romanisti e dai dialettologi, pur derivando da un modello comune, presentano diverse contraddizioni tra di loro; basta osservare la tabella comparativa per accorgersene: alcuni registrano il segno  $\chi$  come fricativa velare sorda (il ROHLFS in questo caso usa una  $h$  ), altri come fricativa mediopalatale sorda; il ROHLFS, nella sua *Grammatica storica della lingua italiana*, usa il segno  $\eta$  con valore velare mentre altrove, giustamente, ha valore cacuminale; la  $n$  palatale nei vari alfabeti risulta trascritta con diacritici diversi: a volte con la pipa, altre volte con un accento circonflesso o con un accento acuto. Non mancano casi in cui lo stesso studioso utilizzi un segno diverso per indicare lo stesso suono; il ROHLFS, sempre nella *Grammatica storica della lingua italiana*, utilizza  $\check{c}$  per indicare l'affricata postpalatale sorda, mentre nella *Grammatica storica dei dialetti italogreci* utilizza  $\check{c}$  per lo stesso suono che invece dovrebbe avere il valore di /tʃ/. Il TEKAŦČIĆ poi assegna due valori leggermente diversi alle varianti di uno stesso segno:  $\vartheta$  col valore fricativa dentale sonora e  $\theta$  col valore di fricativa dentale sorda.

Vi sono casi in cui vengono utilizzati gli stessi segni (sia graficamente che foneticamente) dell'Alfabeto Fonetico Internazionale; a parte l'utilizzo di  $\theta$ ,  $\varphi$ ,  $\beta$  segnalo la  $\gamma$  utilizzata dal TEKAŦČIĆ e la  $\gamma$  utilizzata spesso per indicare la  $l$  palatale nei dialetti galloitalici. Viceversa sono in uso segni graficamente identici a quelli dell'Alfabeto Fonetico Internazionale, ma con valore diverso: è il caso di  $\delta$  che in dialettologia indica la fricativa interdentale sonora, mentre nell'IPA sta ad indicare un'approssimante dentale sonora (la corrispettiva sorda è  $\vartheta$  ); e ancora  $f$ , che in dialettologia rappresenta una  $s$  sonora mentre nell'IPA ha il valore di fricativa mediopalatale sorda. Purtroppo quest'alternanza di coincidenze e discordanze (a volte solo grafiche, altre volte solo fonetiche), potrebbe rischiare di ingenerare a prima vista un po' di confusione nella corretta lettura

di un testo in trascrizione fonetica; anche se tuttavia è facile riconoscere una trascrizione in IPA da un'altra.

Indubbiamente il sistema IPA ha dei pregi innegabili, come la scarsità di diacritici, la presenza di nuovi segni che suggeriscono i suoni che essi rappresentano con la somiglianza alle lettere cui si ispirano ( $\text{ʔ} < n$ ,  $\text{ʔ} < r$ ), altri non sono che il capovolgimento di lettere latine ( $\text{ʔ} < y$ ,  $\text{ʔ} < h$ ) o antiche varianti grafiche di esse (“ʔ” di “s” e “ʔ” di “z”); il tutto costituisce un alfabeto di notevole eleganza ma con notevolissime difficoltà tipografiche e, a volte, di leggibilità.

Qualunque sistema di trascrizione non può che essere convenzionale e l'eccessiva accuratezza nella notazione potrebbe risultare di scarsa utilità pratica. L'alfabeto IPA, è vero, offre una vasta gamma di segni utili per trascrivere ogni lingua, ma nel momento in cui lo studio è limitato ad un certo gruppo di lingue, foneticamente più o meno affini, un segno del comune alfabeto latino può essere usato al posto di un più complesso simbolo dell'IPA, purché se ne definisca prima il valore e purché non vi sia opposizione fonematica: in francese graficamente possiamo usare sempre “r” al posto di “ʔ” giacché non v'è opposizione tra “r” e “ʔ”; la *d* intervocalica spagnola può essere trascritta semplicemente “d” invece che “ð” poiché tra vocali è facilmente prevedibile il suo valore di fricativa interdentale sonora. Sarebbe poi opportuno stabilire in quali casi scegliere una trascrizione analitica o una sintetica; alcuni esempi: il segno “č” si potrebbe utilizzare tranquillamente al posto di “tʃ” o “tš”, analogamente “ǰ” al posto di “dʒ” o “dž” e, se vogliamo, a seconda che fossero sentiti come pronunciati rispettivamente con minore o maggiore fusione<sup>11</sup>. In base allo stesso modo potremmo analizzare e scomporre graficamente il segno “c̣” (occlusiva postpalatale sorda) in “kχ”, dove “χ” in dialettologia indica la fricativa mediopalatale sorda, per questo il suono “c̣” può essere considerato una specie di affricato.

Le trascrizioni che si basano su questo criterio sono chiamate “larghe” e sono utilizzabili previa enunciazione di una regola di distribuzione degli allofoni. Sicuramente si prestano a questo principio i sistemi di trascrizione usati in dialettologia, però andrebbero superati tutti quei problemi relativi ad una loro certa disomogeneità e alle varianti, a volte ingiustificate, che si possono riscontrare nelle opere dei vari studiosi, per non parlare delle diverse denominazioni utilizzate per classificare un suono, denominazioni spesso in contraddizione con quelle usate da altri studiosi. Si noti, infatti, che, a proposito delle articolazioni palatali, i sistemi dei romanisti e dei dialettologi differiscono dal sistema IPA; essi infatti considerano i suoni “š” e “ž” come sibilanti mediopalatali e i suoni “č” e “ǰ” come occlusive postpalatali, mentre il sistema IPA le considera rispettivamente come fricative alveopalatali e affricate palatali.

Al di là dei diversi modi di denominare l'articolazione di un suono, è opportuno stabilire quale sia l'alfabeto fonetico più pratico. Quello del Centro per gli Studi Dialettali Italiani (CDI) risulta essere molto omogeneo e logico, in primo luogo esso risulta essere privo di caratteri estranei all'alfabeto latino, utilizza lo stesso segno (opportunamente accompagnato dai diacritici) per indicare tutti quei suoni che si producono nella stessa zona articolatoria e, infine, ogni diacritico ha sempre lo stesso valore per ogni carattere cui è associato.

Un ultimo problema che si presenta al momento della raccolta dei materiali consiste nello stabilire quale sia la pronuncia migliore da registrare, e cioè: quale sotto l'aspetto geografico (individuare il luogo con la pronuncia meno alterata), quale sotto l'aspetto sociale (classi colte o mediobasse?), quale sotto l'aspetto cronologico (individui giovani o anziani?), quale sotto l'aspetto stilistico (lingua colloquiale o formalizzata?); la cosa migliore sarebbe una raccolta ad ampio respiro che tenga conto di tutte queste varianti, che in un secondo tempo, una volta acquisite, verranno elaborate e messe a confronto da un punto di vista fonetico, lessicale e sintattico da cui poi si trarranno delle statistiche e delle conclusioni.

---

<sup>11</sup> Un modo semplice per stabilire se un suono affricato è più o meno fuso, consiste nell'ascoltare alla rovescia il suono in questione; nel caso di /tʃ/ e /dž/ dovremmo sentire qualcosa di simile a /št/ e /žd/ in presenza di poca fusione a livello fonetico.