

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Uno spoglio fonetico della base di dati-audio "The Græcanic Lexicon" dell'Università di Patrasso**

**This is a pre print version of the following article:**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/50808> since

*Publisher:*

Esagrafica

*Terms of use:*

Open Access

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

(Article begins on next page)

## UNO SPOGLIO FONETICO DELLA BASE DI DATI AUDIO "THE GRÆCANIC LEXICON" DELL'UNIVERSITÀ DI PATRASSO

Antonio Romano

Dip. di Scienze del Linguaggio - Università di Torino  
antonio.romano@unito.it

Nell'ambito della conservazione di lingue minacciate di estinzione, si inseriscono oggi numerosi progetti di raccolta e organizzazione di dati audio originali e di buona qualità, miranti a fornire archivi sonori di riferimento

Tra questi si annovera da qualche anno a questa parte una base di dati dei dialetti ellenofoni d'Italia, *The Graecanic Lexicon*, realizzata dall'Università di Patrasso con parlanti autoctoni della Bovesia e della Grecia Salentina.

A partire da una selezione di entrate lessicali, analizzate per tipo sillabico-accentuale e nella loro strutturazione in sillabe e *cluster* vocalici e consonantici, ho effettuato una serie di confronti con materiali simili da me raccolti direttamente e con i dati disponibili per il greco moderno.

Inoltre, sfruttando i *file* audio messi a disposizione, ho proceduto all'analisi di alcuni fatti fonetici presenti nel corpus, studiando le caratteristiche del sistema vocalico di queste varietà in relazione con quelli di varietà salentine romanze.

### 1. THE GRÆCANIC LEXICON (GL)<sup>1</sup>

#### 1.1 Descrizione

Il *GL* è un dizionario elettronico dei dialetti ellenofoni dell'Italia meridionale. Una sua prima versione, prevista per il CD-ROM *Electronic Dictionary of Pronunciation and Usage of Modern Greek for Foreigners* realizzato presso *Wire Communications Laboratory (WCL)* dell'Università di Patrasso, era originariamente (al momento della realizzazione di questo studio) disponibile *on-line*<sup>2</sup>.

La base di dati messa a disposizione include approssimativamente 9000 lemmi con sviluppo delle voci in Italiano e Greco moderno e con possibilità d'accesso da queste due lingue<sup>3</sup>.

---

<sup>1</sup> *The Graecanic Lexicon* (1996-2003): <http://www.wcl2.ee.upatras.gr/Project/Grec/>

<sup>2</sup> Direttore del progetto è George Kokkinakis (v. Kokkinakis et al., 2000). Il lavoro è stato svolto in collaborazione con un'*équipe* di linguisti di Atene (A. Botinis, M. Katsoyiannou & C. Petropoulou).

<sup>3</sup> Ovviamente risulta prioritariamente implementata la modalità d'accesso da grecanico/griko (sulla base però di una grafia "non collaudata").

Tra le motivazioni che hanno portato alla realizzazione di questa raccolta di dati ve n'è evidentemente una legata alla conservazione di una parte del patrimonio linguistico dei due gruppi di comunità italiane alloglotte coinvolte<sup>4</sup>. Ma un altro aspetto dell'opera realizzata (forse meno soggetto a distorsioni di prospettiva di quanto non sia l'ambizione di fotografare una realtà linguistica così complessa)<sup>5</sup> c'è soprattutto quello di contribuire alla conservazione delle caratteristiche idiofoniche della pronuncia e delle caratteristiche antropologiche di alcune voci di queste aree geolinguistiche<sup>6</sup>.

### 1.2 Costituzione del GL

Il GL è formato da 5983 lemmi registrati in Bovesia (con 3 locutori) e 2981 lemmi in Salento (con 2 locutori) ma prevede anche 359 esempi d'uso che contribuiscono a determinare una base di più di 20000 *file (.wav)*<sup>7</sup>.

Si tratta di segnali campionati a 16 kHz in codifica PCM a 16 bit, complessivamente di buona qualità per eseguire delle misure acustiche<sup>8</sup>.

## 2. ANALISI FONETICA DEI MATERIALI DISPONIBILI

### 2.1 Operazioni preliminari

Prima di eseguire dei trattamenti sui dati effettivamente accessibili *on-line* è stato necessario procedere con attente verifiche di coerenza dei materiali, basandosi su una rappresentazione fonetica. Stanti le consistenti differenze tra i gruppi di parlate coinvolti, una prima verifica è stata condotta sulle voci di una singola area (escludendo quindi le voci relative al Salento) e poi di una singola località (Bova), operando secondo le seguenti tappe:

- ❖ Selezione di forme lessicali singole (eliminazione degli omofoni);
- ❖ Selezione di parole isolate (no polirematiche, frasi e sintagmi);
- ❖ Verifica dell'esatta fonetizzazione di forme ortografiche *trascurate*;
- ❖ Sistematizzazione della trascrizione ortografica e sostituz. grafema-fonema;
- ❖ Applicazione di tecniche di conteggio statistico<sup>9</sup>.

---

<sup>4</sup> Si tratta della Bovesia (in Calabria) e della Grecia Salentina (in Puglia). Per un quadro linguistico generale delle due aree si veda Rohlfs (1950\*2001). Il progetto GL è stato intrapreso, con finanziamenti dall'Unione Europea, « with the conviction that it is important to capture and put on record the language that the indigenous Greek speakers of these areas are handing down to their descendants today » (v. *Introduzione* pubblicata nel sito *web*).

<sup>5</sup> Per questi temi, rimando a un recente lavoro in cui ho preso parte: v. Romano et al. (2002).

<sup>6</sup> Più critica sembra invece la proposta di una sua utilità come "dizionario d'uso" in quanto: (1) presenta un lessico specializzato del mondo rurale e pastorale; (2) non include parole grammaticali. Si presterebbe male, quindi, a un recupero funzionale delle strutture in esso presenti da parte delle giovani generazioni.

<sup>7</sup> Non tutto il materiale raccolto è stato pubblicato e non tutte le modalità d'accesso previste sono state implementate (mancano ad es. i dati audio della sezione salentina).

<sup>8</sup> I segnali, non normalizzati, presentano globalmente un buon rapporto *S/N*. Le informazioni disponibili sul sito precisano che: « Recordings were made initially on high quality analog recorder [...] then digitized at processing ».

<sup>9</sup> Con rif. a Muller (1968); Muljačić (1969); Schmid (1996); Molinu & Romano (1999).

Sono state così conteggiate 14086 sillabe e 32619 segmenti di cui 14638 vocali (10225 atone) e 17981 consonanti, mentre le 4413 forme lessicali sono state analizzate come tipi sillabico-accidentali (bisillabiche parossitoniche, bisillabiche ossitoniche, trisillabiche proparossitoniche etc.), sulla base delle loro occorrenze.

Le diverse strutture isolate e conteggiate e i materiali sonori associati sono stati confrontati con i dati numerici e le analisi acustiche già discussi in studi precedenti (statistiche condotte sui lessici *griko* e salentino romanzo e analisi fonetiche su dati sonori originali in Romano, 2001).

## 2.2 Analisi per tipo sillabico-accidentale

L'analisi dei tipi sillabico-accidentali è stata effettuata su un'ulteriore selezione di 3158 tipi sillabico-lessicali più frequenti (con copertura pari a ca. il 70% delle forme).

Il numero medio di sillabe per tipo è di 3,03 e si mostra quindi più vicino al dato relativo al greco mod. (3,22) di quanto non lo siano il dato corrispondente del griko (2,56) e del romanzo salentino (2,91)<sup>10</sup>.

I tipi sillabico-lessicali più frequenti sono quelli costituiti da sillabe CV (ad es. #CV.C'V.CV.#, #CV.CV.C'V.CV.#, #C'V.CV.# etc.) ma non mancano tipi inconsueti per varietà greche (come quelli con geminazione funzionale di consonanti interne: *pìria* 'orgoglio' vs. *pìrria* 'pettirosso', *ghièno* 'guarire (v. tr.)' vs. *ghiènno* 'parto (n.)' etc.)<sup>11</sup> mentre sono molto più rari dei tipi che siano insoliti per l'italiano o per le varietà italo-romanze in genere. I rapporti percentuali con cui si presentano i vari tipi mostrano una netta dominanza dei trisillabi su bisillabi e quadrisillabi, attestati con tassi di occorrenza simili (mentre nelle altre due varietà dominano bi- e tri-sillabi a scapito dei quadrisillabi che presentano tassi ridimensionati)<sup>12</sup>.

---

<sup>10</sup> I lessici griko e greco mod. analizzati sono disponibili *on-line* rispettivamente ai siti:

<http://atlante.clio.it/grecia/griko.html> (v. Tommasi, 1996) e

<http://www.xanthi.ilsp.gr/dictionaries/> (v. ILSP, 2003).

Un valido ausilio descrittivo delle proprietà fonetiche del griko è in Profili (1986) mentre per il greco mod. si è fatto riferimento alle descrizioni fonetiche e fonologiche di Arvaniti (1999a) e di Mackridge & Holton (1997) (oltre che al sintetico quadro offerto da Wells, 1998).

<sup>11</sup> Non mancano le geminate iniziali tipiche delle varietà romanze a contatto, come in *ttèni* 'pettine', qui ottenute però di solito per assimilazione (cfr. in questo caso greco mod. χτένι).

<sup>12</sup> Analoghi conteggi sul greco moderno conducono a percentuali molto simili a quelle osservate nel *GL*. Il greco mod. presenta però un po' meno trisillabi (35,6%), soprattutto a favore di più pentasillabi (9,5%) ed esasillabi (2,1%) qui scarsamente o per niente rappresentati (queste stime riproducono dei dati ottenuti a partire dai lessici citati in nota 10). La varietà romanza, presentando un tasso leggermente inferiore di proparossitoni, sembra prediligere notevolmente i parossitoni (più di quanto non facciano le varietà greche), mentre non sorprende il consistente aumento nelle occorrenze degli ossitoni in *GL*, in greco mod. e - soprattutto - in griko. Ancora una volta, si può notare come il lessico *GL* si presenti strutturalmente più vicino al lessico greco di quanto non faccia quello griko che, per questi aspetti, sembrerebbe discostarsi più da quest'ultimo che dal lessico romanzo considerato.

	GL		Griko		Sal. romanzo		Greco Mod.	
parossitoni	1892	<b>59,9%</b>	813	<b>53,6%</b>	1580	<b>85,8%</b>	1480	<b>59,1%</b>
ossitoni	635	<b>20,1%</b>	489	<b>32,2%</b>	52	<b>2,8%</b>	534	<b>21,3%</b>
propaross.	631	<b>20,0%</b>	215	<b>14,2%</b>	209	<b>11,4%</b>	491	<b>19,6%</b>
N tipi	3158		1517		1841		2505	

Tabella 1. Numero di occorrenze e frequenza relativa dei tipi accentuali del *GL* e dei lessici griko e romanzo salentini. Nell'ultima colonna sono anche riportati i risultati di un conteggio effettuato, con gli stessi criteri, in un lessico greco moderno.

	GL		Griko		Sal. romanzo		Greco Mod.	
CV	6485	<b>68%</b>	3566	<b>68%</b>	2787	<b>72%</b>	4755	<b>60%</b>
CVC	1582	<b>17%</b>	1112	<b>21%</b>	681	<b>18%</b>	1193	<b>15%</b>
CCV	404	<b>4%</b>	208	<b>4%</b>	126	<b>3%</b>	614	<b>8%</b>
V	312	<b>3%</b>	124	<b>2%</b>	302	<b>8%</b>	481	<b>6%</b>
VC	153	<b>2%</b>	32	<b>1%</b>	47	<b>1%</b>	186	<b>2%</b>
CCVC	130	<b>1%</b>	149	<b>3%</b>	63	<b>2%</b>	129	<b>2%</b>

Tabella 2. Principali tipi sillabici e loro occorrenze nei lessici analizzati.

### 2.3 Analisi dei tipi sillabici

I tipi sillabici considerati derivano dalla destrutturazione dei 3158 tipi sillabico-lessicali più frequenti (questi tipi coprono 9519 sillabe, pari al 75% ca. delle 14086 sillabe del *GL*). Si è avuta cura però nel distinguere i tipi, sulla base della loro occorrenza nelle diverse posizioni all'interno di parola e in relazione alla posizione dell'accento. Il confronto tra i tassi riscontrati nei diversi lessici viene qui semplificato per motivi di spazio come in Tab. 2.

Sono stati osservati anche i ranghi d'occorrenza di consonanti e *cluster*, concludendo in una sostanziale convergenza tra le tendenze presenti in *GL* e nel lessico griko, cui corrisponde naturalmente una certa autonomia rispetto al romanzo salentino (che predilige un ordine leggermente diverso) ma soprattutto rispetto al greco moderno (in cui sono attestate delle combinazioni impensabili per i due lessici italo-greci)<sup>13</sup>.

### 2.4 Misurazioni acustiche relative ad alcuni interessanti fenomeni fonetici

Tra i numerosi fenomeni fonetici presenti nel *GL*, per i quali sarebbero possibili analisi approfondite (vista la cospicuità delle realizzazioni disponibili), sono stati privilegiati, fino al momento della pubblicazione di questo studio, prevalentemente quelli relativi a:

- ❖ le caratteristiche del sistema vocalico;
- ❖ le opposizioni tra occl. dentali brevi e cacuminali lunghe ([d/ð] vs. [d̪/d̪]);
- ❖ le interessanti oscillazioni nelle realizzazioni di /zz/ ([z:/d:z]).

<sup>13</sup> Nel lessico greco antico mancano quasi tutti quei nessi consonantici formati da due o più ostruenti che sono invece comuni in greco (ad es. nel greco mod.: λέξη, λεπτό, βγαίνο, κτίριο, έκπτωση etc.). Alcune forme del greco mod. che sollevano dubbi generali di sillabazione secondo i modelli classici (v. Schmid, 1996) sono state analizzate tenendo conto di soluzioni convenzionali provvisorie (es.: έκπτωση #VC.CCV.CV#, άσθμα #VC.CCV#, παράκαμψη #CV.C'V.CV<sup>N</sup>.CCV#).

Per motivi di spazio, mi limito a commentare i risultati delle misurazioni effettuate sui timbri vocalici. Sono state misurate le prime due formanti di 450 vocali accentate (e questo numero rappresenta solo il 3,4% di tutte le vocali di questo tipo presenti nel *database*) nonché 200 vocali finali inaccentate (pari all'1,6% delle vocali finali), definendo le aree di esistenza delle tre voci maschili disponibili in *GL* (si tratta di voci molto caratteristiche ma anche molto diverse tra di loro, come prova il diverso utilizzo dello spazio articolatorio suggerito anche dal *mapping* acustico nei diagrammi  $F_1$ - $F_2$  di Fig. 1).

È stato anche effettuato il confronto tra le aree di dispersione vocaliche in posizione accentata e non-accentata finale per le produzioni del locutore TS (Bova), riscontrando solo leggerissimi fenomeni di centralizzazione, associati a una scarsissima propensione alla riduzione vocalica.

Le misure vocaliche di *GL* sono state confrontate con quelle di due varietà salentine (una grika e una romanza)<sup>14</sup> mettendo in evidenza delle somiglianze interlinguistiche anche maggiori, in funzione delle caratteristiche generali delle voci considerate (v. Fig. 2)<sup>15</sup>.

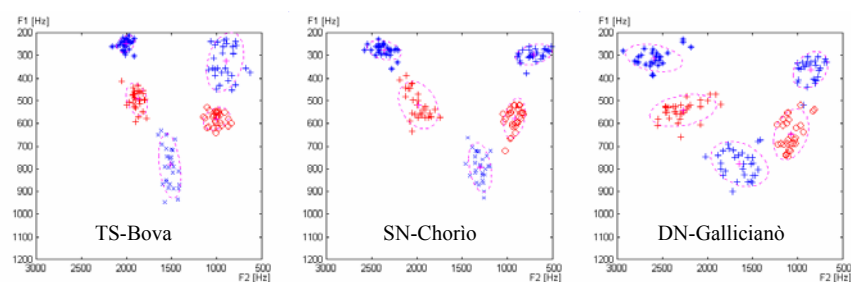


Figura 1. Diagrammi di dispersione delle vocali in posizione accentata (confronto tra le tre varietà di *GL* rappresentate da voci maschili).

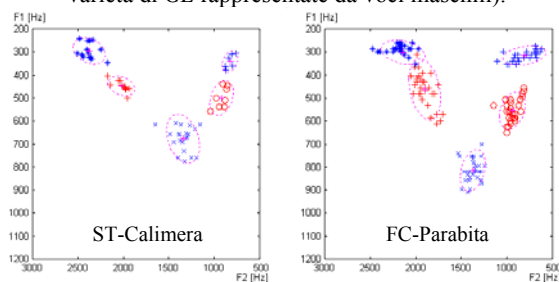


Figura 2. Diagrammi di dispersione delle vocali in posizione accentata (confronto tra varietà). A sinistra il *griko* di Calimera (LE) - si tratta però di un esiguo campione di 60 realizzazioni vocaliche in parole isolate -, e a destra il romanzo salentino di Parabita (LE) [v. nota 14].

<sup>14</sup> Questo confronto si basa su dati originali solo in parte già pubblicati (Romano, 2001).

<sup>15</sup> Un riscontro tra le caratteristiche dei sistemi vocalici qui presentati e quelli descritti per parlanti della varietà standard di greco moderno è possibile grazie a Fourakis et al. (1999).

### 3. CONCLUSIONI

La base di dati sonori analizzata si propone come un *corpus* per l'approfondimento di numerosi aspetti fonetici delle varietà dialettali italo-greche. Nel presente contributo sono state descritte sommariamente solo alcune sue caratteristiche e i risultati di alcune misurazioni condotte a partire dai materiali audio in essa contenuti, come saggio della sua fruibilità per ricerche di questo tipo.

### Bibliografia

- Arvaniti, A. (1999) Illustrations of the IPA: Modern Greek, *Journal of the International Phonetic Association*, 19, 167-172.
- Falcone, G. (1971) Testi bovesi (Galliciano di Condofuri) in trascrizione fonetica, *Parole e Metodi - BALI*, 2, 151-166.
- Fourakis, M., Botinis, A. & Katsaiti, M. (1999) Acoustic Characteristics of Greek Vowels, *Phonetica*, 56, 28-43.
- ILSP. Institute for Language and Speech Processing of Athens (2003):  
<http://www.xanthi.ilsp.gr/dictionaries/>
- Kokkinakis, G. et al. (2000) Electronic Dictionary of Pronunciation and Usage of the Græcanic Dialect of Southern Italy, In *Proc. of the 2<sup>nd</sup> Int. Conf. on Language Resources and Evaluation - LREC* (Atene, 2000).
- Mackridge, P.A & Holton, D. (1997) *Greek: Comprehensive Grammar of the Modern Language*, London, Routledge.
- Molinu, L. & Romano, A. (1999) La syllabe dans un parler roman de l'Italie du Sud (variété salentine de Parabita - Lecce), In *Actes des II<sup>èmes</sup> Journées d'études linguistiques "Syllabes"* (Nantes, 1999), 148-153.
- Muljačić, Ž. (1969) *Fonologia generale e fonologia della lingua italiana*, Bologna, Il Mulino.
- Muller, Ch. (1968) *Initiation à la statistique linguistique*, Paris, Larousse.
- Profili, O. (1986) Description du système phonétique et phonologique du parler grico de Corigliano d'Otranto, *Studi Linguistici Salentini*, 14.
- Rohlf, G. (1950\*2001) *Grammatica storica dei dialetti Italogreci* (Calabria, Salento), München, Beck (ed. it. Galatina, Congedo, 1977; rist. 2001).
- Romano, A. (2001) *Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento: approche linguistique et instrumentale*, Lille, Presses Univ. du Septentrion.
- Romano, A., Manco, F. & Saracino, C. (2002) Un giorno a Martano: riflessioni sulla situazione linguistica della Grecia Salentina, *Studi Ling. Salentini*, 26, 61-109.
- Schmid, S. (1996) A typological view of syllable structure in some Italian dialects, In P.M. Bertinetto et al. (editors), *Certamen Phonologicum, Atti del 3<sup>o</sup> Phonology Meeting* (Cortona, 1996), Torino, Rosenberg & Sellier, 247-265.
- Tommasi, S. (1996) *Katalisti o kosmo. Materiali per un seminario sul griko, Ghetonia-Calimera*, Galatina, Ed. Salentina (Lessico Griko:  
<http://atlante.clio.it/grecia/griko.html>)
- Wells, J.C. (1995-1998) "SAMPA for Greek":  
<http://www.phon.ucl.ac.uk/home/sampa/grk-uni.htm>