# Protocol for Transcription of Griko[*]

## 1. Introduction

This document is the manual used for transcribing of Griko audio files. It contains information on the conventions employed in the transcription of the sound files, as well as information on how data enrichment more in general was carried out.

Transcription was carried out manually. In addition, each word in the corpus is assigned a Part-of-Speech tag, a lemma, and a translation into Italian. This information is provided in separate tiers:

a.  Text Tier: contains the transcription of the sound file.

b.  Part-of-Speech (PoS) Tag Tier: contains the Part of Speech Tag of each item in the Text Tier (see "Protocol for Part-of-Speech tagging of Griko").

c.  Italian Gloss: contains a word-for-word translation of the Griko sentence in Italian.

d.  Griko Lemma: contains lemmas for each word. The version of Griko lemmas we used is the one provided in the following dictionary: Greco, Carmine (2001). *Lessico di Sternatia, paese della Grecia salentina: italiano-griko-neogreco, griko-italiano-neogreco, neogreco-griko*. Lecce: Edizioni Del Grifo.

All four tiers were created manually using Praat (http://www.fon.hum.uva.nl/praat/). Within this program, the sound file was divided into sentences, which were separated by boundaries. Within each set of boundaries, spaces were used to indicate word boundaries in the text tier. In addition to space,  we used the apostrophe and the hyphen, in order to mark word boundaries within which the phenomenon of *raddoppiamento sintattico* (phonosyntactic gemination) takes place (see page 4).

In order to ensure a one-to-one correspondence between items on the transcription tier and items on the other tiers, spaces are used on those tiers as well. Whenever itemization discrepancies across different tiers occurred, the following conventions were used:

1.  when a phonological word contains two morphological words of different syntactic category, e.g. with prepositions fusing with the definite article (text-tag discrepancy), the tag of the two words is separated by period (".").

2.  If a Griko word corresponds to more than one word in Italian (text-gloss discrepancy), then underscore between the two Italian words is used. E.g. *'irta* and *sono_venuta*.

Only the speech of the informants is transcribed, and not that of the interviewer. Whenever additional speakers were present, an additional tier was used to notate their speech. Intonational breaks are transcribed as "#". Whenever transcription was not possible, this was notated on the text tier by "[…]".

Each sentence in the transcription tier is preceded by a number, followed by a space. The number corresponds to the number of the test sentence in the questionnaire. The end of a sentence is notated by a period or a question mark, as appropriate depending on sentence type. The only

other punctuation mark that appears on the text tier is the apostrophe and the hyphen, which mark contexts in which phonosyntactic gemination takes place. No other punctuation marks are used on the transcription tier. On all other tiers, only the underscore and the period are used, in the way mentioned above.

## Synopsis of symbols used

[Text Tier]

| | |
|---|---|
| . | end of test sentence |
| ? | end of test sentence (question) |
| ' | phonosyntactic gemination |
| - | phonosyntactic gemination |
| # | intonational break |
| […] | material in the sound file is not transcribed |

[Gloss Tier]

| | |
|---|---|
| _ | separates two words in Italian corresponding to a single one in Griko |

[PoS Tag Tier]

| | |
|---|---|
| . | separates two PoS tags realized as a single word in Griko |

The transcription is orthographic. Since there is a (limited) tradition of written Griko, we decided to forego a phonetic or phonological transcription, which would be foreign to native speakers of Griko. We relied on a version of the orthographic conventions adopted in texts written in Griko, like the ones employed in e.g. the magazine *Spitta* (http://www.rizegrike.com/spitta.php). The orthographic conventions used for Griko closely recall conventions adopted for Standard Italian. Notes explicating the conventions used are provided on page 4.

# ORTHOGRAPHIC CONVENTIONS USED FOR GRIKO

| Transcription in GRIKO | Correspondence to simplified IPA | NOTES |
|---|---|---|
| a; à | [a] | |
| b | [b] | |
| c | [k] before [a] , [o], [u]<br><br>[tʃ] before [i] and [e] | e. g. *Carlo* [ˈKarlo]<br><br>*ceràsi* [tʃeˈrasi]; *cilìa* [tʃiˈlia];<br><br>**<cia> = [tʃa];** e.g. *cialatèdda* [tʃalatˈed:a];<br>**<cio> = [tʃo]; e.g.** *ciofàli* [tʃoˈfali];<br>**<ciu> = [tʃu]; e.g.** *ciumpì* [tʃumˈpi];<br>**<cìa> = [tʃˈia];**<br>**<cìo> = [tʃˈiu];** |
| ch | [x] | e.g. *rùcho* [rˈuxo] |
| d | [d] | |
| e, è | [e] | |
| f | [f] | |
| g | [g] before [a] , [o], [u]<br><br>[dʒ] before [i] and [e] | e.g. *garrofèddo* [gar:oˈfed:o];<br><br>**<ge> = [dʒe];**<br>**<gi> = [dʒi];**<br><br>**<gia> = = [dʒa] e.g.** *sangìa* **[sanˈdʒia];**<br>**<gio> = = [dʒo]; ;**<br>**<giu> = [dʒu];** |
| gh | [g] | e.g. *ègghene* [ˈeg:ene] |
| i; ì | [i] | |
| j | [ɣ] | |
| k | [k] | |
| l | [l] | |
| m | [m] | |
| n | [n] | |
| o; ò | [o] | |
| p | [p] | |
| r | [r] | |
| s | [s] | |
| t | [t] | |
| u; ù | [u] | |
| v | [v] | |
| z | [dz] | e.g. *ziò* [zˈio] |
| ts | [ts] | e.g. *tsìlo* [ˈtsilo] |
| sc | [ʃ] | e.g. *scìmmata* [ˈʃim:ata] |
| gn | [ɲ] | e.g. *signurèdda* [siɲuˈred:a] |

**NOTES**

1. The symbol <g> represents the plosive /g/, unless it precedes a front vowel (⟨i⟩ or ⟨e⟩). In this case it represents the affricate /dʒ/. When the plosive pronunciation occurs before a front vowel, ⟨gh⟩ is used, so that <ghe> represents [ge] and <ghi> represents [gi].

2. <c> represents the affricate /tʃ/ before front vowels ⟨i⟩ and ⟨e⟩.  In some words of Italian origin, before non-front vowels (<a>, <o>, <u>) <c> spells the unvoiced plosive /k/, e.g. *Carlo* ['Karlo].

3. Thus, the letter <i> may function as a mere indicator that the preceding ⟨c⟩ or ⟨g⟩ is affricate, e.g. *cia* (/tʃa/), *ciu* (/tʃu/), *gia* (/dʒa/)*, giu* (/dʒu/).

4. The symbols <ch> always represent [x]. E.g. *cheretìmmata* [xere'tim:ata]; *chàri* ['xari]. Plosive [k] is transcribed as <k>.

5. For every word that has two or more syllables an accent diacritic is used, to indicate the location of stress, e.g. *cheretìmmata.*

6. *Raddoppiamento fonosintattico* (or phonosyntactic doubling, PD): the phenomenon of syntactic gemination, i.e. the lengthening of word-initial consonants related to the presence of a particular set of preceding elements. PD is notated by an apostrophe or by a hyphen, e.g.: *si' putèka* [sip:u'teka], *èssu- mu* [es:um:u].