

# Protocol for Part-of-Speech Tagging of Griko\*

## 1. Introduction

This document is the manual used for performing part-of-speech (PoS) tagging of Griko texts. All aspects of the data enrichment process, namely transcription, tagging, lemmatization and glossing in Italian were carried out manually, using Praat (<http://www.fon.hum.uva.nl/praat/>); see also “Protocol for transcription of Griko”. The categories used for PoS tagging are the following:

1. N [Noun]
2. Adj [Adjective]
3. V [Verb]
4. Adv [Adverb]
5. P [Adposition]
6. C [Complementizer]
7. Pr [Pronoun]
8. D [Determiner]
9. Prt [Particle]
10. Num [Numeral]

The specifications for values and attributes that were ascribed to each category are explicated in separate subsections below.

## 2. General remarks

1. In the transcription, the category of the word appears first. Specifications for other attributes are separated with a plus (“+”) sign.
2. For each category, there exist obligatory and optional attributes. A value for the obligatory attributes is always specified. Regarding optional attributes, when no value is provided, the value is set to default (which is provided for particular categories and optional attributes).
3. The size of the internal composition of each tag is constant for each category, but not identical across categories. For instance, for Griko nouns a 4-character tag is minimally needed, whereas for finite verbs the tags are 9-character long.
4. In case a specification cannot be given with certainty, e.g. in case the gender of a particular noun is unclear the value ‘unspecified’ (“U”) is provided.
5. In case characterization for a particular attribute does not apply for a given category, 0 (zero) is used.

---

\*The manual for part-of-speech tagging of Griko was developed by Marika Lekakou, Valeria Baldissera and Antonis Anastasopoulos during the data enrichment process of the scientific project “Documentation and analysis of an endangered language: aspects of the grammar of Griko”, funded by the John Latsis Foundation (project duration January to December 2013).

### 3. Specifications, Attributes and Values for each category

#### 3.1 Noun

Abbreviation: N

Specification: Features

Obligatory attributes: Gender, Number, Case

Values for obligatory attributes

Gender: Masc/Fem/Neu

Number: S/Pl

Case: Nom/Gen/Acc/Voc

Optional Attribute: Type.

Since most nouns in our corpus are common, we do not specify the type; common is treated as default. Thus, only proper names come with a fifth specification, namely Prop (for Proper).

Example: the tag for *Maria* is N+Fem+S+Nom+Prop.

The case ascribed to a noun does not always reflect morphological distinctions, but may rely on the syntactic context. For instance, nouns realizing the syntactic role of object will be tagged as realizing accusative case, even if there is no discrete morphological marking for accusative case on the noun. This was deemed necessary for several reasons, one of them being the lack of syntactic annotation of the corpus.

#### 3.2 Adjective

Abbreviation: Adj

Specification: Features, Degree, Position

Obligatory Attributes: Gender, Number, Case

Values for obligatory attributes:

Gender: Masc/Fem/Neu

Number: S/Pl

Case: Nom/Gen/Acc/Voc

Optional attributes: Position, Degree, Nominalization

Values for optional attributes:

Position: Post(nominal)

Degree: Comp(arative)/Sup(erlative)

Nominalization: NM (Nominalized)

The default value for Position is Preposed (reflecting the order Adj-N). When Postposed, an adjective will receive the specification Post (i.e. post-posed).

Example, the tag for the adjective in *petia mincià* (“children young”) is the following: Adj+Neu+Pl+Nom+Post.

The default value for degree is Positive. When the adjective is of comparative or superlative degree, the values Comp and Super are used.

The default value for Nominalized is negative. So NM (Nominalized) only appears in the marked case. When nominalized, the adjective is neither preposed nor postposed, as there is by definition no overt noun with respect to which the adjective is ordered. So NM could be seen as another value for Position.

### 3.3 Verb

Abbreviation: V.

Specification: Features, Type.

Obligatory attributes for all members of category V: Finiteness, Voice, Type.

Values for obligatory attributes:

Finiteness: Fin(ite)N(on)Fin(ite)

Voice: Act(ive)/N(on)Act(ive)

Type: M(ain)/Aux(iliary)

Subtypes of Aux:

- a. Mod = modal auxiliary verb
- b. PRF = perfect auxiliary, e.g. 'have' and 'be' in compound tenses (i.e. present and past perfect).
- c. PASS = passive auxiliaries, e.g. 'be' and 'come'.
- d. ASP = aspectual auxiliaries, e.g. *steo*.

#### 3A. Attributes of finite verbs (VFin)

Tense: Past/NonPast

NonPast is the present tense form, used in Griko also as a future tense.

Aspect: Perf(ective)/Imperf(ective)

Mood: Ind (Indicative)/Imp (Imperative)/Subj (Subjunctive)

The only morphological distinction that exists in Griko is between Imperative and Nonimperative mood. We tag as subjunctive the verbs that follow the subjunctive complementizer *na*. These verbs invariably combine perfective aspect and nonpast tense.

Number: S(ingular)/P(lural)

Person: 1/2/3

For example, a finite main verb like *teli* ("wants") is tagged as follows:  
V+fin+M+Act+Nonpast+Imperf+Ind+S+3

#### 3B. Attributes of nonfinite verbs (VNfin)

Subtype: Inf(initive)/Part(iciple)

We characterize all non-finite verb forms that are not infinitives as participles (subsuming gerunds too). This is meant purely as a descriptive label.

Aspect: Perf(ective)/Imperf(ective)

The characterization reflects the morphological specification of the stem.

Number: S(ingular)/P(lural)

Gender: Masc(uline)/Fem(inine)/Neu(ter)

Griko passive participles inflect for gender and number. In the all other cases (active participle, infinitive), the distinctions don't apply, so 0 is used for these attributes.

For example: a VNfin such as *vriskonta* in *pao vriskonta* would be tagged in the following way: V+Nfin+M+Act+Part+Imperf+0+0.

### 3.4 Adverb

Abbreviation: Adv.

Specification: Type, Features.

Obligatory attributes: Type.

Values for obligatory attributes:

Type: Temp(oral), Loc(ative), Interr(ogative), Asp(ectual), Epist(emic), Quant(ificational), QuantNeg (Negative Quantificational).

Subtype: Temp(oral)/Loc(ative).

The specification of an adverb as interrogative makes possible its further specification as temporal or locative.

Optional Attribute: Degree.

Value for Degree: Comp/Super

Default degree specification is positive, unless otherwise stated.

Example, *pu* is tagged as Adv+Interr+Loc, *pote* as Adv+Interr+Temporal.

### 3.5 Adposition

Abbreviation: P

Specification: Feature

Attribute: P/Pfus(ed)

P is used for simple P's, Pfus for when P is fused with the definite article (D) that follows it. In the latter case, we include the information of the D head too. This is a case where a single word corresponds to two tags, separated by a ".".

Examples: *atsè* is tagged as P, *s(t)i* is tagged as P+Pfus.D+Det+Fem+S+Acc.

### 3.6 Complementizer/Conjunction

Abbreviation: C

Specification: Type and Subtype.

Attributes of type: Sub(ordinating)/Coord(inating)

Co-ordinating conjunctions correspond to "and", "or".

Sub-ordinating conjunctions introduce embedded clauses.

Attributes of subordinating (Sub) C: Decl(arative), Inter(rogative), Rel(ative), Caus(al), Temp(oral), Cond(itional), Subj(unctive), Def(ault).

Def(ault) occurs whenever the value/function of the all-purpose complementizer *ka* is unclear.

Examples: *ce*: C+Coord, *na*: C+Sub+Subj.

### 3.7 Pronoun

Abbreviation: Pr

Specification: Type, Features

Attributes for Type: Pers(onal)/ Dem(onstrative)/ Inter(rogative)/ Quant(ificational)/ Poss(essive)

Attributes for Features:

Strength: W(eak)/Str (ong)

Person: 1/2/3

Gender: Masc/Fem/Neu

Number: S/P

Case: Nom/Acc/Gen/Voc

Strength and person specifications are only applicable for personal pronouns.

Example: *cini* (“those”) is tagged as Pr+Dem+0+0+Pl+Masc+Nom.

Optional attributes: Position, Clitic Doubling.

Default value for Position is proclisis (weak personal pronouns precede finite verbs in Griko as in Standard Modern Greek). Encl(isis) is specified when the pronoun follows the verb.

Default value for Clitic Doubling is no occurrence of clitic doubling. When doubling occurs, dou(bling) is additionally specified.

### 3.8 Determiner

Abbreviation: D

Specification: Type, features.

Values for Type: Def(inite)/Indef(inite)

Values for Features:

Gender: Masc/Fem/Neu

Number: S/Pl

Case: Nom/Gen/Acc/Voc

Example: *i* (definite feminine singular) is tagged as D+Det+Fem+S+Acc.

### 3.9 Particle

Abbreviation: Prt

Specification: Type, Subtype

Attributes: Neg/Other

Attributes for Subtype Neg:

Ind(icative)/N(on)Ind(icative)/Sent(ential) U(nknown)

In our corpus, all particles are negative. In Griko, as in Standard Modern Greek, sentential negative markers are sensitive to the mood (indicative/nonindicative) of the verb. Negative particles that occur in clausal ellipsis contexts are characterized as Sent(ential). For example, *ndè* is tagged as Prt+Neg+Sent.

### 3.10 Numeral

Abbreviation: Num

Example: *ettà* (“seven”).