

A Treebank of Asia Minor Greek

Eleni Vligouridou, Inessa Iliadou, Çağrı Çöltekin

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



LREC-COLING  2024

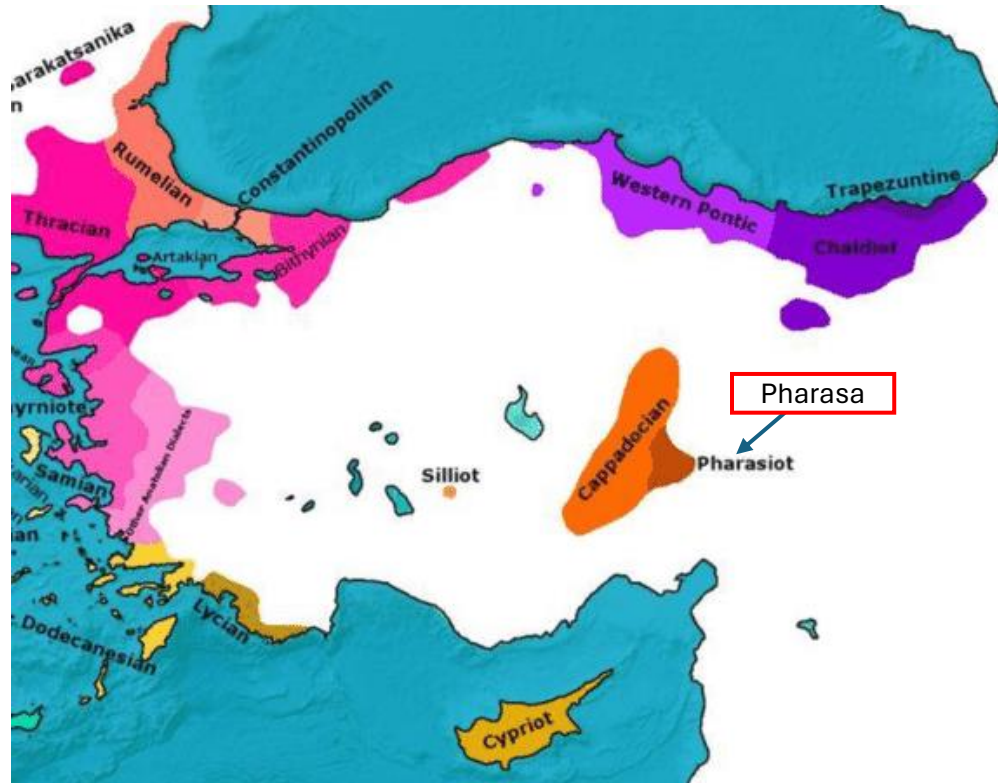
EU
RA
RESOURCES

ICCL
International
Committee on
Computational
Linguistics

Agenda

- ✓ A brief introduction to dialect.
- ✓ The annotation practice
- ✓ Interesting linguistic phenomena
- ✓ Treebank statistics
- ✓ Summary / future

Introduction to AMG Treebank – Pharasiot Location



Until 1923:

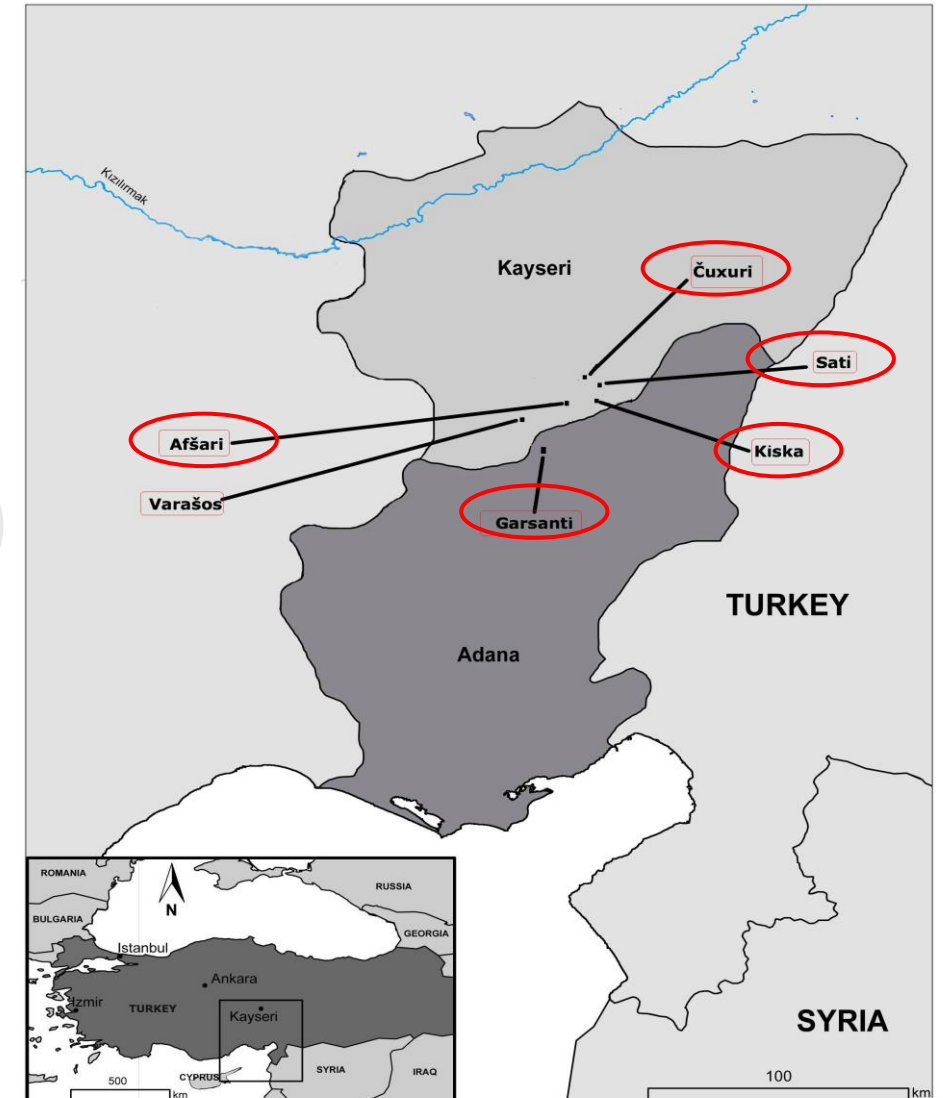


Figure 1.4 , pg 9 **Bagriaçık (2018)**

Introduction to AMG Treebank – Pharasiot Location-endangered status

Today's Greek villages with PhG speakers



Endangered status

- ✓ 1923 population exchange between Greece and Turkey
- ✓ Local's negative attitude towards refugees



diminishing the number of PhG speakers

- ✓ Unavailability of resources

Bagriaçık (2018):

25 known speakers

2 first-generation that passed away

Introduction to AMG Treebank – Pharasiot

Related dialects-relation to modern/ancient standard Greek

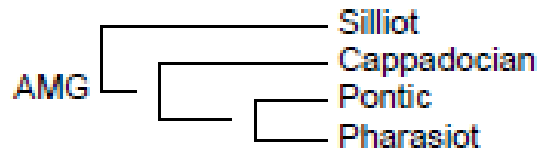
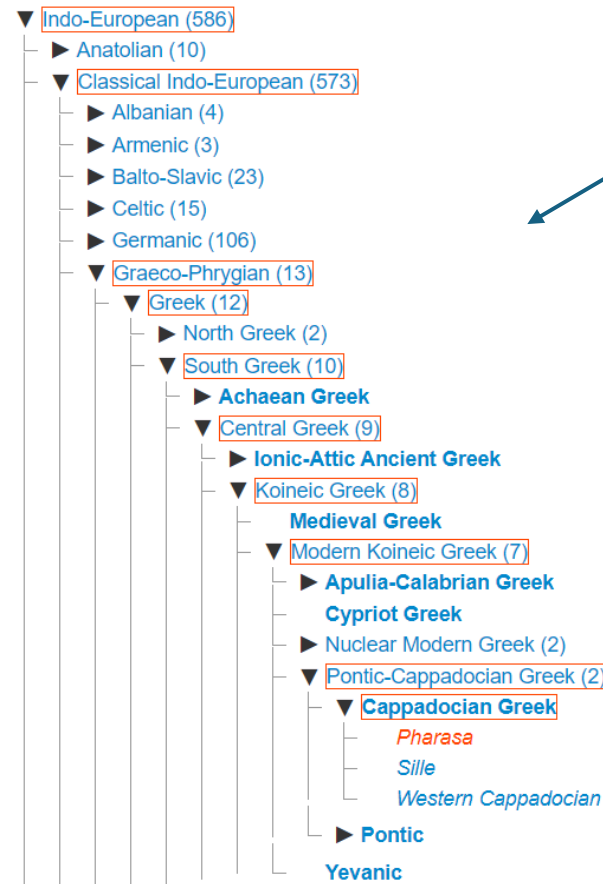


Figure 1: The classification of AMG dialects by Dawkins (1916).



Glottolog's classification

Relations with:

Modern standard Greek:

-One of the modern standard Greek dialects.

Ancient standard Greek

Passive verbs generating Simple Past indicative forms with endings of '-θην', '-θης' and '-θη'. (Andriotis 1948)

Turkish

Language contact.

Annotation practice

UD framework

Two annotators: Native Greek speakers

IAA: 25 sentences- annotated independently

- ✓ 87.54 % of POS tags (Cohen's κ : 0.8581)
- ✓ 70.93 % of dependency labels (disregarding the matching head, Cohen's κ : 0.6841).
- ✓ 80.83 % agreement on head identification
- ✓ 60.38 % agreement on full dependency labels (head and relation label)

Difficulties:

- ✓ Data collection
- ✓ Unavailability of digital resources
- ✓ Multilingual aspects

Interesting linguistic phenomena

✓ **Phonetics and Phonology:**

1. Substitution of the unstressed η with ε (ψεό instead of ύψηλό)
2. Fusion of diphthongs with words like (δεβοσύνη instead of
διαβολοσύνη)

✓ **Morphology:**

Agglutinative morphology

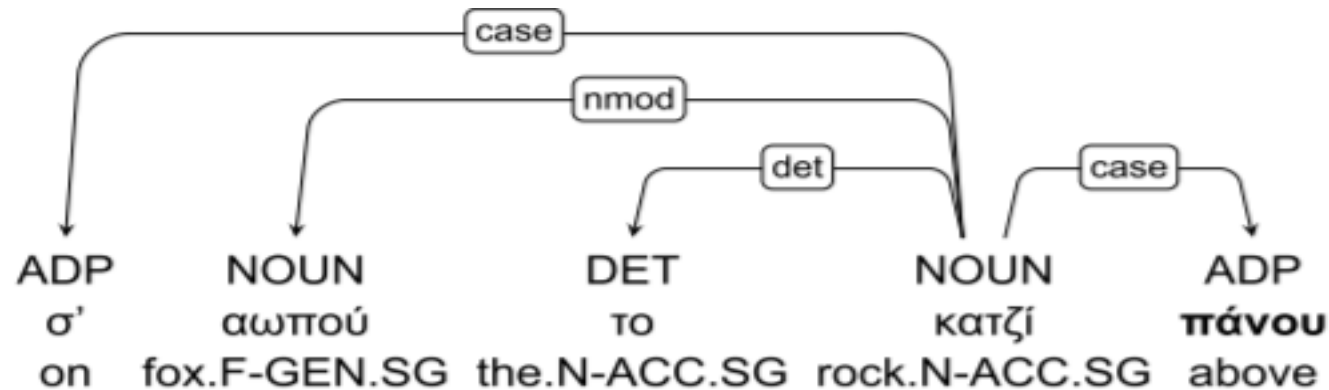
a. ρυιθί -ου -v
hen.N -GEN -PL
'of (the) hens' [AMG (Pharasiot)]

b. ρυιθί -ων
hen.N -GEN.PL
[Standard Modern Greek]

Interesting linguistic phenomena

✓ **Syntax:**

1. Adpositions:



2. τί' or 'di' with reported speech : ' *Είπεν 'τι κι τζ' αωπός:* '

✓ **Lexical:**

Lexical borrowings from: Armenian (κάτζίν), the Pontic Dialect (άπός), Latin (γουργούρι) and Turkish (bey).

Treebank statistics

Feature	Frequency
Number	2362
Case	1774
Gender	1733
PronType	930
Person	684
Aspect	591
VerbForm	591
Voice	591
Mood	589
Tense	553
Definite	401
Poss	57
NumType	15
Foreign	13

Table 1: Morphological features and their frequencies.

POS Tag	Frequency
PUNCT	728
NOUN	683
DET	658
VERB	591
PRON	326
ADP	256
CCONJ	172
ADV	162
AUX	158
SCONJ	152
ADJ	60
NUM	58
PART	55
INTJ	39
X	12
PROPN	7
SYM	1

Table 2: POS tag distribution.

58.62% neutral gender
 28.07 % M.Greek treebank
 17.69 % A.Greek treebanks

Dependency label	Frequency
punct	728
det	618
root	350
obl	238
case	234
obj	229
nsubj	219
conj	195
nmod	189
advmod	188
mark	174
cc	171
aux	113
amod	57
parataxis	55
nummod	46
cop	45
vocative	42
advcl	37
iobj	37
discourse	32
ccomp	31
xcomp	29
expl	16
dislocated	13
appos	8
nsubj:pass	8
acl:relcl	5
acl	4

Table 3: Dependency label distribution.

Treebank statistics

- ✓ Adpositions: **06.25 %** of the adpositions following the verb
 - 0.54 %** for modern standard Greek treebanks
 - 3.08 %** the Ancient Greek treebanks
- ✓ Word order: **23.62 %** of the obj or ccomp relations follow the verb.
 - 19.50 %** for modern standard Greek
 - 45.87 %** for Ancient Greek treebanks
- 52.03 %** subjects (nsubj or csubj) to the right of their verbs
 - 26.43 %** for modern Greek
 - 29.94 %** for ancient Greek treebanks.

Summary/Future

- ✓ Creation of a treebank for a critically endangered dialect which provides a valuable tool for linguistic analysis

- ✓ Future plans:
 - A more comprehensive representation of AMG Dialects
 - Collaboration with other researchers and experts of the field.
 - Use of treebank for machine learning and natural language processing purposes.