

Neither Ancient, nor Modern: Challenges for the creation of a Digital Infrastructure for Medieval Greek

Epistemic Networks and GRID + Web 2.0 for Arts and Humanities
30-31 January 2008 • Imperial College Internet Centre • Imperial College London

Notis Toufexis • nt262@cam.ac.uk

University of Cambridge
Grammar of Medieval Greek project



Creative Commons

Overview of this talk

- Perspective of a research team
 - “Small” subject
- What kind of research are we conducting?
- What kind of data do we use & what are their problems?
- What it is the output of our research?

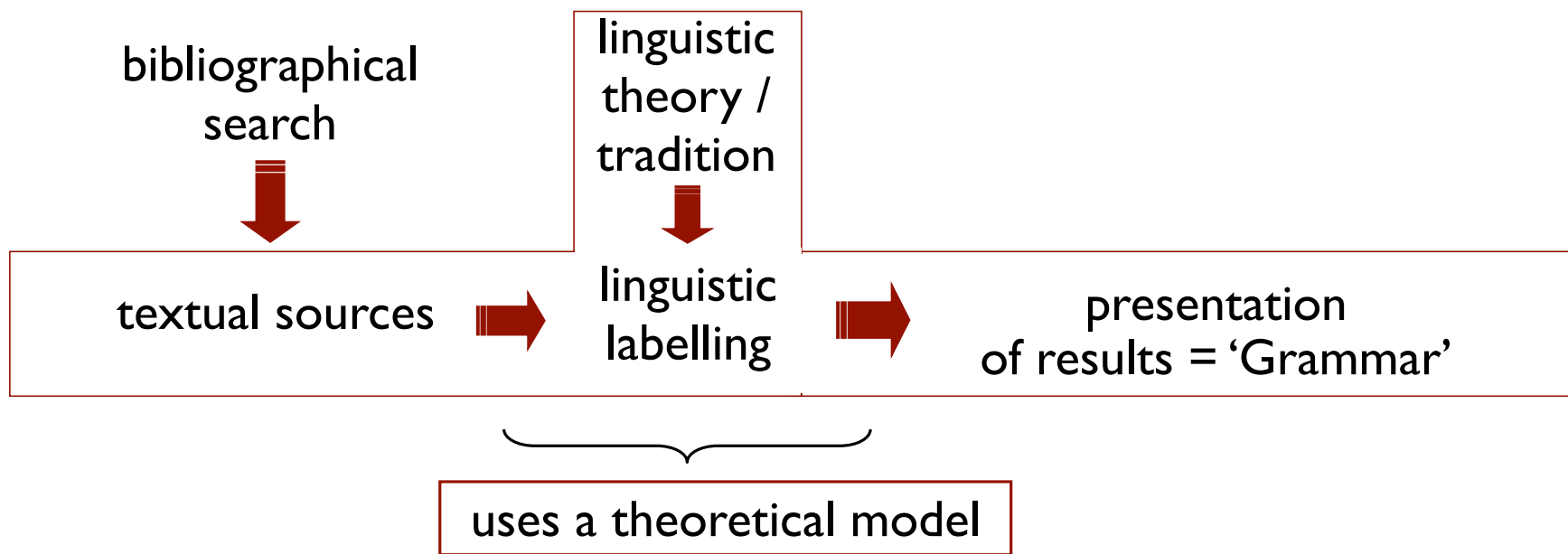
Grammar of Medieval Greek

- 5-year research project, funded by the AHRC
- 2 full-time research associates, 2 consultants, 2 project leaders
- Covers the period of 1100-1700
 - Uses a wide as possible corpus of **published** material as primary sources
- Aims to produce a Grammar of vernacular Greek in book form

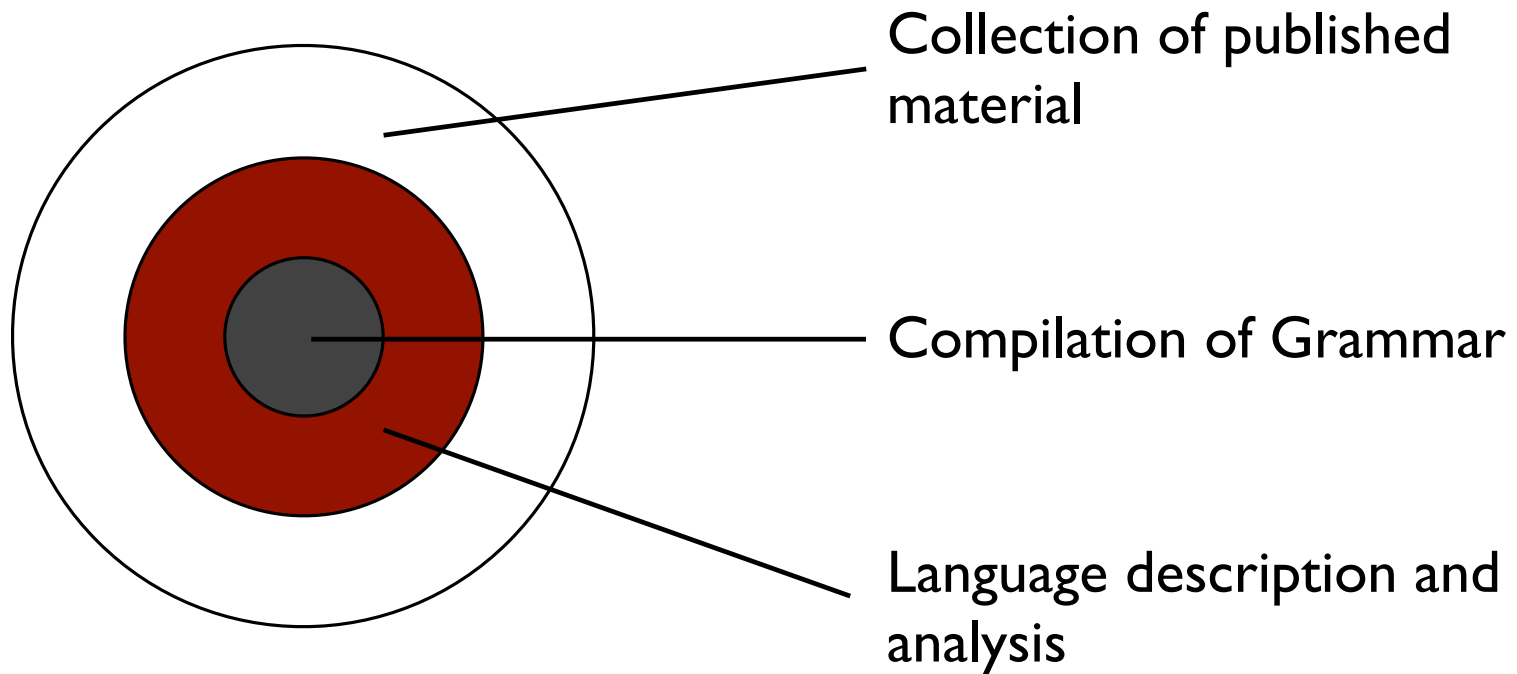
(Computing) resources in the domain of Medieval Greek

- Dictionary of Medieval Greek not concluded yet
- No reliable electronic corpus available
 - Very early stages of automated linguistic processing
- No register of authors & works compiled
- Not all key-texts available in modern editions

Compiling a Grammar for an extinct language



Workflow for the creation of the Grammar



Language description = metalinguistic annotation

- Application of markup

“Unambiguous declarative metalanguage inserted or otherwise associated with primary linguistic data to identify items for processing that software cannot otherwise reliably treat.”

McCarthy 2002: 104

Use of a database tool for collection and annotation of text excerpts

- Relational database
- Implemented with the use of FileMaker Pro
- The database is not designed to model the texts we are using but the task of the ‘grammarians’ in interpreting the linguistic evidence found in them
- Users of database are experts
- Primary data are recorded and labelled as defined by the theoretical model

The world of megabytes

- Present population of the database:
 - ca. 20.000 richly annotated text excerpts
 - ca. 1.300 bibliographical entries (many annotated)
 - ca. 1.300 records about individual sources (literary documents and literary texts)
- Is quantity the only criterion of “good” research?

The Grammar database

Annotation | Lemma overview | Paradigm overview in all texts | Paradigm overview in this text | Parad. in same case overview | Source overview

Show source **Πτοχ. III** Previous Next

Κ' εἰς μακελλεῖον Πτοχ. III 273.31 app. cr. (P) Edit Relookup text

Extended contexts: 12th c./14th c. texts: EXC_ID_20283 Source(s): TXT_134 Pub(s):

Lemma: μακελλεῖον without aut(s): μακελλεῖον

POS: NOUN ART ADJ PRON NUM PART Person: 1 2 3

Case: SG PL INDET NOM GEN DAT ACC VOC INDET

Gender: masculine feminine neuter non-spec. Other0

Reports: Paradigm: Current Found set
 Lemma: Current Found set
 Article: Current Found set

New excerpt from same source

Noun | Adjective & Participle | Article | Pronoun | Numeral

1 2 3

irregular residual comparison Other0

proparoxytone paroxytone oxytone

oxytone/paroxytone paroxytone/oxytone paroxytone/proparoxytone proparoxytone/paroxytone indifferent

Absence of accent shift Absence of agreement Accent shift Dialectal variant Double plural Foreign origin Gender change New ending Pluralia tantum Singularia tantum

hapax rare variable frequent normal exclusive form

see notes calc. not possible


PAR_65 View paradigm details

Neut. in -ιον, paroxytone / oxytone, no loss of /o/ (σχολεῖον / σχολεῖον)

check again dubious uncertain rare form Non-standard ending Rare form Dialectal variation Interesting for syntax

	SINGULAR	PLURAL
NOM	σχολεῖον(ν) / σχολεῖον(ν)	σχολεῖα / σχολεῖα
GEN	σχολεῖου / σχολεῖου	σχολεῖων(ν) / σχολεῖων(ν)ι
DAT	σχολεῖν	σχολεῖσιν
ACC	σχολεῖον(ν) / σχολεῖον(ν)	σχολεῖα / σχολεῖα
VOC	σχολεῖον(ν) / σχολεῖον(ν)	σχολεῖα / σχολεῖα

Created by nt262 on 29/01/2008 18:54:31. Modified by nt262 on 29/01/2008 18:56:14



Challenges

Textual multiplicity

```
graph TD; A[Textual multiplicity] --> B[Medieval texts are unfixed with multiple versions]; A --> C[Different editorial practices and interpretations]; B --- D[readings in the app. cr.]; C --- E[text differs from edition to edition];
```

Medieval texts
are unfixed with multiple
versions

readings in
the app. cr.

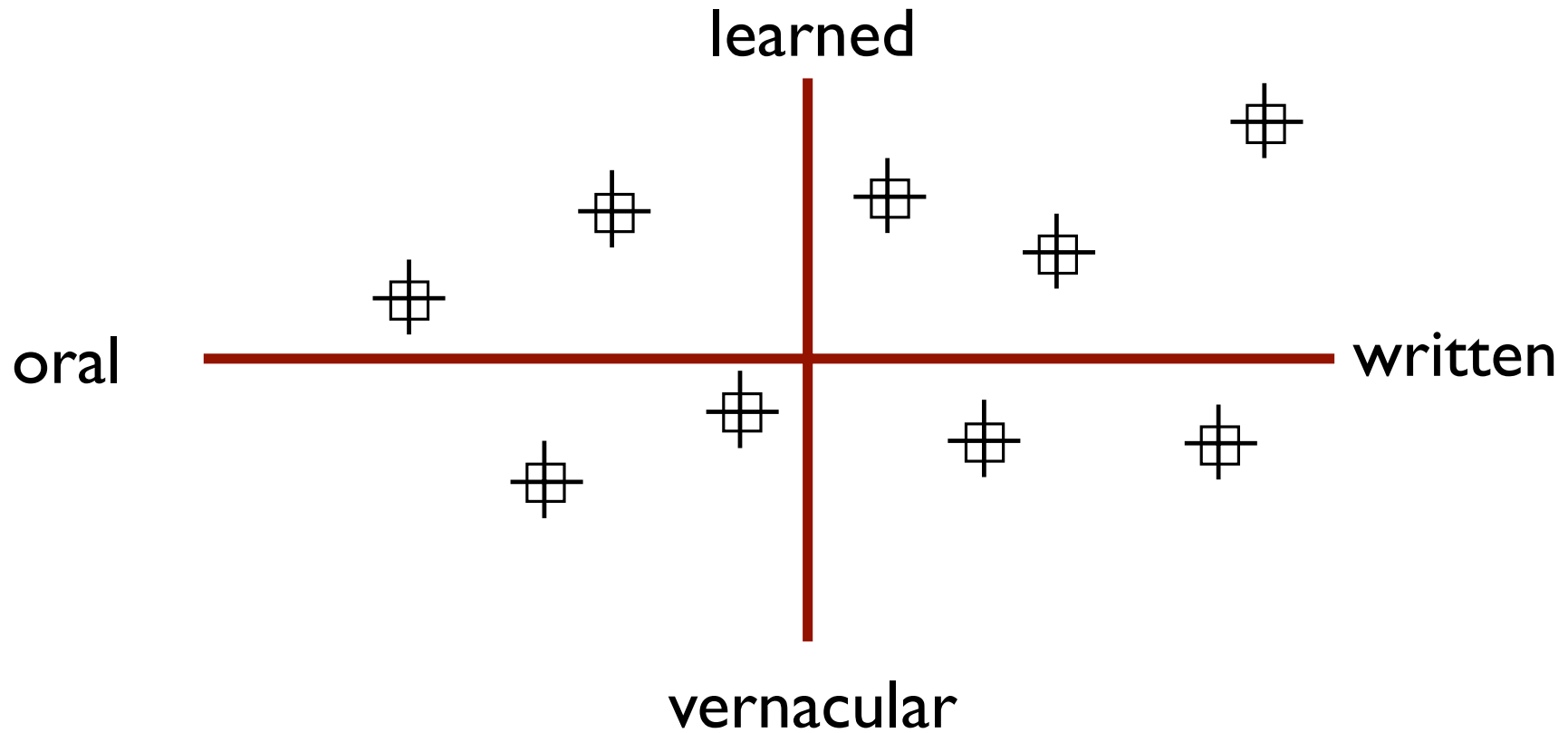
Different editorial
practices and
interpretations

text differs
from edition to edition

Imperfect information

- “Fuzzy” dating of manuscripts (ante / post terminus) and texts (“surely after 1541, not before 1590”)
- Alternative dating – certainty & precision
- Unknown provenance (areal distribution?)
- Unknown authors

Linguistic variation



Interpretation is everywhere

- Transcription of manuscript
 - Collation of manuscripts and creation of edition
 - Linguistic annotation of the edited text
 - Interpretation of collected annotations
- There is no such thing as a “uniform” opinion

Distribution of “raw” Grammar material

- A very small percentage of collected excerpts will be used in the published Grammar
- Why not make the rest available? (Creative commons)

The next stage?

Attitudes towards data distribution

- Reluctance to distribute “unchecked” material
- Tendency to consider collected material own property
 - Whose intellectual property are the annotations?
- Who will have access to my data?

Dissemination challenges

- “Data” are fairly useless without documentation
 - How do I document the appraisal procedure?
- Are “raw” database data (flat text or XML files) more important from the theoretical model and the design elements?
- Longevity guarantee?

How about Web 2.0?

Editions and reference material are available online

The screenshot displays the 'Simple Text Search' interface. On the left is a blue sidebar with 'User Profile', 'TLG Texts Settings' (Results per page: 5, Lines of context: 3, Previous links: No links, Beta excerpts: Show, With diacritics?, Order by date?), 'Full Corpus Search' (Simple, Advanced), and 'TLG Canon Keyword Search' (Search by Author, Work, Publication). The main area is titled 'Simple Text Search' with a 'Help' icon. It features a search input field containing 'θα', a 'switch to betacode input' checkbox, and a 'NEW Lemmatized Search Engine (BETA 1.0)' label. Below the input are buttons for 'Word Index', 'Textual Search', 'Advanced', and 'Start Over'. The search results section shows 'Results for θα (searched 10 out of 136 instances)'. Two results are visible: result 6 and result 7, both from 'Astrologica, Zodiologium (e cod. Ath. Soc. Hist. 210, fol. 38v), {4350.099} Volume 10 page 213'. Result 6 shows the text 'Ὁ γεννηθεὶς νεώτερος ἐν καιρῷ ζωδίου τοῦ Κριστοῦ ἐπὶ γ' ἀσ- (4) θενήσῃ· καὶ θα σκορπίση τὸ πρᾶγμα τῶν γονέων του· καὶ θέ (5) νὰ εἶναι ὑπὸ πολλῶν ἀγαπημένους· καὶ θέ νὰ τὸν σουρεύσουν εἰς'. Result 7 shows the text 'νὰ εἶναι ὑπὸ πολλῶν ἀγαπημένους· καὶ θέ νὰ τὸν σουρεύσουν εἰς γυναῖκα ἕξαναπανθρευμένην· καὶ θα τοῦ πονέσουν τὰ μάτια του· γυναῖκα θέ ν' ἀγαπήσῃ, νὰ ζημιωθῇ ἐξ αὐτῆς, καὶ ἂς βλέπεται καλά.'

Επιστημολογικό Κριτήριο

- Πληροφορίες
- Αναζήτηση
- Σύνθετη Αναζήτηση
- Αποσπασίματα Στίχων

Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω

μακελλείο

Βρες

Επιλεγείς αναζήτησης

 Αναζήτηση και στα σώμα των λημάτων

κατά

 Πρόσθετ. Αίτιαση

Αναζήτηση για: μακελλείο

1 εγγραφή

« < > »



μακελλίων το- μακελλίου- μακελλειού- μακελλειών.

1) Σφαγείο, κρεοπωλείο:

εμφάνιζονταν σαν πρόβατα στο μακελλειό κατά (Διακριούτ. 103⁷).

τα μακελλειά γινεύονται και παίρνει κρέας (Γεωργολ., θαν. 556- Πρόβλ. Η 273-31 γρ. Ρ ερωτ. υπ).

2) (Μεταφ.) μεγάλη εργασία ανθρώπων:

ανδραγαθίας, μακελλειά, χίσις πολλών αιμάτων (Αχιλ. Πρόλ. 34).

Φρ.

1) δίψα του μακελλείου = προορίζομαι για θάνατο, είμαι μελλοθάνατος

(Αλεξ. 737).

2) Κάνω κ. μακελλείο, βλ. κάρνω Φρ. 61.

+ Η λ. σε ποίηση:

(Σουμμ., Ρημτικ. 175).

[-αρχ. ουσ. μακελλών ή μγην. ουσ. μακελλος + κατάλ. -είον. Ο τ. -ειό στο Βλάχ. και σήμ. γρ. -ἰ-ἰ. Η λ. τον 4.-5. αι. και σε Γλωσσόφρ.]

< Προηγούμενο > >> Επόμενο >

Μεταβείτε στη σελίδα: Βρες

Why can I embed video & sound on my site and not chunks of text?

The screenshot shows the TLG Simple Text Search interface. On the left is a blue sidebar with navigation options like 'User Profile', 'TLG Texts Settings', and 'TLG Canon Keyword Search'. The main area is yellow and contains a search bar with the text 'θα', a 'switch to betacode input' checkbox, and buttons for 'Word Index', 'Textual Search', 'Advanced', and 'Start Over'. Below the search bar, it indicates 'Your browser is currently in Greek input mode.' The search results section shows 'Results for θα (searched 10 out of 136 instances)'. Two results are visible, both from 'Astrologica, Zodiologium (e cod. Ath. Soc. Hist. 210, fol. 38v), {4350.099} Volume 10 page 213 line 5. (Browse)'. The first result shows a snippet of Greek text with a red highlight on 'θα' and a small video icon. To the right of the snippet is a video embed box with a URL, an 'Embed' button, and a 'Customise' link. The second result shows a similar snippet with a red highlight on 'θα' and a small video icon, with a corresponding video embed box to its right.

User Profile

Simple Text Search

NEW Lemmatized Search Engine (BETA 1.0)

Search for: switch to betacode input

Your browser is currently in Greek input mode.

[Word Index](#) [Textual Search](#) [Advanced](#) [Start Over](#)

Search Results

Results for **θα**
(searched 10 out of 136 instances)

6. *Astrologica, Zodiologium (e cod. Ath. Soc. Hist. 210, fol. 38v), {4350.099} Volume 10 page 213 line 5. (Browse)*

Ὁ γεννηθεις νεώτερος ἐν καιρῷ ζωδίου τοῦ Κριοῦ ἐπὶ γ' ἀσ- (4)
θενήσῃ· καὶ **θα** σκορπίση τὸ πρᾶγμα τῶν γονέων του· καὶ **θὲ** (5)
νά εἶναι ἐπὶ πολλῶν ἀγαπημένος· καὶ **θὲ** νὰ τὸν σουρεύουσιν εἰς

URL

Embed [Customise](#)
<object width="425" height="355"><param name="mov

7. *Astrologica, Zodiologium (e cod. Ath. Soc. Hist. 210, fol. 38v), {4350.099} Volume 10 page 213 line 7. (Browse)*

νά εἶναι ἐπὶ πολλῶν ἀγαπημένος· καὶ **θὲ** νὰ τὸν σουρεύουσιν εἰς
γυναῖκα ἕξαναπανθρευμένην· καὶ **θα** τοῦ πονέουσιν τὰ μάτια του.
γυναῖκα **θὲ** ν' ἀγαπήσῃ, νὰ ζημωθῇ ἐξ αὐτῆς, καὶ ἄς βλέπεται καλά.

URL

Embed [Customise](#)
<object width="425" height="355"><param name="mov