

**Building a database  
for the compilation of the**



***Corpora and Language***

An Arts and Humanities Research Council Workshop  
Faculty of Classics, Cambridge

12-13 January 2006

Notis Toufexis • nt262@cam.ac.uk



Arts & Humanities  
Research Council

<http://www.toufexis.info>



UNIVERSITY OF  
CAMBRIDGE

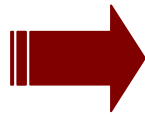
# Compiling a Grammar for an extinct language

*Download*

bibliographical  
search



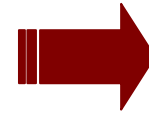
textual sources



linguistic  
theory /  
tradition  
*from*



linguistic  
labelling



presentation  
of results =  
'Grammar'

uses a theoretical model

<http://www.toufexis.info>

# “Model of a Grammar”

## *Download*

---

- Linguistic features that should be grouped and analysed together
  - Abstract concepts based on linguistic theory
    - Terminology to refer to them
- Specific to Medieval Greek:
  - Variation and its interpretation
  - The role of register, of editorial practices etc.
  - Diachrony vs synchrony
  - ...

*<http://www.toufexis.info>*

# What do we model?

*Download*

---

- Our 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
  - The database represents a model of Grammar itself
    - It records and labels primary data as defined by the theoretical model

*from*  
<http://www.toufexis.info>



# Entities

*Download*

---

- Literary texts

- EDITION

- TEXT

- AUTHOR



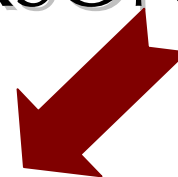
- Non-literary texts

- PUBLICATION

- DOCUMENT

- PERSON

*from*



**TEXT EXCERPT**

*<http://www.toufexis.info>*

# TEXT EXCERPT: first interpretative step

*Download*

---

- EXCERPTS record at an individual source level linguistic phenomena
  - These phenomena are defined by the model of our Grammar
  - The choice of excerpts *from* is based on an act of interpretation by the researcher
- The full list of EXCEPTS collected for each phenomenon will be used in compiling the Grammar

*<http://www.toufexis.info>*

# Compilation of the Grammar: second interpretative step

---

*Download*

- Presenting the collecting excerpts into one “narrative”: the Grammar in book form
  - How about contradictory evidence?
    - False description of phenomenon in secondary literature
    - Different individual *from* interpretations by members of the team
    - Inadequacy of the used model itself →  
Revision
  - Rich excerpts in the book will allow personal interpretation by the user

*<http://www.toufexis.info>*

# Searching and browsing

*Download*

---

- The relational model allows for the presentation of the material linked, selected and ordered in many ways
  - The database can be used for developing a sense of the material as a whole and pointing to spots where further examination is needed

*from*  
<http://www.toufexis.info>

# Look & feel of the database

Download

The screenshot displays the Toufexis database interface. The main window title is "Text excerpts: non literary documents". Below the title, there are search filters and a list of results. The first result is a private letter from an unknown source, dated 1499/09/29. The interface includes a sidebar with navigation options, a main content area with search results, and a right-hand pane with a list of linguistic phenomena. A large watermark "From" is overlaid on the center of the screenshot.

http://www.toufexis.info

# Beyond the book form: making the Grammar available electronically

---

- Text-driven approach:
  - Make the “text” of the Grammar available on the Internet
    - XML-markup for representation of structured information
      - Good examples of the possibilities in Perseus

<http://www.toufexis.info>

# Beyond the book form: making the Grammar available electronically

---

- “Structured information”
  - OHCO: Ordered hierarchy of content objects
    - Hierarchy for presentation
      - Model of publication
    - Hierarchy for interpretation
      - Model of conceptual model

<http://www.toufexis.info>

# Beyond the book form: making the Grammar available electronically

---

- Linguistic phenomena found in Medieval Greek go back to Classical and Koine Greek
  - Linking on the basis of linguistic items not always possible / desirable

<http://www.toufexis.info>

# Beyond the book form: making the Grammar available electronically

---

- Alternative approach: make the model / the categories available electronically
  - Expressed as standoff markup that can be linked to
  - Can incorporate different theory-driven and multilingual terminology
  - Would allow markup of phenomena in a diachronic perspective

<http://www.toufexis.info>

# Example: Liquid interchange

---

*Download*

- FileMaker Pro: Grammar of Medieval Greek
- Kühner-Gerth@Perseus
- An external taxonomy of phenomena would allow linking of all relevant discussion of this phenomenon in a diachronic perspective

*from*

<http://www.toufexis.info>