

Modern Greek Dialectology and the Grammar of Medieval Greek

IO MANOLESSOU
Academy of Athens
manolessou@academyofathens.gr

0. Introduction

The Grammar of Medieval Greek is a five-year research project (2004-2009) whose aim is to describe in detail, for the first time, the Greek language between the years 1100-1700. As such, it is not a dialectological project *per se*, nor was its infrastructure initially geared towards the coverage of dialectal data; however, given its spatiotemporal coverage and the lack of previous systematic research in these domains, it has in fact become the largest project in Greek historical dialectology, as well as the largest digital repository of historical dialectal material. It is for this reason that this short presentation, which has no claims on originality, is included in the present volume (constant reference is made to various publications where the points touched upon here are discussed in detail).

In what follows, a short overview of the project (participants, aims, scope) will be given, followed by a description of its infrastructure (corpus of data, databases) and its contribution to Modern Greek dialectological research, on the basis of concrete examples.

1. The Grammar of Medieval Greek project

The Grammar of Medieval Greek project is located at the University of Cambridge, Faculty of Modern and Medieval Languages¹¹, and is funded by the Arts and Humanities Research Council. Its duration, as already mentioned, is 5 years, 2004-2009- this means that the basic research has already been finished and the results are in the process of being written up. By 2011 it is projected that these results will be published, by Cambridge University Press, in a volume entitled *A Reference Grammar of Medieval and Early Modern Greek*, with 6 authors, which are also the 6 members of the project: Professors David Holton and Geoffrey Horrocks (principal investigators) and Tina Lendari, Notis Toufexis, Marjolijne Janssen and Io Manolessou. The work of the project is assisted by an Advisory Board of international scholars, experts in the field of Medieval Greek texts, which convenes once a year.

The scope of the project is the description of the Greek language between 1100 and 1700, which covers roughly the time-period covered by the Dictionary of Medieval Greek Vernacular Literature of Emmanuel Kriaras (Kriaras 1967-). For this 6 century span, the aim is to describe the language of all Greek-speaking areas, i.e. from Italy to the West to Eastern Asia Minor in the East, Albania and Bulgaria in the North to Cyprus in the South. The project attempts to document this chronological and geographical variation, and track the spread and distribution of specific phenomena across these two parameters, time and space. This is the applicability of the project for the research on Modern Greek dialects: since, for the period under investigation, there is no standardized vernacular language, inevitably description is given by area. The end result is not "the Medieval Greek language" with local variants here and there, but a large diachronic and diatopic map of phonological, morphological and syntactic isoglosses.

2. Temporal coverage

The periodisation of the phase(s) of the Greek language between the end of the Koine and the Modern era is a notoriously difficult issue, and at times various scholars have proposed different dates both for its beginning (some placing in the 3rd and some in

¹¹ See www.mml.cam.ac.uk/Greek/grammarofmedievalgreek and Holton (forthcoming).

the 6th c.) and for its ending (some placing it in the 15th and some in the 17th c)¹². For the purposes of the project, the time-frame of Medieval Greek coincides with the limits defined by the lexicon of Kriaras (1967-) and therefore encompasses those texts written between the 11th and the 18th c., although notice is taken of developments dated close before and after those limits. The final product is therefore more accurately termed "Grammar of Medieval and Early Modern Greek", treating as it does a large corpus of texts from the 16th and 17th c. The reasons for the adoption of this time-frame are the following:

a) The aim to describe, as far as possible, the evolution of the Greek language during the medieval period. Therefore, primary weight is accorded to the type of texts which is usually termed "vernacular" i.e. close to spoken language and not imitating learned, archaising models. The abundant sources written in high registers during this period are only examined for comparative/corroborating purposes. Crucially, vernacular, "low-register" Greek texts become available only after the 12th c., with the rise of vernacular literature¹³.

b) The period prior to the 10th c. is, if not well, at least tolerably covered by the Grammar of non-literary papyri of the Roman and Byzantine periods (Gignac 1976-1981)- it is the language of the subsequent centuries for which no modern grammatical descriptions are available. In fact this was the main reason the project was proposed: the strange fact that although Greek is one of the most well-studied languages, both in its ancient form and in its modern one, for this specific period there simply does not exist any comprehensive grammatical/ linguistic description (Manolessou 2008, Holton & Manolessou 2010 and references therein).

c) the chosen time-period is a time of great linguistic variation, both chronological and geographical, which needs careful documentation and constitutes a necessary step for the investigation of the history of Modern Greek and the Modern Greek dialects. So in fact the project hopes to contribute not only to the research on Medieval Greek but also to the genesis of Modern Greek, constituting a necessary platform for the a future Historical Grammar of Modern Greek (this is another serious lack in the research on the Greek language).

3. Textual coverage

In order to ensure the reliability and authenticity of the data under examination, and to document, as fully as possible, the processes of spatiotemporal spread of the various phenomena, great importance has been accorded to the investigation of non-literary texts (notarial documents, wills, deeds of sale, marriage contracts, private letters etc.). The Grammar of Medieval Greek has prioritized research on such texts, for a variety of reasons: First of all, they are usually dated, named and of known provenance, so the linguistic information they provide is more precise than that of literary texts, which are usually of unknown author, and unspecifiable date and geographical origin. Secondly, they are usually transmitted in a single witness, never copied or copied only once, whereas literary texts in their present form are usually the result of several layers of consecutive copying which has distorted the linguistic picture, introducing features from different times and areas or deleting "outdated" features. Furthermore, non-literary texts are most frequently published as diplomatic editions, which provide an exact picture of the manuscript, spelling conventions, abbreviations etc. without any editorial interventions, while literary texts come to us through the intermediary of the editor, with all its

¹² For more details on the alternative periodisations of Medieval Greek, see Babiniotis (2002: 80-83), Holton & Manolessou (2010: 540-541).

¹³ For the notions "high" vs. "low" register in Medieval Greek see Toufexis (2008), and for the necessity relying on low-register texts for research on language change in Medieval Greek see Manolessou (2008).

advantages and shortcomings (corrections, normalization of orthography etc.)¹⁴. Thus the research conducted by this project is quite different from what has been done in the past concerning the investigation of medieval Greek, which was to a large extent based on the "classic" vernacular literary texts, like the Romances of Chivalry, *Digenis Akritis*, *Ptochoprodromos* etc..

However, the evidence of non-literary texts does not suffice to give a full picture of Medieval Greek, for three main reasons:

a) geographical distribution: Non-literary texts are not available from all Greek-speaking areas. Some areas are over-represented, with hundreds of available primary sources, and some are really under-represented, with almost no available texts at all. Here is a chart showing the geographical distribution of the collected texts so far. The areas under Frankish and Venetian occupation provide very rich legal archives, whereas the areas under Turkish occupation have very little to give. Crete is by far the best documented area, closely followed by the Heptanese and the Cyclades¹⁵.

c) chronological distribution: After the 15th c., non-literary evidence is abundant (mainly from the areas mentioned above). But from the 12th to the 15th non-literary documentation is very scarce. It comes mainly from two areas where a large number of monastery archives are preserved, S. Italy and Athos, which are of no great use concerning distribution, since it cannot be guaranteed that a text written or preserved there was in fact written by a native of the area.

d) genre limitations: some linguistic phenomena, especially in morphology, are very hard to come by in non-literary texts. For example, the 2nd person plural imperfect is extremely rare, especially in the passive- despite examining literally hundreds of texts, the collected attestations remain less than a dozen. First and second person verbal forms, unreal, counterfactual and future formations, genitive plural of feminine and neuter adjectives, are some of the forms that present the greatest difficulty during data collection, due to the nature of the available texts (mainly factual narratives and statements).

Nevertheless, the special attention accorded to texts of ascertainable local provenance has ensured that the Grammar of Medieval Greek is the basic source for anyone interested in the historical dialectology of later Greek: the first attestations of dialectal phenomena, the geographical distribution of linguistic features not common to all forms of Greek, as well as the spread of features which will ultimately form part of Standard Modern Greek can be investigated by using the corpus and the tools developed for the project. It is important to note that systematic research on the history of the Modern Greek dialects has never been undertaken before, and that, in comparison to most Modern languages, Greek is lagging seriously behind¹⁶.

4. The corpus

The textual corpus of the Grammar of Medieval Greek project consists of the following types:

A) .xml, .html, or .doc editions of texts, some, but not all, including apparatus criticus - ca. 2.500.000 words. This is an average-sized historical corpus¹⁷, created from the following sources: i) the TLG (*Thesaurus Linguae Graecae* - www.tlg.uci.edu), which is expanding towards the Medieval period and now contains several literary works (such as the *Chronicle of the Morea*, works of Cretan Drama, etc.) and, more importantly, non-

¹⁴ On the primary importance of datable and geographically localizable texts for linguistic research, and on the value and problems of non-literary sources for the investigation of Medieval Greek in particular, see Manolessou (2001), Manolessou (2008) and Markopoulos (2009).

¹⁵ A rough idea of the statistical distribution of sources by area is provided in Manolessou (2008b).

¹⁶ On historical dialectological research in Greece as compared to other countries see Manolessou (2008b).

¹⁷ For the notion "historical corpus", and the various historical corpora available for Greek as compared to those for other major languages, see Manolessou & Toufexis (forthcoming).

literary texts, such as the ca. 20 volumes of the Archives of Athos ii) other electronic versions of texts available through the Internet iii) texts donated by the modern editors or publishing houses themselves, in a generous gesture towards the project and iv) direct typing/transcribing from the printed editions by the members of the project. Scanning of printed texts and transformation to machine-readable form has been attempted only in a very small scale, as the results of OCR (optical character recognition) for polytonic texts of mixed linguistic form or not following standard orthography has proved unsatisfactory.

This section of the corpus is searchable: it can thus be used to locate specific phenomena, endings, collocations etc. However, because it is neither parsed nor tagged (a huge task which would require a research project on its own) there are limits on the types of searches than can be performed. For example, if one is interested in the "old" 3rd declension genitive inflectional suffix /os/ (e.g. τῆς γυναικός, τοῦ ἀνδρός, but later also τῆς κοπελός, τῆς Πατρός) there is no way to distinguish it from the 2nd declension nominative inflectional suffix /os/ (ὁ ἄνθρωπος, ὁ λόγος etc). Furthermore, because, as already mentioned, the non-literary texts exist in diplomatic editions, and because literary texts exist in many types of accentuation conventions (monotonic, Modern Greek polytonic, Ancient Greek polytonic), there is no unified spelling that one can use for electronic searches, and all alternative variants must be thought of (in this case, -ος, -ως, -οσ, -ωσ, -ὸς, -ὄς, -ὠς, -ώς, -ῶς, -όσ and so on...). Despite its limitations, the electronic corpus has proved invaluable in giving a rough idea of statistical frequency of the various phenomena and features.

B) .pdf texts of mostly non-literary texts, created from the following sources: i) downloading of out-of copyright publications from Anemi- the Digital Library of Modern Greek provided by the University of Crete (www.anemi.lib.uoc.gr) and from the Internet Archive (www.archive.org) and ii) photocopy and scanning of in-copyright printed publications. Because these files are image-file .pdfs, they are not searchable. However, they are useful for quick and remote access to publications, and for storing permanently and together indispensable sources for the history of Greek. In fact the project has created the largest archive of historical dialectal publications in Greece, with hundreds of digitized publications of local documents, many of them published in rare out of print periodicals or books.

C) .tiff and .gif images of medieval literary manuscripts. Thanks to the generous collaboration of an older project undertaken by the University of Sydney (under the direction of M. and E. Jeffreys), one of whose aims was to create microfilms of most major medieval Greek manuscripts, the Grammar of Medieval Greek is in possession of digitized image versions of several important manuscripts, indispensable for checking the validity of editions.

D) The project is also in possession of a considerable body of texts NOT in electronic form, i.e. just printed books or photocopies- a small library dedicated to research on Medieval Greek.

5. The database

The database of the Project can be divided in two parts. The first part is bibliographical: an effort has been made to create exhaustive bibliographies (searchable through keywords) on the following topics:

A) publications of literary texts. The aim is to create a Register of Authors and Works of the period under investigation, complementing that of the Kriaras Medieval Dictionary and providing: Standardised English language abbreviations for all works, information on the various alternative, old and new editions of texts, and basic information on each text (verse, genre, dating etc.).

B) Publications of non-literary texts: considerable effort has been expended in locating all available publications from the various areas of the Greek speaking world.

Using the Project's bibliography, one can easily locate all historical documents from a specific geographical area one is interested in¹⁸.

C) Linguistic publications pertaining to specific phenomena or areas of the period under investigation. Several of these are also available in scanned .pdf format.

The second, and most important part of the database, is the tool for excerpting phonological, morphological and syntactic phenomena. The project researchers go through the texts which constitute the corpus and excerpt chunks of text which are representative of a list of pre-determined phenomena and features. The project Database contains ca. 32.000 textual excerpts, all providing detailed linguistic and source information, and allowing comparisons as to date, place, lemma or linguistic category. It can be used to search for and group phenomena according to author, period, area, linguistic environment, grammatical category or lemma (lexical item). This electronic tool has facilitated and speeded up the project's research considerably. However, problems still remain:

a) **Skewed representation** (chronological and geographical). As discussed above, only areas under Frankish or Venetian control provide sufficient documentation- they heavily outweigh evidence from Turkish-occupied areas such as Macedonia, Thessaly etc. Similarly, the earliest centuries (11th, 12th) are almost exclusively represented by the two large archive depositories of Athos and S. Italy. Thus the absence of a phenomenon or feature from the corpus and the database does not by definition entail its absence from the area or period in question.

b) **obscurity of written sources**: Due to the very nature of historical linguistic investigation, which is based exclusively on written texts, linguistic information is frequently obscured by the nature of the written record. Thus, phonetic information is undetectable through the spelling (for example, it cannot be determined from the orthography whether the strong palatalisation, known as "tsitakismos" of velar consonants represented as <κ> or <τσ>, is in fact [c], [ts], [tʃ] or something else). Also, conservative spelling may conceal phonetic evolutions, such as the deletion of final /n/ or synizesis. Morphological and syntactic information is concealed through the conscious effort of authors/scribes to achieve a more archaic style or to avoid strongly characterized dialectal features.

c) **No statistical data available**. As already discussed, the corpus of the project does not lend itself to large-scale computerized searches (since the texts are not parsed or tagged, and several of them are in fact image files). And the texts constituting the corpus have been only partially excerpted, i.e. only representative samples have been recorded in the database- they are not entered word-by-word. Therefore, the information provided in the final grammar will be approximative only as far as frequency is concerned, with general descriptions such as "rare", "frequent", "absent so far from this type of text" etc.

6. A Sample of historical dialectological research

The following constitutes a sample of the work than can be done using the corpus, the tools and the methodology described above. The phenomenon in question, dental palatalisation, is dialectally restricted, and appeared at some point during the Medieval or Early Modern period.

[n] > [ɲ] and [l] > [ʎ] before [j, i]. The dental nasal /n/ and the lateral /l/ undergo palatalisation before front vowels and semivowels. One may distinguish palatalisation before [j], a semivowel resulting from synizesis of /i/, which occurs in all areas displaying the phenomenon of synizesis and is datable accordingly, and palatalisation before the front vowel /i/ which is dialectally restricted.

Evidence for the first type of palatalisation, before [j], is difficult to establish, since the Greek alphabet has no way of denoting it. However, because in Modern Greek it

¹⁸ The work done by the Research Centre for the History of Greek Law of the Academy of Athens has been an invaluable aid in the compilation of this bibliography.

appears in all dialects as well as in the Standard language (Newton 1972: 137), it must be comparatively early, perhaps simultaneous with the phenomenon of synizesis. Thus, spellings denoting synizesis through change of <ε> to <ι> or accent shift after /l/ and /n/ in all probability (but without certainty) already involve a palatalised [ʎ] or [ɲ]:

πλιόν BERG., *Apokopos* A 459 (Vejleskov)
τῆς ἐλιές / τῖς ἐλιές (1573, Ikaria, TSELIKAS 2000a: 2, 18.15)
λιόντα *Cypr. Canz.* 1.1 (Siapkaras-Pitsillidès)
νιούτσικε TRIV., *Ist. Re Skotsias* 270
τὴν νιότην [LIMEN.], *Than. Rod.* 466 trans. (Lendari)
βουνιά FALIER., *Thrin. Path. Stav.* 223
γονίους MACH., *Chron.* V 65.16 (Pieris/Nikolaou-Konnari)

Furthermore, in Crete and Cyprus the spelling <γν> and <γλ> is occasionally used in order to denote a palatalised sound, in imitation of Italian and French <gn> and Italian <gl> respectively, thus providing more direct evidence of the phenomenon:

ἀναμεγλιά TROILOS, *Rodol.* 2.465 (Aposkiti)
Ἐγλιάν / Ἐγλιάν (1679, Cyprus, PERDIKIS 1998: 16, 41.17)
περβογλοῦ (1699, Cyprus, PERDIKIS 1998: 39, 95.6)
καγλιοντουνα / καλλιόν νά *Cypr. Canz.* 53.8 app. cr. (Siapkaras-Pitsillidès)
ιστιν κεριγιαν / εἰς τὴν Κερύγνεια MACH., *Chron.* R 100.37
εγνια / ἐγνιά (1640, Cyprus, PERDIKIS 1998: 4, 11.9)
εἰς τα χρόγνια / εἰς τὰ χρόνια (1642, Crete, PAPADOPOULOS/FLORENDIS 1990: 21, 16.41)

Another graphematic indication of the existence of palatalised [ɲ] involves /m/ followed by the semivowel [j] in cases of synizesis: the combination [mj]+V results in [mɲ] + V, through consonantisation of the palatal semivowel to a palatal nasal. This outcome is frequently spelt <μνι> in Cretan literary and non-literary texts, and rarely in texts from the Cyclades and Cyprus:

ἐπεθυμνιάς FALIER., *Ist. On.* 75
ἐχλώμνιαεν *Thysia Avr.* 198 app. cr. (B)
μνιαν hora / μνιάν ὥρα CHORT., *Erof.* I. 408 trans. X (Legrand)
μνιά νύκτα / μιὰ νύκτα *Rim. kor.* (A) 2 app. cr. (Caracausi)
ἀσίμνια / ἀσήμια (1532, Crete, KAKLAMANIS/LAMBAKIS 2003: 152, 274.25)
τὰ κορμνιά μας (1549, Crete, MARMARELI/DRAKAKIS 2005: 9, 10.15)
νὰ ζημνιώση MACH., *Chron.* V 316.2 (Dawkins)
τοὺς Ρομνιοὺς / τοὺς Ρωμιοὺς (1614, Tinos, HOFMANN 1936: 1, 59.41)

Evidence for the second type of palatalisation, before [i], comes from the testimony of grammars of the period. Thus Girolamo Germano (PERNOT 1907: 51) and Simon Portius (MEYER 1889: 9-10, 88) state that in certain Greek-speaking areas the sounds [l] and [n] are pronounced like Italian <gli>, <gni>, giving the examples σώνει < sógni>, and παρακαλεῖ <paracagli>. Unfortunately it is not specified which areas present the phenomenon (apart from Chios) but it is emphasised that it is a dialectal phenomenon best avoided.

Direct evidence of the phenomenon is provided by the spellings <gn>, <gn> in the Latin alphabet and <γλ>, <γν> in the Greek, which are quite frequent in Crete:

thegli / θέλει CHORT., *Erof.* I.379 trans. X (Legrand)
i angegli / οἱ ἀγγέλοι *Thysia Avr.* 6 trans. M
ossa bugli / ὠσάν πουλι *Pal. N. Diath.* 3421 app. cr.

na to pglithigni / νὰ τὸ πληθύνει FOSK., *Fort.* I.9 app. cr.

ftignia / φτήνεια (1653, Crete, PANOPOULOU 1991: 2, 429.21)

mu fagnisticchie / μοῦ 'φανίστηκε *Thysia Avr.* 593 app. cr. (M)

tu eogniu / τοῦ αἰωνίου (1661, Chandax / Crete, MAVROMATIS 1986: 10, 102.1)

ci gnictes / τσι νύκτες CHORT., *Erof.* I.403 (Legrand)

narghignisso / ν' ἀρχινήσω FOSK., *Fort.* I.97 trans.

For other, published examples of historical dialectological work conducted with the means provided by the Grammar of Medieval Greek project, see Manolessou & Toufexis (2009), which discusses the phenomenon of the change of /l/ to /r/ and vice versa in Medieval and Modern Greek and Manolessou (2010), which provides a detailed account of the Medieval form of the Cypriot verb system.

7. Conclusion

The Grammar of Medieval Greek project has been the means of creating several primary tools for the investigation of the history of later Greek: bibliographies, textual corpora, and databases containing annotated examples. This material is currently being used in order to compile the *Reference Grammar of Medieval and Early Modern Greek*, but it is hoped that it will also be used in the future for research on the history and analysis of the Modern Greek dialects.

References

- Babiniotis, G. (2002): *Συνοπτική ιστορία της ελληνικής γλώσσας*. 5th edition, Athens.
- Gignac, F. T. (1976-1981). *A Grammar of the Papyri of the Roman and Byzantine Periods*. Milano: Cisalpino-Goliardica.
- Holton, D. (forthcoming): "The Cambridge Grammar of Medieval Greek Project: Aims, Scope, Research Questions". In G. Mavromatis, ed., *Neograeca Medii Aevi VI*, Ioannina, October 2005.
- Holton, D. & I. Manolessou (2010): "Medieval and Modern Greek". In E. Bakker (ed.), *A Companion to the Ancient Greek Language*, Oxford: Wiley/Blackwell, 539-563.
- Kriaras, J. (ed.) (1967-). *Λεξικό της Μεσαιωνικής δημόδους γραμματείας*, 1100-1669, vol. 1-15. Thessaloniki.
- Manolessou, I. (2003): "Οι μη λογοτεχνικές πηγές ως μαρτυρίες για τη γλώσσα της μεσαιωνικής περιόδου". *Λεξικογραφικόν Δελτίον* 24: 61-88.
- Manolessou, I. (2008): "On Historical Linguistics, Linguistic Variation and Medieval Greek". *BMGS* 32: 63-79.
- Manolessou, I. (2008b) «Μεσαιωνική γραμματική και ιστορική διαλεκτολογία: παρατηρήσεις με αφετηρία την κυπριακή διάλεκτο». Στο *Πρακτικά του Ε' Διεθνούς Συνεδρίου Νεοελληνικής Διαλεκτολογίας*, Αθήνα: Ακαδημία Αθηνών, 425-448.
- Manolessou, I. (2010) "Κυπριακή γραμματική Γ': Ρήματα". Στο Α.Βοσκός, Δ. Γούτσος & Α. Μόζερ (επιμ.), *Η ελληνική γλώσσα στην Κύπρο από την αρχαιότητα ως σήμερα*, Αθήνα: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, 68-87.
- Manolessou, I. & N. Toufexis (2009): "Phonetic Change in Medieval Greek: focus on liquid interchange". In *Proceedings of the 8th International Conference on Greek Linguistics, Ioannina, August 30th- September 2nd 2007* [e-book], 291-306.
- Manolessou, I. & N. Toufexis (forthcoming): "Corpus Linguistics in historical dialectology: a case-study of Cypriot". In *Proceedings of the 3rd International Conference on Modern Greek Linguistics and Linguistic Theory*, Nicosia.
- Markopoulos, Th. (2009): "Η έρευνα της γλωσσικής αλλαγής στα κείμενα της Μεσαιωνικής Ελληνικής: μεθοδολογική διερεύνηση των πηγών". In *Proceedings of the 8th International Conference on Greek Linguistics, Ioannina, August 30th- September 2nd 2007* [e-book], 987-998.
- Meyer, W. (ed.) (1889): *Simon, Portius, Grammatica Linguae Graecae Vulgaris*, 1638. Paris: Vieweg.
- Newton, B. D. (1972): *The Generative Interpretation of Dialect*. Cambridge: CUP.
- Toufexis, N. (2008): "Diglossia and Register Variation in Medieval Greek." *BMGS* 32: 203-19.
- Pernot, H. (ed.) (1907): *Girolamo Germano, Grammaire et vocabulaire du grec vulgaire, publiés d'après l'édition de 1622*. Fontenay-sous-Bois (Seine).