



**EURALEX XIX**  
Congress of the  
European Association  
for Lexicography

**Lexicography for inclusion**

**7-11 September 2021**  
Ramada Plaza Thraki  
Alexandroupolis, Greece

[www.euralex2020.gr](http://www.euralex2020.gr)

**Proceedings Book  
Volume 1**

Edited by Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

**EURALEX Proceedings**

ISSN 2521-7100

ISBN 978-618-85138-1-5

Edited by: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

**2020 Edition**

# Creating a DTD template for Greek dialectal lexicography: the case of the Historical Dictionary of the Cappadocian dialect

Karasimos A.<sup>1</sup>, Manolissou I.<sup>1</sup>, Melissaropoulou D.<sup>2</sup>

<sup>1</sup> Academy of Athens, Greece

<sup>2</sup> Aristotle University of Thessaloniki, Greece

## Abstract

This article reports on the compilation of a full dictionary, both print and digital, of Cappadocian Greek, one of the major Modern Greek dialects. This bilingual (Cappadocian Greek-Standard Modern Greek) dictionary is one of the products of the ‘DiCaDLand’ dialectological project, funded by the Hellenic Foundation of Research and Innovation (<http://cappadocian.upatras.gr/en>). Its compilation is based on the powerful professional dictionary editing software TLex Suite, after extensive parameterization in order to meet the needs and the particularities of both the project and the dialectal variety in study. More specifically, we present a sophisticated and state-of-the-art e-lexicographic annotation template capable of handling and describing the complex data of an obsolescent and “aberrant” dialect, without written tradition, heavily influenced by language contact (with Turkish), and presenting considerable variation and serve as a model for future approaches to Greek digital lexicography.

**Keywords:** e-lexicography; dialectology; historical e-dictionary; Cappadocian Greek

## 1 Introduction

An increasing number of scholarly publications and media reports have described, in the last few years, an ongoing decline in linguistic variation on a cross-linguistic level, a phenomenon known as language death. Similar gloomy prognostics emerge for the majority of Modern Greek dialects. The status of geographically determined linguistic variation is changing rapidly under the influence of mobility, migration, and the role of mass media. Thus, the documentation and analysis of existing variation is becoming crucial from the point of view of both linguistic analysis and the history of culture. Furthermore, documentation and analysis of non-standard forms within a language continuum allows for further predictions concerning the notion of language and the limits of linguistic variation and change. In this spirit, the preservation of the Asia Minor Greek linguistic heritage through innovative products has become vital. The term “innovative products” refers here to the digital tools, procedures and methodologies which, although gradually starting to be implemented in Cultural Heritage preservation, have not yet become entrenched in this domain in Greece, nor has a scientifically significant role been acknowledged for them. The combination of state-of-the-art Digital Humanities tools with linguistic research on unique and endangered cultural data forms the core of the project presented. The use of digital tools and methodologies is of major importance for the preservation of the Greek dialectal landscape as well, for which no extended use of digital humanities tools has been reported (see however Galiotou, Karanikolas, Manolissou et al. 2014; Galiotou, Karanikolas & Ralli 2018; Melissaropoulou et al. 2015; Themistokleous et al. 2012). The project DiCaDLand, funded by the HFRI (ΕΛΙΔΕΚ), aims at the thorough documentation and study of an Asia Minor Greek linguistic variety, more specifically the Cappadocian dialect. Apart from primary linguistic research, one of the main objectives of the project is to also produce two state-of-the-art major reference works, namely an interactive electronic dialectal atlas (which will constitute the first such effort in the domain of Greek Linguistics) and a comprehensive historical dictionary of the Cappadocian dialects (again lacking until now). The purpose of this paper is to describe the electronic Historical Dictionary of the Cappadocian Dialects (background, aims, methodology, implementation).

The article is structured as follows: Section 2 describes the shift of Greek dialectology in the digital era and its major steps from a lexicographic perspective. Section 3 offers a brief description of the corpus on which the Dictionary is based. Section 4 provides an analysis of the infrastructure of the project, i.e. the DWS used to compile it and the parameterized DTD especially constructed for the purposes of the DiCaDLand project. Section 5 summarizes our results.

## 2 Background

### 2.1 Greek dialectology in the digital era

#### 2.1.1 Dialectal lexicography and e-lexicography in Greece

In current practice, there is an operational-/pre-theoretical distinction between Modern Greek dialects proper, and Modern Greek patois, depending on the degree of “aberrance” from Standard Modern Greek, and on the consequent possibility of mutual intelligibility between standard and dialect speakers. To the first category belong the following dialects: Pontic, Cappadocian (including Pharasa and Silli), South Italian (Grico-Greco), and Tsakonian, and to the

second all the other dialectal varieties of Greek (Cypriot, Cretan, Northern, Cycladic, Old Athenian etc.). Academic-level dialectal dictionaries exist for all major dialects *except* Cappadocian, namely: Papadopoulos (1955-1958) for Pontic, Karanastasis (1984-1992) for S. Italian, and Kostakis (1986-1987) for Tsakonian, while for other dialects of Greek no lexicographical works of similar high academic quality exist, except the still on-going *Historical Dictionary of Modern Greek* of the Academy of Athens (Manolessou & Bassea-Bezantakou 2017; for an overview of MG dialectal lexicography, see Katsouda 2012). This means that Cappadocian is the only major MG dialect that has not yet been investigated on a lexicographical level, something which constitutes an important lack that is constantly felt in the field of MG dialectology. Up to now, this gap is being filled by the smaller “glossaries” used as an appendix to grammatical descriptions of Cappadocian as a whole (Dawkins 1916: 580-663) or of individual Cappadocian dialects (e.g. Andriotis 1948; Costakis 1963; Mavrochalyvidis & Kesisoglou 1960), or by amateur lexicographical endeavours (e.g. Kotsanidis 2006). However, none of them includes the very rich dialectal material that has only recently surfaced, from current 3<sup>rd</sup> generation native speakers, thanks to the research by M. Janse and D. Papazachariou (see below 2.2). Furthermore, these glossaries do not adopt a unified system of presentation (graphematic representation – principles of lemmatization – system of semantic analysis and sense ordering), and can only give a fragmentary picture of the dialect, accessible only with great difficulty even to the specialist reader (scattered in many publications, with non-obvious cross-glossary correspondence between entries).

In general, Greek dialectal lexicography, until the end of the '80s, was mostly in the hands of non-professional linguists, and its aims were more cultural-folkloristic than purely academic-linguistic. There was also a special focus on the diachronic dimension, as it was felt that evidence of the “archaicity” of a dialect and of a closer connection to Ancient Greek would add validation and prestige to an otherwise culturally threatened and undervalued variety. Synchronic and electronic dialectal lexicography are in their first steps, although they could provide an answer to the danger of extinction faced by many Greek dialectal varieties, as well as the lack of adequate funding for large-scale print lexicographical projects.

In the domain of dialectal e-lexicography, most attempts are limited to the digitisation of already extant print dictionaries (retro-digitisation) through scanning, with, and more usually without OCR (due to the special difficulties presented by the Greek alphabet and its various symbols, see below section 4.3.1). Several of these older print dictionaries or glossaries can be found, for example, in digital depositories like the Internet Archive ([www.archive.org](http://www.archive.org)) or especially for Modern Greek, Anemi (<https://anemi.lib.uoc.gr/>). Simple digitisation of a print dictionary, however, is a practice which is still quite distant from the basic principles and presupposition of true electronic lexicography. As more complex attempts in the direction of dialectal e-lexicography, all very recent, one may mention the following (see Karasimos 2019 for an overview): a) the online dictionary “Syntyshes” for the Cypriot dialect, offered by the University of Cyprus (<http://lexcy.library.ucy.ac.cy/sintixies.aspx>, see Katsoyannou & Armotistis 2019), which offers a digital lemma list with search capabilities, as well as sound clips for each word, created through speech synthesis b) the digital tridialectal dictionary of Pontic, Cappadocian and Aivaliot offered by the University of Patras (Karanikolas et al. 2013; Ξυδόπουλος, Δημελά, Μελισσαροπούλου, Παπαναγιώτου & Ράλλη 2015), which offers a sample of 7500 entries from 3 Asia Minor dialects, with complex search capabilities and a wealth of lexicographic information supported by digitised written or oral documentation and c) the retro-digitisation of the *Historical Dictionary of Modern Greek*, both of the Standard Language and the Dialects, offered by the Academy of Athens (<http://repository.academyofathens.gr/kendi/index.php/gr>, see Manolessou & Katsouda forthcoming), which again offers multiple search capabilities for all the entries (A-D) of the print dictionary, as well as the possibility to download a pdf version of a page or a whole volume.

## 2.2 Cappadocian Greek: a short overview

The Cappadocian dialect constitutes a special case within Modern Greek dialectology. First of all, it is the most highly differentiated dialectal variety of Greek, due to the very long time of separation from evolutions involving the rest of the Greek-speaking world (11<sup>th</sup> c.), and to the very strong influence of Turkish. The dialect is often employed in the literature as a prototypical example of ‘heavy borrowing’ in terms of Thomason & Kaufman’s borrowing scale, with reference to ‘overwhelming long-term cultural pressure’ (Thomason & Kaufman 1988: 50). As a result, it presents a high number of unsolved theoretical and diachronic-historical problems, of great linguistic interest, and has attracted international attention (see e.g. Thomason & Kaufman 1988: 215-222; Johanson 2002:104; Winford 2010; Ralli 2009). Secondly, it is a dialect which was until recently considered extinct, after the end of the life-span of the 1<sup>st</sup> generation speakers, relocated in mainland Greece following the compulsory exchange of populations in 1923. However, in recent years it has been discovered (see Janse 2009) that the dialect is still very much alive, retained by 3<sup>rd</sup> generation speakers in several villages in Thessaly, Macedonia and Thrace and therefore, a wealth of data concerning it, not available to earlier research and not taken into consideration in previous studies of the dialect, is for the first time available. The combination of the factors described above entail that research on Cappadocian is at the forefront of modern dialectology in Greece and abroad (see e.g. Karatsareas 2013; Melissaropoulou 2016; 2019a; 2019b; Ralli 2009; Galiotou et al. 2014). Cappadocian has also formed part of two recent major research projects undertaken by the University of Patras, which had as their aim to collect and preserve as much of the old (written) and new (oral) material as possible (Ralli 2015).

Cappadocian had been under Turkish influence from the 11<sup>th</sup> century until 1923, namely until the exchange of populations that followed the treaty of Lausanne, when spoken in Asia Minor (today’s central Turkey), in an area that covered approximately thirty-two communities. From that period, it was spoken in a situation of regressive bilingualism, since Turkish was the dominant language of the political authorities and was spoken by the overwhelming majority of the population in all aspects of life (cf. Vryonis 1971: 457–59). The dialect is subdivided into two basic groups, North and

South Cappadocian (cf. Dawkins 1916) and an intermediate one, namely Central Cappadocian (cf. Janse forthcoming) showing intra-dialectal divergence. The different zones reflect, following Dawkins (1916: 209–211), different degrees of Turkish influence, which can be attributed to the large extension of the area in combination with other demographic and geographic factors, the most prominent of which would have been the presence of the Turkish population in each different community as well as the existence of Greek schools.

### 3 Corpus of data

The sources of the DiCadLand dictionary are twofold: on the one hand, all the available written sources starting from the 19th century onwards (dictionaries, glossaries, linguistic descriptions, collections of primary texts such as folk-tales, songs, narrations, riddles etc.), most of which were collected and digitized thanks to an earlier project, AMiGre, and are available online (<http://amigredb.philology.upatras.gr/>) were taken into account, in order to allow the systematisation of all existing intra-Cappadocian variation. Considerable effort was expended for the homogenization of this material, which came in a multiplicity of transcription systems, especially the older sources. On the other hand, the corpus was considerably enlarged through the addition of oral recordings. A few date back to the 1930s (available online through the depository Gallica of the Bibliothèque Nationale de France), but most of them derive from the very recent and rich dialectal material from current 3<sup>rd</sup> generation native speakers (descendants of Cappadocian refugees) collected the last decade. Special emphasis was placed on the exploitation of this new oral material, so that the dictionary under preparation will not simply be a depository of already available but disparate data, but an opportunity for the presentation of new data. This allows also for a “diachronic” examination of the evidence, as we have the possibility to examine side-by-side data which may be divided by more than 100 years. However, the dichotomy between written and oral sources has given rise to a major problem: whereas older sources are roughly equally distributed with respect to geographical provenance (i.e. data is available for almost all Cappadocian settlements, ca. 20 in number), oral data, from current speakers, are available only from 2-3 major communities, and mostly from that of Misti, which was the largest. This creates an imbalance in the lexicographical treatment of words, phenomena and senses. Another issue requiring special attention is the fact that the older material was in part collected by amateurs, or at a time when linguistic science had not yet been sufficiently developed, and therefore it is to a certain extent less reliable than the oral material, containing many inaccuracies which can no longer be assessed.

### 4 TLex software parameterization in the service of Greek dialectal lexicography

The electronic availability of well-organized lexical material is of quintessential importance for the transformation of dialectal comparative linguistics into a quantitative and collaborative field of research. To this end, the format of the Dicadland dictionary attempts to conform to standards of state-of-the-art academic-level Dictionary Writing Systems, after careful evaluation of available options.

Its realization requires adherence to the most recent advances in the domain of electronic lexicography and dialect mapping on the one hand (see e.g. Granger & Paquot 2012), and historical and dialectal lexicography on the other (see e.g. Reichmann 2012; Manolessou 2016), filtered by digital humanities methodologies.

#### 4.1 DWS and selection criteria

The Historical Dictionary of Cappadocian Greek is being built using the powerful professional dictionary editing software TLex Suite, one of the most widely-used state-of-the-art DWS internationally (for an overview of DWSs see Abel 2012 and for the main features of TLex and its application to dialectal data see Joffe, McLeod & de Schryver 2008; de Schryver 2011). The DTD construction is aided by the experience acquired by the research team in smaller-size and scope dialectal e-dictionaries, such as the tri-dialectal Asia Minor dictionary (Galiotou, Karanikolas, Manolessou et al. 2014; Galiotou, Karanikolas & Ralli 2018). From a typological viewpoint, this dictionary is being structured as a bilingual one due to the Cappadocian dialect’s considerable distance from the standard form of Modern Greek (among others, Geeraerts 1989: 294-295; Bejoint 2000: 39; Marellò 2004:351).

One may consider that there are three available DWS construction options (Krek, Abel, Tiberius 2015), namely: (a.) purchasing a commercial off-the-shelf software/app/platform, (b.) using a free app or open-source web-based platform, and (c.) building from scratch a tailored app/software for one’s own needs. As far as the first option is concerned, the most widely used and tested applications are the following, although a few of them are currently no longer available (see Abel 2012): IDM DPS (Digital Publishing System) <https://www.idmgroup.com/content-management/dps-info.html> (Grundy & Rawlinson 2016), TLex (TshwaneLex) <https://tshwanedje.com/tshwanelex/> (de Schryver 2007; 2011), iLex (Erlandsen Media Publishing) <http://groupbanker.dk/generic-en/index.htm> (Erlandsen 2010), and ABBYY Lingvo Content <http://www.lingvo.ru/content/> (Kuzmina & Rylova 2010). In the second category, one may find quite a large number of freely available programs online, but most are meant for relatively simple lexicographical projects and therefore cannot match the potential and range of professional commercial applications. Some of the free programs with the best capabilities are: Lexonomy (<https://www.lexonomy.eu/>), Dictionary Editor and Browser (<https://deb.fi.muni.cz/index.php>), Matapuna (<https://sourceforge.net/projects/matapuna/>), Dictionary System DWS (<http://dictionary-system.hvalur.org/index.php?lang=en>), and WeSay <https://software.sil.org/wesay/> (Perlin 2012). As for the third category, many academic and research institutions have opted for the construction of a custom-made DWS, usually based on an already extant general database construction program (e.g. Oracle/MySQL/Filemaker) or a general XML editor (e.g. Oxygen, XMetal), parameterized for lexicographic use. Examples include the DWS of the Institute of



Dutch Lexicology for its collection of historical and local dictionaries (Tiberius, Niestadt & Schoonheim 2014), and the DWS of Institute of Czech language at the Academy of Sciences of the Czech Republic (Barbierik et al. 2014).

#### 4.1.1 Selection criteria

The selection criteria between the three alternative options may be summarised as follows:

- a) Cost: Freely available online programs prove superior on the basis of this criterion, as their cost is literally zero. Commercial projects are obviously the most expensive ones, but custom-made programs also come at considerable cost, since the general programs on which they are based are also commercial ones. Cost also depends on the number of licences to be purchased, as well as the option between personal or institutional use. Some programs (such as TLex) offer a special discount for lexicographic work on endangered languages or varieties.
- b) Time: commercial programs fare the best with respect to the time factor, as the long-term experience behind them and their large professional support teams have already solved most of the challenges in the domain of e-lexicography. At the other end of the spectrum, custom-made programs are the most time-consuming, since everything needs to be rebuilt from scratch. Free online programs constitute an intermediate solution, since, as they are designed to serve a wide variety of research aims with various specifications, they may require a high degree of parameterization.
- c) Capabilities: Commercial programs are the ones to offer the widest range of capabilities as compared to free programs. This concerns not only available functions and greater degree of customisation, but also higher storage capacity, easier interconnection with other software programs, and better support (manuals, tutorials, updates). The only great advantage of free programs is that they are web-based, something which allows parallel processing by several users without need for a server or special installations. Custom-made programs are obviously the ones which allow for the widest range of specialised capabilities, limited only by the time and the funds one is willing to invest in their construction.
- d) User-friendliness: with respect to this criterion, each category has its own advantages and disadvantages. More specifically, free programs, being in general simpler and with fewer capabilities, usually come with a relatively uncomplicated and intuitive interface and fewer interactive buttons than commercial programs, which may have quite a complex GUI. On the other hand, commercial programs may be able to perform automatically a great number of routines which may either be impossible in a free program, or may need to be executed manually with multiple repetitions, or with special programming scripts. As a simple example, for text formatting commercial programs offer an environment similar to that of specialised word-processing applications, with a wide variety of options for fonts, sizes, colours, symbols etc., whereas free programs usually have a built-in predetermined and limited range of options, or require the construction of a Cascading Style Sheet (CSS). Furthermore, professional programs standardly allow for multiple viewing alternatives of the dictionary under construction: in database format, in XML format, or, more importantly, as final exported text or front-end interface (with a WYSIWYG window), so the dictionary compiler immediately realises the impact of his choices. In free programs one usually sees only the back-end interface in database view, and the final (usually html) outcome becomes visible only after the completion of the project. Finally, as also mentioned above, free programs cannot offer the strong user support provided by commercial applications, as they are usually unable to go beyond a FAQ page or a users' forum. As far as custom-made programs are concerned, anything is possible, but previous experience has demonstrated that the more special functions a program has, the more complex it becomes, and without the services of a professional software engineer the final product has few chances of being user-friendly.

#### 4.1.2 Proposal

On the basis of the above, it was deemed that the construction of a custom-made program *ab initio* should be avoided. On the one hand it would be extremely time-consuming, given that the DiCaDLand project needs to be completed in a specific time-frame (3 years), and on the other it would force the research team in a direction beyond the main aims of the project, which are primarily linguistic rather than technical. Furthermore, given that the main issues and requirements of electronic lexicography are similar worldwide, any solution arrived at would most probably be a "re-discovery" of already solved problems (a 're-invention of the wheel' in the terms of de Schryver 2007; 2011). Also, the construction of a custom-made electronic tool would hardly be a cost-effective option, as on the one hand the general programs required for its construction (database server, xml writer etc.) would not come free of charge and on the other a technical specialist would need to be hired. After extensive testing, the possibility of using a freely available online program was also rejected, as the options they offered could not cover the range of needs and specifications of the DiCaDLand dictionary.

#### 4.2 Why TLex?

After comparative evaluation, it was decided to construct the Historical Dictionary of the Cappadocian dialects using the powerful TLex suite, which presents several advantages (cf. De Schryver & De Pauw 2007; De Schryver 2011). More specifically, it has already been tried and tested on more than 40 national and academic lexicographic projects, it offers a surprisingly wide range of parameterization for even the most "non-standard" linguistic varieties (including endangered languages), and it offers the possibility of direct export to Microsoft Word (in .rtf format), or, even better, to professional desktop publishing programs such as Indesign and QuarkXpress. Additionally, it offers the possibility of import/export from spreadsheets such as Excel, allowing us to automatically import material previously collected in such formats and very importantly, it provides its own integrated Corpus Query system, which can be used for example extraction on the basis of the transcribed oral corpus.

Furthermore, as de Schryver (2011) has already pointed out, and as discussed above, the TLEx suite, as any other dedicated DWS software as compared to custom-made solutions, guarantees reduced project completion time, thanks to (amongst others): various levels of automation (e.g. automatic cross-reference tracking and updating of homonym and sense numbers, easy entry of any phonetic symbol through macros, fast full-dictionary text search, automatic checking for various dictionary errors, immediate WYSIWYG, full Unicode support, customisable styles (font, colour, etc.) for every field in the dictionary and language of the metalanguage (cf. De Schryver & Joffe 2005a), increased consistency in the treatment of articles, thanks to features such as the article filter, and finally improved teamwork and easy multi-user adaptation.

### 4.3 Creating and parameterizing a DTD for the Cappadocian Dictionary

The DTD of the Cappadocian Dictionary under implementation was constructed through extensive parametrization and customization of the TLEx software, presented in the following subsections. Although the TLEx Suite comes with a built-in template both for monolingual and bilingual dictionaries, it was deemed necessary for our project's needs to construct a new DTD, which would be better adapted to the needs of modern Greek dialectology. The construction and compilation of this dictionary benefits from the established research and the expertise of research conducted at the Research Centre for Modern Greek Dialects of the Academy of Athens for the publication of print dialectal dictionaries, with special adaptations in order to meet the needs of a specific Asia Minor dialectal group as well as digitization.

Element types	Class	List item	Internal ID	List item	Internal ID
Dictionary	Document	Ανακού	44	Bagr.	73
Language	Section/L	Αξός	29	Cost.	70
Lemma	Entry	Αραβανί	42	Dawk.	62
Etymology	User	Αραβισσός	48	Dawk.Song.	63
Sense	User	Αφσάρι	49	Greg.	74
Example	User	Γαριπτσός	50	Karats.	75
Subentry	Subentry	Γούρδονος	43	Lag.	64
Definition	User	Δίλα	51	Αλεκτ.	77
TE (Translation Eq...	User	Κίσκα	53	Αναστασ.	78
References	SmartRe	Καπαδοκία	41	Ανδρ.	67
Form	User	Καρατζάβιραν	52	ΑΠΥ-Βαῖρ.	129
POS	User	Καρατζάρεν	130	ΑΠΥ-ΑΠΘ	138
Example_phrase	User	Μαλακοπή	20	ΑΠΥ-ΕΝΔ	127
Example_Paraim	User	Μισθί	18	ΑΠΥ-Καρατσ.	128
Example_Riddle	User	Ουλαγάτς	28	Αρχέλ.	79
Example_Song	User	Ποτάμια	26	ΕΚΠΑ 2142	95
Period	User	Σίλατα	45	Ελευθερ.	80
Synonym	User	Σύλλη	24	Θεοδ.Παραδ.	81
Comment	User	Σατί	54	Θεοδ.Τραγ.	82
Forms	User	Σεμέντρα	47		
Senses	User	Σινασσός	21		
		Τελμησσός	27		
		Τζαλίλα	56		

Figure 1: Controlled vocabularies from the proposed DTD schema including geodata and sources

In more detail, all major Greek dialectal dictionaries share the same tripartite structure, following the model of the *Historical Dictionary of Modern Greek* (ILNE) of the Academy of Athens, the earliest and largest Greek scientific lexicographic endeavour (Manolessou & Bassea-Bezantakou 2017; Manolessou & Katsouda forthcoming). This structure comprises, apart from the headword, (i) a Forms field, where the variant dialectal forms are set out, with phonetic transcription and geographical distribution (ii) an Etymology field, where the word's origin (native/loanword) and dating are recorded and (iii) a Senses field, with definitions, examples, quotations and documentation from oral and written corpora, and also including "special" types of examples, such as proverbs, songs and riddles.

Additionally, parameterization involved also issues of alphabetic representation and phonetic transcription, as well as the incorporation of extensive bibliographic references and cross-references, as is expected in a fundamentally academic publication. Finally, the parameterization needed to take into consideration the necessity of a parallel print and digital publication and the combination with an online digital dialectal atlas also under preparation by the same project (<http://cappadocian.upatras.gr/en/node/10>), which among other things entailed the switching of the software's metalanguage to Greek and the preparation of alternative attribute lists (full vs. abbreviated).

#### 4.3.1 Alphabet, script, font and encoding

The Greek alphabet and the multiple non-standard alphabetic symbols used in Greek dialectal literature for the encoding of phonetic features absent from Standard Modern Greek (e.g.  $\sigma = [j]$ ,  $\alpha = [\text{æ}]$ ; for an overview see Manolessou, Beis & Bassea-Bezantakou 2012). From a typographical/orthographic point of view, Standard Modern Greek can (and is) very easily rendered through the Greek alphabet, which has a standardized orthography and suffices to represent all the sounds of the standard phonological system. Correspondingly, when it comes to electronic transcription/encoding, Standard

Modern Greek presents no problems, as the Greek alphabet is included in the Unicode Standard, in two blocks (Greek and Greek Extended), and therefore machine-readable texts conforming to XML and TEI standards can easily be produced.

However, Modern Greek dialects include a number of sounds (phonemes/allophones), both vocalic and consonantal, which do not exist in Standard Modern Greek and therefore are non-representable through the Greek alphabet, at least in its standard Unicode form. The representation of these dialectal sounds can of course be effectuated through the International Phonetic Alphabet (IPA), also included in the Unicode Standard. However, the ILNE, as a national Historical Dictionary, the major Greek dialectal dictionaries, and the DiCaDLand dictionary under discussion are not addressed only to the linguistic community, which is familiar with the IPA, but also to students, philologists, historians, specialists in folklore studies and, to a certain extent, to the general public. The exclusive use of the IPA would render these dictionaries virtually unreadable for non-expert linguists, not only in the listing of variant forms, but especially in the examples and quotations section. For this reason, the ILNE has developed, since the first volume published in 1933, a custom system of notation. The form that these special phonetic characters for the representation of Modern Greek dialects take is mostly letters of the Greek or the Latin alphabet modified by diacritics such as the acute, the breve, the dot or the caron.

The special dialectal characters are not included in the Unicode Standard, and therefore are unidentifiable on computers where the special font is not installed. For the representation of these special characters, the solution adopted by the DiCaDLand project was to use only Unicode symbols, as combinations of letter characters + combining diacritics. This of course presents several problems in the visual end result, which have mainly to do with the limited combinatory capabilities of most widely available Unicode fonts which include the Greek Extended block, when it comes to the superposition of diacritic marks. The height and width of most glyphs entails that the diacritic will appear not on top of them, but to their right. Additionally, when a character is modified by more than one diacritic (for example, a vowel + diaeresis + acute accent), it is quite difficult to get the diacritics to stack one on top of the other. Finally, not all fonts include all the diacritics required. Luckily, the latest Microsoft fonts, incorporating the ClearType technology, perform better when it comes to the rendering of combining characters, and so the extent of the problem may be reduced in the future, as technology improves. On this issue see also Armosti et al. (2014). TLex gives the option of keyboard shortcuts (macros) for special characters, and it has been extensively used for the fast keying of special characters

#### 4.3.2 Forms

Lemma: γλυτώνω LemmaSign=γλυτώνω, PartOfSpeech=ρήμα, Modified=2020-05-24 11:46:31, Created=2020-04-28

Forms

- Form: Form\_Gr=γλυτώνω, IPA=ɣli'tono, Geography=Συνασός Φάρασα Φλογητά
- Form: Form\_Gr=γλυτώνω, IPA=ɣli'tono, Geography=Φερτάκινα
- Form: Form\_Gr=γυλώνω, IPA=ɣy'lono, Geography=Ανακού Αξός Αραβάνι Γούρδονος Ουλαγιάς Σίλατα Τελεμα
- Form: Form\_Gr=γυλώνω, IPA=ɣy'lono, Geography=Ανακού Αραβάνι Μισθί Σίλη Φλογητά
- Form: Form\_Gr=γυλώνω, IPA=ɣy'lono, Geography=Σίλη
- Form: Form\_Gr=κουλτώνω, IPA=kul'tonu, Geography=Φλογητά
- Form: Form\_Gr=κουλτώνω, IPA=kul'tono, Geography=Φλογητά
- Form: Form\_Gr=κουλτώνω, IPA=kul'tono, Geography=Σίλατα
- Form: Form\_Gr=κουλτώνω, IPA=kul'tono, Geography=Φλογητά
- Form: Form\_Gr=γουλώνω, IPA=gu'lono, Geography=Ανακού Ουλαγιάς Σίλατα Φάρασα Φερτάκινα Φλογητά
- Form: Form\_Gr=γουρώνω, IPA=gu'rtonu, Geography=Μισθί Σίλη
- Form: Form\_Gr=γούλωνα, IPA=ɣy'ltona, Geography=Αξός Μισθί
  - POS: GramCateg=Παρατατ.
- Form: Form\_Gr=γούλωνα, IPA=ɣy'ltona, Geography=Μισθί Φλογητά
- Form: Form\_Gr=γυλώσσα, IPA=ɣy'lɔsa, Geography=Κίσκα Συνασός Φάρασα
  - POS: GramCateg=Αόρ.
- Form: Form\_Gr=γούλωσσα, IPA=ɣy'lɔsa, Geography=Αξός Αραβάνι Γούρδονος Μισθί Σίλη Φλογητά
- Form: Form\_Gr=γουλώσσα, IPA=gu'lɔsa, Geography=Ουλαγιάς
- Form: Form\_Gr=κούλωσσα, IPA=ku'lɔsa, Geography=Φλογητά
- Form: Form\_Gr=κούλωσσα, IPA=ku'lɔsa, Geography=Φλογητά
- Form: Form\_Gr=γούλωσσα, IPA=ɣy'lɔsa, Geography=Μισθί Φλογητά

γλυτώνω (ρ.) γλυτώνω [ɣli'tono] Συνασ., Φάρασα, Φλογ. γυλώνω [ɣy'lono] Φερτάκι. γυλώνω [ɣy'lono] Ανακ., Αξ., Αραβάνι, Γούρδ., Ουλαγ., Σίλατ., Τελεμ., Φερτάκι. γυλώνω [ɣy'lono] Ανακ., Αραβάνι, Μισθ., Σίλ., Φλογ. γυλώνω-νου [ɣy'lono:u] Σίλ. κουλτώνω [kul'tonu] Φλογ. κουλτώνω [ku'l'tono] Φλογ. κουλτώνω [ku'l'tono] Σίλατ. κουλτώνω [ku'l'tonu] Φλογ. γουλτώνω [gu'l'tono] Ανακ., Ουλαγ., Σίλατ., Φάρασα, Φερτάκι., Φλογ. γουρώνω [gu'rtonu] Μισθ., Σίλ. Παρατατ. γούλωνα [ɣy'ltona] Αξ., Μισθ. γούλωνα [ɣy'ltona] Μισθ., Φλογ. Αόρ. γλώσσα [ɣli'tɔsa] Κίσκα., Συνασός, Φάρασα. γούλωσσα [ɣy'lɔsa] Αξ., Αραβάνι, Γούρδ., Μισθ., Σίλ., Φλογ. γούλωσσα [gu'lɔsa] Ουλαγ. κούλωσσα [ku'lɔsa] Φλογ. κούλωσσα [ku'lɔsa] Φλογ. γούλωσσα [ɣy'lɔsa] Μισθ., Φλογ. γούλωσσα [ɣy'lɔsa] Σίλ. γούλωσσα [ɣy'lɔsa] Σίλ. γούλωσσα [ɣy'lɔsa] Σίλ. Υποτ. γλώσσα [ɣli'tɔsa] Συνασός, Φάρασα, Φλογ. γούλωσσα [ɣy'lɔsa] Αξ., Αραβάνι, Γούρδ., Μισθ., Σίλ., Τελεμ., Φλογ. γυλώσσα [ɣy'lɔsa] Μισθ., Σίλ. κούλωσσα [ku'lɔsa] Φλογ. γούλωσσα [ɣy'lɔsa] Αραβάνι, Μισθ. Προστ. γύτω [ɣy'lto] Ανακ., Συνασός. γούλω [ɣy'lto] Μισθ., Σίλ. γούλω [ɣy'lto] Ανακ., Αξ., Σίλ., Τελεμ., Φλογ. κούλω [ku'lto] Ανακ., Φλογ. γούτω [ɣy'to] Τελεμ. Παθ. γουλιέμαι [ɣy'l'iɛme] Αξ. Αόρ. γούλωσ'κα [ɣy'lɔs'ka] Σίλ. Υποτ. γούλωθώ [ɣy'lto'tho] Φάρασα. γούλωσ'κα [ɣy'lɔs'ka] Σίλ.

Figure 2: Sample lemma: structure (above) and preview (below)

Given that this is a dialectal dictionary of a non-standardized variety, form variation is very common. Virtually all lexemes present several variant forms, as each individual Cappadocian village has its own local variant. This extreme variability, present also in all other Greek dialectal dictionaries, was treated through a complex “Form” element, which includes the option for multiple embedded sub-forms. Each sub-form provides tripartite information: the form in the Greek alphabet with the special diacritics (if needed), the phonetic transcription in the IPA, and the geographical distribution information, available to the compiler as a controlled vocabulary list<sup>1</sup>. Additionally, the compiler can also insert, optionally, bibliographic information on the specific source form which the form is extracted (again, as a controlled list).

#### 4.3.3 Etymology

Etymology is the only element which has been completely left as a free-text field. Unfortunately TEI standardization (<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-etym.html>) and consequently TLEX implementation is not yet finalized in this domain (see e.g. Bowers & Romary 2016), due to the fact that the etymological information that will be required for each word is to a certain degree unpredictable. As it stands, this field provides, in

<sup>1</sup> The DiCaDLand project also includes the creation of a digital map - in the long term, the geographical information provided for each form (and sense, see below) will be visually represented on the interactive geo-dialectal map currently under implementation, using QGIS mapping.



free-text form, the word's etymological provenance (inherited, loanword etc.), its dating (ancient, medieval, modern etc.), its morphological analysis (stem, suffix etc.) and possibly, bibliographic references and cross-references, as well as quotations from old (ancient, medieval) textual sources for documentation purposes. TLex does not provide the possibility to “embed” XML objects within a free text.

#### 4.3.4 Senses

As in the case of “Forms” above, the element “Senses” has the option of multiple embedded sub-senses, with several attributes, following standard lexicographical practice. This possibility was already provided for in the built-in TLex templates. Similarly, the standard templates included basic fields such as “definition” and “example”. The parameterization implemented for our dictionary centered around the presentation of examples. The simple “example” was expanded through several sub-categories. These were “phrases” (idioms and collocations), “proverbs”, “songs” and “riddles”. Each of these categories required special treatment: apart from their literal translation, they also need an extra field for their interpretation (meaning of the proverb, solution of the riddle etc.), and some required special formatting (verses, in the case of songs and occasionally riddles). Furthermore, each definition and each example is followed, as in the case of forms, by geographical information. Finally, all senses are dated as to their first appearance (ancient, medieval etc.) through a controlled list.

The screenshot displays the TLex dictionary interface. On the left, there are navigation panels for 'Cappadocian Greek' and 'Standard Modern Greek'. The main area shows a lemma entry for 'ένος' (enos) with two senses. Sense 1 is 'ένος' (genitive of 'ένας'), and Sense 2 is 'ένος' (adjective). Each sense includes a definition, morphological analysis, and several examples with their source references. The examples are categorized into 'phrases', 'proverbs', 'songs', and 'riddles'. The interface also includes a search bar, a list of attributes, and a corpus view.

Figure 3: Lemma sample with an extensive use of different kinds of examples to document the senses

## 5 Concluding Remarks

The construction of the electronic Historical dictionary for the Cappadocian is meant to contribute significantly to the thorough documentation and study of this dialect, which holds a prominent position not both in Greek dialectology and in international language contact studies. The proposed template provides contributions in several scientific fields. More specifically, in Greek Linguistics, this study constitutes a much-needed holistic approach to the Cappadocian dialect. In *Lexicography*, it produces an online, freely accessible dictionary, following the latest international standards for electronic lexicography (DWS platform, .xml output) and dialectal lexicography, for the first time in Greece. It provides a standard system of transcription, fully compatible with the Unicode standard while at the same time following the principles of the IPA; it implements, for the first time in Greece, new methodologies in dialectal lexicography, applying in them on a dialect possessing many special or unique features rendering lexicographical treatment extremely difficult (e.g. loss of gender, morphological opacity of the vocabulary due to long-term isolated evolution, high level of foreign influence, unreliable primary sources requiring constant re-evaluation). Finally, this effort constitutes an important step for the development of Digital Humanities in Greece and salvages a significant amount of cultural data from an endangered dialect, which is of great value for Greek civilisation and historical memory.

## 6 References

Abel, A. (2012). Dictionary Writing Systems and Beyond. In S. Granger and M. Paquot (Eds.), *Electronic Lexicography*. Oxford: Oxford University Press, pp. 83-106.

- Abel, A., Krek, S., Tiberius, K. (2015). *Bibliography – Dictionary Writing Systems (DWS) & related software*. ENEL Cost Action documentation, available at [http://www.elexicography.eu/wp-content/uploads/2015/04/Bibliography\\_DWS\\_CQS\\_v7\\_web.pdf](http://www.elexicography.eu/wp-content/uploads/2015/04/Bibliography_DWS_CQS_v7_web.pdf) [accessed 05/11/2018]
- Andriotis, N. P. (1948). *Το γλωσσικό ιδίωμα των Φαράσων* [The dialect of Pharasa]. Athens: Ikaros.
- Armosti S., Christodoulou K., Katsoyannou, M. & Themistocleous, C. (2014). Addressing writing system issues in dialectal lexicography: the case of Cypriot Greek. In Dyck *et al.* (Eds), *Dialogue on Dialect Standardization*. Cambridge: Cambridge Scholars Publishingpp. 23-38.
- Barbierik, K., Dēngeová, Z., Holcová Habrová, M., Jarý, V., Liška, T., Lišková, M., Virius, M. (2014). Simple and Effective User Interface for the Dictionary Writing System. In *Proceedings of the XVI EURALEX International Congress: The User in Focus*, pp. 125-136.
- Béjoint, H. (2000). *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.
- Bowers, J. & L. Romary (2016). Deep Encoding of Etymological Information in TEI. *Journal of the Text Encoding Initiative [Online]*, Issue 10 | 2016. URL: [<http://journals.openedition.org/jtei/1643>] [accessed 03/08/2019]
- Chambers, J.K. & P. Trudgill. (1980). *Dialectology*. Cambridge: CUP.
- de Schryver, G-M. (2007). *Oxford Bilingual School Dictionary: Northern Sotho and English*. Cape Town: Oxford University Press Southern Africa.
- de Schryver, G-M. (2011). Why Opting for a Dedicated, Professional, Off-the-shelf Dictionary Writing System Matters. In K. Akasu and S. Uchida (Eds.), *ASIALEX 2011 Proceedings. Lexicography: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference*. Kyoto, Japan, August 22-24, 2011 Kyoto: The Asian Association for Lexicography, pp. 647–656.
- de Schryver, G-M & De Pauw, G. (2007). Dictionary writing system (DWS) + corpus query package (CQP): The case of Tshwane Lex. *Lexikos*, 17, pp. 226–246.
- de Schryver, G-M, Joffe, D., & M. MacLeod (2008). The TshwaneLex Electronic Dictionary System (Software Demonstration). In E. Bernal and J. Decesaris (Eds.), *Proceedings of the XIII Euralex International Congress*. Barcelona: IULA, Documenta Universitaria, pp. 421-424.
- Erlandsen, J. (2010). iLEX, a general system for traditional dictionaries on paper and adaptive electronic lexical resources. *Proceedings of the 14th EURALEX International Congress*, pp. 306-316.
- Galiotou E, N. Karanikolas, I. Manolessou *et al.* (2014). Asia Minor Greek: Towards a computational processing. In *Procedia- Social and Behavioral Sciences 147*, pp. 458 – 466. Elsevier.
- Galiotou E., Karanikolas N. & A. Ralli. (2018). Preservation and Management of Greek Dialectal Data. In Ioannides, M., E. Fink *et al.* (Eds.) *Digital heritage: progress in cultural heritage: documentation, preservation and protection. 7th International Conference Euromed 2018*. Berlin/Heidelberg: Springer, pp. 752- 761.
- Dawkins, R. M. (1916). *Modern Greek in Asia Minor: a Study of the Dialects of Silli, Cappadocia and Phárasa with Grammar, Texts, Translations and Glossary*. Cambridge: Cambridge University Press.
- Geeraerts, D. (1989). Principles in monolingual lexicography. In F. J. Hausmann, O. O. Reichmann, H. E. Wiegand & L. Zgusta (Eds.), *Wörterbücher/ Dictionaries/ Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie*. Berlin: de Gruyter, pp. 287-296.
- Granger, S. & M. Paquot. (2012). *Electronic Lexicography*. Oxford: Oxford University Press.
- ILNE, (1933-). *Ιστορικό Λεξικό της Νέας Έλληνικής, της τε κοινῶς ὀμιλουμένης καὶ τῶν ιδιωμάτων* [Historical Dictionary of Modern Greek, both of the Standard language and the dialects], Athens: Academy of Athens.
- Grundy, V. & D. Rawlinson. (2016). The Practicalities of Dictionary Production; Planning and Managing Dictionary Projects; Training of Lexicographers. In P. Durkin (ed.) *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press, pp. 561-578.
- Janse, M. (forthcoming). Καππαδοκικά [Cappadocian]”. In C. Tzitzilis (ed.), *Η ελληνική γλώσσα και οι διάλεκτοι της* [Greek and its dialects].Thessaloniki: Institutou Neoellinikon Spoudon (Manolis Triantafyllides Foundation).
- Johanson, L. (2002). *Structural factors in Turkic language contacts*. London: Curzon.
- Karanastasis, A. (1984-1992). *Ιστορικό Λεξικό των Έλληνικών ιδιωμάτων της Κάτω Ιταλίας* [Historical Dictionary of the Greek dialects of South Italy]. Athens: Academy of Athens.
- Karasimos, A. (2019). Οι νεοελληνικές διάλεκτοι στον ψηφιακό χάρτη: ανασκόπηση και προτάσεις. [Modern Greek dialects in digital maps: a review and suggestions]. In G. Karla, I. Manolessou, N. Pantelidis (eds.) *Λέξεις / Τιμητικός τόμος για την Χριστίνα Μπάσσα-Μπεζαντάκου* [Words | Studies in honor of Christina Bassea-Bezantakou]. Athens: Kardamitsa publications, pp. 109-142.
- Karatsareas, P. 2013. Understanding diachronic change in Cappadocian Greek: the dialectological perspective. *Journal of Historical Linguistics* 3(2), pp. 192–229.
- Katsoyannou M. & S. Armostis (2019). Διαλεκτική λεξικογραφία στην Κύπρο: Προκλήσεις και προοπτικές [Dialectal Lexicography in Cyprus: Challenges and prospects], In G. Karla, I. Manolessou, N. Pantelidis, (eds.) *Λέξεις / Τιμητικός τόμος για την Χριστίνα Μπάσσα-Μπεζαντάκου* [Words | Studies in honor of Christina Bassea-Bezantakou] Athens: Kardamitsa publications, pp. 187-210.
- Katsouda, G. 2012. Διαλεκτική Λεξικογραφία: Επισκόπηση [Dialectal Lexicography: An overview]. *Lexicographikon Deltion* 26, pp. 77-159.
- Kostakis A. (1963). *Le parler grec d’Anakou*. Athenes: Kentro Mikrasiatikon Spoudon.

- Kostakis, A. (1986-1987). *Λεξικό της Τσακωνικής διαλέκτου*. [Dictionary of the Tsakonian dialect]. Athens: Academy of Athens.
- Kotsanidis, L. (2006). *Το γλωσσικό ιδίωμα του Μιστί Καππαδοκίας* [The Cappadocian dialect of Misti]. Kilkis: Gnomi.
- Kuzmina, V. & A. Rylova (2010). ABBYY Lingvo electronic dictionary platform and Lingvo Content dictionary writing system. In S. Granger & M. Paquet (Eds.) *E-Lexicography in the 21st Century: New Challenges, New Applications*. UCL Presses Universitaires de Louvain, pp. 419-423.
- Manolessou, I. (2016). Ο νέος τόμος του Ιστορικού λεξικού της νέας ελληνικής της Ακαδημίας Αθηνών: διαχρονικές προοπτικές [The new volume of the Historical Dictionary of Modern Greek of the Academy of Athens: diachronic perspectives]. *Studies in Greek Linguistics* 36, pp. 239-249.
- Manolessou, I., S. Beis & Ch. Bassea-Bezantakou (2012). Η φωνητική απόδοση των νεοελληνικών διαλέκτων [The phonetic transcription of Modern Greek Dialects]. *Lexikographikon Deltion* 26, pp. 161-222.
- Manolessou, I. & Ch. Bassea-Bezantakou (2017). “The Historical Dictionary of Modern Greek”. In Wandl-Vogt E. & A. Dorn. (eds.) *Dialekt | Dialect 2.0. Langfassungen, 7. Kongress der Internationalen Gesellschaft für Dialektologie und Geolinguistik (SIGD)*. Wien: Praesens Verlag, pp. 13-37.
- Manolessou, I. & Katsouda G. (forthcoming). The making of the Historical Dictionary of Modern Greek: Problems and solutions in the domain of historical and dialectal lexicography. *Proceedings of the 10<sup>th</sup> International Conference on Historical Lexicography and Lexicology*, Fryske Akademy, Leeuwarden 12-14 June 2019.
- Marello, C. (2004). Lexicography in Italy: Specific themes and trends. *International Journal of Lexicography* 17.4, pp. 349-356.
- Mavrochalyvidis, G. & I. Kesisoglou. (1960). *Τὸ γλωσσικὸ ἰδίωμα τῆς Ἀξοῦ (Le dialecte d’Axo)* [The Axós dialect]. Athens: Institut Français d’Athènes.
- Melissaropoulou, D. (2016). Variation in word formation in the light of the language contact factor: the case of Cappadocian Greek. *Journal of Language Sciences* 55, pp. 55-67.
- Melissaropoulou, D. (2019a). *Γλωσσική ποικιλία και γλωσσική αλλαγή στο σχηματισμό συγκριτικών παραθετικών δομών: δεδομένα από την Καππαδοκική* [Language variation and change in the formation of comparative constructions: evidence from Cappadocian Greek]. In A. Archakis, N. Koutsoukos, G. Xydopoulos & D. Papazachariou (eds.), *Γλωσσική Ποικιλία. Μελέτες αφιερωμένες στην Αγγελική Ράλλη*. Athens: ΚΑΡΑ ΕΚΔΟΤΙΚΗ, pp. 329-346.
- Melissaropoulou, D. (2019b). Morphological pattern replication phenomena as instances of typological shift. In G. Karla, I. Manolessou, N. Pantelidis (eds.), *Λέξεις | Τιμητικός τόμος για την Χριστίνα Μπασέα-Μπεζαντάκου* [Words | Studies in honor of Christina Bassea-Bezantakou]. Athens: Kardamitsa publications, pp. 317-340.
- Melissaropoulou, D., Galiotou, E., Dimela, E., Karanikolas, N., Papanagiotou, Ch., Xydopoulos, G. & A. Ralli (2015). Υλοποίηση του πρώτου πολυμεσικού τριδιαλεκτικού διαδικτυακού λεξικού [Implementation of the first multimedia tri-dialectal e-dictionary]. In M. Tzakosta (ed.) *Η διδασκαλία των νεοελληνικών γλωσσικών ποικιλιών και διαλέκτων στην πρωτοβάθμια και δευτεροβάθμια εκπαίδευση. Θεωρητικές προτάσεις και διδακτικές εφαρμογές* [Teaching Modern Greek Dialectal varieties in primary and secondary education]. Athens: Gutenberg, pp. 259-280.
- Papadopoulos, A. A. (1958-1961). *Ιστορικὸν Λεξικὸν τῆς Ποντικῆς Διαλέκτου* [Historical Lexicon of the Pontic Dialect]. Ἀρχεῖον Πόντου. Athens: Epitropi Pontiakon Meleton.
- Papazachariou D., Vassalou, N., & M. Janse (2016). Methodological Principles of Mišótika Cappadocian Data Collection. In I. Kappa & M. Tzakosta (Eds.), *7th International Conference on Modern Greek Dialects and Linguistic Theory*, Rethymno: University of Crete, pp. 32–32.
- Ralli, A. (2009). Morphology meets dialectology: insights from Modern Greek dialects *Morphology* 19, pp. 87–105.
- Ralli A. (ed.) (2015). *Πρόγραμμα Θαλής, Πόντος-Καππαδοκία-Αἰβαλί. Στα χνάρια της Μικρασιατικής Ελληνικής*. Πάτρα: Πανεπιστήμιο Πατρών.
- Reichmann, O. (2012). *Historische Lexikographie*. Berlin: De Gruyter.
- Themistocleous, C., Katsogiannou, M., Armosti, S., Christodoulou, K. (2012). *Cypriot Greek Lexicography: An Online Lexical Database*. In: *Proceedings of Euralex 2012*, pp. 889–891.
- Tiberius, C., Niestadt, J., Schoonheim, T. (2014). The INL Dictionary Writing System. *Slovenščina 2.0*, 2 (2), pp. 72–93.
- Thomason, S.G. & T. Kaufman. (1988). *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Trudgill, P. (2003). Modern Greek dialects: A preliminary classification. *Journal of Greek Linguistics* 4(1), pp. 45-63.
- Vryonis, S. (1971). *The Decline of Medieval Hellenism in Asia Minor and the Process of Islamization from the Eleventh through the Fifteenth Century*. Berkeley: University of California Press.
- Winford, D. (2010). Contact and borrowing. In Hickey, R. (ed.), *The Handbook of Language Contact*. Malden, MA/Oxford: Wiley-Blackwell, pp. 170-187.
- Xydopoulos, G.I., Dimela, E., Melissaropoulou, D., Papanagiotou, Ch. & A. Ralli. (2015). ΛΕΠΟΚΑΜ, ένα πολυμεσικό λεξικό των διαλέκτων του Πόντου, της Καππαδοκίας και του Αἰβαλιού: Σχεδιασμός και υλοποίηση [LEPOKAM, a multimodal dictionary of the dialects of Pontus, Cappadocia and Aivali: Designing and Implementation]. In A. Ralli (ed.), *Πρόγραμμα Θαλής: Πόντος, Καππαδοκία, Αἰβαλί: Στα χνάρια της Μικρασιατικής Ελληνικής* [Thalis project: Pontus, Cappadocia, Aivali: in search of Asia Minor Greek. Patras: Laboratory of Modern Greek Dialects, University of Patras, pp. 99-114.



**Acknowledgements**

This research has been co-financed by the General Secretariat for Research and Technology and the Hellenic Foundation for Research and Innovation. Many thanks to Angela Ralli, Metin Bağrıaçık and Petros Karatsareas for providing access to oral dialectal material. Further information available at <http://cappadocian.upatras.gr/en>