

Structuring a Multimedia Tri-dialectal Dictionary

Nikitas N. Karanikolas, Eleni Galiotou,
George J. Xydopoulos, Angela Ralli,
Konstantinos Athanasakos, and
George Koronakis



TSD 2013, Plzeň – CZ, September 1 – 5, 2013

THALIS program, project AMiGre



- In this paper, we discuss the design and implementation of a multimedia tridialectal dictionary of Greek dialects in Asia Minor (Pontic, Cappadocian, Aivaliot) within the framework of the THALIS program,
- “Pontus, Cappadocia, Aivali: in search of **Asia Minor Greek**” (AMiGre)

TSD 2013, 3-Sep-13

Motivation



- the three Asia Minor dialects, Pontic, Cappadocian and Aivaliot are not sufficiently documented,
- they are on the way to extinction,
- with the exception of the old Papadopoulos' (1958) historical dictionary of Pontic, there are only glossaries containing words and idiomatic phrases accompanied by their meaning in Standard Modern Greek,
- in most of these glossaries, lemmas are stored in a very unsystematic way and crucial information, such as pronunciation or usages, is missing,
- some verbs are listed in their past tense form while others appear in the present tense,
- there is no distinction made between words and phrases.

TSD 2013, 3-Sep-13

Functional Requirements (1/2)



- Provide the dialectal area or the source where the lemma has been extracted from,
- access to a graphic representation of each lemma in a conventionally-adopted character set,
- pronunciation (phonetic form),
- grammar (categorial and morphological information),
- origin (etymology),
- meaning (synonymic and/or descriptive definitions),
- usages (thematic and register labels)
- authentic examples of use

TSD 2013, 3-Sep-13

Functional Requirements (2/2) and Dimensions



- Cross-reference links to other entries, related either through derivational processes or through semantic relations (*)
- Synonymy and Homonymy links to other lemmas (in the same dialect)
- Equivalence links to lemmas of other dialects (*)
- 2,500 entries from each of the three dialects of Asia Minor Greek (a total of 7,500 entries)

TSD 2013, 3-Sep-13

Design – lemma structure



- headword, dialect (dialectal region), morphological information/process and etymology are primary information with single values that together define and are dependent on the lemma;
- each lemma can have many different realizations and each one of them is characterized by a slightly different phonetic realization dependent on the microdialectal region it originates from (the specific area within the wider dialectal region where the lemma's realization occurs);
- each lemma can possibly have different meanings (i.e. polysemy), or be homonymous with other, semantically distinct, lemmas;
- for each meaning, different usage examples are essential.

TSD 2013, 3-Sep-13

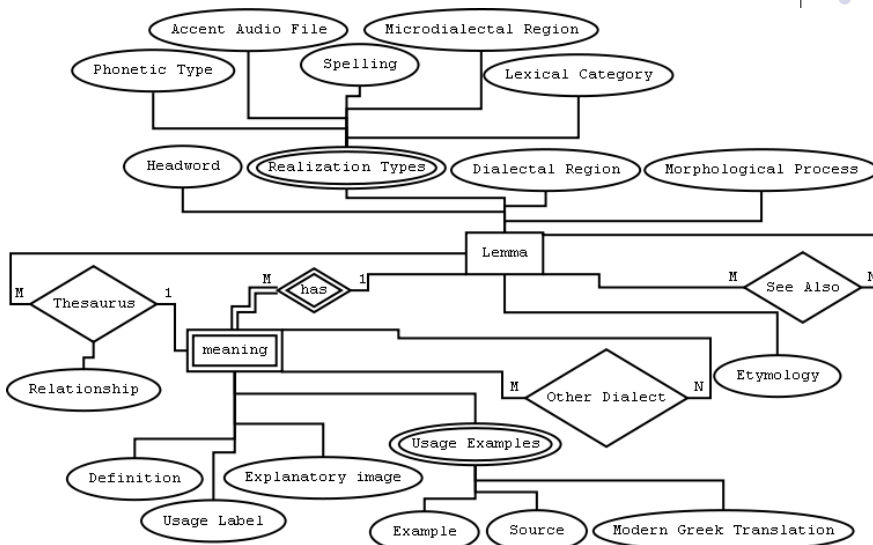
Design - relations



- Cross reference (See also) links can be available for connecting lemmas that are semantically / pragmatically / morphologically / etymologically related to each other (*);
- Synonyms (words with similar meanings) and Antonyms (words with opposite meanings) are two semantic relations that apply between lemmas. Both relations relate a lemma meaning with a lemma (the referenced one);
- There are meanings of different lemmas from different dialects that share the same definition (have the same meaning). This relation can be shortly named (labelled) “Other Dialect” (*).

TSD 2013, 3-Sep-13

AMiGre - ERD



Data Dictionary for “Lemma”



Attribute	Definition	Data format	Example
Headword	The canonical form of the word	String containing only capital letters of the Greek alphabet	ΑΛΛΟΥΓΥΡΙΣΤΡΑ
Etymology	Basic information about the origin of the word.	String written in Greek with accents (polytonal)	Από το ρήμα αλλουγορίζου (from the verb aluji'rizu)
Morphological Process	Different processes involved in word-formation.	A value from a predefined list of morphological processes	Σύνθετο (Compound noun)
Dialectal Region	The region/dialect in which the lemma is found	A value from a predefined list of Dialects	Αϊβαλί (AIVALI)

TSD 2013, 3-Sep-13

Data Dictionary for “Realization Types”



subAttribute	Definition	Data format	Example
Phonetic Type	Phonetic transcription of (the examined) pronunciation of the word.	String containing letters of the International Phonetic Alphabet (IPA).	alujɪ'ristra
Accent Audio	Audio file of the authentic pronunciation of the word	String containing a file path	http://amigre.gr/xyzR1.wav
Spelling	Non-standard graphic representation of pronunciation according to the orthographic rules of Standard Greek, combined with diacritics to annotate any phonological alternations.	String containing the letters of the Greek alphabet and other diacritic symbols (accent, hyphens, parentheses and apostrophes)	αλλουγορίστρα
Microdialectal Region	Name of a specific area within the wider dialectal region of the lemma in which the realization form is found	Value from a predefined list of microdialectal regions	
Lexical Category	Part of Speech & Gender	Value from a predefined list of lexical categories	Ουσιαστικό Θηλυκό (noun feminine)

TSD 2013, 3-Sep-13

Data Dictionary for “Meaning”



Attribute	Definition	Data format	Example
Definition	Short description of the meaning of a lemma	String in StandardModern Greek	Γυναίκα που περιφέρεται εδώ κι εκεί ('woman who goes around')
Explanatory Image File	Image illustrating the meaning of a lemma.	String containing a file path	http://amigre.gr/xyzM1.png
Usage Label	Formal indication of the context (stylistic/register/other) in which the lemma is used.	A value from a predefined list of domains	ΥΠΟΤΙΜΗΤΙΚΟ (pejorative)

TSD 2013, 3-Sep-13

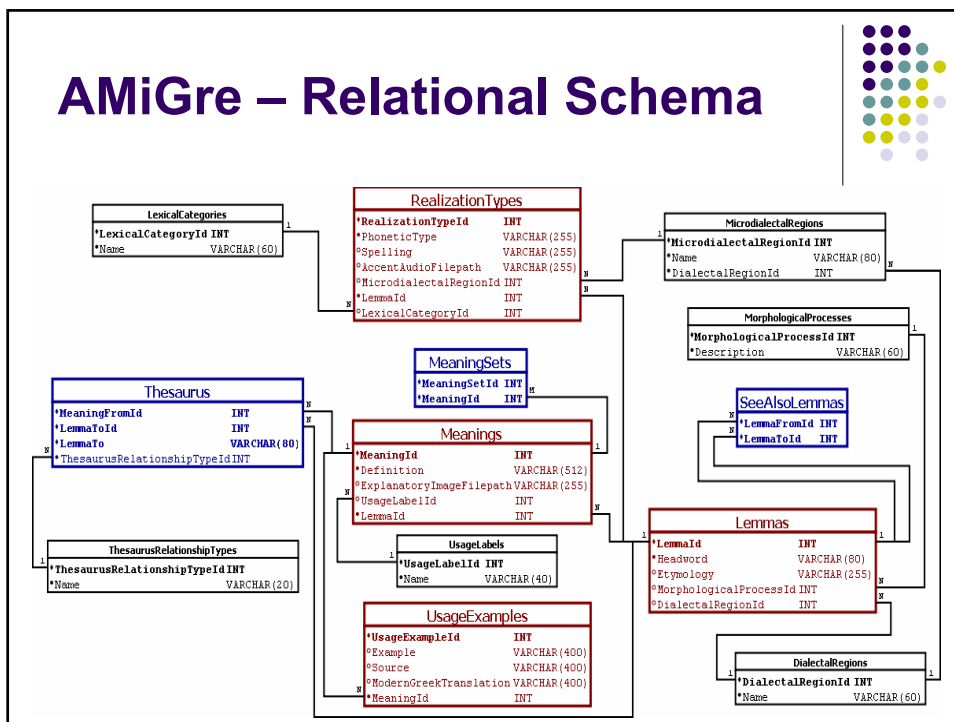
Data Dictionary for “Usage examples”



sub Attribute	Definition	Data format	Example
Usage example	Example (phrase or sentence) demonstrating the usage of the lemma under one specific meaning, in the original dialect	The whole example (the whole string) is written with the letters of the Greek alphabet and other diacritic symbols (accent, hyphens, parentheses and apostrophes)	Ξιπόρτσι πάλ' -η- γ' - αλλουγυρίστ ρα
Standard Modern Greek Translation	Translation of the usage example into Standard Modern Greek	String in Standard Modern Greek	Πάλι βγήκε η αλλουγυρίστ ρα.
Source	Reference to the source from which the usage example was extracted	String (can be a book, a URL, etc)	

TSD 2013, 3-Sep-13

AMiGre – Relational Schema



The realization types of lemma ΑΛΛΟΥΓΥΡΙΣΤΡΑ



* Λέξη κεφαλή: (Headword)

Ετυμολογία: (Etymology)

Μορφολογική Διεργασία: (Morphological Process)

* Διαλεκτική Περιοχή: (Dialectal Region)

Τύποι Πραγμάτωσης (Realizations)

Δημιουργία Νέου Τύπου Πραγμάτωσης

Κωδικός	Φωνητικός Τύπος	Αρχείο Ήχου Προφ	Φωνητική Ορθογραφία	Μικροδιαλεκτική Περιοχή	Λεξική Κατηγορία
	9 aluj'ristra		αλλουγυρίστρα		Ουσιαστικό Θηλυκό
	(Phonetic Type)		(Spelling)		(Lexical Category)

TSD 2013, 3-Sep-13

The (two) meanings of lemma ΑΛΛΟΥΓΥΡΙΣΤΡΑ



* Λέξη κεφαλή:
(Headword)

Ετυμολογία:
(Etymology)

Μορφολογική Διεργασία:
(Morphological Process)

* Διαλεκτική Περιοχή:
(Dialectal Region)

Τύποι Πραγμάτωσης
(Meanings)

Δημιουργία Νέας Σημασίας

Κωδικός	Ορισμός	Χρηστικό Σημάδι	Επεξηγηματική Εικόνα	Πλήθος παραδειγ
12	γυναίκα που περιφέρεται εδώ κι εκεί			1
13	πόνος με πρήξιμο γύρω από το νύχι	Ιατρική		1
	(Definition)	(Usage Label)		

TSD 2013, 3-Sep-13

Usage Examples and Synonyms of the first meaning of lemma ΑΛΛΟΥΓΥΡΙΣΤΡΑ



* Ορισμός:
(Definition)

Επεξηγηματική Εικόνα: Επιλογή εικόνας... Προβολή εικόνας Αφαίρεση εικόνας

Χρηστικό Σημάδι:

Παραδείγματα Χρήσης

Προσθήκη Νέου Παραδείγματος Χρήσης

Κωδικός	Παράδειγμα Χρήσης	Μετάφραση στην ΚΝΕ	Πηγή
5	Ξιπάρτα πάλι η γ' αλλουγυρίστρα	Πάλι βγήκε η αλλουγυρίστρα.	
	(Usage Example)	(Modern Greek Translation)	

Θησαυρός
(Thesaurus)

Προσθήκη Νέου Συνώνυμου/Αντώνυμου

Κωδικός Λήμματος	Κωδικός Σχέσης	Λέξη - Κεφαλή	Φωνητική Ορθογραφία	Διαλεκτική Περιοχή	Σχέση
11	1	ΠΟΡΤΟΥΓΥΡΑ	πουρτουγύρα	Αιβαλί	Συνώνυμο
12	1	ΛΙΩΣΤΡΑ	λ'ιώστρα	Αιβαλί	Συνώνυμο
13	1	ΞΟΡΤΑ	σόρτα	Αιβαλί	Συνώνυμο
		(Headword)	(Spelling)	(Dialectal Region)	(Relationship)

Interesting points and Future work



- The dictionary contains three dialects. However, its design permit to **incorporate more dialects** in the future.
- Future work includes the implementation of an **advanced retrieval component** and the introduction of a innovate module for **automatic (or semi-automatic) identification of the “other dialect” relations**. The latter, could be based on the similarity of lemma meanings’ definitions.

Structuring a Multimedia Tri-dialectal Dictionary (AMiGre)



- Thank you for your attention
- We will try to answer your Questions

TSD 2013, 3-Sep-13

Structuring a Multimedia Tri-dialectal Dictionary (AMiGre)



Springer LNAI 8082 pages 509 - 518

Nikitas N. Karanikolas

nnk@teiath.gr

<http://users.teiath.gr/nnk/>

Eleni Galiotou

egali@teiath.gr

TSD 2013, 3-Sep-13