Dialectal lexicon building: requirements and technical specifications

Professor

Nikitas N. Karanikolas

TEI of Athens

Department of Informatics

nnk@teiath.gr

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

Dialectal lexicon building: requirements and technical specifications

- An International Lecture.
- Based on:
  Structuring a Multimedia Tri-dialectal Dictionary
  by
  Nikitas N. Karanikolas, Eleni Galiotou,
  George J. Xydopoulos, Angela Ralli,
  Konstantinos Athanasakos and George Koronakis.
  TSD'2013: Text, Speech and Dialogue.
  Pilsen, Czech Republic. Springer, LNAI 8082.
- This research is done through the AMiGre project:
  The AMiGre project has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Life-long Learning" of the National Strategic Reference Framework (NSRF) – Research Funding Program: Thalis. Investing in knowledge society through the European Social Fund.

THALIS program, project AMiGre

- AMiGre was a project within the framework of THALIS program.
- Project title: “Pontus, Cappadocia, Aivali: in search of Asia Minor Greek” (AMiGre).
- One of the deliverables of the AMiGre project was the design and implementation of a multimedia tri-dialectal dictionary for three Greek dialects in Asia Minor (Pontic, Cappadocian, Aivaliot).

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.

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 (categorical and morphological information),
- origin (etymology),
- meaning (synonymic and/or descriptive definitions),
- usages (thematic and register labels)
- authentic examples of use

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)

Synonymy

- A synonym is a word or phrase that means exactly or nearly the same as another word or phrase in the same language.
- Words that are synonyms are said to be synonymous, and the state of being a synonym is called synonymy.
- An example of synonyms are the words *begin*, *start*, *commence*, and *initiate*.

Antonymy

- In lexical semantics, opposites are words that lie in an inherently incompatible binary relationship.
- Examples of opposite pairs are:
  - big versus small,
  - long versus short,
  - precede versus follow.
- The notion of incompatibility here refers to the fact that one word in an opposite pair entails that it is not the other pair member.
- For example, something that is long entails that it is not short. It is referred to as a 'binary' relationship because there are two members in a set of opposites.
- The term **antonym** (and the related **antonymy**) is commonly taken to be synonymous with opposite.

Homonymy

- **homonymy** is the case where a group of words share the same pronunciation but have different meanings, whether spelled the same or not.
- A more restrictive definition sees homonyms as words that are simultaneously **homographs** (words that share the same spelling, regardless of their pronunciation) and **homophones** (words that share the same pronunciation, regardless of their spelling). In other words, homonyms are words that have same pronunciation and spelling but different meanings.
- The pair *left* (past tense of leave) and *left* (opposite of right) are homonyms between each other.
- the words *read* (peruse) and *reed* (waterside plant) would be considered homophones.

Polysemy

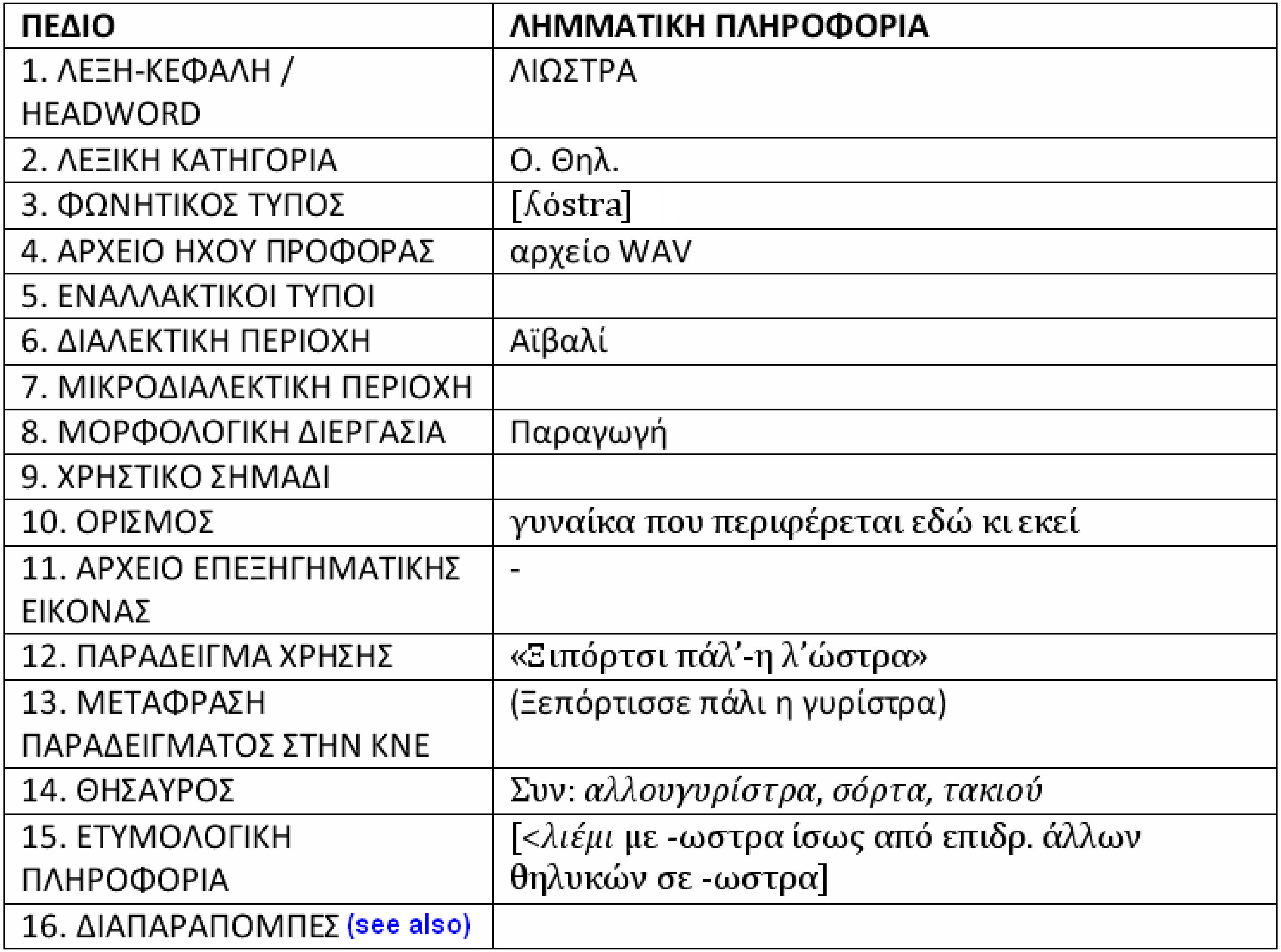
- **Polysemy** (from Greek: πολυ-, poly-, "many" and σῆμα, sêma, "sign") is the capacity for a word to have multiple meanings, usually related by contiguity of meaning within a semantic field.
- Lexicographers define polysemes within a single dictionary lemma, numbering different meanings, while homonyms are treated in separate lemmata.

Example given by Linguists and Lexicographers

ΠΕΔΙΟ (FIELD)	ΛΗΜΜΑΤΙΚΗ ΠΛΗΡΟΦΟΡΙΑ (LEMMA VALUES)
1. ΛΕΞΗ-ΚΕΦΑΛΗ / HEADWORD	ΒΡΟΥΛΟ
2. ΛΕΞΙΚΗ ΚΑΤΗΓΟΡΙΑ (lexical category)	(Ο. ουδ.)
3. ΦΩΝΗΤΙΚΟΣ ΤΥΠΟΣ (phonetic type)	[‘vrulu]
4. ΑΡΧΕΙΟ ΗΧΟΥ ΠΡΟΦΟΡΑΣ (digital record)	WAV
5. ΕΝΑΛΛΑΚΤΙΚΟΙ ΤΥΠΟΙ (alternative types)	Βρόλους [‘vrolus] (Παμφ. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ ΣΥΝΔΕΣΗ)
6. ΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (dialectal region)	Αϊβαλί
7. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (microdialect)	
8. ΜΟΡΦΟΛΟΓΙΚΗ ΔΙΕΡΓΑΣΙΑ (morphological process)	-
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ (usage)	1. ΦΥΤΟΛΟΓΙΑ
10. ΟΡΙΣΜΟΣ (definition)	Βούρλο
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	JPG
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ (usage example)	«Έκουψα καμπόσα βρούλα κι πέρασα αρμαθιά τα ψάρια πό πιασα σήμιρα»
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ*	(‘Εκοψα μερικά βούρλα και ....)
14. ΘΗΣΑΥΡΟΣ (thesaurus)	
15. ΕΤΥΜΟΛΟΓΙΚΗ ΠΛΗΡΟΦΟΡΙΑ (etymology)	[ΕΤΥΜ ελνστ. βροῦλον]
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ	2. ---
10. ΟΡΙΣΜΟΣ	Ανόητος
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	JPG
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ	«Ντιπ για ντιπ βρούλου τούτου του πιδί».
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ	(Τελείως ανόητο αυτό το παιδί)
14. ΘΗΣΑΥΡΟΣ	---

* (modern Greek translation)

Another example by Linguists and Lexicographers



ΠΕΔΙΟ	ΛΗΜΜΑΤΙΚΗ ΠΛΗΡΟΦΟΡΙΑ
1. ΛΕΞΗ-ΚΕΦΑΛΗ / HEADWORD	ΛΙΩΣΤΡΑ
2. ΛΕΞΙΚΗ ΚΑΤΗΓΟΡΙΑ	Ο. Θηλ.
3. ΦΩΝΗΤΙΚΟΣ ΤΥΠΟΣ	[ʎóstra]
4. ΑΡΧΕΙΟ ΗΧΟΥ ΠΡΟΦΟΡΑΣ	αρχείο WAV
5. ΕΝΑΛΛΑΚΤΙΚΟΙ ΤΥΠΟΙ	
6. ΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ	Αϊβαλί
7. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ	
8. ΜΟΡΦΟΛΟΓΙΚΗ ΔΙΕΡΓΑΣΙΑ	Παραγωγή
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ	
10. ΟΡΙΣΜΟΣ	γυναίκα που περιφέρεται εδώ κι εκεί
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	-
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ	«Ξιπόρτσι πάλ'-η λ'ώστρα»
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ	(Ξεπόρτισσε πάλι η γυρίστρα)
14. ΘΗΣΑΥΡΟΣ	Συν: *αλλουγυρίστρα, σόρτα, τακιού*
15. ΕΤΥΜΟΛΟΓΙΚΗ ΠΛΗΡΟΦΟΡΙΑ	[<*λιέμι* με -ωστρα ίσως από επιδρ. άλλων θηλυκών σε -ωστρα]
16. ΔΙΑΠΑΡΑΠΟΜΠΕΣ (see also)	

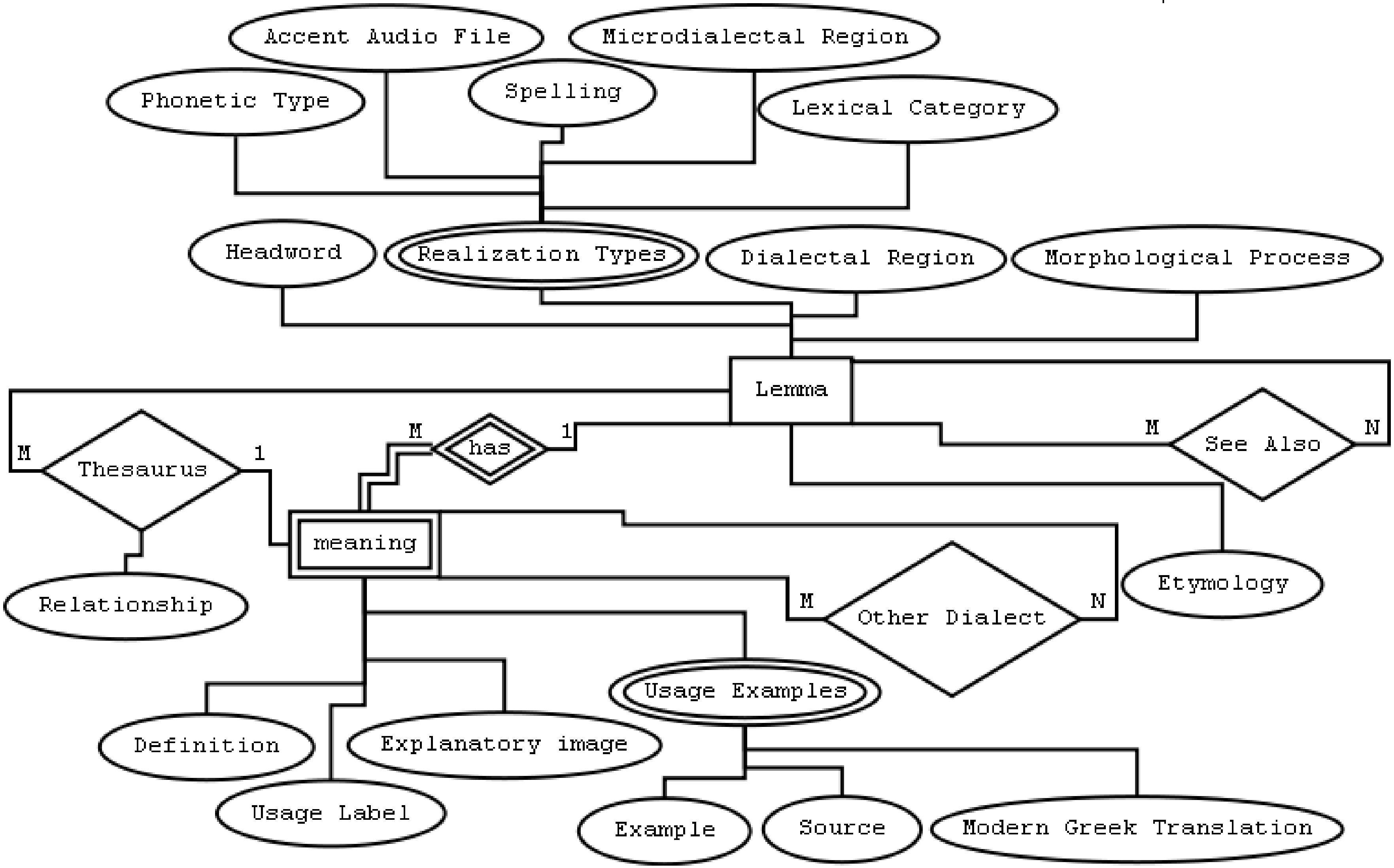
Design – lemma structure

- 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 micro-dialectal 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.

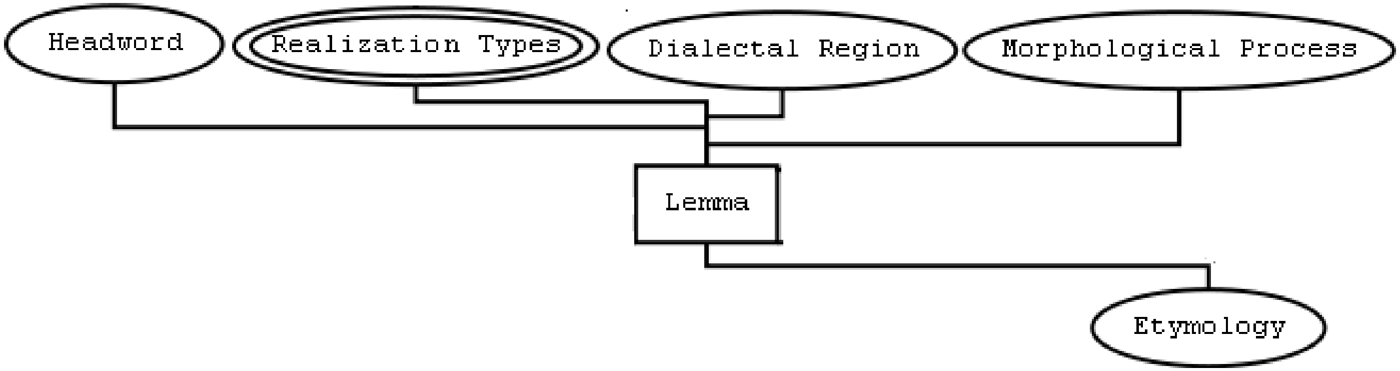
Design - relations

- 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); Synonym and Antonym links are restricted between a lemma meaning and a lemma from the same dialect.
- There are meanings of different lemmas from different dialects that share the same definition (have the same meaning). This relation can be shortly named (labeled) “Other Dialect” (*).
- In contrast with the rest relations, “Other Dialect” is a symmetrical relation.

AMiGre – Data Schema – ERD



Data Dictionary for “Lemma”



Headword

Realization Types

Dialectal Region

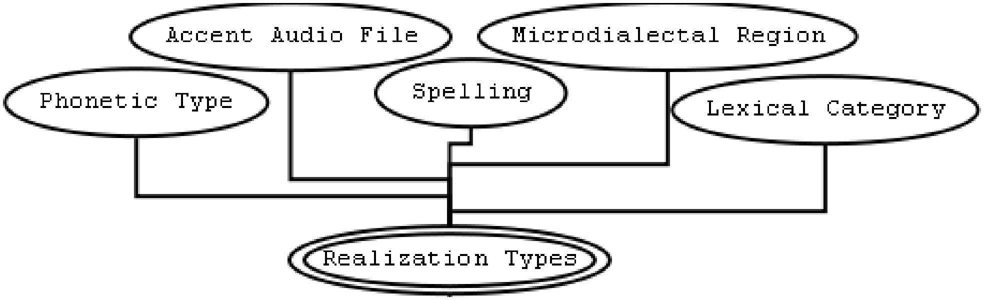
Morphological Process

Lemma

Etymology

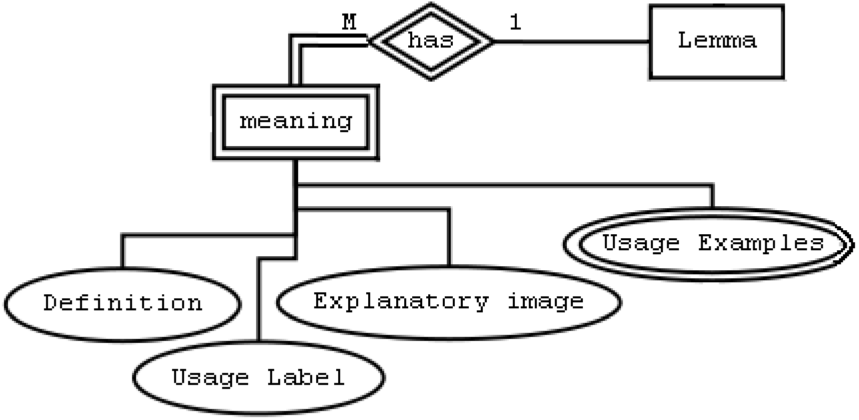
Attribute	Definition	Data format	Example
Headword	The canonical form of the word	String containing only capital letters of the Greek alphabet	ΑΛΛΟΥΓΥΡΙΣΤΡΑ
Etymology	Basic information about the origin of the word.	String written in Greek with accents (polytonal)	Από το ρήμα αλλουγυρίζου (from the verb 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)

Data Dict. for “Realization Types”



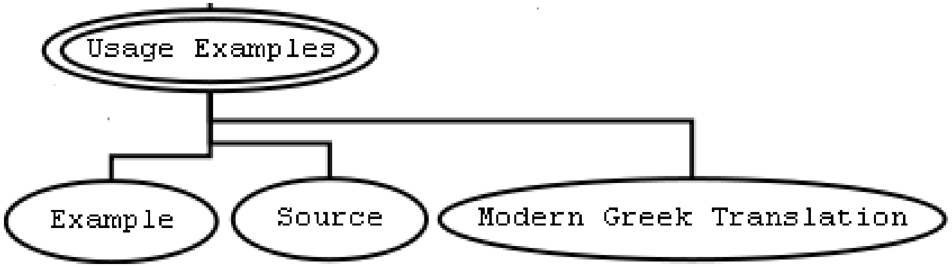
subAttribute	Definition	Data format	Example
Phonetic Type	Phonetic transcription of (the examined) pronunciation of the word.	String containing letters of the International Phonetic Alphabet (IPA).	aluji'ristra
Accent Audio	Audio file of the authentic pronunciation of the word	String containing a file 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)	αλλουγυρίστρα
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)

Data Dict. 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)

Data Dict. 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’-αλλουγυρίστρα
Standard Modern Greek Translation	Translation of the usage example into Standard Modern Greek	String in Standard Modern Greek	Πάλι βγήκε η αλλουγυρίστρα.
Source	Reference to the source from which the usage example was extracted	String (can be a book, a URL, etc)	

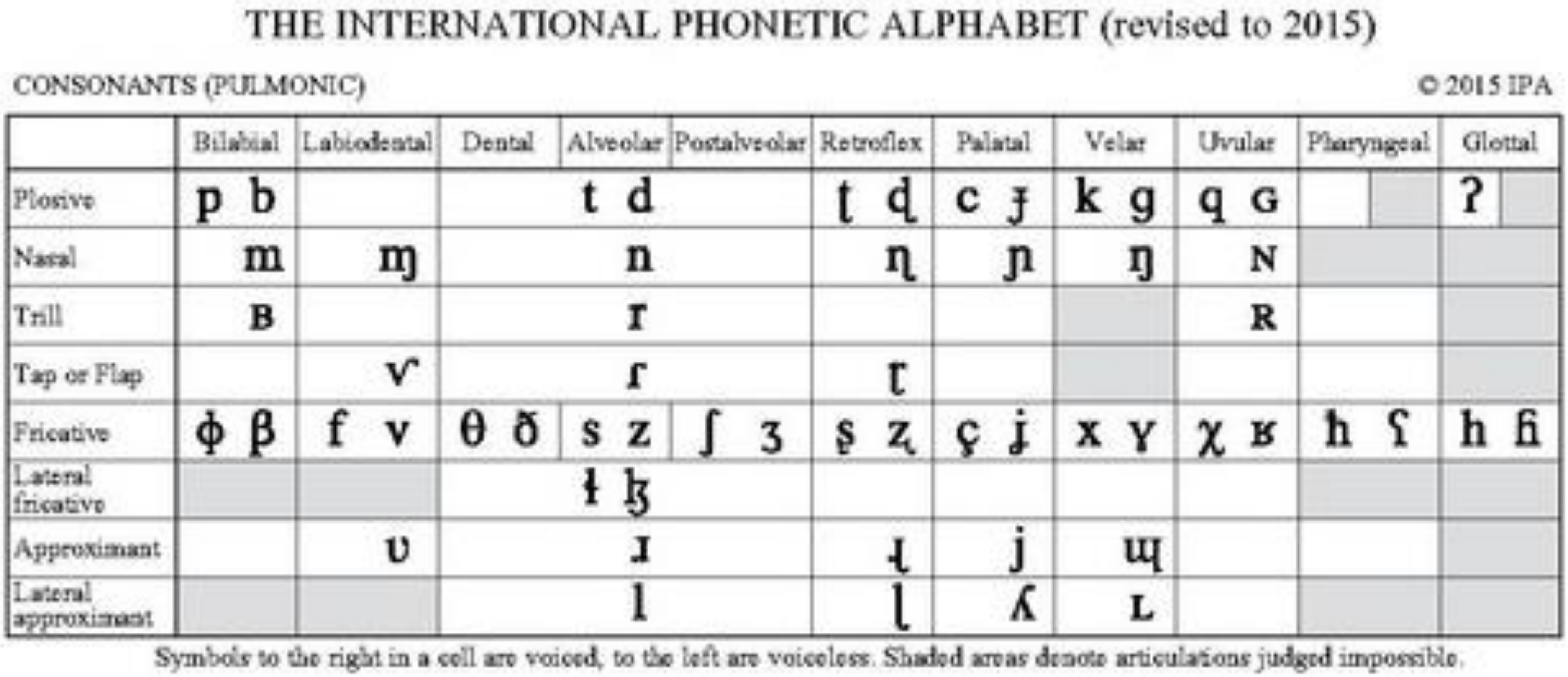
Some points to mention

- Etymology
  - Greek Polytonal
  - Some loan characters from other alphabets in case of loan words (e.g. some characters from the Turkish alphabet)
- Phonetic type
  - IPA
- Spelling (Phonetic Orthography)
  - Greek
  - Accents
  - Hyphen, parentheses, apostrophe,
  - some characters with umlaut

International Phonetic Alphabet (IPA)

- The International Phonetic Alphabet (IPA) is an alphabetic system of phonetic notation based primarily on the Latin alphabet.
- It was devised by the International Phonetic Association as a standardized representation of the sounds of spoken language.
- The IPA is used by lexicographers, foreign language students and teachers, linguists, speech-language pathologists, singers, actors, constructed language creators, and translators.

IPA Consonants (Pulmonic)



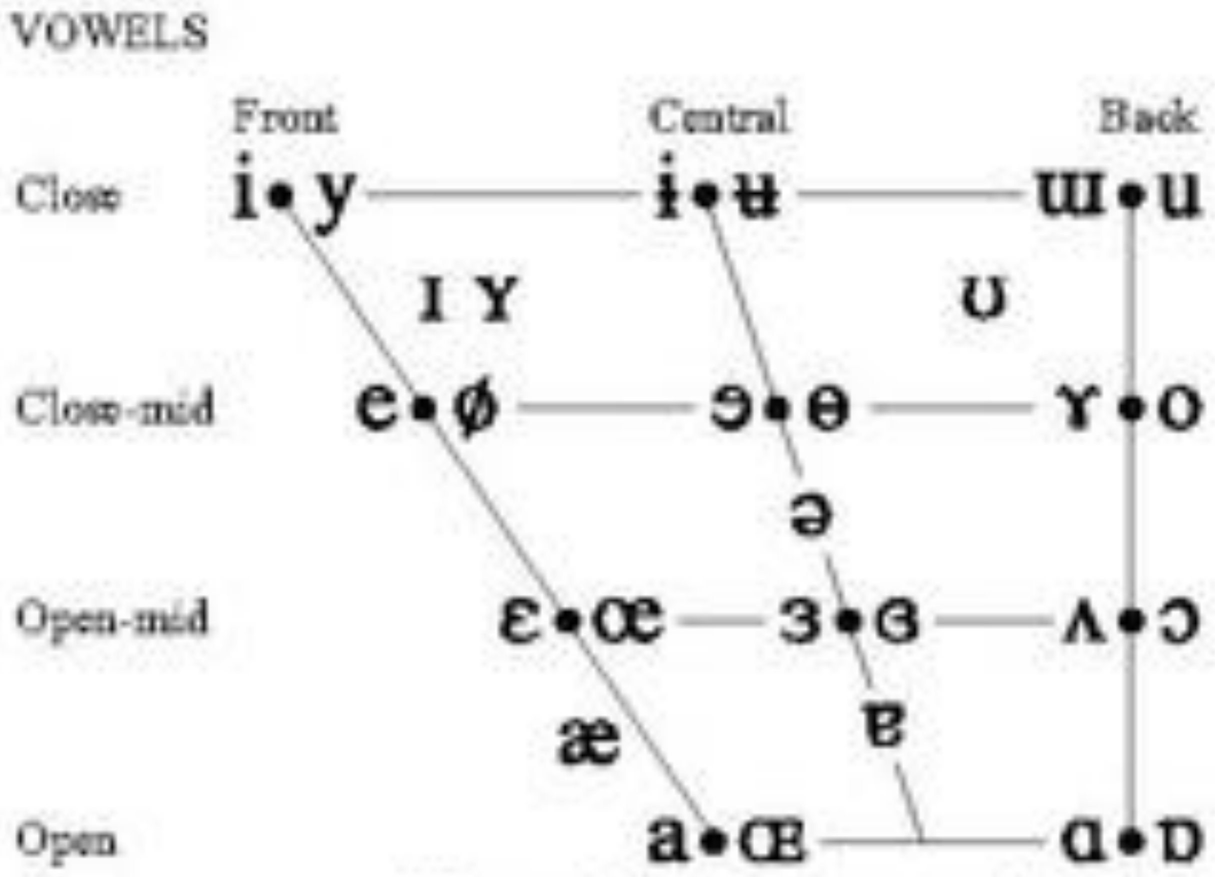
THE INTERNATIONAL PHONETIC ALPHABET (revised to 2015)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ	▨	ʔ ▨
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ	▨	▨
Trill	ʙ			r				▨	ʀ		▨
Tap or Flap		ⱱ		ɾ		ɽ		▨			▨
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative	▨	▨		ɬ ɮ						▨	▨
Approximant		ʋ		ɹ		ɻ	j	ɰ			▨
Lateral approximant	▨	▨		l		ɭ	ʎ	ʟ		▨	▨

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

IPA Vowels

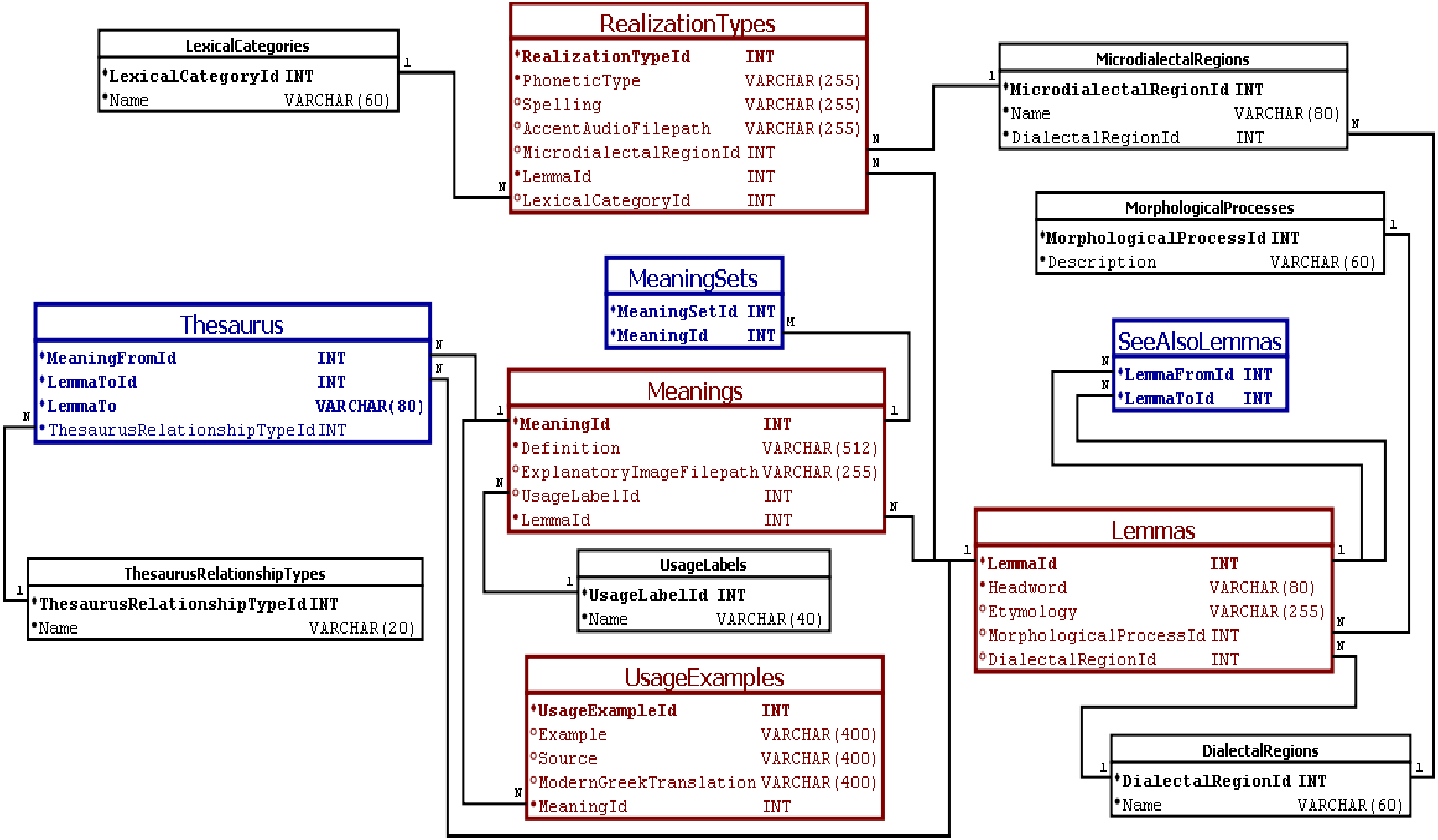


Where symbols appear in pairs, the one to the right represents a rounded vowel.

SAMPA

- SAMPA (Speech Assessment Methods Phonetic Alphabet) is a machine-readable phonetic alphabet.
- It was originally developed under the ESPRIT project 1541, SAM (Speech Assessment Methods) in 1987-89.
- It applied first to the European Communities languages Danish, Dutch, English, French, German, and Italian (by 1989).
- Llater, it applied to Norwegian and Swedish (by 1992).
- Subsequently it applied to Greek, Portuguese, and Spanish (1993).
- It has now been extended to Bulgarian, Estonian, Hungarian, Polish, and Romanian (1996).

AMiGre – Relational Schema



LexicalCategories	
*LexicalCategoryId	INT
*Name	VARCHAR(60)

RealizationTypes	
*RealizationTypeId	INT
*PhoneticType	VARCHAR(255)
°Spelling	VARCHAR(255)
°AccentAudioFilepath	VARCHAR(255)
°MicrodialectalRegionId	INT
*LemmaId	INT
°LexicalCategoryId	INT

MicrodialectalRegions	
*MicrodialectalRegionId	INT
*Name	VARCHAR(80)
*DialectalRegionId	INT

MorphologicalProcesses	
*MorphologicalProcessId	INT
*Description	VARCHAR(60)

MeaningSets	
*MeaningSetId	INT
*MeaningId	INT

Thesaurus	
*MeaningFromId	INT
*LemmaToId	INT
*LemmaTo	VARCHAR(80)
*ThesaurusRelationshipTypeId	INT

SeeAlsoLemmas	
*LemmaFromId	INT
*LemmaToId	INT

Meanings	
*MeaningId	INT
*Definition	VARCHAR(512)
°ExplanatoryImageFilepath	VARCHAR(255)
°UsageLabelId	INT
*LemmaId	INT

Lemmas	
*LemmaId	INT
*Headword	VARCHAR(80)
°Etymology	VARCHAR(255)
°MorphologicalProcessId	INT
°DialectalRegionId	INT

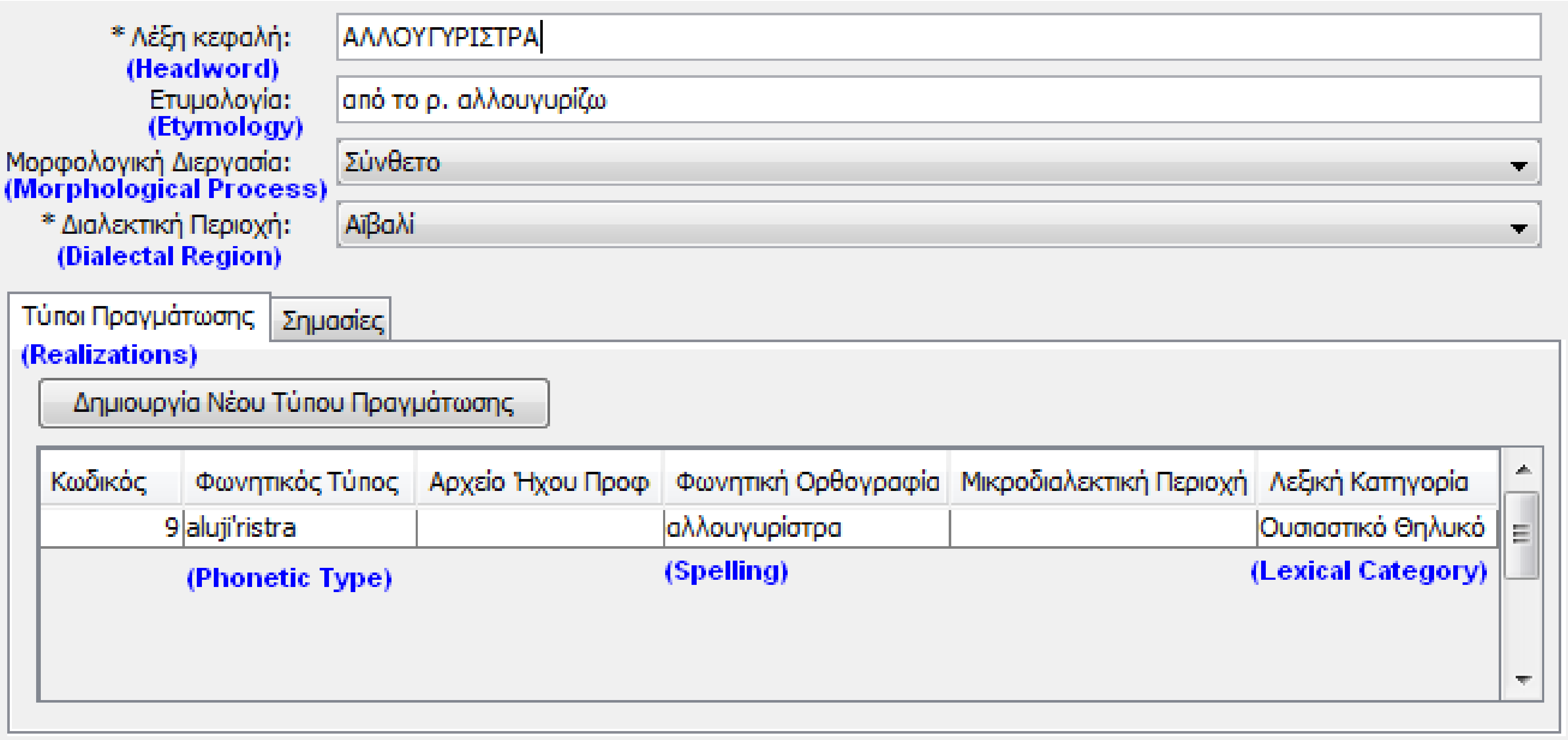
UsageLabels	
*UsageLabelId	INT
*Name	VARCHAR(40)

ThesaurusRelationshipTypes	
*ThesaurusRelationshipTypeId	INT
*Name	VARCHAR(20)

UsageExamples	
*UsageExampleId	INT
°Example	VARCHAR(400)
°Source	VARCHAR(400)
°ModernGreekTranslation	VARCHAR(400)
*MeaningId	INT

DialectalRegions	
*DialectalRegionId	INT
*Name	VARCHAR(60)

The realization types of lemma ΑΛΛΟΥΓΥΡΙΣΤΡΑ

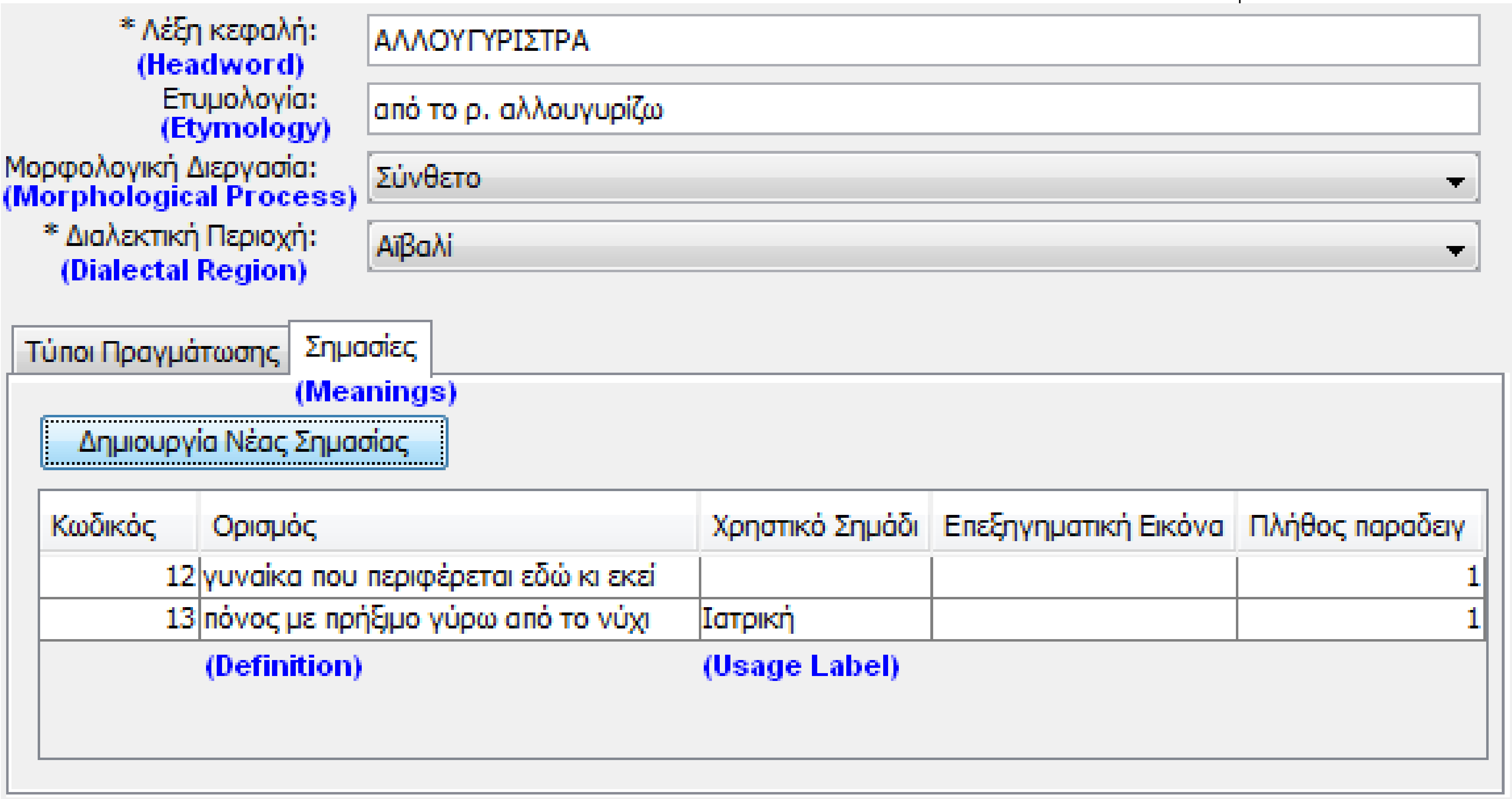


* Λέξη κεφαλή: **(Headword)**	ΑΛΛΟΥΓΥΡΙΣΤΡΑ
Ετυμολογία: **(Etymology)**	από το ρ. αλλουγυρίζω
Μορφολογική Διεργασία: **(Morphological Process)**	Σύνθετο
* Διαλεκτική Περιοχή: **(Dialectal Region)**	Αϊβαλί

(Realizations)

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

The (two) meanings of lemma ΑΛΛΟΥΓΥΡΙΣΤΡΑ



* Λέξη κεφαλή: **(Headword)** ΑΛΛΟΥΓΥΡΙΣΤΡΑ

Ετυμολογία: **(Etymology)** από το ρ. αλλουγυρίζω

Μορφολογική Διεργασία: **(Morphological Process)** Σύνθετο

* Διαλεκτική Περιοχή: **(Dialectal Region)** Αϊβαλί

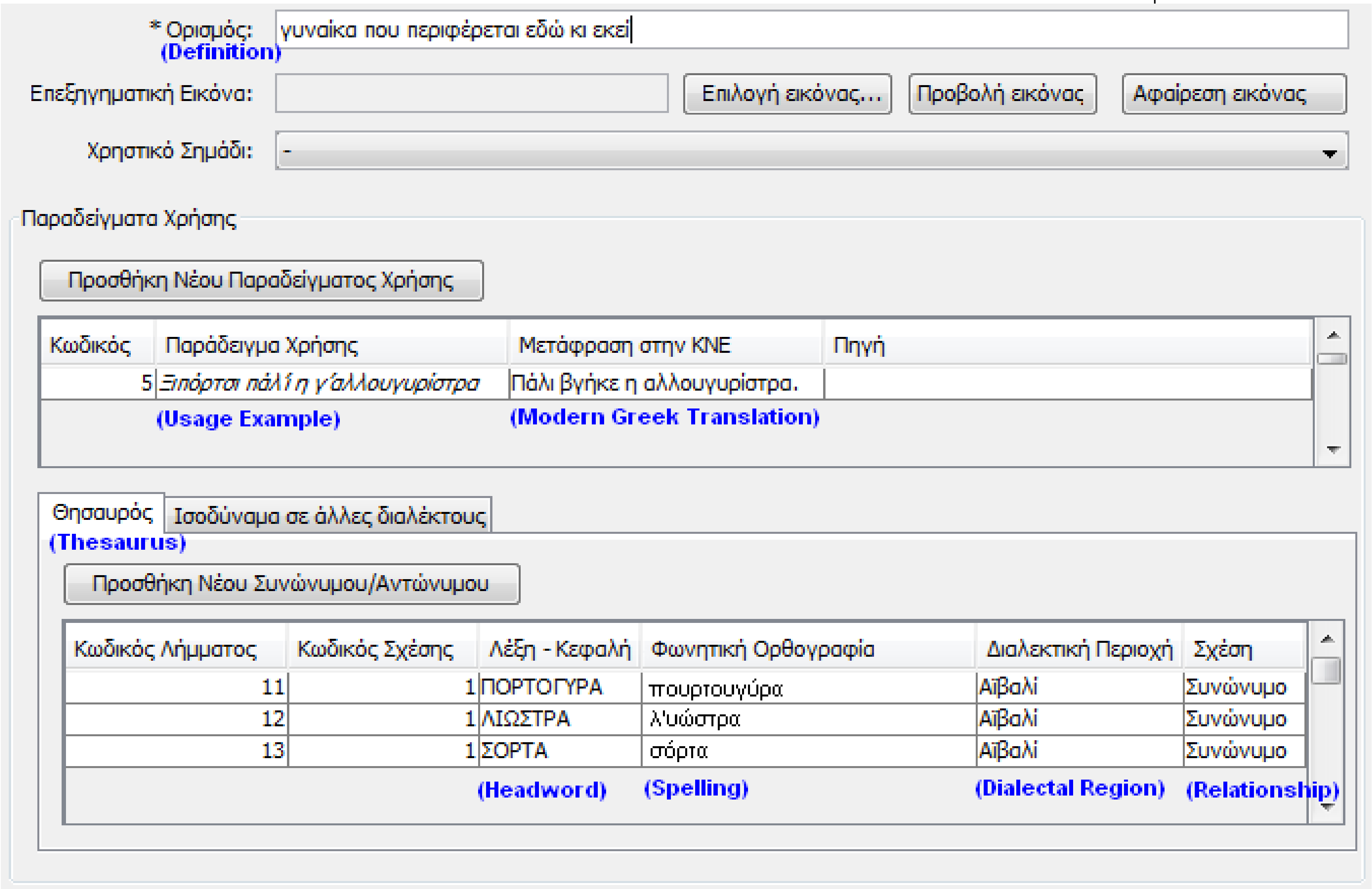
Τύποι Πραγμάτωσης | Σημασίες **(Meanings)**

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)**

Usage Examples and Synonyms of the first meaning of lemma ΑΛΛΟΥΓΥΡΙΣΤΡΑ



* Ορισμός: (Definition)	γυναίκα που περιφέρεται εδώ κι εκεί
Επεξηγηματική Εικόνα:	
Χρηστικό Σημάδι:	-

Επιλογή εικόνας...
Προβολή εικόνας
Αφαίρεση εικόνας

Παραδείγματα Χρήσης

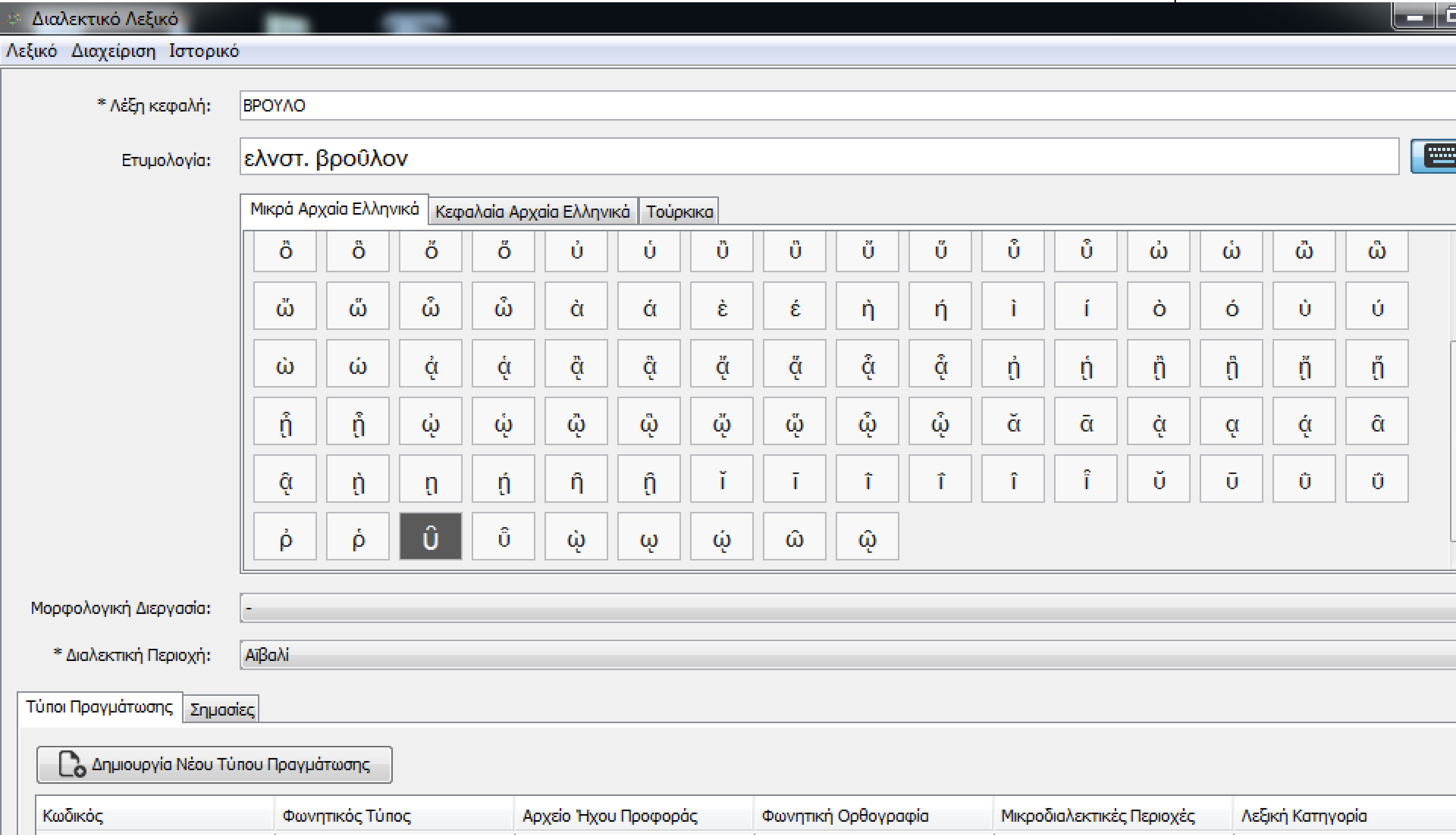
Προσθήκη Νέου Παραδείγματος Χρήσης

Κωδικός	Παράδειγμα Χρήσης	Μετάφραση στην ΚΝΕ	Πηγή
5	*Ξπόρτσι πάλ'ί η γ'αλλουγυρίστρα*	Πάλι βγήκε η αλλουγυρίστρα.	
	(Usage Example)	(Modern Greek Translation)	

Θησαυρός
(Thesaurus)
Ισοδύναμα σε άλλες διαλέκτους
Προσθήκη Νέου Συνώνυμου/Αντώνυμου

Κωδικός Λήμματος	Κωδικός Σχέσης	Λέξη - Κεφαλή	Φωνητική Ορθογραφία	Διαλεκτική Περιοχή	Σχέση
11	1	ΠΟΡΤΟΓΥΡΑ	πουρτουγύρα	Αϊβαλί	Συνώνυμο
12	1	ΛΙΩΣΤΡΑ	λ'υώστρα	Αϊβαλί	Συνώνυμο
13	1	ΣΟΡΤΑ	σόρτα	Αϊβαλί	Συνώνυμο
		(Headword)	(Spelling)	(Dialectal Region)	(Relationship)

Keyboards for entering Etymology



Διαλεκτικό Λεξικό

Λεξικό
Διαχείριση
Ιστορικό

* Λέξη κεφαλή: ΒΡΟΥΛΟ

Ετυμολογία: ελνστ. βροῦλον

Μικρά Αρχαία Ελληνικά
Κεφαλαία Αρχαία Ελληνικά
Τούρκικα
ὂ ὃ ὄ ὅ ὐ ὑ ὒ ὓ ὔ ὕ ὖ ὗ ὠ ὡ ὢ ὣ
ὤ ὥ ὦ ὧ ὰ ά ὲ έ ὴ ή ὶ ί ὸ ό ὺ ύ
ὼ ώ ᾀ ᾁ ᾂ ᾃ ᾄ ᾅ ᾆ ᾇ ᾐ ᾑ ᾒ ᾓ ᾔ ᾕ
ᾖ ᾗ ᾠ ᾡ ᾢ ᾣ ᾤ ᾥ ᾦ ᾧ ᾰ ᾱ ᾲ ᾳ ᾴ ᾶ
ᾷ ῂ ῃ ῄ ῆ ῇ ῐ ῑ ῒ ΐ ῖ ῗ ῠ ῡ ῢ ΰ
ῤ ῥ ῦ ῧ ῲ ῳ ῴ ῶ ῷ

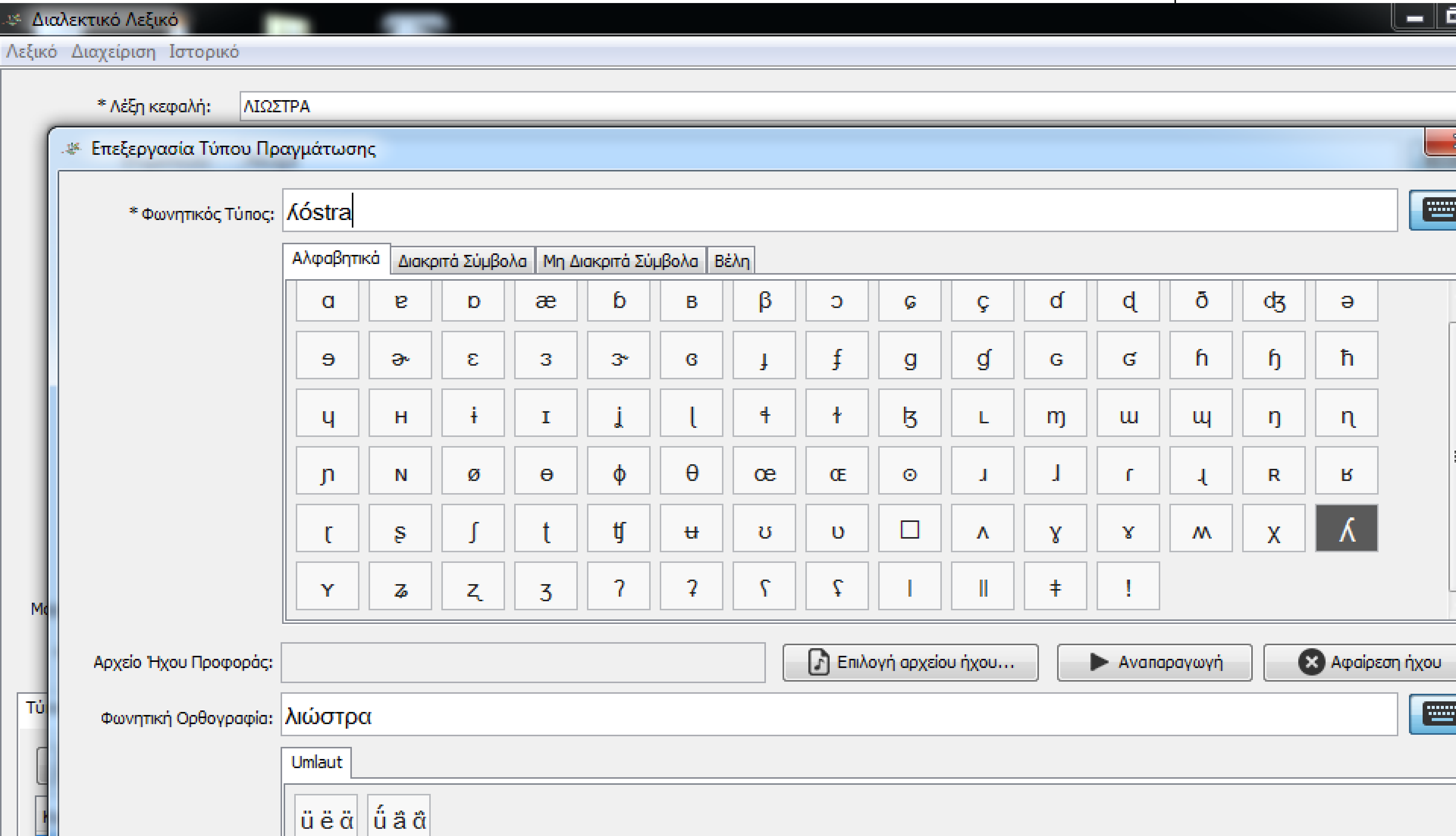
Μορφολογική Διεργασία: -

* Διαλεκτική Περιοχή: Αϊβαλί

Τύποι Πραγμάτωσης
Σημασίες
Δημιουργία Νέου Τύπου Πραγμάτωσης

Κωδικός	Φωνητικός Τύπος	Αρχείο Ήχου Προφοράς	Φωνητική Ορθογραφία	Μικροδιαλεκτικές Περιοχές	Λεξική Κατηγορία

Keyboards for entering Phonetic type and Spelling


Διαλεκτικό Λεξικό
Λεξικό Διαχείριση Ιστορικό
Επεξεργασία Τύπου Πραγμάτωσης
Αλφαβητικά
Διακριτά Σύμβολα
Μη Διακριτά Σύμβολα
Βέλη
ɑ ɐ ɒ æ ɓ ʙ β ɔ ɕ ç ɗ ɖ ð ʤ ə
ɘ ɚ ɛ ɜ ɝ ɞ ɟ ʄ ɡ ɠ ɢ ʛ ɦ ɧ ħ
ɥ ʜ ɨ ɪ ʝ ɭ ɬ ɫ ɮ ʟ ɱ ɯ ɰ ŋ ɳ
ɲ ɴ ø ɵ ɸ θ œ ɶ ʘ ɹ ɺ ɾ ɻ ʀ ʁ
ɽ ʂ ʃ ʈ ʧ ʉ ʊ ʋ □ ʌ ɣ ɤ ʍ χ ʎ
ʏ ʑ ʐ ʒ ʔ ʡ ʕ ʢ ǀ ǁ ǂ ǃ
Επιλογή αρχείου ήχου...
Αναπαραγωγή
Αφαίρεση ήχου
Umlaut
ü ë ä
ű â ᾶ

* Λέξη κεφαλή: ΛΙΩΣΤΡΑ

* Φωνητικός Τύπος: ʎóstra

Αρχείο Ήχου Προφοράς:

Φωνητική Ορθογραφία: λιώστρα

Printing lemmas

Λέξη Κεφαλή: ΑΓΓΕΙΟ
Ετυμολογία: [μσν. αγγείον]
Διαλεκτική Περιοχή: Αϊβαλί

Τύποι Πραγμάτωσης:
1)
Φωνητικός Τύπος: /aŋ□íu/
Φωνητική Ορθογραφία: αγγείου
Λεξική Κατηγορία: Ουσιαστικό Ουδέτερο

Σημασίες:
1)
Ορισμός: ουροδοχείο
Χρηστικό Σημάδι: Οικοκυρική
Επεξηγηματική Εικόνα: AA01003.jpg
Παραδείγματα Χρήσης:
1)
Παράδειγμα Χρήσης: *Μουρή Βασιλ'κώ, φέρι τ_αγγείου απάνου γιατ_έφαγα πουλ'ύ καρπούζ τσι του βράδ θα_ν_έχουμι πλαλ'τήρια!*
Μετάφραση στη ΚΝΕ: Μαρή Βασιλική, φέρε το ουροδοχείο επάνω γιατί έφαγα πολύ καρπούζι και το βράδυ προβλέπεται να έχουμε τρεχάματα!.

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.

Dialectal lexicon building: requirements and technical specifications

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

Nikitas N. Karanikolas

nnk@teiath.gr

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