

Искуства у изградњи вишедијалекатског речника

УДК 811.14'06'374'282.2.

САЖЕТАК: Рад представља искуства стечена током комплетног циклуса изграђивања електронског дијалекатског речника. Иновативност представљеног приступа огледа се у осмишљавању речника који је вишедијалекатски (тројезични), при чему су три дијалекта супротстављена једном циљном језику. Наш приступ отвара могућности које не би биле доступне при раду са три једнојезична дијалекатска речника. Поред наведеног, током успостављања и активирања система искрсавали су веома специфични захтеви за унапређивање система, иницирани од стране корисника, који су се односили посебно на фазу анализе. Стечена искуства и изнађена решења током тежње ка крајњој оптимизацији система у циљу испуњавања свих захтева могу бити корисна за истраживања и пројекте сличне намене.

КЉУЧНЕ РЕЧИ: рачунарска дијалектологија, дијалекатска лексикографија, електронски речници, лексички ресурси, савремени грчки дијалекти, малоазијски грчки.

РАД ПРИМЉЕН: 8. фебруар 2017.

РАД ПРИХВАЋЕН: 28. април 2017.

Никитас Н. Караниколас
nnk@teiath.gr

Елени Галиоту
egali@teiath.gr

*Технолошко-образовни
институт у Атини
Одсек за информатику*

*Превод с енглеског:
Зоран Ристовић*

1. Увод

У претходном раду (Karaniokolas et. al., 2013), представљен је принцип изградње и начин примене мултимедијалног електронског речника три грчка дијалекта у Малој Азији (понтског, кападочанског и

ајваличког). Представљен је лингвистички и лексикографски приступ, као и принципи изградње макро и микроструктуре речника. Приказан је и концептуални модел вишедијалекатског речника и одговарајућа релациона шема. Према наведеним параметрима осмишљен је и имплементиран систем који садржи леме и релевантне лексикографске информације из три малоазијска грчка дијалекта.

Међутим, током развојног циклуса система, уз помоћ високо квалификованих корисника, искрсавали су веома специфични захтеви за побољшања који су омогућили да постигнемо крајњи циљ, а то је одличан систем. У овом раду приказујемо искуства стечена током изградње система и представљамо унапређени дизајн вишедијалекатског речника и његове додатне могућности. Сматрамо да наш систем може да се примени и на друге грчке дијалекте, а побољшани дизајн може да послужи као база за вишедијалекатске речнике других језика. Систем се може користити за вишедијалекатске речнике других језика ако се обезбеди повезивање са одговарајућим виртуелним тастатурама.

Редослед излагања организован је према следећем плану. Одељак 2. образлаже мотивацију за изградњу речника три дијалекта, а рад на речнику представљен је у одељку 3. Одељци 4. и 5. описују почетне захтеве и нацрт који представља одговор на те захтеве. Одељак 6. износи детаље о првој имплементираној верзији система. Одељак 7. представља захтеве за побољшањима и како је на те захтеве одговорено приликом имплементације. Резултат наведених фаза оптимизације је систем кога одликује функционална изузетност – његов нацрт је тема одељка 8. док су закључна разматрања дата у одељку 9.

2. Мотивација

Понтски, кападочански и ајвалички су три грчка дијалекта Мале Азије која нису довољно документована и којима прети изумирање. До сада је мало пажње поклоњено овим дијалектима. Најинтересантнији изузетак представља Пападопулосов историјски речник понтског дијалекта (Papadopoulos, 1958). Кападочански се помиње у неким другим радовима (Thomason, 2001; Thomason и Kaufman, 1988). Постоје и неки глосари малоазијских грчких дијалеката који садрже речи и идиоматске изразе као и њихово значење на стандардном савременом грчком. Међутим, у већини ових глосара леме су изложене на веома несистематичан начин док суштинске информације као што су изговор и употреба недостају. Поред тога постоји и недоследност у навођењу

глагола – неки глаголи наводе се искључиво у прошлом времену а неки у садашњем. Услед наведених околности темељна анализа грчких дијалеката Мале Азије намеће се као витални приоритет да би се добио увид у природу и механизме језичких промена у домену дијалекатских варијација.

Ови и други релевантни социјални фактори мотивисали су нас да покренемо пројекат АМиGre, у оквиру програма THALIS. Акроним пројекта потиче од кровног наслова иницијативе: „Понт, Кападокија, Ајвалик: у потрази за грчким Мале Азије” (енг. Pontus, Cappadocia, Aivali: in search of Asia Minor Greek). Један од резултата пројекта било је осмишљавање и имплементација мултимедијалног вишедијалекатског речника три грчка дијалекта Мале Азије (понтског, кападочанског и ајваличког), о којима ће бити речи у даљем излагању.

Дијалекатски речници обично се третирају као једнојезични синхронијски речници. У нашем случају (АМиGre), уместо да креирамо три једнојезична дијалекатска речника, одлучили смо да их посматрамо заједно и конципирамо тројезични речник (три малоазијска дијалекта наспрам стандардног савременог грчког језика). У техничком смислу ово је најинтересантнија мотивација која доноси и занимљиву новину јер се омогућава унакрсно референцирање на релацији лема-лема, било да су у питању леме истог дијалекта или различитих дијалеката. Ово не би било изводљиво да су у питању три једнојезична дијалекатска речника.

3. Релевантни радови

Електронска лексикографија модерног грчког језика се донедавно није бавила стварањем дијалекатских речника. Онлајн речници развијени на Порталу за грчки језик ([Online, 2016](#)) обухватају рачунарске верзије Георгакасовог *Грчко-енглеског речника*, Триндафилиадесовог *Речника стандардног савременог грчког речника* и обратни речник који су приредили Анастизијади и Симеониди. Поред тога, Портал пружа могућност приступа рачунарској верзији Кријарасовог *Малог речника средњовековне народне грчке књижевности*. Институт за обраду језика и говора развио је двојезичне онлајн речнике (грчко-енглески, грчко-немачки, грчко-руски, грчко-турски и грчко-арапски). Поменути речници се стално развијају и допуњавају а доступни су путем ([ILSP, 2016](#)). Поред тога, развијају се и алати за обраду природних језика намењени подршци лексикографских

апликација. У прилог томе, у раду (Tsalidis et. al., 2010) се говори о инфраструктурним алатима који се користе за кодирање морфолошких, синтаксичких и семантичких информација, као и о алатима за коректуру, као што је провера спеловања, поделе речи на крају ретка, итд. Кад је реч о грчким дијалектима, једини нама познати рачунарски обрађени речник је онлајн лексичка база података за кипарски грчки (Themistocleous, 2012). Окружење овог онлајн речника нуди механизме за напредну претрагу као и за претварање текста у говор ради изговора кипарског грчког.

4. Почетни захтеви

Дијалекатски речници обично се третирају као једнојезични синхронијски речници (Béjoint, 2000; Geeraerts, 1989), услед ограничења у микроструктури (општој организацији леме) (Landau, 2001; Zgusta, 1971). Узимајући у обзир да је наша основна намера била да се осмисли и изгради онлајн речник, његова макроструктура не би подлежала физичким ограничењима која су карактеристична за штампане речнике и они би могли да нуде (виртуелно) „вишеструке макроструктуре“ које би одражавале различите опције претраге које смо били у могућности да иградимо (Burke, 2003). С тога, будући да нису постојала ограничења у макроструктури, определили смо се за изградњу тројезичног речника (три малоазијска грчка дијалекта наспрам стандардног савременог грчког језика) (Xydopoulos и Ralli, 2012), уместо три једнојезична дијалекатска речника. Речник је назван TDGDAM (енг. Tri-Dialectal Greek Dictionary of Asia Minor – тродијалекатски речник малоазијског грчког) и он треба да представља лингвистички компетентан тродијалекатски речник у електронском облику. Један од основних захтева TDGDAM-а био је да корисници могу приступити графичком (заснованом на облицима) представљању сваке леме које им допушта да рукују изговором, значењем, употребом и односом са другим лемама. Уређивање овако представљених лема треба да буде омогућено а за то је било потребно да се користе усвојени скупови карактера. Између осталог, свака лема би требало да садржи податке о дијалекатској зони и извору из којег је екстрахована. Овакав тип речника представља иновацију не само када је у питању грчки језик са својим дијалектима, него и када су у питању међународни стандарди, о чему ће бити више речи у даљем излагању.

ΠΕΔΙΟ (FIELD)	ΛΗΜΜΑΤΙΚΗ ΠΛΗΡΟΦΟΡΙΑ (LEMMA VALUES)
1. ΛΕΞΗ-ΚΕΦΑΛΗ / HEADWORD	ΒΡΟΥΛΟ
2. ΛΕΞΙΚΗ ΚΑΤΗΓΟΡΙΑ (lexical category)	(Ο. ουδ.) (noun, neuter)
3. ΦΩΝΗΤΙΚΟΣ ΤΥΠΟΣ (phonetic type)	[ˈvrulo]
4. ΑΡΧΕΙΟ ΗΧΟΥ ΠΡΟΦΟΡΑΣ (digital record)	WAV
5. ΕΝΑΛΛΑΚΤΙΚΟΙ ΤΥΠΟΙ (alternative types)	Βρόλους [ˈvrolus] (Παμφ. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ ΣΥΝΔΕΣΗ) (Pamph. <i>microdialectal region</i>)
6. ΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (dialectal region)	Αϊβαλί (Aivali)
7. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (microdialect)	
8. ΜΟΡΦΟΛΟΓΙΚΗ ΔΙΕΡΓΑΣΙΑ (morphological process)	-
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ (usage)	1. ΦΥΤΟΛΟΓΙΑ (phytology)
10. ΟΡΙΣΜΟΣ (definition)	Βούρλο (bulrush - reedy plant)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	JPG
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ (usage example)	«Έκουσα καμπόσα βρούλα κι πέρασα αρμαθιά τα ψάρια πόπιασα σήμερα» (a dialectal sentence using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ (modern greek translation)	(Έκοψα μερικά βούρλα και) (the same sentence in Modern Greek)
14. ΘΗΣΑΥΡΟΣ (thesaurus)	
15. ΕΤΥΜΟΛΟΓΙΚΗ ΠΛΗΡΟΦΟΡΙΑ (etymology)	[ΕΤΥΜ ελνστ. βρούλον] (originates from the Hellenistic βρούλον)
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ	2. ΥΠΟΤΙΜΗΤΙΚΟ (pejorative)
10. ΟΡΙΣΜΟΣ	Ανόητος (fatuous)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	JPG
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ	«Ντιπ για ντιπ βρούλου τούτου του πιδι». (a dialectal sentence using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ	(Τελείως ανόητο αυτό το παιδί) (the same sentence in Modern Greek)
14. ΘΗΣΑΥΡΟΣ	---

Слика 1. Нацрт структурног приказа леме ΒΡΟΥΛΟ

Узимајући у обзир предвиђени географски и временски обухват, TDGDAM је осмишљен као локални дијалекатски речник несинхронијске природе који би требало да обухвати одреднице из различитих просторних зона и временских периода (Penhalurik, 2009). Како је одлучено на самом почетку, леме које улазе у TDGDAM требало би да долазе, директно или индиректно, из говорног или писаног материјала сваког од заступљених дијалекатских варијетета (Keymeulen, 2010).

У погледу микроструктуре TDGDAM-а, наш циљ је био да укључимо формалне информације о изговору (фонетски облик), граматички (категоријалне и морфолошке информације), пореклу (етимологија), значењу (синонимске и дескриптивне дефиниције), употреби (ознаке теме и регистра) као и да обезбедимо везе ка мултимедијалним изворима (унутрашњим или спољашњим у односу на TDGDAM) да би обогатили семантички и прагматички опис лема (Barbato и Varvaro, 2004; Rys и Keymeulen, 2009; Хυδopoulos и Ralli, 2012). Како би се избегле разлике и произвољности у писању у оквиру истог дијалекта (Durkin,

2010; *Xydopoulos, 2012*), одреднице се не појављују у „полуфонетској“ транскрипцији већ се наводе записане великим словима у ортографском облику. Наиме, ортографски облик записан великим словима одудара од писања у стандардном дијалекту; он није прескриптиван када је у питању ортографија дијалекта и допушта да се у микроструктури појаве било који алтернативни ортографски облици (*Markus и Heuberger, 2007; Xydopoulos, 2012*). Коначно, аутентични примери употребе сматрани су основним саставним делом одредница које се појављују написане у нестандартном облику, одсликавајући што приближније изговор уз употребу дијакритика, али избегавањем „полуфонетске транскрипције“ (*Rys и Keumeulen, 2009*).

ΠΕΔΙΟ (FIELD)	ΛΗΜΜΑΤΙΚΗ ΠΛΗΡΟΦΟΡΙΑ
1. ΛΕΞΗ-ΚΕΦΑΛΗ / HEADWORD	ΛΙΩΣΤΡΑ
2. ΛΕΞΙΚΗ ΚΑΤΗΓΟΡΙΑ (lexical category)	Ο. Θηλ. (noun, feminine)
3. ΦΩΝΗΤΙΚΟΣ ΤΥΠΟΣ (phonetic type)	[ˈliɔstra]
4. ΑΡΧΕΙΟ ΗΧΟΥ ΠΡΟΦΟΡΑΣ (digital record)	αρχείο WAV
5. ΕΝΑΛΛΑΚΤΙΚΟΙ ΤΥΠΟΙ (alternative types)	
6. ΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (dialectal region)	Αϊβαλί (Aivali)
7. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (microdialectal region)	
8. ΜΟΡΦΟΛΟΓΙΚΗ ΔΙΕΡΓΑΣΙΑ (morphological process)	Παραγωγή (derivation)
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ (usage)	1. ΥΠΟΤΙΜΗΤΙΚΟ (pejorative)
10. ΟΡΙΣΜΟΣ (definition)	γυναίκα που περιφέρεται εδώ κι εκεί (woman who strolls)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ (picture)	-
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ (usage example)	«Ξιπόρτσι πάλι-η λ'ώστρα» (dialectal phrase using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ (Modern Greek translation)	(Ξιπόρτισε πάλι η γυρίστρα) (the same in Modern Greek)
14. ΘΗΣΑΥΡΟΣ (thesaurus)	Συν: αλλουγυρίστρα, σόρτα, τακού (Synonymy: αλλουγυρίστρα, σόρτα, τακού)
15. ΕΤΥΜΟΛΟΓΙΚΗ ΠΛΗΡΟΦΟΡΙΑ (etymology)	[<λιέμι με -ωστρα ίσως από επιδρ. άλλων θηλυκών σε -ωστρα] (λιέμι with affix ώστρα)
16. ΔΙΑΠΑΡΑΠΟΜΠΕΣ (see also)	

Слика 2. Нацрт структурног приказа леме ΛΙΩΣΤΡΑ

Кад су у питању могућности унакрсног повезивања ставки TDGDAM-а, дефинисали смо три основне потребе: унакрсно референцирање ка другим одредницама, које су повезане било кроз деривационе процесе или кроз семантичке релације; повезивање лема са еквивалентним значењима из различитих дијалеката; синонимске и антонимске релације.

Слике 1, 2 и 3 представљају нацрт структурних описа са подједнаким бројем лема које би TDGDAM требало да садржи. На основу овог

ΠΕΔΙΟ	ΛΗΜΜΑΤΙΚΗ ΠΛΗΡΟΦΟΡΙΑ
1. ΛΕΞΗ-ΚΕΦΑΛΗ / HEADWORD	ΑΛΛΟΥΓΥΡΙΣΤΡΑ
2. ΛΕΞΙΚΗ ΚΑΤΗΓΟΡΙΑ	(Ον. Θηλ.) (noun, feminine)
3. ΦΩΝΗΤΙΚΟΣ ΤΥΠΟΣ	[aluji'ristra]
4. ΑΡΧΕΙΟ ΗΧΟΥ ΠΡΟΦΟΡΑΣ	WAV
5. ΕΝΑΛΛΑΚΤΙΚΟΙ ΤΥΠΟΙ	---
6. ΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ	Αϊβαλί (Aivali)
7. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ	---
8. ΜΟΡΦΟΛΟΓΙΚΗ ΔΙΕΡΓΑΣΙΑ	Σύνθετο (composite)
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ	1. ΥΠΟΤΙΜΗΤΙΚΟ (pejorative)
10. ΟΡΙΣΜΟΣ	γυναίκα που περιφέρεται εδώ κι εκεί (woman who strolls)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	---
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ	"Ξιπόρτσι πάλι η γ' αλλουγυρίστρα" (dialectal phrase using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ	(Πάλι βγήκε η αλλουγυρίστρα.) (the same in Modern Greek)
14. ΘΗΣΑΥΡΟΣ	ΣΥΝ λυώστρα, σόρτα (Synonymy: λυώστρα, σόρτα)
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ	2. ΙΑΤΡΙΚΗ (medicine)
10. ΟΡΙΣΜΟΣ	πόνος με πρήξιμο γύρω από το νύχι (pain & turgescence around nail)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	JPG
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ	"Έχου μνιαν αλλουγυρίστρα στο δαχτύλι μ τσι πουν'εί" (dialectal phrase using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ	(Έχω μια αλλουγυρίστρα στο δάχτυλο και με πονάει.) (the same in Modern Greek)
14. ΘΗΣΑΥΡΟΣ	---
15. ΕΤΥΜΟΛΟΓΙΚΗ ΠΛΗΡΟΦΟΡΙΑ	[ΕΤΥΜ από το ρ. αλλουγυρίζω] (from the verb αλλουγυρίζω)
16. ΔΙΑΠΑΡΑΠΟΜΠΕΣ	---

Слика 3. Нацрт структурног приказа леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ

и сличних нацрта структурних приказа конципирани смо TDGDAM систем.

Раније коришћени појмови синонимије (Synonym, 2016) и антонимије (Opposite, 2016), као и појмови хомонимије (Homonym, 2016) и полисемије (Polysemy, 2016) које ћемо користити касније јасно су дефинисани. Њихове дефиниције могу се пронаћи на интернету.

5. Нацрт

На основу анализе представљене у одељку о почетним захтевима и на основу нацрта структурних приказа (видети слике 1, 2 и 3) као резултат је добијена следећа структура лема:

- Одредница, дијалекат (дијалекатско подручје), морфолошка информација или процес и етимологија су примарне информације са јединственим вредностима које удружене формирају лему и зависе од ње.

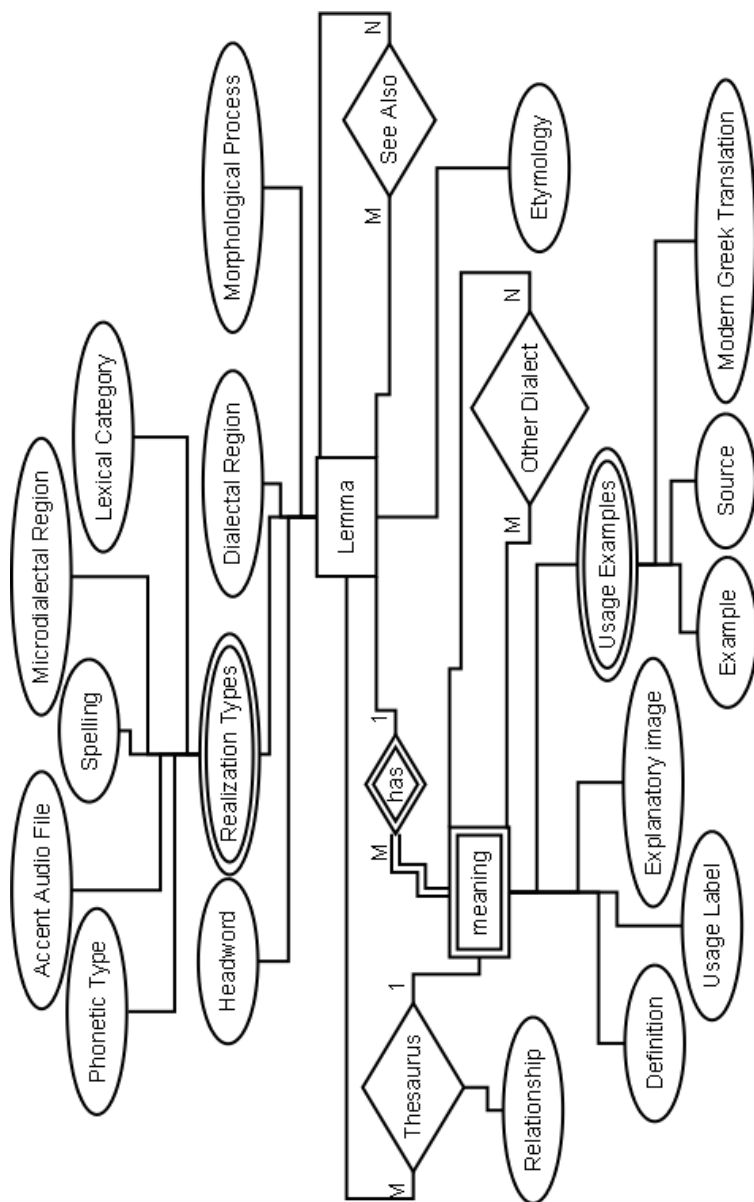
- Свака лема може да има бројне различите реализације а сваку од њих карактерише у извесној мери различита фонетска реализација директно зависна од микродијалекатског подручја са којег потиче – то су специфичне зоне у оквиру ширег дијалекатског подручја у оквиру којег се јавља дата лема.
- Свака лема потенцијално може имати различита значења (појава полисемије), или испољавати хомонимију са другим семантички различитим лемама.
- За свако појединачно значење, различити примери употребе су од суштинског значаја.

Када су у питању релације (између лема и њихових значења) донели смо следеће закључке:

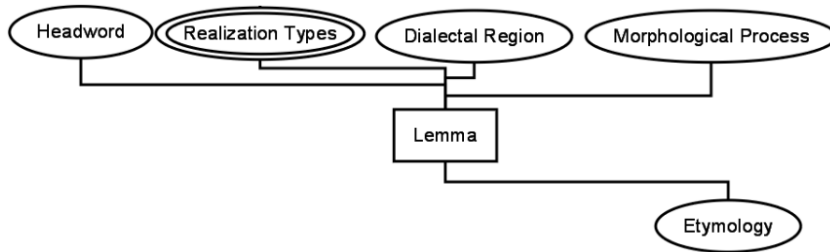
- Унакрсне референце („види и“) могу да буду доступне за повезивање лема које су семантички, прагматички, морфолошки или етимолошки повезане једна са другом.
- *синоними* и *антоними* су две семантичке релације примењиве између различитих лема. Обе наведене релације односе се на значење једне леме у односу на другу, реферисану лему. Синонимске и антонимске везе су примењиве на односе између значења леме и леме који потичу из истог дијалекта.
- Постоје значења различитих лема из различитих дијалеката која деле исту дефиницију. Ова релација обележена је са „други дијалекат“. Насупрот другим релацијама, „други дијалекат“ је симетрична релација.

Целокупна идеја (структура лема и њихових релација) строго је дефинисана дијаграмом ентитет-везе на слици 4.

Четири речника података који следе (табеле 1, 2, 3 и 4) објашњавају четири дела (подшеме) целокупног дијаграма ентитет-везе.



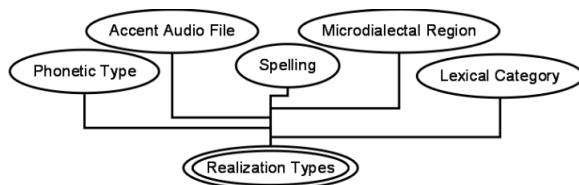
Слика 4. Дијаграм ентитет-везе



Атрибут	Дефиниција	Формат података	пример
<i>Headword</i> (одредница)	Канонски облик речи	Ниска која садржи само велика слова грчког алфабета	ΑΛΛΟΥΓΓΡΙΣΤΡΑ
<i>Etymology</i> (етимологија)	Основне информације о пореклу речи	Ниска записана на грчком са полигоналним акцентима	Από το ρήμα αλλουγυρίζου (од глагола aluji'rizu)
<i>Morphological Process</i> (морфолошки процес)	Различити процеси укључени у формирање речи	Вредност из унапред утврђене листе морфолошких процеса	Σύνθετο (сложеница)
<i>Dialectal region</i> (дијалекатски регион)	Регион или дијалекат у оквиру којих је лема пронађена	Вредност из унапред утврђене листе дијалеката	Αἰβάλι (AIVALI)

Табела 1. Речник података за *Lemma* (лема)

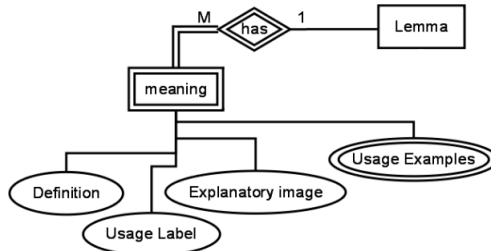
Једна од могућих имплементација концептуалног модела ентитет-везе коришћењем релационе базе података приказана је релационом шемом на слици 5 која садржи тринаест табела. Међутим, од примарног значаја је седам табела. Осталих шест су помоћне табеле које садрже скупове постојећих доступних вредности за нека поља из осталих седам важних табела. Табеле од примарног значаја су на слици 5 посебно истакнуте подебљавањем ивица и повећавањем фонта њиховог назива. Четири од седам главних табела односе се на релационе еквиваленте главног концептуалног ентитета *Lemma* (лема), слабог ентитета *Meaning* (значење) као и на два сложена атрибута са вишеструким вредностима *Realization Types* (типови реализације) и *Usage Examples* (примери



Атрибут	Дефиниција	Формат података	Пример
<i>Phonetic Type</i> (фонетски тип)	Фонетска транскрипција изговора речи који се изучава	Ниска која садржи карактере Међународног фонетског алфабета (IPA).	aluji'ristra
<i>Accent Audio</i> (аудио акцента)	Аудио датотека која садржи аутентични изговор речи	Ниска која садржи путању датотеке	http://amigre.gr/xyzR1.wav
<i>Spelling</i> (спелинг)	Нестандардна граfiјска репрезентација изговора у складу са правописним правилима стандардног грчког, комбинована са дијакритицима да би се аотирале фонолошке алтернације	Ниска која садржи слова грчког алфабета и друге дијакритичке симболе (акценте, цртице, заграде и апострофе)	αλλουγυρίστρα
<i>Microdialectal Region</i> (микро-дијалекатски регион)	Име специфичне зоне унутар ширег дијалекатског подручја у коме је реализовани облик леме пронађен	Вредност из унапред прописане листе микро-дијалекатских подручја	
<i>Lexical Category</i> (лексичка категорија)	Врста речи и род	Вредност из унапред прописане листе лексичких категорија	Ουσιαστικό Θηλυκό (именица женског рода)

Табела 2. Речник података за *Realization Types* (типови реализације)

употребе). Три преостале важне табеле су релациони еквиваленти концептуалних релација – *See Also* (види и), *Thesurus* (тезаурус) и *Other Dialect* (други дијалекат).

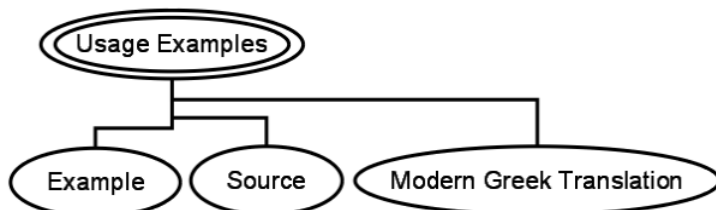


Атрибут	Дефиниција	Формат података	Пример
<i>Definition</i> (дефиниција)	Кратак опис значења леме	Ниска из стандардног савременог грчког	Γυναίκα που περιφέρεται εδώ κι εκεί (“жена која иде наоколо”)
<i>Explanatory Image File</i> (датотека слике за објашњење)	Слика која илуструје значење леме	Ниска која садржи путању датотеке	http://amigre.gr/xyzM1.png
<i>Usage Label</i> (обележје коришћења)	Формална индикација контекста у коме се лема користи (стил, регистар и др.)	Вредност из унапред прописане листе домена	ΥΠΟΤΙΜΗΤΙΚΟ (пежоратив)

Табела 3. Речник података за *Meaning* (значење)

Само табела *Meaning Sets* (скупови значења), као имплементација концептуалне релације *Other Dialect* (други дијалекат), захтева додатно објашњење. Ова релација је симетрична по природи, тј. кад год се за значење одређене леме из неког дијалекта утврди да је еквивалентно значењу неке леме из неког другог дијалекта, онда важи и обрнуто. Управо табела „скупови значења“ и логика саме апликације осигуравају ту симетрију. Преостале две релације, „види и“ и „тезаурус“, нису само по себи симетричне. Ово се читава у релационој шеми као и у логици саме апликације. Последица тога је да корисник мора да дефинише релацију

у оба смера ако је у неком случају релација „види и“ или „тезаурус“ симетрична.



Атрибут	Дефиниција	Формат података	Пример
<i>Usage example</i> (примери употребе)	Фраза или реченица која служи као пример употребе леме специфичног значења у оригиналном дијалекту	Цео пример (цела ниска) је записан словима грчког алфабета и другим дијакритичким симболима (акценти, цртице, заграде и апострофи)	Ξιπόρτσι πάλι-η-γ'-αλλουγυρίστρα
<i>Standard Modern Greek Translation</i> (превод на стандардни савремени грчки)	Превод примера употребе на стандардни савремени грчки	Ниска на стандардном савременом грчком	Πάλι βγήχε η αλλουγυρίστρα
<i>Source</i> (извор)	Референца извора из кога је узет пример употребе	Ниска (може да буде књига, URL, итд.)	

Табела 4. Речник података за *Usage Examples* (примери употребе)

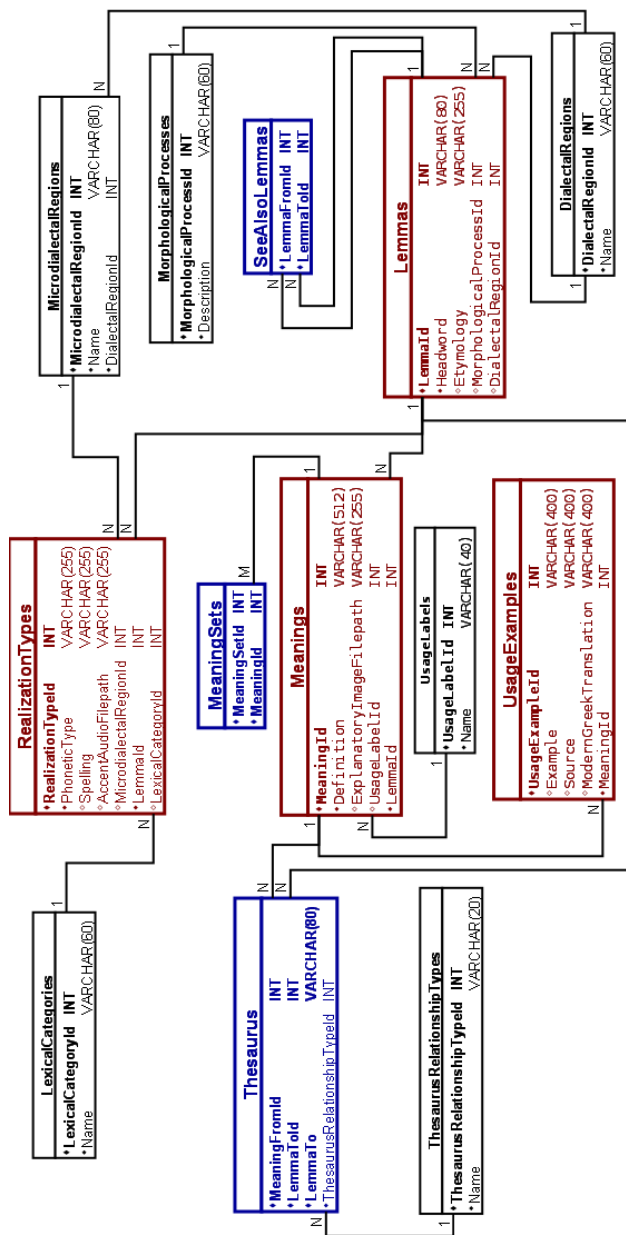
Међународни фонетски алфавет (енг. The International Phonetic Alphabet - IPA) који се користи у табели 2 – подшема *Realization Types* (типови реализације) – је алфаветски систем фонетске нотације који је примарно заснован на латиници. Њега је развила Међународне фонетска асоцијација за стандардизовано представљање звукова говорног језика. IPA стандарде користе лексикографи, студенти и наставници страних језика, лингвисти, патолози говорног језика, певачи, глумци, креатори вештачких језика и преводиоци. Сlike 6 и 7 представљају најкорисније IPA листе.

Други приступ фонетској нотацији је SAMPA (Фонетски алфабет заснован на методама процене говора – енг. Speech Assessment Methods Phonetic Alphabet) који представља машински читљив фонетски алфабет. Он је првобитно био развијен у оквиру пројекта ESPRIT 1541, SAM (Методe процене говора – енг. Speech Assessment Methods) у периоду 1987-1989. У почетку је примењен на дански, холандски, енглески, француски и италијански (1989), а касније је примењен и на норвешки и шведски (1992). Потом је примена проширена на грчки, португалски и шпански (1993) да би у скорије време био проширен и на бугарски, естонски, мађарски, пољски и румунски (1996).

6. Имплементација

GUI (графички кориснички интерфејс, енг. Graphical User Interface) верзија система заснива се на два формулара: „главни формулар“ и „формулар за значења“. Слика 8 представља главни формулар леме АЛЛОУГГҮРИСТРА. Главни формулар је подељен у три дела. Горњи део пружа информацију о одредници, етимологији, морфолошким процесима и дијалекту. Средишњи део је панел са две картице. Прва картица тог панела служи за приказ и уређивање реализација, док други служи за листе значења лема. Доњи део формулара служи за листу референци типа „види и“. Детаљнији опис средишњег дела формулара дат је на слици 9 која приказује њену другу картицу која даје листу значења за исту лему (представљену на слици 8).

Формулар значења леме позива се преко дугмета за активацију када корисник селекује неку ставку из листе значења која се налази у главном формулару. Слика 10 представља формулар значења који даје нека од значења леме АЛЛОУГГҮРИСТРА. Формулар значења је подељен у три дела. Горњи део даје дефиницију значења, а могу бити приложени



Слика 5. Логичка релациона шема

CONSONANTS (PULMONIC)

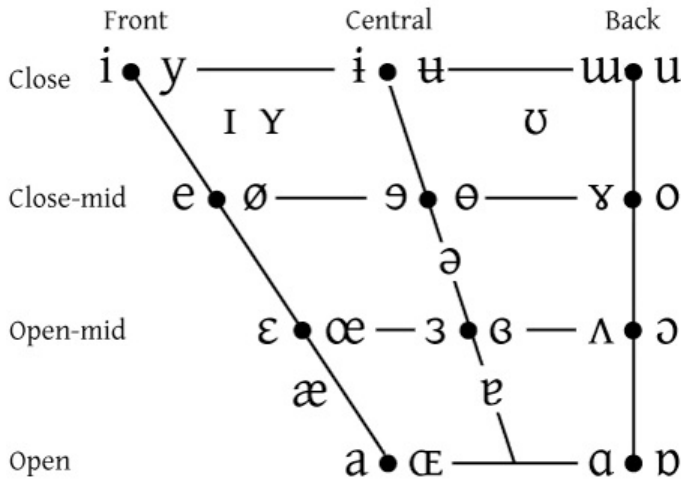
© 2005 IPA

	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			ʀ					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			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Слика 6. Међународни фонетски алфабет (ревидирана верзија из 2005) – сугласници (пулмонални)

VOWELS



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

Слика 7. Међународни фонетски алфабет – самогласници

* Λέξη κεφαλή:

(Headword)
 Ετυμολογία:

Μορφολογική Διεργασία:

(Morphological Process)
 * Διαλεκτική Περιοχή:

(Dialectal Region)

Τύποι Πραγμάτωσης

(Realizations)

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

Βλέπε Επίσης **(See also)**

Слика 8. Главни формулар леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ – картица реализација је у првом плану

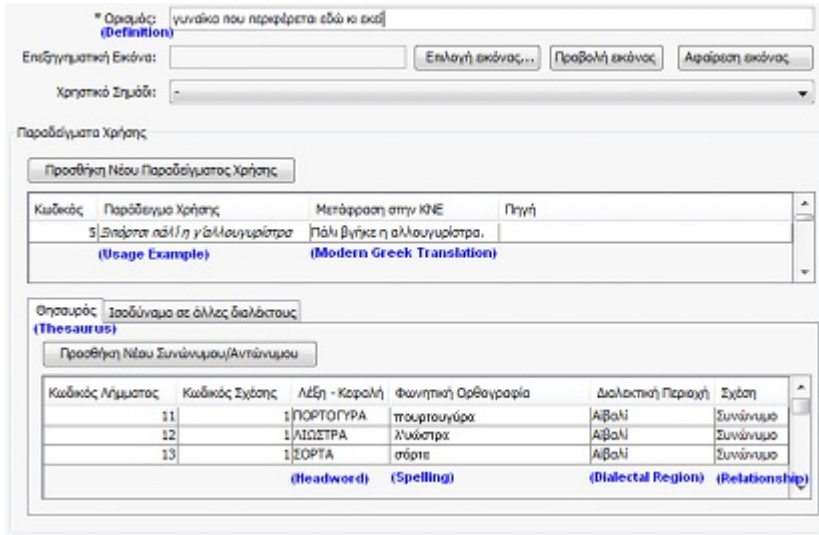
Τύποι Πραγμάτωσης

(Meanings)

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)		

Слика 9. Главни формулар леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ – картица значења

и слика и додатна информацију о употреби. Средишњи део формулара је панел у коме су наведени примери употребе. Доњи део је реализован као панел са две картице. Прва картица се користи за приказ и уређивање релација синонимије и антонимије (тезаурус), док се друга картица користи за руковање еквивалентима у другим дијалектима.



Слика 10. Формулар значења леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ – нека значења са припадајућим тезаурусом

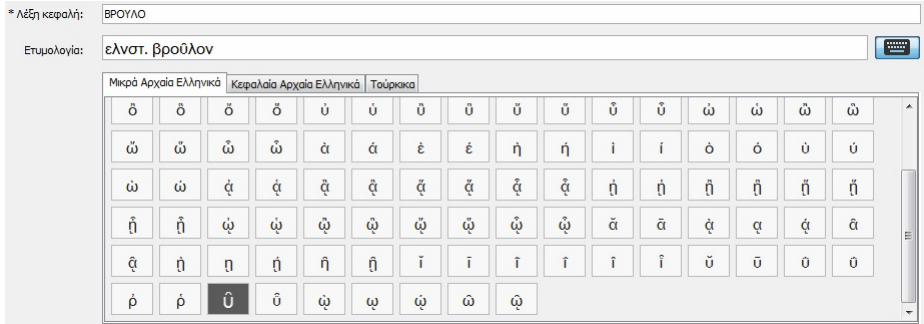
У складу са оним што је речено у одељцима о анализи (захтевима) и нацрту, реализовани систем нуди кориснику следеће скупове карактера за уређивање одговарајућих поља:

- Етмологија
 - грчки политонални;
 - позајмљени карактери из других алфавета уколико су у питању позајмљенице (нпр. карактери из турског алфавета);
- Фонетски тип
 - IPA;

- Спелинг (фонетска ортографија)
 - савремени грчки;
 - акценти;
 - цртице, заграде и апострофи.

7. Завршна оптимизација

1. Као што је добро познато, корисници најчешће изражавају своје мишљење у завршним фазама развојног циклуса система (*прихватање, инсталација, развој*) као и током процеса одржавања. И у случају нашег електронског речника било је тако. У нашем случају корисници су изнели гледиште да би атрибут порекла (етимологија) леме требало да се денотира поштујући изворе и конвенције језика из којег и потиче, то јест који је извор одређене леме. Према томе, у случају вишедијалекатског речника атрибут етимологије може да садржи речи из било ког од језика који су имали утицаја на одређени дијалекат. Да би се овоме удовољило понудили смо решење да се обезбеди (у каснијем развоју) виртуална тастатура за сваки језик који је вршио утицај. У случају три дијалекта који су заступљени у оквиру система AMiGre алфabetи који су остварили утицај су савремени грчки, старогрчки и турски. Слика 11 представља виртуалну тастатуру са три картице за мала слова старогрчког, велика слова старогрчког и за турски. Они су, заједно са савременим грчким за који је на располагању физичка тастатура омогућавали корисницима да унесу жељену етимологију за сваку појединачну лему без икаквих ограничења.
2. Када су у питању атрибут *Phonetic Type* (фонетски тип) и интонација самогласника постоје две могућности. Прва могућност је навођење акцента пре наглашеног слога. Друга могућност је да се употреби наглашена верзија самогласника. Обе опције су семантички еквивалентне. Међутим, прва опција је лакша за употребу и имплементацију у систему јер нема потребе да се подржавају обе верзије IPA симбола које се користе за самогласнике, са и без акцента. Тако је током иницијалне развојне фазе наш систем допуштао само вертикалну акцентуацију пре слога за обележавање интонираног самогласника. Због тога је на слици 8 *Phonetic Type* обележен са “alɟiˈrɪstra”. Након иницијалне развојне



Слика 11. Виртуална тастатура за старогрчки и турски

фазе и током фазе одржавања система указала се потреба да се подрже акцентоване верзије самогласника. Како би се ова потреба задовољила, проширили смо систем тако да обухвати и виртуалне тастатуре за лако уметање IPA симбола, са или без акцената. Сlike 12, 13, 14 и 15 представљају виртуалне тастатуре коришћене за уношење вредности атрибута *Phonetic Type*.



Слика 12. Алфабетски (IPA) симболи

Комбинације IPA симбола (слика 12) са неким од дијакритичких симбола (слика 13) творе акцентоване IPA симболе. Слика 16

* Λέξη κεφαλή: ΛΙΩΣΤΡΑ
 (Headword)
 Ετυμολογία: ΛΙΩΣΤΡΑ
 (Etymology)
 Μορφολογική Διεργασία: Παράγωγο
 (Morphological Process)
 * Διαλεκτική Περιοχή: Αίβαλι
 (Dialectal Region)

Τύποι Πραγμάτωσης Σημασίες
 (Realizations)

Δημιουργία Νέου Τύπου Πραγμάτωσης

Κωδικός	Φωνητικός Τύπος	Αρχείο Ήχου	Προφ	Φωνητική Ορθογραφία	Μικροδιαλεκτική Περιοχή	Λεξική Κατηγορία
4689	λίωστρα			λιώστρα		Ουσιαστικό Θηλυκό
	(Phonetic Type)			(Spelling)		(Lexical Category)

Βάλτε Επίσης (See also)

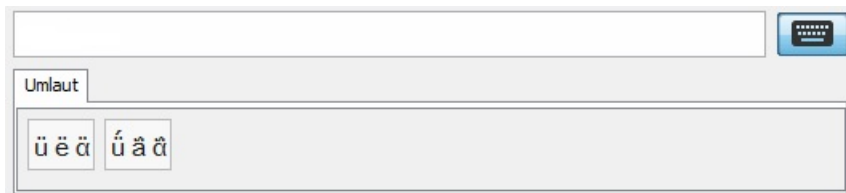
Νέα συσχέτιση με Λήμμα

Κωδικός	Λέξη-Κεφαλή	Ετυμολογία	Μορφολογική Διεργ...	Διαλεκτική Περιοχή	Φωνητικές Ορθογ...	Σημασίες	Παραδείγματα
	(Headword)				(Spellings)	(Meanings)	

Слика 16. Главни формулар леме ΛΙΩΣΤΡΑ са акцентованим IPA симболима у пољу *Phonetic Type*

пртице, заграде и апострофе). Овакав закључак важио је за прву имплементирану верзију система. Од тренутка када је систем ушао у фазу израде, почели смо да примамо веома специфичне захтеве за побољшањима која су се тицала атрибута спеловања. Корисници су указивали да неке речи, будући да потичу из других језика, имају самогласнике који не постоје у стандардном систему спеловања циљног језика, а то је савремени грчки. Будући да дијалекти који су били у фокусу истраживања имају велики број позајмљеница из других језика, ове захтеве корисника нисмо могли да занемаримо. Зато је било неопходно да смислимо како да омогућимо корисницима система да разумеју како се изговарају дијалекатске речи, а да не морају да ишчитавају фонетске IPA симболе. Најједноставнији начин да се то постигне био је да се дозволи уметање графемских представа (слова) које се користе у језику порекла позајмљеница да би се представили они самогласници који не постоје у циљном језику (савременом грчком). Резултат наведеног био је мањи број самогласника са графемским представљањем са прегласом (и, е и

α са прегласом – ü, ë и ʰα). Ова слова била су укључена у једну виртуалну тастатуру предвиђену за уређивање атрибута спеловања (фонетске транскрипције). Слика 17 представља ову виртуелну тастатуру.



Слика 17. Виртуелна тастатура за графеме које представљају додатне самогласнике

4. Као што је описано у ранијем излагању, свака лема може имати различите реализације и сваку од њих карактерише унеколико различита фонетска реализација која зависи од микродијалекатског подручја са којег потиче. Овим смо се руководили приликом осмишљавања и изградње система у којем сваку реализацију карактерише једна микродијалекатска зона која представља уже подручје у оквиру шире дијалекатске регије. Међутим, у фази изградње система корисници су указивали да поједине реализације могу да постоје у више микродијалекатских зона (у оквиру главног дијалекатског подручја). Како би изашли у сусрет оваквим касно изреченим захтевима модификовали смо нацрт шеме података и саму апликацију. Атрибут *Microdialectal region* (микродијалекатска зона) модификован је тако да може да прими више вредности а одговарајућа GUI ставка је измењена како би обухватила листу вредности, при чему домен за сваку појединачну вредност чини скуп микродијалекатских зона које постоје за дијалекат одређене леме. Слика 18 приказује реализације понтске леме ОММА-ТОТЗАТЗИ за коју трећа наведена реализација (трећи ред) има три микродијалекатске зоне: Τραπεζούνα, Χαλδία, Σάντα.

Τύποι Πραγμάτωσης (Realizations)		Σημασίες	
Δημιουργία Νέου Τύπου Πραγμάτωσης			
Κωδικός	Φωνητικός Τύπος	... Φωνητική Ορθογραφία	Μικροδιαλεκτικές Περιοχές
4396	omatodzádʒ	ομματοτζάτʒ	Χαλδία
4397	matodzádʒi(n)	ματοτζάτʒi(v)	Οινόη
4398	matodzádʒ	ματοτζάτʒ	Τραπεζούντα, Χαλδία, Σάντα
4399	matodzæci	ματοτζάκι	Αμισός
4400	matodʒidʒ	ματοτζίτʒ	Κατύωρα
(Phonetic type)		(Spelling)	(Microdialectal regions)

Слика 18. Реализације понтске леме ОММАТОТΖΑΤΖΙ

- Joш једно својство система које није било дефинисано у фази анализе али се појавило током развојне фазе било је садржај листе референци *see also* (види и) а која се односи на циљну леме која може да има више значења или више типова реализације. Решење за које смо се определили било је да представимо дефиницију сваког значења реферисане леме и спеловање сваке реализације реферисане леме у листи ставки „види и“. Доњи део слике 16 представља празну листу референци леме али може се видети да је множина коришћена у насловима релеватних колона – *Spellings* (Спеловања) и *Meanings* (Значења).
- Joш једно својство представљеног система вредно помена је да везе за релације синонимије и антонимије могу да упућују на леме истог дијалекта које могу а и не морају бити присутне у систему. Слика 19 представља значење леме ΛΙΩΣΤΡΑ са припадајућим „тезаурусом“, то јест табелом која садржи везе синонимије и антонимије. На овој слици можемо видети три везе које се односе на леме ΣΟΡΤΑ, ΤΑΚΙΟΥ и ΑΛΛΟΥΓΓΥΡΙΣΤΡΑ од којих је само трећа заступљена у систему. Поређењем слика 10 и 19 може се увидети да смо у табели тезауруса новије верзије (доњи део слике 19) заменили *Spelling* (Спеловање) са *Etymology* (Етимологија). Треба напоменути да је такође додата и нова колона – *Source type* (Тип извора) у табели примера употребе (средишњи део слике 19).

* Ορισμός:

(Definition)

Επεξηγηματική Εικόνα:

Χρηστικό Σημάδι:

Παραδείγματα Χρήσης

Κωδικός	Παράδειγμα Χρήσης	Μετάφραση στην ΚΝΕ	Πηγή	Τύπος Πηγής
5	Ἐπιόρται πάλι ἡ λιώστρα	Ξεπóρτισε πάλι ἡ λιώστρα.		
	(Usage Example)	(Modern Greek Translation)	(Source)	(Source type)

Θησαυρός:

(Thesaurus)

Κωδικός Λήμματος	Κωδικός Σχέσης	Λέξη - Κεφαλή	Ετυμολογία	Διαλεκτική Περιοχή	Σχέση
0	1	ΣΟΡΤΑ		Αἰβαλί	Συνώνυμο
0	1	ΤΑΚΙΟΥ		Αἰβαλί	Συνώνυμο
105	1	ΑΛΛΟΓΥΡΙΣΤΡΑ	αλλουγυρίζου	Αἰβαλί	Συνώνυμο
		(Headword)	(Etymology)	(Dialectal Region)	(Relationship)

Слика 19. Формулар значења леме ΛΙΩΣΤΡΑ – значење и припадајући тезаурус

7. До сада нисмо дали ниједан пример еквивалента у другом дијалекту. Слика 20 је главни формулар кападочанске леме ANΤΕΤ која има једно значење. На слици 21 дат је доњи део формулара значења који приказује картицу *equivalent in other dialects* (еквивалент у другим дијалектима). Као што се може видети на слици 21 једино доступно значење кападочанске леме ANΤΕΤ има еквивалентно значење у ајваличкој лему ANΤΕΤΙ.

8. Током фазе изградње описаног система приметили смо да су корисници одступали од прописа за писање примера употребе лема. Корисници су обично користили *copy/paste* могућност у оквиру оперативног система како би унели карактере које апликација није директно обезбеђивала за атрибут *usage example* (пример употребе). На пример, вредност ζλαγεύω το μήλον унета је као пример употребе за лему ΑΣΛΑΕΥΩ. Ова вредност има грчке карактере који одговарају правилима али такође садржи и један турски карактер (други карактер у ниски). Фраза Αζλαγεύω το μήλον

* Λέξη κεφαλή: ANΤΕΤ
 (Headword)
 Ετυμολογία: τουρκ. adet
 (Etymology)
 Μορφολογική Διεργασία: -
 (Morphological Process)
 * Διαλεκτική Περιοχή: Καπαδοκία
 (Dialectal Region)

Τύποι Πραγμάτωσης Σημασίες
 (Meanings)

Δημιουργία Νέας Σημασίας Επεξεργασία Διαγραφή

Κωδικός	Ορισμός	Χρηστικό Σημάδι	Επεξηγηματική Εικόνα	Πλήθος παραδειγ
2706	έθιμο			0
	(Definition)	(Usage Label)		

Βλέπε Επίσης
 (See also)
 Νέα συσχέτιση με Λήμμα


Слика 20. Главни формулар кападочанске леме ANΤΕΤ

као вредност за пример употребе, заједно са вредношћу “τουρκ. Aşlamak” (то јест, „од турског aşlamak“) у оквиру атрибута *etymology* (етимологија) исте леме може да указује на то како би матерњи говорници дијалекта могли да напишу дијалекатску реч у њиховим документима (Αζλαγεύω). Међутим, ова индикација скривена је унутар једног од значења леме. Претпостављамо да би било боље уколико би се обезбедио други атрибут, назван *indicative writing* (индикативни испис), за сваку реализацију леме. На овај начин поменути „индикативни испис“ био би директно доступан у главном формулару леме, а уз то би се могао издвојити за сваку појединачну реализацију леме (микродијалекатски регион). Ово је једино својство које није имплементирано у завршној фази оптимизације система јер је потреба за њим уочена сувише касно, али сматрамо да би оно било од значаја за наредне вишедијалекатске речнике.

8. Унапређење и проширење система

Шема података (ERD) за подршку унапређеног система приказана је на слици 22.

Θησαυρός: (equivalent in other dialects)

 Επιλογή Ισοδύναμου από άλλη διάλεκτο

Κωδικός Σημ	Ορισμός Σημασίας	Χρηστικό	Κωδικός Λήμμ	Λέξη - Κεφαλή	Ετυμολογία Λήμμ	Μορφολογική Διερ	Διαλεκτική Περιοχή
186	έθιμο, συνήθεια		178	ΑΝΤΕΤΙ	τουρκ. âdet		Αιβαλί
(meaning code)	(definition)		(lemma code)	(headword)	(etymology)	(morphological process)	(dialect)

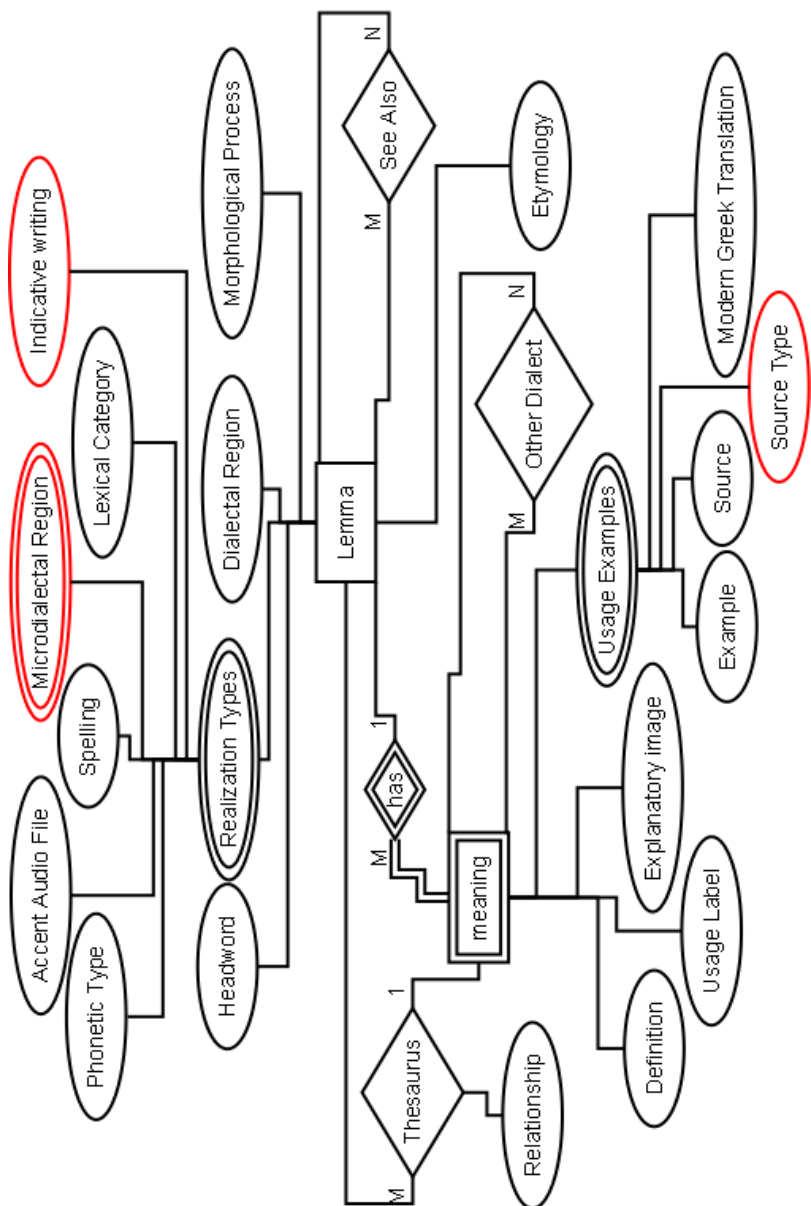
Слика 21. формулар значења леме АΝΤΕΤ – еквиваленти у другим дијалектима приказани су у предњем плану

9. Закључак

Планирана макроструктура система TDGDAM обухвата око 2.500 одредница за сваки од три дијалекта малоазијског грчког (што је укупно око 7.500 одредница). Ове одреднице прикупљене су искључиво из три поменута дијалекта и оне не обухватају елементе речника савременог грчког језика, осим у појединачним случајевима одступања од стандардне употребе. Њихово навођење дато је у складу са алфабетским а не ономасиолошким принципом, док је доступност обезбеђена преко динамичких опција претграживања (Xydopoulos, 2012).

Захвалност

Истраживање је финансијски подржано од стране Европске Уније (Европски социјални фонд, енг. European Social Fund – ESF) и грчких националних фондова кроз програм „Образовање и целоживотно учење“ који спроводи Национални стратешки референтни оквир – Програм за подршку истраживању: „Талис – улагање у друштво знања“. Изражавамо захвалност Ангели Рали, професору лингвистике на Одсеку за филологију Универзитета Патрас, која је координатор целокупног АМиGre пројекта. Дубоку захвалност дугујемо Георгију Ксидопулосу, ванредном професору лингвистике на Одсеку за филологију на Универзитету Патрас, који је поставио лингвистичке захтеве за дијалекатски електронски речник.



Слика 22. Унапређена шема података (ERD)

Литература

- “The Dictionaries of ILSP”. Institute for Language and Speech Processing. Приступљено 22.12.2016. <http://www.xanthi.ilsp.gr/dictionaries/>
- “Homonym”. Wikipedia. Приступљено 22.12.2016. <https://en.wikipedia.org/wiki/Homonym>
- “Online Dictionaries”. Portal for the Greek Language. Приступљено 22.12.2016. http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/index.html
- “Opposite (semantics)”. Wikipedia. Приступљено 22.12.2016. [https://en.wikipedia.org/wiki/Opposite_\(semantics\)](https://en.wikipedia.org/wiki/Opposite_(semantics))
- “Polysemy”. Wikipedia. Приступљено 22.12.2016. <https://en.wikipedia.org/wiki/Polysemy>
- “Synonym”. Wikipedia. Приступљено 22.12.2016. <https://en.wikipedia.org/wiki/Synonym>
- Barbato, Marcello и Alberto Varvaro. “Dialect dictionaries”. *International Journal of Lexicography* Vol. 17 no. 4 (2004): 429–439
- Béjoint, Henri. *Modern lexicography: An introduction*. Oxford: Oxford University Press, 2000
- Burke, Sean. M. “The Design of Online Lexicons”. У *A practical guide to lexicography*, Piet van Sterkenburg, 240–249. Amsterdam: John Benjamins, 2003.
- Durkin, Philip. “Assessing Non-standard Writing in Lexicography”. У *Varieties of English in writing: The written word as linguistic evidence*, Raymond Hickey, 43–60. Amsterdam: John Benjamins, 2010.
- Geeraerts, D. “Principles in Monolingual Lexicography”. У *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie*, Franz Josef Hausmann et al., 287–296. Berlin: Walter de Gruyter, 1989.
- Karanikolas, Nikitas N., Eleni Galiotou, George J. Xydopoulos, Angela Ralli, Konstantinos Athanasakos et. al.. “Structuring a Multimedia Tri-dialectal Dictionary”. У *LNAI 8082: International Conference on Text, Speech and Dialogue*, Ivan Habernal и Vaclav Matousek, 509–518. Berlin Heidelberg: Springer, 2013.
- Keymeulen, J. Van. “Pilot project: A Dictionary of the Dutch Dialects”. Paper presented at the 14th EURALEX International Congress, Fryske Akademy, Ljouwert/Leeuwarden, 6–10 July 2010
- Landau, Sidney. *The art and craft of lexicography (2nd edition)*. Cambridge: Cambridge University Press, 2001

- Markus, Manfred и Reinhard Heuberger. “The architecture of Joseph Wright’s English Dialect Dictionary: preparing the computerized version”. *International Journal of Lexicography* Vol. 20 no. 4 (2007): 355–368
- Papadopoulos, A. *Historical Dictionary of the Pontic Dialect*. Athens: Epitropi Pontiakon Meleton, 1958
- Penhallurik, Rob. “Dialect Dictionaries”. У *The Oxford History of English Lexicography Volume II*, A. P. Cowie, 290–313. Oxford: Oxford University Press, 2009.
- Rys, Kathy и Jacques Van Keymeulen. “Intersystemic correspondence rules and headwords in Dutch dialect lexicography”. *International Journal of Lexicography* Vol. 22 no. 2 (2009): 129–150.
- Themistocleous, C. “Cypriot Greek Lexicography: An Online Lexical Database”. Paper presented at the 15th EURALEX International Congress, University of Oslo, Oslo, 7-11 August 2012.
- Thomason, Sarah G. *Language Contact. An Introduction*. Edinburgh: Edinburgh University Press, 2001.
- Thomason, Sarah G. и Terrence Kaufman. *Language Contact, Creolization, and Genetic Linguistic*. Berkeley: University of California Press, 1988.
- Tsalidis, Christos, Mavina Pantazara, Panagiotis Minos и Elena Mantzari. “NLP Tools for Lexicographic Applications in Modern Greek”. У *eLexicography in the 21st Century: New Challenges, New Applications (Proceedings of eLex2009)*, S. Granger и M. Paquot, 457–462. Louvain-la-Neuve: UCL Presses, 2010.
- Xydopoulos, George J. “Metalexikografikes paratiriseis sta leksika Benardi ke Syrkou (Metalexigraphic comments to Benardi and Syrkou dialectal dictionaries)”. *Patras Working Papers in Linguistics* Vol. 2 no. 1 (2012): 96–113
- Xydopoulos, George. J. и Angela Ralli. “Greek dialects in Asia Minor: Setting lexicographic principles for a tridialectal dictionary”. Paper presented at the 5th MGDLT Conference, Ghent, Belgium, September 2012
- Zgusta, Ladislav. *Manual of lexicography*. The Hague: Mouton, 1971