

Καταγραφή, ανάλυση και ψηφιοποίηση των διαλεκτικών συνθέτων της νέας ελληνικής με τη χρήση της ψηφιακής βάσης δεδομένων ΔΙΑΣΥΝ

Abstract

In this paper, we present a digital database that has been developed in the framework of the research project “Documenting, analyzing and digitizing Greek dialectal compounds”. The aim of the project is to digitize and study the Modern Greek dialectal compounds with the help of an innovative electronic database (*Dicomp*), particularly developed for the purposes of the project. By drawing data from the entire range of the Modern Greek dialectal varieties we aspire to answer a number of important research questions. In particular: (a) which compound structures exist in the Greek dialects; (b) whether compounding in Standard Modern Greek diverges from that of the dialects; (c) which is the dialect with the lowest degree of compounding and why; (d) what kind of evolution has emerged concerning the compounding process from Ancient Greek to Modern Greek; (e) how productive is the morphological process of compounding in Greek as compared to that of derivation. *Dicomp* has been constructed with Drupal CMS (v.7), an open source content management system (CMS) that is actively supported by thousands of programmers and it presents a number of benefits such as high performance, high security against malicious electronic attacks, and automatic backup.

1. Εισαγωγή

Παρά τις εκτενείς μελέτες για τη μορφολογική διαδικασία της σύνθεσης στην κοινή νεοελληνική (βλ., μεταξύ άλλων, Αναστασιάδη Συμεωνίδη 1996· Ράλλη 2007· Ralli 2013α· 2013β), αποτελεί κοινή διαπίστωση ότι ελάχιστη έρευνα έχει διεξαχθεί στην περιοχή των νεοελληνικών διαλέκτων. Η αναγκαιότητα καταγραφής και μελέτης των νεοελληνικών διαλεκτικών συνθέτων είναι σήμερα επιτακτική καθώς: (α) η σύνθεση είναι η κατεξοχήν εκφραστική διαδικασία στην ελληνική, γνωστή ήδη από τον Όμηρο (βλ. Τσερέπη 1902, για μια πλειάδα ομηρικών συνθέτων)· (β) η ελληνική είναι η μόνη γλώσσα στην Ευρώπη που εμφανίζει τόσα σύνθετα και τόσο μεγάλη ποικιλία στη δομή τους (βλ. Ralli 2009· Ralli 2013β για τα [Ρήμα Ρήμα] παρατακτικά σύνθετα)· (γ) η συγκεκριμένη διαδικασία έχει σχεδόν αγνοηθεί από τις απλές περιγραφικές γραμματικές. Χαρακτηριστικό παράδειγμα αποτελεί η γραμματική του Τριανταφυλλίδη (1941), όπου αφιερώνονται ελάχιστες σελίδες στη σύνθεση εν συγκρίσει με την κλίση· (δ) έως σήμερα δεν έχει υπάρξει επίσημη καταγραφή των διαλεκτικών συνθέτων, με εξαίρεση κάποια στοιχεία σε μελέτες θεωρητικής ανάλυσης (ενδεικτικά Αναστασιάδη-Συμεωνίδη 1996· Ράλλη 2007· Ralli 2009· Ralli 2013α· 2013β· Andreou 2014 κτλ.)· (ε) οι διάλεκτοι εμφανίζουν μεγάλο ενδιαφέρον, γιατί περιέχουν στοιχεία που αποτυπώνουν την εξέλιξη της γλώσσας και απουσιάζουν από την κοινή νεοελληνική. Πιστεύουμε ότι η ψηφιοποίηση των νεοελληνικών διαλεκτικών συνθέτων και η διαχείριση της δομής τους μέσω μιας ηλεκτρονικής βάσης δεδομένων θα διευκολύνει την αποτύπωση της δομής τους και θα συμβάλει στην καλύτερη μελέτη των ιδιοτήτων τους.

Υπό αυτό το πρίσμα, η κατασκευή της ηλεκτρονικής βάσης ΔΙΑΣΥΝ στο πλαίσιο του ερευνητικού προγράμματος “Καταγραφή, ανάλυση και ψηφιοποίηση των διαλεκτικών συνθέτων της νέας ελληνικής” αποτελεί, για τα ελληνικά δεδομένα, την πρώτη συστηματική προσπάθεια επιστημονικής προσέγγισης και μελέτης των

συνθέτων που απαντούν στις διαλεκτικές ποικιλίες της γλώσσας. Η ΔΙΑΣΥΝ περιέχει έως τώρα περισσότερα από 13.000 σύνθετα, τα οποία έχουν αναλυθεί πλήρως με βάση τα μορφολογικά, φωνολογικά και σημασιολογικά τους χαρακτηριστικά, ενώ καλύπτει όλες τις διαλεκτικές ποικιλίες της νέας ελληνικής (π.χ. κρητική, κυπριακή, επτανησιακή, ποντιακή, κατωιταλιώτικη, τσακωνική, δωδεκανησιακά ιδιώματα, κυκλαδίτικα ιδιώματα, μικρασιατικά ιδιώματα, βόρεια ιδιώματα κ.ά.). Διαθέτει επίσης μεταγλωσσικές πληροφορίες, όπως τη διάλεκτο προέλευσης, την πηγή από την οποία αντλήθηκαν τα σύνθετα κτλ. Τα δεδομένα που χρησιμοποιήθηκαν για την καταχώρηση των λημμάτων προέρχονται από τις προφορικές και γραπτές πηγές του Εργαστηρίου Νεοελληνικών Διαλέκτων (ΕΝΕΔΙ) του Πανεπιστημίου Πατρών.¹

Η ανάπτυξη της ψηφιακής βάσης ΔΙΑΣΥΝ στοχεύει σε μια διεξοδική μελέτη του φαινομένου της σύνθεσης, όπως απαντά στη νεοελληνική διαλεκτική ποικιλία, προσβλέποντας στην απάντηση μιας σειράς θεωρητικών ερωτημάτων, ενώ τα δεδομένα θα προσεγγιστούν τόσο δια-διαλεκτικά όσο και σε σχέση με την κοινή νεοελληνική. Συγκεκριμένα:

(α) Εξετάζοντας τα συστατικά των συνθέτων και τις δομές που απαντούν στις νεοελληνικές διαλέκτους, είναι δυνατόν να διερευνηθεί ο βαθμός διαφοροποίησής τους, τόσο σε σχέση με την κοινή νεοελληνική όσο και σε σχέση με άλλες ευρωπαϊκές γλώσσες.

(β) Σύμφωνα με τη Ralli (2009· 2013α· 2013β) η ελληνική είναι η μόνη ευρωπαϊκή γλώσσα που παρουσιάζει τόσο μεγάλο αριθμό δομών σύνθεσης. Ως εκ τούτου, είναι ενδιαφέρον να εξεταστεί εάν αυτός ο πλούτος οφείλεται σε ενδογενείς ή εξωγενείς παράγοντες (π.χ. γλωσσική επαφή).

(γ) Δεδομένου ότι η σύνθεση είναι εξαιρετικά παραγωγική διαδικασία στην ελληνική γλώσσα θα διερευνηθεί αν υπάρχουν διάλεκτοι που παρουσιάζουν χαμηλό βαθμό σύνθεσης και αν αυτό οφείλεται σε κάποιο βαθμό στη γεωγραφική απομόνωση ή σε διακριτή ιστορική εξέλιξη.

(δ) Συγκρίνοντας τη σύνθεση των νεοελληνικών διαλέκτων με αυτή της αρχαίας ελληνικής, θα αποδειχθεί ποιες διάλεκτοι διατηρούν στοιχεία παλαιότερων φάσεων της γλώσσας, σε σχέση με τη συγκεκριμένη διαδικασία.

(ε) Τέλος, συγκρίνοντας την παραγωγικότητα μεταξύ της παραγωγής και της σύνθεσης, θα αποκαλυφθεί ποια διαδικασία παρουσιάζει μεγαλύτερο βαθμό παραγωγικότητας καθώς και οι λόγοι για τους οποίους αυτό συμβαίνει.

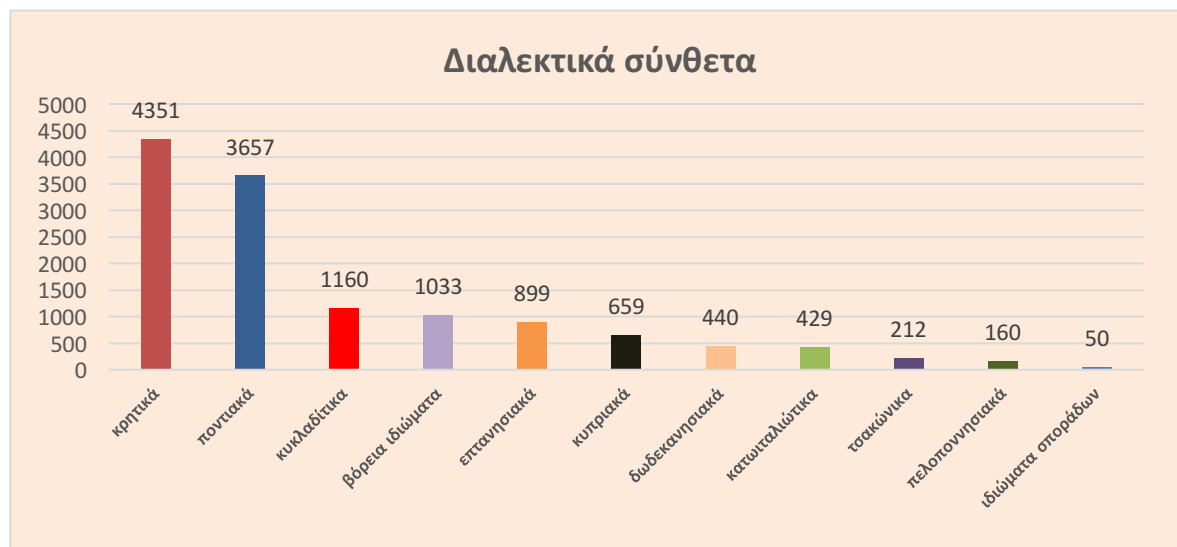
Το παρόν άρθρο οργανώνεται ως εξής: στην Ενότητα 2 παρουσιάζεται διεξοδικά η ψηφιακή βάση συνθέτων ΔΙΑΣΥΝ και τα πεδία γλωσσολογικής ανάλυσης που την απαρτίζουν. Στην Ενότητα 3 καταγράφονται τα βασικότερα τεχνικά χαρακτηριστικά της βάσης, ενώ στην Ενότητα 4 γίνεται αναφορά στη συνεισφορά που αναμένεται να έχει η βάση τόσο στη θεωρητική όσο και στην εφαρμοσμένη γνώση. Το άρθρο κλείνει με τον επίλογο και τις βιβλιογραφικές αναφορές.

2. Πεδία γλωσσολογικής ανάλυσης της ΔΙΑΣΥΝ

Η ψηφιακή βάση ΔΙΑΣΥΝ περιέχει 50 πεδία, τα οποία εντάσσονται σε τρεις άξονες γλωσσολογικής ανάλυσης, τη μορφολογία (πρωτίστως, εφόσον η σύνθεση είναι μορφολογική διαδικασία), τη φωνολογία και τη σημασιολογία. Τα πεδία διακρίνονται

¹ <<http://www.lmgd.philology.upatras.gr/index.php/el/>>

σε τέσσερις ομάδες: (α) βασικές πληροφορίες· (β) πληροφορίες λεξημάτων· (γ) πληροφορίες παραγωγικής προσφύματοποίησης· (δ) μεταπληροφορίες.



Γράφημα 1: Αριθμός συνθέτων ανά διάλεκτο ή διαλεκτική ομάδα

▽ Βασικές πληροφορίες	
02. Διάλεκτος/Ιδιώμα:	Κρητικά Ανατολικοκρητικό ιδίωμα (Ηράκλειο, Λασιθί, Ανατολικές επαρχίες Ρεθύμνης)
03. Σημασία:	εργαλείο με το οποίο κοπανίζουν το λεπιδόχωμα που έχει στρωθεί στη στέγη.
04. Σχέση συστατικών:	Σχέση εξάρτησης
05. Προφορά:	/ðo.ma.to.ko.pa.'ni.ða/
06. Θέση του τόνου:	Παραλήγουσα
07. Μορφολογική Διαδικασία:	Σύνθεση/Παραγωγή/Κλίση
08. Γραμματική Κατηγορία:	Ουσιαστικό
09. Γένος:	Θηλυκό
10. Περιορισμός Γυνού Θέματος:	Όχι
11. Δείκτης Σύνθεσης:	Ναι [-o-]
12. Δομή Συστατικών:	[Ουσιαστικό] + [Ουσιαστικό]
13. Μορφολογική Δομή:	[[Θέμα] + [Λέξη]]
14. Παράγωγο Συνθέτου:	Όχι
45. Ενδο-/Εξωκεντρικό Σύνθετο:	Ενδοκεντρικό
15. Δεσμευμένο Θέμα:	Όχι
16. Θέση Κεφαλής:	Δεξιό Συστατικό
▽ Λέξημα Α	
17. Λέξημα Α:	δῶμα (~ δωματ)
18. Καταγωγή:	Ελληνική
19. Γραμματική Κατηγορία:	Ουσιαστικό
20. Γένος:	Ουδέτερο
22. Σημασία:	η ταρότσα.
▽ Λέξημα Β	
23. Λέξημα Β:	κοπανίδα < κόπανο
24. Καταγωγή:	Ελληνική
25. Γραμματική Κατηγορία:	Ουσιαστικό
26. Γένος:	Θηλυκό
28. Σημασία:	1. εργαλείο με το οποίο χτυπούν τις ρωγμές του δώματος για να μην τις διαπερνά η βροχή. 2. ο γυρίνος.
▽ Πληροφορίες προσφύματοποίησης	
37. Παραγωγικό Επίθημα Α:	-ίδα
38. Καταγωγή:	Ελληνική
50. Συνδυασμός:	Λέξημα Β
▽ Μετα-πληροφορίες	
40. Πηγή:	Περί του γλωσσικού ιδιώματος της Κρήτης

Εικόνα 1: Το σύνθετο δωματοκοπανίδα

Πριν ξεκινήσει η συζήτηση αναφορικά με τα πεδία της βάσης, και προκειμένου να υπάρξει καλύτερη κατανόηση από τον αναγνώστη, είναι καλό να δοθούν με τη μορφή εικόνων (screenshots) τα εξής: (α) ο έως τώρα αριθμός συνθέτων ανά διάλεκτο (Γράφημα 1), τα οποία προέρχονται από ένα σύνολο 13.050 συνθέτων· (β) ένα πλήρως αναλυμένο σύνθετο (Εικόνα 1) από την κρητική διάλεκτο, έτσι όπως εμφανίζεται στη ψηφιακή βάση και το οποίο έχει αντληθεί από τον Πάγκαλο (1955).

2.1 Βασικές πληροφορίες

Όπως φαίνεται στον Πίνακα 1, στην πρώτη ομάδα των βασικών πληροφοριών εμφανίζονται 17 πεδία, τα οποία αφορούν τα γενικά χαρακτηριστικά του συνθέτου. Η κάθε βασική πληροφορία μπορεί να διαθέτει περισσότερες της μίας τιμές, τη σωστή καταχώριση της οποίας (ή και των οποίων) καλείται να επιλέξει ο λημματογράφος.

Σε πρώτη φάση, ο λημματογράφος εισάγει τον βασικό τύπο του λήμματος (ονομαστική ενικού αριθμού για τα ονόματα και α' πρόσωπο ενικού αριθμού ενεστώτα για τα ρήματα) και επιλέγει μεταξύ των τιμών μιας λίστας τη διάλεκτο ή το ιδίωμα στο οποίο ανήκει το σύνθετο (πεδία 1 και 2). Στη συνέχεια, εισάγει τη σημασία του συνθέτου (πεδίο 3). Αν το σύνθετο παρουσιάζει περισσότερες της μίας σημασίες (κυριολεκτικές, μεταφορικές κτλ.), δηλώνονται όλες.

Πληροφορία	Επιλογές
1. Σύνθετο	Ελεύθερο κείμενο
2. Διάλεκτος/ιδίωμα	Λίστα επιλογών
3. Σημασία	Ελεύθερο κείμενο
4. Σχέση συστατικών	Σχέση εξάρτησης Σχέση υπόταξης Σχέση υπόταξης και παράταξης
5. Προφορά	Ελεύθερο κείμενο
6. Θέση του τόνου	Λήγουσα Παραλήγουσα Προπαραλήγουσα Τέταρτη από το τέλος συλλαβή Διπλός τόνος
7. Μορφολογική διαδικασία	Σύνθεση Σύνθεση-κλίση Σύνθεση-παραγωγή Σύνθεση-παραγωγή-κλίση
8. Γραμματική κατηγορία	Ουσιαστικό Επίθετο Ρήμα Επίρρημα Αντωνυμία
9. Γένος	Αρσενικό Θηλυκό Ουδέτερο Διγενές Τριγενές

Πληροφορία	Επιλογές
10. Περιορισμός γυμνού θέματος	Ναι Όχι Παραβίαση Μερική παραβίαση
11. Δείκτης σύνθεσης	Ναι Όχι
12. Δομή συστατικών	Ελεύθερο κείμενο
13. Μορφολογική δομή	Λίστα επιλογών
14. Παράγωγο από σύνθετο	Ελεύθερο κείμενο
15. Ενδο/εξωκεντρικό σύνθετο	Ενδοκεντρικό Εξωκεντρικό
16. Δεσμευμένο θέμα	Ναι Όχι
17. Θέση κεφαλής	Δεξιό συστατικό Αριστερό συστατικό Όλα τα συστατικά Παραγωγικό επίθημα

Πίνακας 1: Πληροφορίες ομάδας 1

Η σχέση των συστατικών μεταξύ τους (πεδίο 4) είναι σημαντική για τη δομική και σημασιολογική ανάλυση του σύνθετου σχηματισμού και εμφανίζει τρεις τιμές: (α) σχέση εξάρτησης, π.χ. αγγειοπλυνίσκω ‘πλένω πιάτα’ (Κύπρος)· (β) σχέση παράταξης, π.χ. ασκελοριζώνω ‘αναπτύσσομαι και βγάζω ρίζες’ (Κρήτη)· (γ) σχέση εξάρτησης και παράταξης, π.χ. συνωραντρόνο ‘αντρόγυνο (άντρας και γυναίκα) που παντρεύτηκε πρόσφατα’ (Κρήτη). Ακολούθως, εισάγεται η προφορά του συνθέτου με βάση τα σύμβολα του Διεθνούς Φωνητικού Αλφαβήτου, π.χ. /bi.zi.ˈlɔ.spu.rus/ ‘ο σπόρος του μπιζελιού’ (Μακεδονία) και η θέση του τόνου (πεδία 5 και 6).

Καθώς η ελληνική γλώσσα έχει πλούσια μορφολογία, κρίθηκε απαραίτητο να διαχωριστούν οι πιθανοί συνδυασμοί μορφολογικών διαδικασιών (πεδίο 7). Οι διαδικασίες αυτές είναι οι ακόλουθες: (α) σύνθεση, π.χ. αδα-πέραν ‘εδώ πέρα’ (Πόντος)· (β) σύνθεση-κλίση, π.χ. κακ-ο-νι(ός)² ‘άσχημος νέος’ (Ιθάκη)· (γ) σύνθεση-παραγωγή, π.χ. διπλ-ο-πόδ-ι ‘με διπλωμένα πόδια’ (Κρήτη)· (δ) σύνθεση-παραγωγή-κλίση, π.χ. αγγελ-ο-μουσου-άτ(ος) ‘ο έχων αγγελικά χαρακτηριστικά’ (Κάρπαθος). Στη συνέχεια, ακολουθούν πληροφορίες αναφορικά με τη γραμματική κατηγορία του συνθέτου και το γένος (πεδία 8 και 9). Ο περιορισμός του γυμνού θέματος (Bare-stem Constraint) έχει αναπτυχθεί θεωρητικά μέσα από το άρθρο των Ralli & Karasimos (2009) και αναφέρεται στην τάση απαλοιφής των παραγωγικών επιθημάτων που βρίσκονται στο αριστερό συστατικό των συνθέτων (πεδίο 10). Ο εν λόγω μορφολογικός περιορισμός, ανάλογα με την εφαρμογή ή μη εφαρμογή του, εμφανίζει τέσσερις τιμές: (α) ναι, όταν ενεργοποιείται και εφαρμόζεται σε ένα σύνθετο, π.χ. αποσπεροβόσκω < αποσπέρ-ας + βόσκω ‘βόσκω το βράδυ’ (Κρήτη)· (β) όχι, όταν δεν υπάρχει λόγος ενεργοποίησής του, π.χ. ηλιοκουτέ < ήλιε + κουτέ ‘ηλιοκαμένος’ (Τσακωνικά Πελοποννήσου)· (γ) παραβίαση, όταν το παραγωγικό επίθημα στο αριστερό συστατικό παραμένει ορατό, π.χ. βοσκαροράβδι < βοσκ-άρι + ραβδί ‘το ραβδί του βοσκού (Κάλυμνος)’· (δ) μερική παραβίαση, όταν το αριστερό συστατικό

² Σε παρένθεση δίνεται το κλιτικό επίθημα.

έχει δύο παραγωγικά επιθήματα και μόλις το ένα έχει απαλειφθεί από τη δομή, π.χ. σφιχτομανταλώνω < σφιχ-τ-ά + μανταλώνω ‘κλειδώνω σφιχτά’ (Κρήτη).

Στη συνέχεια, δηλώνεται ο δείκτης σύνθεσης (πεδίο 11), η δομή των συστατικών με τη μορφή της γραμματικής τους κατηγορίας, π.χ. Ουσιαστικό + Ουσιαστικό, Ρήμα + Ουσιαστικό κ.ο.κ. (πεδίο 12), και η μορφολογική δομή του συνθέτου με τη μορφή των μορφολογικών κατηγοριών, π.χ. Θέμα + Θέμα, Θέμα + Λέξη κ.ο.κ. (πεδίο 13). Επιπλέον, η συγκεκριμένη ομάδα περιλαμβάνει τις εξής πληροφορίες: (α) αν το σύνθετο παράγεται από άλλο ήδη υπαρκτό σύνθετο (πεδίο 14), π.χ. αγγελόκρουσμα < αγγελοκρούζω ‘τρομάρα’ (Κεφαλονιά)· (β) αν το σύνθετο είναι ενδοκεντρικό ή εξωκεντρικό (πεδίο 15)· (γ) αν υπάρχει δεσμευμένο θέμα (πεδίο 16). Σημαντική πληροφορία που είναι καταχωρισμένη στη βάση είναι και η θέση της κεφαλής (πεδίο 17), δηλαδή του συστατικού που είναι υπεύθυνο για τη γραμματική κατηγορία και τη βασική σημασία του συνθέτου (Ralli 2013α, 104–125). Γι’ αυτή την πληροφορία, ο λημματογράφος επιλέγει τη σωστή τιμή μεταξύ τεσσάρων διαφορετικών τιμών: (α) δεξιό συστατικό, π.χ. αλατσ-ο-κόπανος ‘πέτρινος βώλος για το κοπάνισμα του αλατιού’ (Νάξος)· (β) αριστερό συστατικό, π.χ. πορ-ό-λαγκος ‘η έξοδος του λαγκαδιού’ (Κρήτη)· (γ) όλα τα συστατικά, π.χ. κλουθ-ου-γυρίζου ‘γυρίζω γύρω από κάτι’ (Λέσβος)· (δ) παραγωγικό επίθημα, π.χ. κουτσοκέρ-η-ς ‘αυτός που έχει σπασμένο κέρατο’ (Καλαβρία).

2.2 Πληροφορίες λεξημάτων

Η δεύτερη ομάδα περιέχει 15 πεδία που καλύπτουν τις πληροφορίες των λεξημάτων, δηλαδή των θεμάτων ή λέξεων που αποτελούν το σύνθετο. Οι πληροφορίες και οι τιμές τους δίνονται στον Πίνακα 2:

Πληροφορία	Επιλογές
18. Λέξημα Α	Ελεύθερο κείμενο
19. Καταγωγή	Λίστα επιλογών
20. Γραμματική κατηγορία	Ουσιαστικό Επίθετο Ρήμα Επίρρημα Αντωνυμία
21. Γένος	Αρσενικό Θηλυκό Ουδέτερο Διγενές Τριγενές
22. Σημασία	Ελεύθερο κείμενο
23. Λέξημα Β	Ελεύθερο κείμενο
24. Καταγωγή	Λίστα επιλογών
25. Γραμματική κατηγορία	Ουσιαστικό Επίθετο Ρήμα Επίρρημα Αντωνυμία

Πληροφορία	Επιλογές
26. Γένος	Αρσενικό Θηλυκό Ουδέτερο Διγενές Τριγενές
27. Σημασία	Ελεύθερο κείμενο
28. Λέξημα Γ	Ελεύθερο κείμενο
29. Καταγωγή	Λίστα επιλογών
30. Γραμματική κατηγορία	Ουσιαστικό Επίθετο Ρήμα Επίρρημα Αντωνυμία
31. Γένος	Αρσενικό Θηλυκό Ουδέτερο Διγενές Τριγενές
32. Σημασία	Ελεύθερο κείμενο

Πίνακας 2: Πληροφορίες ομάδας 2

Στις βασικές πληροφορίες των λεξημάτων (πεδία 18, 23, 28) περιλαμβάνεται η καταγωγή (πεδία 19, 24, 29), η γραμματική κατηγορία (πεδία 20, 25, 30), το γένος (πεδία 21, 26, 31) και η σημασία (πεδία 22, 27, 32). Σε περίπτωση που το λέξημα είναι δάνεια λέξη, δηλώνεται και η λέξη από την οποία κατάγεται, π.χ. ατζί < *incik*. Επιπλέον, στα πεδία που αφορούν τη σημασία του συνθέτου δηλώνονται όλες οι πιθανές σημασίες του λεξήματος.

2.3 Πληροφορίες παραγωγικής προσφυματοποίησης

Δεδομένου ότι σε έναν μορφολογικά πολύπλοκο σχηματισμό η παραγωγή μπορεί να συνεμφανίζεται με τη σύνθεση, υπάρχει ανάγκη η βάση να περιλαμβάνει πληροφορίες για τα παραγωγικά προσφύματα που απαντούν στο λήμμα.

Πληροφορία	Επιλογές
33. Πρόθημα	Ελεύθερο κείμενο
34. Καταγωγή	Λίστα επιλογών
35. Συνδυασμός	Λέξημα Α Λέξημα Β Λέξημα Γ Σύνθετο θέμα Α Σύνθετο θέμα Β Σύνθετο θέμα Γ Σύνθετο θέμα Δ
36. Παραγωγικό επίθημα Α	Ελεύθερο κείμενο
37. Καταγωγή	Λίστα επιλογών

Πληροφορία	Επιλογές
38. Συνδυασμός	Λέξημα Α Λέξημα Β Λέξημα Γ Σύνθετο θέμα Α Σύνθετο θέμα Β Σύνθετο θέμα Γ Σύνθετο θέμα Δ
39. Παραγωγικό επίθημα Β	Ελεύθερο κείμενο
40. Καταγωγή	Λίστα επιλογών
41. Συνδυασμός	Λέξημα Α Λέξημα Β Λέξημα Γ Σύνθετο θέμα Α Σύνθετο θέμα Β Σύνθετο θέμα Γ Σύνθετο θέμα Δ
42. Παραγωγικό επίθημα Γ	Ελεύθερο κείμενο
43. Καταγωγή	Λίστα επιλογών
44. Συνδυασμός	Λέξημα Α Λέξημα Β Λέξημα Γ Σύνθετο θέμα Α Σύνθετο θέμα Β Σύνθετο θέμα Γ Σύνθετο θέμα Δ
45. Παραγωγικό επίθημα Δ	Ελεύθερο κείμενο
46. Καταγωγή	Λίστα επιλογών
47. Συνδυασμός	Λέξημα Α Λέξημα Β Λέξημα Γ Σύνθετο θέμα Α Σύνθετο θέμα Β Σύνθετο θέμα Γ Σύνθετο θέμα Δ

Πίνακας 3: Πληροφορίες ομάδας 3

Οι πληροφορίες της προσφυματοποίησης παραγωγής δεν περιορίζονται μόνο στη δήλωση των παραγωγικών προθημάτων ή επιθημάτων (πεδία 33, 36, 39, 42, 45) που υπάρχουν στο σύνθετο, αλλά παρέχονται πληροφορίες τόσο για την καταγωγή τους όσο και για τη συνδυαστικότητά τους σε κάθε σύνθετο χωριστά. Χαρακτηριστικά, στο λ. αργαστηροδιάστρα ‘γυναίκα που γνωρίζει καλά την υφαντική τέχνη’, από την κρητική διάλεκτο, τα παραγωγικά επιθήματα -τηρι και -τρα δεν συνδέονται με τον ίδιο τρόπο στη σύνθετη δομή. Συγκεκριμένα, το επίθημα -τηρι προσαρτάται στο πρώτο λέξημα του σύνθετου τύπου, δηλαδή στο θέμα αορίστου του ρήματος αργάζομαι (αργασ-), ενώ το -τρα συνδέεται με το δεύτερο λέξημα της δομής, δηλαδή με το θέμα αορίστου του ρήματος διάζομαι (διασ-).

2.4 Μεταπληροφορίες

Η τελευταία ομάδα περιλαμβάνει 3 πεδία. Αυτά είναι τα σχόλια (πεδίο 48) όπου ο χρήστης μπορεί να κάνει διάφορες παρατηρήσεις για το λήμμα που επεξεργάστηκε, η πηγή (προφορική ή γραπτή)³ από την οποία αντλήθηκε το σύνθετο (πεδίο 49) και παραδείγματα χρήσης που περιέχουν και διασαφηνίζουν το υπό ανάλυση σύνθετο (πεδίο 50). Σχηματικά:

Μεταπληροφορία	Επιλογές
48. Σχόλια	Ελεύθερο κείμενο
49. Πηγή	Λίστα επιλογών
50. Παράδειγμα χρήσης	Ελεύθερο κείμενο

Πίνακας 4: Μεταπληροφορίες ομάδας 4

3. Τεχνικά χαρακτηριστικά ΔΙΑΣΥΝ

Η δημιουργία της βάσης δεδομένων βασίστηκε στην πλατφόρμα διαχείρισης υλικού γενικού σκοπού (CMS)⁴ ανοιχτού κώδικα Drupal 7, η οποία χρησιμοποιείται παγκοσμίως από χιλιάδες προγραμματιστές (Tomlinson & VanDyke 2010). Η επιλογή του συγκεκριμένου ηλεκτρονικού εργαλείου δεν ήταν τυχαία, αφού προσφέρει μια πλειάδα πλεονεκτημάτων, όπως, μεταξύ άλλων, τα ακόλουθα: (α) εύκολη διαχείριση από τους χρήστες· (β) διαβάθμιση χρηστών· (γ) αυτόματα αντίγραφα ασφαλείας (backup)· (δ) πρόσβαση μέσω web.

Επιπλέον, η πλατφόρμα Drupal 7 έχει αποδειχθεί ότι είναι μια αξιόπιστη λύση για την κατασκευή ψηφιακών βάσεων δεδομένων που στοχεύουν στην καταγραφή, ψηφιοποίηση και ανάλυση γλωσσικού υλικού, αφού έχει εφαρμοστεί με επιτυχία σε παρεμφερή ερευνητικά προγράμματα του Εργαστηρίου Νεοελληνικών Διαλέκτων, όπως το Morilan Project⁵ που υλοποιήθηκε κατά το διάστημα 2012–2015 καθώς και το Immigrec,⁶ το οποίο ολοκληρώθηκε το 2018.

3.1 Εύκολη διαχείριση από τους χρήστες

Η βάση είναι φιλική στον χρήστη (user friendly) και η σχεδίασή της τον απαλλάσσει από τα περιττά στοιχεία. Η αρχική σελίδα, όπως φαίνεται στην Εικόνα 2, προσφέρει στον χρήστη πολλές δυνατότητες, όπως: (α) εύκολη αναζήτηση λημμάτων στη βάση συγκεκριμένων κριτηρίων· (β) πρόσβαση και φόρτωση των λημμάτων που έχουν καταχωριστεί· (γ) πρόσβαση στο ιστορικό καταχωρίσεων.

3.2 Διαβάθμιση χρηστών

Οι χρήστες, ανάλογα με τα δικαιώματα που έχουν στη βάση, χωρίζονται σε τέσσερις κατηγορίες. Οι ρόλοι τους είναι οι εξής, αρχίζοντας από τον χρήστη που έχει τα λιγότερα δικαιώματα:

(α) Καταχωριστής (linguist): Καταχωρίζει νέα λήμματα και βιβλιογραφία. Επιπλέον, έχει τη δυνατότητα να επεξεργάζεται τα λήμματα που έχει καταχωρίσει ο ίδιος.

³ Αξίζει να σημειωθεί ότι στη βιβλιογραφία που έχει επισημειωθεί στη ΔΙΑΣΥΝ κωδικοποιούνται 19 μεταπληροφορίες, όπως, μεταξύ άλλων, το όνομα του συγγραφέα και η χρονολογία έκδοσης.

⁴ Content Management System.

⁵ <<http://www.morilan.upatras.gr>>

⁶ <<https://immigrec.com>>

ΔΙΑΣΥΝ Διαλεκτικά Σύνθετα

Σύνθετα

Πηγές

Χρήστες

Ιστορικό Αλλαγών

Αν

Ο λογαριασμός μου

Προσθήκη συνθέτου

Τίτλος

Μορφολογική Διαδικασία

Γραμματική Κατηγορία

Γένος

Κατηγοριακή Δομή

Γενική Δομή

Καταχωρητής

Διάλεκτος

Εισαγωγή σε Λεξικό

Κωδ.	Τίτλος	Διάλεκτος	Μορφολογική Διαδικασία	Γραμματική Κατηγορία	Γένος	Κατηγοριακή Δομή	Γενική Δομή
13589	Βαρταλαιώ	Βορειοελλαδικά ιδιώματα, Λέσβος	Σύνθεση/ Παραγωγή/ Κλίση	Ρήμα		[ρῆμα] + [ρῆμα]	[[θέμα] + [Λέξη]]
12880	κωλοφωτιά	Πελοποννησιακά, Ηλεία	Σύνθεση/κλίση	Ουσιαστικό	Θηλυκό		[[θέμα] + [Λέξη]]
11842	ἄλλουρόκατος	Δωδεκανησιακά Ιδιώματα, Κάρπαθος	Σύνθεση/ Παραγωγή/ Κλίση	Επίθετο	Διγενές	[Επίθετο] + [Ουσιαστικό]	[[θέμα] + [θέμα]]
11843	ἄλλουρόπετρα	Δωδεκανησιακά Ιδιώματα, Κάρπαθος	Σύνθεση/κλίση	Ουσιαστικό	Θηλυκό	[Επίθετο] + [Ουσιαστικό]	[[θέμα] + [θέμα]]
10998	ἡστροσκιασμένος	Κυκλαδītika ιδιώματα, Μύκονος	Σύνθεση/ Παραγωγή/ Κλίση	Επίθετο	Τριγενές	[Επὶρρημα] + [ρῆμα]	[[θέμα] + [θέμα]]
11011	ἡσοχεϊά	Κυκλαδītika ιδιώματα, Μύκονος	Σύνθεση/κλίση	Ουσιαστικό	Ουδέτερο	[Ουσιαστικό] + [Ουσιαστικό]	[[θέμα] + [θέμα]]
11012	ἡσοκούκοδο	Κυκλαδītika ιδιώματα, Μύκονος	Σύνθεση/κλίση	Ουσιαστικό	Ουδέτερο	[Ουσιαστικό] + [Ουσιαστικό]	[[θέμα] + [Λέξη]]

Επαναφορά

Εικόνα 2: Αρχική σελίδα της βάσης ΔΙΑΣΥΝ

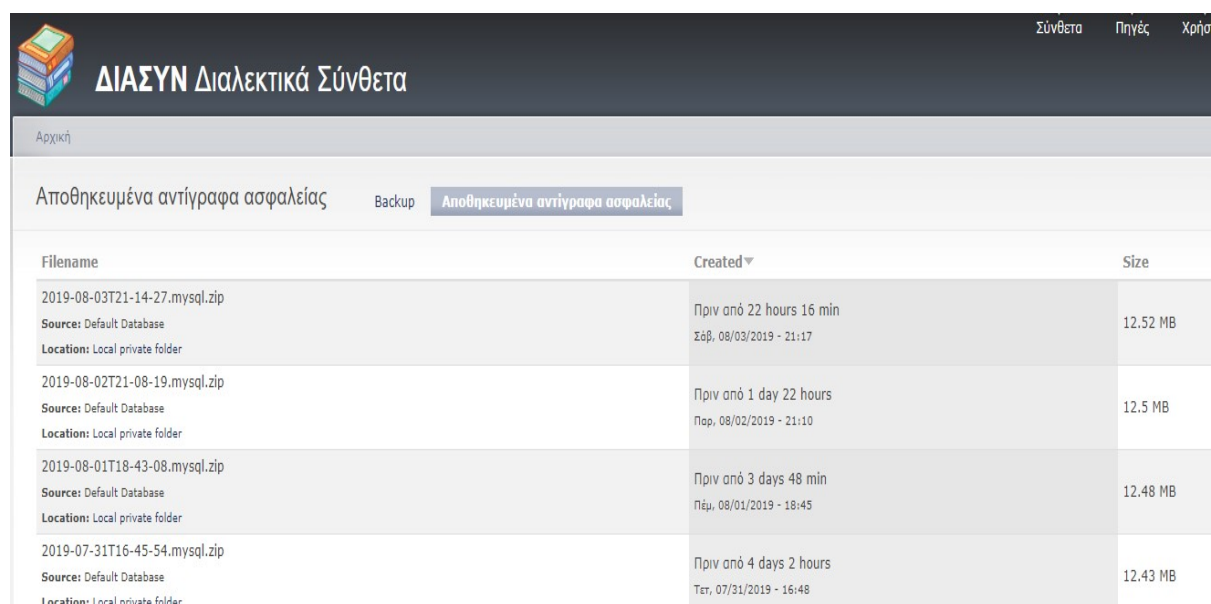
(β) *Διαχειριστής καταχωριστών (manager)*: Έχει τα δικαιώματα του καταχωριστή, αλλά μπορεί να διαγράψει λήμματα και βιβλιογραφία κατά βούληση και να επεξεργαστεί λήμματα άλλων χρηστών. Τέλος, έχει πρόσβαση στα στατιστικά της βάσης.

(γ) *Υποδιαχειριστής (sub-administrator)*: Έχει τα ίδια δικαιώματα με τον διαχειριστή καταχωριστών. Επιπρόσθετα, έχει τη δυνατότητα να προσθέσει ή να διαγράψει χρήστες και να εισαγάγει, τροποποιήσει ή ακόμα και να διαγράψει πεδία της βάσης. Τέλος, έχει πρόσβαση στα αντίγραφα ασφαλείας της βάσης.

(δ) *Γενικός διαχειριστής (administrator)*: Έχει όλα τα παραπάνω δικαιώματα. Είναι υπεύθυνος για τη λήψη αποφάσεων και την πορεία των εργασιών της βάσης.

3.3 Αυτόματα αντίγραφα ασφαλείας (backup)

Η βάση δημιουργεί με αυτόματο τρόπο και σε τακτά χρονικά διαστήματα αντίγραφα ασφαλείας, προκειμένου να διασφαλιστεί ότι το περιεχόμενό της παραμένει ασφαλές σε περίπτωση καταστροφής του υλικού και των δεδομένων. Υπεύθυνοι για την αποθήκευση των αντιγράφων ασφαλείας τοπικά σε υπολογιστικό μηχάνημα του Εργαστηρίου Νεοελληνικών Διαλέκτων είναι ο Υποδιαχειριστής και ο Γενικός διαχειριστής.



The screenshot shows the 'ΔΙΑΣΥΝ Διαλεκτικά Σύνθετα' web application. The top navigation bar includes 'Σύνθετα', 'Πηγές', and 'Χρήση'. Below the header, there's a 'Backup' section with a tab 'Αποθηκευμένα αντίγραφα ασφαλείας'. The main content area displays a table of backups with columns 'Filename', 'Created', and 'Size'.

Filename	Created	Size
2019-08-03T21-14-27.mysql.zip Source: Default Database Location: Local private folder	Πριν από 22 hours 16 min Σάβ, 08/03/2019 - 21:17	12.52 MB
2019-08-02T21-08-19.mysql.zip Source: Default Database Location: Local private folder	Πριν από 1 day 22 hours Παρ, 08/02/2019 - 21:10	12.5 MB
2019-08-01T18-43-08.mysql.zip Source: Default Database Location: Local private folder	Πριν από 3 days 48 min Πέμ, 08/01/2019 - 18:45	12.48 MB
2019-07-31T16-45-54.mysql.zip Source: Default Database Location: Local private folder	Πριν από 4 days 2 hours Τετ, 07/31/2019 - 16:48	12.43 MB

Εικόνα 3: Σύστημα αυτόματης δημιουργίας αντιγράφων ασφαλείας

3.4 Πρόσβαση μέσω web

Ο χρήστης μπορεί να συνδεθεί στη βάση μέσω συγκεκριμένου URL, με μοναδική προϋπόθεση να γνωρίζει τα προσωπικά του στοιχεία πρόσβασης, ενώ η σύνδεση μπορεί να γίνει από οποιοδήποτε σημείο αρκεί ο χρήστης να έχει σύνδεση στο διαδίκτυο. Η βάση είναι συμβατή με Windows, Linux, Mac και Android και ένα από τα προτερήματά της είναι η παράλληλη επεξεργασία ή καταχώριση δεδομένων από πολλαπλούς χρήστες.

4. Συνεισφορά στη θεωρητική και εφαρμοσμένη γνώση

Η υλοποίηση του ερευνητικού προγράμματος “Καταγραφή, ανάλυση και ψηφιοποίηση των διαλεκτικών συνθέτων της νέας ελληνικής” αναμένεται να έχει σημαντικό αντίκτυπο τόσο στη θεωρητική όσο και στην εφαρμοσμένη γνώση.

Ο όγκος πληροφοριών που θα προκύψει από τη λημματογράφηση των διαλεκτικών συνθέτων θα βοηθήσει να απαντηθούν μια πλειάδα ερωτημάτων σχετικά με τη διαδικασία της σύνθεσης (βλ. Ενότητα 1). Επιπλέον, για πρώτη φορά η επιστημονική κοινότητα θα έχει στη διάθεσή της έναν τεράστιο όγκο δεδομένων από όλες τις διαλεκτικές ποικιλίες της νέας ελληνικής, ο οποίος θα εμφανίζει ομοιογένεια στην ανάλυση, θα λαμβάνει υπόψη τις τελευταίες εξελίξεις στον χώρο της ελληνικής και διεθνούς βιβλιογραφίας και θα προσφέρεται για γλωσσολογική μελέτη.

Η ΔΙΑΣΥΝ έχει αποφασιστεί να είναι ανοιχτής πρόσβασης (open access) και, ως εκ τούτου, πιστεύουμε ότι θα αποτελέσει σημαντικό έργο αναφοράς σε μελέτες θεωρητικού, διαλεκτολογικού και ιστορικού χαρακτήρα. Παράλληλα, θα συμβάλει στην προσπάθεια διάσωσης μέρους της γλωσσικής μας κληρονομιάς, η οποία σήμερα αντιμετωπίζει άμεσο κίνδυνο εξαφάνισης.

Τέλος, αξίζει να σημειωθεί ότι το υλικό της βάσης προβλέπεται να χρησιμοποιηθεί για την έκδοση του πρώτου για τα ελληνικά δεδομένα λεξικού διαλεκτικών συνθέτων.

5. Συμπεράσματα

Η ηλεκτρονική βάση ΔΙΑΣΥΝ αποτελεί μια σημαντική πηγή δεδομένων που περιέχει ομοιογενές και αυστηρά ταξινομημένο γλωσσικό υλικό (περισσότερα από 13.000 σύνθετα και περισσότερα από 350.000 αναλυμένα πεδία ανάλυσης) από όλες τις γλωσσικές ποικιλίες της νέας ελληνικής. Με την ολοκλήρωση του προγράμματος υλοποίησής της, φιλοδοξούμε να προσφέρουμε στην επιστημονική κοινότητα έναν τεράστιο όγκο γλωσσικών δεδομένων που θα ενθαρρύνει τη μελέτη της σύνθεσης στις νεοελληνικές διαλέκτους τόσο σε συγχρονικό όσο και σε διαχρονικό επίπεδο.

Ο στόχος είναι τριπλός: (α) να παρουσιαστεί ένα μεθοδικά ταξινομημένο γλωσσικό υλικό, αναλυμένο σύμφωνα με τις τελευταίες επιστημονικές εξελίξεις στον χώρο της γλωσσολογίας, το οποίο θα διευκολύνει την εξέταση της σύνθεσης στις διαλέκτους της νέας ελληνικής· (β) να αποτελέσει πρότυπο για ανάλογες μελλοντικές προσπάθειες καταγραφής και ανάλυσης των νεοελληνικών διαλέκτων· (γ) να υπάρξει συμβολή στην καταγραφή και διαφύλαξη της γλωσσικής μας κληρονομιάς.

Η ΔΙΑΣΥΝ έχει σχεδιαστεί με τη χρήση νέων αναπτυσσόμενων τεχνολογιών (Drupal), έχοντας ως πρότυπο αντίστοιχες βάσεις δεδομένων που έχουν κατασκευαστεί με στόχο την καταγραφή και ανάλυση της γλωσσικής ποικιλίας (βλ. σχετικά Ράλλη, Παπαζαχαρίου & Καρασίμος 2010· Τσιμπούρης, Ανδρέου & Κουτσούκος 2015· Tsimplouris κ.ά. 2019).

Μετά την ολοκλήρωση του προγράμματος, η ΔΙΑΣΥΝ θα είναι ελεύθερα προσβάσιμη σε όσους ενδιαφέρονται να ασχοληθούν με τη μελέτη της σύνθεσης των νεοελληνικών διαλέκτων, παρέχοντας μια πλειάδα γλωσσικών και μεταγλωσσικών πληροφοριών, όπως η γραμματική κατηγορία, η μορφολογική δομή, η σχέση των συστατικών μεταξύ τους, η θέση της κεφαλής, η πηγή άντλησης του υλικού κ.ά.

Βιβλιογραφία

- Αναστασιάδη-Συμεωνίδη, Α. 1996. “Η νεοελληνική σύνθεση”, στο Γ. Κατσιμαλή & Φ. Καβουκόπουλος (επιμ.), *Ζητήματα νεοελληνικής γλώσσας: διδακτική προσέγγιση*. Ρέθυμνο: Πανεπιστήμιο Κρήτης, 97–120.
- Andreou, M. 2014. *Headedness in Word Formation and Lexical Semantics: Evidence from Italiot and Cypriot*. Διδ. διατρ. Πανεπιστήμιο Πατρών.
- Πάγκαλος, Γ. 1955. *Περί του γλωσσικού ιδιώματος της Κρήτης*. Αθήνα.
- Ράλλη, Α. 2007. *Η σύνθεση λέξεων: διαγλωσσική, μορφολογική προσέγγιση*. Αθήνα: Πατάκης.
- Ράλλη, Α., Δ. Παπαζαχαρίου & Α. Καρασίμος. 2010. “Εργαστήριο Νεοελληνικών Διαλέκτων και η βάση δεδομένων GREED”, στο Α. Ράλλη, Β. Joseph, Μ. Janse & Α. Karasimos (επιμ.), *Proceedings of the 4th International Conference of Modern Greek Dialects and Linguistic Theory*. Πάτρα: Πανεπιστήμιο Πατρών, 7–14.
- Ralli, A. 2009. “Modern Greek V V dvandva compounds: A linguistic innovation in the history of the Indo-European languages”, *Word Structure* 2(1), 48–67.
- . 2013α. *Compounding in Modern Greek*. Dordrecht: Springer.
- . 2013β. “Compounding and its locus of realization: Evidence from Greek and Turkish”, *Word Structure* 6, 181–200.
- Ralli, A. & Α. Karasimos. 2009. “The role of constraints in compound formation: The case of bare-stem constraint”, *Lingue e Linguaggio* 8, 53–74.
- Tomlinson, T. & J. Van Dyke. 2010. *Pro Drupal 7 Development*. USA: Apress.
- Τριανταφυλλίδης, Μ. 1941. *Νεοελληνική γραμματική (της δημοτικής)*. Αθήνα.
- Τσερέπης, Γ. 1902. *Τα σύνθετα της ελληνικής γλώσσας*. Αθήνα: Σακελλαρίου.
- Τσιμπούρης, Χ., Μ. Ανδρέου & Ν. Κουτσούκος. 2015. “Σχεδιασμός και ανάπτυξη μίας μορφολογικής βάσης δεδομένων: η βάση MORILAN”, *Patras Working Papers in Linguistics* 4, 43–57.
- Tsimpouris, C., S. Bompolas, C. Papanagiotou, G. Chairetakis, V. Alexelli & A. Ralli. 2019. “A data archiving and browsing multimodal system for the language of Greek immigrants in Canada”, *Patras Working Papers in Linguistics* 5, 193–208.

Λέξεις-κλειδιά: σύνθεση, νεοελληνικές διαλέκτοι, ψηφιοποίηση, βάση δεδομένων, γλωσσολογική επισημείωση, εργαλεία ανοιχτού κώδικα