

# Σχεδιασμός και ανάπτυξη μίας μορφολογικής βάσης δεδομένων: Η βάση MORILAN

---

Χάρης Τσιμπούρης, Μάριος Ανδρέου, Νίκος Κουτσούκος  
*Πανεπιστήμιο Πατρών*

## Περίληψη

Στο παρόν άρθρο παρουσιάζεται η βάση δεδομένων MORILAN, η οποία αναπτύχθηκε στο πλαίσιο του προγράμματος «Μορφολογία και γλωσσική επαφή: Ελληνικές διάλεκτοι σε επαφή με την Τουρκική και την Ιταλική». Η συγκεκριμένη βάση σχεδιάστηκε και υλοποιήθηκε με πρώτο στόχο να αποθηκευτεί και να ταξινομηθεί το διαλεκτικό υλικό που συλλέχθηκε κατά την εκπόνηση του ερευνητικού προγράμματος, αλλά και με απώτερο στόχο να δημιουργηθεί μία σύγχρονη βάση δεδομένων για τη μελέτη και εκπόνηση άλλων προγραμμάτων που σχετίζονται με τις νεοελληνικές διαλέκτους και τη μορφολογία.

**Λέξεις-κλειδιά:** βάσεις δεδομένων, εργαλεία ανοικτού κώδικα, Νεοελληνικές διάλεκτοι, μορφολογική επισημείωση

## 1. Εισαγωγή

Η μορφολογία της Νέας Ελληνικής και των διαλέκτων έχει γίνει αντικείμενο εκτενούς έρευνας τα τελευταία χρόνια (βλ. Ράλλη 2005, Ralli 2003· 2006· 2009 και τις αναφορές που υπάρχουν μέσα σε αυτά τα έργα). Ωστόσο, ένα βασικό κενό που εντοπίζεται στη σχετική έρευνα είναι η έλλειψη έργων υποδομής πάνω στα οποία θα μπορεί να βασιστεί η ανάλυση των γλωσσολογικών φαινομένων. Για παράδειγμα, ιδιαίτερα φτωχή είναι η ύπαρξη βάσεων δεδομένων οι οποίες περιέχουν οργανωμένα γλωσσικά δεδομένα (βλ. σχετικά Karanikolas, Galiotou & Ralli 2014, Καρασίμος κ.ά. υπό έκδοση).

Στο πλαίσιο του ερευνητικού προγράμματος «Μορφολογία και γλωσσική επαφή: Ελληνικές διάλεκτοι σε επαφή με την Τουρκική και την Ιταλική» προσπαθήσαμε να καλύψουμε αυτό το κενό δημιουργώντας μία βάση δεδομένων η οποία έχει διττό στόχο: (α) να καλύψει τις ανάγκες του συγκεκριμένου προγράμματος, και (β) να αποτελέσει ένα σημαντικό έργο υποδομής για την καταγραφή της ποικιλίας και τη μελέτη των διαλέκτων στο σύνολό τους.

Ως προς το σχεδιασμό, η συγκεκριμένη βάση δεδομένων έχει βασιστεί σε αντίστοιχες βάσεις δεδομένων οι οποίες έχουν ως στόχο την καταγραφή είτε ενδογλωσσικής είτε εξωγλωσσικής ποικιλίας (βλ. σχετικά Ράλλη, Παπαζαχαρίου & Καρασίμος 2010), ενώ ως προς το κομμάτι της υλοποίησης, έχει βασιστεί στα διεθνή πρότυπα ανάπτυξης βάσεων δεδομένων. Στις επόμενες ενότητες θα επικεντρωθούμε κυρίως στον σχεδιασμό της βάσης δεδομένων MORILAN (ενότητα 2) και την υλοποίησή της (ενότητα 3). Στην τελευταία ενότητα, συνοψίζουμε τα χαρακτηριστικά της βάσης, και προτείνουμε τη μελλοντική εξέλιξή της μέσα από την επέκτασή της.

## 2. Ο σχεδιασμός της βάσης δεδομένων MORILAN

Η βάση MORILAN<sup>1</sup> σχεδιάστηκε από τους ερευνητές με σκοπό την εισαγωγή και ανάλυση διαλεκτικών δεδομένων, τα οποία παρέχουν πληροφορίες σχετικά με τα βασικά ερωτήματα του προγράμματος, δίνοντας έμφαση στις διαλέκτους Grecanico (Griko και Bovese), Επτανησιακά, Ποντιακά, Καπαδοκικά και Αϊβαλιώτικα.<sup>2</sup> Τα βασικά ερωτήματα είναι τα εξής (για ανάλυση των συγκεκριμένων στόχων, δεξ Ράλλη 2015):

- i. Εξέταση της βάσης σχηματισμού λέξεων.
- ii. Διερεύνηση της σχέσης μορφής και σημασίας στις λεξικές δομές σε συστήματα που έχουν έρθει σε γλωσσική επαφή.
- iii. Μελέτη της σχέσης δανεισμού και λεξικών πληροφοριών.
- iv. Εντοπισμός των πιθανών διαφορών μεταξύ των δανείων που ανήκουν σε διαφορετικές γραμματικές κατηγορίες.
- v. Μελέτη της δανεισιμότητας των λειτουργικών στοιχείων.
- vi. Ανίχνευση της -πιθανής- διαφοράς του βαθμού δανεισιμότητας, όπως προκύπτει από τη σύγκριση μεταξύ κλιτικών και παραγωγικών μορφημάτων.

Δεδομένου ότι τα ερωτήματα αυτά καλύπτουν ένα ευρύ φάσμα της μορφολογικής θεωρίας και οι απαντήσεις στα ερωτήματα αυτά απαιτούν μία εις βάθος ανάλυση των δεδομένων (γραπτών και/ή προφορικών), οι ερευνητές κατέληξαν στο συμπέρασμα ότι η βάση έπρεπε να σχεδιαστεί με τέτοιο τρόπο, ώστε να περιλαμβάνει όλες τις πληροφορίες οι οποίες θα βοηθήσουν στην κατά το δυνατόν πληρέστερη ανάλυση των δεδομένων, καθώς και στην ευκολότερη ανάκτηση των σχετικών πληροφοριών. Στην συνέχεια δίνεται μία λεπτομερής περιγραφή των μεταδεδομένων τα οποία κωδικοποιούνται στη βάση MORILAN.

### 2.1 Διαδικασία επιλογής χαρακτηριστικών (μεταδεδομένων)

Η βάση σχεδιάστηκε με σκοπό να προσφέρει μία διεξοδική ανάλυση της μορφολογικής δομής τόσο της Νέας Ελληνικής, όσο και των υπό μελέτη διαλέκτων. Ως εκ τούτου, η βάση καλύπτει όλες τις μορφολογικές κατηγορίες της Νέας Ελληνικής και των Νεοελληνικών διαλέκτων και κάθε κλιτός, παράγωγος και σύνθετος σχηματισμός *αναλύεται* (decomposed) πλήρως στα συστατικά του. Για παράδειγμα, για ένα παράγωγο σχηματισμό δίνονται λεπτομερείς πληροφορίες για τη βάση από την οποία προέρχεται, καθώς και για το παραγωγικό και το κλιτικό επίθημα το οποίο φέρει. Αυτές οι πληροφορίες κωδικοποιούνται σε χαρακτηριστικά (μεταδεδομένα) τα οποία παίρνουν συγκεκριμένες τιμές.

Επίσης, είναι πολύ σημαντικό να αναφερθεί ότι, καθώς η Ελληνική είναι μία γλώσσα με πλούσια μορφολογία, έγινε κατηγοριοποίηση των πληροφοριών μορφολογικής ανάλυσης και διαχωρίστηκαν οι πιθανοί συνδυασμοί μορφολογικών διαδικασιών. Οι διαδικασίες αυτές είναι οι εξής:

<sup>1</sup> Η βάση δεδομένων, κατά την περίοδο συγγραφής του παρόντος κειμένου περιέχει 9.970 λήμματα από 13 καταχωρητές.

<sup>2</sup> Για παρόμοια θέματα σε σχέση με τις διαλέκτους της Μικράς Ασίας, βλ. Κολιοπούλου, Μαρκόπουλο & Παντελίδη (2015) και Galiotou κ.ά. (2014).

- i. *Κλίση-Ακλισία*: περιλαμβάνει λήμματα τα οποία είτε είναι μονομορφηματικά (χωρίς κλιτικό επίθημα) είτε αποτελούνται από μία και μόνο απλή βάση και ένα κλιτικό επίθημα.
- ii. *Παραγωγή-Κλίση*: περιλαμβάνει λήμματα τα οποία αποτελούνται είτε από μία βάση ένα παραγωγικό επίθημα και ένα κλιτικό είτε από μία βάση και ένα παραγωγικό επίθημα (π.χ. επιρρήματα).
- iii. *Σύνθεση-Παραγωγή*: περιλαμβάνει λήμματα τα οποία αποτελούνται από μία σύνθετη βάση και ένα παραγωγικό επίθημα.
- iv. *Σύνθεση-Παραγωγή-Κλίση*: περιλαμβάνει λήμματα τα οποία εμφανίζουν όλες τις μορφολογικές διαδικασίες.

Πρέπει να τονίσουμε ότι οι μορφολογικές διαδικασίες εμφανίζονται με την μορφή *εξαρτήσεων* (dependencies). Για παράδειγμα, εάν στο μεταδεδομένο *μορφολογική διαδικασία*, ο χρήστης επιλέξει την τιμή *Κλίση-Ακλισία*, τότε στη βάση εμφανίζονται μόνο οι πληροφορίες οι οποίες αφορούν στη συγκεκριμένη διαδικασία. Έτσι, δεν εμφανίζονται πληροφορίες οι οποίες αφορούν στην Παραγωγή ή τη Σύνθεση. Με τον τρόπο αυτό η εισαγωγή δεδομένων γίνεται γρηγορότερα.

Σε αυτό το σημείο, πρέπει να γίνει σαφές ότι δόθηκε ιδιαίτερη βαρύτητα στον φιλικό σχεδιασμό της *διεπαφής* (user interface). Προβλέφθηκαν όλες οι πιθανές εκδοχές/επιλογές που μπορεί να ανακύψουν από τη μορφολογική μελέτη των διαλέκτων και οι τιμές αρκετών μεταδεδομένων δίνονται υπό τη μορφή λιστών και έτσι ελαχιστοποιείται κατά το δυνατόν η πληκτρολόγηση πληροφοριών από τους ερευνητές. Η εισαγωγή ελεύθερου κειμένου γίνεται μόνο όπου αυτό είναι άκρως απαραίτητο (π.χ. σημασία λήμματος). Επίσης, αποφεύγονται λάθη κατά την εισαγωγή των λημμάτων τα οποία οφείλονται σε διαφορετικό τρόπο κωδικοποίησης της ίδιας πληροφορίας (όπως για παράδειγμα την κωδικοποίηση της τιμής γένους ΑΡΣΕΝΙΚΟ ως «Αρσενικό» ή «Αρσ.», κτλ.). Τέτοια λάθη μπορούν να δημιουργήσουν σημαντικά προβλήματα κατά τη διαδικασία ανάκτησης της συγκεκριμένης πληροφορίας ή κατά την δημιουργία στατιστικών αποτελεσμάτων.

Για την επιτυχή υλοποίηση των παραπάνω, δεδομένου ότι δεν υπήρχε μία επισταμένη και ολοκληρωμένη έρευνα για τις συγκεκριμένες διαλέκτους, η διαδικασία προϋπέθετε συνοπτικά τα εξής:

- i. Μελέτη των έργων αναφοράς (λεξικά και γραμματικές) ως προς τη μορφολογική δομή των συγκεκριμένων διαλέκτων.

Η μελέτη αυτή είχε ως στόχο την αναγνώριση όλων των διαδικασιών (κλίση, παραγωγή, σύνθεση) οι οποίες εμφανίζονται στις υπό εξέταση διαλέκτους, καθώς και την αναγνώριση των στοιχείων που εμπλέκονται σε αυτές (π.χ. λεξική και μορφολογική κατηγορία συστατικών).

- ii. Μελέτη των μορφολογικών κατηγοριών των διαλέκτων (ειδικότερα των προθημάτων και των επιθημάτων).

Δεδομένου ότι οι συγκεκριμένες κατηγορίες ανήκουν σε σχετικά κλειστές τάξεις στοιχείων, η αναγνώριση και εισαγωγή τους σε λίστες τιμών κρίθηκε χρήσιμη για την

ταχύτερη εισαγωγή των δεδομένων στη βάση και για την ελαχιστοποίηση πιθανών λαθών (βλ. και πιο πάνω).

Είναι βασικό να τονιστεί ότι έπρεπε να γίνει και πρωτογενής έρευνα για τις διαλέκτους, δεδομένου ότι οι γραμματικές περιγραφές είτε είναι παλιές, είτε είναι βασισμένες σε παραδοσιακά κριτήρια περιγραφής. Ως εκ τούτου, η έρευνα αυτή εμπλούτισε τις λίστες προθημάτων και επιθημάτων που εισήχθησαν στη βάση και οδήγησε στη δημιουργία ενός ενιαίου καταλόγου ο οποίος περιλαμβάνει όλες τις μορφολογικές κατηγορίες (προθήματα, επιθήματα) για όλες τις υπό εξέταση διαλέκτους.

## 2.2 Λίστα μεταδεδομένων

Κρίθηκε απαραίτητο τα λήμματα να μην ορίζονται μονοσήμαντα από αυτό καθαυτό το λήμμα, αλλά με βάση έναν μοναδικό αύξοντα κωδικό, καθώς έτσι ήταν εφικτή η καταχώριση λημμάτων με την ίδια γραφή, αλλά με διαφορετική χρήση ή μορφολογικό χαρακτήρα. Στην συνέχεια, γίνεται αναλυτική περιγραφή των πληροφοριών που μπορούν να καταχωρηθούν για κάθε λήμμα ξεχωριστά.

Οι πληροφορίες χωρίζονται σε 6 βασικές ομάδες και συγκεκριμένα στα εξής: (α) βασικές πληροφορίες, (β) καταγωγή, (γ) σύνθεση, (δ) παραγωγή, (ε) κλίση και (στ) μετα-πληροφορίες. Ο διαχωρισμός ήταν απαραίτητος για τους εξής δύο βασικούς λόγους: (α) κάποιες κατηγορίες ενεργοποιούνται και απενεργοποιούνται αυτόματα ανάλογα με την εκάστοτε μελετώμενη μορφολογική διαδικασία, όπως αναφέρθηκε στην προηγούμενη ενότητα, και (β) για να βελτιωθεί η ευκολία χρήσης του συστήματος από τον μελετητή.

### 2.2.1 Βασικές πληροφορίες

Η ομάδα αυτή περιλαμβάνει τις βασικές πληροφορίες για ένα λήμμα. Σχηματικά περιλαμβάνονται στον παρακάτω πίνακα:

Μεταδεδομένο/ Χαρακτηριστικό <sup>3</sup>	Τιμές
1. Λήμμα	Ελεύθερο κείμενο
2. Σημασία	Ελεύθερο κείμενο
3. Μορφ. διαδικασία	Κλίση-Ακλισία, Παραγωγή-Κλίση, Σύνθεση- Κλίση, Σύνθεση-Παραγωγή-Κλίση
4. Γραμ. κατηγορία	Επίθετο, Ουσιαστικό, Ρήμα, Επίρρημα, Αντωνυμία, Γερούνδιο, Απαρεμφατικός τύπος, Έκφραση, Φράση, Πρόθεση, Σύνδεσμος, Μόριο, Αριθμητικό
5. Γένος	Ουδέτερο, Τριγενές, Αρσενικό, Θηλυκό, Χωρίς γένος, Διγενές
6. Κλιτική τάξη	KT1 ρήματα, KT2α ρήματα, KT2β ρήματα, KT1, KT2, KT3, KT4, KT5, KT6, KT7, KT8, KT9, KT10, Άκλιτο, Μεικτή κλιτική τάξη-Επίθετα, Ετερόκλιση

Πίνακας 1: Μεταδεδομένα ομάδας 1

<sup>3</sup> Τα μεταδεδομένα παρουσιάζονται με αύξοντα αριθμό που συνεχίζει σε όλους τους πίνακες, όπως είναι δηλωμένα μέσα στη βάση δεδομένων.

Αρχικά, ο ληματογράφος εισάγει το λήμμα στον *βασικό τύπο* (citation form), ήτοι α' πρόσωπο ενικού ενεστώτα για ρήματα και ονομαστική ενικού αριθμού για ονόματα, και ακολούθως την πλήρη σημασία του (πεδία 1 και 2). Εάν υπάρχουν περισσότερες της μίας σημασίες, τότε δηλώνονται όλες. Η επιλογή της μορφολογικής διαδικασίας επιτρέπει στον ληματογράφο να περάσει στη βάση τις πληροφορίες οι οποίες είναι σχετικές με τη διαδικασία αυτή (πεδίο 3). Ακολούθως εισάγονται πληροφορίες γραμματικής κατηγορίας, γένους και κλιτικής τάξης (πεδία 4, 5 και 6).

## 2.2.2 Καταγωγή

Η δεύτερη ομάδα μεταδεδομένων καλύπτει πληροφορίες σχετικά με την καταγωγή του λήμματος στο σύνολό του. Εάν μία λέξη είναι δάνεια στο σύνολό της, τότε δίνονται πληροφορίες για την πρωτότυπη δάνεια λέξη και το γένος της (όπου ισχύει).

<i>Μεταδεδομένο/Χαρακτηριστικό</i>	<i>Τιμές</i>
7. Καταγωγή λήμματος	Τουρκική, Ιταλική, Ρομανικές διάλεκτοι, Ελληνική, Γαλλική, Λατινική, Άλλη
8. Πρωτότυπη δάνεια	Ελεύθερο κείμενο
9. Γένος δάνειας λέξης	(βλ. μεταδεδομένο 5)

Πίνακας 2: Μεταδεδομένα ομάδας 2

## 2.2.3 Σύνθεση

Δεδομένου ότι τα σύνθετα περιλαμβάνουν δύο τουλάχιστον βάσεις και στο εσωτερικό των συνθέτων εμφανίζονται και δάνειες λέξεις, η βάση MORILAN σχεδιάστηκε με τρόπο ο οποίος επιτρέπει την λεπτομερή καταχώριση πληροφοριών για κάθε συστατικό ξεχωριστά, όπως ακριβώς γίνεται και για το λήμμα στο σύνολό του στην ομάδα 1:

<i>Μεταδεδομένο/Χαρακτηριστικό</i>	<i>Τιμές</i>
10. Βάση 1	Ελεύθερο κείμενο πολλών γραμμών
11. Σημασία	Ελεύθερο κείμενο πολλών γραμμών
12. Γραμματική κατηγορία	(βλ. μεταδεδομένο 4)
13. Γένος	(βλ. μεταδεδομένο 5)
14. Κλιτική τάξη	(βλ. μεταδεδομένο 6)
15. Καταγωγή βάσης	(βλ. μεταδεδομένο 7)
16. Πρωτότυπη δάνεια λέξη	Ελεύθερο κείμενο
17. Γραμματική κατηγορία δάνειας λέξης	(βλ. μεταδεδομένο 4)
18. Γένος δάνειας λέξης	(βλ. μεταδεδομένο 5)
19. Βάση 2	Ελεύθερο κείμενο
20. Σημασία	Ελεύθερο κείμενο πολλών γραμμών
21. Γραμματική κατηγορία	(βλ. μεταδεδομένο 4)
22. Γένος	(βλ. μεταδεδομένο 5)

23. Κλιτική τάξη	(βλ. μεταδεδομένο 6)
24. Καταγωγή βάσης	(βλ. μεταδεδομένο 7)
25. Πρωτότυπη δάνεια λέξη	Ελεύθερο κείμενο
26. Γραμματική κατηγορία δάνειας λέξης	(βλ. μεταδεδομένο 4)
27. Γένος δάνειας λέξης	(βλ. μεταδεδομένο 5)
28. Παρατακτικό σύνθετο	Ναι/Όχι

Πίνακας 3: Μεταδεδομένα ομάδας 3

Μία σημαντική πληροφορία η οποία δηλώνεται στη βάση είναι το κατά πόσο το σύνθετο είναι παρατακτικό ή όχι (μεταδεδομένο 28), χωρίς ωστόσο να δηλώνονται περαιτέρω σημασιολογικά χαρακτηριστικά του συνθέτου. Η συγκεκριμένη πληροφορία είναι σημαντική τόσο για τη δομική όσο και για τη σημασιολογική ανάλυση του συνθέτου.

## 2.2.4 Παραγωγή

Η ομάδα αυτή καλύπτει τη διαδικασία της παραγωγής και ως εκ τούτου περιλαμβάνει πληροφορίες τόσο για τη βάση όσο και για τα προσφύματα (πρόθηματα και επιθήματα) τα οποία εμφανίζονται στο λήμμα. Αξίζει να αναφερθεί ότι οι πληροφορίες αυτές είναι λεπτομερείς, καθώς προβλέπεται η εισαγωγή πληροφοριών σχετικά με το αν η βάση ή κάποιο πρόσφυμα είναι ξενικής καταγωγής.

<i>Μεταδεδομένο/Χαρακτηριστικό</i>	<i>Τιμές</i>
29. Βάση	Ελεύθερο κείμενο
30. Σημασία βάσης	Ελεύθερο κείμενο πολλών γραμμών
31. Γραμματική κατηγορία βάσης	(βλ. μεταδεδομένο 4)
32. Γένος βάσης	(βλ. μεταδεδομένο 5)
33. Κλιτική τάξη βάσης	(βλ. μεταδεδομένο 6)
34. Καταγωγή βάσης	(βλ. μεταδεδομένο 7)
35. Πρωτότυπη δάνεια λέξη	Ελεύθερο κείμενο
36. Γραμ. κατηγορία δάνειας λέξης	(βλ. μεταδεδομένο 4)
37. Γένος δάνειας λέξης	(βλ. μεταδεδομένο 5)
38. Παραγωγικό πρόθημα 1	Λίστα προσφυσμάτων
39. Καταγωγή προθήματος	(βλ. μεταδεδομένο 7)
40. Παραγωγικό πρόθημα 2	Λίστα προσφυσμάτων
41. Καταγωγή προθήματος	(βλ. μεταδεδομένο 7)
42. Παραγωγικό πρόθημα 3	Λίστα προσφυσμάτων
43. Καταγωγή προθήματος	(βλ. μεταδεδομένο 7)
44. Παραγωγικό επίθημα 1	Λίστα προσφυσμάτων
45. Καταγωγή επιθήματος	(βλ. μεταδεδομένο 7)
46. Πρωτότυπο παραγωγικό επίθημα	-

47. Παραγωγικό επίθημα 2	Λίστα προσφυμάτων
48. Καταγωγή επιθήματος	(βλ. μεταδεδομένο 7)
49. Πρωτότυπο παραγωγικό επίθημα	-
50. Παραγωγικό επίθημα 3	(επιλογή από λίστα προσφυμάτων)
51. Καταγωγή επιθήματος	(βλ. μεταδεδομένο 7)
52. Πρωτότυπο παραγωγικό επίθημα	-

Πίνακας 4: Μεταδεδομένα ομάδας 4

### 2.2.5 Κλίση

Η ομάδα 5 περιλαμβάνει πληροφορίες για τη διαδικασία της κλίσης με έμφαση σε συγκεκριμένους ρηματικούς και ονοματικούς τύπους, όπως φαίνεται από τον αμέσως επόμενο πίνακα:

Μεταδεδομένο/Χαρακτηριστικό	Τιμές
53. Αόριστος ενεργ.-παθ.	Ελεύθερο κείμενο
54. Περιφρ. κλιτός σχηματισμός	Ελεύθερο κείμενο
55. Τύπος γενικής εν.-πληθ.	Ελεύθερο κείμενο
56. Κλιτικό επίθημα	Επιλογή μεταξύ των: Ο, ΟΣ, Ω, Ι, ΗΣ, ΟΜΑΙ/ΟΥΜΑΙ

Πίνακας 5: Μεταδεδομένα ομάδας 5

### 2.2.6 Άλλες μεταπληροφορίες

Η τελευταία ομάδα περιλαμβάνει πληροφορίες σχετικά με την περιοχή στην οποία συναντάει κανείς το λήμμα (πεδίο 57) και την πηγή στην οποία βρίσκεται, με λεπτομερή δήλωση της βιβλιογραφικής αναφοράς (πεδίο 58). Επίσης, ορίζεται το φαινόμενο το οποίο αφορά το συγκεκριμένο λήμμα (πεδίο 62) και άλλες παρατηρήσεις είτε γενικές (πεδίο 63) είτε σχετικά με τη βιβλιογραφική αναφορά (πεδίο 64), εάν αυτές υπάρχουν.

Μεταδεδομένο/Χαρακτηριστικό	Τιμές
57. Περιοχή	Λίστα περιοχών
58. Βιβλιογραφική Αναφορά	Λίστα με βιβλιογραφικές αναφορές
62. Φαινόμενο	-
63. Γενικές Παρατηρήσεις	Ελεύθερο κείμενο
64. Σχόλια βιβλιογραφικής αναφοράς	Ελεύθερο κείμενο

Πίνακας 6: Μεταδεδομένα ομάδας 6

Θα πρέπει να σημειωθεί ότι έχει γίνει πρόβλεψη για μία αναλυτική παρουσίαση του τύπου της πηγής από την οποία έχει συλλεχθεί η πληροφορία (γραπτό κείμενο, ερωτηματολόγιο ή προφορικό υλικό), αλλά δεν εμφανίζεται στη βάση στην παρούσα φάση (αφορά στα πεδία 59, 60 και 61 τα οποία δεν εμφανίζονται στον παραπάνω πίνακα).

## 2.3 Βιβλιογραφική βάση

Στο πλαίσιο του προγράμματος έγινε καταγραφή της υπάρχουσας βιβλιογραφίας σχετικά με τις διαλέκτους οι οποίες εξετάζονται στο πρόγραμμα. Οι ερευνητές συνέλεξαν, μελέτησαν και ταξινομήσαν όλα τα δημοσιευμένα κείμενα (γραμματικές, λεξικά, λογοτεχνικά έργα γραμμένα στις διαλέκτους, βιβλιοκριτικές, επιστημονικά άρθρα) με στόχο να δημιουργήσουν έναν κατάλογο βιβλιογραφίας αναφορικά με τις διαλέκτους και να θέσουν τις βάσεις για την ανάλυσή τους.

Ο ενιαίος κατάλογος βιβλιογραφίας με έργα τα οποία αποτελούν τις πηγές για την παρούσα έρευνα ενσωματώθηκε στη βάση δεδομένων MORILAN, ώστε να υπάρχει η δυνατότητα αυτόματης συμπλήρωσης της βιβλιογραφικής αναφοράς από τους ληματογράφους για κάθε λήμμα το οποίο καταχωρείται.

Παρόλο που η συγκεκριμένη διαδικασία συνήθως ακολουθεί την έρευνα, δεδομένου ότι η ερευνητική ομάδα έχει εμπειρία στη μελέτη διαλέκτων, αποφασίστηκε να δημιουργηθεί εξ αρχής ένας κατάλογος για να αποφευχθούν λάθη από τους ληματογράφους, αλλά και για να γίνει πιο εύκολη και γρήγορη η διαδικασία.

Η *επισημειωμένη βιβλιογραφία* (annotated bibliography) δημιουργήθηκε αρχικά σε υπολογιστικό φύλλο excel και στη συνέχεια ενσωματώθηκε στη βάση δεδομένων του προγράμματος. Περιλαμβάνει βιβλιογραφικές πηγές οι οποίες έχουν ταξινομηθεί με βάση τα διεθνή δεδομένα σε 3 τύπους βιβλιογραφίας ως εξής:

- i. *Βιβλίο, Μονογραφία* (book)
- ii. *Συμβολή σε τόμο* (book section)
- iii. *Συμβολή σε περιοδικό* (journal)

Αντίστοιχα, οι μεταπληροφορίες είναι οι εξής:

- i. *Είδος βιβλιογραφικής αναφοράς* (bibliographic type)
- ii. *Συγγραφέας* (author)
- iii. *Τίτλος* (title)
- iv. *Έτος έκδοσης* (year)
- v. *Πόλη* (city)
- vi. *Εκδότης* (publisher)
- vii. *Επιμελητής* (editor)
- viii. *Τίτλος βιβλίου* (book title)
- ix. *Όνομα περιοδικού* (journal name)
- x. *Τόμος* (volume)
- xi. *Σελίδες* (pages)
- xii. *Κωδικοποίηση διαλέκτων* (dialect code), κωδικοποίηση διαλέκτων οι οποίες είναι σχετικές με το κείμενο.
- xiii. *Λέξεις κλειδιά* (keywords), κωδικοποίηση σχετικών με το κείμενο επιπέδων γλωσσικής ανάλυσης και γλωσσικών φαινομένων.
- xiv. *Είδος πηγής* (source type), (α) πρωτογενής (β) λεξικό/γλωσσάρι (γ) γραμματική (δ) γλωσσολογική μελέτη (ε) άλλο.
- xv. *Ερευνητικό ερώτημα* (research question), κωδικοποίηση πληροφοριών σχετικά με τα επιμέρους ερευνητικά ερωτήματα τα οποία απασχολούν το πρόγραμμα.



- xvi. *Τοποθεσία αντικείμενου* (item location), κωδικοποίηση χώρου (π.χ. βιβλιοθήκη) στον οποίο βρίσκεται το κείμενο.
- xvii. *Περίληψη* (abstract), κωδικοποίηση πληροφοριών σχετικών με το θέμα του κειμένου.
- xviii. *Ψηφιοποιημένο* (digitized), επιλογή μεταξύ (α) Ναι (β) Όχι.
- xix. *Εξωτερικός σύνδεσμος* (external link), κωδικοποίηση του κατά πόσο το κείμενο είναι προσβάσιμο ηλεκτρονικά, π.χ. σε κάποιο σκληρό δίσκο ή στο διαδίκτυο.
- xx. *Σχόλια* (comments), παράθεση άλλων σχολίων τα οποία, για παράδειγμα, μπορεί να αφορούν στο κατά πόσο το κείμενο μπορεί να γίνει αντικείμενο δανεισμού ή όχι.

Κατά τη συγγραφή του παρόντος άρθρου, έχει γίνει καταχώρηση 1.280 γραπτών πηγών.

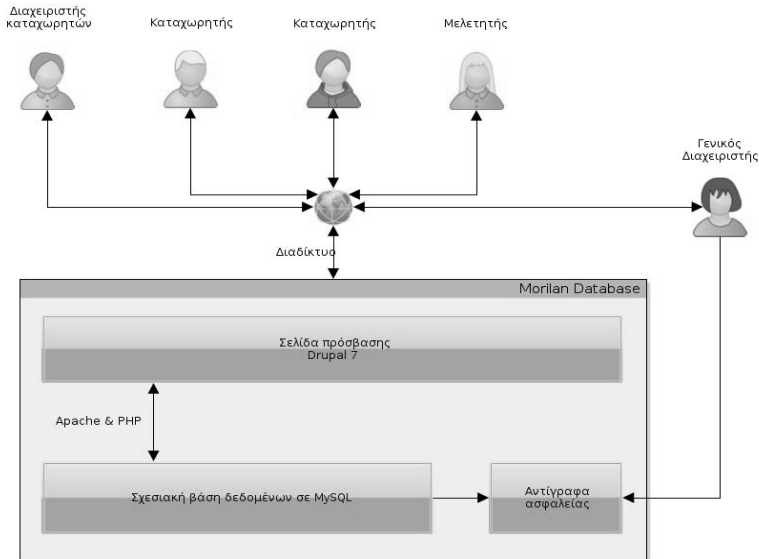
### 3. Υλοποίηση διεπαφής

Η κατασκευή της βάσης δεδομένων, αλλά και της εφαρμογής διεπαφής που την συνοδεύει κατασκευάστηκε στην πλατφόρμα διαχείρισης υλικού γενικού σκοπού (Content Management System) ανοιχτού κώδικα Drupal, έκδοσης 7, η οποία υποστηρίζεται ενεργά από χιλιάδες προγραμματιστές παγκοσμίως (Tomlinson & VanDyke 2010). Η πλατφόρμα Drupal αποτελεί το *λογισμικό* (software) και μπορεί να εγκατασταθεί σε ποικιλία επιλογών *λειτουργικού συστήματος* (operating system) και *υλικού* (hardware). Στο πλαίσιο του προγράμματος επιλέχθηκε λειτουργικό σύστημα και εφαρμογές ανοιχτού κώδικα (LAMP, Linux, Apache, MySQL, PHP) και είναι πλήρως ελεύθερα από περιορισμούς πνευματικών δικαιωμάτων.

Η επιλογή της συγκεκριμένης πλατφόρμας λογισμικού βασίστηκε στους εξής βασικούς λόγους:

- i. εύκολη διαχείριση χρηστών,
- ii. εύκολη διαχείριση και αναζήτηση λημμάτων,
- iii. διαβάθμιση πρόσβασης χρηστών σε προστατευμένες λειτουργίες,
- iv. δημιουργία αυτόματων αντιγράφων ασφαλείας,
- v. πρόσβαση μέσω web.

Οι χρήστες (καταχωρητές και ερευνητές) αλλά και διαχειριστές του συστήματος (γενικός διαχειριστής της βάσης και διαχειριστής καταχωρητών) αποτελούν ένα σύστημα τα οποίο σχηματικά παρουσιάζεται ως εξής:



Εικόνα 1: Διασύνδεση χρηστών με το σύστημα

### 3.1 Διαβάθμιση χρηστών

Κρίθηκε απαραίτητη για λόγους ασφαλείας η διαβάθμιση των χρηστών της βάσης στους παρακάτω ρόλους, όπως αναλυτικά εξηγούνται στην συνέχεια (ξεκινώντας από τον χρήστη με τις ελάχιστες δυνατότητες πρόσβασης):

- i. *Ανώνυμος χρήστης*: Έχει πρόσβαση μόνο στην σελίδα σύνδεσης. Δεν μπορεί να δει τα λήμματα όπως έχουν καταχωρηθεί και να προχωρήσει σε επεξεργασία αυτών.
- ii. *Καταχωρητής (linguist)*: Δικαιούται να καταχωρήσει νέα λήμματα ή περαιτέρω βιβλιογραφία που είναι απαραίτητη για την μελέτη του. Επίσης, μπορεί να επεξεργαστεί τα λήμματα τα οποία έχει καταχωρήσει ο ίδιος. Δεν έχει δικαιώματα να διαγράψει οποιαδήποτε καταχώριση για λόγους ασφαλείας. Εφόσον κρίνει πως ένα λήμμα πρέπει να διαγραφεί, ενημερώνει τον διαχειριστή, που τον επιβλέπει στα πλαίσια της μελέτης του.
- iii. *Διαχειριστής καταχωρητών (manager)*: Έχει τα ίδια δικαιώματα πρόσβασης με τον καταχωρητή, αλλά επιπλέον μπορεί να διαγράψει λήμματα ή βιβλιογραφία, αν το κρίνει απαραίτητο. Επίσης, έχει πρόσβαση σε στατιστικά σχετικά με τα καταχωρημένα λήμματα, δηλαδή ποιος χρήστης καταχώρησε πόσα λήμματα. Τέλος, έχει πρόσβαση στην αυτόματη εξαγωγή όλων των λημμάτων σε μορφή CSV (Comma separated values) για εισαγωγή σε λογιστικό ή στατιστικό πρόγραμμα όπως το Microsoft Excel ή R/SPSS.
- iv. *Γενικός διαχειριστής (administrator)*: Έχει πλήρη δικαιώματα σε όλες τις καταχωρήσεις, αλλά και τους χρήστες, προσθέτοντας ή διαγράφοντας κατά βούληση. Επίσης, έχει δικαίωμα να αλλάξει ή να προσθέσει στα διαθέσιμα

πεδία καταγραφής, στα λήμματα ή στην βιβλιογραφία. Αναλαμβάνει να υποστηρίξει τεχνικά την πλατφόρμα σε περίπτωση δυσλειτουργιών ή κακόβουλης χρήσης.

### 3.2 Διαχείριση λημμάτων

Ένας λόγος που χρησιμοποιήθηκε η πλατφόρμα του Drupal είναι η εύκολη διαχείριση ψηφιακής ύλης, δηλαδή λήμματα και βιβλιογραφία στο πλαίσιο του έργου μας. Συγκεκριμένα, η αρχική σελίδα όλων των συνδεδεμένων χρηστών αποτελεί και την αναζήτηση λημμάτων. Με αφετηρία την συγκεκριμένη σελίδα μπορεί να αναζητήσει ένα λήμμα προς επιβεβαίωση πριν το καταχωρήσει ξανά ή να φορτώσει όλες τις λεπτομέρειες που το απαρτίζουν.

Στην εικόνα 2, εμφανίζεται ενδεικτικό παράδειγμα της σελίδας αναζήτησης λήμματος. Γίνεται σαφές πως έχουν τοποθετηθεί συγκεκριμένα στοιχεία αναζήτησης αλλά όχι η πλήρης λίστα μεταδεδομένων. Αντίστοιχα, η σελίδα προβολής ενός λήμματος σε απλές και καθαρές γραμμές σχεδίασης, χωρίς να επιβαρύνει την οθόνη του καταχωρητή/μελετητή με περιττά στοιχεία παρουσιάζεται στην εικόνα 3.

Κωδ.	Λήμμα	Σημασία	Μορφολογική Διαδικασία	Γραμματική Κατηγορία	Γένος	Κλιτική Τύξη	Επιβεβαίωση
11455	(α)ριβεντέρτα	εις το επανδέν	Κλίση - Ακλίσια	Έκφραση	Χωρίς γένος	Άκλιτο	Όχι
5008	απρούνω, προυίνω	1. Γνωμαι καχεκτικό...	Παραγωγή - Κλίση	Ρήμα	Χωρίς γένος	KT1, ρήματα	Όχι
12431	κα(ν)μυνοτσόδρος	ο διοικητής, ο κυβερ...	Κλίση - Ακλίσια	Ουσιαστικό	Αρσενικό	KT1	Όχι
9106	ναμκόρης	αγόρας	Κλίση - Ακλίσια	Ουσιαστικό	Αρσενικό	KT2	Όχι
8928	ραγούτ	τιμηelió	Κλίση - Ακλίσια	Ουσιαστικό	Ουδέτερο	KT6	Όχι
5009	προυίνω	1. Γνωμαι καχεκτικό...	Παραγωγή - Κλίση	Ρήμα	Χωρίς γένος	KT1, ρήματα	Όχι
10888	(α)κιάεκα	μηλαλόνη	Κλίση - Ακλίσια	Ουσιαστικό	Αρσενικό	KT1	Όχι
10190	(c)annetsicula	γροσδι, μαλλί του λ...	Κλίση - Ακλίσια	Ουσιαστικό	Θηλυκό	KT3	Όχι
6912	(α)βερατσούρα	άνογμα, τρίπα, σχα...	Κλίση - Ακλίσια	Ουσιαστικό	Θηλυκό	KT3	Όχι
1406	(α)γαντζόρω	αγκαστράνω, δένω	Κλίση - Ακλίσια	Ρήμα	Χωρίς γένος	KT1, ρήματα	Όχι
6242	(α)γκαστάρω	ασακτό ακίντο, κομμ...	Κλίση - Ακλίσια	Ρήμα	Χωρίς γένος	KT1, ρήματα	Όχι
6243	(α)γκιάστο	απάκτρω, κτήμα. Η ν...	Κλίση - Ακλίσια	Ουσιαστικό	Ουδέτερο	KT5	Όχι
9299	(α)γκορνιστόρος	κουρδιατής πιόνων	Κλίση - Ακλίσια	Ουσιαστικό	Αρσενικό	KT1	Όχι
9491	(α)μάκα	φιλοδύρημα, τζόμα...	Κλίση - Ακλίσια	Ουσιαστικό	Θηλυκό	KT3	Όχι
9505	(α)μαλάτος	άρρωστος, ασθενής, α...	Κλίση - Ακλίσια	Επίθετο	Τριγενές	Μεικτή κλιτική τάξη-Επίθετο	Όχι
4761	(α)μαμάρισμα	η απελευθέρωση, η πα...	Παραγωγή - Κλίση	Ουσιαστικό	Ουδέτερο	KT8	Όχι
9739	(α)μαμάρω	απολάω, αφήνω, χαλαρ...	Κλίση - Ακλίσια	Ρήμα	Χωρίς γένος	KT1, ρήματα	Όχι
1416	(α)μαλόρω (ή αμαλόρω)	1)αμαλω, 2)δίνω σ...	Κλίση - Ακλίσια	Ρήμα	Χωρίς γένος	KT1, ρήματα	Όχι
6476	(α)μαρούρος	ερωτοπαθής, ερωτήρη...	Κλίση - Ακλίσια	Επίθετο	Τριγενές	Μεικτή κλιτική τάξη-Επίθετο	Όχι
10206	(α)μπαονόρα	ναρίς, πολύ πραι	Κλίση - Ακλίσια	Επίρρημα	Χωρίς γένος	Άκλιτο	Όχι
9528	(α)μπαονόρα	από ναρίς	Κλίση - Ακλίσια	Έκφραση	Χωρίς γένος	Άκλιτο	Όχι

Εικόνα 2: Σελίδα προβολής και αναζήτησης λημμάτων

The screenshot shows the website interface for 'Μορφολογία και γλωσσική επαφή' (Morphology and Linguistic Contact). The page title is '(α)ριβεντέρισι' and it is categorized as 'Προβολή' (Profile) and 'Επεξεργασία' (Edit). The page contains several sections:

- Βασικές Πληροφορίες:**
  - 2. Σημασία: εις το επανιδείν
  - 3. Μορφολογική Διαδικασία: Κλίση - Ακυσία
  - 4. Γραμματική Κατηγορία: Έκφραση
  - 6. Κλιτική Τάξη: Ακίπτο
  - 5. Γένος: Χωρίς γένος
- Καταγωγή:**
  - 7. Καταγωγή Λήμματος: Ιταλή
  - 8. Πρωτότυπη Δόνηση Λέξης: απινδενεισι
  - 9. Γένος Δόνησης Λέξης: Χωρίς γένος
- Μετά-πληροφορίες:**
  - 57. Περιοχή: Κέρκυρα (Νευκίμυ)
  - 58. Βιβλιογραφική Αναφορά: Η παλιά Κορφιάτικη γλώσσα. Book | Αγγελόπουλος, Μ. | 2008 | Κέρκυρα | Ενημέρωση | Κέρκυρα | Λεξιλόγιο | Λεξικό | Σύλλογος Κεφαλλήνων Πάτρων |
  - 64. Σχόλια βιβλιογραφικής αναφοράς: α 251

At the bottom, there is a section for 'Έλεγχος επιβεβαιώθηκε' (Checked/verified) and a form for adding a new comment, including fields for 'To όνομά σας' (Your name), 'Γενικός διαχειριστής' (Administrator), and a 'Comment' text area.

Εικόνα 3: Σελίδα προβολής λήμματος

### 3.3 Δυνατότητα σχολίων

Ένα ακόμη πλεονέκτημα της συγκεκριμένης πλατφόρμας είναι η ευκολία διαχείρισης σχολίων κάτω από κάθε λήμμα. Η συγκεκριμένη υποστήριξη δίνει την δυνατότητα στον καταχωρητή/μελετητή να καταχωρήσει ένα σχόλιο για τον ίδιο (ως σημείωση για περαιτέρω μελέτη) ή για τα άλλα μέλη της ομάδας (για προβληματισμό και συζήτηση). Ως εκ τούτου, κάθε λήμμα έχει την δυνατότητα να υποστηρίξει ένα *νήμα ελεύθερης συζήτησης* (forum thread) από όλα τα μέλη της ομάδας.

### 3.4 Ιστορικό αλλαγών

Καθώς η εργασία καταχώρισης αποτελεί μια επίπονη διαδικασία, κρίθηκε απαραίτητο να διασφαλιστεί η προσπάθεια όλων των μελών μεμονωμένα, αλλά και συνολικά. Στο πλαίσιο αυτό, ορίσαμε την διασφάλιση της αναφοράς εργασίας κάθε λήμματος σε συγκεκριμένο καταχωρητή, μεταξύ άλλων και για να αναζητηθούν πιθανά σφάλματα στην πάροδο του χρόνου. Για να επιτευχθεί η συγκεκριμένη αξίωση, η αποθήκευση κάθε λήμματος δημιουργεί νέα έκδοση/αντίγραφο όλων των πληροφοριών. Με μικρή επιβάρυνση στην βάση δεδομένων, είμαστε σε θέση να δούμε τις αλλαγές στην πάροδο του χρόνου. Για παράδειγμα, στην παρακάτω εικόνα εμφανίζεται το ιστορικό αλλαγών σε συγκεκριμένο λήμμα:

Αναθεώρηση	Compare	Λειτουργίες
την: Τρί, 13/01/2010 13:35, από: [User Name]	<input type="radio"/>	τρέχουσα αναθεώρηση
την: Σάβ, 28/01/2010 14:53, από: [User Name]	<input type="radio"/>	Επαναφορά Διαγραφή
την: Τρί, 03/06/2010 19:35, από: [User Name]	<input type="radio"/>	Επαναφορά Διαγραφή
την: Πέμ, 18/01/2010 20:20, από: [User Name]	<input type="radio"/>	Επαναφορά Διαγραφή
την: Τετ, 17/01/2010 19:06, από: [User Name]	<input type="radio"/>	Επαναφορά Διαγραφή
την: Δευ, 15/01/2010 12:55, από: [User Name]	<input type="radio"/>	Επαναφορά Διαγραφή

Εικόνα 4: Ιστορικό αλλαγών λήμματος

Ενώ, αν επιλέξουμε δύο καταχωρήσεις του ίδιου λήμματος σε διαφορετικές χρονικές στιγμές και πατήσουμε σύγκριση (κουμπί ‘Compare’), θα βρεθούμε στην εξής εικόνα:

Αναθεώρησης του (α)γαντζάρω	Προβολή	Επεξεργασία	Αναθεώρησης
την: Σάβ, 28/01/2010 14:53, από: [User Name] Κα...			την: Τρί, 13/01/2010 13:35, από: [User Name] Κα...
< Previous difference			
Basic Marked down			
<b>Αλλαγές στο 64. Σχόλια Βιβλιογραφικής αναφοράς</b>			<b>σελ. 3</b>
<b>Αλλαγές στο 63. Γενικές Παρατηρήσεις</b>			
<p>Αν υποθέσουμε ότι η λέξη έχει προέλθει από το ιταλικό ουσιαστικό gancio (II), τότε η μορφολογική διαδικασία που έχει πραγματοποιηθεί για το σχημασμό του ρήματος αγαντζάρω είναι η παραγωγή-κλίση. Χρησιμοποιείται δηλαδή το παραγωγικό επίθημα -αρ(ω) και προσκολλάται στη δάνεια βάση gancio για να ενσωματωθεί τη λέξη στη διάλεκτο.</p>			<p>Αν υποθέσουμε ότι η λέξη έχει προέλθει από το ιταλικό ουσιαστικό gancio (II), τότε η μορφολογική διαδικασία που έχει πραγματοποιηθεί για το σχημασμό του ρήματος αγαντζάρω είναι η παραγωγή-κλίση. Χρησιμοποιείται δηλαδή το παραγωγικό επίθημα -αρ(ω) και προσκολλάται στη δάνεια βάση gancio για να ενσωματωθεί τη λέξη στη διάλεκτο.</p>

Εικόνα 5: Σύγκριση λήμματος πριν και μετά την επεξεργασία του

Γίνεται πλέον σαφές ότι στο παραπάνω παράδειγμα (εικόνα 5) ο συγκεκριμένος καταχωρητής αποφάσισε μετά από την μελέτη του, να συμπληρώσει επιπλέον σχόλια στο πεδίο 64, και συγκεκριμένα τον αριθμό σελίδας που βρήκε την βιβλιογραφική αναφορά. Στην εικόνα 5, φαίνεται ως «σελ. 3» με έντονα κόκκινα γράμματα όπως έχει προστεθεί στην δεξιά ενημερωμένη έκδοση του λήμματος σε σχέση με την αριστερή έκδοση, από όπου απουσιάζει. Αν και η παρούσα ανάλυση (ή καταγραφή των αλλαγών) θα μπορούσε να θεωρηθεί υπερβολικά λεπτομερής για την γλωσσολογική μελέτη, ωστόσο παρέχει ένα εξαιρετικά δυνατό εργαλείο στους διαχειριστές του συστήματος, χωρίς να επιφορτίζει στο ελάχιστο τους μελετητές-γλωσσολόγους να διεκπεραιώσουν την εργασία τους. Σε περίπτωση αναγνώρισης λάθους, μπορεί να γίνει επαναφορά του λήμματος σε προηγούμενη έκδοση αλλά και εξακρίβωση μιας εσφαλμένης κατάστασης.

Η διασφάλιση του συστήματος στο σύνολο, επιβεβαιώνεται με ημερήσια αντίγραφα ασφαλείας της βάσης δεδομένων τα οποία παραμένουν σε βάθος 10 ημερών. Το συγκεκριμένο διάστημα κρίθηκε επαρκές, για να καλύψει πιθανά προβλήματα *κατάρρευσης του υλικού* (hardware failure), αλλά και επιθέσεων από κακόβουλους χρήστες που κατέλαβαν το σύστημα παράνομα (με χρήση ιού ή κενού ασφαλείας).

### 3.5 Πρόσβαση μέσω διαδικτύου

Η δυνατότητα πρόσβασης μέσω διαδικτύου παρέχει πολλά πλεονεκτήματα, όπως μεταξύ άλλων: (α) ο καταχωρητής/μελετητής μπορεί να συνδεθεί στο σύστημα χρησιμοποιώντας τον προσωπικό του υπολογιστή απλά γνωρίζοντας το URL σύνδεσης και τα προσωπικά του στοιχεία πρόσβασης, (β) δεν χρειάζεται εγκατάσταση εφαρμογής στον υπολογιστή, οπότε γίνεται πλήρης αποσύνδεση λειτουργικού συστήματος και υπηρεσίας, δηλαδή ο χρήστης μπορεί να χρησιμοποιεί Windows, Linux, Mac, ταμπλέτες iPad/Android, αλλά

να μην στερείται δυνατότητες στο ελάχιστο, (γ) οποιαδήποτε αλλαγή στα πεδία ή σε δυνατότητες της υπηρεσίας γίνονται κεντρικά και όλοι οι χρήστες είναι αυτόματα ενήμεροι, (δ) οποιαδήποτε αλλαγή καταχωρείται σε ασφαλές κεντρικό σημείο και δεν υπάρχει ο φόβος να χαθούν πληροφορίες σε πιθανή κλοπή ή καταστροφή υλικού του χρήστη, (ε) παράλληλη πρόσβαση στο σύστημα από πολλαπλούς χρήστες που μπορούν να καταχωρούν ή/και να επεξεργάζονται παράλληλα όλα τα στοιχεία της πλατφόρμας, και (στ) ο καταχωρητής/μελετητής μπορεί να εργαστεί εκτός γραφείου έχοντας απλώς σύνδεση στο δίκτυο μέσω κινητής συσκευής ή ηλεκτρονικής ταμπλέτας.

#### 4. Καινοτομίες της βάσης και μελλοντική χρήση

Στο πλαίσιο του προγράμματος MORILAN αναπτύχθηκε μία μορφολογική βάση δεδομένων που αφορά κυρίως σε πέντε διαλεκτικά συστήματα, αλλά έχει δυνατότητες επέκτασης σε όλες τις ποικιλίες της Νέας Ελληνικής. Η συγκεκριμένη βάση διαθέτει τα καινοτομικά χαρακτηριστικά τα οποία μπορούν να κωδικοποιηθούν ως εξής:

- i. Η βάση έχει κατασκευαστεί εξ ολοκλήρου με διαθέσιμα εργαλεία ανοικτής πρόσβασης.
- ii. Η βάση περιλαμβάνει λεπτομερή ανάλυση δεδομένων σύμφωνα με όλες τις πιθανές *μορφολογικές κατηγορίες* και περιέχει πληροφορίες από όλες τις *μορφολογικές διαδικασίες* (κλίση, παραγωγή, σύνθεση) της Νέας Ελληνικής. Πολλές πληροφορίες που αφορούν στην ανάλυση γίνονται με επιλογή από κλειστή λίστα τιμών και έτσι ο ερευνητής έχει συγκεκριμένο στόχο στην ανάλυσή του και κερδίζει χρόνο στη μελέτη του συστήματος.
- iii. Η εμφάνιση των πληροφοριών κατά τη λημματογράφηση γίνεται με *εξαρτημένες σχέσεις* (dependencies) ανάλογα με την αρχική επιλογή που θα κάνει ο λημματογράφος/καταχωρητής. Η συγκεκριμένη επιλογή έχει τα εξής πλεονεκτήματα: (α) υποχρεώνει τον λημματογράφο να περάσει όλες τις πληροφορίες που απαιτούνται για την ολοκληρωμένη ανάλυση του λήμματος, (β) διευκολύνει την προσπάθεια που πρέπει να καταβάλει στη λημματογράφηση, καθώς ελαχιστοποιεί τις πληροφορίες που πρέπει να επεξεργαστεί, και (γ) δίνει μόνο τις απαραίτητες πληροφορίες στον μελετητή.
- iv. Η βάση περιλαμβάνει πλήθος δεδομένων από διαλεκτικές ποικιλίες που έχουν δεχθεί επίδραση από την Τουρκική και την Ιταλική και δεν είναι εύκολα προσβάσιμες στους ερευνητές. Επομένως, έχει ενδιαφέρον τόσο για τη μορφολογική έρευνα όσο και για την έρευνα των διαλέκτων.
- v. Η βάση έχει σχεδιαστεί, ώστε τα δεδομένα να είναι ταξινομημένα και αναλυμένα τόσο για τους στόχους της παρούσας έρευνας όσο και για μελλοντικές μελέτες. Για παράδειγμα, ενώ σε κάποια πεδία υπάρχουν αναλυτικές πληροφορίες, σε κάποια άλλα, όπως το *φαινόμενο*, αναμένεται να συμπληρωθούν πληροφορίες μετά την ανάλυση συγκεκριμένων φαινομένων.

Από τα παραπάνω γίνεται σαφές ότι η βάση MORILAN ξεπερνά τους στόχους ενός συγκεκριμένου προγράμματος και βάζει τις βάσεις για μελλοντική χρήση και επέκταση στο πλαίσιο άλλων ερευνητικών προγραμμάτων.

## Βιβλιογραφία

- Καρασίμος, Α., Γαλιώτου, Ε., Καρανικόλας, Ν., Κορωνάκης, Γ., Αθανασάκος, Κ., Παπαζαχαρίου, Δ. & Α. Ράλλη. (υπό έκδ.). Προκλήσεις επισημείωσης ενός πολυ-διαλεκτικού, πολυ-επίπεδου σώματος γραπτών και προφορικών κειμένων των Νεοελληνικών Διαλέκτων. Στο Α. Ράλλη, Ν. Κουτσούκος & Σ. Μπόμπολας (επιμ.), *Proceedings of the 6<sup>th</sup> International Conference on Modern Greek Dialects & Linguistic Theory*. Patras: University of Patras.
- Κολιοπούλου, Μ., Μαρκόπουλος, Θ., & Ν., Παντελίδης. (2015). Πόντος, Καπαδοκία, Αϊβαλί: προκλήσεις ενός ψηφιακού σώματος γραπτού υλικού. Στο Kotzoglou et al. (επιμ.), *Proceedings of ICGL11*. Rhodes: University of Aegean, 750-759.
- Ράλλη, Α. (2005). *Μορφολογία*. Αθήνα: Πατάκης.
- Ράλλη, Α. (2015). *Τεχνική αναφορά προγράμματος «Μορφολογία και γλωσσική επαφή : Ελληνικές διάλεκτοι σε επαφή με την Τουρκική και την Ιταλική» - Aristeia I/643*. Διαθέσιμο Στο <http://nemertes.lis.upatras.gr/jspui/handle/10889/8958>.
- Ράλλη, Α., Παπαζαχαρίου, Δ. & Α. Καρασίμος. (2010). Εργαστήριο Νεοελληνικών Διαλέκτων και η βάση δεδομένων GREED. Στο Α. Ράλλη, Β. Joseph, Μ. Janse & Α. Karasimos (επιμ.), *Proceedings of 4<sup>th</sup> International Conference of Modern Greek Dialects and Linguistic Theory*. Patras: University of Patras, 7-14.
- Galiotou, E., Karanikolas, N., Manolessou, I., Pantelidis, N., Papazachariou, D., Ralli, A. & G. Xydopoulos. (2014). Asia Minor Greek: Towards a Computational Processing. Στο G. Giannakopoulos, D. P. Sakas, D. S. Vlachos and D. Kyriaki-Manessi (επιμ.), *Procedia-Social and Behavioral Sciences 147* (3<sup>rd</sup> International Conference on Integrated Information-IC-ININFO): 458-466. Διαθέσιμο Στο <http://dx.doi.org/10.1016/j.sbspro.2014.07.138>.
- Karanikolas, N., Galiotou, E. & A. Ralli. (2014). Towards a unified exploitation of electronic dialectal corpora: problems and perspectives. Στο Petr Sojka, Aleš Horák, Ivan Kopeček & Karel Pala (επιμ.), *TSD 2014*. Switzerland: Springer, 257-266.
- Ralli, A. (2003). Morphology in Greek Linguistics: The State-of-the-art. *Journal of Greek Linguistics* 4: 77-130.
- Ralli, A. (2006). Syntactic and Morphosyntactic Phenomena in Modern Greek Dialects: The State of the Art. *Journal of Greek Linguistics* 7 (1): 121-159.
- Ralli, A. (2009). Morphology meets Dialectology: Insights from Modern Greek Dialects. *Morphology* 19 (1): 87-105.
- Tomlinson, T. & J. Van Dyke. (2010). *Pro Drupal 7 Development* (3<sup>rd</sup> ed.). Apress: USA.