Αθήνα, Παρασκευή 1 Οκτωβρίου 2010

GREED CORPUS: ΟΡΓΑΝΩΣΗ ΚΑΙ ΕΡΓΑΛΕΙΑ ΓΛΩΣΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΨΗΦΙΑΚΗΣ ΒΑΣΗΣ ΕΛΛΗΝΙΚΩΝ ΔΙΑΛΕΚΤΩΝ ΤΟΥ ΕΡΓΑΣΤΗΡΙΟΥ ΝΕΟΕΛΛΗΝΙΚΩΝ ΔΙΑΛΕΚΤΩΝ

Δημήτρης Παπαζαχαρίου

Εργαστήριο Νεοελληνικών Διαλέκτων

Τμήμα Φιλολογίας, Πανεπιστήμιο Πατρών

Ψηφιακή Βάση Ελληνικών Διαλέκτων

- Χαρακτηριστικά: α) Ανεξαρτησία από λειτουργικά συστήματα και εμπορικά λογισμικά

β) Τρισδιάστατη οργάνωση βάσης ως προς:

i) τα *Δεδομένα*

ii) τα *Μεταδεδομένα* που χαρακτηρίζουν τα δεδομένα

iii) το Λογισμικό, που μας επιτρέπει την α) διαχείριση των δεδομένων και μεταδεδομένων, και β) αναζήτηση των δεδομένων σε σχέση με τα μεταδεδομένα που τα περιγράφουν

Α) Δεδομένα (πρωτογενή)

- Ερευνητικά προγράμματα + Διπλωματικές εργασίες προπτυχιακών και μεταπτυχιακών φοιτητών της ειδίκευσης σε διαλεκτικά φαινόμενα από την Κύπρο μέχρι την Θράκη, οι οποίες προϋπέθεταν ηχογραφήσεις διαλεκτικών συνομιλιών

- Σύνολο: **505 ώρες** μαγνητοφωνημένου διαλεκτικού λόγου

Διαλεκτικός χάρτης

Α) Δεδομένα (επεξεργασμένα)

- Απομαγνητοφωνήσεις των ηχητικών αρχείων

  (ορθογραφική μεταγραφή)

Β) Γλωσσικά μεταδεδομένα

- Κωδικοποίηση (annotation)

  (Φωνητική & Προσωδιακή, Φωνολογική, Μορφολογική)

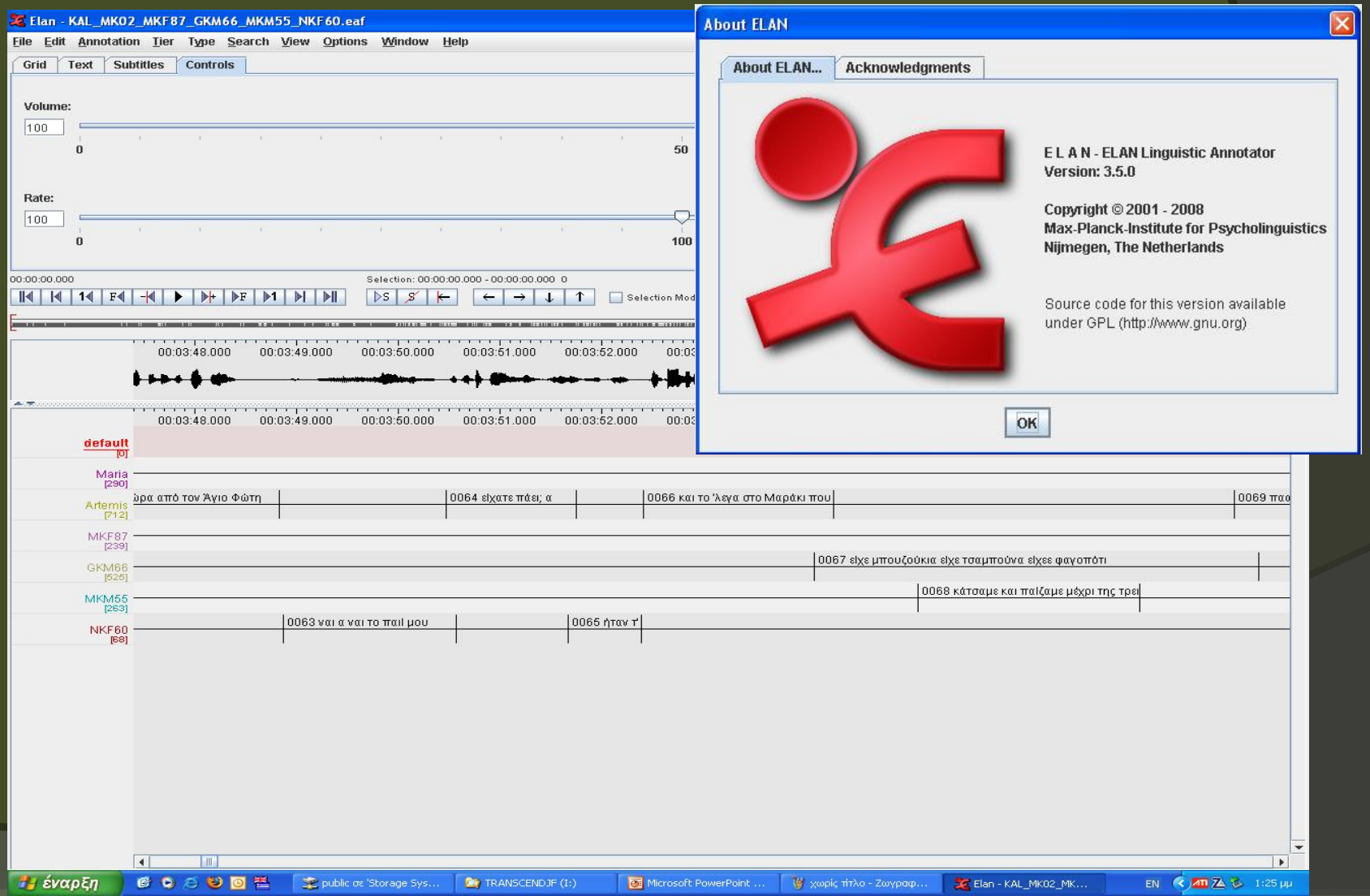
Β) Μη Γλωσσικά Μεταδεδομένα

Β) Μη γλωσσικά μεταδεδομένα: υποκατηγοριοποίηση

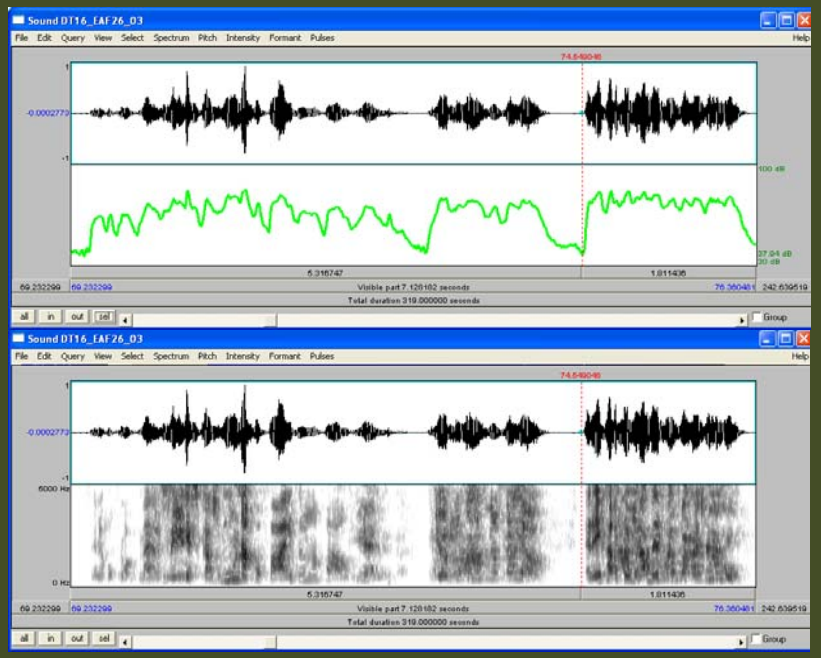
Λογισμικό – Κριτήρια Επιλογής

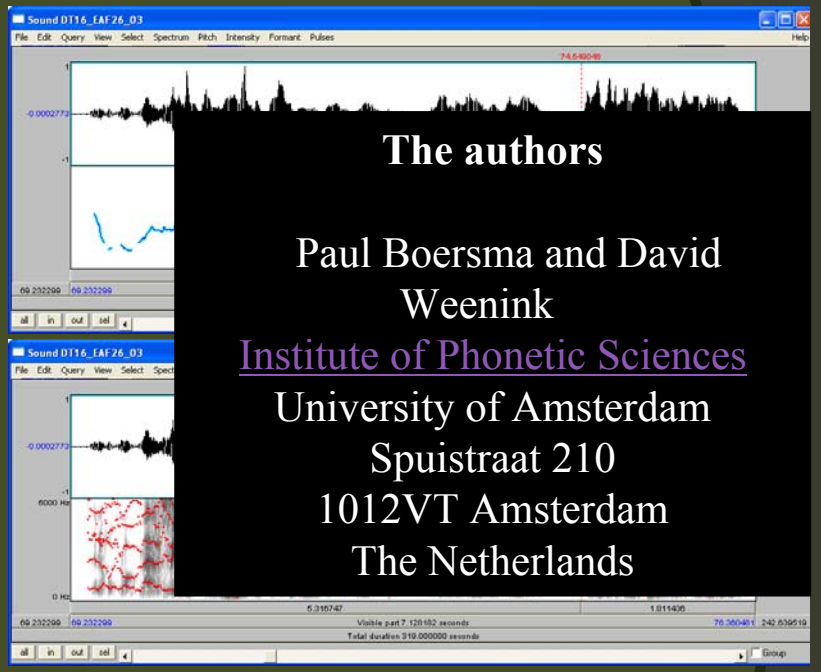
- Λογισμικά αιχμής και ευρείας αποδοχής
- Λογισμικά ανοιχτού κώδικα και ελεύθερα στην χρήση
- Πληθώρα επιλογών για παραμετροποίηση
- Δυνατότητα συνδυασμού των λογισμικών μεταξύ τους
- Συνεχή υποστήριξη από τους δημιουργούς/ προγραμματιστές
- Υποστήριξη ελληνικής γλώσσας και unicode γραμματοσειρών

Λογισμικό για μεταγραφή (ELAN)



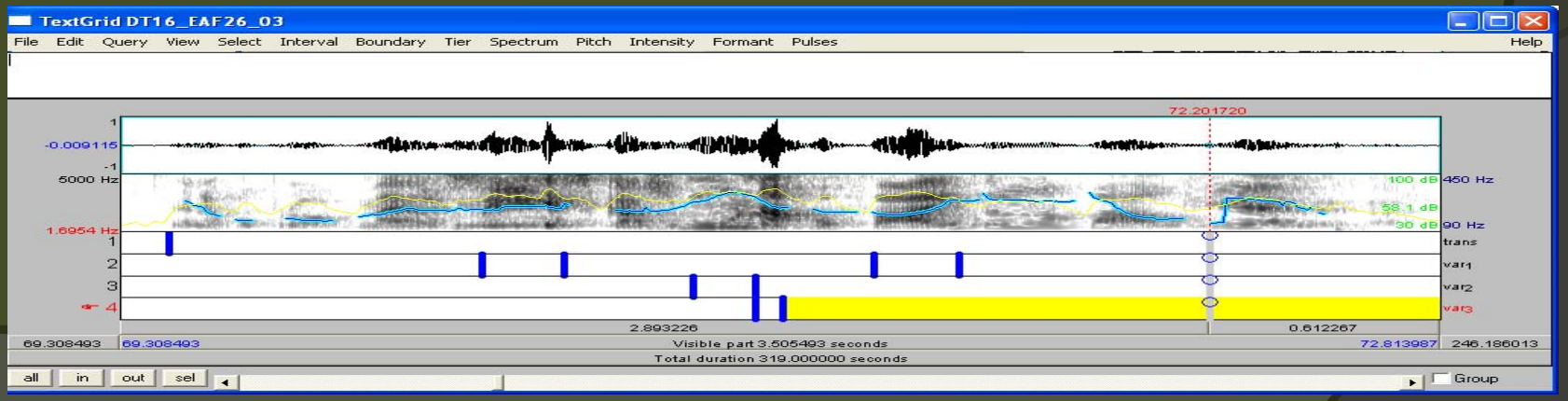
Λογισμικό για Φωνητική / Φωνολογική Ανάλυση (Praat)



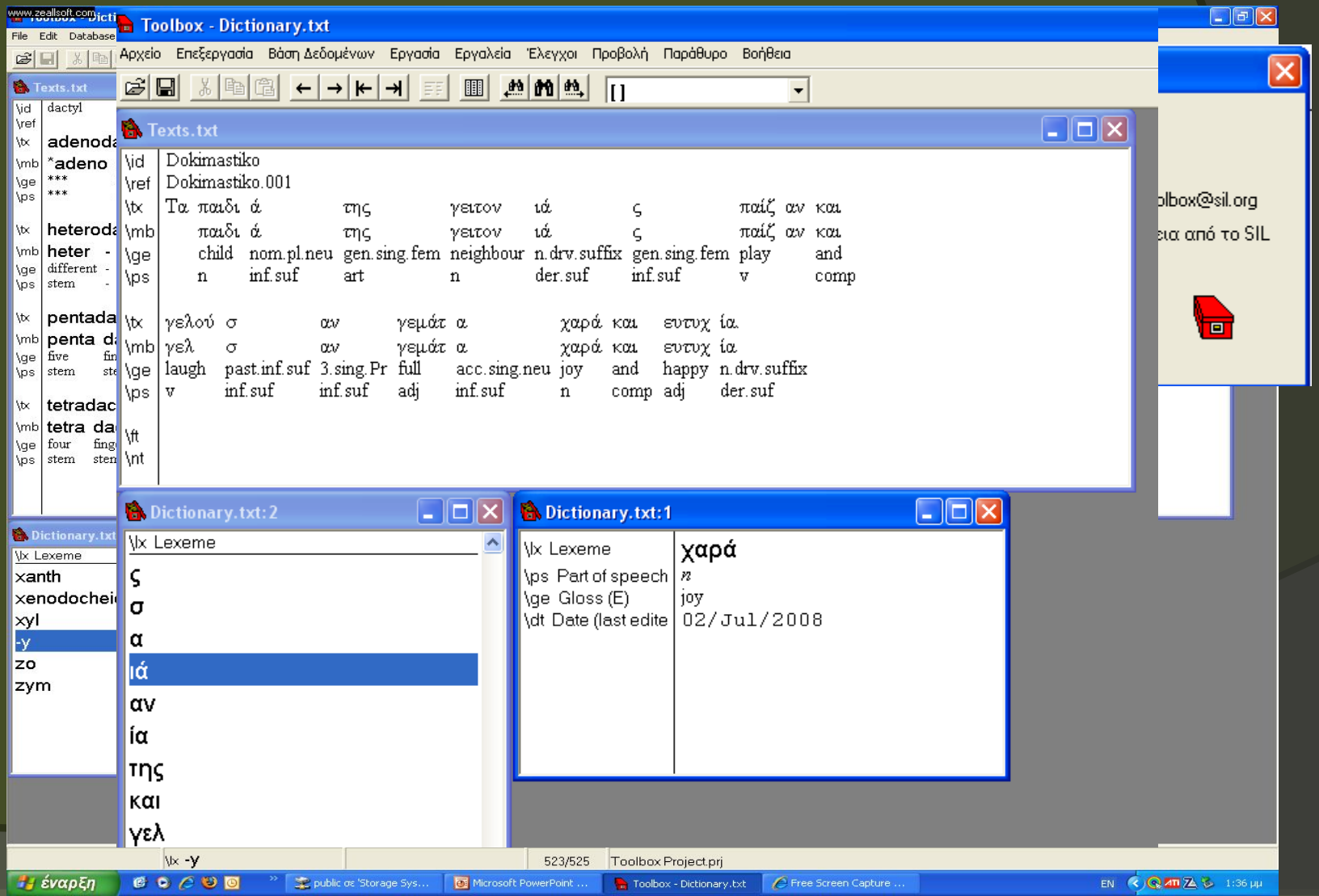


The authors

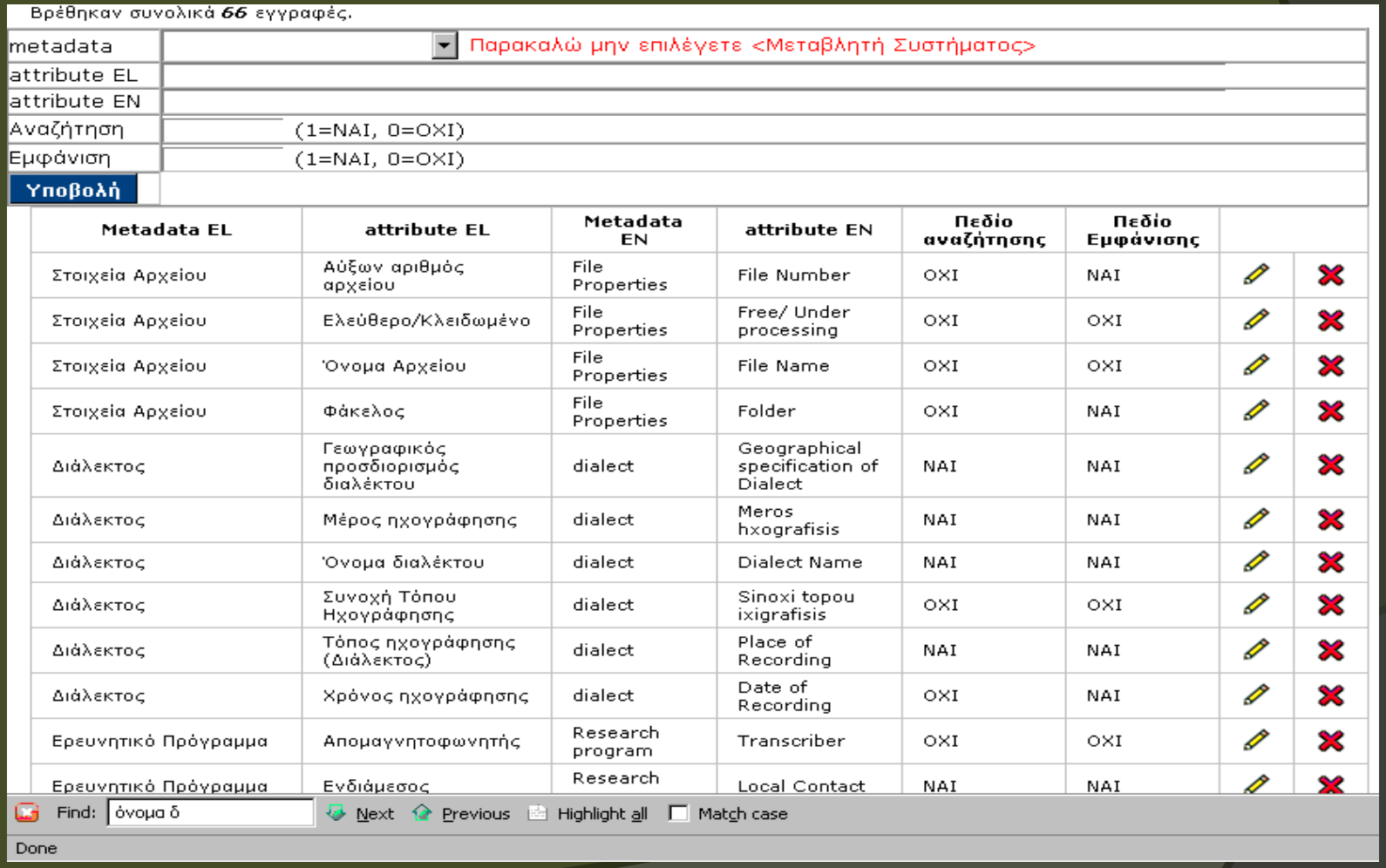
Paul Boersma and David Weenink
Institute of Phonetic Sciences
University of Amsterdam
Spuistraat 210
1012VT Amsterdam
The Netherlands



Λογισμικό για Μορφολογική Ανάλυση (Toolbox)



Γ) Περιβάλλον Βάσης Δεδομένων



Βρέθηκαν συνολικά **66** εγγραφές.

metadata		Παρακαλώ μην επιλέγετε <Μεταβλητή Συστήματος>
attribute EL		
attribute EN		
Αναζήτηση		(1=ΝΑΙ, 0=ΟΧΙ)
Εμφάνιση		(1=ΝΑΙ, 0=ΟΧΙ)

Υποβολή

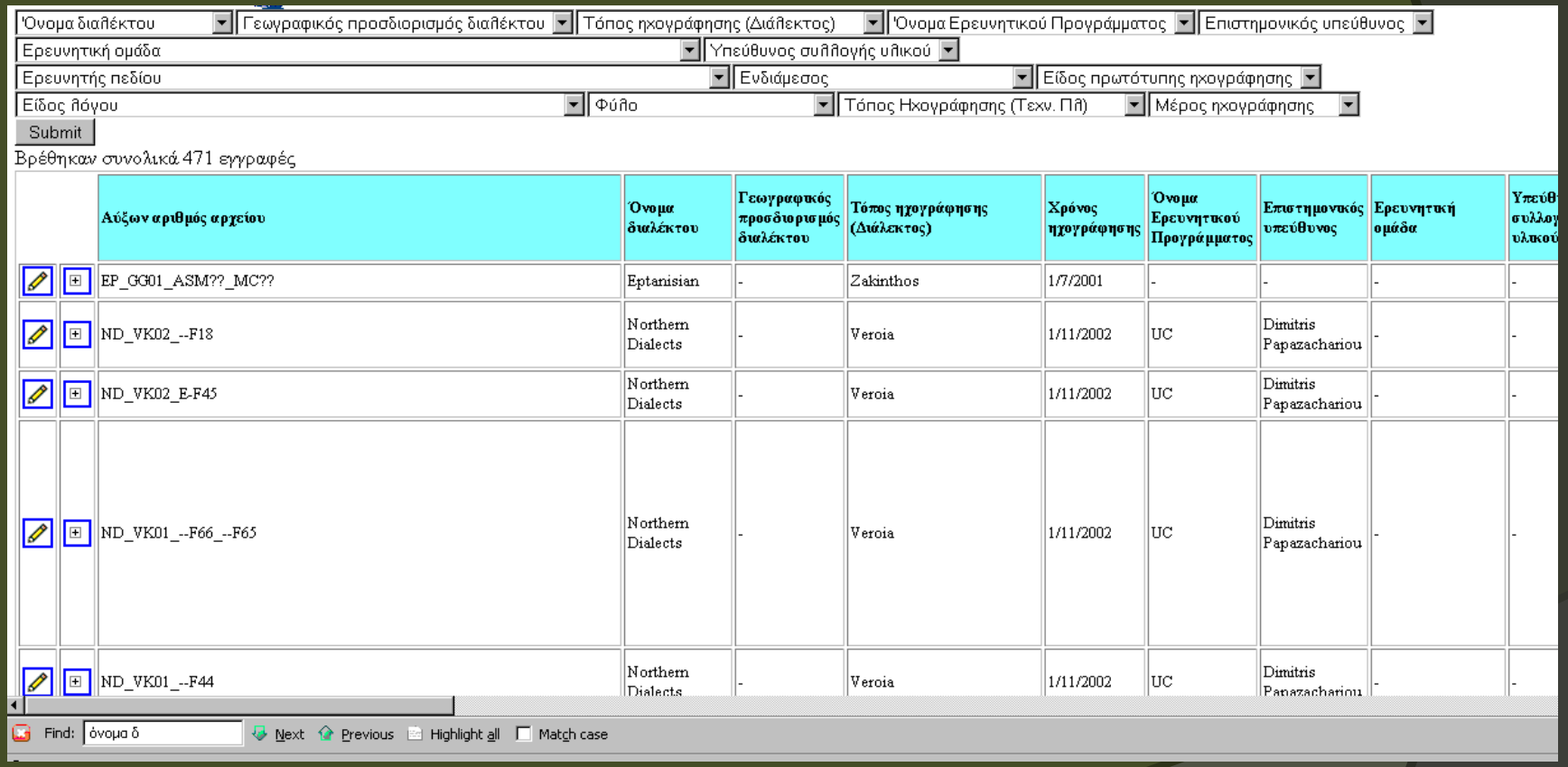
Metadata EL	attribute EL	Metadata EN	attribute EN	Πεδίο αναζήτησης	Πεδίο Εμφάνισης		
Στοιχεία Αρχείου	Αύξων αριθμός αρχείου	File Properties	File Number	ΟΧΙ	ΝΑΙ		
Στοιχεία Αρχείου	Ελεύθερο/Κλειδωμένο	File Properties	Free/ Under processing	ΟΧΙ	ΟΧΙ		
Στοιχεία Αρχείου	Όνομα Αρχείου	File Properties	File Name	ΟΧΙ	ΟΧΙ		
Στοιχεία Αρχείου	Φάκελος	File Properties	Folder	ΟΧΙ	ΝΑΙ		
Διάλεκτος	Γεωγραφικός προσδιορισμός διαλέκτου	dialect	Geographical specification of Dialect	ΝΑΙ	ΝΑΙ		
Διάλεκτος	Μέρος ηχογράφησης	dialect	Meros hxografisis	ΝΑΙ	ΝΑΙ		
Διάλεκτος	Όνομα διαλέκτου	dialect	Dialect Name	ΝΑΙ	ΝΑΙ		
Διάλεκτος	Συνοχή Τόπου Ηχογράφησης	dialect	Sinoxi topou ixigrafisis	ΟΧΙ	ΟΧΙ		
Διάλεκτος	Τόπος ηχογράφησης (Διάλεκτος)	dialect	Place of Recording	ΝΑΙ	ΝΑΙ		
Διάλεκτος	Χρόνος ηχογράφησης	dialect	Date of Recording	ΟΧΙ	ΝΑΙ		
Ερευνητικό Πρόγραμμα	Απομαγνητοφωνητής	Research program	Transcriber	ΟΧΙ	ΟΧΙ		
Ερευνητικό Πρόγραμμα	Ενδιάμεσος	Research	Local Contact	ΝΑΙ	ΝΑΙ		

Find: όνομα δ | Next | Previous | Highlight all | Match case

Done

Ο Διαχειριστής της βάσης δεδομένων μπορεί να εισάγει όσες υποκατηγορίες metadata (attributes) επιθυμεί.

Γ) Αναζήτηση δεδομένων



Τα πεδία αναζήτησης εμφανίζονται ως combo boxes και παράγονται στη σελίδα δυναμικά. Αυτό σημαίνει ότι ανάλογα με τη συμπλήρωση των προηγούμενων σελίδων, από τον διαχειριστή του συστήματος, παράγονται δυναμικά τα πεδία αναζήτησης καθώς και τα πεδία εμφάνισης.

Μελλοντικοί Στόχοι

- Παλαιά κείμενα και ιστορικά έγγραφα (αρχεία εικόνας)
- Μηχανή αναζήτησης που να συνδυάζει και γλωσσικές και μη γλωσσικές πληροφορίες
- Διαβαθμισμένα επίπεδα πρόσβασης για τους χρήστες
- Περισσότερες μεταγραφές / περισσότερη κωδικοποίηση