

ΚΩΔΙΚΟΙ ΓΙΑ ΤΟΥΣ ΕΛΛΗΝΙΚΟΥΣ ΧΑΡΑΚΤΗΡΕΣ: ΠΡΟΒΛΗΜΑΤΑ ΚΑΙ ΣΥΓΧΡΟΝΕΣ ΛΥΣΕΙΣ*

Η ΚΩΔΙΚΟΠΟΙΗΣΗ των ελληνικών χαρακτήρων για ψηφιακή επεξεργασία παρουσιάζει ορισμένα ειδικά προβλήματα. Θα προσπαθήσουμε να αναλύσουμε αυτή την ιδιομορφία και να συζητήσουμε τις λύσεις που προσφέρουν οι διεθνείς πίνακες Unicode και ISO 10646.

Ο ΡΟΛΟΣ ΤΩΝ ΚΩΔΙΚΩΝ ΓΙΑ ΧΑΡΑΚΤΗΡΕΣ

Στους ηλεκτρονικούς υπολογιστές οι κωδικοί των χαρακτήρων απεικονίζουν τα στοιχεία, γράμματα και διακριτικά σημεία της ψηφιακής γραφής και αποτελούν μέρος του τομέα στήριξης μιας γλώσσας με λογισμικό προγραμματισμού για ψηφιακή επεξεργασία και ηλεκτρονική επικοινωνία. Στο σύνολό του αυτός ο τομέας απαρτίζεται και από άλλα στοιχεία, από τα οποία τα κυριότερα είναι η μεθόδευση εισαγωγής (input) στα ηλεκτρονικά μηχανήματα, απεικόνισης (rendering) και ορθογραφικού ελέγχου. Η εφαρμογή των ηλεκτρονικών προγραμμάτων για χρήση σε μια συγκεκριμένη γλώσσα και ένα εθνικό περιβάλλον περιλαμβάνει την μετατροπή του προγράμματος διασύνδεσης (interface) και την χρησιμοποίηση τοπικών, χρονικών, ημερολογιακών και άλλων συμβάσεων. Η πραγματοποίηση όμως όλων αυτών εξαρτάται από την διαμόρφωση ενός εύχρηστου κώδικα χαρακτήρων, πράγμα που αποτελεί το αντικείμενο της ομιλίας μου.

Μέχρι πρόσφατα όλες σχεδόν οι απεικονίσεις των χαρακτήρων μιας γλώσσας στα

ηλεκτρονικά συστήματα χρησιμοποιούσαν τον ίδιο κώδικα, ενσωματωμένο κατά κανόνα στον μηχανισμό του πληκτρολογίου και της οθόνης, όπως φαίνεται στην εικόνα 1. Έτσι, όταν ο χειριστής του υπολογιστή κτυπούσε το πλήκτρο του λατινικού γράμματος «d», το πληκτρολόγιο μετέδιδε τον κωδικό αριθμό 100, η εφαρμογή δεχόταν τον κωδικό 100 και μπορούσε να τον επανεμφανίσει, αναμεταδίδοντάς τον ως χαρακτήρα «d». Αυτή όμως η αρχιτεκτονική δόμηση καθήλωνε τον χειριστή σε μια μόνο πληκτρολογική διάταξη, μια επιλογή χαρακτήρων και μια μοναδική ορατή μορφή εκάστο χαρακτήρα.

Με την εξάπλωση της εικονοηφιδωτής οθόνης (bitmapped screen) η δόμηση αυτή πέρασε σε αχρηστία και οι περιορισμοί της ξεπεράστηκαν. Η πιο σύγχρονη προσέγγιση του θέματος παρουσιάζει την επεξεργασία χαρακτήρων με σειρά διαδικασιών, η καθεμία από τις οποίες διαθέτει δική της απεικόνιση (εικ. 2). Κατανοώντας τον ρόλο αυτών των διαφορετικών απεικονίσεων, θα αποσαφηνίσουμε και τον συγκεκριμένο ρόλο του κώδικα χαρακτήρων.

Κωδικοί πλήκτρων και μέθοδοι εισαγωγής
Κάθε πλήκτρο αντιπροσωπεύεται από έναν ιδιαίτερο κωδικό πλήκτρου (key code), ο οποίος δηλώνει μόνον την ειδική θέση του σ' ένα πληκτρολόγιο. Μ' αυτή την λογική, μολονότι πρόκειται για τον ίδιο χαρακτήρα «2», ο κωδικός του πλήκτρου του στο κανονικό διαφέρει από τον κωδικό του «2» στο ξεχωριστό αριθμητικό πληκτρολόγιο, όπως ακριβώς και ο κωδικός του αριστερού

* Οι απόψεις που διατυπώνονται ανήκουν στον ομιλητή και όχι απαραίτητα στην κοινοπραξία του Unicode, στο ISO ή άλλη οργάνωση.

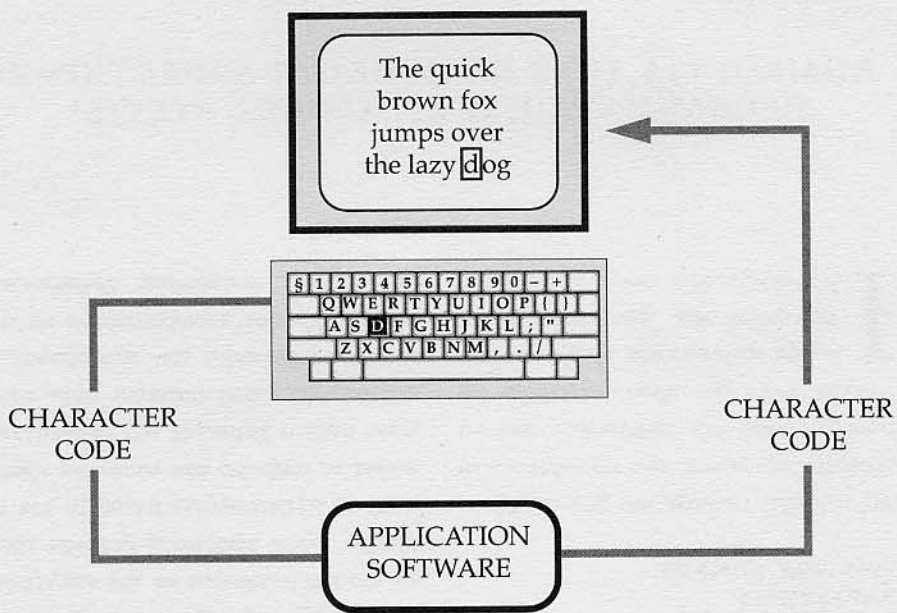


Fig. 1.

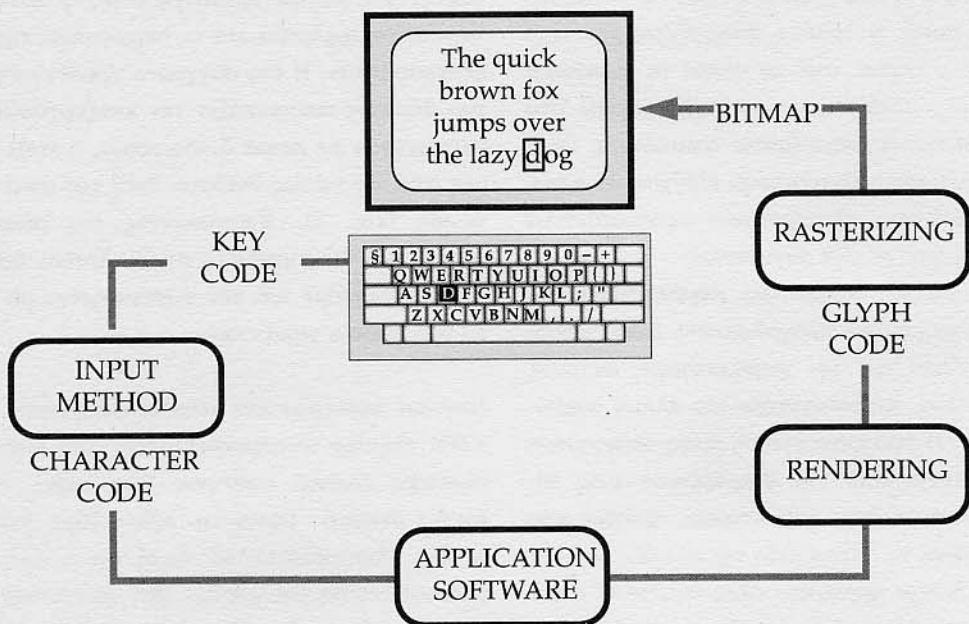


Fig. 2.

πλήκτρου «κεφαλαία» (shift) διαφέρει από εκείνον του δεξιού, ενώ η λειτουργική χρησιμότητα και των δύο είναι η ίδια. Αντίθετα, οι χαρακτήρες «b» και «B» ή «2» και «@» στο κανονικό πληκτρολόγιο δηλώνονται με τον ίδιο κωδικό, ενώ είναι διαφορετικοί. Η μεθόδευση εισαγωγής μετατρέπει τα σύνολα των κωδικών πλήκτρων σε κωδικούς χαρακτήρων. Έτσι, για παράδειγμα και χωρίς αυτός να είναι ο μοναδικός τρόπος για την απόδοση του χαρακτήρα «B», η πληκτρολογική δραστηριότητα με διαδοχικό κτύπημα του αριστερού πλήκτρου «κεφαλαία», του πλήκτρου «b» και, στην συνέχεια, της διαδοχικής απελευθέρωσής τους, δηλαδή, της απελευθέρωσης του πλήκτρου «b» και του αριστερού πλήκτρου «κεφαλαία», αποδίδει πάντοτε το «B» άσχετα από την σειρά προτεραιότητας των χειρισμών.

Οι μέθοδοι εισαγωγής χαρακτήρων με βάση τους κωδικούς πλήκτρων προσφέρουν μεγάλη χρηστική ευελιξία, όπως δείχνει η χρησιμοποίησή τους για την κινέζικη γλώσσα που περιλαμβάνει αμέτρητους χαρακτήρες κοινής χρήσης. Επειδή η κατασκευή πληκτρολογίων, όπου έκαστο πλήκτρο θα αντιστοιχούσε σ' έναν χαρακτήρα, είναι πρακτικά ανεφάρμοστη, χρησιμοποιούνται σειρές συνδυασμών κτυπημάτων σε πληκτρολόγια συμβατικού μεγέθους. Μερικές σχετικές μεθοδεύσεις αποδίδουν φωνητική αντιστοιχία στην εμφάνιση των χαρακτήρων, ενώ άλλες επεμβαίνουν στην δομή της ρίζας ή των κοντυλιών. Η μεθόδευση εισαγωγής μπορεί σε μερικές περιπτώσεις να αποδώσει καταλόγους εναλλακτικών επιλογών με την χρήση ενός πρόσθετου κτυπήματος πλήκτρου. Ο χειριστής μπορεί να επιλέξει μια μέθοδο εισαγωγής χωρίς μετατροπή του προγραμματισμού εφαρμογής.¹

Αν και το ελληνικό πληκτρολόγιο είναι τυποποιημένο, ένας χειριστής που έχει συνηθίσει να χρησιμοποιεί την ταξινόμηση QWERTY του λατινικού πληκτρολογίου, μπορεί να προτιμά διαφορετική διάταξη πλήκτρων. Με παρόμοιο τρόπο άλλοι χειριστές μπορεί να προτιμούν την πληκτρολόγηση τόνων πριν από τα αντίστοιχα γράμματα, πράγμα που θυμίζει τους περιορισμούς στον χειρισμό των παραδοσιακών γραφομηχανών, και άλλοι τον μετατονισμό, όπως γίνεται στα χειρόγραφα. Αυτή η πλευρά του θέματος δεν αφορά τον ηλεκτρονικό προγραμματισμό ή την κωδικοποίηση των χαρακτήρων, αφού αντιμετωπίζεται από την μέθοδο εισαγωγής.

Κωδικοί χαρακτήρων και ηλεκτρονικός προγραμματισμός εφαρμογών

Οι κωδικοί δηλώνουν μια αφηρημένη έννοια, τον «λογικό» χαρακτήρα, ο οποίος παραμένει ίδιος, άσχετα από τις πολλές και ποικίλες γραφικές απεικονίσεις του. Οι διάφορες δηλαδή γραμματοσειρές, τα σχέδια, τα μεγέθη και τα χρώματα του γράμματος «B» αποδίδουν λογικά τον χαρακτήρα «B», «B», «B», «xB» κ.ο.κ.

Οι κωδικοί των χαρακτήρων αντιπροσωπεύουν τα ουσιαστικά γλωσσικά στοιχεία της γραφής. Αν, για παράδειγμα, ένας χειριστής ηλεκτρονικού υπολογιστή αναζητεί την λέξη «τηλέφωνο» σ' ένα κείμενο, δεν ενδιαφέρεται αν θα την βρει γραμμένη με μικρούς, μεγάλους, έντονους ή πλάγιους χαρακτήρες. Οι εφαρμογές όμως απευθύνονται στους χαρακτήρες και όχι στις μεθόδους εισαγωγής ή της γραφικής εμφάνισής τους. Αυτό, βέβαια, δεν σημαίνει ότι οι εφαρμογές μπορούν να τους επεξεργαστούν, χωρίς να αναφερθούν στην γλώσσα, την οποία αντιπροσωπεύουν, αφού, όπως είναι γνωστό, το ίδιο αλφάβητο μπορεί με διαφορετικούς κανόνες να αποδίδει γραπτά περισσότερες από μια γλώσσες. Αλλά το θέμα αυτό υπερβαίνει τα όρια αυτής της ομιλίας.

1. Joseph Becker, «Multilingual Word Processing», *Scientific American*, 251, 7(1984), σελ. 96.

```

111111100
111111110
110000011
110000011
110000011
111111100
111111110
110000111
110000011
110000011
111111110
111111100

```



Εικ. 3.

Η πλήρης κωδικοποίηση ενός κειμένου εκτείνεται πέρα από τους χαρακτήρες. Συνήθως, κάθε χαρακτήρας φέρει το όνομα της γραμματοσειράς ή του σχεδιασμού του, τα κείμενα χωρίζονται σε παραγράφους, επικεφαλίδες και υποσημειώσεις. Υπάρχουν, επίσης, και πολλά άλλα δομικά στοιχεία. Όλες αυτές οι πληροφορίες μπορούν να χρησιμοποιηθούν με την βοήθεια ηλεκτρονικών γλωσσών, όπως η SGML,² ή άλλων, που η αποκλειστική χρήση τους προστατεύεται από τις εταιρείες παραγωγής ηλεκτρονικών προγραμμάτων, όπως το Framemaker ή το Word. Δεν αποτελούν, ωστόσο, μέρος της κωδικοποίησης χαρακτήρων.

Κωδικοί γλύφων και απεικόνιση

Απεικόνιση (rendering) είναι η διαδικασία μετατροπής των κωδικών χαρακτήρων σε εικόνες. Χαρακτηριστικά, ο απεικονιστικός μηχανισμός δέχεται ένα σύνολο χαρακτήρων, τις θέσεις και τις πληροφορίες για την γραμματοσειρά και τον σχεδιασμό της και μετατρέπει τους κωδικούς των χαρακτήρων σε κωδικούς γλύφων, οι οποίοι φέρουν τα ονόματα διαφόρων συγκεκριμένων σχημάτων. Η διαδικασία αυτή συχνά αναφέρεται ως «υποκατάσταση γλύφων» (glyph substitution).

Σε μερικές περιπτώσεις, όταν το σχέδιο ενός χαρακτήρα υπαγορεύεται από την

θέση που κατέχει σε μια λέξη, φράση ή ακόμη και ολόκληρη αράδα,³ χρησιμοποιούνται για αυτόν διαφορετικές γλύφοι ανάλογα με το περιεχόμενό του. Άλλοτε πάλι, ένα σύνολο χαρακτήρων μπορεί να απεικονιστεί ως ενιαία γλύφος, η οποία ονομάζεται σύμπλεγμα (ligature) στην τυπογραφία. Στην συμπλεγματική γραφή του λατινικού αλφαβήτου, το συνηθέστερο σύμπλεγμα είναι η ενοποίηση των γραμμάτων «f» και «i» σε «fi». Μπορεί ακόμη ένας χαρακτήρας να μην έχει αντίστοιχη γλύφο, αφού οι τόνοι στα κεφαλαία μερικών σχεδίων δεν εμφανίζονται. Εξάλλου, σε μερικές μορφές ελληνικής γραφής, το όμικρον ακολουθούμενο από το ύψιλον σχηματίζει το είδος του τυπογραφικού συμπλέγματος «ϝ». Αντίθετα, ένας μοναδικός «λογικός» χαρακτήρας είναι δυνατό να απεικονίζεται από περισσότερες από μια γλύφους, όπως συμβαίνει στην περίπτωση, όπου το σύμπλεγμα της λατινικής γραφής «æ», το οποίο απεικονίζεται με τις δύο διαδοχικές γλύφους «a» και «e» χωρίς εσωτερική διαπλοκή, όταν η γραμματοσειρά που χρησιμοποιείται δεν διαθέτει την συμπλεγματική γλύφο. Στην προκειμένη περίπτωση πρέπει να σημειωθεί η διαφορά μεταξύ των συμπλεγμάτων «æ» και «fi». Το πρώτο αντιστοιχεί ορθογραφικά στην ελλη-

2. Charles F. Goldfarb, *The SGML Handbook*, Oxford 1990.

3. Για την ικανοποιητική στοίχιση της αραβικής γραφής δεν προστίθενται διαστήματα μεταξύ των λέξεων ή των γραμμάτων, αλλά επιμηκύνονται μερικά γράμματα —συνήθως το τελευταίο μιας αράδας.

νική δίφθογγο «αι» —όπως στις λέξεις *aesthete* ή *encyclopaedia*— και, βέβαια, δεν μπορεί να αποδώσει την συμπτωματική διαδοχή των χωριστών φθόγγων «a» και «e» —όπως στις λέξεις «aerobic» ή «paella». Το δεύτερο είναι καθαρά προαιρετική αισθητική τυπογραφική επιλογή για την διάπληξη του «f» και του «i».

Σε μερικές γλώσσες τέτοιες παραλλαγές είναι απαραίτητες και υποχρεωτικές. Στην αραβική γραφή η ενιαία απόδοση των λέξεων είναι κανόνας και η μορφή ενός εκάστου γράμματος υπαγορεύεται από το προηγούμενο και το επόμενο του, μολονότι η ηλεκτρονική ή συμβατική αραβική τυπογραφία δεν βρίσκεται σε υψηλό επίπεδο, καθώς προβλέπει μόνον την διάκριση της αρχικής, μέσης και τελικής θέσης των στοιχείων. Αλλά και τα ελληνικά περιέχουν περιστασιακές παραλλαγές, όπως αυτή του αρχικού ή μέσου πεζού «σ» και του τελικού «ς» ή του αρχικού «β» και του μέσου πεζού «β» ορισμένων γραμματοσειρών. Ακόμη, μερικές γραμματοσειρές περιέχουν πολυάριθμα προαιρετικά συμπλέγματα. Για την ποιοτικά ανώτερη απεικόνιση των ελληνικών στους κωδικούς χαρακτήρων, που απεικονίζουν τα διακριτικά σύμβολα χωριστά, θα έπρεπε ο συνδυασμός του βασικού χαρακτήρα και των διακριτικών του να απεικονίζεται ως σχεδιαστικό σύμπλεγμα με τον τρόπο που περιγράφεται παραπάνω, δηλαδή οι σχεδιαστές τυπογραφικών στοιχείων θα έπρεπε να σχεδιάζουν ειδικές γλύφους για κάθε συνδυασμό. Εδώ πρέπει να τονιστεί ότι τέτοια συμπλέγματα χρειάζεται να περιλαμβάνονται στο σύνολο των γλύφων και όχι απαραίτητα των χαρακτήρων.

Τέλος, ένα κείμενο είναι δυνατό να απεικονίζεται με πολλαπλούς ορθούς τρόπους, διότι εκτός από τις γραμματοσειρές και τα μεγέθη τους μπορούν να τροποποιούνται και τα συμπλέγματα, καθώς και άλλες σχεδιαστικές επιλογές. Η λεγόμενη «μετα-

γλωττιστική απεικόνιση» (transliteration rendering) μπορεί να είναι πραγματικά χρήσιμη διαδικασία για ποικίλα κειμενογραφικά περιεχόμενα, όπως την ανάγνωση μιας εβραϊκής λέξης με λατινικούς χαρακτήρες, για παράδειγμα, σε περίπτωση που δεν είναι γνωστοί οι εβραϊκοί.

Ψηφιακή απεικόνιση οθόνης (bitmaps) και σάρωση (rasterizing)

Ο κωδικός μιας γλύφου κάνει εφικτή την αναγνώριση του σχεδίου της, αλλά, για να γίνει αυτή εκτυπώσιμη ή για να εμφανιστεί στην οθόνη του υπολογιστή, πρέπει προηγουμένως να μετατραπεί σ' ένα μωσαϊκό εικονοψηφίδων (bitmap). Για παράδειγμα, ο μαύρος κεφαλαίος χαρακτήρας «B» των 7 στιγμών της γραμματοσειράς Helvetica σχηματίζεται ως σύνολο δυαδικών ψηφίων

11111110	01111111	1011000	01111000
00111100	00011111	11110011	11111101
10000111	11000001	11100000	11111111
11011111	1100	(9×12).	

Το σχήμα αυτό γίνεται αντιληπτό ευκολότερα, αν παραταχθεί σ' ένα ορθογώνιο, αποτελούμενο από 0 και 1 ή, ακόμη καλύτερα, με τα αντίστοιχα μαύρα και άσπρα εικονοστοιχεία, όπως δείχνει η εικόνα 3.

ΣΧΕΔΙΑΣΜΟΣ ΚΩΔΙΚΩΝ ΧΑΡΑΚΤΗΡΩΝ

Οι κωδικοί χαρακτήρων διέπονται από ορισμένους απλούς κανόνες:

α'. Η απεικόνισή τους πρέπει να είναι μονοσήμαντη, δηλαδή να αντιστοιχεί ένας κωδικός σε κάθε χαρακτήρα και ένας χαρακτήρας σε κάθε κωδικό. Η εφαρμογή αυτού του κανόνα ενδέχεται να είναι περίπλοκη και, μάλιστα, σε γλώσσες που η ιστορική εξέλιξή τους περιέλαβε στάδια, κατά τα οποία το αλφάβητό τους χρησιμοποιήθηκε με ποικίλους τρόπους. Στα κλασσικά λατινικά, για παράδειγμα, δεν υπήρχε διαφοροποίηση ανάμεσα στο I και το J ή στο U και το V.

β'. Ο χώρος τους πρέπει να διατίθεται με οικονομία. Παρόλο το κατερχόμενο κόστος της μνήμης των ηλεκτρονικών υπολογιστών και της αποτελεσματικής συμπίεσης των δεδομένων, δεν πρέπει να γίνεται σπατάλη ψηφιοστοιχείων.

γ'. Για να είναι εύκολη η χρήση τους κατά την διενέργεια υπολογισμών, οι κωδικοί χαρακτήρων πρέπει να είναι ισομεγέθεις ως προς τον αριθμό δυαδικών ψηφίων —αφού ο σύγχρονος τεχνικός εξοπλισμός λειτουργεί αποδοτικότερα με στοιχεία 8, 16, 32 ή 64 δυαδικών ψηφίων.

δ'. Η μεθόδευση της απεικόνισής τους πρέπει να είναι απλή και συνεπής ως προς τον καθορισμό, για παράδειγμα, ενός χαρακτήρα, σε αντιδιαστολή με μια παραλλαγή του και την τοποθέτηση μη γραμμικών στοιχείων —και εδώ τίθεται το ερώτημα κατά πόσον θα έπρεπε τα ενδεχομένως υπάρχοντα διακριτικά να εμφανίζονται πριν ή μετά από τον χαρακτήρα του γράμματος, στον οποίο αντιστοιχούν.

Οι παλαιότερες τεχνολογίες έδιναν πρωταρχική σημασία στην εξοικονόμηση χώρου και η κωδικοποίηση χαρακτήρων κάλυπτε μόνο τους απλούς λατινικούς. Με ταχύτητες μετάδοσης των 110 baud ή και χαμηλότερες, κάθε δυαδικό ψηφίο ήταν υπολογίσιμο και γι' αυτό τον λόγο τα παλαιά τηλέτυπα χρησιμοποιούσαν 5 από αυτά κατά χαρακτήρα, επιτρέποντας έτσι 32 ξεχωριστούς κωδικούς, πράγμα που δεν επαρκούσε για το σύνολο των πεζών και των κεφαλαίων, για να μην αναφέρουμε τα αριθμητικά. Το πρόβλημα αντιμετωπιζόταν τότε με την βοήθεια μηχανισμού εναλλαγής της κατάστασης πληκτρολογίου μεταξύ πεζών και κεφαλαίων (shift). Ωστόσο, τέτοιοι μηχανισμοί δυσκόλευαν την επεξεργασία μιας ακολουθίας χαρακτήρων, καθώς το σχετικό πρόγραμμα έπρεπε να σαρώνει ολόκληρο το κείμενο από την αρχή κάθε φορά. Βέβαια, στην εποχή του τηλέτυπου οι κωδικοί χρησι-

μοποιούνταν σχεδόν αποκλειστικά για μεταδόσεις και σπάνια για επεξεργασίες. Αργότερα, οι κωδικοί των 6 δυαδικών ψηφίων επέτρεψαν την χρήση 64 χαρακτήρων, περιλαμβανομένων του λατινικού αλφαβήτου, των αριθμητικών και μερικών χαρακτήρων γραφικών, αλλά όχι τα πεζά γράμματα. Τελικά, αναπτύχθηκαν κώδικες των 7 και 8 δυαδικών ψηφίων, οι οποίοι καλύπτουν τις περισσότερες (αλλά όχι όλες) της ανάγκες των αγγλικών. Ιδιαίτερα ο κώδικας ASCII, με 7 δυαδικά ψηφία και 96 γραφικούς χαρακτήρες,⁴ αποδείχθηκε πολύ ικανοποιητικός. Αν και σήμερα ο κώδικας ASCII έχει 8 δυαδικά ψηφία, το όγδοο δεν χρησιμοποιείται.

Η κωδικοποίηση μη λατινικών γραφών

Εκτός από τα αγγλικά, οι άλλες γλώσσες απαιτούν περισσότερους χαρακτήρες από αυτούς που διαθέτει ο ατονικός κώδικας ASCII. Για την πλειονότητα των λατινογενών γλωσσών δεν είναι πολυάριθμοι οι πρόσθετοι αναγκαίοι χαρακτήρες. Προσωρινά το πράγμα αντιμετωπίστηκε από τον International Standards Organization (ISO), με τον κώδικα ISO 646,⁵ ο οποίος υποκαθιστούσε μερικούς «εθνικούς» χαρακτήρες με χαρακτήρες ASCII.

Χωρίς να επεκταθούμε στο ιστορικό αυτής της υπόθεσης, αρκεί να αναφερθεί ότι όλες οι κύριες γλώσσες της Δυτικής Ευρώπης μπορούν πλέον να κωδικοποιη-

4. American National Standards Institute, *Coded Character Set—7-Bit American Standard Code for Information Interchange* και το EBCDIC: *Extended Binary Coded Decimal Interchange Code*, ένα εξειδικευμένο πρότυπο που σχεδιάστηκε στην IBM και παλαιότερα το χρησιμοποιούσαν ευρέως σε κεντρικές μονάδες μηχανογράφησης.

5. ISO 646, International Organization for Standardization, *International Standard – Information Processing – ISO 7-bit coded character set for information interchange*, ISO 646, 1991. Αυτό και άλλα ISO πρότυπα σύνολα χαρακτήρων υποστηρίζονται από την IEC JTC1/SC2. Πληροφορίες για τα πρότυπα ISO στην ηλεκτρονική διεύθυνση: <http://www.iso.ch/>.

θούν στο ενιαίο πρότυπο ISO 8859-1 με τα 8 δυαδικά ψηφία του, το οποίο ονομάζεται συνήθως Latin-1. Άλλες λατινογενείς γραφές καλύπτονται από τον πίνακα κωδικοποίησης Latin-2 ή ISO 8859-2 και άλλους της σειράς 8859.⁶ Το πρότυπο Latin-1 παρουσιάζει το πλεονέκτημα ότι είναι ικανό να απεικονίζει ακριβώς οποιοδήποτε κείμενο μπορεί να απεικονιστεί στο αυστηρά στερεότυπο υποσύνολο του πίνακα ASCII.

Και οι μη λατινογενείς γραφές μπορούν να επωφελούνται από το ίδιο τέχνασμα της επέκτασης του ASCII. Έτσι, η ελληνική μονοτονική γραφή εξυπηρετείται με τον κώδικα ISO 8859-7, γνωστό και ως ELOT-928, Greek-8 ή ECMA-118. Για τα πολυτονικά ελληνικά όμως δεν υπάρχει αρκετός χώρος. Τα κυριλλικά, τα αραβικά και τα εβραϊκά εξυπηρετούνται κι αυτά από την σειρά 8859: εξάλλου, υπάρχουν και κώδικες των 8 δυαδικών ψηφίων, οι οποίοι δεν είναι υπερσύνολα του ASCII. Για ιδιαίτερες εφαρμογές συχνά είναι αρκετό να καθορίσουμε επακριβείς κανόνες μεταγλώττισης, όπως του Κώδικα Βήτα (Beta Code) του *Thesaurus Linguae Graecae*.⁷

Οι γλώσσες της Ανατολικής Ασίας, όπως η κινεζική, η ιαπωνική και η κορεατική, έχουν ασύγκριτα μεγαλύτερα σύνολα χαρακτήρων, τα οποία συχνά περιέχουν πολλές δεκάδες χιλιάδες, χωρίς να περιλαμβάνουν τα «ἄπαξ λεγόμενα» και τις διά-

φορες σχεδιαστικές ποικιλίες. Φυσικά, δεν είναι δυνατό να χωρέσουν στους κώδικες των 8 δυαδικών ψηφίων και γι' αυτό τον λόγο η κωδικοποίησή τους έχει εφαρμόσει διάφορες τεχνικές, οι οποίες όμως δεν είναι καθόλου εύχρηστες.

Ο ΧΕΙΡΙΣΜΟΣ ΠΟΛΛΑΠΛΩΝ ΓΡΑΦΩΝ

Η μέχρι τώρα συζήτηση αναφερόταν στην απεικόνιση μιας μοναδικής γραφής κατά περίπτωση, πράγμα που ανταποκρίνεται στις ανάγκες της προσωπικής ή εμπορικής αλληλογραφίας, αλλά όχι στο διεθνές περιβάλλον, όπου τα προβλήματα αφθονούν.

Για να γίνει σαφής η κατάσταση αυτή, πρέπει να υποθέσουμε την μετάδοση ενός μηνύματος από μια χώρα σε μιαν άλλη. Για να διαβαστεί το μήνυμα, ο παραλήπτης του πρέπει να γνωρίζει τον κώδικα που χρησιμοποιήθηκε για την μετάδοση, καθώς επίσης να διαθέτει κάποιον τρόπο μετατροπής αυτού του κώδικα σε γλύφους ή σε μεταγλωττίσεις, σε ορισμένες περιπτώσεις. Ακόμη κι αυτό το στοιχειώδες λειτουργικό επίπεδο, το οποίο αναφέρεται και ως πολυεντοπισμός (multi-localization), δεν έχει ευρεία υποστήριξη.

Ένα άλλο πρόβλημα προκύπτει από το γεγονός ότι πολλά κείμενα περιέχουν πολλές γλώσσες. Σε μερικές περιπτώσεις η χρήση κωδικών 8 δυαδικών ψηφίων εξασφαλίζει εύκολη λύση του προβλήματος, αφού όλες οι γραφές της Δυτικής Ευρώπης καλύπτονται από το πρότυπο Latin-1, ώστε να είναι δυνατή η συνύπαρξη του γερμανικού διακριτικού Umlaut (¨) και της γαλλικής βαρείας (´) (accent grave), και να μπορούν τα ελληνικά και τα αγγλικά να καλύπτονται από το πρότυπο Greek-8, ώστε να είναι δυνατή η συνύπαρξη του γράμματος γ και του q. Ωστόσο, κανένας κώδικας δεν καλύπτει συγχρόνως τα γαλλικά και τα ελληνικά, με αποτέλεσμα να είναι αδύνατη η συνύπαρξη της γαλλικής

6. ISO 8885. International Organization for Standardization, 8-bit Single-Byte Coded Graphic Character Sets - Part 1: Latin Alphabet No. 1, ISO 8859-1: 1987. Part 2: Latin alphabet No. 2, ISO 8859-2, 1987. Part 3: Latin alphabet No. 3, ISO 8859-3, 1988. Part 4: Latin alphabet No. 4, 1988. Part 5: Latin/Cyrillic alphabet, ISO 8859-5, 1988. Part 6: Latin/Arabic alphabet, ISO 8859-6, 1987. Part 7: Latin/Greek alphabet, ISO 8859-7, 1987. Part 8: Latin/Hebrew alphabet, ISO 8859-8, 1988. Part 9: Latin alphabet No. 5, ISO 8859-9, 1980.

7. «Beta Code Manual», Thesaurus Linguae Graecae Project, διατίθεται από την ηλεκτρονική διεύθυνση: [gopher://tlg.cwis.uci.edu:7011/oo/Beta%20Code%20Summary](http://tlg.cwis.uci.edu:7011/oo/Beta%20Code%20Summary).

βαρείας και του γράμματος γ στο ίδιο κείμενο, όταν χρησιμοποιείται ενιαίος κώδικας. Υπάρχουν, πάντως, πολλοί τρόποι αντιμετώπισης αυτού του προβλήματος.

Επισφαλείς μεθοδεύσεις

Ενας απλός και πρόχειρος, αλλά μή αυτοδύναμος τρόπος επεξεργασίας των πολύγλωσσων κειμένων —ή και προσαρμογή μιας αγγλόφωνης λογισμικής εφαρμογής σε διαφορετική γραφή— είναι η χρήση του μηχανισμού «γραμματοσειρά», που υπάρχει σε πολλές εφαρμογές. Αυτός υλοποιείται με την απλή αντικατάσταση των λατινικών γλύφων με μη λατινικούς, ενώ το λογισμικό επεξεργάζεται τους κωδικούς σαν να εξακολουθούσαν να απεικονίζουν λατινικά γράμματα. Οι κειμενογράφοι επιτρέπουν στον χειριστή να επιλέγει όποια γραμματοσειρά θέλει, για οποιονδήποτε χαρακτήρα ή ακολουθία χαρακτήρων, ώστε μια γραμματοσειρά που εμφανίζει το ελληνικό γράμμα «γ» αντί του λατινικού «g», να δίνει την εντύπωση ότι μπορεί να αποδώσει στα ελληνικά.⁸

Ομως, η λύση αυτή είναι κακή για πολλούς λόγους. Δεν υπάρχει φυσιολογικός τρόπος εισαγωγής των τόνων ή, κατά την ηλεκτρονική ορολογία, δεν υπάρχουν κατάλληλες μέθοδοι εισαγωγής. Καθώς δεν υπάρχουν και οι κατάλληλες μέθοδοι απεικόνι-

σης, δεν είναι δυνατό να συμπεριληφθούν κανόνες αυτής για ιδιαίτερες γλωσσικές προδιαγραφές (π.χ. του τελικού «ς» στα ελληνικά). Ο συλλαβισμός είναι ανακριβής όχι εξαιτίας του κώδικα χαρακτήρων αλλά επειδή η εφαρμογή αδυνατεί να διαπιστώσει το γεγονός ότι οι χρησιμοποιούμενοι χαρακτήρες αποδίδουν την ελληνική γραφή. Η εκχώρηση χαρακτήρων είναι συναρτημένη με την γραμματοσειρά, η οποία δεν διέπεται συνήθως από λογική και δεν είναι εύκολο να αλλάξει. Η κωδικοποίηση εξαρτάται ουσιαστικά από το γεγονός πως περιέχεται σε μια εφαρμογή, όπως ένας κειμενογράφος, ο οποίος επιτρέπει την επιλογή γραμματοσειρών. Οι καταχωρίσεις βάσης δεδομένων, τα πεδία λογιστικών φύλλων και τα ονόματα αρχείων δεν είναι, κατά πάσα πιθανότητα, σε θέση να χρησιμοποιήσουν αυτές τις γραμματοσειρές. Ακόμη, καθώς πρόκειται για ειδική μεθόδευση, η λειτουργικότητα της κωδικοποίησης εξαρτάται από το αν ο μεταδότης και ο λήπτης χρησιμοποιούν τον ίδιο κειμενογράφο και τις ίδιες γραμματοσειρές.

Με λίγα λόγια, η μεθόδευση αυτή δεν είναι ικανοποιητική. Αλλά, καθώς είναι εύκολη η εφαρμογή της και επειδή δεν προσφέρονται άλλες αποδοτικότερες, χρησιμοποιείται ευρύτατα.

Αλλαγή κώδικα

Η αξιοποίηση των πολλαπλών υπαρχόντων κωδίκων —ενός για κάθε είδους γραφή— σε συνδυασμό με την πρόβλεψη για κάποια τυποποιημένη εναλλαγή μεταξύ τους, αποτελεί αποτελεσματικότερο τρόπο ενέργειας από την χρήση των αποκλειστικών μεθόδων των διαφόρων κειμενογράφων. Αλλωστε, το υπάρχον διεθνές πρότυπο ISO 2022⁹ προσφέρει αυτήν ακριβώς την δυνατότητα.

9. ISO 2022. International Organization for Standardization, *ISO 7-bit coded character sets – Code extension techniques*, ISO 2022, 1986.

8. Για πληροφορίες σχετικά με παρόμοιες προχειρές λύσεις βλ. “Soc.Culture.Greek FAQ—Technical Information” στην ηλεκτρονική διεύθυνση: <http://www.cis.ohio-state.edu/hypertext/faq/usenet/greek-faq/technical/faq.html>, ή Frequently requested Mac Fonts—Greek Fonts” at: http://jasper.ora.com/comp.fonts/FAQ/cf_39.ht, ή και “Electronic Resources for Classicists: The Second Generation,” by Maria C. Pantelia at: <http://www.circe.unh.edu/classics/resources.html>. Επίσης, πολλές μεθοδεύσεις αναλύονται στα άρθρα του «Offline» στο *Religious Study News* του Σεπτεμβρίου του 1994, τα οποία αναδημοσιεύτηκαν στο ηλεκτρονικό περιοδικό *Offline* 46 και βρίσκονται και στην ηλεκτρονική διεύθυνση: <http://scholar.cc.emory.edu/scripts/societies/RSN/Offline/off46.html>.

Η προσέγγιση αυτή θα μπορούσε να αποδώσει, αλλά η εφαρμογή της παρουσιάζει επιπλοκές. Για να προσδιοριστεί ποιος χαρακτήρας απεικονίζεται από έναν κωδικό, το μήνυμα πρέπει να υποβληθεί σε επεξεργασία ευθύς εξ αρχής, ώστε να ερευνηθούν οι κωδικοί για κώδικες εναλλαγής. Επίσης, δεδομένου ότι οι σηματοδότες του χαρακτηρισμού κωδικοποίησης πρέπει να διατηρηθούν, είναι το ίδιο περίπλοκη η λεπτή διαδικασία της κοπής ενός τμήματος κειμένου από ένα σημείο και η προσάρτησή του σ' ένα άλλο. Αλλωστε, ένας μοναδικός λογικός χαρακτήρας ενδέχεται να έχει πολλαπλές απεικονίσεις στο ίδιο κείμενο, αφού τίποτε δεν εμποδίζει την διαφορετική κωδικοποίηση του ίδιου χαρακτήρα σε διαφορετικούς κώδικες. Τέλος, αντίθετα από την προηγούμενη μεθόδευση, οι εφαρμογές του προτύπου ISO 2022 είναι λίγες και δεν προβλέπεται να αυξηθούν πολύ.

Ο σύνθετος κώδικας DIS 10646

Τα προβλήματα που παρουσιάζουν οι διάφορες μεθοδεύσεις οδήγησαν τους ενδιαφερόμενους για τα διεθνή πρότυπα και τυποποιήσεις να δεχθούν την πρόκληση για την δημιουργία ενός ενιαίου προτύπου κωδικοποίησης, το οποίο θα είχε το τεράστιο μέγεθος που απαιτείται για την ενσωμάτωση όλων των ήδη υπαρχόντων —συγκεκριμένα, θα έπρεπε να διαθέτει κώδικα 32 δυαδικών ψηφίων.

Η κωδικοποίηση αυτή ήταν το πρότυπο International Draft Standard DIS 10646 και αντιπροσώπευε ένα παράδειγμα της λειτουργίας διεθνών προτύπων προς αποφυγήν. Αποτελούσε πρόταση πολιτικής λύσης σ' ένα τεχνικό πρόβλημα, αντί να επιδιώκει την εξεύρεση τεχνικής. Η πολιτική λύση περιλάμβανε την διαίρεση του συνολικού κωδικού χώρου σε περιοχές εθνικών ομάδων, οι οποίες θα διαχειρίζονταν τον χώρο που θα τους αναλογούσε, με όποιον

τρόπο επέλεγαν, ενσωματώνοντας σ' αυτόν τα ποικίλα υπάρχοντα εθνικά πρότυπα. Έτσι, το κόμμα θα μπορούσε να απεικονίζεται από δεκάδες διαφορετικά πρότυπα —όπως το κόμμα των Η.Π.Α., το κυριλλικό, το αραβικό, κ.τ.λ.— χωρίς να διαφοροποιείται σε καμία περίπτωση το σχέδιο και η σημασία αυτού του συμβόλου. Από την άλλη πλευρά, θα ήταν πιθανό να παραμείνουν έξω από την κωδικοποίηση αυτή μερικά συστήματα γραφής, επειδή κανένα κράτος δεν θα τύχαινε να αναλάβει την επικύρωση της νομιμότητάς τους ή να επισημοποιήσει ρητά την κατάργηση κάποιου απαρχαιωμένου τομέα ενός συστήματος γραφής.

Με τον σκοπό να διευκολύνεται η επεξεργασία χαρακτήρων, το πρότυπο DIS 10646 υποστήριζε από τεχνική άποψη μόνον προσυνδυασμένους (precomposed) χαρακτήρες, όπως, για παράδειγμα, το δασυνόμενο άλφα «ά», και δεν διέθετε κινητά διακριτικά (floating diacritics), όπως η ανεξάρτητη από το γράμμα δασεία «'». Το αποτέλεσμα ήταν στην πραγματικότητα πως οι συνδυασμοί που δεν είχαν προβλεφθεί, δεν ήταν απλώς δύσκολο, αλλά αδύνατο να απεικονιστούν.

Ακόμη, το πρότυπο DIS 10646 δεν διέκρινε με συνέπεια τους χαρακτήρες από τις γλύφους, πράγμα που σημαίνει ότι είχε περιληφθεί σ' αυτό μεγάλος αριθμός σχεδίων ειδικής μορφής (presentation forms), δηλαδή χαρακτήρων, που είναι στην ουσία κωδικοί γλύφων, όπως οι συμφραστικές (contextual) παραλλαγές των αραβικών γραμμάτων.

Ασφαλώς, ο κώδικας αποκλειστικά προσυνδυασμένων χαρακτήρων των 32 δυαδικών ψηφίων υποσχόταν την επίλυση του περίπλοκου προβλήματος που αντιμετώπιζε ο κάθε «κολλών τε καὶ ἀφαιρών» συντάκτης κειμένων, αλλά με τίμημα την παρὰβίαση σημαντικών σχεδιαστικών αρχών, όπως καταγράφονται παρακάτω:

α'. Δεν θα απεικονίζε χαρακτήρες μονοσήμαντα, αφού τα διάφορα εθνικά πρό-

τυπα περιλαμβάνουν συχνά τους ίδιους χαρακτήρες και επειδή οι ειδικές μορφές συγχέουν την ταυτότητα των χαρακτήρων.

β'. Θα κατασπαταλούσε χώρο στον κώδικα, καθώς τα ταυτόσημα στοιχεία του συστήματος της κινέζικης, της ιαπωνικής και της κορεατικής γραφής θα επαναλαμβάνονταν σε τρεις εθνικούς χώρους, ενώ, άσχετα από την διαφοροποιημένη εννοιολογική τους σημασία, είναι τα ίδια και στις τρεις γλώσσες και, από την άλλη πλευρά, αμέτρητοι προσυνδυασμένοι χαρακτήρες θα διεκδικούσαν χώρο δυσανάλογα μεγάλο προς την συχνότητα εμφάνισής τους.

γ'. Τόσο οι πολλαπλές απεικονίσεις ενός χαρακτήρα, όσο και η απουσία μιας συνεπούς αντιμετώπισης πραγμάτων, θα καθιστούσαν την επεξεργασία του κώδικα δυσχερή.

δ'. Η ενιαία και συνεπής αντιμετώπιση πραγμάτων θα αποκλειόταν από το γεγονός ότι οι εθνικές ομάδες θα όριζαν εννοιολογικά την σημασία των χαρακτήρων τους.

Πρέπει να σημειωθεί, επίσης, ότι ορισμένες πολύ συζητήσιμες παραχωρήσεις δέσμευαν το πρότυπο DIS 10646 στην χρήση αναχρονιστικών ηλεκτρονικών υπολογιστών, αλλά δεν χρειάζεται να επεκταθούμε σε αυτό το θέμα.

Η λύση Unicode

Μια διαφορετική προσέγγιση προτάθηκε από τον οργανισμό με το όνομα Unicode,¹⁰ ο οποίος επιστράτευσε τεράστιο δυναμικό εμπειρογνομόνων στο θέμα επεξεργασίας

πολύγλωσσων κειμένων από εταιρείες, όπως οι Xerox, Apple, NeXT, Adobe, Ecological Linguistics και Electronic Book Technologies. Μεταξύ των υποστηρικτών του Unicode περιλαμβάνονταν ηγετικές εταιρείες σχεδιασμού και κατασκευής τεχνικού και λογισμικού υλικού, όπως οι IBM, Sun, HP, DEC, Unisys, Compaq, Microsoft, Novell, Lotus, Sybase και Taligent. Συμπεριλάμβανε και επώνυμους προηγμένους χρήστες επεξεργαστών πολύγλωσσων κειμένων, όπως η ομάδα Research Library Group (μια κοινπραξία των σημαντικών βιβλιοθηκών ερεύνης των Η.Π.Α.), το ίδρυμα Getty Art History Foundation, η βιβλιοθήκη Royal Library της Βρεταννίας και το πρακτορείο Reuters. Αν πρέπει να σημειωθεί κάποια θεμελιακή αδυναμία αυτής της οργάνωσης, μπορούμε να αναφέρουμε την ανεπαρκή εκπροσώπηση κρατών εκτός των Η.Π.Α., γεγονός ουσιαστικής σημασίας για μια προσπάθεια διεθνούς βεληνεκού.

Το πρότυπο της Unicode βασίστηκε στο παρακάτω σύνολο σχεδιαστικών αρχών:¹¹

α'. Όλοι οι χαρακτήρες έχουν πλάτος 16 δυαδικών ψηφίων και το σύνολο των 65.536 κωδικών μπορεί να χρησιμοποιηθεί για κωδικοποίηση χαρακτήρων χωρίς παραχωρήσεις σε απαρχαιωμένους τύπους μηχανμάτων.

β'. Οι κωδικοί χαρακτήρων αντιπροσωπεύουν λογικούς, αφηρημένους χαρακτήρες και όχι συγκεκριμένα γραφικά σχέδια (γλύφους).

γ'. Οι γλύφοι, οι οποίες σχηματίζονται από πολλαπλά λογικά στοιχεία —όπως γράμματα με διακριτικά ή διγράμματα, όπως το «ch», το οποίο θεωρείται ως ενιαίο γράμμα στο ισπανικό αλφάβητο— απεικονίζονται ως πολλαπλοί χαρακτήρες.

10. *The Unicode Standard: Worldwide Character Encoding*, Version 1.0, Volume 1, Addison-Wesley, 1990. Το Unicode Consortium έχει επίσης την ηλεκτρονική διεύθυνση <http://www.unicode.org> με συνεχώς ενημερωμένη πληροφόρηση. Η διεύθυνση του Unicode Consortium είναι P.O. Box 700519, San Jose, CA 95170-0519, USA. Επίσης, Jürgen Bettels, F. Avery Bishop, "Unicode: A Universal Character Code," *Digital Technical Journal*, 5 (Καλοκαίρι 1993), σελ. 3.

11. Αυτές οι αρχές διατυπώνονται στο κείμενο Unicode 1.0 του Δεκεμβρίου 1990, αλλά όχι σε νεότερες εκδόσεις του προτύπου της ISO.

δ'. Οι χαρακτήρες είναι διεθνείς και δεν ανήκουν σε συγκεκριμένες γλώσσες. Οι αποφάσεις για κωδικοποιήσεις βασίζονται σε τεχνικά και όχι πολιτικά κριτήρια. Ιδιαίτερα τα θέματα απεικόνισης, όπως οι εθνικές επιλογές γραμματοσειρών ή καλλιγραφικών αποδόσεων, τοποθετούνται έξω από την αρμοδιότητα του Unicode.

ε'. Οποια κι αν είναι η εθνική ιθαγένεια των χαρακτήρων ιδίου σχήματος και λειτουργικότητας, η απεικόνισή τους είναι μία και ενιαία —η διαδικασία αυτή ονομάζεται ενοποίηση (unification) και έχει ιδιαίτερη σημασία για τους χαρακτήρες των γλωσσών της Ανατολικής Ασίας. Έτσι, αφιερώνεται μόνον ένας κωδικός σε κάθε στερεότυπο και ίδιας χρήσης σημείο στίξης, διακριτικό ή γράμμα, ασχέτως του πόσες διαφορετικές γλώσσες το περιέχουν.

στ'. Οι κωδικοί χαρακτήρων αποτελούν στοιχεία κειμένου και όχι οδηγίες κειμενογραφίας. Διεργασίες, όπως η απεικόνιση, διαλογή, συμπίεση και μετάδοση δεν καλύπτονται από το Unicode.

ζ'. Η ακρίβεια για τις μετατροπές στις διάφορες κωδικοποιήσεις χαρακτήρων υποστηρίζεται από το Unicode, αλλά οι σχετικές διευκολύνσεις έρχονται σε δεύτερη μοίρα σε σχέση με άλλες σχεδιαστικές αρχές.

Μία ακόμη αρχή που δεν διατυπώνεται ρητά, αλλά εννοείται σαφώς στο σκεπτικό του προτύπου Unicode, είναι ότι οι κωδικοί χαρακτήρων πρέπει να περιλαμβάνουν τις γραπτές γλώσσες ως ιστορικά και τρέχοντα δεδομένα, και όχι ως πολιτικά ή λογοτεχνικά θέματα. Μολονότι, για παράδειγμα, τα κουρδικά δεν αναγνωρίζονται ως εθνική γλώσσα κυρίαρχου κράτους, αποτελούν υπάρχον σύστημα γραφής και δικαιούνται να έχουν κωδικούς των χαρακτήρων τους, όπως οι άλλες επίσημες εθνικές γλώσσες. Εξάλλου, αν και ως αποτέλεσμα των μεταρρυθμίσεων των αρχών του αιώνα έχει καταργηθεί ή περιπέσει σε αχρηστία ένας

αριθμός γραμμάτων του ρωσικού αλφαβήτου, η κωδικοποίησή τους κρίνεται απαραίτητη, ώστε να εμπλουτίζουν κείμενα ιστορικού περιεχομένου.

Ασφαλώς, η παραπάνω αρχή διαφέρει από την προσέγγιση της αρμόδιας επιτροπής για το πρότυπο 10646, η οποία είναι προσανατολισμένη προς την εξυπηρέτηση των γλωσσικών αναγκών των διαφόρων εθνικών κυβερνήσεων και όχι των ίδιων των γλωσσών. Και εδώ πρέπει να παρατηρηθεί ότι, γενικά, οι κυβερνήσεις κρατών δεν έχουν επιδείξει ζωτικό ενδιαφέρον για ιστορικά θέματα, για γλώσσες μειονοτήτων ή για την επιστημονική μελέτη των γλωσσών. Ακόμη, πολλές από τις κυβερνήσεις δεν διαθέτουν τις απαιτούμενες τεχνικές δυνατότητες για πλήρη ανάλυση όλων των γλωσσικών ιδιωμάτων που χρησιμοποιούνται στην επικράτειά τους. Αντίθετα, η αρμόδια ομάδα του Unicode έχει επεξεργαστεί χρήσιμες κωδικοποιήσεις της γραφής πολλών μειονοτικών γλωσσών ή γλωσσών ιστορικού ενδιαφέροντος.¹² Σ' αυτές περιλαμβάνονται η θιβητιανή, η γλαγολιτική, η οποία θα μπορούσε να έχει θεωρηθεί παραλλαγή του κυριλλικού αλφαβήτου, η αρχαία σφηνοειδής περσική και, μεταξύ αυτών που έχουν ιδιαίτερο ενδιαφέρον για την ελληνική γλώσσα, η γραμμική Β, της οποίας έχει προταθεί η κωδικοποίηση, και η κυπριακή συλλαβική γραφή που προς το παρόν μελετάται.

Για την πραγμάτωση αυτών των αρχών του Unicode χρειάστηκαν ευρύτατες γλωσσολογικές αναλύσεις, αλλά και συμβιβασμοί, λόγω της τρέχουσας γραφής τους και των ιδιοτυπιών του γραπτού λόγου. Για παράδειγμα, αν και το ελληνικό, το λατινικό και

12. *Exploratory Proposals*, Unicode Technical Report #3, Unicode, Inc., 1992. Η διευθέτηση του πίνακα ASCII διατίθεται στην ηλεκτρονική διεύθυνση: http://ecsdg.lu.se/pub/mail/unicode/TechReports/UTR_3.ascii.

το κυριλλικό «А» θα μπορούσαν να είχαν ενοποιηθεί σ' ένα ενιαίο λογικό «А» —μια απόλυτα νόμιμη ενοποίηση από ιστορική και δομική άποψη— η μακρόχρονη πρακτική της γραφής έχει καταστήσει το ελληνικό, το λατινικό και το κυριλλικό αλφάβητο χωριστές τυπογραφικές και κωδικοποιητικές ενότητες χαρακτήρων.

Ουσιαστικότεροι συμβιβασμοί απαιτήθηκαν για την διατήρηση της αρχής της εύκολης μετατροπής απεικονίσεων μεταξύ των υπαρχόντων προτύπων. Χαρακτηριστικά, η σερβοκροατική γλώσσα¹³ περιέχει ένα φθόγγο, ο οποίος στην λατινική γραφή (κροατικά) εμφανίζεται ως το δίγραμμο «Lj», ενώ στην κυριλλική (σερβικά) ως το απλό γράμμα «Љ». Για να παραμείνει εύκολη η κωδική χαρτογράφηση αυτού του γράμματος, το πρότυπο Unicode παραβιάζει τις αρχές του και παραχωρεί έναν κωδικό χαρακτήρα στο «Lj». Σε άλλες περιπτώσεις και για την απόδοση συγκεκριμένων φθόγγων, έχει χρησιμοποιηθεί η τυποποίηση συμβόλων του Διεθνούς Φωνητικού Αλφαβήτου (International Phonetic Alphabet ή IPA), τα οποία μοιάζουν με παραλλαγές ελληνικών ή λατινικών γραμματοσειρών.

Παρ' όλα αυτά, το Unicode 1.0 στο σύνολό του είναι ένα πρότυπο που παρουσιάζει συνέπεια και αποτελεσματικότητα.

Unicode και 10646

Για ένα χρονικό διάστημα, υπήρξε κίνδυνος να προκύψουν δύο χωριστά πρότυπα, το ισχυρό από πολιτική, αλλά αδύνατο από τεχνική άποψη ISO 10646, και το τεχνικά ισχυρό, αλλά πολιτικά αδύνατο Unicode.

Ευτυχώς, η υιοθεσία του DIS 10646 αναχαιτίστηκε από έναν αριθμό σημαντικών μελών του ISO λόγω των τεχνικών αντιρρή-

σεων που διατυπώθηκαν από την κοινόπραξία του Unicode και τα μέλη της. Επιπρόσθετα, ήταν σαφές η δέσμευση ενός μεγάλου τμήματος της βιομηχανίας ηλεκτρονικών υπολογιστών στο πρότυπο Unicode, το οποίο είχε σκοπό να χρησιμοποιήσει, ασχέτως οποιασδήποτε διαδικασίας, την καθιέρωση προτύπων de jure.

Κατάληξη αυτής της κατάστασης ήταν μια συμβιβαστική λύση, η οποία υιοθετήθηκε ως ISO 10646.¹⁴ Το πρότυπο αυτό είναι πολύ συναφές με το Unicode ως προς την βάση κωδικοποίησής του με σχήμα των 16 δυαδικών ψηφίων —δηλαδή, το Βασικό Πολύγλωσσο Επίπεδο (Base Multilingual Plane)— αλλά περιλαμβάνει, επίσης, και προβλέψεις για το σχήμα των 32 δυαδικών, το οποίο συγγενεύει περισσότερο με τον τρόπο αντιμετώπισης πραγμάτων του DIS 10646. Σ' αυτό τον συμβιβασμό, διατηρήθηκαν, δυστυχώς, πολλές κακές προσεγγίσεις του DIS και των τυπολογικών χαρακτηριστικών εθνικής αντίληψης, όπως οι προσυνδυασμένοι χαρακτήρες και οι μορφές απεικονίσεώς τους.

Με αυτό, φθάνουμε στην συζήτηση των συγκεκριμένων θεμάτων κωδικοποίησης της ελληνικής γλώσσας και, ιδιαίτερα, στο πολύ σημαντικό θέμα των προσυνδυασμένων και «κινητών» διακριτικών.

ΘΕΜΑΤΑ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΕΛΛΗΝΙΚΩΝ

Όπως είδαμε παραπάνω, είναι αναγκαίο να αποφασιστεί τι θα απεικονίζει μια κωδικοποίηση και πώς θα το απεικονίζει με τον καλύτερο δυνατό τρόπο. Από την άλλη πλευρά, δεν είναι απαραίτητο να αποφασιστεί ο τρόπος εισαγωγής των χαρακτήρων ή το πώς θα εμφανίζονται στην οθόνη του ηλεκτρονικού υπολογιστή. Η διάκριση μεταξύ αυτών των δύο έχει ουσιαστική σημασία.

13. Οι πολιτικές εξελίξεις του νοτιοσλαβικού χώρου περιλαμβάνουν και την αμφισβήτηση της σερβικής και κροατικής γλωσσικής ενότητας.

14. International Organization for Standardization, *Universal Multiple-octet Coded Character Set (UCS)*, ISO/IEC 10646-1, 1993 (E), 1993.

Θέματα ιστορικής φύσης

Η μακραίωνη ιστορία της ελληνικής γραφής ξεκινά από τους αρχαίους χρόνους και φθάνει στην εποχή μας. Σ' αυτό το χρονικό διάστημα εξελίχθηκαν οι συμβάσεις της ορθογραφίας, αλλά ακόμη και το ίδιο το αλφάβητο. Η σύγχρονη χρήση της γραφής είναι εύκολο να προσδιοριστεί και δεν αποτελεί αντικείμενο αμφισβήτησης. Ο κατάλογος χαρακτήρων της νέας ελληνικής απλουστεύτηκε με την μονοτονική μεταρρύθμιση. Ίσως, τα μόνα θέματα που απομένουν για συζήτηση να είναι αν το τονικό σύμβολο και τα διαλυτικά πρέπει να προσδεθούν στον χαρακτήρα, στον οποίο αντιστοιχούν, και αν το τελικό σίγμα πρέπει να κωδικοποιηθεί χωριστά. Τα κανονικά πολυτονικά ελληνικά είναι κάπως πιο σύνθετα από δομική άποψη.

Αντίθετα, αν στραφούμε στην γραφή της Αρχαϊκής εποχής, τα πράγματα περιπλέκονται σοβαρά. Μερικά γράμματα, όπως το F που αργότερα ονομάστηκε δίγαμμα, το κόππα Q, το σαμπί X και το T¹⁵ έχουν περιέλθει σε αχρηστία, πράγμα που δεν παρουσιάζει σημαντική δυσκολία κωδικοποίησης, αφού το καθένα από αυτά μπορεί να αντιστοιχεί σ' έναν κωδικό. Το πρόβλημα γίνεται δυσκολότερο, όταν πρόκειται για μορφές γραμμάτων, των οποίων η έννοια είναι δυνατό να διαφέρει κατά περίπτωση. Για παράδειγμα, το Η συμβόλιζε το δασύ πνεύμα στην πρώιμη ορθογραφία της γλώσσας —έπειτα, η προφορική αυτή διαφοροποίηση συμβολιζόταν από την δασεία— και αργότερα κατέληξε να γίνει το φωνήεν ήτα. Ένας θιασώτης της ενοποίησης θα μπορούσε, ίσως, να ισχυριστεί ότι η δασεία και το Η δεν είναι παρά απλές παραλλαγές του ίδιου θεμελιακού χαρακτήρα, οπότε έπρεπε να δεχθούμε ότι είναι ανύπαρκτη η γραφική

απόδοση της φιλής στην αρχαϊκή απεικόνιση.¹⁶ Το ίδιο θα ίσχυε και για τα ιωνικά Ι και Ι. Ακόμη, ένα σχήμα που μοιάζει πολύ με του μυ, απέδιδε τον φθόγγο «σ» και, κατά πάσα πιθανότητα, ονομαζόταν «σαν».¹⁷ Παρόμοια φαινόμενα περιέχονται στην πρώιμη κυριλική γραφή.¹⁸

Ένα άλλο ακανθώδες ζήτημα πηγάζει από τις διαφορές των ορθογραφικών κανόνων στην πορεία του χρόνου. Το σύνολο, σχεδόν, των κειμένων που χρησιμοποιούνται σήμερα από τους μελετητές των αρχαίων —με εξαίρεση τους επιγραφολόγους και τους παπυρολόγους— έχει υποβληθεί σε τυποποιήσεις και αντικατοπτρίζουν αλεξανδρινές συμβάσεις και κανόνες. Αυτό που συχνά λησμονείται είναι ότι οι κανόνες και οι συμβάσεις δεν ήταν πάντοτε ίδιες και ο τόνος, για παράδειγμα, συνόδευε κάποτε το πρώτο γράμμα των διφθόγων και όχι το δεύτερο, όπως σήμερα, πράγμα που δυσχεραίνει ουσιαστικά τους συνδυασμούς γραμμάτων και τόνων. Άλλοτε, πάλι, παρόμοια με τα τονικά σημεία κοσμούσαν μη τονιζόμενα φωνήεντα. Κάποιες φορές η φιλή γραφόταν επάνω από το σύμφωνο ρω «ρ». Αντίστροφα, σύγχρονοι κανόνες και συμβάσεις επιβάλλονται σήμερα σε αρχαία κείμενα. Αν και στην ιστορία δεν συναντάται επισεσυρμένη ή πεζή μορφή του γράμματος δίγαμμα F, τίποτε δεν φαίνεται να απαγορεύει την γραφή ενός πεζού στην μέση μιας λέξης ή ενός κεφαλαίου στην αρχή μιας πρότασης, όπως ακριβώς χρησιμοποιούμε το πεζό ή το κεφαλαίο γάμμα ακόμη και σε

16. Alfred Schmitt, *Der Buchstabe η im Griechischen*, Münster 1952.

17. Lilian H. Jeffery, *The local scripts of archaic Greece: a study of the origin of the Greek alphabet and its development from the eighth to the fifth centuries B.C.*, Oxford University Press, 1990.

18. David J. Birnbaum, "Standardizing Characters, Glyphs, and SGML Entities for Encoding Early Cyrillic Writing," *Computer Standards and Interfaces*, 1996.

15. Antonio Lillo, *The ancient Greek numeral system: a study of some problematic forms*, Bonn 1990. (δεν το έχω δει).

κείμενα που μας είναι γνωστά από αποκλειστικά κεφαλαιογράμματα επιγραφές.¹⁹

Δεν υπάρχει αμφιβολία πως η ιστορία της γραπτής ελληνικής γλώσσας είναι πολύπλοκη και θα ήταν παράλογο να περιμένουμε από την απλή κωδικοποίηση χαρακτήρων την λύση όλων των προβλημάτων που παρουσιάζει. Για να κατανοήσουμε τις διαστάσεις του ζητήματος, αρκεί να σκεφθούμε πως το δίγαμμα Γ, το έκτο γράμμα του πρώιμου ελληνικού αλφαβήτου, συγχέοταν αργότερα με το σύμπλεγμα Σ ή ακόμη και με το τελικό σίγμα ς, σε σημείο ώστε το αριθμητικό έξι να καταλήξει να γράφεται ως ΣΤ'!

Συγγενή αρχαϊκά αλφάβητα

Πολλές είναι οι πρώιμες γραφές γλωσσών των ακτών της Μεσογείου, που μοιάζουν με το ελληνικό αλφάβητο. Σ' αυτές περιλαμβάνεται η γραφή της Φρυγίας, της Καρίας, της Λυκίας, της Λυδίας, της Ετρουρίας και, φυσικά, του Λατίου. Αμεσα ή έμμεσα,²⁰ όλες προέρχονται από το φοινικικό αλφάβητο και περιέχουν πολλά κοινά γράμματα.²¹ Με εξαίρεση την λατινική, δεν έχω υπ' όψη μου μελέτες που αφορούν τους κωδικούς χαρακτήρων αυτών των γλωσσών.

Η ελληνική γραφή στον ρωμαϊκό κόσμο

Στον ρωμαϊκό κόσμο υπάρχουν περιπτώσεις, όπου, με περιορισμένες τροποποιήσεις, το ελληνικό αλφάβητο εξυπνέτησε ως γραφή

άλλων γλωσσών. Μια από αυτές είναι η κοπτική γραφή της αιγυπτιακής γλώσσας,²² που βασίζεται στο ελληνικό αλφάβητο με προσθήκη επτά ακόμη συμβόλων και χρησιμοποιήθηκε παράλληλα με το σύστημα γραφής της δημοτικής αιγυπτιακής. Κατά την αρχική επεξεργασία του προτύπου Unicode, είχε προταθεί η ενοποίησή της με την ελληνική γραφή. Επίσης, μερικές γαλατικές (κελτικές) επιγραφές χρησιμοποιούν το ελληνικό αλφάβητο²³ και, σε διάφορες εποχές, η ελληνική γραφή χρησιμοποιήθηκε σε ειδικές περιπτώσεις για την απόδοση μεμονωμένων λέξεων ή φράσεων άλλων γλωσσών. Ένα σχετικά μεγάλο παράδειγμα αυτής της χρήσης περιλαμβάνεται στο βιβλίο *Ἐξαπλᾶ* του Ωριγένη, όπου μια στήλη είναι το κείμενο της εβραϊκής Βίβλου σε ελληνική γραφή. Υπάρχει, ακόμη, το παράδειγμα ενός αραβικού αποσπάσματος του 8ου αιώνα.²⁴

Αλφάβητα βασισμένα στο ελληνικό

Πολλά σημαντικά αλφάβητα σχηματίστηκαν με βάση το ελληνικό. Το κυριλλικό είναι το σημαντικότερο — αν και φαίνεται ότι η γλαγολιτική, η πρωιμότερη μορφή σλαβικής γραφής, βασιζόταν κι αυτή στην ελληνική²⁵ —, καθώς επίσης και το αρμενικό και το γεωργιανό. Εξάλλου, το γοτθικό αλφάβητο του Wulfila — το οποίο δεν πρέπει να συγχέεται με την γερμανική γραφή που είναι γνωστή ως Fraktur —, ήταν ένας συνδυασμός του ελληνικού και του λατινικού αλφαβήτου με μερικά πρόσθετα ειδικά σύμβολα. Όλα αυτά τα συστήματα γραφής

19. Παρατήρηση του Γιάννη Χαραλάμπους στο ISO/IEC J TC1/SC2/WG2 N1019 (6 Απριλίου 1994).

20. Η αμεσότητα ή μη της προέλευσης είναι θέμα ανοικτής διαμάχης! Βλ. Martin Bernal, *Cadmean Letters: the Transmission of the Alphabet to the Aegean and Further West before 1400 B.C.*, 1990, και P. Kyle McCarter, Jr., *The antiquity of the Greek alphabet and the early Phoenician scripts*, Harvard Semitic Monographs αρ. 9, Scholars Press, 1975.

21. David Diringer, *The alphabet: a key to the history of mankind*, 3η έκδ., Funk & Wagnalls, 1968, and Ignace J. Gelb, *A study of writing*, αναθεωρημένη έκδοση, University of Chicago Press, 1963.

22. Υπογραμμίζεται ότι η κοπτική ήταν αρχικά σύστημα γραφής μόνο και όχι γλώσσα, ενώ τώρα ονοματίζει την ίδια την γλώσσα.

23. Την πληροφορία μου έδωσαν οι Keith Dix και Paul Amblard.

24. B. Violet, «Ein zweisprachiges Psalmfragment aus Damaskus», *Berichtiger Sonderabzug aus der Orientalistischen Litteratur-Zeitung*, 1901, (IV, 384), Berlin 1902.

25. André Vaillant, *Manuel du vieux slave*, Institut d'études slaves, τόμ. 1, Paris 1948, σελ. 21.

εμφανίζονται σήμερα αρκετά διαφοροποιημένα από το ελληνικό, ώστε να μην ενοποιηθούν με αυτό στο πρότυπο Unicode.

Νεότερες χρήσεις της ελληνικής γραφής

Μολονότι το ελληνικό αλφάβητο χρησιμοποιήθηκε κυρίως για την γραφή της ελληνικής γλώσσας, υπάρχουν πολυάριθμες ενδιαφέρουσες περιπτώσεις χρησιμοποίησής του χωρίς ουσιαστικές τροποποιήσεις για την γραφή μη ελληνικών γλωσσών στην πολιτιστική περιφέρεια της ελληνικής παιδείας ή στον χώρο των ορθόδοξων χριστιανών.²⁶

Αλβανικά (τόσκικα)

Για μεγάλο χρονικό διάστημα, το κύριο σύστημα γραφής της τοσκικής γλώσσας της Νότιας Αλβανίας βασιζόταν στο ελληνικό.²⁷ Το Συνέδριο του Μοναστηρίου, το 1908, αποφάσισε τελικά την καθιέρωση μιας λατινικής ορθογραφίας, τόσο για τα τόσκικα, όσο και για τα γκέγκικα αλβανικά.²⁸

26. Jürgen Kristophson, «Das griechische als Überlieferungsträger für fremdes Sprachgut», *Folia Neohellenica*, 6 (1984), σελ. 37.

27. Albanian Bible, *Psalteri, kequem mbas ebra iste-se viete skip nde gegeniste prei Konstantinit Kristoforidit, Elbasanist*, Constantinople 1872. Αυτή η αλβανική Καινή Διαθήκη έχει ένα συγκριτικό πίνακα με ορθογραφικές συμβάσεις για τα αλβανικά στην λατινική (δύο διαφορετικές), στην κυριλλική, στην ελληνική και στην αραβική γραφή. Επίσης βλ., Demetrio Camarda, *Appendice al saggio de Grammatologia comparata sulla lingua albanese*, Prato 1866, και *The Gospel in many tongues: specimens of 872 languages in which the British and Foreign Bible Society has published or circulated some portion of the Bible*, British and Foreign Bible Society, 1930, 1965. Armin Hetzer, ed. *Das dreisprachige Wörterverzeichnis von Theodoros Anastassiu Kavalliotis aus Moschopolis, gedruckt 1770 in Venedig: albanisch-deutsch-neugriechisch-aromunisch* (Balkan-Archiv. Beiheft, τόμ. 1) Hamburg, H. Burdke, 1981. Κωνσταντίνος Χριστοφορίδης, *Λεξικόν τῆς Ἀλβανικῆς Γλώσσης*, Ἀθῆναι 1904, και T. Moschopolitis, *Das Lexikon Tetraglosson des Daniil Moschopolitis*, ed. Jürgen Kristophson, στο περιοδικό *Zeitschrift für Balkanologie*, 9, 11. Πρόκειται για λεξικό της ελληνικής, αλβανικής, αρωμουνικής και της μακεδονικής γλώσσας —«βουλγαρικής» στο πρωτότυπο— με ελληνικούς χαρακτήρες.

28. A. Buda, ed., *Alfabeti i gjuhes shquipe dhe*

Νοτιοσλαβικά (μακεδονικά)

Διάφορες νοτιοσλαβικές διάλεκτοι, σχεδόν όλες παραπλήσιες με την σύγχρονη σλαβική μακεδονική γλώσσα,²⁹ έχουν επιζήσει σε ελληνική γραφή.³⁰ Κατά τα φαινόμενα, οι συγγραφείς τέτοιων κειμένων ήταν γνώστες της ελληνικής και δεν είχαν σχετιστεί με την κυριλλική γραφή. Τα σλαβικά μακεδονικά γράφονται σήμερα στα κυριλλικά.

Αρωμουνικά (βλάχικα)

Λίγα δείγματα της αρωμουνικής γλώσσας υπάρχουν σε ελληνική γραφή και τα περισσότερα από αυτά είναι παρόμοια με την περίπτωση των μακεδονικών.³¹ Ένα πιο πρόσφατο δείγμα είναι η γραμματική της αρωμουνικής που έγραψε βλαχόφωνος Έλληνας, η οποία και δημοσιεύτηκε στην Θεσσαλονίκη.³² Δεν υπάρχει μέχρι σήμερα καθιερωμένη ορθογραφία της αρωμουνικής, αλλά φαίνεται πιθανότατο ότι τελικά θα υιοθετηθεί ένα σύστημα βασισμένο στο λατινορωμουνικό πρότυπο.³³

Γκαγκάουζ

Η ονομασία γκαγκάουζ ανήκει σε μια τουρκική γλώσσα ή διάλεκτο, η οποία ομιλείται

Kongresi i Manasterit (14–22 nendor 1908): studime, materiale, dokumente, Tirane 1972.

29. Η οποία, ως σλαβική γλώσσα δεν προέρχεται από την αρχαία μακεδονική γλώσσα.

30. «Dva balgarski rakopisa s gracko pismo», *Balgarski starini*, 6, Sofia, 1920, André Mazon και André Vaillant, *L'Evangelaire de Kulakia, un parler slave de Bas-Vardar*, Bibliothèque d'études balkaniques, αρ. 6, Droz 1938, T. Moschopolitis, ό.π., Max Demeter Preyfass, *Die Druckerei von Moschopolis, 1731–1769: Buchdruck und Heiligenverehrung in Erzbistum Achrida*, Wiener Archiv für Geschichte des Slawentums und Osteuropas, Bd. 13, Vienna 1989, σελ. 56ff.

31. Μίλτος Γαρίδης, «Ο Μητροπολίτης Παΐσιος και η βλάχικη επιγραφή του Κλιννοβού: αλφάβητο και έθνικό πρόβλημα», *Τὰ Ιστορικά*, 2, 3 (1985), σελ. 183. Επίσης, βλ. T. Kavalliotis, ό.π., T. Moschopolitis, ό.π., και M. D. Preyfass, ό.π.

32. Αντώνης Μιχ. Κολτσίδας, *Γραμματική και Λεξικό τῆς Κοντοσβλάχικης γλώσσας*, Θεσσαλονίκη, εκδ. Ἀφῶν Κυριακίδη, 1978.

33. Προσωπική επικοινωνία με τον Victor Friedman.

από έναν περιορισμένο αριθμό κατοίκων περιοχών της Μολδαβίας, της Ουκρανίας, της Ρουμανίας, της Βουλγαρίας και ίσως μερικών ασιατικών πρώην σοβιετικών δημοκρατιών και της Ελλάδας. Γραφόταν με το ελληνικό αλφάβητο.³⁴

Σουργκούτσι

Η τουρκική γλώσσα ή διάλεκτος σουργκούτσι ήταν η ομιλία μιας μικρής ομάδας ορθόδοξων χριστιανών της Βόρειας Ελλάδας. Δεν υπάρχουν πολλές πληροφορίες γι' αυτήν, φαίνεται όμως ότι χρησιμοποιούσε την ελληνική γραφή.³⁵

Ουρούμ (ελληνικά ταρταρικά)

Πρόκειται για άλλη μια δευτερεύουσα γλώσσα της περιοχής Ντονέτσκ της Ουκρανίας, η οποία γραφόταν με το ελληνικό αλφάβητο.³⁶

Τουρκικά (καραμανλικά)

Τέλος, την τουρκική γλώσσα μ' αυτή την ειδική ονομασία μιλούσαν στην Μικρά Ασία πολλοί ορθόδοξοι χριστιανοί και την έγραφαν με το ελληνικό αλφάβητο.³⁷

Αν και αυτές οι γλώσσες δεν έχουν πια ιδιαίτερη σημασία, διατηρούν, ωστόσο, ειδικό ενδιαφέρον για τους ιστορικούς και τους γλωσσολόγους, οι οποίοι μπορούν να βρουν σ' αυτές στοιχεία και πληροφορίες για θέματα σχετικά με εθνικές μειονότητες και θρησκείες που δεν υπάρχουν σε άλλες πηγές.

34. Ingvar Svanberg, «The Turkish-speaking ethnic groups in Europe», *Europa Ethnica*, 5, σελ. 65.

35. I. Svanberg, ό.π., και S.B. Bernstein, «Les langues turques de la peninsule balkanique et l'Union des Langues Balkaniques», *Actes du premier Congrès International des Études Balkaniques et sud-est Européennes*, 1966, Académie Bulgare des Sciences, Sofia, 5, 73, 1968.

36. Baruch Podol'skii, *A Greek Tatar-English glossary*, Wiesbaden, O. Harrassowitz, 1985.

37. Michael G. Miller, *Karamanli-Turkish Texts: The Historical Changes in their Script and Phonology*, Indiana 1974 (δεν το έχω δει), και Severien Salaville και Eugène Dalleggio, *Karamanlidika: bibliographie analytique d'ouvrages en langue turque imprimés en caractères grecs*, Institut Français d'Athènes, Centre d'études d'Asie Mineure, Archives musicales de folklore, Athènes 1952.

Για τους γλωσσολόγους είναι ενδιαφέρον ότι γλωσσολογικά στοιχεία μπορούν να εξαχθούν από την ελληνική, τα οποία δεν καταγράφονται στην επικρατούσα γραφή. Για παράδειγμα, επειδή η αραβική γραφή δεν έχει τόνους, τα αραβικά κείμενα σε ελληνική γραφή έχουν προσφέρει ουσιαστική βοήθεια στην μελέτη του τονισμού της γλώσσας.³⁸ Με παρόμοιο τρόπο, караμανλικά κείμενα θα μπορούσαν να αποσαφηνίσουν ζητήματα τονισμού και χρήσης φωνέντων στα τουρκικά, αφού τα αραβικής γραφής οθωμανικά γραπτά σημειώνουν ατελώς τα φωνήεντα και καθόλου τους τόνους. Είναι, λοιπόν, ευνόητοι οι ουσιαστικοί λόγοι που υπαγορεύουν την προσπάθεια για ηλεκτρονική απεικόνιση αυτών των κειμένων.

Σημεία και σύμβολα ελληνικής γραφής

στις μη ελληνικές γλώσσες

Οι μη κανονικές χρήσεις των ελληνικών γραμμάτων παρουσιάζουν μια σειρά από ιδιαιτερότητες όσον αφορά την ελληνική ορθογραφία. Για την απεικόνισή τους ο κωδικός χαρακτήρων δεν πρέπει να περιορίζεται στην συμβατική χρήση της ελληνικής γραφής. Επιπλέον, κάποια χειρόγραφα και άλλα ιστορικά αρχεία γραμμένα στα ελληνικά πιθανώς να χρησιμοποιούν ιδιότυπη ορθογραφία, η οποία, ασφαλώς, δεν μπορεί να απεικονιστεί με κωδικοποιημένους χαρακτήρες που περιορίζονται σε «γραμματικούς» συνδυασμούς.

Υπάρχουν διάφορα μη ελληνικά διακριτικά για ειδικές χρήσεις, όπως, για παράδειγμα, δύο τελείες πάνω από το βήτα («β̇») και το δέλτα («δ̇») υποδηλώνουν την σκληρή προφορά τους. Μια τελεία πάνω ή μια γραμμή κάτω από το σίγμα («σ̇» ή «σ̣») μπορεί να προσδιορίζει την προφορά του παχέος sh.

38. J. Blau, «Middle and old Arabic material for the history of stress in Arabic», *Bulletin of the School of Oriental and African Studies* 35 (1972), σελ. 476.

Αντίθετα, υπάρχουν διάφορα συμβατικά ελληνικά διακριτικά που χρησιμοποιούνται με μη συμβατικούς τρόπους, όπως η μέση δασεία, το υπογεγραμμένο ιώτα αλλά και η επίθεση διαίρεσης σε γράμματα που κανονικά δεν την επιδέχονται. Επιπλέον, ορισμένες ενώσεις γραμμάτων, που παλαιότερα αποτελούσαν εναλλακτικά συμπλέγματα της ελληνικής ορθογραφίας, έχουν μετατραπεί σε ξεχωριστά σύμβολα, όπως το ου «ϛ» και το στ «ς», ενώ κάποια συνδυαστικά σύμβολα χρησιμοποιούνται συχνά μεταξύ ζευγών γραμμάτων όπως το τσ ή το ιζ. Σε μερικά από αυτά τα συστήματα αναμιγνύονται στοιχεία από τα ελληνικά, τα λατινικά ή τα κυριλλικά.

Σύμβολα για επιστημονικές εφαρμογές

Μερικοί χαρακτήρες της ελληνικής γραφής χρησιμοποιούνται διεθνώς σε διάφορους επιστημονικούς ή άλλους τομείς και, επειδή η ειδική αξιοποίησή τους δεν ακολουθεί πάντοτε τον κανονικό τρόπο χρήσης τους σε άλλες γλώσσες ή και στα ίδια τα ελληνικά, η κωδικοποίησή τους και οι συνδυασμοί τους δεν μπορούν να περιοριστούν στην αυστηρή ορθογραφική και γραμματική πειθαρχία της ελληνικής γλώσσας και χρήσης. Οι φιλόλογοι συχνά προσδιορίζουν τα μακρά και τα βραχέα φωνήεντα με μια ευθεία ή με ανεστραμμένη καμπύλη γραμμή πάνω από το αντίστοιχο γράμμα και σημειώνουν την εναλλακτική δυνατότητα ή περιπτώσεις αμφιβολίας με τον συνδυασμό και των δύο γραμμών. Οι επιγραφολόγοι χρησιμοποιούν ειδικά σύμβολα στην θέση γραμμάτων που έχουν φθαρεί. Οι γλωσσολόγοι και οι λεξικογράφοι χρησιμοποιούν ποικιλία συμβόλων για να δηλώσουν συγκεκριμένες αποδόσεις, διευκρινίσεις και αλλαγές της προφοράς ελληνικών, λόγω διαφορετικής διαλέκτου, ή άλλων γραμμάτων, όπως τον συνδυασμό «ντ» ως «d», «nd» ή «nt», την υπογράμμιση του «σ» ή την επίθεση μιας κάθετης

γραμμής σ' αυτό για να διαβαστεί ως «sh» κ.τ.λ. Ακόμη, για την πιστή απόδοση κειμένων που περιέχουν ιδιαίτερες γραπτές ή προφορικές παραλλαγές, διαλεκτικούς όρους και λάθη, είναι πολλές φορές αναγκαία η επιστράτευση ειδικών συμβόλων ή συνδυασμών διακριτικών και γραμμάτων κατά παράβαση της κανονικής ορθογραφίας, γραμματικής και χρήσης.³⁹

Δεν είναι, βέβαια, απαραίτητη η κωδικοποίηση όλων ανεξαιρέτως αυτών των σημείων, αλλά σε πολλά απ' αυτά είναι δυνατό και εύλογο να δοθούν κωδικοί χαρακτήρων.

Άλλες χρήσεις

Τέλος, υπάρχουν και άλλες ειδικές χρήσεις ιδιαίτερων σημείων και συμβόλων που συνοδεύουν το ελληνικό αλφάβητο —οι σημειώσεις ύμνων⁴⁰— αλλά και η χρησιμοποίηση των ιδίων των γραμμάτων του ελληνικού αλφαβήτου ως ειδικών σημείων,⁴¹ πράγμα πολύ σπάνιο για να δικαιολογεί ειδική εργασία κωδικοποίησης.

ΚΙΝΗΤΑ ΔΙΑΚΡΙΤΙΚΑ

Το εύλογο συμπέρασμα που προκύπτει από την συζήτηση των παραπάνω παραδειγμάτων είναι ότι τα διακριτικά και τονικά σημεία είναι στην ουσία τόσο κινητά και ανεξάρτητα στοιχεία της γραφής, όσο και τα γράμματα, αφού είναι σαφές ότι τα ελληνικά διακριτικά και οι τόνοι μπορούν να χρησιμοποιηθούν με ιδιαίτερους και ανορθόδοξους

39. Προσωπική επικοινωνία με τον Nikos Nicholas.

40. Max Haas, *Byzantinische und slavische Notationen*, series *Palaeographie der Musik*, 1, Köln 1973. (δεν το έχω δει).

41. Erice, *Scritture, libri e testi nelle aree provinciali di Bisanzio*, atti del seminario di Erice (18-25 Σεπτεμβρίου 1988), ed. Guglielmo Cavallo, et al., series *Biblioteca del Centro per il collegamento degli studi medievali e umanistici nell'Università di Perugia* 5, Spoleto: Centro italiano di studi sull'alto Medioevo, 1991, *passim*.

τρόπους, και ότι μη ελληνικά διακριτικά μπορούν να συνοδεύουν ελληνικά γράμματα.

Όπως έχει ήδη αναφερθεί, υπάρχουν δύο προσεγγίσεις στην απεικόνιση διακριτικών σε κωδικούς χαρακτήρων. Η μια προβλέπει τον συνδυασμό διακριτικών και γραμμάτων σε ενιαίους, «προσυνδυασμένους» χαρακτήρες, ενώ η άλλη υιοθετεί την απεικόνιση των διακριτικών ως χωριστών, «κινήτων» ή, ας πούμε, «συμπλέξιμων» στοιχείων, τα οποία μπορούν να συνδυάζονται κατά περίπτωση με άλλα στοιχεία. Εδώ πρέπει να υπενθυμίσουμε ότι τόσο τα συμπλέγματα της πρώτης προσέγγισης, όσο και οι συνδυασμοί της δεύτερης, μπορούν να απεικονίζονται με κατάλληλα σχεδιασμένες γλύνφους. Επίσης, πρέπει να αναφερθεί ότι τα κινητά διακριτικά είναι ευχρηστότερα και μπορούν να χρησιμοποιούνται, όπου και όπως είναι εκάστοτε επιθυμητό. Αλλωστε, πολλά κοινά τονικά και διακριτικά σημεία χρησιμοποιούνται σε συνδυασμό με ελληνικά, λατινικά, κυριλλικά και άλλα γράμματα, όπως τα σημεία στίξης.

Οι προσυνδυασμένοι χαρακτήρες περιορίζουν τον χρήστη στους συνδυασμούς που θεωρεί απαραίτητους ένας συντάκτης κειμένου με συνηθισμένες απαιτήσεις. Αντίθετα, τα κινητά διακριτικά επιτρέπουν συνδυασμούς κατά βούληση και ακόμη και χωρίς προφανή χρησιμότητα ή καν νόημα. Αλλά, αυτή ακριβώς είναι η χρηστική βάση της φύσης των κωδικών για χαρακτήρες. Η επιλογή ενός ρω με υπογεγραμμένη στερείται νοήματος, όσο και η ακολουθία χαρακτήρων «xttx», αλλά οι κωδικοί χαρακτήρων δεν έγιναν, βέβαια, για να υπαγορεύουν τις χρηστικές επιλογές και να περιορίζουν την ελευθερία της έκφρασης. Ωστόσο, οι προσυνδυασμένοι κωδικοί μοιάζουν αρχικά να ευνοούν τον εύκολο χρηστικό χειρισμό, καθώς τα προγράμματα λογισμικού δεν είναι απαραίτητο να γνωρίζουν τους κανόνες συνδυασμών. Στην πραγματικότητα

όμως η εντύπωση αυτού του πλεονεκτήματος είναι απατηλή στα σύγχρονα λογισμικά προγράμματα, όπου η απεικόνιση είναι χωριστή διεργασία που πρέπει να είναι ικανή να αντιμετωπίζει συνδυασμούς και συμπλέγματα. Ο συνδυασμός ενός διακριτικού μ' έναν χαρακτήρα γράμματος δεν είναι, τελικά, διαφορετικός από τον συνδυασμό ενός «f» και ενός «i» για την δημιουργία του συμπλέγματος «fi». Παρόμοια, οι μέθοδοι εισαγωγής διαχωρίζονται από τους εσωτερικούς κώδικες και έτσι κανένα από τα δύο συστήματα δεν υπερτερεί σε αυτό το σημείο.

Ο προγραμματισμός εφαρμογής διεργασιών, όπως η ταξινόμηση σε αλφαβητική σειρά, ο ορθογραφικός έλεγχος ή η ανάλυση μετρικής δόμησης, δεν είναι απλή υπόθεση σε καμία από τις δύο προσεγγίσεις. Μολονότι τα κινητά διακριτικά συχνά απλοποιούν τις διεργασίες, άλλοτε οι προσυνδυασμένοι χαρακτήρες είναι το ίδιο ή και περισσότερο χρήσιμοι. Πάντως, όπως και αν έχουν τα πράγματα, στις ιδιαίτερα επιτηδευμένες διεργασίες είναι μάλλον ενδεδειγμένη η χρησιμοποίηση των δικών τους απεικονίσεων:

1. Η ταξινόμηση, όπως έδειξε ο LaBonté, δεν μπορεί να βασίζεται άμεσα στους κωδικούς χαρακτήρων.⁴²

2. Η ταχεία αλυσιδωτή αναζήτηση, όπως αυτή που ακολουθεί το πρότυπο των Boyer και Moore,⁴³ μπορεί να προσαρμοστεί και στις δύο μεθοδεύσεις απεικόνισης.

42. Alain LaBonté, *Quand "Z" vient-il avant "a"?: Algorithme de tri respectant langues et cultures*, Direction générale des technologies de l'information, Ministère des Communications, Gouvernement du Québec, Canada 1990. Επίσης, βλ. του ιδίου «Multiscript Ordering for Unicode», *Proc. Fourth Unicode Implementors Workshop*, Unicode, Inc. 1992.

43. R. S. Boyer και J. S. Moore, «A fast string searching algorithm», *Commun. of the Association for Computing Machinery*, 20, 10 (1977), σελ. 762.

Τα κινητά διακριτικά απαιτούν κάπως μεγαλύτερο χώρο αποθήκευσης από τους προσυνδυασμένους χαρακτήρες.

3. Ο χώρος αποθήκευσης και στις δύο περιπτώσεις εξισώνεται περίπου μετά την πραγματοποίηση συμπίεσης των δεδομένων.⁴⁴ Από την άλλη πλευρά, οι προσυνδυασμένοι χαρακτήρες καταλαμβάνουν μεγαλύτερο χώρο στους πίνακες κωδικών. Ο αριθμός των πρόσθετων προσυνδυασμένων ελληνικών χαρακτήρων του μονοτονικού συστήματος είναι σχετικά μικρός, αλλά αυτό το σύστημα αποκλείει πολλούς άλλους χρήσιμους συνδυασμούς. Επιπλέον, η υποστήριξη της αρχής του προσυνδυασμού για τα ελληνικά αποτελεί ενθάρρυνση για την εφαρμογή της ίδιας αρχής και σε γλωσσικές περιπτώσεις, όπως των κορεατικών, που μπορούν να επιβαρύνουν σημαντικά το μέγεθος των πινάκων.

Τελικά, τόσο οι κινητοί χαρακτήρες διακριτικών, όσο και οι χαρακτήρες προσυνδυασμών, εξυπηρετούν ικανοποιητικά τις τρέχουσες ελληνικές κειμενογραφικές ανάγκες για δικόγραφα, εμπορικές επιστολές, τεχνικά γραπτά κ.τ.λ. Από την άλλη πλευρά, μόνον τα κινητά διακριτικά προσφέρουν ουσιαστικές λύσεις για τα προβλήματα ιστορικών, φιλολογικών, γλωσσολογικών και άλλων ειδικών ερευνητικών αναγκών των κειμένων.⁴⁵

Οι προσυνδυασμένοι χαρακτήρες είναι σαφώς ακατάλληλοι για την εξυπηρέτηση πολλών άλλων γλωσσών πέρα από την ελληνική.

44. Εχω συγκρίνει πειραματικά τα μεγέθη κωδικοποίησης ελληνικών κειμένων με προσυνδυασμένους χαρακτήρες και με κινητά διακριτικά, χρησιμοποιώντας γνωστούς αλγόριθμους συμπίεσης.

45. Το θέμα έχει γνωστοποιηθεί από γλωσσολόγους στους σχεδιαστές του Unicode, πολλοί από τους οποίους είναι οι ίδιοι γλωσσολόγοι, στο ηλεκτρονικό περιοδικό *Linguist* 2.264, .283, .295, .296, .299, .352, του Ιουνίου-Ιουλίου 1991. Αυτό βρίσκεται στην ηλεκτρονική διεύθυνση <http://tamvm1.tamu.edu/~linguist/>

Συγχώνευση του Unicode και του ISO 10646

Η κατάληξη του θέματος ήταν να περιληφθούν τα κινητά διακριτικά στην τελική μορφή του προτύπου ISO 10646, αλλά με βαρύ αντίτιμο. Περιλήφθηκε, επίσης, και μεγάλος αριθμός προσυνδυασμένων χαρακτήρων, πράγμα που περιπλέκει την διεργασία και καταλαμβάνει υπολογίσιμο χώρο στους πίνακες κωδικών, ο οποίος θα μπορούσε να χρησιμοποιηθεί αποδοτικότερα για πληρέστερη κάλυψη των διάφορων γλωσσών ολόκληρου του κόσμου. Το χειρότερο είναι πως τα κινητά διακριτικά θεωρούνται προαιρετικό μέρος του προτύπου με ιεράρχηση Επιπέδου 3 (Level 3) και, κατά πάσα πιθανότητα, αυτό θα δημιουργήσει σε πολλούς σχεδιαστές λογισμικών την εντύπωση ότι δεν είναι αναγκαίο να συμπεριλάβουν κινητά διακριτικά στα συστήματά τους. Το γεγονός θα είναι ατυχές για τους χρήστες της ελληνικής γραφής, όπως είναι και για τους χρήστες άλλων γλωσσών. Παρόλα αυτά, το ISO 10646 είναι το μοναδικό διεθνές πρότυπο που καθιστά τουλάχιστον δυνατή την απεικόνιση των χαρακτήρων όλων των γλωσσών του κόσμου.

ΣΥΜΠΕΡΑΣΜΑ

Το πρότυπο ISO 10646 αποτελεί ουσιαστικό βήμα προόδου για την επεξεργασία πολύγλωσσων κειμένων. Προσφέρει ηλεκτρονική γραφή για όλες τις κύριες, καθώς και για τις περισσότερες δευτερεύουσες γλώσσες της υφής. Αυτή την στιγμή εκδηλώνεται έντονη δραστηριότητα για τις γλώσσες που απασχολούν τους μελετητές, και το ISO 10646 προσφέρει σ' αυτές ουσιαστική βοήθεια.

Τα προβλήματα αυτού του προτύπου προκύπτουν ακριβώς από τις πολυάριθμες δυνατότητες για προαιρετικές επιλογές, τις οποίες παρέχει. Τα κινητά διακριτικά που, όπως είδαμε, είναι ουσιώδη για την αποτελεσματική αντιμετώπιση πολλών προβλημά-

των επεξεργασίας πολύγλωσσων κειμένων γενικά και ελληνικών ειδικά, παρέχονται ως προαιρετικές επιλογές. Το ISO 10646 προβλέπει ακόμη την επέκταση σε κώδικα των 32 δυαδικών ψηφίων, μια δυνατότητα που η αξιοποίησή της θα γίνει αναγκαία, επειδή ακριβώς έγιναν αποδεκτοί οι προσυνδυασμένοι χαρακτήρες, καθώς επίσης και εξαιτίας ορισμένων θεμάτων των κινεζικών, ιαπωνικών και κορεατικών χαρακτήρων, για τα οποία δεν προσφέρεται ο χώρος εδώ για διεξοδική συζήτηση.

Η εφαρμογή του ISO 10646 καθιερώνεται τώρα. Η τεχνολογία για τις μεθόδους εισαγωγής και απεικόνισης υπάρχει, μολονότι δεν έχει ακόμη αξιοποιηθεί ευρέως. Υπάρχουν, επίσης, οι γραμματοσειρές, τουλάχιστον για τα αλφαβητικά μέρη. Το πρόγραμμα Windows NT — η επόμενη γενεά λειτουργικού συστήματος της εταιρείας Microsoft — χρησιμοποιεί εσωτερικά το πρότυπο Unicode/10646, πράγμα που δεν συμβαίνει στο Windows 95 που, ωστόσο, προσφέρει ένα βαθμό υποστήριξης στο Unicode.⁴⁶ Παράλληλα, είναι πιθανότατο το Unicode να υιοθετηθεί ως πρότυπο κωδικοποίησης χαρακτήρων από τον Ιστό Παγκόσμιας Εμβέλειας (World Wide Web).⁴⁷ Από την άλλη πλευρά, ο σχεδιασμός πολλών σύγχρονων ηλεκτρονικών προϊόντων δεν περιλαμβάνει κινητά διακριτικά και είναι λιγοστά τα προγράμματα εφαρμογών

για τους τελικούς χρήστες, τα οποία όμως πληθαίνουν.⁴⁸

Το ουσιαστικότερο ζήτημα στον σχεδιασμό κωδικών είναι ο διαχωρισμός των θεμάτων του ίδιου του κώδικα από τα θέματα απεικόνισης και εισαγωγής. Μολονότι το ISO 10646 δεν αντιμετωπίζει εντελώς αποτελεσματικά αυτό το ζήτημα, λύνει πολλά προβλήματα σχετικά με τις διεργασίες για ελληνικά κείμενα, τα οποία αναφέρθηκαν σε αυτή την έκθεση. Οι ενδιαφερόμενοι για την ελληνική γλώσσα έχουν συμφέρον να υποστηρίξουν το πρότυπο Unicode και να διεκδικήσουν την ευρεία εφαρμογή του, ώστε να περιληφθούν σε αυτό όλοι οι συμπλέξιμοι χαρακτήρες.

ΕΥΧΑΡΙΣΤΙΕΣ

Για τις χρήσιμες συμβουλές και τις εποικοδομητικές κριτικές παρατηρήσεις τους, οφείλω ευχαριστίες στους Victor Friedman, καθηγητή του University of Chicago, Glenn Adams της Stonehand, Inc., και David J. Birnbaum, καθηγητή του University of Pittsburgh. Επίσης, οφείλω αναγνώριση για τις πρόσθετες χρήσιμες πληροφορίες τους στους Paul Amblard, T. Keith Dix, Lance Eccles, Robert Ingria, Jörg Knappen, Jim Marchand, P. Kyle McCarter, Christopher Miller, Nikos Nicholas, Tito Orlandi, Erland Sommarskog, Timothy Spalding, Bruce Tindall και Bryan Williams.

Languages and Writing Systems» στην ηλεκτρονική διεύθυνση <http://www.w3.org/hypertext/WWW/International/Overview.html>. Για πρόσθετες οδηγίες βλ. «WInter=Web Internationalization & Multilingualism» στο <http://dorado.crpht.lu:80/~carrasco/winter/>

48. Η εφαρμογή του Unicode «Unicode Product Catalog» μπορεί να βρεθεί στην ηλεκτρονική διεύθυνση <http://www.stonehand.com/unicode/products.html> ή από το Unicode Consortium.

46. Βλ. την ηλεκτρονική διεύθυνση <http://198.105.232.5/DEVONLY/TECH/GLOBAL/UNICW95C.HTM>.

47. Πρόχειρες μεθοδεύσεις της χρήσης των ελληνικών στο WWW μπορούν να βρεθούν σήμερα στην ηλεκτρονική διεύθυνση <http://scholar.cc.emory.edu/scripts/TELA/TELA-foreign-chars.html>. Για προτεινόμενες βελτιώσεις βλ. «W3 Internationalization/Localization: Non-western Character sets,

ΥΣΤΕΡΟΓΡΑΦΟ

Το 1995, χρονιά που έγινε το Συνέδριο, υπήρχε η ελπίδα πως το διεθνές πρότυπο iso 10646/Unicode θα έλυνε πολλά προβλήματα της διεθνούς κωδικοποίησης χαρακτήρων. Σήμερα, το 1998, το iso 10646 διατίθεται στους χρήστες μέσα από μια μεγάλη ποικιλία εφαρμογών και, παράλληλα, η ανεπιφύλακτη υποστήριξή του από τις εφαρμογές και τα λειτουργικά συστήματα της εταιρείας Microsoft εξασφαλίζει τον ρόλο του και την συνέχειά του.

Δυστυχώς, όπως άλλωστε είχα σημειώσει το 1995, η πλήρης στήριξή του iso 10646 σπανίζει. Οι ελληνικοί σύνθετοι χαρακτήρες-γλύφοι καλύπτουν όλες τις ανάγκες ενός κανονικού πολυτονικού νεοελληνικού ή και αλεξανδρινού κειμένου, αλλά πολλές σημαντικές εφαρμογές —ερευνητικού χαρακτήρα, ιδιαίτερα— απαιτούν ένα πλήρες σύνολο ελεύθερων διακριτικών. Αντί να επιμεινούν για την πλήρη εφαρμογή του iso 10646, πολλοί οργανισμοί εθνικών προτύπων επέμεναν να εξασφαλίσουν ένα δήθεν «πλήρες» σύνολο γλύφων για την γλώσσα τους. Το αποτέλεσμα είναι πως πολλοί χρήσιμοι

συνδυασμοί διακριτικών και χαρακτήρων δεν είναι εφικτοί, ενώ ένας μεγάλος αριθμός σπάνιων ή αρχαϊκών συνδυασμών σπαταλούν κωδικές θέσεις στο πρώτο επίπεδο του ISO 10646, το λεγόμενο BMP (Basic Multilingual Plan).

Αυτό, ίσως, αποτελεί μια νίκη των «πραγματοστών», αλλά έχει ως συνέπεια την αδιαφορία των κατασκευαστών λογισμικού για τις ανάγκες των ερευνητών. Αλλωστε, πολλοί ερευνητές δεν έχουν επωφεληθεί από την υποστήριξη που παρέχουν οι νεότερες εφαρμογές.

Ίσως θα έπρεπε οι τυπογράφοι και οι ερευνητές να επιμεινούν στην σημασία του συνδυασμού χαρακτήρων με ελεύθερα διακριτικά και να εξηγήσουν την σπουδαιότητά τους στους υπόλοιπους χρήστες. Ταυτόχρονα, σωματεία όπως η ΕΕΤΣ, θα πρέπει τουλάχιστον να προσπαθήσουν να κυκλοφορήσουν τις γραμματοσειρές τους με το πρότυπο του Unicode, συμπεριλαμβάνοντας όλες τις γλύφους, αλλά και τα ελεύθερα διακριτικά, προσφέροντας μαζί και τον σχετικό οδηγό του πληκτρολογίου για την σωστή τους χρήση.

ΒΙΒΛΙΟΓΡΑΦΙΑ

AZIZ T. ATIYA (επιμ.), *The Coptic Encyclopedia*, Macmillan 1991, s.v. Alphabets, Coptic; Alphabet in Coptic, Greek.

ISO 8879, International Organization for Standardization, *Text and Office Systems - Standard Generalized Markup Language*, ISO 8879, 1986.

DONALD E. KNUTH, *The Metafont book*, Addison-Wesley, 1984.

Ιωάννης Δράκοντος Σταματάκος, *Λεξικὸν τῆς Νέας Ἑλληνικῆς γλώσσης: Καθαρευούσης καὶ Δημοτικῆς καὶ εἰς τὴν Ἀρχαίαν*, Ἀθῆναι 1952.