

---

gr. pctrans, wintrans, mactrans:

## Λύσεις που προτείνουμε για τα προβλήματα της αυτόματης αποκατάστα- σης ελληνικού κειμένου

---

Υπό **Αμαλίας Χατζηευγενιάδου**,  
του Τμήματος  
Υπολογιστών του Πανεπιστημίου του Bradford

Τα προγράμματα gr, pctrans, wintrans, mactrans σκοπό έχουν να επαναφέρουν το εκλατινισμένο ελληνικό κείμενο στην πρωτότυπη γραφή του, να αποκαταστήσουν δηλ. το ελληνικό κείμενο. Και τα τέσσερα αυτά προγράμματα εκτελούν την ίδια λειτουργία, όπως παρουσιάζουν διαφορετικά αποτελέσματα ανάλογα με τον υπολογιστή που χρησιμοποιείται. Η εν λόγω λειτουργία είναι αυτή της επαναφοράς και πραγματοποιείται στον πυρήνα του προγράμματος.

Ο διαχωρισμός δράσης - απόδοσης επιτρέπει στους χρήστες του συστήματος να επιλέξουν το κατάλληλο πρόγραμμα, χωρίς η επιλογή αυτή να επηρεάζει τον πυρήνα. ο ίδιος ο πυρήνας με τη σειρά του περιλαμβάνει ένα πλήθος λειτουργιών που και αυτές είναι γραμμένες ως ανεξάρτητα εκτελέσιμα υποπρογράμματα. Στο εξής θα αναφερόμαστε στα υποπρογράμματα αυτά ως μόντουλα. Η συνολική διαδικασία της μετατροπής διασπάται μ' αυτό τον τρόπο σε επί μέρους εργασίες, η κάθε μία από τις οποίες μπορεί να αναπτυχθεί, να μεταφερθεί και να συντηρηθεί ανεξάρτητα από τις άλλες.

Με τον όρο «εκλατινισμένο κείμενο», εννοούμε ένα κείμενο που έχει μετατραπεί χαρακτήρα προς χαρακτήρα σε λατινική γραφή από ένα ελληνικό πρωτότυπο. Υπάρχουν δύο βασικοί τρόποι/μέθοδοι για να επιτύχουμε μια τέτοια μετατροπή, η μεταγραφή και ο μεταγραμματισμός (βλ. σχετική εισήγηση της κ. Evelyn Cornel). Παράλληλα υπάρχει ένα πλήθος από ομάδες κανόνων που ορίζουν την αντιστοιχία ανάμεσα στους ελληνικούς και λατινικούς χαρακτή-

ρες, ένα πλήθος διαφορετικών απεικονίσεων. Οι απεικονίσεις αυτές συνιστούν τα λεγόμενα πρότυπα μεταγραμματισμού (transliteration standards), τα οποία μπορεί να έχουν οριστεί σε εθνικό επίπεδο ή επίπεδο οργανισμού ή ινστιτούτου. Στην πράξη βέβαια η χρήση του όρου μεταγραμματισμός εδώ είναι κάπως χαλαρή, αφού συνήθως αυτά τα πρότυπα περιλαμβάνουν κανόνες τόσο μεταγραμματισμού όσο και μεταγραφής (βλ. τη σχετική εισήγηση της κ. Evelyn Cornel).

Η αντιστοιχία μεταξύ ελληνικών και λατινικών χαρακτήρων ακολουθεί κάθε φορά ένα τέτοιο πρότυπο μεταγραμματισμού και λαμβάνει χώρα στον πυρήνα του προγράμματος. Έτσι έχουμε ένα εκτελέσιμο υποπρόγραμμα για τη μετατροπή σύμφωνα με το πρότυπο της βιβλιοθήκης του Κογκρέσου και ένα διαφορετικό για το πρότυπο του ΕΛΟΤ. Συνολικά ο πυρήνας περιλαμβάνει επτά τέτοια μόντουλα για επτά διαφορετικά πρότυπα: ISO R843, Library of Congress, ELOT 743, Oxford, British Library pre 75, Regole Italiane di catalogazione per Autori, Regeln für die alphabetische Katalogisierung. Οι χρήστες του συστήματος μπορούν να διαλέξουν το πρότυπο που θα εφαρμόσει το πρόγραμμα προσθέτοντας στην εντολή μια από τις επτά κωδικοποιημένες συντομογραφίες: ISO, LC, ELOT, OX, BL, IT, RAK, π.χ.

\* gr - s ISO file

\* mactrans -s ELOT file

\* pctrans-s IT file

\* wintrans-s ISO file

\* wintrans-s RAK file

\* mactrans-s BL file, κλπ, όπου το file περιέχει το προς αποκατάσταση κείμενο.

Ένας άλλος τομέας καταμερισμού εργασιών είναι αυτός που αναφέρεται στη μετατροπή ή τη μη μετατροπή του κειμένου. Ένα κείμενο μπορεί να περιέχει στοιχεία που δεν εκλατινίστηκαν, επομένως δεν χρειάζονται αποκατάσταση, π.χ. οι αραβικοί αριθμοί, ή όταν στο εκλατινισμένο κείμενο παραμένουν λέξεις με ελληνική γραφή. Εκτός από τον πολύπλοκο μηχανισμό της μετατροπής μέσα στον πυρήνα, λοιπόν, υπάρχει και ένα μόντουλο που αποδίδει άθικτα τμήματα. Η απόφαση σχετικά με το αν το συγκεκρι-



μένο τμήμα θα πρέπει να μετατραπεί ή όχι, λαμβάνεται εσωτερικά στον πυρήνα.

Με την υποδιαίρεση του συστήματος σε μόντουλα επιτυγχάνουμε τη μείωση της απαιτούμενης για αποθήκευση μνήμης, προκειμένου να αποθηκεύσουμε και να τρέξουμε τα προγράμματα, καθώς επίσης και τη μείωση της εργασίας από την άποψη του προγραμματισμού. Οι περισσότερες εργασίες που εμπλέκονται στο σύστημα κωδικοποιήθηκαν μόνο μια φορά και χρησιμοποιούνται απ' όλα τα προγράμματα. Έχουμε λύσει, μ' αυτό τον τρόπο βασικά προβλήματα που προκύπτουν από την ποικιλία των διαθέσιμων προτύπων και συσκευών. Εν τούτοις υπάρχουν επί πλέον προβλήματα. Μερικά τα έχουμε αντιμετωπίσει ήδη, ενώ παράλληλα σχεδιάζουμε την επίλυση άλλων.

Έχω ταξινομήσει αυτά τα προβλήματα σε τέσσερις κατηγορίες:

1. Αυτά που σχετίζονται με τα υπάρχοντα πρότυπα μεταγραμματισμού

2. Αυτά που σχετίζονται με την ελληνική γλώσσα

3. Αυτά που σχετίζονται με το προσωπικό των βιβλιοθηκών

4. Αυτά που σχετίζονται με τις συσκευές, υπολογιστές, δίκτυα κλπ.

1. Σχετικά με τα πρότυπα μεταγραμματισμού

Αν το πρότυπο είναι αντιστρέψιμο, τότε οι κανόνες, που έχουν χρησιμοποιηθεί για την εκλατίνιση ενός ελληνικού κειμένου, μπορούν να χρησιμοποιηθούν αντεστραμμένοι, προκειμένου το κείμενο αυτό να αποκατασταθεί στην αρχική μορφή του. Κάτι τέτοιο θα επιτευχθεί αν κάθε χαρακτήρας ή συνδυασμός χαρακτήρων μέσα στο κείμενο αντιστοιχεί σε ένα μοναδικό ελληνικό χαρακτήρα ή συνδυασμό. Για παράδειγμα, αν κάποιο πρότυπο μετατρέπει το γράμμα πι σε λατινικό p, τότε μπορούμε να μετατρέψουμε το λατινικό p σε πι. Όσον αφορά τώρα τους συνδυασμούς χαρακτήρων θα πρέπει κάθε φορά να εξετάζουμε ένα χαρακτήρα μέσα στο πλαίσιο του. Το σύστημα το πετυχαίνει αυτό διαβάζοντας σειρές συμβόλων, αντί απλών χαρακτήρων. Έτσι αν το ελληνικό ψι αντιστοιχεί στο συνδυασμό ps, θα πρέπει να εξεταστεί αν το p στο λατινικό κείμενο συνοδεύεται από s, οπότε θα είναι ψι και όχι πι και σίγμα.

Δυστυχώς τα πρότυπα στην πλειοψηφία τους δεν είναι αντιστρέψιμα και περιέχουν αμφιλεγόμενα σημεία που απαιτούν περισσότερη και πολυπλοκότερη επεξεργασία για να αποσαφηνιστούν, χωρίς αυτό να αποτελεί εγγύηση για το τελικό αποτέλεσμα. Αμφιλεγόμενα σημεία είναι οι περιπτώσεις που ένας λατινικός χαρακτήρας αντιστοιχεί σε περισσότερα από ένα ελληνικά σύμβολα. Στο πρότυπο της Βρετανικής Βιβλιοθήκης για παράδειγμα υπάρχει ένας κανόνας σύμφωνα με τον οποίο το έψιλον όπως και το ήτα αποδίδονται με το λατινικό χαρακτήρα e. Αν επιχειρήσω να αντιστρέψω αυτόν τον κανόνα έχω ένα αμφιλεγόμενο σημείο. Σε ορισμένες περιπτώσεις αυτό μπορεί να αποσαφηνιστεί με βάση το πλαίσιο, δηλαδή τους χαρακτήρες πριν και μετά το γράμμα e, αλλά όχι πάντα.

Αν αυτή η μέθοδος αποτύχει, το σύστημα παρουσιάζει όλες τις εναλλακτικές λύσεις ως ισοδύναμες, π.χ. η λέξη «μπορεί» θα μετατραπεί σε μπορεί. Ο λόγος είναι ότι ο συνδυασμός στο τέλος μιας λέξης είναι πιο πιθανό να είναι έψιλον-γιώτα απ' ότι ήτα-γιώτα. Στην περίπτωση της λέξης «metera» όμως, δεν μπορούμε να κάνουμε καμιά τέτοια υπόθεση και το αποτέλεσμα θα είναι μ<ε/η> τ<ε/η>ρα, σύμφωνα με το πρότυπο της Βρετανικής Βιβλιοθήκης. Οι αναγνώστες σ' αυτή την περίπτωση θα πρέπει να εντοπίσουν ένα λογικό συνδυασμό για να φτιάξουν μια ελληνική λέξη με βάση το γλωσσικό τους αισθητήριο. Αναγνωρίζουμε βέβαια ότι μια τέτοια μορφή επιδέχεται βελτίωση.

Στην τελική φάση του προγράμματος θα χρησιμοποιήσουμε μια πιο πολύπλοκη μέθοδο χειρισμού αμφιλεγόμενων σημείων, με τη βοήθεια ενός λεξικού, δηλαδή μιας λίστας λέξεων που θα χρησιμοποιείται ως σημείο αναφοράς. Θα περιγράψω την αρχή, βάσει της οποίας λειτουργεί το λεξικό αυτό με ένα παράδειγμα. Ας υποθέσουμε την εκλατινισμένη λέξη metera: από τη λέξη αυτή προκύπτουν τέσσερις υποψήφιες λέξεις:

- μετερα

- μητηρα

- μητερα

- μετηρα

το σύστημα θα ψάξει στο λεξικό για κάθε μια απ' αυτές, αν κάποια βρεθεί, το σύστημα θα αποδώσει το



metaera με τη συγκεκριμένη λέξη που βρέθηκε στο λεξικό και θα απορρίψει τις άλλες. Σε περίπτωση που εντοπιστούν περισσότερες από μία θα εμφανιστούν στο κείμενο εναλλακτικές λύσεις ή μπορεί να απαιτηθεί περαιτέρω επεξεργασία για τον εντοπισμό της σωστής λέξης. Αν όμως δεν βρεθεί καμιά λέξη, θα αναγκαστούμε να την αποδώσουμε με τον παραπάνω ανεπαρκή τρόπο.

Ένα άλλο πρόβλημα στην ίδια κατηγορία είναι η μετατροπή ονομάτων. Τα πρότυπα, αντιστρέψιμα και μη, δεν προβλέπουν ιδιαίτερη μεταχείριση για τα ονόματα τόπων ή προσώπων. Τις περισσότερες φορές η εκλατίνιση ονομάτων δεν υπακούει σε κανόνες που επιβάλλονται από κάποιο πρότυπο. Μπορεί να έχει γίνει με βάση τα καθιερωμένα αρχεία ονομάτων των βιβλιοθηκών, τα λεξικά, ή μπορεί να είναι μεταγραφές ή μορφές που έχουν καθιερώσει οι ίδιοι οι συγγραφείς. (βλ. σχετική εισήγηση της κ. Evelyn Cornel).

Στο HELEN μεταχειριζόμαστε τα ονόματα σε επίπεδο λέξης και όχι χαρακτήρα. Θα χρειαστούμε ένα λεξικό ονομάτων, που θα περιέχει τα ονόματα μεταγεγραμμένα στη μορφή που υπαγορεύει το πρότυπο ISO R843. Η επιλογή του προτύπου αυτού σήμερα γίνεται αναγκαία για δύο λόγους:

α) η δημιουργία ενός αρχείου/λεξικού με εκλατινισμένα τα ελληνικά ονόματα είναι απαραίτητα, αφού κανένα ελληνικό σύνολο χαρακτήρων δεν είναι ακόμα κοινώς αποδεκτό.

β) το ISO R843 είναι απολύτως αντιστρέψιμο.

Το λεξικό περιέχει ένα όνομα σε κάθε σειρά στην μορφή:

όνομα <Tab> ISO μεταγραμματισμός <Tab> χώρα

Όταν το πρόγραμμα εντοπίσει ένα όνομα, θα το αναζητήσει στην πρώτη στήλη του λεξικού, θα το αντικαταστήσει με την ISO μορφή του και μετά θα το μετατρέψει ως συνηθισμένο κείμενο. Η παράθεση χώρας είναι προαιρετική αλλά μπορεί να χρησιμεύσει όταν η μετατροπή γίνεται από την ελληνική στη λατινική γραφή. Μπορεί να βοηθήσει το πρόγραμμα να εντοπίσει ποιά από το διαθέσιμο πλήθος γραφών του συγκεκριμένου ονόματος είναι η κατάλληλη. Αυτό το επί πλέον πεδίο δεδομένων μέσα στο λεξικό μπορεί να αποδειχθεί χρήσιμο για την αντιμετώπιση

και άλλων προβλημάτων. Για παράδειγμα στην περίπτωση που θέλουμε ορισμένες λέξεις ή υποπεδία να μην είναι δεκτικές μεταγλώττισης, βάζουμε τη θέση του πεδίου αυτού τη λέξη με ένα αστερίσκο, σηματοδοτώντας τη μη μετατροπή της λέξης αυτής. Π.χ.

όνομα <Tab> UNIX<Tab> \*

Η ίδια τακτική μπορεί να ακολουθηθεί και για ένα άλλο πρόβλημα που μπορεί να προκύψει σε ένα κείμενο κατά τη μετατροπή, εννοούμε κάποια τμήματα του κειμένου, τα οποία δεν θα πρέπει να μεταγλωττιστούν, αλλά να μεταφραστούν. Σ' αυτήν την περίπτωση οι προς μετάφραση λέξεις επισημαίνονται στο λεξικό με ένα θαυμαστό π.χ.

Unix<Tab> sel'da<Tab> (σύμβολο θαυμαστικού)

όπου το δεύτερο πεδίο περιέχει τη μετάφραση στα ελληνικά μεταγραμματισμένη σύμφωνα με το πρότυπο ISO.

## 2. Σχετικά με τη γλώσσα

Το ελληνικό σύστημα γραφής στην μακραίωνη ιστορική του εξέλιξη έχει αλλάξει μορφές εγκαταλείποντας κάθε φορά κάποια σύμβολα. Η τελευταία σημαντική αλλαγή έγινε πριν από λίγα χρόνια με την εγκατάλειψη του πολυτονικού συστήματος και την υιοθέτηση του μονοτονικού. Τα πρότυπα δεν επηρεάζονται ιδιαίτερα εξ' αιτίας της αλλαγής αυτής, μια και τα περισσότερα δεν περιλαμβάνουν τόνους. Παράλληλα όμως πολλά απ' αυτά χρησιμοποιούν τη δασεία. Το πρότυπο της βιβλιοθήκης του Κογκρέσσου, φερ' ειπείν, τοποθετεί δασεία, σύμφωνα με τους κανόνες, ακόμα και αν το πρωτότυπο ελληνικό κείμενο είναι μονοτονικό. Στην παρούσα φάση του προγράμματος αγνοούμε τη δασεία γενικώς, στο μέλλον όμως σκεπτόμαστε να προσθέσουμε ένα μέσο με το οποίο ο χρήστης θα μπορεί να καθορίζει αν προτιμά μονοτονική ή πολυτονική μετατροπή. Ακόμα και στην περίπτωση που το πρότυπο προβλέπει τη χρήση παλαιότερων συμβόλων και το κείμενο τα περιέχει, θα είναι δυνατή η απόδοση των συμβόλων αυτών, αν και εφ' όσον οι χρήστες το ζητήσουν.

## 3. Σχετικά με το προσωπικό βιβλιοθήκης

Στο HELEN έχει προβλεφθεί ακόμη η περίπτωση, κατά την οποία οι υπεύθυνοι καταλογογράφησης του ξενόγλωσσου υλικού να μην είναι σε θέση να



γνωρίζουν τους κανόνες με τους οποίους έγινε η καταλογογράφηση ή το εκλατινισμός του κειμένου. Ως εκ τούτου οι χρήστες μπορεί να μην γνωρίζουν ποιο πρότυπο να επιλέξουν όταν χρησιμοποιούν ένα από τα προγράμματα gr, mactrans, pctrans, wintrans.

Το σύστημα αντιμετωπίζει αυτό το πρόβλημα με τη βοήθεια ενός ανεξαρτήτου εκτελέσιμου υποπρογράμματος που σκοπό έχει να ανιχνεύσει το πρότυπο που χρησιμοποιήθηκε και το οποίο τίθεται αυτόματα σε λειτουργία, στην περίπτωση που οι χρήστες δεν επιλέξουν κάποιο συγκεκριμένο πρότυπο. Είναι ένα μόντουλο αυτόματης αναγνώρισης που λειτουργεί με βάση την αναγνώριση σχηματισμών. Ορισμένοι συνδυασμοί γραμμάτων εμφανίζονται μόνο σε ένα πρότυπο ή μια μικρή ομάδα προτύπων. Για παράδειγμα ο συνδυασμός: cch υπάρχει μόνο στο ιταλικό πρότυπο και αντιστοιχεί στο ελληνικό κάππα-χι., αυτός ο σχηματισμός είναι σπάνιος, αλλά έχουμε τη δυνατότητα να κάνουμε αυτή την τεχνική πιο αποτελεσματική χρησιμοποιώντας την σε περισσότερο πολύπλοκο επίπεδο. Αν ένα κείμενο, φερ' ειπείν, περιέχει οξεία, μπορεί να έχει προκύψει από μετατροπή με βάση το ISO ή το ΕΛΟΤ. Αν επιπλέον περιέχει και περισπωμένη ή βαρεία τότε θα είναι σίγουρα προϊόν που προέρχεται από την εφαρμογή του ISO. Αν επίσης έχει και h ως χι, τότε είναι προϊόν του ISO, διαφορετικά του ΕΛΟΤ.

Η μέθοδος αυτή είναι απολύτως αξιόπιστη, μόνο όταν συναντούνται οι περιπτώσεις αυτές που χαρακτηρίζουν ή αποτελούν τις, ας πούμε, ιδιοτυπίες των επί μέρους προτύπων. Για το λόγο αυτό χρειαζόμαστε κάθε φορά όσο το δυνατόν πιο πολλούς και πιο συνηθισμένους σχηματισμούς. Και βέβαια όσο μεγαλύτερο είναι το κείμενο, τόσο πιθανότερο είναι να ανιχνεύσουμε σ' αυτό κάποιο από τους σχηματισμούς αυτούς.

Ένα ακόμα πρόβλημα παρουσιάζεται όταν εντοπίσουμε μέσα στο κείμενο διαφορετικά σχήματα που αλληλοαντικρούονται, δηλαδή οδηγούν σε διαφορετικά πρότυπα. Να έχουμε για παράδειγμα ένα κείμενο, τμήμα του οποίου να οδηγεί στο ISO και ένα άλλο στο πρότυπο της Βρετανικής Βιβλιοθήκης. Αυτό μπορεί να οφείλεται σε ανθρώπινα λάθη που συνέβησαν κατά τις διάφορες φάσεις σχηματισμού του

καταλόγου. Φυσιολογικά όταν εντοπιστεί ένα πρότυπο, το πρόγραμμα καλεί το αντίστοιχο μόντουλο και ακτελεί μεταγραμματισμό αντίστοιχα. Αν εντοπιστούν επί πλέον στοιχεία που οδηγούν σε άλλη ή σε άλλα πρότυπα, τότε το πρόγραμμα μπορεί να τα παρουσιάσει ως εναλλακτικές λύσεις, δίδοντας ένα μήνυμα:

Υπάρχουν ενδείξεις των ISO, BL, RAK

και να αφήσει τους χρήστες να διαλέξουν, ή να τα δοκιμάσουν όλα κατά βούληση. Εναλλακτικά, το πρόγραμμα μπορεί να επεξεργαστεί τις ενδείξεις εσωτερικά π.χ. αν υπάρχει μια ένδειξη που οδηγεί στο ISO, 3 στο BL, και 4 στο RAK, τότε θα διαλέξει το πρότυπο με τις περισσότερες ενδείξεις. Σε οποιαδήποτε περίπτωση, με ή χωρίς εσωτερική επεξεργασία, το πρόγραμμα θα πρέπει να παρουσιάζει στους χρήστες ένα μήνυμα του τύπου:

Δεν έχει επιλεγεί πρότυπο από το χρήστη.

Το πρότυπο που χρησιμοποιήθηκε μπορεί να ήταν το X

Αν δεν εντοπιστεί κανένα συγκεκριμένο πρότυπο κατά τη διάρκεια της επεξεργασίας, το πρόγραμμα χρησιμοποιεί αυτόματα το ISO R843 δίνοντας το μήνυμα:

Δεν έχει επιλεγεί πρότυπο από το χρήστη

Το πρότυπο που χρησιμοποιήθηκε, μπορεί να ήταν ISO

Ακόμα και στην περίπτωση που η βιβλιοθήκη έχει χρησιμοποιήσει με συνέπεια ένα μοναδικό πρότυπο, δεν αποκλείεται να έχουν γίνει τυπογραφικά λάθη. Μπορεί να υπάρχουν λάθος χαρακτήρες στο κείμενο, δηλ. χαρακτήρες που δεν περιλαμβάνονται στο πρότυπο. Αυτοί οι χαρακτήρες παραλείπονται. Μερικές φορές όμως η παράλειψη/αποβολή ενός χαρακτήρα μπορεί να επιφέρει σύγχυση.

Για την αντιμετώπιση αυτού του προβλήματος το σύστημα παρέχει τη δυνατότητα της επιλογής verbose. Οι λάθος χαρακτήρες θα εμφανιστούν στην ελληνική απόδοση συνοδευόμενοι από πιθανές εναλλακτικές λύσεις π.χ. η λέξη «thamnos» σύμφωνα με το ISO R843, με την εντολή:

gr«or mactrans or wintrans or pctrans»-s ISO file

με verbose. Αυτή η μέθοδος δεν δίνει πάντα λύση στο πρόβλημα, αλλά μπορεί να βοηθήσει σε ορισμένες περιπτώσεις. Κάθε λάθος χαρακτήρας υφίσταται επεξεργασία σε ένα άλλο μόντουλο, το οποίο μπορεί



να βελτιωθεί ακόμα περισσότερο στο μέλλον. το πρόγραμμα δεν χρησιμοποιεί verbose εκτός αν οι χρήστες το ζητήσουν. Άλλωστε κι' αυτό μπορεί να προκαλέσει σύγχυση, διακόπτοντας τη ροή του κειμένου. Η λύση αυτή μπορεί να βελτιωθεί με τη χρήση ενός λεξικού, όπως εκείνο για τα αμφιλεγόμενα σημεία. Η λέξη «thamnos», για παράδειγμα, θα αναλυθεί ως εξής:

θαμβος  
θαμνος  
θαμυος

#### 4. Σχετικά με τον εξοπλισμό

Τα ελληνικά σύνολα χαρακτήρων που χρησιμοποιούνται από διάφορους κατασκευαστές υπολογιστών ποικίλουν (βλ. σχετική εισήγηση του κ. Mic Ridley). Το καθένα τοποθετεί τους ελληνικούς χαρακτήρες σε διαφορετικές θέσεις. Έτσι αυτό που διαβάζουμε ως μικρό άλφα σε μια οθόνη, μπορεί να παρουσιαστεί ως κεφαλαίο ήτα τονούμενο σε μια άλλη. Επίσης τα ελληνικά σύνολα χαρακτήρων είναι δυνατόν να υποστηρίζουν διαφορετικούς αριθμούς χαρακτήρων. Για παράδειγμα το DgTimesPoly υποστηρίζει πολύ περισσότερους χαρακτήρες από ότι το ΕΛΟΤ 928. Το πρώτο είναι πολυτονικό και μπορεί να λειτουργήσει και ως μονοτονικό, ενώ το δεύτερο είναι αποκλειστικά μονοτονικό. Για να αποφύγουμε ασάφειες και σύγχυση σχετικά με τον τρόπο που οι χαρακτήρες αποθηκεύονται, ερμηνεύονται και παρουσιάζονται σε έναν υπολογιστή, πρέπει να εξακριβώσουμε ότι χρησιμοποιούμε το κατάλληλο σύνολο χαρακτήρων.

Οι διάφορες μέθοδοι ερμηνείας χαρακτήρων συνδέονται με τον πυρήνα του συστήματος δια μέσου των προγραμμάτων gr, mactrans κλπ. Το gr ερμηνεύει τους χαρακτήρες για μηχανή UNIX, το mactrans για τα Micintosh, ενώ το wintrans, pctrans αντιστοιχεί στα PCs - Windows, DOS αντίστοιχα. Μπορούμε να αναπτύξουμε όσα τέτοια προγράμματα χρειαζόμαστε για άλλες μηχανές με ελάχιστο κόπο, εφόσον γνωρίζουμε το συγκεκριμένο σύνολο χαρακτήρων που χρησιμοποιούν. Ο πυρήνας θα διατηρείται κατ' ουσία ανέγγιχτος.

Υπάρχουν όμως επί πλέον προβλήματα που το σύστημά μας δεν μπορεί να λύσει. Τέτοια είναι η

παραμόρφωση που υφίσταται ένα νούμερο όταν μεταφερθεί μέσα στο δίκτυο. Επίσης τα πληκτρολόγια συχνά δεν περιέχουν όλα τα σύμβολα που περιλαμβάνει ένα πρότυπο. Για παράδειγμα το μακρόν (το σύμβολο που χρησιμοποιείται πάνω από το ε), που υπαγορεύει το πρότυπο της Βιβλιοθήκης του Κογκρέσσου για την απεικόνιση του γράμματος ήτα, συχνά δεν μπορεί να αναπαραχθεί και αντικαθίσταται από circumflex. Όταν αυτές οι μικρές παραλλαγές υποστούν επεξεργασία από το πρόγραμμα, μπορεί να δημιουργήσουν προβλήματα, εφ' όσον το πρόγραμμα θα ακολουθεί το πρότυπο πιστά, οπότε θα προβλέπει μακρόν. Ένα επιπρόσθετο ζήτημα εδώ είναι ότι το ελληνικό σύστημα γραφής είναι δύσκολο να συνδυαστεί με άλλα έστω και λατινογενή. Συχνά αυτό είναι αδύνατο γιατί κάθε άλλο σύστημα γραφής περιλαμβάνει τα δικά του διακριτικά σημεία που μαζί με τα ελληνικά σύμβολα, υπερβαίνουν το χώρο που διαθέτει το εκτεταμένο ASCII σύνολο χαρακτήρων (βλ. αναλυτική παρουσίαση του προβλήματος στην εισήγηση του κ. Mick Ridley). Οι εργασίες των επιτροπών ISO 10646 και Unicode επικεντρώνονται στα δημιουργία ενός μεγαλύτερου συνόλου χαρακτήρων, ώστε να μπορούν να περιληφθούν όλα τα συστήματα γραφής.

-----  
\*Σύνολα χαρακτήρων (character sets): ο τρόπος με τον οποίο οι χαρακτήρες αποθηκεύονται και αναπαρίστανται μέσα στον υπολογιστή, συχνά περιγράφεται ως ένα πλέγμα όπου κάθε χαρακτήρας, καταλαμβάνει μια συγκεκριμένη θέση στην οποία αναφερόμαστε με τη βοήθεια ενός αριθμού.