



Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

**Σχολή Τεχνολογικών Εφαρμογών
Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων**



Πτυχιακή Εργασία

Στατιστική ανάλυση των δεδομένων της Γραμμική Α σε σχέση με την Γραμμική Β με την χρήση της αναγνώρισης προτύπων.

ΣΠΟΥΔΑΣΤΕΣ

ΠΑΠΑΒΑΣΙΛΕΙΟΥ ΚΑΤΕΡΙΝΑ ΑΜ:940

ΦΡΑΓΓΕΔΑΚΗ ΓΙΩΡΓΙΑ ΑΜ:886

ΥΠ. ΚΑΘΗΓΗΤΗΣ

ΠΑΠΑΔΟΥΡΑΚΗΣ ΓΕΩΡΓΙΟΣ

Ευχαριστούμε:

Τον κ. Χαλκιαδάκη Ιωάννη για την πολύτιμη βοήθεια του στη χρήση του Matlab.

Τον κ. Τσικριτζή Μηνά που μας εμπιστεύτηκε τα δεδομένα που είχε συλλέξει και σας παρουσιάσαμε στην παρούσα έρευνα.

Τον κ. Παπαδουράκη Γεώργιο για την ανάθεση και επίβλεψη της πτυχιακής.

ABSTRACT

Our target is to examine if linear A is related and in which degree with linear B. This we will try to achieve it with the use of intelligent systems and concretely with the use of pattern recognition. More specifically, we try to create a system that we will train it with linear B, which is already known and deciphered, in order to import the unknown linear A and observe resemblances and disparities.

In order to achieve the particular research, we will use matlab. The matlab initially was a program of computers for persons that use numerical calculation, specifically in the linear algebra (tables). Henceforth, it constitutes a most powerful tool in the visualization, in the programming and in the communications. In the potential of matlab are included modern algorithms, possibilities of handling enormous quantities of data and powerful programming tools.

ΣΥΝΟΨΗ

Στόχος μας είναι να εξετάσουμε αν σχετίζεται και σε τι βαθμό η γραμμική A με την γραμμική B . Αυτό θα προσπαθήσουμε να το πετύχουμε με την χρήση των ευφών συστημάτων και συγκεκριμένα με την χρήση της αναγνώρισης προτύπων. Ποιο αναλυτικά, προσπαθούμε να δημιουργήσουμε ένα σύστημα το οποίο θα το εκπαιδεύσουμε στην γραμμική B , που είναι ήδη γνωστή και αποκωδικοποιημένη, ώστε όταν εισάγουμε την άγνωστη γραμμική A να παρατηρήσουμε ομοιότητες και ανομοιότητες.

Για να επιτευχθεί η συγκεκριμένη έρευνα θα χρησιμοποιήσουμε το matlab. Το MATLAB αρχικά ήταν ένα πρόγραμμα υπολογιστών για ανθρώπους που χρησιμοποιούν αριθμητικούς υπολογισμούς, ειδικά στη γραμμική άλγεβρα (πίνακες). Πλέον αποτελεί ένα ισχυρότατο εργαλείο στην οπτικοποίηση, στον προγραμματισμό, στην έρευνα, στην επιστήμη των μηχανικών, και στις επικοινωνίες. Στο δυναμικό του Matlab συμπεριλαμβάνονται μοντέρνοι αλγόριθμοι, δυνατότητες χειρισμού τεράστιων ποσοτήτων δεδομένων, και ισχυρά προγραμματιστικά εργαλεία.

Πίνακας περιεχομένων

1. Η εποχή του χαλκού	10
1.1 Η ανακάλυψη της γραμμικής Β	10
1.2 Η χρήση της γραφής στους μινωίτες.....	19
1.3 Η αποκρυπτογράφηση της γραμμικής Β.....	21
1.4 Τα προκαταρκτικά της αποκρυπτογράφησης:	23
1.5 Η αποκρυπτογράφηση της γραμμικής Β από τον Ventris	24
1.6 Μια μικρή περίληψη των παραπάνω.....	24
2. Αναγνώριση προτύπων	32
2.1 Αναγνώριση προτύπων	32
2.2 Μεθοδολογίες αναγνώρισης προτύπων.....	32
2.3 Ιστορική αναδρομή	33
2.4 Σύστημα αναγνώρισης προτύπων	33
2.5 Απλοί ταξινομητές	34
2.6 Ταξινομητές ελάχιστης απόστασης.....	34
2.7 Μέτρα απόστασης.....	34
3. Υλοποίηση ομαδοποίησης	37
3.1 Δημιουργία ομάδων	37
3.2 Εφαρμογές	37
3.3 Ιδανικές ομάδες.....	38
3.4 Μεθοδολογίες ομαδοποίησης.....	38
3.5 Ιεραρχική ομαδοποίηση	39
3.6 Αλγόριθμος απλής σύνδεσης	39
3.7 Αλγόριθμος πλήρους σύνδεσης.....	39
3.8 Σύγκριση απλής και πλήρους σύνδεσης.....	39
3.9 Αλγόριθμος σύνδεσης μέσω τιμών	40
4. Διαχωριστική ομαδοποίηση	41
4.1 Διαχωριστική ομαδοποίηση	41
5. MATLAB	42
5.1 Τί είναι το Matlab	42
5.2 Πώς δουλεύει το Matlab	42
5.3 Εισαγωγή.....	43
5.4 Προ-επεξεργασία δεδομένων	43
5.5 Ομαδοποίηση	44
5.6 Αλγόριθμος ομαδοποίησης kmeans	44
5.7 Ιεραρχική ομαδοποίηση	45
6. ΥΛΟΠΟΙΗΣΗ	49
6.1 Δεδομένα που μας δόθηκαν	49
7. ΙΕΡΑΡΧΙΚΗ ΟΜΑΔΟΠΟΙΗΣΗ	54
7.1 Ιεραρχική ομαδοποίηση του δείγματος b3344.....	54
7.2 Ιεραρχική ομαδοποίηση του δείγματος b11725.....	57
7.3 Ιεραρχική ομαδοποίηση του δείγματος b8460.....	60
7.4 Ιεραρχική ομαδοποίηση του δείγματος a1859.....	63
7.5 Ιεραρχική ομαδοποίηση των δειγμάτων b3344, b1725 & b8460	66
7.6 Ιεραρχική ομαδοποίηση των δειγμάτων b11725,b8460	68
8. ΚΜΕΑΝΣ	76
8.1 Υλοποίηση του αλγόριθμου ομαδοποίησης kmeans.....	76
8.2 Kmeans στο δείγμα b11725 με 10 ομάδες.....	84
8.3 Kmeans στο δείγμα b8460 με 11 ομάδες.....	94
8.4 Kmeans στο δείγμα a1859 με 10 ομάδες.....	111
8.5 Kmeans στα δείγματα b3344, b11725 & b8460 με 33 ομάδες	127
9. ΤΑΞΙΝΟΜΗΣΗ	139
9.1 Σύνολα εκπαίδευσης και ελέγχου	139

10. ΣΥΜΠΕΡΑΣΜΑΤΑ	144
10.1 Διάφορα συμπεράσματα	144
Βιβλιογραφία	151

Πίνακας εικόνων

1. Η ΕΠΟΧΗ ΤΟΥ ΧΑΛΚΟΥ ΣΤΟ ΑΙΓΑΙΟ	10
Σχήμα 1.2.1	15
Σχήμα 1.2.2	15
Σχήμα 1.2.3	16
Σχήμα 1.2.4	16
Σχήμα 1.2.5	17
Σχήμα 1.2.6	18
Σχήμα 1.2.7	18
Σχήμα 1.2.8	19
Σχήμα 1.4.1	22
Σχήμα 1.4.2	22
Σχήμα 1.6.2.1	25
5. Matlab	15
Σχήμα 5.7.1	42
Σχήμα 5.7.2	42
7. Ιεραρχική ομαδοποίηση του δείγματος b3344	49
Σχήμα 7.1.1	49
Σχήμα 7.2.1	52
Σχήμα 7.3.1	55
Σχήμα 7.4.1	58
Σχήμα 7.5.1	61
Σχήμα 7.5.2	62
Σχήμα 7.6.1	69
Σχήμα 7.6.2	69
8. Υλοποίηση του αλγόριθμου ομαδοποίησης kmeans	71
Σχήμα 8.1.1	72
Σχήμα 8.1.2	74
Σχήμα 8.1.3	76
Σχήμα 8.1.4	78
Σχήμα 8.2.1	80
Σχήμα 8.2.2	82
Σχήμα 8.2.3	84
Σχήμα 8.2.4	86
Σχήμα 8.2.5	88
Σχήμα 8.3.1	90
Σχήμα 8.3.2	92
Σχήμα 8.3.3	94
Σχήμα 8.3.4	96
Σχήμα 8.3.5	98
Σχήμα 8.3.6	100
Σχήμα 8.3.7	102
Σχήμα 8.3.8	104
Σχήμα 8.4.1	106
Σχήμα 8.4.2	108
Σχήμα 8.4.3	110
Σχήμα 8.4.4	112
Σχήμα 8.4.5	114
Σχήμα 8.4.6	116
Σχήμα 8.4.7	118
Σχήμα 8.4.8	120
Σχήμα 8.5.1	123
Σχήμα 8.5.2	126

Σχήμα 8.5.3	129
Σχήμα 8.5.4	132

Λίστα πινάκων

2. Αναγνώριση προτύπων	27
2.7.1 Πίνακας.....	31
6. ΥΛΟΠΟΙΗΣΗ	44
6.1.1 Πίνακας.....	45
6.1.2 Πίνακας.....	46
6.1.3 Πίνακας.....	47
6.1.4 Πίνακας.....	48
7. ΙΕΡΑΡΧΙΚΗ ΟΜΑΔΟΠΟΙΗΣΗ	49
7.1.1 Πίνακας.....	50
7.1.2 Πίνακας.....	51
7.2.1 Πίνακας.....	53
7.2.2 Πίνακας.....	54
7.3.1 Πίνακας.....	56
7.3.2 Πίνακας.....	57
7.4.1 Πίνακας.....	59
7.4.2 Πίνακας.....	60
7.5.1 Πίνακας.....	65
7.5.2 Πίνακας.....	66
7.5.3 Πίνακας.....	67
7.5.4 Πίνακας.....	67
7.5.5 Πίνακας.....	67
7.5.6 Πίνακας.....	68
7.6.1 Πίνακας.....	70
8. ΚΜΕΑΝΣ	71
8.1.1 Πίνακας.....	71
8.1.2 Πίνακας.....	71
8.1.3 Πίνακας.....	73
8.1.4 Πίνακας.....	73
8.1.5 Πίνακας.....	75
8.1.6 Πίνακας.....	75
8.1.7 Πίνακας.....	77
8.1.8 Πίνακας.....	77
8.2.1 Πίνακας.....	79
8.2.2 Πίνακας.....	79
8.2.3 Πίνακας.....	81
8.2.4 Πίνακας.....	81
8.2.5 Πίνακας.....	83
8.2.6 Πίνακας.....	83
8.2.7 Πίνακας.....	85
8.2.8 Πίνακας.....	85
8.2.9 Πίνακας.....	87
8.2.10 Πίνακας.....	87
8.3.1 Πίνακας.....	89
8.3.2 Πίνακας.....	89
8.3.3 Πίνακας.....	91
8.3.4 Πίνακας.....	91
8.3.5 Πίνακας.....	93
8.3.6 Πίνακας.....	93
8.3.7 Πίνακας.....	95
8.3.8 Πίνακας.....	95
8.3.9 Πίνακας.....	97
8.3.10 Πίνακας.....	97
8.3.11 Πίνακας.....	99

8.3.12 Πίνακας.....	99
8.3.13 Πίνακας.....	101
8.3.14 Πίνακας.....	101
8.3.15 Πίνακας.....	103
8.4.1 Πίνακας.....	105
8.4.2 Πίνακας.....	105
8.4.3 Πίνακας.....	107
8.4.4 Πίνακας.....	107
8.4.5 Πίνακας.....	109
8.4.6 Πίνακας.....	109
8.4.7 Πίνακας.....	111
8.4.8 Πίνακας.....	111
8.4.9 Πίνακας.....	113
8.4.10 Πίνακας.....	113
8.4.11 Πίνακας.....	115
8.4.12 Πίνακας.....	115
8.4.13 Πίνακας.....	117
8.4.14 Πίνακας.....	117
8.4.15 Πίνακας.....	119
8.4.16 Πίνακας.....	119
8.5.1 Πίνακας.....	121
8.5.2 Πίνακας.....	122
8.5.3 Πίνακας.....	124
8.5.4 Πίνακας.....	125
8.5.5 Πίνακας.....	127
8.5.6 Πίνακας.....	128
8.5.7 Πίνακας.....	130
8.5.8 Πίνακας.....	131
9. ΤΑΞΙΝΟΜΗΣΗ	133
9.1 Πίνακας.....	134
9.2 Πίνακας.....	135
9.3 Πίνακας.....	136
9.4 Πίνακας.....	137
10. ΣΥΜΠΕΡΑΣΜΑΤΑ	138
10.1 Πίνακας.....	139
10.2 Πίνακας.....	139
10.3 Πίνακας.....	139
10.4 Πίνακας.....	139
10.5 Πίνακας.....	141
10.6 Πίνακας.....	142
10.7 Πίνακας.....	143

Κεφάλαιο 1

Η ΕΠΟΧΗ ΤΟΥ ΧΑΛΚΟΥ ΣΤΟ ΑΙΓΑΙΟ

1.1 Η ανακάλυψη της γραμμικής Β

Η αποκάλυψη της εποχής του χαλκού στην περιοχή του αιγαίου άρχισε με τις πρωτοποριακές ανακαλύψεις του Ερρίκου Σλήμαν στην Τροία και τις Μυκήνες προς το τέλος του 19^{ου} αιώνα. Ο πιο μεγάλος προκλασικός πολιτισμός άκμασε στην μυκηναϊκή περίοδο. Αυτή μπορεί να χρονολογηθεί χονδρικά μεταξύ του 16^{ου} και 12^{ου} αι. π.χ. Ο Arthur Evans τόσο εντυπωσιάστηκε από το επίπεδο του πολιτισμού που είχαν επιτελέσει αυτοί οι μυκηναίοι στην ηπειρωτική Ελλάδα, ώστε σχημάτισε τη γνώμη ότι ένας τέτοιος πολιτισμός δεν μπορούσε να οργανωθεί χωρίς τη γνώση της γραφής. Ωστόσο ούτε στην Τροία ούτε στις Μυκήνες οι ανασκαφείς του Σλήμαν είχαν αποφέρει έστω και μια επιγραφή. Μπορεί να αμφισβητηθεί το αν η γνώμη του Evans ήταν δικαιολογημένη, αλλά η διαίσθηση του αποδείχθηκε ότι ήταν ορθή, και ήταν αυτός που πέτυχε να βρει την απόδειξη, παρόλο που αυτό τον οδήγησε σε τόσο σημαντικές καινούργιες ανακαλύψεις ώστε αργότερα να χάσει το ενδιαφέρον του που τον είχε παρακινήσει στην έρευνά του.

Πρόσεξε ότι οι πωλητές αρχαιοτήτων στην Αθήνα πωλούσαν κάποτε εγχάρακτες πέτρες, που προορίζονταν σαφώς για να χρησιμοποιούνται ως σφραγίδες. Δεν έμοιαζαν με οποιεσδήποτε μεταγενέστερες σφραγίδες και καλύπτονταν με μικρές εικόνες αντικειμένων καταταγμένων με έναν τέτοιο τρόπο ώστε να αποτελούν μάλλον ένα σύστημα γραφής. Ίσως είναι δύσκολο σε τέτοιες περιπτώσεις να είναι κανείς βέβαιος κατά πόσο τα σύμβολα είναι πράγματι γραφή, ή είναι απλώς μια εικονογραφική αναπαράσταση ενός ονόματος. Αλλά ο Evans θεώρησε ότι το σύστημα σε αυτούς τους σφραγιδολίθους έμοιαζε περισσότερο με γραφή κι οι έρευνες του τον οδήγησαν στο συμπέρασμα ότι προέρχονταν από τη μεγαλόνησο Κρήτη.

Την εποχή αυτή η Κρήτη κατεχόταν ακόμη από τους Τούρκους και διαδοχικές επαναστάσεις των ελλήνων στη διάρκεια του 19^{ου} αιώνα υπήρξαν ανεπιτυχείς ίσαμε το 1899, όταν τελικά οι Τούρκοι αποσύρθηκαν. Ο Evans είχε ήδη ταξιδεύσει σε μεγάλο μέρος του νησιού και είχε αποφασίσει που να σκάψει. Ο αρχαιολογικός χώρος που διάλεξε ήταν η Κνωσός, μερικά μίλια στο εσωτερικό από την κύρια πόλη του νησιού, που τώρα είναι γνωστή ως Ηράκλειο. Οι ελληνικές παραδόσεις μιλούσαν για ένα βασιλιά Μίνωα που στους προϊστορικούς χρόνους είχε κυβερνήσει από την Κνωσό μια θαλασσινή αυτοκρατορία στο Αιγαίο. Φαινόταν επομένως ένας χώρος που υποσχόταν πολλά αν ανασκαπτόταν και οι εντόπιοι ανασκαφείς είχαν ήδη αποκαλύψει εκεί ενδιαφέροντα ευρήματα. Όταν οι Τούρκοι έφυγαν, ο Evans κατόρθωσε να τον αγοράσει κι άρχισε εκεί ανασκαφές το 1900. Έγινε αμέσως σαφές ότι είχε βρει ένα μείζονα χώρο της εποχής του χαλκού και ανταμείφτηκε με την ανακάλυψη μεγάλου αριθμού ενεπίγραφων πήλινων πινακίδων. Η γραφή ήταν πιο πολύ αναπτυγμένη από ό,τι σε σφραγιδολίθους και δεν μπορούσε να υπάρξει αμφιβολία ότι επρόκειτο για αληθινή γραφή. Αλλά οι χαρακτήρες δεν έμοιαζαν με καμιά από τις γνωστές γραφές, και παρόλο που ο Evans ξεκίνησε με μεγάλες ελπίδες για την αποκρυπτογράφηση της, το έργο του τελείωσε αρκετά γρήγορα, γιατί επισκιάστηκε από περισσότερο συναρπαστικά ευρήματα.

Ο Evans είχε ελπίσει να βρει ένα μυκηναϊκό χώρο στην Κρήτη να ανταγωνίζεται τις Μυκήνες στην ηπειρωτική χώρα. Βεβαίως, το τεράστιο συγκρότημα κτηρίων που ανασκαψε στην Κνωσό πρέπει να υπήρξε ένα μείζον ανάκτορο και η περίοδος ακμής του τοποθετούνταν στη Μυκηναϊκή περίοδο. Αλλά ήταν διαφορετικού τύπου από το χώρο των Μυκηνών και αυτό που ήταν εντελώς απροσδόκητο ήταν ότι χρονικά έφθανε πολύ πιο πίσω. Ο βασιλιάς της Κνωσού ζούσε με κάποια πολυτέλεια πολύ πριν χτιστούν τα τείχη των Μυκηνών. Στην πραγματικότητα, τώρα γίνεται αποδεκτό ότι στην Κρήτη υπήρχε πολιτισμός σε υψηλό επίπεδο έως και διακόσια χρόνια πριν η ηπειρωτική χώρα αρχίσει να τη μιμείται. Δεν ήταν πια δυνατό αυτός να αποκαλείται μυκηναϊκός και ο Evans δημιούργησε το νέο όρο «Μινωικός» για να περιγράψει τον πολιτισμό της εποχής του χαλκού της Κρήτης.

Πολλοί αρχαιολόγοι ακολούθησαν τον Evans στην Κρήτη και νέα σημαντικά ανάκτορα ανασκάφθηκαν στη Φαιστό στα νότια και τα Μάλλια ανατολικότερα κατά μήκος της βόρειας παραλίας. Και οι δύο αυτοί χώροι και μερικοί άλλοι, ιδίως η Αγία Τριάδα, μερικά μόνο μίλια μακριά

από τη Φαιστό, απέδωσαν μικρές ποσότητες πήλινων πινακίδων, αλλά αυτές ήταν μάλλον διαφορετικές από εκείνες της Κνωσού. Στην αρχή ο Evans νόμισε ότι αυτό μπορούσε να οφειλόταν σε μια ειδική βασιλική γραφή στην Κνωσό, αλλά αργότερα προσέχτηκε ότι οι διαφορές σχετίζονταν με τη χρονολογία. Οι πρωιμότερες επιγραφές ήταν αυτές των σφραγιδόλιθων, που σπάνια βρισκόνταν σε πηλό, ο Evans ονόμασε αυτή τη γραφή «ιερογλυφική» εξαιτίας μιας υποτιθέμενης ομοιότητας με την πρώιμη αιγυπτιακή γραφή γνωστή με το όνομα αυτό, αλλά δεν υπάρχει λόγος να πιστέψουμε ότι συγγενεύουν. Λίγο αργότερα οι εικόνες των αντικειμένων τυποποιούνται περισσότερο κι έτσι γίνονται λιγότερο αναγνωρίσιμες, ιδιαίτερα όταν γράφονται σε πηλό. Ο Evans ονόμασε τη γραφή αυτή γραμμική Α, επειδή τα σύμβολα ήταν απλά περιγράμματα. Λίγα δείγματα από αυτή βρέθηκαν στην Κνωσό, επειδή εκεί ο όγκος των επιγραφών ανήκε σε μεταγενέστερη παραλλαγή της γραφής αυτής, την οποία ονόμασε γραμμική Β. Αυτή περιορίστηκε στην τελευταία φάση της νεοανακτορικής περιόδου, που μπορούμε να χρονολογήσουμε το 1450-1375 π.χ. περίπου.

Οι πήλινες πινακίδες δεν είχαν ψηθεί τότε που έγιναν, αλλά ξεράθηκαν στον ήλιο, ώστε διατηρήθηκαν μόνο αν έτυχε να βρίσκονται σε κτήριο που είχε καεί. Έτσι πινακίδες βρέθηκαν μόνο σε στρώματα όπου έγινε καταστροφή και θα πρέπει να χρονολογούνται εντελώς προς το τέλος της περιόδου που χρησιμοποιήθηκε το κτήριο. Ατυχώς, από την εποχή του Evans συνεχίζεται μια παρατεταμένη συζήτηση για την χρονολογία της τελικής καταστροφής της Κνωσού και παρόλο που οι περισσότεροι επιστήμονες δέχονται μια χρονολογία κάπου γύρω στο 1375 π.χ., έχει προταθεί σοβαρά ότι θα ήταν δυνατή μια χρονολογία στον 13^ο αιώνα. Δεν υπάρχει τίποτα ανάμεσα στα ευρήματα που να καθορίζει τη συζήτηση, αλλά γενικά μια χρονολογία μέσα στον 14^ο αιώνα δείχνει ακόμη πιθανότητες.

Ο Evans μελέτησε τις πήλινες του πινακίδες της γραμμικής Β και έβγαλε μερικά προφανή συμπεράσματα. Αλλά αν και ετοίμασε μια έκδοση όλου του τότε γνωστού ιερογλυφικού υλικού, η ετοιμασία του για την έκδοση των πινακίδων της γραμμικής Β δεν είχε ακόμη συμπληρωθεί όταν η έκρηξη των βαλκανικών πολέμων και ύστερα ο πρώτος παγκόσμιος πόλεμος έστρεψαν την προσοχή του σε άλλα ζητήματα. Μετά τον πόλεμο δημοσίευσε το εκτενές του έργο για το ανάκτορο της Κνωσού, που με αυτοπεποίθηση ονόμασε ανάκτορο του Μίνωα. Περιλάμβανε ένα τμήμα για τις πινακίδες και ορισμένες από αυτές εικονογραφούνταν, αλλά ο τεράστιος όγκος των πινακίδων έμενε ακόμη αδημοσίευτος κι επομένως απρόσιτος στους επιστήμονες. Μόλις το 1952, έντεκα χρόνια μετά το θάνατο του Evans, ένας παλιός του φίλος και συνάδερφος, ο Sir John Myres, δημοσίευσε τελικά τον τόμο που ο Evans είχε σχεδιάσει και κατά ένα μεγάλο μέρος συντάξει γύρω στο 1911-12.

Η μεγάλη αυτή καθυστέρηση ήταν για πολλούς λόγους ατυχής. Μια πρωιμότερη δημοσίευση θα έκανε προσιτές πάρα πολλές πληροφορίες, ώστε να μπορούσε να αρχίσει πολύ νωρίτερα σοβαρή δουλειά για την αποκρυπτογράφηση, ακόμη κι εκείνοι που κατάφεραν να δουν το υλικό στο μουσείο του Ηρακλείου εμποδίζονταν από τον κανόνα ότι κανένας δεν μπορεί να προλάβει στη δημοσίευση την πρώτη δημοσίευση των ευρημάτων από τον ευρετή. Όταν έγινε δυνατό στους επιστήμονες να εργαστούν στο πρωτότυπο υλικό, αποκαλύφθηκε γρήγορα ότι η έκδοση ήταν ελαττωματική και ατελής. Τρεις χωριστές συλλογές τμημάτων πινακίδων που βρέθηκαν στις ανασκαφές του Evans έχουν έκτοτε έρθει στο φως στο μουσείο του Ηρακλείου, αλλά καμία από αυτές δεν εμφανίζεται στη δημοσίευση του 1952. Η μελέτη τους αποκάλυψε αμέσως ότι δεν είχε γίνει σοβαρή προσπάθεια για να ενωθούν τα θραύσματα κι έτσι να αποκατασταθούν οι πινακίδες ώστε να είναι πλήρεις, ή μάλλον σχεδόν πλήρεις. Το καθήκον αυτό επωμίστηκε μια αφοσιωμένη παρέα επιστημόνων διαφόρων εθνικοτήτων, οι οποίοι εργάστηκαν μαζί ως ομάδα για να αποκαταστήσουν τις πινακίδες και να δημοσιεύσουν ένα πλήρες και αξιόπιστο κείμενο.

Αλλά από το 1939 μια νέα μεγαλύτερη ανακάλυψη είχε μεταβάλει ολοσχερώς την κατάσταση όσον αφορά τη γραμμική Β. Το έτος εκείνο μια από κοινού Αμερικανό-Ελληνική αποστολή υπό τον Carl W. Blegen του Πανεπιστημίου του Σινσινάτι άρχισε να σκάβει ένα χώρο στα νότιο-δυτικά της ηπειρωτικής Ελλάδας, λίγο βόρεια της σύγχρονης πόλης της Πύλου, ακριβώς στο εσωτερικό από τον Κόλπο του Ναβαρίνου, ενός από τα πιο ωραία φυσικά λιμάνια της μεσογείου. Αποδείχθηκε ότι ήταν ένα μυκηναϊκό ανάκτορο που καταστράφηκε από φωτιά στο τέλος του 13^{ου} αιώνα π.χ. Κατά ευτυχή συγκυρία η πρώτη τάφος που ανοίχτηκε από τους ανασκαφείς συνάντησε ό,τι γνωρίζουμε τώρα ως Αρχείο, αφού περιείχε εκατοντάδες πήλινων πινακίδων, που σκληρύνθηκαν από φωτιά που το κατέστρεψε. Μόλις συνέλεξαν τα πρώτα θραύσματα από το έδαφος, έγινε δυνατό να αναγνωριστεί ότι ήταν γραμμένα στην ίδια γραμμική Β, γραφή που ήταν ήδη καλά γνωστή από την Κνωσό.

Η είδηση αυτή ίσως δεν δημιούργησε την αίσθηση που θα έπρεπε να είχε προξενήσει, καθώς ο κόσμος το 1939-40 είχε στο νου του πιο σοβαρά ζητήματα από τη γραφή της εποχής του χαλκού. Για πρώτη φορά βλέπαμε ότι η γραφή γραμμική Β δεν περιοριζόταν στην Κνωσό, ή ακόμα στην Κρήτη, αλλά ότι χρησιμοποιούνταν στην ηπειρωτική χώρα, γιατί είναι εντελώς απίθανο ένα τέτοιο αρχείο να μεταφέρθηκε από εκεί όπου είχε γραφεί. Ωστόσο αν ήταν απλώς, όπως ο Evans διαβεβαίωνε, μια τροποποιημένη εκδοχή της γραμμικής Α, μιας καθαρά κρητικής γραφής, σήμαινε αυτό ότι η γλώσσα της Κρήτης χρησιμοποιούνταν επίσης στην ηπειρωτική Ελλάδα; Αποτελούσε την απόδειξη που ο Evans αναζητούσε για να δείξει ότι όλη η νότια Ελλάδα ήταν κάποτε υπό Μινωικό έλεγχο;

Η δυσκολία ήταν ότι οι περισσότεροι επιστήμονες την εποχή αυτή πίστευαν ότι οι μυκηναίοι που ήταν γνωστοί στους αρχαιολόγους ήταν οι Αχαιοί που περιέγραψε ο Όμηρος ως κύριους της Ελλάδας την εποχή του Τρωικού Πολέμου. Φυσικά οι ποιητές, όπως οι μυθιστοριογράφοι, έχουν την τάση να κάνουν τους χαρακτήρες τους να μιλούν τη δική τους γλώσσα, αλλά το γεγονός ότι οι περισσότεροι από τους χαρακτήρες του Ομήρου έχουν ονόματα που κάτι σημαίνουν στα ελληνικά υπονοούσε ότι η ελληνική μιλιόταν ήδη στην Ελλάδα κατά τη μυκηναϊκή εποχή, αν οι ιστορίες του Ομήρου δεν ήταν καθαρός μύθος. Έτσι τι έκανε ο βασιλιάς της μυκηναϊκής Πύλου, ο Νέστορας, αν μπορούσε να εμπιστευθούμε τον Όμηρο, κρατώντας τους λογαριασμούς του σε μια ξένη γλώσσα; Μια εύκολη απάντηση στο ερώτημα δίνεται από το παράλληλο του μεσαίωνα, όταν βασιλιάδες σε όλη την Ευρώπη κρατούσαν τους λογαριασμούς τους στα λατινικά, όποια γλώσσα κι αν μιλούσαν οι ίδιοι. Ωστόσο κι άλλες ανακαλύψεις από άλλους χώρους στην ηπειρωτική Ελλάδα έχουν τώρα αλλάξει εντελώς την εικόνα. Η γραμμική Β θεωρείται τώρα ότι είναι η γραφή των μυκηναϊκών ανακτόρων στην ηπειρωτική Ελλάδα, κι αυτό που χρειάζεται εξήγηση είναι η διείσδυση της στην Κρήτη. Η λύση στο πρόβλημα ήρθε το 1952-3, με την απόδειξη ότι η γλώσσα των πινακίδων της γραμμικής Β ήταν ελληνική. Ο Evans θα δοκίμαζε βαθιά έκπληξη μαθαίνοντας ότι το μινωικό του ανάκτορο στην τελευταία φάση της ύπαρξης του είχε κάνει χρήση της ελληνικής γλώσσας.

Πριν ακόμη βρεθούν οι πινακίδες της γραμμικής Β στην ηπειρωτική Ελλάδα, ήταν γνωστό ότι μεγάλα πιθάρια με ζωγραφισμένες επιγραφές σε αυτή τη γραφή είχαν βρεθεί στην ηπειρωτική Ελλάδα. Η μεγαλύτερη συλλογή προέρχεται από τις Θήβες, στα βορειοδυτικά της Αθήνας, αλλά και άλλοι σύγχρονοι χώροι έχουν προμηθεύσει δείγματα. Τα πιθάρια αυτά χρησιμοποιούνταν συχνά για την μεταφορά ελαιολάδου και κρασιού, και υπέθεσαν ότι ήταν δοχεία για τα κρητικά προϊόντα. Η υπόθεση αυτή έχει τώρα επιβεβαιωθεί με τρόπο αξιοσημείωτο. Παρατηρήθηκε ότι μερικές από τις λέξεις πάνω στα πιθάρια βρίσκονταν επίσης στις πινακίδες της γραμμικής Β της Κνωσού, στις οποίες φαίνονταν να είναι τοπωνύμια, και θα ήταν φυσικό για τον εξαγωγέα να αναγράψει το όνομά του και την διεύθυνσή του στο προϊόν του. Αλλά πολύ πρόσφατα μια ανάλυση του πυλού που χρησιμοποιήθηκε για να κατασκευάσουν αυτά τα πιθάρια αποκάλυψε ότι σχεδόν σίγουρα αυτά προέρχονται από την Κρήτη.

Λίγες πήλινες πινακίδες έχουν βρεθεί στις ίδιες τις Μυκίνες, το πρώτο σημαντικό εύρημα του άγγλου αρχαιολόγου Α.Ι.Β. Wace το 1952, ο οποίος ανάσκαψε μερικές μεγάλες οικίες έξω από τα τείχη της Ακρόπολης, και βρήκε σε αυτές συλλογές πινακίδων γραμμικής Β. Αυτό δεν αποδεικνύει ότι η γραμμική Β ήταν σε ευρεία χρήση σε ολόκληρο τον πληθυσμό, γιατί τέτοιες οικίες θα πρέπει να κατοικούνταν από μέλη του βασιλικού οίκου. Κάπως περισσότερες, μάλλον πολύ κατεστραμμένες πινακίδες προήλθαν από μια οικία μέσα στα τείχη, αλλά δεν βρέθηκε ίχνος από το κύριο αρχείο του ανακτόρου. Αφού το ανάκτορο βρίσκεται στην κορυφή του λόφου, ο χώρος του για πολύ χρόνο είχε μείνει εκτεθειμένος στις καιρικές συνθήκες, και τα αρχεία του κατά πάσαν πιθανότητα έχουν καταστραφεί. Αλλά είναι φρόνιμο να σκεφτούμε ότι αν ο Σλήμαν ήξερε για ποιο πράγμα να ψάξει, μπορεί να ήταν ο πρώτος που θα έβρισκε πινακίδες της γραμμικής Β. Καθώς εμφανίζονται μέσα από το έδαφος, είναι πολύ εύκολο να θεωρήσει κανείς τα θραύσματα πινακίδων ως κομμάτια χονδροειδών αγγείων, που οι πρώτοι ανασκαφείς απέρριπταν χωρίς καμία σκέψη.

Στην Τίρυνθα, μόνο λίγα χιλιόμετρα μακριά από τις Μυκίνες, υπήρχε ένα πελώριο κάστρο με ογκώδη τείχη. Μπορεί να είχε σκοπό να φρουρεί το λιμάνι, αλλά η θάλασσα έχει τώρα αποτραβηχτεί από την περιοχή. Δύσκολα θα πιστεύαμε ότι αποτελούσε, όπως υπονοεί ο Όμηρος, την έδρα ενός ανεξάρτητου βασιλείου. Πρέπει να ήταν κατά κάποιον τρόπο υπό τον έλεγχο του βασιλιά των Μυκηνών, που μπορεί να ήταν επικυρίαρχος έχοντας εξασφαλίσει την υποταγή διαφόρων μικρότερων ηγεμόνων. Από το 1971 και μετά ανασκαφές στην κάτω πόλη έξω από τα τείχη του κάστρου

αποκάλυψαν μερικές αποσπασματικές πινακίδες της γραμμικής Β. Φαίνεται σαν αυτές να κατέβηκαν από την αρχική τους θέση κι ό,τι μπορούμε να πούμε τη στιγμή αυτή είναι ότι κάπου σε αυτόν το χώρο πρέπει να υπήρξε ένα μείζων αρχείο, από το οποίο όμως μόνο θραύσματα έχουν αποκαλυφθεί.

Η κατάσταση στις Θήβες είναι μάλλον διαφορετική. Το πρόβλημα εδώ είναι ότι ο ίδιος χώρος έχει συνεχώς κατοικηθεί για τουλάχιστον 4000 χρόνια και σήμερα είναι μια ακμάζουσα επαρχιακή πόλη. Αυτή έχει χτιστεί πάνω σε διαδοχικά στρώματα κατοίκησης, τουρκική, φράγκικη, βυζαντινή, ρωμαϊκή, ελληνιστική, κλασική, αρχαϊκή, μυκηναϊκή και ακόμα παλαιότερης. Σπάνια είναι δυνατό εδώ να βρει κανείς μέρος για να σκάψει, και μόνον όταν είναι ανάγκη να χτιστεί ένα καινούργιο οικοδόμημα έχουν οι αρχαιολόγοι τη δυνατότητα να ερευνήσουν τι βρίσκεται κάτω από το έδαφος. Σωστικές ανασκαφές αυτού του είδους έχουν ως τώρα αποδώσει δύο μικρές συλλογές πινακίδων γραμμικής Β, και μια ομάδα πήλινων σφραγίδων, μικρούς σβόλους πηλού σφραγισμένους με σφραγίδα και σε μερικές περιπτώσεις ενεπίγραφους με λίγες λέξεις στη γραμμική Β. Η μαρτυρία αυτή καθιστά σχεδόν βέβαιο ότι κάπου κάτω από το κέντρο της σύγχρονης πόλης υπάρχει ένα αρχείο πινακίδων. Οι Θήβες ήταν ολοφάνερα η τοποθεσία ενός ανακτόρου που ήλεγχε ένα μεγάλο βασίλειο σε αυτή την περιοχή της Ελλάδας, και τα αρχεία του θα ήταν πολύ σημαντικά, αν μονάχα μπορούσαμε να τα εντοπίσουμε. Αλλά για την ώρα πολύ λίγα μπορούμε να πούμε για αυτό το βασίλειο.

Τα ευρήματα δείχνουν ότι η γραφή δεν ήταν σε ευρεία χρήση στη μυκηναϊκή Ελλάδα. Δεν έχουν βρεθεί πινακίδες σε μικρότερους χώρους κι όλες εκείνες που έχουν βρεθεί είναι είτε σε ανάκτορα ή τόσο κοντά σε ανάκτορα ώστε μπορούν να θεωρηθούν ως εξαρτήματα τους. Δεν υπάρχει ίχνος οποιασδήποτε ιδιωτικής χρήσης γραφής. Αυτό έρχεται σε αντίθεση με την ιστορία του ελληνικού αλφαβήτου, το οποίο ήδη τον 8^ο αιώνα π.χ. χρησιμοποιούνταν από ιδιώτες για να γράψουν ξένοιαστους στίχους πάνω σε ένα αγγείο, και στους 2 επόμενους αιώνες άρχισε να τίθεται σε χρήση για νόμους που γράφονταν πάνω σε πέτρα σε μέρη όπου όλοι μπορούσαν να τους διαβάσουν. Τίποτα τέτοιο δεν έχει βρεθεί στη γραμμική Β. Η γραφή φαίνεται ότι ήταν αποκλειστικά γραφειοκρατικό εργαλείο, αναγκαία μέθοδος για να κρατούν διοικητικούς λογαριασμούς και αρχεία, αλλά ποτέ δεν χρησιμοποιήθηκε για ιστορικούς ή ακόμη για ελαφρότερους σκοπούς. Όπως θα δούμε, το περιεχόμενο των πινακίδων της γραμμικής Β είναι σχεδόν χωρίς εξαίρεση κατάλογοι ανθρώπων, ζώων, αγροτικών προϊόντων και αντικειμένων βιοτεχνίας.

1.2 Η χρήση της γραφής στους μινωίτες

Οι έρευνες στην Κρήτη προς το τέλος του περασμένου αιώνα και τις αρχές του εικοστού αποκάλυψαν ότι ένας ιδιαίτερος πολιτισμός είχε ακμάσει εκεί κατά την εποχή του χαλκού. Η αρχαιολογική σκαπάνη έφερε στο φως μεγάλα συγκροτήματα κτιρίων (που συμβατικά αποκαλούνται ανάκτορα): στην Κνωσό, στη Φαιστό και στα Μάλια. Άλλες σημαντικές θέσεις που έγιναν γνωστές ως τις αρχές της δεκαετίας του 1900 είναι το Παλαίκαστρο, τα Γουρνιά, η Τύλισος και η Αγία Τριάδα. Πιο πρόσφατα, δύο μείζονες οικισμοί ανακτορικού χαρακτήρα αποκαλύφθηκαν στη Ζάκρο και στα Χανιά. Η μινωική περίοδος διαιρέθηκε από τον Evans σε τρεις χρονολογικές φάσεις: την πρωτομινωική (ΠΜ), τη μεσομινωική (ΜΜ) και την υστερομινωική (ΥΜ). Μόνο στην μεσομινωική και στην υστερομινωική περίοδο χρησιμοποιήθηκε η γραφή στην περιοχή του Αιγαίου.

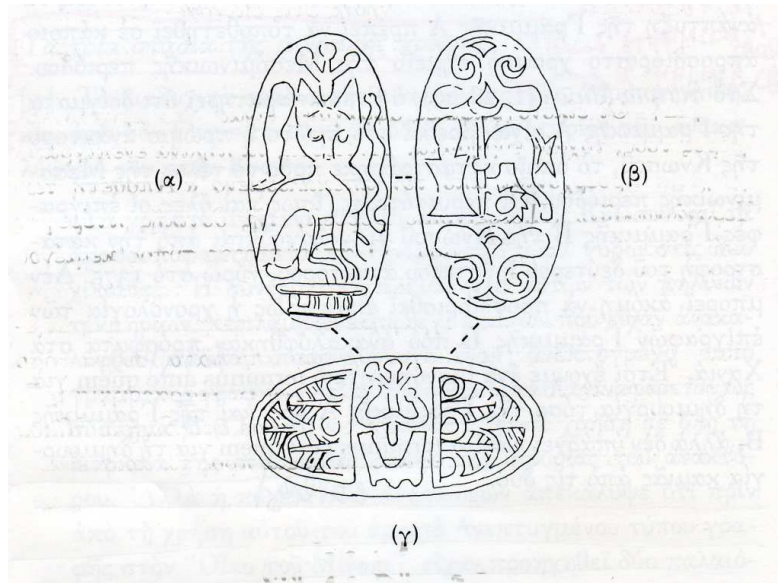
Οι μινωίτες ήταν οι εφευρέτες της γραφής στο χώρο του Αιγαίου. Όχι μόνο επινόησαν αυτή την τέχνη, αλλά και σε όλη τη διάρκεια της εποχής του χαλκού διατήρησαν την πρωτοβουλία για την ανάπτυξη και βελτίωσή της. Τα μινωικά γραφικά συστήματα παραλήφθηκαν συν τω χρόνω και από πολλούς λαούς, όπως τους κύπριους και τους μυκηναίους, οι οποίοι τα χρησιμοποίησαν για τους δικούς τους σκοπούς. Το ερώτημα αν και κατά πόσον υπήρξαν οι μινωίτες οφειλέτες σε κάποιο ξένο γραφικό σύστημα κατά τη δημιουργία των δικών τους συστημάτων γραφής δεν μπορεί ακόμα να απαντηθεί ικανοποιητικά: φαίνεται όμως βέβαιο ότι μερικά από τα πρωιμότερα σύμβολα που χρησιμοποιήθηκαν στη μινωική γραφή προέρχονταν από την Αίγυπτο.

Δυο θεμελιακά ερωτήματα θα πρέπει να απαντηθούν: σε ποιές περιπτώσεις χρησιμοποιούσαν οι λαοί του Αιγαίου τη γραφή; Ποια στάδια της ιστορίας της γραφής αντιπροσωπεύουν τα διάφορα γραφικά τους συστήματα; Η πιθανή απάντηση στο πρώτο ερώτημα είναι ότι η επινόηση της γραφής στην Κρήτη δεν υπαγορεύτηκε από κάποια επιθυμία επικοινωνίας, αλλά από οικονομική αναγκαιότητα. Φαίνεται ότι οι Μινωίτες, και μετά από αυτούς οι Μυκηναίοι, χρησιμοποιούσαν τη γραφή για δύο κύριους σκοπούς: για τον προσδιορισμό της ιδιοκτησίας και για την τήρηση αρχείων.

Μόνον οι επιγραφές που βρέθηκαν σε κάποια μινωικά λατρευτικά αντικείμενα, και ειδικά στις αποκαλούμενες «τράπεζες προσφορών» ξεφεύγουν κάπως από αυτές τις δύο κατηγορίες. Δεν διαθέτουμε, προς το παρόν, καμιά απτή απόδειξη ότι οι Μινωίτες ή οι Μυκηναίοι έγραψαν ποτέ κάποιο λογοτεχνικό ή ιστορικό κείμενο, ή ακόμη ότι έστειλαν κάποια επιστολή από μια πόλη σε άλλη, ωστόσο οι αναλογίες με άλλους σύγχρονους πολιτισμούς καθιστούν πολύ πιθανό να χρησιμοποιούσαν πράγματι τη γραφή και για τέτοιους σκοπούς. Σε όλη την περίοδο, και ειδικά κατά την ύστερη εποχή του χαλκού, η κρητομυκηναϊκή γραφή περιορίστηκε σε ελάχιστα κέντρα, μπορούμε άλλωστε να υποθέσουμε ότι ο αριθμός των ανθρώπων που γνώριζαν γραφή και ανάγνωση ήταν σε όλες τις περιόδους πολύ περιορισμένος.

Τα τρία στάδια της μινωικής γραφής: τα διαδοχικά στάδια της μινωικής γραφής αποσαφηνίστηκαν από τον Evans το 1909 με βάση τις μαρτυρίες των ανακαλύψεων του στην Κνωσό. Τα γραπτά μνημεία από το ανάκτορο της Κνωσού και τα άμεσα παραρτήματα του ανέρχονται τώρα γύρω στις δύο χιλιάδες. Η συντριπτική πλειοψηφία αυτών των πήλινων τεκμηρίων, περιλαμβανομένων και αυτών που είχαν ανακαλυφθεί πρώτα, παρουσιάζουν ένα ανεπτυγμένο τύπο γραμμικής γραφής, η οποία ήταν σε χρήση σε όλη τη διάρκεια της τελευταίας φάσης της ιστορίας του ανακτόρου. Αλλά η πορεία των ανασκαφών αποκάλυψε ότι πριν από τη χρήση αυτού του αρκετά ανεπτυγμένου τύπου γραφής στον οίκο του Μίνωα είχαν προηγηθεί 2 παλαιότεροι τύποι: ένας τύπος που περιγράφεται ως σύστημα A(class A), με γραμμικούς επίσης χαρακτήρες, κι ένας άλλος, ακόμη πρωιμότερος, τύπος συμβατικής εικονογραφικής μορφής που θυμίζει τα αιγυπτιακά ιερογλυφικά. Ο Evans ονόμασε το πρώτο στάδιο της κρητικής γραφής εικονογραφικό ή ιερογλυφικό χωρίς διάκριση(χρησιμοποιείται μόνο ο όρος εικονογραφικό για να αποφεύγεται η σύγχυση με τα αιγυπτιακά ή χεττιτικά ιερογλυφικά). Η γενική εικόνα που έδωσε ο Evans σε αυτή την πρώιμη δημοσίευση είναι αξιοθαύμαστα σαφής, και στις γενικές της γραμμές ισχύει ακόμα και σήμερα. Στην Κρήτη της μέσης και ύστερης εποχής του χαλκού εμφανίστηκαν τρία μεγάλα συστήματα γραφής. Το πιο πρώιμο (εικονογραφικό) άνησε από τις αρχές αυτής της περιόδου. Η ανάπτυξη της γραμμικής A πρέπει να τοποθετηθεί σε κάποιο απροσδιόριστο χρονικό σημείο της μεσομινωικής περιόδου. Ο Evans παρατηρεί ότι δείγματα της γραμμικής A είχαν βρεθεί στο πρώτο ή πρώιμο ανάκτορο της Κνωσού, το οποίο καταστράφηκε προς το τέλος της μεσομινωικής περιόδου. Οι περισσότερες, ίσως και όλες οι επιγραφές της γραμμικής B της Κνωσού χρονολογούνται από την καταστροφή του δεύτερου ή ύστερου ανακτόρου, γύρω στο 1375. Δεν μπορεί ακόμη να προσδιοριστεί επακριβώς η χρονολογία των επιγραφών της γραμμικής B που ανακαλύφθηκαν πρόσφατα στα Χανιά. Έτσι έχουμε ένα ικανοποιητικό terminus ante quem για τη δημιουργία τόσο της γραμμικής A όσο και της γραμμικής B, αλλά δεν υπάρχει κανένα terminus post quem για τη δημιουργία καμιάς από τις δύο.

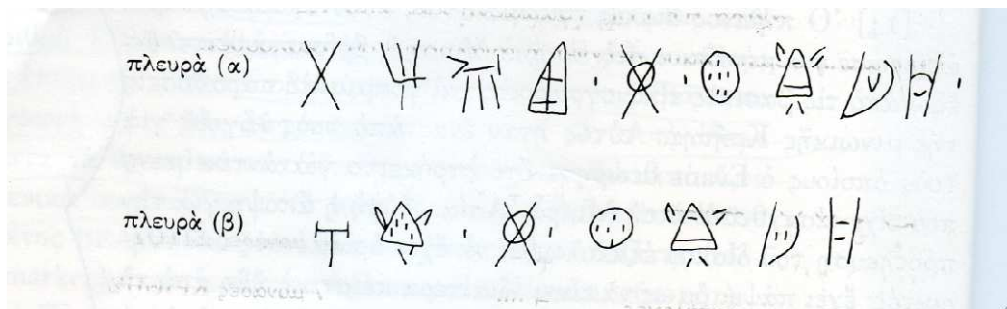
Η εικονογραφική γραφή: οι κρητικές εικονογραφικές επιγραφές ανήκουν κυρίως σε δύο είδη, αυτές που έχουν σκαλισθεί σε σφραγιδόλιθους και αυτές που έχουν χαραχθεί σε τεμάχια πηλού, όπως πινακίδες, δελτάρια και τετράπλευρες ράβδους. Οι επιγραφές σε σφραγίδες ίσως να μην ανήκουν σε κάποιο διαφορετικό χρονολογικό στάδιο από εκείνο των επιγραφών σε πηλό: είναι πιθανό ότι η μεγαλύτερη σχηματοποίηση των εικόνων στις πρώτες οφείλεται στο δυσκατέργαστο του υλικού που χρησιμοποιήθηκε. Μια εντυπωσιακή σφραγίδα ενισχύει την υπόθεση ότι αυτός ο τύπος επιγραφής διατηρούσε ακόμη το ρόλο του προσδιοριστικού ιδιοκτησίας.



Σχήμα 1.2.1: Σφραγίδα με κρητικά ιερογλυφικά

Ο Evans διατύπωσε την άποψη ότι οι ομάδες συμβόλων κάθε πλευράς αποδίδουν επίσημα ονόματα και τίτλους. Οι ίδιες ομάδες συμβόλων επαναλαμβάνονται και σε άλλα εικονογραφικά κείμενα, κάποτε σε διαφορετικούς συνδυασμούς ή σε διαφορετική διάταξη. Αν κρίνουμε από ανάλογα στάδια ανάπτυξης άλλων συστημάτων γραφής, απομένει μικρή αμφιβολία πως ορισμένα τουλάχιστον από τα σύμβολα της κρητικής εικονογραφικής γραφής είναι φωνητικά, δηλαδή αποδίδουν τον πραγματικό ήχο ή συλλαβές της ομιλούμενης γλώσσας. Η σφραγίδα αυτή είναι σημαντική για την μελέτη της γραφής κι από μια άλλη άποψη: η πλευρά(γ) εμφανίζει ένα αμάλγαμα από περισσότερα του ενός σύμβολα, έτσι που να σχηματίζεται ένα σύμπλεγμα ή μονόγραμμα – πρακτική συνήθης στις επιγραφές της γραμμικής Α και ίσως λιγότερο συχνή σε εκείνες της γραμμικής Β. Οι εικονογραφικές σφραγίδες είναι, στο μεγαλύτερο μέρος τους, τυχαία ευρήματα. Αντίθετα, τα πήλινα ενεπίγραφα μνημεία προέρχονται από τον αποκαλούμενο «αποθέτη των ιερογλυφικών» της δυτικής πτέρυγας του ανακτόρου της Κνωσού. Ο απλούστερος τύπος κειμένου από αυτόν τον αποθέτη δεν κάνει τίποτε άλλο παρά να καταγράφει έναν αριθμό εικονογραφημένων αντικειμένων, ανήκει επομένως στο είδος της απογραφής που απαντάται συχνά τόσο στη γραμμική Α όσο και στη γραμμική Β.

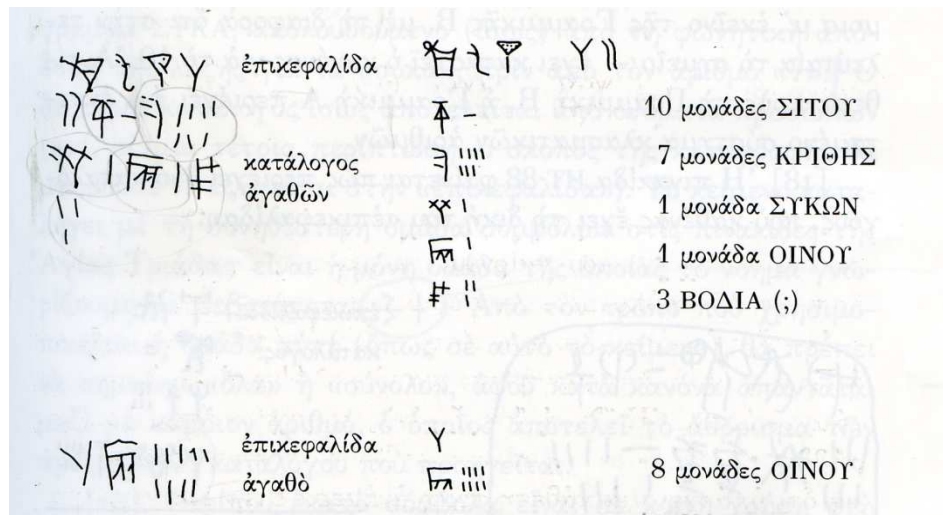
Η γραμμική γραφή Α: κατά τη μεσομινωική περίοδο, η εικονογραφική γραφή μετασηματίστηκε (ίσως στη διάρκεια μιας μακράς χρονικής περιόδου) σε ένα πιο κομψό, συστηματικότερο και πιο ρέον είδος γραφής, στην οποία τα εικονογραφικά στοιχεία τυποποιήθηκαν ακόμη περισσότερο. Η νέα αυτή γραφή ονομάστηκε από τον Evans γραμμική Α. εώς πρόσφατα πιστεύαμε ότι η γραμμική Α είχε αντικαταστήσει τα εικονογράμματα σε ολόκληρη την Κρήτη προς το τέλος της μεσομινωικής περιόδου. Χαρακτηριστικό δείγμα της μετάβασης από την εικονογραφική γραφή στη γραμμική Α αποτελούν οι επιγραφές σε πήλινη ράβδο η οποία βρέθηκε στα Μάλια και χρονολογείται σε κάποιο ύστερο στάδιο της μεσομινωικής περιόδου.



Σχήμα 1.2.2: Επιγραφές σε πήλινη ράβδο με γραμμική Α

Τα σύμβολα σε αυτή τη ράβδο έχουν πλέον μεγάλες ομοιότητες με το σχήμα που παίρνουν τελικά στη γραμμική Α, κι έτσι μπορεί να υποθεθεί ότι η διαδικασία ανάπτυξης της τελευταίας είχε αρχίσει λίγο πρωύτερα. Δεν υπάρχουν ενδείξεις ότι η γραμμική Α δημιουργήθηκε για να καλύψει της ανάγκες μιας γλώσσας. Η ίδια η ύπαρξη ενός τεκμηρίου όπως η πλήρην ράβδος των Μαλίων σημαίνει πως δεν θα πρέπει να θεωρήσουμε ότι υπάρχει μια κάθετη τομή ανάμεσα στην εικονογραφική γραφή και τη γραμμική Α, αλλά μάλλον ότι υπήρξε βαθμιαία μετάβαση από τη μια προς την άλλη. Οι ανακαλύψεις του Leví στη Φαιστό δείχνουν ότι εκεί η μετάβαση από την εικονογραφική γραφή στη γραμμική Α συντελέστηκε νωρίτερα από ό,τι στα βόρεια και ανατολικά της Κρήτης.

Το συνολικά γνωστό υλικό σε γραμμική Α είναι πολύ μικρό. Εκτός από λίγες επιγραφές σε λίθινες τράπεζες προσφορών κι άλλα λατρευτικά αντικείμενα, τα περισσότερα από τα κείμενα σε γραμμική Α που διαθέτουμε προέρχονται από το αρχείο της Αγίας Τριάδας, θέση κοντά στη Φαιστό, που καταστράφηκε από φωτιά στη διάρκεια του 15^{ου} αιώνα π.χ.. Το αρχείο αποτελείται από 150 περίπου μικρές πλήρινες πινακίδες (μερικές από τις οποίες τώρα είναι πολύ αποσπασματικές), που καταγράφουν δοσοληψίες ή εναποθηκεύσεις. Μερικά παραδείγματα θα διευκρινίσουν τους κύριους τύπους επιγραφών που υπήρχαν στα αρχεία της Αγίας Τριάδας. Η παρακάτω πινακίδα, γραμμένη και από τις δύο πλευρές, έχει απλή δομή, όπως δείχνει η ανάλυση που ακολουθεί:

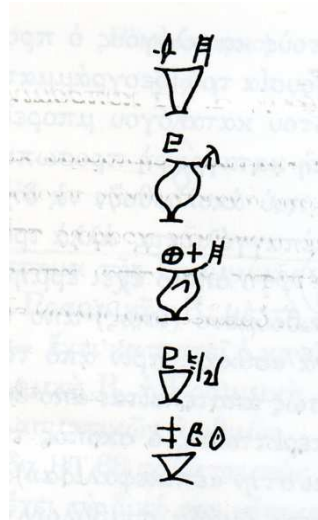


Σχήμα 1.2.3: Πινακίδα HT 114 γραμμένη σε γραμμική α

Είναι γνωστό τι εκφράζουν οι αντίστοιχες «επικεφαλίδες», αλλά ο αριθμός των πιθανών νοημάτων είναι πολύ περιορισμένος: έτσι, θα μπορούσαν να δηλώνουν τον τόπο όπου παρήχθησαν, συλλέχθηκαν ή καταγράφηκαν τα αγαθά, ή τα πρόσωπα που αναμείχθηκαν στην παραγωγή, διανομή ή παράδοση τους. Το νόημα των ιδεογραμμάτων της γραμμικής Α δεν είναι γνωστό με βεβαιότητα, και για τους σκοπούς αυτής της εργασίας ή χρήσης τους έχει συναχθεί από τον τρόπο με τον οποίο χρησιμοποιούνται παρόμοια ιδεογράμματα στη γραμμική Β. Το αριθμητικό σύστημα μπορεί να συγκριθεί με εκείνο της εικονογραφικής γραφής:

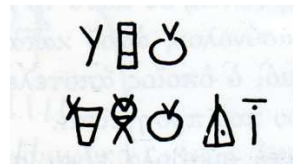
	Εικονογραφική	Γραμμική Α
1	∩ ῥ	
10	•	• ῥ -
100	∖	o
1000	◇	⊕

Σχήμα 1.2.4: : Το αριθμητικό σύστημα



Σχήμα 1.2.6: Ιδεογράμματα

Η πινακίδα αυτή περιέχει μια «επικεφαλίδα» άδηλης σημασίας, και ακολουθεί ένας κατάλογος αγγείων. Εφόσον όλα τα ιδεογράμματα παριστάνουν προφανώς αγγεία διαφόρων σχημάτων, ενδέχεται οι ομάδες συμβόλων



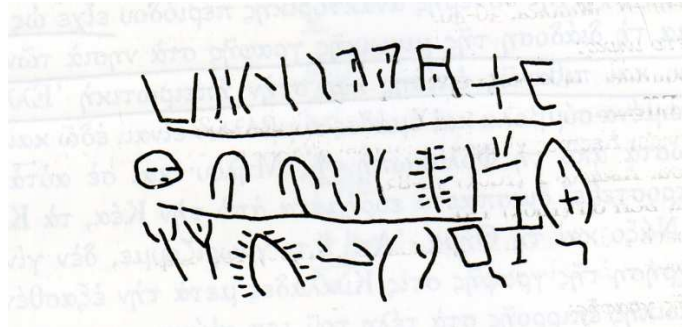
Σχήμα 1.2.7: ομάδες συμβόλων

να αποδίδουν ομοίως τα ονόματα των αγγείων, μόνο που σε αυτές τις περιπτώσεις τα ονόματα γράφονται φωνητικά. Το σημαντικότερο χαρακτηριστικό της επιγραφής είναι η αναγραφή μιας ομάδας συμβόλων με μικρούς χαρακτήρες πάνω από κάθε ιδεόγραμμα. Είναι πιθανό, αλλά όχι βέβαιο, ότι καθεμιά από τις ομάδες αυτές αποτελεί τη φωνητική απόδοση του ονόματος του αγγείου στο οποίο αναφέρεται.

Η διάδοση της γραμμικής Α. Η κυπρομινωική γραφή : η εξάπλωση του κρητικού εμπορίου κατά την πρώιμη και στην αρχή της ύστερης ανακτορικής περιόδου είχε ως αποτέλεσμα τη διάδοση της μινωικής γραφής στα νησιά των Κυκλάδων, και πιθανώς επίσης και στην ηπειρωτική Ελλάδα. Μεμονωμένα σύμβολα και ομάδες συμβόλων είναι, εδώ και καιρό, γνωστά από τη Φυλακωπή της Μήλου και σε αυτά ήδη έχουν προστεθεί διάσπαρτα ευρήματα από την Κέα, τα Κύθηρα, τη Νάξο και τη Θήρα. Από ό,τι γνωρίζουμε, δεν γίνεται πλέον χρήση της γραφής στις Κυκλάδες μετά την εξασθένηση της μινωικής επιρροής στα τέλη του 15^{ου} αιώνα.

Τα δείγματα της γραφής που προέρχονται από την Κύπρο είναι πολύ μεγαλύτερης σημασίας. Ήδη από το 1909, ο Evans ασχολήθηκε με τις επιγραφές σε πήλινες σφαίρες από την Έγκωμη της ανατολικής Κύπρου που χρονολογούνται στην εποχή του χαλκού. Αφού εξέτασε τη στενή ομοιότητα ανάμεσα σε μερικά σύμβολα της Έγκωμης και σε εκείνα των γραμμικών συστημάτων της Κρήτης, κατέληξε στο συμπέρασμα ότι και οι τρεις αυτές γραφές συνδέονται μεταξύ τους. Ήταν επίσης σαφές ότι μια ευθεία γραμμή συνέδεε την κυπριακή γραφή της εποχής του χαλκού, την οποία ο Evans ονόμασε κυπρομινωική, με το συλλαβάριο που ήκμασε στην Κύπρο κατά την κλασική περίοδο. Οι φωνητικές αξίες του κλασικού κυπριακού συλλαβαρίου είχαν γίνει γνωστές κάμποσα χρόνια πριν ο Evans γράψει τον πρώτο τόμο του *Scripta Minoa*.

Τώρα μπορούν να προστεθούν νέα δείγματα ενεπίγραφων πήλινων σφαιρών από την Κύπρο, πέρα από εκείνα που ήταν γνωστά στον Evans, ενώ η ανακάλυψη μεγάλων επιγραφών από την Έγκωμη, που χρονολογούνται στην ύστερη περίοδο της εποχής του χαλκού, μαρτυρεί τη χρήση της γραφής εκεί για σκοπούς διάφορους από την απλή τήρηση αρχείων. Η παλαιότερη σωζόμενη κυπρομινωική πινακίδα, που χρονολογείται στον 16^ο αιώνα, παρουσιάζει έναν τύπο γραφής που σαφώς προέρχεται από τη μινωική γραμμική Α:



Σχήμα 1.2.8: Κυπρομυνοική πινακίδα

Έτσι φαίνεται πιθανό ότι η γραφή εισήχθη στην Κύπρο από μινωίτες εμπόρους πριν από τη μυκηναϊκή επέκταση στην ανατολική μεσόγειο.

1.3 Η αποκρυπτογράφηση της γραμμικής Β

Η παρακάτω εξήγηση είναι απαραίτητη για να καταλάβουμε πως αποκρυπτογραφήθηκε η γραμμική Β. Το πρώτο βήμα στην προσπάθεια αποκρυπτογράφησης μιας άγνωστης γραφής είναι η ανάλυση των κειμένων. Είναι απαραίτητο να γνωρίζουμε τι είδους γραφής είναι και τι μπορεί να συναγάγει κανείς σχετικά με το περιεχόμενο των επιγραφών. Όλες οι γραφές μπορούν να ταξινομηθούν σε έναν από τους τρεις τύπους: α) φωνητικό, β) ιδεογραφικό, γ) μεικτό. Οι φωνητικές γραφές αντιπροσωπεύουν με τα σημεία τους, τους ήχους της γλώσσας. Δεν δίνουν φυσικά λεπτομερή εικόνα των ήχων αυτών, κατ' αρχάς θα δημιουργούνται σύγχυση αν ο κάθε ομιλητής έγραφε ακριβώς όπως μιλούσε, γιατί τότε η ίδια άρθρωση θα καταγραφόταν με πολλούς και διαφορετικούς τρόπους. Υπάρχει γι' αυτόν τον λόγο ένα συμβατικό στοιχείο στις γραφές που εξαλείφει τις περισσότερες από τις ατομικές διαφορές ανάμεσα στους ομιλητές. Δεύτερο, για να αναπαραστήσουμε ακόμη και χονδρικά την κλίμακα των ήχων που χρησιμοποιούνται θα ήταν απαραίτητο ένα πολύ μεγαλύτερο αλφάβητο από τα 24 γράμματά του, πολλές γλώσσες κάνουν χρήση διακριτών σημείων επάνω στα γράμματα για να υποδείξουν ιδιαίτερες αξίες, αλλά ακόμη κι έτσι όλες οι φωνητικές γραφές είναι μόνο μια επαρκής υποδήλωση για να επιτρέψει σε κάποιον που γνωρίζει τη γλώσσα να αναπαραστήσει τη λέξη για λογαριασμό του. Τα τμήματα στα οποία αναλύεται η ροή του λόγου για παράσταση μπορεί να ποικίλουν σε μέγεθος. Η αλφαβητικές γραφές σκοπεύουν στο ιδανικό ενός σημείου για κάθε ήχο, αν και η ελληνική, για παράδειγμα, συχνά ξεφεύγει από το ιδανικό αυτό. Άλλες γλώσσες στο παρελθόν έχουν κάνει χρήση συλλαβικών γραφών, όπου το κάθε σημείο αναπαριστά μια συλλαβή που μπορεί να προφερθεί. Αυτές μπορούν να ποικίλουν από τον απλό τύπο όπου κάθε σημείο δηλώνει ένα σύμφωνο που ακολουθείται από φωνήεν, έως στους πιο περίπλοκους τύπους όπου υπάρχουν σημεία για φωνήεντα που ακολουθούνται από σύμφωνο και για ομάδες συμφώνου, φωνήεντος και συμφώνου και ούτω καθεξής.

Οι ιδεογραφικές γραφές βασικά έχουν ένα σημείο για κάθε λέξη και το σημείο αυτό αντιπροσωπεύει συνήθως την έννοια της λέξης, όχι κατά κύριο λόγο την φωνητική της μορφή. Μπορούν γι' αυτό να μεταφέρονται εύκολα από μια γλώσσα σε μια άλλη. Έτσι οι Ιάπωνες δανειστήκαν την κινέζικη ιδεογραφική γραφή και έγραψαν τα ίδια σημεία για τις ίδιες σημασίες, αλλά τους έδωσαν εντελώς διαφορετικούς ήχους. Το πρόβλημα με τις ιδεογραφικές γραφές είναι ότι χρειάζονται τεράστιους αριθμούς σημείων, που πρέπει να είναι πολύ πολύπλοκα για να είναι διαφορετικά και γι' αυτό είναι δύσκολο να τα μάθει και να τα γράφει κανείς. Ωστόσο το πλεονέκτημα για τον αποκρυπτογράφο είναι ότι είναι εύκολα αναγνωρίσιμα από τον μεγάλο αριθμό διάφορων σημείων και από την πολυπλοκότητά τους. Τα ιδεογράμματα αρχίζουν συνήθως ως εικόνες αντικειμένων και μπορούν παρά την ανάπτυξη τους να παραμένουν ακόμα αναγνωρίσιμα, αλλά μπορούν επίσης να εξελιχτούν σε μη αναγνωρίσιμα σχήματα, η σημασία των οποίων είναι απλώς συμβατική. Μια οικεία σειρά ιδεογραμμάτων, τα αριθμητικά 1, 2, 3, 4, 5, μπορεί στα ελληνικά να διαβαστεί ως ένα, δύο, τρία, τέσσερα, πέντε, αλλά τα ίδια σημεία μπορούν να χρησιμοποιηθούν σε άλλες γλώσσες με εντελώς διαφορετικές φωνητικές αξίες, αν και η σημασία τους παραμένει η ίδια. Μικτά συστήματα δεν είναι ασυνήθιστα, δηλαδή αυτά όπου μερικά σημεία είναι ιδεογραφικά και μερικά φωνητικά. Όταν διαβάζουμε 1^{ος} ως πρώτος και 3^{ος} ως τρίτος, χρησιμοποιούμε ιδεογράμματα

με ένα φωνητικό συμπλήρωμα. Αυτό βοηθούσε για να μας εμποδίζει να δίνουμε τον λανθασμένο ήχο στο ιδεόγραμμα και σε μερικές γλώσσες μας επιτρέπει να υποδεικνύουμε κλιτές μορφές. Παραδείγματα μικτών γραφών είναι τα Χετιτικά στην αρχαιότητα, και στους νεότερους χρόνους στα Ιαπωνικά.

Η εμφάνιση του τρίτου και τελευταίου είδους αιγαιακής γραφής, της «γραμμικής γραφής Β» του Evans, συνήθως τοποθετείται στον 15^ο αιώνα π.χ.. Μπορούμε όμως βάσιμα να υποθέσουμε ότι εμφανίστηκε σε κάποια αρκετά παλαιότερη χρονολογία. Επειδή η δομή της γραμμικής Α παραμένει ακόμη ατελώς γνωστή, δεν είναι εύκολο να εικάσουμε γιατί χρειάστηκε ένα νέο σύστημα γραφής. Από ό,τι μπορούμε να αντιληφθούμε, η γραμμική Β διαφέρει από τη γραμμική Α σε δύο σημαντικά σημεία: στην εσωτερική δομή του συστήματος και στην εξωτερική μορφή των συμβόλων.

Σε ό,τι αφορά τη δομή, είδαμε ότι οι κατάλογοι στις πινακίδες της γραμμικής Α συνήθως αποτελούνται από μεμονωμένα σύμβολα (που φαίνεται να είναι ιδεογράμματα) και από ομάδες συμβόλων (που μοιάζουν να αναπαριστούν συλλαβές της ομιλούμενης γλώσσας). Μόνο σπάνια διαπλέκονται ιδεογράμματα με φωνητικά σύμβολα στην ίδια φράση. Αντίθετα η γραμμική Β χρησιμοποιεί συχνά ομάδες φωνητικών συμβόλων που ακολουθούνται από ένα ιδεόγραμμα, έτσι που και τα φωνητικά σύμβολα και το ιδεόγραμμα να αναφέρονται στο ίδιο αντικείμενο. Αντιπροσωπευτικό δείγμα αυτού του είδους «διπλής γραφής» όπως αποκαλείται μερικές φορές, αποτελεί το πρώτο κείμενο γραμμικής Β που θα εξετάσουμε παρακάτω.

Σε ό,τι αφορά την εξωτερική μορφή των δύο συστημάτων, και τα δύο χρησιμοποιούν σχεδόν τον ίδιο αριθμό συλλαβικών συμβόλων. Ορισμένα είναι κοινά και στα δύο συστήματα, ενώ κάθε γραφή διαθέτει και πολλά σύμβολα που ανήκουν αποκλειστικά σε αυτήν.

Βάσει των παραπάνω στοιχείων, αλλά και βάσει άλλων παραμέτρων, ο Evans διαπίστωσε ότι τα τρία κύρια είδη της κρητικής γραφής, αν και συνδέονται στενά μεταξύ τους, δεν αποτελούσαν στάδια μιας απλής διαδικασίας μετεξέλιξης, σύμφωνα με την οποία η γραμμική Α αντικατέστησε την εικονογραφική γραφή και αυτή, με τη σειρά της, εκτοπίστηκε από τη γραμμική Β. Κατά τη διατύπωση του ίδιου του Evans:

«Το προφανές συμπέρασμα ότι η γραμμική γραφή Β, η οποία στην Κνωσό εκτοπίζει την άλλη, είναι απλώς μεταγενέστερη εξέλιξη της γραμμικής Α, δεν εξηγεί ικανοποιητικά τα φαινόμενα τα οποία έχουμε να αντιμετωπίσουμε. Είναι αλήθεια ότι τα γενικά χαρακτηριστικά αυτών των ύστερων κνωσιακών γραπτών μνημείων είναι πιο προχωρημένα. Τα κείμενα είναι συχνά πληρέστερα και οι πινακίδες μεγαλύτερες, υπάρχει μικρότερη επιλογή χαρακτήρων κι ένα λιγότερο πολύπλοκο σύστημα συνδυασμού συμβόλων. Ταυτόχρονα το συμπέρασμα ότι η γραμμική Β αναπτύχθηκε απλώς από την άλλη[Α] αποκλείεται από το γεγονός ότι αρκετά από τα σύμβολα που ανήκουν στην γραμμική Β δεν απαντώνται στην Α, και από το ότι μερικά από τα σύμβολα που είναι κοινά και στα δυο "συλλαβοσυστήματα" εμφανίζονται σε μια πρωτόγονη μορφή σε πινακίδες της γραμμικής Β. Τα δύο συστήματα, που περιέχουν μεγάλη αναλογία κοινών στοιχείων, θα πρέπει επομένως να θεωρηθούν, σε γενικές γραμμές, παράλληλα μεταξύ τους, και είναι πιθανό, όπως ήδη προτάθηκε, η χρήση της γραμμικής Β στο αναδομημένο ανάκτορο της Κνωσού να συνέπεσε λίγο-πολύ χρονικά με τη συνεχιζόμενη χρήση της γραμμικής Α στη Φαιστό και την Αγία Τριάδα.»

Μεταγενέστερες ανασκαφές στην Κρήτη και περαιτέρω προβληματισμός πάνω στο είδος των γραφικών αυτών συστημάτων έχουν επιβεβαιώσει την ακρίβεια της διάγνωσης του Evans. Η ιστορία της κρητικής γραφής δε θα πρέπει να θεωρηθεί μια ενιαία εξέλιξη σε όλα τα μέρη του νησιού, τουναντίον ίσως υπήρξε μεγάλο ποσοστό αλληλοεπικάλυψης, ενώ διαφορετικές σχολές γραφέων κινούντο με διαφορετικές ταχύτητες, αν και προς την αυτή γενική κατεύθυνση: της απλοποίησης και τυποποίησης των εικονογραφικών συμβόλων. Επιπλέον μολονότι η γραμμική Β έχει πολλά κοινά χαρακτηριστικά με την αρχαιότερη της γραμμική Α, άλλα στοιχεία της εμφανίζουν μεγαλύτερη ομοιότητα με την εικονογραφική γραφή. Συνεπώς είναι προτιμότερο να θεωρήσουμε τη γραμμική Β συγγενή της γραμμικής Α, αλλά όχι άμεση διάδοχο της.

Ο Evans χρονολογεί τις επιγραφές γραμμικής Β της Κνωσού στο τέλος της ύστερης κνωσιακής ανακτορικής περιόδου, που σήμερα συνήθως τοποθετείται γύρω στο 1375. Αυτή, ή κάποια κοντινή προς αυτή, χρονολογία παραμένει ακόμη η πιο πιθανή, παρά τις αντιρρήσεις που διατύπωσε ο Palmer για τις μεθόδους και τα επιχειρήματα του Evans. Πάντως οι μεγάλες ποσότητες επιγραφών γραμμικής Β που βρέθηκαν στην Πύλο το 1939, καθώς και στα μεταπολεμικά χρόνια, θα πρέπει να τοποθετηθούν σε μια πολύ νεότερη χρονολογία. Οι επιγραφές αυτές ανήκουν σε κάποια ύστερη φάση

της ΥΕ ΙΙΒ (γύρω στο 1200), αυτή δε είναι και η κατά προσέγγιση χρονολόγηση των λίγων κειμένων που ανακαλύφθηκαν στις Μυκήνες τη δεκαετία του 1950 και στην Τίρυνθα τη δεκαετία του 1970. Η χρονολογία των πινακίδων από τη Θήβα της Βοιωτίας δεν μπορεί να προσδιορισθεί με βεβαιότητα προτού δημοσιευθεί πλήρως η σχετική κεραμική, βάσει των σημερινών μας γνώσεων, φαίνεται να είναι λίγο-πολύ σύγχρονες με τις πινακίδες της Πύλου.

Εκτός από τις επιγραφές που είναι χαραγμένες σε πήλινες πινακίδες και σφραγίσματα, και οι οποίες αποτελούν τον κύριο όγκο του υπάρχοντος υλικού σε γραμμική Β, είναι γνωστά και μικρά «γραπτά» κείμενα, ζωγραφισμένα δηλαδή πάνω σε πήλινα αγγεία που προέρχονται από διάφορες ηπειρωτικές τοποθεσίες, όπως η Ελευσίνα, οι Μυκήνες, ο Ορχομενός, η Θήβα και η Τίρυνθα, καθώς επίσης και από την Κνωσό και τα Χανιά στην Κρήτη. Οι χρονολογίες κατασκευής τους κυμαίνονται ανάμεσα στον 14^ο και τον 13^ο αιώνα.

1.4 Τα προκαταρκτικά της αποκρυπτογράφησης:

Σε πρώτη ματιά, οι πινακίδες της γραμμικής Β ελάχιστα υπόσχονται κάποια πιθανότητα επιτυχούς αποκρυπτογράφησης τους, λόγω της παντελούς ελλείψεως δίγλωσσου κειμένου. Οι πινακίδες φέρουν μικρές επιγραφές που προέρχονται από τον γραφειοκρατικό μηχανισμό των ανακτόρων της εποχής του χαλκού, και αποτελούνται, σχεδόν αποκλειστικά, από καταλόγους ανθρώπων, ζώων και αγαθών, καταγραφές δοσοληψιών, περιγραφές γαιοκατοχών, και άλλα παρόμοια. Η πρώτη δυσκολία που αντιμετώπιζε ο επίδοξος αποκρυπτογράφος προερχόταν από την υλική κατάσταση των πινακίδων, καθώς οι περισσότερες είναι σπασμένες, φθαρμένες ή καμένες, σε μικρότερο ή μεγαλύτερο βαθμό. Οι επιγραφές, άλλωστε, είναι εξαιρετικά σύντομες και λακωνικές, και αποκαλύπτουν πολύ λίγα σχετικά με τη συντακτική δομή τους. Μια Τρίτη δυσκολία έχει σχέση με το αντικείμενο των κειμένων: καθώς ασχολούνται εκτεταμένα με ανθρώπους και τοποθεσίες, αναπόφευκτα ένας μεγάλος αριθμός των λέξεων που περιέχουν θα είναι κύρια ονόματα. Ενώ όμως τα κύρια ονόματα υπήρξαν η αφετηρία για την αποκρυπτογράφηση της περσικής σφηνοειδούς γραφής τον 19^ο αιώνα, δεν μπόρεσαν να προσφέρουν βοήθεια στην περίπτωση της γραμμικής Β. Μερικά από τα περσικά κείμενα ήταν δυνατό να αναγνωρισθούν ως «βασιλικές» επιγραφές, και φυσικά ήταν ήδη γνωστά τα ονόματα πολλών περσών βασιλέων. Καμιά ένδειξη όμως δεν υπήρχε ως προς τη γλώσσα στην οποία ήταν γραμμένα τα κείμενα της γραμμικής Β, ενώ τα ονόματα που περιείχαν ήταν πιθανότερο να είναι ονόματα εντελώς ταπεινών ανθρώπων παρά βασιλέων. Είναι αλήθεια πως τα τοπωνύμια έπαιξαν πραγματικά κάποιο ρόλο στην αποκρυπτογράφηση της γραμμικής Β, αν και όχι στο αρχικό στάδιο.

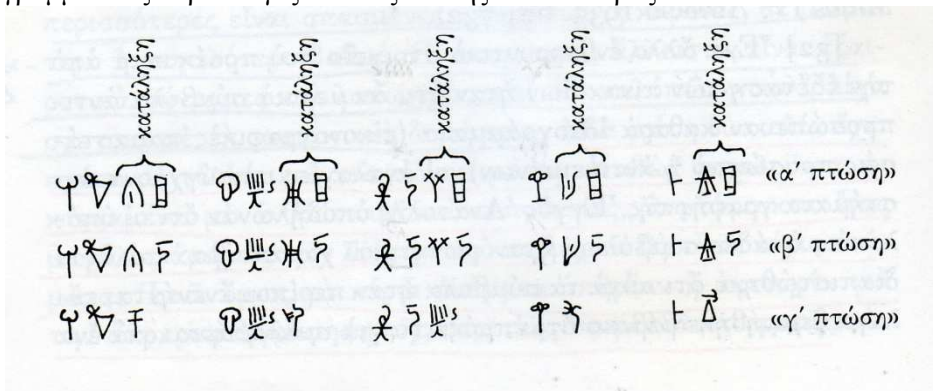
Παράλληλα, όμως, μια προσεκτικότερη και συστηματικότερη εξέταση του υλικού της γραμμικής Β απεκάλυψε μερικά χαρακτηριστικά, που φαίνονταν τουλάχιστον να μειώνουν αρκετά τον αριθμό των πιθανών γλωσσών. Αλλά το πιο σημαντικό στοιχείο είναι ότι η δομή μιας επιγραφής γραμμικής Β είναι τέτοια ώστε να δίνει ένα μέρος συμφραζομένων εντός του οποίου πρέπει να αναζητηθεί το νόημα του συνόλου, και εκτός του οποίου δεν μπορεί να γίνει κατανοητό. Το φάσμα των πιθανών νοημάτων περιορίζεται από το ιδεόγραμμα που υπάρχει στα περισσότερα πλήρη κείμενα. Πολλά χρόνια πριν επιτευχθεί μια πειστική συνολική αποκρυπτογράφηση της γραφής, ο Evans κατάφερε να κατατάξει εκατοντάδες πινακίδες της Κνωσού σε σαφώς προσδιορισμένες ομάδες, στηριζόμενος αποκλειστικά και μόνο στη μαρτυρία των ιδεογραμμάτων.

Ένα άλλο ενθαρρυντικό στοιχείο που προέκυπτε από εξέταση των πινακίδων ήταν ότι, αν μερικά σύμβολα αντιπροσώπευαν καθαρά ιδεογράμματα (εικονογραφικές παραστάσεις προσώπων ή αντικειμένων), οι αναλογίες με σύγχρονα συστήματα γραφής της Εγγύς Ανατολής υποδήλωναν ότι οι υπόλοιπες ομάδες συμβόλων ήταν φωνητικού χαρακτήρα, και όταν διαπιστώθηκε ότι τα σύμβολα ήταν περίπου ενενήντα, εύλογα θεωρήθηκε βέβαιο ότι επρόκειτο για συλλαβάριο: για ένα σύστημα δηλαδή στο οποίο κάθε φωνητικό σύμβολο εκφράζει μια συλλαβή της συγκεκριμένης γλώσσας.

Για τον λόγο αυτό θεωρήθηκε πιθανό ότι τα σύμβολα της γραμμικής Β τα οποία εμφανίζονταν σε ομάδες ήταν «συλλαβογράμματα». Αν το σύστημα, στο σύνολό του, αποτελούσε πράγματι συλλαβάριο, θα μπορούσε να αναλυθεί με κάποιον συστηματικό τρόπο ώστε να δώσει ορισμένες προκαταρκτικές ενδείξεις σχετικά με τις αρχές πάνω στις οποίες είχε δομηθεί η γλώσσα. Οι ομάδες των συμβόλων έπρεπε να μελετηθούν για να διαπιστωθεί αν περιείχαν ενδείξεις γραμματικής κλίσης. Αν τα πορίσματα αυτής της μελέτης ήταν θετικά, θα κατείχαμε κάποιες πληροφορίες σχετικά με τη

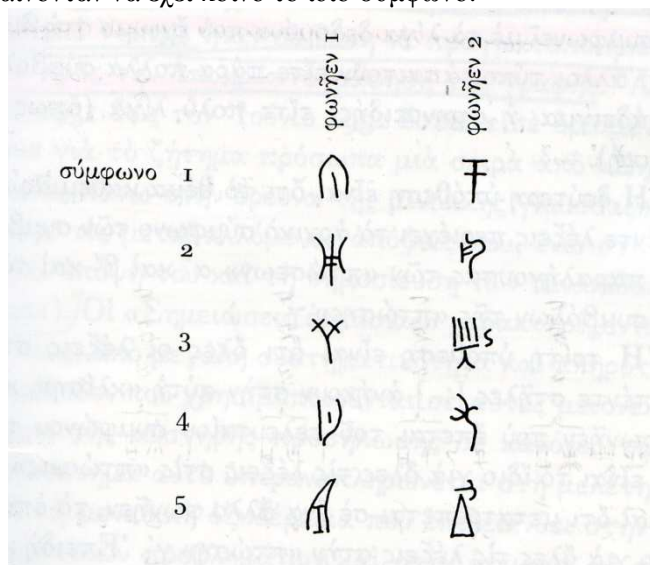
συντακτική δομή. Επειδή όμως οι διαθέσιμες επιγραφές ήταν πολύ σύντομες, οι πληροφορίες αυτές ήταν μάλλον απίθανο να μας επιτρέψουν άμεσα να προσδιορίσουμε τη γλώσσα, αλλά τουλάχιστον θα μπορούσαμε να αποκλείσουμε από την έρευνα σειρά ολόκληρη γλωσσών τελείως διαφορετικής δομής.

Το πρώτο ουσιαστικό βήμα προς τη μεθοδική απομόνωση των ομάδων συμβόλων που παρουσίαζαν σημεία γραμματικής κλίσης έγινε από την Alice Kober. Η Kober παρουσίασε μια σειρά από πιθανά παραδείγματα ουσιαστικών, στα οποία η μεσαία και η τελευταία σειρά έμοιαζαν να αποτελούν γραμματικές παραλλαγές των λέξεων της επάνω σειράς:



Σχήμα 1.4.1: Τα τρίδιμα της Alice Kober

Σε ορισμένες λέξεις, η αλλαγή τύπου ήταν τόσο ομαλή ώστε μπορούσε να προβλεφθεί, επειδή δε αυτές οι λέξεις εμφανίζονταν ως ουσιαστικά και φαινόταν να κλίνονται σύμφωνα με τον ίδιο κανόνα, η A.Kober αποτόλμησε να τα εντάξει στην «πρώτη κλίση» της «μινωικής» γλώσσας. Βάσει των κανόνων κλίσης που είχε εντοπίσει, η Kober κατέληξε σε κάποια προσωρινά συμπεράσματα σχετικά με τις φωνητικές συγγένειες των ίδιων συμβόλων. Έτσι, στην πρώτη στήλη του παραπάνω πίνακα, αν κάθε σύμβολο παριστούσε μια ανοιχτή συλλαβή του τύπου ba ή ti, ήταν πολύ πιθανό ότι το τρίτο σύμβολο στις γραμμές 1,2 και 3 θα έπρεπε να αρχίζει με το ίδιο σύμφωνο: αυτό το σύμφωνο θα έπρεπε να αντιπροσωπεύει τον τελευταίο ήχο του θέματος του ουσιαστικού, ενώ το φωνήεν θα έπρεπε να αντιπροσωπεύει την κατάληξη. Εφαρμόζοντας αυτή την αρχή σε όλες τις στήλες της, η Kober συνέθεσε αυτό που η ίδια αποκάλεσε «προσωρινό φωνητικό σχήμα» με πέντε ζεύγη συμβόλων, καθένα από τα οποία φαινόταν να έχει κοινό το ίδιο σύμφωνο:



Σχήμα 1.4.2: : προσωρινό φωνητικό σχήμα με πέντε ζεύγη συμβόλων της Alice Kober

Τη σχηματική αυτή κατάληξη την περιγράφει ως εξής η Kober: «οι υποθέσεις στις οποίες στηρίζεται [αυτό το σχήμα] είναι οι ακόλουθες: η γραμμική Β είναι ένα απλό συλλαβάριο, όπου κάθε

φωνητικό σύμβολο παριστάνει ένα συνδυασμό συμφώνου-φωνήεντος, με την πιθανότητα ότι κάποιο από αυτά τα δυο μπορεί να είναι νεκρό(zero), είναι δηλαδή δυνατό να είναι καθαρά σύμβολα συμφώνου ή φωνήεντος. Η υπόθεση αυτή δεν έχει αποδειχθεί, ούτε μπορεί να αποδειχθεί προς το παρόν. Ενδέχεται, μάλιστα, να είναι εντελώς λανθασμένη. Χρησιμοποιείται από τη συγγραφέα όπου είναι αναγκαίο, στην υποστήριξη ενός συλλογισμού, ώστε να διερευνηθούν οι πιθανές επιπτώσεις στο πλαίσιο ενός συλλαβαρίου, αφού ο αριθμός των φωνητικών συμβόλων της γραμμικής Β προσεγγίζει πολύ εκείνον του κυπριακού συλλαβαρίου, το οποίο ανήκει στον τύπο που περιγράψαμε. Όταν είναι αναγκαίο να θεωρηθεί δεδομένος κάποιος τύπος συλλαβαρίου, η μόνη λογική επιλογή είναι, κατά τη γνώμη μου να αναζητηθεί κάποιος τύπος που είναι απλός και ταυτόχρονα συμφωνεί με τα λίγα δεδομένα που έχουμε στη διάθεση μας, άλλοι τύποι απαιτούν είτε πάρα πολλά σύμβολα(για παράδειγμα, η σφηνοειδής) είτε πολύ λίγα(όπως η σανσκριτική).

Η δεύτερη υπόθεση είναι ότι το θέμα κάθε μιας από τις πέντε λέξεις περιέχει το αρχικό σύμφωνο των συμβόλων της παραλήγουσας των «πτώσεων» α' και β' και των τελικών συμβόλων της «πτώσης» γ'.

Η Τρίτη υπόθεση είναι ότι όλες οι λέξεις στις πρώτες πέντε στήλες ανήκουν στην αυτή «κλίση», και ότι το φωνήεν που έπεται του τελευταίου συμφώνου του θέματος είναι το ίδιο για όλες τις λέξεις στις «πτώσεις» α' και β', και ότι μετατρέπεται σε ένα άλλο φωνήεν, το οποίο είναι ίδιο για όλες τις λέξεις στην «πτώση» γ'. Επειδή και οι τρεις υποθέσεις είναι αναπόδευκτες, [το προσωρινό φωνητικό σχήμα] δεν έχει καμία αξία. Εάν, ωστόσο, καταστεί δυνατό να συγκροτηθεί ένα σχέδιο κλίσης της γραμμικής Β και εάν καταφέρουμε να δομήσουμε ένα λογικά πλήρες φωνητικό σχήμα με βάση αυτό το σχέδιο κλίσης, χρησιμοποιώντας υποθέσεις όμοιες με τις παραπάνω, και εάν βρούμε ότι οι φωνητικές σχέσεις επαληθεύονται αμοιβαία, τότε η αξία της αρχικής υπόθεσης θα μπορούσε να αποδειχθεί από αυτό και μόνο το γεγονός. Σε αυτό όμως το στάδιο, καμία τέτοια απόδειξη δεν είναι δυνατή.»

Διαπιστώνουμε ότι, σχετικά με το κάθε ζεύγος συμβόλων στο προσωρινό φωνητικό σχήμα της Alice Kober, τα συμπεράσματα της έχουν αποδειχθεί σωστά.

1.5 Η αποκρυπτογράφηση της γραμμικής Β από τον Ventris:

Ο Michael Ventris(12 Ιουλίου 1922-6 Σεπτεμβρίου 1956) ήταν άγγλος αρχιτέκτονας και ο άνθρωπος που αποκρυπτογράφησε την γραφή γραμμική Β.

Από μικρή ηλικία έδειξε έφεση στις γλώσσες, 9 ετών μιλούσε ήδη γαλλικά, ελβετικά, γερμανικά, λατινικά και αρχαία ελληνικά. Στα 14 του, το 1936, ενώ βρισκόταν σε έκθεση μινωικής τέχνης έμαθε ότι η γραμμική Β δεν είχε αποκρυπτογραφηθεί. Ο Μπλέγκεν ανακαλύπτει, το 1939, το ανάκτορο του Νέστορα στην Πύλο κι εκεί βρίσκονται πινακίδες γραμμένες σε γραμμική Β. Αυτές οι πινακίδες, που θα βοηθήσουν πολύ τον Βέντρις στο μέλλον, έθεσαν υπό αμφισβήτηση τη θεωρία του Έβανς ότι η γραμμική Β υπήρχε μόνο στην Κρήτη. Είχε ανακαλύψει τις πρώτες πινακίδες γραμμένες σε γραμμική Β στην Κνωσό το 1900.

Το 1940 άρχισε τις σπουδές του ως αρχιτέκτονας και παράλληλα μελετούσε τη γραμμική Β και προσπαθούσε να την αποκρυπτογραφήσει έχοντας τα ετρουσκικά σαν βάση, δημοσίευσε μάλιστα μια σχετική μελέτη του ('introducing the Minoan language') στον American journal of archaeology. Το 1942 παντρεύτηκε την Λοις Νοξ Νίβεν και λίγο αργότερα κατατάχτηκε στη RAF όπου συμμετείχε σε βομβαρδισμούς γερμανικών στόχων.

Το 1946 άφησε την αεροπορία και συνέχισε τις αρχιτεκτονικές του σπουδές, για ν' αποφοιτήσει 3 χρόνια αργότερα. Βρήκε δουλειά στο υπουργείο παιδείας ως μέλος ομάδας σχεδιασμού νέων σχολικών κτιρίων. Η αποκρυπτογράφηση της γραμμικής Β δεν είχε σταματήσει να τον στοιχειώνει και κατά τα τέλη του 1949 φτιάχνει ένα ερωτηματολόγιο που το αποστέλλει σε όλους τους ειδικούς (γλωσσολόγους, ιστορικούς και αρχαιολόγους) της εποχής του που τους απασχολούσε το ίδιο θέμα, προτείνοντας τους την πρωτοχρονιά του 1950 να κάνουν μια ανεπίσημη ανταλλαγή απόψεων. Συγκέντρωσε τις απαντήσεις του σ' ένα βιβλίο, πρόσθεσε περιληπτικά τις δικές του ιδέες και μια γενική κριτική, και το κυκλοφόρησε σε αυτό το στενό κύκλο λόγιων με τον τίτλο 'the language of Minoan and Mycenaean civilizations' ή 'Mid-Century Report'. Οι περισσότεροι θεωρούσαν ότι η γραμμική Β είναι μια ινδοευρωπαϊκή γλώσσα πιο κοντά στην χετιτική και κάποιοι, όπως και ο ίδιος ο Βέντρις, πίστευαν ότι ήταν μια λιγότερο γνωστή αιγιακή γλώσσα που πιθανότατα

είχε σχέση με την ετρουσκική. Ένας από τους ελάχιστους που υποστήριζαν ότι η γραμμική Β ήταν ελληνική γραφή ήταν ο Κτιστόπουλος, παραλήπτης κι αυτός του ερωτηματολογίου του Βέντρης.

Η Alice Kober, αν και δε θέλησε να συμμετάσχει στο ερωτηματολόγιο του Βέντρης, κράτησε επικοινωνία μαζί του. Οι παρατηρήσεις της έπαιξαν σημαντικό ρόλο στην αποκρυπτογράφηση του Βέντρης. Η Κόμπερ διαπίστωσε την ύπαρξη μοτίβων τριών σημείων στη γραφή (που αργότερα ο Βέντρης ονόμασε Kober triplets) τα οποία έδειχναν την ύπαρξη κλίσεων σε κάποιες λέξεις στις πινακίδες, αλλά βρήκε και ότι η γραμμική Β ήταν συλλαβογραφική. Η Κόμπερ πέθανε το 1950 από καρκίνο και ο Βέντρης συνέχισε την έρευνα.

Έχει ενδιαφέρον ότι όπως ο Σαμπολιόν (Jean Francis Champollion) αποκρυπτογράφησε τα ιερογλυφικά με βάση τα ονόματα βασιλέων, έτσι ο Βέντρης ξεκίνησε την αποκρυπτογράφηση της γραμμικής Β με βάση ονόματα πόλεων που υπήρχαν μόνο στις πινακίδες από την Κρήτη. Τελικά ο Βέντρης, στις 1 Ιουλίου 1952, ανακοίνωσε στο BBC την αποκρυπτογράφηση της γραμμικής Β. Ενώ μέχρι τότε θεωρούνταν ο 7^{ος} ή 8^{ος} αιώνας π.χ. ως ημερομηνία έναρξης της ελληνικής γραφής, με τα νέα δεδομένα για τη γραμμική Β η ημερομηνία τοποθετείται στο 1500π.χ., πολλούς αιώνες πριν τα ελληνικά του Ομήρου.

Μετά την ανακοίνωση στο BBC τον πλησίασε ο John Chadwick, κλασικός φιλόλογος και γλωσσολόγος. Η πλειοψηφία της ακαδημαϊκής κοινότητας δέχτηκε και συμφώνησε με το έργο της αποκρυπτογράφησης, αλλά ορισμένοι διατήρησαν τις επιφυλάξεις τους ή εξέφραζαν αντιρρήσεις, ανάμεσα στους τελευταίους ήταν και ο γερμανός Κλάφενμπαχ. Ωστόσο ο Βέντρης επιβεβαίωσε την αποκρυπτογράφησή του με 400 νέες πινακίδες που ανακάλυψε ο Καρλ Μπλέγκεν στο παλάτι του Νέστορα στην Πύλο. Από τη συνεργασία του Βέντρης με τον Τσαντγουίκ προέκυψε το 1953 η μελέτη 'αποδείξεις για ύπαρξη ελληνικής διαλέκτου στα μυκηναϊκά αρχεία' (evidence for Greek dialect in the Mycenaean archives).

Η ολοένα πιο έντονη απασχόληση του Βέντρης με τη γραμμική Β τον απομάκρυνε αρκετά από την γυναίκα του και τον αποξένωνε από την οικογένεια του γενικότερα. Τον Σεπτέμβριο του 1956, σε ηλικία μόλις 34 ετών ο Μαικλ Βέντρης πέθανε σε τροχαίο δυστύχημα, έπεσε με το αυτοκίνητο του σε ένα παρκαρισμένο φορτηγό σε ένα σκοτεινό δρόμο. Ο Τσαντγουίκ συνέχισε τη δουλειά τους και δημοσίευσε το 1956 το βιβλίο 'έγγραφα σε μυκηναϊκά ελληνικά'. Είχε πεί για τον Βέντρης ότι διέθετε την ικανότητα να διακρίνει τάξη σε ένα σύνολο πληροφοριών που για άλλους μοιάζει σαν χάος. Το σύνολο των σημειώσεων του Βέντρης δημοσιεύτηκε το 1988 με τίτλο 'work notes on Minoan language research and other unedited papers'.

Computer techniques in the study of the Minoan Linear A script: Ήταν ο τίτλος που έδωσε ο Dr. David Packard το 1971 στο άρθρο του. Η βασισμένη στην τεχνολογία έρευνα του για τη γραμμική Α μεταγενέστερα αναπτύχθηκε, επεκτάθηκε και δημοσιεύθηκε αναλυτικά στο "Minoan Linear A" το 1974. Η εργασία του Packard ήταν μια διεξοδική, λεπτομερής εξέταση όλων των στοιχείων της γραμμικής Α μέχρι εκείνη τη στιγμή. Η μελέτη του Packard ήταν, και είναι, τόσο μεγάλης αξίας, διότι ο ίδιος περιορίζεται στον έλεγχο των αξιών των σημείων/συλλαβών της γραμμικής Α, χωρίς να κάνει κανένα σχόλιο για την ταυτότητα της μινωικής γλώσσας.

Μετά την αποκρυπτογράφηση της γραμμικής Β(μυκηναϊκά ελληνικά) από τον Michael Ventris το 1952, όλη η προσοχή των μελετητών στράφηκε στην γραμμική Α. Πριν η γραμμική Α θεωρηθεί με βεβαιότητα γραφή είναι απαραίτητο να αναρωτηθούμε πόσο έγκυρη είναι μια μεταβίβαση φωνητικών αξιών από τη γραμμική Β στη γραμμική Α; Πολλοί μελετητές, όπως και ο Packard, χρησιμοποίησαν μια 'ευθεία' μεταφορά των φωνητικών αξιών από τη γραμμική Β στα ομομορφικά σημεία της γραμμικής Α ως 'υπόθεση εργασίας', με την προειδοποίηση ότι κάτι τέτοιο δεν πρέπει να γίνεται σε κάθε περίπτωση.

Το 1975, ο Olivier ήταν σε θέση ν αποδείξει ότι 10 ομομορφικά σημεία (DA, I, JA, KI, PI, RI, RO, TA, TE, SU) των γραμμικών Α και Β είχαν με βεβαιότητα την ίδια φωνητική αξία και στις δύο γραφές.

Το 1984, ο Godart συμπεριέλαβε τρία ακόμα σημεία(PA, SE, TO).

Το 1989, ο Duboux έφτιαξε μια λίστα με 30 ομομορφικά σημεία των 2 γραφών που όμως είχαν την ίδια φωνητική αξία(A, E, I, U, DA, JA, KE, KI, MA, ME, MI, NA, NI, PA, PI, PO, QA, QE, RA, RI, RO, SE, SI, SU, TA, TE, TI, TO, TU and WA). Αυτά τα 30 σημεία, ενώ αποτελούν λιγότερο από το ήμισυ του συλλαβάριου της Γραμμικής A, είναι ωστόσο σημαντικά να αποδείξουν ότι το συλλαβάριο είναι βασικά το ίδιο με εκείνο της Γραμμικής B. Τριάντα από τις εξήντα βασικές φωνητικές συλλαβές της Γραμμικής B, μπορεί να αποδειχθεί ότι υπήρχαν στη Γραμμική A. Είναι φυσικά γεγονός ότι μερικά σημεία της γραμμικής A είχαν διαφορετική φωνητική αξία από αυτά της γραμμικής B. Όταν η γραφή ταξιδεύει, προκύπτουν αλλαγές για να καλύψουν τις ανάγκες της νέας γλώσσας. Η γραμμική γραφή A στην Κρήτη, ωστόσο, δεν ταξίδεψε, αλλά μάλλον αναπτύχθηκε για να φιλοξενήσει, να εξυπηρετήσει μια άλλη γλώσσα. Η πιθανότητα πολλών ίδιων σημείων της γραμμικής A και B να έχουν την ίδια φωνητική αξία έχει αναγνωρισθεί.

Το 1959 ο Chadwick περιέγραψε τη διαφορά των δύο γραφών παρόμοια με αυτή του ελληνικού και ρωμαϊκού αλφάβητου. Το 1978 οι Pope και Raison είπαν πως τα δύο συλλαβάρια διαφέρουν στον ίδιο βαθμό που διέφεραν τα αγγλικά και γαλλικά αλφαβητάρια τον 18^ο αιώνα. Το 1988 ο Palaima είπε ότι “τα φωνητικά ρεπερτόρια των γραμμικών A και B είναι δικαιολογημένα κοντά το ένα στο άλλο πολύ περισσότερο από ότι προηγούμενες στατιστικές συγκρίσεις έχουν δείξει”.

Για να προχωρήσουμε περαιτέρω με μια σοβαρή μελέτη της γραμμικής A και της μινωικής γλώσσας, είναι απαραίτητο να δούμε πόσο μεγάλο μέρος της συλλογής της γραμμικής A μπορεί να διαβαστεί σύμφωνα με τις φωνητικές αξίες της γραμμικής B και να δούμε τι επίπεδο ηχητικής σχέσης μπορεί να εξακριβωθεί μεταξύ των γλωσσών που εκφράστηκαν από τη γραμμική B και τη γραμμική A. Για να γίνει αυτό είναι αναγκαίο να εδραιωθεί μια στατιστική σχέση μεταξύ της γραμμικής B, γραφή της μυκηναϊκής Ελλάδας και της γραμμικής A, γραφή της μινωικής Κρήτης. Εφόσον η ακριβής στατιστική σχέση μεταξύ της γραμμικής A και της γραμμικής B μπορεί να αποδειχθεί, να εδραιωθεί, σε σχέση με το πώς οι ηχητικές αξίες συμπεριφέρονται σε κάθε γραφή, τότε θα είναι δυνατόν να αποδειχθεί σε ποσοστιαίους όρους πόσο πολύ οι ηχητικές αξίες των γραφών διαφέρουν μεταξύ τους κι αυτό με τη σειρά του ν αποδείξει πως οι φωνητικές αξίες δε μπορούν να μεταφερθούν πίσω από τη γραμμική B στη γραμμική A. Το επόμενο και πιο δύσκολο στάδιο θα είναι να προσδιοριστεί ποια σημεία δεν συμπεριφέρονται με τον ίδιο τρόπο στις δύο γραφές κι έτσι δεν εναρμονίζονται με την τεκμηριωμένη σχέση. Αλλά αυτό είναι για το μέλλον. Για το παρόν είναι απαραίτητο να επαναληφθεί η στατιστική ανάλυση που πραγματοποιήθηκε από τον Packard 25 χρόνια πριν, και να εκσυγχρονιστεί υπό το φως των πρόσφατα ανακαλυφθέντων επιγραφών της γραμμικής A.

Συνεπώς προτείνεται να επαναληφθεί η ανάλυση του Packard και να δούμε εάν μπορούν να γίνουν παρατηρήσεις οι οποίες θα βοηθήσουν σε μελλοντικές έρευνες της γραμμικής A και της μινωικής γλώσσας.

Το 1974, η στατιστική ανάλυση του Packard βασίστηκε στη συχνότητα εμφάνισης του κάθε σημείου της γραμμικής A και B στην αρχική, μεσαία και τελική θέση. Αυτή η πληροφορία έδειξε ένα επίπεδο ομοιότητας μεταξύ του πώς ομομορφικά σημεία συμπεριφέρονται σε σχέση με τη θέση τους σε μια ομάδα σημείων/λέξη, τόσο στη γραμμική A όσο και τη γραμμική B. Ο Packard συγκέντρωσε όλες τις πληροφορίες που είχε στη διάθεση του το 1974 και συνάγαγε συμπεράσματα όσον αφορά τη φωνητική σχέση μεταξύ των γλωσσών που αποδίδονται από τη γραμμική A και B. Η μελέτη του Packard ήταν μια διεξοδική ανάλυση όλων των συλλαβογραμμάτων, τόσο των φωνητικών αξιών σύμφωνα + φωνήεν όσο και των καθαρών φωνηέντων. Αυτό του επέτρεψε να καταδείξει τη διαφορά μεταξύ των δύο γραφών περίπου στο 20%.

Το 1989, ο Duhoux επανέλαβε τη μελέτη της συχνότητας του κάθε καθαρού φωνήεντος στη γραμμική B, και το αντίστοιχο του σημείο στη γραμμική A.

Πριν από την παρουσία των στατιστικών στοιχείων και την εξαγωγή συμπερασμάτων σχετικά με την αντιστοιχία των γλωσσών που εκφράστηκαν από τη γραμμική A και B, είναι απαραίτητο να δείξουμε σε ποιές εργασίες βασίζονται τα στοιχεία:

1) **PACKARD 1974:**

Linear A = J. Raison and M. Pope. Index du Linéaire A, 1971.
Knossos Linear B = A. Morpugo. Mycenaeae Graecitatis Lexicon, 1963.

2) **DUHOUX 1989:**

Linear A = GORILA 1-5, 1976-1985; A. Karetsou - L. Godart and J.-P. Olivier "Inscriptions en Linéaire A du Sanctuaire de Sommet Minoen du Mont Iouktas" Kadmos 24, 1985, 89-147; and A. Karetsou "Two New Inscriptions in Linear A from the Peak Sanctuary of Iouktas", Eilapini, 1987, 84-91.

Linear B = J.-P. Olivier, "Linéaire B et Ordinateur Electronique", L'Antiquité Classique 34 (1965) 387-397.

3) **OWENS 1996:**

Linear A = GORILA 1-5, 1976-1985; and GORILA 6 "Gareth Owens Record of Inscriptions in Linear A" 1985-19957.

Στατιστικά στοιχεία.

PACKARD 1974 (LA = RAISON - POPE 1971: KNLB = MORPUGO 1963)

	I	M	F	TOTAL	
A	81.50	10.80	07.70	72	LA
	90.03	02.05	07.92	341	KNLB
E	50.00	00.00	50.00	2	LA
I	44.40	28.60	27.00	65	LA
	36.75	39.76	23.49	166	KNLB
O	83.30	16.70	00.00	6	LA
	56.73	12.28	30.99	171	KNLB
U	77.30	13.60	09.10	26	LA
	21.51	41.40	37.10	128	KNLB

Vowel total = 171 LA and 1088 KNLB

DUHOUX 1989 (LA = GORILA 1985 + KGO 1985 + K 1987; LB = OLIVIER 1965)

	I	M	F	TOTAL	
A	89.74	05.12	05.12	156	LA
	93.50	02.81	03.67	1496	LB
E	64.00	08.00	28.00	25	LA
	77.67	11.25	11.07	1111	LB
I	55.06	31.01	13.92	158	LA
	42.21	28.44	29.34	552	LB
O	85.18	07.40	07.40	27	LA
	55.47	07.86	36.66	643	LB
U	56.36	25.45	18.18	55	LA
	14.54	37.19	48.26	777	LB

VOWEL TOTAL = 421 LA AND 4579 LB

Τα αριθμητικά στοιχεία που επιτυγχάνονται παραπάνω, επαναλαμβάνονται τώρα παρακάτω για να δείξουν την αυξανόμενη συσχέτιση μεταξύ της γραμμικής Α και της γραμμικής Β, από το 1974 έως το 1989. Αυτό είναι μια καλή ένδειξη ότι η υπόθεση του Packard ήταν σωστή και ότι η στατιστική στάση συνεχίζεται προς τα κάτω, δηλαδή να δείξει ένα αυξανόμενο επίπεδο σχετικότητας μεταξύ των γλωσσών που εκφράζονται από τις δύο γραφές.

	<u>P1974</u>	<u>D1989</u>
A	81.50-10.80-07.70	89.74-05.12-05.12
	90.03-02.05-07.92	93.50-02.81-03.67
DIFF =	08.53-08.75-00.22	03.76-02.29-01.45
E	50.00-00.00-50.00	64.00-08.00-28.00
	81.70-09.38-08.93	77.67-11.25-11.07
DIFF =	31.07-09.38-41.17	13.67-03.25-16.93
I	44.40-28.60-27.00	55.06-31.01-13.92
	36.75-39.76-23.49	42.21-28.44-29.34
DIFF =	07.65-11.16-03.51	12.85-02.57-15.42
O	83.30-16.70-00.00	85.18-07.40-07.40
	56.73-12.28-30.99	55.47-07.86-36.66
DIFF =	16.57-04.42-30.99	29.71-00.46-29.26
U	77.30-13.60-09.10	56.36-25.45-18.18
	21.51-41.40-37.10	14.54-37.19-48.26
DIFF =	55.79-27.80-28.00	41.82-11.74-30.08

OWENS 1996 (LA =GORILA 5 1985+6 1995; KNLB=KT5 1989+JOINS 1-5)

	I	M	F	TOTAL	
A	88.41	06.09	05.59	164	LA
	87.19	02.07	10.74	726	LB
DIFF =	01.22	04.02	05.15		
E	64.00	08.00	28.00	25	LA
	75.12	11.41	13.45	587	LB
DIFF =	11.12	03.41	14.55		
I	55.28	31.05	13.66	161	LA
	35.06	49.14	15.80	348	LB
DIFF =	22.22	18.09	02.12		
O	85.20	07.40	07.40	27	LA
	63.92	19.13	16.95	413	LB
DIFF =	11.28	11.73	09.55		
U	58.62	24.14	17.24	58	LA
	33.98	42.06	23.95	359	LB
DIFF =	24.64	17.92	06.71		

VOWEL TOTAL = 435 LA AND 2433 KNLB

% DIFFERENCE BETWEEN LINEAR A AND LINEAR B

	P1974	D1989	OWENS 1996
A	08.53-08.75-00.22	03.76-02.29-01.45	01.22-04.02-05.15
E	31.07-09.38-41.17	13.67-03.25-16.93	11.12-03.41-14.55
I	07.65-11.16-03.51	12.85-02.57-15.42	22.22-18.09-02.12
O	16.57-04.42-30.99	29.71-00.46-29.26	11.28-11.73-09.55
U	55.79-27.80-28.00	41.82-11.74-30.08	24.64-17.92-06.71
M	23.92-12.30-20.78	20.36-04.06-18.61	14.06-11.03-07.62
AM	19.00	14.34	10.90

Τα στατιστικά στοιχεία που συγκεντρώθηκαν παραπάνω μπορούν πλέον να παρουσιάζονται με μορφή πίνακα για να καταστεί δυνατή η σύγκριση και να δειχθεί μια αυξανόμενη τάση συσχέτισης ως αποτέλεσμα μιας αυξανόμενης βάσης δεδομένων με επιγραφές σε γραμμική A και B.

P1974 M 23.92-12.30-20.78 AV.M= 19.00 (171 LA + 1088 KNLB)
 D1989 M 20.36-04.06-18.61 AV.M= 14.34 (421 LA + 4579 LB!)
 OWENS 1996 M 14.06-11.03-07.62 AV.M= 10.90 (435 LA + 2433 KNLB)

Το 1974, ο Packard ήταν σε θέση να αποδείξει μια περίπου 20% διαφορά μεταξύ της θέσης των καθαρών φωνηέντων στην αρχική, μεσαία και τελική θέση στη γραμμική A και B. Αυτό του επέτρεψε να δείξει ότι κάποια φραστική σχέση μπορεί να υπάρχει μεταξύ των γλωσσών που εκφράζονται με τη γραμμική A και B. Το 1989, ο Duhoux ήταν σε θέση να παρουσιάσει μια διαφορά γύρω στο 14,34%, αλλά οι δικοί του αριθμοί της γραμμικής B ήταν βασισμένοι σε παλαιότερες εκδόσεις και το υλικό του δεν ήταν μόνο από την Κρήτη, εισάγοντας έτσι μια χρονολογική και γεωγραφική διαφοροποίηση από τη μινωική γραμμική A. Το 1995, το πρωτότυπο έργο του Packard σε σχέση με τα καθαρά φωνήεντα επαναλήφθηκε βασισμένο σε ένα πολύ διευρυμένο 'σώμα' της μινωικής γραφής A και της μυκηναϊκής γραφής B. Η μελέτη που εκπονήθηκε το 1996 από τον Owens, βασίζεται σε όλο το υλικό της γραμμικής A και B από την Κρήτη, δηλαδή βασίζεται σε μια βάση δεδομένων σχεδόν διπλάσια από αυτή που διέθετε ο Packard. Ο Packard ήταν σε θέση να καταδείξει μια διαφορά μεταξύ της γραμμικής A και B στο 19%, ενώ ο Owens κατέδειξε την ανάλογη διαφορά στο 10,90%.

Συμπερασματικά, είναι δυνατόν να πούμε ότι η συσχέτιση μεταξύ των φωνητικών αξιών της γραμμικής A και της γραμμικής B ανέρχεται σήμερα σε μια διαφορά της τάξεως του 10,90%. Προβλέπεται ότι ο αριθμός αυτός δε θα αλλάξει δραματικά ως αποτέλεσμα των νέων ανακαλύψεων. Η σχέση και ο συσχετισμός μεταξύ γραμμικής A και γραμμικής B όπως καταδεικνύεται από τον Owens (1996) μπορεί να αποδειχθεί στατιστικά πως είναι σε μια περιοχή 9-13%.

1.6 Μια μικρή περίληψη των παραπάνω

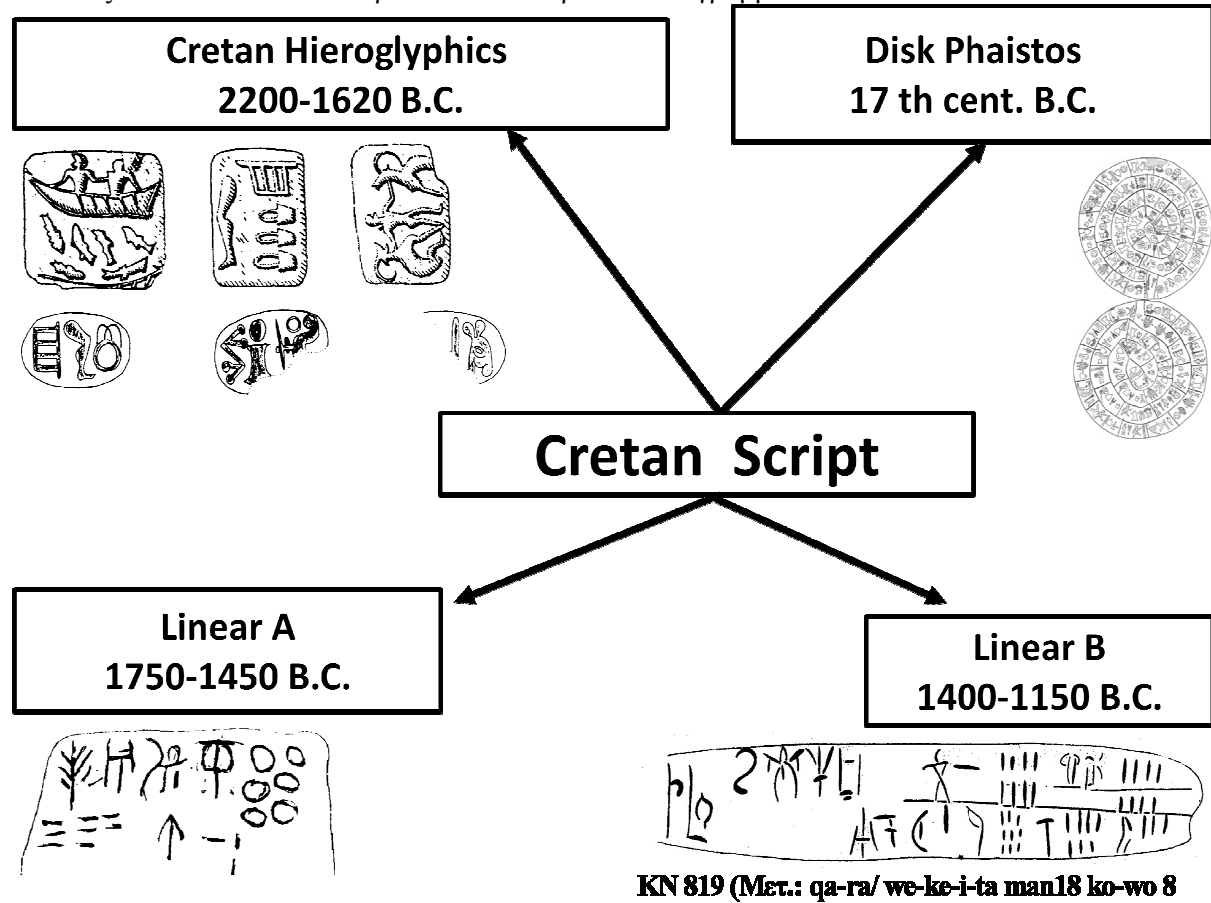
Η πρώτη στατιστική εμφάνιση των συχνοτήτων των δειγμάτων της γραμμικής B και του δείγματος της γραμμικής A που υπήρχαν μέχρι το 1974, με τις θέσεις αρχή, μέση και τέλος των συλλαβών, μελετήθηκε και εκδόθηκε από τον κύριο David Packard και τις εκδόσεις University of California. Νέα στατιστική μελέτη με ποιο συμπληρωμένα τα δείγματα εκδόθηκε το 2000 από τον κύριο Μηνά Τσικριτζή και τις εκδόσεις Βικελαία δημοτική βιβλιοθήκη. Λαμβάνοντας υπόψη μας τα δείγματα των παραπάνω εργασιών, επεξεργαστήκαμε στον ηλεκτρονικό υπολογιστή τις σχετικές συχνότητες εμφάνισης των συλλαβών στην αρχή, μέση και τέλος των λέξεων.

1.6.1 Μια μικρή αναφορά στις αιγαιακές γραφές

Η επινόηση της γραφής στην Κρήτη φαίνεται να υπαγορεύεται από την ανάγκη εξυπηρετήσεως των διοικητικών αναγκών του πολιτισμού της, όπως προκύπτει από τις λογιστικές πινακίδες και τα άλλα οικονομικά τεκμήρια, αλλά και την ανάγκη τελετουργικής επικοινωνίας που καταγράφεται στις τράπεζες προσφορών. Σύμφωνα με τα αρχαιολογικά δεδομένα, στο πλαίσιο του μινωικού και του μυκηναϊκού πολιτισμού, αναπτύχθηκαν τρία συλλαβικά συστήματα γραφής, τα οποία συνετέλεσαν για μία περίπου χιλιετία στην γραπτή επικοινωνία μεταξύ των "πρωτοεγγράμματων" κοινοτήτων του Αιγαίου. Στο χρονικό διάστημα 2200 έως 1150 π.Χ., τα διάφορα ευρήματα γραφής κατατάσσονται και διακρίνονται στον παρακάτω πίνακα.

1.6.2 Αιγαιακές Γραφές

Οι διάφορες γραφές που αναπτύχθηκαν στην περιοχή της Κρήτης, στο χρονικό διάστημα από το 2200 π.Χ. έως και το 1100 π.Χ. διακρίνονται στο παρακάτω διάγραμμα



Σχήμα 1.6.2.1: Αιγαιακές γραφές

Η γραμμική Β είναι σε ικανοποιητικό βαθμό κατανοητή σήμερα, μετά την αποκρυπτογράφηση της από τον Michael Ventris και την φιλολογική επιμέλεια του John Chadwick, οι οποίοι έδειξαν ότι η επιγραφές έγραφαν ελληνικά πολλούς αιώνες πριν τον Όμηρο. Αποτέλεσμα ήταν να μετατεθεί η ιστορία της ελληνικής γραφής και γλώσσας τουλάχιστον στον 14^ο αιώνα π.Χ. από τον 8^ο αιώνα π.Χ. που θεωρείτο μέχρι πρότινος.

Η γραμμική Β λοιπόν είναι γνωστή. Μελετάμε κατά πόσο μπορούμε να προσεγγίσουμε την γραμμική Α με την χρήση της Γραμμικής Β. Αυτό θα το προσπαθήσουμε με την χρήση των ευφών συστημάτων και συγκεκριμένα με την αναγνώριση προτύπων.

ΚΕΦΑΛΑΙΟ 2

2.1 Αναγνώριση προτύπων.

Η αναγνώριση προτύπων είναι ο επιστημονικός κλάδος που ασχολείται με την περιγραφή και κατάταξη αντικειμένων σε έναν αριθμό κατηγοριών. Τα υπό κατάταξη αντικείμενα καλούνται πρότυπα. Η αναγνώριση προτύπων θεωρείται ένα βασικό χαρακτηριστικό των ανθρώπων καθώς και άλλων ζώντων οργανισμών. Ένα πρότυπο είναι η περιγραφή ενός αντικειμένου. Καθημερινά αναγνωρίζουμε πρότυπα γύρω μας όπως: αντικείμενα, το πρόσωπο ενός φίλου μέσα στο πλήθος και μπορούμε να καταλάβουμε αν είναι θυμωμένος ή χαρούμενος από την έκφρασή του καθώς και να αναγνωρίσουμε την φωνή του χωρίς να τον βλέπουμε.

Ανάλογα με την φύση των προτύπων προς αναγνώριση μπορούμε να διαχωρίσουμε τον ανθρώπινο τρόπο αναγνώρισης προτύπων σε δύο κατηγορίες: την αναγνώριση σαφών προτύπων όπως είναι χαρακτήρες, εικόνες, αντικείμενα και ήχους και την αναγνώριση αφηρημένων προτύπων όπως η λύση ενός μαθηματικού προβλήματος ή ενός φιλοσοφικού επιχειρήματος. Η αναγνώριση αφηρημένων προτύπων μπορεί να θεωρηθεί σαν εννοιολογική αναγνώριση σε αντιδιαστολή με την οπτική και ακουστική αναγνώριση που είναι αναγνώριση σαφών προτύπων. Η διαδικασία αναγνώρισης σαφών προτύπων εμπλέκει τον καθορισμό και ταξινόμηση αφενός προτύπων στο χώρο όπως είναι τυπωμένη χαρακτήρες, άνθρωποι, δακτυλικά αποτυπώματα, εικόνες, χάρτες καιρού, φυσικά αντικείμενα και αφετέρου προτύπων στο χρόνο δηλαδή χρονοσειρές όπως είναι ακουστικές κυματογράφος, ηλεκτροεγκεφαλογραφήματα, και σήματα ραντάρ.

Με απλά λόγια η αναγνώριση προτύπων μπορεί να οριστεί ως η κατηγοριοποίηση δεδομένων εισόδου στο σύστημα σε αναγνωρίσιμες κατηγορίες μέσω της εξαγωγής σημαντικών ορισμάτων ή χαρακτηριστικών από τα δεδομένα εισόδου, παραλείποντας τις άσχετες πληροφορίες.

Η πρόγνωση από μετεωρολογικούς χάρτες μπορεί να θεωρηθεί σαν ένα πρόβλημα αναγνώριση προτύπων. Το σύστημα κάνει μια πρόγνωση βασισμένο σε πληροφορίες που εξάγει από τους χάρτες. Η ιατρική διάγνωση μπορεί να θεωρηθεί σαν ένα πρόβλημα αναγνώριση προτύπων. Τα συμπτώματα είναι οι εισοδοί και το σύστημα εξάγει μία ιατρική γνωμάτευση αναλύοντας τις εισόδους του.

2.2 Μεθοδολογίες αναγνώρισης προτύπων.

Όπως προαναφέραμε υπάρχουν δύο είδη προτύπων: πρότυπα στο χώρο και πρότυπα στο χρόνο. Χρησιμοποιώντας εβραίος τον όρο αναγνώριση προτύπων, μπορούμε να ανακαλύψουμε σε κάθε ευφυή κάποια μορφή αναγνώριση προτύπων. Δεν είναι δυνατών να υπάρχει μόνο ένας τρόπος προσέγγισης της θεωρίας αναγνώριση προτύπων για τον τεράστιο αριθμό και εύρος των εφαρμογών που υπάρχουν. Γενικά, θα μπορούσαμε να χωρίσουμε τις κύριες μεθοδολογίες αναγνώριση προτύπων σε δύο μεγάλες κατηγορίες: α) στατιστική αναγνώριση προτύπων, β) συντακτική ή δομημένη αναγνώριση προτύπων. Η στατιστική αναγνώριση προτύπων προϋποθέτει μία στατιστική βάση για την ταξινόμηση των αντικειμένων και στηρίζεται στην πιθανολογική φύση των αντικειμένων – προτύπων. Η περιγραφή των αντικειμένων πραγματοποιείται με την εξαγωγή χαρακτηριστικών γνωρισμάτων που το περιγράφουν και την δημιουργία ενός διανύσματος χαρακτηριστικών στοιχείων για κάθε αντικείμενο. Το πρόβλημα ανάγεται στην κατάταξη του διανύσματος χαρακτηριστικών κάθε αντικειμένου στην κατηγορία που ανήκει χρησιμοποιώντας μαθηματικές – στατιστικές μεθόδους, γραμμική άλγεβρα και θεωρία πιθανοτήτων.

Σε πολλές περιπτώσεις οι σχέσεις μεταξύ των χαρακτηριστικών στοιχείων ενός αντικειμένου φέρουν σπουδαία δομημένη πληροφορία, που μπορεί να χρησιμοποιηθεί για την ταξινόμηση και περιγραφή του αντικειμένου. Η μέθοδος αυτή καλείται συντακτική ή δομημένη αναγνώριση προτύπων. Τυπικές προσεγγίσεις της συντακτικής ή δομημένης αναγνώριση προτύπων είναι η δημιουργία πολύπλοκων ιεραρχικών περιγραφών των προτύπων, τα οποία δημιουργούνται από απλούστερα υπό-πρότυπα. Στο απλούστερο επίπεδο του μοντέλου απλά στοιχεία περιγραφής εξάγονται από τα δεδομένα εισόδου.

Τεχνικές οι οποίες χρησιμοποιούνται για την υλοποίηση αλγορίθμων συντακτικής ή δομημένης αναγνώριση προτύπων είναι δέντρα αποφάσεων, λογική κανόνες και γραμματικές. Το τελικό αποτέλεσμα είναι μία σειρά κανόνων που περιγράφουν πλήρως την διαδικασία ταξινόμησης ή μια γραμματική που περιγράφει πλήρως το αντικείμενο.

Η συντακτικές μεθοδολογίες είναι πολύπλοκες και πολύ ευαίσθητες στην παρουσία θορύβου και γενικά είναι πολύ δύσκολο να ανταποκριθούν αποτελεσματικά σε μικρές παραλλαγές των προτύπων ή και σε ελλιπείς πληροφορίες. Χρησιμοποιούνται κυρίως σε εφαρμογές όπου η δυνατότητες της στατιστικής αναγνώριση προτύπων είναι περιορισμένες. Ένα μεγάλο πρόβλημα των συντακτικών μεθοδολογιών είναι ότι δεν είναι πάντα εύκολο να οριστούν πλήρως και χωρίς αμφιβολία τα πρότυπα του προβλήματος. Για παράδειγμα, δεν είναι καθόλου εύκολο να περιγραφεί με λογικούς κανόνες τι είναι ένα δέντρο και να διαχωριστεί αποτελεσματικά από άλλα παρεμφερή αντικείμενα. Η ανθρώπινη αντίληψη μπορεί εύκολα να διαχωρίσει τα δέντρα και να τα αναγνωρίσει. Αντίθετα, είναι αδύνατη η περιγραφή τους, και συνεπώς ο διαχωρισμός τους, με λογικούς κανόνες.

Η στατιστική αναγνώριση προτύπων είναι πολύ ισχυρά θεμελιωμένη μαθηματικά ενώ η συντακτική αναγνώριση προτύπων είναι περισσότερο βασισμένη σε λογικούς και διαισθητικούς κανόνες. Το πλήθος και η σειρά στο διάνυσμα των χαρακτηριστικών στοιχείων ενός προτύπου είναι πάντα σταθερό στην στατιστική προσέγγιση, αντίθετα με την δομημένη προσέγγιση όπου το πλήθος και η σειρά των χαρακτηριστικών στοιχείων μεταβάλλονται από πρότυπο σε πρότυπο.

Εμείς θα περιοριστούμε στην στατιστική αναγνώριση προτύπων. Έτσι θα εξεταστεί η αναγνώριση προτύπων ως πρόβλημα ταξινόμησης δηλαδή κατάταξης μιας εισόδου σε μία κατηγορία. Βέβαια υπάρχουν και νέες μεθοδολογίες αναγνώριση προτύπων οι οποίες χρησιμοποιούν έμπειρα συστήματα, νευρωνικά δίκτυα, γενετικούς αλγόριθμους ή συνδυασμούς αυτών.

2.3 Ιστορική αναδρομή

Η βάσης της αναγνώριση προτύπων πρωτοτέθηκαν από τον Πλάτωνα και τον Αριστοτέλη που πρώτοι έκαναν την διάκριση μεταξύ της ουσιώδους ιδιότητας (που μοιράζονται μεταξύ τους τα μέλη μιας κατηγορίας) και τις επουσιώδους ιδιότητας (που διαφέρει για τα μέλη μιας κατηγορίας). Η αναγνώριση προτύπων μπορεί να οριστεί σαν την διαδικασία που βρίσκει τέτοιες ουσιώδους ιδιότητες μέσα σε μια κατηγορία αντικειμένων. Η προβληματισμοί τους παραμένουν βασικοί για την επιστημολογία. Ο Αριστοτέλης κατασκεύασε ένα σύστημα ταξινόμησης των ζώων χωρίζοντάς τα αρχικά σε αυτά που έχουν κόκκινο αίμα (σπονδυλωτά) και σε αυτά που δεν έχουν (ασπονδύλωτα). Στην συνέχεια χώρισε τις δύο ομάδες σε μικρότερες χρησιμοποιώντας άλλα χαρακτηριστικά. Στην συνέχεια Θεόφραστος έκανε μια ανάλογη ταξινόμηση των φυτών. Η ταξινόμηση ήταν τόσο καλή που μόλις τον 18ο αιώνα ο Carolus Linnaeus κατασκεύασε συστηματικές ταξινομήσεις ζώων, φυτών, πετρωμάτων και ασθενειών χρησιμοποιώντας της νέες γνώσης που έφερε ο αιώνας των μεγάλων εξερευνησεων. Παρόμοια ο Hertzprung και Russell ταξινόμησαν τα άστρα σε κατηγορίες στηριζόμενοι σε δύο μεταβλητές: στην λαμπρότητα και στην θερμοκρασία της επιφάνειάς τους.

Η πρώτη συστηματική προσπάθεια για την μαθηματική μορφοποίηση του προβλήματος έγινε από τον Fisher το 1936, αλλά έπρεπε να έρθει η ανακάλυψη των υπολογιστών για να θεωρηθεί σαν μία ανεξάρτητη επιστήμη. Η αναγνώριση προτύπων χρησιμοποιείται σε πολλές επιστημονικές περιοχές και σήμερα υπάρχουν δεκάδες επιστημονικά περιοδικά, εκατοντάδες βιβλία και πρακτικά συνεδρίων πάνω σε τομείς της αναγνώριση προτύπων. Κάποιοι επιστημονικοί κλάδοι όπως η στατιστική, η μηχανική μάθηση και τα τεχνητά νευρωνικά δίκτυα έχουν επεκτείνει κατά πολύ το γνωστικό αντικείμενο της αναγνώριση προτύπων, ενώ ορισμένοι άλλοι, όπως η υπολογιστική όραση και η αναγνώριση φωνής στηρίζονται σχεδόν ολοκληρωτικά σε αυτή. Η γνωσιολογική επιστήμη, η ψυχοβιολογία και νευροεπιστήμη μελετούν τους μηχανισμούς αναγνώρισης προτύπων των ανθρώπων και των ζώων. Τεχνικές αναγνώρισης προτύπων έχουν εφαρμοστή έμμεσα ή άμεσα σε όλες σχεδόν τις επιστημονικές περιοχές.

2.4 Σύστημα αναγνώρισης προτύπων

Ο σχεδιασμός ενός αυτόματου συστήματος περιλαμβάνει διάφορα βήματα. Ως πρώτο βήμα πραγματοποιείτε η αναπαράσταση των προτύπων με μία διαδικασία συλλογής δεδομένων. Στις περισσότερες εφαρμογές αναγνώρισης προτύπων εμφανίζονται δύο τύποι προτύπων: πρότυπα στο χρόνο (χρονοσειρές) και πρότυπα στο χώρο (γεωμετρικά αντικείμενα). Το δεύτερο βήμα στην αναγνώριση προτύπων είναι η εξαγωγή χαρακτηριστικών γνωρισμάτων από τα δεδομένα εισόδου και η ελάττωση της διάστασης των διανυσμάτων προτύπων. Αυτή η διαδικασία ονομάζεται εξαγωγή χαρακτηριστικών γνωρισμάτων. Τα χαρακτηριστικά γνωρίσματα μιας κατηγορίας είναι η χαρακτηριστικές ιδιότητες οι οποίες είναι κοινές για όλα τα πρότυπα που ανήκουν σε αυτή την κατηγορία και ονομάζονται ενδοσυνολικά γνωρίσματα. Τα χαρακτηριστικά τα οποία

αντιπροσωπεύουν τις διαφορές μεταξύ κατηγοριών ονομάζονται διασυνολικά γνωρίσματα. Τα στοιχεία των ενδοσυνολικών γνωρισμάτων τα οποία είναι κοινά για όλες τις κατηγορίες δεν περιέχουν καμία διαχωριστική πληροφορία και μπορούν να αγνοηθούν. Εάν μπορεί να καθοριστεί ένα πλήρες σύνολο διαχωριστικών γνωρισμάτων για κάθε κατηγορία από τις μετρήσεις των δεδομένων, τότε η αναγνώριση και η ταξινόμηση των προτύπων είναι πολύ εύκολη. Όμως, στα περισσότερα πρακτικά προβλήματα αναγνώρισης αυτός ο καθορισμός είναι πολύ δύσκολος, αν όχι αδύνατος. Ευτυχώς, πολλές φορές είναι δυνατό να βρεθούν ορισμένα διαχωριστικά χαρακτηριστικά γνωρίσματα από τα δεδομένα εισόδου. Το τρίτο βήμα στο σχεδιασμό ενός συστήματος αναγνώρισης προτύπων είναι ο καθορισμός βέλτιστων διαδικασιών απόφασης, οι οποίες είναι αναγκαίες για την αναγνώριση και την ταξινόμηση. Μετά την δημιουργία του διανύσματος χαρακτηριστικών γνωρισμάτων του προτύπου, το σύστημα πρέπει να αποφασίσει σε ποια κατηγορία ανήκει το πρότυπο. Έστω ότι το σύστημα αναγνώρισης προτύπων είναι σχεδιασμένο κατά τρόπο τέτοιο, ώστε να αναγνωρίζει M διαφορετικές κατηγορίες προτύπων C_1, C_2, \dots, C_M . Τότε ο χώρος προτύπων αποτελείται από M περιοχές, κάθε μία από τις οποίες περιέχει τα σημεία προτύπων αυτής της κατηγορίας. Το πρόβλημα της αναγνώρισης προτύπων ανάγεται στην δημιουργία των ορίων απόφασης τα οποία διαχωρίζουν τις M κατηγορίες βασισμένα στα διανύσματα χαρακτηριστικών γνωρισμάτων των προτύπων.

2.5 Απλοί ταξινομητές

Υπάρχουν τουλάχιστον δύο προσεγγίσεις για το σχεδιασμό ενός ταξινομητή. Η πρώτη προσέγγιση είναι θεωρητική. Αρχικά, δημιουργείτε ένα μαθηματικό μοντέλο του προβλήματος και στην συνέχεια βάση του μοντέλου σχεδιάζεται ένας βέλτιστος ταξινομητής. Η δεύτερη προσέγγιση είναι η πρακτική εφαρμογή. Αρχικά πραγματοποιείτε η υπόθεση μίας πιθανής λύσης, σύμφωνα με τα δείγματα των κατηγοριών του προβλήματος και κατόπιν πραγματοποιείτε η βέλτιστη αναπροσαρμογή της λύσης σύμφωνα με τα πρακτικά δεδομένα του προβλήματος. Η δεύτερη προσέγγιση είναι εμπειρική μέθοδος, χρησιμοποιείτε εβραίος σε πρακτικές εφαρμογές αναγνώρισης προτύπων και είναι η προσέγγιση που θα ακολουθήσουμε.

Η πιο απλή και εμπειρική προσέγγιση στο πρόβλημα της αναγνώρισης προτύπων είναι η ιδέα της ταξινόμησης με την χρήση συναρτήσεων απόστασης. Η χρήση συναρτήσεων απόστασης είναι φυσικό επακόλουθο του γεγονότος ότι ο πιο προφανής τρόπος καθορισμού ενός μέτρου ομοιότητας μεταξύ των προτύπων, τα οποία μπορούν να θεωρηθούν σαν σημεία στο ευκλείδεια χώρο, είναι ο υπολογισμός της απόστασης μεταξύ τους. Για να είναι αποτελεσματική η ταξινόμηση με χρήση συναρτήσεων απόστασης θα πρέπει οι κατηγορίες του προβλήματος να σχηματίζουν ευδιάκριτες ομάδες στο χώρο προτύπων.

2.6 Ταξινομητές ελάχιστης απόστασης.

Η αναγνώριση προτύπων χρησιμοποιώντας συναρτήσεις απόστασης είναι μια από τις πρωτοπόρες τεχνικές που αναπτύχθηκαν στον τομέα της αναγνώρισης προτύπων. Είναι βασικά μια απλή τεχνική ταξινόμησης, η οποία όμως είναι αποτελεσματική σε προβλήματα που οι κατηγορίες τους είναι σχετικά καλά διαχωρίσιμες.

2.7 Μέτρα απόστασης

Διαισθητικά μπορεί να εκτιμηθεί ότι αντικείμενα τα οποία βρίσκονται «κοντά» στο χώρο προτύπων είναι όμοια μεταξύ τους, ενώ αντικείμενα τα οποία βρίσκονται «μακριά» είναι ανόμοια. Για την μαθηματική έκφραση της απόστασης μεταξύ αντικειμένων στο χώρο προτύπων πρέπει να οριστεί το μέτρο ομοιότητας μεταξύ των προτύπων. Ο πιο προφανής τρόπος μέτρησης της απόστασης μεταξύ δύο σημείων είναι η γνωστή μέθοδος της ευκλείδειας απόστασης. Η ευκλείδεια απόσταση μεταξύ των προτύπων x και z διάστασης n , δίνετε από την σχέση:

$$\begin{aligned} D_2(x, z) &= \|x - z\|_2 = \sqrt{(x - z)^T (x - z)} \\ &= \left[\sum_{i=1}^n (x_i - z_i)^2 \right]^{\frac{1}{2}}. \end{aligned} \quad (4.8)$$

Για παράδειγμα, εάν $x = [1, 2, 2, 0]^T$ και $z = [2, 1, 2, 2]^T$ τότε η Ευκλείδεια απόσταση τους είναι:

$$D_2(x, z) = \sqrt{(1-2)^2 + (2-1)^2 + (2-2)^2 + (0-2)^2} = \sqrt{6}.$$

Η ευκλείδεια απόσταση είναι πιθανά το πιο ευρέως χρησιμοποιούμενο μέτρο απόστασης, για μετρήσεις ανομοιότητας μεταξύ χαρακτηριστικών διανυσμάτων, αλλά δεν είναι πάντα κατάλληλη για όλες τις εφαρμογές. Είναι χρήσιμη όταν το πρόβλημα εμπεριέχει πρότυπα με συνεχής τιμές. Όμως σε μία ιατρική εφαρμογή όπου το ένα χαρακτηριστικό γνώρισμα αντιπροσωπεύει το επίπεδο ζαχάρου στο αίμα και το άλλο το βάρος του ασθενούς, δεν φαίνεται λογικό να σχηματιστεί διάνυσμα και να χρησιμοποιηθεί η ευκλείδεια απόσταση για τον διαχωρισμό κατηγοριών ασθενών.

Πολλά προβλήματα περιλαμβάνουν ακέραιες τιμές. Το γεγονός ότι οι διαφορές των χαρακτηριστικών διανυσμάτων υψώνονται στο τετράγωνο έχει σαν αποτέλεσμα να υπερτονίζονται αυτές, ενώ απαιτούνται και επιπλέον υπολογισμοί. Μια απλή προσέγγιση είναι να αθροιστούν οι απόλυτες διαφορές μεταξύ των στοιχείων των χαρακτηριστικών διανυσμάτων. Το μέτρο απόστασης που προκύπτει ονομάζεται Ιπποδάμεια μετρική ή μετρική πρώτης τάξης (στην διεθνή βιβλιογραφία είναι γνωστό με τα ονόματα Manhattan distance ή city block distance ή taxi distance) και ορίζεται από την σχέση:

$$D_1(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\|_1 = \sum_{i=1}^n |x_i - z_i|.$$

Σχήμα 2.7.2:

Τα διανύσματα της Ιπποδάμειας μετρικής περιλαμβάνουν μόνο ακέραιες τιμές και η αναπαράστασή τους στο χώρο γίνεται μόνο σε συγκεκριμένα σημεία.

Σε ορισμένες περιπτώσεις δεν έχουμε αριθμητικά δεδομένα αλλά χαρακτηριστικά του αντικειμένου τα οποία περιγράφονται ποιοτικά. Πολλά από τα ποιοτικά δεδομένα μπορούν να εκφραστούν σε διαδική μορφή όπως 'ΝΑΙ-ΟΧΙ', 'ΠΑΡΟΝ-ΑΠΩΝ'. Το διάνυσμα χαρακτηριστικών που προκύπτει παίρνει τιμές 1 ή 0. Το πιο γνωστό διαδικό μέτρο απόστασης είναι η απόσταση hamming, η οποία μετράει τον αριθμό των θέσεων του δυαδικού διανύσματος όπου τα πρότυπα διαφέρουν.

Για το μαθηματικό ορισμό της απόστασης Hamming πρέπει να εισάγουμε τον γνωστό δυαδικό τελεστή Exclusive OR, του οποίου το σύμβολο είναι \oplus και ορίζεται ως:

$$0 \oplus 0 = 0, 1 \oplus 1 = 0, 0 \oplus 1 = 1, 1 \oplus 0 = 1.$$

Η απόσταση Hamming μεταξύ δύο διανυσμάτων ορίζεται ως:

$$D_H(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\|_1 = \sum_{i=1}^n x_i \oplus z_i \quad (4.10)$$

Η απόσταση Hamming είναι μια υποπερίπτωση της Ιπποδάμειας απόστασης και ταυτίζονται όταν τα διανύσματα είναι δυαδικά.

Η απόσταση Hamming μεταξύ των διανυσμάτων $\mathbf{x} = [1, 0, 0, 1, 1]^T$ και $\mathbf{z} = [1, 1, 0, 1, 0]^T$ είναι:

$$D_H(\mathbf{x}, \mathbf{z}) = |1 - 1| + |0 - 1| + |0 - 0| + |1 - 1| + |1 - 0| = 2.$$

Όλες οι μετρικές που εξετάστηκαν μέχρι τώρα είναι ειδικές περιπτώσεις της απόστασης του Minkowsky η οποία ορίζεται ως:

$$D_s(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\|_s = \left[\sum_{i=1}^n |x_i - z_i|^s \right]^{\frac{1}{s}}$$

Με $s=2$ σχηματίζεται η ευκλείδεια μετρική και με $s=1$ η Ιπποδάμεια. Στην περίπτωση που $s=\infty$ έχουμε την ειδική περίπτωση της απόστασης του Chebyshev η οποία ορίζεται ως:

Δηλαδή είναι η μέγιστη απόσταση μεταξύ των αντίστοιχων στοιχείων των δυο διανυσμάτων.

Η Chebyshev απόσταση μεταξύ των διανυσμάτων $\mathbf{x} = [1, 2, 2, 0]^T$ και $\mathbf{z} = [2, 1, 2, 2]^T$ είναι ίση με:

$$D_{\infty}(\mathbf{x}, \mathbf{z}) = \max \{|1 - 2|, |2 - 1|, |2 - 2|, |0 - 2|\} = \\ = \max \{1, 1, 0, 2\} = 2.$$

Εκτός από τις παραπάνω αναφερόμενες μετρικές συναρτήσεις υπάρχουν και άλλα χρήσιμα μέτρα απόστασης. Ένα πολύ χρήσιμο μέτρο απόστασης το οποίο λαμβάνει υπόψην τους στατικούς δείκτες είναι η απόσταση mahalanobis

$$D_M(\mathbf{x}, \mathbf{z}) = (\mathbf{x} - \mathbf{z})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{z}),$$

Στον παρακάτω πίνακα παρουσιάζονται συνοπτικά τα διάφορα μέτρα απόστασης, οι μαθηματική ορισμοί τους καθώς και παρατηρήσεις πάνω στην χρησιμότητά τους.

Ιπποδάμεια	$D_1(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^n x_i - z_i $	Ακέραια στοιχεία
Ευκλείδεια	$D_2(\mathbf{x}, \mathbf{z}) = [\sum_{i=1}^n (x_i - z_i)^2]^{\frac{1}{2}}$	Συνεχή στοιχεία
Hamming	$D_H(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^n x_i \oplus z_i$	Διαδικά στοιχεία
Minkowsky	$D_s(\mathbf{x}, \mathbf{z}) = [\sum_{i=1}^n x_i - z_i ^s]^{\frac{1}{s}}$	Γενικός τύπος
Mahalanobis	$D_M(\mathbf{x}, \mathbf{z}) = (\mathbf{x} - \mathbf{z})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{z})$	Στατιστική

ΠΙΝΑΚΑΣ 2.7.1: Συνοπτικός πίνακας μέτρων απόστασης

ΚΕΦΑΛΑΙΟ 3

3.1 Δημιουργία ομάδων

Σε ένα μεγάλο αριθμό επιστημονικών κλάδων χρειάζεται να ταξινομηθούν δεδομένα βάση της ομοιότητας που υπάρχει στα δεδομένα. Σε πολλές περιπτώσεις υπάρχει πολύ μικρή γνώση για την μορφή και τους νόμους που διέπουν τα δεδομένα και πρέπει να γίνουν όσο το δυνατόν λιγότερες αυθαίρετες υποθέσεις γι' αυτά. Αρχικά, απαιτείται μια πρώτη εξέταση των εσωτερικών σχέσεων μεταξύ των δεδομένων, έτσι ώστε να γίνουν κάποιες πρώτες υποθέσεις για την δομή τους.

Η ομαδοποίηση είναι εργαλείο διερευνητικής ανάλυσης δεδομένων που προσπαθεί να εκτιμήσει τις σχέσεις που υπάρχουν μεταξύ προτύπων οργανώνοντας τα σε ομάδες ή κατηγορίες με τέτοιο τρόπο ώστε πρότυπα που ανήκουν σε μία ομάδα να έχουν περισσότερη ομοιότητα μεταξύ τους από ότι πρότυπα που ανήκουν σε διαφορετικές ομάδες. Τα αποτελέσματα της ομαδοποίησης μπορεί να χρησιμοποιηθούν για την εξαγωγή υποθέσεων που αφορούν τα δεδομένα για την ταξινόμηση νέων δεδομένων, για τον έλεγχο ομοιογένειας των δεδομένων καθώς και για συμπίεση δεδομένων.

Η πλειονότητα των τεχνικών τείνουν να δημιουργούν ομάδες συγκεκριμένης μορφής. Η ανθρώπινη αντίληψη είναι η καλύτερη τεχνική ομαδοποίηση στο δισδιάστατο και τρισδιάστατο χώρο, αλλά δυστυχώς τα περισσότερα πραγματικά προβλήματα αφορούν ομαδοποιήσεις σε υψηλότερες διαστάσεις. Άλλωστε τα πραγματικά δεδομένα σπάνια ακολουθούν τις ιδανικές δομές.

Οι αλγόριθμοι ομαδοποίησης τείνουν να δημιουργούν ομάδες συγκεκριμένης μορφής. Έτσι, από το ίδιο σύνολο δεδομένων δημιουργούνται διαφορετικές ομαδοποιήσεις ανάλογα με την μέθοδο που κάθε φορά χρησιμοποιείτε. Στην πραγματικότητα οι αλγόριθμοι θα δημιουργήσουν ομάδες ακόμα και αν τα δεδομένα είναι εντελώς τυχαία. Αυτό οφείλετε στο ότι οι ομάδες δεν είναι ένα χαρακτηριστικό των δεδομένων, το οποίο μπορεί να ελεγχθεί με ένα αντικειμενικό κριτήριο, αλλά μια οπτική του ερευνητή πάνω στα δεδομένα.

Το μεγαλύτερο ρίσκο που προκύπτει κατά την εφαρμογή μεθόδων ομαδοποίησης είναι ότι μπορεί να έχει σαν επακόλουθο αντί της εύρεσης μίας «φυσικής» δομής των δεδομένων, την επιβολή μιας αυθαίρετης και τεχνητής δομής. Τα αποτελέσματα μίας μεθόδου ομαδοποίησης πρέπει να γίνονται πάντα δεκτά με σκεπτικισμό. Εάν τα αποτελέσματα μίας τεχνικής ομαδοποίησης αυξάνουν την κατανόηση μας για το πρόβλημα, τότε υπάρχει μεγάλη πιθανότητα οι ομάδες να έχουν επιλεγεί σωστά. Εάν το αποτέλεσμα έρχεται σε αντίθεση με την διαίσθηση μας ή συσκοτίζει αντί να φωτίζει το πρόβλημα τότε δεν υπάρχουν πολλές πιθανότητες να επιτευχθούν οι στόχοι μας. Εάν δεν είναι δυνατόν να βρεθεί ένα αξιοπρεπές σχήμα ομαδοποίησης αυτό αποτελεί ένδειξη ότι δεν έχει γίνει πλήρως κατανοητό το πρόβλημα και πρέπει να γίνει νέα διερεύνησή του. Στην περίπτωση που δύο διαφορετική διαχωρισμοί φαίνονται εξίσου σωστή, τότε τον τελευταίο λόγο έχει ένας εμπειρογνώμονας.

3.2 Εφαρμογές

Τεχνικές ομαδοποίησης έχουν εφαρμοστή σε πολλά και διαφορετικά ερευνητικά προβλήματα. Για παράδειγμα, στην ιατρική ομαδοποιήσεις ασθενειών και συμπτωμάτων έχουν σαν αποτέλεσμα την δημιουργία πολύ χρήσιμων ταξινόμησεων. Στην ψυχιατρική, η σωστή διάγνωση ομάδων συμπτωμάτων όπως η παράνοια και η σχιζοφρένεια είναι βασική προϋπόθεση για μία επιτυχημένη θεραπεία. Στην βιολογία χρησιμοποιούνται τεχνικές ανάλυσης ομάδων για την ταξινόμηση διαφόρων οργανισμών. Στην αρχαιολογία ερευνητές έχουν εφαρμόσει τεχνικές ομαδοποίησης για την ταξινόμηση πέτρινων εργαλείων και ταφικών κτερισμάτων. Στην ανάλυση εικόνας χρησιμοποιείτε για την εύρεση ομάδων εικονοκυττάρων με παρόμοια χαρακτηριστικά όπως χρώμα ή υφή.

3.3 Ιδανικές ομάδες

Οι ιδιότητες που πρέπει να έχουν τα πρότυπα που ανήκουν σε μία ομάδα και τι πρέπει να ισχύει για να μπορεί αυτή η ομάδα να θεωρηθεί ιδανική είναι:

- 1) Μια ομάδα αποτελείται από ένα σύνολο παρόμοιων προτύπων. Πρότυπα από διαφορετικές ομάδες έχουν διαφορετικά χαρακτηριστικά.

- 2) Γενικά η απόσταση μεταξύ των προτύπων που ανήκουν σε μία ομάδα είναι μικρότερη από την απόσταση μεταξύ προτύπων που ανήκουν σε διαφορετικές ομάδες.
- 3) Οι ομάδες αποτελούν συνδεδεμένες περιοχές στο χώρο προτύπων με σχετικά μεγάλη πυκνότητα προτύπων, και διαχωρίζονται από τις άλλες ομάδες με περιοχές με χαμηλή πυκνότητα προτύπων.

Είναι προφανές ότι ο αρχικός σχεδιασμός ενός συστήματος αναγνώρισης προτύπων καθορίζει το τι σημαίνει ομάδα για την εφαρμογή και της απαιτήσεις από την μέθοδο ομαδοποίησης που θα χρησιμοποιηθεί. Ένα από τα πρώτα βήματα για την κατασκευή ενός συστήματος δεδομένων είναι η δημιουργία μίας αντίληψης για το πώς σχηματίζονται οι ομάδες. Για το σκοπό αυτό μπορούν να χρησιμοποιηθούν μαθηματικά μοντέλα που περιέχουν εκ των προτέρων γνώση για το πρόβλημα ή πρόχειρες στατιστικές αναλύσεις.

3.4 Μεθοδολογίες ομαδοποίησης

Οι περισσότεροι αλγόριθμοι ομαδοποίησης μπορούν να ταξινομηθούν σε δύο μεγάλες κατηγορίες: σε ιεραρχικούς και σε διαχωριστικούς αλγόριθμους. Ένας ιεραρχικός αλγόριθμος ομαδοποίησης επιβάλλει μια ιεραρχική δομή στα δεδομένα, δηλαδή τα δεδομένα χωρίζονται σε μεγάλες ομάδες που με την σειρά τους χωρίζονται σε υποομάδες κλπ. Το τελικό αποτέλεσμα ενός ιεραρχικού αλγορίθμου είναι ένα δενδρόγραμμα.

Ο χωρισμός σε ομάδες μπορεί να ξεκινήσει από τις μικρότερες ομάδες προς τις μεγαλύτερες είτε αντίστροφα. Στην πρώτη περίπτωση ο αλγόριθμος ξεκινά με N ομάδες, μία για κάθε πρότυπο και σταδιακά οι ομάδες ενοποιούνται μέχρις ότου φτιαχτεί μια ομάδα που περιέχει όλα τα πρότυπα. Η προσέγγιση αυτή καλείτε συσσωρευτική. Στην δεύτερη περίπτωση ο αλγόριθμος ξεκινά με μια ομάδα που περιέχει όλα τα πρότυπα και με μία διαδικασία βημάτων οι ομάδες διασπώνται μέχρις ότου οι ομάδες περιέχουν μόνο λίγα πρότυπα. Η προσέγγιση αυτή καλείτε διαμοιραστική. Οι ιεραρχικές μεθοδολογίες ομαδοποίησης είναι πολύ διαδεδομένες στις βιολογικές επιστήμες, όπου τα δεδομένα για παράδειγμα φυτά και ζώα συχνά αντιπροσωπεύουν μια ταξινόμια.

Στην διαχωριστική ομαδοποίηση ο στόχος είναι η δημιουργία ενός συνόλου ομάδων οι οποίες διαχωρίζουν τα δεδομένα σε παρόμοιες ενότητες. Πρότυπα τα οποία βρίσκονται σε μικρή απόσταση θεωρούνται ότι είναι όμοια και ο στόχος των διαχωριστικών αλγορίθμων είναι να ομαδοποιηθούν τέτοια δεδομένα. Στους περισσότερους διαχωριστικούς αλγορίθμους το σύνολο των ομάδων που θα δημιουργηθούν προκαθορίζετε. Η διαχωριστικές μέθοδοι χρησιμοποιούν συναρτήσεις κριτηρίων όπως μέθοδοι ελαχιστοποίησης τετραγώνων, εκτιμητές πυκνότητας και πλησιέστερους γείτονες. Οι διαχωριστικές μεθοδολογίες ομαδοποίησης χρησιμοποιούνται σε περιπτώσεις που οι κατηγορίες του προβλήματος δεν σχηματίζουν τόσο ευδιάκριτες ομάδες, αλλά υπάρχει κάποια αλληλοεπικάλυψη μεταξύ τους. Σε αυτές τις περιπτώσεις κάθε πρότυπο μπορεί να ανήκει σε πολλές ομάδες ταυτόχρονα με ένα βαθμό σιγουριάς για κάθε ομάδα που κυμαίνεται από 0 έως 1.

Είναι σημαντικό να γίνει διάκριση μεταξύ μεθόδων ομαδοποίησης και αλγορίθμων ομαδοποίησης. Η ίδια μέθοδος ομαδοποίησης μπορεί να υλοποιηθεί διαφορετικά έχοντας σαν αποτέλεσμα την δημιουργία διαφορετικών αλγορίθμων ομαδοποίησης.

3.5 Ιεραρχική ομαδοποίηση

Μια ιεραρχία μπορεί να αντιπροσωπευτεί από μια δενδρική δομή. Σε μια ιεραρχική ομαδοποίηση εάν δύο πρότυπα ανήκουν στην ίδια ομάδα σε ένα επίπεδο, τότε θα ανήκουν στην ίδια ομάδα σε κάθε υψηλότερο επίπεδο. Ένας συσσωρευτικός ιεραρχικός αλγόριθμος ομαδοποίησης ακολουθεί την παρακάτω μορφή:

- 1) Καταχώρηση κάθε ένα από τα N πρότυπα σε μία μοναδική ομάδα, με αποτέλεσμα την δημιουργία N ομάδων.
- 2) Να βρεθούν οι ομάδες με την μεγαλύτερη ομοιότητα μεταξύ τους και να συγχωνευτούν σε μία νέα ομάδα.
- 3) Επανάληψη του βήματος δύο έως ότου όλα τα πρότυπα να ανήκουν στην ίδια ομάδα.

Χρησιμοποιώντας διαφορετικές μεθόδους για τον καθορισμό της ομοιότητας μεταξύ των ομάδων προκύπτουν διαφορετικοί αλγόριθμοι. Ένας τρόπος μέτρησης της ομοιότητας μεταξύ ομάδων είναι ο ορισμός μίας συνάρτησης μέτρησης της απόστασης μεταξύ των ομάδων. Τα μέτρα απόστασης μεταξύ ομάδων που χρησιμοποιούνται βασίζονται πάνω στα γνωστά μέτρα απόστασης όπως στην ευκλείδεια και την υποδάμια απόσταση.

3.6 Αλγόριθμος απλής σύνδεσης

Ο αλγόριθμος απλής σύνδεσης είναι γνωστός με πολλά ονόματα όπως μέθοδος πλησιέστερης γητνίασης ή ελάχιστη μέθοδος. Στο αλγόριθμο απλής σύνδεσης η απόσταση μεταξύ ομάδων ορίζεται σαν την ελάχιστη απόσταση μεταξύ δύο προτύπων διαφορετικών ομάδων. Με άλλα λόγια η απόσταση μεταξύ δύο ομάδων καθορίζεται από την απόσταση των πιο κοντινών προτύπων διαφορετικών ομάδων.

Έτσι η απόσταση μεταξύ των ομάδων C_i και C_j ορίζεται από την σχέση

$$D_{SL}(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y),$$

όπου $d(x, y)$ είναι η συνάρτηση απόστασης μεταξύ των προτύπων x, y .

Για την πρακτική υλοποίηση του αλγόριθμου πρέπει να βρεθεί η απόσταση κάθε ομάδας με κάθε άλλη και τα αποτελέσματα δημιουργούν ένα συμμετρικό δισδιάστατο πίνακα απόστασης. Η διαστάσεις του πίνακα απόστασης είναι αρχικά $n \times n$, όπου n είναι ο αριθμός των προτύπων. Οι στήλες και οι γραμμές του πίνακα απόστασης αντιπροσωπεύουν τις ομάδες προτύπων και οι τιμές του πίνακα την απόσταση μεταξύ των ομάδων.

3.7 Αλγόριθμος πλήρους σύνδεσης

Ο αλγόριθμος πλήρους σύνδεσης ονομάζεται επίσης μέγιστη μέθοδος ή μέθοδος απομακρυσμένης γειτνίασης. Στο αλγόριθμο πλήρους σύνδεσης η απόσταση μεταξύ δύο ομάδων ορίζεται ως η μέγιστη απόσταση μεταξύ δύο προτύπων όπου κάθε ένα από τα δύο πρότυπα ανήκει σε διαφορετική ομάδα. Με άλλα λόγια η απόσταση μεταξύ ομάδων καθορίζεται από την μέγιστη απόσταση μεταξύ δύο προτύπων που ανήκουν σε διαφορετικές ομάδες.

Έτσι η απόσταση μεταξύ των ομάδων C_i και C_j ορίζεται από την σχέση:

$$D_{CL}(C_i, C_j) = \max_{x \in C_i, y \in C_j} d(x, y),$$

όπου $d(x, y)$ είναι η συνάρτηση απόστασης μεταξύ των προτύπων x, y .

3.8 Σύγκριση απλής και πλήρους σύνδεσης

Ο αλγόριθμος απλής σύνδεσης και ο αλγόριθμος πλήρους σύνδεσης διαφέρουν μεταξύ τους στον τρόπο καθορισμού της ομοιότητας των προτύπων, τα οποία ανήκουν σε διαφορετικές κατηγορίες, για την συνένωσή τους. Είναι φυσικό οι δύο αλγόριθμοι να έχουν σαν αποτέλεσμα διαφορετικές ομαδοποιήσεις για τα ίδια δεδομένα. Αρα ποια μέθοδος θα πρέπει να χρησιμοποιηθεί; Δυστυχώς δεν υπάρχουν ξεκάθαρες οδηγίες για ένα χρήστη. Αρκετές προσπάθειες έχουν πραγματοποιηθεί για να αιτιολογήσουν την χρήση του αλγόριθμου απλής σύνδεσης σε μαθηματική βάση αλλά ο αλγόριθμος μειονεκτεί από το φαινόμενο της αλυσίδας όπου απομακρυσμένα πρότυπα τοποθετούνται στην ίδια ομάδα επειδή έχουν ένα κοινό γειτονικό πρότυπο. Περιληπτικά, ο αλγόριθμος απλής σύνδεσης έχει καλά αποτελέσματα σε περιπτώσεις όπου πραγματικές ομάδες είναι επιμήκεις στο χώρο. Αντίθετα ο αλγόριθμος πλήρους σύνδεσης έχει καλά αποτελέσματα όταν οι ομάδες είναι συμπαγείς και παρόμοιες σε μέγεθος.

Επιπρόσθετοι ιεραρχικοί αλγόριθμοι ομαδοποίησης μπορούν να χρησιμοποιήσουν άλλες συναρτήσεις όπως μέσες τιμές και το κέντρο βάρους (centroid). Ο αλγόριθμος που χρησιμοποιεί συναρτήσεις μέσω τιμών ονομάζεται σύνδεσης μέσω τιμών (average linkage) όπως ο αλγόριθμος που χρησιμοποιεί συναρτήσεις κέντρων βάρους ονομάζεται σύνδεση κέντρων βάρους (centroid linkage).

3.9 Αλγόριθμος σύνδεσης μέσω τιμών

Ο αλγόριθμος απλής σύνδεσης δημιουργεί ομάδες επιμήκεις ενώ ο αλγόριθμος πλήρης σύνδεσης παράγει πιο συμπαγείς ομάδες. Ο αλγόριθμος σύνδεσης μέσω τιμών είναι μια προσπάθεια του συμβιβασμού μεταξύ των άκρων των αλγόριθμων απλής και πλήρους σύνδεσης. Ο αλγόριθμος σύνδεσης μέσω τιμών ονομάζεται επίσης αλγόριθμος μέσω τιμών χωρίς βάρη και είναι ένας από τους πιο δημοφιλείς αλγόριθμους ομαδοποίησης. Στον αλγόριθμο σύνδεσης μέσω τιμών η απόσταση μεταξύ δύο ομάδων υπολογίζεται από την μέση απόσταση των προτύπων κάθε διαφορετικής ομάδας.

Εάν C_i είναι μια ομάδα με n_i μέλη και C_j είναι μια ομάδα με n_j μέλη, τότε η απόσταση μεταξύ δύο ομάδων ορίζεται ως:

$$D_{AL}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{x \in C_i, y \in C_j} d(x, y),$$

ΚΕΦΑΛΑΙΟ 4

4.1 Διαχωριστική ομαδοποίηση

Οι τεχνικές διαχωριστικής ομαδοποίησης διαχωρίζουν ένα σύνολο N προτύπων και δημιουργούν k ομάδες, όπου $k \ll N$. Ο επιθυμητός αριθμός των ομάδων k συνήθως προκαθορίζεται από τον χρήστη. Αντίθετα με τις ιεραρχικές τεχνικές η διαχωριστικές τεχνικές επιτρέπουν τα πρότυπα να μετακινούνται από μια ομάδα σε άλλη. Με αυτόν τον τρόπο ένας κακός αρχικός διαχωρισμός μπορεί να διορθωθεί αργότερα. Ο τρόπος με τον οποίο γίνεται ο διαχωρισμός στις περισσότερες τεχνικές είναι με την μεγιστοποίηση κάποιας συναρτήσεως κριτηρίων. Όμως η εύρεση του βέλτιστου διαχωρισμού δεν είναι μαθηματικά εφικτή. Για παράδειγμα, υπάρχουν $1,93 \cdot 10^8$ διαφορετικοί διαχωρισμοί 19 προτύπων σε 3 ομάδες. Έτσι, οι περισσότερες διαχωριστικές τεχνικές χρησιμοποιούν ορισμένες ευριστικές μεθόδους ώστε να βρεθεί μια ικανοποιητική λύση. Το αποτέλεσμα όμως είναι η δημιουργία υποβελτιστοποιημένων διαχωρισμών. Μεταξύ όλων των κριτηρίων τα οποία χρησιμοποιούνται για την δημιουργία ενός διαχωρισμού, το κριτήριο του τετραγωνικού σφάλματος είναι το πιο δημοφιλές. Βέβαια, η χρησιμοποίηση αυτού του κριτηρίου έχει σαν αποτέλεσμα την δημιουργία ομάδων υπερελλειψοειδούς μορφής. Επίσης, στις τεχνικές διαχωριστικής ομαδοποίησης χρησιμοποιείται συνήθως η ευκλείδεια μετρική για τον υπολογισμό αποστάσεων.

Υπάρχουν τρεις ευρέως διαδεδομένες τεχνικές διαχωριστικής ομαδοποίησης: ο αλγόριθμος forgy, ο αλγόριθμος kmeans και ο αλγόριθμος isodata. Εμείς θα ασχοληθούμε με τον αλγόριθμο kmeans.

ΚΕΦΑΛΑΙΟ 5

MATLAB

5.1 Τί είναι το Matlab

Το MATLAB είναι ένα πρόγραμμα υπολογιστών για ανθρώπους που χρησιμοποιούν αριθμητικούς υπολογισμούς, ειδικά στη γραμμική άλγεβρα (πίνακες). Ξεκίνησε ως ένα πρόγραμμα "Εργαστηρίου Πινάκων" ("MATrixLABoratory") που είχε σκοπό να παρέχει αλληλεπιδρώσα προσπέλαση στις βιβλιοθήκες Linpack και Eispack. Από τότε έχει αναπτυχθεί αρκετά, για να γίνει ένα ισχυρότατο εργαλείο στην οπτικοποίηση, στον προγραμματισμό, στην έρευνα, στην επιστήμη των μηχανικών, και στις επικοινωνίες. Στο δυναμικό του Matlab συμπεριλαμβάνονται μοντέρνοι αλγόριθμοι, δυνατότητες χειρισμού τεράστιων ποσοτήτων δεδομένων, και ισχυρά προγραμματιστικά εργαλεία. Το Matlab δεν είναι σχεδιασμένο για συμβολικούς υπολογισμούς, αλλά αντισταθμίζει αυτή την αδυναμία του επιτρέποντας στο χρήστη να συνδέεται άμεσα με το Maple. Η επιφάνεια αλληλεπίδρασης βασίζεται κυρίως σε κείμενο, γεγονός που μπορεί να συγχύσει μερικούς χρήστες. Το Matlab έρχεται ως πακέτο του βασικού προγράμματος, με πολλές "εργαλειοθήκες", που πωλούνται ξεχωριστά.

Το λογισμικό MATLAB, που παίρνει το όνομά του από τις λέξεις MATrix LABoratory, είναι ένα σύγχρονο ολοκληρωμένο μαθηματικό πακέτο που χρησιμοποιείται εκτενώς στα πανεπιστήμια και στη βιομηχανία. Είναι ένα διαδραστικό (interactive) πρόγραμμα για αριθμητικούς υπολογισμούς και για κατασκευή γραφημάτων, αλλά παρέχει επίσης και τη δυνατότητα προγραμματισμού, κάτι που το καθιστά ένα χρησιμότερο εργαλείο για όλους όσους ασχολούνται με τις θετικές επιστήμες (και όχι μόνο). Σε αντίθεση με τα λογισμικά MAPLE και MATHEMATICA, το MATLAB στις αρχικές του εκδοχές δεν έκανε συμβολικούς υπολογισμούς. Στις νεότερες εκδοχές του, το πακέτο περιλαμβάνει εργαλειοθήκες που επιτρέπουν συμβολικούς υπολογισμούς. Όπως υποδηλώνεται και από το όνομα του, το MATLAB είναι ειδικά σχεδιασμένο για υπολογισμούς με πίνακες, όπως η επίλυση γραμμικών συστημάτων, η εύρεση ιδιοτιμών και ιδιοδιανυσμάτων, η αντιστροφή τετραγωνικού πίνακα κλπ. Επιπλέον το πακέτο αυτό είναι εφοδιασμένο με πολλές επιλογές για γραφικά (δηλ. την κατασκευή γραφικών παραστάσεων) και προγράμματα γραμμένα στη δική του γλώσσα προγραμματισμού για την επίλυση άλλων προβλημάτων όπως η εύρεση των ριζών μη γραμμικής εξίσωσης, η επίλυση μη γραμμικών συστημάτων, η επίλυση προβλημάτων αρχικών τιμών με συνήθεις διαφορικές εξισώσεις κ.α. Η γλώσσα προγραμματισμού του MATLAB δίνει την ευχέρεια στον χρήστη να το επεκτείνει με δικά του προγράμματα. Συχνά θα λέμε η MATLAB (εννοώντας τη γλώσσα προγραμματισμού) και όχι το (πακέτο) MATLAB. Το MATLAB είναι σχεδιασμένο για την αριθμητική επίλυση προβλημάτων σε *αριθμητική πεπερασμένης ακρίβειας* (finite-precision arithmetic). Με άλλα λόγια, δεν βρίσκει την ακριβή λύση αλλά μια προσεγγιστική λύση ενός προβλήματος. Αυτή είναι και η βασική του διαφορά από τα συστήματα συμβολικών υπολογισμών όπως το Maple και το Mathematica. Ας σημειωθεί ότι ο καλύτερος (και ουσιαστικά ο μόνος) τρόπος εκμάθησης της MATLAB είναι η συστηματική ενασχόληση με αυτή και η διερεύνησή της από τον ίδιο τον χρήστη. Το πακέτο είναι εφοδιασμένο με ένα εκτενές σύστημα βοήθειας όπου κάθε εντολή επεξηγείται αναλυτικά και με αντιπροσωπευτικά παραδείγματα. Η πιο σημαντική εντολή της MATLAB είναι η **help** (βοήθεια)!

5.2 Πώς δουλεύει το Matlab

Το Matlab δουλεύει εκτελώντας τις μαθηματικές εντολές που εισάγετε στο παράθυρο εντολών. Η προκαθορισμένη επιλογή είναι, κάθε έξοδος να τυπώνεται απευθείας στο παράθυρο. Ακόμα, σας επιτρέπεται να εκχωρείτε ένα όνομα σε μία έκφραση για δική σας ευκολία. Να θυμάστε ότι το όνομα που εκχωρείτε είναι μόνο ένα όνομα, και δεν αναπαριστά καμία μαθηματική μεταβλητή (όπως θα έκανε στο Maple, για παράδειγμα). Κάθε όνομα πρέπει να έχει μία τιμή κάθε στιγμή. Αν προσπαθήσετε να διαβάσετε την τιμή ενός μη προσδιορισμένου ονόματος, θα πάρετε μήνυμα λάθους. Σχεδόν τα πάντα στο Matlab είναι πίνακες, είτε μοιάζουν με τέτοιους είτε όχι. Αυτό μπορεί να σας πάρει λίγο μέχρι να το συνηθίσετε. Θα εισάγουμε τις λειτουργίες με το στυλ πίνακα μαζί με τα αριθμητικά τους ισοδύναμα, ώστε να μπορείτε να καταλάβετε τα καλούπια που παρουσιάζονται στο συντακτικό.

5.3 Εισαγωγή

Στην Αναγνώριση προτύπων, αλλά και στα Νευρωνικά Δίκτυα έχουμε να κάνουμε με σύνολα δεδομένων. Τα σύνολα αυτά μπορεί να έχουν μετρήσεις για χιλιάδες πρότυπα και κάθε πρότυπο να περιγράφεται από μερικές δεκάδες χαρακτηριστικές τιμές. Η συλλογή, επεξεργασία και τεκμηρίωση τους, αποτελεί το πιο μεγάλο και το πιο δύσκολο μέρος του προβλήματος. Στο κεφάλαιο αυτό θα γίνει μια μικρή εισαγωγή πάνω στην χρήση του Matlab για μια βασική επεξεργασία των δεδομένων.

5.4 Προ-επεξεργασία δεδομένων

Πριν ξεκινήσει η διαδικασία της κατασκευής ενός ταξινομητή ή η διαδικασία της εκπαίδευσης ενός Νευρωνικού Δικτύου, θα πρέπει τα σύνολα αυτά να αναλυθούν για να αποκτηθεί μια αίσθηση των στοιχείων – πληροφοριών που περιέχουν, καθώς και να πραγματοποιηθεί μια σειρά από εργασίες πάνω τους. Τυπικές προ-επεξεργασίες μπορεί να είναι:

1. Η κατάλληλη κωδικοποίηση των χαρακτηριστικών γνωρισμάτων σε χαρακτηριστικές τιμές. Για παράδειγμα αν τα χαρακτηριστικά γνωρίσματα περιέχουν κείμενο ή χρώμα θα πρέπει αυτό να κωδικοποιηθεί με αριθμητικές τιμές, αν ένα χαρακτηριστικό γνώρισμα εκφράζει μια περιοδικότητα να εκφραστεί κατάλληλα κ.λ.π.
2. Να επιλεγούν οι χαρακτηριστικές τιμές που έχουν την απαιτούμενη πληροφορία και θα χρησιμοποιηθούν για την αναγνώριση καθώς και οι χαρακτηριστικές τιμές που θα αγνοηθούν.
3. Να μετασχηματιστούν οι τιμές μέσα από μια μαθηματική συνάρτηση και/ή να κατασκευαστούν νέες από τις υπάρχουσες. Το κομμάτι αυτό εξαρτάται πολύ από την εφαρμογή. Αν το πρόβλημα έχει να κάνει με εικόνες θα χρησιμοποιηθούν μέθοδοι από την επιστήμη της Επεξεργασίας και Ανάλυσης εικόνας, αν έχει να κάνει με χρονοσειρές μέθοδοι DSP (για βιοσειρές, σήματα ραντάρ κ.λ.π) ή οικονομικές (για οικονομικά μεγέθη) κ.λ.π.
4. Μια ειδική περίπτωση της παραπάνω είναι να περιοριστεί το εύρος των τιμών σε ένα κλειστό διάστημα (π.χ. στο [-1,+1] ή στο [0+1]) ή να αποκτήσουν κοινά στατιστικά χαρακτηριστικά (π.χ. όλες οι χαρακτηριστικές τιμές να έχουν μέσο όρο 0 και τυπική διασπορά 1).
5. Να αντιμετωπιστούν διάφορες ασάφειες που πιθανά να υπάρχουν στα δεδομένα. Μπορεί να υπάρχουν εμφανώς λανθασμένες μετρήσεις ή κάποιες μετρήσεις να μην υπάρχουν καθόλου για κάποια πρότυπα. Για να γίνουν τα παραπάνω θα πρέπει να γίνει κάποια αρχική μελέτη στα δεδομένα. Στην μελέτη αυτή θα πρέπει να απαντηθούν οι παρακάτω *βασικές ερωτήσεις*:
 - a. **Πόσα είναι τα δεδομένα:** Όσο περισσότερα δεδομένα υπάρχουν τόσο το καλύτερο. Θα πρέπει να υπάρχουν αρκετά δεδομένα για κάθε διαφορετική κατηγορία του προβλήματος. Αν δεν ισχύει αυτό η μόνη ικανοποιητική λύση είναι να συλλεχθούν νέα. Συχνά τα δεδομένα είναι είτε ακριβό είτε αδύνατο να συλλεχθούν. Σε αυτή την περίπτωση είτε παράγουμε τεχνητά (που έχουν τα ίδια στατιστικά χαρακτηριστικά με τα αρχικά), είτε επαναχρησιμοποιούμε τα δεδομένα χρησιμοποιώντας διαφορετικούς τρόπους χωρισμού των αρχικών σε υποσύνολα (bootstrapping, leave-one-out τεχνικές).
 - b. **Πώς είναι κωδικοποιημένα και τι εκφράζουν τα δεδομένα:** Οι χαρακτηριστικές τιμές μπορεί να είναι:
 - **Αριθμητικές:** (ordinal data) Μετρούν το μέγεθος ενός προτύπου π.χ. μήκος, συγκέντρωση, πλήθος. Κάποιες αριθμητικές τιμές έχουν μια μονάδα μέτρησης ενώ κάποιες άλλες εκφράζουν λόγους αριθμητικών τιμών και είναι χωρίς μονάδες. Οι τελευταίες συχνά είναι ήδη σε ένα κλειστό διάστημα είναι προτιμότερες και συχνά φέρουν περισσότερη πληροφορία.
 - **Περιγραφικές:** Δείχνουν κατηγορία, δηλαδή αν το στοιχείο ανήκει σε ένα σύνολο. π.χ. χρώμα (0=μαύρο, 1=άσπρο, 2=πράσινο), εισόδημα (0=μικρό, 1=μέτριο, 2=κανονικό, 3=υψηλό). Μια σημαντική παρατήρηση είναι ότι στην πρώτη περίπτωση η αριθμητική αντιστοίχιση είναι εντελώς τυχαία, ενώ στην δεύτερη έχει κάποιο νόημα. Μια καλύτερη λύση για την πρώτη 4 περίπτωση, είναι να χρησιμοποιηθεί κάποιος άλλος τρόπος κωδικοποίησης (π.χ. 101 = άσπρο, 010= μαύρο, ...).
 - **Περιοδικές:** Έχουν μια περιοδική φύση. π.χ. μέρα, μήνας. Πάλι απαιτείται μια προσοχή στην κωδικοποίησή τους.
 - c. **Τι στατιστικά χαρακτηριστικά έχουν και πως σχετίζονται στατιστικά:** Πώς «μοιάζουν» τα δεδομένα; Ακολουθούν κάποια στατιστική κατανομή, ποια είναι και τι

χαρακτηριστικά έχει; Υπάρχει κάποια στατιστική εξάρτηση μεταξύ των μεταβλητών; Η περιγραφική στατιστική μπορεί να συνοψίσει τις μετρήσεις με λίγους ευκολονόητους αριθμούς (μέσους όρους, διασπορές κλπ.). Τα στατιστικά διαγράμματα βοηθούν στην οπτικοποίηση των πληροφοριών που μας παρέχουν τα δεδομένα. Άλλες στατιστικές αναλύσεις δίδουν περισσότερα αποτελέσματα.

- d. **Πώς είναι κωδικοποιημένες οι κατηγορίες:** Για προβλήματα ταξινόμησης. Τι ποσοστό από κάθε κατηγορία υπάρχει στα δεδομένα; Αντιστοιχεί με αυτό που υπάρχει στην πραγματικότητα; Οι πιθανότητες εμφάνισης των διαφόρων κατηγοριών δεν είναι ίδιες. Σε μια βιομηχανική εφαρμογή τα ελαττωματικά προϊόντα είναι λιγότερα από τα μη ελαττωματικά. Για να μπορεί ένας ταξινομητής να μάθει τις διαφορές μεταξύ των κατηγοριών, θα πρέπει να υπάρχει αρκετή πληροφορία για όλες τις κατηγορίες. Για τον λόγο αυτό θα πρέπει να χρησιμοποιηθούν σύνολα με περίπου ίσο πλήθος στοιχείων από κάθε κατηγορία. Η γνώση των πραγματικών ποσοστών όμως συχνά είναι ζωτική και θα πρέπει να εκτιμάται

5.5 Ομαδοποίηση

Η ομαδοποίηση είναι ένας τρόπος να εξεταστούν οι ομοιότητες και οι ανομοιότητες μεταξύ προτύπων(δεδομένων). Τα διάφορα πρότυπα συχνά σχηματίζουν ομάδες, όπου τα χαρακτηριστικά των αντικειμένων(προτύπων) στην ίδια ομάδα είναι παρόμοια και τα χαρακτηριστικά των αντικειμένων που ανήκουν σε διαφορετικές ομάδες είναι ανόμοια. Θα εξετάσουμε 2 αλγόριθμους ομαδοποίησης: τον αλγόριθμο ομαδοποίησης kmeans και την ιεραρχική ομαδοποίηση.

5.6 Αλγόριθμος ομαδοποίησης kmeans

Ο αλγόριθμος ομαδοποίησης kmeans θεωρεί ότι τα πρότυπα έχουν κάποιες θέσεις και κάποιες αποστάσεις μεταξύ τους. Τα διαχωρίζει λοιπόν σε K καθορισμένες από πριν ομάδες, έτσι ώστε τα πρότυπα που ανήκουν σε μια ομάδα να είναι όσο το δυνατόν πιο κοντά μεταξύ τους κι όσο το δυνατόν πιο μακριά από πρότυπα που ανήκουν σε άλλες ομάδες. Κάθε ομάδα που προκύπτει αντιπροσωπεύεται από ένα κεντρικό σημείο (κέντρο). Ο αλγόριθμος kmeans δουλεύει ως εξής: αρχικά επιλέγουμε τον αριθμό των ομάδων που θέλουμε να χωρίσουμε τα πρότυπα μας, έστω K ομάδες. Επιλέγουμε τυχαία K πρότυπα τα οποία χρησιμοποιούνται ως αρχικά κέντρα (αντιπρόσωποι των K ομάδων). Υπολογίζουμε τις αποστάσεις των υπόλοιπων προτύπων από τα K πρότυπα-κέντρα που επιλέξαμε(με συναρτήσεις απόστασης). Το κάθε πρότυπο ενσωματώνεται στην ομάδα της οποίας το κέντρο είναι πιο κοντά του. Όταν τελειώσει η ομαδοποίηση εκλέγονται νέα κέντρα σε κάθε ομάδα(με διάφορες μεθόδους). Υπολογίζονται εκ νέου οι αποστάσεις των προτύπων από τα νέα κέντρα, σ αυτό το σημείο πολλά πρότυπα αλλάζουν ομάδα διότι κάποιο νέο κέντρο μπορεί να βρίσκεται πιο κοντά του. Όταν τελειώσει και αυτή η ομαδοποίηση ξανά εκλέγονται νέα κέντρα κι αυτό επαναλαμβάνεται μέχρι να μην έχουμε καμία μετακίνηση προτύπου σε διαφορετική ομάδα από αυτήν που ανήκει κι έτσι ο αλγόριθμος καταλήγει στις τελικές ομάδες.

Η συνάρτηση kmeans παίρνει σαν είσοδο τις n μετρήσεις ενός πίνακα x (μια γραμμή ανά πρότυπο) και επιστρέφει έναν πίνακα (nx1) IDX που περιέχει τον αριθμό της ομάδας που ανήκει κάθε πρότυπο. Η συνάρτηση kmeans υλοποιεί τον αλγόριθμο. Η γενική μορφή της είναι:

$$kmeans(X, k)$$

$$[IDX, C] = kmeans(X, k)$$

$$[IDX, C, sumD, D] = kmeans(X, k)$$

$$[...] = kmeans(..., 'param 1', val 1, ...)$$

Η συμπεριφορά του αλγόριθμου μπορεί να τροποποιηθεί αλλάζοντας κάποια παράμετρο. Οι πιο σημαντικές παράμετροι δίδονται παρακάτω:

- 'distance': Καθορίζει την συνάρτηση απόστασης που θα χρησιμοποιηθεί. Πιθανές τιμές: 'sqEuclidean', 'cityblock', 'cosine', 'correlation', 'Hamming'.
- 'start': Καθορίζει τον τρόπο επιλογής των αρχικών θέσεων των κέντρων. Τα κέντρα τοποθετούνται κάπου στην αρχή και σε κάθε βήμα του αλγόριθμου αλλάζουν θέση μέχρις ότου τερματιστεί ο αλγόριθμος. Πιθανές τιμές: 'sample' (τυχαίες θέσεις), 'uniform' (ομοιόμορφα δεν μπορεί να δοθεί για απόσταση hamming), 'cluster' (εκτελεί

μια γρήγορη ομαδοποίηση με το 10% των στοιχείων και sample για να θέσει τα κέντρα), είτε ένας πίνακας με τα αρχικά κέντρα.

- 'replicates': Πόσες φορές θα εκτελεστεί ο αλγόριθμος με διαφορετικά αρχικά κέντρα. Επιλέγεται η ομαδοποίηση με το μικρότερο sumd.
- 'maxiter': Μέγιστο πλήθος βημάτων. Εξ ορισμού 100

Παράδειγμα:

```
[a,b,c,d]=kmeans(meas,3,'distance','cosine','start','cluster','replicates',15,'maxiter',1500)
```

Για να αξιολογήσουμε την ομαδοποίηση του kmeans έχουμε 2 τρόπους.

- 1) Ο πρώτος είναι οπτικός και αφορά τα διαγράμματα silhouette (όσο πιο κοντά στο 1 τόσο καλύτερη ταξινόμηση έχω) και
- 2) ο δεύτερος είναι μετρίσιμος και αφορά τον τρόπο που εργάζεται ο Kmeans. Συγκεκριμένα ο αλγόριθμος φτιάχνει τις ομάδες του με τέτοιο τρόπο ώστε να ελαχιστοποιεί το άθροισμα των αποστάσεων μεταξύ των ομάδων που κατασκευάζει.

Στις πολλές δοκιμές που θα κάνουμε λοιπόν με τον Kmeans θα διαλέξουμε την ομαδοποίηση που παρουσιάζει το μικρότερο άθροισμα. Το άθροισμα αυτό εμφανίζεται αν στον kmeans βάλουμε ως παράμετρο 'display' την τιμή 'iter':

- iter: ο αριθμός της επανάληψης που βρίσκεται ο αλγόριθμος για την συγκεκριμένη επιλογή αρχικών κέντρων.
- Phase: ο αλγόριθμος όπως είναι δομημένος αποτελείται από 2 φάσεις.
- Num: ο αριθμός των προτύπων που αλλάζουν ομάδα στο συγκεκριμένο βήμα του αλγορίθμου. Όταν κανένα πρότυπο δεν αλλάζει ομάδα ο αλγόριθμος τερματίζεται.
- Sum: το συνολικό άθροισμα των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους είναι το νούμερο που μας ενδιαφέρει να ελαχιστοποιήσουμε.

Τοπικά ελάχιστα: Ο αλγόριθμος των K-μέσων μπορεί να φτάσει σε ένα τοπικό ελάχιστο – δηλαδή σε ένα σημείο όπου δεν θα μπορεί να ελαχιστοποιηθεί περαιτέρω το άθροισμα των αποστάσεων των προτύπων από το κέντρο της ομάδας τους. Τα αρχικά σημεία εκκίνησης των κέντρων είναι τυχαία. Είναι δυνατόν να μετακινηθεί ένα κέντρο σε ένα σημείο που και η πιο μικρή μετακίνηση του να αυξήσει το άθροισμα των αποστάσεων των σημείων από το κέντρο, αλλά παρόλα αυτά μια καλύτερη λύση υπάρχει. Δεν είναι εύκολο δε να διαπιστωθεί αν ένα σημείο είναι σημείο τοπικού ελαχίστου. Μόνος τρόπος να υπερνικηθεί το πρόβλημα είναι να ξεκινήσει ο αλγόριθμος από την αρχή με νέα τυχαία σημεία αρχής των κέντρων και να διαπιστωθεί εάν προκύπτει καλύτερη λύση. Η παράμετρος 'replicates' θέτει το πόσες φορές θα εκτελεστεί ο αλγόριθμος. Η παράμετρος 'display' με πιθανές τιμές ['off' | {'notify'} | 'final' | 'iter'] καθορίζει τι θα τυπωθεί σε κάθε replicate και σε κάθε βήμα του αλγορίθμου. Θα ζητήσουμε πέντε εκτελέσεις και αποτελέσματα στο τέλος κάθε εκτέλεσης.

Θα επιστρέψουμε στην υλοποίηση του αλγορίθμου kmeans μετά την ιεραρχική ομαδοποίηση για λόγους που αναφέρονται παρακάτω.

5.7 Ιεραρχική ομαδοποίηση

Ο αλγόριθμος των k-μέσων παράγει ένα ενιαίο χωρισμό των προτύπων σε ομάδες, αλλά υπάρχει το μειονέκτημα ότι θα πρέπει να γνωρίζουμε τον πραγματικό αριθμό των ομάδων που χωρίζονται τα πρότυπα. Ένα άλλο μειονέκτημα αποτελεί το γεγονός ότι αν δηλώσουμε στον αλγόριθμο του k-means ότι θέλουμε να χωρίσουμε τα πρότυπα με βάση έναν αριθμό ομάδων που διαφέρει από τον πραγματικό και πάλι ο αλγόριθμος θα προχωρήσει στην ομαδοποίηση σύμφωνα με τα δεδομένα που του έχουμε δώσει. Η δεύτερη μέθοδος που θα χρησιμοποιήσουμε για να ομαδοποιήσουμε τα δεδομένα μας είναι η ιεραρχική ομαδοποίηση. Ο αλγόριθμος αυτός παράγει μια ιεραρχία ομάδων, όπου οι ομάδες σ' ένα επίπεδο ενώνονται σε νέες στο επόμενο υψηλότερο επίπεδο. Αυτό μας επιτρέπει να αποφασίσουμε ποιο επίπεδο συγκέντρωσης είναι το καταλληλότερο για τα δεδομένα σε μια εφαρμογή.

Συναρτήσεις ιεραρχικής ομαδοποίησης:

Y=pdist(x,'metric') όπου x είναι τα δεδομένα μου και metric είναι οι μετρικές αποστάσεις ('euclidean', 'mahalanobis', 'cityblock', 'cosine', 'hamming'). Η συνάρτηση pdist υπολογίζει την απόσταση του κάθε προτύπου απ' όλα τα υπόλοιπα (εφαρμοζόμενο για κάθε πρότυπο).

S=squareform(Y) μετασχηματίζει την έξοδο της παραπάνω εντολής σε μορφή πίνακα.

Z=linkage(Y,'method') όπου method είναι οι μέθοδοι της ιεραρχικής ομαδοποίησης ('single', 'complete', 'average', 'centroid', 'ward'). Η συνάρτηση linkage κάνει συνένωση των προτύπων σε ομάδες με την ακόλουθη λογική: αρχικά συνδέονται τα 2 πρότυπα που απέχουν λιγότερο δημιουργώντας την πρώτη ομάδα η οποία λαμβάνει κωδικό τον αμέσως επόμενο από αυτόν που έχει το τελευταίο πρότυπο (πχ στο δείγμα b3344 το τελευταίο πρότυπο έχει κωδικό 27 άρα η πρώτη ομάδα έχει κωδικό 28).

Το χαρακτηριστικό της ομάδας που δημιουργήθηκε και μας ενδιαφέρει άμεσα (πέρα από το ποια πρότυπα την αποτελούν) είναι το μέτρο της απόστασής της από τα υπόλοιπα πρότυπα, καθώς αυτό θα είναι κριτήριο για την συνένωση της με κάποιο άλλο πρότυπο ή κάποια άλλη ομάδα. Η επόμενη συνένωση δημιουργείται μεταξύ των προτύπων ή και της ομάδας με μοναδικό κριτήριο την μικρότερη απόσταση. Με τον τρόπο αυτόν δημιουργούνται οι ομαδοποιήσεις με τελικό στάδιο την δημιουργία μιας ομάδας η οποία θα περιέχει όλα τα πρότυπα.

P=cophenet(Z,Y) ένας τρόπος να εκτιμηθεί η ποιότητα της ομαδοποίησης είναι να συγκριθεί το αποτέλεσμα της συνάρτησης linkage με τα αρχικά στοιχεία του πίνακα εγγύτητας που παράγει η συνάρτηση pdist. Αυτό γίνεται με τη συνάρτηση cophenet. Όσο πιο κοντά στο 1 είναι η τιμή της cophenet, τόσο καλύτερη η ομαδοποίηση.

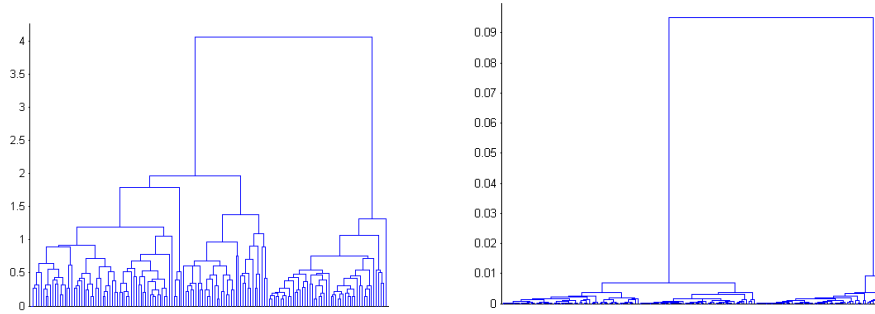
H=dendrogram(Z) οπτική αναπαράσταση των δεδομένων με δενδρογράμματα

Ένας τρόπος να καθοριστούν οι κύριες ομάδες ενός δέντρου ομαδοποίησης είναι να συγκριθεί το ύψος κάθε σύνδεσης ως προς τον κατακόρυφο άξονα, με τα ύψη των γειτονικών συνδέσεων κάτω από αυτήν. Εάν μια σύνδεση έχει περίπου το ίδιο ύψος με τις γειτονικές συνδέσεις της αυτό δείχνει ότι υπάρχουν ομοιότητες μεταξύ των αντικειμένων που ενώνονται σε αυτό το επίπεδο της ιεραρχίας και ο σύνδεσμος καλείται συνεπής. Εάν το ύψος μιας σύνδεσης διαφέρει κατά πολύ από το ύψος των γειτονικών της συνδέσεων, αυτό σημαίνει πως υπάρχουν μεγάλες διαφορές μεταξύ των ομάδων σε αυτό το επίπεδο και ο σύνδεσμος καλείται ασυνεπής. Ένας ασυνεπής σύνδεσμος υποδηλώνει πως υπάρχει ένα φυσικό σύνορο μεταξύ των ομάδων σε αυτό το σημείο και είναι ένα κατάλληλο μέρος για να διαχωριστούν οι ομάδες.

I=inconsistent(Z) η συνάρτηση inconsistent συγκρίνει κάθε σύνδεσμο με τους συνδέσμους 2 επίπεδα κάτω στην ιεραρχία. Οι σύνδεσμοι στα κατώτερα επίπεδα παίρνουν τιμή 0. Εάν το αντικείμενο είναι συνεπές μ εκείνους γύρω από αυτό θα έχει έναν χαμηλό συντελεστή ασυνέπειας. Εάν είναι ασυνεπές θα έχει έναν υψηλό συντελεστή ασυνέπειας. Η συνάρτηση αυτή δημιουργεί έναν πίνακα I, κάθε γραμμή του οποίου περιέχει στοιχεία για κάθε σύνδεση. Η πρώτη στήλη περιέχει το μέσο όρο των αποστάσεων, η δεύτερη στήλη περιέχει την τυπική απόκλιση, η τρίτη στήλη περιέχει τον αριθμό των συνδέσεων που παίρνουν μέρος στον υπολογισμό και η τέταρτη στήλη περιέχει τον συντελεστή ασυνέπειας.

T=cluster(Z,0.8) η συνάρτηση cluster χωρίζει το δενδρόγραμμα σε ομάδες. Οι ομάδες 'κόβονται' εκεί που συντελεστής ασυνέπειας παίρνει τιμή 0,8.

Πρώτα υπολογίζουμε όλες τις αποστάσεις μεταξύ των προτύπων (αποστάσεις ενός προτύπου από όλα τα υπόλοιπα, εφαρμοζόμενο για κάθε πρότυπο) χρησιμοποιώντας την Ευκλείδεια απόσταση. Για την μέθοδο υπολογισμού των αποστάσεων μεταξύ των ομάδων θα χρησιμοποιηθεί η παράμετρος 'average'. Η pdist() επιστρέφει τις αποστάσεις σε ένα διάνυσμα που περιέχει μόνο το άνω τριγωνικό μέρος του πίνακα αποστάσεων (το κάτω μέρος είναι ακριβώς το ίδιο). Για να δούμε τον πίνακα αποστάσεων χρησιμοποιούμε την συνάρτηση squareform(). Η συνάρτηση αυτή μόνο οπτικό πλεονέκτημα μας δίνει και δεν θα χρησιμοποιηθεί για οποιονδήποτε παρακάτω υπολογισμό. Με τη συνάρτηση linkage() γίνονται οι συνδέσεις σε ομάδες μεταξύ των προτύπων. Μια οπτική αναπαράσταση της διαδικασίας που περιγράψαμε στην συνάρτηση linkage μπορούμε να έχουμε με την χρήση της συνάρτησης dendrogram()

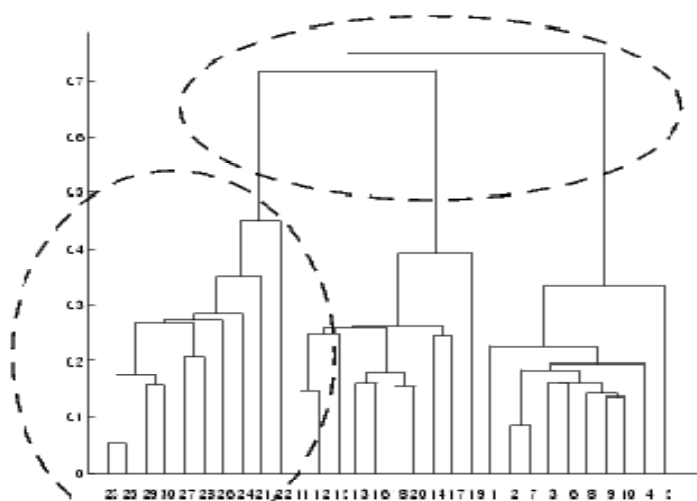


Σχήμα 5.7.1 : Ιεραρχική ταξινόμηση με average linkage α) Με Ευκλείδεια απόσταση, β) Με απόσταση συνημιτόνου.

Η συνάρτηση dendrogram() λαμβάνει ως παραμέτρους τον πίνακα Z το οποίο είναι απλά το όνομα του πίνακα που έχει παράγει η linkage(), και έναν αριθμό, ο οποίος όταν είναι 0 παρουσιάζει το σύνολο των συνδέσεων ενώ αν δεν βάλουμε καμία τιμή, μας παρουσιάζει τις 30 συνδέσεις που γίνονται στο ανώτερο επίπεδο. Ένας τρόπος για να εκτιμηθεί η ποιότητα της ομαδοποίησης, είναι να συγκριθεί το αποτέλεσμα των συνδέσεων της linkage() με τα αρχικά στοιχεία του πίνακα εγγύτητας που παράγει η συνάρτηση pdist(). Εάν η ομαδοποίηση είναι καλή, η απόσταση σύνδεσης των αντικειμένων στο ιεραρχικό δέντρο πρέπει να έχει ισχυρή σχέση με τις αποστάσεις μεταξύ των αντικειμένων στον πίνακα εγγύτητας. Η συνάρτηση cophenet() συγκρίνει αυτά τα δύο σύνολα τιμών και υπολογίζει ένα συντελεστή συσχετισμού τους. Όσο πιο κοντά στο 1 είναι η τιμή του συντελεστή σχέσης P, τόσο καλύτερη η ομαδοποίηση.

Χωρισμός σε ομάδες: Ένας τρόπος να καθοριστούν οι «κύριες ομάδες» ενός δέντρου ομαδοποίησης είναι να συγκριθεί το ύψος κάθε σύνδεσης ως προς τον κατακόρυφο άξονα, με τα ύψη των γειτονικών συνδέσεων κάτω από αυτήν. Εάν μια σύνδεση έχει περίπου το ίδιο ύψος με τις γειτονικές συνδέσεις της, αυτό δείχνει ότι υπάρχουν ομοιότητες μεταξύ των αντικειμένων που ενώνονται σε αυτό το επίπεδο της ιεραρχίας και ο σύνδεσμος καλείται συνεπής. Εάν το ύψος μιας σύνδεσης διαφέρει κατά πολύ από το ύψος των γειτονικών της συνδέσεων, αυτό σημαίνει πως υπάρχουν μεγάλες διαφορές μεταξύ των ομάδων σε αυτό το επίπεδο και ο σύνδεσμος καλείται ασυνεπής. Ένας ασυνεπής σύνδεσμος υποδηλώνει πως υπάρχει ένα φυσικό σύνορο μεταξύ των ομάδων σε αυτό το σημείο και είναι ένα κατάλληλο μέρος για να διαχωριστούν οι ομάδες.

These links show inconsistency when compared to links below them



These links show consistency

Σχήμα 5.7.2: Συνεπείς σύνδεσμοι και φυσικός χωρισμός σε ομάδες

Η σχετική συνέπεια κάθε σύνδεσης σε ένα ιεραρχικό δέντρο ομαδοποίησης εκφράζεται με τον συντελεστή ασυνέπειας. Ο συντελεστής συγκρίνει το ύψος μιας σύνδεσης με το ύψος των συνδέσεων που βρίσκονται ακριβώς από κάτω, σε όποιο βάθος εμείς επιθυμούμε (η προκαθορισμένη τιμή είναι δύο επίπεδα). Εάν το αντικείμενο είναι συνεπές με εκείνους γύρω από αυτό, θα έχει έναν χαμηλό συντελεστή ασυνέπειας, ενώ αν είναι ασυνεπές, θα έχει έναν υψηλότερο συντελεστή ασυνέπειας. Η συνάρτηση `inconsistent()` συγκρίνει κάθε σύνδεσμο με τους συνδέσμους δύο επίπεδα κάτω στην ιεραρχία (το βάθος σύγκρισης μπορεί να αλλάξει). Οι σύνδεσμοι στα κατώτερα επίπεδα παίρνουν τιμή 0 (συνδέσεις μεταξύ προτύπων). Κάθε γραμμή του πίνακα I περιέχει στοιχεία για μια σύνδεση. Η πρώτη στήλη περιέχει το μέσο όρο των αποστάσεων, η δεύτερη την τυπική απόκλιση, η Τρίτη τον αριθμό των συνδέσεων που παίρνουν μέρος στον υπολογισμό και η τέταρτη τον συντελεστή ασυνέπειας για την σύνδεση. Η συνάρτηση `cluster()` μπορεί να χρησιμοποιήσει αυτήν την πληροφορία για να χωρίσει το δενδρόγραμμα σε ομάδες. Για να κοπουν ομάδες εκεί που ο συντελεστής παίρνει τιμή 0.8 Αν θέλουμε να προκαθορίσουμε το πλήθος των ομάδων θα δώσουμε στην συνάρτηση `cluster()` μια ακέραια τιμή μεγαλύτερη του 1 (άρα δεν θα μπορεί να είναι συντελεστής ασυνέπειας).

ΚΕΦΑΛΑΙΟ 6

ΥΛΟΠΟΙΗΣΗ

6.1 Δεδομένα που μας δόθηκαν

- 1) Οι 27 συχνότερα επαναλαμβανόμενες συλλαβές του δείγματος γραμμικής Β της Κνωσού με 3344 συλλαβές.
- 2) Οι 27 συχνότερα επαναλαμβανόμενες συλλαβές του δείγματος γραμμικής Β της Κνωσού με 11725 συλλαβές.
- 3) Οι 27 συχνότερα επαναλαμβανόμενες συλλαβές του δείγματος της γραμμικής Β της Πύλου με 8460 συλλαβές
- 4) Οι 24 συχνότερα επαναλαμβανόμενες συλλαβές του δείγματος της γραμμικής Α με 1859 συλλαβές

Οι 27 συχνότερα επαναλαμβανόμενες συλλαβές του δείγματος γραμμικής Β της Κνωσού με 3344 συλλαβές.

A/A	Συλλαβές	Αρχή σχετική επί τις εκατό συχνότητα	Μέση σχετική επί τις εκατό συχνότητα	Τέλος σχετική επί τις εκατό συχνότητα	Συνολική σχετική επί τις εκατό συχνότητα
1	ro	0,001664	0,039239	0,153396	0,071864
2	jo	0,001664	0,015458	0,174473	0,070993
3	ta	0,043261	0,058264	0,101874	0,070557
4	a	0,241265	0,002378	0,003513	0,065331
5	to	0,019967	0,03805	0,105386	0,058362
6	ra	0,031614	0,082045	0,022248	0,046603
7	u	0,024958	0,03805	0,059719	0,042683
8	ka	0,063228	0,052319	0,010539	0,039634
9	no	0,013311	0,01308	0,076112	0,036585
10	wo	0,021631	0,019025	0,063232	0,03615
11	ri	0,018303	0,078478	0,005855	0,035714
12	ko	0,046589	0,027348	0,031616	0,033972
13	ja	0,018303	0,046373	0,026932	0,031794
14	wa	0,02995	0,048751	0,016393	0,031794
15	da	0,048253	0,041617	0,009368	0,031359
16	na	0,016639	0,057075	0,01171	0,029617
17	re	0,019967	0,049941	0,01171	0,027875
18	so	0	0,007134	0,066745	0,027439
19	te	0,033278	0,040428	0,010539	0,027439
20	ke	0,02995	0,041617	0,008197	0,026132
21	si	0,031614	0,039239	0,005855	0,024826
22	ma	0,036606	0,033294	0,005855	0,023955
23	ti	0,018303	0,039239	0,009368	0,022648
24	e	0,073211	0,007134	0,001171	0,022213
25	sa	0,033278	0,032105	0,003513	0,021777
26	wi	0,041597	0,02497	0,003513	0,021341
27	i	0,041597	0,027348	0,001171	0,021341

ΠΙΝΑΚΑΣ 6.1.1 : Οι 27 συλλαβές του b3344

Οι 27 συχνότερα επαναλαμβανόμενες συλλαβές του δείγματος γραμμικής Β της Κνωσού με 11725 συλλαβές.

A/A	Συλλαβές	Αρχή σχετική επί της εκατό συχνότητα	Μέση σχετική επί της εκατό συχνότητα	Τέλος σχετική επί της εκατό συχνότητα	Συνολική σχετική επί της εκατό συχνότητα
1	jo	0,001386	0,022407	0,162906	0,061373
2	a	0,192148	0,004596	0,006883	0,054473
3	ta	0,02448	0,049411	0,077629	0,05181
4	ro	0,005081	0,053433	0,081836	0,049752
5	ra	0,029099	0,075266	0,027916	0,048178
6	ja	0,005081	0,043378	0,089101	0,047815
7	e	0,127945	0,020684	0,016826	0,047573
8	to	0,030947	0,032462	0,071128	0,044305
9	te	0,030485	0,059178	0,030593	0,04261
10	ko	0,050808	0,037058	0,034034	0,039705
11	re	0,017552	0,066648	0,017591	0,038252
12	u	0,016166	0,043091	0,049713	0,038131
13	o	0,071594	0,010342	0,043977	0,037042
14	ke	0,037875	0,054295	0,012237	0,036678
15	wo	0,021709	0,032462	0,053537	0,036315
16	ri	0,011547	0,07268	0,003824	0,034863
17	ka	0,054965	0,036196	0,014149	0,034136
18	we	0,01709	0,0316	0,047801	0,032926
19	no	0,005081	0,021833	0,061185	0,0299
20	pi	0,028637	0,029876	0,021797	0,026994
21	wa	0,025404	0,031887	0,019503	0,026268
22	i	0,034642	0,023844	0,018356	0,024936
23	ti	0,016166	0,043091	0,006883	0,024573
24	si	0,015704	0,034473	0,015296	0,023484
25	po	0,054042	0,016949	0,006501	0,023363
26	pa	0,044804	0,023269	0,004971	0,023121
27	me	0,029561	0,029589	0,003824	0,021426

ΠΙΝΑΚΑΣ 6.1.2 : Οι 27 συλλαβές του b11725

Οι 27 συχνότερα επαναλαμβανόμενες συλλαβές του δείγματος της γραμμικής Β της Πύλου με 8460 συλλαβές

A/A	Συλλαβές	Αρχή σχετική επί της εκατό συχνότητα	Μέση σχετική επί της εκατό συχνότητα	Τέλος σχετική επί της εκατό συχνότητα	Συνολική σχετική επί της εκατό συχνότητα
1	jo	0,003614	0,023746	0,159315	0,060503
2	e	0,137349	0,025661	0,019201	0,053566
3	a	0,173494	0,00766	0,002595	0,0505
4	te	0,036747	0,060896	0,038921	0,047596
5	ro	0,007831	0,055917	0,070057	0,047435
6	re	0,023494	0,076216	0,022314	0,045337
7	ra	0,029518	0,069322	0,024909	0,044853
8	ja	0,003012	0,041746	0,085106	0,044853
9	ta	0,01988	0,042512	0,066424	0,043885
10	u	0,013855	0,04481	0,060716	0,041465
11	wo	0,025904	0,039448	0,057084	0,041304
12	ke	0,040361	0,059364	0,017125	0,041142
13	o	0,074096	0,011107	0,052413	0,04082
14	ko	0,046988	0,03447	0,036326	0,038399
15	to	0,031325	0,033704	0,046705	0,037109
16	ka	0,055422	0,035236	0,013492	0,033882
17	we	0,018675	0,029108	0,050856	0,033075
18	ri	0,008434	0,064726	0,010379	0,032753
19	pi	0,025904	0,031406	0,026466	0,028396
20	no	0,006024	0,022214	0,052932	0,027428
21	wa	0,029518	0,029108	0,022314	0,027106
22	i	0,034337	0,01915	0,025428	0,025169
23	ti	0,019277	0,042129	0,005708	0,024685
24	me	0,03494	0,031789	0,004152	0,02404
25	po	0,051807	0,014937	0,006227	0,022104
26	si	0,009036	0,032555	0,018682	0,021943
27	pa	0,039157	0,021065	0,004152	0,020652

ΠΙΝΑΚΑΣ 6.1.3 : Οι 27 συλλαβές του b8460

Οι 24 συχνότερα επαναλαμβανόμενες συλλαβές του δείγματος της γραμμικής Α με 1859 συλλαβές

A/A	Συλλαβές	Αρχή σχετική επί της εκατό συχνότητα	Μέση σχετική επί της εκατό συχνότητα	Τέλος σχετική επί της εκατό συχνότητα	Συνολική σχετική επί της εκατό συχνότητα
1	a	0,195	0,003802	0,01662	0,066822
2	na	0,01	0,074144	0,088643	0,058275
3	i	0,08	0,051331	0,036011	0,055944
4	ta	0,0425	0,064639	0,058172	0,055944
5	ti	0,0175	0,057034	0,085873	0,052836
6	ja	0,0525	0,032319	0,080332	0,052059
7	da	0,0675	0,057034	0,022161	0,050505
8	ma	0,04	0,055133	0,038781	0,045843
9	ka	0,07	0,028517	0,038781	0,044289
10	ku	0,0575	0,047529	0,022161	0,043512
11	ra	0,0175	0,04943	0,058172	0,041958
12	re	0,0075	0,038023	0,077562	0,039627
13	ki	0,04	0,047529	0,024931	0,03885
14	si	0,045	0,032319	0,041551	0,03885
15	te	0,0225	0,028517	0,069252	0,038073
16	ru	0,025	0,039924	0,044321	0,036519
17	sa	0,04	0,047529	0,01662	0,036519
18	di	0,03	0,043726	0,027701	0,034965
19	ri	0,0225	0,043726	0,022161	0,03108
20	pa	0,05	0,020913	0,019391	0,029526
21	du	0,03	0,030418	0,027701	0,029526
22	mi	0,015	0,038023	0,027701	0,027972
23	su	0,0175	0,032319	0,024931	0,025641
24	ni	0,005	0,036122	0,030471	0,024864

ΠΙΝΑΚΑΣ 6.1.4 : Οι 27 συλλαβές του a1859

ΚΕΦΑΛΑΙΟ 7**7.1 Ιεραρχική ομαδοποίηση του δείγματος b3344**

Με τις παρακάτω συναρτήσεις υλοποιούμε την ιεραρχική ομαδοποίηση όπως αναφέραμε και στην εισαγωγή.

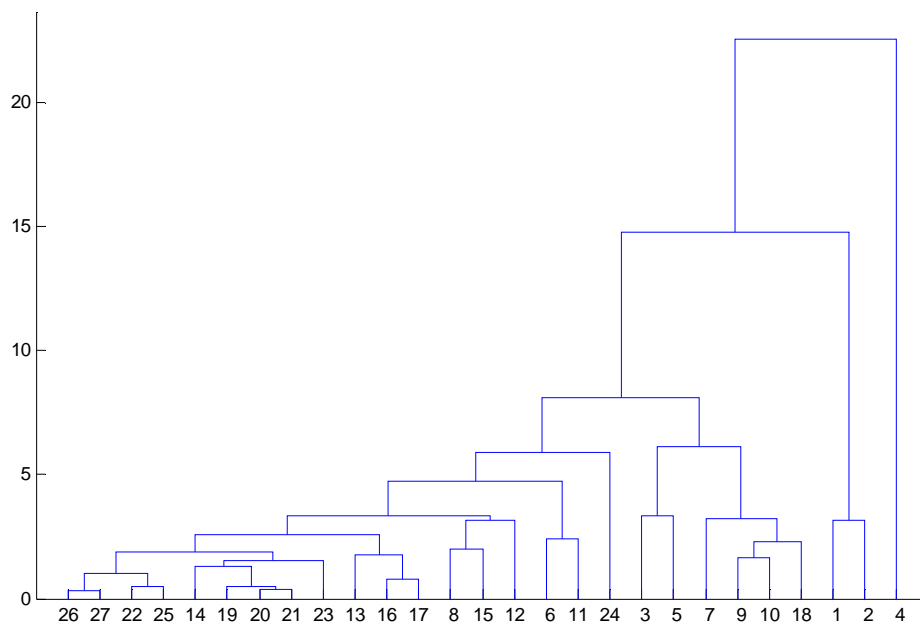
```
Y= pdist(b3344,'euclidean')
```

```
S= squareform(Y)
```

```
Z= linkage(Y,'average')
```

```
P= cophenet(Z,Y) (0.9469)
```

```
H= dendrogram (Z)
```

Δενδρόγραμμα του δείγματος b3344

Σχήμα 7.1.1: Δενδρόγραμμα του b3344

I=inconsistent(Z)			
Μέσος όρος των αποστάσεων	Τυπική απόκλιση	Ο αριθμός των συνδέσεων που παίρνουν μέρος στον υπολογισμό	Συντελεστή ασυνέπειας
0.3338	0	1	0
0.3952	0	1	0
0.4766	0	1	0
0.4520	0.0804	2	0.7071
0.8063	0	1	0
0.6084	0.3592	3	1.1317
0.9163	0.5763	2	0.7071
1.4262	0.1448	2	0.7071
1.6452	0	1	0
1.2785	0.6678	2	0.7071
1.4741	0.4344	3	0.9313
2.0215	0	1	0
1.9727	0.4632	2	0.7071
2.4025	0	1	0
2.0713	0.4491	3	1.1429
3.1789	0	1	0
2.6036	0.8233	2	0.7071
2.7577	0.6470	2	0.7071
3.0282	0.3896	3	0.7345
3.3351	0	1	0
3.4891	1.1837	3	1.0656
5.3099	0.7911	2	0.7071
4.2174	1.6331	3	1.1539
6.6867	1.2198	3	1.1494
8.6831	5.8242	3	1.0471
18.6621	5.4876	2	0.7071

ΠΙΝΑΚΑΣ 7.1.1: Η συνάρτηση inconsistent στο δείγμα b3344

t= cluster(z,0.8)					
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
wi	ja	ra	e	u	ro
i	na	ri		no	jo
	re			wo	
				so	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
a	ma	wa	ka	ta	
	sa	te	ko	to	
		ke	da		
		si			
		ti			

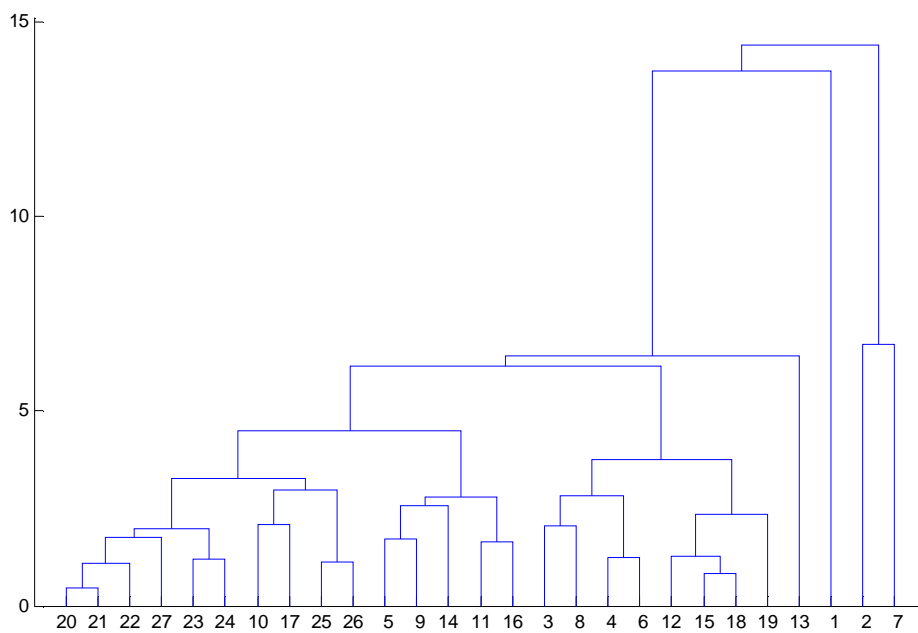
ΠΙΝΑΚΑΣ 7.1.2: Η ομαδοποίηση του b3344 μετά την υλοποίηση της συνάρτησης cluster

7.2 Ιεραρχική ομαδοποίηση του δείγματος b11725

Με τις παρακάτω συναρτήσεις υλοποιούμε την ιεραρχική ομαδοποίηση όπως αναφέραμε και στην εισαγωγή.

```
Y= pdist(b11725,'euclidean')
S= squareform(Y)
Z= linkage(Y,'average')
P= cophenet(Z,Y) (0.8885)
H= dendrogram (Z)
```

Δενδρόγραμμα του δείγματος b11725



Σχήμα 7.2.1: Δενδρόγραμμα του b11725

I=inconsistent(Z)			
Μέσος όρος των αποστάσεων	Τυπική απόκλιση	Ο αριθμός των συνδέσεων που παίρνουν μέρος στον υπολογισμό	Συντελεστή ασυνέπειας
0.4504	0	1	0
0.8153	0	1	0
0.7698	0.4517	2	0.7071
1.1300	0	1	0
1.2102	0	1	0
1.2555	0	1	0
1.0453	0.3253	2	0.7071
1.6537	0	1	0
1.7289	0	1	0
1.4234	0.4726	2	0.7071
1.6520	0.3996	3	0.8415
2.0680	0	1	0
2.1082	0	1	0
1.8147	0.7629	2	0.7071
2.1430	0.5856	2	0.7071
2.3331	0.5997	3	0.7595
2.0491	0.7843	3	0.9877
2.0778	0.9331	3	0.9833
2.7483	0.6715	3	0.7640
2.9761	0.7105	3	1.0899
3.5138	0.8791	3	1.1122
4.7964	1.2270	3	1.1008
6.2856	0.1959	2	0.7071
6.7285	0	1	0
10.0931	5.1888	2	0.7071
11.6313	4.2580	3	0.6510

ΠΙΝΑΚΑΣ 7.2.1: Η συνάρτηση inconsistent στο δείγμα b11725

t=cluster(z,0.8)					
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ti	ro	po	u	ra	o
si	ja	pa	wo	te	
			we	re	
			no	ke	
				ri	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
jo	a	pi	ta	ko	
	e	wa	to	ka	
		i			
		me			

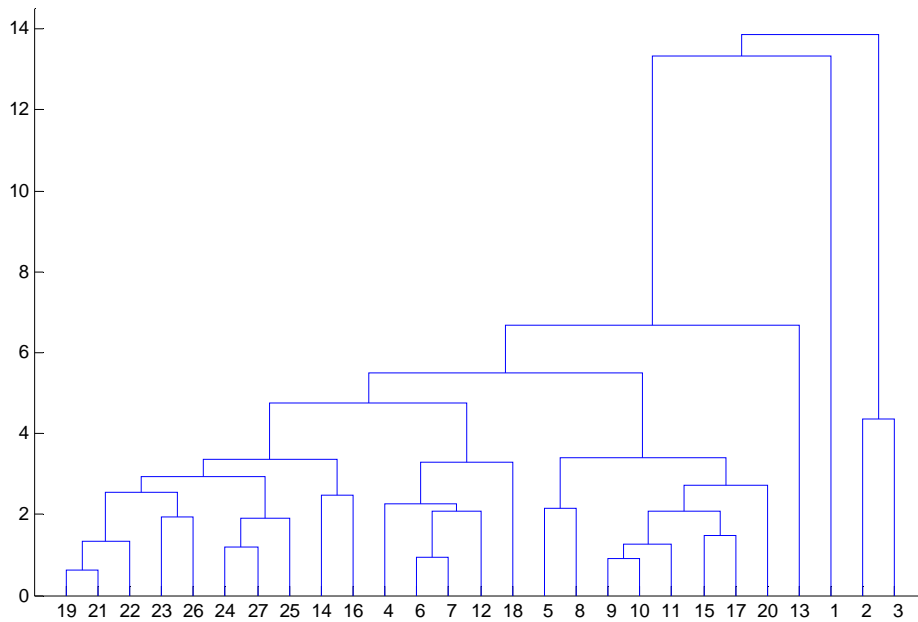
ΠΙΝΑΚΑΣ 7.2.2: Η ομαδοποίηση του b11725 μετά την υλοποίηση της συνάρτησης cluster

7.3 Ιεραρχική ομαδοποίηση του δείγματος b8460

Με τις παρακάτω συναρτήσεις υλοποιούμε την ιεραρχική ομαδοποίηση όπως αναφέραμε και στην εισαγωγή.

```
Y= pdist(b8460,'euclidean')  
S= squareform(Y)  
Z= linkage(Y,'average')  
P= cophenet(Z,Y) (0.9135)  
H= dendrogram (Z)
```

Δενδρόγραμμα του δείγματος b8460



Σχήμα 7.3.1: Δενδρόγραμμα του b8460

I=inconsistent(Z)			
Μέσος όρος των αποστάσεων	Τυπική απόκλιση	Ο αριθμός των συνδέσεων που παίρνουν μέρος στον υπολογισμό	Συντελεστή ασυνέπειας
0.6103	0	1	0
0.8945	0	1	0
0.9528	0	1	0
1.2011	0	1	0
1.0846	0.2689	2	0.7071
0.9780	0.5200	2	0.7071
1.4651	0	1	0
1.5579	0.5046	2	0.7071
1.9297	0	1	0
1.6072	0.4218	3	1.1249
1.5219	0.8049	2	0.7071
2.1382	0	1	0
2.1696	0.1111	2	0.7071
2.4769	0	1	0
1.9416	0.6020	3	1.0097
2.4112	0.4660	2	0.7071
2.4698	0.5198	3	0.9145
2.7681	0.7353	2	0.7071
2.9308	0.4469	3	0.9835
2.7546	0.6235	3	1.0110
4.3768	0	1	0
3.8074	0.8293	3	1.1533
4.5441	1.0664	3	0.8809
6.0707	0.8304	2	0.7071
9.9989	4.7250	2	0.7071
10.5262	5.3319	3	0.6256

ΠΙΝΑΚΑΣ 7.3.1: Η συνάρτηση inconsistent στο δείγμα b8460

t=cluster(z,0.8)					
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ta	pi	no	me	ko	ro
u	wa		po	ka	ja
wo	i		pa		
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
te	o	jo	e	to	ti
re			a	we	si
ra					
ke					
ri					

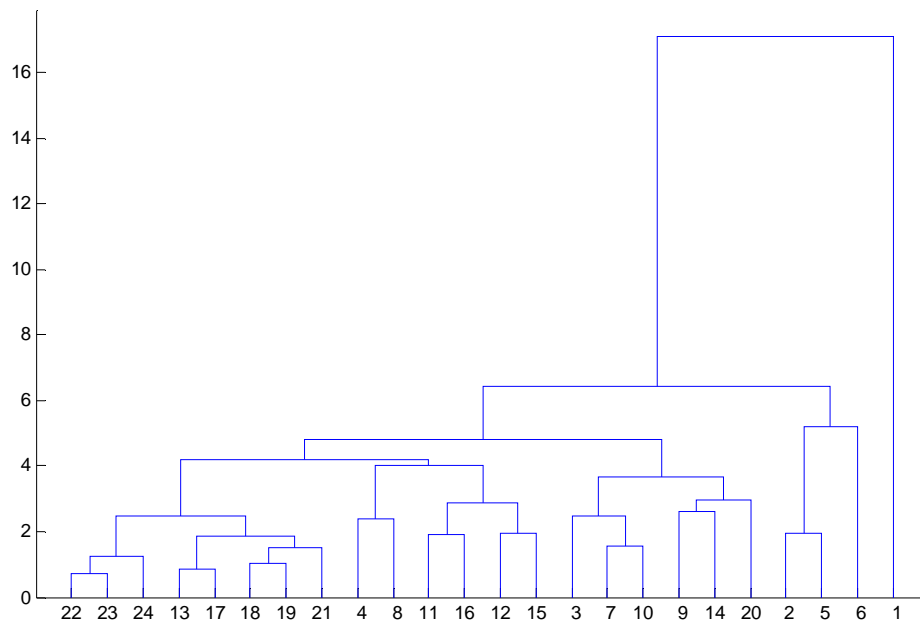
ΠΙΝΑΚΑΣ 7.3.2: Η ομαδοποίηση του b8460 μετά την υλοποίηση της συνάρτησης cluster

7.4 Ιεραρχική ομαδοποίηση του δείγματος a1859

Με τις παρακάτω συναρτήσεις υλοποιούμε την ιεραρχική ομαδοποίηση όπως αναφέραμε και στην εισαγωγή.

```
Y= pdist(a1859,'euclidean')  
S= squareform(Y)  
Z= linkage(Y,'average')  
P= cophenet(Z,Y) (0.9267)  
H= dendrogram (Z)
```

Δενδρόγραμμα του δείγματος a1859



Σχήμα 7.4.1: Δενδρόγραμμα του a1859

I=inconsistent(Z)			
Μέσος όρος των αποστάσεων	Τυπική απόκλιση	Ο αριθμός των συνδέσεων που παίρνουν μέρος στον υπολογισμό	Συντελεστή ασυνέπειας
0.7203	0	1	0
0.8631	0	1	0
1.0101	0	1	0
0.9904	0.3820	2	0.7071
1.2726	0.3711	2	0.7071
1.5468	0	1	0
1.4231	0.5132	3	0.8729
1.9184	0	1	0
1.9654	0	1	0
1.9668	0	1	0
2.3972	0	1	0
1.8646	0.6009	3	0.9946
2.0159	0.6634	2	0.7071
2.6014	0	1	0
2.2496	0.5323	3	1.1535
2.7884	0.2645	2	0.7071
3.0460	0.5994	3	1.0537
3.0913	0.8317	3	1.1084
3.5631	0.9587	3	0.6789
4.2308	0.5619	3	1.0147
3.5761	2.2779	2	0.7071
5.4727	0.8515	3	1.1247
11.7646	7.5437	2	0.7071

ΠΙΝΑΚΑΣ 7.4.1: Η συνάρτηση inconsistent στο δείγμα α1859

t=cluster(z,0.8)				
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα
ki	mi	ra	i	ta
sa	su	ru	da	ma
	ni		ku	
6η ομάδα	7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα
a	di	re	ka	na
	ri	te	si	ti
	du		pa	ja

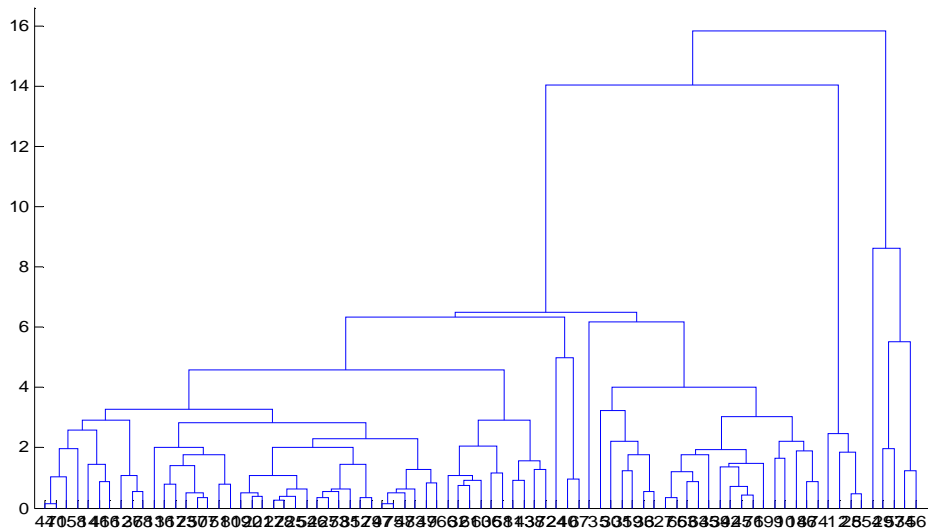
ΠΙΝΑΚΑΣ 7.4.2: Η ομαδοποίηση του a1859 μετά την υλοποίηση της συνάρτησης cluster

7.5 Ιεραρχική ομαδοποίηση των δειγμάτων b3344, b1725 & b8460

Με τις παρακάτω συναρτήσεις υλοποιούμε την ιεραρχική ομαδοποίηση όπως αναφέραμε και στην εισαγωγή.

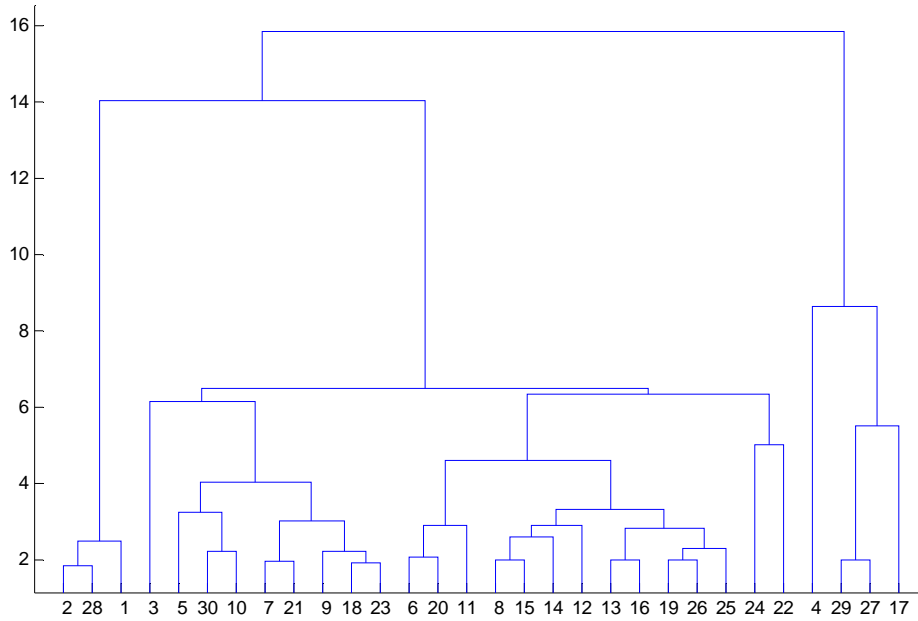
```
Y= pdist(bola,'euclidean')
S= squareform(Y)
Z= linkage(Y,'average')
P= cophenet(Z,Y) (0.8876)
H= dendrogram (Z)
```

Δενδρόγραμμα των δειγμάτων b3344, b1725 & b8460



Σχήμα 7.5.1: Δενδρόγραμμα των b3344, b1725 & b8460

Δενδρόγραμμα των δειγμάτων b3344, b1725 & b8460



Σχήμα 7.5.2: Δενδρόγραμμα που παρουσιάζει τις 30 συνδέσεις που γίνονται στο ανώτερο επίπεδο των δειγμάτων b3344, b1725 & b8460

I=inconsistent(Z)			
Μέσος όρος των αποστάσεων	Τυπική απόκλιση	Ο αριθμός των συνδέσεων που παίρνουν μέρος στον υπολογισμό	Συντελεστή ασυνέπειας
0,1276	0	1	0
0,1283	0	1	0
0,2819	0	1	0
0,3271	0	1	0
0,3338	0	1	0
0,342	0	1	0
0,3463	0	1	0
0,3325	0,0716	2	0,7071
0,3952	0	1	0
0,4252	0	1	0
0,4518	0	1	0
0,452	0,0804	2	0,7071
0,3209	0,2724	2	0,7071
0,438	0,1296	2	0,7071
0,5315	0	1	0
0,4411	0,1519	2	0,7071
0,5628	0	1	0
0,5859	0,0529	2	0,7071
0,577	0,0897	2	0,7071
0,5164	0,1884	2	0,7071
0,5708	0,2059	2	0,7071
0,7456	0	1	0
0,7873	0	1	0
0,8063	0	1	0
0,8497	0	1	0
0,8675	0	1	0
0,8701	0	1	0
0,8945	0	1	0
0,8228	0,1091	2	0,7071
0,9171	0	1	0

0,9607	0	1	0
0,5887	0,6521	2	0,7071
0,9946	0,1339	2	0,7071
0,8134	0,3986	2	0,7071
0,7521	0,3077	3	101241
1,1678	0	1	0
0,8137	0,437	3	0,8947
1,2441	0	1	0
1,2563	0	1	0
0,919	0,3189	3	1,0907
1,3004	0	1	0
1,0354	0,451	2	0,7071
0,9145	0,4488	3	1,0985
0,8015	0,5842	3	1,117
1,1646	0,4165	2	0,7071
1,4231	0,0973	2	0,7071
1,2571	0,3206	3	0,9256
1,6452	0	1	0
1,191	0,5982	3	0,9409
1,3167	0,4903	3	0,8944
1,4805	0,3901	2	0,7071
1,1479	0,9844	2	0,7071
1,3842	0,7306	2	0,7071
1,7289	0,2245	3	0,9334
1,5141	0,6567	2	0,7071
1,9787	0	1	0
1,5169	0,4536	3	1,0621
1,88	0,1765	2	0,7071
1,4428	0,5457	3	1,1517
1,9194	0,284	3	1,0311
1,9883	0,3315	2	0,7071
1,8577	0,5345	3	0,8416
2,1551	0,44	2	0,7071
2,0159	0,5765	3	1,0309
2,3796	0,4156	3	1,0754

2,2012	0,9685	3	0,7196
2,1759	0,6805	3	1,0679
2,3927	0,5665	3	1,1205
2,729	0,7161	2	0,7071
3,0084	0,254	3	1,1433
3,4287	0,5253	3	1,1319
3,6035	0,8919	3	1,1256
2,9831	2,8602	2	0,7071
2,9126	2,2835	3	1,1397
5,0937	1,5138	2	0,7071
5,3161	0,9048	3	1,1264
6,3336	0,1686	3	0,9951
7,073	2,2034	2	0,7071
7,6717	5,8786	3	1,0846
12,8443	3,759	3	0,8008

ΠΙΝΑΚΑΣ 7.5.1: Η συνάρτηση inconsistent στα δείγματα b3344, b1725 & b8460

t=cluster(z,0.8)						
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα	7η ομάδα
te b3344	u b3344	pi b11725	ti b3344	po b11725	ri b3344	ja b11725
ke b3344	wo b8460	wa b11725	ti b11725	po b8460	ri b11725	ja b8460
si b3344		pi b8460	ti b8460			
		wa b8460				
8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα	13η ομάδα	14η ομάδα
si b11725	to b11725	u b11725	ja b3344	ra b3344	no b3344	ta b11725
si b8460		wo b11725		ra b11725	wo b3344	
		we b11725		re b8460		
		to b8460		ra b8460		
		we b8460				
15η ομάδα	16η ομάδα	17η ομάδα	18η ομάδα	19η ομάδα	20η ομάδα	21η ομάδα
wa b3344	ko b3344	to b3344	e b11725	ta b3344	a b3344	ro b3344
ke b11725	ko b11725		e b8460			jo b3344
ke b8460	ko b8460					jo b11725
						jo b8460
22η ομάδα	23η ομάδα	24η ομάδα	25η ομάδα	26η ομάδα	27η ομάδα	28η ομάδα
ma b3344	ta b8460	i b11725	na b3344	wi b3344	re b11725	ro b11725
sa b3344	u b8460	i b8460	re b3344	i b3344	ri b8460	ro b8460
me b11725				pa b11725		
me b8460				pa b8460		
29η ομάδα	30η ομάδα	31η ομάδα	32η ομάδα	33η ομάδα		
te b11725	so b3344	ka b3344	a b11725	e b3344		
te b8460	no b11725	da b3344	a b8460	o b11725		
	no b8460	ka b11725		o b8460		
		ka b8460				

ΠΙΝΑΚΑΣ 7.5.2: Η ομαδοποίηση των b3344, b11725 & b8460 μετά την υλοποίηση της συνάρτησης cluster

Έτσι όπως κατηγοριοποιήθηκαν τα πρότυπα των τριών δειγμάτων της γραμμικής Β μετά την ιεραρχική ομαδοποίηση παρατηρούμε ότι:

- 1) Το 15% των συλλαβών με ίδια φωνητική αξία και από τα τρία δείγματα ταξινομούνται στις ίδιες ομάδες .
- 2) Το 63% των συλλαβών με ίδια φωνητική αξία από τα δείγματα b11725 και b8460 ταξινομούνται στις ίδιες ομάδες. Διαπιστώνουμε με αυτό το ποσοστό ότι υπάρχουν πολλά κοινά χαρακτηριστικά μεταξύ των επιγραφών της Κνωσού (b11725) και της Πύλου (b8460).
- 3) Ένα πολύ μικρό ποσοστό των συλλαβών του δείγματος b3344 ταξινομείται στις ίδιες ομάδες με τα δύο άλλα δείγματα και κυρίως με το δείγμα b11725, αυτό συμβαίνει ίσως επειδή το πλήθος των συλλαβών του δείγματος b3344 (3344 συλλαβές) είναι πολύ μικρότερο από τα άλλα δύο δείγματα αλλά και επειδή το δείγμα b3344 αποτελεί ένα αρχείο με τυχαίες λέξεις από όλη τη γραμμική β από όλες τις περιοχές εμφάνισής της. Ο λόγος που το δείγμα b3344 ταυριάζει λίγο περισσότερο με το δείγμα b11725 είναι επειδή στο b3344 εκτός από τις τυχαίες

λέξεις που επιλέχθηκαν από όλη τη γραμμική β, επιλέχθηκαν κι εκείνες που παρουσίαζαν μεγαλύτερη επανάληψη, πάνω από πέντε φορές, στα αρχεία της Κνωσού (η επιλογή αυτή κρίθηκε αναγκαία επειδή η μεγάλη συχνότητα μιας λέξης σε μια γραφή δημιουργεί μεγαλύτερη πιθανότητα να εντοπισθεί σε άλλη συγγενική γραφή).

Τα δείγματα δεν απαρτίζονται από τις ίδιες φωνητικές συλλαβές-αξίες κι αυτό γιατί κάθε δείγμα είτε προέρχεται από άλλη περιοχή(Κνωσό, Πύλο) είτε αναφέρεται σε διαφορετικού είδους αρχείο(κύρια ονόματα, τοπωνύμια). Οπότε ανάλογα με την περιοχή και το είδος του κειμένου άλλες συλλαβές εμφανίζονται συχνότερα από κάποιες άλλες και εφόσον ασχολούμαστε μόνο με τις 27 συχνότερα επαναλαμβανόμενες συλλαβές από κάθε δείγμα, είναι λογικό να μην έχουμε τα ίδια φωνητικά σημεία σε κάθε δείγμα. Παρατηρώντας την ιεραρχική ομαδοποίηση των συλλαβών θα προσπαθήσουμε να βγάλουμε συμπεράσματα για τις ομάδες που δημιουργούνται: ποιες από τις ομάδες των τριών δειγμάτων περιλαμβάνουν τις ίδιες συλλαβικές αξίες και ποιών δειγμάτων η ομαδοποίηση ταιριάζει περισσότερο.

Παρατηρώντας τις ομαδοποιήσεις που έκανε το matlab σε κάθε δείγμα της γραμμικής β ξεχωριστά:

t=cluster(z,0.8)					
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
wi	ja	ra	e	u	ro
i	na	ri		no	jo
	re			wo	
				so	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
a	ma	wa	ka	ta	
	sa	te	ko	to	
		ke	da		
		si			
		ti			

ΠΙΝΑΚΑΣ 7.5.3: Η ομαδοποίηση του b3344 μετά την υλοποίηση της συνάρτησης cluster

t=cluster(z,0.8)					
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ti	ro	po	u	ra	o
si	ja	pa	wo	te	
			we	re	
			no	ke	
				ri	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
jo	a	pi	ta	ko	
	e	wa	to	ka	
		i			
		me			

ΠΙΝΑΚΑΣ 7.5.4: Η ομαδοποίηση του b11725 μετά την υλοποίηση της συνάρτησης cluster

t=cluster(z,0.8)					
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ta	pi	no	me	ko	ro
u	wa		po	ka	ja
wo	i		pa		
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
te	o	jo	e	to	ti
re			a	we	si
ra					
ke					
ri					

ΠΙΝΑΚΑΣ 7.5.6: Η ομαδοποίηση του b8460 μετά την υλοποίηση της συνάρτησης cluster

Διαπιστώνουμε πως οι συλλαβές των δειγμάτων b11725 και b8460 ομαδοποιούνται παρόμοια. Οι συλλαβικές αξίες:

Συλλαβές που ομαδοποιούνται ίδια από τα δείγματα b11725 & b8460									
ti	ka	ro	ra	o	jo	a	pi	u	po
si	ko	ja	te			e	wa	wo	pa
			re				i		
			ke						
			ri						

Και στα δύο δείγματα ομαδοποιούνται μαζί. Να επισημάνουμε πως οι 27 συλλαβές του b11725 και 27 συλλαβές του b8460 προέρχονται από διαφορετικές περιοχές. Οι μεν 27 συλλαβές του b11725 είναι από την Κνωσό, οι δε 27 συλλαβές του b8460 είναι από την Πύλο. Το ότι αρχεία από διαφορετικές περιοχές ταυτίζονται είναι ενθαρρυντικό.

7.6 Ιεραρχική ομαδοποίηση των δειγμάτων b11725,b8460

Θα εφαρμόσουμε ιεραρχική ομαδοποίηση μόνο στα 2 δείγματα της γραμμικής β (b11725&β8460) γιατί όπως διαπιστώσαμε στην παραπάνω παράγραφο ταιριάζουν περισσότερο αλλά και γιατί τις ομάδες που θα προκύψουν θα τις χρησιμοποιήσουμε στο κεφάλαιο ταξινόμηση για λόγους που εξηγούνται στο ίδιο κεφάλαιο.

Με τις παρακάτω συναρτήσεις υλοποιούμε την ιεραρχική ομαδοποίηση όπως αναφέραμε και στην εισαγωγή.

Y= pdist(b11725-b8460,'euclidean')

S= squareform(Y)

Z= linkage(Y,'average')

P= cophenet(Z,Y) (0.9018)

H= dendrogram (Z)

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα	7η ομάδα	8η ομάδα	9η ομάδα
pi b11725	po b11725	ti b11725	ja b11725	ta b8467	ra b11725	me b11725	ka b11725	ke b11725
wa b11725	po b8467	ti b8467	ja b8467	u b8467	re b8467	me b8467	ka b8467	ke b8467
pi b8467				wo b8467	ra b8467			
wa b8467								
10η ομάδα	11η ομάδα	12η ομάδα	13η ομάδα	14η ομάδα	15η ομάδα	16η ομάδα	17η ομάδα	18η ομάδα
no b11725	re b11725	e b11725	o b11725	jo b11725	i b11725	pa b11725	si b11725	ro b11725
no b8467	ri b11725	e b8467	o b8467	jo b8467	i b8467	pa b8467	si b8467	ro b8467
	ri b8467							
19η ομάδα	20η ομάδα	21η ομάδα	22η ομάδα	23η ομάδα				
u b11725	te b11725	ko b11725	ta b11725	a b11725				
wo b11725	te b8467	ko b8467	to b11725	a b8467				
we b11725								
to b8467								
we b8467								

ΠΙΝΑΚΑΣ 7.5.2: Η ομαδοποίηση των b11725 & b8460 μετά την υλοποίηση της συνάρτησης cluster

ΚΕΦΑΛΑΙΟ 8**8.1 Υλοποίηση του αλγόριθμου ομαδοποίησης kmeans**

Σ' αυτό το σημείο θα χρησιμοποιήσουμε τις ομάδες που μας έδωσε η ιεραρχική ομαδοποίηση και θα ζητήσουμε από τον αλγόριθμο Kmeans να μας ομαδοποιήσει τα δεδομένα - πρότυπα μας σύμφωνα με αυτές τις ομάδες. Για να αξιολογήσουμε την ομαδοποίηση του kmeans έχουμε 2 τρόπους.

- 1) Ο πρώτος είναι οπτικός και αφορά τα διαγράμματα silhouette (όσο πιο κοντά στο 1 τόσο καλύτερη ταξινόμηση έχω) και
- 2) ο δεύτερος είναι μετρίσιμος και αφορά τον τρόπο που εργάζεται ο Kmeans. Συγκεκριμένα ο αλγόριθμος φτιάχνει τις ομάδες του με τέτοιο τρόπο ώστε να ελαχιστοποιεί το άθροισμα των αποστάσεων μεταξύ των ομάδων που κατασκευάζει.

Στις πολλές δοκιμές που θα κάνουμε λοιπόν με τον Kmeans θα διαλέξουμε την ομαδοποίηση που παρουσιάζει το μικρότερο άθροισμα. Πρέπει να ελαχιστοποιήσουμε δηλαδή το συνολικό άθροισμα των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους.

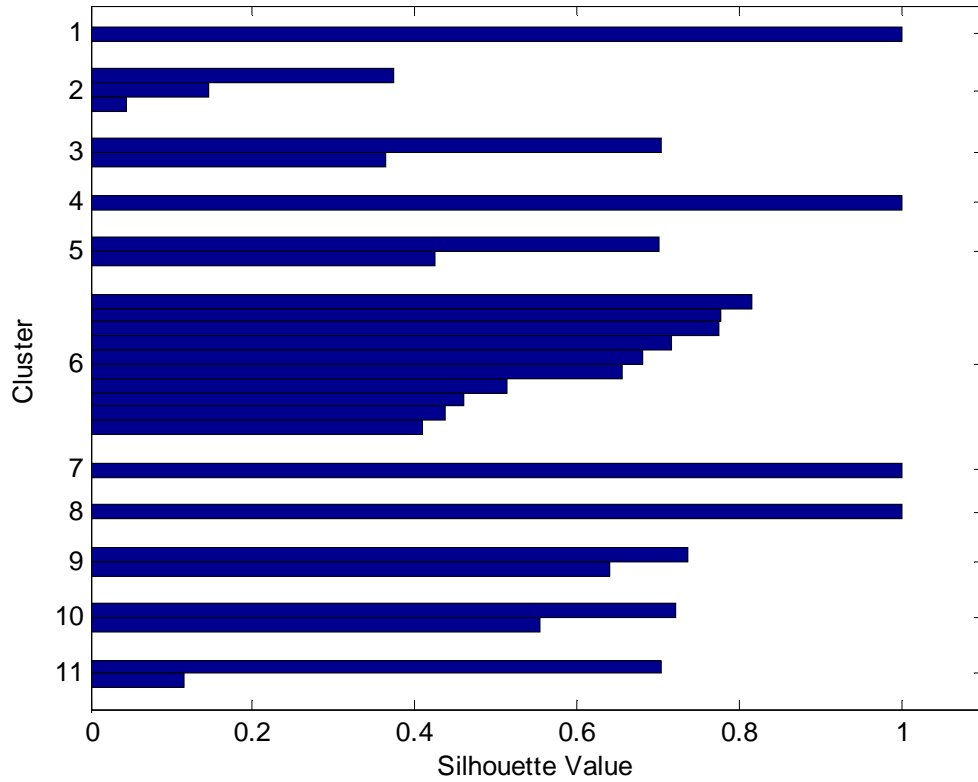
Kmeans στο δείγμα b3344 με 11 ομάδες**1^η δοκιμή**

[idx,c]=kmeans(b3344,11,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	27	816.562
2	1	5	568.215
3	2	5	409.203
4	2	1	394.515
4 iterations, total sum of distances = 39.4515			

ΠΙΝΑΚΑΣ 8.1.1: Πρώτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b3344.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
a	e	ka	ro	no	ja
	wi	da		so	wa
	i				na
					re
					te
					ke
					si
					ma
					ti
					sa
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
ko	jo	ra	ta	u	
		ri	to	wo	

ΠΙΝΑΚΑΣ 8.1.2: ομαδοποίηση του δείγματος b3344 σε 11 ομάδες μετά την πρώτη δοκιμή του kmeans.



Σχήμα 8.1.1: Το διάγραμμα εγγύτητας Silhouette του δείγματος b3344 στην πρώτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα b3344 με 11 ομάδες**2^η δοκιμή**

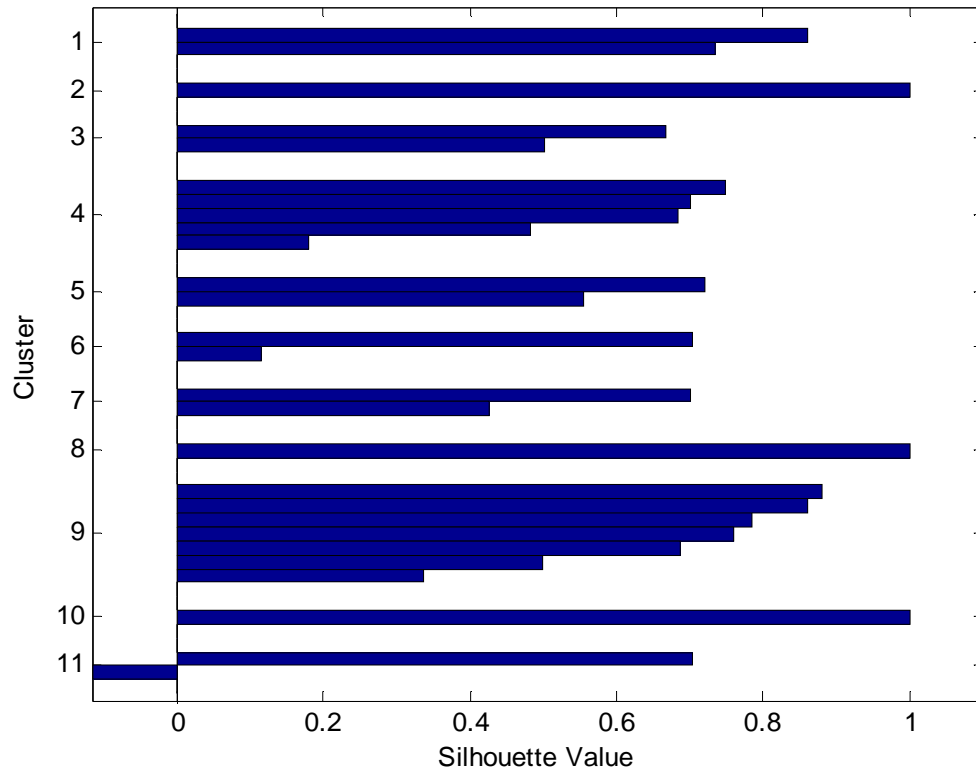
```
[idx,c]=kmeans(b3344,11,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	27	436.188
2	1	4	371.766
3	2	3	353.263
4	2	2	297.815
4 iterations, total sum of distances = 29.7815			

ΠΙΝΑΚΑΣ 8.1.3: Δεύτερη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b3344.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ro	ko	ra	ja	ta	u
jo		ri	wa	to	wo
			na		
			re		
			ti		
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
no	e	te	a	ka	
so		ke		da	
		si			
		ma			
		sa			
		wi			
		i			

ΠΙΝΑΚΑΣ 8.1.4: ομαδοποίηση του δείγματος b3344 σε 11 ομάδες μετά την δεύτερη δοκιμή του kmeans.



Σχήμα 8.1.2: Το διάγραμμα εγγύτητας Silhouette του δείγματος b3344 στην δεύτερη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα b3344 με 11 ομάδες**3^η δοκιμή**

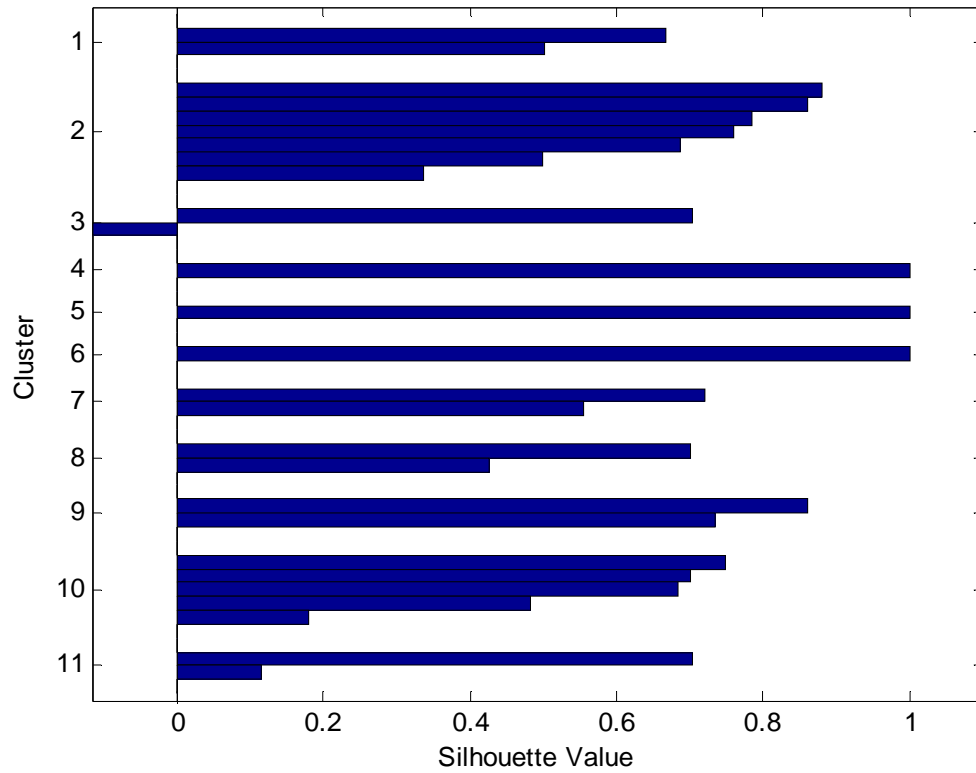
```
[idx,c]=kmeans(b3344,11,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	27	931.708
2	1	4	830.065
3	2	6	353.263
4	2	2	297.815
4 iterations, total sum of distances = 29.7815			

ΠΙΝΑΚΑΣ 8.1.5: Τρίτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b3344.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ra	te	ka	e	ko	a
ri	ke	da			
	si				
	ma				
	sa				
	wi				
	i				
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
ta	no	ro	ja	u	
to	so	jo	wa	wo	
			na		
			re		
			ti		

ΠΙΝΑΚΑΣ 8.1.6: ομαδοποίηση του δείγματος b3344 σε 11 ομάδες μετά την τρίτη δοκιμή του kmeans.



Σχήμα 8.1.3: Το διάγραμμα εγγύτητας Silhouette του δείγματος b3344 στην τρίτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα b3344 με 11 ομάδες**4^η δοκιμή**

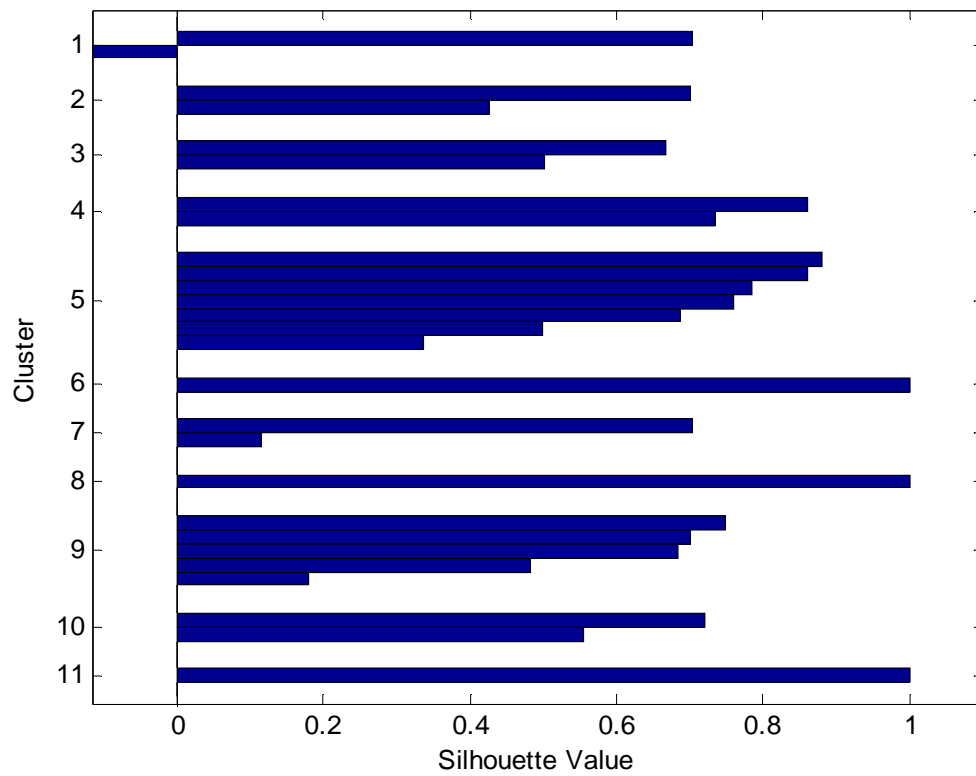
```
[idx,c]=kmeans(b3344,11,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	27	105.804
2	1	5	884.363
3	1	1	872.691
4	2	8	308.583
5	2	3	297.815
5 iterations, total sum of distances = 29.7815			

ΠΙΝΑΚΑΣ 8.1.7: Τέταρτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b3344.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ka	no	ra	ro	te	a
da	so	ri	jo	ke	
				si	
				ma	
				sa	
				wi	
				i	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
u	e	ja	ta	ko	
wo		wa	to		
		na			
		re			
		ti			

ΠΙΝΑΚΑΣ 8.1.8: ομαδοποίηση του δείγματος b3344 σε 11 ομάδες μετά την τέταρτη δοκιμή του kmeans.



Σχήμα 8.1.4: Το διάγραμμα εγγύτητας Silhouette του δείγματος b3344 στην τέταρτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

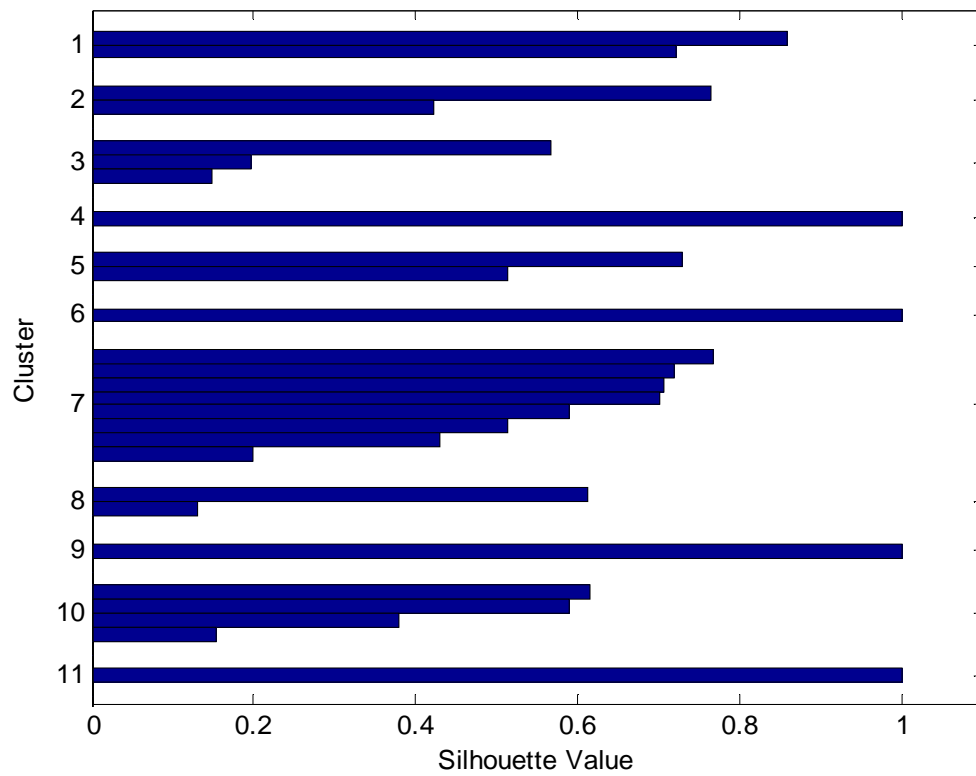
8.2 Kmeans στο δείγμα b11725 με 10 ομάδες**1^η δοκιμή**

[idx,c]=kmeans(b11725,11,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	27	101.535
2	2	6	53.204
3	2	2	500.056
3 iterations, total sum of distances = 50.0056			

ΠΙΝΑΚΑΣ 8.2.1: Πρώτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b11725.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
po	re	ra	a	ti	o
pa	ri	te		si	
		ke			
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
ta	ko	e	pi		
ro	ka		wa		
ja			i		
to			me		
u			jo		
wo					
we					
no					

ΠΙΝΑΚΑΣ 8.2.2: ομαδοποίηση του δείγματος b11725 σε 11 ομάδες μετά την πρώτη δοκιμή του kmeans.



Σχήμα 8.2.1: Το διάγραμμα εγγύτητας Silhouette του δείγματος b11725 στην πρώτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

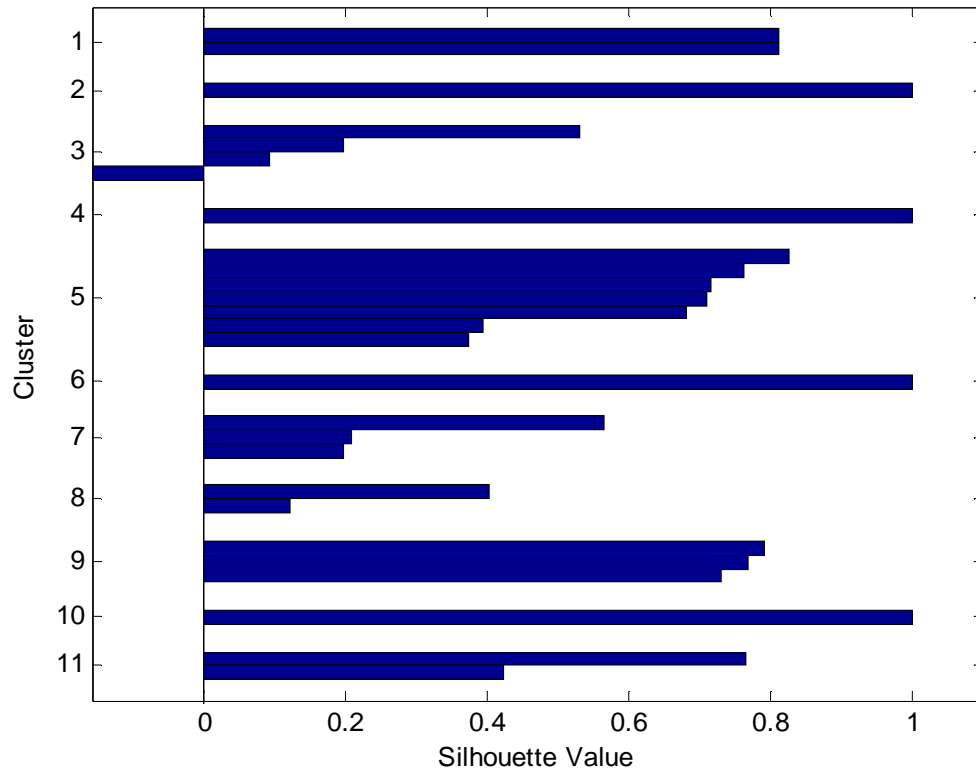
Kmeans στο δείγμα b11725 με 11 ομάδες**2^η δοκιμή**

[idx,c]=kmeans(b11725,11,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	27	678.685
2	1	1	65.91
3	1	1	632.531
4	1	1	619.429
5	2	5	55.102
6	2	3	521.316
7	2	2	445.031
8	2	2	436.367
9	2	1	423.699
9 iterations, total sum of distances = 42.3699			

ΠΙΝΑΚΑΣ 8.2.3: Δεύτερη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b11725.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ro	a	ko	e	pi	jo
ja		o		wa	
		ka		i	
		po		ti	
				si	
				pa	
				me	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
ra	ta	u	no	re	
te	to	wo		ri	
ke		we			

ΠΙΝΑΚΑΣ 8.2.4: ομαδοποίηση του δείγματος b11725 σε 11 ομάδες μετά την δεύτερη δοκιμή του kmeans.



Σχήμα 8.2.2: Το διάγραμμα εγγύτητας Silhouette του δείγματος b11725 στην δεύτερη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

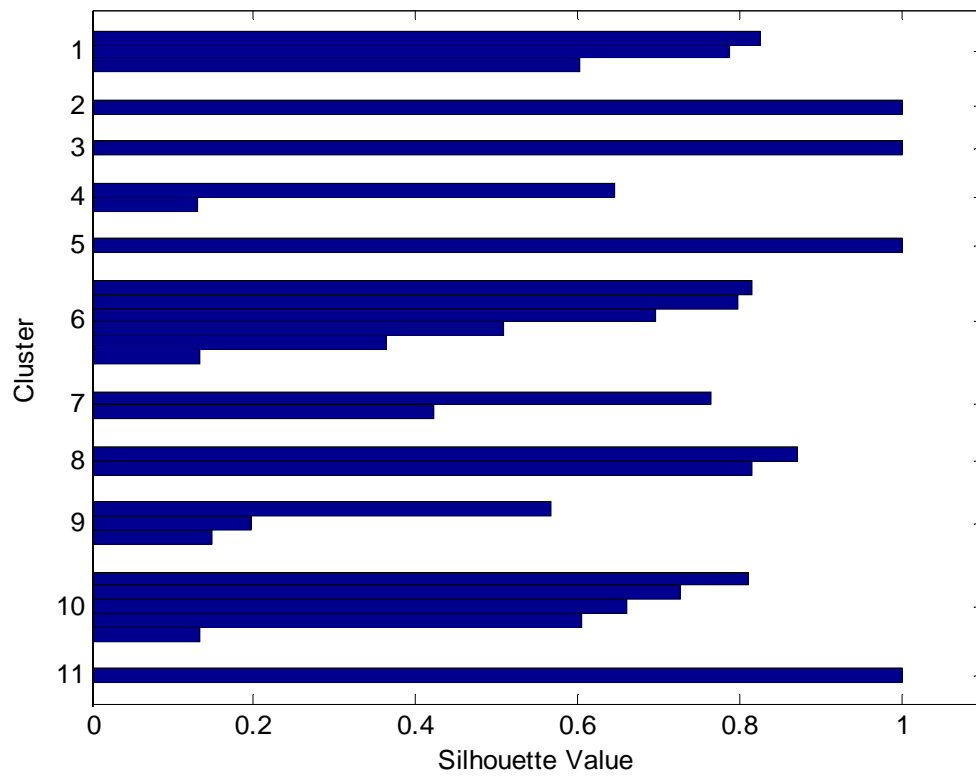
Kmeans στο δείγμα b11725 με 11 ομάδες**3^η δοκιμή**

[idx,c]=kmeans(b11725,11,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	27	151.878
2	1	2	145.088
3	1	1	134.934
4	1	4	71.708
5	1	2	596.924
6	1	1	528.438
7	2	6	351.603
8	2	2	319.619
8 iterations, total sum of distances = 31.9619			

ΠΙΝΑΚΑΣ 8.2.5: Τρίτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b11725.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ta	o	a	ko	jo	pi
ro			ka		wa
ja					i
					ti
					si
					me
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
re	po	ra	to	e	
ri	pa	te	u		
		ke	wo		
			we		
			no		

ΠΙΝΑΚΑΣ 8.2.6: ομαδοποίηση του δείγματος b11725 σε 11 ομάδες μετά την τρίτη δοκιμή του kmeans.



Σχήμα 8.2.3: Το διάγραμμα εγγύτητας Silhouette του δείγματος b11725 στην τρίτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα b11725 με 11 ομάδες**4^η δοκιμή**

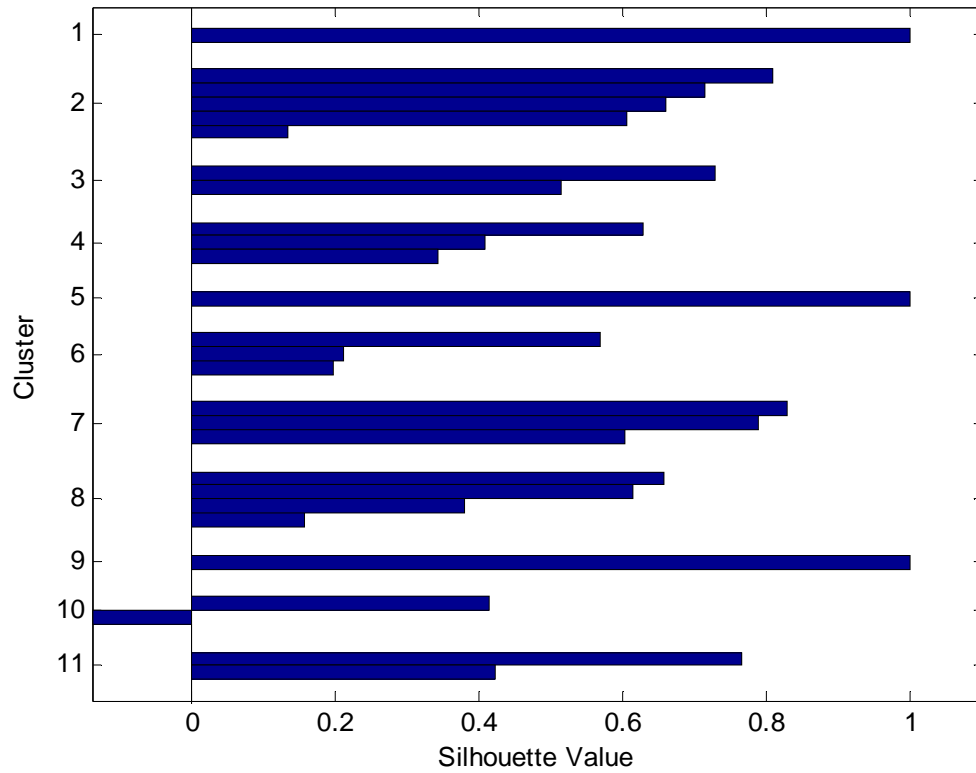
```
[idx,c]=kmeans(b11725,11,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	27	172.492
2	1	3	113.339
3	2	8	551.415
4	2	2	504.793
5	2	3	350.724
5 iterations, total sum of distances = 35.0724			

ΠΙΝΑΚΑΣ 8.2.7: Τέταρτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b11725.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
e	to	ti	ka	jo	ra
	u	si	po		te
	wo		pa		ke
	we				
	no				
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
ta	pi	a	ko	re	
ro	wa		o	ri	
ja	i				
	me				

ΠΙΝΑΚΑΣ 8.2.8: ομαδοποίηση του δείγματος b11725 σε 11 ομάδες μετά την τέταρτη δοκιμή του kmeans.



Σχήμα 8.2.4: Το διάγραμμα εγγύτητας Silhouette του δείγματος b11725 στην τέταρτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

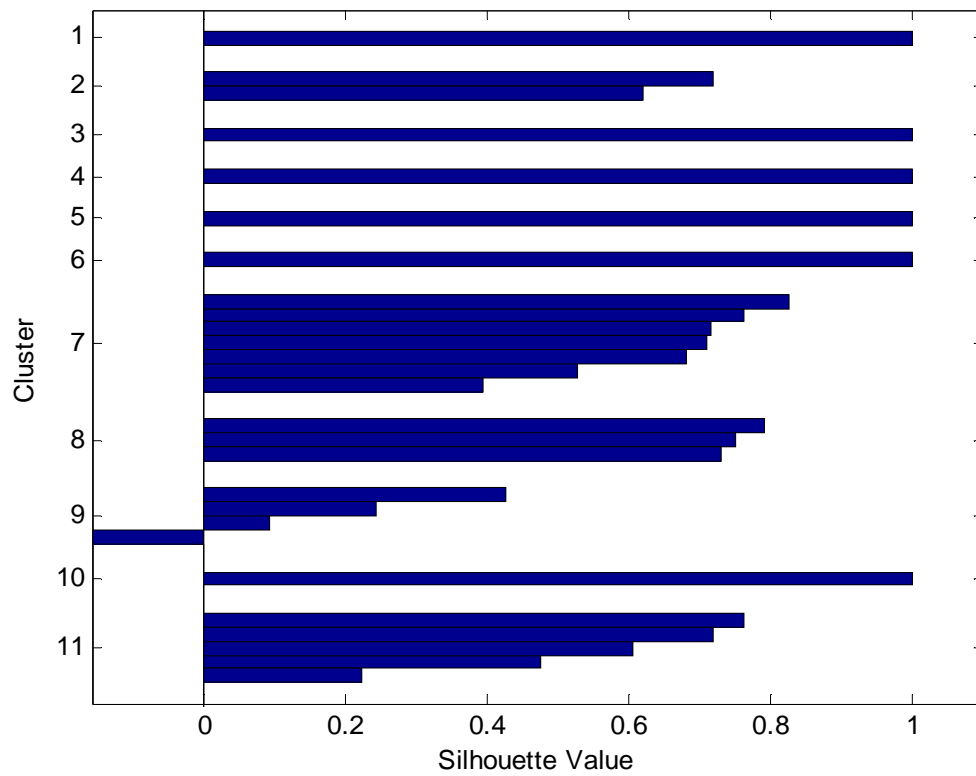
Kmeans στο δείγμα b11725 με 11 ομάδες**5^η δοκιμή**

[idx,c]=kmeans(b11725,11,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	27	754.303
2	1	2	610.157
3	2	8	470.504
3 iterations, total sum of distances = 47.0504			

ΠΙΝΑΚΑΣ 8.2.9: Πέμπτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b11725.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
jo	ro	to	no	e	ta
	ja				
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
pi	u	ko	a	ra	
wa	wo	o		te	
i	we	ka		re	
ti		po		ke	
si				ri	
pa					
me					

ΠΙΝΑΚΑΣ 8.2.10: ομαδοποίηση του δείγματος b11725 σε 11 ομάδες μετά την πέμπτη δοκιμή του kmeans.



Σχήμα 8.2.5: Το διάγραμμα εγγύτητας Silhouette του δείγματος b11725 στην πέμπτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

8.3 Kmeans στο δείγμα b8460 με 11 ομάδες**1^η δοκιμή**

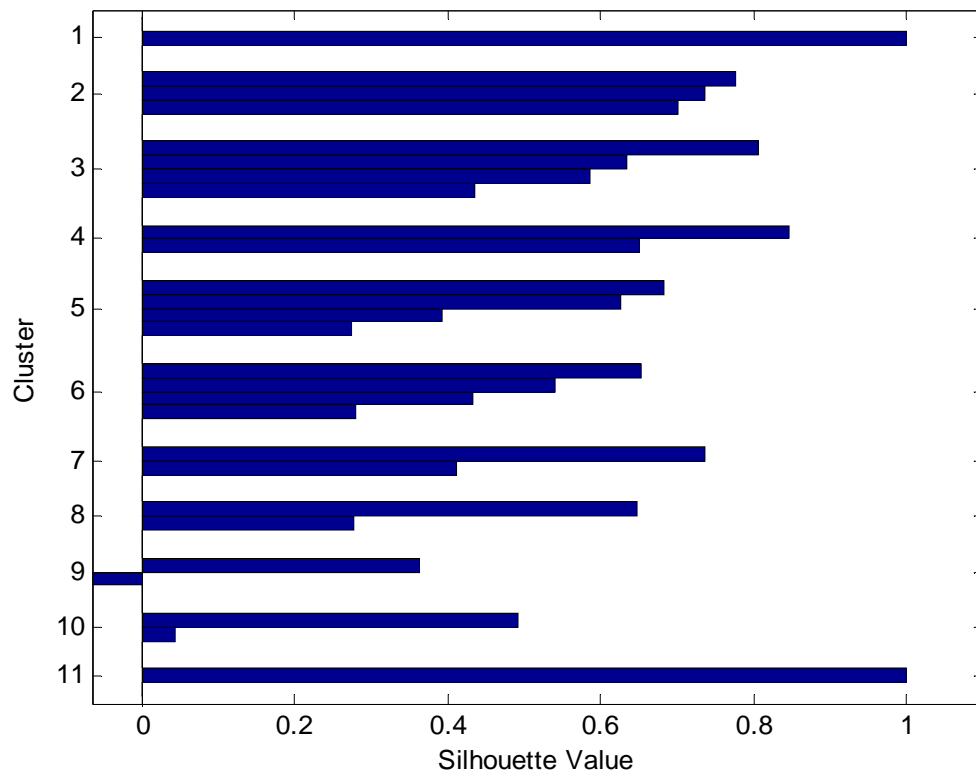
```
[idx,c]=kmeans(b8460,11,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	27	886.032
2	1	1	605.079
3	2	5	405.157
4	2	1	396.851
5	2	1	387.177
5 iterations, total sum of distances = 38.7177			

ΠΙΝΑΚΑΣ 8.3.1: Πρώτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b8460.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
jo	ta	te	e	pi	ka
	u	re	a	wa	me
	wo	ra		i	po
		ke		si	pa
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
we	ko	ri	ro	o	
no	to	ti	ja		

ΠΙΝΑΚΑΣ 8.3.2: ομαδοποίηση του δείγματος b8460 σε 11 ομάδες μετά την πρώτη δοκιμή του kmeans.



Σχήμα 8.3.1: Το διάγραμμα εγγύτητας Silhouette του δείγματος b8460 στην πρώτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

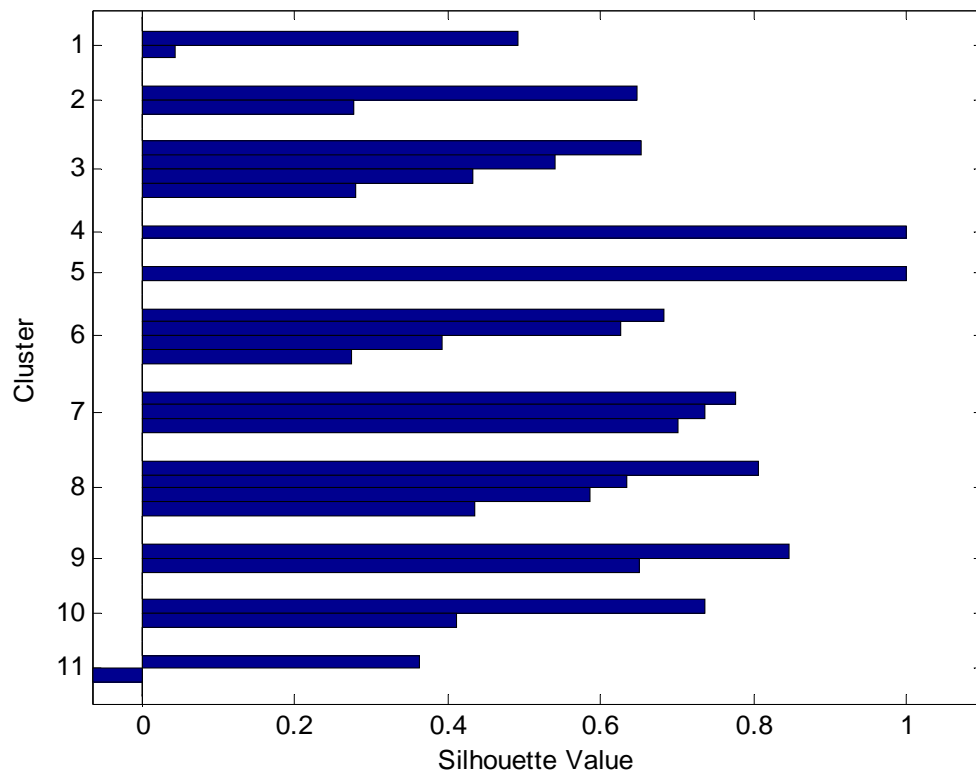
Kmeans στο δείγμα b8460 με 11 ομάδες**2^η δοκιμή**

[idx,c]=kmeans(b8460,11,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	27	834.692
2	1	3	679.337
3	1	2	534.058
4	1	1	513.949
5	2	4	456.746
6	2	2	405.157
7	2	1	396.851
8	2	1	387.177
8 iterations, total sum of distances = 38.7177			

ΠΙΝΑΚΑΣ 8.3.3: Δεύτερη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b8460.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ro	ko	ka	jo	o	pi
ja	to	me			wa
		po			i
		pa			si
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
ta	te	e	we	ri	
u	re	a	no	ti	
wo	ra				
	ke				

ΠΙΝΑΚΑΣ 8.3.4: ομαδοποίηση του δείγματος b8460 σε 11 ομάδες μετά την δεύτερη δοκιμή του kmeans.



Σχήμα 8.3.2: Το διάγραμμα εγγύτητας Silhouette του δείγματος b8460 στην δεύτερη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

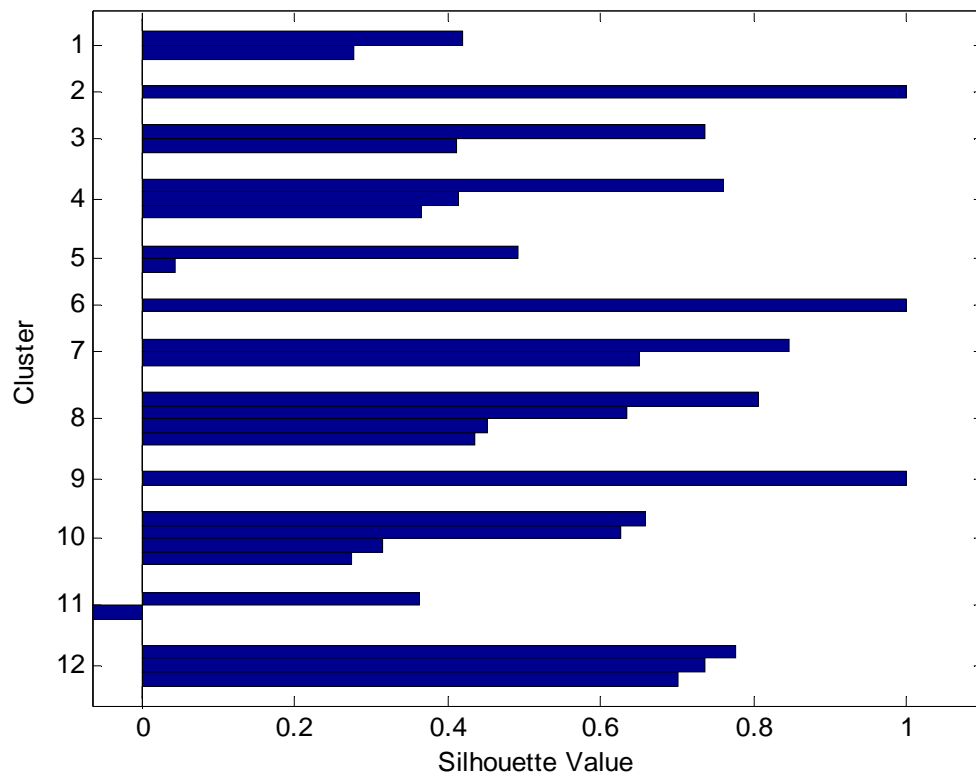
Kmeans στο δείγμα b8460 με 12 ομάδες**3^η δοκιμή**

[idx,c]=kmeans(b8460,12,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	27	166.668
2	1	3	705.794
3	1	1	436.956
4	1	1	417.292
5	2	4	354.996
6	2	2	345.889
6 iterations, total sum of distances = 34.5889			

ΠΙΝΑΚΑΣ 8.3.5: Τρίτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b8460.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ko	o	we	me	ro	jo
to		no	po	ja	
			pa		
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
e	te	ka	pi	ri	ta
a	re		wa	ti	u
	ra		i		wo
	ke		si		

ΠΙΝΑΚΑΣ 8.3.6: ομαδοποίηση του δείγματος b8460 σε 12 ομάδες μετά την τρίτη δοκιμή του kmeans.



Σχήμα 8.3.3: Το διάγραμμα εγγύτητας Silhouette του δείγματος b8460 στην τρίτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα b8460 με 12 ομάδες**4^η δοκιμή**

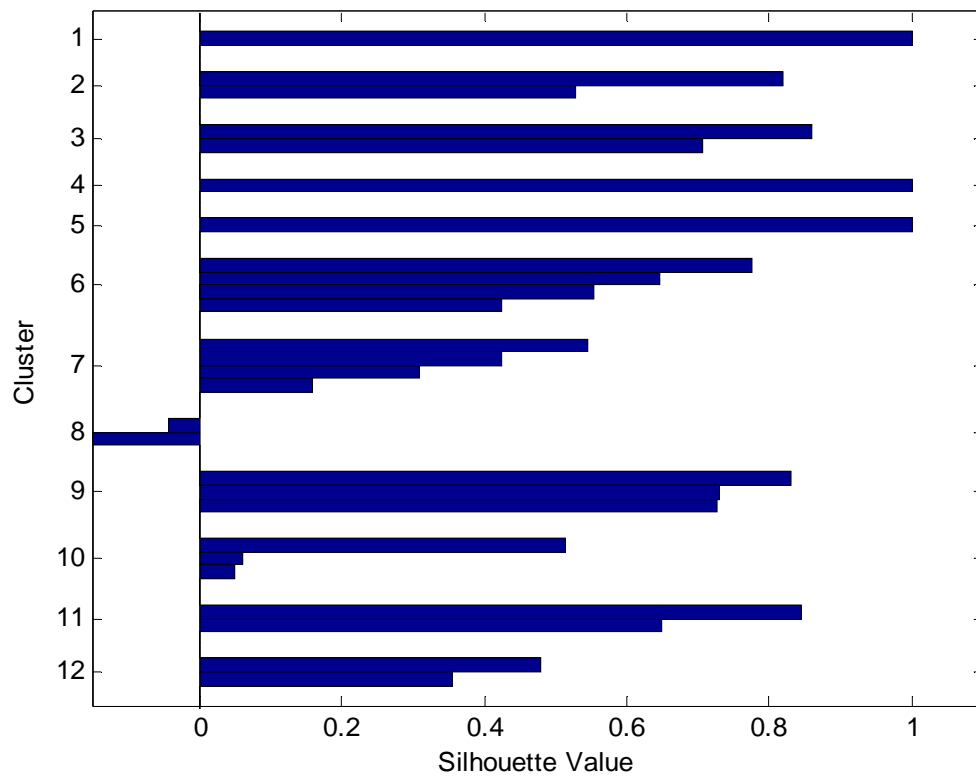
```
[idx,c]=kmeans(b8460,12,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	27	942.777
2	1	1	75.499
3	1	1	505.015
4	2	5	409.327
5	2	2	355.235
5 iterations, total sum of distances = 35.5235			

ΠΙΝΑΚΑΣ 8.3.7: Τετάρτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b8460.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
o	we	re	ri	jo	ro
	no	ra			ja
					ta
					u
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
ka	te	pi	wo	e	ti
me	ke	wa	ko	a	si
po		i	to		
pa					

ΠΙΝΑΚΑΣ 8.3.8: ομαδοποίηση του δείγματος b8460 σε 12 ομάδες μετά την τέταρτη δοκιμή του kmeans.



Σχήμα 8.3.4: Το διάγραμμα εγγύτητας Silhouette του δείγματος b8460 στην τέταρτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα b8460 με 12 ομάδες**5^η δοκιμή**

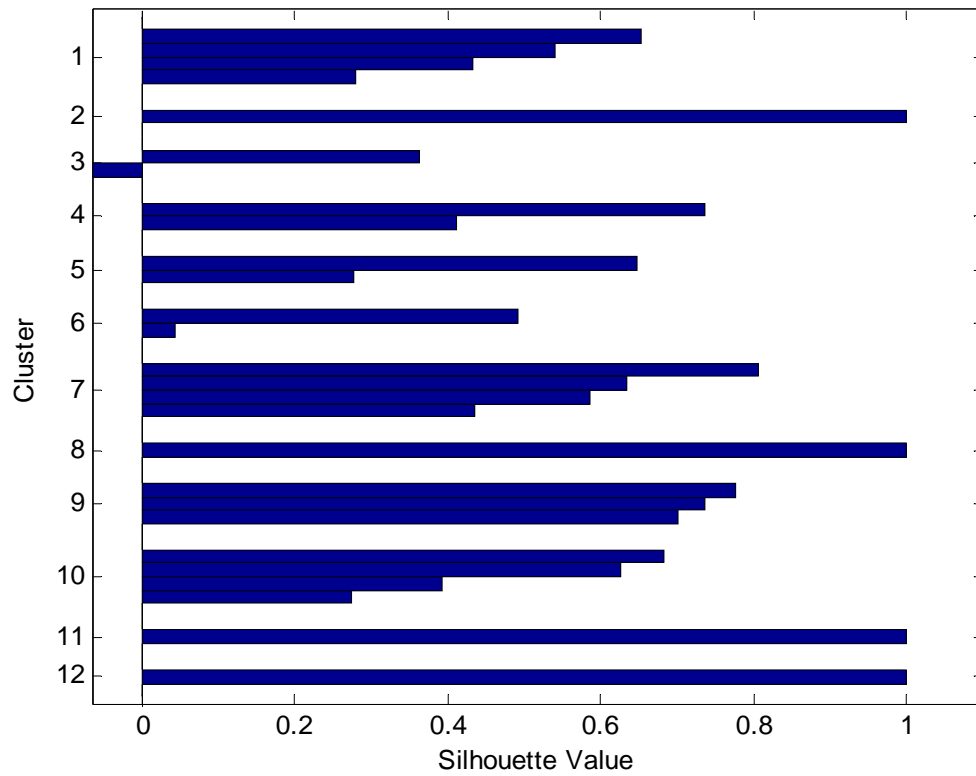
```
[idx,c]=kmeans(b8460,12,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	27	49.073
2	1	3	387.185
3	2	2	349.443
4	2	2	291.396
4 iterations, total sum of distances = 29.1396			

ΠΙΝΑΚΑΣ 8.3.9: Πέμπτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b8460.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ka	jo	ri	we	ko	ro
me		ti	no	to	ja
po					
pa					
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
te	a	ta	pi	e	o
re		u	wa		
ra		wo	i		
ke			si		

ΠΙΝΑΚΑΣ 8.3.10: ομαδοποίηση του δείγματος b8460 σε 12 ομάδες μετά την πέμπτη δοκιμή του kmeans.



Σχήμα 8.3.5: Το διάγραμμα εγγύτητας Silhouette του δείγματος b8460 στην πέμπτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα b8460 με 12 ομάδες**6^η δοκιμή**

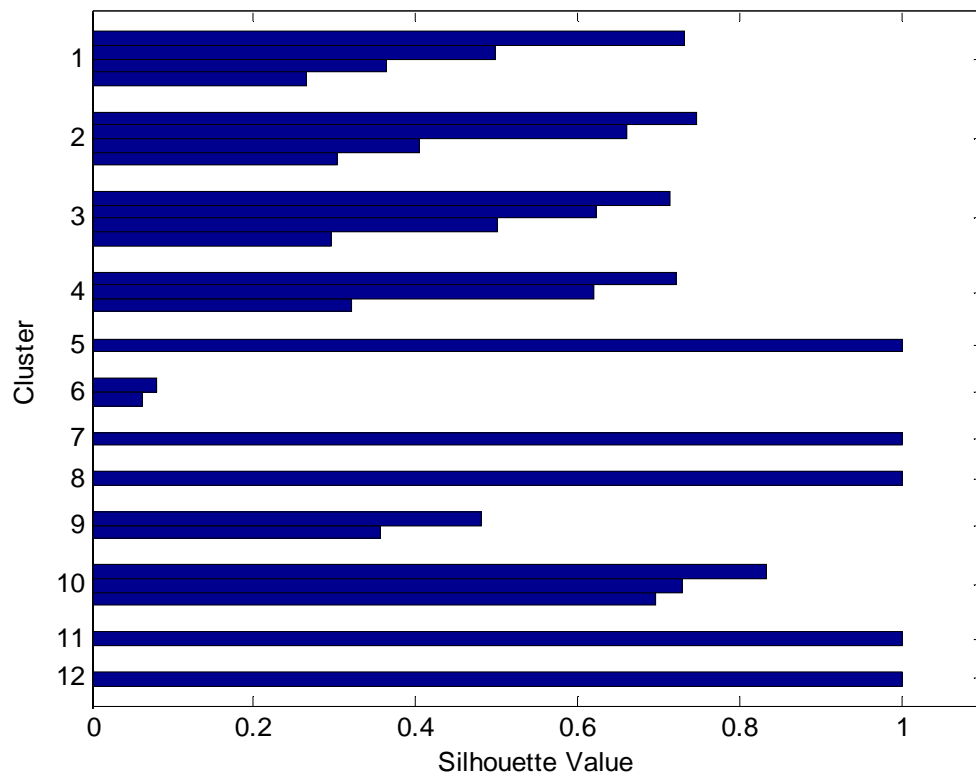
```
[idx,c]=kmeans(b8460,12,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	27	108.028
2	1	1	100.799
3	1	2	819.522
4	1	1	661.607
5	1	1	429.774
6	2	5	310.342
7	2	2	288.715
7 iterations, total sum of distances = 28.8715			

ΠΙΝΑΚΑΣ 8.3.11: Έκτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b8460.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
wo	ro	te	me	jo	ko
to	ja	re	po		ka
we	ta	ra	pa		
no	u	ke			
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
e	ri	ti	pi	o	a
		si	wa		
			i		

ΠΙΝΑΚΑΣ 8.3.12: ομαδοποίηση του δείγματος b8460 σε 12 ομάδες μετά την έκτη δοκιμή του kmeans.



Σχήμα 8.3.6: Το διάγραμμα εγγύτητας Silhouette του δείγματος b8460 στην έκτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

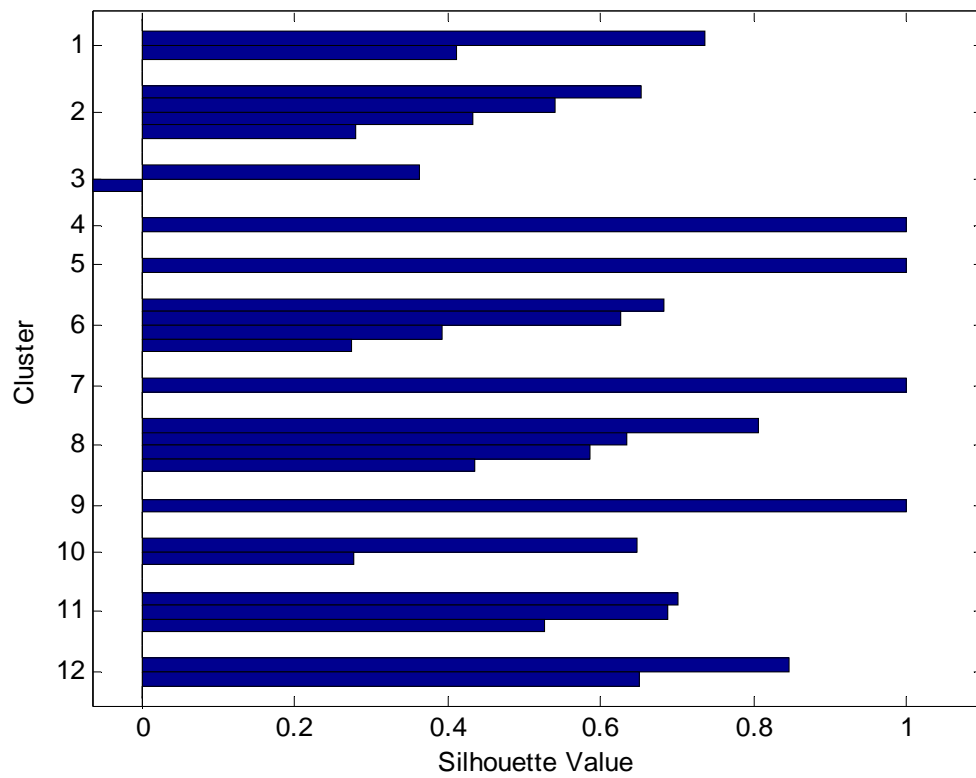
Kmeans στο δείγμα b8460 με 12 ομάδες**7^η δοκιμή**

[idx,c]=kmeans(b8460,12,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	27	168.703
2	1	3	919.168
3	1	2	54.735
4	2	4	42.458
5	2	1	402.924
6	2	2	389.059
7	2	1	364.318
7 iterations, total sum of distances = 36.4318			

ΠΙΝΑΚΑΣ 8.3.13: Έβδομη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b8460.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
we	ka	ri	o	jo	pi
no	me	ti			wa
	po				i
	pa				si
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
ro	te	ja	ko	ta	e
	re		to	u	a
	ra			wo	
	ke				

ΠΙΝΑΚΑΣ 8.3.14: ομαδοποίηση του δείγματος b8460 σε 12 ομάδες μετά την έβδομη δοκιμή του kmeans.



Σχήμα 8.3.7: Το διάγραμμα εγγύτητας Silhouette του δείγματος b8460 στην έβδομη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα b8460 με 12 ομάδες**8^η δοκιμή**

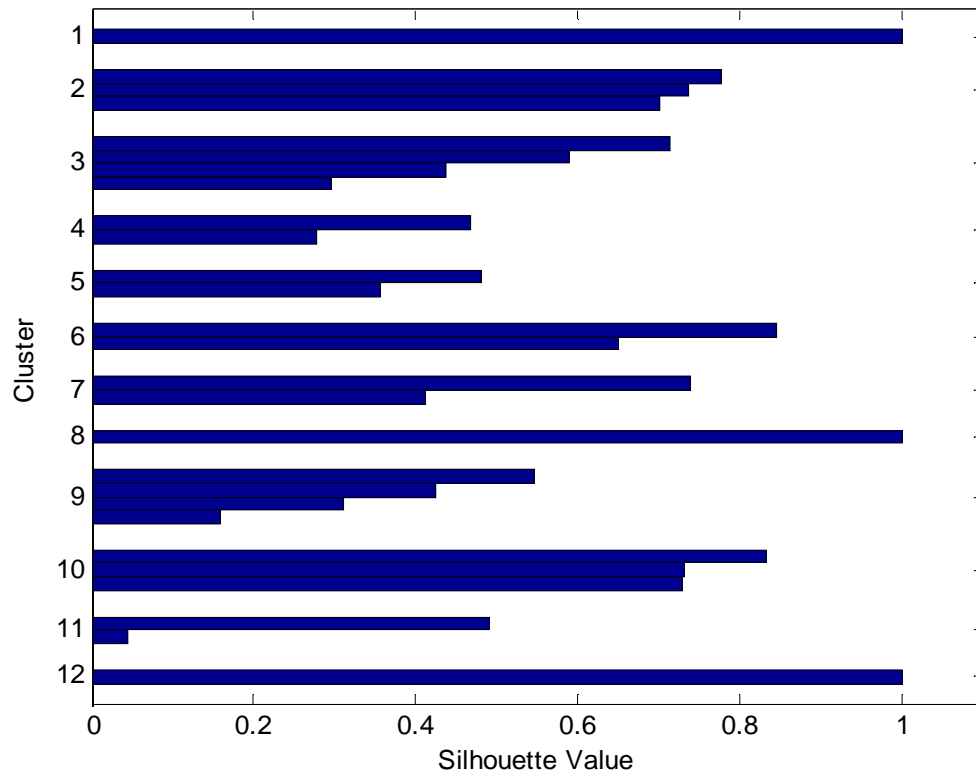
```
[idx,c]=kmeans(b8460,12,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	27	112.327
2	1	1	452.476
3	2	6	345.476
4	2	1	337.741
5	2	1	330.048
5 iterations, total sum of distances = 33.0048			

ΠΙΝΑΚΑΣ 8.3.15: Όγδοη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα b8460.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ri	ta	te	ko	ti	e
	u	re	to	si	a
	wo	ra			
		ke			
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
we	o	ka	pi	ro	jo
no		me	wa	ja	
		po	i		
		pa			

ΠΙΝΑΚΑΣ 8.3.16: ομαδοποίηση του δείγματος b8460 σε 12 ομάδες μετά την όγδοη δοκιμή του kmeans.



Σχήμα 8.3.8: Το διάγραμμα εγγύτητας Silhouette του δείγματος b8460 στην όγδοη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

8.4 Kmeans στο δείγμα α1859 με 10 ομάδες**1^η δοκιμή**

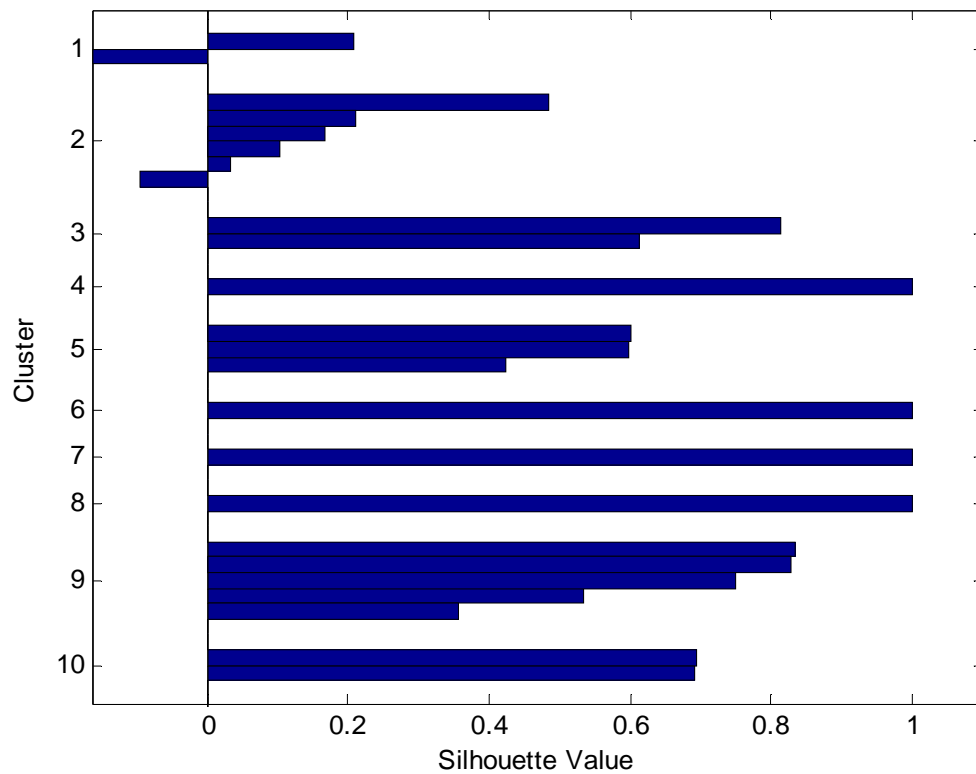
```
[idx,c]=kmeans(a1859,10,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	24	519.342
2	1	2	364.602
3	1	1	311.474
4	2	2	300.178
5	2	1	299.789
5 iterations, total sum of distances = 29.9789			

ΠΙΝΑΚΑΣ 8.4.1: Πρώτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα α1859.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
i	ma	na	a	ra	ja
ka	ki	ti		re	
	si			te	
	ru				
	sa				
	di				
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
ta	pa	ri	da		
		du	ku		
		mi			
		su			
		ni			

ΠΙΝΑΚΑΣ 8.4.2: ομαδοποίηση του δείγματος α1859 σε 10 ομάδες μετά την πρώτη δοκιμή του kmeans.



Σχήμα 8.4.1: Το διάγραμμα εγγύτητας Silhouette του δείγματος a1859 στην πρώτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα α1859 με 10 ομάδες**2^η δοκιμή**

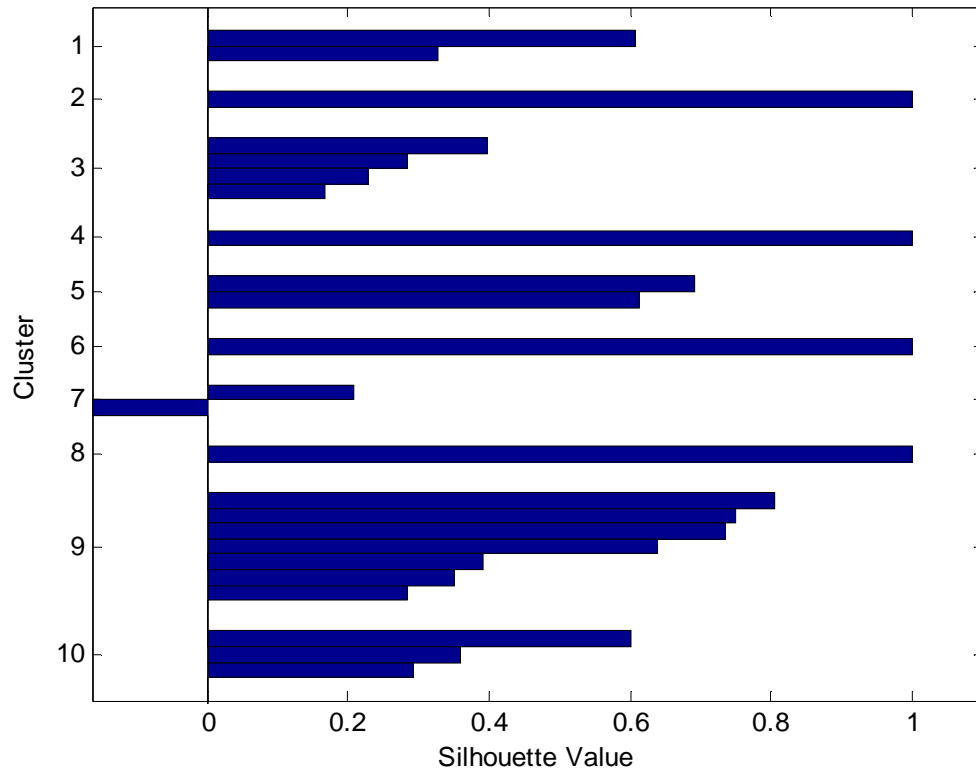
```
[idx,c]=kmeans(a1859,10,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	24	511.262
2	1	2	399.976
3	2	5	339.421
3 iterations, total sum of distances = 33.9421			

ΠΙΝΑΚΑΣ 8.4.3: Δεύτερη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα α1859.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ta	na	ki	ja	da	ti
ma		si		ku	
		sa			
		pa			
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
i	a	ru	ra		
ka		di	re		
		ri	te		
		du			
		mi			
		su			
		ni			

ΠΙΝΑΚΑΣ 8.4.4: ομαδοποίηση του δείγματος α1859 σε 10 ομάδες μετά την δεύτερη δοκιμή του kmeans.



Σχήμα 8.4.2: Το διάγραμμα εγγύτητας Silhouette του δείγματος a1859 στην δεύτερη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα α1859 με 10 ομάδες**3^η δοκιμή**

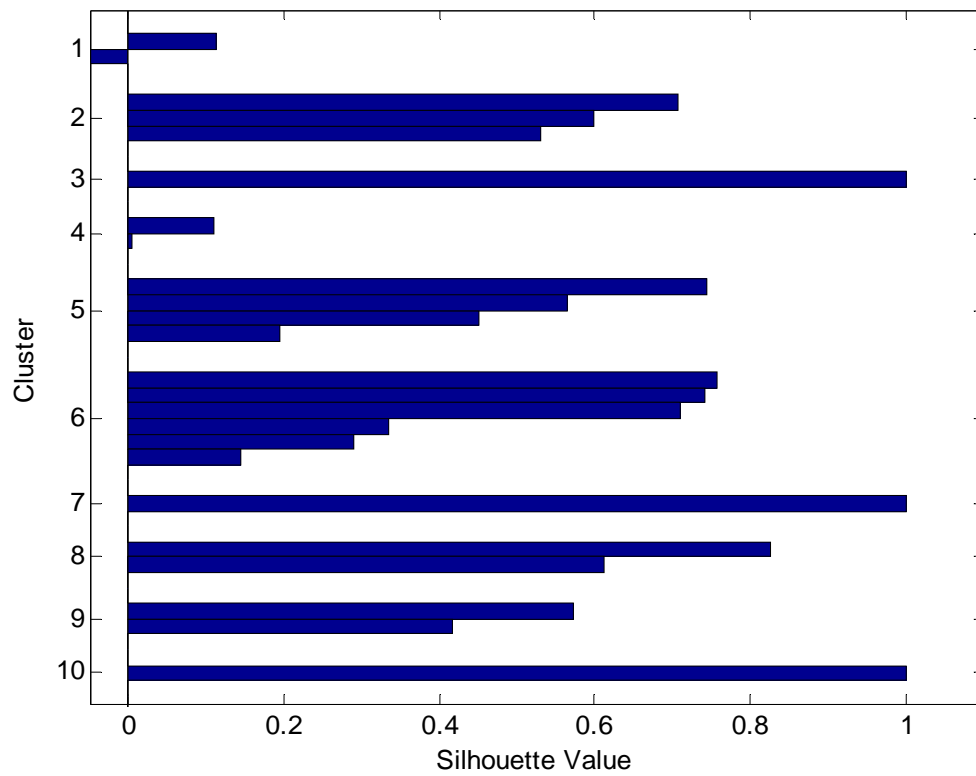
```
[idx,c]=kmeans(a1859,10,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	24	598.523
2	1	2	484.828
3	1	1	476.443
4	2	5	385.443
5	2	1	336.939
5 iterations, total sum of distances = 33.6939			

ΠΙΝΑΚΑΣ 8.4.5: Τρίτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα α1859.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ka	ra	i	ta	ma	ru
si	re		ja	ki	ri
	te			sa	du
				di	mi
					su
					ni
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
a	na	da	pa		
	ti	ku			

ΠΙΝΑΚΑΣ 8.4.6: ομαδοποίηση του δείγματος α1859 σε 10 ομάδες μετά την τρίτη δοκιμή του kmeans.



Σχήμα 8.4.3: Το διάγραμμα εγγύτητας Silhouette του δείγματος a1859 στην τρίτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα α1859 με 10 ομάδες**4^η δοκιμή**

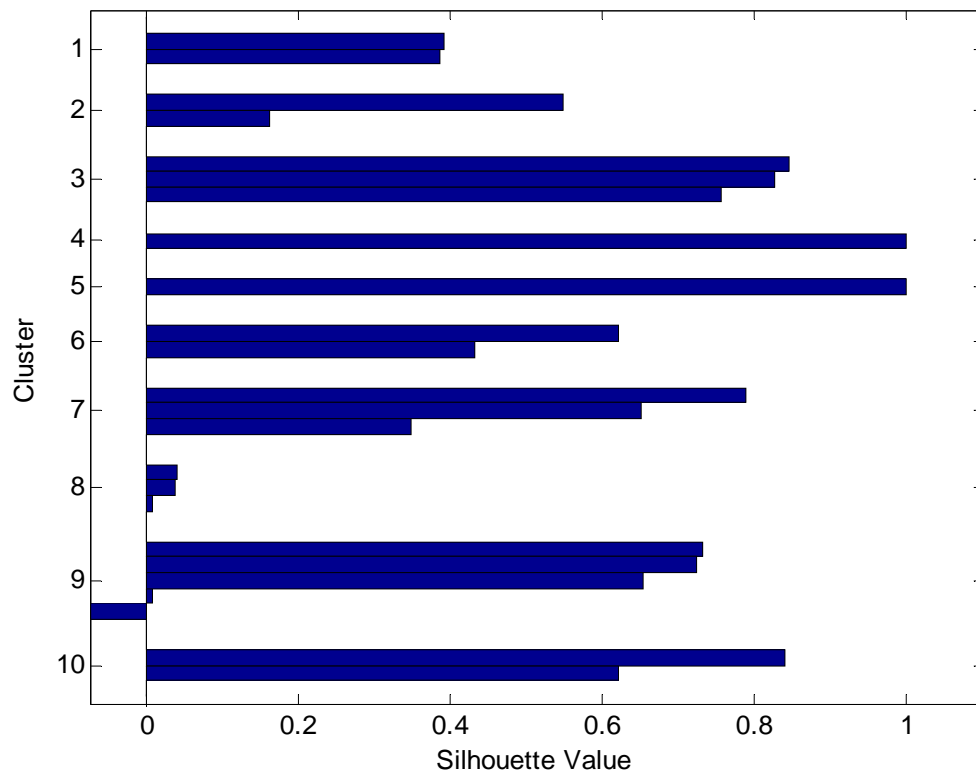
```
[idx,c]=kmeans(a1859,10,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	24	204.007
2	1	3	162.987
3	1	2	116.475
4	1	1	434.872
5	2	5	287.851
5 iterations, total sum of distances = 28.7851			

ΠΙΝΑΚΑΣ 8.4.7: Τέταρτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα α1859.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ra	ta	mi	ja	a	re
ru	ma	su			te
		ni			
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
i	ka	ki	na		
da	si	sa	ti		
ku	pa	di			
		ri			
		du			

ΠΙΝΑΚΑΣ 8.4.8: ομαδοποίηση του δείγματος α1859 σε 10 ομάδες μετά την τέταρτη δοκιμή του kmeans.



Σχήμα 8.4.4: Το διάγραμμα εγγύτητας Silhouette του δείγματος a1859 στην τέταρτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα α1859 με 10 ομάδες**5^η δοκιμή**

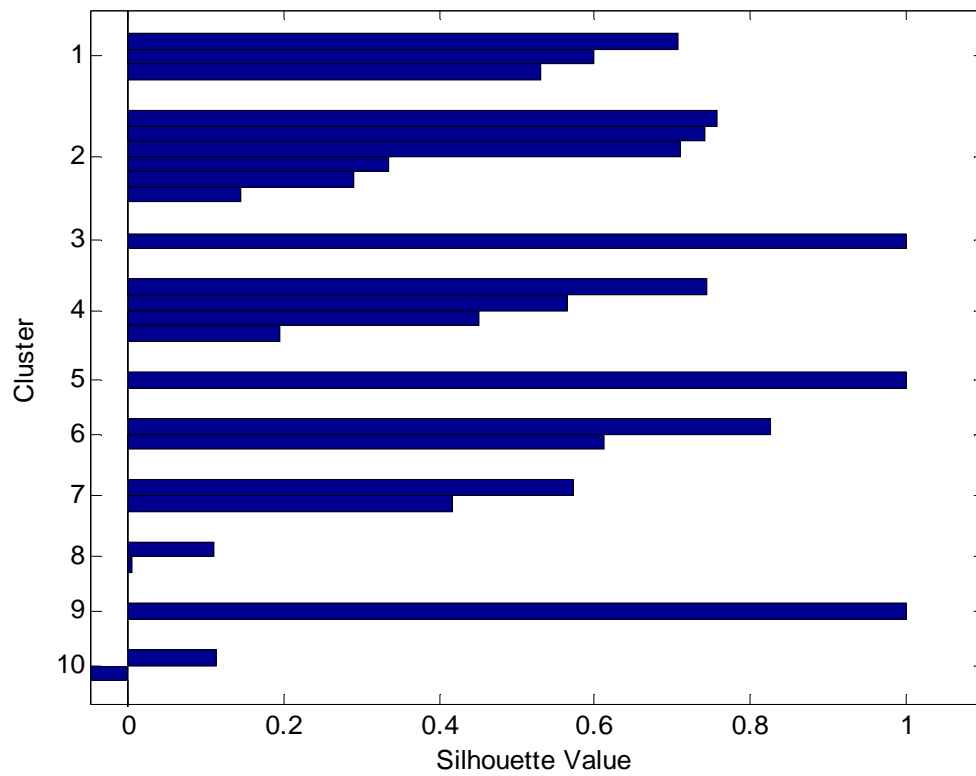
```
[idx,c]=kmeans(a1859,10,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	24	472.858
2	1	2	424.022
3	2	4	33.976
4	2	1	336.939
4 iterations, total sum of distances = 33.6939			

ΠΙΝΑΚΑΣ 8.4.9: Πέμπτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα α1859.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ra	ru	i	ma	pa	na
re	ri		ki		ti
te	du		sa		
	mi		di		
	su				
	ni				
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
da	ta	a	ka		
ku	ja		si		

ΠΙΝΑΚΑΣ 8.4.10: ομαδοποίηση του δείγματος α1859 σε 10 ομάδες μετά την πέμπτη δοκιμή του kmeans.



Σχήμα 8.4.5: Το διάγραμμα εγγύτητας Silhouette του δείγματος a1859 στην πέμπτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα α1859 με 10 ομάδες**6^η δοκιμή**

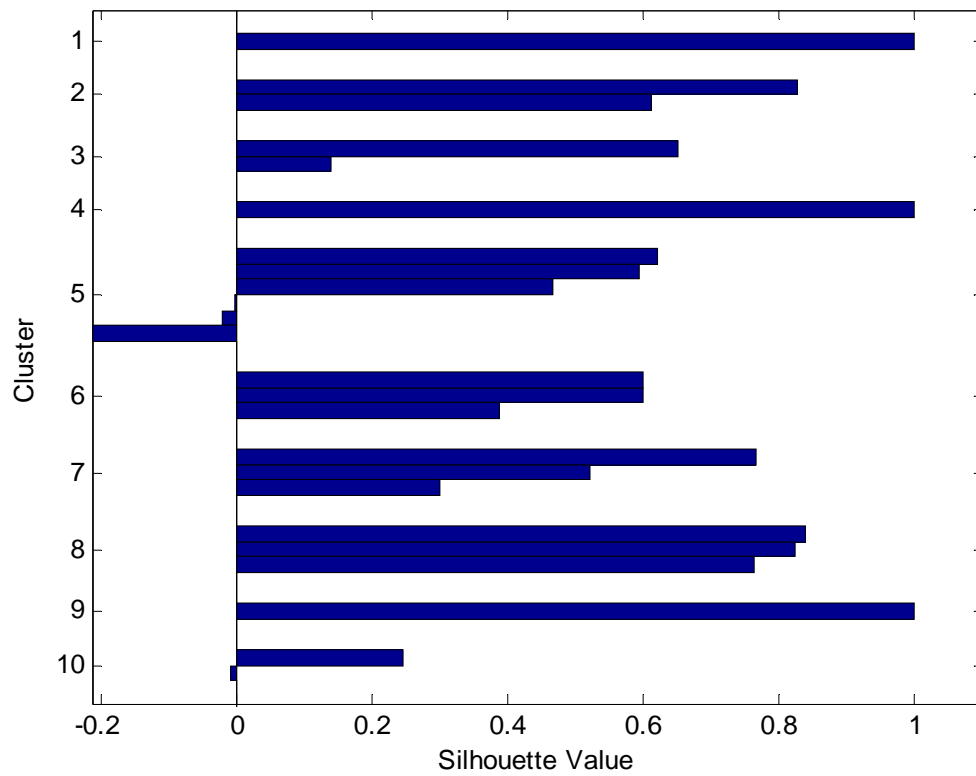
```
[idx,c]=kmeans(a1859,10,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	24	117.802
2	1	1	417.694
3	2	4	37.364
4	2	4	295.922
4 iterations, total sum of distances = 29.5922			

ΠΙΝΑΚΑΣ 8.4.11: Έκτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα α1859.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
pa	na	ta	a	ki	ra
	ti	ma		ru	re
				sa	te
				di	
				ri	
				du	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
i	mi	ja	ka		
da	su		si		
ku	ni				

ΠΙΝΑΚΑΣ 8.4.12: ομαδοποίηση του δείγματος α1859 σε 10 ομάδες μετά την έκτη δοκιμή του kmeans.



Σχήμα 8.4.6: Το διάγραμμα εγγύτητας Silhouette του δείγματος a1859 στην έκτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα α1859 με 10 ομάδες**7^η δοκιμή**

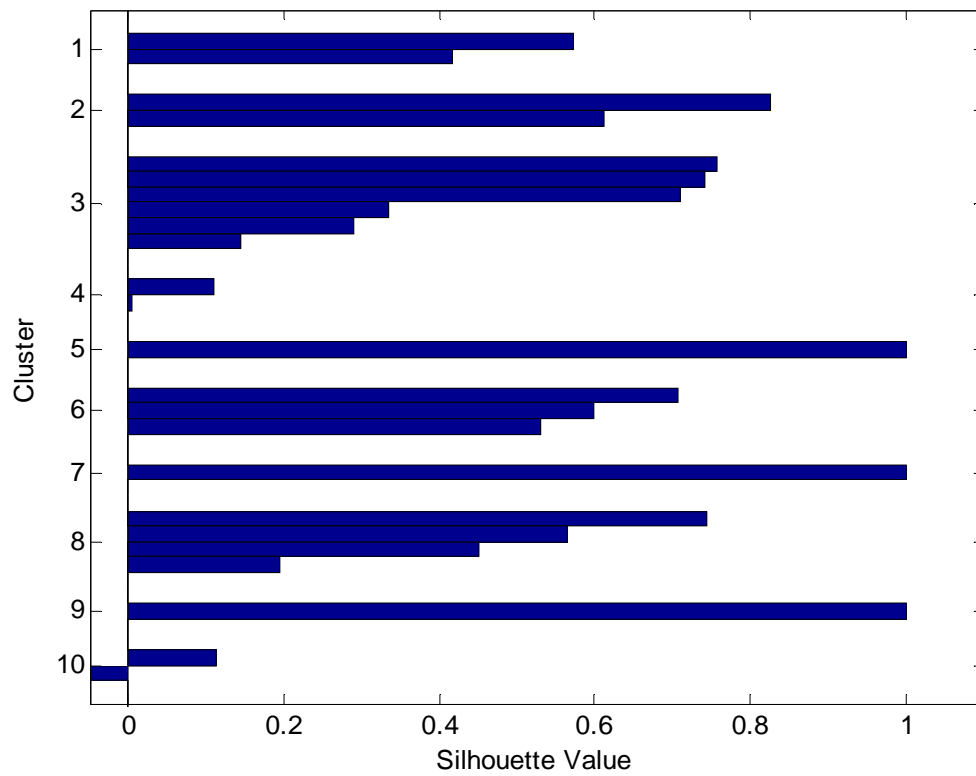
```
[idx,c]=kmeans(a1859,10,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	24	487.807
2	2	6	388.473
3	2	3	336.939
3 iterations, total sum of distances = 33.6939			

ΠΙΝΑΚΑΣ 8.4.13: Έβδομη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα α1859.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
da	na	ru	ta	pa	ra
ku	ti	ri	ja		re
		du			te
		mi			
		su			
		ni			
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
a	ma	i	ka		
	ki		si		
	sa				
	di				

ΠΙΝΑΚΑΣ 8.4.14: ομαδοποίηση του δείγματος α1859 σε 10 ομάδες μετά την έβδομη δοκιμή του kmeans.



Σχήμα 8.4.7: Το διάγραμμα εγγύτητας Silhouette του δείγματος a1859 στην έβδομη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στο δείγμα α1859 με 10 ομάδες**8^η δοκιμή**

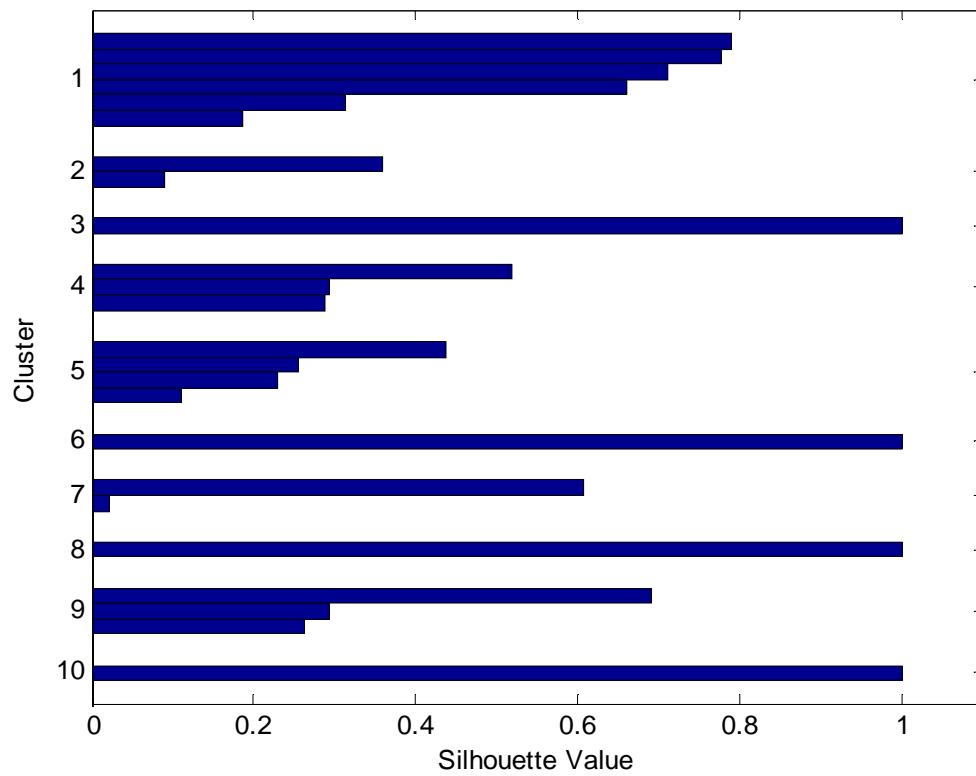
```
[idx,c]=kmeans(a1859,10,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	24	49.263
2	1	1	456.572
3	2	4	344.357
3 iterations, total sum of distances = 34.4357			

ΠΙΝΑΚΑΣ 8.4.15: Όγδοη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στο δείγμα α1859.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
di	si	ti	ra	ku	a
ri	ru		re	ki	
du			te	sa	
mi				pa	
su					
ni					
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
ta	na	i	ja		
ma		da			
		ka			

ΠΙΝΑΚΑΣ 8.4.16: ομαδοποίηση του δείγματος α1859 σε 10 ομάδες μετά την όγδοη δοκιμή του kmeans.



Σχήμα 8.4.8: Το διάγραμμα εγγύτητας Silhouette του δείγματος a1859 στην όγδοη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

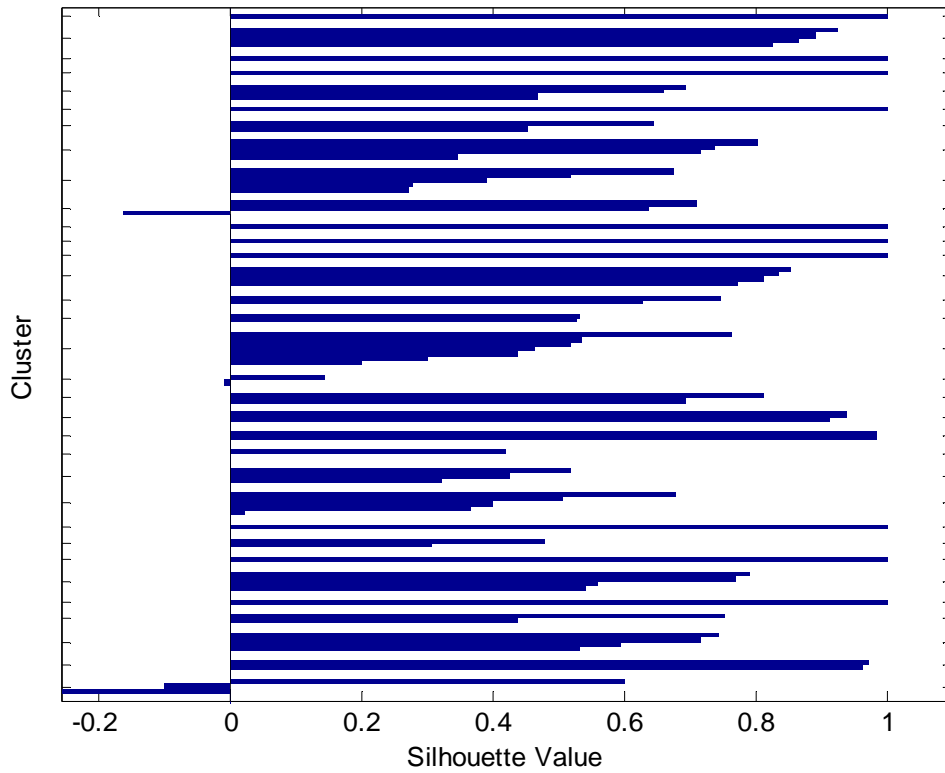
8.5 Kmeans στα δείγματα b3344, b11725 & b8460 με 33 ομάδες**1^η δοκιμή**

[idx,c]=kmeans(bola,33,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	81	149.356
2	1	3	144.832
3	1	1	140.69
4	1	1	133.724
5	2	9	663.303
6	2	2	641.052
7	2	2	638.001
8	2	2	629.716
8 iterations, total sum of distances = 62.9716			

ΠΙΝΑΚΑΣ 8.5.1: Πρώτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στα δείγματα b3344, b11725 & b8460.

1η ομάδα		2η ομάδα		3η ομάδα		4η ομάδα		5η ομάδα		6η ομάδα		7η ομάδα	
ro	b3344	wi	b3344	ri	b3344	da	b3344	jo	b3344	ri	b11725	i	b11725
		i	b3344					jo	b11725			i	b8460
		pa	b11725					jo	b8460				
		pa	b8460										
8η ομάδα		9η ομάδα		10η ομάδα		11η ομάδα		12η ομάδα		13η ομάδα		14η ομάδα	
ro	b11725	ti	b3344	a	b3344	ri	b8460	ko	b3344	re	b11725	pi	b11725
ja	b11725	ti	b11725	a	b11725							wa	b11725
ro	b8460	si	b11725	a	b8460							pi	b8460
ja	b8460	ti	b8460									wa	b8460
		si	b8460										
15η ομάδα		16η ομάδα		17η ομάδα		18η ομάδα		19η ομάδα		20η ομάδα		21η ομάδα	
te	b11725	na	b3344	te	b3344	ta	b11725	ko	b11725	o	b11725	ka	b11725
te	b8460	re	b3344	ke	b3344	to	b11725	ko	b8460	o	b8460	ka	b8460
				si	b3344								
				ma	b3344								
				sa	b3344								
				me	b11725								
				me	b8460								
22η ομάδα		23η ομάδα		24η ομάδα		25η ομάδα		26η ομάδα		27η ομάδα		28η ομάδα	
ta	b3344	so	b3344	u	b3344	wa	b3344	no	b3344	ka	b3344	ra	b3344
to	b3344	no	b11725	u	b11725			wo	b3344			ra	b11725
		no	b8460	ta	b8460							re	b8460
				u	b8460							ra	b8460
				wo	b8460								
29η ομάδα		30η ομάδα		31η ομάδα		32η ομάδα		33η ομάδα					
ja	b3344	ke	b11725	wo	b11725	e	b11725	e	b3344				
		ke	b8460	we	b11725	e	b8460	po	b11725				
				to	b8460			po	b8460				
				we	b8460								

ΠΙΝΑΚΑΣ 8.5.2: ομαδοποίηση των δειγμάτων b3344, b11725 & b8460 σε 33 ομάδες μετά την πρώτη δοκιμή του kmeans.



Σχήμα 8.5.1: Το διάγραμμα εγγύτητας Silhouette των δειγμάτων b3344, b11725 & b8460 στην πρώτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

Kmeans στα δείγματα b3344, b11725 & b8460 με 33 ομάδες**2^η δοκιμή**

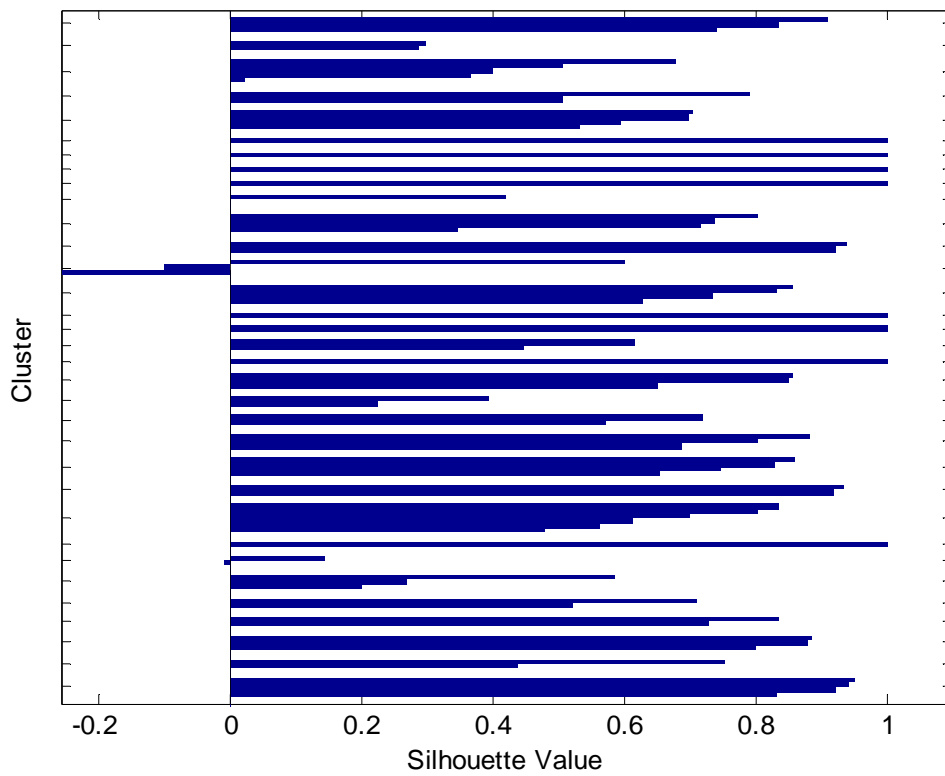
```
[idx,c]=kmeans(bola,33,'distance','sqeuclidean','display','iter')
```

iter	phase	num	sum
1	1	81	75.177
2	1	7	65.671
3	1	2	642.212
4	1	2	619.037
5	2	14	438.393
6	2	3	386.661
6 iterations, total sum of distances = 38.6661			

ΠΙΝΑΚΑΣ 8.5.3: Δεύτερη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στα δείγματα b3344, b11725 & b8460.

1η ομάδα		2η ομάδα		3η ομάδα		4η ομάδα		5η ομάδα		6η ομάδα		7η ομάδα	
ko	b3344	no	b3344	u	b3344	si	b11725	wo	b11725	ra	b3344	ka	b3344
ko	b11725	wo	b3344	u	b11725	si	b8460	we	b11725				
ko	b8460			ta	b8460			to	b8460				
				u	b8460			we	b8460				
				wo	b8460								
8η ομάδα		9η ομάδα		10η ομάδα		11η ομάδα		12η ομάδα		13η ομάδα		14η ομάδα	
wa	b3344	a	b11725	ta	b3344	ro	b11725	o	b11725	e	b3344	ma	b3344
				to	b3344	ja	b11725	o	b8460	po	b11725	sa	b3344
						ro	b8460			po	b8460	me	b11725
						ja	b8460					me	b8460
15η ομάδα		16η ομάδα		17η ομάδα		18η ομάδα		19η ομάδα		20η ομάδα		21η ομάδα	
ja	b3344	so	b3344	na	b3344	a	b3344	da	b3344	re	b11725	ri	b3344
				re	b3344			ka	b11725	ri	b8460	ri	b11725
								ka	b8460				
22η ομάδα		23η ομάδα		24η ομάδα		25η ομάδα		26η ομάδα		27η ομάδα		28η ομάδα	
te	b3344	wi	b3344	e	b11725	pi	b11725	a	b8460	ta	b11725	ra	b11725
ke	b3344	i	b3344	e	b8460	wa	b11725			to	b11725	re	b8460
si	b3344	pa	b11725			i	b11725					ra	b8460
		pa	b8460			pi	b8460						
						wa	b8460						
						i	b8460						
29η ομάδα		30η ομάδα		31η ομάδα		32η ομάδα		33η ομάδα					
te	b11725	no	b11725	ti	b3344	ke	b11725	ro	b3344				
te	b8460	no	b8460	ti	b11725	ke	b8460	jo	b3344				
				ti	b8460			jo	b11725				
								jo	b8460				

ΠΙΝΑΚΑΣ 8.5.4: ομαδοποίηση των δειγμάτων b3344, b11725 & b8460 σε 33 ομάδες μετά την δεύτερη δοκιμή του kmeans.



Σχήμα 8.5.2: Το διάγραμμα εγγύτητας Silhouette των δειγμάτων b3344, b11725 & b8460 στην δεύτερη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

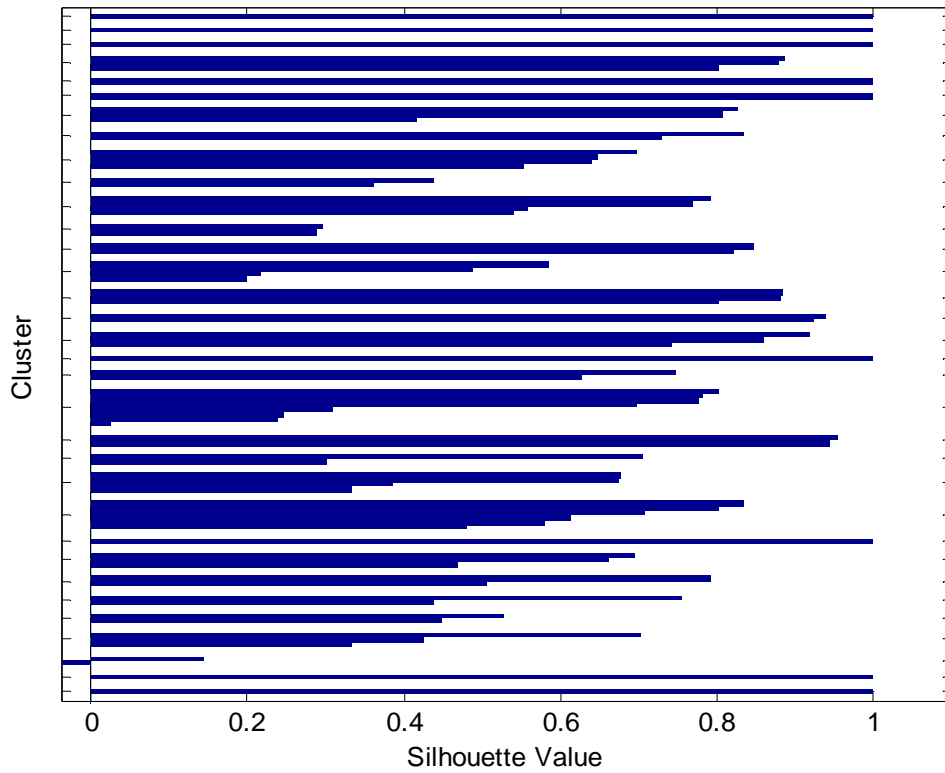
Kmeans στα δείγματα b3344, b11725 & b8460 με 33 ομάδες**3^η δοκιμή**

[idx,c]=kmeans(bola,33,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	81	931.734
2	1	7	820.796
3	1	1	797.235
4	2	16	644.498
5	2	10	333.778
5 iterations, total sum of distances = 33.3778			

ΠΙΝΑΚΑΣ 8.5.5: Τρίτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στα δείγματα b3344, b11725 & b8460.

1η ομάδα		2η ομάδα		3η ομάδα		4η ομάδα		5η ομάδα		6η ομάδα		7η ομάδα	
a	b3344	to	b3344	wa	b3344	ti	b3344	ro	b3344	ri	b8460	ro	b11725
						ti	b11725					ja	b11725
						ti	b8460					ja	b8460
8η ομάδα		9η ομάδα		10η ομάδα		11η ομάδα		12η ομάδα		13η ομάδα		14η ομάδα	
no	b11725	wo	b11725	u	b11725	ra	b3344	no	b3344	a	b11725	u	b3344
no	b8460	we	b11725	u	b8460	ra	b11725	wo	b3344	a	b8460	to	b11725
		to	b8460			re	b8460					ta	b8460
		we	b8460			ra	b8460					wo	b8460
15η ομάδα		16η ομάδα		17η ομάδα		18η ομάδα		19η ομάδα		20η ομάδα		21η ομάδα	
te	b3344	o	b11725	ko	b3344	re	b11725	te	b11725	ma	b3344	e	b11725
ke	b3344	o	b8460	ko	b11725			te	b8460	sa	b3344	e	b8460
si	b3344			ko	b8460					wi	b3344		
										i	b3344		
										pa	b11725		
										me	b11725		
										me	b8460		
										pa	b8460		
22η ομάδα		23η ομάδα		24η ομάδα		25η ομάδα		26η ομάδα		27η ομάδα		28η ομάδα	
ri	b3344	ka	b3344	pi	b11725	so	b3344	jo	b3344	si	b11725	ke	b11725
ri	b11725	da	b3344	wa	b11725			jo	b11725	si	b8460	ke	b8460
		ka	b11725	i	b11725			jo	b8460				
		ka	b8460	pi	b8460								
				wa	b8460								
				i	b8460								
29η ομάδα		30η ομάδα		31η ομάδα		32η ομάδα		33η ομάδα					
na	b3344	e	b3344	ta	b11725	ta	b3344	ja	b3344				
re	b3344	po	b11725	ro	b8460								
		po	b8460										

ΠΙΝΑΚΑΣ 8.5.6: ομαδοποίηση των δειγμάτων b3344, b11725 & b8460 σε 33 ομάδες μετά την τρίτη δοκιμή του kmeans.



Σχήμα 8.5.3: Το διάγραμμα εγγύτητας Silhouette των δειγμάτων b3344, b11725 & b8460 στην τρίτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

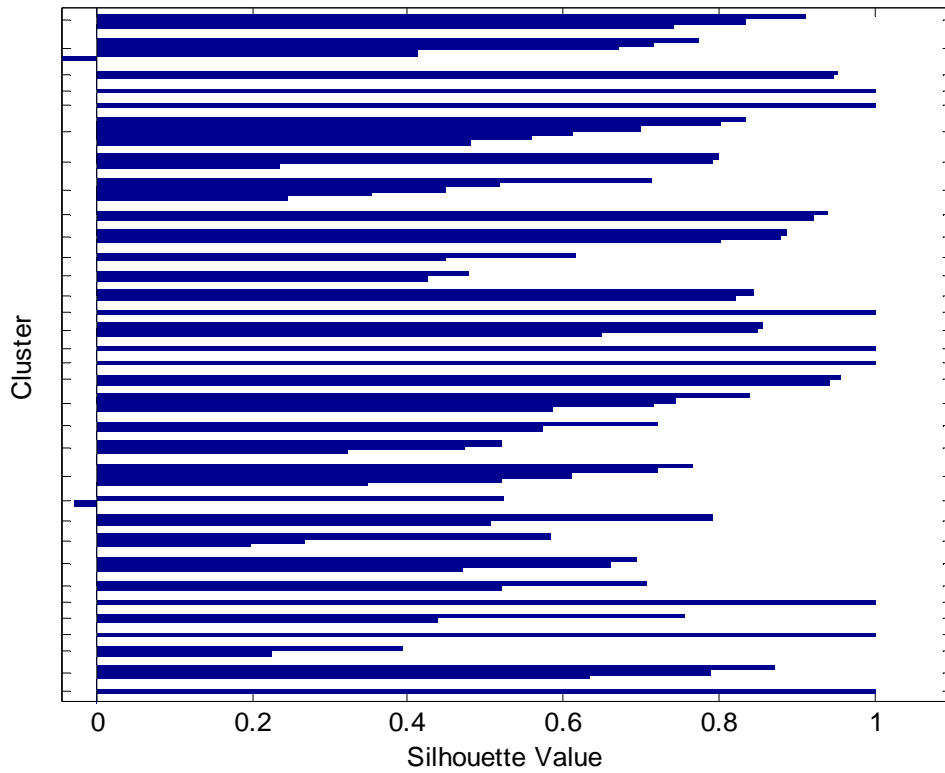
Kmeans στα δείγματα b3344, b11725 & b8460 με 33 ομάδες**4^η δοκιμή**

[idx,c]=kmeans(bola,33,'distance','sqeuclidean','display','iter')			
iter	phase	num	sum
1	1	81	976.996
2	1	7	710.715
3	1	4	638.615
4	1	1	635.167
5	2	12	554.969
6	2	4	53.935
7	2	8	483.731
8	2	3	383.861
8 iterations, total sum of distances = 38.3861			

ΠΙΝΑΚΑΣ 8.5.7: Τέταρτη δοκιμή στην προσπάθεια ελαχιστοποίησης του συνολικού αθροίσματος (sum) των αποστάσεων των προτύπων από τα κέντρα των ομάδων τους στα δείγματα b3344, b11725 & b8460.

1η ομάδα		2η ομάδα		3η ομάδα		4η ομάδα		5η ομάδα		6η ομάδα		7η ομάδα	
ko	b3344	ta	b11725	po	b11725	wa	b3344	ja	b3344	pi	b11725	ma	b3344
ko	b11725	ro	b11725	po	b8460					wa	b11725	sa	b3344
ko	b8460	ja	b11725							i	b11725	me	b8460
		ro	b8460							pi	b8460		
		ja	b8460							wa	b8460		
										i	b8460		
8η ομάδα		9η ομάδα		10η ομάδα		11η ομάδα		12η ομάδα		13η ομάδα		14η ομάδα	
u	b3344	o	b11725	ti	b3344	na	b3344	no	b3344	a	b11725	me	b11725
to	b11725	o	b8460	ti	b11725	re	b3344	wo	b3344	a	b8460		
ta	b8460			ti	b8460								
u	b8460												
wo	b8460												
15η ομάδα		16η ομάδα		17η ομάδα		18η ομάδα		19η ομάδα		20η ομάδα		21η ομάδα	
da	b3344	ro	b3344	e	b3344	e	b11725	wi	b3344	ri	b3344	so	b3344
ka	b11725					e	b8460	i	b3344	ri	b11725	no	b11725
ka	b8460							pa	b11725			no	b8460
								pa	b8460				
22η ομάδα		23η ομάδα		24η ομάδα		25η ομάδα		26η ομάδα		27η ομάδα		28η ομάδα	
u	b11725	ta	b3344	si	b11725	ra	b11725	jo	b3344	te	b11725	a	b3344
wo	b11725	to	b3344	si	b8460	re	b8460	jo	b11725	te	b8460		
we	b11725					ra	b8460	jo	b8460				
to	b8460												
we	b8460												
29η ομάδα		30η ομάδα		31η ομάδα		32η ομάδα		33η ομάδα					
ke	b11725	ra	b3344	re	b11725	te	b3344	ka	b3344				
ke	b8460			ri	b8460	ke	b3344						
						si	b3344						

ΠΙΝΑΚΑΣ 8.5.8: ομαδοποίηση των δειγμάτων b3344, b11725 & b8460 σε 33 ομάδες μετά την τέταρτη δοκιμή του kmeans.



Σχήμα 8.5.4: Το διάγραμμα εγγύτητας Silhouette των δειγμάτων b3344, b11725 & b8460 στην τέταρτη δοκιμή του kmeans που δείχνει πόσο όμοιο είναι ένα πρότυπο μίας ομάδας σε σχέση με πρότυπα άλλων ομάδων.

ΚΕΦΑΛΑΙΟ 9

Ταξινόμηση

9.1 Σύνολα εκπαίδευσης και ελέγχου

Ένα άλλο ζήτημα που θα αντιμετωπιστεί εδώ είναι το ζήτημα του χωρισμού των δεδομένων σε δύο υποσύνολα, το σύνολο εκπαίδευσης και το σύνολο ελέγχου. Κάποια από τα δεδομένα (όσο το δυνατόν περισσότερα) θα χρησιμοποιηθούν για την διαδικασία της εκπαίδευσης (Training set) και κάποια για την διαδικασία του υπολογισμού της απόδοσης (test set) δηλαδή ο υπολογισμός του πόσο καλός είναι ο ταξινομητής για τις διάφορες κατηγορίες. Εάν χρησιμοποιηθούν όλα τα διαθέσιμα δεδομένα για την εκπαίδευση τότε α) δεν έχει νόημα ο υπολογισμός της απόδοσης, και β) η πραγματική απόδοση είναι μικρότερη. Ο λόγος για την συμπεριφορά αυτή είναι ότι τα δεδομένα που υπάρχουν (κατά 99.99%) είναι προϊόν μιας διαδικασίας δειγματοληψίας και δεν είναι το σύνολο των δεδομένων ενός προβλήματος. Καλό είναι να υπάρχουν ισόποσο πλήθος δεδομένων από κάθε κατηγορία. Έτσι αν σε ένα ιατρικό πρόβλημα υπάρχουν 95% υγιείς και 5% ασθενείς, τότε επί συνόλου 100 δεδομένων θα πρέπει να υπάρχουν γύρω στους 50 ασθενείς (και όχι 5) για την εκπαίδευση. Με το δεδομένο ότι τα δεδομένα συχνά δεν υπάρχουν σε επαρκείς ποσότητες ο διαχωρισμός αυτός μπορεί να δυσκολεύει πολύ την κατάσταση.

Η συνάρτηση classify

Η συνάρτηση `classify` κατασκευάζει ταξινομητές χρησιμοποιώντας διάφορους αλγόριθμους διαχωρισμού. Η γενική μορφή της συνάρτησης είναι:

classify

```
class = classify(test,training,group)
```

```
[class,err] = classify(...)
```

```
[...] = classify(...,'type')
```

```
[...] = classify(...,'type',prior)
```

Στην βασική της μορφή **class = classify(test,training,group)** ταξινομεί κάθε σειρά του πίνακα **test** έχοντας εκπαιδευτεί πρώτα από τα δεδομένα του πίνακα **training**. Ο πίνακας **group** δείχνει σε ποια ομάδα ανήκει κάθε δείγμα του πίνακα **training**. Οι πίνακες **test** και **training** πρέπει να περιέχουν τον ίδιο αριθμό στηλών, ενώ οι πίνακες **training** και **group** τον ίδιο αριθμό γραμμών. Ο πίνακας **class** περιέχει την ομάδα (λαμβάνοντας υπ' όψη τον πίνακα `group`) που ταξινομείται κάθε γραμμή του πίνακα `sample`.

Στην μορφή `[class,err] = classify(...)` η συνάρτηση επιστέφει επιπλέον το ποσοστό των λανθασμένων ταξινομήσεων. Η παράμετρος `'type'` καθορίζει τον αλγόριθμο διαχωρισμού. Η παράμετρος `prior` ορίζει τις πιθανότητες εμφάνισης κάθε ομάδας στην πραγματικότητα¹ (`prior` πιθανότητα) που μπορεί να διαφέρει από την κατανομή των ομάδων στον πίνακα `training`.

Η συνάρτηση `classify` χρησιμοποιεί την πληροφορία του πίνακα `group` και γνωρίζει από τα πριν την ομάδα κάθε προτύπου. Άρα είναι μια μορφή μάθησης με επίβλεψη. 1 δηλαδή τις πιθανότητες που συναντούμε αυτά τα είδη λουλουδιών στη φύση. 3.3 14. `mb = (meas - repmat(mean(meas),size(meas),1),1) / repmat(std(meas),size(meas),1)` Υπάρχει και έτοιμη συνάρτηση η οποία εκτελεί την ίδια εργασία: 15. `mb = zscore(meas) classify class = classify(sample,training,group) [class,err] = classify(...) [...] = classify(...,'type') [...] = classify(...,'type',prior)`.

Στο στάδιο αυτό θα προσπαθήσουμε να εκπαιδέσουμε το σύστημα μας στο δείγμα `Bola` (δηλαδή και στα τρία δείγματα της γραμμικής `B` έτσι ώστε όταν εισάγουμε το δείγμα της γραμμικής `A` να μας δώσει αποτελέσματα για το ποια πρότυπα της `A` ταιριάζουν με ποια της `B`. Δηλαδή εκπαιδεύουμε το σύστημα μας με τα δείγματα της γραμμικής `B` και το 'τεστάρουμε' με το μοναδικό δείγμα της γραμμικής `A`. Βέβαια εμείς δεν μπορούμε να διαπιστώσουμε πόσο καλά δούλεψε το σύστημα μας.

```
Bola(b3344,b11725 και b8460) =training
```

```
A1859 =test
```

Οι 33 ομάδες που χωρίσαμε όλα τα δείγματα της γραμμικής `B` με τη χρήση της ιεραρχικής ομαδοποίησης = `group`

```
Εντολή ταξινόμησης στο matlab= classify(test,training,group)
```

class=classify(TEST, TRAINING, GROUP)	
class =	
'ff'	a A1859
'bb'	na A1859
'ee'	i A1859
'cc'	ta A1859
'bb'	ti A1859
'i'	ja A1859
'ee'	da A1859
'cc'	ma A1859
'gg'	ka A1859
'ee'	ku A1859
'w'	ra A1859
'g'	re A1859
'o'	ki A1859
'p'	si A1859
'i'	te A1859
'j'	ru A1859
'o'	sa A1859
'o'	di A1859
'k'	ri A1859
'e'	pa A1859
'c'	du A1859
'k'	mi A1859
'h'	su A1859
'h'	ni A1859

ΠΙΝΑΚΑΣ 9.1: Η ταξινόμηση που πραγματοποίησε το matlab στη γραμμική A μετά την εφαρμογή της εντολής classify.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα	7η ομάδα
te b3344	u b3344	pi b11725	ti b3344	po b11725	ri b3344	ja b11725
ke b3344	wo b8460	wa b11725	ti b11725	po b8460	ri b11725	ja b8460
si b3344		pi b8460	ti b8460			
		wa b8460				
a	b	c	d	e	F	g
		du A1859		pa A1859		re A1859
8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα	13η ομάδα	14η ομάδα
si b11725	to b11725	u b11725	ja b3344	ra b3344	no b3344	ta b11725
si b8460		wo b11725		ra b11725	wo b3344	
		we b11725		re b8460		
		to b8460		ra b8460		
h	i	we b8460	k	l	M	n
		j				
su A1859	ja A1859	ru A1859	ri A1859			
ni A1859	te A1859		mi A1859			
15η ομάδα	16η ομάδα	17η ομάδα	18η ομάδα	19η ομάδα	20η ομάδα	21η ομάδα
wa b3344	ko b3344	to b3344	e b11725	ta b3344	a b3344	ro b3344
ke b11725	ko b11725		e b8460			jo b3344
ke b8460	ko b8460					jo b11725
						jo b8460
o	p	q	r	s	T	u
ki A1859	si A1859		ra A1859			
sa A1859						
di A1859						
22η ομάδα	23η ομάδα	24η ομάδα	25η ομάδα	26η ομάδα	27η ομάδα	28η ομάδα
ma b3344	ta b8460	i b11725	na b3344	wi b3344	re b11725	ro b11725
sa b3344	u b8460	i b8460	re b3344	i b3344	ri b8460	ro b8460
me b11725				pa b11725		
me b8460				pa b8460		
v	w	x	y	z	Aa	bb
						na A1859
						ti A1859
29η ομάδα	30η ομάδα	31η ομάδα	32η ομάδα	33η ομάδα		
te b11725	so b3344	ka b3344	a b11725	e b3344		
te b8460	no b11725	da b3344	a b8460	o b11725		
	no b8460	ka b11725		o b8460		
		ka b8460				
cc	dd	ee	ff	gg		
ta A1859		i A1859	a A1859	ka A1859		
ma A1859		da A1859				
		ku A1859				

ΠΙΝΑΚΑΣ 9.2: Πως ταξινομήθηκε η γραμμική Α στις 33 ομάδες που χωρίστηκαν τα πρότυπα της γραμμικής Β με την ιεραρχική ομαδοποίηση.

Επειδή στο κεφάλαιο 7.5 παρατηρήσαμε πως τα δείγματα b11725 και b8460 ομαδοποιούνται παρόμοια θα ήταν καλό να προσπαθήσουμε μια ταξινόμηση του δείγματος a1859 της γραμμικής α σε αυτά μόνο τα 2 δείγματα της γραμμικής β.

Στο στάδιο αυτό θα προσπαθήσουμε να εκπαιδύσουμε το σύστημα μας στο δείγμα b11725&b8460 (δηλαδή και στα δύο δείγματα της γραμμικής Β έτσι ώστε όταν εισάγουμε το δείγμα της γραμμικής Α να μας δώσει αποτελέσματα για το ποια πρότυπα της Α ταιριάζουν με ποια της Β. Δηλαδή εκπαιδύουμε το σύστημα μας με τα δείγματα της γραμμικής Β και το 'τεστάρουμε' με το μοναδικό δείγμα της γραμμικής Α. Βέβαια εμείς δεν μπορούμε να διαπιστώσουμε πόσο καλά δούλεψε το σύστημα μας.

b11725&b8460 =training

A1859 =test

Οι 23 ομάδες που χωρίσαμε τα δύο δείγματα της γραμμικής Β με τη χρήση της ιεραρχικής ομαδοποίησης = group

Εντολή ταξινόμησης στο matlab= classify(test,training,group)

class=classify(test,training,group)	
a 1859	'w'
na 1859	'r'
i 1859	'h'
ta 1859	't'
ti 1859	'd'
ja 1859	'v'
da 1859	'h'
ma 1859	't'
ka 1859	'm'
ku 1859	'h'
ra 1859	'e'
re 1859	'd'
ki 1859	'i'
si 1859	'u'
te 1859	'e'
ru 1859	's'
sa 1859	'i'
di 1859	'a'
ri 1859	'a'
pa 1859	'b'
du 1859	'a'
mi 1859	'q'
su 1859	'q'
ni 1859	's'

ΠΙΝΑΚΑΣ 9.1: Η ταξινόμηση που πραγματοποίησε το matlab στη γραμμική Α μετά την εφαρμογή της εντολής classify.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα	7η ομάδα	8η ομάδα	9η ομάδα
pi b11725	po b11725	ti b11725	ja b11725	ta b8467	ra b11725	me b11725	ka b11725	ke b11725
wa b11725	po b8467	ti b8467	ja b8467	u b8467	re b8467	me b8467	ka b8467	ke b8467
pi b8467				wo b8467	ra b8467			
wa b8467								
a	b	c	d	e	f	g	H	i
								ki 1859
di 1859	pa 1859		ti 1859	ra 1859			i 1859	sa 1859
ri 1859			re 1859	te 1859			da 1859	
du 1859							ku 1859	
10η ομάδα	11η ομάδα	12η ομάδα	13η ομάδα	14η ομάδα	15η ομάδα	16η ομάδα	17η ομάδα	18η ομάδα
no b11725	re b11725	e b11725	o b11725	jo b11725	i b11725	pa b11725	si b11725	ro b11725
no b8467	ri b11725	e b8467	o b8467	jo b8467	i b8467	pa b8467	si b8467	ro b8467
	ri b8467							
j	k	l	m	n	o	p	Q	r
			ka 1859				mi 1859	na 1859
							su 1859	
19η ομάδα	20η ομάδα	21η ομάδα	22η ομάδα	23η ομάδα				
u b11725	te b11725	ko b11725	ta b11725	a b11725				
wo b11725	te b8467	ko b8467	to b11725	a b8467				
we b11725								
to b8467								
we b8467								
s	t	u	v	w				
ru 1859	ta 1859	si 1859	ja 1859	a 1859				
ni 1859	ma 1859							

ΠΙΝΑΚΑΣ 9.2: Πως ταξινομήθηκε η γραμμική Α στις 23 ομάδες που χωρίστηκαν τα δύο πρότυπα της γραμμικής Β (b11725&b8460) με την ιεραρχική ομαδοποίηση.

ΚΕΦΑΛΑΙΟ 10**10.1 Διάφορα συμπεράσματα**

Έτσι όπως κατηγοριοποιήθηκαν τα πρότυπα των τριών δειγμάτων της γραμμικής Β μετά την ιεραρχική ομαδοποίηση παρατηρούμε ότι:

- 1) Το 15% των συλλαβών με ίδια φωνητική αξία και από τα τρία δείγματα ταξινομούνται στις ίδιες ομάδες .
- 2) Το 63% των συλλαβών με ίδια φωνητική αξία από τα δείγματα b11725 και b8460 ταξινομούνται στις ίδιες ομάδες. Διαπιστώνουμε με αυτό το ποσοστό ότι υπάρχουν πολλά κοινά χαρακτηριστικά μεταξύ των επιγραφών της Κνωσού (b11725) και της Πύλου (b8460).
- 3) Ένα πολύ μικρό ποσοστό των συλλαβών του δείγματος b3344 ταξινομείται στις ίδιες ομάδες με τα δύο άλλα δείγματα και κυρίως με το δείγμα b11725, αυτό συμβαίνει ίσως επειδή το πλήθος των συλλαβών του δείγματος b3344 (3344 συλλαβές) είναι πολύ μικρότερο από τα άλλα δύο δείγματα αλλά και επειδή το δείγμα b3344 αποτελεί ένα αρχείο με τυχαίες λέξεις από όλη τη γραμμική β από όλες τις περιοχές εμφάνισής της. Ο λόγος που το δείγμα b3344 ταιριάζει λίγο περισσότερο με το δείγμα b11725 είναι επειδή στο b3344 εκτός από τις τυχαίες λέξεις που επιλέχθηκαν από όλη τη γραμμική β, επιλέχθηκαν κι εκείνες που παρουσίαζαν μεγαλύτερη επανάληψη, πάνω από πέντε φορές, στα αρχεία της Κνωσού (η επιλογή αυτή κρίθηκε αναγκαία επειδή η μεγάλη συχνότητα μιας λέξης σε μια γραφή δημιουργεί μεγαλύτερη πιθανότητα να εντοπισθεί σε άλλη συγγενική γραφή).

Τα δείγματα δεν απαρτίζονται από τις ίδιες φωνητικές συλλαβές-αξίες κι αυτό γιατί κάθε δείγμα είτε προέρχεται από άλλη περιοχή(Κνωσό, Πύλο) είτε αναφέρεται σε διαφορετικού είδους αρχείο(κύρια ονόματα, τοπωνύμια). Οπότε ανάλογα με την περιοχή και το είδος του κειμένου άλλες συλλαβές εμφανίζονται συχνότερα από κάποιες άλλες και εφόσον ασχολούμαστε μόνο με τις 27 συχνότερα επαναλαμβανόμενες συλλαβές από κάθε δείγμα, είναι λογικό να μην έχουμε τα ίδια φωνητικά σημεία σε κάθε δείγμα. Παρατηρώντας την ιεραρχική ομαδοποίηση των συλλαβών θα προσπαθήσουμε να βγάλουμε συμπεράσματα για τις ομάδες που δημιουργούνται: ποιες από τις ομάδες των τριών δειγμάτων περιλαμβάνουν τις ίδιες συλλαβικές αξίες και ποιών δειγμάτων η ομαδοποίηση ταιριάζει περισσότερο.

Παρατηρώντας τις ομαδοποιήσεις που έκανε το matlab σε κάθε δείγμα της γραμμικής β ξεχωριστά:

t=cluster(z,0.8)					
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
wi	ja	ra	e	u	ro
i	na	ri		no	jo
	re			wo	
				so	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
a	ma	wa	ka	ta	
	sa	te	ko	to	
		ke	da		
		si			
		ti			

ΠΙΝΑΚΑΣ 7.5.3: Η ομαδοποίηση του b3344 μετά την υλοποίηση της συνάρτησης cluster

t=cluster(z,0.8)					
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ti	ro	po	u	ra	o
si	ja	pa	wo	te	
			we	re	
			no	ke	
				ri	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
jo	a	pi	ta	ko	
	e	wa	to	ka	
		i			
		me			

ΠΙΝΑΚΑΣ 7.5.4: Η ομαδοποίηση του b11725 μετά την υλοποίηση της συνάρτησης cluster

t=cluster(z,0.8)					
1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ta	pi	no	me	ko	ro
u	wa		po	ka	ja
wo	i		pa		
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
te	o	jo	e	to	ti
re			a	we	si
ra					
ke					
ri					

ΠΙΝΑΚΑΣ 7.5.6: Η ομαδοποίηση του b8460 μετά την υλοποίηση της συνάρτησης cluster

Διαπιστώνουμε πως οι συλλαβές των δειγμάτων b11725 και b8460 ομαδοποιούνται παρόμοια. Οι συλλαβικές αξίες:

Συλλαβές που ομαδοποιούνται ίδια από τα δείγματα b11725 & b8460									
ti	ka	ro	ra	o	jo	a	pi	u	po
si	ko	ja	te			e	wa	wo	pa
			re				i		
			ke						
			ri						

και στα δύο δείγματα ομαδοποιούνται μαζί. Να επισημάνουμε πως οι 27 συλλαβές του b11725 και 27 συλλαβές του b8460 προέρχονται από διαφορετικές περιοχές. Οι μεν 27 συλλαβές του b11725 είναι από την Κνωσό, οι δε 27 συλλαβές του b8460 είναι από την Πύλο. Το ότι αρχεία από διαφορετικές περιοχές ταυτίζονται είναι ενθαρρυντικό.

Τις δοκιμές που κάναμε με τον kmeans σε κάθε δείγμα ξεχωριστά θα τις συγκρίνουμε με την ομαδοποίηση του κάθε δείγματος. Από αυτή τη σύγκριση θα δούμε ποιες συλλαβές συνεχίζουν να ομαδοποιούνται μαζί και δεν αλλάζουν ομάδα. Α) δοκιμές kmeans στο b3344- ιεραρχική ομαδοποίηση του b3344. Β) δοκιμές kmeans στο b11725- ιεραρχική ομαδοποίηση του b11725. Γ) δοκιμές kmeans στο b8460- ιεραρχική ομαδοποίηση του b8460.

Α) δοκιμές kmeans στο b3344-ιεραρχική ομαδοποίηση του b3344.

1 ^η ομάδα	2 ^η ομάδα	3 ^η ομάδα	4 ^η ομάδα	5 ^η ομάδα	6 ^η ομάδα
a	e	ka	ro	no	ja
	wi	da		so	wa
	i				na
					re
					te
					ke
					si
					ma
					ti
					sa
7 ^η ομάδα	8 ^η ομάδα	9 ^η ομάδα	10 ^η ομάδα	11 ^η ομάδα	
ko	jo	ra	ta	u	
		ri	to	wo	

1 ^η ομάδα	2 ^η ομάδα	3 ^η ομάδα	4 ^η ομάδα	5 ^η ομάδα	6 ^η ομάδα
ro	ko	ra	Ja	ta	U
jo		ri	Wa	to	Wo
			Na		
			Re		
			Ti		
7 ^η ομάδα	8 ^η ομάδα	9 ^η ομάδα	10 ^η ομάδα	11 ^η ομάδα	
no	e	te	A	ka	
so		ke		da	
		si			
		ma			
		sa			
		wi			
		i			

1 ^η ομάδα	2 ^η ομάδα	3 ^η ομάδα	4 ^η ομάδα	5 ^η ομάδα	6 ^η ομάδα
ra	te	ka	e	ko	a
ri	ke	da			
	si				
	ma				
	sa				
	wi				
	i				
7 ^η ομάδα	8 ^η ομάδα	9 ^η ομάδα	10 ^η ομάδα	11 ^η ομάδα	
ta	no	ro	ja	u	
to	so	jo	wa	wo	
			na		
			re		
			ti		

1 ^η ομάδα	2 ^η ομάδα	3 ^η ομάδα	4 ^η ομάδα	5 ^η ομάδα	6 ^η ομάδα
ka	no	ra	Ro	te	A
da	so	ri	Jo	ke	
				si	
				ma	
				sa	
				wi	
				i	
7 ^η ομάδα	8 ^η ομάδα	9 ^η ομάδα	10 ^η ομάδα	11 ^η ομάδα	
u	e	ja	Ta	ko	
wo		wa	To		
		na			
		re			
		ti			

. Β) δοκιμές kmeans στο b11725- ιεραρχική ομαδοποίηση του b11725.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
po	re	ra	a	ti	o
pa	ri	te		si	
		ke			
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
ta	ko	e	pi		
ro	ka		wa		
ja			i		
to			me		
u			jo		
wo					
we					
no					

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ro	a	ko	e	pi	jo
ja		o		wa	
		ka		i	
		po		ti	
				si	
				pa	
				me	
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
ra	ta	u	no	re	
te	to	wo		ri	
ke		we			

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ta	o	a	ko	jo	pi
ro			ka		wa
ja					i
					ti
					si
					me
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
re	po	ra	to	e	
ri	pa	te	u		
		ke	wo		
			we		
			no		

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
e	to	ti	ka	jo	ra
	u	si	po		te
	wo		pa		ke
	we				
	no				
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
ta	pi	a	ko	re	
ro	wa		o	ri	
ja	i				
	me				

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
jo	ro	to	no	e	ta
	ja				
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	
pi	u	ko	a	ra	
wa	wo	o		te	
i	we	ka		re	
ti		po		ke	
si				ri	
pa					
me					

Γ) δοκιμές kmeans στο b8460- ιεραρχική ομαδοποίηση του b8460.

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
jo	ta	te	e	pi	ka
	u	re	a	wa	me
	wo	ra		i	po
		ke		si	pa
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
we	ko	ri	ro	o	
no	to	ti	ja		

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ro	ko	ka	jo	o	pi
ja	to	me			wa
		po			i
		pa			si
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
ta	te	e	we	ri	
u	re	a	no	ti	
wo	ra				
	ke				

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ko	o	we	me	ro	jo
to		no	po	ja	
			pa		
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
e	te	ka	pi	ri	ta
a	re		wa	ti	u
	ra		i		wo
	ke		si		

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
o	we	re	ri	jo	ro
	no	ra			ja
					ta
					u
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
ka	te	pi	wo	e	ti
me	ke	wa	ko	a	si
po		i	to		
pa					

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ka	jo	ri	we	ko	ro
me		ti	no	to	ja
po					
pa					
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
te	a	ta	pi	e	o
re		u	wa		
ra		wo	i		
ke			si		

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
wo	ro	te	me	jo	ko
to	ja	re	po		ka
we	ta	ra	pa		
no	u	ke			
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
e	ri	ti	pi	o	a
		si	wa		
			i		

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
we	ka	ri	o	jo	pi
no	me	ti			wa
	po				i
	pa				si
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
ro	te	ja	ko	ta	e
	re		to	u	a
	ra			wo	
	ke				

1η ομάδα	2η ομάδα	3η ομάδα	4η ομάδα	5η ομάδα	6η ομάδα
ri	ta	te	ko	ti	e
	u	re	to	si	a
	wo	ra			
		ke			
7η ομάδα	8η ομάδα	9η ομάδα	10η ομάδα	11η ομάδα	12η ομάδα
we	o	ka	pi	ro	jo
no		me	wa	ja	
		po	i		
		pa			

Διαπιστώνουμε από όλες τις δοκιμές της ομαδοποίησης *kmeans* πως ο αλγόριθμος αυτός όσες ομάδες και να του ζητήσεις να φτιάξει, θα τις φτιάξει άσχετα από το αν είναι οι σωστές ή όχι. Σίγουρα χρειάζονται πάρα πολλές δοκιμές όχι μόνο στις ομάδες αλλά και στις παραμέτρους του αλγόριθμου *kmeans* για να βρεθούν και να σχηματιστούν σωστές ομάδες, κάτι το οποίο προσπαθήσαμε στη συγκεκριμένη έρευνα. Βέβαια το πλήθος των δοκιμών μπορεί να μας δείξει ποιές συλλαβές είναι αυτές που μετακινούνται σε άλλες ομάδες και ποιές μένουν σταθερές. Δηλαδή και χωρίς να φτάσουμε στον ιδανικό αριθμό ομάδων μπορούμε να έχουμε μια εκτίμηση για τα πρότυπα-συλλαβές και πως αυτά συμπεριφέρονται.

Το μεγαλύτερο πρόβλημα που αντιμετωπίσαμε στην συγκεκριμένη έρευνα είναι ότι δεν ξέραμε από την αρχή σε ποιες κατηγορίες – ομάδες χωρίζονταν τα δείγματά μας, και η προσπάθεια μας περιορίστηκε στο να εκπαιδεύσουμε δίκτυα με τέτοιο τρόπο που να γνωρίσουν τις ομοιότητες και ανομοιότητες των διανυσμάτων ώστε να φτιάξουν στη συνέχεια κατηγορίες – ομάδες. Αυτό το πρόβλημα προσπαθήσαμε να το αντιμετωπίσουμε με όσες δυνατότητες μας επέτρεπε η αναγνώριση προτύπων. Θεωρούμε ότι θα είχαμε καλύτερα αποτελέσματα με την χρήση τεχνητών νευρωνικών δικτύων γιατί θα προσέγγιζε αποτελεσματικότερα μια τόσο πολύπλοκη έρευνα. Ένα άλλο πρόβλημα ήταν τα πολύ μικρά σε μέγεθος δείγματα που μας δόθηκαν γεγονός που δεν επέτρεπε στα δίκτυα μας να εκπαιδευτούν σωστά. Συμπεραίνουμε πως τα αποτελέσματα θα ήταν περισσότερο και καλύτερα ευδιάκριτα με τη χρήση μεγαλύτερων δειγμάτων που σημαίνει περισσότερη πληροφορία στα δίκτυα. Τα αποτελέσματα που πήραμε με την χρήση της αναγνώρισης προτύπων πιστεύουμε πως είναι ικανοποιητικά σε μεγάλο βαθμό για τα γνωστά δείγματα της Γραμμικής Β και όχι τόσο ικανοποιητικά για το άγνωστο δείγμα της Γραμμικής Α λαμβάνοντας υπόψη την αντίστοιχη στατιστική μελέτη του κ. Τσικριτζή.

Βιβλιογραφία

- 1) J.T.Hooker, Linear B: an introduction α' έκδοση, Bristol Classical press, U.K.(1980)
- 2)John Chadwick, Linear B and related scripts
1987 the trustees of the British Museum
Published by the British Museum Publications Ltd
- 3) Μηνάς Δ. Τσικριτσής, Γραμμική Α – Στατιστικά
- 4)Γεώργιος Παπαδουράκης, αναγνώριση προτύπων, Α.Τ.Ε.Ι. Κρήτης Τμήμα εφαρμοσμένης πληροφορικής και πολυμέσων, θεωρία αναγνώρισης προτύπων
- 5)Ιωάννης Χαλκιαδάκης, αλγόριθμος k-means – ιεραρχική ομαδοποίηση, Α.Τ.Ε.Ι. Κρήτης Τμήμα εφαρμοσμένης πληροφορικής και πολυμέσων, εργαστήριο αναγνώρισης προτύπων
- 6) 'COMPUTER TECHNIQUES IN THE STUDY OF THE MINOAN LINEAR A SCRIPT'
(1971-1996) BACK TO THE FUTURE? GARETH OWENS