ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
Εργαστήριο Ευφυών Συστημάτων

Αναγνώριση Στατικών Χαρακτηριστικών στο πλαίσιο της μετάφρασης Νοηματικών Γλωσσών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Ιωάννη Ι. Κουλιεράκη

Επιβλέπων: Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Αθήνα, 2020

Εθνικο Μετσοβιο Πολυτεχνειο
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
Εργαστήριο Ευφυών Συστημάτων

Αναγνώριση Στατικών Χαρακτηριστικών στο πλαίσιο της μετάφρασης Νοηματικών Γλωσσών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Ιωάννη Ι. Κουλιεράκη

Επιβλέπων: Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 11 Μαΐου 2020.

.....	.....	.....
Ανδρέας Σταφυλοπάτης	Γεώργιος Στάμου	Ευθυμίου Ελένη
Καθηγητής Ε.Μ.Π	Καθηγητής Ε.Μ.Π	Ερευνήτρια Α΄, Διευθύντρια Ερευνών
		Ινστιτούτο Επεξεργασίας Λόγου, «Ε.Κ. Αθηνά»

Αθήνα, Φεβρουάριος 2020.

.....
(Ιωάννης Ι. Κουλιεράκης
Διπλωματούχος σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών Ε.Μ.Π.

Ευχαριστίες

Η παρούσα διπλωματική εκπονήθηκε σε στενή συνεργασία του Εργαστηρίου Ευφυών Συστημάτων του Εθνικού Μετσόβιου Πολυτεχνείου με το Ινστιτούτο Επεξεργασίας του Λόγου (ΙΕΛ). Θα ήθελα, λοιπόν, να ευχαριστήσω θερμά την κα Ευθυμίου για την πρωτοβουλία αυτής της πολύ ενδιαφέρουσας συνεργασίας, την πρόσβαση στη πολύ μεγάλη συλλογή δεδομένων του ΙΕΛ, το ενδιαφέρον που έδειξε καθ΄ όλη τη διάρκεια αυτής, για την φιλοξενία της. Για την τεχνική καθοδήγηση, θα ήθελα να ευχαριστήσω τον κ. Γεώργιο Σιόλα για την πολύτιμη, άμεση και πάντα καίρια βοήθεια που μου προσέφερε όποτε ήταν απαραίτητο. Τέλος, θα ήθελα να ευχαριστήσω την τριμελή επιτροπή και ιδιαίτερα τον Καθηγητή μου, κ. Γεώργιο-Ανδρέα Σταφυλοπάτη για το σεβασμό και την εμπιστοσύνη του ως προς τη δουλειά μου.

Ιωάννης Ι. Κουλιεράκης

Περίληψη

Η παρούσα εργασία εξετάζει μία μέθοδο βασισμένη σε μοντέλα Μηχανικής Μάθησης με σκοπό την αναγνώριση των κύριων δομικών χαρακτηριστικών που συνθέτουν μεμονωμένες λέξεις μίας νοηματικής γλώσσας. Αρχικά, γίνεται αναφορά στις ιδιαιτερότητες των νοηματικών γλωσσών συγκριτικά με τον προφορικό λόγο αλλά και μία παρουσίαση για μερικές από τις συλλογές καταγραφής νοηματικών γλωσσών διεθνώς. Η μεθοδολογία που αναπτύσσεται εστιάζει στην παρακολούθηση και ανάλυση του κυρίαρχου χεριού κάθε νοηματιστή αλλά είναι άμεσα επεκτάσιμη και στο δευτερεύον χέρι. Το πρόβλημα της μετάφρασης από μία νοηματική γλώσσα σε γραπτό λόγο αντιμετωπίζεται με τη χρήση του συστήματος επισημείωσης HamNoSys ως ενδιάμεσο στάδιο. Το πρώτο μέρος της μεθόδου αφορά την διαλογή των χρήσιμων στιγμιοτύπων κάθε βίντεο μέσα από την ανάπτυξη ενός προστακτικού αλγορίθμου και την μετάβαση από την επισημείωση σε επίπεδο βίντεο στην επισημείωση σε επίπεδο στιγμιοτύπων με χρήση αλγορίθμων μη επιβλεπούμενης μάθησης. Τέλος, το τροποποιημένο σύνολο δεδομένων χρησιμοποιείται για την εκπαίδευση βαθιών νευρωνικών δικτύων με εξειδίκευση το κάθε ένα στην αναγνώριση επί μέρους χαρακτηριστικών. Ολοκληρώνοντας, εξηγείται πώς ο συνδυασμός αυτών των δικτύων μπορεί να αποτελέσει ένα πολύ καλό σύστημα συστάσεων σε επίπεδο λέξεων και γίνονται προτάσεις για σχεδιασμό Dataset που θα μπορέσουν να χρησιμοποιηθούν για την ανάπτυξη μελλοντικών συστημάτων με δυνατότητες μετάφρασης νοηματικών γλωσσών.

Λέξεις Κλειδιά: Νοηματική Γλώσσα, Openpose, HamNoSys, Βαθιά Νευρωνική Μάθηση, Μηχανική Μάθηση, Αυτόματη Μετάφραση, Αναγνώριση Χειρονομιών, Εντοπισμός Χεριού

Abstract

The present paper examines a method based on Machine Learning models to identify the key structural features that make up single words of a Sign Language. Initially, a simple reference is made to the specific features of Sign Languages compared to spoken ones but also a presentation on some of the Sign Language collections from around the world. The methodology developed focuses on monitoring and analyzing the dominant hand of each signer but is also immediately extensible to the non-dominant hand. The problem of translating from a sign language into a written language is addressed using the HamNoSys notation system as an intermediate stage. The first part of the method is to sort out the useful snapshots of each video by developing an deterministic algorithm and the transition from video-level notation to snapshot-level notation using unsupervised learning algorithms. Finally, the modified dataset is used to train deep neural networks specializing in the recognition of individual features. In conclusion, it is explained how the combination of these networks can be a very good word-level recommendation system, and suggestions are made for designing a Dataset that can be used to develop future systems with sign language translation capabilities.

Key Words: Sign Language, Openpose, HamNoSys, Deep Learning, Machine Translation, Gesture Recognition, Hand Location

Περιεχόμενα

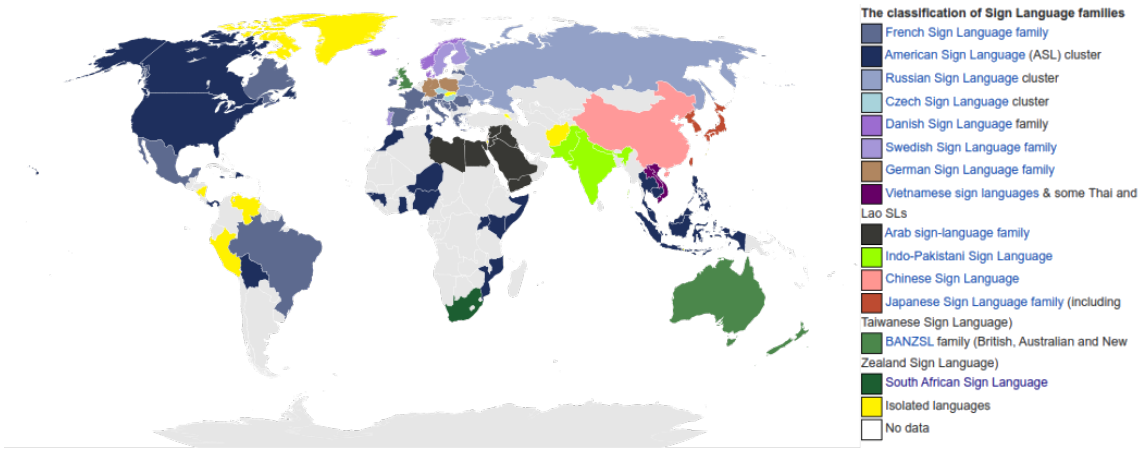
Κεφάλαιο 1

Εισαγωγή

Η Νοηματική Γλώσσα αποτελεί ένα μέσο οπτικής και μη λεκτικής ενσώματης επικοινωνίας με έμφαση κυρίως στα χέρια που χρησιμοποιείται σε όλες τις περιοχές του κόσμου κατεξοχήν από άτομα με δυσκολία ή απουσία ακοής. Στη χώρα μας, η ύπαρξη της Νοηματικής Γλώσσας έγινε ιδιαίτερα διάσημη μέσα από τα «Δελτία Ειδήσεων στη Νοηματική». Το πρώτο δελτίο ειδήσεων για κωφούς στη νοηματική γλώσσα προβλήθηκε το 1982 από την ΕΡΤ και το 2007 δια νόμου έγινε υποχρεωτική η ύπαρξη δελτίου ειδήσεων στη νοηματική τουλάχιστον επτά λεπτών για κάθε ενημερωτικό τηλεοπτικό σταθμό. Πέραν όμως της τηλεοπτικής ενημέρωσης η κοινότητα των κωφών αντιμετωπίζει μεγάλο εμπόδια στην επικοινωνία με τον υπόλοιπο κόσμο καθώς έστω και τα βασικά στοιχεία γύρω από τη Νοηματική Γλώσσα παραμένουν εντελώς άγνωστα για την πλειοψηφία του κόσμου.

Αρχικά, υπάρχει η λανθασμένη εντύπωση ότι ο όρος «Νοηματική Γλώσσα» αναφέρεται σε ένα ενοποιημένο διεθνές πρωτόκολλο επικοινωνίας. Η ανάγκη δημιουργίας ενός κοινού κώδικα επικοινωνίας στις κοινότητες ατόμων με προβλήματα ακοής οδήγησε στην ύπαρξη πολλών διαφορετικών Νοηματικών Γλωσσών ανά τον κόσμο όπως συνέβη αντίστοιχα με τις ομιλούμενες γλώσσες. Επιπλέον, με τον ίδιο τρόπο που σε μία προφορική γλώσσα μικρότερες ομάδες της ίδιας εθνότητας αναπτύσσουν διαλέκτους λόγω απομόνωσης από το υπόλοιπο έθνος, έτσι και οι Νοηματικές Γλώσσες εμφανίζουν ιδιομορφίες όταν χρησιμοποιούνται από κοινότητες κωφών που ζουν απομονωμένες η μία από την άλλη. Για παράδειγμα από την American Sign Language (ASL) προέκυψε η η Black American Sign Language (BASL) ως ένα είδος διαλέκτου που αναπτύχθηκε λόγω έντονου διαχωρισμού στα σχολεία του νότου μεταξύ των μαύρων από τους λευκούς μέχρι το 1954. Η διεθνής ένωση κωφών ξεκινώντας από το 1973, ένα γλωσσικό κατασκεύασμα με την ονομασία «Διεθνής Νοηματική Γλώσσα» (International Sign Language) [21] με σκοπό την δημιουργία μίας υποτυπώδους γλώσσας που θα είναι εύκολα κατανοητή ανάμεσα σε δύο άτομα που είναι χρήστες του νοηματικού λόγου αλλά δεν γνωρίζουν κάποια κοινή νοηματική γλώσσα. Συνολικά, υπάρχουν περισσότερες από 150 διαφορετικές Νοηματικές Γλώσσες σε όλο τον κόσμο αλλά οι περισσότερες μπορούν να ομαδοποιηθούν σε οικογένειες γλωσσών καθώς παρουσιάζουν πολλά κοινά μεταξύ τους. Οι βασικότερες οικογένειες Νοηματι-

κών Γλωσσών είναι οι Γαλλική, Αμερικανική, Ρωσική, Τσέχικη, Δανική, Σουηδική, Γερμανική, Βιετναμέζικη, Αραβική, Κινέζικη, Ιαπωνική και Βρετανική-Αυστραλιανή-Νεοζηλανδική. Το γεγονός ότι οι Νοηματικές Γλώσσες δύο περιοχών μπορεί να σχετίζονται ετυμολογικά και ανήκουν στην ίδια οικογένεια δεν συνεπάγεται ότι το ίδιο συμβαίνει και με τις αντίστοιχες προφορικές γλώσσες των περιοχών αυτών. Για παράδειγμα στο Σχήμα 1.1 φαίνεται ότι η Ελληνική Νοηματική Γλώσσα (ΕΝΓ), όπως και οι υπόλοιπες Ευρωπαϊκές νοηματικές γλώσσες της Μεσογείου ανήκουν στην οικογένεια των Γαλλικών.



Σχήμα 1.1: Κατηγοριοποίηση οικογενειών Νοηματικών Γλωσσών Παγκοσμίως

Γενικότερα, είναι παράδοξο να γίνεται οποιαδήποτε συνειρμική σύνδεση μεταξύ κάποιας νοηματικής γλώσσας και της αντίστοιχης ομιλούμενης, εφόσον ο λόγος ύπαρξης της πρώτης είναι η αδυναμία επαφής ορισμένων ανθρώπων με τη δεύτερη. Όπως θα φανεί στα επόμενα κεφάλαια, η διαφορά στη φύση των δύο τρόπων επικοινωνίας ενδεχομένως καθιστά το πρόβλημα της μετάφρασης από μία νοηματική γλώσσα σε γραπτό λόγο μίας προφορικής γλώσσας ένα ιδιαίτερα σύνθετο πρόβλημα σε πολλά επίπεδα. Στη βάση όλης της ανάλυσης μελετάμε την αντιστοιχία μίας γλώσσας που μεταδίδεται μέσω της κίνησης στο χώρο και συνεπώς θα πρέπει να έχει δομικά μία πολυδιάστατη περιγραφή για κάθε της λεξιλογική οντότητα, σε μία γλώσσα που αρκεί μία διάσταση που εκτυλίσσεται σειριακά στο χρόνο. Επιπλέον, για τον ίδιο λόγο η σύνταξη των νοηματικών γλωσσών διαφέρει σημαντικά μίας και δεν οφείλει να υπακούει στη σειριακή παράθεση πληροφορίας πάνω στην οποία βασίζεται κάθε ομιλούμενη γλώσσα. Η διαδικασία μετάφρασης μίας νοηματικής γλώσσας σε κάποια γραπτή γλώσσα είναι θεωρητικά πιο περίπλοκο πρόβλημα από τη μετάφραση της Κινεζικής Γλώσσας από φωνή απευθείας σε κείμενο γραμμένο στα Ελληνικά.

Προσπαθώντας να επιτευχθεί η μετάφραση μέσω ενός συστήματος μηχανικής μάθησης θα πρέπει να υπάρξουν πολύ μεγάλα σύνολα δεδομένων που θα προσφέρονται για εκπαίδευση ή θα πρέπει να χρησιμοποιηθεί μία ενδιάμεση γραπτή έκφραση των νοηματικών γλωσσών. Θα αναλύσουμε ποιες είναι μερικές από τις διασημότερες καταγραφές νοηματικών γλωσσών και γιατί οι περισσότερες παρότι έχουν πολύ μεγάλη

αξία λεξιλογικά, δεν μπορούν να χρησιμοποιηθούν ως σύνολα δεδομένων για μηχανική μάθηση. Το ερευνητικό κομμάτι θα εστιάσει στη μετάφραση των βίντεο νοηματικής γλώσσας σε μία γραπτή απεικόνιση σε επίπεδο λέξης χρησιμοποιώντας το σύστημα HamNoSys που δημιουργήθηκε για να εξυπηρετήσει τους σκοπούς της σύνθεσης νοημάτισης από εικονικούς χαρακτήρες (avatar). Η σημαντικότητα αυτής της εφαρμογής έγκειται στο γεγονός ότι θα μπορέσει να αποτελέσει αργότερα ένα πολύ καλό εργαλείο συστάσεων σε επίπεδο επισημείωσης όλο και μεγαλύτερων datasets που με τη σειρά τους θα χρησιμοποιηθούν για την εκπαίδευση όλο και καλύτερων συστημάτων μετάφρασης. Ολοκληρώνοντας την εργασία θα παρουσιαστεί μία αναλυτική μέθοδος για το σωστό σχεδιασμό ενός τέτοιων συνόλων δεδομένων.Ακολουθώντας αυτή τη διαδικασία θα μπορέσει να υπάρξει μία ραγδαία εξέλιξη στο κομμάτι αναγνώρισης νοηματικών γλωσσών σχεδόν σαράντα χρόνια αργότερα από τις αντίστοιχες εξελίξεις στην αναγνώριση φωνής.

Η σημαντικότητα μίας τέτοιας πορείας θα είχε τεράστιο κοινωνικό αντίκτυπο καθώς θα απαλλάξει τα άτομα με εκ γενετής απουσία ακοής από το τεράστιο συγκριτικό μειονέκτημα που έχουν ζώντας σε ένα κόσμο όπου η οπτική μετάδοσης πληροφορίας σχεδόν πάντα πραγματοποιείται μέσω του γραπτού κειμένου το οποίο όμως βασίζεται στην γραπτή απεικόνιση του προφορικού λόγου. Είναι άδικο να υπάρχει η απαίτηση από πλευράς κοινωνίας να γίνεται κατανοητός ο γραπτός λόγος από άτομα με προβλήματα ακοής όταν η γλώσσα χρησιμοποιεί κάποιο αλφάβητο. Τα αλφάβητα χρησιμοποιούνται για να περιγράψουν τους διαφορετικούς ήχους που περιέχει μία προφορική γλώσσα. Συνεπώς, το να κρίνεται κάποιος που δεν έχει τη δυνατότητα ακοής πάνω στην κατανόηση του προφορικού λόγου οδηγεί στο να θεωρείται ότι υστερεί νοητικά. Έτσι, ιδιαίτερα σε παλαιότερες κοινωνίες τα άτομα με εκ γενετής προβλήματα ακοής ήταν συνυφασμένα με τα άτομα με μη λεκτικό αυτισμό. Κάτι τέτοια εν μέρη μαρτυρά το γεγονός ότι ο πρώτος σύλλογος κωφών στη Μεγάλη Βρετανία το 1886 είχε την ονομασία "National Association for the Deaf and Dumb". Σήμερα, αν και τέτοιοι χαρακτηρισμοί έχουν αρχίσει να εκλείπουν, ελάχιστα πράγματα γίνονται διεθνώς από πλευράς υποδομών για τα άτομα με προβλήματα ακοής. Αντιθέτως, στην ίδια περίοδο είναι εκπληκτική η πρόοδος που έχει πραγματοποιηθεί για τη διευκόλυνση των ατόμων με προβλήματα όρασης. Μπορεί οι απαιτούμενες εφαρμογές για την εξυπηρέτηση των κωφών και βαρήκοων ατόμων να μην είναι τεχνικά τόσο απλές όσο η χρήση μεγαφώνων για σημαντικές ανακοινώσεις και η γραφή braile αλλά πλέον κάποιες πρωτογενείς υλοποιήσεις είναι εφικτές κρίνοντας το στάδιο όπου έχει φτάσει μέχρι στιγμής η έρευνα πάνω στην τεχνητή νοημοσύνη όπως και τα αποθηκευτικά και επεξεργαστικά μέσα.

Κεφάλαιο 2

Γλωσσολογική Προσέγγιση στην Νοηματική Γλώσσα

2.1 Δομικά χαρακτηριστικά των Νοημάτων

Η συστημική ανάλυση των νοηματικών γλωσσών ξεκινά το 1960 από τον Dr. Stokoe [28]. Επηρεασμένος από την ανάλυση των φωνητικών γλωσσών ανέπτυξε ένα σύστημα για την ανάλυση των νοηματικών γλωσσών που αποτελούνταν από δομικά μέρη πιο στοιχειώδη από την έννοια της λέξης/νόημα (sublexical items) παρόμοια με την έννοια του φωνήματος σε σχέση με τις φωνητικές γλώσσες.Πρότεινε τους όρους «χειρόνημα» και «χειρολογία» (που αποτελούν ανεπίσημη μετάφραση των τεχνιτών όρων "chereme","cherology") κατά αντιστοιχία των όρων «φώνημα» και «φωνολογία» που προέρχονταν από την επεξεργασία φωνής. Ωστόσο, ο προφορικός λόγος διαφέρει σημαντικά από τη νοημάτιση ως προς το μέσον εκφοράς του λόγου. Ο προφορικός λόγος μπορεί να περιγραφή πλήρως ως η διακύμανση ενός μεγέθους, της πίεσης του αέρα, συναρτήσει του χρόνου. Αν αγνοηθεί η παράμετρος της χρονικής διάρκειας ως πλεονάζουσα, τότε οι διαφορετικοί ήχοι που μπορεί να εμφανιστούν στα πλαίσια του προφορικού λόγου μπορούν να περιγραφούν από ένα φώνημα ο καθένας. Αντιθέτως, οι νοηματικές γλώσσες προκειμένου να περιγραφούν απαιτείται, πρωτίστως, μία χωρική περιγραφή. Θεωρώντας το χέρι ως ένα τρισδιάστατο αντικείμενο γνωρίζουμε ότι η στοιχειώδης περιγραφή του θα περιλαμβάνει **θέση** στο χώρο και **προσανατολισμό**. Επιπλέον, δεδομένου μίας τοποθέτησης του χεριού στο χώρο, χρειάζεται να περιγραφεί η **διάταξη των δακτύλων**. Επιπλέον, οι νοηματικές γλώσσες περιέχουν λέξεις οι οποίες μπορεί να αποδίδονται στατικά ή δυναμικά. Συνεπώς, είναι απαραίτητο για την καταγραφή αυτών, ένα χαρακτηριστικό που εμπεριέχει την μεταβολή στο χρόνο δηλαδή την **κίνηση**. Η κίνηση είναι ένα χαρακτηριστικό που μπορεί να εμφανιστεί ως γενικά με τη μορφή κίνησης ολόκληρου του χεριού αλλά και πιο εσωτερικά ως κίνηση μόνο της παλάμης Με αυτά τα χαρακτηριστικά μπορεί να δοθεί με πληρότητα η δράση του ενός χεριού στα πλαίσια της νοημάτισης. Εντούτοις, σημαντικό μέρος των νοημάτων αφορούν και τα 2 χέρια. Για αυτό το λόγο χρειάζεται

ένας τρόπος περιγραφής των αντίστοιχων στοιχείων για το δευτερεύον χέρι όπως η **ύπαρξη ή μη συμμετρίας** και τυχόν **επαφή**. Ολοκληρώνοντας την φωνολογική ανάλυση για λόγους πληρότητας πρέπει να υπάρξει ένας τρόπος αναφοράς στην περιγραφή των υπολοίπων σημείων του σώματος, όταν αυτά συνεπικουρούν στην εκφορά νοημάτων. Τα χαρακτηριστικά αυτά διεθνώς αναφέρονται ως non-manuals. Συνοψίζοντας τα φωνολογικά χαρακτηριστικά περιγραφής νοηματικών γλωσσών μπορούν να κατηγοριοποιηθούν ως εξής:

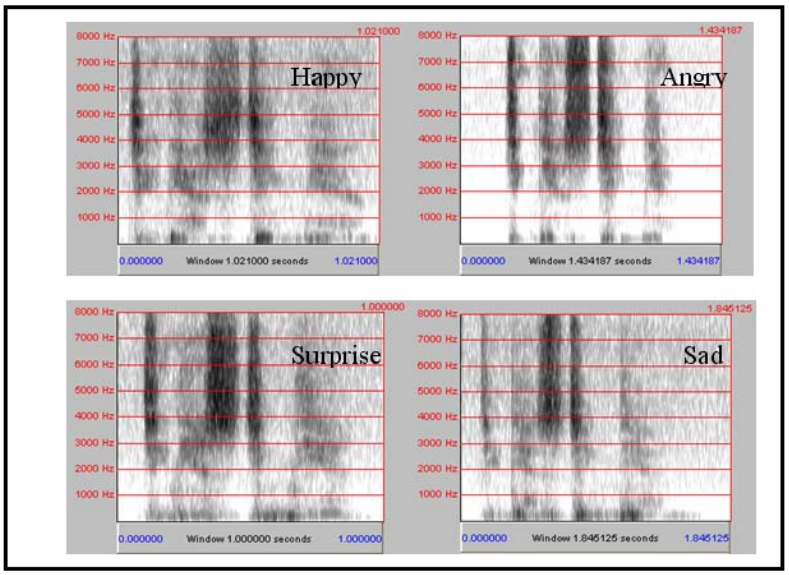
- σχήμα χεριού
- κατεύθυνση παλάμης
- θέση χεριού
- κίνηση
- non-manuals

Αναλυτικότερα, η τεχνική υλοποίηση μία φωνολογικής περιγραφής γίνεται στην Ενότητα 2.3. Στο σημείο αυτό αξίζει να σημειωθεί ότι μία βασική διαφορά της φωνολογικής προσέγγισης των νοηματικών γλωσσών από εκείνη των προφορικών έγκειται στο γεγονός στην περίπτωση της Νοηματικής υπάρχει μεγάλη συσχέτιση ανάμεσα στα φωνολογικά χαρακτηριστικά και στην «τονικότητα» και στην «ένταση». Κατά τον προφορικό λόγο όταν δηλώνεται κάποιο συναίσθημα ή απορία, τα φωνήματα παραμένουν πανομοιότυπα (Σχήμα 2.1). Από την άλλη πλευρά, στις νοηματικές γλώσσες τα χαρακτηριστικά όπως η ταχύτητα των κινήσεων και οι εκφράσεις του προσώπου που δυνητικά δηλώνουν ένταση ή απορία έχουν ήδη οριστεί ως «φωνολογικά» χαρακτηριστικά. Το ευθέως ανάλογο θα ήταν μία ομιλούμενη γλώσσα όπου για παράδειγμα η λέξη «περπατώ», θα προφερόταν διαφορετικά ανάλογα με την συναισθηματική κατάσταση του υποκειμένου ή τον ρυθμό βαδίσματος. Σε καμία περίπτωση δεν υπονοείται ότι οι προφορικές γλώσσες είναι ορθότερα δομημένες. Ωστόσο, κατά ένα αόριστο τρόπο, είναι γεγονός ότι οι νοηματικές γλώσσες «φωνολογικά» βρίσκονται πολύ πιο κοντά στην έννοια που περιγράφουν ως προς την ανθρώπινη αντίληψη, σε σχέση με τις ομιλούμενες. Με άλλα λόγια είναι πολύ πιο εύκολα για κάποιον ομιλούντα να καταλάβει τα συμφραζόμενα μία πρότασης σε κάποια νοηματική παρά σε μία άγνωστη προφορική γλώσσα. Μάλιστα, έχει ενδιαφέρον ότι και το αντίστροφο ισχύει. Δύο ομιλούντες που θέλουν να επικοινωνήσουν και δεν γνωρίζουν κάποια κοινή γλώσσα, ενδεχομένως θα προσφύγουν σε κάποια αυτοσχέδια μέθοδο επικοινωνίας μέσω χειρονομιών.

2.2 Λεξιλογικά Χαρακτηριστικά

2.2.1 Ταξινομητές (Classifiers)

Κάθε νοηματική γλώσσα, όπως και κάθε ομιλούμενη χαρακτηρίζεται σε μεγάλο βαθμό από το λεξιλόγιό της. Γενικότερα, Η ιδιαιτερότητα των νοηματικών γλωσσών έγκειται στο γεγονός ότι είναι οπτικές γλώσσες. Έτσι, με τον ίδιο τρόπο που εμφανίζονται τα επιφωνήματα σε μία ομιλούμενη γλώσσα και έπειτα τα παράγωγα αυτών (π.χ. «γαβ» και «γαβγίζω», "boo" και "boohooing"), έτσι και το λεξιλόγιο μίας Νοηματικής Γλώσσας, για τις έννοιες που έχουν κάποια υλική υπόσταση, το νόημα τείνει



Σχήμα 2.1: Σπεκτρογράμματα της ίδιας πρότασης με διαφορετικά συναισθήματα [11]

να ομοιάζει οπτικά στην έννοια που περιγράφεται (Σχήμα 2.13). Βέβαια, το φαινόμενο αυτό είναι σαφώς συνηθέστερο στην Νοηματική Γλώσσα σε σχέση με τις ομιλούμενες. Μάλιστα, στην ιδιότητα αυτή της νοηματικής γλώσσας να χρησιμοποιεί το χώρο νοημάτισης ως τον φυσικό τρισδιάστατο χώρο βασίζεται η έννοια του **ταξινομητή**[1] ((classifier)).

Οι ταξινομητές είναι κατηγορίες μορφολογικών οντοτήτων, τόσο κινούμενων όσο και ακίνητων, που υποδηλώνονται με την απεικόνιση κάποιας σημαντικής εικονικής όψης αυτών των οντοτήτων με χειροκίνητη άρθρωση. Εμφανίζονται σε συνδυασμό με ρήματα ή ουσιαστικά εκφράζοντας σχήμα, κίνηση και τοποθεσία. Η σχετική ιδιότητα που καθορίζει τη μορφή του ταξινομητή μπορεί να είναι η τρισδιάστατη απεικόνιση του σχήματος ενός αντικειμένου. Για παράδειγμα η χειρομορφή «Ο»(βλ. Παράρτημα Β′) στατικά μπορεί να αντιπροσωπεύει ένα στρογγυλό αντικείμενο. Διαφορετικά, μπορεί να είναι ένα δισδιάστατο ή τρισδιάστατο αντικείμενο που να απεικονίζεται μέσω του περιγράμματός του. Συγκεκριμένα, η έννοια του καθρέπτη θα μπορούσε να νοηματιστεί με τα 2 χέρια να κινούνται συμμετρικά διαγράφοντας ένα το πλαίσιο ενός καθρέπτη. Τέλος, ένας ταξινομητής μπορεί να αναπαριστά την απεικόνιση της κίνησης ενός αντικειμένου όπως η χαρακτηριστική κίνηση που γίνεται κατά το χειρισμό ενός εργαλείου (π.χ. κατσαβίδι). Οι ταξινομητές είναι μέρος του λεξιλογίου που βρίσκεται εκτός του πυρήνα των νοηματικών γλωσσών και βρίσκονται (αν και ποικίλλουν βαθμούς και με διάφορες λεξικές διαφορές) σε κάθε νοηματική γλώσσα που έχει μελετηθεί μέχρι

[1] Δεν πρέπει να συγχέεται με την έννοια του ταξινομητή της Μηχανικής Μάθησης που εμφανίζεται στο Κεφάλαιο 5

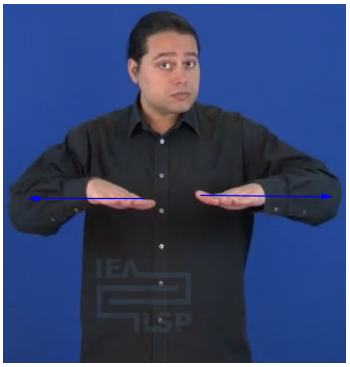
σήμερα.

Ο όρος του ταξινομητή λεξιλογικά εισήχθη στην Νοηματική από τον Frishberg το 1975. Ωστόσο, η έννοια προϋπήρχε λεξιλογικά και εμφανίζεται σε κάποιες μη Ινδοευρωπαϊκές ομιλούμενες Γλώσσες όπως η γλώσσα Κάντο. Προέρχεται από μία οικογένεια γλωσσών που συγχωνεύθηκαν σε μία και ομιλείται από τους γηγενής Αμερικανούς που προέρχονται από περιοχές στο Ανατολικό Τέξας, τη Λουιζιάνα και τα νότια τμήματα της Οκλαχόμα και του Αρκάνσας. Χρησιμοποιεί τους ταξινομητές σε συνδυασμό με ρήματα για να αποδώσει σχήματα.

Ανάλογα με την εννοιολογική κατηγορία οι ταξινομητές διαχωρίζονται σε **κατηγοριοποίησης** και **Μεγέθους-Σχήματος**.Οι ταξινομητές κατηγοριοποίησης μπορεί να εκφράζουν οντότητα, μέρος του σώματος ή χειρισμό.

Οι ταξινομητές που εκφράζουν οντότητες είναι η πιο γενική περίπτωση ταξινομητών και χρησιμοποιούνται για την αναφορά σε ομάδες αντικειμένων με βάση τα χωρικά τους χαρακτηριστικά τα οποία, παράλληλα, μπορεί να είναι είτε στατικά είτε κινούμενα ανάλογα το νοηματικό πλαίσιο. Για παράδειγμα η χειρομορφή «B|» (βλ. Παράρτημα Β′) μπορεί να χρησιμοποιηθεί για να περιγράψει οποιοδήποτε αντικείμενο με λεία επιφάνεια όπως είναι η λεία επιφάνεια ενός τραπεζιού (Σχήμα 2.2), με τη χειρομορφή «C» δίνεται συχνά η έννοια του κυλινδρικού αντικειμένου (Σχήμα 2.3β′) και αντίστοιχα με τη χειρομορφή «1» η έννοια ενός μακριού ή λεπτού αντικειμένου. Το παράδειγμα που φαίνεται στο Σχήμα 2.3 είναι μία πολύ χαρακτηριστική περίπτωση χρήσης ταξινομητών. Αρχικά, γίνεται αναφορά στη λέξη ποτήρι στο Σχήμα 2.3α′. Αφού έχει οριστεί το αντικείμενο (ποτήρι), μπορεί ο νοηματιστής να συνδυάσει τον ταξινομητή για το κυλινδρικό σχήμα σε συνδιασμό με την κίνηση ενός αντικειμένου που πέφτει (Σχήμα 2.3β′) για να αποδώσει το ότι το ποτήρι έπεσε. Έπειτα, αποδίδοντας το νόημα της «έκρηξης» (δεν φαίνεται στο σχήμα) έχει ορίσει οπτικά το προσκήνιο ενός ποτηριού που έπεσε και θραύσματα απλώθηκαν στο πάτωμα. Τέλος, το ρήμα «περπατώ» είναι το κύριο ρήμα της δεύτερης πρότασης και ο νοηματιστής δίνει ιδιαίτερη έμφαση στη δράση του πατήματος των πελμάτων (Σχήμα 2.3ε′ δανειζόμενος το ρήμα «πατώ» (Σχήμα 2.4) και χρησιμοποιώντας το σαν ταξινομητή αφού το αποδίδει κινώντας τα δύο χέρια και επαναλαμβάνοντας περισσότερες από μία φορές τη δράση. Επιπλέον, αξίζει να σημειωθεί ότι τα 2 νοήματα που αποδίδονται αυτούσια από το λεξιλόγιο («ποτήρι»,«περπατώ») πραγματοποιούνται σε ουδέτερο χώρο νοημάτισης ενώ οι ταξινομητές εντοπίζονται χωρικά κοντά στο κατώτερο σημείο του χώρου νοημάτισης δίνοντας έμφαση στα γεγονότα που συμβαίνουν στο χαμηλότερο σημείο του πλαισίου αναφοράς της διήγησης, δηλαδή στην ύπαρξη κομματιών γυαλιού από το σπασμένο ποτήρι και τα πέλματα που πατούν στο πάτωμα.

Αξίζει να παρατηρηθεί, ακόμα, ότι το ουσιαστικό «ποτήρι» οπτικά ομοιάζει στον αντίστοιχο ταξινομητή, όπως είναι αναμενόμενο, λόγω της χειρομορφής «C» που κυριαρχεί και στα δύο. Ωστόσο φωνολογικά είναι πολύ διαφορετικά νοήματα διότι στη μία περίπτωση πρόκειται για νόημα που πραγματοποιήται με τα δύο χέρια ενώ το άλλο μόνο με το κυρίαρχο, αντίστοιχα. Ιδιαίτερο γλωσσολογικό ενδιαφέρον έχουν οι περιπτώσεις όπου μία λέξη κάποιας Νοηματικής Γλώσσας χρησιμοποιείται σαν ταξινομητής σε κάποια άλλη. Για παράδειγμα στο Σχήμα 2.5α′ φαίνεται η απεικόνιση του



Σχήμα 2.2: «Τραπέζι (επιφάνεια)» (Ελληνική Νοηματική)

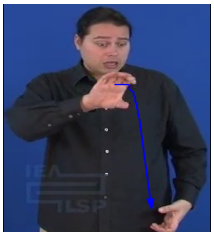
νοήματος «καρέκλα» και την ίδια στιγμή στο Σχήμα 2.5β′ φαίνεται ο όρος «καρέκλα» εκφρασμένος ως οντολογικός ταξινομητής μέσα από ένα καρέ της έκφρασης «σειρά από 7 καρέκλες». Η χειρομορφή του ταξινομητή είναι η ίδια με τη χειρομορφή που χρησιμοποιείται για το νόημα «καρέκλα» της Γερμανικής Νοηματικής Γλώσσας που φαίνεται στο Σχήμα 2.5γ′.

Οι ταξινομητές που αναφέρονται σε μέρη του σώματος θα μπορούσαν να αναφερθούν και ως ταξινομητές οντοτήτων. Στο Σχήμα 2.3ε′ ο ταξινομητής μπορεί να θεωρηθεί επίσης και ταξινομητής μέρους του σώματος και συγκεκριμένα του πέλματος. Εκείνο στο οποίο διαφοροποιούνται τα δύο είδη ταξινομητών είναι πως η δεύτερη κατηγορία είναι τόσο ειδική και συνεπώς δεν απαιτεί τον ορισμό του αντικειμένου αναφοράς.

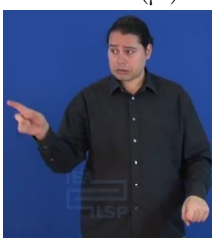
Οι ταξινομητές χειρισμού εμφανίζονται με ρήματα που αφορούν το κράτημα ή τη κίνηση μέσω του χεριού ενός αντικειμένου. Σε αντίθεση με τους ταξινομητές της οντότητας και του σώματος, αντιπροσωπεύουν την οντότητα που αναφέρονται έμμεσα, καθώς αντιπροσωπεύουν μόνο το τμήμα του χειριζόμενου αντικειμένου, για παράδειγμα, το στέλεχος ενός λουλουδιού ή τη λαβή ενός μαχαιριού. Με άλλα λόγια, κωδικοποιούν μια εικονική όψη που συνδέεται με μια ενέργεια που εμπλέκει το θέμα ενός ρήματος, αλλά δεν αντανακλούν τα χαρακτηριστικά του θέματος περ σε. Μερικές φορές το θέμα είναι απλά ένα αντικείμενο που κρατείται ή μεταφέρεται. Για παράδειγμα η λέξη μηχανικός αναπαριστάται ως ένα κατσαβίδι που βιδώνει κάτι. Το κυρίαρχο χέρι αναπαριστά έναν ταξινομητή του κατσαβιδιού ενώ το δευτερεύον μένει σταθερό αναπαριστώντας το αντικείμενο που βιδώνεται. Το ενδιαφέρον είναι ότι η λέξη «μηχανικός» έχει δύο μορφές με χρησιμοποιώντας ταξινομητή οντότητας και ταξινομητή χειρισμού, αντίστοιχα.

Οι ταξινομητές Μεγέθους-Σχήματος, όπως δηλώνει το όνομά τους, χρησιμοποιούνται για να περιγράψουν οικογένειες σχημάτων με παρόμοιο σχήμα (Σχήμα 2.7β′) ή για να δηλώσουν μέγεθος (Σχήματα 2.7α′,2.7γ′). Οι περιγραφητές αυτοί μπορεί να είναι στατικοί ή ιχνογραφικοί (δυναμικοί). Οι έννοιες σχήματος και μεγέθους κατηγοριοποιούνται μαζί διότι σχεδόν πάντα αποδίδονται μέσα από το ίδιο νόημα. Συνχά, οι ταξινομητές αυτοί μπορεί να θεωρούνται και ταξινομητές οντότητας, ιδιαίτερα στην περίπτωση των στατικών νοημάτων. Η κύρια διαφορά ανάμεσα στους περιγραφητές

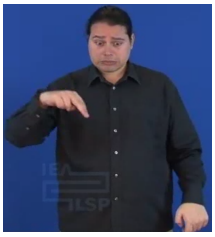
(α′) ποτήρι



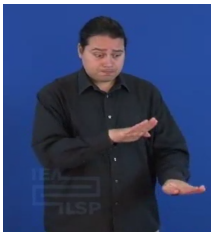
(β′) ποτήρι που πέφτει (ταξινομητής)



(γ′) Εσύ



(δ′) περπατώ



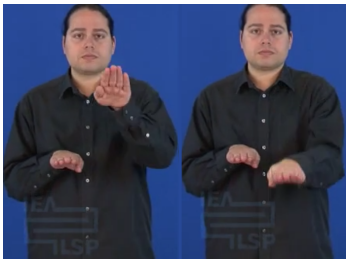
(ε′) πέλματα που πατούν (ταξινομητής)

Σχήμα 2.3: Αποσπάσματα από πρότασης της ΕΝΓ.
πιστή μετάφραση:«Ποτήρι (classifier) Σπάει Περπάτα (classifier) Προσοχή»
ελεύθερη μετάφραση:«Ένα ποτήρι έπεσε στο πάτωμα και έσπασε, πρόσεχε όταν περπατάς πού πατάς»

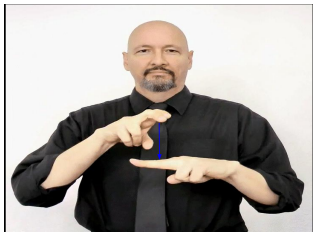
σχήματος και τους οντολογικούς ταξινομητές ή τους ταξινομητές μερών του σώματος είναι εμφανής όταν οι πρώτοι είναι δυναμικοί. Οι δυναμικοί περιγραφητές Σχήματος-Μεγέθους συνδυάζονται με ουσιαστικά λειτουργώντας σαν επίθετα σε αντίθεση με τους οντολογικούς ταξινομητές που, όπως αναφέραμε, συνδυάζονται με ρήματα.

2.2.2 Δακτυλικό Αλφάβητο

Όσο πλούσιο και αν είναι το λεξιλόγιο μίας Νοηματικής Γλώσσας, ακόμα και με την εκφραστική ελευθερία που προσφέρουν οι classifiers, είναι αδύνατον να υπάρξει μία πλήρης αντιστοίχιση ανάμεσα σε κάποια νοηματική γλώσσα και μία ομιλούμενη γλώσσα πολύ απλά διότι δεν είναι δυνατόν να υπάρχει ένα νόημα για οποιαδήποτε οντότητα εκφράζεται μέσα από μία γλώσσα. Η πιο συχνή περίπτωση λέξεων οι οποίες είναι αδύνατον να υποστηρίζονται πλήρως από μία Νοηματική Γλώσσα αποτελούν τα



Σχήμα 2.4: «Πατώ/βήμα» (Ελληνική Νοηματική)



(α′) καρέκλα (ΑΝΓ)

(β′) ταξινομητής για την έννοια «καρέκλα» (ΑΝΓ)

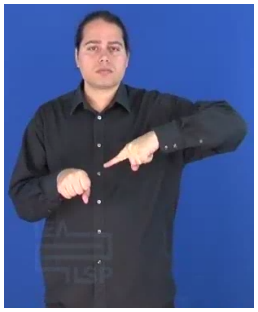
(γ′) καρέκλα (ΓΝΓ)

Σχήμα 2.5: Το ουσιαστικό «καρέκλα» ως ουσιαστικό και ως ταξινομητής σε διαφορετικές Νοηματικές Γλώσσες

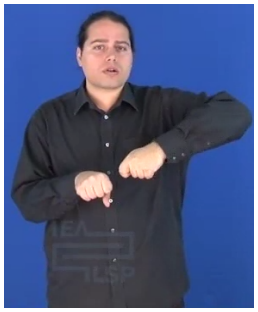
κύρια ονόματα και οι συντομογραφίες. Για το γεφύρωμα αυτού του χάσματος κάθε Νοηματική Γλώσσα εμπεριέχει ένα **δακτυλικό αλφάβητο** για την απόδοση του αντίστοιχου αλφαβήτου γραφής της ομιλούμενης γλώσσας. Βέβαια, το δακτυλικό αλφάβητο αποδίδεται σχεδόν πάντα με χρήση των χειρομορφές που εμφανίζονται σε μία Νοηματική Γλώσσα. Κάθε γράμμα συνήθως αντιστοιχίζεται αμφιμονοσήμαντα με μία χειρομορφή η οποία αποδίδεται στατικά με ελάχιστες εξαιρέσεις (π.χ. το γράμμα J στο Αμερικάνικο Δακτυλικό Αλφάβητο). Ωστόσο, υπάρχουν Νοηματικές Γλώσσες που διαφέρουν καθολικά στον τρόπο απεικόνισης του Δακτυλικού Αλφαβήτου. Η Βρετανική Νοηματική Γλώσσα χρησιμοποιεί και τα δύο χέρια για κάθε γράμμα ώστε να ομοιάζει οπτικά κατά το δυνατόν περισσότερο. Παραδείγματα Δακτυλικών αλφαβήτων παρατίθενται στο Παράρτημα Β′. Ειδικότερα, οι περιπτώσεις που μπορεί να εμφανιστεί ο συλλαβισμός σε δακτυλικό αλφάβητο (fingerspelling) είναι οι εξής:

- Απόδοση κυρίων ονομάτων γράμμα προς γράμμα. Για να συστηθεί κάποιος σε μία Νοηματική Γλώσσα θα πρέπει να «δακτυλοσυλαβίσει» (fingerspell) το όνομά του στο δακτυλικό αλφάβητο και συνήθως αυτό είναι το πρώτο πράγμα που μαθαίνει ένας ομιλούντας ξεκινώντας την εκμάθηση μίας Νοηματικής Γλώσσας.

- Νοήματα που χρησιμοποιούν το πρώτο γράμμα της αντίστοιχης λέξης της προφορικής γλώσσας. Για παράδειγμα στην Αμερικανική Νοηματική Γλώσσα για τη λέξη «νερό» χρησιμοποιείται η χειρομορφή W (εκ του **w**ater) (Σχήμα 2.8).



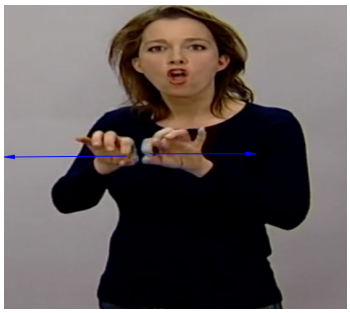
(α΄) Χρήση ταξινομητή οντότητας



(β΄) Χρήση ταξινομητή χειρισμού

Σχήμα 2.6: Η λέξη «μηχανικός» εκφρασμένη με δύο τρόπους χρησιμοποιώντας ταξινομητή για την έννοια «κατσαβίδι»

(α΄) Μικρό στρογγυλό αντικείμενο



(β΄) σωληνοειδές

(γ΄) μεγάλο σωληνοειδές

Σχήμα 2.7: Παραδείγματα ταξινομητών Σχήματος-Μεγέθους στην Δανική Νοηματική Γλώσσα

- Νοήματα που περιγράφονται απλά από το αρχικό γράμμα της προφορικής γλώσσας. Στο Σχήμα 2.9 φαίνεται το νόημα της Βρετανικής Νοηματικής Γλώσσας για τη λέξη «κόρη» που ταυτίζεται με το δακτυλικό γράμμα *D* (εκ του **D**aughter).
- Ονόματα και Εξειδικευμένοι όροι. Για παράδειγμα, προφανώς δεν υπάρχει κάποιο νόημα για τεχνικούς όρους όπως είναι το perceptron που αποτελεί τεχνικό όρο της Μηχανικής Μάθησης.

2.3 Τεχνολογίες Επισημείωσης

Το Hamburg Notation System (HamNoSys) [8] είναι ένα από τα πιο διαδεδομένα σύστηματα επισημείωσης για νοηματικές γλώσσες. Αναπτύχθηκε από το Institute of German Sign Language and Communication of the Deaf, Hambourg University

Σχήμα 2.8: «Νερό» (Αμερικανική Νοηματική Γλώσσα)

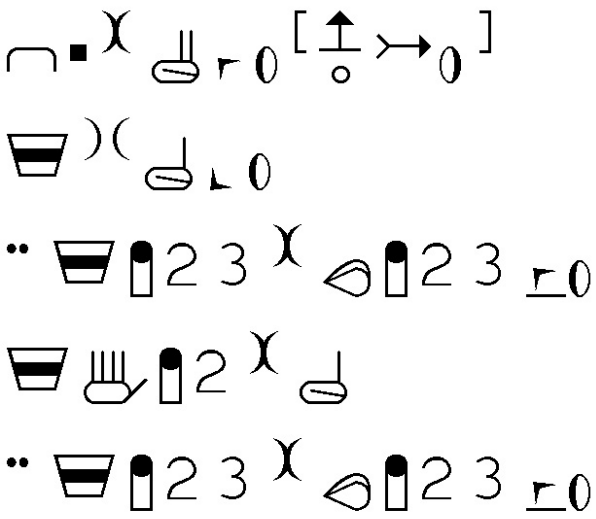
Σχήμα 2.9: «Κόρη» (Βρετανική Νοηματική Γλώσσα)

και είναι μία συνέχεια της δουλειάς του Dr. Stokoe [28] από τη δεκαετία του 1960, ο οποίος ανέπτυξε το αντίστοιχο σύστημα για τις ανάγκες της γραπτής περιγραφής της Αμερικάνικης Νοηματικής Γλώσσας. Το σύστημα αυτό είναι εμπνευσμένο από την ανάλυση σε φωνήματα που χρησιμοποιείται για την αναπαράσταση των ομιλούμενων γλωσσών. Για αυτό το λόγο, το HamNoSys αναφέρεται συνήθως στη βιβλιογραφία ως μία «φωνολογική» (phonological) περιγραφή.
Το σύστημα αυτό έχει μεγάλη απήχηση στο πεδίο ανάπτυξης συστημάτων σύνθεσης Νοηματικής Γλώσσας με χρήση τεχνολογίας Avatar. Πρακτικά, αφορά μία ερευνητική περιοχή που αφορά την αντίστροφη διαδικασία από αυτή που ακολουθά η παρούσα εργασία καθώς πρόκειται για μία διαδικασία μετατροπής κειμένου σε βίντεο. Δημιουργήθηκε με γνώμονα έξι βασικές βασικές προϋποθέσεις:

• Διεθνής χρήση: Πρωταρχικός σκοπός του συστήματος είναι να λειτουργεί σαν ένα ενιαίο σύστημα επισημείωσης που θα έχει τη δυνατότητα να περιγράψει οποιαδήποτε Νοηματική Γλώσσα. Έτσι, για παράδειγμα σε επίπεδο χειρομορφών, η περιγραφή

δεν βασίζεται στο δακτυλικό αλφάβητο κάποιας συγκεκριμένης γλώσσας αλλά στην οπτική περιγραφή του σχήματος του χεριού. Έτσι, το HamNoSys πέρα από βάσεις δεδομένων της Γερμανικής Νοηματικής Γλώσσας έχει χρησιμοποιηθεί ακόμα για άλλες γλώσσες όπως η Ελληνική, η Δανική, η Ελβετική-Γερμανική, η Αμερικανική και η νοηματική της Νέας Ζηλανδίας.

•Εικονική Αναπαράσταση (Iconicity): Για τους σκοπούς της συμβολικής αναπαράστασης, οι δημιουργοί του HamNoSys επέλεξαν να μην χρησιμοποιήσουν ήδη υπάρχοντες χαρακτήρες αλλά ανέπτυξαν μια σειρά από πρωτότυπους ειδικούς οπτικούς χαρακτήρες (glyphs) οι οποίοι εμφανισιακά θυμίζουν ιερογλυφικά (Σχήμα 2.10). Η επιλογή αυτή έγινε αφενώς διότι οι απαιτούμενοι χαρακτήρες για όλα τα δομικά στοιχεία είναι πάρα πολλοί σε αριθμό για να χρησιμοποιηθεί κάποιο προϋπάρχον αλφάβητο και αφετέρου η χρήση ειδικών συμβόλων βοηθά ιδιαίτερα στην απομνημόνευση και ανάγνωση αυτών. Άλλωστε, η επιλογή αυτή είναι σαφής αν σκεφτεί κανείς ότι σε αντίθεση με την γραφή των ομιλούμενων γλωσσών, η συγκεκριμένη περίπτωση αποτελεί μία απεικόνιση οπτικής πληροφορίας σε οπτική πληροφορία.



Σχήμα 2.10: Παράδειγμα αναπαράστασης HamNoSys

• Οικονομία: Το σύστημα έχει σκοπό να κάνει σύμπτυξη πληροφορίας όπου αυτό είναι εφικτό. Έτσι, χρησιμοποιούνται ειδικοί χαρακτήρες που περιγράφουν το είδος της συμμετρίας ανάμεσα στα δύο χέρια ώστε να παραλείπεται η περιγραφή για το μη κυρίαρχο χέρι σε περιπτώσεις συμμετρικών νοημάτων ή νοημάτων που κάνουν χρήση του ενός χεριού μόνο. Το κυριότερο χαρακτηριστικό του HamNoSys, προκειμένου να αποφεύγεται ο βερμπαλισμός, είναι η ύπαρξη σύμβασης για την ουδέτερη κατάσταση στα περισσότερα χαρακτηριστικά. Πιο συγκεκριμένα, όταν δεν γίνεται λόγος για τη θέση του αντίχειρα, το λύγισμα των δακτύλων ή τη διάταξη του χεριού, σε κάθε περίπτωση το αντίστοιχο χαρακτηριστικό θεωρείται ότι έχει την συνηθισμένη του μορφή.

Ομοίως, η απουσία κίνησης υποδηλώνεται με την παράλειψη αναφοράς στο αντίστοιχο χαρακτηριστικό.

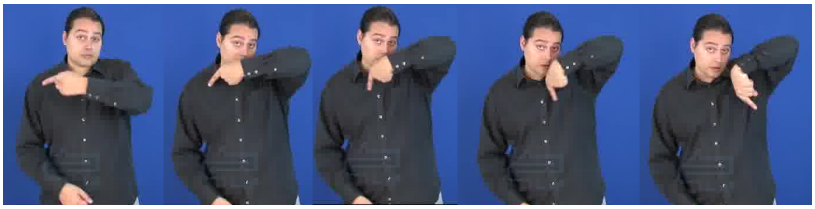
• Δυνατότητα ολοκλήρωσης σε προγράμματα Η/Υ: Το HamNoSys έχει μεγάλη συμβατότητα με ένα πλήθος προγραμμάτων που χρησιμοποιούνται για επισημείωση σε Η/Υ. Τα ποιο χαρακτηριστικά παραδείγματα αυτών είναι το ELAN και το iLex. Τα προγράμματα αυτά παρέχουν τη δυνατότητα σε κάποιον επισημειωτή να προσθέσει τη μετάφραση κατά λέξη σε ένα βίντεο νοηματικής με παρόμοιο τρόπο όπως γίνεται σε αντίστοιχα προγράμματα υποτιτλισμού. Επιπλέον, υποστηρίζεται η μετατροπή του HamNoSys από συμβολική αναπαράσταση σε μία αναπαράσταση αρχείου μορφής XML. Αυτή η μορφή είναι που χρησιμοποιείται εκτενώς στα πλαίσια αναπαράστασης Νοηματικής Γλώσσας μέσω Avatar. Τέλος, οι δημιουργοί του συστήματος έχουν αναπτύξει μία γραμματοσειρά συμβατή με προγράμματα Windows για την απεικόνισή του.

• Συντακτικό: Η επισημείωση έχει κατά βάση ένα καλώς ορισμένο συντακτικό. Παρότι ο σκοπός του συστήματος είναι να πραγματοποιήσει την περιγραφή μίας οπτικής αναπαράστασης, οι δημιουργοί του έχουν ορίσει ένα συγκεκριμένο τρόπο γραφής ο οποίος σχεδόν πάντα ακολουθείται πιστά από όλους τους επισημειωτές. Ωστόσο, το πλούσιο εικονικό αλφάβητο δίνει μεγάλη εκφραστική ελευθερία στον επισημειωτή, με αποτέλεσμα να εμφανίζονται μικρές διαφορές στον τρόπο που διαφορετικές ομάδες περιγράφουν κάποιες διατάξεις ή κινήσεις.

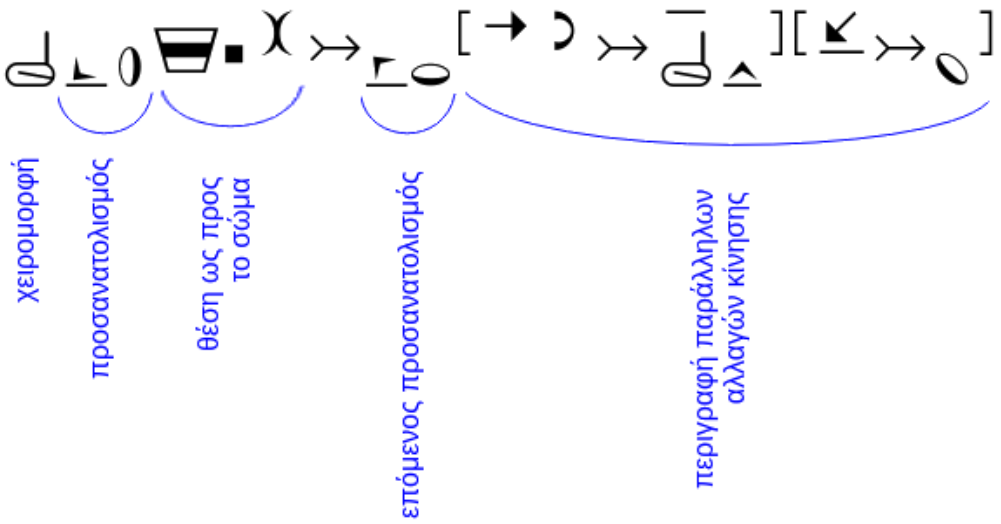
• Συμβατότητα: Το HamNoSys, παρότι είναι σε ανάπτυξη εδώ και μερικές δεκαετίες καταφέρνει να παραμένει συμβατό με όλες τις προηγούμενες εκδόσεις. Παράλληλα, συνεχώς προστίθενται περισσότερα χαρακτηριστικά για την όλο και καλύτερη αναπαράσταση των Νοηματικών Γλωσσών.

Συντακτικά το HamNoSys διαφοροποιείται ανάλογα με το αν πρόκειται για νόημα που περιλαμβάνει ένα μόνο χέρι ή και τα δύο. Στην περίπτωση όπου μονάχα το κυρίαρχο χέρι νοηματίζει, η έκφραση ξεκινά με την αρχική διάταξη του χεριού (σχήμα, προσανατολισμός και θέση), έπειτα, περιγράφεται η κίνηση και τέλος επαναλαμβάνονται όσα στοιχεία μπορεί να αλλάξουν. Οι αλλαγές που εμφανίζονται εντός ενός νοήματος μπορεί να είναι είτε λόγω κίνησης του χεριού είτε λόγω αλλαγής στη διάταξη της παλάμης. Όταν υπάρχει κίνηση, η περιγραφή μπορεί να είναι από απλή αναφορά στο σχήμα της τροχιάς της κίνησης (π.χ. κίνηση προς τα μέσα) μέχρι μία πολύ εκτενής περιγραφή με μία αλληλουχία από χαρακτήρες που περιγράφουν αναλυτικά την μετάβαση σε καινούργια διάταξη του χεριού, την κίνηση και παρενθέσεις ή άγκιστρα ανάλογα με το αν πρόκειται για αλλαγές που συμβαίνουν σειριακά ή παράλληλα, αντίστοιχα. Για παράδειγμα στο Σχήμα 2.11 φαίνεται ένας νοηματιστής της Ελληνικής Νοηματικής που αναπαριστά τη λέξη «ΕΜΕΙΣ» και στο Σχήμα 2.12 φαίνεται η αντίστοιχη αναπαράσταση σε HamNoSys. Παρατηρούμε ότι πρόκειται για ένα απλό οπτικά νόημα το οποίο έχει μία ιδιαίτερα περίπλοκη περιγραφή.

Για τα νοήματα που απαιτούν και τα δύο χέρια, προστίθεται στην αρχή της αναπαράστασης ένα σύμβολο που περιγράφει την ύπαρξη ή μη και το είδος της συμμετρίας που εμφανίζεται ανάμεσά τους. Η μη ύπαρξη συμμετρίας σημαίνει, φυσικά, ότι τα χέρια θα διαφέρουν ως προς το σχήμα ή/και το προσανατολισμό ή/και την κίνηση.



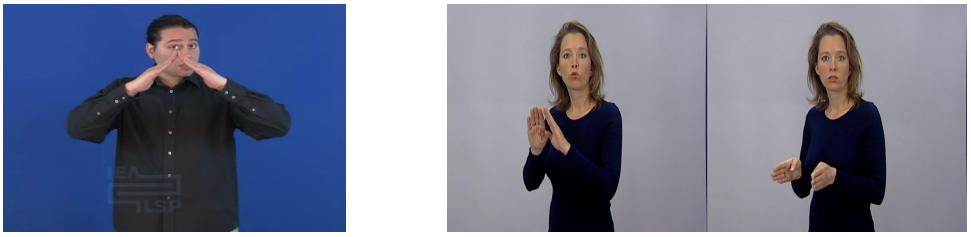
Σχήμα 2.11: «ΕΜΕΙΣ» (Ελληνική Νοηματική)



Σχήμα 2.12: «ΕΜΕΙΣ» εκφρασμένο στο HamNoSys

Συνήθως, πρόκειται για διαφορετικές χειρομορφές ή για περιπτώσεις όπου το κυρίαρχο χέρι κινείται ενώ το δευτερεύον μένει ακίνητο και λειτουργεί επικουρικά παίζοντας το ρόλο του αντικειμένου. Όταν δεν υπάρχει συμμετρία, προκύπτει η ανάγκη να δοθεί η περιγραφή του κάθε χεριού ξεχωριστά. Έτσι, το HamNoSys χρησημοποιεί μία παραλλαγή του τελεστή «συν» (+). Η έκφραση για την περιγραφή των 2 χεριών γίνεται εντός παρένθεσης, πρώτα γίνεται αναφορά στο κυρίαρχο χέρι, ακολουθεί ο τελεστής «+» και τέλος η περιγραφή του δευτερεύοντος χεριού. Ένα πολύ χαρακτηριστικό παράδειγμα συμμετρικού νοήματος είναι διεθνώς η λέξη **«ΣΠΙΤΙ»**. Στην ΕΝΓ αποτελεί ένα σχεδόν στατικό νόημα όπου η συμμετρία των χεριών χρησιμοποιείται για να αναπαρασταθεί το σχήμα της σκεπής (Σχήμα 2.13α′), ενώ στη ΔΝΓ και τη ΓΝΓ είναι ένα δυναμικό νόημα που αναπαριστά όλο σχήμα ενός σπιτιού (Σχήμα 2.13β′). Επιπροσθέτως, πολύ απλά παραδείγματα στατικών ασύμμετρων νοημάτων είναι οι αριθμοί από το 6 έως το 9 καθώς η αρίθμηση γίνεται με επέκταση των δακτύλων όπως είναι οικείο και στους μη χρήστες νοηματικών γλωσσών (Σχήμα 2.14α′). Επιπλέον, οι έννοιες «ΑΔΙΕΞΟΔΟ»/«ΕΜΠΟΔΙΟ»/«ΤΕΡΜΑΤΙΖΩ» είναι ένα παράδειγμα όπου το

δευτερεύον χέρι λειτουργεί επικουρικά μένοντας ακίνητο. Εν προκειμένω, το δευτερεύον χέρι αναπαριστά ένα κατακόρυφο εμπόδιο που βρίσκεται ακίνητο και το κυρίαρχο χέρι κινείται μέχρι που προσκρούει στο δευτερεύον αποδίδοντας οπτικά το ζητούμενο (Σχήμα 2.14β′).



(α′) (β′)

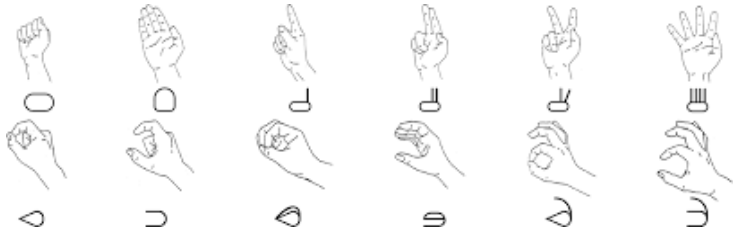
Σχήμα 2.13: Παράδειγμα συμμετρικού νοήματος: «ΣΠΙΤΙ» (α′) Ελληνική Νοηματική (β′) Δανική Νοηματική

(α′) (β′)

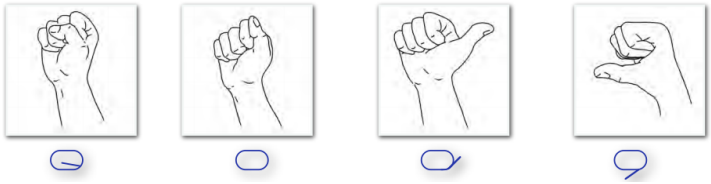
Σχήμα 2.14: Παραδειγμα ασύμμετρων Νοημάτων (α′) 7 (β′) Εμπόδιο/Αδιέξοδο (Ελληνική Νοηματική)

Για την περιγραφή των χειρομορφών το HamNoSys αφιερώνει τρεις κατηγορίες χαρακτηριστικών. Το βασικό χαρακτηριστικό που ποτέ δεν απουσιάζει είναι η βασική χειρομορφή για την οποία υπάρχουν 12 επιλογές (Σχήμα 2.15). Στη συνέχεια, εάν η θέση του αντίχειρα διαφέρει από την καθιερωμένη, προσδιορίζεται η θέση του αντίχειρα. Οι δυνατές επιλογές για τη θέση του αντίχειρα είναι «εγκάρσια»,«ουδέτερη»,«ανοικτή» και «εμπρόσθια» (Σχήμα 2.16). Το τελευταίο κύριο χαρακτηρηστικό είναι το πόσο λυγισμένα είναι τα υπόλοιπα 4 δάκτυλα. Εδώ οι επιλογές είναι «ουδέτερα», «λυγισμένα», «διπλά λιγισμενα», «αγκυλωμένα», «διπλά αγκυλωμένα» και «ευθεία» (Σχήμα 2.17). Στην περίπτωση της ουδέτερης θέσης στα δύο τελευταία χαρακτηριστικά, απλά αγνοείται η αναφορά στο αντίστοιχο χαρακτηριστικό. Τέλος, όπου απαιτείται μπορεί να γίνει αναφορά σε κάθε δάκτυλο ξεχωριστά. Για παράδειγμα στην τέταρτη γραμμή από το Σχήμα 2.10 περιγράφεται ότι το κάτω μέρος από το δείκτη (δάκτυλο 2) ακουμπά το δείκτη από το δευτερεύον χέρι (το οποίο έχει τη χειρομορφή 1).

Το επόμενο χαρακτηριστικό είναι η περιγραφή του προσανατολισμού του χεριού. Αρχικά, περιγράφεται η κατεύθυνση των δακτύλων. Ως κατεύθυνση των δακτύλων



Σχήμα 2.15: Οι βασικές χειρομορφές στο HamNoSys

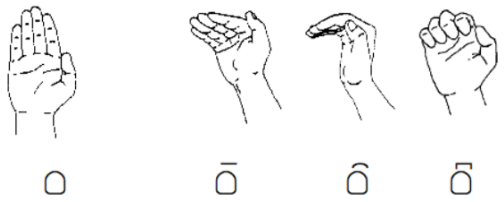


Σχήμα 2.16: Παράδειγμα διατάξεων του αντίχειρα στο HamNoSys στην πλαίσια της «κατηγορίας γροθιά». Στην εικόνα εμφανίζονται με την σειρά οι θέσεις «εγκάρσια»,«ουδέτερη»,«ανοικτή»,«εμπρόσθια», αντίστοιχα.

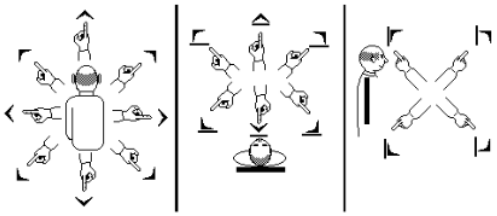
ορίζεται το πού δείχνει ο νοηματιστής αν έχει εκτεταμένο το δείκτη. Συνεπώς, υπάρχουν 3 βαθμοί ελευθερίας που περιγράφουν τον προσανατολισμό των δακτύλων (Σχήμα 2.18). Επιπλέον, από τη στιγμή που το χέρι είναι ένα γεωμετρικό στερεό απαιτείται ακόμα ένας βαθμός ελευθερίας για να γνωρίζουμε την πλήρη διάταξη του χεριού. Για αυτό το λόγω δίνεται ακόμα ο προσανατολισμός της παλάμης με αναφορά τα δάκτυλα (Σχήμα 2.19). Φυσικά, ο ορισμός δεν έχει μία άρτια μαθηματική θεμελίωση, ωστόσο μπορούμε εύκολα διαισθητικά να αντιληφθούμε ότι για κάθε κατεύθυνση των δακτύλων, θεωρούμε τον «κάτω» προσανατολισμό της παλάμης ως προς τη θέση όπου η παλάμη «κοιτάει» είτε μακρυά από το σώμα είτε προς τα κάτω.

Το τρίτο χαρακτηριστικό που περιγράφεται στο HamNoSys είναι η θέση του κυρίαρχου χεριού 2.20. Ομοίως με το σχήμα και τον προσανατολισμό, η θέση ως προς το σώμα περιγράφεται από δύο αναφορές στο HamNoSys. Η πρώτη αφορά στον εντοπισμό του χεριού ως προς το πλάτος και ύψος της θέσης του χεριού. Δίνεται σε σχέση με κάποιο σημείο του κορμού ή του κεφαλιού. Το δεύτερο στοιχείο αφορά την απόσταση του χεριού από το κορμό ή το κεφάλι ως προς το βάθος. Εάν το δεύτερο στοιχείο απουσιάζει, τότε πρόκειται για μία «φυσιολογική» απόσταση του χεριού από το σώμα. Ενδεχομένως, σε κάποιες περιπτώσεις απουσιάζει εξ ολοκλήρου η πληροφορία για τη θέση. Τότε, θεωρείται ότι το χέρι βρίσκεται σε ένα «ουδέτερο χώρο νοημάτισης» (natural signing space).

Στις περιπτώσεις όπου δεν έχουμε συμμετρική νοημάτιση, για λόγους απλούστευσης δεν περιγράφεται ξεχωριστά η θέση για το μη κυρίαρχο χέρι αλλά περιγράφεται ως προς το κυρίαρχο. Χρησιμοποιούνται έννοιες όπως η επαφή ή διασταύρωση των δύο χεριών. Άλλωστε, όπως αναλύσαμε παραπάνω κατά την ασύμμετρη νοημάτιση το δευτερεύον χέρι λειτουργεί προσθέτοντας οπτική πληροφορία αλληλεπιδρώντας ή



Σχήμα 2.17: Παράδειγμα ως προς το λύγισμα των δακτύλων στο HamNoSys στα πλαίσια της «κατηγορίας παλάμη». Εδώ φαίνονται με τη σειρά «ουδέτρα», «ευθεία», «διπλά λυγισμένα» και «αγκύλα»

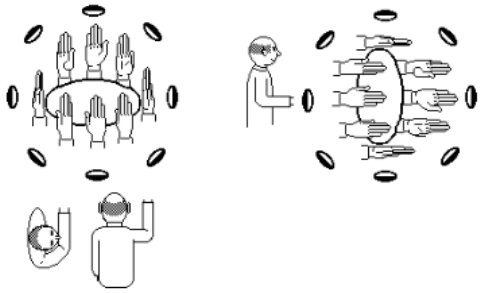


Σχήμα 2.18: Δυνατές διατάξεις για τον προσανατολισμό των δακτύλων.

συμπληρώνοντας το κυρίαρχο χέρι.

Το τελευταίο και πιο δύσκολο κύριο στοιχείο ενός νοήματος είναι η απόδοση των αλλαγών κατά τη διάρκεια αυτού. Σε πρώτο επίπεδο η κίνηση του χεριού μπορεί να περιγραφεί με παρόμοιο τρόπο με αυτόν του προσανατολισμού. Για ακόμη μία φορά έχουμε το κβαντισμό στις τρεις διευθύνσεις του χώρου (μέσα/έξω, δεξιά/αριστερά και πάνω κάτω) και τις ενδιάμεσες (διαγώνιες) αυτών (Σχήμα 2.21). Για αναλυτικότερες περιγραφές το HamNoSys περιέχει πληθώρα συμβόλων που αναφέρονται στο μέγεθος ή την ένταση της κίνησης. Επιπλέον, είναι διαθέσιμη μία οικογένεια από συνήθης τροχιές κίνησης όπως για παράδειγμα το ζιγκ-ζαγκ, ο κυματισμός, ο ορολογιακός και αντιορολογικός κύκλος κτλ. Πιο σύνθετα σχήματα μπορούν να περιγραφούν με τη χρήση παρενθέσεων ή αγκυλών δηλώνοντας ότι δύο κινήσεις πραγματοποιούνται σειριακά ή παράλληλα, αντίστοιχα. Επιπλέον, άλλα σύμβολα χρησιμοποιούνται για να δηλώσουν κίνηση επί τόπου. Η κινήσεις αυτές αφορούν τον καρπό. Τέλος, το HamNoSys εμπεριέχει σύμβολα που υποδηλώνουν επαναλήψεις κινήσεων είτε δηλώνοντας τον αριθμό αυτών είτε αόριστα (π.χ. «μερικές φορές»).

Στο Σχήμα 2.12 φαίνεται μία αναπαράσταση που χρησιμοποιεί το μεγαλύτερο φάσμα του HamNoSys αναφορικά με την κίνηση. Στο κομμάτι της παράλληλης κίνησης έχουμε κίνηση δεξιά του χεριού και παράλληλα ημικυκλική κίνηση του καρπού που καταλήγει σε τροποποίηση της αρχικής χειρομορφής με το δείκτη λυγισμένο «ευθεία», ενώ ακολουθεί ακόμα μία παράλληλη κίνηση προς τα κάτω αριστερά καθώς παράλληλα πραγματοποιείται περιστροφή της παλάμης ώστε να καταλήξει στραμένη προς τα έξω



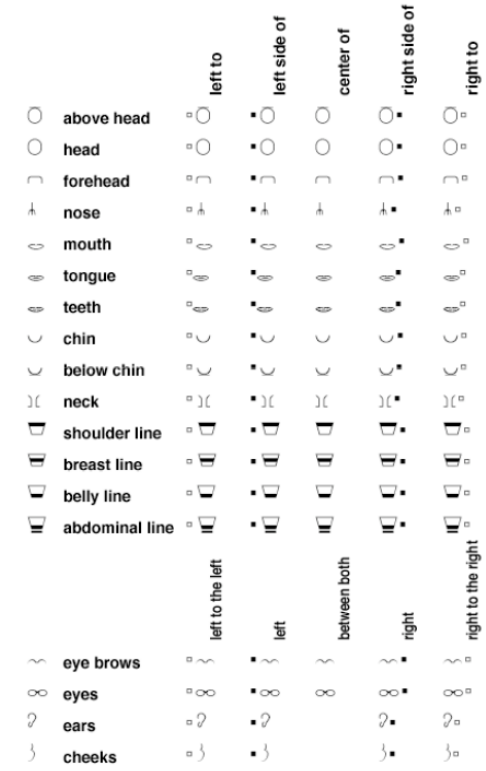
Σχήμα 2.19: Δυνατές διατάξεις για τον προσανατολισμό της παλάμης ως την κατεύθυνση των δακτύλων.

στο τέλος του νοήματος [2].

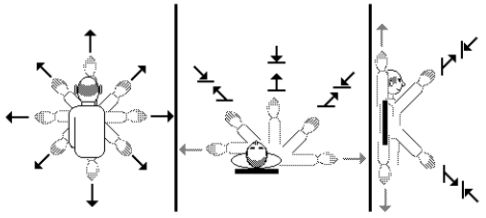
Το τελευταίο στοιχείο που περιγράφεται από το HamNoSys είναι τα χαρακτηριστικά που δεν αφορούν τα χέρια (non-manual). Είναι το στοιχείο του οποίου περιγραφή αποτελεί τη δυσκολότερη καθώς συνήθως αφορά κινήσεις του προσώπου. Θα πρέπει να υπάρχει η δυνατότητα να γίνει περιγραφή κάθε σχήματος του στόματος, φουσκώματος των μάγουλων, έκφρασης στα μάτια και κίνησης στα φρύδια. Επιπλέον, τα χαρακτηριστικά αυτά επεκτείνονται σε κινήσεις του κεφαλιού και των ώμων.

Ολοκληρώνοντας την ανάλυση του HamNoSys πρέπει να γίνει αναφορά στον τελεστή «μεταξύ» (between) ο οποίος συμβολίζεται με την κάθετο «\». Ο τελεστής αυτός μπορεί να εφαρμοστεί σε οποιαδήποτε από τις πέντε ομάδες χαρακτηριστικών δηλώνοντας μία κατάσταση που βρίσκεται ενδιάμεσα σε δύο άλλες. Για παράδειγμα μία χειρομορφή με ελαφρώς ανοικτά δάκτυλα θα επισημειωθεί ώς .

[2]Στο Σχήμα 2.11 ο νοηματιστής είναι αριστερόχειρας, συνεπώς όλοι προσανατολισμοί απεικονίζονται κατοπτρικά αντεστραμμένοι.



Σχήμα 2.20: Θέσεις ως προς το σώμα και το κεφάλι στο HamNoSys



Σχήμα 2.21: Περιγραφές ευθείων κινήσεων στο HamNoSys

Κεφάλαιο 3

Βάσεις Δεδομένων σε διάφορες Νοηματικές Γλώσσες

3.1 Δανική Νοηματική Γλώσσα

Στα πλαίσια της παρούσας εργασίας η κύρια βάση δεδομένων που χρησιμοποιήθηκε για την εκπαίδευσης των μοντέλων της μηχανικής μάθησης, είναι μία συλλογή από βίντεο που αναπτύσσεται από το «Κέντρο για τη Νοηματική Γλώσσα και την επικοινωνία βασισμένη σε νεύματα, University College Capital» σε στενή συνεργασία με την Δανική Ένωση Κωφών.

Σκοπός του έργου είναι η δημιουργεία ενός λεξικού της Δανική Νοηματικής Γλώσσας το οποίο είναι ελεύθερο για χρήση μέσω της ιστοσελίδας www.tegnsprog.dk. Εκεί, μπορεί οποιοσδήποτε να αναζητήσει τα νοήματα που αντιστοιχούν σε λέξεις της ομιλούμενης Δανικής Γλώσσας και να δει τα βίντεο με τα αντίστοιχα λήμματα της Δανικής Νοηματικής Γλώσσας. Για λέξεις που έχουν περισσότερους από έναν τρόπο νοημάτισης, όπως η λέξη "SØNDAG" (Κυριακή), εμφανίζονται όλες οι παραλλαγές αυτής. Επίσης, για λέξεις που μπορεί να έχουν διαφορετικές ερμηνείες που αντιστοιχούν σε διαφορετικά νοήματα. Ιδιαίτερα χρήσιμη αποτελεί η δυνατότητα της αντίστροφης αναζήτησης. Σε περίπτωση που ο ενδιαφερόμενος δεν γνωρίζει την σημειολογία κάποιου νοήματος, μπορεί να κάνει αναζήτηση με βάση τα φωνολογικά χαρακτηριστικά.

Για να επιτευχθεί αυτό, κάθε βίντεο πέρα από τη μετάφρασή του, έχει επισημειωθεί επιπλέον σε επίπεδο «φωνολογικών» χαρακτηριστικών με χρήση του HamNoSys. Οι υπεύθυνοι για την επισημείωση επέλεξαν να χρησιμοποιήσουν μία παραλλαγή της μορφής XML του HamNoSys ώστε να είναι πιο προσιτή ως προς την συντακτική επεξεργασία από το σύστημα. Επίσης, η μέθοδος του HamNoSys έχει τροποποιηθεί σε σχέση με το πως ορίστηκε στην Ενότητα 2.3 για την περιγραφή της χειρομορφής. Κάθε χειρομορφή περιγράφεται μέσω ενός μόνον όρου βασισμένου στο αντίστοιχο δακτυλικό αλφάβητο όπως φαίνεται στο Σχήμα Β΄.2.

Συνολικά, υπάρχουν 2837 βίντεο διάρκειας περίπου 4 ωρών από λήμματα της Δα-

νικής Νοηματικής Γλώσσας εκ των οποίων τα 2715 είναι επισημειωμένα όπως περιγράφηκε παραπάνω. Επιπλέον, αποτελεί ένα πολύ καλό σύνολο δεδομένων για τους σκοπούς της μηχανικής μάθησης καθώς εμφανίζεται μεγάλος αριθμός διαφορετικών νοηματιστών και από τα δύο φύλα, οι οποίοι είναι στημένοι σε διαφορετικές γωνίες ως προς το φακό της κάμερας.

3.2 Ελληνική Νοηματική Γλώσσα

Οι μεγαλύτερες δράσεις για την καταγραφή της Ελληνικής Νοηματικής Γλώσσας έχουν πραγματοποιηθεί από το Ινστιτούτο Επεξεργασίας Λόγου (ΙΕΛ) το οποίο υπάγεται στο Ερευνητικό Κέντρο Καινοτομίας στις Τεχνολογίες της Πληροφορίας, των Επικοινωνιών και της Γνώσης «Αθηνά». Από τις αρχές του 2000 το ΙΕΛ έχει υλοποιήσει τις συλλογές Νόημα και Πολύτροπον [7][12].

3.2.1 Νόημα

Το έργο «Νόημα» ξεκίνησε ως μία συλλογή βίντεο με σκοπό τη δημιουργία ενός Λεξικού της Ελληνικής Νοηματικής Γλώσσας. Είναι το πρώτο ηλεκτρονικό λεξικό ελληνικών νοημάτων με μετάφραση στα Νέα Ελληνικά. Παράλληλα, αποτελεί την πρώτη ελληνική παραγωγή DVD-ROM. Περιέχει 3000 βιντεοσκοπημένες λέξεις και απευθύνεται σε γνώστες της νοηματικής αλλά και σε σπουδαστές που μαθαίνουν την ΕΝΓ ως δεύτερη γλώσσα.

Ακολουθώντας μια εργονομία που έχει ως στόχο τη βέλτιστη χρηστικότητα του προϊόντος ως προς τις δύο ομάδες χρηστών, κάθε λήμμα-νόημα συνοδεύεται από ερμήνευμα που περιλαμβάνει αφενός ένα μεταφραστικό ισοδύναμο και, αν χρειάζεται, περαιτέρω ερμηνεία στα Νέα Ελληνικά και αφετέρου συνώνυμα και αντώνυμα στην ΕΝΓ.

Η ενσωμάτωση ερμηνευμάτων και μεταφράσεων στα Νέα Ελληνικά βοηθά τους χρήστες που δεν είναι ομιλητές της ΕΝΓ να κατανοήσουν τη σημασία του νοήματος, ενώ παράλληλα επιτρέπει στους φυσικούς ομιλητές της ΕΝΓ να πλουτίσουν το λεξιλόγιό τους στα Νέα Ελληνικά. Το λεξικό διαθέτει επίσης κατάταξη των λημμάτων σε σημασιολογικές κατηγορίες, η οποία έχει γίνει για να διευκολύνει την εκμάθηση ομάδων νοημάτων που συνδέονται μεταξύ τους εννοιολογικά ή πραγματολογικά. Το σημαντικότερο συστατικό του λεξικού όμως παραμένει το πολυμεσικό υλικό, εφόσον τα λήμματα αντιπροσωπεύονται από βιντεοσκοπημένα νοήματα της ΕΝΓ.

Η αναζήτηση των λημμάτων είναι δυνατή με τρεις τρόπους:

- Με βάση την σειρά χειρομορφών από τις οποίες αποτελείται κάθε νόημα: στην περίπτωση αυτή, ο χρήστης πατάει με το ποντίκι επάνω στη χειρομορφή που επιλέγει από ένα πίνακα.
- Με βάση την κατάταξη των λημμάτων σε κατηγορίες: εδώ ο χρήστης μπορεί να δει συγκεντρωμένα ονόματα ζώων ή φυτών, λέξεις σχετικές με την οικογένεια, την μαγειρική, την επικοινωνία κλπ.

- Με βάση την αλφαβητική σειρά των μεταφράσεων των νοημάτων στα νέα ελληνικά, δηλαδή με επιλογή της λέξης από ένα κατάλογο, όπως και σε άλλα λεξικά ηλεκτρονικής μορφής.

Συνολικά, το Νόημα απαρτίζεται από 3195 διαφορετικές λέξεις και παρομοίως με το αντίστοιχο Δανικό σύνολο δεδομένων κάθε λέξη που έχει διαφορετικούς τρόπους απόδοσης, εμφανίζεται με όλες τις υπάρχουσες παραλλαγές.

3.2.2 Πολύτροπον

Την περίοδο 2013-2015, σε συνέχεια του έργου «Νόημα», το ΙΕΛ ανέπτυξε τη συλλογή «Πολύτροπον» ως μία βάση δεδομένων για γλωσσολογική ανάλυση και για μεθόδους μηχανικής μάθησης. Τα βίντεο αναπαριστούν ολόκληρες προτάσεις της νοηματικής γλώσσας. Οι προτάσεις είναι σχεδιασμένες με τρόπο ώστε κάθε μία να αναδεικνύει οπτικά μία καινούργια λεκτική οντότητα. Για λέξεις που έχουν περισσότερους από έναν τρόπους νοημάτισης, υπάρχει ο αντίστοιχος αριθμός βίντεο που αναπαριστούν την ίδια πρόταση με την λέξη αυτή νοηματισμένη με όλους τους πιθανούς τρόπους. Σε όλα τα παραδείγματα εμφανίζεται ο ίδιος Έλληνας νοηματιστής. Κάθε προσθήκη είναι επισημειωμένη τόσο σε επίπεδο λήμματος σύμφωνα με τη σειρά που επιτάσσει το συντακτικό της ΕΝΓ, όσο και ως προς τη μορφή της αρχικής πρότασης. Για παράδειγμα, για κάποιο παράδειγμα δίνεται η ερμηνεία της πρότασης «Εγώ αγαπώ το ποδόσφαιρο» αλλά επίσης η επισημείωση περιγράφει ότι η σειρά εμφάνισης των νοημάτων εντός του βίντεο είναι «ΕΓΩ,ΠΟΔΟΣΦΑΙΡΟ,ΑΓΑΠΩ». Οι προτάσεις που καλείται ο νοηματιστής να αποδώσει, τις περισσότερες φορές δίνεται οπτικά με τη μορφή της αφήγησης μέσω απλών εικόνων. Με αυτόν τον τρόπο ο νοηματιστής δεν μπαίνει στην διαδικασία της μετάφρασης από την ομιλούμενη Ελληνική στην ΕΝΓ, δεν εκβιάζεται ως προς το λεξιλόγιο που θα χρησιμοποιήσει και το αποτέλεσμα του λόγου είναι πιο φυσικό και ενδεικτικό της γλώσσας. Έτσι, δημιουργείται κατά βάση ένα πολύ καλό σύνολο δεδομένων για τους σκοπούς της συστηματικής γλωσσολογικής και συντακτικής ανάλυσης της γλώσσας.

3.3 Γερμανική Νοηματική Γλώσσα

3.3.1 SigNum

Η βάση δεδομένων SIGNUM δημιουργήθηκε στο πλαίσιο ενός ερευνητικού έργου στο Ινστιτούτο Αλληλεπίδρασης Ανθρώπου-Μηχανής, το οποίο βρίσκεται στο πανεπιστήμιο RWTH Aachen της Γερμανίας. Το έργο SIGNUM Signer-Independent Continuous Sign Language Recognition for Large Vocabulary Using Subunit Models) χρηματοδοτήθηκε από το Γερμανικό Ίδρυμα Ερευνών (Deutsche Forschungsgemeinschaft) και αποσκοπούσε στην ανάπτυξη ενός αυτόματου συστήματος αναγνώρισης Νοηματικής Γλώσσας μέσω βίντεο.

Στην αρχή του έργου, σύμφωνα με τους δημιουργούς, κανένα από τα σύνολα δεδομένων Νοηματικών Γλωσσών που βρέθηκαν στη βιβλιογραφία δεν πληρούσε τις απαιτήσεις για αναγνώριση συνεχούς αναγνώρισης συνεχούς νοημάτισης από διαφορετικούς νοηματιστές. Σε αντίθεση με την αναγνώριση ομιλίας, στην πραγματικότητα δεν υπήρχε τυποποιημένο σημείο αναφοράς. Για το λόγο αυτό αποφάσισαν να δημιουργήσουν μια νέα βάση δεδομένων, η οποία θα πρέπει να είναι διαθέσιμη για άλλους ενδιαφερόμενους ερευνητές μετά το τέλος του έργου.

Το SIGNUM περιέχει τόσο μεμονωμένα όσο συνεχή νοήματα από διάφορους νοηματιστές. Το λεξιλόγιο αρκείται σε 450 βασικά νοήματα της Γερμανικής Νοηματικής Γλώσσας (ΓΝΓ) που αντιπροσωπεύουν διάφορες λέξεις της ομιλούμενης Γερμανικής. Πάνω σε αυτό το λεξιλόγιο, δημιουργήθηκαν συνολικά 780 προτάσεις. Κάθε πρόταση εκτείνεται από 2 εώς 11 λέξεις. Δεν υπάρχουν συνειδητές παύσεις μέσα στις προτάσεις αλλά οι προτάσεις αποδίδονται σε ξεχωριστά βίντεο. Ολόκληρο το σύνολο δεδομένων αποδόθηκε από μία φορά από κάθε έναν από τους 25 νοηματιστές. Συνεπώς, δημιουργείται μία βάση δεδομένων από 19500 προτάσεις, 33210 βίντεο με συνολική διάρκεια περίπου 55 ώρες. Λόγω του τεράστιο όγκου δεδομένων δεν είναι διαθέσιμο για κατέβασμα διαδικτυακά αλλά ο ενδιαφερόμενος θα πρέπει να κάνει αίτημα για αποστολή της βάσης σε υλική μορφή, δηλαδή κάποιο εξωτερικό σκληρό δίσκο.

3.3.2 RWTH-PHOENIX-Weather

Οι δημιουργοί της βάσης δεδομένων κατέγραψαν όλα τα δελτία καιρού από τις καθημερινές ειδησεογραφικές εκπομπές "Tagesschau" και "Heute-Journal" του δημόσιου τηλεοπτικού σταθμού "Phoenix" κατά την περίοδο δύο ετών (2009 - 2010). Συνολικά 190 προγνώσεις καιρού του ενάμιση λεπτού επισημειώθηκαν σε επίπεδο λέξεων. Τα δελτία καιρού σχηματίζουν ένα πολύ συμπυκνωμένοι σύνολο δεδομένων με την έννοια ότι το λεξιλόγιο που χρησιμοποιείται είναι περιορισμένο σε βασικούς όρους. Μερικές μόνο εξαιρέσεις αποτελούν γεωγραφικού προσδιορισμού όπως οι «Άλπεις», ο «Ρίνος» κτλ.

Αυτή η ιδιότητα είναι μεγάλο πλεονέκτημα στο πεδίο της μηχανικής μάθησης. Ωστόσο, τα βίντεο του RWTH-PHOENIX-Weather δεν έχουν μαγνητοσκοπηθεί σε συνθήκες εργαστηρίου όπως όλα τα ανωτέρω. Τα τηλεοπτικά στούντιο είχαν ως μέριμνα περισσότερο το οπτικό αποτέλεσμα τοποθετώντας τους νοηματιστές με σκούρα ρούχα μπροστά σε σκούρα φόντο που καθιστά την ανάλυση εικόνας δυσκολότερη. Η μεγαλύτερη δυσκολία είναι ότι υπάρχει πολύ μεγάλη ταχύτητα νοημάτισης καθώς ο νοηματιστής προσπαθεί να συμβαδίζει με την ταχύτητα ομιλίας του κεντρικού παρουσιαστή. Επιπλέον, η νοημάτιση γίνεται από ακούοντες υπό πίεση, με αποτέλεσμα η δομή των προτάσεων να βρίσκεται πιο κοντά στο συντακτικό της ομιλούμενης γλώσσας παρά σαν μία φυσική έκφραση της ΓΝΓ.

Η επισημείωση αποτελείται από

- αυτοματοποιημένη μετάφραση μέσω αναγνώρισης φωνής εφαρμοσμένης στον εκφωνητή

- προτάσεις από λέξεις με δεδομένα όρια μεταξύ προτάσεων
- μεμονωμένες λέξεις
- «φωνολογικά» χαρακτηριστικά λέξεων

Συνολικά, η βάση δεδομένων περιλαμβάνει 1980 προτάσεις της ΓΝΓ από 7 διαφορετικούς νοηματιστές με συνολική διάρκεια περίπου 3 ώρες.

3.3.3 DGS-Korpus

Η ομάδα του DGS-Corpus (Σύνολο δεδομένων Γερμανικής Νοηματικής Γλώσσας) αποτελείται από κωφά και ακούοντα μέλη του «Ινστιτούτου για την Γερμανική Νοηματική Γλώσσα και Επικοινωνίας Κωφών» του Πανεπιστημίου του Αμβούργου. Πρόκειται για την ίδια ομάδα που υποστηρίζει την πιο διάσημη τεχνολογία επισημείωσης, το HamNoSys, που αναλύεται στο Κεφάλαιο 2.3.

Ένας από τους σκοπούς του έργου είναι να αναπτύξει ένα σύνολο που αντιπροσωπεύει το καθημερινό «λόγο» ενός χρήστη της ΓΝΓ. Για αυτό το σκοπό έγινε μία συλλογή δεδομένων από 330 κωφούς από 12 περιοχές της Γερμανίας. Μέσω αυτής της διαδικασίας συλλέχθηκαν χαρακτηριστικά ακόμα και από τις τοπικιστικές διαφορές της ΓΝΓ. Η συλλογή δεδομένων αποτελείται από διάφορα μέρη όπως συνομιλίες δύο ατόμων για διάφορες θεματικές ή ακόμα διάφορες διηγήσεις δραστηριοτήτων που προέρχονται από εικόνες ή μικρά βίντεο που προβλήθηκαν στον νοηματιστή, παρόμοια με τη διαδικασία που αναφέρθηκε στο «Πολύτροπον». Συνολικά, υπάρχει υλικό περίπου 500 ωρών χρήσης της ΓΝΓ που αντιστοιχεί στην εμφάνιση περίπου 3 εκατομμυρίων νοημάτων αγγίζοντας την πληροφορία που συναντά κανείς σε αντίστοιχες συλλογές δεδομένων ομιλίας.

3.4 Κορεάτικη Νοηματική Γλώσσα

3.4.1 KETI

Το σύνολο δεδομένων KETI αποτελεί μία συλλογή βίντεο για την καταγραφή της Κορεάτικης Νοηματικής Γλώσσας μέσα από παραδείγματα που αφορούν σε διάφορες καταστάσεις έκτακτης ανάγκης [13]. Σε πολλές κοινωνικές καταστάσεις, οι άνθρωποι με προβλήματα ακοής χρειάζονται αναγκαστικά βοήθεια από επαγγελματίες διερμηνείς νοηματικής γλώσσας για να επικοινωνούν με τους ακροατές ακόμα και όταν πρέπει να αποκαλύπτουν τις πολύ ιδιωτικές και ευαίσθητες πληροφορίες τους. Επιπλέον, οι άνθρωποι με προβλήματα ακοής είναι πιο ευάλωτοι σε διάφορες καταστάσεις έκτακτης ανάγκης λόγω των εμποδίων επικοινωνίας λόγω της απουσίας της ακοής. Για αυτό το λόγο οι δημιουργοί του KETI δίνουν παραδείγματα από περιπτώσεις σχετικά γενικών συζητήσεων για καταστάσεις έκτακτης ανάγκης. Επιλέχθηκαν 105 προτάσεις και 419 λέξεις που θα μπορούσαν να χρησιμοποιηθούν σε διάφορες καταστάσεις έκτακτης ανάγκης. Το σύνολο δεδομένων KETI αποτελείται από 14.672 βίντεο υψηλής ευκρίνειας

(HD) που καταγράφηκαν σε 30 καρέ ανά δευτερόλεπτο και από δύο γωνίες κάμερας: μπροστά και πλάγια. Καταγράφηκαν 524 διαφορετικά νοήματα που προέκυψαν από την προαναφερθείσα διαδικασία και εκτελούνται από δεκατέσσερις διαφορετικούς νοηματιστές με προβλήματα ακοής για να αντικατοπτρίζουν τις ατομικές διαφορές για το ίδιο νόημα. Για κάθε νόημα, καταγράφουμε πρώτα ένα ¨οδηγό βίντεο' ενός νοηματιστή ¨εμπειρογνώμονα' για να καταργηθούν πιθανές ασάφειες των νοημάτων. Μετά την παρακολούθηση του οδηγού βίντεο, οι δεκατέσσερις υπογράφοντες με προβλήματα ακοής κατέγραψαν καθένα από τα 524 νοήματα. Ως αποτέλεσμα, κάθε νοηματιστής κατέγραψε συνολικά 1048 βίντεο για το σύνολο δεδομένων. Η συνολική διάρκεια όλων των βίντεο αθροίζονται σε περίπου 24 ώρες.

Κεφάλαιο 4

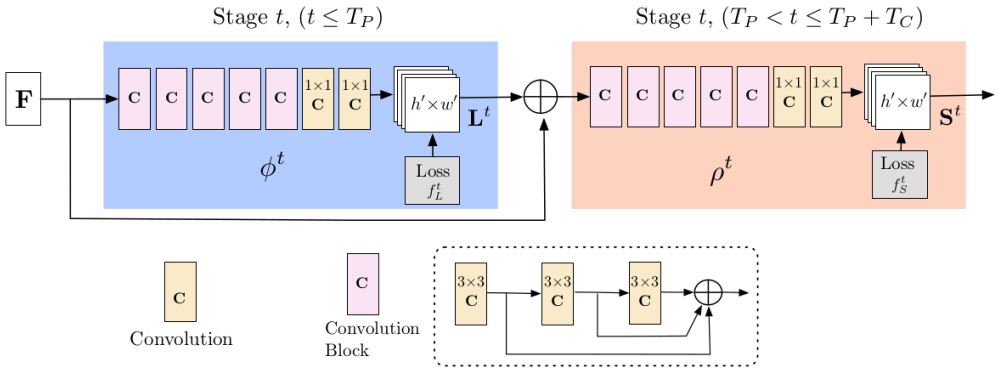
openpose

Το Openpose [4][5][26][31] είναι ένα έργο που υλοποιήθηκε από το Berkeley Artificial Intelligence Research lab (BAIR) σε στενή συνεργασία με το Perceptual Computing lab του Carnegie Mellon University (το οποίο υποστηρίζει το λογαριασμό του έργου στο Github). Το Openpose έχει τη δυνατότητα να υλοποιεί «εκτίμηση διάταξης του ανθρωπίνου σώματος σε πραγματικό χρόνο και για πολλαπλό αριθμό ατόμων ταυτόχρονα». Στα πλαίσια της εφαρμογής έχει αναπτυχθεί τόσο ένα εκτελέσιμο πρόγραμμα για την εξαγωγή χαρακτηριστικών από βίντεο ή εικόνα που προέρχονται είτε από αρχείο είτε κατευθείαν από συσκευή κάμερας, αλλά επίσης, διατίθεται και σχετική διεπαφή προγραμματισμού (API) για ενσωμάτωση σε C και Python . Στην περίπτωση της εισόδου από κάμερα, το αξιοσημείωτο αυτής της εφαρμογής είναι το πρώτο σύστημα ανοικτού κώδικα που έχει τη δυνατότητα ανίχνευσης των χαρακτηριστικών σε πραγματικό χρόνο ($25 \frac{frames}{second}$), ενώ επεκτείνεται στην ανίχνευση λεπτομερειών επί του ανθρωπίνου σώματος πέραν του βασικού σκελετού, όπως είναι τα πέλματα, τα χέρια και το πρόσωπο.

Η ανίχνευση της διάταξης του σώματος (pose) γίνεται με τον εντοπισμό χαρακτηριστικών σημείων (keypoints) όπως οι ώμοι, οι αγκώνες κτλ. Η μέθοδος που χρησιμοποιείται πίσω από το σύστημα βασίζεται στα **πεδία μερικής συνάφειας** (Part Affinity Fields). Στο Σχήμα 4.1 φαίνεται η αρχιτεκτονική του μοντέλου πίσω από την εφαρμογή. Αρχικά, ένα βαθύ νευρωνικό δίκτυο που αποτελείται από επιμέρους επίπεδα Συνελικτικών Νευρωνικών Δικτύων δέχεται την εικόνα/καρέ ως είσοδο (Σχήμα 4.2α′). Το πρώτο στάδιο έχει ως αποτέλεσμα την εξαγωγή ενός χάρτη πιθανοφάνειας, με μέγεθος ίδιο με εκείνο της εισόδου για κάθε χαρακτηριστικό σημείο (Σχήμα 4.2β′). Με άλλα λόγια το σύστημα έχει εκπαιδευτεί να αναγνωρίζει ξεχωριστά κάθε σημείο του σώματος (π.χ. δεξί αγκώνα, αριστερό γόνατο κτλ). Έπειτα με τη χρήση Part Affinity Fields που αποτελούν μία μέθοδο υπολογισμού διανυσμάτων ροής, υπολογίζονται τα διανύσματα για κάθε πιθανή ακμή «Σχήμα 4.2γ′». Κάθε ακμή φυσικά αναπαρίσταται με ένα ευθύγραμμο τμήμα που ενώνει 2 σημεία κλειδιά που αναπαριστούν κάποιες συγκεκριμένες αρθρώσεις και αντιστοιχεί σε κάποιο μέρος του σώματος στο οποίο ανήκουν αυτές οι αρθρώσεις. Για παράδειγμα μία ακμή μπορεί να είναι το δεξί μπράτσο ως το ευθύγραμμο τμήμα που ενώνει τον δεξί ώμο με το δεξί

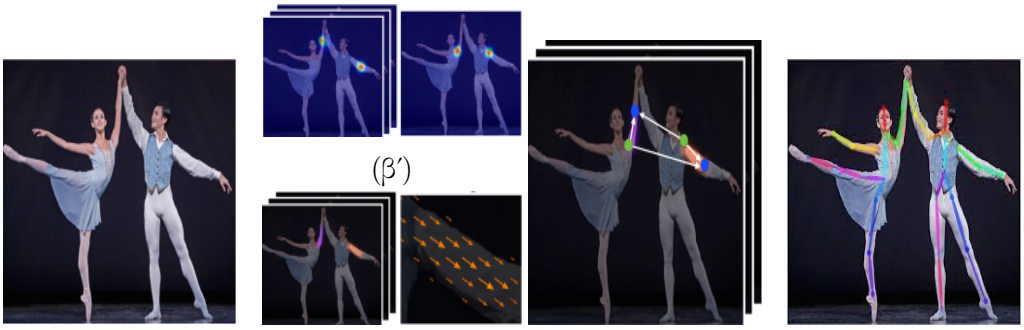
αγκώνα. Τέλος, με τη χρήση ενός αναλυτή (parser) καταλήγει σε όλες τις πιθανές ακμές (Σχήμα 4.2δ′) και το σύστημα στο τελικό στάδιο, εμφανίζει το αποτέλεσμα που είναι πιθανότερο δίνοντας τα χαρακτηριστικά σημεία εκφρασμένα ξεχωριστά για κάθε άτομο (Σχήμα 4.2ε′). Αξίζει να σημειωθεί ότι η απόδοση αυτής της μεθόδου καταφέρνει να είναι σχεδόν ανεπηρέαστη από τον αριθμό των ατόμων που εμφανίζονται καθώς κάθε κατηγορία σημείων εντοπίζεται με μία σάρωση του δικτύου.

Πέραν αυτού του πλεονεκτήματος, το openpose μοιάζει να μένει ανεπιρέαστο από το τι υπάρχει στο προσκήνιο. Κανένα από τα δύο χαρακτηριστικά δεν αξιοποιείται στα πλαίσια της παρούσας εργασίας καθώς όλα τα βίντεο προς επεξεργασία περιέχουν έναν μόνο νοηματιστή κινηματογραφημένο μπροστά από μία μπλέ οθόνη. Ωστόσο, σε πολλές εργασίες [3][27][25] όπου γίνεται χρήση δεδομένων με σύνθετο προσκήνιο δίνεται ιδιαίτερη έμφαση στη σημαντικότητα, την ευαισθησία και την δυσκολία στην προσπάθεια εντοπισμού του νοηματιστή από το προσκήνιο. Μελλοντικά η αναγνώριση διαφορετικών ανθρώπων μπορεί να φανεί ιδιαίτερα χρήσιμη σε εφαρμογές όπου πέραν από τον νοηματιστή υπάρχουν και άλλοι άνθρωποι στο πλάνο του βίντεο.



Σχήμα 4.1: Η αρχιτεκτονική του μοντέλου πίσω από το Openpose

Επιπλέον, λόγω του ότι κάθε σημείο εντοπίζεται χωρικά ανεξάρτητα από τα υπόλοιπα, δίνεται η δυνατότητα στη μέθοδο να μπορεί να λειτυργήσει ακόμα και όταν μόνο ένα μέρος του σώματος είναι ορατό στην εικόνα. Αυτή η ιδιότητα είναι μείζονος σημασίας για την παρούσα εφαρμογή μιας και σχεδόν πάντα ο νοηματιστής κινηματογραφείται από τη μέση και επάνω. Έτσι, το openpose αποτελεί μία επανάσταση ως προς το πεδίο εφαρμογής του καθώς έλυσε τα προβλήματα που ταλάνιζαν τις εφαρμογές που βασίζονταν στον αισθητήρα Kinect που ήταν η πιο διάσημη λύση μέχρι πρότινος. Πέραν από την ικανότητα να λειτουργεί «βλέποντας» μόνο ένα μέρος του σώματος, το openpose κατάφερε να ανεξαρτητοποιήσει την υλοποίηση της εύρεσης χαρακτηριστικών σημείων του σώματος από την κατοχή συγκεκριμένου υλικού εξοπλισμού όπως στην περίπτωση του Kinect. Πρόκειται για μία υλοποίση σε επίπεδο λογισμικού.Για τον ίδιο λόγο φαίνεται να απειλεί και την χρήση της τεχνολογίας MoCap (Motion Capture) καθώς η ακρίβεια που προσφέρει είναι, για τις περισσότερες εφαρμογές, ικανοποιητική ώστε να επιλέξει κανείς να αποφύγει τον πολύ ακριβό ε-



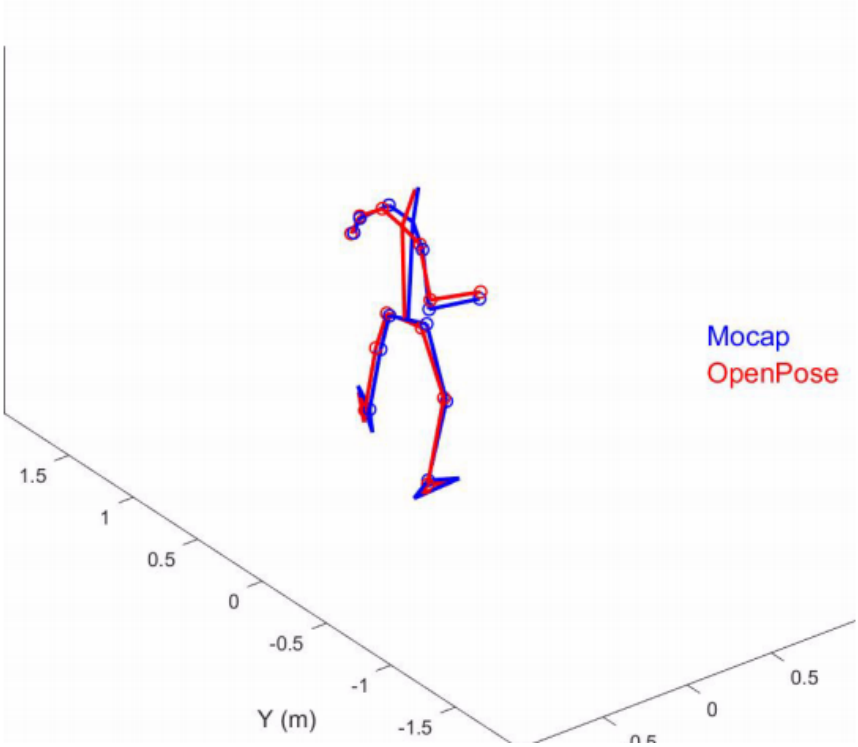
Σχήμα 4.2: Παράδειγμα σταδίων του Openpose (α′) Είσοδος (β′) Χάρτες Πιθανοφάνειας (γ′) Πεδία Μερικής Συνάφειας (δ′) Αποτελέσματα αναλυτή (ε′) Τελικό Αποτέλεσμα

ξοπλισμό της στολής που απαιτεί το Mocap. Στο Σχήμα 4.3 φαίνεται μέσα από ένα παράδειγμα η σύγκριση μεταξύ openpose και mocap [22].

Από την άλλη πλευρά, η μη ύπαρξη σχετικού εξοπλισμού κάνει την δυνατότητα της «ανάλυσης σε πραγματικό χρόνο» εντελώς σχετική με το υλικό που διαθέτει ο ερευνητής. Για παράδειγμα, η εκτέλεση του openpose πάνω σε μία κάρτα RTX 2060 έχει ως αποτέλεσμα την ανίχνευση του βασικού σκελετού με μικρότερη συχνότητα από ότι ο αισθητήρας Kinect, ενώ (τη στιγμή της συγγραφής της εργασίας) κοστίζει περίπου μιάμιση φορά όσο αυτός. Βέβαια, για την περίπτωση, όπως την παρούσα, όπου απαιτείται η εξαγωγή χαρακτηριστικών για τα δάκτυλα, η απόδοση πέφτει στα $5\frac{frames}{sec}$ αλλά το Kinect (αυτή τη στιγμή) δεν προσφέρει αντίστοιχη λειτουργία ώστε να τίθεται θέμα επιλογής ανάμεσα στα δύο. Επιπλέον, η αντίστοιχη υλοποίηση σε MoCap αυξάνει το κόστος ακόμα περισσότερο και πάσχει από το πρόβλημα της απόκρυψης ορισμένων σημείων από την κάμερα λόγω επικάλυψης. Το openpose δίνει λύση και σε αυτό το πρόβλημα εντάσσοντας συμφραζόμενα στο μοντέλο (εξάρτηση στο χρόνο) στην περίπτωση της εισόδου βίντεο. Με άλλα λόγια στις περιπτώσεις που κάποιο σημείο κρύβεται από κάποιο μέλος του σώματος, γίνεται εκτίμηση από προηγούμενα καρέ.

Απο πλευράς τεχνικών χαρακτηριστικών, αν το πρόγραμμα κληθεί με τα σχετικά ορίσματα για πλήρη ανάλυση του σώματος, τότε για κάθε απλή (δισδιάστατη) φωτογραφία/καρέ επιστρέφονται:

- Για το **σώμα** συνολικά παράγονται 18 σημεία (Σχήμα 4.4α′). Το σημείο αναφοράς (υπαριθμόν 0) πρόκειται για το κέντρο του προσώπου και έπειτα η αρίθμηση συνεχίζει με τη βάση του λαιμού. Τα χέρια και πόδια περιγράφονται το καθένα από 3 σημεία, ένα για κάθε άρθρωση. Ο κορμός του ατόμου, σε αντίθεση με άλλες υλοποιήσεις ορίζεται αφαιρετικά ως η περιοχή κάτω από το λαιμό έως τις αρθρώσεις των ισχίων. Επιπλέον, τα τελευταία 4 σημεία προσδιορίζουν τη θέση των ματιών και των αυτιών.

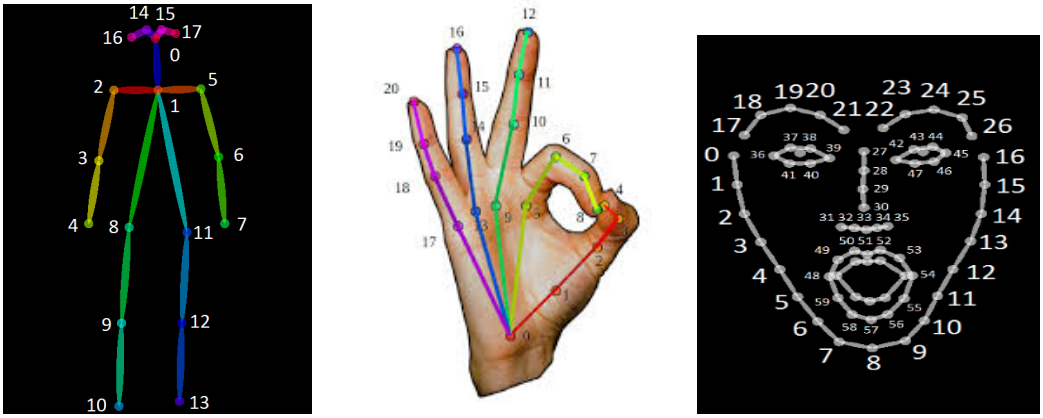


Σχήμα 4.3: Παράδειγμα σύγκρισης απόδοσης Mocap και Openpose Openpose

- Για κάθε **χέρι** παράγονται συνολικά 21 σημεία. Κάθε δάκτυλο αναπαρίσταται από 3 σημεία για τις αρθρώσεις και 1 σημείο για την απόληξη αυτού. Το επιπλέον σημείο (υπ΄ αριθμόν 0) αντιστοιχεί στον καρπό και θεωρητικά θα πρέπει να ταυτίζεται με τα σημεία υπ΄ αριθμόν 4 και 7, για το δεξί και αριστερό χέρι, αντίστοιχα.
- Για το **πρόσωπο** χρησιμοποιούνται 70 σημεία καθώς σε αυτή την περίπτωση γίνεται μία πολύ αναλυτική απεικόνιση ως προς τις βασικές γραμμές το περιγράφουν περίγραμμα, φρύδια, μάτια, μύτη, στόμα.

Κάθε ένα από αυτά τα σημεία επιστρέφεται σε μορφή αρχείου κειμένου (.xml ή .json) ώστε να είναι προσιτά για περαιτέρω επεξεργασία αλλά επίσης υπάρχει η δυνατότητα άμεσης εξαγωγής ως εικόνα ή βίντεο (render), των χαρακτηριστικών σημείων με τα αντίστοιχα ευθύγραμμα τμήματα που τα συνδέουν. Όπως αναφέρθηκε παραπάνω, τα χαρακτηριστικά σημεία δίνονται ξεχωριστά για κάθε άτομο ως 4 διανύσματα (σώμα, δεξί χέρι, αριστερό χέρι, πρόσωπο). Κάθε διάνυσμα δίνεται ως αλληλουχία συντεταγμένων των αντίστοιχων σημείων, με τη σειρά αρίθμησης όπως φαίνεται στο Σχήμα 4.4. Τέλος, η ανάλυση δεν αφορά τα χέρια ή/και το πρόσωπο το αντίστοιχο διάνυσμα απεικονίζεται ως κενό (μηδενικής διάστασης).

Κάθε σημείο έχει 3 διαστάσεις. Οι πρώτες δύο εκφράζουν την τετμημένη και την τεταγμένη, αντίστοιχα, του σημείου και η τρίτη είναι μία ποσοτικοποιήση για την



(α′) (β′) (γ′)

Σχήμα 4.4: Αναπαράσταση των χαρακτηριστικών σημείων όπως παράγονται από το Openpose (α′) Σώμα/Πόδια (β′) Χέρι (γ′) Πρόσωπο

«σιγουριά» του συστήματος πως έχει εντοπίσει σωστά το σημείο αυτό. Η τελευταία τιμή προκύπτει από μία συνάρτηση softmax με αποτέλεσμα να κυμαίνεται στο διάστημα $[0, 1]$. Στην περίπτωση που μέρος του σώματος απουσιάζει από την εικόνα/καρέ τα αντίστοιχα σημεία του εκφράζονται με μηδενικά σε όλες τις θέσεις.

Επίσης, ολόκληρη η υλοποίηση επεκτείνεται και σε τρισδιάστατες εικόνες/βίντεο στις περιπτώσεις χρήσης κάμερας με αισθητήρα βάθους.

Κεφάλαιο 5

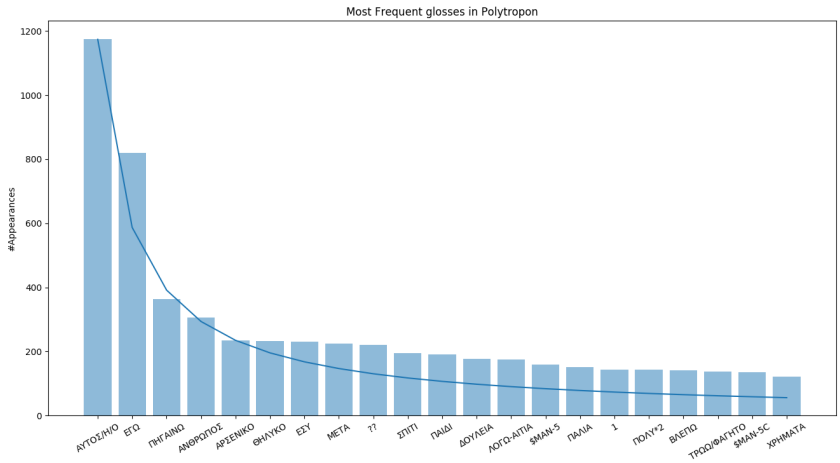
Πειράματα

Όπως αναφέρθηκε στο Κεφάλαιο 3 οι περισσότερες βάσεις δεδομένων σχετικά με τις Νοηματικές Γλώσσες έχουν δημιουργηθεί με σκοπό την καταγραφή αυτών. Έτσι, τα δεδομένα της Δανικής Νοηματικής και το Νόημα+, παρόμοια με ένα λεξικό κάποιας προφορικής γλώσσας, αναπαριστούν κάθε λέξη μία φορά πλην μερικών εξαιρέσεων. Η δομή αυτή είναι ιδανική τόσο για την απαθανάτιση μίας γλώσσας όσο και για την εκμάθηση αυτής από έναν άνθρωπο. Σαν συνέχεια αυτής της λογικής, το Πολύτροπον παρουσιάζει για κάθε λεξιλογική οντότητα μία μικρή πρόταση που την εμπεριέχει δημιουργώντας μία δομή ιδανική για τη διδασκαλία του λεξιλόγιο της Ελληνικής Νοηματικής Γλώσσας. Ωστόσο, αυτή η δομή δεν διευκολύνει την μηχανική μάθηση καθώς αν θεωρήσουμε κάθε λέξη ως μία κλάση. Σε αυτήν την περίπτωση θα έχουμε μία μοναδικότητα (singleton) σε κάθε πρόταση.

Στο Σχήμα 5.1 φαίνεται ότι λεξιλογικά, η συχνότητα εμφάνισης των λέξεων ακολουθεί την κατανομή $[\frac{1}{1^s}, \frac{1}{2^s}, \frac{1}{3^s}, ...]$, με κάποιο μικρό $s > 0$, η οποία είναι χαρακτηριστική γενικά για οποιαδήποτε φυσική γλώσσα σύμφωνα με το θεμελιώδη κανόνα της υπολογιστικής γλωσσολογίας, το νόμο του Zipf [33][1]. Ωστόσο, παρατηρώντας το Σχήμα 5.3 φαίνεται ότι επιλέγοντας κάποιο πλήθος n από τις συχνότερες λέξεις του Πολύτροπον, τότε τα παραδείγματα προτάσεων εντός αυτού όπου δεν υπάρχουν «άγνωστες» λέξεις θα είναι μεταξύ n και $2n$ επιβεβαιώνοντας ότι το Πολύτροπον ανήκει και αυτό στην κατηγορία συνόλων που ταιριάζουν καλύτερα στον ανθρώπινο τρόπο μάθησης.

Ο Dr. Koller σε εργασίες του [16][19][20][18] που εστίασαν πάνω στο ζήτημα της μετάφρασης μέσα από βίντεο Νοηματικής γλώσσας, πέρα από το το Δανικό Dataset χρησιμοποιήθηκαν το **SigNum** και το **RWTH-PHOENIX-Weather Multisigner 2014**. Αυτά τα δύο σύνολα δεδομένων πέρα από το ότι αποτελούν παραδείγματα συνεχούς λόγου, έχουν μία αντιστοιχία προτάσεων και λεξιλογίου περίπου επτά προς ένα. Γενικότερα, σε κάθε περίπτωση είναι ωφέλιμο να επιδιώκουμε την μεγαλύτερη κατά το δυνατόν σχέση μεταξύ παραδειγμάτων και κλάσεων.

Από τη στιγμή που τα δικά μας διαθέσιμα σύνολα δεν έχουν τον ιδανικό λόγο λέξεων ως προς τις προτάσεις θα ήταν χρήσιμο να υπάρξει κάποιο τέχνασμα ώστε να εξάγουμε κλάσεις θεμελιοδέστερες της λέξης. Σε αυτό εξυπηρετούν τα «φωνο-

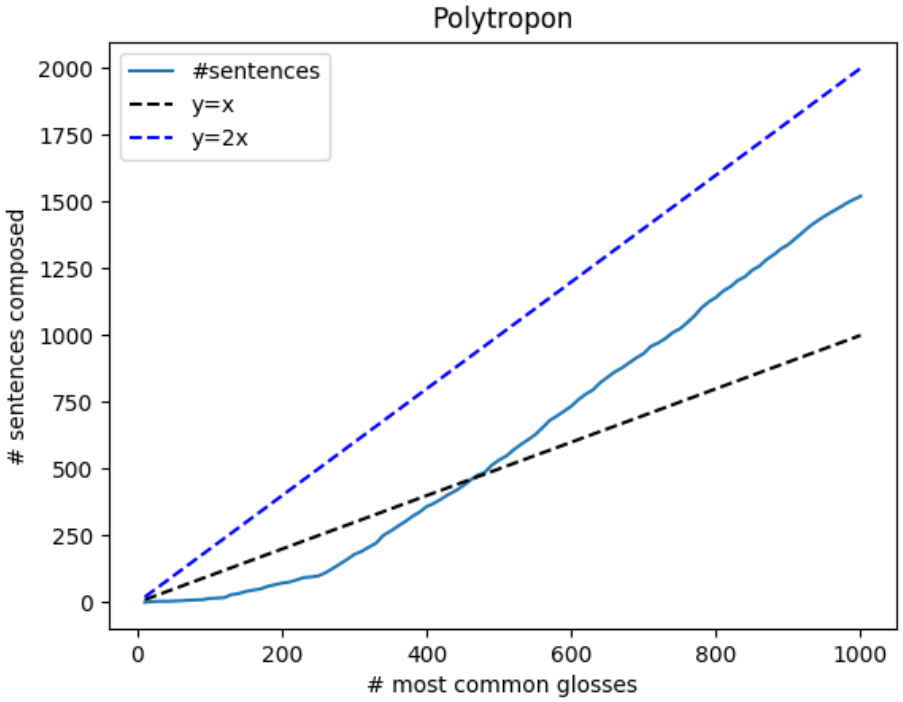


Σχήμα 5.1: Οι 20 συχνότερες λέξεις που εμφανίζονται στο Πολύτροπον

λογικά» χαρακτηριστικά που παρουσιάστηκαν στο Κεφάλαιο 2. Για αυτό το λόγο εκμεταλλευόμαστε το γεγονός ότι τόσο στο Δανικό Dataset όσο και στο Νόημα+, πέρα από τη μετάφρασή τους, για όλες τις λέξεις δίνονται οι αντίστοιχες περιγραφές σε HamNoSys. Με αυτόν τον τρόπο, έχουμε τη δυνατότητα να εκπαιδεύσουμε ένα απλούστερο σύστημα για κάθε φωνολογικό χαρακτηριστικό (χειρομορφή, προσανατολισμός, κίνηση, θέση). Μάλιστα, εφόσον τρία από τα τέσσερα αποτελούν στατικά χαρακτηριστικά, μπορούμε να θεωρήσουμε ως ένα στοιχείο του συνόλου εκπαίδευσης ένα καρέ ενός βίντεο αντί ενός ολόκληρου βίντεο όπως θα πρέπει να γίνεται στην περίπτωση της αναγνώρισης μεμονωμένων λέξεων.

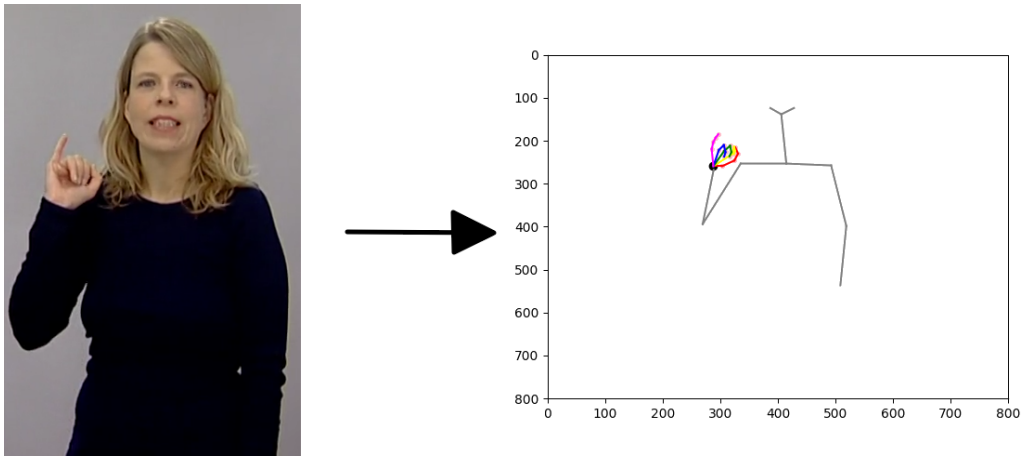
5.1 Εξαγωγή Χαρακτηριστικών

Η επεξεργασίας εικόνας στην μηχανική μάθηση είναι πλέον συνηφασμένη με τη χρήση συνελικτικών νευρωνικών δικτύων. Στα πλαίσια της παρούσας εργασίας η χρήση συνελικτικών δικτύων γίνεται έμμεσα μέσω του Openpose. Όπως αναφέρθηκε στο Κεφάλαιο 4, το Openpose αποτελεί στην ουσία ένα πάρα πολύ μεγάλο συνελικτικό δίκτυο, εκπαιδευμένο πάνω σε μεγάλο αριθμό εικόνων από ανθρώπους από πολλές οπτικές γωνίες. Σύμφωνα με το [14] το Openpose μπορεί να χρησιμοποιηθεί ως ένα εργαλείο προεπεξεργασίας των δεδομένων που οδηγεί σε πολύ καλά αποτελέσματα στην αυτόματη μετάφραση της Νοηματικής Γλώσσας. Σε κάθε περίπτωση, από πλευράς διαστατικότητας υπάρχει μία τεράστια συμπίεση δεδομένων χωρίς να υπάρχει επί της ουσίας αλλοίωση ως προς την πληροφορία για τη διάταξη του ανθρωπίνου σώματος. Αρκεί κανείς να αναλογιστεί ότι πρόκειται για ένα μετασχηματισμό εικόνας σε διάνυσμα. Σε μία ενδεικτική περίπτωση παλιάς τεχνολογίας, το βίντεο έχει ανάλυση $720 \times 540 (=388800)$pixels το οποίο στα πλαίσια της επεξεργασίας μεταφράζεται στον



Σχήμα 5.2: Αναπαράσταση της σχέσης επιλογής κάποιου πλήθος συχνότερων λέξεων και του αριθμού των προτάσεων από το Πολύτροπον που αυτές μπορούν να συνθέσουν

αντίστοιχο αριθμό διαστάσεων. Με τη χρήση του Openpose και την παραδοχή ότι απεικονίζεται ένα άτομο, η αντίστοιχη διαστατικότητα μεταφράζεται σε ένα διάνυσμα διαστάσεων $130 \times 3 (=390)$, στην ακραία περίπτωση όπου εξαντλούνται οι δυνατότητες του προγράμματος. Στην περίπτωσή που εξετάζουμε και χρησιμοποιούμε το Δανικό Dataset δεν χρησιμοποιούμε τα 70 σημεία του προσώπου, το δευτερεύον χέρι και τα σημεία κάτω από τη μέση. Το τελικό μας διάνυσμα για κάθε εικόνα/καρέ έχει διάσταση $33 \times 3 (=99)$. Κατά συνέπεια, μέσω αυτού του εργαλείου το πρόβλημα ξεκινάει με μία συμπίεση των δεδομένων εισόδου κατά περίπου 4000 φορές! Καθώς η ανάλυση του βίντεο αυξάνεται, προφανώς, η διαστατικότητα του διανύσματος μένει σταθερή προσφέροντας, παράλληλα, μεγαλύτερη αξιοπιστία στην εκτίμηση κάθε σημείου. Με τον τρόπο αυτό, το μοντέλο που στηρίζεται στην μετατροπή της εισόδου σε σημεία κλειδιά, γίνεται ανεξάρτητο του μεγέθους της αρχικής εισόδου.



Σχήμα 5.3: Μετατροπή εικόνας μέσω του Openpose

5.2 Κατάτμηση Δεδομένων

Για κάθε βίντεο δίνεται η μετάφρασή του στο HamNoSys. Όπως αναφέρθηκε στο Κεφάλαιο 3, η συλλογή από βίντεο της Δανικής Νοηματικής Γλώσσας περιέχει μεμονωμένες μαγνητοσκοπήσεις λέξεων. Από τη στιγμή που γίνεται αναγνώριση σε επίπεδο στιγμιότυπου θα πρέπει να γίνει μεταφορά της επισημείωσης για κάθε καρέ του βίντεο. Σε αυτή τη διαδικασία εμφανίζονται δύο εμπόδια. Αρχικά, ο τρόπος της επισημείωσης μας δίνει πληροφορία για την σειρά με την οποία εμφανίζεται κάθε φωνολογικό χαρακτηριστικό αλλά όχι για το χρόνο όπου απεικονίζεται κάθε ένα από αυτά, δηλαδή την αντιστοιχία στιγμιοτύπων και επισημείωσης. Με άλλα λόγια το πρόβλημα ανήκει στην περίπτωση της ημιεπιβλεπούμενης μάθησης (semisupervised learning). Για παράδειγμα για το βίντεο που απεικονζεται στο Σχήμα 5.4 μας δίνεται από την επισημείωση ότι εμφανίζονται με τη σειρά οι χειρομορφές «5» και «*s*» ( , ), αντίστοιχα. Ωστόσο, δεν γνωρίζουμε εκ των προτέρων σε ποιες εικόνες εφανίζεται η κάθε μία χειρομορφή. Επιπλέον, αξίζει να παρατηρηθεί ότι υπάρχουν εικόνες που δεν απεικονίζουν καμία από τις δύο χειρομορφές αλλά αποτελούν μεταβατικά στιγμιότυπα πριν και μετά την εκφορά του νοήματος (out of sign), είτε εσωτερικά κατά τη μετάβαση από μία θέση ή χειρομορφή στην επόμενη στα πλαίσια του ίδιου νοήματος (intra sign).

Αυτά τα προβλήματα αντιμετοπίζονται από μερικές εργασίες [17][32][2] με τη χρήση Κρυφών Μαρκοβιανών Μοντέλων για την ευθυγράμμιση του βίντεο με την επισημείωση και η υπόθεση μίας επιπλέον «άχρηστης» κλάσης για κάθε φωνολογικό χαρακτηριστικό ώστε μετά την εκπαίδευση να αντιστοιχίζονται τα στιγμιότυπα που απεικονίζουν μία μεταβατική κατάσταη. Επιπλέον, όπως και στο [30][23] γίνεται χρήση μη επιβλεπούμενης μάθησης για την ομαδοποίηση των κοινών χειρομορφών με τους αντίστοιχους όρους που τις περιγράφουν στο HamNoSys.

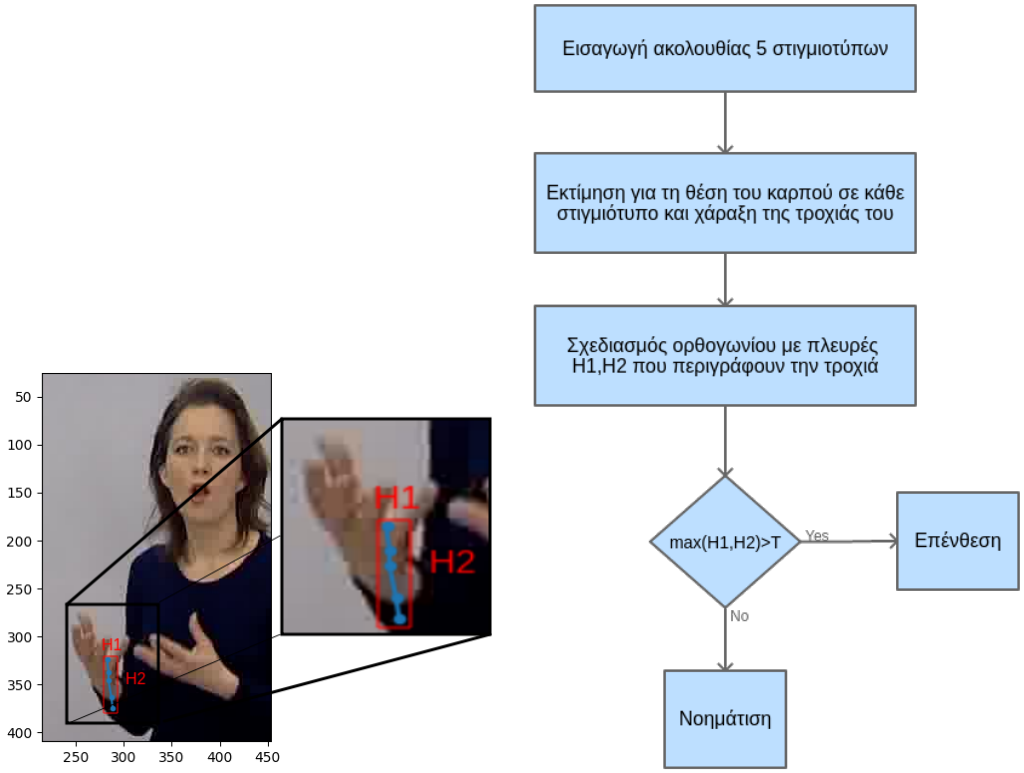
Αναφορικά με το ζήτημα διαχείρισης των μεταβατικών στιγμιοτύπων, το [6] χω-

Σχήμα 5.4: Παράδειγμα ακολουθίας Δανικής Νοηματικης για το νόημα «Μαζί μας»

ρίζει τις κινήσεις σε «επένθεσης» και «νοημάτισης». Ως επένθεση ορίζεται η κίνηση ανάμεσα σε δύο διαδοχικά νοήματα και ως «νοημάτισης» οι κινήσεις που αποτελούν μέρος ενώς νοήματος. Ο διαχωρισμός γίνεται με κριτήριο την ταχύτητα της κίνησης του χεριού και βασίζεται στην παραδοχή ότι κατά την νοημάτιση τα χέρια κινούνται πιο σιγά σε σχέση με τη διάρκεια της επένθεσης. Ως ταχύτητα, χρησιμοποιείται η μετατόπιση του «κέντρου μάζας» του χεριού σε διαδοχικά καρέ. Συνεπώς, τείθεται ένα κατώφλι ταχύτητας ως κριτήριο διαχωρισμού ανάμεσα στις 2 περιπτώσεις κίνησης. Στη συνέχεια, γίνεται καταγραφή της κίνησης του κυρίαρχου χεριού για πέντε διαδοχικά καρέ και σχεδιάζεται το παραλληλόγραμμο που περιγράφει την καταγεγραμένη τροχιά (Σχήμα 5.5α΄). Τέλος, η ακολουθία κατηγοριοποιείται ως κίνηση νοημάτισης ή επένθεσης ανάλογα με το αν η μεγαλύτερη πλευρά του περιγεγραμμένου ορθογωνίου είναι μικρότερη ή μεγαλύτερη από το κατώφλι, αντίστοιχα (Σχήμα 5.17). Ωστόσο, η υπολογιστική αυτή μέθοδος δεν επιρεάζεται από τις στατικές αλλαγές του χεριού που πραγματοποιούνται κατά τη μετάβαση από μία χειρομορφή σε κάποια άλλη, το οποίο στην περίπτωση που εξετάζει η συγκεκριμένη εργασία είναι απολύτως θεμιτό καθώς εστιάζει περισσότερο σε περιπτώσεις όπου το χέρι απλά χαράζει σχηματα κινούμενο στο χώρο, με κάθε σχήμα να είναι εντοπισμένο σε άλλη θέση. Σε άλλες περιπτώσεις όπως στο [15] γίνεται χρήση της οπτικής ροής επί του χεριού ώστε να ποσοτικοποιήται οποιαδήποτε αλλαγή μεταξύ δύο εικόνων.

Εστιάζοντας στο χέρι του νοηματιστή οι αλλαγές που εμφανίζονται μπορεί να είναι είτε στατικές είτε δυναμικές. Με τον όρο δυναμικές περιγράφουμε τις κινήσεις που οφείλονται στην μεταφορά του χεριού εντός τους χώρου νοημάτισης. Με τον όρο

στατική αλλαγή αναφέρεται η περίπτωση όπου η παλάμη μένει στο ίδιο σημείο και αλλαγές οφείλονται στην αλλαγή της μορφής του χεριού.



(α′) Τροχιά χεριού και περιγεγραμμένο ορθογώνιο

(β′) Διάγραμμα Ροής Αλγορίθμου

Σχήμα 5.5: Περιγραφή αλγορίθμου για αναγνώριση του είδους κίνησης

Εφαρμόζοντας αυτή τη μέθοδο, καταφέρνουμε με μεγάλη επιτυχία να διαγράψουμε τα στιγμιότυπα που αντιστοιχούν σε στιγμές εκτός της διάρκειας κάποιου νοήματος, οι οποίες από τη στιγμή που όλα τα παραδείγματα προς επεξεργασία είναι μονολεκτικά, αφορούν τη μετάβαση προς και από το χώρο νοημάτισης στην αρχή και το τέλος του βίντεο, αντίστοιχα. Η βασική διαφοροποίησή μας από το [6] είναι η χρήση του Openpose που μας αποδεσμεύει από τη διαδικασία αναγνώρισης και παρακολούθησης του χεριού [24] η οποία θα προσέθετε αβεβαιότητα στο τελικό αποτέλεσμα της κατάτμησης. Επίσης, αντί της ανίχνευσης ολόκληρου του πεδίου του χεριού και έπειτα η χρήση του κέντρου μάζας, που συνηθίζεται στις κλασσικές μεθόδους όρασης υπολογιστών, εφαρμόζουμε την μέθοδο των τροχιών ξεχωριστά σε κάθε ένα από τα 21 σημεία του χεριού. Το κατώφλι T αφορά πλέον το άθροισμα των μέγιστων πλευρών των περιγεγραμμένων ορθογωνίων αυτών. Ο κανόνας αυτός μοντελοποιείται από την εξίσωση (5.1), προσφέρει τη μέγιστη δυνατή ακρίβεια για την αποφυγή τόσο δυναμικών όσο και στατικών μεταβολών στο χέρι και ισχύει για τις περιπτώσεις αναγνώρισης

χειρομορφής και προσανατολισμού του χεριού.

Μάλιστα, στις περιπτώσεις αυτές, προκειμένου να αυξηθεί η αξιοπιστία της τελικής επισημείωσης δεδομένων, θέτουμε ακόμα ένα κατώφλι T_σ που αφορά την συνολική βεβαιότητα ως προς το σωστό εντοπισμό των σημείων του χεριού. Αυτό επιτυγχάνεται θεωρώντας την πληροφορία που δίνεται από το Openpose μέσω της τρίτης διάστασης ως μέτρο πιθανότητας (βλ. Κεφάλαιο 4). Έπειτα, υπολογίζουμε για κάθε στιγμιότυπο την κατά Bayes συνολική πιθανότητα του κυρίαρχου χεριού. Για υπολογιστικούς λόγους, όπως συνηθίζεται, χρησιμοποιούμε αντί του γινομένου πιθανοτήτων, το άθροισμα των λογαρίθμων αυτών, όπως φαίνεται στην εξίσωση (5.2). Αυτός ο κανόνας είναι χρήσιμος για να εξαιρεθούν περιπτώσεις όπου περνούν εσφαλμένα τον πρώτο περιορισμό διότι το Openpose κάνει λάθος εκτίμηση λόγω θολώματος από κίνηση (motion bluring) αλλά κυρίως οφελεί στην αποφυγή περιπτώσεων κακής εκτίμησης λόγω του ότι ολόκληρο το χέρι ή κάποια δάκτυλα κρύβονται στο πλάνο. Οι περιπτώσεις αυτές οδηγούν σε παραμορφωμένα αποτελέσματα όπου η εκτιμούμενη διάταξη είναι εντελώς αφύσικη. Έτσι, μειώνεται μεν ποσοτικά το τελικό σύνολο δεδομένων, έχει δε μεγαλύτερη αξιοπιστία.

Οι δύο παραπάνω κανόνες εξασφαλίζουν ότι κάθε στιγμιότυπο χαρακτηρίζεται από στατικότητα (μικρή κινητικότητα) και μεγάλη βεβαιότητα ότι η εκτίμηση για την θέση κάθε σημείου αντιστοιχεί πολύ κοντά στην πραγματική. Ωστόσο, κάθε βίντεο από το σύνολο δεδομένων ξεκινά και κλείνει με τον νοηματιστή να έχει τα χέρια σταυρωμένα στη μέση (Σχήμα 5.6). Αυτό το στήσιμο του σώματος ανήκει στις περιπτώσεις που περνούν και τους δύο προηγούμενους κανόνες καθώς το χέρι παραμένει ακίνητο ενώ το Openpose εκτιμά ορθά με μεγάλη ακρίβεια κάθε σημείο. Για να εξαιρεθούν οι περιπτώσεις αυτές αρκεί να προσθέσουμε έναν τελευταίο κανόνα που απαιτεί σε ένα δεδομένο στιγμιότυπο ο καρπός του κυρίαρχου χεριού να βρίσκεται ψηλότερα από τον αγκώνα και μαθηματικοποιείται μέσω της εξίσωσης (5.3).

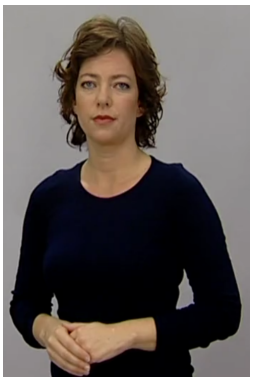
Τελικά, για το Δανικό σύνολο από τα 363131 καρέ, τα 59819 (16%) περνούν τους τρεις περιορισμούς και αποτελούν το σύνολο πάνω στο οποίο εκπαιδεύονται τα μοντέλα για την αναγνώριση των χειρομορφών και του προσανατολισμού του κυρίαρχου χεριού.

Έστω $\mathbf{X}_\tau = \begin{bmatrix} x_{0,\tau} & y_{0,\tau} & \sigma_{0,\tau} \\ x_{1,\tau} & y_{1,\tau} & \sigma_{1,\tau} \\ \vdots & \vdots & \vdots \\ x_{11,\tau} & y_{11,\tau} & \sigma_{11,\tau} \\ x^h_{0,\tau} & y^h_{0,\tau} & \sigma^h_{0,\tau} \\ x^h_{1,\tau} & y^h_{1,\tau} & \sigma^h_{1,\tau} \\ \vdots & \vdots & \vdots \\ x^h_{20,\tau} & y^h_{20,\tau} & \sigma^h_{20,\tau} \end{bmatrix}$ το τροποποιημένο διάνυσμα εξόδου του Openpose, όπου $\begin{bmatrix} x_{i,\tau} & y_{i,\tau} & \sigma_{i,\tau} \end{bmatrix}$ αφορά το $i^{\sigma\tau o}$ σημείο κλειδί του σώματος και με τον εκθέτη h γίνεται αναφορά στα αντίστοιχα σημεία του κυρίρχου χεριού.

$$\sum_{i=0}^{20} \max(H_{1i}, H_{2i}) < T \tag{5.1}$$

$$\sum_{i=0}^{20} log(\sigma_{i,\tau}^{hand}) < T_\sigma \tag{5.2}$$

$$\begin{cases} y_{3,\tau} < y_{4,\tau},\ righthanded \\ y_{6,\tau} < y_{7,\tau},\ lefthanded \end{cases} \tag{5.3}$$

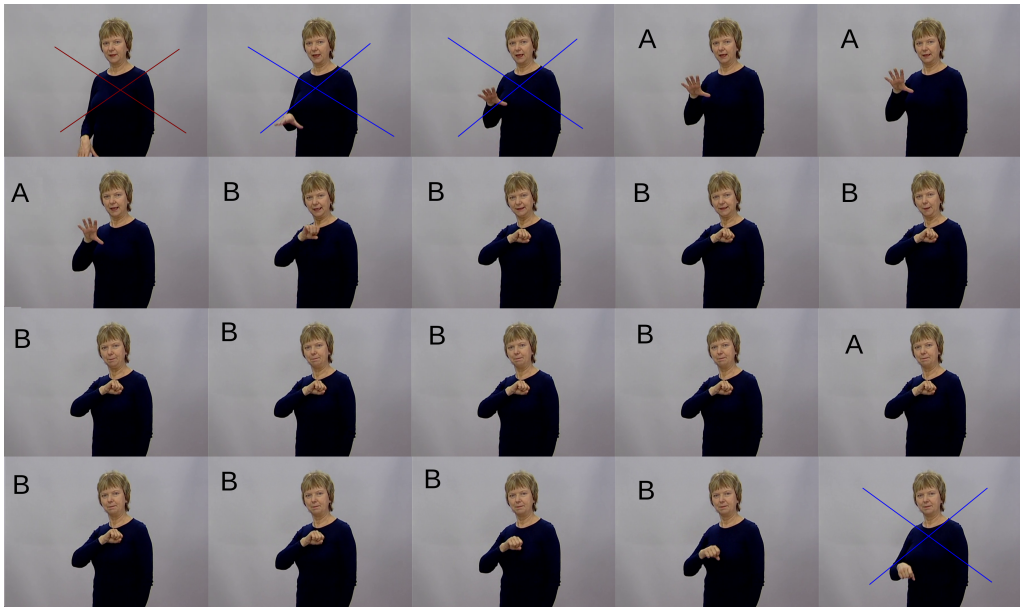


Σχήμα 5.6: Αρχική στάση νοηματίστριας

Στο σημείο αυτό, θεωρώντας ότι η μέθοδος εκκαθάρισης έχει οδηγήσει στα θεμιτά αποτελέσματα, έχουν παραμείνει τα στιγμιότυπα στα οποία αναφέρεται η επισημείωση και κανένα από τα στιγμιότυπα εκτός νοημάτισης. Για χάριν απλότητας, έστω ότι εξετάζεται η περίπτωση δημιουργίας του δατασετ για την περίπτωση αναγνώρισης χειρομορφών. Από τη στιγμή που κάθε βίντεο περιέχει μόνο μία λέξη, γνωρίζουμε ότι θα απεικονίζονται το πολύ δύο διαφορετικές χειρομορφές, αντίστοιχα.

Στην περίπτωση όπου το βίντεο απεικονίζει μόνο μία χειρομορφή, τότε στο τελικό σύνολο δεδομένων , υποθέτουμε ότι όλα τα στιγμιότυπα που πέρασαν τους παραπάνω κανόνες αναπαριστούν τη συγκεκριμένη χειρομορφή.

Στην περίπτωση όπου δύο διαφορετικές χειρομορφές εμφανίζονται, γνωρίζουμε για τα στιγμιότυπα που έχουν απομείνει ότι τα πρώτα τα αναπαριστούν τη μία χειρομορφή και τα υπόλοιπα την άλλη. Ωστόσο, δεν γνωρίζουμε σε ποιο σημείο γίνεται η συγκεκριμένη μετάβαση. Για το σκοπό αυτό, γίνεται χρήση μη επιβλεπούμενης μάθησης. Συγκεκριμένα, χωρίζουμε τα δεδομένα προς επισημείωση σε δύο συστάδες υποθέτοντας ότι μορφολογικά ίδιες χειρομορφές θα κατηγοριοποιηθούν μαζί. Έτσι, έχοντας δύο συστάδες και δύο χειρομορφές προς αντιστοίχιση για τις οποίες γνωρίζουμε τη σειρά εμφάνισής τους, κατηγοριοποιούμε στην πρώτη χειρομορφή τα στοιχεία της συστάδας που εμφανίζεται χρονικά κατά μέσο όρο νωρίτερα και αντίστοιχα στη δεύτερη χειρομορφή την άλλη συστάδα. Ο αλγόριθμος που επιλέγεται είναι το Μοντέλο Μίξης Γκαουσιανών (Gaussian Mixture Model). Το μοντέλο αυτό επιλέγεται πολύ συχνά σε συνδυασμό με Κρυφά Μαρκοβιανά Μοντέλα για την ευθυγράμμιση των επισημειώσεων με τα καρέ του βίντεο [23][30].



Σχήμα 5.7: Υποθετικό αποτέλεσμα μετά την κατάτμηση μίας ακολουθίας με 2 χειρομορφές

Στο Σχήμα 5.7 παρουσιάζεται ένα υποθετικό παράδειγμα της διαδικασίας κατάτμησης όπως προέκυψε από την επεξεργασία της ακολουθίας του Σχήματος 5.4. Με κόκκινο X έχει διαγραφεί το πρώτο στιγμιότυπο του βίντεο λόγω του ότι ο καρπός βρίσκεται χαμηλότερα από το ύψος του αγκώνα. Με μπλε X έχουν διαγραφεί τα στιγμιότυπα που αφορούν την μετάβαση από την αρχική θέση του χεριού στον ουδέτερο νοηματικό χώρο και στο τέλος προς τα πίσω. Από τα στιγμιότυπα που προέκυψαν θεωρητικά έχουν δημιουργηθεί οι συστάδες A,B. Εφόσον η συστάδα A βρίσκεται κατά μέσο όρο νωρίτερα από τη B, κάνουμε την αντιστοίχιση $A \rightarrow$  ,$B \rightarrow$  .

Επιπλέον, αξίζει να παρατηρήσουμε ότι στο συγκεκριμένο παράδειγμα έχουμε θεωρήσει ότι ο αλγόριθμος κάνει ένα ένα λάθος κατά την κατάτμηση θεωρώντας ότι ένα στιγμιότυπο θα ομαδοποιηθεί στην κατηγορία A παρότι ξεκάθαρα ανήκει στην B. Στην πραγματικότητα η μέθοδος είναι αρκετά ακριβής ώστε να αποφεύγει τέτοια εξόφθαλμα λάθη, ωστόσο, σε περιπτώσεις όπου οι δύο χειρομορφές μοιάζουν αρκετά μεταξύ τους ή κυρίως όταν τα μεταβατικά στιγμιότυπα δεν έχουν την απαραίτητη έντονη κινητικότητα ώστε να διαγραφούν, τέτοια λάθη μπορεί να συμβούν. Το θετικό στοιχείο είναι ότι τα λάθη αυτά δεν συμβαίνουν για πολλά συνεχόμενα στιγμιότυπα. Το ίδιο λάθος μπορεί να συμβεί αν ο νοηματιστής σταυρώσει τα χέρια του στο κλείσιμο του βίντεο πάνω από το ύψος της μέσης καταστρατηγώντας τον τρίτο κανόνας. Η λύση σε αυτό είναι αφού έχουμε θέσει τις τελικές τιμές των 2 κατωφλίων, μπορούμε σαρώνοντας πολύ γρήγορα χειροκίνητα τα στιγμιότυπα κάθε κλάσης να διαγράψουμε τέτοιες αστοχίες καθώς ξεχωρίζουν ανάμεσα στα υπόλοιπα λόγο ομοιομορφίας. Το

πλήθος τέτοιων λαθών επηρεάζει το τελικό αποτέλεσμα της αναγνώρισης κατά περίπου μόλις 0.3% και μετατρέπει την χειροκίνητη διόρθωση σε μία διαδικασία λίγων ωρών. Η αντίστοιχη διαδικασία αν γινόταν εξ αρχής χειροκίνητα θα χρειάζονταν δεκάδες εργατοώρες.

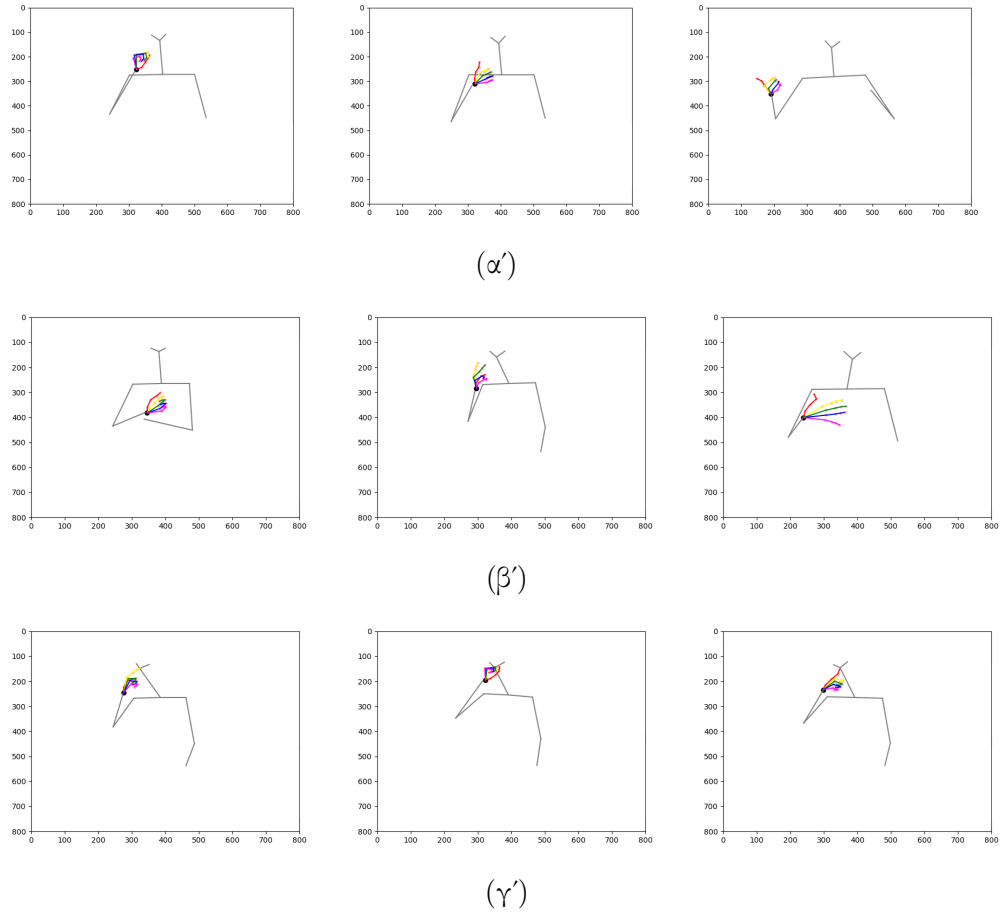
5.3 Εκπαίδευση

Έχοντας διαχωρίσει τα δεδομένα στιγμιότυπα βάσει της μεθόδου στην Ενότητα 5.2, έχουμε δημιουργήσει ένα σύνολο εκπαίδευσης (Training Set) για κάθε ένα από τα στατικά φωνολογικά χαρακτηριστικά (θέση, προσανατολισμός και σχήμα παλάμης). Κοιτώντας τα τελικά αποτελέσματα η κατάτμηση κάθε στατικού χαρακτηριστικού έχει μεγάλη επιτυχία η οποία όπως θα φανεί στις επόμενες ενότητες θα οδηγήσει σε αναλόγως μεγάλες αποδόσεις στην αναγνώριση κάθε χαρακτηριστικού. Έχοντας δημιουργήσει τις κλάσεις για κάθε χαρακτηριστικό, παρατηρούμε ότι όπως φαίνεται στα Σχήματα 5.8.

5.3.1 Σχήμα Χεριού

Για την αναγνώριση του σχήματος του χεριού χρησιμοποιήθηκαν μόνο τα σημεία που δίνονται από το Openpose για την περιγραφή του αντίστοιχου κυρίαρχου χεριού. Συνεπώς, το μέγεθος του διανύσματος εισόδου του μοντέλου είναι 21×2. Παρόμοια διαδικασία μελετήθηκε από το [14] χρησιμοποιώντας και τις τρεις συνιστώσες κάθε σημείου κλειδιού (keypoint). Η διαφορά σε σχέση με την παρούσα εργασία είναι ότι δεν γίνεται διαχωρισμός στα επιμέρους καρέ του βίντεο αλλά χρησιμοποιείται ολόκληρη η ακολουθία με σκοπό την εκπαίδευση σε επίπεδο προτάσεων. Επί της ουσίας, στην δική μας περίπτωση γίνεται χρήση της πληροφορίας που δίδεται μέσω της τρίτης διάστασης κάθε σημείου στο στάδιο της διαδικασίας κατάτμησης. Επιπλέον, η συμπερίληψη αυτής στα δεδομένα εκπαίδευσης μεγαλώνει το μέγεθος του διανύσματος και κατά αναλογία αυξάνεται το μέγεθος των παραμέτρων του μοντέλου. Δοκιμάζοντας και τις δύο περιπτώσεις δεν υπάρχει διαφορά στην απόδοση του κάθε μοντέλου και για αυτό επιλέξαμε να λάβουμε υπόψιν μόνο τις δύο χωρικές διαστάσεις.

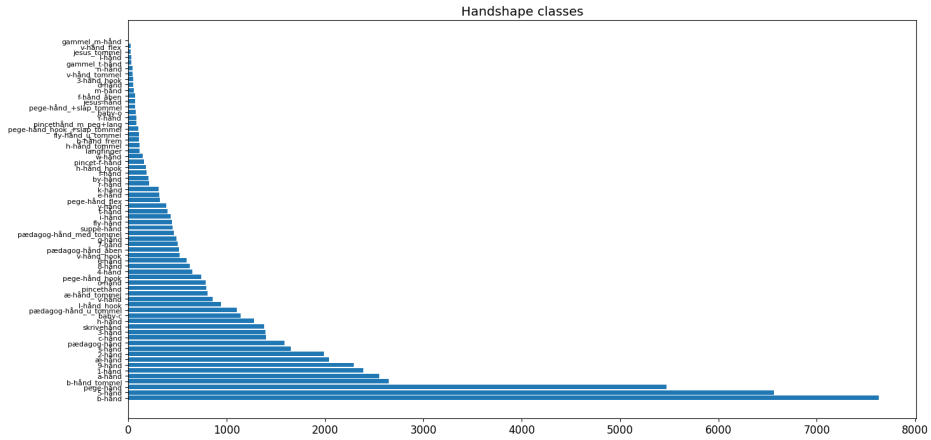
Οι συνιστώσες αυτές, όπως επιστρέφονται από το Openpose, είναι εκφρασμένες σε σύστημα συντεταγμένων που έχει ως σημείο αναφοράς την επάνω αριστερή γωνία της εικόνας και ως μονάδα μέτρησης χρησιμοποιείται το μέγεθος ενός εικονοστοιχείου (pixel). Πρέπει να σημειωθεί, λοιπόν, ότι η ανίσωση 5.3 είναι εκφρασμένη κατά τη διαισθητική σύμβαση ότι ο κατακόρυφος άξονας έχει φορά από κάτω προς τα επάνω. Υλοποιώντας στην πράξη τον συγκεκριμένο κανόνα επί της εξόδου του Openpose, θα πρέπει να αντιστραφεί η φορά των ανισώσεων. Επιπλέον, γίνεται η σύμβαση ότι το κυρίαρχο χέρι είναι το δεξί χέρι. Προκειμένου το μοντέλο να είναι συνεπές στις περιπτώσεις όπου ο νοηματιστής είναι αριστερόχειρας, όπως στην περίπτωση του «Πολύτροπον», αρκεί να εφαρμοστεί μετασχηματισμός κατοπτρικής συμμετρίας. Για αυτό το λόγο, ακόμα και η επισημείωση του Πολύτροπον στο HamNoSys έχει γίνει σε επίπεδο προσανατολισμών με βάση κάποιον δεξιόχειρα νοηματιστή.



Σχήμα 5.8: Παραδείγματα διαφορετικών περιπτώσεων κάθε στατικού χαρακτηριστικού (α′) σχήμα χεριού (s/) (β′) προσανατολισμός (προς τα μέσα) (γ′) θέση χεριού (μύτη)

Στο Σχήμα 5.9 φαίνεται η κατανομή των κλάσεων της χειρομορφής με βάση το πλήθος εμφανίσεων εντός του συνόλου δεδομένων. Παρατηρούμε ότι η κατανομή ακολουθεί και πάλι την ακολουθία που περιμένουμε από κάποια λεξιλογική οντότητα μέσα σε φυσικό λόγο. Δυστυχώς, αυτό είναι μία ιδιαιτερότητα που κάποιος που ασχολείται με ένα τέτοιο πρόβλημα δεν μπορεί να κάνει κάτι. Μία λύση σε αυτό θα ήταν να γίνει προσαύξηση των δεδομένων μεγεθύνοντας και περιστρέφοντας τα δεδομένα στις υποεκπροσωπούμενες κλάσεις. Ωστόσο, κάτι τέτοιο δε θα μπορούσε να λειτουργήσει διότι ούτως ή άλλως η μέθοδος, λόγω κανονικοποίησης των δεδομένων, δεν επηρεάζεται από μεγέθυνση ή σμίκρινση. Επιπλέον, οι κλάσεις με λίγα παραδείγματα στην πραγματικότητα αποτελούνται έχουν μόνο μερικές εμφανίσεις.

Τα δεδομένα στην τελική τους μορφή αποτελούνται από 21 διατεταγμένα ζεύγη με αποτέλεσμα να μην υπάρχει λόγος για τη χρήση ενός συνελικτικού ή αναδρομικού



Σχήμα 5.9: Αριθμός εμφανίσεων κάθε χειρομορφής

νευρωνικού δικτύου. Έτσι, χρησιμοποιήθηκε ένα απλό βαθύ νευρωνικό δίκτυο τύπου MLP (MultiLayer Perceptron).

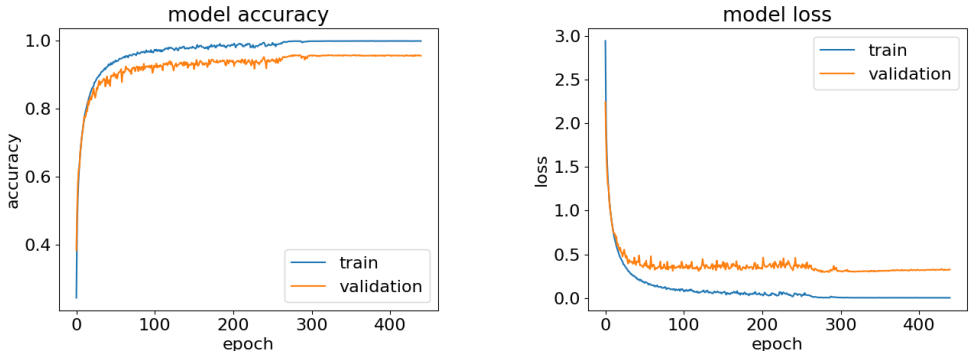
Το διάνυσμα χαρακτηριστικών κάποιου στιγμιότυπου τ πάνω στο οποίο θα εκπαιδευτεί το μοντέλο, κανονικοποιείται σε κάθε διάσταση όπως φαίνεται στην εξίσωση 5.4, θέτοντας το μέσο όρο στο (0,0) και τη διακύμανση ίση με 1.

$$\mathbf{X^*}_\tau = \left[\frac{\mathbf{x^h_{i,\tau}} - \mathbf{\bar{x}^h_\tau}}{\mathbf{s_{x^h,\tau}}} \quad \frac{\mathbf{y^h_{i,\tau}} - \mathbf{\bar{y}^h_\tau}}{\mathbf{s_{y^h,\tau}}} \right]_{\mathbf{i=0,1,...,20}} \tag{5.4}$$

,όπου $s_{x^h,\tau}, s_{y^h,\tau}$ οι τυπικές αποκλίσεις των x^h_τ, y^h_τ, αντίστοιχα.

Δοκιμάζοντας συνδυασμούς από αρχιτεκτονικές για το μοντέλο αναγνώρισης σχήματος χεριού καταλήξαμε ότι το μοντέλο με την μεγαλύτερη ακρίβεια είναι εκείνο με αρχιτεκτονική που αποτελείται από 5 κρυφά επίπεδα, με 128 νευρώνες στο κάθε ένα με συνάρτηση ενεργοποίησης ReLu και συνολικό χρόνο εκπαίδευσης 440 εποχές. Για την περίπτωση της αναγνώρισης σχήματος χεριού, οι διαφορετικές κλάσεις είναι 66 στο πλήθος. Στο Σχήμα 5.10 φαίνεται η απόδοση του μοντέλου αναγνώρισης σχήματος χεριού. Παρατηρούμε, ότι η απόδοση του μοντέλου μετά τις 380 εποχές τείνει να σταθεροποιείται με μερικές μικρές διακυμάνσεις. Οι μετρικές επί του συνόλου επικύρωσης (validation set) ακολουθούν την συμπεριφορά των μετρικών στο σύνολο εκπαίδευσης (training set), χωρίς το μοντέλο να έχει τάση προς το overfiting. Στις 440 εποχές, η ακρίβεια επί του συνόλου εκπαίδευσης έχει φθάσει στο 99,9% και συνεπώς δεν υπάρχει νόημα να συνεχιστεί η εκπαίδευση πέραν αυτού του σημείου (Σχήμα 5.10α΄). Η τελική ακρίβεια πάνω στο σύνολο ελέγχου (test set) είναι **95.7%**.

Στο Σχήμα 5.14 φαίνεται η μήτρα συγχέσεων (Confussion Matrix) για τις κλάσεις των σχημάτων χεριού. Το μεγάλο πλήθος των κλάσεων σε συνδυασμό με την πολύ υψηλή ακρίβεια αναγνώρισης έχει ως αποτέλεσμα να μην είναι ευδιάκριτο στον πίνακα



(α′) (β′)

Σχήμα 5.10: Απόδοση Μοντέλου αναγνώρισης σχήματος χεριού

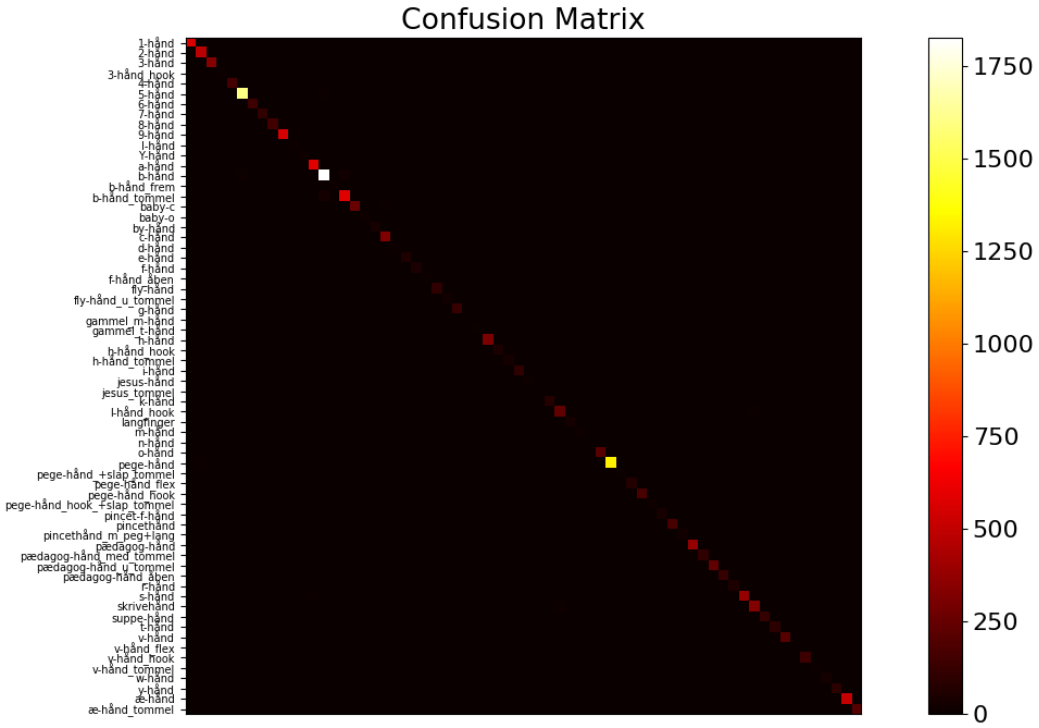
ποιες κλάσεις συγχέονται με ποιες.

5.3.2 Προσανατολισμός Παλάμης

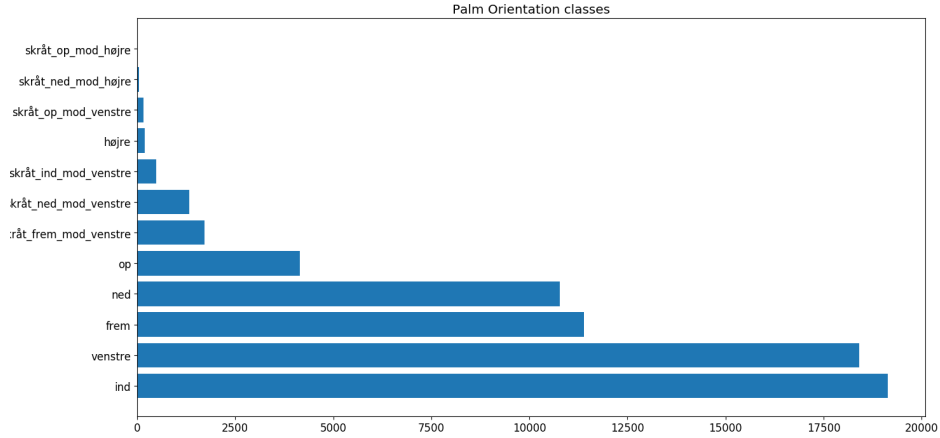
Για το μοντέλο αναγνώρισης προσανατολισμού της παλάμης χρησιμοποιήσαμε την ίδια μέθοδο κατάτμησης χωρίς καμία αλλαγή στη μέθοδο της μη επιβλεπούμενης μάθησης που οδηγεί στην απομόνωση περίπου των ίδιων στιγμιοτύπων τα οποία ομαδοποιούνται τώρα σύμφωνα με το προσανατολισμό που δίνεται από την επισημείωση. Οι διαφορές που μπορεί να εμφανίζονται αφορούν περιπτώσεις όπου έχουμε μία χειρονομία σε διαφορετικούς προσανατολισμούς ή το αντίστροφο. Όπως αναφέρθηκε στο Κεφάλαιο 2.3, το HamNoSys περιγράφει τον προσανατολισμός της παλάμη σε σχετικό σύστημα αναφοράς ως προς την κατεύθυνση προς την οποία δείχνουν τα δάκτυλα.

Στο σχήμα 5.12 φαίνεται η κατανομή των κλάσεων του προσανατολισμού χεριού. Ομοίως με πριν οι κλάσεις κατανέμονται έντονα ανισόρροπα. Ωστόσο σε αυτήν την περίπτωση, όλες σχεδόν οι κλάσεις εκπροσωπούνται από τουλάχιστον μερικές εκατοντάδες παραδείγματα. Μόνο τέσσερις προσανατολισμοί έχουν ελάχιστα παραδείγματα. Οι προσανατολισμοί αυτοί είναι οι «κλίση προς τα πάνω δεξιά», «κλίση προς τα κάτω δεξιά» και «δεξιά» (højre=δεξιά) οι οποίοι αντιστοιχούν σε αφύσικες θέσεις του χεριού για κάποιον δεξιόχειρα νοηματιστή.

Έχοντας ολοκληρώσει την διαδικασία κατάτμησης των στιγμιοτύπων βάσει του προσανατολισμού της παλάμης, χρησιμοποιούμε το τελικό dataset για την εκπαίδευση του δεύτερου μοντέλου. Το μοντέλο αυτό είναι ένα νευρωνικό δίκτυο με τα ίδια αρχιτεκτονικά χαρακτηριστικά με εκείνα του δικτύου αναγνώρισης σχήματος χεριού. Φυσικά, τα δύο μοντέλα διαφέρουν ως προς το πλήθος των εξόδων λόγω διαφορετικού πλήθους κλάσεων και στην είσοδο λόγω διαφορετικού διανύσματος κατάστασης. Από την περιγραφή του τρόπου επισημείωσης στο Κεφάλαιο 2.3 προκύπτει ότι ο προσανατολισμός της παλάμης μπορεί να έχει 16 βαθμούς ελευθερίας. Ωστόσο, κάποιοι εξ αυτών αντιστοιχούν σε αφύσικες στάσεις του χεριού όπως την περίπτωση όπου (σε



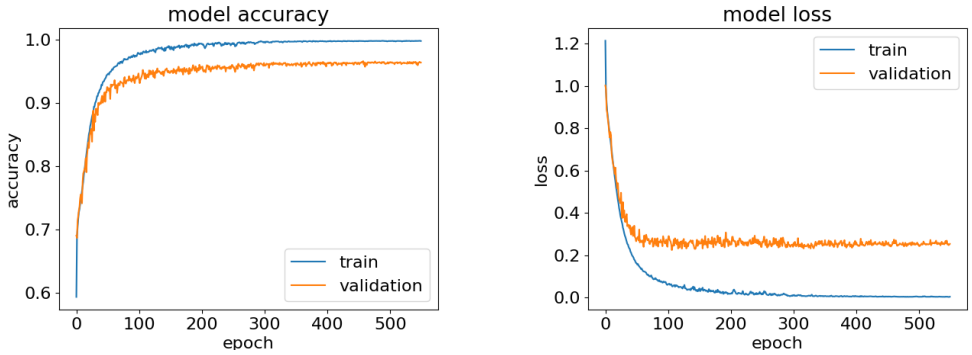
Σχήμα 5.11: Μητρα Συγχέσεων για τις κλάσεις του μοντέλου αναγνώρισης σχήματος χεριού (βλ. σελ. 66)



Σχήμα 5.12: Αριθμός εμφανίσεων κάθε προσανατολισμού χεριού

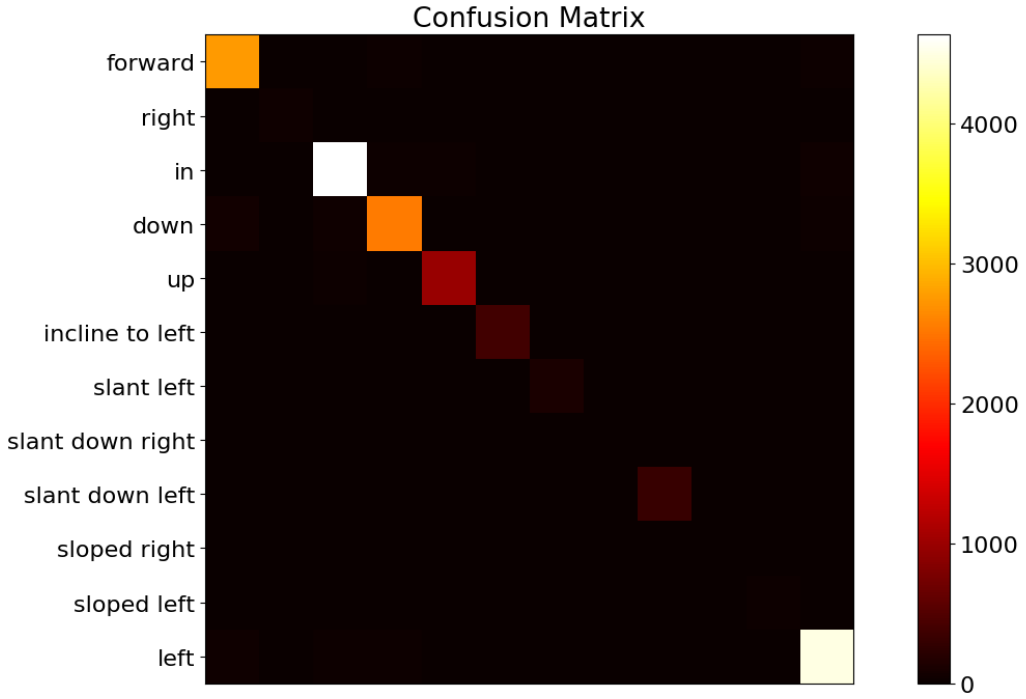
δεξιόχειρα νοηματιστή) τα δάκτυλα έχουν διεύθυνση προς τα αριστερά ενώ η παλάμη είναι στραμμένη διαγώνια επάνω δεξιά ως προς τα δάκτυλα. Ο τελικός αριθμός των

κλάσεων είναι 12 με ορισμένες από αυτές να είναι ιδιαίτερα σπανιότερες από άλλες όπως συνέβη και στην προηγούμενη περίπτωση. Τελικά, μετά από παρόμοια διάρκεια εκπαίδευσης η τελική ακρίβεια του μοντέλου είναι 96.4% ενώ το μοντέλο έχει παρόμοια συμπεριφορά με την προηγούμενη περίπτωση όπως φαίνεται στο Σχήμα 5.13.



(α′) (β′)

Σχήμα 5.13: Απόδοση Μοντέλου αναγνώρισης προσανατολισμού χεριού



Σχήμα 5.14: Μητρα Συγχέσεων για τις κλάσεις του μοντέλου αναγνώρισης προσανατολισμού χεριού

5.3.3 Θέση Χεριού

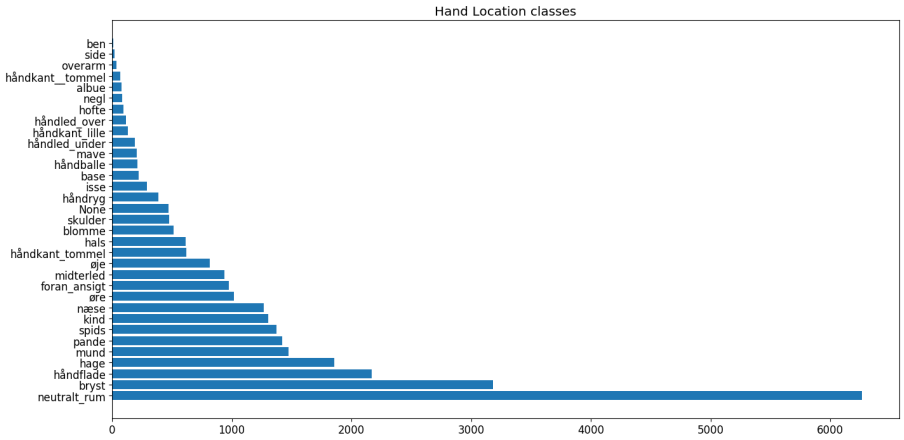
Το τελευταίο στατικό χαρακτηριστικό αποτελεί η θέση του χεριού στο χώρο νοημάτισης. Το συγκεκριμένο χαρακτηριστικό απαιτεί μία τροποποίηση στον τρόπο κατάτμησης των δεδομένων σε σχέση με τις προηγούμενες 2 περιπτώσεις. Ο λόγος είναι ότι το ένα από τα 3 κριτήρια για την κατάτμηση των στιγμιοτύπων αφορά τη συνολική κινητικότητα των χαρακτηριστικών σημείων του χεριού. Προηγουμένως, απαιτήθηκε σχετικά μικρή κινητικότητα αλλά αρκετά μεγάλη προκειμένου να συμπεριληφθούν τα διαστήματα όπου η κίνηση είναι μέρος της νοημάτισης. Όμως, στην παρούσα περίπτωση τα στιγμιότυπα που απεικονίζουν το χέρι στην επισημειωμένη θέση χαρακτηρίζονται από σχεδόν μηδαμινή κινητικότητα αφού ακόμα και η κίνηση κατά τη νοημάτιση αποτελεί μία μετάβαση θέσεων εντός του χώρου. Έτσι, ακολουθούμε την ίδια μέθοδο επισημείωσης θέτοντας το κατώφλι στην κίνηση πολύ χαμηλά. Με αυτόν τον τρόπο συλλαμβάνουμε μόνο τα στιγμιότυπα όπου το χέρι έχει καταλήξει στην περιγραφούμενη θέση ως προς το σώμα. Έτσι, ο συνολικός αριθμός στιγμιοτύπων που απομένουν είναι πολύ μικρότερος.

Ως προς την κατανομή των κλάσεων (Σχήμα 5.15), σε αυτήν την κατηγορία, όπως και στις υπόλοιπες είναι ανισόρροπα κατανεμημένες λόγω του ότι ακολουθούν το νόμο του Zipf. Ωστόσο, η πρώτη σε συχνότητα εμφάνισης κλάση τώρα έχει αναλογικά πολύ μεγαλύτερο πλήθος εμφανίσεων. Η αιτία είναι ότι αφορά την περίπτωση νοημάτισης στον «ουδέτερο χώρο νοημάτισης». Τα περισσότερα νοήματα όπως αναλύσαμε στο Κεφάλαιο 2 χρησιμοποιούν καθένα από τον χώρο νοημάτισης και την κίνηση για λόγους διαχωρισμού μεταξύ νοημάτων αλλά κυρίως προκειμένου να εξυπηρετείται ένας πιο γλαφυρός οπτικά τρόπος απόδοσης. Ωστόσο, στα περισσότερα παραδείγματα που παρουσιάστηκαν παραπάνω δεν γίνεται χρήση κάποιου ιδιαίτερου μέρους του σώματος αλλά γενικά τα χέρια είναι τοποθετημένα σε ένα ουδέτερο σημείο κοντά στο ύψος του στήθους και λίγο εκτεταμένα προς τα εμπρός ώστε να είναι περισσότερο βολικό για τον νοηματιστή. Με αυτόν τον τρόπο νοήματα στοιχειωδών και συχνών εννοιών αποτυπώνονται με έναν τρόπο που είναι ποιο ξεκούραστος τόσο για το νοηματιστή όσο και για τον παρατηρητή που δεν χρειάζεται να ακολουθεί πολλές απότομες κινήσεις.

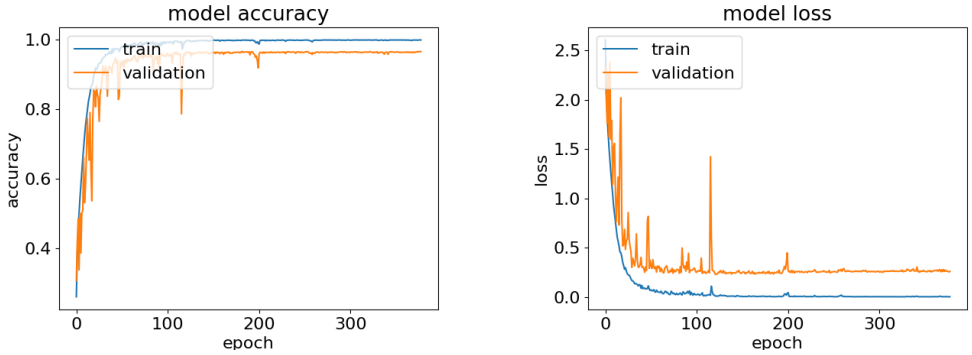
Σε αυτό το σημείο πλέον δεν μπορεί να υπάρξει περαιτέρω μειώση στη διαστατικότητα του διανύσματος εισόδου. Κάθε σημείο που δίνεται από το Openpose είναι απαραίτητο καθώς η περιγραφή της θέσης του χεριού αφορά οποιοδήποτε σημείο είτε του σώματος είτε του δευτερεύοντος χεριού. Ως προς την επιλογή του μοντέλου, αρχικά επιλέγεται ένα νευρωνικό με 6 επίπεδα από 256 νευρώνες το κάθε ένα, το οποίο επιτυγχάνει ακρίβεια **96,8%**. Ωστόσο, η αναγνώριση της θέσης του χεριού είναι μία περίπτωση που βρίσκεται διαισθητικά πολύ κοντά στη μέθοδο του k nearest neightbours με την έννοια ότι έχει καθαρά ένα χαρακτήρα εντοπισμού μέσα σε ένα γεωμετρικό χώρο. Η μέθοδος knn πετυχαίνει ποσοστό 96.3% για $k = 3$.

5.3.4 Κίνηση

Το τελευταίο χαρακτηριστικό που απομένει για να έχουμε έναν πλήρη ταξινομητή λέξεων της νοηματικής είναι να δημιουργήσουμε ένα μοντέλο αναγνώρισης κίνησης.



Σχήμα 5.15: Αριθμός εμφανίσεων κάθε θέσης χεριού

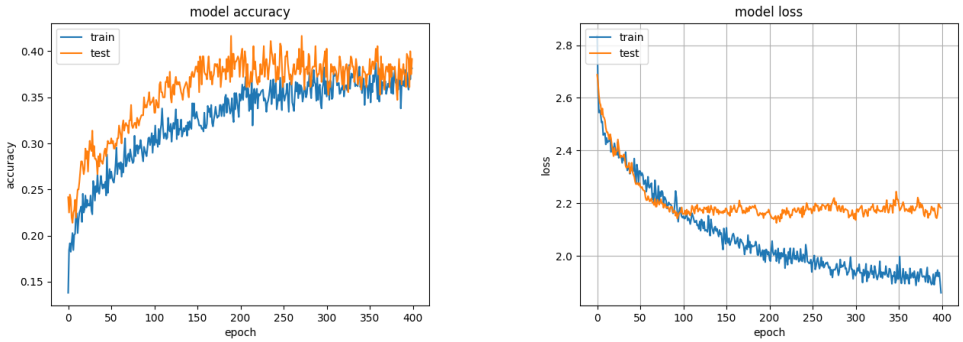


(α′) (β′)

Σχήμα 5.16: Απόδοση Μοντέλου αναγνώρισης θέσης χεριού

Δυστυχώς, πλέον δεν μιλάμε για ένα χαρακτηριστικό που μπορεί να εντοπιστεί στατικά εντός ενός στιγμιότυπου του βίντεο. Για αυτό το λόγο ένα σύστημα αναγνώρισης κινήσεων θα πρέπει να λαμβάνει ως είσοδο ολόκληρη την ακολουθία εικόνων, γεγονός που συνεπάγεται ότι από κάθε βίντεο αντλείται ένα στοιχείο του συνόλου δεδομένων σε αντίθεση με τις προηγούμενες περιπτώσεις όπου από κάθε βίντεο εξαγάγαμε τουλάχιστον 5 στοιχεία. Έτσι αντί να έχουμε μερικές χιλιάδες παραδείγματα από τις συνηθέστερες κλάσεις, έχουμε συνολικά 1600 συνολικά. Το πλήθος αυτό φαινομενικά είναι αρκετά μεγάλο για τις συνολικά 21 κλάσεις, ωστόσο υπάρχουν δύο βασικοί παράγοντες για τους οποίους η αναγνώριση κινήσεων πρακτικά δεν μπορεί να ξεπεράσει το 40% για το δεδομένο dataset. Ο βασικότερος είναι ότι η κίνηση έχει μία αρκετά αόριστη περιγραφή τις περισσότερες φορές. Εισάγοντας μέσα από το Ham-

NoSys όρους όπως "repeat from start several" που αφορά επαναλήψεις κινήσεων, η απόδοση ενός νοήματος μπορεί να ποικίλει σημαντικά από νοηματιστή σε νοηματιστή. Γενικότερα, πέραν της αοριστίας στην περιγραφή των κινήσεων, ένα σημαντικό εμπόδιο είναι η επιλογή από την πλευρά των νοηματιστών να θυσιάζουν την καθαρότητα χάριν συντομίας. Για παράδειγμα, η κίνηση «προς τα εμπρός» μπορεί να είναι άλλοτε μία σαφής κίνηση από το στήθος προς τα έξω ή απλά ένα τίναγμα του χεριού ανάλογα την έμφαση που θέλει ο νοηματιστής να δώσει στην εν λόγω λέξη. Επιπλέον, η μικρή ανάλυση στο χρόνο όπως είναι τα 25 καρέ/δευτερόλεπτο που χρησιμοποιούνται στην καταγραφή βίντεο, δεν είναι επαρκής στη μελέτη μικροκινήσεων όπως είναι το νόημα «μαμά» που πραγματοποιείται ακουμπώντας μερικές φορές το πιγούνι με τον αντίχειρα.



(α′) (β′)

Σχήμα 5.17: Απόδοση Μοντέλου αναγνώρισης τροχιάς χεριού

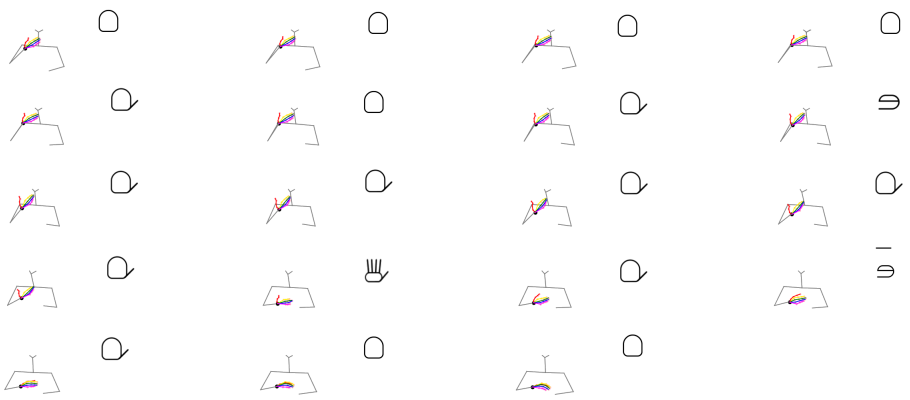
Για το μοντέλο αναγνώρισης τροχιάς χεριού πρέπει να απομονώσουμε από κάθε βίντεο ένα διάνυσμα κατάστασης το οποίο θα απεικονίζει τη τροχιά του χεριού στο χώρο νοημάτισης και εντός του χρονικού πλαισίου όπου πραγματοποιείται η κίνηση κατά τη νοημάτιση. Άλλες εργασίες [14][9], γενικότερα χρησιμοποιούν μηχανισμούς προσοχής για τον εντοπισμό των σημείων ενδιαφέροντος. Οι μηχανισμοί προσοχής (attention mechanisms) κλασσικά χρησιμοποιούνται στα πλαίσια της αναγνώρισης αντικειμένων εντός μίας εικόνας. Σε εφαρμογές όπως η δική μας οι μηχανισμοί αυτοί χρησιμοποιούνται σε χωροχρονικό πλαίσιο για τον εντοπισμό σημείων ενδιαφέροντος στα πλαίσια της αναγνώρισης σημείων ενδιαφέροντος στο χρόνο. Από τη στιγμή όπου η χρήση του OpenPose μας έχει λύσει το πρόβλημα της εύρεσης σημείων ενδιαφέροντος εντός ενός στιγμιότυπου, οι μηχανισμοί προσοχής θα είχαν ενδιαφέρον καθαρά στην διάσταση του χρόνου. Κάτι τέτοιο θα είχε ιδιαίτερο ενδιαφέρον στην περίπτωση της αναγνώρισης συνεχούς νοηματικού λόγου. Στην περίπτωση που εξετάζουμε κάθε βίντεο είναι το πολύ 3,5 δευτερόλεπτα και έχει συμμετρική δομή χρονικά με την έννοια ότι το κύριο μέρος του νοήματος εμφανίζεται στη μέση. Έτσι αρκεί να να δημιουργήσουμε ένα διάνυσμα κατάστασης με διαστάσεις $70 \times 40 \times 3$ όπου τα 40 σημεία αντιστοιχούν στα 19 σημεία του σώματος που είναι ορατά στην κάμερα συν τα 21 σημεία του κυρίαρχου χεριού και 70 τα συνολικά στιγμιότυπα της ακολουθίας. Σε όσα

βίντεο τα στιγμιότυπα που περνούν τους κανόνες είναι περισσότερα από 70 κρατάμε τα 70 που βρίσκονται στη μέση της ακολουθίας. Διαφορετικά προσθέτουμε μηδενικά στο τέλος της ακολουθίας (zero pading).

Για την αναγνώριση σε επίπεδο ακολουθίας συνήθως χρησιμοποιούνται τρισδιάστατα συνελικτικά νευρωνικά δίκτυα [10]. Ωστόσο, στην περίπτωσή μας εφόσον οι 2 χωρικές διαστάσεις έχουν ήδη επεξεργαστεί από το OpenPose και δεν υπάρχει πληροφορία με τη μορφή εικόνας, χρησιμοποιούμε μονοδιάστατα συνελικτικά νευρωνικά δίκτυα. Είναι μία αρχιτεκτονική δικτύου που δε συναντάμε πολύ συχνά σε σχέση με τα δισδιάστατα και τρισδιάστατα συνελικτικά δίκτυα που συναντάμε στην αναγνώριση εικόνας και βίντεο, αντίστοιχα. Τα συνελικτικά νευρωνικά δίκτυα χρησιμοποιούνται σε δεδομένα όπου κάθε συνιστώσα του διανύσματος κατάστασης δεν αντιστοιχούν σε κάποιο μεμονωμένο χαρακτηριστικό. Το πλεονέκτημα που έχουν είναι ότι δεν επηρεάζονται από τη μεγέθυνση και μετατόπιση του προτύπου εντός του χώρου κατάστασης. Στην προκειμένη περίπτωση το μέγεθος που καταλαμβάνει το πρότυπο αντιστοιχεί αντιστρόφως ανάλογα στην διάρκεια της κίνησης. Επιπλέον, μετατόπιση αντιστοιχεί στην χρονική μετατόπιση, δηλαδή σε ποια χρονική στιγμή αρχίζει η ζητούμενη κίνηση εντός της ακολουθίας.

Η αρχιτεκτονική του βέλτιστου δικτύου που ερευνήσαμε είναι 2 επίπεδα μονοδιάστατων συνελικτικών νευρωνικών με 32 φίλτρα με μέγεθος πυρήνα 5 και dropout 0.9 μετά από κάθε συνελικτικό δίκτυο ώστε να αποφευχθεί το οερφιττινγ. Το τελευταίο επίπεδο είναι ένα MLP δίκτυο με 32 κρυφά επίπεδα. Η τελική απόδοση επί το συνόλου αξιολόγησης είναι 39,4%.

5.4 Μεταφορά Μάθησης



Σχήμα 5.18: Παράδειγμα χειρομορφής B| αναγνωρισμένης με βάση το μοντέλο αναγνώρισης χεριού

Για να ολοκληρώσουμε την ανάλυσή μας θα θέλαμε μπορούσαμε να μεταφέρου-

με την πληροφορία από το σύστημα που εκπαιδεύσαμε βασισμένο στην συλλογή του λεξικού της Δανικής Νοηματικής Γλώσσας, σε κάποια άλλη γλώσσα. Συγκεκριμένα, θα θέλαμε να εξετάσουμε την απόδοση του μοντέλου πάνω στην Ελληνική Νοηματική Γλώσσα σε μία συλλογή όπως το «Πολύτροπον» ή το «Νόημα +». Θεωρητικά μία τέτοια ανάλυση θα μπορούσε να γίνει μεταφράζοντας ένα προς ένα τα τα φωνολογικά χαρακτηριστικά από το τροποποιημένο HamNoSys του πανεπιστημίου της Κοπεγχάγης στην τυποποιημένη μορφή του HamNoSys. Πάνω σε αυτό τη μέθοδο προκύπτουν δύο βασικά εμπόδια. Πρώτον, το Νόημα+ δεν είναι επισημειωμένο απευθείας σε HamNoSys[1] αλλά έχει γίνει μεταφορά της μετάφρασης από μία παλαιότερη συλλογή το 'Dicta-Sign¨. Ωστόσο, ο τρόπος δημιουργίας του «Νόημα+» και κατ΄ επέκταση του «Πολύτροπον» έχει γίνει ώστε να υπάρχει η μετάφραση από κάθε λέξη της Ελληνικής Γλώσσας στην ΕΝΓ. Για λόγους πληρότητας, λεξιλογικά έχει γίνει μία πολλή καλή δουλειά ώστε για κάθε λέξη να υπάρχουν όλα τα πιθανά νοήματα που μπορεί να την αναπαριστούν. Για παράδειγμα, η λέξη «κλείνω» έχει αναπαρασταθεί ως «κλείνω ένα καπάκι», «κλείνω ένα κατάστημα» και «κλείνω μία συμφωνία». Προσπαθώντας να προσεγγίσουμε το πρόβλημα από τη δική μας σκοπιά, βρίσκοντας κάποιο βίντεο κοιτάμε ότι μεταφράζεται ως «κλείνω», έπειτα κοιτώντας τον κατάλογο με τις επισημειώσεις βρίσκουμε μία περιγραφή του βίντεο σε HamNoSys η οποία εν δύο στις τρεις φορές θα είναι εσφαλμένη. Επιπλέον, για κάθε νόημα που έχει περισσότερους από έναν τρόπους εκφοράς εμφανίζονται όλες οι εκδοχές. Έτσι, η σχολαστικότητα στο σχεδιασμό του λεξικού οδηγεί παραδόξως σε μία κατάσταση όπου πολύ μεγάλο ποσοστό των νοημάτων είτε δεν έχουν ακόμη επισημείωση είτε δεν γνωρίζουμε αν η επισημείωση που βρίσκουμε είναι η σωστή.

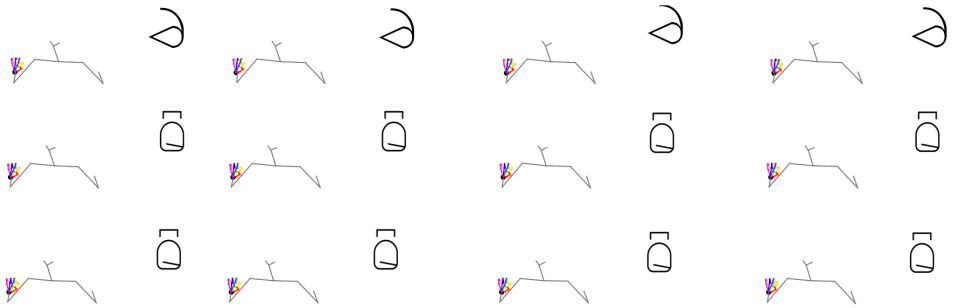
Έστω, όμως, ότι έχουμε ένα λεξικό με επισημείωση ένα προς ένα με τα βίντεο. Για ακόμα μία φορά η μεγάλη πληροφορία που είναι χρήσιμη στην ανθρώπινη μάθηση, έχετε σε αντιπαράθεση με την μηχανική μάθηση. Συγκεκριμένα, κοιτώντας την επισημείωση που υπάρχει πάνω στην ΕΝΓ υπάρχει η εμφάνιση 220 διαφορετικών περιγραφών για το σχήμα του χεριού! Ο λόγος είναι ότι το HamNoSys παρέχει μεγάλες ελευθερίες στην περιγραφή μέσα από την ύπαρξη 3 παραμέτρων για την αυτήν (γενικό σχήμα, θέση αντίχειρα και καμπυλώτητα δακτύλων), την δυνατότητα μεμονωμένη περιγραφής για την διάταξη κάθε δακτύλου και την ύπαρξη του τελεστή «\» (μεταξύ) που αναλύσαμε στην Ενότητα 2.3. Μάλιστα, ο κάθε επισημειωτής μπορεί να αντιλαμβάνεται με διαφορετικό τρόπο.

Εφόσον, λοιπόν δε μπορούμε να είμαστε βέβαιοι για το μεγαλύτερο μέρος της επισημείωσης δεν μπορούμε να έχουμε κάποιο ποσοτικό δείγμα της επιτυχίας της μεταφοράς μάθησης. Θα είχε γενικά, όμως, ενδιαφέρον να μελετήσουμε την απόδοση μεμονωμένα σε κάποια νοήματα.

Στο Σχήμα 5.18 φαίνεται το νόημα «αδιαφορώ» αφού έχει περάσει την μέθοδο φιλτραρίσματος στιγμιοτύπων όπως αναλύθηκε στις προηγούμενες ενότητες. Φυσικά, όπως αναφέραμε εφόσον ο νοηματιστής είναι αριστερόχειρας, έχουμε εφαρμόσει αρχικά μετασχηματισμό κατοπτρικής συμμετρίας (mirroring) στα σημεία εξόδου του

[1]Κατά την συγγραφή της παρούσας εργασίας, η απευθείας επισημείωση του «Νόημα+» έχει πλέον ξεκινήσει.

OpenPose. Για το νόημα αυτό χρησιμοποιείται η χειρομορφή ¨B|' η οποία στην ΔΝΓ αντιστοιχεί στην κλάση του μοντέλου μας 'b-hånd' που αποτελεί την πρώτη σε αριθμό εμφανίσεων κλάση. Ωστόσο, το μοντέλο μας ανά στιγμιότυπο ακροβατεί ανάμεσα στην συγκεκριμένη κλάση και στην ¨B/' (χειρομορφή ¨B' με ανοικτό αντίχειρα). Πράγματι, παρατηρώντας τα στιγμιότυπα είναι αρκετά σαφές ότι ο νοηματιστής έχει ανοικτό τον αντίχειρα. Αυτό είναι ένα πολύ χαρακτηριστικό παράδειγμα του γεγονότος ότι αφενός κάθε επισημειωτής μπορεί να έχει διαφορετική άποψη για το σχήμα κάθε χειρομορφής και ότι ένας νοηματιστής δεν πρόκειται ποτέ να μείνει σε ένα πολύ αυστηρό τρόπο «διάρθρωσης» ενός νοήματος.



Σχήμα 5.19: Παράδειγμα χειρομορφής Έ' αναγνωρισμένης με βάση το μοντέλο αναγνώρισης χεριού

Πέρα από τη χειρομορφή «B|» που είναι η συχνότερη και το σύστημα είναι λογικό να την αναγνωρίζει, παρατηρούμε στο Σχήμα 5.19 ότι η χειρομορφή «E» που αντιστοιχεί στην κλάση του μοντέλου "e-hånd" λειτουργεί πάρα πολύ καλά. Σε κάποια στιγμιότυπα, παρότι η διαφορά είναι ανεπαίσθητη τα ερμηνεύει ως "9-hånd". Η διαφορά στις 2 κλάσεις έγκειται στην πράξη στο αν ο αντίχειρας ακουμπάει στον δείκτη με τα δάκτυλα τεντωμένα ή όχι. Λόγω κακής ανάλυσης το Openpose δεν επιστρέφει μεγάλες διαφορές στα διαδοχικά καρέ αλλά το είναι σε θέση να αναγνωρίσει ορθά την χειρομορφή.

5.5 Συμπεράσματα

Εν κατακλείδι, οι αποδόσεις στα τρία στατικά χαρακτηριστικά είναι ιδιαίτερα υψηλές. Πέραν της απλότητάς της, η συνολική διαδικασία είναι υπολογιστικά ιδιαίτερα ελαφριά ως προς το τελικό προϊόν. Για να χρησιμοποιηθούν τα αποτελέσματα της εργασίας ως ταξινομητές σε βίντεο αρκεί να ακολουθηθούν τα βήματα της μεθόδου με τις αντίστοιχες τροποποιήσεις. Αρχικά, το σύστημα θα πρέπει να περνάει διαδοχικά κάθε καρέ από το Openpose και να διατηρεί στη «μνήμη» τα προηγούμενα 4 καρέ του βίντεο. Με αυτό τον τρόπο αποδίδεται σε κάθε στιγμιότυπο μία τιμή στο μέτρο της μέσης συνολικής κινητικότητας με τον ίδιο τρόπο που αναλύθηκε κατά τη διαδικασία

της κατάτμησης. Έπειτα από κατωφλίωση το μοντέλο θα αναγνωρίζει ή θα αγνοεί, αντίστοιχα, το συγκεκριμένο στιγμιότυπο. Το βαρύτερο υπολογιστικά στάδιο είναι με σημαντική διαφορά το στάδιο της εξαγωγής χαρακτηριστικών, δηλαδή η μετατροπή της εικόνας μέσω του Openpose σε σημεία κλειδιά, δεδομένου ότι απαιτείται δημιουργία του σκελετού τόσο του σώματος όσο και των χεριών. Μία κάρτα γραφικών όπως είναι η "RTX-2060" μπορεί να παράξει 5 καρέ/δευτερόλεπτο που αποτελεί πολύ μικρότερο ρυθμό σε σχέση με τα 24 καρέ/δευτερόλεπτο που είναι ο συνηθέστερος στην καταγραφή βίντεο. Ωστόσο, ο αλγόριθμος κάθε στιγμή δε χρησιμοποιεί πληροφορία από το μέλλον για την εξαγωγή χαρακτηριστικών όπως για παράδειγμα το μοντέλο bidirectional LSTM, συνεπώς θα μπορούσε να εφαρμοστεί για on the fly αναγνώριση, δηλαδή μπορεί να εφαρμοστεί παράλληλα με την καταγραφή του νοηματιστή.

Είναι σημαντικό να επισημάνουμε ότι τα παραδείγματα της Ελληνικής Νοηματικής που δείξαμε αποτελούν μία ισχυρή ένδειξη ότι τα υπάρχοντα λεξικά των διαφόρων νοηματικών γλωσσών μπορούν επίσης να αποτελέσουν επαρκή σύνολα εκπαίδευσης αν επισημειωθούν μεθοδικά με ένα σύστημα που χρησιμοποιεί μερικές δεκάδες τυποποιημένες χειρομορφές. Με αυτόν τον τρόπο χωρίς να χρειάζονται εκ νέου μαγνητοσκοπημένες συλλογές, μπορεί κάθε νοηματική γλώσσα να αποκτήσει ένα αξιόπιστο σύστημα επισημείωσης των φωνολογικών χαρακτηριστικών. Αυτό με τη σειρά του μπορεί να αποτελέσει μία βάση για μελλοντικές εργασίες στην κατεύθυνση της ρεαλιστικής σύνθεσης νοημάτισης. Επιπλέον, τα τρία μοντέλα που αναπτύξαμε θα μπορούσαν να αποτελέσουν ένα πολύ καλό pre-trained μοντέλο το οποίο επεκτείνοντάς το με κάποιο αναδρομικό νευρωνικό δίκτυο να λειτουργήσει σαν μοντέλο μετάφρασης συνεχούς νοημάτισης.

Κεφάλαιο 6

Προτάσεις για Μελλοντικές Εφαρμογές και Βελτιώσεις των Datasets

Τα αποτελέσματα της πειραματικής διαδικασίας υποδεικνύουν ότι σε περιπτώσεις όπου υπάρχει διαχωρισμός του λόγου σε γλωσσικές οντότητες μίας Νοηματικής Γλώσσας η μέθοδος που αναπτύχθηκε έχει εξαιρετικά αποτελέσματα σε ότι αφορά την αναγνώριση των στατικών χαρακτηριστικών (σχήμα, προσανατολισμός, θέση). Αυτό θα μπορούσε να φανεί μελλοντικά ιδιαίτερα χρήσιμο σε δύο ερευνητικές κατευθύνσεις.

Πρώτον, είναι ιδιαίτερα διαδεδομένη η πεποίθηση ότι τα μοντέλα που είναι εκ των προτέρων εκπαιδευμένα να αναγνωρίζουν κάτι, δίνουν καλύτερα αποτελέσματα όταν αποτελούν μέρος ενός μοντέλου σε μία παρόμοια εφαρμογή που εκπαιδεύεται εκ νέου, σε σχέση με ένα τυχαία αρχικοποιημένο μοντέλο. Σχετικές εργασίες όπως η [19] χρησιμοποιούν συνελικτικά νευρωνικά δίκτυα (CNN) αρχιτεκτονικής τύπου GoogleNet [29] εκπαιδευμένα σε εφαρμογές αναγνώρισης εικόνας άσχετες με το ζητούμενο όπως είναι το σύνολο ILSVRC2014. Με τον ίδιο τρόπο, θα μπορούσε ένας παράλληλος συνδυασμός από τα 3 μοντέλα να αποτελέσουν τη βάση για ένα μεγαλύτερο αναδρομικό νευρωνικό δίκτυο (RNN) για αναγνώριση σε επίπεδο αναγνώρισης λέξεων εφαρμοσμένο σε ένα ανάλογο σύνολο δεδομένων.

Δεύτερον, η επισημείωση των βίντεο θα μπορούσε να γίνει πολύ απλούστερη με τη χρήση των μοντέλων αυτών στο προσκήνιο. Ο επισημειωτής ούτως ή άλλως μέχρι στιγμής διαιρεί την ακολουθία στιγμιοτύπων σε τμήματα που αντιστοιχούν σε μεμονωμένες γλωσσικές οντότητες. Έχοντας δουλέψει πάνω σε ένα εκτενές λεξικό κάποιας Νοηματικής Γλώσσας, όπως εκείνα που παρουσιάστηκαν στο Κεφάλαιο 3, ο συνδυασμός των 3 μοντέλων όταν εφαρμόζεται επί των τμημάτων αυτών, είναι σε θέση να λειτουργήσει σαν ένα πολύ αξιόπιστο σύστημα συστάσεων (recommendation system). Η σημαντικότητα αυτής της εφαρμογής είναι πολύ σημαντική αν αναλογιστεί κανείς τη διαδικασία για το σωστό τρόπο δημιουργίας ενός ρεαλιστικού συνόλου καταγραφής της Νοηματικής Γλώσσας. Για την μετάφραση μεταξύ δύο ομιλούμενων

γλωσσών αρκεί να γραφεί ένα κείμενο σε κάποια γλώσσα και έπειτα γραφεί το ίδιο κείμενο στη δεύτερη γλώσσα. Ωστόσο, όπως αναλύθηκε στο Κεφάλαιο 2 οι νοηματικές γλώσσες είναι γλωσσολογικά εντελώς διαφορετικά δομημένες σε σχέση με τις προφορικές, καθώς τείνουν να προσανατολίζονται περισσότερο στην οπτική περιγραφή μίας κατάστασης παρά σε μια ευδιάκριτη αλληλουχία λέξεων με ένα αυστηρό συντακτικό και με ένα σαφή διαχωρισμό σε μέρη του λόγου. Η απαίτηση κάποιος να μεταφέρει τη σημασία μίας γραπτής πρότασης στη νοηματική εκβιάζει το νοηματιστή να χρησιμοποιήσει νοήματα που δε θα χρησιμοποιούσε, σε μία διαφορετική σειρά από ότι θα χρησιμοποιούσε υπό φυσιολογικές συνθήκες. Τέτοια σύνολα δεδομένων μπορεί να έχουν ερευνητικό ενδιαφέρον από πλευράς μηχανικής μετάφρασης αλλά στην πραγματικότητα χάνουν την ιδιότητα της εφαρμογής σε πραγματικές συνθήκες. Αντιθέτως, ο ενδεδειγμένος τρόπος δημιουργίας ενός συνόλου ρεαλιστικής καταγραφής μίας νοηματικής γλώσσας είναι η δημιουργία οπτικού περιεχομένου τύπου κόμικ, μίας φωτογραφίας ή ενός μικρού βίντεο που απεικονίζουν την κατάσταση προς περιγραφή. Ο νοηματιστής μαγνητοσκοπείται να αποδίδει το περιεχόμενο του αντίστοιχου οπτικού μέσου αφήγησης και ο επισημειωτής στο τέλος καλείται να μεταφράσει κάθε νόημα με την σχετική λέξη της ομιλούμενης γλώσσας. Συνεπώς, σε αντίθεση με τη μετάφραση ομιλούμενων γλωσσών, ο μεταφραστής δεν δεν είναι σε θέση να γνωρίζει εκ των προτέρων τις λέξεις που θα πρέπει να χρησιμοποιήσει και για αυτό ίσως ένα αξιόπιστο σύστημα συστάσεων είναι ιδιαίτερα βοηθητικό.

Ωστόσο, είναι ελάχιστα τα σύνολα που έχουν συγκεντρωθεί και πληρούν τόσο τις προϋποθέσεις για μία ρεαλιστική αναπαράσταση κάποιας νοηματικής γλώσσας όσο και τη μεγάλη επαναληψημότητα λέξεων (κλάσεων) ώστε να προσφέρονται για τους σκοπούς της μηχανικής μάθησης. Μέσα από την έρευνα που πραγματοποιήθηκε για την παρούσα εργασία δύο σύνολα της βιβλιογραφίας μπορούσαν να λειτουργήσουν ως ένα σύνολο εκπαίδευσης σε επίπεδο λέξεων. Το πρώτο αποτελεί η συλλογή RWTH-PHOENIX-Weather (βλ. Ενότητα 3.3.2) που είναι, ίσως, η διασημότερη συλλογή τέτοιου είδους. Είναι ένα πολύ καλό παράδειγμα συλλογής δεδομένων μεγάλης συνολικής διάρκειας και παράλληλα μεγάλος αριθμός από επαναλήψεις λέξεων. Ο λόγος για αυτό ήταν ότι τα δελτία καιρού μαγνητοσκοπούνταν καθημερινά ενώ το λεξιλόγιο που συναντάται σε αυτά αποτελείται κυρίως από όρους που αφορούν καιρικά φαινόμενα και συνεπώς είναι περιορισμένα και επαναλαμβάνονται καθημερινά στο δελτίο. Το δεύτερο αφορά το σύνολο KETI της Κορεάτικης Νοηματικής Γλώσσας (βλ. Ενότητα 3.4) που δημιουργήθηκε για τις ανάγκες του [14]. Η συγκεκριμένη συλλογή έγινε υπό την αιγίδα ενός «ειδικού» μαγνητοσκοπώντας κωφούς νοηματιστές, γεγονός που δείχνει ότι η διαδικασία ήταν η ενδεδειγμένη. Περιορίζοντας και εδώ την θεματική σε «καταστάσεις εκτάκτου ανάγκης» γίνεται εφικτό να υπάρξει μειωμένο λεξιλόγιο με μεγάλο πλήθος εμφανίσεων ανά λέξη. Κάθε πρόταση μαγνητοσκοπήται από δέκα διαφορετικούς νοηματιστές προσθέτοντας γενικότητα στην επιτυχία αναγνώρισης από τη πλευρά της εφαρμογής. Επιπροσθέτως, μία πολύ καλή στρατηγική που έγινε από πλευράς συλλογής δεδομένων, ήταν η χρήση 2 καμερών σε κάθε λήψη προσφέροντας μία επιπλέον οπτική γωνία και το διπλασιασμό των δεδομένων χωρίς όμως να διπλασιάζεται η απαιτούμενη δουλειά επί της επισημείωσης.

Στο μέλλον, κάποιος που θα θελήσει να δημιουργήσει ένα dataset για τους σκοπούς της μηχανικής μετάφρασης, θα πρέπει να λάβει υπόψιν τα προβλήματα και τις ανάγκες που εμφανίζονται στην συστηματική μετάβαση από μία νοηματική σε κάποια γραπτή γλώσσα. Το μείζον πρόβλημα είναι ότι σε αντίθεση με τον προφορικό λόγο, η χρήση «φωνολογικών» χαρακτηριστικών είναι μία πολλή καλή λύση για τη σύνθεση αλλά όχι για την αναγνώριση. Η αιτία για αυτό είναι ότι ένα άβαταρ που νοηματίζει ακολουθώντας μηχανικά τα 5 βασικά χαρακτηριστικά για κάθε νόημα (σχήμα, προσανατολισμό, θέση, τροχιά, non-manuals) θα γίνεται (μάλλον) κατανοητό παρότι θα υστερεί φυσικότητας, ενώ ένας νοηματιστής που καλείται να νοηματίσει με φυσικό τρόπο, κατά πάσα πιθανότητα θα καταστρατηγήσει τον τυπικό τρόπο εκφοράς ενός νοήματος για χάρη της ταχύτητας και της γλαφυρότητας. Γίνεται, λοιπόν, σαφές ότι κάποιο σύνολο δεδομένων που αποσκοπεί στην εκπαίδευση ενός στοιχειώδους συστήματος μηχανικής μετάφρασης με αντίκρισμα σε πραγματικές εφαρμογές, θα πρέπει να αποσκοπεί στην αναγνώριση σε επίπεδο μεμονωμένων νοημάτων (λέξεων).

Για να επιτευχθεί αυτό, θα πρέπει να δημιουργηθεί ένα σύνολο από απλές προτάσεις που θα εμφανίζουν τη μεγαλύτερη δυνατή επαναληψιμότητα από μεμονωμένα νοήματα. Ιδανικά, ένας τρόπος για να επιτευχθεί αυτό είναι να επιλεγεί μία τράπεζα από μερικές δεκάδες λέξεις/νοήματα και συνδυάζοντας αυτές να δημιουργηθούν μερικές εκατοντάδες προτάσεις. Στην πράξη, ίσως, η συγκεκριμένη μέθοδος να μην είναι απόλυτα πραγματοποιήσιμη. Εντούτοις, η ύπαρξη λέξεων με πολύ μικρό πλήθος εμφανίσεων εντός του dataset δεν συνεπάγεται ένα μη χρηστικό σύνολο δεδομένων, απλώς τέτοια στοιχεία του συνόλου δεδομένων αναμένεται να μην αναγνωρίζονται με μεγάλη επιτυχία, όπως άλλωστε συμβαίνει σε πολύ μεγάλο αριθμό από εφαρμογές της μηχανικής μάθησης.

Μία πολύ καλή στρατηγική δημιουργίας ενός καλού dataset θα ήταν η επιλογή μίας συγκεκριμένης θεματικής επηρεασμένης από καταστάσεις της πραγματικής ζωής. Το σύνολο ΚΕΤΙ έχει επιλέξει την επικοινωνία σε έκτακτες περιπτώσει όπως πυρκαγιά. Παρομοίως, ο σχεδιασμός αντίστοιχων συλλογών δεδομένων εμπνευσμένες από καταστάσεις της καθημερινότητας όπου είναι μεγάλης σημασίας η σωστή και απρόσκοπτη επικοινωνία με ένα άτομο χωρίς τη δυνατότητα ακοής, οδηγεί αβίαστα σε μία υλοποίηση με τα χαρακτηριστικά που περιγράψαμε και με ένα σημαντικό κοινωνικό αντίκρισμα. Για παράδειγμα ας θεωρήσουμε ότι γίνεται μία συλλογή προτάσεων με θέμα «επίσκεψη στον γιατρό» που περιλαμβάνει μία σειρά από σύντομες προτάσεις που την επίσκεψη σε ένα παθολόγο. Χωρίς ιδιαίτερη προσπάθεια ο σχεδιαστής των παραδειγμάτων μπορεί ακόμα και γραπτά να περιγράψει το νοηματικό πλαίσιο της κάθε πρότασης χωρίς να περιορίσει εκφραστικά το νοηματιστή. Είναι αρκετό να δημιουργήσει προτάσεις της μορφής «Εξήγησε ότι πονάς/χτύπησες στο ̔μέρος του σώματος̓» ή «Εξήγησε ότι έχεις ̔λίστα από συμπτώματἀ». Με αυτό τον τρόπο αυτόματα δημιουργείται ένα σύνολο που αυτόματα επαναλαμβάνει τους όρους που είναι σημαντικότερο να αναγνωριστούν σωστά στο τελικό προϊόν. Επιπροσθέτως, η ύπαρξη πολλαπλών νοηματιστών βοηθά στην αύξηση των δεδομένων και στην αξιοπιστία της τελικής εφαρμογής λόγω καθολικότητας. Ομοίως, είναι πολύ θετικό στοιχείο η χρήση περισσότερων από μία καμέρες ανά μαγνητοσκόπηση, αν είναι εφικτό, ή απλούστερα μικρές διαφοροποιήσεις

στη γωνία λήψης σε κάθε γύρισμα. Δευτερεύοντα χαρακτηριστικά που θα μπορούσε να λάβει υπόψιν ο συντονιστής μίας τέτοιας προσπάθειας κατά την κινηματογράφηση, είναι να γεμίζει το κάδρο κάνοντας κοντινά πλάνα τα οποία δείχνουν αισθητικά λιγότερο ωραία, εκμεταλλεύονται, όμως, στο μέγιστο την πληροφορία σε pixel που μπορεί να αποδώσει η κάθε κάμερα.

Bibliography

[1] Eduardo Altmann and Martin Gerlach. Statistical laws in linguistics. *arXiv: 1502.03296*, 02 2015.

[2] Britta Bauer and Hermann Hienz. Relevant features for video-based continuous sign language recognition. 2000.

[3] Patrick Buehler, Mark Everingham, and Andrew Zisserman. Learning sign language by watching TV (using weakly aligned subtitles). 2009.

[4] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

[5] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017.

[6] Ananya Choudhury, Anjan Kumar Talukdar, Manas Kamal Bhuyan, and Kandarpa Kumar Sarma. Movement Epenthesis Detection for Continuous Sign Language Recognition. *Journal of Intelligent Systems*, 26(3):471–481, 2017.

[7] Eleni Efthimiou and Stavroula-Evita Fotinea. GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI. *Universal Acess in Human Computer Interaction. Coping with Diversity*, pages 657–666, 2007.

[8] Thomas Hanke. HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts.

[9] J. Huang, W. Zhou, H. Li, and W. Li. Attention-based 3d-cnns for large-vocabulary sign language recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(9):2822–2832, Sep. 2019.

[10] Jie Huang, Wengang Zhou, Houqiang Li, and Weiping Li. Sign language recognition using 3d convolutional neural networks. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, June 2015.

[11] Yun Jin, Yan Zhao, Chengwei Huang, and Li Yong Zhao. Study on the emotion recognition of whispered speech. *2009 WRI Global Congress on Intelligent Systems*, 3:242–246, 2009.

[12] K. Karpouzis, G. Caridakis, S. E. Fotinea, and E. Efthimiou. Educational resources and implementation of a Greek sign language synthesis architecture. 2007.

[13] Sang-Ki Ko, Chang Kim, Hyedong Jung, and Choongsang Cho. Neural sign language translation based on human keypoint estimation. *Applied Sciences*, 9:2683, 07 2019.

[14] Sang-Ki Ko, Jae Gi Son, and Hyedong Jung. Sign language recognition with recurrent neural network using human keypoint detection. pages 326–328, 2018.

[15] Oscar Koller, Jens Forster, and Hermann Ney. Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141:108–125, 2015.

[16] Oscar Koller, Hermann Ney, and Richard Bowden. Automatic Alignment of HamNoSys Subunits for Continuous Sign Language Recognition. 2016.

[17] Oscar Koller, Hermann Ney, and Richard Bowden. Deep Hand: How to Train a CNN on 1 Million Hand Images When Your Data is Continuous and Weakly Labelled. 2016.

[18] Oscar Koller, Sepehr Zargaran, and Hermann Ney. Re-Sign : Re-Aligned End-to-End Sequence Modelling with Deep Recurrent CNN-HMMs. 2019.

[19] Oscar Koller, Sepehr Zargaran, Hermann Ney, and Richard Bowden. Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. 2017.

[20] Oscar Koller, Sepehr Zargaran, Hermann Ney, and Richard Bowden. Deep Sign: Enabling Robust Statistical Continuous Sign Language Recognition via Hybrid CNN-HMMs. 2018.

[21] Cesare Magarotto. Towards an international language of gestures. *Unesco Courier*, 1974.

[22] Nobuyasu Nakano, Tetsuro Sakura, Kazuhiro Ueda, and Leon Omura. Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras.

[23] Vassilis Pitsikalis, Stavros Theodorakis, Christian Vogler, and Petros Maragos. Advances in phonetics-based sub-unit modeling for transcription alignment and sign language recognition. 2011.

[24] Anastasios Roussos, Stavros Theodorakis, Vassilis Pitsikalis, and Petros Maragos. Dynamic Affine-Invariant Shape-Appearance Handshape Features and Classification in Sign Language Videos. 14:231–271, 2017.

[25] Haifeng Sang and Hongjiao Wu. A sign language recognition system in complex background. pages 453–461, 10 2016.

[26] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*, 2017.

[27] Thad Starner, Joshua Weaver, and Alex Pentland. Real-time american sign language recognition using desk and wearable computer based video. 1998.

[28] William C Stokoe. Sign language structure (studies in linguistics. *Occasional paper*, 8, 1960.

[29] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[30] Stavros Theodorakis, Vassilis Pitsikalis, and Petros Maragos. Dynamic-static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition. *Image and Vision Computing*, 32(8):533–549, 2014.

[31] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *CVPR*, 2016.

[32] Mahmoud M. Zaki and Samir I. Shaheen. Sign language recognition using a combination of new vision based features. 2011.

[33] George Kingsley Zipf. On the number, circulation-sizes, and the probable purchasers of newspapers. *The American Journal of Psychology*, 61(1):79–89, 1948.

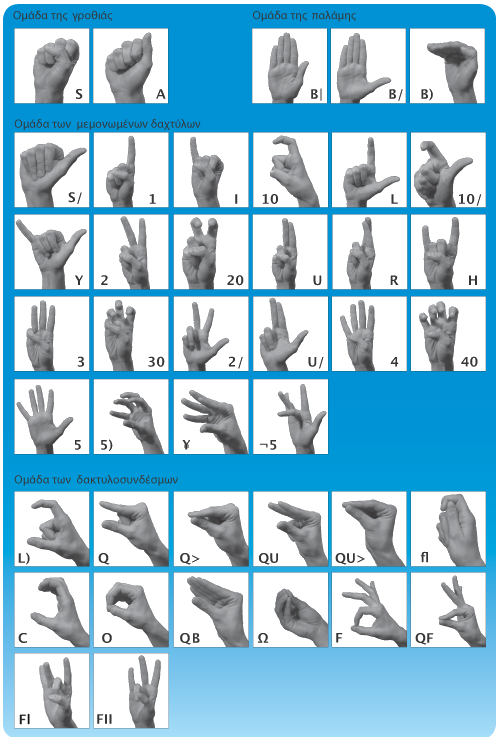
Παράρτημα Α′

HamNoSys Handshapes

Thomas Hanke, 2010-06-10, Drawings by Heiko Ziener, Olga Jerjorski, Andreas Hanß

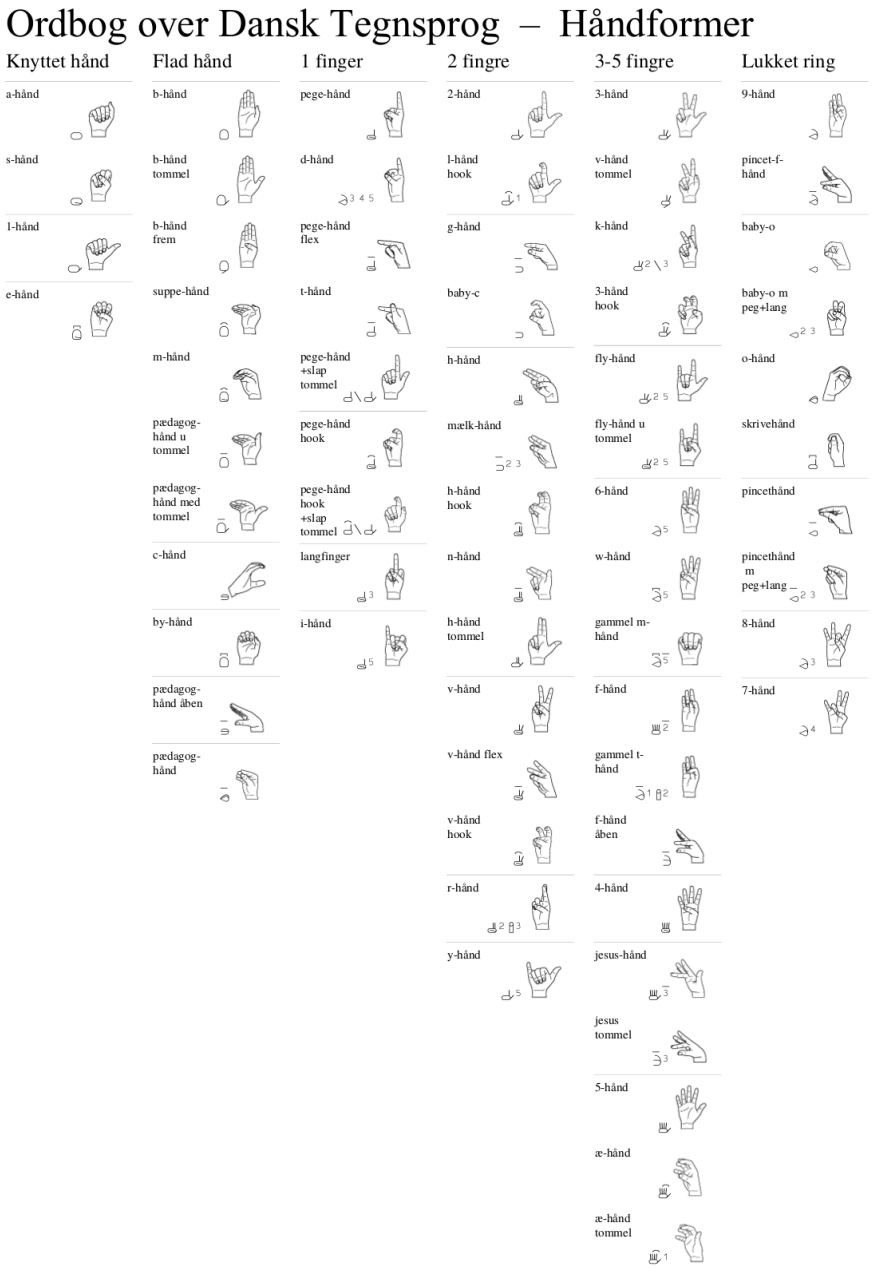
Παράρτημα Β΄

Χειρομορφές και Δακτυλικά Αλφάβητα διαφορετικών Νοηματικών Γλωσσών

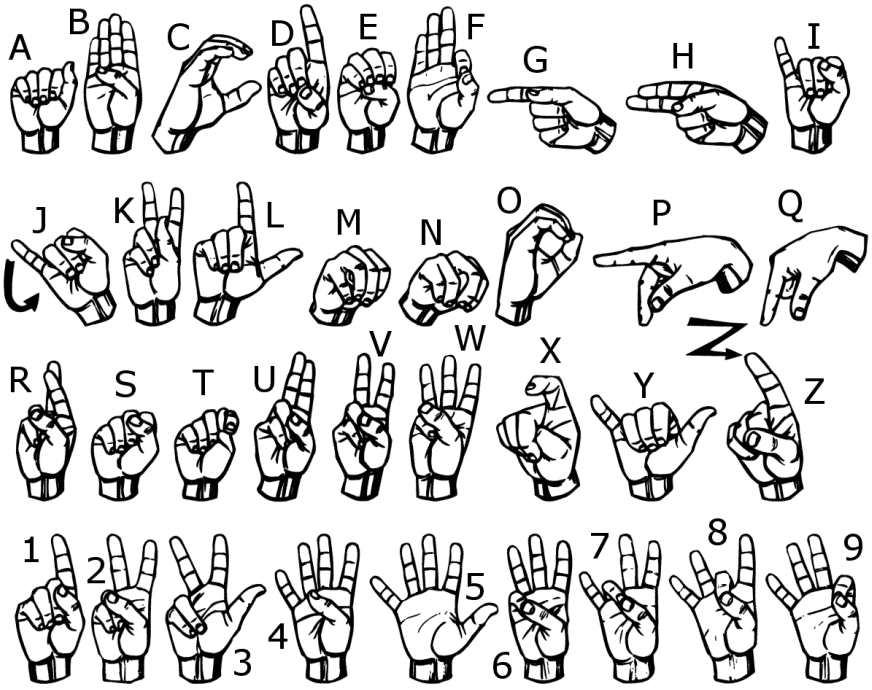


(α΄) Χειρομορφές Ελληνικής/Κυπριακής Νοηματικής Γλώσσας

(β΄) Ελληνικό Δακτυλικό Αλφάβητο



Σχήμα Β΄.2: Χειρομορφές της Δανικής Νοηματικής Γλώσσας



Σχήμα Β΄.3: Δακτυλικό Αλφάβητο Αμερικάνικής Νοηματικής Γλώσσας

Σχήμα Β΄.4: Δακτυλικό Αλφάβητο Βρετανικής Νοηματικής Γλώσσας