

# Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακό Πρόγραμμα Σπουδών  
*Πληροφοριακά και Επικοινωνιακά Συστήματα*

## Μεταπτυχιακή Διατριβή



Εργαλεία ανίχνευσης γλωσσικών προκαταλήψεων  
σε υπερκείμενα τα οποία συν-δημιουργούνται με  
τεχνολογία Wiki

Ιωάννα Λουκοπούλου

Επιβλέπων Καθηγητής  
Σπυρίδων Παπαδάκης

Δεκέμβριος 2016

# **Ανοικτό Πανεπιστήμιο Κύπρου**

**Σχολή Θετικών και Εφαρμοσμένων Επιστημών**

**Μεταπτυχιακό Πρόγραμμα Σπουδών**

***Πληροφοριακά και Επικοινωνιακά Συστήματα***

## **Μεταπτυχιακή Διατριβή**

**Εργαλεία ανίχνευσης γλωσσικών προκαταλήψεων  
σε υπερκείμενα τα οποία συν-δημιουργούνται με  
τεχνολογία Wiki**

**Ιωάννα Λουκοπούλου**

**Επιβλέπων Καθηγητής  
Σπυρίδων Παπαδάκης**

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων για απόκτηση μεταπτυχιακού τίτλου σπουδών στα Πληροφοριακά Συστήματα από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών του Ανοικτού Πανεπιστημίου Κύπρου.

**Δεκέμβριος 2016**

ΛΕΥΚΗ ΣΕΛΙΔΑ

## Περίληψη

Η ερευνητική δραστηριότητα σε θέματα γλώσσας και φύλου εστιάζει στο ερώτημα αν σε μια ανδροκρατούμενη κοινωνία η γλώσσα και η λεγόμενη από ορισμένους ερευνητές «γλωσσική επίδραση» εκφράζει μια τέτοια κατάσταση. Η εργασία αυτή εστιάζεται στην ανάπτυξη υπολογιστικού εργαλείου εντοπισμού του βαθμού εμφάνισης του φαινομένου στερεοτυπικών αντιλήψεων - γλωσσικής προκατάληψης στην Διαδικτυακή εγκυκλοπαίδεια Wikipedia στην περίπτωση του κοινωνικού φύλου (gender).

Ζητούμενο ήταν η διερεύνηση της ύπαρξης είτε ο σχεδιασμός, η ανάπτυξη και αξιολόγηση κατάλληλων υπολογιστικών εργαλείων τα οποία μπορούν να συμβάλλουν στη σύγκριση ψηφιακών κειμένων, υπερκειμένων και υπερμέσων. Για το σκοπό αυτό εξετάστηκε πιλοτικά η χρήση του αναπτυχθέντος εργαλείου για την περίπτωση βιογραφιών προσωπικοτήτων στη Wikipedia για την ανίχνευση διαφορών ως προς το θέμα της γλωσσικής προκατάληψης.

Από τα αποτελέσματα διαπιστώθηκε ότι τα προτεινόμενα εργαλεία μπορούν να χρησιμοποιηθούν στη διάγνωση του βαθμού της γλωσσικής προκατάληψης σε σχέση με το φύλο σε υπερκείμενα - υπερμέσα τα οποία συγγράφονται διαμορφώνονται σε ψηφιακά περιβάλλοντα όπως αυτά τύπου wiki και να συμβάλλουν στην αντιμετώπιση του προβλήματος αυτού.

Ως επέκταση της διατριβής προτείνεται η δημιουργία εργαλείου που θα κάνει προτάσεις για τη διόρθωση των σημείων που εντοπίζονται προβλήματα τέτοιου τύπου.

## Summary

The research was executed on the subject of “language and gender”, specifically our research is to explore if in a male dominated society the “language effect” is what articulates this domination. Our dissertation is specifically focused on the development of a software that would be able to indicate the correlation of variables, which are related to our subject. Overall, our desire is to develop and improve existing software tools that would accurately observe, analyze and correct the degree of stereotypical attitudes that result from language prejudice.

Initially our study began with the utilization of existing software tools to examine current capabilities of analyzing and comparing digital documents and/or multimedia. Specifically our research focused primarily (as a pilot experiment) on the comparison and examination of various biographies, both male and female, to expose the effects of language prejudice. The biographies were used from the prominent internet site “Wikipedia”.

From the results of our study, it was revealed that the existing software tools available were actually able to accurately compare and analyze the degree of the “language effect”. Overall, digital documents and multimedia were easily scrutinized and we could effortlessly identify the specific language used to influence stereotypical attitudes toward gender.

Therefore, the main purpose of our dissertation will focus on the development of a software, which will not only identify the language prejudice used but will also make corrections, or at least suggested replacements in order to avoid gender injustice.

## Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή, κ. Σπύρο Παπαδάκη, για την επιστημονική καθοδήγηση και επίβλεψη καθώς και τη ψυχολογική στήριξη κατά την εκπόνηση της διπλωματικής μου διατριβής.

Επίσης θα ήθελα να ευχαριστήσω την κ. Παλάζη Χρυσάνθη για τη συμβολή της στην επεξεργασία του θέματος μέσα από την κοινωνική και ψυχολογική διάσταση του.

Επίσης θα ήθελα να ευχαριστήσω την επιβλέποντα καθηγήτρια, κ. Jana Otterbacher η οποία μου ανέθεσε το συγκεκριμένο θέμα και επιλέξαμε μαζί τις κατηγορίες κειμένων που θα επεξεργαζόμασταν. Η βοήθειά της ήταν σημαντική για το προγραμματιστικό περιβάλλον της διατριβής με τη βιβλιογραφία που μου πρότεινε, τις εξειδικευμένες γνώσεις της και την ερευνητική προσφορά της σε αυτό τον τομέα της επιστήμης μας.

Ευχαριστώ τους συμφοιτητές μου και τους καθηγητές του τμήματος Π.Ε.Σ για την πρόκληση επικαιροποίησης και συμπλήρωσης γνώσεων που μοιραστήκαμε σε όλη την πορεία του προγράμματος μεταπτυχιακής εκπαίδευσης.

# Περιεχόμενα

<b>Εισαγωγή.....</b>	<b>1</b>
1.1 Οριοθέτηση του προβλήματος.....	2
1.2 Λόγοι επιλογής του θέματος.....	3
1.3 Σκοποί και Στόχοι .....	3
1.4 Μεθοδολογία έρευνας.....	3
1.5 Συμβολή της έρευνας στην κεκτημένη γνώση.....	4
1.6 Περιορισμοί της έρευνας.....	4
1.7 Δομή της Μεταπτυχιακής Διατριβής.....	5
<b>Κεφάλαιο 2 .....</b>	<b>6</b>
<b>Βιβλιογραφική Ανασκόπηση.....</b>	<b>6</b>
2.1 Η γλώσσα ως μέσο επικοινωνίας .....	6
2.3 Η στερεοτυπική συμπεριφορά.....	8
2.4 Το φαινόμενο της γλωσσικής προκατάληψης .....	9
2.5 Παράγοντες γλωσσικής διαφοροποίησης ανάμεσα στα δύο φύλα πριν την ενηλικίωση.....	11
2.6 Ο γλωσσικός σεξισμός στα δύο φύλα.....	12
2.7 Ο κοινωνικός ρόλος στη διαμόρφωση της γλωσσικής συμπεριφοράς .....	13
2.8 Η έμφυλη γλωσσική προκατάληψη στα Μ.Μ.Ε.....	13
2.9 Γλωσσική προκατάληψη στο Διαδίκτυο και στα Μέσα κοινωνικής δικτύωσης....	14
2.10 Η τάξη του κόσμου και οι αλλαγές στο γλωσσικό επίπεδο .....	18
2.11 Το φύλο της γλώσσας.....	19
2.12 Η γλωσσική προκατάληψη στην ελληνική γλώσσα .....	19
2.12.1 Γλωσσική προκατάληψη στη σύνταξη της ελληνικής γλώσσας .....	21
2.12.2 Σημασιολογικά στερεότυπα.....	21
2.12.3 Προτάσεις για την αποφυγή του γλωσσικού σεξισμού .....	22
2.13 Το συλλογικό εγκυκλοπαιδικό εγχείρημα Βικιπαίδεια .....	23
2.14 Αντικειμενικότητα και ποιότητα πληροφοριών στη Wikipedia .....	25
2.15 Μελέτες έμφυλης γλωσσικής προκατάληψης .....	27
2.16 Προηγούμενες έρευνες.....	28
Harvard IV-4 categories:.....	33
<b>Κεφάλαιο 3 .....</b>	<b>37</b>
<b>Μέθοδος.....</b>	<b>37</b>
<b>3.1 Η περιγραφή ερευνητικού προβλήματος.....</b>	<b>37</b>

3.2 Αιτιολόγηση της συγκεκριμένης επιλογής (πλεονεκτήματα και μειονεκτήματα)..	38
3.3 Επιλογή δείγματος.....	38
3.4 Ερευνητικό Υλικό – Εργαλεία .....	39
<b>Κεφάλαιο 4 .....</b>	<b>42</b>
4.1 Εξαγωγή πληροφορίας .....	42
4.2 Δέντρα απόφασης/ταξινόμησης.....	43
4.3 Ερευνητική Διαδικασία.....	44
4.4 Βήματα Αλγόριθμου.....	45
4.4.1 Η λειτουργία Compare .....	47
4.5 Ποσοτικά αποτελέσματα .....	48
4.5.1 Βιογραφίες ηθοποιών .....	49
4.5.2 Βιογραφίες πολιτικών.....	53
<b>Κεφάλαιο 5 .....</b>	<b>57</b>
<b>Σχολιασμός αποτελεσμάτων .....</b>	<b>57</b>
5.1 Αναφορικά ως προς τα δύο φύλα.....	57
5.2 Ως προς την έκταση των βιογραφιών.....	58
5.3 Σύνοψη των ευρημάτων και ερμηνεία τους.....	59
5.4 Συζήτηση, προτάσεις για περαιτέρω έρευνα.....	60
<b>Παράρτημα Α.....</b>	<b>62</b>
<b>Διοικητικά έγγραφα .....</b>	<b>62</b>
A.1 Άρθρο και γλωσσική προκατάληψη.....	62
<b>Παράρτημα Β.....</b>	<b>63</b>
<b>Code semantics.....</b>	<b>63</b>
B.1 Functions.....	63
B.2 Source code for nltk.ccg.logic Functions.....	64
<b>Παράρτημα Γ .....</b>	<b>65</b>
<b>Sentiment Greek Lexicon.....</b>	<b>65</b>
Γ.1 Περιγραφή .....	65
<b>Παράρτημα Δ .....</b>	<b>70</b>
<b>Περιγραφή Εργαλείων .....</b>	<b>70</b>
Δ.1 Notepad++ .....	70
Δ.2 The Unique Word List Text Wizard .....	71
Δ.3 Δείγμα αγγλικού λεξικού AFINN.....	71
<b>Βιβλιογραφία.....</b>	<b>84</b>



## Εικόνες

Εικόνα 1: Γράφημα με τα φαινόμενα σεξισμού ανά φορέα (Γενική Γραμματεία Ισότητας) .....	22
Εικόνα 2: Stanford's NLP tools –the name entity recognition .....	30
Εικόνα 3: A parse tree of sentiment analysis. Assign a positive/negative sentiment to each part. .....	31
Εικόνα 4: Sentiment tree, in <a href="http://nlp.stanford.edu/sentiment/treebank.html">http://nlp.stanford.edu/sentiment/treebank.html</a> .....	31
Εικόνα 5: Sentiment tree, in <a href="http://nlp.stanford.edu/sentiment/treebank.html">http://nlp.stanford.edu/sentiment/treebank.html</a> .....	32
Εικόνα 6: Harvard IV-4 categories .....	33
Εικόνα 7: Package <code>algorithmia.nlp.SentimentAnalysis</code> .....	36
Εικόνα 8: Sentiment analysis process on product reviews .....	42
Εικόνα 9: Κλάσεις ταξινόμησης δείγματος.....	43
Εικόνα 10: Sentiment classification techniques.....	44
Εικόνα 11: Αποτέλεσμα του ανιχνευτή λέξεων γλωσσικής προκατάληψης.....	47
Εικόνα 12: Αποτέλεσμα του ανιχνευτή λέξεων γλωσσικής προκατάληψης .....	47

# Κεφάλαιο 1

## Εισαγωγή

Η διατριβή αυτή ερευνά την ανίχνευση γλωσσικών προκαταλήψεων σε περιβάλλοντα ψηφιακών συνεργατικών κειμένων τύπου wiki όπως η Διαδικτυακή εγκυκλοπαίδεια Wikipedia και ιδιαίτερα αναφορικά με την περίπτωση του κοινωνικού φύλου (gender). Για το σκοπό αυτό σχεδιάστηκε και αναπτύχθηκε ένα υπολογιστικό εργαλείο εντοπισμού του βαθμού εμφάνισης του φαινομένου στερεοτυπικών αντιλήψεων-γλωσσικής προκατάληψης. Σύμφωνα με τον προτεινόμενο αλγόριθμο υπολογίζονται λέξεις που ενέχουν γλωσσικά έμφυλα στερεότυπα με την αξιοποίηση και αξιολόγηση κατάλληλων υπολογιστικών εργαλείων τα οποία μπορούν να συμβάλλουν στη σύγκριση ψηφιακών κειμένων, υπερκειμένων και υπερμέσων.

Για την αξιολόγηση των προτεινόμενων εργαλείων αξιολογούνται βιογραφίες προσωπικοτήτων για την ανίχνευση διαφορών ως προς το θέμα της γλωσσικής προκατάληψης στο περιβάλλον της Wikipedia. Ειδικότερα στην παρούσα μελέτη προτείνουμε έναν αλγόριθμο τον οποίο χρησιμοποιούμε στη διάγνωση του βαθμού της γλωσσικής προκατάληψης σε σχέση με το φύλο σε υπερκείμενα - υπερμέσα τα οποία συγγράφονται διαμορφώνονται σε ψηφιακά περιβάλλοντα όπως αυτά τύπου wiki. Παρατηρώντας την παραγωγή λόγου σε οποιαδήποτε γλώσσα στο Διαδίκτυο με οποιαδήποτε μορφή λαμβάνει αυτή μπορούμε να ανιχνεύσουμε χρήση στερεοτυπικού λόγου. Άλλωστε η γλώσσα, ως το κύριο μέσο παραγωγής και άρθρωσης λόγου, θεωρείται ακριβώς και ως το κύριο μέσο κοινοποίησης των στερεοτύπων και μετάδοσής τους από γενιά σε γενιά εκ μέρους του Τύπου και των ΜΜΕ (Otterbacher, 2015). Με αφετηρία αυτή τη πραγματικότητα η διατριβή αυτή έρχεται να καλύψει το κενό που υπάρχει στην ανίχνευση γλωσσικής προκατάληψης με έμφαση στον εντοπισμό διαφορών που εκφράζεται στα δύο φύλα.

Ειδικότερα σχεδιάστηκε ένα εργαλείο το οποίο καταγράφει συστηματικά και μονοσήμαντα σε κείμενα στο Διαδίκτυο και ειδικά στην διαδικτυακή εγκυκλοπαίδεια Wikipedia λέξεις που ενέχουν έμφυλη γλωσσική προκατάληψη. Η «προκατάληψη» (bias) είναι μια κατ' ουσία εξωλογική τοποθέτηση που φτάνει τα όρια της ισχυρά θεμελιωμένης και θεμελιωμένης πεποίθησης. Συγκεκριμένα τα έμφυλα στερεότυπα οφείλονται σε δομημένες πεποιθήσεις κοινωνιολογικής φύσης και αναφοράς, που μπορεί να είναι, είτε αναχρονιστικές είτε δεν υποβάλλονται σε διεργασίες που έχουν να κάνουν με τον ορθό λόγο αλλά προέρχονται από άκριτες ιδέες και προσδοκίες, επομένως συχνά δεν ανταποκρίνονται στην πραγματικότητα. Έτσι δημιουργείται το κοινωνικό φύλο (gender) το οποίο συνδέεται με το ιστορικό συγκείμενο και αυτό με τη σειρά του εκφράζεται με μια ορισμένη, στερεοτυπική γλώσσα η οποία συχνά εμποδώνει προκαταλήψεις, αναπαράγοντάς τις μέσα στην πορεία του χρόνου.

Η γλωσσική προκατάληψη, στην παρούσα έρευνα, ορίζεται ως μια συστηματική διαφορά στη γλώσσα την οποία χρησιμοποιούμε όταν περιγράφουμε άτομα από διάφορες κοινωνικές ομάδες. (Maass, 1999). Η αναπαραγωγή αυτής με ψηφιακό τρόπο αποτελεί πεδίο έρευνας το οποίο έχει πολλαπλούς λόγους μείζονος σημασίας. Ξεκινάμε επομένως την έρευνά μας βασιζόμενοι στη θέση ότι ο άνθρωπος σε οποιαδήποτε τάξη και αν βρίσκεται επηρεάζεται από σειρά προκαταλήψεων σε κάθε άποψη της γλώσσας, που αναφέρονται στην αυθαίρετη προτίμηση και υποτίμηση, συμβόλων, λέξεων και συνδυασμών τους και οι οποίες σαν αποτέλεσμα έχουν να δημιουργούν σύγχυση στην επικοινωνία και άρα υποβάθμισή της. Η ερευνητική δραστηριότητα σε θέματα γλώσσας και φύλου εστιάζεται στο ερώτημα αν σε μια ανδροκρατούμενη κοινωνία η γλώσσα και η λεγόμενη από ορισμένους ερευνητές «γλωσσική διεπίδραση» εκφράζει κατάσταση γλωσσικής προκατάληψης. Σύμφωνα με αυτές τις απόψεις, για την απάλειψη της ρητής χρήσης του θηλυκού γένους, που γεννά τον προβληματισμό και σηματοδοτεί μια πραγματική και όχι μόνο γλωσσική περιθωριοποίηση (Παυλίδου Θ.-Σ., 2006 ).

([http://www.greek-language.gr/greekLang/modern\\_greek/tools/lexica/glossology/show.html?id=141](http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/glossology/show.html?id=141)).

Θα λέγαμε ότι ένας κοινός τόπος της διεξαγόμενης έρευνας πάνω σε αυτές τις θεματολογίες – και βέβαια κατεξοχήν στις φεμινιστικής αφετηρίας έρευνες-είναι η διασαφήνιση του ρόλου που έχει η γλώσσα στην ύπαρξη της προαναφερόμενης ανισότητας, ανισοτιμίας και ενδεχομένως ευρύτερης κοινωνικής περιθωριοποίησης ανάμεσα στα δύο φύλα και το εγχείρημα της αποκατάστασης αυτών των αρνητικών καταστάσεων διαμέσου της κριτικής επιστημονικής παρέμβασης στα αρνητικά αυτά γλωσσικά φαινόμενα. Όμως η εύπλαστη και ιδιαίτερα ρευστή φύση της γλώσσας είναι αυτή που καθορίζει την πολυπλοκότητα του τρόπου επικοινωνίας και εντέλει την πολιτική ορθότητα που αυτή θα καταφέρει να επιβάλει μέσα από τους διάφορους τρόπους έκφρασής της.

## 1.1 Οριοθέτηση του προβλήματος

Η γλωσσική προκατάληψη, στην παρούσα έρευνα, εξετάζεται ως μια συστηματική διαφορά στη γλώσσα, την οποία χρησιμοποιούμε, όταν περιγράφουμε άτομα από διάφορες κοινωνικές ομάδες. Η ανίχνευση γλωσσικής προκατάληψης σε οποιαδήποτε γλώσσα στο Διαδίκτυο με οποιαδήποτε μορφή λαμβάνει αυτή αποτελεί πεδίο έρευνας και παραγωγής αποτελεσμάτων ικανών έτσι ώστε να αναδειχθούν οι αιτίες που γεννούν και συντηρούν αυτές τις αντιλήψεις, έτσι ώστε τελικά να ελεγχθούν και να αντιμετωπιστούν. Η δημιουργία ενός αλγόριθμου που θα υπολογίζει τον αριθμό των λέξεων που ενέχουν γλωσσικής προκατάληψης αποτέλεσε το αντικείμενο του πεδίου έρευνάς μας.

## 1.2 Λόγοι επιλογής του θέματος

Η επιλογή του θέματος αποφασίστηκε κατόπιν ανάγκης ενασχόλησης με το συγκεκριμένο θέμα. Η καταγραφή και η προσπάθεια ερμηνείας λέξεων που ενέχουν έμφυλη γλωσσική προκατάληψη σε κείμενα στο Διαδίκτυο και ειδικά στην Διαδικτυακή εγκυκλοπαίδεια, ερευνητικά αποτέλεσε μια πρόκληση συγκερασμού του γνωστικού αντικειμένου σπουδών μου της Πληροφορικής με αυτές άλλων επιστημών όπως της Κοινωνιολογίας και της Ψυχολογίας. Άλλωστε η γλώσσα θεωρείται το κύριο μέσο κοινοποίησης των στερεοτύπων και μετάδοσής τους και είναι κοινή πεποίθηση ότι ο άνθρωπος σε οποιαδήποτε τάξη ή κατηγορία βρίσκεται, παρουσιάζει σειρά προκαταλήψεων οι οποίες αναπαράγονται με το λόγο. Με αφετηρία αυτή τη πραγματικότητα η διατριβή αυτή έρχεται να καλύψει το κενό που υπάρχει στην ανίχνευση γλωσσικής προκατάληψης με έμφαση στον εντοπισμό διαφορών που εκφράζεται στα δύο φύλα.

## 1.3 Σκοποί και Στόχοι

Ο σκοπός της διπλωματικής διατριβής αυτής είναι ο σχεδιασμός ενός αλγόριθμου υπολογιστικού εργαλείου ανίχνευσης του βαθμού ουδετερότητας ενός κειμένου σε σχέση με τις στερεοτυπικές αντιλήψεις ως προς το φύλο σε περιβάλλοντα συνεργατικής δημιουργίας υπερκειμένων - υπερμέσων όπως τα Wiki.

### Ερευνητικά ερωτήματα

- i. Πως με βάση τις θεωρίες της κοινωνικής ανθρωπολογίας και κοινωνικής ψυχολογίας, μπορούμε να αναπτύξουμε έναν αλγόριθμο για την ανίχνευση γλωσσικής προκατάληψης ως προς το φύλο;
- ii. Σε τι βαθμό η υλοποίηση του αλγορίθμου οδηγεί στην ανίχνευση διαφορών μεταξύ ορισμένων κοινωνικών ομάδων και προσώπων τη χρήση αφηρημένης και υποκειμενικής γλώσσας στο ψηφιακό περιεχόμενο σε περιβάλλοντα τύπου Wiki όπως η Wikipedia;
- iii. Ποια είναι τα αποτελέσματα της υλοποίησης του αλγορίθμου σε μία σειρά από βιογραφίες αναγνωρισμένων προσώπων και πως μπορεί να βελτιωθεί ώστε να είναι πιο ακριβής;

## 1.4 Μεθοδολογία έρευνας

Στη μεθοδολογία που ακολουθήσαμε στη μεταπτυχιακή διατριβή χρησιμοποιήσαμε συνδυασμό ποσοτικής και ποιοτικής προσέγγισης.

- i. Αξιοποιήσαμε ειδικά Λεξικά, τα οποία διαθέτουν λέξεις που ενέχουν κατά την τρέχουσα χρήση τους προκατάληψη. Οι λέξεις αυτές έχουν θετική ή αρνητική φόρτιση, ενώ τα λεξικά χρησιμοποιήσαμε εδώ πρέπει να καλύπτουν όλες τις υπό μελέτη γλώσσες, συγκεκριμένα επί του προκειμένου τα ελληνικά και τα αγγλικά.
- ii. Διερευνήσαμε, την επιλογή και πιθανώς επέκταση υπάρχοντος λογισμικού που θεωρήθηκε κατάλληλο για εντοπισμό πανομοιότυπων λέξεων σε υπερκείμενα στην Βικιπαίδεια.
- iii. Διερευνήσαμε, την επιλογή και πιθανώς επέκταση κατάλληλου λογισμικού με το οποίο αυτόματα συνέκρινε δύο (αλφαβητικά ταξινομημένες) λίστες λέξεων και εντόπισε όλες τις κοινές λέξεις που υπάρχουν μεταξύ τους.
- iv. Έγινε τελικός έλεγχος και καταμέτρηση.

Η επιβεβαίωση και επαληθευσιμότητα του αλγορίθμου των αποτελεσμάτων του και των συμπερασμάτων που αυτός δίνει μπορεί να γίνει δειγματοληπτικά σε ένα από τα κείμενα που επιλέξαμε και το οποίο έχουμε επεξεργαστεί κατά τη διάρκεια της έρευνας έτσι ώστε να διακρίνει την αλήθεια των αποτελεσμάτων και την ορθή λειτουργία του αλγορίθμου.

## 1.5 Συμβολή της έρευνας στην κεκτημένη γνώση

Υπάρχει αναγκαιότητα και σπουδαιότητα στην παρούσα έρευνα και ταυτόχρονα καινοτομία γιατί προσπαθήσαμε να δημιουργήσουμε έναν αλγόριθμο ο οποίος να μελετά, να υπολογίζει τις γλωσσικές προκαταλήψεις οι οποίες συμβάλλουν σε κοινωνικά στερεότυπα. Με βάση τα αποτελέσματα τα οποία συγκεντρώνονται και αφορούν το πλήθος των λέξεων οι οποίες ενέχουν γλωσσικής προκατάληψης, θα μπορούν να χρησιμοποιηθούν από ερευνητές άλλων επιστημών για να παρακολουθήσουν και να παρέμβουν σε ένα πιθανό έλεγχο και εντέλει άρση και αλλαγή του τρόπου με τον οποίο συγγράφονται κείμενα σε οποιοδήποτε μέσο και συγκεκριμένα αφορούν πρόσωπα και των δύο φύλων.

## 1.6 Περιορισμοί της έρευνας

Η αναγκαιότητα της παρούσας έρευνας περιορίζεται από την έλλειψη ενδιαφέροντος για την ανάπτυξη και έρευνα του θέματος με το οποίο διαπραγματεύεται. Όπως ανέφερα και στην εισαγωγή είναι πάρα πολύ δύσκολο να συγκεραστούν διαφορετικές αντιλήψεις και εν τέλει να αρθούν αδιέξοδα και στερεότυπα που έχουν να κάνουν με τη γλώσσα και τις προκαταλήψεις που ενέχει αυτή. Επίσης ένας σημαντικός περιορισμός είναι ότι τα αποτελέσματα της έρευνας προήλθαν από ένα μικρό δείγμα βιογραφιών και στις δύο γλώσσες με αποτέλεσμα τη μη πλήρη ποιοτική και ποσοτική ανάλυση τους.

Κατά την εκτέλεση του αλγορίθμου παρατηρήθηκαν δυσκολίες που οφείλονται κυρίως σε καταμέτρηση και προσπάθεια ανίχνευσης προκατειλημμένης γλώσσας στις βιογραφίες. Ο σχεδιασμός πιο ευαίσθητων σε λέξεις που ενέχουν γλωσσική προκατάληψη ίσως βοηθήσει στην πιο λεπτομερή καταγραφή του προβλήματος και τελικά της ανάδειξής του.

## 1.7 Δομή της Μεταπτυχιακής Διατριβής

Η μεταπτυχιακή διατριβή δομείται ως εξής:

Στο Κεφάλαιο 1 παρουσιάζεται το αντικείμενο της έρευνας, το θεωρητικό υπόβαθρο, οι σκοποί και οι στόχοι της μεταπτυχιακής διατριβής.

Στο Κεφάλαιο 2 παρουσιάζεται η ανασκόπηση της βιβλιογραφίας για τον καθορισμού του θεωρητικού πλαισίου που αφορά την έρευνα. Αναλύονται οι έννοιες που αφορούν το θέμα της γλωσσικής προκατάληψης των στερεότυπων και επίσης, αναλύονται οι σχετικές προηγούμενες μέθοδοι που αφορούν αυτό το θέμα.

Στο Κεφάλαιο 3 καλύπτεται η μεθοδολογία που ακολουθήθηκε σε αυτή τη διατριβή για να απαντήσει στα ερευνητικά ερωτήματα. Αναφέρονται λεπτομέρειες για την επιλογή της μεθόδου, τη διαδικασία συλλογής των δεδομένων αλλά και της ανάλυσης των αποτελεσμάτων που δικαιολογούν την επιλογή τους για τη συμβολή στην εγκυρότητα και την αξιοπιστία της έρευνας.

Στο Κεφάλαιο 4 παρουσιάζονται τα αποτελέσματα της έρευνας σε σχέση με τα ερευνητικά ερωτήματα έτσι ώστε να αξιοποιηθεί το προηγούμενο κεφάλαιο. Αναφέρονται καταγράφονται και αξιολογούνται οι μετρήσεις που προέκυψαν από την εφαρμογή του αλγόριθμου.

Στο Κεφάλαιο 5 συγκεντρώνονται τα στοιχεία και τα ευρήματα αυτής της έρευνας. επισημαίνεται η συνεισφορά της στο ερευνητικό πεδίο και αναφέρονται οι λόγοι που προέκυψαν αυτά τα ευρήματα και θέτονται οι περιορισμοί μελέτης. Τέλος, επανεξετάζονται τα ερευνητικά ερωτήματα και θέτει εισηγήσεις για περαιτέρω έρευνα.



# Κεφάλαιο 2

## Βιβλιογραφική Ανασκόπηση

### 2.1 Η γλώσσα ως μέσο επικοινωνίας

Η γλώσσα, ως έναρθρος και δομημένος λόγος, ανήκει στις μορφές επικοινωνίας που ανέπτυξε ο άνθρωπος από τα πανάρχαια χρόνια. Μέσα στην διάρκεια του χρόνου, αφιέρωσε σε αυτή μεγάλο μέρος της δημιουργικής του ενέργειας, διαμορφώνοντας δεκάδες επιμέρους γλώσσες και διαλέκτους, γραπτές και προφορικές. Η γλώσσα είναι το αποτέλεσμα των μοναδικών ικανοτήτων του ανθρώπινου είδους, που υπάρχουν στα κέντρα λόγου τα οποία διαθέτει και τα οποία παράγουν τον ανθρώπινο λόγο. Έτσι μέσω της γλώσσας, αναδεικνύονται τα νοητικά χαρακτηριστικά του ανθρώπου. Μέσω της γλώσσας αφήνουμε πίσω μας το παρόν, μιλάμε για το παρελθόν και ανακαλύπτουμε το μέλλον ( Νοάμ Τσόμσκι, 1950).

Ασφαλώς, οι φυσικές γλώσσες παγκοσμίως είναι πολλές, με πολυποίκιλα γνωρίσματα. Αυτό δημιουργεί μια τροχοπέδη στην επικοινωνία ανάμεσα σε διαφορετικούς λαούς μια και η σημερινή εποχή, εποχή της πληροφορίας, της πληροφορικής, της «επανάστασης της τηλεματικής», επιβάλλεται, πιο πολύ, μια παγκοσμιοποιημένη γλώσσα. Τη δεκαετία του '50 ο καθηγητής της Γλωσσολογίας στο Τεχνολογικό Ινστιτούτο της Μασαχουσέτης Νοάμ Τσόμσκι διατύπωσε μια επαναστατική θεωρία για τη γλώσσα η οποία στηρίζεται στη μελέτη των ομοιοτήτων των διαφόρων γλωσσών δηλ. των γλωσσικών παγκοσμιοτήτων παρά των διαφορών τους. Απέδειξε ότι ο άνθρωπος έχει μια συγγενή ικανότητα να αναγνωρίζει τις «παγκοσμιοότητες» που χαρακτηρίζουν μια φυσική γλώσσα. Όταν εκτίθεται σε μια γλώσσα που εμπεριέχει αυτές τις παγκοσμιοότητες τότε τις μαθαίνει εύκολα και με ευχαρίστηση. Αντίθετα μια γλώσσα που καταστρατηγεί αυτούς τους κανόνες δεν είναι δυνατό να μαθευτεί.

## 2.2 Η βιολογική εξέλιξη της γλώσσας στα δύο φύλα

Θεμελιακής σημασίας διαπίστωση είναι ότι η γλώσσα είναι αναπόσπαστα συνδεδεμένη με την εξελικτική πορεία του ανθρώπινου γένους. Από αρχή του κόσμου η επικοινωνία μεταξύ των ανθρώπων γινόταν με νοηματικές κινήσεις των χεριών και με άναρθρες κραυγές. Οι άνθρωποι μένοντας σε σπηλιές και εκτελώντας τα αρσενικά και τα θηλυκά διαφορετικές εργασίες, συν το γεγονός ότι αυτές οι εργασίες ήταν επαναλαμβανόμενες για πάρα πολλά χρόνια, προσδιόρισε σε μεγάλο βαθμό και τη διαφορετική ανάπτυξη του εγκεφάλου του αρσενικού από το θηλυκό.

Συγκεκριμένα στον εγκέφαλο των αρσενικών, αναπτύχθηκε περισσότερο το αριστερό ημισφαίριο, στο οποίο είναι εγκατεστημένο και το κέντρο της ομιλίας. Ο εγκέφαλος των θηλυκών αναπτύχθηκε πιο ομοιόμορφα και στο σύνολο του. Αναπτύχθηκε και το αριστερό και το δεξί μέρος του εγκεφάλου καθώς επίσης και η επικοινωνία μεταξύ των δύο ημισφαιρίων (Shorsetal, 2001). Με βάση ιατρικές μελέτες, το αριστερό ημισφαίριο του εγκεφάλου είναι υπεύθυνο για την λογική, επίτευξη στόχων και την πρακτική εφαρμογή τους ενώ το δεξί είναι επιφορτισμένο με ότι αφορά το συναίσθημα, την καλαισθησία και τη φαντασία (Δάλλα, 2012). Συμπερασματικά των προηγουμένων, πρέπει να αναφερθεί ότι η γλώσσα δηλώνει και την έννοια της συνεργασίας, διότι για να παραχθεί ο λόγος υπάρχει συνεργασία των νευρικών κέντρων του εγκεφάλου (Taylor et al., 2000).

Η βιολογική εξέλιξη της γλώσσας στα δύο φύλα έχει ερευνηθεί σε διάφορες μελέτες, όπως για παράδειγμα, των Bendow & Stanley (1980) στην οποία για τα «προικισμένα» στα μαθηματικά παιδιά, οι γλωσσικές ικανότητες τους, δεν παρουσίαζαν καμιά διαφορά. Στη μελέτη των Cook (1985) παρατηρείται ότι τα αγόρια μιλούν πιο πολύ και παίρνουν πρωτοβουλίες στη συνομιλία πιο νωρίς.

Η υπόθεση για την ταχύτερη βιολογική ωρίμανση των κοριτσιών είναι σχετικά αποδεκτή, δεν μπορεί να αποκλειστεί η πιθανότητα μιας πιο πρώιμης νευρολογικής ωρίμανσης, μολονότι, οι διαφορές των φύλων στους ρυθμούς ωρίμανσης είναι σχετικές και όχι μεγάλες. Ξεκινώντας λοιπόν από το γεγονός ότι ο εγκέφαλος δεν αναπτύχθηκε με τον ίδιο τρόπο στους άνδρες και στις γυναίκες, μπορούμε να καταλάβουμε για ποιο λόγο μετά από χιλιετηρίδες εξέλιξης, η γλώσσα η οποία διαμορφώθηκε εμπεριέχει πολλούς όρους, χρήσης αρσενικού γένους (Κλαίρης Φ. – Μπαμπινιώτης Γ., 2005).



Η χρήση του αρσενικού γένους είναι συντριπτική στα διοικητικά έγγραφα. Οι αρσενικοί τύποι ονομάτων, ειδικότερα, προηγούνται πάντα των θηλυκών, οι οποίοι κατά κανόνα δηλώνονται με την προσθήκη κατάληξης και όχι ολογράφως (Γκασσούκα - Γεωργαλίδου, 2014). Οι άνδρες επομένως, έχοντας αναπτύξει περισσότερο το αριστερό ημισφαίριο του εγκεφάλου, ανέπτυξαν τη γλώσσα και τη διαπραγματευτική ικανότητα και έβαλαν, με ένα περισσότερο ή λιγότερο συνειδητό τρόπο, τη σφραγίδα τους στη γλώσσα. Με αυτό τον τρόπο διατήρησαν στο διηνεκές, την εξουσία πάνω στις γυναίκες, υπενθυμίζοντας, μέσω των όρων αρσενικού γένους ότι η δύναμη ανήκει σε αυτούς. Επίσης ήθελαν να υπενθυμίζουν μέσω της γλώσσας στην καθημερινή συνομιλία των ανθρώπων, την διαφορετικότητα των ρόλων, την τοποθέτηση ορίων και ένδειξη κατεύθυνσης για το κάθε φύλο χωριστά. Η Spender (1980) βλέπει τους εκπροσώπους του ανδρικού φύλου επικυρίαρχους και στην εννοιολογική γλώσσα, εφόσον αυτή έχει να κάνει με την κατοχή θέσεων εξουσίας. Οι εκπρόσωποι αυτοί λοιπόν ρύθμισαν αυτό που θεωρείται κοινωνικά αποδεκτό, δηλαδή τις κανονιστικές αρχές κατά αυτόν τον τρόπο. Ο έλεγχος της γλώσσας λειτουργεί αμφίδρομα, δηλαδή αλληλεπιδραστικά, ως προς τον έλεγχο της ίδιας της πραγματικότητας (Γκασσούκα - Γεωργαλίδου, 2014).

Οι λέξεις που εκφράζουν τα μείζονος σημασίας για την ύπαρξη και λειτουργία μιας κοινωνίας, μεταφράζονται με εικόνες αρσενικών προτύπων π.χ ανδρεία, ηρωισμός, λεβεντιά κλπ. Σε αντίθεση με τους όρους που αναφέρονται στην γλώσσα και έχουν να κάνουν με πράγματα εκτός δύναμης και λογικής, οι οποίοι είναι κατά κύριο λόγο θηλυκού γένους, για παράδειγμα: ομορφιά, καλαισθησία, χάρη, αρμονία κλπ.

## 2.3 Η στερεοτυπική συμπεριφορά

Προσπαθώντας να δώσουμε έναν ορισμό στη λέξη και τον ειδικό όρο στερεότυπο, θα πρέπει να μιλήσουμε για a priori δομημένες πεποιθήσεις κοινωνιολογικής φύσης και αναφοράς, που μπορεί να είναι είτε αναχρονιστικές είτε δεν υποβάλλονται σε διεργασίες που έχουν να κάνουν με τον ορθό λόγο αλλά προέρχονται από άκριτες ιδέες και προσδοκίες, επομένως συχνά δεν ανταποκρίνονται στην πραγματικότητα (Judd et al., 1993). Από τις ιδέες αυτές και ιδιαίτερα όταν τις μοιράζονται μεγάλες ομάδες ανθρώπων, προέρχεται και η στερεοτυπική συμπεριφορά, καθώς και η κατανόησή της (Mc Garty et al., 2002).

Επομένως, η στερεοτυπική συμπεριφορά γεννάται από άκριτα, ανεπεξέργαστα και χωρίς συνολική περίσκεψη «συμπεράσματα» τα οποία, αφορούν κοινωνικές κατηγορίες ή ακόμη και μεμονωμένα άτομα. Το στερεότυπο μπορεί, για παράδειγμα, ειδικότερα να εκδηλώνεται με την μορφή της παραμορφωτικής, υπεραπλουστευτικής ή ενέχουσας υπερβολής θεώρησης αυτών των κοινωνικών κατηγοριών. Σαφές είναι ότι ένα στερεότυπο ως μια μορφή άκριτης -και συχνά αρνητικής- κριτικής αξιολόγησης παρέχει ένα υπόβαθρο για μια σειρά από δυσάρεστες συνέπειες όσον αφορά τις σχέσεις μεταξύ των ποικίλων ομάδων οι οποίες εμφανίζονται στις κοινωνίες (Operario et al., 2001). Στη διαμόρφωση των στερεοτύπων και των αντίστοιχων συμπεριφορών αυτονόητα μεγάλο ρόλο διαδραματίζει το κοινωνικό και πολιτιστικό συγκείμενο, το οποίο παρέχει συχνά ένα σύνολο τέτοιων έτοιμων αντιλήψεων προς το άτομο. Όπως έχουν επισημάνει ειδικοί ερευνητές (Van Knippenberg, 1999), η δημιουργία κατηγοριοποιήσεων και ακόμη και «στερεοτύπων» με την ευρύτερη έννοια δεν είναι απλώς μια θετική αλλά μια απαραίτητη διαδικασία από την οποία διέρχεται ο νέος άνθρωπος κατά την αντιληπτική του ανάπτυξη και ωρίμανση καθώς και την διαδικασία της κοινωνικοποίησής του. Αυτό που έχει ιδιαίτερη βαρύτητα και είναι ζητούμενο είναι αυτή η φυσιολογική διαδικασία να μην εκτραπεί σε μια διαδικασία άκριτης απόρριψης ή εσωτερικής σύγκρουσης με ευρύτερες κοινωνικές κατηγορίες. Τα στερεότυπα συνδέονται με την εθνική ή φυλετική καταγωγή, με το φύλο, και άλλα χαρακτηριστικά που συχνά καθίστανται φανερά όπως ο σεξουαλικός προσανατολισμός. Σε όλες αυτές τις κατηγορίες αποδίδονται πολλές φορές εξωπραγματικά και με εκτενέστατη αλλοίωση της πραγματικότητας γνωρίσματα με τρόπο που περνάει μέσα από διεργασίες αβασάνιστης γενίκευσης (Κωστούλα-Μακράκη, 2001).

## **2.4 Το φαινόμενο της γλωσσικής προκατάληψης**

Η γλώσσα, ως το κύριο μέσο παραγωγής και άρθρωσης λόγου, θεωρείται ακριβώς και ως το κύριο μέσο κοινοποίησης των στερεοτύπων και μετάδοσής τους από γενιά σε γενιά εκ μέρους του Τύπου και των ΜΜΕ (Otterbacher, 2015). Επιπλέον, η ανίχνευση του στερεοτυπικού λόγου διατυπώνεται κατ' ανάγκην και εκφράζεται και πάλι μέσα από τη γλώσσα και ιδιαίτερα την αφαιρετική επιστημονική γλώσσα από μερίδα εκπροσώπων της κοινωνικής ψυχολογίας. Πριν οριστεί η γλωσσική προκατάληψη, ας μνημονεύσουμε και το ευρύτερο φαινόμενο της «προκατάληψης» (bias).

Πρόκειται για μια κατ' ουσία εξωλογική τοποθέτηση που φτάνει τα όρια της ισχυρά θεμελιωμένης και εμπεδωμένης πεποίθησης (<http://www.merriam-webster.com/dictionary/bias>).

Αυτή η πεποίθηση φέρει έντονο το χαρακτήρα της συστηματικής αντιπαράθεσης προς μια διατύπωση επιχειρημάτων που ακολουθούν τις παραδεδεγμένες αρχές της Λογικής. Λογίζεται λοιπόν ως μια τοποθέτηση η οποία δεν ακολουθεί μεθοδολογία μορφοποίησης επιστημονική, ενώ συνιστά το αποτέλεσμα ορισμένων ψυχολογικών παραγόντων, ιδιαίτερα αρνητικής προδιάθεσης. Οι προκαταλήψεις αφορούν μια σειρά αντικειμένων: το φύλο, τη φυλετική ή εθνική καταγωγή και προέλευση, το παρουσιαστικό, κανονιστικές αρχές που προέρχονται από το κοινωνικό ή πολιτιστικό συγκείμενο και φτάνουν ως την αξιολογική κρίση του ανθρώπινου προσώπου. Θα έπρεπε εδώ να επισημανθεί ότι ο βαθμός συνειδητότητας κατά τον σχηματισμό της προκατάληψης ποικίλλει κατά περίπτωση (<http://som.yale.edu/getting-conscious-about-unconscious-bias>) (Chagpar, 2013). Αναφερόμενοι στις προκαταλήψεις που αφορούν ειδικά τη γλώσσα και τα γλωσσικά φαινόμενα, βασιζόμαστε στην πεποίθηση και αποδοχή με την οποία ο άνθρωπος σε οποιαδήποτε τάξη ή κατηγορία βρίσκεται (π.χ αγράμματος ή εγγράμματος), παρουσιάζει σειρά προκαταλήψεων σε κάθε άποψη της γλώσσας, που αναφέρονται στην αυθαίρετη προτίμηση και υποτίμηση, συμβόλων, λέξεων και συνδυασμών τους. Επίσης δημιουργείται σύγχυση στην επικοινωνία είτε εξαιτίας της μη χρήσης γραμματικών κανόνων, είτε μη αναγκαιότητας και ίσως χρησιμότητας της ανάλυσης του γραπτού κειμένου η οποία θα έπρεπε να ισχύει βασισμένη σε επιστημονικές προτάσεις.



## 2.5 Παράγοντες γλωσσικής διαφοροποίησης ανάμεσα στα δύο φύλα πριν την ενηλικίωση

1. Η γλωσσική διαφοροποίηση ανάμεσα στα δύο φύλα αρχίζει πριν ακόμη γεννηθεί το μωρό. Πριν το παιδί γεννηθεί, αρχίζει να λαμβάνει ερεθίσματα από ένα κόσμο που δε γνωρίζει, θα συνεχίσει μετά τη γέννηση με την επικοινωνία, που θα έχει με το υποκείμενο που το προσέχει και συνήθως είναι η μητέρα. Η κοινωνική συμπεριφορά και η απάντηση σε όλα τα ερεθίσματα, θα έχει ολοκληρωθεί πριν το μωρό συμπληρώσει ένα χρόνο.
2. Στα πρώτα χρόνια, εναρμονίζεται με το βιολογικό του ρυθμό, αλλά και το ρυθμό ζωής. Στη συνέχεια, τα μικρά απαντούν στα εξωτερικά ερεθίσματα με μονοσύλλαβες λεξούλες και με εκφράσεις του προσώπου.
3. Έχει παρατηρηθεί, ότι οι ενήλικοι μιλούν με διαφορετικό τρόπο προς τα αγόρια και τα κορίτσια. Αυτή η διαφορετική αντιμετώπιση εσωτερικεύεται από τα παιδιά. Την εσωτερίκευση αυτή, μπορεί να τη δει κανείς στα παιχνίδια των παιδιών, παρατηρώντας και το γλωσσικό κώδικα που χρησιμοποιούν. Στις περιόδους αυτές, εδραιώνονται καλά μέσα στα παιδιά και οι γλωσσικές διαφοροποιήσεις μεταξύ των δύο φύλων, όπως και η σχετική συμπεριφορά. Μέχρι τα παιδιά να μπουν στο σχολείο, έχουν αφομοιώσει πλήρως τα μοντέλα των δύο φύλων.
4. Είναι χαρακτηριστική η παρατήρηση, ότι τα αγόρια κυρίως τα καταφέρνουν στα μαθηματικά και στη γεωμετρία, ενώ τα κορίτσια στη φιλολογία. Τη γλωσσική ικανότητα και την ικανότητα επικοινωνίας, την αποκτάμε βρέφη, μέσα από τη αλληλεπίδραση με το περιβάλλον μας.

Έτσι η καθημερινή χρήση της γλώσσας είναι μια πραγματικότητα που όμως δεν αναιρεί την ειδικότερη μελέτη των γλωσσικών φαινομένων, δηλαδή η γλώσσα πρέπει να νοηθεί ως κάτι ρευστό και εν τη γενέσει του, αλλά η μελέτη αυτού του φαινομένου είναι κάτι θεμιτό ως μια διαδικασία διεξαγόμενη από την επιστήμη (Κωστούλα - Μακράκη, 2001). Ας σημειωθεί καταληκτικά πως όλες οι παραπάνω προκαταλήψεις θα έπρεπε να ενταχθούν στην κατηγορία των γλωσσικών, δηλαδή σχετικών με την γλώσσα και όχι γλωσσολογικών προκαταλήψεων (linguistic biases). Η γλωσσολογική προκατάληψη είναι ένα διαφορετικό φαινόμενο και μπορεί να οριστεί ως μια «*συστηματική ασυμμετρία στην επιλογή των λέξεων, ως συνάρτηση της κοινωνικής κατηγορίας στην οποία συγκαταλέγεται ο στόχος της*» (Beukeboom, 2014). Μέσα από την συνομιλία και τις εκδηλώσεις της, ο άνθρωπος συνηθίζει να πληροφορείται για την κοινωνική θέση, το επάγγελμα, την καταγωγή και την προσωπικότητα του συνομιλητή του, όλα εκ των οποίων αποτελούν αντικείμενα στερεοτυπικού λόγου.

Μέσα δηλαδή στον λόγο που παράγει ο εκάστοτε ομιλητής, πραγματοποιείται και κάποια αποστολή κοινωνικών μηνυμάτων ως συνδηλώσεις αυτού του λόγου. Ο κάθε συνομιλητής λαμβάνει τα μηνύματα αυτά ως μια μορφή πληροφορίας. Επομένως παρατηρείται ότι οι γλωσσικές προκαταλήψεις αυτού του είδους εμφανίζονται και κατά την παραγωγή λόγου, είτε σε επίπεδο δομής είτε σε επίπεδο περιεχομένου (Κωστούλα - Μακράκη, 2001).

## 2.6 Ο γλωσσικός σεξισμός στα δύο φύλα

Το γλωσσικό αυτό φαινόμενο αφορά συγχρόνως τη Γλωσσολογία, τη Ψυχολογία και την Κοινωνιολογία. Όπως ο ρατσισμός, έτσι και ο σεξισμός, σημαίνει πάντα εύνοια προς μια ομάδα σε βάρος μιας άλλης. Είναι συχνά περίπλοκος στις εκδηλώσεις του, διεισδυτικός μέσα στις πολλές συμπεριφορές ανθρώπων, που του παραδίνονται και είναι πολύ δύσκολο να αναγνωριστεί, να ουδετεροποιηθεί και να αλλάξει (Heatherington, 1980).

Σύμφωνα με έρευνα του Πανεπιστημίου Κρήτης, «Το φύλο στην Ιστορία της Εκπαίδευσης», ορίζεται ότι η θηλυκότητα συνδέεται με την «κοινωνικότητα», την «εκφραστικότητα» και τη «συντροφικότητα». Η αρρενωπότητα με την «τραχύτητα», την «εργαλειακότητα», την «υπεροχή» και τον «έλεγχο». Το να μαθαίνεις να είσαι «θηλυκό» ή «αρσενικό» σε μία κοινωνία, σημαίνει μεταξύ άλλων μαθαίνεις να χρησιμοποιείς την ανάλογη γλώσσα. Τα μικρά κορίτσια μαθαίνουν να μιλούν σαν «κυρίες». Αν δεν ακολουθούν τους κανόνες, ή απομονώνονται ή γελοιοποιούνται. Τα μικρά αγόρια από την άλλη, μαθαίνουν να μιλούν «αντρίστικα».

Στη γλωσσική συμπεριφορά των γυναικών, βλέπουμε να πραγματώνεται, από πολύ νωρίς, το πρότυπο του «ωραίου» και «ασθενούς» φύλου. Μια γυναίκα δεν είναι «σωστό» να βρίζει, αλλά θα ήταν πολύ καλύτερο να σωπαίνει (γυναίξι πάσαις κόσμον η σιγή φέρει). Ο Ellis (1895) επίσης δέχεται, ότι οι γυναίκες διακρίνονται σε ένα μόνο τομέα γλωσσικών ικανοτήτων, τη λογοτεχνική γραφή. Οι ερμηνείες, που μερικές φορές διατυπώνονται για το γιατί οι γυναίκες μιλούν «σωστά», σύμφωνα δηλαδή με το επίσημο κοινωνικό πρότυπο, αναφέρονται στη δειλία και στο συντηρητισμό τους και αυτό σε αντίθεση με τους άντρες, που υποτίθεται πως τολμούν να αμφισβητήσουν το κοινωνικό κατεστημένο, μέσα από χρήσεις της γλώσσας, όπως η αργκό (Μπέκα Αποστολία, 2008).

## 2.7 Ο κοινωνικός ρόλος στη διαμόρφωση της γλωσσικής συμπεριφοράς

Η γλωσσική συμπεριφορά των ανθρώπων συνδέεται με διάφορες προδιαγραφές κυρίως σύμφωνα με τον κοινωνικό ρόλο, που καλούνται να αναλάβουν. Στη γλωσσική συμπεριφορά των γυναικών, βλέπουμε να πραγματώνεται ήδη από πολύ νωρίς το πρότυπο του «ωραίου» και «ασθενούς φύλου». Η φράση «σαν αγοροκόριτσο η μιλιά σου», είναι μία προειδοποίηση αρνητικού προτύπου για τα κορίτσια. Άνδρες και γυναίκες, είναι κοινωνικά άνισοι και η κοινωνία οριοθετεί για αυτούς διαφορετικούς κοινωνικούς ρόλους, αλλά περιμένει και διαφορετική συμπεριφορά. Η γλώσσα, απλά αντικατοπτρίζει αυτή την κοινωνική διαφορά. Οι γυναίκες πρέπει να είναι πιο «σωστές», έστω και γλωσσικά. Τα taboo μιας κοινωνίας, είναι οι φόβοι και οι συγκρούσεις της, που η κοινωνία δύσκολα αντιμετωπίζει.

Η σύνδεση μεταξύ της γλώσσας των γυναικών και των taboo, είναι από όλους παραδεκτή. Αλλά τα taboo συνδέονται και με τη γλώσσα των ανδρών. Οι απαγορεύσεις των taboo και οι έλεγχοι που ασκούν, έχουν να κάνουν με την κοινωνική δομή και το πολιτιστικό περιβάλλον.

## 2.8 Η έμφυλη γλωσσική προκατάληψη στα Μ.Μ.Ε

Μια ενδιαφέρουσα και σημαντική πτυχή της σεξιστικής χρήσης της γλώσσας στο δημόσιο χώρο από τα Μέσα Μαζικής Επικοινωνίας, που απαιτεί ιδιαίτερη προσοχή, είναι οι στερεότυπες αναφορές σε γυναίκες με έμφαση στα σωματικά τους προσόντα και στην οικογενειακή τους κατάσταση. Είναι χαρακτηριστική η αντιμετώπιση των γυναικών σε ισχυρές δημόσιες θέσεις που παρουσιάζονται συχνά ως "ξανθιές", "ωραίες", "χήρες" ή "μητέρες" (βλ. Λιάνη, Τσιλέρ). Ανάλογες αναφορές στους δημόσιους άνδρες είναι σχεδόν ανύπαρκτες. Ταυτόχρονα οι γελοιογραφίες και τα σκίτσα στα διάφορα έντυπα αντανakλούν με τη σειρά τους μια ανδροκρατούμενη πραγματικότητα, καθώς απεικονίζουν το ανθρώπινο είδος ως κατ' εξοχήν ανδρικό, ενώ η χρήση των γυναικείων μορφών είναι περιορισμένη. Για να διαπιστωθεί αυτό αρκεί να ξεφυλλίσουμε μια οποιαδήποτε εφημερίδα και να προσέξουμε τις γελοιογραφίες της. Για να αποφευχθεί η ανδροκρατία σ' αυτό το επίπεδο είναι επιτακτική η χρήση περισσότερων γυναικείων μορφών στα έντυπα, που, φυσικά, δεν πρέπει να παίζουν διακοσμητικό ρόλο. Η Βίλμα Παπασάββα, στο περιοδικό Μετρό (τεύχος 6, Απρίλιος 1996), θίγει το θέμα της απεικόνισης των γυναικών στα τηλεοπτικά σήριαλ, ως «ηλίθιες, σεξομανείς, κουτσομπόλες, υπερφίαλες "Κατίνες"». Φυσικά η δημιουργία μιας τέτοιας γυναικείας εικόνας γίνεται μέσα από την εμφάνιση (ρούχα, μαλλιά, μακιγιάζ), αλλά και από τη γλωσσική τους συμπεριφορά.

Το λεξιλόγιο, ο τόνος της φωνής τους, καθώς και η παραγλωσσική τους συμπεριφορά (δηλαδή οι κινήσεις, χειρονομίες και το ύφος τους) ορίζουν αυτόν τον "υπερφίαλο" χαρακτήρα τους. Οι λέξεις «υστερική» και «κυκλοθυμική» επιβεβαιώνουν ένα στερεότυπο, γιατί αυτές οι λέξεις χρησιμοποιούνται συνήθως για να χαρακτηρίσουν τη γυναικεία συμπεριφορά. Η δύναμη των ΜΜΕ στη διαιώνιση τέτοιων αρνητικών γυναικείων χαρακτήρων δεν πρέπει να υποτιμάται, ειδικά όταν αυτά τα σήριαλ προβάλλονται σε ώρες πολύ κατάλληλες για τα παιδιά.

## **2.9 Γλωσσική προκατάληψη στο Διαδίκτυο και στα Μέσα κοινωνικής δικτύωσης**

Η λεγόμενη «Κοινωνία της Πληροφορίας» (Information Society) είναι μια μορφή κοινωνίας η οποία παράγει και διαχειρίζεται μεγάλο όγκο πληροφοριών, επιφέροντας έτσι σημαντικές μεταβολές στους όρους επικοινωνίας, τη συμβολική παραγωγή, τη χωροχρονική οργάνωση της κοινωνίας, την ανάδυση μιας νέας οικονομίας βασισμένης στην παροχή και χρήση πληροφοριών και τέλος στον καταμερισμό της εργασίας (Κάλλας, 2006). Ευρύτερα είναι ένας νέος τρόπος κοινωνικής ύπαρξης, στον οποίο υπάρχει οργανωμένη παραγωγή, αποθήκευση, ανάκτηση και χρήση πληροφορίας. Σε αυτή την νέα τάξη πραγμάτων εμφανίζονται και δημιουργούνται δίκτυα τα οποία παράγουν μια δικτυωμένη κοινωνία, εντός της οποίας συσσωρεύονται και αναπαράγονται θεσμοί από την ήδη υπάρχουσα. Αποτέλεσμα αυτής της αλλαγής σε μακροσκοπικό επίπεδο είναι η γενικότερη αναδιαμόρφωση των θεσμών που διέπουν την κοινωνία σε πολιτικό, οικονομικό και πολιτισμικό επίπεδο. Ενώ σε μικροσκοπικό και μεσοσκοπικό επίπεδο επιφέρει αλλαγές στο οικογενειακό περιβάλλον και σε κάθε άτομο ξεχωριστά (Pinter, 2006).

Ένα από τα ορόσημα που συνέβαλαν σε αυτόν το μετασχηματισμό της κοινωνίας ήταν το Διαδίκτυο. Το Διαδίκτυο συνιστά ένα μέσο το οποίο αναπτύσσεται στο πλαίσιο της Κοινωνίας της Πληροφορίας αλλάζοντας έτσι τους όρους της υπάρχουσας κοινωνίας, διαμορφώνοντας μια νέα τάξη πραγμάτων, στην οποία ο άνθρωπος καλείται να προσαρμοστεί. Ένας σύντομος ορισμός του κυβερνοχώρου, είναι η σύγκληση όλων των υπολογιστών και των τηλεπικοινωνιακών δικτύων μέσω ενός κοινού πρωτόκολλου μεταφοράς δεδομένων (Κουρτή, 2003). Αφορά ένα ιδεατό μοντέλο (Αβούρης, 2000) που συγκροτείται σε ένα νοητικό χώρο στον οποίο συνδέονται μεταξύ τους δίκτυο, λογισμικό και χρήστες, που έχουν ως απότοκο τη δημιουργία μιας «κυβερνο – κουλτούρας» αποτελούμενης από μηχανές και ανθρώπινη κοινωνικότητα (Bell, 2001). Σύμφωνα με τη θεωρία της Κυβερνητικής (Wiener, 1948) η οποία δέχεται την επικοινωνία ως κομμάτι του ελέγχου και αντίστροφα, ο κυβερνοχώρος δεν μπορεί να ελεγχθεί. Είναι σημαντικό να κατανοήσουμε την κοινωνική πλευρά του και ως εκ τούτου την τοποθέτησή του, ανάμεσα στα νέα μέσα επικοινωνίας.

Πολύ χαρακτηριστικό είναι ότι ο εφευρέτης του WEB Tim Berners – Lee, αποτιμά το δημιούργημά του, περισσότερο ως κοινωνικό παρά ως τεχνολογικό κατασκεύασμα (Κουρτή, 2003). Ως αποτέλεσμα των πιο πάνω, η επικοινωνία μεταξύ των ατόμων εδραιώνεται και αναπτύσσεται και εντός των πλαισίων του κυβερνοχώρου με επακόλουθο την επίδραση ή καλύτερα την επέκταση από την «εικονική» ζωή, στην «πραγματική» και αντίστροφα. Αν δεχτούμε όμως αυτή τη μετάβαση του «εικονικού» εαυτού στον «πραγματικό» τότε, με πιο τρόπο υπερνικάτε το εμπόδιο της φυσικής υπόστασης του ατόμου; Ο άνθρωπος αποτελείται από μυαλό και σώμα, το διαδίκτυο του δίνει τη δυνατότητα να ταξιδέψει μόνο νοητικά, αφού το σώμα του παραμένει καθηλωμένο μπροστά από την οθόνη του υπολογιστή. Με βάση λοιπόν αυτή την προοπτική αναπτύσσεται το ερώτημα κατά πόσο μπορούμε να δεχτούμε την ύπαρξη μιας νέας κοινωνίας.

Μπορούμε να υποστηρίξουμε ότι η Κοινωνία της Πληροφορίας μας προσφέρει σχέσεις ασώματες, άυλες οι οποίες απορρίπτουν την προσωπική, φυσική επαφή; Μεγάλο μέρος των οπαδών του Διαδικτύου πιστεύουν πως το Διαδίκτυο επιφέρει την ισότητα καταρρίπτοντας τις κοινωνικές δομές και προβάλλοντας μια ωραιοποιημένη ταυτότητα. Είναι όμως αυτή η νέα κοινωνία υπαρκτή ή μήπως ένα σύνολο στο οποίο καθένας από εμάς μπορεί να κατακερματίσει την ταυτότητα του σε διάφορες προσωπικότητες που στην ουσία δεν του ανήκουν στο έπακρο; Ως αποτέλεσμα των πιο πάνω καταλήγουμε στη θεώρηση ότι η κοινωνία αλλάζει και γίνεται ατομικοκεντρική. Το άτομο εκδηλώνει μια κουλτούρα ναρκισσισμού, στρέφεται προς τον εαυτό του και αντιμετωπίζει δυσκολία στο να δημιουργήσει σχέσεις και δεσμούς. Το Διαδίκτυο έρχεται να καλύψει αυτό το κενό αναπτύσσοντας σχέσεις, όχι όμως δεσμούς (Mann, 2000). Για παράδειγμα οι νέοι του σήμερα μέσα από τη δυνατότητα που τους παρέχεται για επικοινωνία μέσω μηχανών απομακρύνονται μεταξύ τους αφού η προσωπική επαφή αντικαταστάθηκε από το τηλέφωνο και το τηλέφωνο τώρα πια από το απρόσωπο μήνυμα. Ίσως η τεχνολογία να μην ευθύνεται για την απομάκρυνση, ούτε και η κοινωνία των δικτύων που έχει συσταθεί, μα ο τρόπος ζωής, οι γρήγοροι ρυθμοί που επιβάλλουν πολλές ασχολίες ταυτόχρονα κατακερματίζοντας το άτομο.

Ο άνθρωπος λοιπόν όντας κοινωνικό ον προσπαθεί να βρει την ευκαιρία να κοινωνικοποιηθεί ακόμα και αν αυτό θα γίνει παράλληλα με άλλες ασχολίες. Τα δίκτυα λοιπόν, προσφέρουν αυτή τη δυνατότητα και αυτός την αξιοποιεί άλλοτε σωστά και άλλοτε λανθασμένα, δημιουργώντας κατ' αυτό τον τρόπο το λεγόμενο φαινόμενο του «καθημερινού εγώ» (Sunstein, 2002). Το άτομο επιλέγει να διαβάσει μόνο ότι επιθυμεί περιορίζοντας κατά αυτό τον τρόπο την ελευθερία του λόγου. Αντίθετα με τα πιο πάνω οι διαδικτυακές κοινωνικές ομάδες αποτελούν «μικρογραφία δημόσιων σφαιρών» οι οποίες κατά τον Preston (2001) δεν μπορούν να μεταβάλουν τις υπάρχουσες κοινωνικές δομές. Η πιο πάνω άποψη αντιτίθεται στο ρεύμα του διαδικτυακού – ακτιβισμού ο οποίος ενεργοποιείται και γίνεται γνωστός διαμέσου του Διαδικτύου επιφέροντας ισχυρότερες μεταβολές από ότι στη μη – διαδικτυακή του διάσταση (McCaughy & Ayers, 2003). Παράδειγμα αυτού μπορούν να αναφερθούν οι επιπτώσεις των Κοινωνικών Μέσων στις ερωτικές σχέσεις του ατόμου.



Αναμένουμε λοιπόν ότι το άτομο δε θα θυματοποιείται στην «πραγματική» ζωή ως αποτέλεσμα της εικονικής του δράσης. Λαμβάνοντας όμως υπόψη την γενικότερη αναδιάταξη και τη δημιουργία μιας νέας τάξης πραγμάτων στο ευρύτερο κοινωνικό - επικοινωνιακό πλαίσιο αναμένουμε σημαντικές και έντονες διαφοροποιήσεις, που άπτονται στον επηρεασμό που δεχόμαστε από τα κοινωνικά δίκτυα, με την έννοια της ηθικοκοινωνικής και συμπεριφορικής εξέλιξης του ατόμου εντός αυτού του πλαισίου (Christakis & Fowler, 2009). Συνεπώς, το Διαδίκτυο ανάγεται σε ένα κοινωνικό μοτίβο στο οποίο οι δομές της επικοινωνίας συνεχίζουν να υφίστανται τροποποιώντας την κοινωνική πραγματικότητα και μετατρέποντας την σε ένα πιο σύνθετο πλαίσιο. Για όλα τα πιο πάνω το Διαδίκτυο καθίσταται ένα ιδιαίτερα ενδιαφέρον πλαίσιο μελέτης, λόγω των ιδιαίτερων χαρακτηριστικών του καθώς και λόγω των αλλαγών που επιφέρει στους ρόλους, τα άτομα αλλά και τις κοινωνικές δομές. Αποτελεί λοιπόν μια ψηφιδωτή κοινωνική μορφή μιας παγκόσμια «ιστό - κουλτούρας» στην οποία η πραγματικότητα ορίζεται δυνητικά και τα σύμβολα αναπαριστούν εμπειρίες (Castells, 1998). Ως αποτέλεσμα του πιο πάνω η διαδικτυακή ταυτότητα είναι πολυσυλλεκτική αφού ταυτόχρονα εμπερικλείει τόσο το πραγματικό όσο και το δυνητικό στοιχείο, ένα στοιχείο το οποίο οργανώνεται κατά βάση τεχνολογικά, αλλά αναπτύσσεται κοινωνικά διαμέσου των πληροφοριών, των αναπαραστάσεων και της εμπλοκής των ατόμων με άλλους (Castells, 2001). Έτσι η έμφαση δίνεται στη φύση του μέσου αφού αν και «εικονικό» ζει και αναπτύσσεται διαμέσου πραγματικών συμπεριφορών, κατασκευάζοντας έτσι την πραγματικότητα, την κοινωνία και τον ίδιο τον εαυτό (Δεμερτζής, 2000). Παρά το γεγονός ότι το Διαδίκτυο είναι ένας τεχνητός, εικονικός κόσμος, οι συμπεριφορές δεν παύουν να είναι ουσιαστικά όσο πραγματικές είναι και στον έξω κόσμο. Η επικοινωνία στο Διαδίκτυο καθίσταται, όπως προαναφέρθηκε, τεχνολογικά διαμεσολαβούμενη. Τα άτομα εκφράζουν κοινωνικά την ταυτότητά τους σε ένα άυλο, θα έλεγε κανείς, χωροχρόνο, εδραιώνοντας ή απορρίπτοντας σε αρκετές των περιπτώσεων άτομα και καταστάσεις, δημιουργώντας έτσι μια νέα κοινωνική «εικονική» πραγματικότητα που όμως εμπίπτει σε «πραγματικές» συμπεριφορές, εμπειρίες και καταστάσεις.

Είναι ένας χώρος ο οποίος επιτρέπει την κοινωνική αλληλεπίδραση με «γενικευμένο άλλο» (Mead, 1934), το κοινωνικό περιβάλλον δηλαδή, που ορίζει τη συμπεριφορά του ατόμου και κατευθύνει τη δράση του προς την αποδοχή. Μια αφηρημένη με άλλα λόγια κατασκευή στο μυαλό του ατόμου, που του προσφέρει συνολική και γενικευμένη αίσθηση της πραγματικότητας μέσα στην οποία καλείται να δράσει, αναπτύσσοντας παράλληλα σχέσεις με τους «σημαντικούς άλλους» (Mead, 1934) όπως κοινωνικές ομάδες, φίλους, οικογένεια, παρέχοντας ταυτόχρονα τη δυνατότητα για τη «διαχείριση των εντυπώσεων». Σύμφωνα με τον Goffman (1959) το ζητούμενο σε κάθε πράξη επικοινωνίας είναι η «διαχείριση των εντυπώσεων» καθώς και ο «ορισμός της κατάστασης». Στην πρώτη περίπτωση αναφερόμαστε σε όλα αυτά που δίνονται αλλά και σε όλα αυτά που προδίδονται από το άτομο ορίζοντάς το. Το βασικό σημείο λοιπόν στην επικοινωνία είναι ο τρόπος που το άτομο διαχειρίζεται όλες αυτές της πληροφορίες συγκροτώντας κατ' αυτόν τον τρόπο την εικόνα του εαυτού του.

Στη δεύτερη περίπτωση το άτομο προσπαθεί να ορίσει τη συμπεριφορά του άλλου με τέτοιο τρόπο που να του παρέχονται οι απαραίτητες πληροφορίες με στόχο να κατανοήσει τα προσδοκώμενα από τον άλλον. Αυτές οι βασικές αρχές επικοινωνίας μεταφέρονται και στα Μ.Μ.Ε, ωστόσο το διαδίκτυο παρέχει τη δυνατότητα στο άτομο να διαχειρίζεται με περισσότερη ευκολία τις εντυπώσεις, αφού αυτά που προδίδονται είναι αυτά που το ίδιο το άτομο επιλέγει να δώσει. Το άτομο έχει τη δυνατότητα να επιλέξει με περισσότερη προσοχή τις εντυπώσεις (Wallace, 1999) δημιουργώντας μια πιο ελκυστική ταυτότητα και μια πιο αγαπητή αυτοεικόνα. Είναι κατ' ακρίβεια ένα «παιχνίδι πληροφόρησης» που ενέχει διεργασίες αποκάλυψης και απόκρυψης στοιχείων (Goffman, 1959) με στόχο την κατασκευή και την επεξεργασία του ίδιου του εαυτού (Giddens, 1991). Ως αποτέλεσμα αυτού ένα προφίλ στα Κοινωνικά Μέσα ή αλλιώς στις ιστοσελίδες κοινωνικής δικτύωσης, όπως είναι για παράδειγμα το Facebook, δεν αποτελεί απλά την προέκταση του εαυτού στα Κοινωνικά Μέσα και στον χώρο του διαδικτύου. Σε πολλές περιπτώσεις υποδηλώνει και την ωραιοποίηση του εαυτού διαμέσου κολακευτικών φωτογραφιών. Αποτελεί λοιπόν έναν χώρο στο οποίο το άτομο μπορεί να διαχειριστεί και να τροποποιήσει την ταυτότητα του, με τέτοιο τρόπο που να μην γίνει αναφορά σε παραπλανητική εξαπάτηση αλλά σε πειραματισμό με τον υφιστάμενο εαυτό (Turkle, 1995- Wallace, 2001). Έτσι το άτομο στις ιστοσελίδες κοινωνικής δικτύωσης αλλά και γενικότερα στον κυβερνοχώρο είναι σαν να φορά μια προστατευτική μάσκα που του επιτρέπει να αποκαλύπτει ευκολότερα τον επιθυμητό του εαυτό (Peace, 2000. Kollock & Smith, 1999). Το ζήτημα που τίθεται σε αυτό το σημείο είναι κατά πόσο στον «εικονικό» κόσμο το άτομο είναι πραγματικά ελεύθερο και απαλλαγμένο από την κοινωνική επιρροή και τις αναπαραστάσεις που του δημιουργούνται στον «πραγματικό» κόσμο. Η προηγούμενη ανάλυση μας υποβάλλει την ιδέα ότι η διαδικτυακή γλώσσα βρίθκει προκαταλήψεων και στερεοτυπικού λόγου, ιδιαίτερα αυτών τα οποία μελετήσαμε. Μια θεμελιακής σημασίας παρατήρηση πάνω σε αυτό το ζήτημα είναι η γνωστική προέλευση των προκαταλήψεων οι οποίες καταγράφονται στον κυβερνοχώρο. Μάλιστα, αυτή η προέλευση δεν αφορά διατύπωση προσωπικών απόψεων σε ιστοχώρους κοινωνικής δικτύωσης, αλλά ιστοχώρους συλλογικής διαμόρφωσης, όπως το IMDb και ασφαλώς η Wikipedia (Otterbacher, 2015).



## 2.10 Η τάξη του κόσμου και οι αλλαγές στο γλωσσικό επίπεδο

Η τάξη του κόσμου βρίσκεται σε διαρκή κίνηση και αλλαγή, με αποτέλεσμα να απαιτούνται και αντίστοιχες αλλαγές στον τρόπο έκφρασης και στην «κανονικότητα της αρσενικής ιδιότητας σε σχέση με τους κοινωνικούς τίτλους». Χρησιμοποιούμε, δηλαδή, απ' τη μια τις υπάρχουσες διαδικασίες της γραμματικής και του συντακτικού για να περιγράψουμε τη γυναικεία πραγματικότητα και από την άλλη, ανακυκλώνουμε και επαναπροσδιορίζουμε όρους σημασιολογικά, έτσι ώστε αυτοί να μπορούν να καλύψουν τις ανάγκες του γυναικείου φύλου. Τέλος, η δημιουργία και καθιέρωση καινούριων λέξεων και εννοιών είναι επίσης μια απαραίτητη τακτική στην πορεία μας για την εξάλειψη του σεξισμού τόσο στη γλώσσα όσο και στη σκέψη. (Φραγκουδάκη, 1995). Για παράδειγμα στη δημόσια πραγματικότητα εντάσσεται η χρήση των τίτλων "κύριος", "κυρία" και "δεσποινίς".

Παραδοσιακά ο αρσενικός τίτλος αναφέρεται σε κάθε άντρα που έχει περάσει την εφηβική ηλικία, συνήθως όταν είναι άνω των 18, ενώ η χρήση των δύο θηλυκών τύπων γίνεται με κριτήριο την οικογενειακή κατάσταση της γυναίκας, δηλαδή το αν είναι άγαμη (δεσποινίς) ή παντρεμένη (κυρία). Αυτή η διαφορά δηλώνει ότι οι γυναίκες για να γίνουν κυρίες πρέπει όχι να ενηλικιωθούν, όπως οι άντρες, αλλά να παντρευτούν έναν άντρα. Αντίστοιχοι όροι υπάρχουν και σε πολλές άλλες γλώσσες, και αλλαγές σ' αυτήν την κατεύθυνση έχουν διεκδικηθεί από πολλές γυναικείες οργανώσεις τόσο στην Ελλάδα όσο και στις ευρωπαϊκές και αγγλόφωνες χώρες. Παραδείγματα αυτής της διάκρισης των δύο φύλων είναι:

Ελληνικά Δεσποινίς/Κυρία

Αγγλικά Miss/Mrs Ms

Γαλλικά Mademoiselle/Madame

Ισπανικά Senorita/Senora



## 2.11 Το φύλο της γλώσσας

Πορίσματα ερευνών που έγιναν σε διάφορα μέρη του κόσμου απέδειξαν ότι το λεξιλόγιο που χρησιμοποιούνταν στην επίλυση σχολικών ασκήσεων επηρέαζε δραστικά τα αποτελέσματα (Sorrels, 1983). Όταν το ζητούμενο ήταν τα παιδιά να ζωγραφίσουν εικόνες που να αντιστοιχούν στους αγγλικούς όρους "social man" ή "political man", αυτά ζωγράφιζαν άντρες, αν και η αγγλική λέξη "man" χρησιμοποιείται για να χαρακτηρίσει την ανθρωπότητα. Ας αναλογιστούμε όμως την ελληνική "ουδέτερη" λέξη "άνθρωπος" σε μία πρόταση όπως "Ο άνθρωπος έφαγε τα μούτρα του στις σκάλες". Ποια ή ποιος θα σκεφτόταν ποτέ έναν άνθρωπο θηλυκού γένους να είναι η παθούσα στην προκειμένη περίπτωση; Η σεξιστική γλώσσα επομένως δεν είναι μόνο το σύμπτωμα, αλλά και η αιτία.

Το κίνημα ενάντια στη μη σεξιστική γλώσσα πρέπει να διαχωριστεί από αυτό του πολιτικώς ορθού (Cerf, 1992), που ανθίζει, τον τελευταίο καιρό, ιδιαίτερα στην Αμερική και σύμφωνα με το οποίο σχεδόν κάθε λέξη με αρνητικά υπονοούμενα αντικαθίστανται από μια θετικότερη. Η χρήση αρσενικού ως του επικρατέστερου γένους και αυτού που αντιπροσωπεύει το ανθρώπινο είδος ισχύει στερεότυπα σε όλες τις ευρωπαϊκές γλώσσες που έχουν γραμματικό γένος. Αυτή είναι η "γενετική" ή "ρυθμιστική" χρήση του αρσενικού τύπου (Μακρή - Τσιλιπάκου, 1995).

Γαλλικά: Tous les etudiants Tous/ toutes les etudiants /etudiautes
Ισπανικά: To dos estudiantas To dos /to das los/las studentcs
Ελληνικά: Όλοι οι φοιτητές Όλοι/Όλες οι φοιτητές/τριες

## 2.12 Η γλωσσική προκατάληψη στην ελληνική γλώσσα

Η ελληνική γλώσσα είναι κυρίαρχο μέσο για την εξελικτική πορεία του ανθρώπου. Μελέτες πάνω στα αρχαία ελληνικά, εκ μέρους ιδιαίτερα του γλωσσολόγου Charles Higounet (1970), έχουν δείξει ότι υπάρχει σχέση μεταξύ της ελληνικής γλώσσας, των μαθηματικών, της φυσικής, της μουσικής, της αστρονομίας. Επίσης οι άνθρωποι που γνωρίζουν και μιλούν καλά τα αρχαία ελληνικά φαίνεται ότι έχουν και καλύτερη υγεία, διότι η σωστή προφορά της γλώσσας, βοηθάει στην σωστή αναπνοή και δημιουργεί καλό βιορυθμό στο ανθρώπινο σώμα. Στην ελληνική γλώσσα κάθε λέξη μεταφέρει πληροφορίες για το αντικείμενο που περιγράφει.

Αν επιχειρήσουμε να εξετάσουμε την ελληνική γλώσσα διαπιστώνουμε ότι από την πρώτη της γέννηση, κατά την προομηρική εποχή, έως σήμερα έχουν συμβεί μεγάλες και αυταπόδεικτες μεταβολές τόσο στον προφορικό όσο και στο γραπτό λόγο. Οι μεταβολές αυτές ήταν αναπόφευκτες αφού κάθε ομιλούμενη γλώσσα εξελίσσεται για να μπορεί να εκφράζει τις μεταβαλλόμενες ανθρώπινες σχέσεις και εμπειρίες. Οι μεταβολές αυτές ήταν:

α)	φωνητικές
β)	μορφολογικές
γ)	συντακτικές
δ)	λεξιλογικές

Αδιάψευστη μαρτυρία γι αυτό είναι άλλωστε και οι μεταβολές στο ίδιο το ελληνικό αλφάβητο. Η ελληνική γλώσσα, έχει μεγάλη δυνατότητα επιλογών έκφρασης και σε επίπεδο λεξιλογικό αλλά και σύνταξης. Αποτέλεσμα αυτής της ιδιότητας είναι η παρεχόμενη ευχρηστία και ευελιξία στους χρήστες αυτής της γλώσσας για εκφορά λόγου και έκφρασης, με ειδικότερο σκοπό τη γλωσσική αποτύπωση της συναισθηματικής και ψυχικής ή ψυχολογικής τους φόρτισης, είτε είναι επιθυμητή μια πιο τέλεια ακρίβεια λόγου είτε μια νοηματική ελλειπτικότητα και υπαινικτικότητα (Λασκαράτου, 1999). Γι αυτό ορισμένοι ειδικοί επιστήμονες ισχυρίζονται ότι η χώρα μας ειδικότερα έχει οξύτερο πρόβλημα σε ό,τι αφορά τις γλωσσικές προκαταλήψεις, ενώ αυτό αποδίδεται στην έλλειψη ικανής ανάπτυξης της αντίστοιχης επιστημονικής έρευνας σε σύγκριση με γλώσσες άλλων χωρών της Ευρώπης. Για παράδειγμα η χρήση του ενικού και του πληθυντικού, με χαρακτηριστικό παράδειγμα το λεγόμενο «πληθυντικό ευγενείας», είναι ένα φαινόμενο που θεωρείται πως εισήχθη τους τελευταίους αιώνες στην ελληνική κοινωνία (Γ.Καρβέλας, 2002). Η διαφορετική σηματοδότηση στη γλώσσα κατά την επικοινωνία του λόγου δημιουργεί πλήθος αντιφάσεων και προκαταλήψεων. Για παράδειγμα ο σημασιολογικός προσανατολισμός των επιθέτων μέσα σε μια πρόταση μπορεί να είναι οξύμωρος, προκαλώντας σε μια πιθανή υπολογιστική προσέγγιση συναισθηματικής ανάλυσης λανθασμένα αποτελέσματα π.χ ωμό κρέας (ουδέτερος) / ωμός τρόπος (αρνητικός), ικανή συνθήκη (ουδέτερος) / ικανός υπάλληλος (θετικός).

Επιπρόσθετα, πολλές προκαταλήψεις μας προκύπτουν από τη λανθασμένη σε πολλά σημεία σχολική διδασκαλία, καθώς και από παραδοξότητες που καλλιεργούνται στο κοινωνικό γίγνεσθαι και επιπρόσθετα, από την έλλειψη έμπρακτου ενδιαφέροντος για απεύθυνση των ειδικών αυτών προς το ευρύ κοινό (Πετρούνιας, 2002).

### **2.12.1 Γλωσσική προκατάληψη στη σύνταξη της ελληνικής γλώσσας**

Στο επίπεδο της σύνταξης ο κανόνας στα βιβλία γραμματικής και συντακτικών ορίζει ότι ο αρσενικός τύπος προηγείται του θηλυκού και του ουδέτερου. Για παράδειγμα "άντρες και γυναίκες" είναι περισσότερο συνηθισμένο από ό,τι το αντίστροφο, ενώ η έκφραση "αντρόγυνο" δίνει προτεραιότητα στον αρσενικό τύπο. Χαρακτηριστικές εκφράσεις όπου ο αρσενικός τύπος προηγείται του αντίστοιχου θηλυκού είναι οι ακόλουθες: Ο Τάσος και η Γκόλφω, Ο Αδάμ και η Εύα, Ο Αντώνιος κι η Κλεοπάτρα, Ο βοσκός κι η βοσκοπούλα, Ο κύριος και η κυρία Παπαδοπούλου.

Είναι χαρακτηριστικό ότι μόνο ορισμένες στερεότυπες εκφράσεις μπαίνει ο θηλυκός τύπος πριν από τον αρσενικό, π.χ. "η νύφη κι ο γαμπρός", "η Χιονάτη κι οι επτά νάνοι" και "κυρίες και κύριοι", οι οποίες υποδηλώνουν καταστάσεις στις οποίες η γυναίκα πρωταγωνιστεί λόγω της ομορφιάς της ή της σχέσης της με έναν άντρα.

### **2.12.2 Σημασιολογικά στερεότυπα**

Η αναζήτηση της έννοιας των λέξεων "γυναίκα" και "άντρας" σε λεξικά ελληνικής γλώσσας οδηγεί στην ανακάλυψη ότι αυτές ορίζονται και συνδυάζονται με άλλες που αποτελούν ένα σημασιολογικό πεδίο, το οποίο είναι αρνητικό για τον όρο "γυναίκα" και θετικό για τον όρο "άντρας". Ένα παράδειγμα είναι το Λεξικό Συνωνύμων (Δαγκίτσης, 1986) σύμφωνα με το οποίο συνώνυμα των δύο όρων είναι τα ακόλουθα:

Γυναίκα: η γυναίκα του δρόμου, του πεζοδρομίου, η τιποτένια, η γυναικάρια, η αλόγα, η ανδρογυναίκα, η αρκούδα, η βουβάλα κ. τ. λ.

Άντρας: αντρείος, αντρειωμένος, αντρόκαρδος, γενναϊόκαρδος, γενναίος, γενναϊόψυχος, παλικάρι (Πετρούλα Τσοκαλίδου, 1996).

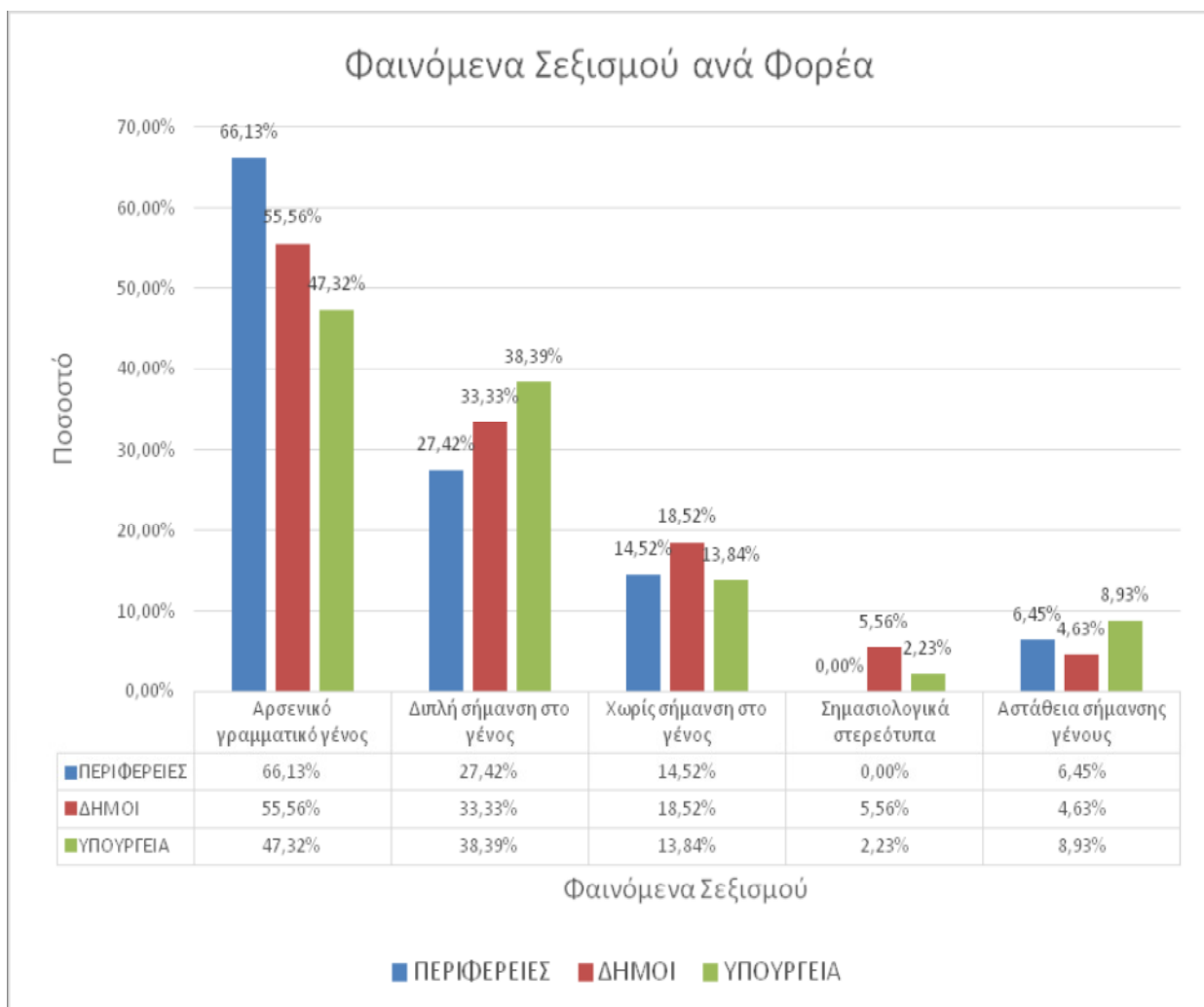
## 2.12.3 Προτάσεις για την αποφυγή του γλωσσικού σεξισμού

Γενίκευση των όποιων «καλών πρακτικών», ώστε η παρουσία των φύλων στις γλωσσικές αναπαραστάσεις να γίνει κατά το δυνατόν ισότιμη σηματοδοτώντας παράλληλα και την ευαισθησία του Πανεπιστημίου σε θέματα γλωσσικού σεξισμού. Δύο είναι οι βασικές κατηγορίες πρακτικών που χρησιμοποιούνται ευρύτατα προς αυτή την κατεύθυνση:

Η προσθήκη του θηλυκού γένους εκεί όπου κατά κανόνα χρησιμοποιείται γενικευτικά μόνο το αρσενικό, για παράδειγμα το (2) αντικαθιστά το (1):

(1) Οι φοιτητές πρέπει να έχουν μαζί τους τη φοιτητική τους ταυτότητα.

(2)\* Οι φοιτητές/τριες πρέπει να έχουν μαζί τους τη φοιτητική τους ταυτότητα.\*



Εικόνα 1: Γράφημα με τα φαινόμενα σεξισμού ανά φορέα (Γενική Γραμματεία Ισότητας)

## 2.13 Το συλλογικό εγκυκλοπαιδικό εγχείρημα Βικιπαίδεια

Η Wikipedia είναι ένα εγχείρημα μη κερδοσκοπικού χαρακτήρα το οποίο οφείλει την ύπαρξή του στο Ίδρυμα Wikimedia Foundation ([https://en.wikipedia.org/wiki/Wikimedia\\_Foundation](https://en.wikipedia.org/wiki/Wikimedia_Foundation)), το οποίο ιδρύθηκε το 2003 με πρωτοβουλία του δημιουργού της Wikipedia Jimmy Wales. Ο Wales μέσω του Wikimedia είχε την πρόθεση να έχει τη δυνατότητα χρηματοδότησης μιας σειράς εγχειρημάτων μη-κερδοσκοπικού χαρακτήρα, ανάμεσα στα οποία βρισκόταν βέβαια και η ίδια η Wikipedia.

Παρόμοιες προσπάθειες με τη Wikipedia οι οποίες εντάσσονται στα πλαίσια του Wikimedia είναι το Wiktionary (Βικι- λεξικό), Wikispecies, Wikiversity και το Wikinews.

Η Wikipedia δεν είναι τίποτα παραπάνω και τίποτα λιγότερο από μια εγκυκλοπαίδεια βασισμένη στο Διαδίκτυο (web- based) και στηριζόμενη σε εργασία εθελοντών οι οποίοι συνεισφέρουν με αφιλοκερδή πρόθεση τη δουλειά τους. Το συλλογικό αυτό έργο έχει προσελκύσει μια μεγάλη σειρά τέτοιων ανθρώπων, οι οποίοι εργάζονται από διάθεση ερασιτεχνική ίσως, αλλά με δρομοδείκτη ένα πρότυπο επαγγελματικό. Το τελευταίο εξασφαλίζεται ακριβώς δια της επίγνωσης των ιδρυτών και των βασικών συντελεστών της Wikipedia ότι ένα τμήμα των άρθρων της είναι εξαιρετικό, αλλά άλλο τμήμα της μπορεί να είναι πολύ χαμηλής ποιότητας (“rubbish”). Αυτήν την επίγνωση την εξωτερικεύει η Wikipedia μέσα από μια σειρά ετικέτες τις οποίες «προσαρτά» στα άρθρα της. Αυτές οι ετικέτες μπορεί να σηματοδοτούν την έλλειψη βιβλιογραφικών πηγών ή αναφορών, τη χρήση υπερβολικά εξειδικευμένης ή τεχνικής γλώσσας (jargon), τέτοια που αδυνατεί να παρακολουθήσει εύκολα το ευρύ κοινό, την μεροληπτική και στερούμενη ουδετερότητας ανάπτυξη του θέματος, τον αμφιλεγόμενο χαρακτήρα κάποιου ζητήματος και πολλά άλλα.

Οι εγγεγραμμένοι χρήστες (registered users) της Wikipedia ακολουθούν μια σταθερά ανοδική πορεία: το 2011 (Brailas, 2011) ήταν ήδη 15.281.656 και το 2014 έφτασαν τα 20 εκατομμύρια (Orlowitz, 2014), ενώ ο βαθμός συμμετοχής των χρηστών της Wikipedia φαίνεται ότι ακολουθεί τον γνωστό κανόνα 90-9-1, όπου το 90% απλώς παρακολουθεί, το 9% έχουν συμβάλει κάποιο υλικό στο εγχείρημα (contributors), ενώ το 1% συνιστά τον λεγόμενο σκληρό πυρήνα του (hard core).

Όσον αφορά τις βασικές αρχές της σύνταξης άρθρων στην Wikipedia, φαίνεται πως είναι τρεις συνολικά. Πρόκειται για την *ουδετερότητα (neutrality)*, την *επαληθευσιμότητα (verifiability)* και τη *μη-πρωτοτυπία της έρευνας* που επικαλείται ο συντάκτης. (<https://www.nngroup.com/articles/participation-inequality/>). Όποιος συντάκτης δεν ακολουθεί αυτές τις βασικές αρχές, που άλλωστε είναι εν πολλοίς και βασικές αρχές επιστημονικά συνταγμένου κειμένου, μπορεί να ανακληθεί στην τάξη από τους υπόλοιπους εγγεγραμμένους ή μη χρήστες οι οποίοι συμβάλλουν υλικό στην Wikipedia.



Όπως αναφέρθηκε και παραπάνω, η Wikipedia είναι ένα συλλογικό εγχείρημα που στηρίζεται και σε ότι αφορά τη λειτουργία, αλλά και σε ό,τι αφορά τη χρηματοδότηση (crowdfunding) στο πλήθος των ανώνυμων ή επώνυμων εθελοντών της, οι οποίοι συμερίζονται το ιδρυτικό της όραμα της ελεύθερης πρόσβασης στο σύνολο της κατακτημένης ανθρώπινης γνώσης.

Στην αγγλική γλώσσα, για να εκφραστεί αυτό χρησιμοποιήθηκε ένας όρος που έγινε ευρύτερα γνωστός από ένα άρθρο που δημοσιεύτηκε στο γνωστό περιοδικό IT «Wired» τον Ιούνιο του 2006 (<http://www.wired.com/2006/06/crowds/>) με τίτλο «*The rise of crowdsourcing*». Έλληνες ερευνητές όπως ο Αλ. Μπράιλας στην παρουσίαση «*Wikipedia: A culture of Participation*» (Brailas, 2016) μεταφράζουν αυτόν τον όρο ως *πληθοπορισμός*. Με βάση την άποψη του Jeff Howe (Howe, 2008), η Wikipedia συνιστά κάτι αντίστοιχο με το ελεύθερο λογισμικό ή λογισμικό ανοικτού κώδικα για την περίπτωση της εγκυκλοπαίδειας. Σε κάθε περίπτωση, το νόημα είναι η απεύθυνση στο πλήθος (και η προαναφερόμενη γιγαντιαία ανταπόκρισή του), το οποίο, σε αντίθεση με τις συμβατικές εγκυκλοπαίδειες, παράγει το αναγνώσιμο υλικό. Ο όρος αυτός θυμίζει άλλωστε και το γνωστό *outsourcing*, μια επιχειρησιακή πρακτική σύμφωνα με την οποία κάποια λειτουργία της παραγωγής ανατίθεται σε κάποιο «εξωτερικό» συνεργαζόμενο μέρος με την επιχείρηση αυτή, με πιο χαρακτηριστικό ίσως παράδειγμα την παραγωγή κώδικα για επιχειρήσεις του δυτικού κόσμου και ιδιαίτερα των ΗΠΑ. Για να επιστρέψουμε στο αναγνώσιμο υλικό, αυτό, εκτεινόμενο ήδη περίπου σε 3.736.024 (Brailas, 2011b) δημοσιευμένα άρθρα στην αγγλική γλώσσα το 2011, ([https://en.wikipedia.org/wiki/Wikipedia:Size\\_of\\_Wikipedia](https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia)) πλέον έχει αγγίξει τα 5,151,587 άρθρα τα οποία, εάν δημοσιεύονταν σε συμβατικούς τόμους μιας εγκυκλοπαίδειας, θα συγκροτούσαν περίπου 1575 τόμους. Μια από τις γνωστότερες συμβατικές εγκυκλοπαίδειες, η Britannica, διαθέτει 32 τόμους. Τα μεγέθη είναι προφανώς ριζικά διαφορετικά, όπως βεβαίως διαφορετικές είναι και οι παράμετροι σύνταξης αυτών των άρθρων.

Η λειτουργία που επιτελεί η Wikipedia είναι πολύ σημαντική για την κοινωνία. Η μέχρι σήμερα εκπονημένη έρευνα (Head & Eisenberg, 2010; Head & Eisenberg, 2009) έχει καταδείξει την μεγάλη έκταση της χρήσης της από φοιτητές, αλλά και από καθηγητές κατά την προετοιμασία του διδακτικού τους υλικού. Άλλωστε η Wikipedia βρίσκεται συνήθως στις πρώτες θέσεις της αναζήτησης με βάση τις δημοφιλέστερες ως σήμερα μηχανές αναζήτησης (google), επομένως χρησιμοποιείται από εκατοντάδες εκατομμύρια χρήστες μηνιαία.

Ενδεικτικά, τον Μάιο του 2015, με βάση τα στοιχεία του Wikimedia Foundation, υπήρχαν 430 εκατομμύρια μοναδικοί επισκέπτες (unique visitors) (<http://reportcard.wmflabs.org/>).



## 2.14 Αντικειμενικότητα και ποιότητα πληροφοριών στη Wikipedia

Στην Wikipedia μπορούν να δημοσιεύσουν άρθρα άνθρωποι από κάθε κατηγορία, από τις περισσότερο έως τις λιγότερο γραμματισμένες. Εκτός από τους εγγεγραμμένους χρήστες, στους οποίους αναφερθήκαμε παραπάνω, υπάρχει η δυνατότητα γραφής ή επεξεργασίας άρθρου (article editing) από οποιονδήποτε χρήστη. Αυτό έχει ως λογική συνέπεια την ενσωμάτωση άρθρων τα οποία εμπεριέχουν πολύ χαμηλής ποιότητας ή πολύ μεγάλης μεροληψίας και χαμηλής επιστημονικής ουδετερότητας υλικό, δηλαδή κάτι που παραβιάζει τις θεμελιώδεις αρχές και της επιστημονικής γραφής αλλά και της ίδιας της Wikipedia. Αυτή η λογική συνέπεια έχει στην πραγματικότητα ήδη προβλεφθεί από τους ιδρυτές της Wikipedia, αλλά και από τους συντάκτες της, οι οποίοι έχουν συγγράψει ένα ειδικό άρθρο για το θέμα της «αξιοπιστίας της Wikipedia» ([https://en.wikipedia.org/wiki/Reliability\\_of\\_Wikipedia](https://en.wikipedia.org/wiki/Reliability_of_Wikipedia)).

Σχετικά με την αντικειμενικότητα της Βικιπαίδεια, έχει διατυπωθεί ένας εμπειρικός γενικός κανόνας ότι «ένα άρθρο είναι τόσο αξιόπιστο όσο ο συντάκτης του». Αυτή η αρχή έχει πράγματι κάποια εγκυρότητα, αλλά δεν είναι τελείως έγκυρη. Αυτός που θα περιηγηθεί στην Βικιπαίδεια θα παρατηρήσει στα περισσότερα άρθρα της μια σειρά από επισημάνσεις οι οποίες έχουν τοποθετηθεί από διάφορους χρήστες της, καθώς επίσης και μακροσκελείς συζητήσεις οι οποίες αφορούν «αμφιλεγόμενα»(controversial) σημεία άρθρων της. Αυτές οι συζητήσεις, οι αντιπαραθέσεις επιχειρημάτων, αλλά και οι επισημάνσεις (ελλιπής τεκμηρίωση, μεροληπτική ανάπτυξη του θέματος κ.λπ..) είναι μια ένδειξη του πώς λειτουργεί η συλλογικότητα ενός τέτοιου εγχειρήματος σε αντίστιξη με την υπευθυνότητα του εκάστοτε συντάκτη.

Το περιβάλλον wiki συνίσταται στα τέσσερα παρακάτω στοιχεία (Brailas, 2016):

- 1) **Επεξεργασία** από τον οποιοδήποτε (*Edit Button*)
- 2) **Σελίδες ιστορικού** με όλες τις αλλαγές
- 3) **Σελίδα συζήτησης** λήμματος για διάλογο
- 4) **Διάφορες σελίδες συζήτησης** για την κοινότητα

Όπως επιγραμματικά αναφέρθηκε στην προηγούμενη ενότητα, ο συγγραφέας Jeff Howe στο βιβλίο του για τον πληθοπορισμό (Howe, 2008) εξέφρασε την άποψη ότι η Βικιπαίδεια καταδεικνύει το γεγονός ότι η αρχή αυτή δεν βρίσκει την αποκλειστική εφαρμογή της στο λογισμικό ανοικτού κώδικα και μάλιστα ότι κατέστησε το *crowdsourcing* δημοφιλές, κάτι που υποβοηθήθηκε σημαντικά από τη συνεργατική φύση του διαδικτύου.

Για το λογισμικό αυτό έχει γραφτεί χαρακτηριστικά (Linus' Law):

«πολλά μάτια κάνουν κάθε σφάλμα ρηχό» (δηλ. εύκολα ανιχνεύσιμο). Συμπερασματικά, θα μπορούσε να ειπωθεί ότι η Βικιπαίδεια δεν διαθέτει a priori αξιόπιστα ή απροκατάληπτα άρθρα, όχι λόγω των αρχών που την διέπουν, αλλά ακριβώς λόγω του χαρακτήρα της που επιτρέπει, θεωρητικά τουλάχιστον, σε οποιονδήποτε να επεξεργαστεί ή να συγγράψει ένα άρθρο.

Οι συντάκτες της Βικιπαίδεια πιθανότατα θα επισημάνουν τα προβληματικά άρθρα και τις εστίες των προβλημάτων αυτών, όμως δεν υπάρχει άλλη δικλείδα ασφαλείας άλλη από την συλλογική αυτενέργεια των χρηστών της. Έτσι, είναι ο αναγνώστης της Βικιπαίδεια που θα πρέπει να έχει ο ίδιος την γνώση και την κατάρτιση ως προς το πώς να τα αξιολογήσει. Αξιοσημείωτο είναι ότι μια πολύ εύστοχη ανάπτυξη για το πώς πρέπει να γίνεται η αξιολόγηση των γραπτών σε ότι αφορά την αξιοπιστία που παρουσιάζουν έχει πραγματοποιηθεί ήδη σε ειδικό άρθρο της Βικιπαίδεια ([https://en.wikipedia.org/wiki/Wikipedia:Identifying\\_reliable\\_sources](https://en.wikipedia.org/wiki/Wikipedia:Identifying_reliable_sources)).

## 2.15 Μελέτες έμφυλης γλωσσικής προκατάληψης

Στο τρίτομο σύγγραμμα Bussmann, H. & M. Hellinger. 2001-2003. *Gender Across Languages*. 3 τόμ. Amsterdam: John Benjamins, διερευνάται μια σειρά από τριάντα στον αριθμό διαφορετικές γλώσσες και καταγράφεται με συστηματικό τρόπο η κατηγοριοποίηση του κοινωνικού φύλου (gender) σε όλα αυτά τα διαφορετικά υπόβαθρα (γενετικά, κοινωνικοπολιτισμικά και γενετικά) που αυτές παρουσιάζουν. Βαρύνουσα σημασία δίνεται στις ιδιότητες που διαθέτει η κάθε γλώσσα και οι οποίες έχουν αντίκτυπο στην σχέση που απασχολεί και αυτήν την εργασία, δηλαδή αυτήν μεταξύ γλώσσας και κοινωνικού φύλου (<https://benjamins.com/#catalog/books/impact.10/main>).

Το δεύτερο βιβλίο (Eckert P. & S. McConnell-Ginet. 2003. *Gender and Language*. Cambridge: Cambridge University Press) είναι μια εισαγωγική μελέτη από δύο θεωρούμενους ως κορυφαίους ειδικούς στον τομέα και καλύπτει τα θέματα που αφορούν το κοινωνικό φύλο (gender) όπως και τον τρόπο που μπορεί να επιδράσει η γλωσσολογία στην δόμηση του κοινωνικού του νοήματος ([https://books.google.gr/books?id=JnZGzMkBaqkC&redir\\_esc=y](https://books.google.gr/books?id=JnZGzMkBaqkC&redir_esc=y)). Άλλα θέματα με τα οποία καταπιάνεται είναι οι σεξουαλικές ταυτότητες και το γενικότερο ζήτημα της αλλαγής, στην συνάφειά του με τις γλωσσολογικές και κοινωνικές αρχές. (<http://www.cambridge.org/us/academic/subjects/languages-linguistics/sociolinguistics/language-and-gender-2nd-edition?format=PB>).

Η τρίτη μελέτη (Holmes, J. & M. Meyerhoff. 2003. *Handbook of Language and Gender*. Οξφόρδη: Blackwell) είναι μια συλλογή άρθρων γλωσσολογικού όπως και κοινωνιογλωσσολογικού ενδιαφέροντος, με ένα ευρύ πεδίο θεματικών και με διεπιστημονικό χαρακτήρα. Οι επιστημονικές περιοχές που καλύπτονται από τα άρθρα αυτά είναι αυτή της ανθρωπολογίας, της κοινωνικής ψυχολογίας, της οργανωτικής επικοινωνίας (organizational communication), της κοινωνιολογίας, των σπουδών πολιτισμού και των σπουδών φύλου με εστίαση στις σύγχρονες εξελίξεις που αφορούν τον βασικό άξονα κοινωνικό φύλο και γλώσσα.

([https://is.muni.cz/el/1423/podzim2012/SAN230/um/the\\_handbook\\_of\\_language\\_and\\_gender.pdf](https://is.muni.cz/el/1423/podzim2012/SAN230/um/the_handbook_of_language_and_gender.pdf)).

Ένα από τα πιο θεμελιακά σημεία των μελετών αυτών είναι η ανάδειξη του λεγόμενου κοινωνικού φύλου σε αντιπαραβολή με το βιολογικό φύλο (sex / gender). Το κοινωνικό φύλο συνδέεται με το ιστορικό συγκείμενο και αυτό με τη σειρά του εκφράζεται από μια ορισμένη, στερεοτυπική γλώσσα η οποία συχνά εμπεδώνει τις προκαταλήψεις, αναπαράγοντάς τις μέσα στην πορεία του χρόνου. Τα στερεότυπα συνδέονται, όπως έχει προαναφερθεί, και με την εκφορά λόγου, και ιδιαίτερα τα στερεότυπα που είναι εθνικής ή φυλετικής υφής. Η άρθρωση λόγου, τα συντακτικά, φωνητικά, δομικά, εννοιολογικά χαρακτηριστικά του και η όλη ιδιοσυγκρασία του όπως αυτή εκδηλώνεται με παροιμιακές και άλλες ιδιότυπες εκφράσεις υποβοηθούν στην συγκρότηση και την διαίωνιση των στερεοτύπων (Σακκάτου, 2011). Η γλώσσα θα λέγαμε λοιπόν πως εκφράζει τα στερεότυπα τα οποία υπάρχουν σε μια λανθάνουσα κατάσταση (Κωστούλα - Μακράκη, 2001).

## 2.16 Προηγούμενες έρευνες

Ξεκινώντας το κεφάλαιο, τίθεται το ερώτημα εάν έχουν ήδη υλοποιηθεί στην υπάρχουσα έρευνα απόπειρες καταμέτρησης των γλωσσικών προκαταλήψεων σε ιστοχώρους όπως η Βικιπαίδεια και, ιδιαίτερα, στην ίδια την Βικιπαίδεια, εφόσον και σε αυτή την διατριβή θα υλοποιηθεί μία τέτοια τεχνική προσπάθεια. Μία από τις σημαντικότερες τέτοιες προσπάθειες που εντοπίστηκαν στην πορεία της συγγραφής αυτής της εργασίας φαίνεται πως υπάρχει στο άρθρο:

*"It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia"* (Wagner, Cl. et al., 2015), όπου και πραγματοποιείται μια προσπάθεια ανίχνευσης και καταμέτρησης ειδικά ανάμεσα στα φύλα (<http://arxiv.org/abs/1501.06307>). Η μεθοδολογία συνίσταται ως ένα βαθμό στην εύρεση των περισσότερο εμφανιζόμενων λέξεων σε βιογραφίες που αφορούν ανδρικά και γυναικεία πρόσωπα. Αφού ευρεθούν αυτές, γίνεται μια αποτίμηση και ανακαλύπτεται ότι υφίσταται γλωσσική προκατάληψη σε ότι αφορά το "κοινωνικό φύλο" (gender). Αυτό προκύπτει από την φύση των περισσότερο εμφανιζόμενων λέξεων, οι οποίες, πιο συγκεκριμένα, φαίνεται ότι αφορούν ρόλους οι οποίοι δεν αντιστοιχίζονται σε αυτούς που η νεωτερική κοινωνία έχει αναθέσει στην γυναίκα -υπό την επίδραση και του φεμινιστικού κινήματος.

Συνολικά, η διερεύνηση που υλοποιείται αφορά τέσσερα είδη μεροληψίας (bias):

κάλυψης	(coveragebias)
δομική	(structuralbias)
λεξιλογική	(lexicalbias)
ορατότητας	(visibilitybias)

Μια δεύτερη έρευνα, συγγενικότερη προς την παρούσα εργασία και ιδιαίτερα από τεχνική άποψη, είναι η εξής: "*Linguistic Models for Analyzing and Detecting Biased Language*" ( <https://web.stanford.edu/~jurafsky/pubs/neutrality.pdf>) (Recasens, M., C. Danescu-Niculescu-Mizil&DanJurafsky,2013).

Το πρώτο βήμα στο οποίο προχωρούν οι ερευνητές του Stanford, που συντάσσουν το άρθρο αυτό, είναι η εύρεση του "*NPOV Corpus*" (Neutral Point of View), δηλαδή του σώματος των άρθρων τα οποία έχουν βρεθεί σε σελίδα "*dispute*" (αντιπαράθεσης) σε ό,τι αφορά το NPOV. Οι σελίδες αυτές έχουν τοποθετηθεί από την Βικιπαίδεια σε αυτόν τον υπερσύνδεσμο (URL): [https://en.wikipedia.org/wiki/Category:All NPOV disputes](https://en.wikipedia.org/wiki/Category:All_NPOV_disputes).

Το δεύτερο βήμα στο οποίο προχωρούν οι εν λόγω ερευνητές είναι να ξεχωριστούν από αυτά τα άρθρα εκείνα τα οποία έχουν τουλάχιστον μια τέτοια αιτιολόγησης "*επεξεργασία*" (edit), πριν πραγματοποιηθεί η εν λόγω επεξεργασία.

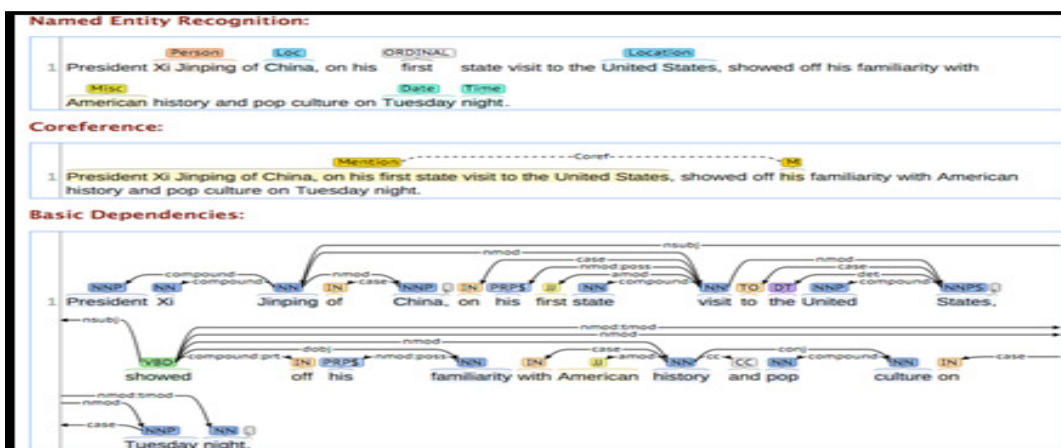
Οι ερευνητές αυτοί έχουν και ένα πρόσθετο κριτήριο, την ύπαρξη μιας τουλάχιστον "αυθεντικής" επιπλέον λέξης για την επιλογή των προτάσεων που λαμβάνουν. Με αυτόν τον τρόπο μπορεί να ληφθεί ένα σύνολο λέξεων οι οποίες φαίνεται ότι προσθέτουν προκατάληψη στο κείμενο και, παραπέρα, να δημιουργηθεί ένα "*λεξικό προκατάληψης*". Εφόσον ληφθούν αυτές οι λέξεις, και για επαλήθευση της μεθόδου, δρομολογείται ένα είδος δοκιμής της προβλεπτικής ικανότητας της παραπάνω μεθόδου, μέσω κάποιας στατιστικής παλινδρόμησης (regression).

Ουσιαστικά χρησιμοποιούνται προτάσεις με λέξεις που είναι γνωστό ότι είναι στερεοτυπικές/μεροληπτικές και από αυτό συνάγεται ποιο ποσοστό είναι «επιτυχές». Τέλος, το εργαλείο επεξεργασίας της πληροφορίας, εφόσον αυτή έχει ανακτηθεί, είναι το "*StanfordCoreNLP*" (<http://stanfordnlp.github.io/CoreNLP/>).

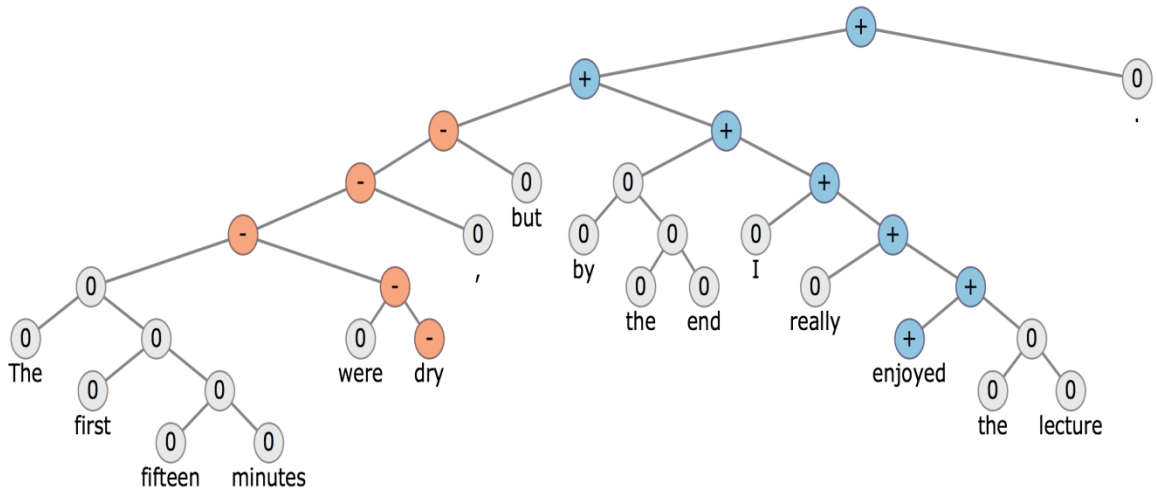
Ένα τρίτο άρθρο σχετικό με την εργασία αυτή το οποίο έχει συνταχθεί το 2011 (Reagle, J. - L. Rhue, 2011) από τους Joseph Reagle και Lauren Rhue και φέρει τον τίτλο «Προκατάληψη φύλου στην Βικιπαίδεια και την Britannica»

( <http://ijoc.org/index.php/ijoc/article/view/777> ).

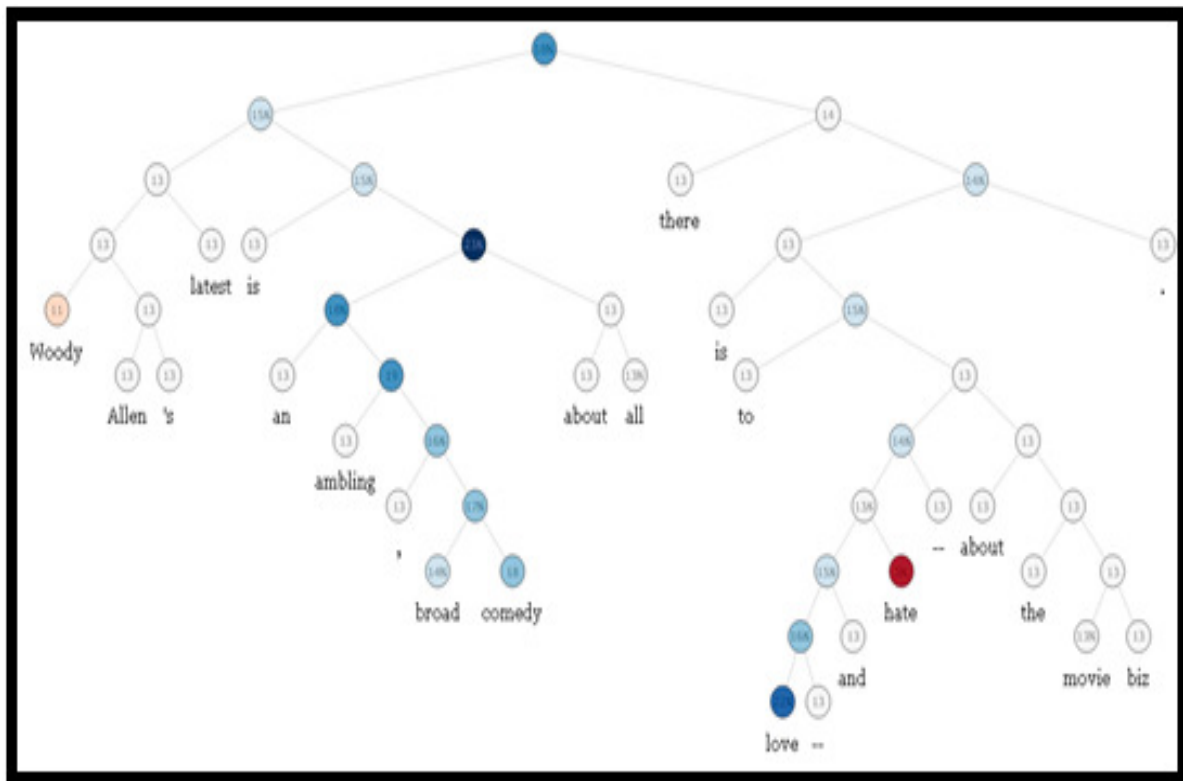
Αυτοί οι δύο πανεπιστημιακοί ερευνητές προχώρησαν σε μία εις βάθος σύγκριση βιογραφιών που ανέκτησαν από τις δύο εγκυκλοπαίδειες, τόσο μεταξύ τους όσο και από ένα σύνολο άλλων έξι πηγών (όπως το περιοδικό “Time”). Η μέθοδος που ακολουθείται από τους δύο ερευνητές είναι αρκετά περίπλοκη και σύνθετη και περιλαμβάνει προγραμματισμό (μέσω της γλώσσας Python), υπολογισμό στατιστικών στοιχείων, πραγματοποίηση στατιστικών παλινδρομήσεων. Αναζήτησαν την κάλυψη ή διασπορά σε τοπικό επίπεδο των Wikipedians, το ποσοστό ελλειπουσών βιογραφιών σημαντικών προσωπικοτήτων ανάμεσα στα φύλα και ανάλυση της έκτασης των άρθρων. Επίσης οι στατιστικές παλινδρομήσεις επιχειρούν μεταξύ άλλων να εκτιμήσουν, δηλαδή να «προβλέψουν» στατιστικά, την έκταση ενός άρθρου με βάση το φύλο, αλλά και με βάση άλλες παραμέτρους. Τα κύρια ευρήματα των ερευνητών είναι ότι η Βικιπαίδεια παρουσιάζει καλύτερες επιδόσεις όσον αφορά τις βιογραφίες σε σχέση με την εκπροσώπηση του φύλου, την κάλυψη των προσωπικοτήτων και έκταση του άρθρου, ενώ το μόνο σημείο στο οποίο υστερεί είναι οι «ελλείπουσες» βιογραφίες γυναικών, οι οποίες είναι σχετικά περισσότερες στην Βικιπαίδεια από ότι στην Britannica. Αυτό το τελευταίο αποδίδεται και στον περισσότερο «ισορροπημένο» χαρακτήρα της Britannica, που προκύπτει από την συστηματική και προπαρασκευασμένη της υφή.



Εικόνα 2: Stanford's NLP tools –the name entity recognition



Εικόνα 3: A parse tree of sentiment analysis. Assign a positive/negative sentiment to each part.



Εικόνα 4: Sentiment tree, in <http://nlp.stanford.edu/sentiment/treebank.html>





Εικόνα 5: Sentiment tree, in <http://nlp.stanford.edu/sentiment/treebank.html>

## Harvard IV-4 categories:

( <http://www.wjh.harvard.edu/~inquirer/homecat.htm>)

Three semantic dimensions.

These categories reflect Charles Osgood's semantic differential findings regarding basic language universals. An earlier version had three different "intensity" levels for each category, but these were combined. A word may be more than one dimension, if appropriate. For example, "celebration" in the Harvard dictionary is [PositivPstvAffilActiveRitual](#)

[Pstv](#) 1045 positive words, an earlier version of [Positiv](#).

A subset of 557 words are also tagged [Affil](#) for words indicating affiliation or supportiveness.

[Ngstv](#) 1160 negative words, an earlier version of [Negativ](#).

A subset of 833 words are also tagged [Hostile](#) for words indicating an attitude or concern with hostility or aggressiveness.

[Strong](#) 1902 words implying strength.

A subset of 689 words are tagged [Power](#), indicating a concern with power, control or authority.

[Weak](#) 755 words implying weakness.

Εικόνα 6: Harvard IV-4 categories

Το αρχείο semantic analysis περιέχει λέξεις που ενέχουν γλωσσικής προκατάληψης:

[www.wjh.harvard.edu/~inquirer/spreadsheet\\_guide.htm](http://www.wjh.harvard.edu/~inquirer/spreadsheet_guide.htm)

```

import algorithmia.*;
// import com.algorithmia.*;
// import com.algorithmia.algo.*;
// import com.algorithmia.data.*;
import com.google.gson.*;
import java.util.*;

import edu.stanford.nlp.ling.CoreAnnotations;
import edu.stanford.nlp.neural.rnn.RNNCoreAnnotations;
import edu.stanford.nlp.pipeline.Annotation;
import edu.stanford.nlp.pipeline.StanfordCoreNLP;
import edu.stanford.nlp.sentiment.SentimentCoreAnnotations;
import edu.stanford.nlp.trees.Tree;
import edu.stanford.nlp.util.CoreMap;

// Depends on: edu.stanford.nlp / stanford-corenlp / 3.5.2
// Depends on:
//   <dependency conf="default" rev="3.5.2" name="stanford-
corenlp-models" org="edu.stanford.nlp"
maven:classifier="models">
//       <artifact name="stanford-corenlp-models" type="jar"
url="http://repo1.maven.org/maven2/edu/stanford/nlp/stanford
-corenlp/3.4/stanford-corenlp-3.4-models.jar"/>
// </dependency>

/**
 * Performs sentiment analysis
 */
public class SentimentAnalysis extends Algorithm {
    public SentimentAnalysis() {
        setCachable(true);
    }
    private static StanfordCoreNLP pipeline = null;
    // The input and output of apply() automatically turns into JSON
    public Integer apply(String input) throws Exception {
        if(input == null) {
            return null;
        }
        // Initialize stanford corenlp pipeline for sentiment
        if (pipeline == null) {
            Properties props = new Properties();
            props.setProperty("annotators", "tokenize, ssplit, parse,
sentiment");
            // props.setProperty("annotators", "tokenize, ssplit, pos,

```

```

lemma, ner, parse, dcoref, sentiment");
    pipeline = new StanfordCoreNLP(props);
}

// Trim long lines
String line;
if (input.length() < 1000) {
    line = input;
} else {
    line = input.substring(0,1000);
}

// Compute sentiment
int mainSentiment = 0;
HashMap<Integer,Integer> sentiments = new
HashMap<Integer,Integer>();
if(line != null && line.length() > 0) {
    int longest = 0;
    Annotation annotation = pipeline.process(line);

    // Iterate over each sentence
    for(CoreMap sentence :
annotation.get(CoreAnnotations.SentencesAnnotation.class)) {

        // Get the sentiment for this sentence
        Tree tree =
sentence.get(SentimentCoreAnnotations.SentimentAnnotatedTree.
class);
        int sentiment=
RNNCoreAnnotations.getPredictedClass(tree);

        // Let each sentence vote for the most popular sentiment
        if (sentiments.containsKey(sentiment)) {
            int curVotes = sentiments.get(sentiment);
            sentiments.put(sentiment,curVotes+1);
        } else {
            sentiments.put(sentiment,1);
        }
        //String partText = sentence.toString();
        //if (partText.length() > longest) {
        //    mainSentiment = sentiment;
        //    longest = partText.length();
        //}

    }
    return returnWinner(sentiments);
}

return mainSentiment;
}

```

```
public int returnWinner(HashMap<Integer,Integer> votes) {  
    int maxValue = 0;  
    int maxKey = 0;  
    for(Integer key : votes.keySet()) {  
        int value = votes.get(key);  
        if(value > maxValue) {  
            maxValue = value;  
            maxKey = key;  
        }  
    }  
    return maxKey;  
}
```

Εικόνα 7: Package algorithmia.nlp.SentimentAnalysis

# Κεφάλαιο 3

## Μέθοδος

### 3.1 Η περιγραφή ερευνητικού προβλήματος

Στην εργασία αυτή, ο αντικειμενικός στόχος είναι η εύρεση μιας μεθόδου προκειμένου να ανιχνευθεί η στερεοτυπική, προκατειλημμένη γλώσσα μελετώντας περιπτώσεις βιογραφιών οι οποίες είναι δημοσιευμένες στην Βικιπαίδεια (Wikipedia). Το πρώτο ερώτημα που τίθεται αφορά τον τρόπο αναγνώρισης της προκατειλημμένης γλώσσας. Σε προηγούμενες έρευνες, που κυρίως έχουν διεξαχθεί από ακαδημαϊκούς, έχει προταθεί μία σειρά από διαφορετικές μεθοδολογίες, όπως αναπτύχθηκε στην πρώτη παράγραφο αυτού του κεφαλαίου.

Η μεθοδολογία που χρησιμοποιήθηκε στηρίχτηκε στα εξής βήματα:

- i. Χρησιμοποιήθηκε συνδυασμός ποσοτικής και ποιοτικής προσέγγισης.
- ii. Αξιοποιήθηκαν ειδικά Λεξικά, τα οποία διαθέτουν λέξεις που ενέχουν κατά την τρέχουσα χρήση τους προκατάληψη. Οι λέξεις αυτές έχουν θετική ή αρνητική φόρτιση, ενώ τα λεξικά που θα αξιοποιηθούν εδώ πρέπει να καλύπτουν όλες τις υπό μελέτη γλώσσες, συγκεκριμένα επί του προκειμένου τα ελληνικά και τα αγγλικά.
- iii. Διερευνήθηκε η επιλογή και πιθανώς επέκταση υπάρχοντος λογισμικού που θα θεωρηθεί κατάλληλο για εντοπισμό πανομοιότυπων λέξεων σε υπερκείμενα στην Βικιπαίδεια;
- iv. Διερευνήθηκε η επιλογή και πιθανώς επέκταση κατάλληλου λογισμικού με το οποίο αυτόματα θα συνέκρινε τις δύο (αλφαβητικά ταξινομημένες) λίστες λέξεων και θα εντόπιζε όλες τις κοινές λέξεις που υπάρχουν μεταξύ τους
- v. Έγινε τελικός έλεγχος και καταμέτρηση.

## 3.2 Αιτιολόγηση της συγκεκριμένης επιλογής (πλεονεκτήματα και μειονεκτήματα)

Αιτιολογώντας την επιλογή του θέματος αυτής της διατριβής μπορούμε να αναφέρουμε αρκετές θετικές εκτιμήσεις για την τελική αποτύπωση αποτελεσμάτων καθώς και την ποιοτική και ποσοτική τους επεξεργασία. Τα πλεονεκτήματα που έχει είναι ότι προσφέρει ευρύ πεδίο έρευνας το οποίο καλύπτει αρκετές επιστημονικές περιοχές. Η πληροφορική θα επεξεργαστεί το γραπτό λόγο και θα αποτυπώσει στερεοτυπικές συμπεριφορές που εμφανίζονται σε αυτόν. Όμως μειονέκτημα μπορεί να παραμένει το μικρό και περιορισμένο δείγμα που αναγκαζόμαστε να επιλέξουμε και το οποίο θα δώσει αποτελέσματα για συγκεκριμένες κατηγορίες. Περιορίζεται έτσι η γενίκευση και αναγκαστικά κατηγοριοποιείται ο λόγος. Μετά τη διασαφήνιση των προηγούμενων, ένα ζήτημα που τίθεται αφορά την αναγνώριση, την ανίχνευση της γλώσσας που ενέχει προκαταλήψεις στο σώμα των κειμένων των βιογραφιών της Wikipedia. Το ερώτημα το οποίο ήμασταν επιφορτισμένοι να απαντήσουμε αφορά το εάν και πώς μπορεί να συγκροτηθεί μια συγκεκριμένη μεθοδολογία ανίχνευσής της που να μπορεί να εφαρμοστεί σε κάθε περίπτωση.

## 3.3 Επιλογή δείγματος

Το δείγμα που έχει χρησιμοποιηθεί αφορά τον έλεγχο της γλωσσικής προκατάληψης με μελέτη περίπτωσης. Η περίπτωση που επιλέχθηκε να ελεγχθεί είναι βιογραφίες που κατανέμονται σε μια σειρά υποκατηγοριών. Συγκεκριμένα, αναλύθηκαν οι παρακάτω βιογραφίες :

- Βιογραφίες πολιτικών ανδρών
- Βιογραφίες πολιτικών γυναικών
- Βιογραφίες ηθοποιών ανδρών
- Βιογραφίες ηθοποιών γυναικών

Από κάθε μία από τις ανωτέρω υποκατηγορίες έχει πραγματοποιηθεί μια διαλογή από δεκατρείς (13) βιογραφίες από ευρύτερα γνωστά πρόσωπα που ανήκουν σε αυτές.

Ας σημειωθεί επίσης ότι η έρευνα στο τμήμα αυτό εντοπίζεται σε δύο διαφορετικές γλώσσες: ελληνικά και αγγλικά. Συνεπώς, έχουν επιλεγεί ισάριθμες βιογραφίες από κάθε μία από τις παραπάνω υποκατηγορίες. Τα προηγούμενα συνεπάγονται ότι συνολικά η ανάλυση γλωσσικής προκατάληψης διεξάγεται σε ένα σύνολο 104 βιογραφιών. Τα αποτελέσματα τα οποία συγκεντρώνονται αφορούν το πλήθος των λέξεων οι οποίες ενέχουν γλωσσική προκατάληψη, σε κάποια από τις δύο επιλεγμένες γλώσσες. Ένα δεύτερο αποτέλεσμα το οποίο συγκεντρώνεται αφορά τις ετικέτες NPOV (Neutral Point Of View). Όταν ένα άρθρο βιογραφίας φέρει την ετικέτα NPOV, αυτό είναι ισοδύναμη πρόταση με το ότι κάποιος ή κάποιιο συντάκτης της Wikipedia θεώρησαν ότι το άρθρο αυτό φαίνεται ότι έχει προβλήματα με την ουδετερότητά του και επιπλέον ότι ως την στιγμή που η βιογραφία ανακτήθηκε για τους σκοπούς της εργασίας αυτής, το εν λόγω πρόβλημα δεν είχε ακόμη επιλυθεί. Με άλλη διατύπωση, οι προβληματισμοί του αρχικού συντάκτη που είχε τοποθετήσει την ετικέτα, όπως αυτοί εξηγήθηκαν σε κατάλληλο χώρο συζήτησης της Wikipedia (NPOV:Dispute) δεν έχουν επιλυθεί, ώστε να δρομολογηθεί η αφαίρεση της ετικέτας αυτής. ([https://en.wikipedia.org/wiki/Wikipedia:Tagging\\_pages\\_for\\_problems#Removing\\_tags](https://en.wikipedia.org/wiki/Wikipedia:Tagging_pages_for_problems#Removing_tags)). Συμπερασματικά, εάν πράγματι υπάρχει ετικέτα NPOV σε μια βιογραφία της Wikipedia, αυτό σημαίνει ότι έχει οικοδομηθεί ένα status συναίνεσης (Consensus) (<https://en.wikipedia.org/wiki/Wikipedia:Consensus>) μεταξύ των συντακτών της ως προς την κατηγοριοποίηση αυτή.

### 3.4 Ερευνητικό Υλικό – Εργαλεία

Στην παρούσα εργασία, αξιοποιήθηκε η ήδη συσσωρευμένη γνώση όπως έχει αποκρυσταλλωθεί σε ειδικά Λεξικά, τα οποία διαθέτουν λέξεις οι οποίες θεωρούνται ότι ενέχουν κατά την τρέχουσα χρήση τους προκατάληψη. Οι λέξεις αυτές μπορεί να έχουν θετική ή αρνητική φόρτιση, ενώ τα λεξικά που θα αξιοποιηθούν εδώ πρέπει να καλύπτουν όλες τις υπό μελέτη γλώσσες, συγκεκριμένα επί του προκειμένου τα ελληνικά και τα αγγλικά.

Έτσι, στα ελληνικά υπάρχει ένα τέτοιο λεξικό που έχει καταρτισθεί από Έλληνες ερευνητές, το «*Greek Sentiment Lexicon*». Το λεξικό αυτό περιλαμβάνει ως όλον 2315 λέξεις, ενώ είναι διαθέσιμο στο Διαδίκτυο σε μορφή .tsv στην γνωστή σελίδα github.



Από την ίδια σελίδα είναι ανακτήσιμο και το δεύτερο λεξικό, (<https://github.com/woorm/afinn-111> ) το οποίο καλύπτει την προκατειλημμένη γλώσσα στην αγγλική και το οποίο επιλέχθηκε κατά πρώτον για την πληρότητα, κατά δεύτερον για την ποσοτική αναλογία ως προς το “*Greek Sentiment Lexicon*” και κατά τρίτον για την υψηλή ποιότητά του. Το λεξικό αυτό στην πληρέστερη εκδοχή του έχει την ονομασία *AFINN-111*, συμπεριλαμβάνει 2477 λέξεις, ενώ η λεξικογράφηση και έκδοσή του έχει πραγματοποιηθεί εκ μέρους του Τεχνικού Πανεπιστημίου της Δανίας ([http://www2.imm.dtu.dk/pubdb/views/publication\\_details.php?id=6010](http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010)).

Επομένως διαθέτοντας αυτά τα λεξικά, υπολείπεται η σύγκριση και εντοπισμός πανομοιότυπων λέξεων μεταξύ του λεξικού και της βιογραφίας στην Βικιπαίδεια.

Η αναζήτησή μας κατέληξε στην χρήση ορισμένων λογισμικών, τα οποία μπορούν να φέρουν εις πέρας, συνεργαζόμενα, ένα τέτοιο στόχο. Τα λογισμικά αυτά μπορούν να ευρεθούν και χρησιμοποιηθούν και online και είναι τα ακόλουθα:

- 1) Wordlist Maker v 1.2 (<http://design215.com/toolbox/wordlist.php>)
- 2) Unique Word List Text Wizard (<http://tools.perceptus.ca/text-wiz.php?ops=4>)

Στα δύο αυτά εργαλεία εισάγεται το κείμενο της βιογραφίας, έτσι ώστε να προκύψει μια λίστα από τις αλφαβητικά ταξινομημένες «μοναδικές λέξεις» (*unique words*) που περιέχονται σε αυτήν. Αφού έχει ολοκληρωθεί αυτό το βήμα, αυτό που απαιτείται είναι ένα εργαλείο που μπορεί αυτόματα να συγκρίνει τις δύο (αλφαβητικά ταξινομημένες) λίστες λέξεων και να εντοπίσει όλες τις κοινές λέξεις που υπάρχουν μεταξύ τους. Αυτήν την εργασία μπορεί να την πραγματοποιήσει το πολύ αξιόλογο λογισμικό *Notepad++* (<https://notepad-plus-plus.org/download/v6.9.1.html>). Ένα από τα πιο εξειδικευμένα πρόσθετα (*plugins*) που παρέχει το λογισμικό αυτό είναι το plugin *Compare*, το οποίο κάθε κάτοχός του μπορεί να εγκαταστήσει αναζητώντας το στην λίστα με τα Πρόσθετα από το κύριο μενού του λογισμικού. Έτσι, λοιπόν, σε κάθε περίπτωση, αρκεί να έχουν ανοιχθεί το αντίστοιχο λεξικό και η αλφαβητικά ταξινομημένη λίστα των μοναδικών λέξεων της βιογραφίας, προκειμένου μέσα από το γραφικό περιβάλλον με επιλογή της εντολής *Compare* να ανιχνευθούν όλες οι λέξεις της βιογραφίας που υπάρχουν στο λεξικό. Μετά από αυτά έπεται ένας τελικός έλεγχος και καταμέτρηση, προκειμένου να καταρτισθούν οι πίνακες οι οποίοι θα παρουσιαστούν στο επόμενο κεφάλαιο.

Ο τελικός έλεγχος που μόλις μνημονεύθηκε είναι απαραίτητος, ώστε να μην παραλειφθεί ή, αντιθέτως, συμπεριληφθεί κάποια λέξη που δεν θα έπρεπε. Αυτό που ενδιαφέρει είναι η λέξη που υπάρχει τόσο στο λεξικό όσο και στο κείμενο να μην επέχει άλλη θέση από συντακτικής άποψης στο σώμα της βιογραφίας, δηλαδή να μην συνιστά άλλο «μέρος του λόγου» (ουσιαστικό αντί για επίθετο κ.ο.κ.).

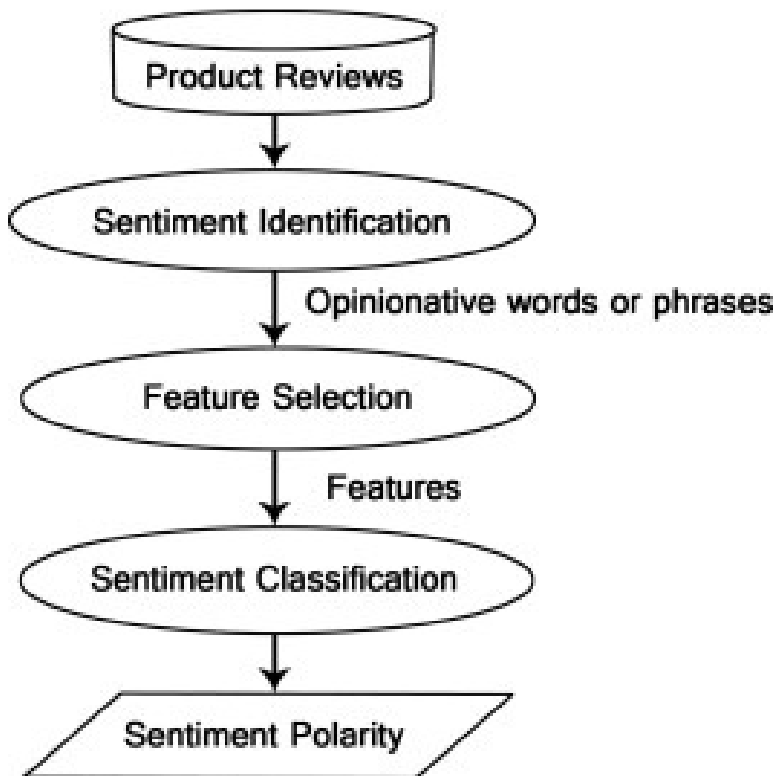
# Κεφάλαιο 4

## Αποτελέσματα

### 4.1 Εξαγωγή πληροφορίας

Ορίζεται ως η αυτόματη εξαγωγή δομημένης πληροφορίας (structured Information) από αδόμητο κείμενο, κυρίως σε φυσική γλώσσα. Η εξαγωγή πληροφορίας καλύπτει μια ευρεία ερευνητική περιοχή, από τον απλό εντοπισμό στοιχείων από ιστοσελίδες με χρήση προτύπων (patterns) και κανονικών γραμματικών (regular grammars), μέχρι την σημασιολογική ανάλυση της γλώσσας για εξαγωγή νοήματος και εννοιών, όπως οι ερευνητικές περιοχές της αποσαφήνισης εννοιών λέξεων (word sense disambiguation) ή της ανάλυσης συναισθήματος (sentiment analysis).

(<http://www.sciencedirect.com/science/article/pii/S2090447914000550>)



Εικόνα 8: Sentiment analysis process on product reviews

## 4.2 Δέντρα απόφασης/ταξινόμησης

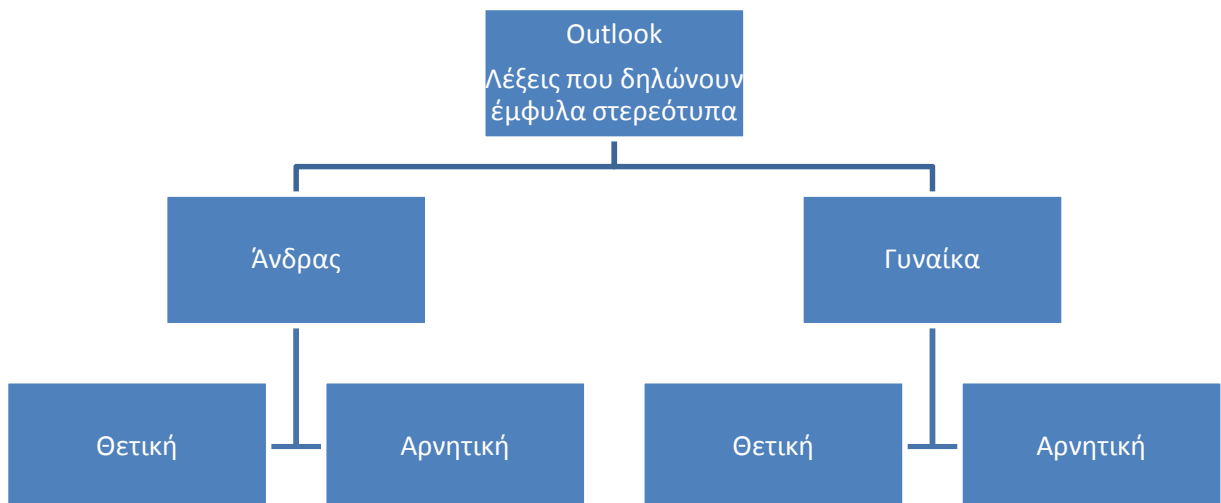
Τα «Δένδρα Απόφασης/ταξινόμησης» (*Decision / Classification trees*) αποτελούν ένα είδος ταξινομητή είναι ένα εργαλείο για ταξινόμηση δεδομένων, αλλά και για την πρόβλεψη σημαντικών αποτελεσμάτων. Η γενική ιδέα στα Δένδρα Απόφασης είναι η εξής:

Αρχικά, βασική προϋπόθεση αποτελεί το ότι κάθε δείγμα (objects ή cases) του data set πρέπει να μπορεί να εκφράζεται ως μία συλλογή από τα χαρακτηριστικά (attributes) του data set.

Ο αλγόριθμος λαμβάνει ως είσοδο κάθε case ως ένα input vector με τις τιμές των μεταβλητών και την αντίστοιχη κλάση που ανήκει.

Αυτές οι cases – input vectors αποτελούν το *training set* του ταξινομητή.

Στη συνέχεια, παρατηρώντας και συγκρίνοντας αν τα vectors, των οποίων οι τιμές είναι κοντά η μία στην άλλη, ανήκουν ή όχι στην ίδια κλάση, κατασκευάζει ένα σετ από κανόνες αποφάσεων (decision rules) με σκοπό την ταξινόμηση μελλοντικών δειγμάτων (*test set*) στις γνωστές κλάσεις.

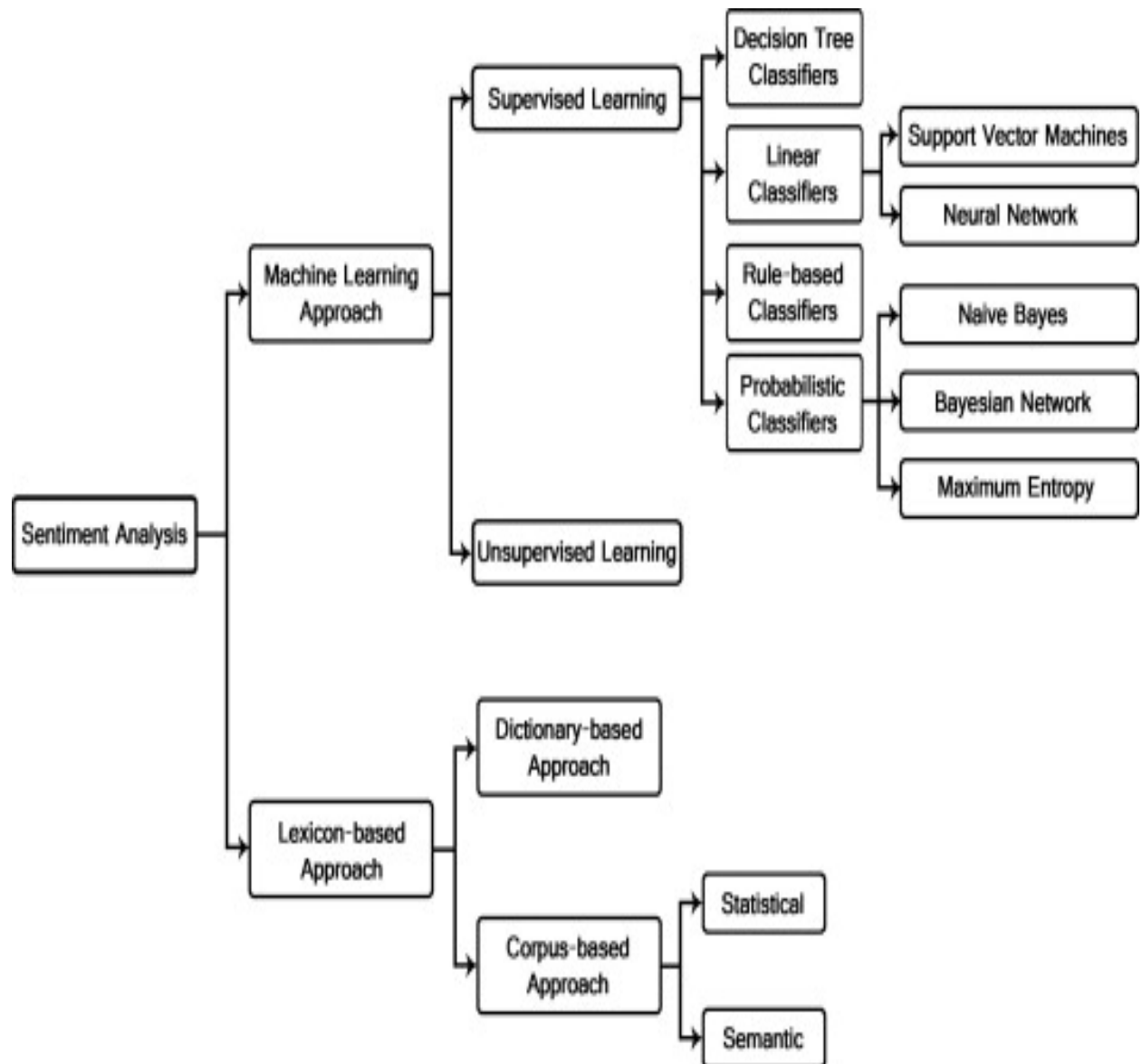


Εικόνα 9: Κλάσεις ταξινόμησης δείγματος

## 4.3 Ερευνητική Διαδικασία

Οι διάφορες προσεγγίσεις και οι πιο δημοφιλείς αλγόριθμοι του Sentiment Analysis απεικονίζονται στο παρακάτω σχήμα.

(<http://www.sciencedirect.com/science/article/pii/S2090447914000550>)



Εικόνα 10: Sentiment classification techniques

Οι τεχνικές αυτές μπορούν να διαιρεθούν σε α) προσέγγιση μηχανικής μάθησης, β) προσέγγιση λεξικού και γ) υβριδική προσέγγιση.

Στον αλγόριθμο που υλοποιήσαμε χρησιμοποιήσαμε τη δεύτερη προσέγγιση (Lexicon-based Approach). Στηρίζεται σε προκαθορισμένες λέξεις χαρακτηρισμένες για την εκδήλωση συναισθήματος με βάση το λεξικό στο οποίο υπάρχουν (Dictionary-based Approach). Στη συνέχεια με βάση τα δεδομένα που προέρχονται από το λεξικό κάνουμε στατιστική και semantic ανάλυση (Corpus-based Approach).

## 4.4 Βήματα Αλγόριθμου

Βήμα 1: Αναζήτηση βιογραφίας και επιλογή της.

Βήμα 2: Αναζήτηση αξιόπιστων λεξικών ανίχνευσης γλωσσικής προκατάληψης. Τα λεξικά που χρησιμοποιήσαμε ήταν:

Ελληνικά: «*Greek Sentiment Lexicon*», περιλαμβάνει 2315 λέξεις.

Αγγλικά: *AFINN-111*, συμπεριλαμβάνει 2477 λέξεις.

Βήμα 3: Αναζήτηση εργαλείων λογισμικού όπου εισάγαμε το κείμενο της βιογραφίας, έτσι ώστε να προκύψει μια λίστα από τις αλφαβητικά ταξινομημένες «μοναδικές λέξεις» (*unique words*) που περιέχονται σε αυτήν. Τα εργαλεία που χρησιμοποιήσαμε ήταν:

Wordlist Maker v 1.2

Unique Word List Text Wizard

Βήμα 4: Αντιγραφή της αλφαβητικά ταξινομημένης λίστας στο Excel. Έτσι δημιουργήσαμε το data set .

Βήμα 5: Άνοιγμα του *Notepad++*. Αντιγραφή του data set της βιογραφίας στο *Notepad++*.

Βήμα 6: Άνοιγμα στο *Notepad++* του αντίστοιχου λεξικού.

Βήμα 7: Συγκρίναμε την αλφαβητικά ταξινομημένη λίστα λέξεων του data set βιογραφίας με το λεξικό και εντοπίσαμε όλες τις κοινές λέξεις που υπάρχουν μεταξύ τους. Το εργαλείο που χρησιμοποιήσαμε ήταν το *Notepad++* με τη βοήθεια ενός από τα πιο εξειδικευμένα πρόσθετα (*plugins*) το *Compare*.

Βήμα 8: Τελικός έλεγχος και καταμέτρηση, προκειμένου να καταρτισθούν οι πίνακες. Για τους πίνακες χρησιμοποιήθηκε το Word.

Βήμα 9: Επανάληψη για κάθε data set βιογραφίας των προηγούμενων βημάτων.

10

- Αναζήτηση βιογραφίας και επιλογή της

20

- Αναζήτηση αξιόπιστων λεξικών ανίχνευσης γλωσσικής προκατάληψης

30

- Αναζήτηση εργαλείων λογισμικού όπου εισάγαμε το κείμενο της βιογραφίας, έτσι ώστε να προκύψει μια λίστα από τις αλφαβητικά ταξινομημένες «μοναδικές λέξεις» (*unique words*) που περιέχονται σε αυτήν.

40

- Αντιγραφή της αλφαβητικά ταξινομημένης λίστας στο Excel. Έτσι δημιουργήσαμε το data set.

50

- Άνοιγμα του *Notepad++*. Αντιγραφή του data set της βιογραφίας στο *Notepad++*.

60

- Άνοιγμα στο *Notepad++* του αντίστοιχου λεξικού

70

- Συγκρίση της αλφαβητικά ταξινομημένης λίστας λέξεων του data set βιογραφίας με το λεξικό και εντοπισμός όλων των κοινών λέξεων που υπάρχουν μεταξύ τους. Το εργαλείο που χρησιμοποιήσαμε ήταν το *Notepad++* με τη βοήθεια ενός από τα πιο εξειδικευμένα πρόσθετα (*plugins*) το *Compare*.

80

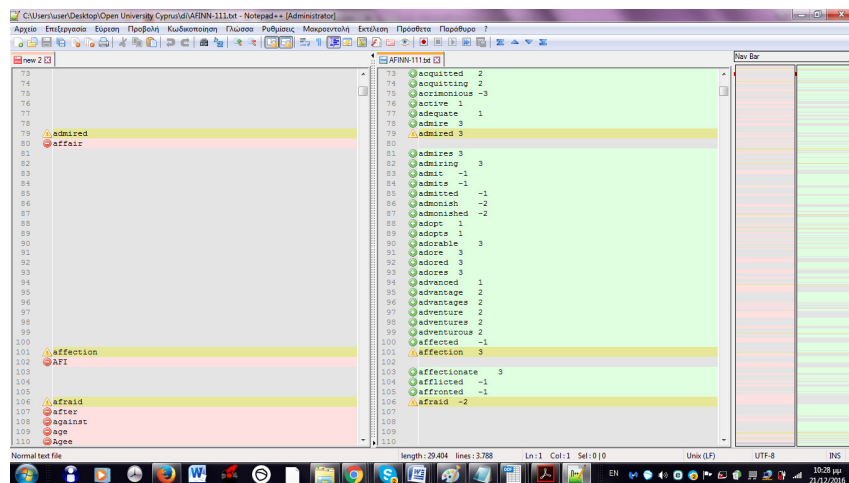
- Τελικός έλεγχος και καταμέτρηση, προκειμένου να καταρτισθούν οι πίνακες. Για τους πίνακες χρησιμοποιήθηκε το Word.

90

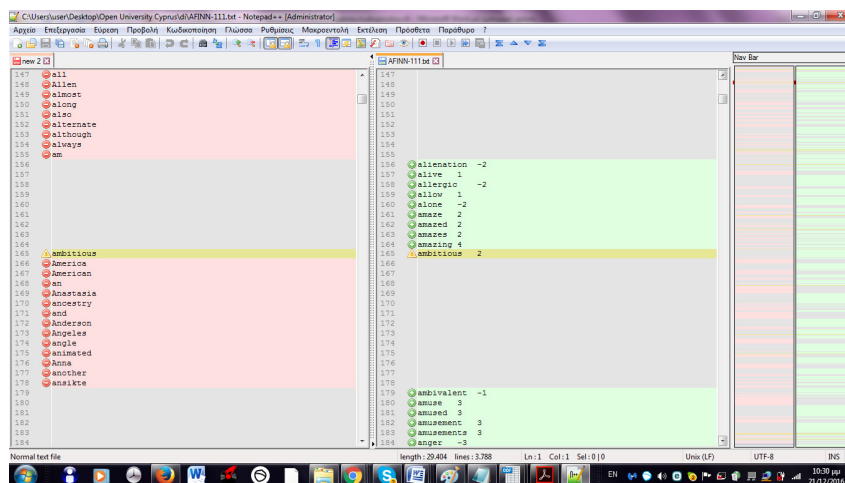
- Επανάληψη για κάθε data set βιογραφίας των προηγούμενων βημάτων.

## 4.4.1 Η λειτουργία Compare

Είναι μια εντολή η οποία περιέχεται στα Πρόσθετα του Notepad++. Αφού έχουμε ανοίξει τη λίστα του λεξικού που χρησιμοποιήσαμε και την ταξινομημένη λίστα από το data set της βιογραφίας, επιλέγουμε από τα πρόσθετα την εντολή compare η οποία συγκρίνει τις δύο λίστες και εμφανίζει τις κοινές λέξεις δηλ. αυτές που περιέχονται στο λεξικό σημασιολογικής ανάλυσης και στην ταξινομημένη λίστα λέξεων του κειμένου που θέλουμε να ερευνήσουμε, εν προκειμένω της βιογραφίας. Παρατηρώντας τα παρακάτω στιγμιότυπα υπάρχουν δύο ξεχωριστά περιβάλλοντα επεξεργασίας (δύο ξεχωριστά πάνελ) όπου είναι ανοιχτά δύο διαφορετικά σύνολα αρχείων σε κάθε περιβάλλον. Η αποτύπωση παραπέμπει σε κοινές λέξεις, συγκεκριμένα πρόκειται για βιογραφία στα αγγλικά και το λεξικό που χρησιμοποιήσαμε ήταν το AFINN.



Εικόνα 11: Αποτέλεσμα του ανιχνευτή λέξεων γλωσσικής προκατάληψης



Εικόνα 12: Αποτέλεσμα του ανιχνευτή λέξεων γλωσσικής προκατάληψης



## 4.5 Ποσοτικά αποτελέσματα

Τα αποτελέσματα αυτά έχουν να κάνουν, με το πλήθος των λέξεων οι οποίες ενέχουν προκατάληψη και οι οποίες εντοπίζονται στις βιογραφίες που ανακτήθηκαν από την Βικιπαίδεια-Wikipedia. Στην επόμενη παράγραφο θα παρατεθούν οι σχετικοί πίνακες, όπου στην αριστερή στήλη υπάρχει το ονοματεπώνυμο της προσωπικότητας που βιογραφείται στην Wikipedia, ενώ στην δεξιά στήλη παρατίθεται το πλήθος των λέξεων που υπάρχουν ως καταχωρήσεις στο αντίστοιχο λεξικό, μετά την σχετική καταμέτρηση και έλεγχο. Όπου καταγράφεται ετικέτα NPOV, θα σημειώνεται δίπλα στον αριθμό αυτόν στην δεξιά στήλη. Επίσης για κάθε υποκατηγορία από τις οκτώ, τα τρία κορυφαία ευρήματα δίνονται με έντονους (bold) χαρακτήρες. Σημειώνεται εδώ ότι για την αναζήτηση των ετικετών NPOV που δεν έχουν αφαιρεθεί, αρκεί η αναζήτησή τους στον χώρο όπου καταγράφονται οι ετικέτες σε μια βιογραφία.

Αν επιθυμούσαμε να αναζητήσουμε την ύπαρξή τους σε κάποιον προγενέστερο χρόνο, θα μπορούσαμε να αξιοποιήσουμε τις δυνατότητες που δίνει το API της Wikipedia.

([https://en.wikipedia.org/w/index.php?title=Laurence\\_Olivier&action=history&year=2016&tagfilter=NPOV](https://en.wikipedia.org/w/index.php?title=Laurence_Olivier&action=history&year=2016&tagfilter=NPOV)). Στην τελευταία παράγραφο θα δοθεί ένας τελικός σχολιασμός των αποτελεσμάτων, ορισμένα από τα οποία αποδεικνύονται ιδιαίτερα ενδιαφέροντα.

## 4.5.1 Βιογραφίες ηθοποιών

Στις αγγλόφωνες βιογραφίες ανδρών ηθοποιών, παρατηρείται ότι οι ηθοποιοί με τα κορυφαία πλήθη γλωσσικής προκατάληψης είναι οι: α) Marlon Brando, β) Laurence Olivier, γ) Brad Pitt.

Όνομα ηθοποιού	Πλήθος λέξεων γλωσσικής προκατάληψης
Eddie Redmayne	5
Anthony Hopkins	59
Michael Douglas	119
<b>Laurence Olivier</b>	<b>271</b>
<b>Marlon Brando</b>	<b>331</b>
Gregory Peck	73
Daniel Day-Lewis	67
Robert De Niro	62
Chuck Norris	83
<b>Brad Pitt</b>	<b>121</b>
Johnny Depp	101
Bruce Willis	80
Al Pacino	84
<b><u>Σύνολο</u></b>	<b>1456</b>

Στις ελληνόφωνες βιογραφίες ανδρών ηθοποιών, παρατηρείται ότι τα κορυφαία «scores» καταγράφονται για τους κάτωθι ηθοποιούς: α) Θανάσης Βέγγος, β) Δημήτρης Χορν, γ) Κώστας Καζάκος

Όνομα ηθοποιού	Πλήθος λέξεων γλωσσικής προκατάληψης
Γιώργος Μεσσάλας	4
Ντίνος Ηλιόπουλος	12
<b>Θανάσης Βέγγος</b>	<b>71</b>
Βλαδίμηρος Κυριακίδης	2
Αλέκος Αλεξανδράκης	11
<b>Δημήτρης Χορν</b>	<b>19</b>
<b>Κώστας Καζάκος</b>	<b>13</b>
Λάμπρος Κωνσταντάρης	9
Στέφανος Ληναίος	2
Χάρης Ρώμας	1
Διονύσης Παπαγιαννόπουλος	4
Χριστόφορος Παπακαλιάτης	1
Δημήτρης Παπαμιχαήλ	6
<b><u>Σύνολο</u></b>	<b>155</b>

Στις αγγλόφωνες βιογραφίες γυναικών ηθοποιών, παρατηρείται ότι τις περισσότερες ανιχνεύσεις συναντούμε στις εξής: α) Angelina Jolie, β) Meryl Streep, γ) Marilyn Monroe

Όνομα ηθοποιού	Πλήθος λέξεων γλωσσικής προκατάληψης
Liza Minelli	43
Julie Andrews	95
Ingrid Bergman	107
Holly Hunter	18
Bette Davis	18
Diane Keaton	141
Vivian Leigh	163
<b>Meryl Streep</b>	<b>202</b>
Sophia Loren	47
Cameron Diaz	42
<b>Marilyn Monroe</b>	<b>200</b>
Catherine Zeta-Jones	111
<b>Angelina Jolie</b>	<b>215</b>
<b>Σύνολο</b>	<b>1402</b>

Στις ελληνόφωνες βιογραφίες γυναικών ηθοποιών, παρατηρείται ότι τις περισσότερες ανιχνεύσεις συναντούμε στις εξής: α) Ρένα Βλαχοπούλου, β) Αλίκη Βουγιουκλάκη, γ) Μελίνα Μερκούρη

Όνομα ηθοποιού	Πλήθος λέξεων γλωσσικής προκατάληψης
Γεωργία Βασιλειάδου	10
Τζένη Καρέζη	5
<b>Ρένα Βλαχοπούλου</b>	<b>44</b>
Έλλη Λαμπέτη	20
Κατίνα Παξινού	12
Ειρήνη Παππά	2
<b>Μελίνα Μερκούρη</b>	<b>22</b>
Κάτια Δανδουλάκη	8
Κατερίνα Λέχου	0
<b>Αλίκη Βουγιουκλάκη</b>	<b>32</b>
Ξένια Καλογεροπούλου	1
Σαπφώ Νοταρά	8
Έλλη Αλεξίου	0
<b><u>Σύνολο</u></b>	<b>164</b>

## 4.5.2 Βιογραφίες πολιτικών

Στις ελληνόφωνες βιογραφίες ανδρών πολιτικών ,η τριάδα των κορυφαίων «επιδόσεων» καταγράφεται για τους: α) Κωνσταντίνος Καραμανλής, β) Ανδρέας Γ. Παπανδρέου, γ) Χαρίλαος Τρικούπης

Όνομα πολιτικού	Πλήθος λέξεων γλωσσικής προκατάληψης
Χαρίλαος Τρικούπης	19
<b>Ελευθέριος Βενιζέλος</b>	<b>37</b>
Γεώργιος Παπανδρέου	25
<b>Κωνσταντίνος Γ. Καραμανλής</b>	<b>54</b>
<b>Ανδρέας Γ. Παπανδρέου</b>	<b>39</b>
Κωνσταντίνος Μητσοτάκης	15
Κώστας Καραμανλής	16
Γεώργιος Α. Παπανδρέου	22
Κωνσταντίνος Γ. Σημίτης	15
Αλέξης Τσίπρας	15
Γεώργιος Γεννηματάς	5
Απόστολος Κακλαμάνης	6
Ευάγγελος Βενιζέλος	12
<b><u>Σύνολο</u></b>	<b>280</b>

**Στις ελληνόφωνες βιογραφίες γυναικών πολιτικών**, οι περισσότερες ανιχνεύσεις υπάρχουν στις: α) Μελίνα Μερκούρη, β) Λιάνα Κανέλλη, γ) Ντόρα Μπακογιάννη

<b>Όνομα πολιτικού</b>	<b>Πλήθος λέξεων γλωσσικής προκατάληψης</b>
Ελένη Σκούρα	1
Καλλιρρόη Παρρέν	8
Λίνα Τσαλδάρη	2
Βάσω Παπανδρέου	3
Άννα Συνοδινού	5
<b>Μελίνα Μερκούρη</b>	<b>22</b>
Άννα Ψαρούδα Μπενάκη	6
Αλέκα Παπαρήγα	6
Μαρία Δαμανάκη	3
<b>Λιάνα Κανέλλη</b>	<b>12</b>
<b>Ντόρα Μπακογιάννη</b>	<b>9</b>
Λούκα Κατσέλη	6
Μαριέττα Γιαννάκου	8
<b><u>Σύνολο</u></b>	<b>91</b>

Στις αγγλόφωνες βιογραφίες γυναικών πολιτικών, τα ανώτατα αποτελέσματα λαμβάνονται για τις εξής: α) Benazir Bhutto, β) Emma Goldman, γ) Benazir Bhutto

Όνομα πολιτικού	Πλήθος λέξεων γλωσσικής προκατάληψης
<b>Benazir Bhutto</b>	<b>346</b>
Dolores Ibárruri	216
<b>Margaret Thatcher</b>	<b>274</b>
Indira Gandhi	252
Isabel Martínez de Perón	124
Madeleine Albright	108
Xiang Xing	142
Angela Merkel	145
Golda Meir	118
Hillary Clinton	249
Victoria Woodhull	108
Mother Teresa	172
<b>Emma Goldman</b>	<b>328</b>
<b><u>Σύνολο</u></b>	<b>2020</b>



Στις αγγλόφωνες βιογραφίες ανδρών πολιτικών, το ανώτατο πλήθος ανιχνεύσεων υφίσταται στις εξής: α) Fidel Castro, β) John Kennedy γ) Nelson Mandella

Όνομα πολιτικού	Πλήθος λέξεων γλωσσικής προκατάληψης
<b>Nelson Mandela</b>	<b>343</b>
Mahatma Gandhi	332
<b>John Kennedy</b>	<b>359</b>
Vladimir Lenin	342
Vladimir Putin	247[NPOV Tag]
Barack Obama	247
<b>Fidel Castro</b>	<b>407</b>
Mao Zedong	298
David Cameron	300
Recep Tayyip Erdogan	331
Francois Mitterand	257
Willi Brandt	194
Mateo Renzi	180
<b><u>Σύνολο</u></b>	<b>3837</b>

Ως συμπληρωματικό αποτέλεσμα, η βιογραφία του Vladimir Putin φέρει ετικέτα *NPOV* (Neutral Point Of View). Σημειώνεται ότι η βιογραφία αυτή είναι η μοναδική από τις 104 που εξετάστηκαν με τέτοιο εύρημα.

# Κεφάλαιο 5

## Σχολιασμός αποτελεσμάτων

### 5.1 Αναφορικά ως προς τα δύο φύλα

Βασισμένοι σε όλα τα ανωτέρω, μπορούμε να προβούμε σε κάποιον αναλυτικό σχολιασμό των αποτελεσμάτων. Ο σχολιασμός μπορεί να γίνει από ορισμένα διαφορετικά μεταξύ τους πρίσματα, κάθε ένα από τα οποία μπορεί να δώσει μια αξιόλογη εμβάθυνση και ερμηνεία των αποτελεσμάτων. Εστιάζοντας, έτσι, πρώτον, στη διαφορετική κάλυψη αναφορικά προς τα δύο φύλα, παρατηρείται ότι:

- Στις τέσσερις πρώτες κατηγορίες βιογραφιών (ηθοποιών), δεν υπάρχουν σημαντικές διαφορές στο πλήθος των ανιχνεύσεων. Στη μία περίπτωση συγκεκριμένα, στους Έλληνες και Ελληνίδες ηθοποιούς, είναι οι Ελληνίδες ηθοποιοί των οποίων οι βιογραφίες έχουν το προβάδισμα στην προκατειλημμένη γλώσσα, κατά ποσοστό 5,806%. Αντιθέτως, το προβάδισμα στις αγγλόφωνες βιογραφίες το έχουν οι άνδρες ηθοποιοί, κατά 3,852%.
- Μεταβαίνοντας στην κατηγορία των πολιτικών, η εικόνα αλλάζει πάρα πολύ. Στις βιογραφίες των Ελλήνων ανδρών πολιτικών συναντάται, σύμφωνα με τους υπολογισμούς, 207,7% περίπου περισσότερη προκατειλημμένη γλώσσα σχετικά με τις γυναίκες πολιτικούς. Αντίστοιχα, το ποσοστιαίο προβάδισμα στις αγγλόφωνες βιογραφίες είναι περίπου στο 89,95%.

Πού θα μπορούσαμε να αποδώσουμε κάτι τέτοιο; Οπωσδήποτε σημαντικό ρόλο έχει η έκταση των βιογραφιών.

## 5.2 Ως προς την έκταση των βιογραφιών

Στον παρακάτω πίνακα παρατηρείται μια δειγματοληπτική κατάσταση όπου αντιπαραβάλλονται βιογραφίες με σχεδόν ίδιο αριθμό λέξεων ανά ζεύγη: i) Benazir Bhutto και Fidel Castro, ii) Indira Gandhi και Francois Mitterand

Όνομα πολιτικού	Συνολικές λέξεις βιογραφίας
Benazir Bhutto	15717
Fidel Castro	14083
Indira Gandhi	9528
Francois Mitterand	10585

Η έκταση της βιογραφίας του Francois Mitterand είναι κατά περίπου 11,09% μεγαλύτερη από αυτήν της Indira Gandhi, παρά το ότι παρουσιάζουν περίπου το ίδιο πλήθος προκατειλημμένων λέξεων σε απόλυτους όρους. Αντίστοιχα, η βιογραφία της B. Bhutto είναι εκτενέστερη κατά 11,6% αυτής του Fidel Castro, παρά την ισόποση χρήση προκατειλημμένης γλώσσας και στις δύο σε απόλυτα νούμερα. Αυτή η εναλλάξ διαφοροποίηση μας δίνει το έναυσμα για μια ακόμη παρατήρηση από διαφορετική σκοπιά. Οι κορυφαίες στην κατηγορία τους περιπτώσεις της Benazir Bhutto, της Emma Goldman και κατεξοχήν του Fidel Castro, που άλλωστε διαθέτει το ανώτατο πλήθος προκατειλημμένης γλώσσας, αλλά και του Vladimir Putin, η βιογραφία του οποίου φέρει την ετικέτα NPOV έχουν ένα κοινό χαρακτηριστικό.

Πρόκειται δηλαδή για προσωπικότητες οι οποίες δεν συνάδουν προς την δυτική κοσμοθεώρηση η οποία, όπως έχει επισημανθεί από ερευνητές, είναι σε μεγάλο βαθμό κυρίαρχη στους συντάκτες των άρθρων της Wikipedia (<https://www.theguardian.com/technology/2015/sep/15/wikipedia-view-of-the-world-is-still-written-by-the-west>) (Reagle, J. - L. Rhue, 2011).

Ο Fidel Castro, για παράδειγμα, εκτός των άλλων, αποκαλείται «δικτάτορας» εφάμιλλος του Χίτλερ και του Μουσσολίνι, ενώ η πολιτική του χαρακτηρίζεται από ολοκληρωτισμό («totalitarianism»), ενώ η Emma Goldman συνδέθηκε με περισσότερο ριζοσπαστικές απόψεις στα πλαίσια της ευρύτερης ιδεολογίας του αναρχισμού οι οποίες παρομοίως δεν συνάδουν με τις δυτικές αντιλήψεις περί πολιτικής.

Ένα τελευταίο στοιχείο που ίσως πρέπει να προστεθεί σε όλα τα παραπάνω είναι και η φύση των επαγγελματιών αυτών.

Δε διαλανθάνει της προσοχής μας το γεγονός ότι το επάγγελμα του πολιτικού υπήρξε για πολλούς αιώνες –και στις περισσότερες χώρες ως και τα μέσα του 20<sup>ου</sup> αιώνα– κυριαρχούμενο ολοκληρωτικά από το ανδρικό φύλο, πολύ περισσότερο από αυτό του ηθοποιού. Αυτή η παρατήρηση συνεπάγεται ότι οι άνδρες πολιτικοί τείνουν, έως και σήμερα, να αποτελούν τη μεγάλη πλειοψηφία, (<http://www.theguardian.com/politics/2015/may/10/who-are-the-new-mps-middle-aged-white-men-again>) και ακόμη, να έχουν σημαντικότερες θέσεις στην άσκησή της, με αποτέλεσμα να εφελκύνουν περισσότερο το ενδιαφέρον των βιογράφων.

### **5.3 Σύνοψη των ευρημάτων και ερμηνεία τους**

Σε αυτήν την εργασία κατεξοχήν μελετήθηκε η ανίχνευση της προκατειλημμένης γλώσσας σε βιογραφίες που διατίθενται online και συγγράφονται με βάση την τεχνική του πληθοπορισμού, στην διαδικτυακή εγκυκλοπαίδεια Wikipedia ή Βικιπαίδεια. Ιδιαίτερη έμφαση δόθηκε στην εξήγηση του πώς οι προκαταλήψεις περνούν και εκφράζονται μέσα από την κάθε γλώσσα, καθώς και στον τρόπο που δημιουργούνται στερεοτυπικές αντιλήψεις για ποικιλόμορφες ομάδες, αρχής γενομένης από το φύλο, την εθνότητα, την φυλή, την κοινωνική καταγωγή κ.ά. Στην έρευνα που διεξάχθηκε αναπτύχθηκε ένας συστηματικός τρόπος ανίχνευσης της προκατειλημμένης γλώσσας, μάλιστα με «πρώτες ύλες» εργαλεία και λογισμικά που έχουν διατεθεί ελεύθερα από τους κατασκευαστές τους προς χρήση στο Διαδίκτυο: λεξικά προκατάληψης, εργαλεία εξαγωγής αλφαβητικά ταξινομημένων λιστών λέξεων και λογισμικά σύγκρισης λιστών λέξεων, όπως και μερική χρήση του API της Wikipedia είναι τα κύρια στοιχεία αυτής της συστηματικής προσέγγισης που καταδείχθηκε. Η έρευνα που διεξάχθηκε σε μια σειρά από οκτώ διαφορετικές υποκατηγορίες βιογραφιών που εκτείνονται σε δύο διαφορετικές γλώσσες έδειξε ότι στο επάγγελμα του ηθοποιού δεν καταγράφονται σημαντικές διαφορές μεταξύ των δύο φύλων σε ό,τι αφορά την συνολική και σε απόλυτους όρους χρήση στερεοτυπικής γλώσσας με θετική ή αρνητική φόρτιση αδιακρίτως. Τουναντίον, για το επάγγελμα του πολιτικού, συναντήθηκε σε απόλυτα νούμερα κατά πολύ εκτενέστερη και για τις δύο γλώσσες χρήση προκατειλημμένης γλώσσας στους άνδρες πολιτικούς. Όλα τα παραπάνω αποτελέσματα ειδώθηκαν μέσα από διαφορετικά πρίσματα και επισημάνθηκε η μερική επαλήθευση της έρευνας ως προς το ότι σε γενικές γραμμές η Βικιπαίδεια παρουσιάζει βελτιωμένη επίδοση όσον αφορά την προκατάληψη ως προς το φύλο, ενώ, αντιθέτως, παρουσιάζει σημαντική προκατάληψη όσον αφορά πρόσωπα τα οποία δεν συνάδουν προς τον δυτικό τρόπο ζωής και ειδικότερα τις δυτικές πολιτικές αξίες (φιλελεύθερη δημοκρατία, ελεύθερη αγορά, έννομη τάξη, δυτική επικυριαρχία).

Οι προσωπικότητες λοιπόν οι οποίες φαίνεται ότι αψηφούν αυτές τις αντιλήψεις παρουσιάζουν πολύ συχνά ανώτατα επίπεδα προκατειλημμένου λεξιλογίου, και μάλιστα με ιδιαίτερα αρνητική φόρτιση.

## 5.4 Συζήτηση, προτάσεις για περαιτέρω έρευνα

Η παρούσα έρευνα προσπάθησε να δημιουργήσει ένα υπολογιστικό εργαλείο για την εξόρυξη γνώμης, πιο συγκεκριμένα για την ανίχνευση γλωσσικής προκατάληψης διερευνώντας το βαθμό υποκειμενικότητας σε κείμενα του Διαδικτύου με βάση το φύλο. Επιλέχθηκε η Βικιπαίδεια γιατί, όπως αναφέρθηκε και στη βιβλιογραφική ανασκόπηση, τα άρθρα που διαθέτει δε τυγχάνουν αξιοπιστίας σε οποιονδήποτε τα επεξεργαστεί ή τα συγγράψει. Έτσι τα ευρήματα της έρευνας θα ενημερώσουν τον αναγνώστη ότι θα πρέπει να έχει ο ίδιος τη γνώση και την κατάρτιση ως προς το πώς να τα αξιολογήσει. Όπως είδαμε και από τη βιβλιογραφική ανασκόπηση υπάρχουν μελέτες και προηγούμενες έρευνες για τον έλεγχο της υποκειμενικότητας σε κείμενα με τις διάφορες τεχνικές μηχανικής μάθησης, στις οποίες αναφερθήκαμε. Η επιλογή της lexicon approach στη συγκεκριμένη έρευνα μας βοήθησε με τα εργαλεία που χρησιμοποιήσαμε να αναλύσουμε ένα δείγμα 104 βιογραφιών. Η μη υποκειμενική στάση των άρθρων της Βικιπαίδεια απέναντι σε κάποιες προσωπικότητες συγκεκριμένης πολιτικής αναφοράς, κατά την εκτίμησή μου αποτελεί ένα σπουδαίο εύρημα. Το δείγμα υπήρξε όσο γίνεται κατανεμημένο ως προς τις κατηγορίες που προκαθορίστηκαν. Περιορισμός αποτελεί όμως το μέγεθος του δείγματος. Τα ποσοτικά αποτελέσματα μαρτυρούν τη γλωσσική προκατάληψη ως προς το φύλο. Έτσι, ο περιορισμός αυτός θέτει μια πρόταση, που σε κάθε κατηγορία, όταν ο αλγόριθμος που δημιουργήσαμε εφαρμοστεί σε μεγαλύτερο δείγμα θα προσφέρει καλύτερα ποιοτικά αποτελέσματα. Επίσης άλλος σημαντικός περιορισμός που παρατηρήθηκε ήταν ότι κατά την εκτέλεση του αλγορίθμου παρατηρήθηκαν δυσκολίες που οφείλονται κυρίως στην επαναληπτική μη αυτοματοποιημένη καταμέτρηση λέξεων που ενέχουν γλωσσικής προκατάληψης. Η πρόταση για ένα ολοκληρωμένο προγραμματιστικό περιβάλλον εφαρμογής του συγκεκριμένου αλγορίθμου θα επηρεάσει τη παρούσα χρονοβόρα εκτέλεσή του. Όμως ο εμπλουτισμός και ο σχεδιασμός πιο ευαίσθητων σε λέξεις που ενέχουν γλωσσική προκατάληψη λεξικών είναι απαραίτητος για την αντιμετώπιση του φαινομένου της γλωσσικής προκατάληψης.

Όπως αναφέρθηκε στη βιβλιογραφική επισκόπηση τα συστήματα εξόρυξης γνώμης με διάφορες τεχνικές είναι ευρέως διαδεδομένα για την αποτύπωση συναισθηματικής ανάλυσης στην αγγλική γλώσσα. Η ελληνική γλώσσα όμως, όπως αναφέρθηκε περιέχει αρκετούς σύνθετους τύπους σύνταξης, γραμματικής οι οποίοι πολλές φορές δημιουργούν αμφισημία και λάθος εκτίμηση της συναισθηματικής ανάλυσης. Η υποκειμενικότητα που περιέχει το οποιοδήποτε λεξικό συναισθηματικής ανάλυσης και ειδικά στην ελληνική γλώσσα σε κάθε λέξη που σε αυτή αναφέρεται, δε μπορεί απόλυτα να εξασφαλίσει ακρίβεια στην εξόρυξη γνώμης συναισθηματικής ανάλυσης, καθώς δεν περιέχει όλους τους πιθανούς τρόπους με τους οποίους αυτές μπορούν να εκφραστούν. Αυτός ο περιορισμός δημιουργεί την ανάγκη μιας νέας πρότασης ειδικά για την ελληνική γλώσσα. Η δόμηση συντακτικών και σημασιολογικών κανόνων οι οποίοι να αποδίδουν αξιοπιστία στα κείμενα που θα εφαρμοστούν είναι μια πρόταση για τη επιτυχή ερμηνεία του λεξικού πόρου στο αντίστοιχο σύστημα συναισθηματικής ανάλυσης που θα εφαρμοστεί.

# Παράρτημα Α

## Διοικητικά έγγραφα

### Α.1 Άρθρο και γλωσσική προκατάληψη

«ΕΝΤΑΞΗ ΤΗΣ ΙΣΟΤΗΤΑΣ ΤΩΝ ΦΥΛΩΝ ΣΤΗ ΓΛΩΣΣΑ ΤΩΝ ΔΙΟΙΚΗΤΙΚΩΝ ΕΓΓΡΑΦΩΝ» Σελίδα 28

Περίπτωση: Άρθρο

Η ελληνική γλώσσα έχει οριστικό και αόριστο άρθρο.

Στην περίπτωση επαγγελματικών ονομάτων υπάρχει κοινή για το θηλυκό και το αρσενικό κατάληξη σε -ος και εκείνων για τα οποία επιλέγεται ο λόγιος τύπος του αρσενικού, το θηλυκό ξεχωρίζει από το αρσενικό μόνο εφόσον εκφέρεται με άρθρο. Σα δύο γένη συμπίπτουν όμως και πάλι στην ονομαστική και γενική του πληθυντικού.

**Πρόβλημα:**

Στο σύνολο των εγγράφων που αναλύθηκαν οι περιπτώσεις ασάφειας ως προς την πρόθεση της/του συντάκτριας/τη να χρησιμοποιήσει το αρσενικό γένος επιλύονται με τη χρήση του αρσενικού άρθρου σε διαφορετικά σημεία του ίδιου κειμένου.

Έτσι, για παράδειγμα *οι υπάλληλοι, οι εκπαιδευτικοί, οι ασθενείς, οι πρόσφυγες, οι ανά τον κόσμο ομογενείς* προσδιορίζονται σε επόμενες παραγράφους του ίδιου κειμένου ως *άντρες εφόσον στην αιτιατική του πληθυντικού αναφέρονται ως τους υπαλλήλους, τους εκπαιδευτικούς, τους ασθενείς, τους πρόσφυγες, τους ανά τον κόσμο ομογενείς*.

Στη συντριπτική τους πλειοψηφία λοιπόν, τα ονόματα που περιγράφουν επαγγελματική ή άλλη ιδιότητα λοιπόν, αναλύονται ως γένους αρσενικού επιβεβαιώνοντας την κυρίαρχη τάση της ελληνικής δημόσιας διοίκησης να αναφέρεται ή/και να απευθύνεται και στις γυναίκες αποδέκτριες των ρυθμίσεων που περιλαμβάνονται στα έγγραφα -ή και των ίδιων των εγγράφων- μέσα από την προσφώνηση/ αναφορά σε άντρες εκπροσώπους των εν λόγω κατηγοριών.

**Πρόταση:**

*Σε κάθε περίπτωση κατά την οποία δημιουργείται αμφισημία ως προς το γένος του ονόματος αναφοράς σε άτομα και ομάδες ανθρώπων λόγω της κοινής μορφολογικής κατάληξης, να αναγράφονται τα άρθρα και στα δύο γένη. την περίπτωση δε της μορφολογικά κοινής ονομαστικής και γενικής του πληθυντικού, να προστίθεται η διευκρίνιση: 'γυναίκες και άντρες'.*

Δηλαδή, *τις/τους υπαλλήλους, τις/τους εκπαιδευτικούς, τις/τους ασθενείς, τις/του πρόσφυγες, τις/τους ανά τον κόσμο ομογενείς & οι πρόσφυγες, γυναίκες και άντρες, οι ανά τον κόσμο γυναίκες και άντρες ομογενείς, γυναικών και αντρών υπαλλήλων, των εκπαιδευτικών, γυναικών και αντρών, των ασθενών, γυναικών και αντρών*

# Παράρτημα Β

## Code semantics

### B.1 Functions

Tokenize and tag some text:

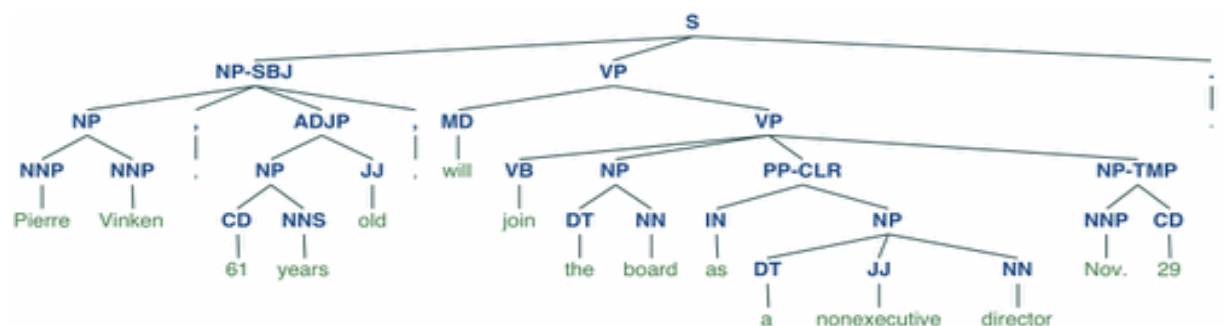
```
>>> import nltk
>>> sentence = """At eight o'clock on Thursday morning
... Arthur didn't feel very good."""
>>> tokens = nltk.word_tokenize(sentence)
>>> tokens
['At', 'eight', "o'clock", 'on', 'Thursday', 'morning',
'Arthur', 'did', "n't", 'feel', 'very', 'good', '.']
>>> tagged = nltk.pos_tag(tokens)
>>> tagged[0:6]
[('At', 'IN'), ('eight', 'CD'), ("o'clock", 'JJ'), ('on', 'IN'),
('Thursday', 'NNP'), ('morning', 'NN')]
```

Identify named entities:

```
>>> entities = nltk.chunk.ne_chunk(tagged)
>>> entities
Tree('S', [(('At', 'IN'), ('eight', 'CD'), ("o'clock", 'JJ'), ('on', 'IN'),
('Thursday', 'NNP'), ('morning', 'NN'))],
Tree('PERSON', [(('Arthur', 'NNP')])], ('did', 'VBD'), ("n't", 'RB'), ('feel',
'VB'), ('very', 'RB'), ('good', 'JJ'), ('.', '.')])
```

Display a parse tree:

```
>>> from nltk.corpus import treebank
>>> t = treebank.parsed_sents('wsj_0001.mrg')[0]
>>> t.draw()
```





## B.2 Source code for nltk.ccg.logic Functions

```
def compute_composition_semantics(function, argument):
    assert isinstance(argument, LambdaExpression), "" + str(argument) + "` must be a
    lambda expression" return Lambda Expression(argument.variable, Application
    Expression(function, argument.term).simplify())
```

```
def compute_substitution_semantics(function, argument):
    assert isinstance(function, LambdaExpression) and isinstance(function.term,
    LambdaExpression), "" + str(function) + "` must be a lambda expression with 2
    arguments"
    assert isinstance(argument, LambdaExpression), "" + str(argument) + "` must be a
    lambda expression" new_argument = ApplicationExpression(argument,
    VariableExpression(function.variable)).simplify()new_term=
    ApplicationExpression(function.term,new_argument).simplify()return
    LambdaExpression(function.variable, new_term)
```

[http://www.nltk.org/\\_modules/nltk/ccg/logic.html#compute\\_function\\_semantics](http://www.nltk.org/_modules/nltk/ccg/logic.html#compute_function_semantics)

# Παράρτημα Γ

## Sentiment Greek Lexicon

### Γ.1 Περιγραφή

Είναι ένα ελληνικό λεξικό που θα χρησιμοποιηθεί για την ανάλυση συναισθήματος .

Αυτό το λεξικό δημιουργήθηκε με την υποστήριξη του EC-funded FP7 Project . Το λεξικό δημιουργήθηκε από τον Adam Tsakalidis (CERTH-ITI, now University of Warwick) και με τη συνεργασία του Symeon Papadopoulos (CERTH-ITI) και τη συμβολή της Ourania Voskaki (Centre for Greek Language) και της Kyriaki Ioannidou (Centre for Greek Language) και Christina Boididou (CERTH-ITI).

Κάθε σειρά του λεξικού διαχωρίζεται σε πεδία ως εξής:

- Κατάληξη: Στις περιπτώσεις των επιθέτων, τα τρία γένη (αρσενικό, θηλυκό, ουδέτερο) παρουσιάζεται με την κατάληξη (-ος -η -ο).
- POS\_x: Μέρος του λόγου. Οι τιμές  $x$  ( $x = 1,2,3,4$ ).
- Subjectivity\_x: κλάση υποκειμενικότητας. ανατεθεί από σχολιαστής  $x$  . Πιθανές τιμές είναι αντικειμενική ( OBJ), ασθενώς υποκειμενική ( SUBJ-) και έντονα υποκειμενική ( SUBJ+).
- Polarity\_x: τάξη πολικότητας:. Πιθανές τιμές είναι θετικά ( POS), Αρνητική ( NEG), τόσο θετικές όσο και αρνητικές ( BOTH) και Ουδέτερο ( N/A).
- Anger\_x:Θυμός : Πιθανές τιμές είναι μεταξύ 1 και 5, και N/A.
- Disgust\_x: Αηδία . Πιθανές τιμές είναι μεταξύ 1 και 5, και N/A.
- Fear\_x: Φόβος. Πιθανές τιμές είναι μεταξύ 1 και 5, και N/A.
- Happiness\_x: Ευτυχία . Πιθανές τιμές είναι μεταξύ 1 και 5, και N/A.
- Sadness\_x: Θλίψη. Πιθανές τιμές είναι μεταξύ 1 και 5, και N/A.
- Surprise\_x: Έκπληξη. Πιθανές τιμές είναι μεταξύ 1 και 5, και N/A.

αγανάκτηση	NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG	
NEG	5	5	5	5	4	3	5					
αγανακτώ	VERB	VERB	VERB	VERB	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG	
NEG	5	5	5	5	4	3	5					
αγάπη	NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ+	SUBJ+	SUBJ+	POS	POS	POS	POS
1	1	1	1	1	1	1						
αγαπημένος -η -ο		PART	PART	PART	PART	SUBJ+	SUBJ+	SUBJ+	SUBJ+	POS	POS	
POS	POS	1	1	1	1	1	1	1				
αγαπητός -ή -ό	ADJ	ADJ	ADJ	ADJ	SUBJ+	SUBJ+	SUBJ+	SUBJ+	POS	POS	POS	
POS	1	1	1	1	1	1	1					
αγαπώ	VERB	VERB	VERB	VERB	SUBJ+	SUBJ+	SUBJ+	SUBJ+	POS	POS	POS	POS
1	1	1	1	1	1	1						
άγγελος		NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ-	SUBJ+	OBJ	POS	BOTH	POS
N/A	1	1	1	N/A	1	1	1					
αγέλη	NOUN	NOUN	NOUN	NOUN	SUBJ-	SUBJ-	SUBJ-	OBJ	NEG	NEG	NEG	N/A
4	3	5	N/A	3	3	5						
αγευσία		NOUN	NOUN	NOUN	NOUN	OBJ	OBJ	SUBJ+	OBJ	N/A	N/A	NEG
N/A	N/A	N/A	1	N/A	N/A	N/A	5					
αγκουσεύω	VERB	VERB	VERB	VERB	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG	
NEG	3	1	1	3	2	1	1					
αγνός -ή -ό	ADJ	ADJ	ADJ	ADJ	SUBJ+	SUBJ-	SUBJ+	SUBJ+	BOTH	POS	POS	
POS	1	1	1	1	1	1	1					

αγορεύω	VERB	VERB	VERB	VERB	SUBJ+	SUBJ-	SUBJ-	OBJ	NEG	NEG	NEG
N/A	3	2	2	N/A	4	2	5				
αγρίεμα	NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ-	SUBJ+	SUBJ+	NEG	NEG	NEG
NEG	5	5	4	5	3	1	4				
αγριεύω	VERB	VERB	VERB	VERB	SUBJ+	SUBJ-	SUBJ+	SUBJ+	NEG	NEG	NEG
NEG	5	5	4	5	3	1	4				
αγροτοπατέρας		NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG
NEG	NEG	4	3	2	3	4	3	4			
άγχος	NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG
2	1	1	1	3	1	2					
αγωνία	NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ+	SUBJ+	SUBJ+	BOTH	NEG	NEG
BOTH	1	1	1	1	1	1	1				
αγωνιώ	VERB	VERB	VERB	VERB	SUBJ+	SUBJ+	SUBJ+	SUBJ+	BOTH	NEG	NEG
BOTH	1	1	1	1	1	1	1				
αδειάζω	VERB	VERB	VERB	VERB	SUBJ+	SUBJ-	SUBJ+	SUBJ-	NEG	NEG	NEG
BOTH	4	3	2	1	4	3	1				
άδειασμα	NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ-	SUBJ+	SUBJ-	NEG	NEG	NEG
BOTH	4	2	2	1	4	3	1				
άδειος -α -ο	ADJ	ADJ	ADJ	ADJ	SUBJ+	OBJ	OBJ	SUBJ-	NEG	N/A	N/A
NEG	3	N/A	N/A	1	3	N/A	N/A				
αδελφάτο	NOUN	NOUN	NOUN	NOUN	SUBJ-	SUBJ-	SUBJ-	SUBJ+	BOTH	NEG	NEG
NEG	3	1	3	3	3	2	3				

αδελφή	NOUN	NOUN	NOUN	NOUN	SUBJ-	SUBJ-	SUBJ-	SUBJ-	POS	BOTH	POS
NEG	1	1	1	3	1	2	1				
αδελφικός -ή -ό	ADJ	ADJ	ADJ	ADJ	ADJ	SUBJ+	OBJ	SUBJ+	SUBJ+	POS	N/A
POS	POS	1	N/A	1	1	1	N/A	1			
αδελφός	NOUN	NOUN	NOUN	NOUN	SUBJ-	SUBJ-	SUBJ-	SUBJ-	POS	POS	POS
POS	1	1	1	1	1	1	1				
αδελφοσύνη	NOUN	NOUN	NOUN	NOUN	SUBJ-	SUBJ+	SUBJ+	SUBJ-	POS	POS	POS
POS	1	1	1	1	1	1	1				
αδελφούλης	NOUN	NOUN	NOUN	NOUN	SUBJ-	SUBJ+	SUBJ+	SUBJ+	POS	POS	POS
POS	1	1	1	1	1	1	1				
αδερφός	NOUN	NOUN	NOUN	NOUN	SUBJ-	SUBJ-	SUBJ-	SUBJ-	POS	POS	POS
POS	1	1	1	1	1	1	1				
αδερφούλης	NOUN	NOUN	NOUN	NOUN	SUBJ-	SUBJ+	SUBJ+	SUBJ+	POS	POS	POS
POS	1	1	1	1	1	1	1				
αδιάθετος -η -ο	ADJ	ADJ	ADJ	ADJ	OBJ	OBJ	OBJ	SUBJ+	N/A	N/A	N/A
NEG	N/A	N/A	N/A	1	N/A	N/A	N/A				
αδιάλυτος -η -ο	ADJ	ADJ	ADJ	ADJ	SUBJ-	OBJ	SUBJ-	OBJ	NEG	N/A	NEG
N/A	1	N/A	1	N/A	2	N/A	2				
αδιάντροπος -η -ο	ADJ	ADJ	ADJ	ADJ	ADJ	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG
NEG	NEG	4	2	5	3	5	5	5			
αδιαφορία	NOUN	NOUN	ADJ	NOUN	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG
NEG	4	1	4	3	5	1	1				

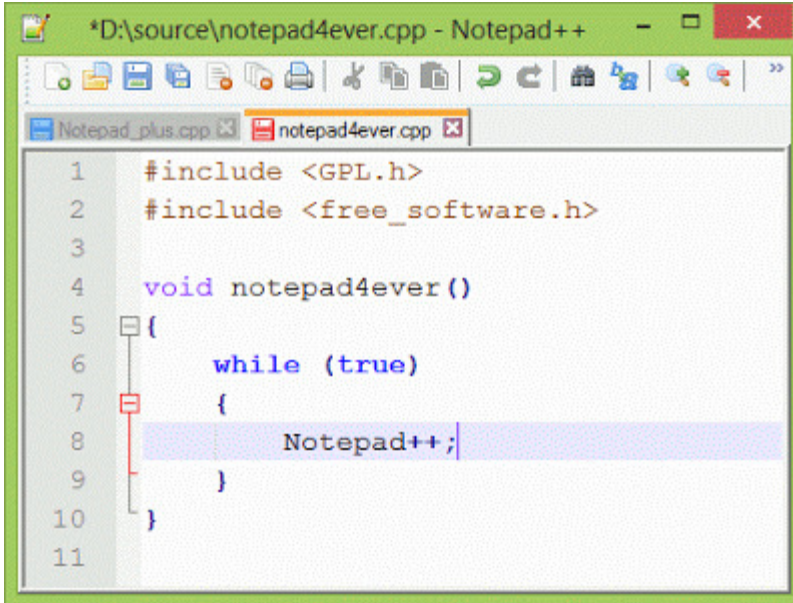
άδολος-η-ο	ADJ	ADJ	ADJ	ADJ	SUBJ+	OBJ	SUBJ+	SUBJ+	POS	N/A	POS	
POS	1	N/A	1	1	1	N/A	1					
αέρας	NOUN	NOUN	NOUN	NOUN	OBJ	OBJ	OBJ	OBJ	N/A	N/A	N/A	N/A
	N/A	N/A	N/A	N/A	N/A	N/A	N/A					
αεροβατώ	VERB	VERB	VERB	VERB	SUBJ+	OBJ	SUBJ+	SUBJ-	NEG	N/A	NEG	
NEG	1	N/A	4	1	3	N/A	1					
αηδής-ής-ές	NOUN	ADJ	ADJ	ADJ	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG	
NEG	4	1	3	3	5	5	5					
αηδία	NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG	NEG
	4	1	3	2	5	5	5					
αηδιάζω	VERB	VERB	VERB	VERB	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG	
NEG	4	1	3	3	5	5	5					
άθλος	NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ+	SUBJ+	SUBJ-	POS	POS	POS	POS
	1	1	1	1	1	1	1					
αιδημοσύνη	NOUN	NOUN	NOUN	NOUN	SUBJ+	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG	
BOTH	4	1	4	1	5	1	4					
αιδώς	NOUN	NOUN	NOUN	NOUN	SUBJ-	SUBJ+	SUBJ+	SUBJ+	NEG	NEG	NEG	
BOTH	4	4	4	1	5	4	4					
αίμα	NOUN	NOUN	NOUN	NOUN	SUBJ+	OBJ	SUBJ-	SUBJ-	NEG	N/A	NEG	
BOTH	4	N/A	2	4	4	N/A	4					

# Παράρτημα Δ

## Περιγραφή Εργαλείων

### Δ.1 Notepad++

Είναι ένα εργαλείο που περιέχει πολλές εφαρμογές στο χρήστη όπως μετάφραση από σε γλώσσα καθώς επίσης για να επεξεργαζόμαστε κώδικα. Είναι ιδανικό για νέους χρήστες που θέλουν να πειραματιστούν στην χρήση και επεξεργασία κώδικα και γενικότερα στον προγραμματισμό. Επίσης λειτουργεί εξαιρετικά και ως ένα ισχυρός επεξεργαστής κειμένου και διαθέτει την Ελληνική γλώσσα.

A screenshot of the Notepad++ application window. The title bar shows the file path '\*D:\source\notepad4ever.cpp - Notepad++'. The window contains two tabs: 'Notepad\_plus.cpp' and 'notepad4ever.cpp'. The code in the active tab is as follows:

```
1  #include <GPL.h>
2  #include <free_software.h>
3
4  void notepad4ever ()
5  {
6      while (true)
7      {
8          Notepad++;
9      }
10 }
11
```

## Δ.2 The Unique Word List Text Wizard

Είναι ένα εργαλείο στο οποίο εισάγοντας το κείμενο ελέγχει για πιθανόν διπλές λέξεις και εμφανίζει το κείμενο που δε θα περιέχει επανάληψη λέξεων. Εμφανίζει δηλαδή μοναδικές λέξεις. Επίσης χωρίζονται οι λέξεις από κενά και αλλαγές γραμμών.



## Δ.3 Δείγμα αγγλικού λεξικού AFINN

abandon	-2
abandoned	-2
abandons	-2
abducted	-2



abduction	-2
abductions	-2
abhor	-3
abhorred	-3
abhorrent	-3
abhors	-3
abilities	2
ability	2
aboard	1
absentee	-1
absentees	-1
absolve2	
absolved	2
absolves	2
absolving	2
absorbed	1
abuse	-3
abused	-3

abuses	-3
abusive	-3
accept	1
accepted	1
accepting	1
accepts	1
accident	-2
accidental	-2
accidentally	-2
accidents	-2
accomplish	2
accomplished	2
accomplishes	2
accusation	-2
accusations	-2
accuse	-2
accused	-2
accuses	-2

accusing	-2
ache	-2
achievable	1
aching	-2
acquit	2
acquits	2
acquitted	2
acquitting	2
acrimonious	-3
active	1
adequate	1
admire	3
admired	3
admires	3
admiring	3
admit	-1
admits	-1
admitted	-1

admonish	-2
admonished	-2
adopt	1
adopts	1
adorable	3
adore	3
adored	3
adores	3
advanced	1
advantage	2
advantages	2
adventure	2
adventures	2
adventurous	2
affected	-1
affection	3
affectionate	3
afflicted	-1

affronted	-1
afraid	-2
aggravate	-2
aggravated	-2
aggravates	-2
aggravating	-2
aggression	-2
aggressions	-2
aggressive	-2
aghast	-2
agog	2
agonise	-3
agonised	-3
agonises	-3
agonising	-3
agonize	-3
agonized	-3
agonizes	-3

agonizing	-3
agree	1
agreeable	2
agreed	1
agreement	1
agrees	1
alarm	-2
alarmed	-2
alarmist	-2
alarmists	-2
alas	-1
alert	-1
alienation	-2
alive	1
allergic	-2
allow	1
alone	-2
amaze	2

amazed	2
amazes	2
amazing	4
ambitious	2
ambivalent	-1
amuse	3
amused	3
amusement	3
amusements	3
anger	-3
angers	-3
angry	-3
anguish	-3
anguished	-3
animosity	-2
annoy	-2
annoyance	-2
annoyed	-2

annoying	-2
annoys	-2
antagonistic	-2
anti	-1
anticipation	1
anxiety	-2
anxious	-2
apathetic	-3
apathy	-3
apeshit	-3
apocalyptic	-2
apologise	-1
apologised	-1
apologises	-1
apologising	-1
apologize	-1
apologized	-1
apologizes	-1



apologizing	-1
apology	-1
appalled	-2
appalling	-2
appease	2
appeased	2
appeases	2
appeasing	2
applaud	2
applauded	2
applauding	2
applauds	2
applause	2
appreciate	2
appreciated	2
appreciates	2
appreciating	2
appreciation	2

apprehensive -2

approval 2

approved 2

approves 2

ardent 1

arrest -2

arrested -3

arrests -2

arrogant -2

ashame -2

ashamed -2

ass -4

assassination -3

assassinations -3

asset 2

assets 2

assfucking -4

asshole -4

astonished	2
astound	3
astounded	3
astounding	3
astoundingly	3
astounds	3
attack	-1
attacked	-1
attacking	-1
attacks	-1
attract	1
attracted	1
attracting	2
attraction	2
attractions	2
attracts	1
audacious	3
authority	1

avert -1

averted-1

averts -1

avid 2

avoid -1

avoided -1

avoids -1

await -1

awaited -1

awaits -1

award 3

awarded 3

awards3

awesome 4

# Βιβλιογραφία

- Brailas, A. V. (2011), " Using Wikipedia in a Course Assignment: Implications for Wikipedia Literacy in Higher and Secondary Education". Conference Proceedings EDEN 2011 Open Classroom: *Never Waste a Crisis! Inclusive Excellence, Innovative Technologies and Transformed Schools as Autonomous Learning Organizations*, Athens, 27-29 October 2011.
- Burgers, C., & Beukeboom, C.J. (in press), "Stereotype transmission and maintenance through interpersonal communication: "The Irony Bias. *Communication Research*. doi: 10.1177/0093650214534975
- Craig McGarty, Vincent Y. Yzerbyt & Russell Spears (2002). "Stereotypes as Explanations:" *The Formation of Meaningful Beliefs about Social Groups. 1<sup>st</sup> hardback edition*. Cambridge University Press.
- Dijksterhuis, A., & van Knippenberg, A. (1999), "On the parameters of associative strength:" *Central tendency and variability as determinants of stereotype accessibility*. *Personality and Social Psychology Bulletin*, 25, 527–536
- Eckert, Penelope & Sally McConnell -Ginet, 2013, "Language and Gender." *2<sup>nd</sup> Edition*. Cambridge University Press.
- Franco, F. M., & Maass, A. (1999), "Intentional control over prejudice: "When the choice of the measure matters." *European Journal of Social Psychology*, 29(4), 469-477
- Gould E, Beylin A, Tanapat P, Reeves A, Shors TJ (1999), "Learning enhances adult neurogenesis in the hippocampal formation." *Nat Neurosci* 2:260–265.
- Head, Alison H. and Michael B. Eisenberg(2009), "*Lessons Learned: How College Students Seek Information in the Digital Age*", Project Information Literacy First Year Report with Student Survey Findings, University of Washington's Information School
- Head, Alison J. and Michael B. Eisenberg, "Truth Be Told: "How College Students Evaluate and Use Information in the Digital Age", Project Information Literacy Progress Report, University of Washington's Information School, November 1, 2010.

- Hellinger, M. & Busmann, H. (επιμ.) (2001-2003). "*Gender across Languages*". 3 τόμοι. Amsterdam: John Benjamins
- Higournet, Ch.(1970), "*Η γραφή*", Δαίδαλος Ι. Ζαχαρόπουλος.
- Holmes, J. & M. Meyerhoff, 2005. "*The Handbook of Language and Gender*." 1<sup>st</sup> Edition. Wiley-Blackwell.
- Jeff Howe, 2009. Crowdsourcing: "*Why the Power of the Crowd Is Driving the Future of Business*." 0 Edition. Crown Business.
- Judd, Charles M.; Park, Bernadette, "*Definition and assessment of accuracy in social stereotypes*." *Psychological Review*, Vol100(1), Jan 1993,109- 128.<http://dx.doi.org/10.1037/0033-295X.100.1.109>
- Operario, Don, and Susan T. Fiske (2001) "*Causes and consequences of stereotypes in organizations*." In Manuel London (ed.), *How People Evaluate Others in Organizations*. Mahwah, NJ: Lawrence Erlbaum, pp. 45-62.
- Otterbacher, Jahna, "*Linguistic Bias in Collaboratively Produced Biographies: Crowdsourcing Social Stereotypes?*" In: Ninth International AAAI Conference on Web and Social Media. 2015.
- Reagle, Joseph & Lauren Rhue(2011), "*Gender Bias in Wikipedia and Britannica*", *International Journal of Communication*, Vol. 5.
- Recasens, M., C. Danescu-Niculescu-Mizil & Dan Jurafsky(2013), "*Linguistic Models for Analyzing and Detecting Biased Language*".
- Ritzer, George 1996, *The McDonaldization of Society: "An Investigation Into the Changing Character of Contemporary Social Life*." Revised Edition. Pine Forge Press.
- Spender, D. (1980), "*Man Made Language*". London: Routledge.
- Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R, & Updegraff, J. A. (2000), "*Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight*". *Psychological Review*, 107, 441-429.
- Wagner, Cl., D. Garcia, M. Jadidi & M. Strohmaier (2015), "*It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia*".
- Αβούρης, Ν. (2000), "*Εισαγωγή στη επικοινωνία ανθρώπου υπολογιστή*". Διάυλος
- Γκασσούκα, Μ. (2014), "*Οδηγός χρήσης μη σεξιστικής γλώσσας στα διοικητικά έγγραφα*", Γενική Γραμματεία Ισότητας των Φύλων.

- Δάλλα, Χ. (2012). "Γυναίκες και άνδρες: μια ερευνητική προσέγγιση των διαφορών σε επίπεδο ψυχικής υγείας", Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών .
- Καρβέλας, Γ. (2002). «Πληθυντικός ευγενείας - Πώς η αρχαιοελληνική δημοκρατικότητα απωθήθηκε από την ξενική δουλοπρέπεια».
- Κλαίρης Χρ. & Μπαμπινιώτης Γ., "Γραμματική της Νέας Ελληνικής. Δομολειτουργική-Επικοινωνιακή", Ελληνικά Γράμματα, Αθήνα, 2005
- Κούρτη, Ε. (2003), "Η επικοινωνία στο Διαδίκτυο". Ελληνικά Γράμματα
- Κωστούλα-Μακράκη, Ν. (2001), "Γλώσσα και κοινωνία." Εκδ. Μεταίχμιο
- Λασκαράτου, Χ. (1999), "Ελευθερία και Πειθαρχία" στο: Κοπιδάκης, Μ.Ζ. (επιμ.) Ιστορία της Ελληνικής Γλώσσας, Αθήνα: ΕΛΙΑ.
- Παυλίδου, Θ.-Σ. 2006. Γλώσσα-γένος-φύλο: Προβλήματα, αναζητήσεις και ελληνική γλώσσα. Στο Γλώσσα-Γένος-Φύλο, 2η έκδ., επιμ. Θ.-Σ. Παυλίδου, 15-64. Θεσσαλονίκη: Ινστιτούτο Νεοελληνικών Σπουδών [Ίδρυμα Τριανταφυλλίδη].
- Πετρούνιας, Ε. (2002), "Νεοελληνική γραμματική και συγκριτική (αντιπαραθετική) ανάλυση". Εκδ. Ζήτη
- Σακκάτου, Α.(2011), "Σtereότυπα και προκαταλήψεις σε γηγενείς μαθητές και μαθητές πολιτισμικά διαφορετικών ομάδων." Μια ερευνητική προσέγγιση στερεοτυπικών και προκαταληπτικών συμπεριφορών», Μ.Σc., Πανεπιστήμιο Πατρών

## Online Articles

- Brailas, A. (2016), Crowdsourcing & Wikipedia: "A culture of participation." Ανακτήθηκε 20/11/2015 από το: [https://figshare.com/articles/Crowdsourcing\\_Wikipedia\\_A\\_culture\\_of\\_participation/3124996](https://figshare.com/articles/Crowdsourcing_Wikipedia_A_culture_of_participation/3124996).
- Cambridge University Press. 2016, "Language and Gender" Sociolinguistics Cambridge University Press. Ανακτήθηκε 20/11/2015 από το: <http://www.cambridge.org/us/academic/subjects/languages-linguistics/sociolinguistics/language-and-gender-2nd-edition>.

- Google Books. 2016, "*Language and Gender*" - Penelope Eckert, Sally McConnell-Ginet  
Google. [ONLINE] Available Ανακτήθηκε 15/10/2015 από  
το: [https://books.google.gr/books?id=JnZGzMkBaQkC&redir\\_esc=y](https://books.google.gr/books?id=JnZGzMkBaQkC&redir_esc=y)
- John Benjamins Publishing Company. 2016. John Benjamins Publishing  
Company.[ONLINE]Available Ανακτήθηκε 20/11/2015 από  
το: <https://benjamins.com/#catalog/books/impact.10/main>.
- Merriam-Webster. 2016. Bias | Definition of Bias by Merriam-Webster. Ανακτήθηκε  
15/5/2015 από το: [ONLINE] Available at: <http://www.merriam-webster.com/dictionary/bias>.
- Participation Inequality: The 90-9-1 Rule for Social Features. 2016. Participation  
Inequality: The 90-9-1 Rule for Social Features. Ανακτήθηκε 20/11/2015 από  
το: <https://www.nngroup.com/articles/participation-inequality/>.
- THE McDONALDIZATION. 2016, "*THE McDONALDIZATION*". Ανακτήθηκε 20/11/2015  
από το: <http://web.simmons.edu/~chen/nit/NIT'96/96-171-Larney.html>.
- Wikipedia. 2016. Wikimedia Foundation - Wikipedia, the free encyclopedia.[ONLINE]  
Ανακτήθηκε 15/05/2015 από το:  
[https://en.wikipedia.org/wiki/Wikimedia\\_Foundation](https://en.wikipedia.org/wiki/Wikimedia_Foundation)
- Wikipedia.2016. Wikipedia.[ONLINE] Available Ανακτήθηκε 20/11/2015 από  
το: <https://www.wikipedia.org/>.
- Wikipedia. 2016. Wikipedia:Size of Wikipedia - Wikipedia, the free encyclopedia.  
[ONLINE] Available. Ανακτήθηκε 20/11/2015 από το:  
[https://en.wikipedia/a.org/wiki/Wikipedia:Size\\_of\\_Wikipedia](https://en.wikipedia/a.org/wiki/Wikipedia:Size_of_Wikipedia)
- WIRED: WIRED. 2016. The Rise of Crowdsourcing Ανακτήθηκε 20/11/2015 από το:  
<http://www.wired.com/2006/06/crowds/>.
- Yale School of Management. 2016. Getting conscious about unconscious bias | Yale  
School of Management. [ONLINE] Available Ανακτήθηκε 20/11/2015 από  
το: <http://som.yale.edu/getting-conscious-about-unconscious-bias>
- YouTube. 2016. Wikipedia: "*Verifiability and Neutral point of view*". Ανακτήθηκε  
20/11/2015 από το: <https://www.youtube.com/watch?v=Fa0Nmv9qsd8>.
- Συμεωνίδης, Β.(2016). Βασίλης Συμεωνίδης, προσωπικός δικτυακός  
τόπος, φιλολογικά. Ανακτήθηκε 20/11/2015 από  
το: <http://users.sch.gr/symfo/ekpedefsi.htm>.



Fylopedia, (2016). "Φύλο και γλώσσα". Ανακτήθηκε 20/11/2015 από το: <http://www.fylopedia.uoa.gr/index.php?>

Μακρή-Τσιλιπάκου, Μ. (2010). Η "γυναικεία γλώσσα» και η γλώσσα των γυναικών". Στο Β. Καντσά, Β. Μουτάφη, & Ε. Παπαταξιάρχης (Επιμ.), Φύλο και κοινωνικές επιστήμες στη σύγχρονη Ελλάδα (σ. 119-146). Εκδόσεις Αλεξάνδρεια. Ανακτήθηκε 7/11/2016 από το:

[http://www.aegean.gr/gender-postgraduate/documents/Praktika\\_Synedriou/%CE%9A%CE%B5%CE%AF%CE%BC%CE%B5%CE%BD%CE%BF%20%CE%9C%CE%B1%CE%BA%CF%81%CE%AE-%CE%A4%CF%83%CE%B9%CE%BB%CE%B9%CF%80%CE%AC%CE%BA%CE%BF%CF%85.pdf](http://www.aegean.gr/gender-postgraduate/documents/Praktika_Synedriou/%CE%9A%CE%B5%CE%AF%CE%BC%CE%B5%CE%BD%CE%BF%20%CE%9C%CE%B1%CE%BA%CF%81%CE%AE-%CE%A4%CF%83%CE%B9%CE%BB%CE%B9%CF%80%CE%AC%CE%BA%CE%BF%CF%85.pdf)

Μπέκα, Α. (2008). "Η γλώσσα των δύο φύλων: Μοντέλα επικοινωνίας στη χρήση της γλώσσας βάσει του φύλου", μεταπτυχιακή εργασία. Ανακτήθηκε 7/11/2016 από το <http://www.eduportal.gr/fyla-glossa/>

Τσοκαλίδου, Π. (1996). "Το φύλο της γλώσσας, Οδηγός μη-σεξιστικής γλώσσας για το δημόσιο ελληνικό λόγο", Σύνδεσμος Ελληνίδων Επιστημόνων. Ανακτήθηκε 7/11/2016 από το:

[http://www.pi.ac.cy/pi/files/epimorfosi/isotita\\_fylou/dimgymn/arxeia/filoglosa/Biblia/TsakalidouP\\_to\\_fylo\\_tis\\_glosas.pdf](http://www.pi.ac.cy/pi/files/epimorfosi/isotita_fylou/dimgymn/arxeia/filoglosa/Biblia/TsakalidouP_to_fylo_tis_glosas.pdf)

Προτάσεις για την αποφυγή του 'Γλωσσικού Σεξισμού' στα κείμενα του ΕΚΠΑ από το Γραφείο ΘΕ.ΦΥΛ.ΙΣ. Ανακτήθηκε 7/11/2016 από το <http://school.med.uoa.gr/enimerosi/enimerwsh-apo-th-dioikhsh-ths-sxolhs/699-2013-05-08-08-49-09.html>

ΠΑΡΑΤΗΡΗΤΗΡΙΟ ΙΣΟΤΗΤΑΣ ΚΥΠΡΟΥ (2016).

Έρευνα για το γλωσσικό σεξισμό στη γλώσσα των εγγράφων της δημόσιας διοίκησης της Κυπριακής Δημοκρατίας. Ανακτήθηκε 7/11/2016 από το:

<https://www.ucy.ac.cy/genderstudies/documents/%CE%95%CE%A1%CE%95%CE%A5%CE%9D%CE%912.pdf>

Πανεπιστήμιο Κρήτης, Το φύλο στην Ιστορία της Εκπαίδευσης.

Ανακτήθηκε 7/11/2016 από το:

[https://opencourses.uoc.gr/courses/pluginfile.php/11523/mod\\_resource/content/0/1.%CE%97%20%CE%B3%CE%BB%CF%8E%CF%83%CF%83%CE%B1%20%CE%B4%CE%B5%CE%BD%20%CE%B5%CE%AF%CE%BD%CE%B1%CE%B9%20%CE%AC%CE%BC%CE%BF%CE%B9%CF%81%CE%B7%20%CF%84%CE%BF%CF%85%20%CF%86%CF%8D%CE%BB%CE%BF%CF%85.pdf](https://opencourses.uoc.gr/courses/pluginfile.php/11523/mod_resource/content/0/1.%CE%97%20%CE%B3%CE%BB%CF%8E%CF%83%CF%83%CE%B1%20%CE%B4%CE%B5%CE%BD%20%CE%B5%CE%AF%CE%BD%CE%B1%CE%B9%20%CE%AC%CE%BC%CE%BF%CE%B9%CF%81%CE%B7%20%CF%84%CE%BF%CF%85%20%CF%86%CF%8D%CE%BB%CE%BF%CF%85.pdf)

## Πηγές από το Διαδίκτυο

(n.d.). Ανάκτηση από <http://www.merriam-webster.com/dictionary/bias>.

(n.d.). Ανάκτηση από <http://som.yale.edu/getting-conscious-about-unconscious-bias>.

(n.d.). Ανάκτηση από [https://en.wikipedia.org/wiki/Wikimedia\\_Foundation](https://en.wikipedia.org/wiki/Wikimedia_Foundation).

(n.d.). Ανάκτηση από <https://www.nngroup.com/articles/participation-inequality/>.

(n.d.). Ανάκτηση από <http://www.wired.com/2006/06/crowds/>.

(n.d.). Ανάκτηση από [https://en.wikipedia.org/wiki/Wikipedia:Size\\_of\\_Wikipedia](https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia).

(n.d.). Ανάκτηση από <http://reportcard.wmflabs.org/>.

(n.d.). Ανάκτηση από [https://en.wikipedia.org/wiki/Reliability\\_of\\_Wikipedia](https://en.wikipedia.org/wiki/Reliability_of_Wikipedia).

(n.d.). Ανάκτηση από [https://en.wikipedia.org/wiki/Wikipedia:Identifying\\_reliable\\_sources](https://en.wikipedia.org/wiki/Wikipedia:Identifying_reliable_sources).

(n.d.). Ανάκτηση από <https://benjamins.com/#catalog/books/impact.10/main>.

(n.d.). Ανάκτηση από [https://books.google.gr/books?id=JnZGzMkBaqkC&redir\\_esc=y](https://books.google.gr/books?id=JnZGzMkBaqkC&redir_esc=y).

(n.d.). Ανάκτηση από <http://www.cambridge.org/us/academic/subjects/languages-linguistics/sociolinguistics/language-and-gender-2nd-edition?format=PB>.

(n.d.). Ανάκτηση από [https://is.muni.cz/el/1423/podzim2012/SAN230/um/the\\_handbook\\_of\\_language\\_and\\_gender.pdf](https://is.muni.cz/el/1423/podzim2012/SAN230/um/the_handbook_of_language_and_gender.pdf).

(n.d.). Ανάκτηση από <http://arxiv.org/abs/1501.06307>.

(n.d.). Ανάκτηση από <https://web.stanford.edu/~jurafsky/pubs/neutrality.pdf>.

(n.d.). Ανάκτηση από <http://stanfordnlp.github.io/CoreNLP/>.

(n.d.). Ανάκτηση από <http://ijoc.org/index.php/ijoc/article/view/777>.

(n.d.). Ανάκτηση από [https://en.wikipedia.org/wiki/Wikipedia:Tagging\\_pages\\_for\\_problems#Removing\\_tags](https://en.wikipedia.org/wiki/Wikipedia:Tagging_pages_for_problems#Removing_tags).

(n.d.). Ανάκτηση από <https://en.wikipedia.org/wiki/Wikipedia:Consensus>.

(n.d.). Ανάκτηση από <https://github.com/woorm/afinn-111>.

(n.d.). Ανάκτηση από [http://www2.imm.dtu.dk/pubdb/views/publication\\_details.php?id=6010](http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010).

(n.d.). Ανάκτηση από <http://design215.com/toolbox/wordlist.php>.

(n.d.). Ανάκτηση από <http://tools.perceptus.ca/text-wiz.php?ops=4>.

(n.d.). Ανάκτηση από <https://notepad-plus-plus.org/download/v6.9.1.html>.

(n.d.). Ανάκτηση από [https://en.wikipedia.org/w/index.php?title=Laurence\\_Olivier&action=history&year=2016&tagfilter=NPOV](https://en.wikipedia.org/w/index.php?title=Laurence_Olivier&action=history&year=2016&tagfilter=NPOV).

(n.d.). Ανάκτηση από <https://www.theguardian.com/technology/2015/sep/15/wikipedia-view-of-the-world-is-still-written-by-the-west>.

(n.d.). Ανάκτηση από <http://www.theguardian.com/politics/2015/may/10/who-are-the-new-mps-middle-aged-white-men-again>.

(n.d.). Ανάκτηση από [https://en.wikipedia.org/wiki/Wikimedia\\_Foundation](https://en.wikipedia.org/wiki/Wikimedia_Foundation).

(n.d.). Ανάκτηση από [http://www.greek-language.gr/greekLang/modern\\_greek/tools/lexica/glossology/show.html?id=141](http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/glossology/show.html?id=141).

(n.d.). Ανάκτηση από [http://www.greek-language.gr/greekLang/modern\\_greek/tools/lexica/glossology/show.html?id=141](http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/glossology/show.html?id=141).

(n.d.). Ανάκτηση από <http://www.sciencedirect.com/science/article/pii/S2090447914000550>.

(n.d.). Ανάκτηση από <http://www.sciencedirect.com/science/article/pii/S2090447914000550>.

(n.d.). Ανάκτηση από <http://www.sciencedirect.com/science/article/pii/S2090447914000550>.

(n.d.). Ανάκτηση από , <http://www.wjh.harvard.edu/~inquirer/homecat.htm>.

(n.d.). Ανάκτηση από <http://www.wjh.harvard.edu/~inquirer/homecat.htm>.

(n.d.). Ανάκτηση από <http://www.wjh.harvard.edu/~inquirer/homecat.htm>.