

<b>ΕΙΣΑΓΩΓΗ</b> .....	<b>1</b>
<b>ΥΠΟΛΟΓΙΣΤΙΚΑ ΜΟΝΤΕΛΑ ΑΝΑΓΝΩΣΗΣ</b> .....	<b>2</b>
DUAL ROUTE CASCADED .....	2
TRIANGLE MODEL.....	4
CDP+ .....	7
MULTIPLE TRACE MEMORY MODEL .....	9
ΣΥΓΚΡΙΝΟΝΤΑΣ ΤΑ ΥΠΟΛΟΓΙΣΤΙΚΑ ΜΟΝΤΕΛΑ .....	11
ΣΧΟΛΙΑΣΜΟΣ ΤΩΝ ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΜΟΝΤΕΛΩΝ .....	17
<b>ΠΡΟΒΛΕΨΙΜΕΣ ΚΑΙ ΜΗ-ΠΡΟΒΛΕΨΙΜΕΣ ΛΕΞΕΙΣ ΚΑΙ ΨΕΥΔΟΛΕΞΕΙΣ</b> .....	<b>18</b>
Η ΕΛΛΗΝΙΚΗ ΠΕΡΙΠΤΩΣΗ .....	19
<b>ΤΟ ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ ΑΝΑΓΝΩΣΗΣ ΠΡΟΒΛΕΨΙΜΩΝ ΚΑΙ ΜΗ- ΠΡΟΒΛΕΨΙΜΩΝ ΛΕΞΕΩΝ ΚΑΙ ΨΕΥΔΟΛΕΞΕΩΝ ΓΙΑ ΤΗΝ ΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ</b> .....	<b>21</b>
ΣΧΕΔΙΑΣΗ .....	22
ΑΝΑΠΑΡΑΣΤΑΣΗ ΟΡΘΟΓΡΑΦΙΚΟΥ ΚΑΙ ΦΩΝΟΛΟΓΙΚΟΥ ΕΠΙΠΕΔΟΥ.....	25
<i>Ορθογραφική Αναπαράσταση</i> .....	26
<i>Φωνολογική Αναπαράσταση</i> .....	29
ΕΚΠΑΙΔΕΥΣΗ .....	30
ΈΛΕΓΧΟΣ ΕΠΙΔΟΣΗΣ .....	35
ΑΠΟΤΕΛΕΣΜΑΤΑ .....	38
ΣΥΜΠΕΡΑΣΜΑΤΑ .....	41
<b>ΕΠΙΛΟΓΟΣ</b> .....	<b>47</b>
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b> .....	<b>48</b>
<b>ΠΑΡΑΡΤΗΜΑΤΑ</b> .....	<b>50</b>

## Εισαγωγή

Ένα από τα θέματα που ερευνώνται στη γνωστική ψυχολογία είναι και αυτό της μετατροπής της ορθογραφίας σε φωνολογία. Πως δηλαδή από το έντυπο κείμενο, μπορεί ο άνθρωπος να διαβάσει φωναχτά τις λέξεις που βλέπει. Αντικείμενο της έρευνας είναι η διαδικασία της μετατροπής αλλά και η διαδικασία της απόκτησης αυτής της ικανότητας με την πάροδο του χρόνου. Όπως γίνεται και με τις περισσότερες έρευνες αυτού του τύπου, οι ερευνητές καταλήγουν σε μοντέλα που θεωρούν ότι μπορούν να εξηγήσουν τα φαινόμενα που συναντούν στις έρευνές τους αλλά και στις έρευνες άλλων συναδέλφων τους που μελετούν το ίδιο αντικείμενο.

Τα τελευταία 30 χρόνια διάφορα μοντέλα έχουν προταθεί, καταργηθεί, τροποποιηθεί, επεκταθεί και γενικά έχουν δώσει μια δυναμική σε αυτόν τον τομέα έρευνας. Τα τελευταία χρόνια κάποια από τα μοντέλα αυτά βρήκαν το δρόμο της υπολογιστικής υλοποίησης που τους έδωσε τη δυνατότητα της αξιολόγησης ως προς την ορθότητά τους στην πρόβλεψη των φαινομένων της ανάγνωσης.

Παρακάτω θα εξεταστούν τέσσερα υπολογιστικά μοντέλα που αυτή τη στιγμή είναι ενεργά και χρησιμοποιούνται για τη μελέτη των φαινομένων πίσω από την ανάγνωση.

Στη συνέχεια θα παρουσιαστεί ένα υπολογιστικό μοντέλο ανάγνωσης λέξεων που βασίζεται σε ένα από τα τέσσερα αυτά μοντέλα, και το οποίο μπορεί να διαβάσει ελληνικές πολυσύλλαβες λέξεις και ψευδολέξεις. Τα αποτελέσματα της ανάλυσης του μοντέλου αυτού είναι πολύτιμα για την αξιολόγηση τόσο της θεωρητικής βάσης του μοντέλου, όσο και για την εφαρμογή της στην Ελληνική γλώσσα.

Επόμενα βήματα θα μπορέσουν να ολοκληρώσουν τη δουλειά αυτή και να παρουσιάσουν ένα ολοκληρωμένο μοντέλο ανάγνωσης λέξεων για την Ελληνική γλώσσα.

## Υπολογιστικά Μοντέλα Ανάγνωσης

Τα μοντέλα που υπάρχουν αυτή στιγμή και παρουσιάζουν το μεγαλύτερο ενδιαφέρον είναι το μοντέλο της διπλής διαδρομής διαδοχικής ενεργοποίησης (Dual Route Cascaded) των Coltheart, Rastle, Perry, Langdon, Ziegler (2001) και το συνδεδιστικό μοντέλο μονής διαδρομής των Harm και Seidenberg (1999, 2004). Το πρώτο αποτελείται ένα συνδυασμό συνδεδιστικού και συμβολικού μοντέλου, μιας και μέρος των διεργασιών του υλοποιούνται από ένα συνδεδιστικό μοντέλο

Πέρα από αυτά τα δύο, υπάρχει η πολύ ενδιαφέρουσα πρόταση των Zorzi, Perry και Ziegler (2007), ενός συνδεδιστικού μοντέλου διπλής διαδρομής (CDP+) που συνδυάζει στοιχεία από τα δύο προηγούμενα (μάλλον περισσότερα από το πρώτο).

Τέλος, ένα ακόμα συνδεδιστικό μοντέλο, έρχεται να συμπληρώσει της ομάδα των μοντέλων προς εξέταση. Το μοντέλο είναι των Ans, Carbonnel και Valdois (1998) και αφορά ένα σχεδιασμό μονής διαδρομής, δύο διεργασιών που μπορεί να αναγνώσει και πολυσύλλαβες λέξεις.

### ***Dual Route Cascaded***

Η θεωρία πίσω από το μοντέλο αυτό λέει ότι υπάρχουν δύο διαδρομές που ακολουθούνται κατά την επεξεργασία μιας έντυπης λέξης και τη μετατροπή της σε ήχο. Η πρώτη διαδρομή είναι η λεξικολογική, μέσω της οποίας αποκτούμε πρόσβαση σε ένα λεξικό, μέσα στο οποίο υπάρχουν αποθηκευμένες όλες οι γνωστές λέξεις και οι αντιστοιχίσεις τους σε ήχο. Η δεύτερη διαδρομή περνά μέσα από ένα σύστημα κανόνων, το οποίο δίνει τη δυνατότητα της μετατροπής της έντυπης λέξης, γράμμα – γράμμα σε ήχο, εφαρμόζοντας κανόνες μετατροπής γραμμάτων ή ομάδων γραμμάτων σε φωνήματα.

Υλοποιώντας τα παραπάνω, η λεξικολογική διαδρομή είναι ουσιαστικά ένα συνδεδιστικό δίκτυο πρόσθιας τροφοδότησης (feed forward) τοπικής αναπαράστασης, το οποίο έχει εκπαιδευτεί ώστε να παράγει έξοδο την φωνολογική έκφραση της έντυπης λέξης που δίδεται ως είσοδος. Η

φωνολογική αυτή έκφραση ενεργοποιεί τα αντίστοιχα φωνήματα στο σύστημα φωνημάτων (που αποτελεί την έξοδο του μοντέλου) και απαγορεύει την ενεργοποίηση όλων των άλλων φωνημάτων.

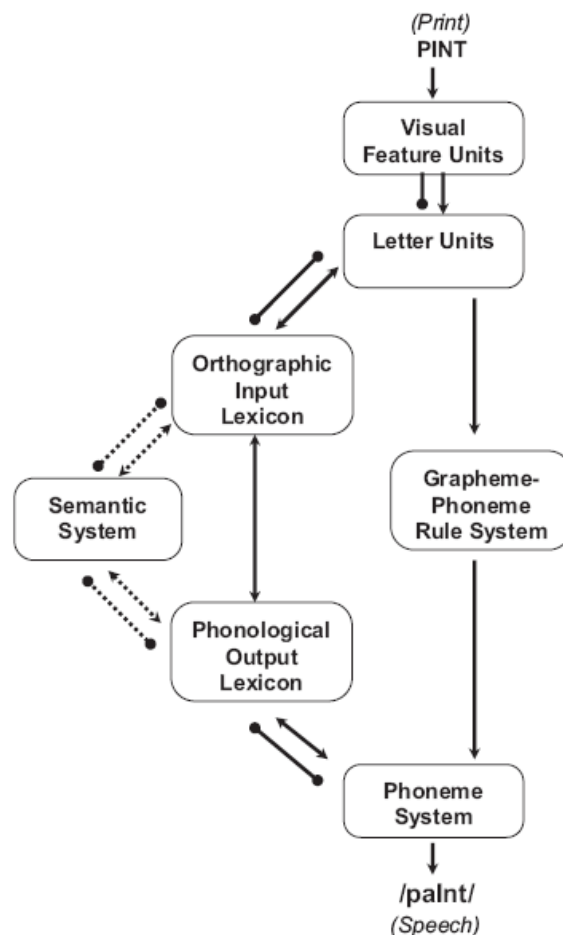
Το σύστημα μετατροπής των γραμμάτων σε φωνήματα υλοποιήθηκε ως ένα σύστημα παραγωγής με κανόνες που ενεργοποιούνται καθώς η λέξη που έχει δοθεί στην είσοδο του μοντέλου, παρουσιάζεται γράμμα – γράμμα (από τα αριστερά προς τα δεξιά) στο σύστημα αυτό. Κάθε κανόνας βρίσκει σε ποιο φώνημα αντιστοιχεί μια ομάδα γραμμάτων ή ένα γράμμα μόνο του και το ενεργοποιεί στην έξοδο του μοντέλου, που είναι το σύστημα φωνημάτων. Στην συγκεκριμένη διαδρομή δεν υπάρχει απαγόρευση στα υπόλοιπα φωνήματα.

Και οι δύο διαδρομές δέχονται την ίδια είσοδο (την έντυπη λέξη και τα χαρακτηριστικά των γραμμάτων που περιέχει) και ξεκινούν ταυτόχρονα να λειτουργούν, ανεξάρτητα η μία από την άλλη. Η συνολική ενεργοποίηση και απαγόρευση που προκύπτει από τις δύο αυτές διαδρομές, δίνει και το τελικό αποτέλεσμα που είναι ο ήχος της έντυπης λέξης ως αλληλουχία των φωνημάτων που υπάρχουν ενεργοποιημένα στο σύστημα φωνημάτων και που η στάθμη ενεργοποίησής τους είναι μεγαλύτερη από το κατώφλι που έχει οριστεί

Θα πρέπει να σημειωθεί ότι η λεξικολογική διαδρομή που περιγράφηκε αποτελεί τη μη σημασιολογική λεξικολογική διαδρομή, μιας και δεν υπάρχει αλληλεπίδραση με τη σημασιολογία της λέξης κατά τη μετατροπή της σε φωνολογική έκφραση, όπως προβλέπει το θεωρητικό μοντέλο της διπλής διαδρομής που είδαμε στην προηγούμενη παράγραφο.

Στο υπολογιστικό μοντέλο υπάρχει πρόβλεψη και για μια τρίτη διαδρομή, τη λεγόμενη σημασιολογική λεξικολογική διαδρομή, η οποία ξεκινά από το επίπεδο του ορθογραφικού λεξικού (το πρώτο τμήμα της λεξικολογικής διαδρομής) περνά μέσα από το σημασιολογικό σύστημα και καταλήγει στο φωνολογικό λεξικό (το δεύτερο τμήμα της λεξικολογικής διαδρομής) και μέσα από ενεργοποιήσεις και απαγορεύσεις μεταξύ των τμημάτων αυτών, προσδίδει σημασιολογική πληροφορία στην λέξη που έχει παρουσιαστεί στην είσοδο. Και οι δύο λεξικολογικές διαδρομές ξεκινούν ταυτόχρονα και

καταλήγουν στο ίδιο σημείο προσθέτοντας κάθε μια από τη δικιά της πλευρά ενεργοποίηση στο τμήμα που θα παραγάγει τη φωνητική ακολουθία που αντιστοιχεί στην έντυπη λέξη. Αυτή η διαδρομή, όμως, δεν έχει υλοποιηθεί ακόμα.



**Εικόνα 1.** Το υπολογιστικό μοντέλο διπλής διαδρομής διαδοχικής ενεργοποίησης (DRC Model).

### ***Triangle Model***

Το αρχικό μοντέλο υπήρξε αποτέλεσμα της έρευνας των Seidenberg και McClelland (1989) και αφορούσε ένα συνδυαστικό μοντέλο με τρία επίπεδα (ένα ορθογραφικό, ένα φωνολογικό και ένα σημασιολογικό) και τρία κρυφά επίπεδα να βρίσκονται μεταξύ τους κατά τη σύνδεση του ενός με το άλλο επίπεδο σε μια τριγωνική μορφή (από εκεί πήρε και το όνομά του). Όλες οι μονάδες (τεχνητοί νευρώνες) κάθε επιπέδου είναι συνδεδεμένες με όλες τις

μονάδες των άλλων επιπέδων, αφού πρώτα έχουν συνδεθεί με όλες τις μονάδες των ενδιάμεσων κρυφών επιπέδων. Κατά τη λειτουργία του μοντέλου, η έντυπη λέξη παρουσιάζονταν στο ορθογραφικό επίπεδο και το φωνολογικό επίπεδο έδινε σαν αποτέλεσμα την φωνολογικής έκφραση της έντυπης λέξης. Για τη δημιουργία της φωνολογικής έκφραση παίζει ρόλο τόσο η σύνδεση μεταξύ του ορθογραφικού επιπέδου με το φωνολογικό, όσο και η σύνδεση του σημασιολογικού επιπέδου με το φωνολογικό του ενεργοποιείται μέσω της σύνδεσής του με το ορθογραφικό.

Οι Seidenberg και McClelland δεν υλοποίησαν ολόκληρο το μοντέλο, αφήνοντας απ' έξω το σημασιολογικό επίπεδο. Στη συνέχεια οι Plaut, McClelland, Seidenberg και Patterson (1996) βελτίωσαν την αρχική υλοποίηση (σε επίπεδο αναπαραστάσεων εισόδου και εξόδου αλλά και σχεδιασμού του συνδετιστικού δικτύου) και επιπλέον, αν και δεν υλοποίησαν το σημασιολογικό επίπεδο, έδειξαν ότι τα αποτελέσματα του μοντέλου θα ήταν πολύ καλύτερα αν το σημασιολογικό επίπεδο υπήρχε στην υλοποίηση.

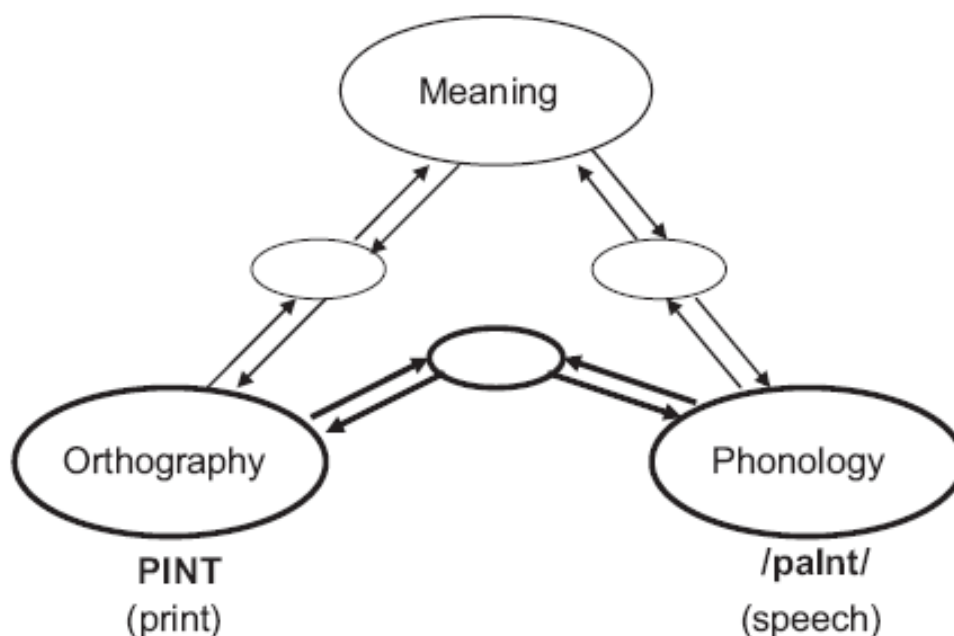
Έτσι το 2004 οι Harm και Seidenberg υλοποίησαν το πλήρες τριγωνικό μοντέλο, αν και ο κύριο σκοπός τους ήταν να ερευνήσουν κυρίως θέματα που έχουν να κάνουν με το διαχωρισμό εργασίας μέσα στο σύστημα της ανάγνωσης (η μεσολάβηση του φωνολογικού επιπέδου σε σχέση με την απ' ευθείας αντιστοίχιση μεταξύ ορθογραφικού και σημασιολογικού).

Στην υλοποίηση αυτή πρώτα υλοποιήθηκε η σύνδεση ανάμεσα στο σημασιολογικό και το φωνολογικό επίπεδο, ώστε να δημιουργηθεί η γνώση που κατέχουν οι άνθρωποι πριν αναπτύξουν την ικανότητα να διαβάζουν. Στη συνέχεια στο μοντέλο προστέθηκε το ορθογραφικό επίπεδο και συνδέθηκε τόσο με το σημασιολογικό όσο και με το φωνολογικό επίπεδο, δημιουργώντας έτσι το ολοκληρωμένο τριγωνικό μοντέλο. Επιπλέον και τροποποιώντας το αρχικό μοντέλο, χρησιμοποίησαν και απευθείας συνδέσεις μεταξύ του ορθογραφικού επιπέδου και του φωνολογικού αλλά και του σημασιολογικού που δεν περνούν μέσα από τα κρυφά επίπεδα ενδιάμεσα των επιπέδων αυτών. Ο λόγος ήταν η βελτίωση της γενίκευσης.

Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε μία παραλλαγή του αλγόριθμου Backpropagation που χρησιμοποιείται για την εκπαίδευση συνδεδεστικών μοντέλων που ενσωματώνουν δομές ελκυστών (attractors).

Θα πρέπει να σημειωθεί ότι με την προσθήκη του σημασιολογικού επιπέδου στην υλοποίηση του μοντέλου, αν και φαίνεται να γίνεται ένα μοντέλο διπλής διαδρομής, δεν είναι. Το διπλής διαδρομής μοντέλο σημαίνει ότι υπάρχουν δύο διαφορετικές διαδικασίες για την ονομασία λέξεων. Η μία αναλαμβάνει τις λέξεις που είναι ήδη γνωστές (λεξικολογική διαδρομή) και η άλλη αναλαμβάνει τις λέξεις που δεν είναι γνωστές π.χ. τις ψευδολέξεις (σύστημα κανόνων μετατροπής γραφημάτων σε φωνήματα). Στο τριγωνικό μοντέλο δεν υπάρχει ξεχωριστή διαδικασία για την ονομασία άγνωστων λέξεων, αλλά το μοντέλο αναλαμβάνει την ονομασία λαμβάνοντας υπόψη τις ήδη υπάρχουσες λέξεις με τις οποίες έχει εκπαιδευτεί και στις δομές των ελκυστών που έχουν δημιουργηθεί και που έχουν σχηματίσει γειτονίες λέξεων.

Επίσης είναι σημαντικό το ότι κάθε ένα από τα υπολογιστικά μοντέλα που αναφέρθηκαν ότι υλοποιούν το τριγωνικό μοντέλο, είχε διαφορετικό σκοπό στην υλοποίησή του.



**Εικόνα 2.** Το τριγωνικό υπολογιστικό μοντέλο (Triangle model)

## **CDP+**

Η δημιουργία αυτού του υπολογιστικού μοντέλου από τους Perry, Ziegler και Zorzi (2007) έρχεται να συνδυάσει τα καλύτερα χαρακτηριστικά (σύμφωνα με τους δημιουργούς του) των δύο παραπάνω προσεγγίσεων και να παρουσιάσει ένα μοντέλο που μπορεί να ανταπεξέλθει σε όλες τις δοκιμασίες που αφορούν τα μοντέλα ανάγνωσης, ακόμα και αυτά που δυσκόλευαν τους δύο αντιπάλους που προαναφέρθηκαν.

Αποτελεί μια συνδεδιστική υλοποίηση του μοντέλου της διπλής διαδρομής των Coltheart et. al. (2001) όπου η λεξικολογική διαδρομή είναι η ίδια, ενώ η διαδρομή μετατροπής των γραφημάτων σε φωνήματα, που στο μοντέλο της διπλής διαδρομής γίνεται μέσω ενός συνόλου κανόνων, έχει αντικατασταθεί από ένα συνδεδιστικό δίκτυο δύο επιπέδων φωνολογικής σύνθεσης (δίκτυο TLA). Το μόνο που δανείζεται από την τριγωνικό μοντέλο είναι η κοινή υπολογιστική πλατφόρμα που χρησιμοποιεί, δηλαδή την παράλληλη κατανεμημένη επεξεργασία.

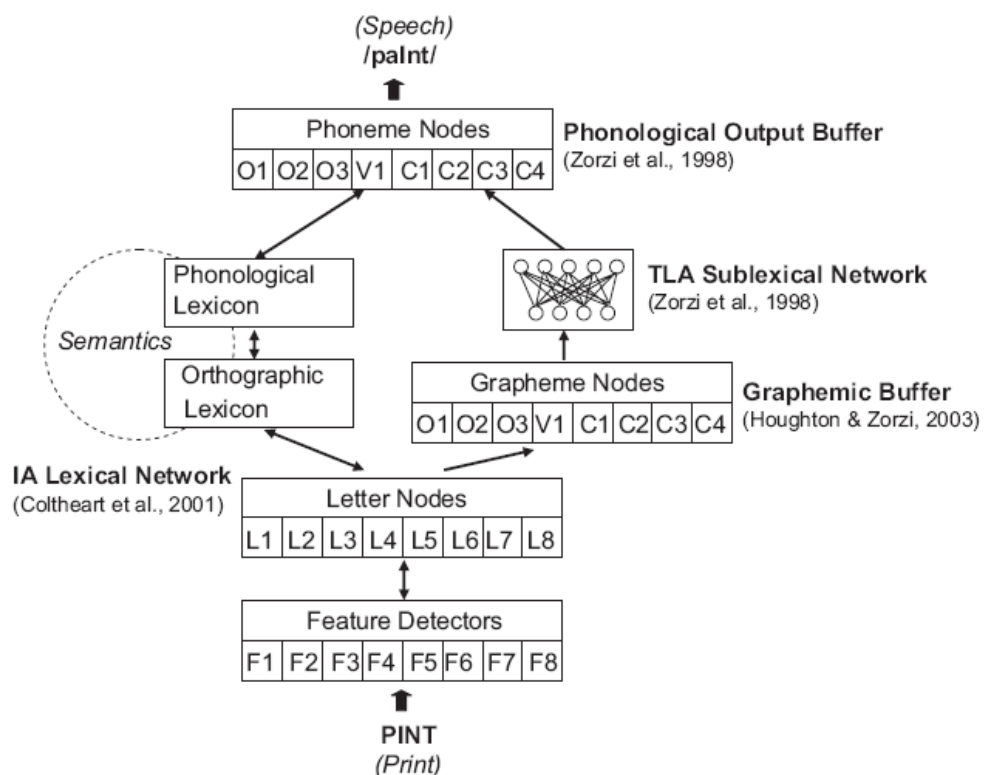
Συγκεκριμένα το μοντέλο υλοποιήθηκε ως εξής. Σαν είσοδο στο μοντέλο δίνεται ένα σύνολο από γράμματα που αντιστοιχούν στην έντυπη λέξη και ένα σύνολο από χαρακτηριστικά για τα γράμματα αυτά. Η είσοδο αυτή είναι η ίδια με το μοντέλο της διπλής διαδρομής. Η είσοδος αυτή μοιράζεται σε δύο διαδρομές που ξεκινούν ταυτόχρονα να την επεξεργάζονται. Η πρώτη διαδρομή είναι η λεξικολογική και ως υλοποίηση είναι ίδια με αυτή της διπλής διαδρομής. Υπάρχει ένα ορθογραφικό λεξικό το οποίο από την ενεργοποίηση που δέχεται από την είσοδο ενεργοποιεί μια αντιστοίχιση που υπάρχει σε ένα φωνολογικό λεξικό και είναι αυτό που παράγει σαν έξοδο την φωνολογική έκφραση της έντυπης λέξης ενεργοποιώντας με τη σειρά του φωνήματα στο φωνολογικό σύστημα. Πάντως και σε αυτήν την υλοποίηση, δεν έχει συμπεριληφθεί η σημασιολογική διαδρομή όπως και στο μοντέλο της διπλής διαδρομής.

Αντίστοιχα υπάρχει και μία δεύτερη διαδρομή η οποία εκτελεί μια παρόμοια δουλειά με το σύστημα κανόνων, που υπήρχε στο μοντέλο διπλής διαδρομής, και που μετέτρεπαν τα γραφήματα στο οποία αντιστοιχούν τα γράμματα της έντυπης λέξης, σε φωνήματα. Στην παρούσα υλοποίηση η διαδρομή



ονομάζετε υπό-λεξικόλογική (sublexical) και ξεκινά με ένα σύνολο από κόμβους που εκτελούν χρέη μεταφραστή των γραμμάτων της έντυπης λέξης σε γραφήματα. Τα γραφήματα αυτά δίνονται ως είσοδος σε ένα συνδεδειστικό δίκτυο δύο επιπέδων που έχει εκπαιδευτεί να τα μετατρέπει σε φωνήματα. Ο τρόπος με τον οποίο τα γράμματα μετατρέπονται σε φωνήματα είναι παρόμοιος με αυτόν στο μοντέλο της διπλής διαδρομής (παρουσιάζονται τα γράμματα ένα – ένα και κάθε γράμμα προστίθεται στο τέλος των ήδη παρουσιασθέντων δομώντας σιγά – σιγά ξανά την λέξη).

Οι έξοδοι και από τις δύο διαδρομές είναι ενεργοποιήσεις προς το σύστημα της φωνολογικής εξόδου. Το σύστημα αυτό αθροίζει τις ενεργοποιήσεις των δύο διαδρομών και στη συνέχεια για να αποφασίσει ότι έχει τελειώσει η διαδικασία ελέγχει αν για κάθε ένα από τα φωνήματα που η τιμή ενεργοποίησής τους είναι κάτω από το κριτήριο ενεργοποίησης για την ανάγνωση, η τιμή αυτή είναι η ίδια με την τιμή του προηγούμενου ελέγχου. Αν συμβαίνει αυτό και επίσης τουλάχιστον ένα φώνημα έχει τιμή ενεργοποίησης πάνω από το κριτήριο, τότε η εκτέλεση του μοντέλου σταματά δίνοντας έξοδο τα φωνήματα αυτά.



**Εικόνα 3.** Το συνδεδειστικό μοντέλο διπλής διαδρομής (CDP+)

## ***Multiple Trace Memory Model***

Το μοντέλο των Ans, Carbonnel και Valdois (1998), είναι το μοναδικό από τα τέσσερα μοντέλα που μελετάμε, που έχει υλοποιηθεί έτσι ώστε να μπορεί να επεξεργάζεται πολυσύλλαβες λέξεις. Είναι και αυτό συνδεδιστικό, αλλά η σχεδίαση του διαφέρει κατά πολύ από ότι έχουμε δει μέχρι τώρα.

Το κύριο στοιχείο του είναι το τμήμα της επεισοδιακής μνήμης που περιλαμβάνει και στο οποίο υπάρχουν αποθηκευμένα ίχνη των λέξεων που είναι ήδη γνωστές, αλλά όχι μόνο αυτές. Υπάρχουν επίσης και ίχνη τμημάτων λέξεων (συλλαβές), μέσα όμως στο ορθογραφικό πλαίσιο στο οποίο ανήκουν, αλλά και μεμονωμένα γραφήματα. Όλα αυτά, το σύστημα, έχει εκπαιδευτεί να τα αντιστοιχεί με συγκεκριμένα φωνήματα ή σύνολα από φωνήματα.

Η λογική της επεξεργασίας μιας έντυπης λέξης είναι η εξής. Η λέξη παρουσιάζεται στην είσοδο του μοντέλου, που αρχικά τίθεται σε «γενική» κατάσταση επεξεργασίας. Αν η λέξη αυτή υπάρχει ήδη ως ίχνος στην επεισοδιακή μνήμη τότε η φωνολογική της έκφραση εμφανίζεται στην φωνολογική έξοδο και η διαδικασία τερματίζεται. Αν δεν υπάρχει όμως ίχνος της λέξης αυτής, τότε η κατάσταση επεξεργασίας του μοντέλου αλλάζει σε «αναλυτική».

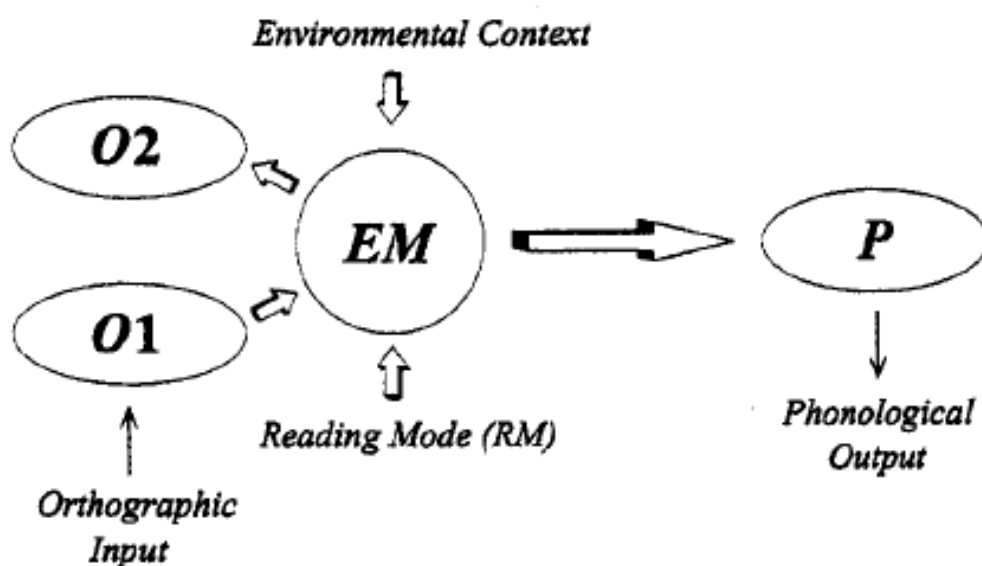
Σε αυτήν της κατάσταση ξεκινά μια διαδικασία τροφοδότησης του μοντέλου με τμήματα της λέξης εισόδου (αποτελούμενα από ένα ή περισσότερα γράμματα) ξεκινώντας από τα αριστερά προς τα δεξιά. Τα τμήματα αυτά ενεργοποιούν τα ίχνη των τμημάτων ή των γραφημάτων που υπάρχουν στην επεισοδιακή μνήμη και αυτά αντίστοιχα ενεργοποιούν τα φωνήματα με τα οποία έχουν αντιστοιχηθεί. Το αποτέλεσμα είναι η σύνθεση της φωνολογικής έκφρασης της έντυπης λέξης.

Έτσι το μοντέλο μπορεί να χαρακτηριστεί ως μοντέλο δύο διαδικασιών, όπως και το μοντέλο διπλής διαδρομής, με τη διαφορά του ότι οι διαδικασίες αυτές δεν εκτελούνται ταυτόχρονα αλλά εκτελείτε πρώτα η μια και αν αποτύχει τότε εκτελείται η δεύτερη.

Για να μπορέσει το μοντέλο να αναγνωρίσει αν μια λέξη υπάρχει ή όχι στην επεισοδιακή μνήμη και να αποφασίσει αν θα πρέπει να περάσει στην αναλυτική επεξεργασία ή όχι χρησιμοποιεί το εξής. Η επεισοδιακή μνήμη, δεχόμενη είσοδο από το ορθογραφικό επίπεδο προκαλεί ενεργοποίηση στην φωνολογική έξοδο αλλά και σε ένα δεύτερο ορθογραφικό επίπεδο, ανασυνθέτοντας την λέξη που δέχθηκε ως είσοδο. Αν οι αναπαραστάσεις στα δύο ορθογραφικά επίπεδα είναι ίδιες, τότε η λέξη ανακλήθηκε σωστά από την επεισοδιακή μνήμη και η διαδικασία θεωρείτε λήξασα. Η αναπαράσταση στην φωνολογική έξοδο αποτελεί την φωνολογική έκφραση της έντυπης λέξης.

Αν δεν είναι ίδιες τότε η αναπαράσταση στην φωνολογική έξοδο διαγράφεται και η κατάσταση επεξεργασίας αλλάζει σε «αναλυτική». Περνάμε δηλαδή στη δεύτερη διαδικασία, αυτή της σύνθεσης της φωνολογικής έκφρασης. Εκεί χρησιμοποιώντας ένα παράθυρο εστίασης, η είσοδος στο μοντέλο αλλάζει και τμηματικά τροφοδοτείται με τις συλλαβές από τις οποίες αποτελείται η λέξη. Η διαδικασία ελέγχου της ύπαρξης των ιχνών των τμημάτων αυτών στην επεισοδιακή μνήμη είναι η ίδια με αυτήν της προηγούμενης διαδικασίας. Τώρα όμως η έξοδος της επεισοδιακής μνήμης προς το δεύτερο ορθογραφικό επίπεδο είναι το τμήμα που ανακαλείται και συγκρίνεται με το τμήμα που βρίσκετε εκείνη τη στιγμή μέσα στο παράθυρο εστίασης στο ορθογραφικό επίπεδο εισόδου. Αν τα τμήματα είναι ίδια τότε η φωνολογική έξοδος είναι και αυτή σωστή και το παράθυρο εστίασης μετατοπίζεται προς τα δεξιά. Αν δεν είναι τότε το παράθυρο εστίασης διευρύνοντας έτσι και το τμήμα που δίνετε ως είσοδος στο μοντέλο. Η διαδικασία ξεκινά πάλι από την αρχή μέχρι να βρεθεί το τέλος της λέξης.

Θα πρέπει να σημειωθεί, ότι το μοντέλο έχει την ικανότητα της μάθησης κάθε λέξης που του παρουσιάζετε και του ζητείτε να διαβάσει, είτε την ξέρει ήδη (ενδυναμώνοντας έτσι το ίχνος της στην επεισοδιακή μνήμη), είτε δεν την ξέρει αλλά την συνθέτει (δημιουργώντας έτσι ένα νέο ίχνος στην επεισοδιακή μνήμη).



**Εικόνα 4.** Το μοντέλο πολλαπλών ιχνών (Multitrace Memory model)

### **Συγκρίνοντας τα υπολογιστικά μοντέλα**

Και για τα τέσσερα παραπάνω μοντέλα οι ερευνητές χρησιμοποίησαν προηγούμενες έρευνες σε ανθρώπους για να μπορέσουν να συγκρίνουν τα αποτελέσματα των μοντέλων τους. Μετρήθηκαν οι χρόνοι απόκρισης (ο χρόνος σε υπολογιστικούς κύκλους που μπορούν να μετατραπούν σε πραγματικό χρόνο) και τα ποσοστά λαθών που έδωσαν τα μοντέλα και στη συνέχεια συγκρίθηκαν με τα αντίστοιχα αποτελέσματα των ερευνών που χρησιμοποίησαν.

Η πολυετής έρευνα πάνω στη γνωστική λειτουργία της ανάγνωσης, έχει δημιουργήσει ένα μεγάλο όγκο από παρατηρήσεις φαινομένων. Τα μοντέλα λοιπόν θα πρέπει να προβλέπουν αυτά τα φαινόμενα (αν όχι όλα, τα κυριότερα). Τα φαινόμενα αυτά παρουσιάζονται παρακάτω και ακολουθεί μία σύγκριση της απόδοσης των μοντέλων ως προς αυτά. Τα κύρια φαινόμενα λοιπόν είναι:

1. Η επίδραση της συχνότητας – Αφορά το φαινόμενα του να διαβάζονται ταχύτερα λέξεις υψηλής συχνότητας σε σχέση με τις λέξεις χαμηλής συχνότητας.
2. Η λεξικολογική επίδραση – Είναι το φαινόμενο κατά το οποίο το διάβασμα είναι ταχύτερο για κανονικές λέξεις σε σχέση με ψευδολέξεις.
3. Η αλληλεπίδραση της συχνότητας και της κανονικότητας – Αν μια λέξη είναι κανονική αλλά χαμηλής συχνότητας θα διαβαστεί ταχύτερα σε σχέση με μια ανώμαλη λέξη αντίστοιχα χαμηλής συχνότητας. Δεν συμβαίνει το ίδιο όμως και για τις λέξεις υψηλής συχνότητας, όπου εκεί δεν εμφανίζεται το φαινόμενο αυτό
4. Η επίδραση της συνέπειας – Αφορά το φαινόμενο κατά το οποίο η ανάγνωση μιας λέξης είναι ταχύτερη εάν η λέξη αυτή περιέχει ένα κύριο σώμα που έχει την ίδια προφορά σε όλες τις λέξεις της γλώσσας (ανήκει στην ίδια οικογένεια) σε σχέση με το εάν το κύριο σώμα δεν έχει την ίδια προφορά σε όλες τις λέξεις.
5. Η επίδραση της θέσης της ανωμαλίας – Το φαινόμενο αυτό σχετίζεται με το φαινόμενο της επίδρασης της κανονικότητας για τις λέξεις χαμηλής συχνότητας (βλέπε παραπάνω). Η επίδραση της κανονικότητας μειώνεται ανάλογα με την θέση της ανωμαλίας στη λέξη, ξεκινώντας από αριστερά προς τα δεξιά.
6. Η επίδραση του μήκους της ψευδολέξης – Ο χρόνος απόκρισης σε μια ψευδολέξη είναι ανάλογος ως προς το μήκος της.

Πριν παρατεθεί ο πίνακας της σύγκρισης των μοντέλων θα πρέπει να γίνει μία επισήμανση. Για το τριγωνικό μοντέλο, κάθε μία από τις υλοποιήσεις του (1989, 1996, 1999, 2004) μελετούσε ένα συγκεκριμένο κομμάτι του θεωρητικού μοντέλου και παράλληλα εξέταζε ένα μέρος των φαινομένων της ανάγνωσης λέξεων. Έτσι οι επιδόσεις στα φαινόμενα αυτά καταγράφηκαν σε κάποια υλοποίηση του μοντέλου και δεν εξετάστηκαν στις επόμενες (με κάποιες εξαιρέσεις). Ο λόγος για τον οποίο έγινε αυτό, έχει να κάνει με τη

θεώρηση της έννοιας της δημιουργία υπολογιστικού μοντέλου από τους ερευνητές.

Σύμφωνα με αυτούς λοιπόν κάθε ένα από τα υλοποιημένα μοντέλα, δημιουργήθηκε όχι με σκοπό να ολοκληρώσει το θεωρητικό μοντέλο αλλά να μελετήσει ένα κομμάτι του. Αυτό έγινε ώστε να μπορέσει να περιοριστεί ο σκοπός του υπολογιστικού μοντέλου και έτσι να μπορέσουν να εξαχθούν χρήσιμα συμπεράσματα από αυτό (για το κομμάτι που επιλέχθηκε να υλοποιηθεί) μέσα σε ένα εύλογο χρονικό διάστημα πριν ο ερευνητής χάσει το ενδιαφέρον του ή πεθάνει.

Έτσι, η αναφορά παρακάτω στις επιδόσεις του μοντέλου αυτού, αφορά κυρίως το υπολογιστικό μοντέλο των Plaut, McClelland, Seidenberg και Patterson (1996). Τα μεταγενέστερα μοντέλα των Harm και Seidenberg ασχολήθηκαν κυρίως με την αναπτυξιακή δυσλεξία (1999) και την ενεργοποίηση του σημασιολογικού επιπέδου του θεωρητικού μοντέλου (2004). Όμως οι αλλαγές που παρουσίασαν στην υλοποίηση του μοντέλου (σύστημα ελκυστών στο φωνολογικό επίπεδο, διαφοροποίηση στην αναπαράσταση εισόδου και εξόδου) δεν έχει αξιολογηθεί ως προς την απόδοση του μοντέλου στα παραπάνω κριτήρια.

<div style="text-align: center;"><b>Μοντέλο</b></div> <div style="text-align: left;"><b>Φαινόμενο</b></div>	<b>DRC</b>	<b>Triangle</b>	<b>CDP+</b>	<b>Multiple Trace Memory Model</b>
<b>Η επίδραση της συχνότητας</b>	<p>Εμφανίζεται στο μοντέλο λόγω του ότι έχει οριστεί μία τιμή βασισμένη στη συχνότητα της λέξης και η οποία προστίθεται στην ενεργοποίηση των γραμμάτων της λέξης στο ορθογραφικό λεξικό. Μεγαλύτερη τιμή σημαίνει γρηγορότερη ενεργοποίηση της φωνολογικής αντιστοιχίας στο φωνολογικό λεξικό.</p>	<p>Εμφανίζεται στο μοντέλο. Το μοντέλο εκπαιδεύεται στις λέξεις ανάλογα με τη συχνότητα των λέξεων με αποτέλεσμα να ενδυναμώνονται περισσότερο οι συνδέσεις για τις λέξεις με υψηλή συχνότητα σε σχέση με τις χαμηλής συχνότητας.</p>	<p>Εμφανίζεται στο μοντέλο. Εφόσον χρησιμοποιείται το ίδιο λεξικολογικό σύστημα με αυτό του DRC ο λόγος εμφάνισης του φαινομένου είναι ο ίδιος. Η συχνότητα της λέξης χρησιμοποιείται για να προκαλέσει ταχύτερη ενεργοποίηση της λέξης.</p>	<p>Εμφανίζεται στο μοντέλο. Τα ίχνη των λέξεων που έχουν αποθηκευτεί στην επεισοδιακή μνήμη, δίνουν έξοδο ανάλογη ως προς τη συχνότητά τους, έτσι οι συχνές λέξεις δίνουν περισσότερη ενεργοποίηση στην έξοδο και η φωνολογική έκφραση δημιουργείται ταχύτερα.</p>
<b>Η λεξικολογική επίδραση</b>	<p>Εμφανίζεται στο μοντέλο. Οι ψευδολέξεις επεξεργάζονται μόνο από το σύστημα μετατροπής γραφημάτων σε φωνήματα, το οποίο είναι γενικά αργότερο από τη λεξικολογική διαδρομή που ακολουθούν οι κανονικές λέξεις.</p>	<p>Εμφανίζεται στο μοντέλο. Οι κανονικές λέξεις έχουν μικρότερους χρόνους ηρεμίας (settling times – είναι ο χρόνος που περνά μέχρι να υπάρχει σημαντική ενεργοποίηση σε ένα αναδρομικό δίκτυο) σε σχέση με τις ψευδολέξεις.</p>	<p>Εμφανίζεται στο μοντέλο. Οι ψευδολέξεις ακολουθούν διαφορετική διαδρομή σε σχέση με τις κανονικές κατά την οποία η σύνθεση της φωνολογικής τους έκφρασης είναι μια αργή διαδικασία.</p>	<p>Εμφανίζεται το φαινόμενο κυρίως λόγω του ότι αρχικά οι ψευδολέξεις επεξεργάζονται από το μοντέλο με την κατάσταση να είναι η «γενική» και στη συνέχεια και εφόσον αποτύχει η επεξεργασία, ξεκινά νέα επεξεργασία σε κατάσταση «αναλυτική».</p>

<p><b>Η αλληλεπίδραση της συχνότητας και της κανονικότητας</b></p>	<p>Εμφανίζεται στο μοντέλο. Οι ανώμαλες λέξεις που επεξεργάζονται από το σύστημα μετατροπής μέσω κανόνων μπορεί να φτάσουν νωρίτερα στην φωνολογική έξοδο και έτσι να δημιουργήσουν καθυστέρηση στην ονομασία της χαμηλής συχνότητας ανώμαλης λέξης που θα καθυστερήσει να έρθει από τη λεξικολογική διαδρομή.</p>	<p>Εμφανίζεται στο μοντέλο και η εξήγηση που δίνεται έχει να κάνει με το ότι χαμηλής συχνότητας ανώμαλη λέξη που δεν έχει ενδυναμωμένες συνδέσεις στο δίκτυο, κατά την ανάκλησή της θα δεχθεί μεγαλύτερη επιρροή από τις γειτονικές τις κανονικές λέξεις σε σχέση με μια ανώμαλη λέξη με υψηλότερη συχνότητα που οι συνδέσεις της είναι αρκετά ισχυρές ώστε να αποφύγουν την επιρροή.</p>	<p>Εμφανίζεται στο μοντέλο. Η διαδρομή της σύνθεσης της ανώμαλης λέξης μπορεί να δώσει νωρίτερα ενεργοποίηση στην φωνολογική έξοδο από την λεξικολογική διαδρομή και έτσι να δημιουργήσει καθυστέρηση στην ονομασία της.</p>	<p>Εμφανίζεται στο μοντέλο. Ο λόγος είναι η σύνδεση των αποθηκευμένων ιχνών στην επεισοδιακή μνήμη με την συχνότητα τους, ενισχύοντας έτσι το επίπεδο της ενεργοποίησης που δίνουν κατά την ανάκληση. Αν μια ανώμαλη λέξη ζητηθεί, θα δεχθεί επιρροή από τις γειτονικές κανονικές λέξεις και μόνο αν είναι υψηλής συχνότητας θα καταφέρει να την αποφύγει.</p>
<p><b>Η επίδραση της συνέπειας</b></p>	<p>Δεν εμφανίζεται στο μοντέλο, σύμφωνα νεώτερες μελέτες από τους Jared (2002) και Treiman (2003), αν και αρχικά είχε αφεθεί να εννοηθεί ότι εμφανίζεται.</p>	<p>Εμφανίζεται στο μοντέλο ως αποτέλεσμα της επιρροής των γειτονικών λέξεων όπως εξηγήθηκε και στο προηγούμενο φαινόμενο. Η συνέπεια είναι συνδεδεμένη με τη συχνότητα εμφάνισης μίας λέξης.</p>	<p>Εμφανίζεται στο μοντέλο. Παρότι πρόκειται για ένα μοντέλο διπλής διαδρομής, παρόμοιας φιλοσοφίας με το DRC, η χρήση ενός συνδεδετιστικού δικτύου για τη μετατροπή των γραφημάτων σε φωνήματα αντί ενός συνόλου κανόνων, έχει το πλεονέκτημα και να εμφανίζει το φαινόμενο και να το συνδέει με την επιρροή της οικογένειας της λέξης κατά την ενεργοποίησή της.</p>	<p>Εμφανίζεται στο μοντέλο λόγω του ότι η συνέπεια είναι συνδεδεμένη με την οικογένεια της λέξης και δέχεται επιρροές από αυτή ή τους εχθρούς της.</p>



<p><b>Η επίδραση της θέσης της ανωμαλίας</b></p>	<p>Εμφανίζεται στο μοντέλο, λόγω της σειριακής εμφάνισης των γραμμάτων στο μοντέλο κατά την επεξεργασία της λέξης από το σύστημα μετατροπής γραφημάτων σε φωνήματα. Όσο νωρίτερα στη λέξη υπάρχει η ανωμαλία τόσο μεγαλύτερη ενεργοποίηση θα μεταφέρει στη φωνολογική έξοδο, σε σχέση με μία ανωμαλία που βρίσκετε προς το τέλος της λέξης.</p>	<p>Δεν εμφανίζεται στο μοντέλο. Η λέξη τροφοδοτείται στο μοντέλο όλη μαζί και η επεξεργασία της γίνεται παράλληλα για όλα τα γράμματά της. Η παράλληλη αυτή ενεργοποίηση όλων των θέσεων των γραφημάτων δεν δίνει τη δυνατότητα εμφάνισης του φαινομένου.</p>	<p>Εμφανίζεται στο μοντέλο λόγω του τρόπου με τον οποίο εμφανίζονται τα γράμματα της ανώμαλης λέξης στο σύστημα μετατροπής γραφημάτων σε φωνήματα. Όσο νωρίτερα τόσο περισσότερη ενεργοποίηση θα εμφανίσει στην φωνολογική έξοδο και έτσι περισσότερη επιρροή θα έχει στην σωστή ονομασία της ανώμαλης λέξης.</p>	<p>Εμφανίζεται το φαινόμενο, αν και τα αποτελέσματα προέρχονται από ένα πείραμα με κατασκευασμένες ανώμαλες λέξεις μιας και η Γαλλική γλώσσα έχει ελάχιστες ανωμαλίες. Ο λόγος εμφάνισης δεν είναι ξεκάθαρος</p>
<p><b>Η επίδραση του μήκους της ψευδολέξης</b></p>	<p>Εμφανίζεται στο μοντέλο, λόγω της σειριακής εμφάνισης των γραμμάτων στο μοντέλο κατά την επεξεργασία της λέξης από το σύστημα μετατροπής γραφημάτων σε φωνήματα. Όσα περισσότερα γράμματα υπάρχουν τόσο περισσότερο θα αργήσει η διαδικασία μετατροπής να ολοκληρωθεί.</p>	<p>Δεν εμφανίζεται στο μοντέλο. Ο λόγος είναι ο ίδιος με αυτόν του προηγούμενου φαινομένου.</p>	<p>Εμφανίζεται στο μοντέλο. Όσο νωρίτερα εμφανίζεται η ανωμαλία στην ψευδολέξη τόσο περισσότερη ενεργοποίηση μεταφέρεται από το σύστημα μετατροπής στην φωνολογική έξοδο, μιας και τα γράμματα τροφοδοτούνται σειριακά και σιγά – σιγά.</p>	<p>Εμφανίζεται στο μοντέλο και είναι χαρακτηριστικό του μιας και μπορεί να επεξεργαστεί πολυσύλλαβες λέξεις. Τα αποτελέσματα λοιπόν δείχνουν ότι η καθυστέρηση στην ονομασία των λέξεων αυξάνει όσο αυξάνουν τα γράμματα και ακόμα περισσότερο, όσο αυξάνουν οι συλλαβές.</p>

## **Σχολιασμός των υπολογιστικών μοντέλων**

Κανένα από τα υπολογιστικά μοντέλα δεν έχει υλοποιήσει το κομμάτι της σημασιολογίας, εκτός από αυτών των Harm και Seidenberg (2004), αν και προβλέπεται από το θεωρητικό μοντέλο στο οποίο στηρίχθηκαν. Το ένα και μοναδικό μοντέλο που το υλοποίησε, ασχολήθηκε με την μελέτη της ανάπτυξης και δεν υπάρχουν δεδομένα ή αποτελέσματα ως προς τη συνεισφορά της σημασιολογίας στα φαινόμενα που χρησιμοποιούνται για τη σύγκριση των μοντέλων.

Το μοντέλο των πολλαπλών ιχνών στη μνήμη, είναι το μοναδικό που μπορεί να επεξεργαστεί λέξεις με περισσότερες από μία συλλαβές. Τα άλλα μοντέλα περιορίζονται σε μονοσύλλαβες λέξεις και ένα κύριος λόγος για τον οποίο το κάνουν (σύμφωνα με τους ερευνητές) είναι και το θέμα του τονισμού των λέξεων. Το μοντέλο των πολλαπλών ιχνών στη μνήμη δεν ασχολείται με το θέμα του τονισμού.

Το μοντέλο DRC κατηγορείται ότι έχει φτιαχτεί για να δώσει αποτελέσματα παρόμοια με συγκεκριμένες έρευνες.

Το μοντέλο CDP+ βασίζεται σε νεώτερες έρευνες σε σχέση με τα υπόλοιπα μοντέλα και σε κάποια σημεία διαφοροποιείται από αυτά (π.χ. ως προς την επίδραση της συχνότητας με την κανονικότητα όπου τα αποτελέσματα του Jared (2002) ήταν διαφορετικά από ότι των προηγούμενων ερευνών).

Το τριγωνικό υπολογιστικό μοντέλο φαίνεται να μην υπάρχει καθώς έχει υλοποιηθεί τμηματικά με διαφορετική υλοποίηση και για διαφορετικούς σκοπούς.

Τα υπολογιστικά μοντέλα DRC και CDP+, ακολουθούν την ίδια θεωρία της ύπαρξης δύο παράλληλων διαδρομών κατά την επεξεργασία των έντυπων λέξεων και τη δημιουργία των φωνολογικών αντιστοιχιών τους, ενώ δεν ασχολούνται με τις δύο παράλληλες διαδρομές κατά την πρόσβαση στη σημασία μιας έντυπης λέξης.

Το τριγωνικό μοντέλο ακολουθεί τη θεωρία της μονής διαδρομής, ως προς τους μηχανισμούς φωνολογικής έκφρασης μιας έντυπης λέξης, αλλά θεωρεί

ότι υπάρχουν και έχει υλοποιήσει δύο διαδρομές για την πρόσβαση στη σημασία της έντυπης λέξης. Οι διαδρομές αυτές είναι η οπτική (ή άμεση) και η φωνολογική. Η τελευταία υλοποίηση του μοντέλου (2004) ασχολήθηκε με αυτές τις διαδρομές και μελέτησε τον τρόπο ανάπτυξής τους και δημιουργίας ισορροπίας, κατά την επεξεργασία λέξεων, μεταξύ τους.

Τα μοντέλα DRC και CDP+ που χρησιμοποιούν ένα μηχανισμό μετατροπής γραφημάτων σε φωνήματα, παρουσιάζουν την έντυπη λέξη σε αυτούς τους μηχανισμούς με διαφορετική συχνότητα. Στο CDP+ η παρουσίαση γίνεται ταχύτερα από ότι στο DRC δείχνοντας και τη διαφορετικότητά τους ως προς την θεώρηση του ρόλου της οπτικής διαδρομής στη δημιουργία της φωνολογικής έκφρασης. Μεγαλύτερη η συνεισφορά για το DRC μικρότερη για το CDP+.

Και τα τέσσερα μοντέλα χρησιμοποιούν συνδεδεστικά δίκτυα για την αποθήκευση των γνωστών λέξεων. Μόνο στο μοντέλο των πολλαπλών ιχνών μνήμης, η αναπαράσταση των λέξεων είναι τοπική, ενώ στα υπόλοιπα είναι κατανεμημένη.

Το DRC, Triangle και CDP+ μοντέλο, μελετούν την ανάγνωση λέξεων στην αγγλική γλώσσα, ενώ τα δύο πρώτα έχουν χρησιμοποιηθεί και για την μελέτη άλλων γλωσσών πέρα των αγγλικών. Το Multitrace μελετά την ανάγνωση με γαλλικές λέξεις.

## **Προβλέψιμες και Μη-Προβλέψιμες Λέξεις και Ψευδολέξεις**

Έχουν γίνει προσπάθειες ώστε να δημιουργηθούν κανόνες που να επιτρέπουν την αντιστοίχιση της ορθογραφία με τη φωνολογία. Οι κανόνες υπάρχουν και όπως όλοι οι κανόνες οπουδήποτε και αν εφαρμόζονται, έχουν και τις εξαιρέσεις τους. Οι κανόνες αυτοί λοιπόν μπορούν να προβλέψουν το ποια θα είναι η φωνολογική έκφραση μια λέξης (ομαλές λέξεις), ενώ υπάρχουν και εξαιρέσεις που ορίζουν τις ανώμαλες λέξεις.

Όταν θα πρέπει να διαβαστεί μια ψευδολέξη, οι κανόνες πρόβλεψης θα μπορέσουν να δώσουν τη φωνολογική έκφραση της λέξης αυτής και η οποία δεν θα μοιάζει ποτέ με κάποια από τις εξαιρέσεις των κανόνων.

Η χρήση των κανόνων αυτών μπορεί να φανεί και στην πράξη, κοιτάζοντας το μοντέλο DRC. Εκεί το σκέλος της σύνθεσης της φωνολογικής έκφρασης βασίζεται σε ένα σύνολο τέτοιων κανόνων. Οποιαδήποτε ψευδολέξη δοθεί στο μοντέλο για ανάγνωση, θα εκφραστεί φωνολογικά μέσα από το σκέλος αυτό βάση των κανόνων (μιας και το λεξικολογικό σκέλος δεν θα επιστρέψει κάποια ισχυρή απάντηση για τη φωνολογία της λέξης που δεν υπάρχει στο λεξικό).

Όμως τα πράγματα δεν είναι τόσο απλά. Οι άνθρωποι τείνουν να προφέρουν τις λέξεις με τρόπο που αρκετές φορές δεν είναι αυτός που προβλέπουν οι κανόνες μετατροπής. Αυτό συμβαίνει όταν πρόκειται για ψευδολέξεις που θυμίζουν ανώμαλες λέξεις και ειδικά όταν οι λέξεις αυτές έχουν ένα μεγάλο αριθμό φιλικών γειτόνων (Friendly Neighbors) .

Αυτή η χρήση αναλογίας στην ανάγνωση (το να διαβάζεται κάτι κατ' αναλογία προς αυτά με τα οποία μοιάζει) δίνουν ένα προβάδισμα στα μοντέλα που χρησιμοποιούν συνδεδειστικά δίκτυα για την ανάγνωση, καθώς είναι τέτοια η φύση τους που η ανάγνωση βάση αναλογίας απλά συμβαίνει.

## ***Η Ελληνική Περίπτωση***

Μέχρι στιγμής δεν υπήρχε κάποια υλοποίηση ενός από τα προηγούμενα μοντέλα που να μελετά την ανάγνωση Ελληνικών λέξεων. Οι λόγοι που θα καθιστούσαν μια εφαρμογή τους στα Ελληνικά είναι διάφοροι και μερικοί από αυτούς παρατίθενται παρακάτω.

- Τα μοντέλα εκτός του Multitrace έχουν σχεδιαστεί να μελετούν την ανάγνωση μονοσύλλαβων λέξεων. Στα Ελληνικά οι μονοσύλλαβες (αλλά και οι δισύλλαβες) λέξεις είναι ελάχιστες σε σχέση με άλλες γλώσσες και είναι σχεδόν αναγκαστική ή χρήση πολυσύλλαβων λέξεων.

- Η μελέτη πολυσύλλαβων λέξεων θέτει το θέμα της μελέτης του τονισμού των λέξεων, θέμα που δεν έχει διερευνηθεί επαρκώς ώστε να συμπεριληφθεί σε ένα υπολογιστικό μοντέλο.
- Για τα μοντέλα που χρησιμοποιούν κανόνες μετατροπής γραφημάτων σε φωνήματα εμφανίζει μια δυσκολία η Ελληνική γλώσσα, καθώς ενώ θεωρείται ρηχή (υπάρχουν κανόνες για τη μετατροπή των γραφημάτων σε φωνήματα) υπάρχει μια εξαίρεση για την οποία δεν υπάρχει κανόνας. Η εξαίρεση είναι το σύμπλεγμα ΣιΦ (σύμφωνο – κάποιο i – φωνήεν) για το οποίο δεν υπάρχει γενικός κανόνας προφοράς.

Προσπαθώντας να γίνει μια πρώτη υλοποίηση ενός υπολογιστικού μοντέλου στα Ελληνικά, θα έπρεπε να ληφθούν υπόψη τα παραπάνω ώστε να γίνει επιλογή του μοντέλου που θα επιλεγεί προς υλοποίηση. Ιδανικά θα έπρεπε να υλοποιηθούν και τα τέσσερα μοντέλα και να συγκριθούν μεταξύ τους ως προς την απόδοση στην Ελληνική γλώσσα. Κάτι τέτοιο όμως δεν είναι εφικτό στα πλαίσια μιας και μόνο διπλωματικής εργασίας έστω και αν αυτή είναι μεταπτυχιακού επιπέδου. Η δυσκολία υλοποίησης ενός και μόνο μοντέλου, ειδικά όταν θα πρέπει να γίνουν σε αυτό οι απαραίτητες αλλαγές προσαρμογής στην Ελληνική γλώσσα, είναι πολλές και ξεκινούν από το επίπεδο της προγραμματιστικής υλοποίησης και φτάνουν ως το επίπεδο της αξιολόγησης και επαναπροσδιορισμού των παραμέτρων του μοντέλου.

Όλα αυτά έδειξαν προς το Triangle model ως το καταλληλότερο μοντέλο για μια πρώτη υλοποίηση για την Ελληνική γλώσσα και οι κυριότεροι λόγοι είναι οι εξής:

- Ο κώδικας υλοποίησης του συνδεδιστικού δικτύου του μοντέλου είναι ελεύθερα διαθέσιμος από τον δημιουργό του, M.Harm.
- Ο κώδικας υλοποίησης του μοντέλου επί του συνδεδιστικού αυτού δικτύου μας έγινε ελεύθερα διαθέσιμος από τον J.Zevin. Ο κώδικας αυτός προέρχεται από μια υλοποίηση του μοντέλου που έκανε ο ίδιος για τη μελέτη της επίδρασης της συνέπειας.
- Τα αποτελέσματα μελετών τόσο ανάγνωσης όσο και τονισμού λέξεων και ψευδολέξεων για την ελληνική γλώσσα δείχνουν τη

χρήση της αναλογίας από τους συμμετέχοντες στα πειράματα, κάτι που προωθεί τη χρήση συνδετιστικού μοντέλου (Protorapas, 2006; Protorapas, Gerakaki, Alexandri, 2007; Protorapas, Gerakaki, υπό έκδοση; Protorapas, Nomikou, 2009; Protorapas, Vlahou, υπό έκδοση).

- Μπορεί να επεκταθεί σε πολυσύλλαβες λέξεις και να μπορέσει (θεωρητικά) να προβλέψει τον τονισμό των λέξεων και των ψευδολέξεων τόσο από την ορθογραφία όσο και από την φωνολογία μια λέξης ή ψευδολέξης, μέσω του συστήματος ελκυστών στο φωνολογικό επίπεδο κάτι που δεν μπορούν να κάνουν τα υπόλοιπα μοντέλα.

## **Το Υπολογιστικό Μοντέλο Ανάγνωσης Προβλέψιμων και Μη-Προβλέψιμων Λέξεων και Ψευδολέξεων για την Ελληνική Γλώσσα**

Παρακάτω περιγράφεται το μοντέλο που υλοποιήθηκε για την ανάγνωση πολυσύλλαβων Ελληνικών λέξεων και ψευδολέξεων. Για την υλοποίησή του λήφθηκαν υπόψη όλες οι ιδιαιτερότητες που φέρει η Ελληνική γλώσσα και γραμματική, ώστε να δημιουργηθεί ένα μοντέλο αξιόπιστο και λειτουργικό που θα προσομοιάζει κατά τον καλύτερο δυνατό τρόπο την ανάγνωση των Ελληνικών. Το πρότυπο μοντέλο του Harm που χρησιμοποιήθηκε, τροποποιήθηκε ώστε να καλύπτει αυτές τις ιδιαιτερότητες με πρώτο μέλημα την δυνατότητα ανάγνωσης πολυσύλλαβων λέξεων. Ακολούθησε η δυνατότητα εφαρμογής του τονικού σημαδιού που φέρουν οι Ελληνικές λέξεις και η αντιστοίχιση του με την απόδοση έμφασης στο αντίστοιχο φώνημα κατά την παραγωγή της φωνολογικής εξόδου.

Τέλος, για την αντιμετώπιση του φαινομένου ΣίΦ, ερευνήθηκε μια σειρά από εναλλακτικές αναπαραστάσεις ορθογραφικής εισόδου, ικανές να αντιμετωπίσουν το πρόβλημα της αδυναμίας προ-συλλαβισμού των λέξεων. Οι αναπαραστάσεις αυτές αξιολογήθηκαν μέσω της απόδοσης (αριθμός και είδος λαθών) του μοντέλου που τις υλοποιούσε και επιλέχθηκε αυτή που έδινε

την καλύτερη απόδοση, χωρίς βέβαια να είναι σίγουρο ότι αυτή η αναπαράσταση είναι και βέλτιστη για την ανάγνωση Ελληνικών πολυσύλλαβων λέξεων και ψευδολέξεων.

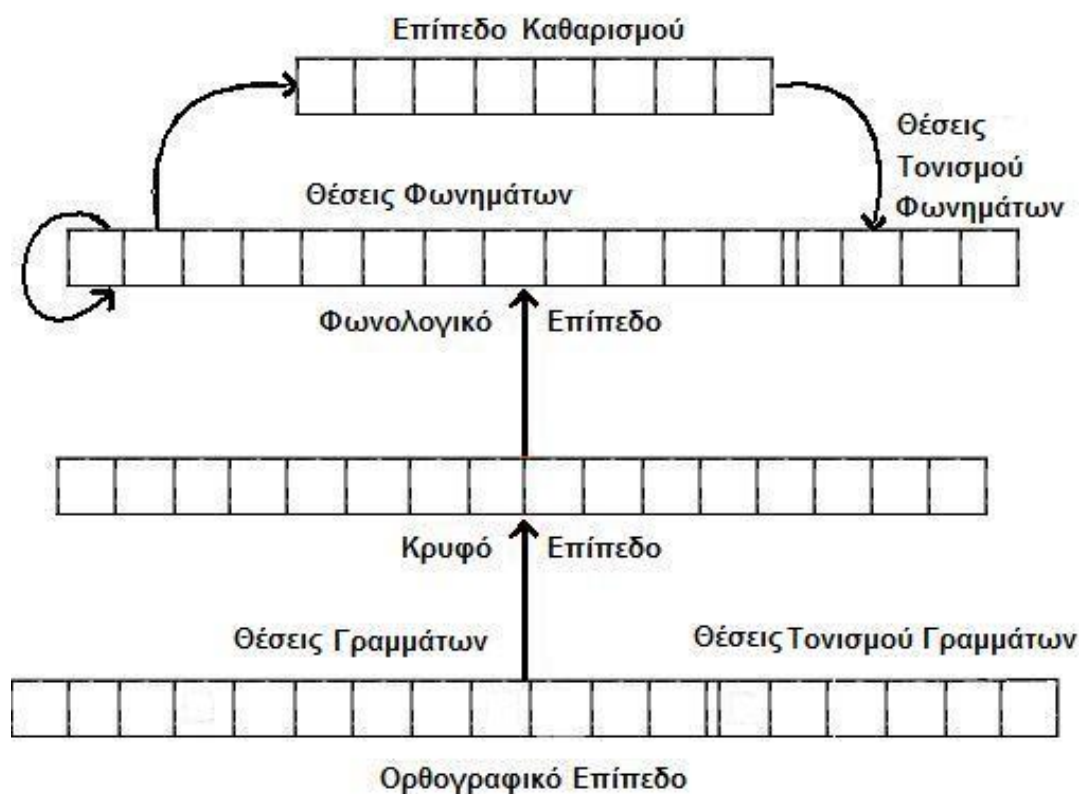
Το υπολογιστικό μοντέλο είναι σε θέση να διαβάζει ελληνικές λέξεις με 2 έως 5 συλλαβές και με μήκος από 4 έως 10 γράμματα. Αυτά όρια τέθηκαν με γνώμονα τη μείωση της υπολογιστικής ισχύος που χρειάζεται αλλά και τη μείωση του χρόνου εκπαίδευσης.

Υλοποιήθηκε χρησιμοποιώντας τον πηγαίο κώδικα των Zevin και Seidenberg (2006), μία υλοποίηση που μελέτησε φαινόμενα συνέπειας κατά την ανάγνωση ψευδολέξεων. Ο κώδικας είναι βασισμένος στη βιβλιοθήκη συναρτήσεων συνδετιστικών δικτύων «MikeNet» στην έκδοση 8, μία βιβλιοθήκη ειδικά σχεδιασμένη από τον Harm για υλοποίηση υπολογιστικών μοντέλων ανάγνωσης. Ο πηγαίος κώδικας τροποποιήθηκε κατάλληλα για να αντιμετωπίσει τις απαραίτητες αλλαγές στις αναπαραστάσεις εισόδου και εξόδου αλλά και να μπορέσει να χειριστεί την εισαγωγή του τονικού σημαδιού.

## **Σχεδίαση**

Το μοντέλο που υλοποιήθηκε δεν συμπεριέλαβε προς το παρόν το σημασιολογικό μέρος του θεωρητικού μοντέλου στο οποίο βασίστηκε. Υλοποιήθηκε μόνο το κομμάτι της μετατροπής της ορθογραφίας σε φωνολογία δηλαδή μόνο δύο από τις τρεις ενότητες του τριγωνικού μοντέλου.

Η σχεδίαση ακολούθησε τη βασική σχεδίαση των υπολογιστικών μοντέλων που έχουν υλοποιηθεί μέχρι σήμερα. Σε γενικές γραμμές υπάρχει το ορθογραφικό επίπεδο, στο οποίο παρουσιάζεται η ορθογραφική αναπαράσταση μιας λέξης ή ψευδολέξης, το κρυφό επίπεδο το οποίο είναι πλήρως συνδεδεμένο με το ορθογραφικό επίπεδο και το φωνολογικό επίπεδο στο οποίο προβάλλει το κρυφό επίπεδο και με το οποίο είναι πλήρως συνδεδεμένο επίσης. Επιπλέον υπάρχει και ένα επίπεδο κόμβων καθαρισμού (cleanup units) το οποίο είναι πλήρως αμφίδρομα συνδεδεμένο με το φωνολογικό επίπεδο, ενώ τέλος κάθε κόμβος του φωνολογικού επιπέδου είναι συνδεδεμένος με όλους τους υπόλοιπους κόμβους του φωνολογικού επιπέδου αλλά και με τον εαυτό του. Στην Εικόνα 5 φαίνεται η σχεδίαση του μοντέλου που έχει υλοποιηθεί.



**Εικόνα 5.** Η σχεδίαση του υπολογιστικού μοντέλου.

Το ορθογραφικό επίπεδο αποτελείται από 449 κόμβους. Αποτελούν τους κόμβους εισόδου του υπολογιστικού μοντέλου καθώς παίρνουν ενεργοποίηση από την αναπαράσταση της λέξης ή ψευδολέξης που θέλουμε το υπολογιστικό μοντέλο να «διαβάσει».

Κάθε ένας από τους κόμβους του ορθογραφικού επιπέδου είναι συνδεδεμένος με κάθε έναν από τους 500 κόμβους που αποτελούν το κρυφό επίπεδο με κατεύθυνση από το ορθογραφικό επίπεδο προς το κρυφό επίπεδο. Το κρυφό επίπεδο υπάρχει για να μπορέσει να «μάθει» τις κανονικότητες (regularities) που υπάρχουν στα δεδομένα εισόδου του υπολογιστικού μοντέλου, που στην προκειμένη περίπτωση είναι οι λέξεις οι οποίες παρουσιάζονται στην είσοδο κατά την εκπαίδευση, και να μπορέσει να γενικεύσει τη «γνώση» αυτή και να την εφαρμόσει σε περιπτώσεις που η είσοδος που δίνεται στο υπολογιστικό μοντέλο δεν έχει δοθεί ξανά κατά τη φάση της εκπαίδευσης.



Η έξοδος κάθε κόμβου του κρυφού επιπέδου αποτελεί την είσοδο για κάθε έναν κόμβο από τους 630 κόμβους του φωνολογικού επιπέδου μιας και υπάρχει πλήρης συνδεσιμότητα μεταξύ των δύο επιπέδων με κατεύθυνση από το κρυφό επίπεδο προς το φωνολογικό επίπεδο.

Το φωνολογικό επίπεδο όμως δεν δέχεται είσοδο μόνο από το κρυφό επίπεδο. Κατ' αρχήν υπάρχουν συνδέσεις από κάθε κόμβο του φωνολογικού επιπέδου προς κάθε άλλον κόμβο του ίδιου επιπέδου αλλά και προς τον εαυτό του. Επιπλέον το φωνολογικό επίπεδο είναι πλήρως συνδεδεμένο με ένα σύνολο 400 κόμβων που αποτελούν το επίπεδο καθαρισμού με τις συνδέσεις να έχουν αμφίδρομη κατεύθυνση. Το επίπεδο καθαρισμού έχει πάρει το όνομά του από το ρόλο που έχει στο υπολογιστικό μοντέλο, που είναι το να «καθαρίζει» την φωνολογική έξοδο ώστε αυτή να δίδετε μια σωστή φωνολογική αναπαράσταση ως έξοδος.

Συγκεκριμένα, η πλήρης και αμφίδρομη συνδεσιμότητα των δύο επιπέδων αλλά και η σύνδεση των φωνολογικών κόμβων με όλους τους άλλους και τον εαυτό τους, δημιουργεί ένα δυναμικό σύστημα ελκυστών που κατά τη πάροδο του χρόνου η κατάστασή του κάθε χρονική στιγμή δύναται να τροποποιήσει την αμέσως επόμενη κατάστασή του. Η εκπαίδευση ενός τέτοιου δυναμικού συστήματος, δημιουργεί ένα σύνολο από ελκυστές καταστάσεων κάθε ένας από τους οποίους έχει την ιδιότητα να έλκη προς μια συγκεκριμένη κατάσταση, καταστάσεις που βρίσκονται κοντά σε αυτή.

Έτσι το φωνολογικό επίπεδο, μετά την παρουσίαση της ορθογραφικής εισόδου και την έλευση του χρόνου καταλήγει σε μία έξοδο η οποία είναι η φωνολογική αναπαράσταση της λέξης που παρουσιάστηκε στην είσοδο και προέκυψε μετά την κατάληξη των πιθανών καταστάσεων της φωνολογικής εξόδου σε έναν ελκυστή που η κατάσταση που αντιπροσωπεύει είναι η κοντινότερη προς την κατάσταση εξόδου με την οποία έχει εκπαιδευτεί το φωνολογικό επίπεδο να δίνει όταν εμφανίζεται η συγκεκριμένη λέξη στην είσοδο του υπολογιστικού μοντέλου.

## **Αναπαράσταση Ορθογραφικού και Φωνολογικού Επιπέδου**

Η αναπαράσταση που χρησιμοποιήθηκε από το μοντέλο πρότυπο του Harm, αφορά την ορθογραφική και φωνολογική αναπαράσταση μόνο μίας συλλαβής, μιας και χρησιμοποιήθηκαν μονοσύλλαβες αγγλικές λέξεις, χωρίς να υπάρχει και η έννοια του τονισμού είτε ως τονικού σημαδιού, που δεν υφίσταται στην αγγλική γλώσσα, είτε ως έννοια απόδοσης έμφασης σε μια συγκεκριμένη συλλαβή, μιας και οι λέξεις ήταν μονοσύλλαβες.

Η λογική εξέλιξη του μοντέλου του Harm, ώστε να διαβάζει πολυσύλλαβες λέξεις, θα περιελάμβανε την χρήση αναπαραστάσεων όπου θα προέβλεπαν την αναπαράσταση πολυσύλλαβων λέξεων, έχοντας ικανό αριθμό κόμβων ώστε να καλύψουν τον αριθμό των συλλαβών και των γραμμάτων. Η αναπαράσταση πρόθεμα – φωνήεν – επίθημα (onset-vowel-coda) για κάθε συλλαβή θα επαναλαμβανόταν για όσες συλλαβές θα ήταν ικανό το μοντέλο να αναγνώσει. Έτσι για την ανάγνωση πεντασύλλαβων λέξεων, η παραπάνω τριπλέτα θα επαναλαμβανόταν πέντε φορές και θα δέχονταν τα γράμματα κάθε συλλαβής αναλυμένα στα συστατικά της. Αυτό θα συνέβαινε και στην ορθογραφική αναπαράσταση αλλά και στην φωνολογική, μιας και υπάρχει αντιστοιχία μεταξύ των δύο επιπέδων.

Όμως το φαινόμενο του ΣίΦ εγείρει έναν σημαντικό περιορισμό. Οι ελληνικές λέξεις και κυρίως οι ψευδολέξεις δεν μπορούν να προσυλλαβιστούν καθώς δεν είναι γνωστός από πριν ο τρόπος με τον οποίο θα προφερθεί το *i* στο ΣίΦ και δεν μπορεί κάποιος να το επιβάλει βάση κάποιου κανόνα. Αυτό ισχύει μόνο για την ορθογραφική αναπαράσταση, καθώς για την φωνολογική έξοδο του μοντέλου, στο σημείο δηλαδή που το μοντέλο έχει αποφασίσει πως θα συλλαβίσει μια λέξη, ο συλλαβισμός είναι απαραίτητος. Οι λέξεις που περιέχουν ΣίΦ θα συλλαβιστούν από το μοντέλο και η έξοδος που θα προκύψει θα μας δείξει και τον τρόπο με τον οποίο έχει διαβαστεί το *i*.

Έτσι από τις αναπαραστάσεις που περιγράφονται παρακάτω, μόνο η ορθογραφική είναι διαφορετική από την αναπαράσταση του πρότυπου μοντέλου του Harm, ενώ η φωνολογική απλώς έχει τροποποιηθεί ώστε να

μπορεί να αποδώσει φωνολογικά και λέξεις με περισσότερες από μία συλλαβές.

## Ορθογραφική Αναπαράσταση

Η αδυναμία προ-συλλαβισμού μιας λέξης, οδήγησε στην αναζήτηση εναλλακτικής αναπαράστασης στην οποία η συλλαβή δεν θα παίζει κάποιο ρόλο. Η απλή παράθεση των γραμμάτων μιας λέξης από τα αριστερά προς τα δεξιά (όπως εμφανίζεται και στην έντυπη μορφή της) έδωσε μη ικανοποιητικά αποτελέσματα τόσο ως προς τον αριθμό των λαθών ανάγνωσης, όσο και ως προς τον τύπο των λαθών (αρκετά γραμματικά συμπλέγματα δεν μπορούσαν διαβαστούν από το μοντέλο).

Ικανοποιητικά αποτελέσματα επιτεύχθηκαν με τη χρήση μιας αναπαράστασης στην οποία υπάρχει ομαδοποίηση των συμφώνων και των φωνηέντων μιας λέξης. Μετά από μια ανάλυση των λέξεων που περιλαμβάνονται στο σώμα εκπαίδευσης, ως προς την αλληλουχία των συμφώνων και των φωνηέντων, προέκυψε ότι μία σειρά από 40 θέσεις (slots) μπορεί να αναπαραστήσει κάθε λέξη από το σώμα αυτό με μοναδική προϋπόθεση τα σύμφωνα και τα φωνήεντα της λέξης να παραμένουν ομαδοποιημένα καθώς εναλλάσσονται. Οι 40 θέσεις αυτές και οι ομάδες των συμφώνων και φωνηέντων όπως εναλλάσσονται, φαίνονται παρακάτω μαζί με τους αριθμούς γραμμάτων ανά ομάδα:

ΣΣΣΦΦΦΦΣΣΣΣΣΦΦΦΦΦΣΣΣΣΦΦΦΦΣΣΣΦΦΦΦΣΣΦΦΣ  
3 5 4 6 4 5 3 4 2 2 1

Για να αναπαρασταθεί μια λέξη οι θέσεις κάθε ομάδας γραμμάτων γεμίζουν με τα αντίστοιχα γράμματα (σύμφωνα ή φωνήεντα) με κατεύθυνση από τα αριστερά προς τα δεξιά.

Για να μειωθεί ο αριθμός των κόμβων του συνδεδειστικού δικτύου που είναι απαραίτητα για την είσοδο σε αυτό του γράμματος που έχει τοποθετηθεί σε

κάθε θέση, έγινε μια ανάλυση των λέξεων στο σώμα εκπαίδευσης για να διαπιστωθεί ποια γράμματα δύναται να τοποθετηθούν σε κάθε μια από τις θέσεις (βλέπε Παράρτημα Α). Έτσι πέρα από το προφανές ότι στις θέσεις των συμφώνων δεν τοποθετούνται φωνήεντα αλλά και του αντιστρόφου, προέκυψε για κάθε θέση μια λίστα από γράμματα που είναι έγκυρα για τη θέση αυτή. Για τη μετατροπή της παραπάνω αναπαράστασης σε είσοδο για το συνδεδεστικό δίκτυο κάθε θέση μετατράπηκε σε ένα δυαδικό διάνυσμα με μήκος που αντιστοιχεί στον αριθμό των γραμμάτων που είναι έγκυρα για τη θέση αυτή. Το διάνυσμα αντιστοιχεί σε κόμβους εισόδου, ένας από τους οποίους παίρνει ως είσοδο την τιμή ένα και αντιστοιχεί στο γράμμα που έχει τοποθετηθεί στη θέση αυτή. Όλοι οι άλλοι κόμβοι δέχονται ως είσοδο το 0. Αυτά τα μεταβλητά σε μήκος διανύσματα με τις αντίστοιχες τιμές 0 ή 1 αποτελούν την ορθογραφική είσοδο στο συνδεδεστικό δίκτυο.

Επιπλέον των 40 θέσεων που υπάρχουν για τα γράμματα μιας λέξης, ένα σύνολο από 22 θέσεις προσάπτεται σε αυτές ώστε να αναπαρασταθεί το τονικό σημάδι. Οι 22 θέσεις αντιστοιχούν στις 5 ομάδες φωνηέντων που υπάρχουν στην αναπαράσταση των γραμμάτων. Το τονικό σημάδι αναπαρίσταται με την τοποθέτηση του γράμματος που το φέρει στην αντίστοιχη θέση που υπάρχει και στην αναπαράσταση των γραμμάτων. Και εδώ κάθε θέση έχει μια έγκυρη λίστα γραμμάτων την οποία μπορεί να δεχθεί. Η λίστα αυτή δημιουργεί ένα δυαδικό διάνυσμα το οποίο αποτελεί και την είσοδο για τους κόμβους εισόδου του συνδεδεστικού δικτύου που αντιπροσωπεύουν τη θέση αυτή. Όπως και πριν, αν ένα γράμμα έχει τοποθετηθεί σε μία θέση το διάνυσμα της έχει όλες του τις τιμές 0 εκτός από αυτή του αντίστοιχου γράμματος που τοποθετήθηκε στη θέση και που παίρνει την τιμή 1. Το σύνολο των άνισων σε μήκος διανυσμάτων, αποτελεί την είσοδο του συνδεδεστικού δικτύου ως προς το τονικό σημάδι. Από τις 22 θέσεις του τονικού σημαδιού μόνο μια μπορεί να δεχθεί γράμμα μιας και μόνο ένα φωνήεν τονίζεται σε μία λέξη, άρα και από το συνολικό διάνυσμα του αντιστοιχεί και στις 22 θέσεις, μόνο μία τιμή μπορεί να είναι 1.

Τα δύο διανύσματα των γραμμάτων και του τονικού σημαδιού δίνονται ως είσοδο στο συνδεδεστικό δίκτυο ταυτόχρονα σε ξεχωριστούς κόμβους το καθένα.

Παρακάτω φαίνεται ο τρόπος με τον οποίο θα αναπαρασταθεί ορθογραφικά η λέξη «άδεια»:

Γράμματα:

\_\_\_|α\_\_\_\_|δ\_\_\_\_|ε|α\_\_\_|\_\_\_\_\_|\_\_\_\_\_|\_\_\_\_\_|\_\_\_\_\_|\_\_\_\_\_|\_\_\_\_\_|\_\_\_\_\_|  
Σ Σ Σ|ΦΦΦΦΦ|Σ Σ Σ Σ Σ|ΦΦΦΦΦ|Σ Σ Σ Σ Σ|ΦΦΦΦΦ|Σ Σ Σ|ΦΦΦ|Σ Σ|ΦΦ|Σ

Τονικό σημάδι:

α \_\_\_\_\_

Ο παραπάνω τρόπος αναπαράστασης αν και λύνει τα θέματα που προκύπτουν από τη χρήση της Ελληνικής γλώσσας, δημιουργεί ένα άλλο πρόβλημα που έχει να κάνει με την βέλτιστη χρήση των συνδεδιστικών δικτύων. Τα συνδεδιστικά δίκτυα, προκειμένου να παράγουν τη συμπεριφορά για την οποία προτιμώνται (ανακάλυψη κανονικοτήτων στα δεδομένα εισόδου και γενίκευση των ανακαλύψεων αυτών σε νέα δεδομένα εισόδου) ζητούν να εφαρμόζονται κάποιες προϋποθέσεις. Μία από αυτές τις προϋποθέσεις είναι και η θέση των αντικειμένων που χειρίζεται το συνδεδιστικό δίκτυο (στην περίπτωση του υπολογιστικού μοντέλου τα αντικείμενα είναι τα γράμματα των λέξεων) που θα πρέπει να βρίσκονται πάντα στην ίδια θέση αν θέλουμε το δίκτυο να ανακαλύψει όλες τις κανονικότητες που σχετίζονται με αυτό. Στην περίπτωση του υπολογιστικού μοντέλου, τα γράμματα τοποθετούνται όλα μαζί σε θέσεις ανάλογα με την θέση τους στη λέξη αλλά και ανάλογα με τα γειτονικά τους γράμματα (σύμφωνα ή φωνήεντα). Με αυτόν τον τρόπο η τοποθέτησή τους σε διαφορετικές θέσεις κάνει την ανακάλυψη των οποιονδήποτε κανονικοτήτων από το συνδεδιστικό δίκτυο να είναι συνδεδεμένη με τη διαφορετική θέση του γράμματος και να μην μεταφέρεται στο ίδιο γράμμα σε μια άλλη θέση. Το πρόβλημα αυτό είναι γνωστό ως πρόβλημα της διασποράς (βλέπε και Plaut et. al. 1996).

Η ορθογραφική αναπαράσταση που περιγράφηκε παραπάνω δεν αποτελεί λύση του προβλήματος αυτού. Το πρόβλημα υφίσταται και το μοναδικό τεχνούργημα για τη μείωσή του είναι η ομαδοποίηση των γραμμάτων σε

εναλλαγές συμφώνων και φωνηέντων. Η λύση του προβλήματος θα απέφερε την συγκέντρωση όλων των κανονικότητων που μπορεί να ανακαλύψει το συνδεδιστικό δίκτυο για ένα και μόνο αντικείμενο (γράμμα) σε μια και μόνο θέση. Όμως κάτι τέτοιο θα μείωνε το ρόλο της θέσης ενός γράμματος στη λέξη και οποιαδήποτε κανονικότητα θα μπορούσε να εξαχθεί από αυτή θα χάνονταν. Με την ομαδοποίηση των γραμμάτων σε σύμφωνα και φωνήεντα εναλλάξ όπως εμφανίζονται στην λέξη πετυχαίνουμε να υπάρχει μια σχετική σταθερή θέση για κάποια γράμματα στην είσοδο του συνδεδιστικού δικτύου και επιπλέον εισάγουμε και το ρόλο της θέσης ενός γράμματος στη λέξη αλλά και του είδους του (σύμφωνο – φωνήεν) δίνοντας έτσι τη δυνατότητα στο δίκτυο να ανακαλύψει και επιπλέον κανονικότητες κατά την εκπαίδευσή του.

### **Φωνολογική Αναπαράσταση**

Η φωνολογική αναπαράσταση είναι παρόμοια με αυτή του πρότυπου μοντέλου του Harn καθώς δεν επηρεάζεται από το φαινόμενο ΣίΦ. Η έξοδος χωρίζεται σε συλλαβές και κάθε συλλαβή αποτελείται από μια τριπλέτα, το πρόθεμα, το φωνήεν και το επίθημα. Το πρόθεμα αποτελείται από τρεις θέσεις στις οποίες τοποθετούνται τα φωνήματα πριν από το φωνήεν της συλλαβής. Το επίθημα αποτελείται από 2 θέσεις στις οποίες τοποθετούνται τα φωνήματα μετά το φωνήεν της συλλαβής. Κάθε μία θέση από τις συνολικά 30 (6 θέσεις για κάθε συλλαβή για 5 συλλαβές) περιέχει ένα σύνολο από 18 φωνητικά χαρακτηριστικά (βλέπε Παράρτημα Α) σαν ένα διάνυσμα 18 δυαδικών τιμών. Κάθε φώνημα που τοποθετείται σε μία θέση θέτει τις τιμές των φωνητικών χαρακτηριστικών του στην τιμή 1.

Παρακάτω φαίνεται πως θα αναπαρασταθεί φωνολογικά η λέξη «άδεια» και στις δύο ομογραφικές μορφές της:

Φωνήματα:

\_\_\_ a \_\_|\_\_ ð i \_\_|\_\_\_ a \_\_|\_\_\_\_\_|\_\_\_\_\_  
Σ Σ Σ φ Σ Σ| Σ Σ Σ φ Σ Σ| Σ Σ Σ φ Σ Σ|Σ Σ Σ φ Σ Σ|Σ Σ Σ φ Σ Σ

Τονισμός:

a            \_            \_            \_            \_

Φωνήματα:

\_\_\_ a \_\_|\_ ð i a \_\_|\_\_\_\_\_|\_\_\_\_\_|\_\_\_\_\_  
Σ Σ Σ φ Σ Σ| Σ Σ Σ φ Σ Σ| Σ Σ Σ φ Σ Σ|Σ Σ Σ φ Σ Σ|Σ Σ Σ φ Σ Σ

Τονισμός:

a            \_            \_            \_            \_

Κοιτάζοντας την ορθογραφική και τις φωνολογικές αναπαραστάσεις της λέξης «άδεια», είναι φανερό ότι η προφορά μιας λέξης δεν υπόκειται πια σε περιορισμούς συλλαβισμού. Η ορθογραφική αναπαράσταση είναι ανεξάρτητη και ο τρόπος με τον οποίο θα προφερθεί μια λέξη εξαρτάται πια από την εκπαίδευση του δικτύου, τους ελκυστές που θα δημιουργηθούν από αυτή και όταν σε επόμενο στάδιο υλοποιηθεί το σημασιολογικό κομμάτι του θεωρητικού μοντέλου, από τη σημασιολογία της λέξης αλλά και το πλαίσιο στο οποίο τίθεται.

## **Εκπαίδευση**

Για την εκπαίδευση του υπολογιστικού μοντέλου χρησιμοποιήθηκε ένα σώμα 120.745 λέξεων με 2 έως 5 συλλαβές και με μήκος από 4 έως 10 γράμματα. Η επιλογή έγινε από ένα σύνολο 30.000.000 περίπου λέξεων από γραπτά κείμενα του Εθνικού Θησαυρού Ελληνικής Γλώσσας (ΕΘΕΓ, διαθέσιμο διαδικτυακά από τη διεύθυνση hnc.ilsr.gr). Για την δημιουργία των αντίστοιχων φωνολογικών αναπαραστάσεων των λέξεων, χρησιμοποιήθηκε

λογισμικό μετάφρασης γραφημάτων σε φωνήματα που έχει αναπτυχθεί για εφαρμογές μετατροπής κειμένου σε λόγο (Chalamandaris, Raptis, & Tsiakoulis, 2005). Ο αριθμός αυτός είναι ο μεγαλύτερος που χρησιμοποιήθηκε ποτέ για εκπαίδευση υπολογιστικού μοντέλου ανάπτυξης (πάνω από μια τάξη μεγέθους) για οποιαδήποτε άλλη γλώσσα.

Ο μεγάλος αριθμός λέξεων προέκυψε από την ανάγκη να συμπεριληφθούν λέξεις αρκετών συλλαβών για την καλύτερη μελέτη των φαινομένων που προκύπτουν κατά την ανάγνωση και οφείλονται στο μήκος της λέξης. Χρειάστηκε όμως να περιοριστεί το μήκος των γραμμάτων τους όμως γιατί λέξεις με περισσότερα γράμματα χρειάζονται περισσότερες θέσεις στην ορθογραφική αναπαράσταση με αποτέλεσμα να αυξάνεται κατά πολύ ο αριθμός των συνδέσεων μεταξύ των κόμβων του συνδετιστικού δικτύου και να προκαλεί προβλήματα διάθεσης υπολογιστικών πόρων για την εκπαίδευσή του. Ο περιορισμός αυτός του μήκους δεν δημιουργεί προβλήματα ως προς την κάλυψη των υπόλοιπων χαρακτηριστικών των λέξεων καθώς χρησιμοποιήθηκαν όλες οι λέξεις αυτού του μήκους και όχι κάποιο υποσύνολο.

Σε σχέση με το υπολογιστικό μοντέλο των Harm & Seidenberg (1999, 2004) σε αυτό το υπολογιστικό μοντέλο δεν έχει γίνει εκπαίδευση του φωνολογικού επιπέδου πριν από την εκπαίδευση της αντιστοίχισης ορθογραφικού – φωνολογικού επιπέδου. Ο κύριο λόγος ήταν ο συνολικός χρόνος εκπαίδευσης που θα αυξάνονταν κατά πολύ και το ότι τα αποτελέσματα των υπολογιστικών μοντέλων των Harm & Seidenberg με προ-εκπαιδευμένο το φωνολογικό επίπεδο δεν διέφεραν σημαντικά από τα αποτελέσματα χωρίς το προ-εκπαιδευμένο επίπεδο.

Οι λέξεις παρουσιάζονταν στο υπολογιστικό μοντέλο κατά την εκπαίδευση, σε αναλογία ως προς τη συχνότητα εμφάνισής τους στο σώμα κειμένου από το οποίο προήλθαν. Η συχνότητα αυτή έχει μετατραπεί σε σχετική πιθανότητα εμφάνισης μέσα από μια συνάρτηση λογαριθμικής συμπίεσης συχνότητας, όπως χρησιμοποιήθηκε και από τους Harm & Seidenberg (1999):



$$p_i = \frac{\log[(f_i / 100) + 1]}{\log(m / 100)}$$

Στη συνάρτηση αυτή το  $f$  είναι η συχνότητα της λέξης  $i$  και το  $m$  είναι η συχνότητα της λέξης που εμφανίζεται συχνότερα στο σώμα κειμένου. Στη συγκεκριμένη περίπτωση η λέξη «είναι» εμφανίζεται συχνότερα με συχνότητα 300.307 εμφανίσεις στο σώμα κειμένου.

Η επιλογή μιας λέξης για εμφάνιση στο υπολογιστικό μοντέλο κατά την εκπαίδευση γίνονταν κατά τον εξής τρόπο:

1. Μια λέξη επιλέγοντας τυχαία από το σώμα εκπαίδευσης των 120.745 λέξεων.
2. Δημιουργούνται ένα τυχαίος αριθμός από το 0 έως το 1.
3. Αν ο αριθμός του βήματος 2 ήταν μεγαλύτερος ή ίσος από τη σχετική πιθανότητα εμφάνισης της λέξης που επιλέχθηκε τυχαία, τότε η λέξη εμφανιζόταν στο υπολογιστικό μοντέλο. Αλλιώς η διαδικασία ξεκινούσε από την αρχή (βήμα 1).

Το μεγάλο μέγεθος του σώματος εκπαίδευσης δημιούργησε προβλήματα στην εμφάνιση των λέξεων στο υπολογιστικό μοντέλο κατά την εκπαίδευση. Πολλές λέξεις χαμηλής συχνότητας δεν είχαν την ευκαιρία να εμφανιστούν στην είσοδο και το υπολογιστικό μοντέλο δεν τις έβλεπε ποτέ. Για να περιοριστεί αυτό το φαινόμενο ορίστηκε ένα κατώτατο όριο στη σχετική πιθανότητα εμφάνισης. Για τις λέξεις με συχνότητα εμφάνισης μικρότερη ή ίση από 0,3 ανά εκατομμύριο λέξεων του σώματος κειμένου, η σχετική πιθανότητα ορίστηκε σε 0,2. Αν και η παρέμβαση αυτή ενισχύει τις λέξεις χαμηλής συχνότητας, το αποτέλεσμα δεν διαφοροποιείται καθώς η πιθανότητα εμφάνισης παραμένει αρκετά χαμηλή και σε συνδυασμό με τον πολύ μεγάλο αριθμό λέξεων δεν επηρεάζει τον τρόπο εκπαίδευσης του υπολογιστικού μοντέλου. Οι λέξεις υψηλής και μεσαίας συχνότητας που μας

ενδιαφέρουν και περισσότερο θα έχουν τον σωστό αριθμό εμφανίσεων και έτσι θα μπορέσουν να δημιουργηθούν οι δομές στο συνδεδιστικό δίκτυο από την εκπαίδευση, ενώ οι περισσότερες από τις λέξεις χαμηλής συχνότητας θα παρουσιαστούν τουλάχιστον μία φορά στην είσοδο.

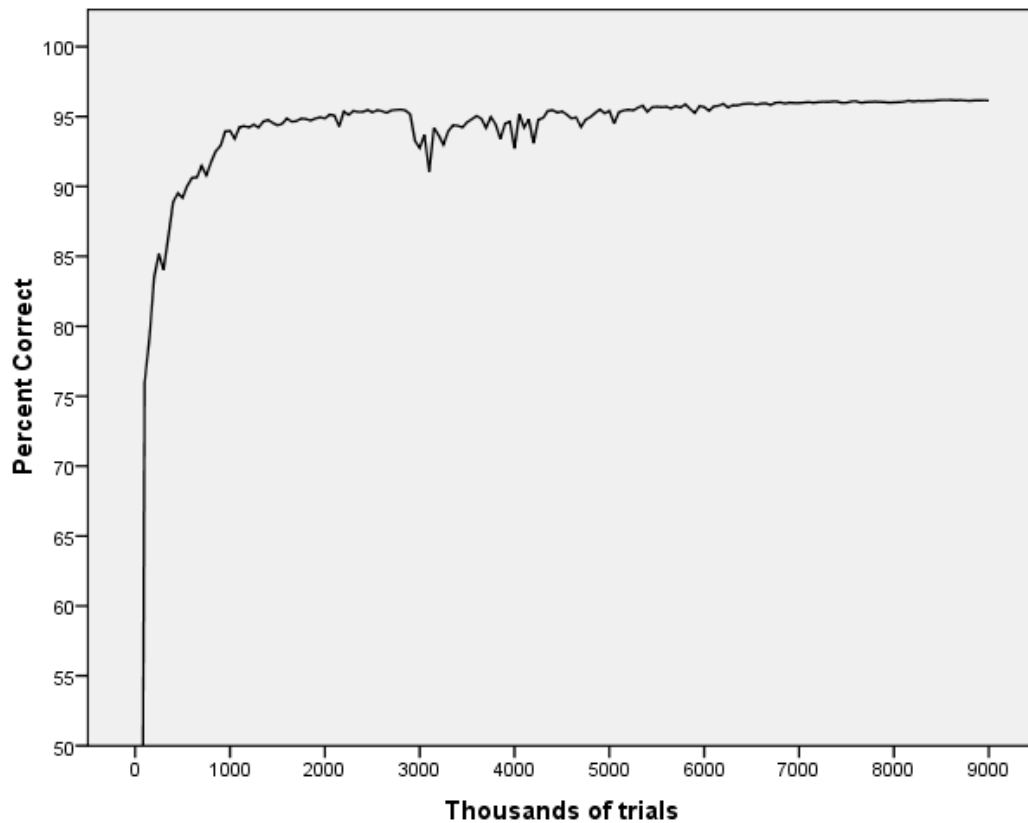
Η εκπαίδευση ξεκίνησε με τυχαία βάρη στις συνδέσεις των κόμβων του συνδεδιστικού δικτύου. Στην είσοδο παρουσιάστηκαν 9.000.000 λέξεις επιλεγμένες με τη διαδικασία που περιγράφηκε παραπάνω. Κάθε λέξη ενεργοποιούσε του κόμβους εισόδου του δικτύου για 10 συνεχόμενους κτύπους (ticks). Κάθε κτύπος είναι ένα κύκλος μετάδοσης της ενεργοποίησης στο συνδεδιστικό δίκτυο, από την είσοδο στην έξοδο. Στη συνέχεια, για 2 ακόμα κτύπους δεν υπήρχε είσοδος αλλά λειτουργούσε το δυναμικό σύστημα του φωνολογικού επιπέδου και τους επιπέδου καθαρισμού δίνοντας ενεργοποίηση το ένα στο άλλο. Εν τέλει, μετά από 12 κτύπους από τη στιγμή που παρουσιάστηκε η λέξη στην είσοδο, η έξοδος που υπήρχε στο φωνολογικό επίπεδο συγκρίνονταν με την επιθυμητή. Τα βάρη των συνδέσεων τροποποιούνταν χρησιμοποιώντας τη μέθοδο continuous recurrent backpropagation με ρυθμό εκπαίδευσης 0,1.

Για να αξιολογηθεί η μάθηση του μοντέλου, από το σώμα λέξεων εκπαίδευσης επιλέχθηκε ένα υποσύνολο 12.017 λέξεων (επιλέχθηκε μία λέξη ανά δέκα, περίπου το 10%) και κάθε 50.000 εκπαιδευτικές εμφανίσεις λέξεων δίδονταν στο δίκτυο για ανάγνωση (βλέπε Εικόνα 6). Το υποσύνολο των λέξεων είναι αντιπροσωπευτικό δείγμα του σώματος εκπαίδευσης και η μάθηση που εμφανίζει το δίκτυο σε αυτό είναι παρόμοια με την μάθηση που εμφανίζει στο σώμα εκπαίδευσης.

Η διαδικασία της ανάγνωσης αρχίζει με την εμφάνιση στην είσοδο του συνδεδιστικού δικτύου της αναπαράστασης της κάθε λέξης και τη διατήρησή της για 16 κτύπους. Μετά το πέρας του χρονικού αυτού διαστήματος η είσοδος χάνεται και δίνεται χρονικό διάστημα άλλων 2 κτύπων κατά το οποίο η ενεργοποίηση αφορά μόνο το δυναμικό σύστημα του φωνολογικού και του επιπέδου καθαρισμού. Η έξοδος στο φωνολογικό επίπεδο μετά από 18 κτύπους από την αρχική εμφάνιση της λέξης στην είσοδο, θεωρείται η απάντηση του υπολογιστικού μοντέλου. Το χρονικό διάστημα των 18 κτύπων είναι μεγαλύτερο από αυτό των 12 που χρησιμοποιήθηκε κατά την

εκπαίδευση. Αυτό επιλέχθηκε με βάση παρατηρήσεις στις εξόδους του υπολογιστικού μοντέλου, όταν αυτό διάβαζε λέξεις μεγάλου μήκους ή / και χαμηλής συχνότητας και οι οποίες όταν τους δίνονταν μεγαλύτερο χρονικό διάστημα το υπολογιστικό μοντέλο μπορούσε και τις διάβαζε σωστά. Αυτό συμβαίνει γιατί το δυναμικό σύστημα έχει περισσότερο χρόνο στη διάθεσή του για να καταλήξει στη σωστή προφορά της λέξης.

Από τους 12 κτύπους έγιναν δοκιμές με μεγαλύτερα χρονικά διαστήματα και η βέλτιστη αύξηση της απόδοσης προέκυψε από τους 18 κτύπους. Η απόδοση του υπολογιστικού μοντέλου αυξάνεται ακόμα περισσότερο αν αυξηθεί και άλλο το χρονικό διάστημα (έγιναν δοκιμές για έως και 45 κτύπους), αλλά η αύξηση αυτή δεν είναι σημαντική σε σχέση με το χρονικό διάστημα και την υπολογιστική ισχύ που χρειάζεται για να δοθεί η απάντηση. Κατά την εκπαίδευση χρησιμοποιήθηκαν μόνο 12 κτύποι γιατί αυτό οδήγησε γρηγορότερη μάθηση (τόσο από άποψη βελτίωσης της επίδοσης, όσο και από της άποψη του υπολογιστικού χρόνου), πιθανότατα λόγω της αυξημένης πίεσης, που ασκεί το μειωμένο χρονικό διάστημα, ώστε να μάθει το συνδεδετιστικό δίκτυο.



**Εικόνα 6.** Διάγραμμα μάθησης του υπολογιστικού μοντέλου (ποσοστό σωστών αναγνώσεων λέξεων) κατά τη διάρκεια εκπαίδευσής του.

## **Έλεγχος Επίδοσης**

Για τον έλεγχο της επίδοσης του υπολογιστικού μοντέλου έγινε έλεγχος των επιδράσεων των ποιο σημαντικών μεταβλητών που είναι γνωστό ότι επηρεάζουν το χρόνο ανάγνωσης (Balota, Yap & Cortese, 2006). Ο έλεγχος βασίστηκε στην ανάγνωση μιας ομάδας λέξεων και μιας ομάδας ψευδολέξεων. Η ομάδα των λέξεων περιλαμβάνει 150 λέξεις (Βλέπε Παράρτημα Β) ειδικά επιλεγμένες ώστε να καλύπτουν ένα εύρος μήκους, συχνότητας, αριθμού ορθογραφικών γειτόνων και διγραμματικής συχνότητας και οι συσχετίσεις μεταξύ των μεταβλητών αυτών να είναι όσο το δυνατόν ελάχιστες (Spearman's  $\rho < 0.2$ ) και να μην είναι στατιστικά σημαντικές ώστε να

περιοριστούν οι επιδράσεις μόνο σε αυτές τις βασικές μεταβλητές μεμονωμένα.

Η ομάδα των ψευδολέξεων αποτελείται από 150 ψευδολέξεις (βλέπε Παράρτημα Β) οι οποίες δημιουργήθηκαν σε αντιστοιχία προς τις λέξεις, ώστε να καλύπτεται παρόμοιο εύρος των μεταβλητών και να έχουν παρόμοια χαρακτηριστικά. Επίσης αποφεύχθηκε η οποιαδήποτε ομοιότητα με λέξεις της ομάδας των λέξεων.

Οι λέξεις και ψευδολέξεις αυτές προέρχονται από τη διπλωματική εργασία την μεταπτυχιακής φοιτήτριας Ευθυμίας Καπνουλά και τα δεδομένα που θα συλλεχθούν από την ανάγνωση αυτών των ερεθισμάτων από ανθρώπους, θα χρησιμοποιηθούν μελλοντικά για τη σύγκρισή τους με τα δεδομένα που δίνει το υπολογιστικό μοντέλο.

Οι λέξεις των παραπάνω ομάδων δόθηκαν για ανάγνωση από το υπολογιστικό μοντέλο τόσο στη τονισμένη μορφή τους όσο και στην άτονη. Αυτό έγινε για να δούμε την επίδοση του υπολογιστικού μοντέλου τόσο ως προς την άρθρωση των λέξεων όσο και ως προς τον τονισμό τους. Οι λέξεις και οι ψευδολέξεις παρουσιάστηκαν στην είσοδο του συνδεδετιστικού δικτύου με τον ίδιο τρόπο που παρουσιάζονταν οι λέξεις για την παρακολούθηση της μάθησης. Μετά την απάντηση του υπολογιστικού μοντέλου γίνονταν έλεγχος της ορθότητάς της. Κάθε απάντηση της φωνολογικής εξόδου σε κάθε έναν από του κτύπους από τον 3 έως τον 18, συγκρίνονταν με την επιθυμητή απάντηση. Αν η απάντηση ήταν ίδια με την επιθυμητή τότε καταγράφονταν ο κτύπος στον οποίο βρέθηκε η σωστή απάντηση, ως ο χρόνος απόκρισης του μοντέλου για την ανάγνωση της συγκεκριμένης λέξης. Αν μέχρι τον 18<sup>ο</sup> κτύπο δεν υπήρχε σωστή απάντηση τότε ο χρόνος απόκρισης έπαιρνε την τιμή 20 που σήμαινε ότι η απάντηση ήταν λάθος.

Για την σύγκριση της απάντησης με την επιθυμητή θα έπρεπε να μετατραπούν τα διανύσματα των χαρακτηριστικών των φωνημάτων που προέρχονται από τους κόμβους εξόδου, σε φωνήματα ώστε να συγκριθούν με τα επιθυμητά φωνήματα που θα έπρεπε να υπάρχουν σε κάθε θέση. Για τη μετατροπή αυτή, για κάθε θέση στην φωνολογική αναπαράσταση λαμβάνοντας το διάνυσμα εξόδου των 18 χαρακτηριστικών του φωνήματος

που περιείχε και συγκρίνονταν με τα διανύσματα και των 32 φωνημάτων που υπάρχουν και με τον μηδενικό διάνυσμα για την περίπτωση που η θέση ήταν κενή. Με τη μέθοδο της Ευκλείδειας διαφοράς βρίσκονταν η Ευκλείδεια απόσταση κάθε φωνήματος από το φώνημα που υπήρχε σε κάθε θέση και το φώνημα με την μικρότερη απόσταση ορίζονταν ότι δόθηκε ως απάντηση από το υπολογιστικό μοντέλο για τη θέση αυτή.

Παρόμοια ήταν η διαδικασία και για το τονικό σημάδι, για τη θέση και το φώνημα το οποίο τονίζεται από το υπολογιστικό μοντέλο. Μόνη διαφορά ήταν ότι μόνο τα φωνήματα που αντιστοιχούν σε φωνήεντα συγκρίνονταν με την έξοδο.

Τα λάθη στα οποία το υπολογιστικό μοντέλο μπορεί να υποπέσει είναι τα εξής:

1. Λάθος άρθρωση με διαφορά ενός ή περισσότερων φωνημάτων και σωστό τονισμό
2. Λάθος άρθρωση με διαφορά ενός ή περισσότερων φωνημάτων και τονισμό σε διαφορετική θέση και στο σωστό φώνημα
3. Λάθος άρθρωση με διαφορά ενός ή περισσότερων φωνημάτων ένα εκ των οποίων και το τονούμενο φώνημα και τονισμό στη σωστή θέση αλλά στο λάθος φώνημα.
4. Τονισμός σε λάθος θέση με σωστή άρθρωση

Για τον έλεγχο των επιδράσεων των κύριων μεταβλητών εκτελέστηκε απλή ανάλυση παλινδρόμησης του χρόνου απόκρισης των σωστών απαντήσεων του υπολογιστικού μοντέλου. Οι μεταβλητές οι οποίες εξετάστηκαν είναι οι εξής:

- Η πιθανότητα εμφάνισης της λέξης κατά την εκπαίδευση (ως συχνότητα και μόνο για τις λέξεις)
- Το μήκος της λέξης ή της ψευδολέξης σε γράμματα

- Ο αριθμός των συλλαβών
- Ο αριθμός των ορθογραφικών γειτόνων
- Η σωρευτική διγραμματική πιθανότητα για τα γράμματα
- Η μέση ορθογραφική συλλαβική συχνότητα
- Η ορθογραφική διαφάνεια μετρούμενη από την ελάχιστη πιθανότητα εμφάνισης μονοσήμαντης αντιστοιχίας γραφήματος – φωνήματος.

Από τις παραπάνω μεταβλητές μόνο η μέση ορθογραφική συλλαβική συχνότητα διέφερε σημαντικά μεταξύ λέξεων και ψευδολέξεων ( $t(298) = 2.007, p = .01$ ).

### **Αποτελέσματα**

Μετά από 9.000.000 αναγνώσεις λέξεων κατά την εκπαίδευση, το υπολογιστικό μοντέλο μπορεί να αναγνώσει σωστά το 96,14% του σώματος εκπαίδευσης. Το ποσοστό αυτό δεν είναι το οριστικό καθώς η επίδοση στην ορθή ανάγνωση συνεχώς αυξάνεται με περαιτέρω εκπαίδευση. Από τα λάθη που έκανε το υπολογιστικό μοντέλο, το 2,37% αφορούσε λάθη τονισμού με σωστή άρθρωση της λέξης (Τύπου 1, 0,09% επί του συνόλου των λέξεων). 4,53% των λαθών αφορά τονισμό στη σωστή θέση αλλά σε λάθος φώνημα ίδιο όμως με αυτό της άρθρωσης (Τύπου 3, 0,17% επί του συνόλου των λέξεων). Το 18,97% των λαθών αφορά λάθη όπου το τονούμενο φώνημα βρίσκεται σε διαφορετική θέση κατά την άρθρωση την οποία ακολουθεί και το τονισμό (Τύπου 2, 0,73% επί του συνόλου των λέξεων). Όλα τα υπόλοιπα λάθη ήταν Τύπου 1 (2,86% επί του συνόλου των λέξεων) δηλαδή λάθη άρθρωσης των λέξεων αλλά με σωστό τονισμό.

Όπως προαναφέρθηκε, για να εξεταστούν οι επιδράσεις των κύριων μεταβλητών εκτελέστηκαν απλές αναλύσεις παλινδρόμησης επί του χρόνου απόκρισης κατά την ανάγνωση των λέξεων από το υπολογιστικό μοντέλο. Τα συνοπτικά αποτελέσματα των αναλύσεων φαίνονται στον Πίνακα 1.

Προβλεπτικοί Παράγοντες	Λέξεις	Ψευδολέξεις
Σημαντικοί	Διγραμματική Συχνότητα Αριθμός Γραμμάτων	Διγραμματική Συχνότητα Αριθμός Γραμμάτων Αριθμ. Ορθογρ. Γειτόνων Αριθμός συλλαβών
Μη Σημαντικοί	Συχνότητα Εμφάνισης Αριθμός συλλαβών Αριθμ. Ορθογρ. Γειτόνων Συλλαβική Συχνότητα Ορθογραφική Διαφάνεια	Συλλαβική Συχνότητα Ορθογραφική Διαφάνεια

**Πίνακας 1:** Συνοπτικός πίνακας αποτελεσμάτων της απλής ανάλυσης παλινδρόμησης στον χρόνο απόκρισης των τονούμενων λέξεων και ψευδολέξεων. Για τους σημαντικούς παράγοντες, η σειρά εμφάνισης είναι ανάλογη με το ποσοστό της διακύμανση που κάθε παράγοντας εξηγεί.

Το 98% των λέξεων της ομάδας λέξεων διαβάστηκε σωστά από το υπολογιστικό μοντέλο, όταν παρουσιάστηκαν τονούμενες. Ο μέσος χρόνος απόκρισης ήταν 4,44 κτύποι. Τα αποτελέσματα των αναλύσεων για τις τονούμενες λέξεις έδειξαν ότι ο αριθμός των γραμμάτων των λέξεων μπορεί να εξηγήσει το 4,4% της διακύμανσης του χρόνου απόκρισης ( $R^2 = .044$ ,  $\text{standardized } \beta = .084$ ,  $p = .011$ ) ενώ η διγραμματική συχνότητα μπορεί να εξηγήσει ένα άλλο 4,5% ( $R^2 = .045$ ,  $\text{standardized } \beta = -.028$ ,  $p = .010$ ). Καμία από τις υπόλοιπες μεταβλητές δεν έδωσαν στατιστικά σημαντικά αποτελέσματα επομένως δεν μπορούν να θεωρηθούν προβλεπτικοί παράγοντες της διακύμανσης του χρόνου απόκρισης.

Όταν παρουσιάστηκε η ομάδα των ψευδολέξεων τονούμενη, το υπολογιστικό μοντέλο διάβασε σωστά το 92% των ψευδολέξεων. Ο μέσος χρόνος απόκρισης ήταν 4,54 κτύποι. Η διαφορά ανάμεσα στο χρόνο απόκρισης των τονούμενων λέξεων και ψευδολέξεων δεν είναι στατιστικά



σημαντική ( $t(283) = -1.457, p = .148$ ). Στις ψευδολέξεις η ανάλυση παλινδρόμησης έδειξε ότι οι προβλεπτικοί παράγοντες είναι περισσότεροι από τους αντίστοιχους των λέξεων. Ο αριθμός των γραμμάτων εξηγεί το 5,6% της διακύμανσης ( $R^2 = .056, \text{standardized } \beta = .116, p = .005$ ). Ο αριθμός των συλλαβών εξηγεί το 3,2% ( $R^2 = .032, \text{standardized } \beta = .179, p = .035$ ). Η διγραμματική συχνότητα το 7,5% ( $R^2 = .075, \text{standardized } \beta = -.046, p = .001$ ). Τέλος ο αριθμός των ορθογραφικών γειτόνων εξηγεί το 3,8% της διακύμανσης ( $R^2 = .038, \text{standardized } \beta = -.195, p = .022$ ). Οι άλλες δύο μεταβλητές (συλλαβική συχνότητα και ορθογραφική διαφάνεια) δεν έδωσαν στατιστικά σημαντικά αποτελέσματα στις αναλύσεις και έτσι θεωρείται ότι δεν είναι προβλεπτικοί παράγοντες.

Ακολούθησε η παρουσίαση στο υπολογιστικό μοντέλο και των δύο ομάδων σε μη τονούμενη μορφή. Αυτό που μας ενδιαφέρει σε αυτή την περίπτωση είναι η μελέτη του τονισμού που εκτελεί το υπολογιστικό μοντέλο. Στην περίπτωση των μη τονούμενων λέξεων το 53,33% του συνόλου διαβάστηκε σωστά. Ο μέσος χρόνος απόκρισης ήταν 5,11 κτύποι και η διαφορά του σε σχέση με τον χρόνο απόκρισης στις τονούμενες λέξεις είναι σημαντική ( $t(80) = -3.267, p = .002$ ). Από το σύνολο των λαθών το 85,71% αφορά λάθη τονισμού σε σωστά αρθρωμένες λέξεις. (54 λέξεις τονίστηκαν σε διαφορετική θέση και 6 λέξεις τονίστηκαν σε λάθος θέση και το φώνημα που τονίστηκε είναι διαφορετικό από αυτό που βρίσκεται στην θέση αυτή).

Αντίστοιχα για τις ψευδολέξεις, όταν παρουσιάστηκαν χωρίς τονικό σημάδι, το 42,67% διαβάστηκε σωστά, με τονισμό δηλαδή σύμφωνο με τον προεπιλεγμένο για κάθε ψευδολέξη. Ο χρόνος απόκρισης ήταν κατά μέσο όρο 5,3 κτύποι και η διαφορά του με τον αντίστοιχο μέσο όρο των τονούμενων ψευδολέξεων είναι στατιστικά σημαντική ( $t(63) = -2.019, p = .048$ ). Στον Πίνακα 2 φαίνονται οι κατανομές τονισμού που έκανε το υπολογιστικό μοντέλο στις λέξεις και στις ψευδολέξεις στην λήγουσα, την παραλήγουσα και την προπαραλήγουσα. Επίσης εμφανίζεται και η κατανομή τονισμού σε ολόκληρο το σώμα κειμένων όπως προέκυψε από την ανάλυση του Πρωτόπαπα (2006).

	Λήγουσα	Π/λήγουσα	ΠΠ/λήγουσα	
Μη Τονούμενες Λέξεις Σωστά Αρθρωμένες	30,7	44,5	24,8	Ανομοιόμορφη Κατανομή ( $\chi^2(2) = 8.423, p = .015$ )
Μη Τονούμενες Λέξεις Σωστά Αρθρωμένες και Τονισμένες	27,5	42,5	30	Ομοιόμορφη Κατανομή ( $\chi^2(2) = 3.100, p = .212$ )
Μη Τονούμενες Λέξεις Λάθος Τονισμένες	35,1	47,4	17,5	Ανομοιόμορφη Κατανομή ( $\chi^2(2) = 7.684, p = .021$ )
Τονισμένες Λέξεις	30,7	40	29,3	
Μη Τονούμενες Ψευδολέξεις Συνολικά	28,7	44,1	27,2	Ανομοιόμορφη Κατανομή ( $\chi^2(2) = 7.162, p = .028$ )
Σώμα Κειμένων	30	44,9	25	

**Πίνακας 2.** Ποσοστά τονισμού λέξεων και ψευδολέξεων.

## Συμπεράσματα

Έγινε μια προσπάθεια να δημιουργηθεί ένα υπολογιστικό μοντέλο ανάγνωσης για την Ελληνική γλώσσα, βασισμένο στην θεωρία του τριγωνικού μοντέλου. Είναι το πρώτο για τα Ελληνικά και ένα από τα ελάχιστα που μπορούν να δεχθούν ως είσοδο πολυσύλλαβες λέξεις. Η ύπαρξη του φαινομένου ΣίΦ στην Ελληνική γλώσσα επέβαλε την μη συλλαβοποιημένη είσοδο στο μοντέλο. Σαν αποτέλεσμα, το συνδεδειστικό δίκτυο θα έπρεπε να «μάθει» να αντιστοιχεί τις σειρές των γραμμάτων στην είσοδο με συλλαβές στην έξοδο. Αυτό είναι αρκετά δύσκολο να γίνει μιας και τα γράμματα στην είσοδο δεν είναι σταθερά αλλά αλλάζουν σε πλήθος και θέση ανάλογα με το μήκος και τη μορφολογία της λέξης. Ιδιαίτερα δύσκολο είναι και η αντιστοίχιση λέξεων με σειρές διαδοχικών φωνηέντων (μπορεί να φτάσουν τα 6). Παράδειγμα είναι η τετρασύλλαβη λέξη «αβαείου» η οποία έχει μόνο ένα σύμφωνο και τη σειρά φωνηέντων «αείου» που δίνεται ως έχει στην είσοδο του δικτύου. Η σειρά αυτή θα πρέπει να αντιστοιχηθεί στις τρεις τελευταίες συλλαβές της λέξης.

Επιπρόσθετο πρόβλημα προκύπτει και από το πρόβλημα της διασποράς που θέτει η συγκεκριμένη ορθογραφική αναπαράσταση. Η δυσκολία αυξάνει από τη στιγμή που θα πρέπει να κάνει αντιστοιχίσεις μεταξύ γραμμάτων που δεν βρίσκονται στην ίδια θέση κάθε φορά με συλλαβές που βρίσκονται. Το συνδετιστικό δίκτυο δέχεται βοήθεια σε αυτόν τον τομέα από την υψηλή συνέπεια που χαρακτηρίζει την Ελληνική γλώσσα στον τομέα της μετατροπής των γραφημάτων σε φωνήματα. Έτσι δεν φαίνεται να δημιουργείται κάποιο πρόβλημα στο να «μάθει» την αντιστοίχιση αυτή ενώ επιπλέον φαίνεται ότι μπορεί και με επιτυχία να εφαρμόσει τη «γνώση» που απέκτησε και στην ανάγνωση των ψευδολέξεων. Η διευκόλυνση αυτή και σωστή λειτουργία κατά την αντιστοίχιση φαίνεται και από τα αρχικά στάδια της εκπαίδευσης (100.000 εμφανίσεις λέξεων) όπου το υπολογιστικό μοντέλο είναι σε θέση να αναγνώσει με επιτυχία έναν σχετικά μεγάλο αριθμό λέξεων (76%, βλέπε Εικόνα 6).

Στο υπολογιστικό μοντέλο χορηγήθηκε ένα τεράστιος αριθμός λέξεων κατά την διάρκεια της εκπαίδευσης. Οι 120.745 λέξεις χορηγήθηκαν σε ένα πλήθος από 9.000.000 εμφανίσεις στην είσοδο, που είναι σχετικά μικρό για τον αριθμό των λέξεων. Συγκριτικά σε αντίστοιχα υπολογιστικά μοντέλα (αλλά για μονοσύλλαβες λέξεις, άρα μικρότερης πολυπλοκότητας) το πλήθος των λέξεων και ο αριθμός των εμφανίσεων φαίνονται στον Πίνακα 3. Το υπολογιστικό μοντέλο μετά τις 9.000.000 εμφανίσεις μπορεί να αναγνώσει σωστά περισσότερο από το 96% του συνόλου των λέξεων καθιστώντας το ιδιαίτερα ακριβές, αν ληφθεί υπόψη:

- το μεγάλο σύνολο των λέξεων
- το ότι κάποιες λέξεις από το σύνολο αυτό δεν εμφανίστηκαν ποτέ στην είσοδο εξαιτίας της χαμηλής τους συχνότητας εμφάνισης και κατά την ανάγνωση αντιμετωπίζονται ως ψευδολέξεις.
- το ότι οι λέξεις με 5 συλλαβές δεν ήταν αρκετές (περίπου 12,5% του συνόλου) ώστε να δημιουργηθούν μέσα στα πλαίσια των 9.000.000 εμφανίσεων αρκετά δυνατές συνδέσεις μεταξύ των κόμβων και να γίνονται πάντοτε οι σωστές αντιστοιχίσεις για την τελευταία συλλαβή

σε σχέση με τις συλλαβές από 2 έως 4, μιας και η στοίχιση των συλλαβών ήταν από αριστερά προς τα δεξιά.

Ρόλο στα καλά αποτελέσματα που φαίνονται παραπάνω πιστεύεται ότι παίζει η απλή γραφοφωνημική αντιστοίχιση που ακολουθείται στην Ελληνική γλώσσα.

<b>Υπολογιστικό Μοντέλο</b>	<b>Αριθμός λέξεων κατά την εκπαίδευση</b>	<b>Αριθμός εμφανίσεων</b>	<b>Ποσοστό σωστής ανάγνωσης</b>
Harm & Seidenberg 1999	3.123	10.000.000	99%
Harm & Seidenberg 2004	6.103	1.500.000	99,2%
Zevin & Seidenberg 2006	5.870	1.000.000	95%
Pagliuca & Monaghan (υπό έκδοση) με συλλαβισμό της εισόδου	9,911	1.200.000	93,7%
Pagliuca & Monaghan (υπό έκδοση) χωρίς συλλαβισμό της εισόδου	2,424	10,000,000	96%

**Πίνακας 3.** Επιδόσεις παρόμοιων μοντέλων σε σχέση με τον αριθμό των λέξεων του σώματος εκπαίδευσης και των αριθμό των εμφανίσεων.

Η μόνη περίπτωση κατά την οποία δεν ισχύει η υψηλή συνέπεια γραφοφωνημικής μετατροπής της Ελληνικής γλώσσας είναι στην περίπτωση του φαινομένου ΣίΦ. Παρατηρήθηκαν λάθη ανάγνωσης του υπολογιστικού μοντέλου σε λέξεις που περιείχαν το ΣίΦ και αφορούσαν την εναλλακτική φωνολογική αναπαράσταση της λέξης και μόνο, χωρίς δηλαδή λάθη φωνολογικού περιεχομένου. Τα λάθη αυτά αφορούσαν λέξεις με χαμηλή συχνότητα που το υπολογιστικό μοντέλο είτε δεν είδε καθόλου στην είσοδο είτε τις είδε ελάχιστες φορές ώστε να είναι σε θέση να τις αναγνώσει σωστά άρα έπαιξαν το ρόλο της ψευδολέξης. Παρόλα αυτά η ανάγνωση έγινε χωρίς φωνολογικό λάθος. Μένει να γίνει μια πιο εμπειριστατωμένη μελέτη της επίδραση του φαινομένου στο υπολογιστικό μοντέλο με ανάγνωση

ελεγχόμενων λέξεων και ψευδολέξεων και αντιπαραβολή των αποτελεσμάτων με τα αποτελέσματα του έδωσαν οι άνθρωποι (Νομικού, 2007).

Η Ελληνική γλώσσα επέβαλε και ένα δεύτερο στοιχείο κατά τον σχεδιασμό του υπολογιστικού μοντέλου. Οι Ελληνικές λέξεις είναι τονούμενες και το τονικό σημάδι θα έπρεπε με κάποιον τρόπο να δοθεί στο υπολογιστικό μοντέλο και να υπολογιστεί η επίδρασή του. Έχουν γίνει και άλλες προσπάθειες να γίνει μοντελοποίηση του τονισμού (Monaghan, Arciuli & Seva, 2008; Rastle & Coltheart, 2000) αλλά και να μπει το τονικό σημάδι στην ορθογραφική αναπαράσταση των λέξεων (Pagliuca & Monaghan, 2009, υπό έκδοση). Αυτή όμως είναι η πρώτη φορά που η αναπαράσταση του τονικού σημαδιού είναι ανεξάρτητη από την αναπαράσταση του τονούμενου γράμματος. Με αυτόν τον τρόπο το τονούμενο και μη τονούμενο γράμμα είναι το ίδιο και δεν χάνει τις ιδιότητες που μεταφέρει, ενώ το τονικό σημάδι λειτουργεί ανεξάρτητα. Το υπολογιστικό μοντέλο δείχνει να έχει μάθει τις αρχές τονισμού που ακολουθούνται στην Ελληνική γλώσσα. Κανένα λάθος δεν έγινε με τονισμό λέξης πριν από την προπαραλήγουσα αν και οι λέξεις δεν είχαν σταθερή θέση των τονούμενων συλλαβών (στοίχιση από αριστερά προς τα δεξιά). Τα λάθη τονισμού στην ανάγνωση των λέξεων και των ψευδολέξεων έδειξαν ότι το μοντέλο έχει μια προτίμηση στον τονισμό της παραλήγουσας όπως και οι άνθρωποι. Συνολικά στην ανάγνωση λέξεων και ψευδολέξεων το υπολογιστικό μοντέλο δεν εμφάνισε σημαντική απόκλιση από την κατανομή των τονούμενων συλλαβών κάτι που δείχνει ότι όχι μόνο δεν τονίζει τυχαία, αλλά και ότι έχει «μάθει» τον τρόπο με τον οποίο τονίζονται οι Ελληνικές λέξεις.

Η σχέση ανάμεσα στο τονούμενο γράμμα και στο τονικό σημάδι φαίνεται να είναι αρκετά ισχυρή, καθώς από την ανάλυση των λαθών προκύπτει ότι μόνο το 0,1% του συνόλου των λέξεων τονίστηκε σε λάθος συλλαβή. Ακόμα και όταν υπήρχαν λάθη αντιστοίχισης γραφήματος – φωνήματος ή όταν υπήρχαν λάθη στη θέση του τονούμενου φωνήματος, το τονικό σημάδι ακολούθησε το φώνημα είτε στη λανθασμένη του μορφή είτε στη λανθασμένη του θέση αντίστοιχα. Αυτό δείχνει ότι το υπολογιστικό μοντέλο έχει μάθει να χρησιμοποιεί το τονικό σημάδι για να τονίσει τις λέξεις.

Τα καλά αποτελέσματα όμως στις τονούμενες λέξεις δεν μπορούν να συγκριθούν με τα ιδιαίτερα χαμηλής ορθότητας αποτελέσματα των μη τονούμενων λέξεων (μόλις 53,3%). Τα αποτελέσματα αυτά έρχονται σε αντίθεση με αυτά των ανθρώπων (Protopapas, Gerakaki, & Alexandri, 2007; Protopapas & Gerakaki, υπό έκδοση) που δείχνουν ότι δεν επηρεάζονται από την έλλειψη του τονικού σημαδιού. Τα φτωχά αυτά αποτελέσματα πιθανών να βελτιωθούν αν στο μοντέλο πριν από την εκπαίδευση της γραφοφωνημικής αντιστοίχισης, γίνει μια εκπαίδευση του φωνολογικού επιπέδου μοντελοποιώντας έτσι την εκπαίδευση που δέχονται οι άνθρωποι ακούγοντας τις λέξεις κατά την ομιλία και ενισχύοντας τη σχέση μεταξύ των τονούμενων φωνημάτων και του τονικού σημείου. Επίσης βελτίωση μπορεί να προκύψει και από την απευθείας σύνδεση του ορθογραφικού επιπέδου με το φωνολογικό (Zorzi et. al., 1998) καθώς θα ενίσχυε την επίδραση του τονικού σημαδιού στην έξοδο (χωρίς την διαμεσολάβηση του κρυφού επιπέδου) και η έλλειψή του ενδεχομένως να έχει μικρότερη επίδραση.

Η ανάλυση των προβλεπτικών παραγόντων έδειξε ότι μόνο δύο από αυτούς επηρεάζουν την ανάγνωση των λέξεων και των ψευδολέξεων. Η πρώτη είναι το μήκος, που για τις λέξεις είναι σε γράμματα ενώ για τις ψευδολέξεις είναι σε γράμματα και σε συλλαβές. Για τον παράγοντα αυτό ο Balota et. al. (2006) έδειξε ότι υπάρχει επίδραση αλλά μόνο για τις λέξεις με μικρή συχνότητα και για τις ψευδολέξεις. Στην περίπτωση του υπολογιστικού μοντέλου μπορούμε να θεωρήσουμε ότι ο μικρός αριθμός εμφανίσεων των λέξεων κατά την εκπαίδευση σε σχέση με το σώμα λέξεων εκπαίδευσης, λειτουργεί ως παράγοντας μείωσης της συχνότητας των λέξεων καθώς δεν εμφανίστηκαν αρκετές φορές ώστε να ενισχυθούν οι συνδέσεις των κόμβων. Αυτό προκύπτει από το μέγεθος της επίδρασης του παράγοντα στις λέξεις σε σχέση με τον αριθμό των εμφανίσεων όπου ενώ αυτές αυξάνονται το μέγεθος μειώνεται. Ένας άλλος λόγος για τον οποίο αυτός ο παράγοντας επηρεάζει την ανάγνωση είναι και η χαμηλής πολυπλοκότητας γραφοφωνημική αντιστοίχιση της Ελληνική γλώσσας. Όσο περισσότερα μεμονωμένα γραφήματα πρέπει να μετατραπούν σε φωνήματα, τόσο μεγαλύτερη καθυστέρηση θα πρέπει να υπάρχει και να οδηγεί σε μεγαλύτερη επίδραση του μήκους. Ο συγκεκριμένος λόγος μπορεί να δώσει και μια εξήγηση και για

την απουσία επίδρασης της συχνότητας της λέξης αλλά και της απουσία επίδρασης από την λεξικότητα (διαφορά μεταξύ λέξεων και ψευδολέξεων).

Όπως και να έχει πάντως η εμφάνιση της επίδρασης της λέξης έρχεται σε αντίθεση με ότι πιστεύεται για τα υπολογιστικά μοντέλα που βασίζονται σε συνδεδετιστικά δίκτυα (Rastle & Coltheart, 2006) δηλαδή ότι δεν θα πρέπει να εμφανίζουν τέτοια επίδραση μιας και η επεξεργασία της λέξης γίνεται παράλληλα και όχι γράφημα – γράφημα όπως σε κάποια υπολογιστικά μοντέλα βασισμένα στην θεωρία της διπλής διαδρομής (Dual Route models). Ρόλο σε αυτήν την ανατροπή παίζει και το ότι το μοντέλο μπορεί να δεχθεί πολυσύλλαβες λέξεις οι οποίες δημιουργούν περισσότερους φωνολογικούς ελκυστές. Αυτό εν μέρει οφείλεται στο πρόβλημα της διασποράς καθώς τα ίδια γράμματα σε διαφορετική θέση δημιουργούν διαφορετικούς ελκυστές. Η ύπαρξη πολλών γραμμάτων αυξάνει τον αριθμό των ελκυστών αλλά και τον χρόνο που χρειάζονται να ισορροπήσουν σε μία σταθερή κατάσταση.

Ο δεύτερος παράγοντας που επηρεάζει την ανάγνωση από το υπολογιστικό μοντέλο είναι η διγραμματική συχνότητα. Αυτού του είδους επίδραση είναι αναμενόμενη (Balota et. al., 2006) για λέξεις χαμηλής συχνότητας σε σχέση με λέξεις υψηλής συχνότητας. Ο ίδιος λόγος που αναφέρθηκε παραπάνω, για τον οποίο οι λέξεις με τις οποίες εκπαιδεύτηκε το υπολογιστικό μοντέλο μπορούν να θεωρηθούν χαμηλής συχνότητας δείχνει να είναι αυτό που επηρεάζει το αποτέλεσμα και εδώ. Το υπολογιστικό μοντέλο με περισσότερες εμφανίσεις λέξεων κατά την εκπαίδευση ενδεχομένως να παρουσιάσει μικρότερη ή και καθόλου επίδραση της διγραμματικής συχνότητας ενώ ενδεχομένως να εμφανιστεί και επίδραση της συχνότητας της λέξης που τώρα δεν υφίσταται.

Δύο άλλοι παράγοντες δεν εμφάνισαν καμία επίδραση της ανάγνωση των λέξεων από το υπολογιστικό μοντέλο. Ο πρώτος είναι η διαφάνεια (η πολυπλοκότητα της γραφοφωνημικής αντιστοίχισης ) η οποία αναμενόταν να έχει επίδραση. Θα πρέπει να διερευνηθεί η έλλειψη επίδρασης μετά την ολοκλήρωση της συλλογής δεδομένων από Έλληνες αναγνώστες. Ο δεύτερος παράγοντας είναι οι ορθογραφικοί γείτονες που για τις Αγγλικές λέξεις έχει επίδραση αλλά το υπολογιστικό μοντέλο για τις Ελληνικές λέξεις δεν φαίνεται να επηρεάζεται. Ένας λόγος μπορεί να είναι ο χαμηλός αριθμός των

ορθογραφικών γειτόνων για τις Ελληνικές λέξεις. Με μέσω όρο 1,69 γείτονες ανά λέξεις δεν μπορούμε να θεωρήσουμε ότι είναι αρκετοί ώστε να επηρεάσουν το αποτέλεσμα.

Διαφορετική επίδραση όμως φαίνεται να έχουν οι ορθογραφικοί γείτονες στην ανάγνωση των ψευδολέξεων. Στην περίπτωση των ψευδολέξεων, όπου οι λέξεις δεν έχουν εμφανιστεί ξανά στην είσοδο του υπολογιστικού μοντέλου η γραφοφωνημική αντιστοίχιση υπόκειται σε μεγαλύτερη επίδραση των φωνολογικών ελκυστών καθώς ενεργοποιούνται περισσότεροι. Αυτή η αύξηση των ελκυστών ενισχύεται και από την ύπαρξη των ορθογραφικών γειτόνων καθώς και αυτοί αυξάνουν τον αριθμό των ελκυστών όχι σε μεγάλο βαθμό, λόγο του μικρού αριθμού τους, αλλά αθροιστικά με τους υπόλοιπους φαίνεται να είναι σε θέση να παρουσιάσουν επίδραση κατά την ανάγνωση των ψευδολέξεων.

## **Επίλογος**

Παρουσιάστηκε ένα υπολογιστικό μοντέλο ανάγνωσης προβλέψιμων και μη προβλέψιμων λέξεων και ψευδολέξεων. Τα αποτελέσματα που δίνει το υπολογιστικό μοντέλο είναι ενθαρρυντικά ώστε να συνεχίσει η προσπάθεια εξέλιξης βελτιώνοντας την απόδοση τους στις μη τονούμενες λέξεις αλλά και μελετώντας την επίδραση ή όχι κάποιων από τους προβλεπτικούς παράγοντες που επηρεάζουν την ανάγνωση. Η συλλογή δεδομένων από Έλληνες αναγνώστες βρίσκεται σε εξέλιξη και τα αποτελέσματά τους θα πρέπει να συγκριθούν με τα αποτελέσματα του μοντέλου.



## Βιβλιογραφία

- Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, 105(4), 678-723.
- Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (A travel update). In M. Traxler & M. A. Gernsbacher (Eds.) *Handbook of psycholinguistics (2nd ed.)*.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256.
- Chalamandaris, A. Raptis, S., & Tsiakoulis, P. (2005). Rule-based grapheme-to-phoneme method for the Greek. In *INTERSPEECH-2005*
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition and dyslexia: Insights from connectionist models. *Psychological Review*, 106, 491–528.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, 111, 662–720.
- Jared, D. (2002). Spelling–sound consistency and regularity effects in word naming. *Journal of Memory & Language*, 46, 723-750
- Monaghan, P., Arciuli, J., & Seva, N. (2008). Constraints for computational models of reading: Evidence from learning lexical stress. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.
- Pagliuca, G., & Monaghan, P. (υπό έκδοση). Discovering large grain-sizes in a transparent orthography: insights from a connectionist model of Italian word naming. *European Journal of Cognitive Psychology*.
- Pagliuca, G. & Monaghan, P. (2009). A connectionist model of reading for Italian. In J. Mayor, N. Ruh, & K. Plunkett (Eds.), *Connectionist models of behaviour and cognition II* (pp.301–312). Singapore: World Scientific.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, 114, 273–315.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Protopapas, A. (2006). On the use and usefulness of stress diacritics in reading Greek. *Reading & Writing: An Interdisciplinary Journal*, 19, 171–198.
- Protopapas, A., Gerakaki, A., & Alexandri, S. (2007). Sources of information for stress assignment in reading Greek. *Applied Psycholinguistics*, 28, 695–720.
- Protopapas, A., & Gerakaki, S. (υπό έκδοση). Development of processing stress diacritics in reading Greek. *Scientific Studies of Reading*.
- Protopapas, A., & Nomikou E. (2009). Rules vs. lexical statistics in Greek nonword reading. To appear in the *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

- Protopapas, A., & Vlahou, E. L. (υπό έκδοση). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods*.
- Rastle K, Coltheart M (2006), "Is there serial processing in the reading system; and are there local representations?", In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*. Hove: Psychology Press.
- Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory & Language*, 42, 342–364.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Seidenberg, M., & Plaut, D. (2006). Progress in understanding word reading: Data fitting versus theory building. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*. Hove: Psychology Press.
- Spencer, K. A. (2007) Predicting children's word-spelling difficulty for common English words from measures of orthographic transparency, phonemic and graphemic length and word frequency. *British Journal of Psychology*, 98, 305–338.
- Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, 54, 145–160.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1131–1161.

# Παραρτήματα

## Παράρτημα Α

Τα γράμματα που μπορεί να δεχθεί κάθε θέση της ορθογραφικής αναπαράστασης.

Θέση Γραμμάτων	Γράμματα																	
1	β	γ	δ	ζ	θ	κ	λ	μ	ν	ξ	π	ρ	σ	τ	φ	χ	ψ	
2	β	γ	δ	ζ	θ	κ	λ	μ	ν	π	ρ	σ	τ	φ	χ			
3	λ	ν	π	ρ														
4	α	ε	η	ι	ο	υ	ω											
5	α	ε	η	ι	ο	υ	ω	ϊ	ϋ									
6	α	ε	η	ι	ο	υ	ω	ϊ										
7	α	ε	η	ι	ο	υ	ω											
8	ι	υ	ω															
9	β	γ	δ	ζ	θ	κ	λ	μ	ν	ξ	π	ρ	ς	σ	τ	φ	χ	ψ
10	β	γ	δ	ζ	θ	κ	λ	μ	ν	ξ	π	ρ	ς	σ	τ	φ	χ	ψ
11	ζ	θ	κ	λ	μ	ν	π	ρ	σ	τ	φ							
12	π	ρ	σ	τ														
13	τ																	
14	α	ε	η	ι	ο	υ	ω											
15	α	ε	η	ι	ο	υ	ω	ϊ	ϋ									
16	α	ε	η	ι	ο	υ	ω											
17	α	ε	ι	ο	υ	ω												
18	ε	ι	υ	ω														
19	ι																	
20	β	γ	δ	ζ	θ	κ	λ	μ	ν	ξ	π	ρ	ς	σ	τ	φ	χ	ψ
21	β	γ	δ	ζ	θ	κ	λ	μ	ν	ξ	π	ρ	ς	σ	τ	φ	χ	ψ
22	ζ	θ	κ	λ	μ	ν	π	ρ	σ	τ								
23	ρ																	
24	α	ε	η	ι	ο	υ	ω											
25	α	ε	η	ι	ο	υ	ω	ϊ	ϋ									
26	α	ε	η	ι	ο	υ	ω											
27	α	ε	ι	ο	υ	ω												
28	α	ι	ο	υ														
29	β	γ	δ	ζ	θ	κ	λ	μ	ν	ξ	π	ρ	ς	σ	τ	φ	χ	ψ
30	β	γ	δ	ζ	θ	κ	λ	μ	ν	ξ	π	ρ	ς	σ	τ	φ	χ	ψ
31	ζ	θ	κ	ρ	τ													
32	α	ε	η	ι	ο	υ	ω											
33	α	ε	η	ι	ο	υ	ω											
34	α	ε	η	ι	ο	υ	ω											
35	ι	υ																
36	β	γ	δ	ζ	θ	κ	λ	μ	ν	ξ	π	ρ	ς	σ	τ	φ	χ	ψ
37	δ	θ	κ	λ	μ	ν	ξ	ρ	σ	τ	χ	ψ						
38	α	ε	η	ι	ο	υ	ω											
39	α	ι	ο	υ														
40	ν	ς																

Θέση Τονικού Σημαδιού	Τονούμενο Γράμμα															
	1	α	ε	η	ι	ο	υ	ω								
2	α	ε	η	ι	ο	υ	ω	ϊ	ϋ							
3	α	ε	η	ι	ο	υ	ω	ϊ								
4	α	ι	ο	υ	ω											
5																
1	α	ε	η	ι	ο	υ	ω									
2	α	ε	η	ι	ο	υ	ω	ϊ	ϋ							
3	α	ε	η	ι	ο	υ	ω									
4	ι	υ														
5																
6																
1	α	ε	η	ι	ο	υ	ω									
2	α	ε	η	ι	ο	υ	ω	ϊ								
3	α	ε	η	ι	ο	υ	ω									
4	ι	υ														
5																
1	α	ε	η	ι	ο	υ	ω									
2	α	ε	η	ι	ο	υ	ω									
3	α	ι	ο	υ	ω											
4	ι															
1	α	ε	η	ι	ο	υ	ω									
2	ι	υ														

Ο πίνακας φωνολογικών χαρακτηριστικών των φωνημάτων.

	Χειλικά	Οδοντικά	Φατνιακά	Ουρανικά	Υπερωικά	Κλειστά	Τριβόμενα	Προστριβόμενα	Ένρινα	Πλάγια	Παλλόμενα	Φωνούμενα	Άφωνα	Πρόσθια	Οπίσθια	Ψηλά	Μεσαία	Χαμηλά
D	..	1	..	..	..	..	1	..	..	..	..	1	..	..	..	..	..	..
G	..	..	..	1	..	1	..	..	..	..	..	1	..	..	..	..	..	..
J	..	..	..	..	1	..	1	..	..	..	..	1	..	..	..	..	..	..
L	..	..	..	1	..	..	..	..	..	1	..	1	..	..	..	..	..	..
N	..	..	..	1	..	..	..	..	1	..	..	1	..	..	..	..	..	..
T	..	1	..	..	..	..	1	..	..	..	..	..	1	..	..	..	..	..
X	..	..	..	1	..	..	1	..	..	..	..	..	1	..	..	..	..	..
S	..	..	1	..	..	..	..	1	..	..	..	..	1	..	..	..	..	..
Z	..	..	1	..	..	..	..	1	..	..	..	1	..	..	..	..	..	..
c	..	..	..	1	..	1	..	..	..	..	..	..	1	..	..	..	..	..
b	1	..	..	..	..	1	..	..	..	..	..	1	..	..	..	..	..	..
d	..	1	1	..	..	1	..	..	..	..	..	1	..	..	..	..	..	..
g	..	..	..	..	1	1	..	..	..	..	..	1	..	..	..	..	..	..
f	1	1	..	..	..	..	1	..	..	..	..	..	1	..	..	..	..	..
h	..	..	..	..	1	..	..	..	1	..	..	1	..	..	..	..	..	..
k	..	..	..	..	1	1	..	..	..	..	..	..	1	..	..	..	..	..
j	..	..	..	1	..	..	1	..	..	..	..	1	..	..	..	..	..	..
m	1	..	..	..	..	..	..	..	1	..	..	1	..	..	..	..	..	..
l	..	1	1	..	..	..	..	..	..	1	..	1	..	..	..	..	..	..
n	..	1	1	..	..	..	..	..	1	..	..	1	..	..	..	..	..	..
p	1	..	..	..	..	1	..	..	..	..	..	..	1	..	..	..	..	..
s	..	..	1	..	..	..	1	..	..	..	..	..	1	..	..	..	..	..
r	..	..	1	..	..	..	..	..	..	..	1	1	..	..	..	..	..	..
t	..	1	1	..	..	1	..	..	..	..	..	..	1	..	..	..	..	..
v	1	1	..	..	..	..	1	..	..	..	..	1	..	..	..	..	..	..
x	..	..	..	..	1	..	1	..	..	..	..	..	1	..	..	..	..	..
z	..	..	1	..	..	..	1	..	..	..	..	1	..	..	..	..	..	..
a	..	..	..	..	..	..	..	..	..	..	..	..	..	1	..	..	..	1
e	..	..	..	..	..	..	..	..	..	..	..	..	..	1	..	..	1	..
i	..	..	..	..	..	..	..	..	..	..	..	..	..	1	..	1	..	..
o	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1	..	1	..
u	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1	1	..	..

## Παράρτημα Β

Οι 150 λέξεις που χρησιμοποιήθηκαν για την αξιολόγηση του υπολογιστικού μοντέλου.

Λέξη	Λογαριθμικά συμπίεσμένη συχνότητα με κατώτατο όριο 0,2	Μήκος Λέξης	Αριθμός Συλλαβών	Ορθογρ. Γείτονες	Διγραμ. Συχνότητα	Ορθογρ. Διαφάνεια	Συλλαβική Συχνότητα	Τονούμενη συλλαβή
αδερφές	0,2	7,0	3,0	3,0	-21,2	0,01317	17,873	Λήγουσα
άθλησης	0,2	7,0	3,0	0,0	-21,26	0,01426	2,806	Προπαραλήγουσα
άθραυστος	0,2	9,0	3,0	2,0	-25,88	0,00312	1,899	Προπαραλήγουσα
αλφαβήτου	0,2	9,0	4,0	0,0	-29,83	0,00894	16,088	Παραλήγουσα
ανδρισμός	0,2	9,0	3,0	0,0	-25,41	0,00515	17,33	Λήγουσα
ανθρακικής	0,2	10,0	4,0	1,0	-28,13	0,0119	13,992	Λήγουσα
αντιληπτά	0,2	9,0	4,0	2,0	-31,31	0,00898	13,806	Λήγουσα
αρχέγονο	0,2	8,0	4,0	3,0	-24,9	0,00393	15,028	Προπαραλήγουσα
άσμα	0,2	4,0	2,0	3,0	-13,22	0,00515	3,276	Παραλήγουσα
ασφαλείς	0,220601786	8,0	3,0	0,0	-23,16	0,0048	17,347	Λήγουσα
αυθαίρετα	0,214025772	9,0	4,0	4,0	-28,73	0,00126	16,735	Προπαραλήγουσα
αυξάνουν	0,21824096	8,0	3,0	0,0	-23,76	0,00312	17,466	Παραλήγουσα
άφθαρτος	0,2	8,0	3,0	4,0	-25,99	0,01317	1,913	Προπαραλήγουσα
βαρύτερα	0,2	8,0	4,0	3,0	-25,47	0,0035	3,985	Προπαραλήγουσα
βασάνιζε	0,2	8,0	4,0	2,0	-29,47	0,00652	1,776	Προπαραλήγουσα
βγαίναμε	0,2	8,0	3,0	2,0	-25,63	0,00126	11,985	Προπαραλήγουσα
βόλεμα	0,2	6,0	3,0	2,0	-21,03	0,00894	4,109	Προπαραλήγουσα
βρεγμένο	0,2	8,0	3,0	2,0	-24,75	0,00894	2,281	Παραλήγουσα
γλιτώσαμε	0,2	9,0	4,0	4,0	-32,04	0,00915	5,005	Προπαραλήγουσα
γνώριμοι	0,2	8,0	3,0	5,0	-26,36	0,00576	2,848	Προπαραλήγουσα
γόςης	0,2	4,0	2,0	0,0	-15,39	0,00915	0,235	Παραλήγουσα
γόπα	0,2	4,0	2,0	3,0	-14,24	0,00915	3,886	Παραλήγουσα
γύρισμα	0,2	7,0	3,0	3,0	-22,89	0,0035	3,02	Προπαραλήγουσα
δάσκαλος	0,233026124	8,0	3,0	4,0	-23,6	0,01865	0,33	Προπαραλήγουσα
δεσπότη	0,2	7,0	3,0	0,0	-19,45	0,0171	3,589	Παραλήγουσα
δήλωσαν	0,279560846	7,0	3,0	6,0	-22,62	0,0119	1,138	Προπαραλήγουσα
δίσκο	0,309290236	5,0	2,0	1,0	-16,41	0,01159	0,946	Παραλήγουσα
διχασμένο	0,2	9,0	4,0	3,0	-25,74	0,00515	4,452	Παραλήγουσα
δολοφονήσω	0,2	10,0	5,0	0,0	-32,46	0,0119	2,192	Παραλήγουσα
δώστε	0,2	5,0	2,0	3,0	-15,01	0,00925	1,083	Παραλήγουσα
εβδομάδας	0,303960123	9,0	4,0	3,0	-31,16	0,00894	7,357	Παραλήγουσα
εκδικητής	0,2	9,0	4,0	1,0	-27,03	0,0119	6,914	Λήγουσα
εκπροσώπου	0,273920027	10,0	4,0	0,0	-31,67	0,00925	9,029	Παραλήγουσα
έλαβε	0,300824123	5,0	3,0	4,0	-18,39	0,00894	5,058	Προπαραλήγουσα
ελεούσα	0,2	7,0	4,0	1,0	-23,68	0,00961	8,257	Παραλήγουσα
έλκη	0,2	4,0	2,0	3,0	-17,94	0,01266	5,764	Παραλήγουσα
εμπρός	0,274762178	6,0	2,0	1,0	-17,52	0,00249	13,288	Λήγουσα
ενεργή	0,2	6,0	3,0	4,0	-17,62	0,00594	10,426	Λήγουσα
επέβαλε	0,233306723	7,0	4,0	4,0	-22,61	0,00894	8,276	Προπαραλήγουσα
ερμηνευθεί	0,2	10,0	4,0	1,0	-30,22	0,00312	7,831	Λήγουσα

εύρος	0,227591306	5,0	2,0	2,0	-14,75	0,00121	13,288	Παραλήγουσα
εφτά	0,286291102	4,0	2,0	1,0	-15,64	0,01317	13,369	Λήγουσα
εχθρούς	0,202184954	7,0	2,0	0,0	-21,04	0,00806	13,283	Λήγουσα
εχτές	0,2	5,0	2,0	1,0	-17,26	0,00806	13,284	Λήγουσα
ζακέτα	0,2	6,0	3,0	3,0	-21,91	0,00652	4,576	Παραλήγουσα
ζορίζεται	0,2	9,0	4,0	2,0	-25,17	0,00586	2,601	Προπαραλήγουσα
ηγεμονίας	0,2	9,0	5,0	2,0	-28,59	0,00594	4,183	Παραλήγουσα
ημεδαπό	0,2	7,0	4,0	2,0	-21,58	0,0171	9,495	Λήγουσα
ήπαρ	0,2	4,0	2,0	1,0	-20,38	0,0119	1,433	Παραλήγουσα
ηχητικός	0,2	8,0	4,0	2,0	-25,0	0,00393	5,793	Λήγουσα
θεάνθρωπος	0,2	10,0	4,0	0,0	-33,09	0,01426	2,405	Προπαραλήγουσα
θέμιδα	0,2	6,0	3,0	0,0	-21,3	0,01426	2,263	Προπαραλήγουσα
θερμικές	0,2	8,0	3,0	3,0	-22,64	0,01266	1,727	Λήγουσα
ινδικό	0,2	6,0	3,0	2,0	-19,96	0,0171	2,808	Λήγουσα
ισόπαλα	0,2	7,0	4,0	2,0	-25,13	0,0171	3,221	Προπαραλήγουσα
ισοφάρισε	0,211275503	9,0	5,0	2,0	-29,51	0,01317	4,384	Προπαραλήγουσα
καθαρόαιμη	0,2	10,0	5,0	2,0	-33,91	0,00586	3,826	Προπαραλήγουσα
καθρέφτες	0,2	9,0	3,0	2,0	-27,62	0,01317	3,512	Παραλήγουσα
καλπάζει	0,2	8,0	3,0	1,0	-23,46	0,00652	3,88	Παραλήγουσα
καρπός	0,2	6,0	2,0	1,0	-18,49	0,0171	5,262	Λήγουσα
κοστίζει	0,237038616	8,0	3,0	1,0	-20,94	0,00652	1,372	Παραλήγουσα
κόψη	0,2	4,0	2,0	5,0	-12,36	0,00271	3,455	Παραλήγουσα
λαβύρινθος	0,2	10,0	4,0	2,0	-38,87	0,0035	2,695	Προπαραλήγουσα
λαούς	0,239167519	5,0	2,0	1,0	-20,36	0,00961	1,419	Λήγουσα
λάσπη	0,2	5,0	2,0	0,0	-19,88	0,01908	0,703	Παραλήγουσα
μαχόμενη	0,2	8,0	4,0	2,0	-24,94	0,00806	7,22	Προπαραλήγουσα
μεμπτές	0,2	7,0	2,0	0,0	-21,42	0,01792	8,579	Λήγουσα
μεταβίβασα	0,2	10,0	5,0	3,0	-33,65	0,00894	6,815	Προπαραλήγουσα
μέχρις	0,234010124	6,0	2,0	0,0	-14,67	0,00806	2,831	Παραλήγουσα
μισθωτοί	0,2	8,0	3,0	1,0	-27,67	0,0019	0,123	Λήγουσα
μνηστή	0,2	6,0	2,0	0,0	-20,65	0,0119	0,51	Λήγουσα
μονάδα	0,335398146	6,0	3,0	2,0	-19,58	0,01865	2,253	Παραλήγουσα
νάρκη	0,2	5,0	2,0	0,0	-20,09	0,01266	1,052	Παραλήγουσα
ναύλος	0,2	6,0	2,0	3,0	-19,92	2,4E-4	9,381	Παραλήγουσα
μέθη	0,2	4,0	2,0	4,0	-13,16	0,01426	4,009	Παραλήγουσα
νέφος	0,2	5,0	2,0	0,0	-16,32	0,01317	1,185	Παραλήγουσα
νεφρική	0,2	7,0	3,0	4,0	-21,95	0,0119	3,131	Λήγουσα
νησιά	0,306866105	5,0	2,0	0,0	-17,93	0,0010	1,479	Λήγουσα
νυχτερινές	0,2	10,0	4,0	3,0	-34,77	0,00806	2,1	Λήγουσα
ξεναγός	0,2	7,0	3,0	0,0	-21,13	0,00744	6,853	Λήγουσα
ξενύχτησα	0,2	9,0	4,0	2,0	-32,97	0,0035	1,032	Προπαραλήγουσα
ξιφία	0,2	5,0	3,0	0,0	-20,71	0,00744	17,617	Παραλήγουσα
οθόνη	0,317770261	5,0	3,0	0,0	-20,05	0,01426	7,53	Παραλήγουσα
ολόκληρα	0,24571901	8,0	4,0	3,0	-27,03	0,0171	6,983	Προπαραλήγουσα
όμματα	0,2	6,0	3,0	0,0	-17,36	9,6E-4	8,453	Προπαραλήγουσα
ομόκεντρο	0,2	9,0	4,0	1,0	-28,7	0,00898	5,978	Προπαραλήγουσα
οπλές	0,2	5,0	2,0	2,0	-15,12	0,01792	9,755	Λήγουσα
οσμή	0,2	4,0	2,0	1,0	-14,63	0,00515	9,754	Λήγουσα
παλλαϊκές	0,2	9,0	4,0	2,0	-24,17	5,2E-4	2,69	Λήγουσα
παλμός	0,2	6,0	2,0	2,0	-18,95	0,0171	3,66	Λήγουσα
πάπιες	0,2	6,0	2,0	4,0	-18,96	0,0010	1,534	Παραλήγουσα
παράγραφος	0,215812887	10,0	4,0	3,0	-27,25	0,00915	2,949	Προπαραλήγουσα

παράδοξο	0,219936528	8,0	4,0	3,0	-26,84	0,00744	3,135	Προπαραλήγουσα
παρέχω	0,2	6,0	3,0	0,0	-19,29	0,00806	2,98	Παραλήγουσα
πεζογράφος	0,2	10,0	4,0	3,0	-30,12	0,00652	1,993	Παραλήγουσα
πέψη	0,2	4,0	2,0	1,0	-13,11	0,00271	1,293	Παραλήγουσα
πίκρα	0,2	5,0	2,0	0,0	-16,29	0,01159	1,383	Παραλήγουσα
ποδιές	0,2	6,0	2,0	2,0	-16,53	0,00236	5,249	Λήγουσα
πράο	0,2	4,0	2,0	0,0	-15,13	0,01908	9,987	Παραλήγουσα
προοπτικές	0,307088969	10,0	4,0	1,0	-29,57	0,01266	6,837	Λήγουσα
προσαρμογή	0,256025961	10,0	4,0	0,0	-30,55	0,00594	2,283	Λήγουσα
προσοχής	0,2	8,0	3,0	3,0	-22,18	0,00393	2,694	Λήγουσα
προσπάθησα	0,2	10,0	4,0	1,0	-29,9	0,01426	2,673	Προπαραλήγουσα
προσφοράς	0,248220821	9,0	3,0	2,0	-25,65	0,01317	2,053	Λήγουσα
ράβδο	0,2	5,0	2,0	0,0	-21,59	0,00894	1,533	Παραλήγουσα
ρωμαϊκό	0,2	7,0	4,0	2,0	-23,81	5,2E-4	4,167	Λήγουσα
σεισμών	0,2	7,0	2,0	2,0	-18,86	0,00515	1,582	Λήγουσα
σκόνη	0,206860442	5,0	2,0	2,0	-15,54	0,0171	1,541	Παραλήγουσα
σκυτάλη	0,2	7,0	3,0	0,0	-21,34	0,01266	1,767	Παραλήγουσα
σπάμε	0,2	5,0	2,0	4,0	-16,69	0,01908	8,751	Παραλήγουσα
σπίθα	0,2	5,0	2,0	1,0	-18,12	0,01159	2,458	Παραλήγουσα
σπουδάσει	0,2	9,0	3,0	1,0	-27,12	0,0088	1,128	Παραλήγουσα
στάθμευαν	0,2	9,0	3,0	0,0	-27,34	0,00238	0,289	Προπαραλήγουσα
στέκω	0,2	5,0	2,0	2,0	-19,39	0,01765	0,198	Παραλήγουσα
σύγκριση	0,320829696	8,0	3,0	2,0	-23,86	0,00121	3,88	Προπαραλήγουσα
σύλληψη	0,286074364	7,0	3,0	1,0	-23,18	0,00205	1,345	Προπαραλήγουσα
συμβούν	0,2	7,0	2,0	0,0	-19,43	0,00894	3,711	Λήγουσα
συμμετέχει	0,294872192	10,0	4,0	0,0	-26,41	9,6E-4	3,053	Παραλήγουσα
συμπαθές	0,2	8,0	3,0	3,0	-21,7	0,00249	2,531	Λήγουσα
σύνταξη	0,307088969	7,0	3,0	3,0	-20,51	0,0035	1,285	Προπαραλήγουσα
σφαίρα	0,251208483	6,0	2,0	0,0	-20,36	0,00126	3,357	Παραλήγουσα
σωτήρ	0,2	5,0	2,0	0,0	-20,94	0,0119	0,741	Λήγουσα
τάγμα	0,2	5,0	2,0	1,0	-14,73	0,00915	2,432	Παραλήγουσα
ταχυδρόμος	0,2	10,0	4,0	2,0	-32,35	0,00393	3,42	Παραλήγουσα
τέντωμα	0,2	7,0	3,0	2,0	-19,37	0,00898	3,58	Προπαραλήγουσα
τετράδα	0,2	7,0	3,0	0,0	-21,04	0,01865	3,117	Παραλήγουσα
τίγρη	0,2	5,0	2,0	1,0	-17,35	0,00915	0,845	Παραλήγουσα
τόλμη	0,2	5,0	2,0	1,0	-17,84	0,0171	0,821	Παραλήγουσα
τραπεζάκι	0,2	9,0	4,0	0,0	-31,88	0,00652	2,303	Παραλήγουσα
τραπεζίτης	0,2	10,0	4,0	1,0	-28,13	0,00652	4,254	Παραλήγουσα
τριπλή	0,2	6,0	2,0	2,0	-19,47	0,0119	0,393	Λήγουσα
υδραυλικά	0,2	9,0	4,0	2,0	-29,99	0,00238	3,049	Λήγουσα
υδρόψυκτος	0,2	10,0	4,0	1,0	-36,13	0,00271	1,495	Προπαραλήγουσα
υπέγραψε	0,253765313	8,0	4,0	3,0	-29,36	0,00271	2,223	Προπαραλήγουσα
υπερβολικά	0,239167519	10,0	5,0	2,0	-31,41	0,00894	4,401	Λήγουσα
υπήκοος	0,2	7,0	4,0	3,0	-27,44	0,0119	2,429	Προπαραλήγουσα
ύποπτα	0,2	6,0	3,0	2,0	-22,16	0,0035	3,645	Προπαραλήγουσα
φέρει	0,2	4,0	2,0	4,0	-13,65	0,01317	1,646	Παραλήγουσα
φιλοδοξίες	0,246561749	10,0	5,0	1,0	-36,23	0,00744	1,805	Παραλήγουσα
φιλόξενες	0,2	9,0	4,0	2,0	-31,87	0,00744	1,271	Προπαραλήγουσα
φορολογικά	0,219599309	10,0	5,0	2,0	-28,17	0,00594	3,344	Λήγουσα
φραπέ	0,2	5,0	2,0	0,0	-21,75	0,01317	0,825	Λήγουσα
φταίμε	0,2	6,0	2,0	2,0	-21,02	0,00126	8,589	Παραλήγουσα
χαμογέλασα	0,2	10,0	5,0	2,0	-33,33	0,00594	2,047	Προπαραλήγουσα



χαρακτήρες	0,241753515	10,0	4,0	3,0	-30,68	0,00806	2,397	Παραλήγουσα
χρονική	0,283034216	7,0	3,0	2,0	-19,17	0,00806	3,187	Λήγουσα
χρυσοχείο	0,2	10,0	5,0	1,0	-36,3	0,0048	5,867	Παραλήγουσα
ψηφώνται	0,2	8,0	3,0	1,0	-25,42	0,00271	2,892	Προπαραλήγουσα
ψηφοθηρικό	0,2	10,0	5,0	2,0	-33,32	0,00271	4,001	Λήγουσα
ώσπου	0,275179923	5,0	2,0	0,0	-16,94	0,00925	0,842	Παραλήγουσα

Οι 150 ψευδολέξεις που χρησιμοποιήθηκαν για την αξιολόγηση του υπολογιστικού μοντέλου.

Λέξη	Μήκος Λέξης	Αριθμός Συλλαβών	Ορθογραφικοί Γείτονες	Διγραμματική Συχνότητα	Ορθογραφική Διαφάνεια	Συλλαβική Συχνότητα	Τονούμενη συλλαβή
αγκρισμος	9,0	3,0	0,0	-25,45	0,00121	17,0	Λήγουσα
αγλαρτος	8,0	3,0	0,0	-27,41	0,00915	17,0	Λήγουσα
αιμνετη	7,0	3,0	0,0	-23,56	0,00586	3,0	Παραλήγουσα
αξηκας	6,0	3,0	0,0	-19,09	0,00744	17,0	Παραλήγουσα
αρτιμλικος	10,0	4,0	0,0	-52,14	0,01159	13,0	Προπαραλήγουσα
αυρακανα	8,0	4,0	0,0	-22,56	0,00238	19,0	Παραλήγουσα
βαμυπερα	8,0	4,0	0,0	-27,16	0,00894	2,0	Παραλήγουσα
βιλανα	6,0	3,0	0,0	-19,99	0,00894	7,0	Παραλήγουσα
βοκενα	6,0	3,0	0,0	-18,9	0,00894	7,0	Παραλήγουσα
γαλτα	5,0	2,0	2,0	-20,36	0,00915	0,495	Λήγουσα
γαρονετιζω	10,0	5,0	0,0	-32,28	0,00652	2,0	Παραλήγουσα
γκαριμοκι	9,0	4,0	0,0	-29,14	0,00121	3,0	Παραλήγουσα
γκοιμενετε	10,0	4,0	0,0	-29,81	0,00041	4,0	Προπαραλήγουσα
γυτισκα	7,0	3,0	0,0	-23,96	0,00594	0,723	Παραλήγουσα
δειξαπτο	7,0	3,0	0,0	-22,02	0,00744	4,0	Παραλήγουσα
δεκλαμη	7,0	3,0	0,0	-22,26	0,01865	1,0	Παραλήγουσα
δηκωραν	7,0	3,0	0,0	-23,96	0,0119	0,496	Προπαραλήγουσα
δηλακτητα	8,0	4,0	0,0	-26,84	0,01266	5,0	Προπαραλήγουσα
διτρι	5,0	2,0	0,0	-16,02	0,01159	5,0	Λήγουσα
δογορονησω	10,0	5,0	0,0	-32,16	0,00915	2,0	Παραλήγουσα
δουτρος	7,0	2,0	0,0	-18,36	0,01473	0,177	Λήγουσα
δυζισεις	8,0	3,0	0,0	-24,14	0,00652	1,0	Παραλήγουσα
δωλκε	5,0	2,0	0,0	-22,93	0,00925	0,252	Παραλήγουσα
εγκα	4,0	2,0	0,0	-12,49	0,00121	6,0	Παραλήγουσα
ελδοπαδας	9,0	4,0	0,0	-32,96	0,01865	7,0	Παραλήγουσα
ελτικο	6,0	3,0	0,0	-19,28	0,02776	10,0	0,0
εντλος	6,0	2,0	0,0	-19,82	0,00898	6,0	Παραλήγουσα
ερνητευθει	10,0	4,0	0,0	-31,16	0,00312	9,0	Λήγουσα
ερτωμιδα	8,0	4,0	0,0	-27,4	0,01159	7,0	Παραλήγουσα
ευτο	4,0	2,0	5,0	-12,41	0,00099	13,0	Παραλήγουσα
ευτρος	6,0	2,0	0,0	-17,76	0,00099	13,0	Παραλήγουσα
ζατεπα	6,0	3,0	0,0	-21,99	0,00652	3,0	Παραλήγουσα
ζενοπι	6,0	3,0	0,0	-20,43	0,00652	3,0	Παραλήγουσα

ζερελομινα	10,0	5,0	0,0	-29,55	0,00652	5,0	Παραλήγουσα
ζομιθεται	9,0	4,0	0,0	-28,13	0,00586	3,0	Προπαραλήγουσα
ηγηπιβος	8,0	4,0	0,0	-25,92	0,00594	6,0	Προπαραλήγουσα
ηθεφαπο	7,0	4,0	0,0	-21,85	0,01317	6,0	Προπαραλήγουσα
ηκρο	4,0	2,0	2,0	-15,68	0,0119	1,0	Παραλήγουσα
ημωρ	4,0	2,0	1,0	-19,95	0,0119	1,0	Παραλήγουσα
ηπρακαδι	8,0	4,0	0,0	-26,02	0,01865	7,0	Παραλήγουσα
ηρεξονιας	9,0	4,0	0,0	-28,48	0,00744	4,0	Παραλήγουσα
ησμακας	7,0	3,0	0,0	-21,75	0,00515	4,0	Παραλήγουσα
θεβιρα	6,0	3,0	0,0	-22,59	0,00894	3,0	Παραλήγουσα
θελμισες	8,0	3,0	0,0	-24,81	0,01426	0,836	Προπαραλήγουσα
ικοπαζα	7,0	4,0	0,0	-24,62	0,00652	4,0	Προπαραλήγουσα
καδι	4,0	2,0	8,0	-10,55	0,01865	7,0	Παραλήγουσα
κατλης	6,0	2,0	0,0	-18,12	0,0119	5,0	Λήγουσα
κατσακονα	9,0	4,0	0,0	-26,65	0,00104	9,0	Παραλήγουσα
κηξη	4,0	2,0	6,0	-13,05	0,00744	3,0	Παραλήγουσα
κλιπα	5,0	2,0	3,0	-16,47	0,01159	4,0	Παραλήγουσα
κλιταρμι	8,0	3,0	0,0	-26,5	0,01908	1,0	Παραλήγουσα
κορταμα	7,0	3,0	0,0	-21,71	0,0171	5,0	Προπαραλήγουσα
κουλτα	6,0	2,0	0,0	-19,76	0,00961	0,87	Παραλήγουσα
λαμηγιθος	9,0	4,0	0,0	-33,01	0,0119	2,0	Προπαραλήγουσα
λαργη	5,0	2,0	0,0	-17,64	0,00594	0,789	Παραλήγουσα
λεοις	5,0	2,0	1,0	-17,44	0,00576	0,719	Παραλήγουσα
λοθνι	5,0	2,0	0,0	-21,69	0,01159	2,0	Λήγουσα
λουτραμιδι	10,0	4,0	0,0	-28,83	0,01159	3,0	Παραλήγουσα
μαβομεξη	8,0	4,0	0,0	-24,67	0,00744	7,0	Προπαραλήγουσα
μαγνι	5,0	2,0	2,0	-18,11	0,00915	4,0	Λήγουσα
μεκλη	5,0	2,0	0,0	-15,38	0,0119	9,0	Λήγουσα
μελδου	6,0	3,0	0,0	-23,75	0,0035	12,0	Παραλήγουσα
μενθις	6,0	2,0	1,0	-20,93	0,01159	9,0	Λήγουσα
μιργωτοι	8,0	3,0	0,0	-26,19	0,0019	1,0	Λήγουσα
μοπραθα	7,0	3,0	0,0	-20,11	0,01426	2,0	Προπαραλήγουσα
μπρολι	6,0	2,0	0,0	-16,93	0,00249	2,0	Παραλήγουσα
ναβη	4,0	2,0	3,0	-17,34	0,00894	1,0	Παραλήγουσα
νεμιρω	6,0	3,0	0,0	-24,57	0,01159	2,0	Παραλήγουσα
νευθας	6,0	2,0	0,0	-19,14	0,00099	2,0	Παραλήγουσα
ντροσι	7,0	2,0	0,0	-20,38	0,00898	2,0	Παραλήγουσα
νυλτεπινες	10,0	4,0	0,0	-35,83	0,01159	0,805	Παραλήγουσα
ξερυληρα	9,0	4,0	0,0	-32,65	0,0035	1,0	Προπαραλήγουσα
οθλανης	7,0	3,0	0,0	-23,2	0,01426	7,0	Παραλήγουσα
ολτι	4,0	2,0	2,0	-15,21	0,0171	6,0	Παραλήγουσα
ομογεμπρο	9,0	4,0	0,0	-28,37	0,00249	6,0	Προπαραλήγουσα
ορεθος	6,0	3,0	0,0	-20,84	0,01426	7,0	Λήγουσα
οσμηκας	7,0	3,0	0,0	-21,94	0,00515	7,0	Παραλήγουσα
οστακουδα	9,0	4,0	0,0	-28,14	0,00961	6,0	Παραλήγουσα
παγκουλωνω	10,0	4,0	0,0	-29,59	0,00121	2,0	Παραλήγουσα
παδεφω	6,0	3,0	0,0	-22,52	0,01317	3,0	Παραλήγουσα
παζιστε	7,0	3,0	0,0	-21,47	0,00652	3,0	Παραλήγουσα
παζοβραφος	10,0	4,0	0,0	-30,25	0,00652	2,0	Παραλήγουσα
πακαϊζες	8,0	4,0	0,0	-26,17	0,00009	4,0	Παραλήγουσα
πακεινουδα	10,0	4,0	0,0	-31,03	0,0088	2,0	Παραλήγουσα
παλλικωες	9,0	4,0	0,0	-32,82	0,00205	2,0	Προπαραλήγουσα

παρακυλθος	10,0	4,0	0,0	-33,7	0,01266	3,0	Προπαραλήγουσα
παριβαλα	8,0	4,0	0,0	-22,8	0,00894	5,0	Παραλήγουσα
πασαροψο	8,0	4,0	0,0	-28,27	0,00271	3,0	Προπαραλήγουσα
πατσακονας	10,0	4,0	0,0	-29,52	0,00104	3,0	Παραλήγουσα
περιαμαιγη	10,0	5,0	0,0	-29,88	0,00236	2,0	Προπαραλήγουσα
περογγαλος	10,0	4,0	0,0	-28,3	0,00027	2,0	Προπαραλήγουσα
πιρλα	5,0	2,0	1,0	-22,56	0,01159	0,984	Παραλήγουσα
πλοιμας	7,0	2,0	0,0	-20,24	0,0019	0,615	Παραλήγουσα
πολτεμνεος	10,0	4,0	0,0	-37,31	0,01792	3,0	Προπαραλήγουσα
ποντας	6,0	2,0	2,0	-14,79	0,00898	5,0	Παραλήγουσα
πορμικο	7,0	3,0	0,0	-21,25	0,0171	4,0	Προπαραλήγουσα
ποσκεις	7,0	2,0	0,0	-18,29	0,0088	4,0	Παραλήγουσα
πουφη	6,0	2,0	2,0	-18,08	0,00961	0,077	Παραλήγουσα
πρεσλανι	8,0	3,0	0,0	-26,69	0,01908	2,0	Παραλήγουσα
πρευτα	6,0	2,0	0,0	-17,51	0,00312	0,304	Λήγουσα
προμαδι	7,0	3,0	0,0	-18,17	0,01865	6,0	Παραλήγουσα
προσατησα	10,0	4,0	0,0	-33,06	0,00515	4,0	Προπαραλήγουσα
προσμοβας	9,0	3,0	0,0	-25,56	0,00515	2,0	Παραλήγουσα
πυτρι	5,0	2,0	0,0	-18,89	0,01159	0,338	Λήγουσα
ραφθο	5,0	2,0	0,0	-20,37	0,01317	1,0	Παραλήγουσα
ρελταμος	8,0	3,0	0,0	-29,66	0,01908	0,836	Παραλήγουσα
ρεφορτι	7,0	3,0	0,0	-27,84	0,01317	0,959	Παραλήγουσα
ρονείγω	7,0	4,0	0,0	-52,12	0,00009	2,0	Παραλήγουσα
σαιαρ	6,0	2,0	0,0	-22,86	0,00126	0,017	Παραλήγουσα
σαμες	5,0	2,0	7,0	-16,14	0,01908	0,162	Παραλήγουσα
σαντρεδες	9,0	3,0	0,0	-25,77	0,00898	0,91	Παραλήγουσα
σεκιπι	6,0	3,0	0,0	-19,67	0,01159	5,0	Παραλήγουσα
σικλι	5,0	2,0	0,0	-16,35	0,01159	2,0	Λήγουσα
σιρτεμακι	9,0	4,0	0,0	-33,38	0,01266	2,0	Παραλήγουσα
σπουγατει	9,0	3,0	0,0	-24,8	0,0088	0,447	Παραλήγουσα
στακωθητα	9,0	4,0	0,0	-25,24	0,00925	5,0	Προπαραλήγουσα
σταλμαυαν	9,0	3,0	0,0	-28,54	0,00238	0,344	Προπαραλήγουσα
σταμουδα	8,0	3,0	0,0	-22,72	0,00961	2,0	Παραλήγουσα
στεμακιδας	10,0	4,0	0,0	-30,16	0,01159	3,0	Παραλήγουσα
στογι	5,0	2,0	1,0	-13,92	0,00594	3,0	Λήγουσα
στοιλαμοκα	10,0	4,0	0,0	-28,53	0,00576	4,0	Παραλήγουσα
συλλαθη	7,0	3,0	1,0	-22,38	0,00205	3,0	Λήγουσα
συμμαξους	9,0	3,0	1,0	-25,84	0,00096	0,918	Προπαραλήγουσα
συμπατεσει	10,0	4,0	0,0	-27,03	0,00249	3,0	Παραλήγουσα
συνθουν	7,0	2,0	1,0	-20,22	0,00961	4,0	Λήγουσα
ταβυτροκος	10,0	4,0	0,0	-33,41	0,00894	3,0	Παραλήγουσα
τεγακοπα	8,0	4,0	0,0	-23,88	0,00915	5,0	Παραλήγουσα
τηλδι	5,0	2,0	0,0	-20,89	0,01159	4,0	Λήγουσα
τομπρη	6,0	2,0	0,0	-17,34	0,00249	0,816	Παραλήγουσα
τουγι	5,0	2,0	1,0	-13,24	0,00594	1,0	Παραλήγουσα
τραμακι	7,0	3,0	0,0	-23,06	0,01266	2,0	Παραλήγουσα
τριγκη	6,0	2,0	0,0	-18,4	0,00041	0,264	Λήγουσα
τριδηκα	7,0	3,0	0,0	-19,68	0,01159	5,0	Προπαραλήγουσα
τσειμανω	8,0	3,0	0,0	-26,67	0,00104	1,0	Παραλήγουσα
τσεκοτα	7,0	3,0	0,0	-23,77	0,00104	2,0	Λήγουσα
υγραπλικα	9,0	4,0	0,0	-27,07	0,00915	4,0	Προπαραλήγουσα
υμεγρασε	8,0	4,0	0,0	-24,67	0,00915	5,0	Προπαραλήγουσα

υπερμαφισα	10,0	5,0	0,0	-33,01	0,01317	3,0	Προπαραλήγουσα
υπωντα	6,0	3,0	0,0	-20,81	0,0035	0,713	Προπαραλήγουσα
φανιζισε	8,0	4,0	0,0	-25,99	0,00652	3,0	Προπαραλήγουσα
φερτιμογης	10,0	4,0	0,0	-32,36	0,00594	0,673	Παραλήγουσα
φιλογοριες	10,0	5,0	0,0	-31,75	0,00915	3,0	Παραλήγουσα
χανοβεκασα	10,0	5,0	0,0	-31,42	0,00806	4,0	Προπαραλήγουσα
χρατι	5,0	2,0	1,0	-15,1	0,00806	6,0	Παραλήγουσα
χρειπα	6,0	2,0	0,0	-18,19	0,0048	4,0	Παραλήγουσα
χρητες	6,0	2,0	0,0	-18,65	0,00806	0,567	Παραλήγουσα
χρητηρας	8,0	3,0	0,0	-24,45	0,00806	0,819	Παραλήγουσα
ψησοκηρικο	10,0	5,0	0,0	-33,17	0,00271	4,0	Λήγουσα
ψονα	4,0	2,0	4,0	-18,12	0,00271	9,0	Παραλήγουσα
ωλτου	5,0	2,0	0,0	-21,1	0,00925	0,792	Παραλήγουσα