



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ**

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΕΦΑΡΜΟΣΜΕΝΑ ΜΑΘΗΜΑΤΙΚΑ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗ»**

**ΜΙΑ ΠΛΗΡΗΣ ΜΕΘΟΔΟΣ ΑΥΤΟΜΑΤΗΣ
ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΗΣ ΝΕΟΕΛΛΗΝΙΚΗΣ ΓΛΩΣΣΑΣ
ΩΣ ΒΑΣΗ ΑΝΑΠΤΥΞΗΣ ΑΝΟΙΚΤΟΥ ΕΚΠΑΙΔΕΥΤΙΚΟΥ
ΛΟΓΙΣΜΙΚΟΥ ΓΙΑ ΤΗ ΔΙΔΑΣΚΑΛΙΑ ΜΕ ΤΗ ΒΟΗΘΕΙΑ
ΥΠΟΛΟΓΙΣΤΗ ΣΤΗΝ ΤΑΞΗ Ή ΣΤΟ ΔΙΑΔΙΚΤΥΟ**

Μπολιάρη Ελένη

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΙΩΑΝΝΙΝΑ, 2020

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των Σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στα Εφαρμοσμένα Μαθηματικά και Πληροφορική που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την --/--/2019 από την Εξεταστική Επιτροπή:

Όνοματεπώνυμο **Βαθμίδα**

Επιβλέπων:

Σωκράτης Μπαλτζής Αφυπηρετήσας Λέκτορας

τα Μέλη ΔΕΠ:

Δημήτριος Νούτσος Καθηγητής

Χάρης Παπαδόπουλος Επίκουρος Καθηγητής

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

«Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς ηθικούς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασίας της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα στην εργασία αυτή»

Ελένη Μπολιάρη

Ευχαριστίες

Ολοκληρώνοντας τον μεταπτυχιακό κύκλο σπουδών και της διατριβής μου επιθυμώ να εκφράσω τις ευχαριστίες μου σε όλους όσους με βοήθησαν να ολοκληρώσω το έργο μου αυτό.

Ιδιαίτέρως ευχαριστώ τον επιβλέποντα αυτής της διπλωματικής κ.κ. Σωκράτη Μπαλτζή.

Τέλος θερμά θέλω να ευχαριστήσω για όλη την στήριξη και τη βοήθεια που μου παρείχαν: η κόρη μου Μαρία-Καλλιρρόη Μπόνια και ο σύζυγος μου Μπόνιας Αναστάσιος.

Ιωάννινα, Ιούνιος 2019

Ελένη Μπολιάρη

ΠΕΡΙΛΗΨΗ

Αντικείμενο αυτής της διατριβής είναι η παρουσίαση ενός πλήρους συστήματος αυτόματης επεξεργασίας της νεοελληνικής γλώσσας (NEΓ) για την ανάπτυξη ανοικτού εκπαιδευτικού λογισμικού, το οποίο υποστηρίζει την εν γένει διδασκαλία σύγχρονη ή ασύγχρονη με τη βοήθεια Η/Υ στην τάξη ή στο διαδίκτυο. Η διατριβή στηρίζεται αποκλειστικά στις δημοσιεύσεις για την επεξεργασία της NEΓ [1-17,19,27] και στην καταγεγραμμένη εμπειρία που αποκτήθηκε από τη σχετική σχεδίαση και υλοποίηση διαδικτυακού εκπαιδευτικού λογισμικού [22,23]. Δίνεται ένα χαρακτηριστικό παράδειγμα των προτεινόμενων υπολογιστικών εργαλείων για τη διδασκαλία της NEΓ με τη βοήθεια Η/Υ στην τάξη ή στο διαδίκτυο. Το παράδειγμα αυτό είναι αντιπροσωπευτικό για τη διδασκαλία οποιουδήποτε άλλου μαθήματος, όπως για παράδειγμα των Μαθηματικών, κ.λπ., ακόμα και για μαθητές, πρώτης σχολικής ηλικίας, με μαθησιακά προβλήματα δυσλεξίας ή δυσαριθμησίας. Επίσης, γίνεται αναφορά στα αντίστοιχα προαναφερθέντα λογισμικά «Λεξικράτης» και «Αριθμοσθένης», τα οποία σχεδιάστηκαν και υλοποιήθηκαν από το Εργαστήριο Επεξεργασίας Φυσικής Γλώσσας και Μαθηματικών Προβλημάτων του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων και διατέθηκαν δοκιμαστικά σε σχολικές μονάδες σε όλη την επικράτεια. Έτσι αναλυτικότερα: Στο Κεφάλαιο 1, δίνονται συνοπτικά οι έννοιες και οι αρχές του Ανοικτού Εκπαιδευτικού Λογισμικού, της Διδασκαλίας με τη βοήθεια Η/Υ στην τάξη ή στο διαδίκτυο και της Επεξεργασίας Φυσικής Γλώσσας. Στο Κεφάλαιο 2, περιγράφεται ένα αποτελεσματικό Σύστημα Επεξεργασίας της NEΓ [1-17,19,28] και αναλύονται οι συνιστώσες του που είναι αφενός οι Μητροειδείς Γραμματικές και οι Χαρακτηριστικοί Εκθέτες αφετέρου το Βασικό Νεοελληνικό Πολυλεξικό (BNΠ) και οι Αλγόριθμοι που το συνοδεύουν. Το BNΠ αποτελείται από: το Βασικό Νεοελληνικό Λεξικό, το Ηλεκτρονικό Υπολογιστικό Νεοελληνικό Μορφολογικό Λεξικό, το Ηλεκτρονικό Υπολογιστικό Νεοελληνικό Σημασιολογικό Λεξικό και το Ηλεκτρονικό Υπολογιστικό Νεοελληνικό Συντακτικό Λεξικό καθένα από αυτά χειρίζεται αντίστοιχα τις λέξεις και τις κωδικοποιημένες πληροφορίες της μορφολογίας, της σημασιολογίας και της σύνταξης της NEΓ. Επίσης γίνεται και σχετική αναφορά σε συναφείς προσεγγίσεις στην Επεξεργασία της NEΓ. Στο Κεφάλαιο 3, αναδεικνύονται τα Υπολογιστικά Λεξικά της Νεοελληνικής οι αλγόριθμοι και οι εφαρμογές τους ως οι βάσεις για διδασκαλία χρησιμοποιώντας τη NEΓ σε ανοικτό περιβάλλον με τη βοήθεια του Η/Υ

στην τάξη ή στο διαδίκτυο. Επίσης, δίνεται η περιγραφή και παραδείγματα της διαδικασίας υλοποίησης του Τεχνολόγου και του Γεννήτορα προτάσεων της ΝΕΓ καθώς και του Διορθωτή και Σημασιολογικού Αναλυτή προτάσεων της ΝΕΓ. Το Κεφάλαιο 4, περιλαμβάνει Συγκρίσεις και Συμπεράσματα καθώς και προτάσεις για περαιτέρω μελέτη και έρευνα. Τέλος στις Βιβλιογραφικές Αναφορές, αυτές ομαδοποιούνται σε ενότητες σύμφωνα με τη σειρά που έγινε η προσέγγιση τους κατά τη διάρκεια της ετοιμασίας αυτής της διατριβής.

Abstract

The aim of this dissertation is to present a complete and comprehensive method of Modern Greek Language (MGL) Processing as a tool of a Computer Assisted Learning Method (CALM) in the classroom or by internet. The suggested method is based solely on the MGL processing publications [1-17,19,28] and on the recorded experience gained from the design and implementation of online educational software [22,23]. This suggested CALM supports a synchronous or an asynchronous open educational software. In addition an example of a Computer Assisted MGL Learning Method in the classroom or by internet based on the suggested system is given. This example constitutes a guideline for the implementation of the suggested CALM either to other topics too, as Mathematics etc, even or to preprimary school pupils with learning disabilities as of dyslexia and malnutrition. Reference is also made to the corresponding online educational software "Lexikratis" and "Arithmosthenes" mentioned above, which were designed and implemented by the Laboratory of Natural Language and Mathematical Problems Processing of the Department of Mathematics of the University of Ioannina. We note that "Lexikratis" and "Arithmosthenes" have been put to the test in school units across the country. So in more detail: In Chapter 1, preliminaries are summarized concerning the principles and concepts of required topics as the Open Educational Software, the CALM in the classroom or by internet, and the Natural Language Processing (NLP). In Chapter 2, an efficient NLP method is described and its components and tools are analyzed [1-17,19,28]. These components and tools are the Template Grammars and the Characteristic Exponents, the Modern Greek Language Multilexicon (MGLM) and its algorithms. MGLM is consisted by the Computational MGL Morphological Lexicon, the Computational MGL Semantics Lexicon and the Computational MGL Syntactical Lexicon. Each one of MGLM Lexicons handles the words and coded information of the structure and function of the MGL morphology, semantics and syntax accordingly. Moreover, reference is also made to related approaches to MGL Processing. In Chapter 3, the Computational Lexicons of MGLM their algorithms and their applications are highlighted as the basic tools of CALM development. In addition description as well as examples of the implementation procedure of the MGL sentence Parser, Generator, Corrector and Semantic Analyzer are given. In Chapter 4 are included comparisons with other NLP and CAL methods, conclusions of the

suggested CALM as well as suggestions for further study and research. Finally the References are grouped in the order used in the preparation of the dissertation.

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη	7
Abstract	9
1 Προαπαιτούμενες Έννοιες και Ορισμοί	
1.1 Εισαγωγή	13
Λέξεις Κλειδιά	
1.2 Ανοικτό εκπαιδευτικό λογισμικό	14
1.3 Διδασκαλία με τη βοήθεια Η/Υ στην τάξη ή στο διαδίκτυο και η Επεξεργασία Φυσικής Γλώσσας	15
2. Ένα αποτελεσματικό Σύστημα Επεξεργασίας της Νεοελληνικής Γλώσσας	
2.1 Εισαγωγή	19
Λέξεις Κλειδιά	
2.2 Μητροειδείς Γραμματικές και Χαρακτηριστικοί Εκθέτες	20
2.3 Το Βασικό Νεοελληνικό Πολυλεξικό και οι Αλγόριθμοι που το συνοδεύουν	24
2.3.1 Το Βασικό Νεοελληνικό Λεξικό	24
2.3.2 Το Νεοελληνικό Υπολογιστικό Μορφολογικό Λεξικό	25
2.3.3. Το Νεοελληνικό Υπολογιστικό Σημασιολογικό Λεξικό	32
2.3.3.1 Εισαγωγή	32
2.3.3.2 Η Σημασιολογική Βάση	33
2.3.3.3 Η κατασκευή των Σημασιολογικών και Σημασιοσυντακτικών κανόνων	36
2.3.4 Το Νεοελληνικό Υπολογιστικό Συντακτικό Λεξικό	43
2.4 Συμπεράσματα και συναφείς προσεγγίσεις στην Επεξεργασία της Νεοελληνικής Γλώσσας	43
3. Τα Υπολογιστικά Λεξικά της Νεοελληνικής οι αλγόριθμοι και οι εφαρμογές τους οι βάσεις για διδασκαλία σε ανοικτό περιβάλλον με τη βοήθεια του Η/Υ στην τάξη ή στο διαδίκτυο	
3.1 Εισαγωγή	47
Λέξεις Κλειδιά	
3.2 Τεχνολόγος και Γεννήτορας προτάσεων της ΝΕΓ	48
3.2.1 Εισαγωγή	48
3.2.2 Παραδείγματα	49
	11

3.2.2.1 Τεχνολόγος προτάσεων της ΝΕΓ	49
3.2.2.2 Γεννήτορας προτάσεων της ΝΕΓ	51
3.3 Διορθωτής και Σημασιολογικός Αναλυτής προτάσεων της ΝΕΓ	53
3.3.1 Εισαγωγή	53
3.3.2 Παραδείγματα	53
3.3.2.1 Διορθωτής προτάσεων της ΝΕΓ	53
3.3.2.2 Σημασιολογικός Αναλυτής προτάσεων της ΝΕΓ	55
4. Συγκρίσεις και Συμπεράσματα - Προτάσεις για περαιτέρω έρευνα	57
5. Βιβλιογραφικές Αναφορές	63

Κεφάλαιο 1

Προαπαιτούμενες Έννοιες και Ορισμοί

1.1 Εισαγωγή

Η εμπλοκή των Νέων Τεχνολογιών γενικότερα ή του εκπαιδευτικού λογισμικού ειδικότερα ως εργαλείο διδασκαλίας στην τάξη ή στο διαδίκτυο γίνεται όλο και πιο επιτακτική. Το εκπαιδευτικό λογισμικό αξιολογείται και πέρα από την παιδαγωγική μέθοδο, την ακρίβεια και την εγκυρότητα των παρεχόμενων πληροφοριών που ακολουθεί ο σχεδιασμός του. Το εκπαιδευτικό λογισμικό αξιολογείται και ως προς την αποτελεσματικότητα της χρήσης του, δηλαδή, ως προς την ευελιξία του, με άλλα λόγια, ως προς τους όσο μεγαλύτερους βαθμούς ελευθερίας που παρέχει στον χειρισμό της αντίστοιχης γνώσης ώστε, αφενός να δίνεται στον εκπαιδευτικό η δυνατότητα να αξιοποιεί τη διδακτική εμπειρία του, αφετέρου στον εκπαιδευόμενο να μαθαίνει και να εμπεδώνει την προσδοκώμενη γνώση, σε κάθε περίπτωση, ανάλογα με τους βιορυθμούς και τις ανάγκες του. Στο κεφάλαιο αυτό δίνονται εκείνοι οι ορισμοί και οι βασικές έννοιες που αφορούν κυρίως: 1. το ανοικτό εκπαιδευτικό λογισμικό, 2. τη διδασκαλία με τη βοήθεια του ηλεκτρονικού υπολογιστή (Η/Υ) στην τάξη ή στο διαδίκτυο και τέλος 3. το εργαλείο της επεξεργασίας της φυσικής γλώσσας και ειδικά της ελληνικής που προσδίδει στο εκπαιδευτικό λογισμικό μεγάλους βαθμούς ελευθερίας στην εφαρμογή του, άρα το καθιστά και επιλέξιμο

εκπαιδευτικό εργαλείο, ως αναπτυξιακά κατάλληλο λογισμικό σε κάθε περίπτωση διδασκαλίας με χρήση την ελληνική γλώσσα.

Λέξεις Κλειδιά: Ανοικτό Εκπαιδευτικό Λογισμικό, Διδασκαλία με τη βοήθεια Υπολογιστή, CALL μέθοδος, Επεξεργασία Φυσικής Γλώσσας

1.2 Ανοικτό εκπαιδευτικό λογισμικό

Με τον όρο «Εφαρμογές των ΤΠΕ στην εκπαίδευση» εννοούμε τα «Υπολογιστικά Περιβάλλοντα» για την υποστήριξη της «Ανθρώπινης Μάθησης», δηλαδή το γνωστικό εργαλείο και το εργαλείο επικοινωνίας και συνεργασίας, που ορίζονται ως λογισμικά ανοικτού¹ τύπου. Τα λογισμικά ανοικτού τύπου είναι: 1. εποικοδομιστικά, με συνήθη μορφή εργαλείου και επιτρέπουν πολλαπλές και όχι εκ των προτέρων καθορισμένες δραστηριότητες και 2. συνεργατικά, με τη μορφή εργαλείου για επικοινωνία, κοινωνική αλληλεπίδραση και συνεργασία [29,30,31].

Ο ακρογωνιαίος λίθος του αποτελεσματικού σχεδιασμού ενός εκπαιδευτικού λογισμικού ανοικτού τύπου είναι η δυνατότητα του συστήματος για επεξεργασία της φυσικής γλώσσας και κατ' επέκταση η υπολογιστική αξιοποίηση των δεδομένων και της γνώσης ή της δομολειτουργίας των κανόνων που διέπουν το εκάστοτε θεματικό αντικείμενο του αντίστοιχου πεδίου διδασκαλίας. Τα εργαλεία της επεξεργασίας της (φυσικής) γλώσσας διδασκαλίας που απαιτούνται, και ειδικά της ελληνικής εν προκειμένω, είναι και το αντικείμενο της διατριβής.

Ένα λογισμικό κατά κανόνα ανοικτού τύπου που χαρακτηρίζεται και ως αναπτυξιακά κατάλληλο λογισμικό [29,30,31]:

1. ανταποκρίνεται στην ηλικία και τις γνώσεις των μαθητών,
2. ανταποκρίνεται στην ηλικία και τις γνώσεις των παιδιών, με προοπτική να ενταχθεί και στο αναλυτικό πρόγραμμα,
3. επιτρέπει τη διερευνητική μάθηση καθώς και τη συνεργατική μάθηση, ακόμα και με τη μορφή παιχνιδιού και
4. χρησιμοποιεί άμεση και κατάλληλη τεχνική ανατροφοδότησης.

Επίσης, το αναπτυξιακά κατάλληλο λογισμικό [29,30,31]:

1. λαμβάνει υπόψη του τις δυσκολίες και τις λανθασμένες αντιλήψεις των μαθητών και προσφέρει ευκαιρίες για την ανασυγκρότησή τους.
2. ενεργοποιεί τον προβληματισμό,

¹ Λογισμικά κλειστού τύπου

3. επιτρέπει επιπλέον,
 - 3.1 την εμπλοκή του παιδιού και την ενεργητική συμμετοχή του και ευνοεί τη συνεχή ενεργοποίηση του ενδιαφέροντός του.
 - 3.2 τη μαθησιακή αξιοποίηση του λάθους, δίνοντας στον εκπαιδευτικό τη δυνατότητα να επέμβει και να τροποποιήσει το λογισμικό.

Τέλος στο ανοικτό τύπου εκπαιδευτικό λογισμικό [29,30,31]:

1. ο υπολογιστής είναι εργαλείο για τον μαθητή και ο δάσκαλος είναι σύμβουλος ή βοηθός. Αφού πρόκειται για υπολογιστικά περιβάλλοντα μάθησης τα οποία ευνοούν τη διερεύνηση, την ανακάλυψη και την οικοδόμηση της γνώσης. Ο μαθητής χρησιμοποιεί το περιβάλλον ως εργαλείο, διερευνώντας τα «αντικείμενα» που περιέχει και ανακαλύπτει πληροφορίες και γεγονότα, κατανοεί έννοιες και τις συσχετίζει μεταξύ τους, με άλλα λόγια οικοδομεί γνώσεις μέσω πειραματισμού και επίλυσης προβλημάτων,
2. ο δάσκαλος είναι σύμβουλος ή βοηθός που απλά καθοδηγεί στη γνώση.

1.3 Διδασκαλία με τη βοήθεια Η/Υ στην τάξη ή στο διαδίκτυο και η Επεξεργασία Φυσικής Γλώσσας

Η χρήση μεθόδων επεξεργασίας φυσικής γλώσσας στον σχεδιασμό εκπαιδευτικών δραστηριοτήτων με τη βοήθεια Η/Υ και ειδικά στην εκμάθηση/διδασκαλία της γλώσσας με τη βοήθεια Η/Υ (Computer Assisted Language Learning – CALL) είναι πρακτική γνωστή από χρόνια [5,15]. Χρησιμοποιήθηκε ευρέως και σε δίγλωσσα κείμενα για να διευκολυνθούν οι μαθητές στην άμεση πρόσβασή τους σε κάθε είδους πληροφορία που αφορούσε τη μετάφραση των λέξεων, τις διαφορετικές σημασίες τους, σε περιπτώσεις πολυσημίας, και εντοπισμό και διόρθωση μορφολογικών ή συντακτικών τους λαθών που εντοπίζονται και διορθώνονται από το σύστημα.

Στην παρούσα διατριβή, όπως έχει ειπωθεί και παραπάνω μελετάται ένα ολοκληρωμένο σύστημα επεξεργασίας της νέας ελληνικής γλώσσας (ΝΕΓ) που χρησιμοποιήθηκε περαιτέρω ως βάση για την ανάπτυξη μιας διδασκαλίας της, με τη βοήθεια του Η/Υ στην τάξη ή στο διαδίκτυο, ως μητρικής ή δεύτερης γλώσσας [1-17,19,28]. Το μοντέλο αυτό τυποποιεί όλες τις γλωσσολογικές διαστάσεις της ΝΕΓ, δηλαδή τη μορφολογία, τη σύνταξη και τη σημασιολογία. Δηλαδή στο σύστημα αυτό:

1. κωδικοποιούνται και αρχειοθετούνται σε κατάλληλα σχεδιασμένες βάσεις δεδομένων τα αντίστοιχα δεδομένα των τριών γλωσσολογικών διαστάσεων της ΝΕΓ,
2. κωδικοποιούνται οι πεπερασμένου πλήθους μορφολογικοί, συντακτικοί,

σημασιολογικοί και σημασιολογικοί κανόνες της ΝΕΓ και 3. περιγράφονται με αποτελεσματικούς και ευέλικτους αλγορίθμους, γραμμικής πολυπλοκότητας οι προαναφερθέντες κανόνες λαμβάνοντας υπόψη την αλληλοεξάρτηση των κανόνων αυτών, και των αντίστοιχων δεδομένων, κατά την τεχνολόγηση ή την παραγωγή των προτάσεων της ΝΕΓ. Έτσι το σύστημα δίνει την απαιτούμενη δυνατότητα στον σχεδιαστή του εκπαιδευτικού λογισμικού να χρησιμοποιήσει τη ΝΕΓ αποτελεσματικά για τη διδασκαλία οποιοδήποτε μαθήματος, ακόμα και για τη διδασκαλία αυτής καθ'αυτής της ΝΕΓ. Στον εκπαιδευτικό και στον εκπαιδευόμενο δίνει την ευκαιρία να πειραματιστεί και να εξοικειωθεί, χρησιμοποιώντας τη ΝΕΓ, με το αντίστοιχα αντικείμενο καθώς το σύστημα εντοπίζει τα λάθη και μπορεί να ανακαλέσει τους αντίστοιχους κανόνες που διέπουν τη σωστή απάντηση στη γλώσσα ή σε θέματα του υπό διδασκαλία αντικείμενου. Η περίπτωση της διδασκαλίας της γλώσσας αποτελεί αναμφισβήτητο το αντιπροσωπευτικότερο παράδειγμα και αναλύεται στη διατριβή διεξοδικά, αφού με αυτόν τον τρόπο περιγράφονται σαφέστερα και αποδίδονται με τρόπο εύληπτο οι δυνατότητες του συστήματος ως προς τον χειρισμό και την αξιοποίηση της ΝΕΓ σε κάθε περίπτωση. Έτσι ο εκπαιδευόμενος πληροφορείται για τους μορφολογικούς, συντακτικούς και σημασιολογικούς κανόνες για τη σωστή χρήση της γλώσσας, παράλληλα με τις όποιες διορθώσεις αφορούν στις απαντήσεις του υπό διδασκαλία και εκμάθηση αντικείμενου, στην τάξη ή στο διαδίκτυο.

Το σύστημα αποτελείται από δύο κύριες συνιστώσες και άλλες τρεις βοηθητικές που στηρίζονται στις δύο πρώτες και στην πραγματικότητα είναι περαιτέρω επεκτάσεις των δυνατοτήτων τους ή βασικές εφαρμογές τους.

Η πρώτη συνιστώσα περιλαμβάνει τον Τεχνολόγο και τον Γεννήτορα των συντακτικών δομών κύριων και ακεραίων προτάσεων της ΝΕΓ. Η δεύτερη συνιστώσα περιλαμβάνει τέσσερα αλληλοεξαρτώμενα Υπολογιστικά Λεξικά της ΝΕΓ στα οποία είναι καταχωρημένα ταξινομημένα και κωδικοποιημένα τα γλωσσολογικά δεδομένα της ΝΕΓ, που αφορούν στη Μορφολογία, στη Σύνταξη και στη Σημασιολογία και αποτελούν το Βασικό Νεοελληνικό Πολυλεξικό (ΒΝΠ). Το ΒΝΠ συνοδεύεται από τους αλγορίθμους του, οι οποίοι: 1. εντοπίζουν τους τύπους των λέξεων της πρότασης, ελέγχουν την ορθότητα του σημασιολογικού συνδυασμού των εννοιών τους, εντοπίζουν τον συντακτικό ρόλο τους και μετατρέπουν την πρόταση της ΝΕΓ στην αντίστοιχη συντακτική δομή της και 2. μετασχηματίζουν τις συντακτικές δομές σε αποδεκτές προτάσεις της ΝΕΓ. Οι βοηθητικές συνιστώσες αποτελούνται από βοηθητικά εργαλεία των οποίων η λειτουργία στηρίζεται στις

δυνατότητες επεξεργασίας που παρέχονται από αυτές των δύο κύριων συνιστωσών. Αυτά είναι: 1. ο διορθωτής προτάσεων, μπορεί να διορθώσει ορθογραφικά λάθη και λάθη μορφολογικών τύπων κλιτών λέξεων που δεν συνάδουν με τον συντακτικό τους ρόλο, επίσης μπορεί να εντοπίσει λάθη σημασιολογικών συνδυασμών εννοιών λέξεων, 2. ο αναλυτής προτάσεων, ο οποίος αποφαινεται αν οι υπό μελέτη προτάσεις έχουν ίδιο ή συνώνυμο υποκείμενο, αντικείμενο ή οποιονδήποτε άλλο δευτερεύοντα όρο, ενώ ως προς το ρήμα μπορεί να αποφανθεί αν είναι και αντώνυμο. Ο αναλυτής προτάσεων μπορεί να δώσει επιπλέον πληροφορίες ως προς το νόημα των υπό μελέτη προτάσεων, τέλος 3. το εργαλείο συλλαβισμού, το οποίο βάσει κανόνων μπορεί να συλλαβίσει όλες τις λέξεις που αναγνωρίζει το σύστημα. Τα εργαλεία αυτά, κυρίως της πρώτης συνιστώσας θα μελετηθούν στα παρακάτω κεφάλαια σαν μία ολοκληρωμένη μέθοδος αλλά και ξεχωριστά.

Κεφάλαιο 2

Ένα αποτελεσματικό Σύστημα Επεξεργασίας της Νεοελληνικής Γλώσσας

2.1 Εισαγωγή

Στο κεφάλαιο αυτό περιγράφεται συνοπτικά και περιεκτικά η αποτελεσματική Μέθοδος για την Επεξεργασία της ΝΕΓ [1-17,19,28]. Η καινοτομία αυτής της μεθόδου έγκειται στο ότι 1. προσεγγίζει σφαιρικά και τις τρεις γλωσσολογικές διαστάσεις για την επεξεργασία της ΝΕΓ, αυτές είναι η μορφολογία, η σημασιολογία και η σύνταξη, 2. μελετάται, αναλύεται και τα μοντέλα που εισάγονται είναι προσαρμοσμένα στις αρχές τις ιδιότητες και τις ιδιαιτερότητες της ΝΕΓ, η οποία διαφέρει από τα διεθνή πρότυπα επεξεργασία της φυσικής γλώσσας τα οποία στηρίζονται στην αγγλική γλώσσα και 3. Οι Μέθοδος για την επεξεργασία της ΝΕΓ που μελετάμε δεν χρησιμοποιεί στατιστικές και πιθανότητες πιθανότητες αλλά εφαρμόζει τους δομολειτουργικούς κανόνες της ΝΕΓ και των τριών γλωσσολογικών διαστάσεων με ακρίβεια 100%. Αναλυτικά στις παρακάτω ενότητες περιγράφονται οι μητροειδείς γραμματικές και οι χαρακτηριστικοί εκθέτες που έχουν εισαχθεί ειδικά για την επεξεργασία της ΝΕΓ [7,8,9,16,17,19,28], καθώς και το Βασικό Νεοελληνικό Πολυλεξικό, με τα επιμέρους Υπολογιστικά Λεξικά του, που χειρίζονται τις κωδικοποιημένες πληροφορίες και των τριών γλωσσολογικών διαστάσεων της ΝΕΓ καθώς και τους αλγορίθμους που τα συνοδεύουν. Βάσει αυτών ορίζονται ο

Τεχνολόγος, ο Γεννήτορας, ο Διορθωτής και ο Σημασιολογικός Αναλυτής προτάσεων της ΝΕΓ [4,19,28].

Λέξεις Κλειδιά: Μητροειδείς Γραμματικές, Χαρακτηριστικοί Εκθέτες, Γραμματικές Ενοποίησης, Μορφολογικός Αναλυτής, Συντακτικός Αναλυτής, Σημασιολογικός Αναλυτής, Οντολογία, Υπολογιστικά Λεξικά, Σημασιολογικές κατηγορίες

2.2 Μητροειδείς Γραμματικές και Χαρακτηριστικοί Εκθέτες

Η πρώτη συνιστώσα όπως προαναφέρθηκε περιλαμβάνει τον Τεχνολόγο και τον Γεννήτορα των συντακτικών δομών κύριων και ακεραίων προτάσεων της ΝΕΓ. Ο τεχνολόγος και ο γεννήτορας όπως ορίζονται στη μελέτη μας αφορούν στις κύριες και ακεραίες προτάσεις της ΝΕΓ, τις οποίες στο εξής θα αναφέρομε ως προς προτάσεις της ΝΕΓ.

Σημειώνομε ότι οι προτάσεις της ΝΕΓ ταξινομούνται και ως προς τη συντακτική δομή τους [7,8,9,16,17,19,28], σε απλές, σύνθετες, επαυξημένες και σύνθετες και επαυξημένες. Αναλυτικότερα:

1. η δομή των απλών προτάσεων αποτελείται από τους κύριους όρους, που είναι το υποκείμενο και το κατηγορημα. Το κατηγορημα ορίζεται να είναι είτε το κατηγορηματικό ρήμα είτε το συνδεδετικό ρήμα και το κατηγορούμενο. Το κατηγορούμενο προσδίδει μία ιδιότητα στο υποκείμενο ή δράστη της ενέργειας, δηλαδή του ρήματος, στην πρόταση. Για παράδειγμα: ο καθηγητής διδάσκει, ο καθηγητής είναι επιεικής, κ.ο.κ.,
2. οι σύνθετες προτάσεις είναι εκείνες που στη δομή τους περιέχονται περισσότερα του ενός υποκείμενα ή κατηγορούμενα. Για παράδειγμα: ο πατέρας, η μητέρα και η γιαγιά ξύπνησαν, ο Παύλος είναι γλύπτης και ζωγράφος.
3. η δομή στις επαυξημένες αποτελείται από προσδιορισμούς των κυρίων όρων της πρότασης. Σημειώνομε ότι στην πραγματικότητα το αντικείμενο προσδιορίζει το ρήμα, ως αποδέκτης της ενέργειας, όπως ακριβώς μπορεί να θεωρηθεί και μία επιρρηματική φράση, π.χ. τα μικρά παιδιά παίζουν πολύ, ο Ίων αγαπά τα ζώα, ο ύποπτος προδοσίας αξιωματικός παγίδευσε τον άτυχο στρατιώτη, κ.ο.κ. και τέλος,
4. η δομή της σύνθετης και επαυξημένης πρότασης αποτελείται από περισσότερα του ενός υποκείμενα ή κατηγορούμενα αλλά και από προσδιορισμούς των όρων της. Για παράδειγμα: ο Μιλτιάδης και ο Σοφοκλής πολέμησαν τους Πέρσες, ο παππούς και το μεγαλύτερο εγγόνι του είναι διακεκριμένοι επιχειρηματίες, κ.ο.κ.

Για την κατασκευή των συνιστωσών αυτού του συστήματος προτάθηκαν μία τροποποιημένη εκδοχή των γραμματικών της Ιεραρχίας Chomsky, οι Μητροειδείς Γραμματικές (μ.γραμματικές) και οι επεκτάσεις τους οι Σπονδυλωτές Μητροειδείς Γραμματικές (σ.μ.γραμματικές). Οι πρώτες παράγουν και αναγνωρίζουν τις συντακτικές δομές των εύχρηστων τύπων των κύριων και ακεραίων προτάσεων της ΝΕΓ, ενώ οι δεύτερες τους κειμενοποιημένους τύπους της ΝΕΓ, δηλαδή, συντακτικών δομών κειμένων που αποτελούνται από κύριες και δευτερεύουσες προτάσεις της ΝΕΓ. Στην παρούσα μελέτη θα ασχοληθούμε μόνο με τις μ.γραμματικές.

Οι Γραμματικές της Ιεραρχίας Chomsky ορίζονται ως μία μαθηματική οντότητα, ένα μαθηματικό σχήμα που περιγράφει την παραγωγή ενός συνόλου, στην προκειμένη περίπτωση αλφαριθμητικών βάσει συγκεκριμένων κανόνων, το σύνολο αυτό ορίζει τη γλώσσα που η γραμματική αυτή παράγει. $G = (V_M, V_T, \Pi, S)$, όπου τα σύμβολα V_M, V_T, Π, S παριστούν αντίστοιχα, το πεπερασμένο σύνολο των μεταβλητών (ή μη τερματικών συμβόλων), το πεπερασμένο σύνολο των τερματικών στοιχείων, το πεπερασμένο σύνολο των κανόνων παραγωγής (ή των παραγωγών) και το αρχικό σύμβολο. Σημειώνουμε ότι τα σύνολα V_M και V_T δεν περιέχουν κοινά στοιχεία ($V_M \cap V_T = \emptyset$).

Οι κανόνες, δηλαδή τα στοιχεία του συνόλου Π είναι (αναδρομικοί) κανόνες «αντικατάστασης» και έχουν τη μορφή: $\alpha \rightarrow \beta$, όπου τα σύμβολα α και β παριστούν αλυσίδες αποτελούμενες από μη τερματικά σύμβολα και τερματικά στοιχεία, δηλαδή, $\alpha, \beta \in (V_M \cup V_T)^+$.

Διακρίνονται τρεις τύποι γραμματικών, που αποτελούν την Ιεραρχία Chomsky ανάλογα με ορισμένους περιορισμούς στους οποίους υποβάλλονται οι αλυσίδες α και β .

Έτσι όταν τα μήκη των αλυσίδων α και β υπόκεινται στον περιορισμό $|\alpha| \leq |\beta|$, η γραμματική λέγεται τύπου 1 ή ευπαθής στα συμφραζόμενα (context sensitive).

Εάν κάθε κανόνας μίας γραμματικής G είναι της μορφής $A \rightarrow \alpha$, όπου $A \in V_M$ και $\alpha \in (V_M \cup V_T)^+$, τότε η γραμματική λέγεται τύπου 2 ή ανεξάρτητη συμφραζομένων (context free).

Εάν η γραμματική έχει κανόνες μόνο της μορφής $A \rightarrow aB$ ή της μορφής $A \rightarrow a$, όπου $\{A, B\} \subseteq V_M$ και $a \in V_T$, τότε λέγεται τύπου 3 ή κανονική (regular). Σημειώνουμε ότι αν οι μεταβλητές στους κανόνες αντικαθίστανται από δεξιά, το μήκος δηλαδή, των παραγόμενων αλυσίδων αυξάνεται προς τα δεξιά τότε οι γραμματικές

λέγονται και δεξιά γραμμικές, αν αντίστοιχα η διαδικασία γίνεται από αριστερά τότε έχουμε αριστερά γραμμικές.

Όταν οι κανόνες δεν υπόκεινται σε κανέναν περιορισμό η γραμματική λέγεται τύπου 0.

Πριν την παρουσίαση των μ.γραμματικών κρίνεται σκόπιμο να αναφερθούμε αφενός στη σημασία των γραμματικών στην επεξεργασία φυσικής γλώσσας (ΕΦΓ) αλλά και γιατί προτάθηκαν οι μ.γραμματικές για την επεξεργασία της ΝΕΓ.

Είναι γνωστό ότι η επεξεργασία φυσικής γλώσσας (ΕΦΓ) έγκειται στην ανάλυση του κειμένου της, η οποία προϋποθέτει τη συντακτική του ανάλυση ή την τεχνολόγησή του, κατά τη διάρκεια της οποίας κάθε πρότασή του «απογραμματικοποιείται», δηλαδή από τις λέξεις της πρότασης εξάγεται μία δενδροειδής δομή. Αυτή η δομή περιγράφει τον συντακτικό ρόλο κάθε λέξης της πρότασης. Η δυνατότητα ή μη εξαγωγής της δενδροειδούς δομής επιβεβαιώνει αν η πρόταση ανήκει ή όχι στη γλώσσα. Στη διαδικασία αυτή τον κεντρικό ρόλο τον παίζει η γραμματική, η οποία είναι ένα σχήμα που περιγράφει τα χαρακτηριστικά των προτάσεων της γλώσσας. Στην περιγραφή των φυσικών γλωσσών οι γραμματικές Chomsky, καθώς και διάφορες τροποποιήσεις και επαυξήσεις που οφείλονται είτε στον Chomsky είτε σε άλλους ερευνητές παρουσίασαν κάποιες ανεπάρκειες. Πρέπει όμως να σημειώσουμε ότι η ανεπάρκεια αυτή παρατηρήθηκε και καταγράφηκε κυρίως στην περίπτωση της αγγλικής γλώσσας (ΑΓ), η οποία διαφέρει σημαντικά από τη ΝΕΓ και από άποψη πληροφοριακού περιεχομένου των λέξεων (γένη, πτώσεις, κλπ.).

Οι ανεπάρκειες αυτές για την ΑΓ εντοπίζονται κυρίως στο ότι στις προταθείσες γραμματικές δεν ενσωματώνονται, σε ανάλογο βαθμό, μορφολογικές και κυρίως σημασιολογικές πληροφορίες, οι οποίες ως γνωστόν είναι αλληλοεξαρτώμενες με αυτές της σύνταξης. Επίσης στο ότι το μοντέλο τους δεν είναι γενικό, «καθολικό», για όλη τη γλώσσα για την οποία σχεδιάστηκαν, αλλά προτείνονται τόσες περιπτώσεις όσες για κάθε περίπτωση πρότασης ξεχωριστά. Χαρακτηριστικό παράδειγμα είναι οι γραμματικές ενοποίησης [27], όπου γίνεται μία προσπάθεια ενσωμάτωσης μορφολογικών και σημασιολογικών πληροφοριών, τα μοντέλα που προτάθηκαν είναι δύσκολα στην υλοποίηση εξαιτίας της πολυπλοκότητας του σχεδιασμού τους, η οποία επιδεινώνεται καθώς δεν προβλέπεται, και πάλι, ένα καθολικό μοντέλο που να καλύπτει όλες τις περιπτώσεις. Για την ΕΦΓ έχουν προταθεί κατά καιρούς και άλλα μοντέλα που χρησιμοποιούν πιθανότητες.

Οι δομολειτουργικές διεργασίες της ΝΕΓ μπορούν να περιγραφούν ως αλγόριθμοι, σε συνδυασμό δε με τη μελέτη, ταξινόμηση, κωδικοποίηση και αρχειοθέτηση των γλωσσολογικών της δεδομένων, κατάλληλα σε βάσεις δεδομένων, μπορούμε να προσεγγίσουμε την επεξεργασία της αναλυτικά και με στόχους ακρίβειας 100% [16,17,19].

Η παρατήρηση ότι στη μη κλιτή ή ημίκλιτη ΑΓ οι πληροφορίες που φέρουν οι λέξεις στην πρόταση σχετικά με τον συντακτικό τους ρόλο καθορίζεται από την θέση τους (πρβλ SVO) ενώ στην κλιτή ΝΕΓ από τις πτώσεις τους οδηγεί στο συμπέρασμα ότι οι κανόνες φρασεοδομής των γραμματικών της ιεραρχίας Chomsky, που είναι και η βάση όλων των επεκτάσεων ή επαυξήσεων αυτών και εκφράζουν τη συντακτική δομή των προτάσεων της ΑΓ, δεν μπορούν να εκφράσουν επαρκώς τη συντακτική δομή των προτάσεων της ΝΕΓ. Για παράδειγμα η πρόταση: «η γάτα τρώει τον ποντικό» στη ΝΕΓ μπορεί να διατυπωθεί διατηρώντας το ίδιο νόημα με 6 διαφορετικούς τρόπους αλλάζοντας τη θέση των λέξεων. Στην ΑΓ όμως: «the cat eats the mouse» διατυπώνεται μόνο με έναν τρόπο. Η πρόταση: «the mouse eats the cat», η μοναδική επιτρεπτή αλλαγή λέξεων, σημαίνει κάτι τελείως διαφορετικό. Επισημαίνουμε το πόσο σημαντική είναι η παθητική φωνή (passive voice) για την ΑΓ, αφού είναι η μοναδική δυνατότητα που έχει να αντιστρέψει τη θέση υποκειμένου και αντικειμένου, χωρίς να αλλάξει το νόημα της πρότασης, εισάγοντας το ποιητικό αίτιο, μία πρακτική η οποία στη ΝΕΓ δεν είναι διαδεδομένη [5,7,19].

Η διατύπωση λοιπόν γραμματικής που παράγει τις συντακτικές δομές των προτάσεων της ΝΕΓ συνεπάγεται την παραγωγή-δυνατότητα επεξεργασίας των συνταγματικών αξόνων τους, καθώς και του ορισμού και ποικιλίας των παραδειγματικών αξόνων σε κάθε περίπτωση. Οι συντακτικές κατηγορίες της ΝΕΓ που συνθέτουν τους συνταγματικούς άξονες είναι τα σημεία όπου εντοπίζεται η αλληλεξάρτηση της σύνταξης, της μορφολογίας και της σημασιολογίας και το αντικείμενο της εν γένει μελέτης και έρευνάς μας για αυτόματη επεξεργασία της ΝΕΓ.

Έτσι, οδηγηθήκαμε στη διατύπωση των μητροειδών γραμματικών (μ. γραμματικών) – χαρακτηριστικοί εκθέτες (χ.ε.) [16,17,19], μίας τροποποιημένης εκδοχής των γραμματικών Chomsky. Η τροποποίηση δεν αφορά τους τύπους των γραμματικών οι οποίοι παραμένουν ως έχουν, αλλά τους κανόνες παραγωγής, οι οποίοι είναι κανόνες συντακτικής δομής και όχι φρασεοδομής, και τους ομαδοποιούν σε μήτρες συντακτικών κανόνων, οι οποίες παράγουν συντακτικές δομές συνήθων και εύχρηστων προτάσεων της ΝΕΓ. Δηλαδή, στο προτεινόμενο τροποποιημένο σχήμα $G = (V_{NT}, V_T, P, S)$, P: είναι το πεπερασμένο σύνολο μητρών συντακτικών

κανόνων, και όχι φρασεοδομής, επιπλέον V_T : είναι το πεπερασμένο σύνολο των συντακτικών κατηγοριών της ΝΕΓ και όχι λέξεις της. Οι μ. γραμματικές παράγουν μία ενδιάμεση γλώσσα, τη γλώσσα των συντακτικών κατηγοριών ή γλώσσα βάση της ΝΕΓ, η οποία προσομοιάζει με τη ΝΕΓ ως προς τη συντακτική δομή της. Το πλεονέκτημα είναι ότι είναι απαλλαγμένη από γραμματικούς τύπους και έννοιες και μελετάται εύκολα η συντακτική συμπεριφορά της αφού είναι μία τυπική γλώσσα που εύκολα προγραμματίζεται [7,17,19].

Οι μ. γραμματικές παράγουν την ενδιάμεση γλώσσα των συντακτικών κατηγοριών ή γλώσσα βάση της ΝΕΓ [7,16,17,19]. Με τον τρόπο αυτόν προσεγγίζουμε από πάνω προς τα κάτω τη μελέτη της δομής των σωμάτων κειμένων σε αντίθεση με την ετικετοποίηση που αποτελεί προσέγγιση από κάτω προς τα επάνω [25].

Οι προτάσεις που παράγονται από τις μ. γραμματικές (μη κειμενοποιημένες) είναι ακέραιες και στην κανονική σειρά των (συντακτικών) όρων τους, προς αποφυγή ορισμού διφορούμενων γραμματικών. Οι αποκλίσεις από την κανονική σειρά των όρων καθώς και ο χειρισμός των ελλειπτικών προτάσεων προσεγγίζονται με μετασχηματιστικούς κανόνες.

Σε κάθε συντακτική δομή της ΝΕΓ ή συντακτική δομή βάθους αντιστοιχεί ένας μεγάλος αριθμός ποικίλων αντίστοιχων δομών επιφανείας αποδεκτών προτάσεων της ΝΕΓ. Η αντιστοίχιση αυτή υλοποιείται με τη βοήθεια του Βασικού Νεοελληνικού Πολυλεξικού (ΒΝΠ) και των αλγορίθμων που το συνοδεύουν και που έχουμε εισαγάγει [2,3,4,5,19]. Δηλαδή, η σύνδεση των συνταγματικών αξόνων με τη μορφολογία και τη σημασιολογία στην πραγματικότητα γίνεται με ένα σύστημα Ηλεκτρονικών Υπολογιστικών Λεξικών της ΝΕΓ και των αλγορίθμων που τα συνοδεύουν.

2.3 Το Βασικό Νεοελληνικό Πολυλεξικό και οι Αλγόριθμοι που το συνοδεύουν

Το Βασικό Νεοελληνικό Πολυλεξικό (ΒΝΠ) αποτελείται από 4 επί μέρους λεξικά στα οποία είναι καταχωρημένες ταξινομημένες και κωδικοποιημένες όλες οι γλωσσολογικές πληροφορίες που απαιτούνται. Τα Λεξικά αυτά θα περιγραφούν στις παρακάτω ενότητες.

2.3.1 Το Βασικό Νεοελληνικό Λεξικό (ΒΝΛ)

Το ΒΝΛ περιέχει τις λέξεις, σε αλφαβητική σειρά – περίπου 7.000 λήμματα, μεταξύ των οποίων τουλάχιστον 200 πολύσημες λέξεις – τις οποίες το προτεινόμενο σύστημα

μας αναγνωρίζει και επεξεργάζεται.

Στα υπάρχοντα λήμματα είναι καταχωρημένα αντιπροσωπευτικά δείγματα όλων των κλιτών και άκλιτων μερών του λόγου. Το BNL εύκολα συντηρείται και μπορεί να επεκταθεί στον χειρισμό 60.000 και πλέον λημμάτων.

2.3.2 Το Νεοελληνικό Μορφολογικό Υπολογιστικό Λεξικό (NΜορφΥπΛ)

Το NΜορφΥπΛ [1,2,3,4,5,19].), το οποίο περιέχει τα λήμματα του BNL και ελέγχει όλα τα είδη της μορφολογικής πληροφορίας, μπορεί:

- να παράγει και να αναγνωρίζει: 3000 ουσιαστικά, 1600 επίθετα, 2100 ρήματα και τις αντίστοιχες μετοχές τους, 50 αντωνυμίες και 250 μη κλιτές λέξεις, καθώς και να επεκταθεί στον χειρισμό 60.000 και πλέον λημμάτων. Λαμβάνοντας υπόψη ότι κάθε ουσιαστικό έχει 8 διακριτούς κλιτούς τύπους, κάθε επίθετο έχει 120, κάθε ρήμα έχει 76 κλπ., το σύστημά μας αναγνωρίζει συνολικά 1.000.000 κλιτούς τύπους κατά προσέγγιση και με ακρίβεια 100%,
- να χειριστεί περιφραστικούς αλλά και πολλαπλούς τύπους και να ελέγχει μορφολογικές αμφισημίες, έτσι ώστε να αποδίδει με σαφήνεια και ακρίβεια τις οποιοσδήποτε πληροφορίες και
- να είναι εύκολα επεκτάσιμο και ευκολοσυντήρητο.

Παραθέτουμε κάποιες οθόνες ενδεικτικές των δυνατοτήτων του NΜορφΥπΛ.

Οθόνη 1: Μορφολογική αναγνώριση πολλαπλών και διαφορετικών τύπων λέξης, και απαιτούμενος χρόνος αναγνώρισής τους

The image displays three screenshots of the NΜορφΥπΛ software interface, illustrating the morphological recognition process for the word 'άβυσσοι' (abysses) in different grammatical contexts. Each screenshot shows a main window with a search input field and a 'Δώστε λέξη:' (Enter word) button. Below the input field, there is a description of the word's grammatical category and a 'Αναγνώριση της Λέξης' (Recognize word) button. To the right, there are several smaller windows or panels, including 'Λογικό Υπολογιστικό Λεξικό Βασικών Εννοιών' (Logical Computational Dictionary of Basic Concepts), 'Νεοελληνικό Συντακτικό Υπολογιστικό Λεξικό' (Neo-Greek Syntactic Computational Dictionary), and 'Νεοελληνικό Γραμματικό Υπολογιστικό Λεξικό' (Neo-Greek Grammatical Computational Dictionary). The screenshots show the software's response to the input word, including the recognition of the word's form and its grammatical properties.

Οθόνη 2: Μορφολογική αναγνώριση τύπου λέξης που ανήκει σε διαφορετικά μέρη του λόγου και απαιτούμενος χρόνος αναγνώρισής του

Οθόνη 3: Μορφολογική αναγνώριση περιφραστικού τύπου λέξης

Οθόνη 4: Μορφολογική αναγνώριση (α) πολλαπλών, περιφραστικών τύπων του υπερθετικού του επιθέτου καλός

Οθόνη 5: Μορφολογική αναγνώριση (β) πολλαπλών, περιφραστικών τύπων του υπερθετικού του επιθέτου καλός

Δώστε λέξη : των καλύτερων

Βασικό Νεοελληνικό Υπολογιστικό Λεξικό

Πτώση γενική, αριθμοί πληθυντικού, γένους αρσενικού βαθμοί περιφραστικού υπερθετικού του κοινού επιθέτου: καλός

Υπάρχουν περισσότερα του ενός αποτελέσματα

Αναγνώριση της λέξης

Λογικό Υπολογιστικό Λεξικό Βασικών Εννοιών

Συνώνυμα - Αντίθετα Ιδιωματισμοί

Δεσπόζοντες Σημασιολογικοί Συνδυασμοί Λέξεων

Νεοελληνικό Συντακτικό Υπολογιστικό Λεξικό

Σχηματισμός Τύπου Λέξης ανάλογα με το Συντακτικό της Ρόλο

Κανόνες

Κλίση Λέξης

Κανόνες

Εμφάνιση Γλωσσαρίου

Εισαγωγή Νέας Λέξης

Εξόδος από το ΕΝΠ

Οθόνη 6: Μορφολογική αναγνώριση των πολλαπλών, περιφραστικών τύπων του υπερθετικού του επιθέτου καλός της οθόνης 4, τα γένη και η κλίση του αρσενικού των μονολεκτικών τύπων του υπερθετικού

Βασικό Νεοελληνικό Υπολογιστικό Πολυλεξικό (BNΠ)

Δώστε λέξη : των καλύτερων

Βασικό Νεοελληνικό Υπολογιστικό Λεξικό

Πτώση γενική, αριθμοί πληθυντικού, γένους αρσενικού βαθμοί περιφραστικού υπερθετικού του κοινού επιθέτου: καλός

Υπάρχουν περισσότερα του ενός αποτελέσματα

Αναγνώριση της λέξης

Λογικό Υπολογιστικό Λεξικό Βασικών Εννοιών

Συνώνυμα - Αντίθετα Ιδιωματισμοί

Δεσπόζοντες Σημασιολογικοί Συνδυασμοί Λέξεων

Νεοελληνικό Συντακτικό Υπολογιστικό Λεξικό

Σχηματισμός Τύπου Λέξης ανάλογα με το Συντακτικό της Ρόλο

Κανόνες

Κλίση Λέξης

Κανόνες

Εμφάνιση Γλωσσαρίου

Εισαγωγή Νέας Λέξης

Εξόδος από το ΕΝΠ

ΝΓΑ Νεοελληνικό Γραμματικό Λεξικό

Επίθετο: καλός

Γλωσσόφυο:

αβαθής
αβύρρανος
αναβούθης
αγγελικός
αγγελιαλάτης
άγιος
άγιος
αγύρτης
αδικημένος

Γένη

Κλίση

Έξοδος

Μονολεκτικός (απόλυτος) τύπος παραθετικού

Περιφραστικός (σχετικός) τύπος παραθετικού

Βαθμός

Γένος

Αρσενικό

Θηλυκό

Ουδέτερο

Υπερθετικός

Γένη του Επίθετου στον Επιλεγμένο Τύπο και Βαθμό

Αρσενικό: άριστος κἀλλιστος

Θηλυκό: άριστη κἀλλιστη

Ουδέτερο: άριστο κἀλλιστο

Κλίση του Επίθετου στον Επιλεγμένο Τύπο, Βαθμό και Γένος

Ενικός Αριθμός

Πληθυντικός Αριθμός

άριστος κἀλλιστος

άριστοι κἀλλισταί

άριστου κἀλλιστου

άριστων κἀλλιστων

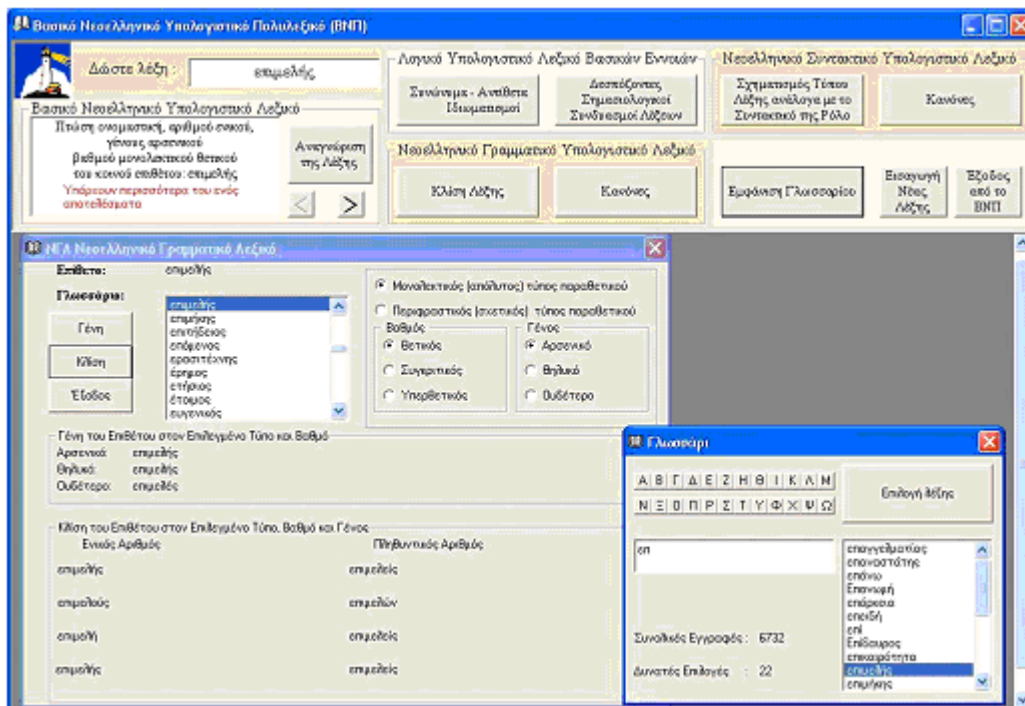
άριστο κἀλλιστο

άριστους κἀλλιστους

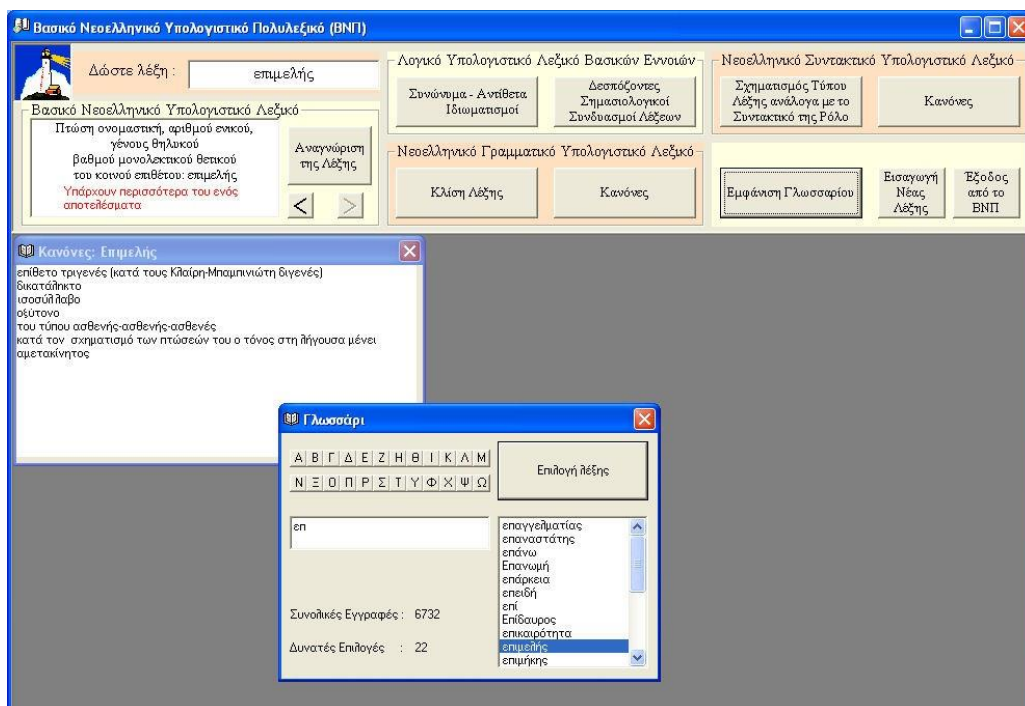
άριστε κἀλλιστε

άριστα κἀλλιστα

Οθόνη 7: Εύρεση και επιλογή του επιθέτου *επιμελής* από το γλωσσάριο, αναγνώριση και κλίση του θετικού βαθμού του αρσενικού



Οθόνη 8: Εύρεση και επιλογή του επιθέτου *επιμελής* από το γλωσσάριο, αναγνώριση και ανάκληση των μορφολογικών κανόνων που διέπουν το επίθετο



Οθόνη 9: Γένη μονολεκτικού υπερθετικού και κλίση του αρσενικού μονολεκτικού υπερθετικού του επιθέτου *επιμελής*

Επίθετο: επιμελής

Γλωσσάριο: επιμελής, επιμήκης, επιτήδειος, επόμενος, ερασιτέχνης, έρμος, ετήσιος, έτοιμος, ευγενικός

Γένη: Γένη, **Κλίση**, Έξοδος

Μονολεκτικός (απόλυτος) τύπος παραθετικού

Βαθμός: Θετικός, Συγκριτικός, Υπερθετικός

Γένος: Αρσενικό, Θηλυκό, Ουδέτερο

Γένη του Επιθέτου στον Επιλεγμένο Τύπο και Βαθμό

Αρσενικά: επιμελέστατος
Θηλυκά: επιμελέστατη
Ουδέτερα: επιμελέστατο

Κλίση του Επιθέτου στον Επιλεγμένο Τύπο, Βαθμό και Γένος

Ενικός Αριθμός	Πληθυντικός Αριθμός
επιμελέστατος	επιμελέστατοι
επιμελέστατου	επιμελέστατων
επιμελέστατο	επιμελέστατους
επιμελέστατε	επιμελέστατοι

Οθόνη 10: Γένη περιφραστικού υπερθετικού και κλίση του θηλυκού περιφραστικού υπερθετικού του επιθέτου *επιμελής*

Επίθετο: επιμελής

Γλωσσάριο: επιμελής, επιμήκης, επιτήδειος, επόμενος, ερασιτέχνης, έρμος, ετήσιος, έτοιμος, ευγενικός

Γένη: Γένη, **Κλίση**, Έξοδος

Περιφραστικός (σχετικός) τύπος παραθετικού

Βαθμός: Θετικός, Συγκριτικός, Υπερθετικός

Γένος: Αρσενικό, Θηλυκό, Ουδέτερο

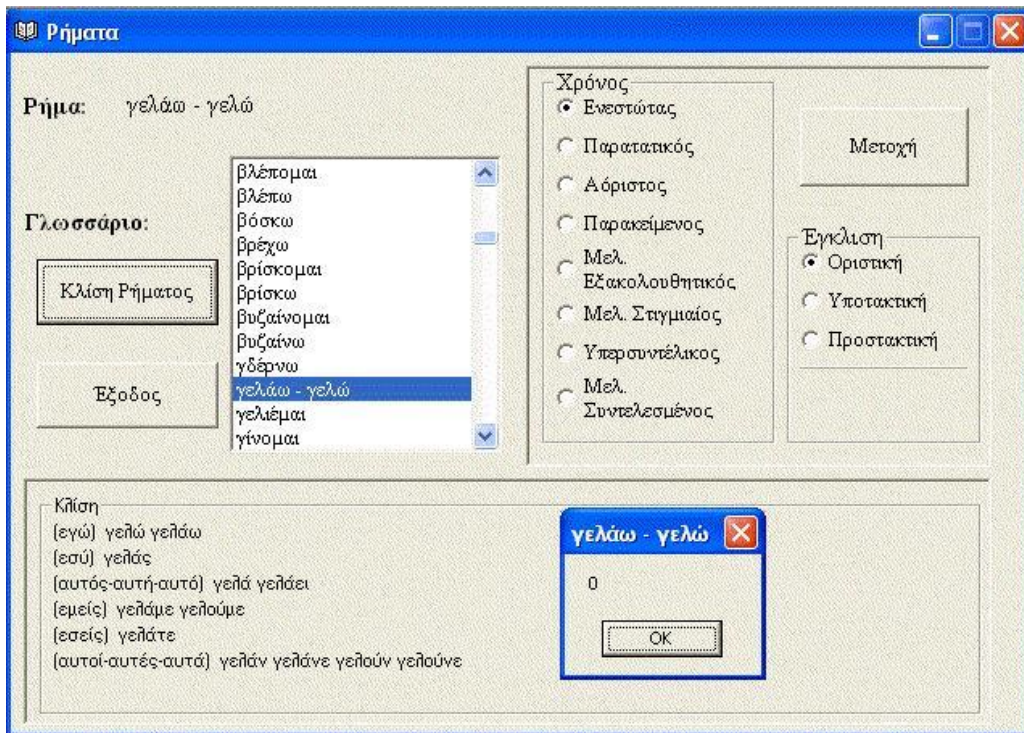
Γένη του Επιθέτου στον Επιλεγμένο Τύπο και Βαθμό

Αρσενικά: ο πιο επιμελής
Θηλυκά: η πιο επιμελής
Ουδέτερα: το πιο επιμελής

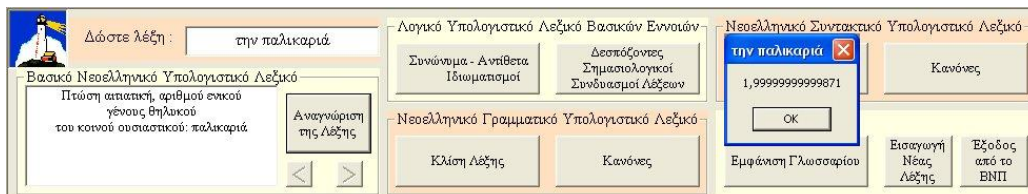
Κλίση του Επιθέτου στον Επιλεγμένο Τύπο, Βαθμό και Γένος

Ενικός Αριθμός	Πληθυντικός Αριθμός
η πιο επιμελής	αι πιο επιμελείς
της πιο επιμελούς	των πιο επιμελών
την πιο επιμελή	τις πιο επιμελείς
· πιο επιμελής	· πιο επιμελείς

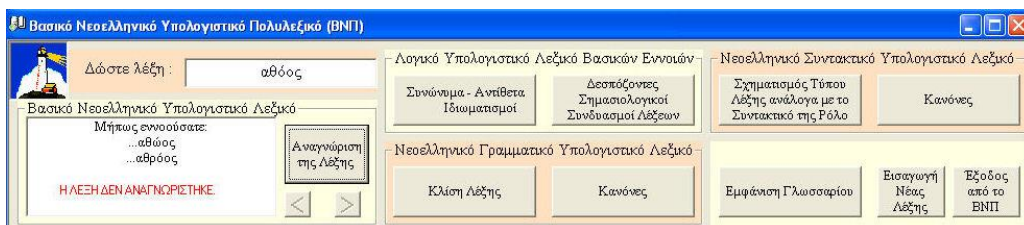
Οθόνη 11: Κλίση της οριστικής του ενεστώτα του ρήματος *γελάω – γελώ* και απαιτούμενος χρόνος κλίσης



Οθόνη 12: Ακριβής μορφολογική αναγνώριση τύπου στηριζόμενη στη συμφωνία πτώσης άρθρου ↔ ουσιαστικού και απαιτούμενος χρόνος αναγνώρισής του



Οθόνη 13: Πιθανές σωστές εκδοχές ανορθόγραφης λέξης



Ηλεκτρονικά λεξικά κυκλοφορούν πολλά. Όμως ηλεκτρονικά μορφολογικά λεξικά που να είναι συγχρόνως και υπολογιστικά δεν υπάρχουν. Τα ηλεκτρονικά λεξικά του χώρου αυτού είναι κυρίως ερμηνευτικά, ετυμολογικά, εγκυκλοπαιδικά κ.ά., τα οποία όπως είναι προφανές έχουν άλλο σχεδιασμό, αντικείμενο και στόχους, αλλά και αν μπορούν να δώσουν μορφολογικές πληροφορίες για κλίσεις ή μορφολογικούς τύπους

λέξεων, αυτοί είναι αποθηκευμένοι, και απλά ανακαλούνται· δεν παράγονται. Τα ηλεκτρονικά υπολογιστικά μορφολογικά λεξικά αναπτύσσονται κυρίως από ερευνητικές ομάδες ως εργαλεία για την αυτόματη επεξεργασία της ΝΕΓ. Εξαιτίας του γεγονότος ότι η νέα ελληνική γλώσσα ανήκει στις υψηλού βαθμού κλιτές γλώσσες, έχουν γίνει πολλές προσπάθειες προς την κατεύθυνση αυτή, αφού η κατασκευή ηλεκτρονικών μορφολογικών επεξεργαστών αποτελεί το πρώτο βήμα οποιασδήποτε προσπάθειας αυτόματης επεξεργασίας της.

Η πλειονότητα αυτών των εργαλείων – μορφολογικοί επεξεργαστές, ετικετοποιητές, ανιχνευτές θεμάτων (θέμα κατάληξη) – χρησιμοποιούν πρακτικές καθιερωμένες παγκόσμια όπως : το μορφολογικό μοντέλο των δύο επιπέδων του Koskenniemi [39], το άμεσο ακυκλικό γράφημα λέξης [40,41], οι τεχνικές μηχανικής εκμάθησης [35,37,38], οι στατιστικές μέθοδοι [42], δοκιμάζοντας την αποτελεσματικότητά τους σε σχέση με τη μορφολογία της ΝΕΓ.

Παρόλη την ποικιλία των θεωρητικών μοντέλων και υλοποιήσεων των μεθόδων, όλα τα συστήματα έχουν ένα κοινό χαρακτηριστικό : βασίζονται σε προσεγγίσεις οι οποίες αναπτύχθηκαν για να αντιμετωπίσουν τη μορφολογία φτωχότερων κλιτικά γλωσσών. Γεγονός που προδικάζει τις δυσκολίες που αναφέρονται στην προσπάθεια χειρισμού δια μέσου αυτών της μορφολογίας της νέας ελληνικής γλώσσας.

Αντίθετα το μοντέλο μορφολογικής ανάλυσης και σύνθεσης του ΒΝΠ σχεδιάστηκε εξ αρχής με βάση τις ιδιαιτερότητες της ελληνικής γλώσσας και μάλιστα το πλήθος και την ποικιλία των μορφολογικών τύπων των λέξεων. Έτσι έχει αποφευχθεί ένα μοντέλο που επικεντρώνεται στη λέξη, και το οποίο αντιστοιχεί και χειρίζεται την κάθε λέξη ξεχωριστά. Το μοντέλο αυτό στην αρχή θα είχε τη δυνατότητα επεξεργασίας περιορισμένου αριθμού μερών του λόγου, σταδιακά όμως θα επεκτεινόταν με όλο και περισσότερους κανόνες για να μπορεί να αντιμετωπίσει όσο το δυνατόν περισσότερες εξαιρέσεις. Οι συνέπειες του χειρισμού επιβαρυνμένων συστημάτων είναι πολλές φορές απρόβλεπτες, αφού οδηγούν στη διακοπή λειτουργίας του (κρέμασμα) ή σε άλλες αδιέξοδες καταστάσεις. Ομοίως αποφύγαμε και τις στατιστικές μεθόδους, αφού το μόνο που θα μπορούσαν να δώσουν ήταν προσεγγιστικές μέθοδοι και προσεγγιστικά αποτελέσματα. Αντίθετα προτείνουμε ένα μοντέλο όσο το δυνατόν γενικότερο, το οποίο αντιμετωπίζει συνολικά τη δομολειτουργία της γλώσσας, έτσι ώστε οι μηχανισμοί αναγνώρισης και παραγωγής μορφολογικών τύπων (μονολεκτικών ή περιφραστικών, 1^{ov} , 2^{ov} ή 3^{ov}) να λειτουργούν ανεξάρτητα από την πολυπλοκότητα των δεδομένων που χειρίζονται κάθε φορά. Τα

δεδομένα, κωδικοποιημένα, είναι αρχειοθετημένα και ταξινομημένα στη βάση δεδομένων του συστήματος. Το αποτέλεσμα αυτής της καινοτομίας οδήγησε στην 100% αναγνώριση όλων των μορφολογικών τύπων της νέας ελληνικής γλώσσας και κατ' επέκταση στην παραγωγή τους με απλούς, αποτελεσματικούς και ταχείς αλγορίθμους. Οι χρόνοι αναγνώρισης είναι πολύ μικροί, ακόμη και αυτοί που αντιστοιχούν σε αποτελέσματα περισσότερα του ενός : βλέπε οθόνες 1, 2 και 12. Ο χρόνος παραγωγής είναι μηδενικός : βλέπε οθόνη 11.

2.3.3 Το Νεοελληνικό Σημασιολογικό Υπολογιστικό Λεξικό (ΝΣημΥπΛ)

2.3.3.1 Εισαγωγή

Το ΝΣημΥπΛ [1,2,3,6,12,19] αποτελείται από το Λογικό Υπολογιστικό Λεξικό των Βασικών Εννοιών (ΛΥπΛΒΕ) και το Υπολογιστικό Λεξικό των Συμφραζομένων της ΝΕΓ (ΛΥπΛΣΝΓ).

Αναλυτικότερα, η κατασκευή του ΝΣημΥπΛ και των λεξικών που το αποτελούν άπτεται της αντιμετώπισης του προβλήματος της σημασιολογικής διάστασης για την επεξεργασία της ΝΕΓ αποτέλεσε το πιο ιδιαίτερο κομμάτι της έρευνας για την κατασκευή του ΒΝΠ [1,2,5,6,10,11,12,19], διότι δεν υπήρχαν καταγεγραμμένοι σημασιολογικοί κανόνες ώστε να μελετηθούν και να αποτελέσουν τη βάση για τη μοντελοποίηση αυτής της γλωσσολογικής διάστασης. Δηλαδή με πεπερασμένου πλήθους σχέσεις να μπορεί να περιγραφεί το σύνολο των δεσποζόντων σημασιολογικών συνδυασμών των εννοιών των λέξεων ως προς τον συντακτικό τους ρόλο για την άμεση σύνδεσή τους με τους παραδειγματικούς άξονες των συνταγματικών αξόνων κάθε φορά. Με άλλα λόγια, εισάγονται πεπερασμένου πλήθους σημασιολογικοί κανόνες οι οποίοι περιγράφονται με τη βοήθεια των σημασιολογικών κατηγοριών (σ.κ.) ώστε να συνδέουν τη σημασιολογία με τη σύνταξη, δηλαδή να περιγράφουν τους αποδεκτούς δεσπόζοντες συνδυασμούς των εννοιών των λέξεων ανάλογα με τον συντακτικό τους ρόλο κάθε φορά. Οι κανόνες αυτοί ορίστηκαν ως σημασιοσυντακτικοί ώστε να διαχωρίζονται τους από τους σημασιολογικούς κανόνες που περιγράφουν κανόνες-μηχανισμούς που εντοπίζουν τις λέξεις που συνδέονται μεταξύ τους με μια σημασιολογική σχέση, για παράδειγμα τα συνώνυμα ή τα αντώνυμα ενός ρήματος ή ενός επιθέτου, τα συνώνυμα ενός ουσιαστικού, κλπ. Επιπλέον σημειώνουμε την ανάγκη οι λέξεις που αναγνωρίζει το σύστημά μας, δηλαδή «ο Θησαυρός» (*Λεξικό της Κοινής Νεοελληνικής, Λεξικό της*

Νέας Ελληνικής Γλώσσας), να αναταξινομηθούν κωδικοποιημένες κατά αυστηρή αλληλουχία των εννοιών τους με προδιαγραφές «Λογικού Λεξικού».

2.3.3.2 Η Σημασιολογική Βάση

Έτσι για την ταξινόμηση αυτή ακολουθήθηκε εκείνη του Αντιλεξικού [44] διακρίνοντας αφενός επίπεδα κωδικοποίησης και εμπλουτίζοντας το όπου χρειαζόταν. Σημειώνουμε ότι η ταξινόμηση αυτή αποτελεί ένα μέρος της σημασιολογικής βάσης του συστήματός μας [1,4,5,6,12,19].

Το πρώτο επίπεδο αυτής της ταξινόμησης διαιρεί τον κόσμο σε 21 Κατηγορίες Ευρύτερων Ενοτήτων Εννοιών/Κεφαλαίων (ΚΕΕΕ/Κ), η κάθε μία από τις οποίες αναφέρεται και σε μία διαφορετική διάσταση της ανθρώπινης πραγματικότητας (βλπ. Πίνακα 1) [1,2,3,6,12,19].

Κάθε μία από τις ΚΕΕΕ/Κ υποδιαιρείται σε Έννοιες/Κεφάλαια (Ε/Κ) (βλπ. Πίνακα 2). Κάθε Ε/Κ αποτελεί μία διακριτή έννοια της κάθε ΚΕΕΕ/Κ στην οποία ανήκει. Μία Ε/Κ μπορεί να είναι αντίθετη ή συμπληρωματική με μία άλλη Ε/Κ της ίδιας ΚΕΕΕ/Κ. Το πλήθος των Ε/Κ είναι 1500 (βλπ. Πίνακα 3).

Πίνακας 1: Το σύνολο των 21 Κατηγοριών Ευρύτερων Ενοτήτων Εννοιών / Κεφαλαίων το οποίο συνθέτει το πρώτο επίπεδο της σημασιολογικής μας βάσης

1. ΓΕΝΙΚΗ		2. ΕΙΔΙΚΗ	
1. ΑΦΗΡΗΜΕΝΗ	2. ΣΥΓΚΕΚΡΙΜΕΝΗ	3. ΥΛΙΚΗ	4. ΑΨΛΗ
1. ΥΠΑΡΞΗ 2. ΣΧΕΣΗ 3. ΕΝΟΤΗΤΑ 4. ΤΑΞΗ 5. ΠΟΣΟΤΗΤΑ 6. ΑΡΙΘΜΟΣ 7. ΧΡΟΝΟΣ	8. ΧΩΡΟΣ 9. ΔΙΑΣΤΑΣΗ 10. ΣΧΗΜΑ 11. ΕΝΕΡΓΕΙΑ 12. ΚΙΝΗΣΗ	13. ΑΝΟΡΓΑΝΗ ΥΛΗ 14. ΟΡΓΑΝΙΚΗ ΥΛΗ	15. ΝΟΥΣ 16. ΒΟΥΛΗΣΗ 17. ΔΡΑΣΗ 18. ΑΞΙΕΣ 19. ΣΥΝΑΙΣΘΗΜΑ 20. ΗΘΟΣ 21. ΘΕΟΣ

Πίνακας 2: Η Κατηγορία Ευρύτερη Ενότητα Εννοιών / Κεφαλαίων «Υπαρξη» περιλαμβάνει 7 Έννοιες/Κεφάλαια

1. Υπαρξη	2. Ανυπαρξία
3. Κατάσταση	
4. Περίσταση	
5. Εσωτερικός Κόσμος	6. Εξωτερικός Κόσμος
7. Εγώ	

Οι Ε/Κ αποτελούν το δεύτερο επίπεδο αυτής της ενότητας της σημασιολογικής βάσης του συστήματός μας. Κάθε μία από τις 1.500 Ε/Κ περιλαμβάνει λέξεις –

ταξινομημένες βάσει του μέρους του λόγου όπου ανήκουν, ιδιωτισμούς και κοινές φράσεις, οι οποίες εννοιολογικά ανήκουν στην ίδια σημασιολογική περιοχή. Τα μέρη του λόγου, τα παραγγέλματα, οι φράσεις και οι ιδιωτισμοί είναι το τρίτο επίπεδο αυτής της ενότητας του συστήματός μας, το οποίο ονομάζουμε Παραγράφους.

Έτσι κάθε Ε/Κ έχει τόσες Παραγράφους όσες είναι οι κλιτές και οι άκλιτες λέξεις, τα παραγγέλματα, οι φράσεις και οι ιδιωτισμοί που συναντώνται στη γλώσσα με την ίδια γενική έννοια.

Σημειώνουμε ότι με τον όρο γενική έννοια αναφερόμαστε σε λέξεις που μπορεί να είναι συνώνυμες ή αντώνυμες με τη γενικότερη ή ειδικότερη έννοια του όρου (βλπ. Πίνακα 3).

Τέλος, σε κάθε Παράγραφο σχηματίζονται ομάδες σημασιολογικά συγγενών λέξεων βάσει της σχέσης των συνωνύμου ή αντωνύμου. Αυτές οι ομάδες ονομάζονται Επιμέρους Έννοιες (ΕΕ) και συνθέτουν το τέταρτο επίπεδο αυτής της ενότητας της σημασιολογικής βάσης του συστήματός μας. Αντιμετωπίσαμε ακόμα και το δυσεπίλυτο πρόβλημα της πολυσημίας (βλπ Πίνακες 1 & 2).

Πίνακας 3: Η Έννοια/Κεφάλαιο «Υπαρξη» περιλαμβάνει μόνο 5 Παραγράφους, όπου α/απ είναι ο αύξοντας αριθμός Παράγραφοι

α/απ	Παράγραφοι	Λέξεις του Βασικού Νεοελληνικού Λεξικού (ΒΝΛ)
1	κ1_ουσιαστικά	ύπαρξη, υπόσταση, το είναι, κ.λ.π.
2	κ3_αμετάβατο_ρήμα	υπάρχω, είμαι, υφίσταμαι, κ.λ.π.
3	κ4_μεταβατικό_ρήμα	δίδω υπόσταση, διατηρώ, κ.λ.π..
4	κ5_τριτοπρόσωπο_ρήμα	υπάρχει, είναι, βρίσκεται, κ.λ.π.
5	κ6_επίθετο	υπαρκτός, υπάρχων, υφιστάμενος, κ.λ.π.

Οι κωδικοποιήσεις που προέκυψαν (βλπ Πίνακες 5 & 6) απετέλεσαν το περιεχόμενο του ΛΥπΛΒΕ. Αντίστοιχα το περιεχόμενο του ΛΥπΛΣΝΓ είναι η καταχώρηση για κάθε λέξη του ΒΝΛ της αντίστοιχης σημασιολογικής κατηγορίας, βάσει της οποίας, ανάλογα με το μέρος του λόγου που ανήκει η λέξη μπορεί να συνδυαστεί με άλλες λέξεις και με συγκεκριμένους συντακτικούς ρόλους κάθε φορά (βλπ Πίνακες 6 & 7). Οι επιτρεπτοί αυτοί συνδυασμοί συνθέτουν τους αντίστοιχους σημασιολογικούς κανόνες. Ο ορισμός και η ταξινόμηση των σ.κ. ορίζουν την εναλλακτική οντολογία όπως ορίστηκε [11].

Πίνακας 4: Οι 11 ομάδες σύμφωνα με τις Επιμέρους Έννοιες των λέξεων που αποτελούν την Παράγραφο Ουσιαστικό της Ε/Κ «Υπαρξη». Όπου α/αεε είναι ο αύξοντας αριθμός της Επιμέρους Έννοιας

α/αεε	Λέξεις/ουσιαστικά του Βασικού Νεοελληνικού Λεξικού (ΒΝΛ)
1.	ύπαρξη, υπόσταση, (το) είναι, οντότητα,
2.	αυθυπαρξία, αυτοτέλεια, αυτοζωή, αυτοζωία,
3.	Προϋπαρξη, προϋπόσταση,
4.	Συνύπαρξη,
5.	Ενούπαρξη,
6.	διατήρηση, παραμονή,
7.	διάσωση, περίσωση, επιβίωση,
8.	όν, οντότητα, πλάσμα,
9.	άτομο, πρόσωπο, ψυχή, κανείς, κανένα, τις, ένας, κάποιος,
10.	αντικείμενο, πράγμα, κάτι, τίποτε,
11.	οντολογία.

Πίνακας 5: Δείγμα σημασιολογικής κωδικοποίησης λέξης του ΒΝΠ

α/α	Λέξη	ΚΕΕΕ/Κ→	Ε/Κ→	Παράγραφος→	Επιμέρους Έννοια
α/α.	άνθρωπος	ΟΡΓΑΝΙΚΗ ΥΛΗ	Άνθρωπος	ουσιαστικό	α/αλ
		14	0543	κ1.1	α/αλ
		14 . 0543 . κ1.1 . α/αλ			

Πίνακας 6: Δείγμα κωδικοποίησης των πολύσημων λέξεων του ΒΝΠ, οι οποίες ανήκουν σε τόσες διαφορετικές Ε/Κ όσες είναι και οι αντίστοιχες έννοιές τους

ΒΝΛ	ΛΥπΛΒΕ
	ΚΕΕΕ/Κ . Ε/Κ . Παράγραφος . α/αλ
α/α. μελετάω-ώ	ΝΟΥΣ . Μάθηση . ρήμα . α/αλ 15 . 0828 . κ2. α/αλ
	ΝΟΥΣ . Εξέταση . ρήμα . α/αλ 15 . 0845 . κ3. α/αλ
	ΝΟΥΣ . Μνεία . ρήμα . α/αλ 15 . 0918 . κ3. α/αλ
	ΒΟΥΛΗΣΗ . Σκοπός . ρήμα . α/αλ 16 . 1010 . κ3. α/αλ

Πίνακας 7: Δείγμα κωδικοποίησης των δεδομένων του ΝΣημΥπΛ, ειδικότερα των στοιχείων του ΛΥπΛΣΝΓ, με παραδείγματα τις σ.κ. <διδασκαλος>, <πολεμιστης> και <συγγραφεις> να είναι υπώνυμα της σ.κ. <ανθρωπος> αλλά να ανήκουν σε διαφορετικά υποδένδρα. Σημειώνουμε ότι η σ.κ. <ανθρωπος> είναι υπώνυμο της σ.κ. <ανθρωπιν_ον>

BNA	ΝΣημΥπΛ	
	ΛΥπΛΒΕ	ΛΥπΛΣΝΓ
	ΚΕΕΕ/Κ . Ε/Κ . Παράγραφος . α/αλ	
α/α. άνθρωπος	ΟΡΓΑΝΙΚΗ ΥΛΗ. Άνθρωπος . ουσιαστικό . α/αλ 14 . 0543 . κ1.1 . α/αλ	<ανθρωπος>
α/α. δάσκαλος	ΝΟΥΣ . Διδάσκαλος . ουσιαστικό . α/αλ 15 . 0831 . κ1.2 . α/αλ	<διδασκαλος>
α/α. στρατιώτης	ΔΡΑΣΗ . Πολεμιστής . ουσιαστικό . α/αλ 17 . 1111 . κ1.3 . α/αλ	<πολεμιστης>
α/α. συγγραφέας	ΝΟΥΣ . Συγγραφέας . ουσιαστικό . α/αλ 15. 0956 . κ1.1 . α/αλ	<συγγραφεις>

2.3.2.3 Η κατασκευή των Σημασιολογικών και Σημασιοσυντακτικών κανόνων

Το ρήμα περιγράφει την ενέργεια στην πρόταση και καθορίζει τη συντακτική δομή και τις λέξεις της πρότασης [11,12]. Δηλαδή, ο συνταγματικός άξονας που εισάγει ορίζει κάθε φορά τους αντίστοιχους σε κάθε συντακτική κατηγορία παραδειγματικούς άξονες. Σημειώνεται ότι με τις μ. γραμματικές και τους χ.ε. και με τις σ.μ. γραμματικές και τους χ.ε., παράγεται και τεχνολογείται αντίστοιχα το σύνολο των εύχρηστων συντακτικών δομών των ακέραιων και κύριων συντακτικών δομών (βάθους) της ΝΕΓ.

Έτσι ορίστηκαν, τυποποιώντας, κωδικοποιώντας και ταξινομώντας, τα κοινά σημασιολογικά χαρακτηριστικά που υπαγορεύονται από τα ρήματα ώστε οι λέξεις, κυρίως ουσιαστικά, που αναγνωρίζει το ΒΝΠ, που συναρμόζονται και ανήκουν στους παραδειγματικούς άξονες των αντίστοιχων κύριων συντακτικών κατηγοριών του, να ομαδοποιούνται για περαιτέρω μηχανική αξιοποίηση. Οι κωδικοποιήσεις αυτές έχουν μια γενική και προτυποποιημένη μορφή ώστε να είναι αποτελεσματικά αξιοποιήσιμες στον αυτόματο χειρισμό τους. Αυτά τα κοινά σημασιολογικά χαρακτηριστικά ορίζουν τις σ.κ., όπως έχουμε ήδη προαναφέρει. Οι σ.κ. ομαδοποιούν και κωδικοποιούν περαιτέρω τις λέξεις του Πολυλεξικού, δημιουργώντας το Υπολογιστικό Λεξικό των Συμφραζομένων της ΝΕΓ (ΛΥπΛΣΝΓ). Με τις σ.κ. συνθέτονται οι πεπερασμένοι πλήθους σημασιοσυντακτικοί κανόνες, με τους οποίους παράγονται και αναγνωρίζονται οι δεσπύζοντες σημασιολογικοί συνδυασμοί των λέξεων ως προς τους συντακτικούς ρόλους τους κάθε φορά [11,12].

Για να επιτευχθεί η γενική και προτυποποιημένη μορφή των σ.κ. μελετήθηκαν

παράλληλα: 1. οι δομές των συνταγματικών αξόνων σε ένα ικανό δείγμα ρημάτων, μεταβατικών, αμετάβατων, μεταβατικών και αμετάβατων, τριτοπρόσωπων, βοηθητικών, πολύσημων και σύνθετων με πρώτο συνθετικό προθήματα, των συνωνύμων, παρασυνωνύμων και των αντωνύμων τους, που δημιουργούσαν αντίστοιχα απλές και επαυξημένες προτασιακές δομές και 2. τα βασικά σημασιολογικά χαρακτηριστικά μεγάλου δείγματος λέξεων, οι οποίες ανήκουν κάθε φορά στον ίδιο παραδειγματικό άξονα, όχι απαραίτητα συνωνύμων ή αντωνύμων, που υπαγορεύεται από το δείγμα των υπό μελέτη ρημάτων. Συγκεκριμένα, σε πρώτη φάση, μελετήσαμε για να ομαδοποιήσουμε - κατηγοριοποιήσουμε τα βασικά σημασιολογικά χαρακτηριστικά επιλεγμένων κάθε φορά ουσιαστικών ώστε αυτά να μπορούν να λειτουργήσουν ως υποκείμενα ή ως αντικείμενα άμεσα ή έμμεσα του υπό μελέτη δείγματος ρημάτων που καλύπτουν μεγάλο εύρος αντιπροσωπευτικών δεσποζόντων σημασιολογικών συνδυασμών λέξεων. Έτσι, πέρα από τη σημασιολογική ταξινόμηση των λέξεων βάσει της δομής του Αντιλεξικού, επιτεύχθηκε: 1. τη σημασιολογική κατηγοριοποίηση των λέξεων για τη γενική και προτυποποιημένη οργάνωση των σημασιολογικών σχέσεων μέσα σε προτασιακές δομές και κατά συνέπεια, 2. ο ολοκληρωμένος ορισμός της σημασιολογικής βάσης του προτεινόμενου συστήματός [1,6,7,19,26,45].

Εξ ορισμού οι σ.κ. ομαδοποιούν λέξεις, ονόματα ουσιαστικά, τα οποία μπορεί να ανήκουν στην ίδια παράγραφο, ακόμα και διαφορετικών Ε/Κ, με κοινά σημασιολογικά χαρακτηριστικά και μπορούν να ανήκουν στον ίδιο παραδειγματικό άξονα, ενός συνταγματικού άξονα ρήματος και μπορούν να εναλλάσσονται αποδεκτά, δηλαδή να εναλλάσσονται ως υποκείμενα ή αντικείμενα, άμεσα ή έμμεσα, στο ίδιο ρήμα, σε συνώνυμά του ή ακόμα και σε αντώνυμά του ή πολλές φορές και σε παρασυνώνυμά τους και συνθέτων τους με ανάλογες σημασίες. Κάθε σ.κ. συμβολίζεται με ελληνικές άτονες και με πεζά γράμματα, μέσα σε ένα ζευγάρι γωνιασμένων αγκυλών <...> λέξεις ή σύζευξη λέξεων που περιγράφουν τα κοινά σημασιολογικά χαρακτηριστικά των λέξεων που ομαδοποιούν [6,11,19] (βλπ Πίνακες: 7, 9, 10 & 11).

Παρατηρήσεις που αφορούν σχέσεις μεταξύ των σ.κ. και έχουν άμεση επίπτωση στην αποτελεσματικότερη αυτόματη αξιοποίηση των σημασιολογικών κανόνων: 1. τα σημασιολογικά χαρακτηριστικά κάποιων ομάδων λέξεων μπορεί να μην είναι ξένα μεταξύ τους αλλά να σχετίζονται με κριτήρια υπερωνύμων και υπωνύμων. Δηλαδή, οι αντίστοιχες σ.κ. τους σχετίζονται ανάλογα και ανήκουν στην ίδια διαδρομή της ίδιας δενδροειδούς δομής

σημασιολογικών κατηγοριών (δ.δ.σ.κ.). Έτσι επεκτείνεται ή περιορίζεται το εύρος των σημασιών και το πλήθος των λέξεων στους παραδειγματικούς άξονες ανάλογα με την κάθε περίπτωση, 2. τα σημασιολογικά χαρακτηριστικά κάποιων ομάδων λέξεων μπορεί να είναι ξένα μεταξύ τους, δηλαδή να μην σχετίζονται με κριτήρια υπερωνύμων και υπωνύμων. Με άλλα λόγια, οι αντίστοιχες σ.κ. τους ή ανήκουν σε διαφορετικά υποδένδρα σ.κ. του ίδιου δένδρου σ.κ. ή ανήκουν σε διαφορετικά δένδρα σ.κ., 3. ορίζουμε τη σύνθετη σημασιολογική κατηγορία (σ.σ.κ.) επεκτείνοντας την έννοια της σ.κ., ως την ομαδοποίηση σ.κ., υπωνύμων σ.κ., που ανήκουν στο ίδιο υποδένδρο σ.κ. ή διαφορετικά υποδένδρα σ.κ. με επιπλέον κοινό σημασιολογικό χαρακτηριστικό και 4. ορίζουμε ως κόμβους-ρίζες υποδένδρων σ.κ. πρώτου επιπέδου τις διαφορετικές ιδιότητες που μπορεί να έχει η έννοια μιας λέξης ή η αντίστοιχη χαρακτηριστική σημασιολογική ιδιότητά της, για παράδειγμα η λέξη άνθρωπος μπορεί να έχει 18 διαφορετικές ιδιότητες ή 18 κόμβους-ρίζες η σ.κ. <ανθρωπιν_ον> στην οποία ανήκει (βλπ Πίνακα: 8).

Για παράδειγμα, ως υποκείμενα του γράφω μπορεί να είναι λέξεις που ανήκουν σε υπώνυμες σ.κ. της σ.κ. <ανθρωπιν_ον> και ανήκουν σε διαφορετικά υποδένδρα σ.κ. της όπως: <ατομο_νεαρης_ηλικιας>, <ατομο_μεγαλης_ηλικιας>, <διαμεν_ον_στην_πολη>, <επαγγελματιας_ενδυματος>, <εχ_ον_περιουσια_κυριοτητα>, κλπ. με ένα επιπλέον κοινό σημασιολογικό χαρακτηριστικό, την ικανότητα γραφής και ανάγνωσης. Τότε εισάγεται μία νέα σ.σ.κ.:<ανθρωπιν_ον_με_ικανοτητα_γραφης_αναγνωσης> που τις ομαδοποιεί. Έτσι έχουμε τους αναδρομικούς κανόνες:

3.1<ανθρωπιν_ον_με_ικανοτητα_γραφης_αναγνωσης>::=

...<ατομο_νεαρης_ηλικιας> | ... | <ατομο_μεγαλης_ηλικιας> |...|
 <ατομο_με_αρνητικη_συμπεριφορα> | ... | <διδασκομενος> | ... |
 <εχ_ον_συγγενεια> | ... | <διαμεν_ον_στην_πολη> | ... |
 <παροχιας_υπηρεσιων_ανθρωπιν_ον> | ... | <εχ_ον_περιουσια_κυριοτητα> | ...

3.2 <παροχιας_υπηρεσιων_ανθρωπιν_ον> ::= ... |

<παροχιας_υπηρεσιων_εκπαιδευσης> | ... | <παροχιας_υπηρεσιων_ενημερωσης>
 | ... | <παροχιας_υπηρεσιων_πολιτισμου> | <παροχιας_υπηρεσιων_διακυβερνησης>
 | ... | <παροχιας_νομικων_υπηρεσιων> | ... | <παροχιας_υπηρεσιων_υγειας> | ... |
 <παροχιας_υπηρεσιων_φυλαξης> | ... |
 <παροχιας_υπηρεσιων_γεωργιας_κτηνοτροφιας> | ...

Συμπερασματικά, πολλές φορές ως υποκείμενα ενός ρήματος μπορεί να γίνουν αποδεκτά και σύνολα ουσιαστικών που η βασική σημασιολογική ιδιότητά τους διευρύνεται ή προς τα επάνω (γενικότερη) ή προς τα κάτω (ειδικότερη), με αντίστοιχες σ.κ. να είναι

υπερώνυμο ή υπώνυμο της ήδη υπάρχουσας, κ.ο.κ. Έτσι δημιουργείται μία δ.δ.σ.κ., για μία ομάδα ρημάτων συνωνύμων, αντωνύμων ή και συνθέτων με ανάλογες σημασίες.

Πίνακας 8: Κόμβοι-Ρίζες υποδένδρων σ.κ., πρώτου επιπέδου, που διέπουν την έννοια της λέξης «άνθρωπος», με χαρακτηριστική σημασιολογική ιδιότητα ανθρώπινο ον ή σ.κ. <ανθρωπιν_ον>

a/a	ΚΟΜΒΟΙ-ΡΙΖΕΣ ΥΠΟΔΕΝΔΡΩΝ σ.κ. της έννοιας της λέξης «άνθρωπος»
1	1.0 ΑΝΘΡΩΠΟΣ
2	1.1 ΑΝΘΡΩΠΟΥ ΥΠΟΣΤΑΣΗ
3	1.2 ΑΝΤΙΚΟΙΝΩΝΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ
4	1.3 ΔΕΚΤΗΣ ΕΚΠΑΙΔΕΥΣΗΣ
5	1.4 ΣΥΓΓΕΝΗΣ
6	1.5 ΚΑΤΟΙΚΟΣ
7	1.6 ΠΑΡΟΧΕΑΣ ΥΠΗΡΕΣΙΩΝ ΓΕΩΡΓΙΑΣ – ΚΤΗΝΟΤΡΟΦΙΑΣ
8	1.7 ΕΜΠΟΡΙΟ – ΕΛΕΥΘΕΡΟ ΕΠΑΓΓΕΛΜΑ
9	1.8 ΥΠΑΛΛΗΛΟΣ
10	1.9 ΠΟΛΕΜΙΚΕΣ ΥΠΗΡΕΣΙΕΣ
11	1.10 ΔΙΑΝΟΗΤΙΚΗ ΕΡΓΑΣΙΑ
12	1.11 ΘΡΗΣΚΕΙΑ
13	1.12 ΑΘΛΗΤΙΣΜΟΣ
14	1.13 ΝΟΜΙΚΗ ΔΙΑΦΟΡΑ
15	1.14 ΥΠΗΡΕΣΙΕΣ ΔΙΑΜΕΣΟΛΑΒΗΣΗΣ
16	1.15 ΠΑΡΟΧΗ
17	1.16 ΚΥΡΙΑΡΧΙΑ – ΙΔΙΟΚΤΗΣΙΑ
18	1.17 Ο ΠΑΡΑΛΑΜΒΑΝΩΝ - ΔΙΔΩΝ ΑΚΙΝΗΤΟ

Κάθε ρήμα εισάγει τον δικό του συνταγματικό άξονα υπαγορεύοντας αντίστοιχους παραδειγματικούς άξονες με σημασιολογικώς συμβατά σύνολα ουσιαστικών ως υποκείμενα και αντικείμενά του. Για να διατυπώσομε πεπερασμένου πλήθους γενικευμένους σημασιοσυντακτικούς κανόνες για αποτελεσματική μηχανική επεξεργασία μελετήσαμε ένα αντιπροσωπευτικό δείγμα ρημάτων ως προς τις δομές που εισάγουν και το ευρύ φάσμα των εννοιών τους, πέραν των ρημάτων² της μορφολογικής διάστασης.

Στην παρούσα εργασία παρουσιάζονται τα αποτελέσματα της μελέτης που προέκυψαν από την τυποποίηση και κωδικοποίηση αντιπροσωπευτικών δειγμάτων ρημάτων και της ταξινόμησης και κωδικοποίησης μόνο των συναρμοζουσών αντιπροσωπευτικών ομάδων δείγματος ουσιαστικών (που λειτουργούν ως υποκείμενα

² Για την αποτελεσματική μορφολογική μηχανική επεξεργασία του μελετήθηκε, τυποποιήθηκε και κωδικοποιήθηκε ένα αντιπροσωπευτικό δείγμα 201 ρημάτων, βάσει των οποίων κλίνονται όλα τα ρήματα της ΝΕΓ (Μπαλτζής et al., 2006· Baldzis et al., 2005).

Πίνακας 9: Δείγμα ορισμού Σημασιολογικών Κατηγοριών με τους κωδικούς τους και Λέξεων ή Σύζευξης Λέξεων που αντιστοιχούν στην ίδια Σημασιολογική Κατηγορία με τους αντίστοιχους κωδικούς τους. Στη δεξιά στήλη όσες Σημασιολογικές Κατηγορίες είναι με έντονα Γράμματα αυτές μόνο στον πίνακά μας αντικαθίστανται από τα υπόνοιμά τους. Η Σύζευξη Λέξεων είναι με έντονα γράμματα διότι στο επόμενο επίπεδο μετατρέπεται σε Σημασιολογική Κατηγορία και αναλύεται περαιτέρω

Κανόνες Παραγωγής ή Αντικατάστασης ορισμού των Σημασιολογικών Κατηγοριών και κωδικοποίησή τους				
1.0	<εμψυχ_ον>	::=	1.1 1.2 1.3 1.4 1.5 1.6	<ανθρωπιν_ον> <ζω_ον> <πετειν_ον> <ψαρι_ον> <εντομ_ον> <ερπειτ_ον>
1.1.0	<ανθρωπιν_ον>	::=	1.1.1 1.1.2 1.1.3 1.1.4 1.1.5 1.1.6 1.1.7 1.1.8 1.1.9 1.1.10 1.1.11 1.1.12 1.1.13 1.1.14 1.1.15 1.1.16 1.1.17 1.1.18	<ανθρωπιν_υποσταση> <παροχεας_υπηρεσιων_ανθρωπιν_ον> <ατομο_με_αρνητικη_συμπεριφορα> <διδασκομενος> <εχ_ον_συγγενεια> <κατοικ_ον> <παροχ_αγροκτηνοτροφικων_υπηρεσιων> <ασκ_ον_εμποριο_ελευθερο_επαγγελμα> <εχ_ον_υπαλληλικη_σχεση> <παροχεας_πολεμικων_υπηρεσιων> <ασκ_ον_διανοητικη_εργασια> <υπηρετ_ον_θρησκεια> <υπηρετ_ον_αθλητισμο> <εχ_ον_νομικη_διαφορα> <παροχεας_υπηρεσιων_διαμεσολαβησης> <παροχεας> <εχ_ον_κυριαρχια_κατεχων_διοκτησια> <παραλαμβαν_ον_διδ_ον_ακινητο>
1.1.2.0	<παροχεας_υπηρεσιων_ανθρωπιν_ον>	::=	1.1.2.1 1.1.2.2 1.1.2.3 1.1.2.4 1.1.2.5 1.1.2.6 1.1.2.7 1.1.2.8	<παροχεας_υπηρεσιων_εκπαιδευσης> <παροχεας_υπηρεσιων_ενημερωσης> <παροχεας_υπηρεσιων_πολιτισμου> <παροχεας_υπηρεσιων_διακυβερνησης> <παροχεας_νομικων_υπηρεσιων> <παροχεας_υπηρεσιων_υγειας> <παροχεας_υπηρεσιων_φυλαξης> <παροχεας_υπηρεσιων_γεωργ_κτηνοτροφι>
1.1.2.1.0	<παροχεας_υπηρεσιων_εκπαιδευσης>	::=	1.1.2.1.1	< διδασκ_ον>
1.1.2.1.1.0	<διδασκ_ον>		→ → →	διδασκικό προσωπικό (15.0831.κ1.19.1) εκπαιδευτικός (15.0831.κ1.19.3) παιδαγωγός (15.0831.κ1.6.1)
1.1.2.1.1.1.0	<διδασκικό_προσωπικό>	::=	1.1.2.1.1.1.1 1.1.2.1.1.1.2 1.1.2.1.1.1.3 1.1.2.1.1.1.4	<διδασκικό_προσωπικό_προσχολικής> <διδασκικό_προσωπικό_α_βαθμίας> <διδασκικό_προσωπικό_β_βαθμίας> <διδασκικό_προσωπικό_γ_βαθμίας>

και αντικείμενά τους) και που μας οδήγησαν στον ορισμό μίας εναλλακτικής αξιοποίησης της έννοιας της οντολογίας [11]. Πρόκειται για μία τμηματικοποιημένη οντολογία η οποία ορίζεται κατά περίπτωση, δηλαδή μία επιμέρους οντολογία που ορίζεται βάσει των ειδικών σημασιολογικών ιδιοτήτων που υπαγορεύει το κάθε ρήμα

για τα υποκείμενα και αντικείμενά του προσαρμόζοντάς την ανάλογα ακόμα και με σύνθεση επιμέρους τμημάτων της.

Έτσι, η ταξινόμηση αυτή των δενδροειδών δομών των κοινών σημασιολογικών χαρακτηριστικών των λέξεων και των επεκτάσεών τους, σε συνδυασμό με την ταξινόμηση των λέξεων σε Ε/Κ και Παραγράφους προσδίδει στο σύστημά μας μία πλήρη, αποτελεσματική, ευέλικτη και ευκολοσυντήρητη σημασιολογική βάση. Με τη βοήθεια της βάσης αυτής, μπορούμε να ορίσουμε το πεπερασμένο σύστημα των σημασιοσυντακτικών κανόνων μας, αφού πεπερασμένα είναι τα ρήματα και ελέγξιμη η ποικιλία των ουσιαστικών κάθε φορά, ώστε να περιγράφομε τους δεσπόμενους σημασιολογικούς συνδυασμούς των εννοιών των λέξεων και κατά συνέπεια των δεσποζουσών σημασιοσυντακτικών δομών της ΝΕΓ, δηλαδή των αντίστοιχων σημασιοσυντακτικών κανόνων τους.

Το τελικό αποτέλεσμα είναι ένα μοντέλο που συνδυάζει τα πλεονεκτήματα ενός θησαυρού για την κωδικοποίηση των σημασιών των λέξεων και την ευελιξία μίας ιεραρχικής και δικτυακής δομής που προσομοιάζει αυτή μιας οντολογίας, τα μέλη της οποίας μπορούν να λειτουργήσουν ως μεταβλητές σε σημασιοσυντακτικούς κανόνες. Η ουσιώδης διαφορά όμως μεταξύ αυτής της γλωσσολογικής οντολογίας και μιας κλασικής οντολογίας είναι πως κάθε οντολογική κλάση (κόμβος στη δενδροειδή **Πίνακας 10:** Η συνέχεια του κλάδου του δένδρου του προηγούμενου πίνακα και η κατάληξή του σε λέξεις.

Κανόνες Παραγωγής Σημασιολογικών Κατηγοριών-Λέξεις που ομαδοποιούν και κωδικοποιήσεις τους			
1.1.2.1.1.1.0	<διδασκικο_προσωπικο_προσχολικης>	→	νηπιαγωγός (15.0831.κ1.6.2) . . .
1.1.2.1.1.1.2.0	<διδασκικο_προσωπικο_α_βαθμιας>	→ →	δασκάλα (15.0831.κ1.2.2) ... δάσκαλος (15.0831.κ1.1.2)
1.1.2.1.1.1.3.0	<διδασκικο_προσωπικο_β_βαθμιας>	→ → → → →	γυμνασιάρχης (15.0831.κ1.18.2) γυμναστής (20.1413.κ1.11.1) καθηγητής (15.0831.κ1.15.1) ... φυσικός (13.0399.κ1.13.9) χημικός (13.0404.κ1.9.1) ...
1.2.2.1.1.1.4.0	<διδασκικο_προσωπικο_γ_βαθμιας>	→ → → → → →	αναπληρωτής καθηγητή(15.0831.κ1.15.10) διδάκτορας (15.0831.κ1.15.6) ... λέκτορας (15.0831.κ1.15.5) ομότιμος καθηγητής (15.0831.κ1.17.1) πρύτανης (15.0831.κ1.18.5)

δομή) δεν αντιστοιχεί σε μία έννοια ή ένα σύνολο συνώνυμων εννοιών όπως στο WordNet αλλά σε μία σημασιολογική ιδιότητα ή σε ένα συνδυασμό σημασιολογικών

ιδιοτήτων στις σύνθετες σημασιολογικές κατηγορίες. Αυτό καθιστά την οντολογία πολύ πιο μικρή άρα και πιο διαχειρίσιμη (εφόσον το πλήθος των χαρακτηριστικών είναι πολύ μικρότερο του πλήθους των εννοιών των καταγεγραμμένων λέξεων), αλλά με ευρύτερο φάσμα καθώς κάθε σημασιολογική ιδιότητα ανασύρει πλήθος εννοιών που ανήκουν σε τελείως διαφορετικές έννοιες κεφάλαια. Επιπλέον η τμηματοποίηση που επιφέρει η χρήση σύνθετων κατηγοριών καθιστά την οντολογία εύκολα επεκτάσιμη και ευέλικτη, καθώς κατά τον τρόπο αυτόν μπορούν να συνδυαστούν έννοιες που ανήκουν σε διαφορετικές δένδροειδείς δομές των σημασιολογικών κατηγοριών. Τέλος, επιτρέπει την ευκολότερη διαχείριση του ιεραρχικού μοντέλου στα ανώτερα επίπεδά του³.

Εν κατακλείδι, η σημασιολογική κατηγορία εμπεριέχει και συνοψίζει ταυτόχρονα σημασιολογική πληροφορία (θησαυρός και δίκτυα σε καταληκτικό επίπεδο), συντακτική πληροφορία (σημασιολογική κατηγορία) και πληροφορία κατανομής (οντολογία⁴ σημασιολογικών δομών) σε έναν ενιαίο φορμαλισμό. Επιπροσθέτως, η ιεραρχική ταξινόμηση των απλών και σύνθετων σημασιολογικών κατηγοριών καθιστά τη γλωσσολογική οντολογία μικρή, ελέγξιμη, ευέλικτη και εύκολα επεκτάσιμη καθιστώντας τους αλγόριθμους τους συστήματος απλούς και αποτελεσματικούς, καθώς έχουν να διαχειριστούν έναν πολύ μικρότερο αριθμό κόμβων σε σχέση με μια οντολογία που οι έννοιές της αναφέρονται στις έννοιες των λέξεων που αναγνωρίζει το σύστημα.

2.3.4 Το Νεοελληνικό Συντακτικό Υπολογιστικό Λεξικό (ΝΣυνΥπΛ)

Το ΝΣυνΥπΛ [2,3,16,17,19], το οποίο ελέγχει τις συμφωνίες των μορφολογικών τύπων των λέξεων της πρότασης ανάλογα με τον συντακτικό ρόλο τους στην πρόταση.

2.3.5 Συμπεράσματα και συναφείς προσεγγίσεις στην Επεξεργασία της Νεοελληνικής Γλώσσας

Η εισαγωγή των μ. γραμματικών - παράγουν συντακτικές δομές της ΝΕΓ - είναι τροποποιημένη εκδοχή των γραμματικών της ιεραρχίας Chomsky. Η τροποποίηση έγκειται στο ότι ομαδοποιεί τους κανόνες παραγωγής σε μήτρες κανόνων συντακτικής δομής της ΝΕΓ και όχι φρασεοδομής και παράγουν τις συνήθεις και

³ Αξίζει εδώ να σημειωθεί πως και στο ανώτερο μέρος της Οντολογίας του EuroWordNet δεν έχουμε έννοιες αλλά σύνολα σημασιολογικών χαρακτηριστικών για την καλύτερη απόδοση του νοήματος των εννοιών, την ευκολότερη ταξινόμησή τους και μείωση του αριθμού των κορυφαίων κόμβων στη δένδροειδή δομή.

⁴ Αξίζει εδώ να σημειωθεί πως ο όρος *οντολογία* χρησιμοποιείται με την ευρύτερη έννοια του όρου που περιλαμβάνει και τις ταξινομιές (is-a taxonomy).

εύχρηστες συντακτικές δομές της ΝΕΓ που αντιστοιχούν σε κύριες και ακέραιες προτάσεις της ΝΕΓ, αφού οι λέξεις της φυσικής γλώσσας στις γραμματικές αυτές αντικαταστάθηκαν από το πεπερασμένο σύνολο των συντακτικών κατηγοριών της ΝΕΓ. Έτσι προέκυψε η Γλώσσα των Συντακτικών Κατηγοριών της ΝΕΓ (Γλώσσα Βάση της ΝΕΓ), μία ενδιάμεση τυπική γλώσσα, που προσομοιάζει με τη ΝΕΓ ως προς τη συντακτική δομή της, με το πλεονέκτημα ότι: 1. είναι απαλλαγμένη από γραμματικούς τύπους και έννοιες και 2. μελετάται εύκολα η συντακτική συμπεριφορά της ως τυπική γλώσσα που εύκολα προγραμματίζεται. Με τον τρόπο αυτόν προέκυψε όλη η ποικιλία των συνήθων εύχρηστων συνταγματικών αξόνων που αντιστοιχούσαν αφενός στις συντακτικές δομές βάθους της ΝΕΓ, αφετέρου από αυτές προέκυψε απεριόριστο σύνολο δομών επιφανείας της ΝΕΓ, δηλαδή κύριες και ακέραιες προτάσεις της. Με τον τρόπο αυτόν δεν χρειαζόταν σε αυτή τη φάση να ανατρέξουμε σε σώματα κειμένων για να εντοπίσουμε συντακτικούς κανόνες παραγωγής συντακτικών δομών της ΝΕΓ. Σημειώνουμε ότι οποιαδήποτε απόκλιση από την κανονική σειρά των λέξεων μίας πρότασης ή ακόμα και οι ελλειπτικές προτάσεις αφορούν μετασχηματισμούς και είναι για περαιτέρω μελέτη και έρευνα.

Οι συνταγματικοί άξονες, που μπορούσαμε να τους παράγουμε και να τους αναγνωρίσουμε πλέον μηχανικά και αποτελεσματικά, περιέχουν τα σημεία, δηλαδή τις συντακτικές κατηγορίες με τους αντίστοιχους παραδειγματικούς άξονες που εισάγουν, όπου συναντώνται η σύνταξη με τη μορφολογία και τη σημασιολογία. Το Βασικό Νεοελληνικό Πολυλεξικό (ΒΝΠ) είναι ένα σύστημα Ηλεκτρονικών Υπολογιστικών Λεξικών της ΝΕΓ που περιέχει κατάλληλα ταξινομημένα, κωδικοποιημένα και αρχειοθετημένα τα γλωσσολογικά δεδομένα της ΝΕΓ και οι αλγόριθμοι που το συνοδεύουν είναι αλγοριθμοποιημένοι οι κανόνες της ΝΕΓ που μετατρέπουν τις συντακτικές δομές βάθους σε αποδεκτές δομές επιφανείας κάθε φορά.

Επισημαίνουμε επίσης ότι, το σημασιολογικό μοντέλο που παρουσιάζεται στην παρούσα, ως μέρος μίας μεθόδου που προτάθηκε για την επεξεργασία της ΝΕΓ, βασίστηκε όπως και όλη η προταθείσα μέθοδος:

1. στην ιδιαιτερότητα της ΝΕΓ έναντι της ΑΓ ως βαριά κλιτής αλλά και με τη δυναμική της σύνθεσης που τη χαρακτηρίζει σε όλες τις γλωσσολογικές διαστάσεις της ως κλιτή και νοητική γλώσσα,
2. στη μελέτη, ταξινόμηση και περιγραφή με αλγορίθμους των κανόνων της ΝΕΓ άρα και των μηχανισμών που είναι καλώς ορισμένοι και με σαφήνεια διατυπωμένοι,

3. στη μελέτη, ταξινόμηση, κωδικοποίηση και κατάλληλη αρχειοθέτηση των γλωσσολογικών δεδομένων σε βάσεις δεδομένων και

4. στον συνδυασμό των προτερημάτων ενός προσαρμοσμένου θησαυρού, ενός εναλλακτικού γλωσσολογικού δικτύου και μίας εναλλακτικής τμηματικοποιημένης γλωσσολογικής οντολογίας σε έναν ενιαίο φορμαλισμό που επιχειρεί να συνδυάσει και να συνοψίσει ταυτόχρονα σημασιολογική πληροφορία, συντακτική πληροφορία και πληροφορία κατανομής αξιοποιώντας όλα τα παραπάνω χαρακτηριστικά.

Συμπερασματικά το BNP και οι αλγόριθμοι που το συνοδεύουν αποτελούν μία καινοτόμο μέθοδο Επεξεργασίας της ΝΕΓ η οποία αφενός αντιμετωπίζει σφαιρικά τη διαχείριση και των τριών γλωσσολογικών διαστάσεων, συγχρόνως δηλαδή, αυτή της μορφολογικής, της σημασιολογικής και της συντακτικής διάστασης, και αφετέρου στηρίζεται αποκλειστικά στην υπολογιστική αξιοποίηση των δομολειτουργικών κανόνων των αντίστοιχων γλωσσολογικών διαστάσεων. Διάφορες άλλες προσεγγίσεις που έγιναν ή γίνονται στηρίζονται σε πρότυπα που έχουν εισαχθεί για την αντιμετώπιση άλλων γλωσσών με διαφορετικά χαρακτηριστικά αυτών της ΝΕΓ ή στηρίζονται σε στατιστικές μεθόδους. Ενδεικτικά αναφέρουμε προσεγγίσεις για την επεξεργασία της ΝΕΓ που έχουν εισαχθεί από: 1. το Εργαστήριο Ενσύρματης Τηλεπικοινωνίας του Πανεπιστημίου Πατρών, συστήματα που αφορούν επί μέρος στη μορφολογία [39,40,41], στη σύνταξη [34] και στη σημασιολογία [42,43], 2. το Εργαστήριο Μηχανικών Λογισμικού και Γνώσης (SKEL | Software and Knowledge Engineering Laboratory) του Εθνικού Κέντρου Επιστημονικών Ερευνών Δημόκριτος (NCSR), μεταξύ άλλων την ετικετοποίηση μερών του λόγου (a part-of-speech tagger, a case tager – [39]) κ.λπ., επίσης εισήγαγαν το SKEL λεξικό με 60.000 λήμματα [38], τέλος 3. το Ινστιτούτο Επεξεργασίας του Λόγου (ILSP), εισήγαγε ένα λεξικό στα πρότυπα του PAROLE/SIMPLE [36], [33], το οποίο σε μία επόμενη έκδοση [32] περιελάμβανε 60.000 καταχωρήσεις, από τις οποίες ένα υποσύνολο 30.000 καταχωρήσεων αφορούσαν στη σύνταξη και ένα υποσύνολο 15.000 καταχωρήσεων στη σημασιολογία. Σημειώνουμε ότι το λεξικό αυτό αποτελεί την πιο ολοκληρωμένη πρόταση υπολογιστικού λεξικού της ΝΕΓ, το οποίο μπορεί να χρησιμοποιηθεί σε μία ευρείας κλίμακας εφαρμογές Επεξεργασίας Φυσικής Γλώσσας, όπως είναι η τεχνολόγηση και παραγωγή προτάσεων, εξαγωγή νοήματος κ.λπ. Σημειώνουμε επίσης ότι είναι η μόνη προσέγγιση που έχει κάποιες ομοιότητες στον τμηματικοποιημένο σχεδιασμό, το εύρος της κωδικοποιημένης πληροφορίας και τη χρήση οντολογίας στην περιγραφή των εννοιών των λέξεων. Το σύστημα του BNP και των αλγορίθμων

του ορίζουν τους αποδεκτούς μορφολογικούς τύπους σε σχέση με συντακτικούς ρόλους και κατηγορίες παράλληλα με τους σημασιολογικούς συνδυασμούς των εννοιών των λέξεων. Η έρευνα συνεχίζεται κυρίως με μεθόδους που βασίζονται στη στατιστική αλλά και πρόσφατα με πειραματισμούς σε μεθόδους που αξιοποιούν τα νευρωνικά δίκτυα [20,21].

Κεφάλαιο 3

Τα Υπολογιστικά Λεξικά της Νεοελληνικής οι αλγόριθμοι και οι εφαρμογές τους οι βάσεις για διδασκαλία σε ανοικτό περιβάλλον με τη βοήθεια του Η/Υ στην τάξη ή στο διαδίκτυο

3.1 Εισαγωγή

Με τη βοήθεια των υπολογιστικών λεξικών του ΒΝΠ και των αλγόριθμων που το συνοδεύουν εισάγονται οι εφαρμογές του Τεχνολόγου και του Γεννήτορα προτάσεων και κειμένων της ΝΕΓ [5,19]. Οι εφαρμογές αυτές αποτελούν τη βάση αφενός για την αυτόματη επεξεργασία της ΝΕΓ, αφετέρου για την περαιτέρω ανάπτυξη εφαρμογών που στηρίζονται σε αυτές, όπως ο Διορθωτής προτάσεων και κειμένων της ΝΕΓ [1,2,3,4,5,10,15,18], ο Σημασιολογικός Αναλυτής προτάσεων και κειμένων της ΝΕΓ καθώς και το υπολογιστικό εργαλείο του Συλλαβισμού λέξεων της ΝΕΓ [13]. Οι εφαρμογές αυτές, επίσης, μπορούν: 1. να λειτουργήσουν ανεξάρτητα ή να ενσωματωθούν με οποιονδήποτε συνδυασμό σε άλλα λογισμικά και εφαρμογές κοινής χρήσης, για παράδειγμα σε έναν γενικό επεξεργαστή κειμένου όπως το Word της Microsoft, κ.ο.κ και 2. να αποτελέσουν τη βάση ανάπτυξης ενός ανοικτού εκπαιδευτικού λογισμικού,, οποιουδήποτε θεματικού αντικειμένου ή ακόμα και να ενσωματωθούν σε αυτό όπως στον Λεξικράτη [22], διαδικτυακό εκπαιδευτικό λογισμικό φωνολογικής ενημερότητας για παιδιά πρώτης σχολικής ηλικίας, κατάλληλο και για παιδιά με μαθησιακά προβλήματα δυσλεξίας, όπου

ενσωματώνεται ο συλλαβισμός⁵, κ.ο.κ.

Στην παρούσα μελέτη παρουσιάζεται ο Τεχνολόγος, ο Γεννήτορας, ο Διορθωτής και ο Σημασιολογικός Αναλυτής μόνο των κυρίων και ακεραίων προτάσεων της ΝΕΓ προσέγγιση που αποτελεί καταρχήν το βασικότερο εργαλείο για την ανάπτυξη ανοικτού εκπαιδευτικού λογισμικού οποιουδήποτε θεματικού αντικειμένου και κυρίως μαθηματικού περιεχομένου. Σημειώνεται ότι ο Τεχνολόγος, ο Γεννήτορας κειμενοποιημένων προτάσεων της ΝΕΓ και κατ' επέκταση και οι αντίστοιχοι Διορθωτής και ο Σημασιολογικός Αναλυτής αυτών δεν εξετάζονται στην παρούσα μελέτη. Όταν αναφερόμαστε σε μαθηματικού περιεχομένου ανοικτό εκπαιδευτικό λογισμό, αυτό μπορεί να εκφράζεται είτε με φυσική γλώσσα, στην περίπτωση μας την ελληνική, είτε χρησιμοποιώντας τα μαθηματικά σύμβολα είτε και τα δύο, δηλαδή συνδυασμό ελληνικής γλώσσας και μαθηματικού συμβολισμού. Για παράδειγμα να αναγνωρίζει εξίσου τις ισοδύναμες εκφράσεις: 1. «η δευτεροβάθμια εξίσωση με συντελεστές $\alpha=3$, $\beta=-4$ και $\gamma=1$ » και 2.1 « $3x^2-4x+1=0$ » ή και 2.2 « $3*x**2-4*x+1=0$ ».

Λέξεις Κλειδιά: Τεχνολόγος, Γεννήτορας, Διορθωτής Σημασιολογικός Αναλυτής προτάσεων της ΝΕΓ

3.2 Τεχνολόγος και Γεννήτορας προτάσεων της ΝΕΓ

3.2.1 Εισαγωγή

Ο Τεχνολόγος των προτάσεων και κατ' επέκταση των κειμένων μίας φυσικής γλώσσας είναι το κατεξοχήν εργαλείο που διέπει την αυτόματη επεξεργασία της. Με τον Τεχνολόγο επιτυγχάνεται η ανάλυση της πρότασης δηλαδή ο εντοπισμός της ενέργειας που εκφράζει, του ρήματος δηλαδή, του δράστη της ενέργειας, δηλαδή του υποκειμένου και του αποδέκτη της ενέργειας, δηλαδή του αντικειμένου και κατ' επέκταση οποιουδήποτε άλλου δευτερεύοντος όρου της πρότασης που σχετίζεται με τους προηγούμενους κύριους όρους της. Ο γεννήτορας των προτάσεων και των κειμένων μίας φυσικής γλώσσας είναι στην πραγματικότητα μία αντίστροφη διαδικασία αυτής του Τεχνολόγου. Στην επόμενη παράγραφο θα περιγραφεί

⁵ το οποίο σχεδιάστηκε και υλοποιήθηκε από το Εργαστήριο Επεξεργασίας Φυσικής Γλώσσας και Μαθηματικών Προβλημάτων του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, με Επιστημονικό Υπεύθυνο τον Λέκτορα Σ.Δ. Μπαλτζή, σε συνεργασία με την ΕΛΕΠΑΠ Παράρτημα Ιωαννίνων και Χρηματοδότηση του Ιδρύματος Μποδοσάκη. Ο Λεξικράτης διατίθεται δωρεά από το Εργαστήριο για χρήση και σε σχολικές μονάδες, όπως και ο Αριθμοσθένης ένα αντίστοιχο εκπαιδευτικό λογισμικό για παιδιά πρώτης σχολικής ηλικίας κατάλληλο και για παιδιά με μαθησιακά προβλήματα δυσαριθμησίας.

παράδειγμα Τεχνολόγου.

3.2.2 Παραδείγματα

3.2.2.1 Τεχνολόγος προτάσεων της ΝΕΓ

Το παράδειγμα του Τεχνολόγου που παρατίθεται στην ενότητα αυτή (βλπ. Οθόνη 14) επιλέγεται να είναι από τη ΝΕΓ, δηλαδή να είναι η τεχνολόγηση μίας κύριας ακέραλης και επαυξημένης πρότασης της ΝΕΓ με κάποιο εσκεμμένο μορφοσυντακτικό λάθος: «ο μαθητής μελετώ την ιστορία». Με αυτό το παράδειγμα δίνεται η δυνατότητα να σκιαγραφηθεί με λεπτομέρεια αφενός η πολυσύνθετη διαδικασία του Τεχνολόγου, που απαρτίζεται από αλληλοεξαρτώμενες επιμέρους διαδικασίες, που χαρακτηρίζονται στο παράδειγμά ως Στάδια και αφετέρου οι δυνατότητες πολλαπλής αξιοποίησης του που διαθέτει. Αυτή η επαυξημένη πρόταση

Οθόνη 14: Τεχνολόγος κύριων και ακεραίων προτάσεων της ΝΕΓ

Τεχνολόγος Προτάσεων

Δώστε πρόταση : ο μαθητής μελετώ την ιστορία

Τεχνολόγηση Πρότασης Νέα Πρόταση Εξοδος από τον Τεχνολόγο

Βασικό Νεοελληνικό Υπολογιστικό Λεξικό - Νεοελληνικό Γραμματικό Υπολογιστικό Λεξικό

ο	αναγνωριστικες	Ρήμα Μεταβατικό και Αμετάβατο,	Κανόνες
μαθητής	αναγνωριστικες	Φωνή Ενεργητική, Συζυγία 2η	
μελετώ	αναγνωριστικες	Εγκλιση Οριστική, Χρόνος Ενεστώτας, Αριθμός Ενικός,	
την	αναγνωριστικες	Πρόσωπο 1ο	
ιστορία	αναγνωριστικες	του Ρήματος: μελετώ - μελετώ	Κλίση Λέξης

Οι Λέξεις της Πρότασης Αναγνωρίστηκαν

Συντακτικό Υπολογιστικό Λεξικό της Νεοελληνικής

<υποκαμενο>: ο μαθητής	αποδεκτό	Συντακτικοί Κανόνες
<κατηγορηματικο_ρημα>: μελετώ	αποδεκτό	
<αντικαμενο>: την ιστορία	αποδεκτό	

Η Πρόταση δεν είναι Συντακτικά Αποδεκτή

Σημασιολογικό Υπολογιστικό Λεξικό της Νεοελληνικής

ο μαθητής	αποδεκτό υποκαμενο του 'μελετώ'	Συνώνυμα Αντίθετα Ιδιωματισμοί
την ιστορία	αποδεκτό αντικαμενο του 'μελετώ'	

Η Πρόταση είναι Σημασιολογικά Αποδεκτή

Δεσπόζοντες Σημασιολογικοί Συνδυασμοί

Η Πρόταση δεν είναι Αποδεκτή

Η Αποδεκτή Πρόταση είναι:

Ως προς τη Δομή της είναι:

Η Συντακτική Δομή της είναι:

Οι Κανόνες Παραγωγής της Συντακτικής Δομής της είναι:

Επαυξημένη

<υποκαμενο><κατηγορηματικο_ρημα><αντικαμενο>

<<επαυξημενη_προταση>>:=<<επαυξημενο_υποκαμενο>><<επαυξημενο_κατηγορημα>>,<<επαυξημενο_υποκαμενο>>:=<υποκαμενο>,<<επαυξημενο_κατηγορημα>>:=<κατηγορηματικο_ρημα><<επαυξημενος_προσδιορισμος_ρηματος>>,<<επαυξημενος_προσδιορισμος_ρηματος>>:=<αντικαμενο>.

Αυτόματη Επεξεργασία της Νεοελληνικής σε Ανοικτό Εκπαιδευτικό Περιβάλλον

ΒΝΠ Τεχνολόγος Γεννήτορας Βοήθεια Εξοδος

Page 1 Sec 1 1/1 At 2,5€

Start Αυτόματη Επεξεργασία της Νεοελληνικής σε Ανοικτό Εκπαιδευτικό Περιβάλλον Document1 - Microsoft... Τεχνολόγος Προτά... Jasc Paint Shop Pro 12:15 μμ

της ΝΕΓ του παραδείγματός αποτελείται από ένα μεταβατικό, που μπορεί να συναντηθεί και σαν αμετάβατο, πολύσημο ρήμα, εδώ με την κωδικοποιημένη έννοια: «ΝΟΥΣ . Μάθηση . ρήμα . α/αλ ή 15 . 0828 . κ2. α/αλ (βλπ. Πίνακα 6), ένα από τα

αποδεκτά υποκείμενα του, όπως επίσης και ένα από τα αποδεκτά αντικείμενα που δέχεται η συγκεκριμένη έννοια της μεταβατικής διάστασης του ρήματος αυτού.

αρχή

1^ο Στάδιο του Τεχνολόγου: δίδεται στο σύστημα του ΒΝΠ η πρόταση στη φυσική γλώσσα, στην προκειμένη περίπτωση η πρόταση της ΝΕΓ, «ο μαθητής διαβάζω την ιστορία»,

2^ο Στάδιο του Τεχνολόγου: με τον Μορφολογικό Αναλυτή του ΒΝΠ, δηλαδή, το Νεοελληνικό Μορφολογικό Υπολογιστικό Λεξικό, αναγνωρίζονται μία προς μία οι λέξεις της πρότασης,

Επιμέρους Απάντηση 1: Οι λέξεις της πρότασης αναγνωρίστηκαν.

3^ο Στάδιο του Τεχνολόγου: με τον Συντακτικό Αναλυτή του ΒΝΠ, δηλαδή, το Νεοελληνικό Συντακτικό Υπολογιστικό Λεξικό, ελέγχονται και αντιστοιχίζονται στις λέξεις της δοθείσας πρότασης ανάλογα με τον συντακτικό τους ρόλο οι αντίστοιχες συντακτικές κατηγορίες, δηλαδή μία διαδικασία ετικετοποίησης, και προκύπτει μία πρόταση συντακτικών κατηγοριών. Στην περίπτωσή μας εντοπίστηκε σφάλμα και μάλιστα στο πρόσωπο του ρήματος, που είναι στο πρώτο πρόσωπο, δεν υπάρχει συμφωνία με το υποκείμενο, θα έπρεπε να εκφέρεται στο τρίτο πρόσωπο. Η αντιστοίχιση των λέξεων με τις συντακτικές κατηγορίες έγινε και αποθηκεύτηκε δείχνοντας το «a» ως «ο μαθητής» και το «e₁» ως «την ιστορία». Το σύστημα αποθήκευσε και το «b» ως «μελετώ», με σφάλμα στο πρόσωπο, το σωστό στο τρίτο πρόσωπο,

Επιμέρους Απάντηση 2: Η πρόταση δεν είναι συντακτικά αποδεκτή,

4^ο Στάδιο του Τεχνολόγου: με τον Σημασιολογικό Αναλυτή του ΒΝΠ, δηλαδή το Νεοελληνικό Σημασιολογικό Υπολογιστικό Λεξικό που αποτελείται από το Λογικό Υπολογιστικό Λεξικό των Βασικών Εννοιών και το Υπολογιστικό Λεξικό των Συμφραζομένων της ΝΕΓ. Τα επιμέρους αυτά λεξικά ελέγχουν τον αποδεκτό σημασιολογικό συνδυασμό των εννοιών των λέξεων ως προς τον συντακτικό του ρόλο και αποθηκεύουν την επιβεβαίωση ότι το δείχνοντας «ο μαθητής» είναι αποδεκτό «a» ως προς το «b» που είναι το «μελετώ» και «την ιστορία» που είναι αποδεκτό «e₁» ως προς το «b» που είναι το «μελετώ»,

Επιμέρους Απάντηση 3: Η πρόταση είναι σημασιολογικά αποδεκτή,

5^ο Στάδιο του Τεχνολόγου: Συμπέρασμα 1, αποδεκτή ή όχι αποδεκτή πρόταση,

Απάντηση: Η πρόταση δεν είναι αποδεκτή,

6° Στάδιο του Τεχνολόγου: Διόρθωση της δοθείσας πρότασης,

Απάντηση: Η αποδεκτή πρόταση είναι «ο μαθητής μελετά την ιστορία»

7° Στάδιο του Τεχνολόγου: Ταξινόμηση της πρότασης ως προς τη δομή της,

Απάντηση: Επαυξημένη

8° Στάδιο του Τεχνολόγου: Αντίστοιχη αλυσίδα των συντακτικών κατηγοριών της πρότασης ή η συντακτική δομή της,

Απάντηση: abe_1 ή $\langle \text{υποκειμενο} \rangle \langle \text{κατηγορηματικο_ρημα} \rangle \langle \text{αντικειμενο} \rangle$

9° Στάδιο του Τεχνολόγου: οι Κανόνες Παραγωγής της Συντακτικής Δομής της πρότασης,

Απάντηση:

$\langle \langle \text{επαυξημενη_προταση} \rangle \rangle ::= \langle \langle \text{επαυξημενο_υποκειμενο} \rangle \rangle \langle \langle \text{επαυξημενο_κατηγορημα} \rangle \rangle$,
 $\langle \langle \text{επαυξημενο_υποκειμενο} \rangle \rangle ::= \langle \langle \text{υποκειμενο} \rangle \rangle$,
 $\langle \langle \text{επαυξημενο_κατηγορημα} \rangle \rangle ::= \langle \langle \text{κατηγορηματικο_ρημα} \rangle \rangle \langle \langle \text{επαυξημενος_προσδιορισμος_ρηματος} \rangle \rangle$,
 $\langle \langle \text{επαυξημενος_προσδιορισμος_ρηματος} \rangle \rangle ::= \langle \langle \text{αντικειμενο} \rangle \rangle$.

10° Στάδιο του Τεχνολόγου: Η μήτρα των συντακτικών κανόνων παραγωγής της πρότασης και οι αντίστοιχοι χαρακτηριστικοί εκθέτες

Απάντηση:

1. Η μήτρα συντακτικών κανόνων παραγωγής της πρότασης είναι

$$p_{3,20} = \{ S \rightarrow AB, A \rightarrow a, B \rightarrow bE_1, E_1 \rightarrow e_1 \},$$

όπου: $S \equiv \langle \langle \text{προταση} \rangle \rangle$, $A \equiv \langle \langle \text{επαυξημενο_υποκειμενο} \rangle \rangle$, $B \equiv \langle \langle \text{επαυξημενο_κατηγορημα} \rangle \rangle$,
 $E_1 \equiv \langle \langle \text{επαυξημενο_αντικειμενο_1} \rangle \rangle$, $a \equiv \langle \langle \text{υποκειμενο} \rangle \rangle$, $b \equiv \langle \langle \text{κατηγορηματικο_ρημα} \rangle \rangle$ και $e_1 \equiv \langle \langle \text{αντικειμενο_1} \rangle \rangle$ και

2. Οι αντίστοιχοι χαρακτηριστικοί εκθέτες είναι,

Πίνακας 12: Η

ΜΗΤΡΕΣ ΣΥΝΤΑΚΤΙΚΩΝ ΚΑΝΟΝΩΝ ΤΗΣ ΝΕΓ	ΑΛΥΣΙΔΕΣ ΣΥΝΤΑΚΤΙΚΩΝ ΚΑΤΗΓΟΡΙΩΝ	ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΙ ΕΚΘΕΤΕΣ																				
		κ	i	j	j	κ	...	κ ₂	i	j	j	κ ₂	ρ ₁	ρ ₂	λ	i	j	j	λ	...	μ	...
$p_{3,20}$	$a^x b^y e_1^z$ ή abe_1	0	-	-	-	-	...	-	-	-	-	-	1	0	-	-	-	-	-	...	1	...

τέλος.

3.2.2.2 Γεννήτορας προτάσεων της ΝΕΓ

Οι διαδικασίες που διέπουν τον Γεννήτορα κύριων και ακεραίων προτάσεων της ΝΕΓ

ακολουθούν ακριβώς την αντίστροφη πορεία των Σταδίων που ακολουθεί ο Τεχνολόγος. Η διαδικασία αυτή αποτυπώνεται στο παρακάτω παράδειγμα του Γεννήτορα απλών και σύνθετων προτάσεων της ΝΕΓ.

αρχή

1^ο Στάδιο Γεννήτορα:

1. Επιλογή της αλυσίδας των χαρακτηριστικών εκθετών (χ.ε.) από τον αντίστοιχο πίνακα της κατάλληλης μήτρας συντακτικών κανόνων παραγωγής της πρότασης των συντακτικών κατηγοριών που θέλουμε να παράγουμε,
2. Εισαγωγή των επιλεγμένων τιμών των χ.ε.:

"κ=2, i=1, j= -, j=1, κ=1, ρ₁=1, ρ₂=0, λ= -, i= -, j= -, j= -, λ= -",

2^ο Στάδιο Γεννήτορα: έλεγχος ορθότητας/ύπαρξης της δοθείσας αλυσίδας χ.ε, «ναι» ή «όχι»,

Επιμέρους Απάντηση 1: «ναι»,

3^ο Στάδιο Γεννήτορα: εύρεση της μήτρας κανόνων, από τον αντίστοιχο πίνακα της συντακτικής δομής της αλυσίδας/πρότασης των συντακτικών κατηγοριών που θέλουμε να παράξουμε,

Επιμέρους Απάντηση 2: $\pi_{2,13} = \{ S \rightarrow AB, A \rightarrow a | avA | auA, B \rightarrow b \}$

4^ο Στάδιο Γεννήτορα: παραγωγή της πρότασης των συντακτικών κατηγοριών,

Επιμέρους Απάντηση 3: "avanaauab",

5^ο Στάδιο Γεννήτορα: τύπος της πρότασης που παράχθηκε.

Επιμέρους Απάντηση 4: σύνθετη πρόταση με περισσότερα από ένα υποκείμενα (4) και κατηγορηματικό ρήμα αμετάβατο,

6^ο Στάδιο Γεννήτορα: οι αλγόριθμοι του ΒΝΠ εμφανίζουν υποκατάλογο του Βασικού Νεοελληνικού Λεξικού του ΒΝΠ για την επιλογή κατηγορηματικού αμετάβατου ρήματος, για παράδειγμα επιλογή του ρήματος «ταξιδεύω»,

Επιμέρους Απάντηση 5: για παράδειγμα επιλογή του ρήματος «ταξιδεύω»,

7^ο Στάδιο Γεννήτορα: με τη βοήθεια του Νεοελληνικού Σημασιολογικού Υπολογιστικού Λεξικού, ακριβέστερα των επιμέρους λεξικών του, δηλαδή, του Υπολογιστικού Λογικού Λεξικού των Βασικών Εννοιών και του Υπολογιστικού Λεξικού των Συμφραζομένων της ΝΕΓ εμφανίζεται μετά υπολογιστική διαδικασία κατάλογος λέξεων με την αντίστοιχη ταξινόμηση του Λεξικού των Βασικών Εννοιών που είναι αποδεκτές ως υποκείμενα του επιλεγέντος ρήματος, στην περίπτωση μας

του «ταξιδεύω».

Επιμέρους Απάντηση 6: επιτρεπτή επιλογή μόνο τεσσάρων λέξεων, για παράδειγμα, ο πατέρας, η μητέρα, ο γιός, η κόρη,

8^ο Στάδιο Γεννήτορα: με τη βοήθεια του Συντακτικού Υπολογιστικού Λεξικού του ΒΝΠ σχηματίζεται η αντίστοιχη αποδεκτή πρόταση της ΝΕΓ,

Επιμέρους Απάντηση 7: «ο πατέρας, η μητέρα, ο γιός και η κόρη ταξιδεύουν»

τέλος

3.3 Διορθωτής και Σημασιολογικός Αναλυτής προτάσεων της ΝΕΓ

3.3.1 Εισαγωγή

Οι Διορθωτής και ο Σημασιολογικός Αναλυτής προτάσεων της ΝΕΓ στηρίζονται αποκλειστικά και μόνο στον Τεχνολόγο που παρουσιάστηκε στην προηγούμενη παράγραφο. Σημειώνεται ότι ακολουθώντας τα Στάδια εφαρμογής του Τεχνολόγου σε προτάσεις της ΝΕΓ, κάποιες από τις πληροφορίες που το σύστημα υπολογίζει, εξάγει και αυτοτροφοδοτεί την εν γένει διαδικασία του - αξιοποιώντας τα δεδομένα και τους κανόνες που τα διέπουν - απομονώνονται επιλεκτικά, ανάλογα με την περαιτέρω εφαρμογή που θέλουν να εξυπηρετήσουν. Οι εφαρμογές αυτές, εν προκειμένω, είναι ο Διορθωτής και ο Σημασιολογικός Αναλυτής προτάσεων της ΝΕΓ

3.3.2 Παραδείγματα

3.3.2.1 Διορθωτής προτάσεων της ΝΕΓ

Συνοπτικά θα σκιαγραφήσουμε τα Στάδια του Διορθωτή προτάσεων της ΝΕΓ και των δυνατοτήτων που παρέχει (βλπ. Οθόνη 15):

αρχή

1^ο Στάδιο: Εισαγωγή της υπό ελέγχου πρότασης ή προτάσεων της ΝΕΓ

2^ο Στάδιο:

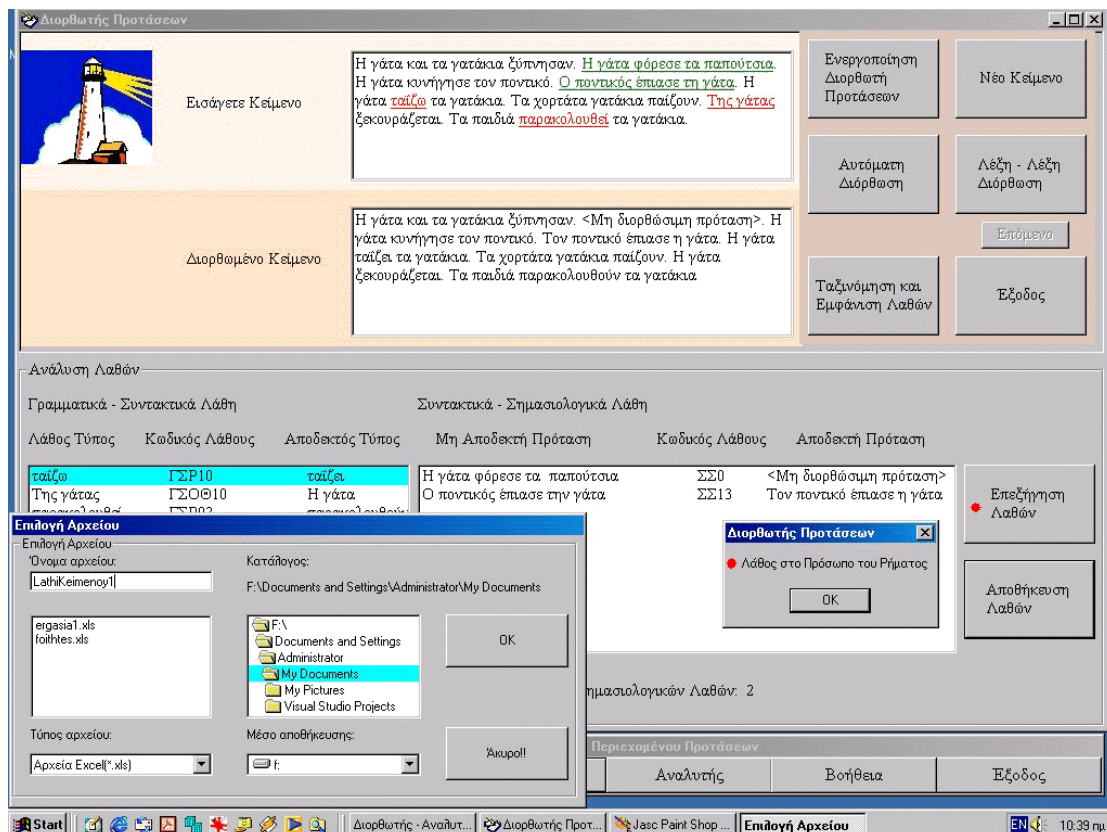
1. Λεξική και Μορφολογική αναγνώριση των λέξεων της πρότασης της ΝΕΓ, μία – μία, με τη βοήθεια αντίστοιχα του Βασικού Νεοελληνικού Λεξικού και του Νεοελληνικού Μορφολογικού Υπολογιστικού Λεξικού,
2. αναγνώριση του συντακτικού ρόλου κάθε λέξης της κάθε πρότασης της ΝΕΓ με τη βοήθεια του Νεοελληνικού Συντακτικού Υπολογιστικού Λεξικού,

3. αναγνώριση του σημασιολογικού αποδεκτού των εννοιών των λέξεων της κάθε πρότασης με τη βοήθεια των επιμέρους λεξικών του Νεοελληνικού Σημασιολογικού Υπολογιστικού Λεξικού, δηλαδή, του Υπολογιστικού Λογικού Λεξικού των Βασικών Εννοιών και του Υπολογιστικού Λεξικού των Συμφραζομένων της ΝΕΓ,

Επιμέρους Απαντήσεις 1:

1. αποδοχή της κάθε πρότασης της ΝΕΓ ή
2. μη αποδοχή της αντίστοιχα, με κωδικοποίηση των λαθών ως προς

Οθόνη 15: Διορθωτής κύριων και ακεραίων προτάσεων της ΝΕΓ



- 2.1 την αναγνώρισή της,
- 2.2 τον αποδεκτό μορφολογικό τύπο της,
- 2.3 τον αποδεκτό μορφολογικό τύπο της σε σχέση με τον συντακτικό ρόλο της και τέλος
- 2.4 τον αποδεκτό σημασιολογικό συνδυασμό των εννοιών των λέξεων σε σχέση με τον συντακτικό ρόλο τους,

3ο Στάδιο: Το σύστημα ταξινομεί και αποθηκεύει σε προσβάσιμα αρχεία τα λάθη σύμφωνα με τους παραπάνω τύπους λαθών, της κάθε πρότασης, για περαιτέρω αξιοποίηση,

4ο Στάδιο: Το σύστημα διορθώνει:

1. Αυτόματα τα λάθη της κάθε πρότασης, όσες προτάσεις δεν διορθώνονται το δηλώνει (βλπ. Οθόνη 15) και
2. Λέξη προς λέξη της κάθε πρότασης,

τέλος.

3.3.2.2 Σημασιολογικός Αναλυτής προτάσεων της NEΓ

Επίσης συνοπτικά θα σκιαγραφήσουμε τα Στάδια του Σημασιολογικού Αναλυτή προτάσεων της NEΓ και των δυνατοτήτων που παρέχει (βλπ. Οθόνη 16):

αρχή

1ο Στάδιο: Εισαγωγή της υπό σημασιολογικής ανάλυσης πρότασης ή προτάσεων της NEΓ,

2ο Στάδιο: Κλήση του Διορθωτή προτάσεων της NEΓ μόνο στα επιμέρους Στάδια 2 και 4.1,

Οθόνη 16: Σημασιολογικός Αναλυτής κύριων και ακεραίων προτάσεων της NEΓ

The screenshot shows the 'Αναλυτής Περιεχομένου Προτάσεων' window. It is divided into several sections:

- Εισάγετε Κείμενο:** A text area containing the sentence: "Η γάτα και τα γατάκια ζήτησαν. Η γάτα φόρεσε τα παπούτσια. Η γάτα κινήθηκε τον ποντικό. Ο ποντικός έπιασε τη γάτα. Η γάτα τσίξ τα γατάκια. Τα χορτάτα γατάκια παίζουν. Της γάτας ξεκουράζεται. Τα παιδιά παρακολουθεί τα γατάκια." The words "Η γάτα", "Ο ποντικός", and "Της γάτας" are highlighted in red.
- Διορθωμένο Κείμενο:** A list of 7 corrected sentences, with the 4th sentence "4. Τον ποντικό έπιασε τη γάτα." highlighted in blue.
- Ανάλυση Περιεχομένου:**
 - Υποκείμενο:** η γάτα
 - Ρήμα:** έπιασε
 - Αντικείμενο:** τον ποντικό
- Επιλέξτε Όρους της Πρότασης:** A grid of checkboxes for selecting grammatical categories. The first checkbox, "Υποκείμενο / Υποκείμενα", is checked.
- Εχετε επιλέξει:** A text box containing "η γάτα".
- με Έννοια:** A text box containing "αρπακτικό_κατουκίδιο".
- που περιέχονται στις εξής Προτάσεις:** A text box containing "1 3 4 5 7".
- Εμφάνιση των Επιλεγμένων Όρων της Πρότασης:** A button.
- Εμφάνιση της Έννοιας των Επιλεγμένων Όρων:** A button.
- Εμφάνιση των Προτάσεων που περιέχουν τους Επιλεγμένους Όρους:** A button.
- Νέα Επιλογή Όρων:** A button.
- Αποθήκευση Αποτελεσμάτων:** A button.

3ο Στάδιο: Αντιστοίχιση σε κάθε λέξη της κάθε πρότασης ανάλογα με τον συντακτικό ρόλο της την αντίστοιχη συντακτική κατηγορία,

4^ο Στάδιο: Αντιστοίχιση σε κάθε λέξη της κάθε πρότασης την αντίστοιχη σημασιολογική κατηγορία της και κατ' επέκταση και την αντίστοιχη έννοια της βάσει του Υπολογιστικού Λογικού Λεξικού των Βασικών Εννοιών ενός εκ των δύο λεξικών που απαρτίζουν το Νεοελληνικό Σημασιολογικό Υπολογιστικό Λεξικό, υπενθυμίζουμε ότι το άλλο είναι το Υπολογιστικό Λεξικό των Συμφραζομένων της ΝΕΓ,

5^ο Στάδιο: το σύστημα ομαδοποιεί, καταμετρά, αποθηκεύει και εμφανίζει από τις δοθείσες προτάσεις, ανάλογα τι θα ζητηθεί, εκείνες τις προτάσεις των οποίων, για παράδειγμα, τα υποκείμενα είναι ίδια ή συνώνυμα ή αντίθετα ή διαφέρουν. Η διαδικασία αυτή μπορεί να επαναληφθεί για οποιονδήποτε όρο ή συνδυασμό όρων των υπό σημασιολογική ανάλυση προτάσεων της ΝΕΓ για περαιτέρω αξιοποίηση.

τέλος.

Συμπερασματικά, με τις δυνατότητες στην επεξεργασία της ελληνικής γλώσσας που παρέχονται από τις παραπάνω εφαρμογές γίνονται αντιληπτά τα πλεονεκτήματα που προσδίδονται στην αξιοποίηση της ΝΕΓ σε ανοικτά εκπαιδευτικά λογισμικά για τη διδασκαλία, στην τάξη ή στο διαδίκτυο, αυτής καθαυτής της ΝΕΓ ή για οποιοδήποτε άλλο αντικείμενο με χρήση της ΝΕΓ.

Κεφάλαιο 4

Συγκρίσεις και Συμπεράσματα - Προτάσεις για περαιτέρω έρευνα

Οι σύγχρονες καθημερινές ανάγκες και τα παγκοσμιοποιημένα πρότυπα επιβάλλουν στην καθημερινή ζωή άρα και στην εκπαίδευση άλλες συνήθειες και προδιαγραφές αντιμετώπισής τους. Οι προδιαγραφές αυτές υπαγορεύονται από την ανάγκη χρήσης και αξιοποίησης των Νέων Τεχνολογιών της διαχείρισης της Πληροφορίας και της Επικοινωνίας. Το τελευταίο επιβάλλει όχι μόνο τη γνώση αλλά και την καθημερινή χρήση των Νέων Τεχνολογιών και ειδικά στην Εκπαίδευση ως αντικείμενο διδασκαλίας και ως μέσο αξιοποίησής τους στη διδασκαλία.

Το ΒΝΠ και οι αλγόριθμοι που το συνοδεύουν είναι ένα σύστημα που προσεγγίζει ταυτόχρονα την επεξεργασία και των τριών γλωσσολογικών διαστάσεων της ΝΕΓ, δηλαδή τη μορφολογία, τη σύνταξη και τη σημασιολογία. Στην παρούσα μελέτη δεν κρίθηκε αναγκαίο να συμπεριληφθεί η επεξεργασία των δευτερευουσών προτάσεων, δηλαδή εκείνη τη διαδικασία που επεξεργάζεται πολυσύνθετα κείμενα της ΝΕΓ. Η προσέγγιση αυτή ακολουθήθηκε διότι θέλαμε να αναδείξουμε μία μέθοδο που ενώ επεξεργάζεται διεξοδικά τη ΝΕΓ σε επίπεδο κυρίων και ακεραίων προτάσεων δίνει τη δυνατότητα να κατασκευαστούν εργαλεία διδασκαλίας, δηλαδή, ανοικτά εκπαιδευτικά λογισμικά, τα οποία αφενός μπορούν να ακολουθήσουν σύγχρονες έγκριτες εκπαιδευτικές μεθόδους σε οποιοδήποτε μαθησιακό επίπεδο, αφετέρου να γίνεται η διδασκαλία, σύγχρονη ή ασύγχρονη, με τη χρήση της ΝΕΓ στην τάξη ή στο

διαδίκτυο με τη βοήθεια εκπαιδευτικού.

Σύμφωνα με τις θεμελιώδεις παιδαγωγικές αρχές η κατανόηση απαιτεί μελέτη. Είναι δεδομένο ότι το αποτέλεσμα της μελέτης η μάθηση απαιτεί την επανάληψη, τη διόρθωση των λαθών και την εξάσκηση, δηλαδή, καλή γνώση της θεωρίας και ανατροφοδότηση ποικιλίας εφαρμογών και ασκήσεων. Αυτό επιβάλλει απλότητα στους μηχανισμούς χρήσης και συγχρόνως δυνατότητες διεξοδικής προσέγγισης και εμπέδωσης της θεωρίας και των εφαρμογών – ασκήσεων σε όλα τα επίπεδα και τις διαστάσεις της γνώσης του περιεχομένου του. Με άλλα λόγια την αποφυγή υπερφόρτωσης του χρήστη-μελετητή με πολύπλοκους κανόνες και εκτενείς προτάσεις που τον μπερδεύουν και τον απροπροσανατολίζουν.

Συνήθως οι μέθοδοι ΕΦΓ εστιάζουν στον τρόπο που ο Η/Υ μπορεί καλύτερα να κατανοήσει και να επεξεργαστεί τη φυσική γλώσσα. Οι μέθοδοι ΕΦΓ που συμβάλλουν στην καλύτερη κατανόηση και επεξεργασίας της γλώσσας είναι συνήθως αυτοί των καταλόγων σημασιολογικών συμφωνιών εννοιών λέξεων, της σήμανσης ή των ετικετοποιητών κειμένων, της αναγνώρισης και σύνθεσης φωνής, των μορφολογικών και συντακτικών αναλυτών καθώς και αυτών της αυτόματης μετάφρασης. Ωστόσο αυτές οι τεχνικές εφαρμόζονται μεμονωμένα στα εκπαιδευτικά λογισμικά, χωρίς να εφαρμόζονται με τρόπο ώστε να τα καθιστούν ανοικτά εκπαιδευτικά λογισμικά που στηρίζονται σε έξυπνες υπολογιστικές διαδικασίες διαχείρισης των γλωσσολογικών και εν γένει δεδομένων τους (1,2,3,4,5,6,19).

Σε αντίθεση οι τεχνικές ΕΦΓ που χαρακτηρίζουν το προταθέν μοντέλο, που μελετούμε, προτείνει μία πλήρη και διεξοδική προσέγγιση επεξεργασίας της ΝΕΓ σε όλα τα γλωσσολογικά της επίπεδα – μορφολογία, σύνταξη και σημασιολογία – σύμφωνα με τις σχετικές οδηγίες και προδιαγραφές που έχει εισηγηθεί το Υπουργείο Εθνικής Παιδείας και Θρησκευμάτων (). Ο Γεννήτορας και ο Τεχνολόγος όπως περιγράφηκαν σε προηγούμενη ενότητα παρέχουν στον μελετητή τη δυνατότητα να συνθέτουν και να εξασκούνται σε άπειρο πλήθος προτάσεων που απεικονίζουν τη χρήση και την εφαρμογή του πλήρους φάσματος των μορφολογικών, συντακτικών και σημασιολογικών κανόνων της ΝΕΓ. Για παράδειγμα, με τον Γεννήτορα κυρίων και ακεραίων προτάσεων της ΝΕΓ ο μελετητής μπορεί να σχηματίσει προτάσεις από ένα κατάλογο λέξεων. Παράλληλα ο Τεχνολόγος πληροφορεί τον μελετητή για τη συντακτική δομή της πρότασης και για τους συντακτικούς κανόνες που την διέπουν, διορθώνοντας τα πιθανά συντακτικά λάθη.

Το ΒΝΠ παρέχει στους χρήστες για κάθε λέξη πληροφορίες μορφολογικές,

συντακτικές και σημασιολογικές είτε αυτή είναι λέξη σε συγκεκριμένη πρόταση κάθε φορά είτε εξετάζεται μεμονωμένα, δίνοντας τη δυνατότητα διατύπωσης μίας μεγάλης ποικιλίας ασκήσεων για περαιτέρω εμπέδωση, είτε σε κάθε μία από τις τρεις γλωσσολογικές διαστάσεις ξεχωριστά είτε και σε συνδυασμούς αυτών, όπως επίσης και οποιασδήποτε αξιοποίησης αυτών των δυνατοτήτων για τη διδασκαλία εν γένει, χωρίς να έχει ανάγκη να ανατρέξει σε άλλα μέσα και πηγές. Αφού:

1. το Νεοελληνικό Μορφολογικό Υπολογιστικό Λεξικό περιέχει, 1.1 όλους τους μορφολογικούς κανόνες για τη μορφολογική αναγνώριση των τύπων των κλιτών μερών του λόγου και των μη κλιτών μερών του λόγου και 2.1 τους αντίστοιχους κανόνες που διέπουν αυτούς,
2. το Νεοελληνικό Συντακτικό Υπολογιστικό Λεξικό περιέχει, όλους τους αλληλοεξαρτώμενους κανόνες που διέπουν τη μορφολογία και τη σύνταξη οι οποίοι ενεργοποιούνται και για τη διόρθωση σε περιπτώσεις λάθους μορφολογικού τύπου κλιτού μέρους του λόγου σε σχέση με τον συντακτικό ρόλο του.
3. το Νεοελληνικό Σημασιολογικό Υπολογιστικό Λεξικό, αποτελείται από τα επιμέρους Υπολογιστικά Λεξικά του, δηλαδή, το Υπολογιστικό Λογικό Λεξικό των Βασικών Εννοιών και το Υπολογιστικό Λεξικό των Συμφραζομένων της ΝΕΓ. Το μεν πρώτο από τα επιμέρους Υπολογιστικά Λεξικά, , δίνει τις αντίστοιχες αποδεκτές σημασιολογικά συναφείς λέξεις και εκφράσεις για την κάθε λέξη της υπό εξέταση πρότασης της ΝΕΓ. Με άλλα λόγια, η δυνατότητα αυτή καθιστά το ΒΝΠΙ Θησαυρό και Λεξικό συγχρόνως, επιτρέποντάς του σχηματισμούς αποδεκτών σημασιολογικών αντικαταστάσεων των λέξεων της πρότασης με συνώνυμες ή αντώνυμες τους λέξεις, αν υπάρχουν. Τέλος, το Υπολογιστικό Λεξικό των Συμφραζομένων της ΝΕΓ ελέγχει για κάθε λέξη τον αποδεκτό σημασιολογικό συνδυασμό των εννοιών της με άλλες ανάλογα με τον συντακτικό ρόλο τους, εντοπίζει τα λάθη, τα επεξηγεί, προτείνει αποδεκτές λύσεις και μετά από την κατάλληλη εντολή αποκαθιστά το λάθος.

Το λεξιλόγιο του συστήματος επεξεργασίας της ΝΕΓ που μελετάμε αποτελείται από λέξεις που αφενός αντιπροσωπεύουν όλες τις κατηγορίες των περιπτώσεων της μορφολογίας και των κανόνων της αφετέρου καλύπτουν ένα ευρύ εύχρηστο εννοιολογικό φάσμα λέξεων και εύχρηστων συντακτικών δομών προτάσεων της ΝΕΓ. Επίσης ακολουθείται ένα απλό και φιλικό στον χρήστη σύστημα διεπαφής και χρήσης των μορφολογικών, συντακτικών, σημασιολογικών, μορφοσυντακτικών και σημασιοσυντακτικών κανόνων. Έτσι εφαρμόζοντας τον Τεχνολόγο και τον

Γεννήτορα προτάσεων της ΝΕΓ σε ένα ανοικτό πειραματικό και δημιουργικό περιβάλλον ενισχύει την συνεργατική μάθηση.

Το σύστημα ΕΦΓ που μελετάμε μπορεί να λειτουργήσει διασυνδεδεμένο με οποιαδήποτε έκδοση των Windows και να λειτουργήσει αξιοποιώντας όλες τις υπηρεσίες του μέσα από εφαρμογές γενικής χρήσης, για παράδειγμα τις εφαρμογές του Microsoft Office, που υποστηρίζεται από την τεχνολογία OLE της Microsoft. Το σημαντικότερο είναι ότι το προτεινόμενο σύστημα επεξεργασίας της ΝΕΓ εφαρμοζόμενο για οποιονδήποτε διδακτικό σκοπό μέσα από εφαρμογές, όπως για παράδειγμα αυτές του Microsoft Office, δίνει τη δυνατότητα στον χρήστη δουλεύοντας σύγχρονα ή ασύγχρονα στην τάξη ή στο διαδίκτυο, να αποθηκεύονται τα αποτελέσματα των απαντήσεων του σε μία βάση δεδομένων για περαιτέρω αξιολόγηση από τον εκπαιδευτικό.

Συνοψίζοντας, για τους παραπάνω λόγους όπως τους εκθέσαμε αλλά και διότι το σύστημα που μελετήθηκε και παρουσιάστηκε για την επεξεργασία της ΝΕΓ είναι

1. αποτελεσματικό ως προς την ακρίβεια των αποτελεσμάτων
2. καινοτόμο στην ποικιλία των υπηρεσιών του,
3. ευέλικτο και απλό στη χρήση εφαρμογών του καθώς και
4. ταχύ στις διαδικασίες του, αφού οι αλγόριθμοι των Τεχνολόγων, των Γεννητόρων και των άλλων εφαρμογών των Διορθωτών και των Σημασιολογικών Αναλυτών είναι μικροί ευέλικτοι και ευκολοσυντήρητοι γραμμικής πολυπλοκότητας.
5. ευκολοσυντήρητο ως προς τις βάσεις δεδομένων του που αφορούν τις λέξεις, τις κλίσεις και τις έννοιες τους όπως και ως προς τις μήτρες κανόνων παραγωγής εύχρηστων συντακτικών δομών, τους δεσπόζοντες σημασιολογικούς συνδυασμούς εννοιών λέξεων καθώς και τους μορφοσυντακτικούς και σημασιοσυντακτικούς κανόνες ,
6. αυτόαρκες, αφού έχει κωδικοποιημένες ταξινομημένες και αρχειοθετημένες έτοιμες για οποιαδήποτε χρήση οποιαδήποτε πληροφορία και των τριών γλωσσολογικών διαστάσεων της γλώσσας στα αντίστοιχα υπολογιστικά λεξικά του.
7. εφαρμόσιμο στη εκπαίδευση σε οποιοδήποτε γνωστικό αντικείμενο. Διότι επισημαίνεται επίσης ότι πέραν του ΒΝΠ και των αλγορίθμων του, η μέθοδος επεξεργασίας των προτάσεων της ΝΕΓ αξιοποίησε και το πλεονέκτημα των μ.γραμματικών, που προτάθηκαν επί τούτου, με τους διακριτά ταξινομημένους κανόνες παραγωγής τους και τους χαρακτηριστικούς εκθέτες τους. Στην

προκειμένη περίπτωση οι κανόνες παραγωγής ήταν οι συντακτικοί κανόνες παραγωγής των συντακτικών δομών κυρίων και ακεραίων προτάσεων της ΝΕΓ και αποτέλεσαν τη βάση για την ανάπτυξη των Τεχνολόγων, των Γεννητόρων, του Διορθωτή, του Σημασιολογικού Αναλυτή των προτάσεων της ΝΕΓ και κατ' επέκταση και του Εργαλείου του Συλλαβισμού των λέξεων της ΝΕΓ. Με ανάλογους κανόνες και περαιτέρω ανάλογο εμπλουτισμό του περιεχομένου του ΒΝΠ το σύστημα μπορεί να λειτουργήσει όχι μόνο για οποιαδήποτε άλλη φυσική γλώσσα αλλά και για οποιοδήποτε άλλο περιεχόμενο άλλου πεδίου διδασκαλίας. Τέλος, η έρευνα που προτείνεται περαιτέρω αφορά κυρίως στην εμπλοκή των μεθόδων μάθησης βάθους, μάθησης μηχανής και τελικά των νευρωνικών δικτύων στην εφαρμογή και λειτουργία του ΒΝΠ και των αλγορίθμων του με απώτερο στόχο την αναβάθμιση της κατασκευής των ανοικτών εκπαιδευτικών λογισμικών που η μέθοδος αυτή που μελετήσαμε υποστηρίζει.

5. Βιβλιογραφικές Αναφορές

Δημοσιεύσεις σε Διεθνή Περιοδικά και Πρακτικά Συνεδρίων - Ανακοινώσεις σε Διεθνή Συνέδρια και Τεχνικές Αναφορές

1. **Baldzis S.D., Eumeridou E. & Kolalas S.** (2007). Efficient NLP Tools for the Generation and Recognition of Greek word Dominant Semantic Combinations. In: *Conference Proceedings II, 3rd Balkan Conference in Informatics: Research in Informatics and Information Technology*, 3-16.
2. **Baldzis, S.D., Kolalas S. & Eumeridou E.** (2005a). The Computational Modern Greek Morphological Lexicon - An Efficient and Comprehensive System for Morphological Analysis and Synthesis. *Literary and Linguistic Computing*, 19(4), 1-35, Oxford University Press.
3. **Baldzis S.D., Kolalas S. & Eumeridou E.** (2005b). Computational Morphological Lexicons: Current tools for Knowledge and Communication Management. In: *Conference Proceedings, 2nd Balkan Conference in Informatics*. Ohrid, 346 - 356.
4. **Baldzis, S.D., Savranidis, Ch. & Kolalas S.** (2001b). Written Modern Greek Sentences Corrector and Content Analyser based on Intelligent Computer Methods. In: *Proceedings of 1st Panhellenic Conference with International Participation on Human – Computer Interaction*. Patras: University of Patras, Advances in Human-Computer Interaction I, 182-191.
5. **Baldzis, S.D., Eumeridou E. & Kolalas S.** (2001c). A Complete and Comprehensive System for Modern Greek Language Processing Proposed as a Modern Greek Language Call Method Developer. *Literary and Linguistic Computing*, 17(4), 373-400, Oxford University Press.
6. **Baldzis, S.D.** (1999). *The Modern Greek Language Semantic Basis for Processing*. In Transscientific Semiotics: III-V, vol.11-1-3: 9-22.
7. **Baldzis, S. D.** (1998). *Generative and Transformational Grammars in Modern Greek Language Processing*, Technical Report No 267(b). Ioannina: Dept of Maths, University of Ioannina.
8. **Baldzis, S. D.** (1995b). Modern Greek Language Processing – Tool1: Template Grammars and Characteristic Exponents. Technical Report 249, Department of

Mathematics, University of Ioannina.

9. **Baldzis, S. D.** (1995b). *An Aspect in Formalization of the Structure of the Modern Greek Language with Grammars of the Chomsky Hierarchy – A Step for Modern Greek Language Processing*. Technical report No 245, Department of Mathematics, University of Ioannina.
10. **Μπαλτζής, Δ. Σ.** (2018). *Υπολογιστικά Λεξικά της Νεοελληνικής το μέσο για τη συνέχεια της γλώσσας στη σύγχρονη παγκοσμιοποιημένη ηλεκτρονική εποχή. Πρακτικά ΣΤ' Ευρωπαϊκού Συνεδρίου Νεοελληνικών Σπουδών της ΕΕΝΣ: Ο ελληνικός κόσμος σε περιόδους κρίσης και ανάκαμψης, 1204-2018*, Lund.
11. **Μπαλτζής, Δ. Σ., Βυζάς Θ. & Ευμοιρίδου Ε.** (2017). *Εναλλακτική αξιοποίηση της έννοιας της οντολογίας στη σημασιολογική κωδικοποίηση μίας φυσικής γλώσσας για αυτόματη επεξεργασία της - Παραδείγματα από τη Νέα Ελληνική Γλώσσα*. Conference Proceedings of the 6th International Conference on Modern Greek Language (ICGL 13), University of Westminster- London.
12. **Μπαλτζής, Δ. Σ., Βυζάς, Θ. & Ευμοιρίδου Ε.** (2015). *Η σημασιολογική συμπεριφορά του ρήματος ως καθοριστικός παράγοντας στην κατασκευή των σημασιολογικών κανόνων της νεοελληνικής για την αυτόματη επεξεργασία της. Πρακτικά Ε' Ευρωπαϊκού Συνεδρίου Νεοελληνικών Σπουδών της ΕΕΝΣ: Συνέχειες, ασυνέχειες, ρήξεις στον ελληνικό κόσμο (1204-2014): οικονομία, κοινωνία, ιστορία, λογοτεχνία. Θεσσαλονίκη: Τόμος Δ' 611-650, http://www.eens.org/EENS_congresses/2014/books/tomo4.pdf.*
13. **Μπαλτζής, Δ. Σ., Μάντζιου Ό. & Αποστόλου Κ.** (2010). *Μια επέκταση του καινοτόμου Ηλεκτρονικού Υπολογιστικού Λεξικού της Νέας Ελληνικής Γλώσσας - ΒΝΠ ως Διαδικτυακό Πολυμεσικό Εκπαιδευτικό Λογισμικό Φωνολογικής Ενημερότητας πρώτης σχολικής ηλικίας. Πρακτικά Δ' Ευρωπαϊκού Συνεδρίου ΝΣ της ΕΕΝΣ: Ταυτότητες στον ελληνικό κόσμο (από το 1204 έως σήμερα). Γρανάδα: Τόμος Ε' 649-888, http://www.eens.org/wordpress/wp-content/uploads/2012/05/Identities-in-the-Greek-world-Granada-2010-Congress-Vol_5-2011-isbn_978-960-99699-7-01.pdf.*
14. **Μπαλτζής, Δ. Σ., Κολαλάς Σ., Ευμοιρίδου Ε., Αλεξάκης Α. & Λούκα Δ.** (2006). *Ένα Καινοτόμο Ηλεκτρονικό Λεξικό της Νέας Ελληνικής Γλώσσας - Πρώτο μέρος: Μορφολογικό. Πρακτικά Γ' Ευρωπαϊκού Συνεδρίου ΝΣ της*

ΕΕΝΣ: *Ο ελληνικός κόσμος ανάμεσα στην εποχή του Διαφωτισμού και στον εικοστό αιώνα*. Βουκουρέστι: Τόμος Β' 341-354.

15. **Μπαλτζής, Δ. Σ., Σαβρανίδης Ν. Χ. & Κολαλάς, Σ. Α.** (2001a). Αυτόματη Επεξεργασία της Νεοελληνικής σε Ανοικτό Περιβάλλον Εκπαιδευτικού Λογισμικού. Πρακτικά του 5^{ου} Πανελληνίου Συνεδρίου Διδακτικής των Μαθηματικών και της Πληροφορικής στην Εκπαίδευση, Θεσσαλονίκη, σελ. 429-435.
16. **Μπαλτζής, Δ. Σ.** (1997). *Σπονδυλωτές Μητροειδείς Γραμματικές και Αυτόματη Επεξεργασία της Νεοελληνικής Γλώσσας*. Πρακτικά του 6ου Πανελληνίου Συνεδρίου Πληροφορικής. Αθήνα (ΕΠΥ): Τόμος Ι, 213-227.
17. **Μπαλτζής, Δ. Σ.** (1995c). *Μητροειδείς Γραμματικές και Αυτόματη Επεξεργασία της Νεοελληνικής*. Πρακτικά του 5ου Πανελληνίου Συνεδρίου Πληροφορικής. Αθήνα (ΕΠΥ): Τόμος ΙΙ, 699-712.

Διατριβές

18. **Κοντοδήμας Δημήτρης**. Μητροειδείς και Σπονδυλωτές Μητροειδείς Γραμματικές – Αποτελεσματικά Εργαλεία για την Αυτόματη Επεξεργασία της Συντακτικής Δομής της Νεοελληνικής Γλώσσας. Μεταπτυχιακή Διατριβή. Τμήμα Μαθηματικών, Πανεπιστήμιο Ιωαννίνων, 2015.
19. **Μπαλτζής, Δ. Σ.** Αυτόματη Επεξεργασία της Νεοελληνικής με Τροποποιημένης Εκδοχής των Γραμματικών της Ιεραρχίας Chomsky και με το Βασικό Νεοελληνικό Πολυλεξικό. Διδακτορική Διατριβή. Τμήμα Μαθηματικών, Πανεπιστήμιο Ιωαννίνων, 1993.
20. **Σικαλόπουλος Δημήτριος**. Νευρωνικά Δίκτυα και επεξεργασία Φυσικής Γλώσσας. Μεταπτυχιακή Διατριβή. Τμήμα Μαθηματικών, Πανεπιστήμιο Ιωαννίνων, 2019.
21. **Σταμούλης Βασίλειος**. Από τη Φιλοσοφική Λογική στη Βιομηχανική Λογική. Μεταπτυχιακή Διατριβή. Τμήμα Μαθηματικών, Πανεπιστήμιο Ιωαννίνων, 2019.

Εκπαιδευτικά Λογισμικά

22. **Μπαλτζής, Δ. Σ.** (2009) Λεξικράτης - Έκπαιδευτικό Λογισμικό, υλοποιήθηκε στο Εργαστήριο Επεξεργασίας Φυσικής Γλώσσας και Μαθηματικών Προβλημάτων του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων ([nlampr-lab.uoi.gr/Παραγόμενο Εκπαιδευτικό Λογισμικό](http://nlampr-lab.uoi.gr/Παραγόμενο_Εκπαιδευτικό_Λογισμικό)), - χρησιμοποιήθηκε ως

εργαλείο διδασκαλίας σε παιδιά με εν γένει Μαθησιακές Δυσκολίες, Δυσλεξία - από την ΕΛΕΠΑΠ Ιωαννίνων και χρηματοδοτήθηκε από το Ίδρυμα Μποδοσάκη.

23. **Μπαλτζής, Δ. Σ.** (2011) Αριθμοσθένης - Έκπαιδευτικό Λογισμικό, υλοποιήθηκε στο Εργαστήριο Επεξεργασίας Φυσικής Γλώσσας και Μαθηματικών Προβλημάτων του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων (nlampp-lab.uoi.gr/ Παραγόμενο Εκπαιδευτικό Λογισμικό), - χρησιμοποιήθηκε ως εργαλείο διδασκαλίας σε παιδιά με εν γένει Μαθησιακές Δυσκολίες - Δυσαριθμησία - από την ΕΛΕΠΑΠ Ιωαννίνων και χρηματοδοτήθηκε από το Ίδρυμα Μποδοσάκη.

Βιβλία - Διδακτικά Εγχειρίδια - URL

24. Jackendoff, R. (1990). *Semantic Structures*, Cambridge, MA: MIT Press.
25. **Jurafsky, D. & Martin, J.H.** (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New Jersey: Prentice Hall.
26. **Kumel, P.** (1979). *Formalization of natural language*. Berlin: Springer.
27. **Shieber, S. M .** (1986). *An Introduction to Unification-Based Approaches to Grammar*. Mass: Microtome Publishing.
28. **Μπαλτζής, Δ. Σ.** (2005). Εισαγωγή στην Επεξεργασία Φυσικής Γλώσσας - Σημειώσεις από τις Παραδόσεις, Πανεπιστημιακές Σημειώσεις, Πανεπιστήμιο Ιωαννίνων.
29. **Παιδαγωγικό Ινστιτούτο.** (1998). Γενικές Προδιαγραφές Εκπαιδευτικού Λογισμικού, Γραφείο Πιστοποίησης και Πολυμέσων.
30. **Παιδαγωγικό Ινστιτούτο.** (29/12/91019). <http://www.pi-schools.gr/software/dimotiko/>
31. Εκπαιδευτικό Λογισμικό – Τεχνολογία στην Εκπαίδευση. (29/12/91019). <https://economu.wordpress.com> > ενσωματώνοντας-τις-τπε-στη-διδασκαλία.

Συμπληρωματικές Βιβλιογραφικές Αναφορές

32. **Anagnostopoulou, D., Despiri, E, Lambropoulou, P., Mantzari, E., and Gavrilidou, M.** (2000). LEXIS – lexicographical Infrastructure: systematizing the data. Proceedings of COMPLEX 2000, Workshop on Computational Lexicography and Multimedia Dictionaries, Greece, pp. 31-4.
33. **Gavrilidou, M., Lambropoulou, P., Mantzari, E., and Rousou, S.** (1998).

- The morphological level of a computational lexicon. Proceedings of the Panhellenic Conference of New Information Technologies NIT'98, Greece, pp. 298-308.
34. **Kermanidis, K., Sgarbas, K., Fakotakis, N., and Kokkinakis, G.** (2000). A PC-PATR-based syntactic description of Modern Greek. *Literary and Linguistic Computing*, 15: 219-312.
 35. Papageorgiou, H., Prokopidis, P., Giouli, V., and Piperidis, P. (2000). A Unified POS Tagging Architecture and its Application to Greek, Athens: LREC, pp. 1455-63.
 36. PAROLE (2000). Available at: <http://www.ub.es/gilcub/SIMLE/simple.html>. 2/10/2000 and <https://www.wordreference.com/engr/parole/simle>. 5/1/2020
 37. **Petasis, G., Paliouras, G., Karkaletsis, V., Spyropoulos, C.D., and Androutsopoulos, I.** (1999). Using machine learning techniques for a part-of-speech tagging in the Greek Language. In Fotiadis, D.I. and Nikolopoulou, S. D. (eds), *Advances in Informatics*. World Scientific, Singapore, pp.273-81.
 38. **Petasis, G., Paliouras, G., Karkaletsis, V., Spyropoulos, C.D., and Androutsopoulos, I.** (2001), A Greek morphological lexicon and its exploitation by a Greek controlled language checker. 8th Panhellenic Conference on Informatics, Cyprus, Vol. I, pp. 80-9.
 39. **Sgarbas, K., Fakotakis, N., and Kokkinakis, G.** (1995). A PC-KIMMO-based morphological description of Modern Greek. *Literary and Linguistic Computing*, 10: 189-201.
 40. **Sgarbas, K., Fakotakis, N., and Kokkinakis, G.** (2000a). Two algorithms for incremental construction of direct acyclic word graphs. *International Journal on Artificial Intelligence Tools*, 4: 369-81.
 41. **Sgarbas, K., Fakotakis, N., and Kokkinakis, G.** (2000b). A straightforward approach to morphological analysis and synthesis, Proceedings COMPLEX 2000, Workshop on Computational Lexicography and Multimedia Dictionaries, Greece, pp. 31-4.
 42. Tambouratzis, G., and Caragiannis, G. (2001). Automatic corpora-based stemming in Greek. *Literary and Linguistic Computing*, 16(4): 445-66

43. **Thanopoulos, A., Fakotakis, N., and Kokkinakis, G.** (2000a). Automatic extraction of semantic relations from specialized corpora. Proceedings of COLING 2000, pp. 836-42.
44. **Thanopoulos, A., Fakotakis, N., and Kokkinakis, G.** (2000b). Automatic Extraction of Semantics Similarity of Words from Row Technical Texts. Athens: LREC, pp. 609-15.
45. **Βοστταντζόγλου, Θ.** (1962). Αντιλεξικόν ή Ονομαστικόν της Νεοελληνικής Γλώσσας. Αθήνα: Εκδόσεις Πατρίς.