

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ  
ΣΧΟΛΗ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΦΙΛΟΛΟΓΙΑΣ

**Αθανάσιος Ν. Καρασίμος**

Διδακτορική Διατριβή

**Υπολογιστική Επεξεργασία της  
Αλλομορφίας  
στην παραγωγή λέξεων της Ελληνικής**

Επόπτρια: καθ. Αγγελική Ράλλη

Μέλη Συμβουλευτική Επιτροπή  
Επίκ. καθ. Δημήτριος Παπαζαχαρίου  
Επίκ. καθ. Κυριάκος Σγάρμπας

Πάτρα 2011

**Επταμελής Εξεταστική Επιτροπή**

καθ. Αγγελική Ράλλη

καθ. Νικόλαος Φακωτάκης

αν. καθ. Ελένη Γαλιώτου

ερευνήτρια Β' Βαθμίδας Ιώ Μανωλέσσου

επικ. καθ. Γεώργιος Μαρκόπουλος

επικ. καθ. Δημήτριος Παπαζαχαρίου

επικ. καθ. Κυριάκος Σγάρμπα

Η υποστήριξη της διδακτορικής διατριβής έγινε την Τρίτη 12 Απριλίου 2011. Στη συγκεκριμένη μορφή έχουμε λάβει υπόψη τα σχόλια και τις παρατηρήσεις της επταμελούς εξεταστικής επιτροπής, την οποία ευχαριστούμε θερμά για τη συμβολή της.

# Δήλωση

Δηλώνω εδώ υπεύθυνα πως αυτή η διδακτορική διατριβή είναι έργο προσωπικής εκπόνησης και δεν περιέχει υλικό που να έχει κατατεθεί για την απόκτηση κάποιου διπλώματος οποιασδήποτε βαθμίδας σπουδών. Το έργο που αναφέρεται σε αυτή τη διδακτορική διατριβή έχει γίνει από εμένα, εκτός των σημείων όπου αναφέρονται οι παραπομπές σε έργα και έρευνες άλλων μέσα στο κείμενο.

*(Αθανάσιος Ν. Καρασίμος)*

*στους δικούς μου ανθρώπους  
& στους δικούς μου «ξένους»*

## ΟΙ ΔΙΚΟΙ ΜΟΥ ΞΕΝΟΙ

Στίχοι/ Μουσική: Χάρις Αλεξίου

Οι δικοί μου ξένοι οι πιο μακρινοί  
είναι αυτοί που ζουν κοντά μου.  
Τους κοιτάζω, τους αγγίζω, τους μιλώ,  
τους ανοίγω την καρδιά μου.

Μα ο καθένας ταξιδεύει μοναχός  
μεσ' στην άγνωστη ψυχή του.  
Ο καθένας στη δική του ξενιτιά  
πολεμάει για τη ζωή του.

Οι δικοί μου ξένοι  
είν' οι αγαπημένοι.

Οι δικοί μου άνθρωποι ζούνε μακριά  
κι από μακριά αγαπάνε.  
Έτσι μεγαλώνει ο κόσμος κι η καρδιά  
και θυμόμαστε όπου πάμε.

Μα ο καθένας ταξιδεύει μοναχός  
κι αδελφή ψυχή γυρεύει  
και στα κρύσταλλα κομμάτια της καρδιάς  
την αλήθεια του λαζεύει.

Οι δικοί μου άνθρωποι  
της καρδιάς μου οι τόποι.

# Συντομογραφίες

ελλ.	Ελληνικά
ΑΕ	Αρχαία Ελληνικά
ΚΝΕ	Κοινή Νεοελληνική
ολλ.	Ολλανδικά
γερμ.	Γερμανικά
λατ.	Λατινικά
τουρκ.	Τουρκικά
ιταλ.	Ιταλικά
αορ.	αόριστος
μελ.	μέλλοντας
ενεσ.	ενεστώτας
πρτ.	παρατατικός
πρκ.	παρακειμένος
υπερσ.	υπερσυντέλικος
εν.	ενικός
πλ.	πληθυντικός

# Ευχαριστίες

Φτάνοντας επιτέλους στην «Ιθάκη» – την ολοκλήρωση της διδακτορικής διατριβής – αναλογίζομαι αυτό το μακρινό και ιδιαίτερο ταξίδι που με οδήγησε αυτή η εργασία. Το ταξίδι αυτό είχε διάφορες στιγμές, διάφορες στάσεις, αλλά προπάντων είχε ιδιαίτερα συναπαντήματα με πολλούς, διαφορετικούς και ιδιαίτερους ανθρώπους. Ίσως η επικοινωνία και η αλληλεπίδραση μαζί τους να είναι το μεγαλύτερο όφελος και κέρδος από αυτή την πορεία και για αυτό αισθάνομαι την ανάγκη να εκφράσω τις ειλικρινείς ευχαριστίες μου για τη συμβολή τους στο παρόν εγχείρημα.

Από την ακαδημαϊκή κοινότητα, θέλω να εκφράσω ένα μεγάλο ευχαριστώ στην επόπτρια καθηγήτρια της συγκεκριμένης διατριβής και διευθύντρια του Εργαστηρίου Νεοελληνικών Διαλέκτων, *καθ. Αγγελική Ράλλη*, για τη συμπαράστασή της κατά τη διάρκεια όλων των ακαδημαϊκών μου σπουδών, για την υπομονή που έδειξε στις συναντήσεις μας και στα γραπτά μου και για τη συνεχή παρότρυνση για βελτίωση. Επιπλέον θέλω να εκφράσω τις θερμές ευχαριστίες και στα άλλα δύο μέλη της τριμελούς επιτροπής μου: τον *επικ. καθ. Δημήτριο Παπαζαχαρίου*, για τις πάντα χρήσιμες συμβουλές και παρατηρήσεις και κυρίως για την παρότρυνση να ασχοληθώ με τον τομέα της υπολογιστικής γλωσσολογίας, και τον *επικ. καθ. Κυριάκο Σγάρμπα* του Τμήμα Ηλεκτρολόγων Μηχανολόγων και Τεχνολογίας Ηλεκτρονικών Υπολογιστικών, για τις συμβουλές και την ενθάρρυνση για την προσπάθειά μου.

Αρωγοί στη διαδρομή του διδακτορικού υπήρξαν και οι καθηγητές του Τομέα της Γλωσσολογίας του Τμήματος Φιλολογίας Πανεπιστημίου Πατρών. Ευχαριστώ ειλικρινά τους *αν. καθ. Αργύρη Αρχάκη*, *επικ. καθ. Γιώργο Ξυδόπουλο* και *δρ. Χριστίνα Μανουηλίδου* για τις συμβουλές τους και την στήριξη τους, καθώς και για τον γενικότερο προβληματισμό σε ζητήματα γλώσσας. Παράλληλα θα ήθελα να ευχαριστήσω την *αν. καθ. Ελένη Γαλιώτου* για τις παρατηρήσεις και τα σχόλια της στη διαμόρφωση του κομματιού της υπολογιστικής μορφολογίας. Επιπροσθέτως τις θερμές μου ευχαριστίες θέλω να εκφράσω για τα λοιπά μέλη της επταμελούς, τον *καθ. Νικόλαο Φακωτάκη*, την *ερευνήτρια Β' βαθμίδος Ιώ Μανωλέσσου* και τον *επικ. καθ. Γεώργιο Μαρκόπουλο* για τα σχόλια και τις παρατηρήσεις τους.

Παράλληλα οφείλω ένα ιδιαίτερο ευχαριστώ στους ερευνητές του Istituto di Linguistica Computazionale 'Antonio Zampolli' του Centro Nazionale dello Ricerche της Πίζας και ειδικότερα στον υποδιευθυντή *dr. Vito Pirrelli*, για τις συμβουλές, τις

παρατηρήσεις, τη μεταλαμπάδευση των γνώσεών του και την βοήθεια στο σχεδιασμό του πειράματος AMIS. Για το ίδιο πείραμα θα ήθελα να ευχαριστώ και τους *dr. Felice Dell' Orletta* και *dr. Simonetta Montemagni* για καθοδήγηση τους, καθώς και τον *Kazuhiro Yoshida* για την πολύτιμη βοήθεια κατά την εκτέλεση του πειράματος. Επιπροσθέτως ευχαριστώ πολύ τους φίλους μου *Γιάννη Καραβίδα* και *Δημήτρη Βίτσιο* για τη σημαντική βοήθεια τους στο προγραμματιστικό κομμάτι του κώδικα του πειράματος. Επίσης ευχαριστώ την υποψήφια διδάκτορα *Βανέσσα Πετροπούλου* για τη συμπαράσταση και κατανόηση του υπολογιστικού γολγοθά της διατριβής, αλλά και τη συμβολή της στο πείραμα του *Linguistica*.

Βεβαίως, νιώθω την ανάγκη να ευχαριστήσω καθένα από τα παιδιά που συνεργαστήκαμε τόσα χρόνια τόσο στο Εργαστήριο Νεοελληνικών Διαλέκτων, όσο γενικότερα στον Τομέα και μοιραστήκαμε ιδιαίτερες και έντονες στιγμές. Θεωρώ ότι υπήρξα τυχερός που απέκτησα φίλους στον χώρο της γλωσσολογίας, καθότι η διαδρομή μιας διδακτορικής διατριβής είναι μια τρελή πορεία και αποκτά εξαιρετικό ενδιαφέρον να την ζεις με «ομοιοπαθείς» που κατανοούν αυτή την ιδιαίτερη εμπειρία. Σας ευχαριστώ όλους... για όλες τις απίθανες στιγμές!

Τέλος, δεν θα βρισκόμουν εδώ αν δεν είχα τους κοντινούς μου ανθρώπους, να με στηρίζουν, να μου δίνουν δύναμη και με ενισχύουν ποικιλοτρόπως και αδιαλείπτως. Θέλω να ευχαριστήσω εκ βάθους καρδιάς τους δικούς μου ανθρώπους, την *οικογένεια* μου, που χωρίς αυτήν πολλά από αυτά θα ήταν ανούσια και η ευτυχία μετρημένη. Το ίδιο και για τους δικούς μου «ξένους», τους *αγαπητούς φίλους της κλίμας*, που φρόντιζαν πάντα με τον τρόπο να μου δίνουν κουράγιο και να μου γεμίζουν τις μπαταρίες...

*Ένα ευχαριστώ μονάχα είναι λίγο...*



# Περιεχόμενα

<b>ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ</b> .....	<b>VI</b>
<b>ΕΥΧΑΡΙΣΤΙΕΣ</b> .....	<b>VII</b>
<b>ΠΕΡΙΕΧΟΜΕΝΑ</b> .....	<b>IX</b>
<b>INDEX ΠΙΝΑΚΩΝ ΚΑΙ ΕΙΚΟΝΩΝ</b> .....	<b>XII</b>
<b>1. ΕΙΣΑΓΩΓΗ</b> .....	<b>1</b>
1.1. ΟΡΙΟΘΕΤΗΣΗ ΘΕΜΑΤΟΣ .....	1
1.2. ΔΙΑΡΘΡΩΣΗ ΔΙΑΤΡΙΒΗΣ.....	3
<b>2. ΑΛΛΟΜΟΡΦΙΑ</b> .....	<b>7</b>
2.1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΛΛΟΜΟΡΦΙΑ .....	7
2.1.1. <i>Η Αλλομορφία ως «είδωλο» της γλωσσικής ιστορίας</i> .....	9
2.2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ ΤΗΣ ΑΛΛΟΜΟΡΦΙΑΣ .....	12
2.2.1. <i>(Μορφο-)Φωνολογικές προσεγγίσεις της Αλλομορφίας</i> .....	14
2.2.2. <i>Μορφολογικές προσεγγίσεις της Αλλομορφίας</i> .....	21
2.3. ΤΟ ΜΟΡΦΟΛΟΓΙΚΟ ΜΟΝΤΕΛΟ ΤΗΣ ΡΑΛΛΗ ΓΙΑ ΤΗΝ ΑΛΛΟΜΟΡΦΙΑ .....	34
2.4. ΚΑΘΟΡΙΣΜΟΣ ΤΟΥ ΘΕΩΡΗΤΙΚΟΥ ΠΛΑΙΣΙΟΥ ΤΗΣ ΑΛΛΟΜΟΡΦΙΑΣ .....	39
2.4.1. <i>Ορισμός της Αλλομορφίας</i> .....	39
2.4.1.1. <i>Είδη αλλομορφίας</i> .....	41
2.4.1.2. <i>Ψευδοαλλομορφα (vs. ελεύθερες παραλλαγές) και υποκατάσταση</i> .....	45
2.4.2. <i>Η θέση της Αλλομορφίας στο Λεξικό</i> .....	49
2.4.2.1. <i>Μορφολεξικοί κανόνες και κανόνες χρήσης αλλομορφίας</i> .....	56
2.4.3. <i>Η εμφάνιση των αλλομόρφων στην κλίση</i> .....	63
2.4.3.1. <i>Η εξάρτηση της αλλομορφικής συμπεριφοράς από την κλίση</i> .....	64
2.4.3.2. <i>Ονοματική αλλομορφία</i> .....	68
2.4.3.3. <i>Ρηματική αλλομορφία</i> .....	73
2.4.3.4. <i>Επιθετική (και Επιρρηματική) αλλομορφία</i> .....	77
2.4.4. <i>Η αρχή της αλλομορφικής συμπεριφοράς</i> .....	79
2.4.4.1. <i>Η εξαίρεση των ουσιαστικών τύπου –παιδί</i> .....	82
2.4.4.2. <i>Η εξαίρεση των ουσιαστικών τύπου –μπακάλη</i> .....	84
2.4.4.3. <i>Άλλες εξαιρέσεις</i> .....	87
2.5. <i>Η ΑΛΛΟΜΟΡΦΙΑ ΩΣ ΚΡΙΤΗΡΙΟ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ</i> .....	88
<b>3. ΠΑΡΑΓΩΓΗ ΚΑΙ ΑΛΛΟΜΟΡΦΙΑ</b> .....	<b>91</b>
3.1. <i>Η ΑΛΛΟΜΟΡΦΙΑ ΤΩΝ ΠΑΡΑΓΩΓΙΚΩΝ ΠΡΟΘΗΜΑΤΩΝ</i> .....	91
3.2. <i>Η ΑΛΛΟΜΟΡΦΙΑ ΤΩΝ ΠΑΡΑΓΩΓΙΚΩΝ ΕΠΙΘΗΜΑΤΩΝ</i> .....	94

3.3. Ο ΣΥΝΔΕΣΜΟΣ ΤΗΣ ΑΛΛΟΜΟΡΦΙΑΣ: ΑΠΟ ΤΑ ΑΛΛΟΜΟΡΦΑ ΤΗΣ ΚΛΙΣΗΣ ΣΤΑ ΑΛΛΟΜΟΡΦΑ ΤΗΣ ΠΑΡΑΓΩΓΗΣ .....	100
3.3.1. Ο περιορισμός της επιλογής μοναδικού αλλομόρφου .....	102
3.4. ΟΙ ΠΕΡΙΠΤΩΣΕΙΣ ΤΩΝ ΑΛΛΟΘΗΜΑΤΩΝ .....	110
<b>4. ΥΠΟΛΟΓΙΣΤΙΚΗ ΜΟΡΦΟΛΟΓΙΚΗ ΑΝΑΛΥΣΗ .....</b>	<b>119</b>
4.1. ΕΙΣΑΓΩΓΗ.....	119
4.2. ΣΥΝΤΟΜΗ ΙΣΤΟΡΙΑ ΤΗΣ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΜΟΡΦΟΛΟΓΙΑΣ.....	121
4.3. ΑΥΤΟΜΑΤΑ ΠΕΠΕΡΑΣΜΕΝΩΝ-ΚΑΤΑΣΤΑΣΕΩΝ .....	123
4.3.1. <i>Pc-Kimmo</i> : Μορφολογικός αναλυτής Δύο-Επιπέδων.....	127
4.3.1.1. Τα βασικά του Ελληνικού Pc-Kimmo (Σγάρμπας 1996).....	127
4.3.1.2. Κριτική στο μοντέλο Pc-Kimmo για τα Ελληνικά.....	128
4.4. ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΤΗΣ ΜΟΡΦΟΛΟΓΙΑΣ .....	132
4.4.1. <i>Μη-Εποπτευόμενη Μορφολογική Μάθηση</i> .....	133
4.4.1.1. Εισαγωγή .....	133
4.4.1.2. Διάφορες προσεγγίσεις Μη-Εποπτευόμενης Μορφολογικής Μάθησης .....	135
4.4.1.3. Η προσέγγιση του Goldsmith (2001).....	136
4.4.1.3.1. Δημιουργία Υποψηφίων Φύλλων σώματος .....	138
4.4.1.3.2. Αξιολόγηση Υποψηφίων (και τα διαγλωσσικά πειράματα) .....	139
4.4.1.4. Κριτική μοντέλων Μη-Εποπτευόμενης Μορφολογικής Μάθησης.....	143
4.4.2. <i>Εποπτευόμενη Μορφολογική Μάθηση</i> .....	145
4.4.2.1. Μέγιστη Εντροπία .....	146
4.4.2.2. Μνημονική Μάθηση.....	150
4.4.2.3. Μάθηση Μετασχηματιστικού Τύπου.....	153
<b>5. ΤΑ ΠΕΙΡΑΜΑΤΑ ΜΕ ΤΟΥΣ ΜΟΡΦΟΛΟΓΙΚΟΥΣ ΑΝΑΛΥΤΕΣ.....</b>	<b>154</b>
5.1. ΤΟ ΠΕΙΡΑΜΑ ΤΟΥ LINGUISTICA ΣΤΑ ΕΛΛΗΝΙΚΑ .....	154
5.1.1. <i>Σχετικά με το Linguistica</i> .....	155
5.1.2. <i>Εντοπίζοντας την Αλλομορφία με το Linguistica</i> .....	156
5.1.3. <i>Τα σώματα κειμένου δοκιμής</i> .....	157
5.1.4. <i>Αποτελέσματα</i> .....	158
5.1.4.1. Προθηματοποίηση .....	161
5.1.4.2. Επιθηματοποίηση .....	162
5.1.4.3. Θέματα/ Ρίζες .....	164
5.1.4.4. Σύνθετα και δεσμευμένα θέματα .....	165
5.1.5. <i>Σύντομος σχολιασμός του πειράματος</i> .....	167
5.2. ΤΟ ΠΕΙΡΑΜΑ ΤΟΥ AMIS ΓΙΑ ΤΗΝ ΟΝΟΜΑΤΙΚΗ ΚΛΙΣΗ ΤΗΣ ΚΝΕ .....	168
5.2.1. <i>Σχετικά με το AMIS</i> .....	169
5.2.2. <i>Αναπαράσταση γλωσσολογικών πληροφοριών σε σύνολα λεξικών χαρακτηριστικών</i> .....	169
5.2.2.1. Μία στατιστική αναπαράσταση της Ονοματικής Αλλομορφίας .....	171
5.2.3. <i>Επιλογή Χαρακτηριστικών</i> .....	173
5.2.4. <i>Δοκιμή Ονοματικού Μοντέλου</i> .....	175
5.2.4.1. Αξιολόγηση επιρροής βασικών χαρακτηριστικών .....	180

5.2.4.2. ΑλλοMantIS – Ένας αναλυτής AMIS πρόβλεψης της αλλομορφίας.....	184
5.2.5. Συμπερασματικά σχόλια.....	187
<b>6. ΥΠΟΛΟΓΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΤΗΣ ΑΛΛΟΜΟΡΦΙΑΣ.....</b>	<b>190</b>
6.1. ΠΡΟΗΓΟΥΜΕΝΕΣ ΕΡΕΥΝΕΣ ΚΑΙ ΠΡΟΤΑΣΕΙΣ.....	190
6.2. Η ΠΡΟΒΛΗΜΑΤΙΚΗ ΤΗΣ ΘΕΜΑΤΟΚΕΝΤΡΙΚΗΣ Η ΠΡΟΣΦΥΜΑΤΟΚΕΝΤΡΙΚΗΣ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΠΡΟΣΕΓΓΙΣΗΣ ΤΗΣ ΑΛΛΟΜΟΡΦΙΑΣ.....	198
6.3. ΠΡΟΤΕΙΝΟΜΕΝΟ ΠΛΑΙΣΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΗΣ ΑΛΛΟΜΟΡΦΙΑΣ.....	200
6.3.1. <i>Ανενεργοί Γλωσσολογικοί Κανόνες – Η πρόταση των κανόνων αλλομορφικής μετατροπής</i> .....	202
6.3.1.1. Ονοματικοί Κανόνες Αλλομορφικής Μετατροπής.....	205
6.3.1.2. Ρηματικοί Κανόνες Αλλομορφικής Μετατροπής.....	208
6.3.1.3. Προθηματικοί Κανόνες Αλλομορφικής Μετατροπής.....	213
6.3.1.4. Φίλτρα Μορφολογικού Περιβάλλοντος.....	219
6.3.2. <i>Προβλεψιμότητα και Κανονικότητα στην Αλλομορφία – Η πρόταση των χαρακτηριστικών κατηγοριοποίησης σε Μέγιστη Εντροπία</i> .....	228
6.3.3. <i>Συνδυαστικότητα μοντέλων προσέγγισης της αλλομορφίας</i> .....	237
<b>7. ΣΥΜΠΕΡΑΣΜΑΤΑ.....</b>	<b>246</b>
7.1. ΓΕΝΙΚΗ ΕΠΙΣΚΟΠΗΣΗ.....	246
7.1.1. <i>Η αλλομορφία ως σύγχρονο μορφολογικό φαινόμενο</i> .....	246
7.1.2. <i>Μη-Εποπτευόμενα Μοντέλα Μορφολογικής Μάθησης για την Αλλομορφία</i> .....	247
7.1.3. <i>Εποπτευόμενη Μορφολογική Μάθηση για τη Μορφολογική Ανάλυση</i> .....	249
7.1.4. <i>Η προβληματική για την υπολογιστική επεξεργασία της Αλλομορφίας και τα συνδυαστικά μοντέλα</i> .....	251
7.2. ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ.....	253
7.2.1. <i>Επαναληπτική μάθηση και διεύρυνση του ΑλλοMantIS</i> .....	253
7.2.2. <i>Υλοποίηση του μοντέλου Μορφολογικής Μάθησης της Αλλομορφίας</i> .....	253
7.2.3. <i>Αυξημένη κάλυψη του φαινομένου στην κλίση και τη σύνθεση</i> .....	254
7.2.4. <i>Πιθανός συνδυασμός για αυτόματη μορφολογική ανάλυση</i> .....	255
7.3. ΠΕΡΙΛΗΨΗ.....	256
<b>ΞΕΝΟΓΛΩΣΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑ.....</b>	<b>258</b>
<b>ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ.....</b>	<b>284</b>
<b>ΠΑΡΑΡΤΗΜΑ.....</b>	<b>292</b>
ΕΥΡΕΤΗΡΙΟ ΟΡΩΝ.....	292
APPENDIX I.....	296
1. Πίνακας παραγωγικών ρηματικών επιθημάτων.....	296
2. Πίνακας παραγωγικών ονοματικών επιθημάτων.....	296

# Index Πινάκων και Εικόνων

Πίνακας 1: Το Σχήμα Συνολικής Κατανομής για τον μέγιστο αριθμό θεμάτων – αλλομόρφων ενός ρήματος των Ιταλικών. Ως S καθορίζεται το βασικό προκαθορισμένο θέμα του ρήματος (Pirrelli & Battista 2000a: 330).	30
Πίνακας 2: Το μορφολογικό συνεχές των διαφόρων αλλαγών σε ένα μόρφημα.	48
Πίνακας 3: Δείγμα λημμάτων από το Σταθερό Λεξικό (Ράλλη 2005).	51
Πίνακας 4: Το παράδειγμα FINIRE από τους Pirrelli & Battista (2000a: 15), όπου το S <sub>x</sub> είναι ο ενδείκτης του κλιτικού παραδείγματος.	53
Πίνακας 5: Μερικώς εμπλουτισμένο δείγμα ονοματικών λημμάτων του λεξικού.	55
Πίνακας 6: Μερικώς εμπλουτισμένο δείγμα ρηματικών λημμάτων του λεξικού.	56
Πίνακας 7: Οι τέσσερις ομάδες διαφορετικών πληροφοριών για τα ουσιαστικά όπως καταχωρούνται στο φίλτρο μορφολογικού περιβάλλοντος.	62
Πίνακας 8: Οι ονοματικές κλιτικές τάξεις με τις αλλομορφικές πληροφορίες.	72
Πίνακας 9: Οι κλιτικές τάξεις του ρηματικού συστήματος κατά τη Ράλλη (2005) και τη δική μας τροποποίηση.	77
Πίνακας 10: Οι συμπληρωματικές κλιτικές τάξεις των επιθέτων της KNE.	78
Πίνακας 11: Γενικός πίνακας του περιορισμού επιλογής μοναδικού αλλομόρφου για ονοματικά θέματα και επιθήματα.	109
Πίνακας 12: Τα ποσοστά επίδοσης του Linguistica για Αγγλικά και Γαλλικά από τη δοκιμή του Goldsmith (2001).	141
Πίνακας 13: Το τοπ-3 των φύλλων σώματος από το πείραμα του Goldsmith για τα Αγγλικά, Γαλλικά και Ισπανικά (2001: 179-182).	142
Πίνακας 14: Τα συγκεντρωτικά αποτελέσματα των τριών σωμάτων κειμένων από το πείραμα του LINGUISTICA.	158
Πίνακας 15: Η δεκάδα υψηλότερης εμφάνισης επιθημάτων στο ΣΚ2+.	161
Πίνακας 16: Δείγμα χαρακτηριστικών με κωδικοποίηση για Μοντέλο Εντροπίας.	170
Πίνακας 17: Στατιστικά στοιχεία από το corpus του πειράματος του AMIS.	175
Πίνακας 18: Συγκεντρωτικός πίνακας των 10 υψηλότερων και χαμηλότερων θετικών και αρνητικών βαρών για κάθε αλλομορφική τάξη.	179
Πίνακας 19: Συγκεντρωτικός πίνακας με τα βασικά χαρακτηριστικά και τα βάρη του ανάλογα με την αλλομορφική τάξη.	183
Πίνακας 20: Επιτρεπόμενες ετεροιώσεις του συστήματος για τα Φινλανδικά (Virpioja & Kohonen 2009: 4).	197
Εικόνα 1: Η διεπαφή του προγράμματος LINGUISTICA.	155
Εικόνα 2: Δείγμα από τα φύλλα σώματος επιθημάτων του ΣΚ2.	162
Εικόνα 3: Φύλλα σώματος από κλιτικά και παραγωγικά επιθήματα (ΣΚ2).	163
Εικόνα 4: Φύλλα σώματος από ουσιαστικά, ρήματα και επίθετα.	165

Εικόνα 5: Φύλλα σώματος από δεσμευμένα θέματα .....	167
Εικόνα 6: Η τροποποίηση του αρχείου event από απλό txt (πάνω μισό) στην κατάλληλη μορφή μέσω της απαραίτητης κωδικοποίησης με το MatLab (κάτω μισό) ( <a href="http://www.mathworks.com/products/matlab/">http://www.mathworks.com/products/matlab/</a> ) .....	177



*“Ἐν ἀρχῇ ἦν ὁ Λόγος, καὶ ὁ Λόγος ἦν πρὸς τὸν Θεόν, καὶ Θεὸς ἦν ὁ Λόγος”*

Κατὰ Ἰωάννη 1,1

# 1. Εισαγωγή

## 1.1. Οριοθέτηση Θέματος

Στην παρούσα διατριβή εξετάζεται το μορφολογικό φαινόμενο της αλλομορφίας στην παραγωγή της Ελληνικής γλώσσας, παρουσιάζεται η θεωρητική ανάλυσή του και προτείνεται ένα υπολογιστικό μοντέλο επεξεργασίας του. Συγκεκριμένα, διερευνώνται οι συστηματικές αλλαγές των μορφημάτων που προκύπτουν αυθαίρετα και απρόβλεπτα σε θεματικό και προσφυματικό επίπεδο κατά τη διαδικασία της παραγωγής και αναλύονται συγκριτικά οι αλλομορφικές συμπεριφορές θεμάτων και επιθημάτων για την εξαγωγή κοινών συμπερασμάτων.

Η βάση δεδομένων για την ανάλυση του φαινομένου δεν αποτελείται όμως μόνο από λέξεις παράγωγες, αλλά και κλιτές και σύνθετες, λόγω της αυξημένης διεισδυτικότητας του φαινομένου. Η επιλογή της σύνδεσης της αλλομορφίας της παραγωγής με αυτή της κλίσης έγινε συνειδητά και καθόλου τυχαία, καθότι η πλειονότητα των μορφολογικών αλλαγών συντελέστηκαν στο επίπεδο της κλίσης με αποτέλεσμα την εκτεταμένη εμφάνιση αλλομόρφων. Αναλύεται ένα φαινόμενο που συμμετέχει ενεργά σε όλες τις διαδικασίες σχηματισμού λέξεων, στο οποίο δεν έχει δοθεί η απαραίτητη προσοχή, ενώ συχνά γίνεται λανθασμένη χρήση του όρου για τον χαρακτηρισμό φωνολογικών αλλαγών.

Η αλλομορφία αποτελεί μία πρόκληση τόσο σε θεωρητικό επίπεδο, αφού αποτυπώνει τις αλλαγές στο μορφολογικό σύστημα μίας γλώσσας μέσω της γλωσσικής

εξέλιξης, αλλά και σε υπολογιστικό επίπεδο, καθώς αποτελεί έναν από τους δυσκολότερους σκοπέλους που καλείται να ξεπεράσει ένας μορφολογικός αναλυτής. Η έρευνα για την αλλομορφία κλείνει σχεδόν τριάντα χρόνια, ωστόσο δεν αποτέλεσε ποτέ το επίκεντρο μελέτης στο πλαίσιο των σύγχρονων μορφολογικών θεωριών (για τη διαπίστωση, βλ. Lieber 1982, Carstairs 1987, Ralli 1988, 2000, 2005, Maiden 1992, 2004, Booij 1997, Drachman 2006, μεταξύ άλλων). Το υπό εξέταση φαινόμενο αποτελεί πρόσφορο έδαφος για την υποστήριξη της αυτονομίας της μορφολογίας, καθώς και για την αλληλένδετη ροή δεδομένων και συμπεριφορών μεταξύ των διαδικασιών σχηματισμού λέξεων, όπως υποστηρίζουν οι Booij (1997) και Ralli (2006). Στη διεθνή βιβλιογραφία το όλο ζήτημα προσεγγίζεται από τους μελετητές είτε από φωνολογικής απόψεως (Drachman, Kager & Malikouti-Drachman 1995, Mascarò 1996a, 1996b, 2007, Thornton 1997, Drachman 2006), είτε από μορφολογικής απόψεως (Lieber 1982, Carstairs 1987, Maiden 1992, 2004, Booij 1997, Ralli 2005, 2006), πάντα στην προσπάθεια να ενταχθεί σε συγκεκριμένο θεωρητικό πλαίσιο.

Στόχος της παρούσας έρευνας είναι να καλύψει το φαινόμενο της αλλομορφίας σε συγχρονική θεωρητική βάση ορίζοντάς την, θέτοντας τα όρια της, αναλύοντας τη συμπεριφορά των αλλομόρφων και παρουσιάζοντας σε υπολογιστική βάση ρεαλιστικές και εφαρμόσιμες τακτικές για την υπολογιστική επεξεργασία. Σε θεωρητικό επίπεδο ασχολούμεθα με τις περιπτώσεις της καθαρής αλλομορφίας, αποκλείοντας όλες τις υπόλοιπες που δεν πληρούν τις προϋποθέσεις του ορισμού της Lieber (1982). Αντίστοιχα, σε υπολογιστικό επίπεδο παρουσιάζουμε μία εφικτή λύση δομημένη σε μοντέλα που δεν είναι πολύπλοκα (βλ. Koskeniemi 1983, Σγάρμπας 1996, Μαρκόπουλος 1998) και βασίζονται σε στρατηγικές μηχανικής μάθησης της μορφολογίας (Jurafsky & Martin 2000, Goldsmith 2001, Schone & Jurafsky 2001, Roark & Sproat 2007). Οι παρατηρήσεις για την κρίση των μοντέλων είναι βασισμένες σε μορφολογικά κριτήρια και στην υπολογιστική λειτουργικότητά τους.

Οι μεμονωμένες προσπάθειες αντιμετώπισης του ζητήματος της αλλομορφίας (Krieger, Pinker & Nerbonne 1993, Goldsmith 2001, Handl *et al.* 2009, Kohonen, Virpioja & Klami 2009, Virpioja & Kohonen 2009) προσέγγισαν το πρόβλημα με προσοχή χωρίς να αποφύγουν τις κακοτοπιές των θεωρητικών προβλημάτων που προκύπτουν από την αλλομορφία. Συνήθως, αντιμετώπισαν με επιτυχία τα φωνολογικά αλλόμορφα, ενώ συχνά απέκλεισαν ομάδες μορφημάτων και αλλομόρφων από την ανάλυση λόγω αδυναμίας επεξεργασίας. Τα θετικά βήματα και πλεονεκτήματα κάθε



μοντέλου τα χρησιμοποιούμε συνδυαστικά με άμεσο στόχο την αποτελεσματική και καθολική επεξεργασία της αλλομορφίας. Κάθε προσπάθεια υλοποίησης ενός υπολογιστικού μοντέλου έχει κάποια αδύναμα σημεία, τα οποία προσπαθούμε να καλύψουμε με παραπλήσιες και συναφείς στρατηγικές.

## 1.2. Διάρθρωση Διατριβής

Η παρούσα έρευνα διαρθρώνεται σε έξι μέρη. Στο αρχικό μέρος (Κεφάλαιο 2<sup>ο</sup>), μετά την αναλυτική αναφορά στον ορισμό και τις θεωρητικές προσεγγίσεις της αλλομορφίας, δίνουμε τη δική μας προσέγγιση στο ζήτημα. Στο δεύτερο μέρος (Κεφάλαιο 3<sup>ο</sup>), παρατίθενται οι θεωρητικές προσεγγίσεις της διαδικασίας της παραγωγής και η ανάλυση της αλλομορφίας σε προθήματα, επιθήματα και θέματα. Στο τρίτο μέρος (Κεφάλαιο 4<sup>ο</sup>), αναφέρονται τα κυρίαρχα υπολογιστικά μοντέλα, ώστε να γίνει κατανοητό για ποιο λόγο επιλέγονται αυτά του επόμενου μέρους (Κεφάλαιο 5<sup>ο</sup>), για την τέλεση των δύο κεντρικών πειραμάτων αναφορικά με την επεξεργασία της αλλομορφίας. Στο πέμπτο μέρος (Κεφάλαιο 6<sup>ο</sup>), πραγματοποιείται η ανάλυση με βάση τα αποτελέσματα των πειραμάτων για τη δόμηση της υπολογιστικής ανάλυσης, ενώ στο τελευταίο μέρος (Κεφάλαιο 7<sup>ο</sup>), επιχειρείται μία σύντομη εξαγωγή συμπερασμάτων και αναφορά στη μελλοντική έρευνα.

Πιο συγκεκριμένα, στο 2<sup>ο</sup> Κεφάλαιο παρατίθεται μία μικρή εισαγωγή στην αλλομορφία (2.1), καθώς και η συσχέτισή της με την εξέλιξη της γλώσσας (2.1.1), ενώ στη βιβλιογραφική επισκόπηση παρουσιάζονται επιλεκτικά οι φωνολογικές (2.2.1) και οι μορφολογικές (2.2.2) προσεγγίσεις. Ακολούθως, παρουσιάζεται το θεωρητικό μοντέλο της Ράλλη (2.3), καθότι υιοθετούμε τις βασικές αρχές του για τη θεωρητική μελέτη της αλλομορφίας. Στη συνέχεια, παρουσιάζουμε το προτεινόμενο θεωρητικό πλαίσιο για την επεξεργασία της αλλομορφίας, όπου ορίζουμε αναλυτικά το φαινόμενο (2.4.1) και παρουσιάζουμε την ένταξη της Αλλομορφίας στο Λεξικό (2.4.2). Παράλληλα περιγράφουμε την αλλομορφική συμπεριφορά των θεμάτων ανά κλιτική τάξη του μοντέλου της Ράλλη (2.4.3) και κάποιων περιορισμών που διέπουν τα αλλόμορφα στη διαδικασία της κλίσης. Κλείνοντας το κεφάλαιο, υποστηρίζουμε ότι η αλλομορφία αποτελεί κριτήριο κατηγοριοποίησης (2.5).

Ακολούθως, στο 3<sup>ο</sup> Κεφάλαιο παρουσιάζονται οι αλλομορφικές συμπεριφορές των παραγωγικών προθημάτων (3.1) και επιθημάτων (3.2). Ειδικότερα, διατυπώνεται ο περιορισμός της επιλογής μοναδικού αλλομόρφου (3.3.2), σύμφωνα με τον οποίο μόνο

ένα αλλόμορφο κάθε μορφήματος συμμετέχει στην παραγωγή. Τέλος, ασχολούμαστε με τις περιπτώσεις των αλλοθημάτων (3.4).

Στο 4<sup>ο</sup> Κεφάλαιο ύστερα από την εισαγωγή και τη σύντομη ιστορία της Υπολογιστικής μορφολογίας (4.1-4.2), παρουσιάζονται τα αυτόματα πεπερασμένων καταστάσεων και ειδικότερα η υλοποίηση του PC-KIMMO στα Ελληνικά (4.3.1), την οποία ακολουθεί μία γενικότερη κριτική (4.3.1.2). Στη συνέχεια γίνεται αναφορά για τη Μηχανική Μάθηση της Μορφολογίας με τους δύο βασικούς τομείς της, τη Μη-Εποπτευόμενη Μορφολογική Μάθηση (4.4.1) και ειδικότερα την κυρίαρχη προσέγγιση του Goldsmith (2001) (4.4.1.3) που ακολουθείται από την κριτική της συγκεκριμένης προσέγγισης (4.4.1.4). Κλείνοντας το κεφάλαιο, παρουσιάζονται οι τρεις βασικές διαφοροποιήσεις της Εποπτευόμενης Μορφολογικής Μάθησης, η Μέγιστη Εντροπία (4.4.2.1), η Μνημονική Μάθηση (4.4.2.2) και τέλος, η Μάθηση Μετασχηματιστικού Τύπου (4.4.2.3).

Στο 5<sup>ο</sup> Κεφάλαιο παρουσιάζονται τα δύο πειράματα που διεξήχθησαν για την ανάλυση της αλλομορφίας. Αρχικά δοκιμάστηκε το πρόγραμμα του LINGUISTICA (5.1), που αποτελεί υλοποίηση της προσέγγισης του Goldsmith (2001). Παρουσιάζεται το πρόγραμμα, τα σώματα κειμένων, ενώ γίνεται σχολαστική αναφορά στα αποτελέσματα (5.1.4) για κάθε ομάδα μορφημάτων και καταλήγουμε με σύντομο σχολιασμό του πειράματος (5.1.5). Στη συνέχεια, γίνεται η παρουσίαση του δεύτερου πειράματος με το AMIS (5.2), το οποίο τροποποιήθηκε για να δοκιμαστεί σε ονοματικό μοντέλο της αλλομορφίας (5.2.4) και δημιουργήθηκε ο ΑλλοMantIS για την πρόβλεψη της αλλομορφίας ονοματικών αλλομόρφων (5.2.4.2).

Το 6<sup>ο</sup> Κεφάλαιο περιέχει τις προτάσεις μας για την υπολογιστική επεξεργασία της αλλομορφίας με βάση τα αποτελέσματα των πειραμάτων. Μετά από μία σύντομη ανασκόπηση των άλλων υπολογιστικών προσπαθειών για την αντιμετώπιση της αλλομορφίας (6.1), αναλύεται η λογική προσέγγισης της προτεινόμενης υλοποίησής μας (6.3). Παρουσιάζεται αρχικά η πρόταση των ανενεργών κανόνων που κωδικοποιούνται σε αλλο-κανόνες (6.3.1) και ακολούθως η εναλλακτική πρόταση με το μοντέλο των χαρακτηριστικών (6.3.2) που στηρίζεται στην κανονικότητα και προβλεψιμότητα της αλλομορφίας για να μαντέψει τις αλλομορφικές αλλαγές. Στο τέλος του κεφαλαίου γίνεται προσπάθεια συνδυασμού των δύο μοντέλων (6.3.3) σε μία πιο αποτελεσματική προσέγγιση.

Κλείνοντας, στο 7<sup>ο</sup> Κεφάλαιο παρατίθενται τα συνολικά συμπεράσματα, τα οποία προκύπτουν από την παρούσα εργασία, σχετικά με τη θεωρητική προσέγγιση της αλλομορφίας (7.1.1), τα μοντέλα μορφολογικής μάθησης (7.1.2-3) και τη γενικότερη προβληματική της υπολογιστικής αλλομορφικής ανάλυσης στα προτεινόμενα συνδυαστικά μοντέλα. Τέλος, επιχειρείται μία αναφορά σε μελλοντικές εργασίες, όπως στην ανάλυση της αλλομορφίας στη διαδικασία της σύνθεσης και στην επανεκπαίδευση και περαιτέρω δοκιμή των υπολογιστικών μοντέλων με βάση τα αποτελέσματα των πειραμάτων της παρούσας διατριβής.



“Allomorphy is a marginal synchronic junk pile” Maiden (1992)

“The motto ‘minimize allomorphy’ remains today’s watchword” Goldsmith (1995)

## 2. Αλλομορφία

### 2.1. Εισαγωγή στην Αλλομορφία

Όταν παρατηρήσουμε την εμφάνιση των μορφημάτων στις διάφορες λέξεις της Ελληνικής γλώσσας διαπιστώνουμε ότι στα περισσότερα μορφήματα η μορφή ποικίλλει ανάλογα με το περιβάλλον στο οποίο εμφανίζονται. Οι διαφορετικές μορφές του ίδιου μορφήματος ονομάζονται **αλλόμορφα** (allomorphs). Αντίστοιχες περιπτώσεις έχουμε στη φωνολογία, όπου οι διαφορετικές εκφορές του ίδιου φωνήματος λέγονται **αλλόφωνα** (allophones). Όπως τα αλλόφωνα, έτσι και τα αλλόμορφα βρίσκονται σε συμπληρωματική κατανομή στο εσωτερικό των μορφολογικών σχηματισμών. Από τις πρώτες ερμηνείες της εμφάνισής τους ήταν ότι οι αλλαγές του μορφήματος προέκυψαν από την εφαρμογή συγκεκριμένων συγχρονικών φωνολογικών κανόνων. Για να καταστεί σαφές, δίνονται τα παραδείγματα (1.) του προθήματος εν-:

- |      |               |            |                                   |
|------|---------------|------------|-----------------------------------|
| (1.) | α. ενέχω      | /enexo/    |                                   |
|      | β. εντάσσομαι | /endosome/ | [t → d/ n_]¹                      |
|      | γ. εμπαθής    | /embaθis/  | [(i.) n → m/ _p, (ii.) p → b/ m_] |

¹ Στο συγκεκριμένο παράδειγμα ο συνδυασμός του παραγωγικού προθήματος με το ρηματικό θέμα επιφέρει φωνολογικές αλλαγές στο θέμα, όπου το αρχικό φώνημα που είναι το άηχο και κλειστό σύμφωνο /t/ λόγω του κανόνα της αφομοίωσης ηχηροποιείται, επειδή προηγείται το ηχηρό σύμφωνο /n/.

δ. ελλιμενίζομαι	/elimenizome/	[(i.) n → l/ _l, (ii.) l → Ø/ _l]
ε. έρρινος	/erinos/	[(i.) n → r/ _r, (ii.) r → Ø/ _r]

Στα παραπάνω παραδείγματα, το παραγωγικό πρόθημα *συν-* κατά τη διαδικασία προθηματοποίησης παραμένει ίδιο (1.α-β), μεταβάλλεται (1.γ) ή αποβάλλεται (1.δ-ε) λόγω εφαρμογής γενικών κανόνων της φωνολογίας που συντελούνται στα όρια μορφημάτων. Παρατηρείται ότι το /n/ εξαρτάται από το αρχικό φώνημα του συνδυαζόμενου θέματος. Συνεπώς, σε κάποιες περιπτώσεις διατηρείται, σε κάποιες άλλες μεταβάλλεται με τον φωνολογικό κανόνα της **αφομοίωσης** (assimilation), ενώ παράλληλα σε κάποιες άλλες αποβάλλεται με τον φωνολογικό κανόνα της **αποβολής** (deletion) μετά την εφαρμογή της αφομοίωσης.

Ανάλογη διαδικασία επισυμβαίνει και με τα επιθήματα, τα οποία πολλές φορές προκαλούν αλλαγές στη μορφή του θέματος, με το οποίο συνδυάζονται. Τα παραδείγματα (2.) είναι από τον Αόριστο κάποιων ρημάτων (Ράλλη 2005: 66):

(2.) α. ράβ-ω	/rav-o/,	έ-ραψ-α	/e-rav-s-a/ → /e-rap-s-a/
β. βρέχ-ω	/vrex-o/,	έ-βρεξ-α	/e-vrex-s-a/ → /e-vrek-s-a/
γ. γράφ-ω	/graf-o/,	έ-γραψ-α	/e-graf-s-a/ → /e-grap-s-a/
δ. λήγ-ω	/liγ-o/,	έ-ληξ-α	/e-liγ-s-a/ → /e-lik -s-a/

Παρατηρώντας τα ρήματα στο (2.) βλέπουμε πως το κλιτικό επίθημα όψης *-σ-*, που εκφράζει το συνοπτικό ποιόν ενεργείας, τρέπει το ληκτικό φώνημα του ρηματικού θέματος. Η αλλαγή αυτή συντελείται λόγω της λειτουργίας του φωνολογικού κανόνα της **ανομοίωσης** (dissimilation), όπου το πρώτο από τα δύο γειτνιάζοντα εξακολουθητικά φωνήματα χάνει την εξακολουθητικότητά του.<sup>2</sup> Αξίζει να αναφέρουμε ότι οι συγκεκριμένες αλλαγές των μορφημάτων στο (1.) και (2.) που δημιουργούνται από την εφαρμογή κάποιου φωνολογικού κανόνα εξαρτώνται άμεσα από το μορφολογικό περιβάλλον και εντάσσονται στο πεδίο της Λεξικής Φωνολογίας (βλ. Kiparsky 1982β, Mohanan 1986, Ralli 1988, 2005). Σύμφωνα με τους υποστηρικτές της Λεξικής Φωνολογίας, οι κανόνες της λαμβάνουν χώρα μεταξύ των ορίων των μορφημάτων, διακρίνονται από προβλεψιμότητα και υπάρχουν εξαιρέσεις κατά την

<sup>2</sup> Ο συγκεκριμένος κανόνας ανομοίωσης δίνεται με βάση το δυαδικό σύστημα των διαφοροποιητικών χαρακτηριστικών, όπως αυτό προτάθηκε από τους Chomsky & Halle (1968) (πρβλ. για τα Ελληνικά, Nespor (1999)).

+ εξακολουθητικό → [-εξακολουθητικό]/\_ [+εξακολουθητικό]  
- κορωνιδικό

εφαρμογή τους. Αντίθετα οι κανόνες της Γενικής Φωνολογίας εφαρμόζονται πάντα και απαρέγκλιτα.

Συχνά από τους ερευνητές γίνεται χρήση του όρου *αλλομορφία* για αλλαγές σε μορφήματα, όπως στα παραδείγματα (1.) και (2.). Θεωρούμε πως δεν πρέπει να τους αποδίδεται αυτός ο όρος, καθότι οι παραλλαγές των μορφημάτων δικαιολογούνται από φωνολογικούς και μορφολογικούς κανόνες. Αντιθέτως, ο όρος της αλλομορφίας θα χρησιμοποιείται για παραλλαγές του ίδιου μορφήματος που δεν υπόκεινται σε καμία εφαρμογή κανόνα. Στη διεθνή βιβλιογραφία αυτού του είδους η αλλομορφία αναφέρεται είτε ως **λεξική αλλομορφία** (lexical allomorphy, πρβλ. Carstairs 1987), είτε ως **μορφολογική (καθαρή) αλλομορφία** (morphological (pure) allomorphy, βλ. Ράλλη 2005: 67).

### 2.1.1. Η Αλλομορφία ως «είδωλο» της γλωσσικής ιστορίας

Η αλλομορφία θέματος στα Ελληνικά αντανακλά πολύ συχνά τη φωνολογική ιστορία της γλώσσας, που δεν είναι πλέον διαφανής συγχρονικά. Παράλληλα καθρεφτίζει φωνολογικά και μορφολογικά φαινόμενα παλαιότερων φάσεων της Ελληνικής γλώσσας που έχουν πάψει να είναι ενεργά. Για παράδειγμα, στα Ελληνικά η δάσυνση συμφώνου πριν από δασυνόμενο φωνήεν (άντι + υγιεινός = άνθυγιεινός) ήταν κάποτε μία ενεργή διαδικασία, αλλά πλέον δεν υφίσταται. Έτσι η φωνολογικώς κυβερνώμενη αλλομορφία γίνεται μορφολογικώς κυβερνώμενη αλλομορφία. Επιπροσθέτως, το φαινόμενο του συλλαβικού αναδιπλασιασμού στους παρελθοντικούς χρόνους παρακειμένου και υπερσυντελικού, όπου συνήθως το πρώτο γράμμα του ρήματος συνθέτετε μία επιπρόσθετη συλλαβή σε συνδυασμό με το φωνήεν /ε/ τύπου CV/ε/, όπως *λέλυκα*, *γέγραφα*, *πεποίηκα*, είναι συγχρονικά ανενεργό· εμφανίζεται πλέον μόνο ως κατάλοιπο σε παθητικές μετοχές του τύπου *εγγεγραμμένος*, *προσκεκλημένος*, *πεπερασμένος*.

Αυτή η φυσική εξέλιξη της αλλομορφίας, κυρίως από το επίπεδο της φωνολογίας σε αυτό της μορφολογίας, συμβαίνει στην πλειονότητα των γλωσσών που εμφανίζουν αλλομορφοί. Για παράδειγμα, ο Tiersma (1983: 60) παραθέτει παραδείγματα από τα Φριζικά<sup>3</sup>, όπου υπήρχε η φωνολογική δημιουργία ημιφώνου πριν από μακρά μεσαία

---

<sup>3</sup> Η ομάδα των φριζικών διαλέκτων συμπεριλαμβάνει διαλέκτους της Γερμανικής οικογένειας γλωσσών στην Ολλανδία και στη Δυτικοσαξονική Γερμανία.

φωνήεντα βραχύνοντας αυτά τα φωνήεντα και σχηματίζοντας **διασπασμένους δίφθογγους** (broken diphthongs).

(3.)	<i>Αρχαία Frisian</i>	<i>Νέα Frisian</i>	<i>Ελληνικά</i>
	ske:n	skjin	‘βαγόνη’
	le:dər	ljedər	‘σκάλα’
	ro:t	rwot	‘αιθάλη’
	ske:lə	skwolə	‘σχολείο’

Όπως σημειώνει κι ο συγγραφέας, η συγκεκριμένη αλλαγή δεν αποτελεί συγχρονικό κανόνα στα Φριζικά, αφού οι ομιλητές οφείλουν να απομνημονεύουν τις λέξεις όπου συντελείται αυτή η αλλαγή παράλληλα με τα μορφολογικά περιβάλλοντα. Ομοίως και στα Ολλανδικά υπάρχει πληθώρα φωνολογικών κανόνων που αποτελούν συγχρονικά είδωλα παλαιότερης ιστορικής φάσης της γλώσσας. Ο Booij (1997: 38) παραθέτει την περίπτωση αποκοπής του φωνήματος *schwa*, όταν αυτό είναι τελικός χαρακτήρας ουσιαστικών, που έχει ως αποτέλεσμα συγχρονικά κάποια ολλανδικά ουσιαστικά (4.) να έχουν δύο μορφές και άλλα όχι.

(4.)	schande ‘ντροπή’	schand-knaap ‘απολίθωμα γιγάντιου χόρτου’
	aarde ‘γη’	aard-bei ‘φράουλα’
	armoede <sup>4</sup> ‘φτώχεια’	armoed-zaaier ‘ρακένδυτος’
	einde ‘τέλος’	eind-oordeel ‘τελική απόφαση’
	vreugde ‘χαρά’	vreugd-vuut ‘μεγάλη φωτιά’

Πολύ εύστοχα η αλλομορφία έχει ονομαστεί ως το «είδωλο» της γλωσσικής ιστορίας από τον Booij (1997). Όπως ακριβώς το είδωλο σε έναν καθρέφτη δεν μπορεί να προσεγγιστεί και να αναλυθεί, έτσι και η αλλομορφία είναι απρόβλεπτη και αυθαίρετη,<sup>5</sup> αφού αποτελεί την αντανάκλαση αλλαγών που έχουν επέλθει σε παλαιότερη περίοδο της γλώσσας (σε προηγούμενες φάσεις της). Άρα τι ήταν προτού να γίνει αλλόμορφο; Ας δούμε μερικά επιπλέον παραδείγματα αυτού του «ειδώλου».

(5.)	α.	τάξη	τάξε-ως	τάξ-εις	(ταξη~ ταξε~ ταξ)	<i>KNE</i>
		πόλη	πόλε-ως	πόλ-εις	(πολη~ πολε~ πολ)	

<sup>4</sup> Σημειώνεται πως η λέξη *armoede* ‘φτώχεια’ έχει τρία αλλόμορφα: *armoede*, *armoed*, *armoe*, όπου το αλλόμορφο *armoed* εμφανίζεται μόνο σε συγκεκριμένα σύνθετα, όπως *armoedzaaier* ‘ρακένδυτος’.

<sup>5</sup> Ως φαινόμενο η αλλομορφία χαρακτηρίζεται ότι είναι αυθαίρετη και απρόβλεπτη χωρίς να υπάκουει σε κανόνες. Εντούτοις η διαδικασία μπορεί να γίνει προβλέψιμη, όπως θα αποδειχθεί σε επόμενο κεφάλαιο.



	κλήση κλήσε-ως	κλήσ-εις	(κλήση~ κλήσε~ κλήσ)
β.	ερωτηθέντ-ς	>	έρωτηθείς, έρωτηθέντ-ος (γεν. εν.) AE
	αναρτηθέντ-ς	>	άναρτηθείς, άναρτηθέντ-ος (γεν. εν.)
	αναφερθέντ-ς	>	άναφερθείς, άναφερθέντ-ος (γεν. εν.)
γ.	ηεχω	>	έχω AE
	θήθημι	>	τίθημι
	χεχαριτωμένος	>	κεχαριτωμένος
δ.	καρδί-α/ καρδί-αι	AE >	καρδιά-Ø / καρδι-ές KNE
	ψυχ-ή/ ψυχ-αί	>	ψυχή-Ø / ψυχ-ές
	ποιητ-ής/ ποιητ-αί	>	ποιητή-ς / ποιητ-ές

Στα ουσιαστικά της ομάδας τύπου *τάξη* (5.α) υφίσταται αλλόμορφο τύπου *ταξε-* στη γενική ενικού και πληθυντικού· ο τύπος *ταξε-* δημιουργήθηκε στα Αρχαία Ελληνικά από το φωνολογικό φαινόμενο της χρονικής αντιμεταχώρησης<sup>6</sup> (το *πόλη-ος* έγινε *πόλε-ως* με την αντιστροφή της βραχύτητας και μακρότητας των δύο γειτονικών φωνηέντων), το οποίο είναι ανενεργό συγχρονικά λόγω της απουσίας ποσοτικής διαφοροποίησης στα φωνήεντα των Νέων Ελληνικών (Σταματάκος 1949: 138-139). Επίσης, οι κατάλοιπες παθητικές μετοχές των Αρχαίων Ελληνικών (5.β) σχημάτιζαν τον τύπο της ονομαστικής ενικού με την αποβολή του συμπλέγματος *-ντ-* προ του *-ς* και την αναπληρωτική έκταση του βραχέου φωνήεντος *ε* στο μακρό φωνήεν *ει* (βλ. Σταματάκος 1949: 71, ον. εν. *αναρτηθεντ-ς* > *αναρτηθείς*. γεν. εν. *αναρτηθέντ-ος*). Ωστόσο συγχρονικά οι συγκεκριμένες μετοχές εμφανίζουν αλλόμορφα τύπου *Xθ~Xθεντ* (*αναρτηθ~ αναρτηθεντ*), αφού κανένας από τους παραπάνω φωνολογικούς κανόνες των ΑΕ δεν υφίσταται πλέον, ώστε να δικαιολογούνται οι αλλαγές του μορφήματος. Επιπροσθέτως, ο νόμος του Grassmann<sup>7</sup> (5.γ), δηλαδή η αποδάσυνση του πρώτου φωνήματος σε λέξεις με δύο δασυνόμενα φωνήματα (κυρίως σύμφωνα), περιγράφει φωνολογικές μεταβολές που δεν ισχύουν στα Νέα Ελληνικά. Παρομοίως, η μετατόπιση των ορίων μεταξύ των μορφημάτων λόγω της επανάλυσης (5.δ) αποτελεί μέρος της γλωσσικής αλλαγής που έχει επέλθει στα Ελληνικά και έχει αφήσει το στίγμα της εμφανίζοντας αλλόμορφα συγχρονικά· όπως παρατηρούμε (5.δ) οι αλλαγές στα όρια των μορφημάτων στο κλιτικό παράδειγμα των Νέων Ελληνικών, σε

<sup>6</sup> Ο συγκεκριμένος κανόνας αναφέρεται ως ποσοτική μετάθεση από τον Μπαμπινιώτη (1985: 9).

<sup>7</sup> Ο Μπαμπινιώτης (1985) το αναφέρει ως ανομοίωση δασέων συμφώνων.

σύγκριση με το αντίστοιχο των Αρχαίων Ελληνικών οδήγησε στην εμφάνιση κλιτικών αλλομόρφων. Είναι σημαντικό και αξιοσημείωτο το γεγονός ότι τα περισσότερα προαναφερθέντα φαινόμενα που είναι ανενεργά συγχρονικά και οδήγησαν στην εμφάνιση αλλομόρφων, συντελούνταν στον τομέα της μορφοφωнологίας.

Η κατανόηση του συγκεκριμένου «ειδώλου» της γλωσσικής ιστορίας δεν είναι απαραίτητη μόνο για τη δικαιολόγηση του φαινομένου συγχρονικά, αλλά και για την ορθότερη χρήση του ορισμού της αλλομορφίας. Κρίνεται αναγκαία η προσφυγή στις ρίζες της αλλομορφίας, στη χρονική περίοδο όπου αυτές οι αλλαγές των μορφημάτων ήταν προβλέψιμες και αποτέλεσμα φωνολογικών ή μορφολογικών κανόνων. Αυτοί οι ανενεργοί κανόνες θα αποτελέσουν τη βάση, όπως θα δούμε σε επόμενο κεφάλαιο, για την υπολογιστική αντιμετώπιση της αλλομορφίας. Έτσι θα μπορούμε να συσχετίζουμε τις μορφές με κανονιστική σχέση και το υπολογιστικό μοντέλο θα μπορεί να παράγει τα αλλόμορφα χωρίς αυτά να καταχωρούνται.

## 2.2. Βιβλιογραφική ανασκόπηση της αλλομορφίας

Όντας ιδιαίτερο φαινόμενο της σύγχρονης μορφολογικής θεωρίας η αλλομορφία παρεκίνησε το ενδιαφέρον κάποιων ερευνητών από τις αρχές της δεκαετίας του 1980. Εντούτοις, όπως ορθώς επισημαίνει η Lieber (1982: 27), είναι ένα ζήτημα αρκετά παραμελημένο. Παρόλη την ενασχόληση κάποιων μελετητών με την αλλομορφία, αυτό το θέμα δεν αποτέλεσε το επίκεντρο κάποιας συστηματικής έρευνας, δεν της δόθηκε η δέουσα προσοχή για τη σημαντική συμμετοχή της στις διαδικασίες σχηματισμού λέξεων, αλλά ούτε έγινε το κέντρο προσοχής στο πλαίσιο της μορφολογίας ή της γενετικής γραμματικής γενικότερα. Κατά τη γνώμη μας, για αυτό ευθύνεται σε μεγάλο βαθμό η ελάχιστη συμμετοχή της αλλομορφίας στην Αγγλική γλώσσα· η παρουσία του και η συμβολή του φαινομένου υποβαθμίστηκε σε αρκετές ερευνητικές αναφορές. Πολύ συχνά συγκεκριμένα αλλομορφικά φαινόμενα έχουν αναλυθεί σε σχέση πάντα με άλλα πιο πιεστικά ζητήματα της σύνταξης, της φωνολογίας και του σχηματισμού λέξεων. Σπάνια η αλλομορφία αποτέλεσε ανεξάρτητα ένα θεωρητικό αντικείμενο ανάλυσης από μόνη της, αλλά τακτικά κατά την παρουσίαση θεωρητικών πλαισίων αποτέλεσε μέρος της συνολικής συζήτησης και προβληματικής. Ακολουθούν κάποια παραδείγματα από τις τρεις διαφορετικές γλωσσολογικές προσεγγίσεις του φαινομένου.

(i.) Η προσπάθεια ανάλυσης της (κλιτικής) αλλομορφίας στο πλαίσιο συντακτικών θεωριών. Για παράδειγμα, σε παλαιότερες συντακτικές θεωρίες γίνεται αναφορά σε έναν τοπικό κανόνα μετακίνησης, στον κανόνα του **περιοδεύοντος προσφύματος** (Affix Hopping). Ο παραπάνω κανόνας δεν σχετίζεται άμεσα με την αλλομορφία, αλλά στην προβληματική του εντάχθηκε και η συντακτική προσέγγιση του φαινομένου των απρόβλεπτων αλλαγών και των ανώμαλων τύπων. Ειδικότερα, συντακτικοί κανόνες οδηγούν το επίθημα αορίστου *-ed* των Αγγλικών στην επικόλλησή του σε ρηματικά θέματα (*tell-ED*, *feel-ED*, *buy-ED*), ενώ στη συνέχεια μετα-συντακτικοί κανόνες αντιστοιχούν τους παραπάνω τύπους με αλλομορφικούς τύπους, για παράδειγμα το *tell-ED* σε *told*, το *feel-ED* σε *felt*, το *buy-ED* σε *bought*, κτλ. Οι συγκεκριμένοι μετα-συντακτικοί κανόνες ελέγχουν την ορθότητα των τύπων στο Λεξικό και εφόσον υπάρχουν ανώμαλοι τύποι, αυτοί αντικαθιστούν τους λανθασμένους τύπους σε *-ed* που σχηματίστηκαν με τη μετακίνηση του επιθήματος.

(ii.) Σε έρευνες που κυριαρχεί η φωνολογία και το μοντέλο της θεωρίας του Βέλτιστου, τόσο οι αυθαίρετες, όσο και οι προβλέψιμες παραλλαγές των μορφημάτων αντιμετωπίζονται κυρίως με φωνολογικά κριτήρια. Σε κάποιες μελέτες βασισμένες στη Θεωρία του Βέλτιστου, για την επιλογή της κατάλληλης μορφής ενός μορφήματος χρησιμοποιούνται οι διάφορες ιδιότητες των συλλαβών του θέματος, η θέση του τόνου, η φωνολογική σύσταση των επιθημάτων. Για παράδειγμα, στα Γαλλικά το επίθετο *beau* ‘όμορφος’ έχει δύο αλλόμορφα (*beau~ bel*) και στις περιπτώσεις που συνδυάζεται με ουσιαστικό με εναρκτήριο φωνήεν, επιλέγεται το αλλόμορφο *bel~ belle*, όπως *bel ami* (\**beau ami*), *belle époque* κτλ., γιατί η επιλογή βασίζεται στις φωνολογικές ιδιότητες συγκεκριμένων συλλαβών (Mascaro 2007). Με αυτές τις προσεγγίσεις θα ασχοληθούμε αναλυτικότερα στην ενότητα 2.2.1.

(iii.) Σε θεωρίες σχηματισμού λέξεων και μορφολογίας η αλλομορφία αναλύεται από ένα σύνολο «περιγραφικών» κανόνων. Για παράδειγμα ο Aronoff (1976) αναφέρεται στους **διορθωτικούς κανόνες** (readjustment rules), ένα είδος μορφολογικών κανόνων που λειτουργούν μόνο με συγκεκριμένα μορφήματα σε περιβάλλοντα άλλων συγκεκριμένων μορφημάτων και μετά από άλλες παραγωγικές διαδικασίες σχηματισμού λέξεων. Παρουσιάζοντας τους κανόνες στα Ελληνικά, ένας τυπικός αλλομορφικός διορθωτικός κανόνας για τη ρηματική κλίση θα μπορούσε να είναι η εισαγωγή των φωνήματος /e/ στο θέμα ρημάτων τύπου *διαρ(ώ)* για τον σχηματισμό κάποιων χρόνων (όπως *διαίρε-σ-α*, *θα διαιρέ-σ-ω*), γιατί το αλλόμορφο *διαίρε* δεν θα

μπορούσε να έχει ανεξάρτητη θέση στο λεξικό, αλλά πρέπει να προκύψει με τη μορφή ενός διορθωτικού κανόνα.

Στις επόμενες υποενότητες παρατίθενται οι απόψεις και οι έρευνες διαφόρων γλωσσολόγων που ασχολήθηκαν με το θέμα της αλλομορφίας τις τελευταίες τρεις δεκαετίες. Η παράθεσή τους γίνεται με βάση το θεωρητικό πλαίσιο της προσέγγισής τους και τη χρονική περίοδο συγγραφής του έργου. Παράλληλα κάθε ανασκόπηση συνοδεύεται συνήθως από κριτική προσέγγιση της ανάλυσης και των προτάσεων.

### 2.2.1. (Μορφο-)Φωνολογικές προσεγγίσεις της Αλλομορφίας

*Drachman, Kager & Malikouti-Drachman (1995), Drachman (2006)*

Οι Drachman, Kager & Malikouti-Drachman (1995) και ο Drachman (2006) ερευνούν με βάση τη Θεωρία του Βέλτιστου (Optimality Theory - OT<sup>8</sup>) κατά πόσο κάποιες αλλαγές μορφημάτων είναι αποτέλεσμα εφαρμογής φωνολογικού κανόνα ή προκύπτουν αυθαίρετα (λεξικά καταχωρισμένα). Διερευνούν την αλλομορφία σε όλες τις διαδικασίες σχηματισμού λέξεων. Η ανάλυσή τους βασίζεται σε ρήματα με πολλαπλή αλλομορφία (*δερν~ δειρ~ δαρ*) και σε ουσιαστικά με αλλόμορφα με την προσθήκη συμφώνου (*παπα~ παπαδ, κυμα~ κυματ*). Παράλληλα εξετάζουν ζεύγη επιθημάτων που τα κατηγοριοποιούν ως αλλόμορφα, όπως *-ιμο<sup>9</sup> vs. -μα, -ινος vs. -ενιος*.

(6.)	<i>vréks-imo (vréx-o)</i>	vs.	<i>skúpiz-ma (skúpiz-o)</i>	<i>/-imo/ ~ /-ma/</i>
	<i>pétr-ino (pétr-a)</i>	vs.	<i>moliv-énjo (móliv-os)</i>	<i>/-ino/ ~ /-enjo/</i>
	<i>kléf-tis (klév-o)</i>	vs.	<i>kaθaris-tís (kaθariz-o)</i>	<i>/-tis/ ~ /-tís/</i>
	<i>ník-o (ník-os)</i>	vs.	<i>filip-e (filip-os)</i>	<i>/-o/ ~ /-e/</i>

(από Drachman, Kager & Malikouti-Drachman 1995: 2-3)

Υποστηρίζουν πως η διαδικασία της αλλομορφίας είναι αυθαίρετη και απρόβλεπτη, αλλά καταφεύγουν σε φωνοτακτικούς, μορφοπροσωδιακούς και προσωδιακούς περιορισμούς, για να προβλέψουν την επιλογή του αλλομόρφου. Οι συγκεκριμένοι περιορισμοί λειτουργούν πάντοτε συνδυαστικά, ώστε να αποκλειστούν οι λανθασμένοι

<sup>8</sup> Αναλυτικότερα, βλ. Prince & Smolensky (1993).

<sup>9</sup> Σύμφωνα με το μοντέλο της Ράλλη (2005) το συγκεκριμένο παραγωγικό επίθημα είναι *-σιμο*.

τύποι και να προκριθεί το κατάλληλο αλλόμορφο, όπως φαίνεται και στο ακόλουθο παράδειγμα για τους τύπους *skupiz-ma* vs. *\*skupiz-imo*.

(7.)

/skupiz-, {-ma, -imo}/	FTBIN	PARSE2	STEM=PRWD	NONFIN
a. (skupiz)-(má)	*!			*
b. (skúpis)-imo		*!		
c. skupis-(ímo)		*!	*	*
d. sku(píz-ma)			*!	*
e. sku(pís-i)mo			*!	
f. (skupis)-(ímo)				*!
g. ☞ (skúpis)-ma				

(από Drachman, Kager & Malikouti-Drachman 1995: 6)<sup>10</sup>

Υποστηρίζουν πως κάποια ρηματικά θέματα επιλέγουν τα αλλόμορφα από τα ζεύγη των παραγωγικών επιθημάτων (6.) με τα οποία συνδυάζονται, καθώς αποδεικνύουν ότι τα πολυσύλλαβα ρηματικά θέματα (από δύο συλλαβές και πάνω) σχηματίζουν ονοματικούς τύπους με το παραγωγικό επίθημα *-μα*, όπως *ανάβω > άναμμα*, *διδάσκω > δίδαγμα*, *σκουπίζω > σκούπισμα*, ενώ τα μονοσυλλαβικά θέματα παράγουν ουσιαστικά με το επίθημα *-ιμο*, όπως *βρέχω > βρέξιμο*, *παίζω > παίξιμο*. Επισημαίνουν πως τα αλλόμορφα που εμφανίζουν τα ουσιαστικά στο κλιτικό παράδειγμα μετέχουν στην παραγωγή και τη σύνθεση από κοινού (π.χ. κρέας [κρε~ κρεατ]: *κρεατ-ίλα* vs. *κρε-ο-κοφτήρας*, αίμα [αιμ~ αιματ]: *αιμάτ-ωμα* vs. *αιμ-ο-δοσία*). Στη θεωρία του Drachman (2006) περί κοινών αλλομόρφων εντάσσονται όλες οι λέξεις που έχουν απώλεια του αρχικού φωνήεντος (8.), επειδή αυτή η διαγραφή δεν είναι συστηματική σε όλα τα περιβάλλοντα.

- (8.) (η)μέρα ημερίσιος (\*μερίσιος), ημερομίσθιο (\*μερομίσθιο)  
μερόνυχτα  
(ε)νοίκι(ο) νοικιάζω, ενοικιαστής (\*νοικιαστής), νοικιάρης  
(ο)μιλώ μίλησε, ομιλητής (\*μιλητής)

(από Drachman 2006: 19)

Συνήθως αποφεύγουν τη χρήση μορφολογικής ορολογίας και στις περιπτώσεις που οι φωνολογικοί περιορισμοί αποτυγχάνουν στην επιλογή αλλομόρφου, καταφεύγουν σε

<sup>10</sup> Οι περιορισμοί είναι FT-BIN (δισύλλαβος πόδας), PARSE-2 (μία από δύο γειτονικές συλλαβές πρέπει να αναλυθεί σε πόδα), STEM=PRWD (το θέμα αντιστοιχεί σε πραγματική λέξη) και NONFIN (η κεφαλή πόδας της PrWd δεν πρέπει να είναι τελική).

σημασιολογικά κριτήρια για την αιτιολόγηση κάποιων αλλομορφικών συμπεριφορών. Για παράδειγμα, το *παραβλέπω* [= κάτι λανθάνει της προσοχής μου] > *παρ-έ-βλεψ-α*, σε σύγκριση με το *παραβλέπω* [= βλέπω κάτι/ κάποιον πολύ ώρα] > *παρα-είδ-α* που έχει βάση με το χαρακτηριστικό +λόγιο. Θεωρούμε ότι πολλές από τις λέξεις τους έχουν λανθασμένο χωρισμό μορφημάτων (π.χ. *κρέα-τα* αντί *κρέατ-α*), είναι κατασκευασμένες και αμφιβόλου γραμματικότητας λέξεις (*!καφεδόμπρικι*, *!αλεπόγουννα*) και υπάρχει παράταιρη σύνδεση μεταξύ λέξεων, αφού ο *ματωμένος* συνδέεται με το *αίμα* και όχι με το *ματώνω*, η *απελπισία* με την *ελπίδα* και όχι με το *απελπίζομαι*, τα *γε(η)ρατειά* με το *γέρος* και όχι με το *γήρας*. Το γενικότερο πρόβλημα της ανάλυσης αυτής είναι ότι εντέλει οι λέξεις εξετάζονται ως ανεξάρτητες μονάδες και παραβλέπονται οι διαδικασίες από τις οποίες προήλθαν στη συγχρονική φάση της γλώσσας.

*Mascarò (1996a, 1996b, 2007)*

Ο Mascarò (1996a, 1996b, 2007) βασιζόμενος στη Θεωρία του Βέλτιστου εισάγει την έννοια της εξωτερικής αλλομορφίας. **Εξωτερικό ή φραστικό αλλόμορφο** (external or phrasal allomorph) είναι η επιφανειακή φωνολογική μορφή της λέξης που εξαρτάται από τις άμεσες επιπτώσεις και εφαρμογές της Φραστικής Φωνολογίας, δηλαδή την επιτέλεση κανόνων, η οποία λαμβάνει χώρα έξω από το πεδίο της λέξης, με αποτέλεσμα η επιλογή των αλλομόρφων να εξαρτάται από τις λέξεις που έπονται ή προηγούνται.

(9.)	α.	<u>au</u> temps (<à le temps)	‘στο χρόνο’	<i>Γαλλικά</i>
		<u>à</u> tout le temps	‘καθ’ όλο το χρόνο’	
		<u>beau</u> train	‘όμορφο τρένο’	
		<u>be</u> l avion	‘όμορφο αεροπλάνο’	
	β.	j-il- <u>oh</u>	‘αυτός έχει δει κάτι’	<i>Tzeltal (Mayan)</i>
		s-nuts- <u>oh</u>	‘αυτός έχει κυνηγήσει κάτι’	
		s-mak’lij- <u>eh</u>	‘αυτός έχει ακούσει κάτι’	
	γ.	l’ ámo	‘ο ιδιοκτήτης’	<i>ΒΔ Καταλανικά</i>
		l’ uraká	‘ο τυφώνας’	
		lo pá	‘το ψωμί’	
		lo mite	‘ο μύθος’	

(από Mascarò 2007: 716)

Για την περιγραφή των μορφοφωολογικών αλλαγών των μορφημάτων καταφεύγει στη χρήση των φωνοτακτικών κανόνων με τη χρήση περιορισμών της Θεωρίας του Βέλτιστου. Ο Mascarò χρησιμοποιεί βασικούς περιορισμούς της Θεωρίας του Βέλτιστου, όπως οι ONSET, NO-CODA, FILL, COMPLEX, PARSE-SEG, MAX, οι οποίοι έχουν σχέση με τη δομή της συλλαβής (βλ. παράδειγμα (10.)). Οι παραπάνω περιορισμοί μπορούν να προβλέψουν την εξωτερική αλλομορφία και την τελική επιλογή του κατάλληλου αλλομόρφου, καθώς θεωρεί πως σε όλες τις περιπτώσεις έχουμε πιθανά υποψήφια αλλομορφα, αλλά μόνο ένα προκρίνεται λόγω των μαρκαρισμένων ( $\pm$ marked) φωνολογικών χαρακτηριστικών του μορφολογικού περιβάλλοντος. Επομένως, σε κάθε περίπτωση επιλέγεται το μη μαρκαρισμένο, όπως στο ακόλουθο παράδειγμα από τα Γαλλικά (Mascarò 1996b [2004: 520]):

(10.)

I.	'bel ami'	ONSET	NO-CODA
☞	bel ami		
	beau ami	*!	

II.	'beau mari'	ONSET	NO-CODA
	bel mari		*!
☞	beau mari		

Χρησιμοποιώντας τέτοια παραδείγματα, καταδεικνύει τον τρόπο επιλογής του αλλομόρφου ανάμεσα από τα υποψήφια, προσπαθώντας παράλληλα να αιτιολογήσει την ύπαρξη της εξωτερικής αλλομορφίας.

Η μελέτη του συγκεκριμένου φαινομένου είναι ενδιαφέρουσα, ωστόσο είναι καθαρά τυπική και περιγραφική στα περισσότερα σημεία. Ο Mascarò δεν εμπλέκεται σε βαθύτερες αναλύσεις λόγω και του ιδιαίτερου χαρακτήρα της εξωτερικής αλλομορφίας και βασίζεται στη θεωρία του Βέλτιστου και τη χρήση περιορισμών για την επιλογή του αλλομόρφου από μία λίστα υποψηφίων μορφών. Η συγκεκριμένη ανάλυση, όπως θα δούμε στις ενότητες 2.4.4 και 3.2.2, μπορεί να εφαρμοστεί και στα Ελληνικά στις περιπτώσεις κανονικής αλλομορφίας. Να σημειωθεί ότι οι φωνολογικές ιδιότητες των μορφημάτων σίγουρα προσδίδουν κάποια προβλεψιμότητα στη

διαδικασία αλλομορφίας και μπορούν να ενισχύσουν την προσπάθεια ενός μορφολογικού αναλυτή, όπως θα δούμε στο πείραμα του 5.2.

*Thornton (1997)*

Η Thornton (1997) ασχολείται με την αλλομορφία θέματος και παραγωγικού επιθήματος στα Ιταλικά, αναφερόμενη παράλληλα στην αλληλεπίδραση της μορφοφωнологίας. Υιοθετεί τη θεωρία του **σχήματος**<sup>11</sup> (schema) των Bybee & Slobin (1982) και Bybee & Moder (1983), σύμφωνα με την οποία προσπαθεί να κατηγοριοποιήσει σε ομάδα τα θέματα που εμφανίζουν αλλομορφία, κάνοντας το διαχωρισμό με φωνολογικά κριτήρια. Ορίζει το πρότυπο μίας κατηγορίας που λειτουργεί ως φυσική τάξη με βάση τη φωνολογική μορφή και αναγκάζεται να δημιουργήσει πολλές υποκατηγοριοποιήσεις και υποπεριπτώσεις με αποτέλεσμα να διαφοροποιεί στο τέλος την κατηγορία από το πρότυπό της. Στόχος της είναι να εντοπίσει συγκεκριμένες φωνολογικές συνθήκες που αποτελούν προϋπόθεση για τη δημιουργία ή όχι κάποιου αλλομόρφου.

(11.)	ministero ‘υπουργείο’	>	ministri <u>ale</u> ‘υπουργικός’
	mondo ‘κόσμος’	>	mondia <u>le</u> ‘παγκόσμιος’
	influenza ‘επιρροή’	>	influenza <u>le</u> ‘επηρεάσιμος’
	pastor ‘πάστορας’	>	pastora <u>le</u> ‘ποιμενικός’
	punto ‘σημείο’	>	punta <u>le</u> ‘ακριβής’
	mens- ‘μήνας’	>	mensua <u>le</u> ‘μηνιαίος’

(από Thornton 1997: 91-92)

Για παράδειγμα στα Ιταλικά η εμφάνιση του /i/ ή του /u/ ανάμεσα στο θέμα και στο επίθημα *-ale* εξαρτάται από την παρουσία ή απουσία πρόσθιου κορωνιδικού συμφώνου στο τέλος του θέματος (11.). Υποστηρίζει ότι συνήθως οι μορφολογικοί ρυθμιστικοί κανόνες δεν μπορούν να προβλέψουν την εμφάνιση αλλομορφίας, αλλά καταλήγει προτείνοντας ένα δείκτη προβλεψιμότητας για τα αλλόμορφα με βάση ένα σύνολο μορφολογικών στοιχείων στους συνδυασμούς θεμάτων με επιθήματα. Τέλος,

<sup>11</sup> Το σχήμα (schema) είναι μία κατάσταση που περιγράφει τις φωνολογικές ιδιότητες μίας μορφολογικής τάξης. Δεν είναι περιορισμός που καθορίζει αυστηρά τι μπορεί ή δεν μπορεί να εμφανιστεί, αλλά είναι περισσότερο ένας χαλαρός τύπος συσχέτισης.



θεωρεί ότι μόνο ο συνδυασμός μορφολογίας και φωνολογίας δεν δίνει πάντα λύση σε αλλομορφικά διλήμματα, και πιστεύει ότι η σημασιολογία του σχηματισμού λέξεων μπορεί να παρέχει κάποιες λύσεις. Για παράδειγμα, αναρωτιέται αν τα επιθήματα *-ano* και *-iano* είναι αλλόμορφα του ίδιου επιθήματος ή δύο διαφορετικά μορφήματα: ωστόσο, εκφράζει και τη σκέψη ότι είναι πιθανό το /i/ να είναι μέρος του θέματος δημιουργώντας έτσι θεματικό αλλόμορφο, π.χ. *montano* ‘βουνίσιος’ vs. *montiano* ‘χαρακτηριστικό από το V. Monti’.

Η Thornton οριοθετεί τα μορφήματα χρησιμοποιώντας κυρίως φωνολογικά κριτήρια. Ένα μειονέκτημα της ανάλυσής της είναι η εκτεταμένη χρήση της φωνολογίας (κυρίως της υπερτεμαχιακής) στις διαδικασίες σχηματισμού λέξεων. Επίσης, ο διαχωρισμός είναι προβληματικός σε κάποιες κατηγορίες, δεδομένου του υψηλού στατιστικού ποσοστού αντιπαραδειγμάτων που η ίδια παραθέτει. Στον προβληματισμό της για τα όρια των αλλομόρφων δίνει πιθανές λύσεις χωρίς να παίρνει κάποια ξεκάθαρη θέση. Για παράδειγμα, σε περιπτώσεις, όπως *tori+ale* vs. *tor+iale*? δέχεται ότι και οι δύο λύσεις είναι σωστές, αλλά προκρίνει τη δεύτερη ανάλυση ως πιο οικονομική για το σύστημα, δηλαδή είναι υπέρ της αλλομορφίας των επιθημάτων. Κλείνοντας, θεωρούμε ότι κάνει μία ενδιαφέρουσα παρατήρηση για τη σχέση σημασιολογίας και αλλομορφίας. Παρόλο που δεν αναλύει επαρκώς τη συγκεκριμένη σχέση, η παρατήρησή της μπορεί να έχει χρήσιμες εφαρμογές.

*Burzio (1999, 2005, 2006)*

Ο Burzio (1999, 2005, 2006) εξετάζει το πρόβλημα της αλλομορφίας στο πλαίσιο της Θεωρίας του Βέλτιστου<sup>12</sup> επικεντρώνοντας την προσοχή του στη σχέση Πιστότητας Εξαγόμενου – Εξαγόμενου (Output-Output Faithfulness), ενώ παράλληλα θεμελιώνει τη δική του πρόταση, τη Θεωρία της Αντιστοιχίας (Correspondence Theory). Υποστηρίζει πως στους πολύπλοκους μορφολογικούς σχηματισμούς συμμετέχουν ως εισαγόμενοι τύποι όλα τα πιθανά αλλόμορφα και η επιλογή του τελικού εξαγομένου γίνεται με την εφαρμογή πολλαπλών περιορισμών. Εξετάζει το ρηματικό σύστημα των Ιταλικών και παρατηρεί ότι υπάρχουν κάποια προκαθορισμένα αλλόμορφα σε

<sup>12</sup> Παρόμοιες προσεγγίσεις βασισμένες στο πρότυπο της Θεωρίας του Βέλτιστου είναι μεταξύ άλλων και οι έρευνες των Bonet (2004), Paster (2005), Kikuchi (2006), Lubowicz *et al.* (2006), Bye (2007), Bonet & Mascaró (2007), Löfstedt (2009), Bonet & Harbour (2010), Nevins (2010), Suh (2010), van der Veer & Booij (to appear). Η παρουσίαση των παραπάνω εργασιών υπερβαίνει τους στόχους της παρούσας έρευνας.

συγκεκριμένες κλιτικές τάξεις. Όταν κάποια ρήματα παρεκκλίνουν του αλλομορφικού προτύπου της συγκεκριμένης συζυγίας και εμφανίζουν διαφορετικές αλλομορφικές συνήθειες, ομαδοποιούνται διαφορετικά εντός των ίδιων συζυγιών. Επίσης, υποστηρίζει πως στις ρομανικές γλώσσες κάποιες ρηματικές συζυγίες εμφανίζουν αντιστάσεις στην εμφάνιση αλλομορφίας λόγω της υψηλής θέσης στην Πιστότητα Εξαγόμενου – Εξαγόμενου, όπως για παράδειγμα η πρώτη κλιτική συζυγία των Ιταλικών σε *-are* (12.), η οποία δεν εμφανίζει αλλομορφία.

(12.)	<i>Συζυγία</i>	<i>Απαρέμφατο</i>	<i>Μετοχή</i>	
	I	adapt-are	adapt-át-o	‘προσαρμόζω’
	I	am-are	am-át-o	‘αγαπώ’
	I	gener-are	gener-át-o	‘παράγω’

(από Burzio 1999: 45)

Θεωρεί ότι η αντίσταση στην εμφάνιση αλλομορφίας εντοπίζεται σε ρηματικές ή ονοματικές κλιτικές τάξεις όλων των γλωσσών. Σημειώνει πως η έλλειψη τόνου από ένα επίθημα δημιουργεί αλλόμορφο και αποκλίνοντας από το θεωρητικό πλαίσιο αναφέρει πως σε μερικές περιπτώσεις από τα υποψήφια αλλόμορφα η τελική επιλογή γίνεται λεξικά χωρίς να δικαιολογείται με κάποιον περιορισμό.

Στις μελέτες του Burzio εντοπίζουμε τη συστηματική χρήση φωνολογικών χαρακτηριστικών και περιορισμών για την ανάλυση του φαινομένου της αλλομορφίας, για παράδειγμα σημειώνει πως η παρουσία ή απουσία του τόνου μπορεί να δημιουργήσει αλλόμορφα. Παρατηρεί ορθώς κάποιες αποκλίσεις στις αλλομορφικές συμπεριφορές των συζυγιών, αλλά δεν εξάγει κάποια γενικά συμπεράσματα, ούτε κάνει συστηματική ταξινόμηση των αλλομορφικών εξαιρέσεων, δηλαδή των αλλομόρφων που διαφοροποιούνται από τα υπόλοιπα της ίδιας συζυγίας. Δέχεται πως υπάρχει ένα προκαθορισμένο αλλόμορφο στο Λεξικό, ενώ συμμετέχουν πολλαπλά αλλόμορφα στις διαδικασίες σχηματισμού λέξεων, τα οποία αποκλείονται βάση συγκεκριμένων περιορισμών. Η ιεράρχηση των περιορισμών που προτείνει για την επιλογή του κατάλληλου αλλομόρφου δεν είναι συνεπής ως προς την αντιμετώπιση των κατηγοριών του θέματος και του επιθήματος, με αποτέλεσμα να εντοπίζονται κενά στη θεωρητική προσέγγιση.

## 2.2.2. Μορφολογικές προσεγγίσεις της Αλλομορφίας

*Lieber (1982)*

Η Lieber (1982) ασχολήθηκε με την αλλομορφία με δεδομένα από τα Λατινικά, Αγγλικά, Γερμανικά και Warlpiri, με σκοπό να υποστηρίξει ότι το φαινόμενο είναι μορφολογικό. Απορρίπτει οποιαδήποτε συντακτική και φωνολογική προσέγγιση του φαινομένου και φέρει επιχειρήματα ενάντια στην προσέγγιση με **διορθωτικούς κανόνες** (readjustment rules, βλ. περισσότερο Lieber 1982: 27-32 για τις R θεωρίες). Καταχωρεί τα αλλόμορφα στο Λεξικό μαζί με τις υπόλοιπες πληροφορίες του λήμματος και τα συνοδεύει από διακριτικά χαρακτηριστικά τύπου [+ΕΝΙΚΟΣ, +ΠΛΗΘΥΝΤΙΚΟΣ, +ΑΟΡΙΣΤΟΣ]. Καταχωρεί τα αλλόμορφα ως αντικείμενα λίστας, η οποία λίστα παρέχει τα εισαγόμενα (inputs) για τις διαδικασίες σχηματισμού λέξεων. Προτείνει ότι η αλλομορφία προηγείται των όποιων διεργασιών δημιουργίας λέξεων και δεν είναι παραγωγική διαδικασία. Αναφέρει ότι η δημιουργία των αλλομόρφων αποτελεί μέρος της γλωσσικής ιστορικής εξέλιξης.

Για την οργάνωση του λεξικού επισημαίνει πως είναι απαραίτητη η ομαδοποίηση των λεξημάτων ανάλογα με την αλλομορφική τους συμπεριφορά. Εισάγει την έννοια των **μορφολεξικών κανόνων** (morpholexical rules) ως ένα είδος πλεοναστικών κανόνων με βάση τους οποίους εντάσσει τα αλλόμορφα σε **αλλομορφικές τάξεις** (allomorphic classes). Οι μορφολεξικοί κανόνες της Lieber περιγράφουν τη σχέση που κυβερνά τα αλλόμορφα ενός μορφήματος, όπως π.χ. στα Λατινικά (13.α) με το θεματικό φωνήεν των διαφορετικών ρηματικών τάξεων και στα Γερμανικά (13.β) με τη σύνδεση αλλομόρφων ενικού και πληθυντικού:

- |       |         |  |                  |
|-------|---------|--|------------------|
| (13.) | α. (i.) | $C_0 V C_0 \sim C_0 V C_0 \bar{a}$ (am, amā)         | <i>Λατινικά</i>  |
|       |         | (ii.) $C_0 V C_0 \sim C_0 V C_0 \bar{e}$ (del, delē) |                  |
|       |         | (iii.) $C_0 V C_0 \sim C_0 V C_0 i$ (cap, capi)      |                  |
|       | β.      | AC2: $X \sim Xs$ (Streik, Streiks)                   | <i>Γερμανικά</i> |
|       |         | AC4: $X \sim Xn$ (Aug, Augen)                        |                  |

(από Lieber 1982: 31, 40)

Οι ιδιότητες των μορφολεξικών κανόνων<sup>13</sup> είναι ότι: (α.) δεν είναι κανονικοί κανόνες παραγωγής λέξεων, (β.) δεν είναι παραγωγικοί, (γ.) δεν μπορούν να μπουν σε

<sup>13</sup> Για τους μορφολεξικούς κανόνες στα Ελληνικά βλ. Ralli (1988), Ράλλη (2005).

σειρά και (δ.) απλώς περιγράφουν αυθαίρετες σχέσεις. Αυτοί οι κανόνες υπήρξαν κάποτε παραγωγικοί και ενεργοί, αλλά πλέον δεν χρησιμοποιούνται και παραμένουν ανενεργοί, όπως για παράδειγμα, ο φωνολογικός κανόνας ‘umlaut’ των Γερμανικών που έχει πλέον μορφολογικοποιηθεί. Όλα τα αλλόμορφα μπορούν να συμμετέχουν σε όλες τις διαδικασίες σχηματισμού λέξεων<sup>14</sup>, ενώ παράλληλα είναι η σημαντική η κωδικοποίηση της πληροφορίας της αυθαίρετης κατάταξης σε αλλομορφική τάξη. Η Lieber υποθέτει ότι είναι απαραίτητο να υπάρξει μία ιεραρχία στη σειρά εφαρμογής των αλλομορφικών κανόνων. Τέλος, υποστηρίζει ότι δεν διαχωρίζονται τα αλλόμορφα που εμφανίζονται σε περιβάλλοντα κλίσης από εκείνα σε περιβάλλοντα παραγωγής ή σύνθεσης, γιατί μπορούν να εμφανιστούν σε κάθε διαδικασία σχηματισμού λέξεων.

Η Lieber απέρριψε τις Διορθωτικές θεωρίες (R(eadjustment) theories), αφού κατέδειξε τις αδυναμίες τους. Η ανάλυσή της παρέχει περιγραφική και ερμηνευτική επάρκεια για το φαινόμενο της αλλομορφίας. Θεωρεί το φαινόμενο ως μη γενετική διαδικασία – κατάλοιπο της ιστορικής εξέλιξης της γλώσσας – αλλά δεν τονίζει την πραγματική διάσταση της συγχρονικής της συμμετοχής στις διαδικασίες. Το ενδιαφέρον της εστιάζεται στην αλλομορφία των θεμάτων και παραβλέπει τις αλλομορφικές συμπεριφορές των προσφυσμάτων. Η δημιουργία των αλλομορφικών της κανόνων είναι οικονομική, αλλά παραμένει εν μέρει προβληματική. Παρόλο που εισάγει την έννοια των αλλομορφικών τάξεων, δεν επεκτείνει τη συγκεκριμένη έννοια για την κατηγοριοποίηση των ουσιαστικών και ρημάτων. Τέλος, δημιουργεί υπεργενικεύσεις βασιζόμενη στα δεδομένα των γλωσσών που αναλύει.

- (14.) α. Buchbinder ‘βιβλιοδέτης’                      Bücherfolge ‘σειρά βιβλίων’  
Vaterland ‘πατρίδα’                                      Vätersitte ‘τα ήθη των προγόνων’  
(από Lieber 1982: 40)

Για παράδειγμα, υποστηρίζει ότι όλα τα αλλόμορφα ενός λήμματος συμμετέχουν στη διαδικασία της σύνθεσης με βάση τα δεδομένα των Γερμανικών (14.α), ενώ αυτό δεν ισχύει σε πολλές γλώσσες, όπως στα Ελληνικά (14.β), αφού στη σύνθεση συμμετέχει μόνο ένα αλλόμορφο στη θέση του πρώτου συνθετικού (βλ. Ράλλη 2005, Καρασίμος 2011).

<sup>14</sup> Αναφέρει συγκεκριμένα (1982: 40): «τοποθετώντας σε λίστα τα διάφορα αλλομορφικά θέματα των γερμανικών κλιτικών παραδειγμάτων ως μέλη διαφόρων αλλομορφικών τάξεων ενισχύεται η άποψη ότι όλα τα αλλόμορφα πρέπει να είναι διάθεσιμα στο λεξικό ως εισαγόμενο σε κάθε είδος σχηματισμού λέξεων. Και αυτό όντως συμβαίνει στην κλίση, στην παραγωγή και στη σύνθεση».

- |       |    |                          |                       |
|-------|----|--------------------------|-----------------------|
| (14.) | β. | <u>κυματο</u> θραύστης   | <u>κυματο</u> μορφή   |
|       |    | <u>παπαδο</u> παίδι      | <u>παπαδο</u> σύναξη  |
|       |    | <u>καρδιο</u> κατακτητής | <u>καρδιο</u> λογικός |

*Carstairs (1987)*

Ο Carstairs (1987) υποστηρίζει την αυτονομία της μορφολογίας με στόχο να προσεγγίσει τη διαδικασία της κλίσης για την εξαγωγή γενικεύσεων σχετικά με τις μορφολογικές συμπεριφορές των θεμάτων. Παράλληλα στοχεύει στην ερμηνεία των διαφόρων αποκλίσεων του κλιτικού συστήματος και ειδικότερα των αλλομορφικών αλλαγών. Πιο συγκεκριμένα εντοπίζει τέσσερις αποκλίσεις – αλλομορφίες, οι οποίες συσχετίζονται με τα περιβάλλοντα συμμετοχής των θεμάτων και τις μορφολεξικές ιδιότητες που μπορεί να φέρουν σε λεξικές δομές (Carstairs 1987: 14):

- (15.) Απόκλιση I: Μία ιδιότητα σε πολλούς μορφικούς εκπροσώπους συνταγματικά.  
 Απόκλιση II: Μία ιδιότητα σε πολλούς μορφικούς εκπροσώπους παραδειγματικά.  
 Απόκλιση III: Πολλές ιδιότητες σε έναν μορφικό εκπρόσωπο συνταγματικά.  
 Απόκλιση IV: Πολλές ιδιότητες σε έναν μορφικό εκπρόσωπο παραδειγματικά.

Υποστηρίζει ότι εάν υπάρχουν αρκετές διαθέσιμες πολλαπλές κλιτικές πραγματώσεις των επιθημάτων (αλλόμορφα) για έναν συγκεκριμένο συνδυασμό μορφοσυντακτικών ιδιοτήτων, το θέμα πρέπει να επιλέξει *μόνο* μία από αυτές τις πραγματώσεις.

- (16.) I. έλελύκετε (ΑΕ): στο β' πληθυντικό υπερσυντέλικου η όψη δίνεται από το πρόθημα *λε-* και το επίθημα *-κ-* περίξ του θέματος *λυ-*.
- II. –en vs. –s (Αγγλικά): το αλλόμορφο του πληθυντικού στα ουσιαστικά των Αγγλικών, όπως *ox-en* 'βόδια'.
- III. –me– (–miy–) vs. –méz (Τουρκικά): το αρνητικό μόρφημα –*méz*– (έναντι του –*me*–) εμφανίζεται μόνο στον ενεστωτικό αόριστο (Present Aorist) και περιέχει μαζί με την έννοια της άρνησης και το χαρακτηριστικό του αορίστου, π.χ. *gelméz* 'δεν έρχεται' (όπου εσωκλείεται και η πληροφορία του γ' προσώπου).
- IV. –eb-i vs. –ni (Γεωργιανά): ο τύπος –*eb-i* δηλώνει τον πληθυντικό με το –*eb*– και την ονομαστική με το –*i*, ενώ στο αλλόμορφο –*ni* οι πληροφορίες πληθυντικού και ονομαστικής υπάρχουν στο ίδιο μόρφημα, π.χ. *kal* 'κόρη' – *kal-eb-i* 'κόρες' και *kal* 'κόρη' – *kal-ni* 'κόρες'.

(από Carstairs 1987: 14-16, 131)

Ωστόσο υπάρχουν περιπτώσεις αλλομορφικών πραγματώσεων, οι οποίες συσχετίζονται με σημασιολογικές και υφολογικές λειτουργίες και δεν επηρεάζονται από συγκεκριμένες μορφοσυντακτικές ιδιότητες. Παραδείγματος χάριν, στα Γερμανικά η λέξη *Wort* ‘λέξη’ εμφανίζει δύο αλλομορφικούς τύπους στον πληθυντικό με διαφορετική σημασία, *Wörter* ‘λέξεις’ vs. *Worte* ‘λόγια’ (ό.π.: 31).

Σε γενικές γραμμές, καθορίζει την αλλομορφία με κριτήριο τις φωνολογικές και τις μη-φωνολογικές διαφοροποιήσεις των μορφημάτων. Οι επιλογές των αλλομόρφων γίνονται είτε με βάση τα φωνολογικά χαρακτηριστικά του περιβάλλοντος, όπως στα Ουγγρικά (17.α) όπου τα επιθήματα *-ol* και *-(a)sz* εμφανίζονται ανάλογα με τα φωνήματα που προηγούνται, είτε με σημασιολογικά-λεξικά κριτήρια, όπως στα Ρωσικά (17.β) τα επιθήματα *-on* και *-y* εμφανίζονται στην αιτιατική πληθυντικού ανάλογα με το αν είναι έμψυχο ή όχι το θέμα.

(17.)	α.	<i>olvas-ol</i>	‘εσύ διαβάζεις’			<i>Ουγγρικά</i>
		<i>ül-sz</i>	‘εσύ κάθεσαι’			
	β.	<i>student</i>	‘φοιτητής’	<i>student-on</i>	αιτ.πλ.	<i>Ρωσικά</i>
		<i>stol</i>	‘τραπέζι’	<i>stol-y</i>	αιτ.πλ.	

(από Carstairs 1987: 72)

Στην προσπάθεια ενίσχυσης της θεωρίας ο Carstairs εισάγει την έννοια της αρχής της **οικονομίας του κλιτικού συστήματος** και τη συνδέει με τις μη-φωνολογικές αλλαγές μορφημάτων. Πιο συγκεκριμένα, για τον Carstairs η θεματική αλλομορφία αποτελεί κεντρικό κριτήριο για τη δημιουργία ενός συστήματος με λιγιστά κλιτικά μακροπαραδείγματα (*macroparadigms*, αναλυτικότερα βλ. Wurzel 1984, 1987). Εντούτοις, ο συγγραφέας προτείνει τα προσφύματα αντί της θεματικής αλλομορφίας ως τη βάση για τον καθορισμό των παραδειγμάτων και των διαφορών τους. Τα μακροπαραδείγματα παρουσιάζονται κυρίως μέσω του ονοματικού κλιτικού συστήματος των Γερμανικών (18).

(18.)	I.	<i>Tag</i>	‘ημέρα’	<i>Tag-en</i>	‘ημέρες’
	III.	<i>Gast</i>	‘καλεσμένος’	<i>Gäst-e</i>	‘καλεσμένοι’
	VII.	<i>Mann</i>	‘άντρας’	<i>Männ-er</i>	‘άντρες’
	XV.	<i>Firma</i>	‘φίρμα’	<i>Firm-en</i>	‘φίρμες’

(από τον Carstairs 1987: 235)

Το βιβλίο του Carstairs (1987) αποτελεί μία σημαντική προσπάθεια για την ανάλυση της κλίσης και την οριοθέτηση της αυτονομίας της μορφολογίας. Οι αρχές που παρουσιάζει είναι καθαρά μορφολογικές, αλλά κάνει ελάχιστη χρήση των δυνατοτήτων που του παρέχει η αλλομορφία. Δεν ακολουθεί τον διαχωρισμό των αλλομόρφων σε εκείνα που προκύπτουν από φωνολογικούς κανόνες και σε εκείνα που οι αλλαγές είναι αυθαίρετες. Προτείνει την κατηγοριοποίηση των αλλομόρφων με βάση το συσχετισμό των ιδιοτήτων και των μορφικών εκπροσώπων (exponents, βλ. θεωρία του Matthews 1974) στο κλιτικό παράδειγμα (βλ. (15.)). Στην προσέγγιση του ονοματικού κλιτικού συστήματος των Γερμανικών, ο Carstairs διαχειρίζεται λανθασμένα την αλλομορφία (βλ. αναλυτικότερα την κριτική του Wurzel 1989) σε περιπτώσεις που παρατηρείται ομοιότητα συμπεριφοράς των θεμάτων (behaviour similarity), δηλαδή σε περιπτώσεις που εμφανίζουν όμοιες πιθανές αλλαγές και συνδυασμούς με κλιτικά επιθήματα, αφού αυτό τον οδηγεί σε προβληματικά μακροπαράδειγματα. Για τη χρήση της ομοιότητας συμπεριφοράς, ο Wurzel (1989: 208) αναφέρει πως εάν εγκατέλειπε τη συνθήκη της ομοιότητας, όλες οι κλιτικές τάξεις των επιθέτων, ουσιαστικών και ρημάτων που είναι αμαρκάριστες με βάση τη θεωρία της κλιτικής μαρκαρισσιμότητας, θα αποτελούσαν ένα μακροπαράδειγμα, ενώ όλες οι λέξεις χωρίς συγκεκριμένα κλιτικά δεδομένα θα καταχωρούνταν στο λεξικό. Οι περιορισμοί και οι αρχές για την κλίση που επεξεργάζεται ο Carstairs με βάση τις διαφορετικές πραγματώσεις πρέπει να τύχουν αναλυτικότερης μελέτης σε διαχτυτικές γλώσσες, αν και υποθέτει γενικευμένη εφαρμογή της προσέγγισης του.

*Maiden (1992, 2004, 2007)*

Ο Maiden (1992, 2004, 2007) υποστηρίζει ότι η αλλομορφία αποτελεί στην ουσία ανωμαλία ενός συστήματος και επομένως, μπορεί να χαρακτηριστεί ως μία ενεργή, αφηρημένη δομική ιδιότητα των μορφολογικών συστημάτων. Η μορφολογική ‘ανωμαλία’ δεν πρέπει να συγχέεται με τη συχνότητα εμφάνισης των ανώμαλων θεμάτων, καθώς σχετίζεται με τους εγγενείς παράγοντες αλλαγής της μορφής. Στόχος του Maiden είναι να δημιουργήσει αφηρημένα πρότυπα της ανωμαλίας του κλιτικού παραδείγματος, δηλαδή γενικά σχήματα αλλομορφικής συμπεριφοράς, τα οποία θα μπορούν να ερμηνεύσουν τις αλλομορφικές αλλαγές. Για την περιγραφή αρκετών αλλομόρφων προτείνει τη χρήση συγκεκριμένων μορφολογικών σχημάτων, όπως για παράδειγμα το ‘n-σχήμα’ της ιταλικής ρηματικής κλίσης (19.α).

(19.)	α. /η / vs. /n/ ('gn' vs. 'n')	<i>Παλαιά Ιταλικά</i>				
	1ΕΝ	2ΕΝ	3ΕΝ	1ΠΛ	2ΠΛ	3ΠΛ
IND.	vegno	vieni	viene	venimo	venite	vegnono
SUBJ	vegna	vegna	vegna	vegnamo	vegna	vegna
IND	rimagno	rimani	rimani	rimanemo	rimanete	rimagnono
SUBJ	rimagna	rimagna	rimagna	rimagnamo	rimagnate	rimagnano
	β. άτονο /ə/ (</a/)	<i>Παλαιά Γαλλικά</i>				
	mener	>	merrai	'οδηγώ'		
	doner	>	durrai/ dorrai	'δίνω'		
	parler	>	parrai	'μιλώ'		
	torner	>	torrai	'επιστρέφω'		

(από Maiden 1992: 292, 295)

Όπως φαίνεται από τα δεδομένα του (19.α), το 'n-σχήμα' αφορά στην ουρανικοποίηση του τελικού συμφώνου του θέματος. Ο Maiden (1992) υποστηρίζει ότι στο 'n-σχήμα' η θέση του τόνου διαδραματίζει καθοριστικό ρόλο για την αλλομορφία (τονισμένα θέματα [rhizotones] vs. μη τονισμένα θέματα [arhizotones]). Στην συνέχεια παραθέτει και διάφορα άλλα σχήματα ουρανικοποίησης που προήλθαν από την επίδραση του τόνου και δημιούργησαν αλλόμορφα. Επιπροσθέτως, χαρακτηριστικά είναι τα παραδείγματα από τα Γαλλικά στο (19.β), στα οποία φαίνεται ότι αρκετοί φωνολογικοί κανόνες, όντας πλέον ανενεργοί, οδηγούν στον σχηματισμό 'ανώμαλων' τύπων.

Συγχρονικά ο συσχετισμός του τόνου με τη διαφοροποίηση του μορφήματος είναι πλέον αδιαφανής και απρόβλεπτος. Για παράδειγμα, στις ρομανικές γλώσσες η θέση του τόνου έχει μορφολογικοποιηθεί και κανένας φωνολογικός κανόνας δεν μπορεί να προκαλέσει μετατόπιση. Σχετικά με τις στρατηγικές που χρησιμοποιούν οι ομιλητές για την αντιμετώπιση των ανωμαλιών ο Maiden αναφέρει χαρακτηριστικά ότι «οι ομιλητές ερχόμενοι αντιμέτωποι με συγκρουόμενα τυπολογικά στοιχεία στις λεξικές ρίζες, καταφεύγουν σε μία ποικιλία στρατηγικών, των οποίων θεμέλιος στόχος είναι να εισάγουν όσο το δυνατόν πιο σταθερά στη γραμματική τις ποικίλες διαφοροποιήσεις που έχουν συναντήσει. Μία στρατηγική είναι να συσχετισθεί η αλλομορφία με μία άλλη μεγαλύτερη (και σημασιολογικά αυθαίρετη) μορφολογική τάξη, δηλαδή συζυγία. Άλλη στρατηγική είναι η αναζήτηση της ταυτοποίησης κοινών κατανεμημένων σχημάτων που κρύβονται κάτω από το εύρος μιας φωνολογικά διαφορετικής αλλομορφίας» (Maiden 1992: 293).



Στη μελέτη του (2004) ο Maiden κάνει λόγο για την υποκατάσταση και υποστηρίζει ότι είναι ένα είδος αλλομορφίας που δεν συμβαίνει τυχαία, αλλά ακολουθεί κάποια πρότυπα (patterns). Τα πρότυπα της υποκατάστασης μπορούν να θεωρηθούν ως μορφολογικά αυτόνομα (βλ. και Aronoff 1994, Pirrelli 2000), γιατί δεν καθορίζονται συγχρονικά από φωνολογικούς, λειτουργικούς/ σημασιολογικούς παράγοντες. Εντούτοις ο Maiden επισημαίνει ότι το χαρακτηριστικό της συνωνυμίας είναι καθοριστικό σε αρκετές περιπτώσεις υποκατάστασης τύπων.

(20.) VALERE ‘αξίζω’ (+ ANDARE ‘πηγαίνω’) *Ligurian (Monaco)*

Present	vagu	vai	va	andamu	ande	van
Pres. subj.	vage	vagi	vage	andamu	ande	vagu
Imperfect	andavu	andavi	andava	andavamu	andavi	andavu
Imprf. subj.	andəssa	andəssi	andassa	andəssəmu	andəssi	andəssu
Future	va.ɔ	va.ɪ	va.ɪ	va.ɪəmu	va.ɪ	va.ɪan
Conditional	va.ɪessa	va.ɪassi	va.ɪossa	va.ɪəssomu	va.ɪəssi	va.ɪəssu

(από Maiden 2004: 245)

Με βάση τα δεδομένα ο Maiden οδηγείται στο συμπέρασμα ότι το παράδοξο της «ομαλής» υποκατάστασης είναι η συνεπαγωγή της μετατροπής των σημασιολογικώς διακριτών λεξημάτων σε απόλυτα συνώνυμα αλλόμορφα. Θεωρεί πως η υποκατάσταση οφείλεται στην διεπαφή συγγενών γλωσσών και κυρίως, στην ύπαρξη συνωνύμων υψηλής συχνότητας, όπου ένα λέξημα A δανείζει τύπους σε συνώνυμο λέξημα B, με αποτέλεσμα το λέξημα A να οδηγείται στην εξάλειψη. Σημειώνει πως τα λεξήματα που υποκαθιστούν τύπους από συνώνυμα λεξήματα έχουν έννοιες που σε μεγάλο βαθμό συμπίπτουν απόλυτα και διακρίνονται από μια σημασιολογική απόχρωση που δεν είναι πάντα σαφώς αντιληπτή από τους ομιλητές (ό.π.: 248).

Ο Maiden (1992, 2004, 2007) επανέφερε στο προσκήνιο την προβληματική γύρω από την αλλομορφία. Βασικό πλεονέκτημα της προσέγγισής του, αποτελεί το γεγονός ότι χρησιμοποιεί ένα αρκετά μεγάλο εύρος γλωσσών, ώστε να θεμελιώσει τις αρχές του και να τεκμηριώσει τη σχέση των ανώμαλων τύπων με τη μορφολογική αλλαγή. Ωστόσο, συχνά ακροβατεί ανάμεσα στη συγχρονία και τη διαχρονία με αποτέλεσμα η οριοθέτηση των μορφημάτων να είναι σε κάποια σημεία ασαφής.

*Booij (1997)*

Ο Booij (1997) υποστηρίζει τη σημαντικότητα της αλλομορφίας θεωρώντας την θεμέλιο λίθο της αυτονομίας της μορφολογίας. Αναφέρει δύο ειδών αλλομορφίες:

α) τη φωνολογική αλλομορφία, π.χ. Ολλ.: *naad* /nat/ ‘ρυτίδα’ – *naad-en* /naden/ ‘ρυτίδες’ και

β) τη μορφολογική ή λεξική αλλομορφία που δεν δικαιολογείται συγχρονικά από κανένα φωνολογικό κανόνα, π.χ. Ολλ.: *eet* ‘τρώω’ – *et-er* ‘ο τρώγων’ (*eet*~ *et*).

Οι αλλομορφικοί κανόνες δεν αφορούν τον σχηματισμό λέξεων (βλ. και Lieber 1982, Ralli 1988, 2005, 2006), αφού δεν προσθέτουν μόρφημα πλήρους σημασίας στη βάση, αλλά εισάγουν μία αλλαγή μορφής χωρίς σημασιακή διαφοροποίηση. Για παράδειγμα, στα Λατινικά: *dele-t-us* (μτχ.) > *delet-io* ‘καταστροφή’, όπου το /t/ της μετοχής *deletus* έχει ενσωματωθεί στη ρίζα, δεν αποτελεί αυτόνομο επίθημα και δεν δηλώνει πλέον τη μετοχή παρελθόντος στο θέμα της λέξης *deletio*. Ερμηνεύει την αλλομορφία με συγχρονικούς φωνολογικούς κανόνες αποδεχόμενος την ιστορική της υπόσταση ως κατάλοιπο. Καταχωρεί τα αλλόμορφα στο Λεξικό, ως στοιχεία ενός προκαθορισμένου μορφήματος, χωρίζοντάς τα σε **προκαθορισμένα** (default) και σε **εξειδικευμένα** (specified) αλλόμορφα και διαφωνεί με το χωρισμό της αλλομορφίας από την υποκατάσταση (suppletion). Υιοθετώντας το θεωρητικό μοντέλο της Lieber (1982), αναφέρει πως δεν υπάρχει διαχωρισμός μεταξύ των αλλομόρφων παραγωγής και σύνθεσης. Χρησιμοποιεί την παραδειγματική αλλομορφία για να πιστοποιήσει την αυτονομία της μορφολογίας, ωστόσο παρατηρεί ότι στην παραδειγματική αλλομορφία τα αλλόμορφα εμφανίζουν διαφορετικές συμπεριφορές στις διαδικασίες σχηματισμού λέξεων και υπακούουν στην αρχή του **Ομοιόμορφου Μαρκαρίσματος Κεφαλής**<sup>15</sup> (Uniform Head Marking) του Stump (1995). Συμμερίζεται την άποψη του Stump (1995: 249) ότι, δηλαδή, οποιαδήποτε ανωμαλία (αλλόμορφο ή υποκατάστατο) μίας ρηματικής ρίζας στο κλιτικό παράδειγμα θα εμφανισθεί και σε σύνθετες μορφές της ρίζας (21.), όπου η αλλομορφία του θέματος δεν επηρεάζεται από το παραγωγικό πρόθημα ή το πρώτο συνθετικό.

(21.) α. Κλίση	Παραγωγή	Λατινικά
regō – rexī	surgō <sup>16</sup> – surrēxī	

<sup>15</sup> Ο ορισμός της αρχής Uniform Head Marking αναφέρει ότι η μορφολογική κεφαλή μίας μορφολογικά πολύπλοκης λέξης κλίνεται πάντα με τον ίδιο τρόπο που κλίνεται η λέξη, όταν απαντάται μόνη της, π.χ. *come* ‘έρχομαι’: *came* ‘ήρθα’ / *became* ‘έγινα’.

<sup>16</sup> Είναι από το *sub-* + *regō*.

κυβερνώ – έχω κυβερνήσει	εξεγείρομαι – έχω εξεγερθεί	
ferre – tulī	conferre – contulī	
φέρνω – έχω φέρει	φέρνω μαζί – έχω φέρει μαζί	
β. Κλίση	Σύνθεση	Ελληνικά
τρώ-ω – έφαγα-α	καλοτρώ-ω – καλοέφαγα-α	
καφέ-ς – καφέδ-ες	παλιοκαφέ-ς – παλιοκαφέδ-ες	

Τέλος, δηλώνει πως οι παραδειγματικές σχέσεις ανάμεσα στις λέξεις φαίνεται να διαδραματίζουν σημαντικό ρόλο στον σχηματισμό κάποιων κανονικοτήτων στους τύπους της αλλομορφίας.

Όπως και αρκετοί ερευνητές, ο Booij (1997) αναφέρει ως αλλομορφία τις αλλαγές ενός μορφήματος ακόμα και όταν αυτές υφίστανται προβλέψιμα και συστηματικά μετά από εφαρμογή φωνολογικών κανόνων. Για να ενισχύσει την άποψή του χρησιμοποιεί παραδείγματα από άλλες μελέτες μειώνοντας τη δυναμική της μορφολογικής του προσέγγισης. Σημαντικό πρόβλημα είναι η έλλειψη πραγματικού διαχωρισμού της αλλομορφίας από την υποκατάσταση, αφού τα θεωρούμε διαφορετικά φαινόμενα. Καταθέτει εύστοχες σκέψεις και αναλύσεις σχετικά με τον παραδειγματικό χαρακτήρα της αλλομορφίας: εντούτοις δεν αξιοποιεί τα αποτελέσματα για να τεκμηριώσει περισσότερο την θεωρία του για την αλλομορφία, μίας και ο σκοπός τους είναι να ενισχύει τη θεωρία της αυτονομίας της μορφολογίας.

*Pirrelli & Battista (2000a, 2000b)*

Οι Pirrelli & Battista (2000a, 2000b) υποστηρίζουν ότι η αλλομορφία είναι το βασικό κριτήριο ταξινόμησης στη ρηματική κλίση των Ιταλικών και ο θεμέλιος λίθος της **παραδειγματικής ακεραιότητας** (paradigmatic integrity), η οποία ορίζει ότι κάθε θέση (slot) στο παράδειγμα καταλαμβάνεται από μία μόνο λεξική μορφή. Δημιουργούν έναν παραδειγματικό πίνακα που ονομάζεται **Σχήμα Συνολικής Κατανομής** (Overall Distribution Schema – ODS), όπου σχετίζονται τα βασικά θέματα με τα αλλομόρφά τους και τα μορφολογικά περιβάλλοντα που συμμετέχουν. Ο συγκεκριμένος πίνακας διαφοροποιείται ανάλογα με τα ρηματικά θέματα και κυρίως με τον αριθμό των αλλομόρφων που εμφανίζουν. Στον πίνακα εμφανίζονται επίσης θέματα υποκατάστασης και θέματα που προέκυψαν από ημιδιαφανείς ή ημιενεργούς φωνολογικούς κανόνες.

FINITE FORMS

	SINGULAR			PLURAL		
	1	2	3	1	2	3
PRESENT SUBJUNCTIVE	S <sub>2</sub>	S <sub>2</sub>	S <sub>2</sub>	S <sub>4</sub>	S <sub>4</sub>	S <sub>2</sub>
PRESENT INDICATIVE IMPERFECT	S <sub>2</sub>	S <sub>3</sub>	S <sub>3</sub>	S <sub>4</sub>	S	S <sub>2</sub>
PRESENT INDICATIVE IMPERFECT SUBJ.	S	S	S	S	S	S
PASSATO REMOTO	S <sub>5</sub>	S	S <sub>5</sub>	S	S	S <sub>5</sub>
PRESENT IMPERATIVE	-	S <sub>3</sub>	S <sub>2</sub>	S <sub>4</sub>	S	S <sub>2</sub>
PRESENT CONDITIONAL	S <sub>6</sub>	S <sub>6</sub>	S <sub>6</sub>	S <sub>6</sub>	S <sub>6</sub>	S <sub>6</sub>
FUTURE INDICATIVE	S <sub>6</sub>	S <sub>6</sub>	S <sub>6</sub>	S <sub>6</sub>	S <sub>6</sub>	S <sub>6</sub>

Πίνακας 1: Το Σχήμα Συνολικής Κατανομής για τον μέγιστο αριθμό θεμάτων – αλλομόρφων ενός ρήματος των Ιταλικών. Ως S καθορίζεται το βασικό προκαθορισμένο θέμα του ρήματος (από Pirrelli & Battista 2000a: 330).

Επισημαίνουν πως το ρηματικό σύστημα των ρομανικών γλωσσών έχει αλλάξει αρκετά λόγω της αλλομορφίας και σε κάποιες συζυγίες η αλλομορφία τείνει να εξαλειφθεί, ενώ σε άλλες αυξάνεται σημαντικά. Στον προβληματισμό για το ποιο θεωρείται το προκαθορισμένο βασικό θέμα ανάμεσα στις διαφορετικές μορφές που εμφανίζουν διάφορα θέματα στον ενεστώτα, δεν αποδέχονται τις προσεγγίσεις με τα φωνολογικά κριτήρια, ενώ αποδέχονται προσεγγίσεις με μορφολογικό υπόβαθρο. Υποστηρίζουν πως η μετάβαση από ένα κλιτικό παράδειγμα με πολλά αλλόμορφα σε ένα κλιτικό παράδειγμα με λιγότερα αλλόμορφα προκαλεί πάντα μία ασθενή ομοιομορφία. Δηλώνουν πως οι αλλαγές στα θέματα αποτελούν στην πλειονότητά τους προϊόντα των **ελάσσονων κανόνων** (minor rules), οι οποίοι είναι συγχρονικά αδιαφανείς και μη παραγωγικοί κανόνες. Σημειώνουν χαρακτηριστικά πως «*οι παραδειγματικές τάξεις πρέπει να θεωρούνται ως κατάλοιπα φωνολογικώς προερχομένων διαδικασιών που έχασαν τη φωνολογική τους διαφάνεια [...], η όλη διαδικασία δεν πρέπει να θεωρείται μορφολογικοποίηση (morphologization) ενός φωνολογικού κανόνα, αλλά ένα εμφανές σημάδι αλληλεπίδρασης μορφολογίας και φωνολογίας*» (Pirrelli & Battista 2000a: 372).

Είναι σαφές ότι οι Pirrelli & Battista (2000a, 2000b) αντιμετωπίζουν την αλλομορφία ως μορφολογικό φαινόμενο. Χρησιμοποιούν την αλλομορφία ως κριτήριο κατηγοριοποίησης των ρημάτων, ωστόσο εισάγουν αρκετά ρηματικά παραδειγματικά σχήματα, ενώ παράλληλα εντοπίζονται περιπτώσεις με αρκετές παρεκκλίσεις. Δεν

παίρνουν ξεκάθαρη θέση εάν υπάρχει σειρά εφαρμογής των αλλομορφικών κανόνων σε αντίθεση με την προγενέστερη έρευνα του Wanner (1972).<sup>17</sup>

*Galani (2003a, 2003b)*

Η Galani (2003a, 2003b) εξετάζει τη ρηματική κλίση της Νέας Ελληνικής προτείνοντας την ύπαρξη αλλομορφίας των **θεματικών φωνηέντων** (theme vowels).<sup>18</sup> Τα θεματικά φωνήεντα της Galani προσδιορίζουν τις ρηματικές κλιτικές τάξεις και είναι φορείς όψης. Σημειώνει ότι τα θεματικά φωνήεντα μπορούν σε κάποιες περιπτώσεις να είναι μηδενικά μορφήματα ή να αποτελούν μέρος των κλιτικών επιθημάτων. Ο τρόπος με τον οποίο πραγματώνονται τα θεματικά φωνήεντα εξαρτάται από τα υπόλοιπα μορφήματα και το περιβάλλον συμμετοχής (22.).

- (22.) (α.) τιμ(STEM) – η(PERFECTIVE) – σ(ACTIVE) – α(1SG.PAST)  
 (β.) καλ(STEM) – ε(PERFECTIVE) – στηκ(NON-ACTIVE) – α(1SG.PAST)  
 (γ.) αφαιρ(STEM) – ου(IMPERFECTIVE) – μαι(1SG.PRESENT.NON-ACTIVE)  
 (δ.) ε(AUGMENT) – γραφ(STEM) – Ø(PERFECTIVE) – σ(ACTIVE) – α(1SG.PAST)
- (από Galani 2003a: 2)

Αντιτίθεται στις υπάρχουσες προσεγγίσεις θεωρώντας ότι η αλλομορφία είναι μία παραγωγική διαδικασία που λαμβάνει χώρα πριν τον σχηματισμό των λέξεων. Τα θεματικά φωνήεντα είναι κατασκεύασμα της συγγραφέως και η κατανομή τους είναι αρκετά πολύπλοκη. Θεωρεί ως προκαθορισμένο δείκτη το /o/, π.χ.: γράφ-ω, γράφ-ο-μαι, και ως αλλομορφα όλες τις υπόλοιπες φωνηεντικές πραγματώσεις των κλιτικών επιθημάτων (δηλαδή τα -ει-ς, -ει, -ου-με, -ε-τε, -ου-ν), οι οποίες κατανέμονται σύμφωνα με τα μορφολογικά και συντακτικά χαρακτηριστικά. Ακολουθώντας κάποιες παραδοχές της Lieber (1982) για την αλλομορφία, προσεγγίζει το πρόβλημα με βάση την Κατανεμημένη Μορφολογία<sup>19</sup> (Distributed Morphology) και υποστηρίζει ότι στο σχηματισμό λέξεων πρέπει να αλληλοεπιδρούν η σύνταξη, η μορφολογία και η

<sup>17</sup> Ο Wanner (1972) εκφράζει σοβαρές αμφιβολίες για την ύπαρξη σειράς στους κανόνες σχηματισμού αλλομόρφων, καθότι θεωρεί ότι αυτοί εφαρμόζονται αυθαίρετα.

<sup>18</sup> Αυτά που ονομάζει θεματικά φωνήεντα είναι φωνήεντα που εφευρίσκει η Galani και δεν εννοεί τα θεματικά φωνήεντα που υπάρχουν στα ΑΕ. Τα συγκεκριμένα φωνήεντα θα μπορούσαν να χαρακτηριστούν ως απλοί δείκτες.

<sup>19</sup> Η συγκεκριμένη προσέγγιση ονομάζεται έτσι διότι ο σχηματισμός λέξεων κατανέμεται σε διάφορους τομείς της γραμματικής (σύνταξη, φωνολογία και σημασιολογία), μίας και δεν αποδέχεται την ύπαρξη αυτόνομου πεδίου της μορφολογίας (για τις βασικές αρχές, πρβλ. Halle & Marantz 1993).

Συναφείς έρευνες για την αλλομορφία βασισμένες στο μοντέλο της Κατανεμημένης Μορφολογίας για την ανάλυση της αλλομορφίας είναι των Bobaljik (2000), Lahne (2006), Bailyn & Nevins (2008), Harizanov & Gribanova (2010). Οι συγκεκριμένες μελέτες δεν θα αναπτυχθούν, αφού δεν σχετίζονται με τους στόχους της παρούσας διατριβής.

φωνολογία. Σημειώνει ότι τα θεματικά φωνήεντα παρέχουν στο θέμα και στο κλιτικό επίθημα τις πληροφορίες της κλιτικής τάξης και πρέπει να ταυτίζονται οι κλιτικοί δείκτες θέματος και επιθήματος, για να συνδυαστούν. Τα θεματικά φωνήεντα είναι μαρκαρισμένα στο λεξικό, για να αποφευχθούν αντιγραμματικές λέξεις και υποστηρίζει πως το ρηματικό σύστημα της ΚΝΕ είναι οργανωμένο με βάση τη μαρκαρισσιμότητα και την ιεραρχία των θεματικών φωνηέντων, των οποίων η αλλομορφία κάνει το σύστημα πιο οικονομικό.

Σύμφωνα με το μοντέλο της Ralli (1988, 2005), τονίζουμε ότι δεν αποδεχόμαστε την ύπαρξη θεματικών φωνηέντων στην ΚΝΕ, καθότι τα θεματικά φωνήεντα δεν ενέχουν τις ιδιότητες και τα χαρακτηριστικά που είχαν στα ΑΕ, ενώ τα συγκεκριμένα αποτελούν αυθαίρετα θεωρητικά κατασκευάσματα της Galani. Ο διαχωρισμός των κλιτικών επιθημάτων παρουσιάζει πάρα πολλά προβλήματα, αφού συχνά τα όρια τους είναι ασαφή και μη συστηματικά (22.), όπως και οι μορφολογικές πληροφορίες που φέρουν. Προτείνει ένα θεματικό φωνήεν (/o/), το οποίο εμφανίζει πολλά αλλόμορφα χωρίς καμία δικαιολογία, δημιουργώντας μία προσέγγιση που ταιριάζει περισσότερο στα Αρχαία Ελληνικά. Εμφανίζει, για παράδειγμα, το θεματικό φωνήεν –ou– (/u/) που είναι πολλαπλός δείκτης με διαφορετικές μορφολογικές ιδιότητες ανάλογα με το περιβάλλον σύνδεσης (23.).

- (23.) (α.) δείκτης ΜΗ-ΣΥΝΟΠΤΙΚΟ.ΕΝΕΡΓΗΤΙΚΗ στο ‘καλούμε’,  
 (β.) δείκτης ΣΥΝΟΠΤΙΚΟ.ΕΝΕΡΓΗΤΙΚΗ στο ‘καλούσα’  
 (γ.) δείκτης ΜΗ-ΣΥΝΟΠΤΙΚΟ.ΜΗ-ΕΝΕΡΓΗΤΙΚΗ στο ‘καλούμαι’

Το κυριότερο πρόβλημα της ανάλυσης είναι ο διαχωρισμός των μορφημάτων και η απουσία συζήτησης για τη θεματική αλλομορφία. Παραδείγματος χάριν, αναφέρεται στα κλιτικά επιθήματα –τηκ– και –θηκ– ως ξεχωριστά, αγνοώντας έτσι απλούς κανόνες της Λεξικής Φωνολογίας. Παρουσιάζει το κλιτικό επίθημα –στηκ– (καλέστηκα), το οποίο ανήκει στην κλιτική κατηγορία των ρημάτων με θεματικό φωνήεν το –ε–, δηλ. υποθέτει ως υπαρκτό τύπο το αφαιρέστηκα και όχι το αφαιρέθηκα(!). Η Galani παρουσιάζει ένα ρηματικό σύστημα με θεματικά φωνήεντα και επί της ουσίας εκμηδενίζει την αλλομορφία των ρηματικών θεμάτων. Για παράδειγμα, το πλένω αναλύεται ως πλ-έν-ω και έ-πλ-υν-α σε ενεστώτα και αόριστο και τα αλλόμορφα πλεν~πλυν δεν υφίστανται, γιατί το θεματικό στοιχείο –εν– έχει το αλλόμορφο –υν–. Κλείνοντας, να αναφέρουμε πως δεν εισάγει κανέναν αλλομορφικό ή μορφολεξικό

κανόνα και δεν καταφέρνει να αποδείξει πως η πρότασή της είναι πιο οικονομική από τις άλλες θεωρητικές προτάσεις.

Anderson (2008)

Ο Anderson (2008) βασιζόμενος στη θεωρία του για μια μορφολογία χωρίς μορφήματα (A-Morphous Morphology, βλ. Anderson 1992) αναλύει τη φωνολογικώς εξαρτημένη αλλομορφία της Surmiran (Rumantsch).<sup>20</sup> Κύριος στόχος του είναι να δικαιολογήσει την παρουσία της αλλομορφίας, αλλά πρωτίστως την επιλογή του κατάλληλου αλλομόρφου σε κάθε μορφολογικό περιβάλλον με φωνολογικά κριτήρια, χρησιμοποιώντας και τη Θεωρία του Βέλτιστου σε θεωρητικό επίπεδο. Υποστηρίζει ότι η αλλομορφία είναι κάθε αλλαγή του θέματος που δεν είναι αποτέλεσμα ενεργού φωνολογικού κανόνα και καταδεικνύει πως οι αλλαγές θεμάτων στα Surmiran δεν υπόκεινται σε κανόνες (24.)

(24.) *eir* ‘πηγαίνω’ *neir* ‘έρχομαι’ (*vu*)*leir* ‘θέλω’ *deir* ‘λέω’ *star* ‘ζω’ *saveir* ‘ξέρω’

1EN	<i>vign</i>	<i>vign</i>	<i>vi</i>	<i>dei</i>	<i>stung</i>	<i>sa</i>
2EN	<i>vast</i>	<i>vignst</i>	<i>vot</i>	<i>deist</i>	<i>stast</i>	<i>sast</i>
3EN	<i>vo</i>	<i>vign</i>	<i>vot</i>	<i>dei</i>	<i>stat</i>	<i>so</i>
1ΠΛ	<i>giagn</i>	<i>nign</i>	<i>lagn</i>	<i>schagn</i>	<i>stagn</i>	<i>savagn</i>
2ΠΛ	<i>gez</i>	<i>niz</i>	<i>lez</i>	<i>schez</i>	<i>stez</i>	<i>savez</i>
3ΠΛ	<i>von</i>	<i>vignan</i>	<i>vottan</i>	<i>deian</i>	<i>stattan</i>	<i>son</i>

(από Anderson 2008: 33-34)

Στο Λεξικό καταχωρεί ένα μοναδικό θέμα (24.) χωρίς μορφοσυντακτικά χαρακτηριστικά, ενώ όσες λέξεις έχουν αλλομορφο καταχωρούνται με συγκεκριμένους συνδυασμούς ιδιοτήτων, δηλαδή ως δέσμες χαρακτηριστικών που καθορίζουν τα περιβάλλοντα συμμετοχής τους, όπως τα [YOU], [PAST], [PLURAL], μεταξύ άλλων. Υπογραμμίζει πως η επιλογή αλλομόρφου κυβερνάται φωνολογικά από τη θέση του τόνου σε κάποιες ρηματικές συζυγίες. Αποδεικνύει πως η αλλομορφία στη Surmiran είναι αποτέλεσμα ιστορικών αλλαγών και ανενεργών φωνολογικών κανόνων, με αποτέλεσμα να υπάρχει ένα φωνολογικό σχήμα δόμησης που να δικαιολογεί την

<sup>20</sup> Γλώσσα της ομάδας Rumantsch που ομιλείται στις κοιλάδες Surmeira και Albula της Ελβετίας.

επιλογή αλλομόρφου, αλλά και να καθορίζει τον τύπο αλλομορφίας (εναλλαγή φωνηέντων ή συμφώνων). Μάλιστα αποκλείει την πιθανότητα εμπλοκής των μορφολογικών κατηγοριών στην όλη διαδικασία. Επισημαίνει πως και η υποκατάσταση είναι φωνολογικά εξαρτώμενη, καθότι οι τύποι που είναι κενοί και υποκαθίστανται, απουσιάζουν για φωνολογικούς λόγους, ενώ παράλληλα τα ρήματα που συμπληρώνουν αυτά τα κενά έχουν αρκετές φωνολογικές ομοιότητες.

Ο Anderson (2008) αποδεχόμενος ότι η αιτία ύπαρξης της αλλομορφίας είναι κανόνες φωνολογικοί που δεν είναι ενεργοί, υπογραμμίζει ορθώς ότι υπάρχουν γλώσσες όπου τα αλλόμορφα επιλέγονται με φωνολογικά κριτήρια. Δεν εμπλέκει καθόλου μορφολογικά κριτήρια στις διαδικασίες επιλογής του κατάλληλου αλλομόρφου. Επίσης, αποφεύγει να παρουσιάσει παραδείγματα από τη διαδικασία παραγωγής και σύνθεσης της γλώσσας, ώστε να αποδείξει με δεδομένα των υπόλοιπων διαδικασιών σχηματισμού λέξεων ότι ο τόνος είναι ο μόνος λόγος επιλογής αλλομόρφου. Επιπροσθέτως, δεν παίρνει ξεκάθαρη θέση για τον τρόπο καταχώρησης των θεμάτων στο Λεξικό, αλλά παραθέτει τα προβλήματα διαφόρων προσεγγίσεων. Τέλος, να σημειωθεί πως δεν αξιοποιεί το φαινόμενο για να εξάγει κάποια συμπεράσματα για τη μορφολογία της γλώσσας, αλλά για να αναλύσει γενικότερα τη γραμματική μιας γλώσσας.

### 2.3. Το μορφολογικό μοντέλο της Ράλλη για την Αλλομορφία

Στην παρούσα εργασία η περιγραφή, ανάλυση και υπολογιστική επεξεργασία του φαινομένου της αλλομορφίας στην παραγωγή υιοθετεί τις βασικές αρχές του μοντέλου της Ράλλη, όπως αυτό έχει παρουσιαστεί σε διάφορες μελέτες. Η Ράλλη (Ralli 1994, 1999, 2000, 2007, Ράλλη 2005) θεωρεί ότι η αλλομορφία συμμετέχει συστηματικά σε όλες τις διαδικασίες σχηματισμού λέξεων. Υποστηρίζει ότι οι περιπτώσεις παραλλαγών του ίδιου μορφήματος που δεν ερμηνεύονται φωνολογικά, αποτελούν φαινόμενα «καθαρής» αλλομορφίας και εισάγει τον όρο **μορφολογική αλλομορφία** για τις συγκεκριμένες παραλλαγές (25.α-β). Αντιθέτως, για τις φωνολογικές αλλαγές μορφημάτων θεωρεί ότι δεν θα έπρεπε να τους αποδίδεται ο όρος αλλομορφία, δεδομένου ότι υπάρχει φωνολογική ερμηνεία για τις μεταβολές της μορφής (25.γ).

- (25.) α. σωμα ~ σωματ  
παπα ~ παπαδ



β.	φεύγ-ω	έ-φυγ-α	θα φύγ-ω
	παίρν-ω	πήρ-α	θα πάρ-ω
	φέρν-ω	έ-φερ-α	θα φέρν-ω
	αγαπ-ώ	αγάπα-γ-α	αγαπ-ούσ-α, αγάπη-σ-α, θα αγαπή-σ-ω
γ.	γράφ-ω	έ-γραψ-α	/e-ɣrap-sa/
	τρέχ-ω	έ-τρεξ-α	/e-trek-sa/

(από τη Ράλλη 2005: 68-69)

Η εισαγωγή του όρου της μορφολογικής αλλομορφίας έναντι του όρου της λεξικής αλλομορφίας (Carstairs 1987), γίνεται επειδή ο όρος «λεξικός/ή/ό» παραπέμπει σε στοιχεία με αδιαφανή δομή ή χωρίς μορφολογική δομή. Για παράδειγμα, η Ράλλη παρατηρεί πως στην κλίση των ουσιαστικών εμφανίζεται συχνά ένα φώνημα (/t/ ή /ð/) ανάμεσα στο θέμα και το κλιτικό επίθημα, για το οποίο δεν υπάρχει καμία φωνολογική ερμηνεία συγχρονικά· τα συγκεκριμένα φωνήματα τα εντάσσει στο θέμα και τις μορφές αυτές τις ονομάζει αλλόμορφα (25.α).

Η αλλομορφία κατά τη Ράλλη (2005) έχει διαχρονική εξήγηση, αφού οφείλεται σε φωνολογικές αλλαγές που δεν ισχύουν ή σε ανενεργούς φωνολογικούς κανόνες, όπως για παράδειγμα ο κανόνας της συναίρεσης. Επίσης, μπορεί να είναι αποτέλεσμα λεξικών δανεισμών που έλαβαν χώρα σε διαφορετικές φάσεις εξέλιξης της γλώσσας. Πιο συγκεκριμένα, σύμφωνα με τη Ralli (1988), ο κανόνας της συναίρεσης, δηλαδή η συγχώνευση του ληκτικού φωνήεντος του θέματος με το αρχικό του κλιτικού επιθήματος σε ένα φωνήεν, δημιούργησε τη μορφή *αγαπα-* (25.β), η οποία θεωρείται αλλόμορφο, αφού είναι αποτέλεσμα εφαρμογής ενός κανόνα της ΑΕ που είναι ανενεργός. Δηλώνει πως το νέο θεματικό αλλόμορφο *αγαπα-* έχει δημιουργηθεί στα Νέα Ελληνικά και στη συνέχεια επεκτάθηκε σε όλο το παράδειγμα λόγω του φαινομένου της **παραδειγματικής ομοιομορφίας** (paradigmatic uniformity). Για παράδειγμα, έχουν επαναναλυθεί όλοι οι τύποι του ενικού (*αγαπάω/ αγαπώ, αγαπάς, αγαπά*) και του πληθυντικού (*αγαπάμε/ αγαπούμε, αγαπάτε, αγαπάν/ αγαπούν*). Οι σχηματισμοί αυτοί μετά την εφαρμογή του κανόνα της συναίρεσης, επαναναλύθηκαν συγχρονικά σε νέο θέμα με την ύπαρξη αλλομόρφου και σε νέο κλιτικό επίθημα (26.).

(26.)	αγαπά-εις	→	αγαπά-ς
	αγαπά-ει	→	αγαπά
	αγαπά-ετε	→	αγαπά-τε

Υιοθετεί το μοντέλο της Lieber (1982) σχετικά με τους αλλομορφικούς κανόνες και την αναπαράστασή τους στο Λεξικό. Δηλώνει τις συσχετιζόμενες μορφές ενός μορφήματος με πλεοναστικούς κανόνες (redundancy rules), δηλαδή κανόνες που δεν δημιουργούν λεξικά στοιχεία. Ειδικότερα, κάνει χρήση των μορφολεξικών κανόνων, για να δηλώσει τις σχέσεις αλλομορφίας μεταξύ των διαφορετικών μορφών ενός λήμματος στο Λεξικό. Υπογραμμίζει ότι ο περιορισμός της εμφάνισης του αλλομόρφου σε συγκεκριμένα μορφολογικά περιβάλλοντα αποτελεί απαραίτητη πληροφορία, η οποία είναι μη προβλέψιμη και για αυτό πρέπει να καταχωρηθεί στο Λεξικό. Η πρόταση της Ralli (1988) είναι η ένταξη της μορφολογικής πληροφορίας της αλλομορφικής εξάρτησης στο μορφολεξικό κανόνα (27.).

- (27.) παπα/ \_\_\_\_\_ ~ παπαδ  
                   [ενικός]  
       φευγ/ \_\_\_\_\_ ~ φυγ  
                   [μη συνοπτικό]

(από Ράλλη 2005: 83)

Για τις μορφολογικές πληροφορίες του μορφολεξικού κανόνα η Ράλλη (ό.π.) διευκρινίζει ότι σε μορφολογικά πολύπλοκους σχηματισμούς, οι οποίοι σχηματίστηκαν στη σύνθεση ή στην παραγωγή, οι μορφολεξικοί κανόνες δεν έχουν εφαρμογή. Για παράδειγμα, αναφέρει πως ο μορφολεξικό κανόνας ενώ ισχύει για το ρήμα *πέφτω* (*πεφτ~ πεσ*) δεν μπορεί να εφαρμοστεί για το προθηματοποιημένο ρήμα *ξεπέφτω* (*ξεπεφτ~ ξεπεσ*). Το κενό που προκύπτει από την έλλειψη της πληροφορίας του μορφολογικού περιβάλλοντος συμμετοχής των αλλομόρφων στη διαδικασία παραγωγής και σύνθεσης, θα το αντιμετωπίσουμε στην αντίστοιχη ενότητα της δικής μας πρότασης (2.4.2).

Η αλλομορφία στην κλίση έχει ενεργό ρόλο τόσο στην παραδειγματική οργάνωση, όσο και στην παραδειγματική αναδόμηση (Ralli 2007). Προτείνει το διαχωρισμό των ρηματικών και ονοματικών κλιτικών τάξεων με βάση την παρουσία ή απουσία αλλομορφίας, καθώς και με βάση τον τύπο αλλομορφικής ποικιλίας. Εντούτοις δεν καταχωρεί τις ρηματικές κλιτικές τάξεις στο λεξικό, όπως πράττει με τις ονοματικές, μίας και η ίδια η ρηματική αλλομορφία ορίζει την κλιτική τάξη.

Διαχωρίζει τα ουσιαστικά σε οκτώ κλιτικές τάξεις σύμφωνα με δύο κριτήρια (Ralli 2000, Ράλλη 2005: 118):

- (1) τη συστηματική αλλομορφική ποικιλία των θεμάτων και

(2) τη μορφή του συνόλου των κλιτικών επιθημάτων που συνδέονται με αυτά.

Όπως παρατίθεται (Πίνακας 8) στην ενότητα 2.4.3.2, οι δύο πρώτες κλιτικές τάξεις περιλαμβάνουν τα αρσενικά ουσιαστικά, οι ακόλουθες δύο τα θηλυκά και τέλος, στις υπόλοιπες τέσσερις εντάσσονται τα ουδέτερα. Παραδείγματος χάριν, το ουσιαστικό *άνθρωπ-ος* εντάσσεται στην πρώτη κλιτική τάξη και δεν έχει κάποιο αλλόμορφο, ενώ αντίθετα το ουσιαστικό *παπά-ς* και *καφέ-ς* εμφανίζουν την ίδια αλλομορφική ποικιλία, τύπου  $X \sim X\delta$ , όπου  $X$  = μεταβλητή θέματος ( $X$  = *παπα*, *καφε* και  $X\delta$  = *παπαδ*, *καφεδ*). Τα ίδια κριτήρια κατηγοριοποίησης σε κλιτικές τάξεις ισχύουν και για τα επίθετα που σε γενικές γραμμές προσαρμόζουν τα κλιτικά τους παραδείγματα στο πρότυπο των ουσιαστικών, με εξαιρέσεις τα επίθετα τύπου *ευγενής/ ευγενές*, που δημιουργούν δύο νέες κλιτικές τάξεις. Οι συγκεκριμένες επιθετικές κλιτικές τάξεις δεν εμφανίζουν κάποιο αλλόμορφο, αλλά συνδυάζονται με κλιτικά επιθήματα, κατάλοιπα των ΑΕ.

Η ανάλυση της Ralli (1988) για τη δομή της ρηματικής κλίσης προσπαθεί να αποφύγει την υπερβολική χρήση μηδενικών κλιτικών επιθημάτων. Οι ρηματικοί τύποι αναλύονται ως διφυείς δομές, με πρώτο συστατικό το ρηματικό θέμα και δεύτερο το κλιτικό επίθημα. Τα ρήματα διαχωρίζονται σε δύο κλιτικές τάξεις με βάση την παρουσία συστηματικής αλλομορφίας ή όχι. Το χαρακτηριστικό της κλιτικής τάξης στα ρήματα είναι προβλέψιμο από την παρουσία ή απουσία της συστηματικής αλλομορφίας. Κρίνεται έτσι περιττή η καταχώρηση αυτής της πληροφορίας στο λήμμα του ρήματος. Σύμφωνα με τη Ralli (1988, 2005), τα ρήματα με αλλομορφία τύπου  $X(\alpha) \sim XV_{\{[i]/[e]\}}$ , όπως το *αγαπώ* και *διαιρώ*, εντάσσονται στη δεύτερη κλιτική τάξη, ενώ όσα δεν υπακούουν στον παραπάνω κανόνα στην πρώτη κλιτική τάξη. Ειδικότερα, οι κλιτικές τάξεις παρουσιάζονται στην ενότητα 2.4.3.3 (Πίνακας 9), όπου αναλύουμε την αλλομορφία ρημάτων στην κλίση.

Συμφωνώντας με τις απόψεις των Mel'čuk (1976) και Carstairs (1987), η Ράλλη θεωρεί ότι η αλλομορφία είναι χαρακτηριστικό και των προσφυμάτων, αν και πιστεύει ότι είναι περιορισμένης έκτασης, επειδή τα προσφύματα αποτελούν κλειστή τάξη (closed-set items). Περισσότερο συχνή η εμφάνιση της αλλομορφίας είναι στα ονοματικά παραγωγικά επιθήματα, όπως το  $-τζη- \sim -τζηδ-$  (28.α), όσο και στα ρηματικά επιθήματα, όπως το  $-αρ- \sim -αρη-$  (28.β). Παρατηρεί εύστοχα ότι στα επιθήματα εντοπίζονται αλλομορφικές συμπεριφορές που παραπέμπουν σε αντίστοιχες των ουσιαστικών, όπως για παράδειγμα η αλλομορφία του  $-τζη(ς)$  είναι παρόμοια με αυτή του *παπά-ς*  $\sim$  *παπάδ-ες*.

- (28.) α. -τζη- ~ -τζηδ-  
 γκαφα-τζή-ς / γκαφα-τζήδ-ες  
 ταξι-τζή-ς / ταξι-τζήδ-ες
- β. -αρ- ~ -αρη-  
 τρατ-άρ-ω / τρατ-άρη-σ-α  
 σουτ-άρ-ω / σουτ-άρη-σ-α

Εξετάζοντας συνδυαστικά την αλλομορφία με την παραδειγματική ομοιομορφία στις νεοελληνικές διαλέκτους, καταδεικνύει τη σημαντικότερη συμμετοχή που έχει η αλλομορφία στη διαδικασία της κλίσης. Υποστηρίζει πως οι αλλομορφικές αλλαγές σε μερικές περιπτώσεις συμβάλλουν στην παραδειγματική ομοιομορφία και τη γραμματική απλοποίηση, ενώ υπάρχουν και περιπτώσεις όπου λειτουργεί ανεξάρτητα. Για παράδειγμα, τα ουσιαστικά της ΚΝΕ έχουν στον πληθυντικό το κλιτικό επίθημα -α σε ονομαστική, αιτιατική και κλιτική, όπως *παιδί-παιδιά, μωρό-μωρά, κύμα-κύματα*, εκτός από τα ουσιαστικά σε -ος, όπως *λάθος-λάθη*. Ωστόσο τα διαλεκτικά δεδομένα, για παράδειγμα στα Λεσβιακά τα ουσιαστικά της τάξης *λάθος-λάθη* τείνουν να ομοιοποιήσουν το κλιτικό τους παράδειγμα εμφανίζοντας τους τύπους *λάθη-α* και *λάθητ-α*, όπως υπάρχουν κατάλοιπα στην ΚΝΕ *στήθια/στήθη, χείλια/χείλη* μεταξύ άλλων. Με βάση τα διαλεκτικά δεδομένα κάνει την υπόθεση ότι η παραδειγματική ομοιομορφία ενεργοποιεί τη μετατόπιση των μορφηματικών ορίων με αποτέλεσμα την εμφάνιση νέας αλλομορφίας. Εντούτοις, επισημαίνει πως στα παράγωγα ουσιαστικά η παρουσία αλλομορφίας παρέχει σημαντική υποστήριξη στη διατήρηση της δομής του παραγωγικού επιθήματος, ώστε να αντισταθεί στην αναδόμηση (29.).

(29.) α. ΚΝΕ

ΕΝ ΟΝ	pso'ma-s	ΠΛ ΟΝ/ ΑΙΤ/ ΚΛ	pso'mað-es
ΓΕΝ/ ΑΙΤ/ ΚΛ	pso'ma	ΓΕΝ	pso'mað-on

β. Λεσβιακά

ΕΝ ΟΝ	psu'ma-s	ΠΛ ΟΝ/ ΑΙΤ/ ΚΛ	psu'mað-is (*psu'm-i)
ΓΕΝ/ ΑΙΤ/ ΚΛ	psu'ma	ΓΕΝ	-

(από Ralli 2007: 23)

Σημειώνει ότι τα παραγωγικά επιθήματα εμφανίζουν υψηλές αντιστάσεις στην παραδειγματική ομοιομορφία, δηλαδή στην πίεση να ενταχθούν σε άλλη κλιτική τάξη. Αναφέρει πως μία εύλογη εξήγηση είναι ότι η αλλομορφία αποτελεί σημαντικό

κομμάτι της ουσίας του επιθήματος και επομένως ιδιαίτερη σημαντική και υποστηρίζει ότι η αλλομορφία βοηθά στην ενίσχυση της λεξικής αναπαράστασης του επιθήματος με την αντίσταση στην αλλαγή τάξης. Όπως ορθά επισημαίνει η Ralli (2007), συμβαίνει το παράδοξο η αλλομορφία να είναι σε αυτές τις περιπτώσεις τροχοπέδη στην παραδειγματική ομοιομορφία. Αντιστέκεται, λοιπόν, στις τάσεις παραδειγματικής εξομάλυνσης παρόλο που οι συνθήκες δεν διαφοροποιούνται σημαντικά από τις αντίστοιχες της θεματικής αλλομορφίας, όπου διαδραματίζει σημαντικό ρόλο στην επίτευξη της ομοιομορφίας.

Συνοπτικά για την αλλομορφία η Ralli (2007: 26) αναφέρει πως η «[...] η αλλομορφία μπορεί να μας πει τον τρόπο οργάνωσης των κλιτών λέξεων σε παραδείγματα, συμβάλλει στη γραμματική απλοποίηση, ενώ επίσης βοηθά τα θέματα και τα παραγωγικά προσφύματα να αντισταθούν στην εξομάλυνσή τους, όταν η διατήρηση δομής κινδυνεύει. [...] η αλλομορφία εμφανίζει μία ιδιαίτερη ανεξαρτησία που μπορεί να προκύψει σε περιπτώσεις παραδειγματικής επαναδόμησης ακόμα και ενάντια στη γραμματική απλοποίηση».

Τέλος, η Ράλλη (2005) αναφέρει και περιπτώσεις αλλομορφίας που προκύπτουν από τον δανεισμό μεταξύ των διαφόρων γλωσσών. Ένα παράδειγμα από τα Ελληνικά δημιουργίας αλλομορφίας λόγω δανεισμού δίνεται στις ακόλουθες λέξεις, όπου η έννοια «σερβίρω» εμφανίζεται στο ρήμα και στα παράγωγά του με τις μορφές *σερβιρ-* και *σερβι-* αντιστοίχως, ανάλογα με τη γλώσσα από την οποία προέρχονται:

- (30.)    *σερβίρ-ω*    (‘servir’, δάνειο από τα Γαλλικά)  
           *σερβι-τόρος* (‘servitore’, δάνειο από τα Ιταλικά)  
           *σερβί-τσιο* (‘servizio’, δάνειο από τα Ιταλικά)

(από Ράλλη 2005: 71)

Σημειώνεται ότι το ρηματικό θέμα *σερβιρ-* εμφανίζεται με το αλλόμορφο *σερβιρι-* σε τύπους, όπως στον αόριστο (*σερβίρισα*), αλλά και ως *σερβιρισ-* σε παράγωγα, όπως *σερβίρισμα* και *σερβιρισμένος* κτλ.

## 2.4. Καθορισμός του θεωρητικού πλαισίου της Αλλομορφίας

### 2.4.1. Ορισμός της Αλλομορφίας

Παρουσιάζονται αρκετές διαφοροποιήσεις του ορισμού στη διεθνή βιβλιογραφία ανάλογα με τη θεωρία που προτείνεται, ή τη γλώσσα που γίνεται αντικείμενο μελέτης

κάθε φορά. Είναι συχνό φαινόμενο η υπερβολική χρήση του όρου αλλόμορφο για κάθε μορφολογική και φωνολογική αλλαγή. Εδώ, κρίνεται απαραίτητος ο ορισμός της έννοιας του αλλομόρφου, καθώς ο χαρακτηρισμός του δεν είναι απλός όσο αυτός του αλλοφώνου. Βασιζόμενοι αρχικά στον ορισμό της Lieber (1982: 27) ορίζουμε ως **αλλόμορφα** τις διαφορετικές ποικιλίες του ιδίου μορφήματος, οι οποίες μοιράζονται εξίσου λεξική πληροφορία, όσο και σημασιολογική αναπαράσταση και δομή, διαφέρουν όμως απρόβλεπτα και αυθαίρετα στη φωνολογική τους μορφή και στα μορφολογικά περιβάλλοντα, στα οποία εμφανίζονται. Στο (31.α) δίνονται λέξεις που εμφανίζουν αλλόμορφα που πληρούν τις προϋποθέσεις του παραπάνω ορισμού. Για παράδειγμα τα ουσιαστικά *παπάς* και *ψυχή* έχουν τα αλλόμορφα *παπα~ παπαδ* και *ψυχη~ ψυχ*, ενώ το ρήμα *παίρνω* έχει τα *παιρν~ παρ~ πηρ*.

Αντιθέτως, εισάγουμε τον όρο **φωνόμορφα**, για να διαφοροποιούμε τις περιπτώσεις της αλλομορφίας από τις προβλέψιμες αλλαγές μορφής, οι οποίες αποτελούν για τους περισσότερους ερευνητές επίσης αλλόμορφα ή φωνολογικώς εξαρτημένα αλλόμορφα. Ο όρος είναι ένας συμφυρμός των λέξεων φωνολογικό αλλόμορφο. Ορίζουμε ως φωνόμορφα τις διαφορετικές ποικιλίες του ιδίου μορφήματος, οι οποίες μοιράζονται την ίδια λεξική πληροφορία, όσο και σημασιολογική αναπαράσταση, διαφέρουν προβλέψιμα στη μορφή τους εξαιτίας της εφαρμογής φωνολογικού κανόνα. Στο παράδειγμα (31.β) η μορφή *έτρεξα* (e-trek-s-a) του ρήματος *τρέχω* έχει φωνολογική ερμηνεία λόγω του κανόνα της ανομοίωσης· το ίδιο ισχύει για τις μορφές *έγραψα* (e-γrap-s-a) και *έραψα* (e-rap-s-a) των ρημάτων *γράφω* και *ράβω* αντίστοιχα.

(31.)	α.	παπάς	παπα	~	παπαδ	
		ψυχή	ψυχη	~	ψυχ	
		παίρνω	παιρν	~	παρ	~ πηρ
		φιλώ	φιλ(α)~		φιλη	
	β.	τρέχω	έτρεξα		/e-trek-s-a/	
		γράφω	έγραψα		/e-γrap-s-a/	
		ράβω	έραψα		/e-rap-s-a/	

Επεκτείνοντας τον αρχικό ορισμό της Lieber, θεωρούμε ότι η αλλομορφία είναι η μελέτη των διαφορετικών παραλλαγών ενός μορφήματος, που μοιράζονται λεξική πληροφορία και σημασιολογική αναπαράσταση· ωστόσο διαφέρουν συγχρονικά απρόβλεπτα και αυθαίρετα στη φωνολογική τους μορφή, βρίσκονται σε

συμπληρωματική κατανομή και η μορφή τους δεν απορρέει από την εφαρμογή κάποιου συγχρονικού φωνολογικού κανόνα. Υποστηρίζουμε ότι είναι μία διαδικασία που χαρακτηρίζεται από κανονικότητα και προβλεψιμότητα, όπως θα αποδείξουμε σε ακόλουθο κεφάλαιο, στο πείραμα του υπολογιστικού μοντέλου, καθότι στις περισσότερες περιπτώσεις η αλλομορφία προήλθε από φωνολογικούς κανόνες που συγχρονικά δεν είναι ενεργοί. Τα επιθήματα επίσης ακολουθούν τις αλλομορφικές συμπεριφορές των θεμάτων. Ταυτόχρονα φαίνεται πως οι ανενεργοί κανόνες, σε συνδυασμό με διάφορες μορφοφωολογικές πληροφορίες, μπορούν *υπολογιστικά* να συνδράμουν στην πρόβλεψη της αλλομορφίας.

#### 2.4.1.1. Είδη αλλομορφίας

Στα Ελληνικά τόσο η **θεματική αλλομορφία** (stem allomorphy), όσο και η **προσφυματική αλλομορφία** (affixal allomorphy) εμφανίζουν κοινές συμπεριφορές σχετικά με τον τύπο της αλλομορφικής αλλαγής του εξεταζόμενου μορφήματος. Έτσι, ορίζουμε τα διαφορετικά είδη αλλομορφίας που παρατίθενται παρακάτω με βάση τον τύπο της αλλομορφικής ποικιλίας και παραθέτουμε τους όρους περιγραφής. Για αυτό εισάγουμε νέους όρους για την κατηγοριοποίηση των διαφορετικών μορφών αλλομορφίας, δίνοντας περιγραφικά παραδείγματα για το κάθε είδος αλλομορφίας. Οι νέοι όροι που προτείνουμε για τα είδη της αλλομορφίας περιγράφουν τον τρόπο διαφοροποίησης των αλλομόρφων μεταξύ τους.

Κατ' αρχάς, πρέπει να αναφέρουμε πως δεν εμφανίζουν όλα τα μορφήματα αλλομορφοί. Εισάγουμε τον όρο **μηδενική αλλομορφία** (zero allomorphy) για τις περιπτώσεις που δεν πληρούν τις προϋποθέσεις του ορισμού και δεν εμφανίζουν αλλομορφοί σε καμία διαδικασία σχηματισμού λέξεων. Τέτοιου είδους μορφήματα είναι καταχωρημένα στο λεξικό χωρίς κανένα αλλομορφοί, μίας και το προκαθορισμένο μόρφημα συμμετέχει σε όλες τις διαδικασίες σχηματισμού λέξεων χωρίς κάποια αλλομορφική αλλαγή του θέματος, όπως π.χ. *άνθρωπ-ος, δάσ-ος, σοφ-ός, βουν-ό, τρέχ-ω*.

Στο κλιτικό σύστημα των αρσενικών και ουδετέρων ουσιαστικών παρατηρείται συχνά η προσθήκη ενός φωνήματος /δ/ ή /t/ στο μόρφημα πριν το κλιτικό επίθημα, το οποίο δεν είναι αποτέλεσμα συγχρονικού φωνολογικού κανόνα. Ακολουθώντας την ανάλυση της Ράλλη (2000, 2005, βλ. υποεν. 2.3), το φώνημα αυτό θεωρείται μέρος του ονοματικού θέματος (*παπαδ, βηματ*), το οποίο έχει δύο διαφορετικές μορφές ανάλογα

με το μορφολογικό περιβάλλον, στο οποίο βρίσκεται. Για τις παραπάνω περιπτώσεις προτείνουμε τον όρο **προσθετική αλλομορφία**, η οποία συνδέει το αρχικό μόρφημα με το αλλόμορφο με ένα πλεοναστικό κανόνα τύπου ΜΟΡΦΗΜΑ ~ ΜΟΡΦΗΜΑΧ, όπου προστίθεται ένα φώνημα στο τέλος του μορφήματος. Αντίστοιχα παραδείγματα απαντώνται στο ρηματικό σύστημα, όπου στη δεύτερη ρηματική τάξη (Ralli 1988, Ράλλη 2005) κατά τον σχηματισμό του αορίστου «προστίθεται» ένα φώνημα /e/ ή /i/ στο ρηματικό θέμα. Για παράδειγμα, τα πρώην συνηρημένα ρήματα, *αγαπώ* και *διαιρώ* που εμφανίζουν τα αλλόμορφα *αγαπ~ αγαπη* και *διαιρ~ διαιρε* αντίστοιχα.

Αντιστρόφως, εισάγουμε τον όρο της **αφαιρετικής αλλομορφίας** για τις περιπτώσεις στην κλίση των αρσενικών και θηλυκών ουσιαστικών, όπου παρατηρούμε την αποβολή του τελευταίου χαρακτήρα του θέματος στο κλιτικό παράδειγμα του πληθυντικού. Τα ουσιαστικά τύπου *καρδιά* (*καρδιά~ καρδι*) και *ψυχή* (*ψυχή~ ψυχη*) αναλύονται με τα φωνήματα /a/ και /i/, ως μέρος του θέματος (Ράλλη 2000, 2005), αφού έχει γίνει επανάλυση των κλιτικών επιθημάτων και θεωρούμε πως υπάρχει ένα μηδενικό κλιτικό επίθημα στον ενικό. Τα φωνήματα /i/ και /a/ αποτελούν μέρος του θέματος, γιατί εμφανίζονται σε όλες τις πτώσεις του ενικού και λόγω της συστηματικής εμφάνισης στο κλιτικό παράδειγμα εντάσσονται στο θέμα. Για τις περιπτώσεις προτείνουμε τον όρο αφαιρετική αλλομορφία που συσχετίζει αλλόμορφα με τον κανόνα ΜΟΡΦΗΜΑΧV~ ΜΟΡΦΗΜΑΧ. Στα ρήματα έχουμε περιπτώσεις αφαιρετικής αλλομορφίας, όπου οι συγχρονικά αδιαφανείς ενεστωτικοί δείκτες (*παίρ-ν-ω*, *κλέπ-τ-ω*) διαγράφονται σε κάποιους χρόνους δημιουργώντας αλλόμορφα, καθώς και σε ρηματικά παραγωγικά επιθήματα, όπως για παράδειγμα η διαγραφή του /n/ από το –*ών(ω)*, *τελει-ών-ω/ τελεί-ω-σ-α*.

Σε αντίθεση με τους δύο προαναφερθέντες αλλομορφικούς τύπους που είναι οι συχνότεροι σε εμφάνιση, υπάρχουν και κάποιοι τύποι που δεν είναι ιδιαίτερα συχνοί στα Ελληνικά. Εισάγουμε τον όρο **αντικαταστατική αλλομορφία**, όπου ο τελικός χαρακτήρας αντικαθίσταται από κάποιον άλλο. Για παράδειγμα, η τέταρτη κλιτική τάξη (Ράλλη 2000, 2005) συμπεριλαμβάνει θηλυκά ουσιαστικά που εμφανίζουν τρία αλλόμορφα (32.α). Το ένα αλλόμορφο περιγράφεται από τον τύπο της αφαιρετικής αλλομορφίας (*πολ–, ταξ–, βασ–*)· το άλλο αλλόμορφο που συμμετέχει στους κλιτικούς σχηματισμούς της γενικής ενικού και πληθυντικού αντικαθιστά το τελευταίο φώνημα με κάποιο άλλο. Θεωρούμε ότι αυτά τα αλλόμορφα εντάσσονται στο είδος της αντικαταστατικής αλλομορφίας με τον γενικό τύπο ΜΟΡΦΗΜΑΧ ~ ΜΟΡΦΗΜΑΖ. Αυτό



το είδος αλλομορφίας εμφανίζεται πιο συχνά στα ρήματα σε διάφορους συνδυασμούς (32.β).

- (32.) α. πολη ~ πολε ~ πολ  
ταξη ~ ταξε ~ ταξ  
βαση ~ βασε ~ βασ  
οψη ~ οψε ~ οψ
- β. ταραζ ~ ταραχ<sup>21</sup>  
αλλαζ ~ αλλαγ  
αρπαζ ~ αρπαγ  
τιναζ ~ τιναγ

Ο γενικός τύπος του μορφολεξικού κανόνα της αντικαταστατικής αλλομορφίας στην ονοματική κλίση εμφανίζει συγκεκριμένη κανονικότητα, όσον αφορά ποιο φώνημα αντικαθίσταται από ποιο. Αντίθετα στη ρηματική κλίση δεν είναι εφικτή η ύπαρξη ενός γενικού τύπου μορφολεξικού κανόνα, λόγω των διαφόρων ζευγών αντικατάστασης φωνημάτων στα ρηματικά αλλόμορφα. Εδώ οφείλουμε να τονίσουμε πως τα παραδείγματα του (32.α) εμφανίζουν δύο αλλόμορφα από δύο διαφορετικά είδη αλλομορφίας, αφού τα συγκεκριμένα θηλυκά εμφανίζουν **πολλαπλή αλλομορφία**.

Επιπροσθέτως, στις Ινδοευρωπαϊκές γλώσσες εμφανίζεται ο κανόνας της **μετάπτωσης** (apophony) ή **ετεροίωσης**<sup>22</sup> (ablaut) που στις περισσότερες γλώσσες δεν είναι πλέον ενεργός (βλ. Beekes 1995, Booij 1997, Szemerényi 1999, καθώς και Kuryłowicz & Mayrhofer 1969, van Coetsem 1993, Κίτσα 2001, Ράλλη 2005). Συγχρονικά οι αλλαγές που έχει επιφέρει η ανενεργή ετεροίωση θεωρούνται πλέον αλλομορφία. Για τις περιπτώσεις (33.) εισάγουμε το είδος της **ετεροιωτικής αλλομορφίας** με τον κανόνα να είναι ΜΟΡΦΗΜΑΧΥ ~ ΜΟΡΦΗΜΑΖΥ. Η ετεροιωτική αλλομορφία μπορεί να θεωρηθεί είδος της αντικαταστατικής αλλομορφίας, ωστόσο αφορά αλλομορφικές αλλαγές που συντελούνται στο εσωτερικό και όχι στο τέλος του μορφήματος. Είναι επίσης απαραίτητη η διαφοροποίηση, αφού δεν ταυτίζονται οι

<sup>21</sup> Περισσότερα για την ανάλυση της ρηματικής αλλομορφίας, βλ. ενότητα 2.4.3, όπου αιτιολογείται ποια είναι η μορφή του δεύτερου αλλομόρφου στα συγκεκριμένα αλλόμορφα.

<sup>22</sup> Ο κανόνας της ετεροίωσης δημιούργησε σε παλαιότερη φάση της γλώσσας διαφορετικής κατηγορίας θέματα, όπως π.χ. *βάλλω* > *-βόλ(ος)*, *φέρω* > *-φορά*, *κτείνω* > *-κτόν(ος)* (ενδεικτικά για τα δεσμευμένα θέματα, βλ. Ράλλη 2005, Petropoulou 2009, Πετροπούλου in progress).

γενικοί αλλομορφικοί κανόνες και στην υπολογιστική γλωσσολογία αντιμετωπίζονται ως διαφορετικές περιπτώσεις.

- (33.)    μεν ~ μειν  
           πλεν ~ πλυν  
           φευγ ~ φυγ

Η ετεροιωτική αλλομορφία εμφανίζεται αποκλειστικά σε έναν περιορισμένο αριθμό ρημάτων, αφού στα Ελληνικά σε αντίθεση με άλλες γλώσσες, όπως Ολλανδικά, Γερμανικά, Φινλανδικά, Βουλγαρικά, Τουρκικά, υπήρξε ενεργός ο κανόνας της ετεροίωσης στα ουσιαστικά *μόνο* στα ΑΕ (λιμήν- λιμένος, πατήρ-πατέρες, κανών- κανόνος). Ωστόσο, εμφανίζονται οι ετεροιωτικές αλλαγές σε περιπτώσεις αλλαγής κατηγορίας από ρήμα σε ουσιαστικό με την αλλαγή του εσωτερικού φωνήεντος του θέματος, όπως π.χ. *πλέκρ-ω* → *πλοκο-ή*, *λέγρ-ω* → *λόγο-ος*, *φέρρ-ω* → *φορο-ά*. Ουσιαστικά όμως τα παραπάνω αποτελούν διαφορετικά λήμματα στο λεξικό της ΚΝΕ.

Τέλος, υπάρχουν περιπτώσεις που έχουμε διπλή εφαρμογή αλλομορφικού κανόνα σε κάποιο λήμμα θέματος. Αυτές τις περιπτώσεις τις εντάσσουμε στο προτεινόμενο είδος της **συνδυαστικής αλλομορφίας**, το οποίο εμφανίζεται μόνο στα ρηματικά θέματα. Υπάρχουν περιπτώσεις όπου έχουμε την ταυτόχρονη εμφάνιση δύο διαφορετικών τύπων αλλομορφίας στο ίδιο αλλόμορφο, όπως π.χ. παίρνω: *παιρν* ~ *πηρ* ~ (*παρ*) / καλώ: *καλ* ~ *καλε* ~ *κλη* ~ *καλεσ*. Πιο συγκεκριμένα στο [*παρ*~/ *πηρ*~] έχουμε την ταυτόχρονη ύπαρξη αφαιρετικής αλλομορφίας (διαγραφή του /n/) και της ετεροιωτικής αλλομορφίας (τροπή του /e/ σε /a/ ή /i/), ενώ στο [*καλ*~/ *κλη*~] έχουμε την ύπαρξη της αφαιρετικής αλλομορφίας (διαγραφή του /a/) και της προσθετικής αλλομορφίας (προσθήκη του /i/). Φυσικά, η συνδυαστική αλλομορφία περιγράφεται από τους αντίστοιχους αλλομορφικούς κανόνες χωρίς να υπάρχει σειρά ιεράρχησης και εφαρμογής τους, η οποία ορίζεται αυθαίρετα. Το ίδιο ισχύει και για το είδος της πολλαπλής αλλομορφίας, όπου ένα θέμα εμφανίζει δύο αλλόμορφα που ανήκουν σε διαφορετικά είδη αλλομορφίας, όπως για παράδειγμα τα ουσιαστικά τύπου *πολη*~ *πολε*~ *πολ* και *μπακαλη*~ *μπακαληδ*~ *μπακαλ*. Στην πολλαπλή αλλομορφία το κάθε αλλόμορφο συνδέεται με τον αντίστοιχο αλλομορφικό κανόνα με το αρχικό μόρφημα.

#### 2.4.1.2. Ψευδοαλλόμορφα (vs. ελεύθερες παραλλαγές) και υποκατάσταση

Όπως αναφέρθηκε προηγουμένως, τα περισσότερα ρήματα περιλαμβάνουν αλλόμορφα, τα οποία αποτελούν εγγενείς απρόβλεπτες αλλαγές. Για το λόγο αυτό καταχωρούνται κάτω από το ίδιο λήμμα στο λεξικό συνδεδεμένα με πλεοναστικούς κανόνες τύπου *αγαπ/ αγαπα ~ αγαπη, διαιρ ~ διαιρε*. Σύμφωνα με τον ορισμό τα αλλόμορφα του ίδιου μορφήματος βρίσκονται σε συμπληρωματική κατανομή. Υπάρχουν, όμως, περιπτώσεις με εναλλαγές τύπων στα ίδια μορφολογικά περιβάλλοντα. Πιο συγκεκριμένα, ρήματα δεύτερης συζυγίας της ΚΝΕ που στα αρχαία Ελληνικά ανήκαν στα συνηρημένα ρήματα σε *-άω (-ᾶ)* έχουν τους τύπους *αγαπ(ώ)* και *αγαπά(ω)* στον Ενεστώτα και τους τύπους *αγάπα-γ(α)* και *αγαπ-ού(σα)* στον Παρατατικό, όπου σύμφωνα με τη Ράλλη (2005)<sup>23</sup> τα δύο αλλόμορφα συμπεριφέρονται ως **ελεύθερες παραλλαγές** (free variants) μοιραζόμενα το ίδιο μορφολογικό περιβάλλον.

- (34.) αγαπ / αγαπα ~ αγαπη  
 κεντ / κεντα ~ κεντη  
 μετρ / μετρα ~ μετρη  
 μιλ / μιλα ~ μιλη

Συμφωνούμε με την άποψη της Ράλλη (2005: 69) ότι κάποιοι ανενεργοί φωνολογικοί κανόνες μπορούν να ενεργοποιηθούν ξανά. Πρέπει να είμαστε εξαιρετικά επιφυλακτικοί για την υιοθέτηση της άποψης για την επανενεργοποίηση κανόνων, αφού έτσι εμπλέκεται η διαχρονία με τη συγχρονία. Στις συγκεκριμένες περιπτώσεις (34.) φαίνεται πως η διαγραφή του /a/ συμβαίνει συστηματικά και προβλέψιμα στο σύνολο των ρημάτων που προέρχονται ή εντάσσονται στη συζυγία των πρώην συνηρημένων. Υποστηρίζουμε πως τα συγκεκριμένα μορφήματα πρέπει να αναφέρονται ως **ψευδοαλλόμορφα**. Εισάγουμε τον όρο ψευδοαλλόμορφα, διότι καταπατούν τον όρο της συμπληρωματικής κατανομής του ορισμού της αλλομορφίας και επομένως δεν πληρούν τις προϋποθέσεις για να χαρακτηριστούν αλλόμορφα. Επιπροσθέτως, τα ψευδοαλλόμορφα μοιάζουν πολύ με τα φωνόμορφα, μόνο που τα

<sup>23</sup> Η Ράλλη (2005: 69) σημειώνει «*πως το αλλόμορφο χωρίς /a/ (αγαπ-) παράγεται από το αλλόμορφο /Χα/ (αγαπα-), το οποίο χάνει το ληκτικό φωνήεν μπροστά από το φωνήεν της κατάληξης με την εφαρμογή ενός φωνολογικού κανόνα απαλοιφής. Αυτός ο φωνολογικός νόμος θυμίζει τα αποτελέσματα του ανενεργού σήμερα αρχαίου νόμου της συναίρεσης. Αν και θεωρητικά δεν αποκλείεται η επανενεργοποίηση ανενεργών νόμων, πρέπει να είμαστε ιδιαίτερα επιφυλακτικοί στην υιοθέτησή τους δεδομένου ότι μοιάζει να μεταφέρεται η διαχρονία στη συγχρονία*».

πρώτα μορφήματα είναι πιθανότερο να αποτελούν αποτέλεσμα κάποιου φωνολογικού κανόνα παρά μία αυθαίρετη και απρόβλεπτη διαφοροποίηση. Τα επιχειρήματα για την υποστήριξη είτε της μίας ή είτε της άλλης συνθήκης δεν είναι ισχυρά. Θεωρούμε ότι οι συγκεκριμένες μορφές βρίσκονται σε κατάσταση αλλαγής και τείνουν προς την κατεύθυνση να αποτελέσουν προϊόντα ενός λεξικού φωνολογικού κανόνα.

- (35.) α. πραγματ-ευ-τή(ς)      β. πραγματ-ευ-τ(ές) / πραγματ-ευ-τάδ(ες)  
 τραγουδ-ιστή(ς)      τραγουδ-ιστ(ές) / τραγουδ-ιστάδ(ες)  
 ψάλ-τη(ς)      ψάλ-τ(ες)      / ψαλ-τάδ(ες)  
 [δεσπότη(ς)]      δεσπότη(ες)      / δεσποτάδ(ες)]

Τα μορφήματα του (35.β) συμπεριφέρονται ως **ελεύθερες παραλλαγές**, αφού μοιράζονται το ίδιο μορφολογικό περιβάλλον. Τυπικά, οι ελεύθερες παραλλαγές καταχωρούνται στο λήμμα του λεξικού με το σύμβολο «/». Πρέπει να σημειωθεί ότι οι ελεύθερες παραλλαγές είναι ιδιαίτερα σπάνιες και κατά τη Ράλλη (2005: 69-70) μπορούν συνήθως να ερμηνευθούν με βάση τις διάφορες διαλεκτικές ποικιλίες. Τα παραδείγματα του (35.β) περιλαμβάνουν μία μικρή ομάδα παράγωγων ουσιαστικών σε *-τη(ς)*<sup>24</sup> και *-ιστη(ς)* που εναρμονίζουν το κλιτικό τους παράδειγμα σύμφωνα με τη δεύτερη κλιτική τάξη και δημιουργούν δύο τύπους στον πληθυντικό αριθμό. Ο ένας τύπος της αφαιρετικής αλλομορφίας (*-τ-*, *-ιστ-*) είναι το καθιερωμένο αλλόμορφο για τα συγκεκριμένα επιθήματα, ενώ ο δεύτερος τύπος είναι με την προσθήκη του συμπλέγματος /αδ/ σε σύγκριση με τα ήδη υπάρχοντα αλλόμορφα ή με άλλα λόγια είναι τα αλλόμορφα *-ταδ-* και *-ισταδ-*. Είναι αξιοσημείωτο ότι το εναλλακτικό αλλόμορφο έχει όλα τα χαρακτηριστικά του παραγωγικού επιθήματος *-α(ς)~ -αδ-*.

Θεωρούμε ότι αποτελούν παράγωγες λέξεις (συνδυασμός δύο παραγωγικών επιθημάτων), από τις οποίες απουσιάζει το κλιτικό παράδειγμα ενικού (*\*ψαλτάς*, *\*τραγουδιστάς*), ενώ το αντίστοιχο του πληθυντικού εντάσσεται ως ελεύθερη παράλλαξη στο αρχικό παράγωγο θέμα (*ψαλ-τ-*, *τραγουδ-ιστ-*). Η ελεύθερη παραλλαγή σε /αδ/ και /ισταδ/ (π.χ. *ψαλτάδες*, *τραγουδιστάδες*) ενέχει δείγματα αρνητικής σημασίας και αυτή είναι η μόνη διαφοροποίηση από τον παράλληλο τύπο (π.χ. *ψάλτες*, *τραγουδιστές*). Χαρακτηρίζουμε τα συγκεκριμένα μορφήματα ως ελεύθερες παραλλαγές, αφού δεν υφίστανται καμία υπόνοια εφαρμογής κάποιου κανόνα και η

<sup>24</sup> Θεωρούμε ότι η λέξη *δεσπότης* είναι μορφολογικά αδιαφανής και το παραγωγικό επίθημα *-τη(ς)* αποτελεί μέρος του θέματος.

προσθήκη του παραγωγικού αλλόμορφου  $-αδ-$  έχει επαναναλυθεί ως μέρος του προηγούμενου επιθήματος  $-τ-$  και  $-ιστ-$ . Όπως και στα ψευδοαλλόμορφα, καταστρατηγείται η προϋπόθεση της εμφάνισης σε διαφορετικό περιβάλλον, αλλά σε αυτά τα μορφήματα η αλλαγή είναι πέραν καμίας αμφιβολίας αυθαίρετη και απρόβλεπτη.

Στην προηγούμενη υποενότητα μελετήσαμε την εμφάνιση διαφορετικών πραγματώσεων του ίδιου μορφήματος, οι οποίες είναι απρόβλεπτες και δεν υπακούουν σε κάποιον φωνολογικό κανόνα. Στη διεθνή βιβλιογραφία σημειώνεται από διάφορους συγγραφείς (Mel'cuk 1976, 1992, 1994, 2006, Anderson 1985, Carstairs-McCarthy 1992, 1994, Booij 1997, Meyer 1997, Veselinova 2000, 2003, 2007, Maiden 2001, 2004· για μια τυπολογική βάση και συλλεκτική βιβλιογραφία πρβλ. Corbett, Brown, Chumakina & Hippisley 2005) ότι η **υποκατάσταση** (suppletion) μπορεί να αποτελέσει μία ακραία μορφή αλλομορφίας. Η υποκατάσταση είναι φαινόμενο κατά το οποίο τα κλιτικά κενά ενός παραδείγματος αντικαθίστανται από τύπους άλλου λεξήματος, ένα φαινόμενο αρκετά συχνό στο ρηματικό κλιτικό σύστημα των Αρχαίων Ελληνικών. Συνήθως οι διάφοροι ερευνητές συμφωνούν σε κάποια σημεία ορισμού της υποκατάστασης. Αποδέχονται ότι η **ασθενής υποκατάσταση** συνδέει δύο μορφήματα που μοιράζονται κάποια φωνολογικά χαρακτηριστικά (π.χ. *buy* και *bought* 'αγοράζω και αγόρασα'), ενώ η **ισχυρή υποκατάσταση** συνδέει δύο τελείως φωνολογικά διαφορετικές μορφές (π.χ. *go* and *went* 'πηγαίνω και πήγα'). Ειδικότερα, σύμφωνα με τον Mel'cuk (1976, 1992) σημαντική παράμετρος είναι αυτή της μοναδικότητας, όπου η υποκατάσταση ενός μορφήματος δεν πρέπει να ομοιάζει φωνολογικά ως προς την μορφή του μορφήματος που καλείται να καλύψει το κενό στο κλιτικό παράδειγμα.

(36.)	α.	θέω	ἔδραμον (αορ.)	δεδράμηκα (πρκ.)	ΑΕ
		βλώσκω	μολοῦμαι (μελ.)	ἔμολον (αορ.)	ΑΕ
		πωλῶ	πέπρακα (πρκ.)	ἐπεπράκειν (υπερσ.)	ΑΕ
		εἶμι	ἦλθον (αορ.)	ἐλήλυθα (πρκ.)	ΑΕ
		ἐσθίω	ἔφαγον (αορ.)	ἐδήδοκα (πρκ.)	ΑΕ
	β.	εἶμαι	θα υπάρχω (μελ.)	υπήρξα (αορ.)	KNE
		τρώω	θα φάω (μελ.)	έφαγα (αορ.)	KNE

Όπως υποστηρίζει και η Ράλλη (2005), δεν κατατάσσουμε την υποκατάσταση στην κατηγορία της αλλομορφίας, αφού οι τύποι υποκαθιστούν άλλους γραμματικά τύπους,

δεν προέρχονται από το ίδιο μόρφημα και η διαδικασία δεν είναι αυθαίρετη (είναι απαραίτητη η συμπλήρωση των κενών από τύπους άλλου λήμματος), αλλά μοναδική. Αντιθέτως «οι τύποι της αλλομορφίας είναι συγγενείς, ανήκουν στο ίδιο μόρφημα και αυτό επιβεβαιώνεται τόσο από την φωνολογική τους μορφή, όσο και από τη σημασία.»<sup>25</sup> (Ράλλη 2005: 73).

Συνοψίζοντας την περιγραφή και τον καθορισμό της αλλομορφίας, παραθέτουμε έναν πίνακα στον οποίο παρουσιάζουμε το μόρφημα με τις διάφορες μορφές που μπορεί να έχει στην καταχώρησή του ως λήμμα στο Λεξικό. Η σειρά με την οποία παρατίθενται οι όροι δεν είναι τυχαία. Στο κάτω μέρος του πίνακα είναι οι μορφηματικές αλλαγές που ερμηνεύονται φωνολογικά (φωνόμορφα) και στην κορυφή του είναι οι μορφές ενός μορφήματος που τις έχει δανειστεί από κάποιο άλλο (υποκατάστατα) και είναι πλησίον του να χαρακτηριστεί ως ανεξάρτητο μόρφημα. Ανάμεσά τους τοποθετούνται οι υπόλοιπες διαφοροποιήσεις με την κατηγορία του αλλομόρφου να είναι σχεδόν στο υψηλότερο σημείο του πίνακα, αφού ο ορισμός του περιγράφει αλλαγές μορφήματος που προέκυψαν αυθαίρετα και απρόβλεπτα πάντα σε συμπληρωματική κατανομή.

Διαφοροποίηση αλλαγών σε μορφήματα
<b>Μόρφημα</b>
Υποκατάστατο
Αλλόμορφο
Ελεύθερη Παραλλαγή
Ψευδοαλλόμορφο
Φωνόμορφο

Πίνακας 2: Το μορφολογικό συνεχές των διαφόρων αλλαγών σε ένα μόρφημα.

<sup>25</sup> Οφείλουμε να επιστημόνουμε πως οι τύποι που δανείζονται ένα μόρφημα κατά την υποκατάσταση είναι από κάποιο λήμμα με συνώνυμη ή σχεδόν ταυτόσημη σημασία. Επομένως το τελευταίο κριτήριο δεν είναι ιδιαίτερο ισχυρό, γιατί ισχύει και στη σχέση μορφήματος – υποκατάστατου.

#### 2.4.2. Η θέση της Αλλομορφίας στο Λεξικό

Σύμφωνα με την πρόταση της Ράλλη (2005), για τα Ελληνικά το Νοητικό Λεξικό (ΝΛ - Mental Lexicon)<sup>26</sup> είναι ο χώρος, στον οποίο καταχωρούνται γενικά τα μορφολογικά συστατικά που δεν κατατέμνονται σε επιμέρους στοιχεία, δηλαδή οι ελάχιστες μονάδες που αποτελούν συνδυασμό μορφής και σημασίας<sup>27</sup>. Πιο συγκεκριμένα, στο ΝΛ καταχωρούνται τα μορφήματα με όλες τις φωνολογικές, μορφολογικές, συντακτικές και σημασιολογικές πληροφορίες που τα χαρακτηρίζουν, καθώς και τα αλλομορφά τους. Όπως διαφαίνεται από τα παραπάνω το συγκεκριμένο ΝΛ διαφέρει από τα κοινά λεξικά, τα οποία περιέχουν κυρίως λεξήματα και πληροφορίες σημασιολογικής ή/ και πραγματολογικής χροιάς.

Βάσει της θεωρίας που πρώτη εισήγαγε η Lieber (1980, 1982), το ΝΛ έπαψε να προσεγγίζεται ως ένας κατάλογος λεξημάτων. Έχει ενεργό ρόλο στη γλωσσολογική θεωρία και περιλαμβάνει τον γραμματικό τομέα σχηματισμού λέξεων (μορφολογία), ο οποίος αποτελεί αυτόνομο τομέα της γραμματικής<sup>28</sup>. Την ίδια προσέγγιση ακολουθούν και οι Selkirk (1982), Kiparsky (1982), Ράλλη (1984), Mohanan (1986) και Ralli (1988). Σε αυτή η φάση, το ΝΛ χωρίζεται σε δύο επίπεδα: (α.) το σταθερό λεξικό που περιλαμβάνει τα μορφήματα και (β.) το δυναμικό λεξικό (dynamical lexicon), όπου σχηματίζονται οι λέξεις.

Στο ΝΛ καταχωρούνται λήμματα με μορφηματικό χαρακτήρα (*ανθρωπ-, μιλ-, ομορφ-*) και μονομορφηματικά (*έξω, εγώ, κατά*), τα οποία φέρουν συγκεκριμένα χαρακτηριστικά και έχουν λειτουργία. Υπάρχουν δύο αντικρουόμενες απόψεις όσον αφορά τα κλιτικά μορφήματα και το ΝΛ. Για τους υποστηρικτές της **υπόθεσης του αδύναμου λεξικαλισμού** (weak lexicalist hypothesis), όπως ο Spencer (1991), τα κλιτικά επιθήματα δεν καταχωρούνται στο λεξικό. Ωστόσο, στην παρούσα διατριβή υιοθετούμε την **υπόθεση του ισχυρού λεξικαλισμού** (strong lexicalist hypothesis) και

---

<sup>26</sup> Πρέπει να αναφερθεί ότι στη βιβλιογραφία δεν υπάρχει συμφωνία ως προς την ορολογία. Ο όρος Νοητικό Λεξικό απαντά και ως **αφηρημένο λεξικό** (abstract lexicon) ή **σταθερό λεξικό** (permanent lexicon), (βλ Lieber 1980 για τη σχετική συζήτηση).

<sup>27</sup> Δεν υπάρχει σύγκλιση απόψεων μεταξύ ψυχολinguιστών, νευρογλωσσολόγων και θεωρητικών μορφολόγων για το περιεχόμενο του νοητικού λεξικού και το είδος των μορφολογικών μονάδων (για μία επισκόπηση των συγκεκριμένων απόψεων βλ. Lieber 1980, Ralli 1988, Pinker 1999, Διακογιώργη 2000α, 2000β Booij 2002, Ράλλη 2005: 74-84).

<sup>28</sup> Περισσότερα σχετικά με την αυτονομία της μορφολογίας και την θέση της σε σχέση με τα άλλα επίπεδα της γλώσσας, βλ. Carstairs (1987), Spencer (1991), Booij (1997, 2003, 2005), Español-Echevarría & Ralli (2000), Ράλλη (2005) μεταξύ άλλων.

εντάσσουμε στο σταθερό λεξικό όλα τα προσφύματα<sup>29</sup>, τόσο τα παραγωγικά προθήματα και επιθήματα, όσο και τα κλιτικά επιθήματα. Αν και ο τρόπος αναπαράστασης των προσφυσμάτων έχει διαφορές από τα λήμματα θεμάτων (ρίζων) και μονομορφηματικών λέξεων, εντούτοις θεωρούμε ότι καταχωρούνται σε ένα χώρο του σταθερού λεξικού που λειτουργεί παράλληλα και συμπληρωματικά με τα υπόλοιπα λήμματα. Έτσι, όταν η μορφολογία κατά τη διάρκεια σχηματισμού λέξεων καταφεύγει στο λεξικό, παίρνει όλες τις απαραίτητες πληροφορίες.

Τα προσφύματα αποτελούν περιορισμένη κατηγορία συγκριτικά με τα θέματα, είναι δεσμευμένα, δεν μπορούν να συνδυαστούν μόνο μεταξύ τους για το σχηματισμό λέξεων, μπορεί να έχουν μηδενική μορφή. Παράλληλα υπάρχει ένα **πλαίσιο υποκατηγοριοποίησης** (subcategorization frame) που ορίζει το είδος των θεμάτων ή και των προσφυσμάτων που επιλέγονται για το σχηματισμό λέξεων (Ράλλη 2005: 76-77). Το πλαίσιο υποκατηγοριοποίησης θα τροποποιηθεί στη συνέχεια και για την καταχώρηση των αλλομόρφων.

Είναι σαφές ότι τα καταχωρημένα λήμματα στο σταθερό λεξικό είναι απλά μορφήματα (θέματα, ρίζες, προσφύματα) και δεν εμπεριέχουν μορφές που να αποτελούν αποτέλεσμα εφαρμογής οιοδήποτε κανόνα. Επομένως, δεν υφίσταται λόγος παρουσίας στο σταθερό λεξικό όλων εκείνων των λεξικών στοιχείων που είναι αποτελέσματα εφαρμογής κανόνων σχηματισμού λέξεων. Είναι σημαντικό να επισημάνουμε τα **κριτήρια επιλογής** (selectional criteria, βλ. Ράλλη 2005: 78) σύμφωνα με τα οποία τα διάφορα προσφύματα επιλέγουν και αποκλείουν κάποια θέματα ή προσφυσματοποιημένα θέματα. Για παράδειγμα, το ρηματικό παραγωγικό επίθημα  $-αρ(\omega)$  επιλέγει θέματα που φέρουν το χαρακτηριστικό [+ ξενικό], όπως σουτ+αρ(ω) → σουτάρω, σερφ+αρ(ω) → σερφάρω.

Στον ακόλουθο πίνακα δίνεται μία πρώτη διάρθρωση λημμάτων του λεξικού, θεμάτων και επιθημάτων, για την καλύτερη κατανόηση και αναπαράσταση των παραπάνω θέσεων. Οι κύριες πληροφορίες που παρουσιάζουμε είναι κυρίως αυτές της μορφολογίας και της φωνολογίας:

Λήμματα θεμάτων	«δέρνω»	«κύμα»
-----------------	---------	--------

<sup>29</sup> Υπάρχουν υποστηρικτές της λεξικής μορφολογίας, όπως οι Aronoff (1976) και Kiparsky (1982) που δεν συμπεριλαμβάνουν τα προσφύματα στο λεξικό, αλλά τα ενσωματώνουν με κανόνες κατά την επιτέλεση των διαδικασιών σχηματισμού λέξεων.



<b>Κατηγορία</b>	Ρήμα	Ουσιαστικό
<b>Μορφολογικές πληροφορίες</b>	αλλόμορφα: δερν~ δειρ~ δαρ	αλλόμορφα: κυμα~ κυματ
<b>Φωνολογικές πληροφορίες</b>	/ðern/, /ðir/, /ðar/, τονική συμπεριφορά	/kima/, /kimat/, τονική συμπεριφορά
<b>Συντακτικές πληροφορίες</b>	Μεταβατικό ρήμα: Y_ A	Y, A, K, ΠΠ
<b>Σημασιολογικές πληροφορίες</b>	Σύνολο χαρακτηριστικών	Σύνολο χαρακτηριστικών

<b>Λήμματα επιθημάτων</b>	<b>«-τζη(ς)»</b>	<b>«-ως»</b>
<b>Τύπος</b>	Παραγωγικό επίθημα	Παραγωγικό επίθημα
<b>Κατηγορία</b>	Ουσιαστικό	Επίρρημα
<b>Μορφολογικές πληροφορίες</b>	αλλόμορφα: -τζη- ~ -τζηδ-	χωρίς αλλόμορφο
<b>Φωνολογικές πληροφορίες</b>	/dzi/, /dzið/, συνοδευτικές τονικές πληροφορίες	/os/, τονικές πληροφορίες
<b>Πλαίσιο υποκατηγοριοποίησης</b>	]o –	]E –
<b>Συντακτικές πληροφορίες</b>	Y, A, K, ΠΠ	ΕπιρΠ
<b>Σημασιολογικές πληροφορίες</b>	Σύνολο χαρακτηριστικών	Σύνολο χαρακτηριστικών
<b>Διακριτικά χαρακτηριστικά</b>		Θέμα <sub>[+λόγιο]</sub> ]E –

Πίνακας 3: Δείγμα λημμάτων από το Σταθερό Λεξικό (Ράλλη 2005).

Αξίζει να τονισθεί ότι καταγράφεται διαφωνία μεταξύ των μορφολόγων (Lieber 1980, 1982, Maiden 1992, Booij 1997, Ράλλη 2005) και των φωνολόγων (Drachman 2005, Mascarò 1996a, 1996b, 2007) σχετικά με τον τρόπο καταχώρησης των αλλομόρφων, καθώς και για τον τύπο αναπαράστασης. Ο Booij (1997) θεωρεί πως υπάρχει ένα **προκαθορισμένο ή βασικό αλλόμορφο** (default allomorph) που συναγωνίζεται τα υπόλοιπα **μη-προκαθορισμένα ή μη-βασικά αλλόμορφα** (non-default allomorphs). Σημειώνει πως αυτά πρέπει να καταχωρούνται σαν μία απλή λίστα στοιχείων και όχι με τη μορφή κανόνων εξαγωγής τους από το βασικό αλλόμορφο. Να υπογραμμίσουμε ότι οι αλλομορφικές αλλαγές δεν είναι πλέον παραγωγικές και δεν μπορούν να ερμηνευθούν σε καμία περίπτωση με φωνολογικούς και μορφολογικούς κανόνες. Κατά τις Lieber (1980, 1982) και Ράλλη (2005) τα αλλόμορφα πρέπει να

τοποθετηθούν σε λίστα και πρέπει να σημειωθούν οι πληροφορίες που να προσδιορίζουν το περιβάλλον εμφάνισης.

Θεωρούμε πως κανένα αλλόμορφο δεν παράγεται από κάποιο άλλο, ενώ συμβατικά ορίζουμε ως βασικό αλλόμορφο αυτό που χρησιμοποιείται στη μορφή του λήμματος, καθώς δεν υπάρχουν κριτήρια για την επιλογή του, όπως συχνότητα εμφάνισης, εύρος μορφολογικού περιβάλλοντος (σε ποιες διαδικασίες συμμετέχει) κ.α. Το βασικό αλλόμορφο τοποθετείται πάντα στο αριστερό μέρος του μορφολεξικού κανόνα (ΒΑΣΙΚΟ/ΠΡΟΚΑΘΟΡΙΣΜΕΝΟ~ ΜΗ-ΠΡΟΚΑΘΟΡΙΣΜΕΝΟ1{~ ΜΗ-ΠΡΟΚΑΘΟΡΙΣΜΕΝΟ2}). Η κατηγοριοποίηση των αλλομόρφων είναι θέμα αναπαράστασης και αναφερόμαστε σε μία τυπική κανονικότητα που δεν κυβερνάται από τη φωνολογία, ούτε βρίσκεται σε κάποια συστηματική σχέση μορφής – σημασίας. Δεν υπάρχουν κριτήρια επιλογής του βασικού. Η συνδετική σχέση ανάμεσα σε δυο αλλόμορφα μπορεί να εκφραστεί μόνο με την μορφή πλεοναστικών κανόνων και δεν υφίσταται ιεραρχία ανάμεσα σε αυτά.

(37.) (α.) παρν~

(β1.) πηρ]V σ]Past]Aor > \*πηρ-σ-αμε

(β2.) πηρ]V,+Past Ø]Aor > πηρ-αμε

Είναι απαραίτητο, λοιπόν, να εισάγουμε πλαίσιο αλλομορφικής πληροφορίας στο λεξικό σε όσα λήμματα εμφανίζουν αλλόμορφα. Εάν τα αλλόμορφα φέρουν μόνο το δείκτη της λεξικής κατηγορίας (37.β1), τότε υπάρχει ο κίνδυνος των αντιγραμματικών λέξεων και επομένως επιβάλλεται η χρήση συνοδευτικών χαρακτηριστικών. Εισάγουμε την **πληροφορία περιβάλλοντος** (context-sensitive information) με τη συνδρομή συμπληρωματικών διακριτικών σε κάποιες περιπτώσεις, για λήμματα μορφημάτων με αλλόμορφα, ώστε να είναι ορθώς προσδιορισμένα. Για τα αλλόμορφα ειδικών περιστάσεων επισημαίνει χαρακτηριστικά ο Booij (1997: 43):

*«Ο μόνος τρόπος για να αντιμετωπίσουμε τους αλλομορφικούς τύπους [...] είναι να τοποθετήσουμε όλα τα θέματα σε λίστα μέσα στο λεξικό και να επισυνάψουμε ένα στοιχείο υποκατηγοριοποίησης στο καθένα από αυτά. Για παράδειγμα, το αλλόμορφο 'schand' από το 'schande' πρέπει να υποκατηγοριοποιηθεί ως στοιχείο που εμφανίζεται στη σύνθεση σε θέσεις μη-κεφαλής (schandknaap, schandvlek, schanddaad), όπου το 'schande' είναι η βασική μορφή, το οποίο εμφανίζεται ως ανεξάρτητο στοιχείο, σε θέσεις κεφαλής στη σύνθεση και στην παραγωγή (schandelijk)».*

Παρομοίως, οι Pirrelli & Battista (2000a) υποστηρίζουν πως η διαφοροποίηση των μορφών ενός θέματος με απρόβλεπτο τρόπο μας αναγκάζει να καταχωρήσουμε στο λεξικό όλες τις μορφές του. Η κωδικοποίηση τους είναι απαραίτητη για την οικονομική χρήση του λεξικού. Όντας πολύπλοκο το ρηματικό σύστημα των Ιταλικών, προτείνουν ένα ιδιαίτερο σύστημα κατηγοριοποίησης, το οποίο αποτελείται από **ενδείκτες** (indexes) που καλύπτουν διάφορους υποδοχείς των παραδειγματικού συστήματος. Οι υποδοχείς αναφέρονται στα περιβάλλοντα της κλίσης που εμφανίζονται διαφορετικά αλλόμορφα, για παράδειγμα Αόριστος, Παρακείμενος, Μετοχή, Γερούνδιο ή πρόσωπα ενός χρόνου. Τα αλλόμορφα έχουν ενδείκτες που αναφέρονται σε ποιο περιβάλλον θα εμφανιστούν και είναι αναμενόμενο ένας ενδείκτης να συμπληρώνει παραπάνω από έναν υποδοχέα σε πολλά ρήματα, π.χ. να εμφανίζεται το ίδιο αλλόμορφο και στα δύο πρώτα πρόσωπα του πληθυντικού ενεστώτα (βλ. Πίνακας 4) ή σε περιβάλλοντα που άλλα ρήματα θα είχαν δύο διαφορετικά αλλόμορφα.

ΕΝΙΚΟΣ			ΠΛΗΘΥΝΤΙΚΟΣ		
1	2	3	1	2	3
/fi'nisco/	/fi'niffi/	/fi'niffe/	/fi'njamo/	/fi'nite/	/fi'niskonono/
finisco	finisci	finisce	finiamo	finite	finiscono
S <sub>2</sub>	S <sub>3</sub>	S <sub>3</sub>	S	S	S <sub>2</sub>

Πίνακας 4: Το παράδειγμα *FINIRE* από τους Pirrelli & Battista (2000a: 15), όπου το S<sub>x</sub> είναι ο ενδείκτης του κλιτικού παραδείγματος.

Είναι συνηθισμένο στην πλειονότητα των ρομανικών και γερμανικών γλωσσών να χρειάζονται τα αλλόμορφα συγκεκριμένα χαρακτηριστικά περιβάλλοντος, καθώς η αλλομορφία τους δεν εμφανίζει κανονικότητα ως προς το περιβάλλον συμμετοχής. Για παράδειγμα, στα Γερμανικά (για περισσότερα βλ. Lieber 1982) τα αλλόμορφα της κλίσης μετέχουν εξίσου στη σύνθεση, δηλαδή όλα τα κλιτικά αλλόμορφα μπορούν να συμμετέχουν στη σύνθεση ως πρώτο συνθετικό. Ο Anderson (2008) εντοπίζει το παραπάνω πρόβλημα και θεωρεί ότι το σύνολο μορφολογικής υποκατηγοριοποίησης κάνει την περιγραφή αλλομόρφων πιο πολύπλοκη από το αντίστοιχο σύνολο της φωνολογικής υποκατηγοριοποίησης για τα Surmiran. Κοινώς, τα αλλόμορφα πρέπει να καταχωρούνται στο Λεξικό ως μία σκέτη λίστα μορφημάτων με τη συνοδεία συνοδευτικών πληροφοριών.

Οφείλουμε να σημειώσουμε πως σε αντίθεση με τα Ελληνικά, σε αρκετές γλώσσες είναι σύνηθες να αλλάζει το αλλόμορφο ακόμα και τέσσερις φορές σε ένα κλιτικό παράδειγμα ρήματος (βλ. Ιταλ. Πίνακας 4). Αυτό καθιστά δύσκολο να προσδιοριστεί η πληροφορία περιβάλλοντος. Εντούτοις, θεωρούμε πως είναι απαραίτητη η

καταχώρηση της πληροφορίας περιβάλλοντος με τρόπο αρκετά λειτουργικό. Η αρχική μας προσέγγιση είναι να τοποθετήσουμε όλες τις πληροφορίες περιβάλλοντος εντός του ΝΛ σε κάθε λήμμα. Στην επόμενη ενότητα θα εισαγάγουμε μία εναλλακτική προσέγγιση της πληροφορίας περιβάλλοντος που θα είναι και η τελική μας πρόταση.

Στα Ελληνικά η καταχώρηση πληροφοριών για τα αλλόμορφα είναι δυνατή χωρίς τη χρήση πλήθους ιδιοτήτων. Για παράδειγμα, αν καταχωρήσουμε ένα αλλόμορφο με την πληροφορία του αορίστου δεν χρειάζεται να προσδιορίσουμε τίποτα περισσότερο για την κλίση. Το ίδιο πρέπει να υποστηριχθεί και για τις περισσότερες κλιτικές τάξεις των ονομάτων, όπου τα αλλόμορφα μπορούν να χωριστούν σε αλλόμορφα ενικού (*καρδιά*) και πληθυντικού αντίστοιχα (*καρδι-*). Το τελευταίο συμμετέχει αποκλειστικά και στην παραγωγή.

Πιο αναλυτικά, σύμφωνα με την πρότασή μας τα θέματα των ουσιαστικών που εμφανίζουν αλλόμορφα καταχωρούνται έχοντας ως *βασικό* αλλόμορφο εκείνο που χρησιμοποιείται για το σύνολο ή την πλειονότητα του κλιτικού παραδείγματος του ενικού και ως μη βασικό αυτό που χρησιμοποιείται κυρίως για τον πληθυντικό. Το μη βασικό χρησιμοποιείται και στην παραγωγή και στη σύνθεση σε αντίθεση με την κατηγοριοποίηση του Booij (1997), ο οποίος καθορίζει το αλλόμορφο με την εκτενέστερη εμφάνιση. Το βασικό αλλόμορφο δίνεται πάντα πρώτο σε κάποιο λήμμα στην καταχώρηση του Λεξικού, π.χ. *παιρν~ πηρ~ παρ, κυμα~ κυματ* (ον. εν. για ουσιαστικά [ΚΑΡΔΙΑ, ΒΗΜΑ] και επίθετα [ΟΜΟΡΦΟΣ, ΕΥΓΕΝΗΣ], 1εν. ορ. ενεστ. για τα ρήματα [ΑΓΑΠΩ]). Υπάρχουν ωστόσο κάποιες περιπτώσεις, όπως τα ουσιαστικά που κλίνονται όπως η *πόλη* και το *κύμα* που χρειάζονται επιπρόσθετη μορφολογική κατηγοριοποίηση.<sup>30</sup> Ο καθορισμός του βασικού αλλομόρφου στη θεωρία αποτελεί βάση για τους υπολογιστικούς κανόνες αλλομορφικής μετατροπής, αφού όπως θα δούμε στην ενότητα 6.2.1 οι μορφολεξικοί κανόνες σε συνδυασμό με τους κανόνες χρήσης αλλομορφίας (βλ. 2.4.2.1) κωδικοποιούνται σε παραγωγικούς υπολογιστικούς κανόνες.

Λήμματα ουσιαστικών	«καρδιά»	«όμορφος»
<b>Κατηγορία</b>	Ουσιαστικό ΚΤ3	Επίθετο ΚΤ1
<b>Μορφολογικές πληροφορίες</b>	αλλόμορφα: καρδια~ καρδι	αλλόμορφα: ομορφ-

<sup>30</sup> Να σημειωθεί πως η κωδικοποίηση της πληροφορίας περιβάλλοντος και τα διακριτικά χαρακτηριστικά είναι πιο ανεπτυγμένα στο επόμενο κεφάλαιο, όπου γίνεται αναλυτικότερη κατανομή των μορφολογικών περιβαλλόντων που εμφανίζεται ένα αλλόμορφο με τη χρήση νέων κανόνων.

Πληροφορίες περιβάλλοντος (context-sensitive information)	καρδια]EN	ομορφ]ΚΛ, ΠΑΡ, ΣΥΝΘ
	καρδ]ΠΛ, ΠΑΡ, ΣΥΝΘ	
Φωνολογικές πληροφορίες	/karðia/, /karði/	/omorf/

Πίνακας 5: Μερικώς εμπλουτισμένο δείγμα ονοματικών λημμάτων του λεξικού.

Στον παραπάνω πίνακα δίνονται δύο παραδείγματα για την κατανόηση της προτεινόμενης πληροφορίας περιβάλλοντος στο Λεξικό, όπου παρατίθενται μόνο οι μορφολογικές και φωνολογικές πληροφορίες, αφού αυτά τα στοιχεία παρουσιάζονται διαφοροποιημένα σε σχέση με τον πίνακα (3.).

Όπως θα δούμε αναλυτικότερα στην επόμενη υποενότητα η παραπάνω πρόταση είναι εν μέρει προβληματική, γιατί θεωρεί ότι τα λήμματα φέρουν πολλές πληροφορίες και περιγράφεται επαρκώς μόνο η επιθηματοποίηση και όχι η προθηματοποίηση. Αν και η σύνθεση δεν αποτελεί αντικείμενο της συγκεκριμένης εργασίας το ίδιο ισχύει και για τη σύνθεση, αφού το μη βασικό αλλόμορφο (βηματ<sub>ΣΥΝΘ</sub>) καλύπτει μόνο τις περιπτώσεις ως α' συνθετικό ή ως ΘΕΜΑ στη θέση β' συνθετικού. Οι ίδιες συνθήκες με τα ουσιαστικά σχετικά με τις μορφολογικές πληροφορίες ισχύουν και για τα λιγιστά επίθετα που εμφανίζουν αλλομορφία.

Στα ρήματα, οι πληροφορίες περιβάλλοντος εξαρτώνται άμεσα από την παρουσία ή απουσία συστηματικής αλλομορφίας. Το συγκεκριμένο κριτήριο χρησιμοποιεί η Ράλλη (2005: 132) για την προβλεψιμότητα της κλιτικής τάξης των ρημάτων. Σε γενικές γραμμές, τα ρήματα με συστηματική αλλομορφία τύπου  $X \sim X_{\{A/E\}}$  χρησιμοποιούν το βασικό αλλόμορφο για τους χρόνους του μη-συνοπτικού, π.χ.: αγαπ-ώ [Ενεστώτας], αγαπ-ούσα [Παρατατικός], θα αγαπ-ώ [Εξακολ. Μέλλοντας], και για τους αντίστοιχους χρόνους στη σύνθεση. Αντίθετα, το δεύτερο αλλόμορφο για τους συνοπτικούς χρόνους, π.χ.: αγάπη-σα [Αόριστος], έχω/ είχα αγαπή-σει [Παρακείμενος/Υπερσυντέλικος], για τους αντίστοιχους χρόνους στη σύνθεση και στην παραγωγή, όπως αγαπη-τικ-ός, αγαπη-μέν-ος. Δίνονται κάποια παραδείγματα από ρηματικά λήμματα, για να καταστεί πιο κατανοητή η αναπαράσταση.

Λήμματα ρημάτων	«αγαπώ»	«παίρνω»
Κατηγορία	Ρήμα ΚΤ2	Ρήμα ΚΤ1
Μορφολογικές πληροφορίες	αλλόμορφα: αγαπ/ αγαπα ~ αγαπη	αλλόμορφα: παιρν ~ πηρ ~ παρ

Φωνολογικές πληροφορίες	/αγαρ(a)/, /αγαρι/	/pern/, /pir/, /par/
Πληροφορίες περιβάλλοντος	αγαπ(α)] <sub>ΚΛ-ΜΗ ΣΥΝ, ΣΥΝΘ</sub>	παιρν)] <sub>ΕΝ, ΠΡΤ, ΕΞΑΚ, ΜΕΛΛ, ΣΥΝΘ</sub>
	αγαπη)] <sub>ΚΛ-ΣΥΝ, ΠΑΡ</sub>	πηρ)] <sub>ΑΟΡ</sub>
		παρ)] <sub>ΜΕΛ, ΠΡΚ, ΥΠΕΡΣ, ΠΑΡ</sub>

Πίνακας 6: Μερικώς εμπλουτισμένο δείγμα ρηματικών λημμάτων του λεξικού.

Συνοψίζοντας, υποστηρίζουμε πως στο Λεξικό καταχωρούνται οι δομές που δεν είναι προβλέψιμες, δεν είναι μορφολογικά διαφανείς και δεν έχουν σχηματισθεί με παραγωγή, κλίση ή σύνθεση. Επομένως, το Λεξικό αποτελείται από μορφήματα, π.χ. *ανθρωπ-*, *-ακι*, *υπο-*, μονομορφηματικές λέξεις, όπως *έξω*, *κάτω*, *παρά*, καθώς και μορφήματα με αδιαφανή δομή, π.χ. *συκοφαντ-*, *λεωφορ-*. Επίσης, τα λήμματα φέρουν αλλόμορφα μαζί με την πληροφορία περιβάλλοντος για τις διαδικασίες σχηματισμού λέξεων.

#### 2.4.2.1. Μορφολεξικοί κανόνες και κανόνες χρήσης αλλομορφίας

Οι υποστηρικτές της θεωρίας του ΝΛ επισημαίνουν πως στο Λεξικό υπάρχουν συσχετικοί δεσμοί μεταξύ των λημμάτων ή των διαφόρων μορφών του ίδιου λήμματος που πρέπει να αναπαριστώνται και να δηλώνονται. Η παρουσίαση των συσχετισμών γίνεται με τους **πλεοναστικούς κανόνες** (redundancy rules)<sup>31</sup>, οι οποίοι συμπληρώνουν την προβλέψιμη ή πλεοναστική πληροφορία, δεν δημιουργούν νέες δομές και δεν αλλάζουν υπάρχουσες δομές. Είναι σαφές ότι δεν είναι κανόνες που δημιουργούν νέες λέξεις. Όπως ορίζει η Ράλλη (2005: 82) «οι πλεοναστικοί κανόνες εκφράζουν διάφορες μορφολογικές σχέσεις ανάμεσα σε λεξικά τεμάχια του νοητικού λεξικού χωρίς αυτές να είναι προβλέψιμες φωνολογικά ή δομικά».

Στη συγκεκριμένη περίπτωση, εστιάζουμε στους συσχετισμούς που συνοδεύουν τα αλλόμορφα. Ειδικότερα, θα αναφερθούμε στους επονομαζόμενους **μορφολεξικούς κανόνες** (morpholexical rules), οι οποίοι περιγράφουν τον συσχετισμό μεταξύ των αλλομόρφων ενός θέματος, προσφύματος, κτλ. Όπως έχουμε προαναφέρει, η δήλωση της αλλομορφίας στους μορφολεξικούς κανόνες συμβολίζεται με την περισπωμένη (~). Τονίζεται από τη διεθνή βιβλιογραφία (Lieber 1980, Kiparsky 1982β, Archangeli 1984, Spencer 1988, Ράλλη 2005) ότι η πληροφορία της περιβαλλοντικής εξάρτησης του

<sup>31</sup> Αρχικά προτάθηκαν από τους Halle (1959), Stanley (1967) και Chomsky & Halle (1968), για να μελετηθούν αργότερα από την Lieber (1980, 1982).

αλλομόρφου πρέπει να αποτυπώνεται στον μορφολογικό κανόνα. Οι μελετητές εντάσσουν, όμως, μόνο τις πληροφορίες για το κλιτικό παράδειγμα, ενώ για τις υπόλοιπες διαδικασίες δεν παραθέτουν καμία πληροφορία. Στο ακόλουθο παράδειγμα δίνονται οι αλλομορφικές σχέσεις μεταξύ ονοματικών και ρηματικών θεμάτων.

- (38.) α. ταμια ~ ταμι  
 β. πολη ~ πολε ~ πολ  
 γ. κλιμα ~ κλιματ  
 δ. μαλων ~ μαλω  
 ε. χτυπ(α) ~ χτυπη  
 στ. παιρν ~ πηρ ~ παρ

Στην ενότητα του ΝΛ (2.4.2) επισημάναμε πως τα αλλόμορφα πρέπει να ορίζονται ως προς το μορφολογικό περιβάλλον στο οποίο εμφανίζονται. Έχουμε προσδιορίσει τα αλλόμορφα ενός λήμματος με βάση τις πληροφορίες για τα περιβάλλοντα συμμετοχής τους. Στο παράδειγμα (39.), το αλλόμορφο *ψυχη*~ της λέξης *ΨΥΧΗ* εμφανίζεται στον ενικό αριθμό, ενώ για τον πληθυντικό χρησιμοποιείται το αλλόμορφο *ψυχ*~. Παράλληλα παρατηρούμε ότι το συγκεκριμένο αλλόμορφο χρησιμοποιείται στην παραγωγή (*ψυχ-ικ-ός*) και στη σύνθεση (*ψυχ-ο-πονιάρης*). Αντίστοιχα, στα ρήματα το αλλόμορφο *αγαπ*- χρησιμοποιείται στους κλιτικούς τύπους του μη-συνοπτικού του ρήματος *αγαπώ*, καθώς και στη σύνθεση, ενώ το αλλόμορφο *αγαπη*- στα παραδείγματα των συνοπτικών χρόνων, των αντίστοιχων στη σύνθεση και στην παραγωγή. Η Ράλλη (1988, 2005) παρατηρεί πως ανάλογα με την περίπτωση η πληροφορία της περιβαλλοντικής εξάρτησης πρέπει να αποτυπώνεται στον μορφολογικό κανόνα, όπως στο ακόλουθο παράδειγμα:

- (39.) α. *ψυχη* / — ~ *ψυχ*  
 [ενικός]  
 β. *αγαπ(α)* / — ~ *αγαπη*  
 [+μη συνοπτικό]

Οι παραπάνω μορφολογικοί κανόνες με τις πληροφορίες περιβάλλοντος παρέχουν το πλαίσιο της υποκατηγοριοποίησης για τη χρήση των αλλομόρφων στην κλίση. Ωστόσο, προκύπτει το πρόβλημα με την καταχώρηση των υπολοίπων πληροφοριών περιβάλλοντος για τις λοιπές διαδικασίες σχηματισμού. Αυτές οι πληροφορίες είναι απαραίτητες σε ένα μοντέλο υπολογιστικής επεξεργασίας, όπου πρέπει να καθορίζεται

το περιβάλλον συμμετοχής κάθε αλλόμορφου. Με βάση την προσέγγισή μας οι μορφολεξικοί κανόνες οφείλουν να είναι περιγραφικοί και συμπτωνόμενοι, για να αποφύγουμε τη χρήση πιο πολύπλοκων περιγραφών, όπως τους ενδείκτες των Pirrelli & Battista (2000a) ή πιο αντιοικονομικών, δηλαδή να καταχωρούνται όλες οι πληροφορίες περιβάλλοντος σε κάθε λήμμα. Ακολουθώντας παρουσιάζουμε τους μορφολεξικούς κανόνες του προηγούμενου παραδείγματος με μία αρχική παρουσίαση της πρότασής μας:

$$\begin{array}{l}
 (40.) \quad \alpha. \text{ ψυχη}] \quad \sim \text{ ψυχ]} \\
 \left( \begin{array}{l} \text{ΚΛ}^{32} \\ \text{Β'Σ\_ΣΥΝΘ} \end{array} \right) [\text{ΕΝΙΚΟΣ}] \quad \left( \begin{array}{l} \text{ΚΛ} \\ \text{Β'Σ\_ΣΥΝΘ} \end{array} \right) [\text{ΠΛΗΘΥΝΤΙΚΟΣ}] \\
 \left( \begin{array}{l} \text{ΠΡΓ} \\ \text{Α'Σ\_ΣΥΝΘ} \end{array} \right) \\
 \\
 \beta. \text{ αγαπ(α)]} \quad \sim \text{ αγαπη]} \\
 \left( \begin{array}{l} \text{ΚΛ} \\ \text{ΠΡΟΘΠΡΓ} \\ \text{Β'Σ\_ΣΥΝΘ} \\ \text{Α'Σ\_ΣΥΝΘ} \end{array} \right) [\text{ΜΗ-ΣΥΝΟΠΤΙΚΟ}] \quad \left( \begin{array}{l} \text{ΚΛ} \\ \text{ΠΡΟΘΠΡΓ} \\ \text{Β'Σ\_ΣΥΝΘ} \\ \text{ΕΠΙΘΠΡΓ} \end{array} \right) [\text{ΣΥΝΟΠΤΙΚΟ}]
 \end{array}$$

Έτσι, παρατηρούμε πως στο παράδειγμα (40.) ο μορφολεξικός κανόνας συνδέει τα δύο αλλόμορφα του λεξήματος ΨΥΧΗ και δηλώνει ότι το αλλόμορφο *ψυχη*~ συναντάται στους τύπους του ενικού στην Κλίση και στη Σύνθεση ως β' συνθετικό, ενώ το αλλόμορφο *ψυχ*~ απαντάται στους τύπους του πληθυντικού στην Κλίση και στη Σύνθεση ως β' συνθετικό, καθώς και σε όλους τους τύπους της Παραγωγής και της Σύνθεση ως α' Συνθετικό. Παρόμοιες είναι και οι πληροφορίες για το μορφολογικό περιβάλλον του ρήματος. Στα μη προκαθορισμένα αλλόμορφα θα μπορούσαμε να μην εισάγουμε τις πληροφορίες των μορφολογικών περιβαλλόντων στα οποία εμφανίζονται, γιατί εννοούνται όλα τα υπόλοιπα περιβάλλοντα που δεν καταχωρούνται στο βασικό αλλόμορφο. Κρίνουμε όμως ότι η τακτική να παραλειφθούν οι πληροφορίες από τα μη-προκαθορισμένα αλλόμορφα καταστρέφει την ομοιομορφία στη δομή ενός κανόνα και προκαλεί ποικίλα προβλήματα στην οποιαδήποτε υπολογιστική του υλοποίηση.

<sup>32</sup> Η κωδικοποίηση του περιβάλλοντος συμμετοχής είναι ΚΛ για Κλίση, Α'Σ\_ΣΥΝΘ για Α'Συστατικό Σύνθεσης, Β'Σ\_ΣΥΝΘ για Β' Συνθετικό Σύνθεσης, ΠΡΓ για Παραγωγή, ΠΡΟΘΠΡΓ για Προθηματοποιημένη Παράγωγή και ΕΠΙΘΠΡΓ για Επιθηματοποιημένη Παραγωγή.

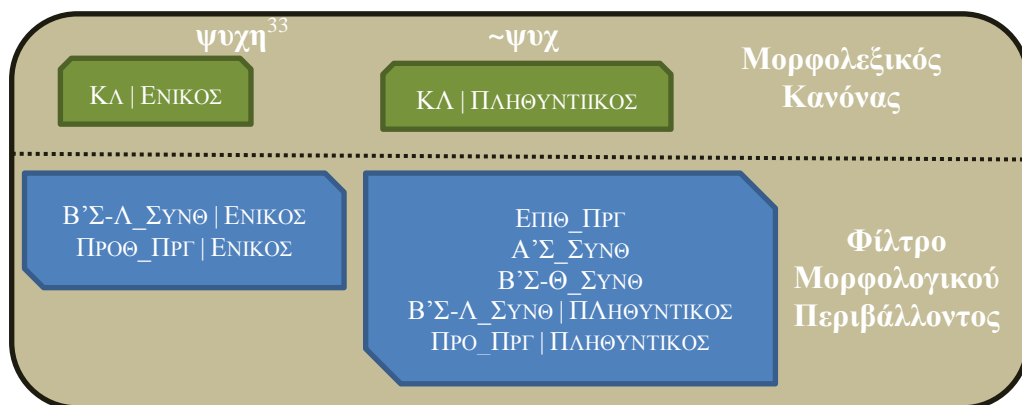


Η Ράλλη (2005: 83) διευκρινίζει πως οι μορφολεξικοί κανόνες δεν έχουν εφαρμογή σε αλλόμορφα που εμφανίζονται σε μορφολογικά πολύπλοκα περιβάλλοντα. Έτσι δίνοντας το ακόλουθο παράδειγμα, αναφέρει πως ο κανόνας για τα αλλόμορφα του ρήματος πέφτω είναι *πεφτ~ πεσ* και πως δεν μπορεί να ισχύσει και για τον πολύπλοκο σχηματισμό του *ξεπέφτω* (\**ξεπεφτ~ ξεπεσ*).

Προκύπτει το ζήτημα για την καταχώρηση της πληροφορίας σχετικά με το μορφολογικό περιβάλλον συμμετοχής των αλλομόρφων για υπολογιστικούς λόγους, καθώς και για θεωρητικούς λόγους. Η αρχική μας πρόταση, όπως παρουσιάστηκε στο (40.) για την καταχώρηση όλων των πληροφοριών περιβάλλοντος στους μορφολεξικούς κανόνες εξασφαλίζει τη συμμετοχή του κατάλληλου αλλόμορφου σε κάθε διαδικασία σχηματισμού λέξεων και εσωκλείει όλες τις απαραίτητες πληροφορίες. Ωστόσο αυτή η πρόταση είναι μάλλον αντικοινωνική, αφού εισάγει τις ίδιες – πολλές – πληροφορίες σε όλα τα λήμματα της ίδιας τάξης. Επομένως αποδεχόμαστε ότι οι μορφολεξικοί κανόνες συνδέουν απλώς δύο αλλόμορφα, ενώ περιλαμβάνουν και τις πληροφορίες που συνοδεύουν τα συγκεκριμένα αλλόμορφα στη διαδικασία της *κλίσης*.

Στις ακόλουθες παραγράφους θα προτείνουμε ένα μοντέλο καταχώρησης πληροφοριών για τα ονοματικά, επιθετικά και ρηματικά αλλόμορφα. Το συγκεκριμένο μοντέλο κρίνεται αναγκαίο και απαραίτητο για την υπολογιστική επεξεργασία της αλλομορφίας, καθώς και για την υπολογιστική ανάλυση και δημιουργία των παράγωγων λέξεων. Εισάγουμε την έννοια των **κανόνων χρήσης αλλομορφίας**, οι οποίοι είναι κανόνες ανάθεσης μορφολογικού περιβάλλοντος για τα αλλόμορφα στη συμμετοχή τους σε *παραγωγή* και *σύνθεση*. Θέτουμε εύλογα τα παρακάτω ερωτήματα. Πώς προκρίνεται η κατάλληλη μορφή ενός μορφήματος κατά τη διαδικασία σχηματισμού ενός παράγωγου ή ενός σύνθετου; Από πού τροφοδοτείται η παραγωγή και η σύνθεση με τις κατάλληλες πληροφορίες περιβάλλοντος για τη σωστή επιλογή; Πού εντάσσονται αυτές οι πληροφορίες και με ποια μορφή είναι κωδικοποιημένες; Οι απαντήσεις σε αυτά τα ερωτήματα εντοπίζονται στους κανόνες χρήσης αλλομορφίας που προτείνουμε. Αυτοί οι κανόνες προκρίνουν το κατάλληλο αλλόμορφο, τροφοδοτούνται από τους μορφολεξικούς κανόνες με τη συνδρομή του φίλτρου μορφολογικού περιβάλλοντος και εντάσσονται στον τομέα της γραμματικής .).

Τοποθετούμε τους συγκεκριμένους κανόνες χρήσης αλλομορφίας εκτός Λεξικού και εντός του πλαισίου της γραμματικής. Οι κανόνες χρήσης αλλομορφίας τροφοδοτούνται από δύο διαφορετικές πηγές: από τους μορφολεξικούς κανόνες για τις πληροφορίες της κλίσης και από το **φίλτρο μορφολογικού περιβάλλοντος** που φέρει τις απαραίτητες πληροφορίες για παραγωγή και σύνθεση. Σκοπός των κανόνων χρήσης αλλομορφίας που προτείνουμε είναι να προκριθεί το κατάλληλο αλλόμορφο για κάθε διαδικασία. Οι συγκεκριμένοι κανόνες δεν είναι κανόνες δημιουργίας στοιχείων και δείχνουν τη σχέση των αλλομόρφων σε συνάρτηση με το περιβάλλον συμμετοχής τους στις διαδικασίες σχηματισμού λέξεων.



Για να καταστήσουμε κατανοητό γιατί είναι απαραίτητη η εισαγωγή των κανόνων χρήσης αλλομορφίας, καθώς και τον λόγο που εντάσσονται στον τομέα της γραμματικής, παρουσιάζουμε συγκεκριμένα παραδείγματα. Εξετάζοντας την ονοματική και ρηματική κλίση, παρατηρούμε πως όλα τα ουσιαστικά που εντάσσονται στο ίδιο κλιτικό παράδειγμα ή εντάσσονται στις ίδιες κλιτικές τάξεις (Ralli 1988, 2000), χρησιμοποιούν τα αλλόμορφα ιδίου τύπου στις ίδιες διαδικασίες. Για παράδειγμα, το ουσιαστικό *ψυχή* ανήκει στην ΚΤ3 και έχει για αλλόμορφα τα *ψυχη~ψυχ'* το αλλόμορφο *ψυχη~* εμφανίζεται στον ενικό προθηματοποιημένων παράγωγων (*αναψυχή*) και συνθέτων με β' συνθετικό λέξη (*ανθοψυχή*) και το αλλόμορφο *ψυχ~* στον πληθυντικό των προηγούμενων τύπων, καθώς και στα επιθηματοποιημένα παράγωγα (*ψυχικός*), στα σύνθετα με α' συνθετικό (*ψυχοπομπός*) και με β' συνθετικό θέμα (*επτάψυχος*). Όλα τα θηλυκά ουσιαστικά της ίδιας κλιτικής τάξης (ΚΤ3), όπως

<sup>33</sup> Η κωδικοποίηση του περιβάλλοντος συμμετοχής είναι ΚΛ για Κλίση, Α'Σ\_ΣΥΝΘ για Α'Συστατικό Σύνθεσης, Β'Σ-Λ\_ΣΥΝΘ για Β' Συνθετικό Λέξη Σύνθεσης, Β'Σ-Θ\_ΣΥΝΘ για Β' Συνθετικό Θέμα Σύνθεσης ΠΡΟΘΠΡΓ για Προθηματοποιημένη Παράγωγή και ΕΠΙΘΠΡΓ για Επιθηματοποιημένη Παραγωγή.

αυλή, αδελφή, ακτή, συμμετέχουν με τα αλλόμορφα ίδιου τύπου στις αντίστοιχες περιπτώσεις των διαδικασιών σχηματισμού λέξεων.

Γενικεύοντας το παραπάνω παράδειγμα, υποστηρίζουμε πως όλα τα ουσιαστικά κάθε κλιτικής τάξης συμμετέχουν με αλλόμορφα του ίδιου τύπου στις αντίστοιχες διαδικασίες χωρίς να αποκλίνουν από το σύνολο της κάθε κλιτικής τάξης.

Το φίλτρο μορφολογικού περιβάλλοντος παρέχει τις σωστές πληροφορίες στον κανόνα χρήσης αλλομορφίας, γιατί ελέγχει τον δείκτη κλιτικής τάξης του λήμματος. Πιο συγκεκριμένα, το φίλτρο μορφολογικού περιβάλλοντος έχει τεσσάρων ειδών σύνολα πληροφοριών για τα ουσιαστικά αναλόγως με τον κλιτικό δείκτη. Το πρώτο είδος είναι για τα ουσιαστικά χωρίς αλλόμορφα, όπου όλες οι πληροφορίες εντάσσονται στη μοναδική μορφή και περιλαμβάνει τα ουσιαστικά των ΚΤ1, ΚΤ6, ΚΤ7. Το δεύτερο είδος είναι για όσα ουσιαστικά έχουν αλλόμορφα, όπου οι κατάλληλες πληροφορίες εντάσσονται στα αντίστοιχα αλλόμορφα και αφορά στα ουσιαστικά των ΚΤ2, ΚΤ3, ΚΤ5. Το τρίτο είδος είναι για τα ουσιαστικά της ΚΤ4 που εμφανίζουν στη γενική ενικού και πληθυντικού διαφορετικό αλλόμορφο, ενώ τα μορφολογικά περιβάλλοντα των υπόλοιπων διαδικασιών σχηματισμού λέξεων είναι όμοια με τα αντίστοιχα του δεύτερου είδους. Τέλος, το τέταρτο είδος είναι για τα ουσιαστικά της ΚΤ8 που εμφανίζουν στη γενική ενικού το αλλόμορφο του πληθυντικού, ενώ οι υπόλοιπες πληροφορίες αντιστοιχούν στις ίδιες διαδικασίες.

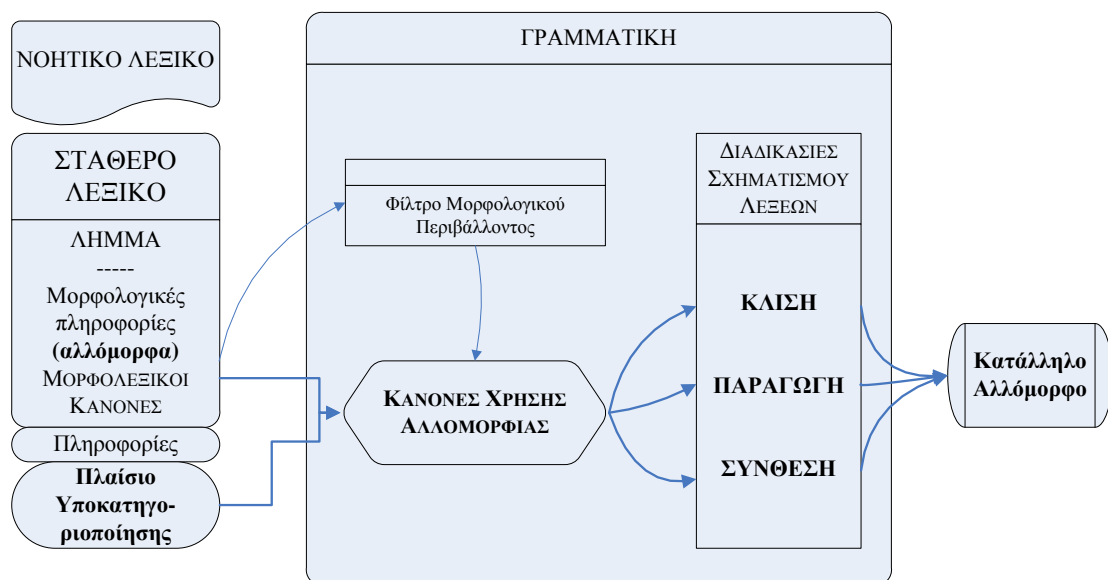
Το φίλτρο τροφοδοτεί τους κανόνες χρήσης αλλομορφίας με τις σωστές πληροφορίες ανάλογα με τις ονοματικές και επιθετικές ομάδες. Στο παρακάτω πίνακα παρουσιάζουμε τις τέσσερις διαφορετικές ομάδες, όπου αναφέρουμε ως ΑΛΛΟΜ1 το βασικό αλλόμορφο και ως ΑΛΛΟΜ2 και ΑΛΛΟΜ3 τα μη-προκαθορισμένα, όπου αυτά υπάρχουν· δεν δίνουμε τις πληροφορίες για την κλίση, καθότι αυτές οι πληροφορίες είναι καταχωρημένες στους μορφολεξικούς κανόνες κάθε λήμματος.

ΟΜΑΔΑ1		ΟΜΑΔΑ2	
ΑΛΛΟΜ1	ΑΛΛΟΜ1	ΑΛΛΟΜ1	ΑΛΛΟΜ2
[Όλα τα περιβάλλοντα]	B'Σ-Λ_ΣΥΝΘ   ΕΝΙΚΟΣ ΠΡΟΘ_ΠΡΓ   ΕΝΙΚΟΣ		ΕΠΙΘ_ΠΡΓ Α'Σ_ΣΥΝΘ B'Σ-Θ_ΣΥΝΘ B'Σ-Λ_ΣΥΝΘ   ΠΛΗΘΥΝΤΙΚΟΣ ΠΡΟ_ΠΡΓ   ΠΛΗΘΥΝΤΙΚΟΣ
ΟΜΑΔΑ3		ΟΜΑΔΑ4	

ΑΛΛΟΜ1	ΑΛΛΟΜ2	ΑΛΛΟΜ3	ΑΛΛΟΜ1	ΑΛΛΟΜ2
B'Σ-Λ_ΣΥΝΘ   ΕΝΙΚΟΣ{-ΓΕΝ}	ΕΠΙΘ_ΠΡΓ Α'Σ_ΣΥΝΘ	B'Σ-Λ_ΣΥΝΘ   ΓΕΝ{ΕΝ, ΠΛΗΘ}	B'Σ-Λ_ΣΥΝΘ   ΕΝΙΚΟΣ{-ΓΕΝ}	ΕΠΙΘ_ΠΡΓ Α'Σ_ΣΥΝΘ
ΠΡΟΘ_ΠΡΓ   ΕΝΙΚΟΣ{-ΓΕΝ}	B'Σ-Θ_ΣΥΝΘ B'Σ-Λ_ΣΥΝΘ   ΠΛΗΘ{-ΓΕΝ}	ΠΡΟΘ_ΠΡΓ   ΓΕΝ{ΕΝ, ΠΛΗΘ}	ΠΡΟΘ_ΠΡΓ   ΕΝΙΚΟΣ{-ΓΕΝ}	B'Σ-Θ_ΣΥΝΘ B'Σ-Λ_ΣΥΝΘ   ΠΛΗΘ, ΓΕΝ_ΕΝ
	ΠΡΟ_ΠΡΓ   ΠΛΗΘ{-ΓΕΝ}			ΠΡΟ_ΠΡΓ   ΠΛΗΘ, ΓΕΝ_ΕΝ

Πίνακας 7: Οι τέσσερις ομάδες διαφορετικών πληροφοριών για τα ουσιαστικά όπως καταχωρούνται στο φίλτρο μορφολογικού περιβάλλοντος.

Ας παρουσιάσουμε με παραδείγματα πώς λειτουργεί το φίλτρο μορφολογικού περιβάλλοντος. Στην περίπτωση του μορφήματος *θάλασσα* το φίλτρο μορφολογικού περιβάλλοντος ελέγχει την ΚΤ, ώστε να τροφοδοτήσει με τις σωστές πληροφορίες τους κανόνες χρήσης. Έχοντας πρόσβαση στις πληροφορίες του Λεξικού, το φίλτρο σημειώνει ότι η *θάλασσα* ανήκει στην ΚΤ3 και επομένως αντιστοιχεί στη λέξη τις πληροφορίες περιβάλλοντος της ομάδας που ανήκει, δηλαδή τη δεύτερη. Ακολούθως οι κανόνες χρήσης αλλομορφίας επιλέγουν το κατάλληλο αλλόμορφο παραδείγματος χάριν, για την επιθηματοποιημένη παραγωγή προκρίνουν το ΑΛΛΟΜ2. Αντίστοιχα στην περίπτωση του μορφήματος *βήμα* το φίλτρο μορφολογικού περιβάλλοντος ελέγχει την ΚΤ και το εντάσσει στην τέταρτη ομάδα. Ακολούθως οι κανόνες χρήσης αλλομορφίας επιλέγουν το κατάλληλο αλλόμορφο που είναι το ΑΛΛΟΜ2 στην περίπτωση που εμφανίζεται ως πρώτο συστατικό σύνθετης λέξης.



Γράφημα 1: Ο συσχετισμός Σταθερού Λεξικού με τους μορφολεξικούς κανόνες.

Στο παραπάνω γράφημα παρουσιάζουμε τη θέση των μορφολεξικών κανόνων (εντός του Λεξικού), καθώς και την αντίστοιχη θέση των κανόνων χρήσης

αλλομορφίας που τροφοδοτούνται από τους μορφολεξικούς κανόνες και το φίλτρο μορφολογικού περιβάλλοντος. Οι κανόνες χρήσης αλλομορφίας εντάσσονται στο χώρο της Γραμματικής και ενεργούν πάντα πριν οποιαδήποτε διαδικασία σχηματισμού λέξεων και είναι υπεύθυνοι για την αποφυγή αντιγραμματικών λέξεων.

Συνοψίζοντας αναφέρουμε ότι υιοθετούμε την άποψη ότι οι μορφολεξικοί κανόνες φέρουν τις πληροφορίες για την κλίση, όπως είδαμε στην ενότητα 2.4.2 στην καταχώρηση του λήμματος. Ωστόσο, για τις πληροφορίες των υπολοίπων διαδικασιών εισάγαμε τους κανόνες χρήσης αλλομορφίας, ώστε να προκρίνονται τα κατάλληλα αλλομόρφα στην παραγωγή και στη σύνθεση. Όπως δείχνει και το γράφημα (1.), οι κανόνες χρήσης αλλομορφίας τοποθετούνται εκτός Λεξικού στον τομέα της Γραμματικής στο επίπεδο της μορφολογίας και χρησιμοποιούνται σε όλες τις διαδικασίες σχηματισμού λέξεων για την τελική επιλογή του κατάλληλου αλλομόρφου. Άρα οι κανόνες χρήσης αλλομορφίας με τις σωστές πληροφορίες μορφολογικού περιβάλλοντος των αλλομόρφων, αν και δεν δημιουργούν νέες δομές και δεν αλλάζουν υπάρχουσες δομές, εντούτοις διασφαλίζουν τις δομές, ώστε είναι γραμματικές με την επιλογή του κατάλληλου αλλομόρφου.

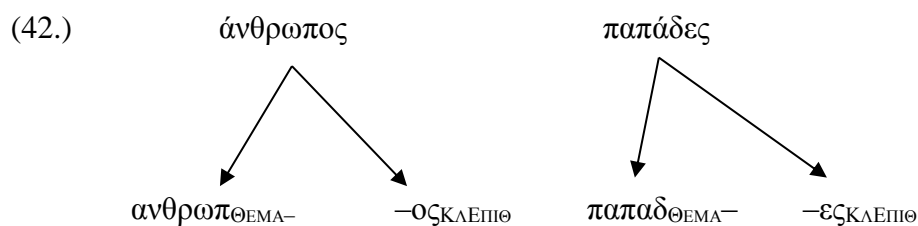
### 2.4.3. Η εμφάνιση των αλλομόρφων στην κλίση

Στην παρούσα ενότητα θα παρουσιάσουμε την άποψή μας για ποιο λόγο θεωρούμε ότι τα αλλομόρφα πρωτοδημιουργούνται διαχρονικά στην κλίση και γιατί αυτή αποτελεί σημείο αναφοράς για την αλλομορφική συμπεριφορά των θεμάτων, αλλά και των επιθημάτων. Θα δείξουμε τις κλιτικές αλλομορφικές συμπεριφορές των θεμάτων (και όχι των κλιτικών προσφυμάτων), καθότι αυτές θα μας είναι απαραίτητες στην ανάλυση της αλλομορφίας στην παραγωγή. Ήδη από τις πρώτες έρευνες της αλλομορφίας (Lieber 1980 και Carstairs 1987) τονίζεται η αλληλένδετη σχέση μεταξύ αυτών των δύο, αφού συμβάλλουν στην αρχή της οικονομίας στο κλιτικό παράδειγμα, η οποία επιτυγχάνεται με τη μία και μόνο κλιτική πραγμάτωση κάθε λέξης μίας τάξης. Μάλιστα η Ralli (2007) υποστηρίζει πως η αλλομορφία συχνά συμβάλλει στην παραδειγματική συνοχή και ακεραιότητα του κλιτικού συστήματος της Ελληνικής αποδεικνύοντας την αμφίδρομη σχέση που διέπει το φαινόμενο της αλλομορφίας με την κλίση.

### 2.4.3.1. Η εξάρτηση της αλλομορφικής συμπεριφοράς από την κλίση

Υποθέτουμε ότι τα αλλόμορφα εμφανίζονται αρχικά στην κλίση που είναι η λιγότερα πολύπλοκη μορφολογική διαδικασία σχηματισμού λέξεων. Αναλόγως προς την αλλομορφική συμπεριφορά του μορφήματος στην κλίση, αναμένουμε και αντίστοιχες συμπεριφορές στις άλλες μορφολογικές διαδικασίες, δηλαδή εφόσον υπάρχουν αλλόμορφα στην κλίση, το ίδιο θα ισχύει και στην παραγωγή και τη σύνθεση. Ήδη στις πρώτες ερευνητικές προσπάθειες για την αλλομορφία οι Lieber (1982) και Carstairs (1987) αναλύουν το φαινόμενο στο επίπεδο της κλίσης. Χαρακτηριστικό είναι ότι καταχωρούν τα αλλόμορφα στο Λεξικό με κλιτικά χαρακτηριστικά<sup>34</sup>, όπως αλλόμορφο ΕΝΙΚΟΥ και ΠΛΗΘΥΝΤΙΚΟΥ (στα Γερμανικά) ή αλλόμορφο παρελθοντικών χρόνων (Αγγλικά, Γερμανικά). Παρόμοια τακτική χρησιμοποιεί και η Ράλλη (2005) στις πληροφορίες που συνοδεύουν τους μορφολεξικούς κανόνες (όπως, +ΕΝΙΚΟΣ, +ΣΥΝΟΠΤΙΚΟ). Ήδη κατά την περιγραφή της αλλομορφίας στη σύνθεση και στην παραγωγή, οι περισσότεροι μορφολόγοι (Lieber 1982, Booij 1997, Pirrelli & Battista 2001, Ράλλη 2005, 2008) χρησιμοποιούν τα αλλόμορφα της κλίσης στην παραγωγή και τη σύνθεση.

Στη διαδικασία της κλίσης τα απαραίτητα συστατικά στην κλίση είναι ένα θέμα και ένα κλιτικό επίθημα (ακόμα και μηδενικής μορφής) στα ουσιαστικά και επίθετα, όπως στο παράδειγμα (42.), ενώ στη ρηματική κλίση το θέμα μπορεί να ακολουθείται από δύο κλιτικά επιθήματα. Η συμπεριφορά του θέματος απέναντι στα κλιτικά επιθήματα δεν καθορίζει την αλλομορφία τους.



Η πρότασή μας για την εξάρτηση της αλλομορφικής συμπεριφοράς των μορφημάτων από την κλίση βασίζεται σε δύο σκέλη: (α.) στην απουσία παραγωγικών επιθημάτων από περαιτέρω επιθηματοποίηση ή από τη σύνθεση και (β.) στην επανάλυση των μορφηματικών ορίων στην κλίση.

<sup>34</sup> Μόνο διάφοροι υποστηρικτές της Θεωρίας του Βέλτιστου (Bonet 2004, Lubowicz *et al.* 2006, Mascaró 2006, Bye 2007, Bonet & Mascaró 2007, μεταξύ άλλων) χαρακτηρίζουν τα αλλόμορφα με φωνολογικά κριτήρια ή ανάλογα με τα φωνολογικά περιβάλλοντα που εμφανίζονται και όχι με τις μορφολογικές διαδικασίες που συμμετέχουν.

Για να υποστηρίξουμε την άποψη ότι τα αλλόμορφα εμφανίζονται αρχικά κατά τη διαδικασία της κλίσης, ας παρατηρήσουμε τη συμπεριφορά των προσφυμάτων των Ελληνικών. Όλα τα παραγωγικά επιθήματα πλην των επιρρηματικών συνδυάζονται με μια βάση για τη δημιουργία παράγωγης λέξης και συνοδεύονται πάντα από κλιτικά επιθήματα. Αρκετά παραγωγικά επιθήματα όμως δεν συμμετέχουν σε διαδικασίες περαιτέρω επιθηματοποίησης ή στη σύνθεση. Δεν αποτελούν δηλαδή συστατικό μίας παράγωγης βάσης, η οποία θα συνδυαστεί με ένα παραγωγικό επίθημα για τη δημιουργία ενός νέου παράγωγου ή με κάποιο θέμα/ λέξη για τη δημιουργία μίας σύνθετης λέξης. Αυτά τα επιθήματα συνδυάζονται εκ δεξιών μόνο με κλιτικά επιθήματα και ποτέ με άλλα παραγωγικά επιθήματα ή θέματα. Με άλλα λόγια, αυτά τα παραγωγικά επιθήματα συμμετέχουν με τα αλλόμορφα τους (εφόσον υπάρχουν αλλόμορφα) μόνο στη διαδικασία της κλίσης.

Τα Ελληνικά χαρακτηρίζονται από πλούσια παραγωγή και από αρκετή συνδυαστικότητα μεταξύ των επιθημάτων. Οι Μελισσαροπούλου (2007) και Melissaropoulou & Ralli (2009) σημειώνουν χαρακτηριστικά πως στα Ελληνικά είναι εφικτή η ύπαρξη έως και πέντε παραγωγικών επιθημάτων στη σειρά σε μία λέξη. Για παράδειγμα, από την παράγωγη λέξη *χορ-εύ(ω)* δημιουργούμε τη λέξη *χορ-ευ-τη(ς)* και ακολούθως μια νέα παράγωγη λέξη, το *χορ-ευ-τ-ικ(ος)*· από την παράγωγη λέξη *χωρ-ίζ(ω)* δημιουργούμε τη λέξη *χωρ-ίσ-τρα* και στη συνέχεια το υποκοριστικό *χωρ-ισ-τρ-ούλα*. Εντούτοις, υπάρχουν αρκετά επιθήματα που δεν επιδέχονται να έπονται άλλα παραγωγικά επιθήματα. Επιθήματα, όπως *-οσύνη*, *-ότητα*, *-ένιος*, *-ίλα*, *-έας* κλπ. (43.), δεν εντοπίζονται σε παράγωγες λέξεις, όπου τα συγκεκριμένα επιθήματα να ακολουθούνται από άλλα παραγωγικά επιθήματα.

(43.)	δίκαι-ος	+	-οσύνη	>	δικαιοσύνη
	μετριόφρων	+	-οσύνη	>	μετριοφροσύνη
	ένα-ς	+	-ότητα	>	ενότητα
	φυσικ-ός	+	-ότητα	>	φυσικότητα
	γρανίτη-ς	+	-ένιος	>	γρανιτένιος
	σίδερ-ο	+	-ένιος	>	σιδερένιος
	καπν-ός	+	-ίλα	>	καπνίλα
	πράσιν-ος	+	-ίλα	>	πρασινίλα
	γράμμα	+	-έας	>	γραμματέας

συγγράφ-ω + -έας > συγγραφέας

Δηλαδή οι παραπάνω παράγωγες λέξεις (43.) αποκλείεται να εμφανιστούν ως βάση σε κάποιο άλλο επιθηματοποιημένο παράγωγο, όπως \*γρानιτενιοσύνη, \*καπνιλίζω, \*δικαιοσυνίτσα, \*φυσικοτητεύω, \*σιδερενιότητα. Κάποια από τα παραπάνω επιθήματα εμφανίζουν αλλόμορφια, όπως -ότητα (-οτητα~ -οτητ), -οσύνη (-οσυνη~ -οσυν), -ίλα (-ίλα~ -ιλ). Τα αλλόμορφα τους εμφανίζονται μόνο όταν συνδυαστούν με κλιτικά επιθήματα στο κλιτικό παράδειγμα κάθε παράγωγης λέξης. Αντίθετα με άλλα παραγωγικά επιθήματα δεν εμφανίζουν αλλόμορφα στην παραγωγή, γιατί δεν συνδυάζονται με άλλα παραγωγικά επιθήματα όντας σε θέση βάσης μαζί με κάποιο θέμα. Αντίθετα με τα παραπάνω, για παράδειγμα, το παραγωγικό επίθημα -τη(ς) που έχει αλλόμορφα τα -τη- ~ -τ-, συμμετέχει ως βάση παράγωγου θέματος με το αλλόμορφο -τ-, π.χ. χορευ-τη(ς) > χορευτ-ικ(ός), ράφ-τη(ς) > ραφτ-ούλη(ς).

Αντίστοιχη είναι η συμπεριφορά των παραγωγικών επιθημάτων του (43.) και στη σύνθεση. Τα συγκεκριμένα επιθήματα δεν συμμετέχουν ποτέ (μαζί με κάποιο θέμα) ως πρώτο συστατικό σύνθετης λέξης. Τα συγκεκριμένα επιθήματα, όπως αρκετά άλλα παραγωγικά που έχουν την ίδια συμπεριφορά, εμφανίζουν αλλομορφία μόνο στην κλίση. Επομένως, όσα παραγωγικά προσφύματα δεν συμμετέχουν σε επιπλέον παραγωγικές διαδικασίες και ακολουθώς σε σύνθετες λεξικές δομές εμφανίζουν τα αλλόμορφα τους, μόνο και όταν συνδυαστούν με κλιτικά επιθήματα. Κανένα θέμα ή παραγωγικό επίθημα – πλην των επιρρηματικών – δεν απουσιάζει από το στάδιο της κλίσης, ενώ υπάρχει μία ομάδα επιθημάτων που δεν συμμετέχει στη διαδικασία παραγωγής σε θέση μη-κεφαλής. Η αλλομορφία που εμφανίζει αυτή η ομάδα θεμάτων και επιθημάτων είναι μόνο στην κλίση, αφού απουσιάζουν από τις υπόλοιπες διαδικασίες.

Αναλυτικότερα για τη σύνθεση, υπάρχουν αρκετά παραγωγικά επιθήματα, τα οποία υφίστανται μορφολογικά και σημασιολογικά, αλλά διαγράφονται επιφανειακά στη σύνθεση, όταν βρεθούν στο αριστερό μέρος ενός συνθέτου, δηλαδή στη θέση του πρώτου συνθετικού. Όπως έχουν δείξει οι Καρασίμος (2001) και Ralli & Karasimos (2008, 2009) στα ρηματικά παρατακτικά σύνθετα, το πρώτο ρήμα χάνει το παραγωγικό επίθημα, εφόσον υπάρχει. Ο **περιορισμός του γυμνού θέματος** (bare-stem constraint) αποτρέπει τα ρηματικά παραγωγικά επιθήματα από το να πραγματοποιούν



επιφανειακά<sup>35</sup>, όταν αποτελούν μέρος μιας παράγωγης ρηματικής βάσης, όπως στα παραδείγματα του (44.α). Ο περιορισμός αποκλείει τα επιθήματα να πραγματοποιούν φωνολογικά και να συμμετέχουν με κάποιο αλλόμορφό τους στη σύνθεση σε θέση α' συνθετικού. Αντίθετα τα παραγωγικά επιθήματα εμφανίζονται στο δεξί μέρος του συνθέτου ως β' συνθετικά. Σε αυτήν τους τη θέση παρουσιάζουν αλλομορφία, γιατί συνδυάζονται και πάλι με κλιτικά επιθήματα.

(44.)	α.	αλων-ίζ(ω)	θερίζω	⇒	αλωνοθερίζω
		κλειδ-ών(ω)	αμπαρώνω	⇒	κλειδαμπαρώνω
		χορ-εύ(ω)	πηδώ	⇒	χοροπηδώ
		σκαλ-ίζ(ω)	ποτίζω	⇒	σκαλοποτίζω
	β.	αέρ-ι(ο)	αγωγός	⇒	αεριαγωγός vs. αεραγωγός
		αγρ-ότη(ς)	σπίτι	⇒	αγροτόσπιτο vs. αγρόσπιτο

Πλην ελαχίστων περιπτώσεων (44.β) τα παραγωγικά επιθήματα δεν ακολουθούνται από θέματα ή λέξεις. Επομένως, απουσιάζουν από τη διαδικασία της σύνθεσης και ο μοναδικός τους συνδυασμός με τη σύνθεση είναι η προσκόλληση του στο δεξί μέρος σύνθετων λέξεων, όπως π.χ. σπирτόκουτο > σπирτοκουτάκι, καρεκλοπόδαρο > καρεκλοποδαράτος δημιουργώντας παράγωγα από σύνθετες βάσεις. Επίσης είναι δυνατόν να συμμετέχουν ως μέρος λέξης σε θέση β' συνθετικού, όπως παλιοκαπνίλα, βρωμοδικαιοσύνη, κωλοσυγγραφέας. Θεωρούμε ότι τα παραγωγικά επιθήματα δεν «εμφανίζουν» αλλομορφία στη σύνθεση, αφού ποτέ δεν ακολουθούνται από κάποιο θέμα, ενώ στις περιπτώσεις που συμμετέχουν ως μέρος λέξης σε θέση β' συνθετικού ισχύουν οι κανόνες της κλίσης.

Ο δεύτερος λόγος αιτιολόγησης της αιτιακής σχέσης ανάμεσα στην εμφάνιση αλλομόρφων και στην κλίση είναι σε σημαντικό βαθμό διαχρονικός. Θεωρούμε ότι η κλίση είναι το επίπεδο, στο οποίο πρωτοεμφανίζονται χρηστικά τα αλλόμορφα, τα οποία συμμετέχουν και στις υπόλοιπες – πιο πολύπλοκες – μορφολογικές διαδικασίες, όπως θα δούμε στο επόμενο κεφάλαιο. Αναφέρουμε τον όρο «πρωτοεμφανίζονται/πρωτοχρησιμοποιούνται», καθότι διαχρονικά όλες οι αλλαγές στα μορφήματα που οδήγησαν συγχρονικά στην εμφάνιση των αλλομόρφων, συντελέστηκαν στην κλίση είτε λόγω της επανάλυσης, είτε της απενεργοποίησης φωνολογικών κανόνων.

<sup>35</sup> Όπως υποστηρίζουμε (Ralli & Karasimos 2008, 2009) τα μορφολογικά και σημασιολογικά χαρακτηριστικά του ρήματος υφίστανται, απλώς η φωνολογική μορφή δεν πραγματώνεται.

Τέτοιες περιπτώσεις είναι τα ουσιαστικά τύπου *καρδιά* και *ψυχή* που στην αρχαιότητα αναλύονταν ως *καρδι-α* και *ψυχ-η* λόγω διαφορετικών μορφηματικών ορίων και συγχρονικά αναλύονται ως *καρδιά-Ø* και *ψυχή-Ø*. Επίσης αλλαγές στο εσωτερικό θέματος ρημάτων, όπως *πλέν-ω* και *πλύν-αμε*, προήλθαν από την απενεργοποίηση του νόμου της ετεροίωσης (βλ. Οικονόμου 1971: 189). Τα αλλόμορφα δημιουργήθηκαν σε προηγούμενη φάση της γλώσσας στην κλίση λόγω της μετατόπισης ορίων ή της εφαρμογής φωνολογικών κανόνων και με την απενεργοποίησή τους εισήλθαν αυτόματα στο λεξικό, όπου συνδέονται απλά με τους μη παραγωγικούς πλεοναστικούς κανόνες. Είναι εμφανής η άμεση και αλληλένδετη σύνδεση της αλλομορφίας και της κλίσης. Εξάλλου αποτελεί συστηματική τάση στις μορφολογικές προσεγγίσεις του φαινομένου της αλλομορφίας να εξετάζεται στο επίπεδο της κλίσης και να χαρακτηρίζονται τα αλλόμορφα με κλιτική ορολογία από τους μελετητές (Lieber 1982, Carstairs 1987, Ralli 1988, 2000, Booij 1997, Pirrelli & Battista 2000a, 2000b). Καθίσταται πλέον σαφές ότι η αλλομορφία εξαρτάται άμεσα από την κλίση και για αυτό δεν είναι τυχαίο ότι η καταχώρηση των αλλομόρφων στο λεξικό γίνεται με κλιτικούς όρους στο πλαίσιο υποκατηγοριοποίησης των μορφολεξικών κανόνων.

#### 2.4.3.2. Ονοματική αλλομορφία

Η ονοματική κλίση της ΚΝΕ έχει αποτελέσει εντατικό πεδίο έρευνας ήδη από τις αρχές του 20<sup>ου</sup> αιώνα (βλ. Χατζιδάκι 1905, 1907). Οι διάφορες μελέτες έγιναν στο πλαίσιο της παραδοσιακής γραμματικής έχοντας ως βάση τρεις κλιτικές τάξεις και κατένειμαν τα ουσιαστικά με κριτήρια το γένος και τον αριθμό συλλαβών (πρβλ. Τριανταφυλλίδης 1941, Τσοπανάκης 1994), που προσπάθησαν να σπάσουν το πρότυπο των Αρχαίων Ελληνικών με το χωρισμό των τάξεων με βάση κυρίως το γένος. Νεότεροι ερευνητές (Μπαμπινιώτης & Κοντός 1967, Κλαίρης & Μπαμπινιώτης 1996) ομαδοποιούν τις τρεις κλιτικές τάξεις με κριτήριο τις κοινές μορφές των ουσιαστικών ανάλογα με τις πτώσεις. Κριτική ασκήθηκε από τη Ράλλη (Ralli 2000, Ράλλη 2005: 116-118) για τα διάφορα προβλήματα που εμφανίζουν αυτές οι προσεγγίσεις. Σημαντικές ερευνητικές προσπάθειες στην ονοματική κλίση είναι επίσης των Seiler (1958), Ruge (1969), Μαλικούτη (1970), Luraghi (2005) μεταξύ άλλων.

Στην παρούσα διατριβή ακολουθούμε το πρότυπο της Ralli (2000), που περιλαμβάνει οκτώ κλιτικές τάξεις. Η παραμετροποίηση της ονοματικής κλίσης γίνεται με κριτήρια (i.) τη συστηματική αλλομορφική συμπεριφορά των θεμάτων και (ii.) τη

μορφή του συνόλου των κλιτικών επιθημάτων. Η παρούσα μελέτη βασίζεται σε θεωρητικό πλαίσιο, όπου η αλλομορφία αποτελεί βασικό κριτήριο κατηγοριοποίησης, αφού κάτι τέτοιο είναι απαραίτητο για την επιτυχή αντιμετώπιση της αλλομορφίας στην υπολογιστική επεξεργασία (βλ. αναλυτικότερα κεφάλαιο 6°).

Σε προηγούμενη εργασία μας (Καρασίμος 2001) ομαδοποιήσαμε τις οκτώ κλιτικές τάξεις σύμφωνα με τον τύπο της αλλομορφίας, διαχωρίζοντας παράλληλα τα ουσιαστικά με μηδενική αλλομορφία, τα ουσιαστικά με απλή αλλομορφία και τα ουσιαστικά με μεικτή αλλομορφία. Ο όρος απλή αλλομορφία ανταποκρίνεται στα μορφήματα που έχουν ένα αλλόμορφο, ενώ η μεικτή αλλομορφία αναφέρεται σε μορφήματα που έχουν δύο αλλόμορφα.

Ουσιαστικά χωρίς αλλομορφία σύμφωνα με το μοντέλο της Ralli (2000) είναι αυτά της πρώτης (άνθρωπ-ος), πέμπτης (βουν-ό), έκτης (παιδί) και έβδομης κλιτικής τάξης (δάσ-ος). Τα ουσιαστικά με απλή αλλομορφία είναι εκείνα που σχηματίζουν<sup>36</sup> αλλόμορφα είτε προσθέτοντας κάποιο σύμφωνο στο τέλος του θέματος (/ð/ ή /t/) είτε αφαιρώντας το τελικό φωνήεν του θέματος (/a/ ή /i/): τέτοιες κλιτικές τάξεις είναι η δεύτερη (ταμια~ ταμι, παπα~ παπαδ), είναι η τρίτη (ψυχη~ ψυχ, καρδια~ καρδι) και η όγδοη (κυμα~ κυματ). Τέλος, τα ουσιαστικά με μεικτή αλλομορφία είναι αυτά της τέταρτης κλιτικής τάξης (πολη~ πολε~ πολ) που συνδυάζουν και τους δυο προαναφερθέντες τρόπους στη δημιουργία αλλομόρφου.

Στον ακόλουθο πίνακα παραθέτουμε τις οκτώ κλιτικές τάξεις των ουσιαστικών της Κοινής Νεοελληνικής σύμφωνα με το πρότυπο της Ralli (2000) με ελάχιστες τροποποιήσεις. Τα κλιτικά επιθήματα διαχωρίζονται με παύλα (–), ενώ τα αλλόμορφα με την περισπωμένη (~). Επίσης, στην κωδικοποίηση των αλλομορφικών κανόνων το X συμβολίζει είτε όλο το θέμα είτε το θέμα πλην του τελευταίου χαρακτήρα.

### Κλιτική Τάξη 1 (KT1)

Τύπος Θέματος ανθρώπ<sub>ΑΡΣ</sub>

ψηφ<sub>ΘΗΛ</sub>

Γένος Αρσενικό, Θηλυκό

Μορφολεξικός Κανόνας X~ [μηδενική αλλομορφία]

<sup>36</sup> Με τη λέξη «σχηματίζουν» περιγράφουμε τη διαφοροποίηση του βασικού αλλομόρφου από το άλλο. Δεν προϋποθέτουμε κανένα παραγωγικό κανόνα σχηματισμού αλλομόρφων.

**Κλιτική Τάξη 2α (KT2α)**

Τύπος Θέματος ανδρα ~ ανδρ

αλητη ~ αλητ

Γένος Αρσενικό

Μορφολοξικός Κανόνας ΧΦ<sub>[/a/, /i/]</sub> ~ Χ [αφαιρετική αλλομορφία]**Κλιτική Τάξη 2β (KT2β)**

Τύπος Θέματος αγα ~ αγαδ

τεμπελη ~ τεμπεληδ [~ τεμπελ<sup>37</sup>]

καναπε ~ καναπεδ

παππου ~ παππουδ

Γένος Αρσενικό

Μορφολοξικός Κανόνας Χ ~ Χδ [προσθετική αλλομορφία]

**Κλιτική Τάξη 3α (KT3α)**

Τύπος Θέματος καρδια ~ καρδι

ψυχη ~ ψυχ

Γένος Θηλυκό

Μορφολοξικός Κανόνας ΧΦ<sub>[/a/, /i/]</sub> ~ Χ [αφαιρετική αλλομορφία]**Κλιτική Τάξη 3β (KT3β)**

Τύπος Θέματος μαμα ~ μαμαδ

μαΐμου ~ μαΐμουδ

Γένος Θηλυκό

Μορφολοξικός Κανόνας Χ ~ Χδ [προσθετική αλλομορφία]

**Κλιτική Τάξη 4 (KT4)**

Τύπος Θέματος βαση ~ βασε ~ βασ

Γένος Θηλυκό

Μορφολοξικός ΧΦ<sub>[/i/]</sub> ~ ΧΦ<sub>[/ε/]</sub> ~ Χ [συνδυαστική αλλομορφία]**Κλιτική Τάξη 5 (KT5)**<sup>37</sup> Για τη συγκεκριμένη περίπτωση θα αναφερθούμε εκτενέστερα στην ενότητα 2.4.5.2.

Τύπος Θέματος αβγ

Γένος Ουδέτερο

Μορφολοξικός Κανόνας X~ [μηδενική αλλομορφία]

### Κλιτική Τάξη 6 (KT6)

Τύπος Θέματος αυτι [~ αυτ<sup>38</sup>]

Γένος Ουδέτερο

Μορφολοξικός Κανόνας X~ [μηδενική αλλομορφία]

### Κλιτική Τάξη 7 (KT7)

Τύπος Θέματος δασ

Γένος Ουδέτερο

Μορφολοξικός Κανόνας X~ [μηδενική αλλομορφία]

### Κλιτική Τάξη 8α (KT8α)

Τύπος Θέματος βημα ~ βηματ

προϊον ~ προϊόντ

συμβαν ~ συμβαντ

φωνηνεν ~ φωνηνεντ

Γένος Ουδέτερο

Μορφολοξικός Κανόνας X ~ Χτ [προσθετική αλλομορφία]

### Κλιτική Τάξη 8β (KT8β)

Τύπος Θέματος τερας ~ τερατ

κλει-σιμο ~ κλει-σιματ

Γένος Ουδέτερο

Μορφολοξικός Κανόνας XΣ<sub>[/s/]</sub> ~ XΣ<sub>[/t/]</sub> [αντικαταστατική αλλομορφία]

XΣ<sub>[/o/]</sub> ~ XΣ<sub>[/at/]</sub> [αντικαταστατική αλλομορφία]

### Κλιτική Τάξη 2.4

{συνδυαστική κλιτική τάξη}

### (KT2.4)

Τύπος Θέματος βασιλεα ~ βασιλε ~ βασιλ

<sup>38</sup> Για τη συγκεκριμένη περίπτωση αλλομόρφου θα αναφερθούμε στην ενότητα 2.4.5.3.

## Γένος Αρσενικό

Αλλομορφικός τύπος  $X\Phi_{[e]} \Phi_{[a]} \sim X\Phi_{[e]} \sim X$  [συνδυαστική αλλομορφία]

Πίνακας 8: Οι ονοματικές κλιτικές τάξεις με τις αλλομορφικές πληροφορίες.

Πιο συγκεκριμένα οι τροποποιήσεις που προσθέτουμε στο μοντέλο είναι οι ακόλουθες: αρχικά χωρίζουμε την KT2 σε δύο υπο-τάξεις, αφού εντοπίζεται διαφορετικός τύπος αλλομορφίας. Η KT2α συμπεριλαμβάνει ουσιαστικά με αφαιρετική αλλομορφία που προέκυψε από την επανάλυση των ορίων μεταξύ θέματος και κλιτικού επιθήματος, για παράδειγμα *ταμί-ας*  $\Rightarrow$  *ταμία-ς*, *ναύτ-ης*  $\Rightarrow$  *ναύτη-ς*. Αντιθέτως η KT2β, όπως επισημάναμε παλαιότερα (Καρασίμος 2001), συμπεριλαμβάνει ουσιαστικά ξενικής κυρίως προέλευσης που ανέπτυξαν ένα /ð/ για την αποφυγή χασμωδίας στον πληθυντικό που χαρακτηρίζεται ως παρέκταση του θέματος (αναλυτικότερα βλ. Μαλικούτη 1970: 43, Θωμαδάκη 1994). Ακριβώς ίδια είναι η αιτιολόγηση της αλλομορφίας των ουσιαστικών της KT3 και ορθώς μπορεί να χαρακτηριστεί ως δίδυμη της KT2, για αυτό και ο διπλός χωρισμός της. Επιπροσθέτως, η KT8 σε αντίθεση με τις υπόλοιπες κλιτικές τάξεις των ουδετέρων παρουσιάζει ιδιόζουσα αλλομορφία. Τα ουσιαστικά της όγδοης κλιτικής τάξης σχηματίζουν αλλόμορφα με την προσθήκη ενός /t/ στο θέμα της ονομαστικής. Η αντίθεση αυτής της κλιτικής τάξης σε σύγκριση με τα ουσιαστικά της δεύτερης και τρίτης κλιτικής τάξης είναι ότι το /t/ εμφανίζεται και στη γενική ενικού (*κύματ-ος*, *ονόματ-ος*). Να σημειωθεί πως σύμφωνα με τους Οικονόμου (1971) και Σετάτο (1973: 22) το τελικό /t/ υπήρχε σε όλες τις πτώσεις της AE/ KNE και δεν πραγματωνόταν, όταν δεν ακολουθούσαν από κλιτικό επίθημα. Διαφορετική προσέγγιση για τη συγκεκριμένη KT ακολουθούν οι Κλαίρης & Μπαμπινιώτης (1996) που θεωρούν ως αλλόμορφα τα *κυμ~ κυματ*. Παράλληλα υπάρχει και μία μικρή ομάδα ουσιαστικών (KT8β) που έχουν θέμα που λήγει σε -ας, ενώ διατηρεί το αλλόμορφο σε -τ- από τα AE (Οικονόμου 1971, Chantraine 1998, Sihler 2009).

Τέλος, να αναφερθεί πως μία μικρή ομάδα θεμάτων περιέχει το εγγενές χαρακτηριστικό [+λόγιο] που συμβάλλει σημαντικά στη διαμόρφωση της κλίσης τους. Συνήθως ανήκουν σε μη παραγωγικές κλιτικές τάξεις των Νέων Ελληνικών. Σε αυτή την κατηγορία ανήκει η KT4 που υπόκειται σε τάσεις αναδιοργάνωσης του παραδείγματός με αποτέλεσμα τη δημιουργία τριών αλλομόρφων τύπου *βαση ~ βασε ~ βασ*. Ακόμα πιο πολύπλοκη είναι η αναδιοργάνωση των παραδειγμάτων τύπου *βασιλέας (βασιλιάς)*.

- (45.) α. αλιέας (αλιεα ~ αλιε ~ αλι)  
 βασιλέας (βασιλεα ~ βασιλε ~ βασιλ)  
 γονέας (γονεα ~ γονε ~ γον)  
 β. βαφέας (βαφεα ~ βαφε ~ βαφ)  
 ιερέας (ιερεα ~ ιερε ~ ιερ)  
 κουρέας (κουρεα ~ κουρε ~ κουρ)  
 συγγραφέας (συγγραφεα ~ συγγραφε ~ συγγραφ)  
 σφαγέας (σφαγεα ~ σφαγε ~ σφαγ)

Τα παραπάνω αποτελούν κατάλοιπα των αρχαίων ουσιαστικών μονοθέματων καταληκτικών σε –ευ(ς) (45.α) και των παραγωγικών ουσιαστικών σε –εα(ς) (45.β, Οικονόμου 1971: 63-64, Sihler 2009). Τα συγκεκριμένα έχουν ήδη επαναναλυθεί αλλομορφικά, με αποτέλεσμα να έχουν πλέον τρία αλλόμορφα, όπως ακριβώς και τα θηλυκά ουσιαστικά της ΚΤ4. Σύμφωνα με τη Ράλλη (2005) αποτελούν περιπτώσεις λεξικοποίησης τα παράγωγα ουσιαστικά σε –εας. Αποτελούν μία αρκετά μεγάλη ομάδα και πρέπει να ενταχθούν κλιτικά σε κάποια τάξη. Για την ανάλυση των συγκεκριμένων λέξεων υπάρχουν δύο πιθανές λύσεις: (α.) να δημιουργήσουμε μία καινούργια κλιτική τάξη, και (β.) να θεωρήσουμε ότι είναι στο στάδιο λεξικοποίησης και αναδιάρθρωσης του κλιτικού παραδείγματος εντάσσοντας αυτές τις λέξεις σε άλλα κλιτικά παραδείγματα. Επιλέγουμε τη δεύτερη λύση, γιατί την κρίνουμε ως επαρκέστερη, αφού μας παρέχεται η δυνατότητα να ενταχθούν σε δύο υπάρχουσες κλιτικές τάξεις. Θεωρούμε πως ο ενικός αριθμός των κλιτών τύπων σχηματίζεται σύμφωνα με το πρότυπο της ΚΤ2 και ο πληθυντικός αριθμός με αντίστοιχο πρότυπο της ΚΤ4. Ευκολότερη υπόθεση είναι και τα παράγωγα ουσιαστικά σε –σιμο τα οποία τα εντάσσουμε στην υποκατηγορία της ΚΤ8(β), αφού έχουν παρόμοια αλλομορφική συμπεριφορά με εκείνη των λέξεων τύπου *τερας*~ *τερατ* και το ίδιο κλιτικό παράδειγμα.

### 2.4.3.3. Ρηματική αλλομορφία

Η Ελληνική ρηματική κλίση παραμένει στο επίκεντρο της γλωσσολογικής έρευνας εδώ και μισό αιώνα υπό το πρίσμα διαφορετικών προσεγγίσεων. Στο πλαίσιο της παραδοσιακής γραμματικής οι Τριανταφυλλίδης (1941), Τσοπανάκης (1994) προσέγγισαν το ρηματικό κλιτικό σύστημα καθαρά περιγραφικά κατηγοριοποιώντας τα ρήματα με φωνολογικά κριτήρια, όπως ο ληκτικός χαρακτήρας του θέματος, ο τόνος

και ο χαρακτηρισμός ως πρώην συνηρημένα. Ακολούθως από την παράδοση της γενετικής ανάλυσης με την Philippaki-Warburton (1970, 1973, 1991) και την προσέγγιση του Μπαμπινιώτη (1972) καταλήξαμε στη θεωρία του ισχυρού λεξικαλισμού με τη Ράλλη (Ralli 1988, Ράλλη 1993, 2005).

Το μοντέλο της Ράλλη για τα ρήματα της ΚΝΕ περιλαμβάνει δύο βασικές κλιτικές τάξεις, οι οποίες είναι προβλέψιμες στα ρήματα, σε αντίθεση με τα ουσιαστικά και για αυτό δεν αποτελεί εγγενές χαρακτηριστικό καταχώρησης. Η ένταξη κάθε ρήματος στην αντίστοιχη κλιτική τάξη κρίνεται αποκλειστικά από την παρουσία ή απουσία **συστηματικής αλλομορφίας**. Όπως υποστηρίζει η Ralli (1988, 2005) τα ρήματα που εμφανίζουν συστηματική αλλομορφία είναι κυρίως αυτά που ανήκαν στην ομάδα των πρώην συνηρημένων με αλλομορφικό τύπο  $X(a) \sim X\Phi_{[a/, /e]}$  και αποτελούν τη δεύτερη κλιτική τάξη (PKT2), ενώ τα ρήματα που δεν εμφανίζουν τέτοιου είδους αλλομορφία ανήκουν στην πρώτη κλιτική τάξη (PKT1).

(46.)	α. ανοίγ-ω	ανοίξ-ω (ανοιγ-σω)
	β. πλέκ-ω	πλέξ-ω (πλεκ-σω)
	γ. αρπάξ-ω	αρπάξ-ω (αρπαγ-σω)
	δ. αγαπ(ά)-ω	αγαπή-σω
	ε. απαντ(ά)-ω	απαντή-σω
	στ. χωρ(ά)-ω	χωρέ-σω

Ο παραπάνω αλλομορφικός κανόνας είναι πλεοναστικός και όλα τα θέματα που δεν περιγράφονται από τον κανόνα και δεν εμφανίζουν συστηματική αλλομορφία ανήκουν αυτόματα στην πρώτη κλιτική τάξη (46.α-γ), ενώ τα ρηματικά θέματα που εμφανίζουν τη συγκεκριμένη αλλομορφική συμπεριφορά (46.δ-στ) αποκτούν την τιμή της δεύτερης κλιτικής τάξης. Έτσι διασφαλίζεται η γραμματικότητα του κλιτικού συστήματος, αφού τα κλιτικά επιθήματα συνδέονται μόνο με θέματα που έχουν τον κατάλληλο κλιτικό δείκτη. Ο κλιτικός δείκτης είναι ένας δείκτης αναφοράς που πρέπει να ταυτίζεται με τον αντίστοιχο των κλιτικών επιθημάτων για την αποφυγή των αντιγραμματικών λέξεων. Για παράδειγμα, το εναλλακτικό κλιτικό επίθημα παρατατικού -γα δεν μπορεί να συνδυαστεί με ρηματικά θέματα της ΚΤ1, όπως δένω  $\Rightarrow$  \*δέναγα, γράφω  $\Rightarrow$  \*γράφαγα, κάνω  $\Rightarrow$  \*κάναγα.

Όπως έχουμε αναφέρει ήδη, σύμφωνα με τη Ράλλη (2005) τα βασικά κριτήρια διαχωρισμού και ταξινόμησης των ρημάτων είναι η παρουσία ή η απουσία συστηματικής αλλομορφίας  $X(a) \sim X\{[i/e]\}$ , καθώς και τα χαρακτηριστικά της



αλλομορφικής συμπεριφοράς που παρουσιάζουν τα ρηματικά θέματα. Αντίθετα, προτείνουμε τις ακόλουθες κατηγορίες με τα συγκεκριμένα χαρακτηριστικά προχωρώντας σε περαιτέρω κατηγοριοποίηση και επεκτείνοντας την ταξινόμηση της Ράλλη (2005: 138). Η συγκεκριμένη κατηγοριοποίηση δεν ανταποκρίνεται σε διαφορετικές κλίσεις και συνδυασμούς με κλιτικά επιθήματα, αλλά μία προσπάθεια ευρύτερης και πιο περιγραφικής ανάλυσης της ρηματικής αλλομορφίας:

(47.) Κατηγορία I: Μηδενική αλλομορφία

π.χ. διώκ-ω, μπλέκ-ω

Κατηγορία IIa: Συστηματική αλλομορφία I

π.χ. αγαπ-ώ ~ αγάπη-σα, τραγουδ-ώ ~ τραγούδη-σα

Κατηγορία IIb: Συστηματική αλλομορφία II

π.χ. διαιρ-ώ ~ διαίρε-σα, τελ-ώ ~ τέλε-σα

Κατηγορία IIIa: Μη συστηματική απλή αλλομορφία

π.χ. φεύγ-ω ~ φύγ-ω,

Κατηγορία IIIb: Μη συστηματική πολλαπλή αλλομορφία

π.χ. παίρν-ω ~ πάρ-ω ~ πήρ-α, καλ-ώ ~ καλε-σα ~ κληθ-ώ

Κατηγορία IV: Υποκατάσταση

π.χ. τρώ-ω ~ έ-φαγ-α, λέ-ω ~ είπ-α

Η δεύτερη κατηγορία αποτελεί την KT2 της Ράλλη (2005), ενώ οι υπόλοιπες την KT1. Σε προηγούμενη μελέτη μας (Καρασίμος 2001) προσπαθήσαμε να συνδυάσουμε τα κριτήρια της αλλομορφίας με την τάση κατηγοριοποίησης της παραδοσιακής γραμματικής που βασίζεται στους ληκτικούς χαρακτήρες του θέματος και τα κλιτικά επιθήματα με τα οποία συνδυάζονται. Η σύζευξη της ρηματικής αλλομορφίας και της παραδοσιακής γραμματικής έδωσε πιο ομογενοποιημένες τάξεις με ακριβείς αλλομορφικούς πλεοναστικούς κανόνες, εντούτοις δημιούργησε αρκετές κλιτικές τάξεις ή υποτάξεις με αποτέλεσμα να χαρακτηρίζεται από αντιοικονομικότητα (βλ. Καρασίμος 2001).<sup>39</sup> Αντιοικονομικό κρίνεται και το σύστημα της παραδοσιακής

<sup>39</sup> Είναι εμφανές ότι ο διαχωρισμός των ρηματικών κλιτικών τάξεων χρήζει περαιτέρω έρευνας και ορθότερης ταξινόμησης, διότι η αλλομορφία της KT1 παρουσιάζει πολλές αλλαγές ρηματικών θεμάτων· ίσως η προσέγγιση των Pirrelli & Battista (2000a, 2000b) με την **αλλομορφική ταξινόμηση** ('indexing') και τα **κλιτικά κελιά** (slots) με τη **συνγχώνευση** (merging) να είναι υλοποιήσιμη και οικονομική για τα Ελληνικά.

γραμματικής που περιλαμβάνει 12 κλιτικές τάξεις με πολλές υποκατηγορίες στα ρήματα (Τριανταφυλλίδης 1941).

### Κλιτική Τάξη 1 (KT1α)

Ενδεικτικός Τύπος Ρήματος	μπλέκω (μπλεκ~)
Αλλομορφικός Τύπος	Μηδενική αλλομορφία
Μορφολεξικός Κανόνας	X

### Κλιτική Τάξη 1β (KT1β)

Ενδεικτικός Τύπος Ρήματος	χάνω (χαν~ χα)
Αλλομορφικός Τύπος	Μη συστηματική απλή αλλομορφία I
Μορφολεξικός Κανόνας	{διάφοροι κανόνες}

### Κλιτική Τάξη 1γ (KT1γ)

Ενδεικτικός Τύπος Ρήματος	παίρνω (παιρν~ παρ~ πηρ)
Αλλομορφικός Τύπος	Μη συστηματική πολλαπλή αλλομορφία II
Μορφολεξικός Κανόνας	{διάφοροι κανόνες}

### Κλιτική Τάξη 1δ (KT1δ)

Ενδεικτικός Τύπος Ρήματος	είμαι (~υπηρχ)
Αλλομορφικός Τύπος	Υποκατάσταση
Μορφολεξικός Κανόνας	X ~ Y

### Κλιτική Τάξη 2α (KT2α)

Ενδεικτικός Τύπος Ρήματος	αγαπώ (αγαπ(α)~ αγαπη)
Αλλομορφικός Τύπος	Συστηματική αλλομορφία I
Μορφολεξικός Κανόνας	X(Φ <sub>[a]</sub> ) ~ XΦ <sub>[i]</sub>

### Κλιτική Τάξη 2β (KT2β)

Ενδεικτικός Τύπος Ρήματος	διαιρώ (διαιρ~ διαιρε)
Αλλομορφικός Τύπος	Συστηματική αλλομορφία II
Μορφολεξικός Κανόνας	X ~ XΦ <sub>[ε]</sub>

Πίνακας 9: *Οι κλιτικές τάξεις του ρηματικού συστήματος κατά τη Ράλλη (2005) και τη δική μας τροποποίηση.*

Στον παραπάνω πίνακα παραθέτουμε την ομαδοποίηση των ρημάτων με συνοπτικό τρόπο βασισμένοι στο μοντέλο της Ράλλη και τις απαραίτητες τροποποιήσεις λόγω της επιμέρους αλλομορφικής κατηγοριοποίησης σύμφωνα με την πρότασή μας. Για την Κλιτική Τάξη 1 προτείνουμε τέσσερις υποτάξεις, για τις οποίες δεν υπάρχει συγκεκριμένος αλλομορφικός κανόνας, αλλά μερικώς διαχωρισμός αφού δεν υπάρχει συγκεκριμένο είδος αλλομορφίας (π.χ. προσθετική, αφαιρετική) παρά μόνο τύπος (π.χ. μη συστηματική, απλή και πολλαπλή). Να σημειωθεί ότι *δεν αποτελεί* στόχο της παρούσας έρευνας να διευθετήσει το ζήτημα των ρηματικών κλιτικών τάξεων, αλλά να παρουσιάσει μία γενική εικόνα της αλλομορφίας που διέπει το ρηματικό κλιτικό σύστημα.

#### 2.4.3.4. Επιθετική (και Επιρρηματική) αλλομορφία

Όπως επισημαίνει η Ράλλη (2005) τα επίθετα της ΚΝΕ ακολουθούν τα κλιτικά παραδείγματα των αντίστοιχων ουσιαστικών και εντάσσονται στις αντίστοιχες κλιτικές τάξεις. Αυτό ισχύει σε γενικές γραμμές για επίθετα, όπως *αγαθός, αγαθή, αγαθό* και *νέος, νέα, νέο*. Πιο συγκεκριμένα, τα αρσενικά επίθετα κλίνονται με βάση τις καταλήξεις της ΚΤ1 (*αγαθός/ νέος*), τα θηλυκά επίθετα εφαρμόζουν το κλιτικό παράδειγμα της ΚΤ3 (*αγαθή/ νέα*) και, τέλος, τα ουδέτερα επίθετα κλίνονται σύμφωνα με το πρότυπο της ΚΤ5 (*αγαθό/ νέο*). Έτσι, προσαρμόζοντας τα παραδείγματά τους στα ουσιαστικά μόνο τα θηλυκά επίθετα εμφανίζουν αλλομορφο, με τον αλλομορφικό κανόνα  $X\Phi_{[a/i]} \sim X$  να περιγράφει τον τύπο αλλαγής.

Εντούτοις, υπάρχουν δυο ομάδες επιθέτων, για τις οποίες κρίνεται απαραίτητη η εισαγωγή νέων κλιτικών τάξεων (*ταχύς, ταχεία, ταχύ* και *διεθνής, διεθνές*). Πρόκειται για ομάδες επιθέτων που χαρακτηρίζονται από το χαρακτηριστικό [+λόγιο] με αποτέλεσμα να ανήκουν σε μία περιφερειακή κλίση της γλώσσας (βλ. Καρασίμος 2001, Αναστασιάδη-Συμεωνίδη & Χειλά-Μαρκοπούλου 2003, Ράλλη 2005). Για αυτό το λόγο και παρατηρείται υψηλό ποσοστό λαθών στη χρήση τους, συστηματική έλλειψη τύπων σε κάποια κλιτικά παραδείγματα, καθώς και ύπαρξη εναλλακτικών κλιτικών επιθημάτων, αφού το σύστημά τους είναι σε κατάσταση αναδιάρθρωσης. Όπως ορθώς παρατηρεί η Ράλλη (2005: 122) τα επίθετα τύπου *διεθνής, διεθνές* (τα επονομαζόμενα τριγενή και δικατάληκτα), αν και περιφερειακά, αποτελούν μεγάλη ομάδα και

αντιστέκονται στις τάσεις αναδιάρθρωσης. Στον αντίποδα, τα επίθετα τύπου *βαθός*, *βαθιά*, *βαθύ* αποτελούν μικρή ομάδα και αλλάζουν το κλιτικό τους παράδειγμα. Θεωρούμε ότι αν και η ομάδα αυτή είναι μικρή και κλειστή, περιλαμβάνει μερικά βασικά και συχνότατα επίθετα, που μάλιστα δεν έχουν και συνώνυμα (*βαθός*, *βαρύς*, *ελαφρύς*, *μακρύς*, *πλατύς*).

Στον πίνακα που ακολουθεί παρατίθενται τα νέα κλιτικά πρότυπα που προστέθηκαν για να καλύψουμε τα κενά. Οι ΚΤ9 και ΚΤ10 έχουν προταθεί από την Ράλλη (2000, 2005), ενώ εισάγουμε (Καρασίμος 2001) μια νέα κλιτική τάξη, την ΚΤ11 ως πρότυπο για τα επίθετα σε *-υς* (το θηλυκό εφαρμόζει το πρότυπο της ΚΤ3, ενώ το ουδέτερο ακολουθεί το πρότυπο της ΚΤ6<sup>40</sup>).

### Κλιτική Τάξη 9 (ΚΤ9)

Τύπος Θέματος διεθ<sub>ΝΑΡΣ, ΘΗΛ</sub>

Γένος Αρσενικό, Θηλυκό

Αλλομορφικός κανόνας X~ [μηδενική αλλομορφία]

### Κλιτική Τάξη 10 (ΚΤ10)

Τύπος Θέματος διεθ<sub>ΝΟΥΔ</sub>

Γένος Ουδέτερο

Αλλομορφικός κανόνας X~ [μηδενική αλλομορφία]

### Κλιτική Τάξη 11 (ΚΤ11)

Τύπος Θέματος βαρ<sub>υ</sub> ~ βαρ<sub>ε</sub> ~ βαρ

Γένος Αρσενικό

Αλλομορφικός Κανόνας XΦ<sub>[/ω]</sub> ~ XΦ<sub>[/ε]</sub> ~ X

Πίνακας 10: Οι συμπληρωματικές κλιτικές τάξεις των επιθέτων της ΚΝΕ.

Κλείνοντας να σημειώσουμε πως η ΚΤ11 έχοντας λιγότερες αντιστάσεις στις τάσεις αναδιάρθρωσης εμφανίζει τύπους σε ορισμένα επίθετα, όπως σε *-ιοι* (ον.πλ., αιτ.πλ., κλ.πλ.) και *-ίων* (γεν.πλ.). Η συγκεκριμένη κλιτική τάξη εμφανίζει πολλές

<sup>40</sup> Τα ουδέτερα επίθετα σε *-υ* ακολουθούν το πρότυπο της ΚΤ6, η οποία συμπεριλαμβάνει εκτός των ουδέτερων ουσιαστικών σε *-ι*, τα ουδέτερα ουσιαστικά σε *-υ*, όπως *δίχτυ*, *δάκρυ*, *στάχυ*, *δόρυ*, *βράδυ*. Να σημειωθεί πως μόνο σε κάποια κατάλοιπα εμφανίζεται στη γενική πληθυντικού των επιθέτων τύπος σε *-έων* αντί *-ιών* που θεωρούμε πλέον στην ΚΝΕ ότι είναι αναλογικός σχηματισμός από τον αντίστοιχο της ΚΤ11 των αρσενικών.

ομοιότητες με τα ουσιαστικά της συνδυαστικής κατηγορίας ΚΤ2.4 που περιλαμβάνει τα ουσιαστικά τύπου *γονέας*. Επομένως μια εφικτή λύση είναι να ενταχθούν τα συγκεκριμένα επίθετα σε αυτή την κατηγορία. Εντούτοις θεωρούμε ότι συγχρονικά οφείλουμε να τα εισάγουμε σε μια ανεξάρτητη κλιτική τάξη, γιατί έχουν κάποιους τύπους που παρεκκλίνουν από το πρότυπο της συνδυαστικής κατηγορίας. Είναι πιθανό λόγω αναδιοργάνωσης του κλιτικού συστήματος σε μελλοντικό στάδιο της γλώσσας να καταργηθεί η ΚΤ11 και τα συγκεκριμένα επίθετα να ενσωματωθούν στην ΚΤ2.4.

#### 2.4.4. Η αρχή της αλλομορφικής συμπεριφοράς

Κατά το σχηματισμό του κλιτικού παραδείγματος των ουσιαστικών, επιθέτων και ρημάτων πολύ συχνά τα μορφήματα συμμετέχουν με παραπάνω από μία μορφές, δηλαδή έχουν αλλόμορφα. Όπως έχουμε αναφερθεί ήδη, θεωρούμε πως τα αλλόμορφα «*πρωτοχρησιμοποιούνται*» κατά τη διαδικασία της κλίσης, όντας η πιο απλή διαδικασία σχηματισμού λέξεων. Η επανάλυση των ορίων μεταξύ των μορφημάτων – κυρίως ανάμεσα σε θεματικά μορφήματα και κλιτικά επιθήματα – και οι ανενεργοί φωνολογικοί κανόνες, όπως παρουσιάσαμε στην ενότητα 2.1.1, δημιουργήσαν αλλαγές σε μορφήματα που συγχρονικά θεωρούνται αλλόμορφα. Έτσι παρατηρούμε στα Ελληνικά, ότι αν ένα μόρφημα έχει αλλομορφική συμπεριφορά, θα εμφανίσει αλλόμορφα στην κλίση και θα χρησιμοποιήσει ένα από αλλόμορφά του στις διαδικασίες της παραγωγής και της σύνθεσης. Αντιθέτως, εάν ένα μόρφημα δεν εμφανίσει αλλομορφία στην κλίση, τότε ούτε στις άλλες δύο διαδικασίες σχηματισμού λέξεων θα εμφανίσει κάποιο αλλόμορφο. Την παραπάνω παρατήρηση ονομάζουμε ως **αρχή της αλλομορφικής συμπεριφοράς**. Όπως θα δούμε παρακάτω, στα Ελληνικά τα περισσότερα μορφήματα υπόκεινται σε αυτήν την αρχή, κυρίως τα επίθετα και τα ρήματα, όπως και τα ουσιαστικά με κάποιες ελάχιστες, συστηματικές και δικαιολογημένες παραβιάσεις της αρχής.

#### (48.) Αρχή της Αλλομορφικής Συμπεριφοράς

*Τα μορφήματα εμφανίζουν αλλομορφική συμπεριφορά ή όχι στις διαδικασίες σχηματισμού λέξεων ανάλογα με τη συμπεριφορά τους στην κλίση.*

Έτσι σχεδόν όλα τα αλλόμορφα, σύμφωνα με την αρχή, έχουν κοινή αλλομορφική συμπεριφορά σε όλες της διαδικασίες σχηματισμού λέξεων και επομένως μία τέτοια συμπεριφορά δεν είναι συμπτωματική και τυχαία. Επιπλέον, θεωρούμε πως η *αρχή της*

αλλομορφικής συμπεριφοράς διασφαλίζει την ομοιομορφία μεταξύ των διαφόρων σχηματισμών λέξεων (49.) και συμβάλλει στο να διέπεται η γλώσσα από την αρχή της οικονομίας λόγω της χρήσης των ίδιων αλλομόρφων σε όλες τις διαδικασίες σχηματισμού λέξεων. Παράλληλα συντελεί στην προβλεψιμότητα της αλλομορφίας, αφού μπορούμε να προβλέψουμε την απουσία ή παρουσία αλλομορφίας από τη συμπεριφορά κάθε μορφήματος στην κλίση. Πιστεύουμε πως η ύπαρξη της συγκεκριμένης αρχής δικαιολογείται άμεσα από την πλειονότητα των μορφημάτων της Ελληνικής γλώσσας, τόσο των θεμάτων/ ριζών, όσο και των παραγωγικών επιθημάτων. Όπως έχει ήδη επισημανθεί, τα ουσιαστικά, επίθετα και ρήματα της Ελληνικής δεν εμφανίζουν καινούργια αλλόμορφα στην παραγωγή πλην ελαχίστων εξαιρέσεων, αλλά χρησιμοποιούν τα ήδη υπάρχοντα (αυτά της κλίσης), εφόσον αυτά υπάρχουν. Παράλληλα τα παραγωγικά επιθήματα συμβαδίζουν με τα θέματα στη μη εμφάνιση νέων αλλομόρφων, όταν ακολουθούνται από παραγωγικά επιθήματα:

(49.) α. [ουσιαστικά]

άνθρωπ(ος)	άνθρωπ(οι)	ανθρώπ-ιν(ος)	ανθρωπ-άκι
ψυχή	ψυχ(ές)	ψυχ-ούλα	ψυχ-ικ(ός)
κύμα	κύματ(α)	κυματ-άρα	κυματ-ίζ(ω)

β. [επίθετα]

άσπρ(ος)	άσπρ(οι)	ασπρ-ούλη(ς)	ασπρ-άδι
διεθν(ής)	διεθν(είς)	διεθν-ικ(ός)	διεθν-ιστή(ς)
βαθ(ύς)	βαθ(είς)	βαθέ-ος	βαθ-ύτερ(ος) <sup>41</sup>

γ. [ρήματα]

δέν(ω)	δέ(σ-αμε)	δε-μέν(ος)	δε-τήρα(ς)
αγαπ(ώ)	αγάπη(σα)	αγαπη-τ-ικ(ός)	αγαπη-μέν(ος)
αφαιρ(ώ)	αφαίρε(σα)	αφαίρε-ση	αφαιρέ-τη(ς)

Παρατηρούμε στα παραπάνω παραδείγματα, ότι τα διάφορα μορφήματα κατά τη διαδικασία της παραγωγής χρησιμοποιούν αλλόμορφα ή τη μοναδική μορφή που έχουν στην κλίση. Επισημαίνεται πως δεν χρησιμοποιούν κάποια καινούργια μορφή, καθώς συνδυάζονται με κάποιο παραγωγικό επίθημα. Επιπροσθέτως, είναι αυτονόητο πως

<sup>41</sup> Θεωρούμε ότι το –υ– είναι μέρος του συγκριτικού επιθήματος και όχι του θέματος, μίας και το επίθημα –ύτερ(ος) εμφανίζεται και σε επίθετα που δεν ανήκουν στην ΚΤ11, όπως *καλός – καλύτερος, πρώτος – πρωτότερος, τρανός – τρανύτερος* (βλ. Ralli 1988, Καρασίμος 2001).

στην προθηματοποίηση ισχύει η ίδια αρχή, ειδικά αφού όλες οι αλλομορφικές αλλαγές συμβαίνουν σχεδόν πάντα<sup>42</sup> στο τέλος του μορφήματος (π.χ. *κυμα~ κυματ*) και ελάχιστες φορές στο μέσον (π.χ. *παιρν~ πηρ~ παρ*). Υποθέτουμε ότι η αλλομορφική συμπεριφορά των μορφημάτων εξαρτάται από την κλίση, γιατί οι μεγαλύτερες αλλαγές έγιναν διαχρονικά στη συγκεκριμένη διαδικασία (βλ. υποεν. 2.4.3). Για τα ουσιαστικά των ΚΤ2α και ΚΤ3α άλλαξαν τα όρια μεταξύ θεμάτων και κλιτικών επιθημάτων σε σύγκριση με παλαιότερη περίοδο της γλώσσας, όπως *καρδιά* από *καρδί-α* > *καρδιά-Ø* (με αλλόμορφο στον πληθυντικό το *καρδι-ες*) και *ταμία-ς* από *ταμί-ας* > *ταμία-ς* (με αλλόμορφο στον πληθυντικό το *ταμί-ες*). Αυτές οι αλλαγές των ορίων επηρεάζουν και τις διαδικασίες παραγωγής και σύνθεσης, με αποτέλεσμα την εμφάνιση θεματικών αλλομόρφων.

Εντούτοις, όπως θα δούμε στις επόμενες υποενότητες (2.4.4.1 και 2.4.4.2) υπάρχουν δύο συστηματικές παραβιάσεις της ανωτέρω αρχής, αυτή των ουσιαστικών τύπου *παιδί* και *μπακάλης*. Τα ουσιαστικά τύπου *παιδί* (ΚΤ6), ενώ δεν παρουσιάζουν αλλομορφία στην κλίση, εμφανίζουν αλλόμορφο στην παραγωγή και στη σύνθεση. Αντίθετα τα ουσιαστικά τύπου *μπακάλης* (ΚΤ2β), παρόλο που έχουν αλλόμορφο στην κλίση, εμφανίζουν διαφορετικό αλλόμορφο σε παραγωγή και σύνθεση.

Η αρχή της αλλομορφικής συμπεριφοράς εφαρμόζεται σε όλες τις γραμματικές κατηγορίες. Υποστηρίζουμε πως η κλίση αποτελεί τον κεντρικό άξονα δοκιμής της αλλομορφικής συμπεριφοράς, γιατί όπως έχει αναφερθεί και προηγουμένως η πλειονότητα των αλλαγών στα μορφήματα που συγχρονικά προκύπτουν αυθαίρετα και απρόβλεπτα συντελέστηκαν στη διαδικασία της κλίσης. Εντούτοις, μπορεί κάποιος να ισχυριστεί πως η αλλομορφία δεν έχει για απαρχή τη διαδικασία της κλίσης. Αν τα επιχειρήματα για την αλληλένδετη σχέση αλλομορφίας και κλίσης δεν είναι ισχυρά και αποδεκτά, αυτό δεν καταργεί την αρχή της αλλομορφικής συμπεριφοράς. Για αυτό και αποδίδουμε την αρχή και με εναλλακτικό ορισμό.

---

<sup>42</sup> Σε ελάχιστα ρήματα παραμένουν αλλομορφικοί τύποι που δημιουργήθηκαν από τη χρονική αύξηση σε αρκτικό φωνηεντικό θέμα (π.χ. *εχ~ ειχ*) και από τον ενεστωτικό αναδιπλασιασμό (πειθ(ω) – πεπεισ-μέν(ος)). Όπως συμβαίνει σε όλες τις αλλομορφικές περιπτώσεις, τα παραπάνω μορφοφονολογικά φαινόμενα είναι ανενεργά στην ΚΝΕ.

**(50.) Αρχή της Αλλομορφικής Συμπεριφοράς (επανορισμός)**

*Τα μορφήματα εμφανίζουν την ίδια αλλομορφική συμπεριφορά, δηλαδή παρουσία ή απουσία αλλομορφίας σε όλες τις διαδικασίες σχηματισμού λέξεων χωρίς κάποια εξάρτηση από κάποια διαδικασία.*

Στον επανορισμό (50.) της αρχής της αλλομορφικής συμπεριφοράς αφαιρούμε την «εξάρτηση» της αλλομορφικής συμπεριφοράς από τη διαδικασία της κλίσης και τη συγκριτική συσχέτισή της με τις υπόλοιπες διαδικασίες. Στόχος του καινούργιου ορισμού είναι να παρουσιάσουμε μία αρχή της μορφολογίας στα Ελληνικά και να δοκιμάσουμε την αρχή σε κλίση, παραγωγή και σύνθεση συγκριτικά. Δεχόμαστε ότι η αρχή είναι καθολική σε όλες τις διαδικασίες σχηματισμού λέξεων χωρίς καμία εξάρτηση από κάποια συγκεκριμένη διαδικασία. Σύμφωνα με την αρχή, κάθε μόρφημα εμφανίζει την ίδια αλλομορφική συμπεριφορά σε κλίση, παραγωγή και σύνθεση. Στα Ελληνικά η παρουσία ή απουσία αλλομορφίας κάποιου μορφήματος είναι συστηματική σε κάθε συμμετοχή του. Για παράδειγμα η λέξη ΠΑΠΑΣ εμφανίζει κάποιο αλλόμορφο σε όλες τις διαδικασίες, όπως *παπά-ς/ παπάδ-ες* στην κλίση, *παπαδ-άκι* στην παραγωγή και *παπαδ-ο-παιδί* στη σύνθεση.

**2.4.4.1. Η εξαιρέση των ουσιαστικών τύπου –παιδί**

Σύμφωνα με τη Ralli (2000) η ΚΤ6 περιλαμβάνει ουδέτερα ουσιαστικού τύπου *παιδί* ( $X\Phi_{/i}$ ), τα οποία δεν έχουν αλλόμορφο στο κλιτικό τους παράδειγμα. Εντούτοις, εμφανίζουν αλλόμορφο στις άλλες διαδικασίες σχηματισμού λέξεων με την αποβολή του τελικού φωνήεντος ( $X\Phi_{/i} \sim X$ ).

(51.)	παιδί	παιδι(ά)	παιδ-άκι	παίδ-αρ(ος)
	μέλι	μέλι(α)	μελ-άκι	μελ-ίσι(ος)
	ραβδί	ραβδι(ά)	ραβδ-άκι	ραβδ-άρα
	γουδί	γουδι(ά)	γουδ-άκι	γουδ-άρα

Για τον Τσοπανάκη (1994: 256-257) τα ουδέτερα της ΚΤ6 έχουν προέλθει από τα μεσαιωνικά υποκοριστικά σε *-ιον*, όπως για παράδειγμα *ὁ παῖς – παιδός > τό παιδ-ίον*, *ἡ μάνδρ-α > το μανδρ-ίον*, *ἡ αὔλαξ > τό αὔλακ-ιον*. Οι περισσότεροι ερευνητές (Χατζιδάκις 1900, Georgacas 1948, Μπαμπινιώτης 1977, Τσοπανάκης 1994, Σταματάκος 1999, Φίλος υπό εκδ.) επισημαίνουν πως συνέβηκε μία ασυνήθιστη αναλογική σιώπηση του *-ον* και τα συγκεκριμένα ουσιαστικά μετατράπηκαν σε



οξύτονα ή παροξύτονα σε *-ι*. Ωστόσο διατήρησαν «ανέπαφες» τις υπόλοιπες κλιτικές μορφές, καθώς και τονικές συμπεριφορές στις πλάγιες πτώσεις. Επιπλέον, σύμφωνα με τον Τσοπανάκη (1994) στην παραπάνω ομάδα εντάχθηκαν και ξένες λέξεις που τονίζονται στην προπαραλήγουσα, όπως *φίλντισι* (τουρκ. *fildişi*: ελεφαντόδοντο), *γιούσουρι* (αραβ. *yüsgü*: κοράλλι), λιγοςτές τουρκικές που τονίζονται στην παραλήγουσα, όπως *γιουρούσι* (τουρκ. *yüğüyüş*: έφοδος, επίθεση), *κανταϊφι* (τουρκ. *kadayif*: είδος σιροπιαστού γλυκού) και λέξεις από διάφορες γλώσσες που τονίζονται στη λήγουσα, όπως *μπουγιουρντί* (τουρκ. *buyurdu*<sup>43</sup>: έγγραφο με δυσάρεστο περιεχόμενο), *πατσουλί* (γαλ. *patchouli*: φτηνό άρωμα), *τσαπαρί* (ιταλ. *chiaparini*: μπετονιά με πολλά αγκίστρια).

Έτσι με την ενσωμάτωση ξένων λέξεων που κλίνονταν, όπως οι λέξεις του (28.) δημιουργήθηκε μία νέα και ανεξάρτητη κλιτική τάξη, η ΚΤ6, που μοιάζει αρκετά με την ΚΤ5. Εντούτοις, ένα από τα αρχικά στάδια δημιουργίας των συγκεκριμένων λέξεων ήταν η επανάλυση και μετατόπιση των μορφηματικών ορίων· το παραγωγικό επίθημα *-ιον* έγινε *-ί* και στη συνέχεια με μετατόπιση των ορίων ενώθηκε με το θέμα, π.χ. *παιδ-ί* > *παιδί*, *καράβ-ι* > *καράβι*. Η αναλογική σιώπηση του *-ον* κατά τους ιστορικούς γλωσσολόγους εξακολουθεί και υφίσταται ως μηδενικό κλιτικό μόρφημα (∅) στο μοντέλο μας στις πτώσεις της ονομαστικής και αιτιατικής του ενικού. Υπάρχει όμως εμφανές κλιτικό μόρφημα στις αντίστοιχες πτώσεις του πληθυντικού (*-α*). Θεωρούμε πως η διαφοροποίηση των ορίων των μορφημάτων στη συγχρονική ανάλυση των ουσιαστικών σε σύγκριση με την προσέγγιση των παραδοσιακών γραμματικών δεν προκαλεί αυθαίρετες αλλαγές στα μορφήματα, ώστε να δικαιολογείται η ύπαρξη αλλομορφων. Στην ανάλυση του υιοθετούμε (Ralli 2000) υποθέτουμε ότι το *-ι* είναι μέρος του θέματος και ακολουθείται από ένα μηδενικό κλιτικό επίθημα.<sup>44</sup> Για την επιβεβαίωση του μηδενικού κλιτικού επιθήματος παραθέτουμε παραδείγματα από σύνθεση. Οι Nespor & Ralli (1996) υποστήριξαν πως στη σύνθεση εφαρμόζεται ένας ειδικός κανόνας τονισμού για κάθε β' συνθετικό που είναι ΘΕΜΑ και όχι ΛΕΞΗ. Σύμφωνα με τον συγκεκριμένο κανόνα ο τόνος τοποθετείται στην προπαραλήγουσα στα σύνθετα τύπου ΘΕΜΑ ΘΕΜΑ ή ΛΕΞΗ ΘΕΜΑ, όπως για παράδειγμα *κουκλό3.σπι2.το1*, *σπιρτό3.κου2.το1*, *σαπιοκά3.ρα2.βο1*

<sup>43</sup> Η τουρκική λέξη είναι ένα ρήμα με κλίση: *buyur+du+{μηδενικό μόρφημα}*: ΔΙΑΤΑΖΩ + ΑΟΡΙΣΤΟΣ + Γ'ΕΝ = "διέταξε".

<sup>44</sup> Υποθέτουμε επίσης ότι το μηδενικό κλιτικό επίθημα είναι η μη-επιφανειακή πραγμάτωση του αποσιωπημένου *-ον*.

*παπαδο3.παι2.δι1*. Η λέξη *παπαδοπαίδι* φαίνεται ότι τονίζεται στην παραλήγουσα. Οι Nespor & Ralli υποστηρίζουν ότι δεν καταστρατηγείται ο κανόνας γιατί υπάρχει μία λήγουσα συλλαβή, το μηδενικό κλιτικό επίθημα [-ø], που δεν είναι φωνολογικώς πραγματοποιμένη στην επιφανειακή δομή. Αν δεν υπήρχε το μηδενικό κλιτικό επίθημα, τότε η λέξη θα τονιζόταν είτε *παπαδόπαιδι*, είτε *παπαδοπαιδί*, γιατί το β' συνθετικό θα ήταν ΛΕΞΗ και ο τόνος δεν μετακινείται. Το συγκεκριμένο παράδειγμα επιβεβαιώνει τον διαχωρισμό των ουσιαστικών της ΚΤ6 σε  $XV_{/i} + \emptyset$  (Ralli 2000).

Στην παραγωγή, όμως, (όπως και στη σύνθεση) υιοθετεί και ακολουθεί την συμπεριφορά όλων των ουσιαστικών που στο κλιτικό τους παράδειγμα το θέμα έχει για ληκτικό χαρακτήρα φωνήεν, όπως *καρδιά*, *ψυχή*, *κύμα*, *παπά-ς*. Τα ουσιαστικά με ληκτικό χαρακτήρα φωνήεν παρουσιάζουν αλλόμορφο με διαγραμμένο το τελικό φωνήεν (*καρδι~*, *ψυχ~*) είτε με την προσθήκη συμφώνου (*κυματ~*, *παπαδ~*). Η ΚΤ6 ενώ δεν έχει αλλόμορφο στην κλίση, «παρακάμπτει» την αρχή της αλλομορφικής συμπεριφοράς και εμφανίζει αλλόμορφο στην παραγωγή και στη σύνθεση. Ο τύπος του αλλομόρφου της ΚΤ6 υπάγεται στην αφαιρετική αλλομορφία, όπως *παιδ-ικ(ός)*, *καραβ-άκι*, *μελ-ένι(ος)*. Η παράβαση της αρχής οφείλεται στη διαδικασία της επανάλυσης, γιατί αν το *-ι* είναι κλιτικό επίθημα, τότε οι καταλήξεις της συγκεκριμένης κλιτικής τάξης είναι {-ι, -ιου, -ι}ΕΝ, {-ια, -ιων, -ια}ΠΛΗΘ, τότε δεν θα εμφάνιζε αλλομορφία σε καμία διαδικασία. Σε προηγούμενες έρευνές μας (Καρασίμος 2009, 2011) υποστηρίζουμε πως η μορφή των αλλαγών του κλιτικού παραδείγματος στα συγκεκριμένα ουσιαστικά (*παιδ-ίον* -> *παιδ-ι* -> *παιδί*, βλ. Debrunner 1908, Mirambel 1978, Πετρούνιας 1984) διατηρεί το τελικό-ι στην κλίση, με αποτέλεσμα να μηδενίζονται οι πιθανότητες για εμφάνιση των αλλομορφων και στο κλιτικό παράδειγμα.

#### 2.4.4.2. Η εξαίρεση των ουσιαστικών τύπου –μπακάλη

Κατά την ανάλυση της ΚΤ2 έχουμε εντοπίσει (Καρασίμος 2001, 2009) μία μικρή ομάδα αρσενικών ουσιαστικών που διαφοροποιεί την αλλομορφική της συμπεριφορά από τα υπόλοιπα της κατηγορίας του. Κατά την ανάλυση της αλλομορφίας στη σύνθεση επισημάναμε πως «τα ουσιαστικά τύπου μπακάλης φαίνεται να μην υιοθετούν τα αλλόμορφο της κλίσης στη σύνθεση και να εμφανίζουν ένα νέο αλλόμορφο διαγράφοντας τον φωνηεντικό χαρακτήρα του θέματος».

(52.) μπακάλη(ς)                      μπακάληδ(ες)                      μπακάλ-ικ(ος)

τεμπέλη(ς)	τεμπέληδ(ες)	τεμπελ-άκ(ος)
ντελάλη(ς)	ντελάληδ(ες)	ντελάλ-ικ(ος)
βεζύρη(ς)	βεζύρηδ(ες)	βεζυρ-ούλη(ς)

Η υποκατηγορία των  $X\eta(\varsigma)\sim X\eta\delta\sim X$  φαίνεται να είναι πιο περίπλοκη και πιο απρόβλεπτη από την αντίστοιχη του τύπου *παιδί*. Όπως υποστηρίζει η Ralli (2000), η ομάδα των  $X\eta(\varsigma)$  που ανήκουν στη δεύτερη κλιτική τάξη έχουν καταχωρημένα αλλόμορφα του τύπου  $X\eta\delta$  (*μπακάλη-ς/ μπακάληδ-ες*). Επομένως, περιμένουμε να συναντήσουμε το  $X\eta\delta$  αλλόμορφο στην παραγωγή και τη σύνθεση, ως τη μοναδική αλλομορφική βάση (βλ. εν. 3.3.1), και να μην εμφανίσουν κάποιο νέο αλλόμορφο. Εντούτοις, αυτή η ομάδα παρουσιάζει ένα νέο αλλόμορφο. Αυτό αποτελεί ιδιαίτερη συμπεριφορά στο Ελληνικό ονοματικό κλιτικό σύστημα, αφού έχει ήδη παρουσιάσει αλλόμορφο στην κλίση. Τα ουσιαστικά αυτής της υποκατηγορίας έχουν τα χαρακτηριστικά [+ΔΑΝΕΙΟ, +ΤΟΥΡΚΙΚΟ].<sup>45</sup> Ωστόσο, υπάρχουν και άλλα ουσιαστικά της ίδιας ΚΤ με τα χαρακτηριστικά [+ΔΑΝΕΙΟ, +ΤΟΥΡΚΙΚΟ] που εμφανίζουν αλλόμορφο στον πληθυντικό τύπου  $XV_{\alpha\epsilon\delta}$  (π.χ. *νταλγά(ς)/ νταλγάδ(ες), καναπέ(ς)/ καναπέδ(ες)*). Αυτά τα ουσιαστικά δεν δημιουργούν καινούργιο αλλόμορφο στην παραγωγή (*νταλγάδικος, καναπεδάκι*) και χρησιμοποιούν το ήδη υπάρχον. Τα συγκεκριμένα αρσενικά ουσιαστικά έχουν τα ακόλουθα χαρακτηριστικά: (α.) είναι δάνεια ενσωματωμένα στο κλιτικό σύστημα των Ελληνικών, (β.) είναι όλα τουρκικής προέλευσης, (γ.) τονίζονται όλα στην παραλήγουσα και (δ.) είναι δηλωτικά επαγγέλματος ή ιδιότητας.<sup>46</sup>

Τα ουσιαστικά του (52.) παραβιάζουν την *αρχή της αλλομορφικής συμπεριφοράς*. Όπως έχουμε υποστηρίξει (Καρασίμος 2009), η αιτία παράβασης είναι φωνολογική. Πιο συγκεκριμένα, δύο φωνολογικά χαρακτηριστικά διαφοροποιούν τα εξεταζόμενα παραδείγματα, σε αντίθεση με τα υπόλοιπα της ίδιας ΚΤ: α.) η διαφορετική τονισμένη συλλαβή, και β.) το φωνολογικό βάρος του Η-ΤΕΛΕΥΤΑΙΟΥΧΑΡΑΚΤΗΡΑ. Πιο συγκεκριμένα, όλα τα παραδείγματα τονίζονται στη παραλήγουσα σε σύγκριση με τα υπόλοιπα ουσιαστικά της ίδιας κλιτικής τάξης με αλλόμορφα του τύπου  $X\eta\sim X\eta\delta-$ , τα οποία τονίζονται στη λήγουσα. Παράλληλα, το /i/ ως υψηλό φωνήεν και μάλιστα άτονο

<sup>45</sup> Υπάρχουν ελάχιστες λέξεις αραβικής και περσικής καταγωγής σε αυτή την ομάδα λέξεων, οι οποίες έχουν επίσης εισέλθει στα Ελληνικά από τα Τουρκικά.

<sup>46</sup> Σε αυτή την ομάδα των κλιτών λέξεων, εντάσσονται και τα παράγωγα ουσιαστικά τύπου *-(ι)άρης*, όπως *ζηλιάρης-ζηλιάρηδες (ζηλιάρικο, ζηλιάρογατος), σκουπιδιάρης-σκουπιδιάρηδες (σκουπιδιάρικο)*.

έχει την τάση να διαγράφεται πιο εύκολα σε μορφολογικούς σχηματισμούς σύμφωνα με την ιεραρχία ηχηρότητας.<sup>47</sup>

Το δεύτερο χαρακτηριστικό είναι το σημαντικότερο στοιχείο για την αιτιολόγηση της παράβασης του περιορισμού. Τα συγκεκριμένα ουσιαστικά εισέρχονται στο κλιτικό σύστημα της Ελληνικής με τελευταίο χαρακτήρα θέματος το σύμφωνο, ενώ το /i/ εμφανίζεται μόνο στη βάση των Ελληνικών (53.α) σε αντίθεση με τα υπόλοιπα τουρκικά δάνεια σε -ής (53.β) που το /i/ αποτελεί μέρος της τουρκικής λέξης.

(53.)	Τουρκικά		Ελληνικά
α.	<i>bakkal</i>	>	μπακάλη-ς
	<i>vezir</i>	>	βεζίρη-ς
	<i>tembel</i>	>	τεμπέλη-ς
	<i>muhacir</i>	>	ματζίρη-ς
	<i>sakat</i>	>	σακάτη-ς
	<i>çifit/ çifut</i> <sup>48</sup>	>	τσιφούτη-ς
β.	<i>kadi</i>	>	κατή-ς
	<i>batakçi</i>	>	μπαταξή-ς
	<i>bekri</i>	>	μπεκρή-ς
	<i>dayi</i>	>	νταή-ς
	<i>ispahi</i>	>	σπαχή-ς
γ.	<i>nargile</i>	>	ναργιλέ-ς
	<i>kahve</i>	>	καφέ-ς
	<i>köfte</i>	>	κεφτέ-ς
	<i>baklava</i>	>	μπακλαβά-ς
	<i>zorba</i>	>	ζορμπά-ς
	<i>kavgā</i>	>	καυγά-ς

<sup>47</sup> Σύμφωνα με την Selkirk (1984) η ιεραρχία ηχηρότητας είναι  $a > o > u > \varepsilon > i$ , όπου το  $a$  είναι το πιο ηχηρό και το  $i$  το λιγότερο ηχηρό. Όμως η συγγραφέας σημειώνει ότι «μερικές φορές το /i/ αποδεικνύεται πιο δυνατό από το /ε/, οπότε έχουμε διαφορετική ιεραρχία η οποία είναι  $a > o > u > i > \varepsilon$ ». Παράλληλα οι Holton, Mackridge & Φιλίππκη (1999 : 23) υιοθετούν την παραπάνω άποψη, αναφέροντας ότι «ο πιο σπουδαίος παράγοντας που επηρεάζει την επιλογή του φωνήεντος που θα απαλειφθεί είναι η θέση του φωνήεντος αυτού στην ιεραρχία ηχηρότητας (αντηχητικότητας). Τα πέντε φωνήεντα της Ελληνικής ταξινομούνται σε μία κλίμακα ηχηρότητας, από το πιο ηχηρό (πιο αντηχητικό) /a/ ως το λιγότερο ηχηρό (λιγότερο αντηχητικό) /i/».

<sup>48</sup> Η αρχική σημασία της λέξης 'çifit' ήταν 'Εβραίος'.

Τα ουσιαστικά του τύπου  $XV_{\{[a/\epsilon]\}}$  (53.γ) και οι αντίστοιχες τουρκικές λέξεις έχουν τα ίδια ληκτικά φωνήεντα του θέματος, όπως *μπακλαβάς*, *νταλκάς*, *ναργιλές*, *καφές*, όπως και τα αντίστοιχα ουσιαστικά  $XV_{[i]}$  (53.β), για παράδειγμα *κατής*, *μπεκρής*, *νταής*. Κατά τον εξελληνισμό των δανείων υποθέτουμε πως το παραγωγικό επίθημα  $-η(ς)\sim -ηδ-$  συνδυάστηκε με τις τουρκικές λέξεις της ομάδας (53.α), για να εισαχθούν οι λέξεις στο κλιτικό σύστημα των Ελληνικών. Οι συγκεκριμένες λέξεις έχουν επαναναλυθεί και το παραγωγικό επίθημα αποτελεί πλέον μέρος του θέματος και είναι μορφολογικά αδιαφανές. Με αυτό τον τρόπο το κλιτικό τους παράδειγμα ομοιάζει πλήρως με τα ουσιαστικά σε  $-ής$  της ΚΤ2. Η διαφοροποίηση των ουσιαστικών της ομάδας (53.α) από τα άλλα ουσιαστικά με  $\delta$ -αλλόμορφο έγκειται στο γεγονός ότι η αρχική μορφή της τουρκικής λέξης (όπου ο ληκτικός χαρακτήρας ήταν σύμφωνο) του επιτρέπει να συμμετάσχει στην παραγωγή και στη σύνθεση και έτσι εμφανίζει ένα νέο αλλόμορφο σε αυτές τις διαδικασίες. Με άλλα λόγια το  $-η-$  δεν παρουσιάζεται στην παραγωγή και στη σύνθεση, γιατί το θέμα χωρίς  $-η-$  πληροί τις προϋποθέσεις του περιορισμού βέλτιστης συλλαβής (βλ. εν. 3.3.1) και δεν υπάρχει κανένας φωνολογικός λόγος παρουσίας, όπως στην κλίση. Είναι μία συστηματική «παράβαση» της αρχής αλλομορφικής συμπεριφοράς με την εμφάνιση ενός νέου αλλομόρφου, που ομοιάζει με τη μορφή της τουρκικής λέξης - προέλευσης.

#### 2.4.4.3. Άλλες εξαιρέσεις

Όπως επισημαίνει ο Drachman (2006) έχουμε και κάποιες άλλες λέξεις που παραβιάζουν την αρχή αλλομορφικής συμπεριφοράς. Παραθέτουμε τα παραδείγματά του:

(54.)	ΚΛΙΣΗ	ΠΑΡΑΓΩΓΗ	ΣΥΝΘΕΣΗ
α.	καφέ(ς)	καφεδ-άκι	καφ-ό-μπρικο/ καφεδόμπρικο <sup>49</sup>
β.	παπά(ς)	παπαδ-άκι	παπαδ-ο-παίδι/ παπα-Γιάννης
γ.	κρέα(ς)	[κρεατ-άκι]	1. κρε-ο-φαγία, κρεατ-ο-πωλείο 2. κρεατ-ό-πιτα, κρεατ-ο-σανίδα
δ.	αίμα	[αιματ-άκι]	1. αιμ-ο-σφαίρια, αιμ-ο-δοσία 2. αιματ-ο-κυλώ, αιματ-ο-βαμμένος

<sup>49</sup> Καμία από τις δύο μορφές δεν βρέθηκαν σε κάποιο λεξικό ή στο σώμα ΕΘΕΓ του ΙΕΛ (<http://hnc.ilsp.gr>), ενώ στη μηχανή αναζήτησης Google ήταν αποδεκτό κυρίως το *καφεδόμπρικο*.

ε. μάλαμα                      μαλαματ-ένιος                      μαλαμ-ο-καπνισμένος<sup>50</sup>

(από Drachman 2006: : 12-13)

Τα ουσιαστικά *κρέας* και *αίμα* φαίνεται πράγματι πως σε συγκεκριμένες περιπτώσεις παραβιάζουν την αρχή αλλομορφικής συμπεριφοράς για διαφορετικούς λόγους το καθένα. Το *κρέας* εμφανίζει ένα νέο αλλόμορφο, το *κρε~*, κατά το σχηματισμό κάποιων συνθέτων (54.γ1), ενώ παράλληλα χρησιμοποιεί πάντα στην παραγωγή και στα περισσότερα σύνθετα το κλιτικό αλλόμορφο *κρεατ~* (54.γ2). Κατάλοιπο της αρχαίας κλίσης<sup>51</sup> το *κρε~* εμφανίζεται σε συγκεκριμένες λέξεις που σχηματίστηκαν σε πολύ παλαιότερες φάσεις της Ελληνικής γλώσσας: η ύπαρξη της διπλοτυπίας φανερώνει ότι το αλλόμορφο *κρεατ~* είναι κυρίαρχο και τείνει να εξαλείψει το αρχαιοελληνικό θέμα (*κρεοφαγία* vs. *κρεατοφαγία*, *κρεοπόλης* vs. *κρεατοπόλης*). Από την άλλη, η λέξη *αίμα*, όπως έχει επισημανθεί από τον Καρασίμο (2001, 2011), εμφανίζει νέο αλλόμορφο μόνο στη σύνθεση παραβιάζοντας τον περιορισμό, γιατί οι σύνθετες λέξεις που δημιουργήθηκαν είτε στα Αρχαία Ελληνικά είτε στα Γαλλικά έχουν το αλλόμορφο *αιμ~*. Αντίθετα αυτές που σχηματίστηκαν στα Αγγλικά ή στα Νέα Ελληνικά χρησιμοποιούν το ήδη υπάρχον *αιματ~*. Είναι από τις ξεκάθαρες περιπτώσεις αλλομορφίας που οφείλονται στο δανεισμό.

## 2.5. Η Αλλομορφία ως Κριτήριο Κατηγοριοποίησης

Σε προηγούμενες υποενότητες παρατηρήσαμε πως η αλλομορφική συμπεριφορά των κλιτών λέξεων καθορίζει τη συμμετοχή τους σε συγκεκριμένες κλιτικές ομάδες. Ήδη οι Ralli (1988, 2006, Ράλλη 2005), Booij (1997, 2005), Rubah & Booij (2001), μεταξύ άλλων, επισημαίνουν πως η αλλομορφία δεν είναι απλώς η έλλειψη κανονικότητας μέσα σε μία γλώσσα που δημιουργεί «ανώμαλους» τύπους και εξαιρέσεις, αλλά ένα σημαντικό κριτήριο για την ταξινόμηση των λέξεων σε κατηγορίες με ίδια χαρακτηριστικά και αυτό λαμβάνει χώρα σε όλες τις διαδικασίες σχηματισμού λέξεων.

<sup>50</sup> Στα λεξικά υπάρχει μόνο η μορφή *μαλαματοκαπνισμένος* και όχι *!μαλαμοκαπνισμένος*.

<sup>51</sup> Στα αρχαία Ελληνικά δεν εμφανίζεται το αλλόμορφο *κρεατ*. Η κλίση της λέξης *κρέας* είναι η ακόλουθη (βλ. Οικονόμου 1977: 85) και έχει μοναδικό θέμα το *κρε*:

ΕΝΙΚΟΣ	τό κρέας	ΠΛΗΘΥΝΤΙΚΟΣ	τά κρέα
	τοῦ κρέως		τῶν κρεῶν
	τῷ κρεα		τοῖς κρέασι
	τό κρέας		τά κρέα
	(ῶ) κρέας		(ῶ) κρέα

Σε αντίθεση με την άποψη της Galani (2003a, 2003b), στα Νέα Ελληνικά δεν υφίστανται θεματικά φωνήεντα και κυρίως δεν υπήρξαν ποτέ τα συγκεκριμένα θεματικά φωνήεντα που αναφέρει στην Ελληνική γλώσσα. Η θεωρία της Ralli (1988) διαχωρίζει τα ρήματα της ΚΝΕ επίσης σε δύο κλιτικές τάξεις με κριτήριο κατηγοριοποίησης, όχι πλέον την εμφάνιση ή όχι θεματικού φωνήεντος, αλλά την παρουσία ή απουσία συστηματικής αλλομορφίας. Στις ονοματικές κλιτικές τάξεις τα κριτήρια κατηγοριοποίησης είναι τρία: (α) η μορφή των επιθημάτων, (β) το γένος και (γ) η αλλομορφία. Κατά τη Ralli (2007) τα πάμπολλα δεδομένα από τις διαλέκτους της ΚΝΕ ενισχύουν την άποψη ότι η αλλομορφία παίζει σημαντικό ρόλο στην κλίση.

Επεκτείνοντας την προσέγγιση της Ralli (2007) θα αποδείξουμε πως και στην παραγωγή η αλλομορφία διαδραματίζει σημαντικό ρόλο (βλ. κεφάλαιο 3). Μεταξύ των άλλων λειτουργιών της αλλομορφίας, κατηγοριοποιεί και τα παραγωγικά επιθήματα σε κλιτικές τάξεις, εφόσον οι παράγωγες λέξεις είναι κλιτές. Όπως παρουσιάζεται στο επόμενο κεφάλαιο, ο τύπος της αλλομορφίας ενός παραγωγικού επιθήματος είναι το μοναδικό κριτήριο και προϋπόθεση, για να ενταχθεί σε κάποια κλιτική τάξη. Για παράδειγμα, τα παραγωγικά επιθήματα *-οσύνη*, *-οτητα*, *-εια*, *-ουλα* παρουσιάζουν αφαιρετική αλλομορφία τύπου  $X\Phi \sim X$  (*-οσυν~*, *-οτητ~*, *-ει~*, *-ουλ*) και επομένως, συνδυάζονται με τις καταλήξεις της ΚΤ3 (όπως τα ουσιαστικά *καρδιά* και *ψυχή*). Επίσης, κάποια ρηματικά επιθήματα, όπως *-ίζ(ω)*, *-ευ(ω)*, *-ών(ω)*, συνδυάζονται με τις ρηματικές κλιτικές καταλήξεις της ΚΤ1 λόγω της μη συστηματικής τους αλλομορφίας.

Μετά τη διαδικασία σχηματισμού λέξεων στο επίπεδο της παραγωγής και τη μετάβαση στο επίπεδο της κλίσης, το θέμα μαζί με το/τα παραγωγικό/ά επιθήμα/τα συμπεριφέρεται ως μία ολότητα στην οποία κυριαρχούν τα χαρακτηριστικά του επιθήματος λόγω της **μεταπήδησης χαρακτηριστικών**<sup>52</sup> (percolation principle, βλ. Lieber 1980, Selkirk 1982, Di Sciullo & Williams 1987, Spencer 1991, Ράλλη 1993, 2005). Για να κυριαρχήσουν τα χαρακτηριστικά του επιθήματος πρέπει να υπάρχουν λήμματα παραγωγικών επιθημάτων με χαρακτηριστικά αλλομορφίας και κλιτικής τάξης – όπως ισχύει με τα θέματα – και τα παραγωγικά επιθήματα να βρίσκονται σε θέση κεφαλής του σχηματισμού. Το παραγωγικό επίθημα κληροδοτεί τα

<sup>52</sup> Ο Ευδόπουλος το μεταφράζει ως διαπήδηση χαρακτηριστικών (Κρύσταλ 2003).

χαρακτηριστικά (αλλομορφίας και κλιτικής τάξης) στον παραγωγικό σχηματισμό, ώστε να ενταχθεί σε κάποια ΚΤ.

Ειδικότερα, συγκρίνοντας τη θεματική αλλομορφία με την επιθηματική αλλομορφία, παρατηρούμε κάποιες κανονικότητες. Τα επιθήματα δείχνουν παρόμοια αλλόμορφα με αυτά των θεμάτων, με τα οποία τα προηγούμενα μοιράζονται μερικά κοινά χαρακτηριστικά. Παραδείγματος χάριν, το επίθημα *-α(ς)* ‘κάποιος που ενεργεί’ είναι παρόμοιο με τα ουσιαστικά τύπου *παπάς*: ακολουθούν τον κανόνα εισαγωγής-δ για τον σχηματισμό του πληθυντικού και των παραγώγων.

(55.) α.	<i>παπάς(ς)</i>	<i>παπάδ(ες)</i>	<i>-α(ς)</i>	<i>παπλωματ-ά(ς)</i>	<i>παπλωματάδ(ες)</i>
	<i>καρδιά</i>	<i>καρδι(ές)</i>	<i>-άλα</i>	<i>κρεμ-άλα</i>	<i>κρεμάλ(ες)</i>
	<i>βουν(ό)</i>	<i>βουν(ά)</i>	<i>-εί(ο)</i>	<i>γραφ-εί(ο)</i>	<i>γραφεί(α)</i>
	<i>κύμα</i>	<i>κύματ(α)</i>	<i>-μα</i>	<i>πέταγ-μα</i>	<i>πετάγματ(α)</i>
β.	<i>τεμπέλη(ς)</i>	<i>τεμπέληδ(ες)</i>	<i>-άρης</i>	<i>περιβολ-άρη(ς)</i>	<i>περιβολάρηδ(ες)</i>
	<i>παιδί</i>	<i>παιδ-άκι</i>	<i>-ούλι</i>	<i>σακ-ούλι</i>	<i>σακουλ-άκι</i>

Αντίστοιχα στα υπόλοιπα παραδείγματα (55.α) εντοπίζουμε περιπτώσεις, όπου τα επιθήματα φαίνεται να «αντιγράφουν» την αλλομορφία των θεμάτων, όπως τα επιθήματα *-άλα* και *-μα* που εμφανίζουν ίδιο τύπο αλλομορφίας με τα θέμα της *καρδιά* και *κύμα* αντίστοιχα. Αυτή η κανονικότητα εμφανίζεται επίσης και ανάμεσα στα ρήματα και τα ρηματικά επιθήματα. Αυτό συμβαίνει συνήθως λόγω της επανάλυσης των λέξεων, όπου ένα πρώην ανεξάρτητο επίθημα είναι πλέον μέρος του θέματος. Η πιο εντυπωσιακή κανονικότητα είναι αυτή των λέξεων του (55.β). Όπως έχουμε ήδη τονίσει προηγουμένως, είναι μερικά ουσιαστικά που παράγουν νέα αλλόμορφα κατά τη διαδικασία της παραγωγής (και σύνθεσης). Παρομοίως, τα επιθήματα, όπως *-άρη(ς)*, *-ούλι*, συμπεριφέρονται ακριβώς, όπως τα ουσιαστικά *τεμπέλη(ς)* και *παιδί*. Αυτό σημαίνει ότι δημιουργούν ένα αλλόμορφο παρόλη την παρουσία του (*ταξιδ-ιάρη(ς)* *ταξιδ-ιάρηδ(ες)*, αλλά *ταξιδ-ιάρ-ικός*) ή την απουσία του (*σακ-ούλι* *σακ-ούλι(α)*, αλλά *σακ-ουλ-άκι*) στην κλίση. Αυτό καταδεικνύει ότι η αλλομορφία συμμετέχει σε όλα τα επίπεδα σχηματισμού λέξεων με την ίδια κανονικότητα.



*“Irregular allomorphy is not merely the residual product of failure to analyze alternation into underlying invariance, but an autonomous principle rooted, like ‘one meaning – one form’, in the nature of the relationship between signans and signatum”*

Maiden (1992)

## 3. Παραγωγή και Αλλομορφία

### 3.1. Η αλλομορφία των παραγωγικών προθημάτων

Τα προθήματα είναι κατηγορία προσφυσμάτων τοποθετημένα πριν από κάποιο θέμα/ ρίζα/ λέξη και δεν αλλάζουν ποτέ τη γραμματική κατηγορία της βάσης. Είναι δεσμευμένα μορφήματα. Σε μία πρώτη προσέγγιση, συγκαταλέγονται από τη Ράλλη (2005) στα προθήματα των Νέων Ελληνικών το *α-*, *δυσ-*, *ευ-* και το *ξε-*, εκ των οποίων τα τρία πρώτα είναι αρχαιοελληνικά και το τελευταίο είναι προϊόν επανάλυσης.<sup>1</sup>

(56.)	<i>α-/αν-</i>	<i>α-κατάλληλος, α-όρατος, άν-εργος, αν-αίτιος</i>
	<i>δυσ-</i>	<i>δυσ-νόητος, δύσ-τροπος, δύσ-βατος</i>
	<i>ευ-</i>	<i>εύ-λογος, ευ-κατάστατος, εύ-στροφος</i>
	<i>ξε-/ξ-</i>	<i>ξε-δίνω, ξε-χωρίζω, ξ-αφρίζω, ξ-αγρυπνώ</i>

Η ομάδα των Αρχαίων Ελληνικών προθέσεων έχει αποτελέσει σημείο τριβής ανάμεσα στους μορφολόγους, εάν ο συνδυασμός τους με κάποιο θέμα/ ρίζα/ λέξη αποτελεί μέρος της διαδικασίας της σύνθεσης ή της επιθηματοποίησης. Οι συγγραφείς

<sup>1</sup> Οι Τριανταφυλλίδης (1941), Καραντζόλα & Γιαννουλοπούλου (2000), Ralli (2002, 2004), Ευθυμίου (2001, 2002, 2003), Ralli (2004), Ράλλη (2005), Γαβριηλίδου (2008) υποστηρίζουν ότι το *ξε-* δημιουργήθηκε από επανάλυση του προθήματος *εκ-* και της χρονικής αύξησης *ε-*.

των παραδοσιακών γραμματικών (βλ. Τριανταφυλλίδης 1941, Mirambel 1978, Πετρούνιας 1984, Τσοπανάκης 1994 μεταξύ άλλων) και ο Μπαμπινιώτης (1969) εκφράζουν την άποψη ότι οι σχηματισμοί με αρχαιοελληνικές προθέσεις ανήκουν στη σύνθεση. Αντιθέτως, η Ralli (2002, 2004) εντάσσει τις προθέσεις στην κατηγορία των προθημάτων και επομένως οι σχηματισμοί τους είναι μέρος της παραγωγής, παρόλο που υπήρξαν ελεύθερα μορφήματα σε πολύ παλαιότερο στάδιο της γλώσσας. Πιο συγκεκριμένα, η δεύτερη ομάδα προθημάτων συμπεριλαμβάνει όλες τις προθέσεις των ΑΕ που είναι οι *ανα-*, *αντι-*, *απο-*, *δια-*, *εισ-*, *εν-*, *εκ-*, *επι-*, *κατα-*, *μετα-*, *παρα-*, *περι-*, *προ-*, *προσ-*, *συν-*, *υπερ-* και *υπο-*.

(57.)	<i>ανα-/αν-</i>	ανα-ψυχή, ανα-μορφώνω, αν-έρχομαι
	<i>αμφι-</i>	αμφι-βάλλω, αμφι-θέατρο, αμφί-βιος
	<i>αντι-/αντ- ~ανθ-</i>	αντί-λογος, αντ-έχω, ανθ-ίσταμαι
	<i>απο-/απ- ~αφ-</i>	απο-καλώ, απ-έχω, αφ-οπλίζω
	<i>δια-/δι-</i>	δια-λαλώ, δια-τυμπανίζω, δι-αγωνίζομαι
	<i>εισ-</i>	εισ-ρέω, εισ-ηγούμαι, εισ-φέρω
	<i>εκ-/εξ-</i>	εκ-κρεμώ, εκ-τοπίζω, εξ-έρχομαι
	<i>εν-/εμ-/εγ-/ελ-/ερ-</i>	έν-νομος, εμ-μένω, εγ-καλώ, ελ-λιμενίζομαι, έρ-ρινος
	<i>επι-/επ- ~εφ-</i>	επι-σκιάζω, επ-αινώ, εφ-ησυχάζω
	<i>κατα-/κατ- ~καθ-</i>	κατα-χρεώνω, κάτ-εργο, καθ-αιρώ
	<i>μετα-/μετ- ~μεθ-</i>	μετα-κομίζω, μετ-οικώ, μεθ-εόρτια
	<i>παρα-/παρ-</i>	παρα-μιλώ, παρα-τραβώ, παρ-ενοχλώ
	<i>περι-</i>	περι-τριγυρίζω, περι-φράζω, περι-αύλιο
	<i>προ-</i>	προ-πορεύομαι, προ-ωθώ, πρό-ναος
	<i>προσ-</i>	προσ-γειώνω, προσ-εδαφίζω, πρόσ-καιρος
	<i>συν-/συμ-/συλ-/</i>	συν-τρέχω, συμ-πλέκω, συλ-λογή
	<i>συγ-/συρ-/συσ-</i>	συγ-κρατώ, συρ-ρέω, σύσ-σωμος
	<i>υπερ-</i>	υπερ-βάλλω, υπερ-εκτιμώ, υπερ-έχω
	<i>υπο-/υπ- ~υφ-</i>	υπο-θάλπτω, υπ-αρχηγός, υφ-υπουργός

Κατάλοιπο των Αρχαίων Ελληνικών είναι επίσης και μία τρίτη ομάδα μορφημάτων που τα εντάσσουμε πλέον στο προθήματα και δεν απαντώνται στα Νέα Ελληνικά ως ελεύθερα μορφήματα, καθώς δεν συμμετέχουν ούτε σε τυποποιημένες φράσεις. Τα

συγκεκριμένα μορφήματα χαρακτηρίζονται ως [+λόγιο] είναι τα ακόλουθα: *αλεξι-*, *αμφι-*, *αρχι-*, *ημι-*, *παν-*.

(58.)	<i>αλεξι-</i>	αλεξι-κέραυνο, αλεξι-σφαιρος, αλεξι-πυρος
	<i>αρχι-/αρχ-</i>	αρχι-διάκονος, αρχ-ίατρος, αρχι-επίσκοπος
	<i>ημι-</i>	ημί-θεος, ημι-κύκλιο, ημί-χρονο
	<i>παν-/παγ-/παμ-</i>	πάν-θεο, παγ-κόσμιος, παμ-μέγιστος
	<i>παλ-</i>	παλ-λαϊκός
	<i>υψι-</i>	υψί-φωνος, υψι-πετής, υψι-κάμινος

Τέλος, υπάρχει μία ομάδα μορφημάτων που αποκαλούνται **προσφυματοειδή** (affixoids, Fleischer 1965, Marchand 1967, Scalise 1994), στην οποία ανήκουν και τα **προθηματοειδή** (prefixoids). Με τη συγκεκριμένη ορολογία δηλώνουμε μορφήματα που μοιάζουν να συμμετέχουν στη διαδικασία της σύνθεσης, εμφανίζονται ως ανεξάρτητα λεξήματα, αλλά έχουν συγκεκριμένη λειτουργία και πιο περιορισμένη σημασία (για περισσότερα, Booij 2005b, 2010, Δημελά 2005, 2009α, 2009β, 2010, Ralli 2009). Για παράδειγμα στις ΚΝΕ και στις διαλέκτους θεωρούνται προθηματοειδή, τα *ακρο-* (*ακροπατώ*, *ακροσυγγενεύω*), *ψιλο-* (*ψιλοάσχετος*, *ψιλοκουράζομαι*), *χαμο-* (*χαμοπατώ*, *χαμοπερπατώ*) μεταξύ άλλων. Καθ' όσον, δεν αποτελούν ξεκάθαρη περίπτωση προθημάτων<sup>2</sup>, δεν θα ενταχθούν στο συγκεκριμένο ζήτημα που διερευνούμε.

Εξετάζοντας τα προθήματα των Νέων Ελληνικών παρατηρούμε ότι η αλλομορφία τους είναι περιορισμένη, αφού μόλις το 24% των τριών κατηγοριών προθημάτων έχουν αλλόμορφα, ενώ τα περισσότερα από αυτά έχουν δύο τύπους λόγω του φωνολογικού κανόνα της αποβολής φωνήεντος. Τέλος, υπάρχουν και τρία προθήματα (*εν-*, *συν-*, *παν-*) με αρκετά φωνόμορφα λόγω της εφαρμογής του κανόνα της αφομοίωσης. Και τα αλλόμορφα των έξι προθημάτων (*αντι-*, *απο-*, *επι-*, *κατα-*, *μετα-*, *υπο-*) είναι προϊόντα

<sup>2</sup> Σε γενικές γραμμές, κατά τη Ράλλη (2005: 43) προκύπτουν κάποια κριτήρια χαρακτηρισμού προθημάτων που οφείλουν να πληρούν σε μεγάλο βαθμό τις προϋποθέσεις, για να καταχωρηθούν ως προθήματα. Κατ' αρχάς, εμφανίζονται πλέον μόνο ως δεσμευμένα μορφήματα σε σχηματισμούς λέξεων και δεν είναι ελεύθερα και αμέτοχα σε σχηματισμούς, όπως οι προθέσεις της Νέας Ελληνικής *με*, *σε*, *για*. Παράλληλα, εάν ήταν ελεύθερα μορφήματα θα έπρεπε να υπάρχει και στις δύο θέσεις (α' και β' συστατικό) της σύνθεσης (ΜΟΡΦΟΛΟΓΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ). Επιπροσθέτως, η χρήση τους ως προθέσεις εντοπίζεται σε ελάχιστες και στερεότυπες εκφράσεις – κατάλοιπες από τα Αρχαία Ελληνικά – και έχουν πλέον μηδενική συχνότητα εμφάνισης σε νέους φραστικούς σχηματισμούς (ΠΑΡΑΓΩΓΙΚΟΤΗΤΑ). Τέλος, η σημασία που προσδίδουν στο συνδυαζόμενο θέμα/λέξη δεν είναι σταθερή, σαφής και προβλέψιμη, σε αντίθεση με τις νεοελληνικές προθέσεις ή (σχεδόν πάντα) με τα συνθετικά που είναι σαφής η σημασία τους (ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ).

του ίδιου φωνολογικού κανόνα των Αρχαίων Ελληνικών, τον νόμο της δάσυνσης, ο οποίος είναι συγχρονικά ανενεργός.

(59.) αντι-/ανθ-	ανθ-ύγιεινός, ανθ-ΐσταμαι, ανθ-Έλληνας
απο-/αφ-	αφ-όρίζω, αφ-όπλίζω, αφ-αΐμαξη
επι-/εφ-	εφ-ήσυχάζω, εφ-όρμώ, εφ-ΐππος
κατα-/καθ-	καθ-άγιάζω, καθ-όδος, καθ-Έλξη
μετα-/μεθ-	μεθ-έορτιος, μεθ-αΰριο, μεθ-όδεύω
υπο-/υφ-	υφ-ύπουργός, υφ-Ήλιος, υφ-ήγεσία

Ο κανόνας της δάσυνσης εφαρμοζόταν στα όρια μορφημάτων, όπου το δασυνόμενο φωνήεν της λέξης που ακολουθούσε μετέφερε τη δάσυνσή (/<sup>h</sup>/) του στο τελικό σύμφωνο (ένα εκ των /t/, /k/, /p/) του προηγούμενου μορφήματος. Έτσι, το σύμφωνο μεταβαλλόταν στο αντίστοιχο δασύ σύμφωνο των Αρχαίων Ελληνικών, δηλαδή το /t/ σε /t<sup>h</sup>/, το /k/ σε /k<sup>h</sup>/ και το /p/ σε /p<sup>h</sup>/. Στα Νέα Ελληνικά, όμως, δεν υφίσταται πλέον η δάσυνση και επομένως οι αλλαγές του μορφήματος αποτελούν αλλόμορφα.

Συγχρονικά εντοπίζεται μία ασυνέπεια στη χρήση των αλλομόρφων των προθημάτων. Κατά τον σχηματισμό νεολογισμών με κάποιο πρόθημα παρατηρείται το φαινόμενο της χρήσης και των δύο μορφών του προθήματος, ανάλογα με το επίπεδο χρήσης των Ελληνικών του ομιλητή. Παρότι υπάρχει η λέξη *ανθελληνικός*, είναι δυνατόν να βρεθεί και η μορφή *αντιελληνικός*. Επίσης, συναντάται ο τύπος *καθορμώ* (ορμώ υπερβολικά εναντίον), καθώς και ο τύπος *καταορμώ*, όπως η λέξη *πενταήμερη* παράλληλα με την *πενθήμερη*. Επομένως, παρατηρείται η δημιουργία διτυπίας, για την οποία ευθύνεται συνήθως η άγνοια των Αρχαίων Ελληνικών από τους ομιλητές που λειτουργούν σύμφωνα με τους συγχρονικούς μηχανισμούς της γλώσσας. Τύποι που συγχρονικά χαρακτηρίζονται ως μη γραμματικοί, συχνά αποτελούν μελλοντικά γραμματικούς τύπους. Ενδεικτικό παράδειγμα η λέξη *πενταήμερος* που πλέον υπερισχύει σε συχνότητα χρήσης της αντίστοιχης *πενθήμερος*.

### 3.2. Η αλλομορφία των παραγωγικών επιθημάτων

Τα επιθήματα είναι μία κλειστή ομάδα προσφυσμάτων που τοποθετείται μετά από θέματα (*ανθρώπ-ιν-ος*) ή δεσμευμένα θέματα (*πυροβोल-ικ-ό*). Και ενώ τα κλιτικά επιθήματα έχουν γραμματικές πληροφορίες, τα παραγωγικά επιθήματα μεταφέρουν

λεξικές πληροφορίες. Σε σύγκριση με τα προθήματα, τα παραγωγικά επιθήματα είναι σαφώς πιο μεγάλη κατηγορία με ποικιλία, έντονη αλλομορφική συμπεριφορά και συνδυαστικότητα μεταξύ τους. Χαρακτηριστικά η Ράλλη (2005) αναφέρει πως τα παραγωγικά επιθήματα εμφανίζουν ποικίλα και ενδιαφέροντα χαρακτηριστικά, τα οποία δεν έχουν ακόμα μελετηθεί επαρκώς στο σύνολό τους, αν και υπάρχει έντονη ερευνητική κινητικότητα στο συγκεκριμένο τομέα τα τελευταία χρόνια. Τα παραγωγικά επιθήματα επιφέρουν σημαντικές αλλαγές στη βάση, τόσο στη γραμματική της κατηγορία (συνήθως), όσο και στη σημασία της. Στις υποενότητες που έπονται παρατίθεται μία μικρή μα χαρακτηριστική ομάδα παραγωγικών επιθημάτων<sup>3</sup> από κάθε κατηγορία. Παράλληλα δηλώνεται η αλλαγή βάσης και η (πιθανή) αλλομορφική τους συμπεριφορά. Για πρακτικούς λόγους, δίνεται και το κλιτικό επίθημα που ακολουθεί.

Ιδιαίτερα, τα επιθήματα ουσιαστικών είναι η μεγαλύτερη υποομάδα των παραγωγικών επιθημάτων της Ελληνικής. Για την καλύτερη και πιο οργανωμένη παρουσία της δίνονται ομαδοποιημένα σύμφωνα με τις συνηθισμένες κατηγορίες των σύγχρονων γραμματικών και γλωσσολογικών προσεγγίσεων.

#### (60.) Παραγωγικά Επιθήματα ονοματικής κατηγορίας

##### (i.) Γενικά επιθήματα ουσιαστικών

ΒΑΣΗ <sub>[P]</sub>	-τή(ς) ~ -τ-	γητευ-τή-ς γητευ-τ-ές	(γητεύ-ω)
ΒΑΣΗ <sub>[O/P]</sub>	-εί(ο)	γραφ-εί-ο	(γράφ-ω)
ΒΑΣΗ <sub>[O]</sub>	-ώνα(ς) ~ -ων-	αμπελ-ώνα-ς μπελ-ών-ες	(αμπελ~ [αμπελι])
ΒΑΣΗ <sub>[E/O]</sub>	-[ό]τητα ~ -[ό]τητ-	αρχαι-ότητα αρχαι-ότητ-ες εθν-ότητα εθν-ότητ-ες	(αρχαί-ος) (έθν-ος)

##### (ii.) Ειδικά επιθήματα – Υποκοριστικά

<sup>3</sup> Ο πλήρης κατάλογος των παραγωγικών επιθημάτων της ΚΝΕ βρίσκεται στο Παράρτημα – Appendix I. Για περισσότερες πληροφορίες ανατρέξτε εκεί.

ΒΑΣΗ <sub>[E/O]</sub>	-άκι	μαυρ-άκι	(μαύρ-ος)
ΒΑΣΗ <sub>[O]</sub>	-άκη(ς) ~ -άκηδ-	κοσμ-άκη-ς κοσμ-άκηδ-ες	(κόσμ-ος)
ΒΑΣΗ <sub>[O]</sub>	-ίτσα ~ -ίτσ-	αυλ-ίτσα αυλ-ίτσ-ες	(αυλή)
ΒΑΣΗ <sub>[E]</sub>	-ούτσικ(ος)	μικρ-ούτσικ-ος	(μικρ-ός)

(iii.) Ειδικά επιθήματα – Μεγεθυντικά

ΒΑΣΗ <sub>[O]</sub>	-αρά(ς) ~ -αράδ-	παιδ-αρά-ς παιδ-αράδ-ες	(παιδ~ [παιδί])
ΒΑΣΗ <sub>[O]</sub>	-ούκλα ~ -ούκλ-	ψαρ-ούκλα ψαρ-ούκλ-ες	(ψαρ~ [ψάρι])
ΒΑΣΗ <sub>[O]</sub>	-ούρα ~ -ούρ-	παιχτ-ούρα παιχτ-ούρ-ες	(παιχ-τη~παιχ-τ)
ΒΑΣΗ <sub>[O]</sub>	-αρ(ος)	γάτ-αρ-ος	(γάτ-ος)

(iv.) Ειδικά επιθήματα – Εθνικά/ Πατριδωνυμικά

ΒΑΣΗ <sub>[O]</sub>	-έζ(ος)	κιν-έζ-ος	(Κίνα)
ΒΑΣΗ <sub>[O]</sub>	-ι(ός)	θεσσαλονικ-ι-ός	(Θεσσαλονίκη)
ΒΑΣΗ <sub>[O]</sub>	-ίτη(ς) ~ -ίτ-	ανατολ-ίτη-ς ανατολ-ίτ-ες	(Ανατολη~ Ανατολ)
ΒΑΣΗ <sub>[O]</sub>	-ιν(ος)	πατρ-ιν-ός	(Πατρα~ Πατρ)

(v.) Ειδικά επιθήματα – Επαγγελματικά

ΒΑΣΗ <sub>[O/P]</sub>	-έα(ς) ~ -ε- ~ Ø	δρομ-έα-ς δρομ-έ-ων – δρομ-Ø-είς σφαγ-έα-ς σφαγ-έ-ων – σφαγ-Ø-είς	(δρόμ-ος) (σφαγ~ [σφάζ-ω])
ΒΑΣΗ <sub>[O]</sub>	-ά(ς) ~ -αδ-	γαλατ-ά-ς γαλατ-άδ-ες	(γαλατ~ [γάλα])
ΒΑΣΗ <sub>[O]</sub>	-τζή(ς) ~ -τζηδ-	ταξι-τζή-ς	(ταξί)

		ταξι-τζήδ-ες	
ΒΑΣΗ <sub>[0]</sub>	-ιέρη(ς) ~ -ιέρηδ-	νταλικ-ιέρη-ς	(νταλικ~ [νταλικά])
		νταλικ-ιέρηδ-ες	

Παρατηρώντας τις λέξεις στα παραδείγματα του (60.), αξίζει να σημειώσουμε ότι τα παραγωγικά επιθήματα εμφανίζουν σημαντική αλλομορφία. Υπάρχει και μία μικρή ομάδα επιθημάτων που δεν εμφανίζει αλλομορφία, όπως τα επιθήματα *-εί(ο)* (γραφείο, ραφείο, φαρμακείο), *-ισμ(ος)* (λογισμός, βουδισμός, αμερικανισμός), *-αδορ(ος)* (κομπιναδόρος, αβανταδόρος, λουστραδόρος). Οι παράγωγες λέξεις αρσενικού γένους έχουν επιθήματα με αφαιρετική αλλομορφία, όπως *-τη(ς)* (χορευτής, αθλητής, πωλητής), *-ώνα(ς)* (στρατώνας, περιστερώνας, κοιτώνας), *-τήρα(ς)* (αλήτρας, βραστήρας, γεμιστήρας), εφόσον αποβάλλουν το τελικό φωνήεν (/i/ ή /a/). Με παρόμοια συχνότητα είναι και τα επιθήματα με την προσθετική αλλομορφία, όπως *-ά(ς)* (γαλατάς, παπλωματάς, σαπουνάς), *-τζή(ς)* (ταξιτζής, παπατζής, φαναρτζής), *-άρη(ς)* (βρωμιάρης, βαρκάρης, γελαδάρης), αφού εμφανίζουν αλλομορφα με την προσθήκη του συμφώνου /δ/.

Τα παράγωγα θηλυκά έχουν πάντα αλλομορφα της αφαιρετικής αλλομορφίας, όπως *-ρια* (καθηγήτρια, φοιτήτρια, μαθήτρια), *-ισσα* (φουρνάρισσα, αρχόντισσα, ορειβάτισσα), *-οσύνη* (καλοσύνη, μεγαλοσύνη, μετριοφροσύνη). Μοναδική εξαίρεση στην αλλομορφική κανονικότητα των θηλυκών επιθημάτων είναι το επίθημα *-ού* που λειτουργεί κατ' αντιστοιχία με το αντίστοιχο αρσενικό επίθημα επαγγελματιών *-α(ς)*, αφού συνδυάζεται με άλλα αρσενικά επιθήματα (π.χ. με το *-τζή(ς)*) και έχει αλλομορφα με το *-δ-*, π.χ. *γλωσσού - γλωσσούδ-ες*, *παπλωματού - παπλωματούδ-ες*, *καταφερτζού - καταφερτζούδες* κατά το *μαϊμού - μαϊμούδες*.

Τέλος, τα παραγωγικά επιθήματα ουδετέρου γένους εμφανίζουν σχετικά περιορισμένη αλλομορφία, αφού ο μοναδικός τύπος σχηματισμού αλλομόρφων είναι αυτός με την προσθήκη του /t/ στο επίθημα. Τέτοιες περιπτώσεις είναι τα παραγωγικά επιθήματα *-σιμο~ -σιματ-* (*βγάλσιμο, δέσιμο, γδάσιμο*) και *-μα~ -ματ-* (*κλάμα, άγγιγμα, πάτημα*). Σε πολλές λέξεις το επίθημα *-μα~ -ματ-* θεωρείται πλέον μέρος του θέματος, όπως *βήμα* (> βαίνω), *τάμα* (> τάζω), *άγημα* (> άγω), *σύνθημα* (> συνθέτω). Επίσης, τα επιθήματα τύπου  $X\Phi_{/i/}$ , όπως *-ιλίκι* (αρχηγιλίκι, παζαριλίκι, υπουργιλίκι), *-τήρι* (εργαστήρι, καπνιστήρι, κουρδιστήρι) στις περαιτέρω επιθηματοποιημένες περιπτώσεις εμφανίζουν αλλομορφα με την αποβολή του τελικού /i/.

## (61.) Παραγωγικά Επιθήματα ρηματικής κατηγορίας

ΒΑΣΗ <sub>[O]</sub>	-αρ(ω) ~ -αρι- ~ -αρισ-	σαлт-άρ-ω	(σάλτ-ο)
		σαлт-άρι-σ-α – σαлт-αρισ-μέν-ος	
ΒΑΣΗ <sub>[E/O]</sub>	-αιν(ω) ~ -αν-	ζεστ-αίν-ω	(ζεστ-ός)
		ζέστ-αν-α	
		ανασ-αίν-ω	(ανασ~ [ανάσα])
		ανάσ-αν-α	
ΒΑΣΗ <sub>[E/O/Επ]</sub>	-εύ(ω)	νοστιμ-εύ-ω	(νόστιμ-ος)
		φυτ-εύ-ω	(φυτ-ό)
		κοντ-εύ-ω	(κοντ-ά)
ΒΑΣΗ <sub>[E/O/Επ]</sub>	-ίζ(ω) ~ -ισ-	ασπρ-ίζ-ω	(άσπρ-ος)
		άσπρ-ισ-(σ)α	
		ψιθυρ-ίζ-ω	(ψίθυρ-ος)
		ψιθύρ-ισ-(σ)α	
		εγγ-ίζ-ω	(εγγ-ύς)
		έγγ-ισ-(σ)α	

Τα παραγωγικά επιθήματα ρηματικής κατηγορίας είναι πολύ περιορισμένα και επομένως είναι ιδιαίτερα εύκολο να ομαδοποιηθούν σύμφωνα με την εμφάνιση αλλομόρφων. Σχεδόν όλα τα ρηματικά παραγωγικά επιθήματα έχουν αλλομορφα που καλύπτουν τα περισσότερα είδη της αλλομορφίας. Δείγμα πολλαπλής αλλομορφίας είναι το επίθημα *-άρ(ω)~ -αρι-~ -αρισ-*, π.χ. *παρκάρω, σκανάρω, σερφάρω*, ενώ αντίθετα το *-ών(ω)* διαγράφει το τελικό σύμφωνο στο αλλομόρφό του, π.χ. *κλειδώνω, κουμπώνω, λαδώνω*. Επιπροσθέτως, έχουμε περιπτώσεις με ετεροιωτική αλλομορφία, όπως αυτή του επιθήματος *-αιν(ω)* με την αντικατάσταση του /e/ σε /a/, π.χ. *ακριβαίνω, θερμαίνω, ρυπαίνω*, παράλληλα με την περίπτωση του επιθήματος *-ίζ(ω)* που υπόκειται στην αντικαταστατική αλλομορφία με το /s/ να παίρνει τη θέση του /z/, π.χ. *ακοντίζω, αλατίζω, πλουτίζω*. Δεν εντοπίζεται, εντούτοις, κανένα επίθημα με συστηματική αλλομορφία που να έχει ομοιότητες με την αλλομορφία των ρημάτων τύπου *αγαπώ* και *αφαιρώ*.



## (62.) Παραγωγικά επιθήματα επιθετικής κατηγορίας

ΒΑΣΗ <sub>[E/O]</sub>	-ικ(ος)	απλο- <i>ικ</i> -ός	(απλ-ός <sup>4</sup> )
		ακτιν- <i>ικ</i> -ός	(ακτιν~ [ακτίνα])
ΒΑΣΗ <sub>[O]</sub>	-ένι(ος)	μαρμαρ- <i>ένι</i> -ος	(μάρμαρ-ο)
ΒΑΣΗ <sub>[O]</sub>	-ώδ(ης)	χειμαρρ- <i>ώδ</i> -ης	(χείμαρρ-ος)
ΒΑΣΗ <sub>[O/P]</sub>	-ατ(ος)	καρυδ- <i>άτ</i> -ος	(καρυδ~ [καρύδι])
		φευγ- <i>άτ</i> -ος	(φεύγ-ω)
ΒΑΣΗ <sub>[E/O]</sub>	-αι(ος)	ωρι- <i>αί</i> -ος	(ωρ~ [ώρα])
		τριτ- <i>αί</i> -ος	(τρίτ-ος)
ΒΑΣΗ <sub>[E]</sub>	-τέ(ος)	παραδο- <i>τέ</i> -ος	(παραδίν-ω)

Μετά την ομάδα των παραγωγικών επιθημάτων των ουσιαστικών, αυτή των επιθέτων είναι η δεύτερη κατηγορία σε πλήθος επιθημάτων. Αντίθετα με τα ονοματικά παραγωγικά επιθήματα, τα παραγωγικά επιθήματα επιθέτων εμφανίζουν ελάχιστη αλλομορφία. Παρόλο που αρκετά επιθήματα είναι κατάλοιπα της ΑΕ, όπως *-ώδ(ης)*, *-τέ(ος)*, *-αί(ος)*, *-ιμ(ος)*, δεν εμφανίζουν αλλομορφία, αφού δεν έγινε καμία επανάλυση των ορίων τους. Από τις ελάχιστες περιπτώσεις αλλομορφίας είναι το υποκοριστικό *-ούλη(ς)* ~ *-ούληδ-* (*μικρούλης*, *ατακτούλης*, *πατερούλης*) και το επίθημα *-ή(ς)* ~ *-Ø-* (*θαλασής*, *μολυβής*, *ουρανής*). Εξετάζοντας το επίθημα *-ώδ(ης)* (*πνευματώδης*, *θυελλώδης*, *αλματώδης*) δίνονται δύο τρόποι οριοθέτησης του επιθήματος: είτε θεωρούμε ότι η κατάληξη είναι *-ς* και κλίνεται σύμφωνα με την ΚΤ2, είτε είναι *-ης* και κλίνεται με βάση την ΚΤ9. Ο πρώτος διαχωρισμός παρουσιάζει αλλόμορφο στον πληθυντικό, ενώ ο δεύτερος όχι. Θεωρούμε ορθότερο τον δεύτερο διαχωρισμό, γιατί έτσι καλύπτουμε και τις περιπτώσεις του ουδετέρου γένους σε *-ώδ(ες)* (*πνευματώδες*, *θυελλώδες*, *αλματώδες*) έχοντας ως επιθηματική βάση το *-ωδ-* και μη προϋποθέτοντας το αλλόμορφο του αρσενικού, ή γενικότερα τον τύπο του αρσενικού/ θηλυκού για το σχηματισμό του ουδετέρου.

## (63.) Παραγωγικά επιθήματα επιρρηματικής κατηγορίας

ΒΑΣΗ <sub>[E]</sub>	-α	όμορφ- <i>α</i>	(όμορφ-ος)
---------------------	----	-----------------	------------

<sup>4</sup> Το συγκεκριμένο παράδειγμα δίνεται από τη Ράλλη (2005: 133). Εκτεταμένος έλεγχος των παράγωγων λέξεων σε *-ικ(ος)* έδειξε πως δεν υπάρχουν άλλα παραδείγματα με βάση επιθέτου, ενώ προκαλεί προβληματισμό η ύπαρξη του *-ο-* ανάμεσα στη βάση και το επίθημα. Μπορεί να συνδυαστεί μόνο σε ήδη επιθηματοποιημένες βάσεις επιθετικού υποκορισμού για το σχηματισμό του ουδετέρου επιθέτου, όπως π.χ. *μικρούλη-ς* > *μικρούλικος*, *μαυρούλη-ς* > *μαυρούλικος*.

ΒΑΣΗ <sub>[E]</sub>	-ως	διεθν-ώς	(διεθν-ής)
ΒΑΣΗ <sub>[E/O/Επ]</sub>	-θεν	άλλο-θεν θεό-θεν κάτω-θεν	(άλλ-ος) (θε-ος) (κάτω)
ΒΑΣΗ <sub>[O]</sub>	-ιστί	ελλην-ιστί	(έλληνα-ς)

Τα μοναδικά επιθήματα επιρρημάτων που είναι ενεργά στα Νέα Ελληνικά είναι το  $-α^5$  και το  $-ως$ . Τα υπόλοιπα είναι κατάλοιπα από τα Αρχαία Ελληνικά και συνήθως οι λέξεις που τα περιέχουν, έχουν απωλέσει τη μορφολογική τους διαφάνεια. Καθότι τα επιρρήματα ανήκουν στην κατηγορία των άκλιτων λέξεων, δεν συνδυάζονται με κλιτικά επιθήματα και δεν υπόκεινται σε περαιτέρω επιθηματοποίηση, ώστε να εμφανίσουν αλλόμορφα.

### 3.3. Ο σύνδεσμος της αλλομορφίας: από τα αλλόμορφα της κλίσης στα αλλόμορφα της παραγωγής

Σημαντικές έρευνες έχουν εκπονηθεί για να ορισθούν οι διαφορές μεταξύ της κλίσης και της παραγωγής, καθώς και ο καθορισμός σαφών κριτηρίων διαχωρισμού των δύο φαινομένων ανάλογα με το θεωρητικό υπόβαθρο του κάθε ερευνητή (για περισσότερες πληροφορίες, βλ. Bybee 1985, Ράλλη 1986, Dressler 1989, Anderson 1992, Plank 1994, Booij 1995, 2005, Payne 1997, Rice 2000, Stump 2001a, 2001b, μεταξύ άλλων). Στην παρούσα εργασία αποδεχόμαστε ότι η κλίση και η παραγωγή αποτελούν γειτνιάζοντα σημεία του μορφολογικού συνεχούς, των οποίων τα όρια είναι διακριτά και συχνά διαβαθμισμένα. Εμφανίζουν αρκετά κοινά φαινόμενα που, όπως για παράδειγμα την αλλομορφία. Συμφωνούμε με τον Booij (1997) πως η θεματική αλλομορφία δεν περιορίζεται στα κλιτικά συστήματα και παραδείγματα, αλλά διαδραματίζει σημαντικό ρόλο στην παραγωγή και τη σύνθεση. Ο Booij χρησιμοποιεί ως βάση τα Ολλανδικά και εντούτοις διατείνεται ότι ισχύει στις περισσότερες γλώσσες,

<sup>5</sup> Τα επίθετα σε  $-ως$  αποτελούν μία μικρή, περιορισμένη και καθόλου παραγωγική ομάδα που είναι κατάλοιπο του κλιτικού συστήματος των Αρχαίων Ελληνικών. Σύμφωνα με τον Τριανταφυλλίδη (1941, Λεξικό) τα συγκεκριμένα επίθετα σχηματίζουν επιρρήματα σε  $-έως$  (ταχύς > ταχέως, ευθύς > ευθέως, οξύς > οξέως) και σε  $-ιά$  (μακρύς > μακριά, βαθύς > βαθιά, παχύς > παχιά). Το επίθημα  $-ιά$  προέκυψε από συνίτηση για αποφυγή χασμωδίας από τον τύπο του πληθυντικού του ουδετέρου, όπου το μακρ-έα έγινε μακρ-ιά. Υποθέτουμε ότι λόγω επανάλυσης έχει δημιουργηθεί ένα νέο επίθημα επιρρηματος, το  $-ία$  για τα συγκεκριμένα επίθετα.

αφού υπογραμμίζει ότι η παρουσία της αλλομορφίας στην κλίση είναι σημαντική για την εμφάνιση στις υπόλοιπες διαδικασίες σχηματισμού λέξεων.

Στα Ελληνικά η εμφάνιση του φαινομένου της αλλομορφίας στις παραδειγματικές σχέσεις των λέξεων είναι συστηματική και η επιλογή του αλλομόρφου για κάθε μορφολογική διαδικασία είναι προβλέψιμη. Γνωρίζουμε ανά πάσα στιγμή, εφόσον έχουμε τη γνώση από την αλλομορφική συμπεριφορά ενός θέματος στην κλίση, αν θα εμφανίζει αλλόμορφο στην παραγωγή και αν ναι, τότε ποια θα είναι η μορφή του.

(64.)	α.	άνθρωπ-ος	(ΚΛ) <sup>6</sup>
		ανθρώπ-ιν(ος)	(ΠΡΓ)
		ανθρωπ-ότητα	(ΠΡΓ)
		ανθρωπ-ο-κυνηγός	(ΣΝΘ)
	β.	κύμα	(ΚΛ)
		κυματ-άκι	(ΠΡΓ)
		κυματ-ίζ(ω)	(ΠΡΓ)
		κυματ-ο-θραύστης	(ΣΝΘ)
	γ.	όμορφ-ος	(ΚΛ)
		ομορφ-ούλη(ς)	(ΠΡΓ)
		ομορφ-ότερ(ος)	(ΠΡΓ)*
		ομορφ-άντρας	(ΣΝΘ)
	δ.	πηδ(ώ)	(ΚΛ)
		πήδηγ-μα	(ΠΡΓ)
		πηδηχ-τής	(ΠΡΓ)
		πηδ-ο-γελώ	(ΣΝΘ)

Ελέγχοντας τα παραδείγματα του (64.) παρατηρούμε ότι στην παραγωγή και στη σύνθεση έχουμε μόνο ένα αλλόμορφο που συμμετέχει ως θεματική βάση και δεν εμφανίζονται όλα τα πιθανά αλλομορφικά θέματα που έχουμε στην κλίση. Αυτόν τον περιορισμό θα εξετάσουμε στην ακόλουθη υποενότητα, αν και υπάρχουν κάποιες ανακολουθίες στα ρηματικά θέματα. Η συγκεκριμένη συμπεριφορά υπάρχει και στη διαδικασία της σύνθεσης. Έτσι για παράδειγμα, το ουσιαστικό *κύμα* εμφανίζει στην κλίση τα μορφήματα *κυμα~ κυματ*, με το αλλόμορφο *κυματ~* να συμμετέχει στην

<sup>6</sup> ΚΛ= κλίση, ΠΡΓ= παραγωγή, ΣΝΘ= σύνθεση, ΠΡΓ\*= παραγωγή αλλά με επιφύλαξη.

παραγωγή (*κυματ-άκι*) και στη σύνθεση (*κυματ-ο-μορφή*). Έχουμε, δηλαδή, από τις δύο υποψήφιες μορφές *κυμα~* και *κυματ~* μόνο τη μία ως τελική επιλογή.

### 3.3.1. Ο περιορισμός της επιλογής μοναδικού αλλομόρφου

Η διαδικασία της παραγωγής εμπεριέχει αρχές και περιορισμούς για το είδος των θεμάτων των παράγωγων λέξεων που δέχονται ως βάση. Το πλέον συνηθισμένο είδος περιορισμού εισόδου (input constraint) είναι οι προϋποθέσεις μορφολογικής βάσης κατά την επιλογή τους από τα παραγωγικά επιθήματα, δηλαδή η βάση πρέπει να εντάσσεται στις επιτρεπόμενες κατηγορίες με τις οποίες μπορούν να συνδυαστούν. Συγκρίνοντας τις διαδικασίες της κλίσης, της παραγωγής και της σύνθεσης με κριτήριο την αλλομορφία παρατηρούμε διάφορες τάσεις στις γλώσσες. Υπάρχουν γλώσσες, όπως τα Γερμανικά, όπου τα αλλόμορφα της κλίσης εμφανίζονται και στις υπόλοιπες διαδικασίες (65.α), ενώ σε άλλες γλώσσες, όπως στα Ολλανδικά και στα Ελληνικά η συμπεριφορά είναι διαφορετική (65.β). Φαίνεται πως υπάρχει ένας περιορισμός που σε κάποιες γλώσσες αποτρέπει την εμφάνιση όλων των αλλομόρφων σε όλες τις διαδικασίες.

(65.) α. Γερμανικά

Land ~ Länder	Landkarte	-	Länderspiele
‘χώρα – χώρες’	‘χάρτης’		‘αγώνας μεταξύ χωρών’
Vater ~ Väter	Vaterland	-	Väteraufbruch
‘πατέρας – πατέρες’	‘πατρίδα’		‘πρωτοβουλία πατέρων’
Buch ~ Bücher	Buchhandlung	-	Bücherregal
‘βιβλίο – βιβλία’	‘βιβλιοπωλείο’		‘ράφι για βιβλία’

β. Ελληνικά

άνθρωπ(ος) – άνθρωπ(οι)	ανθρωπ-ισμ(ός)	ανθρώπ-ιν(ος)
παπά(ς) ~ παπάδ(ες)	παπαδ-άκι	παπαδ-ίστικ(ος)
ψυχή ~ ψυχ(ές)	ψυχ-άρα	ψυχ-ικός
βήμα ~ βήματ(α)	βηματ-ίζ(ω)	βηματ-άρα

Εξετάζοντας τα παραδείγματα των Γερμανικών παρατηρούμε ότι στη σύνθεση εμφανίζονται στη θέση του πρώτου συστατικού και τα δύο αλλόμορφα που απαντώνται στην κλίση, δηλαδή και το αλλόμορφο του ενικού και το αντίστοιχο του πληθυντικού. Αντιθέτως, τα παραδείγματα των Ελληνικών πιστοποιούν πως στη διαδικασία της

παραγωγής συμμετέχει μόνο ένα από τα αλλόμορφα που τυχόν διαθέτει ένα μόρφημα. Η μη-εμφάνιση και των δύο (ή τριών) αλλομόρφων ως βάση σε παράγωγη λέξη των Ελληνικών μπορεί να χρησιμοποιηθεί ως επιχείρημα υπέρ της άποψης ότι τα αλλόμορφα συνδέονται με την κλίση. Εγείρεται το ερώτημα *γιατί δεν έχουμε όλα τα πιθανά αλλόμορφα ως βάση στην παραγωγή*.

Σημειώνουμε ότι αναφερόμαστε στη διαδικασία της επιθηματοποίησης αποκλειστικά, καθώς μας ενδιαφέρουν οι μηχανισμοί παραγωγής που συντελούνται ανάμεσα στο θέμα και ένα παραγωγικό επίθημα. Στην περίπτωση της προθηματοποίησης, η επιλογή κατάλληλου αλλομόρφου για τη βάση δεν εξαρτάται ούτε συσχετίζεται με το πρόθημα. Επιπλέον, δεν λαμβάνουμε υπ' όψιν τις περιπτώσεις παράγωγων λέξεων που δεν περιλαμβάνουν παραγωγικό επίθημα, όπως τα παράγωγα από μετάπλαση και από αλλαγή τόνου.

Κατά την άποψή μας, μία απάντηση στην παραπάνω ερώτηση πρέπει να αναζητηθεί στη διαδικασία των περιορισμών. Η απουσία αλλομόρφων στην παραγωγή δεν είναι τυχαία και αποσπασματική και δεν είναι ανεξάρτητη από το περιβάλλον που εμφανίζεται. Προτείνουμε ότι στις ονοματικές και επιθετικές βάσεις παραγωγής επιτρέπεται η εμφάνιση ενός και *μόνου* αλλομόρφου λόγω της λειτουργίας ενός περιορισμού που λαμβάνει χώρα κατά τη διαμόρφωση εισόδου (input configuration) και εμποδίζει την εμφάνιση των άλλων αλλομόρφων ως βάση σε παράγωγη λέξη. Για αυτόν εισάγουμε τον όρο **περιορισμός της επιλογής μοναδικού αλλομόρφου**. Έχουμε ήδη δει ότι οι παράγωγες λέξεις της Ελληνικής υπόκεινται γενικά στον περιορισμό με τον αποκλεισμό κάποιων αλλομόρφων της κλίσης. Εφόσον οι περιορισμοί δεν μπορούν να εφαρμόζονται τυχαία, αλλά για συγκεκριμένη αιτία, προτείνουμε ότι *ο περιορισμός της επιλογής μοναδικού αλλομόρφου* διασφαλίζει τη σύνδεση μεταξύ των παράγωγων λέξεων.

#### (66.) Περιορισμός της επιλογής μοναδικού αλλομόρφου

*Ο συσχετισμός μεταξύ παραγωγών λέξεων με κοινή βάση διασφαλίζεται με την εμφάνιση μόνο μίας μορφής του λεξήματος – βάσης.*

Θεωρούμε πως η παρουσία αυτού του περιορισμού εντοπίζεται στις δομές όλων των ονοματικών και επιθετικών και δικαιολογείται για μορφολογικούς και φωνολογικούς λόγους κυρίως. Ελέγχοντας προσεχτικά τη φωνολογική δομή της βάσης μίας παράγωγης λέξης παρατηρείται ότι στατιστικά η ύπαρξη θέματος με τελικό χαρακτήρα σύμφωνο και όχι φωνήεν είναι το πλέον σύνηθες. Τα ουσιαστικά των Ελληνικών με

αλλομορφία σχηματίζουν δύο μορφήματα, ένα με ληκτικό χαρακτήρα φωνήεν και ένα με σύμφωνο, ενώ οι κλιτικές τάξεις χωρίς αλλομορφία έχουν στην πλειονότητά τους ονοματικά θέματα της μορφής XC με μοναδική συστηματική εξαίρεση την έκτη κλιτική τάξη (XV<sub>/i</sub>).<sup>7</sup> Παράλληλα και τα ρήματα που εμφανίζουν συστηματική αλλομορφία έχουν έναν τύπο με ληκτικό σύμφωνο και έναν με ληκτικό φωνήεν (X ~ XV): τα υπόλοιπα ρήματα μπορεί να έχουν όλους τους αλλομορφικούς τύπους με θεματικό χαρακτήρα σύμφωνο.

- (67.) α. **άνθρωπ-ος**  
           **δάσ-ος**  
           **μωρ-ό**
- β. **παπά-ς**           **παπάδ-ες**  
      **ποιητή-ς**       **ποιητ-ές**  
      **θάλασσα**       **θάλασσ-ες**  
      **κύμα**           **κύματ-α**
- γ. **αγαπ-ώ**           **αγάπη-σα**  
      **χτυπ-ώ**           **χτύπη-σα**  
      **διαρ-ώ**           **διαίρε-σα**
- δ. **παίρν-ω**       **πήρ-α**           **πάρ-ω**  
      **περν-ώ**       **πέρα-σα**  
      **τραβ-ώ**       **τράβηγ-σα (τράβηξα)**

Δομικά, πάνω από τα τέσσερα πέμπτα των παραγωγικών επιθημάτων που συνδυάζονται με ονοματικές και επιθετικές βάσεις αρχίζουν με φωνήεν. Για τη διατήρηση της ιδανικής συλλαβικής δομής ΣΦ<sup>8</sup>, υποθέτουμε ότι η βάση είναι μόρφημα με ληκτικό χαρακτήρα σε σύμφωνο. Επομένως, αν και δεν αποτελεί απαραίτητη προϋπόθεση, προκρίνεται το αλλόμορφο σε σύμφωνο. Για παράδειγμα το επίθημα –ιζ(ω), όταν ενώνεται με το λέξιμα ΚΥΜΑ που έχει τα αλλόμορφα *κυμα* ~ *κυματ* ενώνεται με το αλλόμορφο *κυματ*~, για να διατηρηθεί η ιδανική συλλαβική δομή. Η ιδιαίτερα παραγωγική υποκατηγορία των υποκοριστικών και μεγεθυντικών έχει όλα τα παραγωγικά επιθήματα να αρχίζουν από φωνήεν, όπως –*ακι*, –*αρα*, –*ουτσικος*, –*ουκλα*,

<sup>7</sup> Τα ουσιαστικά τύπου *παιδί* δεν έχουν κλιτικό αλλόμορφο, αλλά εμφανίζουν αλλόμορφο σε παραγωγή και σύνθεση. Αναλυτικότερα βλ. στην ενότητα 2.4.4.1.

<sup>8</sup> Για περισσότερες πληροφορίες: βλ. Nespor & Vogel 1986, Nespor 1999.

–ουλης, –ιτσα, –ώνα μεταξύ άλλων. Να υπογραμμιστεί ότι το παραγωγικό επίθημα δεν επιλέγει το κατάλληλο αλλόμορφο για τη βάση του παράγωγου, αλλά υπακούει στον περιορισμό της διασφάλισης της βέλτιστης συλλαβικής δομής. Εξάλλου υφίστανται περιπτώσεις με τη βέλτιστη συλλαβική δομή να καταστρατηγείται λόγω της απουσίας εναλλακτικού μορφήματος, όπως π.χ. *λα(ός) > λα-ϊκ(ος)*, *μυστήρι(ο) > μυστηρι-ακ(ός)*, *ράβ(ω) > ράφ-τη(ς)*. Συνοψίζοντας, δηλώνουμε πως δεν εντοπίζονται συστηματικά κάποια φωνολογικά χαρακτηριστικά ή κριτήρια για την τελική επιλογή του αλλομόρφου.

Στο δεξί μέρος ενός επιθηματοποιημένου παράγωγου βρίσκεται η κεφαλή, αφού στην παραγωγή των Ελληνικών εφαρμόζεται ο **νόμος της δεξιόστροφης κεφαλής** (righthand head rule) του Williams (1981).<sup>9</sup> Στις δομές των παραγώγων η ιδιότητα της κεφαλής αποδίδεται στα παραγωγικά επιθήματα. Παραδείγματος χάριν, στα Ελληνικά το αλλόμορφο *κυματ~* συμμετέχει στην παραγωγή (και στη σύνθεση) σε θέση μη-κεφαλής και αυτή η πληροφορία καταχωρείται στο πεδίο πληροφοριών μορφολογικού περιβάλλοντος.

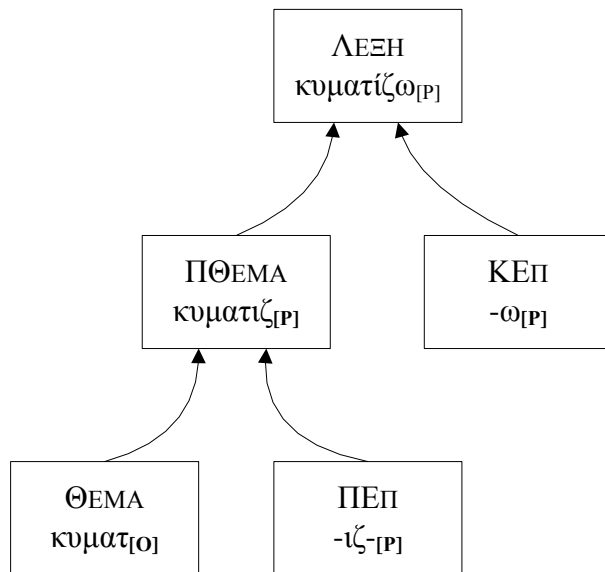
Υποστηρίζουμε ότι στα Ελληνικά κατά τη διαδικασία της παραγωγής για όλες τις ονοματικές και επιθετικές βάσεις *μόνο ένα* αλλόμορφο λεξήματος μπορεί να εμφανιστεί σε θέση *μη-κεφαλής*<sup>10</sup>. Ο περιορισμός της επιλογής μοναδικού αλλομόρφου εφαρμόζεται και για όλες τις βάσεις των ρημάτων της ΚΤ2. Αναφερόμαστε αποκλειστικά στα επιθηματοποιημένα παράγωγα, όπου η θέση της μη κεφαλής βρίσκεται πάντα στο αριστερό μέρος της παράγωγης λέξης. Τονίζουμε ότι το μορφολογικό χαρακτηριστικό της μη-κεφαλής εισάγεται για δύο λόγους: (α) γιατί ο περιορισμός της ιδανικής συλλαβικής δομής μπορεί να καταστρατηγηθεί και (β) γιατί είναι απαραίτητο στην υπολογιστική ανάλυση η χρήση της μη-κεφαλής για την αποφυγή δημιουργίας αντιγραμματικών λέξεων.<sup>11</sup>

<sup>9</sup> Για την εφαρμογή του νόμου στα Ελληνικά, πρβλ. Ralli (1988). Η καθολικότητα του συγκεκριμένου νόμου αμφισβητείται έντονα από αρκετούς γλωσσολόγους (βλ. Joseph & Wallace 1984).

<sup>10</sup> Σε μία μορφολογική διαδικασία, **κεφαλή** (head) αποκαλείται το συστατικό που καθορίζει τη μορφολογική κατηγορία του σχηματισμού. Υπάρχει και η σημασιολογική κεφαλή που δεν ταυτίζεται συνήθως με τη μορφολογική.

<sup>11</sup> Αυτό το χαρακτηριστικό είναι απαραίτητο κυρίως για τα επιθήματα που όταν αλλάζουν θέση (από θέση κεφαλής σε θέση μη-κεφαλής) πρέπει να προκρίνουν το κατάλληλο αλλόμορφο. Για περισσότερα στο κείμενο που ακολουθεί.

(68.)



Το σύνολο των παράγωγων λέξεων της Ελληνικής αποτελείται από ένα θέμα που βρίσκεται σε θέση μη-κεφαλής. Στο (68.) το θέμα *κυματ~* είναι σε θέση μη-κεφαλής και ενώνεται με το ρηματικό παραγωγικό επίθημα *-ιζ-* που είναι σε θέση κεφαλής. Στα μορφήματα της μη-κεφαλής ενός παραγώγου που εμφανίζουν αλλομορφία δεν επιτρέπεται να εμφανίσουν όλες τις πιθανές τους μορφές στη συγκεκριμένη θέση. Επομένως κάθε αλλόμορφο θέματος λαμβάνει αυτή την πληροφορία περιβάλλοντος από τους κανόνες χρήσης αλλομορφίας, ώστε να αποκλείονται οι πιθανότητες εμφάνισης αντιγραμματικών λέξεων.

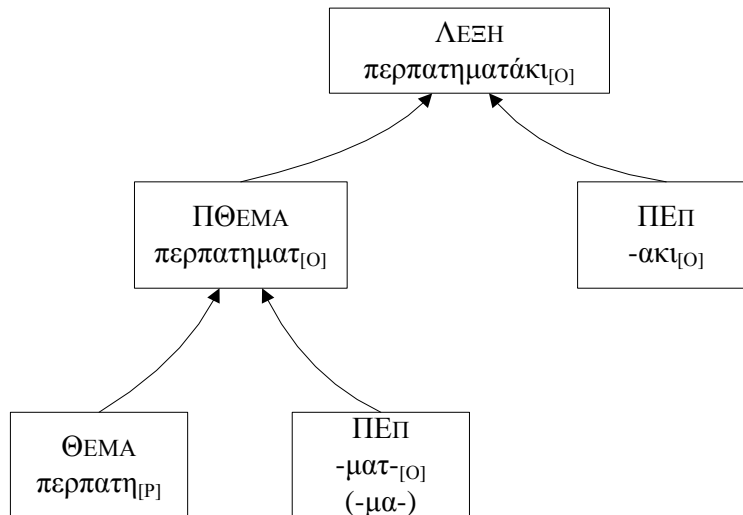
Από τον συγκεκριμένο περιορισμό δεν εξαιρούνται και τα παραγωγικά επιθήματα, αφού παρατηρούμε πως τα επιθήματα «*υιοθετούν*» τις ίδιες αλλομορφικές συμπεριφορές με τα θέματα και στο συγκεκριμένο περιορισμό. Αναμένουμε στο λεξικό να μοιράζονται τις ίδιες πληροφορίες περιβάλλοντος για τα αλλόμορφα τους. Πιο συγκεκριμένα, όταν ένα παραγωγικό επίθημα, το οποίο πρόκειται να ενωθεί με κάποιο θέμα ή ακόμα και με παράγωγο θέμα<sup>12</sup> και ακολουθείται από ένα επιπλέον παραγωγικό επίθημα, τότε κι αυτό θα βρεθεί στη δεύτερη διαδικασία παραγωγής σε θέση μη-κεφαλής και προάγει μόνο ένα αλλόμορφο. Επί της ουσίας, δεν είναι το παραγωγικό επίθημα που εμφανίζει μόνο ένα αλλόμορφο, αλλά το παράγωγο θέμα που κληρονομεί τα χαρακτηριστικά του επιθήματος. Για παράδειγμα, η λέξη *περπατηματάκι* αποτελείται από το αλλόμορφο του ρήματος *περπατώ* (*περπάτη~*), το παραγωγικό επίθημα *-μα*

<sup>12</sup> Παράγωγο θέμα ονομάζεται μία βάση που έχει ενωθεί με τουλάχιστον ένα παραγωγικό επίθημα, π.χ. *ανθρωπιν-*, *χορευ-*, *ομορφοτερ-*.



(ματ~), το παραγωγικό επίθημα *-ακι* και ένα μηδενικό κλιτικό επίθημα *Ο*: από το ρήμα *περπατώ* έχουμε το *περπάτημα* και στη συνέχεια παράγουμε το *περπατηματάκι*.

(69.)



Δομικά, οι συγκεκριμένοι τύπου παράγωγες λέξεις αποτελούν το συνδυασμό παράγωγου θέματος και παραγωγικού επιθήματος. Σε δομές όπως αυτή του (69.) το παράγωγο θέμα *περπατημ-* είναι σε θέση μη-κεφαλής, καθώς σύμφωνα με το νόμο δεξιόστροφης κεφαλής, το παραγωγικό επίθημα *-ακι* είναι η κεφαλή. Στο πρώτο στάδιο της παραγωγής το παραγωγικό επίθημα *-μα* είναι σε θέση κεφαλής και επομένως και τα δύο αλλομορφά του είναι διαθέσιμα. Εντούτοις, επειδή η διαδικασία της παραγωγής επαναλαμβάνεται στο δεύτερο στάδιο, τότε μόνο το αλλομορφο *-περπατηματ~* με την πληροφορία περιβάλλοντος<sub>[ΘΕΣΗ ΜΗ-ΚΕΦΑΛΗΣ]</sub> θα πραγματοποιηθεί. Στην ουσία παράγονται δύο θέματα, *περπατημα* και *περπατηματ*, αλλά χρησιμοποιείται μόνο το δεύτερο ως ΠΘέμα σύμφωνα με τις απαιτήσεις του περιορισμού. Στην περίπτωση που η λέξη *περπατηματάκι* επιδέχεται περαιτέρω επιθηματοποίηση, τότε και το επίθημα *-άκι* θα υιοθετήσει τη συμπεριφορά του προηγούμενου επιθήματος. Θα εφαρμοστεί ο περιορισμός της επιλογής μοναδικού αλλομορφου και το νέο παράγωγο θέμα θα έχει κληρονομήσει τα κατάλληλα χαρακτηριστικά.

Το πλεονέκτημα που μας παρέχει ο περιορισμός της επιλογής μοναδικού αλλομόρφου είναι η προβλεψιμότητα της τελικής επιλογής του κατάλληλου αλλομόρφου. Η διαδικασία της αλλομορφίας είναι αυθαίρετη και απρόβλεπτη όσον αφορά στην ύπαρξη του αλλομόρφου και επειδή τα αλλομορφα είναι *απρόβλεπτα* ως προς τη μορφή τους καταχωρούνται στο Λεξικό μαζί με τις υπόλοιπες πληροφορίες του λήμματος. Παρόλα αυτά ο περιορισμός επιτρέπει τη χρήση μόνου ενός αλλομόρφου

στο θέμα, ενώ παράλληλα προβλέπει και ποιο είναι το αλλόμορφο που θα συμμετάσχει στη διαδικασία. Επομένως, με τον συγκεκριμένο περιορισμό μπορούμε να προβλέψουμε τα περιβάλλοντα εμφάνισης των αλλομόρφων στην παραγωγή και για αυτό καταχωρούνται στους κανόνες χρήσης αλλομορφίας εκτός Λεξικού. Η πρόβλεψη περιβάλλοντος εμφάνισης των αλλομόρφων είναι εξαιρετικά σημαντική για την υπολογιστική επεξεργασία του φαινομένου. Όπως θα αναλύσουμε σε επόμενο κεφάλαιο, η αλλομορφία αποτελεί τροχοπέδη σε μορφολογικούς αναλυτές τόσο στο επίπεδο της ανάλυσης όσο και στο αντίστοιχο του σχηματισμού. Ο παραπάνω περιορισμός μας επιτρέπει υπολογιστικά να κάνουμε με αυτοματοποιημένες διαδικασίες την καταχώρηση των αλλομόρφων, καθώς και την αυτόματη επιλογή του κατάλληλου αλλόμορφου για τον σχηματισμό παραγώγων, τόσο αυτών που καταχωρούνται στο λεξικό όσο και αυτών που δημιουργούνται εκτός λεξικού.

Αυτός ο περιορισμός εφαρμόζεται κυρίως σε ονοματικές και επιθετικές βάσεις παραγώγων, αλλά παρατηρείται και μία συστηματική παρουσία και σε ρηματικά αλλόμορφα. Πιο συγκεκριμένα, για τα ουσιαστικά και τα επίθετα περιμένουμε ότι τα αλλόμορφα που δεν συμμετέχουν στην ονομαστική ενικού, είναι αυτά που θα εμφανιστούν στη διαδικασία της παραγωγής. Επίσης, τα ρήματα με συστηματική αλλομορφία σε X[ $\{H/E\}$ ] εμφανίζουν το αλλόμορφο του αορίστου στις παράγωγες λέξεις με ρηματική βάση.

Σχηματικά ο περιορισμός της επιλογής μοναδικού αλλομόρφου αναπαρίσταται στον ακόλουθο πίνακα, όπου έχουμε τρεις κατηγορίες μορφημάτων: (α.) τα μορφήματα με δύο αλλόμορφα, το βασικό και το μη-προκαθορισμένο, όπως π.χ. *θαλασσα~ θαλασσ*, *καφε~ καφεδ*, *αγαπ(α)~ αγαπη*, (β.) τα μορφήματα χωρίς αλλομορφία, όπως π.χ. *ανθρωπ-*, *διεθν-* και (γ.) τα μορφήματα που έχουν παραπάνω δύο αλλόμορφα, όπως π.χ. *πολη~ πολε~ πολ*, *μπακαλη~ μπακαληδ~ μπακαλ*, *παιρν~ παρ~ πηρ*. Τα μορφήματα της πρώτης ομάδας συμμετέχουν στην παραγωγή πάντα με το δεύτερο αλλόμορφο, ενώ η ομάδα των μορφημάτων που δεν έχει αλλόμορφα συμμετέχει με τη μοναδική μορφή που διαθέτει. Τέλος, η τρίτη ομάδα μορφημάτων με τα τρία αλλόμορφα συμμετέχει με το τρίτο αλλόμορφο. Υπολογιστικά, καθορίζουμε ως τρίτο αλλόμορφο αυτό που εμφανίζεται στις διαδικασίες παραγωγής (και σύνθεσης), δηλαδή τα αλλόμορφα τύπου *πολ~* και *μπακαλ~* για τις συγκεκριμένες ομάδες ονομάτων. Στον ακόλουθο πίνακα εντάσσεται και η σύνθεση αλλά μόνο για τις περιπτώσεις των πρώτων συνθετικών και

για την εφαρμογή του περιορισμού *μόνο* για τα ονοματικά και επιθετικά μορφήματα, αφού για τα ρηματικά θέματα ισχύουν διαφορετικές.

	ΜΟΡΦΗΜΑ	ΑΛΛΟΜΟΡΦΟ	ΜΟΡΦΗΜΑ	ΜΟΡΦΗΜΑ	ΑΛΛΟΜΟΡΦΑ
ΚΛΙΣΗ	↓	(↓) <sup>13</sup>	↓	↓	↓ ↓
ΠΑΡΑΓΩΓΗ	∅	↓	↓	∅	∅ ↓
ΣΥΝΘΕΣΗ	∅	↓	↓	∅	∅ ↓

Πίνακας 11: Γενικός πίνακας του περιορισμού επιλογής μοναδικού αλλομόρφου για ονοματικά θέματα και επιθήματα.

Υποστηρίζουμε ότι ο παραπάνω περιορισμός εμφανίζεται στα Ελληνικά χωρίς εξαιρέσεις στις ονοματικές και επιθετικές βάσεις. Ειδικότερα στην παραγωγή δεν εντοπίζεται κανένα αντιπαράδειγμα που να αντιβαίνει τα όσα ορίζει ο περιορισμός. Εντούτοις, ο Drachman (2006) παρουσιάζει κάποια παραδείγματα που φαίνεται να αποτελούν εξαιρέσεις στον υπό παρουσίαση περιορισμό. Θεωρεί ότι όλα τα αλλόμορφα είναι πιθανοί συμμετέχοντες σε θέση μη-κεφαλής ενός παράγωγου και σε θέση πρώτου συνθετικού ενός σύνθετου (70.α).

- (70.) α. κρέ-ας/ κρέατ-α κρεατ-ερό, κρεατ-ίλα, κρεατ-ινό  
 κρε-ο-πώλης, κρε-ο-φάγος, κρε-ο-κοφ-τήρας  
 κρεατ-ο-πωλείο, κρεατ-ο-σανίδα, κρεατ-ό-πιτα
- αίμ-α/ αίματ-α αιματ-άκι, αιμάτ-ωμα  
 αιμ-ο-σφαίρια, αιμ-ο-δοσία, αιμ-ο-ρραγία  
 αιματ-ο-βαμμένος, αιματ-ο-κύλισμα, γλυκ-ο-αίματ-ος
- β. (η)μέρα ημερίσιος (\*μερίσιος), ημερομίσθιο (\*μερομίσθιο)  
 μερόνυχτα
- (ε)νοίκι(ο) νοικιάζω, ενοικιαστής (\*νοικιαστής), νοικιάρης
- (ο)μιλώ μίλησε, ομιλητής (\*μιλητής)

(από Drachman 2006: 14, 19)

<sup>13</sup> Να επισημανθεί ότι υπάρχει μία μικρή ομάδα λεξημάτων που το αλλόμορφο της συμμετέχει μόνο στην παραγωγή και στη σύνθεση, όπως παρουσιάστηκε από τον περιορισμό του κλιτικού αλλομόρφου.

Για τις μεμονωμένες παραβάσεις του περιορισμού στη σύνθεση αναφέρουμε ότι τα δάνεια έχουν διαφορετική συμπεριφορά και συχνά παρεκκλίνουν από τους κανόνες, τις αρχές και τους περιορισμούς της γλώσσας. Για την περίπτωση του λεξήματος ΑΙΜΑ με αλλόμορφα *αιμα~ αιματ~ αιμ* έχουμε ήδη αναφερθεί σε παλαιότερη μελέτη (Καρασίμος 2001), όπου σημειώνουμε την ιδιαίτερη συμπεριφορά των συνθέτων του λεξήματος ΑΙΜΑ. Οι παράγωγες λέξεις του λεξήματος ΑΙΜΑ έχουν ως μοναδική βάση το αλλόμορφο *αιματ*, καθώς αυτό προβλέπεται από τον περιορισμό της επιλογής μοναδικού αλλομόρφου για τα ουσιαστικά της ΚΤ8. Στη σύνθεση υπάρχουν δύο ομάδες συνθέτων με το λέξιμα *αίμα*: (α.) λέξεις με *α'* συνθετικό το αίμα (*αιμ~*), οι οποίες προέρχονται από τα Αρχαία Ελληνικά και τα Γαλλικά μέσω των Διεθνών Ελληνικών και (γ.) αυτές με *α'* συνθετικό το *αίμα* (*αιματ~*) που προέρχονται από τα Αγγλικά ή σχηματίστηκαν κανονικά στα Νέα Ελληνικά. Επιπροσθέτως, διαφωνούμε με τον χαρακτηρισμό της συγκεκριμένης ομάδας λέξεων (70.β) ως αλλόμορφα, γιατί η αλλαγή του θέματος με τη διαγραφή του αρχικού φωνήεντος υπόκειται σε φωνολογικό κανόνα και δεν αποτελεί αλλομορφία σύμφωνα με τον ορισμό μας.

Κλείνοντας, να επισημανθεί πως θεωρούμε ότι ο περιορισμός επιλογής του μοναδικού αλλομόρφου εφαρμόζεται καθολικά στη διαδικασία της παραγωγής στην Ελληνική γλώσσα, δεν έχει εξαιρέσεις (με μικρή επιφύλαξη για τις Νεοελληνικές διαλέκτους) και επιτρέπει την πρόβλεψη του μορφήματος της βάσης. Σε γλώσσες, όπως τα Γερμανικά (Lieber 1982), δεν υφίσταται ο συγκεκριμένος περιορισμός, αλλά θεωρούμε ότι αξίζει να μελετηθεί εάν υφίσταται ο περιορισμός και σε άλλες γλώσσες με έντονη αλλομορφία.

### 3.4. Οι περιπτώσεις των αλλοθημάτων

Σε κάποιες μελέτες αναφέρονται ελάχιστα ζεύγη επιθημάτων ή προθημάτων, τα οποία χωρίζονται σε δύο ομάδες: (α) τα ζεύγη επιθημάτων που έχουν διαφορετική μορφή, αλλά φέρουν σχεδόν τα ίδια σημασιολογικά χαρακτηριστικά, συνδυάζονται με βάσεις ίδιου τύπου και εμφανίζονται συνήθως σε ίδια μορφολογικά περιβάλλοντα και (β) τα ζεύγη επιθημάτων που συνδυάζονται με βάσεις ίδιου τύπου, αλλά εμφανίζονται πάντα σε διαφορετικά μορφολογικά περιβάλλοντα λόγω κάποιων φωνολογικών ιδιοτήτων. Σε σχέση με τα ζεύγη επιθημάτων της πρώτης ομάδας, η Ράλλη (2005) υπογραμμίζει ότι έχουν τις ίδιες πληροφορίες και επομένως μπορούν να συνδυαστούν με τα ίδια θέματα. Παραδείγματος χάριν, οι ομιλητές της Νέας Ελληνικής

εξακολουθούν να σχηματίζουν επιρρηματικούς τύπους σε *-ως* με βάσεις που δεν είναι χαρακτηρισμένες ως [+λαϊκές], ενώ ταυτόχρονα χρησιμοποιούν και το νεοελληνικό επίθημα σε *-α*, όπως φαίνεται στα παραδείγματα (71.).

(71.)	ΠΕΠ <sub>Επιρ</sub> [-ως]	ΠΕΠ <sub>Επιρ</sub> [-α]		ΘΕΜΑ <sub>Επιθ</sub>
	σπανίως	σπάνια	<	σπάνι-ος
	βεβαίως	βέβαια	<	βέβαι-ος
	καλώς	καλά	<	καλ-ός
	κακώς	κακά	<	κακ-ός
	παραλλήλως	παράλληλα	<	παράλληλ-ος
	γενικώς	γενικά	<	γενικ-ός

Τα ζεύγη επιρρηματικών λέξεων σε *-ως* και *-α* είναι πάρα πολλά στα Ελληνικά και είναι συστηματική η παρουσία τους όσον αφορά τη χρήση τους. Ωστόσο, όπως τονίζει και η Ράλλη (2005), οι ομιλητές δεν αποδέχονται λέξεις όπως *\*φαγαδίκως* (αντί *φαγάδικα*) και *\*χορατατζιδίκως* (αντί *χορατατζίδικα*) καθότι αυτές οι βάσεις δεν είναι μαρκαρισμένες ως λόγιες (ή αλλιώς ως *-λαϊκές*). Στην παρούσα ενότητα δεν θα μελετήσουμε περιπτώσεις τέτοιων ζευγών, καθότι τα ζεύγη της δεύτερης ομάδας φαίνεται να φέρουν κάποιες από τις ιδιότητες που απαντώνται στα αλλόμορφα, αλλά αποτελούν ξεκάθαρες περιπτώσεις διαφορετικών μορφημάτων. Είναι ξεκάθαρο για τη δική μας θεωρητική προσέγγιση ότι τα παραγωγικά επιρρηματικά επιθήματα *-ως* και *-α* δεν είναι αλλόμορφα του ίδιου μορφήματος.

Μια διαφορετική προσέγγιση των ελαχίστων ζευγών επιθημάτων της δεύτερης ομάδας για τα Ελληνικά έγινε από τους Drachman, Kager & Malikouti-Drachman (1994) και Drachman (2006). Οι παραπάνω συγγραφείς εξετάζουν τέσσερα τέτοια ζεύγη (6.), τα οποία δρουν ανταγωνιστικά ως προς τα περιβάλλοντα συμμετοχής. Μελετούν ζεύγη των επιθημάτων με ίδια ή παρόμοια μορφολογικά και σημασιολογικά χαρακτηριστικά και τα κατηγοριοποιούν ως αλλόμορφα, όπως *-ιμο*<sup>14</sup> vs. *-μα*, *-ινος* vs. *-ενιος*.

(6.)	vréks-imo (vréx-o)	vs.	skúpiz-ma (skúpiz-o)	/-imo/ ~ /-ma/
	pétr-ino (pétr-a)	vs.	moliv-énjo (móliv-os)	/-ino/ ~ /-enjo/
	kléf-tis (klév-o)	vs.	kaθaris-tís (kaθariz-o)	/-tis/ ~ /-tís/

<sup>14</sup> Σύμφωνα με το μοντέλο της Ράλλη (2005) και τη δική μας ανάλυση το συγκεκριμένο παραγωγικό επίθημα είναι *-σίμο* και όχι *\*-ίμο*.

ník-o (ník-os) vs. filíp-e (filíp-os) /-o/ ~ /-e/ (ΚΛ)

(από Drachman, Kager & Malikouti-Drachman 1995: 2-3)

Για τους Drachman, Kager & Malikouti-Drachman τα ζεύγη /-imo/ και /-ma/, /-inos/ και /-enios/ είναι αλλόμορφα, όπως και τα υπόλοιπα ζεύγη του (6.), αφού δεν μπορούν να εμφανιστούν στο ίδιο μορφολογικό περιβάλλον και η κατανομή τους είναι συμπληρωματική, π.χ. *πιάσιμο* και όχι *\*πιάσμα*, *πέτρινος* και όχι *\*πετρένιος*. Σημειώνουμε, όμως, μερικά αντιπαραδείγματα, όπως *δέσιμο* και *δέμα*, *κλάψιμο* και *κλάμα*, *μαρμάρινος* και *μαρμαρένιος*. Ακολουθώντας τις αρχές της θεωρίας του Βέλτιστου και της Προσωδιακής Μορφολογίας χρησιμοποιούν ως κύρια κριτήρια τον αριθμό συλλαβών και τη θέση του τόνου για την τελική επιλογή του ενός επιθήματος έναντι του άλλου. Αποδέχονται ότι είναι αλλόμορφα που κυβερνώνται αποκλειστικά από φωνολογικά κριτήρια και μπορούν να προβλέψουν με ποια θέματα θα συνδυαστούν τα επιθήματα. Τα φωνολογικά χαρακτηριστικά της βάσης διαδραματίζουν σημαντικό ρόλο για την επιλογή κάποιου επιθήματος έναντι κάποιου άλλου ανταγωνιστικού.

Στην παρούσα ενότητα θα ασχοληθούμε με ζεύγη επιθημάτων, όπως αυτά του παραδείγματος (6.), τα οποία δεν αποδεχόμαστε ως αλλόμορφα. Δηλαδή, για παράδειγμα, το επίθημα /-(s)imo/ δεν είναι αλλόμορφο του επιθήματος /-ma/ ή το αντίστροφο. Δεν τα εντάσσουμε στην κατηγορία των αλλομόρφων, επειδή δεν εφαρμόζεται ο ορισμός της αλλομορφίας και αποτελούν δύο διαφορετικά λήμματα του Λεξικού. Επίσης τα επιθήματα των παραδειγμάτων (71.) δεν είναι αλλόμορφα<sup>15</sup>, αφού εμφανίζονται συστηματικά με τα ίδια θέματα-βάσεις. Αποτελούν διαφορετικά λήμματα που φέρουν σχεδόν παρόμοιες μορφολογικές ιδιότητες, αλλά προέρχονται από διαφορετικές περιόδους της γλώσσας.

Εξετάζοντας τα παραδείγματα του (6.), παρατηρούμε κατ' αρχάς ότι σε κάποιες περιπτώσεις τα υπό εξέταση επιθήματα εμφανίζουν αλλόμορφα (σύμφωνα με τον ορισμό μας), όπως για παράδειγμα το επίθημα [-σιμο]: *-σιμο~ -σιματ-* || το [-μα]: *-μα~ -ματ-* και το [-τη(ς)]: *-τη~ -τ-*. Οι Drachman, Kager & Drachman-Malikouti δεν κάνουν καμία αναφορά σε αυτές τις μορφές ως αλλόμορφα, αφού δεν τις αναφέρουν καθόλου. Εγείρεται το ερώτημα, εάν το [-(σ)ιμο] είναι αλλόμορφο του [-μα] (-(σ)ιμο~

<sup>15</sup> Η Ράλλη (2005) δεν τα θεωρεί αλλόμορφα, αλλά διαφορετικά επιθήματα.

-μα) ή το αντίστροφο (-μα~ -(σ)ιμο), τότε ως τι πρέπει να καταχωρηθούν τα πραγματικά αλλόμορφα των επιθημάτων; Ξεκαθαρίζουμε λοιπόν ότι τα συγκεκριμένα ζεύγη δεν είναι αλλόμορφα, αφού είναι σαφές πως τα συγκεκριμένα ζεύγη επιθημάτων δεν αποτελούν διαφορετικές εκφάνσεις του ίδιου μορφήματος, αλλά δύο διαφορετικά επιθήματα που μοιράζονται παρόμοια μορφολογικά και σημασιολογικά χαρακτηριστικά. Αντίστοιχα παραδείγματα παραθέτει και η Ράλλη (2005: 156-157), όπου δίνει ζεύγη, όπως -άδα και -οσύνη (αγριάδα, πρασινάδα, δικαιοσύνη, ιεροσύνη), τα οποία φυσικά δεν τα αναφέρει ως αλλόμορφα. Θεωρεί πως η χρήση του ενός μπλοκάρει τη χρήση του άλλου, αφού η δυνατότητα συνδυασμού θεμάτων με το επίθημα -οσύνη παρεμποδίζει την επέκταση και εμφάνιση του -άδα με τα ίδια θέματα.<sup>16</sup>

Οι Drachman, Kager & Drachman-Malikouti (1995) δίνουν ελάχιστα αντιπαραδείγματα μόνο για το ζεύγος επιθημάτων /-inos/ vs. /-enios/, τα οποία σύμφωνα με τους συγγραφείς δικαιολογούνται από την αλλαγή της σειράς εφαρμογής συγκεκριμένων περιορισμών της Θεωρίας του Βέλτιστου. Εντούτοις, σημειώνουμε ότι τα αντιπαραδείγματά (72.) εμφανίζονται σε όλα τα ζεύγη και αντιβαίνουν στη συμπεριφορά των αλλομόρφων να μην συνδυάζονται με ίδιες βάσεις και να έχουν συστηματικές ελεύθερες ποικιλίες.

(72.)	α.	δέσιμο    δέμα	<	δένω
		κλάψιμο    κλάμα	<	κλαίω
		λιώσιμο    λιώμα	<	λιώνω
		πλέξιμο    πλέγμα	<	πλέκω
		πλύσιμο    πλύμα	<	πλένω
	β.	μαρμάρινος    μαρμαρένιος	<	μάρμαρο
		βελούδινος    βελουδένιος	<	βελούδο
		λουλούδινος    λουλουδένιος	<	λουλούδι
		συρμάτινος    συρματένιος	<	σύρμα
		χωμάτινος    χωματένιος	<	χώμα

<sup>16</sup> Είναι αξιοσημείωτο το γεγονός ότι τα δύο ανταγωνιστικά μορφήματα δεν προέρχονται από την ίδια χρονική περίοδο (π.χ. το -άδα που είναι βενετσιάνικο και το -οσύνη που είναι αρχαίο). Χρηζουν συστηματικότερης ανάλυσης οι λόγοι που εμπόδισαν την παραγωγή ουσιαστικών όσο το ένα από τα δύο δεν υπήρχε ακόμα.

Θεωρούμε πως ζεύγη επιθημάτων που φέρουν παρόμοια μορφολογικά και σημασιολογικά χαρακτηριστικά χρήζουν ιδιαίτερης εξέτασης. Για τις συγκεκριμένες περιπτώσεις ζευγών επιθημάτων εισάγουμε τον όρο του αλλοθήματος. Ορίζουμε ως **αλλοθήματα** δύο προσφύματα τα οποία αποτελούν διαφορετικές καταχωρήσεις μορφήματος στο Λεξικό, διαφέρουν φωνολογικά στη μορφή, αλλά έχουν παρόμοιες (σχεδόν ίδιες) σημασιολογικές ιδιότητες και εμφανίζονται σε διαφορετικά μορφολογικά περιβάλλοντα, κυρίως λόγω μορφολογικών – και συχνά φωνολογικών – περιορισμών. Τα αλλοθήματα είναι κυρίως παραγωγικά (και κλιτικά) επιθήματα που βρίσκονται σε συμπληρωματική κατανομή, δηλαδή όπου συμμετέχει το ένα απουσιάζει το άλλο. Στον ορισμό φέρουν το συστατικό ΑΛΛΟ-, όπως οι όροι ΑΛΛΟφωνο και ΑΛΛΟμορφο, για να δηλωθεί η συμπεριφορά της συμπληρωματικής κατανομής. Επίσης θεωρούμε ότι είναι πολύ σπάνιες και ειδικές περιπτώσεις, όπου δύο αλλοθήματα μπορούν να μοιράζονται το ίδιο μορφολογικό περιβάλλον, όπως στα παραδείγματα (72.) τα ζεύγη *-μα vs. -σιμο* και *-ινος vs. -ένιος*. Όπως θα δούμε ακολούθως, οι παραβιάσεις αυτές οφείλονται κυρίως σε διαφορετικές χρονικές φάσεις της γλώσσας, καθώς και σε διαλεκτικές διαφοροποιήσεις που έχουν ενσωματωθεί στην ΚΝΕ.

Οι Μαλικούτη-Drachman & Drachman (1989) και Drachman, Kager & Malikouti-Drachman (1995) επισημαίνουν φωνολογικούς περιορισμούς στη διαδικασία της παραγωγής για την επιλογή επιθημάτων στις περιπτώσεις των αλλοθημάτων, όπως για παράδειγμα τον αριθμό συλλαβών του θέματος. Στον αντίποδα, από τους μορφολογικούς περιορισμούς στην παραγωγή (ενδεικτικά Stotskii 1971, Kracht 2003, Ράλλη 2005, Booij 2005α), ιδιαίτερα σημαντικός θεωρείται ο μορφολογικός **περιορισμός του φραγμού** (blocking constraint), που έχει προταθεί αρχικά από τον Aronoff (1976: 43-44). Ο περιορισμός του φραγμού δεν επιτρέπει στη γλώσσα να δημιουργήσει λέξεις που ήδη υπάρχουν. Έτσι, στα Ελληνικά δεν έχουμε παράγωγες λέξεις, όπως *\*βασιλευτής*, *\*νίκηση*, *\*κλαδώνω*, *\*κλαύση* κτλ., αφού υπάρχουν στη γλώσσα ήδη οι λέξεις *βασιλιάς*, *νίκη*, *κλαδεύω*, *κλάμα*. Η Ράλλη (2005) υποστηρίζει ότι στον περιορισμό αυτό οφείλεται η απουσία των παραπάνω παραγώγων, αφού η παρουσία των αντίστοιχων λέξεων *βασιλιάς*, *νίκη*, κτλ. λειτουργεί αποτρεπτικά. Θα μπορούσαμε να υποστηρίξουμε ότι ο περιορισμός του φραγμού δημιουργεί ζεύγη προσφύματων που ανταγωνίζονται έχοντας την ίδια λειτουργία και σημασία. Ενδεικτικά παρατίθενται τα ακόλουθα ζεύγη:

(73.) α. ΒΑΣΗ                                 *-άδα*                                 β. ΒΑΣΗ                                 *-οσύνη*



	πράσιν(ος)	πρασιν-άδα	καλ(ός)	καλ-οσύνη
	γρήγορ(ος)	γρηγορ-άδα	κακ(ός)	κακ-οσύνη
	γλυκ(ός)	γλυκ-άδα	δίκαι(ος)	δικαι-οσύνη
γ.	ΒΑΣΗ	-ίζ(ω)	δ. ΒΑΣΗ	-άρ(ω)
	αλώνι	αλων <sup>17</sup> -ίζω	βάρκα	βαρκ-άρω
	αλάτι	αλατ-ίζω	βίγλα	βιγλ-άρω
	άνθ(ος)	ανθ-ίζω	κορνίζα	κορνιζ-άρω
ε.	ΒΑΣΗ	-τατος	στ. ΒΑΣΗ	-ιστος
	άνω	ανώ-τατος	καλ(ός)	κάλλ-ιστος
	άπω	από-τατος		βέλτ-ιστος
	μακάρι(ος)	μακαρι-ότατος		κράτ-ιστος

Τα παραπάνω ζεύγη επιθημάτων συναγωνίζονται στη δημιουργία παραγώγων –τα χαρακτηρίζουμε ως αλλοθήματα – και φέρουν κάποια χαρακτηριστικά ως προς το είδος βάσης με τις οποίες συνδέονται. Το κάθε επίθημα από τα ζεύγη των αλλοθημάτων φέρει χαρακτηριστικά τα οποία προκρίνουν το συγκεκριμένο επίθημα ως τελική επιλογή. Έτσι, το παραγωγικό επίθημα –άδα στα παραδείγματα (73.α) προέρχεται από το βενετσιάνικης προέλευσης επίθημα –*ada* και είναι [+ξενικής προέλευσης], ενώ το επίθημα –οσύνη (73.β) προέρχεται από τα Αρχαία Ελληνικά και είναι λιγότερο παραγωγικό. Η ύπαρξη λέξεων σε –άδα αποτρέπει την εμφάνιση των αντίστοιχων λέξεων σε –οσύνη (\*πρασινοςύνη, \*γρηγοροσύνη) ή και το αντίστροφο (\*καλάδα, \*ταπεινάδα). Παράλληλα, το επίθημα –ότητα που είναι επίσης αρχαιοελληνικής προέλευσης μπορεί σε αρκετές περιπτώσεις να εμφανιστεί σε ίδια περιβάλλοντα με το –οσύνη, όπως *αθωοσύνη/αθωότητα*, *καθαροσύνη/καθαρότητα*, *λευκοσύνη/λευκότητα*, *ταπεινότητα/ταπεινοσύνη* κτλ. Θα μπορούσε και το –ότητα να είναι αλλόθημα με το –άδα, αλλά όχι με το –οσύνη γιατί μοιράζονται αρκετά περιβάλλοντα.

Το επίθημα –άρ(ω) προέρχεται από το ιταλικό απαρεμφατικό επίθημα –*are*, είναι και αυτό μαρκαρισμένο ως [+ξενικής προέλευσης] και συνδυάζεται με θέματα ξενικών λέξεων, ενώ το –ίζ(ω), που προέρχεται από τα Αρχαία Ελληνικά, συνδυάζεται σχεδόν πάντα με ελληνικά θέματα. Τα δύο αυτά επιθήματα θεωρούμε ότι είναι δυνατόν να χαρακτηριστούν ως αλλοθήματα, αφού δεν μπορούν να εμφανιστούν στα ίδια

<sup>17</sup> Όλα τα παράγωγα της συγκεκριμένης ομάδας έχουν ως βάση το αλλόμορφο της παραγωγής με την αποβολή του τελικού –ι.

περιβάλλοντα και φέρουν όμοια σημασιολογικά χαρακτηριστικά. Επιπροσθέτως, για το σχηματισμό του υπερθετικού βαθμού των επιθέτων (73.ε-στ) εξακολουθεί και υφίσταται το αρχαιοελληνικό επίθημα *-ιστ(ος)*, το οποίο είναι μαρκαρισμένο ως [+αρχαιοπρεπές] και έτσι εμποδίζει σε κάποιες περιπτώσεις την επέκταση του *-οτατ(ος)* σε όλες τις βάσεις επιθέτων, όπως παραδείγματος χάριν *\*καλότατος*, *\*κακότατος*, *\*αισχροτάτος*. Τα μορφολογικά χαρακτηριστικά των επιθημάτων διαφοροποιούν το ένα αλλόθημα έναντι του άλλου, καθότι επιλέγουν συγκεκριμένες βάσεις που συνδυάζονται μαζί τους.

Τέλος, κατά τον συνδυασμό βάσης και επιθήματος εμφανίζονται σημασιολογικές επιδράσεις ή με την ευρύτερη έννοια σημασιολογικοί περιορισμοί. Ο Booij (2005a: 67) παραθέτει ότι το πρόθημα των Αγγλικών *in-* επιβάλλει την ερμηνεία της αντιστρεψιμότητας στο ρήμα, δηλαδή έχουμε αναίρεση της διαδικασίας που εκφράζεται από το ρήμα. Το συγκεκριμένο τύπο σημασιολογικής επίδρασης αποκαλεί ο Booij **εξαναγκασμό τύπου** (type coercion). Σύμφωνα με τον εξαναγκασμό τύπου το πρόσφυμα εξαναγκάζει τη βάση να ανήκει σε ένα συγκεκριμένο σημασιολογικό τύπο και έτσι επιβάλλει μία συγκεκριμένη ερμηνεία στην προσφυματοποιημένη λέξη ως ολότητα. Με άλλα λόγια, ο συγκεκριμένος περιορισμός υποδεικνύει πως αν μια βάση δεν πληροί τις σημασιολογικές προϋποθέσεις που θέτει ένα πρόθημα ή επίθημα, τότε δεν είναι δυνατόν να συνδυαστεί με αυτά.

Στα Ελληνικά, δύο επιθήματα που χαρακτηρίζονται ως αλλοθήματα πρέπει να υπόκεινται ή να μην υπόκεινται στη σημασιολογική επίδραση του εξαναγκασμού τύπου. Για παράδειγμα, το επίθημα *-ιν(ος)* απαιτεί από τη βάση να φέρει τη σημασιολογική ερμηνεία της ύλης/ του υλικού για να συνδυαστεί μαζί της (*μάρμαρο* > *μαρμάρινος*, *πέτρα* > *πέτρινος*, *ξύλο* > *ξύλινος*). Ομοίως, το αλλόθημά του *-ένι(ος)* επιβάλλει τον εξαναγκασμό τύπου στην ονοματική βάση, να ανήκει στη σημασιολογική ομάδα της ύλης/ υλικού (*ζάχαρη* > *ζαχαρένιος*, *λάστιχο* > *λαστιχένιος*, *μετάξι* > *μεταξένιος*). Η υπέρσχυση του ενός προσφύματος έναντι κάποιου άλλου βασίζεται αποκλειστικά σε φωνολογικά ή μορφολογικά κριτήρια, αφού συνήθως το σημασιολογικό κριτήριο δεν διαφοροποιείται. Έτσι η διαφορά των αλλοθημάτων *-σιμο* και *-μα* στην επιλογή θέματος είναι βασισμένη σε φωνολογικά κριτήρια, δηλαδή στον αριθμό συλλαβών του ρήματος κατά τους Drachman, Kager & Drachman-Malikouti (1995). Παράλληλα τα αλλοθήματα *-ίζ(ω)* και *-άρ(ω)* ενώνονται με βάσεις διαφορετικού διακριτικού μορφολογικού χαρακτηριστικού, δηλαδή με ελληνικές και

ξένες αντίστοιχα, σύμφωνα με τη Ράλλη (2005).<sup>18</sup> Ο εξαναγκασμός τύπου είναι ένας περιορισμός που έχει μικρό εύρος δράσης στα ελληνικά δεδομένα των προσφυμάτων, αφού είναι λιγοστά τα επιθήματα που συνδυάζονται με βάσεις που φέρουν συγκεκριμένα σημασιολογικά χαρακτηριστικά, όπως τα εθνικά παραγωγικά επιθήματα, π.χ. *-ανος*, *-ωτης*, που συνδυάζονται με βάσεις που δηλώνουν χώρες, πόλεις ή περιοχές ή το επίθημα *-λόι* που απαιτεί η βάση να είναι [+έμφυχο] (με εξαίρεση το *κομπολόι*). Ο συγκεκριμένος περιορισμός δεν είναι απαραίτητη προϋπόθεση που πρέπει να τηρούν τα αλλοθήματα, αλλά είναι αναγκαία η παρουσία ή απουσία του και στα δύο αλλοθήματα.

---

<sup>18</sup> Η Ράλλη (2005) σημειώνει ότι σε ελάχιστες περιπτώσεις το επίθημα *-ίζ(ω)* μπορεί να ενωθεί με ιταλικής προέλευσης θέματα, όπως *paragalo > παπαγαλίζω*.



**HAL:** *I know you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.*

**Dave:** *Where the hell did you get that idea, HAL?*

**HAL:** *Dave, although you took thorough precautions in the pod against my hearing you, I could see your lips move.*

A. Clark (2001, A Space Odyssey)

## 4. Υπολογιστική Μορφολογική Ανάλυση

### 4.1. Εισαγωγή

Οι υπολογιστικές προσεγγίσεις στη γλωσσολογία και ειδικότερα στη μορφολογία και τη σύνταξη είναι γενικά κατανοητές με τυπικές εφαρμογές, όπως τυπικές γραμματικές, στοχαστικά μοντέλα και αλγόριθμους, όπως προσάρτηση ετικέτας (tagging) και μορφοσυντακτική ανάλυση (parsing). Οι προσεγγίσεις αυτές μπορούν να ποικίλλουν και να εκτείνονται από την οπτική βασικής θεωρητικής έρευνας, όπως για παράδειγμα την υπολογιστική πολυπλοκότητα των αλγορίθμων για τη χρήση κάποιας κλάσης γραμματικών. Επίσης περιλαμβάνουν και τις προσεγγίσεις βασικής εφαρμοσμένης έρευνας, όπως για την υλοποίηση καλύτερων τεχνικών για τα στατιστικά γλωσσικά μοντέλα στα συμφραζόμενα της αυτόματης αναγνώρισης ομιλίας (Roark & Sproat 2007). Πρέπει να επισημάνουμε πως στις υπολογιστικές προσεγγίσεις είναι ιδιαίτερα συχνό σε ένα μεγάλο εύρος προβλημάτων να καταφεύγουμε σε γενικές τεχνικές και αλγορίθμους. Το πλέον κλασικό δείγμα αποτελούν τα αυτόματα

πεπερασμένων καταστάσεων (finite-state automata) που χρησιμοποιούνται σε διάφορες τυπικές συσκευές για κωδικοποίηση πολλών μοντέλων, από μορφολογικές γραμματικές σε στατιστικά συντακτικές προσαρτήσεις ετικέτας.

Ας υποθέσουμε πως θέλουμε να μελετήσουμε την Ελληνική προσεγγίζοντάς την υπολογιστικά και βασισμένοι στο μοντέλο της Ράλλη (2005). Κατά τη μελέτη συναντάμε ποικίλες λέξεις και μορφές λεξημάτων και μία από αυτές είναι η λέξη *γράψαμε*, την οποία δεν αναγνωρίζουμε αμέσως· ωστόσο μπορούμε να εικάσουμε λόγω της μορφής του ότι συσχετίζεται και είναι τύπος του ρήματος *γράφω*. Ύστερα από μικρή διερεύνηση της γραμματικής της γλώσσας, βεβαιωνόμαστε για τη σχέση του τύπου *γράψαμε* με το ρήμα *γράφω* και έχουμε όλες τις πληροφορίες για να τον χαρακτηρίσουμε, αφού είναι ο τύπος για το Α΄ΠΡΟΣΩΠΟ, ΠΛΗΘΥΝΤΙΚΟΣ, ΣΥΝΟΠΤΙΚΟ, ΕΝΕΡΓΗΤΙΚΗ ΦΩΝΗ. Επιπροσθέτως με τη βοήθεια της γραμματικής μαθαίνουμε να χωρίζουμε τη λέξη στα συστατικά της που είναι το ρηματικό θέμα *γραφ* (το οποίο τρέπεται σε *γραπ* λόγω του φωνολογικού κανόνα της ανομοίωσης λόγω *-σ-*), το κλιτικό επίθημα της ενεργητικής συνοπτικής όψης *-σ-* και το κλιτικό επίθημα πρώτου προσώπου πληθυντικού *-αμε*.

Ο συσχετισμός μορφών λέξεων και ο εντοπισμός της δομής των λέξεων είναι το κέντρο και η βάση της υπολογιστικής μορφολογικής ανάλυσης. Ο στόχος της συσχέτισης μίας δοθείσας λέξης με μία κανονική μορφή ονομάζεται **λημματισμός** (lemmatization)· έτσι για παράδειγμα στα Ελληνικά, θα λημματίσουμε το *γράψαμε* στο *γράφω*. Όπως και στο Νοητικό Λεξικό, έτσι και εδώ λήμμα αποτελεί η μορφή της λέξης που βρίσκουμε σε ένα τυπικό λεξικό. Από την άλλη ο εντοπισμός της δομής και ο διαχωρισμός στα συστατικά της λέξης λέγεται **ανάλυση συστατικών** (decomposition)· έτσι στην παραπάνω περίπτωση θα εντοπίσουμε τα συστατικά και θα αντιστοιχίσουμε στα Α΄ΠΡΟΣΩΠΟ, ΠΛΗΘΥΝΤΙΚΟΣ, ΣΥΝΟΠΤΙΚΟ, ΕΝΕΡΓΗΤΙΚΗ ΦΩΝΗ.

Για την καλύτερη κατανόηση της αλληλένδετης δυναμικής ανάμεσα στον λημματισμό και την ανάλυση συστατικών θα παρουσιάσουμε δύο συναρτήσεις. Για παράδειγμα, εάν έχουμε μία συνάρτηση  $A$  που να μετασχηματίζει μία λέξη στην αντίστοιχη μορφολογική ανάλυση συστατικών, μπορούμε να συνδυάσουμε αυτή τη συνάρτηση με μία άλλη, τη συνάρτηση  $\Lambda$  που καθορίζει σε ποιο λήμμα πρέπει να τοποθετηθεί κάθε αναλυόμενο θέμα. Επομένως, μπορούμε να λημματίσουμε μία δοσμένη λέξη με την εφαρμογή της συνάρτησης  $A \circ \Lambda$  σε αυτή τη λέξη.

Έτσι σε αυτή την περίπτωση για το παράδειγμα της Ελληνικής, η συνάρτηση  $\Lambda$  θα αντιστοιχίσει το *γράψαμε* σε *γραπ-σ-αμε*, προσαρτώντας ετικέτες στη λέξη όπως  $\Lambda'$  ΠΡΟΣΩΠΟ, ΠΛΗΘΥΝΤΙΚΟΣ, ΑΟΡΙΣΤΟΣ, ΕΝΕΡΓΗΤΙΚΗ ΦΩΝΗ με το θέμα γραφ:

γραφ+σ<sub>ΣΥΝΟΠΤΙΚΟ</sub>+αμε $\Lambda'$  ΠΡΟΣΩΠΟ, ΠΛΗΘΥΝΤΙΚΟΣ, ΕΝΕΡΓΗΤΙΚΗ

ενώ η συνάρτηση  $\Lambda$  θα αντιστοιχίσει το θέμα γραφ στην κανονιστική μορφή πρώτου προσώπου ενεστώτα ενεργητικής φωνής αφαιρώντας το υπόλοιπο υλικό χαρακτηριστικών, και αποδίδοντας τον τύπο ως

γράφω<sub>ΕΝΕΣΤΩΤΑΣ,  $\Lambda'$  ΠΡΟΣΩΠΟ, ΕΝΙΚΟΣ, ΕΝΕΡΓΗΤΙΚΗ</sub>

Στόχος λοιπόν της υπολογιστικής μορφολογικής ανάλυσης είναι να λαμβάνει λεξικές μορφές και να προσπαθεί να τις συσχετίσει με άλλες λεξικές μορφές, ενώ παράλληλα προσπαθεί να εξάγει πληροφορίες χαρακτηριστικών από τις δοθείσες μορφές.

Είναι σύνηθες στη μορφολογία να υφίσταται έντονες συζητήσεις, αντεγκλήσεις και διαφορές για την *κλιτική* και *παραγωγική* μορφολογία. Οι Roark & Sproat (2007: 24-27) υποστηρίζουν ότι εν μέρει αυτές οι διαφορές έχουν μεταπηδήσει και στην υπολογιστική μορφολογική ανάλυση με τις αντίστοιχες προσεγγίσεις. Επισημαίνουν εύστοχα ότι η υπολογιστική προσέγγιση της μορφολογίας εμπίπτει συχνά σε σημαντικές εμπλοκές που θα πρέπει να προκαλέσουν το ενδιαφέρον της θεωρητικής μορφολογίας.

## 4.2. Σύντομη ιστορία της Υπολογιστικής Μορφολογίας<sup>1</sup>

Η Υπολογιστική Μορφολογική Ανάλυση (ΥΠΑ) πραγματοποίησε τα πρώτα βήματα στα πρώιμα στάδια έρευνας της υπολογιστικής γλωσσολογίας: έχει κιάλας συμπληρώσει περισσότερο από μισό αιώνα ιστορίας, μίας και οι πρώτες προσπάθειες εμφανίστηκαν πριν τα πρώτα ευρήματα των ερευνών της Μηχανικής Μετάφρασης. Οι Roark & Sproat (2007) αναφέρονται τιμητικά στους Andron (1962), Woyna (1962), Bernard-Georges *et al.* (1962), Boussard & Berthaud (1965), Vauquois (1965), Schweiger & Mathe (1965), Matthews (1966), Brand *et al.* (1969), Hutchins (2001). Ειδικά ο Hutchins (2001: 8-9) αναφέρει ότι στην περίοδο 1953-1956 μπήκαν τα πρώτα θεμέλια στην υπολογιστική μορφολογία διαμέσου του πεδίου της Μηχανικής Μετάφρασης. Στην πάροδο αυτών των χρόνων πολλές εφαρμογές δημιουργήθηκαν

<sup>1</sup> Η παρούσα υποενότητα είναι βασισμένη σε σημαντικό βαθμό στο κεφάλαιο 4 (A Brief History of Computational Morphology) των Roark & Sproat (2007).

βασιζόμενες εν μέρει ή πλήρως στην υπολογιστική μορφολογική ανάλυση: αυτές αποτελούν μία συλλογή από θεματοποιητές (stemmers, όπως Porter 1980), μορφολογικούς αναλυτές (parsers, όπως Koskenniemi 1983, Allen *et al.* 1987, Coker *et al.* 1990), ορθογραφικούς διορθωτές (spelling correction, όπως McIlroy 1982, Hankamer 1986), συστήματα ανάκτησης πληροφορίας και συστήματα φυσικής γλώσσας (information retrieval and nature language generation, όπως μεταξύ άλλων από Koskenniemi 1983, Meya-Lloport 1987, Choueka 1990).

Παρόλη τη σημαντικότητα των προαναφερθέντων πρωτεργατών, η πιο ενδιαφέρουσα και ρηξικέλευθη εργασία στην υπολογιστική γλωσσολογία από θεωρητικής και υπολογιστικής απόψεως ήταν η έρευνα που βασιζόταν κυρίως σε μεθόδους **πεπερασμένων καταστάσεων** (finite-states). Η κυριαρχούσα μορφολογία πεπερασμένων καταστάσεων υπήρξε η προσέγγιση του Koskenniemi (1983), βασισμένη σε μετατροπείς πεπερασμένων καταστάσεων. Ο Koskenniemi (1983), στηριζόμενος στη θεωρητική έρευνα των Kaplan & Kay (1981) της Xerox PARC, κατάφερε την πρακτική εφαρμογή τους με τη **μορφολογία δύο-επιπέδων** ή **διεπίπεδη μορφολογία** (two-level morphology). Ο PC-KIMMO, ένας μορφολογικός αναλυτής δύο-επιπέδων, εξακολουθεί να είναι *state-of-the-art* στην υπολογιστική γλωσσολογία παρόλη την πολυπλοκότητά του. Διαφορετικές προσεγγίσεις στην υπολογιστική μορφολογία κατά την περίοδο κυριαρχίας των πεπερασμένων αυτομάτων αποτέλεσαν το **μορφοτακτικό μοντέλο πεπερασμένων καταστάσεων** (morphotactics finite-state, πρβλ. Allen *et al.* 1987, Roark & Sproat 2007) και τα **μοντέλα αποκοπής επιθημάτων** (suffix stripping models, πρβλ. Porter 1980, McIlroy 1982).

Όπως σημειώνουν οι Roark και Sproat (2007: 102) η Μορφολογία Δύο-Επιπέδων του Koskenniemi έχει χρησιμοποιηθεί για τη δημιουργία μορφολογικών αναλυτών σε αρκετές γλώσσες, συμπεριλαμβανομένων και των Ελληνικών (βλ. Σγάρμπας 1997, Μαρκόπουλος 1998). Ο PC-KIMMO είναι ένα σύστημα που βασίζεται στη διατριβή του Koskenniemi (1983) και περιέχει τρεις κεντρικούς άξονες: α) τα **δέντρα** (trees), η απεικόνιση των λεξιλογίων, β) τα **λεξικά συνέχειας** (continuation lexica), η απεικόνιση των μορφολογικών αλληλουχιών, και γ) τους **μετατροπείς πεπερασμένων συστημάτων** (finite-state transducers – FST), που παρουσιάζουν τις μορφές, αλλαγές και διαδικασίες της επιφανειακής-λεξικής μορφολογίας. Στην ακόλουθη υποενότητα θα δούμε αναλυτικότερα τη μέχρι πρότινος δημοφιλέστερη υπολογιστική μορφολογική προσέγγιση.



### 4.3. Αυτόματα Πεπερασμένων-Καταστάσεων

Στην υπολογιστική ανάλυση, οι **εκφράσεις κανονικότητας<sup>2</sup> ή κανονικές εκφράσεις** (regular expressions) είναι κάτι παραπάνω από μία κατάλληλη μεταγλώσσα για την αναζήτηση σε κείμενα. Σύμφωνα με τους Jurafsky & Martin (2000: 33) οι κανονικές εκφράσεις είναι ένας τρόπος, για να περιγράψει ένα **αυτόματο πεπερασμένων-καταστάσεων** (finite-state automaton – FSA)<sup>3</sup>· είναι ένα μοντέλο συμπεριφοράς που αποτελείται από έναν συγκεκριμένο αριθμό καταστάσεων, από μεταβάσεις ανάμεσα σε αυτές τις καταστάσεις και δράσεις. Παρομοιάζεται με ένα **γράφημα** ροής, όπου είναι δυνατή η επιθεώρηση της πορείας που ακολουθεί η λογική εφόσον συγκεκριμένες προϋποθέσεις εκπληρώνονται. Κάθε έκφραση κανονικότητας<sup>3</sup> μπορεί να υλοποιηθεί από ένα αυτόματο πεπερασμένων καταστάσεων, ενώ παράλληλα κάθε FSA μπορεί να περιγράψει πλήρως από μία κανονική έκφραση ή φράση κανονικότητας. Δευτερευόντως, όπως έχει ήδη αναφερθεί οι κανονικές εκφράσεις μαζί με τα FSA αποτελούν ικανοποιητικούς τρόπους περιγραφής μίας τυπικής γλώσσας που ονομάζεται **γλώσσα κανονικότητας<sup>4</sup> ή κανονική γλώσσα** (regular language). Το ακόλουθο σχήμα περιγράφει αυτή τη σχέση.

---

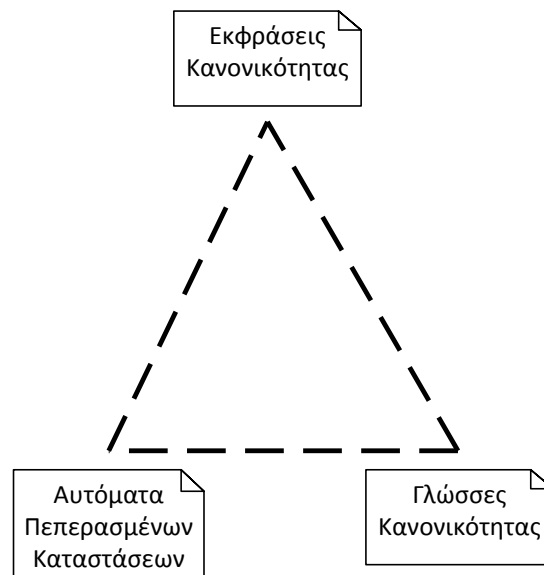
<sup>2</sup> Οι κανονικές εκφράσεις ή εκφράσεις κανονικότητας (regular expressions, regex, regexp) στην πληροφορική είναι ένας επαρκής και αποτελεσματικός τρόπος για να εντοπίζουμε ακολουθίες γραμμάτων μέσα σε ένα κείμενο, όπως λέξεις και συνδυασμούς γραμμάτων. Συνήθως οι εκφράσεις κανονικότητας γράφονται σε τυπική γλώσσα (formal language) στον προγραμματισμό και ερμηνεύονται με επεξεργαστές εκφράσεων κανονικότητας ή μορφολογικών αναλύσεων. Χρησιμοποιούνται από όλους τους επεξεργαστές κειμένου και στην κωδικοποίηση μηχανών αναζήτησης.

Για παράδειγμα η έκφραση `.έλος` βρίσκει λέξεις όπως *βέλος*, *μέλος*, *τέλος*, η έκφραση `[βμ]?έλος` βρίσκει τις λέξεις *βέλος*, *μέλος* και *έλος*, ενώ η έκφραση `βέλος|μέλος` βρίσκει τις λέξεις *μέλος* και *βέλος*.

Ενδεικτική βιβλιογραφία, Sipser (1998), Friedl (2002), Stubblebine (2003), Gruber & Neven (2008), Gruber & Holzer (2008).

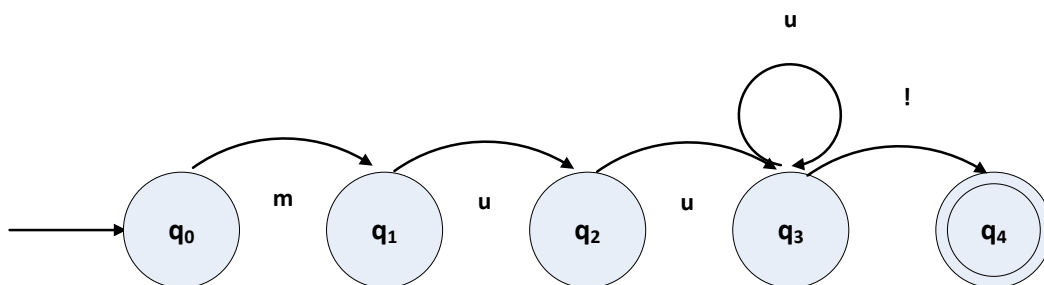
<sup>3</sup> Σε αυτές δεν συμπεριλαμβάνονται οι κανονικές εκφράσεις που απαιτούν τη χρήση της μνήμης (memory feature). Για περισσότερες πληροφορίες βλ. Jurafsky & Martin (2000) *Chapter 2* και βιβλιογραφία προηγούμενης υποσημείωσης.

<sup>4</sup> Οι γλώσσες κανονικότητας (regular languages) είναι περιβαλλοντικά ελεύθερες γραμματικές (context-free grammars) που παράγουν μία διαδοχή χαρακτήρων (strings) που αποτελούνται από μη-τερματικά σύμβολα, τερματικές και κενές λέξεις. Μπορούν να περιγραφούν από ντετερμινιστικά ή μη-ντετερμινιστικά πεπερασμένα αυτόματα που παράγουν ή αποδέχονται προτάσεις που προέρχονται από αυτές τις γλώσσες (βλ. Salomaa 1985, Yu 1997, Jurafsky & Martin 2000).



Γράφημα 2: Η σχέση μεταξύ αυτόματων πεπερασμένων-καταστάσεων, γλωσσών κανονικότητας και κανονικών εκφράσεων, όπως αποδίδεται από τους Jurafsky & Martin (2000: 33).

Ενώ αρχικά τα αυτόματα πεπερασμένων-καταστάσεων χρησιμοποιούνται κατά κόρον για την υλοποίηση εκφράσεων κανονικότητας, πλέον τα συγκεκριμένα αυτόματα έχουν ένα μεγάλο εύρος εφαρμογών και χρήσεων. Για την καλύτερη κατανόηση θα δημιουργήσουμε ένα FSA που να ανταποκρίνεται στη «γλώσσα μίας αγελάδας» και περιγράφει πλήρως το μουγκανητό. Θεωρούμε ότι ο ήχος που παράγει μία αγελάδα είναι το /mu:!/ , όπου το /u/ είναι συνήθως ιδιαίτερα διαρκές και δηλώνεται ως muu!, muuu!, muuuu!, muuuuu!, ... Το ακόλουθο σχήμα αποτελεί την υλοποίηση της κανονικής έκφρασης ενός μουγκανητού αγελάδος σε αυτόματο.



Γράφημα 3: Ένα αυτόματο πεπερασμένων-καταστάσεων για μία αγελάδα!

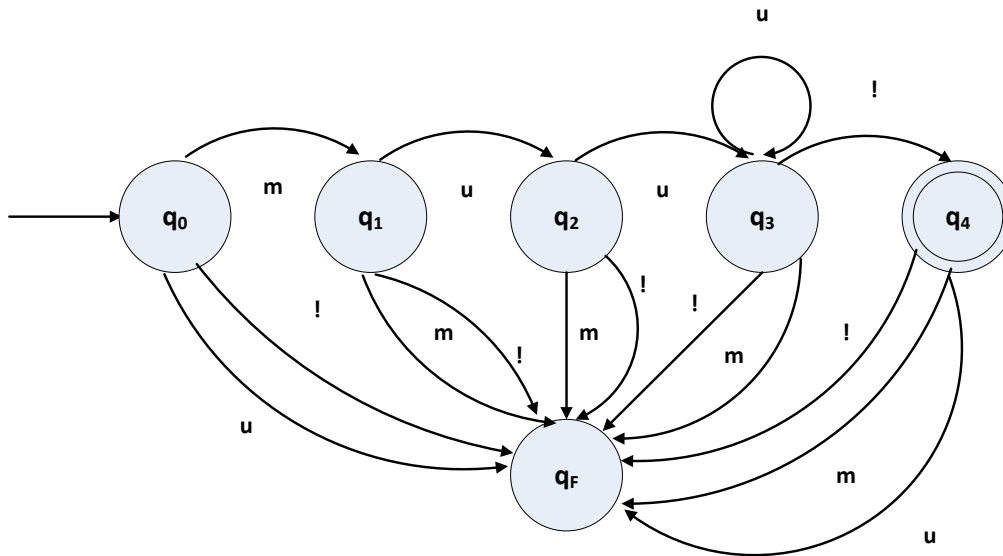
Το παραπάνω αυτόματο περιγράφει τη «γλώσσα μίας αγελάδας». Περιέχει πέντε καταστάσεις (κόμβους στο σχήμα), από τις οποίες η κατάσταση 0 είναι η **αρχική κατάσταση** (start state) και η κατάσταση 4 είναι η **τελική ή αποδεκτή κατάσταση** (final or accepting state) που αναπαρίσταται πάντα με διπλό κύκλο. Οι υπόλοιποι κόμβοι είναι οι ενδιάμεσες καταστάσεις με τα βέλη να αναπαριστούν τις τέσσερις

**μεταβάσεις** (transitions) από κατάσταση σε κατάσταση. Όταν μία ακολουθία χαρακτήρων δεν πληροί τις προϋποθέσεις (π.χ.  $\mu!$ ,  $\mu u a!$ ), ώστε να αναλυθεί από ένα αυτόματο και αποτυγχάνει να φτάσει στην τελική κατάσταση λόγω πρόωρης ολοκλήρωσης (βλ.  $\mu!$ ) ή μη αντιστοίχισης με κάποια κατάσταση (βλ.  $\mu u a!$ ), τότε λέμε πως το αυτόματο απορρίπτει ή αποτυγχάνει να αναλύσει αυτές τις ακολουθίες χαρακτήρων.

Πιο τυπικά, το παραπάνω αυτόματο πεπερασμένων-καταστάσεων πρέπει να οριστεί από τις ακόλουθες πέντε παραμέτρους:

- ❖  $Q$ : ένα πεπερασμένο σύνολο από  $N$ -καταστάσεις  $q_0, q_1, q_2 \dots q_N$
- ❖  $\Sigma$ : ένα πεπερασμένο αλφάβητο εισόδου από σύμβολα
- ❖  $q_0$ : η αρχική κατάσταση
- ❖  $F$ : το σύνολο των τελικών καταστάσεων  $F \subseteq Q$
- ❖  $\delta(q,i)$ : η συνάρτηση μετάβασης ή ο πίνακας μεταβάσεων ανάμεσα σε καταστάσεις. Δοθείσης μίας κατάστασης  $q \in Q$  και ενός συμβόλου εισόδου  $i \in \Sigma$ , η  $\delta(q,i)$  επιστρέφει μία νέα κατάσταση  $q' \in Q$ . Επομένως, η  $\delta$  είναι η σχέση από το  $Q \times \Sigma$  στο  $Q$ .

Άρα για το αυτόματο του γραφήματος (3.) έχουμε για  $Q = \{q_0, q_1, q_2, q_3, q_4\}$ ,  $\Sigma = \{m, u, !\}$ ,  $F = \{q_4\}$  και  $\delta(q,i)$  που περιγράφεται από έναν πίνακα μεταβάσεων. Το ανωτέρω αυτόματο ωστόσο θα αποτύχει – όπως και κάθε πεπερασμένο αυτόματο – αν δεν υπάρχει νόμιμη μετάβαση από ένα δοσμένο συνδυασμό καταστάσεων και κειμένου εισόδου. Δηλαδή, σε περιπτώσεις όπου δίνεται ένα σύμβολο που δεν υπάρχει στο αλφάβητο ή μία ακολουθία που δεν ικανοποιείται από τις παρούσες καταστάσεις τότε η ολοκλήρωση είναι αδύνατη. Τις «κενές» αυτές περιπτώσεις έχουμε τη δυνατότητα να τις αντιμετωπίσουμε μεταβιβάζοντάς τες σε μία «κενή» κατάσταση που την ονομάζουμε **κατάσταση αποτυχίας** (fail or sink state). Υπό αυτή την προσέγγιση μπορούμε να διευρύνουμε το παρόν αυτόματο με μία ακόμα κατάσταση, αλλά και τις μεταβάσεις που χρειάζονται, ώστε να καλύψουμε πλήρως την αναβάθμιση του αυτόματου. Το ακόλουθο σχήμα είναι βασισμένο στο σχήμα 63 με την προσθήκη της κατάστασης  $q_F$  και των νέων τόξων μεταβάσεων.



Γράφημα 4: Ένα πεπερασμένο αυτόματο με κατάσταση αποτυχίας  $q_F$ .

Κλείνοντας, να αναφέρουμε πως εκτός από τα ντετερμινιστικά αυτόματα (βλ. Γράφημα 4), όπου δεν υφίστανται σημεία επιλογής και ο αλγόριθμος ξέρει πάντα τι να κάνει με οποιοδήποτε κείμενο εισόδου, υπάρχουν και τα **μη-ντετερμινιστικά αυτόματα πεπερασμένων καταστάσεων** (non-deterministic FSAs). Τα συγκεκριμένα αυτόματα εμφανίζουν μεταβάσεις που επιστρέφουν σε προηγούμενη κατάσταση και έτσι δεν είναι καθορισμένη η διαδρομή αναγνώρισης από την αρχή, δηλαδή δεν είναι μία διαδρομή προκαθορισμένη και χωρίς επιλογές. Αυτές οι επιλογές ενέχουν τον κίνδυνο να οδηγηθεί σε λανθασμένο μονοπάτι και να αποτύχει η αναγνώριση. Όντας σημαντικό πρόβλημα αυτό στα μη-ντετερμινιστικά αυτόματα, οι λύσεις είναι (α.) το *back-up*, όπου τοποθετείται κάποιος δείκτης στο κείμενο εισόδου για να επανέλθουμε σε περίπτωση λανθασμένου μονοπατιού, (β.) η *προορατότητα* (look-ahead), όπου βλέπουμε το υπόλοιπο κείμενο προτού αποφασίσουμε για το μονοπάτι που θα ακολουθηθεί για την αναγνώριση και (γ.) ο *παράλληλισμός*, όπου σε κάθε σημείο επιλογής παρατηρούμε παράλληλα κάθε εναλλακτικό μονοπάτι.

Εξέλιξη των αυτομάτων πεπερασμένων καταστάσεων είναι οι μετατροπείς πεπερασμένων καταστάσεων, που εκτός από το κείμενο εισόδου έχουν και κείμενο εξόδου (text output) με το αντίστοιχο πεπερασμένο σετ αλφαβήτου εξόδου (συμβολίζεται με το  $\Gamma$ ). Οι μετατροπείς δέχονται δύο υποδοχές, μία για το κείμενο εισόδου και μία για το κείμενο εξόδου· μετατρέπει το κείμενο εισόδου της κασέτας εισόδου σε κείμενο στην κασέτα εξόδου και όντας συνήθως μη-ντετερμινιστικό το

κείμενο εισόδου αντιστοιχεί συχνά σε πλείονα του ενός κείμενα εξόδου. Σε γενικές γραμμές, ένας μετατροπέας υπολογίζει τη σχέση ανάμεσα σε δύο φορμαλιστικές/τυπικές γλώσσες. Για αυτό και χρησιμοποιούνται συχνά σε φωνολογικές και μορφολογικές αναλύσεις στις έρευνες και εφαρμογές της επεξεργασίας φυσικής γλώσσας. Πρωτοπόροι στο συγκεκριμένο τομέα θεωρούνται οι Roland Kaplan και Martin Kay της Xerox, ο Kimmo Koskenniemi και ο Lauri Karttunen που συνέβαλαν στη δημιουργία του πολύπλοκου και state-of-the-art μορφολογικού αναλυτή δύο επιπέδων, PC-KIMMO.

#### 4.3.1. PC-KIMMO: Μορφολογικός αναλυτής Δύο-Επιπέδων

Η προσέγγιση της Διεπίπεδης Μορφολογίας του Koskenniemi (1983) αποτέλεσε βάση για την ανάπτυξη ενός μεγάλου εύρους μορφολογικών αναλυτών για ένα σημαντικό μέγεθος γλωσσών. Η πλέον αναγνωρίσιμη υλοποίηση αυτής της προσέγγισης είναι ο PC-KIMMO, ένας μορφολογικός αναλυτής δύο-επιπέδων από τον Karttunen (1983) και στη συνέχεια από την ομάδα του Antworth που έχει περιγράψει και αναλυθεί από διάφορους, όπως από τους Antworth (1990, 1993), Sproat (1992), Beesley & Karttunen (2003), Karttunen & Beesley (2005), Roark & Sproat (2007). Ο PC-KIMMO έχει εξελιχθεί σε μία εξειδικευμένη γλώσσα προγραμματισμού από ειδικό περιβάλλον προγραμματισμού. Κάθε ακολουθία χαρακτήρων παρουσιάζεται ως μετατροπέας πεπερασμένων καταστάσεων με ένα μοναδικό μονοπάτι που διασταυρώνεται με μία μηχανή με την κατάλληλη φορμαλιστική γλώσσα.

##### 4.3.1.1. Τα βασικά του Ελληνικού PC-KIMMO (Σγάμπας 1996)<sup>5</sup>

Για τη μορφολογική περιγραφή και ανάλυση μίας γλώσσας και δη της Ελληνικής με τη χρήση του PC-KIMMO είναι αναγκαία η παρουσία ενός λεξικού που αντιπροσωπεύει τη μορφοτακτική δομή της γλώσσας και η συνοδεία από κανόνες δύο επιπέδων που θα αντιμετωπίσουν τις μορφοφονολογικές αλλαγές που λαμβάνουν χώρα στα όρια μορφημάτων. Αυτοί οι δύο ακρογωνιαίοι λίθοι του συστήματος αποθηκεύονται ως ξεχωριστά αρχεία με τη μορφή συστημάτων πεπερασμένων καταστάσεων· ειδικότερα το λεξικό αποτελείται από απλά πεπερασμένα αυτόματα

---

<sup>5</sup> Επίσης ο Μαρκόπουλος (1998) ασχολήθηκε με μία υλοποίηση του PC-KIMMO για το ονοματικό σύστημα της Ελληνικής, αλλά οι αναφορές στην έρευνά του θα γίνουν στο κομμάτι της κριτικής.

(FSAs) και οι κανόνες από απλούς μετατροπείς πεπερασμένων καταστάσεων (FSTs). Ο Σγάρμπας (1996: 67) περιγράφει ότι το λεξικό αποτελείται από εγγραφές μορφημάτων με δείκτες σε άλλα μορφήματα, με τους δείκτες να δηλώνουν τα μορφήματα που επιτρέπεται να συνδυαστούν με τα συγκεκριμένα μορφήματα: οι κανόνες αποτελούνται από πίνακες που καθορίζουν τη σχέση μεταξύ καταστάσεων – μεταβάσεων. Παράλληλα υφίσταται και ένα αρχείο που αφορά στη μορφολογική γραμματική και τους κανόνες που συσχετίζονται με τις γραμματικές κατηγορίες των μορφημάτων.

Στο λεξικό καταχωρούνται οι λέξεις χωρισμένες σε μορφήματα, συνήθως με τη συνοδεία συμβόλων για την εξυπηρέτηση εφαρμογής των κανόνων και δεν έχουν τη σωστή ορθογραφική μορφή, η οποία επιτυγχάνεται στην επιφανειακή μορφή. Τα μορφήματα εντάσσονται σε υπολεξικά με κριτήρια τα μορφοτακτικά τους χαρακτηριστικά (κλίση, τάξη, γένος, κτλ.). Για παράδειγμα, τα θέματα *βουν-*, *μωρ-*, *παλτ-* αποτελούν το υπολεξικό N6 και οι αντίστοιχες καταλήξεις ενικού *-ο* και *-ου* στο E6 που συνδυάζονται μεταξύ τους για το σχηματισμό του κλιτικού παραδείγματος του ενικού. Η καταχώρηση των μορφημάτων στα υπολεξικά περιλαμβάνει πληροφορίες σχετικά με την κατηγορία του μορφημάτων, των πιθανών συνδυασμών με άλλα μορφήματα, λοιπές μορφολογικές πληροφορίες και πιθανή μετάφραση.

Στο αρχείο των κανόνων ορίζονται εκτός από τους κανόνες δύο επιπέδων, το αλφάβητο και βασικά σύμβολα βοήθειας, ώστε να περιγράψουν τις μορφοφωολογικές αλλαγές που συντελούνται στα όρια μορφημάτων κατά τις διαδικασίες σχηματισμού λέξεων. Τέλος, ο Σγάρμπας (1996: 77-82) σημειώνει ότι η μορφολογική γραμματική λειτουργεί προαιρετικά, αλλά σημαντικώς συνεπικουρικά στη λειτουργία του μορφολογικού αναλυτή. Χρησιμοποιεί έναν αναλυτή που έχει χτιστεί πάνω στη θεωρία της ενοποίησης (unification) και αποτελείται από σχεδιάτυπα χαρακτηριστικών (feature templates), κανόνες και περιορισμούς χαρακτηριστικών (feature constraints). Για επιπρόσθετες πληροφορίες για τη μορφολογία δύο επιπέδων στα Ελληνικά, πρβλ. Sgarbas, Fakotakis & Kokkinakis (1995, 2000), Σγάρμπας, Φακωτάκης & Κοκκινάκης (1999).

#### 4.3.1.2. Κριτική στο μοντέλο PC-KIMMO για τα Ελληνικά

Στην Ελληνική υλοποίηση του PC-KIMMO από τον Σγάρμπα (1996) έγινε μία συστηματική προσπάθεια καταγραφής και απόδοσης του κλιτικού συστήματος των

ουσιαστικών, επιθέτων, μετοχών και ρημάτων. Βασιζόμενος σε παλαιότερες γραμματικές και παραδοσιακές προσεγγίσεις εντοπίζει 89 κατηγορίες για τα ουσιαστικά, μίας και εκτός από τις διαφορετικές καταλήξεις σε κάθε πτώση, η θέση του τόνου πολύ συχνά αλλάζει κατά την κλίση, με τη συγκεκριμένη θέση να απέχει ελάχιστα από την ονοματική κατηγοριοποίηση του Τριανταφυλλίδη (1941). Οργανώνει με μεγάλη προσοχή τις κατηγορίες με τελεστές τονισμού για τη δημιουργία οικονομικότερου μοντέλου. Εντούτοις δεν καταφέρνει να περιορίσει σημαντικά τις κατηγορίες, αν αναλογιστούμε ότι η ονοματική κλίση της Ελληνικής έχει οκτώ κλιτικές τάξεις σύμφωνα με το μοντέλο της Ράλλη (2000, 2005). Ταυτόχρονα για το σχηματισμό των κλιτικών παραδειγμάτων δίνονται λανθασμένα κλιτικά επιθήματα και παράλληλα εξαλείφεται κάθε είδους αλλομορφία, αφού κάτι τέτοιο θα δυσκόλευε περισσότερο το σύστημα.

(74.)	N16	πρύτανης	>	πρυταν-	+ης, +η, +η, +εις, +εων, +εις, +εις
	N17	δουλευτής	>	δουλευ-	+της, +τη, +τη, +τες, +των, +τες, +τες
	N25	καφές	>	καφ-	+ες, +ε, +ε, +εδες, +εδων, +εδες, +εδες
	N45	ψυχή	>	ψυχ-	+η, +ης, +η, +η, +ες, +ων, +ες, +ες
	N55	αλεπού	>	αλεπ-	+ου, +ους, +ου, +ου, +ουδες, +ουδων, +ουδες, +ουδες
	N73	παιδί	>	παιδ-	+ι, +ιου, +ι, +ι, +ια, +ιων, +ια, +ια
	N84	όνομα	>	ονο-	+μα, +ματος, +μα, +μα, +ματα, +ματων, +ματα, +ματα

(από τον πίνακα 4.3 του Σγάρμπα 1996: 88-89)

Από τα παραπάνω παραδείγματα έχουν αφαιρεθεί οι τελεστές τονισμού, καθώς και οι τόνοι στα χωρισμένα μορφήματα. Τα παραδείγματα είναι επιλεγμένα με βασικό κριτήριο την παρουσία αλλομορφίας. Παρατηρείται ότι τόσο οι περιπτώσεις της προσθετικής αλλομορφίας (βλ. N25, N55, N84), όσο και εκείνες της αφαιρετικής αλλομορφίας (βλ. N16, N17, N45, N73) αντιμετωπίζονται με τον ίδιο τρόπο, δηλαδή οι οποιοσδήποτε αλλομορφικές αλλαγές του θέματος προσμετρώνται στα κλιτικά επιθήματα. Πιο συγκεκριμένα, στο παράδειγμα του *καφές* το σύστημα αντί να καταχωρήσει τα θέματα *καφε~* και *καφεδ~* με τα αντίστοιχα κλιτικά επιθήματα, τα κλιτικά επιθήματα «εκτείνονται» με αποτέλεσμα να δημιουργούνται ψεύτικες ομάδες από κλιτικά επιθήματα. Επιπροσθέτως, υπάρχουν κάποια παραδείγματα στα οποία

εμπλέκονται παράγωγες λέξεις μέσα στα κλιτικά παραδείγματα, όπου το παραγωγικό επίθημα είναι ενοποιημένο με το κλιτικό επίθημα.

Η ανάλυση του ρηματικού κλιτικού συστήματος από τον μορφολογικό αναλυτή PC-KIMMO είναι σαφώς πιο οικονομική από την αντίστοιχη του ονοματικού, αφού διαθέτει κατηγορίες με ένα σαφώς οικονομικό σύστημα, καθότι τα ρήματα έχουν μεγαλύτερη συστηματικότητα στις αλλαγές του θέματος/ ρίζας. Και στη συγκεκριμένη ομάδα η αντιμετώπιση της αλλομορφίας γίνεται με τον ίδιο τρόπο· η προσθετική και η αφαιρετική αλλομορφία αντιμετωπίζεται με την αύξηση των κλιτικών επιθημάτων, τα οποία στους παρελθόντες χρόνους με συνοπτικότητα συμπεριλαμβάνουν και την όψη.

(75.)	V3	γελώ	γελ-	>	+ασα [EN-OP-AO], +αστηκα [ΠΑ-OP-AO]
	V4	λαλώ	λαλ-	>	+ησα [EN-OP-AO], +ηθηκα [ΠΑ-OP-AO]
	V5	καλώ	καλ-	>	+εσα [EN-OP-AO], +εστηκα [ΠΑ-OP-AO]

(από τον πίνακα 4.4 του Σγάρμπα 1996: 92-93)

Οι τελεστές στο ρηματικό κλιτικό σύστημα είναι σαφώς πιο πολύπλοκοι, γιατί πρέπει να περιγράψουν τις αλλαγές στο ενεστωτικό θέμα, την πιθανή μετακίνηση τόνου και τέλος τις περιπτώσεις του κατάλοιπου χρονικής αύξησης. Εντούτοις, η αντιμετώπιση πιο πολύπλοκων αλλομόρφων (π.χ. *παιρν~ πηρ~ παρ~*) αυξάνει τους «κανόνες» σχηματισμού και το πιθανότερο είναι η αποτυχία δημιουργίας και ανάλυσης των συγκεκριμένων κλιτικών τύπων. Σε γενικές γραμμές, όπως έχει ειπωθεί, η ανάλυση του Σγάρμπα (1996) δεν ακολουθεί σύγχρονα γλωσσολογικά πρότυπα, αλλά τις υποδείξεις και καθοδηγήσεις παραδοσιακών γραμματικών για τα Ελληνικά με αποτέλεσμα να φαίνεται ότι καταφεύγει σε αντιοικονομικές λύσεις, οι οποίες υπολογιστικά αποδίδουν συνήθως ορθά αποτελέσματα.

Αντίθετα, ο Μαρκόπουλος (1998) για την υπολογιστική επεξεργασία του Νεοελληνικού ονόματος υιοθετεί το μοντέλο της Ralli (1994, 1997) με σαφώς πιο οικονομικούς προσανατολισμούς και με ελάχιστες διαφοροποιήσεις στο κλιτικό σύστημα. Διαφοροποιείται στην εισαγωγή αλλομόρφων, όπου εντάσσει στο λεξικό μορφολογικά αλλόμορφα, φωνολογικά αλλόμορφα και αλλόμορφα τονισμού, καθότι οι μετατροπείς πεπερασμένων καταστάσεων εμφανίζουν πρόβλημα με τον τονισμό. Ειδικότερα για την ονοματική παραγωγή, εισάγει αρκετούς φωνολογικούς κανόνες που λαμβάνουν χώρα στα όρια μορφημάτων, ενώ για την κωδικοποίηση των διάφορων παραγωγικών σχηματισμών και την αποφυγή αντιγραμματικής δημιουργίας λέξεων



στην λημματική καταχώρηση του λεξικού εισάγει κωδικούς παραγωγικών προθημάτων και επιθημάτων με τα οποία μπορούν να συνδυαστούν.

$$(76.) \quad \left( \begin{array}{l} \text{ΓΚ: όνομα} \\ \text{Γένος: αρσενικό} \\ \text{Πτώση: [ ]} \\ \text{Αριθμός: [ ]} \\ \text{ΚΤ: 1} \\ \text{ΚΠΠ: \{1, 2, 3\}} \\ \text{ΚΠΕ: \{5H, 1A, 6B\}} \end{array} \right) \quad \begin{array}{l} /συν/ : \quad [ \text{ΚΠΠ: 1} ] \\ \\ /ότητα/ : \quad \left( \begin{array}{l} \text{A\_ΓΚ: όνομα} \\ \text{A\_ΓΕΝ: θηλυκό} \\ \text{A\_ΚΤ: 3} \\ \text{ΚΠΕ: 5H} \end{array} \right) \end{array}$$

(από Μαρκόπουλο 1998: 82-83)

Ο τρόπος κωδικοποίησης των παραγωγικών επιθημάτων (ΚΠΕ) και προθημάτων (ΚΠΠ) είναι απαραίτητος για τον μετατροπέα πεπερασμένων καταστάσεων, όπως και η δέσμη των πληροφοριών που φέρουν τα συγκεκριμένα προσφύματα, όπως πιθανή αλλαγή κατηγορίας (A\_ΓΚ), πιθανή αλλαγή γένους (A\_ΓΕΝ), καθώς και αλλαγή κλιτικής τάξης (A\_ΚΤ). Ωστόσο η αντιστοίχιση των κωδικών ΚΠΕ μεταξύ θεμάτων και επιθημάτων είναι προβληματική για τις περιπτώσεις παράγωγων συνθέτων, όπως *παιδεραστής* > *παιδεραστία*, όπου αντίστοιχα δεν υφίσταται το ζεύγος *εραστής* > *\*εραστία*. Παράλληλα καθίσταται το σύστημα εξαιρετικά πολύπλοκο και αντιοικονομικό με την εισαγωγή τέτοιων κωδικοποιήσεων, καθότι δεν υπάρχουν κανονικότητες τύπου *όλα τα ουσιαστικά της X κλιτικής τάξης συνδυάζονται πάντα με το X παραγωγικό επίθημα* και επομένως αυτές οι αντιστοιχίες πρέπει να γίνουν χειρωνακτικά και μετά από σχολαστικό έλεγχο.

Οι Galiotou και Ralli (2002, 2003, 2005, 2007, Ralli & Galiotou 2005) δοκίμασαν το PC-KIMMO για την υπολογιστική επεξεργασία των συνθέτων στα Ελληνικά. Παρόλη την επιτυχία στην ανάλυση διαφόρων τύπων συνθέτων, κατά την επεξεργασία τους αντιμετώπισαν συγκεκριμένα προβλήματα που οφείλονται κυρίως στην προσέγγιση των αυτομάτων πεπερασμένων καταστάσεων. Τα προβλήματα που συνάντησαν κατά την υλοποίηση της επεξεργασίας των Ελληνικών συνθέτων ήταν ο τόνος και το μαρκάρισμα των χαρακτηριστικών. Για τον τόνο αναφέρουν (Galiotou & Ralli 2003: 4) χαρακτηριστικά ότι:

*«Σύμφωνα με τον Antworth (1990) “τα υπερτεμαχιακά στοιχεία όπως ο τονισμός, η διάρκεια και ο μουσικός τόνος πρέπει να αναπαριστώνται με σύμβολα, τα οποία*

διασπείρονται ανάμεσα στα υπόλοιπα σύμβολα στο ίδιο επίπεδο. Επομένως, ο τονισμός και η μετακίνηση του τόνου αναπαριστώνται με τη χρήση τελεστών στο λεξικό επίπεδο, το οποίο χαρτογραφείται με το σύμβολο μηδέν (0) στο επιφανειακό επίπεδο. Αυτή η λύση αποδεικνύεται ιδιαίτερα προβληματική για τα Ελληνικά, εφόσον με στόχο την κατάλληλη αντιμετώπισή της διαδικασίας τονισμού της προπαραλήγουσας η συλλαβικοποίηση είναι απαραίτητη, κάτι που όμως δεν είναι δυνατό με το λογισμικό του PC-KIMMO».

Ταυτόχρονα, υπογραμμίζουν ότι οι δομές των χαρακτηριστικών σε συνδυασμό με τον μηχανισμό ενοποίησης επιτρέπει την τέλεση περασμάτων χαρακτηριστικών (feature passing) για τη διαφύλαξη των ορθώς σχηματισμένων λέξεων. Εντούτοις, δεν δίνεται η δυνατότητα απευθείας σύνδεσης φωνολογίας και γραμματικής, καθιστώντας αδύνατο τον ορισμό φωνολογικών κανόνων που δεσμεύονται από δομικές ιδιότητες και γλωσσικά συστημικά, όπως για παράδειγμα η αντιμετώπιση του συνδετικού φωνήεντος σε παρατακτικά σύνθετα τύπου *αγγλοαμερικάνος* που αποτυγχάνει να τα αναλύσει (βλ. Galiotou & Ralli 2005: 136).

#### 4.4. Μηχανική Μάθηση της Μορφολογίας

Τα πρώτα βήματα της ανάλυσης λέξεων έγιναν στο πλαίσιο της υπολογιστικής σύνταξης με μορφολογικούς αναλυτές βασισμένους σε σύνολα χειροκίνητων κανόνων. Ωστόσο με τη μετατόπιση της υπολογιστικής έρευνας της ανάλυσης λέξεων στον τομέα της μορφολογίας, οι αλλαγές στις ερευνητικές κατευθύνσεις ήταν σημαντικές. Στην τελευταία δεκαετία του προηγούμενου αιώνα το ενδιαφέρον επικεντρώθηκε σε στατιστικές μεθόδους μορφολογικής ανάλυσης που συνδυάζουν τις πιθανότητες με κανόνες ή βάρη (weights) σχετικά με τα χαρακτηριστικά. Σύμφωνα με τους Roark & Sproat (2007: 137) η απουσία σωμάτων κειμένων με πλήρη μορφολογικό χαρακτηρισμό των λέξεων – κάτι αντίστοιχο του περίφημου συντακτικού σώματος κειμένων Penn Treebank<sup>6</sup> – απέτρεπε τους ερευνητές να δημιουργήσουν έναν μορφολογικό αναλυτή ικανό να συναγωνιστεί τον ιδιαίτερα πολύπλοκο μοντέλο του Koskenniemi (1983). Παράλληλα μία βασική αιτία κρύβεται πίσω από το γεγονός ότι η ικανοποιητικός καλή απόδοση των αναλυτών με «χειροποίητους» μορφολογικούς

---

<sup>6</sup> Για το περίφημο σώμα κειμένων (corpus) του Penn Treebank που περιλαμβάνει πλήθος γλωσσών συντακτικά αναλυμένες σε δεντρικές δομές, βλ. στην ιστοσελίδα <http://www.cis.upenn.edu/~treebank/> και στο Wallis (2008).

κανόνες εμπόδιζε τους ερευνητές να οδηγηθούν σε στατιστικές προσεγγίσεις· σε αντίθεση με τη σύνταξη, θεωρούν ότι είναι πιθανό να καλυφθεί μεγάλος μέρος της μορφολογίας, ακόμα και για μορφολογικά πολύπλοκες γλώσσες, όπως τα Φινλανδικά (και τα Ελληνικά) με υπολογιστικούς κανόνες «*γραμμένους στο χέρι*». Πρόσφατα, εντούτοις, το ενδιαφέρον επικεντρώθηκε στις στατιστικές αναλύσεις της μορφολογίας και της μορφολογικής εκμάθησης και πιο συγκεκριμένα στην **μη-εποπτευόμενη** (unsupervised) ή **ελαφρώς εποπτευόμενη** (lightly supervised) εκμάθηση από σώματα κειμένων. Οι συγκεκριμένες τακτικές εφαρμόστηκαν αρχικά σε συγκολλητικές γλώσσες. Ορθώς (όπως θα φανεί και από το πείραμα του επόμενου κεφαλαίου) οι Roark & Sproat (2007: 177) υπογραμμίζουν ότι οι συγκολλητικές γλώσσες, όπως τα Κορεατικά και Τουρκικά, είναι ιδανικές υποψήφιες για μία τέτοια προσέγγιση, μίας και αποτελούν μία μακρά ακολουθία μορφημάτων που υπόκειται σε *λεξικοσυντακτικούς* περιορισμούς και κάθε μόρφημα αντιστοιχεί σε σθεναρή συγκεκριμένη δέσμη μορφοσυντακτικών χαρακτηριστικών. Αντιθέτως, τέτοιες προσεγγίσεις δύσκολα μπορούν να αποδώσουν ικανοποιητικώς θετικά αποτελέσματα σε γλώσσες με πλούσιο κλιτικό σύστημα.

Τα τελευταία χρόνια η προσπάθεια έχει επικεντρωθεί σχεδόν εξ ολοκλήρου σε συστήματα εποπτευόμενης ή μη μορφολογικής ανάλυσης· η έρευνα είναι πλουσιότατη και ρηξικέλευθη με ιδιαίτερα ικανοποιητικά αποτελέσματα, τα οποία εντούτοις δεν δύνανται να αποτελέσουν το μεγάλο αντίπαλο δέος που θα επικρατήσει έναντι της μορφολογικής ανάλυσης της μεθόδου του PC-KIMMO. Βασικός άξονας δράσης των αυτόματων μεθόδων είναι η ανακάλυψη των μορφολογικών διαφοροποιήσεων με τις μη-επιβλεπόμενες μεθόδους να κυριαρχούν.

#### 4.4.1. Μη-Εποπτευόμενη Μορφολογική Μάθηση

##### 4.4.1.1. Εισαγωγή

Κρίνεται απαραίτητο να δοθεί ο ορισμός του αντικειμένου της **Αυτόματης Μορφολογικής Μάθησης** (Automated Morphology Learning). Προσεγγίζοντας γενικά το ζήτημα της μορφολογικής εκμάθησης υφίστανται δύο βασικοί στόχοι που πρέπει να καλυφθούν. Ο πρώτος στόχος αποτελεί και τη θεμελιώδη αρχή του κλάδου της **Επεξεργασίας Φυσικής Γλώσσας** (Natural Language Processing [NLP]) και είναι η ανακάλυψη μέσω σώματος κειμένων των εναλλακτικών μορφών ενός μορφήματος/μίας λέξης και της σχέσης μεταξύ των· για παράδειγμα να εντοπισθούν οι διάφορες

μορφές της λέξης *γράφω*, όπως είναι *γράφεις*, *έγραψα*, *γράφουμε*, *γράφθηκε*, *γραμμένος*. Ο σκοπός της εύρεσης ενός συνόλου συσχετισμένων μορφολογικά φορμών δεν αφορά, όμως, μόνο λέξεις του κλιτικού συστήματος, αλλά και παράγωγες και σύνθετες λέξεις. Ο δεύτερος στόχος – και για τον τομέα της **Παραγωγής Φυσικής Γλώσσας** (Natural Language Generation [NLG]) – είναι η εκμάθηση των διαφόρων προσφυμάτων από το σώμα κειμένων, ώστε το σύστημα να μπορεί να παράγει νέες λέξεις που δεν έχει συναντήσει στο δείγμα του. Έτσι για παράδειγμα, ο χαρακτηρισμός του επιθήματος – *μέν(ος)* ως μετοχή ρήματος θα οδηγήσει στη δημιουργία μετοχής με άλλα ρήματα, όπως με το ρήμα *χτυπώ* (και την απαραίτητη επιλογή αλλομόρφου) για το σχηματισμό του *χτυπημένος*. Από μορφολογικής άποψης ο σκοπός είναι να καταλήξει σε ένα σύνολο κανόνων για την παραγωγή νέων λεξικών μορφών που δεν έχουν ήδη απαντηθεί.

Σαφώς επηρεασμένοι από τις σύγχρονες τάσεις της γλωσσικής κατάκτησης (language acquisition) και της ψυχολinguistics, οι ερευνητές που επικεντρώθηκαν στην αυτόματη μάθηση μορφολογίας υλοποίησαν τις ιδέες τους είτε δημιουργώντας γενικεύσεις μέσω κανόνων είτε λειτουργώντας υπό το πρίσμα της αναλογίας και των αναλογικών σχηματισμών<sup>7</sup>. Τα πρώτα δείγματα της αυτόματης μάθησης της μορφολογίας έγιναν από τους Rumelhart & McClelland (1986) που πρότειναν ένα θεωρητικό πλαίσιο διασυνδεδετισμού, όπου τα ρήματα δίνονται σε ζευγάρια τύπων (ενεστώτας και αόριστος), για να οδηγηθούν σε μία γενίκευση από τα συγκεκριμένα ρήματα σε ρήματα που δεν έχουν συναντήσει. Παράλληλα, ο Goldsmith (2001) θεωρεί άξιες αναφοράς άλλες έρευνες που έλαβαν χώρα εκείνη την περίοδο πάνω στην εποπτευόμενη εκμάθηση της μορφολογίας, όπως αυτές των Pacak & Pratt (1976), Koch, Küstner & Rüdiger (1989) και Wothke & Schmidt (1992). Κριτική ασκήθηκε κατά καιρούς σε όλες τις παραπάνω τεχνικές όσον αφορά στην τεχνική γενίκευσης και στα προβλήματα που προέκυπταν από τη λανθασμένη γενίκευση.

Η μη-εποπτευόμενη μορφολογική μάθηση συχνά έρχεται αντιμέτωπη με την «ιδιότητα» των μορφημάτων να αλλάζουν μορφή είτε εξαιτίας μορφολογικών ή φωνολογικών κανόνων, είτε για λόγους αυθαίρετους και απρόβλεπτους. Άμεσος στόχος της είναι να εντοπίσει και να κατηγοριοποιήσει τις διαφοροποιήσεις των

<sup>7</sup> Το συγκεκριμένο ζήτημα ανάγεται στην γενικότερη προβληματική για τον τρόπο εκμάθησης και κατάκτησης μίας γλώσσας από τους φυσικούς ομιλητές. Παλαιότερες συζητήσεις και διαφωνίες που αποτέλεσαν προπομπή της μη-εποπτευόμενης εκμάθησης της μορφολογίας έγιναν από τους Rumelhart & McClelland (1986) και Pinker & Parker (1988)· οι σύγχρονες τάσεις συζητούνται αναλυτικότερα από τους Pirrelli (2007a, 2007b), Cambell & Yang (2007), Yvon & Stroppa (2007) και Westermann & Plunkett (2007).

μορφημάτων σε ένα συγκεκριμένο σώμα κειμένων· ταυτόχρονα, όμως, πρέπει να εξάγει γενικεύσεις από τα μορφήματα. Αυτό είναι εξαιρετικά δύσκολο και πολύπλοκο ζήτημα που αντιμετωπίζουν επιδερμικά οι τεχνικές της μη-εποπτευόμενης εκμάθησης, γιατί οι απρόβλεπτες αλλαγές της αλλομορφίας απέχουν από την τεχνική ανάλυση των συγκεκριμένων μοντέλων. Οι Roark & Sproat (2007) προτείνουν ότι είναι λογικό να εφαρμόζονται εκ των υστέρων επιβλεπόμενες τεχνικές για την γενίκευση μεταξύ μορφημάτων. Ωστόσο, και για τους δύο στόχους η αλλομορφία δημιουργεί σοβαρά προβλήματα. Για την αντιμετώπιση της αλλομορφίας, επικεντρώνεται το ενδιαφέρον στην εύρεση συσχετισμένων μορφολογικών μορφών της ίδιας λέξης, όπως *κύμα* και *κύματα* (kima~ kimat), που δεν έχουν παραχθεί από κανέναν φωνο- και μορφολογικό κανόνα.

Η πλειονότητα της πρόσφατης έρευνας στοχεύει στο πεδίο της μη-εποπτευόμενης εκμάθησης της μορφολογίας και ειδικότερα την τελευταία δεκαετία με πρωτεργάτη τον Goldsmith (2001), του οποίου η προσέγγιση εξακολουθεί και θεωρείται *state-of-the-art*. Άλλες σημαντικές έρευνες είναι εκείνες των Yarowsky & Wicentowski (2001), Schone & Jurafsky (2001), Creutz & Lagus (2002) και Johnson & Martin (2003). Προτού παρουσιάσουμε αναλυτικότερα (βλ. ενότητα 4.4.1.3) την ρηξικέλυθη έρευνα του Goldsmith (2001), θα παραθέσουμε επιγραμματικά τις υπόλοιπες γενικές προσεγγίσεις που προηγήθηκαν.

#### 4.4.1.2. Διάφορες προσεγγίσεις Μη-Εποπτευόμενης Μορφολογικής Μάθησης

Η έρευνα στη Μη-Εποπτευόμενη Μορφολογική Μάθηση (M-EMM) έχει χωριστεί σε τέσσερις μεγάλες κατηγορίες που αποτέλεσαν τον προπομπό του συγκεκριμένου τομέα<sup>8</sup>. Η *πρώτη προσέγγιση* προσπαθεί να αναγνωρίσει τα όρια μορφημάτων και να τα κατηγοριοποιήσει σε θέματα, επιθήματα και προθήματα, όπως υποστηρίχτηκε από τους Harris (1955, 1967) και Hafer & Weiss (1974). Οι συγκεκριμένοι συγγραφείς προσπαθούν με την προβλέψιμη εντροπία προθήματος/ επιθήματος να τοποθετήσουν όρια μέσα στις λέξεις, να καθορίσουν μορφήματα και τέλος να τα χαρακτηρίσουν κατηγοριακά. Η *δεύτερη προσέγγιση* προτείνει τα διγράμματα (digrams) και τα τριγράμματα (trigrams), τα οποία είναι μέρος της μορφής του μορφήματος (πρβλ. Janssen 1992, Klenk 1992 και Flenner 1994, 1995). Σύμφωνα με αυτή την προσέγγιση,

---

<sup>8</sup> Ο παραπάνω διαχωρισμός έγινε από τον ίδιο τον Goldsmith (2001).

η τοπική πληροφορία που αποτυπώνεται από ένα σύνολο πιθανοτήτων και συχνοτήτων εμφάνισης και γειτνίασης σε κάθε ακολουθία φωνημάτων (δύο ή τριών αναλόγως την προσέγγιση, εξού και τα δι-/τριγράμματα) είναι επαρκής για να τοποθετηθούν τα όρια μορφημάτων. Η  *τρίτη προσέγγιση* στοχεύει στην αξιοποίηση του προτύπου των φωνολογικών σχέσεων ανάμεσα σε ζεύγη συσχετισμένων λέξεων, όπως δόθηκε από τους Dzeroski & Etjavec (1997). Στόχος τους είναι να ανακαλυφθούν τα μοτίβα που κρύβονται μέσα σε μία λέξη βασισμένη σε μορφολογικές αξίες. Στην ακόλουθη ενότητα θα συζητηθεί η τέταρτη προσέγγιση, αυτή του Goldsmith (2001), που είναι βασισμένη στο θεωρητικό πλαίσιο του Rissanen (1989) του **Ελαχίστου Μήκους Περιγραφής** (Minimum Description Length - MDL).

#### 4.4.1.3. Η προσέγγιση του Goldsmith (2001)

Η υλοποίηση της θεωρίας του Ελαχίστου Μήκους Περιγραφής από τον Goldsmith ονομάζεται *Linguistica*<sup>9</sup> και είναι πρόγραμμα εντοπισμού θεμάτων/ ριζών και διαφοροποιήσεων της προσφουματοποίησης. Στις μέρες μας θεωρείται ως το πρόγραμμα με τις περισσότερες ετεροαναφορές για την αυτόματη μορφολογική κατάκτηση. Αποτελεί ένα *state-of-the-art* πρόγραμμα για την μη-εποπτευόμενη μορφολογική μάθηση αποτελώντας φυσικά το *gold standard* για τις συγκρίσεις και τον συναγωνισμό με άλλα συστήματα και/ ή προγράμματα.

Το σύστημα του Goldsmith βασίζεται για την ανάλυσή του σε ένα τεράστιο corpus μη-σχολιασμένων/ μη-μεταγραμμένων κειμένων (unannotated corpora) και στοχεύει στον τεμαχισμό των λέξεων με τρόπο που θα εγγίζει την ανάλυση ενός πραγματικού μορφολόγου. Κατά την επεξεργασία των δεδομένων παράγει μία ομάδα **φύλλων σώματος** (signatures) μαζί με λέξεις που ανήκουν σε αυτά τα φύλλα. Ένα *φύλλο σώματος* είναι μία ομάδα από προσφύματα (είτε προθήματα, είτε επιθήματα) που συνδυάζονται με μία δοσμένη ομάδα ριζών ή θεμάτων (Goldsmith 2001, Roark & Sproat 2007). Για παράδειγμα, ένα φύλλο επιθημάτων στα αγγλικά μπορεί να είναι το *NULL.ed.ing.s*, το οποίο συνδυάζεται με θέματα όπως *jump* (πηδώ), *laugh* (γελώ), *walk* (περπατώ), *move* (μετακινώ), *prove* (αποδεικνύω) κ.α.: όλα αυτά τα θέματα παίρνουν τα επιθήματα του παραπάνω φύλλου, για να δημιουργήσουν λέξεις, όπως *jumpρω*,

<sup>9</sup> Το πρόγραμμα του LINGUISTICA διατίθεται δωρεάν στην επίσημη ιστοσελίδα της ομάδας του Goldsmith στο <http://linguistica.uchicago.edu/>. Υποστηρίζεται και από παλαιότερες εκδόσεις, καθώς και από συνοδευτικό υλικό. Αναλυτικότερα θα αναφερθούμε στο επόμενο κεφάλαιο και στο πρώτο πείραμα.

*jumped*, *jumping* και *jumps* (αντίστοιχα και τα υπόλοιπα ρήματα). Άλλα φύλλα σώματος είναι τα *e.ed.ing*, *NULL.s*, *NULL.ing.s*, *NULL.er.est.ly* μεταξύ άλλων. Όπως μπορούμε να παρατηρήσουμε τα φύλλα σώματος είναι όπως τα παραδείγματα λέξεων, που συνήθως περιέχουν μαζί κλιτικά και παραγωγικά επιθήματα. Επομένως, το βασικό σχήμα για το πώς λειτουργούν αυτά τα φύλλα σώματος είναι το (77.):

(77.)

$$\left\{ \begin{array}{l} \Theta\text{EMA}_1 \\ \Theta\text{EMA}_2 \\ \Theta\text{EMA}_3 \\ \Theta\text{EMA}_4 \end{array} \right\} \quad \left\{ \begin{array}{l} \text{ΕΠΙΘΗΜΑ}_1 \\ \text{ΕΠΙΘΗΜΑ}_2 \end{array} \right\}$$

όπου τα μέλη του φύλλου των θεμάτων συνδέονται με όλα τα μέλη του άλλου φύλλου. Τα προσφύματα μπορούν να μετέχουν σε πλείονα του ενός φύλλα. Το σύστημα στοχεύει παράλληλα στην αντιμετώπιση των περιπτώσεων που μετά από κάποιο επίθημα ακολουθεί και ένα δεύτερο επίθημα (π.χ. αγγ. *measure-ment-s*, *turn-ing-s*, ελλ. *χορ-ευ-ω*, *ομορφ-ότερ-ος*). Να σημειωθεί ότι οι μονομορφηματικές λέξεις αντιμετωπίζονται ξεχωριστά και προσμετρώνται σε διαφορετικό σύνολο, από τις υπόλοιπες μορφολογικά αναλυμένες λέξεις.

Εξετάζοντας πιο προσεκτικά τα φύλλα σώματος, παρατηρείται πως μερικά σύνολα δεν είναι πάντα ολοκληρωμένα: συνήθως τα κλιτικά επιθήματα παρελθόντων χρόνων απουσιάζουν από τα φύλλα σώματος, ακόμα και από τα ρήματα που σχηματίζουν ομαλώς τους παρελθοντικούς τύπους. Πιο συγκεκριμένα, θα αναλυθεί ένα από τα παραδείγματα που εντοπίστηκαν από τους Roark & Sproat (2007: 120). Επισημαίνουν ότι το φύλο σώματος *NULL.er.ing.s* δοσμένο από τον Goldsmith (2001: 179), το οποίο περιλαμβάνει θέματα, όπως *blow* (φυσώ), *broadcast* (μεταδίδω), *drink* (πίνω), *dwell* (κατοικώ), *feel* (αισθάνομαι), δεν εσωκλείει το επίθημα *-ed*, μίας και τα ρήματα εμφανίζουν αλλομορφία στον αοριστικό τύπο και τη μετοχή (ανώμαλα ρήματα κατά τους συγγραφείς). Εντούτοις, το επίθημα *-ed* απουσιάζει επίσης από φύλλα σώματος με ρηματικά θέματα, όπως *bomb* (βομβαρδίζω) και *farm* (καλλιεργώ), τα οποία έχουν ομαλό τύπο αορίστου/ παθητικής μετοχής (*bombed* και *farmed*), αλλά οι συγκεκριμένοι τύποι δεν εμφανίζονται στο σώμα κειμένων του(!). Ο Goldsmith παραθέτει μερικά προβλήματα με τα φύλλα σώματος και σημειώνει ότι το σύστημά του δεν είναι ικανό

να αντιμετωπίσει διαφοροποιήσεις (π.χ. *αλλόμορφα*), όπως *feel/ felt*, μίας και το σύστημά του ασχολείται μόνο με την προσφύματοποίηση.<sup>10</sup>

Το συμπιεσμένο μήκος ενός σώματος κειμένου καθορίζεται τόσο από τα θέματα και τα επιθήματα, όσο και από το συσχετισμό τους με τα φύλλα σώματος και δίνεται από τον ακόλουθο τύπο:

$$(78.) \sum_{w=t+f} [\log prob(\sigma(w)) + \log prob(t) + \log prob(f | \sigma(w))]^{11}$$

Τα φύλλα σώματος που παράγονται από το σύστημα βασίζονται στις δύο θεμελιώδεις στρατηγικές της θεωρίας που είναι η *παραγωγή υποψηφίων* (candidate generation) και η *αξιολόγηση υποψηφίων* (candidate evaluation). Τα παρακάτω παρουσιάζονται αναλυτικότερα στην επόμενη ενότητα.<sup>12</sup>

#### 4.4.1.3.1. Δημιουργία Υποψηφίων Φύλλων σώματος

Η δημιουργία των φύλλων σώματος περιλαμβάνει δύο βήματα: πρώτα το σύστημα παράγει μερικά υποψήφια φύλλα σώματος (εντάσσοντας μέλη σ' αυτά) και στη συνέχεια αξιολογεί τα υποψήφια, ώστε να προκύψουν τα πραγματικά. Για την παραγωγή υποψηφίων, η μέθοδος διαχωρισμού λέξεων είναι βασισμένη στη **σταθμισμένη αμοιβαία πληροφορία** (weighted mutual information) έχοντας προηγουμένως δοκιμάσει διάφορες ευρετικές στρατηγικές<sup>13</sup> (βλ. Goldsmith 2001: 170-173). Αυτή η μέθοδος αρχίζει να δημιουργεί μία λίστα προσφυμάτων (κυρίως επιθημάτων), ένα αντίστροφο λεξικό (αλφαβητίζοντας από το δεξί μέρος των λέξεων) και χτίζει σύνολα πιθανών επιθημάτων μήκους έως και έξι χαρακτήρων (για παράδειγμα, -ούτσικ[ος]/ μικρ#ουτσικ#ος, όπου η # οριοθετεί τα όρια). Έπειτα χρησιμοποιεί έναν αλγόριθμο που ζυγίζει/ σταθμίζει όλα τα πιθανά επιθήματα για να εντοπίσει τα πραγματικά επιθήματα και τα ομαδοποιεί έπειτα σε φύλλα σώματος.

<sup>10</sup> ΣΗΜΕΙΩΣΗ: όπως θα δούμε στο επόμενο κεφάλαιο και στο πείραμα με το LINGUISTICA, αυτού του είδους η αλλομορφική διαφοροποίηση θα αποτελέσει τεράστιο πρόβλημα αν προσπαθήσουμε να υλοποιήσουμε ένα Μοντέλο Μη-εποπτευόμενης Μορφολογικής Μάθησης (MeMM) εφαρμοσμένο στην Ελληνική γλώσσα, μίας και η αλλομορφία στα Ελληνικά είναι πραγματικά πολύπλοκη και εμφανίζεται σε κάθε επίπεδο σχηματισμού λέξης (κλίση, παραγωγή, σύνθεση)

<sup>11</sup> Όπου  $w$  = αριθμός λέξεων,  $t$  = αριθμός θεμάτων,  $\sigma$  = αριθμός φύλλων σώματος και  $f$  = αριθμός επιθημάτων.

<sup>12</sup> Τα ακόλουθα αποτελούν μία δημιουργική σύμπτυξη από το άρθρο του Goldsmith (2001) και την αντίστοιχη παρουσίαση των Roark & Sproat (2007: 121-124).

<sup>13</sup> Περισσότερες πληροφορίες για τις ευρετικές (heuristics) στην Τεχνητή Νοημοσύνη και την Επεξεργασία Φυσικής Γλώσσας, δείτε στο βιβλίο Russell, Stuart & Norvig, Peter (2003). *Artificial Intelligence: A modern approach* (2<sup>nd</sup> ed.). New Jersey: Prentice Hall.



Ακολουθώς οι λέξεις του σώματος κειμένων αναλύονται σύμφωνα με τα υποψήφια επιθήματα και εάν είναι δυνατόν, διαχωρίζονται οι λέξεις με τον καλύτερο τρόπο σύμφωνα με την μετρική εξίσωση που δίνεται παρακάτω, η οποία εκχωρεί μία πιθανότητα στην ανάλυση της λέξης  $w$  σε ένα θέμα  $w_{l,i}$  και ένα επίθημα  $w_{l+i,l}$ , όπου το  $l$  συμβολίζει το μήκος της λέξης:

$$(79.) P(w = w_{1,i} + w_{i+1,l}) = \frac{1}{\sum_{j=1}^{l-1} H(w_{1,i}, w_{i+1,l})} e^{-H(w_{1,i}, w_{i+1,l})}$$

όπου:

$$(80.) H(w_{1,i}, w_{i+1,l}) = -i \log(\text{freq}(\text{stem} = w_{1,i})) + (l-i) \log(\text{freq}(\text{suffix} = w_{i+1,l}))$$

Στο τέλος ελέγχονται όλα τα επιθήματα, ώστε να διαγραφούν όλα όσα δεν μπορούν να εκπληρώσουν την ανάλυση έστω μίας λέξης. Το αποτέλεσμα είναι η απόδοση ενός συνόλου θεμάτων που συνδυάζονται με ένα σύνολο επιθημάτων εσωκλείοντας και το μηδενικό μόρφημα. Η αλφαριθμητική λίστα των επιθημάτων σχετίζεται με κάθε θέμα συντάσσοντας το αντίστοιχο φύλλο σώματος. Σε αυτό το σημείο επιπλέον ευρετικές τακτικές ενεργούν, ώστε να περικόψουν όλα τα φύλλα σώματος που αντιστοιχούν μόνο σε ένα θέμα ή τα φύλλα σώματος που αποτελούνται από ένα επίθημα. Μετά και από αυτό το ξεσκαρτάρισμα τα εναπομείναντα φύλλα σώματος με τουλάχιστον δύο επιθήματα ως μέλη χαρακτηρίζονται από τον Goldsmith ως **κανονικά φύλλα σώματος** (regular signatures).

#### 4.4.1.3.2. Αξιολόγηση Υποψηφίων (και τα διαγλωσσικά πειράματα)

Μετά την παραγωγή των κανονικών φύλλων σώματος, ο Goldsmith (2001) προτείνει μία μετρική αξιολόγησης βασισμένη στη θεωρία του **ελαχίστου μήκους περιγραφής** του Rissanen (1989), όπου η καλύτερη πρόταση για τα φύλλα σώματος είναι αυτή που περιλαμβάνει την πιο συμπαγή περιγραφή της γλώσσας του σώματος κειμένων.

Τα κανονικά φύλλα σώματος που προέκυψαν από το τελικό ξεσκαρτάρισμα της παραγωγής υποψηφίων μαζί με τα συσχετιζόμενα θέματα αποτελούν την πρόταση του μοντέλου της MeMM για το διαχωρισμό των λέξεων της γλώσσας σε θέματα και επιθήματα: ωστόσο, όπως έχει σημειωθεί η συγκεκριμένη πρόταση πρέπει να αξιολογηθεί, ώστε να επικρατήσουν τα χρήσιμα και απαραίτητα φύλλα, εφόσον δεν είναι δυνατόν όλες οι αναλύσεις λέξεων να είναι ορθές. Έτσι ο Goldsmith βασισμένος στο ελάχιστο μήκος περιγραφής αξιοποιεί μία μετρική αξιολόγησης που θα του

επιτρέψει να έχει μία πιο συγκροτημένη και συμπυκνωμένη περιγραφή της ίδιας της μορφολογίας της γλώσσας. Αυτή η μετρική υλοποιείται με έναν αλγόριθμο, ο οποίος ελαχιστοποιεί το μέγεθος του συμπιεσμένου κειμένου, καθώς και το μέγεθος του μοντέλου περιγράφει την κωδικοποίηση του κειμένου:

$$(81.) \lambda\langle T \rangle + \lambda\langle F \rangle + \lambda\langle \Sigma \rangle$$

Όπου εδώ το  $\lambda\langle T \rangle$  είναι το μήκος της λίστας των δεικτών που έχουν τοποθετηθεί σε  $\langle T \rangle$  θέματα, με το  $T$  να συμβολίζει το σύνολο των θεμάτων και τα  $\langle \rangle$  το πλήθος των στοιχείων που εντοπίζονται στο συγκεκριμένο σύνολο. Αντίστοιχα τα  $\lambda\langle F \rangle$  και  $\lambda\langle \Sigma \rangle$  παρουσιάζουν τα μήκη των λιστών των δεικτών για τα σύνολα των επιθημάτων και φύλλων σωμάτων.

Το κάθε μήκος έχει προκύψει από αλγόριθμους που υπολογίζουν την ολική άθροιση κάποιων μεταβλητών. Πιο συγκεκριμένα, το  $\lambda\langle T \rangle$  δίνεται από τον ακόλουθο:

$$(82.) \sum_{t \in T} (\log(26) * length(t) + \log \frac{[W]}{[t]})$$

Όπου το  $T$  είναι το σύνολο των θεμάτων, το  $[W]$  είναι το πλήθος των λέξεων – δειγμάτων (word tokens) και το  $[t]$  είναι το πλήθος των δειγμάτων του συγκεκριμένου θέματος  $t$ : από τούδε και στο εξής το  $[X]$  θα σημαίνει το πλήθος δειγμάτων του  $X$  στο σώμα κειμένων, ενώ ο παράγοντας  $\log(26)$  προκύπτει από το πλήθος των συμβόλων του λατινικού αλφαβήτου (παρόλο που το Ελληνικό διαθέτει 24).

Αντίστοιχα για το  $\lambda\langle F \rangle$  ο αλγόριθμος δίνεται από την ακόλουθη εξίσωση:

$$(83.) \sum_{f \in suffixes} (\log(26) * length(f) + \log \frac{[W_A]}{[f]})$$

Όπου το  $[W_A]$  συμβολίζει το πλήθος των δειγμάτων των μορφολογικά αναλυμένων λέξεων, ενώ το  $[f]$  είναι το πλήθος των δειγμάτων εμφάνισης του επιθήματος  $f$  στο σώμα κειμένων.

Τέλος η σύσταση των φύλλων σωμάτων μπορεί να οριστεί απλώς από:

$$(84.) \sum_{\sigma \in \Sigma} \log \frac{[W]}{[\sigma]}$$

Όπου το  $\Sigma$  είναι το σύνολο των φύλλων σώματος.

Συνοψίζοντας, το συμπιεσμένο μήκος του σώματος κειμένων όσον αφορά στο μορφολογικό μοντέλο της γλώσσας βασισμένο στις τελικές τροποποιήσεις είναι το:

$$(85.) \sum_{w \in W} [w] \left[ \log \frac{[W]}{[\sigma[w]]} + \log \frac{[\sigma[w]]}{[\text{stem}(w)]} + \log \frac{[\sigma[w]]}{[\text{suffix}(w) \in \sigma(w)]} \right]$$

ή με άλλα λόγια η μέγιστη πιθανότητα παραγόμενη από τις πιθανότητες:

$$(86.) \sum_{w \in W} [w] [-\log(P(\sigma(w))) - \log(P(\text{stem}(w) | \sigma(w))) - \log(P(\text{suffix}(w) | \sigma(w)))]$$

Καταχωρώντας τις προβληματικές περιπτώσεις ο Goldsmith βελτιώνει τις ευρετικές, ώστε να δοκιμάσει το λογισμικό του σε διάφορες γλώσσες, όπως Αγγλικά, Ιταλικά, Γαλλικά, Ισπανικά και Λατινικά. Για κάθε μία γλώσσα από αυτές ο Goldsmith δίνει τα δέκα πιο συχνά φύλλα σώματος, εκ των οποίων αξιολογεί αναλυτικότερα τις αναλύσεις από τα αγγλικά και γαλλικά. Μάλιστα δημιουργεί κατηγορίες για κάθε ανάλυση και εντάσσει τα αποτελέσματα σε *καλά*, *λανθασμένως αναλυμένα*, *ψευδώς αναλυμένα* και *αποτυχόντα να αναλυθούν*. Μη έχοντας κάποιο αντίστοιχο πρόγραμμα για καθιερωμένο, ώστε να συγκρίνει τα αποτελέσματα και να εξάγει συγκριτικά συμπεράσματα, οδηγείται εκ των πραγμάτων σε μία «υποκειμενική» αξιολόγηση αλλά με αντικειμενικά κριτήρια. Τα αποτελέσματα είναι τα ακόλουθα:

Κατηγορίες	Αγγλικά	Γαλλικά
Good	82,9%	83,3%
Wrong Analysis	5,2%	6,1%
Spurious Analysis	8,3%	6,4%
Failed to analyze	3,6%	4,2%

Πίνακας 12: Τα ποσοστά επίδοσης του *Linguistica* για Αγγλικά και Γαλλικά από τη δοκιμή του Goldsmith (2001).

A/A	Αγγλικά	
1.	NULL.ed.ing.s	accent, add, administer, afford, alert, amount, appeal
2.	's.NULL.s	afternoon, airline, architect, assessor, association
3.	NULL.ed.er.ing.s	attack, back, bath, boil, borrow, charm, demand
	Γαλλικά	
1.	NULL.e.es.s	abondant, abstaint, adjacent, attaint, bantou, bleu
2.	NULL.s	abbaye, abdication, abdominale, aberration, abrasif
3.	NULL.ment.s	aggressive, ancienne, annuelle, chimique, classique
	Ισπανικά	
1.	a.as.o.os	abiert, ajen, amig, antigu, compuest, cortesan, cuy
2.	NULL.s	aborrecido, abrasado, abundante, achaque, acosado
3.	a.o.os	afligid, asalt, caballeriz, despiert, dorad, enemig, flac

Πίνακας 13: Το top-3 των φύλλων σωμάτων από το πείραμα του Goldsmith για τα Αγγλικά, Γαλλικά και Ισπανικά (2001: 179-182).

Όπως ορθώς υπογραμμίζουν οι Roark & Sproat (2007: 123) η μέθοδος του Goldsmith είναι “η *de facto gold standard* για έρευνα στην μη-εποπτευόμενη κατάκτηση στη μορφολογία”. σημαντικό ρόλο διαδραμάτισε η ελεύθερη διάθεση του λογισμικού του, η συνεχής υποστήριξή του, καθώς και η μετρική αξιολόγηση που αποτέλεσε πρότυπο για μετέπειτα συγκρίσεις και αναφορές λόγω της απλής μα ισχυρής δυναμικής της. Η απλότητα της δομικής στρατηγικής του συστήματος είναι και παραμένει αξιοσημείωτη, ενώ ταυτόχρονα το μεγάλο πλεονέκτημα, μα και μειονέκτημα του συστήματος. Η παντελής απουσία μορφολογικών και φωνολογικών κανόνων σε συνδυασμό με την έλλειψη συντακτικών και σημασιολογικών πληροφοριών επιτρέπουν το σύστημα να εφαρμοστεί χωρίς εξαίρεση σε κάθε γλώσσα. Εντούτοις, εδώ έγκειται και το πρόβλημα της αποτυχίας σε γλώσσες όπως η Ελληνική (βλ. περισσότερα στην επόμενη υποενότητα και κεφάλαιο).

Ο Goldsmith ακολουθώντας σύγχρονες τάσεις της υπολογιστικής γλωσσολογίας που επηρεάζονται από την ψυχολογολογία, υποστηρίζει την απουσία των παραπάνω πληροφοριών με το γεγονός ότι και ένα παιδί μπορεί να έχει πρόσβαση σε άλλες πληροφορίες, αλλά δεν το βοηθούν να κατακτήσει τη μορφολογία της γλώσσας του. Μάλιστα είναι πεπεισμένος ότι ο αλγόριθμός του δεν θα είχε κάποιο κέρδος στην επίλυση των προβλημάτων που επισημαίνει, εφόσον επωφελούνταν από συντακτικές

και σημασιολογικές πληροφορίες. Αντίθετα, οι Schone & Jurafsky (2001a, 2001b, 2001c) απέδειξαν πως οι συντακτικές και σημασιολογικές πληροφορίες δύνανται να συνδράμουν σημαντικά στη μορφολογική κατάκτηση.

#### 4.4.1.4. Κριτική μοντέλων Μη-Εποπτευόμενης Μορφολογικής Μάθησης

Ωστόσο, αυτό το σύστημα απέχει πολύ από το να χαρακτηριστεί τέλειο. Όπως παρατηρείται, ένα M-EMM δεν δέχεται κανένα μορφολογικό και φωνολογικό κανόνα, δεν περιέχει κανένα προ-δημιουργημένο λεξικό/ λεξιλόγιο και προφανώς δεν χρησιμοποιεί κανένα πλεονέκτημα από οποιαδήποτε (πιο συγκεκριμένα μορφολογική) θεωρία ή θεωρητικό πλαίσιο. Απλώς προσπαθεί να χωρίσει λέξεις με βάση τεράστια σώματα κειμένων με τη συνδρομή σύνθετων αλγόριθμων και των πιθανοτήτων. Διάφοροι ερευνητές (Roark & Sproat 2007, βλ. τόμος *Lingue e Linguaggio* 2007b) διαμαρτύρονται ότι η μέθοδος του Goldsmith δεν εκμεταλλεύεται καμία σημασιολογική και συντακτική πληροφορία: αυτή η διαφωνία κινείται στο ίδιο μήκος κύματος με την προσέγγιση της ψυχολογολογίας και την εναντίωσή της στο γεγονός ότι τα παιδιά και οι ενήλικες έχουν πρόσβαση και σε άλλες πληροφορίες πλην των ομάδων των επιθημάτων και προθημάτων. Ωστόσο, αναλογιζόμενοι ότι δεν γίνεται καμία χρήση μορφολογικού κανόνα ή θεωρίας από το σύστημα του Goldsmith, δεν πρέπει να αναμένουμε και τη χρήση οποιασδήποτε σημασιολογικής και συντακτικής πληροφορίας.

Η αποτυχία σωστής κατάτμησης των λέξεων στα πραγματικά μορφήματα οφείλεται στην απουσία μορφο- και φωνολογικών κανόνων, τη μη-χρήση της Λεξικής Φωνολογίας και την εμφάνιση σπανίων, μαρκαρισμένων και ανώμαλων περιπτώσεων. Αυτό μπορεί να συμβεί και στο ορθογραφικό και στο φωνολογικό επίπεδο λεξικής μεταγραφής:

(87.)	α. έτρεξα	>	ε – τρεχ – σ(α)	[ανομοίωση <sup>14</sup> ]
			θέμα: τρεχ	
	β. etrexa	>	e – treh – s(a)	
			θέμα: treh	

<sup>14</sup> Η ανομοίωση (dissimilation) είναι ένα φωνολογικό φαινόμενο, όπου παρόμοια σύμφωνα ή φωνήεντα χάνουν κάποιο κοινό χαρακτηριστικό, ώστε να είναι λιγότερα όμοια, όπως στο παραπάνω παράδειγμα τα φωνήεντα /x/ και /s/ είναι όμοια καθότι είναι τριβόμενα και το ένα φωνήεν μετατρέπεται σε κλειστό (/x/ σε /k/).

Η αποτυχία ανάλυσης λέξεων είναι στατιστικά αυξημένη σε μορφολογικά πλούσιες γλώσσες, όπως τα Ελληνικά, Φιλανδικά, Σουηδικά, Ουγγρικά και Τουρκικά. Η έντονη παραγωγικότητα της σύνθεσης και της παραγωγής περιπλέκει περισσότερο το ζήτημα με την εκτεταμένη παρουσία της συνδυαστικότητας των επιθημάτων. Σύμφωνα με τους Kurimo *et al.* (2007), το υψηλότερο σκορ σε ένα ΜεΜΜ μοντέλο αξιολόγησης για τα Φιλανδικά και τα Τουρκικά ήταν 65% και 64% αντιστοίχως και το χαμηλότερο 3% και 2%, παρότι κάποια συστήματα του διαγωνισμού του Kurimo *et al.* είχαν ελάχιστη βοήθεια από εποπτευόμενη μορφολογία. Αναμένουμε ότι τα αποτελέσματα στα Ελληνικά θα είναι χειρότερα λόγω της εκτεταμένης αλλομορφίας της γλώσσας (πρβλ. Καρασίμος 2001, Ράλλη 2005, 2007), όπως και λόγω των περίπλοκων συνδυαστικών ιδιοτήτων των επιθημάτων. Οι Μελισσαροπούλου (2007α, 2007β) και Melissaropoulou & Ralli (2008) σημειώνουν ότι στα Ελληνικά, πέντε παραγωγικά επιθήματα μπορούν να βρεθούν στη σειρά μέσα στην ίδια λέξη.<sup>15</sup>

(88.) χορευταρούλικο	>	χορ – ευ – τ – αρ – ουλ – ικ –(ο) θέμα–πε – πε –πε – πε – πε – πε (κε) [5 παρ.επ.(+1)]
κοινωνικότητα	>	κοιν – ων – ι – ικ – οτητα(θ) θέμα– πε– πε– πε– πε(κε) [4 παρ.επ. (+1)]
ποτιστικός	>	ποτ – ισ – τ – ικ(ος) θέμα–πε– πε– πε(κε) [3 παρ.επ. (+1)]
ξαναεπαναλαμβάνω	>	ξανά – επανα – λαμβάν(ω) ππ – ππ θέμα [2 παρ.πρ.]
παρασυμπαραστέκομαι	>	παρά – συν – παρα –στεκ(ομαι) ππ – ππ – ππ θέμα [3 παρ.πρ.]
συμπεριφέρομαι	>	συν – περί – φέρ(ομαι) ππ – ππ θέμα [2 παρ.πρ.]

Επιστρέφοντας στη θεωρία του Goldsmith ένα φύλλο σώματος είναι το σύνολο επιθημάτων που μπορεί να συνδυαστεί με ένα σύνολο θεμάτων. Επομένως, οφείλουμε να δημιουργήσουμε φύλλα σώματος για επιθήματα που να μπορούν να συνδυαστούν με άλλα φύλλα σώματος για επιθήματα. Είναι εύκολο να φανταστούμε πόσο πολύπλοκο μπορεί να γίνει ένα σύστημα με τέτοιο πλέγμα φύλλων σώματος για επιθήματα/ προθήματα· αντιθέτως, οι επιλεκτικοί περιορισμοί και οι συνδυαστικές

<sup>15</sup> ΠΕ = Παραγωγικό Επίθημα, ΚΕ = Κλιτικό Επίθημα, ΠΠ = Παραγωγικό Πρόθημα.

επιλογές των προσφυμάτων καθιστούν τη δημιουργία τέτοιων φύλλων σώματος σχεδόν αδύνατη ή πλήρως ατελής.

#### 4.4.2. Εποπτευόμενη Μορφολογική Μάθηση

Ιστορικά, τα υπολογιστικά μοντέλα της μορφολογίας κατά την υλοποίησή τους παρακινούνταν από δύο βασικούς στόχους, τη μοντελοποίηση στη ψυχολογία και την υποστήριξη των σκοπών της επεξεργασίας φυσικής γλώσσας. Οι προσεγγίσεις της Εποπτευόμενης Μορφολογικής Μάθησης θεωρούνται ότι παρέχουν περισσότερα πλεονεκτήματα από τις παλαιότερες θεωρίες: μεταξύ άλλων είναι δημοφιλείς οι **βασισμένες σε κανόνες** (rule-based), οι **στοχαστικές** (stochastic), οι **διασυνδεδετισμικές** (connectic). Η βασική ιδέα πίσω από κάθε υλοποίηση και πλαίσιο αυτής της προσέγγισης είναι η εξαγωγή κάποιων γενικευμένων προτύπων/ κανόνων/ συμπεριφορών από μία ομάδα δεδομένων εκπαίδευσης. Πιο συγκεκριμένα, περιγράφουν τη σχέση ανάμεσα στα δεδομένα εισόδου και τα εξαγόμενα αποτελέσματα που παρουσιάζονται σε ένα σύνολο παραδειγμάτων, καθώς και τον αλγόριθμο που μαθαίνει από τα δεδομένα εκπαίδευσης, ώστε να προβλέψει αυτό που θα του ζητηθεί από κάθε καινούργιο εισαγόμενο. Έτσι το σύστημα επεξεργάζεται τα εισαγόμενα για να παραχθούν τα εξαγόμενα, ώστε αργότερα να βελτιωθεί η επίδοση του συστήματος εξαιτίας της εμπειρίας που αποκτάται με την εκμάθηση. Μορφολογική ανάλυση σύμφωνα με το πλαίσιο αυτό είναι ο διαχωρισμός των διαφόρων παραδειγμάτων – λέξεων σε μία σειρά από μορφήματα (θέματα και προσφύματα) από τα οποία αποτελούνται. Να σημειωθεί ότι οι συγκεκριμένες προσεγγίσεις για να επιτύχουν το στόχο τους αξιοποιούν διάφορες γλωσσικές πληροφορίες – και όχι μόνο μορφολογικές.

Οι παράμετροι του εκάστοτε ζητήματος καθορίζουν σε γενικές γραμμές τη στρατηγική κάθε τεχνικής, καθότι αναπαρίστανται με διαφορετικό τρόπο οι γνώσεις κάθε γλωσσικού προβλήματος. Πρέπει να γίνει κατανοητό ότι το μέγεθος των δεδομένων εκπαίδευσης του συστήματος είναι απαραίτητο να είναι επαρκές, για αυτό και υπάρχουν μετρήσιμα μεγέθη, ώστε να καλύπτεται μεγάλο εύρος των παραδειγμάτων· παράλληλα το ποσοστό συνεισφοράς των λεξικών πόρων στην αποτελεσματικότητα και βελτίωση της τεχνικής, παρόλη την παρουσία «θορυβώδων» εισαγόμενων με χαμηλή συνέπεια (noise inputs) συμβάλλει στην χάραξη της τεχνικής (βλ. Κουτσομπόγερα 2004, Wicentowski 2004).

Στο πλαίσιο του **Διασυνδεδετισμού** (Connectism) οι Rumelhart & McClelland (1986), Pinker & Prince (1988), Sproat & Egedi (1988) και Plunkett & Juola (2001) προσπάθησαν να δημιουργήσουν ένα υπολογιστικό μοντέλο βασισμένο σε σημαντικό βαθμό στο ψυχολογολογικό φαινόμενο της παιδικής γλωσσικής κατάκτησης, έχοντας ως θεματικό άξονα την εκμάθηση των ρηματικών παρελθόντων χρόνων των Αγγλικών. Έχοντας δημιουργήσει ζευγάρια λέξεων με όμοια κλιτική συμπεριφορά εκπαίδευσαν συστήματα **νευρωνικών δικτύων** (neural networks – NN) που αποτελούσαν προσπάθεια μίμησης διάφορων καθοδηγητικών προτύπων εκμάθησης που παρατηρήθηκαν στα παιδιά. Ο Wicentowski (2002: 22) ορθώς επισημαίνει ότι οι συγκεκριμένες προσεγγίσεις είχαν σχεδιαστεί εξ ολοκλήρου για την αντιμετώπιση φωνολογικών αλλαγών και μετατοπίσεων, κυρίως στο ρηματικό παρελθοντικό σύστημα και ήταν αναποτελεσματικές στην πρόβλεψη αναλύσεων των ανωμάτων μορφών που δεν εντοπίστηκαν στα δεδομένα εκπαίδευσης.

Όπως επισημαίνει ο Wicentowski (2002) είναι αναπόσπαστο μέρος των τεχνικών η αξιολόγηση των αποτελεσμάτων, η μετρησιμότητα και η σύγκριση μεταξύ τους. Εδώ έγκειται και ένα από τα μεγάλα πλεονεκτήματα που παρέχουν οι τεχνικές της Εποπτευόμενης Μορφολογικής Μάθησης, στο γεγονός ότι τα σώματα κειμένων και τα δεδομένα εκπαίδευσης μπορούν να αποτελέσουν κοινό τόπο για όλες τις τεχνικές, αφού η βελτίωσή τους βασίζεται σε ανθρωπίνως (συνήθως) προσχολιασμένα παραδείγματα και δεδομένα.

#### 4.4.2.1. Μέγιστη Εντροπία

Η ιδέα της **Μέγιστης Εντροπίας** (Maximum Entropy) αρχικά ανιχνεύεται σε διάφορα στάδια της επιστήμης των μαθηματικών. Ωστόσο, μόλις τις τελευταίες δύο δεκαετίες τέθηκε το συγκεκριμένο μοντέλο στο πλαίσιο εφαρμογής, αφού πλέον το επέτρεπε και η υπολογιστική ισχύς, ώστε να αντιμετωπιστούν πραγματικά προβλήματα βασισμένα σε στατιστικές εκτιμήσεις και αναγνωρίσεις προτύπων. Εντούτοις από νωρίς ο Jaynes (1957a, 1957b, 1968, 1990) επεσήμανε τη δύναμη της Μέγιστης Εντροπίας στην επιστήμη της Πληροφορικής. Χρησιμοποιείται πλέον στην Επεξεργασία της Φυσικής Γλώσσας λόγω της δυναμικής στην ικανότητα των μοντέλων εντροπίας να συνδυάζουν ταυτόχρονα πολλές και διαφορετικές πληροφορίες. Σύμφωνα με τον Ratnaparkhi (1997) η θεωρία της Μέγιστης Εντροπίας είναι ένας ξεκάθαρος δρόμος για τους ερευνητές να συνδυάσουν στοιχεία/ ευρήματα για τη μοντελοποίηση



των δεδομένων· παράλληλα επισημαίνει ότι είναι ανεξάρτητη από την υπολογιστική συντακτική ανάλυση και μπορεί να εφαρμοστεί απρόσκοπτα και σε άλλα γλωσσικά/γλωσσολογικά ζητήματα.

Την περίοδο που σημειώθηκε αξιοσημείωτη πρόοδος στην κατεύθυνση της δυνατότητας πρόβλεψης των στατιστικών μοντέλων στην επεξεργασία γλώσσας, άρχισε να καθιερώνεται και η προσέγγιση της Μέγιστης Εντροπίας. Για παράδειγμα, τόσο στην μοντελοποίηση της γλώσσας, όσο και στην αναγνώριση ομιλίας, κρίνονται σημαντικές οι έρευνες των Bahl *et al.* (1989) και Della Pietra *et al.* (1994), και των Lucassen & Mercer (1984) με πρωτοπόρες τεχνικές η καθεμία από αυτές. Παράλληλα σημαντικές έρευνες πάνω στην υλοποίηση του μοντέλου στη μορφολογική επεξεργασία της γλώσσας είναι των Lau, Rosenfeld & Roukos (1993), Rosenfeld (1996) και Ratnaparkhi (1996, 1998, 1999), Koeling (2000) και Dell' Orletta *et al.* (2007). Στον τομέα του ενδιαφέροντός μας, της μορφολογικής ανάλυσης, οι Black *et al.* (1992) περιέγραψαν τον τρόπο εξαγωγής γραμματικών κανόνων από σχολιασμένο κείμενο αυτόματα και ενσωμάτωσης των κανόνων σε στατιστικά μοντέλα γραμματικής.

Όλα τα παραπάνω μοντέλα, αν και διαφέρουν στις βασικές λεπτομέρειες, έρχονται αντιμέτωπα με δύο σημαντικούς στόχους των στατιστικών μοντέλων. Για αυτούς τους στόχους οι Berger, Della Pietra & Della Pietra (1997: 40) τόνισαν ότι:

*«...ο πρώτος στόχος είναι να καθοριστεί το σύνολο των στατιστικών που θα συλλάβει τη συμπεριφορά μίας τυχαίας επεξεργασίας, δηλαδή η επιλογή χαρακτηριστικών. Και με δοσμένο το σύνολο των στατιστικών, ο δεύτερος στόχος είναι να συμπεριληφθούν τα γεγονότα αυτά σε ένα ακριβές μοντέλο επεξεργασίας – ένα μοντέλο που να μπορεί να προβλέψει τα μέλλοντα εξαγόμενα της επεξεργασίας –, δηλαδή η επιλογή του μοντέλου».*

Βασισμένοι στη Μέγιστη Εντροπία οι Berger, Della Pietra & Della Pietra (1997) χτίζουν ένα στοχαστικό μοντέλο που να αναπαριστά με ακρίβεια τη συμπεριφορά μίας τυχαίας επεξεργασίας, όπου ένα τέτοιο μοντέλο είναι η μέθοδος εκτίμησης της εξαρτώμενης πιθανότητας που με δεδομένα τα συμφραζόμενα  $x$  θα δώσει το εξαγόμενο  $y$ . Η βασική μονάδα, εντούτοις, είναι το χαρακτηριστικό (feature), μία συνάρτηση  $f$  ( $f: A \times B \rightarrow \{0,1\}$ ), όπου το  $A$  είναι το σύνολο των πιθανών δράσεων και το  $B$  το σύνολο των πιθανών συμφραζομένων. Γενικότερα, ορίζεται ένα σύνολο στοιχείων  $X$  που αφορά στο παρελθόν των γεγονότων (π.χ. λέξεις που προηγούνται, χαρακτηρισμούς λέξεων, μορφολογικά στοιχεία), ανάλογα με το στόχο της έρευνας και ένα σύνολο

στοιχείων  $Y$  που αφορά στο μέλλον των γεγονότων (π.χ. η υπό εξέταση λέξη, ο συνδυασμός χαρακτηριστικών, η σχέση μεταξύ μορφολογικών στοιχείων) και τέλος, ένας ενδεικτικός αριθμός χαρακτηριστικών που να περιγράφουν τη σχέση ανάμεσα στα στοιχεία  $X$  και  $Y$ . Έτσι η σχέση υποδηλώνεται με το  $p(x|y)$  που είναι η πιθανότητα που αντιστοιχεί το μοντέλο στα στοιχεία  $x$  και  $y$ . Αντίστοιχα το  $P$  είναι το σύνολο όλων των κατανομών των εξαρτώμενων πιθανοτήτων· επομένως, το  $p(x|y)$  είναι μέρος του συνόλου  $P$ . Η Κουτσομπόγερ (2004: 22) κατά την αξιολόγηση της συγκεκριμένης θεωρίας για το μορφολογικό χαρακτηρισμό λέξεων, επισημαίνει ορθά πως ένα τέτοιο μοντέλο δεν κάνει καμία υπόθεση για εμφανίσεις γεγονότων για τα οποία δεν έχει ενδείξεις από τα προκαθορισμένα χαρακτηριστικά, ενώ ταυτόχρονα ενσωματώνει σε ικανοποιητικό βαθμό τα εμπειρικά δεδομένα των προκαθορισμένων χαρακτηριστικών. Έτσι το μοντέλο βασίζεται πάνω στην κατανομή πιθανότητας που εμφανίζει τη μέγιστη εντροπία  $p^*$ :

$$(89.) \quad p^* = \arg \max_{p \in P} H(p),$$

όπου  $H(p)$  είναι η εντροπία της  $p$  (βλ. Berger, Della Pietra & Della Pietra (1996) για την Αρχή της Μέγιστης Εντροπίας: *μέγιστη εντροπία* είναι η πιθανολογική κατανομή της οποίας η εντροπία είναι τουλάχιστον όσο μεγάλη είναι των άλλων μελών μίας καθορισμένης ομάδας κατανομών).

Για να γίνει πιο κατανοητή η εκτίμηση των παραμέτρων του μοντέλου της Μέγιστης Εντροπίας παραθέτουμε τον αλγόριθμο IIS, που προτάθηκε από τους παραπάνω συγγραφείς και υλοποιήθηκε από τον Kazuhiro Yoshida με το AMIS<sup>16</sup>.

$$(90.) \quad p_\lambda(y|x) = \frac{1}{Z_\lambda(x)} \exp\left(\sum_{i=1} \lambda_i f_i(x,y)\right)$$

όπου  $f_i$  είναι η σχέση ισοδυναμίας ανάμεσα σε  $x$  και  $y$ , ενώ τα χαρακτηριστικά απεικονίζονται με το  $i$ , τα βάρη των αντίστοιχων χαρακτηριστικών με το  $\lambda_i$  και τέλος το  $Z(x)$  είναι μία σταθερά κανονικοποίησης. Ο συγκεκριμένος αλγόριθμος επαναλαμβάνει τις μετρήσεις του και λειτουργεί συγκριτικά μαζί τους, ώστε να αποδώσει τις τελικές ρυθμιστικές τιμές των βαρών των χαρακτηριστικών ( $\lambda_i$ ). Τελικά, το μοντέλο προκρίνει τις τιμές που έχουν κατανομή σε μέγιστη εντροπία.

<sup>16</sup> Περισσότερες πληροφορίες για το AMIS στην ιστοσελίδα <http://www-tsujii.is.s.u-tokyo.ac.jp/amis/>, αλλά και στο πείραμα του επόμενου κεφαλαίου.

Σημαντική ιδιότητα της θεωρίας αποτελεί η ικανότητα χειρισμού χαρακτηριστικών που παρουσιάζουν ένα βαθμό επικάλυψης ή συμπερίληψης ανάμεσά τους σχετικά με το υπό διερεύνηση πρόβλημα. Ένας υλοποιημένος parser ή tagger αυτής της θεωρίας μαθαίνει τις κατανομές πιθανότητας από ένα ανθρωπίνως σχολιασμένο σώμα κειμένων/ λέξεων (ειδικά για τη συντακτική ανάλυση η βάση του Tree-Bank αποτελεί το καθιερωμένο πεδίο δοκιμών). Είναι κατάλληλη τεχνική για μορφοσυντακτική ανάλυση, καθώς υπολογίζεται η κατανομή ενός εξεταζόμενου μορφολογικού χαρακτηριστικού σε ένα γλωσσικό περιεχόμενο ή σε ένα σύνολο μορφολογικών χαρακτηριστικών (βλ. πείραμα ενότητας 5.2) αξιοποιώντας και συνδυάζοντας διαφορετικά στοιχεία. Τα στοιχεία αυτά μπορεί να είναι ενδείξεις που συντελούν στον προσδιορισμό του σωστού μορφολογικού χαρακτηριστικού. Κάνοντας μία υπόθεση, ο εντοπισμός της αλλομορφικής τάξης εξαρτάται από την επιλογή των κατάλληλων χαρακτηριστικών, ώστε να αξιοποιηθούν από το σύστημα για την πρόβλεψη της αλλομορφίας για άλλες λέξεις. Όπως θα δούμε και στο πείραμα, τέτοια χαρακτηριστικά μπορεί να είναι η κλιτική τάξη, ο συλλαβισμός, ο τόνος, η σειρά των χαρακτήρων και το γένος μεταξύ άλλων.

Σύμφωνα με τους δημιουργούς του αλγόριθμου οι γνωστές και άγνωστες λέξεις αντιμετωπίζονται από κοινού με την ίδια στρατηγική, αφού αποτελούν ένα ακόμα χαρακτηριστικό στο γενικό μοντέλο των πιθανοτήτων. Έτσι κατά την αυτόματη σχολίαση η αναζήτηση έχει σκοπό να βρει την ακολουθία των χαρακτηριστικών με τη μεγαλύτερη πιθανότητα για τη λέξη, την πρόταση ή την μονάδα-στόχο· εξάλλου δεν γνωρίζουμε *a priori* ότι αυτά τα χαρακτηριστικά είναι όλα χρήσιμα ή απαραίτητα, για αυτό και πολλές φορές το μοντέλο ξεκινά με έναν μεγάλο αριθμό χαρακτηριστικών<sup>17</sup>. Εντούτοις με εμπειρικό τρόπο ορίζεται ένα όριο πρόκρισης χαρακτηριστικών, ώστε να αποκλείονται εκείνα που δεν καταφέρνουν να το ξεπεράσουν.

Ειδική αναφορά γίνεται για τα χαρακτηριστικά που έχουν χαμηλή συχνότητα εμφάνισης στο σχολιασμένο σώμα κειμένων και είναι στατιστικά δύσκολο να προβλεφθεί κάτι με βάση τις λιγότες στατιστικές τους. Όλες αυτές οι ευρετικές στρατηγικές χτίζουν την αξιοπιστία του κάθε χαρακτηριστικού στη συχνότητα εμφάνισης με αποτέλεσμα οι άγνωστες, σπάνιες και σποραδικές λέξεις των σωμάτων

---

<sup>17</sup> Σε πραγματικές εφαρμογές η επιλογή των τελικών χαρακτηριστικών είναι αρκετά πολύπλοκη και για αυτό λειτουργούν προσεγγιστικά με βάση τον όγκο δεδομένων εκπαίδευσης (αναλυτικότερα βλ. Berger, Della Pietra & Della Pietra 1997: 49-50).

εκπαίδευσης και δοκιμής να αντιμετωπίζονται παρομοίως, καθότι βασίζεται στην παραδοχή της όμοιας συμπεριφοράς. Σαφώς, όμως, αποτελούν ιδιαίτερες και προβληματικές περιπτώσεις που αντιμετωπίζονται σπασμωδικά και «τυχαία», αφού το μοντέλο κείται ευνοϊκά προς τις περιπτώσεις ομαλότητας, κανονικότητας και μη-χαρακτηρισμού (unmarkedness).

Στα Ελληνικά η συγκεκριμένη τεχνική χρησιμοποιήθηκε από την Κουτσομπόγερα (2004) για τη μορφολογική ανάλυση ιδιαζουσών περιπτώσεων, χρησιμοποιώντας την υλοποίηση *Open NLP Entropy Framework*<sup>18</sup> που βασίζεται στην προσέγγιση του Ratnaparkhi (1998). Χρησιμοποίησε περικείμενο  $\pm 1$  λέξης, αφού αυτό οδήγησε στην μέγιστη βελτίωση της ακρίβειας στην εύρεση της γραμματικής κατηγορίας των λέξεων. Σημειώνει πως σημαντικό πλεονέκτημα του μοντέλου αποτελεί η ακρίβειά του και η χρήση απλών γλωσσικών χαρακτηριστικών ακόμα σε συνθήκες επικάλυψης. Ταυτόχρονα εντυπωσιάζει η μεγάλη ικανότητα αναπαράστασης γνώσης με τη μορφή χαρακτηριστικών, ειδικά για πλούσιες μορφοσυντακτικά γλώσσες, όπως τα Ελληνικά. Στο πείραμά της ο tagger Μέγιστης Εντροπίας σε επίπεδο λέξης και πρότασης είχε ποσοστό επιτυχίας 98,154% και 79,703% αντίστοιχα. Συμφωνούμε με την τελική αξιολόγηση της Κουτσομπόγερα (2004: 25) που αναφέρει χαρακτηριστικά:

*«...Η μεθοδολογία της Μέγιστης Εντροπίας συνδυάζει τα πλεονεκτήματα όλων των τεχνικών Μηχανικής Μάθησης: παρέχει πλούσια αναπαράσταση χαρακτηριστικών, όπως οι βασισμένες σε κανόνες προσεγγίσεις και για κάθε απόφαση παράγει μία κατανομή πιθανότητας μορφολογικών χαρακτηριστικών, όπως οι τεχνικές των δέντρων Απόφασης και τα HMM. Επιπλέον, κάνει άμεση χρήση των λέξεων χωρίς να παρουσιάζεται το πρόβλημα της τμηματοποίησης δεδομένων. Τέλος, συνδυάζει με τον καλύτερο τρόπο – σε σχέση με οποιαδήποτε άλλη τεχνική – διαφορετικές και μη τοπικές πηγές πληροφορίας».*

#### 4.4.2.2. Μνημονική Μάθηση

Η **Μνημονική Μάθηση** (Memory-based Learning) είναι μία απλή σε λειτουργία προσέγγιση και μέθοδος, της οποίας οι ρίζες απαντώνται ήδη στη δεκαετία του 1910. Η θεωρία Μνημονικής Μάθησης υποστηρίζει ότι οι αποφάσεις σχετικά με τα νέα δεδομένα είναι βασισμένες σε επαναχρησιμοποίηση αποθηκευμένων παλαιότερων

<sup>18</sup> Περισσότερες πληροφορίες για το ελεύθερο λογισμικό στο <http://maxent.sourceforge.net>.

εμπειριών. Η παραγωγή προβλέψεων σχετικά με το εξαγόμενο που θα είναι αποτέλεσμα κάποιων χαρακτηριστικών των εισαγόμενων βασισμένα σε δεδομένα, γίνεται με τον εντοπισμό στοιχείων στη μνήμη, το ταίριασμα ενός μοντέλου σ' αυτά τα στοιχεία, ώστε να γίνει η πρόβλεψη με βάση το μοντέλο· ένα μοντέλο μνημονικής μάθησης αποτελείται από τέσσερα συστατικά: μία μετρική απόστασης (distance metric), τον αριθμό των πλησίων γειτόνων (near neighbors), μία συνάρτηση βάρους (weighting function) και ένα μοντέλο.

Με βάση τους Daelemans & van Bosch (2005) και Keuleers & Daelemans (2007) η συγκριμένη ιδέα έχει μακρά ιστορία στην επιστήμη της Σκέψης με τα πρώτα δειλά βήματα να γίνονται στην προ-Chomsky εποχή από τους de Saussure (1916) και Bloomfield (1933). Συναφείς έρευνες ως προς την ιδέα της Μνημονικής Μάθησης εντοπίζουμε στη Γνωστική Γλωσσολογία και Ψυχολογία, όπως αυτές των Skousen (2002), Croft & Cruse (2003) και Estes (1994), ενώ η πρώτη αλγοριθμική υλοποίηση ανήκει στους Fix & Hodges (1951) με τη γενίκευση του **κανόνα πλησιέστερου γείτονα** (nearest neighbor rule). Οι υπόλοιποι αλγόριθμοι που ακολούθησαν και εντάχθηκαν υπό την αιγίδα της Τεχνητής Νοημοσύνης, όσον αφορά στο θεωρητικό τους πλαίσιο δίνονται με διάφορες φράσεις καθορισμού, όπως 'case-based reasoning', 'memory-based reasoning', 'instance-based learning' κτλ. (αναλυτικότερα στους Daelemans & van Bosch 2005).

Τα μοντέλα Μνημονικής Μάθησης βασίζονται για την κλίση νέων μορφών/φορμών απευθείας στην αναλογία των αποθηκευμένων υποδειγμάτων, για αυτό και είναι γνωστά και ως *τεμπέλικα μοντέλα μάθησης* (lazy learning models). Συνήθως στα μοντέλα με σώματα κειμένων εκμάθησης παραλείπονται κάποια παραδείγματα που υπολείπονται του ορίου συχνότητας εμφάνισης. Αυτό κρίνεται ως αρνητικό σημείο των μοντέλων, γιατί μία από τις κεντρικές ιδέες των συγκεκριμένων μοντέλων είναι ότι κάθε υπόδειγμα είναι συναφές στη συμπεριφορά γενίκευσης, ενώ τονίζεται παράλληλα ότι τα υποδείγματα χαμηλής συχνότητας διαδραματίζουν σημαντικό ρόλο στη γενίκευση. Σύμφωνα με τον Baayen (2001) η παραγωγικότητα ενός κλιτικού σχήματος δόμησης αυξάνεται με τον αριθμό των **άπαξ λεγομένων**<sup>19</sup> (hapax legomena)

---

<sup>19</sup> Το *άπαξ λεγόμενο* είναι μία λέξη που εμφανίζεται μόνο μία φορά σε κάποια γραπτή αναφορά, σε κάποιο κείμενο, σε κάποιον συγγραφέα. Αναλογικά γίνεται και σπάνια χρήση των όρων *δισ λεγόμενον*, *τρεις λεγόμενον*, *τετράκις λεγόμενον* για τις αντίστοιχες αναφορές. Έχει απασχολήσει διάφορες έρευνες και επιστήμες αποτελώντας σημαντικό πρόβλημα ερμηνείας, αντιμετώπισης και καταχώρησης. Για τη γλωσσολογία βλ. Malmkjær (2002: 87).

φανερώνοντας το πραγματικό σχήμα δόμησης. Ο πλέον κατάλληλος τρόπος για να αντιμετωπίσουμε όλες αυτές τις περιπτώσεις είναι να συμπεριληφθούν όλα όσα είναι ήδη γνωστά κατά τη διαδικασία εκμάθησης που γίνεται στο μοντέλο (Keuleers & Daelemans 2007: 155). Εικάζεται ότι ακόμα και τα χαμηλής συχνότητας υποδείγματα/μορφές είναι καταχωρημένες στη βάση γνώσης του μοντέλου μάθησης ενός ενήλικα μαθητή.

Οι Keuleers & Daelmans (2007) στοχεύουν στο να μαντέψουν την κλιτική τάξη κάθε εισαγόμενου, καθώς και των τύπων πληθυντικού κατά προσέγγιση από τα υποδείγματα που είναι αποθηκευμένα στη μνήμη του μοντέλου. Αναμφίβολα αποτελεί πλεονέκτημα ότι ο αλγόριθμος εντοπισμού της κλίσης δεν γίνεται πιο πολύπλοκος, όταν προστίθενται όλο και περισσότερες κλιτικές τάξεις. Αν και βασίζουν τη μελέτη τους στα φωνολογικά χαρακτηριστικά για αυτή την πρόβλεψη, αναφέρουν όμως ότι βάσει άλλης έρευνας αναμένεται τα διάφορα σημασιολογικά χαρακτηριστικά να συνδράμουν στην προβλεψιμότητα κάποιων τύπων (Baayen *et al.* 2002, Baayen & Moscoso del Prado Martín 2005). Επιπροσθέτως, για χαρακτηριστικά που δεν έχουν πραγματική τιμή, το μοντέλο δίνει εικονική τιμή για να μην επηρεάσουν τη συνολική επίδοση. Φαίνεται ότι τα μοντέλα Μνημονικής Μάθησης είναι ικανά να ανταπεξέλθουν στις δυσκολίες πολύπλοκων ζητημάτων της μορφοφωнологίας. μία από τις πιο διαδεδομένες και πετυχημένες υλοποιήσεις είναι ο TiMBL (Tilburg Memory-Based Learner, πρβλ. για περισσότερα στους Daelemans *et al.* 2007), ο οποίος δοκιμάστηκε στα Ελληνικά από την Κουτσομπόγερα (2004) για τη βελτίωση της επισημείωσης (tagging).

Βασικά πλεονεκτήματα του μοντέλου Μνημονικής Μάθησης είναι η τεμπέλικη μάθηση, αφού δεν χρειάζεται η δημιουργία ενός καθολικού μοντέλου, καθότι πολλά απλά μοντέλα μπορούν να παρουσιάσουν ένα πιο πολύπλοκο καθολικό μοντέλο, με καλύτερα εστιασμένη μάθηση και αντιμετώπιση των κενών και απουσιών τιμών κάποιων χαρακτηριστικών. Ταυτόχρονα είναι πολύ αποτελεσματικό για συγκριτική εγκυρότητα, χαρακτηρίζεται από ευφυή μάθηση, ενώ με την εκμετάλλευση του αλγορίθμου *k*-nearest neighbors μπορεί να χρησιμοποιηθεί οποιαδήποτε μετρική απόστασης (distance metric strategy). Από την άλλη, μειονέκτημα αποτελεί το γεγονός ότι είναι χρονοβόρα και μνημοβόρα τακτική σε μεγάλους όγκους δεδομένων χωρίς να ξεπερνούν τα μοντέλα μνημονικής μάθησης τον μέσο όρο επίδοσης, ενώ μερικές φορές υπολείπονται των μοντέλων νευρωνικών δικτύων.

#### 4.4.2.3. Μάθηση Μετασχηματιστικού Τύπου

Ο Brill (1995) εισήγαγε τον φορμαλισμό της Μάθησης Μετασχηματιστικού Τύπου (Transformation-based Learning) για την επισημείωση των μερών του λόγου, του οποίου φορμαλισμού κεντρική ιδέα είναι να ξεκινήσει το μοντέλο με τις απλές λύσεις του εκάστοτε προβλήματος, να εφαρμόσει κάποιους μετασχηματισμούς συνεχώς, ώστε αυτοί να πληθαίνουν προς όφελος του συστήματος, να επιλέγονται και να εφαρμόζονται στο πρόβλημα. Προσθέτει σχολιασμούς/ ταξινομικά σχόλια σε κάθε εισαγόμενο. Ο αλγόριθμος σταματάει, όταν οι επιλεγμένοι μετασχηματισμοί δεν τροποποιούν περαιτέρω τα δεδομένα ή δεν υπάρχουν άλλοι μετασχηματισμοί για να επιλεγούν. Σε γενικές γραμμές στόχος του μοντέλου είναι η ταξινόμηση των συνόλων των υποδειγμάτων έχοντας τη δυνατότητα εφαρμογής στις περισσότερες διαστάσεις της γλωσσικής επεξεργασίας<sup>20</sup>, αλλά δυστυχώς σύμφωνα με τους Florian & Ngai (2001, 2007) δεν μπορεί να εφαρμοστεί με επιτυχία στη μορφολογική ανάλυση, αν και οι Brill (1996) και Siefkes (2003) διαφωνούν με αυτή την άποψη.

Αναλύοντας το συγκεκριμένο μοντέλο ο Siefkes (2003) σχολιάζει ως βασικά πλεονεκτήματα τη στρατηγική να μαθαίνει το μοντέλο από όλα τα δεδομένα και να επεξεργάζεται τα δεδομένα αλλάζοντας τη δομή τους με γρήγορη εφαρμογή των κανόνων. Ωστόσο είναι ιδιαίτερα απαιτητικό για την εκμάθηση των κατάλληλων κανόνων· δεν είναι πιθανοτικό (ή πιθανοθεωρητικό) μοντέλο και έτσι αδυνατεί να επιστρέψει πάνω από ένα αποτέλεσμα και δεν μπορεί να βασιστεί σε αξιόπιστη μέτρηση των αποτελεσμάτων. Στα Ελληνικά η Κουτσομπόγερα (2004: 16-21) βελτίωσε την απόδοση ενός επισημειωτή για την αναγνώριση των προβληματικών περιπτώσεων με ελάχιστη διαφορά από τα υπόλοιπα μοντέλα.

---

<sup>20</sup> Σε διάφορους τομείς της ΕΦΓ, υπάρχουν έρευνες, όπως των Ramshaw & Marcus (1995) και Brill *et al.* (2000) στο διαχωρισμό κειμένων και των Brill & Resnik (1994), Brill (1996), Day *et al.* (1997) και Samuel (1998) στο μορφοσυντακτικό πεδίο.

*“A computer lets you make more mistakes faster than any invention in human history – with possible exceptions of handguns and tequila” Mitch Ratcliffe*

## 5. Τα πειράματα με τους μορφολογικούς αναλυτές

Στο προηγούμενο κεφάλαιο αναφερθήκαμε σε διάφορες προσεγγίσεις της αυτόματης ή μη μορφολογικής μάθησης από κάποιο υπολογιστικό μοντέλο. Στο παρόν κεφάλαιο θα δοκιμάσουμε διάφορους μορφολογικούς αναλυτές των οποίων οι αρχές δόμηση διέπονται από διαφορετικό θεωρητικό, για να εντοπίσουμε τις επιδόσεις τους στην αντιμετώπιση του εκτενούς φαινομένου της αλλομορφίας στην Ελληνική Γλώσσα. Είναι προφανές ότι λόγω της διαφορετικής δόμησης και στρατηγικής δεν είναι δυνατόν να δοκιμάσουμε τα ίδια δεδομένα σε όλα τα πειράματα. Για αυτό το λόγο και χρησιμοποιήθηκαν διαφορετικά δεδομένα για κάθε πείραμα· εντούτοις, η φύση των δεδομένων που χρησιμοποιήθηκαν είναι επαρκής και ικανοποιητική για την ασφαλή εξαγωγή συμπερασμάτων και προτάσεων.

### 5.1. Το πείραμα του LINGUISTICA στα Ελληνικά

Στην πιλοτική θεωρητική προσέγγιση του ζητήματος της Μη-Εποπτευόμενης Μορφολογικής Μάθησης (Καρασίμος 2009) είχαμε υποστηρίξει ότι το συγκεκριμένο μοντέλο είναι καταδικασμένο να αποτύχει στον εντοπισμό της αλλομορφίας στο ονοματικό και ρηματικό κλιτικό σύστημα της Ελληνικής όχι μόνο λόγω της πολυπλοκότητας και της συχνότητας του φαινομένου, αλλά και λόγω της παντελούς απουσίας κάποιων μορφοφονολογικών κανόνων από το σύστημα. Το πείραμα με το



Linguistica διεξήχθη με διαφορετικά σώματα κειμένων (Karasimos 2009, Karasimos & Petropoulou 2009, 2010 και Πετροπούλου & Καρασίμος 2010) για τη διασφάλιση της ακεραιότητας των αποτελεσμάτων και συμπερασμάτων για το συγκεκριμένο πρόγραμμα και θεωρητικό πλαίσιο. Να σημειωθεί ότι σύμφωνα με τους δημιουργούς του, το συγκεκριμένο πρόγραμμα μπορεί να τρέξει σχεδόν για οποιαδήποτε γλώσσα, αρκεί να δοθεί ένα ικανοποιητικό σώμα κειμένων για τη μορφολογική ανάλυση και πρόταση του συστήματος για την υπό δοκιμή γλώσσα. Εντούτοις, παρόλο που ορθώς ο Wicentowski (2002: 27) σημειώνει ότι εντοπίζεται ένα αξιοσημείωτο κενό στην έρευνα πάνω στους ανώμαλους τύπους λέξεων, ο Goldsmith (2001) προσπαθεί με ιδιαίτερη απλότητα να εντοπίσει την αλλομορφία σημειώνοντας ότι το εν λόγω σύστημα *δεν είναι* κατασκευασμένο για να επεξεργάζεται αλλαγές τύπου ‘feel/ felt’.

### 5.1.1. Σχετικά με το LINGUISTICA

Το LINGUISTICA<sup>1</sup> είναι ένα πρόγραμμα για την εξερεύνηση της μη-εποπτευόμενης μάθησης μίας φυσικής γλώσσας, με πρωταρχικό στόχο τη μορφολογία και ειδικά τη δομή των λέξεων. Εκτελείται κάτω από διάφορα λειτουργικά συστήματα και είναι γραμμένο στη γλώσσα C++ βασισμένο στο QT development framework, ενώ οι απαιτήσεις μνήμης εξαρτώνται από τον όγκο δεδομένων που διαχειρίζεται. Σε γενικές γραμμές απαιτεί λίγη μνήμη και είναι ταχύτατο για την ανάλυση που απαιτείται.

Signatures	Exemplar	Descr. Length (model)	Corpus Count	Stem Count	Source	Robustness
αόλιος-αοριόλιος	εντερ	59,2227	8	4	Check sign	53
αοριόλιος-αοριόλιος	επισπε	194,8956	31	15	Check sign	289
γενής-γενής	επερο	95,246	6	3	Check sign	36
οποια-οποιας	οποιας	54,6068	9	2	Check sign	45
αοριόλιος-αοριόλιος	ηβ	55,9844	3	3	Known stems to suffixes	92
αοριόλιος-αοριόλιος	βουλκο	71,626	10	5	Check sign	88
αοριόλιος-αοριόλιος	θραυτ	39,9864	4	2	Check sign	27
ροσοκοπία-ροσοκοπία	ιε	86,3202	8	4	Check sign	62
αοριόλιος-αοριόλιος	ιραο	52,3214	6	3	Check sign	46
αοριόλιος-αοριόλιος	κατοκογ	76,6321	15	3	SF1	128
αοριόλιος-αοριόλιος	καθουγ	67,4195	12	3	SF1	113
αοριόλιος-αοριόλιος	αοριόλιος	132,331	20	10	Check sign	148
οποια-οποιας	οποιας	57,7854	10	2	Check sign	74
οποια-οποιας	οποιας	40,6623	4	2	Check sign	20
αοριόλιος-αοριόλιος	κορ	49,4135	6	3	Check sign	39
αοριόλιος-αοριόλιος	κορ	36,8937	4	2	Check sign	22
αοριόλιος-αοριόλιος	αβ	52,7219	6	3	Check sign	42
αοριόλιος-αοριόλιος	αοριόλιος	50,1208	6	2	Check sign	44
αοριόλιος-αοριόλιος	αοριόλιος	41,6952	4	2	SF1	25
αοριόλιος-αοριόλιος	αοριόλιος	76,1617	10	5	SF1	68
αοριόλιος-αοριόλιος	αοριόλιος	96,549	12	6	Check sign	101
αοριόλιος-αοριόλιος	αοριόλιος	43,7572	4	2	Check sign	17
αοριόλιος-αοριόλιος	αοριόλιος	54,1396	8	2	From known stem and suffix	59
αοριόλιος-αοριόλιος	αοριόλιος	47,2446	6	2	Known stems to suffixes	56
αοριόλιος-αοριόλιος	αοριόλιος	45,3425	4	2	SF1	22
αοριόλιος-αοριόλιος	αοριόλιος	42,1713	4	2	Check sign	21

Εικόνα 1: Η διεπαφή του προγράμματος LINGUISTICA.

Ο υπολογιστικός στόχος του συστήματος είναι να εξάγει συμπεράσματα (και επομένως να κατακτήσει τη γνώση) σχετικά με τη δομή που κρύβεται πίσω από τα

<sup>1</sup> Διατίθεται δωρεάν σε διάφορες εκδόσεις και διαφορετικές πλατφόρμες και λειτουργικά συστήματα στο <http://linguistica.uchicago.edu/>.

δεδομένα και ανάμεσα στις λέξεις χωρίς να έχει κάποια εξωγενή άμεση βοήθεια για τη λεξική δομή. Ακολουθώντας τις επιταγές του θεωρητικού πλαισίου, το LINGUISTICA εξερευνά τις πιθανότητες των συνδυασμών των μορφημάτων για το σύνολο των λέξεων χωρίς καμία εσωτερική γνώση της γλώσσας.

Ο διαχωρισμός των μορφημάτων είναι ο πρώτος στόχος επεξεργασίας: το πρόγραμμα αρχικά καθορίζει τα όρια των μορφημάτων μέσα στις λέξεις και ακολούθως αποφασίζει ποια είναι τα θέματα, ποια είναι τα επιθήματα και ούτω καθ' εξής. Η στρατηγική της λειτουργικότητας του LINGUISTICA εντοπίζεται στη λήψη των σωστών αποφάσεων για τον κατάλληλο διαχωρισμό. Για το συγκεκριμένο πείραμα, χρησιμοποιήσαμε την έκδοση 3.2.6 (Μάρτιος 2009) σε περιβάλλον Windows XP.

### 5.1.2. Εντοπίζοντας την Αλλομορφία με το LINGUISTICA

Αναφέρεται ότι το LINGUISTICA έχει τη δυνατότητα να καθορίσει έναν περιορισμένο αριθμό αλλομόρφων στα θέματα/ ρίζες. Σε αρκετές γλώσσες (συμπεριλαμβανομένων των Αγγλικών), συχνά ο τελευταίος χαρακτήρας του θέματος πριν από επίθημα διαγράφεται. Για παράδειγμα, το τελικό φωνήεν του θέματος *-e* διαγράφεται, όταν έπεται μία ομάδα επιθημάτων: *love*, αλλά *lov-ing* και όχι *love-in-sane* και *sanity*, αλλά όχι *sane-ity*. Ο Goldsmith τα αντιμετωπίζει ως αλλόμορφα – γεγονός που δεν ισχύει –, όντας αποτελέσματα κανόνων της μορφοφωнологίας.

Η τακτική του LINGUISTICA είναι να επαναναλυθούν στοιχεία που προηγουμένως έχουν ενσωματωθεί σε κάποιο επίθημα αν και είναι μέρος θέματος και παρέχει την πληροφορία ώστε και τα υπόλοιπα επιθήματα να διαγράψουν αυτό το υλικό, όταν εμφανίζεται πριν από αυτά. Ο Goldsmith (2001) παρουσιάζει το ακόλουθο παράδειγμα: οι λέξεις *love*, *loves*, *loved* and *loving*, οι οποίες αναλύθηκαν ως *lov* + με φύλλο σώματος τα *e.ed.es.ing*, θα αναλυθούν σε *love* και επιθήματα *NULL*, *ed*, *s*, και *ing*. Τα επιθήματα *-ed* και *-ing* θα ενημερωθούν ότι έχουν τη δυνατότητα να διαγράψουν το *e* που προηγείται κατά τη διαδικασία σχηματισμού και αυτό σημειώνεται με το *e* σε άγκιστρα πριν το επίθημα, για αυτό έχουμε: *<e>ing* και *<e>ed*. Επομένως, το νέο φύλλο σώματος για το *love* είναι *NULL.<e>ed.<e>ing.s*, και αυτό το φύλλο σώματος αντιμετωπίζει επιτυχώς τα θέματα λήγοντα σε *-e* και τα αντίστοιχα χωρίς το ληκτικό φωνήεν *-e*.

Επιπροσθέτως, δηλώνεται ότι το *Linguistica* αντιμετωπίζει ουσιαστικά και ρήματα σε τελικό-*y* με τον ίδιο τρόπο: *academy/academies* και *try/tries* αναλύονται έχοντας ως θέμα το *academy* και *try* και κλιτικά επιθήματα τα NULL και *<y>ies*.

### 5.1.3. Τα σώματα κειμένου δοκιμής

Όπως έχουμε ήδη αναφέρει στην κριτική του θεωρητικού πλαισίου, η υπόθεσή μας είναι ότι το LINGUISTICA πρόκειται να έχει σοβαρά προβλήματα στην ανάλυση ενός σώματος ελληνικών κειμένων. Για να ελέγξουμε την υπόθεσή μας, δημιουργήθηκαν τρία σώματα κειμένων *ad hoc*: το πρώτο σώμα (ΣΚ1) έχει 55.897 δείγματα λέξεων (48.000 λέξεις) από την εφημερίδα *Μακεδονία*<sup>2</sup>, το δεύτερο σώμα (ΣΚ2) έχει 30.907 προσεκτικά επιλεγμένες λέξεις<sup>3</sup> και λήμματα (λέξεις με κοινά κλιτικά και παραγωγικά επιθήματα, ομάδες με προθηματοποιημένες λέξεις, κτλ.) και το τρίτο σώμα (ΣΚ3) έχει 281.821 λέξεις από δύο μυθιστορήματα του φανταστικού (Καρασίμος 2002, υπό επεξεργασία). Τα αποτελέσματα από το πρώτο σώμα κειμένων ήταν τελείως απογοητευτικά (ακρίβεια 7,31%): τα αποτελέσματα του τρίτου σώματος κειμένου ήταν ελαφρώς καλύτερα, αλλά η ακρίβειά του εξακολουθούσε να είναι υπερβολικά χαμηλή (ακρίβεια 11,49%). Απεναντίας, τα αποτελέσματα του δεύτερου σώματος είναι πιο ξεκάθαρα, συγκεκριμένα και σαφή, αν και η ακρίβεια παρέμεινε σε χαμηλά επίπεδα (ακρίβεια 23,41%).

Το σύστημα κατάφερε να εντοπίσει αρκετά κλιτικά παραδείγματα, λίγα παραγωγικά επιθήματα – καθώς και κάποια δεσμευμένα θέματα. Επιπροσθέτως, ο πρωτεύων στόχος μας ήταν ο εντοπισμός της αλλομορφίας: το LINGUISTICA εντόπισε δύο ειδών αλλομορφίας, εκ των οποίων μόνο μία ήταν σωστή, τα αλλόμορφα τύπου *παιδι~παιδ!* Να σημειωθεί πως δοκιμαστήκαν τα σώματα κειμένων σε άτονες εκδόσεις με μηδαμινή βελτίωση: τα φύλλα σώματος που είχαν πραγματικά προσφύματα αλλά συνδυάζονταν με σύνολο λέξεων που περιείχαν ψεύτικα θέματα συγκαταλέγονταν στις σωστές αναλύσεις (ανθ.υγιεινός, ανθ.ρωπος, ανθ.ρακας), διότι η ανάλυση βασίστηκε

<sup>2</sup> <http://www.komvos.edu.gr/dictionaries/corpora/Corpora.htm>. Για την ομοιομορφία του σώματος επιλέχθηκαν αποσπάσματα μόνο από τους τομείς *πολιτική* και *οικονομικά*.

<sup>3</sup> Δοκιμάστηκε παράλληλα και μία επεκταμένη έκδοση του ΣΚ2 με επιπλέον 4.100 λέξεις από τα δεδομένα της Πετροπούλου (in progress) για την εξαγωγή κάποιων συμπερασμάτων για τα δεσμευμένα θέματα. Τα αποτελέσματα από το ΣΚ2+ αναφέρονται στον πίνακα με αστερίσκο (\*) και περιλαμβάνουν όλες τα λεξήματα με δεσμευμένα θέματα της Ελληνικής. Επίσης, στο ΣΚ3 αφαιρέθηκαν όλα τα ονόματα και τοπωνύμια, καθότι ήταν κατασκευασμένες λέξεις εκτός διαδικασιών της γλώσσας.

στα προσφύματα. Σε αντίθετη περίπτωση θα σημειώνονταν χαμηλότερα ποσοστά σωστής ανάλυσης.

Αναλύσεις	Corpus 1 «Μακεδονία»	Corpus 2 «Στοχευμένο»	Corpus 3 «Σκήπτρο x2»
Σωστή Ανάλυση	07,31%	23,41% (*27,72%)	11,49%
Λανθασμένη Ανάλυση	05,22%	42,30% (*40,01%)	49,62%
Μη-Ανάλυση <sup>4</sup>	86,38%	29,76% (*26,21%)	31,44%
Ψεύτικη Ανάλυση	01,09%	04,53% (*06,06%)	07,45%

Πίνακας 14: Τα συγκεντρωτικά αποτελέσματα των τριών σωμάτων κειμένων από το πείραμα του LINGUISTICA.

#### 5.1.4. Αποτελέσματα

Παρατηρώντας τα αποτελέσματα<sup>5</sup> του LINGUISTICA, είχαμε την ακόλουθη top δεκάδα (91.) φύλλων σώματος για τα επιθήματα:

- (91.) (i.) NULL.δες.δων,  
 (ii.) άρεις.άρετε.άρισα.άρουμε.άρω,  
 (iii.) άρα.ες.ης.ικός.ων,  
 (iv.) άτων.ατάκι.ατάρα,  
 (v.) ά.άκι.ου.ο.ων,  
 (vi.) ά.άδες.άς,  
 (vii.) ές.εδάκι,  
 (viii.) ά.άς.ατζή,  
 (ix.) νες.νων και  
 (x.) NULL.είς.

Το πρώτο φύλλο σώματος (91.i) συνδυάζεται με θέματα που έχουν το δ-αλλόμορφο (μεζε-, κουβα-, μαμα-, παπα-), το δεύτερο συμπεριλαμβάνει θέματα ξενικής προέλευσης, που σχηματίζουν ρήματα σε -άρ(ω) (σκορ-, σοκ-, σκαν-, σερφ-, παρκ-) και το πέμπτο συνδυάζεται μόνο με ουσιαστικά της KT5 (βουν-ό, μωρ-ό, νερ-ό).

Τα αποτελέσματα προήλθαν από την εφαρμογή ενός προηγμένου συστήματος με προχωρημένες ευρετικές (πρβλ. Goldsmith 2001). Ο ίδιος σημειώνει ότι το γενικό

<sup>4</sup> Αναφερόμαστε στις λέξεις που αναλύονται από το σύστημα ως μονομορφηματικές.

<sup>5</sup> Η ανάλυση των αποτελεσμάτων έγινε πάνω στα δεδομένα που προέκυψαν από το δεύτερο σώμα κειμένων, μίας και ήταν αυτό με την υψηλότερη ακρίβεια.

σχέδιο της μορφολογίας των Αγγλικών και άλλων Ευρωπαϊκών γλωσσών παρουσιάζει σε γενικές γραμμές αρκετή κανονικότητα στην περιγραφή τους. Εντούτοις, τα αποτελέσματα των Αγγλικών, εφόσον εξεταστούν σχολαστικά, φανερώνουν κάποια ουσιώδη λάθη στην ανάλυση λέξεων. Ο συγγραφέας του LINGUISTICA προσπάθησε με αρκετή επιτυχία να διορθώσει τα περισσότερα των λαθών με επιπρόσθετες ευρετικές και να τις αξιολογήσει με τη μετρική MDL. Τα αποτελέσματα των Ελληνικών δεν χρειάζονται προσεχτικότερη ματιά, αφού τα σφάλματα αποτελούν τον κανόνα και όχι την εξαίρεση. Τα σφάλματα μπορούν να οργανωθούν με τον ακόλουθο τρόπο, ώστε να είναι κατανοητά σε ομάδες:

(α.) Η ένωση δύο επιθημάτων σε ένα: για παράδειγμα, ιδιαίτερα συχνά βρήκαμε το *-ικός* και *αντισυν-* παράλληλα εντοπίστηκαν και μεγαλύτερες ενώσεις επιθημάτων, όπως *-ευτικός* και *-αρισμένος*.

(β.) Η συστηματική συμπερίληψη του τελευταίου χαρακτήρα θέματος σε ένα σύνολο από λανθασμένα επιθήματα. Στα Ελληνικά, παραδείγματος χάριν, το ιδιαίτερα υψηλής συχνότητας εμφάνισης θεματικό τελικό *-τ-* (*κύματ-α*, *σώματ-α*, *βήματ-α*) μπορεί να οδηγήσει το σύστημα σε μία λανθασμένη οριοθέτηση ενός ολόκληρου συνόλου επιθημάτων, όπως το φύλλο *τος.τα.των οτ τακι.ταρα*. Ίδια περίπτωση είναι τα αλλόμορφα τύπου-*δ* (*παπαδ-*, *καναπεδ-*) με την «επανανάλυση» του συμφώνου *δ* στο επίθημα (*παπά-δες*, *καναπέ-δες*), όπως παρουσιάζεται από τις παραδοσιακές γραμματικές.

(γ.) Η εμφάνιση ψεύτικων φύλλων σώματος, που κατά κόρον παράγονται από μικρά θέματα ή επιθήματα και το πρόβλημα της επαναδρομής και εφαρμογής κάποιων φύλλων σώματος πάνω σε πραγματικά επιθήματα. Για παράδειγμα, το κλιτικό επίθημα *-ς* (ON.EN.AΡΣ/ΓΕΝ.EN.ΘΗΛ.) είναι πραγματικό επίθημα, αλλά κάθε λεξική κατάληξη σε *-ους* δεν πρέπει να αναλύεται σαν να συμπεριλαμβάνει αυτό το επίθημα. Αναμενόμενο λάθος από το σύστημα είναι να ταυτίζει το *-ας* από τύπους, όπως *αγαπ-ας*, *μιλ-ας* με αντίστοιχο «ψεύτικο» κλιτικό επίθημα των θηλυκών, όπως *θάλασσ-ας*, *γυάλ-ας*.<sup>6</sup>

(δ.) Απουσία συστηματικότητας στον τρόπο ανάλυσης των κλιτικών επιθημάτων όψης, μίας και σε κάποια ρήματα το *-σ-* αποτελούσε μέρος του θέματος (*αγαπησ-*

<sup>6</sup> Σύμφωνα με την ανάλυση της Ράλλη (2000, 2005) το *-α-* είναι μέρος του θέματος και μόνο το *-ς* αποτελεί κλιτικό επίθημα γενικής ενικού.

α.εξ.ε), ενώ σε άλλες περιπτώσεις μέρος του επιθήματος (*διαιρού-σα.σεξ.σε*). Αντιθέτως ποτέ φώνημα παραγωγικού επιθήματος δεν προσκολλήθηκε σε θέμα.

(ε.) Η αποτυχία χωρισμού των λέξεων που έχουν κοινό θέμα/ επίθημα με τον ίδιο τρόπο ή η συσχέτιση μεταξύ τους: για παράδειγμα, το θέμα *χορ-* με το φύλλο σώματος *ος.οι.ους* δεν σχετίζεται με το *χορ-* με τα επιθήματα *ευω.ευεις.ευει*. κτλ., ενώ το *-τη(ς)* αναλύεται με ποικίλους τρόπους όπως *μαχ.ητης, ασκη.της, μετρητ.ης*. Αυτό είναι ιδιαίτερα σύνηθες στα δεσμευμένα θέματα, όπου στις περισσότερες περιπτώσεις έχουμε τις μορφές με το συνδετικό φωνήεν και εκείνες χωρίς (*ολογία.ολόγος vs. λογία.λόγος, όμορφος.ομορφία vs. μορφος,μορφία*).

(στ.) Στην Ελληνική γλώσσα, θέματα που δεν είναι πανομοιότυπα μπορούν να συσχετίζονται μεταξύ τους, όπως τα αλλόμορφα. Το θέμα *αιμ-* εντοπίζεται με το φύλλο σώματος *α.ατα.ατων*, καθώς και το θέμα *αι-* μπορεί να βρεθεί με τα *ματακι.ματαρα*, αλλά αυτά τα θέματα είναι ίδια στη μορφολογία.

(ζ.) Το σύστημα ποτέ δεν κατάφερε να ξεχωρίσει το συνδετικό φωνήεν *-ο-* των δεσμευμένων θεμάτων και (ελαχίστων) συνθέτων ως ξεχωριστό στοιχείο. Στατιστικά τις περισσότερες φορές (68,11%) το συνδετικό φωνήεν αποτελούσε μέρος του θέματος ή πρώτου συνθετικού (*γλωσσο-*) και λιγότερες φορές μέρος του δεσμευμένου θέματος ή δεύτερου συστατικού (*-ολόγος*) χωρίς να υφίσταται κάποια συστηματικότητα ή προβλεψιμότητα πίσω από αυτή τη στρατηγική.

(η.) Το LINGUISTICA απέτυχε παταγωδώς να αντιμετωπίσει την αλλομορφία. Σε πολύ μεγάλο ποσοστό οριοθετούσε λανθασμένα λόγω της αλλομορφίας, ενώ στις λίγες περιπτώσεις που εντόπιζε και τα δύο αλλόμορφα του μορφήματος (*ανθ~ αντι, υφ~ υπο*) δεν τα συσχέτισε ποτέ. Το ίδιο ισχύει και για τις περιπτώσεις των φωνομόρφων (*συν- συμ- συρ, γραφ- γραψ*).

#### 1. α.ε.εξ.ουμε.ουν.ω

γλυκαιν πεταξ κοιταξ τραβηξ βουτηξ  
ρουφηξ τσεκαρ στοκαρ παρκαρ  
λασκαρ σκαναρ φρεναρ σταμπαρ  
κεντραρ πλασαρ μανταρ πιλοταρ  
σερφαρ ...

(σύνολο 512 θεμάτων)

#### 2. ομαι.ω

συνθλιβ κοβ σκοβ συνεισαγ  
αντιπροσαγ αποκαιγ διδ μεταδιδ  
συνεκδιδ αποδιαβαζ αποβιβαζ βγαζ  
ζυγιαζ ανθιζ απαλειφ ...

#### 6. NULL.δες.δων

αγά βεζιρη γιαγια ζαφιρη ζελε ζουρνατζη  
καναπε καβγατζη καφε κεσε λουλε μαμα  
μινारे μπαχτσε παπα χαλιφη ...

(σύνολο 188 θεμάτων)

#### 7. ζα.ζαμε.ζαν.ζατε.ζε.ζει.ζεις.ζεξ.ζετε.

ζουμε.ζουν.ζω.σαμε.σαν.σατε.σει.σεις  
επηρεα αγια θυσια θαυμα αγορα οπλι  
εντοπι μαυρι αλατι ...

(σύνολο 162 θεμάτων)

#### 8. ατων.ατακι.αταρα

<p>(σύνολο 486 θεμάτων)</p> <p><b>3. α.ε.εξ.μενος.ουμε.ουν.ω</b> επιηρεασ αγιασ θυσιασ εστιασ χαλασ γελασ κρεμασ θαυμασ περασ αγορασ ξεχασ οπλισ εντοπισ θερισ ορισ ... (σύνολο 264 θεμάτων)</p> <p><b>4. α.ος</b> απεικαστικ ορθολογιστικ ναζιστικ αντιεθνικιστικ ρεαλιστικ σοσιαλιστικ καπιταλιστικ κοιμιστικ απεικονιστικ συντονιστικ κομμουνιστικ μαρξιστικ αντιεπενδυτικ αποθηκευτικ συνταραχτικ (σύνολο 232 θεμάτων)</p> <p><b>5. ικα.ικος.ικως</b> απαντηη αντιπολιτ ανεκτ συνδιαλλακτ απαλλακτ συνταρακτ αντιρομαντ δημοτ αποθαρρυντ αντιαγροτ δυναστ συναρπαστ απεργοσπαστ ... (σύνολο 201 θεμάτων)</p>	<p>αιμ αλμ αρμ βημ βιασμ βιδωμ βλεμμ γεμυ δεμ διλημμ ερμ ζητημ θαυμ θρεμμ κραμ κρασμ κρατημ κυμ λαλημ λημμ μειγμ νημ ξερασμσωμ τραυμ χτυπημ ... (σύνολο 137 θεμάτων)</p> <p><b>9. ευα.ευαμε.ευαν.ευατε.ευε.ευει.ευεις.</b> <b>ευες.ευετε.ευουν.ευω.εψαμε.εψαν.εψατε.</b> <b>εψει.εψεις.εψετε</b> αναδ μπερδ λογικ παιν κουρ κιοτ λατρ παντρ πιστ φυτ χορ χων ψαρ φον χηρ χαζ αποθηκ γιατρ γητ δασκαλ δασ λογικ ... (σύνολο 136 θεμάτων)</p> <p><b>10. ά.άκι.ού.ό.ών</b> αβγ βουν γονατ δεντρ ερμαρι ζω ηνι θρανι ιματι κινητ λουστρ μνημει νερ ξυλ οστρακ ξεφωτ παραθυρ στεν τουβλ υπερω φρουρι χειροποιητ ψητο ... (σύνολο 89 θεμάτων)</p>
--	---

Πίνακας 15: Η δεκάδα υψηλότερης εμφάνισης επιθημάτων στο ΣΚ2+.

#### 5.1.4.1. Προθηματοποίηση

Η ανάλυση των προθημάτων της Ελληνικής δεν πρέπει να αποτελέσει δύσκολο αντίπαλο για το LINGUISTICA, αφού είναι μικρό σύνολο με περιορισμένη αλλομορφία. Τα αποτελέσματα έδειξαν ότι στην προθηματοποίηση το ποσοστό επιτυχίας άγγιξε το 40% (38,41%), ποσοστό υψηλό αναλογιζόμενοι τους χαμηλούς μέσους όρους. Κατόρθωσε να δημιουργήσει φύλλα σώματος όπως *συν.αντι* {εργατικός, ένζυμο, εισφορά}, *αντι.κατα* {βάλλω, θέτω, βαινω}, *συν* {έρχομαι, τρέχω, άγω, δράμω}, που περιέχουν αληθινά προθήματα. Εντούτοις, όπως σημειώθηκε προηγουμένως, φύλλα σώματος με δύο προθήματα δημιουργήθηκαν επίσης, όπως *συν.συνεπι* {τηρω, τηρητής, τηρούμαι}, *συν.συνυπο* {δηλώνω, γράφω} και *αντι.συνυπο* {γράφομαι, γεγραμμένος}. Επιπροσθέτως, το σύστημα απέτυχε να ομαδοποιήσει προθήματα με κοινούς χαρακτήρες *α-* και *αν-*, *κατα-* και *κατ-* ή το πιο «φωνομορφικό» πρόθημα *συν-* {*συμ-*, *συγ-*, *συλ-*, *συρ-*, *συσ-*}, αφού το σύστημα δεν ενσωματώνει καμία εφαρμογή φωνολογικών κανόνων, όπως αυτής της διαγραφής και αφομοίωσης. Επιπλέον, είναι πολύ κοινό λάθος στα ψεύτικα προθήματα να συμπεριλαμβάνονται κάποιοι από τους αρχικούς χαρακτήρες του ακόλουθου θέματος ή προθήματος (π.χ. *συνδ-*, *συναρ-*,

συνθηκ-, συναρμ-) ή να χαρακτηρίσει ως προθήματα ανύπαρκτα επίθηματα (γλυ-, λευ- {κός.κό.καίνω.κά.κάθηκα}). Τέλος, το LINGUISTICA δεν μπόρεσε να εντοπίσει καμία αλλομορφική συμπεριφορά των επιθημάτων και φυσικά απέτυχε να συσχετίσει πραγματικές μορφές του ίδιου μορφήματος όπως *κατα-* και *καθ-*, *υπο-* και *υφ-*.

Signatures	Exemplar	Descr. Length (model)	Corpus Count	Stem Count	Source	Robustness
απο-παρα	λειτουργώ	1081.65	272	121	Singleton	1702
NULL-αντι	αγιακά	684.761	126	63	PF1	948
απο-κατα	λουζώ	502.362	92	46	Singleton	789
αντι-απο	λαμπή	371.726	84	38	Singleton	567
αντι-παρα	οικονομικός	296.99	70	31	Singleton	590
NULL-παρ	απομένω	351.043	66	32	PF1	351
κατ-παρ	ακουράζω	332.671	60	30	PF1	417
NULL-συν	αντιπάλητος	308.352	56	28	PF1	331
NULL-απο	αποικίηση	264.387	48	24	PF1	318
απ-συν	αστρία	243.144	43	21	PF1	271
ανθ-απ	όσκεπος	235.335	42	21	PF1	242
απο-συν	διαπέγομαι	167.549	30	15	Singleton	199
αντι-απο-παρα	λαβή	59.8342	26	7	Singleton	146
αντι.παρα.απο	λαγία	41.2301	12	3	Singleton	60
NULL-παρα	αντιπαθώ	135.828	24	12	PF1	146
απο-παρα-συν	τρομάζω	80.5405	21	7	PF1	148
παρα.συν	διασκεδάζω	92.8124	16	8	PF1	115
παρ-συν	ακλουθώ	115.267	20	10	PF1	135
NULL-απ	ανθράκωμα	104.334	18	9	PF1	98
απ-απαθ	όγτος	107.689	18	9	PF1	107
α-αντι	πίεση	95.9887	16	8	PF1	105
απο-κατα-συν	διώκομαι	61.5896	15	5	PF1	100
κατα.συν	θλιβώ	51.1627	8	4	PF1	55
απ-κατ-παρ	αγγέλλομαι	64.6358	15	5	PF1	114
αντι-συν	ένζυμο	81.9933	14	7	PF1	101
αντι-αντι-απο-παρα	κόβομαι	36.4908	12	2	Singleton	49
παρα-συνε	βγαλμένος	42.9316	12	5	Singleton	71
απα-απο	λησμονιά	75.2622	12	6	PF1	76
απ-παρ-συν-συνδι	αλλιάσσομαι	48.6419	12	3	PF1	116
απ.παρ	αιτούμαι	82.8944	15	7	PF1	79
απ.παρ.συν	ακούω	31.9818	6	2	PF1	30
απ.παρ.συνδι	αλλιάσσω	36.2128	6	2	PF1	42
απ.συνδι	αιτητής	33.2553	4	2	PF1	20
απ-αποδι	ομπάος	64.8117	11	5	PF1	78
απο-συνεκ-παρα	δίδουσι	35.0439	10	2	Singleton	34

Command Line	Graphic Display	DCN Stress	DCN Syllabification
απο.κατα			
Stems:			
βίβαση	ζήτηση	γοήτευση	κάλυψη
κρημνίζομαι	λυμαίνομαι	γυμνώνομαι	κόπτομαι
θραύομαι	γοητεύομαι	δεικνύομαι	ζητούμαι
κείνες	διωκτέος	κλεισμένος	γκρεμισμένος
γοητευμένος	δεικτικός	βιβαστικός	γοητευτικός
βιβασμός	κλεισμός	κυλίω	λήγω
κομίζω	κιτρινίζω	λουζώ	βρίσκω
κλείνω	διώνω	γυμνώνω	πασπώνω
καλύπτω	κρύπτω	θραύω	γοητεύω
κυλίσω	καλύπτομαι	καλόκαιρο	κιτρινισμένος
καίγω	διωγμός	καίγω	λευκαίνω
κόπτω	δεικνύω	κόπτω	δεικνύω

Εικόνα 2: Δείγμα από τα φύλλα σώματος επιθημάτων του ΣΚ2.

### 5.1.4.2. Επιθηματοποίηση

Το επιθηματικό σύστημα της Ελληνικής Γλώσσας είναι αρκετά πολύπλοκο· όπως παρουσιάζουν οι Μελισσαροπούλου και Ράλλη (Μελισσαροπούλου 2007, Μελισσαροπούλου & Ράλλη 2009, Melissaropoulou & Ralli 2009), μπορούν να έπονται ενός θέματος έως και έξι επιθήματα (παραγωγικά και κλιτικά στο σύνολο). Σίγουρα μία τέτοια κατάτμηση είναι υπερβολικά δύσκολη για το σύστημα αυξάνοντας κατακόρυφα το βαθμό δυσκολίας για την ανάλυση. Το LINGUISTICA πέτυχε να εντοπίσει μερικά κλιτικά παραδείγματα των ρημάτων (ενεστώτας) και ονοματικών



τάξεων, όπως *ω.εις.ει.ουμε.ουτε.ουν* (γράφω, τρέχω, πίνω, δένω), *ω.ας.α.αμε.ατε.αν* (αγαπώ, μιλώ, διαιρώ), *ο.ου.ων.α.[ακι]* (βουνό, μωρό, νερό) και *ος.ων.η.ων* (δάσος, άλσος, βάρος). Εκτός από άλλα δεκαεπτά φύλλα σώματος, τα υπόλοιπα από αυτά ήταν λανθασμένα (78) ή ψεύτικα επιθήματα (8). Υπήρχε ένας σημαντικά μεγάλος αριθμός από συνδυασμένα επιθήματα (συνήθως παραγωγικά επιθήματα με τις καταλήξεις, τα επιθήματα όψης με τα κλιτικά επιθήματα), όπως *άρεις.αρει.αρω.αρουμε.αρισα, ατζη.ατζης.ατζηδες.ατζηδων* ή *ευτικός.ευτικοί* (χορός, δήμος). Ήταν ιδιαίτερο κοινό λάθος να προσκολλάται χαρακτήρας θέματος ή επιθήματος στο μόρφημα που προηγείται ή έπεται πιο συγκεκριμένα, το φύλλο *γα.ζα* (ανοι-, τυλι-, διαλε-) ή το φύλλο *ινα.να* (γλυκα-, πικρα-, λευκα-, τρελα-).

Signatures	Exemplar	Descr. Leng	Corpus Count	Stem Count
α-ε-εξ-ουμε-ουν-ω	αγριευσ	630.203	513	58
ο-ουμ-ε.ουν-ω	αγαπησ	261.337	69	21
α.ε.εξ.ουμ-ε.ουν	ζεσταν	82.9587	25	5
ε.εξ	αμαν	85.9942	12	6
ομαι-ω	ανθις	2983.7	483	240
α-ε-εξ-μενος-ουμ-ε-ουν-ω	αγιασ	323.603	264	28
α-ος	ανθρωπογραφικ	1448.54	232	116
ικα-ικος-ικως	ανκτ	809.024	201	67
ικα.ικως	αντικοινοβουλευτ	99.3132	14	7
ζα-ζαμε-ζαν-ζατε-ζε-ζει-ζεις-ζεσε-ζεσε-ζουμ-ζουν-ζω-ζαμε-σαν-σατε-σει-σεις-σετε	αγια	255.841	162	9
εια-ειαμ-ε-ειαν-ειατε-εμε-εμει-εμεις-εμεις-εμετε-εουον-εω-ειαμ-ε-εμιαν-εματε-εμει-εμεις-εμετε	πατρ	246.703	136	8
α-αμε-αν-ατε-ε-ει-εις-εξ-εξε-οντας-ουμ-ε-ουν	αγκων	121.836	96	4
α.αμε	αποκταγ	50.2956	6	3
ο-ος	ανθηλι	519.95	83	41
α-α-ησα-ησαμ-ησαν-ησατε-ησε-ησει-ησεις-ησεσε-ησετε-ουν-ουσα-ουσαν-ουσε-ουσεσε-ω	αγαπ	182.601	78	4
α.α.ουσα.ουσαν.ουσε.ουσεσε.ω	κατ	83.9961	21	3
α.α.ς	καβγ	50.2095	6	3
ις-ιτε-σαμ-ε-σαν-σατε-σει-σεις-σετε	αναιρε	117.913	72	4

Command Line	Graphic Display	DCN Stress	DCN Syllabification
εια.ειαμ.ε.ειαν.ειατε.εμε.εμει.εμεις.εμεις.εμετε.εουον.εω.ειαμ.ε.εμιαν.εματε.εμει.εμεις.εμετε			
Stems:			
μπερδ λογικ παιν ψαρ πατρ			
παντρ πιστ φυτ			

Εικόνα 3: Φύλλα σώματος από κλιτικά και παραγωγικά επιθήματα (ΣΚ2).

Ο Goldsmith προσπάθησε να διορθώσει το ζήτημα της ένωσης των επιθημάτων με προχωρημένες στρατηγικές ευρεστικών και εφαρμόζοντας την τεχνική «μικρό-μήκος για μη-θέματα» (short-length for non-stems): ωστόσο, η επεξεργασία των μονοφωνηματικών επιθημάτων (-ς) και προθημάτων (*α-στερητικό*) είναι ένα σημαντικό ζήτημα που προκαλεί σοβαρό πρόβλημα σε ένα σύστημα MeMM. Το σημαντικότερο, όπως προβλέψαμε στις υποθέσεις μας, το LINGUISTICA απέτυχε να εντοπίσει οποιαδήποτε επιθηματική αλλομορφία, μίας και το σύστημα δεν συσχέτισε διαφορετικές επιθηματικές μορφές και συνήθως τα οριοθετούσε λανθασμένα (48,55% αποτυχία). Επομένως, αναγνώρισε ως επιθήματα τα *αρω.αρισα* αντί για τα *αρ~αρι* (*αρ<ι>*), τα *ατζής.ατζήδων.ατζήδες* αντί για τα *τζη~τζηδ* (*τζη<δ>*) κτλ. Ελέγχοντας

προσεκτικά τα δεδομένα, η ακρίβεια του συστήματος για τα παραγωγικά επιθήματα ήταν 14,29%, ενώ 22,23% για τα κλιτικά επιθήματα<sup>7</sup>.

### 5.1.4.3. Θέματα/ Ρίζες

Το LINGUISTICA ακολούθησε την ίδια πεπατημένη για την ανάλυση των ονοματικών θεμάτων. Πρωτίστως να αναφερθεί πως μόνο τα ονοματικά αλλόμορφα τύπου-*παιδί* εντοπίστηκαν από το σύστημα. Σε όλες τις άλλες περιπτώσεις, εάν υφίσταται αφαιρετική αλλομορφία (V-διαγραφή, π.χ. *καρδιά~καρδι*), το σύστημα εντόπιζε μόνο το θέμα με τη V-διαγραφή (*καρδι-*) θεωρώντας το διαγραφόμενο φωνήεν ως επιθήμα. Στον αντίποδα, στις περιπτώσεις της προσθετικής αλλομορφίας (C-εισαγωγή, π.χ. *κύμα~κυματ*), το μοντέλο βάσιζε την ανάλυσή του πάνω στην υπόθεση ότι το σύμφωνο αποτελούσε τον εναρκτήριο χαρακτήρα των επιθημάτων (*κύμα*). Επιπροσθέτως, υπήρχαν κάποια λανθασμένα φύλλα σώματος, όπου οι δύο τελευταίοι χαρακτήρες του θέματος, όπως *νας.να.νες.νων* (σωλη-, πυρη-, αιω-, λιμε-, καντι-) και *μα.ματα.ματος.ματων* (βη-, σω-, πλεγ-, τριμ-) αποτελούσαν εναρκτήριο μέρος επιθήματος. Το σύστημα απέτυχε να συσχετίσει κάποιο θέμα ειδικά με τους τύπους αορίστου ομαλών και ανωμάτων ρημάτων και έδειξε να εμφανίζει μεγάλη ποικιλία στην λανθασμένη ανάλυση.

Παράλληλα, η στατιστική ανάλυση όλων των σωμάτων κειμένου κατέδειξε ότι μόλις το 4% των αλλομόρφων ανιχνεύτηκαν από το LINGUISTICA. Οι μετρήσεις μας είναι παρόμοιες με τις χαμηλές μετρήσεις των συγκρίσεων της ομάδας του Kurimo για τα Φιλανδικά και Τουρκικά (κυρίως σε παλαιότερες δημοσιεύσεις τους για τον διαγωνισμό μορφολογικών αναλυτών, πρβλ. Kurimo *et al.* 2006a, 2006b, Kurimo, Creutz & Varjokallio 2007· στο διαγωνισμό Morpho Challenge οι Kurimo, Virpioja & Tumunen 2009 αναφέρουν ότι οι επιδόσεις των συστημάτων είναι βελτιωμένες). Η υπόθεση για την αποτυχία του LINGUISTICA όσον αφορά στην αλλομορφία των Ελληνικών εκπεφρασμένη από τον Karasimos (2008) δοκιμάστηκε πειραματικά σε δύο φάσεις (Karasimos 2009, Karasimos & Petropoulou 2010, παρούσα εργασία) και αποδείχτηκε με μεγάλη ακρίβεια αληθινή.

---

<sup>7</sup> Θεωρούμε ως πραγματικά φύλλα σώματος, εκείνα που περιέχουν στο σύνολό τους αληθινά επιθήματα. Φυσικά, κάποια φύλλα σώματος δεν συμπεριλαμβάνουν όλες τις καταλήξεις ενός ονοματικού ή ρηματικού παραδείγματος, χωρίς ωστόσο αυτό να θεωρείται λάθος.

Signatures	Mini /	Exemplar	Desc. Length	Corpus Count	Sigs	Remarks
ά.άδες.άς	αλογ	26.731	3	1	Check sigs	8
γα.ξα	άνοι	43.5492	8	4	SF1	29
ικός.ών	αντρ	41.7779	8	4	SF1	32
ινος.ου.ων	αργύρ	40.5731	9	3	SF1	42
ου.ων	άρθρ	34.7184	6	3	SF1	22
ινος.ων.ου.ου	αρκούδ	36.5525	3	1	Known stems to suffixes	18
ζω.ιστή.ιστής	βαπτ	48.4072	12	4	SF1	76
ικός	βασιλ	63.8619	7	7	SF1	24
ικος	βεζίρ	57.0721	6	6	SF1	20
ματάκι.ματάρα	βη	43.5492	8	4	Check sigs	46
ά.άδες.άς.αδάκι	βορι	46.0905	12	3	Known stems to suffixes	63
ευτής.ευτικός	βουή	88.0206	22	11	Check sigs	163
ά.άς.ές.νός.ών	βραδι	42.6864	5	1	Loose fit	20
άδες.άς.ά.ά	γαλατ	40.8237	6	2	Known stems to suffixes	41
ά.άδες.άς.ιού	γιαουρτ	37.0285	4	1	Loose fit	21
άρα.ές.ικός.ών	γραμμ	51.1773	16	4	SF1	78
ά.άς.ές.ών	γωνι	43.9105	12	3	From known stem and suffix	53
είς.ικός	εκδρομ	29.2365	4	2	SF1	20
ατζίδες.ατζής	ετοιμ	23.6773	2	1	Loose fit	5
ές.εδάκι	ζεή	48.2158	10	5	Known stems to suffixes	49
ά.άς.ατζήδες.ατζής	ζουρν	52.709	16	4	From known stem and suffix	96
άκι.ζω.ιστή.ιστής	θερ	37.4848	4	1	SF1	9
ανθείς.ανθώ	θερμ	49.7152	10	5	Check sigs	58
αινα.ανα	θέρμ	55.999	12	6	Check sigs	56
ά.άς.αδάκι	κουβ	28.0923	3	1	Loose fit	8
ά.άς.άτων.αδάκι	κυμ	35.8372	4	1	Loose fit	9
ά.άς.ατζήδες	ζουρμ	42.1972	8	2	Loose fit	54

Εικόνα 4: Φύλλα σώματος από ουσιαστικά, ρήματα και επίθετα.

#### 5.1.4.4. Σύνθετα και δεσμευμένα θέματα

Εξίσου πολύπλοκες μορφές αποτελούν τα σύνθετα και τα δεσμευμένα θέματα για ένα σύστημα που απουσιάζουν κάθε είδους κανόνες με δομική πληροφορία και επομένως η δοσμένη ανάλυση από το σύστημα είναι το πιθανότερο να μην είναι η επιθυμητή, που όντως ισχύει και στη συγκεκριμένη ομάδα λέξεων. Ειδικότερα, μεταξύ των παραχθέντων φύλλων σώματος από την ανάλυση των δεσμευμένων θεμάτων του ΣΚ2+, εντοπίστηκαν ‘πραγματικά’ επιθήματα, όπως τα παραγωγικά *-ία* (π.χ. *θεολογ-ία*, *αρχαιολογ-ία*), *-εί(ο)* (π.χ. *ανθοπωλ-είο*, *φωτογραφ-είο*), *-τη(ς)* (π.χ. *αιμοδό-της*, *καταδό-της*), *-ίσσα* (π.χ. *παλαιοπώλ-ίσσα*), ονοματικά κλιτικά (*-ος*) (π.χ. *βοτανολόγ(ος)*, *αγιογράφ(ος)*) και ρηματικά κλιτικά (*-ω*) (π.χ. *ηχογραφ(ώ)*, *πυροβολ(ω)*). Παράλληλα, εντοπίστηκαν πληθώρα δεσμευμένων θεμάτων, όπως *-ολόγος*, *-ολογία*, *-ογράφος*, *-ομανής*, *-οποία*, *-οτρόφος*, *-οστάτης*, τα οποία είναι σαν τα συμφύματα με το συνθετικό φωνήεν προσκολλημένο. Ταυτόχρονα και χωρίς κάποιο σαφή λόγο και στρατηγικό σχεδιασμό εντοπίσαμε μεταξύ των φύλλων σώματος και μορφήματα, όπως *-φάγος*, *-παθής*, *-σκοπία*, *-ούχος*, *-γένης*, *-άρχης*, *-φόρος*, που μοιάζουν με τα συμφύματα αλλά χωρίς το συνδετικό στοιχείο. Αποτελέσματα σαν αυτά καταδεικνύουν ότι το σύστημα απέτυχε να αναγνωρίσει το συνδετικό φωνήεν *-ο-* ως ξεχωριστό στοιχείο, ούτε τα παραγωγικά και κλιτικά επιθήματα που ήταν συνδυασμένα με τα δεσμευμένα θέματα.

Όντως είναι λογικό και αναμενόμενο, η προσκόλληση του συνδεδειμένου φωνήεντος στα νεοκλασικά θέματα, όπως αναλύθηκε παραπάνω, επέτρεψε στο σύστημα να δώσει πληθώρα 'σωστών' θεμάτων/ ριζών<sup>8</sup> της ΚΝΕ. Για παράδειγμα, *μυθ-*, *ωκεαν-*, *δενδρ-*, *νερ-* ή θεματικά αλλόμορφα, όπως *μουσικ-*, *παιδ-*, *δραματ-*, *χρωματ-*, *στοματ-*, *νηματ-* και ούτω καθ' εξής, εμφανίζονται στο δεξί κομμάτι του σχηματισμού. Ωστόσο, αναγνωρίζονται ως θέματα και ακουλουθίες λέξεων (συνθετικά θέματα), όπως *κρυπτογραφ-*, *σεισμολογ-*, *πρωτογον-*, *βιβλιοκλοπ-*, *τεχνολογ-*, *πλουτοκρατ-*, εξαιτίας των παραγωγικών και κλιτικών επιθημάτων που αναλύθηκαν προηγουμένως.

Εν κατακλείδι, οφείλουμε να σημειώσουμε ότι το σύστημα δεν κατάφερε να διαχειριστεί δεσμευμένα θέματα τύπου *-λογ-*, *-γραφ-*, *-κτον-*, *-μαθη-* κ.α., ούτε το συνθετικό στοιχείο *-ο-*, όπως έχει προταθεί από διάφορες μελέτες (ενδεικτικά Ράλλη 2008, Πετροπούλου υπό επεξεργασία) για λέξεις που περιέχουν δεσμευμένα θέματα στην ΚΝΕ. Συγχρόνως το LINGUISTICA δεν μπόρεσε να αναγνωρίσει κανένα από τα σύνθετα. Η στρατηγική του και η αρχιτεκτονική του είναι δομημένη για την εξαγωγή επιθημάτων (κυρίως) και προθημάτων ακόμα και για γλώσσες με πλούσια μορφολογία. Τα σώματα αγγλικών κειμένων που δοκιμάστηκαν σε αυτό το σύστημα περιείχαν ελάχιστα μονολεκτικά σύνθετα, καθώς και πολύ λίγα νεοκλασικά σύνθετα: οι ερευνητές δεν κάνουν κανένα σχόλιο σχετικά με αυτές τις περιπτώσεις. Δυστυχώς τα τρία σώματα ελληνικών κειμένων που δοκιμάστηκαν αδυνατούν να αποτελέσουν σοβαρό πεδίο αξιολόγησης για σύνθετα, μίας και τα αποτελέσματα είναι εντελώς τυχαία: εξάλλου κάτι τέτοιο δεν αποτελεί και στόχο της παρούσας διατριβής. Σε γενικές γραμμές θα υποστηρίζαμε, το σύστημα δεν ήταν σε θέση να αναγνωρίσει κανένα από τα συστατικά και απέτυχε να αναλύσει κάποιο από αυτά.

---

<sup>8</sup> ΣΗΜΕΙΩΣΗ: χωρίς τα κλιτικά επιθήματα, όπως αυτά εμφανίζονται στα συγκεκριμένα σύνθετα.

Stem	Phonological content	Length ptr to me	Corpus count	Suffix sig
ώτο	15.3878	9.53867	8	NULL. ειδής. λογία. λόγος. σκόπιο
υδρ	15.3878	9.53867	8	αμία. ογόνο. ολογία. ολόγος. οφόρος. όβιος. όψηση. όφιλος
μηχαν	25.6464	9.53867	8	αγράφος. ολογία. ολόγος. οποιώ. ουργία. ουργείο. ουργός. όβιος
θεο	15.3878	9.53867	8	γενής. γονία. ειδής. κράτης. κρατία. μαχία. ποιώ. φόρος
ηλεκτρο	35.905	9.53867	8	βόρος. γόνος. λογία. λόγος. νόμος. σκόπιο. τεχνία. φόρος
καϊκ	20.5171	9.73132	7	αγράφος. οειδής. ουργία. ουργείο. ουργός. οφόρος. ούχος
ουσι	15.3878	9.73132	7	αγράφος. οειδής. οποιώ. οπωλείο. ουργία. ουργείο. ουργός
σεισμο	30.7757	9.73132	7	γενής. γόνος. λογία. λόγος. παθής. σκόπιο
πεταύ	25.6464	9.73132	7	οειδής. οποιία. οποιείο. οποιός. ουργία. ουργείο. ουργός
ελληνο	30.7757	9.73132	7	γενής. μάχος. μαθής. ποιώ. πρέπεια. πρεπής. φωνία
γλωσσο	30.7757	9.73132	7	γονία. γράφος. γραφία. μάθεια. μαθής. πηασία. πηαστία
βιο	15.3878	9.73132	7	γονία. γράφος. λογία. λόγος. νομία. τέχνης. τεχνία
βιβλιο	30.7757	9.73132	7	γράφος. δέτης. δεσία. κλοπή. πώλης. τεχνία. φάγος
ανθρωπο	35.905	9.73132	7	γενής. γονία. ειδές. ειδής. φάγος. φαγία. φοβία
ωο	10.2586	9.95371	6	γονία. γόνος. ειδής. σκοπία. σκόπιο. φόρος
χρωματο	35.905	9.95371	6	γραφία. γόνος. λόγιο. σκοπία. σκόπιο. φόρος
χρωματ	30.7757	9.95371	6	οποιία. οποιείο. οποιός. ουργία. ουργείο. ουργός
καρτ	20.5171	9.95371	6	αγράφος. ομανής. ομανία. οποιία. οποιός. οπωλείο
τουρκο	30.7757	9.95371	6	γενής. κρατία. μαθής. φάγος. φιλία. φοβία
τοπ	15.3878	9.95371	6	άρχης. αρχία. αγράφος. οθεσία. οθετώ. ολογία
στερεο	30.7757	9.95371	6	βάτης. γραφία. ποιώ. σκοπία. σκόπιο. τομία
πυρο	20.5171	9.95371	6	γράφος. γραφία. πάθεια. παθής. τέχνης. φοβία
ξυλι	15.3878	9.95371	6	αγράφος. οειδής. οσάτης. ουργία. ουργείο. ουργός
νηματ	25.6464	9.95371	6	οειδής. οποιία. οποιός. ουργία. ουργείο. ουργός
νεο	15.3878	9.95371	6	γενής. λογία. πηασία. πηαστία. φανής. φοβία
νεκρο	25.6464	9.95371	6	γενής. φάγος. φανής. φιλία. φοβία. φόρα
μυθ	15.3878	9.95371	6	αγράφος. ολόγος. ομανής. ομανία. οποιία. οποιός
μον	15.3878	9.95371	6	άρχης. αρχία. οθεσία. ομανής. ομανία. όφωνος
μικρο	25.6464	9.95371	6	γράφος. πρέπεια. πρεπής. σκοπία. σκόπιο. τεχνία
κακ	15.3878	9.95371	6	αγράφος. ολόγος. οποιός. οποιώ. ουργία. όφωνος
ιερ	15.3878	9.95371	6	άρχης. αρχία. οδικής. οδικείο. ουργία. ουργός
δραματ	30.7757	9.95371	6	αγράφος. ολογία. ολόγιο. ουργία. ουργός
γεροντο	35.905	9.95371	6	κρατία. λογία. λόγος. φιλία. φοβία
γαϊακτ	30.7757	9.95371	6	οποιώ. οπωλείο. οπώλης. στροφία. οφόρος. ούχος
μυτο	20.5171	10.2162	5	νοσσία. κύκος. τεχνία. φάγος. πηνία

Εικόνα 5: Φύλλα σώματος από δεσμευμένα θέματα.

### 5.1.5. Σύντομος σχολιασμός του πειράματος

Σχεδόν η πλειονότητα των μοντέλων μη-εποπτευόμενης μορφολογικής μάθησης αντιμετωπίζουν το οποιοδήποτε γλωσσικό σύστημα που εξετάζουν χωρίς εξωγενείς βοήθειες και εισαγόμενους κανόνες, ωστόσο επιτυγχάνουν ικανοποιητικές επιδόσεις σε γλώσσες με φτωχή μορφολογία. Στην αρχική μας υπόθεση (Karasimos 2008) είχαμε υποστηρίξει ότι πλούσιες μορφολογικά (ειδικότερα σε διαχυτικές) γλώσσες με πολυπλοκότητα και συνδυαστικότητα στην παραγωγή, όπως τα Ελληνικά, θα αμαυρώσουν την εικόνα ενός συστήματος MeMM. Ήδη οι προηγούμενες προσπάθειες να αξιοποιηθεί αυτό το θεωρητικό μοντέλο σε μορφολογικά πλούσιες συγκολλητικές γλώσσες, όπως τα Φιλανδικά και τα Τουρκικά, είχαν εμφανώς μετριότατες αποδόσεις. Σε παρόμοια μονοπάτια επιδόσεων, το δικό μας πείραμα απέδειξε ότι ένα μοντέλο MeMM αναλύει με χαμηλή ακρίβεια την Ελληνική γλώσσα. Με βάση τα αποτελέσματα, υποστηρίζουμε ότι ένα σύστημα: α.) χωρίς χειρωνακτική προ-ανάλυση των γραμματικών μορφημάτων μίας γλώσσας, β.) κάποια χαρακτηρισμένα μορφήματα και προσφύματα και γ.) προ-εισαγμένους φωνολογικούς και μορφολογικούς κανόνες, θα αποτύχει να αγγίξει υψηλές επιδόσεις.

Ένα σύστημα που κατασκευάζει λεξικά πάνω σε μία κοινή ακολουθία μορφημάτων χωρίς κατάλληλους κανόνες ή περιγραφικές διαδικασίες αδυνατεί να αντιμετωπίσει την αλλομορφία. Όπως έχουμε δείξει ήδη, η αλλομορφία στα Ελληνικά είναι κοινό χαρακτηριστικό, είναι συνήθως ο κανόνας και όχι η εξαίρεση. Συμμετέχει με την ίδια συχνότητα σε κάθε διαδικασία σχηματισμού λέξεων<sup>9</sup> επομένως, εγέρθηκε μία λογική απορία απέναντι σε ένα μοντέλο MeMM, για το αν είναι ικανό χωρίς κάποια βοήθεια να αναλύσει, να επεξεργαστεί και να αντιμετωπίσει επιτυχώς τα αλλόμορφα. Παρατέθηκαν πολλά παραδείγματα αλλομόρφων από τη λανθασμένη ανάλυση του LINGUISTICA και έχει τεκμηριωθεί η πολυπλοκότητα των αλλομορφικών αλλαγών παράλληλα με το συστηματικό χαρακτήρα εμφάνισης των αλλομόρφων. Αφού η εισαγωγή κανόνων επεξεργασίας της αλλομορφίας δεν επιτρέπεται σε ένα σύστημα MeMM, ο στόχος για υψηλή ακρίβεια στη μορφολογική ανάλυση δεν επιτυγχάνεται. Συνοψίζοντας επισημαίνουμε πως μόνο τα μοντέλα εποπτευόμενης μορφολογικής μάθησης με κανόνες και εισαγμένη ανθρώπινη γνώση πληρούν τις βασικές προϋποθέσεις να αντιμετωπίσουν με επιτυχία την αλλομορφία.

## 5.2. Το πείραμα του AMIS για την ονοματική κλίση της KNE

Στο πλαίσιο της προβληματικής για τα Μη-Επιβλεπόμενα Μορφολογικής Μάθησης Μοντέλα (Καρασίμος 2009, Karasimos 2009) είχαμε υποστηρίξει ότι ένα μοντέλο Εποπτευόμενης Μορφολογικής Μάθησης έχει αυξημένες πιθανότητες να αντιμετωπίσει την αλλομορφία με επιτυχία. Στην προσέγγιση των Dell' Orletta *et al.* (2007) για την αξιοποίηση των περιορισμών στην πρόβλεψη του υποκειμένου και αντικειμένου στα Ιταλικά και στα Τσέχικα κατέφυγαν στην τεχνική της Μέγιστης Εντροπίας. Έτσι δημιουργήθηκε η ιδέα<sup>9</sup> να δοκιμαστεί ένα μοντέλο για την ονοματική αλλομορφία των Ελληνικών, ώστε να γίνει μία προσπάθεια πρόβλεψης της αλλομορφίας που θα εμφανίσουν τα ουσιαστικά με τη συνδρομή διάφορων χαρακτηριστικών, τα οποία θα είναι μορφολογικά, φωνολογικά και σημασιολογικά. Στόχος είναι ο εντοπισμός των βαρών (weights) που έχουν άμεση εξάρτηση με την αλλομορφία και θα συμβάλλουν στο σχεδιασμό ενός μοντέλου πρόβλεψης. Το μοντέλο

---

<sup>9</sup> Να ευχαριστήσω τον Vito Pirrelli, καθώς και τον Felice Dell' Orletta και Simonetta Montemagni για τον αρχικό σχεδιασμό του συγκεκριμένου πειράματος, καθώς και τον Kazuhiro Yoshida για την πολύτιμη βοήθεια κατά την εκτέλεση του πειράματος.

αυτό δεν θα αφορά μόνο στη θεματική ονοματική αλλομορφία, αλλά και την παραγωγική ονοματική αλλομορφία.

### 5.2.1. Σχετικά με το AMIS

Το συγκεκριμένο λογισμικό είναι βασισμένο στα μοντέλα μέγιστης εντροπίας (Berger, Della Pietra & Della Pietra 1996) και είναι ελεύθερο λογισμικό<sup>10</sup> για την αξιοποίηση των δεσμών χαρακτηριστικών (Yoshida 2006). Προσφέροντάς του ένα σύνολο συμβάντων για εκπαίδευση, το πρόγραμμα δίνει ένα εξαγόμενο που βελτιστοποιεί τις πιθανότητες του σώματος εκπαίδευσης. Το λογισμικό υποστηρίζει τις ακόλουθες λειτουργίες:

(α.) αλγόριθμος παραμετρικής εκτίμησης για δέσμη χαρακτηριστικών (feature forests, βλ. Miyao & Tsujii 2002),

(β.) αλγορίθμους παραμετρικής εκτίμησης με περιορισμένη μνήμη BFGS και GIS (limited-memory BFGS and GIS, πρβλ. Nocedal 1980, Darroch & Ratcliff 1972 αντίστοιχα),

(γ.) πρόβλεψη MAP με τη χρήση κατανομών τύπου Gauss (Chen & Rosenfeld 1999),

(δ.) εκτίμηση των μοντέλων Μέγιστης Εντροπίας με περιορισμούς ανισότητας (Kazama & Tsujii 2003, 2005) με τη χρήση του αλγόριθμου BLMVM (Benson & More 2001) και

(ε.) συμπίεση των αρχείων συμβάντων.

Το πείραμα έτρεξε σε περιβάλλον Linux (σε διανομή CentOS v5.0) με τις απαραίτητες βιβλιοθήκες c++ και g++ για την έκδοση Amis 4.0.

### 5.2.2. Αναπαράσταση γλωσσολογικών πληροφοριών σε σύνολα λεξικών χαρακτηριστικών

Προτού συνεχίσουμε στην μοντελοποίηση και τη δοκιμή του μοντέλου Μέγιστης Εντροπίας, είναι απαραίτητο να αναλογιστούμε την τροποποίηση της μορφολογικής δομής μίας λέξης σε μία αναπαράσταση χαρακτηριστικών υποκειμένου σε

---

<sup>10</sup> Για κατέβασμα του λογισμικού και περισσότερες πληροφορίες στην ιστοσελίδα του: <http://www-tsujii.is.s.u-tokyo.ac.jp/amis/>.

πιθανολογική ανάλυση. Ένα απλό παράδειγμα από αυτού του είδους της μεταβολής φαίνεται από τα παραδείγματα μεζές και άγαμμα του παρακάτω πίνακα. Τα λήμματα δίνονται στην αρχική οριζόντια γραμμή, ενώ στην πρώτη στήλη δίνονται τα διάφορα χαρακτηριστικά και στις υπόλοιπες οι τιμές που λαμβάνουν. Για κάθε λέξη δίνεται ένα εύρος γλωσσολογικών πληροφοριών για την κατηγοριοποίησή του, όπως μέρος του λόγου, μορφολογικά (γένος, κλιτική τάξη), φωνολογικά (συλλαβές, τόνος) και σημασιολογικά χαρακτηριστικά (δάνειο, έμψυχο). Να τονισθεί ότι δεν δίνονται όλα τα πιθανά χαρακτηριστικά στα συγκεκριμένα παραδείγματα. Για παράδειγμα, το λήμμα *άγαμμα* ανήκει στην όγδοη κλιτική τάξη (IC8), είναι ουδετέρου γένους (NEUTRAL) και εμφανίζει αλλομορφία (ALLOM\_T). Ταυτόχρονα χαρακτηρίζεται από τρεις συλλαβές (3SYL) και τόνο στην προπαραλήγουσα. Επίσης, είναι μία λέξη που δεν είναι ούτε παράγωγη, ούτε σύνθετη (NON-DER&NON-COMP). Κάποιες από τις πληροφορίες αυτές θα κωδικοποιηθούν τεχνικά για την αναπαράστασή τους· ζήτημα σημαντικό αποτελεί για τις πληροφορίες που προκρίνονται και πώς οι πολύπλοκες από αυτές θα αναπαρασταθούν.

Χαρακτηριστικά	ΜΕΖΕΣ	ΆΓΑΜΜΑ
Κλιτική Τάξη	IC2B	IC8
Γένος	MASCULINE	NEUTRAL
Αλλομορφία	ALLOM_Δ	ALLOM_T
Τύπος Λέξης	NON-DER&NON-COMP	NON-DER&NON-COMP
Συλλαβές	2SYL	3SYL
Τονισμός	ULTIMATE	ANTIPENULTIMATE
Καταγωγή	TURKISH	GREEK
Έμψυχο	NON-ANIM	NON-ANIM

Πίνακας 16: Δείγμα χαρακτηριστικών με κωδικοποίηση για Μοντέλο Εντροπίας.

Ο πίνακας 15 μας δίνει τη δυνατότητα να αντιληφθούμε την κωδικοποίηση της γλωσσολογικής πληροφορίας στο εσωτερικό μίας λέξης ή για τον ακόμα δυσκολότερο στόχο της δομής μίας πρότασης. Τα στοχαστικά μοντέλα είναι ιδιαίτερος κατάλληλα να καλύψουν τις απαιτήσεις ενός μοντέλου με σύνολο χαρακτηριστικών τέτοιου είδους. Ενασχολούμενοι με τη διαχρονία του ζητήματος της αλλομορφίας είναι ασφαλές να διατυπωθεί ότι τα αλλόμορφα είναι κατάλοιπα μη-ενεργών φωνολογικών και μορφολογικών κανόνων και αλλαγών των Ελληνικών. Έτσι κάνουμε την υπόθεση ότι ίσως οι λέξεις των Ελληνικών να ενέχουν τις απαραίτητες πληροφορίες, ώστε ένα σύστημα με ελάχιστη επιτήρηση να προβλέψει εάν μία λέξη έχει αλλομορφία και εάν ναι, τι είδους αλλόμορφα σχηματίζει.



### 5.2.2.1. Μία στατιστική αναπαράσταση της Ονοματικής Αλλομορφίας

Έχοντας παρουσιάσει αρκετά γλωσσολογικά χαρακτηριστικά είναι σχετικά εύκολο να υλοποιήσουμε μία στατιστική αναπαράσταση του ζητήματος της Ονοματικής Αλλομορφίας. Πρέπει να υπολογίσουμε ποιες είναι οι περισσότερες πιθανότητες για τη λέξη *παπάς* να ενταχθεί σε κάποια αλλομορφική τάξη, με βάση αποκλειστικά τις γλωσσολογικές πληροφορίες της προαναφερθείσας λέξης. Σε αντίθεση με τα διάφορα ζητήματα της σύνταξης, όπου εκτός από τις εγγενείς πληροφορίες της λέξης-στόχου, αξιοποιούνται πλήρως και οι πληροφορίες περιβάλλοντος, στη μορφολογία φαίνεται πως είναι δύσκολο (και πιθανότατα μη απαραίτητο) να αξιοποιηθούν οι περιφερειακές πληροφορίες.

Αναλυτικότερα, ονομάζουμε  $AX_{\text{παπάς}}$  την αναπαράσταση χαρακτηριστικών της λέξης *παπάς*. Με βάση τον πίνακα (16.), το  $AX_{\text{παπάς}}$  ανταποκρίνεται σε μεγάλο βαθμό στα συγκεκριμένα γλωσσολογικά στοιχεία. Η πιθανότητα της λέξης *παπάς* να εμφανίσει αλλομορφία είναι η δεσμευμένη πιθανότητα  $p_{AL}(\text{παπάς}_{\text{αλλομ}} | AX_{\text{παπάς}})$ , τουτέστιν η πιθανότητα να εμφανίσει αλλομορφία με βάση μία συγκεκριμένη ομάδα γλωσσολογικών χαρακτηριστικών. Για να καλύψουμε το διπλό ερώτημα της αλλομορφίας (ναι/όχι, τύπος αλλομορφίας) στοχεύουμε στη δεσμευμένη πιθανότητα  $p_{AC}(\text{παπάς}_{\text{ΑλλομΤάξη}} | AX_{\text{παπάς}})$  που ερμηνεύεται ως η πιθανότητα να ενταχθεί σε συγκεκριμένη αλλομορφική τάξη (υφίσταται ως τάξη η μηδενική αλλομορφία) η λέξη *παπάς* με βάση κάποια χαρακτηριστικά. Ως χαρακτηριστικό-στόχο ορίζουμε το  $\text{παπάς}_{\text{ΑλλομΤάξη}}$ , ενώ το  $AX_{\text{παπάς}}$  είναι το σύνολο των γλωσσολογικών στοιχείων που αποτελούν τη βάση ανάθεσης του στόχου.

Προτού προχωρήσουμε στον καθορισμό των τελικών χαρακτηριστικών πρέπει να παρουσιάσουμε τον τρόπο υπολογισμού της δεσμευμένης πιθανότητας: ποιος είναι, δηλαδή, ο τελικός υποψήφιος που προκρίνεται για τη λέξη που αναζητάμε. Εάν το  $p_{AC}(\text{παπάς}_{AC7} | AX_{\text{παπάς}})$  είναι υψηλότερο από οποιαδήποτε  $p_{AC}(\text{παπάς}_{ACX} | AX_{\text{παπάς}})$ , όπου  $ACX$  οποιαδήποτε άλλη αλλομορφική τάξη, τότε ο *παπάς* εντάσσεται στην έβδομη αλλομορφική τάξη. Η επιλογή της οποιασδήποτε αντίστοιχης αλλομορφικής τάξης ακολουθεί το ίδιο μοτίβο. Για να είμαστε πιο σαφείς, υποθέτουμε ότι η πιθανότητα για τα αρσενικά ουσιαστικά που τονίζονται στην παραλήγουσα και είναι τουρκικής προέλευσης έχουν υψηλότερη πιθανότητα για την  $AC8$  από τα υπόλοιπα ουσιαστικά, δηλαδή  $p_{AC}(\text{μπακάλης}_{AC8} | \text{Male} + |\text{PenUlt} + |\text{Cal\_Tur})$  ή είναι απίθανο να

εμφανίσουν αλλόμορφα σε  $-\delta-$  τα ουδέτερα:  $p_{AC}(\text{σώμα}_{AC6} | \text{non-neutral}) < p_{AC}(\text{σώμα}_{IC6} | \text{neutral})$ . Με βάση αυτή την ανάλυση το *σώμα* δεν θα μπορούσε ποτέ να ενταχθεί στη συγκεκριμένη αλλομορφική τάξη.

Ωστόσο αυτό μπορεί να αποτελέσει σοβαρό πρόβλημα για το σύστημα, όπως επισημαίνουν οι Dell'Orletta *et al.* (2007). Εάν η  $p_{AC}(\text{σώμα}_{AC6} | \text{non-neutral})$  έχει μηδενικές πιθανότητες να εμφανιστεί, τι θα συμβεί εάν έχουμε περιπτώσεις που πιθανότατα να εμφανίσουν την εκδοχή της έκτης αλλομορφικής τάξης σε μη ουδέτερα ουσιαστικά. Πώς θα αντιμετωπίσει ένα πιθανοτικό σύστημα αυτές τις περιπτώσεις; Το αναμενόμενο είναι να αναλυθούν λανθασμένα και αυτό είναι πρόβλημα για όλα τα άπαξ λεγόμενα που θα βρεθούν στο σώμα δοκιμής. Όπως σε όλα τα συστήματα που βασίζονται σε κανόνες, όλοι οι πιθανοτικοί χαρακτηριστές δύνανται να αντιμετωπίσουν περιθωριακές και εξαιρετικές περιπτώσεις όσο είναι εφικτό να αναγνωρίζονται από τα γνώριμα χαρακτηριστικά που τις συνοδεύουν. Αυτό το πρόβλημα συναντάται πιο εύκολα σε συντακτικές μελέτες, μίας και προσεχτικά δομημένα σώματα εκπαίδευσης για συντακτικά φαινόμενα μηδενίζουν αυτές τις περιπτώσεις.

Συνήθως υπάρχουν και άλλες πληροφορίες χαρακτηριστικών που είναι ευαίσθητες για τις συγκεκριμένες περιπτώσεις, ώστε να μπορεί ένα πιθανοτικό μοντέλο να αντιμετωπίσει τις εξαιρέσεις. Σε όλες τις άλλες περιπτώσεις τα γενικά χαρακτηριστικά είναι εκείνα που προκρίνουν τον τελικό νικητή. Αυτό το ζήτημα υπογραμμίζει μία σημαντική εν γένει πτυχή των πιθανοτικών μοντέλων και εκείνων της Μέγιστης Εντροπίας. Πιο συγκεκριμένα, αποτελεί θεμελιώδη προϋπόθεση για την οικοδόμηση ενός γλωσσολογικού μοντέλου να επιλεγθούν εκείνα τα χαρακτηριστικά που σχετίζονται με το στόχο κατάταξης. Σε αντίθετη περίπτωση, το μοντέλο θα επιφορτιστεί με περιττές πληροφορίες, θα γίνει πιο δύσχρηστο και δυσκίνητο με πιθανότητα να οδηγηθεί σε διαφορετικές αναλύσεις. Ορθώς σημειώνουν οι Dell'Orletta *et al.* (2007) ότι «κανένα πιθανοτικό μοντέλο δεν μπορεί να παράγει μηχανικά λογικά αποτελέσματα βασιζόμενο σε λανθασμένη ή ελλιπή παρουσίαση των γλωσσολογικών δεδομένων».

### 5.2.3. Επιλογή Χαρακτηριστικών

Στην προηγούμενη υποενότητα δόθηκαν κάποια χαρακτηριστικά για το μοντέλο, καθώς το πλέον βασικό κομμάτι σχεδιασμού κάθε μοντέλου ΜΕ αποτελεί η επιλογή των χαρακτηριστικών, των οποίων τα ειδικά βάρη θα εκτιμηθούν με βάση τις κατανομές των δεδομένων. Καθότι δεν είχε δοκιμαστεί το συγκεκριμένο μοντέλο σε αμιγώς μορφολογικό ζήτημα βασιστήκαμε στην υπόθεση, ότι ίσως είναι καλό να δοκιμαστούν διάφορα μορφολογικά, φωνολογικά και σημασιολογικά χαρακτηριστικά για την περιγραφή των λέξεων, πάντοτε με τον «κίνδυνο» να δοθούν περιττές πληροφορίες. Ωστόσο μία τέτοια στρατηγική συντελεί στη δόμηση ενός πιθανοτικού μοντέλου με σεβασμό στη συνέπειά του με αξιόπιστες γλωσσολογικές γενικεύσεις. Εξάλλου αυτό θα αποτελέσει σημαντικό βήμα για τη στοιχειοθέτηση κάποιων θεωρητικών γενικεύσεων για την αλλομορφία με βάση συγκεκριμένα εμπειρικά στοιχεία, καθώς και την *πρώτη* προσπάθεια αξιοποίησης ενός τέτοιου πειράματος για τη μορφολογία οποιασδήποτε γλώσσας.

Τα χαρακτηριστικά αποτελούν συναρτήσεις τύπου  $f_{\chi_n}(\lambda, \Sigma)$ , όπου δοκιμάζεται ένα συγκεκριμένο στοιχείο  $\chi_i$  για το χαρακτηριστικό  $\lambda$  της λέξης που εμφανίζεται σε ένα σύνολο χαρακτηριστικών  $\Sigma$ . Για το συγκεκριμένο ΜΕ μοντέλο, επιλέξαμε διαφορετικού τύπου χαρακτηριστικά που δοκιμάζουν μορφολογικές, φωνολογικές και σημασιολογικές διαστάσεις των κατανομών της ονοματικής αλλομορφίας (σε αλλομορφικές τάξεις  $ACx$ ).

*Αλλομορφική Τάξη (AC)*. Είναι το χαρακτηριστικό για το οποίο αναζητείται συσχέτιση με όλα τα υπόλοιπα χαρακτηριστικά. Οι τιμές που λαμβάνει είναι από 1 (μηδενική αλλομορφία) ως 8 (διπλή αλλομορφία τύπου Δ-προσθήκη/ Η-διαγραφή). Πιο συγκεκριμένα, οι τάξεις είναι οι ακόλουθες: AC1 – Μηδενική Αλλομορφία, AC2 – Α- Διαγραφή, AC3 – Η- Διαγραφή, AC4 – Ι- Διαγραφή, AC5 – Δ-Εισαγωγή, AC6 – Τ- Εισαγωγή, AC7 – Η- Διαγραφή/Ε- Αντικατάσταση και AC8 – Δ-Εισαγωγή/Η- Διαγραφή.

*Κλιτική Τάξη (IC)*. Αυτό συμπεριλαμβάνει τις οκτώ κλιτικές ονοματικές τάξεις της Ράλλη (2000, 2005) που συμπεριλαμβάνει ουσιαστικά και παράγωγα ουσιαστικά. Η κλιτική τάξη αποτελεί βασικό χαρακτηριστικό για να περιορίσει σημαντικά το εύρος προβλέψεων της αλλομορφίας, καθότι ο τύπος αλλομορφίας αποτελεί ένα από τα δύο κριτήρια κατηγοριοποίησης των ονοματικών τύπων.

*Συλλαβές*: αριθμός + τύπος (*Syllables*, *Syllable\_XX*). Το διπλό αυτό χαρακτηριστικό ορίζει τον αριθμό των συλλαβών, αλλά και το χειρωνακτικό συλλαβισμό των λέξεων.

*Τόνος* (*Stress*). Όντας βασικό χαρακτηριστικό υποκατηγοριοποίησης των κλιτών λέξεων για τις παραδοσιακές γραμματικές, εισάγουμε το χαρακτηριστικό του τόνου για επιρροή στην αλλομορφική κατηγοριοποίηση. Χαρακτηρίζεται από τρεις τιμές.

*Τελευταίοι χαρακτήρες* (*Character\_X*). Χαρακτηριστικό παρόμοιο με αυτό των συλλαβών, που σχετίζεται άμεσα με την ομοιότητα των τελευταίων χαρακτήρων μίας ρίζας και παραγωγικά επιθήματα που ανήκουν στην ίδια κλιτική τάξη.

*Εμψυχότητα* (*Animacy*). Αποτελεί από τα βασικά σημασιολογικά χαρακτηριστικά, που χαρακτηρίζει εάν ένα ουσιαστικό είναι έμψυχο ή άψυχο. Βάση τυπολογικών ερευνών (πρβλ. Aissen 2003, Croft 2003) αποτελεί κεντρικό άξονα για σημασιολογικά και συντακτικά μοντέλα με χαρακτηριστική την **Ιεραρχία του Χαρακτηρισμού Εμψυχότητας** (*Animacy Markedness Hierarchy*, βλ. Woodford 1999, Dell'Orletta 2007). Η συγκεκριμένη ιεραρχία οδηγεί σε βοηθητικούς περιορισμούς ιδιαίτερα στη σύνταξη, ωστόσο στο παρόν πείραμα ακολουθήσαμε τη στρατηγική των Dell'Orletta (2007), καθώς το χαρακτηριστικό της εμψυχότητας δεν έχει τόσο ειδικό βάρος στη μορφολογία. Έγινε χρήση μίας περιορισμένης έκδοσης, όπου οι άνθρωποι και άλλα έμψυχα όντα χαρακτηρίστηκαν ως *animate* και όλα τα υπόλοιπα ως *inanimate*.

*Καταγωγή – Δάνειο* (*Origin*, *Calque*). Με αυτή τη δυάδα καθορίζεται, εφόσον κάποιο ουσιαστικό δεν είναι Ελληνικό, τι καταγωγής είναι. Σε παλαιότερες έρευνες (Χατζιδάκις 1905, 1907, Τριανταφυλλίδης 1941, Κουκκίδη 1950, Τσοπανάκης 1994, Αναστασιάδη-Συμεωνίδη 1997, Κοντοσόπουλος 2001, Adrados 2003: 416-424, Minniti-Γκώνια 2007) επισημάνθηκε η έντονη παρουσία δάνειων λέξεων στην Ελληνική γλώσσα, σε όλες τις φάσεις της εξέλιξης. [Να σημειωθεί ότι σίγουρα η ετυμολογία των λέξεων αποτελεί *εγκυκλοπαιδική* γνώση που μόνο μία μικρή μειοψηφία της γλωσσικής κοινότητας γνωρίζει και σίγουρα δεν διαδραματίζει ρόλο κατά τη γλωσσική κατάκτηση].

Σύμφωνα με τις μελέτες, υπογραμμίζεται ότι ομάδες δανείων εντάσσονται στο κλιτικό σύστημα της Ελληνικής και εμφανίζουν κοινά χαρακτηριστικά· ακολουθώντας τα βασικά αποτελέσματα των μελετών, ο Καρασίμος (2001) ανέφερε τις ιδιαιτερότητες της αλλομορφίας των δανείων. Ειδικότερα, είναι αξιοσημείωτο πως συγκεκριμένες αλλομορφικές ομάδες είναι αποκλειστικά δάνειες λέξεις, όπως για παράδειγμα τα

ουσιαστικά τύπου *μπακάλης* που είναι όλα τουρκικής καταγωγής (Καρασίμος 2009, 2011). Με δεδομένη τη θεωρητική σημαντικότητα αυτού του χαρακτηριστικού στο ζήτημα της αλλομορφίας, εντάσσεται στο μοντέλο ΜΕ στόχο την πιθανή διαλεύκανση του θολού τοπίου στην ποικίλη αλλομορφία των λέξεων της δεύτερης κλιτικής τάξης και όχι μόνο. Να σημειωθεί πως οι κατηγορίες του τύπου των συλλαβών και των χαρακτήρων δεν ανήκουν στα βασικά χαρακτηριστικά.

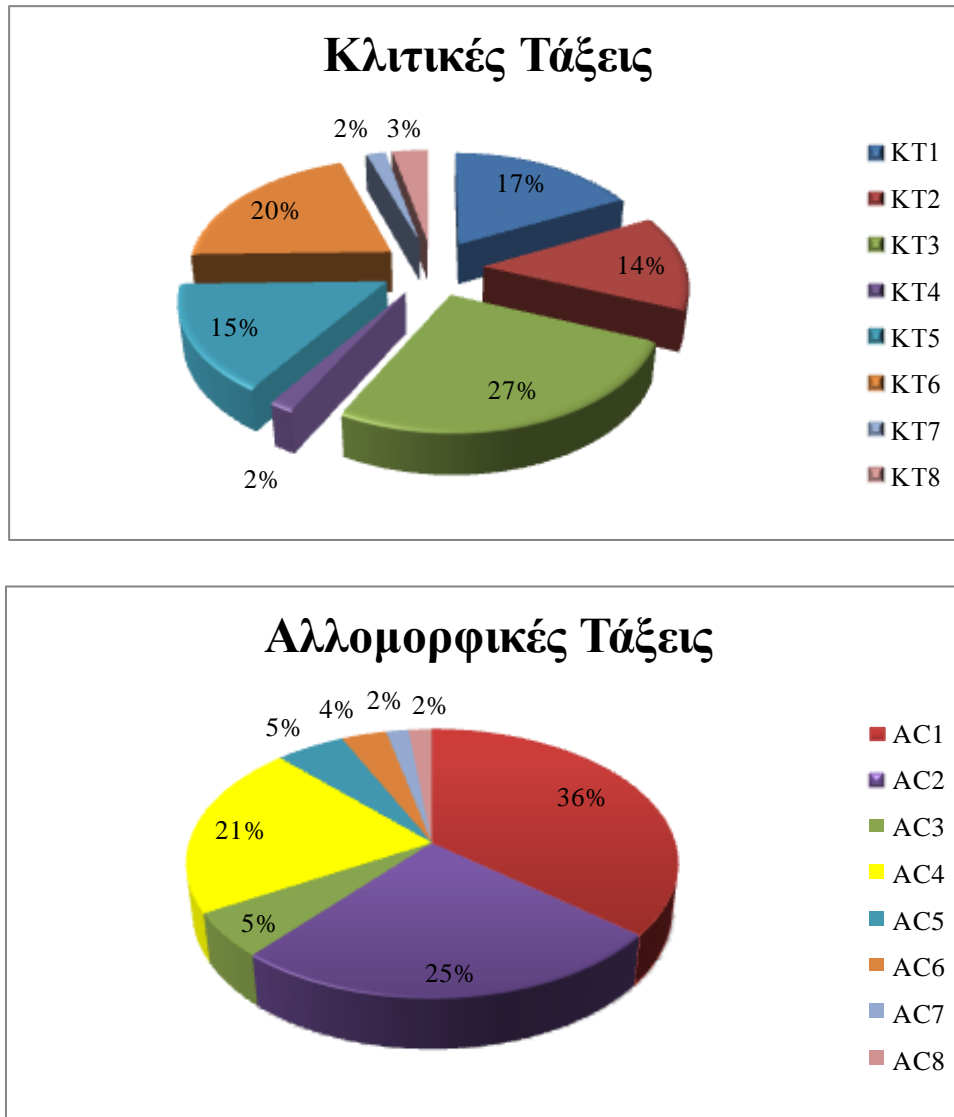
#### 5.2.4. Δοκιμή Ονοματικού Μοντέλου

Το Ελληνικό μοντέλο Μέγιστης Εντροπίας εκπαιδεύτηκε πάνω σε ένα σώμα 4.677 ουσιαστικών και των οκτώ κλιτικών τάξεων που δεν έχουν παραχθεί από κάποια άλλη λέξη ή συγχρονικά είναι μορφολογικά αδιαφανή με παραγωγικά επιθήματα ενσωματωμένα στο θέμα. Με βάση το ηλεκτρονικό σύνθετο λεξικό του Τριανταφυλλίδη<sup>11</sup> εισήχθησαν σχολαστικά όλα τα ουσιαστικά ελέγχοντας κάθε χαρακτηριστικό του μοντέλου με τη βοήθεια του λεξικού. Αποτελεί εντυπωσιακό χαρακτηριστικό γεγονός το ότι η τρίτη κλιτική τάξη των θηλυκών (*ψυχή, καρδιά, μαμά*) είναι η πολυπληθέστερη (26,56%) με την τέταρτη κλιτική τάξη (*πόλη*) να είναι η μικρότερη (1,63%) με ελάχιστες πιθανότητες αύξησης μέσω της παραγωγής και σύνθεσης. Από την άλλη το ένα τρίτο των ουσιαστικών δεν εμφανίζει αλλομορφία (34,55%), η αλλομορφία τύπου Α-Διαγραφή (AC2) εμφάνισε υψηλό ποσοστό (24,31%) και όπως ήταν αναμενόμενο η αντίστοιχη αλλομορφική τάξη (AC7, *πολη~ πολε~ πολ*) της τέταρτης κλιτικής τάξης ήταν η χαμηλότερη (1,63%).

<i>Κλιτικές Τάξεις</i>	<i>Ποσοστό</i>	<i>Αλλομορφικές Τάξεις</i>	<i>Ποσοστό</i>
<b>KT1</b>	17,45%	<b>AC1</b>	34,55%
<b>KT2</b>	13,73%	<b>AC2</b>	24,31%
<b>KT3</b>	26,56%	<b>AC3</b>	5,17%
<b>KT4</b>	1,63%	<b>AC4</b>	20,24%
<b>KT5</b>	15,31%	<b>AC5</b>	5,07%
<b>KT6</b>	20,24%	<b>AC6</b>	3,31%
<b>KT7</b>	1,77%	<b>AC7</b>	1,63%
<b>KT8</b>	3,31%	<b>AC8</b>	1,72%
	<b>100,00%</b>		<b>100,00%</b>

Πίνακας 17: Στατιστικά στοιχεία από το corpus του πειράματος του AMIS.

<sup>11</sup> Για το σύνθετο λεξικό: <http://www.komvos.edu.gr/dictionaries/dictadv/DictAdvSeaS.htm>.



Γράφημα 5: Γραφική αναπαράσταση των στατιστικών για τις κλιτικές και αλλομορφικές τάξεις του σώματος εκπαίδευσης για το πείραμα AMIS.

Η στατιστική ανάλυση των δεδομένων παρουσιάζει μία σημαντική διαφοροποίηση μεταξύ των τάξεων, καθότι υπάρχουν τρεις τάξεις (AC1, AC2, AC4) με υψηλά ποσοστά και οι υπόλοιπες πέντε με εξαιρετικά χαμηλά ποσοστά (όλα κάτω του 6%). Εξακολουθούν, όμως, να παρέχουν τις δέουσες πληροφορίες, για να μπορέσει το μοντέλο να αποδώσει κάποια σημαντικά βάρη για την υλοποίηση του μοντέλου εντοπισμού των αλλομορφικών τύπων. Εδώ να σημειωθεί αυτό που τονίζουν για την ιταλική οι Pirrelli & Battista (2000a, 2000b) ότι στις πολυπληθείς κλιτικές ομάδες υπάρχει λιγότερη κανονικότητα από ότι στις ολιγοπληθείς και έτσι θεωρητικά εντοπίζονται ευκολότερα μοτίβα στις μικρές κλιτικές ομάδες. Και στα Ελληνικά μπορούμε να ισχυριστούμε το ίδιο, καθότι στις ρηματικές κλιτικές τάξεις των πρώην

συντηρημένων ρημάτων είναι ευκολότερη η εξαγωγή μοτίβων από ότι στα υπόλοιπα ρήματα· αντίστοιχα στην όγδοη κλιτική τάξη των ουσιαστικών, που έχει πληθώρα ονομάτων λήγοντα σε –μα, εντοπίζονται κάποιες κανονικότητες.

```

αγώνος 2AC 2IC masculine 3syl penultimate α-γω-νος non_animacy non_calique greek
αγώνος 2AC 2IC masculine 3syl penultimate α-γω-νος non_animacy non_calique greek
αβένος 2AC 2IC masculine 3syl penultimate α-β-νος non_animacy non_calique greek
αβένος 2AC 2IC masculine 3syl penultimate α-β-νος non_animacy non_calique greek
αιθέρος 2AC 2IC masculine 3syl penultimate αι-θ-ρος non_animacy non_calique greek
αιώνος 2AC 2IC masculine 3syl penultimate αι-ω-νος non_animacy non_calique greek
αλαζόνος 2AC 2IC masculine 4syl penultimate α-λα-ζο-νος animacy non_calique greek
αλάτρου 2AC 2IC masculine 3syl penultimate αλ-τη-ρος non_animacy non_calique greek
ανδρόντος 2AC 2IC masculine 4syl penultimate αν-δρ-ι-ο-ντος non_animacy non_calique greek
ανδρός 2AC 2IC masculine 3syl penultimate αν-δρ-ος non_animacy non_calique greek
ανθώνος 2AC 2IC masculine 3syl penultimate αν-θ-νος non_animacy non_calique greek
αίνος 2AC 2IC masculine 3syl antipenultimate α-ι-ο-νος non_animacy non_calique greek
ασπερίδος 2AC 2IC masculine 5syl penultimate α-σπ-ρι-ω-νος animacy non_calique greek
ασπρέδος 2AC 2IC masculine 4syl penultimate α-σπ-ρ-ω-νος non_animacy non_calique greek
αστρός 2AC 2IC masculine 3syl penultimate α-στ-ρος non_animacy non_calique greek
αστίδος 2AC 2IC masculine 3syl penultimate α-τι-δος non_animacy non_calique greek
α5ένος
0 IC_second1 Genre_masculine1 Syllables_three1 Stress_penultimate1 Syllable1_o1 Syllable2_εr1 Syllable3_νq1 Animacy_no1 Calique_no1 Origin_greek1 Character1_o1 Character2_v11 Character3_e
1 IC_second2 Genre_masculine2 Syllables_three2 Stress_penultimate2 Syllable1_o2 Syllable2_εr2 Syllable3_νq2 Animacy_no2 Calique_no2 Origin_greek2 Character1_o2 Character2_v12 Character3_e
0 IC_second3 Genre_masculine3 Syllables_three3 Stress_penultimate3 Syllable1_o3 Syllable2_εr3 Syllable3_νq3 Animacy_no3 Calique_no3 Origin_greek3 Character1_o3 Character2_v13 Character3_e
0 IC_second4 Genre_masculine4 Syllables_three4 Stress_penultimate4 Syllable1_o4 Syllable2_εr4 Syllable3_νq4 Animacy_no4 Calique_no4 Origin_greek4 Character1_o4 Character2_v14 Character3_e
0 IC_second5 Genre_masculine5 Syllables_three5 Stress_penultimate5 Syllable1_o5 Syllable2_εr5 Syllable3_νq5 Animacy_no5 Calique_no5 Origin_greek5 Character1_o5 Character2_v15 Character3_e
0 IC_second6 Genre_masculine6 Syllables_three6 Stress_penultimate6 Syllable1_o6 Syllable2_εr6 Syllable3_νq6 Animacy_no6 Calique_no6 Origin_greek6 Character1_o6 Character2_v16 Character3_e
0 IC_second7 Genre_masculine7 Syllables_three7 Stress_penultimate7 Syllable1_o7 Syllable2_εr7 Syllable3_νq7 Animacy_no7 Calique_no7 Origin_greek7 Character1_o7 Character2_v17 Character3_e
0 IC_second8 Genre_masculine8 Syllables_three8 Stress_penultimate8 Syllable1_o8 Syllable2_εr8 Syllable3_νq8 Animacy_no8 Calique_no8 Origin_greek8 Character1_o8 Character2_v18 Character3_e
α6ένος
0 IC_second1 Genre_masculine1 Syllables_three1 Stress_penultimate1 Syllable1_o1 Syllable2_εr1 Syllable3_νq1 Animacy_no1 Calique_no1 Origin_greek1 Character1_o1 Character2_p11 Character3_e
1 IC_second2 Genre_masculine2 Syllables_three2 Stress_penultimate2 Syllable1_o2 Syllable2_εr2 Syllable3_νq2 Animacy_no2 Calique_no2 Origin_greek2 Character1_o2 Character2_p12 Character3_e
3 IC_second3 Genre_masculine3 Syllables_three3 Stress_penultimate3 Syllable1_o3 Syllable2_εr3 Syllable3_νq3 Animacy_no3 Calique_no3 Origin_greek3 Character1_o3 Character2_p13 Character3_e
0 IC_second4 Genre_masculine4 Syllables_three4 Stress_penultimate4 Syllable1_o4 Syllable2_εr4 Syllable3_νq4 Animacy_no4 Calique_no4 Origin_greek4 Character1_o4 Character2_p14 Character3_e
0 IC_second5 Genre_masculine5 Syllables_three5 Stress_penultimate5 Syllable1_o5 Syllable2_εr5 Syllable3_νq5 Animacy_no5 Calique_no5 Origin_greek5 Character1_o5 Character2_p15 Character3_e
0 IC_second6 Genre_masculine6 Syllables_three6 Stress_penultimate6 Syllable1_o6 Syllable2_εr6 Syllable3_νq6 Animacy_no6 Calique_no6 Origin_greek6 Character1_o6 Character2_p16 Character3_e
0 IC_second7 Genre_masculine7 Syllables_three7 Stress_penultimate7 Syllable1_o7 Syllable2_εr7 Syllable3_νq7 Animacy_no7 Calique_no7 Origin_greek7 Character1_o7 Character2_p17 Character3_e
0 IC_second8 Genre_masculine8 Syllables_three8 Stress_penultimate8 Syllable1_o8 Syllable2_εr8 Syllable3_νq8 Animacy_no8 Calique_no8 Origin_greek8 Character1_o8 Character2_p18 Character3_e

```

Εικόνα 6: Η τροποποίηση του αρχείου event από απλό txt (πάνω μισό) στην κατάλληλη μορφή μέσω της απαραίτητης κωδικοποίησης με το MatLab (κάτω μισό) (<http://www.mathworks.com/products/matlab/>).

Είναι αναμενόμενο ότι ένα μοντέλο με πάνω από 20.000 χαρακτηριστικά για απόδοση βάρους είναι αρκετά βαρύ στατιστικά, μίας και η συνδυαστικότητα των συλλαβών και χαρακτήρων αύξησε εκθετικά το μέγεθος του αρχείου amis.model. Εμφανίστηκαν διάφορες συλλαβές που έπιασαν πολύ υψηλά βάρη, όπως λέξεις που έχουν λήγουσα συλλαβή σε –λης έχουν βάρος 7,25E+02 για την AC5, λέξεις με παραλήγουσα συλλαβή τα– έχουν βάρος 6,51E+02 για την AC4. Παρομοίως, χαρακτηριστικά με λιγότερο εύρος, όπως αυτό της εμψυχότητας ή του τόνου έχουν την αντίστροφη επιρροή σε κάποιες κατηγορίες, όπως το χαρακτηριστικό *non-animacy* με βάρος 2,36E-02 για την AC8 ή τα χαρακτηριστικά του τόνου *stress\_penultimate* και *stress\_antipenultimate* με βάρη 2,04E-02 και 1,82E-02 αντίστοιχα για την AC5. Είναι ιδιαίτερα ενδιαφέρον πόσο βολικά εμπλέκονται τα βάρη αρνητικής επιρροής για την τελική καταχώρηση της αλλομορφίας, μίας και εδώ ο τόνος φαίνεται να αποτελεί τον οριοθέτη ανάμεσα στην AC5 (τύπου *θεριακλής*) και την AC8 (τύπου *μακακλής*).

AC1	Βάρη θετικής επιρροής	Βάρη αρνητική επιρροής	
Syllable3_ψιος	2,28E+01	Syllable2_ι	2,21E-01
Syllable2_ντιο	2,24E+01	Syllable3_δα	2,13E-01
Syllable2_λιο	1,84E+01	Syllable2_ντι	1,93E-01
Syllable3_γιο	1,61E+01	Origin_turkish	1,36E-01
Syllable1_πλιν	1,60E+01	Syllable4_νι	1,18E-01
Syllable4_νιος	1,49E+01	Syllable2_γη	9,80E-02
Syllable2_νιο	1,41E+01	Genre_feminine	9,03E-02

	Syllable2_ος	1,03E+01	Syllable2_ση	5,04E-02
	Syllable2_βδος	9,45E+00	Syllable3_λη	3,09E-02
	Syllable1_γιος	8,97E+00	Character1_α	6,12E-03
<b>AC2</b>	<b>Βάρη θετικής επιρροής</b>		<b>Βάρη αρνητική επιρροής</b>	
	Syllable3_τζας	5,33E+01	Syllable1_δε	1,38E-01
	Syllable3_τη	1,24E+01	Syllable3_δι	1,29E-01
	Syllable2_χια	1,19E+01	Syllable3_μα	1,20E-01
	Syllable2_μπας	1,01E+01	Character2_μ	1,16E-01
	Syllable1_φιρ	9,93E+00	Syllable3_λο	7,79E-02
	Syllable2_γας	8,13E+00	Syllable3_ρο	7,70E-02
	Syllable2_γκας	7,46E+00	Character2_υ	7,10E-02
	Syllable2_ρια	6,88E+00	Syllable3_ο	6,16E-02
	Syllable1_μνα	5,90E+00	Syllable3_νι	3,67E-02
	Syllable3_βια	5,46E+00	Character1_η	4,37E-03
<b>AC3</b>	<b>Βάρη θετικής επιρροής</b>		<b>Βάρη αρνητική επιρροής</b>	
	Syllable2_τρης	5,41E+02	Character3_θ	1,96E-01
	Syllable3_ρη	1,37E+02	Syllable2_τα	1,78E-01
	Syllable1_μεγ	7,62E+01	Syllable2_να	1,64E-01
	Syllable2_δης	5,31E+01	Syllable3_δα	1,24E-01
	Syllable3_πης	4,28E+01	Character2_σ	1,09E-01
	Syllable2_ντης	3,54E+01	Stress_antipenultimate	9,18E-02
	Syllable3_δης	2,87E+01	Syllable2_μα	7,96E-02
	Syllable3_φης	2,58E+01	Syllable3_τα	7,05E-02
	Syllable4_χη	2,19E+01	Origin_italian	2,43E-02
	Syllable1_ζη	2,05E+01	Origin_turkish	1,98E-02
<b>AC4</b>	<b>Βάρη θετικής επιρροής</b>		<b>Βάρη αρνητική επιρροής</b>	
	Syllable4_τα	6,51E+02	Character2_α	2,01E-01
	Syllable4_λα	2,11E+01	Character2_ο	1,82E-01
	Syllable4_ρα	1,85E+01	Syllable3_κο	1,67E-01
	Syllable3_ρισ	1,17E+01	Syllable3_λο	1,67E-01
	Syllable2_χτυ	1,16E+01	Syllable1_ε	1,57E-01
	Syllable2_γκρι	1,16E+01	Syllables_five	1,29E-01
	Syllable3_λε	1,11E+01	Character1_θ	1,22E-01
	Syllable1_σπε	1,06E+01	Syllable1_ι	1,02E-01
	Syllable4_ι	8,95E+00	Character2_υ	2,25E-02
	Syllable5_ρι	8,10E+00	Syllable3_ο	1,09E-02
<b>AC5</b>	<b>Βάρη θετικής επιρροής</b>		<b>Βάρη αρνητική επιρροής</b>	
	Syllable4_λης	7,27E+02	Syllable2_πα	3,78E-01
	Syllable2_δες	1,02E+02	Syllable1_α	3,74E-01
	Syllable1_γιορ	5,62E+01	Syllable2_πι	3,66E-01
	Syllable3_νης	4,24E+01	Calque_no	3,45E-01
	Syllable3_πης	4,11E+01	Origin_greek	3,44E-01
	Syllable3_για	3,69E+01	Syllable1_λα	3,42E-01
	Syllable3_κης	2,74E+01	Character2_υ	2,23E-01
	Syllable2_ριας	2,68E+01	Syllable3_κας	1,40E-01
	Syllable2_σπο	2,00E+01	Stress_antipenultimate	4,99E-02
	Syllable4_γας	1,89E+01	Stress_penultimate	8,16E-02
<b>AC6</b>	<b>Βάρη θετικής επιρροής</b>		<b>Βάρη αρνητική επιρροής</b>	
	Syllable3_ον	7,53E+01	Syllable2_λι	2,86E-01
	Syllable2_λας	7,52E+01	Character3_ι	2,78E-01
	Syllable2_ρον	7,23E+01	Character4_ο	2,51E-01
	Syllable2_ρας	3,51E+01	Character3_ο	2,16E-01
	Syllable2_ζον	3,14E+01	Character4_β	2,04E-01
	Syllable2_ον	3,08E+01	Stress_ultimate	1,97E-01
	Syllable3_κον	2,88E+01	Syllable2_α	1,89E-01
	Syllable1_ον	1,87E+01	Character4_μ	1,45E-01
	Syllable3_νον	1,84E+01	Character3_ν	1,33E-01
	Syllable3_χον	1,79E+01	Animacy_yes	1,12E-01



AC7		Βάρη θετικής επιρροής		Βάρη αρνητική επιρροής	
	Syllable1_τιγ	6,23E+01	Character3_ι		3,48E-01
	Syllable2_βρη	2,23E+01	Character4_ο		3,33E-01
	Syllable2_στη	1,91E+01	Character4_λ		3,30E-01
	Syllable3_σαν	1,86E+01	Character4_δ		3,28E-01
	Syllable4_δρη	1,86E+01	Character4_ζ		2,89E-01
	Syllable3_βδη	1,75E+01	Character2_δ		2,74E-01
	Syllable2_ψη	9,29E+00	Character2_λ		2,07E-01
	Syllable2_ξη	8,38E+00	Character4_α		1,67E-01
	Syllable2_ναρ	8,15E+00	Animacy_yes		1,41E-01
	Syllable1_τερ	8,10E+00	Calque_yes		7,54E-02
AC8		Βάρη θετικής επιρροής		Βάρη αρνητική επιρροής	
	Syllable2_λης	4,59E+01	Syllable2_ρα		3,00E-01
	Syllable2_ης	4,54E+01	Character4_ν		2,96E-01
	Syllable2_γης	4,12E+01	Origin_italian		2,93E-01
	Syllable2_ρης	2,98E+01	Character3_σ		2,34E-01
	Syllable2_ντζε	2,79E+01	Character4_ρ		2,03E-01
	Syllable1_ερ	2,63E+01	Character4_λ		1,84E-01
	Syllable2_τερ	2,35E+01	Calque_no		1,37E-01
	Syllable1_μπου	1,70E+01	Origin_greek		1,36E-01
	Syllable2_της	1,69E+01	Character4_α		1,10E-01
	Syllable1_γιαν	1,54E+01	Animacy_no		2,36E-02

Πίνακας 18: Συγκεντρωτικός πίνακας των 10 υψηλότερων και χαμηλότερων θετικών και αρνητικών βαρών για κάθε αλλομορφική τάξη.

Ένας επιπλέον τρόπος δημιουργίας ενός μοντέλου πρόβλεψης της αλλομορφίας είναι αναμφίβολα και οι αρνητικές μεταβλητές κατανομών βαρών που δίνουν πολύτιμες πληροφορίες σχετικά με συγκεκριμένες αλλομορφικές τάξεις, ελέγχοντας με προσοχή τις κατανομές συγκριτικά. Οι δέκα υψηλότερες κατανομές βαρών θετικής και αρνητικής επιρροής παρουσιάζονται στον πίνακα (18.), όπου τα γκρι κελιά στα χαρακτηριστικά υποδηλώνουν εκείνα με πρωταρχικές επιλογές, όπως αριθμό συλλαβών, τόνο, κλιτική τάξη κτλ., ενώ αντίστοιχα τα γκρι κελιά στις στήλες των τιμών είναι για τα εξαιρετικά υψηλά ή χαμηλά βάρη. Σε όλες τις αλλομορφικές τάξεις με λιγιστά μέλη τα βάρη αρνητικής επιρροής δεν περιέχουν αποκλειστικά διάφορες συλλαβές ή και χαρακτήρες, αλλά σημαντική επιρροή φαίνεται να ασκούν για την αποτροπή τους κύρια και βασικά χαρακτηριστικά.

Ας προσπαθήσουμε τώρα να δούμε συνδυαστικά χαρακτηριστικά που μπορούν να βοηθήσουν στην πρόβλεψη κάποιων αλλομορφικών τάξεων. Για τα ουσιαστικά που δεν εμφανίζουν αλλομορφία εκτός διαφόρων συλλαβών που αποτελούν μερική θετική επιρροή για τη συγκεκριμένη κατηγορία (AC1), παρατηρούμε ότι το χαρακτηριστικό *IC\_seventh* με βάρος 2,53E+00, καθώς και τα *IC\_fitst* με 2,57E+00 και *IC\_fifth* με 2,91E+00 συντελούν στον εντοπισμό της αλλομορφίας. Παράλληλα οι λέξεις που φέρουν το χαρακτηριστικό *origin\_turkish* (1,36E-01) και *genre\_feminine* (9,09E-02) αποτρέπουν το σύστημα να συμπεράνει ότι δεν θα εμφανίσουν αλλομορφία. Δυστυχώς

για τη δεύτερη, τέταρτη και πέμπτη αλλομορφική τάξη φαίνεται πως κυρίαρχο ρόλο παίζουν οι συλλαβές και οι χαρακτήρες με ελάχιστες παρεμβολές είτε θετικής (*stress\_penultimate* 1,44E+00) είτε αρνητικής επιρροής (*stress\_antipenultimate* 7,91E-01) από τα βασικά χαρακτηριστικά. Σαφώς καλύτερη είναι η πρόβλεψη για την τρίτη τάξη αλλομορφίας, όπου αποτρέπουν σημαντικά τα *origin\_italian* (2,43E-02), *origin\_turkish* (1,98E-02) και *stress\_antipenultimate* (9,19E-02).

Ειδικότερα στις τρεις τελευταίες αλλομορφικές τάξεις η εμφάνιση πολλών βασικών χαρακτηριστικών στις αρνητικές κατανομές βαρών και ειδικά στις χαμηλότερες θέσεις είναι αξιοσημείωτη. Να σημειωθεί ότι είναι οι μοναδικές τάξεις που η αλλομορφία τους αντιστοιχεί μόνο σε μία κλιτική τάξη που είναι η τέταρτη, η έκτη και η όγδοη. Για την AC6 ο αριθμός των συλλαβών διαδραματίζει σημαντικό ρόλο, καθώς τα *syllables\_five* (3,50E-01), *syllables\_three* (3,13E-01) και *syllables\_four* (4,22E-01) έχουν αρνητική επιρροή παράλληλα με τον τόνο στη λήγουσα (*stress\_ultimate* 1,48E-01). Ακόμα περισσότερα είναι τα αντίστοιχα βάρη αρνητικής επιρροής στην έβδομη αλλομορφική τάξη με τα κατάλοιπα της αρχαίας Ελληνικής να επηρεάζουν σημαντικά τις αποφάσεις, καθότι αποτρεπτικά λειτουργούν τα χαρακτηριστικά *calque\_yes* (7,54E-02), *animacy\_yes* (1,41E-01) και *syllables\_four* (3,74E-01) μεταξύ άλλων· αντίθετα το χαρακτηριστικό *IC\_fourth* (6,94E+01) συντελεί στη σωστή επιλογή, καθότι είναι και μικρή κατηγορία, αλλά και με αποκλειστική αλλομορφία για τα θηλυκά ουσιαστικά σε αντίθεση με την τρίτη. Τέλος, για την όγδοη κλιτική τάξη είναι αξιοσημείωτη η ύπαρξη συλλαβών τύπου Χης με μεγάλο θετικό βάρος (από 2,98E+02 έως 4,59E+01), ενώ αντίθετα τα χαρακτηριστικά *animacy\_no* (2,36E+02), *calque\_no* (1,37E+01), *origin\_greek* (1,36E-01) και *origin\_italian* (2,93E-01) φιλτράρουν αρνητικά τα ουσιαστικά κατά την καταχώρηση στη συγκεκριμένη τάξη που έχει πολλές ομοιότητες με την αντίστοιχη πέμπτη (ανήκουν στην ίδια κλιτική τάξη και σχηματίζουν αλλόμορφα τύπου-δ).

#### 5.2.4.1. Αξιολόγηση επιρροής βασικών χαρακτηριστικών

Με σχετική υπεροχή και από σημαντική – είτε θετική είτε αρνητική – θέση τα βασικά χαρακτηριστικά του μοντέλου μπορούν να επηρεάσουν και να παρέμβουν στην τελική επιλογή λειτουργώντας συνδυαστικά με τα ατομικά βάρη των υπόλοιπων χαρακτηριστικών. Παρατηρούμε ότι στις ιδιαίτερες αλλομορφικές τάξεις υπάρχει μία διαβάθμιση των «περιορισμών», δηλαδή εκείνων των χαρακτηριστικών που

συμβάλλουν αποτρεπτικά για την επιλογή της συγκεκριμένης αλλομορφικής τάξης για την οποιαδήποτε εξεταζόμενη ή αναλυόμενη λέξη.

AC1	Βάρη θετικής επιρροής		Βάρη αρνητικής επιρροής	
	IC_fifth	2,91E+00	Stress_penultimate	9,64E-01
	Origin_german	2,68E+00	Origin_albanian	7,95E-01
	IC_first	2,57E+00	Origin_arabic	6,98E-01
	IC_seventh	2,53E+00	Origin_slavic	3,44E-01
	Syllables_six	2,46E+00	Origin_turkish	1,36E-01
	Origin_spanish	2,16E+00	Genre_feminine	9,03E-02
	Stress_antipenultimate	2,10E+00		
	Origin_russian	2,02E+00		
	Syllables_five	1,82E+00		
	Origin_hebrew	1,76E+00		
	Syllables_four	1,74E+00		
	Genre_masculine	1,70E+00		
	Origin_italian	1,69E+00		
	Origin_bulgarian	1,58E+00		
	Syllables_two	1,39E+00		
	Genre_neutral	1,37E+00		
	Origin_english	1,33E+00		
	Origin_french	1,30E+00		
	Animacy_no	1,30E+00		
	Calque_yes	1,28E+00		
	Calque_no	1,27E+00		
	Origin_greek	1,27E+00		
	Animacy_yes	1,15E+00		
	Stress_ultimate	1,14E+00		
	Syllables_one	1,02E+00		
	Syllables_three	1,01E+00		
AC2	Βάρη θετικής επιρροής		Βάρη αρνητικής επιρροής	
	Origin_arabic	3,58E+00	Animacy_no	9,88E-01
	Origin_latin	2,86E+00	Stress_antipenultimate	9,84E-01
	Syllables_one	2,64E+00	Origin_greek	9,57E-01
	Origin_spanish	2,20E+00	Calque_no	9,56E-01
	Origin_polish	2,03E+00	Syllables_two	8,60E-01
	Origin_japanese	2,01E+00	Syllables_five	8,20E-01
	IC_second	1,99E+00	Origin_turkish	4,68E-01
	Origin_bulgarian	1,98E+00	Stress_ultimate	3,53E-01
	Origin_romanian	1,96E+00		
	IC_third	1,94E+00		
	Genre_feminine	1,77E+00		
	Origin_english	1,73E+00		
	Origin_albanian	1,67E+00		
	Origin_german	1,58E+00		
	Origin_french	1,23E+00		
	Calque_yes	1,14E+00		
	Stress_penultimate	1,14E+00		
	Origin_italian	1,11E+00		
	Syllables_three	1,11E+00		
	Animacy_yes	1,11E+00		
	Origin_hebrew	1,10E+00		
	Origin_slavic	1,08E+00		
	Genre_masculine	1,05E+00		
	Syllables_four	1,03E+00		
	Origin_russian	1,02E+00		
AC3	Βάρη θετικής επιρροής		Βάρη αρνητικής επιρροής	

	IC_second	1,96E+00	Syllables_four	9,21E-01
	Origin_albanian	1,54E+00	Origin_greek	8,66E-01
	IC_third	1,53E+00	Calque_no	8,64E-01
	Syllables_six	1,51E+00	Animacy_no	8,21E-01
	Syllables_five	1,48E+00	Syllables_three	8,19E-01
	Origin_french	1,41E+00	Genre_masculine	8,00E-01
	Genre_feminine	1,29E+00	Animacy_yes	7,74E-01
	Stress_ultimate	1,28E+00	Syllables_two	7,53E-01
	Syllables_one	1,05E+00	Stress_penultimate	7,33E-01
			Calque_yes	6,13E-01
			Stress_antipenultimate	9,18E-02
			Origin_italian	2,43E-02
			Origin_turkish	1,98E-02
<b>AC4</b>	<b>Βάρη θετικής επιρροής</b>		<b>Βάρη αρνητικής επιρροής</b>	
	Origin_persian	4,52E+00	Syllables_two	9,78E-01
	IC_sixth	2,95E+00	Origin_french	8,43E-01
	Genre_neutral	1,74E+00	Origin_arabic	8,24E-01
	Origin_slavic	1,63E+00	Origin_russian	8,17E-01
	Stress_penultimate	1,40E+00	Origin_italian	8,11E-01
	Stress_ultimate	1,26E+00	Origin_english	8,09E-01
	Syllables_three	1,26E+00	Animacy_yes	6,22E-01
	Origin_turkish	1,25E+00	Syllables_four	5,58E-01
	Animacy_no	1,15E+00	IC_third	4,02E-01
	Calque_yes	1,07E+00	Syllables_five	1,29E-01
	Origin_greek	1,06E+00	Genre_feminine	8,02E-03
	Calque_no	1,05E+00		
<b>AC5</b>	<b>Βάρη θετικής επιρροής</b>		<b>Βάρη αρνητικής επιρροής</b>	
	Origin_slavic	7,78E+00	Origin_french	9,73E-01
	IC_second	2,42E+00	Syllables_three	9,67E-01
	Origin_hebrew	2,19E+00	Origin_english	7,99E-01
	Animacy_yes	1,96E+00	Animacy_no	6,68E-01
	Stress_ultimate	1,84E+00	IC_third	5,73E-01
	Genre_masculine	1,78E+00	Syllables_four	4,68E-01
	Origin_turkish	1,55E+00	Origin_italian	4,49E-01
	Origin_arabic	1,46E+00	Genre_feminine	4,47E-01
	Calque_yes	1,30E+00	Calque_no	3,45E-01
	Syllables_two	1,14E+00	Origin_greek	3,44E-01
			Stress_antipenultimate	2,04E-02
			Stress_penultimate	1,82E-02
<b>AC6</b>	<b>Βάρη θετικής επιρροής</b>		<b>Βάρη αρνητικής επιρροής</b>	
	IC_eighth	5,39E+00	Calque_no	9,66E-01
	Syllables_one	4,13E+00	Origin_greek	9,64E-01
	Genre_neutral	2,25E+00	Stress_antipenultimate	7,91E-01
			Syllables_two	6,51E-01
			Animacy_no	5,47E-01
			Stress_penultimate	4,75E-01
			Syllables_four	4,22E-01
			Syllables_five	3,50E-01
			Syllables_three	3,13E-01
			Stress_ultimate	1,97E-01
			Animacy_yes	1,12E-01
<b>AC7</b>	<b>Βάρη θετικής επιρροής</b>		<b>Βάρη αρνητικής επιρροής</b>	
	IC_fourth	6,94E+00	Origin_french	6,68E-01
	Stress_antipenultimate	1,33E+00	Stress_penultimate	5,28E-01
	Syllables_five	1,25E+00	Calque_no	5,12E-01
	Genre_feminine	1,18E+00	Origin_greek	5,12E-01
			Syllables_two	4,96E-01
			Animacy_no	4,83E-01

		Origin_italian	4,30E-01	
		Syllables_three	3,80E-01	
		Syllables_four	3,74E-01	
		Animacy_yes	1,41E-01	
		Calque_yes	7,54E-02	
<b>AC8</b>	<b>Βάρη θετικής επιρροής</b>	<b>Βάρη αρνητικής επιρροής</b>		
	IC_second	2,10E+00	Calque_yes	8,29E-01
	Origin_turkish	1,68E+00	Origin_arabic	7,73E-01
	Genre_masculine	1,57E+00	Syllables_three	7,30E-01
	Stress_penultimate	1,53E+00	Syllables_four	6,42E-01
	Origin_albanian	1,17E+00	Syllables_two	4,68E-01
	Animacy_yes	1,06E+00	Origin_italian	2,93E-01
			Calque_no	1,37E-01
			Origin_greek	1,36E-01
			Animacy_no	2,36E-02

Πίνακας 19: Συγκεντρωτικός πίνακας με τα βασικά χαρακτηριστικά και τα βάρη του ανάλογα με την αλλομορφική τάξη.

Ο Πίνακας 19 παρουσιάζει τα βάρη θετικής και αρνητικής επιρροής των βασικών χαρακτηριστικών του μοντέλου για κάθε αλλομορφική τάξη ταξινομημένα από την υψηλότερη στη χαμηλότερη. Να σημειωθεί ότι κάποια βασικά χαρακτηριστικά απουσιάζουν από τον παραπάνω πίνακα, μίας και τα βάρη τους είναι 1.0 και επομένως δεν συμβάλλουν καθόλου στο μοντέλο της αλλομορφίας για την τελική του επιλογή. Φαίνεται αξιοσημείωτο το γεγονός ότι οι δύο πρώτες αλλομορφικές τάξεις που έχουν και τα περισσότερα μέλη εμφανίζουν τα περισσότερα βασικά χαρακτηριστικά με βάρη θετικής επιρροής, ενώ όλες οι άλλες πέντε αλλομορφικές τάξεις κατακλύζονται από βασικά χαρακτηριστικά με βάρη αρνητικής επιρροής. Εντούτοις διαθέτουν πολύ συγκεκριμένα χαρακτηριστικά με βάρη θετικής επιρροής, όπως η όγδοη αλλομορφική τάξη (τύπου *μπακαλη~μπακάληδ~μπακαλ*), με τα στοιχεία της δεύτερης κλιτικής τάξης (IC\_second), του παραλήγοντα τόνου (Stress\_penultimate) και την τουρκικής καταγωγής (Origin\_turkish).

Είναι εμφανές ότι κάποια χαρακτηριστικά επιδρούν δυναμικά επηρεάζοντας το μοντέλο προς επιλογή ή προς αποφυγή της καθοριστικής αλλομορφικής τάξης. Παράλληλα κάποια άλλα χαρακτηριστικά επιδρούν σε μικρότερο βαθμό για την τελική επιλογή. Είναι θεμελιώδης συνισταμένη για το μοντέλο ότι κάθε βάρος χαρακτηριστικού συμμετέχει στην τελική επιλογή. Σημαντική επισήμανση είναι ότι εξίσου με τις υψηλές θετικές τιμές, τα χαμηλά βάρη αρνητικής επιρροής θα περιορίσουν τις πιθανότητες για αυτήν την επιλογή. Όπως φαίνεται στον πίνακα 18 σε άμεση συνάρτηση με τον πίνακα 16, κάποιες συλλαβικές δομές θα επηρεάσουν σημαντικά την επιλογή, ωστόσο αποτελούν το ένα τέταρτο των βαρών με τη συμβολή των βασικών χαρακτηριστικών να είναι καίρια. Οι διαφορές σε κάποιες κατηγορίες

μπορούν να προκαλούν εντύπωση με την εμφάνιση δύο εκ διαμέτρου αντίθετων τιμών στην ίδια ομάδα (βλ. animacy features στο AC7): ωστόσο πρέπει να αναλογιστούμε τις τιμές πάντα σε συνδυασμό με τις αντίστοιχες των άλλων αλλομορφικών τάξεων.

#### 5.2.4.2. ΑΛΛΟΜΑΝΤΙΣ – Ένας αναλυτής AMIS πρόβλεψης της αλλομορφίας

Το Ελληνικό μοντέλο Μέγιστης Εντροπίας εκπαιδεύτηκε πάνω σε ένα σώμα ουσιαστικών και των οκτώ κλιτικών τάξεων που δεν έχουν παραχθεί από κάποια άλλη λέξη, όπως αναφέραμε στο 5.2.4. Για την αξιολόγηση της αποτελεσματικότητας του μοντέλου δημιουργήθηκε ένα δεύτερο σώμα παράγωγων ουσιαστικών χρησιμοποιώντας όλα τα επιθήματα που δημιουργούν ουσιαστικά μη εξαιρώντας κανένα. Το συγκεκριμένο σώμα περιέχει 2.755 ουσιαστικά προσεχτικά επιλεγμένα, ώστε να καλύπτουν όλο το εύρος των χαρακτηριστικών έχοντας ως βάση το λεξικό του Ιδρύματος Τριανταφυλλίδη: επισημαίνεται πως το χαρακτηριστικό της καταγωγής (*origin*) δόθηκε από τη βάση και όχι από το επίθημα. Δημιουργήθηκε ένα αντίστοιχο αρχείο με εκείνο του σώματος της εκπαίδευσης, αλλά απουσίαζε η πληροφορία για την ορθή αλλομορφική τάξη.<sup>12</sup>

Με το MatLab χτίστηκε ο ΑΛΛΟΜΑΝΤΙΣ<sup>13</sup>, ένας μορφολογικός αναλυτής με στόχο την πρόβλεψη της αλλομορφίας έχοντας για εισαγόμενα τα δεδομένα από οποιοδήποτε πείραμα σε AMIS. Ο αναλυτής έπαιρνε τα χαρακτηριστικά κάθε λέξης και τα αντικαταστάσε με αντίστοιχα βάρη που δόθηκαν από το AMIS για τα χαρακτηριστικά του σώματος εκπαίδευσης. Αν δεν υπήρχε κάποιο χαρακτηριστικό με κάποιο βάρος στο αρχείο του, τότε του δίδεται η τιμή 1,00. Το σύστημα πολλαπλασιάζει τα βάρη των χαρακτηριστικών για κάθε υποψήφια αλλομορφική τάξη και προκρίνει αυτό με το μεγαλύτερο γινόμενο: το τελευταίο σύμφωνα με το μοντέλο της μέγιστης εντροπίας είναι και ο τελικός νικητής στη μάχη των αλλομορφικών τάξεων και γίνεται ταυτοποίηση από τον αναλυτή.

Είναι αξιοσημείωτο ότι το μοντέλο εκπαιδεύτηκε από ένα σώμα με ουσιαστικά που δεν προέκυψαν από τη διαδικασία της παραγωγής (ή της σύνθεσης) και δοκιμάστηκε/αξιολογήθηκε από ένα σώμα με παράγωγα ουσιαστικά. Αυτή η επιλογή δεν έγινε

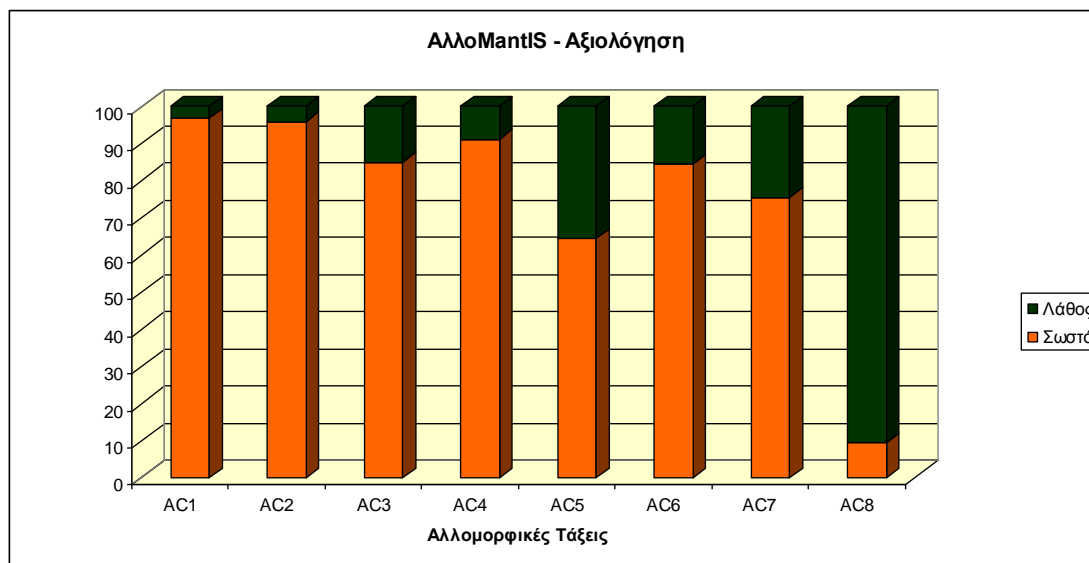
<sup>12</sup> Η πληροφορία για την αλλομορφική τάξη ήταν προσβάσιμη μετά την πρόβλεψη του μοντέλου για τον έλεγχο του ποσοστού ορθότητας της πρόβλεψης.

<sup>13</sup> Ο συμφυρμός ΑΛΛΟΜΑΝΤΙΣ προέκυψε από την αλλομορφία και τον μάντη με λατινικούς χαρακτήρες, ώστε τα κεφαλαία γράμματά του να συμπίπτουν με το πρόγραμμα της Μέγιστης Εντροπίας AMIS.

αυθαίρετα, καθότι ο Καρασίμος (2001, 2009) είχε υποστηρίξει ότι τα παραγωγικά ονοματικά επιθήματα που εμφανίζουν κοινά χαρακτηριστικά με απλά ουσιαστικά, θα ενταχθούν στην ίδια κλιτική ομάδα και επομένως θα εμφανίσουν την ίδια αλλομορφική συμπεριφορά. Ο ΑΛΛΟΜΑΝΤΙΣ προέβλεψε ορθώς την αλλομορφία που εμφανίζει κάποιο παράγωγο ουσιαστικό με ποσοστό 86,4973%. Είναι αναμενόμενο ότι ο ΑΛΛΟΜΑΝΤΙΣ σε περίπτωση που εκπαιδευτεί επιπλέον με σώμα παράγωγων ουσιαστικών, τότε το ποσοστό θα ανέβει σε πολύ υψηλότερα επίπεδα. Κρίθηκε απαραίτητο στο πρωταρχικό στάδιο της δοκιμής του αναλυτή να έχει την ελάχιστη δυνατή βοήθεια από το σώμα εκπαίδευσης, μίας και θα αποκτούσε μία άλλη δυναμική η συστηματική παρουσία συγκεκριμένων ληγουσών συλλαβών στο μοντέλο. Πιο συγκεκριμένα, εφόσον το μοντέλο είχε εκπαιδευτεί με ελάχιστες παράγωγες λέξεις σε *-τζής*, τότε ο ΑΛΛΟΜΑΝΤΙΣ θα είχε σημαντική βοήθεια από την *SyllableX\_τζής* για την πρόβλεψη της αλλομορφίας των νέων λέξεων σε *-τζης* στο σώμα δοκιμής.

Η συνολική ακρίβεια του μοντέλου ήταν 86,49% με το ποσοστό αποτυχίας στο 13,51%. Μία λεπτομερής ανάλυση των αποδόσεων του μοντέλου για κάθε αλλομορφική τάξη παρατίθεται στον ακόλουθο πίνακα και γράφημα. Πάνω από 90% επιτεύχθηκε σε αρκετές τάξεις, όπως στην πρώτη αλλομορφική τάξη 96,92%, στη δεύτερη 95,97% και στην τέταρτη 91,2%, ενώ αντιθέτως οι δύο τάξεις με τα χαμηλότερα ποσοστά ήταν η πέμπτη (64,37%) και η όγδοη (9,28%), με την τελευταία να καταγράφει τραγικά ποσοστά αναλογιζόμενοι τον μέσο όρο επιτυχίας. Είναι αναμενόμενο οι συγκεκριμένες κατηγορίες να καταγράφουν χαμηλά ποσοστά, γιατί έχουν πάμπολλες ομοιότητες μίας και ανήκουν στην ίδια κλιτική τάξη και έχουν το πρώτο τους αλλόμορφο με το *δ*-τύπο. Ενδιαφέρον στοιχείο είναι ότι αναμφίβολα οι μεγαλύτερες αλλομορφικές τάξεις είχαν υψηλότερα ποσοστά επιτυχίας, γεγονός που φανερώνει πως με ένα σώμα εκπαίδευσης με παράγωγα ονόματα, θα ανέβαιναν σε υψηλά σημεία και των υπολοίπων «μικρών» αλλομορφικών τάξεων. Κλείνοντας, είναι σαφές ότι η Μέγιστη Εντροπία είναι σε θέση να επεξεργαστεί βασικές γλωσσολογικές και γλωσσικές ιδιότητες που ενέχονται στο φαινόμενο της αλλομορφίας στην Ελληνική, μπορεί σε μεγάλο ποσοστό να προβλέψει τους αλλομορφικούς τύπους, ενώ παράλληλα υπάρχουν κάποιες μικρές περιφερειακές ομάδες αλλομόρφων όπου φαίνεται να μην τα καταφέρνει με μεγάλη επιτυχία. Αυτές οι ομάδες είναι κυρίως τα ουσιαστικά με τρία αλλόμορφα, όπως *γονέας* (*γονεα~γον~ γονε*) και κυρίως τα ουσιαστικά τύπου *μπακάλης* (*μπακαλη~ μπακαληδ~ μπακαλ*) που εμφανίζουν

διαφορετικό αλλόμορφο στην παραγωγή και ομοιάζουν πολύ με τα υπόλοιπα ουσιαστικά που εμφανίζουν τα αλλόμορφα τύπου-δ (παπαδ, καφεδ, νταηδ).

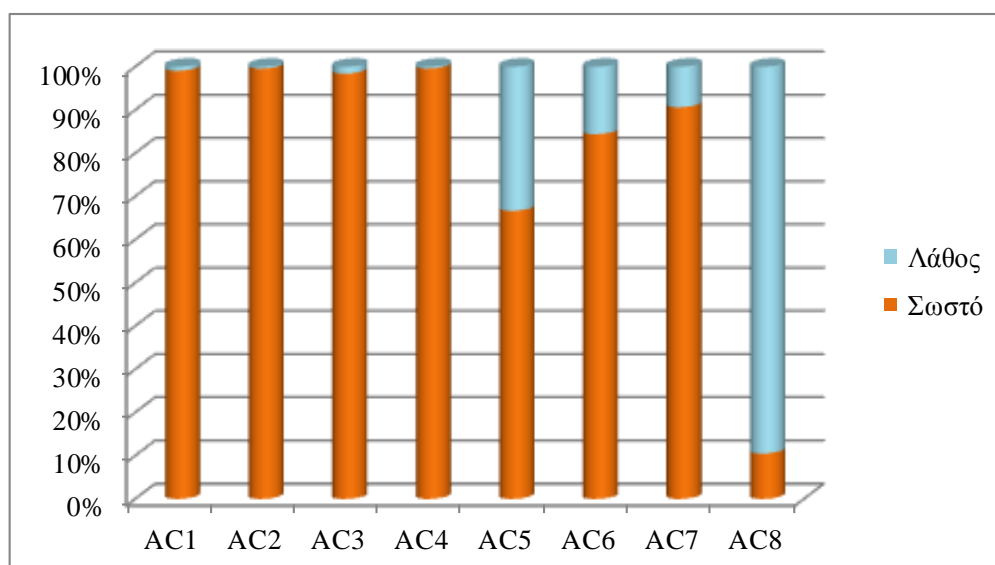


	Σωστό	Λάθος
Αλλομορφική Τάξη AC1	96,92	3,08
Αλλομορφική Τάξη AC2	95,97	4,03
Αλλομορφική Τάξη AC3	85,04	14,96
Αλλομορφική Τάξη AC4	91,2	8,8
Αλλομορφική Τάξη AC5	64,37	35,63
Αλλομορφική Τάξη AC6	84,43	15,57
Αλλομορφική Τάξη AC7	75,53	24,47
Αλλομορφική Τάξη AC8	9,28	90,72
<b>Σύνολο</b>	<b>86,5</b>	<b>14,5</b>

Γράφημα 6: Η λεπτομερής καταγραφή πρόβλεψης του ΑΛΛΟΜΑΝΤΙΣ για κάθε αλλομορφική τάξη.

Για τη βελτίωση του συστήματος, δοκιμάσαμε μία πιο ορθολογική προσέγγιση, ώστε να επιτύχουμε καλύτερη επίδοση. Στην προηγούμενη μορφή του ο ΑλλοMantIS αριθμούσε τις συλλαβές από τα αριστερά προς τα δεξιά (δηλ. γο-syl1, να-syl2, το-syl3), ενώ στην ανανεωμένη έκδοση ο συλλαβισμός ακολούθησε την τακτική του τόνου, δηλαδή η λήγουσα αποτελούσε την πρώτη συλλαβή, τη δεύτερη αποτελούσε η παραλήγουσα και ούτω καθ' εξής. Έτσι οι προθηματοποιημένες και οι σύνθετες λέξεις θα έχουν πάντα τις ίδιες αρχικές συλλαβές, δηλαδή οι λέξεις κλήση και παράκληση θα έχουν syl1\_ση και syl2\_κλη. Το αποτέλεσμα της αναβαθμισμένης έκδοσης του ΑΛΛΟΜΑΝΤΙΣ ήταν το ποσοστό πρόβλεψης να ανέβει στο 91,43% με τις αποτυχημένες περιπτώσεις στο 8,47%. Μάλιστα οι τέσσερις πρώτες αλλομορφικές τάξεις άγγιξαν το 100% με την όγδοη κλιτική τάξη να παραμένει σε τραγικά χαμηλά επίπεδα (10,53%) έχοντας 52,43% των λανθασμένων εκτιμήσεων στην πέμπτη αλλομορφική τάξη.





	Σωστό	Λάθος
Αλλομορφική Τάξη 1	99,09	0,91
Αλλομορφική Τάξη 2	99,59	0,41
Αλλομορφική Τάξη 3	98,43	1,57
Αλλομορφική Τάξη 4	99,6	0,4
Αλλομορφική Τάξη 5	66,67	33,33
Αλλομορφική Τάξη 6	84,42	15,58
Αλλομορφική Τάξη 7	90,64	9,36
Αλλομορφική Τάξη 8	10,53	89,47
Σύνολο	91,43	8,57

Γράφημα 7: Η λεπτομερής καταγραφή της δεύτερης πρόβλεψης του *ΑΛΛΟΜΑΝΤΙΣ* για κάθε αλλομορφική τάξη.

### 5.2.5. Συμπερασματικά σχόλια

Τα πιθανολογικά γλωσσικά μοντέλα, οι αλγόριθμοι εποπτευόμενης μηχανικής μάθησης και σύγχρονα γλωσσολογικά θεωρητικά μοντέλα, όλα εμφανίζονται να υποστηρίζουν την άποψη της γλωσσικής επεξεργασίας σύμφωνα με τους Dell' Orletta *et al.* (2007: 53-54) ότι είναι αποτέλεσμα μίας δυναμικής και on-line ανάλυσης συγκρουόμενων γραμματικών περιορισμών. Επεκτείνοντας αυτόν το συλλογισμό φαίνεται πως συγκεκριμένα μορφολογικά φαινόμενα ή διαδικασίες είναι αποτέλεσμα μίας συνδυαστικής ανάλυσης μορφολογικών χαρακτηριστικών με τη συνεπικουρία (κάποιες φορές) στοιχείων από άλλους τομείς της γλώσσας (φωνολογία, σημασιολογία, κτλ). Βασιζόμενοι στις θεωρητικές παρατηρήσεις της αλλομορφικής συμπεριφοράς εισάγαμε διάφορες γλωσσολογικές πληροφορίες σε ένα μοντέλο υπολογιστικής επεξεργασίας της αλλομορφίας. Η προσέγγιση μας επιτρέπει να ερευνησουμε πιθανά

σχήματα ανάμεσα στην ονοματική κλίση και τα παράγωγα ουσιαστικά που εντάσσονται σε μία από τις οκτώ κλιτικές τάξεις. Επιπροσθέτως, δείχνει να αποτελεί λύση, όχι μόνο σε επίπεδο θεωρίας και αρχών, αλλά με πρακτικό και συνάμα αποτελεσματικό τρόπο για το ζήτημα της αλλομορφίας που δεν της έχει δοθεί η δέουσα σημασία στο θεωρητικό της κομμάτι, ενώ στην υπολογιστική επεξεργασία χαρακτηρίζεται από διάφορους ερευνητές ως κακοτοπιά και στατιστική απόκλιση (Goldsmith 2001), οριακά συγχρονικός σωρός σκουπιδιών (Maiden 1992), δυσπροσπελάσιμος ύφαλος που συστηματικά παρακάμπτεται (Wicentowski 2002), αγνοείται και λανθάνει συστηματικής προσοχής (Kurimo *et al.* 2010).

Στο συγκεκριμένο πείραμα της Μέγιστης Εντροπίας καταφάνηκε πόσο απαραίτητη είναι η ύπαρξη μορφολογικά σχολιασμένων σωμάτων κειμένων για την αποτελεσματική διεξαγωγή μορφολογικών πειραμάτων στην Ελληνική γλώσσα. Δείξαμε ότι ένα πιθανοτικό μοντέλο για την εξερεύνηση σωμάτων κειμένων εφαρμοσμένο σε ένα συγκρίσιμο, με πλούσιο σχολιασμό σώμα λέξεων, μπορεί να εξάγει κάποιες βασικές αρχές πάνω, στις οποίες οφείλει να δομηθεί ένα μοντέλο υλοποίησης υπολογιστικής επεξεργασίας της αλλομορφίας.

Πάνω απ' όλα, περιμένουμε ότι η υπολογιστική αντιμετώπιση δύσκολων και παραμελημένων ζητημάτων της μορφολογίας των Ελληνικών θα δώσει μία ώθηση στην υλοποίηση αποτελεσματικότερων αναλυτών για τη γλώσσα. Δευτερευόντως, θα ενισχυθεί η άποψη ότι τα μαθηματικά μοντέλα της γλώσσας όντως αυξάνονται και αποκτούν δυναμική εφαρμογής σε πλούσιες γλώσσες και σε πολύπλοκες μορφολογικά δομές, θέτοντας συστηματική επεξήγηση σε θεωρητικά ζητήματα υπό τη δοκιμασία ενός εμπειρικού ελέγχου. Στην παρούσα έρευνα η αλλομορφία δοκιμάστηκε σε ένα μαθηματικό μοντέλο γλώσσας που απέδειξε τη δύναμη του σε ένα δύσκολο στόχο, την πρόβλεψη των αλλομορφικών αλλαγών.

Πιο συγκεκριμένα, θέσαμε υπό εμπειρική δοκιμή μία υπόθεση που είχαμε διατυπώσει σε θεωρητικό επίπεδο στο πλαίσιο της προβληματικής για την μη – επιβλεπομένη μορφολογική μάθηση (Καρασίμος 2009), ότι η αλλομορφία ως φαινόμενο κρύβει προβλεψιμότητα και κανονικότητα. Χαρακτηριστικά είχαμε αναφέρει ότι «είναι κρίσιμο για την υπολογιστική εκμάθηση της μορφολογίας, οι ανενεργοί κανόνες να κωδικοποιηθούν στο σύστημα, με στόχο να βοηθήσει το σύστημα να αντιμετωπίσει το φαινόμενο της αλλομορφίας· το ότι τα θέματα και τα επιθήματα μοιράζονται αρκετά κοινά χαρακτηριστικά μόνο τυχαίο δεν είναι και πιθανότατα να

κρύβουν κοινά σχεδιάτυπα από πίσω». Σημαντικότερα, κάποια χαρακτηριστικά αποτελούν είδος μορφολογικών περιοριστικών χαρακτηριστικών στην προβλεψιμότητα της αλλομορφίας. Αυτό μας οδηγεί στο να δομήσουμε μία ολοκληρωμένη ομάδα περιοριστικών χαρακτηριστικών για την αλλομορφία, στην οποία είναι εφικτή η ιεραρχία μεταξύ των. Ωστόσο οι κατανομές των βαρών διαφοροποιούνται δραματικά στις αλλομορφικές τάξεις με λίγα μέλη και οι επιρροές τους μοιάζει να μην είναι όσο καίριες αναμενόταν.

Τα αποτελέσματα και τα συμπεράσματά μας έρχονται σε σύγκρουση με την κοινή πεποίθηση αρκετών γλωσσολόγων ότι τα πιθανοτικά μοντέλα είναι μόνο για την επιφανειακή μελέτη των κατανομών των δεδομένων. Για την ακρίβεια, μοντέλα τέτοιου είδους μπορούν να χρησιμοποιηθούν ως πεδίο ελέγχου και για μορφολογικά φαινόμενα και διαδικασίες. Επιπροσθέτως, η υπολογιστική προσομοίωση της αλλομορφίας και οι δυναμικές των περιοριστικών χαρακτηριστικών της γλώσσας φανερώνουν ένα ιδιαίτερα πολύπλοκο σύστημα που, όμως, κρύβει κανονικότητα και προβλεψιμότητα.

*“A year spent in Artificial Intelligence is enough to make one believe in God”*

Alan Perlis

## 6. Υπολογιστική επεξεργασία της αλλομορφίας

### 6.1. Προηγούμενες έρευνες και προτάσεις

Το «περιθωριακό» φαινόμενο – κατά πολλούς ερευνητές – της αλλομορφίας κάθε άλλο παρά κέντρισε το ενδιαφέρον για την αντιμετώπισή του υπολογιστικά από τους ερευνητές της μορφολογικής ανάλυσης· είναι αναμενόμενο να υφίσταται αυτό καθότι είναι ένα ζήτημα ιδιαίτερα παραμελημένο και μάλιστα σε θεωρητικό επίπεδο, όπως ορθώς επισημαίνει η Lieber (1982: 27). Παρόλη την ενασχόληση κάποιων μελετητών με την αλλομορφία, αυτό το θέμα δεν αποτέλεσε το επίκεντρο κάποιας συστηματικής έρευνας, δεν της αποδόθηκε η δέουσα προσοχή στη σημαντική συμμετοχή της στις διαδικασίες σχηματισμού λέξεων, αλλά ούτε έγινε το κέντρο προσοχής στο πλαίσιο της μορφολογίας ή της γενετικής γραμματικής γενικότερα. Για αυτό ευθύνεται σε μεγάλο βαθμό η ελάχιστη συμμετοχή της αλλομορφίας στην Αγγλική γλώσσα, με αποτέλεσμα να υποβαθμιστεί η παρουσία και η συμβολή του φαινομένου στις διαδικασίες σχηματισμού λέξεων. Σπάνια, όμως, η αλλομορφία αποτέλεσε από μόνη της αντικείμενο υπολογιστικής ανάλυσης (ή έστω συμπληρωματικά σε κάποιο ζήτημα), αλλά σχεδόν πάντοτε παρακαμπτόταν ως ένα περιθωριακό ζήτημα που δεν χρήζει άμεσης επίλυσης.

Στο γενικότερο θεωρητικό πλαίσιο η Ράλλη (2005: 18) αναφέρει ότι η καθολικότητα της μορφολογίας επηρεάζεται ιδιαίτερα από διάφορες παραμέτρους που λειτουργούν σε συγκεκριμένες γλώσσες και με τις ιδιαιτερότητες των γλωσσών εν

γένει. Η μορφολογία των γλωσσών είναι πολύ διαφορετική ανάλογα με τον τύπο της γλώσσας, όπως είναι οι **απομονωτικές** (isolating/ analytic, π.χ. Κινεζικά, Βιρμανέζικα) οι **διαχυτικές** (fusional, π.χ. Ελληνικά, Γερμανικά, Ρωσικά), οι **πολυσυνθετικές** (polysynthetic, π.χ. οι γλώσσες των Μάγια, οι γλώσσες Uto-Aztecan και Eskimo-Aleut), οι **σχεδιοτυπικές** (templatic, π.χ. τα Αραβικά, τα Εβραϊκά) και οι **συγκολλητικές** (agglutinative, π.χ. Αλταϊκές γλώσσες, Φινλανδικά, Κορεάτικα). Συστηματικά η πλειονότητα της υπολογιστικής μορφολογικής επεξεργασίας στοχεύει σε γλώσσες με φτωχή μορφολογία, όπως τα Αγγλικά, ή σε συγκολλητικές γλώσσες, όπως τα Φινλανδικά, Τουρκικά, Ουγγρικά, όπου κάθε μόρφημα αντιστοιχεί σε μία λειτουργία, τα μορφήματα τίθενται παρατακτικά σε μία σειρά με ελάχιστες αλλαγές και τροποποιήσεις των μορφών. Πολλοί μορφολογικοί αναλυτές, όπως ο PC-KIMMO, είναι προσανατολισμένοι πρωτίστως στις συγκολλητικές γλώσσες. Έχοντας ως βάση το πόσο διαφορετική είναι η προσέγγιση της μορφολογίας στις συγκολλητικές γλώσσες σε σύγκριση με τις άλλες γλώσσες, κατανοούμε τη δυσκολία τροποποίησης τέτοιων μορφολογικών αναλυτών. Όπως αναφέρει ο Σγάμπας (1997) στα Ελληνικά η πολυπλοκότητα μεταφέρεται από τα μορφήματα των μορφοτακτικών στην πολυσημία και αμφισημία των καταλήξεων με αποτέλεσμα την αυξημένη δυσκολία υλοποίησης ενός επιτυχημένου μορφολογικού αναλυτή. Ακολουθούν οι παρουσιάσεις των ελάχιστων προσπαθειών για την υπολογιστική αντιμετώπιση της αλλομορφίας.

Παρουσιάζουμε αρχικά το έργο των Jurafsky & Martin (2000) που ασχολούνται με την αλλομορφία στο πλαίσιο της υπολογιστικής επεξεργασίας φυσικής γλώσσας και ακολουθούν κάποιο υπολογιστικό μοντέλο. Αναφέρονται στο φαινόμενο όχι ως αλλομορφία, αλλά ως κάτι ανώμαλο (irregular) με μία λιγότερη ή περισσότερη ιδιοσυγκρασία στη μορφή της λέξης. Στην κλίση περιγράφουν τη μη ομαλότητα συνήθως με φωνολογικούς όρους ([α.] *eat ~ ate* ‘τρώω ~ έφαγα’ (αλλαγή φωνήεντος), [β.] *catch ~ caught* ‘πιάνω ~ έπιασα’ (αλλαγή χαρακτήρων), [γ.] *cut ~ cut* ‘κόβω ~ έκοψα’ (αλλαγή χωρίς μόρφημα)). Οι συγγραφείς θεωρούν πως τα 250 εξεταζόμενα ρήματα και τα 20 ουσιαστικά των Αγγλικών που εμφανίζουν αλλομορφία είναι μία μικρή ομάδα καταχωρήσεων με διαφορετική συμπεριφορά και τα καταγράφουν πλήρως στο λεξικό. Έτσι τοποθετούν τη διαδικασία δημιουργίας αλλομόρφου<sup>1</sup>

<sup>1</sup> Πιο συγκεκριμένα με την προσέγγιση των αυτομάτων πεπερασμένων καταστάσεων (finite-state automata): ένα ομαλό ρήμα θα ξεκινήσει από τον κόμβο  $q_0$  πηγαίνοντας στον κόμβο  $q_1$ , για να πιστοποιήσει την κανονικότητα στο σχηματισμό αορίστου και έτσι παίρνει το επίθημα *-ed*, για να οδηγηθεί με ολοκληρωμένη μορφή στον τελικό κόμβο  $q_2$ , σε αντίθεση με ένα ρήμα που χρησιμοποιεί

παράλληλα με αυτή της κλίσης και παραγωγής προσεγγίζοντας τα αλλόμορφα σαν ξεχωριστά λήμματα αντίθετα με τις προσεγγίσεις της θεωρητικής μορφολογίας. Ωστόσο είναι ένα σημαντικό πρόβλημα που θεωρούν ότι το λύνουν σε σημαντικό βαθμό δημιουργώντας είτε ένα μεγάλο δοχείο που περιέχει αποκλειστικά τα αλλόμορφα (κατά τους συγγραφείς τους ανώμαλους τύπους) είτε παραθέτοντας τα μορφήματα τεμαχισμένα σε φωνήματα τύπου  $g\ o:e\ o:e\ s\ e$  (για το ζευγάρι *goose, geese* ‘χήνα, χήνες’),  $m\ o:i\ u:e^2\ s:c\ e$  (για το ζευγάρι *mouse, mice* ‘ποντίκι, ποντίκια’). Για τα Ελληνικά η πρότασή τους θα ήταν  $p\ a\ p\ a\ \delta:e$  (για το πάπα-ς, παπάδ-ες),  $t\ i\ m\ a:i^3$  (για το τιμ(ά)ω, τίμη-σ(α)),  $f\ \theta\ i:a\ r$  (για το φθείρ-ω, έ-φθαρ-α). Είναι όμως εμφανές ότι τα αλλομορφικά δεδομένα της αγγλικής είναι περιορισμένα και τέτοια στρατηγική μπορεί να είναι επιτυχής· φυσικά κάτι τέτοιο είναι τελείως αντιοικονομικό για τα Ελληνικά.

Οι Krieger, Pinker & Nerbonne (1993) στο πλαίσιο της θεωρίας *Head-Driven Phrase Structure Grammar* (HPSG) των Pollard & Sag (1987) προσπάθησαν να χτίσουν ομοιογενή αυτόματα πεπερασμένων καταστάσεων για την αλλομορφία με τα ίδια χαρακτηριστικά της γλώσσας που περιέγραψαν τους μορφοτακτικούς περιορισμούς. Στόχος τους η αντιμετώπιση της αλλομορφίας (πρωτίστως αυτή που καθορίζεται από φωνολογικά περιβάλλοντα<sup>4</sup>) διατηρώντας τη μέγιστη διαφάνεια των υπολοίπων γραμματικών πληροφοριών. Για την ανάλυση της γερμανικής κλίσης υιοθετούν τη θεωρία WP (word-and-paradigm) του Matthews (1972) σε συνδυασμό πάντα με τη στρατηγική της μορφολογίας των Δύο-Επιπέδων (Koskenniemi 1983), ώστε να αντιμετωπίσουν την έλλειψη διαχείρισης της αλλομορφίας από τα υπάρχοντα πεπερασμένα μοντέλα. Χρησιμοποιούν περιορισμούς ανάμεσα στα όρια των μορφηματικών μονάδων και τις αλλομορφικές πραγματώσεις.

Οι περιορισμοί συνεπικουρούνται από την προβλεψιμότητα των φωνολογικών σχημάτων του θέματος και οι εξαγόμενοι κανόνες βασίζονται στην διεπίπεδη μορφολογία.

---

αλλόμορφο για τον αόριστο όπου θα περάσει αυτόματα από το  $q_0$  στο  $q_2$  με τον τύπο έτοιμο. Σε καλύτερες περιπτώσεις κάποιων άλλων αυτόματων υπάρχει κάποιος κόμβος ανάμεσα στο  $q_0$  και στο  $q_1$ , όπου το ανώμαλο ρήμα επιλέγει το κατάλληλο μόρφημα και συνεχίζει έπειτα απευθείας στο  $q_2$ .

<sup>2</sup> Το σύμβολο  $e$  συμβολίζει το κενό.

<sup>3</sup> Να σημειωθεί ότι απουσιάζει ο συμβολισμός  $a:e$  για το αλλόμορφο τύπου *τιμ~*. Ο συμβολισμός  $x:y$  δηλώνει τη χρήση μετατροπέα με το  $x$  να είναι *lexical symbol* και το  $y$  *surface symbol* και δεν μπορεί να γενικευτεί σε  $x:y:z:\dots$  για να δηλώσει διαφορετικές αλλομορφίες στην επιφανειακή δομή. Σε αυτό το μοντέλο οι αλλομορφίες περιγράφονται με διαφορετικά *links* τύπου  $x:y, x:z, x:w$ , κλπ στον μετατροπέα.

<sup>4</sup> Δεν πρόκειται για περιπτώσεις καθαρής αλλομορφίας, αλλά για αλλαγές που προήλθαν από την εφαρμογή φωνολογικών κανόνων.

(92.) e-επένθεση  $+ : e \Leftrightarrow \{d, t\} \_ \{s, t\}$

s-διαγραφή  $s : \emptyset \Leftrightarrow \{s, z, x, ch\} + : \emptyset \_ t$

(Krieger, Pinker & Nerbonne 1993: 144)

Εισάγουν τα χαρακτηριστικά LEX (για την καταχώρηση του λεξικού) και SURF (για την επιφανειακή δομή της ορθογραφίας) για να καταχωρήσουν τις αλλαγές που υφίσταται κάποιο θέμα, όταν υποχρεούται να εφαρμόσει αλλομορφικούς περιορισμούς.

Τα πεπερασμένα συστήματα έχουν υποστεί έντονη κριτική ήδη από τα πρώτα τους βήματα για διάφορα ζητήματα. Οι Krieger, Pinker & Nerbonne (1993) κατανοούν τα όρια ενός πεπερασμένου συστήματος και επισημαίνουν ότι κλασικοί κανόνες δύο-επιπέδων για την αλλομορφία είναι εφαρμόσιμοι σε φωνολογικά περιβάλλοντα στα όρια λέξεων και μορφημάτων. Για την ακρίβεια, η διαπίστωση ότι συχνά αλλομορφικές αλλαγές είναι «απολιθώματα» που έχουν ξεφύγει από τη γραμμή της φωνολογικής εξάρτησης, εφαρμόζονται σε συγκεκριμένες κατηγορίες λέξεων και σε θέματα με ιδιοσυγκρατικά χαρακτηριστικά, τους οδηγεί να επεκτείνουν τη στρατηγική τους. Έτσι καθορίζουν τους αλλομορφικούς περιορισμούς σε συγκεκριμένες τάξεις των λημμάτων κάνοντας χρήση της στρατηγικής τεχνικής με την κληρονόμηση χαρακτηριστικών από το λεξικό και έτσι απελευθερώνουν τους κανόνες από τα αυστηρά όρια του φωνολογικού περιβάλλοντος.

Σε γενικές γραμμές, προτιμούν ένα σύστημα που να περιλαμβάνει όλες τις διαθέσιμες γλωσσολογικές πληροφορίες. Εφαρμόζουν έναν ομοιογενή φορμαλισμό για τη διαφύλαξη της διαφάνειας, τον καθορισμό της εξάρτησης των αλλομορφών από συγκεκριμένα μορφολογικά περιβάλλοντα και την άμεση εφαρμογή περιορισμών για την φωνολογική αλλομορφία. Επιπροσθέτως, η χρήση των χαρακτηριστικών τους επιτρέπει να παρακάμψουν τη διακλάδωση της διεπίπεδης μορφολογίας και να αντιμετωπίσουν περιπτώσεις καθαρής αλλομορφίας. Σίγουρα ο διαχωρισμός της πραγματικής αλλομορφίας από τα φωνόμορφα θα τους επέτρεπε να διαχωρίσουν τους κανόνες σε δυο διαφορετικές ομάδες με καλύτερα αποτελέσματα, εφόσον ακολουθούσαν τη συγκεκριμένη τακτική. Επίσης, είναι αξιοσημείωτο ότι δεν ασχολούνται καθόλου με την κλιτική αλλομορφία των ουσιαστικών των Γερμανικών, όπου η αλλομορφία είναι καθαρά θεματική και φυσικά δεν βασίζεται σε φωνολογικά περιβάλλοντα. Μία προσέγγιση που εφαρμόζει κανόνες αλλομορφίας κατά τη διάρκεια του σχηματισμού των λέξεων έχει το μειονέκτημα να κάνει τη διαδικασία πιο

περίπλοκη, αλλά και να έχει φανερές δυσκολίες στις διαδικασίες της παραγωγής και της σύνθεσης. Αναγνωρίζουν τα κέρδη της δικής τους προσέγγισης, αλλά κατανοούν τα προβλήματα της καθαρής αλλομορφίας που δεν έχουν αντιμετωπίσει ή έχουν παραβλέψει, όπως το φαινόμενο της ανενεργής ετεροίωσης ή πιο πολύπλοκες αλλομορφικές αλλαγές.

Σε πιο πρόσφατες προσεγγίσεις, βασισμένοι στη θεωρητική προσέγγιση του SLIM και LAG<sup>5</sup> οι Handl *et al.* (2009) δημιούργησαν τον μορφολογικό αναλυτή JSLIM (το J είναι για τη γλώσσα προγραμματισμού Java) που βασίζεται στην αριστερόστροφη ανάλυση<sup>6</sup>· το κομμάτι της εισαγόμενης λέξης που έχει ήδη αναλυθεί συνδυάζεται με το ακόλουθο κομμάτι (λέξη ή επίθημα), το οποίο στη μορφολογική ανάλυση είναι συνήθως το αλλόμορφο του εισαγόμενου (ό.π.: 11). Μάλιστα ο αριθμός των απαιτούμενων βημάτων συνδυασμού για τον επιτυχή διαχωρισμό του εισαγόμενου αντιστοιχεί στον συνολικό αριθμό των αλλομόρφων του εισαγόμενου μείον ένα. Για την ανάλυσή τους ακολουθούν την *έμμεση προσέγγιση*, όπου οι πληροφορίες αναφέρονται στο θέμα/ λήμμα και δεν καταχωρούνται άμεσα σ' αυτό.

Μένοντας πιστοί σε προηγούμενη έρευνα του LAG (βλ. Schüller & Lorenz 1994) ακολουθούν την ίδια μέθοδο για τα αλλόμορφα εμπλουτίζοντάς την. Η στρατηγική τους είναι απλή, καθότι δημιουργούν ένα βασικό λεξικό και λεξικό αλλομόρφων, εντάσσοντας στο τελευταίο και τα μορφήματα που προκύπτουν από φωνολογικούς κανόνες. Το λεξικό των αλλομόρφων παράγεται από το βασικό λεξικό με την εφαρμογή των **αλλο-κανόνων** (allo-rules) πριν την εφαρμογή της ανάλυσης. Κατά τη διάρκεια επεξεργασίας τα αλλόμορφα του λεξικού αλλομόρφων είναι διαθέσιμα ως προ-υπολογισμένες, πλήρως αναλυμένες μορφές παρέχοντας τον μέγιστα δυνατόν απλό διαχωρισμό: η άγνωστη επιφανειακή δομή χωρίζεται από τα αριστερά στα δεξιά στα κατάλληλα αλλόμορφα χωρίς καμία υποβάθμιση σε μορφήματα. Σύμφωνα με τους συγγραφείς η δομή του βασικού λεξικού διασφαλίζεται από το λεξικό των αλλομόρφων, αφού το δεύτερο παρέχει τις πληροφορίες που λείπουν για κάθε καταχώρηση.

<sup>5</sup> Το SLIM είναι ακρωνύμιο από το Surface compositional Linear Internal Machine (βλ. Hausser 2006), ενώ το LAG είναι ακρωνύμιο από το Left-Associative Grammar (βλ. Hausser 1999).

<sup>6</sup> Αριστερόστροφη ανάλυση είναι αυτή που ξεκινάει την ανάλυση των λέξεων από την αρχή της και όχι από το τέλος.



Οι αλλο-κανόνες κωδικοποιούνται με σαφή τρόπο μέσω προτύπων (patterns) και με έμμεσο τρόπο σε κάθε εγγραφή του λεξικού με δείκτη αναφοράς στο σχεδιάτυπο που φέρει τον κατάλληλο άλλο-κανόνα. Στο ακόλουθο σχήμα φαίνεται πως λειτουργεί ένας άλλο-κανόνας στα γερμανικά για αλλόμορφα με umlaut τύπου ‘Bach’. Η δική τους εκδοχή είναι πιο βελτιωμένη από παλαιότερες, καθότι τα αλλόμορφα μίας βασικής μορφής πρέπει να παραχθούν με την εφαρμογή όλων των άλλο-κανόνων.

Γράφημα 8: Άλλο-κανόνας από τα γερμανικά για τη λέξη *Bach* (Handl *et al.* 2009:20)

```
!template[cat: (n' a' v), sem: (pres_,...)]
![sur cor]
lern lernen
erb erben
table A_Bach: [cor]          => [sur, cor, cat, sem]
/(.*?)([ΑΟΥαου])(u?[^αουäöü]*)/ => /#0/ /$0/ (m-g) (m sg);
                                     => /$1$2"$3/ /$0/ (pstem) (m pl).
```

Παρακάμπτοντας πρότερες τακτικές με την εφαρμογή μίας γεννήτριας αλλομόρφων για την επέκταση των σχεδιοτύπων και τη δημιουργία ολοκληρωμένων δομών, καταφεύγουν στο πλεονέκτημα που τους παρέχει η αποθήκευση των αλλομορφικών σχεδιοτύπων ως τιμή ιδιότητας στις εγγραφές του λεξικού. Έτσι ανακαλούνται οι αλλο-κανόνες από τη δομή των χαρακτηριστικών/ ιδιοτήτων της βασικής μορφής, όποτε κρίνεται απαραίτητο. Εφαρμόζεται σε κάθε περίπτωση πριν συμμετάσχει σε διαδικασία σχηματισμού λέξης και διαχωρίζονται οι σημασιολογικές πληροφορίες ανάλογα με τη μορφή.

Κρίνοντας την προσπάθεια, πρέπει να σημειώσουμε πως είναι ιδιαίτερα συστηματική η αντιμετώπιση της αλλομορφίας, ενώ σημειώνουν συνειδητά τη σημαντικότητα του φαινομένου στις μορφολογικές διαδικασίες. Βελτιώνουν προηγούμενες στρατηγικές που είχαν προβλήματα είτε κατά την εκτέλεση των αλλο-κανόνων, είτε κατά την περαιτέρω ανάλυση των λέξεων. Εντούτοις, επιμένουν να διαχωρίζουν τα λεξικά σε βασικά μορφήματα και αλλόμορφα, χωρίς το δεύτερο να τους παρέχει κάποιο ουσιαστικό όφελος. Με έντονες επιρροές από τα αυτόματα πεπερασμένων καταστάσεων, συχνά η κωδικοποίηση των αλλο-κανόνων περιπλέκεται δημιουργώντας προβλήματα σε κάποιες δομές. Να σημειωθεί ότι κατά τη δοκιμή<sup>7</sup> τους στα Ελληνικά η αλλαγή θέσης τόνου σε κάποιο λέξημα εισήγαγε καινούργια καταχώρηση στο λεξικό των αλλομόρφων, ενώ ήταν δυνατή μόνο η κωδικοποίηση με

<sup>7</sup> Η δοκιμή έγινε στο πλαίσιο του workshop *State of the Art in Computational Morphology*, Σεπτέμβριος 2009 στη Ζυρίχη. Ακόμα και η αναβαθμισμένη έκδοση λίγους μήνες αργότερα είχε τα ίδια προβλήματα για τα Ελληνικά.

χαρακτηριστικά της κλίσης. Δεν ήταν επομένως εφικτή η χρήση του περιορισμού της επιλογής μοναδικού αλλόμορφου στην κωδικοποίηση των παραγώγων και συνθέτων της Ελληνικής. Κλείνοντας να σημειωθεί, ότι υποστηρίζουμε πως η παραπάνω προσπάθεια αποτελεί μία αξιόλογη υλοποίηση μοντέλου και συμβάλλει στη συστηματική βελτίωση των μορφολογικών αναλυτών πλούσιων μορφολογικά γλωσσών.

Η πιο δημοφιλής (και μοναδική) υλοποίηση στη ΜεΜΜ είναι ο Allomorfessor (Kohonen, Virpioja & Klami 2009, Virpioja & Kohonen 2009) που δοκιμάστηκε και στους τρεις τελευταίους διαγωνισμούς του Morpho Challenge (Kurimo *et al.* 2010). Η μέθοδός του βασίστηκε στη διαπίστωση ότι το φαινόμενο της αλλομορφίας περιορίζει την ποιότητα ανάλυσης μορφημάτων σε ένα σύστημα, μίας και τα αλλόμορφα ενός μορφήματος εμφανίζονται σε συγκεκριμένα μορφο- και φωνολογικά περιβάλλοντα και οι περισσότερες αλλαγές υφίστανται πλησίον των μορφηματικών ορίων. Το συγκεκριμένο πρόγραμμα υλοποιήθηκε πάνω στον αναλυτή *Morfessor Baseline* χρησιμοποιώντας τη βελτιστοποίηση του *Maximum A Posteriori* (MAP). Το σώμα κειμένων χωρίζεται σε δύο μοντέλα, ένα για τις λέξεις ( $M_W$ ) και ένα για τα μορφήματα ( $M_M$ ): παράλληλα χτίζεται ένα λεξικό λέξεων ( $L_W$ ), ένα λεξικό μορφημάτων ( $L_M$ ) μαζί με μία γραμματική για αυτά ( $G_M$ ). Στόχος του προγράμματος είναι να εντοπίσει και να αναλύσει σωστά τις μεταβολές/ μεταλλάξεις (mutations) στην επιφανειακή δομή των μορφημάτων και έτσι τα αλλομορφήματα (allomorphemes) μπορούν να διαμορφωθούν σύμφωνα με αυτές τις ετεροιώσεις.

Οι μεταβολές/ ετεροιώσεις που διαχειρίζεται ο Allomorfessor είναι η διαγραφή και η αντικατάσταση φωνήματος. Η επηρεασμένη θέση εντοπίζεται με την ταυτοποίηση της  $k$ -πλησίον απόστασης από το γράμμα-στόχο, το οποίο εντοπίστηκε προηγουμένως στο τέλος ενός εικονικού επιθήματος:

Λειτουργία	Σημείωση	Περιγραφή
Αντικατάσταση	$kx y$	αλλαγή του $k$ :th $x$ με $y$
Διαγραφή	$-kx$	απομάκρυνση του $k$ :th $x$
Μόρφημα	Ετεροίωση	Αλλόμορφο
kenkä (παπούτσι)	(k g)	kengä (π.χ. kengä+ssä, στο παπούτσι)
ranta (ακτή)	(-a t n)	rann (π.χ. rann+oi+lla, στις όχθες)
ihminen (άνθρωπος)	(2n s)	ihmisen (π.χ. ihmisen, του ανθρώπου)

Πίνακας 20: *Επιτρεπόμενες ετεροιώσεις του συστήματος για τα Φινλανδικά (Virpioja & Kohonen 2009: 4).*

Οι παραπάνω ετεροιώσεις εφαρμόστηκαν σε όλο το σώμα κειμένων, αλλά με τις αλλομορφικές αλλαγές σε προσφύματα και μικρά θέματα να αποκλείονται από την αξιολόγηση. Σημειώνεται πως αν συμπεριληφθούν θα πέσει η επίδοση του συστήματος δραματικά· μάλιστα για τα Τουρκικά, όπου η περισσότερη αλλομορφία εντοπίζεται στα προσφύματα ή/ και στα μικρά θέματα, μόνο το 2% των περιπτώσεων αλλομορφικής αλλαγής μπορεί να εντοπιστεί!

Για την καλύτερη επίδοση του συστήματος, ο αναλυτής εκπαιδεύτηκε σε περισσότερα δεδομένα. Όπως σημειώνουν οι Kurimo *et al.* (2010) ο Allomorfessor εκπαιδεύτηκε με φιλτραρισμένα δεδομένα (λέξεις που απαντώνται μόνο μία φορά διαγράφηκαν), ώστε το σύστημα να επιτύχει υψηλότερη ανάκληση. Το σύστημα βελτίωσε ελαφρώς τις επιδόσεις στη γλωσσική αξιολόγηση, αλλά στον τομέα της Ανάκτησης Πληροφορίας (Information Retrieval) και στη Στατιστική Μηχανική Μετάφραση (Statistical Machine Translation) απέτυχε να σημειώσει αξιόλογες επιδόσεις. Σε γενικές γραμμές η επίδοσή του ήταν καλή, αλλά όπως τα περισσότερα συστήματα δεν τα κατάφεραν γενικά καλά σε όλες τις γλώσσες του διαγωνισμού· μοναχά στην ανάλυση του αγγλικού λεξικού κατάφεραν να πρωτεύσουν.

Είναι αξιοσημείωτο ότι οι «αλλομορφικοί» κανόνες περιορίζονται σε αυτούς της αντικατάστασης και της αφαίρεσης φωνήματος, ενώ παραβλέπονται οι περιπτώσεις προσθήκης φωνήματος ή πολλών αλλαγών στο θέμα (π.χ. *come~ came, παίρν~ πηρ~ παρ*). Εντούτοις, αποτελεί αρκετά επιτυχημένη προσπάθεια, αναλογιζόμενοι τον αλγόριθμο αλλομορφίας του Linguistica. Να σημειωθεί το γεγονός ότι πρέπει να ενταχθεί στις τακτικές της ελαφρώς εποπτευόμενης μορφολογικής μάθησης. Κινείται προς τη σωστή κατεύθυνση, καθότι τοποθετεί στο επίκεντρο ενός άτυπου διαγωνισμού μοντέλων και θεωριών το ζήτημα της αλλομορφίας που είναι μείζον για κάποιες γλώσσες και δίνει αρκετά σωστά αποτελέσματα. Ωστόσο, βρίσκεται σε αρχικό στάδιο, αφού αντιμετωπίζει μόνο μέρος των θεμάτων και δεν ασχολείται με τα μικρά θέματα και τα προσφύματα.

## 6.2. Η προβληματική της θεματοκεντρικής ή προσφυματοκεντρικής υπολογιστικής προσέγγισης της αλλομορφίας

Στην υπολογιστική μορφολογία, υφίσταται πάντα μία προβληματική για τον τρόπο έναρξης ανάλυσης μίας λέξης, δηλαδή από την αρχή της (συνήθως ρίζα/ θέμα) ή από το τέλος της (συνήθως επίθημα). Η **θεματοκεντρική τεχνική** (stemming) είναι η διαδικασία για τη μείωση συνήθως των κλιτών και παράγωγων λέξεων με την αποκοπή των επιθημάτων, με στόχο την απομόνωση του θέματος ή της ρίζας. Το θέμα/ ρίζα δεν πρέπει να ταυτίζεται με τη μορφολογική ρίζα της λέξης, μίας και συνήθως συσχετίζεται μέσω της λεξικής απεικόνισης (word mapping) με το θέμα, ακόμα κι όταν δεν είναι έγκυρη ρίζα. Ένας τέτοιος αλγόριθμος αποτελεί μακροχρόνιο ζήτημα στον τομέα της υπολογιστικής γλωσσολογίας με την πρώτη αναφορά να γίνεται στα τέλη της δεκαετίας του '60 (Lovins 1968). Η διαδικασία της θεματοκεντρικής τεχνικής ονομάζεται και συγχώνευση (conflation) και είναι χρήσιμη στις μηχανές αναζήτησης. Τα προγράμματα βασισμένα στη συγκεκριμένη τεχνική ονομάζονται θεματοποιητές.

Η πλέον σημαντική προσπάθεια έγινε από τον Porter (1980), του οποίου ο αλγόριθμος είναι ευρέως διαδεδομένος και έγινε *de facto* ο πρότυπος αλγόριθμος για την αγγλική θεματοκεντρική τεχνική. Πολλές εφαρμογές που απορρέουν από τον αλγόριθμο του Porter δημιουργήθηκαν και χορηγήθηκαν δωρεάν, χωρίς ωστόσο να επιτύχουν τις πραγματικές δυνατότητες της τεχνικής: πολλές από αυτές τις εφαρμογές περιέχουν χτυπητές αδυναμίες. Ως εκ τούτου, οι εν λόγω αλγόριθμοι δεν κατάφεραν να επιτύχουν υψηλά ποσοστά ακρίβειας σύμφωνα με τις προσδοκίες των ερευνητών και υποστηρικτών. Για να αντιστρέψει την αρνητική εικόνα από τα σφάλματα των αλγορίθμων, ο Porter (2001) κυκλοφόρησε μία επίσημη ελεύθερη έκδοση του αλγορίθμου, με την ονομασία Snowball<sup>8</sup>, στον οποίο εφάρμοσε μία βελτιωμένη αγγλική έκδοση του αλγορίθμου σε συνδυασμό με αλγόριθμους για πολλές άλλες γλώσσες.

Η λογική πίσω από τις τεχνικές ενός θεματοποιητή ή ενός επιθηματιστή/ αποκοπέα επιθημάτων (suffix stripper) δεν βασίζεται σε κάποιους πίνακες με κλιτικά παραδείγματα ή ομάδες επιθημάτων. Αντιθέτως, έχει καταχωρημένο ένα είδος λίστας από κανόνες, ώστε να αποτελέσει το μονοπάτι για την οδήγηση του αλγορίθμου στη ρίζα. Για παράδειγμα:

---

<sup>8</sup> Η επίσημη ιστοσελίδα είναι <http://snowball.tartarus.org/>.

- (93.) α. εάν η λέξη τελειώνει σε –ed, διάγραψε το ‘ed’ Αγγλικά  
 εάν η λέξη τελειώνει σε –ly, διάγραψε το ‘ly’  
 εάν η λέξη τελειώνει σε –est, διάγραψε το ‘est’  
 εάν η λέξη τελειώνει σε –ing, διάγραψε το ‘ing’
- β. εάν η λέξη τελειώνει σε –ος, διάγραψε το ‘ος’ Ελληνικά  
 εάν η λέξη τελειώνει σε –μένος, διάγραψε το ‘μένος’  
 εάν η λέξη τελειώνει σε –ως, διάγραψε το ‘ως’

Το σημαντικό πλεονέκτημα αυτών των τεχνικών είναι η απλότητά τους σε σχέση με άλλες στρατηγικές, αν και έχουν φανερό αδυναμία για τις λέξεις με ανώμαλους τύπους ή που έχουν παραπάνω της μίας πραγματώσεις. Για αυτό και απαιτείται η συνδρομή από τη **λημματοποίηση** (lemmatization). Στόχος της λημματοποίησης είναι ο καθορισμός του λήμματος για μία δοθείσα λέξη. Για τις πολύπλοκες λέξεις που είναι απαραίτητη η χρήση των συμφραζομένων με τη βοήθεια ενός tagger, η υλοποίηση ενός λημματιστή<sup>9</sup> (lemmatizer) αποτελεί δύσκολο στόχο. Είναι μία τεχνική υψηλής συχνότητας εμφάνισης μορφών για τη σωστή γραμματική κατηγοριοποίηση, καθότι σε αρκετές περιπτώσεις συμπίπτουν ονοματικές, ρηματικές και επιθετικές συνθήκες. Παράλληλα οι δυνατότητές του είναι περιορισμένες καθώς αδυνατεί να εντοπίσει τις λανθασμένες κατηγορίες και να παράγει τις σωστές, εκμεταλλευόμενος τη συνδρομή του επιθηματιστή (suffix stripper).

Κατά την εξέλιξη των αλγορίθμων τα λάθη που εντοπίζονταν ήταν συνήθως από την υπερεφαρμογή των κανόνων που είχαν τη μορφή: (υπόθεση) Suf1 → Word-Suf1 και πάντοτε ένας μόνο κανόνας εφαρμοζόταν για μία συνθήκη, εκείνος με το μακρύτερο ταίριασμα του επιθήματος Suf1. Αυτό είχε ως αποτέλεσμα είτε να πετυχαίνει είτε να αποτυγχάνει ο κανόνας: εκτός από αυτό πολλές φορές εφαρμοζόταν κάποιος κανόνας ακόμα κι αν περνούσε όλη η λίστα των κανόνων χωρίς να επιτύχει κανένας.

Συχνά αυτό δημιουργούσε προβλήματα είτε με την υπερανάλυση λέξεων είτε με την λανθασμένη οριοθέτηση λόγω της επανάλυσης ή της μορφολογικής αδιαφάνειας

<sup>9</sup> Από τους λιγοστούς λημματιστές είναι ο Lemmatizer (<http://lemmatizer.org/>) για Αγγλικά και Ρωσικά, ο Collatinus ([http://offset.sourceforge.net/freeduc/book/book\\_27.html](http://offset.sourceforge.net/freeduc/book/book_27.html)) για τα Λατινικά και ο MorphAdorner (<http://morphadorner.northwestern.edu/>) για τα Αγγλικά βασισμένο σε Java. Για τα Αρχαία Ελληνικά υπάρχει ο λημματιστής Morpheus που συνοδεύει το TLG (για περισσότερες πληροφορίες στις ιστοσελίδες: <http://www.tlg.uci.edu/help/Lemma/aboutlemma.html> και <http://llc.oxfordjournals.org/content/6/4/243.short>).

λέξεων. Επομένως, είναι αναμενόμενο να δούμε περιπτώσεις, όπως έξ-ω (μαζί με το έξ-ι και έξ-η), πάν-ω (μαζί με το Πάν-ος, παν-ί), κοιλ-άδα (αλλά και κατσ-άδα, αγελ-άδα), μαστ-ούρα (μαζί με το μαστ-ός, μαστ-ίζ-ω) κτλ. Η συνδρομή του λημματιστή δεν μπορεί να βοηθήσει σημαντικά, γιατί μόνο με τη σημασιολογία μπορείς να διαχωρίσεις κάποια θέματα με ψευδοθέματα, ώστε να αποτρέψεις λανθασμένες κρίσεις και διαχωρισμούς. Μπορεί να βρίσκει πάντα το ίδιο θέμα στις παράγωγες λέξεις, αλλά μόλις χρειαστεί να συγκρίνει τα λήμματα από τις παράγωγες με τα αντίστοιχα των κλιτών, τότε τα αλλόμορφα θα αποτελέσουν σοβαρό πρόβλημα.

Είναι βέβαιο ότι διάφορες στρατηγικές μη-εποπτευόμενης ή εποπτευόμενης μορφολογικής μάθησης επηρεάστηκαν από τις τακτικές των θεματιστών, λημματιστών και επιθηματιστών. Με βελτιωμένες τεχνικές επιχείρησαν με μεγάλη επιτυχία να παρακάμψουν τους κανόνες και να ξεχωρίσουν τα επιθήματα με αυτόματες διαδικασίες. Εξάλλου, είναι σύνηθες να δημιουργούν δύο λεξικά, ένα κανονικό και ένα αντίστροφο, βασιζόμενοι στην τακτική της ακολουθίας χαρακτήρων για ομαδοποίηση των λέξεων με κοινό θέμα/ ρίζα ή με κοινό επίθημα/ πρόθημα. Αναμφίβολα οι (μη-)εποπτευόμενες τεχνικές υπερτερούν με τη δυνατότητα χρήσης αλλομορφικών κανόνων για τη συσχέτιση δύο διαφορετικών μορφών του ίδιου μορφήματος. Εξάλλου και ο Porter (1980, 2001) κατανοεί την αδυναμία των συγκεκριμένων αλγορίθμων με τους ανώμαλους τύπους και επομένως, εμμέσως, παραδέχεται την αδυναμία θεραπείας της αλλομορφίας με τέτοιες τακτικές.

### 6.3. Προτεινόμενο πλαίσιο Επεξεργασίας της Αλλομορφίας

Στον τομέα της επεξεργασίας φυσικής γλώσσας και ειδικότερα στον τομέα της υπολογιστικής μορφολογίας, η υιοθέτηση ή δημιουργία ενός θεωρητικού πλαισίου οφείλει να συνδέεται άμεσα με τη δυνατότητα εφαρμογής του στο υπολογιστικό μοντέλο. Για τη θεωρητική ανάλυση της αλλομορφίας υιοθετήθηκε και τροποποιήθηκε ελαφρώς το μοντέλο της Ράλλη (2000, 2005) για την αλλομορφία και την παραγωγή στα Ελληνικά. Το προτεινόμενο πλαίσιο μας για την υπολογιστική επεξεργασία στο μορφολογικό επίπεδο της παραγωγής στοχεύει σε μία εφικτή και δυνατή υπολογιστική υλοποίηση που να προσομοιάζει με σημαντική ακρίβεια τη συγκεκριμένη μορφολογική διαδικασία, την πραγματική δομή των παραγώγων λέξεων και την επιτυχή ανάλυση του φαινομένου της αλλομορφίας. Στην παρούσα διατριβή, το σκεπτικό της

υλοποίησης της υπολογιστικής επεξεργασίας της αλλομορφίας στην παραγωγή των Ελληνικών βασίζεται σε δύο διαφορετικά μοντέλα και στη δυνατότητα συνδυασμού των θετικών δυνατοτήτων ενός μοντέλου με προεισαγμένους κανόνες και ενός μοντέλου πρόβλεψης Μέγιστης Εντροπίας. Ο κεντρικός άξονας της υλοποίησης είναι η παρουσίαση δύο διαφορετικών προτάσεων (6.3.1. και 6.3.2) που αφορούν στο κομμάτι σύνδεσης και παραγωγής αλλομορφικών κανόνων και την ικανότητα ενός μοντέλου που να ανακαλύπτει και να προβλέπει τις αλλομορφικές συμπεριφορές των παράγωγων λέξεων.

Η ανάλυσή μας υιοθετεί το θεωρητικό μοντέλο της Ράλλη (2000, 2005) το οποίο στηρίζεται στην άποψη ότι η αλλομορφία είναι ένα φαινόμενο που συμμετέχει με την ίδια συστηματικότητα σε όλες τις μορφολογικές διαδικασίες σχηματισμού λέξεων και ότι στην παραγωγή στα Ελληνικά τα παραγωγικά προσφύματα επιγράφονται με χαρακτηριστικά που κωδικοποιούν την πληροφορία για ενδεχόμενη αλλαγή γραμματικής κατηγορίας, κλιτικής τάξης και γένους (βλ. Μαρκόπουλος 1998: 80-87). Το σκεπτικό της παρούσας έρευνας είναι να δημιουργήσουμε ένα συνδυαστικό μοντέλο υπολογιστικής επεξεργασίας που να υλοποιεί τα βασικά χαρακτηριστικά στοιχεία της θεωρίας. Η πρόταση των κανόνων αλλομορφικής μετατροπής είναι η προσπάθεια για την εφαρμογή των πλεοναστικών κανόνων, όπου υπολογιστικά οι κανόνες θα παράγουν τα αλλόμορφα από ένα βασικό μόρφημα και συνήθως αυτοί οι κανόνες αποτελούν την εφαρμογή ανενεργών γλωσσολογικών κανόνων. Το τελικό αποτέλεσμα των κανόνων είναι η σύνδεση των αλλομόρφων με μία σχέση αντίστοιχη των πλεοναστικών κανόνων, χωρίς να υποστηρίζουμε ότι σε θεωρητικό επίπεδο τα αλλόμορφα παράγονται από κάποια άλλη μορφή του ίδιου μορφήματος. Ταυτόχρονα, βασισμένοι στα αποτελέσματα του πειράματος AMIS εισάγουμε ένα μοντέλο πρόβλεψης της κανονικότητας της αλλομορφίας, ώστε το μοντέλο επεξεργασίας να μαντεύει με μεγάλη ακρίβεια την αλλομορφική συμπεριφορά των παράγωγων λέξεων, τόσο σε επίπεδο θέματος, όσο και παραγωγικού προσφύματος.

Κλείνοντας η υπολογιστική ανάλυση της αλλομορφίας στην παραγωγή στοχεύει στην ενοποίηση των δύο παραπάνω προσεγγίσεων της αλλομορφίας, δηλαδή του σχηματισμού των αλλομόρφων και της πρόβλεψης της αλλομορφικής συμπεριφοράς, ενσωματώνοντάς τα σε ένα μοντέλο εποπτευόμενης μορφολογικής μάθησης που θα φέρει κανόνες παρόμοιους με αυτούς των Handl *et al.* (2009) και Kohonen, Virpioja & Klami (2009), Virpioja & Kohonen (2009). Η φιλοσοφία του προτεινόμενου πλαισίου

είναι να θέσει τις βάσεις για την αντιμετώπιση της αλλομορφίας, μίας και είναι ένα θέμα που λανθάνει της προσοχής των ερευνητών. Από τη στιγμή που οι περισσότερες προσεγγίσεις αποτυγχάνουν να αντιμετωπίσουν με επιτυχία την αλλομορφία, είναι σαφές ότι η επίλυση της αλλομορφικής επεξεργασίας κρύβεται πίσω από τις συνδυαστικές ικανότητες διαφορετικών μοντέλων. Ο Wicentowski (2002, 2004) απέδειξε πως η συνδυαστική λειτουργία εποπτευόμενων και μη-εποπτευόμενων μοντέλων έχει συγκριτικά καλύτερα αποτελέσματα στην κλιτική μορφολογία από ότι κάθε προσέγγιση από μόνη της.

### **6.3.1. Ανενεργοί Γλωσσολογικοί Κανόνες – Η πρόταση των κανόνων αλλομορφικής μετατροπής**

Στην παρούσα προτεινόμενη υπολογιστική πρόταση για την επεξεργασία της αλλομορφίας στοχεύουμε στη δημιουργία υπολογιστικών κανόνων που να παράγουν αλλόμορφα από κάποιο μόρφημα, έχοντας ως στόχο το τελικό αποτέλεσμα να προσομοιάζει τους πλεοναστικούς κανόνες. Οι υπολογιστικοί κανόνες μετατροπής είναι κυρίαρχο συστατικό στα μοντέλα αυτομάτων πεπερασμένων καταστάσεων, αλλά παράλληλα υπεισέρχονται και σε ελαφρώς εποπτευόμενα μορφολογικά μοντέλα για τη διευθέτηση μορφοφωολογικών αλλαγών στα όρια μορφημάτων. Συνήθως οι υπολογιστικοί κανόνες ταυτίζονται με τους γλωσσολογικούς στο επίπεδο εφαρμογής της μορφοφωολογίας (π.χ. το φαινόμενο της αφομοίωσης στα όρια μορφημάτων, βρέχ-ω > βρεχ-σ-αμε [βρέξαμε]). Στο προτεινόμενο μοντέλο οι υπολογιστικοί κανόνες εκμεταλλεύονται σε μερικές περιπτώσεις ανενεργούς γλωσσολογικούς κανόνες, για να σχηματίσουν τα αλλόμορφα. Σε γενικές γραμμές οι συγκεκριμένοι κανόνες δεν αντιστοιχούν σε συγχρονικούς γλωσσολογικούς κανόνες και στοχεύουν να αποδώσουν ως αποτέλεσμα τη μορφή των πλεοναστικών κανόνων. Επιπροσθέτως, προτείνουμε τη χρήση ενός φίλτρου μορφολογικού περιβάλλοντος, ώστε κάθε αλλόμορφο που θα παράγεται από τους υπολογιστικούς κανόνες αλλομορφικής μετατροπής θα έχει την απαραίτητη κωδικοποίηση σχετικά με το περιβάλλον συμμετοχής του.

Στο κεφάλαιο της θεωρητικής ανάλυσης της αλλομορφίας αναφέραμε ότι στην πλειονότητα της αλλομορφίας κρύβονται ανενεργοί μορφολογικοί και φωολογικοί κανόνες. Για την υπολογιστική επεξεργασία της αλλομορφίας, αναλύουμε την αλλομορφική συμπεριφορά των θεμάτων και προσφωμάτων. Πρώτα, παρατηρούμε τα συστηματικά χαρακτηριστικά των αλλομόρφων και δημιουργούμε κάποιους



υπολογιστικούς κανόνες. Προτού παρουσιάσουμε και αναλύσουμε συστηματικότερα τους **κανόνες αλλομορφικής μετατροπής**<sup>10</sup> (allomorphic transformation rules) οφείλουμε να καταστήσουμε σαφές ότι αυτοί οι κανόνες δεν αποτελούν υλοποίηση του θεωρητικού μοντέλου. Είναι αντιοικονομικό για κάθε εισαγωγή ενός λήμματος σε ηλεκτρονικό λεξικό μορφολογικού αναλυτή να εισάγουμε χειρωνακτικά και τα αλλόμορφα, εφόσον αυτά υφίστανται. Οι υπολογιστικοί κανόνες περιγράφουν τη σχέση μεταξύ του αρχικού μορφήματος και του αλλομόρφου, τις αλλαγές που εντοπίζονται και δείχνουν πως παράγεται τελικά το αλλόμορφο· αν και υποστηρίζουμε πως στο Νοητικό Λεξικό τα αλλόμορφα υπάρχουν στα λήμματα και συνδέονται με πλεοναστικούς κανόνες μεταξύ τους, στην προκειμένη περίπτωση οι κανόνες αποτελούν συμβατικό τρόπο για την αντιμετώπιση της αλλομορφίας στην προσπάθεια δημιουργίας ενός αποτελεσματικού μορφολογικού αναλυτή.

Αυτοί οι κανόνες στοχεύουν στην περιγραφή των αλλομορφικών μετατροπών και την κατάλληλη κωδικοποίηση για έναν αποτελεσματικό μορφολογικό αναλυτή· βοηθούν κυρίως το σύστημα να αναλύσει σωστά τις λέξεις και να δημιουργήσει ένα σύνδεσμο ανάμεσα στο αλλόμορφο και το μόρφημα. Σημαντική βοήθεια αρχικά παρέχει η Ιστορική Γλωσσολογία για τη δημιουργία των κανόνων αλλομορφικής μετατροπής, καθότι αρκετά αλλόμορφα παρήχθησαν από κανόνες παλαιότερης ιστορικής περιόδου της Ελληνικής γλώσσας, οι οποίοι συγχρονικά είναι ανενεργοί. Για τις αλλομορφικές αλλαγές που προέκυψαν από τη διαδικασία επανάλυσης, όπως για παράδειγμα *καρδι-α* > *καρδιά*, *ψυχ-η* > *ψυχή*, πρέπει να σχηματίσουμε τους κανόνες από μηδενική βάση.

Πριν από οποιοδήποτε άλλο στάδιο μορφολογικής επεξεργασίας, προέχει να δημιουργηθούν οι κανόνες αλλομορφικής μετατροπής ή αλλιώς αλλο-κανόνες (παρόμοια πρόταση από τους Handl *et al.* 2010). Οι Handl *et al.* (2010) σημειώνουν ότι οι αλλο-κανόνες δημιουργούν ένα αλλομορφικό λεξικό από το βασικό ληματικό λεξικό και εφαρμόζονται πάντα πριν από οποιαδήποτε διαδικασία σχηματισμού λέξεων. Η εφαρμογή των αλλο-κανόνων πριν από τη διαδικασία της παραγωγής παρακάμπτει τα προβλήματα που μπορεί να προκύψουν με την ταυτόχρονη εφαρμογή φωνολογικών κανόνων στα όρια μορφημάτων, όπως μετακίνηση τόνου και φωνολογικά φαινόμενα αλλαγής ή τις δυσκολίες που είχε η τακτική των Krieger, Pinker

---

<sup>10</sup> Εφ' εξής οι υπολογιστικοί κανόνες για την επεξεργασία της αλλομορφίας αναφέρονται ως κανόνες αλλομορφικής μετατροπής και δεν αντιστοιχούν σε γλωσσολογικούς κανόνες.

& Nerbonne (1993). Για τα Ελληνικά, θεωρούμε πως είναι λειτουργικότερο το αλλομορφικό λεξικό να περιέχει όλες τις μορφές του μορφήματος. Ταυτόχρονα οι αλλομορφικοί κανόνες μετατροπής δεν δημιουργούν απλώς τα αλλόμορφα, αλλά κατά την καταχώρησή τους στο αλλομορφικό λεξικό προσκολλούν και ένα χαρακτηριστικό, το ALLOM $X$ , όπου  $X$  η αρίθμηση του αλλομόρφου με 1 πάντα το αρχικό μόρφημα.

- (94.) <WORD καρδιά> => <ALLOM1 ΚΑΡΔΙΑ>, <ALLOM2 ΚΑΡΔΙ>  
 <WORD παπάς> => <ALLOM1 ΠΑΠΑ>, <ALLOM2 ΠΑΠΑΔ>  
 <WORD άνθρωπος> => <ALLOM1 ΑΝΘΡΩΠ>  
 <WORD αγαπώ> => <ALLOM1 ΑΓΑΠ[Α]>, <ALLOM2 ΑΓΑΠΗ>  
 <WORD βρέχω> => <ALLOM1 ΒΡΕΧ>, <ALLOM2 ΒΡΑΧ>

Το χαρακτηριστικό ALLOM είναι απαραίτητο για τη διαδικασία δημιουργίας νέων λέξεων. Κατά την έξοδο ενός αλλομόρφου από το αλλομορφικό λεξικό θα περνάει από ένα φίλτρο που θα του επιτρέπει ή απαγορεύει να συνεχίσει στην ακόλουθη διαδικασία σχηματισμού λέξεων, όπως θα δούμε στην ενότητα 6.2.1.4. Όπως αποδείξαμε με τον περιορισμό της επιλογής μοναδικού αλλομόρφου, μόνο ένα αλλόμορφο συμμετέχει στην παραγωγή και ο αποκλεισμός των υπολοίπων επιτυγχάνεται με τη χρήση φίλτρου μορφολογικού περιβάλλοντος για την αποφυγή αντιγραμματικών λέξεων, όπως π.χ. \**κυμαίζω*, \**χεριούκλα*. Η αλλομορφία στα Ελληνικά συμμετέχει σε όλες τις διαδικασίες σχηματισμού λέξεων και είναι απαραίτητη η κωδικοποίηση αυτής της μορφολογικής πληροφορίας του κάθε μορφήματος. Αυτό μπορεί να γίνει με δύο τρόπους: (α.) εσωτερικά στο αλλομορφικό λεξικό και (β.) εκτός λεξικού και πριν την όποια διαδικασία σχηματισμού λέξης.

Η λύση της εσωτερικής καταχώρησης στο αλλομορφικό λεξικό είναι σύμφωνη με την αρχική μας προσέγγιση των μορφολεξικών κανόνων στο Λεξικό. Για παράδειγμα, το λήμμα ΠΑΠΑΣ έχει δύο αλλόμορφα, το *παπα~* και το *παπαδ~*, όπου τα αλλόμορφα πρέπει να καταχωρηθούν με τις πληροφορίες του μορφολογικού περιβάλλοντος, όπως στο ακόλουθο σχήμα:

- (95.) <Allom1 ΠΑΠΑ> = 
$$\left( \begin{array}{l} \text{ΚΛ\_ΕΝΙΚΟΣ} \\ \text{ΠΡΠΡΓ\_ΕΝΙΚΟΣ} \\ \text{ΣΥΝΘ\_ΒΛΕΞ\_ΕΝΙΚΟΣ} \end{array} \right)$$

όπου  $ΚΛ\_Ενικός =$  Ενικός Κλίσης,  $ΠρΠΡΓ\_Ενικός =$  Ενικός Προθηματοποιημένου Παράγωγου και  $ΣΥΝΘ\_ΒΛΕΞ\_Ενικός =$  Ενικός Σύνθετου με Β' Συστατικό ΛΕΞΗ<sup>11</sup>

$$\langle \text{Allom2 ΠΑΠΙΑΔ} \rangle = \left( \begin{array}{l} ΚΛ\_ΠΛΗΘΥΝΤΙΚΟΣ \\ ΠΡΠΡΓ\_ΠΛΗΘΥΝΤΙΚΟΣ, \\ ΠΡΓ, \\ ΣΥΝΘ\_Α, \\ ΣΥΝΘ\_ΒΘΕΜ, \\ ΣΥΝΘ\_ΒΛΕΞ\_ΠΛΗΘΥΝΤΙΚΟΣ \end{array} \right)$$

όπου  $ΚΛ\_Πληθυντικός =$  Πληθυντικός Κλίσης,  $ΠρΠΡΓ\_Πληθυντικός =$  Πληθυντικός Προθηματοποιημένου Παράγωγου,  $ΠΡΓ =$  Επιθηματοποιημένο Παράγωγο,  $ΣΥΝΘ\_Α =$  Σύνθετο με Α' Συστατικό,  $ΣΥΝΘ\_ΒΘΕΜ =$  Σύνθετο με Β' Συστατικό ΘΕΜΑ και  $ΣΥΝΘ\_ΒΛΕΞ\_Πληθυντικός =$  Πληθυντικός με Β' Συστατικό ΛΕΞΗ.

Η παραπάνω λύση σίγουρα δεν είναι οικονομική για ένα υπολογιστικό σύστημα, καθιστώντας το λεξικό του μορφολογικού αναλυτή ιδιαίτερα πολύπλοκο με κάθε λήμμα να φέρει όλες τις παραπάνω πληροφορίες. Αντιθέτως, η λύση της απεμπλοκής των πληροφοριών μορφολογικού περιβάλλοντος των αλλομορφων από το λεξικό αξιοποιεί τα δεδομένα από τον περιορισμό επιλογής μοναδικού αλλομόρφου και επεκτείνει τις δυνατότητές του, όπως θα δούμε στην αντίστοιχη υποενότητα. Με αυτή την πρόταση μπορούμε να δημιουργήσουμε ένα φίλτρο που λειτουργεί εκτός ηλεκτρονικού ληματικού λεξικού και δίνει τις πληροφορίες περιβάλλοντος εγκαίρως, πριν από κάθε διαδικασία μορφολογικού σχηματισμού.

Προτού δείξουμε πως θα λειτουργήσει το συγκεκριμένο φίλτρο στο μοντέλο, θα δοθεί το σύνολο κανόνων αλλομορφικής μετατροπής για την ονοματική παραγωγή, τόσο για τα θέματα όσο και για τα επιθήματα, καθώς και το αντίστοιχο σύνολο για τις αλλομορφικές μετατροπές στη ρηματική, επιθετική και προθηματική παραγωγή.

### 6.3.1.1. Ονοματικοί Κανόνες Αλλομορφικής Μετατροπής

Οι ακόλουθοι ονοματικοί κανόνες αλλομορφικής μετατροπής είναι υπολογιστικής φύσεως και δημιουργούν αλλομορφα από κάποια άλλα του ίδιου μορφήματος και έχουν καθολική εφαρμογή σε ονοματικά θέματα και επιθήματα. Οι κανόνες αντιστοιχούν

<sup>11</sup> Σχετικά με τα συστατικά και δυνατούς συνδυασμούς των συνθέτων στα Ελληνικά, βλ. Ράλλη (2005, 2008).

στην αλλομορφία που εμφανίζουν οι κλιτικές τάξεις σύμφωνα με τη θεωρία της Ralli (2005) και τις τροποποιήσεις της, όπως αυτές αναφέρθηκαν στο θεωρητικό κομμάτι της παρούσας διατριβής.

(96.) α. Μηδενική Αλλομορφία

Θέμα, Επίθημα → Μηδενική αλλομορφία

{ΑΝΘΡΩΠ → Μηδενική αλλομορφία}, {ΕΙ(ο) → Μηδενική αλλομορφία}

:: κανένας κανόνας ::

β. Α-Διαγραφή Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφο τύπου Α-Διαγραφή

{ΤΑΜΙΑ → ΤΑΜΙ}, {ΟΤΗΤΑ → ΟΤΗΤ}

::  $\alpha : \emptyset \Leftrightarrow \{\text{stem, suffix}\} \alpha : \emptyset \_ \_ [\text{filter: morpho-context::Allom2}] ::$

γ. Η-Διαγραφή Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφο τύπου Η-Διαγραφή

{ΨΥΧΗ → ΨΥΧ}, {ΤΗ → Τ}

::  $\eta : \emptyset \Leftrightarrow \{\text{stem, suffix}\} \eta : \emptyset \_ \_ [\text{filter: morpho-context::Allom2}]$

δ. Ι-Διαγραφή Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφο τύπου Ι-Διαγραφή

{ΠΑΙΔΙ → ΠΑΙΔ}, {ΤΗΡΙ → ΤΗΡ}

::  $\iota : \emptyset \Leftrightarrow \{\text{stem, suffix}\} \iota : \emptyset \_ \_ [\text{filter: morpho-context::Allom2}]$

ε. Δ-Εισαγωγή Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφο τύπου Δ-Εισαγωγή

{ΠΑΠΑ → ΠΑΠΑΔ}, {Α → ΑΔ}

::  $+\delta : \delta \Leftrightarrow \{\text{stem, suffix}\} +\delta : \delta \_ \_ [\text{filter: morpho-context::Allom2}]$

στ. Τ-Εισαγωγή Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφο τύπου Τ-Εισαγωγή

{ΚΥΜΑ → ΚΥΜΑΤ}, {ΜΑ → ΜΑΤ}

::  $+\tau : \tau \Leftrightarrow \{\text{stem, suffix}\} +\tau : \tau \_ \_ [\text{filter: morpho-context::Allom2}]$

ζ. Διπλή Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφο τύπου Η-Διαγραφή & Αλλαγή Φωνήεντος

{ΠΟΛΗ → ΠΟΛΕ ~ ΠΟΛ}, {ΣΗ → ΣΕ ~ Σ}

::  $\eta : \emptyset \Leftrightarrow \{\text{stem, suffix}\} \eta : \emptyset \_ \_ [\text{filter: morpho-context::Allom2}]$

:: η : ε ⇔ {stem, suffix} η: ε \_\_ [filter: morpho-context::Allom3]

η. Συνδυαστική Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφα τύπου Δ-Εισαγωγή & Η-Διαγραφή

{ΜΠΑΚΑΛΗ → ΜΠΑΚΑΛΗΔ~ ΜΠΑΚΑΛ}, {ΑΡΗ → ΑΡΗΔ ~ ΑΡ}

:: + : δ ⇔ {stem, suffix} + : δ \_\_ [filter: morpho-context::Allom3]

:: η : ∅ ⇔ {stem, suffix} η : ∅ \_\_ [filter: morpho-context::Allom2]

Η συγκεκριμένη ομάδα μορφημάτων παρουσιάζει μία συγκεκριμένη και συστηματική εμφάνιση αλλομόρφων, καθιστώντας τη δημιουργία των κανόνων για ένα μορφολογικό αναλυτή αρκετά εύκολη υπόθεση. Εφαρμόζονται πάντα πριν συμμετάσχει το μόρφημα σε κάποια διαδικασία, και πιο ειδικά στο σχηματισμό παράγωγων λέξεων. Κανόνες όπως αυτοί του (96.) μπορούν απευθείας να μετατραπούν για τους αποδέκτες των αυτομάτων πεπερασμένων καταστάσεων πάνω σε ακολουθίες ζευγών συμβόλων.

Κάνοντας χρήση των κανονικών εκφράσεων (regular expressions) ως σχεδιάτυπα (patterns) μπορούμε να μετατρέψουμε σε κωδικοποιημένο κανόνα καθένα από τους παραπάνω κανόνες για ένα οποιοδήποτε εποπτευόμενο μοντέλο μορφολογικής μάθησης, αλλά και για μοντέλο διεπίπεδης μορφολογίας. Στο προτεινόμενο πλαίσιο κάθε κωδικοποιημένος κανόνας χωρίζεται σε δύο μέρη, όπου κάθε μέρος διαθέτει κάποιο καθορισμένο περιεχόμενο και διαχωρίζονται μεταξύ τους με το ⇔. Στο αριστερό μέρος περιγράφεται ο τύπος της αλλομορφικής μετατροπής, π.χ. [+ : τ], ενώ το σύμβολο ⇔ υποδηλώνει ότι ο κανόνας εφαρμόζεται πάντα. Στο δεξί κομμάτι, το πρώτο μέρος της κωδικοποίησης είναι αναγκαίο, καθότι υποχρεώνει το μόρφημα σε κάποια αλλαγή της μορφής του, η οποία γίνεται χωρίς καμία εξαίρεση. Πιο συγκεκριμένα, το τμήμα [{stem, suffix} + : τ] του κανόνα (96.στ) φανερώνει την απαραίτητη προσθήκη ενός φωνήματος /t/ στο τέλος του μορφήματος. Παράλληλα το υπόλοιπο σκέλος του κανόνα καθορίζει τον τύπο του αλλομόρφου, αλλά και το μορφολογικό περιβάλλον στο οποίο επιτρέπεται να συμμετέχει περνώντας πάντα από τη διαδικασία του φιλτραρίσματος για την αποφυγή αντιγραμματικών λέξεων. Για παράδειγμα, το κομμάτι [filter: morpho-context::ALLOM2] δηλώνει ότι το συγκεκριμένο αλλόμορφο (κυματ) φέρει το χαρακτηρισμό ALLOM2 και το μορφολογικό περιβάλλον εμφάνισης καθορίζεται από τις πληροφορίες που φέρει το morpho-context::ALLOM2 στο φίλτρο.

Μόνο δύο κανόνες αλλομορφικού μετασχηματισμού των ονοματικών θεμάτων και επιθημάτων παράγουν τρία αλλόμορφα, αυτά του τύπου *πολη~ πολε~ πολ* και *μπακαλη~ μπακαληδ~ μπακαλ~*. Ο εκάστοτε κανόνας (96.η-ζ) αντιστοιχεί την κατάλληλη ετικέτα ALLOM στο αλλόμορφο, ώστε να μπορεί αργότερα να φιλτραριστεί σωστά από το σύστημα στη διαδικασία επιλογής αλλομόρφου. Στην ανωτέρω κωδικοποίηση χαρακτηρίζουμε ως ALLOM1 το αρχικό μόρφημα και το αλλόμορφο ως ALLOM2 που συμμετέχει στην παραγωγική διαδικασία. Σύμφωνα με τον περιορισμό της επιλογής μοναδικού αλλομόρφου όλα τα ονοματικά θέματα και επιθήματα έχουν πάντα μόνο ένα και συγκεκριμένο αλλόμορφο που αποτελεί μέλος της παραγωγής. Αυτό απλοποιεί την υλοποίηση του υπολογιστικού αλλομορφικού μοντέλου της ονοματικής παραγωγής, καθότι πραγματοποιείται εύκολα η ομαδοποίηση των θεμάτων και επιθημάτων, καθώς και η ανάδειξη του κατάλληλου αλλομόρφου (για περισσότερα βλ. 3.2.2). Στους κανόνες (96.) ορίζεται ως αλλόμορφο παραγωγής το ALLOM2 και είναι το αλλόμορφο που απαντάται στον πληθυντικό. Μοναδική εξαίρεση η όγδοη αλλομορφική τάξη που για την παραγωγή και τη σύνθεση παρουσιάζει διαφορετικό αλλόμορφο και το νέο αλλόμορφο κωδικοποιείται ως ALLOM2 για το σύστημα και το φίλτρο.

### 6.3.1.2. Ρηματικοί Κανόνες Αλλομορφικής Μετατροπής

Οι ρηματικοί κανόνες αλλομορφικής μετατροπής είναι υπολογιστικής φύσεως και δημιουργούν αλλόμορφα από κάποια άλλα του ίδιου μορφήματος και έχουν καθολική εφαρμογή σε ρηματικά θέματα και επιθήματα. Οι κανόνες αντιστοιχούν εν μέρει στην αλλομορφία που εμφανίζουν οι κλιτικές τάξεις σύμφωνα με τη θεωρία της Ralli (2005). Ωστόσο λόγω της έντονης προβληματικότητας που παρουσιάζουν τα ρήματα της PKT1 δίνονται πολλοί και διαφορετικοί υπολογιστικοί κανόνες, ώστε να καλυφθούν επαρκώς οι αλλομορφικές αλλαγές.

(97.) α. Μηδενική Αλλομορφία

Θέμα, Επίθημα → Μηδενική αλλομορφία

{ΙΔΡΥ → Μηδενική αλλομορφία}, {ΕΥ → Μηδενική αλλομορφία}

:: κανένας κανόνας ::

β. Η-Εισαγωγή Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφο τύπου η-Εισαγωγή

{ΑΓΑΠ → ΑΓΑΠΗ}, {ΑΡ → ΑΡΗ}

:: + : η ⇔ {stem, suffix} + : η \_\_\_ [filter: morpho-context::Allom2] ::

γ. Ε-Εισαγωγή Αλλομορφία

Θέμα, ~~Επίθημα~~ → Αλλόμορφο τύπου Ε-Εισαγωγή

{ΑΦΑΙΡ → ΑΦΑΙΡΕ}

:: + : ε ⇔ {stem, suffix} + : ε \_\_\_ [filter: morpho-context::Allom2] ::

δ. ΑΙ-Αντικατάσταση Αλλομορφία 1

Θέμα, Επίθημα → Αλλόμορφο τύπου ΑΙ-Αντικατάσταση

{ΒΑΣΚΑΙΝ → ΒΑΣΚΑΝ}, {ΑΙΝ → ΑΝ}

:: αι : α ⇔ {stem, suffix} 1αι : α \_\_\_ [filter: morpho-context::Allom2] ::

ε. ΑΙ-Αντικατάσταση Αλλομορφία 2

Θέμα, Επίθημα → Αλλόμορφο τύπου ΑΙ-Αντικατάσταση

{ΒΑΘΑΙΝ → ΒΑΘΥΝ}, {ΑΙΝ → ΥΝ}

:: αι : υ ⇔ {stem, suffix} 1αι : υ \_\_\_ [filter: morpho-context::Allom2] ::

στ. Ζ-Αντικατάσταση Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφο τύπου Ζ-Αντικατάσταση

{ΑΓΓΙΖ → ΑΓΓΙΓ}, {ΙΖ → ΑΓ}

:: ζ : γ ⇔ {stem, suffix} ζ : γ \_\_\_ [filter: morpho-context::Allom2] ::

ζ. Ν-Διαγραφή Αλλομορφία

Θέμα, Επίθημα → Αλλόμορφο τύπου Ν-Διαγραφή

{ΧΑΝ → ΧΑ}, {ΩΝ → Ω}

:: ν : ∅ ⇔ {stem, suffix} ν : ∅ \_\_\_ [filter: morpho-context::Allom2] ::

η. ΑΓ-Εισαγωγή Αλλομορφία

Θέμα, ~~Επίθημα~~ → Αλλόμορφο τύπου ΑΓ-Εισαγωγή

{ΠΕΤ → ΠΕΤΑΓ}

:: + : αγ ⇔ {stem, suffix} + : αγ \_\_\_ [filter: morpho-context::Allom2] ::

θ. ΗΓ-Εισαγωγή Αλλομορφία

Θέμα, ~~Επίθημα~~ → Αλλόμορφο τύπου ΗΓ-Εισαγωγή

{ΡΟΥΦ → ΡΟΥΦΗΓ}

:: + : ηγ ⇔ {stem, suffix} + : ηγ \_\_\_ [filter: morpho-context::Allom2] ::

ι. Μεμονωμένη Αλλομορφία

Θέμα, ~~Επίθημα~~ → Αλλόμορφα μεμονωμένου τύπου

{ΠΠΝ → ΗΠΠ ~ ΠΠ ~ Π}

:: ειδικός κανόνας // προκαταχώρηση ::

Ξεκινώντας την υπολογιστική επεξεργασία της αλλομορφίας των ρηματικών θεμάτων και επιθημάτων, η ανάλυση αρχίζει από εκείνη της δεύτερης κλιτικής τάξης (PKT2) που έχει αλλομορφία με την προσθήκη του φωνήεντος /η/, /α/ και /ε/ (97.β-γ). Όπως αναφέρθηκε από τη Ράλλη (2005), η συστηματική ρηματική αλλομορφία εντοπίζεται στη δεύτερη κλιτική τάξη στα πρώην συνηρημένα ρήματα. Τα αλλόμορφα των συγκεκριμένων ρηματικών θεμάτων δημιουργούνται με την προσθήκη φωνήεντος στο αρχικό μόρφημα. Οι περιπτώσεις συστηματικής αλλομορφίας είναι πάρα πολύ συχνές και προσφέρουν τα ίδια ακριβώς πλεονεκτήματα της ονοματικής αλλομορφίας: υπάγονται στον περιορισμό επιλογής μοναδικού αλλομόρφου και επομένως περιμένουμε το αλλόμορφο των συνοπτικών θεμάτων να είναι το μοναδικό που θα εμφανιστεί στην παραγωγή που έχει ως βάση ρηματικό θέμα.

Τα ρήματα της PKT2 φαίνεται να χρησιμοποιούν το ίδιο αλλόμορφο στην παραγωγή, όπως *αγαπη-τικός, αγαπη-τός, αγαπημένος* και *αφαιρέ-της, διαιρέ-της, διαιρε-μένος*. Όντας πρώην συνηρημένα ρήματα, αλλά με διαφοροποιημένο αοριστικό θέμα, οι αλλομορφικές τάξεις PAT8 (97.η) και PAT9 (97.θ) παραμένουν πιστές στις τάξεις της γενικότερης κατηγορίας και χρησιμοποιούν το ίδιο αλλόμορφο στις παράγωγες λέξεις, όπως *τράβηγ-μα, τραβηγ-μένος, τραβηγ<sup>12</sup>-τικός | φώναγ-μα, φωναχ<sup>13</sup>-ταράς | χαλασ-μένος, χαλάσ-τρα*. Αρκετοί συγγραφείς γραμματικών (βλ. Τσοπανάκης, Τριανταφυλλίδης, Κλαίρης & Μπαμπινιώτης) ασχολούνται με τη μορφή των συγκεκριμένων ρημάτων και με την προβληματική του τελικού συμφώνου. Θεωρούν την καταφυγή σε ιστορικές γραμματικές ή σε γραμματικές των Αρχαίων Ελληνικών ως την ασφαλέστερη οδό, όπου ο εντοπισμός είναι πιο εύκολος λόγω της ύπαρξης φωνολογικών κανόνων που δικαιολογούν αυτές τις αλλαγές (βλ. Οικονόμου 1971, Βλάχου 1994). Εντούτοις, συγχρονικά στην υπολογιστική μας πρόταση επιλέγουμε ως βασικό σύμφωνο τον τελικό θέματος χαρακτήρα που εμφανίζεται στα περισσότερα περιβάλλοντα παραγωγής.

<sup>12</sup> Η μετατροπή του /γ/ σε /x/ γίνεται λόγω του φωνολογικού κανόνα της αφομοίωσης, όπου χάνεται το χαρακτηριστικό της ηχηρότητας.

<sup>13</sup> Βλ. προηγούμενη υποσημείωση.



Είναι ιδιαίτερα συνήθης η αλλαγή στο θεματικό φωνήεν του ρήματος που συχνά ακολουθείται και από μία επιπλέον μεταβολή του θέματος, καθιστώντας την αλλομορφία του ιδιαίτερα πολύπλοκη. Όπως σημειώσαμε σε παλαιότερη έρευνα (Καρασίμος 2001), παρατηρείται ότι σε κάποια ρήματα το τελευταίο φωνήεν του θέματος αλλάζει σε κάποιο άλλο χωρίς η αλλαγή αυτή να είναι συστηματική ως προς το φωνήεν, δηλαδή π.χ. το /e/ να αλλάζει πάντα σε κάποιο συγκεκριμένο φωνήεν. Ακολούθως σημειώνει ότι υφίστανται και ρηματικά θέματα με πολλαπλή αλλομορφία, τα οποία αλλάζουν απρόβλεπτα και χωρίς να είναι εφικτή η ομαδοποίηση κάποιων αλλομορφικών αλλαγών. Στα παραδείγματα που έπονται δίνεται ένα σημαντικό δείγμα από τα διάφορα ρήματα της πρώτης κλιτικής τάξης (PKT1).

(98.)	α.	φεύγω:	φευγ~ φυγ	(έ-φυγ-α)
		βρέχω:	βρεχ~ βραχ	(βράχ-ηκ-α)
		χαίρω:	χαιρ~ χαρ	(χάρ-ηκ-α)
		φθείρω:	φθειρ~ φθαρ	(φθάρ-θηκ-α)
	β.	βάλλω:	βαλ(λ)~ βλη	([προσ]βλή-θηκ-α)
		γέρνω:	γερν~ γειρ	(έ-γειρ-α)
		γδέρνω:	γδέρν~ γδαρ	(γδάρ-θηκ-α)
		ελαύνω:	ελαυν~ ελα	([προσ]έλα-σ-α)
		δίνω <sup>14</sup> :	διν~ δο/ω	(έ-δω-σ-α)
		καλώ:	καλ~ καλε~ κλη	(κάλε-σ-α, κλή-θηκ-α)
		βρίσκω:	βρισκ~ βρ~ βρε	(βρ-ήκα, βρέ-θηκ-α)
		άγω:	αγ~ αγαγ	(προσ-αγάγ-ω)
		μαθαίνω:	μαθαιν~ μαθ~ μαθη	(έ-μαθ-α, μαθη-μένος)
		τυχαίνω:	τυχαιν~ τυχ	(έ-τυχ-α)
		πηγαίνω:	πηγαιν~ πηγ~ πα	(πήγ-α, πά-ω)
		λαμβάνω <sup>15</sup> :	λαμβαν~ λαβ~ ληφ	(έ-λαβ-α, ε-λήφθ-η, ειλημ-μένος)

Είναι σαφές ότι σχεδόν για όλα τα ρήματα χρειάζεται ξεχωριστός κανόνας περιγραφής της σχέσης ανάμεσα στα δύο – ή και τρία σε λιγιστές περιπτώσεις –

<sup>14</sup> Όταν είναι σύνθετο τότε τα αλλόμορφα έχουν την μορφή +δίδω: δίδω~ δο~ δος

<sup>15</sup> Το ρήμα *λαμβάνω* εκτός από την ελεύθερη εναλλαγή στο ενεστωτικό θέμα (*λαβαίνω*) για τον σχηματισμό της παθητικής φωνής στο μη συνεχές δημιουργεί αλλόμορφο σύμφωνα με την παρατήρηση (βλ. παραπάνω, η μορφή [liv]).

αλλόμορφα. Είναι αντιοικονομικό για το σύστημα να εισαχθούν ξεχωριστοί κανόνες για αυτές τις περιπτώσεις που στην πλειονότητα είναι άπαξ λεγόμενα. Κρίνεται βολικό για ένα εποπτευόμενο σύστημα να γίνει προεισαγωγή των συγκεκριμένων ρηματικών μορφών με την απαραίτητη κωδικοποίηση ως ένα σύνολο που θα προηγείται του ελέγχου αλλομορφίας. Το σύστημα κατά την ανάλυση ενός ρήματος, θα ελέγχει αρχικά εάν εμπίπτει στο συγκεκριμένο σύνολο ρημάτων και στη συνέχεια θα δοκιμάζει να εφαρμόσει κάποιον άλλο κανόνα. Η κωδικοποίηση των συγκεκριμένων μορφών αποσκοπεί στην απρόσκοπτη ανάλυση των προθηματοποιημένων παραγώγων των ανωτέρω ρημάτων, ενώ για τις περιπτώσεις επιθηματοποίησης θα υπάρχει η απαραίτητη ένδειξη για την επιλογή του κατάλληλου τύπου.

Σύμφωνα με τους Jurafsky & Martin (1999) σε αρκετά συστήματα προκρίνεται η λογική της *εξαγωγής από δεξαμενή* (pool extraction), όπου καταχωρούνται όλες οι προβληματικές περιπτώσεις ρημάτων με ασυνήθιστες μορφηματικές αλλαγές. Το σύστημα ξεκινάει αρχικά την ανάλυση ελέγχοντας τη δεξαμενή των συγκεκριμένων μορφημάτων, προτού συνεχίσει το πέρασμα ελέγχου από κανονικές μορφές τύπων και κανόνων που δεν εμφανίζουν αλλομορφία και σχηματίζουν ομαλώς τους τύπους. Αυτό βέβαια προκαλεί συμφόρηση στο σύστημα, μίας και δεν υφίσταται η παρουσία κανόνων αλλομορφικής μετατροπής και όλα τα αλλόμορφα παρατίθενται μέσα σε ένα σύνολο, τα οποία εισήχθησαν χειρωνακτικώς. Αξιοποιώντας κεντρικά σημεία της τακτικής της εξαγωγής από δεξαμενή εισάγουμε τη χρήση της περιορισμένης δεξαμενής αλλομόρφων με την κατάλληλη κωδικοποίηση στο παρόν προτεινόμενο μοντέλο. Με τη συγκεκριμένη στρατηγική η δεξαμενή θα περιέχει περί τα είκοσι ρήματα με τα αλλόμορφα τους χαρακτηρισμένα, ώστε να ελέγχονται από το εκάστοτε σύστημα, προτού γίνει η μορφολογική ανάλυση ή δημιουργία λέξεων. Ο αριθμός είναι περιορισμός και η εύρυθμη λειτουργία του συστήματος δεν επιβραδύνεται από μία πολύ περιορισμένη δεξαμενή αλλομόρφων.

Επιπροσθέτως υπάρχει και το αλλομορφικό ζήτημα της παραγωγής με ετεροίωση που εφαρμόζεται σε αρκετά ρηματικά θέματα και που απουσιάζει παντελώς από τα ρηματικά επιθήματα. Έχουμε κάποιες περιπτώσεις, όπως *αλείφω > αλοιφή*, *αμείβω > αμοιβή*, *φέρω > φορά*, *κλέβω > κλοπή*, *βάλλω > βολή*. Εδώ δεν αποτελεί πρόβλημα η ετεροίωση, η οποία είναι μία ανενεργή φωνολογική διαδικασία στα Νέα Ελληνικά και επομένως θεωρούνται αλλόμορφα, αλλά η επιλογή διαφορετικού αλλομόρφου για τις τυχόν υπόλοιπες παράγωγες λέξεις (πρβλ. Κίτσα 2001). Σε αντίθεση με τις άλλες

κατηγορίες των ρημάτων, εδώ χρησιμοποιείται το ενεστωτικό θέμα κατά τη διαδικασία παραγωγής. Τα παράγωγα με ετεροίωση φέρουν ένα σημαντικό χαρακτηριστικό, το οποίο μας επιτρέπει να αποφύγουμε τον σκόπελο που ξεπροβάλλει με την αλλαγή του θεματικού φωνήεντος: *όλα τα παράγωγα τέτοιας μορφής δεν ακολουθούνται από παραγωγικά επιθήματα*, καθότι όταν δημιουργήθηκαν η ετεροίωση αποτελούσε μία μορφοφωнологική διαδικασία παραγωγής. Επομένως, κατά την ανάλυση των ετεροιωμένων αλλομόρφων έχουμε τη δυνατότητα να παραπέμπουμε τον αναλυτή στη συγκεκριμένη ομάδα αλλομόρφων από τη στιγμή που θα παρατηρείται η απουσία του επιθήματος παραγωγής. Κατά την υπολογιστική δημιουργία των συγκεκριμένων τύπων, θα απαγορεύεται η παρουσία παραγωγικού επιθήματος και θα αποτρέπονται με αυτόν τον τρόπο οι αντιγραμματικές λέξεις. Παράλληλα κατά τη μορφολογική ανάλυση μετά τον εντοπισμό της απουσίας του παραγωγικού, θα γίνεται η αναζήτηση για αλλαγές φωνήεντος με εντοπισμό της θέσης του με τον τύπο 2X/3X έχοντας ως βάση το αρχικό ρηματικό μόρφημα. Για παράδειγμα για το φέρω > φορά, ο κανόνας θα σχηματιστεί ως [stem] 2ε : ο \_\_ [inflectional suffix], όπου ο αριθμός καθορίζει πόσο απέχει το φωνήεν από το όριο του μορφήματος.

Κλείνοντας την υποενότητα των αλλομορφικών μετασχηματιστικών κανόνων για ρηματικά θέματα και επιθήματα, θεωρούμε ότι κατέστησαν σαφείς η πολυπλοκότητα και προβληματικότητα της κατάστασης, καθότι είναι ξεκάθαρο ότι κληρονομούνται από την κλίση διάφορες αλλαγές που έχουν σημειωθεί κατά την εξέλιξη του ρηματικού συστήματος της γλώσσας. Οι αλλαγές δεν είναι τόσο τακτικές και συγκεκριμένες σε όλα τα ρήματα που ανήκουν στη μη-συστηματική αλλομορφία και έχει ως συνέπεια αρκετές περιπτώσεις να πλησιάζουν πολύ τη φιλοσοφία των *άπαξ λεγομένων*.

### 6.3.1.3. Προθηματικοί Κανόνες Αλλομορφικής Μετατροπής

Οι ακόλουθοι προθηματικοί κανόνες αλλομορφικής μετατροπής είναι υπολογιστικής φύσεως και δημιουργούν αλλομορφα για την ιδιαίτερα περιορισμένη κλειστή ομάδα των παραγωγικών προθημάτων. Οι κανόνες ανταποκρίνονται στην ιδιαίτερα περιορισμένη αλλομορφική συμπεριφορά των προθημάτων που σε όλες τις περιπτώσεις είναι υπολογιστική υλοποίηση ανενεργών φωнологικών κανόνων. Εισάγονται επίσης και κανόνες αντιμετώπισης των κατάλοιπων μετοχών παρακειμένου.

(99.) α. Δασυνόμενη Αλλομορφία 1

Πρόθημα → Αλλομορφία τύπου T-Αντικατάσταση

{ANTI → ANΘ}

:: τV : θ ⇔ {prefix} τV : θ \_\_ [V<sup>h</sup>μόρφημα, πρόθημα:: Allom2]:<sup>16</sup>

β. Δασυνόμενη Αλλομορφία 2

Πρόθημα → Αλλομορφία τύπου Π-Αντικατάσταση

{EΠI → EΦ}

:: πV : φ ⇔ {prefix} πV : φ \_\_ [V<sup>h</sup>μόρφημα, πρόθημα:: Allom2]::

γ. K-Αντικατάσταση Αλλομορφία

Πρόθημα → Αλλομορφία τύπου K-Αντικατάσταση

{EK → EΞ}

:: κ : ξ ⇔ {prefix} κ : ξ \_\_ [Vθέμα, πρόθημα:: Allom2]::

# δ. Αναδιπλασιαστική Αλλομορφία<sup>17</sup>

‘Πρόθημα’ Αναδιπλασιασμού → Αλλομορφία τύπου Αναδιπλασιασμού

{TEΛ → TETEΛEΣ}

:: + : Cε ⇔ + : Cε \_\_ [Cθέμα+μεν:: Allom2]::

# ε. Αποδασυνόμενη Αναδιπλασιαστική Αλλομορφία

‘Πρόθημα’ Αναδιπλασιασμού → Αλλομορφία τύπου Αναδιπλασιασμού

{ΦΩΤIΖ → ΠEΦΩΤIΣ}

:: + : κε ⇔ + : κε \_\_ [xθέμα+μεν:: Allom2]::

:: + : τε ⇔ + : τε \_\_ [θθέμα+μεν:: Allom2]::

:: + : πε ⇔ + : πε \_\_ [φθέμα+μεν:: Allom2]::

Η προθηματοποίηση στα Ελληνικά δεν ενέχει ιδιαίτερα προβλήματα, καθότι η ομάδα των προθημάτων είναι πολύ μικρή και η αλλομορφία της πολύ συστηματική.

<sup>16</sup> Συμβατικά θεωρούμε ότι διαγράφουμε το τελικό φωνήεν (V) προτού γίνει η αντικατάσταση του /t/ με το /θ/. Διαφορετικά ο κανόνας θα έχει τη μορφή 2t : θ με την απαραίτητη προϋπόθεση ότι ο σχηματισμός του αλλομόρφου γίνεται πριν τη διαγραφή του φωνήεντος μέσω φωνολογικού κανόνα. Επίσης, συμβατικά χαρακτηρίζουμε κάποιο αρχικό φωνήεν ως δασυνόμενο (V<sup>h</sup>), καθότι δεν υφίσταται η δάσυνση συγχρονικά.

<sup>17</sup> Οι δύο τελευταίες κατηγορίες κανόνων αναφέρονται στις περιπτώσεις των παθητικών μετοχών παρακειμένων που υφίστανται στα Ελληνικά με τον αναδιπλασιασμό των ΑΕ. Επειδή οι μετοχές είναι ένα κομμάτι της μορφολογίας που βρίσκεται στα όρια παραγωγής και κλίσης, η ανάλυση των συγκεκριμένων μορφών είναι περίπλοκη και συνηθίζεται να αποφεύγεται ο όποιος διαχωρισμός των συγκεκριμένων μετοχών. Π.χ. στις παραδοσιακές γραμματικές υφίσταται η ανάλυση γε-γραμ-μεν-ος, ενώ για τις σύγχρονες μορφολογικές αναλύσεις γεγραμ(-)μεν-ος ή και γεγραμμεν-ος.

Στα Αρχαία Ελληνικά ο κανόνας της δάσυνσης εφαρμόζεται στα όρια μορφημάτων με την ύπαρξη δασυνόμενου φωνήεντος στην αρχή του δεύτερου μορφήματος. Μετά την αποβολή – σε περίπτωση παρουσίας του – τελικού φωνήεντος από το πρώτο μόρφημα, πολλές φορές η δάσυνση του φωνήεντος μεταφερόταν στο σύμφωνο που προηγείται και μπορεί να δασυνθεί. Ο φωνολογικός κανόνας αυτός πλέον δεν υφίσταται και επομένως αρκεί να τον κωδικοποιήσουμε υπολογιστικά, μίας και η συστηματικότητά του είναι δεδομένη (99.α-β). Τα δασυνόμενα σύμφωνα των {τ, π, κ} είναι {θ, φ, χ} αντίστοιχα με την τελευταία δυάδα {κ | χ} να απουσιάζει. Οι δυάδες αντικατάστασης είναι {τ | θ} και {π | φ} για τις περιπτώσεις προθηματοποιημένων λέξεων, όπως *ανθ-υπασπιστής* και *αφ-οπλίζω*. Να υπογραμμιστεί ότι απαιτείται και η σύμπραξη του φωνολογικού κανόνα της αποβολής πριν εφαρμοστεί ο αλλομορφικός κανόνας της δάσυνσης. Για την εφαρμογή του προθηματικού κανόνα αλλομορφικής μετατροπής προϋποτίθεται η αποβολή του τελικού φωνήεντος από το πρόθημα, όπως *επί+ησυχάζω* > *επ+ησυχάζω* (αποβολή φωνήεντος λόγω του αρκτικού φωνήεντος του ρήματος) > *εφ+ησυχάζω* (δασυνόμενη αλλομορφία) > *εφησυχάζω* και *αντί+έλληνας* > *αντ+έλληνας* (αποβολή φωνήεντος λόγω του εναρκτήριου φωνήεντος του ρήματος) > *ανθ+έλληνας* (δασυνόμενη αλλομορφία) > *ανθέλληνας*. Για την ορθή και αποτελεσματική εφαρμογή των κανόνων (99.α-β) οι λέξεις που έφεραν δασεία στα Αρχαία Ελληνικά πρέπει να φέρουν το χαρακτηριστικό V<sup>h</sup>Μόρφημα, ώστε να τεθεί σε εφαρμογή ο υπολογιστικός προθηματικός κανόνας για το σχηματισμό παραγώγων τύπου *αφαίμαξη*, *εφορμό* και *υφυπουργός* αντί των αντιγραμματικών τύπων *\*απαίμαξη*, *\*επορμό* και *\*υφυπουργός*. Το συγκεκριμένο χαρακτηριστικό το φέρουν θεματικά μορφήματα, καθώς και το πρόθημα *υπο-*, καθώς απαιτείται από το φίλτρο μορφολογικού περιβάλλοντος για την επιλογή του κατάλληλου αλλομόρφου.

Στην παρούσα υποενότητα θα εντάξουμε και μία σύντομη ανάλυση των παθητικών μετοχών λόγω των αλλομορφικών μετατροπών που συντελούνται στο αριστερό μέρος του θέματος και θεωρούμε ότι μπορούν να προσεγγιστούν υπολογιστικά με συναφή τρόπο των προθηματικών κανόνων. Οι παθητικές μετοχές παρακειμένου με αναδιπλασιασμό είναι ιδιαίτερα πολύπλοκες, χαρακτηρίζονται λόγιες μετοχές από τους Κλαίρη & Μπαμπινιώτη (1996) και τυγχάνουν λιγοστή προσοχή από τους ερευνητές (αναλυτικότερα βλ. Manolessou 2005). Θα αναφέρουμε επιγραμματικά μία υπολογιστική προσέγγιση για την αντιμετώπιση της επέκτασης του θέματος στις μετοχές, καθότι πρέπει να αντιμετωπιστεί υπολογιστικά η ιδιαίτερη αλλομορφία που

εμφανίζουν. Σε αυτές τις μετοχές έχουμε την προσθήκη μίας ακόμα συλλαβής στην αρχή του θέματος, ακόμα και αν προηγείται κάποιο πρόθημα.<sup>18</sup> Η νέα συλλαβή βασίζεται στην επανάληψη του αρχικού συμφώνου με την προσθήκη του φωνήματος /ε/, όπως για παράδειγμα οι μετοχές *δεδηλωμένος* (δηλώνω), *πεπιεσμένος* (πιέζω), *πεπαιδευμένος* (παιδεύω), *εγκεκριμένος* (εγκρίνω), *συνδεδεμένος* (συνδέω). Είναι εξαιρετικά πιθανή, όμως, και η εφαρμογή του νόμου του Grassmann – ανενεργός φωνολογικός κανόνας συγχρονικά – που αποτρέπει την εμφάνιση δύο δασυνόμενων συμφώνων στην ίδια λέξη αποδασύνοντας το πρώτο. Παραδείγματος χάριν, έχουμε *πεφωτισμένος* αντί \*φεφωτισμένος, *κεχαριτωμένη* αντι \*χεχαριτωμένη, *τεθωρακισμένος* αντι \*θεθωρακισμένος, *πεφωτισμένος* αντί \*φεφωτισμένος. Συγχρονικά όλες οι παραπάνω αλλαγές είναι μορφολογικά αδιαφανείς και επομένως δεχόμαστε ότι είναι αλλομορφα μετοχής των διαφόρων ρηματικών τύπων και το προτεινόμενο πλαίσιο διαθέτει συγκεκριμένους υπολογιστικούς κανόνες για τη δημιουργία των απαραίτητων αλλομόρφων.

Η ταυτόχρονη ακολουθία δύο κανόνων για τον αναδιπλασιασμό και τον νόμο Grassmann καθιστά την ανάλυση πιο δύσκολη, γιατί θα πρέπει να εφαρμοστούν με σειρά προτεραιότητας, δηλαδή πρώτα η προσθήκη της συλλαβής και έπειτα η πιθανή εφαρμογή της αλλαγής του αρχικού συμφώνου. Έτσι κρίνεται ορθότερο η χρήση δύο διαφορετικών υπολογιστικών κανόνων, όπου ο ένας θα περιλαμβάνει την απλή περίπτωση αναδιπλασιασμού (99.δ) και ο άλλος κανόνας θα συνδυάζει την ταυτόχρονη εφαρμογή των δύο ανενεργών φωνολογικών κανόνων χωρίς σειρά εκτέλεσης στο υπολογιστικό μοντέλο (99.ε). Καθίσταται προφανές ότι σε κάποια σημεία η υπολογιστική μορφολογία και το προτεινόμενο πλαίσιο επεξεργασίας καταστρατηγεί σε επίπεδο υπολογιστικών κανόνων ανάλυσης τους αντίστοιχους μορφολογικούς, ώστε να υπερκεράσει κάποια προβλήματα πολυπλοκότητας, αλλά με στόχο πάντα την ορθή μορφολογική ανάλυση στην επιφανειακή δομή. Σύμφωνα με το θεωρητικό πλαίσιο είτε εισάγουμε ως ένα μόρφημα με το κλιτικό επίθημα είτε χωρίζουμε τις μετοχές αναγνωρίζοντας το συστατικό *-μεν-*. Υπολογιστικά όμως δημιουργούμε κανόνα για να αναπαραστήσουμε τον αναδιπλασιασμό της συλλαβής και μάλιστα συνδυάζοντας δύο φαινόμενα σε έναν υπολογιστικό κανόνα (αναδιπλασιασμός και νόμος Grassmann). Έτσι ο υπολογιστικός κανόνας στοχεύει απλώς στη δημιουργία του αλλομόρφου (πε-

<sup>18</sup> Για το σχηματισμό των μετοχών παρακειμένου με αναδιπλασιασμό (γεγραμμένος) ή με αύξηση (ηγιασμένος) βλ. ενδεικτικά Debrunner (1917), Τζωρτζάτος (1967), Οικονόμου (1971).

φωτισ-μέν-ος) και όχι να ακολουθήσει τη σειρά σχηματισμού του μορφήματος (φε+φωτισ > πε+φωτισ > πεφωτισ). Υπάρχουν, όμως, μετοχές που έχουν λεξικοποιηθεί πλήρως ή υπάρχει πλήρης μορφολογική αδιαφάνεια και μη σύνδεση με το ρήμα ώστε το μόνο ορατό μόρφωμα πλην του θέματος να είναι το κλιτικό επίθημα, όπως *ειρημεν-ος*, *ειλημμέν-ος*, *συνημμέν-ος*, καθότι οι ομιλητές δεν μπορούν να τα εντάξουν συχρονικά στην παραγωγή από τα ρήματα *λέγω*, *λαμβάνω* και *συνάπτω* αντίστοιχα.

Ένα ακόμα κατάλοιπο φαινόμενο των Αρχαίων Ελληνικών είναι η χρονική αύξηση που παρ' όλες τις προβλέψεις και την ύπαρξη του κανόνα της τρισυλλαβίας για τον τόνο, εξακολουθεί και εφαρμόζεται σύμφωνα με τους κανόνες των ΑΕ σε προθηματοποιημένους ρηματικούς τύπους (βλ. περισσότερα Joseph & Janda 1988, Ralli 1988, Μαλικούτη – Drachman & Drachman 1992, 1993, Ράλλη 2005, Revithiadou & Spyropoulos 2005, Κουτσούκος 2010). Ανάλογα με το μορφολογικό πλαίσιο του καθενός, η προσέγγιση της χρονικής αύξησης είναι διαφορετική ως φαινόμενο. Σύμφωνα με τον Κουτσούκο (2010: 184) η χρονική αύξηση μπορεί να προβλεφθεί με γραμματικά κριτήρια με σημαντικό ρόλο στην εμφάνιση των σημασιολογικών χαρακτηριστικών βάσης και προθήματος. Μία εφικτή λύση είναι να μην αποτελεί σφιχτή δομή ένα πρόθημα με την ακόλουθη βάση, ώστε όταν εφαρμόσουμε τους μηχανισμούς κλίσης να μπορεί το σύστημα να σχηματίσει τύπους, όπως *προέβαλε*, *προέκυψε*, *απέστειλα*, *επέβαλε*. Είναι αναμενόμενο η εμφάνιση προβλημάτων, καθότι συχνά προκαλείται και διαγραφή φωνήεντος από το πρόθημα με την προσθήκη της χρονικής αύξησης. Παράλληλα κατά την μορφολογική ανάλυση θα πρέπει να ελέγχεται μετά τον διαχωρισμό του προθήματος/ των προθημάτων η ύπαρξη του /e/ ως χρονική αύξηση και όχι ως πιθανό αρκτικό φωνήεν θέματος. Έτσι θα πρέπει να κωδικοποιηθεί ως  $\{\$prefix\}^{19} + : \varepsilon \_ \{true\ stem\}$ , ώστε να αποφύγουμε διαχωρισμούς τύπου *\*απ-ε-ρχομαι*. Η χρονική αύξηση είναι ένα σημαντικό ζήτημα για την υπολογιστική επεξεργασία της ρηματικής κλίσης και δυστυχώς διεισδύει και στις παρυφές της διαδικασίας της παραγωγής, προκαλώντας κάποια προβλήματα κατά την ανάλυση των παραγώγων.

Στο γενικότερο προβληματισμό της επεξεργασίας με υπολογιστικούς κανόνες, παρόμοιοι κανόνες μετατροπής έχουν αναφερθεί και συζητηθεί ήδη στην βιβλιογραφία της υπολογιστικής μορφολογίας (βλ. Krieger, Pinker & Nerbone 1993, Gaussier 1995,

<sup>19</sup> Ο συμβολισμός  $\$X\$$  είναι για την προαιρετικότητα εμφάνισης του προθήματος.

Jurafsky & Martin 1999, Bernhard 2006, Handl *et al.* 2010 μεταξύ άλλων) για τα Αγγλικά και τα Γερμανικά. Οι περιγραφές τους αφορούν συνήθως ρηματικούς αλλομορφικούς τύπους, όπως αλλαγή τελευταίου χαρακτήρα (send~ sent), μηδενική αλλαγή (cut~ cut), αλλαγή τύπου ετεροίωσης (drink~ drunk) και περιγράφονται συνήθως με κανόνες προτύπων (pattern rules) ή προσεγγίζονται σε θεωρητικό επίπεδο. Προκαλεί έκπληξη ότι η πλειονότητα των ερευνητών αναφέρονται στις αλλομορφικές αλλαγές, αλλά ποτέ δεν προσπαθούν να εφαρμόσουν τη θεωρητική τους προσέγγιση στο υπολογιστικό μοντέλο που δημιουργούν ή υποστηρίζουν· σε άλλες περιπτώσεις απουσιάζει η βούληση υλοποίησης θεωρητικών προσεγγίσεων του ζητήματος από τους ίδιους τους δημιουργούς ενός μορφολογικού αναλυτή. Είναι κοινός τόπος κυρίως των ερευνητών της Μη-εποπτευόμενης Μορφολογικής Μάθησης και των συστημάτων πεπερασμένων καταστάσεων, είτε να αποφεύγουν το πρόβλημα της αλλομορφίας είτε να ελαχιστοποιούν τις επιπτώσεις του ζητήματος (βλ. υποενότητα 6.1). Ενδεικτικά οι Kohonen, Virpioja & Klami (2009) και Virpioja & Kohonen (2009) σημειώνουν πως αποκλείουν συγκεκριμένες κατηγορίες από την αλλομορφική στρατηγική του Allomorfessor· η στρατηγική τους υποδηλώνει πως αναγνωρίζουν τις ιδιαιτερότητές της και ασχολούνται μόνο με την αντιμετώπιση των προβλεπόμενων μορφηματικών αλλαγών.

Ένα ελαφρυντικό γεγονός για τη γενικότερη τακτική των ερευνητών είναι ότι ρηματική και ονοματική αλλομορφία είναι ελάχιστη στα Αγγλικά και αρκετά περιορισμένη στα Γερμανικά· τα αλλόμορφα των Αγγλικών συνήθως δεν συμμετέχουν στη διαδικασία της παραγωγής, ενώ τα αντίστοιχα των Γερμανικών έχουν κάποια συμβολή στις σύνθετες δομές. Εξάλλου στην κυρίαρχη θεωρία της MeMM, οι κανόνες μετασχηματισμού προσθέτουν επιβλεπόμενες τακτικές σε ένα αντίστοιχο μοντέλο, ενώ οι στρατηγικές των πεπερασμένων συστημάτων είναι ούτως ή άλλως κανονοκεντρικές. Οι υποστηρικτές της καθαρής θεωρίας MeMM (Goldsmith, Gaussier, Bernhard) υποστηρίζουν σθεναρά ότι ένα μοντέλο MeMM είναι μη αποδεκτό να βασίζεται σε *ελαφρώς* ανθρωπίνως σχεδιασμένη ανάλυση για τα μορφήματα μίας γλώσσας· είναι επίσης απορριπτέο για ένα μοντέλο MeMM να βοηθιέται κατά την αναγνώριση των ριζών/ προσφυμάτων/ λέξεων και να εσωκλείει προ-σχεδιασμένους κανόνες για τη σωστή μορφολογική ανάλυση (Goldsmith 2001: 156-167). Τέτοιες τακτικές υποστηρίχτηκαν επίσης από τους Pacak, Pratt & White (1976), Koch, Küstner & Rüdiger (1989) και Wothke & Schmidt (1992).



Υπογραμμίζουμε ότι μοναχά τακτικές που βασίζονται ή συμπεριλαμβάνουν βασικούς μορφολογικούς και φωνολογικούς κανόνες μπορούν να επιτύχουν υψηλά ποσοστά ορθής ανάλυσης. Τα αυτόματα πεπερασμένων καταστάσεων ως πολύπλοκα συστήματα παρέχουν τη δυνατότητα για άμεση εφαρμογή τέτοιων κανόνων. Παράλληλα και τα διάφορα μοντέλα Εποπτευόμενης Μορφολογικής Μάθησης είναι δομημένα, ώστε να ενταχθούν κανόνες και στην προκειμένη κανόνες αλλομορφικού μετασχηματισμού. Η αλλομορφία όντας είδωλο της γλωσσικής εξέλιξης της Ελληνικής κρύβει πίσω από τις αλλαγές που προκαλούνται στα μορφήματα, κυρίως κανόνες ανενεργούς συγχρονικά. Δεν πρέπει να θεωρηθεί ότι οι ανενεργοί κανόνες κωδικοποιούνται με την ίδια ευκολία και με την ίδια συστηματικότητα· εντούτοις θα αποτελέσουν τη θεμέλιο λίθο για μία επιτυχή προσέγγιση υπολογιστικής ανάλυσης της αλλομορφίας στα Ελληνικά.

#### 6.3.1.4. Φίλτρα Μορφολογικού Περιβάλλοντος

Στην εισαγωγή της συγκεκριμένης προτεινόμενης υπολογιστικής πρότασης αναφέραμε πως υπάρχουν δύο βασικοί τρόποι καταχώρησης για την κωδικοποίηση της πληροφορίας του μορφολογικού περιβάλλοντος εμφάνισης του μορφήματος. Αντίστοιχα στο θεωρητικό πλαίσιο ανάλυσης της αλλομορφίας παραθέσαμε αναλυτικά τον τρόπο καταχώρησης των αλλομόρφων στο Νοητικό Λεξικό, τους πλεοναστικούς κανόνες που συνδέουν τα μορφήματα και την αναπαράσταση των μορφολογικών πληροφοριών.

Οι Handl *et al.* (2010) χτίζουν τη στρατηγική τους για την αντιμετώπιση της αλλομορφίας στη διαφορετική καταχώρηση των πληροφοριών. Κάθε μόρφημα καταχωρείται στο λεξικό μαζί με την κωδικοποίηση του κλιτικού προτύπου που ακολουθεί. Παράλληλα οι πληροφορίες του μορφολογικού περιβάλλοντος ενσωματώνονται σε κάθε κλιτικό πρότυπο, ενώ με τους αλλο-κανόνες τα αλλόμορφα εισάγονται στο αντίστοιχο λεξικό των αλλομόρφων άνευ συνοδείας συναφών δεδομένων. Ακολουθώντας εναλλακτική κωδικοποίηση, προτείναμε τα αλλόμορφα να καταχωρίζονται μαζί με τα μορφήματα σε ένα λεξικό μαζί με την πληροφορία του αλλομορφικού τύπου που θα περιέχει και τη γραμματική κατηγορία. Έτσι θα έχουμε όλες τις απαραίτητες πληροφορίες για το φίλτρο μορφολογικού περιβάλλοντος. Φυσικά το φίλτρο θα αποτελείται από τρία βασικά υποφίλτρα, ένα για κάθε διαδικασία σχηματισμού λέξεων (κλίση, παραγωγή και σύνθεση). Ακολούθως και τα τρία

υποφίλτρα, θα ενέχουν υποκατηγορίες φίλτρων για την ολοκλήρωση κάποιας διαδικασίας (π.χ. ανάλογα με τη γραμματική κατηγορία και τη διαδικασία, όπως προθηματοποίηση ή επιθηματοποίηση). Ταυτόχρονα θα επιτρέπεται η οποιαδήποτε επικοινωνία μεταξύ των φίλτρων, όπως για παράδειγμα μετά τη σύνθεση και την παραγωγή είναι πάντα (πλην ελαχίστων εξαιρέσεων) το πέρασμα από τη διαδικασία της κλίσης. Στο φίλτρο της παραγωγής οι περιπτώσεις είναι τρεις: (α.) προθηματοποιημένα παράγωγα, (β.) επιθηματοποιημένα παράγωγα και (γ.) παράγωγα άνευ παραγωγικού προσφύματος.

Η γενική ιδέα στρατηγικής του φίλτρου είναι δομημένη στην προγραμματιστική λογική του *if...then..., else..* (εάν..., τότε..., αλλιώς). Ως πρώτη διαβάθμιση του αρχικού φίλτρου είναι ο καταμερισμός των διαδικασιών για το σχηματισμό λέξεων, ώστε να ενεργοποιείται το κατάλληλο υποφίλτρο για την εκάστοτε διαδικασία. Το γενικό σχέδιο του φίλτρου είναι το ακόλουθο:

```

for all INFLECTION do
  if ... then,
  ...
  end if
end for
for all DERIVATION do
  if ... then,
  ...
  end if
end for
for all COMPOUNDING do
  if ... then,
  ...
  end if
end for

```

} ΚΛΙΣΗ

} ΠΑΡΑΓΩΓΗ

} ΣΥΝΘΕΣΗ

Γράφημα 9: Γενικό σχήμα του φίλτρου μορφολογικού περιβάλλοντος.

Λόγω του περιορισμένου θέματος της παρούσας διατριβής θα αναλυθούν μόνο τα φίλτρα της παραγωγής. Μία εξέταση σε βάθος και ένας εκτενέστερος σχολιασμός των φίλτρων της κλίσης και της σύνθεσης είναι εκτός αντικειμένου της έρευνας μας, η οποία προσανατολίζεται στην υπολογιστική επεξεργασία της παραγωγής. Σε γενικές γραμμές μπορούμε να αναφέρουμε πως για την κλίση οι συνθήκες για την επιλογή αλλομόρφου θα είναι *if ENIKOS, then ALLOM1, else {ΠΛΗΘΥΝΤΙΚΟΣ}, ALLOM2.*

Βέβαια η συνθήκη αυτή δεν είναι πλήρως σωστή, γιατί έχουμε τις περιπτώσεις της ΚΤ4 και ΚΤ8 που έχουν διαφορετικό αλλόμορφο στη γενική ενικού, ενώ η τέταρτη διαφοροποιείται και στην αντίστοιχη πτώση του πληθυντικού.

```

for all DERIVATION do,
  for all PRE_DERIVATION do
  ... ..
  end for
  for all SUF_DERIVATION do
  ... ..
  end for
  for all NO_AF_DERIVATION do
  ... ..
  end for
end for

```

Γράφημα 10: Γενικό σχήμα του υποφίλτρου μορφολογικού περιβάλλοντος της παραγωγής.

Στην περίπτωση της προθηματοποίησης οι συνθήκες περάτωσης για την επιλογή των αλλομόρφων είναι απλές και συγκεκριμένες:

```

for all PRE_DERIVATION do
  if one allomorphs, then use ALLOM1,
  else if two allomorphs, then use ALLOM2 with VhMorph
  else use ALLOM1.
  end if
end if
end for

for all PRE_DERIVATION do
  if one allomorphs, then use ALLOM1, ## Allom1=ξε
  else if two allomorphs, then use ALLOM2 with VhMorph
  ## 2 Allom for επι, then use Allom2=εφ with ορμώ[Vh]
  else use ALLOM1. ## Allom1=επι
  end if
end if
end for

```

Γράφημα 11: Το βασικό σχήμα φίλτρου για την προθηματοποιημένη παραγωγή μαζί με 2 παραδείγματα για τις συνθήκες εφαρμογής του.

Όπως καταγράφεται στο (11.), οι προθηματοποιημένες λέξεις δεν συναντούν ιδιαίτερα προβλήματα. Εφόσον ένα πρόθημα δεν διαθέτει δύο αλλόμορφα, όπως συμβαίνει στην πλειονότητα αυτών, τότε στην παραγωγή επιλέγεται το μοναδικό

μόρφημα. Αντίθετα τα προθήματα με δύο αλλόμορφα, συμμετέχουν με την κανονική τους μορφή σε όλες τις διαδικασίες πλην εκείνων των περιβαλλόντων που το μόρφημα που ακολουθεί φέρει το χαρακτηριστικό *VhMorph*, όπως το *ορμώ*, *οπλίζω*, *ησυχάζω*. Το συγκεκριμένο στοιχείο είναι απαραίτητο από τους προθηματικούς κανόνες (99.α-β), ώστε να συμμετέχει το αλλόμορφο ALLOM2 στα συγκεκριμένα περιβάλλοντα (π.χ. *ανθ-υγιεινός*, *εφ-ησυχάζω*, *υφ-υπουργός*). Η σύνδεση προθήματος με ρηματικά θέματα πρέπει να είναι χαλαρή και πλήρως διαφανής για τη διαδικασία της κλίσης, διότι εξακολουθούν και εμφανίζονται περιπτώσεις προθηματοποιημένων ρημάτων με εσωτερική αύξηση (βλ. Κουτσούκος 2010).

```

for all SUF_DERIVATION do
  for NOM_DERIV do
    if one allomorph, then use ALLOM1,
      else if two/three allomorphs, then use ALLOM2.
    end if
  end if
end for
for VERB_DERIV do
  if one allomorph, then use ALLOM1,
    else if two allomorphs, then
      use ALLOM2 with ConstraintSUA20
      use ALLOM $X$  if specific condition
    end if
  end if
end for
for ADJ_DERIV do
  if one allomorph, then use ALLOM1,
    else if two allomorphs, then use ALLOM2.
  end if
end if
end for
for ADV_DERIV do
  use ALLOM1.
end for
end for

```

<sup>20</sup> Το ConstraintSUA είναι η κωδικοποίηση για τον περιορισμό επιλογής μοναδικού αλλομόρφου που καθορίζει τη χρήση ενός μόνο αλλομόρφου στην παραγωγή (αναλυτικότερα βλ. 3.3.1).

Γράφημα 12: Το βασικό σχήμα φίλτρου για την επιθηματοποιημένη παραγωγή.

Σαφώς διαφορετική είναι η προσέγγιση στην επιθηματοποιημένη παραγωγή, καθότι διακρίνεται από τη βάση ή το παραγωγικό επίθημα. Στις περιπτώσεις της ονοματικής και επιθετικής αλλομορφίας στην παραγωγή, ο σχεδιασμός του φίλτρου βασίζεται πάνω στον περιορισμό της επιλογής μοναδικού αλλομόρφου. Ο περιορισμός αναλύεται σε δύο κατευθύνσεις, στα μορφήματα με ένα αλλόμορφο και σε εκείνα με δύο ή και τρία αλλόμορφα. Η συνθήκη για τα μορφήματα χωρίς αλλομορφία είναι απλή, ενώ για τα μορφήματα με δύο ή τρία αλλόμορφα χαρακτηρίζουμε ως ALLOM2 αυτό που συμμετέχει στην παραγωγή. Για λόγους τακτικής χαρακτηρίζουμε ως ALLOM2 το αλλόμορφο που εμφανίζεται στον πληθυντικό με μοναδική εξαίρεση αυτό της όγδοης αλλομορφικής τάξης όπου έτσι χαρακτηρίζουμε το ‘νέο’ αλλόμορφο παραγωγής και σύνθεσης που διαφοροποιείται από το αντίστοιχο του πληθυντικού (*μπακάληδ-ες* vs. *μπακάλ-ικος*).

```

for all SUF_DERIVATION do
  for NOM_DERIV do
    if one allomorph, then use ALLOM1,## Allom1=ανθρωπ
      else if two/three allomorphs, then use ALLOM2.
        ##2 Allom for ψυχη, then use Allom2=ψυχ
      end if
    end if
  end for
end for
end for

```

Γράφημα 13: Παραδείγματα φίλτρου ονοματικής επιθηματοποιημένης παραγωγής.

Τα επίθετα δεν διαφέρουν καθόλου από τα ουσιαστικά και έτσι δεν εντοπίζεται κάποιο ζήτημα που χρήζει εναλλακτική προσέγγιση. Η πλειονότητα των ιδιαίτερα συχνών επιθέτων κλίνονται με βάση κάποιο ονοματικό κλιτικό σύστημα, με εξαιρέσεις τα επίθετα τύπου *αληθ-ης*, *αληθ-ες* που δεν έχουν αλλόμορφα και τα *βαθ-*, *βαθ-εια*, *βαθυ* που αλλομορφικά το αρσενικό και ουδέτερο προσομοιάζει στα αντίστοιχα της KT4, αλλά εντάσσονται σε νέα κλιτική τάξη, την KT11. Φυσικά για τα επιρρήματα δεν τίθεται κάποιο ζήτημα, καθότι παραμένουν χωρίς συνθήκες αλλομόρφων, μίας και δεν έχουν.

Τέλος, το ρηματικό παραγωγικό μοντέλο για θέματα και επιθήματα παραμένει το δυσκολότερο κομμάτι της υλοποίησης του φίλτρου. Η έλλειψη καθολικής ρηματικής συστηματικότητας, περιπλέκει τη δομή του φίλτρου μορφολογικού περιβάλλοντος, καθότι δεν ακολουθούν όλα τα ρήματα της πρώτης ρηματικής κλιτικής τάξης τις υποδείξεις του περιορισμού επιλογής μοναδικού αλλομόρφου· ο περιορισμός αφορά ονοματικά θέματα και επιθήματα, αλλά εντούτοις τα ρήματα της ΡΚΤ2 τον εφαρμόζουν στα δικά τους παραγωγικά παραδείγματα. Στην υποενότητα 6.3.1.2. αναλύθηκαν στην πλειονότητά τους οι ρηματικοί αλλομορφικοί κανόνες μετατροπής, οι οποίοι καθορίζουν ως το δεύτερο αλλόμορφο του κάθε ρήματος, εκείνο που εμφανίζεται (συνήθως) στον αόριστο και κυρίως στην παραγωγή. Εντούτοις, υφίσταται μία ιδιαίτερα ασυνήθιστη συνθήκη στο φίλτρο και αφορά στα αλλόμορφα που αναλύθηκαν στους κανόνες (97.) και συμμετέχουν με ποικίλους τρόπους στην παραγωγή. Για παράδειγμα, *κλεφτης* (*κλεβ-*), *κλέψιμο* (*κλεβ-*), *κλοπη* (*κλοπ-*), ενώ υπάρχει και το *κλαπ-* (από το *κλάπ-ηκ#*)· *καλεσμένος* (*καλεσ-*), *καλεσμα* (*καλεσ-*), *κλήση* (*κλη-*) [βλέπε και {προσ}κεκλημένος]· *βαλτός* (*βαλ-*), *βάλσιμο* (*βαλ-*), *βλήμα* (*βλη-*), *βολή* (*βολ-*). Το πρόβλημα των συγκεκριμένων ρημάτων δεν εντοπίζεται στον υπολογιστικό σχηματισμό αλλομόρφων, αλλά στη διαδικασία επιλογής του κατάλληλου τύπου για την παραγωγή.

```

for VERB_DERIV do
    if one allomorph, then use ALLOM1,## Allom1=ιδρύ
    else if two allomorphs, then
        use ALLOM2 with ConstraintSUA
        ## 2 Allom=αγαπάω, then use Allom2=αγαπη
        use ALLOM3 if specific condition
        ## 3 Allom= καλώ, then use Allom2=καλεσ
            then use Allom3=κλη
    end if
end if
end for

```

Γράφημα 14: Παραδείγματα φίλτρου ρηματικής επιθηματοποιημένης παραγωγής.

Η συστηματική ανάλυση των παραγωγικών επιθημάτων εξάγει κάποια συμπεράσματα και δίνει πληροφορίες για τα μορφολογικά περιβάλλοντα, όπως για

παράδειγμα το *-σιμο*<sup>21</sup> συντάσσεται πάντα με μονοσυλλαβικά θέματα. Ίσως στις παραπάνω περιπτώσεις να συνέδραμαν και συμπληρωματικές πληροφορίες, όπως ο τόνος και οι συλλαβές, αλλά η ενδελεχής εξέταση δεν μας παρέχει τη δυνατότητα για την καλύτερη αντιμετώπιση στο φίλτρο. Εντούτοις υφίσταται ένα θεωρητικό ζήτημα σε αρκετά από αυτά τα παράγωγα. Ποια από αυτά είναι σημασιολογικά και κυρίως μορφολογικά διαφανή; Για πολλούς συγγραφείς γραμματικών αρκετά ουσιαστικά παραμένουν και συγχρονικά μορφολογικά διαφανή, όπως *βήμα* από *βαίνω*, *γεύμα* από το *γεύομαι*, *βλήμα* από *βάλλω*, *στέμμα* από *στέφω*, *τμήμα* από *τέμνω*, κτλ., όπου αναγνωρίζουν πάντα το παραγωγικό επίθημα *-μα~ -ματ-*. Σίγουρα οι αναλύσεις μερικών παραγώγων με ορατά συγχρονικά επιθήματα επιδέχεται αρκετή κριτική και είναι πολύ πιθανό ότι αρκετοί μορφολόγοι θα διαφωνήσουν με διαχωρισμούς τύπου *τιμη-μα*, *βλη-μα*, *βη-μα*. Σίγουρα τα προαναφερθέντα παραδείγματα αποτελούν σκοπέλους για τα υπολογιστικά συστήματα παραγωγής λέξεων.

```
for all NO_AF_DERIV do
  if one allomorph, then use Allom1
    else if two allomorpha, then use Allom2 or
    else if three allomorpha, then use Allom3
    end if
  end if
end for
```

```
for all NO_AF_DERIV do
  if one allomorph, then use Allom1
    else if two allomorpha, then use Allom1 or
    ## 2 Allom for γραψω, then use Allom1=γραφ
    else if three allomorpha, then use Allom3
    ## 3 Allom for τρεπω, then use Allom3=τροπ
    end if
  end if
end for
```

<sup>21</sup> Διαφωνούμε με τον διαχωρισμό του τελικού *-ο* από το παραγωγικό επίθημα (*\*-σιμ(ο)*), καθότι κάτι τέτοιο θα μας απέτρεπε να το εντάξουμε στην ΚΤ8, όπου υφίσταται μηδενικό κλιτικό μόρφωμα στην ονομαστική, αιτιατική και κλιτική ενικού. Σε αντίθετη περίπτωση θα πρέπει να δημιουργηθεί νέα κλιτική τάξη.

Γράφημα 15: Το βασικό σχήμα φίλτρου για την παραγωγή χωρίς προσφύματα.

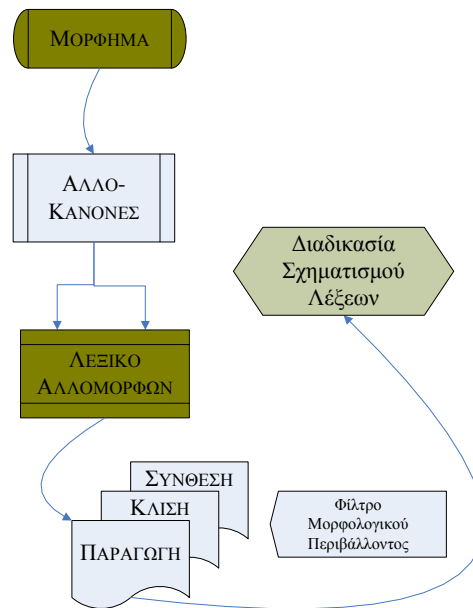
Η πιο περιορισμένη ομάδα παραγώγων είναι στατιστικά τα σχηματιζόμενα με την απουσία παραγωγικού προθήματος ή επιθήματος. Όπως παρουσιάστηκε στη θεωρητική ανάλυση της παραγωγής, το ανενεργό μορφοφονολογικό φαινόμενο της ετεροίωσης και η μετάπλαση μπορούν να σχηματίσουν παράγωγες λέξεις χωρίς την παρουσία παραγωγικού επιθήματος. Εδώ έχουμε τρεις υποσυνθήκες, όπου:

(α.) τα ρήματα χωρίς αλλομορφία σχηματίζουν κάποιο παράγωγο με την αλλαγή γραμματικής κατηγορίας, όπως *γράφω – γραφή*, χωρίς να θεωρούμε ότι ένας από τους δύο τύπους είναι η βάση της παραγωγής, αφού συνδέονται με έναν πλεοναστικό μεταπλαστικό κανόνα (βλ. Plag 2002: 134-144, Ράλλη 2005: 141).

(β.) αντίστοιχα τα ρήματα της ΚΤ2 σχηματίζουν τα παράγωγα της μετάπλασης με τη χρήση του πρώτου αλλομόρφου (ALLOM1), δηλαδή το μόρφημα που εμφανίζεται πρωτίστως στο κλιτικό παράδειγμα του ενεστώτα. Τέτοιες περιπτώσεις είναι *κυνηγός – κυνηγώ, αγάπη – αγαπώ*. Με τη μετάπλαση συνδέονται δύο τύποι με πλεοναστικό κανόνα και αδυνατούμε να ισχυριστούμε με βεβαιότητα ποιος είναι ο αρχικός τύπος και επομένως αναμένεται να χρησιμοποιείται πάντα το πρώτο αλλόμορφο.

(γ.) τα ρήματα με τρία αλλόμορφα με το τρίτο αλλόμορφο να είναι κατάλοιπο της ετεροίωσης δημιουργούν παράγωγες λέξεις τύπου *κλέβω – κλοπή, φέρνω – φορά, τρέπω – τροπή*. Σε αυτές τις περιπτώσεις «εμφανίζεται» ένα νέο αλλόμορφο για αυτές τις παράγωγες λέξεις, καθώς και για όλες όσες θα παραχθούν από αυτές, αλλά με την εφαρμογή της προθηματοποίησης ή της επιθηματοποίησης, όπως *τροπή – εκτροπή, κλοπή – κλοπιμαίος*.





Γράφημα 16: Η διαδικασία δημιουργίας, καταχώρησης και επιλογής των αλλομόρφων σύμφωνα με την πρόταση των αλλομορφικών κανόνων μετατροπής.

Εξετάζοντας συνολικά το μοντέλο των κανόνων αλλομορφικού μετασχηματισμού εντοπίζουμε δύο προβληματικές υποενότητες. Τόσο οι αλλο-κανόνες μίας μικρής ομάδας ρημάτων με σχεδόν ατομικευμένη αλλομορφία, όσο και οι συνθήκες του φίλτρου για την ίδια ομάδα αποτελούν ιδιαίζουσα πρόκληση. Οι κανόνες είναι απαραίτητοι όχι μόνο για το συσχετισμό των αλλομόρφων, αλλά και για τον καθορισμό του μορφολογικού περιβάλλοντος εμφάνισης. Οι τακτικές που καταχωρούν τα λήμματα με δείκτες αναφοράς σε κλιτικά πρότυπα συμβάλλουν στην απρόσκοπτη και χωρίς εκπλήξεις υλοποίηση ενός μορφολογικού αναλυτή που να αντιμετωπίζει την κλίση μίας γλώσσας. Η ενίσχυση της τακτικής αυτής και στις υπόλοιπες διαδικασίες, της παραγωγής και της σύνθεσης περιπλέκεται ιδιαίτερα, γιατί πλέον οι συνθήκες είναι πολλές και διαφορετικές.

Οι κανόνες αλλομορφικής μετατροπής παρέχουν τη δυνατότητα για ένα οικονομικότερο μοντέλο με δυναμική και αποτελεσματικότητα, καθότι παρέχει τις βασικές δομές για την επίλυση των προβλημάτων που προκαλεί η έντονη αλλομορφία στα Νέα Ελληνικά. Ο διαχωρισμός των πληροφοριών σε διάφορα κομμάτια (components), τα οποία λειτουργούν διαδραστικά μεταξύ τους και διασφαλίζουν την ομαλή λειτουργία και κυρίως την αποφυγή των αντιγραμματικών περιπτώσεων με την επιλογή λανθασμένου αλλομόρφου. Εξάλλου είναι εύκολο να γίνει πολύπλοκο, όταν ακολουθείται το σύστημα του δείκτη αναφοράς σε κλιτικό πρότυπο. Το μοντέλο του Σγάρμπα (1996) για τα Ελληνικά περιλαμβάνει 89 διαφορετικά κλιτικά παραδείγματα

για ουσιαστικά, γεγονός που καταδεικνύει την ανάγκη για έναν μορφολογικό αναλυτή με σαφώς πιο συμπυκνωμένο σύστημα ανάλυσης δομημένο στις αρχές της σύγχρονης γλωσσολογίας και της οικονομίας. Οι Handl *et al.* (2010) διασαφηνίζουν πως ο διαχωρισμός της αλλομορφίας στα κλιτικά πρότυπα τους επιτρέπει να αποκλείσουν προβλήματα στις αναλύσεις των γερμανικών συνθέτων.

Κλείνοντας, να τονιστεί ξανά ότι γίνεται συστηματική προσπάθεια για εφαρμογή των βασικών αρχών του θεωρητικού πλαισίου για την υπολογιστική επεξεργασία της αλλομορφίας: οι προαναφερθέντες υπολογιστικοί κανόνες αλλομορφικής μετατροπής είναι είδος αποτύπωσης των πλεοναστικών κανόνων που περιγράφουν τη σχέση μεταξύ των αλλομορφών στο Νοητικό Λεξικό. Οι υπολογιστικοί κανόνες δημιουργούν αλλομορφα σε αντίθεση με τους πλεοναστικούς κανόνες των αλλομορφών που απλώς συνδέουν δύο – ή και παραπάνω – αλλομορφα του ίδιου μορφήματος. Ωστόσο το τελικό αποτέλεσμα είναι ίδιο, αφού και οι υπολογιστικοί κανόνες καταλήγουν να δημιουργούν αλλομορφα και να τα παραθέτουν με μορφή αντίστοιχη των κανόνων του Νοητικού Λεξικού.

### **6.3.2. Προβλεψιμότητα και Κανονικότητα στην Αλλομορφία – Η πρόταση των χαρακτηριστικών κατηγοριοποίησης σε Μέγιστη Εντροπία**

Το δεύτερο σκέλος της προτεινόμενης υπολογιστικής πρότασης για την επεξεργασία της αλλομορφίας στοχεύει στην παρουσίαση ενός μοντέλου εποπτευόμενης μορφολογικής μάθησης με Μέγιστη Εντροπία που έχει ως στόχο τη χρήση χαρακτηριστικών για την πρόβλεψη της αλλομορφίας. Ένα μοντέλο που εκπαιδεύεται από μορφολογικά και φωνολογικά χαρακτηριστικά των ελληνικών λέξεων φαίνεται πως προβλέπει την παρουσία ή απουσία της αλλομορφίας, καθώς και τον τύπο αλλομορφικής μετατροπής. Συνήθως τα μοντέλα πρόβλεψης χρησιμοποιούνται στη σύνταξη στην προσπάθεια εντοπισμού του υποκειμένου και αντικειμένου ενός ρήματος. Εδώ προτείνουμε την εκμετάλλευση των δυνατοτήτων της Μέγιστης Εντροπίας για τον εντοπισμό της κανονικότητας των αλλομορφών παράγωγων λέξεων. Στο προτεινόμενο μοντέλο δεν καταστρατηγούμε τον ορισμό για την αλλομορφία, αφού εκμεταλλευόμαστε τις ιδιότητες των αποτελεσμάτων της διαδικασίας και όχι την ίδια τη διαδικασία. Σε γενικές γραμμές, θα παρουσιάσουμε τη λογική του μοντέλου πρόβλεψης με καθολικά χαρακτηριστικά ή καλύτερα τροποποιημένα λεξικά χαρακτηριστικά με βάση τα αποτελέσματα του πειράματος AMIS.

Όπως προαναφέρθηκε, η αλλομορφία ορίζεται ως το φαινόμενο που περιγράφει αυθαίρετες και απρόβλεπτες αλλαγές μορφημάτων· εντούτοις, υπάρχει προβλεψιμότητα και κανονικότητα στο επίπεδο της διαδικασίας της αλλομορφίας και όχι σε αυτό του ορισμού της. Τα αλλόμορφα στα Ελληνικά οφείλονται στην απενεργοποίηση φωνολογικών και μορφολογικών κανόνων και στις μετακινήσεις ορίων μορφημάτων λόγω της επανάλυσης. Επομένως, η μορφή των συγκεκριμένων αλλομόρφων δεν προβλέπεται συγχρονικά από κανόνες. Το κέρδος της ιστορικής ανάλυσης της αλλομορφίας έχει τονιστεί ήδη στην προηγούμενη υποενότητα και είναι σημαντικό για την εξαγωγή συμπερασμάτων όσον αφορά στις μορφολογικές αλλαγές. Είναι κρίσιμο για την υπολογιστική εκμάθηση της μορφολογίας, μίας και οι ανενεργοί κανόνες μπορούν να κωδικοποιηθούν στο σύστημα, με στόχο να βοηθήσει το σύστημα να αντιμετωπίσει το φαινόμενο της αλλομορφίας.

Σε παλαιότερες έρευνές μας (Καρασίμος 2009, Karasimos 2009) συγκρίνοντας τη θεματική αλλομορφία με την επιθηματική αλλομορφία παρατηρήθηκαν κάποιες κανονικότητες και σημαντικές ομοιότητες ανάμεσα στις κατηγορίες των θεμάτων και των παραγωγικών επιθημάτων, αφού είναι λήμματα του ίδιου Λεξικού (Ralli 1988, Ράλλη 2005). Τα επιθήματα παρουσιάζουν παρόμοια αλλομορφική συμπεριφορά με αυτά των θεμάτων, με τα οποία τα προηγούμενα μοιράζονται μερικά κοινά χαρακτηριστικά και εντάσσονται στις ίδιες κλιτικές τάξεις. Παραδείγματος χάριν, το επίθημα *-α(ς)* ‘κάποιος που ενεργεί’ είναι παρόμοιο με τα ουσιαστικά τύπου *παπάς*· ακολουθούν τον κανόνα *δ-εισαγωγής* για τον σχηματισμό πληθυντικού και παραγωγής.

(100.)	α. παπά(ς)	παπάδ(ες)	- <b>α(ς)</b>	παπλωματ-ά(ς)	παπλωματάδ(ες)
	θάλασσα	θάλασσ(ες)	- <b>άλα</b>	κρεμ-άλα	κρεμάλ(ες)
	μωρ(ό)	μωρ(ά)	- <b>εί(ο)</b>	μεγαλ-εί(ο)	μεγαλεί(α)
	βήμα	βήματ(α)	- <b>μα</b>	άγγιγ-μα	αγγίγματ(α)
	β. μπακάλη(ς)	μπακάληδ(ες)	- <b>άρη(ς)</b>	βρωμι-άρη(ς)	βρωμιάρηδ(ες) <sup>22</sup>
	παιδί	παιδι(ά)	- <b>ούλι</b>	σακ-ούλι	σακούλι(α) <sup>23</sup>

Αυτή η κανονικότητα εμφανίζεται επίσης και ανάμεσα στα ρήματα και τα ρηματικά επιθήματα, ενώ διέπει και τη σχέση των επιθετικών θεμάτων με τα επιθετικά επιθήματα. Οι παράγωγες λέξεις εντάσσονται και αυτές σε μία από τις κλιτικές τάξεις

<sup>22</sup> Αλλά στην περαιτέρω παραγωγή *βρωμιάρ-ικος*, κατά το *μπακάλ-ικος*.

<sup>23</sup> Αλλά στην περαιτέρω παραγωγή *σακουλ-άκι*, κατά το *παιδ-άκι*.

κάθε κατηγορίας· επομένως τα κριτήρια που είναι η ομάδα των κλιτικών επιθημάτων με την οποία συνδυάζονται και η αλλομορφία που εμφανίζουν, εφαρμόζονται και σε αυτές τις περιπτώσεις. Είναι, λοιπόν, απαραίτητη προϋπόθεση για ένα παράγωγο ουσιαστικό να συνάδει αλλομορφικά με τα υπόλοιπα μη παράγωγα ουσιαστικά της ίδιας κλιτικής τάξης. Στόχος της προτεινόμενης υπολογιστικής πρότασης είναι να εκμεταλλευτούμε τις ανωτέρω παρατηρήσεις από το θεωρητικό μας πλαίσιο.

Επιπλέον, το φαινόμενο της επανάλυσης των λέξεων προκαλεί αλυσιδωτές αλλαγές, καθώς ένα πρώην ανεξάρτητο επίθημα είναι πλέον μέρος του θέματος ή μέρος του κλιτικού επιθήματος να εντάσσεται συγχρονικά στο θέμα. Η πιο εντυπωσιακή κανονικότητα είναι αυτή του (100.β). Όπως έχουμε ήδη τονίσει στην υποενότητα 2.4.4, υπάρχουν μερικά ουσιαστικά που εμφανίζουν νέα αλλόμορφα κατά τη διαδικασία της παραγωγής (και σύνθεσης) διαφορετικά αυτών της κλίσης. Παρομοίως, τα επιθήματα, όπως *-άρη(ς)*, *-ούλι*, συμπεριφέρονται ακριβώς όπως τα ουσιαστικά *μπακάλη(ς)* και *παιδί*. Αυτό σημαίνει ότι δημιουργούν ένα αλλόμορφο παρόλη την παρουσία του (*ταξιδ-ιάρη(ς)*, *ταξιδ-ιάρηδ(ες)*, αλλά *ταξιδ-ιάρ-ικός*) ή την απουσία του (*σακ-ούλι* *σακ-ούλι(α)*, αλλά *σακ-ουλ-άκι*) στην κλίση. Όλες αυτές οι αποδείξεις καταδεικνύουν ότι η αλλομορφία συμμετέχει σε όλα τα επίπεδα σχηματισμού λέξεων με την ίδια κανονικότητα. Επομένως, τα αλλόμορφα εμφανίζονται όσο απρόβλεπτα τα καθόριζε η Lieber (1982)· εξακολουθούν να σχηματίζονται συγχρονικά απρόβλεπτα, αλλά εσωκλείουν προβλεψιμότητα.

Αυτές οι κανονικότητες μπορούν να μας βοηθήσουν να κατηγοριοποιήσουμε καλύτερα την αλλομορφική συμπεριφορά θεμάτων και επιθημάτων. Μας παρέχουν επίσης κοινούς κανόνες που μπορούν να κωδικοποιηθούν σε ένα μοντέλο Εποπτευόμενης Μάθησης με εκ των προτέρων χειρωνακτική ανάλυση μορφημάτων, όπως παρουσιάζουμε στους αλλομορφικούς κανόνες της ενότητας 6.2.1. Το πείραμα με το AMIS και η εφαρμογή με τον ΑλλοMantIS βασίστηκε και δομήθηκε πάνω στο συγκεκριμένο θεωρητικό πλαίσιο, στην παραδοχή της προβλεψιμότητας και στις παρατηρήσεις της κανονικότητας. Τα αποτελέσματα δικαίωσαν και με χειροπιαστό και στατιστικό τρόπο, ότι η θεωρία ορθώς πρόβλεψε την κανονικότητα και την προβλεψιμότητα της ονοματικής αλλομορφίας.

Το πείραμα του AMIS (βλ. ενότητας 5.2.), το οποίο απέδειξε σημαντικά στοιχεία κανονικότητας και προβλεψιμότητας στη διαδικασία της αλλομορφίας, μας παρέχει τις βασικές και απαραίτητες δομές για να υλοποιήσουμε μια πρόταση χαρακτηριστικών.

Εφόσον στη θεωρία ισχυριζόμαστε ότι υφίσταται προβλεψιμότητα στην αλλομορφία και το πείραμα μας παρέχει και τις αποδείξεις, πρέπει να την υλοποιήσουμε και υπολογιστικά. Η επεξεργασία των δεδομένων φανερώνει ότι πιθανότατα να κρύβονται πίσω από τα μορφήματα και την ακολουθία φωνημάτων κάποια πρότυπα που συμβάλλουν στην αντιμετώπιση του φαινομένου. Στους πίνακες 16 και 17 παρουσιάσαμε τα κυρίαρχα στοιχεία για κάθε αλλομορφική ονοματική τάξη, καθώς και τα αποτρεπτικά βάρη αρνητικής επιρροής για κάθε μία από αυτές.

Πρωταρχικό ρόλο για την επιλογή αλλομόρφου σύμφωνα με τα αποτελέσματα του πειράματος διαδραμάτισαν οι συλλαβές, καθώς και κάποιοι χαρακτήρες, ενώ αντιθέτως άλλες πιο βασικές πληροφορίες συνετέλεσαν ως αποτρεπτικοί παράγοντες για την αποφυγή καταχώρησης σε αλλομορφικές τάξεις λέξεις που δεν ανήκουν σε αυτές. Η αλλομορφία είναι ένα αμιγές μορφολογικό φαινόμενο που, όμως, για τα συστήματα εποπτευόμενης μάθησης της μορφολογίας απαιτούνται και περιφερειακές μη μορφολογικές πληροφορίες, όπως φωνολογικές και σημασιολογικές. Το σύστημα ανάλυσης βασισμένο σε χαρακτηριστικά (μορφολογικά, φωνολογικά και σημασιολογικά) θεωρείται ελαφρώς πολύπλοκο, αλλά και ιδιαίτερα αντικοινωνικό με την κάθε λέξη να συνοδεύεται από τόσες χειρωνακτικά δοσμένες πληροφορίες. Στην ουσία, όμως, το σύστημα δεν είναι όσο αντικοινωνικό φαίνεται, ενώ κινείται πάνω στις βασικές γραμμές ενός υπολογιστικού μοντέλου χαρακτηριστικών.

Κρίνεται, λοιπόν, απαραίτητη για έναν μορφολογικό αναλυτή τέτοιου είδους η συνδρομή και άλλων εργαλείων, όπως οι συντακτικοί αναλυτές και φωνολογικοί διαχωριστές. Ήδη από τα πρώτα βήματα των συνδυαστικών μορφολογικών αναλυτών (Zampolli 1990), οι ερευνητές κατέφευγαν στη βοήθεια εργαλείων από άλλα επίπεδα της γλώσσας (Calzonari, Mammini & Monachini 2004, Declerck, Kessler & Krieger 2005, Monachini *et al.* 2005). Είναι συνήθης τακτική για τους αυτόματους σχολιασμούς σωμάτων κειμένων να καταφεύγουν στη λύση των εργαλείων που δουλεύουν παράλληλα (για τα Ελληνικά βλ. ενδεικτικά Κουτσομπόγερα 2004).

Τα χαρακτηριστικά του ονοματικού μοντέλου είναι η κλιτική τάξη, το γένος, οι συλλαβές, ο τόνος, ο αριθμός συλλαβών, τα τελευταία φωνήματα, η εμφυσχότητα και η καταγωγή της λέξης. Κάποια από τα παραπάνω μπορούν να εξαχθούν με αυτοματοποιημένες διαδικασίες. Πιο συγκεκριμένα, με τη συνδρομή ενός φωνολογικού διαχωριστή μπορούμε να ξεχωρίσουμε τις συλλαβές, τους τελευταίους χαρακτήρες, καθώς και τη θέση του τόνου. Παράλληλα, ένας συντακτικός αναλυτής

μπορεί να δώσει το γένος από τα συμφραζόμενα με τη συνδρομή ενός μορφολογικού εξαγωγέα κλιτικών επιθημάτων (inflectional suffix stripping), καθώς και με αρκετές πιθανότητες την κλιτική τάξη. Ελάχιστες πληροφορίες δεν μπορούν να ενταχθούν αυτόματα, όπως αυτές της προέλευσης και εμψυχότητας αποκλείονται (εν μέρει) από τις αυτόματες τεχνικές σχολιασμού κειμένων. Δεν αποτελεί στόχο της παρούσας εργασίας να ασχοληθεί με τον ιδιαίτερα αναπτυσσόμενο τομέα του αυτόματου σχολιασμού κειμένων (automated annotation for corpora) και επομένως δεν θα επεκταθούμε περαιτέρω στον τρόπο αυτόματης προσθήκης πληροφοριών και χαρακτηριστικών στις λέξεις.

Ένα μοντέλο μορφολογικής ανάλυσης που είναι βασισμένο στην πρόταση των χαρακτηριστικών ανήκει κυρίως στην κατηγορία των συστημάτων Εποπτευόμενης Μορφολογικής Μάθησης που εκμεταλλεύεται την τακτική της Μέγιστης Εντροπίας. Μία προσέγγιση με χρήση κάποιων χαρακτηριστικών είναι καθολική, μίας και πρέπει να είναι πλήρως παραμετροποιήσιμη. Είναι βέβαιο ότι κάθε κατηγορία σε κάθε γλώσσα θα χρειαστεί κάποιες τροποποιήσεις των βασικών χαρακτηριστικών (π.χ. η ύπαρξη του γένους ή ο αριθμός των κλιτικών τάξεων). Τα χαρακτηριστικά πρέπει να επιλεγούν με προσοχή και ενδελεχή μελέτη της εκάστοτε κατηγορίας, γιατί πάντα ελλοχεύει ο κίνδυνος της υπερφόρτωσης ενός συστήματος είτε με περιττές, είτε με ελλιπείς πληροφορίες. Για παράδειγμα, στην ανάλυση του ΑλλοMantIS η πιο προβληματική ομάδα παράγωγων ουσιαστικών ήταν εκείνη της όγδοης αλλομορφικής τάξης τύπου *μπακάλης*: στη θεωρητική ανάλυση καταλήξαμε στο συμπέρασμα ότι διαφοροποιούνται από τα υπόλοιπα ουσιαστικά της ΚΤ, μίας και είναι όλα τους τουρκικής προέλευσης, τονίζονται στην παραλήγουσα και είναι έμψυχα. Η τελευταία πληροφορία είναι μοναδική, μίας και δεν υπήρχε σε καμία άλλη ομάδα με τη θετική βαρύνουσα επιρροή. Εφόσον δεν συνέβαλε με τον τρόπο της στη σωστή πρόβλεψη των παράγωγων ουσιαστικών, θεωρούμε ότι είναι περιττή πληροφορία και προφανώς δεν είναι απαραίτητη η παρουσία της στο μοντέλο. Στο ακόλουθο παράδειγμα παρουσιάζουμε τα μοντέλα χαρακτηριστικών για ουσιαστικά, επίθετα και ρήματα. Μόνο το πρώτο έχει δοκιμαστεί πειραματικά και η τελική επιλογή τους είναι πλέον βεβαιωμένη στατιστικά. Υποθέτουμε ότι τα αντίστοιχα χαρακτηριστικά των λοιπών μοντέλων θα είναι εξίσου αποτελεσματικά στην πρόβλεψη της αλλομορφίας.

(101.) α. *Ονοματικό μοντέλο*

Κλιτική Τάξη (ΚΤ): IC1 IC2 IC3 IC4 IC5 IC6 IC7 IC8

Γένος: masculine, feminine, neutral

Αριθμός Συλλαβών: Syl1, Syl2, Syl3, Syl4, Syl5, Syl6, Syl7

Συλλαβές: Syl1\_ultimate, Syl2\_penultimate, Syl3\_antipenultimate

Τόνος: stress\_ultimate, stress\_penultimate, stress\_antipenultimate

Χαρακτήρες: Ch1\_last, Ch2\_last-1, Ch3\_last-2, Ch4\_last-3

Καταγωγή: greek, turkish, italian, english, french, etc.

Εμψυχότητα: ~~animaey, non\_animaey~~

### β. Επιθετικό μοντέλο

Κλιτική Τάξη (KT): IC1 IC3 IC5 IC9 IC10 IC11

Γένος: masculine, feminine, neutral

Αριθμός Συλλαβών: Syl1, Syl2, Syl3, Syl4, Syl5, Syl6, Syl7

Συλλαβές: Syl1\_ultimate, Syl2\_penultimate, Syl3\_antipenultimate

Τόνος: stress\_ultimate, stress\_penultimate, stress\_antipenultimate

Χαρακτήρες: Ch1\_last, Ch2\_last-1, Ch3\_last-2, Ch4\_last-3

~~Καταγωγή: greek, turkish, etc.~~

~~Εμψυχότητα: animaey, non\_animaey~~

### γ. Ρηματικό μοντέλο

Κλιτική Τάξη (KT): IC1 IC2A IC2B

Αριθμός Συλλαβών: Syl1, Syl2, Syl3, Syl4, Syl5, Syl6, Syl7

Συλλαβές: Syl1\_ultimate, Syl2\_penultimate, Syl3\_antipenultimate

Τόνος: stress\_ultimate, stress\_penultimate, stress\_antipenultimate

Χαρακτήρες: Ch1\_last, Ch2\_last-1, Ch3\_last-2, Ch4\_last-3

Μεταβατικότητα: no\_transitive, one\_transitive, two\_transitive

Τύπος Υποκειμένου: no\_subj, anim\_subj, non\_anim\_subj, both\_subj

Το μοντέλο για την αλλομορφία των επιθέτων δεν διαφέρει συγκριτικά με το ονοματικό μοντέλο. Διαγράφουμε τα χαρακτηριστικά της καταγωγής και της εμψυχότητας, διότι δεν περιμένουμε ότι θα βελτιώσει τα ποσοστά πρόβλεψης. Παρατηρώντας προσεκτικά τα δεδομένα εντοπίσαμε ότι η εμψυχότητα δεν διαδραμάτισε κάποιο σημαντικό ρόλο θετικής ή αρνητικής επιρροής, ενώ παρόμοια ήταν και η συμπεριφορά των χαρακτηριστικών καταγωγής. Τα βασικά χαρακτηριστικά ενός ρηματικού μοντέλου πρόβλεψης θα διαφέρουν σε λιγιστά σημεία από το μοντέλο

ουσιαστικών και επιθέτων. Τα χαρακτηριστικά που θα πρέπει να διέπουν το μοντέλο, ώστε να είναι ικανό για υψηλή προβλεψιμότητα είναι σίγουρα ο *αριθμός συλλαβών*, ο *τόνος*, η *κλιτική τάξη* (με πιθανή αύξηση των τάξεων του μοντέλου της Ralli 1988, Ράλλη 2005), οι *συλλαβές*, οι *τελευταίοι χαρακτήρες* (πλην των κλιτικών επιθημάτων), η *μεταβατικότητα*, ο *τύπος υποκειμένου*. Η πεπατημένη του ονοματικού μοντέλου μάς παρέχει τους βασικούς άξονες πάνω στους οποίους θα χτίσουμε και το παρόν μοντέλο.

(102.) α. *Ονοματικό μοντέλο για ΔΟΜΙΣΜΟΣ*

Κλιτική Τάξη (KT): IC1

Γένος: masculine

Αριθμός Συλλαβών: Syl3

Συλλαβές: Syl1\_σμος, Syl2\_μι, Syl3\_δο

Τόνος: stress\_ultimate

Χαρακτήρες: Ch1\_μ, Ch2\_σ, Ch3\_ι, Ch4\_μ

Καταγωγή: greek

~~Εμψυχότητα: non\_animacy~~

β. *Επιθετικό μοντέλο για ΦΕΥΓΑΤΗ*

Κλιτική Τάξη (KT): IC3

Γένος: feminine

Αριθμός Συλλαβών: Syl3

Συλλαβές: Syl1\_τη, Syl2\_γα, Syl3\_φευ

Τόνος: stress\_penultimate

Χαρακτήρες: Ch1\_η, Ch2\_τ, Ch3\_α, Ch4\_γ

~~Καταγωγή: greek~~

~~Εμψυχότητα: animacy, non\_animacy~~

γ. *Ρηματικό μοντέλο για ΔΩΡΙΖΩ*

Κλιτική Τάξη (KT): IC1

Αριθμός Συλλαβών: Syl3

Συλλαβές: Syl1\_ζω, Syl2\_ρι, Syl3\_δω

Τόνος: stress\_penultimate

Χαρακτήρες: Ch1\_ζ, Ch2\_ι, Ch3\_ρ, Ch4\_ω

Μεταβατικότητα: two\_transitive



## Τύπος Υποκειμένου: anim\_subj

γράφω

IC1 AC1 2syl stress\_penultimate syl1\_φω syl2\_γρα one\_transitive both\_subj

αγαπώ

IC2a AC2 3syl stress\_ultimate syl1\_πω syl2\_γα syl3\_α one\_transitive anim\_subj

κοιμάμαι

IC2a AC2 3syl stress\_ultimate syl1\_μαι syl2\_μα syl3\_κοι no\_transitive anim\_subj

Γράφημα 17: Δείγμα σχεδιασμού του ρηματικού μοντέλου για το ΣΕ του AMIS.

Παράλληλα, για το σχεδιασμό ενός καθολικού μοντέλου, το σώμα εκπαίδευσης (ΣΕ) του συστήματος είναι βαρύνουσας σημασίας. Στο πείραμα της Μέγιστης Εντροπίας το ΣΕ του μοντέλου ήταν βασισμένο σε λέξεις που δεν είχαν σχηματιστεί από διαδικασία παραγωγής ή σύνθεσης, ενώ το σώμα δοκιμής (ΣΔ) αποτελούταν αποκλειστικά από προθηματοποιημένες και επιθηματοποιημένες παράγωγες λέξεις. Είναι βασική προϋπόθεση για οποιοδήποτε μοντέλο πρόβλεψης και εποπτευόμενης μάθησης, τα σώματα κειμένων τόσο της εκπαίδευσης όσο και αυτό της δοκιμής να σχετίζονται άμεσα με τον στόχο και το σκοπό του συστήματος. Ο κανόνας για τα ΣΕ και ΣΔ είναι απλός, δηλαδή πρέπει να περιέχουν όμοιο υλικό και μην εσωκλείεται υλικό του ΣΕ στο ΣΔ, καθότι μία τέτοια επιλογή αποτελεί μη επιτρεπτή βοήθεια στο σύστημα. Εφόσον στόχος της παρούσας διατριβής είναι η αντιμετώπιση της αλλομορφίας στην παραγωγή, θα πρέπει τουλάχιστον να ενσωματωθούν και κάποια παράγωγα ουσιαστικά στο ΣΕ, καθότι επιβεβαιώθηκε από το AMIS ότι κυρίως οι τελευταίες συλλαβές είναι κρίσιμες για την πρόβλεψη της αλλομορφίας σε παραπλήσιες τάξεις (βλ. *νταής* vs. *μπακάλης*).

Αντιθέτως με όσα υποστηρίχτηκαν παραπάνω, η πλήρης απουσία παράγωγων λέξεων από το ΣΕ έγινε εσκεμμένα και έγκειται σε δύο καίρια σημεία. Πρώτον, αφού υποστηρίζουμε ότι η αλλομορφία συμμετέχει εξίσου σε όλες τις διαδικασίες, τα δεδομένα της μίας θα συνδράμουν και στις υπόλοιπες δύο. Δεύτερον, επιθυμούσαμε να παρέχουμε τη μικρότερη δυνατή βοήθεια για το σύστημα πρόβλεψης, ώστε να γίνει ακόμα καλύτερη εκτίμηση της σημαντικότητας του τελικού αποτελέσματος πρόβλεψης που ήταν πάνω από 91%. Το πείραμα πιστοποίησε πως η αλλομορφία της κλίσης συνδέεται άμεσα με την παραγωγή και αναμένεται να γίνει το ίδιο και με τη σύνθεση. Ένα σώμα εκπαίδευσης εντούτοις πρέπει να είναι ισορροπημένο, ώστε να περιέχει υλικό από όλες τις διαδικασίες σχηματισμού και να παρέχεται ένα εύρος

ικανοποιητικών δεδομένων για την ασφαλή δοκιμή και κρίση του συστήματος. Να τονίσουμε πως είμαστε πεπεισμένοι ότι εφόσον υπήρχαν και δεδομένα από παράγωγες λέξεις στο ΣΕ του πειράματος AMIS, τότε η συνολική πρόβλεψη της αλλομορφίας θα ήταν ακόμα υψηλότερη αγγίζοντας ποσοστά άνω του 96%, μίας και το προτεινόμενο μοντέλο θα είχε εκπαιδευτεί και από λέξεις που ανήκουν στην ίδια διαδικασία σχηματισμού λέξεων με εκείνες του ΣΔ.

Μια ολοκληρωμένη προσέγγιση της αλλομορφίας από ένα υπολογιστικό μοντέλο μορφολογικής ανάλυσης για κάθε διαδικασία σχηματισμού λέξεων οφείλει να περιέχει δείγματα από κάθε διαφορετική διαδικασία. Επομένως, ένα επαρκές ΣΕ θα χτιστεί πάνω σε κλιτικές, παράγωγες και σύνθετες λέξεις, ώστε να περιέχονται ομάδες λέξεων με διαφορετική μορφολογική δομή. Σίγουρα υπάρχουν κάποιες ενστάσεις για το μοντέλο, καθότι δεν δοκιμάστηκε με ρήματα και επίθετα. Για τα επίθετα δεν περιμένουμε ιδιαίτερες διαφορές: αντιθέτως αναμένεται το ποσοστό να αγγίζει το απόλυτο, αφού δεν υπάρχει τόση ποικιλία στην αλλομορφία, όπως στην κλίση, ενώ τα περισσότερα επίθετα εντάσσονται σε αλλομορφικές τάξεις για τις οποίες ο ΑλλοMantIS σκόραρε πάνω από 99%. Για τα ρήματα πιστεύουμε πως ο ΑλλοMantIS θα εξακολουθήσει να έχει υψηλό ποσοστό πρόβλεψης, αν και εδώ τα χαρακτηριστικά θα είναι τελείως διαφορετικά.

Η συνολική αρχιτεκτονική αυτής της πρότασης βασίζεται στην μοντελοποίηση ενός γραμματικού φαινομένου ιδιαίτερα έντονου στην Ελληνική γλώσσα έχοντας ως βάση σώματα κειμένων ή λέξεων. Σίγουρα η απουσία ελεύθερων προσβάσιμων σχολιασμένων σωμάτων δυσκολεύει το σχεδιασμό ενός τέτοιου μοντέλου, ωστόσο το συγκεκριμένο εμπόδιο δεν πρέπει να αποτελέσει τροχοπέδη για μία υπολογιστική προσέγγιση που επιτυγχάνει στην επεξεργασία του ζητήματος της αλλομορφίας με τεχνικές εποπτευόμενης μορφολογικής μάθησης.

Η επιτυχής πρόβλεψη της αλλομορφίας από ένα μοντέλο εποπτευόμενης μορφολογικής μάθησης αποτελεί αναμφίβολα ένα από τα δύο βασικά βήματα επίλυσης του προβλήματος της αλλομορφίας. Χωρίς την οριοθέτηση των περιβαλλόντων και καθοδήγηση για τις αλλομορφικές αλλαγές δεν είναι δυνατή η συσχέτιση δύο διαφορετικών μορφών του ίδιου μορφήματος και η επιλογή αλλομόρφου σε μη επιτρεπόμενο περιβάλλον δεν μπορεί να αποκλειστεί. Επομένως, κρίνεται απαραίτητο να συνδυαστούν οι δύο προσεγγίσεις, ώστε η αλλομορφία στην παραγωγή και κατ' επέκταση και στις υπόλοιπες διαδικασίες σχηματισμού λέξεων να γίνει αντικείμενο

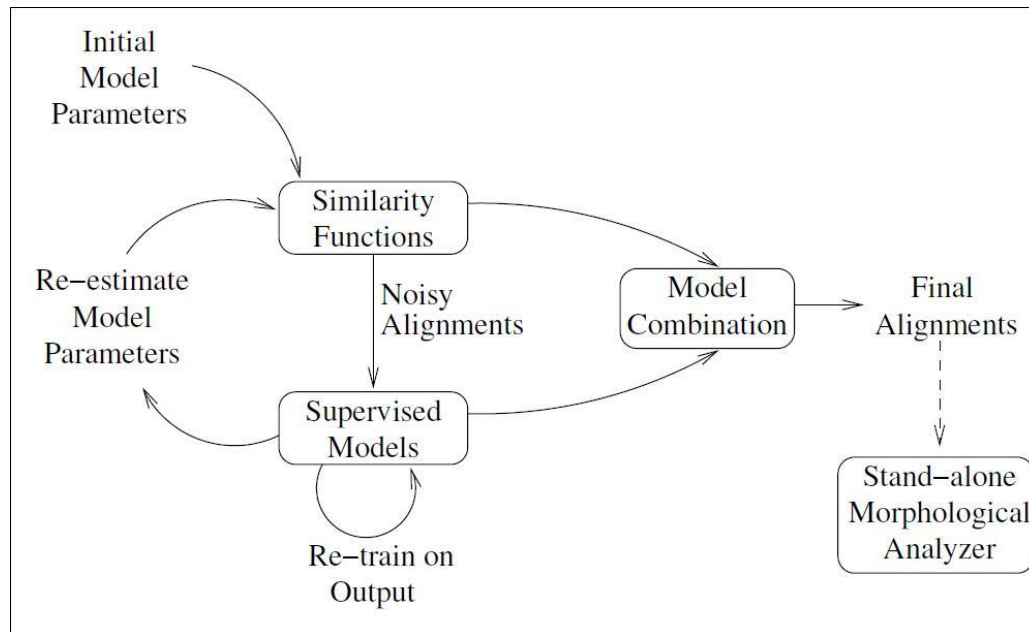
υπολογιστικής επεξεργασίας με επιτυχία. Ένας τέτοιος συνδυασμός θα αναπτυχθεί στην ενότητα που έπεται.

### 6.3.3. Συνδυαστικότητα μοντέλων προσέγγισης της αλλομορφίας

Το τρίτο μέρος του προτεινόμενου υπολογιστικού μοντέλου για την επεξεργασία της αλλομορφίας στοχεύει στην προσπάθεια συνδυασμού των δύο πρώτων μερών της πρότασής μας, δηλαδή της δημιουργίας κανόνων αλλομορφικής μετατροπής και του μοντέλου χαρακτηριστικών πρόβλεψης της αλλομορφίας. Κάθε μία πρόταση ως ανεξάρτητο θεωρητικό κομμάτι έχει από ένα σοβαρό μειονέκτημα και με τον συνδυασμό λειτουργούν οι δύο προτάσεις συμπληρωματικά εξαλείφοντας τις αδυναμίες. Οι υπολογιστικοί κανόνες αλλομορφικής μετατροπής υστερούν σε αυτοματοποίηση, καθώς πρέπει να αντιστοιχηθούν χειρωνακτικά σε κάθε μόρφημα (θέμα και πρόσφυμα) που εισάγεται στο λεξικό. Αντίστοιχα το μοντέλο τη πρόβλεψης της αλλομορφικής συμπεριφοράς δεν μπορεί να καθορίσει το περιβάλλον εμφάνισης των αλλομόρφων που προέβλεψε. Έτσι η συνδυαστικότητα των μοντέλων θα παρακάμψει τα αδύνατα σημεία των προτεινόμενων μοντέλων των προηγούμενων ενοτήτων, αφού το μοντέλο πρόβλεψης θα ενσωματώνει αυτόματα την αντιστοίχιση του κατάλληλου υπολογιστικού κανόνα σε κάθε μόρφημα του λεξικού, ενώ παράλληλα οι κανόνες θα παρέχουν όλες τις απαραίτητες πληροφορίες με τη συνδρομή του φίλτρου για τα περιβάλλοντα εμφάνισης των αλλομόρφων.

Όπως παρουσιάστηκε εμπειρικά στην ενότητα 5.1, τα μοντέλα μη-εποπτευόμενης μάθησης αδυνατούν να επεξεργαστούν με επιτυχία την αλλομορφία. Παράλληλα, το μοντέλο της Μέγιστης Εντροπίας φανέρωσε σημαντικά δείγματα προς τη σωστή κατεύθυνση για την αλλομορφική ανάλυση. Στις ενότητες 6.3.1 και 6.3.2 παρουσιάστηκαν δύο μοντέλα σε θεωρητικό επίπεδο, για το πώς μπορεί να αντιμετωπιστεί υπολογιστικά το φαινόμενο. Ωστόσο, όπως ορθώς επισημαίνει ο Wicentowski (2002) ο παραδοσιακός τρόπος συνδυασμού τεχνικών (δύο διαφορετικά μοντέλα σε παράλληλη ή σε σειριακή ακολουθία) μπορεί να εφαρμοστεί για τη συγχώνευση των δύο μοντέλων. Για την καλύτερη επίδοση των συστημάτων, το εξαγόμενο από τα συγχωνευμένα συστήματα μπορεί να χρησιμοποιηθεί για να επανεκπαιδευτούν κατ' επανάληψη οι παράμετροι του κάθε συστατικού του συστήματος. Αυτή τη στρατηγική συνιστά ο Wicentowski (2002, 2004) για την βελτίωση των επιδόσεων, μίας και το τελικό εξαγόμενο μπορεί να χρησιμοποιηθεί ως

θορυβώδη δεδομένα εκπαίδευσης για διάφορους αλγόριθμους εποπτευόμενης μάθησης.

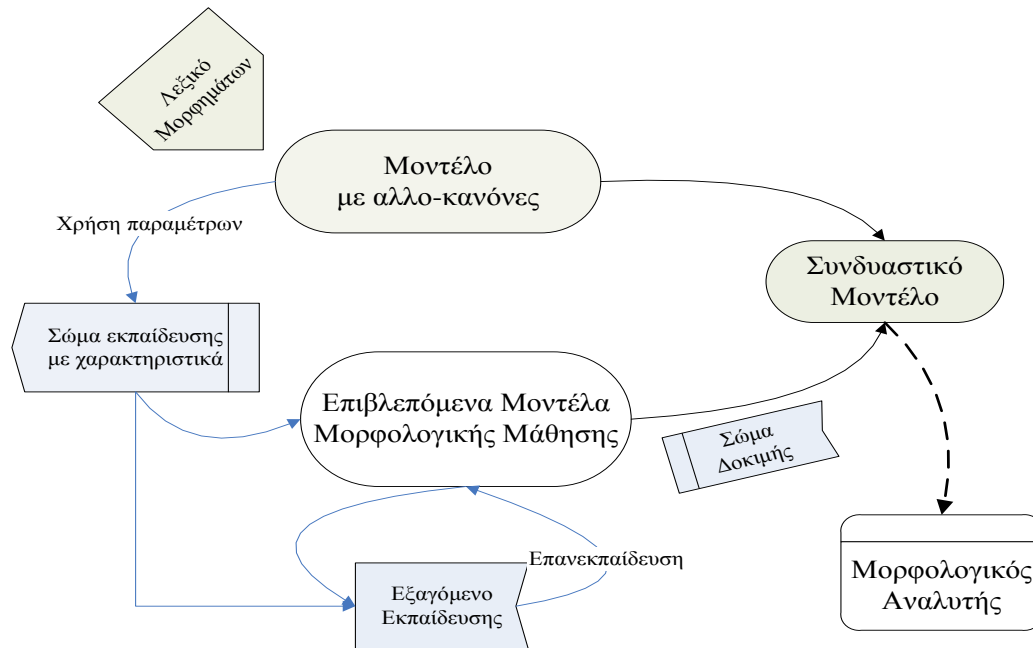


Γράφημα 18: Το μοντέλο επαναληπτικής επανεκπαίδευσης του Wicentowski (2002: 164).

Το παραπάνω γράφημα (19.) αποτελεί μία επισκόπηση του συνδυαστικού μοντέλου που δοκίμασε ο Wicentowski (2002, 2004) που δείχνει σχηματικά την γενική αρχιτεκτονική της επαναληπτικής επανεκπαιδευτικής επεξεργασίας. Αρχικά, οι πρωταρχικές παράμετροι του μοντέλου με τις συναρτήσεις ομοιότητας (similarity functions) χρησιμοποιούνται για τον καθορισμό των μοντέλων. Οι αναλύσεις για κάθε συνάρτηση ομοιότητας συνδυάζονται και χρησιμοποιούνται, για να εκπαιδεύσουν τα εποπτευόμενα μοντέλα. Τα συγκεκριμένα μοντέλα μπορούν να αξιοποιηθούν για την καλύτερη εκτίμηση των αρχικών παραμέτρων (re-estimate model parameters) του μοντέλου των συναρτήσεων ομοιότητας. Η τελική ανάλυση είναι ο συνδυασμός των επανακαθορισμένων συναρτήσεων ομοιότητας με το εξαγόμενο από τα κατ' επανάληψη εκπαιδευμένα επιβλεπόμενα μοντέλα.

Η δική μας προσέγγιση ελάχιστα παρεκκλίνει αρχιτεκτονικά από την πρόταση του Wicentowski (2002: 162-165): στόχος της συνδυαστικής πρότασής μας είναι η ομαλή σύνδεση δύο διαφορετικών μοντέλων για την επιτυχή υπολογιστική ανάλυση της αλλομορφίας. Οι διαφορές εντοπίζονται στον τρόπο δόμησης, αλληλεπίδρασης, αλλά και επιλογής των μοντέλων που θα αποτελέσουν τη βάση του συνδυασμού. Αρχικά από το λεξικό των μορφημάτων με την εφαρμογή των αλλο-κανόνων δημιουργούνται κάποιες από τις βασικές παραμέτρους που θα χρησιμοποιηθούν στο εποπτευόμενο

μοντέλο μάθησης. Με βάση την κατηγοριοποίηση των αλλο-κανόνων δομείται το σώμα εκπαίδευσης του συστήματος, ώστε να εκπαιδευτεί με τα σωστά δεδομένα και την κατάλληλη αντιστοίχιση με την αλλομορφική τάξη. Η εκτίμηση των επιλογών θα γίνει από το σώμα δοκιμής, ώστε οι ρυθμίσεις κάθε μοντέλου να είναι οι βέλτιστες δυνατές κατά την ενοποίησή τους στο συνδυαστικό μοντέλο.



Γράφημα 19: Το προτεινόμενο συνδυαστικό μοντέλο με βάση την επανεκπαίδευση σε επιβλεπόμενα μοντέλα μορφολογικής μάθησης.

Ο Wicentowski (2002, 2004) στην ανάλυση της πρότασής του υπογραμμίζει ότι το μεγαλύτερο πρόβλημα του συνδυασμού των μη-επιβλεπόμενων μοντέλων ομοιότητας (unsupervised similarity models) με τα επιβλεπόμενα είναι ότι το δυναμικό εύρος των διάφορων μετρήσεων είναι αρκετά διαφορετικό και όχι υποκείμενο σε απευθείας συνδυασμό (είτε πρόκειται για πολλαπλασιαστικό είτε για προσθετικό). Είναι κατανοητό ότι η δομή του συνδυαστικού προτεινόμενου μοντέλου καθίσταται ιδιαίτερα δύσκολη, όταν πρόκειται για δύο υπολογιστικές δομές, που είναι δομημένες και χτισμένες πάνω σε τελείως διαφορετικές στρατηγικές, ωστόσο έχουν ως κοινή σταθερά την εποπτευόμενη.

Για να επανέλθουμε στο πρόβλημα, είναι απίθανο να συνδυάσεις δύο μοντέλα πολλαπλασιαστικά, αν βασίζονται σε διαφορετικές προσεγγίσεις. Η τελική ανάλυση της παραγωγής και η προτεινόμενη ρίζα ή πρόσφυμα καθορίζεται από το μοντέλο των βαρών της μέγιστης εντροπίας και την εφαρμογή των κανόνων του άλλου μοντέλου. Σύμφωνα με τον Wicentowski (2002) ο συνδυασμός επιβλεπόμενων μοντέλων με

ρυθμιζόμενα βάρη μέσω της επανεκπαίδευσης βελτιώνει τη συνολική επίδοση του συστήματος. Εάν ένα μοντέλο μπορεί και προβλέπει τις αυθαίρετες αλλαγές της αλλομορφίας, αλλά δεν έχει το υπόβαθρο να συσχετίσει τα αλλόμορφα μεταξύ τους και να καθορίσει τα περιβάλλοντα εμφάνισης, τότε είναι επίσης καταδικασμένο να αποτύχει. Δηλαδή εάν το σύστημα καταφέρει να προβλέψει ότι η λέξη *κύμα* θα έχει αλλόμορφο *κυματ*, αλλά αδυνατεί να συσχετίσει τις μορφές *κύματα*, *κυματάκι*, *κυματίζω* με το *κύμα*, τότε θεωρούμε ότι έχει αποτύχει. Από την άλλη εάν ένα μοντέλο χρειάζεται τη συνεχή χειρωνακτική βοήθεια από τον χρήστη, ώστε σε κάθε καταχώρηση να υποδεικνύεται και ο αντίστοιχος αλλο-κανόνας, τότε το μοντέλο δεν απέχει και πολύ από την πολυπλοκότητα ενός αυτόματου συστήματος πεπερασμένων καταστάσεων. Στόχος είναι να καλύψει το ένα μοντέλο τις αδυναμίες του άλλου, ώστε συνεπικουρικά να προβάλουν μία επιτυχή λύση στην υπολογιστική ανάλυση της αλλομορφίας.

Η δομή του συνδυασμένου μοντέλου είναι βασισμένη στην αλληλουχία του εκάστοτε μοντέλου, το οποίο λειτουργεί παράλληλα με το άλλο και με δυνατότητα, όμως, παρεμβολής στις παραμέτρους. Η λογική του μοντέλου είναι ελαφρώς επηρεασμένη από το μινιμαλιστικό πνεύμα (Chomsky 1995) και από την προσέγγιση της Di Sciullo (2005) όπου στο πλαίσιο της ασυμμετρίας υποστηρίζεται ότι η μορφολογία, η σύνταξη, η φωνολογία κτλ. είναι παράλληλα υποσυστήματα στον υπολογιστικό χώρο της γλώσσας που συνήθως δεν είναι ορατά. Στο συγκεκριμένο μοντέλο εμείς χρησιμοποιούμε την ιδέα της παραλληλότητας στην αρχή του σχεδιασμού, αλλά στο τέλος έχουμε την ενοποίηση των δυο μοντέλων, όχι αθροιστικά αλλά παρατακτικά.

Το μοντέλο με τους αλλο-κανόνες χτίζεται με βάση το θεωρητικό πλαίσιο της Ralli (Ralli 1988, 2000, Ράλλη 2005), όπου η αλλομορφία αποτελεί βασικό κριτήριο κατηγοριοποίησης για ρήματα, ουσιαστικά και επίθετα και στο γενικότερο πλαίσιο των Handl *et al.* (2009). Κάθε κανόνας δημιουργεί ένα συγκεκριμένο αλλόμορφο από ένα υπάρχον μόρφημα, αλλά ταυτόχρονα συσχετίζει και δύο διαφορετικές μορφές του ίδιου μορφήματος. Η κωδικοποίηση των κανόνων είναι απαραίτητη για το μοντέλο της Μέγιστης Εντροπίας.

Το επιβλεπόμενο μοντέλο μορφολογικής μάθησης της Μέγιστης Εντροπίας χρησιμοποιεί για εκπαίδευσή του ένα σύνολο κλιτικών θεμάτων από κάθε γραμματική κατηγορία. Το σώμα εκπαίδευσης συνοδεύεται μαζί με όλες τις απαραίτητες πληροφορίες και παραμέτρους και από τον δείκτη της αλλομορφίας τους, όπως αυτός

ορίζεται από τους κανόνες. Αν και αυτές οι παράμετροι είναι γνωστές από πριν, οι αλλομορφικοί δείκτες σε κάθε μόρφημα δίνονται με βάση τους κανόνες, γιατί θα πρέπει να ταυτίζονται, καθότι αυτή η πληροφορία είναι κρίσιμη για το συνδυαστικό μοντέλο. Στη συνέχεια το μοντέλο εκπαιδεύεται με το συγκεκριμένο σώμα, γίνεται εκτίμηση των παραμέτρων και επανεκπαιδεύεται με το εξαγόμενο για την καλύτερη οριοθέτηση των ειδικών βαρών. Τέλος, δοκιμάζεται από ένα σώμα δοκιμής με παράγωγες λέξεις, για την εξακρίβωση της υψηλής επίδοσης στην πρόβλεψη του φαινομένου της αλλομορφίας.

Στη συνδυαστική στρατηγική των μοντέλων, η αρχική ανάλυση θα γίνει από το επιβλεπόμενο μοντέλο δημιουργώντας το λεξικό των μορφημάτων με τα απαραίτητα χαρακτηριστικά, ώστε να σχηματιστεί ένα σώμα δοκιμής, παραπλήσιο του σώματος εκπαίδευσης. Το εξαγόμενο του επιβλεπόμενου μοντέλου θα έχει προβλέψει τις αλλομορφικές μετατροπές, για να εισαχθούν τα αποτελέσματα στο μοντέλο με τους αλλο-κανόνες. Το μοντέλο θα δημιουργήσει τα αλλόμορφα σύμφωνα με την πρόβλεψη καταχωρώντας τα στο λεξικό και θα ελέγξει, ώστε να μην υπάρχουν μορφές του ίδιου μορφήματος σε διαφορετικά λήμματα. Τέλος, θα έχει τη δυνατότητα στην περίπτωση σχηματισμού λέξεων από τον μορφολογικό αναλυτή να παρέχει τους απαραίτητους δείκτες μορφολογικού περιβάλλοντος μέσω του φίλτρου για την ορθή πρόκριση του κατάλληλου αλλομόρφου.



Γράφημα 20: *Η εσωτερική δομή ενός μορφολογικού αναλυτή που βασίζεται στο συνδυαστικό μοντέλο.*

Το παραπάνω σχήμα παρουσιάζει την εσωτερική δομή και λειτουργία του συνδυαστικού μοντέλου με την υλοποίησή του ως μορφολογικού αναλυτή. Σε γενικές γραμμές έχοντας ήδη εκπαιδευτεί το εποπτευόμενο μοντέλο, ο μορφολογικός αναλυτής προβλέπει τον τύπο αλλομορφίας κάθε μορφήματος. Στη συνέχεια με τη χρήση των αλλο-κανόνων καταγράφονται όλα τα αλλόμορφα στο λεξικό για τη χρήση τους στις διάφορες διαδικασίες. Ενσωματώνοντας τους αλλο-κανόνες σε κάποιο ΜεΜΜ ή σε ελαφρώς εποπτευόμενο μοντέλο, η στρατηγική της υπολογιστικής ανάλυσης θα προσομοιάζει πολύ την αντίστοιχη του Allomorphessor που με αρκετή επιτυχία αντιμετώπισε ομάδα αλλομόρφων παρακάμπτοντας τα μικρά θέματα και προσφύματα. Με καλύτερο σχεδιασμό μία παραλλαγή του Allomorphessor ή ενός επιπρόσθετου εργαλείου σε ένα ΜεΜΜ, η αλλομορφική ανάλυση θεμάτων και προσφυμάτων θα έχει σαφώς πιο βελτιωμένα αποτελέσματα, αλλά μόνο στον τομέα της υπολογιστικής ανάλυσης των παραγώγων. Στον τομέα της υπολογιστικής ανασύνθεσης των παραγώγων δεν έχει καμία τύχη, αφού απουσιάζουν οι σχετικές πληροφορίες για το σχηματισμό παραγώγων.

Επανερχόμενοι στο σώμα εκπαίδευσης και έχοντας καθορίσει τις γενικές παραμέτρους του εποπτευόμενου μοντέλου γίνεται η επιλογή των δεδομένων για το σώμα εκπαίδευσης με τη συμμετοχή κλιτικών θεμάτων και την απουσία λοιπών λέξεων. Στη συγκεκριμένη μέθοδο επιλογής, μόνο κλιτικές ρίζες με την αφαίρεση των κλιτικών επιθημάτων εντάσσονται στο ΣΕ, ώστε να εκπαιδευτεί μόνο σε μία αλλομορφία κάθε φορά, καθότι στις παράγωγες λέξεις μπορεί να ενέχουν δύο ή και παραπάνω αλλόμορφα αναλόγως τον αριθμό των επιθημάτων (π.χ. γκρινι-[ι]αρ-ικ-ο, με το αλλόμορφο γκρινι~ από γκρίνια και [ι]αρ~ από -αρη). Ωστόσο θεωρούμε πως θα πρέπει να υφίσταται και μία μικρή ομάδα παραγώγων, για την καλύτερη εξισορρόπηση των ειδικών βαρών του συστήματος ιδιαίτερα για το χαρακτηριστικό των συλλαβών.

Εξάλλου, όπως φανέρωσε και το πείραμα με το AMIS, δεν χρειάζεται διαφορετική αντιμετώπιση η αλλομορφία των θεμάτων από εκείνη των επιθημάτων, αφού οι ομάδες λέξεων των ΣΕ και ΣΔ διαφοροποιούνται ως προς τη διαδικασία σχηματισμού· εντούτοις, το σύστημα εκπαιδευμένο με κλιτές λέξεις προέβλεψε την αλλομορφία του 91% των παραγώγων λέξεων. Για το συνδυαστικό μοντέλο του Wicentowski (2002) είναι απαραίτητος ο διαχωρισμός σε δύο μονάδες την WFBase και WFAffix· τα



συγκεκριμένα υπομοντέλα αντιμετωπίζουν ρίζες και προσφύματα αντίστοιχα, ενώ στα σημεία προσφύματοποίησης δεν καταχωρούνται οι πιθανότητες εμφάνισης. Οι πιθανότητες των αλλομορφικών αλλαγών δεν καταχωρίζονται εκεί, γιατί δεν εξαρτώνται από το περιβάλλον. Για τα Ελληνικά δεδομένα, δεν είναι απαραίτητος αυτός ο διαχωρισμός και επομένως είναι περιττή η παρουσία δύο διαφορετικών σωμάτων εκπαίδευσης· σε περίπτωση παρόμοιας στρατηγικής στα Ελληνικά τα αλλόμορφα εξαρτώνται για την εμφάνισή τους από το περιβάλλον που συμμετέχουν.

Ιδανικά, είναι εφικτό για ένα μορφολογικό αναλυτή να καλύπτει και τους δύο τομείς της υπολογιστικής μορφολογίας, την μορφολογική ανάλυση και την παραγωγή φυσικής γλώσσας. Ο πρωταρχικός στόχος είναι να είναι εξίσου επιτυχημένα και τα δύο υποσυστήματα του μορφολογικού αναλυτή. Για το δεύτερο υποσύστημα η βελτίωση δεν είναι εφικτή μέσω μη-επιβλεπόμενων τεχνικών, αλλά με σαφή καθορισμό των κανόνων και του φίλτρου μορφολογικού περιβάλλοντος από τον σχεδιαστή. Αντιθέτως, για το πρώτο υποσύστημα, η επαναληπτική επανεκπαίδευση στοχεύει στη βελτίωση του πυρήνα του μορφολογικού μοντέλου, το οποίο δεν εξυπηρετεί μόνο το αρχικό εποπτευόμενο μοντέλο των χαρακτηριστικών, αλλά και έχει άμεσες επιπτώσεις στο συνολικό συνδυαστικό μορφολογικό μοντέλο. Καθώς προχωρούν οι επαναλήψεις, το επιβλεπόμενο μοντέλο επανεκπαιδύεται στο εξαγόμενο προηγούμενων επαναλήψεων και έτσι φιλτράρονται καλύτερα τα βάρη, καθαρίζοντας το σύστημα από περιττούς θορύβους. Η τελική ανάλυση είναι εκείνη που βασίζεται στον πληρέστερο κύκλο βελτιστοποιημένης επανεκπαίδευσης, η οποία τερματίζει στο σημείο που δεν επιδέχεται επιπλέον καλύτερευση. Για το μοντέλο του Wicentowski (2002: 166-176) η βέλτιστη επίδοση επιτυγχάνεται μετά την πέμπτη ή έκτη επανάληψη.

Επιπροσθέτως, οφείλουμε να συλλογιστούμε και τις πιθανές βελτιώσεις που μπορεί να παρέχουν σε ένα συνδυασμένο σύστημα μία σύμπλεξη από διαφορετικές τεχνικές της εποπτευόμενης μορφολογικής μάθησης. Ο Wicentowski (2002) εντός του συνδυασμένου μοντέλου χρησιμοποιεί διαφορετικά επιβλεπόμενα μοντέλα. Θα ήταν ενδιαφέρον να ελέγξουμε πόσα παρόμοια επιβλεπόμενα μοντέλα μπορούν να συνδράμουν στη συνολική προσπάθεια αντιμετώπισης της αλλομορφίας. Κάποια μοντέλα βασίζονται στην παραδοχή ότι η εκμάθηση είναι η αποθήκευση παραδειγμάτων στη μνήμη και η επεξεργασία είναι η αναλογική συσχέτιση των καταχωρημένων παραδειγμάτων (βλ. Saussure 1916, Bloomfield 1933).

Δύο ιδιαίτερα επιτυχημένες προσπάθειες στα επιβλεπόμενα μοντέλα μάθησης είναι αυτές των Keuleers & Daelemans (2007) και των Yvon & Stroppa (2007), όπου οι πρώτοι βασίζονται στην τεχνική του *k-πλησιέστερου-γείτονα* (k-nearest-neighbor) ενός μοντέλου μνημονικής μάθησης και οι δεύτεροι στην τεχνική της *μάθησης καθοδηγούμενης από παραδείγματα* (Paradigm-Driven Learning) σε μοντέλο αναλογικής μάθησης. Παλαιότερες προσπάθειες έδειξαν ότι τα μοντέλα μνημονικής μάθησης αντιμετωπίζουν με μεγάλη επιτυχία προβλήματα μη-κανονικότητας στην κλίση, όπως την αντιμετώπιση του αορίστου στα Αγγλικά (Eddington 2000), την πρόβλεψη του ονοματικού πληθυντικού στα Γερμανικά (Hahn & Nakisa 2000), την ανάλυση του ονοματικού πληθυντικού των Ολλανδικών (Keuleers *et al.* 2007) μεταξύ άλλων. Ταυτόχρονα προκαταρκτικά πειράματα των Pirrelli & Yvon (1999a, 1999b) και Stroppa & Yvon (2005) κατέδειξαν ότι οι ιδιαίτερα παραγωγικές περιπτώσεις της παραγωγικής επιθηματοποίησης μπορούν να αναλυθούν από ένα μοντέλο αναλογικής μάθησης· επί της αρχής, οι Yvon & Stroppa (2007: 223) βεβαιώνουν ότι μπορεί να επιτύχει σε λιγότερο παραγωγικές μορφολογικά πολύπλοκες μορφές με τακτικές όπως αντίστροφους-σχηματισμούς και «δημιουργικές» αναλογίες. Εντούτοις, εξαιτίας της σποραδικότερης φύσης<sup>24</sup> των «παραγωγικών παραδειγμάτων», φαίνεται να απαιτείται ένα πιο δυνατό μοντέλο από αυτό των τυπικών αναλογιών για την επίτευξη μίας καλύτερης γενικής επίδοσης.

Όπως σημειώνουν και οι συγγραφείς των δύο προσεγγίσεων (Keuleers & Daelemans 2007, Yvon & Stroppa 2007), τα γενικότερα μοντέλα πρέπει να τεθούν υπό δοκιμή σε περισσότερα γλωσσικά δεδομένα από γλώσσες με διαφορετική μορφολογία από αυτές που δοκιμάστηκαν. Σίγουρα αποτελεί εξαιρετικά ενδιαφέρον να δοκιμαστούν οι συγκεκριμένες θεωρίες των επιβλεπόμενων μοντέλων στα Ελληνικά δεδομένα. Παρόλη τη μη δοκιμή τους στα Ελληνικά, τα ανωτέρω μοντέλα εποπτευόμενης μάθησης έχουν τη δυνατότητα να εκπαιδευτούν με τα ίδια δεδομένα και να συμβάλουν στην υλοποίηση ενός συνδυαστικού μοντέλου που να εκμεταλλεύεται τα πλεονεκτήματα κάθε επιβλεπόμενης τακτικής. Με άλλα λόγια, τα βάρη από τα επιβλεπόμενα μοντέλα μετά από κάθε επανάληψη θα επιλέγονται με άμεσο στόχο την μεγιστοποίηση της επίδοσης του συστήματος. Σε πρακτικό επίπεδο

---

<sup>24</sup> Χαρακτηρίζεται έτσι καθότι δεν έχει τη συστηματικότητα των κλιτικών παραδειγμάτων, αφού όλα τα ουσιαστικά που κλίνονται με τον ίδιο τρόπο, δεν αναμένεται να εμφανίσουν τα ίδια παραγωγικά παραδείγματα, αλλά αντιθέτως να εμφανίζουν διάφορες σποραδικές παραγωγικές ομάδες.

έχει δοκιμαστεί με αρκετή επιτυχία σε γλώσσες με αρκετή επιθηματοποίηση από τον Wicentowski (2002), αν και τα μοντέλα δεν ομοιάζαν των παραπάνω.

*“We are not certain, we are never certain. If we were we could reach some conclusions and we could, at last, make others take us seriously”*

Albert Camus

*“Please be good enough to put your conclusions and recommendations on one sheet of paper in the very beginning of your report, so I can even consider reading it”*

Winston Churchill

## 7. Συμπεράσματα

### 7.1. Γενική επισκόπηση

#### 7.1.1. Η αλλομορφία ως σύγχρονο μορφολογικό φαινόμενο

Η παρούσα διατριβή διερεύνησε σε θεωρητικό επίπεδο το ζήτημα της καθαρής μορφολογικής αλλομορφίας, αυτής που δεν καθορίζεται από μορφολογικές και φωνολογικές συνθήκες και είναι ένα σύγχρονο κατάλοιπο της ιστορικής εξέλιξης και ανενεργών κανόνων (βλ. εν. 2.6, όπως υποστηρίζει η πλειονότητα των ερευνητών), ενώ όπως σημειώνει και η Ralli (2007) αποτελεί μία βασική ιδιότητα της μορφολογικής δομής. Υποστηρίξαμε ότι η αλλομορφία διαδραματίζει σημαντικό ρόλο σε όλες τις διαδικασίες σχηματισμού λέξεων και ειδικότερα στην παραγωγή και ενέχει συστηματικά σχεδιάτυπα που κρύβονται σε διάφορους σχηματισμούς.

Κατά την ανάλυση των αλλομόρφων εντοπίστηκαν δύο περιορισμοί με βάση την αλλομορφία που αποτελούν σημαντικούς αρωγούς στην προσπάθεια υπολογιστικής αντιμετώπισής της. Η αρχής της αλλομορφικής συμπεριφοράς υπαγορεύει ότι το πεδίο εμφάνισης όλων των αλλομόρφων είναι η κλίση με μοναδικές και συστηματικές

εξαιρέσεις τα ουσιαστικά τύπου *παιδί* και *μπακάλης*. Ο παραπάνω περιορισμός φανέρωσε και την άρρηκτη σχέση των διαδικασιών μεταξύ τους, αλλά και τον τρόπο που αποτελεί η αλλομορφία έναν ισχυρό συνδετικό κρίκο. Ταυτόχρονα, ο περιορισμός του μοναδικού αλλομόρφου συμμετέχοντα αποκάλυψε τη συστηματική χρήση ενός και μόνο ονομαστικού και επιθετικού αλλομόρφου στην παραγωγή με ελάχιστες εξαιρέσεις στη ρηματική παραγωγή. Αυτός καθιστά την συστηματοποίηση της αλλομορφίας πιο εύκολη, γεγονός σημαντικό για ένα φαινόμενο που αποτελεί για τους ερευνητές ζήτημα προς αποφυγή.

Παράλληλα, σε θεωρητικό επίπεδο αποκαλύψαμε πως η αλλομορφία παρόλο που χαρακτηρίζεται από απρόβλεπτες και αυθαίρετες αλλαγές των μορφημάτων, ως φαινόμενο χαρακτηρίζεται από κανονικότητα και προβλεψιμότητα (βλ. εν. 2.7), της οποίας η αιτιολόγηση βρίσκεται πίσω από την ιστορική εξέλιξη της γλώσσας. Η συστηματικότητα της αλλομορφίας δίνει σε θεωρητικό επίπεδο αρκετά επιχειρήματα για το ζήτημα της αυτονομίας της μορφολογίας, ενώ σε πρακτικό επίπεδο και συγκεκριμένα σε υπολογιστικό μάς προστατεύει από πολύπλοκες στρατηγικές επίλυσης, ενώ παρέχει και κοινά χαρακτηριστικά σε όλες τις διαδικασίες που επιτρέπουν την επεξεργασία της ως συνολικό φαινόμενο και όχι ως ένα πρόβλημα με διαβαθμισμένες διαδικασιακές αλλαγές.

### **7.1.2. Μη-Εποπτευόμενα Μοντέλα Μορφολογικής Μάθησης για την Αλλομορφία**

Η Υπολογιστική Μορφολογία είναι ένας ταχύτατα αναπτυσσόμενος τομέας της Υπολογιστικής Γλωσσολογίας. Η θεωρία της Μη-Εποπτευόμενης Μάθησης της Μορφολογίας είναι μία πρόσφατη προσέγγιση στη μορφολογική ανάλυση των προβλημάτων και φαίνεται να επιτυγχάνει ικανοποιητικές επιδόσεις σε γλώσσες με φτωχή κλιτική μορφολογία, με σχετικά απλές διαδικασίες παραγωγής και περιορισμένη σύνθεση. Αντίθετα, οι προσπάθειες εφαρμογής της θεωρίας σε μορφολογικά πλούσιες γλώσσες, όπως είναι τα Φιλανδικά και τα Τουρκικά, χαρακτηρίζεται με επιείκεια ως μετριότατη προσπάθεια, αναλογιζόμενοι και τις επιδόσεις διαφορετικών υπολογιστικών μοντέλων (για γενικές επιδόσεις, βλ. Kurimo *et al.* 2006a, 2006b, Kurimo, Creutz & Varjokallio 2007, Kurimo, Virpioja & Turunen 2009, ενώ ειδικότερα για την αλλομορφική τους συμπεριφορά βλ. Kohonen, Virpioja & Klami 2009, Virpioja & Kohonen 2009, Kurimo *et al.* 2010). Τα παραπάνω αποτελέσματα επιβεβαιώθηκαν

με ακόμα χαμηλότερες επιδόσεις για την Ελληνική γλώσσα έχοντας δοκιμαστεί σε τρία διαφορετικά σώματα κειμένων (βλ. ενότητα 5.1). Υποστηρίζουμε με βεβαιότητα ότι μία προσέγγιση υπολογιστικής μορφολογικής υλοποίησης χωρίς: α.) την εκ των προτέρων χειρωνακτική ανάλυση κάποιων μορφημάτων της υπό εξέταση γλώσσας, β.) τον εντοπισμό αρκετών προσφυμάτων και θεμάτων και γ.) (προ)εισαγμένους μορφολογικούς και φωνολογικούς κανόνες είναι καταδικασμένο να αποτύχει ιδίως σε πλούσιες μορφολογικά γλώσσες που δεν ανήκουν στην ομάδα των συγκολλητικών. Ένα σύστημα που χτίζει λεξικά (κανονικά και αντίστροφα) βασισμένο σε ακολουθίες φωνημάτων και σε συστηματικές αντικαταστάσεις, προσθήκες και αφαιρέσεις χωρίς τους κατάλληλους κανόνες δεν έχει τη δυνατότητα να αντιμετωπίσει με επιτυχία τις πολύπλοκες και συνδυαστικές συμπεριφορές των παραγωγικών επιθημάτων, δεσμευμένων θεμάτων και θεμάτων/ ριζών. Ειδικότερα, όπως έχει αποδειχθεί το φαινόμενο της αλλομορφίας στα Ελληνικά είναι ιδιαίτερα εκτεταμένο. Η αλλομορφία συμμετέχει με την ίδια συχνότητα και στατιστική διασπορά σε κάθε διαδικασία σχηματισμού λέξεων. Έτσι εγείρεται η φυσική ερώτηση εάν ένα ΜεΜΜ μοντέλο έχει τη δυνατότητα να αναλύσει και να αντιμετωπίσει επιτυχώς την προσφυματική και θεματική αλλομορφία στην Ελληνική.

Παρουσιάσαμε ένα σημαντικό μέγεθος δεδομένων με αλλόμορφα σε θεωρητικό και υπολογιστικό επίπεδο και καταδείξαμε την περιπλοκότητα των αλλομορφικών αλλαγών, την υψηλή συνδυαστικότητα των παραγωγικών επιθημάτων και την κανονικότητα των απλών και δεσμευμένων θεμάτων. μίας και η εισαγωγή κανόνων οιασδήποτε μορφής απαγορεύεται για ένα αμιγές ΜεΜΜ μοντέλο, τότε ο στόχος της επιτυχούς αλλομορφικής αντιμετώπισης δεν θα επιτευχθεί ποτέ. Από μία πιο θεωρητική οπτική γωνία, η παρούσα εργασία δεν είχε να κάνει με το καίριο ερώτημα που εγείρεται από τους ερευνητές (για περισσότερα πρβλ. Pirrelli 2007a): ένας νέος ομιλητής μίας γλώσσας μαθαίνει τη γλώσσα και διαχωρίζει τα μορφήματα με τρόπο παρόμοιο ενός ΜεΜΜ μοντέλου;

Η παρούσα διατριβή αιτιολόγησε θεωρητικά τις αδυναμίες ενός ΜεΜΜ μοντέλου και πιστοποίησε και πειραματικά πως στις περιπτώσεις που απουσιάζουν τα δεδομένα των σωμάτων εκπαίδευσης τα μη-επιβλεπόμενα μοντέλα αδυνατούν να διαχειριστούν με επιτυχία όχι μόνο το υπολογιστικά ιδιαίτερος απαιτητικό φαινόμενο της αλλομορφίας, αλλά και τη διαδικασία της παραγωγής. Πιθανότατα, τα συγκεκριμένα μοντέλα να αποδίδουν καλύτερα με τα κλιτικά παραδείγματα, καθώς όλες οι επιτυχίες

που έχουν καταχωρηθεί στη διεθνή βιβλιογραφία (μεταξύ άλλων Goldsmith 2001, Wicentowski 2002, Kurimo *et al.* 2010) αφορούν στη διαδικασία της κλίσης. Δεν είναι τυχαίο το γεγονός ότι η παραγωγή απουσιάζει σχεδόν πάντα από τις υπολογιστικές αναλύσεις λόγω των σποραδικών παραδειγμάτων και την έλλειψη συστηματικής ομοιογένειας. Το πείραμα με τη δημοφιλέστερη υλοποίηση της MeMM θεωρίας, το LINGUISTICA, κατέδειξε τη δυσκολία ανάλυσης της περιοχής της παραγωγής και μάλιστα υπό το πρίσμα της αλλομορφίας (βλ. εν. 5.1.5). Κλείνοντας, αυτό που θέλουμε να υπογραμμίσουμε με βεβαιότητα είναι ότι μόνο τα επιβλεπόμενα μορφολογικά μοντέλα μάθησης με κανόνες και εισαγμένη ανθρώπινη γνώση μπορούν να υπηρετήσουν ως βάση για την υπολογιστική επεξεργασία του μορφολογικού φαινομένου της αλλομορφίας στην παραγωγή των Νέων Ελληνικών.

### **7.1.3. Εποπτευόμενη Μορφολογική Μάθηση για τη Μορφολογική Ανάλυση**

Η ύπαρξη μορφολογικά σχολιασμένων σωμάτων κειμένου επιτρέπει μία σαφώς επιτυχή και σημαντική ανάλυση μίας γλώσσας από ένα επιβλεπόμενο μοντέλο μορφολογικής μάθησης. Ένα υπολογιστικό μοντέλο κατάκτησης της μορφολογίας δεν πρέπει να αξιολογείται μόνο πάνω στη βάση του βαθμού ακρίβειας των προβλέψεών του, αλλά παράλληλα να συγκρίνονται οι ικανότητές του να αντιγράφει τις γενικεύσεις ενός φυσικού ομιλητή κατά την κατάκτηση και επεξεργασία της μορφολογίας (Yvon & Stroppa 2007: 203). Είναι εμφανές ότι το δεύτερο σκέλος μίας τέτοιας παραδοχής είναι αρκετά δύσκολο, εντούτοις η επιλογή των μοντέλων εποπτευόμενης μάθησης έγιναν με βάση αυτό το κριτήριο, αφού επιτρέπουν την εισαγωγή κανόνων, ανθρώπινης γνώσης, καθώς και εκπαίδευσης του συστήματος.

Ένα σώμα λέξεων με προσεκτικά επιλεγμένες λέξεις δημιουργήθηκε για να δοκιμαστεί η δυνατότητα ενός εποπτευόμενου μοντέλου Μέγιστης Εντροπίας στην αντιμετώπιση της αλλομορφίας. Η επιλεκτική χρήση κλιτών λέξεων για εκπαίδευση σε σύγκριση με τις παράγωγες λέξεις του σώματος δοκιμής προσέδωσαν μεγάλη εγκυρότητα στο τελικό υψηλό ποσοστό πρόβλεψης της αλλομορφίας, ενός συστήματος που δεν επανεκπαιδεύτηκε από το εξαγόμενό του και δεν περιελάβανε παραπλήσια σώματα στην εκπαίδευση και τη δοκιμή. Είναι αξιοσημείωτο ότι δοκιμάστηκε ένα σύστημα που στοχεύει κυρίως σε συντακτικές εφαρμογές, ωστόσο η τροποποίησή του σε μορφολογικά δεδομένα λειτούργησε ιδιαίτερα αποτελεσματικά.

Η υλοποίηση του ΑΛΛΟΜΑΝΤΙΣ βασίστηκε σε μία προσεχτικά διαλεγμένη συλλογή κλιτών εισαγόμενων με τις κατάλληλες γλωσσολογικές παραμέτρους, ένα σύνολο άγνωστων εισαγόμενων παράγωγων λέξεων και μία λίστα από τα κατάλληλα βάρη για την τελική πρόβλεψη. Οι γενικεύσεις του εξαγόμενου περιείχαν μία υψηλή πρόβλεψη για την αλλομορφία παρουσιάζοντας παράλληλα και δύο βασικές ιδιότητες του φαινομένου, μία συστηματική κανονικότητα και μία «μη αναμενόμενη» προβλεψιμότητα. Η ενδεδειγμένη παρατήρηση των αποτελεσμάτων (βλ. 5.2.4.2) φανέρωσε πως οι μικρές ομάδες μορφημάτων με ξέχωρη διαφορά σίγουρα αποτελούν μία ανοικτή πληγή για κάθε (σχεδόν) υπολογιστικό μοντέλο. Στην παρούσα διατριβή δεν έγινε προσπάθεια για χρήση της τακτικής της επανάληψης των εξαγόμενων, αλλά εντούτοις δοκιμάστηκε η αξιολόγηση των παραμέτρων με αποτέλεσμα στη δεύτερη προσπάθεια του ο ΑΛΛΟΜΑΝΤΙΣ να σημειώσει καλύτερη επίδοση, αλλάζοντας τη σειρά χαρακτηρισμού των συλλαβών.

Με τα μοντέλα ΜεΜΜ να αδυνατούν στο γενικό τους πλαίσιο να χαρακτηριστούν ως επαρκώς αποδοτικά για να πραγματοποιήσουν προβλέψεις σχετικά με τις μορφολογικές (και συντακτικές) ιδιότητες που συνοδεύουν το κάθε μόρφημα (Yvon & Stoppa 2007) αναζητούμε τα επιβλεπόμενα μοντέλα που θα ταιριάζουν περισσότερο στο ζήτημα της Ελληνικής, αφού κάποιες γλώσσες απαιτούν διαφορετική προσέγγιση σύμφωνα με τους Yvon & Stoppa (2007). Αντιθέτως ο Wicentowski (2002: 193) δεν παρατήρησε σημαντικές διαφορές στις επιδόσεις των μοντέλων σε εύρος δοκιμής 32 γλωσσών. Φαίνεται πως η λύση κρύβεται πίσω από τις συνδυαστικές ιδιότητες των μοντέλων για την τελική επικράτηση.

Η εποπτευόμενη προσέγγιση μας φανέρωσε διάφορα σχήματα που κρύβονται πίσω από τον συνδυασμό ονοματικής κλίσης και παραγωγής και μας υποψιάζει για πιθανά αντίστοιχα σχεδιάτυπα στη ρηματική κλίση και παραγωγή. Επιπροσθέτως, δείχνει να αποτελεί λύση, όχι μόνο σε επίπεδο θεωρίας και αρχών, αλλά με πρακτικό και συνάμα αποτελεσματικό τρόπο για το ζήτημα της αλλομορφίας που δεν της έχει δοθεί η δέουσα σημασία στο θεωρητικό της κομμάτι, ενώ στην υπολογιστική επεξεργασία χαρακτηρίζεται ως περιορισμένο φαινόμενο έλλειψης κανονικότητας και ομαλότητας τύπων με ελάχιστη έκταση. Είναι σαφές ότι τέτοια συστήματα δεν απέχουν πολύ από τις απαιτήσεις των ψυχολογολόγων για προσομοίωση της ανθρώπινης εκμάθησης της γλώσσας, ενώ παράλληλα δεν παρεκκλίνουν ιδιαίτερα από την μορφολογική θεωρία που υιοθετούμε για την παραγωγή και την αλλομορφία.



#### 7.1.4. Η προβληματική για την υπολογιστική επεξεργασία της Αλλομορφίας και τα συνδυαστικά μοντέλα

Είναι αξιοσημείωτη η συστηματική απουσία υπολογιστικής επεξεργασίας της αλλομορφίας από τους ερευνητές, ενώ αρκετές φορές χαρακτηρίζεται ως κατάλοιπο, ως μία ανωμαλία του συστήματος που απλώς διαταράσσει την άριστη λειτουργία ενός μορφολογικού αναλυτή. Η κύρια ιδιότητα των αλλομόρφων να διαφοροποιούνται με *απρόβλεπτο και αυθαίρετο* τρόπο αποκλείει την εφαρμογή πραγματικών κανόνων για την αντιμετώπισή τους. Καθίσταται παράλληλα σαφές ότι μία αυτοματοποιημένη τεχνική προσπαθεί να επιτύχει κάτι οξύμωρο, να προβλέψει κάποια μορφήματα που έχουν αλλάξει *απρόβλεπτα!* Ακόμα και οι περιπτώσεις των «αλλομόρφων» που εξαρτώνται από φωνολογικές συνθήκες αλλοιώνουν τις συνολικές επιδόσεις και σε περιπτώσεις που καταφεύγουν σε επιβλεπόμενες τακτικές (Daelemans, Berck & Gillis 1996).

Συχνά μετατοπίζεται το πρόβλημα της αλλομορφίας σε μελλοντικές εκδόσεις των αναλυτών και ο τρόπος αντιμετώπισής της στους ερευνητές της υπολογιστικής μορφολογίας δείχνει αδιαφορία για τα προβλήματα που προκαλεί. Από την αντίπερα όχθη οι θεωρητικοί μορφολόγοι δεν δείχνουν την απαιτούμενη προσοχή στο φαινόμενο. Η ευθύνη για την επεξεργασία της αλλομορφίας πρέπει να επωμιστεί και στους υπολογιστικούς γλωσσολόγους, αλλά παράλληλα και στους μορφολόγους, ώστε να είναι εφικτή η σωστή ανάλυση του φαινομένου σε θεωρητικό και υπολογιστικό επίπεδο. Δεν είναι δίκαιο να απαιτείται η αντιμετώπιση της αλλομορφίας υπολογιστικά από τη στιγμή που δεν αποτελεί συστηματικό αντικείμενο μελέτης από τους θεωρητικούς μορφολόγους, χαρακτηρίζεται ως περιθωριακό φαινόμενο και υφίσταται ασάφεια για το τι χαρακτηρίζεται ως πραγματικό αλλόμορφο και τι φωνολογικό αλλόμορφο. Μάλιστα τείνει τα τελευταία χρόνια λόγω της πληθώρας των ερευνών της αλλομορφίας υπό το πρίσμα της Θεωρίας του Βέλτιστου (μεταξύ άλλων Drachman, Kager & Malikouti-Drachman 1995, Bonet 2004, Paster 2005, Kikuchi 2006, Lubowicz *et al.* 2006, Bye 2007, Bonet & Mascaró 2007, Bonet & Harbour 2010, Nevins 2010, Suh 2010, van der Veer & Booij to appear) η αλλομορφία να θεωρείται φωνολογικό φαινόμενο ή ότι οι αλλαγές του κυβερνούνται από φωνολογικές συνθήκες και περιβάλλοντα. Από την άλλη πλευρά, τα μοντέλα υπολογιστικής μορφολογίας αποτυγχάνουν στην ανάλυσή της και επειδή σχεδόν πάντα στιγματίζουν τη συνολική

εικόνα μίας καλής επίδοσης επιλέγουν συνειδητά να μην ασχοληθούν με το φαινόμενο ή να το αποκλείουν από οποιαδήποτε διαδικασία αντιμετώπισης και αξιολόγησης.

Η αναβάθμιση της αλλομορφίας σε μείζον ζήτημα της υπολογιστικής ανάλυσης θα επιφέρει καίριες αλλαγές στον τρόπο διαχείρισής της, αλλά παράλληλα θα φέρει στο προσκήνιο ένα σημαντικό πρόβλημα που τόσα χρόνια «κρύβεται κάτω από το χαλί». Ο χαρακτηρισμός ενός αναλυτή ως επιτυχημένου δεν μπορεί να βασίζεται σε μισά δεδομένα ή σε επιλεγμένα δεδομένα προσπαθώντας πάντοτε να καλυφθούν οι ομαλές και συστηματικές ομάδες μορφημάτων ή τα (κλιτικά και σπανίως παραγωγικά) παραδείγματα με κανονικότητες και προβλέψιμες φωνολογικές αλλαγές. Πρέπει να γίνει κατανοητό πως η αλλομορφία είναι αναπόσπαστο μέρος της μορφολογικής δομής όλων των σχηματισμών στα Ελληνικά και έχει την ίδια βαρύνουσα σημασία με τις διάφορες φωνολογικές και μορφολογικές αλλαγές που υφίστανται στους συνδυασμούς μορφημάτων.

Ένα μεμονωμένο μοντέλο με αρκετά ή πλήρως αυτοματοποιημένες τεχνικές αδυνατεί να φέρει εις πέρας το δύσκολο στόχο της αλλομορφικής ανάλυσης. Είναι βέβαιο ότι ένα μοντέλο, όπως τα αυτόματα πεπερασμένων καταστάσεων, όπου το μεγαλύτερο μέρος της υπολογιστικής γνώσης για τη γλώσσα προέρχεται από προ-εισαγμένη ανθρώπινη γνώση και μορφοφωνολογικούς κανόνες μπορεί να διαχειριστεί με επιτυχία την αλλομορφία· ωστόσο τέτοια συστήματα χαρακτηρίζονται τρομερά πολύπλοκα και δύσχρηστα. Τέτοιο σύστημα είναι ο PC-KIMMO, που εξακολουθεί τριάντα χρόνια μετά να αποτελεί *state-of-the-art*, αλλά γίνεται πάντα αναφορά για την πολύπλοκη και αντιοικονομική δομή του.

Κοινή στρατηγική των λιγοστών προσπαθειών για την υπολογιστική ανάλυση του φαινομένου είναι να συνδυάζουν τακτικές διαφορετικών μοντέλων ή προσεγγίσεων, για να επιτύχουν το σκοπό τους. Οι συγγραφείς του Allomorphessor εντάσσουν τακτικές εποπτευόμενης μορφολογικής μάθησης σε ένα ΜεMM μοντέλο, ενώ οι συγγραφείς του JSLIM συνδυάζουν θεωρητικές προσεγγίσεις και έχουν σαφείς επιρροές από διάφορες στρατηγικές των πεπερασμένων συστημάτων. Υποστηρίξαμε πως τα επιβλεπόμενα μοντέλα μορφολογικής μάθησης έχουν τη δυνατότητα κυρίως λειτουργώντας συνδυαστικά να επιτύχουν τον δύσκολο στόχο, καθότι έτσι καλύπτει το ένα μοντέλο τις αδυναμίες του άλλου και συνεπικουρικά αυξάνουν τα ποσοστά σωστής μορφολογικής ανάλυσης. Εξάλλου, είναι αναμενόμενη αυτή η τακτική, αφού τα αλλόμορφα υπήρξαν παλαιότερα προϊόντα φωνολογικών (κυρίως) κανόνων, αλλά

εντάσσονται συγχρονικά στη μορφολογία. Από θεωρητικής άποψης απαιτείται η συνδρομή και άλλων επιπέδων στη γλώσσα, έτσι και από υπολογιστικής άποψης κρίνεται απαραίτητη μία προσέγγιση με συνδυαστικό μοντέλο υπολογιστικών προσεγγίσεων, ώστε να επιτύχουμε την υπολογιστική επεξεργασία της αλλομορφίας με όσο το δυνατόν απλό, αυτοματοποιημένο, επιτυχημένο και ακριβή τρόπο

## **7.2. Μελλοντική εργασία**

### **7.2.1. Επαναληπτική μάθηση και διεύρυνση του ΑλλοMantIS**

Στη διεθνή βιβλιογραφία της Μηχανικής Μάθησης γίνεται συνεχής αναφορά για την αποτελεσματικότητα των επαναληπτικών επανεκπαιδεύσεων για τα μοντέλα, ειδικότερα τα ειδικά βάρη που συσχετίζονται με τα επιβλεπόμενα μοντέλα. Στο παρόν μοντέλο δεν έγινε προσπάθεια επαναληπτικής επανεκπαίδευσης, με στόχο τη βελτίωση των ειδικών βαρών και παραμέτρων του μοντέλου και την αύξηση της ακρίβειας του ΑλλοMantIS. Μελλοντικός στόχος είναι η δοκιμή με συστηματικές και επαναλαμβανόμενες εκπαιδεύσεις του μοντέλου για την αύξηση της ακρίβειας πρόβλεψης με στόχο την επιβεβαίωση της γενικής τάσης για βελτίωση των επανεκπαιδευμένων μοντέλων.

Παράλληλα, στόχος μας είναι να δοκιμάσουμε το μοντέλο Μέγιστης Εντροπίας για την πρόβλεψη της ρηματικής αλλομορφίας και κατά πόσο είναι εξίσου αποδοτικό και αποτελεσματικό για τη συγκεκριμένη ομάδα αλλομόρφων. Είναι απαραίτητη η δημιουργία ενός ικανοποιητικά μεγάλου σώματος εκπαίδευσης με ρήματα που δεν έχουν παραχθεί από άλλη ρίζα, καθώς και ενός σώματος δοκιμής με παράγωγα ρήματα για την αξιολόγηση του ΑλλοMantIS. Επιπροσθέτως, εφόσον το αποτέλεσμα της επανεκπαίδευσης είναι θετικό για το ονοματικό μοντέλο του ΑλλοMantIS, θα δοκιμαστεί και στο αντίστοιχο ρηματικό. Παράλληλα θα πρέπει να γίνει και ένα επιθετικό μοντέλο, το οποίο θεωρούμε ότι θα έχει επιδόσεις εφάμιλλες του μοντέλου ονοματικής αλλομορφίας, καθώς διαφοροποιούνται ελάχιστα.

### **7.2.2. Υλοποίηση του μοντέλου Μορφολογικής Μάθησης της Αλλομορφίας**

Αποδείξαμε πως σε θεωρητικό επίπεδο είναι απαραίτητη η αντιμετώπιση της αλλομορφίας στους μορφολογικούς αναλυτές, ενώ παράλληλα δείξαμε με ένα υποτυπώδες επιβλεπόμενο μοντέλο Μέγιστης Εντροπίας πως να προβλέψουμε την

αλλομορφία για τα ουσιαστικά. Επόμενο στάδιο αυτής της σχεδίασης είναι ένας αναβαθμισμένος μορφολογικός αναλυτής που θα βασίζεται σε σημαντικό βαθμό σε ένα καλύτερα δομημένο επιβλεπόμενο μοντέλο που θα συνδυάζει όλα τα μοντέλα αλλομορφίας (ονοματικό, ρηματικό, επιθετικό) και θα μπορεί να προσδίδει την πλειονότητα των παραμέτρων αυτόματα, ώστε να ελαχιστοποιηθεί ο ανθρώπινος σχολιασμός των λέξεων στα σώματα ανάλυσης και δοκιμής.

Η υλοποίηση του συγκεκριμένου μοντέλου θα γίνει στο πλαίσιο ενός γενικού μοντέλου μορφολογικής ανάλυσης και θα είναι ένα από τα βασικά υποσυστήματα του αναλυτή. Παρατηρώντας τις σύγχρονες εφαρμογές μορφολογικής ανάλυσης είναι συνήθως τακτική να αποτελείται από μικρότερα υποσυστήματα για την καλύτερη επίδοση του συστήματος, αλλά και για την προσπάθεια συνδρομής άλλων εργαλείων. Ο Wicentowski (2002) υποθέτει ότι πιθανότατα όλα τα μοντέλα μηχανικής μορφολογικής μάθησης έχουν την ανάγκη συνδρομής συντακτικών και σημασιολογικών εργαλείων με άμεσο στόχο τη μείωση της ανθρώπινης παρέμβασης στα στάδια μετά την εκπαίδευση, αλλά και την συνδρομή τους για ελάττωση τυχόν αμφισημιών, λανθασμένων αναλύσεων και αντιγραμματικών λέξεων. Να υπογραμμιστεί ξανά πως το ζήτημα της επανανάλυσης και μορφολογικής αδιαφάνειας θα εξακολουθήσει να δημιουργεί προβλήματα. Στόχος της υλοποίησης είναι ένας μορφολογικός αναλυτής που να διαχωρίζει ορθά τα μορφήματα, να αντιστοιχεί τα αλλομορφα στα αντίστοιχα μορφήματα και να αντιμετωπίζει την παραγωγή των Ελληνικών με μεγάλη ακρίβεια. Είναι ενδιαφέρον να δοκιμαστεί το μοντέλο πρόβλεψης της αλλομορφίας και σε άλλες γλώσσες που εμφανίζουν αρκετά υψηλή συχνότητα εμφάνισης αλλομόρφων.

### 7.2.3. Αυξημένη κάλυψη του φαινομένου στην κλίση και τη σύνθεση

Ενώ το μοντέλο είναι εμφανώς αποτελεσματικό στην ανάλυση της αλλομορφίας των παράγωγων λέξεων της Ελληνικής, γεννιέται το ερώτημα εάν θα είναι το μοντέλο το ίδιο αποτελεσματικό και επιτυχημένο στην κλίση και στη σύνθεση. Στο κεφάλαιο 2 υποστηρίξαμε ότι η αλλομορφία συμμετέχει εξίσου σε όλες τις διαδικασίες σχηματισμού λέξεων και με παρόμοιο τρόπο. Επομένως, θεωρούμε ότι το μοντέλο είναι σχεδιασμένο να ανταπεξέλθει με επιτυχία στην αλλομορφία της κλίσης και της σύνθεσης. Εξάλλου το σώμα δοκιμής αποτελούταν από κλιτές λέξεις και επομένως, αφού το μοντέλο πέτυχε 91% ακρίβεια πρόβλεψης στις παράγωγες λέξεις, θα μπορέσει

σίγουρα να έχει υψηλότερο ποσοστό σε ΣΔ κλιτών λέξεων μίας και τα σώματα λέξεων θα αποτελούνται από λέξεις της ίδιας διαδικασίας σχηματισμού λέξεων.

Ταυτόχρονα είναι ενδιαφέρον να δούμε και την υλοποίηση του μοντέλου στις σύνθετες λέξεις και σε αυτές με δεσμευμένα θέματα. Σε θεωρητικό επίπεδο ο Καρασίμος (2001, 2011) είχε παρατηρήσει κοινά σχεδίοτυπα ανάμεσα στην κλίση και τη σύνθεση. Είναι σημαντικό να διερευνηθούν πιο συστηματικά οι αλλομορφικές συμπεριφορές των πρώτων συστατικών συνθέτων με τις αντίστοιχες των δεύτερων συνθετικών, κυρίως αυτών που συμμετέχουν με θέμα και όχι με λέξη, δηλαδή με σύνθετα τύπου ΘΕΜΑ ΘΕΜΑ (και ΛΕΞΗ ΘΕΜΑ)<sup>1</sup>. Κρίνεται επίσης απαραίτητο να διερευνηθεί πιο συστηματικά ο περιορισμός επιλογής του μοναδικού αλλομόρφου στο φαινόμενο της σύνθεσης και να δοκιμαστεί μία πιθανή τροποποίησή του για τη συγκεκριμένη διαδικασία (βλ. Καρασίμος 2011) τόσο των ονοματικών όσο και των ρηματικών θεμάτων. Η γενικότερη μελέτη του φαινομένου της αλλομορφίας σε θεωρητικό επίπεδο θα δώσει σημαντικά συμπεράσματα, ενώ επιπροσθέτως θα προσδώσει στην υπολογιστική ανάλυση όλα τα απαραίτητα εφόδια για τη δημιουργία ενός καθολικού μοντέλου αλλομορφικής αντιμετώπισης στα Ελληνικά.

#### 7.2.4. Πιθανός συνδυασμός για αυτόματη μορφολογική ανάλυση

Πολύς λόγος γίνεται για τον συνδυασμό μοντέλων (μη-εποπτευόμενων με εποπτευόμενων) με στόχο την καλύτερη δυνατή μορφολογική ανάλυση μίας γλώσσας. Ωστόσο είναι αξιοσημείωτο πως μοναχά ο PC-KIMMO συνδυάζει και τους δύο επιμέρους τομείς της μορφολογικής ανάλυσης, την ανάλυση και τη σύνθεση λέξεων· ειδικότερα ο τομέας της ανάλυσης επεξεργάζεται τις λέξεις στα κατάλληλα μορφήματα και ο τομέας της σύνθεσης συνδυάζει μορφήματα για το σχηματισμό λέξεων από κλιτές λέξεις έως σύνθετες· επιπλέον στους μορφολογικούς επεξεργαστές που είναι βασισμένοι σε μετατροπείς (FSA/ FST) αρκεί να προγραμματιστεί η μία διαδικασία και λόγω της συμμετρίας θα λειτουργεί ορθά και η άλλη. Έκτοτε οι περισσότεροι μορφολογικοί αναλυτές ασχολούνται μόνο με το πρώτο σκέλος και ελαχιστότατες προσπάθειες έγιναν για το δεύτερο. Χαρακτηριστικά οι Rajeev, Rajendran & Sherly (2008) αναφέρουν πόσο δύσκολο είναι να αντιμετωπίσουν τα προβλήματα και τους

---

<sup>1</sup> Δεν μας ενδιαφέρουν άμεσα οι περιπτώσεις όπου το δεύτερο συνθετικό είναι λέξη, δηλ. ΛΕΞΗ ΛΕΞΗ ή ΘΕΜΑ ΛΕΞΗ, γιατί σε αυτές τις περιπτώσεις δεν έχουμε καμία διαφορά από την κλίση όσον αφορά στο κλιτικό παράδειγμα και την αλλομορφική συμπεριφορά (βλ. Ράλλη 2007).

περιορισμούς στη σειρά των προσφυμάτων, καθώς και τους περιορισμούς στο συνδυασμό τους σε μία πλούσια και συγκολλητική γλώσσα, όπως τα Malayalam (για τα Ελληνικά, βλ. Melissaropoulou & Ralli to appear). Γίνεται σαφές ότι η πολυσυνδυαστικότητα των επιθημάτων αποτελεί ιδιαίτερα υψηλό βαθμό δυσκολίας για ένα σύστημα που συνθέτει παράγωγες λέξεις. Οι έρευνες των Goldsmith (2001), Schone & Jurafsky (2001) και Wicentowski (2002) παρουσιάζουν μη επιβλεπόμενες μεθόδους που εντοπίζουν αυτόματα υποψήφια προσφύματα για το αντίστοιχο σύνολο των προσφυμάτων. Είναι λοιπόν απαραίτητη και η συνδρομή των μη επιβλεπόμενων μοντέλων, καθώς θα κάνουν ένα πρώτο γενικό καθορισμό προσφυμάτων και πιθανότατα θα εντοπίσουν και διάφορους γενικούς περιορισμούς. Επομένως, η συνδυαστικότητα δεν απαιτείται μόνο στο επίπεδο υπολογιστικών θεωριών για την ανάλυση, αλλά των δύο επιμέρους τομέων ενός μορφολογικού αναλυτή.

### 7.3. Περίληψη

Η παρούσα διδακτορική διατριβή παρουσίασε ένα συστηματικό και συγκεκριμένο τρόπο προσέγγισης και ανάλυσης της αλλομορφίας σε θεωρητικό, ενώ παράλληλα απέδειξε επιτυχώς τη δυνατότητα επεξεργασίας του φαινομένου σε υπολογιστικό επίπεδο στα Ελληνικά. Προηγούμενες έρευνες δεν ασχολήθηκαν με την αλλομορφία στο σύστημα παραγωγής των Ελληνικών, ενώ οι γενικότερες προσπάθειες μηχανικής μάθησης και ανάλυσης της αλλομορφίας ήταν εξαιρετικά περιορισμένες και μερικώς επιτυχημένες στη μορφολογική ανάλυση. Αντίθετα, η εργασία μας όρισε το φαινόμενο της αλλομορφίας με αυστηρά μορφολογικά κριτήρια, καθόρισε τα περιβάλλοντα εμφάνισης και παρουσίασε με παραδείγματα τη συστηματική συμμετοχή σε όλες τις διαδικασίες σχηματισμού λέξεων. Πιο συγκεκριμένα παρουσίασε συγκριτικά τις αλλομορφικές συμπεριφορές θεμάτων και προσφυμάτων, περιορισμούς που διέπουν την επιλογή αλλομόρφου και την πεποίθηση της προβλεψιμότητας και κανονικότητας του φαινομένου.

Εξετάζοντας την αλλομορφία υπολογιστικά επισημάνθηκε η έλλειψη καθορισμένης στρατηγικής για την αντιμετώπισή της, καθώς και η επιλεκτική επεξεργασία των αλλομόρφων. Μετά την επιλογή των στρατηγικών της μηχανικής μάθησης, η αλλομορφία των παραγώγων δοκιμάστηκε αρχικά στο LINGUISTICA, το *gold standard* υλοποιημένο μοντέλο MeMM, όπου ο συγκεκριμένος αλγόριθμος απέτυχε όχι μόνο να αναλύσει σωστά τις παράγωγες λέξεις, αλλά και να εντοπίσει

αλλομορφα. Επιπλέον, δοκιμάστηκε ένα επιβλεπόμενο μοντέλο Μέγιστης Εντροπίας, το AMIS, για την πρόβλεψη της αλλομορφίας των ονοματικών αλλομόρφων σε επίπεδο θεμάτων και επιθημάτων. Ο ΑΛΛΟΜΑΝΤΙΣ κατάφερε να επιτύχει μία *state-of-the-art* επίδοση και έθεσε τις αρχικές βάσεις για την υπολογιστική πρόβλεψη των αλλομορφικών αλλαγών. Για την τελική υπολογιστική προσέγγιση της αλλομορφίας προτάθηκε ένα συνδυαστικό μοντέλο, το οποίο θα συνδυάζει διαφορετικές στρατηγικές, αυτής της χρήσης των αλλο-κανόνων σε συνεργασία με το μοντέλο χαρακτηριστικών για την πρόβλεψη της αλλομορφίας. Θεωρούμε ότι η παρούσα διατριβή κάλυψε και παρουσίασε με συστηματικότητα την υπολογιστική αντιμετώπιση της αλλομορφίας στην παραγωγή της ΚΝΕ κρίνοντας με βάση τα αποτελέσματα των πειραμάτων.

## Ξενόγλωσση Βιβλιογραφία

- AISSEN Judith 2003. Differential Object marking: iconicity vs. economy. *Natural Language and Linguistic Theory* **21**, pp. 435-483.
- ALLEN Jonathan, HUNNICUTT M. Sharon. & KLATT Dennis 1987. *From Text to Speech: the MITalk System*. Cambridge: Cambridge University Press.
- ANDERSON Stephen 1985. Inflectional Morphology. In T. Shopen (ed.) *Language Typology and Syntactic Fieldwork, Volume III*, pp. 4-56. Cambridge: Cambridge University Press.
- ANDERSON Stephen 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.
- ANDERSON Stephen 2008. Phonologically conditioned Allomorphy in the Morphology of Surmiran (Rumantsch). *On-line Proceedings of the 6<sup>th</sup> Mediterranean Morphology Meeting*, pp. 32-49. University of Patras. <http://morbo.lingue.unibo.it/mmm>
- ANDRON D. 1962. *Analyse morphologique du substantif russe*. Tech.rep., Centre d'Etudes pour la Traduction Automatique, Université de Grenoble 1.
- ANTWORTH Evan 1990. *PC-KIMMO: a two-level processor for morphological analysis*. Occasional publications in Academic Computing 16. Dallas TX: Summer Institute of Linguistics.
- ANTWORTH Evan 1993. Glossing text with the PC-KIMMO morphological parser. *Computers and Humanities* **26**, pp. 475-484.
- ARCHANGELI Diana 1984. *Underspecification in Yawelmani Phonology and Morphology*. PhD Dissertation. MIT.
- ARONOFF Mark 1976. *Word formation in Generative Grammar*. Cambridge, MA: MIT Press.
- BAAYEN Harald 2001. *Word frequency distributions*. Dordrecht: Kluwer.
- BAAYEN Harald, SCHREUDER Robert, DE JONG Nivja & KROTT Andrea 2002. Dutch Inflection: The rules that prove the exception. In S. Nooteboom, F. Weerman & F. Wijnen (eds.) *Storage and Computation in the Language Faculty*. Dordrecht: Kluwer.



- BAAYEN Harald & MOSCOSO DEL PRADO Martín Fermín 2005. Semantic density and past-tense formation in three Germanic languages. *Language* **81**, pp. 666-689.
- BAHL Lalit, BROWN Peter, DESOUZA Peter & MERCER Robert 1989. A tree-based statistical language model for natural language speech recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing* **37** (7), pp. 1001-1008.
- BAILY John & NEVINS Andrew 2008. Russian Genitive Plurals are Impostors. In A. Bachrach & A. Nevins (eds.) *Inflectional Identity*, pp. 237-270. Oxford: Oxford University Press. [🔴]
- BAUER Laurie & VALERA Salvador (eds.) 2005. *Approaches to Conversion/ Zero-Derivation*. München: Waxmann.
- BEEKES Robert 1995. *Comparative Indo-European Linguistics: An Introduction*. Amsterdam: John Benjamins.
- BEESLEY Kenneth & KARTTUNEN Lauri 2003. *Finite State Morphology*. CSLI Publications.
- BENSON Steven & MORE Jorge 2001. *A Limited-Memory Variable-Metric Algorithm Bound-Constrained Minimization*. Argonne National Laboratory, Mathematics and Computer Science Division.
- BERGER Adam, DELLA PIETRA Stephen & DELLA PIETRA Vincent 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* **22** (1), pp. 37-71.
- BERNARD-GEORGES A., LAURENT G. & LEVENBACH D. 1962. *Analyse morphologique du verbe allemand*. Technical report, Centre d' Etudes pour la Traduction Automatique, Université de Grenoble 1.
- BERNHARD Delphine 2006. *Apprentissage de connaissances morphologiques pour l'acquisition automatique de ressources lexicales*. PhD Thesis. Université Joseph Fourier - Grenoble I.
- BLACK Ezra, BROWN Peter, DELLA PIETRA Stephen, DELLA PIETRA Vincent, MERCER Robert & ROUKOS Salim 1992. Towards history-based grammars: Richer Models for Probabilistic Parsing. In *Proceedings of DAPRA, Speech and Natural Language Workshop*, pp. 134-139. New York: Arden House.
- BLOOMFIELD Leonard 1933. *Language*. New York: H. Holt and Company.

- BOBALJIK Jonathan 2000. The Ins and Outs of Contextual Allomorphy. In K. K. Grohman & C. Struijke (eds.) *University of Maryland Working Papers in Linguistics*, Volume 10, pp. 35-71. [🔴]
- BONET Eulàlia 2004. Morph insertion and Allomorphy in Optimality Theory. *International Journal of English Studies* 4 (2), pp. 73-104, Universidad de Murcia: Servicio de Publicaciones.
- BONET Eulàlia & MASCARÓ Joan 2007. Allomorph Selection and Lexical Preferences: Two case studies. *Lingua* 117 (6), pp. 903-927.
- BONET Eulàlia & HARBOUR Daniel 2010. Contextual Allomorphy. In J. Trommer (ed.) *The Handbook of Exponence*. Oxford: Oxford University Press. [🔴]
- BOOIJ Geert 1995. *The Phonology of Dutch*. Oxford: Clarendon Press.
- BOOIJ Geert 1997. Allomorphy and the Autonomy of Morphology. *Folia Linguistica* XXXI/1-2, pp. 25-56.
- BOOIJ Geert 2002. *The Morphology of Dutch*. Oxford/ New York: Oxford University Press.
- BOOIJ Geert 2003. Periphrastic Word Formation. In G. Booij, H. De Cesaris, A. Ralli & S. Scalise (eds.) *Proceedings of the 3<sup>rd</sup> Mediterranean Morphology Meeting*, pp. 15-28. Barcelona: Pompeu Fabra.
- BOOIJ Geert 2005a. *The Grammar of Words. An Introduction to Linguistic Morphology*. Oxford: Oxford University Press.
- BOOIJ, Geert 2005b. Compounding and derivation: evidence for Construction Morphology. In W. Dressler *et al.* (eds.) *Morphology and its demarcations*, pp. 109-132. Amsterdam / Philadelphia: John Benjamins.
- BOOIJ, Geert 2010. *Construction Morphology*. Oxford: Oxford University Press.
- BOUSSARD A. & BERTHAUD M. 1965. *Présentation de la synthèse morphologique du français*. Technical report, Centre d' Etudes pour la Traduction Automatique, Université de Grenoble 1.
- BRAND I., KLIMONOW G. & NÜNDEL S. 1969. Lexico-morphologische Analyse. In S. Nündel, G. Klimonow, I. Starke & I. Brand (eds.) *Automatische Sprachübersetzung: Russisch-Deutsch*, pp. 22-64. Akademie-Verlag, Berlin.

- BRILL Eric 1995. Transformation-based Error-Driven Learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics* **21**(4), pp. 543-565.
- BRILL Eric 1996. *Learning to Parse with Transformations*. Dordrecht: Kluwer.
- BRILL Eric & RESNINK Phillip 1994. A rule-based approach to prepositional phrase attachment disambiguation. In *Proceedings of the 15th conference on Computational linguistics, Volume 2 (COLING '94)*. Kyoto. [🔴]
- BURZIO Luigi 1999. Italian Participial Morphology and Correspondence Theory. In G. Booij, A. Ralli & S. Scalise (eds.) *Proceedings of the 1st Mediterranean Meeting of Morphology*, pp. 42-53. University of Patras Press.
- BURZIO Luigi 2005. Sources of Paradigm Uniformity. In L. J. Downing, T. A. Hall, R. Raffelsiefen (eds.) *Paradigms in Phonological theory*, pp. 65-106. Oxford: Oxford University Press.
- BURZIO Luigi 2006. Lexicon versus Grammar in English Morphophonology: Modularity Revisited. *Korean Journal of English Language and Linguistics* **6-3**, pp. 437-464.
- BYBEE Joan 1985. *Morphology. A study of the relation between Meaning and Form*. Amsterdam: John Benjamins.
- BYBEE Joan & SLOBIN Dan 1982. Rules and schemas in the development and use of the English past tense. *Language* **58**, pp. 265-289.
- BYBEE Joan & MODER Carol Lynn 1983. Morphological classes as natural categories. *Language* **59**, pp. 251-270.
- BYE Patrick 2007. Allomorphy – Selection, not optimization. In M. Krämer, S. Blaho & P. Bye (eds.) *Freedom of Analysis*. Berlin: Mouton de Gruyter.
- CALZOLARI Federico, MAMMINI Michele & MONACHINI Monica 2004. Unifying Lexicons in view of a Phonological and Morphological Lexical DB. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Volume IV, pp. 1107-1110. Paris: The European Language Resources Association (ELRA).
- CARSTAIRS Andrew 1987. *Allomorphy in Inflection*. London: Croom Helm Linguistics Series.

- CARSTAIRS-MCCARTHY Andrew 1992. *Current Morphology* (Linguistic Theory Guides). London: Routledge
- CARSTAIRS-MCCARTHY Andrew 1994. Inflection classes, gender and the principle of contrast. *Language* **70**, pp. 737-788.
- CHARNIAK Eugene 2000. A Maximum-Entropy-Inspired parser. In *ACM International Conference Proceedings 4* (Proceedings of the 1<sup>st</sup> North American Chapter of Association for Computational Linguistics Conference), pp. 132-139. Seattle, Washington: Morgan Kaufmann Publishers Inc.
- CHEN Stanley & ROSENFELD Roland 1999. A Gaussian prior for smoothing maximum entropy models. *Technical Reports CMUCS-99-108*. Carnegie Mellon University.
- CHOMSKY Noam 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- CHOMSKY Noam & HALLE Morris 1968. *The Sound Pattern of English*. New York: Harper and Row.
- CHOUEKA Yaacov 1990. Reponsa: An operational full-text retrieval system with linguistic components for large corpora. In A. Zampolli (ed.) *Computational Lexicology and Lexicography*. Pisa: Giardini.
- COKER Cecil, CHURCH Ken & LIBERMAN Meyer 1990. Morphology and rhyming: Two powerful alternatives to letter-to-sound rules for speech synthesis. In G. Bailly & C. Benoit (eds.) *Proceedings of the ESCA Workshop on Speech Synthesis*, pp. 83-86.
- CORBETT Greville, BROWN Dunstan, CHUMAKINA Marina & HIPPISELY Andrew 2005. Resources for Suppletion: A Typological Database and a Bibliography. In G. Booij, E. Guevara, A. Ralli, S. Sgroi & S. Scalise (eds.) *Morphology and Linguistic Typology*, On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4), pp. 35-44. University of Bologna [📄].
- CREUTZ Mathias & LAGUS Krista 2002. Unsupervised discovery of morphemes. In *Proceedings of the ACL-2002 Workshop on Morphological and Phonological Learning*, pp. 21-30.
- CROFT William 2003. *Typology and Universals*. Second Edition. Cambridge: Cambridge University Press.

- CROFT William & CRUSE D. Alan 2003. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- DAELEMANS Walter, BERCK Peter & GILLIS Steven 1996. Unsupervised discovery of phonological categories through supervised learning of morphological rules. In *Proceedings of the 16<sup>th</sup> conference on Computational Linguistics, Volume 1*, pp. 95-100. Morristown, NJ: Association of Computational Linguistics.
- DAELEMANS Walter & VAN BOSCH Antal 2005. *Memory-based Language Processing*. Cambridge: Cambridge University Press.
- DAELEMANS Walter, ZAVREL Jakub, VAN DER SLOOT Ko & VAN DEN BOSCH Antal 2007. *TiMBL 6.1 Reference Guide [ILK Technical Report]*. Tilburg, The Netherlands: Tilburg University. [🔴]
- DARROCH Hohn & RATCLIFF Douglas 1972. Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics* **43**, pp. 1470-1480.
- DAY David, ABERDEEN John, HIRSCHMANN Lynette, KOZIEROK Robyn, ROBINSON Patricia & VILAIN Marc 1997. Mixed-initiative Development of Language Processing Systems. In *Fifth Conference on Applied Natural Language Processing*, pp. 348-355. Association for Computational Linguistics. [🔴]
- DE SAUSSURE Ferdinand 1916. *Cours de linguistique générale*. Lausanne, Paris: Payot.
- DECLERCK Thierry, KESSLER Mirjam & KRIEGER Ulrich 2005. *Evaluation of initiatives for morpho-syntactic and syntactic annotation*. LIRICS Deliverable D3.1, pp. 1-62. DFKI, Saarbrücken. [🔴]
- DELL'ORLETTA Felice, LENCI Alessandro, MONTEMAGNI Simonetta & PIRRELLI Vito 2007. Corpus-based modeling of grammar variation. In A. Sansò (ed.) *Language Resources and Linguistics Theory*, pp. 38-55. [Materiali Linguistici 59]. Milano: Franco Angeli.
- DELLA PIETRA Stephen, DELLA PIETRA Vincent, GILLET John, LAFFERTY John, PRINTZ Harry & UREŠ Lubos 1994. Inference and Estimation of a Long-range Trigram Model. *Lecture Notes in Artificial Intelligence* **862**, pp. 78-92. Springer Verlag.
- DI SCIULLO Anna-Maria 2005. *Assymetry in Morphology*. Linguistic Inquiry Monograph 46. Cambridge, Massachusetts: MIT Press.

- DI SCIULLO Anna-Maria & WILLIAMS Edwin 1987. *On the Definition of Word*. Cambridge, Massachusetts: MIT Press.
- DRACHMAN Gaberell, KAGER Rene & MALIKOUTI-DRACHMAN Angeliki 1995. Greek Allomorphy: an Optimality Account. In M. Dimitrova-Vulchanova & L. Hellan (eds), *Papers from the First Conference on Formal Approaches to South Slavic Languages*, pp. 345-361.
- DRACHMAN Gaberell 2005. A note on ‘shared’ allomorphs. *Journal of Greek Linguistics* **6**, pp. 5-38.
- DRESSLER Wolfgang 1989. Prototypical differences between Inflection and Derivation. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* **42** (1), pp. 3-10.
- DZEROSKI Saso & ERJAVEC Tomaz 1997. Induction of Slovene nominal paradigms. In N. Lavrac & S. Dzeroski (eds.) *Inductive Logic Programming, 7<sup>th</sup> International Workshop (ILP 1997)*, pp. 17-20. Prague, Berlin: Springer.
- EDDINGTON David 2000. Analogy and the dual-route model of morphology. *Lingua* **110**, pp. 281-289.
- ESPAÑOL-ECHEVARRÍA, Manuel & RALLI Angela 2000. Case mismatches in Greek: Evidence for the Autonomy of Morphology. *Acta Linguistica Hungarica, Volume* **47** (1-3), pp. 179-203.
- ESTES William K. 1994. *Classification and cognition*. Oxford: Oxford University Press.
- FABB Nigel 1988. English Suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory* **6**, pp. 527-539.
- FIX Evelyn & HODGES J. L. 1951. *Discriminatory analysis. Nonparametric discrimination: consistency properties. [aka An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation]*. Randolph Field, Texas. USAF School of Aviation Medicine. {via Silverman & Jones (1989) Review in *International Statistical Review* **57** (3), pp. 233-247}.
- FLEISCHER, Wolfgang 1969. *Wortbildung der deutschen Gegenwartssprache*. Leipzig: VEB Bibliographisches Institut.
- FLENNER Gudrun 1994. Ein quantitatives Morphsegmentierungssystem für spanische Wortformen. In U. Klenk (ed.) *Computatio Linguae II*, pp. 31-62. Stuttgart: Steiner Verlag.





- FLENNER Gudrun 1995. Quantitative Morphsegmentierung im Spanischen auf phonologischer Basis. *Sprach und Datenverarbeitung* **19**(2), pp. 63-79.
- FLIER Michael 1974. The Glide shift in Russian deverbal derivation. *Russian Linguistics* **1**, pp. 15-31.
- FLORIAN Radu, HENDERSON John & NGAI Grace 2000. Coax Confidences from an old friend: Probabilistic Classifications from Transformation Rule Lists. In *Proceedings of Joint Sidgat Conference of Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000)*, pp. 26-34. Hong Kong: University of Science and Technology. [🔴]
- FLORIAN Radu & NGAI Grace 2001. Multidimensional transformational-based learning. In *Proceedings of the 5<sup>th</sup> Conference on Computational Natural Language Learning*, pp. 1-8.
- FLORIAN Radu & NGAI Grace 2007. *Fast Transformation-based Learning Toolkit*. Johns Hopkins University. On-line [<http://www.cs.jhu.edu/~rflorian/fntbl/tbl-toolkit/tbl-toolkit.html>].
- FRIEDL Jeffrey 2002. *Mastering Regular Expressions*. O' Reilly Media.
- GALANI Alexandra 2003a. Analysis of Theme Vowels in Modern Greek within Distributed Morphology. *Interlinguistica* **14**, pp. 399-412.
- GALANI Alexandra 2003b. Allomorphy: Theme Vowels in Modern Greek, in *Camling Proceedings*, Cambridge, Cambridge University Press.
- GALIOTOU ELENI & RALLI ANGELA 2002. Parsing Deficiencies of the PC-KIMMO System. In *Proceedings of the 2<sup>nd</sup> Hellenic Conference on Artificial Intelligence (Companion Volume)*, pp. 53-64. Thessaloniki: Aristotle University Press.
- GALIOTOU ELENI & RALLI ANGELA 2003. Morpho-phonological modeling in Natural Language Processing. *International XII Turkish Symposium of Artificial Intelligence and Neural Networks – TAINN*.
- GALIOTOU ELENI & RALLI ANGELA 2005. Morpho-phonological modeling in Natural Language Processing. *International Journal of Computational Intelligence* **1** (3), pp. 155-158.
- GAMBELL Timothy & YANG Charles 2007. Computational Learning and Language Acquisition: A view from word segmentation. *Lingue e Linguaggio* **VI.2** (2007), pp. 139-150. Bologna: Il Mulino.

- GAUSSIER Eric 1995. Analyse textuelle, indexation automatique et construction référentielles. *Troisième Journée sur l'Information Elaborée*. Ile-Rousse, France
- GEORGACAS Demetrius John 1948. On nominal endings –is and –in in later Greek. *CPh* **43**, pp. 243-260.
- GOLDSMITH John 2001. Unsupervised Learning of the Morphology of a Natural Language. *Computational Linguistics* **27**(2), pp. 153-198.
- GÖKSEL Asli & KERSLAKE Celia 2005. *Turkish A Comprehensive Grammar*. New York: Routledge.
- GRUBER Hermann & NEVEN Frank 2008. Succinctness of the Complement and Interaction of Regular Expressions. *Proceedings of the International Symposium of Theoretical Aspects of Computer Science (STACS 2008)*, pp. 325-336.
- GRUBER Hermann & HOLZER Markus 2008. Finite-state automata, Digraph connectivity and Regular Expression size. *Proceedings of the 35<sup>th</sup> International Colloquium on Automata, Languages and Programming (ICALP 2008)*, pp. 39-50.
- HAFER Margaret & WEISS Stephen 1974. Word segmentation by letter successor varieties. *Information Storage and Retrieval* **10**, pp. 371-385.
- HAHN Ulrike & NAKISA Ramin Charles 2000. German Inflection: Single Route or Double Route? *Cognitive Psychology* **41**, pp. 313-360.
- HALLE Morris 1959. *The Sound Pattern of Russian*. The Hague, Mouton.
- HALLE Morris & MARANTZ Alec 1993. Distributed Morphology and the Pieces of Inflection, in K. Hale and J.S. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honour of Sylvain Bromberger*, pp.111-176. Cambridge, Massachusetts: MIT Press.
- HANDL Johannes, KABASHI Besim, PROISL Thomas & WEBER Carsten 2009. JSLIM – Computational Morphology in the Framework of the SLIM Theory of Language. In C. Mahlow & M. Pioteowski (eds.) *State of the Art in Computational Morphology*, pp. 10-27. Berlin: Springer, Verlag Heidelberg.
- HANKAMER Jorge 1986. Finite state morphology and left to right phonology. In *Proceedings of the West Coast Conference on Formal Linguistics, Volume 5*, pp. 41-52. Stanford: Stanford Linguistic Association.



- HARIZANOV Boris & GRIBANOVA Vera 2010. The role of morphological and phonological factors in Bulgarian allomorph selection. *Morphology at Santa Cruz*. [🔴]
- HARRIS Zellig 1955 [1970]. From phoneme to morpheme. *Language* **31**, pp. 190-222.
- HARRIS Zellig 1967 [1970]. Morpheme boundaries within words: Report on a computer test. *Transformations and Discourse Analysis Papers* **73**. University of Pennsylvania.
- HAUSSER Roland 1999. *Foundations of Computational Linguistics*. Heidelberg: Springer.
- HAUSSER Roland 2006. *A Computational Model of Natural Language Communication: Interpretation, Inference, and Production in Database Semantics*. Heidelberg: Springer.
- HAY Jennifer & PLAG Ingo 2004. What Constrains Possible Suffix Combinations? On the Interaction of Grammatical Processing Restrictions in Derivational Morphology. *Natural Language and Linguistics Theory* **22**, pp. 565-596
- HUTCHINS John 2001. Machine translation over 50 years. *Histoire, Epistemologie, Langage* **22** (1), pp. 7-31. [🔴]
- JANSSEN Axel 1992. Segmentierung französischer Wortformen in Morphe ohne Verwendung eines Lexikons. In U. Klenk (ed.) *Computatio Linguae*, pp. 74-95. Stuttgart: Steiner Verlag.
- JAYNES Edwin Thompson 1957a. Information Theory and Statistical Mechanics. *Physical Review* **106** (4), pp. 620-630.
- JAYNES Edwin Thompson 1957b. Information Theory and Statistical Mechanics II. *Physical Review* **108** (2), pp. 171-190.
- JAYNES Edwin Thompson 1968. Prior Probabilities. *IEEE Transactions on System Science and Cybernetics* **4** (3), pp. 227-241 [🔴]
- JAYNES Edwin Thomson 1990. Notes on Present Status and Future Prospects. In W.T. Grandy & L.H. Schick (eds.) *Maximum Entropy and Bayesian Methods*, pp. 1-13. Dordrecht: Kluwer.
- JOHNSON Howard & MARTIN James 2003. Unsupervised learning of morphology for English and Inuktitut. In *Proceedings of the 2003 Human Language Technology*

- Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2003), Companion Volume*, pp.43-45.
- JOSEPH Brian & WALLACE Rex 1984. Lexical Relatedness, Head of the Word and the misanalysis of Latin. *Ohio-State Working Papers in Linguistics* **29**, pp. 30-49.
- JOSEPH Brian & JANDA Richard 1998. The how and why of diachronic morphologization and demorphologization. In M. Hammond & M. Noonan (eds.) *Theoretical Morphology. Approaches to Modern Linguistics*, pp. 193-210. New York: Academic Press
- JURAFSKY Daniel & MARTIN James 2000. *Speech and Language Processing (An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition)*. Prentice Hall Inc.
- KAPLAN Ronald & KAY Martin 1981. Phonological Rules and finite-state transducers. Presented at the *Winter Meeting of Linguistic Society of America* (New York).
- KAPLAN Ronald & KAY Martin 1994. Regular Models of Phonological Rule Systems. *Computational Linguistics* **20** (3), pp. 331-378.
- KARASIMOS Athanasios 2009. Comments on Unsupervised Morphology Learning Model: The case of Greek Allomorphy. Paper presented at the *19<sup>th</sup> International Symposium in Theoretical and Applied Linguistics (ISTAL19)*. 3-5 April 2009, Thessaloniki.
- KARASIMOS Athanasios & PETROPOULOU Evanthia 2009. Computational analysis of suffixes and bound stems of the Greek Language: A crash test with *Linguistica*. In A. Ralli (ed.) *Patras Working Papers in Linguistics* (Special Issue: Morphology), pp. 128-148.
- KARASIMOS Athanasios & PETROPOULOU Evanthia 2010. A crash test with *Linguistica* in Modern Greek: the case of derivational suffixes and bound stems. In *Proceedings of LREC 2010*, pp. 182-189. Valetta, Malta. [🔴]
- KARTTUNEN Lauri 1983. KIMMO: a general morphological parser. *Texas Linguistic Forum* **22**, pp. 163-186.
- KARTTUNEN Lauri & BEESLEY Kenneth 2005. Twenty-five years of finite-state morphology. In A. Arppe *et al.* (eds.) *Inquiries into Words, Constraints and Contexts* (Festschrift for Kimmo Koskenniemi on his 60<sup>th</sup> birthday), pp. 71-83. Stanford, California: CSLI Publications.

- KAZAMA Jun'ichi & TSUJII Jun'ichi 2003. Evaluation and Extension of Maximum Entropy Models with Inequality Constraints. In *Proceedings of EMNLP 2003*. [
- KAZAMA Jun'ichi & TSUJII Jun'ichi 2005. Maximum Entropy Models with Inequality Constraints: A Case Study on Text Categorization. *Machine Learning* **60** (1-3), pp. 159-194. Hingham, MA, USA: Kluwer Academic Publishers
- KEULEERS Emmanuel & DAELEMANS Walter 2007. Memory-based Learning of Inflectional Morphology: A methodological case study. *Lingue e Linguaggio* **VI.2**, pp. 151-174. Bologna: Il Mulino.
- KEULEERS Emmanuel, SANDRA Dominiek, DAELEMANS Walter, GILLIS Steven, DURIEUX Gert & MARTENS Evelyn 2007. Dutch plural inflection: The exception that proves analogy. *Cognitive Psychology* **54** (4), pp. 283-318.
- KIKUCHI Seiichiro 2006. On Galician Definite Article Allomorphy. *On'in Kenkyu [Phonological Studies]* **9**, pp. 41-48. Tokyo: Kaitakusha. [
- KIPARSKY Paul 1982a. *Lexical Phonology and Morphology*. Seoul: Linguistics in the Morning Calm, Hanshin.
- KIPARSKY Paul 1982b. From Cyclic Phonology to Lexical Phonology. In H. van der Hulst & N. Smith (eds.) *The Structure of Phonological Representation I*, pp. 131-175. Dordrecht: Foris.
- KLENK Ursula 1992. Verfahren morphologischer Segmentierung und die Wortstruktur im Spanischen. In U. Klenk (ed.) *Computio Linguae*. Stuttgart: Steiner Verlag.
- KOCH Sabine, KÜSTNER Andreas & RÜDIGER Barbara 1989. Deutsche Wortformesegmentierung ohne Lexicon. *Sprache und Datenverarbeitung* **18**, pp. 35-44.
- KOELING Rob 2000. Chunking with Maximum Entropy Models. *Proceedings of CoNLL 2000 and LLL-2000*, pp. 139-141. Lisbon. [
- KOHONEN Oskar, VIRPIOJA Sami & KLAMÌ Michaela 2009. Allomorfessor: Towards Unsupervised Morpheme Analysis. In *Evaluating systems for Multilingual and Multimodal Information Access*, Lecture Notes in Computer Science volume 5706, pp. 975-982. Springer Publications. [
- KORNFILT Jaklin 1997. *Turkish (Descriptive Grammars)*. New York: Routledge.

- KOSKENNIEMI, Kimmo 1983. *Two-level morphology: a general computational model for word-form recognition and production*. Ph.D. Thesis. Helsinki: University of Helsinki.
- KRACHT Markus 2003. Constraints on Derivation. *Grammars* 6, pp. 89-113.
- KRIEGER Hans-Ulrich, PIRKER Hannes & NERBONNE John 1993. Feature-based Allomorphy. *Proceedings of the 31<sup>st</sup> annual meeting of the Association for Computational Linguistics*, pp. 140-147. ACL: Morristown, NJ, USA.
- KURIMO Mikko, CREUTZ Mathias, VARJOKALLIO Matti, ARISOY Elbu & SARACLAR Murat 2006a. Unsupervised segmentation of Words into Morphemes – Challenge 2005 An Introduction and Evaluation Report. In *PASCAL Challenge Workshop on Unsupervised segmentation of words into morphemes*. Venice, Italy. [📄]
- KURIMO Mikko, CREUTZ Mathias, VARJOKALLIO Matti, ARISOY Elbu & SARACLAR Murat 2006b. Unsupervised segmentation of words into morphemes – Morpho Challenge 2005 Application to Automatic Speech Recognition. In *Interspeech 2006*, pp. 1-4. Pittsburg, USA. [📄]
- KURIMO Mikko, CREUTZ Mathias & VARJOKALLIO Matti 2007. Unsupervised Morpheme Analysis Evaluation by a Comparison to a Linguistic Gold Standard Morpho Challenge 2007. In A. Nardi & C. Peters (eds.) *Working Notes of the CLEF 2007 Workshop*, pp. 1-12. Budapest, Hungary. [📄]
- KURIMO Mikko, VIRPIOJA Sami & TURUMEN Ville 2009. Overview and Results of Morpho Challenge 2009. In *Working Notes for CLEF 2009 Workshop*, pp. 1-27. Corfu, Greece. [📄]
- KURIMO Mikko, VIRPIOJA Sami, TURUNEN Ville & LAGUS Krista 2010. Morpho Challenge competition 2005-2010: Evaluations and results. *Proceedings of the 11th Meeting of the ACL Special Interest Group on Computational Morphology and Phonology*, pp. 87-95. Uppsala, Sweden: Association of Computational Linguistics. [📄]
- KURIŁOWICZ Jerzy & MAYRHOFER Manfred 1969. *Indogermanische Grammatik*. Heidelberg: Winter Verlag.
- LAHIRI Aditi, RIAD Thomas & JACOBS Haike 1999. Diachronic Prosody. In H. van der Hulst *Word Prosodic Systems and the Languages of Europe*. Berlin/ New York: Mouton de Gruyter, pp. 335-422.

- LAHNE Antje 2006. When Features are not deleted: Contextual Allomorphy in Sanskrit Noun Declension. In G. Müller & J. Trommer (eds.) *Subanalysis of Argument Encoding in Distributed Morphology*, pp. 143-160. Linguistische Arbeitsberichte 84, Universität Leipzig. [📄]
- LAMPROPOULOS Aristomenis, GALIOTOU Eleni, MANOLESSOU Ioanna & RALLI Angela 2007. A Finite-State Approach to the Computational Morphology of Early Modern Greek. In *Proceedings of 7<sup>th</sup> WSEAS International Conference on Applied Computer Science*, pp. 242-245. Velice.
- LAU Raymond, ROSENFELD Roni & ROUKOS Salim 1993. Adaptive Language modeling using the maximum entropy principle. *Proceedings of Human Language Technology Workshop*, pp. 108-113. ARPA.
- LIEBER Rochelle 1980. *On the Organization of the Lexicon*. PhD Dissertation, MIT.
- LIEBER, Rochelle 1982. Allomorphy. *Linguistic Analysis*. Volume 10, Number 1, 1982, pp. 27-52.
- LÖFSTEDT Ingvar 2009. Perceptually driven Allomorphy: The Swedish non-neutral singular definite article. In A. Schardl, M. Walkow & M. Abdurrahman (eds.) *Nels 38: Proceedings of the 38<sup>th</sup> Meeting of North East Linguistics Society*, pp. 87-98. University of Ottawa: Createspace. [📄]
- LOVINS Julie Beth 1968. Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics* **11**, pp. 22–31.
- ŁUBOWICZ Anna, GO Nathan, HUANG Nancy & McDONALD Sara 2006. Polish Nickname Formation: The case of Allomorph selection. In *Proceedings of Formal Approaches to Slavic Languages 15*. University of Michigan: Michigan Slavic Publications, Ann Arbor. [📄]
- LURAGHI Silvia 2005. The evolution of the Greek nominal paradigms from Mycenaean to Modern Greek. *Classica et Medievalia* **55**, pp. 361-380.
- MAHLOW Cerstin & PIOTROWSKI Michael (eds.) 2009. *State of the Art in Computational Morphology*. Proceedings of Workshop on Systems and Frameworks for Computational Morphology, SFCM 2009. Zurich, Switzerland, September 2009. Springer: Communications in Computer and Information Science 41.
- MAIDEN Martin 1992. Irregularity as a Determinant of Morphological Change. *Journal of Linguistics* **28**, pp. 285-312.

- MAIDEN Martin 2001. Morphological bootstrapping? Paradigmatic autonomy and the rise of suppletion. Paper presented at the *International Conference on Hybrid Learning 2001*. Melbourne, Australia.
- MAIDEN Martin 2004. When lexemes become allomorphs – On the genesis of suppletion. *Folia Linguistica* **XXXXCIII**/3-4, pp. 227-256. Mouton de Gruyter, Berlin: Societas Linguistica Europaea.
- MAIDEN Martin 2007. Morphemes and ‘phonologically conditioned allomorphy’. Paper presented at *First Oxford Workshop on Romance Verb Morphology*. Oxford: Trinity College.
- MALMKJÆR Kirsten (ed.) 2002. *The Linguistics Encyclopedia*. (2<sup>nd</sup> ed.). New York: Routledge.
- MANOLESSOU Ioanna 2005. From Participles to Gerunds. In M. Stavrou & A. Terzi (eds.) *Advances in Greek Generative Syntax*, pp. 241-283. Amsterdam: John Benjamins.
- MANOVA Stela & DRESSLER Wolfgang 2005. The morphological technique of conversion. In L. Bauer & S. Valera (eds.) *Approaches to Conversion/ Zero-Derivation*, pp. 67-101. München: Waxmann.
- MARANTZ Alec 1982. Re Reduplication. *Linguistic Inquiry* **25**, pp. 609-667.
- MARCHAND, Hans 1967. Expansion, Transposition and Derivation. *La Linguistique* **1**, pp. 13-26.
- MARCHAND, Hans 1969. *The Categories and Types of Present-Day English Word-Formation: A Synchronic-Diachronic Approach*. München: C.H. Beck. 2nd edition.
- MASCARÒ Joan 1996α. External allomorphy and contractions in Romance. *Probus* **8**, pp. 181-205.
- MASCARÒ Joan 1996β. External allomorphy as emergence of the unmarked. In J. Durand & B. Laks (eds.) *Current Trends in Phonology: Models and methods* **2**, pp. 473-483. Salford Manchester: University of Salford, European Studies Research Institute. [Reprinted in J. McCarthy (ed.) *Optimality Theory in Phonology: A reader*, pp. 513-522. Oxford: Blackwell (2004)].
- MASCARÒ Joan 2007. External allomorphy and lexical representation. *Linguistic Inquiry* **38** (4), pp. 715-735.

- MATTHEWS Peter 1966. A procedure for morphological encoding. *Mechanical Translation* **9**, pp. 15-21.
- MATTHEWS Peter 1972. *Inflectional Morphology: A theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.
- MATTHEWS Peter 1974. *Morphology: An introduction to the theory of Word-Structure*. Cambridge: Cambridge University Press.
- MCILROY M. Douglas 1982. Development of a spelling list. *IEEE Transactions on Communications* **30** (1), pp. 91-99.
- MEL'CUK Igor 1976. On suppletion. *Linguistics* **170**, pp. 207-218.
- MEL'CUK Igor 1992. Suppletion. In W. Bright (ed.) *International Encyclopedia of Linguistics*, pp. 97-99. New York: Oxford University Press.
- MEL'CUK Igor 1994. Suppletion: Toward a logical analysis of the concept. *Studies in Language* **18**, pp. 339-410.
- MEL'CUK Igor 2006. *Trends in Linguistics*. In D. Beck (ed.) *Aspects of the Theory of Morphology (series)*. Berlin: Mouton de Gruyter.
- MELISSAROPOULOU Dimitra & RALLI, Angela 2010. Structural combinatorial properties of Greek derivational suffixes. *Morphology* (submitted).
- MEYA-LLOPORT M. 1987. Morphological analysis of Spanish for retrieval. *Literary and Linguistic Computing* **2**, pp. 228-244.
- MEYER Paul Georg 1997. *Coming to know (Studies in the lexical semantics and pragmatics of academic English)*. Tübingen: Gunter Narr Verlag.
- MIYAO Yusuke & TSUJII Jun'ichi 2002. Maximum Entropy estimation for feature forests. In M. Marcus (ed.) *Proceedings of the Second International Conference on Human Language Technology Research (HLT 2002)*, pp. 292-297. [📄]
- MOHANAN K.P. 1986. *The Theory of Lexical Phonology. Studies in Natural Language and Linguistic Theory*. Dordrecht: D. Reidel.
- MONACHINI Monica, CALZONARI Nicoletta, CHOUKRI Khalid, FRIEDRICH Jochen, MALTESE Giulio, MAMMINI Michele, ODIJK Jan & ULIVIER Marisa 2006. Unified Lexicon and Unified Morphosyntactic Specifications for Written and Spoken Italian. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pp. 1852-1857. Paris: The European Language Resources Association (ELRA). [📄]

- NESPOR Marina & VOGEL Irene 1986. *Prosodic Phonology*. Dordrecht: Foris
- NEVINS Andrew 2010. Phonologically-conditioned allomorph selection. In C. Ewen, E. Hume, M. van Oostendorp & K. Rice (eds.) *The Companion to Phonology*. Blackwell Publications. [🔴]
- NOCEDAL Jorge 1980. Updating Quasi-Newton matrices with limited storage. *Mathematics on Computation* **35**, pp. 773-783.
- PACAK Milos, PRATT Arnold & WHITE William 1976. Automated morphosyntactic analysis of medical language. *Information Processing and Management* **12/1**, pp. 71-76.
- PASTER Mary 2005. Subcategorization vs. Output Optimization in Syllable-counting Allomorphy. In J. Alderete, C. hye Han, & A. Kochetov (eds.) *Proceedings of the Twenty-Fourth West Coast Conference on Formal Linguistics*, pp. 326–333. Somerville, MA: Cascadilla Proceedings Project.
- PAYNE Thomas 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge, New York: Cambridge University Press.
- PETROPOULOU Evanthia 2009. Neoclassical compounds in Modern Greek and English: A comparative view. In A. Karasimos *et al.* (eds.) *Proceedings of the 1<sup>st</sup> Patras International Conference of Graduate students in Linguistics*, pp. 86-97. Patras: University of Patras Press Center.
- PHILIPPAKI-WARBURTON Irene 1970. *On the Verb in Modern Greek*. The Hague: Mouton.
- PHILIPPAKI-WARBURTON Irene 1973. Modern Greek Verb Conjugation: Inflectional Morphology in a Transformational Grammar. *Lingua* **32**, pp. 193-226.
- PHILIPPAKI-WARBURTON Irene 1991. The analysis of the Verbal Complex in Modern Greek. *Μελέτες για την Ελληνική Γλώσσα 1990*, σ. 119-130. Thessaloniki, Greece.
- PILCH Herbert 1988. Lexical indeterminacy. In E. G. Stanley & T. F. Hoad (eds.) *Words: For Robert Burchfield's sixty fifth birthday*, pp. 140-156. D.S. Brewer.
- PINKER Steven 1999. *Words and Rules: The Ingredients of Language*. New York: Basic Books.
- PINKER Steven & PRINCE Alan 1988. On language and connectionism Analysis of a parallel distributed processing model of language acquisition. In S. Pinker & J.



- Mehler (eds.) *Connections and Symbols (Special Issue: Cognition)*, pp. 73-193. Cambridge, MA: MIT Press.
- PIRRELLI Vito 2000. *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale de l' italiano*. Pisa-Rome: Istituti editoriali e poligrafici internazionali.
- PIRRELLI Vito 2007a. Psycho-Computational issues in Morphology Learning and Processing: An overture. *Lingue e Linguaggio* **VI.2**, pp. 131-138. Bologna: Il Mulino.
- PIRRELLI Vito 2007b. On the cognitive autonomy of morphological processing. In N. Hathout & F. Montermini (eds.), *Morphologie à Toulouse. Actes du colloque international de Morphologie 4èmes Décembrettes*, München, Lincom Europa (LSTL 37), pp. 245-269.
- PIRRELLI Vito & YVON François 1999a. Analogy in the Lexicon: a probe into analogy-based machine learning of language. In *Proceeding of 6<sup>th</sup> International Symposium on Human Communication*. Cuba: Santiago de Cuba.
- PIRRELLI Vito & YVON Francois 1999b. The hidden dimensions: paradigmatic approaches to data-driven natural language processing. *Journal of Experimental and Theoretical Artificial Intelligence* **11** (Special Issue on Memory-based Language Processing), pp. 391-408.
- PIRRELLI Vito & BATTISTA Marco 2000α. The paradigmatic Dimension of Stem Allomorphy in Italian Verb Inflection. *Italian Journal of Linguistics (Rivista di Linguistica) Volume 12* (II), pp. 307-380.
- PIRRELLI Vito & BATTISTA Marco 2000β. On the interaction of paradigmatic and syntactic stem alternation in Italian conjugation. *Acta Linguistica Hungarica* **47** (1-4), pp. 289-314.
- PLAG Ingo 1999. *Morphological Productivity. Structural Constraints in English Derivation*. Berlin/ New York: Mouton de Gruyter.
- PLAG Ingo 2003. *Word-formation in English*. Cambridge: Cambridge University Press.
- PLANK Frans 1994. Inflection and Derivation. In R. E. Asher (ed.) *The encyclopedia of Languages and Linguistics* **3**, pp. 1671-1678. Oxford: Pergamon Press.


- PLUNKETT Kim & Juola Patrick 2001. A Connectionist Model of English Past-Tense and Plural Morphology. In M.H. Christiansen & N. Carter (eds.) *Connectionist Psycholinguistics*, pp. 106-137. Ablex Publishing Corporation.
- POLLARD Carl & Sag Ivan 1987. *Information-Based Syntax and Semantics, Volume I*. Stanford: CSLI.
- PORTER Martin 1980. An Algorithm for suffix stripping. *Program* **14** (3), pp. 130-137.
- PORTER Martin 2001. *Snowball: A language for stemming algorithms*. Available on: <http://snowball.tartarus.org/texts/introduction.html>
- PRINCE Alan & SMOLENSKY Paul 1993. *Optimality Theory: Constraint interaction in Generative Grammar*. Rutgers University Center for Cognitive Science. ROA Version 2002.
- QUBLAN AL-MOZAINY Hamza, BLEY-VROMAN Robert & MCCARTHY John 1985. Stress shift and Metrical Structure. *Linguistic Inquiry* **16**, pp. 135-144.
- RAJEEV R., RAJENDRAN N. & SHERLY Elisabeth 2008. A suffix-stripping based Morph Analyzer for Malayalam Language. *Proceedings of 20<sup>th</sup> Kerala Science Congress*, pp. 482-494.
- RALLI Angela 1988. *Eléments de la Morphologie du Grec Moderne: La Structure du Verbe*. PhD Dissertation, University of Montreal.
- RALLI Angela 1994. Feature Representation s and Feature-Passing Operations in Greek Nominal Inflection. *Proceedings of the 8<sup>th</sup> Symposium on English and Greek Linguistics*, pp. 19-46. Thessaloniki: English Department of Aristotle University of Thessaloniki.
- RALLI Angela 1997. Inflectional Features and the Morphological Component Hypothesis. *Proceedings of GLOW 1997*, pp. 93-103. Rabat.
- RALLI Angela 1999. Inflectional Features and the Morphological Module Hypothesis. In A. Kakouriotes (ed.) *Working Papers in Theoretical and Applied Linguistics* **6**, pp. 111-142. Thessaloniki: English Department of Aristotle University of Thessaloniki.
- RALLI Angela 2000. A Feature-based Analysis of Greek Nominal Inflection. *Glossologia* **11-12**, pp. 201-228.

- RALLI Angela 2002. Prefixation vs. Compounding. In A.-M. Di Sciullo (ed.) *Asymmetry in Grammar: Phonology, Morphology and Language Acquisition*, pp. 37-64. Amsterdam: John Benjamins.
- RALLI Angela 2004. Stem-Based vs. Word-Based Morphological Configurations: the case of Modern Greek preverbs. *Lingue e Linguaggio II*, pp. 241-275. Bologna: Il Mulino.
- RALLI Angela 2007. On the Role of Allomorphy in Inflectional Morphology: Evidence from the Dialectal Varieties of Lesvos, Kydonies and Moschonisia. In G. Sica (ed.) *Open problems in Linguistics and Lexicography*. Milano: Polimetrica.
- RALLI Angela 2009. Hellenic Compounding. In R. Lieber & P. Stekauer (eds.), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. pp. 453-463.
- RALLI Angela & GALIOTOU Eleni 2005. Greek Compounds: A Challenging Case for the Parsing Techniques of PC-KIMMO v.2. *International Journal of Computational Intelligence* 1 (2), pp. 128-138.
- RALLI Angela & KARASIMOS Athanasios 2008. Constituency in compounding and its relation with derivational suffixation. *On-line Proceedings of 8<sup>th</sup> International Conference of Greek Linguistics*, Workshop in compounding. University of Ioannina.
- RALLI Angela & KARASIMOS Athanasios 2009a. The Role of Constraints in Greek Compound Formation. *Lingue e Linguaggio VIII.1*, pp. 53-74. Bologna: Il Mulino.
- RALLI Angela & KARASIMOS Athanasios 2009b. The Bare-Stem Constraint in Greek Compound Formation. *Gengo Kenkyu* 135, pp. 29-48.
- RAMSHAW Lance & MITCH Marcus 1995. Text Chunking Using Transformation-Based Learning. In D. Yarovsky & K. Church (eds.) *Proceedings of the Third Workshop on Very Large Corpora*, pp. 82-94. New Jersey: Association for Computational Linguistics. [🔴]
- RATNAPARKHI Adwait 1996. A Maximum Entropy Model for Part-of-Speech Tagging. In E. Brill & K. Church (eds.) *Proceedings of Empirical Methods in Natural Language Processing*, pp. 133-142. University of Pennsylvania.
- RATNAPARKHI Adwait 1998. *Maximum Entropy Models for Natural Language Ambiguity Resolution*. PhD Thesis. University of Pennsylvania.

- RATNAPARKHI Adwait 1999. Learning to Parse Natural Language with Maximum Entropy Models. In C. Claire & R. Mooney (eds.) *Machine Learning* **34**, pp. 151-175. The Netherlands: Kluwer Academic Publishers. [🔴]
- REVITHIADOU Anthi & SPYROPOULOS Vassilios 2009. The morphology of past in Greek. *Studies in Greek Linguistics* **29**, pp. 108-122.
- RIEHMANN Suzanne 1998. Type-based Derivational Morphology. *Journal of Comparative Germanic Linguistics* **2**, pp. 44-77.
- RICE Keren 2000. *Morpheme order and semantic scope*. Cambridge: Cambridge University Press.
- ROARK Brian & SPROAT Richard 2007. *Computational Approaches to Morphology and Syntax*. Oxford: Oxford University Press.
- ROSENFELD Roni 1996. A maximum entropy approach to adaptive statistical language modeling. *Computer Speech and language* **10**, pp. 187-228.
- RUBACH Jerzy & BOOIJ Geert 2001. Allomorphy in Optimality Theory: Polish Iotation. *Language* **77** (Volume 1), pp. 26-60.
- RUGE Hans 1969. *Zur Entstehung der neugriechischen Substantiv-deklination*. Stockholm: Almqvist & Wicksell.
- RUMELHART David & MCCLELLAND, James 1986. On learning the past tense of English verbs. In J. McClelland & D. Rumelhart (eds.) *Parallel Distributed Processing*, Volume **2**, pp. 216-271. Cambridge, MA: MIT Press.
- SALOMAA Arto 1985. *Computation and Automata, Encyclopedia of Mathematics and its Applications*, Volume 25. Cambridge.
- SAMUEL Ken 1998. Lazy Transformation-based Learning. In *Proceedings of the 11<sup>th</sup> International Florida Artificial Intelligence Research Symposium Conference*, pp. 235-239. [🔴]
- SCALISE Sergio 1994. *Morfologia*. Bologna: Il Mulino.
- SCHONE Patrick & JURAFSKY Daniel 2001a. Knowledge-free induction of inflectional morphologies. In *Proceedings of the 2<sup>nd</sup> Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pp. 1-9.
- SCHONE Patrick & JURAFSKY Daniel 2001b. Language-independent Induction of Part of Speech Class Labels using only Language Universals. In *Proceedings of the Workshop "Text Learning: Beyond Supervision" of UCAI 2001*, pp. 53-60.

- SCHONE Patrick & JURAFSKY Daniel 2001c. Is Knowledge-Free Induction of Multiword Unit Dictionary Headwords a solved problem? In *Proceedings of Empirical Methods of Natural Language Processing*. Pittsburgh, PA.
- SCHÜLLER Gerard & LORENZ Oliver 1994. LA-Morph – Ein linksassoziatives Morphologiesystem. In *Linguistische Verifikation*, pp. 103-119.
- SCHVEIGER Paul & MATHE Jacob 1965. Analyse d' information de la déclinaison du substantif en hongrois (du point de vue de la transduction automatique). *Cahiers de Linguistique Théorique et Appliquée* **2**, pp. 263-265.
- SEILER Hansjakob 1958. Zur Systematik und Entwicklungsgeschichte der griechischen Nominaldeklinatation. *Glotta* **37**, pp. 41-67.
- SELKIRK Elisabeth 1982. *The Syntax of Words*. Cambridge, Mass: MIT Press.
- SELKIRK Elisabeth 1984. On the major class features and Syllable Theory. In M. Aronoff and R. T. Oehrle (eds.) *Language Sound Structure*, pp. 107-136. Cambridge, MA: MIT Press.
- SGARBAS Kyriakos, FAKOTAKIS Nikos & KOKKINAKIS George 1995. A PC-KIMMO-based Morphological Description of Modern Greek. *Literary and Linguistic Computing* **10** (3), pp. 189-201. Oxford: Oxford University Press.
- SGARBAS Kyriakos, FAKOTAKIS Nikos & KOKKINAKIS George 2000. A Straightforward Approach to Morphological Analysis and Synthesis. *Proceedings COMPLEX 2000, Workshop on Computational Lexicography and Multimedia Dictionaries*, pp. 31-34.
- SIEFKES Christian 2003. *Transformation-based Learning*. On-Line notes [<http://www.cs.jhu.edu/~rflorian/fntbl/tbl-toolkit/tbl-toolkit.html>]
- SIPSER Michael 1998. *Introduction to the Theory of Computation*. (Chapter 1: Regular Languages, pp. 31-90). PWS Publishing.
- SKOUSEN Royal 2002. An overview of analogical modeling. In R. Skousen, D. Lonsdale & D. Parkinson (eds.) *Analogical Modeling*, pp. 11-26. Amsterdam: John Benjamins.
- SPENCER Andrew 1991. *Morphological Theory*. Blackwell Textbooks in Linguistics Series. Oxford: Blackwell Publications.
- SPENCER Andrew 1998. Morphological operations. In A. Spencer & A. Zwicky *The Handbook of Morphology*. Blackwell, pp. 11-148.

- SPROAT Richard 1992. *Morphology and Computation*. Cambridge: MIT Press.
- SPROAT Richard & EGEDI Dania 1988. Connectionist networks and natural language morphology. In *Conference on Grammar and Language Processing*, pp. 67-75.
- STANLEY Richard 1967. Redundancy rules in Phonology. *Language* **43**, pp. 393-436.
- ŠTEKAUER Pavol 1996. *A theory of Conversion in English*. Grove/ Atlantic: Peter Lang Publishing.
- STOTSKII È. D. 1971, Formal Grammars and Constraints on Derivation, *Probl. Peredachi Inf.* **7**(1), pp. 87–101.
- STROPPA Nicolas & YVON Francois 2005. *Formal models of analogical relationships*. Technical Report 2006D008, École Nationale Supérieure des Télécommunications.
- STUBBLEBINE Tony 2003. *Regular Expressions Pocket Reference*. O' Reilly.
- STUMP Gregory 1995. The Uniformity of Head Marking in Inflectional Morphology. In G. Booij & J. van Marle (eds.) *Yearbook of Morphology 1994*, pp. 245-295. Netherlands: Kluwer Academic Publishers.
- STUMP Gregory 2001a. Inflection. In A. Spencer and A.M. Zwicky (eds.), *The handbook of Morphology*, pp. 13-43. Oxford: Blackwell Publishers.
- STUMP Gregory 2001b. *Inflectional Morphology*. Cambridge: Cambridge University Press.
- SUH Yunju 2010. Korean Suffix Allomorphy in Optimality Theory. In *Proceedings of SUNY CUNY NYU Mini Conference* **8**. [🔴]
- SZEMERÉNYI Oswald 1999. *Introduction to Indo-European Linguistics*. Oxford: Oxford University Press.
- THORNTON Anna-Maria 1997. Stem allomorphs, suffix allomorphs, interfixes or different suffixes? On Italian derivatives with antesuffixal glides. In G. Booij, S. Scalise & A. Ralli (eds.) *Proceedings of the 1<sup>st</sup> Mediterranean Meeting of Morphology*, pp. 86-98. Mytilene, Greece
- VAN COETSEM Frans 1993. Ablaut and Reduplication in the Germanic Verb. *Indogermanische Bibliothek Vol. 3*. Heidelberg: Winter Verlag.
- VAN DER VEER Bart & BOOIJ Geert [to appear]. Allomorphy in OT: the Italian mobile diphthongs. In Bernard Tranel (ed.) *Understanding Allomorphy: Perspectives from OT*. London: Equinox. [🔴]

- VAUQUOIS Bernard 1965. *Présentation d' un programme d' analyse morphologique russe*. Technical report, Centre d' Etudes pour la Traduction Automatique, Université de Grenoble 1.
- VEILLON G., VEYRUNES J. & VANQUOIS B. 1967. Un métalangage de grammaires transformationnelles: applications aux problèmes de transferts et de génération syntaxiques. *Proceedings of the International Conference on Computational Linguistics (1967)*, pp. 1-50.
- VESELINOVA Ljuba 2000. Suppletion in the paradigms of copula verbs. In B. Fox, D. Jurafsky & L. Michaelis (eds.) *Cognition and Function in Language*, pp. 65-83. Stanford: Centre for the Study of Language and Information.
- VESELINOVA Ljuba 2003. *Suppletion in verb paradigms: bits and pieces of the puzzle*. PhD Thesis. Stockholm: John Benjamins.
- VESELINOVA Ljuba 2007. Suppletion from a typological perspective. In M. Miestamo & B. Wälchli (eds.) *New Challenges in Typology*. Berlin: Mouton de Gruyter.
- VIRPIOJA Sami & KOHONEN Oskar 2009. Unsupervised Morpheme Discovery with Allomorphs. In *Working Notes for the CLEF 2009 Workshop*. Corfu, Greece. []
- WANNER Dieter 1972. The Derivation of Inflectional Paradigms in Italian. In J. Casagrande and B. Sasiuk (eds.) *Generative Studies in Romance Languages*. Rowley, Massachusetts: Newbury House.
- WALLIS Sean 2008. Searching treebanks and other structured corpora. Chapter 34 in A. Lüdeling. & M. Kytö (eds.) *Corpus Linguistics: An International Handbook*. Handbücher zur Sprache und Kommunikationswissenschaft series. Berlin: Mouton de Gruyter
- WESTERMANN Gert & PLUNKETT Kim 2007. Connectionist Models of Inflection Processing. *Lingue e Linguaggio VI.2 (2007)*, pp. 291-311. Bologna: Il Mulino.
- WICENTOWSKI Richard 2002. *Modeling and Learning Multilingual Inflectional Morphology in a Minimally Supervised Framework*. PhD Thesis. Baltimore, Maryland: The Johns Hopkins University.
- WICENTOWSKI Richard 2004. Multilingual noise-robust supervised morphological analysis using the WordFrame model. In *Proceedings of the 7<sup>th</sup> Meeting of the ACL Special Interest Group in Computational Phonology: Current Themes in*

- Computational Phonology and Morphology*, pp. 70-77. Morristown, New Jersey: Association for Computational Linguistics.
- WILLIAMS Eddie 1981. On the notions of 'Lexical Related' and 'Head of the Word'. *Linguistic Inquiry*, pp. 245-274.
- WOODFORD Helen 1999. Animacy Hierarchy effects on object agreement. In P. Kotey (ed.) *New Dimensions in African Linguistics and Languages* (Trends in African Linguistics **3**), pp. 203-216.
- WOTHKE Klaus & SCHMIDT Richard 1992. A morphological segmentation procedure for German. *Sprache und Datenverarbeitung* **16**(1), pp. 15-28.
- WOYNA A. 1962. *Morphological analysis of Polish verbs in terms of machine translation*. Technical Report, Machine Translation Research Project, Georgetown University.
- WURZEL, Wolfgang Ulrich 1984. *Flexionmorphologie und Natürlichkeit*. Studia Germanica XXI. Berlin: Akademie-Verlag.
- WURZEL, Wolfgang Ulrich 1987. Paradigmenstrukturbedienungen: Aufbau und Veränderung von Flexionsparadigmen. In A. G. Ramat, O. Carruba & G. Bernini (eds.) *Papers from the 7<sup>th</sup> International Conference on Historical Linguistics*, pp. 629-644. Amsterdam: Benjamins.
- WURZEL, Wolfgang Ulrich 1989. Review of Andrew Carstairs, Allomorphy in Inflection. In G. Booij & J. van Marle (eds.) *Yearbook of Morphology* **2**, pp. 205-227. Dordrecht: Foris.
- XYDOPOULOS Georgios 2009. *Online Glossary of linguistic terminology: Greek-English & English-Greek*. <http://users.uoi.gr/gjxydo/lexicon/glossary.html>
- YAROWSKY David & WICENTOWSKI Richard 2001. Minimally supervised morphological analysis by multimodal alignment. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 160-167.
- YOSHIDA Kazuhiro 2006. *AMIS – A maximum entropy estimator for feature forests*. AMIS Manual. Department of Computer Science, University of Tokyo. Available at: <http://www-tsujii.is.s.u-tokyo.ac.jp/amis/manual.html>.
- YU Zheng 1997. Regular Languages. In G. Rozenberg and A. Salomaa (eds.) *Handbook of Formal Languages*, Volume **1**, p. 41-110. Springer.



- YVON Francois & STROPPA Nicolas 2007. Proportions in Lexicon: (Re)Discovering Paradigms. *Lingue e Linguaggio* **VI.2** (2007), pp. 201-226. Bologna: Il Mulino.
- ZAMPOLLI Antonio 1990. Summary of the Activities of the Instituto di Linguistica Compuazionale. *Computers and Humanities* **24**, pp. 411-416. Dordrecht: Kluwer.

# Ελληνική Βιβλιογραφία

- ADRADOS Francisco Rodriguez 2003. *Ιστορία της Ελληνικής Γλώσσας*. Αθήνα: εκδ. Παπαδήμα.
- ΑΝΑΣΤΑΣΙΑΔΗ-ΣΥΝΕΩΝΙΔΗ Άννα 1997. *Εισαγωγή στη Μορφολογία: Παραγωγή της Νέας Ελληνικής*. Θεσσαλονίκη: Πανεπιστημιακές παραδόσεις.
- ΑΝΑΣΤΑΣΙΑΔΗ-ΣΥΜΕΩΝΙΔΗ Άννα & ΧΕΙΛΑ-ΜΑΡΚΟΠΟΥΛΟΥ Δέσποινα 2003. Συγχρονικές και Διαχρονικές τάσεις στο γένος της Ελληνικής: μία θεωρητική προσέγγιση. Στο Α. Αναστασιάδη-Συμεωνίδη, Α. Ράλλη & Δ. Χειλά-Μαρκοπούλου (επιμ.) *Το Γένος*, σσ. 13-56. Αθήνα: εκδόσεις Πατάκης.
- ΒΛΑΧΟΥ Βασιλική 1994. *Τα ρήματα (ομοιά και ανώμαλα) της Αρχαίας Ελληνικής*. Αθήνα: εκδόσεις Gutenberg.
- ΓΑΒΡΙΗΛΙΔΟΥ Ζωή 2009. Από την στέρηση στην επίταση: η περίπτωση του α-/αν. *Πρακτικά του Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας*, σσ. 674-483. [🔴]
- ΓΑΛΙΩΤΟΥ Ελένη & ΡΑΛΛΗ Αγγελική 2002. Υπολογιστική Επεξεργασία των Συνθέτων της Νέας Ελληνικής. *Μελέτες για την Ελληνική Γλώσσα 2001*, σσ. 135-147. Θεσσαλονίκη: εκδοτικός οίκος αφών Κυριακίδη.
- CHANTRAINE Pierre 1998. *Ιστορική μορφολογία της ελληνικής γλώσσας*. Αθήνα: εκδ. Καρδαμίτσα.
- DEBRUNNER Albert 2008 (1917). *Ο σχηματισμός των λέξεων στην Αρχαία Ελληνική* (Griechische Wortbildungslehre). Μτφ. Η. Τσιριγκάκης. Επιμ. Ε. Πετρούνας. Επανεκτύπωση από τον εκδοτικό οίκο Bibliolife (2009). Ηλεκτρονική διάθεση συγγραμματος: [http://www.greek-language.gr/greekLang/ancient\\_greek/tools/structure/contents.html](http://www.greek-language.gr/greekLang/ancient_greek/tools/structure/contents.html)
- ΔΗΜΕΛΑ Ελεονώρα 2005. *Προθηματοποίηση στην Κρητική Διάλεκτο: η περίπτωση των συν-, σ(ι)ο-, σω-*. Αδημοσίευτη μεταπτυχιακή εργασία. Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών.
- ΔΗΜΕΛΑ Ελεονώρα 2009α. Το ακρο- ως αριστερό συστατικό σύνθετων δομών. In Α. Karasimos et al. (eds.), *1<sup>st</sup> Patras International Conference of Graduate Students in Linguistics*, pp. 210-221.

- ΔΗΜΕΛΑ Ελεονώρα 2009β. Το σιο-/σο- στην Κρητική Διάλεκτο: μία περίπτωση γραμματικοποίησης;. Στο Γ. Γιαννάκης *et al.* (επιμ.), *Ηλεκτρονικά Πρακτικά 8<sup>ου</sup> Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας*, σσ. 784-797. [🔴]
- ΔΗΜΕΛΑ Ελεονώρα 2010. *Η προθηματοποίηση στις Νεοελληνικές Διαλέκτους: Συγχρονική και Διαχρονική προσέγγιση*. Διδακτορική διατριβή. Πανεπιστήμιο Πατρών.
- ΔΙΑΚΟΓΙΩΡΓΗ Κλεοπάτρα 2000α. *Ψυχολογολογία I*. Εγχειρίδιο (σημειώσεις) μαθήματος. Τμήμα Παιδαγωγικού. Εκτυπωτικό Κέντρο Πανεπιστημίου Πατρών.
- ΔΙΑΚΟΓΙΩΡΓΗ Κλεοπάτρα 2000β. *Ψυχολογολογία II*. Εγχειρίδιο (σημειώσεις) μαθήματος. Τμήμα Παιδαγωγικού. Εκτυπωτικό Κέντρο Πανεπιστημίου Πατρών.
- ΕΥΘΥΜΙΟΥ Αγγελική 2001. Το Νεοελληνικό πρόθημα ζε-: Οι έννοιες της Απομάκρυνσης και της Αλλαγής Κατάστασης. *Μελέτες για την Ελληνική Γλώσσα* 2000, σσ. 202-213.
- ΕΥΘΥΜΙΟΥ Αγγελική 2002. Σημασιολογικές παρατηρήσεις για τα Νεοελληνικά προθήματα ζε-, εκ-, από-. *Μελέτες για την Ελληνική Γλώσσα* 2001, σσ. 199-209.
- ΕΥΘΥΜΙΟΥ Αγγελική 2003. Προθήματα ή πρώτα συνθετικά που δηλώνουν επίταση στη Νέα Ελληνική. *Μελέτες για την Ελληνική Γλώσσα* 2002, σσ. 519-528.
- ΘΩΜΑΔΑΚΗ Ευαγγελία 1994. *Μορφολογικά προβλήματα της Νεοελληνικής: Η Κλίση του Ουσιαστικού*. Διδακτορική Διατριβή. Πανεπιστήμιο Αθηνών.
- HOLTON David, MACKRIDGE Peter & ΦΙΛΙΠΠΑΚΗ-WARBURTON Ειρήνη 1999. *Γραμματική της Ελληνικής Γλώσσας*. Μτφρ. Β. Σπυρόπουλος. Αθήνα: εκδόσεις Πατάκης.
- ΚΑΡΑΝΤΖΟΛΑ Ελένη & ΓΙΑΝΝΟΥΛΟΠΟΥΛΟΥ Γιαννούλα 2000. Σημασιολογικά στοιχεία για τη σύνθεση και την παραγωγή στην Πρώιμη Νεοελληνική. *Μελέτες για την Ελληνική Γλώσσα* 1999, σσ. 193-202.
- ΚΑΡΑΣΙΜΟΣ Αθανάσιος 2001. *Αλλομορφία στην κλίση και τη σύνθεση της Νέας Ελληνικής*. Αδημοσίευτη πτυχιακή εργασία, Πανεπιστήμιο Πατρών.
- ΚΑΡΑΣΙΜΟΣ Αθανάσιος 2002. *Το Σκήπτρο του Φοίνικα: Η Αναζήτηση του Σκήπτρου*. [Βιβλίο 1<sup>ο</sup>] Αδημοσίευτο μυθιστόρημα. Κατερίνη.
- ΚΑΡΑΣΙΜΟΣ Αθανάσιος 2009. Μη-Εποπτευόμενη Μορφολογική Μάθηση και το πρόβλημα της Αλλομορφίας στην Ελληνική Γλώσσα. In A. Karasimos *et al.* (eds.) *Proceedings of the 1<sup>st</sup> Patras International Conference of Graduate*

- students in Linguistics*, σσ. 123-138. Πάτρα: Εκτυπωτικό Κέντρο Πανεπιστημίου Πατρών.
- ΚΑΡΑΣΙΜΟΣ Αθανάσιος 2011. Θεματική Αλλομορφία στη Νέα Ελληνική: ο περιορισμός του μοναδικού αλλομόρφου συμμετέχοντα. Στα *Proceedings of the 2<sup>nd</sup> Patras International Conference of Graduate students in Linguistics*. Πάτρα: Εκτυπωτικό Κέντρο Πανεπιστημίου Πατρών.
- ΚΑΡΑΣΙΜΟΣ Αθανάσιος (in progress). *Το Σκήπτρο του Φοίνικα: Τα τρία μονοπάτια της μοίρας*. [Βιβλίο 2<sup>ο</sup>]. Αδημοσίευτο μυθιστόρημα. Κατερίνη. [🔴]demo
- ΚΙΤΣΑ Παρασκευή 2001. *Ζητήματα παραγωγής του Λεξιλογίου: η αλλομορφία στα Νέα Ελληνικά με προοπτικές εφαρμογής στη διδασκαλία της Νέας Ελληνικής σε ξένους*. Διδακτορική Διατριβή. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.
- ΚΛΑΙΡΗΣ Χρήστος & ΜΠΑΜΠΙΝΙΩΤΗΣ Γεώργιος 1996. *Γραμματική της Νέας Ελληνικής, Δομολειτουργική – Επικοινωνιακή: Το όνομα*, τομ.1<sup>ος</sup>, Αθήνα: Ελληνικά Γράμματα.
- ΚΟΝΤΟΣ Γεώργιος, ΓΑΛΙΩΤΟΥ Ελένη, ΡΑΛΛΗ Αγγελική & ΤΖΕΒΕΛΕΚΟΥ Μαρία 1989. Υπολογιστική Επεξεργασία της Ελληνικής Γλώσσας: μία πρώτη προσέγγιση. *Μελέτες για την Ελληνική Γλώσσα 1988*, σσ. 505-526. Θεσσαλονίκη: Εκδοτικός Οίκος αφών Κυριακίδη
- ΚΟΝΤΟΣΟΠΟΥΛΟΣ Νικόλαος 2001. *Διάλεκτοι και ιδιώματα της Νέας Ελληνικής* (2<sup>η</sup> έκδ.). Αθήνα: εκδόσεις Γρηγόρης.
- ΚΟΥΚΚΙΔΗΣ Κωνσταντίνος 1950. *Λεξιλόγιον Ελληνικών Λέξεων παραγόμενων εκ της Τουρκικής*. Αθήνα: Εταιρεία Θρακικών Μελετών.
- ΚΟΥΤΣΟΜΠΟΓΕΡΑ Μαρία 2004. *Η χρήση γλωσσικών εργαλείων στη μορφολογική ανάλυση: αντιμετώπιση προβληματικών περιπτώσεων*. Διπλωματική Εργασία στη «Τεχνολογία II». Αθήνα: Εθνικό Καποδιστριακό Πανεπιστήμιο.
- ΚΟΥΤΣΟΥΚΟΣ Νικόλαος 2010. Σχέση κλίσης και παραγωγής: Η εμφάνιση εσωτερικής αύξησης στους προθηματοποιημένους ρηματικούς τύπους της ΝΕ. Στα *Πρακτικά 5<sup>ης</sup> Συνάντησης Εργασίας Μεταπτυχιακών Φοιτητών του Τμήματος Φιλολογίας του Πανεπιστημίου Αθηνών*, σσ. 177-185. Αθήνα: Εθνικό Καποδιστριακό Πανεπιστήμιο.
- ΚΡΥΣΤΑΛ Ντέιβιντ (2003). *Λεξικό Γλωσσολογίας και φωνητικής*. Μετάφραση: Γ. Ξυδόπουλος. Αθήνα: εκδ. Πατάκης.

- ΜΑΛΙΚΟΥΤΗ Αγγελική 1970. *Μετασηματική μορφολογία του Νεοελληνικού ονόματος*. Διδακτορική Διατριβή. Αθήνα: Βιβλιοθήκη της εν Αθήναις Φιλεκπαιδευτικής Εταιρείας.
- ΜΑΛΙΚΟΥΤΗ-DRACHMAN Αγγελική & DRACHMAN Gaberell 1989. Τονισμός στα Ελληνικά. *Μελέτες για την Ελληνική Γλώσσα 1988*, σσ. 127-145.
- ΜΑΛΙΚΟΥΤΗ-DRACHMAN Αγγελική & DRACHMAN Gaberell 1992. Θεωρητικά προβλήματα τονισμού των κλιτικών στα Νέα Ελληνικά. *Μελέτες για την Ελληνική Γλώσσα 12*, σσ. 83-104.
- ΜΑΛΙΚΟΥΤΗ-DRACHMAN Αγγελική & DRACHMAN Gaberell 1993. Σύγκριση ρηματικού τονισμού Κοινής και διαλέκτων. *Μελέτες για την Ελληνική Γλώσσα 13*, σσ. 143-161.
- ΜΑΡΚΟΠΟΥΛΟΣ Γεώργιος 1998. *Υπολογιστική επεξεργασία του Νεοελληνικού Ονόματος*. Διδακτορική διατριβή. Αθήνα: Εθνικό Καποδιστριακό Πανεπιστήμιο.
- ΜΕΛΙΣΣΑΡΟΠΟΥΛΟΥ Δήμητρα 2007. *Μορφολογική περιγραφή και ανάλυση του μικρασιατικού ιδιώματος της περιοχής Κυδωνίων και Μοσχονησίων: η παραγωγή λέξεων*. Διδακτορική Διατριβή. Πάτρα: Πανεπιστήμιο Πατρών.
- ΜΕΛΙΣΣΑΡΟΠΟΥΛΟΥ Δήμητρα & ΡΑΛΛΗ Αγγελική 2009. Συνδυαστικότητα των παραγωγικών επιθημάτων της Νέας Ελληνικής: μία πρώτη προσέγγιση. *Μελέτες για την Ελληνική γλώσσα 29*, σσ. 97-107.
- ΜΙΝΝΙΤΙ-ΓΚΩΝΙΑ Domenica 2007. Ιταλικά δάνεια στην Ελληνική αργκό και διαδικασίες προσαρμογής τους. *Γλωσσικός Περίπλους* (Μελέτες αφιερωμένες στη Δήμητρα Θεοφανοπούλου-Κοντού), σσ. 200-221. Αθήνα: Καρδαμίτσα – Τομέας Γλωσσολογίας Πανεπιστημίου Αθηνών.
- ΜΙΡΑΜΒΕΛ André 1978. *Η νέα Ελληνική γλώσσα. Περιγραφή και ανάλυση*. Θεσσαλονίκη: Ινστιτούτο Νεοελληνικών Σπουδών Μανόλη Τριανταφυλλίδη.
- ΜΠΑΜΠΙΝΙΩΤΗΣ Γεώργιος 1972. *Το ρήμα της ελληνικής: Δομικά εξελίξεις και κατηγοριοποιήσεις του ρήματος της Ελληνικής – Αρχαίας και Νέας*. Αθήνα: εκδ. Σαριπόλου.
- ΜΠΑΜΠΙΝΙΩΤΗΣ Γεώργιος 1977. *Ιστορική Γραμματικής της Αρχαίας Ελληνικής Γλώσσας, II. Μορφολογία*. Αθήνα: εκδόσεις Γ. Γκέλμπεσης.
- ΜΠΑΜΠΙΝΙΩΤΗΣ Γεώργιος 1985. *Ιστορική Γραμματικής της Αρχαίας Ελληνικής Γλώσσας, I. Φωνολογία*. Αθήνα: εκδόσεις Γ. Γκέλμπεσης.

- ΜΠΑΜΠΙΝΙΩΤΗΣ Γεώργιος 2002. *Λεξικό της Νέας Ελληνικής Γλώσσας*. Αθήνα: Κέντρο Λεξικολογίας Ε.Π.Ε (2005: 2<sup>η</sup> έκδοση εμπλουτισμένη).
- ΜΠΑΜΠΙΝΙΩΤΗΣ Γεώργιος & ΚΟΝΤΟΣ Παναγιώτης 1967. *Συγχρονική Γραμματική της Κοινής Νέας Ελληνικής: Θεωρία και Ασκήσεις*. Αθήνα.
- NESPOR Marina 1999. *Φωνολογία*. Απόδοση: Αγγελική Ράλλη. Αθήνα: εκδόσεις Πατάκη.
- ΝΙΚΟΛΟΥ Καλομοίρα 2003. *Μορφολογική και φωνολογική ανάλυση των μονολεκτικών σύνθετων λέξεων της Ελληνικής*. Αδημοσιεύτη Μεταπτυχιακή Εργασία. Πανεπιστήμιο Αιγαίου.
- ΟΙΚΟΝΟΜΟΥ Μιχαήλ 1971. *Γραμματική της Αρχαίας Ελληνικής*. Θεσσαλονίκη: Ινστιτούτο Νεοελληνικών Σπουδών ΑΠΘ, Ίδρυμα Μανόλη Τριανταφυλλίδη.
- ΠΕΤΡΟΠΟΥΛΟΥ Ευανθία (in progress). *Υπολογιστική επεξεργασία των Νεοκλασικών συνθέτων Ελληνικής και Αγγλικής*. Διδακτορική διατριβή. Πάτρα: Πανεπιστήμιο Πατρών.
- ΠΕΤΡΟΠΟΥΛΟΥ Ευανθία & ΚΑΡΑΣΙΜΟΣ Αθανάσιος 2010. Υπολογιστική Ανάλυση των παραγωγικών προσφυσμάτων και δεσμευμένων θεμάτων της Ελληνικής: Το πείραμα του Linguistica. Στα *Πρακτικά 5ης Εργασίας Μεταπτυχιακών Φοιτητών του Τμήματος Φιλολογίας του Πανεπιστημίου Αθηνών*, σσ. 233-242. Αθήνα: Εθνικό Καποδιστριακό Πανεπιστήμιο.
- ΠΕΤΡΟΥΝΙΑΣ Ευάγγελος 1984. *Νεοελληνική γραμματική και συγκριτική ανάλυση. Μέρος Α: Θεωρία*. Θεσσαλονίκη: University Studio Press.
- ΡΑΛΛΗ Αγγελική 1984. Μορφολογία του Ελληνικού Ρήματος και Θεωρία του Λεξικού: Μερικές Προκαταρκτικές Παρατηρήσεις. *Μελέτες για την Ελληνική γλώσσα* 1984, σσ. 61-80.
- ΡΑΛΛΗ Αγγελική 1986. Κλίση και Παραγωγή. *Μελέτες για την Ελληνική Γλώσσα* 1986, σσ. 29-49.
- ΡΑΛΛΗ Αγγελική 1989. Τα ρηματικά Σύνθετα της Νέας Ελληνικής. *Μελέτες για την Ελληνική Γλώσσα* 1988, σσ. 205-221.
- ΡΑΛΛΗ Αγγελική 1993. Η Θεωρία των Χαρακτηριστικών και η Δομή των Κλιτών Λέξεων της Νέας Ελληνικής. *Μελέτες για την Ελληνική Γλώσσα* 1992, σσ. 193-212. Θεσσαλονίκη: Αφοί Κυριακίδη.
- ΡΑΛΛΗ Αγγελική 2005. *Μορφολογία*. Αθήνα: εκδόσεις Πατάκη.

- ΡΑΛΛΗ Αγγελική 2008. *Η Σύνθεση λέξεων: διαγλωσσική μορφολογική προσέγγιση*. Αθήνα: εκδόσεις Πατάκη.
- ΣΓΑΡΜΠΑΣ Κυριάκος 1997. *Τεχνικές αυτόματης μορφοσυντακτικής ανάλυσης της Ελληνικής*. Διδακτορική διατριβή. Πάτρα: Πανεπιστήμιο Πατρών.
- ΣΓΑΡΜΠΑΣ Κυριάκος, ΦΑΚΩΤΑΚΗΣ Νικόλαος & ΚΟΚΚΙΝΑΚΗΣ Γεώργιος 1999. *Μορφολογική Περιγραφή της Νέας Ελληνικής με το Μοντέλο Δύο Επιπέδων. Πρακτικά 19<sup>ης</sup> Ετήσιας Συνάντησης Εργασίας Τομέα Γλωσσολογία*, σσ. 419-433.
- ΣΕΤΑΤΟΣ Μιχαήλ 1973. *Φωνολογία της κοινής Νεοελληνικής*. Αθήνα: εκδόσεις Παπαζήση.
- SIHLER Andrew 2009. *Συγκριτική γραμματική της ελληνικής και της λατινικής*, Αθήνα: εκδ. Παπαδήμας.
- ΣΟΦΙΑΝΟΣ Νικόλαος 1997. *Γραμματική της κοινής των Ελλήνων γλώσσας*. Επιμέλεια-Εισαγωγή: Θανάσης Χ. Παπαδόπουλος. Αθήνα: εκδόσεις Κέδρος.
- ΣΤΑΜΑΤΑΚΟΣ Ιωάννης 1999. *Ιστορική Γραμματικής της Αρχαίας Ελληνικής*. Αθήνα: Βιβλιοπρομηθευτική (Ι. Σιδηρογάφης & Σία Ο.Ε.) [5<sup>η</sup> έκδοση ανατύπωσης 1949].
- ΤΡΙΑΝΤΑΦΥΛΛΙΔΗΣ Μανόλης 1941. *Νεοελληνική Γραμματική*. Θεσσαλονίκη: Ίδρυμα Μανόλη Τριανταφυλλίδη (4<sup>η</sup> έκδοση 1998).
- ΤΖΑΡΤΖΑΝΟΣ Αχιλλέας 1967. *Γραμματικής της Αρχαίας Ελληνικής Γλώσσας*. Αθήνα: Οργανισμός Εκδόσεων Διδακτικών Βιβλίων.
- ΤΣΟΠΑΝΑΚΗΣ Αγαπητός 1994. *Νεοελληνική Γραμματική*. Θεσσαλονίκη: Αφοί Κυριακίδη.
- ΦΙΛΟΣ, Παναγιώτης (υπό έκδ.). Ο μεταπλασμός των αρσενικών σε -ιος > ις και των ουδετέρων -ιον > -ιν στη μετακλασική ελληνική: φωνολογικές και μορφοφωнологικές παράμετροι. *Πρακτικά της 6<sup>ης</sup> ετήσιας ημερίδας γλωσσολογίας του Πανεπιστημίου Ιωαννίνων*.
- ΧΑΤΖΙΔΑΚΙΣ, Γεώργιος 1900. Περί του σχηματισμού των ονομάτων εις -ις, -ιν αντί -ιος, -ιον εν τα μεταγενέστερα Ελληνικά. *Γλωσσολογικά Έρευναι II*, σσ. 70-82.
- ΧΑΤΖΙΔΑΚΙΣ Γεώργιος 1905. *Μεσαιωνικά και Νέα Ελληνικά*, τόμος Α. Αθήνα: Σακελλαρίου.
- ΧΑΤΖΙΔΑΚΙΣ Γεώργιος 1907. *Μεσαιωνικά και Νέα Ελληνικά*, τόμος Α. Αθήνα: Σακελλαρίου.

*Λεξικό της Κοινής Νεοελληνικής*. Θεσσαλονίκη: Αριστοτέλειο Πανεπιστήμιο  
Θεσσαλονίκης, Ίδρυμα Μανόλη Τριανταφυλλίδη [ΙΝΣ].  
<http://www.komvos.edu.gr/dictionaries/dictonline/DictOnLineTri.htm>





# Παράρτημα

## Ευρετήριο όρων

<b>A</b>	<b>P</b>
Allomorfessor ..... 197, 198, 219	patterns..... <i>Βλέπε</i> σχεδιάτυπο
	PC-KIMMO ....122, 127, 129, 130, 131, 132, 133, 192
<b>C</b>	<b>S</b>
case-based reasoning..... <i>Βλέπε</i> μνημονική μάθηση	schema ..... <i>Βλέπε</i> σχήμα
<b>H</b>	<b>T</b>
Head-Driven Phrase Structure Grammar ..... 193	tagging ..... 119, 153, 262
<b>I</b>	TiMBL ..... 152, 264
instance-based learning ..... <i>Βλέπε</i> μνημονική μάθηση	<b>A</b>
<b>K</b>	αλλόθημα ..... 114
<i>k</i> -nearest neighbors ..... <i>Βλέπε</i> μνημονική μάθηση	αλλο-κανόνες ..... 195, 196
<b>L</b>	αλλομόρφημα..... 197
LINGUISTICA ....4, 136, 138, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 167, 169, 198, 250, 258, 261, 265, 269, 273, 276, 284, 289	αλλομορφία
αξιολόγηση υποψηφίων ..... 138	αντικαταστατική ..... 43
παραγωγή υποψηφίων ..... 138	αφαιρετική ..... 42
LINGUISTICA	εξωτερική ..... 16
παραγωγή υποψηφίων ..... 139, 140	ετεροιωτική..... 44, 98
<b>M</b>	θεματική .....24, 41, 90, 100, 230, 249
Maximum A Posteriori ..... 197	λεξική ..... 9
memory-based reasoning .. <i>Βλέπε</i> μνημονική μάθηση	μηδενική ..... 41
Morfessor Baseline ..... 197	μορφολογική (καθαρή) ..... 9
<b>O</b>	προσθετική..... 42, 70, 71, 97
Open NLP Entropy Framework ..... 150	προσφυματική..... 41
	συνδυαστική ..... 44
	φωνολογική ..... 27, 194
	αλλομορφική τάξη ..... 174, 187, 188
	αλλόμορφο ..1, 2, 7, 9, 10, 12, 14, 15, 17, 18, 28, 29, 30, 31, 32, 33, 37, 40, 42, 43, 45, 46, 51, 52, 53, 54, 55, 56, 57, 58, 59, 63, 64, 65, 67, 69, 72, 73, 77, 79, 80, 81, 84, 85, 90, 93, 94, 97, 98, 100, 102, 103, 104, 108, 110, 111, 112, 113, 138, 157,

158, 160, 161, 165, 167, 169, 171, 173, 181, 192, 193, 195, 196, 197, 201, 204, 205, 209, 211, 212, 213, 214, 221, 223, 224, 225, 227, 230, 231, 241, 242, 243, 244, 249, 254, 255, 258	
εξειδικευμένο.....	28
προκαθορισμένο .....	28
αλλόφωνο.....	7
ανάκτηση πληροφορίας.....	198
ανάλυση συστατικών .....	120
ανομοίωση .....	8, 120
άπαξ λεγόμενα .....	152, 173
αποβολή .....	8, 93, 216
αποφωνία .....	43
αρχή	
αλλομορφικής συμπεριφοράς .....	79, 80
αρχή της Ομοιόμορφης Σχηματισμένης Κεφαλής	28
αυτόματος σχολιασμός κειμένων .....	233
αφομοίωση.....	7, 8, 93, 162, 211

**B**

βάρος	
ειδικό .....	174
βάρος (υπολογιστικό).....	132

**Γ**

γλώσσα	
απομονωτική.....	192
διαχυτική .....	192
πολυσυνθετική.....	192
συγκολλητική .....	133, 192
σχεδιοτυπική.....	192
γλώσσα κανονικότητας.....	123

**Δ**

δέντρα .....	122
δεσμευμέν θέμα .....	158, 161, 166, 167, 168, 256
διασυνδετισμός .....	146
δίγραμμα .....	136
διορθωτικοί κανόνες .....	13
δίφθογγος	
σπασμένος .....	10

**E**

εκφράσεις κανονικότητας .....	123
ελαχίστο μήκος περιγραφής.....	136
ελεύθερη παραλλαγή.....	45, 46, 47
εμφυχότητα .....	175
ενδεικτής.....	53, 58
εξαναγκασμός τύπου.....	116
επεξεργασίας φυσικής γλώσσας.....	134
Επιβλεπόμενη Μάθηση Μορφολογίας	
k-πλησιέστερου-γείτονα .....	245
παραδειγματική-οδηγούμενη μάθηση.....	245
επιβλεπόμενη μορφολογική μάθησ	
κανονοθετική.....	145
επιβλεπόμενη μορφολογική μάθηση	
διασυνδετισμική .....	145
μνημονική μάθηση.....	4
στοχαστική.....	145
επιθηματιστής .....	200
επιθηματοποίηση .....	55, 100, 221, 246
ετεροίωση .....	43, 44, 195, 214, 219, 227
ευρετικές στρατηγικές.....	139

**Θ**

θεματικό φωνήεν.....	31
θεματοκεντρική τεχνική.....	199
Θεωρία του Βέλτιστου .....	14, 16, 33

**I**

ιεραρχία του χαρακτηρισμού εμφυχότητας.....	175
---	-----

**K**

κανόνας	
αλλομορφικής μετατροπής .....	54, 204, 207, 209, 215, 228, 229, 238
αλλομορφικός μετασχηματιστικός.....	204
ελάσσων.....	30
μορφολεξικός.....	56, 63, 64
περιοδούμενος προσφύματος .....	13
πλεοναστικός .....	43, 44, 45, 56, 75, 204, 220
πλησιέστερου γείτονα .....	151
κανονικές εκφράσεις.....	<i>Βλέπε</i> εκφράσεις κανονικότητας

κανονική γλώσσα..... <i>Βλέπε</i> γλώσσα κανονικότητας	
κεφαλή .....	28, 105, 107
μορφολογική.....	105
σημασιολογική..... <i>Βλέπε</i>	
κλιτική τάξη.....	69, 70, 71, 72, 76, 78, 171, 174
κλιτικό κελί.....	76
κριτήριο επιλογής .....	50, 52

## Λ

λεξικά.συνέχειας .....	122
λεξικό	
αλλομορφικό.....	205
δυναμικό .....	49
σταθερό.....	49, 50
λημματισμός .....	120
λημματιστής.....	200
λημματοποίηση .....	120, 200

## Μ

μάθηση μετασηματιστικού τύπου .....	153
μέγιστη εντροπία.147, 148, 149, 150, 169, 170, 176,	
185, 189, 202, 229, 233, 236, 239, 242, 250, 254,	
258	
Μέγιστη Εντροπία.....	147
ΜεΜΜ..... <i>Βλέπε</i> μη-επιβλεπόμενη μορφολογική	
μάθηση.	
μεταβολή	
ετεροίωση .....	197
μεταπήδηση χαρακτηριστικών.....	89
μετρική απόσταση.....	151, 153
μη-επιβλεπόμενη μορφολογική μάθηση	
μικρό-μήκος για μη-θέματα.....	164
μηχανική μετάφραση .....	121
μνημονική μάθηση.....	151
Μνημονική Μάθηση.....	151
μοντέλο αποκοπής επιθημάτων.....	122
μορφολογία	
διεπίπεδη ή διεπίπεδ .....	122
προσωδιακή .....	112
μορφολογική μάθηση	
αυτόματη .....	134
Μορφολογική Μάθηση	
μη-επιβλεπόμενη .....	138
Μη-επιβλεπόμενη .....	220

μορφολογικός αναλυτής.....	122
----------------------------	-----

## N

νευρωνικά δίκτυα.....	146
νόμος της δεξιόστροφης κεφαλής .....	105

## Π

παραγωγής φυσικής γλώσσας.....	134
παραδειγματική ακεραιότητα.....	29
παρέκταση.....	72
πεπερασμένο σύστημα	
αρχική κατάσταση .....	124
δέντρα .....	122
κατάσταση αποτυχίας .....	125
λεξικά συνέχειας.....	122
μετάβαση .....	125
μετατροπέας.....	122
μετατροπέας πεπερασμένων καταστάσεων....	126
μη-ντετερμινιστικό .....	126
μορφοτακτικό μοντέλο πεπερασμένου	
συστήματος .....	122
τελική κατάσταση.....	124
πεπερασμένων καταστάσεων	
αυτόματο.....	122
μετατροπέας.....	122
περιγραφή ελαχίστου μήκους.....	140
περιορισμός	
λεξικοσυντακτικός.....	133
μοναδικού αλλομορφου συμμετέχοντα..	224, 225
μοναδικού συμμετέχοντα αλλομόρφου.....	103
μορφοτακτικός.....	193
του γυμνού θέματος.....	67
του φραγμού .....	114
πλαίσιο υποκατηγοριοποίησης.....	50, 52, 54, 55
πληροφορία περιβάλλοντος .....	52, 54, 55
πληροφορίες (μορφολογικού) περιβάλλοντος.....	60
προθηματοειδές.....	93
προθηματοποίηση .....	8, 103, 222
προσφυματοειδές .....	93

## Σ

σταθμισμένη κοινή πληροφορία .....	139
Στατιστική Μηχανική Μετάφραση .....	198

συγχώνευση .....	76, 199, 239
σύμφυμα .....	166
συναίρεση .....	35, 45, 74
σύστημα εισαγωγής κειμένων.....	122
σύστημα φυσικής γλώσσας.....	122
συστηματική αλλομορφία.....	74, 89, 211
σχεδιάτυπο .....	26, 190, 208, 232, 247, 251, 256
σχήμα.....	18
σχήμα συνολικής κατανομής .....	29
σώμα κειμένων.....	133, 134, 139

## T

ταξινόμηση	
αλλομορφική.....	76
τεμπέλικά μοντέλα μάθησης .....	152

## Y

υπόθεση του ισχυρού λεξικαλισμού.....	49
---------------------------------------	----

υποκατάσταση .....	28, 29, 33, 45, 47, 48
ασθενής.....	47
ισχυρή.....	47

## Φ

φύλλο σώματος...137, 138, 139, 140, 141, 145, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166	
κανονικό .....	140
φωνολογία	
λεξική .....	8, 32, 143
φραστική.....	16
φωνόμορφο .....	40, 46, 93, 194

## Ψ

ψευδοαλλόμορφο .....	45, 46
----------------------	--------

## Appendix I

1. Πίνακας παραγωγικών ρηματικών επιθημάτων			
Αρχικό Θέμα	Παράγωγη λέξη	Παραγωγικό επίθημα	Σημασία
(α.) Ουσιαστικό ή Επίθετο			
κλάψα	κλαψουρίζω	-(ουρ)+ίζ(ω)	υποκορισμός σε ρήμα
μάσα	μασουλίζω	-(ουλ)+ίζ(ω)	
δόξα	δοξάζω	-άζ(ω)	το υποκείμενο είναι ή γίνεται ή ενεργεί εκείνο που σημαίνει η σημασία του αρχικού θέματος
λαμπάδα	λαμπαδιάζω	-ιάζ(ω) <sup>1</sup>	
αρχή	αρχίζω	-ίζ(ω)	
πίστη	πιστεύω	-εύ(ω)	σύνδεση με ξενικές βάσεις
λάδι	λαδώνω	-ών(ω)	
ζέστη	ζεσταίνω	-αίν(ω)	
σουτ	σουτάρω	-αρ(ω)	
(β.) Επίρρημα			
συχνά	συχνάζω	-άζ(ω)	
παράμερα	παραμερίζω	-ίζ(ω)	
κοντά	κοντεύω	-εύ(ω)	
σιμά	σιμώνω	-ών(ω)	

2. Πίνακας παραγωγικών ονοματικών επιθημάτων			
Αρχικό Θέμα	Παράγωγη λέξη	Παραγωγικό μόρφημα	Σημασία
(α.) Ρήμα			
λυτρών-ω	λυτρωτής	-τη(ς)	το πρόσωπο που ενεργεί
τραγουδ-ώ	τραγουδιστής	-ιστή(ς)	
εισπράττ-ω	εισπράκτορας	-τορα(ς)	
συμβουλευέ <sup>2</sup> -ω	συμβουλάτορας	-άτορα <sup>3</sup> (ς)	
γράφ-ω	γραφιάς γραφέας	-ιά(ς) -έα(ς)	η ενέργεια ή το αποτέλεσμα της ενέργειας
ορίζ-ω	ορισμός	-μός	

<sup>1</sup> Το /i/ προέκυψε πιθανότητα από επανανάληψη του επιθήματος (Λεξικό Τριανταφυλλίδη).

<sup>2</sup> Η λέξη προέρχεται από τη λέξη «συμβουλή», ενώ ο συμβουλάτορας ίσως από τη λέξη *σύμβουλος*.

<sup>3</sup> Το /a/ προκαλεί προβληματισμό για την προέλευσή του. Ίσως από επανανάληψη. Κατά το Λεξικό του Τριανταφυλλίδη: μσν. -άτορας < αρχ. αρσ. ουσ. σε -άτωρ, αιτ. άτορα: αρχ. κοσμοκράτωρ, αιτ. κοσμοκράτ-ορα & ΛΑΤ. -ator: ΕΛΝΣΤ. μανδ-ᾶτον (δες μαντάτο) – ΜΣΝ. μαντ-άτωρ < ΛΑΤ. mandatum – mandator `που αναθέτει εντολή´.

βεβαίων-ω λάμπ-ω λήγ-ω	βεβαίωση λάμψη λήξη	-ση -*ψη -*ξη?		
βράζ-ω ψάχν-ω γράφ-ω	βράσιμο ψάξιμο γράψιμο	-σίμ(ο) -*ξιμ(ο) -*ψιμ(ο)		
ψαρεύ-ω	ψάρεμα	-μα		
στρών-ω	στρωσίδι	-ίδι		
βογκώ	βογκητό	-ητό		
ανακατώνω	ανακατωσούρα	-(σ)+ούρα*		
γενν-ώ	γέννα	-α		
κολυμπ-ώ	κολύμπι	-ι		
γελ-ώ	γέλιο	-ιο		
κρεμ-ώ	κρεμάλα	-άλα		
επιθυμ-ώ	επιθυμία	-ία		
μιλ-ώ	μιλιά	-ιά		
γοητεύω <sup>4</sup> δουλεύω προσπαθώ	γοητεία δουλειά προσπάθεια	-[ε]ία -[ε]ιά -[ε]ια		
ανεμίζω	ανεμιστήρας	-τήρα(ς)	το όργανο ή το μέσο μίας ενέργειας ή ο τόπος όπου γίνεται αυτή	
ποτίζω	ποτιστήρι	-(τή)+ρι		
πλένω	πλυντήριο	-(τή)+ρι(ο)		
θερμαίνω	θερμάστρα	-τρα		
κινώ	κίνητρο	-τρ(ο)		
γράφ-ω	γραφείο	-εί(ο)		
(β.) Ουσιαστικά				
παιδί	παιδάκι	-άκι	υποκοριστικά	
Δημήτρη-ς	Δημητράκης	-άκη(ς)		
γέροντα-ς	γεροντάκος	-άκ(ος)		
καπέλ-ο	καπελούδα	-ούδα		
άγγελος	αγγελούδι	-ούδι		
χωριό	χωριουδάκι	-ουδάκι		
κόρη	κορούλα	-ούλα		
αδερφ-ός	αδερφούλης	-ούλη(ς)		
σακί	σακούλι	-ούλι		
βοσκ-ός	βοσκόπουλο βοσκοπούλα	-ουλ(ο) -ούλα		
μήλ-ο	μηλαράκι	-(αρ)+άκι		
φίλ-ος	φιλαράκος	-(αρ)+άκ(ος)		
κράτ-ος	κρατίδιο	-ίδι(ο)		
λα-ός	λαουτζίκος	-ουτζίκ(ος)		
θάλαμ-ος	θαλαμίσκος	-ίσκ(ος)		
κασόνι	κασόνα	-α		μεγεθυντικά
γυναίκα	γυναικάρα	-άρα		
σκύλ-ος	σκύλαρος	-αρ(ος)		

<sup>4</sup> Αξίζει να σημειωθεί η αποβολή του παραγωγικού επιθήματος για την δημιουργία του ουσιαστικού.

χορευτή-ς	χορευταράς	-αρά(ς)	
φωνή	φωνακλάς	-ακλά(ς)	
χέρι	χερούκλα	-ούκλα	
σπίτι	σπιταρόνα	-(αρ)+όνα	
καμπάνα	καμπαναριό	-αρι(ό)	τόπος
τρελ-ός	τρελάδικο	-άδικο	
σουβλατζής	σουβλατζίδικο	-(τζ)+ίδικο	
φύλακας	φυλάκιο	-ιο	
δήμαρχος	δημαρχείο	-είο	
πλατάνι	πλατανιά	-ιά	το μέρος που περιέχει πολλά από εκείνα που φανερώνει η σημασία του θέματος (περιεκτικά ουσιαστικά)
πουρνάρι	πουρναριάς	-ιά(ς)	
ξέν-ος	ξενώνας	-ώνα(ς)	
περιστέρι	περιστεριώνας	-ιώνα(ς) <sup>5</sup>	
Θάσ-ος	Θασίτης	-ίτη(ς)	ο άνθρωπος που κατάγεται από ορισμένο τόπο ή που ανήσει σ' αυτόν
Μοριά-ς	Μοραϊτής	-αϊτή(ς)	
Σπάρτη	Σπαρτιάτης	-ιάτη(ς)	
Ήπειρ-ος	Ηπειρώτης	-ώτη(ς)	
Σουφλί	Σουφλιώτης	-(ι)ώτης	
Αφρική	Αφρικανός	-αν(ός)	
Ψαρά	Ψαριανός	-ιαν(ός)	
Πρέβεζα	Πρεβεζάνος	-άν(ος)	
Πάτρα	Πατρινός	-ιν(ός)	
Σμύρνη	Σμυρνιός	-ι(ός)	
Κόρινθος	Κορίνθιος	-ι(ος)	
Θήβα	Θηβαίος	-αί(ος)	
Κίνα	Κινέζος	-έζ(ος)	
ψωμί	ψωμάς	-άς	επάγγελμα
περιβόλι	περιβολάρης	-άρη(ς)	
σκουπίδι	σκουπιδιάρης	-ιάρη(ς)	
βιβλιοθήκη	βιβλιοθηκάριος	-άρι(ος)	
ψιλικά	ψιλικατζής	-τζή(ς)	
Γιώργος	Γιώργαινα	-αινα	ανδρωνυμικά
Μπουμπούλης	Μπουμπουλίνα	-ίνα	
κεράσι	κερασιά	-ιά	φυτό ή δέντρο
ρέμα	ρεματιά	-ιά	τόπος
καλοκαίρι	καλοκαιριά	-ιά	φυσικό φαινόμενο
νύχτα	νυχτιά	-ιά	χρονική έκταση
παλικάρι	παλικάρι	-ιά	ιδιότητα
βελόνα	βελονιά	-ιά	ενέργεια
φτυάρι	φτυαριά	-ιά	πόσο
λάδι	λαδιά	-ιά	μηλίδα ή μυρωδιά

<sup>5</sup> Αυτό (ι) εμφανίζεται μόνο σε παράγωγα της έβδομης κλιτικής τάξης.



Έλληνα-ς	Ελληνιστής	-(ισ)+τη(ς)	οπαδός ή θιασώτης μίας θεωρίας, εκείνος που διδάσκει μία γλώσσα και τη φιλολογία της
Ιουδαί-ος	Ιουδαϊσμός	-(ισ)+μ(ος)	θεωρία ή αίρεση
δούκα-ς	δουκάτο	-άτ(ο)	εξουσία, ένωση ή επικράτεια
κτήμα	κτηματίας	-ία(ς)	αυτός που έχει ή έκανε κάτι
αλάτι	αλατιέρα	-ιέρα	σκεύος, ιδίως φαγητού
πόρτα	πορτιέρης	-ιέρη(ς)	το πρόσωπο που ενεργεί
πρόεδρο	προεδρικό	-ικό	ιδιότητα
συγγενής	συγγενολόγος	-λό(γ)ος	σύνολο εμφύχων
λάσπη	λασπουριά	-ουριά	περιληπτικά
3. Επίθετο			
σκληρός	σκληράδα	-άδα	ιδιότητα σχετική με το θέμα του επιθέτου
μαύρος	μαυρίλα	-ίλα	
καλός	καλοσύνη	-οσύνη	
απλός	απλότητα	-ότητα	
τραχύς	τραχύτητα	-ύτητα	
κουφός	κουφαμάρα	-αμάρα <sup>6</sup>	
γλυκός	γλύκα	-α	
ζεστός	ζέστη	-η	

### 3. Πίνακας παραγωγικών μορφημάτων (επιθήματα) επιθέτων

Αρχικό Θέμα	Παράγωγη λέξη	Παραγωγικό μόρφημα	Σημασία
1. Ρήμα			
δροσίζ-ω	δροσιτικός	-(τ)+ικός	εκείνος που +ρήμα
λάμπ-ω	λαμπερός	-ερ(ός)	
πειθαρχ-ώ	πειθαρχικός	-ικός	
κιν-ώ	κινητήριο	-(τή)+ρι(ος)	
αγαπ-ώ	αγαπητός	-τός	ό,τι και η παθητική μετοχή ή εκείνος που μπορεί να δεχτεί την ενέργεια του ρήματος
φορολογ-ώ	φορολογητέος	-τέ(ος)	εκείνος που πρέπει να γίνει

<sup>6</sup> Είναι από το ρήμα και όχι από το επίθετο. Δες τα παραδείγματα

συζητ-ώ	συζητήσιμος	-σιμος	εκείνος που μπορεί να γίνει
<b>2. Ουσιαστικό</b>			
ζημιά	ζημιάρης	-άρη(ς)	εκείνος που έχει σχέση με την ιδιότητα του ουσιαστικού ή που έχει σχέση με τη σημασία του ουσιαστικού
πείσμα	πεισματάρης	-άρη(ς)	
δροσιά	δροσάτος	-άτ(ος)	
λάστιχ-ο	λαστιχένιος	-ένι(ος)	εκείνος που είναι από την ύλη που σημαίνει η πρωτότυπη λέξη
πέτρα	πέτρινος	-ίν(ος)	
βροχή	βροχερός	-ερ(ός)	εκείνος που έχει σχέση με το ουσιαστικό
θάλασσα	θαλασσής	-ή(ς)	εκείνος που έχει το χρώμα που φανερώνει η πρωτότυπη λέξη
θώλ-ος	θολωτός	-ωτ(ος)	εκείνος που έχει σχέση ή που ανήκει σε αυτό που φανερώνει η πρωτότυπη λέξη
ήλι-ος	ηλιακός	-ακ(ός)	
αδερφ-ός	αδερφικός	-ικ(ός)	
ψεύτη-ς	ψεύτικος	-ικ(ός)	
Μάη-ς	μαγιάτικος	-(ι)άτικος	
θέατρ-ο	θεατρinίστικος	-(ιν)+ίστικ(ος)	
παλικάρι	παλικαρίσιος	-ίσι(ος)	
ουραν-ός	ουράνιος	-ι(ος)	
άντρα-ς	αντρείος	-εί(ος)	
Ευκλείδη-ς	Ευκλείδειος	-ει(ος)	
πείνα	πειναλέος	-λέ(ος)	
σιωπή	σιωπηλός	-λ(ός)	
κορυφή	κορυφαίος	-αί(ος)	
βαθμ-ός	βαθμιαίος	-ιαί(ος)	
<b>3. Επίθετο</b>			
μικρ-ός	μικρούλης	-ούλη(ς)	υποκοριστική – εκείνος που έχει σε μικρό βαθμό εκείνο που σημαίνει το κυρίως μόρφημα
καλ-ός	καλούτσικος	-ούτσικ(ος)	
μαύρ-ος	μαυριδερός	-ιδερ(ός)	
βαθύ-ς	βαθουλός	-ουλ(ός)	
άγριος	αγριωπός	-ωπ(ός)	εκείνος που μοιάζει μ' αυτό

			που φανερώνει η πρωτότυπη λέξη
4. Επίρρημα			
σήμερα	σημερινός	-ιν(ός)	
αύριο	αυριανός	-ιαν(ός)	

4. Πίνακας παραγωγικών μορφημάτων (επιθήματα) επιρρημάτων			
Αρχικό Θέμα	Παράγωγη λέξη	Παραγωγικό μόρφημα	Σημασία
καλ-ός	καλά	-ά	
άμεσ-ος	αμέσως	-ως	
άλλ-ος	αλλού	-ού	
εδώ	εδώθε	-θε	