



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΕΛΟΠΟΝΝΗΣΟΥ

2^η διάλεξη

Σχεδιασμός, κατασκευή και αξιολόγηση του
ελληνικού γραμματικού διορθωτή

Δομή Παρουσίασης

- Αντικείμενο έρευνας - Θεωρητικό πλαίσιο
- Ιδιαιτερότητες της Ν. Ελληνικής Γλώσσας
- Σχεδιασμός - υλοποίηση εφαρμογής
- Αξιολόγηση εφαρμογής
- Συμπεράσματα

ΥΠΟΛΟΓΙΣΤΙΚΗ ΓΛΩΣΣΟΛΟΓΙΑ

Επεξεργασία Φυσικής Γλώσσας

- Διόρθωση ορθογραφίας, εύρεση πληροφορίας, μηχανική ή αυτόματη μετάφραση, έλεγχος γραμματικής.

Αναγνώριση Φωνής

- Στατιστική επεξεργασία σήματος, κατανόηση φυσικής γλώσσας, νευρωνικά συστήματα, αναγνώριση προτύπων, φωνολογία.

ΙΔΙΑΙΤΕΡΟΤΗΤΕΣ Ν. ΕΛΛΗΝΙΚΗΣ ΓΛΩΣΣΑΣ

Ασάφεια

Σημασιολογική

Συντακτική

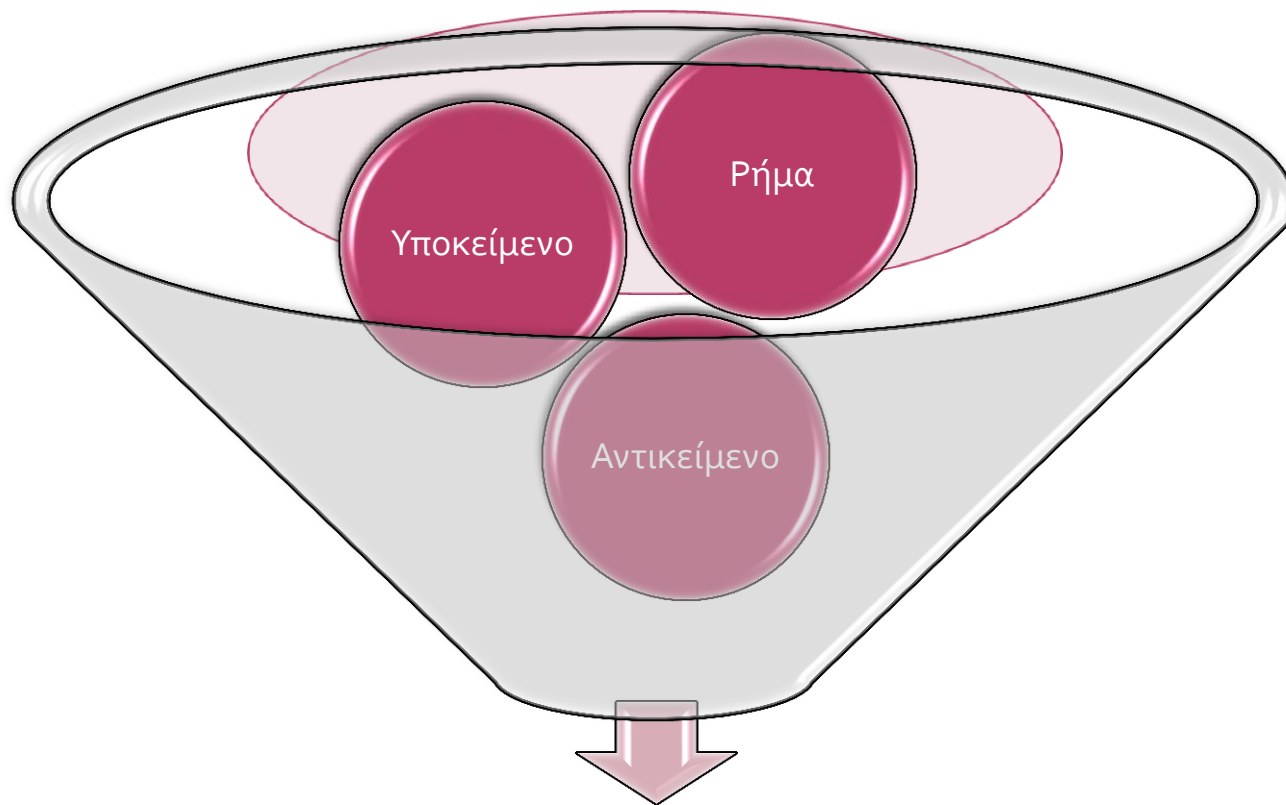
Λεξική

Ασάφεια
&
Η/Υ

κοινοί
ορθογραφικοί
τύποι >
διαφορετικά
λήμματα

Κοινοί
ορθογραφικοί
τύποι:
(διαφορετικά
μορφοσυντακτικά
χαρακτηριστικά)

2) «Ελευθερία» Μετακίνησης όρων



Ο Γιώργος αγαπά τη Μαρία

ΣΧΕΔΙΑΣΜΟΣ ΕΦΑΡΜΟΓΗΣ



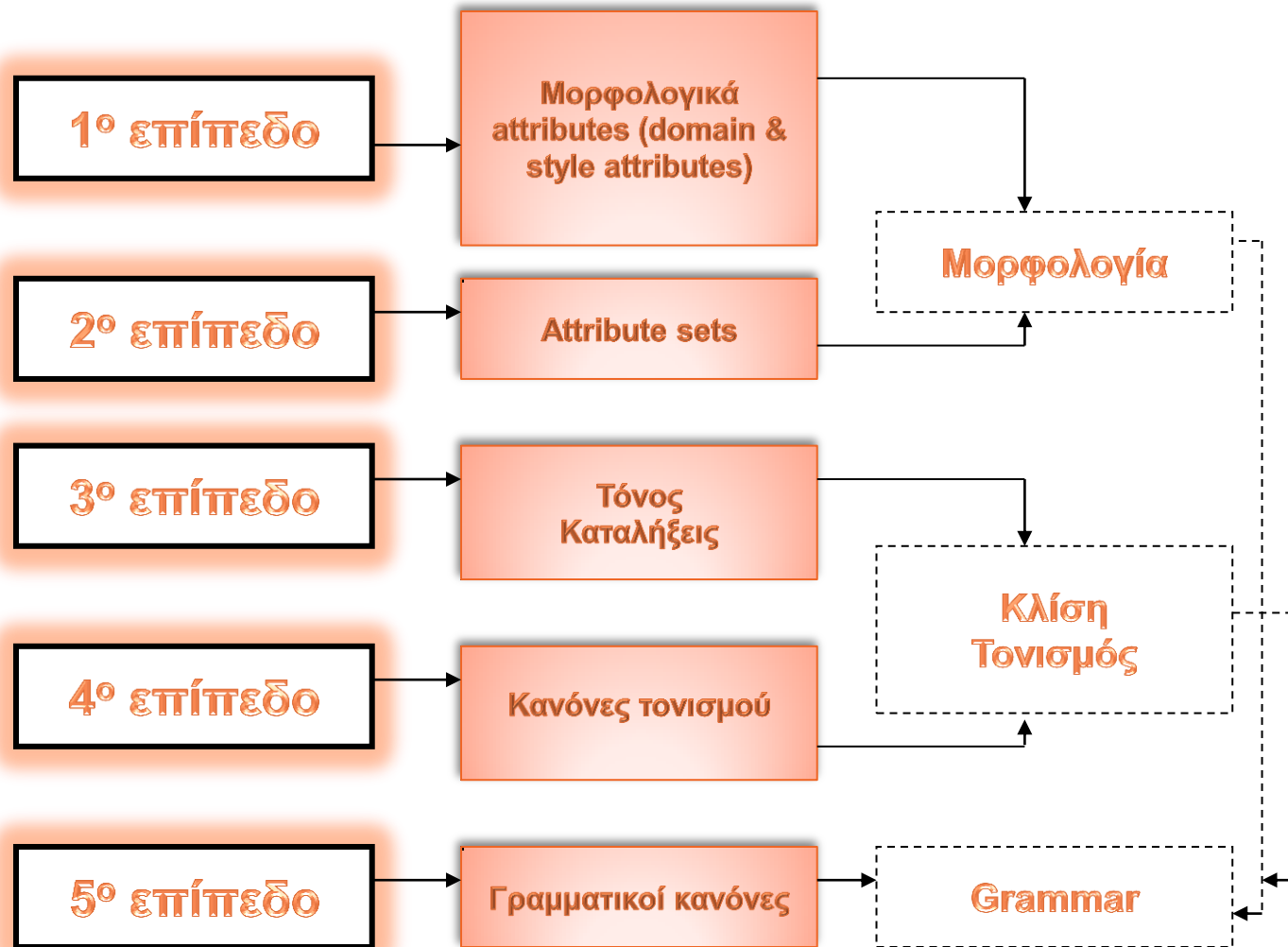
Ηλεκτρονικό λεξικό Neurolinguo

- Βάση για την υλοποίηση του γραμματικού διορθωτή είναι το μορφολογικό υπολογιστικό λεξικό που δεν προορίζεται για ανθρώπους-χρήστες αλλά για υπολογιστικές εφαρμογές όπως ο μορφοσυντακτικός σχολιασμός (morphosyntactic tagging), η συντακτική ανάλυση (parsing), ο σημασιολογικός σχολιασμός (semantic tagging) κ.ά.

Ηλεκτρονικό λεξικό Neurolinguo



Αρχιτεκτονική λεξικού



LexEdit

| Sections | Lemma | Main Morphology | # Meanings | POS | Description |
|----------|---------------------|------------------------------|------------|--------|------------------|
| άλφα | ιστοθεραπεία | {ι-στο-}{θε-ρα-π}[ει-] #... | 1 | (N) | |
| βήτα | ιστοκαλλιέργεια | {ι-στο-}{καλ-λι-}{ερ-γ}... | 1 | (N) | -ειών |
| γάμα | ιστολογία | {ι-στο-}{λο-γ}[ι-] #1# | 1 | (N) | σπάν. πληθ. |
| δέλτα | ιστολογικά | {ι-στο-}{λο-γ}[ι-κ] #1# | 1 | (ADV) | |
| έψιλον | ιστολογικός | {ι-στο-}{λο-γ}[ι-κ] #1# | 1 | (ADJ) | |
| ζήτα | ιστόλυση | {ι-στο-}{λυ-}[σ] #1# | 1 | (N) | -ης κ' -εως |
| ήτα | ιστολυσία | {ι-στο-}{λυ-σ}[ι-] #1# | 1 | (N) | |
| θήτα | ιστολυτικά | {ι-στο-}{λυ-}[τ][ι-κ] #1# | 1 | (ADV) | |
| ιγώτα | ιστολυτικός | {ι-στο-}{λυ-}[τ][ι-κ] #1# | 1 | (ADJ) | |
| κάπα | ιστόρημα | {ι-στο-ρ}[η-μ] #2# | 1 | (N) | |
| λάμδα | ιστορημένος | {ι-στο-ρ}[η-με-ν] #1# | 1 | (PART) | |
| μι | ιστόρηση | {ι-στο-ρ}[η-σ] #1# | 1 | (N) | |
| νι | ιστορία | {ι-στο-ρ}[ι-] #1# | 1 | (N) | |
| ξ | ιστορίζω | {ι-στο-ρ}[ι-ζ] #8# | 1 | (V) | -ισα, -ομαι, -ίς |
| όμικρον | ιστορικά | {ι-στο-ρ}[ι-κ][ο-][τε-ρ] ... | 1 | (ADV) | |
| πι | ιστορισμός | {ι-στο-ρ}[ι-κ][ι-σμ] #1# | 1 | (N) | |
| ρο | ιστορικό | {ι-στο-ρ}[ι-κ] #1# | 2 | (N) | |
| σίγμα | ιστορικός | {ι-στο-ρ}[ι-κ] #3# | 1 | (ADJ) | |
| ταυ | ιστορικός2 | {ι-στο-ρ}[ι-κ] #1# | 1 | (N) | |
| ύψιλον | ιστορικοσυγκριτικά | {ι-στο-ρ}[ι-κ][ο-]<συ-γ... | 1 | (ADV) | |
| φι | ιστορικοσυγκριτικός | {ι-στο-ρ}[ι-κ][ο-]<συ-γ... | 1 | (ADJ) | |
| | ιστορικότητα | {ι-στο-ρ}[ι-κ][ο-πη-τ] #... | 1 | (N) | |
| | ιστοριογνώσια | {ι-στο-ρ}[ι-][ο-]{γνω-σ... | 1 | (N) | |

LexEdit

Lexeme [μαγαζι]

```
{μα-γα-ζ}{N_NEUT_I1{NEUT_I_SING[[i(ACC | NEUT | NOM | SING | VOC) | iou(GEN | NEUT | SING)] ()]ULT1<<Λ(1)()>>() || NEUT_I_PLUR[[ia(ACC | NEUT | NOM | PLUR | VOC) | iou(GEN | NEUT | PLUR)] ()]ULT1<<Λ(1)()>>()}(N)}}.
```

```
{μα-γα-ζ}[α-κ]{N_NEUT_I2_NOGEN{NEUT_IIA[[i(ACC | NEUT | NOM | SING | VOC) | ia(ACC | NEUT | NOM | PLUR | VOC)] ()]PEN1<<Π(1)()>>()}(N)}})
```

Επιπλέον παρέχεται η πληροφόρηση ότι το {μα-γα-ζ} είναι θέμα και η παραγωγική κατάληξη [α-κ] είναι ένθημα (Infix). Κάθε λήμμα, δηλαδή, περιλαμβάνει όλα τα θέματα που ανήκουν σημασιολογικά στην

ίδια κατηγορία τα οποία και χαρακτηρίζονται από τις παραγωγικές καταλήξεις τους.

Ηλεκτρονικό λεξικό

±

| Κλιτικοί τύποι | Μορφολογία | Μορφή | Attribute |
|----------------|----------------|-------------------------------------|------------|
| μα-γα-ζι | {μα-γα-ζ} | ι(ACC NEUT NOM SING VOC.. | N_NEUT |
| μα-γα-ζιού | {μα-γα-ζ} | ιου(GEN NEUT SING)... | N_NEUT |
| μα-γα-ζιά | {μα-γα-ζ} | ια(ACC NEUT NOM PLUR VOC.. | N_NEUT |
| μα-γα-ζιών | {μα-γα-ζ} | ιων(GEN NEUT PLUR.. | N_NEUT |
| μα-γα-ζά-κι | {μα-γα-ζ}[α-κ] | [[ι(ACC NEUT NOM SING VOC.. | N_NEUT DIM |
| μα-γα-ζά-κια | {μα-γα-ζ}[α-κ] | ια(ACC NEUT NOM PLUR VOC).. | N_NEUT DIM |

Πίνακας 6. Παραγωγή μορφολογικών τύπων

Μορφοσυντακτική ανάλυση

<contents>Είναι</contents>{είμαι,**C_P+IND+PASS+PLUR+PRES+SING+SUBJ+V,**}
{είναι,**ACC+GEN+INDEC+N+NEUT+NOM+PLUR+SING+VOC,PSYCH**}

<contents>αναμφισβήτητα</contents>
{αναμφισβήτητα,**ADV+NO_COMP_SUP,**}
{αναμφισβήτητος,**ACC+ADJ+NEUT+NOM+NO_COMP_SUP+PLUR+VOC,**}

<contents>το</contents>
{αυτός,**ACC+NEUT+NOM+PRON+SING,PHOTOGR**}
{ο,**ACC+ART+MASC+SING,**}
{ο,**ACC+ART+NEUT+NOM+SING,**}

<contents>πιο</contents>
{πιο,**ADV,**}

Τυποποίηση λαθών - Σχεδιότυπα

- Η τυποποίηση των γραμματικών λαθών και η κατηγοριοποίησή τους έγινε με τη δημιουργία ειδικών σχεδιοτύπων (templates).
- Η συλλογή των γραμματικών λαθών μέσα από ένα αυθεντικό σώμα κειμένων ειδικής χρήσης (specialized corpus) αποτελεί σημαντικό πόρο του εκπαιδευτικού αυτού εργαλείου, ως εργαλείου ελέγχου και αξιολόγησης του γραμματικού διορθωτή corpus

Τυποποίηση λαθών - Σχεδιότυπα

- εντοπισμός προβληματικών τύπων εκτός της καθιερωμένης νόρμας
- διατήρηση προσωπικού ύφους – έκφρασης
- υποδείξεις «ορθής» γραφής
- επεξήγηση λάθους

Ιδιότητες Grammar Checker

- Ο γραμματικός διορθωτής υπερβαίνει τα υπάρχοντα εργαλεία και εξετάζει και ελέγχει «ορθογραφικά» κάθε λέξη ξεχωριστά, αλλά και σε σχέση με το άμεσο περιβάλλον της, π.χ. την πρόθεση με το ουσιαστικό, το άρθρο με το επίθετο και το ουσιαστικό, το ρήμα με τις αντωνυμίες και το συμπλήρωμά του κ.λπ.
- Ο συντακτικός έλεγχος επικεντρώνεται κυρίως στον εντοπισμό των τύπων που περιγράφονται ως προβληματικοί (μορφολογικά ή υφολογικά).
- Αν η λέξη ή η φράση έχει πράγματι τα χαρακτηριστικά που επιβάλλουν οι χρησιμοποιούμενοι κανόνες συντακτικού ελέγχου, τότε θεωρείται σωστή.
- Στην αντίθετη περίπτωση, είτε προτείνονται τα σωστά χαρακτηριστικά (π.χ. η ίδια λέξη σε άλλη πτώση: *πληγέντων περιοχών* το διορθώνει σε *πληγεισών περιοχών*) είτε προτείνεται μια άλλη λέξη με τα σωστά χαρακτηριστικά (π.χ. *πιο καλύτερος* θα διορθωθεί σε: *πιο καλός*).

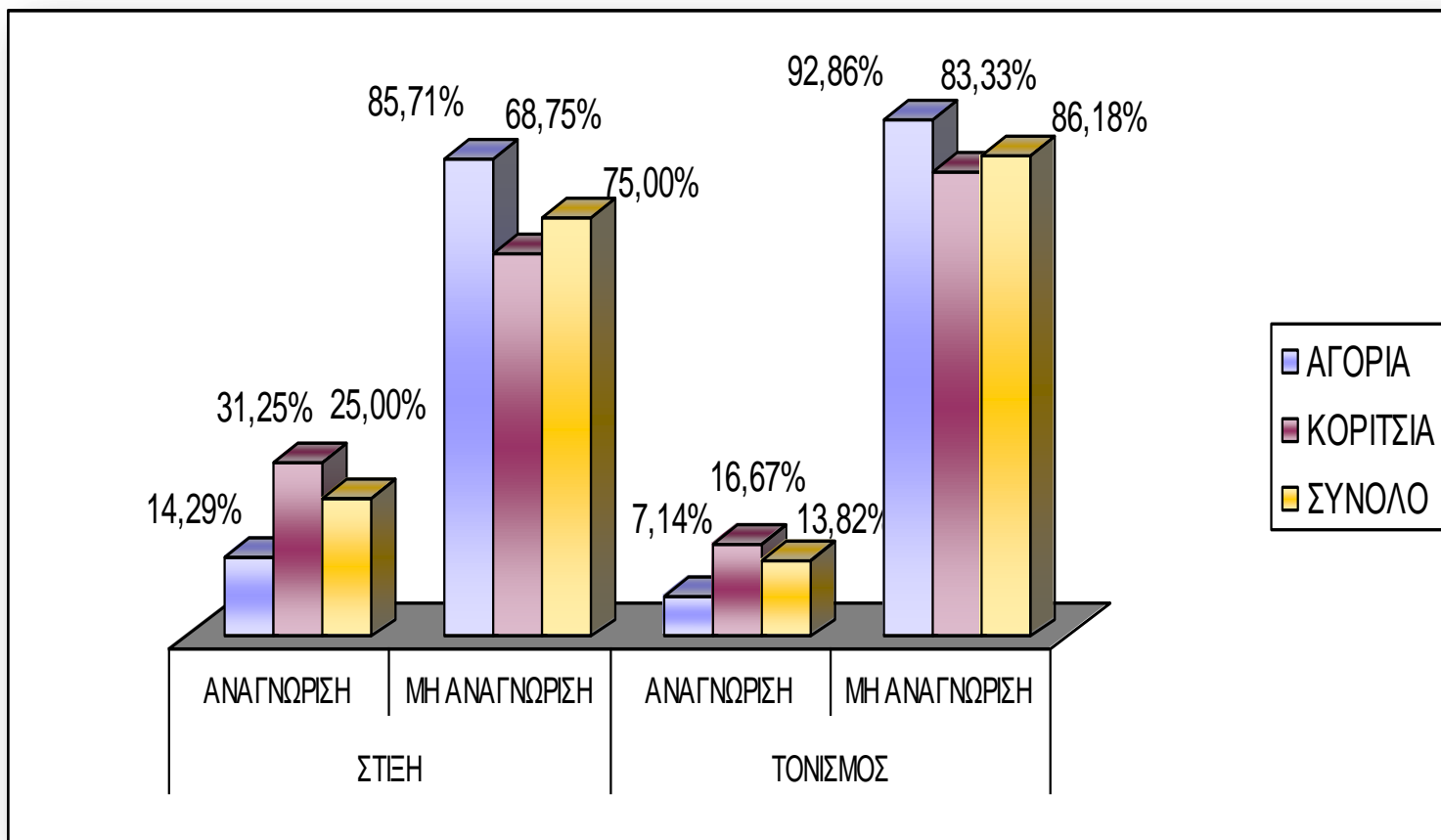
Ιδιότητες Grammar Checker

- Ο συντακτικός διορθωτής έχει σχεδιαστεί έτσι, ώστε να επισημαίνει και την υφολογική διαφοροποίηση των λεκτικών τύπων.
- Αυτό σημαίνει ότι ο γραμματικός διορθωτής βασίζεται στο ηλεκτρονικό μορφολογικό λεξικό στο οποίο η κλιτική αναπαραγωγή κάθε λήμματος επιτρέπει τη μορφολογική ποικιλία των τύπων (π.χ. «γυρνάνε», «γυρνάν», «γυρνούν», «γυρνούνε» ή «αγαναχτούν», «αγαναχτούνε» και «αγανακτούν», «αγανακτούνε»).
- Κατ' αυτό τον τρόπο ο χρήστης θα μπορεί να επιλέγει εκ των προτέρων το ύφος κειμένου που επιθυμεί και, κατά συνέπεια, να έχει και τον ανάλογο έλεγχο ύφους από τον γραμματικό διορθωτή

Ιδιότητες Grammar Checker

- Υπάρχει επιπλέον η δυνατότητα διευκόλυνσης του χρήστη, καθώς θα ενημερώνεται για τη σημασία μιας λέξεως η οποία είναι ομόηχη ή «προϊόν» εννοιολογικής σύγχυσης και, ως εκ τούτου, μπορεί να χρησιμοποιείται λανθασμένα σε συγκεκριμένο περιβάλλον

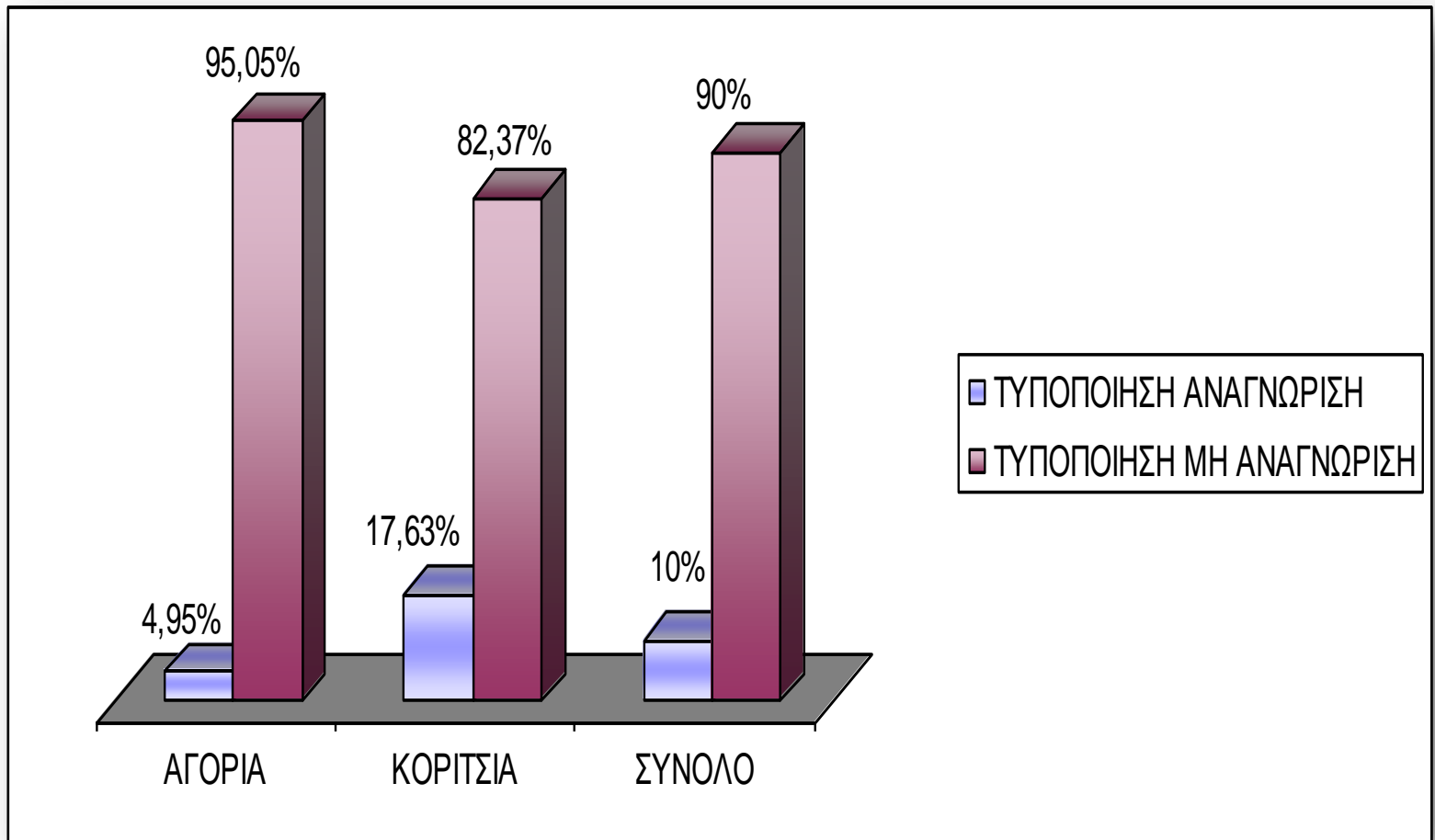
ΚΑΤΗΓΟΡΙΕΣ ΛΑΘΩΝ ΣΤΙΞΗ - ΤΟΝΙΣΜΟΣ



Στίξη - Τονισμός

- Η αναγνωρισιμότητα λάθους λόγω:
 1. της λανθασμένης εφαρμογής των κανόνων του μονοτονικού συστήματος κυμαίνεται σε ποσοστό 13,82%),
 2. της λανθασμένης χρήσης αποστρόφου κυμαίνεται σε ποσοστό 42,11%),
 3. της λανθασμένης χρήσης ενωτικού κυμαίνεται σε ποσοστό 21,05%) και
 4. εξαιτίας της λανθασμένης χρήσης προσωπικής ή κτητικής αντωνυμίας που απαιτεί παρουσία ή απουσία τόνου κυμαίνεται σε ποσοστό 15,79%).

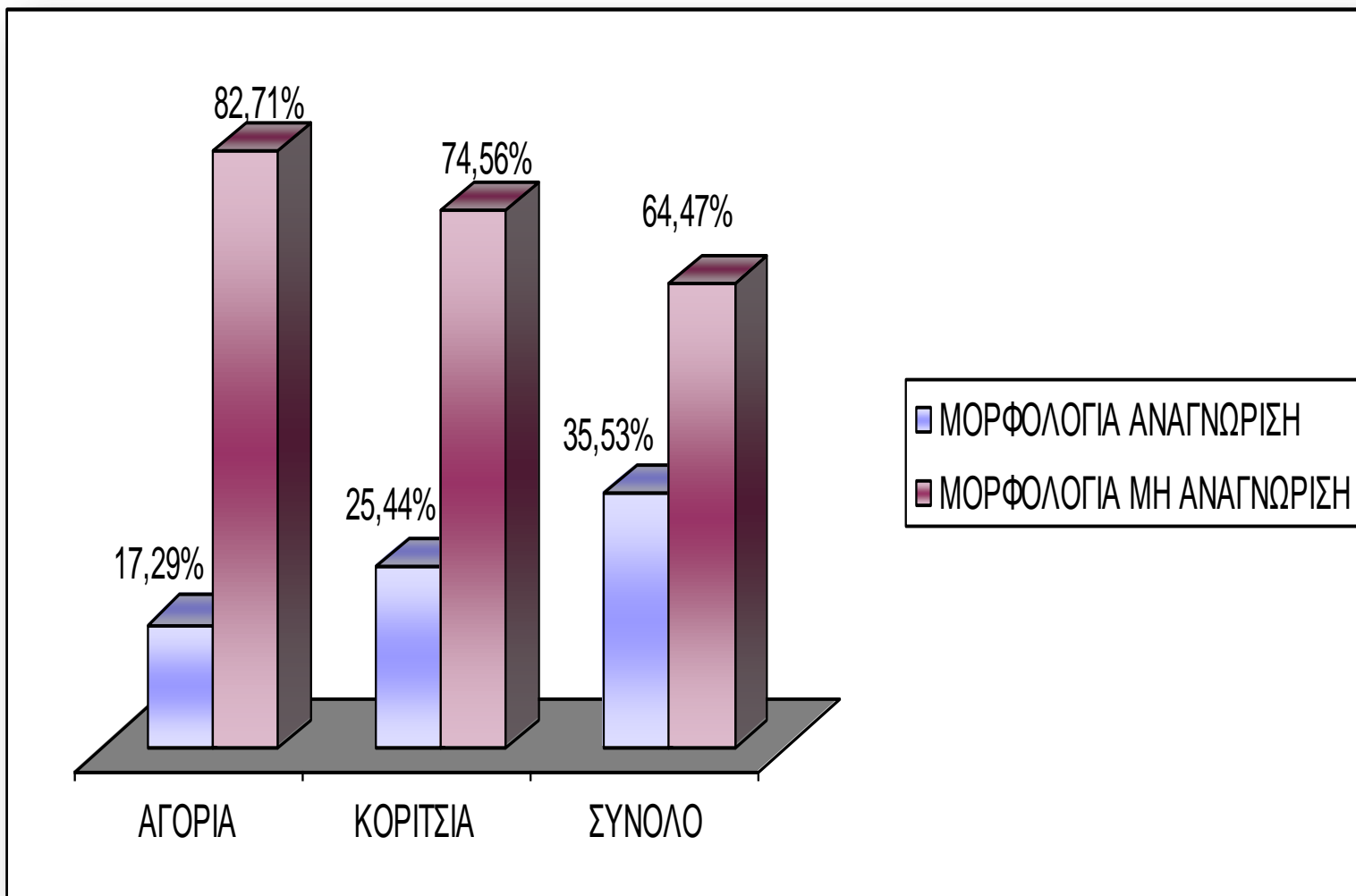
ΤΥΠΟΠΟΙΗΣΗ ΓΡΑΦΗΣ



Τυποποίηση γραφής

- α) σε επίπεδο γραφής προτείνεται η γραφή σε μία ή δύο λέξεις για τα ονοματικά σύνολα που προέρχονται από παλαιότερη περίοδο και έχουν ενσωματωθεί στον λόγο της νεοελληνικής,
- β) σε επίπεδο λεξιλογίου καταγράφονται, στο μεγαλύτερο ποσοστό, οι χρησιμοποιούμενες και στα νέα ελληνικά λόγιες εκφράσεις, με αντιστοίχιση του εσφαλμένου τύπου με τον ορθό,
- γ) σε επίπεδο ορθογραφίας προτείνεται η σύγχρονη γραφή κυρίως για λέξεις με μη ελληνική ετυμολογία και
- δ) σε επίπεδο φωνολογίας η λανθασμένη χρήση (παρουσία ή έλλειψη) του τελικού -v.

ΜΟΡΦΟΛΟΓΙΑ



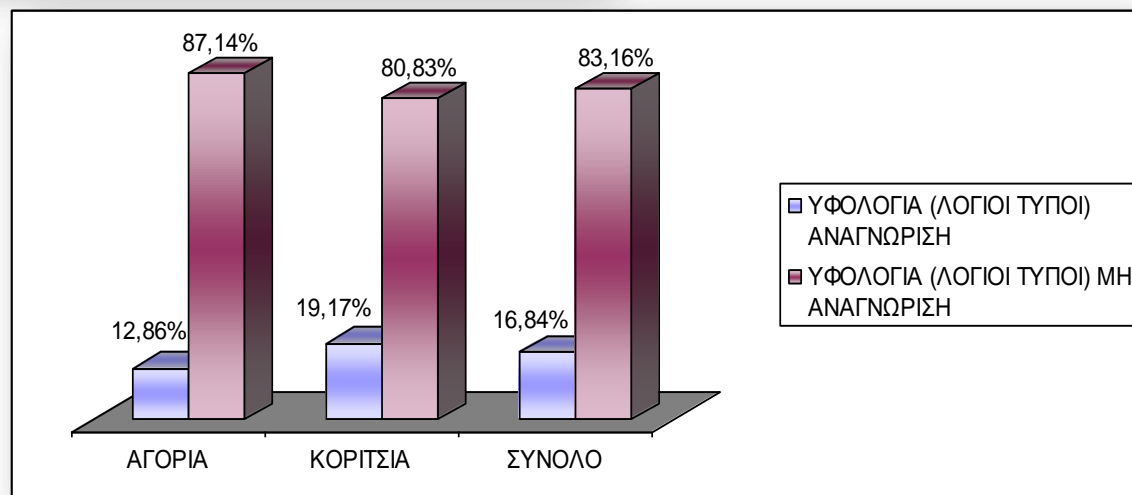
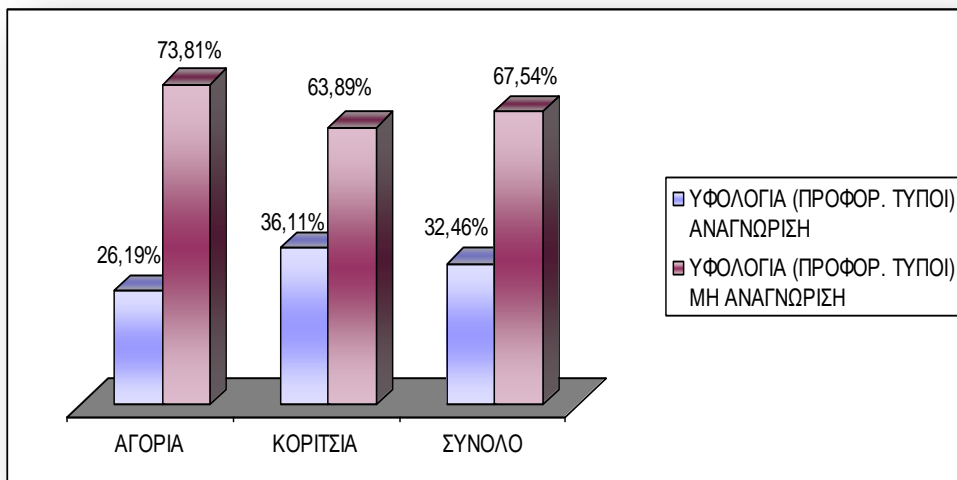
Μορφολογία

... ..

| ΚΛΙΣΗ ΟΥΣΙΑΣΤΙΚΩΝ | | ΚΛΙΣΗ ΡΗΜΑΤΩΝ | | ΚΛΙΣΗ ΕΠΙΘΕΤΩΝ | |
|---|--------|--|--------|----------------|-------|
| Κατηγορίες | % | Κατηγορίες | % | Κατηγορίες | % |
| Ουσιαστικά χωρίς πληθυντικό | 15,79% | Εσφαλμένος ενεστώτας ενεργητικής φωνής | 42,11% | ουδέτερο | 5,26% |
| Ουσιαστικά με σπάνιο πληθυντικό | 21,05% | Εσφαλμένος ενεστώτας παθητικής φωνής | 78,95% | | |
| Ουσιαστικά χωρίς γενική πληθυντικού | 5,26% | Εσφαλμένος αόριστος ενεργητικής φωνής | 78,95% | | |
| Εσφαλμένος πληθυντικός ελληνικών λέξεων | 73,68% | εσφαλμένος αόριστος παθητικής φωνής | 0% | | |
| Εσφαλμένος πληθυντικός ξένων λέξεων | 63,16% | εσφαλμένος παρατατικός ενεργητικής φωνής | 47,37% | | |

Πίνακας 8. Σχεδιάγραμμα μορφολογίας

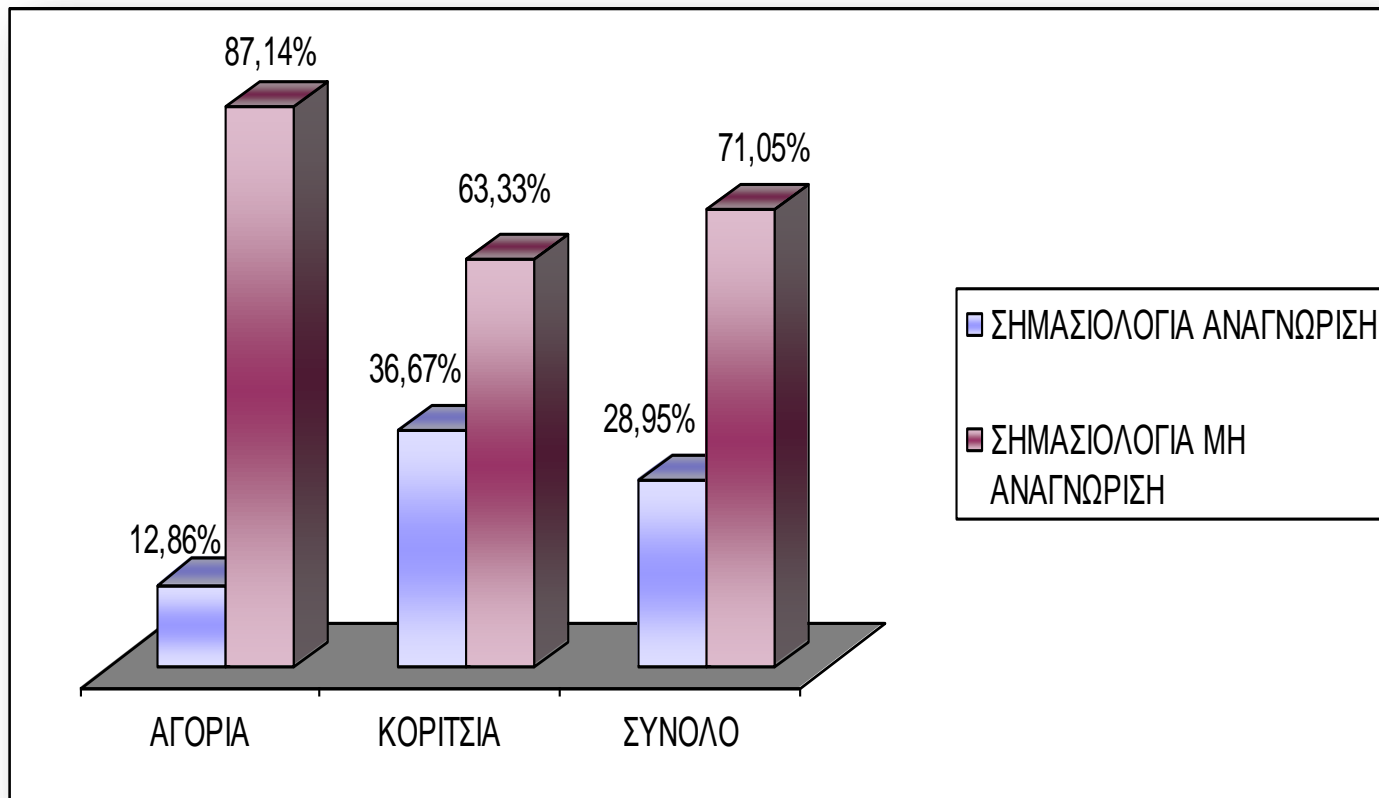
ΥΦΟΛΟΓΙΑ



Υφολογία

- Λήμματα (ουσιαστικά, ρήματα, επίθετα) με προφορικό υφολογικό χαρακτηριστικό (attribute),
- μεμονωμένοι προφορικοί τύπου ουσιαστικών,
- μεμονωμένοι προφορικοί τύποι ρημάτων,
- μεμονωμένοι προφορικοί τύποι επιθέτων,
- μεμονωμένοι προφορικοί τύποι επιρρημάτων,
- μορφολογικοί τύποι λημμάτων με ειδικούς χαρακτήρες (απόστροφος) που δεν είναι δόκιμοι στον γραπτό λόγο

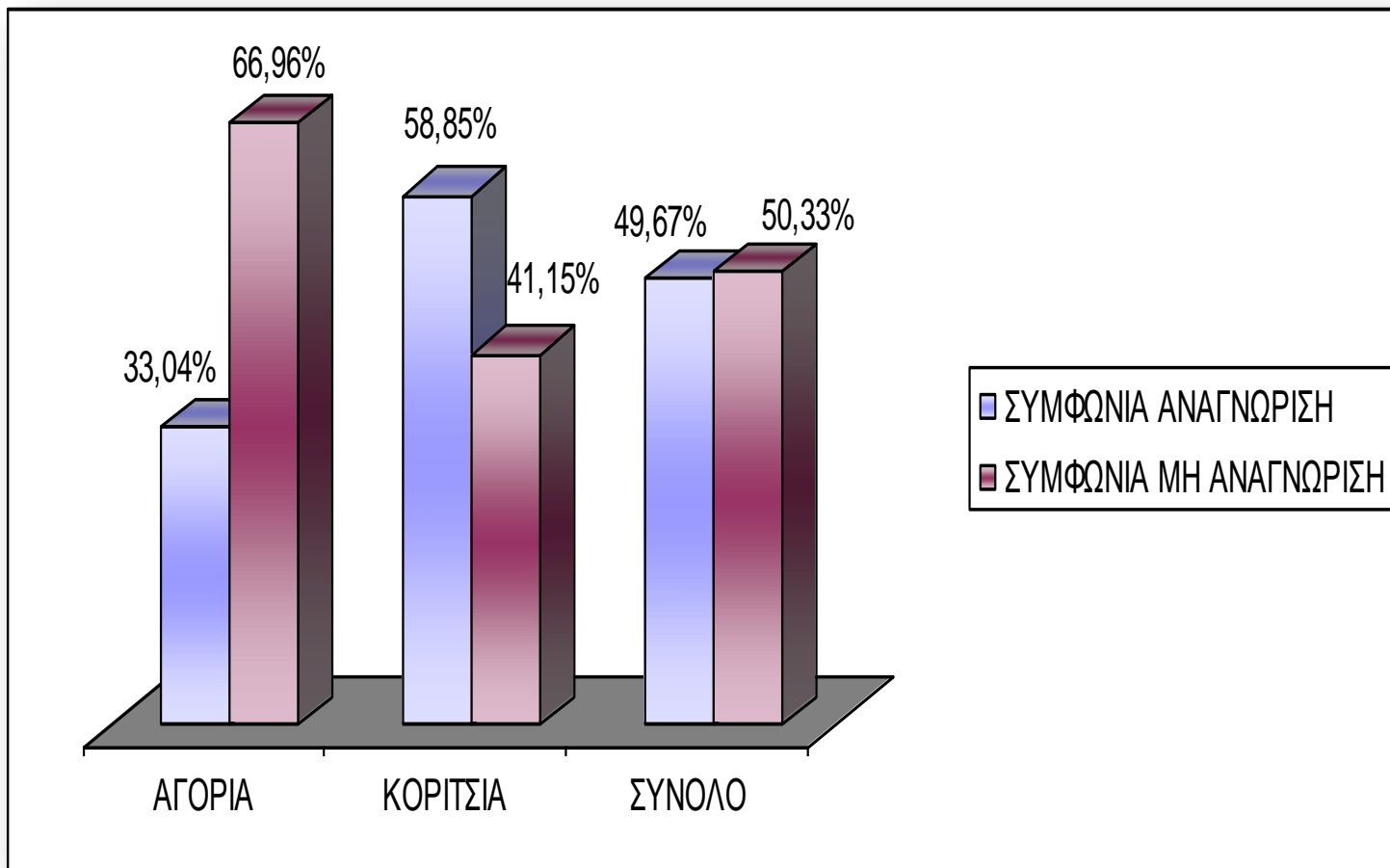
ΣΗΜΑΣΙΟΛΟΓΙΑ



Σημασιολογία

- Σκοπός είναι η διευκόλυνση του χρήστη με την κατανόηση της σημασίας μιας λέξεως η οποία είναι ομόηχη ή «προϊόν» εννοιολογικής σύγχυσης και ως εκ τούτου χρησιμοποιείται λανθασμένα.

ΣΥΜΦΩΝΙΑ

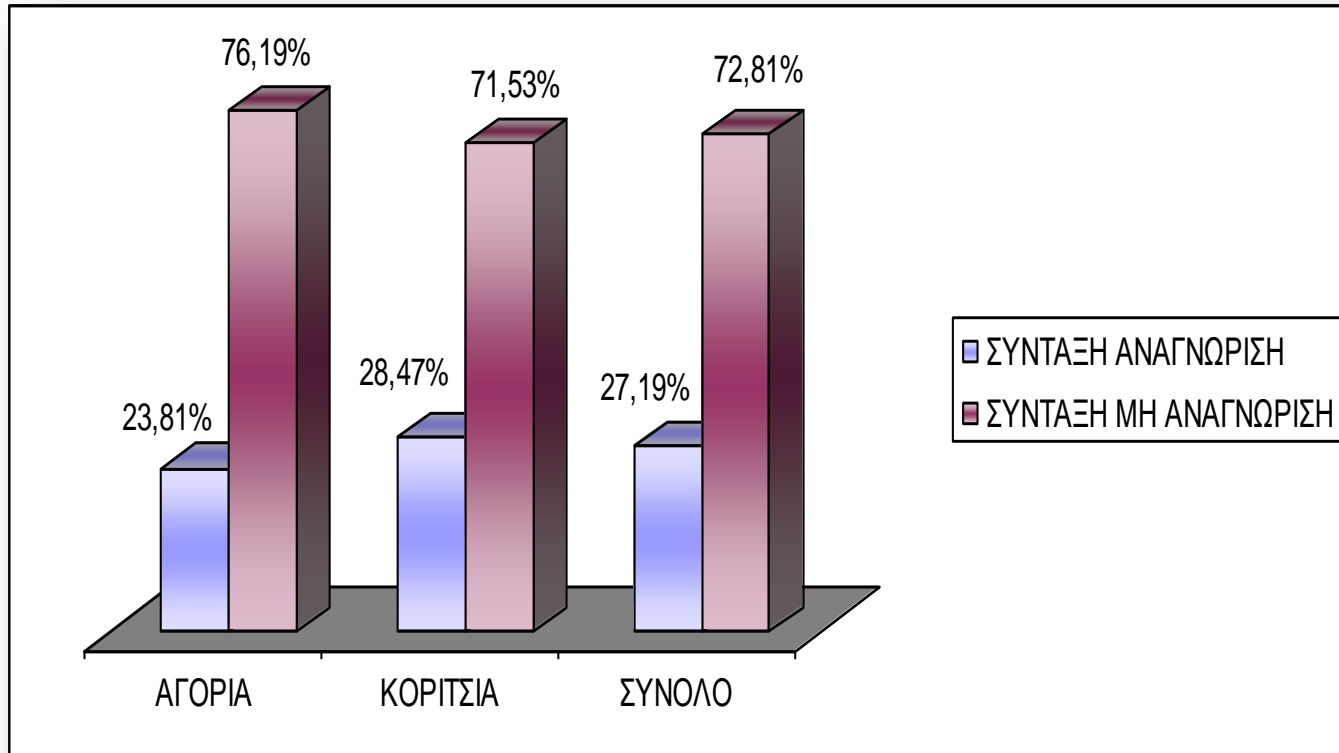


Συμφωνία

| ΣΥΜΦΩΝΙΑ ΣΕ ΡΗΜ. ΦΡΑΣΕΙΣ | | ΣΥΜΦΩΝΙΑ ΣΕ ΟΝΟΜ. ΦΡΑΣΕΙΣ | |
|-----------------------------------|--------|------------------------------------|--------|
| ΚΑΤΗΓΟΡΙΑ | % | ΚΑΤΗΓΟΡΙΑ | % |
| Σχηματισμός περιφραστικών χρόνων | 42,11% | Συμφωνία άρθρου – ουσιαστικού | 21,05% |
| Συμφωνία Υποκειμένου - Ρήματος | 73,68% | Συμφωνία ουσιαστικού – ουσιαστικού | 89,47% |
| Σχηματισμός αορίστου | 57,89% | Συμφωνία επιθέτου – ουσιαστικού | 57,89% |
| Αποσαφήνιση ουσιαστικού - ρήματος | 87,33% | Συμφωνία μετοχής – ουσιαστικού | 84,21% |
| Σχηματισμός προστακτικής | 68,42% | Ονοματικές οντότητες | 15,79% |

Πίνακας 9. Συμφωνία ΡΦ ή ΟΦ

ΣΥΝΤΑΞΗ



Σύνταξη

Στην κατηγορία αυτή περιγράφονται ζητήματα:

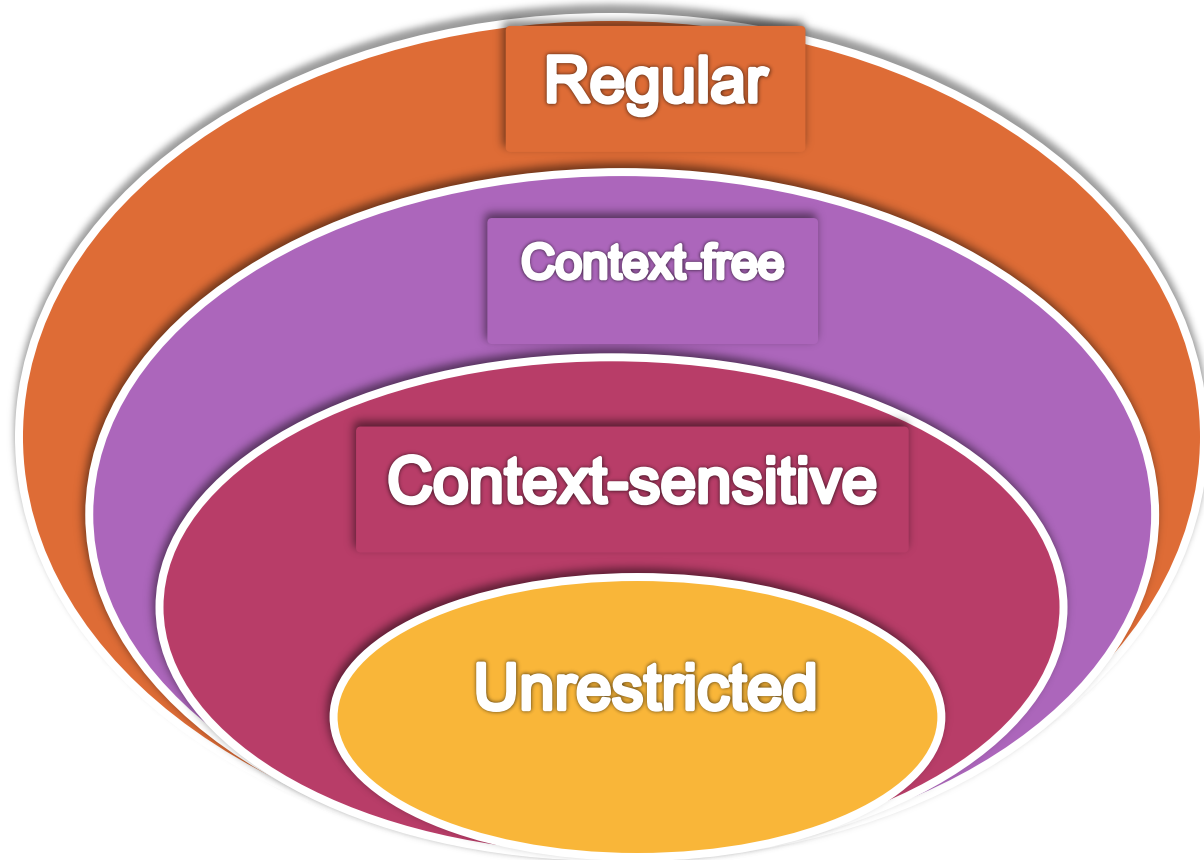
- α) λανθασμένης σύνταξης ρημάτων (με ονομαστική, με γενική, με αιτιατική, με εμπρόθετο, με πρόταση, αποθετικά ρήματα),
- β) λανθασμένης σύνταξης προθέσεων και
- γ) λανθασμένης σύνταξης επιρρημάτων

Grammar checker

- Βάση για την αντιμετώπιση όλων αυτών των ζητημάτων συνιστούν οι ρυθμίσεις της Γραμματικής (αναπροσαρμογή της Μικρής Νεοελληνικής Γραμματικής του Μανόλη Τριανταφυλλίδη), η οποία αποτελεί την επίσημη, κρατικά θεσμοποιημένη, από το 1976, γραμματική κωδικοποίηση της νεοελληνικής γλώσσας.
- έχουν ληφθεί υπόψη οι -ελάχιστες- διαφορές της νέας σχολικής γραμματικής Ε΄ και Στ΄ Δημοτικού)


ΥΛΟΠΟΙΗΣΗ ΕΦΑΡΜΟΓΗΣ

- Γραμματικές υλοποίησης (Chomsky)

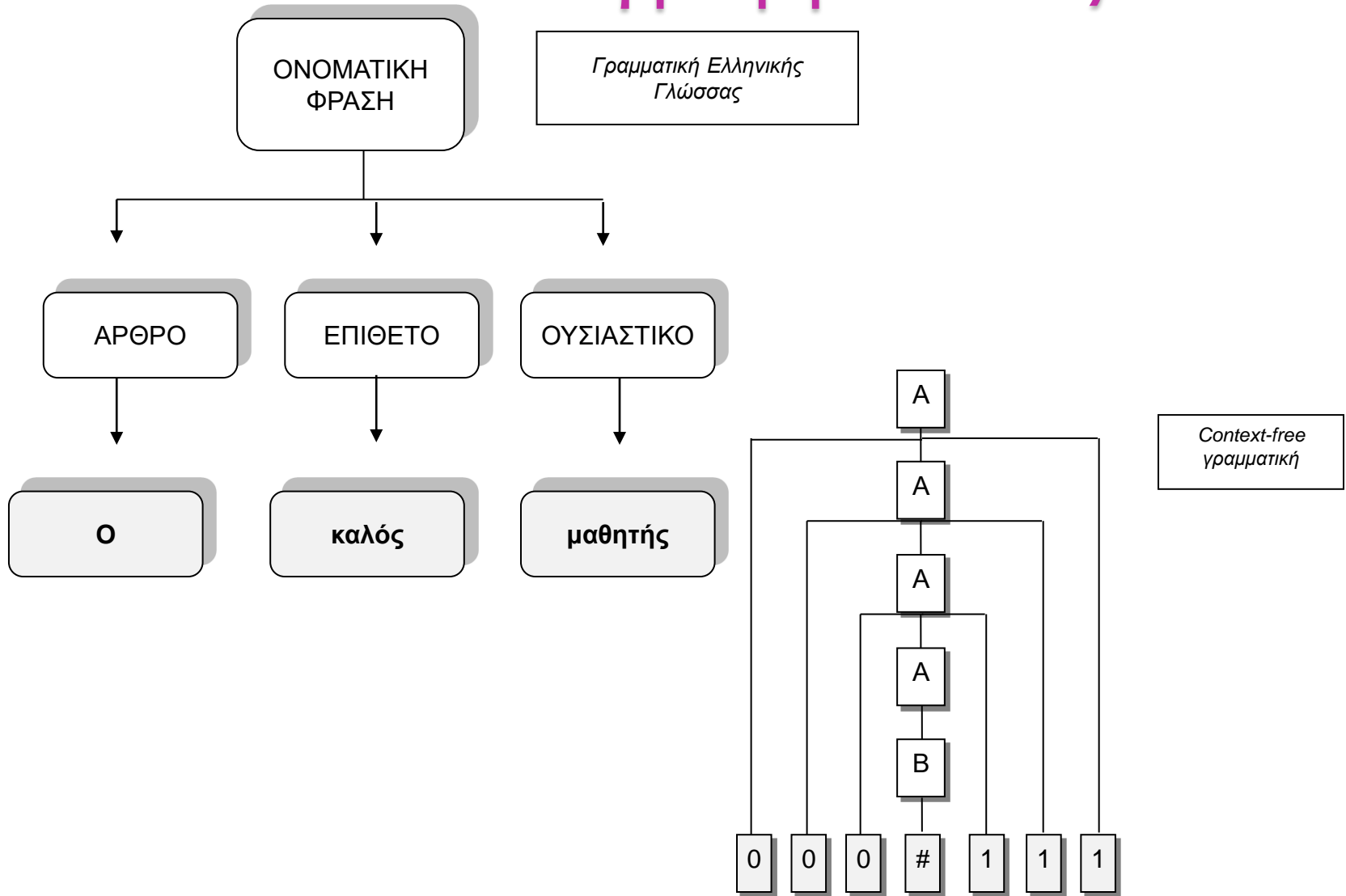


CHOMSKY

- Ο Chomsky (1956) πρώτος περιέγραψε τις γραμματικές αυτές το 1956 και τις ταξινόμησε σε τέσσερις τύπους, που είναι γνωστοί ως η ιεραρχία του Chomsky.
- Η διαφοροποίηση των γραμματικών αυτών έγκειται στο ότι διαδοχικά οι τύποι αυτοί έχουν κανόνες κλιμακούμενης αυστηρότητας και μπορούν, επομένως, να εκφράσουν τις λιγότερο τυπικές γλώσσες.
- Οι γραμματικές αυτές έχουν χαρακτηριστικά που θεωρήθηκαν από αρκετούς πρωτοπόρους της γνωστικής επιστήμης ως εγγενή στοιχεία της γλώσσας

- 
- Στη συντακτική ανάλυση ο γραμματικός διορθωτής αναλύει μικρές φραστικές ενότητες, μικρούς συνδυασμούς λέξεων.
 - Τέτοιες μικρές φραστικές ενότητες είναι το ονοματικό και το ρηματικό σύνολο, που στη συνέχεια χωρίζονται σε μικρότερες ενότητες μέχρις ότου φτάσουμε στις μεμονωμένες λέξεις.
 - Όλη αυτή τη συντακτική γνώση μπορούμε να την μοντελοποιήσουμε με ειδικού τύπου γραμματικές

Context-free γραμματικές



Συμφραστικές γραμματικές (*Context-sensitive grammar*)

- Οι κανόνες έχουν τη μορφή:

$$\alpha A \beta \rightarrow \alpha \gamma \beta$$

(α, β, γ) : ακολουθίες τερματικών και μη τερματικών συμβόλων.

Αυτό σημαίνει ότι, όποτε έχουμε το σύμβολο A με συμφραζόμενα α και β , τότε μπορούμε να το αντικαταστήσουμε με τη συμβολοσειρά τερματικών και μη τερματικών συμβόλων γ .

Context-sensitive grammar

- Η έννοια της συμφραστικής γραμματικής παρουσιάστηκε από τον Chomsky ως ένας τρόπος περιγραφής του συντακτικού των φυσικών γλωσσών στις οποίες η παρουσία ή όχι μιας λέξης σε συγκεκριμένο σημείο της πρότασης εξαρτάται από τις γειτονικές λέξεις, το γλωσσικό περιβάλλον (context).

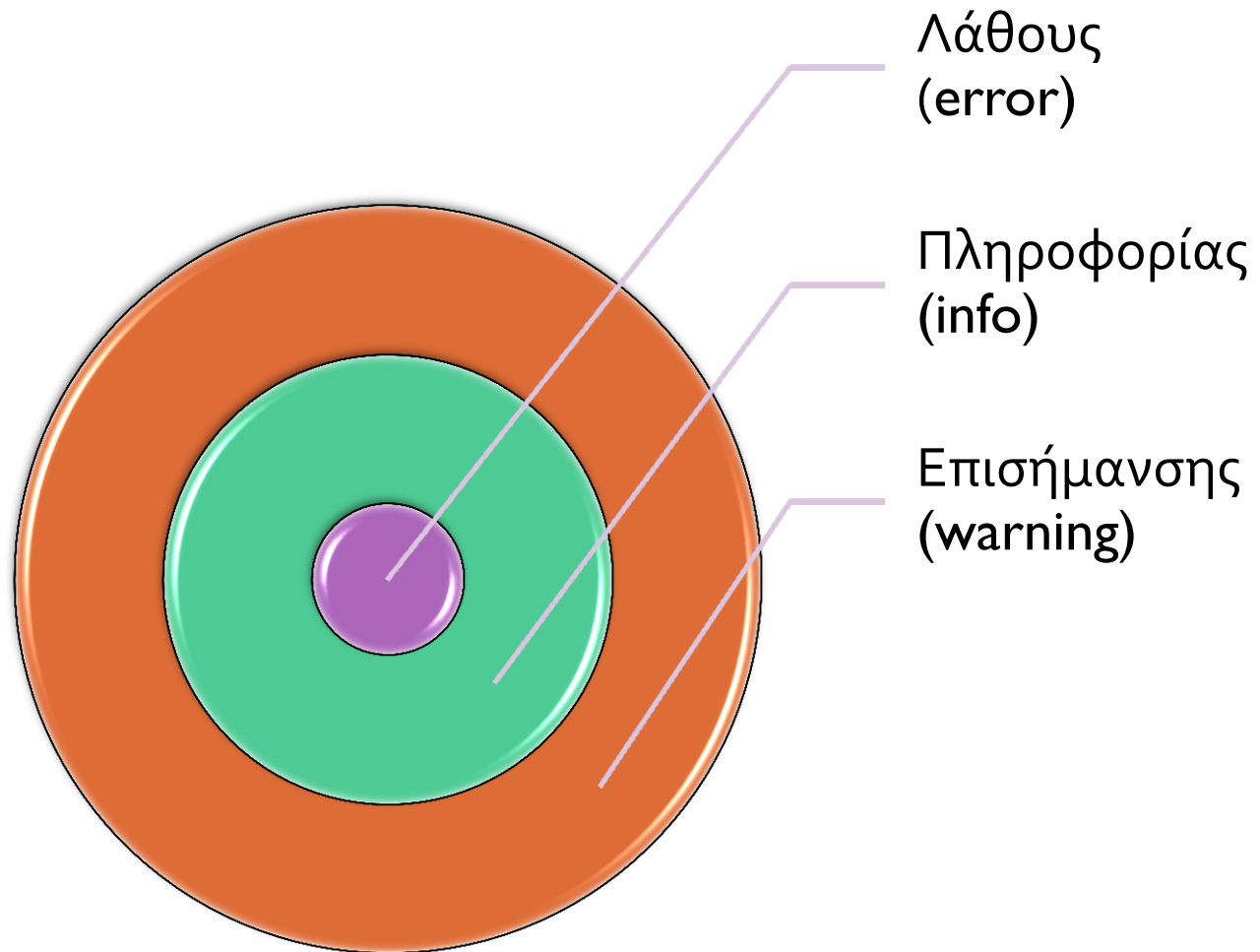
Context-sensitive grammar

- Αυτές οι γραμματικές είναι χρήσιμες σε εφαρμογές όπως η αυτόματη διόρθωση.
- Οι συμφραστικές γραμματικές χρησιμοποιούνται για την ανάλυση φυσικών γλωσσών
- **Η απόπειρα μηχανικής ανάλυσης με κανόνες συμφραστικής γραμματικής (context-sensitive grammar) θα έδινε καλύτερα αποτελέσματα αλλά θα είχε πολύ μεγάλο κόστος σε χρόνο**
- **Αυτό γιατί δύο προτάσεις μπορεί να έχουν κοινή εξωτερική δομή (επιφανειακή δομή) αλλά εντελώς διαφορετικό νόημα (βαθεία δομή), καθώς το σημασιολογικό και όχι μόνο περιεχόμενο μιας λέξης καθορίζει και την ανεύρεση της σημασίας στο επίπεδο της «βαθείας» δομής**

Mnemosyne

- σύστημα επεξεργασίας φυσικής γλώσσας για ανάκτηση δομημένης πληροφορίας
- ενσωματώνει δεδομένα γλωσσολογικής πληροφορίας
- έχει χρησιμοποιηθεί ήδη σε περιβάλλοντα με μεγάλης ποσότητας στοιχεία

Γραμματικός έλεγχος (κανόνες)



Γραμματικός έλεγχος (κανόνες)

- Στην κατηγορία του λάθους (error) περιγράφονται τα λάθη που αποκλίνουν από τη γλωσσική νόρμα,
- στην κατηγορία της πληροφορίας (info) περιγράφονται κυρίως κανόνες που αφορούν στο ύφος και
- στην κατηγορία επισήμανση (warning) οι κανόνες που αφορούν στη σημασιολογική πληροφόρηση του χρήστη για λέξεις με εννοιολογική σύγχυση.
- Ο χρήστης μπορεί να επιλέξει μέρος ή το σύνολο των πιο πάνω ομάδων (χρωματίζονται διαφορετικά) και να λάβει την πληροφορία που επιθυμεί.

Γραμματικός έλεγχος (κανόνες)

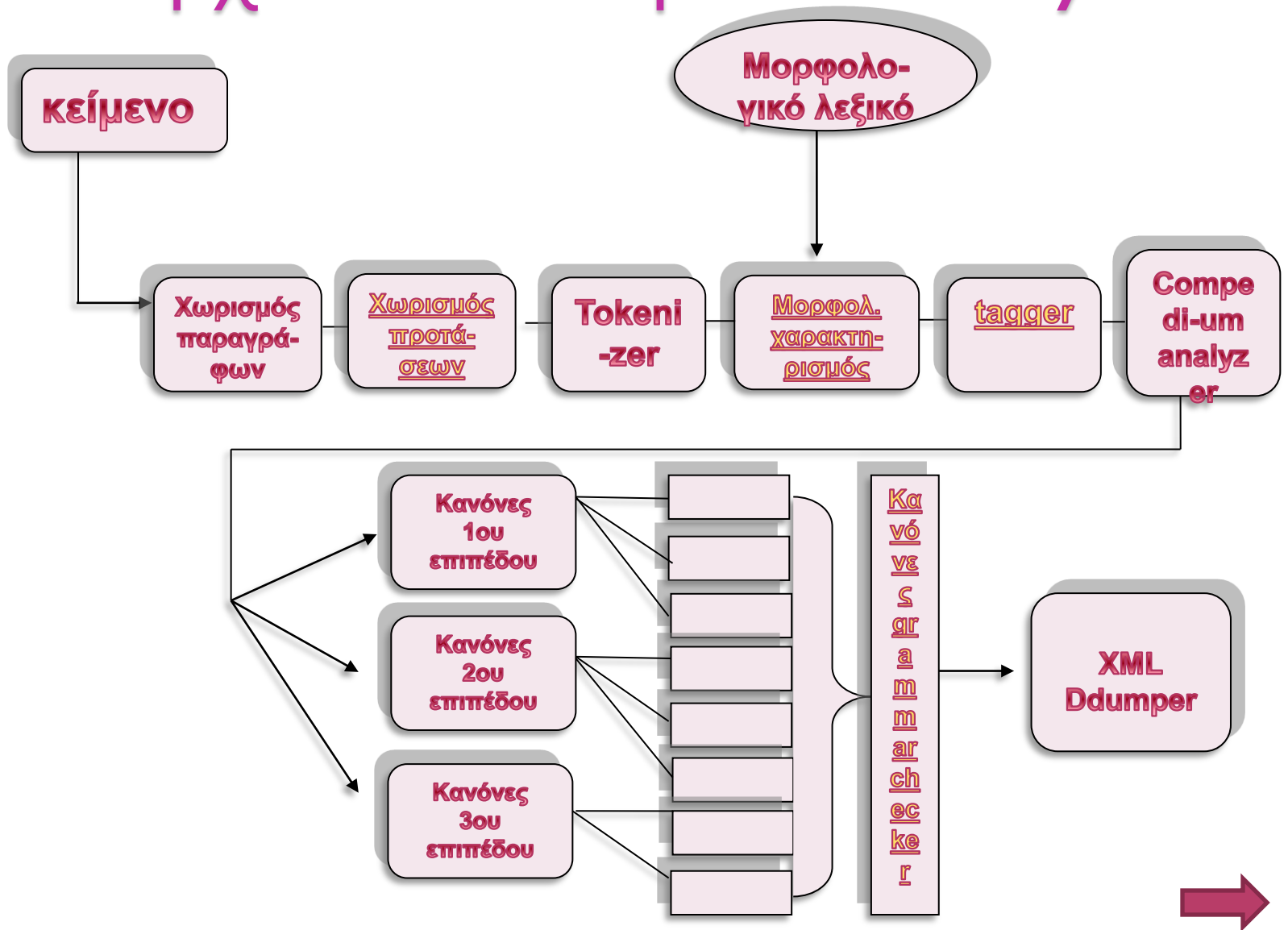
Select: ΓΡΑΜΜΑΤΙΚΟΣ ΕΛΕΓΧΟΣ

Πάει το «φέσι»... σύννεφο στην αγορά της **Πάτρας καθώς λιγιστοί συμπολίτες μπορούν να καλύψουν** με τα πενιχρά τους έσοδα τις απαιτήσεις της καθημερινότητας, ενώ ακόμα χαρότερα είναι τα πράγματα για μεγάλη μερίδα των επιχειρηματιών της πόλης. Έτσι οι οφθαλμοί των πάντων κατά πάντων **πάνε κι** έρχονται με **ότι** και να **συνεπάγεται** αυτό για το βιοτικό επίπεδο των πολιτών αλλά και **την** βιωσιμότητα των επιχειρήσεων που δεν έχουν **ρευσιμότητα για να** πληρώσουν τις υποχρεώσεις τους. Και που **δεν** λειτουργεί και πάλι το παλιό γνωστό «τεφτέρι»: Γονείς αδυνατούν να πληρώσουν τα φροντιστήρια των παιδιών τους, νοικοκυρές ζητούν από τους ιδιοκτήτες μίνι μάρκετ να τα... «γράφουν», επιχειρηματίες χρωστούν σε αποθήκες και προμηθευτές, και πάει λέγοντας. Τα μετρητά είναι είδος υπό εξαφάνιση στις συνοικίες της πόλης, οι ξένοι όμως από όπου εισάγουν προϊόντα οι Πατρινοί καταστηματάρχες δεν... παίζουν. Ζητούν να πληρώνονται με μετρητά κάθε παραγγελία που τους γίνεται από την Ελλάδα. Το γεγονός αυτό φέρνει σε μεγάλο αδιέξοδο τους επιχειρηματίες της πόλης οι οποίοι βλέπουν το «τεφτέρι» τους να είναι γεμάτο με οφθαλμούς αλλά τα ταμεία του άδεια, κάθε φορά που θέλουν να κάνουν παραγγελίες... Έτσι **λοιπόν** κάνουν λίγες προμήθειες (**αφού** οι Τράπεζες έχουν κόψει τα **δανεικά**) με μεγαλύτερο κόστος με **αποτέλεσμα** να μην μπορούν να είναι ανταγωνιστικοί απέναντι στις πολυεθνικές.

- (ΠΛΗΡΟΦΟΡΙΑ)
- (ΛΑΘΟΣ)
- (ΕΠΙΣΗΜΑΝΣΗ)

Ομαδοποίηση κανόνων

«Αρχιτεκτονική» Mnemosyne



Χωρισμός προτάσεων

```
<analysisname="sentence">
```

```
<spanlength="23" offset="0">
```

```
<contents>ο κ. Νίκος ήρθε  
σήμερα.</contents>
```

```
<annotations>
```

```
<tagname="FPARSEN" class="class  
java.lang.String">true</tag>
```

```
<tagname="SSEQNO" class="class  
java.lang.String">0</tag>
```

```
</annotations></span>
```



Tokenizer

- Κάθε συστατικό της πρότασης χαρακτηρίζεται (οι λέξεις, τα σύμβολα, οι παρενθέσεις, οι απόστροφες, τα εισαγωγικά κ.λπ.).
- Σε πρώτο επίπεδο καταγράφεται η ορθογραφία κάθε συστατικού της πρότασης (Ελληνικά γράμματα, αριθμοί, πρώτο γράμμα κεφαλαίο, κ.ά.).
- Σε δεύτερο επίπεδο λέξεις με ιδιαίτερη ορθογραφία μπορούν περαιτέρω να χωριστούν σε 2 ή τρεις λέξεις. Για παράδειγμα μια λέξη που τελειώνει σε κόμμα ή θαυμαστικό ή ερωτηματικό χωρίζεται σε δύο λέξεις.
- Ίδια αντιμετώπιση γίνεται και όταν η λέξη αρχίζει με εισαγωγικά ή/και τελειώνει σε εισαγωγικά οπότε και χωρίζεται σε δύο ή τρεις λέξεις

Μορφολογικός χαρακτηρισμός

<spanlength="3" offset="3">

<contents>του</contents>

<annotations>

<tagname="LEXY" class="classjava.lang.String">{αυτός,GEN+MASC+PRON+SING,}{αυτός,GEN+NEUT+PRON+SING,}{ο,ART+GEN+MASC+NEUT+SING,}</tag>

<tagname="TTEXT" class="class java.lang.String">του</tag>

<tagname="ORTHO" class="classjava.lang.String">NrWrd+WthLtrs+FconWrd+StyleI </tag>

</annotations>

<spanlength="10" offset="7">

<contents>απαντήσεις</contents>

<annotations>

<tagname="LEXY" class="classjava.lang.String">{απάντηση,ACC+FEM+N+NOM+PLUR+VOC,}{απαντάω,ACT+AOR+B_P+SING+SUBJ+V,}</tag>

<tagname="TTEXT" class="class java.lang.String">απαντήσεις</tag>

<tagname="ORTHO" class="classjava.lang.String">NrWrd+WthLtrs+FvolWrd+StyleI </tag>

</annotations>



Άρση ασάφειας

Οι υπολογιστικές μέθοδοι που έχουν αναπτυχθεί με στόχο τη μορφοσυντακτική αποσαφήνιση χωρίζονται γενικά σε δύο κατηγορίες:

- α) Σύμφωνα με τη γλωσσολογική προσέγγιση, οι ειδικοί κωδικοποιούν χειρωνακτικά κανόνες βασισμένους σε γενικεύσεις παραδειγμάτων αποσαφήνισης, τα οποία συνήθως συλλέγονται από σώμα κειμένων μορφοσυντακτικά χαρακτηρισμένων (Ορφανός et al., 1999).
- β) Σύμφωνα με την προσέγγιση της μηχανικής εκμάθησης, ένα στατιστικό μοντέλο για την επίλυση του γλωσσικού προβλήματος επάγεται αυτόματα από σώμα χαρακτηρισμένων κειμένων.

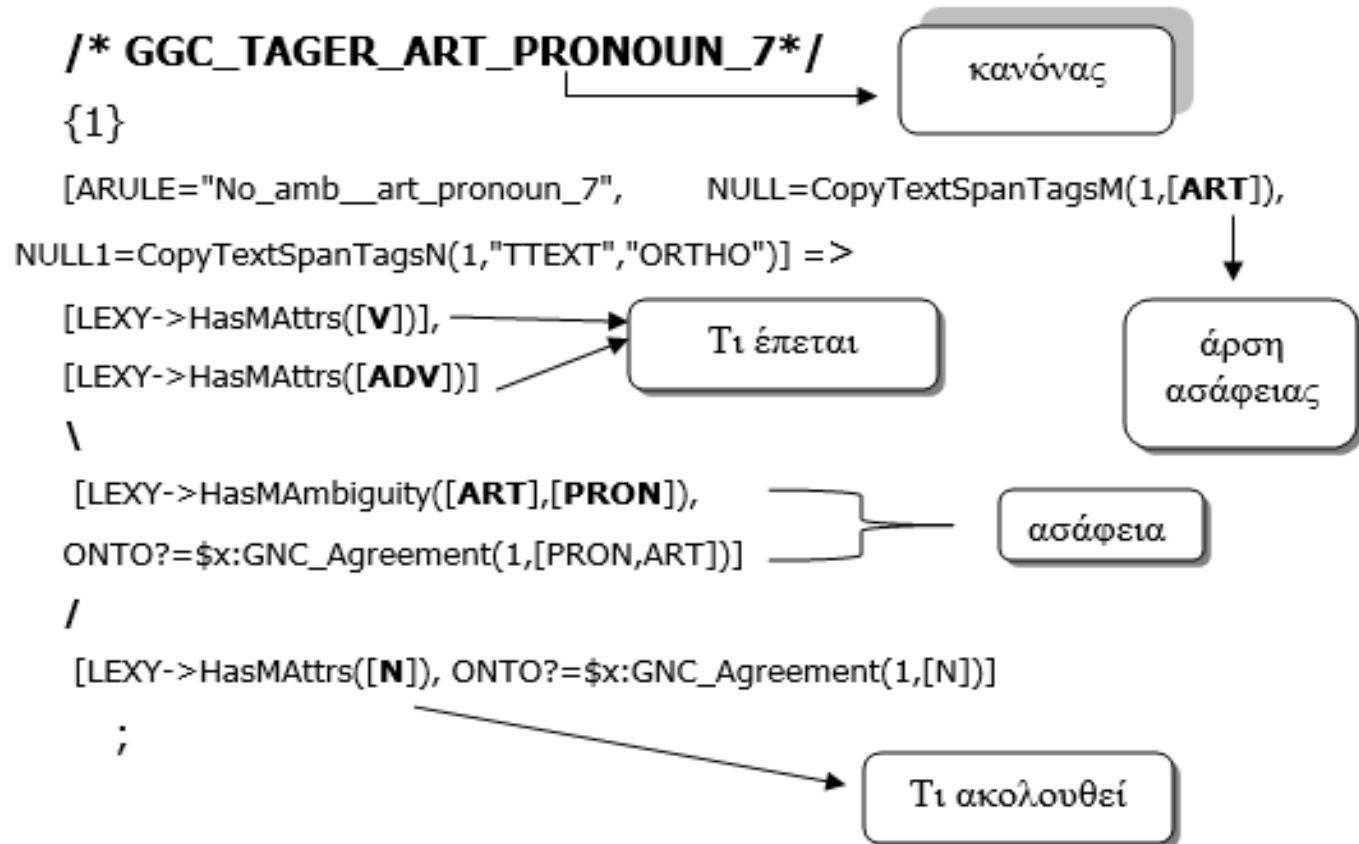
Tagger Mnemosyne

- προσανατολισμένος στην άρση της λεξικής ασάφειας στα νέα ελληνικά
- βασισμένος στο ανάλογο γλωσσικό περιβάλλον των λέξεων
- άρση της λεξικής ασάφειας μόνο με γλωσσολογική πληροφορία
- 100 κανόνες (rules)
- άρση της ασάφειας ως προς το ΜτΛ αλλά και ως προς το γένος και την πτώση της ασαφούς λέξης

Tagger Mnemosyne

- Ο tagger επεκτείνεται και χαρακτηρίζει λέξεις που δεν υπάρχουν στο μορφολογικό λεξικό (άτονες λέξεις, λέξεις με ανορθογραφία) η γνώση των μορφολογικών των οποίων είναι απαραίτητη στο επίπεδο της συντακτικής ανάλυσης
- Ο tagger του mnemosyne δε στοχεύει μόνο στην άρση της ασάφειας ως προς το ΜτΛ αλλά και ως προς το γένος και την πτώση της ασαφούς λέξης
- Για την άρση της λεξικής ασάφειας εξετάζονται τόσο οι προηγούμενες λέξεις (σε αριθμό μέχρι και 4 token) όσο και/ή οι επόμενες (σε αριθμό μέχρι και 4 token). Η μορφή που μπορεί να έχει ένας κανόνας είναι η ακόλουθη

Άρση ασάφειας



Άρση ασάφειας

- ο tagger δεν αίρει την ασάφεια με μοντέλα γενικευμένης ισχύος (π.χ. όταν η ασαφής λέξη ακολουθεί άρθρο είναι ουσιαστικό) αλλά επιτρέπει – μέσω των επιπέδων να αντιμετωπίσει κα πιο ειδικές, σπάνιες περιπτώσεις που όμως επιτρέπουν την καλύτερη διαχείριση των συντακτικών κανόνων.
- Έτσι, στη φράση: *[τις μυστικές τους δυνάμεις]* η ακολουθία των ΜτΛ έχει ως εξής: **[άρθρο], [επίθετο], [ασαφής λέξη], [ουσιαστικό]**. Ομοίως και στη φράση: *[την καλή την κοπέλα]*, η ακολουθία των ΜτΛ έχει την ίδια μορφή: **[άρθρο], [επίθετο], [ασαφής λέξη], [ουσιαστικό]**. Στην πρώτη περίπτωση όμως η ασαφής λέξη είναι αντωνυμία ενώ στη δεύτερη άρθρο.
- Μια γενική και καθολική ισχύς του κανόνα θα δημιουργούσε προβλήματα

Άρση ασάφειας

```
/* GGC_TAGER_PRONOUN_ART_13*/
```

```
{1} → 1° επίπεδο
```

κανόνας

```
[ARULE="No_amb_pronoun_art_13", NULL=CopyTextSpanTagsM(1,[PRON]),  
NULL1=CopyTextSpanTagsN(1,"TTEXT","ORTHO")] =>
```

```
(  
  [LEXY->HasMAttrs([ADJ]), ONTO?=$x:GNC_Agreement(1,[ADJ])] |  
  [LEXY->HasMAttrs([PRON]), ONTO?=$x:GNC_Agreement(1,[PRON])]  
)
```

άρση
ασάφειας

Τι προηγείται

```
[LEXY->HasMAmbiguity([ART],[PRON])]
```

ασάφεια

```
[LEXY->HasMAttrs([N]), ONTO?=$x:GNC_Agreement(1,[N])]
```

συμφωνία

```
;
```

Τι ακολουθεί

Κανόνες mnemosyne

Section

```
/* GGC_TELIKO_N_remove_n_3*/  
{3}
```

Κεφαλή
του
κανόνα

```
[ARULE="GGC_TELIKO_N_remove_n_3", TTEXT=$x_I_3,  
EVTEXT=TagEvent("gevent.wrong","GEVENT","FINAL_N","%f",  
"ERMSG", "Το 'τον' δεν πρέπει να έχει τελικό -v.  
Αντικαταστήστε το 'τον' με το 'το'.")]=>
```

```
\  
[TTEXT->$x_I_3:Match("τον"),  
LEXY->CanMatch("o",[ART,ACC,MASC,SING])]
```

Αντικείμενο
έρευνας

```
/  
[LEXY->HasMAttrs([MASC,SING,ACC]),  
ORTHO->AnyOfOAttrs([Style2])]  
;  
end
```

Κανόνες

Κάθε συντακτικός κανόνας έχει πέντε στοιχεία:

α) την κεφαλή του κανόνα αριστερά του συμβόλου "=>", για παράδειγμα [VTEXT="μήνας"] ή [VTEXT="ρήμα"]. Αυτό καθορίζει την αντικατάσταση της αναγνωρισμένης έκφρασης από ένα ή περισσότερα εικονικά σημεία ή κανόνες. Για παράδειγμα στη φράση: [αυστηρώς ακατάλληλον] αντικαθίσταται από το VTEXT="__ancient_phrase__". Η δήλωση της πιο πάνω φράσης έχει την ακόλουθη μορφή:

Section

```
/* GGC_pattern_ancient_1*/
```

```
{1}
```

```
[ARULE="pattern_ancient_1",VTEXT="__ancient_phrase__"] =>
```

```
\
```

```
(
```

```
[TTEXT->Match("αυστηρώς")],
```

```
[TTEXT->Match("ακατάλληλον")]
```

```
)
```

```
/
```

```
;
```

```
end
```


Κανόνες

- β) τη αντίστοιχη φράση ή λέξη ή λήμμα που είναι αντικείμενο έρευνας (εν προκειμένω λανθασμένος τύπος) που υπάρχει μεταξύ των συμβόλων 'V' και 'V'
- γ) το αριστερό μέρος της έκφρασης πριν το σύμβολο 'V' (στα πιο πάνω παραδείγματα είναι άδειο) και αποτελεί ένα σύνολο λέξεων, φράσεων, γενικότερα tokens που είναι χρήσιμο για τον προσδιορισμό της έκφρασης αλλά δεν αντικαθίσταται από την κεφαλή του κανόνα.

Κανόνες

- δ) ομοίως το δεξί μέρος της έκφρασης που ακολουθεί το σύμβολο 'V'.
- ε) το είδος των κανόνων που περικλείεται εντός των συμβόλων «και» πριν από το αριστερό πλαίσιο. Αυτό το μέρος δεν είναι απαραίτητο, υπάρχει όμως σε κάθε κανόνα και χρησιμοποιείται για να κατηγοριοποιήσει τα λάθη.
- Στους κανόνες που ορίζονται στον γραμματικό διορθωτή υπάρχουν τρεις ευρείες κατηγορίες κανόνων:
 1. Κανόνες λάθους (**event.wrong**),
 2. κανόνες προειδοποίησης (**gevent.warning**) και
 3. κανόνες πληροφορίας (**gevent.info**).

functions:

| | |
|--------------------------|---------------------------------------|
| HasLemma3 | is method of LEXY (text); |
| HasNotLemma | is method of LEXY (text); |
| HasMAttrs | is method of LEXY (MATTRS); |
| HasNoneMAttrs | is method of LEXY (MATTRS); |
| IsNotOneOfMAttrs2 | is method of LEXY (MATTRS, MATTRS); |
| CanMatch | is method of LEXY (text, MATTRS); |
| IsEqual | is method of ORTHO (OATTRS); |
| AnyOfOAttrs | is method of ORTHO (OATTRS); |
| NoneOfOAttrs | is method of ORTHO (OATTRS); |
| Match | is method of TTEXT (text); |
| Prefix | is method of TTEXT (text); |
| Suffix | is method of TTEXT (text); |
| RLookingAt | is method of TTEXT (text); |
| SuffixCanMatch | is method of LEXY (text, MATTRS); |
| SuffixLemma | is method of LEXY (text, text); |

Κανόνες

- Ο όρος *ORTHO->AnyOfOAttrs([NrWrd])* σημαίνει ότι πρέπει η λέξη να είναι normal Word (λέξη με γράμματα του αλφαβήτου).
- Ο όρος *[LEXY->HasMAttrs ([ADJ,... ACC])* σημαίνει ένα οποιοδήποτε λήμμα που έχει συγκεκριμένα μορφολογικά χαρακτηριστικά (εν προκειμένω επίθετο, ... αιτιατική). Αντίστοιχα μπορεί να περιλαμβάνει και τον όρο *[LEXY->HasNoMAttrs([ADJ,... ACC])* που σημαίνει το λήμμα που δεν πρέπει να έχει τον πιο πάνω χαρακτηρισμό έναν όρο.
- Ο όρος *[TTEXT->Match(«γκαρσόν»)]* ορίζει λέξεις που έχουν τη συγκεκριμένη ακολουθία γραμμάτων (κεφαλαία (ΓΚΑΡΣΟΝ) ή/και πεζά (γκαρσόν/Γκαρσόν), τονισμένα ή άτονα [γκαρσόν]) ενώ στην πιο «αυστηρή» του μορφή ορίζεται *[TTEXT == «επιστήμων»]* που περιλαμβάνει μόνο τον συγκεκριμένο τύπο [επιστήμων].

Κανόνες

- Στους κανόνες αυτούς μπορεί να οριστεί το πρόθημα ($[TTEXT \rightarrow \mathbf{Prefix}$ («εξεφώνησ»)] το οποίο περιλαμβάνει είναι όλες τις παραγόμενες λέξεις που έχουν το πρόθημα «εξεφώνησ». Αναλόγως ορίζεται και η κατάληξη μιας οποιασδήποτε λέξης $[TTEXT \rightarrow \mathbf{Suffix}$ («ουνα») που είναι όλοι οι κλιτικοί τύποι με την κατάληξη [-ουνα].
- Στον κανόνα $[TTEXT \rightarrow \mathbf{Suffix}$ («ουνα»), $LEXY \rightarrow \mathbf{HasMAttrs}$ ($[V, A_P, SING]$)] καλούνται λέξεις με κατάληξη -ουνα και μορφολογικά χαρακτηριστικά [Ρήμα και Α΄ πρόσωπο, ενικός].

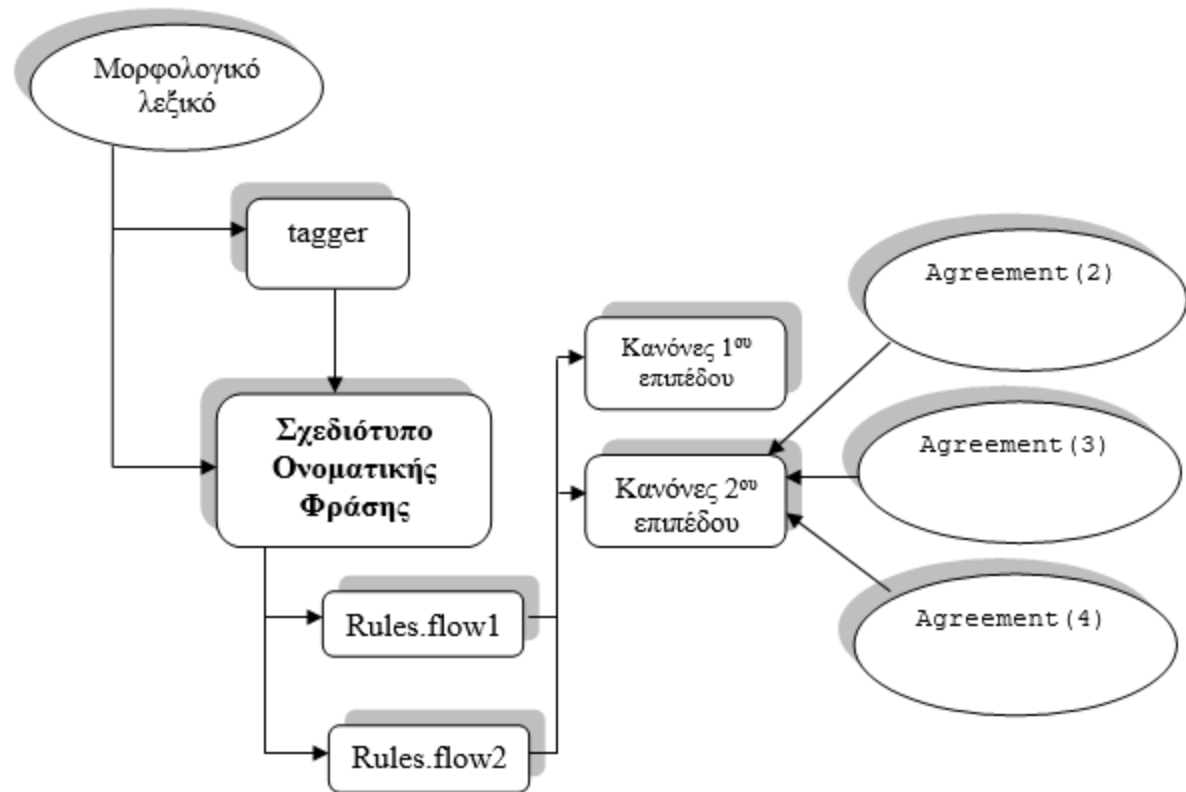
Κανόνες

- Ο όρος $[LEXY \rightarrow HasLemma]$ («γράφω») καλεί το συγκεκριμένο λήμμα (σε όλους τους χρόνους, εγκλίσεις, πρόσωπα, φωνές), ενώ για πιο περιορισμένη αναζήτηση χρησιμοποιείται ο όρος $[LEXY \rightarrow CanMatch]$ («ψηλός», $[FEM]$) με τον οποίο αναζητείται μόνο το θηλυκό του επιθέτου.

Κανόνες

1. Συμφωνία σε γένος και αριθμό και πτώση [*ONTO?* = *\$x:*
GNC_Agreement(1)],
2. συμφωνία σε αριθμό και πτώση [*ONTO?* = *\$x:*
GNC_Agreement(2)],
3. συμφωνία μόνο σε αριθμό [*ONTO?=\$x:**GNC_Agreement(3)*] ή
4. συμφωνία μόνο στη πτώση [*ONTO?=\$x:**GNC_Agreement(4)*]

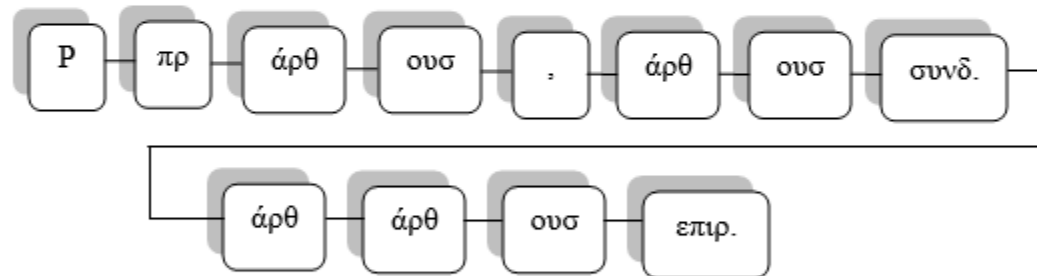
Κανόνες Συμφωνίας



Κανόνες

- Στη φράση: *[πήγαμε στην πόλη, το χωριό και το γήπεδο σήμερα.]* ο διορθωτής επισημαίνει τη λανθασμένη χρήση των άρθρων *[το]* και *[το]* και προτείνει τη χρήση πρόθεσης και άρθρου *[στο]*.
- Ίδια συντακτική ανάλυση όμως προκύπτει και στη φράση: *[πήγαμε στην πόλη, τη Δευτέρα και την Τετάρτη κάποτε].*

- Και στις δύο φράσεις η ανάλυση έχει τη μορφή:



- Στη δεύτερη φράση όμως δεν έχει ισχύ ο κανόνας που περιγράφηκε. Ως εκ τούτου -σε άλλο σύνολο κανόνων (alias.flow) και σε άλλο επίπεδο (2^ο)- αναγνωρίζεται το ονοματικό σύνολο [την Τετάρτη] -όπως και όλες οι ημερομηνίες- ως VTEXT=Date. Με αυτό τον τρόπο σε φράσεις με την ίδια συντακτική ανάλυση αλλά με παρουσία VTEXT=Date δεν εφαρμόζεται ο κανόνας.

Κανόνες

Επίπεδα ανάλυσης

Section

```
/* GGC_TELIKO_N_remove_n_12 */
```

```
{1}
```

```
[ARULE="GGC_TELIKO_N_remove_n_12",  
VTEXT"__αρχαιαμετοχή__"] =>
```

```
\
```

```
[LEXY->CanMatch("μη",[PARTICLE])]
```

```
/
```

```
(
```

```
[LEXY->HasMAttrs([LEARNED])] |
```

```
[LEXY->HasMAttrs([PART,LEARNED])])
```

```
)
```

```
;
```

```
end
```

Rules

```
/* GGC_oral_noun_2*/
```

```
{1}
```

```
[ARULE="GGC_oral_noun_2",  
EVTEXT=TagEvent("gevent.info","GEVENT","ORAL_NOUN","%f","ERMSG","
```

Το λήμμα 'βουλευτίνα' χρησιμοποιείται στον προφορικό λόγο.

Αντικαταστήστε το 'βουλευτίνα' με 'βουλευτής'.")] =>

```
\
```

```
[LEXY->CanMatch("βουλευτής",[N, FEM, ORAL])]
```

```
/
```

```
;
```

Rules

```
/* GGC_oral_noun_4*/
```

```
{1}
```

```
[ARULE="GGC_oral_noun_4",
```

```
EVTEXT=TagEvent("gevent.info","GEVENT","ORAL_NOUN","%f","ERMSG","To
```

```
συγκεκριμένο λήμμα χρησιμοποιείται στον προφορικό λόγο. Αντικαταστήστε με τη  
φράση 'συνάδελφος'.")] =>
```

```
\
```

```
[LEXY->HasLemma3("συναδέρφισσα")]
```

```
/
```

```
;
```

Rules

```
/* GGC_oral_noun_7*/ // ok 22/8/2012
```

```
{1}
```

```
[ARULE="GGC_oral_noun_7",
```

```
EVTEXT=TagEvent("gevent.info","GEVENT","ORAL_NOUN","%f","ERMSG","
```

Το συγκεκριμένο λήμμα χρησιμοποιείται στον προφορικό λόγο.

Αντικαταστήστε με τη φράση 'λυκειάρχης'.")]

```
\
```

```
[TTEXT->Match("λυκειάρχισσα")] |
```

```
[TTEXT->Match("λυκειάρχισσας")] |
```

```
[TTEXT->Match("λυκειάρχισσες")]
```

```
/
```

```
;
```

Rules

```
* GGC_oral_noun_10*/
```

```
{1}
```

```
[ARULE="GGC_oral_noun_10",
```

```
EVTEXT=TagEvent("gevent.info","GEVENT","ORAL_NOUN","%f","ERMSG",
```

```
Το συγκεκριμένο λήμμα χρησιμοποιείται στον προφορικό λόγο.
```

```
Αντικαταστήστε με το λήμμα 'μάγειρας'.")]
```

```
\ [TTEXT=="μάγερας"] |
```

```
[TTEXT=="μάγερα"] |
```

```
[TTEXT=="μάγερες"]
```

```
/
```

```
;
```

Rules

```
/* GGC_oral_noun_14*/
```

```
{2}
```

```
[ARULE="GGC_oral_noun_14",
```

```
EVTEXT=TagEvent("gevent.info","GEVENT","ORAL_NOUN","%f","ERMSG","Τα ονόματα μηνών εκφέρονται στον γραπτό λόγο στη λογιότερη μορφή τους (Ιούνιος, Ιανουάριος, κ.λπ.). Σε περίπτωση που λόγοι ύφους επιβάλλουν την πρόκριση της δημοτικής μορφής των μηνών, προσθέστε το άρθρο 'του' πριν τον μήνα.") =>
```

```
    [ORTHO->IsEqual([WthDgts])]
```

```
    \
```

```
        [TTEXT=="Οκτώβρη"] |
```

```
        [TTEXT=="Σεπτέμβρη"] |
```

```
        [TTEXT=="Ιούνη"] |
```

```
        [TTEXT=="Ιούλη"] |
```

```
        [TTEXT=="Γενάρη"] |
```

```
        [TTEXT=="Φλεβάρη"] |
```

```
        [TTEXT=="Μάρτη"] |
```

```
        [TTEXT=="Απρίλη"] |
```

```
        [TTEXT=="Νοέμβρη"] |
```

```
        [TTEXT=="Οκτώβρη"] |
```

```
        [TTEXT=="Δεκέμβρη"]
```

```
    /
```

```
    ;
```


Rules

```
/* GGC_oral_noun_16*/ // ok 22/8/2012
```

```
{1}
```

```
[ARULE="GGC_oral_noun_16",
```

```
EVTEXT=TagEvent("gevent.info","GEVENT","ORAL_NOUN","%f","ERMSG",
```

```
Το λήμμα 'πνέμα' χρησιμοποιείται στον προφορικό λόγο. Αντικαταστήστε  
το 'πνέμα' με 'πνεύμα'.)] =>
```

```
\
```

```
[TTEXT ->Prefix("πνέμ")] |
```

```
[TTEXT ->Prefix("πνεμάτ")]
```

```
/
```

```
;
```

Rules

```
/* GGC_oral_verb_6*/
```

```
{|}
```

```
[ARULE="GGC_oral_verb_6",
```

```
EVTEXT=TagEvent("gevent.info","GEVENT","ORAL_VERB","%f","E
```

```
RMSG","Η συγκεκριμένη κατάληξη 'ουνα' χρησιμοποιείται στον  
προφορικό λόγο. Σβήστε το 'α'.")]
```

```
\
```

```
[TTEXT ->Suffix("ουνα"), LEXY->HasMAttrs([V,A_P])]
```

```
/
```

```
;
```

Rules

```
/* GGC_oral_verb_4_1*/ // ok 24/8/2012
```

```
{4} // 4 επίπεδο για να μην μπερδεύεται με το γελάγανε
```

```
[ARULE="GGC_oral_verb_4_1",
```

```
EVTEXT=TagEvent("gevent.info","GEVENT","ORAL_VERB","%f","ERMSG","Οι τύποι σε  
'-ανε' και '-ούνε' συνηθίζονται κυρίως στον προφορικό λόγο και σε ανεπίσημο ύφος.  
Σε γραπτά κείμενα και επίσημο ύφος συνηθίζεται η κατάληξη '-ουν'.")] =>
```

```
\
```

```
[LEXY->HasMAttrs([V, C_P, PLUR]),TTEXT->Suffix("άνε")] |
```

```
[LEXY->HasMAttrs([V, C_P, PLUR]),TTEXT->Suffix("ούνε")]
```

```
/
```

```
i
```

Rules

```
/* GGC_ancient_phrase_49_1*/ //15-10-2012
{1}
[ARULE="ancient_phrase_49_1",
EVTEXT=TagEvent("gevent.wrong","GEVENT","ancient_phrase","%f","ERMSG","Ο
σηματισμός της λόγιας φράσης είναι λανθασμένος. Αντικαταστήστε το 'παρόν' με το
'παρών'.")] =>
    [TTEXT->Match("το")]
\
    [VTEXT=="_QUOTE_"?],
    [TTEXT->Match("παρόν")],
    [VTEXT=="_QUOTE_"?
/
    [LEXY->HasNoneMAttrs([NEUT]), // για να μην χτυπάει το παρόν π.χ. σύστημα
        []{0,5},
        [LEXY->HasLemma3("δίνω")]
i

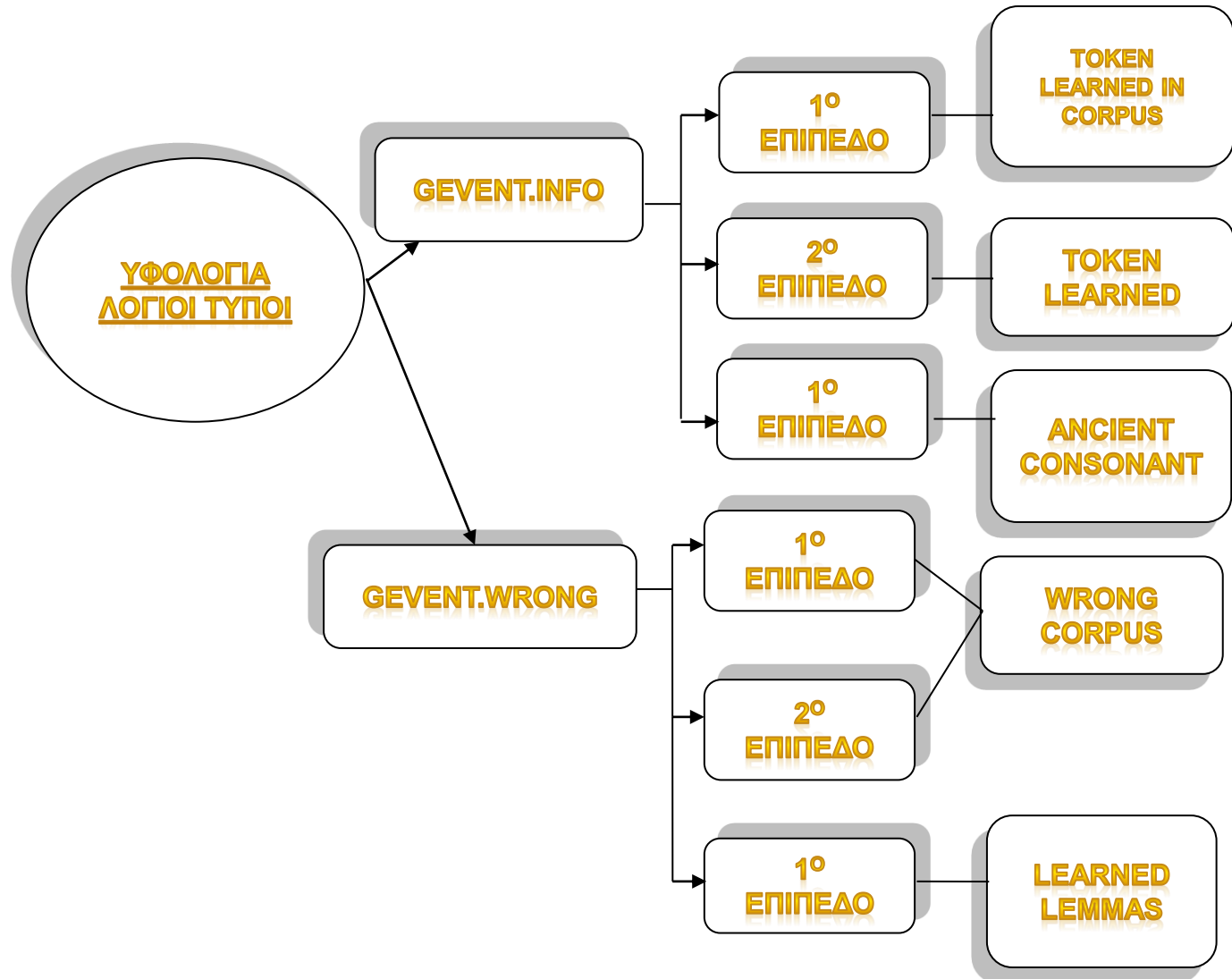
// αυτός ο κανόνας για να χτυπάει: το παρόν θα δώσει καιο υπουργός ....
```

Rules

```
/* GGC_TELIKO_N_add_n_2*/
{2,4} // για τη φράση: τον γενικό γραμματέα
[ARULE="GGC_TELIKO_N_add_n_2",
EVTEXT=TagEvent("gevent.wrong","GEVENT","FINAL_N","%f","ERMSG","Το 'το' πρέπει να
έχει τελικό -ν. Αντικαταστήστε το 'το' με το 'τον'.")] =>
\
    [TTEXT->$x_l_3:Match("το"), LEXY->CanMatch("ο",[ART,ACC, MASC, SING])]
/
    (
[LEXY->HasMAttrs([AD],MASC,SING,ACC), ORTHO->AnyOfOAttrs([Style2])] |
[LEXY->HasMAttrs([PART,MASC,SING,ACC]), ORTHO->AnyOfOAttrs([Style2])]
    ),
    [LEXY->HasMAttrs([PRON, GEN])?], // το καλό μας συμμαθητή
    [LEXY->HasMAttrs([N,MASC,SING,ACC])]
;

```

Κανόνες υφολογίας (λόγιοι τύποι)



Κανόνες υπολογίας (λόγιοι τύποι)

Section

```
/* GGC_pattern_ancient_I*/
```

```
{I}
```

```
[ARULE="pattern_ancient_I",VTEXT="__ancient_phrase__"] =>
```

```
(
```

```
    [LEXY->HasLemma3("επίδομα")] |
```

```
    [LEXY->HasLemma3("φύλλο")]
```

```
),
```

```
\
```

```
[TTEXT=="αδείας"]
```

```
/
```

```
;
```

```
end
```

Κανόνες υπολογίας (λόγιοι τύποι)

Section

```
/* GGC_pattern_2_ancient_2_2 */
```

```
{2}
```

```
[ARULE="GGC_pattern_ancient_2_2",  
  EVTEXT=TagEvent("gevent.info","GEVENT","ANCIENT_T  
  YPE","%f","ERMSG","Ο τονισμός είναι λόγιος. Τονίστε στην  
  προπαραλήγουσα.")] =>
```

```
\
```

```
[TTEXT == "αδειάς"]
```

```
|
```

```
[TTEXT == "ακριβείας"]
```

```
/
```

```
;
```

```
end
```


Κανόνες υφολογίας (προφορικοί τύποι)

- α) λήμματα (**ORAL**),
- β) καταλήξεις (προφορικός υφολογικός χαρακτηρισμός)

/ GGC_oral_adj_1_1*/*

[.....] Ο τύπος '-ούχα/ -γόνα' χρησιμοποιείται στον προφορικό λόγο. Αντικαταστήστε με '-ούχος / -γόνος'.") =>

[LEXY->HasMAttrs([FEM]), ONTO?=\$x:GNC_Agreement(2,[FEM])],

[]{0,4},

[TTEXT->Match("είναι")],

[]{0,3}

\

[LEXY->SuffixCanMatch("ούχος",[AD],FEM)],TTEXT->Suffix("ούχα"),
ONTO?=\$x:GNC_Agreement(2,[AD])]

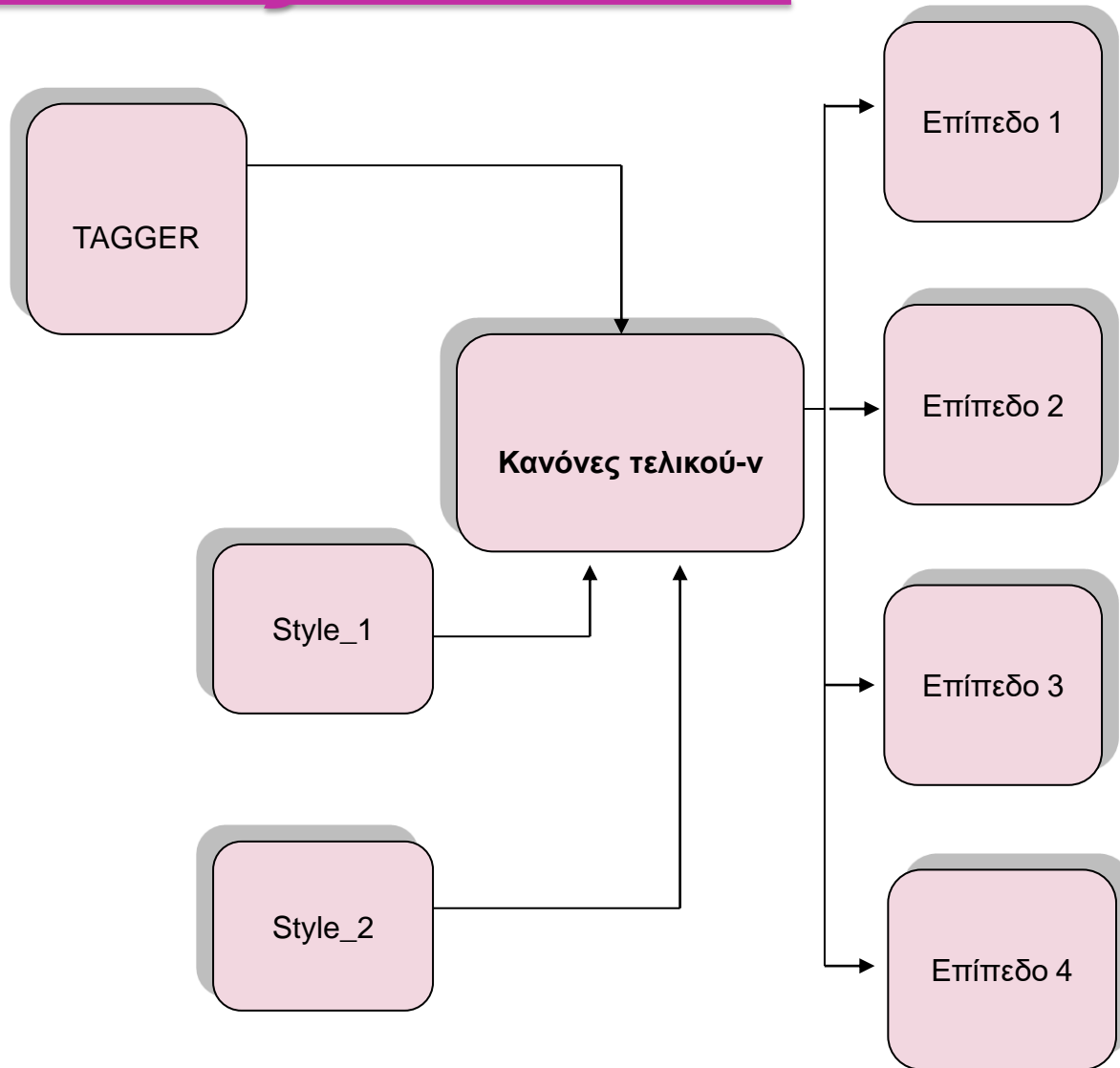
[LEXY->SuffixCanMatch("γόνος",[AD],FEM)],TTEXT->Suffix("γόνα"),
ONTO?=\$x:GNC_Agreement(2,[AD])]

/

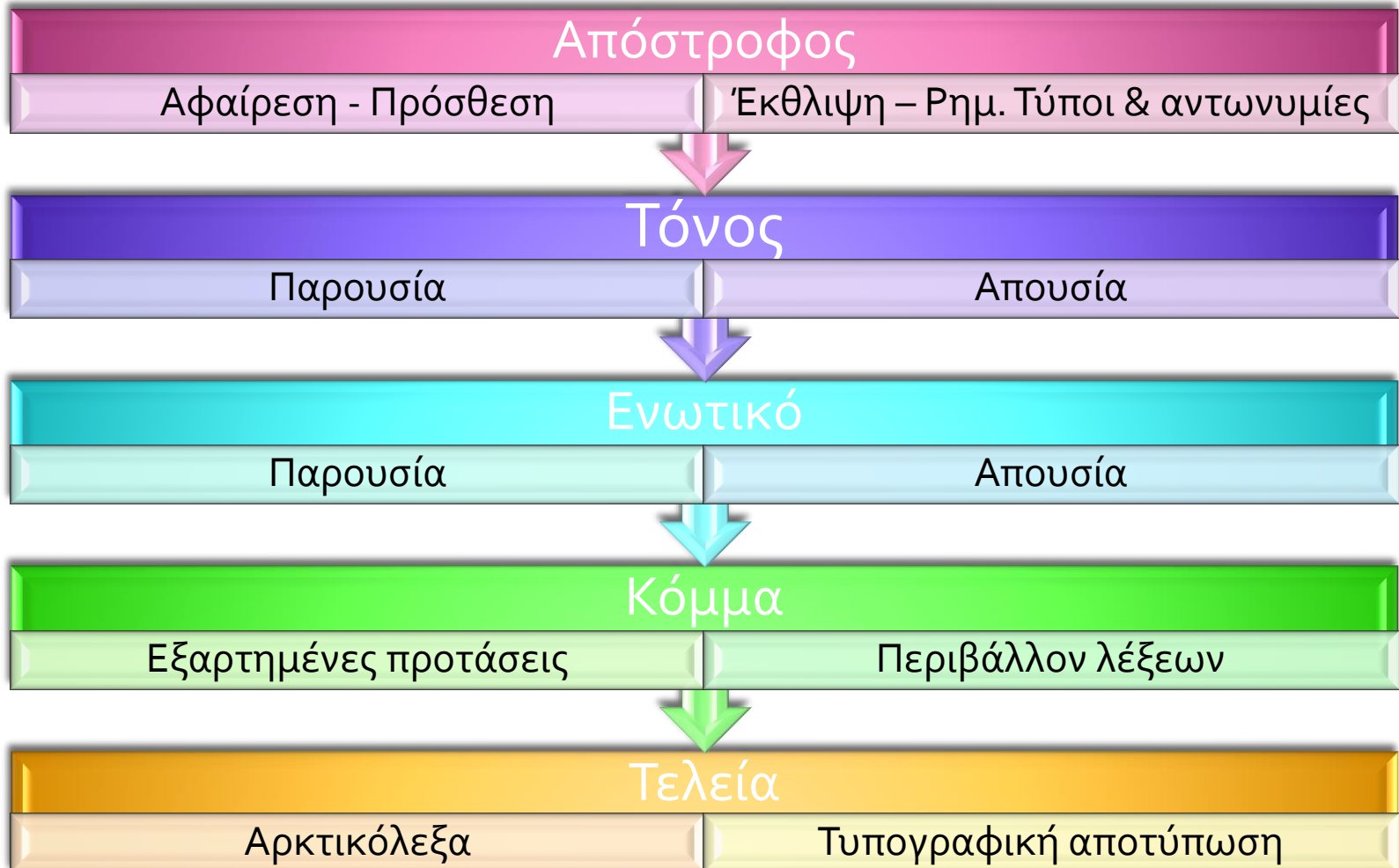
;

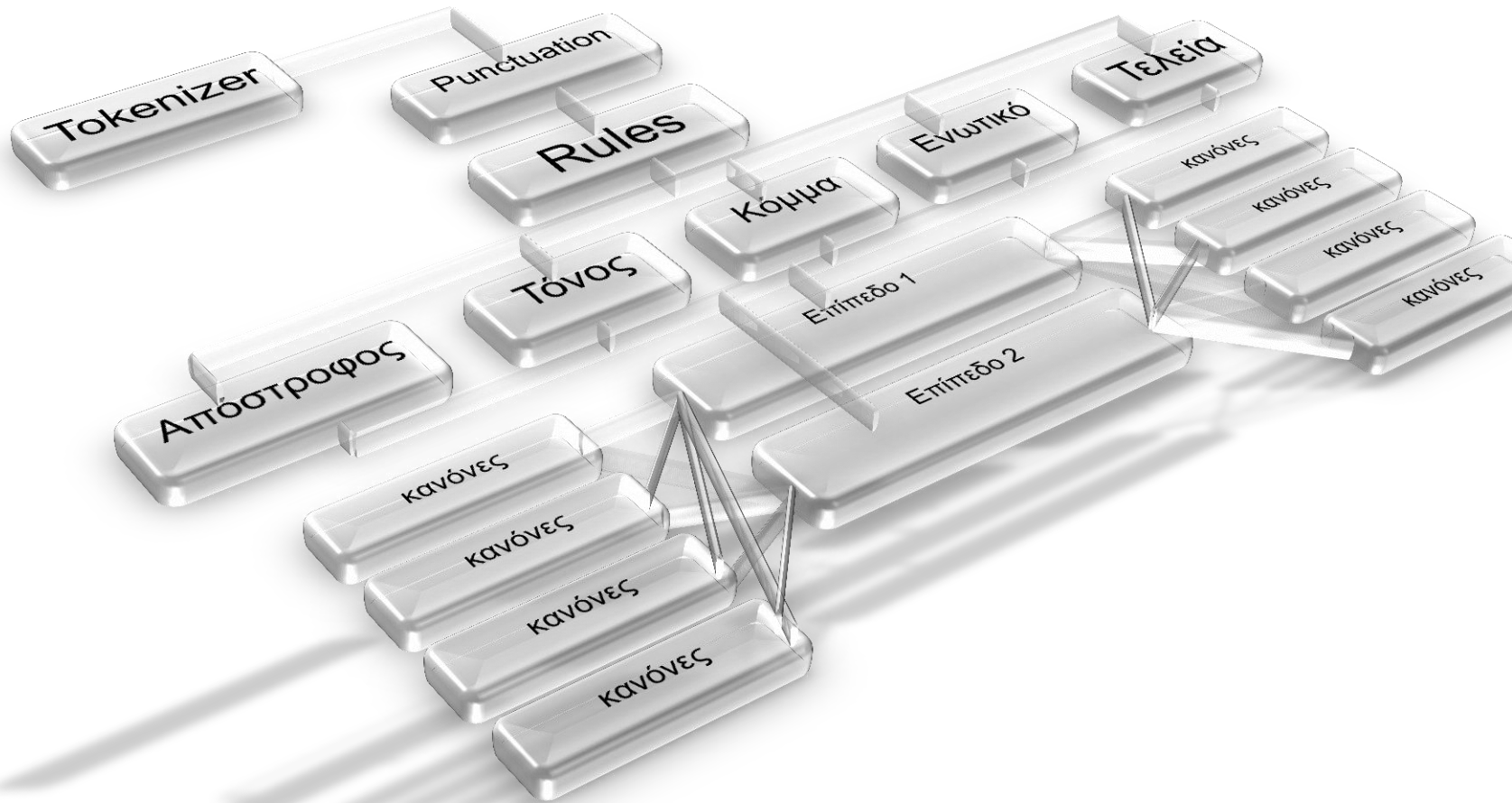
end

Κανόνες τελικού -v



Κανόνες για τόνους, ορθογραφικά σημεία και σημεία στίξης





Εννοιολογική σύγχυση

- Ενημέρωση:
 - α) πιθανή άστοχη επιλογή λήμματος
 - β) λήμμα με το οποίο συγχέεται
 - γ) για τη σημασία και των δύο λημμάτων
- αποδελτίωση γλωσσικού περιβάλλοντος
- λειτουργία επιπέδων

* GGC_OMOPRON_allpart_7_2 */

{1}

[ARULE="GGC_OMOPRON_allpart_7_2",
VTEXT="__δηκτικός__"] =>

\
[LEXY->CanMatch("δηκτικός",[AD],NEUT)]

/

[] {0,3},

(

[LEXY->HasLemma3("ύφος")] |

[LEXY->HasLemma3("κείμενο")] |

[LEXY->HasLemma3("σχόλιο")] |

[LEXY->HasLemma3("πνεύμα")] |

[LEXY->HasLemma3("χιούμορ")]

)

```
/* GGC_OMOPRON_allpart_7_3 */
```

```
{2}
```

```
[ARULE="GGC_OMOPRON_allpart_7_3",  
  EVTEXT=TagEvent("gevent.warning","GEVENT",  
  OMOPRON,"%f","ERMSG","Το λήμμα  
'δεικτικός' συγχέεται με το 'δηκτικός'. Δεικτικός  
σημαίνει 'αυτός που χρησιμεύει για να δείχνει',  
δηκτικός σημαίνει 'πικρόχολος'.")]
```

```
\
```

```
[LEXY->HasLemma3("δεικτικός")] |
```

```
[LEXY->HasLemma3("δηκτικός")]
```

```
/
```

```
;
```



Τυποποίηση γραφής

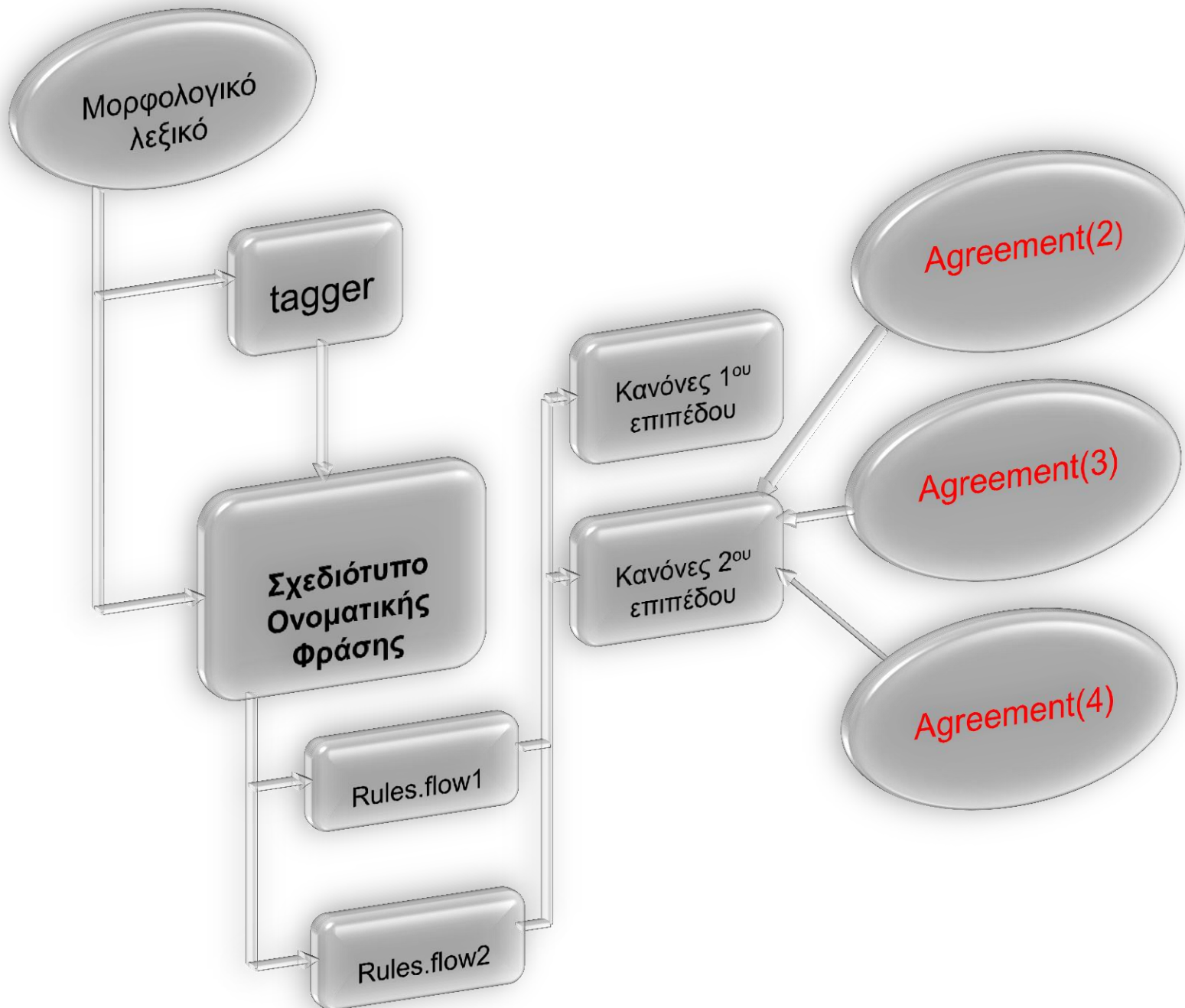


Ισοδυναμία γραφής

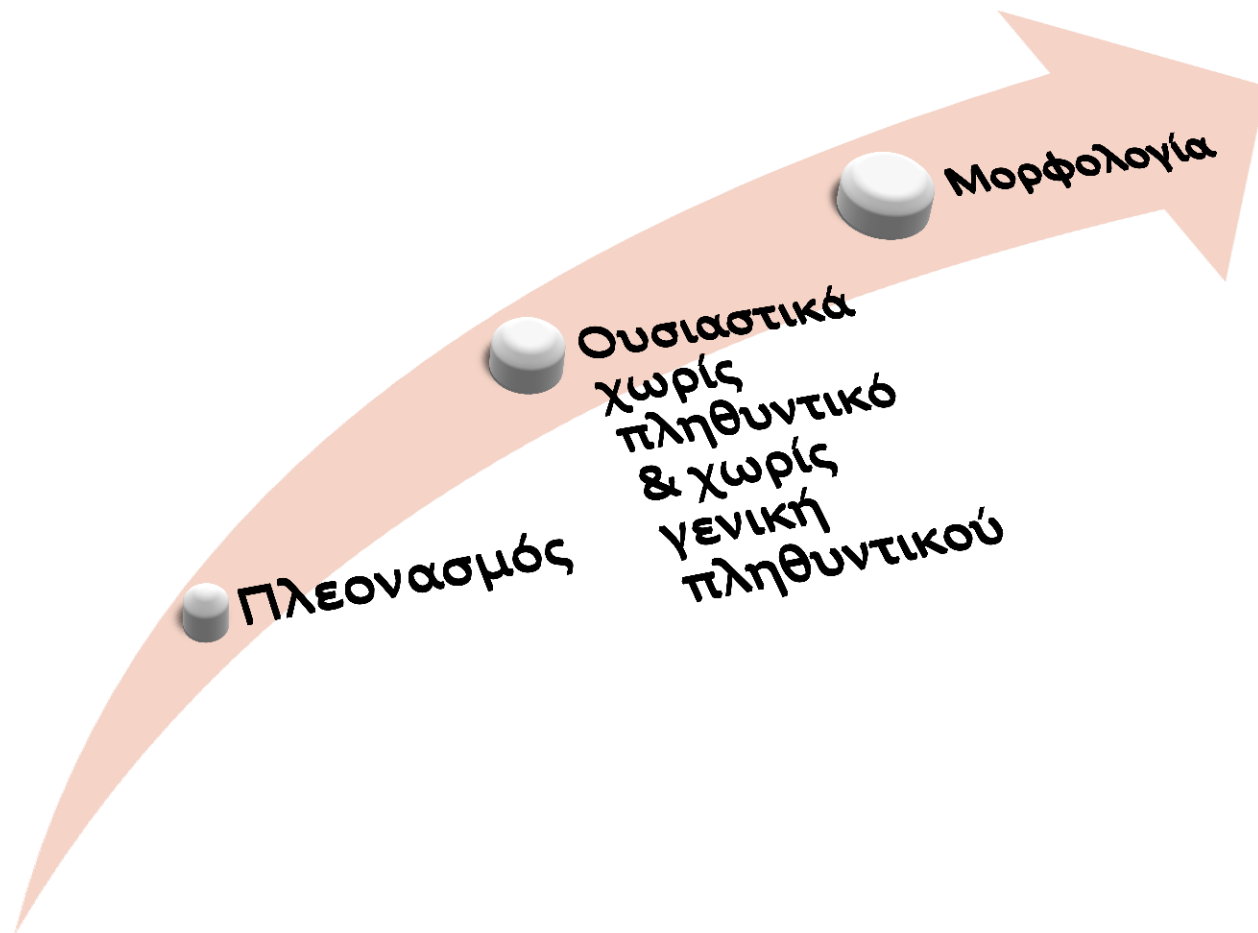


Συμφωνία





Σχεδιότυπα



Στατιστική ανάλυση

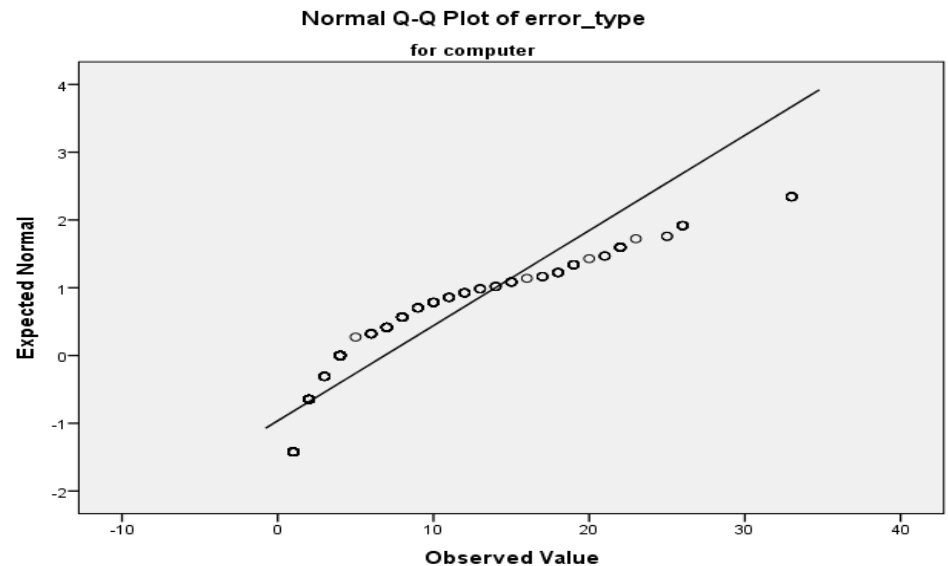
- Η στρεβλότητα (Skewness) για τα λάθη που εντόπισε ο άνθρωπος-διορθωτής είναι $\gamma_1 = -1,55$ και για τον γραμματικό διορθωτή $\gamma_1 = -4,42$.

Έλεγχος της κανονικότητας της συχνότητας λάθους (Q-Q plot)



Άνθρωπος [$D(557) = ,26; p <,001$]

γραμματικός διορθωτής
[$D(557) = ,26; p <,001$]



Στατιστικά κριτήρια

Μη παραμετρική ανάλυση: λόγω της απόκλισης από την κανονικότητα

Kolmogorov-Smirnov

Έλεγχος υποθέσεων για την κατανομή των τιμών μεταβλητής (Dytham, 2011).

McNemar τεστ

Έλεγχος διαφορών ύστερα από την επίδραση κάποιου παράγοντα (Israel, 2009).

Kruskal-Wallis

Έλεγχος της υπόθεσης ότι τα δείγματα της μελέτης προέρχονται από πληθυσμούς με την ίδια διάμεσο (Sheskin, 2004· Dytham, 2011).

Mann-Whitney U

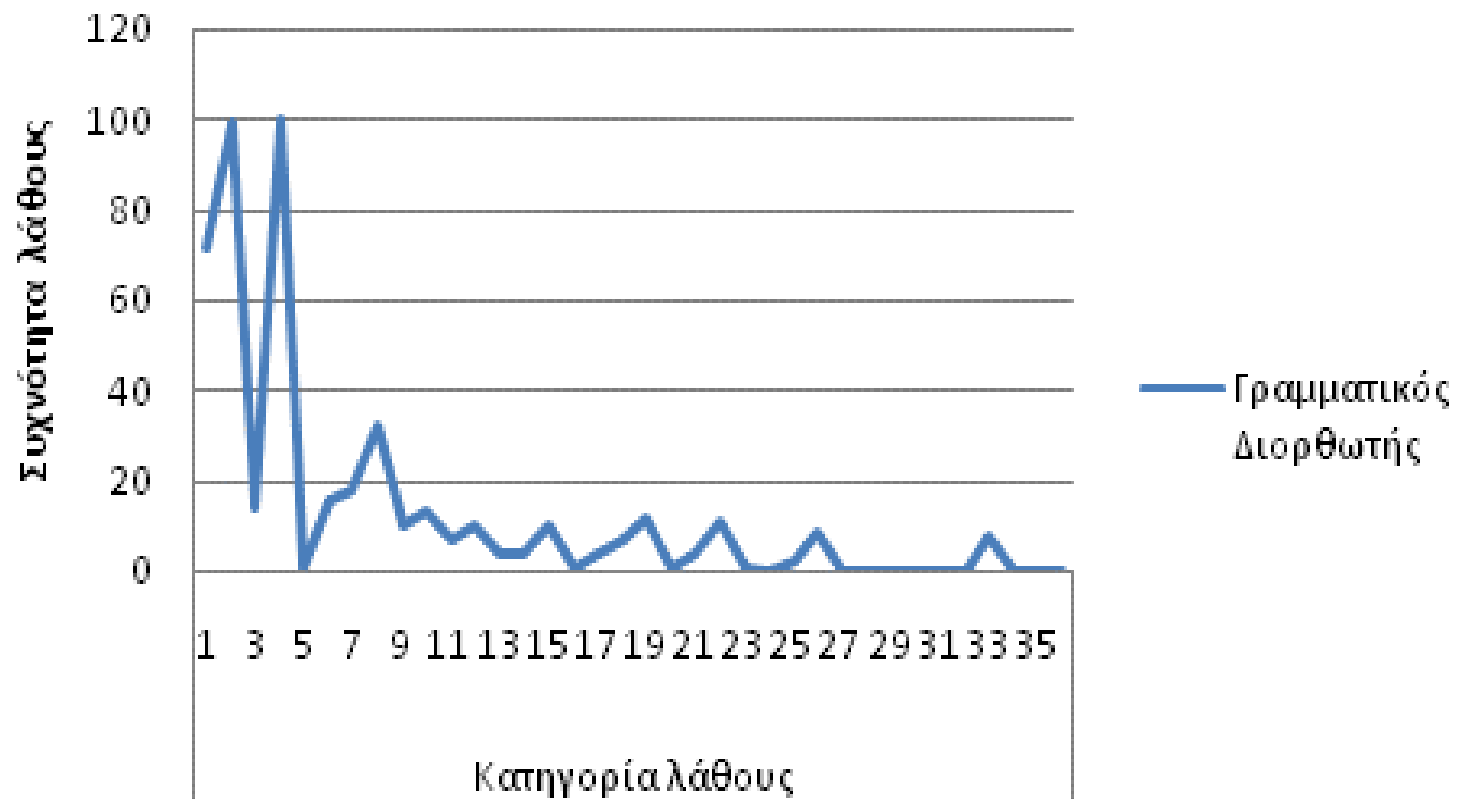
Έλεγχος διαφορών μεταξύ δύο ανεξάρτητων δειγμάτων (Field, 2009).



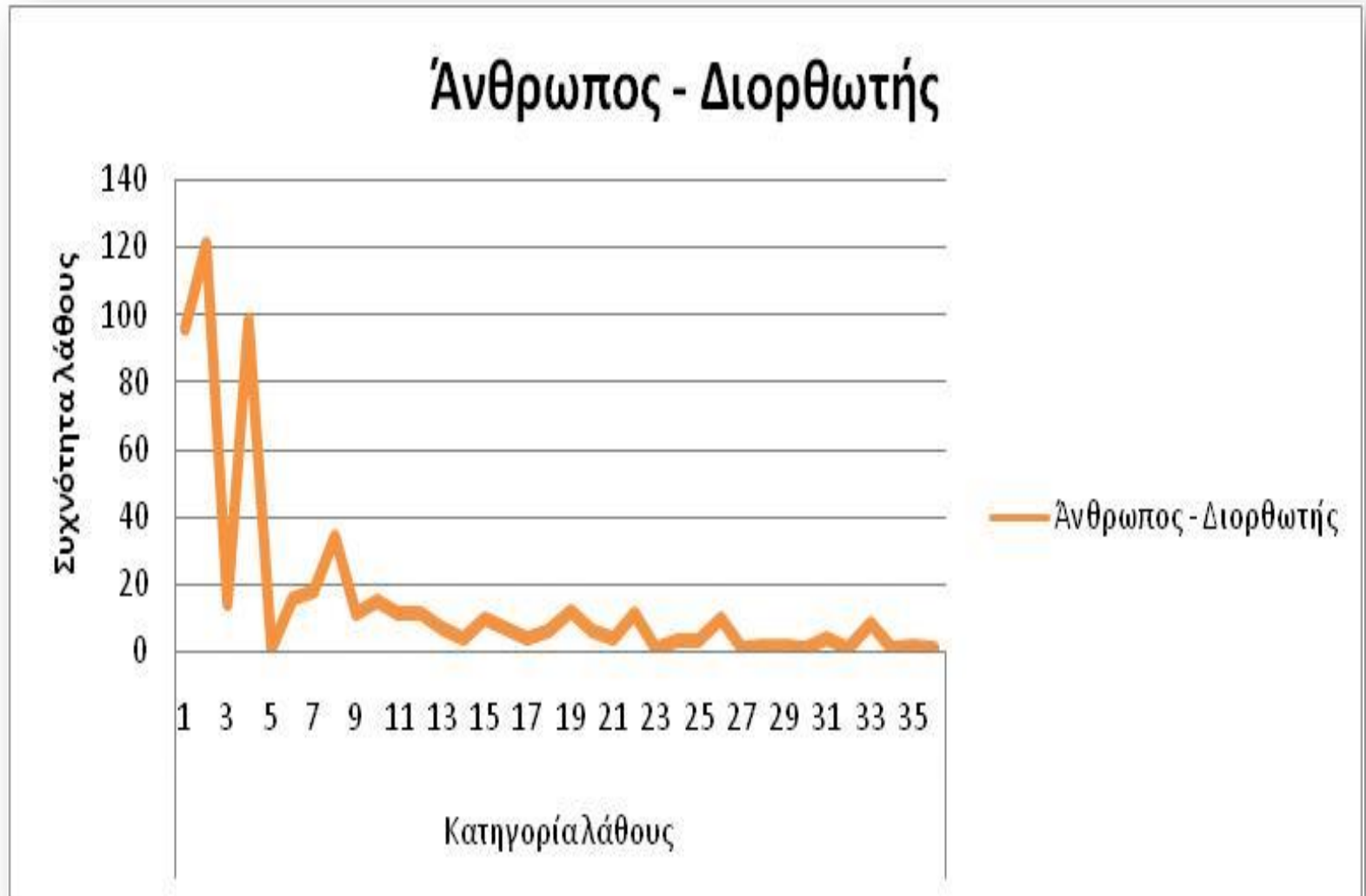
Ευρήματα ανάλυσης λαθών

Αξιολόγηση

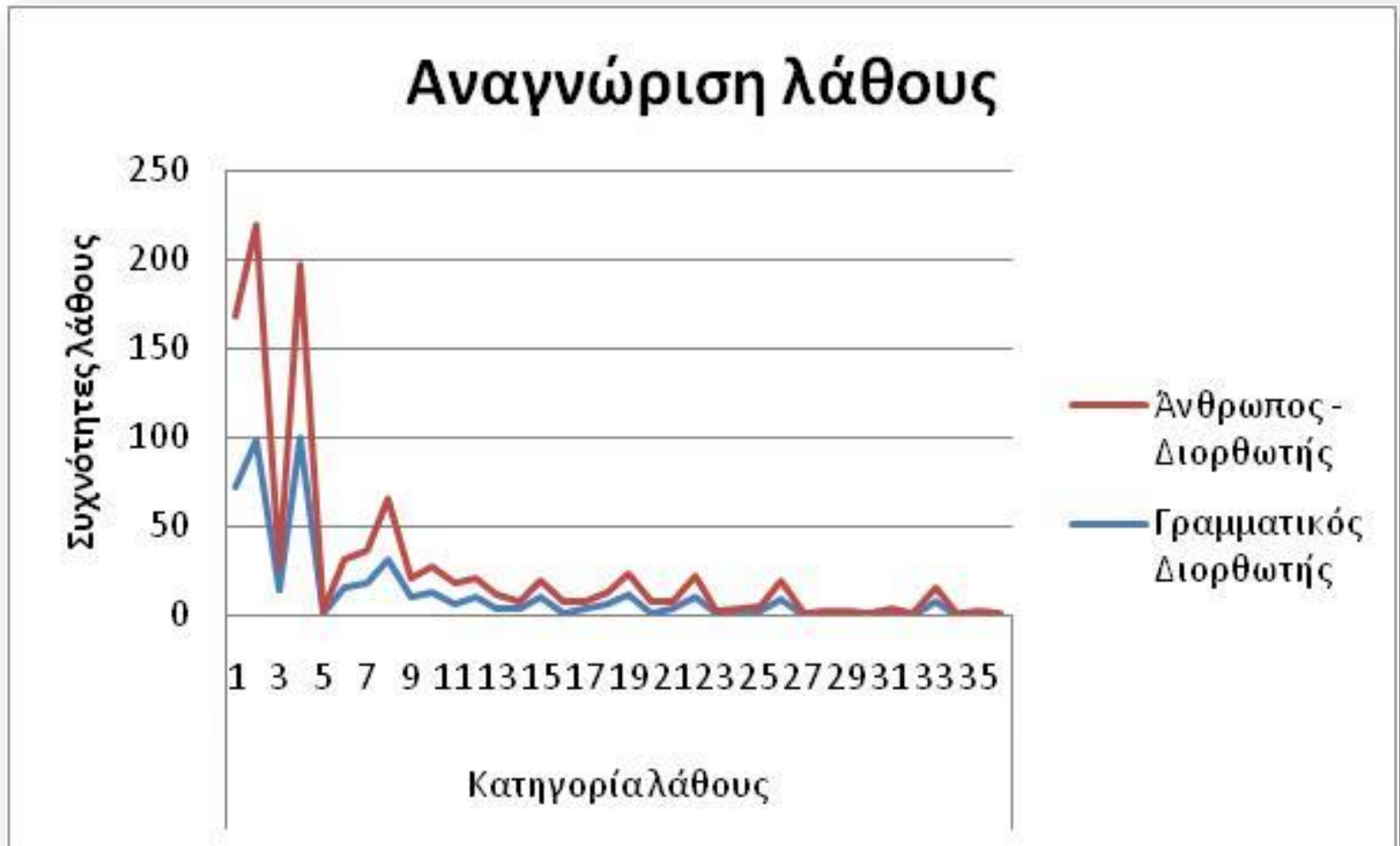
Γραμματικός Διορθωτής



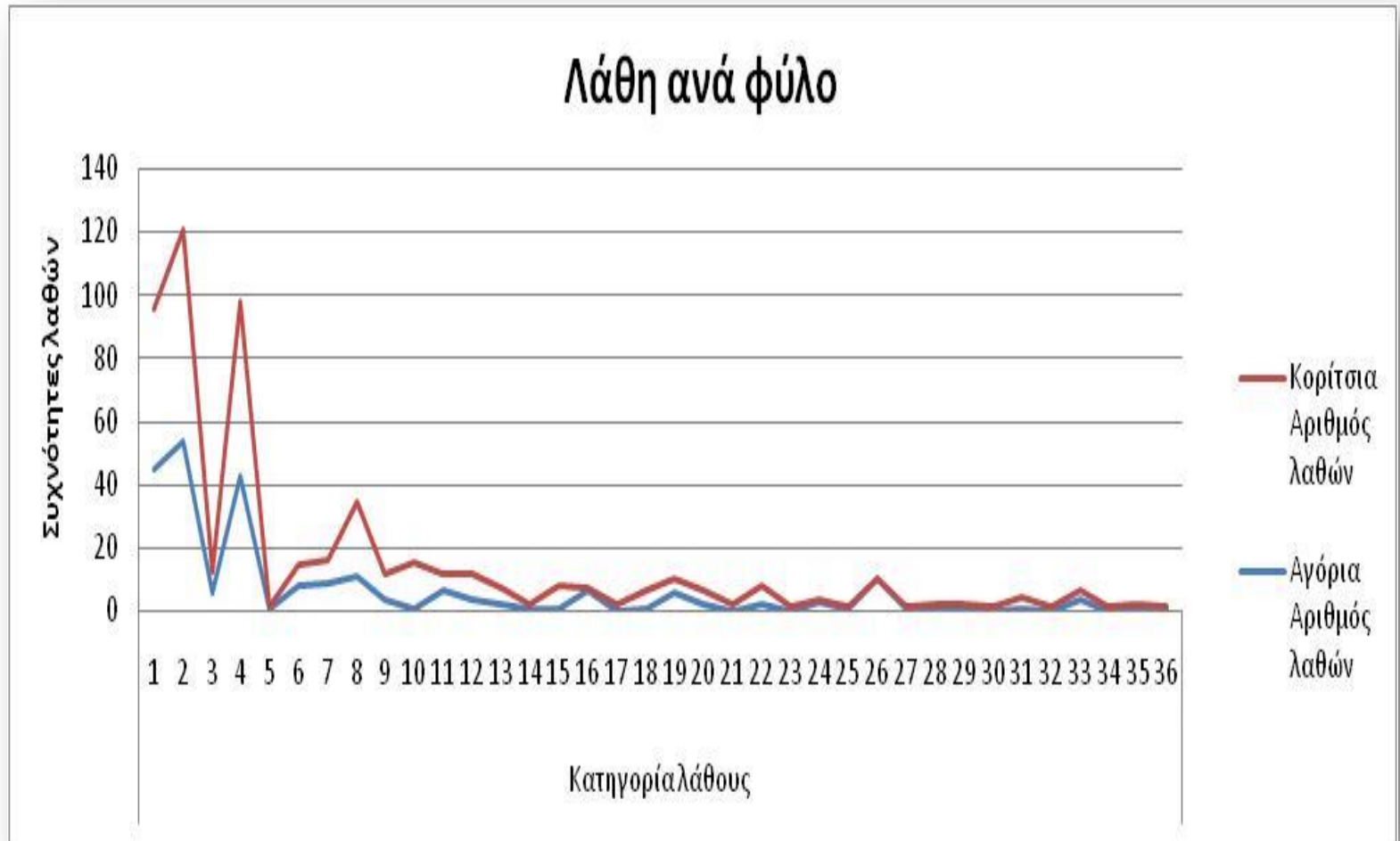
Αξιολόγηση



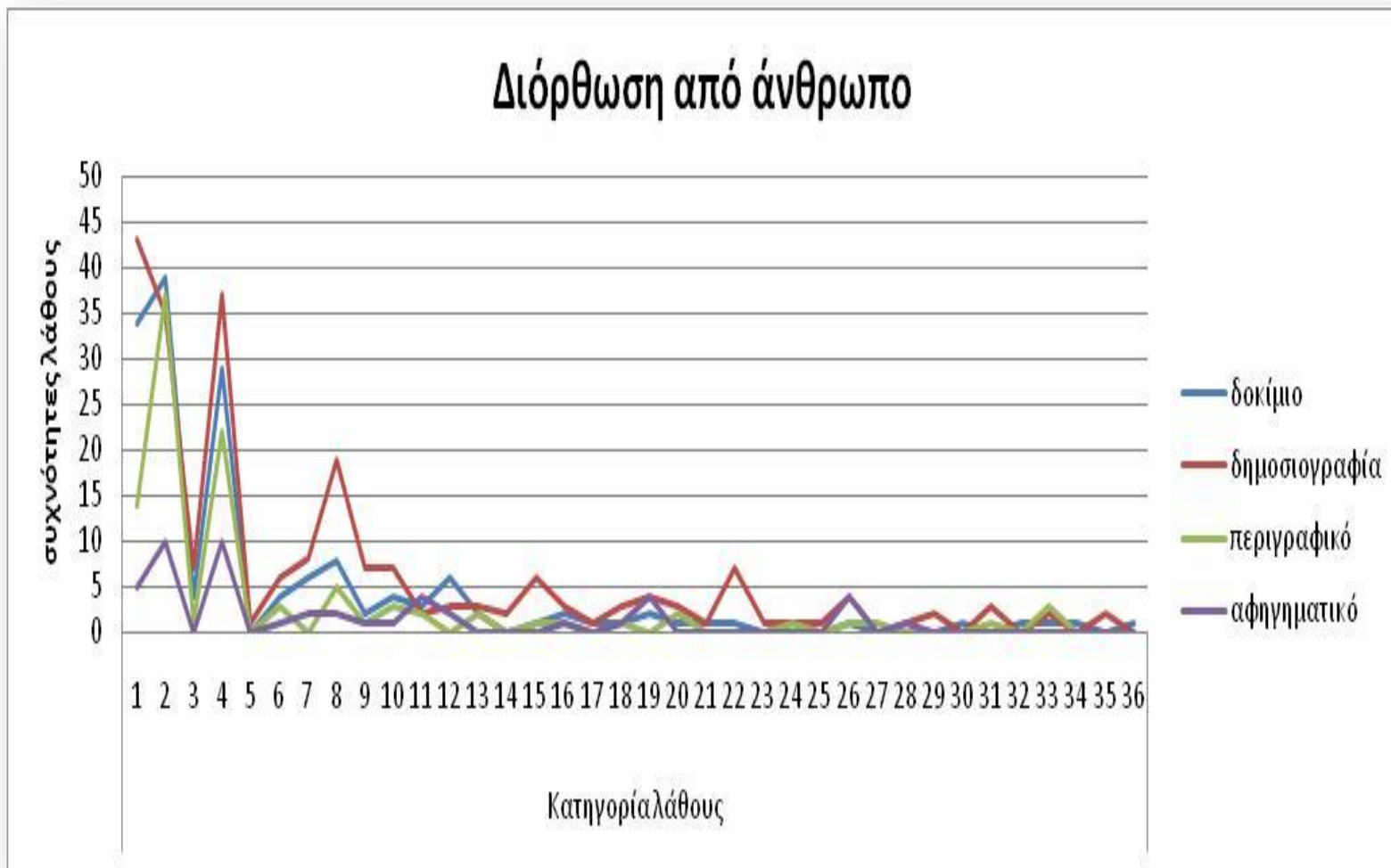
Αξιολόγηση



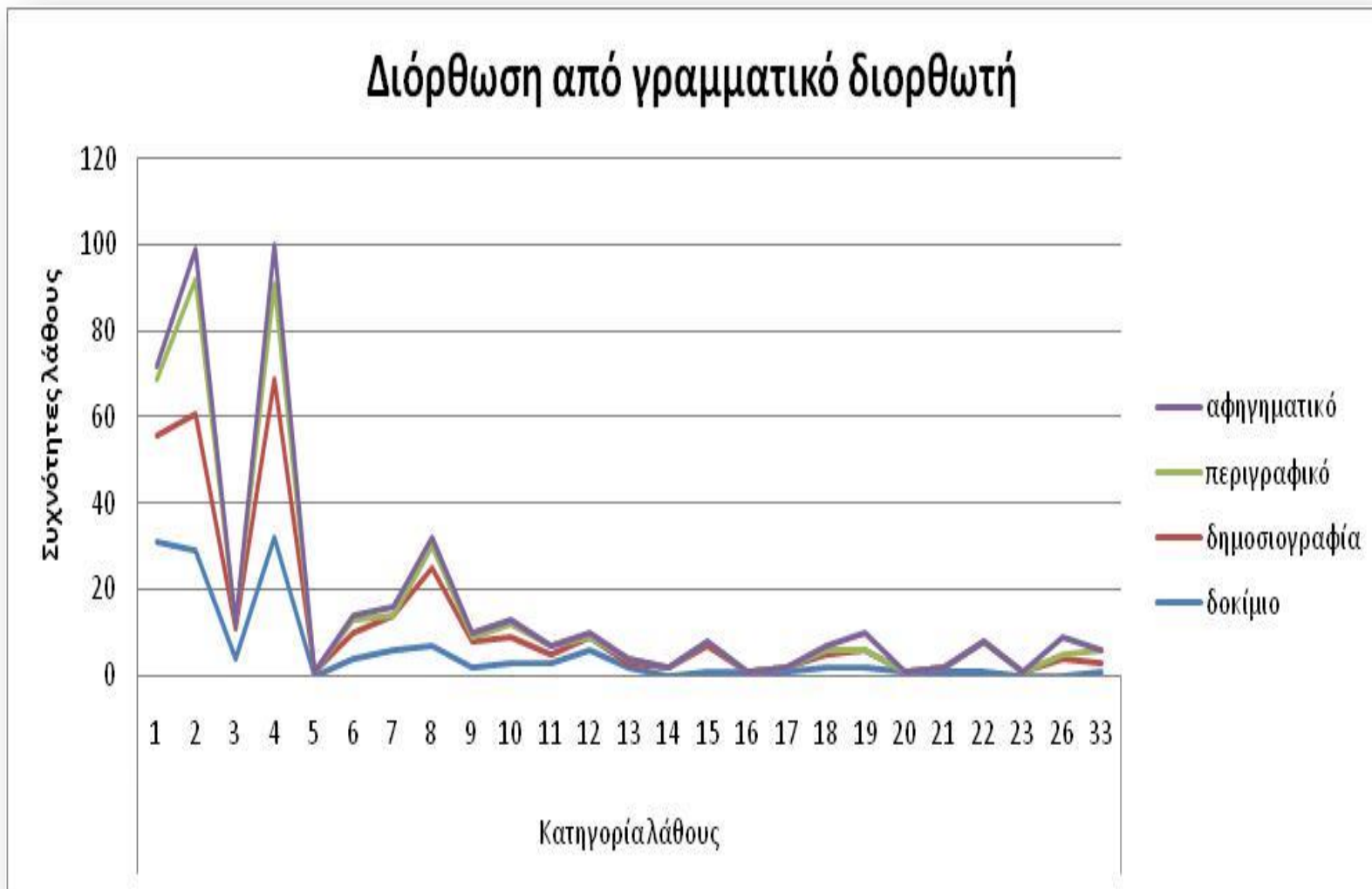
Λάθη ανά φύλο



Συχνότητα λάθους ανά κειμενικό είδος



Συχνότητα λάθους ανά κειμενικό είδος



Συμπεράσματα

- προσεγγίζει τη διόρθωση του ανθρώπου
- ξεπερνά τον ορθογραφικό διορθωτή
- προσπάθεια κωδικοποίησης στοιχείων του μηχανισμού της σκέψης
- υψηλός βαθμός ακρίβειας
- πληρότητα των δεδομένων

Συμπεράσματα

- καινοτόμα εργαλεία
- όμοια σχεδιάσματα με ξενόγλωσσους διορθωτές
- δυνατότητα ενσωμάτωσης σε Ms-office
- **αδυναμίες: ελλειπτικός λόγος, νοηματικά λάθη, «αισθητική απόλαυση»**



Ο Γραμματικός διορθωτής

Κείμενο

Ανάλυση

Ο Γραμματικός διορθωτής

στο web



Ευχαριστώ για την προσοχή σας!