



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΕΛΟΠΟΝΝΗΣΟΥ

ΥΠΟΛΟΓΙΣΤΙΚΗ ΓΛΩΣΣΟΛΟΓΙΑ

4^η διάλεξη

Π. ΓΑΚΗΣ

Υπολογιστικά Λεξικά

- Αναζήτηση μπορεί να γίνει και βάσει άλλων πληροφοριών, εκτός του ίδιου του λήμματος (απλές ή σύνθετες αναζητήσεις), όπως για παράδειγμα αναζήτηση σε:
 1. γραμματική κατηγορία + μορφολογική πληροφορία (όλα τα ουσιαστικά που δεν έχουν πληθυντικό),
 2. επίπεδα χρήσης (λόγιες και μη λόγιες λέξεις),
 3. θεματικό πεδίο (οι λέξεις που χρησιμοποιούνται στην ιατρική ορολογία,
 4. ετυμολογική πληροφορία (οι λέξεις που έχουν αρχαιοελληνική προέλευση)
 5. συντακτική πληροφορία.
- Υπάρχει επιπλέον η δυνατότητα αναζήτησης **πολυλεκτικών σχηματισμών και φράσεων ως αυτόνομων ενοτήτων** που στα παραδοσιακά λεξικά αποτελούν ενδολήμματα ή υπολήμματα λεξικογραφικών άρθρων

Σχεδιασμός λεξικού

- Το μοντέλο προσέγγισης που υιοθετήθηκε για το συγκεκριμένο λεξικό (Gakis et al., 212) σχεδιάστηκε ειδικά και εξαρχής για τη νεοελληνική γλώσσα.
- Ο σχεδιασμός του έγινε βάσει αλγορίθμων απλών στον σχεδιασμό τους αλλά αποτελεσματικών ως προς τη διαδικασία παραγωγής και αναγνώρισης οποιοδήποτε τύπου καθώς και ως προς την ταχύτητα πρόσβασης στα δεδομένα με δυνατότητα διαρκούς συντήρησης, εμπλουτισμού και βελτίωσης.

Σχεδιασμός λεξικού

- Σε πρώτο επίπεδο τυποποιήθηκε το σύνολο των αρχών της μορφοποίησης και της μορφολογικής περιγραφής των καταλήξεων της νεοελληνικής γλώσσας με την ταξινόμησή τους σε διακριτούς τύπους των ξεχωριστών μερών του λόγου
- Ταξινομώντας τα κλιτά λήμματα σε κατηγορίες, ανάλογα με το μέρος του λόγου, **ελαχιστοποιούμε τις πληροφορίες που απαιτούνται να αποθηκευτούν και μεγιστοποιούμε** συγχρόνως την **ταχύτητα των διαδικασιών αναγνώρισης και παραγωγής** των τύπων τους. Με τον τρόπο αυτόν το λεξικό εμπλουτίζεται πιο εύκολα καθώς η μόνη πληροφορία που χρειάζεται να καταχωρηθεί για κάθε καινούρια λέξη είναι η κλίση και τα θέματά της.

LexEdit

- Για την αυτοματοποίηση και απλοποίηση της κωδικοποιημένης λεξιλογικής πληροφορίας υλοποιήθηκε ένα ειδικό εργαλείο το LexEdit.

LexEdit

giota.nlx - Neuro Lexicon Editor

File Edit Morphos Lemma Dictionary View Help

Sections

- άλφα
- βήτα
- γάμα
- δέλτα
- έψιλον
- ζήτα
- ήτα
- θήτα
- γιώτα
- κάπα
- λάμδα
- μι
- νι
- ξ
- όμικρον
- πι
- ρο
- σίγμα
- ταυ
- ύψιλον
- φι

Lemma	Main Morphology	# Meanings	POS	Description
ισοθεραπεία	{ι-στο-}{θε-ρα-η}[ει-] #...	1	(N)	
ιστοκαλλιέργεια	{ι-στο-}{καλ-λι-}{ερ-γ}...	1	(N)	-ειών
ιστολογία	{ι-στο-}{λο-γ}[ι-] #1#	1	(N)	σιάν. πληθ.
ιστολογικά	{ι-στο-}{λο-γ}[ι-κ] #1#	1	(ADV)	
ιστολογικός	{ι-στο-}{λο-γ}[ι-κ] #1#	1	(ADJ)	
ιστόλυση	{ι-στο-}{λυ-}[σ] #1#	1	(N)	-ης κ' -εως
ιστολυσία	{ι-στο-}{λυ-}[σ][ι-] #1#	1	(N)	
ιστολυτικά	{ι-στο-}{λυ-}[τ][ι-κ] #1#	1	(ADV)	
ιστολυτικός	{ι-στο-}{λυ-}[τ][ι-κ] #1#	1	(ADJ)	
ιστόρημα	{ι-στο-ρ}[η-μ] #2#	1	(N)	
ιστορημένος	{ι-στο-ρ}[η-με-ν] #1#	1	(PART)	
ιστόρηση	{ι-στο-ρ}[η-σ] #1#	1	(N)	
ιστορία	{ι-στο-ρ}[ι-] #1#	1	(N)	
ιστορίζω	{ι-στο-ρ}[ι-ζ] #8#	1	(V)	-ισα, -ομαι, -ις
ιστορικά	{ι-στο-ρ}[ι-κ][ο-][τε-ρ] ...	1	(ADV)	
ιστορικός	{ι-στο-ρ}[ι-κ][ο-][τε-ρ] ...	1	(ADJ)	
ιστορικός2	{ι-στο-ρ}[ι-κ] #1#	1	(N)	
ιστορικός2	{ι-στο-ρ}[ι-κ] #3#	1	(ADJ)	
ιστορικός2	{ι-στο-ρ}[ι-κ] #1#	1	(N)	
ιστορικοσυγκριτικά	{ι-στο-ρ}[ι-κ][ο-]<συ-γ...	1	(ADV)	
ιστορικοσυγκριτικός	{ι-στο-ρ}[ι-κ][ο-]<συ-γ...	1	(ADJ)	
ιστορικότητα	{ι-στο-ρ}[ι-κ][ο-τη-τ] #...	1	(N)	
ιστοριογνώσια	{ι-στο-ρ}[ι-][ο-]{γνω-σ...	1	(N)	

Ready NUM

- Στο συγκεκριμένο στιγμιότυπο οθόνης έχει επιλεγεί το γράμμα [Ι]. Οι πληροφορίες που παρουσιάζονται λεπτομερώς είναι: α) το **λήμμα ή η ετικέτα στην πρώτη στήλη**, β) η **μορφολογία των λημμάτων**, δηλ. τα **ιδρυτικά μορφήματα** (εκτός από το επίθημα), γ) ο **αριθμός εννοιών**, δ) το **μέρος του λόγου (POS)** και ε) μια **περιγραφή** (ή σχόλια) στην τελευταία στήλη

giota.nlx - Neuro Lexicon Editor

File Edit Morphos Lemma Dictionary View Help

Sections	Lemma	Main Morphology	# Meanings	POS	Description
άλφα	ισοθεραπεία	{i-sto-}{the-ra-pi}[ei-] #...	1	(N)	
βήτα	ιστοκαλλιέργεια	{i-sto-}{kala-li-}{er-y}...	1	(N)	-είων
γάμμα	ιστολογία	{i-sto-}{lo-y}[i-] #1#	1	(N)	σπάν. πληθ.
δέλτα	ιστολογικά	{i-sto-}{lo-y}[i-k] #1#	1	(ADV)	
έψιλον	ιστολογικός	{i-sto-}{lo-y}[i-k] #1#	1	(ADJ)	
ζήτα	ιστόλυση	{i-sto-}{lu-}[s] #1#	1	(N)	-ης κ' -εως
ήτα	ιστολυσία	{i-sto-}{lu-s}[i-] #1#	1	(N)	
θήτα	ιστολυτικά	{i-sto-}{lu-}[t][i-k] #1#	1	(ADV)	
ιότα	ιστολυτικός	{i-sto-}{lu-}[t][i-k] #1#	1	(ADJ)	
κάπα	ιστόρημα	{i-sto-r}[h-i-m] #2#	1	(N)	
λάμδα	ιστορημένος	{i-sto-r}[h-me-v] #1#	1	(PART)	
μι	ιστόρηση	{i-sto-r}[h-s] #1#	1	(N)	
νι	ιστορία	{i-sto-r}[i-] #1#	1	(N)	
ξι	ιστορίζω	{i-sto-r}[i-z] #8#	1	(V)	-ισα, -ομαι, -ικ
omicron	ιστορικά	{i-sto-r}[i-k][o-][te-p] ...	1	(ADV)	
πι	ιστορισμός	{i-sto-r}[i-k][i-om] #1#	1	(N)	
ρο	ιστορικό	{i-sto-r}[i-k] #1#	2	(N)	
σίγμα	ιστορικός	{i-sto-r}[i-k] #3#	1	(ADJ)	
ταυ	ιστορικός2	{i-sto-r}[i-k] #1#	1	(N)	
ύψιλον	ιστορικοσυγκριτικά	{i-sto-r}[i-k][o-]<su-y...	1	(ADV)	
φι	ιστορικοσυγκριτικός	{i-sto-r}[i-k][o-]<su-y...	1	(ADJ)	
	ιστορικότητα	{i-sto-r}[i-k][o-tp-t] #...	1	(N)	
	ιστοριογνώσια	{i-sto-r}[i-][o-]{gnw-s...	1	(N)	

Ready NUM

Είδη κωδικοποιημένης πληροφορίας

- Για κάθε λήμμα όπως επίσης και για τους παραγόμενους λεκτικούς τύπους ορίζονται τα πιο κάτω είδη πληροφορίας:
- **α. μορφοσυντακτική**
σε επίπεδο λήμματος περιγράφεται η γραμματική κατηγορία (μέρος του λόγου). Έτσι προκύπτουν τα παρακάτω 11 χαρακτηριστικά (attributes): επίθετο (ADJ), επίρρημα (ADV), άρθρο (ART), σύνδεσμος (CONJ), επιφώνημα (INTERJ), ουσιαστικό (N), μετοχή (PART), μόριο (PARTICLE), πρόθεση (PREP), αντωνυμία (PRON), ρήμα (V).

Είδη κωδικοποιημένης πληροφορίας

- **α. μορφοσυντακτική**

σε επίπεδο λεκτικού τύπου (και ανά μέρος του λόγου) προκύπτουν τα: γένος, πτώση, φωνή, χρόνος, πρόσωπο, βαθμός, έγκλιση, κλιτό – άκλιτο, ισχυρός – ασθενής τύπος.

Είδη κωδικοποιημένης πληροφορίας

- **β. μορφηματική**

Κάθε λεκτικός τύπος υφίσταται μορφηματική κατάτμηση σε: πρόθημα, επίθημα, κατάληξη

- **γ. συλλαβισμού**

Για κάθε λεκτικό τύπο υπάρχει ορθός συλλαβισμός. Πιο συγκεκριμένα περιγράφεται ορθά το φαινόμενο της συνίζησης ([rotírja]) και αντιδιαστέλλονται συνιζημένοι και ασυνιζητοι τύποι ([rotí-rja] / [mistí-ri-a]). Συγκεκριμένα, εξετάζονται περιπτώσεις όπου το μορφολογικό (και ιστορικό) κριτήριο αποφασίζει για την επιλογή του συνιζημένου ή του ασυνιζητου τύπου. Στο λήμμα {δικαστήριο} στον πληθυντικό ο τύπος συλλαβίζεται χωρίς συνίζηση [δι-κα-στη-ρι-α], ενώ στο λήμμα {ποτήρι} στον πληθυντικό ο τύπος συλλαβίζεται με συνίζηση [πο-τη-ρια].

Είδη κωδικοποιημένης πληροφορίας

- **δ. υφολογική (επίπεδο ύφους)**

Η υφολογική κωδικοποίηση περιγράφεται για λεκτικούς τύπους με διαφορετικά της κοινής νέας ελληνικής γλώσσας υφολογικά χαρακτηριστικά. Σε πολλές περιπτώσεις η πληροφορία αυτή είναι ενσωματωμένη μέσα στο κλιτικό σύστημα. Τα υφολογικά χαρακτηριστικά που περιγράφονται είναι: *προφορικό, αργκό, αρχαιοπρεπής, διαλεκτικός, επίσημος- ανεπίσημος, λόγιος, παρωχημένος, σπάνιος, λαϊκότροπο.*

Είδη κωδικοποιημένης πληροφορίας

- **ορολογική (θεματικά πεδία)**

Σε επίπεδο λήμματος ορίζεται επιπλέον πληροφορία για το αν ο τύπος αποτελεί μέρος ειδικού λεξιλογίου: λογιστική, γεωπονική, ανθρωπολογία, αρχαιολογία, αρχιτεκτονική, αστρολογία, αστρονομία, βιολογία-γενετική, βοτανική, διοίκηση επιχειρήσεων, χημεία, κινηματογράφος, μαγειρική-ζαχαροπλαστική, κοσμογραφία, χορός, διακοσμητικές τέχνες, μουσική, εκκλησιαστικοί όροι, οικολογία, οικονομία, ηλεκτρολογία-ηλεκτρονική, εθνολογία, καλές τέχνες, λαογραφία, γεωγραφία, γεωλογία, γραφιστική, ιστορία, πληροφορική, δημοσιογραφία, νομικοί όροι, γλωσσολογία, λογοτεχνία, μαθηματικά, μηχανολογία, ιατρικοί όροι, μετεωρολογία, στρατιωτικοί όροι, ορυκτολογία, μυθολογία, ναυτικοί όροι, ζωγραφική, παιδαγωγική, φιλολογία, φιλοσοφία, φωτογραφία, φυσική, πολιτικοί όροι, ψυχολογία, ρητορική, γλυπτική, κοινωνιολογία, αθλητισμός, τεχνολογία, θέατρο, θεολογία – θρησκευσιολογία, τυπογραφία, ζωολογία

Είδη κωδικοποιημένης πληροφορίας

- **στ. κλίσης**
- **ζ. τονισμού**
- **η. εναλλακτικοί, παράλληλοι ή παράγωγοι τύποι**

Στο επίπεδο αυτό περιγράφονται τα παράγωγα (υποκοριστικά/μεγεθυντικά), παραθετικά επιθέτων, κτλ.

Μοντέλο ανάπτυξης

- Το σύστημα βασίζεται σε προεπιλεγμένες λίστες χαρακτηριστικών και κανόνων με συγκεκριμένο αριθμό τιμών (οι λίστες μπορούν εύκολα να εμπλουτιστούν και να ξαναφορτωθούν στο σύστημα). Η 'αρχιτεκτονική' του λεξικού κτίζεται σε ένα μοντέλο 5 επιπέδων

1^ο Επίπεδο

- Στο πρώτο στάδιο κωδικοποίησης της πληροφορίας ορίζονται πρώτες λίστες: τα 67 μοναδικά attributes (Part-of-Speech & morphological attributes).
- Οποιαδήποτε τιμή δοθεί εδώ θα κληρονομηθεί αυτόματα και σε όλους τους τύπους που θα αναπτυχθούν παρακάτω για το λήμμα αυτό π.χ. (NOUN: ουσιαστικό).
- Στη συνέχεια ορίζονται 79 tag attributes (domain attributes & style attributes) που προσδιορίζουν επιμέρους και πιο εξειδικευμένη πληροφορία ενός λήμματος και αναφέρονται είτε στο κατά περίπτωση θεματικό πεδίο, π.χ. (tag_attribute {ARCHEOL}{"Αρχαιολογία"}), είτε σε υφολογική πληροφορία π.χ. (tag_attribute {ORAL}{"Προφορικό"}) είτε σε χρήση π.χ. (tag_attribute {RARE}{"Σπάνιο"})

1^ο Επίπεδο

- Στο μορφολογικό λεξικό υπάρχουν και tags που προσδιορίζουν επιπλέον πληροφορία για ξένες λέξεις, συνώνυμα, ομόηχα, λήμματα με εννοιολογική σύγχυση ή ομόηχα: Ο ορισμός αυτός είναι της μορφής:
`(\tag{SIGXISIENNOIA}{\"Εννοιολογική σύγχυση\"},`
`\tag{SYNONYMS}{\"Συνώνυμα\"}.`

1^ο Επίπεδο

Lemma's Definition ΣΣ

Lemma Info | Morphology | Word-forms | Tags | Meanings

Lexeme:

Description:

Attributes:

Tag Attributes:

Preview first lemma's definition

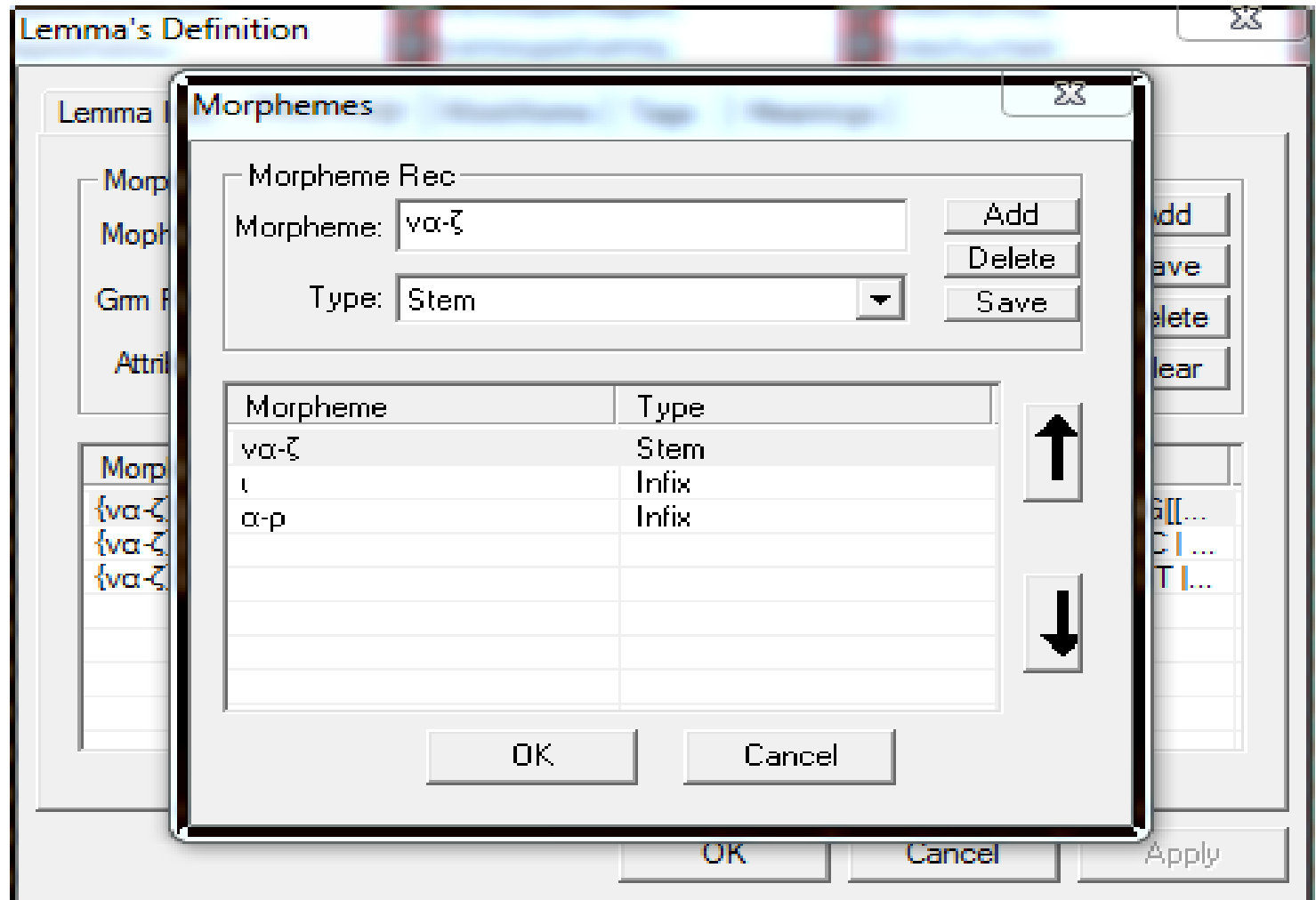
2^ο Επίπεδο

- Σε δεύτερο επίπεδο το κτίσιμο του λεξικογραφικού editor περιλαμβάνει κατόπιν συνδυασμό 77 μορφολογικών χαρακτηριστικών: `\attribute_sets`: π.χ. `attribute_set {ACC_PLUR} {"Αιτιατική, πληθυντικός"}`, ή `attribute_set {B_P_ORAL} {"β' πρόσωπο, προφορικός λόγος"}`, για καθένα από το οποίο ορίζονται αναλυτικά τα μορφολογικά χαρακτηριστικά του. Για παράδειγμα στο `attribute_set {ACC_PLUR}` περιλαμβάνονται τα μοναδικά attributes (`ACC`=αιτιατική και `PLUR` = πληθυντικός). Τα μορφολογικά αυτά χαρακτηριστικά θα περιγράψουν αργότερα ευρύτερες ομάδες λέξεων (όλες τις λέξεις που έχουν μορφολογικό χαρακτηριστικό αιτιατική και πληθυντικός).

3^ο επίπεδο

- Σε μετέπειτα επίπεδο ορίζεται η θέση του τόνου του λήμματος και οι καταλήξεις (suffixes). Υπάρχουν 19 stress rules και 191 κατηγορίες suffix rules.
- Οι κανόνες τονισμού περιγράφουν τη θέση του τόνου και την πιθανή μετακίνησή του κατά την κλιτική παραγωγή όπως επίσης και την απουσία τόνου.
- Έτσι: α) π.χ. για άτονες λέξεις όπως το λήμμα {γη}, έχουμε τον κανόνα `stress_rule{NO_STR}{\"άτονες λέξεις\"}`, β) για την ομάδα λέξεων που ο τόνος μετακινείται από την προπαραλήγουσα στην παραλήγουσα και στη συνέχεια ξανά στην προπαραλήγουσα (άν-θρω-πος, αν-θρώ-που, αν-θρω-πο) υπάρχει ο κανόνας `stress_rule{UNTER_PEN}{\"προπαραλήγουσα, παραλήγουσα, προπαραλήγουσα\"}` στον οποίο υπάγονται οι επιμέρους θέσεις τονισμού που περιγράφηκαν

Ορισμός μορφημάτων



4^ο επίπεδο

- Σε αυτό το επίπεδο ορίζονται και οι καταλήξεις (suffixes). Για παράδειγμα το suffix{α} είναι η κατάληξη -α με μορφολογικά χαρακτηριστικά (attribute names): α' προσώπου, ενικού, ρημάτων. Στο ίδιο επίπεδο περιγράφονται και ευρύτεροι κανόνες (suffixes rules) που είναι ομαδοποίηση πολλών καταλήξεων για να περιγραφεί μια πιο ευρεία κλητική παραγωγή.

Παράδειγμα κωδικοποίησης

- Για παράδειγμα για τον παρελθοντικούς χρόνους συμφωνόληκτων ρημάτων ενεργητικής φωνής με αύξηση ορίζεται ο κανόνας `suffix_rule {ACT_AORa}` στον οποίο περιγράφονται το α' πρόσωπο ενικού με suffix `{-α}`, το β' πρόσωπο ενικού με suffix `{-ες}`, το γ' πρόσωπο ενικού με suffix `{-ε}`, το γ' πρόσωπο πληθυντικού με suffix `{-αν}` μαζί με τα ανάλογα μορφολογικά χαρακτηριστικά τους (attribute names).
- Η περιγραφή των suffixes μπορεί να προσδιορίσει και δισύλλαβες ή και υπερδισύλλαβες καταλήξεις (για ρηματικού τύπους **κυρίως**), γεγονός που επιτρέπει στην ορθότερη αναζήτηση των θεμάτων των λημμάτων.

```
\suffix_rule{GEN_EWS}{"γενική -εως (πόλεως, βασιλέως)"}
```

```
\suffix{ε-ως}
```

```
{\attribute_name{GEN}{0}
```

```
\attribute_name{SING}{o} }
```

- Σε περιπτώσεις που τύπος με ίδια μορφολογικά – γραμματικά χαρακτηριστικά διαφοροποιείται στην κατάληξη λόγω διαφορετικής υφολογικής πληροφόρησης (λόγια μορφή, διαλεκτικός τύπος), στον ίδιο suffix rule περιλαμβάνονται και οι 2 τύποι.
- Για την ακρίβεια τόσο για τους τόνους όσο και για τις καταλήξεις υπάρχουν προεπιλεγμένες λίστες. Έτσι το σύστημα προσφέρει δύο επιλογές: α) την **αυτόματη επιλογή** από τις λίστες (μέσω των κλιτικών και των συνδυασμών) και β) τη **χειρωνακτική** (custom κανόνες) που ο λεξικογράφος ορίζει το δικό του συνδυασμό για λεξικούς τύπους που μέχρι τώρα δεν υπήρχαν και η εξέλιξη της γλώσσας – ως ζωντανός οργανισμός – δημιουργήσει.

5^ο επίπεδο

- Με δεδομένη την ύπαρξη: α) της μορφολογικής πληροφόρησης (attributes), β) του ορισμού του τόνου (stress_value) και γ) των δυνατών κλιτικών μορφημάτων (suffix_values) ορίζονται στο τελευταίο επίπεδο **οι 306 γραμματικοί κανόνες** (grammar_rules) για την παραγωγή ομαδοποιημένων κατηγοριών λημμάτων ανάλογων με αυτούς που ορίζονται στην παραδοσιακή γραμματική (π.χ. αρσενικά προπαροξύτονα σε -ος, ενεστώτας βαρύτονων ρημάτων κλπ.)

Παράδειγμα δήλωσης

- Έτσι για την παραγωγή των οξύτονων θηλυκών επιθέτων σε -ή η περιγραφή έχει την ανάλογη μορφή

```
\grammar_rule{ADJ_FEM_HI} {\attribute_name{ADJ}{0} }  
    {\grform  
    {\suffix_value{FEM_H_SING}}}  
    {\stress_value{ULTI}}}  
    \attributes  
    {\attribute_name{INFORMAL}{0}} }  
    {\grform  
    {\suffix_value{PLUR_ESWN}}}  
    {\stress_value{ULTI}}}  
    \attributes  
    {\attribute_name{FEM}{0}}
```

Παράδειγμα δήλωσης

- Η περιγραφή των suffixes μπορεί να προσδιορίσει και δισύλλαβες ή και υπερδισύλλαβες καταλήξεις (για ρηματικούς -κυρίως-τύπους).

```
\suffix_rule{GEN_EWS}{«γενική -εως  
(πόλεως, βασιλέως)» }
```

```
\suffix{ε-ως}
```

```
{\attribute_name{GEN}{0}
```

```
\attribute_name{SING}{o} }
```

Ορισμός καταλήξεων

Choose/Create Grammar Rule

Grammar Form Record

GR Forms:

Attributes: !

Grammar Rules:

Γραμματική Φόρμα	Ιδιότητες Γραμματικής Φόρμας
ADJ_MASC_FEM_HSA_SING{MAS...	ADJ_MASC_FEM_HSA_SING(ADJ)

Κανόνες Γραμματικών Φο... | Περιγραφή Κανόνα Γραμματικών Φορμών

ADJ_COMP	επίθετα,συγκριτικός, σε -τερος, -η, -ο προπαροξύτονα (ο
ADJ_FEM_A1	θηλ. -ά για οξύτονα επίθ. -ος, -ιά, -ό κ' -ύς, -ιά, -ύ (ξανθιά,
ADJ_FEM_A1_SING	ενικός θηλ. -ιά για επίθ. -ος, -η κ' -ιά, -ό (κακός, -ή κ' -ιά, -έ
ADJ_FEM_A2a	θηλ. -α για παροξύτονα επίθ. -ύς, -εία, -ύ (οξεία)
ADJ_FEM_A2b	θηλ. επίθ. παροξ. -α που δεν κατεβάζουν τον τόνο στη γ
ADJ_FEM_A3_LEARNED	λόγια παροξ. τύποι προπαρ. θηλ. επίθ. -α,κυρίως για παρ.
ADJ_FEM_A4	οξύτονα θηλ. επίθ. -ά (σιδηρά)
ADJ_FEM_H1	οξύτονα θηλ. επίθ. -ή (ενεργή)
ADJ_FEM_H3_LEARNED	λόγια τύποι επίθ. θηλ. που κατεβάζουν τον τόνο (βεβαιο
ADJ_FEM_OS1_LEARNED	λόγια οξ. θηλ. επίθ. -ός (η ενεργός)
ADJ_FEM_OS2_LEARNED	λόγια παροξ. θηλ. επίθ. -ος (η νόθος)
ADJ_FEM_OS3_LEARNED	λόγια προπαροξ. θηλ. επίθ. -ος (η εμπύρετος)

OK Cancel

Παραγωγή κλιτών τύπων

- Με δεδομένη την περιγραφή όλων των γραμματικών κανόνων (grammar rules) στον λεξιλογικό editor η μορφή των λημμάτων έχει ως εξής:

Lexeme [μαγαζί]

{μα-γα-ζ} {{N_NEUT_I1{NEUT_I_SING[[ι(ACC | NEUT | NOM | SING | VOC) | ιου(GEN | NEUT | SING)] ()]ULT1<<Λ(1)()>()>() || NEUT_I_PLUR[[ια(ACC | NEUT | NOM | PLUR | VOC) | ιων(GEN | NEUT | PLUR)] ()]ULT1<<Λ(1)()>()>()}}(N)}}.

{μα-γα-ζ}[α-κ] {{N_NEUT_I2_NOGEN{NEUT_IIA[[ι(ACC | NEUT | NOM | SING | VOC) | ια(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 Infix. Κάθε λήμμα δηλαδή περιλαμβάνει όλα τα θέματα που ανήκουν σημασιολογικά στην ίδια κατηγορία και χαρακτηρίζει τις παραγωγικές καταλήξεις τους

Παραγωγή κλιτών τύπων

- Η παραγωγή των λέξεων έχει την αντίστοιχη μορφή:



Word forms	Morphology	Form	Attribute
μα-γα-ζί	{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-ζιάκι	{μα-γα-ζ}[α-κ]	[[ι(ACC NEUT NOM SING VOC..	N NEUT DIM
μα-γα-ζιάκια	{μα-γα-ζ}[α-κ]	ια(ACC NEUT NOM PLUR VOC)..	N NEUT DIM



Ηλεκτρονικό λεξικό

alpha.nlx - Neuro Lexicon Editor

File Edit Morphos Lemma Dictionary View Help

Sections

- άλφα
- βήτα
- γάμα
- δέλτα
- έψιλον
- ζήτα
- ήτα
- θήτα
- γιώτα
- κάπα
- λάμδα
- μι
- νι
- ξι
- όμικρον
- πι
- ρο
- σίγμα
- ταυ
- ύψιλον
- φι

α_ΕΠΦ	αβανγκαρντιστής	αβαράρω	αβάσκαντο	αβγατίζω
α_ΟΥΣ	αβανεύω	αβαρέλιαστος	αβάσκαντος	αβγατίσμα
αβαειο	αβάνης	αβαρεσιά	αβασκαντούρι	αβγάτος
αβάθεια	αβανιά	αβάρετος_1	αβάσταγα	αβγίλα
αβαθή	αβανιάζω	αβάρετος_2	αβασταγή	αβγό
αβαθής	αβανιάρης	αβαρής	αβασταγιά	αβγοβολώ
αβαθμιδίωτα	αβάνς	αβαρία	αβασταγό	αβγοειδής
αβαθμιδίωτος	αβάντα	αβαριάτος	αβάσταγος	αβγοειδώς
αβαθμολόγητος	αβανταδόρκος	αβαρος	αβασταζιά	αβγοθήκη
άβαθος	αβανταδόρος	αβάς	αβάσταχτα	αβγοκαλάμαρο
αβαθύλωτος	αβαντάζ	αβασάνιστα	αβάσταχτος	αβγοκάσα
αβαθύρριζος	αβαντάρω	αβασάνιστος	αβάτευτος	αβγοκόβω
αβαθώς	αβάντε_ΕΠΡ	αβασία	άβατο	αβγοκούλουρα
αβάκα	αβάντε_ΟΥΣ	αβασίλευτος_1	άβατος	αβγοκούλουρο
άβακας	αβάντζα	αβασίλευτος_2	αβατσίνωτος	αβγολέμονο
αβάκιο	αβαντζάρω	αβάσιμα	αβαυκάλιστα	αβγολόγος
αβακοειδής	αβάντζο	αβάσιμος	αβαυκάλιστος	αβγολογώ
αβακωτός	αβάντι_ΕΠΡ	αβασιμότητα	άβαφος	αβγομαντεία
αβαλσάμωτος	αβάντι_ΟΥΣ	αβάσιστα	αβάφτιστος	αβγόπιτα
αβαλτος	αβάπτιστος	αβάσιστος	άβαφτος	αβγοπώλης
αβανγκάρντ_ΕΠΘ	άβαπτος	αβασκαίνω	άβγαλτα	αβγοσαλάτα
αβανγκάρντ_ΟΥΣ	αβάρα_ΕΠΦ	αβάσκαμα	άβγαλτος	αβγόσουπα
αβανγκαρντισμός	αβάρα_ΟΥΣ	αβασκανία	αβγαταίνω	αβγόσχημος

Ready NUM

Lex Edit

lemma's Definition ✕

Lemma Info | Morphology | Word-forms | Tags | Meanings

Lexeme: άβατος

Description: -ος, -η, -ο όχι επίρρ. Κ θρησκ. όλοι λήμμα

Attributes: ADJ

Tag Attributes: ECCLES

Preview first lemma's definition

Modify Modify

OK Άκυρο Εφαρμογή

Lex Edit

Lemma's Definition

ΣΣ

Word	Morphology	Form	Attributes
ά-βα-του	<α->{βα-}[τ]	SFX_OUWN...	(ADJ GEN MASC NO_CO...
ά-βα-των	<α->{βα-}[τ]	SFX_OUWN...	(ADJ GEN MASC NO_CO...
ά-βα-τους	<α->{βα-}[τ]	SFX_OUWN...	(ACC ADJ MASC NO_CO...
ά-βα-τος	<α->{βα-}[τ]	SFX_OS3[[ο...	(ADJ MASC [NOM_SING] ...
ά-βα-το	<α->{βα-}[τ]	SFX_OS3[[ο...	([ACC_SING] ADJ MASC ...
ά-βα-τε	<α->{βα-}[τ]	SFX_OS3[[ο...	(ADJ MASC NO_COMP_SU...
ά-βα-τοι	<α->{βα-}[τ]	SFX_OS3[[ο...	(ADJ MASC NOM NO_CO...
ά-βα-τη	<α->{βα-}[τ]	FEM_H_SIN...	(ACC ADJ FEM NOM NO...
ά-βα-της	<α->{βα-}[τ]	FEM_H_SIN...	(ADJ FEM GEN NO_COM...
ά-βα-τες	<α->{βα-}[τ]	PLUR_ESW...	(ACC ADJ FEM NOM NO...
ά-βα-των	<α->{βα-}[τ]	PLUR_ESW...	(ADJ FEM GEN NO_COM...
ά-βα-το	<α->{βα-}[τ]	NEUT_O_SI...	(ACC ADJ NEUT NOM N...
ά-βα-του	<α->{βα-}[τ]	NEUT_O_SI...	(ADJ GEN NEUT NO_CO...
ά-βα-τα	<α->{βα-}[τ]	NEUT_O_P...	(ACC ADJ NEUT NOM N...
ά-βα-των	<α->{βα-}[τ]	NEUT_O_P...	(ADJ GEN NEUT NO_CO...

OK Ακυρο Εφαρμογή

Lex Edit

Morphemes

26

Morpheme Rec

Morpheme:

Type:

Add

Delete

Save

Morpheme	Type
α-	Prefix
βα-	Stem
τ	Infix



OK

Cancel

Lex Edit

Choose/Create Grammar Rule

ΣΣ

Grammar Form Record

GR Forms: ADJ_OSHO3(SFX_OUWNOUS[[ου(GEN | SING) | ων(GEN | PLUR) | ους(AI

Attributes: ADJ_OSHO3(ADJ)

Grammar Rules:

Γραμματική Φόρμα	Ιδιότητες Γραμματικής Φόρμας
ADJ_OSHO3(SFX_OUWNOUS[[ου(...	ADJ_OSHO3(ADJ)

Κανόνας Γραμματικών Φο...	Περιγραφή Κανόνα Γραμματικών Φορμών
ADJ_COMP	επίθετα, συγκριτικός, σε -τερος, -η, -ο προπαροξύτονα (ο
ADJ_FEM_A1	θηλ. -ά για οξύτονα επίθ. -ος, -ιά, -ό κ' -ύς, -ιά, -ύ (ξανθιά,
ADJ_FEM_A1_SING	ενικός θηλ. -ιά για επίθ. -ος, -η κ' -ιά, -ό (κακός, -ή κ' -ιά, -έ
ADJ_FEM_A2a	θηλ. -α για παροξύτονα επίθ. -ύς, -εία, -ύ (οξεία)
ADJ_FEM_A2b	θηλ. επίθ. παροξ. -α που δεν κατεβάζουν τον τόνο στη γ
ADJ_FEM_A3_LEARNED	λόγιοι παροξ. τύποι προπαρ. θηλ. επίθ. -α, κυρίως για παρ.
ADJ_FEM_A4	οξύτονα θηλ. επίθ. -ά (σιδηρά)
ADJ_FEM_H1	οξύτονα θηλ. επίθ. -ή (ενεργή)
ADJ_FEM_H3_LEARNED	λόγιοι τύποι επίθ. θηλ. που κατεβάζουν τον τόνο (βεβαιο
ADJ_FEM_OS1_LEARNED	λόγια οξ. θηλ. επίθ. -ός (η ενεργός)
ADJ_FEM_OS2_LEARNED	λόγια παροξ. θηλ. επίθ. -ος (η νόθος)
ADJ_FEM_OS3_LEARNED	λόγια προπαροξ. θηλ. επίθ. -ος (η εμπύρετος)

OK

Cancel

Lex Edit

Choose/Create Stress Rule

Stress Record

Stresses: UNTEP1<<Πη(1)()>()>

Attributes: ()

Κανόνας Τονισμού	Περιγραφή Κανόνα Τονισμού
NO_STR	άτονες λέξεις (γη)
NO_STR_NO_STR_...	άτονο, άτονο, παραλήγουσα
NO_STR_PEN	άτονο, παραλήγουσα
PEN1	τόνος στο τελευταίο φωνήεν της παραλήγουσας
PEN2	τόνος στο προτελευταίο φωνήεν της παραλήγουσας
PEN_NO_STR	παραλήγουσα, άτονο
PEN_PEN_UNTEP	παραλήγουσα, παραλήγουσα, προπαραλήγουσα
PEN_ULT	παραλήγουσα, λέγουσα
PEN_UNTEP	παραλήγουσα, προπαραλήγουσα
ULT1	τόνος στο 1ο απ' το τέλος φωνήεν της λέξης
ULT2	τόνος στο προτελευταίο φωνήεν λέγουσας

OK

Cancel

Lex Edit

Suffixes ✕

Suffix:

Attributes:

Κατάληξη	Ιδιότητες
ου	(GEN SING)
ων	(GEN PLUR)
ους	(ACC PLUR)

Σημασία Λεξικού

- Η σπουδαιότητα του λεξικού εντοπίζεται στη δυνατότητα να υποστηρίξει επιπλέον εφαρμογές.
- Αποτελεί επαναχρησιμοποιούμενο εργαλείο, βασική γλωσσολογική τεχνολογία και υποδομή για την ελληνική γλώσσα για την οποία δεν έχουν γίνει σπουδαία βήματα.
- Απώτερος σκοπός είναι το λεξικό αυτό να λειτουργήσει είτε ανεξάρτητα είτε σε συνδυασμό με άλλα υπολογιστικά εργαλεία και να χειριστούμε την ελληνική γλώσσα υπολογιστικά -αυτόματα- σε όλες τις διαστάσεις της.

Σπουδαιότητα του λεξικού

- Το λεξικό αυτό αποτελεί και τον θεμέλιο λίθο για την υποστήριξη πολλών εφαρμογών (όλων;::::), όπως του γραμματικού διορθωτή (Grammar Checker) που ελέγχει και διορθώνει περιπτώσεις συντακτικής ασυμφωνίας, περιπτώσεις εννοιολογικής σύγχυσης, ζητήματα κλίσης, ζητήματα σύνταξης και γενικά περιπτώσεις λαθών που απαιτούν πιο εξειδικευμένη διαχείριση από αυτή που προσφέρει ο ορθογραφικός διορθωτής.
- Η αξία και ο ρόλος του μορφολογικού λεξικού για την υποστήριξη ενός γραμματικού διορθωτή είναι αυτονόητη, καθώς η μορφολογία είναι το πρώτο επίπεδο γλώσσας που εξετάζεται και το συντακτικό επίπεδο βασίζεται και εξαρτάται από τη μορφολογία των λέξεων



Αυτόματη επεξεργασία λέξεων/λεξικών μονάδων

Αυτόματη επεξεργασία

- Για να επεξεργαστούμε γραπτά κείμενα στις φυσικές γλώσσες, πρέπει να **μοντελοποιήσουμε φαινόμενα** (σε **τυπογραφικό** και **μορφολογικό** επίπεδο).
- Κατά τη διάρκεια της **τυπογραφικής ανάλυσης**, πρέπει να **οριοθετήσουμε τη λειτουργία κάθε χαρακτήρα** μέσα στην πρόταση (για παράδειγμα, η τελεία μπορεί να χρησιμοποιηθεί ως σημείο στίξης, ως διαχωριστικό αριθμών ή συντομογραφιών, αρκτικολέξων κτλ.). Επειδή οι τυπογραφικοί κανόνες κάθε γλώσσας είναι διαφορετικοί, η τυπογραφική ανάλυση κάθε γλώσσας είναι διαφορετική
- Έπειτα, πρέπει να ορίσουμε τις ελάχιστες μονάδες που θα επεξεργαστούμε, δηλαδή τις λεξικές μονάδες του κειμένου (απλές ή σύνθετες), και να τις μετατρέψουμε στον λημματικό τους τύπο. Αυτός είναι ο ρόλος της **μορφολογικής ανάλυσης**

Μορφολογική ανάλυση

- Αυτή η επεξεργασία σε λεξικό επίπεδο μπορεί να πραγματοποιηθεί και μέσω της χρήσης ενός **υπολογιστικού λεξικού**.
- **Απλή λέξη** είναι «**μια ακολουθία χαρακτήρων χωρίς διαχωριστικό**», ενώ
- **Σύνθετη λέξη** ορίζεται **κάθε ακολουθία λέξεων που εμφανίζει κάποιο βαθμό παγίωσης** και περιγράφει μία σημασιολογική μονάδα στη γενική γλώσσα ή στα ειδικά λεξιλόγια, π.χ. *μαύρη αγορά, κρέμα γάλακτος*.
- Στο μορφολογικό επίπεδο πρέπει επίσης κάθε λεξική μονάδα να ενταχθεί σε μια γραμματική κατηγορία, λ.χ. ουσιαστικό (κοινό/προσηγορικό ή κύριο, **όπου ανήκουν οι ονοματικές οντότητες**) και να ενταχθεί σε μια κλιτική τάξη, εφόσον ανήκει στα κλιτά στοιχεία (λ.χ. ουσιαστικά, επίθετα, ρήματα).

Ορισμός λέξης/ λεξικής μονάδας (γλωσσολογία)

- Σε μια **γλωσσολογική περιγραφή**, ελάχιστη σημασιολογική μονάδα θεωρείται το μόρφημα, το οποίο μπορεί να ταυτίζεται με τη λέξη ως προς τα όριά του π.χ. *μπαρ* (μία λέξη, ένα μόρφημα), αλλά στη νέα ελληνική (ΝΕ) συνήθως αποτελεί τμήμα της, π.χ. *διαγράφω* (μία λέξη, τρία τουλάχιστον μορφήματα)

Ορισμός λέξης/ λεξικής μονάδας (ηλεκτρονική περιγραφή)

- Για μια **ηλεκτρονική περιγραφή** όμως, οι λέξεις είναι οι ελάχιστες σημασιολογικές μονάδες και απαντούν στα κείμενα με πεζά ή/και κεφαλαία γράμματα και με ελληνικούς ή/και λατινικούς χαρακτήρες
- Επιπλέον εμφανίζονται και κάποια σύμβολα που είναι αναγκαίο να εντοπίζονται, εφόσον έχουν συγκεκριμένη γλωσσική σημασία και κάποτε χρησιμοποιούνται για να αναπαραστήσουν συντομευμένες μορφές λέξεων, π.χ. % 'τοις εκατό', \$ 'δολλάριο', € 'ευρώ'

Ορισμός λέξης/ λεξικής μονάδας (ηλεκτρονική περιγραφή)

- Τέλος, οι λέξεις συνήθως δεν εμφανίζονται στα κείμενα στον λημματικό τους τύπο αλλά σε έναν από τους **κλιτικούς τύπους**
- Ο εντοπισμός των λέξεων/ λεξικών μονάδων ενός κειμένου προϋποθέτει τη **λημματοποίησή τους**, δηλαδή την αναγωγή της καθεμιάς στον αντίστοιχο **λημματικό τύπο**
- Στη συνέχεια απαιτείται κάθε κλιτικός τύπος να συνοδεύεται από τις συγκεκριμένες κλιτικές πληροφορίες που τον χαρακτηρίζουν

Μορφολογία (Γλωσσολογία)

- Μια λεξική μονάδα μπορεί να είναι **απλή** (simple) π.χ. νερό, ή **μη απλή**, δηλ. με εσωτερική δομή, π.χ. νερ-άκι, νερ-ουλός, νερ-ό-μυλος.
- Οι μη απλές λεξικές μονάδες μπορεί να είναι είτε **παράγωγες** (derived), δηλ. **προθηματοποιημένες** (prefixed) π.χ. **κατα**-γράφω, ή **επιθηματοποιημένες** (suffixed) π.χ. γραπ-**τός**) είτε **σύνθετες** (compound), π.χ. **γραφ**-ο-μηχανή.

Μορφολογία (Υπολογιστική Γλωσσολογία)

- Υπολογιστική Γλωσσολογία: είναι δυνατό απλή λεξική μονάδα να θεωρείται αυτή που καταγράφεται **μονολεκτικά**, ανεξάρτητα από το αν είναι εσωτερικά δομημένα, π.χ. νερό, νεράκι, νερόμυλος, ενώ σύνθετη να θεωρείται η **πολυλεκτική λεξική μονάδα**, δηλ. που αποτελείται από δύο ή περισσότερες λέξεις, π.χ. βαρύ ύδωρ, παιδική χαρά, φακοί επαφής, νόμος-πλαίσιο

Ορισμός απλής λέξης

- απλή λέξη είναι «μια ακολουθία χαρακτήρων χωρίς διαχωριστικό» (Silberztein, 1993): εντάσσονται λέξεις όπως βιβλίο, καλός, παίζω, αλλά και **δάνειες λέξεις** συμμορφωμένες ή μη στο μορφολογικό σύστημα της ΝΕ
- απλές θεωρούνται **και οι παράγωγες** λέξεις, π.χ. σκουπ-ίζω, καθώς και τα μονολεκτικά **σύνθετα**, που σε άλλα θεωρητικά πλαίσια θεωρούνται **κατασκευασμένες** λέξεις, π.χ. σιδηρ-ό-δρομος.

Ορισμός απλής λέξης

- **ΠΡΟΒΛΗΜΑ**: αυτός ο τυπογραφικού τύπου ορισμός του Silberztein δεν επιτρέπει την αναγνώριση όλων των λέξεων. Για παράδειγμα, δύσκολα εντάσσονται στον ορισμό αυτό οι λέξεις που υπόκεινται στα φαινόμενα της **έκθλιψης**, της **αφαίρεσης** και της **αποκοπής**, όπως τ' (τ' άστρα), θ' (θ' αγοράσω) – 'ρθει (θα 'ρθει) – φέρ' (φέρ' το), απ' (απ' το πρωί), **τις** **οποίες πρέπει να μετατρέψουμε στη μορφή που έχουν στην υποκείμενη δομή** (το, θα, έρθει, φέρε, από αντίστοιχα). Επίσης ο παραπάνω ορισμός **δεν καλύπτει** περιπτώσεις όπως των εμπρόθετων άρθρων π.χ. στο(ν), στους, στις

Παραγωγή και σύνθεση

- Ο όρος derivation (παραγωγή) είναι αμφίσημος, αφού **μπορεί να αναφέρεται είτε γενικά στις διαδικασίες κατασκευής των λεξικών μονάδων, λ.χ. να περιλαμβάνει και τη σύνθεση, είτε ειδικά μόνο στη διαδικασία της προσφυματοποίησης (affixation).**
- Προσφυματοποίηση: εισαγωγή προθέματος, κατάληξης, ενθέματος
 1. Πρόθεμα (prefix): **ξε**-πλένω
 2. Ένθεμα (infix): δι-**έ**-σχισα
 3. Κατάληξη (suffix): κατεργαρ-**άκος**

Πώς κατασκευάζονται οι λέξεις;

- Στη ΝΕ, όπως και σε πολλές άλλες γλώσσες, υπάρχουν τέσσερα διαφορετικά είδη **κατασκευής των λέξεων**:
 1. η **προθηματοποίηση** (prefixation) και
 2. η **επιθηματοποίηση** (suffixation), που συγκροτούν την **παραγωγή** (derivation),
 3. η **σύνθεση** (compounding) και
 4. η **μετατροπή ή μετάπλαση/μεταπλάσμος** (conversion), δηλ. η αλλαγή γραμματικής κατηγορίας χωρίς την παρέμβαση επιθήματος, π.χ. μέτρο → μετρώ, αρθρογράφος → αρθρογραφώ, νομικός (επίθετο) → νομικός (επιστήμων) και νομική (επιστήμη)

Πώς κατασκευάζονται οι λέξεις;

- οι βασικοί μηχανισμοί, όμως, εμπλουτισμού του λεξιλογίου είναι η **παραγωγή** και η **σύνθεση**
- Όλα τα είδη των λεξικών μονάδων που παράγονται με τις παραπάνω μορφολογικές διαδικασίες θα θεωρηθούν απλές λεξικές μονάδες

Παραγωγή

- Ορίζεται η διαδικασία κατασκευής λέξεων συνήθως με την εφαρμογή **προσφύματος** (ή **παραθήματος**) (affix) πάνω σε **λεξικό μόρφημα** ή **λέξημα**
- Τα προσφύματα, που αποτελούν **μη ελεύθερα γραμματικά μορφήματα**, μπορεί να προηγούνται της βάσης, οπότε ονομάζονται **προθήματα** (prefix) ή να έπονται, οπότε ονομάζονται **επιθήματα** (suffix).
- Η παραγωγή (προθηματοποίηση και επιθηματοποίηση) μπορεί να δημιουργήσει λέξεις που είτε **διατηρούν την ίδια γραμματική κατηγορία με τη βάση** (ενδοκεντρική δομή) είτε **δεν τη διατηρούν** (εξωκεντρική δομή): π.χ. **ντύνω** → **ξεντύνω** (ρήμα και στις δυο περιπτώσεις), **παιδί** → **παιδάκι**, **μήλο** → **μηλιά** (ουσιαστικό και στις δύο περιπτώσεις), **αλλά** **κλήρος** → **άκληρος**, **καιρός** → **πρόσκαιρος**, **μάρμαρο** → **μαρμάρινος** (επίθετο που παράγεται από ουσιαστικό)

Παραγωγή

- Η **σημασία** της παράγωγης λέξης μπορεί να είναι **συνθετική** (compositional), δηλ. να ισούται με το **άθροισμα** των σημασιών των στοιχείων που συμμετέχουν, π.χ. μαρμάρinos `από μάρμαρο` μπορεί όμως να είναι ποικίλου βαθμού **συνθετικότητας** έως και **απρόβλεπτη**, π.χ. συγγράφω (ενώ είναι ένας ο δημιουργός του κειμένου).

Προθήματα

- Πρόκειται για κατασκευή λέξης με τη βοήθεια των προθημάτων, γραμματικών μορφημάτων που τίθενται μπροστά από τη βάση
- Το πρόθημα και η βάση **συνενώνονται** (αντινομάρχης) **ή ενώνονται** με μία παύλα (αντικαρκινικός)
- Τα προθήματα που ενώνονται με ρήματα προκαλούν προβλήματα τόσο στο επίπεδο της σύνταξης όσο και στο επίπεδο της μορφολογίας (κλίσης, π.χ. αύξηση). Ειδικά για τη ΝΕ τα προθήματα προέρχονται από τις προθέσεις της ΑΕ, π.χ. ανα-, αντι-, αμφι-, απο-, δια-, εισ-, εν-, εκ-, επι-, κατα-, μετα-, παρα-, περι-, προσ-, συν-, υπερ-, υπο-
- Η σημασία της προθηματοποιημένης λέξης **δεν** είναι πάντα **διαφανής**

Προθηματοποίηση και σύνθεση

- Σημαντικό θεωρητικό ζήτημα αποτελεί η διάκριση ανάμεσα σε προθηματοποίηση και σύνθεση, δηλαδή η διάκριση ανάμεσα σε πρόθημα και **πρώτο συνθετικό**. Τα προθήματα θεωρούνται γραμματικά μορφήματα, ενώ τα α' συνθετικά θεωρούνται λεξικά μορφήματα

Προθηματοποίηση και σύνθεση

- Οι προθηματοποιημένες λέξεις είναι δύσκολες στην επεξεργασία για διάφορους λόγους:
 1. Είναι **παραγωγικές** αλλά όχι πάντα προβλέψιμες στον σχηματισμό τους.
 2. Ο τρόπος γραφής τους ποικίλλει και εξαρτάται από την κάθε λέξη (*αντιπυρηνικός, αντι-πολεμικός*).
 3. Οι συντακτικές και σημασιολογικές τους ιδιότητες διαφέρουν από τις ιδιότητες των αντίστοιχων ρημάτων χωρίς το πρόθημα, π.χ.:
 - *Ο υδραυλικός έκανε τη βρύση*
 - **Ο υδραυλικός παραέκανε τη βρύση*
 - Η Μαρία κάνει παρέα με τον Πέτρο*
 - Η Μαρία παρακάνει παρέα με τον Πέτρο*

Προθηματοποίηση και σύνθεση

- Πρέπει να τονίσουμε ότι στη ΝΕ η προθηματοποίηση μπορεί να πραγματοποιηθεί και με περισσότερα του ενός προθήματα που εφαρμόζονται σε μια λεξική μονάδα. Ωστόσο η ανάλυση είναι πάντα **διτεμαχιακή**, π.χ. δια-γράφω → ξανα-διαγράφω)

Προθηματοποίηση

- Με το πρόθημα ξε- που είναι ιδιαίτερα παραγωγικό στη νέα ελληνική **φτιάχνονται κυρίως ρήματα** (π.χ. γράφω > ξε-γράφω, φουσκώνω > ξε-φουσκώνω), αλλά **ορισμένες φορές και επίθετα** (π.χ. πνοή > ξέ-πνοος). Η σημασία που δίνεται στις παράγωγες λέξεις με το ξε- είναι:

- α) στέρηση, αντίθεση με αυτό που δηλώνει το θέμα της λέξης (π.χ. κουράζω > **ξε**-κουράζω),
- β) επίταση, ολοκλήρωση αυτού που δηλώνει το θέμα (π.χ. δίνω > **ξε**-δίνω) και
- γ) χρονική διάρκεια (π.χ. νυχτ- > **ξε**-νυχτώ, καλοκαιρ- > **ξε**-καλοκαιριάζω).

Προθηματοποίηση

- Με το πρόθημα **α-** φτιάχνονται επίθετα και η σημασία που δίνεται στη νέα λέξη είναι η **στέρωση** αυτού που δηλώνει το θέμα της. Π.χ. στέγη > ά-στεγος, πόνος > ά-πονος, θεός > ά-θεος. Όταν το θέμα της λέξης αρχίζει από φωνήεν, τότε το πρόθημα παίρνει τη μορφή αν-· π.χ. ικανός > αν-ίκανος.
- Με το πρόθημα **συν-** φτιάχνονται ρήματα (π.χ. εργάζομαι > συν-εργάζομαι), και ουσιαστικά (π.χ. άνθρωπος > συν-άνθρωπος), στα οποία το πρόθημα δίνει τη σημασία του **‘μαζί’**, της συνεργασίας, της συνύπαρξης.
- Με το πρόθημα **παρα-** φτιάχνονται ρήματα (π.χ. τρώω > παρα-τρώω), στα οποία το πρόθημα δίνει τη σημασία της **υπερβολής**.
- Με το πρόθημα **δυσ-** παράγονται ουσιαστικά (π.χ. λεξ- > δυσ-λεξία), επίθετα (π.χ. κάμπτ- > δύσ-καμπτος) και ρήματα (π.χ. λειτουργώ > δυσ-λειτουργώ), στα οποία το πρόθημα δίνει τη **σημασία της αρνητικής, κακής ιδιότητας**.
- Με το πρόθημα **ευ-** παράγονται επίθετα (π.χ. καμπτ- > εύ-καμπτος, στόχ- > εύ-στοχος) και η σημασία που τους δίνεται με το πρόθημα είναι της θετικής, **εύκολης ιδιότητας**.

Επιθήματα

- Τα επιθήματα αποτελούν **μη ελεύθερα (δεσμευμένα)** γραμματικά μορφήματα που εφαρμόζονται στην τελική θέση λεξικού μορφήματος που αποτελεί τη βάση και πριν τα γραμματικά μορφήματα κλίσης, λ.χ. **-ίζ-** στο ουσιαστικό βάση **σκούπα** για να κατασκευαστεί το ρήμα **σκουπ-ίζ-ω**.
- Η επιθηματοποίηση αποτελεί σημαντική διαδικασία κατασκευής λέξεων στη ΝΕ (**Όπως είπαμε, θεωρούνται απλές λέξεις σύμφωνα με το θεωρητικό μας πλαίσιο**). Για τον λόγο αυτό οι παράγωγες λέξεις απαντούν συχνά στα σώματα κειμένων

Επιθήματα

- Υπάρχουν εξειδικευμένα επιθήματα που κατασκευάζουν:
 1. ρήματα, π.χ. σκούπα → σκουπίζω, όμορφος → ομορφαίνω,
 2. ουσιαστικά, π.χ. πορτοκάλι → πορτοκαλιά, παιδί → παιδάκι, κατευθύνω → κατεύθυνση, άρτιος → αρτιότητα,
 3. επίθετα, π.χ. τράπεζα → τραπεζικός, πληρώνω → πληρωτέος, καλός → καλούτσικος, και
 4. επιρρήματα, π.χ. ωραίος → ωραία

Επιθήματα

- Ορισμένα επιθήματα που παράγουν ρήματα:
 - εύ-ω: ψάρ-ι > ψαρ-εύ-ω
 - ών-ω: κλειδ-ί > κλειδ-ών-ω
 - ίζ-ω: αφρός > αφρ-ίζ-ω
- Ορισμένα επιθήματα που παράγουν ουσιαστικά:
 - τη(ς): χορεύω > χορευ-τη(ς)
 - σιμ-ο: δένω > δε-σιμ-ο
 - ότητ-α: άνθρωπος > ανθρωπ-ότητ-α
 - οσύν-η: καλός > καλ-οσύν-η
 - άδ-α: κόκκινος > κοκκιν-άδ-α

Επιθήματα

- Στην κατηγορία των επιθημάτων που παράγουν ουσιαστικά ανήκουν και τα υποκοριστικά και τα μεγεθυντικά:
 - άκ-ι: μωρό > μωρ-άκ-ι
 - ούλ-ι: μικρός > μικρ-ούλ-ι
 - ίτσ-α: κοπέλα > κοπελ-ίτσ-α
 - άρ-α: κορίτσ-ι > κοριτσ-άρ-α
- Ορισμένα επιθήματα που παράγουν επίθετα:
 - ικ-ος: νόμος > νομ-ικ-ός
 - ιν-ος: ξύλο > ξύλ-ιν-ος
 - ίσι-ος: βουνό > βουν-ίσι-ος
 - ιάτικ-ος: Σάββατο > σαββατ-ιάτικ-ος
 - ίστικ-ος: κορίτσι > κοριτσ-ίστικ-ος
 - ένι-ος: χρυσάφι > χρυσαφ-ένι-ος
- Επιθήματα επιρρημάτων:
 - ά: καλός > καλ-ά
 - ώς: ακριβής > ακριβ-ώς

Επιθήματα

- Για την κατασκευή του παραγώγου απαιτείται η αφαίρεση του **κλιτικού μορφήματος** της βάσης (συνήθως ένας ή δύο χαρακτήρες) και κατόπιν η εφαρμογή του επιθήματος, δηλ. τα κλιτικά μορφήματα δεν μετέχουν στην παραγωγή.
- Ενώ γενικά τα επιθήματα αλλάζουν τη γραμματική κατηγορία της βάσης, λ.χ. σκούπα (ουσιαστικό) → σκουπίζω (ρήμα), κάποια επιθήματα όπως τα επιθήματα κατασκευής **υποκοριστικών** και **μεγεθυντικών** διατηρούν τη γραμματική κατηγορία της βάσης
- Για παράδειγμα:
- αυτοκίνητο, .N:Nns (Noun, Nominative, singular)
αυτοκινητάκι, .N+Dim:Nns
- άγγελος, .N:Nms
αγγελούδι, .N+Dim:Nn

Σύνθεση

- Ως **σύνθεση** (compounding) ορίζεται η διαδικασία με την οποία κατασκευάζονται λέξεις ή συνδυασμοί λέξεων από τη συνένωση θεμάτων, π.χ. φιλ-ό-ζω(ος), ή ενός θέματος και μίας λέξης π.χ. ντοματ-ο-σαλάτα ή, τέλος, δύο λέξεων π.χ. ξαναβρέχω, ζώνη ασφαλείας
- Τα θέματα και οι λέξεις ανήκουν στη γραμματική κατηγορία των ουσιαστικών, των επιθέτων, των ρημάτων και των επιρρημάτων
- Η κατασκευή ενός συνθέτου μπορεί να οδηγήσει σε μονολεκτικό σύνθετο ή σε πολυλεκτικό. Όπως είπαμε, σύμφωνα με το θεωρητικό μας πλαίσιο, τα **μονολεκτικά σύνθετα θα θεωρηθούν απλές λέξεις**, ενώ **σύνθετα θα θεωρηθούν μόνο τα πολυλεκτικά**

Σύνθεση

Τα περισσότερα μέρη του λόγου μπορούν να είναι πρώτα και δεύτερα συνθετικά· π.χ.

- στη λέξη *καρεκλοπόδαρο* το πρώτο συνθετικό, δηλ. το *καρεκλ-*, είναι ουσιαστικό και το δεύτερο συνθετικό, το *-ποδαρ-* είναι επίσης ουσιαστικό·
- στη λέξη *ανεβοκατεβαίνω* το πρώτο συνθετικό, δηλ. το *ανεβ-*, είναι ρήμα και το δεύτερο συνθετικό, το *-κατεβ-* είναι επίσης ρήμα·
- στη λέξη *πισώπλατα* το πρώτο συνθετικό, δηλ. το *πισω-* είναι επίρρημα·
- στη λέξη *παραπέρα* το δεύτερο συνθετικό, το *-πέρα*, είναι επίρρημα·

Σύνθεση

Οι σύνθετες λέξεις μπορούν να ανήκουν σε όλα τα βασικά μέρη του λόγου· π.χ.

- η σύνθετη λέξη *καρεκλοπόδαρο* είναι ουσιαστικό,
- η σύνθετη λέξη *στενόμακρος* είναι επίθετο,
- η σύνθετη λέξη *ανεβοκατεβαίνω* είναι ρήμα,
- η σύνθετη λέξη *καλόκαρδα* είναι επίρρημα.

Σύνθεση

- Οι σύνθετες λέξεις μπορεί να έχουν κατάληξη που είναι η ίδια με την κατάληξη του δεύτερου συνθετικού · π.χ.
 - στη λέξη παλιοκαρέκλα η κατάληξη -α είναι ίδια με την κατάληξη του δεύτερου συνθετικού καρέκλ-α.
- Άλλες φορές, όμως, η κατάληξη της σύνθετης λέξης μπορεί να είναι διαφορετική · π.χ.
 - στη λέξη καρεκλοπόδαρο η κατάληξη -ο **δεν είναι ίδια** με την κατάληξη του δεύτερου συνθετικού ποδάρ-ι.

Μονολεκτικά σύνθετα

- Τα θέματα που μετέχουν στα μονολεκτικά σύνθετα μπορεί να είναι **ελεύθερα** (free), **ελευθερώσιμα** (freeable) ή **δεσμευμένα** (bound) (Αναστασιάδη-Συμεωνίδη, 1986β), π.χ. ξανα-βρέχ-ω όπου **το πρώτο θέμα είναι ελεύθερο και το δεύτερο ελευθερώσιμο**, αδελφ-ο-κτόνο-ς όπου **το πρώτο θέμα είναι ελευθερώσιμο και το δεύτερο δεσμευμένο**
- **Ελευθερώσιμο** είναι το θέμα που μπορεί να αποτελέσει λέξη, εφόσον συμπληρωθεί με το ή τα κλιτικά μορφήματα,
- **δεσμευμένο** είναι το θέμα το οποίο αποτελεί πάντα τμήμα λέξης, δηλ. δεν έχει αυτονομία, ελευθερία

Μονολεκτικά σύνθετα

- μεταξύ των συστατικών των μονολεκτικών συνθέτων συνήθως παρεμβάλλεται ένα **συνδετικό φωνήεν**, το οποίο τις περισσότερες φορές έχει τη μορφή -ο-, π.χ. ανοιγ-ο-κλείνω, νυχ-ο-κόπτης
- Η σημασία της σύνθετης λέξης μπορεί να είναι **συνθετική**, δηλ. **προβλέψιμη** σε ποικίλο βαθμό όπως π.χ. αλατοπίπερο
- Το σύστημα κανόνων της σύνθεσης επιτρέπει τη δημιουργία πολλών ειδών συνθέτων ανάλογα με τις συντακτικές σχέσεις των δύο συνθετικών: **παρατακτικά σύνθετα** π.χ. αλατοπίπερο, **προσδιοριστικά** π.χ. αγριοπερίστερο, **κτητικά** π.χ. ερυθρόδερμος, **αντικειμενικά** π.χ. καντηλανάφτης
- Επίσης τα σύνθετα κατατάσσονται σε **ενδοκεντρικά** (endocentric) και **εξωκεντρικά**

Ενδοκεντρικά

- Ενδοκεντρικά: η κεφαλή και στις δύο περιπτώσεις βρίσκεται εντός της εσωτερικής δομής, δηλαδή είναι ένα από τα δύο συστατικά ("αγριογούρουνο" > "γουρούνι" και "πατατοσαλάτα" > "σαλάτα").
- Επίσης, όπως αναφέρθηκε, για τη ΝΕ στα υποτακτικά σύνθετα η κεφαλή βρίσκεται πάντα δεξιά.

Εξωκεντρικά

- Εξωκεντρικά: στα εξωκεντρικά σύνθετα θεωρείται ότι η κεφαλή απουσιάζει
Π.χ. [κοκκινομάλλης]Επιθ. > [κόκκινο]Επιθ. – [μαλλί]όνομα
- Οι λόγοι είναι οι εξής: **το πρώτο συστατικό: είναι Επίθετο**, ενώ το **δεύτερο συστατικό: Όνομα**. Δεύτερον, ως προς το γένος, το δεύτερο συστατικό είναι γένους ουδετέρου ("μαλλί"), ενώ το σύνθετο είναι γένους αρσενικού. Δηλαδή ο "κοκκινομάλλης" δεν είναι "ένα είδος μαλλιού", αλλά "αυτός που έχει κόκκινα μαλλιά".
- Επομένως, παρατηρούμε ότι στο συγκεκριμένο σύνθετο τόσο η **γραμματική κατηγορία** (επίθετο) όσο το **γένος** (αρσενικό) και η **βασική σημασία** ("αυτός που έχει κόκκινα μαλλιά") δεν προκύπτουν από τα επιμέρους συστατικά του συνθέτου.

Μονολεκτικά σύνθετα

- Ιδιαίτερο ενδιαφέρον παρουσιάζει και η σύνθεση και η **αποσύνθεση** αυτών των ακολουθιών. Σε κάποιες περιπτώσεις, το α' συνθετικό μπορεί να αποκοπεί από το ρήμα και να λειτουργήσει ως επίρρημα:

*Η Μαρία ξανατραγούδησε
= Η Μαρία τραγούδησε ξανά*

*Η Μαρία στραβοπατάει
= Η Μαρία πατάει στραβά*

*Η Μαρία κουτσοπίνει
Η Μαρία πίνει κουτσά

Πολυλεκτικά σύνθετα

- Στο χώρο των **πολυλεκτικών μονάδων** (multiword units) τίθεται ένα σημαντικό ζήτημα: ο διαχωρισμός των πολυλεκτικών μονάδων από τις ελεύθερες ακολουθίες.
- Ως **πολυλεκτικό σύνθετο** ορίζεται κάθε ακολουθία λέξεων που εμφανίζει κάποιο βαθμό παγίωσης και περιγράφει μία σημασιολογική μονάδα στη γενική γλώσσα π.χ. *μαύρη αγορά, κρέμα γάλακτος*, ή στα επιστημονικά και τεχνικά λεξιλόγια, π.χ. *βαρύ ύδωρ, αγωγή του πολίτη.*

Σύνθετη λέξη

- Σύνθετη λεξική μονάδα ορίζεται η ακολουθία δύο τουλάχιστον απλών λέξεων με ένα τουλάχιστον διαχωριστικό. Ως διαχωριστικά ορίζονται **ΤΟ ΚΕΝΟ**, το **ΕΝΩΤΙΚΟ** και η **απόστροφος**. Λόγου χάρη οι ακολουθίες κινητό τηλέφωνο, παιδί-θαύμα και αντ' αυτού αποτελούν σύνθετες λεξικές μονάδες
- Ο ορισμός αυτός διαχωρίζει τις σύνθετες λέξεις (πολυλεκτικά σύνθετα) από τις απλές και ο οποίος βασίζεται σε καθαρά τυπογραφικά κριτήρια
- **ΟΡΙΣΜΟΣ ΑΝΑΓΚΑΙΟΣ ΑΛΛΑ ΉΧΙ ΕΠΑΡΚΗΣ**

Σύνθετη λέξη

- Διεθνής και ελληνική βιβλιογραφία έχει πληθώρα κριτηρίων:
 1. μια ακολουθία απλών λέξεων είναι σύνθετη όταν τουλάχιστον μία από τις **συντακτικές ή σημασιολογικές της ιδιότητες δεν συνάγεται από τις ιδιότητες των συστατικών** της. Για παράδειγμα, υπάρχουν πολυλεκτικές σύνθετες μονάδες οι οποίες απαντούν είτε μόνο στον ενικό είτε μόνο στον πληθυντικό αριθμό, περιορισμοί που αποτελούν ένδειξη παγιωμένης ακολουθίας

Το πολυλεκτικό σύνθετο όνομα *εργατικά χέρια* είναι συλλογικό ανθρώπινο ουσιαστικό το οποίο μπορεί να λειτουργήσει ως υποκείμενο ρημάτων όπως το *διεκδικώ*. Συνεπώς, το **πολυλεκτικό σύνθετο περιγράφει έννοια διαφορετική από την έννοια του ουσιαστικού-κεφαλή**. Τα δύο ουσιαστικά (*χέρια – εργατικά χέρια*) πρέπει να φέρουν διαφορετικά σημασιολογικά χαρακτηριστικά και για τον λόγο αυτό **στο ηλεκτρονικό μορφολογικό λεξικό καταγράφονται σε δύο διαφορετικά λήμματα**

Σύνθετη λέξη

- Οι συνηθέστερες μορφές πολυλεκτικών συνθέτων στη ΝΕ (Αναστασιάδη-Συμεωνίδη, 1986) είναι οι δομές που αποτελούνται από:
 - α) **επίθετο + ουσιαστικό**, π.χ. *ψυχρός πόλεμος*,
 - β) **ουσιαστικό + ουσιαστικό σε γενική πτώση**, π.χ. *φακός επαφής*,
 - γ) **ουσιαστικό + ουσιαστικό στην ίδια πτώση ή αλλιώς παραθετικά σύνθετα**, π.χ. *λέξη-κλειδί*

Σύνθετη λέξη

- Η πιο απλή περίπτωση αφορά τις ακολουθίες που είναι **παγιωμένες στο σύνολό τους**, π.χ. αποδιοπομπαίος τράγος. Ωστόσο, αυτή δεν είναι η πιο συνηθισμένη περίπτωση.
- Συχνά, παρατηρούμε στον παραδειγματικό άξονα την υποκατάσταση λεξικών μονάδων με άλλες με την ίδια λειτουργία, όπως π.χ. στις ακολουθίες *υπαίθρια/τοπική/κεντρική/λαϊκή/ελεύθερη/μαύρη αγορά*, όπου παρατηρείται κάποιος βαθμός αδιαφάνειας, αλλά ταυτόχρονα υπάρχει και λεξιλογική ελευθερία (Gross, 1996: 17).
- Επιπλέον, μελετώντας τα παραπάνω παραδείγματα παρατηρούμε ότι η ακολουθία *μαύρη αγορά* είναι πιο παγιωμένη από την *υπαίθρια αγορά*. **Υπάρχουν λοιπόν βαθμοί παγίωσης στις γλώσσες**, ένα συνεχές ανάμεσα στις ελεύθερες και τις παγιωμένες ακολουθίες

Φρασεολογία

- Με τον όρο **φρασεολογία** αναφερόμαστε στο σύνολο των φαινομένων που αφορούν τους σταθερούς, δηλ. μη ελεύθερους συνδυασμούς δύο ή περισσότερων λέξεων που η σημασία τους δεν είναι γενικά συνθετική, δηλ. που δεν είναι προβλέψιμη
- Στην ελληνική βιβλιογραφία πολλοί όροι αναφέρονται σ' αυτές τις πολυλεκτικές μονάδες: **φρασεολογισμοί, ιδιωματισμοί, ιδιωτισμοί, ιδιωτισμικές εκφράσεις, παγιωμένες/τυποποιημένες/στερεότυπες εκφράσεις**. Σ' αυτό το ευρύ φάσμα εντάσσονται και οι **παροιμίες** (βλ. Μυτούλα, 2005, Τσακνάκη, 2005)

Φρασεολογία

- Στον χώρο των πολυλεκτικών ακολουθιών τίθεται και διεθνώς ένα σημαντικό ζήτημα: ο διαχωρισμός των **στερεότυπων εκφράσεων** (idioms) από τις **λεξικές συνάψεις** (collocations) και τις στερεότυπες εκφράσεις, π.χ. **κλείνω το παράθυρο** (ελεύθερη ακολουθία), **κλείνω θέση** (**λεξική σύναψη**), **δεν έκλεισα μάτι** (**στερεότυπη έκφραση**)

Στερεότυπες εκφράσεις

- Οι στερεότυπες εκφράσεις συνοδεύονται από ορισμένα χαρακτηριστικά και διέπονται από ειδικούς περιορισμούς. Ωστόσο κάθε στερεότυπη έκφραση δεν φέρει όλα τα χαρακτηριστικά και όλους τους περιορισμούς.
- Βασικό χαρακτηριστικό (Αναστασιάδη-Συμεωνίδη & Ευθυμίου, 2006), η μη συνθετικότητα της σημασίας, λ.χ. *έσκισε τη γάτα στα δύο* (μη συνθετική σημασία) – *ανοίγω το στόμα μου 'μαρτυρώ μυστικό'* (μερικώς συνθετική).
- Επίσης οι στερεότυπες εκφράσεις μπορεί να χαρακτηρίζονται από έλλειψη, απουσία στοιχείων λ.χ. του οριστικού άρθρου: *παίρνω πόδι*, του ρήματος είναι ή έχω: *κακά τα ψέματα*.
- Ακόμη χαρακτηρίζονται από **λογιότητα** ή **προφορικότητα**, λ.χ. *η ισχύς εν τη ενώσει, ο αποθανών δεδικαίωται* – *τα 'χασα, για όνομα του Θεού!*, *δεν έχω μούτρα να+ρήμα*

Στερεότυπες εκφράσεις

- Τέλος υπάρχουν δομικά σχήματα με τα οποία σχηματίζονται πολλές στερεότυπες εκφράσεις, λ.χ. με υποκοριστικό π.χ. *την προσέχει τη ζωούλα του, με το σχήμα της παρομοίωσης* π.χ. *βρίζει σα χαμάλης, με το σχήμα της επανάληψης* π.χ. *κόσμος και κοσμάκης, είδα κι απόειδα, όπου φύγει φύγει, με το σχήμα όποιος/όπου+ρήμα ως φόρμουλα έναρξης στις παροιμίες* π.χ. *όποιος θέλει τα πολλά χάνει και τα λίγα.*

Στερεότυπες εκφράσεις

- Παρατηρείται επίσης ότι κανένα από τα στοιχεία που απαρτίζουν μια στερεότυπη έκφραση δεν μπορεί να **επικαιροποιηθεί** μεμονωμένα π.χ. *πλήρως τα σπασμένα* - **πλήρως αυτά τα σπασμένα/τα σπασμένα μου.*
- Διαπιστώνεται (βλ. και Μότσιου, 1987) υποχρεωτική παρουσία γραμματικών κατηγοριών λ.χ. του **αριθμού** (ενικός-πληθυντικός) π.χ. *βγάζει μάτι* - **βγάζει μάτια, είναι για τα πανηγύρια* - **είναι για το πανηγύρι, του γένους* π.χ. *η Μαρία έγινε Τούρκος 'θύμωσε'* - **η Μαρία έγινε Τουρκάλα, ορισμένου χρόνου* π.χ. *τα χρειάστηκα 'φοβήθηκα'* - **θα τα χρειαστώ, ορισμένης φωνής* π.χ. *πάτησα την πεπονόφλουδα* - **πατήθηκε η πεπονόφλουδα.*
- Ως προς τις **τροπικότητες**, παρατηρείται υποχρεωτική παρουσία **άρνησης** π.χ. *καρφί δε μου καίγεται, δεν έχω μούτρα να+ρήμα, ερώτησης* π.χ. *τι γυρεύει η αλεπού στο παζάρι;*, **προστακτικής** π.χ. *πέσε, πίτα, να σε φάω.*

Στερεότυπες εκφράσεις

- Παρατηρείται υποχρεωτική παρουσία των αδύνατων τύπων της προσωπικής αντωνυμίας γ' προσώπου, π.χ. την έπαθα, την κάνω 'φεύγω', και απολιθωμένων τύπων και συντάξεων, π.χ. τα ευκόλως εννοούμενα παραλείπονται, τα καλά κόποις κτώνται, ο σώζων εαυτόν σωθήτω, καλώς εχόντων των πραγμάτων, έχω προ οφθαλμών, είναι εν γνώσει μου, χαίρει άκρας υγείας, τρόπος του λέγειν, δράττομαι της ευκαιρίας.
- Παρατηρείται σταθερή **σειρά των όρων** π.χ. είναι διαβόλου κάλτσα - *είναι κάλτσα διαβόλου και καταπάτηση των **επιλεκτικών περιορισμών** π.χ. χάφτει μύγες (**για άνθρωπο**).
- Επίσης τα συστατικά της στερεότυπης έκφρασης δεν δέχονται επιμέρους προσδιορισμούς π.χ. κάνω την πάπια – *κάνω την όμορφη πάπια, και δεν λειτουργεί η **υποκατάσταση στον παραδειγματικό άξονα** π.χ. πυρ και μανία - *φωτιά και μανία.

Λεξικές συνάψεις (collocation)

- Με τον όρο αυτό, ορίζεται η ακολουθία λέξεων με υψηλή συχνότητα συνεμφάνισης στον λόγο και περιορισμό στις λεξιλογικές εναλλαγές και στους συντακτικούς μετασχηματισμούς.
- Σύμφωνα με τους Sinclair & Carter (1991), η **σύναψη** είναι η εμφάνιση δύο ή περισσότερων λέξεων με μικρή απόσταση μεταξύ τους σε ένα κείμενο
- Η González-Rey (2002: 70 & 84) ορίζει τις λεξικές συνάψεις ως συνδυασμούς λέξεων που «προτιμούν» να συνυπάρχουν στον λόγο με αναφορική-δηλωτική λειτουργία (κυριολεξία), διατηρώντας τη σημασία που έχουν ως ξεχωριστές λεξικές μονάδες.

Λεξικές συνάψεις (collocation)

- Τέλος, οι Gurrutxaga & Alegria (2011) διακρίνουν τις λεξικές συνάψεις σε δύο κατηγορίες:

α) τους **ημι-συνθετικούς συνδυασμούς** όπου το ουσιαστικό διατηρεί την κυριολεκτική του σημασία, ενώ το ρήμα λειτουργεί ως **υποστηρικτικό** ή έχει ειδική σημασία μέσα σε αυτό τον συνδυασμό και

β) τους **συνθετικούς συνδυασμούς** με λεξικούς περιορισμούς όπου δεν υφίσταται η δυνατότητα αντικατάστασης του ρήματος με συνώνυμά του.

Κριτήρια διαχωρισμού των λεξικών συνάψεων

- Ως κριτήρια διαχωρισμού των λεξικών συνάψεων από τις ελεύθερες ακολουθίες, τα πολυλεκτικά σύνθετα και τις στερεότυπες εκφράσεις υπάρχει η δυνατότητα (στις λεξικές συνάψεις):
- α) ανάπτυξης ανάμεσα σε δύο στοιχεία, π.χ. γερά, ατσάλινα νεύρα,
- β) σύμπλεξης με άλλα επίθετα, π.χ. δυνατός και μυρωδάτος καφές,
- γ) διαβάθμισης από ποσοτικά επιρρήματα, π.χ. πολύ δυνατός καφές,
- δ) τοποθέτησης του επιθέτου σε θέση κατηγορουμένου, π.χ. ο καφές είναι δυνατός.

Συνεπώς η ακολουθία δυνατός καφές δεν αποτελεί ελεύθερη ακολουθία ή πολυλεκτικό σύνθετο αλλά λεξική σύναψη

Ονοματικές οντότητες

- Οι **ονοματικές οντότητες** (στο εξής ΟΟ) (named entities), μια ευρεία γλωσσική κατηγορία με υψηλή πληροφοριακή αξία, αποτελούν πόλο έλξης των ερευνητών που ασχολούνται με τη Θεωρητική Γλωσσολογία αλλά και τη Γλωσσολογία **σωμάτων κειμένων**, την Υπολογιστική Γλωσσολογία, την έρευνα στην **ορολογία**, την **Ιστορική Γλωσσολογία**, τη σύνταξη **εγκυκλοπαιδικών λεξικών** καθώς και τη **Μεταφρασεολογία**.
- Η αναγνώριση των ΟΟ αποτελεί σημαντική διαδικασία για την **εξαγωγή** και **διαχείριση πληροφορίας**, καθώς οι ΟΟ κατέχουν την πρώτη θέση στις **μηχανές αναζήτησης**, όπως επίσης και στα συστήματα ερωταποκρίσεων.

Ορισμός ονοματικής οντότητας

- ΟΟ η κατηγορία που περικλείει «όλα τα γλωσσικά στοιχεία που περιγράφουν μία ενιαία και συγκεκριμένη οντότητα η οποία ανήκει σε έναν συγκεκριμένο τομέα».
- Η κατηγορία των ΟΟ, που είναι γνωστότερες ως **κύρια ονόματα**, βρίσκεται σε αντίστιξη προς την κατηγορία των **κοινών ή προσηγορικών ονομάτων**, π.χ. *άνθρωπος, πόλη, βουνό, τράπεζα, βιβλίο, επανάσταση.*

Κατηγορίες των ΟΟ

- α) **ανθρωπωνύμια** ή **αγιωνύμια** π.χ. *Μέγας Αλέξανδρος, Μέγας Κωνσταντίνος, Άγιος Δημήτριος,*
- β) **τοπωνύμια** (ονόματα χωρών, **οικωνύμια** (ονόματα πόλεων κτλ.) (βλ. Συμεωνίδης, 2010), ονομασίες περιοχών κτλ.), π.χ. *Ανατολική Ευρώπη, Εύξεινος Πόντος,*
- γ) **ορειωνύμια, υδρωνύμια** π.χ. *Πίνδος, λίμνη Αγίου Βασιλείου,*
- δ) **οργανισμοί**, π.χ. *Αγροτική Τράπεζα, Βρετανικό Μουσείο, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης,*
- ε) **συγγραφικά έργα**, π.χ. *Πράσινη Βίβλος, Εθνικός Ύμνος,*
- στ) **γεγονότα**, π.χ. *Γαλλική Επανάσταση, Ολυμπιακοί Αγώνες.*

Κατηγορίες των ΟΟ

- μπορούν να συμπεριληφθούν οι **χρονικές** και οι **ποσοτικές εκφράσεις**. Επομένως, χαρακτηρίζουμε ΟΟ και:
 - α) τις χρονικές εκφράσεις, π.χ. *Μεγάλη Εβδομάδα, Ναυτική Εβδομάδα,*
 - β) τις ημερομηνίες, π.χ. *22 Σεπτεμβρίου 1979, 22/09/79,*
 - γ) τις ώρες, π.χ. *7 η ώρα, εφτά η ώρα, 7:00,*
 - δ) τις αριθμητικές εκφράσεις, π.χ. *400 χιλιόμετρα, 400 χμ., 5 εκατομμύρια ευρώ, 5.000.000 ευρώ.*
- Επιπρόσθετα, θεωρούμε ΟΟ τις ακολουθίες που περιγράφουν προσωπικά στοιχεία, όπως το **ηλεκτρονικό ταχυδρομείο**, π.χ. *metafrasi@frl.auth.gr* και τις **διευθύνσεις**, π.χ. *Καρόλου Ντηλ 1, Πανεπιστημιούπολη Α.Π.Θ., Θεσσαλονίκη, Τ.Κ. 54124.*
- Τέλος, στις ΟΟ καταχωρίζουμε και τα **μεταφορικά μέσα** τα οποία περιλαμβάνουν ένα ή περισσότερα κύρια ονόματα, όπως το *Express Samina/το Εξπρές Σαμίνα, το αεροπλάνο Airbus A316.*

Άλλες κατηγορίες ΟΟ

- εταιρίες και οργανισμοί,
- φυσικά πρόσωπα,
- τοποθεσίες,
- εκφράσεις ώρας,
- χρηματικά ποσά και ποσοστώσεις,
- Ονόματα κτιρίων ή/και άλλων ανθρώπινων κατασκευών,
- Δρομολόγια πλοίων

Τύπος Ονοματικής	
Οντότητας	Παραδείγματα
ΠΡΟΣΩΠΑ	Ελ. Βενιζέλος, Θ. Κολοκοτρώνης
ΟΡΓΑΝΙΣΜΟΙ	ΟΗΕ, ΟΤΕ
ΤΟΠΟΣ	Κέρκυρα, Όλυμπος, Ανώγια
ΗΜΕΡΟΜΗΝΙΑ	08 Ιουνίου 1982
ΩΡΑ	12:30 π.μ., δύο και μισή μ.μ.
ΧΡΗΜΑΤΑ	13 εκατομμύρια Ευρώ, GBP 2342.43
ΠΟΣΟΣΤΑ	Είκοσι τοις εκατό, 33%
FACILITY	Στήλες του Ολυμπίου Διός, Ακρόπολη, Stonehenge
GPE	Ελλάδα,

Μορφές των ΟΟ

- Μια ΟΟ μπορεί να παρουσιάζει ποικίλες μορφές.
- Ειδικότερα, οι ημερομηνίες εμφανίζονται είτε αριθμητικά (14/12 ή 14-12) είτε με γλωσσικά μέσα, π.χ. δεκατέσσερις Δεκεμβρίου είτε σε συνδυασμό αριθμητικής και λεκτικής μορφής (αλφαριθμητικά), π.χ. 14 Δεκεμβρίου 2013.
- Επιπλέον, μια ΟΟ μπορεί να περιέχει εκτός από ελληνικούς και λατινικούς χαρακτήρες (το *Express Samina*, το αεροπλάνο *Airbus A316*).
- Στο συντακτικό επίπεδο, οι ΟΟ απαντούν στις ερωτήσεις ποιος, πού, πότε.

Δυσκολίες κατηγοριοποίησης ονοματικών οντοτήτων

- Η κατηγοριοποίηση των ΟΟ δεν είναι πάντα ευχερής, καθώς φαινόμενα όπως (α) η **ομωνυμία**, (β) η **πολυσημία** και (γ) η **μετωνυμία** δυσχεραίνουν τη διαδικασία αυτή:
 1. Ο τύπος αρκτικολέξου *Α.Ε.* αποτελεί τη συντομευμένη μορφή δύο διαφορετικών ΟΟ, δηλ. ομώνυμων ΟΟ: α) *Ανώνυμη Εταιρεία* και β) *Ασφαλιστική Εταιρεία*.
 2. Επίσης, εντάσσουμε στην κατηγορία των ομωνύμων και τις περιπτώσεις όπου ο ίδιος τύπος ΟΟ αναφέρεται σε διαφορετικά αντικείμενα αναφοράς, π.χ. ο τύπος *Ελευθέριος Βενιζέλος* μπορεί να αναφέρεται τόσο στον πολιτικό άνδρα, στο ιστορικό πρόσωπο, όσο και στον αερολιμένα της Αθήνας

Δυσκολίες κατηγοριοποίησης ονοματικών οντοτήτων

- Πολύσημες θεωρούμε τις ΟΟ όπου η ίδια μορφή αντιστοιχεί σε πολλαπλές πλευρές του ίδιου αντικειμένου αναφοράς. Λόγου χάρη, το ανθρωπωνύμιο *Αντώνης Σαμαράς* περιγράφει ταυτόχρονα ένα πολιτικό πρόσωπο, τον τ. πρωθυπουργό της Ελλάδας, όπως επίσης και τον τ. πρόεδρο του κόμματος *Νέα Δημοκρατία*. Στις περιπτώσεις αυτές στο λεξικό υπάρχει μία μόνο καταχώριση.
- Τέλος, η μετωνυμία είναι ένα σχήμα λόγου που μαζί με τη μεταφορά χρησιμοποιείται ευρύτατα στη γλώσσα για την κατονομασία νέων αντικειμένων, οντοτήτων κτλ. με ήδη υπάρχον υλικό. (π.χ. *11η Σεπτεμβρίου* εκτός από ημερομηνία αποτελεί ΟΟ η οποία μετωνυμικά περιγράφει ένα γεγονός:

Δυσκολίες εντοπισμού ονοματικών οντοτήτων

- Οι ΟΟ παρουσιάζουν δυσκολίες και στον εντοπισμό τους, καθώς συχνά απαντούν στα κείμενα υπό τη μορφή **αρκτικολέξου** ή **ακρωνυμίου**:

π.χ.

- 2α. ... και κυρίως στις γραμματείες των περιφερειών και σε διοικήσεις ΔΕΚΟ
- 2β. ... απασχόλησε χθες το Ανώτατο Δικαστήριο, όπου προσέφυγε η διοίκηση των ΕΛΤΑ με αίτημα να αναιρεθεί προηγούμενη απόφαση.
- η μορφή με την οποία εμφανίζονται οι ΟΟ δεν είναι πάντοτε σταθερή, ακόμα και μέσα στο ίδιο κείμενο

Δυσκολίες εντοπισμού ονοματικών οντοτήτων

- ζήτημα στον εντοπισμό των ΟΟ τίθεται, όπως είδαμε, από τα ομώνυμα αρκτικόλεξα

π.χ. ο τύπος του αρκτικολέξου *Α.Ε.* αποτελεί τη συντομευμένη μορφή δύο διαφορετικών ΟΟ: α) *Ανώνυμη Εταιρεία* και β) *Ασφαλιστική Εταιρεία*.

- Η δυσκολία στην αυτόματη αναγνώρισή τους μπορεί να παραμείνει ακόμα και όταν ο ίδιος τύπος αρκτικολέξου αντιστοιχεί σε ΟΟ διαφορετικού γένους, όπως ο τύπος αρκτικολέξου *Γ.Σ.* που αντιστοιχεί στο α) *Γυμναστικός Σύλλογος*, ΟΟ αρσενικού γένους, και στο β) *Γενική Συνέλευση*, ΟΟ θηλυκού γένους.

Δυσκολίες εντοπισμού ονοματικών οντοτήτων

- Η δυσκολία εντοπισμού μιας ΟΟ κυρίως από συστήματα αυτόματης αναγνώρισης δομών σε κείμενα οφείλεται και στο γεγονός ότι η κατηγορία αυτή διαθέτει ανοιχτό, απροσδιόριστο και ετερογενή αριθμό λημμάτων, και επομένως καθίσταται αδύνατη η πλήρης καταχώρισή τους σε λεξικά είτε έντυπα είτε ηλεκτρονικά.

Αναγνώριση ονοματικών οντοτήτων

Τρομοκρατικές Ενέργειες

- Δίνεται μία συλλογή κειμένων γύρω από ειδήσεις τρομοκρατικών ενεργειών
- Για κάθε κείμενο καθόρισε:
 - Τον τύπο της ενέργειας
 - Την ημερομηνία
 - Την τοποθεσία
 - κτλ.
- Συμπλήρωσε μία βάση δεδομένων με αυτά τα στοιχεία (templates)

Είσοδος:

- 19 March - A bomb went off this morning near a power tower in San Salvador leaving a large part of the city without energy, but no casualties have been reported. According to unofficial sources, the bomb - allegedly detonated by urban guerrilla commandos - blew up a power tower in the northwestern part of San Salvador at 0650 (1250 GMT).

Ονοματικές Οντότητες

Έξοδος:

- Incident type	bombing
- Date	March 19
- Location	El Salvador: San Salvador (city)
- Perpetrator	urban guerilla commandos
- Physical target	power tower
- Human target	-
- Effect on physical target	destroyed
- Effect on human target	no injury or death
- Instrument	bomb

Δράστης

στόχος

Δυσκολίες

- Κανένα λεξικό δεν μπορεί να συμπεριλάβει όλα τα υπάρχοντα κύρια ονόματα, διευθύνσεις κτλ.
- Νέα κύρια ονόματα δημιουργούνται συνεχώς
- Η ίδια ονομαστική οντότητα μπορεί να αναφέρεται με διάφορες παραλλαγές (Coca Cola - Coke)
- Συνήθως δημιουργούνται ακρωνύμια- συντομεύσεις για τα κύρια ονόματα
- Τα ακρωνύμια-συντομεύσεις δεν είναι πάντα κύρια ονόματα (π.χ. λ.χ. κτλ.)
- Η χρήση κεφαλαίων γραμμάτων δεν είναι πάντα κανόνας

Ασάφεια Ονομάτων - Οντοτήτων

- Ακόμα και αν αναγνωριστεί ένα όνομα-οντότητα συχνά είναι δύσκολο να ταξινομηθεί σωστά
 - Άνθρωπος ή εταιρεία: *Ford, Philip Morris*
 - Άνθρωπος ή τοποθεσία: *Jordan, JFK*
 - Άνθρωπος ή μήνας: *April, June*
 - Ακρωνύμιο ή οργανισμός: MRI (Magnetic Resonance Imaging, Mental Research Institute)
 - ...

Προσεγγίσεις

- Χειρονακτικοί κανόνες
 - Συνήθως αποδίδουν καλύτερα για εξειδικευμένες εφαρμογές
 - Δύσκολο να κατασκευαστούν
 - Domain-specific (Εξαρτώμενοι από την θεματική περιοχή)
- Μηχανική μάθηση
 - Μπορεί να προσαρμοστεί εύκολα σε νέα πεδία
 - Χρειάζεται domain-specific δεδομένα εκπαίδευσης
- Η αναγνώριση ονοματικών οντοτήτων είναι από τις πιο πετυχημένες εργασίες επεξεργασίας φυσικής γλώσσας (ακρίβεια ~95%)

Μεταφραστικές δυσκολίες ονοματικών οντοτήτων

- Όσον αφορά τη μετάφραση, τίθενται ζητήματα τόσο ταύτισης όσο και απόδοσης.
- Όλα τα προαναφερθέντα χαρακτηριστικά των ΟΟ δυσχεραίνουν και πολλές φορές καθιστούν αδύνατη την ταύτιση ή/και απόδοση των ΟΟ σε μια άλλη γλώσσα-στόχο τόσο από τον άνθρωπο-μεταφραστή όσο και από τα συστήματα **αυτόματης ή μηχανικής μετάφρασης**.
- Τα φαινόμενα της **ομωνυμίας, της πολυσημίας και της μετωνυμίας** σε συνδυασμό με τη μη πληρότητα καταγραφής των ΟΟ από τα έντυπα και τα ηλεκτρονικά δίγλωσσα ή πολύγλωσσα λεξικά δυσχεραίνουν το έργο του μεταφραστή

Μεταφραστικές δυσκολίες ονοματικών οντοτήτων

- Α) Ζητήματα ταύτισης: τα ονόματα οργανισμών δεν εμφανίζονται πάντα με ομοιογενή τρόπο και, όταν πρόκειται να τα μεταφράσουμε, αντιλαμβανόμαστε ότι οι μεταφραστικές πρακτικές αποκλίνουν. Υπάρχουν παραδείγματα όπου η ΟΟ περιλαμβάνει το ίδιο κύριο όνομα αλλά με διαφορετικές μορφές. Τίθεται ζήτημα ταύτισης των δύο μορφών

Μεταφραστικές δυσκολίες ονοματικών οντοτήτων

- Β) Ζητήματα **απόδοσης**. Το γεγονός ότι για πληθώρα ΟΟ δεν υφίστανται αντιστοιχίες σε άλλες γλώσσες προκαλεί δυσκολίες στην απόδοσή τους.
- Για παράδειγμα, η ΟΟ της γαλλικής γλώσσας *Aéroport Charles-de-Gaulle* θα μπορούσε να αποδοθεί στην ελληνική γλώσσα είτε αποδίδοντας την απλή λέξη *aéroport* ως *αεροδρόμιο* και παραθέτοντας το κύριο όνομα όπως στη γλώσσα-πηγή (*αεροδρόμιο Charles-de-Gaulle*) είτε κάνοντας και φωνητική μεταγραφή του κύριου ονόματος (*αεροδρόμιο Σαρλ ντε Γκωλ*).
- Επίσης, το αρκτικόλεξο *CNRS* το οποίο αποτελεί τη συντομευμένη μορφή της ΟΟ *Centre National de la Recherche Scientifique* δεν έχει αντιστοιχία στην ελληνική γλώσσα και οποιαδήποτε απόδοσή του δεν θα απέδιδε επαρκώς και το σημασιολογικό φορτίο που φέρει.
- Αντίστοιχα, ο ελληνικός οργανισμός *Παγκρήτια Τράπεζα* δεν θα μπορούσε να αποδοθεί στο σύνολό του σε καμία άλλη γλώσσα γιατί, όπως παρατηρούν και οι Grass & Maurel (2008: 163), κάποιες ενώσεις έχουν εθνικό χαρακτήρα και δεν είναι γνωστές έξω από αυτό το εθνικό περιβάλλον, εκτός αν χαίρουν κάποιας αναγνωρισιμότητας και κατ' επέκταση μεταφράζονται.

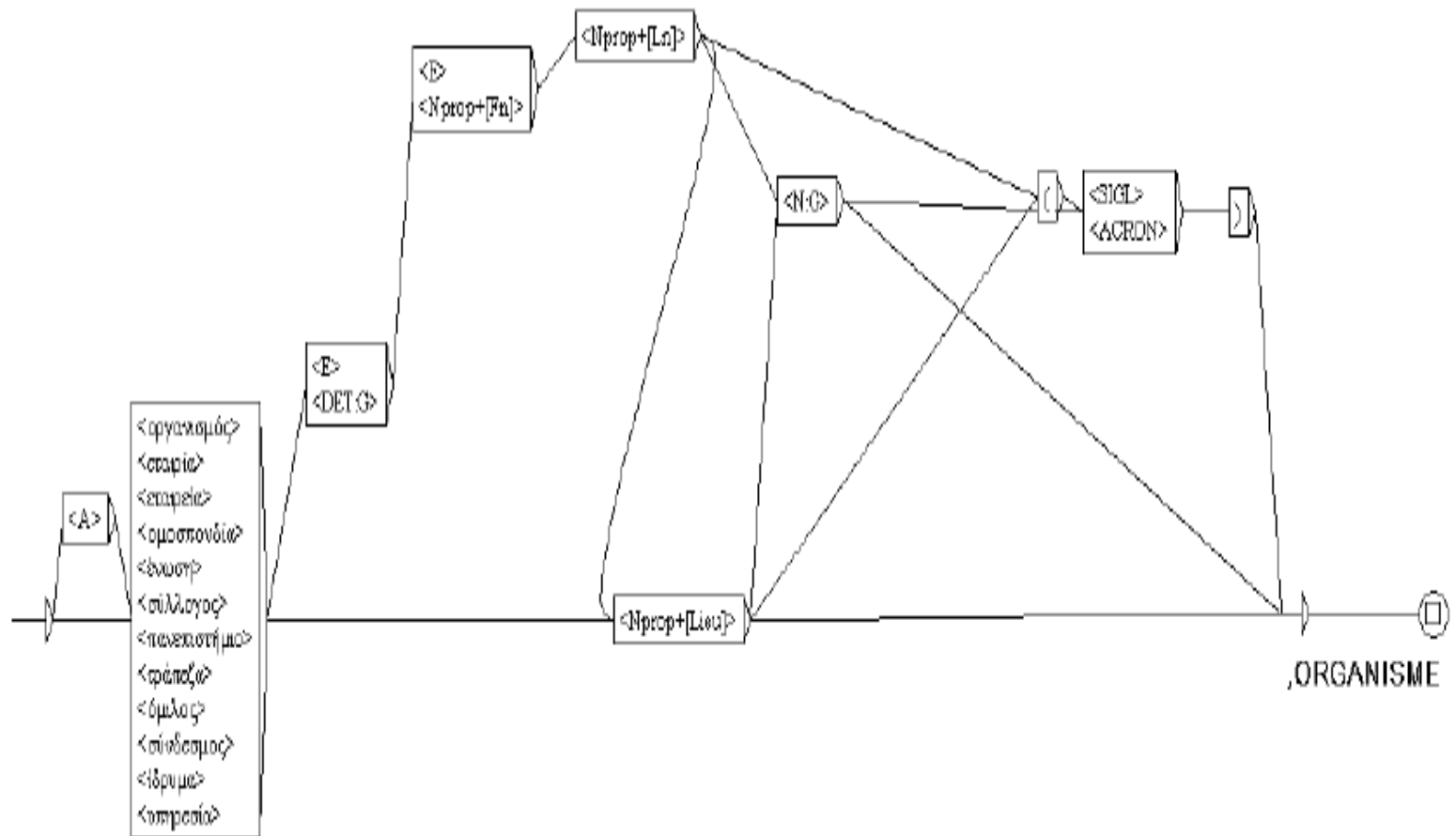
Μεταφραστικές δυσκολίες ονοματικών οντοτήτων

- Το γεγονός όμως ότι δεν υπάρχει ενιαία πρακτική δημιουργεί προβλήματα τόσο στον άνθρωπο-μεταφραστή όσο και στα συστήματα αυτόματης ή μηχανικής μετάφρασης.
- Επίσης οι ΟΟ άλλης γλώσσας μπορούν να εμφανιστούν στη ΝΕ είτε με το λατινικό αλφάβητο είτε με το ελληνικό.

Αυτόματη επεξεργασία ΟΟ

- Έχοντας ως κύριο στόχο τη διαχείριση των ΟΟ μέσω συστημάτων αυτόματης επεξεργασίας φυσικών γλωσσών, προβαίνουμε στην τυποποίηση των δεδομένων για την επίτευξη καλύτερων αποτελεσμάτων.
- Πιο συγκεκριμένα, δημιουργούμε **γράφους** αναγνώρισης ακολουθιών οι οποίοι περιλαμβάνουν όλες τις πιθανές μορφές εμφάνισης μιας δεδομένης ΟΟ
- **Ο γράφος αναγνώρισης και εξαγωγής** των ΟΟ που περιγράφουν οργανισμούς: Οι οργανισμοί συχνά έχουν ως κεφαλή ουσιαστικά όπως *εταιρεία, ομοσπονδία, οργάνωση, ένωση, σύλλογος, ίδρυμα* κ.ά. τα οποία προσδιορίζονται από ένα ή περισσότερα επίθετα (A) και μπορούν να ακολουθούνται από ένα κύριο όνομα τόσο βαφτιστικό όνομα (Nprop+[Fn] όσο και επώνυμο (Nprop+[Ln]), ενδεχομένως από προσδιοριστικό σε γενική (DET:G), από ένα άλλο ουσιαστικό συχνά σε γενική, που πολλές φορές είναι γεωγραφικό όνομα, όπως επίσης και από ένα αρκτικόλεξο (SIGL) ή ακρωνύμιο (ACRON):

Γράφος αναγνώρισης και εξαγωγής οργανισμών



Γράφος αναγνώρισης και εξαγωγής οργανισμών

- Εφαρμόζοντας τον παραπάνω γράφο σε σώμα ελληνικών κειμένων μέσω του αναλυτή κειμένων Unitex (Paumier, 2013) λαμβάνουμε αποτελέσματα με **συμφραστικούς πίνακες**, όπως αυτά που παρατίθενται παρακάτω και τα οποία φέρουν τον χαρακτηρισμό *οργανισμός* (ORGANISME)
- Είναι σημαντικό να αναφέρουμε ότι η αναγνώριση των ακολουθιών στα κείμενα βασίζεται σε μεγάλο βαθμό στην ύπαρξη υπολογιστικών λεξικών

Τα λήμματα του υπολογιστικού λεξικού

- Τα λήμματα του ηλεκτρονικού λεξικού (DelaGR) έχουν την εξής μορφή:
 - *Αύγουστος*, .N+Hum+Nprop+[Fn]:Nms
 - *Αύγουστος*, .N+Nprop+[Ntps]:Nms
 - *Γερμανός*, .N+Hum+Nprop+[Ln]:Nms
 - *Γερμανός*, .N+Hum+Nprop+[Topo]:Nms
 - *Γεωργία*, .N+Hum+Nprop+[Fn]:Nfs:Afs:Vfs
 - *Γεωργία*, .N+Nprop+[Lieu]:Nfs:Afs:Vfs
 - Όπου αρχικό N = ουσιαστικό,
 - **Hum** = ανθρώπινο,
 - **Nprop** = κύριο όνομα,
 - **Fn** = κύριο όνομα βαφτιστικό,
 - **Nms** = ονομαστική αρσενικό ενικός,
 - **Ntps** = ουσιαστικό χρόνου,
 - **Ln** = κύριο όνομα επώνυμο,
 - **Topo** = πατριδωνυμικό ουσιαστικό,
 - **Nfs** = ονομαστική θηλυκό ενικός,
 - **Afs** = αιτιατική θηλυκό ενικός,
 - **Vfs** = κλητική θηλυκό ενικός,

Κανονικοποίηση των ονοματικών οντοτήτων

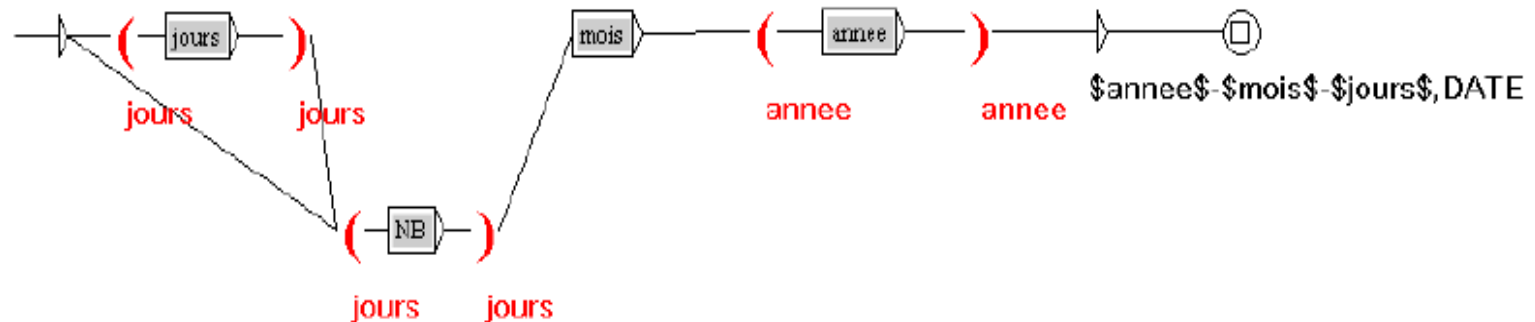
- Για την αποτελεσματικότερη επεξεργασία των ΟΟ, προτείνουμε τη μετατροπή του κειμένου σε δομημένα στοιχεία μέσω της μεθόδου της κανονικοποίησης (normalization).
- Η κανονικοποίηση συνίσταται στην απόδοση μιας μοναδικής αναπαράστασης για κάθε τύπο ονοματικής οντότητας.
- Η κανονικοποίηση κάποιων κατηγοριών ονοματικών οντοτήτων διευκολύνει τη σύγκριση, τη διαλογή και τις έρευνες στα δεδομένα

Κανονικοποίηση των ονοματικών οντοτήτων

- Η κανονικοποίηση των ημερομηνιών επιτυγχάνεται μέσω της μετατροπής όλων των διαφορετικών μορφών εμφάνισης σε μια όμοια μορφή.

Για παράδειγμα, μέσω του παρακάτω γράφου μετατρέπουμε μια ημερομηνία της μορφής 5 Μαρτίου 2013 στη μορφή χρονολογία (annee) – μήνας (mois) – ημέρα (jour)

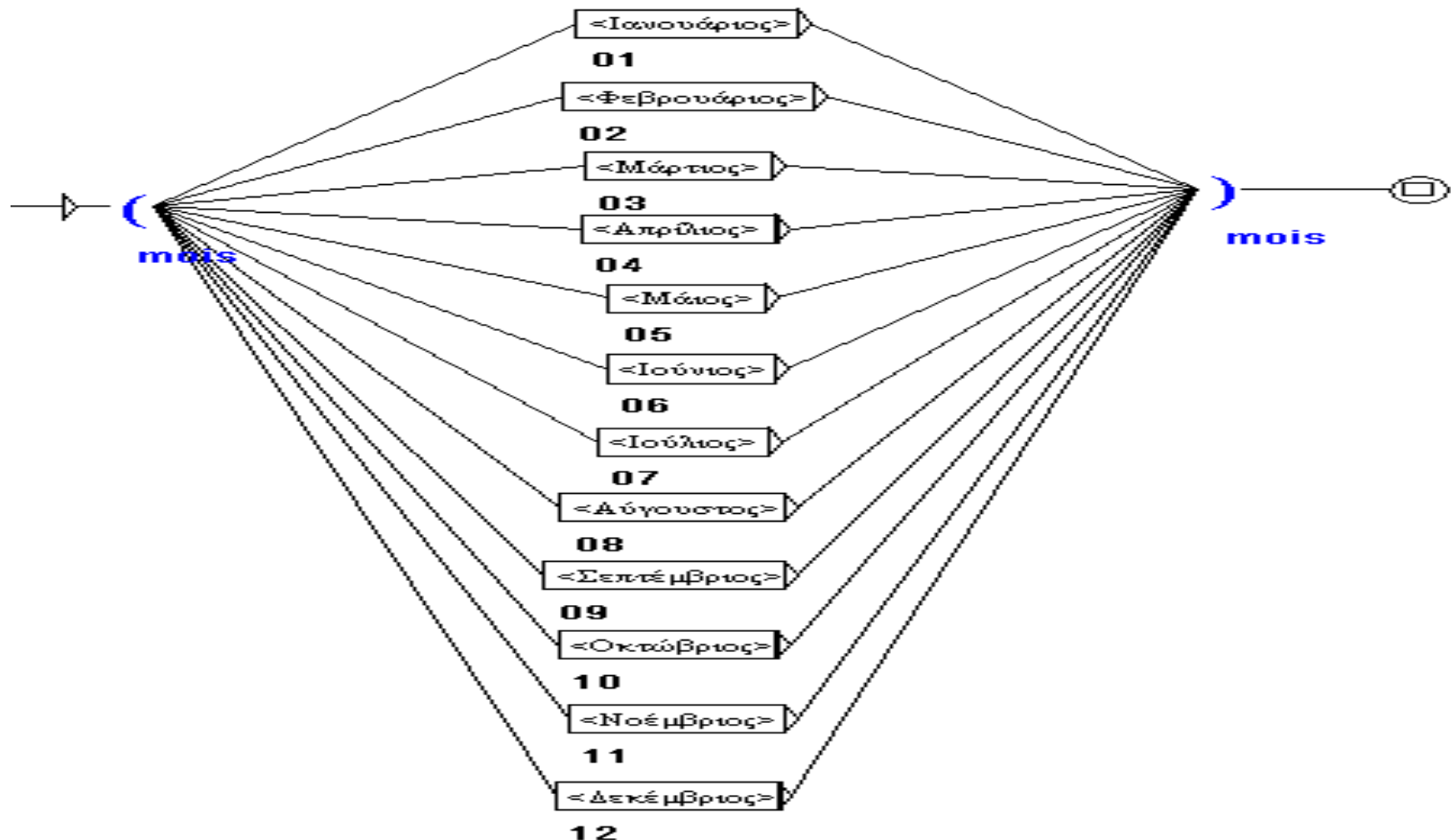
Εξαγωγή και κανονικοποίηση των ημερομηνιών



δηλώνεται η σειρά με την οποία θα εμφανίζονται οι ημερομηνίες κατά την εξαγωγή τους από τα κείμενα, δηλαδή πρώτα η **χρονολογία**, έπειτα ο **μήνας** και στη συνέχεια η **ημέρα**. Με τον παραπάνω γράφο δίνεται και μια επιπλέον πληροφορία ότι η ακολουθία που θα εξαχθεί από το σώμα κειμένων περιγράφει μια ημερομηνία (date).

Εξαγωγή και κανονικοποίηση των ημερομηνιών

- Στον γράφο που ακολουθεί περιέχονται όλες οι πιθανές μορφές εμφάνισης των μηνών, δηλαδή είτε λεκτικά είτε αριθμητικά



Απόσπασμα συμφραστικών πινάκων ημερομηνιών

- Εφαρμόζοντας τον γράφο κανονικοποίησης των ημερομηνιών σε κείμενα της ΝΕ, λαμβάνουμε αποτελέσματα όπως τα παρακάτω:

στις 7 Ιουνίου 2000 2000-06-7,DATE. {S} θέλοντας να συμβάλ
στις 29 Ιουνίου 2002 2002-06-29,DATE. {S} Πριν από την εκτ
x τις 3 Μαρτίου 2003 2003-03-3,DATE η δίκη για την υπόθεση
στις 30 Νοεμβρίου 2002 2002-11-30,DATE, εκείνη την ημέρα
στις 30 Δεκεμβρίου 2002 2002-12-30,DATE, αντιμετώπι με τ
ύ τις 29 Ιουνίου 2002 2002-06-29,DATE, οπότε και άρχισε η
xς {S}1 Οκτωβρίου 1997 1997-10-1,DATE {S}Ο κ. Στάθης Τραπε
J. {S}4 Οκτωβρίου 1997 1997-10-4,DATE {S}Ο κ. Τραπεζαλίδης
στις 24 Απριλίου 1999 1999-04-24,DATE, άλλες 18 οικογένει
στις 3 Μαρτίου 2003 2003-03-3,DATE. {S}Ο Ανέστης Παπαναστ
στις 2 Δεκεμβρίου 2002 2002-12-2,DATE. {S}"Είναι γελοίο ν
ύ τις 20 Ιανουαρίου 2003 2003-01-20,DATE οι τιμές όλων των
ιό τη Δευτέρα 20 Ιανουαρίου 2003 2003-01-20,DATE και μετά
στις 10 Δεκεμβρίου 2003 2003-12-10,DATE {S}ΠΑΡΑΣΚΕΥΗ ΚΑΤΗ
στις 26 Ιουλίου 2002 2002-07-26,DATE, έδωσε τη σπάνια ευκ
στις 16 Μαρτίου 2001 2001-03-16,DATE αναζήτησε τον Α. Οτζ
ίκτη, 16 Μαρτίου 2001 2001-03-16,DATE. {S}Ο δικαστικός επι
στις 11 Σεπτεμβρίου 2001 2001-09-11,DATE, την επιβράδυνση

στις 7 Ιουνίου 2000 2000-06-7,DATE. {S} θέλοντας να συμβάλ
στις 29 Ιουνίου 2002 2002-06-29,DATE. {S} Πριν από την εκτ
x τις 3 Μαρτίου 2003 2003-03-3,DATE η δίκη για την υπόθεση
στις 30 Νοεμβρίου 2002 2002-11-30,DATE, εκείνη την ημέρα
στις 30 Δεκεμβρίου 2002 2002-12-30,DATE, αντιμέτωποι με τ
ό τις 29 Ιουνίου 2002 2002-06-29,DATE, οπότε και άρχισε η
ς {S} 1 Οκτωβρίου 1997 1997-10-1,DATE {S} Ο κ. Στάθης Τραπε
J. {S} 4 Οκτωβρίου 1997 1997-10-4,DATE {S} Ο κ. Τραπεζαλίδης
στις 24 Απριλίου 1999 1999-04-24,DATE, άλλες 18 οικογένει
στις 3 Μαρτίου 2003 2003-03-3,DATE. {S} Ο Ανέστης Παπαναστ
στις 2 Δεκεμβρίου 2002 2002-12-2,DATE. {S} "Είναι γελοίο ν
ό τις 20 Ιανουαρίου 2003 2003-01-20,DATE οι τιμές όλων των
ιό τη Δευτέρα 20 Ιανουαρίου 2003 2003-01-20,DATE και μετά
στις 10 Δεκεμβρίου 2003 2003-12-10,DATE {S} ΠΑΡΑΣΚΕΥΗ ΚΑΤΗ
στις 26 Ιουλίου 2002 2002-07-26,DATE, έδωσε τη σπάνια ευκ
στις 16 Μαρτίου 2001 2001-03-16,DATE αναζήτησε τον Α. Οτζ
ήκη, 16 Μαρτίου 2001 2001-03-16,DATE. {S} Ο δικαστικός επι
στις 11 Σεπτεμβρίου 2001 2001-09-11,DATE, την επιβράδυνση

το σύστημα αναγνώρισε και εξήγαγε τις ημερομηνίες που περιέχονταν στο κείμενο (π.χ. *11 Σεπτεμβρίου 2001, Δευτέρα 20 Ιανουαρίου 2003*), στη συνέχεια προέβη στην κανονικοποίησή τους (δηλαδή στη μετατροπή τους σε όμοια μορφή, π.χ. *2001-09-11, 2003-01-20*) και έπειτα έδωσε και την πληροφορία ότι πρόκειται για ημερομηνία (DATE).

Άλλες πηγές πληροφορίας για τον χρόνο στις ΟΟ

- Απόλυτες ημερομηνίες και χρονικές στιγμές
 - “on April 6, 1995”
- Σχετικές ημερομηνίες και χρονικές στιγμές
 - “last week”
- Χρόνοι ρημάτων
- Γνώση σχετικά με την έμφυτη αλληλουχία των γεγονότων

Μορφοσυντακτικές ιδιότητες ΟΟ

- Σε κάθε οντότητα που αναγνωρίζεται αποδίδεται μορφο-συντακτική πληροφορία ανάλογα με το είδος της
 - Στα noun groups: αριθμός, όνομα, λήμμα
 - Στα verb groups: χρόνος, φωνή, λήμμα
- Αυτή η πληροφορία μπορεί να χρησιμοποιηθεί στα επόμενα στάδια

Σημασιολογικές ιδιότητες ΟΟ

Σε κάθε ονοματική φράση αποδίδεται ένας σημασιολογικός τύπος

```
<np entity="e1"> Sam Schwartz </np>  
<vg> retired </vg>  
as  
<np entity="e2"> executive vice president </np>  
of  
<np entity="e3"> the famous hot dog  
manufacturer </np>,  
<np entity="e4"> Hupplewhite Inc. </np>  
<np entity="e5"> He </np>  
<vg> will be succeeded </vg>  
by  
<np entity="e6"> Harry Himmelfarb </np>.
```

- e1 type: person name: "Sam Schwartz"
- e2 type: position value: "executive vice president"
- e3 type: manufacturer
- e4 type: company name: "Hupplewhite Inc."
- e5 type: person
- e6 type: person name: "Harry Himmelfarb"


ΟΝΟΜΑΤΙΚΕΣ ΟΝΤΟΤΗΤΕΣ ΚΑΙ ΣΥΣΤΗΜΑΤΑ ΕΡΩΤΑΠΟΚΡΙΣΕΩΝ

Ερωτήσεις των οποίων η απάντηση είναι αυστηρά
καθορισμένη (factual questions)

- ο όνομα προσώπου
- ο όνομα οργανισμού
- ο χρονική έκφραση
- ο τοποθεσία
- ο ποσότητα
- ο ορισμός

Εξαγωγή συμβάντων/Σχέσεις ονοματικών οντοτήτων

- Ο εμπλουτισμός μιας τοπικής γραμματικής επιτρέπει την κατασκευή **έμπειρων γραμματικών** (expert grammars) που αναγνωρίζουν πολύπλοκα τεμάχια και ενθαρρύνουν την κατάρτιση συλλογών με πηγές με πιο ευέλικτο και αρθρωτό τρόπο.

- 
- Για παράδειγμα, η ανάλυση της πρότασης:
Ο Αλέξης Τσίπρας δεν θα έχει ακόμα επιστρέψει στις 20 Αυγούστου 2015
 - Σύμφωνα με αυτή τη γραμματική, το όνομα του ατόμου περιέχει το βαφτιστικό και το επώνυμο, η ημερομηνία αποτελείται από μια μέρα και από ένα μήνα και η πρόταση είναι σε αρνητική μορφή.

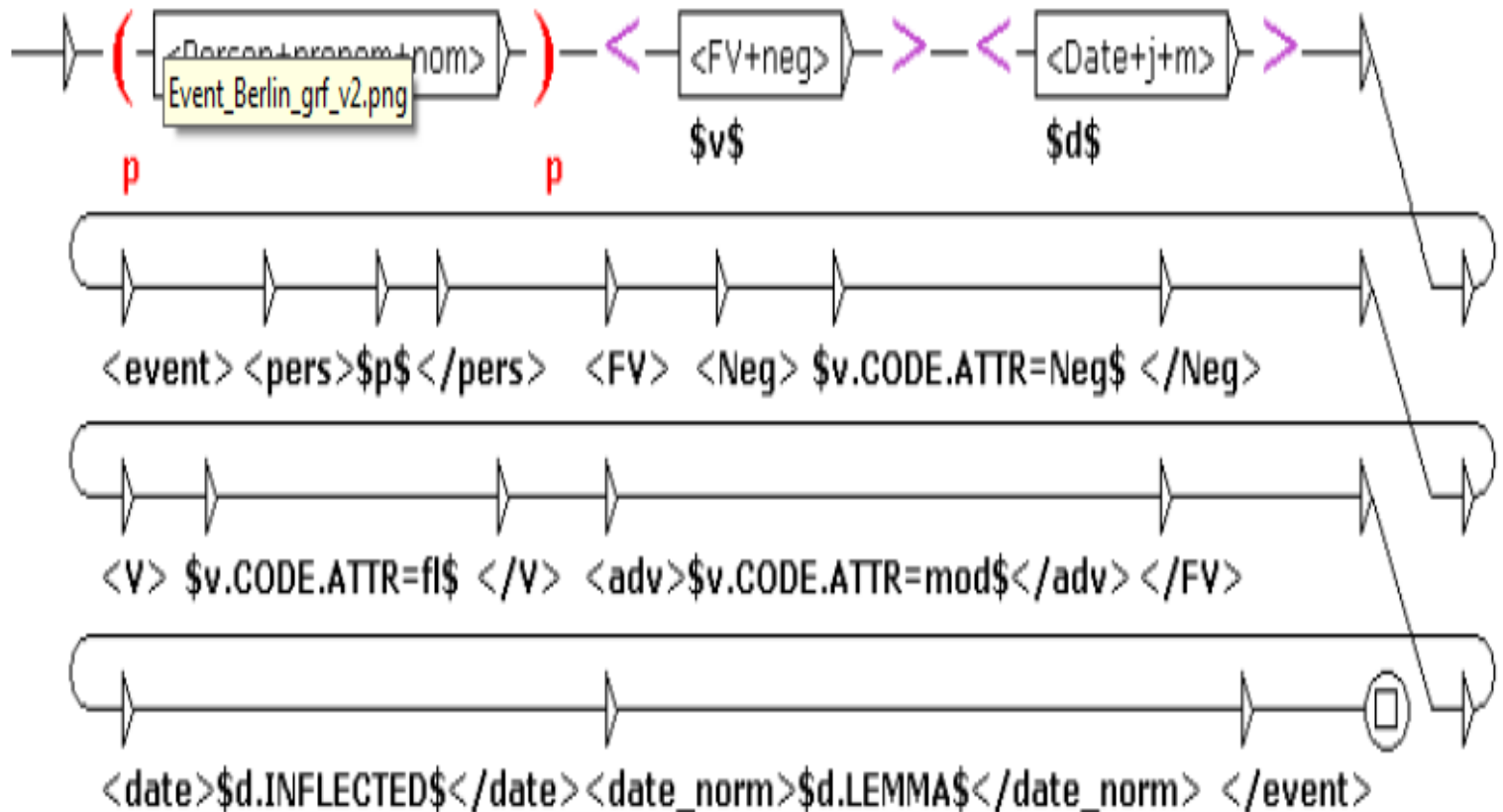
Παράδειγμα γραμματικής

Αλέξης Τσίπρας, Αλέξης Τσίπρας, Person+prenom+prenom_grec+nom+nom_grec+Prenom=Αλέξης+Nom=Τσίπρας:Nms
δεν θα έχει ακόμα επιστρέψει, επιστρέφω.FV+f1=θα έχει επιστρέψει+mod=ακόμα+Neg=δεν+neg
στις 20 Αυγούστου 2015, 2015-08-20.Date+j+m+a

```
<event>
  <pers>Αλέξης Τσίπρας</pers>
  <FV>
    <Neg>δεν</Neg>
    <V>θα έχει επιστρέψει</V>
    <adv>ακόμα</adv>
  </FV>
  <date>στις 20 Αυγούστου 2015</date>
  <date_norm>2015-08-20</date_norm>
</event>
```

Το ίδιο γράφημα δίνει επίσης την πληροφορία ότι το συγκεκριμένο παράδειγμα εκφράζει ένα συμβάν

Παράδειγμα γραμματικής



Σας ευχαριστώ πολύ
για την προσοχή σας

