

## ΚΕΦΑΛΑΙΟ 2

### 2. Επίπεδα ανάλυσης της φυσικής γλώσσας

Για να προωθήσει καλύτερα τους στόχους της η επιστήμη της κλασικής Γλωσσολογίας, όπως όλες οι επιστήμες, αναλύει το αντικείμενο της μελέτης της, την φυσική γλώσσα, σε διακριτά μεταξύ τους επίπεδα, που είναι:

1. Φωνολογικό
2. Μορφολογικό
3. Συντακτικό
4. Σημασιολογικό
5. Πραγματολογικό

Υιοθετώντας αυτή την διαίρεση, η σύγχρονη γλωσσολογία ακολουθεί την παραδοσιακή άποψη. Ωστόσο, μεγάλο μέρος της σύγχρονης έρευνας αφορά τις διεπιφάνειες (*interfaces*) ανάμεσα σε αυτά τα πέντε επίπεδα. Τη διαίρεση αυτή θα υιοθετήσουμε και εμείς εδώ και θα δώσουμε μια πολύ αδρή εικόνα των προβλημάτων που καλείται να αντιμετωπίσει ο θεωρητικός της Υπολογιστικής Γλωσσολογίας (*Computational Linguistics*) και ο γλωσσικός μηχανικός (*language engineer*). Η παρουσίαση διανθίζεται με ασκήσεις είτε επί χάρτου είτε σε γλώσσα Prolog που στόχο έχουν να προθερμάνουν τον αναγνώστη για τα επόμενα κεφάλαια όπου τα ίδια θέματα παρουσιάζονται υπό το φως των ιδεών και των τεχνικών που έχουν αναπτύξει η Υπολογιστική Γλωσσολογία και η Γλωσσική Μηχανική (*Language Engineering*).

#### 2.1 Φωνολογικό επίπεδο

Η γλώσσα πραγματώνεται αρχικά με τον **λόγο** (*speech*). Η **γραφή** έρχεται, αν έρθει ποτέ<sup>1</sup>, δεύτερη και υπολείπεται σε εκφραστικότητα, είναι συντηρητική και περιορισμένης χρήσης (σκεφτείτε την Αγγλική ορθογραφία για παράδειγμα). Ο λόγος παράγεται σε συνεχή ροή. Πραγματικά, είναι ακόμη μεγάλο μυστήριο πως τα μικρά παιδιά ξεχωρίζουν τις λέξεις μέσα στη ροή του λόγου με το οποίο τα περιλούζουν οι μεγάλοι σε εκπληκτικά μικρό χρονικό διάστημα. Ενώ όταν μεγαλώσουν, μαζί με άλλα ωραία πράγματα χάνουν κι αυτό το ταλέντο...

Σε ένα πρώτο επίπεδο ανάλυσης, ο λόγος αποτελείται από ακολουθίες επαναλαμβανόμενων ηχητικών μονάδων τις οποίες αποκαλούμε **φθόγγους**. Για να δείτε τί λέμε στη Γλωσσολογία φθόγγους προφέρετε τις λέξεις “χήρος” και “χώρος” και σκεφθείτε πόσοι είναι οι φθόγγοι που εμφανίζονται και στις δύο λέξεις συνολικά: είναι πέντε ή έξι (το σωστό είναι έξι, η εξήγηση θ’ ακολουθήσει). Το παράδειγμα που μόλις αναφέραμε όμως δείχνει πως η έννοια φθόγγος είναι ανεπαρκής για να καλύψει την κοινή διαίσθηση, δηλαδή ότι κάτι που λέγεται “χι” βρίσκεται στην αρχή των λέξεων “χήρος” και “χώρος”. Για να εκφράσει αυτήν ακριβώς τη διαίσθηση, η Γλωσσολογία χρησιμοποιεί την έννοια του **φωνήματος** (*phoneme*). Το φώνημα είναι μία

<sup>1</sup> Σκεφτείτε πόσες γλώσσες δεν έχουν γραφή. Δεν είναι μόνον οι γλώσσες των aborigines της Αυστραλίας και των Ινδιάνων της Αμερικής (εκεί έχουν εξαφανιστεί και οι ομιλητές της γλώσσας στις περισσότερες περιπτώσεις). Και στα καθ’ ημάς, οι γλώσσες των αρχαιοτέρων κατοίκων της περιοχής δεν είχαν γραφή. Πχ. τα έπη του Ομήρου ήταν δημιούργημα του προφορικού λόγου και της εποχής που η γραφή, και αν υπήρχε ακόμη, ήταν για πολύ περιορισμένη χρήση (πχ. τα λογιστικά του άρχοντα της εποχής).

αφηρημένη φωνολογική οντότητα που έχει διαφοροποιητική αξία μέσα στη γλώσσα. Την διαφοροποιητική τους αξία τη βλέπουμε με την μέθοδο των ελαχίστων ζευγών (*minimal pairs*). Για παράδειγμα συγκρίνετε τα ζεύγη (ζ1) και (ζ2):

- (ζ1) [τ]όπος και [κ]όπος  
(ζ2) μ[έ]λος και Μ[ή]λος

Τα φωνήματα μπορεί να πραγματώνονται στον λόγο από ένα ή περισσότερους φθόγγους. Στην τελευταία περίπτωση μιλάμε για **αλλόφωνα** (*allophones*). Άλλοφωνα του φωνήματος “χι” είναι οι φθόγγοι που βρίσκονται στην αρχή των λέξεων “χήρος” και “χώρος”.

### Ασκηση

A 2.1.1 Γράψτε ένα πρόγραμμα σε Prolog το οποίο να παίρνει λέξεις και να ανακαλύπτει ελάχιστα ζεύγη. Για παράδειγμα, όταν βλέπει τα “κόπος” και “τόπος” να αναγνωρίζει το ελάχιστο ζεύγος <κ,τ> αλλά να απορρίπτει ζεύγη όπως “κόπος” και “κόπτω”.

Σκεφτείτε τώρα την περίπτωση των “κόπος” και “κόπου”. Προφανώς μια απλοϊκή έκδοση του προγράμματός σας θα έβρισκε κάποιο ελάχιστο ζεύγος εδώ. Συγκεντρώστε παρόμοιες περιπτώσεις και γράψτε έτσι τον κώδικα σας ώστε να τις αναγνωρίζει.

Τα φωνήματα και οι φθόγγοι δεν αρκούν για να περιγράψουμε τον λόγο ως φαινόμενο. Σκεφτείτε τη φράση “Ηρθε ο Πέτρος”. Έτσι όπως την γράφουμε δεν λέει και πολλά, θα λέει όμως όταν την εκφέρουμε. Αλήθεια πώς θα θέλατε να την εκφέρετε: ως απλή δήλωση, ως απλή ερώτηση, με θαυμασμό, με απορία, με αγδία, με προσμονή...; Για κάθε μία εκφορά θα χρησιμοποιήσουμε διαφορετικό **επιτονισμό** (*intonation*). Ο επιτονισμός, ο οποίος είναι **υπερτηματικό** (*suprasegmental*) φαινόμενο, αποτελεί οργανικό χαρακτηριστικό του λόγου και πολύ ενδιαφέρον πεδίο μελετών και εφαρμογών.

### Ασκήσεις

A 2.1.2 Ποιά βασικά προβλήματα βλέπετε στην σύνθεση και στην ανάλυση του εν ροή λόγου. Για παράδειγμα, τι είναι πιο δύσκολο: να πετύχουμε, τη σύνθεση ή τη ανάλυση; Σε τί δυσκολεύονται περισσότερο, κατά την γνώμη σας, οι μηχανές που παράγουν φωνή; Στην πραγμάτωση των φθόγγων ή του κατάλληλου επιτονισμού;

A 2.1.3 Μαθαίνετε μια καινούργια γλώσσα. Εκτός και αν είστε ιδιαίτερα προικισμένοι, η εκμάθηση της προφοράς σας δυσκολεύει πολύ). Αν όμως ξεκινούσατε την ίδια διαδικασία στην ηλικία των τριών χρόνων τα πράγματα θα ήσαν απλούστερα. Τι σημαίνει αυτό για τον τρόπο που κατακτάται η γλώσσα από το ανθρώπινο είδος;

## 2.2 Μορφολογικό επίπεδο

Η **Μορφολογία** ασχολείται με τις λέξεις. Υπάρχουν δύο κύριοι τρόποι προσέγγισης: ο **διαδικασιακός** (*procedural*) και ο **δηλωτικός** (*declarative*). Ο διαδικασιακός τρόπος προσέγγισης είναι ίσως πιο κοντά στον τρόπο σκέψης του επιστήμονα που ασχολείται με υπολογιστές και λογισμικό. Το ερώτημα που θέτει είναι πώς παράγονται οι λέξεις και, προφανώς, ανακαλύπτει διαδικασίες για τον σκοπό αυτό. Ο δηλωτικός τρόπος προσέγγισης απαντά στο ερώτημα **ποιές**

λεκτικές δομές είναι σωστές. Δηλαδή, η δηλωτική προσέγγιση διαχωρίζει την περιγραφή του γλωσσικού αντικειμένου από την παραγωγή του ενώ η διαδικασιακή τα αναμιγνύει.

Στην συνέχεια θα αναφερθούμε στις **διαδικασίες παραγωγής λέξεων**, με άλλα λόγια θα προσεγγίσουμε την μορφολογία διαδικασιακά. Ξεκαθαρίζουμε ότι αυτό το κάνουμε μόνο για λόγους παρουσίασης, και όχι για να ταχθούμε υπέρ της μίας ή της άλλης προσέγγισης. Στη συνέχεια θα επανέλθουμε πολλές φορές σ' αυτήν την διάκριση. Διαδικασιακά, λοιπόν, σκεφτόμενοι συνηθίζεται να διακρίνουμε δυό τρόπους για να σχηματίζουμε λέξεις, την **κλίση** και την **παραγωγή**. Σημειώνουμε ότι τα όρια ανάμεσα σε αυτές τις δύο διαδικασίες είναι υποδειγματικά δυσδιάκριτα.

### 2.2.1. Κλίση (Inflection)

Η μορφολογική διαδικασία της κλίσης έχει τα εξής χαρακτηριστικά.

1. Δεν οδηγεί σε αλλαγή του μέρους του λόγου. Για παράδειγμα όταν κλίνουμε την “πηγή” (κατά τα γνωστά, η πηγή, της πηγής, την πηγή, ω πηγή) δεν αλλάζουμε μέρος του λόγου, προφανώς το “ω πηγή” είναι και παραμένει ουσιαστικό.
2. Εφαρμόζεται σε όλα τα μέλη ενός συνόλου με δεδομένα χαρακτηριστικά. Για παράδειγμα, όλα (ή σχεδόν όλα) τα θηλυκά σε -η κλίνονται όπως η “πηγή” και τα περισσότερα έχουν τον ίδιο αριθμό πτώσεων και δύο αριθμούς (ενικό και πληθυντικό).
3. Επιφέρει ελάχιστη και συστηματική αλλαγή στη σημασία της λέξης (πάντα μέσα στο σύστημα μιας συγκεκριμένης γλώσσας). Για παράδειγμα, όταν κλίνουμε ένα ουσιαστικό, έστω “πηγή”, αλλάζουμε τον αριθμό και την πτώση του αλλά συνεχίζουμε να αναφερόμαστε σε όλες εκείνες τις ιδιότητες που χαρακτηρίζουν ένα αντικείμενο ως πηγή

### 2.2.2. Παραγωγή (Derivation)

Σε αντιδιαστολή προς την μορφολογική διαδικασία της κλίσης, η παραγωγή έχει τα εξής χαρακτηριστικά.

1. Συχνά, αλλά όχι πάντα, επιφέρει αλλαγή του μέρους του λόγου. Ιδού μερικές λέξεις που μπορούμε να πάρουμε από το ρήμα “γκρεμίζω” με παραγωγή.

Ρήμα	γκρεμίζω
Ρήμα	ξαναγκρεμίζω
Ομόρριζο ουσιαστικό που δηλώνει το συγκεκριμένο αποτέλεσμα	γκρέμισμα
Ομόρριζο ουσιαστικό που δηλώνει την διαδικασία	γκρέμισμα
Μετοχή που δηλώνει τον τρόπο	γκρεμίζοντας
Μετοχή (ή επίθετο) που δηλώνει κατάσταση (πιθανόν ως αποτέλεσμα της ενεργείας του ρήματος)	γκρεμισμένος

2. Η εφαρμογή της παρουσιάζει κενά. Για παράδειγμα, το ρήμα “γκρεμίζω” δεν σχετίζεται με ομόρριζο ουσιαστικό που δηλώνει τον δράστη. Δείτε τώρα τί γίνεται με το “καταστρέφω” το οποίο είναι σημασιολογικά συναφές προς το “γκρεμίζω”. Το καταστρέφω σχετίζεται με ομόρριζο ουσιαστικό που δηλώνει τον δράστη αλλά δεν έχει αντίστοιχο ουσιαστικό για το συγκεκριμένο αποτέλεσμα.

Ρήμα	καταστρέφω
Ρήμα	ξανακαταστρέφω
Ομόρριζο ουσιαστικό που δηλώνει τον δράστη	καταστροφέας
Ομόρριζο ουσιαστικό που δηλώνει το συγκεκριμένο αποτέλεσμα	; ; ;
Μετοχή που δηλώνει τον τρόπο	καταστρέφοντας
Μετοχή (ή επίθετο) που δηλώνει κατάσταση (πιθανόν ως αποτέλεσμα της ενεργείας του ρήματος)	κατεστραμένος

3. Επιφέρει κατά βάση συστηματική αλλαγή στη σημασία της λέξης αλλά τα προϊόντα της πολύ συχνά **λεξικοποιούνται**, δηλαδή χρησιμοποιούνται (και) με σημασία άλλη από την αρχική. Για παράδειγμα, το επίθημα “-άκι” προσαρτώμενο σε ουσιαστικά δίνει υποκοριστικά (πχ παιδάκι, υπολογιστάκι κλπ). Η λέξη “καλαμάκι” έχει την συστηματική (και συνεπώς προβλεπόμενη) σημασία “μικρό καλάμι” αλλά και την λεξικοποιημένη (και συνεπώς απρόβλεπτη) σημασία “το εξάρτημα που χρησιμοποιούμε για να ρουφήζουμε την πορτοκαλάδα”. Ιδού μερικές παρόμοιες κοινότατες περιπτώσεις: κουτάλα, σουβλάκι, τρυπητό.
4. Σε αντίθεση με την κλίση που δίνει συνήθως συνεχείς λέξεις (υπάρχουν όμως και οι περιφραστικοί χρόνοι, πχ “θα έλθω”), η παραγωγή συχνά δίνει και ασυνεχείς. Παράδειγμα τα “ατομική βόμβα, γαλλικό κλειδί, παιδί-λάστιχο, φακοί επαφής”.

### Ασκηση

#### A 2.2.1.

Γράψτε ένα πρόγραμμα Prolog που να διαβάζει λέξεις όπως “έγραψα”, ”έρραψα”, ”έτριψα” και να αναγνωρίζει τον χρόνο και το λήμμα. Για παράδειγμα

```
: - recover(etripsy,X,Y).
X : past
Y : trivo
```

Το ίδιο πρόγραμμα να λειτουργεί ως **συνθέτης** (generator). Για παράδειγμα

```
: - recover(X,past,trivo).
X : etripsy
```

Προφανώς μπορείτε να έχετε μία βάση δεδομένων με γεγονότα όπως

```
is_tense(etripsy,past,trivo).
```

Αυτό όμως είναι μάλλον αδιάφορο. Πιό ενδιαφέρον είναι να έχετε μία βάση δεδομένων με γεγονότα όπως

```
verb(trivo)
```

και κάποιον κανόνα που να λέει ότι τα ρήματα σε “-βω” και “-φω” έχουν αόριστο σε “-ψα” κλπ.

Φτιάξτε ένα τέτοιο εργαλείο και χρησιμοποιείστε το τόσο για αναγνώριση και για σύνθεση.

## 2.3 Συντακτικό Επίπεδο

Το **Συντακτικό** ασχολείται με τις **φράσεις** και τις **προτάσεις**. Γενικά, όλες οι γραμματικές δίνουν **κανόνες φραστικής δομής** (*phrase structure rules*). Κανόνες δηλαδή που δηλώνουν πώς συντάσσεται μια φράση. Οι κανόνες αυτοί είναι **συμφραστικώς ανεξάρτητοι** (*context free rules*). Μέχρι πολύ πρόσφατα ήταν γενικώς αποδεκτό πως η φυσική γλώσσα περιγράφεται (ή παράγεται, ή τεχνολογείται) επαρκώς από συμφραστικώς ανεξάρτητους κανόνες. Σήμερα έχουν βρεθεί φαινόμενα της φυσικής γλώσσας που απαιτούν συμφραστικώς **εξαρτημένους κανόνες** (*context sensitive rules*) για την **παραγωγή** ή **τεχνολόγηση** ή **περιγραφή** τους.

Παραθέτουμε ένα σύνολο κανόνων για τα Ελληνικά και τα Αγγλικά για ένα απόσπασμα μιας Ελληνικής και μιας Αγγλικής γραμματικής το οποίο μπορεί να παράγει ή να τεχνολογήσει ή να περιγράψει φράσεις όπως τις Π.2.3.1. και Π.2.3.2.

Π.2.3.1. *Ένας σκύλος κοιτάζει μία μικρή γάτα*  
*A dog looks at a small cat.*

Π.2.3.2. *H* άποψη ότι ένα μικρό γκρίζο ποντίκι τρομάζει την γάτα ακούγεται περίεργη.  
*The view that a small grey mouse frightens the cat sounds odd.*

$\Pi \rightarrow O\Phi \ P\Phi$   
 $S \rightarrow NP \ VP$

$O\Phi \rightarrow AP\Theta \ E^* \ O$   
 $NP \rightarrow Det \ A^* \ N$

$O\Phi \rightarrow O\Phi \ óti \ \Pi$   
 $NP \rightarrow NP \ that \ S$

$O\Phi \rightarrow O$   
 $NP \rightarrow N$

$P\Phi \rightarrow P \ O\Phi$   
 $VP \rightarrow V \ NP$

$AP\Theta \rightarrow$  ένας, μία, ένα, η  
 $Det \rightarrow$  a, the

$O \rightarrow$  σκύλος, γάτα, ποντίκι, άποψη  
 $N \rightarrow$  dog, cat, mouse, view

$E \rightarrow$  άσπρος, ραβδωτή, μικρή, μικρό, γκρίζο, περίεργη  
 $A \rightarrow$  white, strapped, small, grey, odd

$P \rightarrow$  κυνηγά, κοιτάζει, ακούγεται, τρομάζει  
 $V \rightarrow$  is-after, looks-at, sounds, frightens

όπου:

$\Pi$ : Πρόταση,

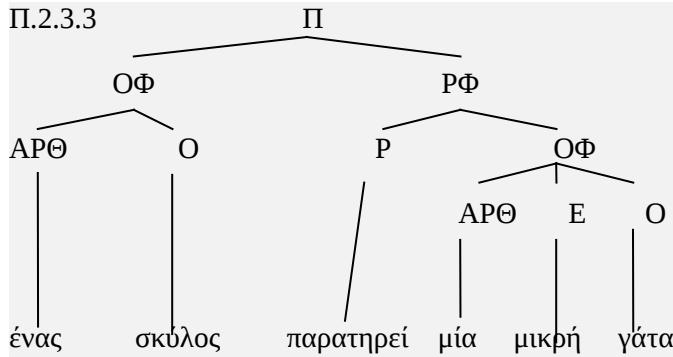
$S$ : sentence

$O\Phi$ : ονοματική φράση,  
 $P\Phi$ : ρηματική φράση,  
 $AP\Theta$ : άρθρο,  
 $E$ : επίθετο,  
 $O$ : όνομα,  
 $P$ : ρήμα

$NP$ : noun phrase  
 $VP$ : verb phrase  
 $Det$ : determiner  
 $A$ : adjective  
 $N$ : noun  
 $V$ : verb

Συνηθίζεται να χρησιμοποιούνται δένδρα ή αγκύλες για να παρασταθεί η φραστική δομή. Και οι δύο παραστάσεις βρίσκονται σε αντιστοιχία με τους κανόνες φραστικής δομής. Για παράδειγμα, η φράση Π.2.3.1. παριστάνεται ως δένδρο με το Π.2.3.3 και με αγκύλες στο Π.2.3.4.

Π.2.3.3



Π.2.3.4. [[ένας [σκύλος]] [παρατηρεί [μία μικρή [γάτα]]]]

Χρησιμοποιούμε τους κανόνες φραστικής δομής για να **τεχνολογήσουμε** (*parse*) ή να **συνθέσουμε** (*generate*) φράσεις και προτάσεις. Στα επόμενα θα ασχοληθούμε κυρίως με την **τεχνολόγηση** (*parsing*) αλλά θα αναφερόμαστε στην **σύνθεση κειμένου** (*generation*) όποτε μας είναι βολικό.

Θα έχετε ήδη αντιληφθεί ότι στη σύνταξη ξαναβρίσκουμε την αντίθεση ανάμεσα στην διαδικασιακή και στην δηλωτική προσέγγιση. Διαδικασιακά οι κανόνες φραστικής δομής είναι δυνατόν να ερμηνευτούν ως κανόνες *επαναγραφής* (*rewrite rules*) ή ως *περιορισμοί* (*constraints*) στην φραστική δομή. Ίσως δεν είναι φανερό με την πρώτη ματιά, αλλά η διαφορά είναι σημαντική. Και να γιατί:

Η διαδικασιακή προσέγγιση δημιουργεί το πρόβλημα της *αριστερής αναδρομής* (θα αναφερθούμε σ' αυτήν σε επόμενο κεφάλαιο). Αριστερή αναδρομή προκύπτει όταν ερμηνεύουμε διαδικασιακά κανόνες φραστικής δομής όπως τον Π.2.3.5, όταν δηλαδή λέμε ότι ο κανόνας αυτός *παράγει* φράσεις.

Π.2.3.5  $A \rightarrow A B \dots$

Και φυσικά χρειαζόμαστε κανόνες που έχουν τη μορφή του Π.2.3.5 όταν ασχολούμαστε με τη φυσική γλώσσα. Σκεφθείτε τι μορφή μπορεί να έχουν οι κανόνες που χρειάζονται για δομές όπως “το παιδί της μαμάς” και “μικρό αθώο πουλάκι”. Στην πρώτη περίπτωση χρειάζεται ο κανόνας  $O\Phi \rightarrow O\Phi \ O\Phi$  και στη δεύτερη  $E\Phi \rightarrow E\Phi$  (όπου  $E\Phi$  : επιθετική φράση). Τέτοιας μορφής κανόνες, όμως, όταν τους ερμηνεύσουμε διαδικασιακά, δηλαδή ως κανόνες παραγωγής, οδηγούν σε ατέρμονες βρόχους. ‘Οχι όμως και όταν τους ερμηνεύσουμε δηλωτικά γιατί τότε δεν θεωρούμε ότι υπάρχει μια μηχανή που αρχίζει και ξαναρχίζει από την αρχή τον ίδιο κανόνα χωρίς να βγαίνει από αυτόν ποτέ. Αυτό που κρατάμε για την ώρα είναι ότι με τον κανόνα φραστικής

δομής αυτόν έχουμε μια απλή περιγραφή η οποία καθορίζει ότι μία γραμματική ΟΦ μπορεί να συνισταται από δύο ΟΦ στην σειρά.

Αν ακόμη διερωτάστε γιατί αυτή η διάκριση ανάμεσα σε διαδικασιακές και δηλωτικές ερμηνείες των γραμματικών, σάς προτείνουμε το ακόλουθο σενάριο.

Φανταστείτε έναν γλωσσολόγο που με την πιρόγα του κατεβαίνει τον Αμαζόνιο, πιάνει φιλίες με μία πολύ εξωτική φυλή και καταγράφει τη γλώσσα της, δηλαδή φωνολογία, λεξιλόγιο, κανόνες και όλα τα συναφή. Ας πούμε ότι η γλώσσα αυτή είναι τόσο εξωτική όσο τα Ελληνικά και έχει δομές σαν και “το παιδί της μαμάς”. Τί πιστεύετε ότι θα καταγράψει αυτός ο γλωσσολόγος; Μία γραμματική που δεν συμφωνεί με τον κοινό νου μόνο και μόνο για να αποφύγει την αριστερή αναδρομή;

Αυτό ο γλωσσολόγος το αφήνει στον μηχανικό που θα κάνει την γραμματική αυτή πρόγραμμα. Με άλλα λόγια, οι δύο γλώσσες, αυτή της γλωσσολογικής περιγραφής και αυτή του προγράμματος, αν και είναι και οι δύο συμφραστικώς ανεξάρτητες δεν μοιάζουν πάντα. Δεν μπορούμε και δεν θέλουμε να θυσιάσουμε την μία για την άλλη.

## Ασκήσεις

### A 2.3.1

Η γραμματική του παραδείγματος Π.2.3.1. υπερ-συνθέτει (*overgenerates*), δηλαδή, παράγει και γραμματικές και μη γραμματικές προτάσεις<sup>2</sup>. Δώστε τουλάχιστον δύο μη γραμματικές προτάσεις που είναι δυνατόν να παραχθούν με αυτήν την γραμματική και εξηγείστε πού βρίσκεται το πρόβλημα.

### A 2.3.2

Στόχος αυτής της άσκησης είναι να μας εξοικειώσει με το φαινόμενο της δομικής αμφισημίας (*structural ambiguity*). Προσθέστε στη γραμματική που δώσαμε πιο πάνω τους κανόνες:

ΠΡΦ	→ ΠΡ ΟΦ
ΟΦ	→ ΑΡΘ Ο ΠΡΦ
ΡΦ	→ Ρ ΟΦ ΠΡΦ
ΠΡ	→ με, από, χωρίς
Ο	→ κυάλια, κύριος
ΑΡΘ	→ τα, τον

Πόσα δέντρα είναι δυνατόν να αντιστοιχίσετε σε προτάσεις όπως  
“είδα τον κύριο με τα κυάλια”

### A 2.3.3

Γράψτε έναν τεχνολογητή σε PROLOG που να αναγνωρίζει προτάσεις σαν την Π.2.3.1 και την Π.2.3.2. Ο τεχνολογητής σας να έχει κανόνες του τύπου:

```
s(L):- np(L1), vp(L2), append(L1,L2,L).  
noun([dog]).
```

όπου *L* το αλφαριθμητικό εισαγωγής (δηλαδή, οι προτάσεις σας) γραμμένο σαν λίστα λέξεων. Μπορεί ο τεχνολογητής σας να λειτουργήσει και ως συνθέτης; Είναι ο τεχνολογητής σας γρήγορος ή επιδέχεται βελτίωσης;

<sup>2</sup> Γραμματικές είναι οι προτάσεις που θεωρούνται ορθές από τους φυσικούς ομιλητές της γλώσσας.

## 2.4 Σημασιολογικό επίπεδο

Η φυσική γλώσσα, σαν συμβολικό σύστημα που είναι, δεν έχει μόνο μορφή, δηλ. μορφολογία και σύνταξη, αλλά και **σημασία**. Η μελέτη της σημασίας της φυσικής γλώσσας (δυστυχώς ή ευτυχώς) μάς φέρνει αντιμέτωπους με δύσκολα, φιλοσοφικά εν κατακλείδι, ερωτήματα όπως κατά πόσον η γλώσσα αντανακλά μια πραγματικότητα έξω από τον άνθρωπο, μια αντικειμενική ας πούμε πραγματικότητα, ή κάτι τι που έχει φιλτραριστεί μέσα από τον ανθρώπινο εγκέφαλο, δηλ. μια υποκειμενική πραγματικότητα. Προφανώς, αυτά είναι ερωτήματα ανοιχτά προς συζήτηση και δεν θα μας απασχολήσουν. Εδώ θα περιοριστούμε σε μια μόνον άποψη και θα ασχοληθούμε κυρίως με το τί μπορούμε να κάνουμε στην PROLOG με αυτήν μας την επιλογή. Εις το εξής, λοιπόν, αποδεχόμαστε ότι η γλώσσα έχει αντικειμενική σημασία η οποία μπορεί να περιγραφεί με λογικά συστήματα που στηρίζονται στην **αναφορά** και στις **τιμές αληθείας**. Θα κάνουμε μία ακόμη επιλογή. Θα χρησιμοποιήσουμε την θεωρία των συνόλων για να ερμηνεύσουμε την έννοια “αναφορά”.

### 2.4.1. Αναφορά και μοντέλα

**Αναφορά** είναι η ιδιότητα που έχουν εκφράσεις της φυσικής γλώσσας να εντοπίζουν ένα σύνολο από οντότητες σε έναν συγκεκριμένο κόσμο. Πάντα ορίζουμε έναν συγκεκριμένο κόσμο για να μελετήσουμε σημασίες και για την ακρίβεια κατασκευάζουμε ένα μαθηματικό αντικείμενο το οποίο το λέμε **μοντέλο** (*model*) το οποίο θεωρούμε ότι απεικονίζει τον κόσμο για τον οποίο ενδιαφερόμαστε. Ένα μοντέλο συνίσταται από το σύνολο **O** των οντοτήτων που περιέχει και από μια συνάρτηση, την **ερμηνευτική συνάρτηση** (*interpretation function*) **E**, η οποία απεικονίζει εκφράσεις σε σύνολα.

Για παράδειγμα, ας πούμε ότι ενδιαφερόμαστε για τον κόσμο που ορίζεται από τον χώρο στον οποίο βρίσκεστε τώρα που διαβάζετε αυτές τις γραμμές και ας ορίσουμε ένα μοντέλο που απεικονίζει αυτόν τον κόσμο. Σ' αυτό το μοντέλο, το σύνολο των οντοτήτων που περιγράφονται ως “άντρες” μπορεί να έχει μηδέν, ένα ή περισσότερα μέλη, για την περίπτωση θα πούμε ότι έχει τρία μέλη. Λέμε ότι η σημασία της έκφρασης άντρας στο συγκεκριμένο μοντέλο είναι αυτό το τριμελές σύνολο των αντρών.

Θα αρχίσουμε τώρα να εισάγουμε και τον συμβολισμό που θα χρησιμοποιούμε εφεξής. Θα χρησιμοποιούμε καλλιγραφικούς χαρακτήρες για τις εκφράσεις και **μαύρους** χαρακτήρες για τις οντότητες που εμπλέκονται στην σημασία των εκφράσεων. Επίσης θα γράφουμε [[έκφραση]] όταν μιλάμε για τη σημασία μιας έκφρασης (και όχι την ίδια την έκφραση). Προφανώς, ο συμβολισμός [[A]] είναι ισοδύναμος με τον συμβολισμό E(A), δηλ. η σημασία μιας έκφρασης είναι η τιμή της Ερμηνευτικής Συνάρτησης για αυτήν την έκφραση. Για παράδειγμα, στο μοντέλο που ορίσαμε γράφουμε:

$$[[\text{άντρας}]] = \{ \text{άντρας1}, \text{άντρας2}, \text{άντρας3} \}$$

Σ' αυτό το μοντέλο δίνουμε και τη σημασία των εκφράσεων γυναίκα, βιβλίο, εφημερίδα.

$$\begin{aligned} [[\text{γυναίκα}]] &= \{ \text{γυναίκα2}, \text{γυναίκα3} \} \\ [[\text{βιβλίο}]] &= \{ \text{βιβλίο1}, \text{βιβλίο2}, \text{βιβλίο3}, \text{βιβλίο4} \} \\ [[\text{εφημερίδα}]] &= \{ \text{εφημερίδα1} \} \end{aligned}$$

Ποιά είναι τότε η σημασία των ρημάτων, πχ. *κοιμάμαι*, *διαβάζω*? Από αυτά τα μονοδύναμα ρήματα, πχ. *το κοιμάμαι*, αναφέρονται σε σύνολα οντοτήτων. Τα πολυδύναμα, πχ. *το μεταβατικό*

διαβάζω αναφέρονται σε σύνολα απαρτιζόμενα από πλειάδες οντοτήτων. Υποθέτουμε ότι τα παραδείγματα που ακολουθούν προκύπτουν από το μοντέλο που χρησιμοποιήσαμε και πιο πάνω.

$$\begin{aligned} [[\text{κοιμάται}]] &= \{\text{άντρας1}, \text{άντρας3}\} \\ [[\text{διαβάζει}]] &= \{\langle \text{άντρας1}, \text{βιβλίο2} \rangle, \langle \text{γυναίκα3}, \text{εφημερίδα1} \rangle\} \\ [[\text{δίνει}]] &= \{\langle \text{άντρας1}, \text{άντρας2}, \text{βιβλίο2} \rangle, \langle \text{γυναίκα3}, \text{άντρας3}, \text{εφημερίδα1} \rangle\} \end{aligned}$$

Και η αναφορά των προτάσεων τί είναι; Δηλαδή σε τι αναφέρεται η πρόταση “ένας άντρας διαβάζει ένα βιβλίο”; Οι προτάσεις δεν αναφέρονται σε σύνολα αλλά σε **τιμές αληθείας** οι οποίες είναι δύο – ας πούμε Αλήθεια και Ψέμα. Θα μπορούσαμε βέβαια να χρησιμοποιήσουμε όποιο άλλο όνομα μας αρέσει, π.χ. Ολυμπιακός και Παναθηναϊκός αρκεί να τηρήσουμε τη συμφωνία ότι η μία από τις δύο τιμές, εδώ η Αλήθεια, θα είναι αναφορά των προτάσεων οι οποίες είτε περιγράφουν σχέσεις που υπάρχουν στο μοντέλο μας ή η τιμή αληθείας τους μπορεί να υπολογισθεί με βάση τους λογικούς συνδέσμους (σύζευξη, άρνηση και τα παράγωγά τους). Βρίσκουμε δηλαδή την αναφορά όλων των επιμέρους εκφράσεων που αποτελούν την πρόταση, τις **συνδυάζουμε** με τον τρόπο που επιβάλει η πρόταση και βλέπουμε άνη σχέση που προκύπτει διακατολογείται από το μοντέλο μας. Δείτε τώρα τα παραδείγματα που ακολουθούν και που στηρίζονται στις σημασίες που δώσαμε μέχρι τώρα.

$$[[\text{ένας άντρας διαβάζει ένα βιβλίο}]] = \text{Αλήθεια}$$

γιατί η πλειάδα **<άντρας1, βιβλίο2>** είναι μέλος του συνόλου-αναφορά του διαβάζει.

$$[[\text{ένας άντρας διαβάζει μια εφημερίδα}]] = \text{Ψέμα}$$

επειδή δεν υπάρχει η κατάλληλη πλειάδα στο σύνολο-αναφορά του διαβάζει.

Η σημασία της πρότασης, είπαμε, προκύπτει από τον συνδυασμό των σημασιών των μερών της, είναι με άλλα λόγια συνάρτηση των επιμέρους σημασιών. Αυτή η ιδιότητα είναι γνωστή ως **συνθετικότητα** (*compositionality*).

Η χρήση μοντέλων για τον υπολογισμό της σημασίας των εκφράσεων της φυσικής γλώσσας έλκει την καταγωγή της από τη σημασιολογία των τυπικών γλωσσών (*formal languages*) που χρησιμοποιούνται στους υπολογιστές. Στην πράξη ένα μοντέλο είναι δομή δεδομένων που περιέχει οντότητες και σχέσεις μεταξύ τους. Η φυσική γλώσσα μπορεί να χρησιμοποιηθεί για να διερευνηθεί το μοντέλο και για να ανακτηθεί πληροφορία από αυτό. Σε επόμενο κεφάλαιο θα δείξουμε μερικές μεθόδους που να υπηρετούν τον σκοπό αυτόν.

## Aσκήσεις

### A 2.4.1.

Δώστε το μοντέλο που χρησιμοποιήθηκε στην παράγραφο 2.4.1. (δηλαδή το σύνολο των οντοτήτων **O** και την ερμηνευτική συνάρτηση **E**).

### A 2.4.2.

Υποθέστε πως η αναφορά των κυρίων ονομάτων είναι μονοσύνολα, πχ.

$$[[\text{Γιάννης}]] = \{\text{άντρας1}\}.$$

Προσαρμόστε το μοντέλο της άσκησης A 2.4.1. έτσι ώστε οι ακόλουθες προτάσεις να έχουν τιμή αληθείας “Αλήθεια”.

Ο Γιάννης έδωσε το βιβλίο στην Ελένη.

Ένας άντρας έδωσε ένα βιβλίο σε μία γυναίκα

Η Μαρία διαβάζει ένα βιβλίο.

*Ο Γιάννης ονειρεύται.*

*Όλοι οι άντρες ονειρεύονται.*

*Τι πράξεις πάνω στα σύνολα χρησιμοποιείτε για να υπολογίσετε την τιμή αληθείας των προτάσεων;*

#### A 2.4.3.

*Γράψτε ένα πρόγραμμα prolog το οποίο να αποδίδει τιμές αληθείας σε προτάσεις με μονοδύναμα ρήματα, π.χ.*

*Ένας άντρας κοιμάται.*

*Δεν χρειάζεται να γίνεται τεχνολόγηση της πρότασης. Υποθέστε κανόνες της μορφής:*

*s(NPsem,VPsem,Ssem):- np(NPsem),vp(VPsem),member(NPsem,VPsem).*

### 2.4.2. Κατηγορικός Λογισμός

Στην παράγραφο 2.4.1 ορίσαμε ένα μοντέλο και μια ερμηνευτική συνάρτηση, ας την πούμε **E**. Η **E** απεικονίζει εκφράσεις της φυσικής γλώσσας (ονόματα, ρήματα, προτάσεις) σε σύνολα και τιμές αληθείας. Δηλαδή, στην παράγραφο 2.4.1. υπολογίσαμε τη σημασία των εκφράσεων της φυσικής γλώσσας με μία κατευθείαν απεικόνιση από την φυσική γλώσσα πρός το μοντέλο.

Ίσως η κατευθείαν απεικόνιση να φαίνεται σύμφωνη με την διαίσθησή μας περί σημασίας της φυσικής γλώσσας. Δεν χρησιμοποιούμε όμως πάντα αυτήν την πρακτική. Αυτό που συνήθως κάνουμε είναι να αναπαριστάνουμε τη σημασία της φυσικής γλώσσας με μια ενδιάμεση τεχνητή γλώσσα. Μετά, υπολογίζουμε την σημασία των εκφράσεων της τεχνητής γλώσσας. Προφανώς, φροντίζουμε ώστε η τεχνητή γλώσσα που διαλέγουμε να έχει επιθυμητές (μαθηματικές) ιδιότητες. Η ενδιάμεση τεχνητή γλώσσα που και χρησιμοποιείται συχνά είναι ο **κατηγορικός λογισμός πρώτης τάξης** (*first order predicate calculus*).

Γιατί όλη αυτή φασαρία; Για έναν βασικά λόγο: για να μπορούμε να είμαστε σίγουροι ότι τα συμπεράσματά μας είναι αξιόπιστα. Ας πιάσουμε τα πράγματα από την αρχή.

Γιατί θέλουμε να βρούμε την σημασία των εκφράσεων της φυσικής γλώσσας; Γιατί θέλουμε να προσεγγίσουμε τον τρόπο με τον οποίο οι άνθρωποι, με τη γλώσσα, συνάγουν σωστά συμπεράσματα από τα εξακριβωμένα δεδομένα τα οποία έχουν. Για παράδειγμα, όταν ξέρουμε ότι ο Κώστας έχει τρία παιδιά ξέρουμε ταυτόχρονα ότι ο Κώστας έχει ένα τουλάχιστον παιδί. Ή, όταν ξέρουμε ότι όλοι οι σκύλοι αγαπούν τα κόκκαλα ξέρουμε ότι και ο δικός μας Κανέλλης (σκύλος είναι ο Κανέλλης) αγαπάει τα κόκκαλα. Με άλλα λόγια, θέλουμε να έχουμε μηχανισμούς για να “συλλογιζόμαστε” τέτοιους που να είμαστε σίγουροι ότι όταν τους τροφοδοτούμε με εξακριβωμένα δεδομένα μας επιστρέφουν σωστά συμπεράσματα. Η φυσική γλώσσα καθαυτή δεν μας προσφέρει αυτή τη δυνατότητα. Ως συμβολικό σύστημα γενικά δεν μας είναι μαθηματικά περιγράψιμη. Γιαυτό τον λόγο χρησιμοποιούμε την αναπαράσταση σε μία απολύτως κατανοητή και ελεγχόμενη γλώσσα όπως ο κατηγορικός λογισμός πρώτης τάξης.

### Ασκήσεις

#### A 2.4.4.

*Δώστε την αναπαράσταση σε κατηγορικό λογισμό της σημασίας των φράσεων:*

*Διαβάζει ο Πέτρος την Ιλιάδα.*

*Διαβάζει ο Πέτρος την Ιλιάδα;*

*Διαβάζει ο Πέτρος την Ιλιάδα!*

*(Συζητείστε το πώς φτάσατε σε αυτήν).*

#### A 2.4.5.

Αν υποθέσουμε ότι ισχύουν οι πιό κάτω αντιστοιχίες ανάμεσα σε εκφράσεις της φυσικής γλώσσας και στον κατηγορικό λογισμό (οι εκφράσεις του κατηγορικού λογισμού δίνονται με λατινικούς χαρακτήρες).

Αζώρ	azor
Πέτρος	petros
Ελένη	eleni
παίζει	(x,paizei(x))
είδε	(y,x,eide(x,y))
έδωσε	(z,y,x,edose(x,y,z))

Γράψτε ένα πρόγραμμα Prolog το οποίο να παράγει αντιστοιχίες σαν την επόμενη.

Ο Αζώρ παίζει paizei(azor)

## 2.5. Πραγματολογικό Επίπεδο

Υπάρχουν όμως πολλές γλωσσικές σημασίες που δεν έχει κατορθώσει να εκφράσει η σημασιολογία που στηρίζεται στην αναφορά και στις τιμές αληθείας. Για παράδειγμα, οι ερωτήσεις, οι διαταγές, οι παρακλήσεις, οι ευχές, δεν έχουν ακόμη βρεί τα σημασιολογικά τους αντίστοιχα σε συστήματα που στηρίζονται στην συνθετικότητα και στις τιμές αληθείας. Γιαυτές τις “δύσκολες” περιπτώσεις υπάρχει το λεγόμενο πραγματολογικό επίπεδο. Πάντως τα όρια ανάμεσα στην Σημασιολογία και στην **Πραγματολογία** είναι εξαιρετικά ασαφή. Η αναφορά μας στην πραγματολογία είναι (όπως άλλωστε βλέπετε) τηλεγραφική και δεν θα επανέλθουμε. Οφείλουμε όμως να σημειώσουμε πως ο κλάδος γνωρίζει άνθηση σήμερα.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

Κλασικές εισαγωγές στην Υπολογιστική Γλωσσολογία είναι τα βιβλία των Allen (1995), Gazdar and Mellish (1989), Covington (1994).

Η χρήση της Prolog ως υπολογιστικής γλώσσας κατεξοχήν κατάλληλης για την Επεξεργασία της Φυσικής Γλώσσας έχει πλέον καθιερωθεί. Τα βιβλία των Bratko (1986) και Covington (1996) αποτελούν κλασικές αναφορές.

Όσον αφορά την Γλωσσολογία μπορείτε να αναφερθείτε στις Θεοφανοπούλου-Κοντού (1989) και Philippaki-Warburton (1992) για μια εισαγωγή στην Μετασχηματιστική Θεωρία αλλά και στις βασικές έννοιες που δόθηκαν εδώ. Επιπλέον, τα βιβλία αυτά συζητούν χαρακτηριστικά της Ελληνικής γλώσσας. Κλασική εισαγωγή στη Μορφολογία είναι ο Spencer (1991) και στη Γραμματική της Φραστικής Δομής ο Borsley (1996). Ο τελευταίος ιδίως κάνει εκτεταμένη συζήτηση των χαρακτηριστικών των διαφόρων φορμαλισμών και της συμπεριφοράς τους στις εφαρμογές.

Κλασική εισαγωγή στη Σημασιολογία αποτελούν οι Allwood et al (198?) και Chierchia and McConnell-Ginet (1990). Η ιδιαιτερότητα αυτών των βιβλίων έγκειται στο ότι παρουσιάζουν ακριβώς τα ζητήματα που ανακύπτουν από την εφαρμογή της τυπικής σημασιολογίας στην φυσική γλώσσα.

Για την πληροφόρησή σας υπάρχει η Ένωση για την Υπολογιστική Γλωσσολογία (*Association for Computational Linguistics (ACL)*) η οποία οργανώνει δύο ετήσια συνέδρια, ένα Ευρωπαϊκό

(EACL) και ένα παγκόσμιο (ACL) καθώς και ένα διετές (COLING). Φυσικά, ό,τι πληροφορίες θέλετε για την Ένωση υπάρχουν στο Διαδίκτυο.

Δήμητρα Θεοφανούλου-Κοντού. 1989. *Μετασχηματιστική Σύνταξη*. Από την Θεωρία στην Πράξη. Αθήνα. Καρδαμίτσας.

Ειρήνη Φιλιππάκη-Warburton. 1992. *Εισαγωγή στην Θεωρητική Γλωσσολογία*. Αθήνα. Νεφέλη

Robert D. Borsley. 1996. *Modern Phrase Structure Grammar*. Oxford: Blackwell

Ivan Bratko. 1986. *Prolog Programming for Artificial Intelligence*. Wokingham: Addison-Wesley.

Michael A. Covington. 1994. *Natural Language Processing for Prolog Programmers*. Prentice Hall: New Jersey, USA

James Allen. 1995. *Natural Language Understanding*, Benjamin / Cummings Publishing Co Inc

Gennaro Chierchia and Sally Mc-Connell-Ginet. 1990. *Meaning and Grammar: an introduction to semantics*. Cambridge, Massachusetts: MIT Press

Gerald Gazdar and Chris Mellish. 1989. *Natural Language Processing in PROLOG*. Wokingham: Addison-Wesley

Andrew Spencer. 1991. *Morphological Theory: an Introduction to Word Structure in Generative Grammar*. Oxford: Blackwell