

ΓΛΩΣΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ

ΣΗΜΑΣΙΟΛΟΓΙΑ - SEMANTICS

Σημασιολογική Επεξεργασία

- Η σημασιολογική επεξεργασία ακολουθεί συνήθως τη συντακτική
- Βασίζεται σε
 - Αναπαράσταση του νοήματος της κάθε λέξης
 - Αναπαράσταση της δομής της πρότασης (συντακτικό δέντρο)
- Παράγει
 - Αναπαράσταση του νοήματος της πρότασης
- Αγνοεί τα συμφραζόμενα και τη γνώση του κόσμου
 - «Αυτός αγαπάει τη Μαίρη» → Ποιος είναι «αυτός»;

Αναπαράσταση του νοήματος μιας πρότασης

- Έχουν προταθεί πολλοί φορμαλισμοί για την αναπαράσταση του νοήματος προτάσεων
- Ο πιο κοινός φορμαλισμός είναι η κατηγορηματική λογική
 - Δεν καλύπτει όλες τις πτυχές της φυσικής γλώσσας
- Για το λόγο αυτό χρησιμοποιούνται
 - Πιο σύνθετες μορφές λογικής
 - Λιγότερο τυπικοί συμβολισμοί (case frames)

Συνθετική Σημασιολογία (Compositional Semantics)

- Συνθετικότητα: Το νόημα ενός συστατικού βρίσκεται αποκλειστικά από το νόημα των υποσυστατικών του
 - π.χ. Το νόημα της πρότασης βρίσκεται συνδυάζοντας το νόημα της ονοματικής φράσης και της ρηματικής φράσης
 - Ο κανόνας συνδυασμού των νοημάτων πρέπει να είναι ανεξάρτητος από το πώς παράγονται τα νοήματα της κάθε φράσης
- Η ανάπτυξη ενός καλού συνθετικού μοντέλου είναι πολύ δύσκολη
- Δεν ισχύει πάντα
 - «Βρέχει καρέκλες»

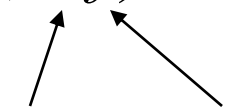
Σύνταξη και Σημασιολογία

- Υπάρχουν δύο τρόποι για να συνδυάσουμε τη συντακτική ανάλυση με τη σημασιολογική ανάλυση
 1. Κάνουμε και τις δύο αναλύσεις ταυτόχρονα εμπλουτίζοντας τη γραμματική με σημασιολογικά στοιχεία
 2. Πραγματοποιούμε πρώτα τη συντακτική ανάλυση και με βάση το δέντρο ανάλυσης δημιουργούμε τη σημασιολογική ερμηνεία

1^η προσέγγιση: Ανάλυση Lambda

- Αναπαράσταση νοήματος σε κατηγορηματική λογική
- $(\lambda x.f)$ είναι μία συνάρτηση με ένα όρισμα

όρισμα συνάρτηση



- $(\lambda x.f)(y)$ είναι ισοδύναμο με το f με αντικατάσταση κάθε εμφάνισης του x με y
 - $(\lambda x.x+1)(2) = 2+1$
 - $(\lambda x.f(x,y))(3) = f(3,y)$
 - $(\lambda x.\lambda y.f(x,y))(2) = (\lambda y.f(2,y))$
 - $(\lambda x.\lambda y.f(x,y))(2)(3) = (\lambda y.f(3,y))(2) = f(3,2)$
- Αναγωγή λάμδα (lambda reduction)

Ανάλυση Lambda: Λεξικό

- Κάθε λέξη στο λεξικό έχει μία σημασιολογική καταχώρηση
 - “jumps”: $(\lambda x. jumps(x))$
 - “loves”: $(\lambda x. \lambda y. loves(y, x))$
 - “Mary”: *mary*
 - “John”: *john*

Ανάλυση Lambda: Γραμματική

- Κάθε κανόνας της γραμματικής έχει σημασιολογικές παραμέτρους
- Η συνάρτηση `reduce` υλοποιεί την αναγωγή λάμδα

```
s(Sem) --> np(NPSem), vp(VPSem), reduce(VPSem, NPSem, Sem).
```

```
vp(Sem) --> v(VSem), np(NSem), reduce(VSem, NSem, Sem).
```

```
np(Sem) --> name(Sem).
```

```
vp(Sem) --> v(Sem).
```


Ανάλυση Lambda: Παράδειγμα

S -> NP VP

NP -> John

NP -> Mary

VP -> V NP

VP -> V

V -> loves

V -> jumps

“jumps”: $(\lambda x. jumps(x))$

“loves”: $(\lambda x. \lambda y. loves(y, x))$

“Mary”: *mary*

“John”: *john*

□ “John jumps.”

■ $(\lambda x. jumps(x))(john) = jumps(john)$

□ “John loves Mary.”

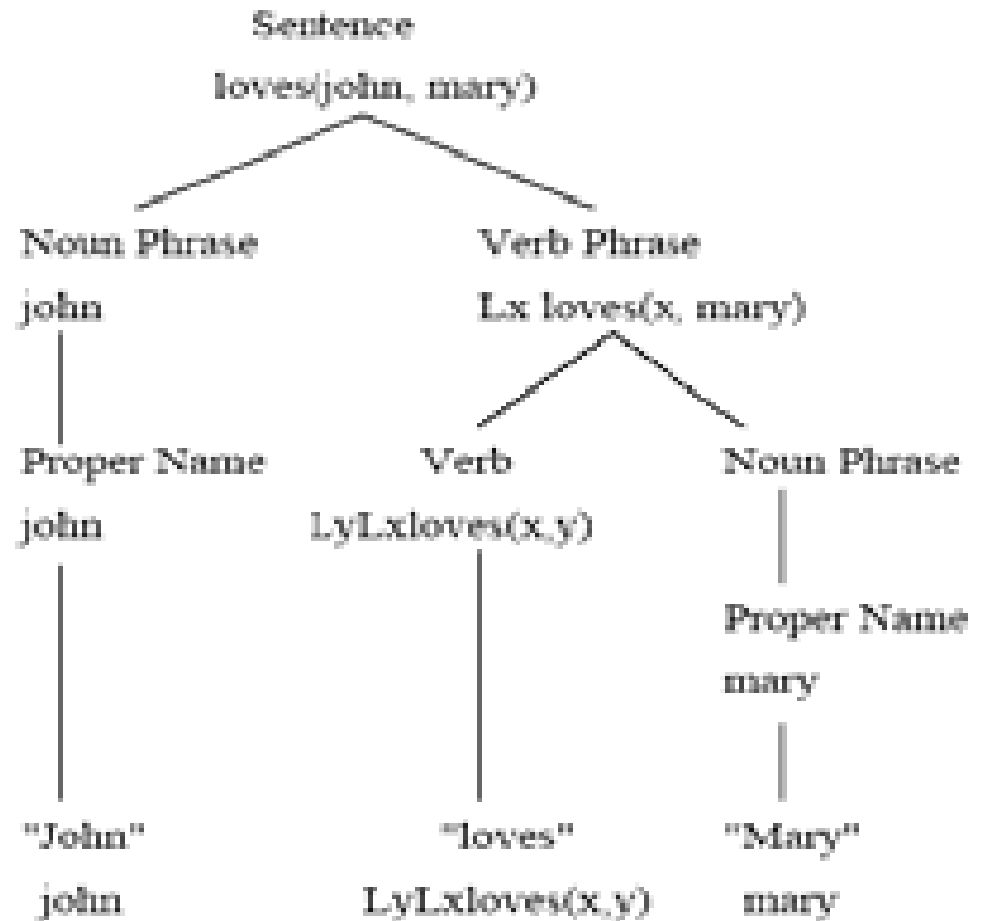
$(\lambda x. \lambda y. loves(y, x))(john)(mary) =$

$(\lambda y. loves(y, mary))(john) =$

$loves(john, mary)$

Δέντρο Ανάλυσης με Σημασιολογική Ερμηνεία

“John loves Mary”



(L = Lambda)

2^η προσέγγιση: Ανεξαρτησία Σύνταξης-Σημασιολογίας



Ασάφεια (Ambiguity)

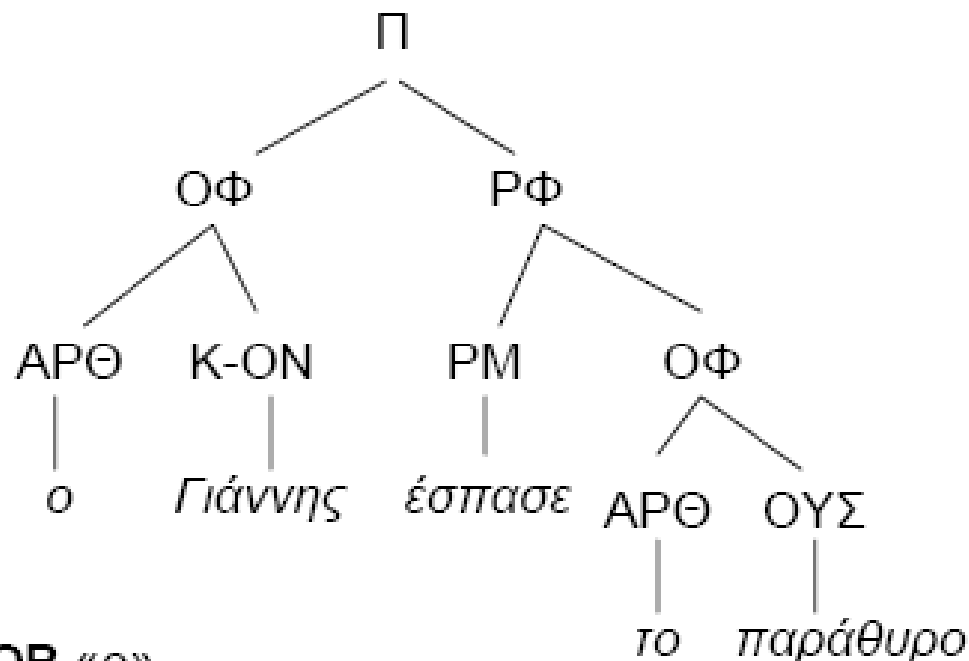
- Η εύρεση του νοήματος μιας πρότασης περιπλέκεται εξαιτίας της ασάφειας της γλώσσας.
 - Συντακτική ασάφεια: Μία πρόταση μπορεί να έχει πολλές διαφορετικές συντακτικές δομές
 - Σημασιολογική ασάφεια: Μία λέξη μπορεί να έχει πολλές διαφορετικές έννοιες
 - Αναφορική ασάφεια: Μία αναφορά μπορεί να μην είναι σαφής
- Οι περισσότερες μορφές ασάφειας επιλύονται από τα επόμενα στάδια ανάλυσης

Επίλυση Ασάφειας

- Η **λογική φόρμα** (logical form) παρέχει αυτή τη δυνατότητα
- Είναι λιγότερο ακριβής και τυπική σε σχέση με τη κατηγορηματική λογική
- Είναι περισσότερο ευέλικτη και σε μεγαλύτερη αντιστοιχία με τη συντακτική δομή της πρότασης

Έξοδος Συντακτικής Ανάλυσης

«Ο Γιάννης έσπασε
το παράθυρο.»



(Π ΥΠΟΚ (ΟΦ Κ-ΡΜ «σπάζω» ANTIK (ΟΦ ΑΡΘ-ΟΡ «ο» Κ-ΟΝ «Γιάννης» ΑΡΘ-ΟΡ «το» ΟΥΣ «παράθυρο»))

Ιεραρχία Τύπων

- Κάθε λέξη έχει τουλάχιστον ένα σημασιολογικό τύπο που υποδεικνύει τις ιδιότητες της λέξης
- Σε μία ιεραρχία τύπων οι θυγατρικοί τύποι κληρονομούν τις ιδιότητες των μητρικών τύπων.



Ποντίκι: ΤΡΩΚΤΙΚΑ ή ΗΛ.ΣΥΣΚΕΥΕΣ

Ιδιότητες Τύπων

- Κάθε τύπος προσδιορίζει συγκεκριμένες σημασιολογικές ιδιότητες
 - ΦΥΣ.ΑΝΤΙΚ: πιάνονται, βλέπονται
 - ΥΛΙΚΑ: δημιουργούνται, καταστρέφονται
 - ΖΩΝΤΑΝΑ: ζουν, πεθαίνουν
 - ΦΥΤΑ: τρώγονται, ανθίζουν
 - ΑΝΘΡΩΠΟΙ: διαβάζουν, τρώνε, πίνουν
 - ΤΡΩΚΤΙΚΑ: τρώνε, τρώγονται?

Γραμματικές Πτώσεων (Case Grammars)

- Περιλαμβάνουν ένα σύνολο σημασιολογικών ρόλων (cases, semantic roles)
- Οι σημασιολογικοί ρόλοι μπορούν να αντιστοιχιστούν σε διαφορετικές συντακτικές δομές
 - *Ο Γιάννης έσπασε το παράθυρο με το σφυρί.*
 - *Το σφυρί έσπασε το παράθυρο.*
 - *Το παράθυρο έσπασε.*
- Υποκινητής της ενέργειας: *ο Γιάννης*
- Όργανο που χρησιμοποιήθηκε: *το σφυρί*
- Αντικείμενο που δέχτηκε την ενέργεια: *το παράθυρο*

Οι Σημαντικότεροι Σημασιολογικοί Ρόλοι

- AGENT: ο υποκινητής μιας ενέργειας
- THEME: το αντικείμενο που δέχεται μια ενέργεια
- INSTRUMENT: το αντικείμενο που χρησιμοποιείται σε μια ενέργεια
- EXPERIENCER: το άτομο που αντιλαμβάνεται κάτι ή βρίσκεται σε μία ψυχολογική κατάσταση
- BENEFICIARY: το άτομο για το οποίο έγινε μία ενέργεια
- TO-LOC: τελικός τόπος
- TO-POSS: τελικός κάτοχος
- FROM-LOC: αρχικός τόπος
- FROM-POSS: αρχικός κάτοχος

Παραδείγματα

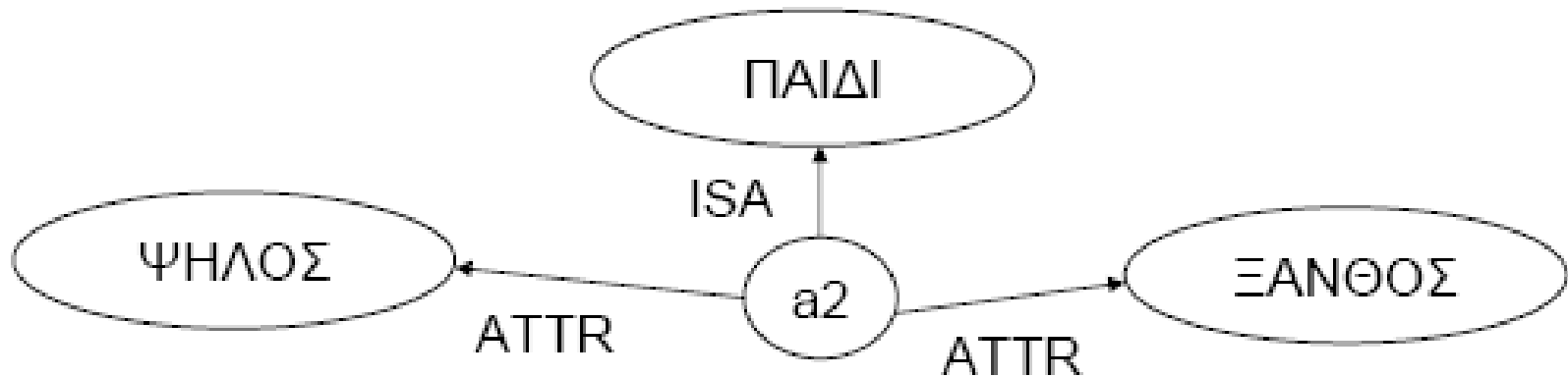
- Ο Γιάννης έσπασε το παράθυρο με το σφυρί.
AGENT THEME INSTRUMENT
- Ο αέρας στέγνωσε τα ρούχα.
INSTRUMENT THEME
- Ο Γιάννης κοιμήθηκε.
EXPERIENCER
- Έδωσα στο Γιάννη το βιβλίο.
TO-POSS THEME (AGENT: εγώ)
- Έδωσα το βιβλίο στο Γιάννη για τη Μαρία.
THEME TO-POSS BENEFICIARY
- Περπάτησα από εδώ ως το σχολείο.
FROM-LOC TO-LOC

Λογική Φόρμα

- Μία λογική φόρμα περιγράφει τη σημασιολογική ερμηνεία ενός αντικειμένου ή μιας ενέργειας
 - Τελεστής: υποδεικνύει το είδος της δομής που περιγράφεται
 - Ετικέτα: ένα μοναδικό σύμβολο για τη δομή που περιγράφεται
 - Τύπος: ο σημασιολογικός τύπος της δομής που περιγράφεται
 - Προσδιορισμοί: μία λίστα από δομές λογικής φόρμας

Σημασιολογικό Δίκτυο: Αντικείμενα

π.χ. Το ψηλό ξανθό παιδί.



Λογική Φόρμα Αντικειμένων

□ *Το παιδί*

- (ANT b1 ΠΑΙΔΙ)

□ *Το μεγάλο παιδί*

- (ANT b1 ΠΑΙΔΙ (ATTR b1 ΜΕΓΑΛΟΣ))

τελεστής ετικέτα τύπος προσδιορισμός

□ *Ένα ψηλό ξανθό αγόρι*

- (ANT b2 ΑΓΟΡΙ (ATTR b2 ΨΗΛΟΣ)
 (ATTR b2 ΞΑΝΘΟΣ))

□ *Το ποντίκι*

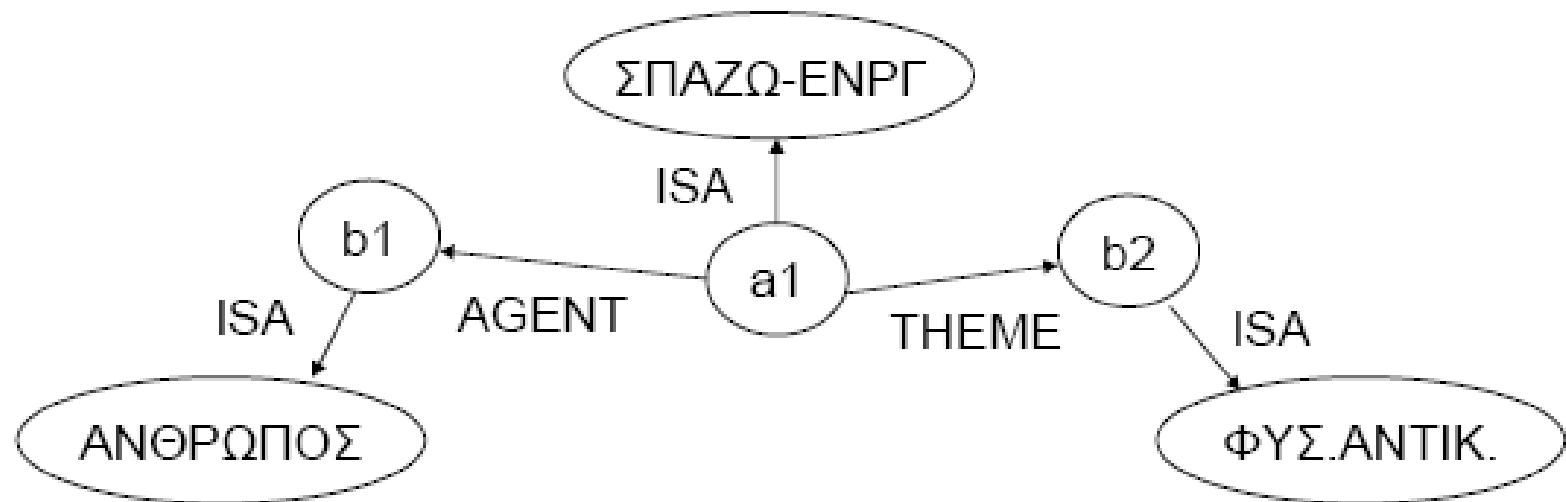
- (ANT b3 {ΤΡΩΚΤΙΚΟ, ΗΛ.ΣΥΣΚΕΥΗ})

□ *Ο Γιάννης*

- (ONOMA b4 ΑΝΘΡΩΠΟΣ «Γιάννης»)

Σημασιολογικό Δίκτυο: Ενέργειες

- π.χ. *Ο Γιάννης έσπασε το παράθυρο.*



Λογική Φόρμα Ενεργειών

- *Ο Γιάννης έσπασε το παράθυρο με το σφυρί.*
 - (ENRG a1 ΣΠΑΖΩ
(AGENT a1 (ΟΝΟΜΑ b4 ΑΝΘΡΩΠΟΣ «Γιάννης»))
(THEME a1 (ANT b5 ΦΥΣ.ΑΝΤΙΚ.))
(INSTRUMENT a1 (ANT b6 ΣΦΥΡΙ)))
- *Το σφυρί έσπασε το παράθυρο.*
 - (ENRG a2 ΣΠΑΖΩ
(INSTRUMENT a2 (ANT b6 ΣΦΥΡΙ))
(THEME a2 (ANT b5 ΦΥΣ.ΑΝΤΙΚ.)))
- *Το παράθυρο έσπασε.*
 - (ENRG a3 ΣΠΑΖΩ (THEME a2 (ANT b5 ΦΥΣ.ΑΝΤΙΚ.)))

Σημασιολογική Ερμηνεία με Λογική Φόρμα

- Η λογική φόρμα παρέχει τη δυνατότητα κατασκευής μερικών περιγραφών (partial descriptions) για τα συστατικά της πρότασης
- Οι μερικές περιγραφές για τα συστατικά της πρότασης δημιουργούνται βάσει κανόνων
- Οι μερικές περιγραφές συγχωνεύονται για τη δημιουργία του νοήματος ολόκληρης της πρότασης
- Ό,τι δεν μπορεί να ερμηνευτεί αφήνεται για τα μετέπειτα στάδια επεξεργασίας

Παραγωγή Μερικών Περιγραφών

- * Δημιουργία μίας νέας μοναδικής ετικέτας
- **T(συστατικό)**
 - Επιστρέφει το σημασιολογικό τύπο του συστατικού
- **S(συστατικό)**
 - Επιστρέφει τη σημασιολογική ερμηνεία (ολόκληρη την λογική φόρμα) του συστατικού

Κανόνες Επιλεκτικών Περιορισμών (Selectional Restrictions)

Κανόνας νόστιμος.1:

(ΟΦ ΕΠΘ νόστιμος

ΟΥΣ +ΦΑΓΩΣΙΜΟ) → (ANT * Τ(ΟΥΣ) (ATTR * ΕΥΓΕΣΤΟΣ))

Κανόνας νόστιμος.2:

(ΟΦ ΕΠΘ νόστιμος

ΟΥΣ +ΑΝΘΡΩΠΟΣ) → (ANT * Τ(ΟΥΣ) (ATTR * ΟΜΟΡΦΟΣ))

Καινούρια ετικέτα

σημασ. τύπος

Συντακτική δομή με
επιλεκτικούς
περιορισμούς

Μερική Περιγραφή

Εφαρμογή των κανόνων

□ *Το νόστιμο μήλο*

(ANT b1 ΦΡΟΥΤΟ (ATTR b1 ΕΥΓΕΣΤΟΣ))

□ *Ο νόστιμος άντρας*

(ANT b1 ΑΝΘΡΩΠΟΣ (ATTR b1 ΟΜΟΡΦΟΣ))

Κανόνες (συνέχεια)

Κανόνας σπάζω.1

(Π Κ-PM σπάζω

ΥΠΟΚ +ΑΝΘΡΩΠΟΣ) → (ENERG * ΣΠΑΖΩ (AGENT * S(ΥΠΟΚ)))

Κανόνας σπάζω.2

(Π Κ-PM σπάζω

ΑΝΤΙΚ +ΦΥΣ.ΑΝΤΙΚ.) → (ENERG * ΣΠΑΖΩ (THEME * S(ΑΝΤΙΚ)))

Ο Γιάννης έσπασε το παράθυρο →

(ENERG a1 ΣΠΑΖΩ (AGENT a1 (ΟΝΟΜΑ b1 ΑΝΘΡΩΠΟΣ «Γιάννης»))

(THEME a1 (ANT b2 ΦΥΣ.ΑΝΤΙΚ.))

Κανόνες (συνέχεια)

Κανόνας σπάζω.3

(Π Κ-PM σπάζω

ΥΠΟΚ +ΦΥΣ.ΑΝΤΙΚ-ΑΝΘΡΩΠΟΣ

ΑΝΤΙΚ +ΦΥΣ.ΑΝΤΙΚ)

→ (ENERG * ΣΠΑΖΩ (INSTRUMENT * S(ΥΠΟΚ)))
(THEME * S(ΑΝΤΙΚ))

Κανόνας σπάζω.4

(Π Κ-PM σπάζω

ΠΦ (ΠΡΘ με

ΟΥΣ +ΦΥΣ.ΑΝΤΙΚ-ΑΝΘΡΩΠΟΣ))

→ (ENERG * ΣΠΑΖΩ (INSTRUMENT * S(ΠΦ)))

Το σφυρί έσπασε το παράθυρο →

(ENERG a1 ΣΠΑΖΩ (INSTRUMENT a1 (ANT b3 ΕΡΓΑΛΕΙΟ))

(THEME a1 (ANT b4 ΦΥΣ.ΑΝΤΙΚ.)))

Το παράθυρο έσπασε με το σφυρί →

(ENERG a1 ΣΠΑΖΩ (INSTRUMENT a1 (ANT b5 ΕΡΓΑΛΕΙΟ))

Ενοποίηση Περιγραφών

- Οι περιγραφές που προκύπτουν από την εφαρμογή των κανόνων συγχωνεύονται εφόσον
 - Ο τελεστής, η ετικέτα και ο τύπος είναι ταυτόσημα ή δεν έχουν ακόμα πάρει τιμή
 - Οι προσδιορισμοί είναι συμβατοί

- Ασύμβατοι προσδιορισμοί (εσωτερικές πτώσεις):
(ENPT * ΤΡΕΧΩ (THEME * (ONOMA b1 ΑΝΘΡΩΠΟΣ «Γιάννης»)))
(ENPT * ΤΡΕΧΩ (THEME * (ANT b2 ΛΟΓΙΣΜΙΚΟ)))

Επίλυση Ασάφειας

- Η εφαρμογή των κανόνων παραγωγής μερικών περιγραφών μπορεί να επιλύσει σημασιολογικές ασάφειες

- *Το μαύρο ποντίκι*

(ANT b1 {ΤΡΩΚΤΙΚΟ, ΗΛ.ΣΥΣΚΕΥΗ} (ATTR B1 ΜΑΥΡΟΣ))

- *Το ασύρματο ποντίκι*

(ANT b1 ΗΛ.ΣΥΣΚΕΥΗ (ATTR b1 ΑΣΥΡΜΑΤΟΣ))

Ιδιωματισμοί

- Για την αναπαράσταση του νοήματος των ιδιωματισμών πρέπει να ορίσουμε σαφείς λεξιλογικούς περιορισμούς στους κανόνες

(Π ΥΠΟΚ +ΑΝΘΡΩΠΟΣ

K-PM *κάνω*

ΑΝΤΙΚ (ΟΦ ΑΡΘ *την*
ΟΥΣ *πάπια*))

→ (ENRG * ΦΟΒΑΜΑΙ (EXPERIENCER * S(ΥΠΟΚ)))

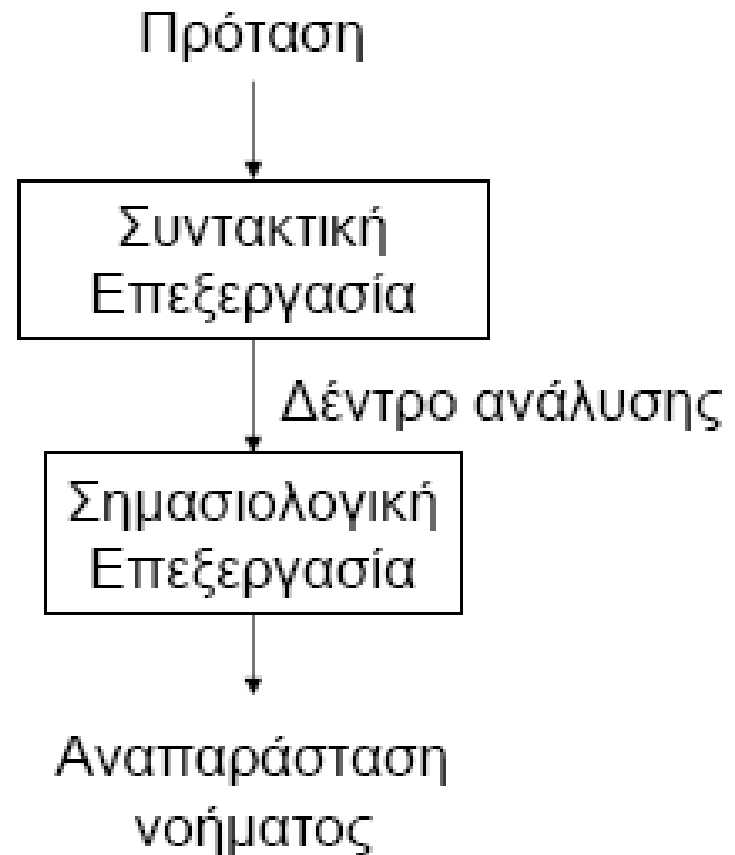
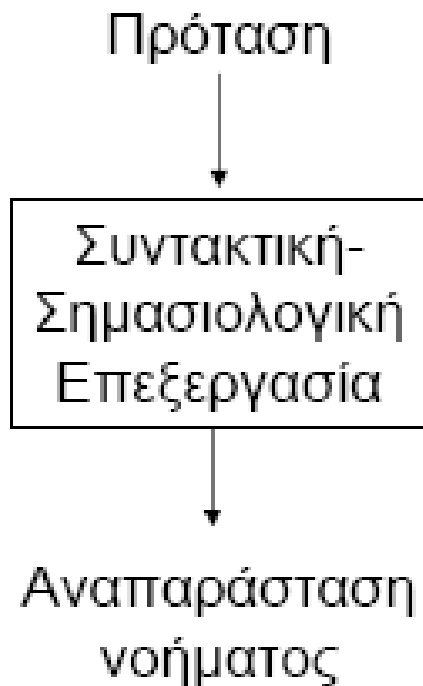
(Π ΥΠΟΚ +ΑΝΘΡΩΠΟΣ

K-PM *κάνω*

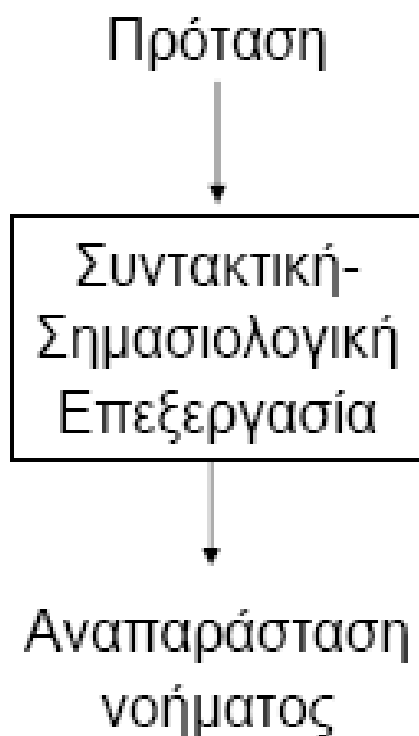
ΑΝΤΙΚ +ΔΙΑΔΙΚΑΣΙΑ

→ (ENRG * ΔΗΜΙΟΥΡΓΩ (AGENT * S(ΥΠΟΚ))
(THEME * S(ΑΝΤΙΚ)))

Σύγκριση Προσεγγίσεων

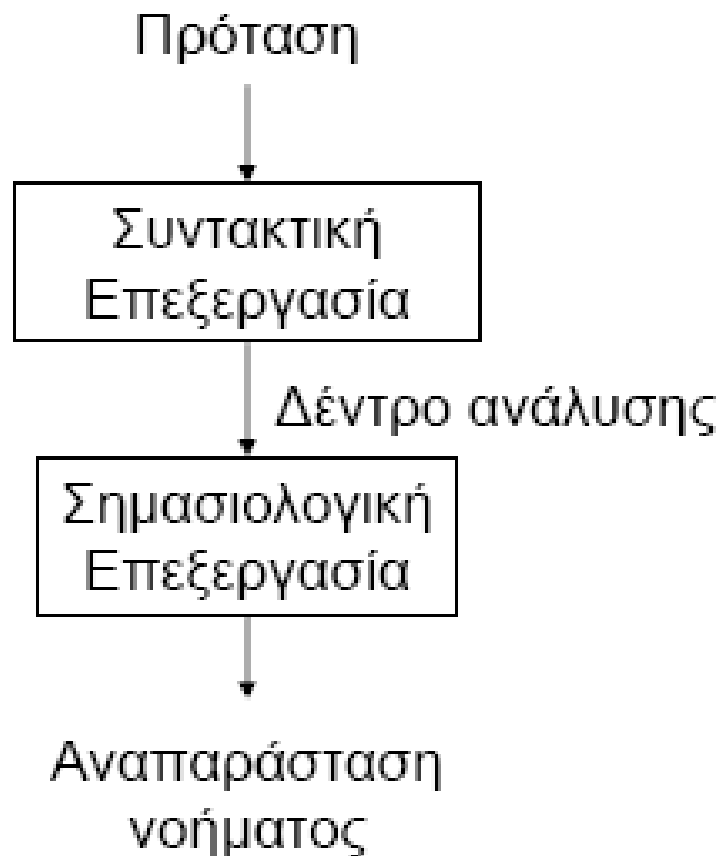


1^η προσέγγιση



- ☑ Η συντακτική ασάφεια επιλύεται πιο εύκολα και νωρίτερα
- ☒ Η γραμματική που προκύπτει είναι πιο περίπλοκη

2^η προσέγγιση



- ✓ Πιο αρθρωτή προσέγγιση
- ✓ Το κάθε τμήμα μπορεί να βελτιωθεί ανεξάρτητα
- ✗ Η συντακτική ασάφεια οδηγεί σε πολλά πιθανά δέντρα ανάλυσης