Συναναφορά

Μαρίνα Βασιλείου

Ινστιτούτο Επεξεργασίας του Λόγου/Ε.Κ. Αθηνά

ΜΠΣ Τεχνογλωσσία: Επεξεργασία Σωμάτων Κειμένων

Περιεχόμενα

1. Ορισμός σχέσεων
2. Ανασκόπηση στο χώρο
3. Σχήμα επισημείωσης για ελληνικά δεδομένα
4. Εφαρμογές
5. Μαθαίνοντας το TrEd & το GDT_Bridging
6. Επισημείωση στην τάξη

❶ Ορισμός σχέσεων

Οι λέξεις αναφέρονται σε οντότητες

Ο **ήλιος** μισοσκεπάστηκε από τα **σύννεφα**.

Οι **γείτονες** ποτέ δε ρίχνουν τα **σκουπίδια τους** μέσα στον **κάδο**.

Ο **Αυγερινός** εμφανίζεται στον ουρανό τις πρώτες πρωινές ώρες, ενώ ο **Αποσπερίτης** μετά τη δύση του ήλιου.

Πού βρίσκονται οι οντότητες;

- Στον πραγματικό κόσμο (real world)

  ↓

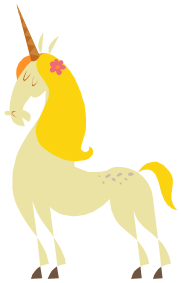
  - Ήλιος
  - Σύννεφα
  - Σκουπίδια
  - Κάδος
  - Αφροδίτη

ή

- Στο καταστασιακό περιβάλλον (discourse)

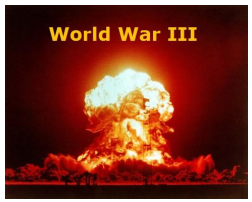
  ↓

  - Μονόκερως

  - Άγιος Βασίλης

  - Γ' Παγκόσμιος Πόλεμος

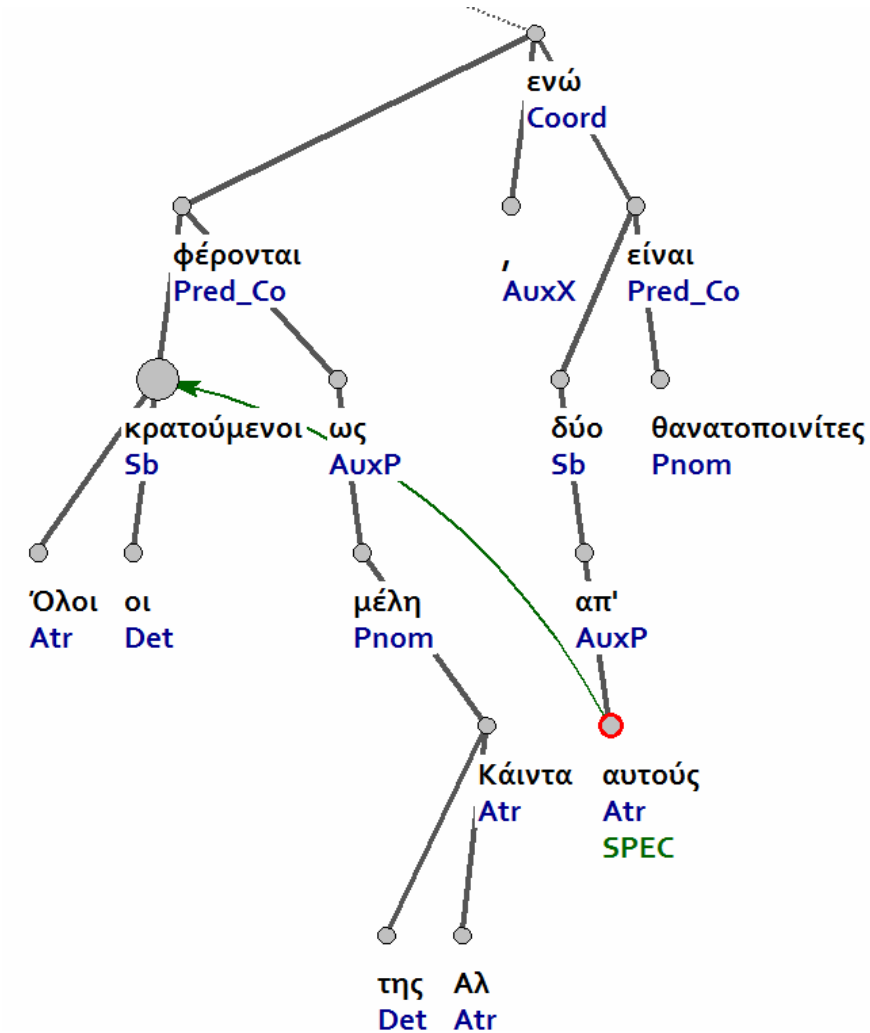
    

❶ Συναναφορά (Coreference)

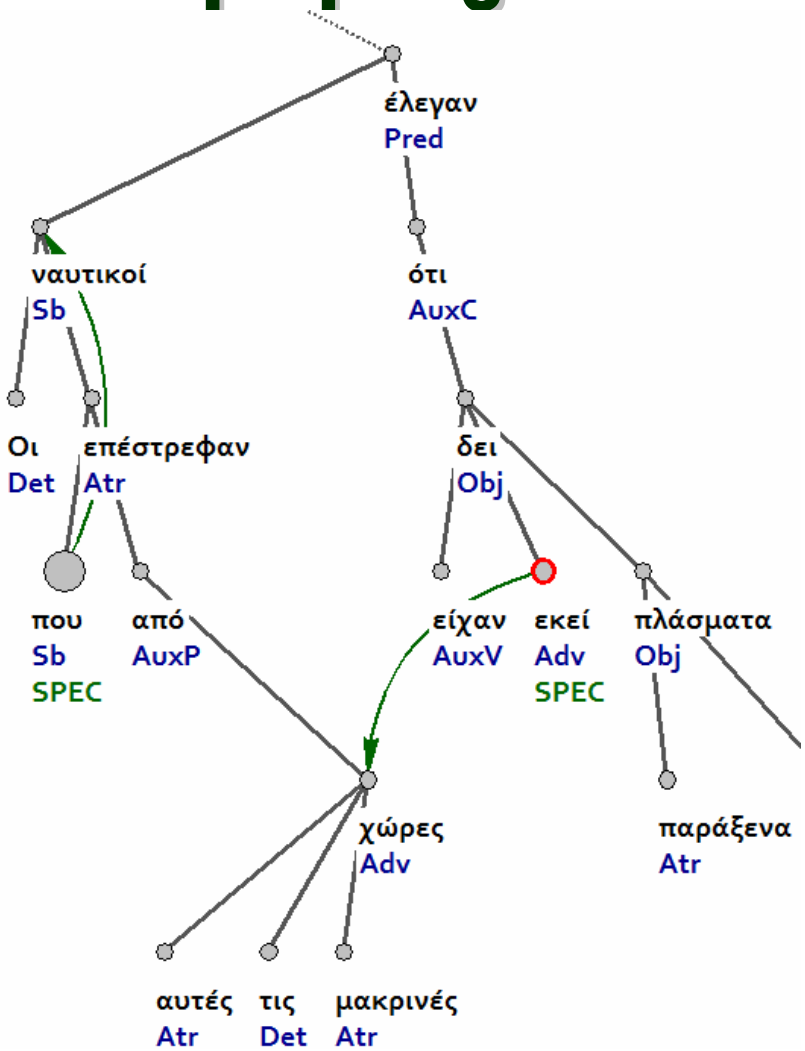
Σχέση μεταξύ (τουλάχιστον) 2 εκφράσεων που αναφέρονται στην **ίδια οντότητα**

Το ΔΝΤ δημοσιοποίησε την έκθεση αξιολόγησης για την Ελλάδα. Μία ώρα νωρίτερα, ο επικεφαλής της αποστολής του Ταμείου στη χώρα, Πολ Τόμσεν, είχε μιλήσει στους δημοσιογράφους για την έκθεση αυτή, η οποία αναμενόταν εδώ και καιρό.

Παραδείγματα συναναφοράς



Όλοι οι **κρατούμενοι** φέρονται ως μέλη της Αλ Κάιντα, ενώ δύο απ' **αυτούς** είναι θανατοποινίτες.



Οι **ναυτικοί που** επέστρεφαν από αυτές τις μακρινές **χώρες** έλεγαν ότι είχαν δει **εκεί** πλάσματα παράξενα ...

Αναφορά ή Συναναφορά (εν κειμένω);

Αναφορά

Σχέση μεταξύ αναφορικού στοιχείου & σημείου αναφοράς (antecedent):

Αλυσίδα με **μόνο δύο** στοιχεία

Συναναφορά

Σχέση μεταξύ εκφράσεων που αναφέρονται στην ίδια οντότητα:

Αλυσίδα με **τουλάχιστον δύο** στοιχεία

Μέσα στις επόμενες εβδομάδες θα συνεδριάσει το διοικητικό συμβούλιο του **ΔΝΤ$_i$** στο πλαίσιο της 5ης αξιολόγησης του ελληνικού προγράμματος, δήλωσε σήμερα ο εκπρόσωπός **του$_i$** Τζέρι Ράις απαντώντας σε σχετική ερώτηση, κατά την καθιερωμένη συνέντευξη Τύπου, στα γραφεία του Ταμείου στην Ουάσιγκτον. Η άποψη του **Ταμείου$_k$** για την αξιολόγηση του χρέους παραμένει η ίδια, καθώς **pro$_k$** αναμένει ότι το χρέος θα παραμείνει σε «υψηλά επίπεδα».

Δείτε τη διαφορά

Μέσα στις επόμενες εβδομάδες θα συνεδριάσει το διοικητικό συμβούλιο του **ΔΝΤ$_i$** στο πλαίσιο της 5ης αξιολόγησης του ελληνικού προγράμματος, δήλωσε σήμερα ο εκπρόσωπός **του$_i$** Τζέρι Ράις απαντώντας σε σχετική ερώτηση, κατά την καθιερωμένη συνέντευξη Τύπου στα γραφεία του **Ταμείου$_i$** στην Ουάσιγκτον. Η άποψη του **Ταμείου$_i$** για την αξιολόγηση του χρέους παραμένει η ίδια, καθώς **pro$_i$** αναμένει ότι το χρέος θα παραμείνει σε «υψηλά επίπεδα».

Άλλες 2 σχέσεις μεταξύ οντοτήτων

❷ **Οιονεί συναναφορά (Near Identity)**

- Σχέση μεταξύ 2 εκφράσεων που διαφέρουν ως προς ένα **χαρακτηριστικό** ή **πτυχή** μίας οντότητας

❸ **Γεφύρωση (Bridging)**

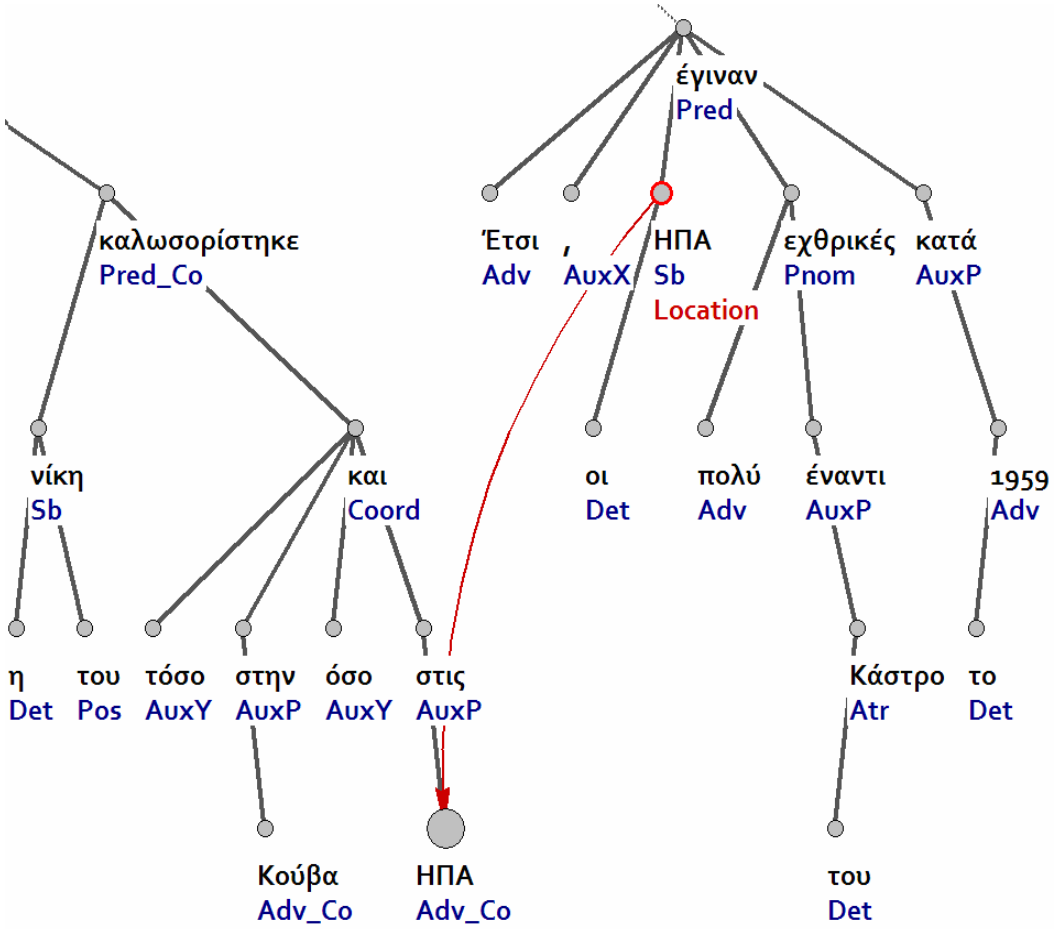
- Σχέση μεταξύ 2 **μη συναναφερόμενων** αλλά συσχετιζόμενων εκφράσεων

Παράδειγμα οιονεί συναναφοράς

Η νίκη του καλωσορίστηκε τόσο στην Κούβα όσο και στις **ΗΠΑ**. ...

Έτσι οι **ΗΠΑ** έγιναν πολύ εχθρικές έναντι του Κάστρο κατά το 1959.

Οι Κουβανοί εξόριστοι έγιναν γρήγορα μια ισχυρή ομάδα λόμπι στις **ΗΠΑ**, ισχύ την οποία διατηρούν μέχρι σήμερα.



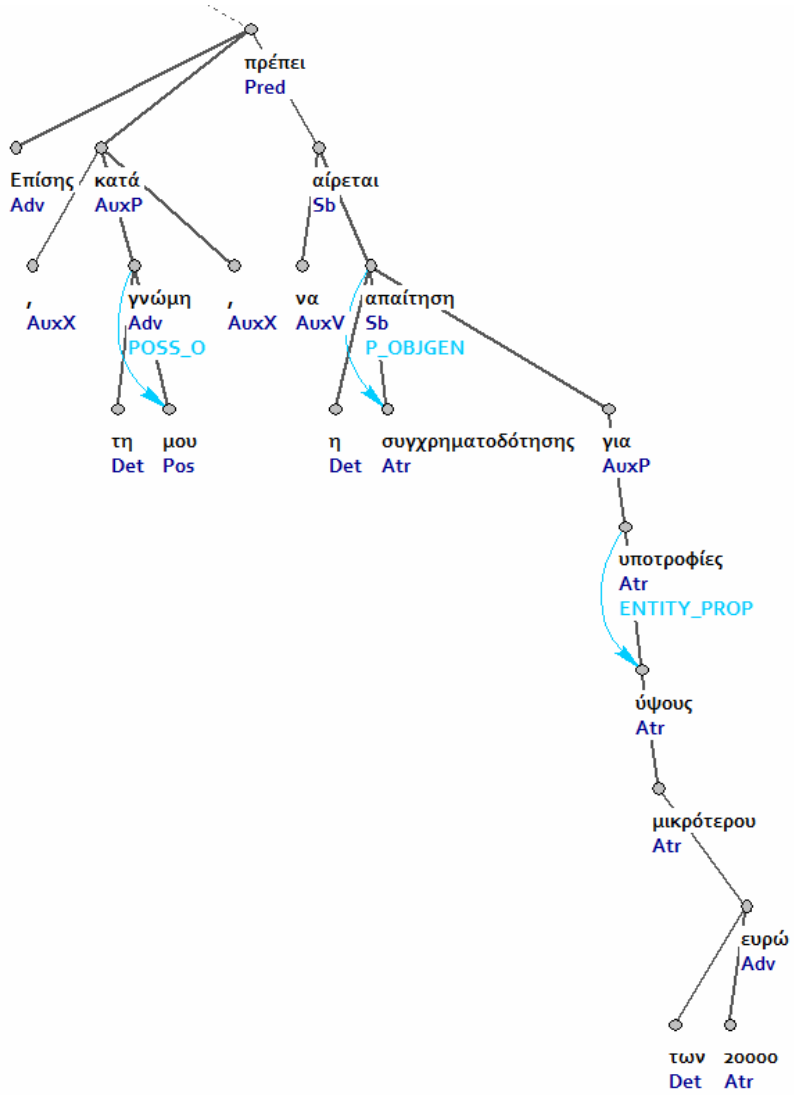
Παραδείγματα γεφύρωσης

Σχέσεις

(α) κτήσης

(β) έκφρασης κατηγορήματος

(γ) ιδιότητας



Επίσης, κατά τη **γνώμη μου**, πρέπει να αίρεται η **απαίτηση συγχρηματοδότησης** για **υποτροφίες ύψους** μικρότερου των 20.000 ευρώ.

- Συναναφορά
- Οιονεί συναναφορά
- Γεφύρωση

} Μηχανισμοί συνοχής

❷ Ανασκόπηση στο χώρο

2 ζητήματα για τη συναναφορά

- **Επισημείωση** σχέσεων συναναφοράς
- **Επίλυση** συναναφοράς

Επισημείωση: Ζητήματα

- Ποια στοιχεία επισημειώνει κανείς (π.χ. μόνο ονοματικά);
- Αποκλείονται κάποια στοιχεία (π.χ. αόριστες αντωνυμίες);
- Ποιες σχέσεις;
- Σε ποιο επίπεδο αναπαράστασης (π.χ. φραστικής δομής ή συντακτικών εξαρτήσεων);
- Τι γίνεται με τα
  - κενά υποκείμενα;
  - διακριτά σημεία αναφοράς (split antecedents);
  - κατηγορούμενα;
  - δομές παράθεσης

Επισημείωση: Σχήματα

Relation & features	UCREL	MUC-7	MATE	ACE	AnCoraCo	PDT	OntoNotes 5.0	GCB
Coreference	✓	✓	✓	✓	✓	✓	✓	✓
Bridging			✓			✓		✓
Split antecedents	✓		✓		✓	✓		✓
Phrases		✓	✓	✓	✓		✓	
Words		✓	✓	✓	✓	✓	✓	✓
Metonymy				✓	✓			✓
Near identity					✓			✓
Null subjects			✓		✓	✓	✓	✓
Discourse deixis					✓	✓	✓	
Generic / specific reference				✓		✓		✓

- UCREL (Fligelstone 1992)
- MUC-7 (Hirschman & Chinchor 1997)
- MATE (Poesio et al. 1999)
- ACE (Doddington et al. 2004)
- AnCoraCo (Recasens & Martí 2010)
- PDT 2.0 (Nědolužko & Mírovský 2011)
- OntoNotes 5.0 (Weischedel et al. 2012)

Επισημείωση: Σώματα κειμένων

- **MUC-6** (1995) & **MUC-7** (1998)
- **ACE-2** (47.984 λέξεις); **ACE2005** (303.833 λέξεις)
- **SemEval 2010** (100.000 λέξεις)
- **MEDLINE** (~55.000 λέξεις)

Επισημείωση: Εργαλεία

ενδεικτικά

- AnCoraPipe (Bertran et al. 2008)

  http://annotation.exmaralda.org/index.php/AnCoraPipe

- TrEd (Pajas & Stepanek 2008)

  http://ufal.mff.cuni.cz/tred/

- BRAT (Nguyen et al. 2011)

  http://brat.nlplab.org

Επίλυση συναναφοράς: Χαρακτηριστικά

- String matching (exact; substring; head; edit distance; head & modifiers; aliasing)
- Γραμματικά χαρακτηριστικά (μέρος του λόγου, μορφοσυντακτικά χαρακτηριστικά)
- Συντακτικές σχέσεις (συντακτικοί ρόλοι, συντακτικές εξαρτήσεις, φραστική δομή, αρχές αναφορικής δέσμευσης)
- Σημασιολογικά χαρακτηριστικά (κατηγόρημα & ορίσματα, σημασιολογικές τάξεις, σημασιολογικές σχέσεις, σχέσεις βάσει του Wordnet)
- Γνώση του κόσμου (πληροφορίες αντλημένες από διαδικτυακές πηγές όπως η wikipedia ή DBPedia)
- Άλλα χαρακτηριστικά (αναφορικότητα, απόσταση)

Επίλυση συναναφοράς: Μέθοδοι

- Κανονιστικές μέθοδοι (rule-based)
- Μέθοδοι μηχανικής μάθησης (machine learning or corpus-based)

Κανονιστικές μέθοδοι

- **RAP** [Resolution of anaphora procedure] (Lappin & Leass (1994))
  - Συμφωνία, συντακτικοί περιορισμοί & σταθμισμένη **salience** (π.χ. συντακτικοί ρόλοι, συχνότητα, εγγύτητα κ.ά.)

- **Multi-pass sieve system** (Raghunathan et al. (2010); Lee et al. (2011))
  - Μία σειρά **φίλτρων** που σταδιακά (υψηλότερη → χαμηλότερη ακρίβεια) ομαδοποιούν τις συναναφερόμενες οντότητες

- **ARKref** (Haghighi & Klein (2009); O' Connor & Heilman (2013))
  - **Συντακτικές** σχέσεις & σχέσεις αναφορικής δέσμευσης, βασικές **σημασιολογικές** έννοιες (π.χ. πρόσωπο, τοποθεσία, ομάδα, αντικείμενο, ουσία κ.ά.)

Μέθοδοι μηχανικής μάθησης

- Χρησιμοποιούν επισημειωμένα σώματα κειμένων για να εκπαιδεύσουν ταξινομητές (classifiers)

- Διατρέχουν τα κείμενα εισόδου με στόχο να δημιουργήσουν συστάδες (clusters) συναναφερόμενων στοιχείων

3 τύποι μεθόδων μηχανικής μάθησης

1. **Mention-Pair models** [1 υποψήφια έκφραση]
   - Εξέταση εκφράσεων **ανά 2** ως προς τη συναναφορά
   - Δημιουργία θετικών και αρνητικών παραδειγμάτων για την εκπαίδευση του ταξινομητή (Soon et al. (2001); Ng & Cardie (2002c))

2. **Entity-Mention models**
   - Εξέταση μίας έκφρασης έναντι μίας ήδη σχηματισμένης **συστάδας** με βάση κάποια χαρακτηριστικά της συστάδας (π.χ. γένος, αριθμός, σημασιολογική τάξη κ.ά.) (Yang et al. (2004b))

3. **Ranking models** [2 υποψήφιες εκφράσεις]
   - Ταυτόχρονη εξέταση όλων των υποψήφιων σημείων αναφοράς & κατάταξη αυτών βάσει πιθανότητας (probability)
   - Εκπαίδευση ταξινομητή με παραδείγματα **ζευγών** (NP + 2 υποψήφια στοιχεία) (Iida et al. (2003); Yang et al. (2008))

Επίλυση συναναφοράς: Εργαλεία

ενδεικτικά

- BART (Versley et al. 2008a & 2008b)

  http://www.bart-coref.org/index.html

- Stanford Deterministic Coreference Resolution System (Raghunathan et al. 2010; Lee et al. 2011)

  http://www-nlp.stanford.edu/software/dcoref.shtml

- ARKref (Haghighi & Klein 2009; O'Connor & Heilman 2013)

  http://www.ark.cs.cmu.edu/ARKref/

Μετρικές αξιολόγησης

Αξιολογούν τις αλυσίδες συναναφοράς που εντοπίζει ένα σύστημα σε ένα κείμενο έναντι των "σωστών" (gold) αλυσίδων

- **MUC** (Vilain et al. 1995)
- **B3** (Bagga & Baldwin 1998)
- **CEAF** (Luo 2005)
- **LMetrics** (Chen & Ng 2013)

❸ Σχήμα επισημείωσης για ελληνικά δεδομένα

Σχήμα επισημείωσης για τα Ελληνικά

- Βασισμένο εν πολλοίς στο σχήμα επισημείωσης PDT 2.0, με αρκετές τροποποιήσεις
- Είσοδος → Δέντρα συντακτικών εξαρτήσεων
- Αναπτύχθηκε & έχει εφαρμοσθεί χειρωνακτικά σε ένα μικρό υποσύνολο του GDT (109 κείμενα, 2.075 προτάσεις, 49.329 tokens)
- Σχέσεις
  - Συναναφοράς
  - Οιονεί συναναφοράς
  - Γεφύρωσης

συνεχίζεται →

- Ονοματικά στοιχεία (ουσιαστικά & αντωνυμίες)
- Αναφορικά επιρρήματα
- Κενά υποκείμενα
- Παραθετικές δομές & δομές παράταξης
- Διακριτά σημεία αναφοράς
- Μετωνυμικές σχέσεις

Περιβάλλοντα συναναφοράς

Περιβάλλον	Παράδειγμα
Ίδια λήμματα	Πρέπει να αναπτυχθεί ένα πνεύμα συμβιβασμού στο Νότιο **Καύκασο** ... πώς θα βελτιωθούν οι σχέσεις της Ένωσης με το Νότιο **Καύκασο**.
Αντωνυμικά στοιχεία, (αναφορικές αντωνυμίες, κλιτικά κ.ά.)	(α) Τρεις νεκροί από τη σφοδρή **κακοκαιρία που** πλήττει τα Βαλκάνια. (β) Οι συγκρούσεις αποτελούν απειλή για τη σταθερότητα και την ανάπτυξη της **περιφέρειας** και των γειτόνων **της**.
Δομές ελέγχου	Η **δόνηση** είχε εστιακό βάθος 5 χιλιόμετρα, ∅ προκαλώντας την κατάρρευση των περισσοτέρων μεσαιωνικών κτισμάτων στο κέντρο της πόλης.
Κενό υποκείμενο παρεμφατικών δομών	Η στρατιωτική **ηγεσία** της Αιγύπτου είχε υποσχεθεί ότι ∅ θα παραδώσει την εξουσία σε μια εκλεγμένη πολιτική κυβέρνηση.
Αναφορικά επιρρήματα	Πιστεύεται ότι είχε αποπλεύσει από τη **Ρόδο**, **όπου** έζησε ο Ίππαρχος.
Discourse coreferents	Το Κοινοβούλιο υποστήριξε την **πρωτοβουλία** αυτή κατά την πρώτη ανάγνωση, αν και παρατήρησε ότι δεν θα πρέπει να περιμένουμε από την **πρόταση** αυτή να επιλύσει όλα τα προβλήματα που παρατηρούνται σήμερα.
Μετωνυμία	**Μία κατηγορία από μόνη της!**

Μετωνυμία: Μηχανισμός αντικατάστασης

Τύπος μετωνυμίας		Παράδειγμα
Within-Sets	One-Many	Ο **Τούρκος** φίλος δεν πιάνεται. Αυτή την παροιμία την έλεγε η μάνα μου. Και να που κάθε λίγο έρχονται οι **Τούρκοι** με τις δηλώσεις τους και την δικαιώνουν.
	Set-Member	Το 1951, ο Λαϊκός Απελευθερωτικός Στρατός της **Κίνας** εισέβαλε στο Θιβέτ ... τα ανθρώπινα δικαιώματα των Θιβετιανών παραβιάζονται μέσω διακρίσεων, σε εφαρμογή μιας πολιτικής που οι **Κινέζοι** αποκαλούν "διαχωρισμός και αφομοίωση".
Abstract-Concrete		Η κοινωνία απαθής βλέπει τα **νιάτα** της να καταστρέφονται ... Η μάστιγα των ναρκωτικών απλώνεται παντού. Πολλά τα κέρδη για τους εμπόρους ... πολλά τα κέρδη από τον έλεγχο των **νέων** και όχι μόνο.
NE-Creation		Πάντα διαβάζω **Ελύτη** κι ας μην καταλαβαίνω την **ποίησή** του.
Constituent	Part-Whole	Ήταν μία προσωπικότητα που ζούσε κι ανέπνεε για το **σανίδι**. Η μεγάλη κυρία του θεάτρου έφυγε προχθές από τη ζωή, συμπληρώνοντας έτσι 65 χρόνια στη **σκηνή**.
	Material-Object	Κατέληξε να πίνει **ποτά** όλο το βράδυ, αν και είχε υποσχεθεί στον εαυτό του να μην ξαναβάλει στο στόμα του **οινόπνευμα**.

Μετωνυμία (2)

Τύπος μετωνυμίας		Παράδειγμα
Container-Content		Καθώς του αρέσουν πάρα πολύ οι **φακές**, έφαγε όλη την **κατσαρόλα**.
Instrument	Instrument-Act	Το **σίδερο** το βαριέμαι, έτσι πολλά ρούχα μένουν χωρίς **σιδέρωμα**.
	Instrument-Agent	... αποχώρησαν οι μουσικοί αφήνοντας μόνο το μαέστρο να διευθύνει μπροστά σε δυο **βιολιστές**. ... Λίγο αργότερα, ο μαέστρος κατέβασε την μπαγκέτα και αποχώρησε και αυτός από τη σκηνή, κάνοντας τα δυο **βιολιά** να σιγήσουν.
Institution	Seat-Institution	Η συνάντηση του **πρωθυπουργού** με τους αντάρτες βουλευτές έγινε νωρίς το πρωί ... το **Μαξίμου** έγινε αποδέκτης οργισμένων αντιδράσεων από μερίδα βουλευτών.
	Institution-Person	Ο κ. **Αρβανιτόπουλος** σημείωσε ότι υπάρχουν «δεσμεύσεις του παρελθόντος» ... η ΟΛΜΕ προειδοποίησε το **υπουργείο Παιδείας** πως δεν θα αποδεχθεί τις απολύσεις εκπαιδευτικών ...
Aspect		Αγαπώ το **Ρίντλει Σκοτ** και περιμένω σαν μικρό παιδί κάθε νέα του ταινία. ... Ακόμα περισσότερο όταν αυτή η αποκαθήλωση της ιλουστρασιόν εικόνας γίνεται από ένα **μαέστρο** της οπτικής «γυαλάδας».

Τύποι οιονεί συναναφοράς

Τύπος	Παράδειγμα
Location	Το **Αφγανιστάν** πρέπει να εφαρμόσει νομοθεσία κατά της βίας εναντίον γυναικών ... Η Υπηρεσία των Ηνωμένων Εθνών για την αποστολή βοήθειας στο **Αφγανιστάν** ... προέτρεψε την Τετάρτη το **Αφγανιστάν** να εφαρμόσει στην ολότητά της και άμεσα τη νομοθεσία για τον τερματισμό της βίας κατά των γυναικών.
Organisation	Ο Larry και ο Sergey αποφασίζουν να αλλάξουν το όνομα της μηχανής αναζήτησης BackRub. Μετά από πολλή σκέψη κατέληξαν στο όνομα **Google**. ... Στις 4 Σεπτεμβρίου η **Google** υποβάλλει αίτηση για ίδρυση στην Καλιφόρνια. ... Είναι ένα πρόσθετο προγράμματος περιήγησης που επιτρέπει την αναζήτηση χωρίς να χρειάζεται να μεταβαίνετε στην αρχική σελίδα του **Google**.
Instantiation	Έχω διαβάσει το **Μονόγραμμα**, αλλά δεν **το** έχω ακούσει.
Similar sets	Πηγές της Αντιπολίτευσης αναφέρουν ότι «οι **κυβερνητικές δυνάμεις** προσπαθούν να αποτρέψουν τις ομάδες του Ερυθρού Σταυρού να γίνουν μάρτυρες της σφαγής που συντελείται από τους **σύρους στρατιώτες** που καταδιώκουν και εκτελούν τους εναπομείναντες αντάρτες». Ο Γενικός Γραμματέας των Ηνωμένων Εθνών, Μπαν Κι Μουν, ανέφερε την Παρασκευή ότι «έχει λάβει μακάβριες πληροφορίες, σύμφωνα με τις οποίες οι **δυνάμεις ασφάλειας** συλλαμβάνουν και βασανίζουν ανθρώπους στη Χομς μετά την αποχώρηση των ανταρτών».
OTHER	... δεκάδες χωριά της Σερβίας έχουν αποκλειστεί από την κακοκαιρία και την έντονη **χιονόπτωση** ... Τα συνεργεία καθαρισμού κατέβαλαν μέχρι αργά χθες το βράδυ έντονες προσπάθειες για να καθαρίσουν το **χιόνι**, χωρίς όμως να καταφέρουν να αποκαταστήσουν την κυκλοφορία ...

Τύποι γεφύρωσης

Τύπος	Παράδειγμα
Contrast	**Άλλοτε** η περιοχή έσφυζε από ζωή, αφού η μεσαία τάξη κατέκλυζε την αγορά. **Σήμερα**, μία στις δύο επιχειρήσεις έχουν κλείσει.
Subset – Set	Σε όλες σχεδόν τις κατοικημένες **περιοχές** του **κέντρου** υπάρχει η ίδια εικόνα ερήμωσης.
Part – Whole	Ο αρχικός ρόλος της Γιορτής του Βιβλίου ήταν η άμεση επαφή ενός κοινού που δεν είχε συνηθίσει να ανοίγει την **πόρτα** ενός **βιβλιοπωλείου** για να βρει τι να διαβάσει.
Possession – Owner	Οι **ιδιοκτήτες** των **καταστημάτων** αρνούνται να ανοίξουν τις Κυριακές.
Genitive subject – predicate	Η **απεργία** των διοικητικών **υπαλλήλων** αρχίζει την Τετάρτη και θα διαρκέσει έως και την Παρασκευή.
Genitive object – predicate	Οι συμβασιούχοι αντιδρούν στην επικείμενη **προκήρυξη διαγωνισμού** για προσλήψεις στον οργανισμό την ώρα που οι δικές τους συμβάσεις λήγουν.
Entity & Property	Στη ζύμη της πίτσας ρίχνουμε ένα **κουταλάκι** του **γλυκού** αλάτι.
Function – Object	Ο **διοικητής** του **ΟΑΕΕ** πέρασε πριν λίγο την πόρτα του κτηρίου του οργανισμού στην οδό Ακαδημίας, που βρίσκεται από νωρίς το πρωί υπό κατάληψη.
Rest	Η διεξαγωγή των προεδρικών **εκλογών** στην **Ουκρανία** ήταν σύμφωνη με τις διεθνείς δεσμεύσεις της χώρας και με τις θεμελιώδεις ελευθερίες. Η **ομηρία** των εργατών έληξε μετά από **διαπραγματεύσεις**.

Κατευθυντικότητα σχέσεων (1)

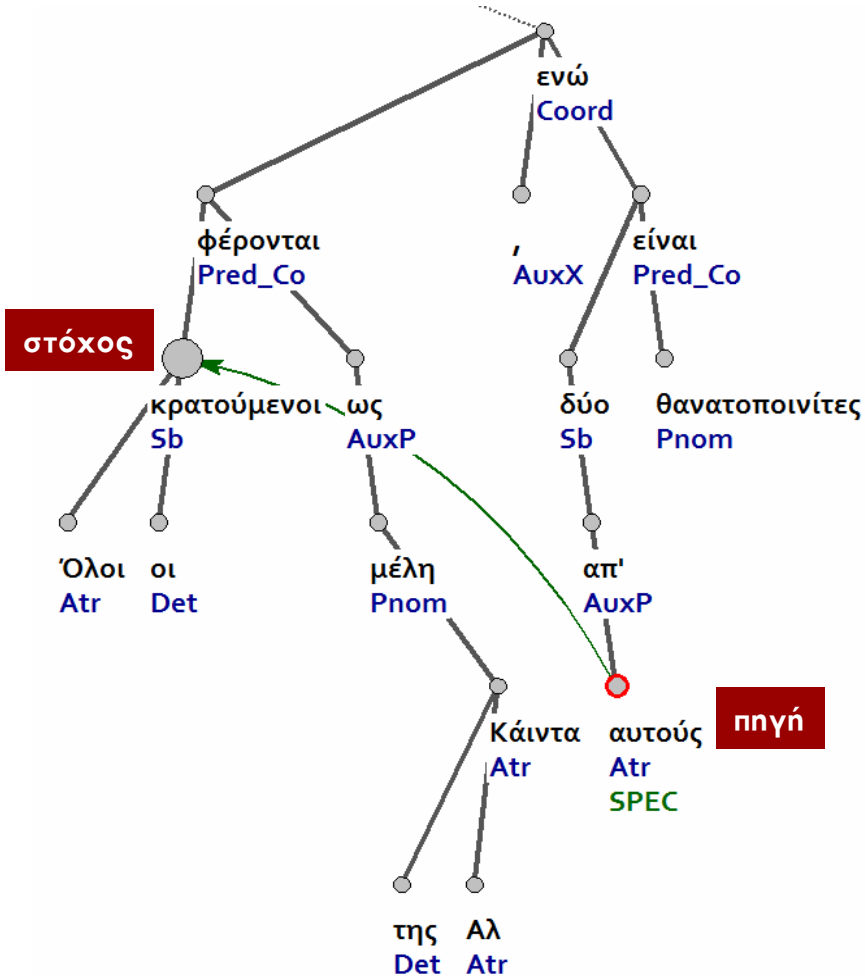
<**στόχος**, πηγή>

- Συναναφορά: αριστερά ← δεξιά
- Οιονεί συναναφορά: αριστερά ← δεξιά

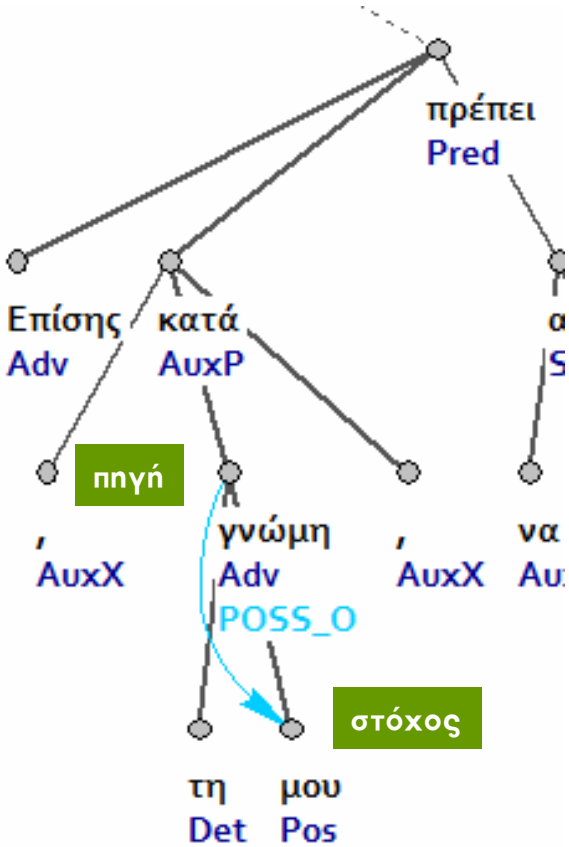
<πηγή, **στόχος**>

- Γεφύρωση: αριστερά → δεξιά

Παραδείγματα



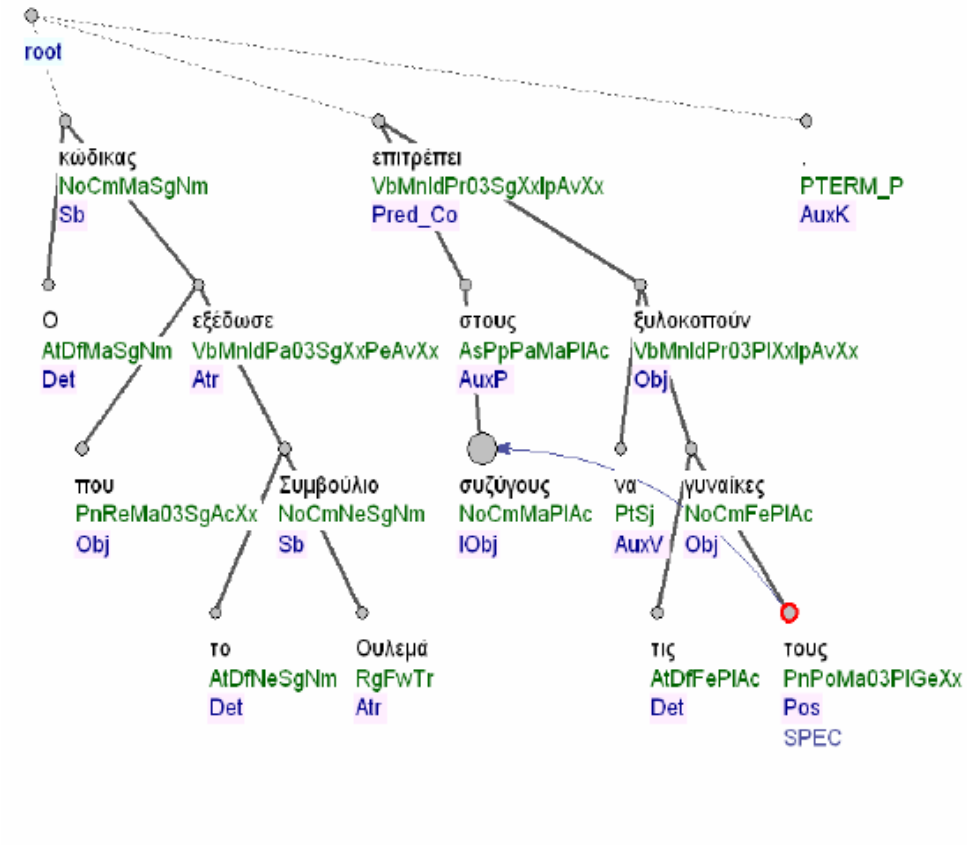
Όλοι οι **κρατούμενοι** φέρονται ως μέλη της Αλ Κάιντα, ενώ δύο απ' **αυτούς** είναι θανατοποινίτες.



Επίσης, κατά τη **γνώμη μου** …

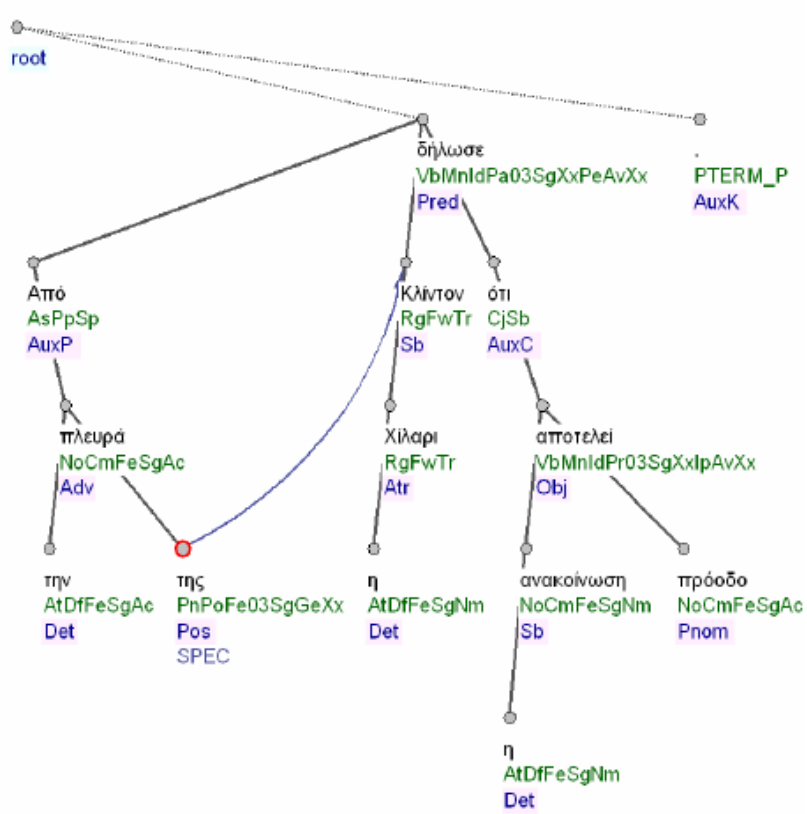
Κατευθυντικότητα σχέσεων (2)

Αναφορά



Ο κώδικας που εξέδωσε το Συμβούλιο Ουλεμά επιτρέπει στους **συζύγους** να ξυλοκοπούν τις γυναίκες **τους**.

Καταφορά



Από την πλευρά **της**, η Χίλαρι **Κλίντον** δήλωσε σε αρμόδια επιτροπή του Κογκρέσου ότι η ανακοίνωση αποτελεί πρόοδο.

Γεφύρωση: Μόνο μια φορά

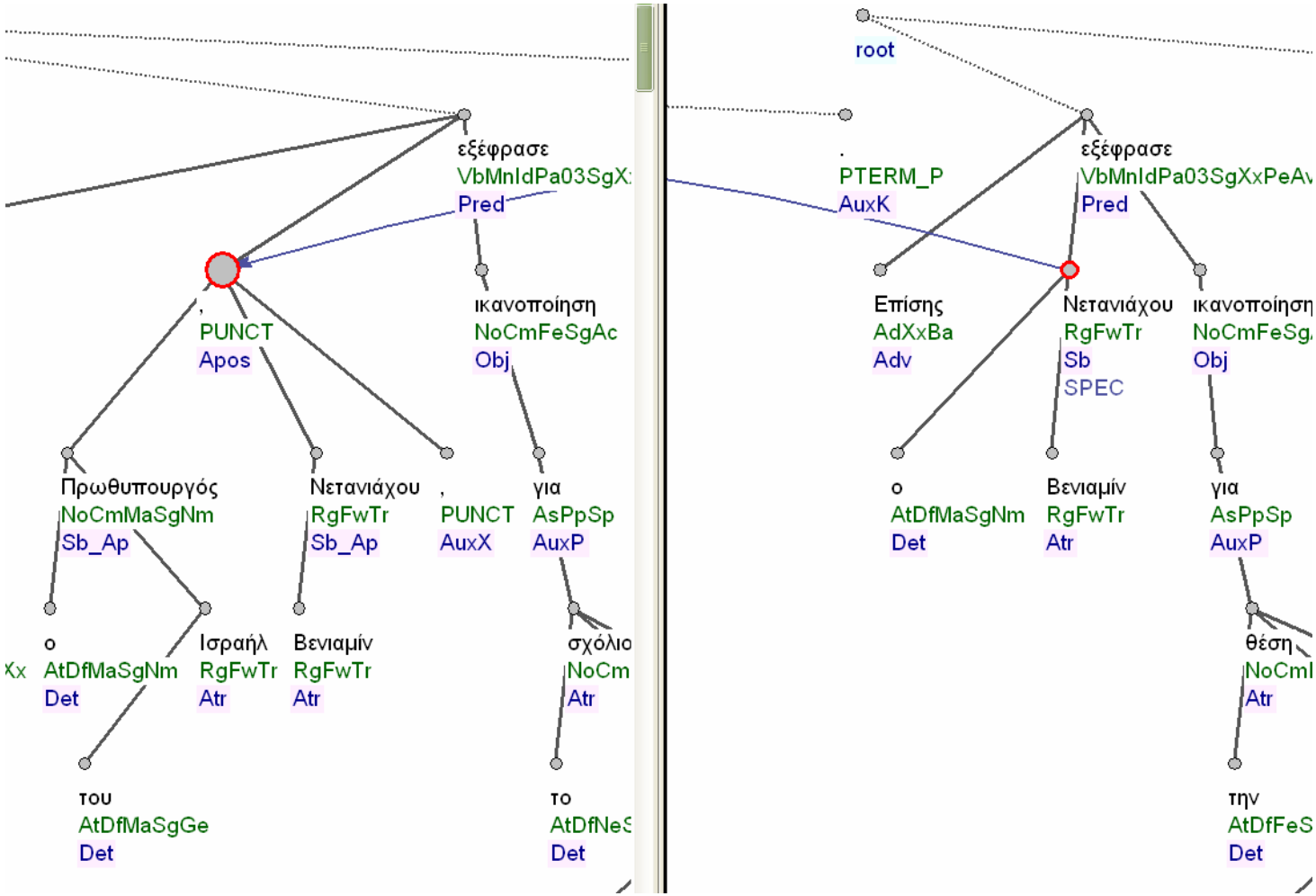
- Οι σχέσεις γεφύρωσης επισημειώνονται **μόνο μία** φορά, ακόμη κι αν επαναλαμβάνονται μέσα στο κείμενο

«Υποστηρίζουμε την επιθυμία των Αρχών να αποφευχθούν οριζόντιες **περικοπές δαπανών** για μισθούς και συντάξεις. Αλλά αυτός είναι και ο λόγος που είναι σημαντικό να προχωρήσουν οι διαρθρωτικές μεταρρυθμίσεις για τον εκσυγχρονισμό των δημοσιονομικών θεσμών στην Ελλάδα, όπως είναι η ενίσχυση της φορολογικής διοίκησης, όπου η πρόοδος εξακολουθεί να υστερεί, έτσι ώστε ο καθένας να πληρώνει το μερίδιο από τους φόρους που του αναλογεί. Μόνο με την πρόοδο σε αυτούς τους τομείς θα είναι δυνατόν να αποφευχθούν περαιτέρω οδυνηρές **περικοπές δαπανών**», ανέφερε.

↑

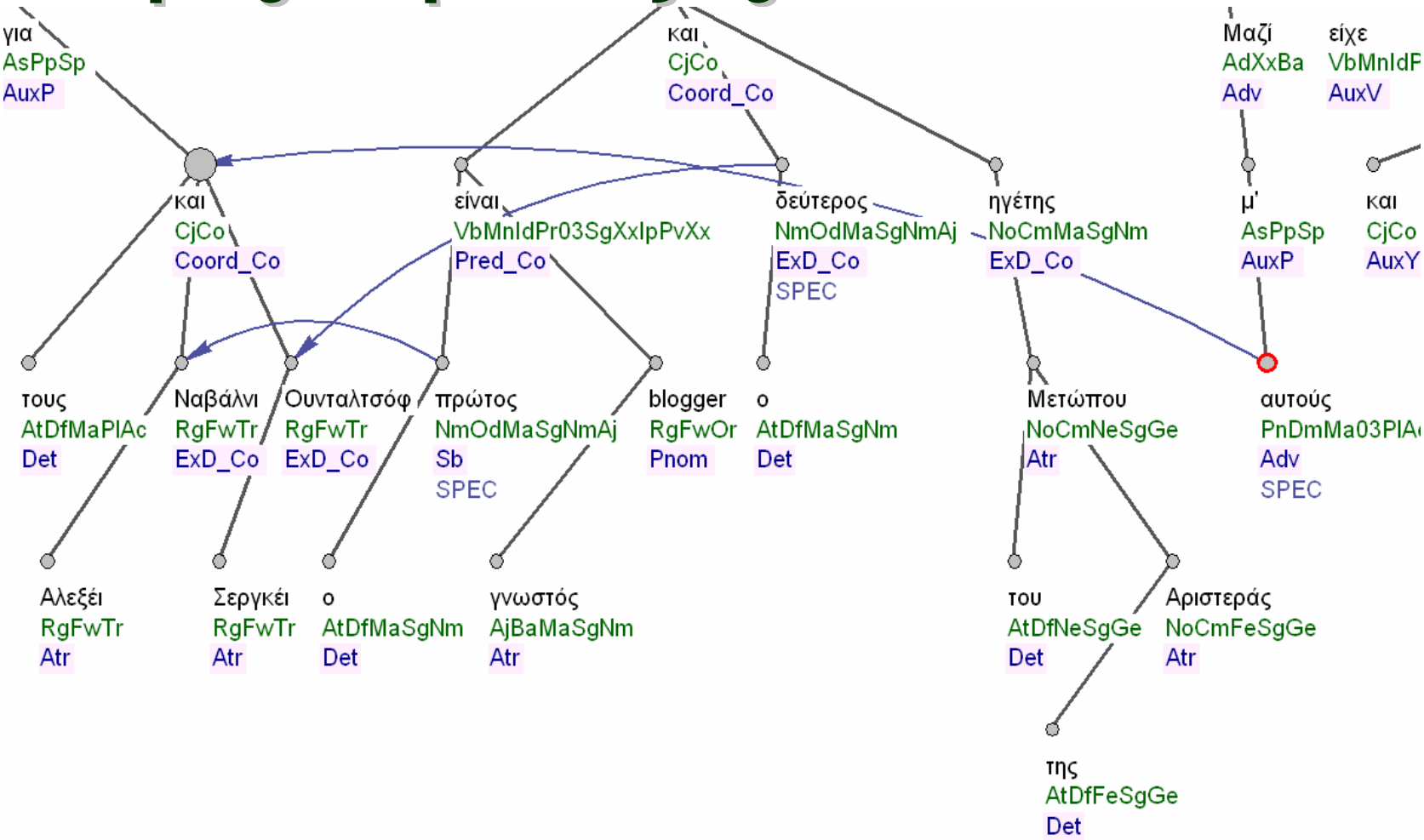
ΔΕΝ επισημειώνεται

Παραθετικές δομές



Από τη πλευρά του ο Πρωθυπουργός του Ισραήλ, Βενιαμίν Νετανιάχου, εξέφρασε ικανοποίηση για το σχόλιο του Μπαράκ Ομπάμα … Επίσης ο Βενιαμίν **Νετανιάχου** εξέφρασε ικανοποίηση …

Δομές παράταξης



Ο λόγος για τους Αλεξέι **Ναβάλνι** **και** Σεργκέι **Ουνταλτσόφ** - ο **πρώτος** είναι γνωστός blogger και ο **δεύτερος** ηγέτης του Μετώπου της Αριστεράς. Μαζί μ' **αυτούς** είχε συλληφθεί και ο ηγέτης του κινήματος «Σαλιντάρνοστ», Ιλιά Ιάτσιν.

❹ Εφαρμογές

Μηχανική μετάφραση (1)

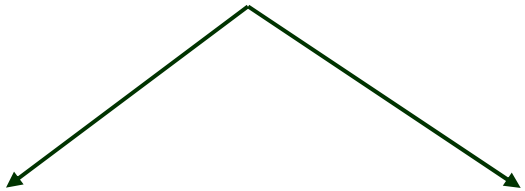
Η εφημερίδα, επίσης, χαρακτηρίζει ανειλικρινή τη δήλωση του Πούτιν ότι δεν μπορεί να επηρεάσει τους εισαγγελείς.

The newspaper also characterises as insincere **Putin's** statement that **it** can not influence public prosecutors.

The newspaper also characterises as insincere **Putin's** statement that **he** can not influence public prosecutors.

Μηχανική μετάφραση (2)

Η ΕΕ πρέπει να δράσει εκτός των συνόρων της για να προωθήσει την ειρήνη και την ανάπτυξη.



The **EU** must act outside **her** borders to promote peace and development.

The **EU** must act outside **its** borders to promote peace and development.

❺ TrEd & GDT_Bridging

Γιατί το TrEd;

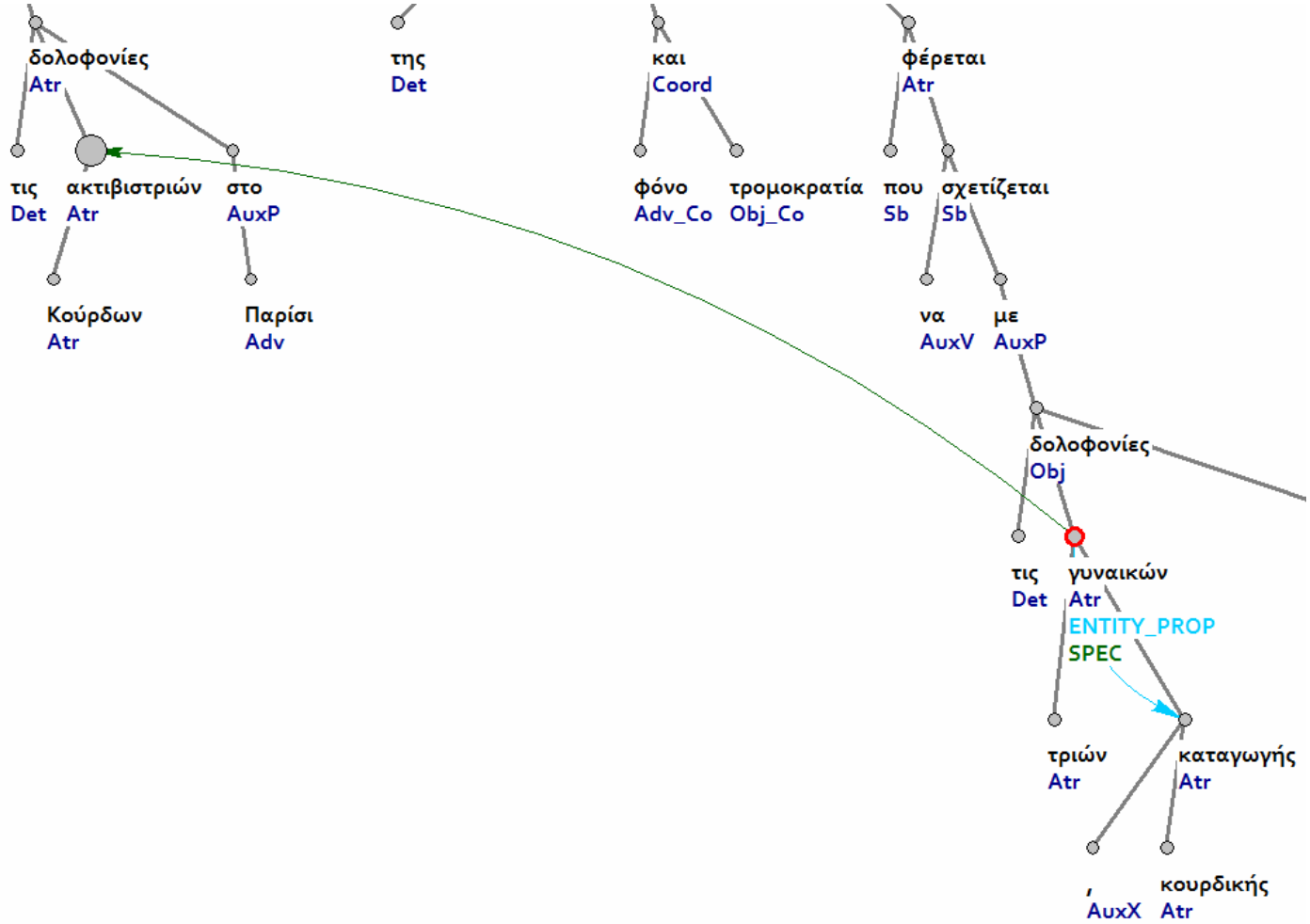
- Είχε χρησιμοποιηθεί και στην επισημείωση των συντακτικών σχέσεων εξάρτησης & υπήρχε η σχετική εμπειρία
- Δυνατότητα άμεσης προσθήκης ενός επιπλέον επιπέδου επισημείωσης
- Ανοικτού κώδικα → δυνατότητα τροποποιήσεων
- Το σχήμα επισημείωσης ομοιάζει κατά πολύ στο PDT

Διαδικασία επισημείωσης

- Κάθε σχέση ορίζεται μεταξύ 2 κόμβων.
- Αναπαρίσταται ως βέλος από τον κόμβο-πηγή στον κόμβο-στόχο.
- Το χρώμα του βέλους εκφράζει το είδος της σχέσης.
- Οι σχέσεις ορίζονται τόσο εντός της περιόδου όσο και εκτός των ορίων αυτής.
- Σε κάθε σχέση πρέπει να οριστεί ο τύπος της.
  - Συναναφορά: SPEC ή GEN (ειδική ή γενικευτική αναφορά)
  - Οιονεί συναναφορά: 5 τύποι
  - Γεφύρωση: 16 τύποι
- Η σχέση συναναφοράς μεταξύ ίδιων λημμάτων μπορεί να επισημειωθεί αμέσως και όχι ανά 2 κόμβους.

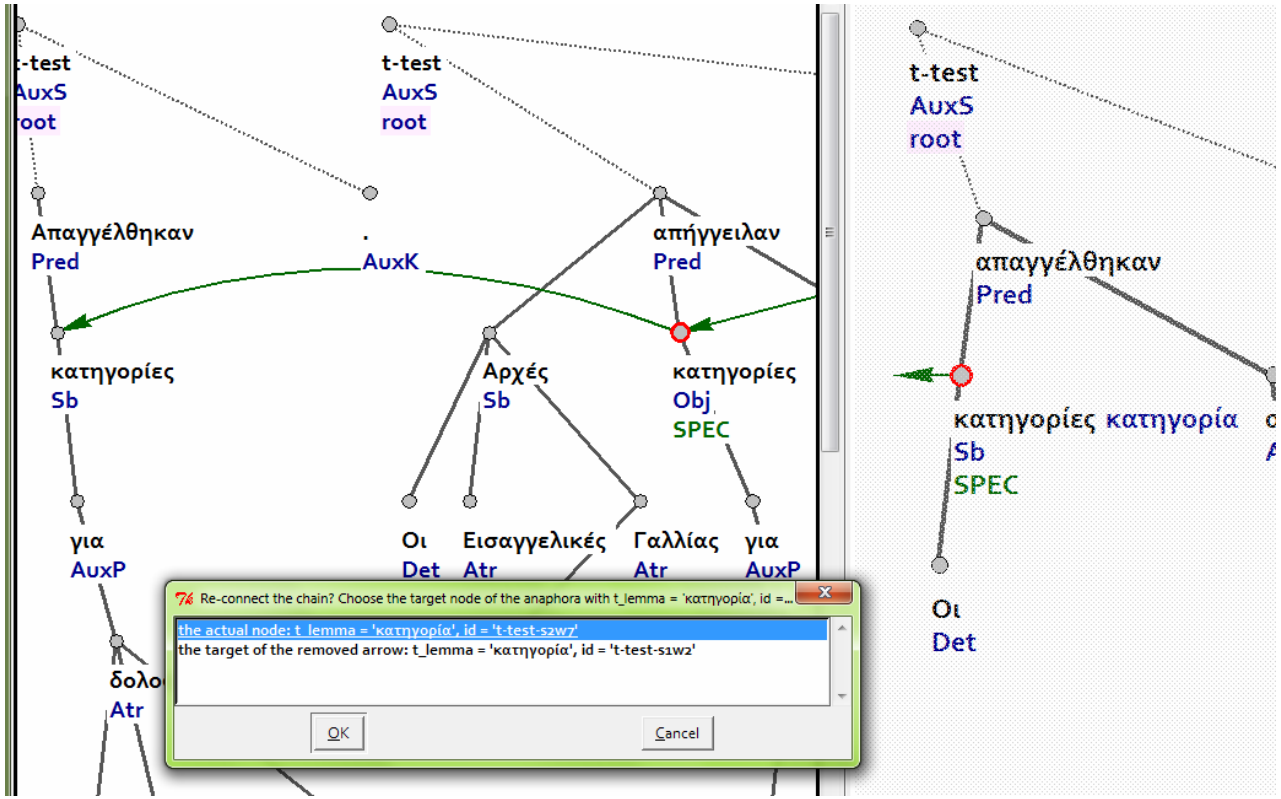
Διαδικασία επισημείωσης (2)

- Από ένα κόμβο είναι δυνατό να ξεκινούν περισσότερα από ένα βέλη.



Διαδικασία επισημείωσης (3)

- Η διαγραφή ενός μέλους σε μία πολυμελή αλυσίδα δε διασπά τη συνέχεια της αλυσίδας, καθώς το TrEd ζητά από τον επισημειωτή να επανασυνδέσει τα υπόλοιπα μέλη.



Πληροφορίες

- Οδηγίες επισημείωσης σχέσεων συναναφοράς

  http://gdt.ilsp.gr/guidelines/coreference/coreferenceannotationguidelines

- Σελίδα για τη συναναφορά

  http://gdt.ilsp.gr/annotators/texno_201405/corefexercise

Βιβλιογραφία (1)

Bagga Amit & Baldwin Breck (1998): **Algorithms for scoring coreference chains**. LREC Workshop on Linguistic Coreference, pp. 563–566

Bertran Manuel, Borrega Oriol, Recasens Marta & Soriano Bàrbara (2008): **AnCoraPipe: A tool for multilevel annotation**. Procesamiento del Lenguaje Natural, 41, 291–292

Cai Jie and Strube Michael (2010): **Evaluation Metrics For End-to-End Coreference Resolution Systems**. The 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2010), Tokyo, Japan, pp. 28–36

Chen Chen and Ng Vincent (2013): **Linguistically Aware Coreference Evaluation Metrics**. International Joint Conference on Natural Language Processing, Nagoya, Japan, pp. 1366–1374

O'Connor Brendan & Heilman Michael (2013): **ARKref: a rule-based coreference resolution system**. CoRR, 1310.1975. Available at: http://arxiv.org/pdf/1310.1975v1.pdf

Doddington George, Mitchell Alexis, Przybocki Mark, Ramshaw Lance, Strassel Stephanie & Weischedel Ralph (2004): **The Automatic Content Extraction (ACE) program—Tasks, data, and evaluation**. LREC 2004, Lisbon, Portugal, pp. 837–840

Fligelstone Steve (1992): **Developing a Scheme for Annotating Text to Show Anaphoric Relations**. In: G. Leitner (ed.), New Directions in Corpus Linguistics. Berlin: Mouton de Gruyter, pp. 153-170

Haghighi Aria & Klein Dan (2009): **Simple Coreference Resolution with Rich Syntactic and Semantic Features**. 2009 Conference on Empirical Methods in Natural Language Processing, Singapore, pp.1152–1161

Hirschman Lynette and Chinchor Nancy (1997): **MUC-7 Coreference Task Definition, Version 3.0**. Available at: http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/co_task.html

Βιβλιογραφία (2)

Luo Xiaoqiang (2005): **On coreference resolution performance metrics**. Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 25–32

Nědolužko Anna and Mírovský Jiří (2011): **Annotating extended textual coreference and bridging relations in the Prague Dependency Treebank**. Technical report. Prague, Czech Republic. Available at: http://ufal.mff.cuni.cz/~nedoluzko/koref_anot/manual_angl_final_printed.pdf

Nguyen Ngan, Kim Jin-Dong and Tsujii Jun'ichi (2011): **Overview of BioNLP 2011 Protein Coreference Shared Task**. BioNLP Shared Task 2011 Workshop, Portland, Oregon, USA, pp. 74-82

OntoNotes Release 5.0. Available at: https://catalog.ldc.upenn.edu/docs/LDC2013T19/OntoNotes-Release-5.0.pdf

Pajas Petr & Stepanek Jan (2008): **Recent Advances in a Feature-Rich Framework for Treebank Annotation**. The 22nd International Conference on Computational Linguistics (Coling 2008), Manchester, UK, pp. 673-680

Poesio Massimo, Bruneseaux Florence & Romary Laurent (1999): **The MATE meta-scheme for coreference in dialogues in multiple languages**. ACL'99 Workshop Towards Standards and Tools for Discourse Tagging, pp. 65-74

Recasens Marta & Martí M. Antònia (2010): **AnCora-CO: coreferentially annotated corpora for Spanish and Catalan**. Language Resources and Evaluation 44(4): 315–345

Versley Yannick, Moschitti Alessandro, Poesio Massimo & Yang Xiaofeng (2008a): **Coreference Systems based on Kernel Methods**. The 22nd International Conference on Computational Linguistics (Coling 2008), Manchester, UK, pp. 961-968

Versley Yannick , Ponzetto Simone, Poesio Massimo, Eidelman Vladimir, Jern Alan, Smith Jason, Yang Xiaofeng & Moschitti Alessandro (2008b): **BART: A Modular Toolkit for Coreference Resolution**. 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco, pp. 962-965

Vilain Marc, Burger John, Aberdeen John, Connolly Dennis & Hirschman Lynette (1995): **A model theoretic coreference scoring scheme**. 6th Message Understanding Conference, pp. 45–52

❻ Επισημείωση στην τάξη