

# COREFERENCE & BRIDGING ANNOTATION GUIDELINES



<b>Title</b>	<b>Coreference &amp; Bridging annotation guidelines</b>
<b>Date</b>	10.6.2014
<b>Version</b>	30
<b>Authors</b>	GCBT [Greek Coreference & Bridging Team]

## Table of Contents

<b>1.</b>	<b>INTRODUCTION .....</b>	<b>5</b>
<b>2.</b>	<b>DEFINING TERMS .....</b>	<b>6</b>
<b>3.</b>	<b>COREFERENCE .....</b>	<b>10</b>
3.1	RELATIVE PRONOUNS .....	11
3.2	NULL SUBJECTS IN CONTROL STRUCTURES .....	12
3.3	NULL SUBJECTS IN FINITE CLAUSES.....	12
3.4	CLITICS.....	13
3.5	OTHER PRONOMINAL ELEMENTS .....	14
3.6	REFERENTIAL ADVERBS .....	14
3.7	DISCOURSE COREFERENTS.....	15
3.8	METONYMY .....	15
<b>4.</b>	<b>METONYMY AS A MECHANISM OF COREFERENCE.....</b>	<b>16</b>
4.1	WITHIN-SETS .....	17
4.2	ABSTRACT-CONCRETE.....	17
4.3	NE-CREATION.....	17
4.4	CONSTITUENT .....	18
4.5	CONTAINER-CONTENT.....	18
4.6	INSTRUMENT .....	18
4.7	INSTITUTION .....	19
4.8	ASPECT .....	19
<b>5.</b>	<b>NEAR IDENTITY .....</b>	<b>20</b>
5.1	LOCATION .....	21
5.2	ORGANISATION .....	21
5.3	INSTANTIATION .....	21
5.4	SIMILAR SETS .....	21
5.5	OTHER .....	22
<b>6.</b>	<b>BRIDGING.....</b>	<b>23</b>
6.1	CONTRAST.....	23
6.2	SUBSET – SET & SET – SUBSET .....	23
6.3	PART – WHOLE & WHOLE – PART.....	24
6.4	POSSESSION – OWNER & OWNER – POSSESSION.....	24
6.5	GENITIVE SUBJECT IN A COMPLEX NP CONSTRUCTION .....	24
6.6	GENITIVE OBJECT IN A COMPLEX NP CONSTRUCTION .....	24
6.7	ENTITY & PROPERTY .....	24
6.8	FUNCTION – OBJECT & OBJECT – FUNCTION .....	25
6.9	REST .....	25
<b>7.</b>	<b>LINGUISTIC EXPRESSIONS .....</b>	<b>26</b>
7.1	NOMINAL & PRONOMINAL HEADS.....	26
7.2	ADVERBIAL HEADS .....	26
7.3	NULL SUBJECTS .....	26
7.4	APPOSITIONAL STRUCTURES.....	27
7.5	COORDINATING STRUCTURES.....	28
<b>8.</b>	<b>CONVENTIONS &amp; CASE-SPECIFIC GUIDELINES .....</b>	<b>29</b>
8.1	ANNOTATION CONVENTIONS: COREFERENCE .....	29
8.2	ANNOTATION CONVENTIONS: BRIDGING.....	30
8.3	ANNOTATION ROUTE.....	32
8.4	NON-ELIGIBLE EXPRESSIONS.....	33
8.5	CASE-SPECIFIC GUIDELINES.....	33
8.5.1	<i>Cataphora with a non-pronominal element.....</i>	<i>33</i>

8.5.2	<i>Copula verb structures</i> .....	35
8.5.3	<i>Null subjects in coordinating structures</i> .....	35
8.5.4	<i>Null subjects of multi-token verbs</i> .....	36
8.5.5	<i>Location names</i> .....	36
8.5.6	<i>Expression-specific bridging relations</i> .....	36
<b>9.</b>	<b>ANNOTATING WITH TRED &amp; GDT_BRIDGING</b> .....	<b>38</b>
9.1	USING THE GDT_BRIDGING MACROS .....	38
9.2	FAQ .....	42
<b>10.</b>	<b>REFERENCES</b> .....	<b>44</b>

## List of Figures

FIGURE 1:	COREFERENCE WITH IDENTICAL LEMMATA .....	10
FIGURE 2:	COREFERENCE WITH A RELATIVE ELEMENT .....	11
FIGURE 3:	COREFERENCE IN CONTROL STRUCTURES .....	12
FIGURE 4:	COREFERENCE WITH THE NULL SUBJECT OF A FINITE CLAUSE .....	12
FIGURE 5:	COREFERENCE WITH PERSONAL CLITICS .....	13
FIGURE 6:	COREFERENCE WITH POSSESSIVE CLITICS .....	13
FIGURE 7:	COREFERENCE WITH DEMONSTRATIVE PRONOUNS .....	14
FIGURE 8:	COREFERENCE WITH REFERENTIAL ADVERBS .....	14
FIGURE 9:	COREFERENCE WITH DISCOURSE COREFERENTS .....	15
FIGURE 10:	COREFERENCE VIA METONYMY .....	15
FIGURE 11:	EXAMPLE OF A NEAR IDENTITY RELATION .....	20
FIGURE 12:	APPOSITIONAL STRUCTURE .....	27
FIGURE 13:	COORDINATING STRUCTURE .....	28
FIGURE 14:	ANAPHORA & CATAPHORA CHAINS WITH A PRONOMINAL ELEMENT .....	29
FIGURE 15:	ANAPHORA CHAIN WITH A NON-PRONOMINAL ELEMENT .....	30
FIGURE 16:	BRIDGING CHAIN .....	30
FIGURE 17:	ANNOTATION ROUTE .....	32
FIGURE 18:	CATAPHORA CHAIN WITH A NON-PRONOMINAL ELEMENT (A) .....	34
FIGURE 19:	CATAPHORA CHAIN WITH A NON-PRONOMINAL ELEMENT (B) .....	34
FIGURE 20:	NULL SUBJECT IN A COORDINATING STRUCTURE .....	35
FIGURE 21:	NULL SUBJECT OF A MULTI-TOKEN VERB .....	36

## List of Tables

TABLE 1:	LIST OF COREFERENCE CONTEXTS .....	11
TABLE 2:	METONYMY RELATIONS .....	16
TABLE 3:	NEAR IDENTITY TYPES .....	20
TABLE 4:	LIST OF BRIDGING RELATION TYPES .....	23
TABLE 5:	LIST OF EXPRESSIONS & THEIR BRIDGING RELATIONS .....	36

TABLE 6: TrED – ANNOTATING COREFERENCE .....	38
TABLE 7: TrED – ANNOTATING NEAR IDENTITY .....	39
TABLE 8: TrED – ANNOTATING BRIDGING .....	39
TABLE 9: TrED – ADDING NULL SUBJECTS .....	39
TABLE 10: TrED – EDITING NODES .....	40
TABLE 11: TrED – MOVING TO & FROM ANTECEDENTS .....	40
TABLE 12: TrED – SWITCH DISPLAY .....	40
TABLE 13: TrED – ADDING COMMENTS .....	41

## 1. INTRODUCTION

The present document provides a set of guidelines for the annotation of coreference and bridging relations in Greek texts. **Coreference** denotes a relation between linguistic expressions, which have the same referent, i.e. they point to the same entity. **Bridging** denotes an associative relation (such as PART-WHOLE or SET-SUBSET) between two non-coreferential linguistic expressions. Both coreference and bridging are considered to be mechanisms contributing to cohesion (and coherence) within a text.

This document is organised as follows: Section 2 contains a list of terms and their definitions that will come up during the annotation process. Sections 3 and 4 specify coreference in detail, while Section 5 exemplifies cases of near identity. Bridging relations are the subject of Section 6. Section 7 identifies the linguistic expressions that are eligible as participants in a linking relation (coreference, near identity or bridging). Section 8 cites the annotation conventions that have been decided upon and provides general and case-specific annotation guidelines. Finally, Section 9 discusses the TrEd manual annotation tool and especially the GDT\_Bridging extension, which is employed for the annotation process of Greek texts represented as sequences of dependency trees. References are provided in Section 10.

## 2. DEFINING TERMS

The following table lists the definitions of terms that will come up during the annotation process.

Term	Definition	Equivalent term
<b>Abstract-Concrete</b>	Metonymy relation, by which an expression with an abstract reference is employed in the place of the corresponding concrete one	
<b>Anaphora chain</b>	A coreferential chain having the form <antecedent, source>	
<b>Antecedent</b>	A word, phrase, or clause referred to endophorically by another expression, which follows it	
<b>Arrow</b>	In the TrEd annotation tool, it represents a linking relation between expressions; it assumes a different colouring depending on the relation type Coreference: green arrow Near identity arrow: red arrow Bridging relation: light blue arrow	
<b>Aspect</b>	Metonymy relation, by which an entity-denoting expression is coreferent with an expression highlighting an aspect of that entity such as role, property, kinship, origin etc.	
<b>Bridging chain</b>	Set of associated non-coreferential mentions; it contains only two members	
<b>Bridging</b>	Associative relation between two non-coreferential mentions	Bridging relation [BR]
<b>Cataphora chain</b>	A coreferential chain having the form <source, postcedent>	
<b>Chain</b>	Set of linked mentions	
<b>Constituent</b>	A broad term capturing two metonymy relations, Part-Whole and Material-Object	
<b>Container-Content</b>	Metonymy relation, by which an expression designating a container is used for referring to the corresponding content	
<b>CONTRAST</b>	Bridging relation type denoting contrast between two linguistic expressions	
<b>Cor</b>	Null subject in a control structure	
<b>Coreference</b>	Identity relation between (at least) two mentions	Coreference relation [CR]
<b>Coreference chain</b>	Set of mentions with the same referent; it contains at least two members	Coreferential chain
<b>Coreferential</b>	Property of linguistic expressions which have the same referent	coreferent
<b>Discourse coreferents</b>	Non-synonymous expressions which are coreferential within a given context	
<b>ENTITY_PROP</b>	Bridging relation type denoting a object-property relation between two linguistic expressions	
<b>FUNCT_P</b>	Bridging relation type denoting a function-object relation between two linguistic expressions	
<b>GEN</b>	A coreference relation feature designating expressions with generic reference	
<b>Instantiation</b>	Near identity relation type	

Term	Definition	Equivalent term
<b>Institution</b>	A broad term covering two metonymy relations, Seat-Institution and Institution-Person	
<b>Institution-Person</b>	Subtype of the "Institution" metonymy relation, by which an institution-designating expression is used with a reference to the person(s) instantiating the given institution	
<b>Instrument</b>	A broad term covering two metonymy relations, Instrument-Act and Instrument-Agent	
<b>Instrument-Act</b>	Subtype of the "Instrument" metonymy relation, by which an act is referred to via the corresponding instrument	
<b>Instrument-Agent</b>	Subtype of the "Instrument" metonymy relation, by which the agent of an act is referred to via the corresponding instrument	
<b>Linking relation</b>	A broad term capturing coreference, near identity & bridging relations	
<b>Location</b>	Near identity relation type	
<b>Material-Object</b>	Subtype of the "Constituent" metonymy relation, by which reference is made to an object through the use of an expression that identifies the basic constituent material of that object	
<b>Mention</b>	Instance of reference to an entity	
<b>Metonymic</b>	Property of a relation between linguistic expressions	
<b>Metonymy</b>	A "replacement" mechanism, by which a linguistic expression, referring to entity A, is used for referring to a different yet closely related entity B	Metonymy mechanism
<b>Near Identity [neid]</b>	A quasi coreferential relation between two linguistic expressions, which only partially share the same referent	
<b>NE-Creation</b>	Metonymy relation, by which a named entity is used for referring to a work created by or related to that entity	
<b>O_POSS</b>	Bridging relation type denoting an owner-possession relation between two linguistic expressions	
<b>OBJGEN_P</b>	Bridging relation type denoting an object-predicate relation between two linguistic expressions, the first of which is in genitive case	
<b>One-Many</b>	Subtype of the "Within sets" metonymy relation, by which a member of a set is used with a reference to the whole set	
<b>Organisation</b>	Near identity relation type	
<b>P_FUNCT</b>	Bridging relation type denoting an object-function relation between two linguistic expressions	
<b>P_OBJGEN</b>	Bridging relation type denoting a predicate-object relation between two linguistic expressions, the second of which is in genitive case	
<b>P_SUBJGEN</b>	Bridging relation type denoting a predicate-subject relation between two linguistic expressions, the second of which is in genitive case	
<b>PART_WHOLE</b>	Bridging relation type denoting a part-whole relation between two linguistic expressions	
<b>Part-Whole</b>	Subtype of the "Constituent" metonymy relation, by which a constituent part of an entity is mentioned, while reference is made to the given entity as a whole	
<b>PersPron</b>	Null subject in a finite clause	

Term	Definition	Equivalent term
PoS	Part of Speech	
POSS_O	Bridging relation type denoting a possession-owner relation between two linguistic expressions	
Postcedent	A word, phrase, or clause referred to endophorically by another expression, which precedes it	
PROP_ENTITY	Bridging relation type denoting a property-object relation between two linguistic expressions	
Referent	Entity to which a linguistic expression refers	
REST	Bridging relation type denoting a relation between two linguistic expressions which can not be subsumed under the other types	
Seat-Institution	Subtype of the "Institution" metonymy relation, by which an institution's seat is used for referring to the institution itself	
SET_SUB	Bridging relation type denoting a set-subset relation between two linguistic expressions	
Set-member	Subtype of the "Within sets" metonymy relation, by which an expression designating a set is used for referring to the contained members	
Similar sets	Near identity relation type	
Singleton	Single mention, standing in no coreference relation with another mention	
Source	<p>The starting point of an arrow</p> <p><b>Annotation conventions:</b></p> <p>In coreference relations involving a pronominal element and a referring expression, the source is always the pronominal element. The target is always the referring expression, i.e. the antecedent, if it precedes the source (anaphora), or the postcedent, if it follows the source (cataphora).</p> <p>In all other cases the CR arrows have a leftward direction, i.e. the source always follows.</p> <p>Neid arrows always have a leftward direction, i.e. the source always follows.</p> <p>In bridging relations (i.e. in associative relations between two non-coreferential expressions) the <b>source</b> is always the first of the two expressions in the order that they appear in the sentence or in the text. The target is always the second of the two expressions.</p>	
SPEC	A coreference relation feature designating expressions with specific reference	
SUB_SET	Bridging relation type denoting subset-set relation between two linguistic expressions	
SUBJGEN_P	Bridging relation type denoting a subject-predicate relation between two linguistic expressions, the first of which is in genitive case	



Term	Definition	Equivalent term
<b>Target</b>	<p>The end point of an arrow</p> <p><b>Annotation conventions:</b></p> <p>In coreference relations involving a pronominal element and a referring expression, the target is always the referring expression, i.e. the antecedent, if it precedes the source (anaphora), or the postcedent, if it follows the source (cataphora). The source is always the pronominal element.</p> <p>In all other cases the CR arrows have a leftward direction, i.e. the target is always the first expression.</p> <p>Neid arrows always have a leftward direction, i.e. the target is always the first expression.</p> <p>In bridging relations (i.e. in associative relations between two non-coreferential expressions) the <b>target</b> is always the second of the two expressions in the order that they appear in the sentence or in the text. The source is always the first of the two expressions.</p>	
<b>WHOLE_PART</b>	Bridging relation type denoting a whole-part relation between two linguistic expressions	
<b>Within-Sets</b>	A broad term capturing two metonymy relations, One-Many and Set-Member	

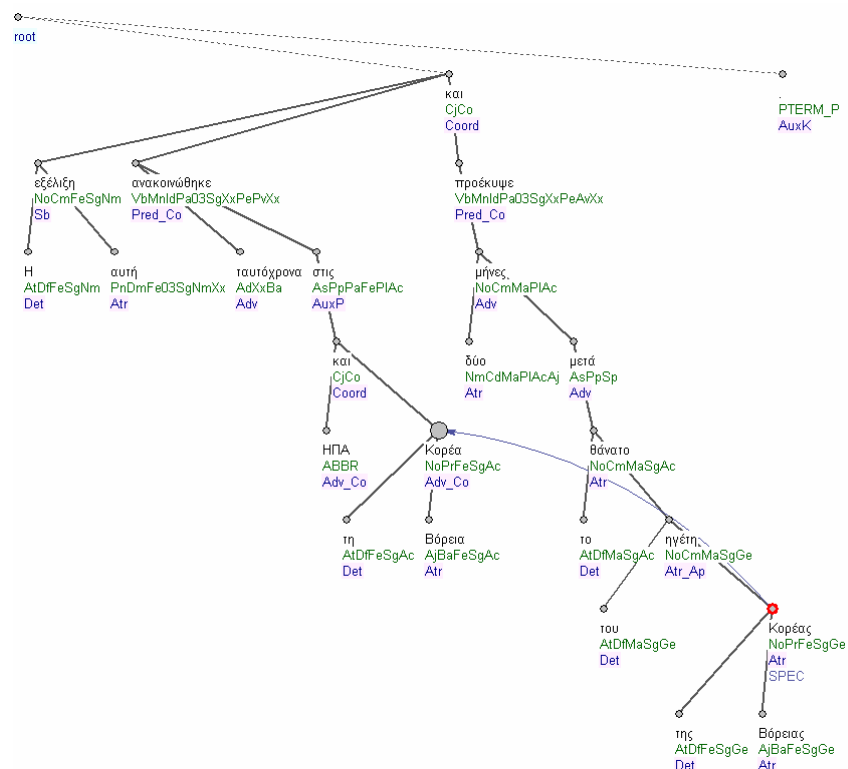
### 3. COREFERENCE

Coreference denotes a relation between (at least) two linguistic expressions, which have the same referent, i.e. they point to the same entity. The linking of these expressions results in a chain, either within the sentence boundaries or spanning more than one sentence<sup>1</sup>. Concrete examples of coreferentially linked expressions are given below, while conventions related to how coreference is annotated are detailed under subsection 8.1.

Coreference typically occurs between identical linguistic expressions, i.e. identical lemmata, since they are considered to normally have the same referent (Figure 1).

Figure 1: Coreference with identical lemmata

Η εξέλιξη αυτή ανακοινώθηκε ταυτόχρονα στις ΗΠΑ και τη Βόρεια **Κορέα** και προέκυψε δύο μήνες μετά το θάνατο του ηγέτη της Βόρειας **Κορέας**.



**NOTE**, however, that the same linguistic expressions do not necessarily share the same referent in all cases due to the presence of different modifiers (cf. the first example) or on the basis of the broader context (cf. the second example).

#### EXAMPLES<sup>2</sup>

Τα κατώτερα κοινωνικά **στρώματα** εμφανίζουν μεγαλύτερη ανησυχία από τα ανώτερα κοινωνικά **στρώματα**.

Ο Σπυρίδων Λάμπρος σχημάτισε **κυβέρνηση** με μη πολιτικά πρόσωπα. ... λίγες μέρες μετά ο Βενιζέλος σχημάτισε **κυβέρνηση** ...

Coreference can also emerge in a variety of other settings, which are listed in Table 1 and exemplified in the subsequent sections.

<sup>1</sup> Due to space considerations mainly within-sentence cases of coreference are exemplified in this section.

<sup>2</sup> Unless otherwise stated, all examples are taken from the GDT corpus or web news sites.

**Table 1:** List of coreference contexts

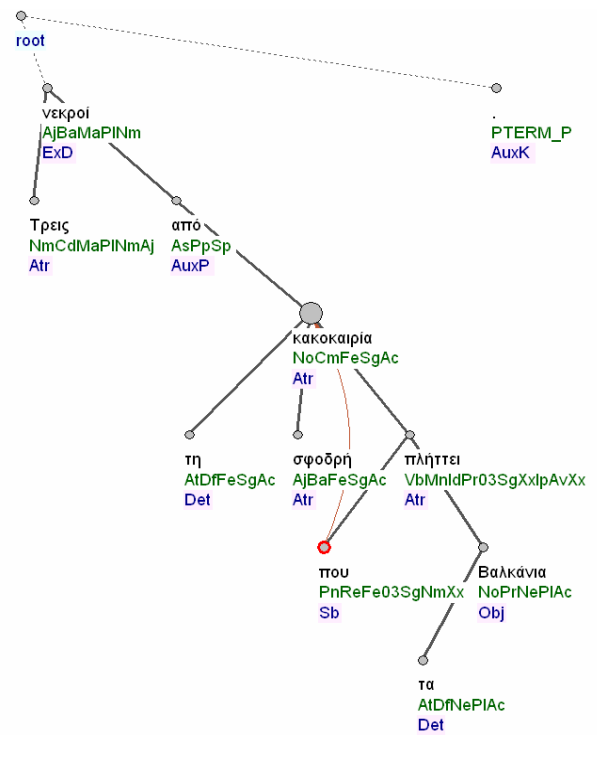
Coreference contexts	
Identical lemmata	
Relative pronouns	
Control structures	
Null subjects in finite clauses	
Clitics	Personal clitics
	Possessive clitics
Other pronominal elements	
Referential adverbs	
Discourse coreferents	
Metonymy	

### 3.1 Relative pronouns

A relative element corefers with the head it refers to (Figure 2).

**Figure 2:** Coreference with a relative element

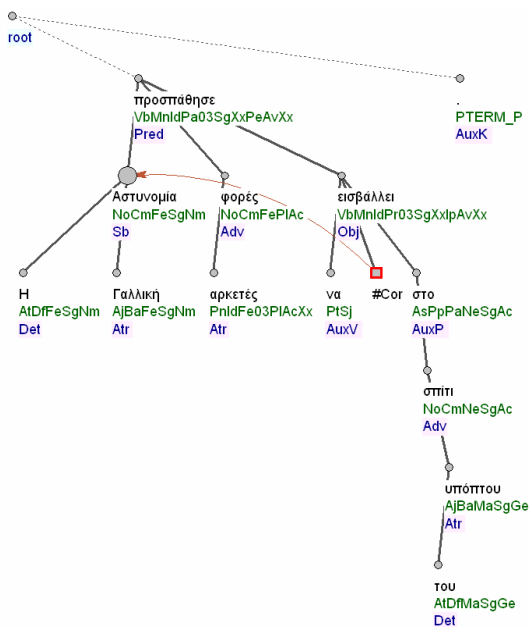
Τρεις νεκροί από τη σφοδρή **κακοκαιρία**  
**που** πλήττει τα Βαλκάνια.



### 3.2 Null subjects in control structures

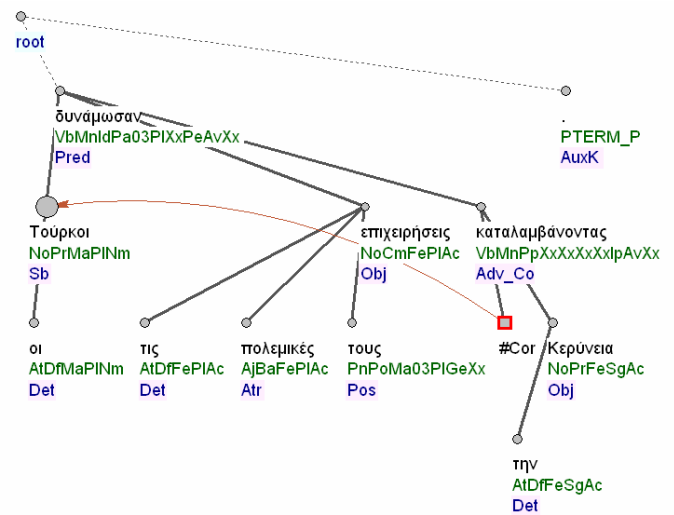
The null subject corefers with its controller (Figure 3).

Figure 3: Coreference in control structures



#### Complement clause

Η Γαλλική **Αστυνομία** προσπάθησε αρκετές φορές να **Cor** εισβάλει στο σπίτι του υπόπτου.



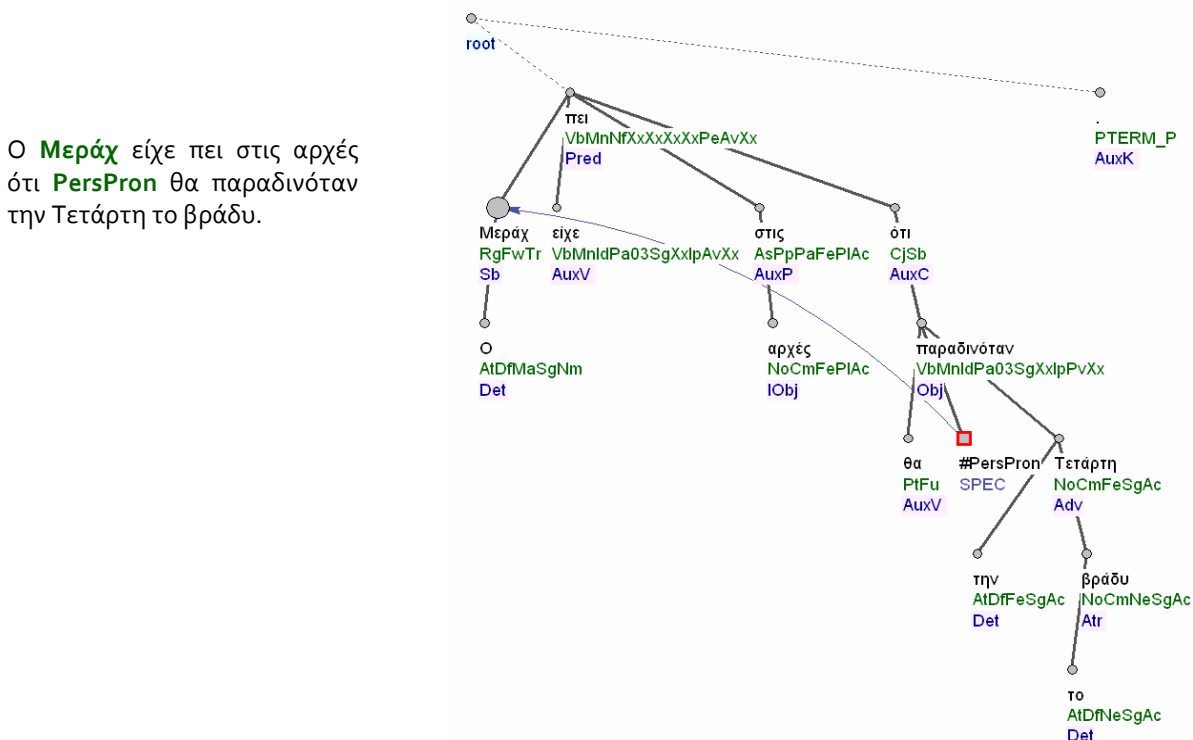
#### Gerund

Οι **Τούρκοι** δυνάμωσαν τις πολεμικές τους επιχειρήσεις, **Cor** καταλαμβάνοντας την Κερύνεια.

### 3.3 Null subjects in finite clauses

The null subject of a finite clause builds a coreferential relation with the linguistic expression it refers to (Figure 4).

Figure 4: Coreference with the null subject of a finite clause



Ο **Μεράχ** είχε πει στις αρχές ότι **PersPron** θα παραδινόταν την Τετάρτη το βράδυ.

### 3.4 Clitics

Clitics, either **personal** (Figure 5) or **possessive** ones (Figure 6) can build a coreference relation.

Figure 5: Coreference with personal clitics

Στην Ουάσιγκτον, ο Λευκός Οίκος εξέφρασε ικανοποίηση για την **απόφαση** της Πιονγκγιάνγκ, χαρακτηρίζοντας **την** θετικό πρώτο βήμα.

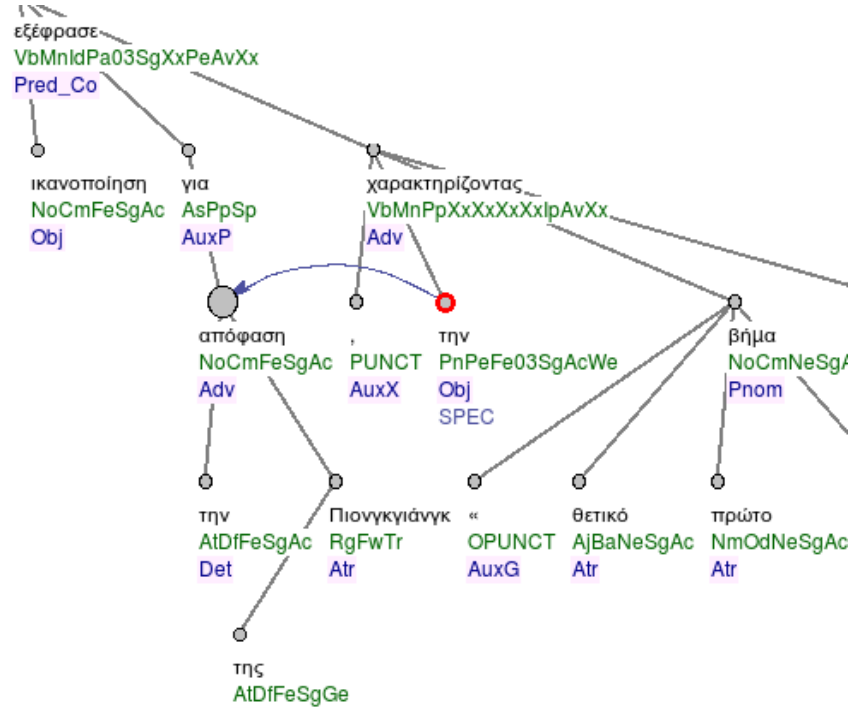
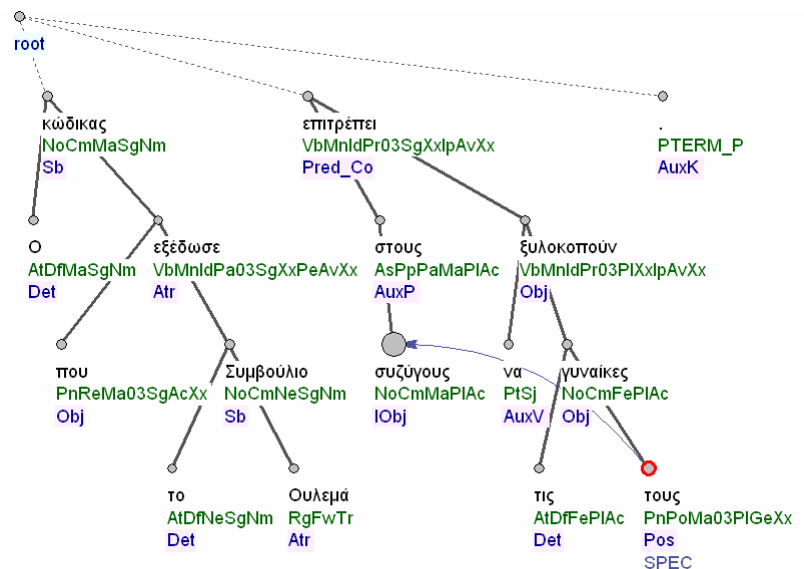


Figure 6: Coreference with possessive clitics

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

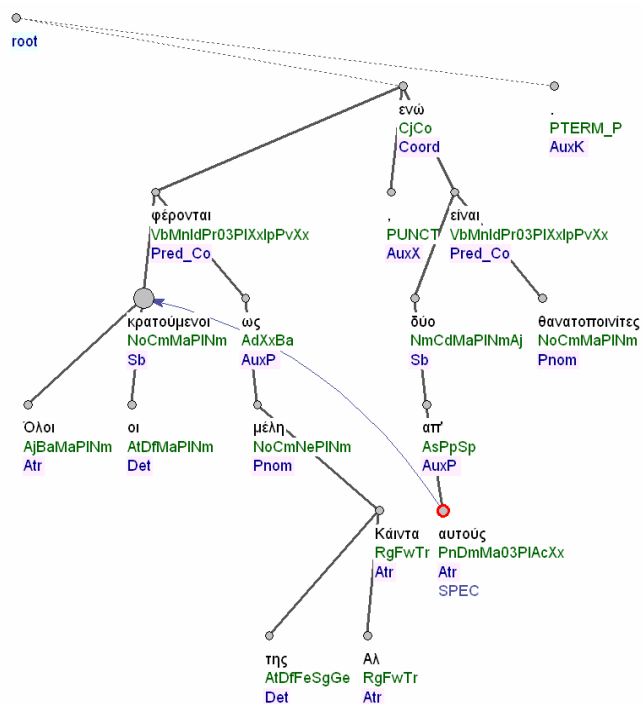


### 3.5 Other pronominal elements

Coreference can be established between a pronominal element (e.g. a demonstrative pronoun) and a linguistic expression (Figure 7).

Figure 7: Coreference with demonstrative pronouns

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

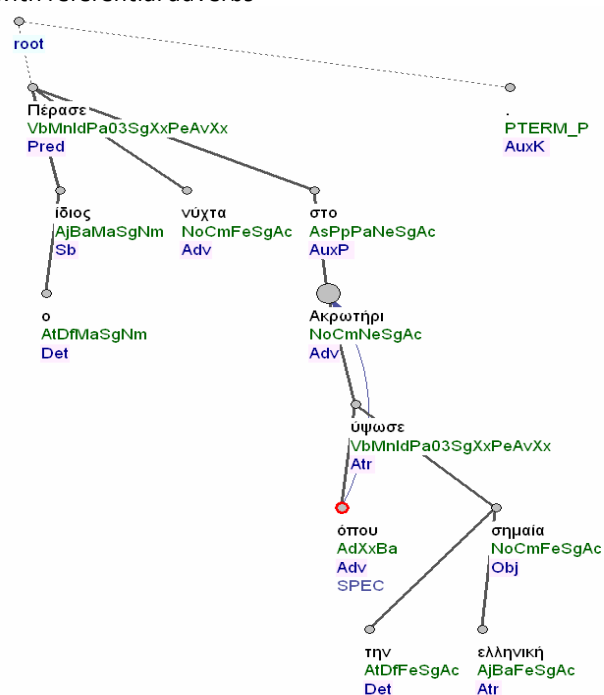


### 3.6 Referential adverbs

A linguistic expression can corefer with an adverb, which has a referential function (Figure 8).

Figure 8: Coreference with referential adverbs

Πέρασε ο ίδιος νύχτα στο **Ακρωτήρι** όπου ύψωσε την ελληνική σημαία.

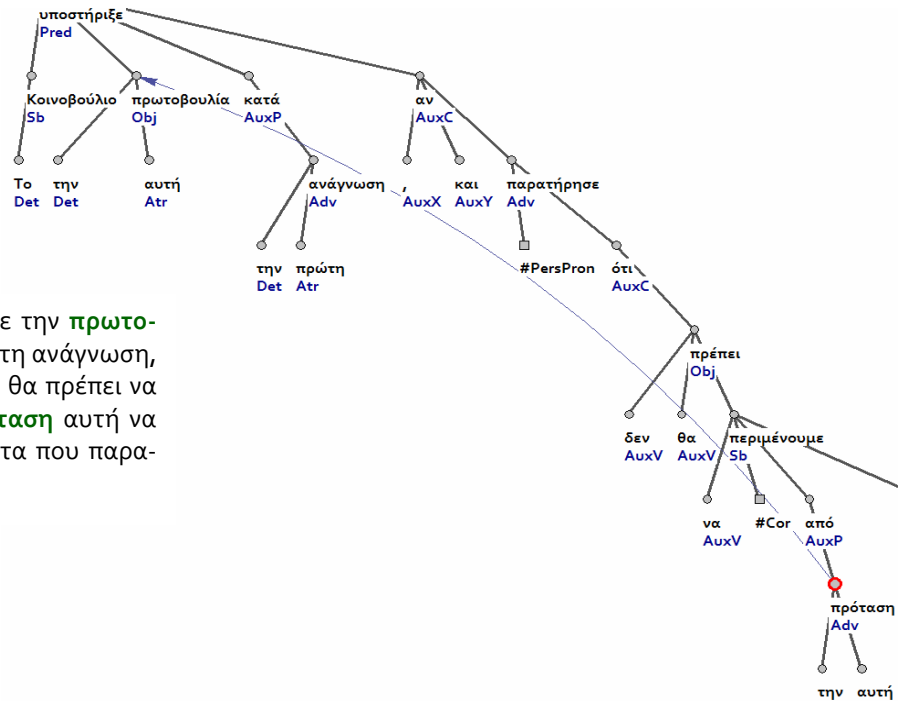


### 3.7 Discourse coreferents

The term applies to those linguistic expressions within a given text which have different meanings yet share the same referent (Figure 9).

Figure 9: Coreference with discourse coreferents

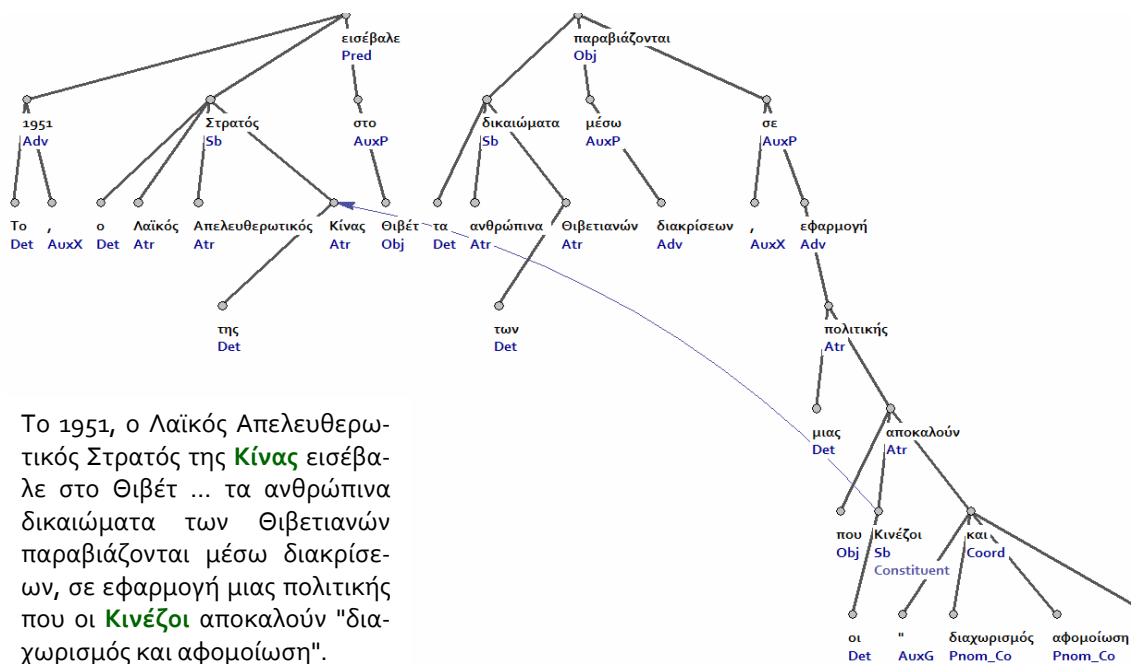
Το Κοινοβούλιο υποστήριξε την **πρωτοβουλία** αυτή κατά την πρώτη ανάγνωση, αν και παρατήρησε ότι δεν θα πρέπει να περιμένουμε από την **πρόταση** αυτή να επιλύσει όλα τα προβλήματα που παρατηρούνται σήμερα.



### 3.8 Metonymy

This is a mechanism covering a broad range of cases exemplified in Section 4 (Figure 10).

Figure 10: Coreference via metonymy



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".

## 4. METONYMY AS A MECHANISM OF COREFERENCE

The current section presents those cases of coreference, which emerge through the use of **metonymy**, a "replacement" mechanism<sup>3</sup>, in the sense that a different expression than the anticipated one appears within a given context.

Consider the following example where the expression "*πρωθυπουργού*" corefers with the locative expression "*Μαξίμου*", which metonymically refers to the Prime minister.

EXAMPLE: Η συνάντηση του **πρωθυπουργού** με τους αντάρτες βουλευτές έγινε νωρίς ... το **Μαξίμου** έγινε αποδέκτης οργανωμένων αντιδράσεων από μερίδα βουλευτών.

Metonymy is a mechanism, by which a linguistic expression, which typically refers to an entity A, can be used for referring to a different entity B. This is feasible because both entities are closely related or *adjacent*. Therefore, reference is made not to entity A, which is linguistically present, but to a different one, entity B, which is "self-evidently" revoked by the first entity due to their relation (Βελούδης, 2005b).



**NOTE** that sometimes one could waver over coreference and bridging when faced with a metonymic relation. The parameter of context is decisive for choosing between the two relations. Consider the following pair of sentences. In the first one the expression "*κεραμίδι*" metonymically corefers with "*σπίτι*", but in the second one it designates a house part, thus building a bridging relation with "*σπίτι*".

### EXAMPLES

*Coreference*: Τόσα χρόνια πάσχιζε να βάλει ένα **κεραμίδι** πάνω από το κεφάλι του και τώρα η Εφορία θα του πάρει το **σπίτι**.

*Bridging [PART-WHOLE]*: Είχε τοποθετήσει κάθε **κεραμίδι** του **σπιτιού** με τα ίδια του τα χέρια.

The following subsections describe a series of metonymic relations, being listed in Table 2, which are representative and occur frequently. It should be kept in mind though that a metonymic relation may in principle emerge between any closely related expressions.

Table 2: Metonymy relations

Metonymy relations				
Within-sets	One-Many		Container-Content	
	Set-member		Instrument	Instrument-Act
Abstract-Concrete				Instrument-Agent
NE-Creation		Institution	Seat-Institution	
Constituent	Part-Whole		Institution-Person	
	Material-Object	Aspect		

<sup>3</sup> The term covers the so-called figures of speech such as synecdoche, metonymy or antonomasia, which are traditionally defined as follows:

*Synecdoche*: figure of speech by which a word is used with its narrower or broader sense.

*Metonymy*: figure of speech by which a sense is expressed via a word with a different, yet closely related, sense.

*Antonomasia*: figure of speech by which a common or proper name is replaced by a synonym or equivalent word.



## 4.1 Within-sets

The specific type is distinguished into two subtypes.

**One-Many:** One member of a set is mentioned with a reference to the whole set. In the following context the expression "Τούρκος", which designates the whole set of Turks and not just a member of this set, corefers with the expression "Τούρκοι".

EXAMPLE: 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:** In this case an expression designating a set is used for referring to the contained members or vice versa.

So in the first example the second instance of the expression "πόλη", the set-designating expression, is coreferent with the member-designating expression, namely "πολίτες".

In the second example the member-designating expression "Κινέζοι" points to the set, being thus coreferent with the set-designating expression "Κίνα".

EXAMPLE: Ομάδες του Ερυθρού Σταυρού στη Συρία παρέδωσαν την Κυριακή προμήθειες σε περιοχές κοντά στην γειτονιά Μπάμπα Αμπ της πόλης Χομς. ... Οι ανησυχίες εντείνονται για τους **πολίτες** που έμειναν αβοήθητοι ... Ακτιβιστές αναφέρουν ότι ολόκληρη η **πόλη** της Χομς είναι αντιμέτωπη με μια ανθρωπιστική καταστροφή.

EXAMPLE: 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".

## 4.2 Abstract-Concrete

An expression with an abstract reference is employed in the place of the corresponding concrete one.

EXAMPLE: 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.

## 4.3 NE-Creation

A named entity is used for referring to a work created by or related to that entity.

In the first example the NE "Ελύτη" does not refer to the poet but rather to his produced poetry; therefore it is coreferent with the expression "ποίηση".

In a similar vein, the expression "Αγίου Πέτρου" in the second example corefers with the expression "Βασιλικής", as both refer to the Vatican church.

EXAMPLES

Πάντα διαβάζω **Ελύτη** κι ας μην καταλαβαίνω την **ποίηση** του.

... την ξύλινη μακέτα της **Βασιλικής** του Αγίου Πέτρου της Ρώμης .. είχε κατασκευάσει ο Αντόνιο ντα Σανγκάλο ο νεότερος ως Πρώτος Αρχιτέκτονας του Βατικανού ... Από την έκθεση φυσικά δεν θα λείπει και η μακέτα του τρούλου του Αγίου **Πέτρου** που κατασκεύασε ο Μιχαήλ Άγγελος ...

## 4.4 Constituent

The specific type is distinguished into two subtypes.

**Part-Whole:** A constituent part of an entity is mentioned, while reference is made to the given entity as a whole. Consider the following context where the expressions "σανίδι" and "σκηνή" both refer to the theatre stage (and do **not** build a bridging (part-whole) relation).

EXAMPLE: 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:** In this case reference is made to an object through the use of an expression that mentions the basic constituent material of that object. So, in the following context the expressions "οινόπνευμα" and "ποτά", both referring to alcoholic drinks, are coreferent.

EXAMPLE: Κατέληξε να πίνει **ποτά** όλο το βράδυ, αν και είχε υποσχεθεί στον εαυτό του να μην ξαναβάλει στο στόμα του **οινόπνευμα**.

## 4.5 Container-Content

In this case an expression designating a container is used for referring to the corresponding content.

So the expression "κατσαρόλα", the container-designating expression, is coreferent with the content-designating expression, namely "φακές" in the first example. The same holds for the expressions "καλάθι" and "ψώνια".

EXAMPLES

Καθώς του αρέσουν πάρα πολύ οι **φακές**, έφαγε όλη την **κατσαρόλα**.

Ακριβότερο φέτος το **καλάθι** της νοικοκυράς<sup>4</sup>... Τα καθημερινά **ψώνια** είναι πλέον εφιάλτης.

## 4.6 Instrument

The specific type includes two subtypes.

**Instrument-Act:** An act is referred to via the corresponding instrument. In the example that follows coreference is established between the expressions "σιδέρωμα" and "σίδερο", since both refer to ironing.

EXAMPLE: Το **σίδερο** το βαριέμαι, έτσι πολλά ρούχα μένουν χωρίς **σιδέρωμα**.

**Instrument-Agent:** The agent of an act is referred to via the corresponding instrument. In the following example the instrument-designating expression "βιολιά" in essence refers to the violinists, being thus coreferent to the agent-designating expression "βιολιστές".

EXAMPLE: Κατά τη διάρκεια μιας ως συνήθως μεθυστικής συναυλίας της Καμεράτα ... 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.

<sup>4</sup> The first part of the example is taken from Βελούδης (2005b).

## 4.7 Institution

The specific type includes two subtypes.

**Seat-Institution:** An institution's seat can be used for referring to the institution itself. In the following example the institution-referring expression "πρωθυπουργού" corefers with the locative expression "Μαξίμου", which metonymically refers to the Prime minister.

EXAMPLE: Η συνάντηση του **πρωθυπουργού** με τους αντάρτες βουλευτές έγινε νωρίς το πρωί ... το **Μαξίμου** έγινε αποδέκτης οργισμένων αντιδράσεων από μερίδα βουλευτών.

**Institution-Person:** An institution-designating expression is used with a reference to the person(s) instantiating the given institution. Consider the following example, where the expression "υπουργείο Παιδείας" essentially refers to the education minister.

EXAMPLE: ... Ο κ. **Αρβανιτόπουλος** σημείωσε ότι υπάρχουν «δεσμεύσεις του παρελθόντος» ... Σε χθεσινή ανακοίνωσή της η ΟΛΜΕ προειδοποίησε το **υπουργείο Παιδείας** πως δεν θα αποδεχθεί τις απολύσεις εκπαιδευτικών ...

## 4.8 Aspect

An entity-denoting expression is coreferent with an expression highlighting an aspect of that entity such as role, property, kinship, origin etc. (cf. the following expression pairs).

EXAMPLES

*Entity-property:* Αγαπώ το **Ρίντλεϊ 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».

*Entity-role:* Ο Έλληνας πρωθυπουργός δεν έχει πολλά να περιμένει από την συνάντηση του με την Γερμανίδα **Καγκελάριο**. Η απάντηση της **Μέρκελ** είναι προβλέψιμη: θα επαινέσει τις προσπάθειες των Ελλήνων και θα τονίσει ότι η ομοσπονδιακή κυβέρνηση αναγνωρίζει τις μεγάλες τους θυσίες.

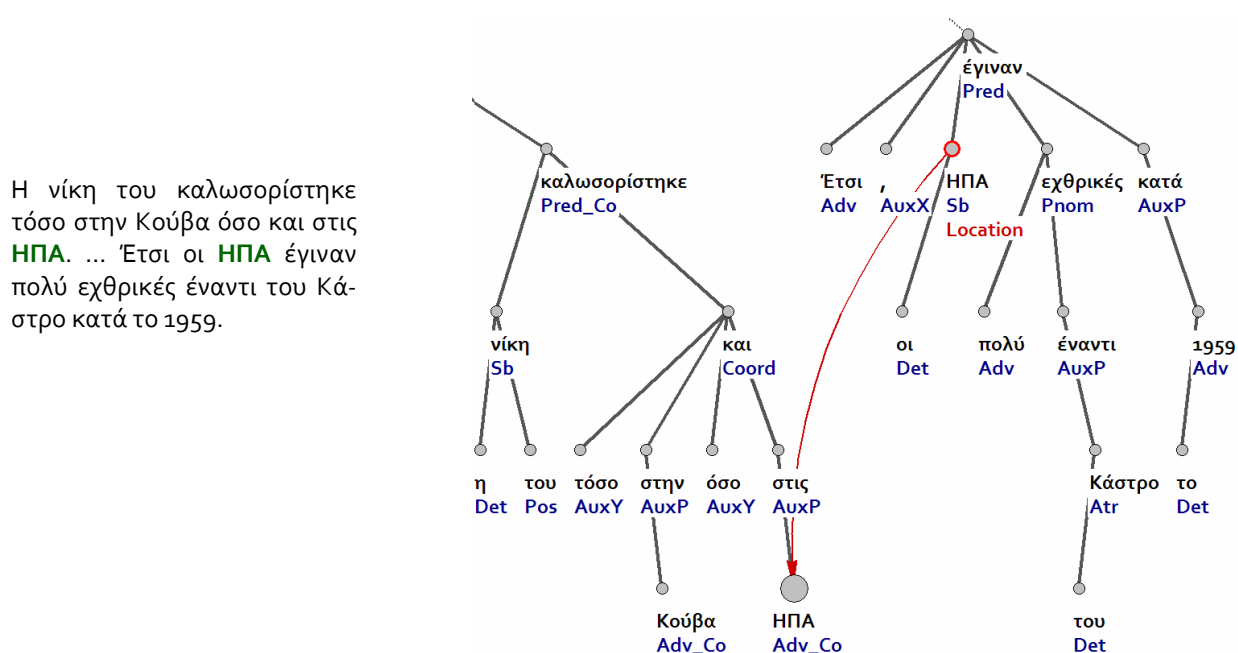
*Entity-origin:* Η Μπενφίκα νίκησε την **Ζενίτ** Αγίας Πετρούπολης με σκορ 2:0. ... Στο δεύτερο ημίχρονο, οι **Ρώσοι** προσπάθησαν να ισοφάρισουν.

## 5. NEAR IDENTITY

The current section brings together exceptional yet frequent cases where two different expressions or instances of the same expression seem to be coreferent as they implicitly refer to the same entity or represent an aspect of a given entity. These expressions are construed as **nearly identical** or **quasi coreferent**. Recasens et al. (2010: 151) point out that nearly identical expressions "are partially the same in that they share most of the important characteristics, but differ in at least one crucial dimension".

Consider the following example where the two instances of the expression "ΗΠΑ" are quasi coreferent, since the first one has a locative reading whereas the second one refers to an aspect of this location, namely the US government.

Figure 11: Example of a near identity relation



Four different types of near identity are distinguished, which are listed in Table 3 and exemplified below.

Table 3: Near identity types

Near identity types
Location
Organisation
Instantiation
Similar sets
OTHER

## 5.1 Location

A location name can be employed for referring to associated aspects of a place such as the population, the administration etc.

So, the three instances of the expression "Αφγανιστάν" are quasi coreferent, since the first mention refers to the government, the second one to the population and the third one to the country.

EXAMPLE: Ηνωμένα Έθνη: 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.

## 5.2 Organisation

An organisation name can be employed for referring not only to an organisation itself but also to related aspects such as the employees, the administration, the facilities, the products etc. Consider the following example of near identity, where the first mention of the expression "Google" refers to the company name, the second one to the company itself and the third one to the search engine.

EXAMPLE: 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**.

## 5.3 Instantiation

Different representations (cf. the first example) or realisations (cf. the second example) of the same entity are construed as nearly identical mentions. The first example is about different representations of the divine entity Zeus; the second example concerns various realisations (printed version or song lyrics) of a poetry collection.

EXAMPLES

*Entity-representation:* Δύο αγάλματα του **Δία** είναι τα πιο γνωστά από την αρχαιότητα, ο **Δίας** του Αρτεμισίου και ο χρυσελεφάντινος **Δίας** στην Ολυμπία.

*Entity-realisation:* Έχω διαβάσει το **Μονόγραμμα**, αλλά δεν **το** έχω ακούσει.

## 5.4 Similar sets

A special case of near identity, yet instantiated through the use of different expressions, is that involving expressions, which designate members of nearly overlapping or conceptually close sets.

In the first example the green-coloured expressions refer to sets, the members of which are not necessarily the same. Yet the expressions, cumulatively representing one of the warring parties, are treated as nearly coreferent.

In the second example the possessive clitic refers to the polish population, as a representative of which the polish Prime minister makes a statement concerning the railway accident. Yet within the given context the population could equal the country itself; therefore the country mentions and the clitic are construed as nearly identical expressions.

## EXAMPLES

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».

Σύγκρουση τρένων στην **Πολωνία**. ... Είναι ένα από τα χειρότερα σιδηροδρομικά δυστυχήματα των τελευταίων 20 ετών. ... Ο πρωθυπουργός της **Πολωνίας** Ντόναλντ Τουσκ δήλωσε «Είναι η πιο τραγική σιδηροδρομική καταστροφή στην ιστορία **μας** εδώ και πάρα πολλά χρόνια».

## 5.5 OTHER

Nearly identical expressions, which cannot be subsumed under the aforementioned cases, fall within this category. These types may form a distinct category in the future depending on the annotation data. Consider the following example:

## EXAMPLE

... 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 ...

## 6. BRIDGING

Bridging denotes an associative relation between two linguistic expressions which are non-coreferential, yet they can be linked to each other. Consequently a two-member chain is created either within the sentence boundaries or across sentences.

The available types of bridging relations are listed in Table 4 and exemplified below. Annotation conventions related to bridging can be found under subsection 8.1.

**Table 4:** List of bridging relation types

Bridging type	Value in Tred
Contrast	CONTRAST
Subset – Set	SUB_SET, SET_SUB
Part – Whole	PART_WHOLE, WHOLE_PART
Possession – Owner	POSS_O, O_POSS
Genitive subject – predicate	P_SUBJGEN, SUBJGEN_P
Genitive object – predicate	P_OBJGEN, OBJGEN_P
Entity & Property	ENTITY_PROP, PROP_ENTITY
Function – Object	FUNCT_P, P_FUNCT
Rest	REST

### 6.1 Contrast

Linguistic expressions which stand under a contrastive relation fall under this category.

EXAMPLE: **Άλλοτε** η περιοχή έσφυζε από ζωή, αφού η μεσαία τάξη κατέκλυζε την αγορά. **Σήμερα**, μία στις δύο επιχειρήσεις έχουν κλείσει.

Available values in TrEd: "CONTRAST"

### 6.2 Subset – Set & Set – Subset

This relation holds between linguistic expressions that denote a set and its subsets.

**Subset–Set:** Σε όλες σχεδόν τις κατοικημένες **περιοχές** του **κέντρου** υπάρχει η ίδια εικόνα ερήμωσης. Η Κυψέλη, για παράδειγμα, μέχρι τη δεκαετία του '80 κατοικούνταν από άτομα της μεσαίας τάξης.

**Set–Subset:** Σε όλες σχεδόν τις κατοικημένες περιοχές του **κέντρου** υπάρχει η ίδια εικόνα ερήμωσης. Η **Κυψέλη**, για παράδειγμα, μέχρι τη δεκαετία του '80 κατοικούνταν από άτομα της μεσαίας τάξης.

Available values in TrEd: "SUB\_SET", "SET\_SUB"

### 6.3 Part – Whole & Whole–Part

This bridging relation type designates linguistic expressions that denote a whole and its parts.

**Part–Whole:** Τα **πόδια** της **καρέκλας** ήταν βαμμένα κόκκινα.

**Whole–Part:** Της **καρέκλας** τα **πόδια** ήταν βαμμένα κόκκινα.

Available values in TrEd: "PART\_WHOLE", "WHOLE\_PART"

### 6.4 Possession – Owner & Owner – Possession

Linguistic expressions denoting the owner and the object possessed fall under this category.

**Possession–Owner:** Των **καταστημάτων** οι **ιδιοκτήτες** αρνούνται να ανοίξουν τις Κυριακές.

**Owner–Possession:** Οι **ιδιοκτήτες** των **καταστημάτων** αρνούνται να ανοίξουν τις Κυριακές.

Available values in TrEd: "POSS\_O", "O\_POSS"

### 6.5 Genitive subject in a complex NP construction

This relation holds between linguistic expressions, which form a complex NP. One of the expressions represents a predicate and the other one, which is in genitive case, functions as the subject of the predicate.

**Predicate–Genitive subject:** η **άμυνα** των **μασκοφόρων** [= αμύνονται οι масοφόροι]

**Genitive subject–Predicate:** Των **μασκοφόρων** η **άμυνα** [= οι масοφόροι αμύνονται]

Available values in TrEd: "P\_SUBJGEN", "SUBJGEN\_P"

### 6.6 Genitive object in a complex NP construction

This relation holds between linguistic expressions, which form a complex NP. One of the expressions represents a predicate and the other one, which is in genitive case, functions as the object of the predicate.

**Predicate–Genitive object:** Η **καταστροφή** της **πόλης** [= καταστρέφουν την πόλη]

**Genitive object–Predicate:** Της **πόλης** η **καταστροφή** [=την πόλη καταστρέφουν]

Available values in TrEd: "P\_OBJGEN", "OBJGEN\_P"

### 6.7 Entity & Property

This relation holds between linguistic expressions denoting an entity and a related property.

**Entity–Property:** Είναι άνθρωπος της **δράσης**.

**Property–Entity:** Είναι της **δράσης** άνθρωπος.

Available values in TrEd: "ENTITY\_PROP", "PROP\_ENTITY"



## 6.8 Function – Object & Object – Function

This relation holds between linguistic expressions denoting an object and the related function or role.

**Function–Object:** Όταν έφυγε ο **δάσκαλος**, η **τάξη** ξέσπασε σε γέλια.

**Object–Function:** Η **τάξη** ξέσπασε σε γέλια, όταν έφυγε ο **δάσκαλος**.

Available values in TrEd: "FUNCT\_P", "P\_FUNCT"

## 6.9 REST

Any type of associative relation between linguistic expressions that cannot be subsumed under the aforementioned types falls within this category. These types may form a distinct category in the future depending on the annotation data.

Available values in TrEd: "REST"

Indicative examples of such relation types are the following:

- **Location**

EXAMPLE: Ο Βλαντιμίρ Πούτιν κέρδισε τις προεδρικές **εκλογές** της **Ρωσίας**.

- **Time**

EXAMPLE: Αυτό είναι το πιο δυνατό καταγεγραμμένο **ξέσπασμα** σιφώνων του μηνός **Μαρτίου**.

- **Generic-Specific**

EXAMPLE: Η έκκληση της υπηρεσίας των Ηνωμένων Εθνών πραγματοποιήθηκε ενόψει της σημερινής Παγκόσμιας Ημέρας για την **Γυναίκα** και μια ημέρα μετά την στήριξη του Προέδρου Καρζάι για τον «κώδικα διαγωγής» που εκδόθηκε από ομάδα κληρικών. Ο κώδικας που εξέδωσε το Συμβούλιο Ουλεμά επιτρέπει στους συζύγους να ξυλοκοπούν τις **γυναίκες** τους...

- **Thematic or situational association**

EXAMPLES

Η **ομηρία** των εργατών έληξε μετά από **διαπραγματεύσεις**.

Στις 23 και 24 Μαΐου οι προεδρικές **εκλογές** στην Αίγυπτο ... Τα τελικά **αποτελέσματα** θα ανακοινωθούν στις 21 Ιουνίου.

Νωρίτερα, ο Κρίστιαν Βουλφ έδωσε την συγκατάθεση του για την **έρευνα** και δεν χρειάστηκε ειδικό **ένταλμα**.

... λόγω **σκανδάλου** για παθητική **δωροδοκία**

## 7. LINGUISTIC EXPRESSIONS

The current section specifies which linguistic expressions have a referring function and therefore should be marked as mentions participating in a coreference, near identity or bridging relation.

### 7.1 Nominal & pronominal heads

Nominal & pronominal heads may be annotated with one of the following PoS tags:

- NoCm, NoPr [common or proper noun]

EXAMPLE: οι **κρατούμενοι**, η **Μόσχα**

- Rg, ABBR, NBABBR, DIG, DATE [acronym, foreign word, abbreviation, digit]

EXAMPLE: η **ΔΕΗ**, οι **Beatles**, τα **5**, το **1972**

- Aj, VbMnPp, Nm [adjective, passive participle, numeral]

EXAMPLE: τους **πλούσιους**, οι **εργαζόμενοι**, ο **τρίτος**

- Ad [adverb]<sup>5</sup>

EXAMPLE: οι **κάτω**

- Pn [pronoun]

EXAMPLE: **εμένα**, **που**, των **οποίων**, **αυτές**, **τον** είδα, στη χώρα **τους**

### 7.2 Adverbial heads

Adverbial heads with a referring function are also eligible linguistic expressions. Their PoS tag is **Ad**.

EXAMPLES

Η ζωή στην **Αθήνα** είναι αγχωτική. Όταν πήγα **εκεί**, δυσκολεύτηκα να ακολουθήσω τους ρυθμούς ζωής.

Οι περισσότεροι κάτοικοι έχουν διαφύγει σε γειτονικές **περιοχές όπου** τους παρέχονται τρόφιμα.

### 7.3 Null subjects

Null subjects are marked **even if they do not participate in a linking relation**. In this case, insert a **new subject node** (cf. section 9) with the value "**PersPron**" to represent a finite clause subject or the value "**Cor**" that designates the null subject in a control structure.

EXAMPLES

*Null subject in a finite clause:* Έντονοι νοτιάδες θα πνέουν στη χώρα ενώ **PersPron** θα είναι ελαφρώς πιο ενισχυμένοι στο Ιόνιο.

*Null subject in a control structure:* Η **εταιρεία** σκοπεύει να **Cor** προβεί σε μαζικές απολύσεις.

<sup>5</sup> Adverbs in such structures assume a nominal function.

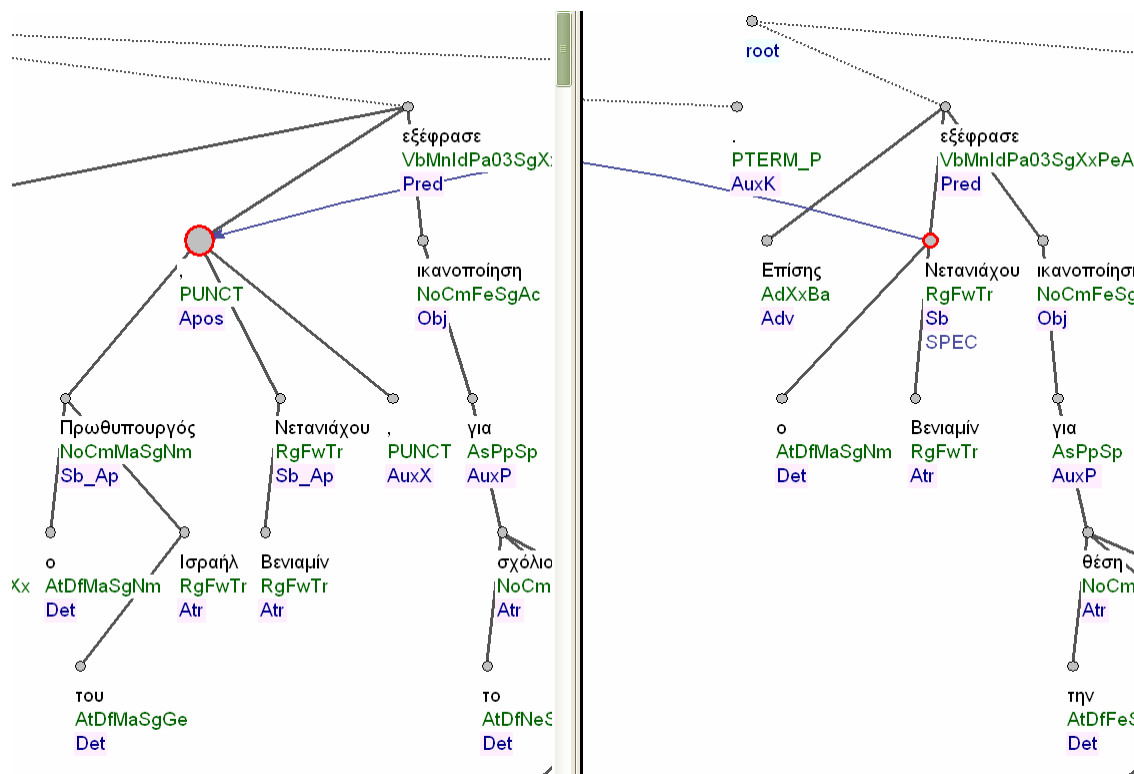
## 7.4 Appositional structures

In the case of appositional structures (with or without a comma), mark the **APOS** node (see Figure 12).

EXAMPLE

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

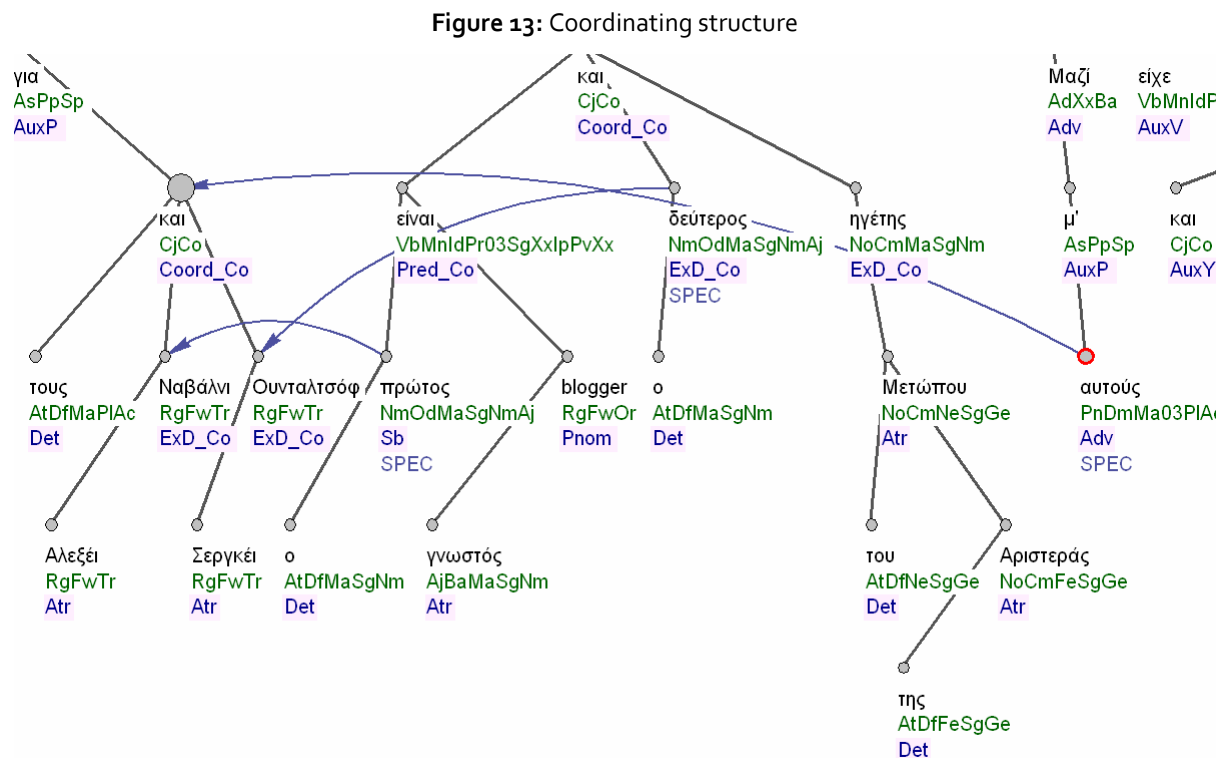
Figure 12: Appositional structure



## 7.5 Coordinating structures

In the case of coordinated elements, mark the bigger structure, i.e. the **COORD** node. The **individual heads** can also be marked if they independently participate in a linking relation (see Figure 13).

EXAMPLE: Ο λόγος για τους Αλεξί **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.



## 8. CONVENTIONS & CASE-SPECIFIC GUIDELINES

The current section is broadly divided into two (2) parts: the 1<sup>st</sup> part cites the conventions that have been decided upon, conventions that are related to technical aspects of the annotation process, i.e. to the way coreference, near identity & bridging should be marked. The 2<sup>nd</sup> part provides general and case-specific guidelines.

## 8.1 Annotation conventions: Coreference

- In coreference relations involving a **pronominal element** (either null or instantiated) and another referring expression, the source is always the pronominal element. The target is always the other referring expression, i.e. the **antecedent**, if it precedes the source (anaphora), or the **postcedent**, if it follows the source (cataphora). So the coreferential chain formed in each case is as follows (based on the surface order of words):

Anaphora chain: <antecedent, source>

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

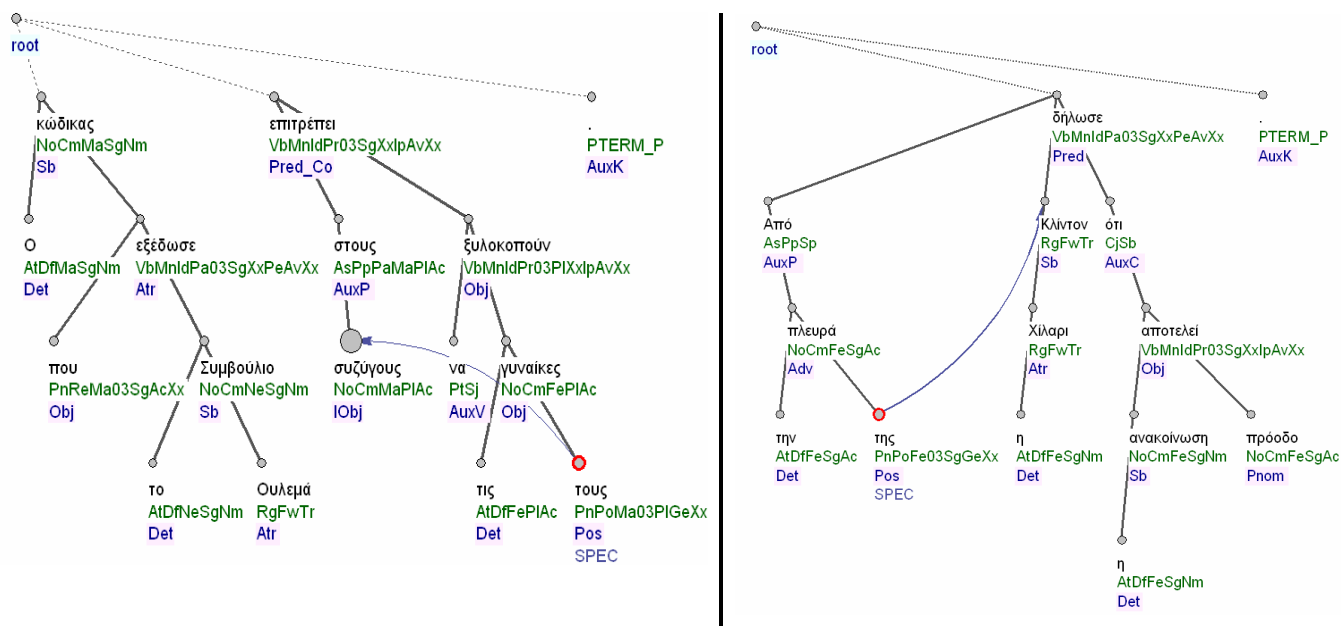
Anaphora chain in the example: <συζύγους, τους>

Cataphora chain: <source, postcedent>

EXAMPLE: Από την πλευρά της, η Χίλαρι Κλίντον δήλωσε σε αρμόδια επιτροπή του Κογκρέσου ότι «η ανακοίνωση αποτελεί πρόοδο».

Cataphora chain in the example: <της, Κλίντον>

**Figure 14:** Anaphora & Cataphora chains with a pronominal element



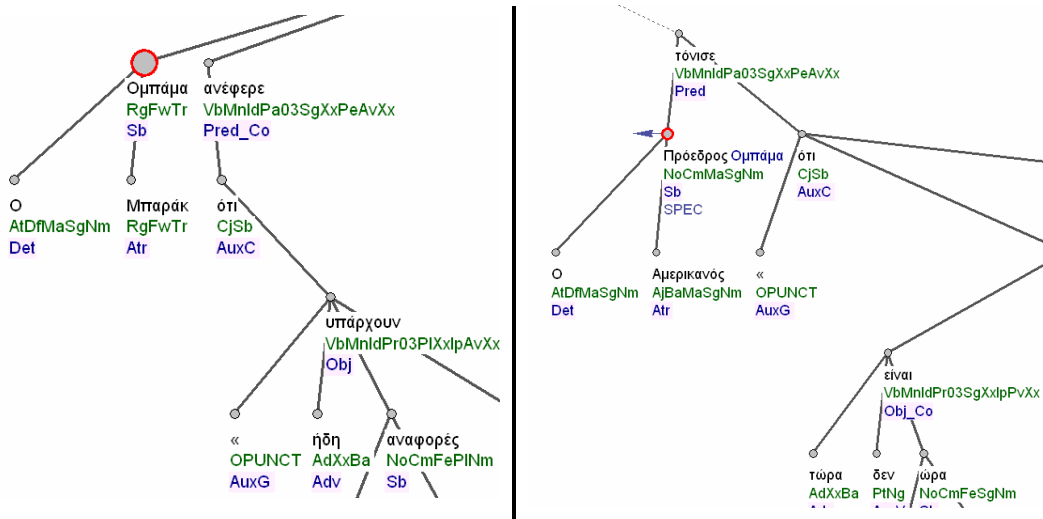
- In all other cases the coreferential arrows have a **leftward** direction, i.e. the source is always to the right and the target always to the left; thus building a coreferential chain with the form <antecedent, source> (based on the surface order of words).

**Anaphora chain:** <antecedent, source>

EXAMPLE: Ο Μπαράκ **Ομπάμα** ανέφερε ότι «ήδη υπάρχουν πολλές αναφορές για πόλεμο με το Ιράν». Ο Αμερικανός **Πρόεδρος** τόνισε «ότι τώρα δεν είναι η ώρα για θεαματικές κινήσεις».

Anaphora chain in the example: <**Ομπάμα**, **Πρόεδρος**>

Figure 15: Anaphora chain with a non-pronominal element



## 8.2 Annotation conventions: Bridging

- Bridging relation arrows always have a **rightward** direction, i.e. the source is always the first of the two expressions in the order that they appear in the sentence or in the text and the target is always the second of the two expressions. The chain created in each case always has the form <source, target>.

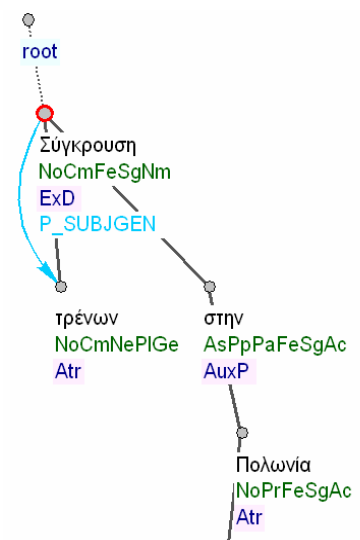
**Bridging chain:** <source, target>

EXAMPLE: **Σύγκρουση** **τρένων** στη Νότια Πολωνία

Bridging chain in the example: <**σύγκρουση**, **τρένων**>

Bridging relation type in the example: "P\_SUBJGEN"

Figure 16: Bridging chain



- Some types of bridging relations are reverses of other bridging relations. One should follow the linear order of words for selecting the appropriate type; thus the chain formed in each case always has the form <source, target>.

EXAMPLE: Κύμα κακοκαιρίας πλήττει τα **Βαλκάνια**, ενώ μέχρι στιγμής τρεις άνθρωποι έχουν χάσει τη ζωή τους και δεκάδες έχουν αποκλειστεί στη **Σερβία**.

Bridging chain: <**Βαλκάνια**, **Σερβία**>

Bridging relation type: "SET\_SUB"

- When two linguistic expressions build a bridging relation but they are repeated within the text, one should mark **only one** bridging relation between the closest instances of the expressions participating in the relation. In the following example only the green-coloured instances should be annotated.

EXAMPLE

*Τρεις νεκροί από τη σφοδρή κακοκαιρία που πλήττει τα Βαλκάνια.*

*Κύμα κακοκαιρίας πλήττει τα **Βαλκάνια**, ενώ μέχρι στιγμής τρεις άνθρωποι έχουν χάσει τη ζωή τους και δεκάδες έχουν αποκλειστεί στη **Σερβία**.*

*Στην ανατολική πόλη Ποζαρεβάκ ένα ηλικιωμένο ζευγάρι πάγωσε από το κρύο, καθώς ο υδράργυρος έπεσε στους - 15 βαθμούς Κελσίου.*

*Την ίδια ώρα, δεκάδες χωριά της Σερβίας έχουν αποκλειστεί από την κακοκαιρία και την έντονη χιονόπτωση, ενώ προβλήματα έχουν προκληθεί και σε κεντρικές οδικές αρτηρίες.*

- The same constraint holds when a coreferent linguistic expression appears within the text. So in the following example only the green-coloured expressions should be annotated. The instance "πρώην Προέδρου της Γερμανίας" should be ignored since it is coreferent to "Κρίστιαν Βουλφ".

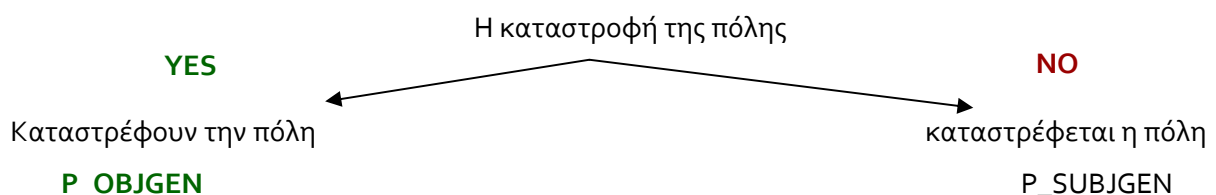
EXAMPLE

*Η Εισαγγελία της Γερμανίας ερευνά την **κατοικία** του **Κρίστιαν Βουλφ**.*

*Η Εισαγγελία της Γερμανίας ξεκίνησε την έρευνά της στην κατοικία του πρώην Προέδρου της Γερμανίας.*

- In order to distinguish between a Genitive subject and a Genitive object in cases of ambiguity, one should replace the nominal predicate with an active verb (cf. the following example).

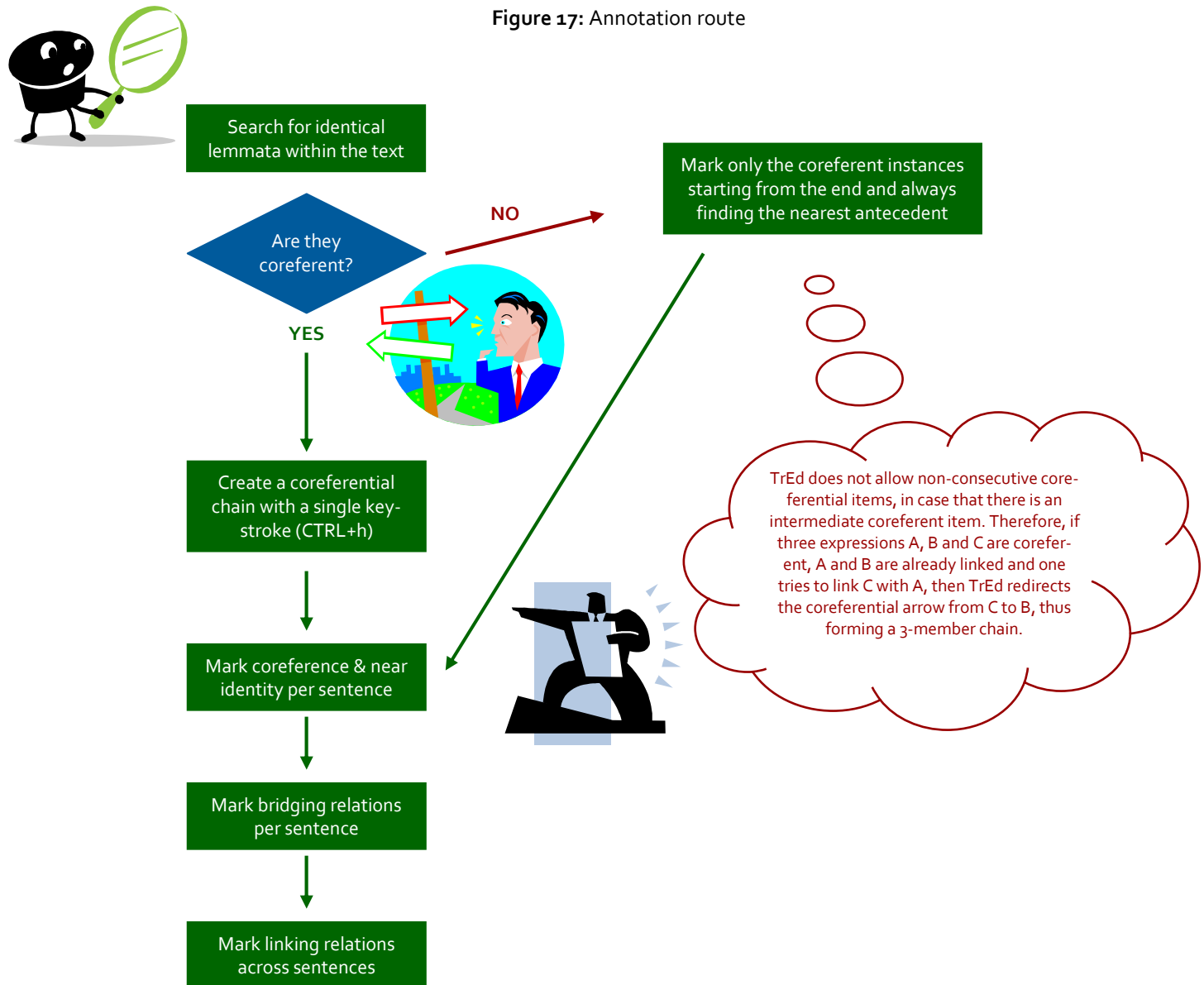
EXAMPLE



### 8.3 Annotation route

The current subsection describes a route suggested for annotators, although one is free to use their one method of annotating.

Figure 17: Annotation route



You may find it helpful & convenient to use a printout of the text while annotating!



## 8.4 Non-eligible expressions

Do **NOT** mark the following linguistic expressions:

- A **singleton**, i.e. a mention that does not participate in a linking relation.

- A pronominal element that refers to a clause.

EXAMPLE: Πλήρωσα πολύ ακριβά για τα λάθη μου και **αυτό** με λύπησε ιδιαίτερα.

- A nominal head that refers to a clause or a proposition.

EXAMPLE: Στο υπουργείο Παιδείας σκέφτονται να επιστρατεύσουν τους διοικητικούς υπαλλήλους και να τους αποσταλούν φύλλα πορείας, ωστόσο στελέχη του υπουργείου φοβούνται πως μία τέτοια **κίνηση** θα έχει πολιτικό κόστος για την κυβέρνηση.

- A pronominal element that is annotated with the PoS tag *pnir* (interrogative pronoun).

EXAMPLE: δεν ξέρω **ποιος** τηλεφώνησε

- Expressions within a copula verb structure, which have an attributive function.

EXAMPLE: Τρεις ισχυρές εκρήξεις ακούστηκαν κοντά στο σπίτι του άνδρα που θεωρείται **ύποπτος** για τους θανάτους επτά ανθρώπων.

## 8.5 Case-specific guidelines

### 8.5.1 Cataphora with a non-pronominal element

Within a specific context usually involving the presence of split antecedents, non-pronominal elements build a cataphora chain (<source, postcedent>) (see Figure 18 and Figure 19).

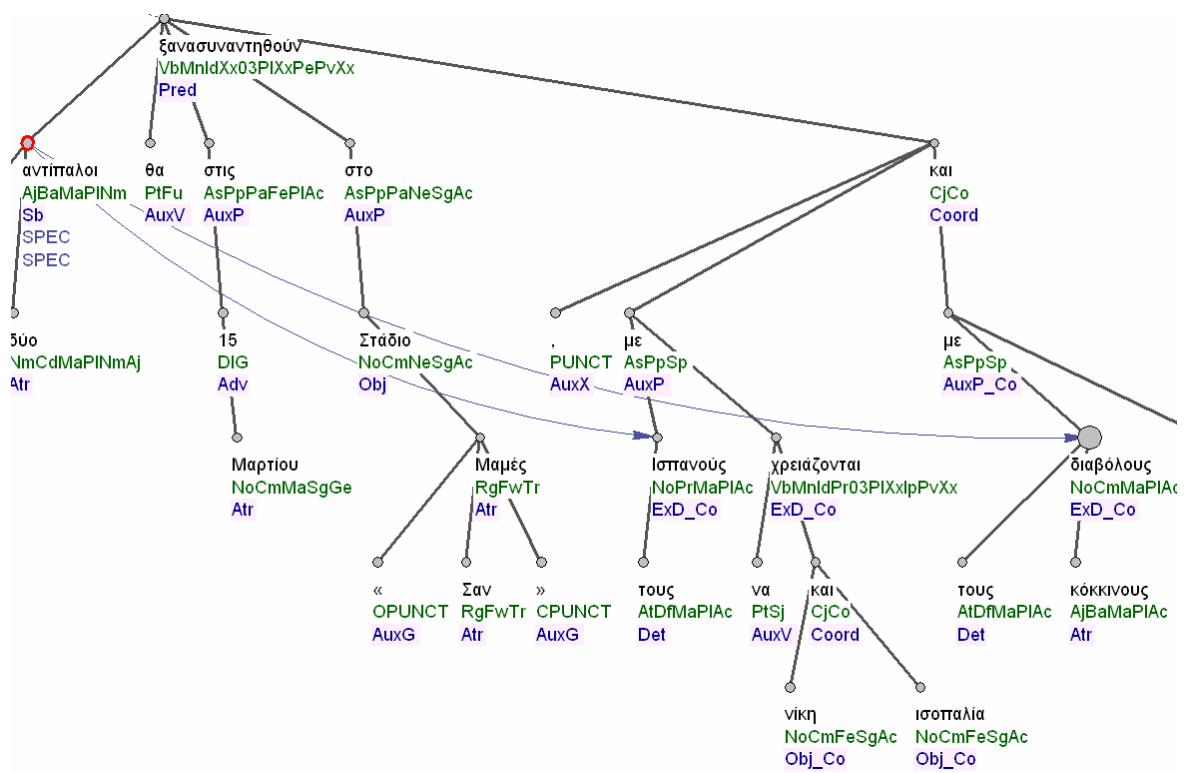
EXAMPLES

Οι δύο **αντίπαλοι** θα ξανασυναντηθούν στις 15 Μαρτίου στο Στάδιο «Σαν Μαμές», με τους **Ισπανούς** να χρειάζονται νίκη και ισοπαλία και με τους κόκκινους **διαβόλους** να χρειάζονται νίκη με διαφορά δύο τερμάτων.

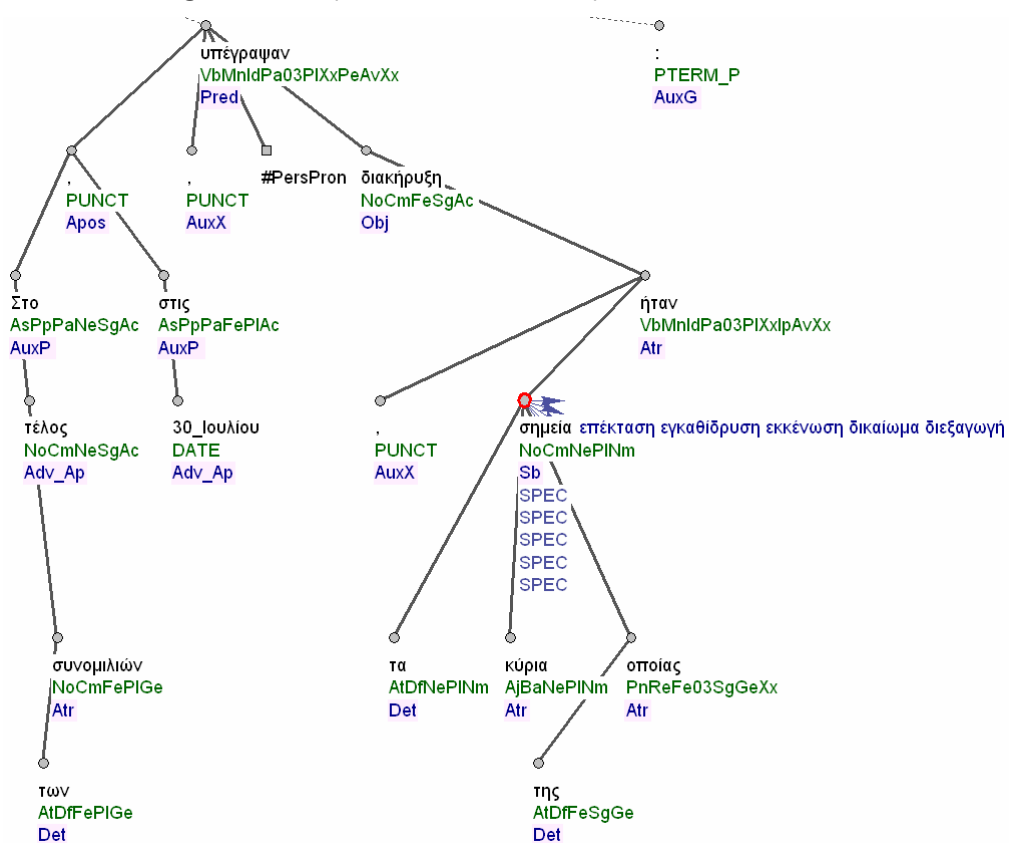
Στο τέλος των συνομιλιών, στις 30\_Ιουλίου, υπέγραψαν διακήρυξη, τα κύρια **σημεία** της οποίας ήταν:

1. Η μη **επέκταση** των περιοχών που είχαν κάτω από τον έλεγχό τους οι αντίπαλες δυνάμεις.
2. Η **εγκαθίδρυση** ζωνών ασφαλείας μεταξύ των αντιμαχόμενων.
3. Η **εκκένωση** των Τουρκικών θυλάκων από την Ειρηνευτική Δύναμη του ΟΗΕ.
4. Το **δικαίωμα** να διαθέτουν οι δύο πλευρές δική τους αστυνομία και δυνάμεις ασφαλείας.
5. Η **διεξαγωγή** διαπραγματεύσεων με συμμετοχή Ελληνοκυπρίων και Τουρκοκυπρίων για οριστική διευθέτηση του Κυπριακού.

**Figure 18:** Cataphora chain with a non-pronominal element (a)



**Figure 19:** Cataphora chain with a non-pronominal element (b)



### 8.5.2 Copula verb structures

Expressions found within a copula verb structure are only annotated when they have a **referring** function.

EXAMPLE

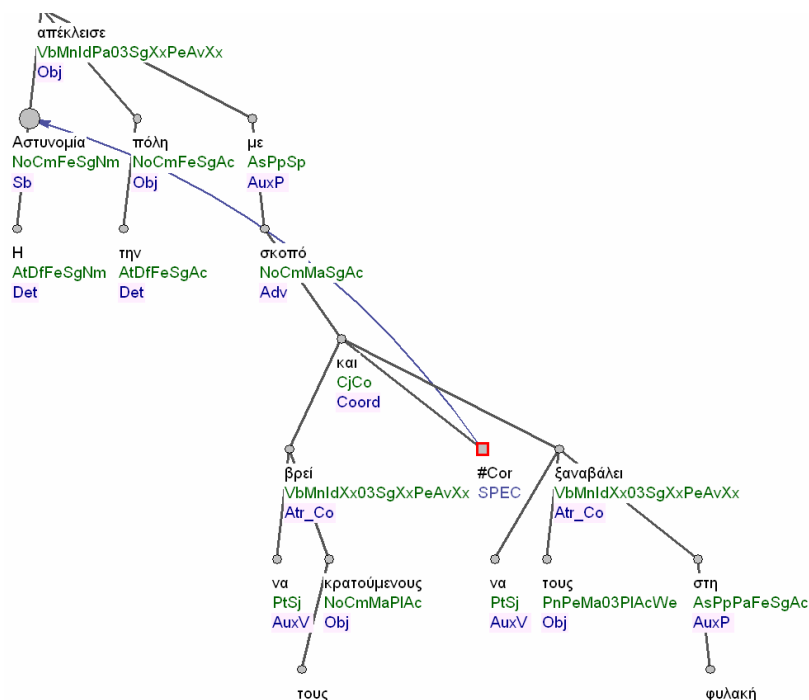
**Σύγκρουση** τρένων στην Νότια Πολωνία ... Η **σύγκρουση** έγινε στη γραμμή Βαρσοβία – Κρακοβία ... Από τη **σύγκρουση** καταστράφηκαν 3 βαγόνια. Ο πρωθυπουργός της Πολωνίας Ντόναλντ Τουσκ δήλωσε «Είναι η πιο τραγική σιδηροδρομική **καταστροφή** στην ιστορία μας εδώ και πάρα πολλά χρόνια».

### 8.5.3 Null subjects in coordinating structures

In coordinating structures missing a subject the new subject node with the appropriate value ("**PersPron**" or "**Cor**") is inserted under the **COORD** node (see Figure 20).

EXAMPLE: Η **Αστυνομία** απέκλεισε την πόλη με σκοπό να βρει τους κρατούμενους **και** να τους ξαναβάλει στη φυλακή.

Figure 20: Null subject in a coordinating structure

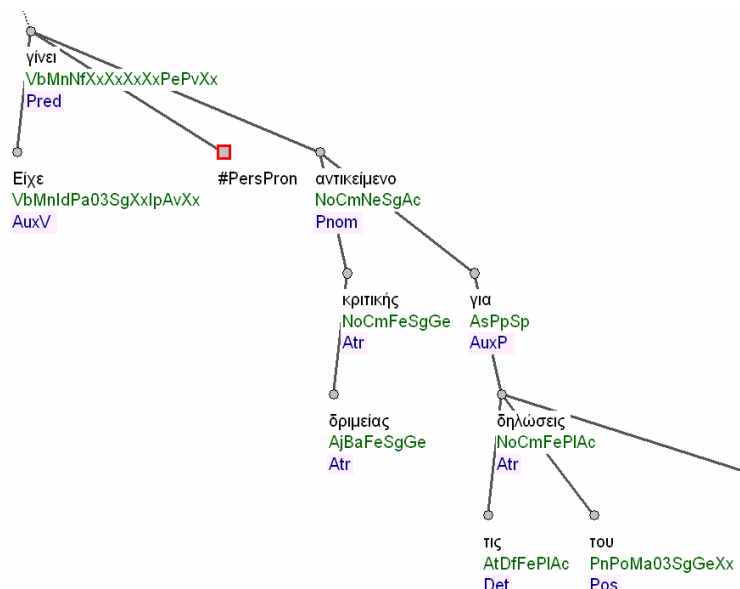


### 8.5.4 Null subjects of multi-token verbs

Null subjects of multi-token verbs are inserted under the lexical verb node (see Figure 21).

EXAMPLE: Είχε **PersPron** γίνει αντικείμενο δριμείας κριτικής για τις δηλώσεις του.

Figure 21: Null subject of a multi-token verb



### 8.5.5 Location names

In structures having the form noun + location name, the **location name** should be marked as the participant in a linking relation.

EXAMPLE

Ομάδες του Ερυθρού Σταυρού στη Συρία παρέδωσαν την Κυριακή προμήθειες σε περιοχές κοντά στην γειτονιά **Μπάμπα Αμπ** της πόλης **Χομς**, αλλά για τρίτη συνεχόμενη ημέρα τους απαγορεύτηκε η είσοδος στην **συνοικία** που ελέγχονταν μέχρι πρόσφατα από τους αντάρτες ... ολόκληρη η πόλη της **Χομς** είναι αντιμέτωπη με μια ανθρωπιστική καταστροφή ... μάρτυρες της σφαγής που συντελείται από τους σύρους στρατιώτες που καταδιώκουν και εκτελούν τους εναπομείναντες αντάρτες στην **πόλη**.

### 8.5.6 Expression-specific bridging relations

Certain expressions within a given context build specific bridging relations. A non-exhaustive list is provided in Table 5.

Table 5: List of expressions & their bridging relations

Word / word pair	Bridging relation	Example
noun + agency name	POSS_O	ομάδα του ΟΗΕ
noun + country name	POSS_O	υπουργός Εξωτερικών της Ελλάδας, εισαγγελία της Γερμανίας
αντιπρόσωπος	FUNCT_P	αντιπρόσωπος της κυβέρνησης

Word / word pair	Bridging relation	Example
αποστολή	POSS_O	αποστολή του Ερυθρού Σταυρού
αρχές	FUNCT_P	αρχές της χώρας
γραμματέας	FUNCT_P	γραμματέας του συμβουλίου
διεύθυνση	FUNCT_P	διεύθυνση του οργανισμού
διευθυντής	FUNCT_P	διευθυντής του οργανισμού
διοίκηση	FUNCT_P	διοίκηση του οργανισμού
διοικητής	FUNCT_P	διοικητής του οργανισμού
εκπρόσωπος	FUNCT_P	εκπρόσωπος της κυβέρνησης
επικεφαλής	FUNCT_P	επικεφαλής της αντιπροσωπείας
επιτροπή	POSS_O	επιτροπή του ΟΗΕ
ηγεσία + country name	FUNCT_P	ηγεσία της Αιγύπτου
ηγεσία + person name	P_SUBJGEN	ηγεσία του Μπαράκ Ομπάμα
ηγέτης	FUNCT_P	ηγέτης της Αιγύπτου
ημίχρονο	PART_WHOLE	ημίχρονο του αγώνα
ημίχρονο + αγώνας	PART_WHOLE	το 1 <sup>ο</sup> ημίχρονο του αγώνα
καθεστώς + person name	POSS_O	καθεστώς του Σαντάμ Χουσεΐν
κυβέρνηση	FUNCT_P	κυβέρνηση της χώρας
κυβερνήτης	FUNCT_P	κυβερνήτης της χώρας
λεπτό	PART_WHOLE	λεπτό του αγώνα
λεπτό + αγώνας	PART_WHOLE	στο 81 <sup>ο</sup> λεπτό του αγώνα
μέτοχος	FUNCT_P	μέτοχος της εταιρείας
ομάδα	POSS_O	ομάδα του Ερυθρού Σταυρού
περιοχή * <sup>6</sup>	PART_WHOLE	περιοχή της Ελλάδας
πηγή	SUB_SET	πηγή του υπουργείου, πηγές της αντιπολίτευσης
πόλη *	PART_WHOLE	πόλη της Ελλάδας
πολιτεία *	PART_WHOLE	πολιτεία των ΗΠΑ
πρόεδρος	FUNCT_P	πρόεδρος της χώρας, πρόεδρος της επιτροπής
πρωθυπουργός	FUNCT_P	πρωθυπουργός της Ελλάδας
πρωτεύουσα *	PART_WHOLE	πρωτεύουσα της Ελλάδας
συμβούλιο	POSS_O	συμβούλιο ασφαλείας του ΟΗΕ
υπουργός	FUNCT_P	υπουργός Εξωτερικών
υφυπουργός	FUNCT_P	υφυπουργός Εξωτερικών
φάση	PART_WHOLE	φάση του Europa League
χωριό *	PART_WHOLE	χωριό της Ελλάδας

<sup>6</sup> In cases marked with an asterisk the same bridging relation holds when the expression is name-instantiated, cf. "Αθήνα" in relation to "Ελλάδα".

## 9. ANNOTATING WITH TRED & GDT\_BRIDGING

The TrEd manual annotation tool receives as input texts represented as sequences of dependency trees. The GDT\_Bridging extension<sup>7</sup> employs a series of macros for annotating the linking relations, namely coreference, near identity and bridging. A detailed presentation of those macros can be found under subsection 9.1, whereas a FAQ section can be found under 9.2.

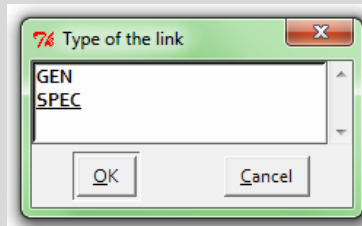
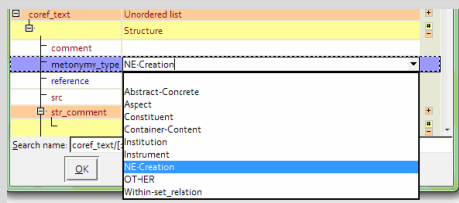
### 9.1 Using the GDT\_Bridging macros

The following tables exemplify how to use the GDT\_Bridging macros for the annotation process.



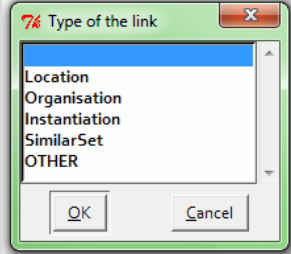
**NOTE** that the keyboard shortcuts are case-sensitive.

**Table 6:** TrEd – Annotating coreference

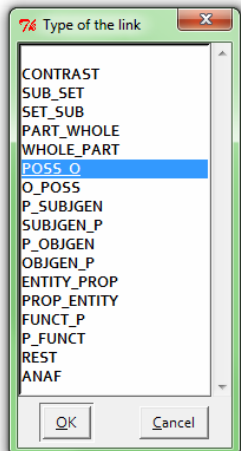
I want to ...	How to do it
<b>Add a coreference arrow</b> <i>[arrow colour: green]</i>	1) select the source node 2) press <b>CTRL</b> and hold 3) select the target node 4) a pop-up window displays two values, <b>SPEC</b> (specific reference) and <b>GEN</b> (generic reference) 5) select the appropriate value 
<b>Create a coreference chain</b>	1) select a mention and press <b>CTRL+h</b> 2) a pop-up window displays two values, <b>SPEC</b> and <b>GEN</b> 3) select the appropriate value 4) for a text with <i>n-mentions</i> , an implicit coreference chain is created by the annotation of <i>n-1</i> coreference relations
<b>Mark metonymy</b>	1) select the source node of a coreference arrow 2) press <b>ENTER</b> 3) the node tab appears 4) select the appropriate value from the drop-down list appearing under the <i>coref_text</i> element 

<sup>7</sup> Available at [http://gdt.ilsp.gr/tools/gdt\\_tred\\_extensions](http://gdt.ilsp.gr/tools/gdt_tred_extensions)

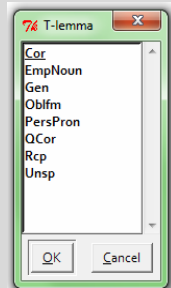
**Table 7:** TrEd – Annotating near identity

I want to ...	How to do it
Add a near identity arrow [arrow colour: red]	<ol style="list-style-type: none"> <li>1) select the source node</li> <li>2) press <b>ALT</b> and hold</li> <li>3) select the target node</li> <li>4) a pop-up window displays the near identity types</li> <li>5) select the appropriate type</li> </ol> 

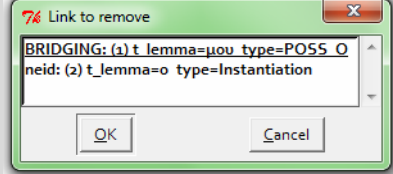
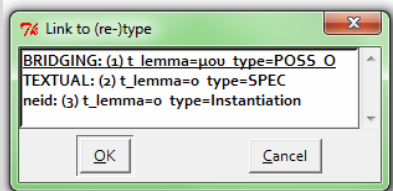
**Table 8:** TrEd – Annotating bridging

I want to ...	How to do it
Add a bridging arrow [arrow colour: light blue]	<ol style="list-style-type: none"> <li>1) select the source node</li> <li>2) press <b>SHIFT</b> and hold</li> <li>3) select the target node</li> <li>4) a pop-up window displays the bridging types</li> <li>5) select the appropriate type</li> </ol> <p><b>NOTE:</b> follow the linear order of words for selecting the relation type.</p> 

**Table 9:** TrEd – Adding null subjects

I want to ...	How to do it
Create a new null subject node	<ol style="list-style-type: none"> <li>1) select a node</li> <li>2) press <b>INSERT</b></li> <li>3) select the appropriate type for the new node</li> </ol> <p>Values to be used</p> <ul style="list-style-type: none"> <li>▪ <b>Cor:</b> null subject in a control structure</li> <li>▪ <b>PersPron:</b> null subject of a finite clause</li> </ul> 

**Table 10:** TrEd – Editing nodes

I want to ...	How to do it	
<b>Remove an arrow</b>	Select the source node and press <b>X</b>  <b>NOTE:</b> If more than one arrows start from the given node, then a pop-up window appears first listing all the relations, which start from the given node, and asking the user to select the relation to be deleted.	
<b>Change the relation type</b>	1) select the source node and press <b>r</b> 2) a pop-up window displays the available types per linking relation  <b>NOTE:</b> If more than one arrows start from the given node, then a pop-up window appears first. It lists all the relations starting from the given node and asks the user to select the relation to be retyped.	
<b>Edit the annotation of a node</b>	1) select a node 2) press <b>ENTER</b> 3) the node tab appears 4) perform the desired modifications	

**Table 11:** TrEd – Moving to & from antecedents


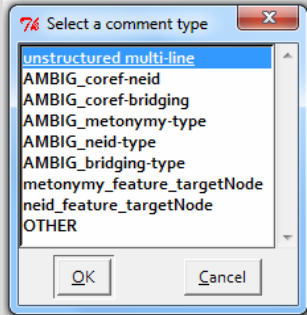
I want to ...	How to do it
<b>Jump to the antecedent (1<sup>st</sup> arrow)</b>	1) select the source node & press <b>a</b> to jump to the 1 <sup>st</sup> antecedent of a given node
<b>Jump to the antecedent (2<sup>nd</sup> arrow)</b>	1) select the source node & press <b>q</b> to jump to the 2 <sup>nd</sup> antecedent of a given node
<b>Jump back from antecedent</b>	1) select the antecedent node & press <b>z</b> to jump to the source node

**Table 12:** TrEd – Switch display

I want to ...	How to do it
<b>Switch displaying context trees</b>	Press <b>e</b> to make neighbouring trees appear and disappear
<b>Switch displaying coreference arrows</b>	Press <b>CTRL+t</b> to make coreference arrows appear and disappear
<b>Switch displaying near identity arrows</b>	Press <b>CTRL+n</b> to make near identity arrows appear and disappear
<b>Switch displaying bridging arrows</b>	Press <b>CTRL+b</b> to make bridging arrows appear and disappear



**Table 13:** TrEd – Adding comments

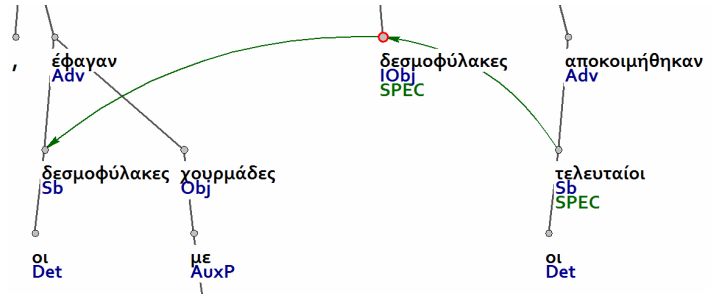
I want to ...	How to do it
<b>Add an arrow comment</b>	<p>1) select the source node and press <b>m</b></p> <p>2) a pop-up window displays the available comment types</p> <p>3) select the appropriate type and (optionally) insert a comment text</p> <p><b>NOTE:</b> If more than one arrows start from the given node, then a pop-up window appears first. It lists all the relations starting from the given node and asks the user to select the relation to be commented.</p> <p> The comment concerns an arrow, i.e. a linking relation, which starts from this node; thus it gets deleted after the deletion of the arrow.</p> 
<b>Add a node comment</b>	<p>1) select the source node and press <b>c</b></p> <p>2) a pop-up window appears for inserting a free text comment</p>
<b>Edit a node comment</b>	<p>1) select a node and press <b>ALT+c</b></p> <p>2) a pop-up window appears where the comment text may be modified</p>

## 9.2 FAQ

### 1. What happens when removing an **intermediate** node of a coreferential chain?

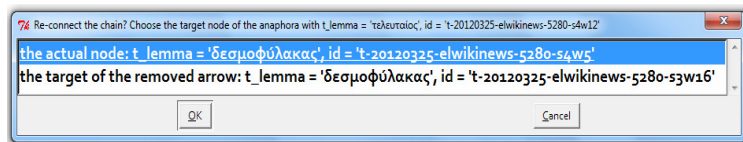
TrEd asks to reconnect the residual chain members. Consider the following example, involving a 3-member anaphora chain: <δεσμοφύλακες, δεσμοφύλακες, τελευταίοι>

Δεκαεννέα κρατούμενοι κατάφεραν να αποδράσουν από φυλακή του Κιρκούκ, στο Ιράκ, αφού οι **δεσμοφύλακες** έφαγαν χουρμάδες με υπνωτικό. Οι κρατούμενοι έδωσαν στους **δεσμοφύλακες** χουρμάδες με υπνωτικό και όταν οι **τελευταίοι** αποκοιμήθηκαν κατάφεραν να αποδράσουν.

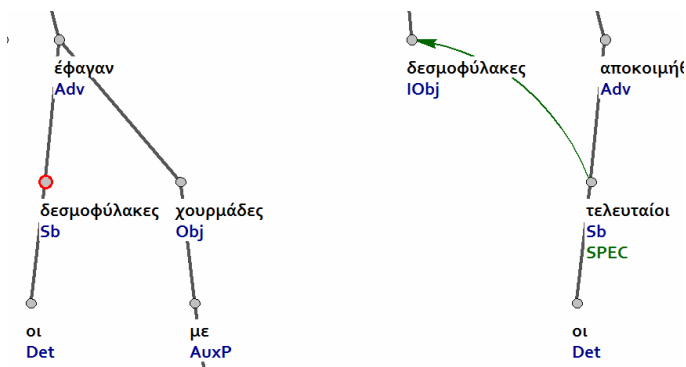


When pressing **x** in order to remove the coref arrow, a pop-up window appears with two options for re-connecting the chain:

#### Option A

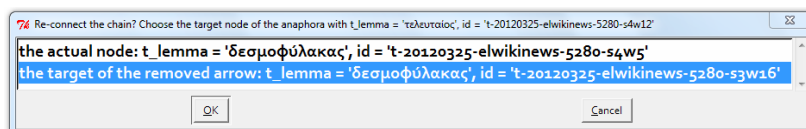


#### Result of option A

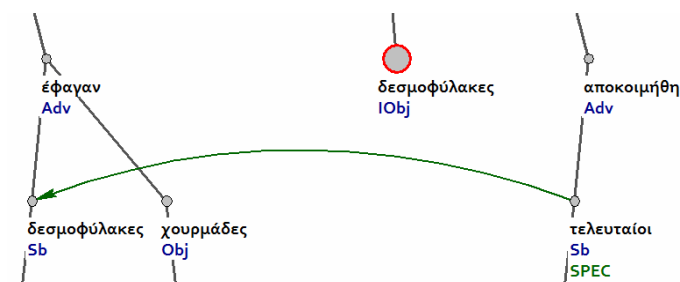


The 1<sup>st</sup> mention of *δεσμοφύλακες* (the first member of the chain) is removed.

#### Option B



#### Result of option B



The 2<sup>nd</sup> mention of *δεσμοφύλακες* (the intermediate node) is removed from the chain and *τελευταίοι* is reconnected to the 1<sup>st</sup> mention of *δεσμοφύλακες*.

**2. Are near identity chains part of a coreferential chain?**

No, near identity chains are annotated as discreet chains.

**3. What is the meaning of "metonymy\_feature\_targetNode" and "neid\_feature\_targetNode" values in the comment menu?**

The metonymy type or near identity type are marked on the source node of a chain, yet they characterise the target node.

**4. What should I choose in cases of ambiguity between coreference and bridging?**

Choose coreference.

## 10. REFERENCES

For compiling the annotation guidelines the following works have been consulted:

AnCora-CO (2008): Coreference Guidelines for Catalan and Spanish, v2.1

Cinková Silvie, Hajič Jan, Mikulová Marie, Mladová Lucie, Nedolužko Anja, Pajas Petr, Panevová Jarmila, Semecký Jiří, Šindlerová Jana, Toman Josef, Urešová Zdeňka and Žabokrtský Zdeněk (2005): Annotation of English on the tectogrammatical level. Reference book. TR-2005-28, ÚFAL MFF UK, Prague

Coreference Corpus Annotation Guidelines for ACL Anthology papers (2012). URL: [take.dfki.de/C12-2103-dataset/AnnotationGuidelines.html](http://take.dfki.de/C12-2103-dataset/AnnotationGuidelines.html)

Desypri Elina (2001) Coreference. MA Thesis

Mikulová Marie, Bémová Allevtina, Hajič Jan, Hajičová Eva, Havelka Jiří, Kolářová Veronika, Kučová Lucie, Lopatková Markéta, Pajas Petr, Panevová Jarmila, Ševčíková Magda, Sgall Petr, Štěpánek Jan, Urešová Zdeňka, Veselá Kateřina and Žabokrtský Zdeněk (2006): Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual. TR-2006-30, ÚFAL MFF UK, Prague

Mírovský Jiří, Pajas Petr and Nedoluzhko Anna (2010): Annotation Tool for Extended Textual Coreference and Bridging Anaphora. Proceedings of the 7<sup>th</sup> International Conference on Language Resources and Evaluation (LREC'10), pp. 168-171, Valletta, Malta

Nědolužko Anna (2011): Rozšířená textová koreference a asociační anafora. Koncepte anotace českých dat v Pražském závislostním korpusu. Studies in Computational and Theoretical Linguistics, Vol. 10, Institute of Formal and Applied Linguistics, Prague, Czech Republic

Nědolužko Anna and Mírovský Jiří (2011): Annotating extended textual coreference and bridging relations in the Prague Dependency Treebank. Technical report. Prague, Czech Republic. URL: <http://ufal.mff.cuni.cz/~nedoluzko/koref anot/manual angl final printed.pdf>

Nědolužko Anna and Mírovský Jiří (2013): How Dependency Trees and Tectogrammatics Help Annotating Coreference and Bridging Relations in Prague Dependency Treebank. Proceedings of the 2<sup>nd</sup> International Conference on Dependency Linguistics (DepLing 2013), pp. 244–251, Prague, Czech Republic

Ogrodniczuk Maciej, Zawisławska Magdalena, Głowińska Katarzyna and Savary Agata (2013): Coreference annotation schema for an inflectional language. In Gelbukh A. (ed.): Proceedings of the 14<sup>th</sup> International conference on Computational Linguistics and Intelligent Text Processing (CICLING 2013), Part I, pp. 394-407, Springer-Verlag, Berlin Heidelberg

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

Recasens Marta, Hovy Eduard and Martí M. Antònia (2010): A typology of near-identity relations for coreference (NIDENT). Proceedings of the 7<sup>th</sup> International Conference on Language Resources and Evaluation (LREC'10), pp. 149-156, Valletta, Malta

Schäfer Ulrich, Spurk Christian and Steffen Jörg (2012): A fully Coreference-annotated Corpus of Scholarly Papers from the ACL Anthology. Proceedings of the 24<sup>th</sup> International Conference on Computational Linguistics (COLING 2012), pp. 1059-1070, Mumbai, India

Βελούδης, Γ. (2005b): Η σημασία πριν, κατά και μετά τη γλώσσα. Αθήνα: Κριτική

Τζάρτζανος Αχιλλέας (1953) Νεοελληνική Σύνταξις (της Κοινής Δημοτικής). Εκδοτικός οίκος Αδελφών Κυριακίδη, Θεσσαλονίκη, Ελλάδα