

# Author Identification

Katerina T. Frantzi

Dept. of Mediterranean Studies

University of the Aegean

Rhodes, Greece

[frantzi@rhodes.aegean.gr](mailto:frantzi@rhodes.aegean.gr)

# The area

Forensic Linguistics:

The interface between language and crime



# terrorism

3 days ago:

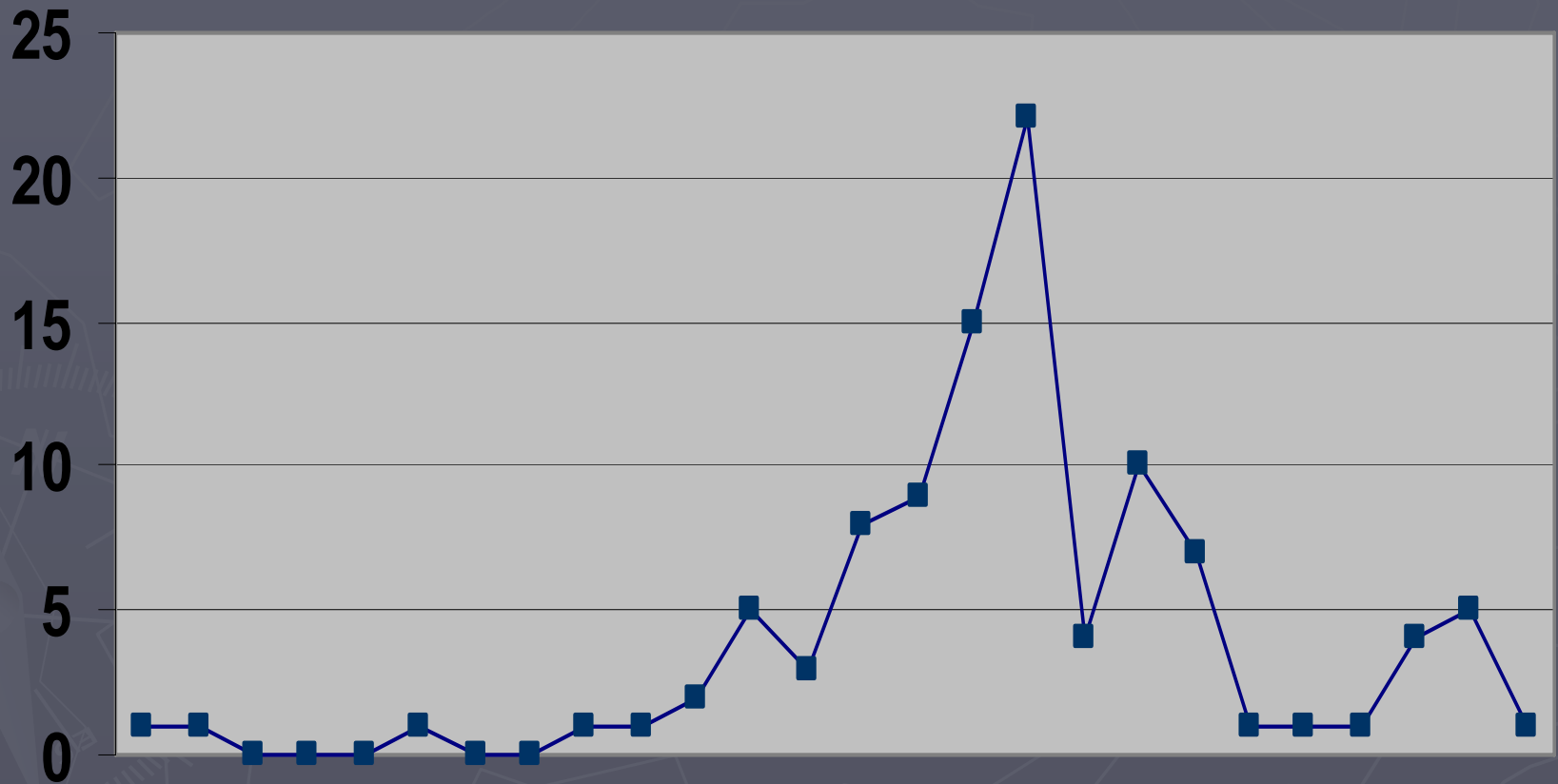
new message from Osama Bin Laden

- Characterised genuine

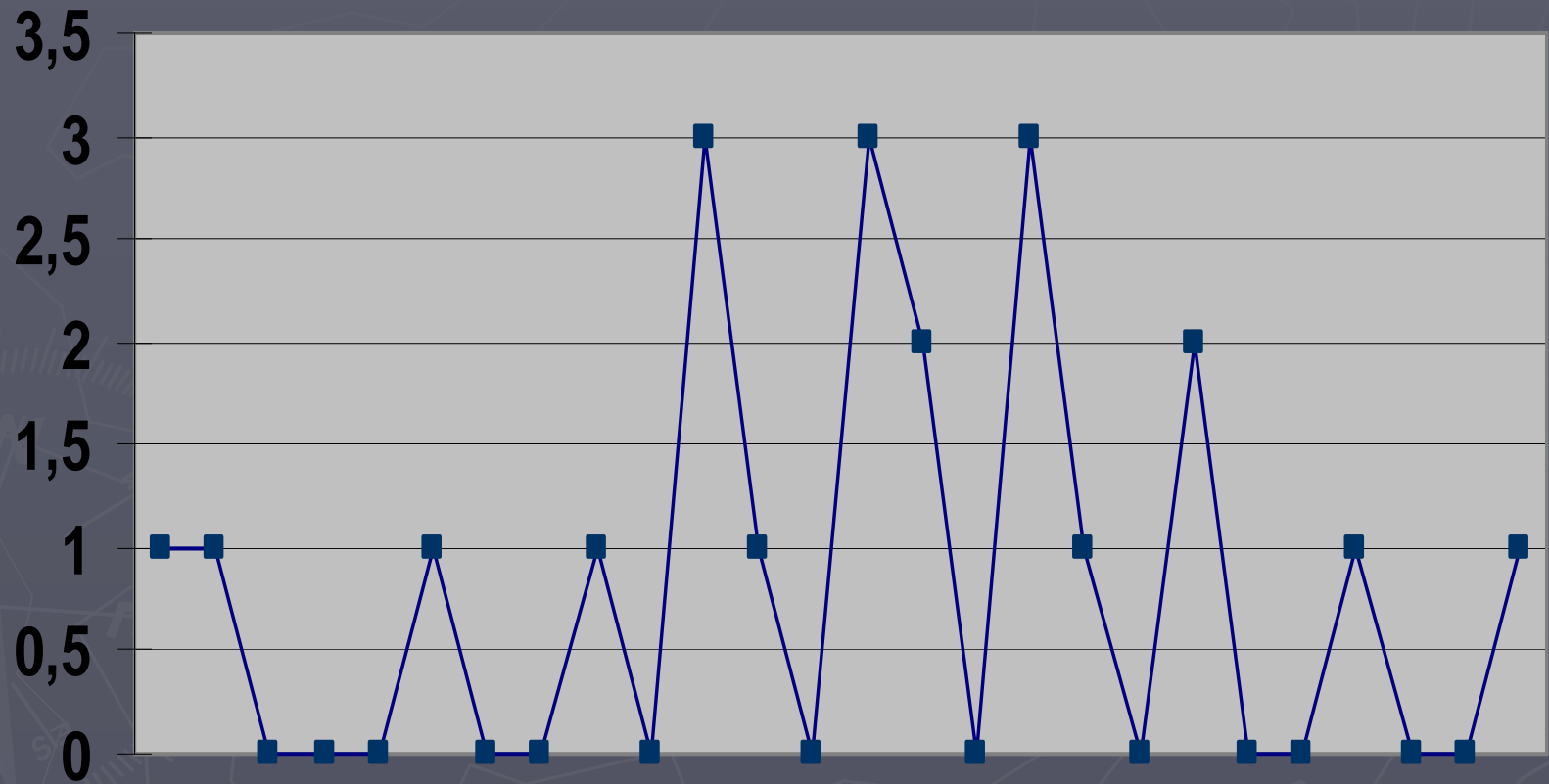
# The object

- ▶ 17N
- ▶ 1975 – 2002
- ▶ 1975 - Richard Welch
- ▶ 2000 - Steven Sonders
- ▶ 84 proclamations

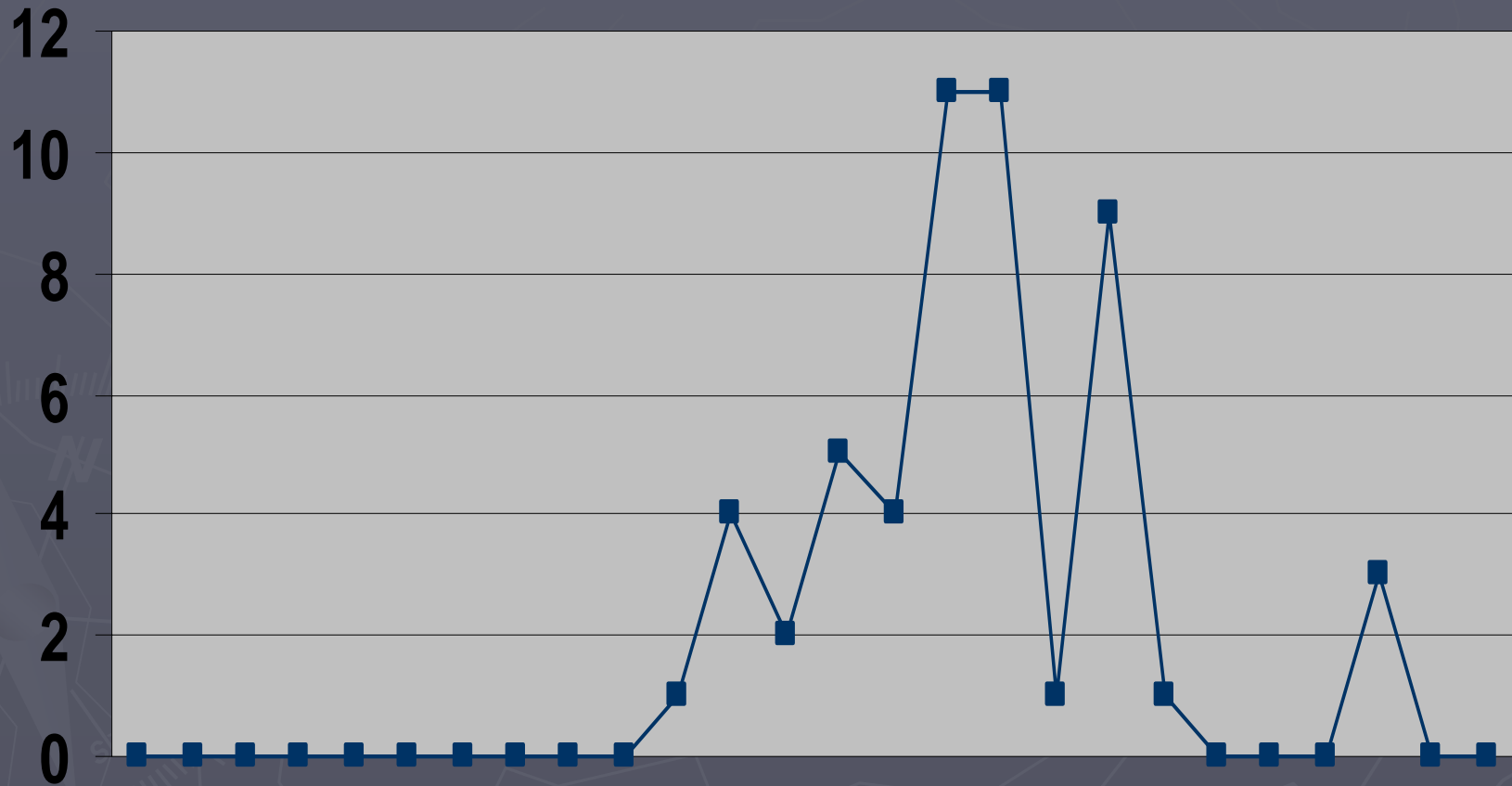
# 17N attacks 1975-2000



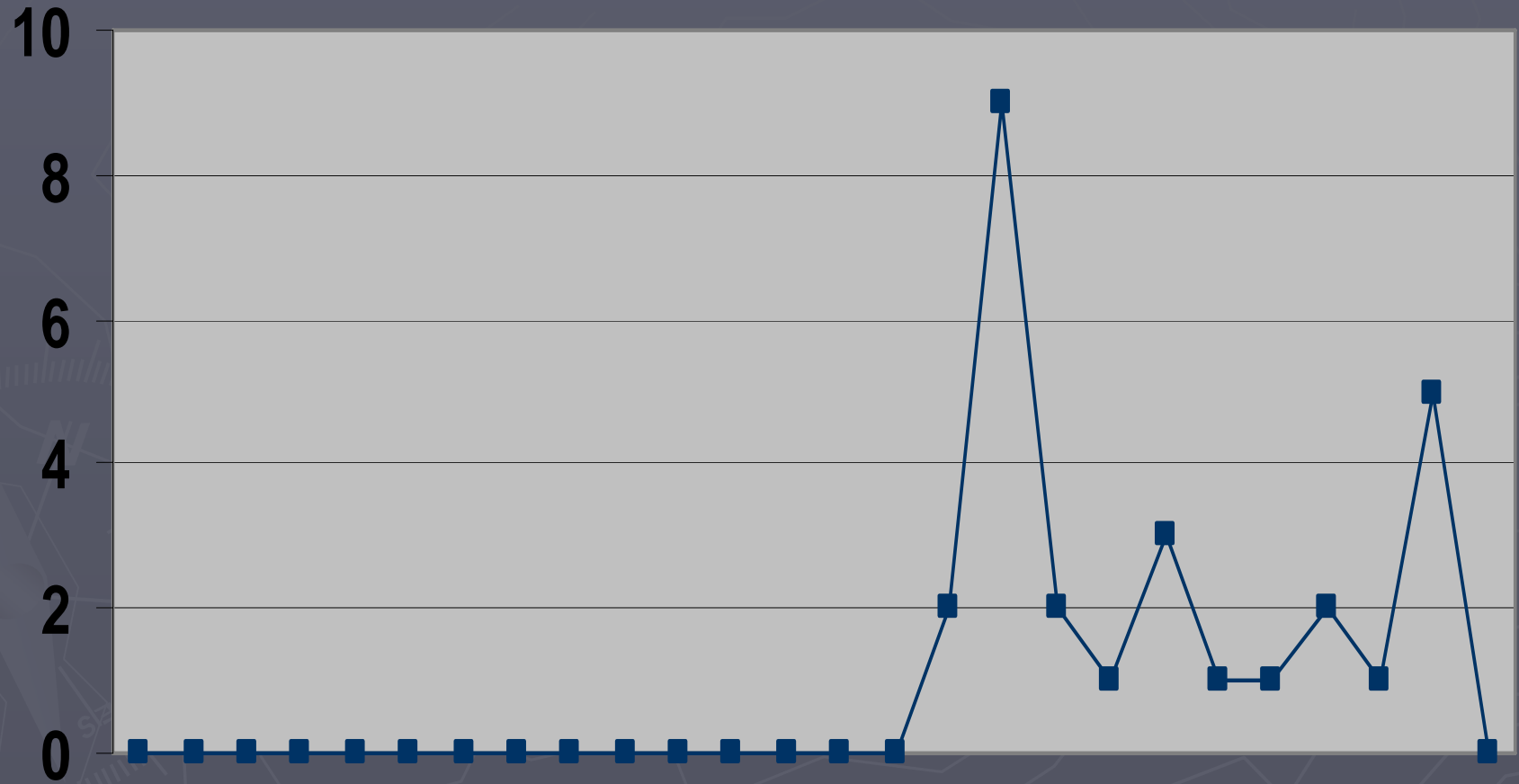
# 17N murder attacks 1975-2000



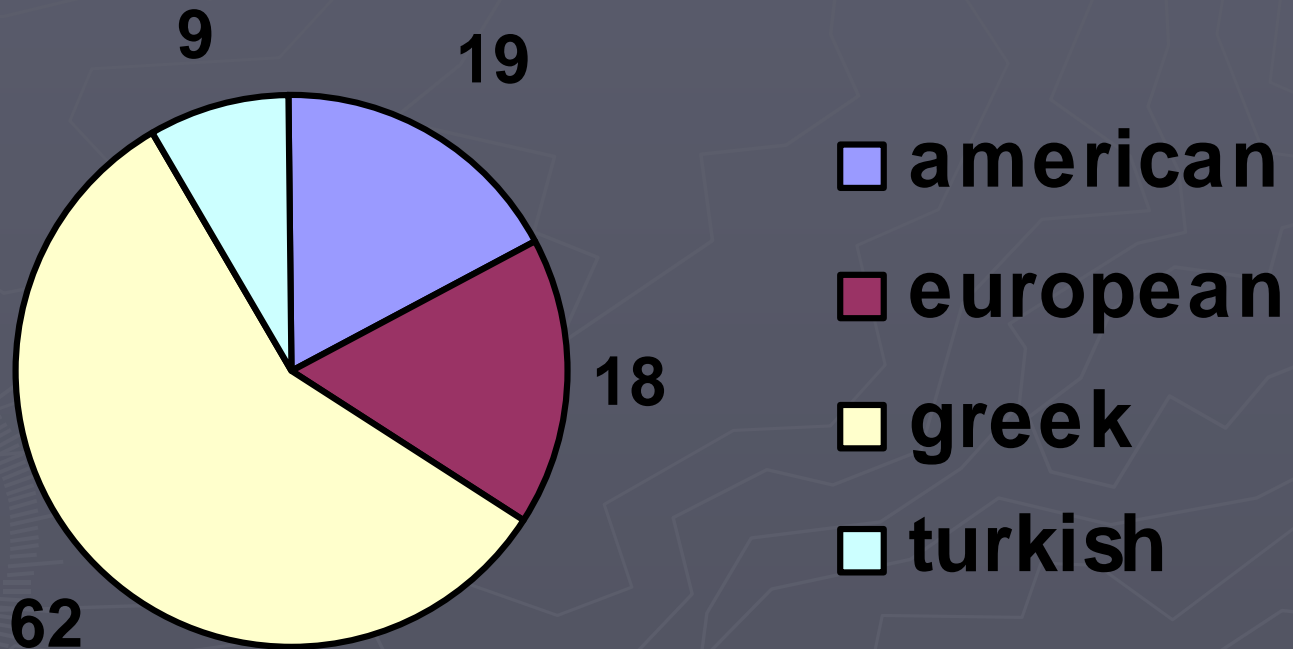
# 17N bomb attacks 1975-2000



# 17N rocket attacks



# 17N targets



None of those accused to be 17N members  
plead guilty for writing the proclamations

Yotopoulos, accused to be the leader and  
main writer, did not confess so

Have all members been arrested?

# The problem

Map each proclamation to a possible writer

Make proclamations groups

A member is characterised by its plea

If we could match a proclamation to a plea  
then we could assign it to its owner

So, we need to check matching among all the  
pleas and the proclamations

Each plea is characterised by a set of attributes' values

The same applies for each proclamation

# Attributes considered

- ▶ Num of sentences
- ▶ Average word length
- ▶ Num of different nouns
- ▶ Num of different adjectives
- ▶ Num of different proper nouns
- ▶ Num of different verbs (active voice)
- ▶ Num of different verbs (passive voice)
- ▶ Num of different adverbs
- ▶ Num of future tense use
- ▶ Conjunction use
- ▶ Preposition use
- ▶ Pronouns use
- ▶ Particles use
- ▶ Determiners use
- ▶ Articles use
- ▶ Question words use
- ▶ Punctuation use
- ▶ Content words use

We find the “distance” between a plea and a proclamation in terms of their attributes values

The smallest distance gives the assignment of a proclamation to a plea, ie. to an accused member

# distance matrix

Particles distance	Determiners distance	Articles distance	Conjunctions distance	Pronouns distance
prepositions distance	Question-words distance	future distance	# of diff Proper noun distance	# of diff adverbs distance
# of sentences distance	# of diff adjectives distance	# of diff passive verbs distance	# of diff nouns distance	# of diff active verbs distance
Compared words distance	? distance	! distance	: distance	Word length distance

The set of pleas of the accused to be 17N  
members

The set of pleas<sub>2</sub> of the accused to be 17N  
members

from the first trial

Can we map pleas<sub>2</sub> to pleas correctly?

# Speech

## Lost information

- ▶ Spelling
- ▶ Number of paragraphs
- ▶ Some punctuation
- ▶ ...

for each pair (plea<sub>a</sub>, plea<sub>b</sub>)

for each cell i assign

feature<sub>i</sub>\_distance(plea<sub>a</sub>, plea<sub>b</sub>)

distance(plea<sub>a</sub>, plea<sub>b</sub>) =

avg(for all i,

feature<sub>i</sub>\_distance(plea<sub>a</sub>, plea<sub>b</sub>))

$\text{feature}_i\text{-distance}(\text{plea}_a, \text{plea}_b) =$   
 $\text{avg}(\text{feature}_i\text{-distance}(\text{plea}_a, \text{plea}_b)[p_j])$

for each  $p_j$  phenomemo of feature<sub>i</sub>

$$\text{feature}_i\text{\_distance}(\text{plea}_a, \text{plea}_b)[p_j] = \frac{|f_{aj} - f_{bj}|}{(f_{aj} + f_{bj})}$$

where  $f_{aj}$ ,  $f_{bj}$  the frequencies of occurrence of  $p_j$  in  $\text{plea}_a$  and  $\text{plea}_b$

# Pleas

Person	words	Person	words
Georgiadis	12479	P_Serifis	4095
Karatsolis	6483	T_Serifis	7078
Kondilis	13688	Sotiropoulou	14723
Kostaris	14257	Tselentis	42778
Koufontinas	5699	Tzortzatos	13906
Papanastasiou	6974	V_Xiros	454
Psaradellis	20957	C_Xiros	12205
G_Serifis	1114	Yotopoulos	18321

# Pleas<sub>2</sub>

Person	Words
Kostaris	1594
Koufontinas	326
Psaradellis	502
G_Serifis	703
Tzortzatos	8111
Yotopoulos	2036

## 3 texts

- ▶  $Y_{p1}$  : Yotopoulos Plea - 18321 words
- ▶  $Y_{p2}$  : Yotopoulos Plea2 - 2036 words
- ▶  $K_{p2}$  : Koufontinas Plea2 - 326 words

Assign ( $Y_{p2}$  to  $K_{p2}$ ) or ( $Y_{p2}$  to  $Y_{p1}$ )

# Particles

$Y_{p1}$		$Y_{p2}$		$K_{p2}$	
va	450	va	60	για	6
για	183	θα	22	va	4
θα	165	για	20	θα	3
ας	6		0		0

# Particles - normalised

$Y_{p1}$		$Y_{p2}$		$K_{p2}$	
va	245.619	va	294.695	$\gamma_{ia}$	184.049
$\gamma_{ia}$	99.885	$\theta_{ia}$	108.055	va	122.699
$\theta_{ia}$	90,060	$\gamma_{ia}$	98.232	$\theta_{ia}$	92.024
aç	3,274		0		0

# Particles

Particles\_distance( $Y_{p1}, Y_{p2}$ )[va]=

$$\frac{|245.619 - 294.695|}{(245.619 + 294.695)} = 0.091$$

Particles\_distance( $Y_{p1}, K_{p2}$ )[va]=

$$\frac{|245.619 - 122.699|}{(245.619 + 122.699)} = 0.334$$

# Particles

	Particles_distance( $Y_{p1}, Y_{p2}$ )	Particles_distance( $Y_{p1}, K_{p2}$ )
Na	0.091	0.334
$\gamma_{ia}$	0.0088	0.251
$\theta_a$	0.0587	0.051
$\alpha_\zeta$	1	1
avg	0.289	0.409

# distance( $Y_{p1}, Y_{p2}$ )

0.289	0.225	0.191	0.552	0.63
0.32	0.225	<i>0.09</i>	0.305	0.34
<i>0.232</i>	0.04	<i>0.17</i>	0.017	0.009
0.616	0.477	1	1	0.036

# distance( $Y_{p1}, K_{p2}$ )

0.409	0.923	0.667	0.922	0.849
0.622	0.776	<i>0.010</i>	0.575	0.474
<i>0.176</i>	0.27	<i>0.115</i>	0.485	0.073
0.798	1	1	1	0.042

$$\text{distance}(Y_{p1}, K_{p2}) = 0.5593$$

$$\text{distance}(Y_{p1}, Y_{p2}) = 0.3382$$

→  $Y_{p1}$  is linked to  $Y_{p2}$

# Distances to $Y_{p1}$

Kostaris	Koufontinas	Psaradellis
0.7842	0.5593	0.5110
G_Serifis	Tzortzatos	Yotopoulos
0.4637	0.7096	<b>0.3382</b>

# Distances to Kou<sub>p1</sub>

Kostaris	Koufontinas	Psaradellis
0.6932	0.3904	0.5204
G_Serifis	Tzortzatos	Yotopoulos
0.4198	0.7258	0.5836

# Distances to $Kos_{p1}$

Kostaris	Koufontinas	Psaradellis
0.3091	0.4601	0.3890
G_Serifis	Tzortzatos	Yotopoulos
0.5329	0.3372	0.7149

# Distances to $P_{p1}$

Kostaris	Koufontinas	Psaradellis
0.4861	0.5580	0.3420
G_Serifis	Tzortzatos	Yotopoulos
0.6012	0.5912	0.4357

# Distances to $G_{S_{p1}}$

Kostaris	Koufontinas	Psaradellis
0.6954	0.4177	0.5840
G_Serifis	Tzortzatos	Yotopoulos
0.3230	0.7582	0.3986

# Distances to $T_{p1}$

Kostaris	Koufontinas	Psaradellis
0.4102	0.7513	0.4950
G_Serifis	Tzortzatos	Yotopoulos
0.5873	0.3644	0.7339

- ▶ Comparisons between pleas and proclamations
- ▶ Comparisons among the proclamations

(IAFL 2006, 14-16 September, Barcelona)

- ▶ Are all attributes of same importance (weights)?
- ▶ Add more attributes (e.g. collocations)
- ▶ Distribution of features
- ▶ ...

What about other text types,  
e.g. novels, poems etc?

We do not know.

# Future Plans

- ▶ Test the use of weights
- ▶ Test more attributes
- ▶ authorship identification on world terrorist organisations proclamations

e.g. Al-Qaida, Grey Wolves etc.

# Some basic Refs

- ▶ G. Kassimeris; 1997. *Europe's Last Red Terrorists. The revolutionary Organizations 17 November*. C. Hurst & Co. Ltd.
- ▶ Α. Παπαχελάς, Τ. Τέλλογλου; 2003. *Φάκελος 17 Νοέμβρη*. Εστία.
- ▶ Γ. Πρεπεντέρης; 2002. *Η αναμέτρηση – Ζωή και θάνατος της 17Ν*. Βιβλιοπωλείον της Εστίας.
- ▶ J. Olsson, 2004. *Forensic Linguistics – An Introduction to language, crime and the law*. Continuum.
- ▶ G.R. McMenamin, 2002. *Forensic Linguistics – Advances in Forensic Stylistics*. CRC Press.
- ▶ ...