

Recent Developments in Forensic Idiolectometry

Recent Developments in Forensic Idiolectometry*

Raquel Casesnoves, Jordi Cicres, Montserrat Forcadell, Núria Gavalda, Fernanda López, Maria Spassova, Maria Teresa Turell (IP).

ForensicLab
 Institut Universitari de Lingüística Aplicada
 Universitat Pompeu Fabra, Barcelona (Spain)
iulaforensic@upf.edu
<http://www.iula.upf.edu/forensiclab/fpresca.htm>

* This research has been conducted through the project "Idiolectometria Forense e Índice de Similitud Idiolectal" (FI2008-03583-FILO), funded by the Spanish Ministry of Science and Technology (2008-2011)


 INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

1

Recent Developments in Forensic Idiolectometry

Outline

1. Introduction
2. Objectives
3. Exploratory hypotheses
4. Corpora
5. Methodology
6. Summary of IIS calculations with methods 1 and 2
7. Results with Method 3
8. Discussion and conclusions


 INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

2

Recent Developments in Forensic Idiolectometry

Outline

1. Introduction
2. Objectives
3. Exploratory hypotheses
4. Corpora
5. Methodology
6. Summary of IIS calculations with methods 1 and 2
7. Results with Method 3
8. Discussion and conclusions

 INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA


3

Recent Developments in Forensic Idiolectometry

1. Introduction

- This paper presents experimental work in progress derived from two research projects on Forensic Idiolectometry.
- Main object of study: the **idiolectal style** in its application to **forensic voice and written text comparison**.
 - “the set of choices that the speaker/writer makes from the linguistic repertoire available to them as users of a specific language.” (Nolan, 1994)
 - the speaker’s “own distinct and individual version of the language they speak and write.” (Coulthard 2004)

→ Idiolectal style is individual and unique.


 INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

4

Recent Developments in Forensic Idiolectometry

Outline

1. Introduction
2. Objectives
3. Exploratory hypotheses
4. Corpora
5. Methodology
6. Summary of IIS calculations with methods 1 and 2
7. Results with Method 3
8. Discussion and conclusions

 INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA


5

Recent Developments in Forensic Idiolectometry

2. Objectives (1)

- These 2 projects carried out by ForensicLab try to:
 - Explore and develop a possible measurement of the linguistic differences between idiolects and each individual’s idiolectal distance.
 - Create an **Index of Idiolectal Similitude (IIS)** that
 - a) compares several linguistic samples.
 - b) calculates the linguistic distance between them.

→ What kind of idiolectal similitude is needed to reliably say that two linguistic samples (spoken or written) have been produced by the same person?

 INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

6

Recent Developments in Forensic Idiolectometry

2. Objectives (2)

- The IIS as a continuum

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
 UNIVERSITAT POMPEU FABRA

7

Recent Developments in Forensic Idiolectometry

2. Objectives (3)

- To develop this IIS, we have explored:
 - Three linguistic levels
 - Phonological
 - Morphosyntactic
 - Discourse-pragmatic
 - Three languages
 - Catalan
 - Spanish
 - English

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
 UNIVERSITAT POMPEU FABRA

8

Recent Developments in Forensic Idiolectometry

Outline

- Introduction
- Objectives
- Exploratory hypotheses**
- Corpora
- Methodology
- Summary of IIS calculations with methods 1 and 2
- Results with Method 3
- Discussion and conclusions

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
 UNIVERSITAT POMPEU FABRA

9

Recent Developments in Forensic Idiolectometry

3. Exploratory hypotheses

We expect:

- Inter-speaker/author variation to be higher than intra-speaker/author variation.
- Despite the existing intra-speaker/author variation, an individual's idiolectal style to be quite stable independently of:
 - the course of time.
 - the use of different genres (results on this hypothesis are not presented here).

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
 UNIVERSITAT POMPEU FABRA

10

Recent Developments in Forensic Idiolectometry

Outline

- Introduction
- Objectives
- Exploratory hypotheses
- Corpora**
- Methodology
- Summary of IIS calculations with methods 1 and 2
- Results with Method 3
- Discussion and conclusions

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
 UNIVERSITAT POMPEU FABRA

11

Recent Developments in Forensic Idiolectometry

4. Corpora (1)

- For most linguistic modules: data from sociolinguistic interviews compiled following the Labovian framework (Labov 2001, Turell 2003), so they include spoken data.
 - Apparent and Real Time (AT/RT).
 - Different genres (spontaneous and semi-spontaneous).
- The morpho-syntactic module of Spanish considers written data.
- A set of linguistic variables was selected for each module and analysed from both a qualitative and quantitative point of view.

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
 UNIVERSITAT POMPEU FABRA

12

Recent Developments in Forensic Idiolectometry

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

Outline

1. Introduction
2. Objectives
3. Exploratory hypotheses
4. Corpora
5. **Methodology**
6. Summary of IIS calculations with methods 1 and 2
7. Results with Method 3
8. Discussion and conclusions

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

13

Recent Developments in Forensic Idiolectometry

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

5. Methodology (1)

- Experimental methodologies applied for the calculation of the IIS:

Method 1 consists in:

1. Classifying variables into generalising rules and idiolectal features according to their linguistic nature and the researcher's preliminary observations.
2. Calculate and compare the percentage of occurrence of each variable for the speakers.
3. Attributing a specific weight to groupings of variables (50%~50%; 30%~70%; 20%~80%)

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

14

Recent Developments in Forensic Idiolectometry

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

5. Methodology (2)

- Experimental methodologies applied for the calculation of the IIS:

Method 2 (applied to Mexican Spanish):

- Is based upon the measurement of the Euclidian distance between two speakers.
- Normalises the result in order to obtain a range between 0 and 1, and calculate the IIS.

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

15

Recent Developments in Forensic Idiolectometry

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

5. Methodology (3)

- Methodology used in this paper for the calculation of the IIS.

Method 3 consists in:

1. Running cross-tabulations in SPSS.
2. Looking for the Adjusted Residual Value (ARV) for each variable
3. Assign a value to each range ARV:
 - ARV <1 → 0
 - ARV >1 & <2 → 1
 - ARV >2 & <3 → 2
 - ...
 - ARV > 5 → 5 (max. ARV conversion)

$$\text{IIS} = \frac{\text{ARV (variable 1)} + \dots + \text{ARV (variable n)}}{(\text{max. ARV conversion}) * n}$$

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

16

Recent Developments in Forensic Idiolectometry

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

5. Methodology (4)

Method 3

- This ARV indicates whether the two speakers analysed have a similar distribution of the variants or not.
- The bigger the difference between the speaker is, the greater the ARV.
- It is a statistical technique, so it tells us if the differences are statistically significant or not.

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

17

Recent Developments in Forensic Idiolectometry

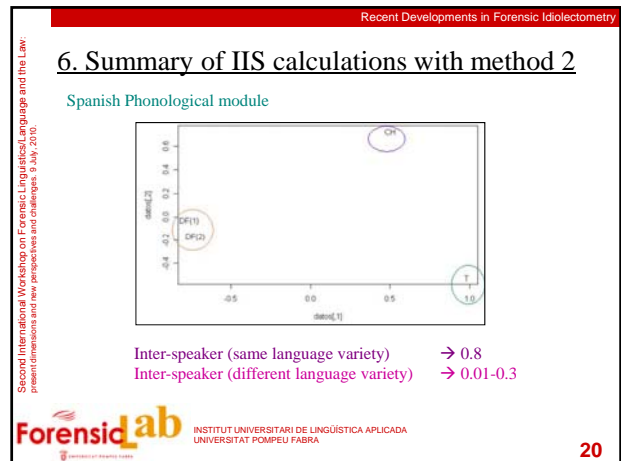
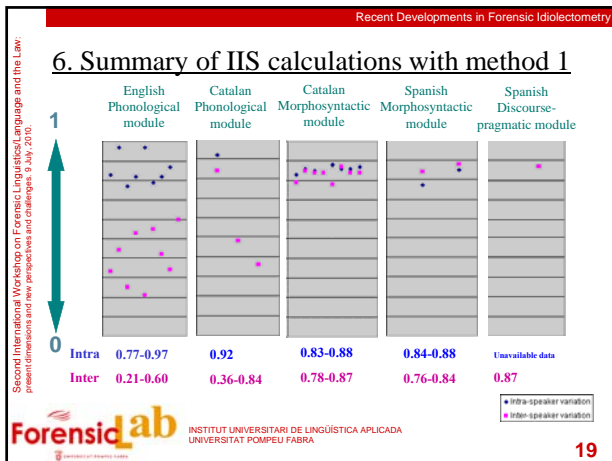
Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

Outline

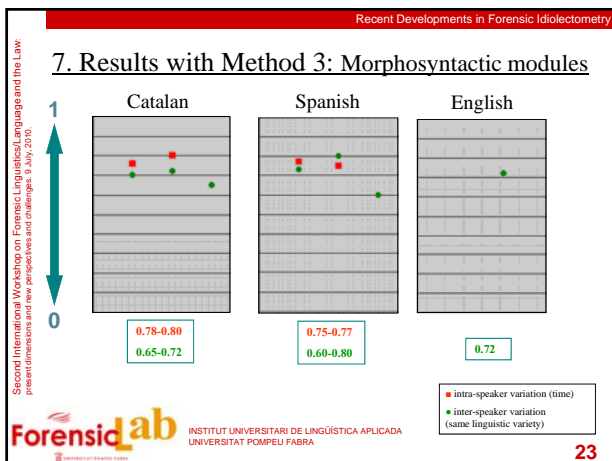
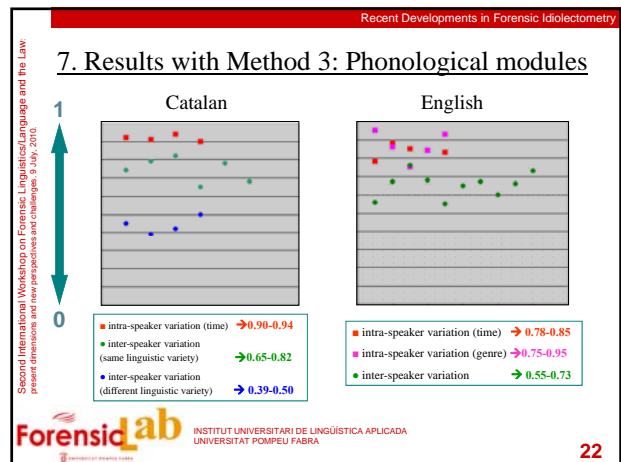
1. Introduction
2. Objectives
3. Exploratory hypotheses
4. Corpora
5. Methodology
6. **Summary of IIS calculations with methods 1 and 2**
7. Results with Method 3
8. Discussion and conclusions

ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

18



- Recent Developments in Forensic Idiolectometry
- ### Outline
1. Introduction
 2. Objectives
 3. Exploratory hypotheses
 4. Corpora
 5. Methodology
 6. Summary of IIS calculations with methods 1 and 2
 7. Results with Method 3
 8. Discussion and conclusions
- ForensiqLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA UNIVERSITAT POMPEU FABRA
- 21



Recent Developments in Forensic Idiolectometry

Outline

1. Introduction
2. Objectives
3. Exploratory hypotheses
4. Corpora
5. Methodology
6. Summary of IIS calculations with methods 1 and 2
7. Results with Method 3
8. Discussion and conclusions

25

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA UNIVERSITAT POMPEU FABRA

Recent Developments in Forensic Idiolectometry

8. Discussion and Conclusions (1)

Intra-speaker/writer IIS values

Inter-speaker/writer IIS values

26

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA UNIVERSITAT POMPEU FABRA

Recent Developments in Forensic Idiolectometry

8. Discussion and Conclusions (2)

1. Preliminary analyses show interesting results that confirm our hypotheses in that for all modules:
 - there seems to be greater inter- than intra-speaker/writer variation.
 - an individual's idiolect seems to be quite stable despite the course of time and the use of different genres.
2. Unexpected results for the morphosyntactic and discourse-pragmatic modules.
 - Possible reason: idiosyncratic nature of variables (less idiolectal than phonological features?)

27

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA UNIVERSITAT POMPEU FABRA

Recent Developments in Forensic Idiolectometry

8. Discussion and Conclusions (3)

- We think that it will be possible to come up with more robust and reliable results by further research, which will allow us to:
 - Refine the selection of the analysed variables, according to linguistic criteria.
 - Extend our experiments to a wider corpus, including data from various language varieties and different genres.
 - Increase the sample and corpus size with real world texts.
 - Extend the application of the IIS procedure to real forensic texts.

28

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA UNIVERSITAT POMPEU FABRA

Recent Developments in Forensic Idiolectometry

Thank you!

29

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA UNIVERSITAT POMPEU FABRA

Recent Developments in Forensic Idiolectometry

Recent Developments in Forensic Idiolectometry *

Raquel Casesnoves, Jordi Cires, Montserrat Forcadell, Núria Gavaldà, Fernanda López, Maria Spassova, Maria Teresa Turell (IP).

ForensicLab
Institut Universitari de Lingüística Aplicada
Universitat Pompeu Fabra, Barcelona (Spain)

iulaforensic@upf.edu
<http://www.iula.upf.edu/forensiclab/fpresca.htm>

* This research has been conducted through the project "Idiolectometria forense e Índice de Similitud Idiolectal" (FI2008-03583-FILO), funded by the Spanish Ministry of Science and Technology (2008-2011)

30

Second International Workshop on Forensic Linguistics/Language and the Law: present dimensions and new perspectives and challenges. 9 July, 2010.

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA UNIVERSITAT POMPEU FABRA