

## Delatats pel llenguatge! La tasca dels lingüistes forenses

**M. Teresa Turell**

[teresa.turell@upf.edu](mailto:teresa.turell@upf.edu)

Directora del ForensicLab

Directora del Màster en Lingüística Forense IDEC-IULA

Institut Universitari de Lingüística Aplicada

Universitat Pompeu Fabra

<http://www.iula.upf.edu/forensiclab>

Consultora associada de FLAIR

Forensic Linguistic Advice, Investigation and Research

Centre for Forensic Linguistics

Aston University

<http://www.forensiclinguistics.net/flair.html>

## Delatats pel llenguatge!

- Per què aquest títol?
- El llenguatge humà proporciona informació oral i escrita diversa que pot delatar la persona que ha emès un missatge quan fa ús del seu idiolecte.
- Aquesta informació pot ser emprada pels lingüistes forenses per ajudar als jutges, magistrats o a un tribunal popular, a entendre:

– què diu exactament un text oral o escrit.

– qui és l'autor d'aquest text.

## Objectius de la conferència

- **Il·lustrar** aquesta premissa a partir de:
  - dos casos d'assassinat (gravacions orals i sms's).
  - un cas d'extorsió (correus electrònics i faxos).
- **Demostrar** com un Pèrit Judicial Lingüista pot presentar evidència lingüística de forma clara i fiable *si* se li permet fer-ho.
- **Explicar** que aquest condicional està bàsicament determinat per dificultats:
  - metodològiques pròpies de la disciplina que s'ocupa d'estudiar la informació lingüística esmentada, és a dir, la **Lingüística Forense**, que també descriuré breument.
  - derivades del nostre sistema judicial.

## Lingüística Forense

## Àmbits d'estudi, recerca i peritatge de la Lingüística Forense (LF)

- **Llenguatge jurídic**
  - Language of the Law
- **Llenguatge judicial**
  - Language of the Legal Process
- **Llenguatge evidencial**
  - Language as Evidence

## Llenguatge evidencial Idiolecte

- L'Idiolecte oral i escrit d'un individu conté:
  - trets lingüístics que poden revelar el seu gènere (sex), edat, professió, nivell educatiu, religió, filiació política, i altres.
  - claus per a la determinació de l'origen geogràfic, l'etnicitat d'un individu (Smith and Shuy 2002), la seva procedència no nadiua en l'ús d'una segona llengua o una llengua estrangera, i fins i tot la seva condició de bilingüe.
- Aquesta evidència lingüística pot ser emprada per ajudar a un jutge, magistrat o a un tribunal popular, a entendre **què diu un text i qui és el seu autor** (Coulthard 2004, 2005a, 2005b).

## Premisses de la LF

- **Comparació forense de textos:**
  - orals i escrits//dubitats i indubitats.
- **Premisses:**
  - S'observa molta més variació entre parlants/escriptors (**inter-variació**) que en la parla i l'escriptura d'un mateix parlant/escriptor (**intra-variació**).
  - L'idiomelecte oral i escrit d'un individu es manté **relativament constant en gèneres textuais diferents**
  - Aquest idiomelecte canvia **relativament poc al llarg del temps**, tot i que pot canviar per factors interns a la llengua o externs, per exemple, pel contacte amb altres llengües.

CDL de Catalunya - 5 de febrer de 2009

## Objectius de la LF Llenguatge evidencial

- **Establir a partir de la comparació de textos:**
  - entre diversos candidats, a quin d'ells correspon(en) el/els text(os) oral(s)/escrit(s) dubitat(s) [és a dir, l'autoria dels quals presenta dubte] – **determinació**.
  - si un(s) text(s) escrit(s) dubitat(s) correspon(en) a un escriptor concret – **atribució**.
  - el perfil del possible autor d'un text oral/escrit desconegut - **perfil**.

CDL de Catalunya - 5 de febrer de 2009

## Dades emprades

- **Lingüístiques.**
- Objecte d'estudi de la LF:
  - llenguatge i les llengües, en les seves múltiples manifestacions.
- No s'han de confondre amb dades estadístiques, ni amb dades jurídiques.
- El **Pèrit Judicial Lingüista** descriu el que observa, però no jutja ni, encara menys, condemna.

CDL de Catalunya - 5 de febrer de 2009

## Mètodes i tècniques de la LF

- Han de permetre:
  - Establir la **significació estadística** dels resultats de les nostres anàlisis.
  - Assegurar en els dictàmens, i davant del jutge en la Vista Oral, la màxima **fiabilitat i validesa** d'aquests.

CDL de Catalunya - 5 de febrer de 2009

## Complementarietat evidencial

- Davant la impossibilitat de quantificar i aplicar proves estadístiques, s'ha de plantejar el caràcter **complementari** de diverses tècniques d'anàlisi lingüística,
  - a) dins la prova pericial lingüística:
    - **complementarietat intra-evidencial**, i
  - b) entre diferents proves derivades d'altres disciplines (medicina, psicologia, enginyeria, etc.):
    - **complementarietat inter-evidencial**.

CDL de Catalunya - 5 de febrer de 2009

## Dificultats

CDL de Catalunya - 5 de febrer de 2009

### Dificultats de consolidació (1)

- Naturalesa de nostre sistema judicial (Pérez 2005):
  - Basat en el Civil Law, també conegut com el dret continental.
  - Estructurat casi exclusivament a partir de la figura del jutge.

CDL de Catalunya - 5 de febrer de 2009

### Dificultats de consolidació (2)

- No sempre es demana una segona prova pericial, tal com requereix la legislació:
  - **Llei d'Enjudiciament Civil**
  - **Llei d'Enjudiciament Criminal**

CDL de Catalunya - 5 de febrer de 2009

### Dificultats de consolidació (3)

- Minsa presència dels pèrits lingüistes en les nostres sales de vistes:
  - Història i idiosincràsia de la nostra cultura jurídica.
  - Desconeixement per part dels actors judicials directes (jutges i magistrats, Ministeri Fiscal, lletrats de la defensa, ciutadans) de l'activitat que poden portar a terme els lingüistes forenses.
- La Lingüística Forense, aquesta gran desconeguda!

CDL de Catalunya - 5 de febrer de 2009

### Situació actual de la LF a Catalunya i Espanya

- Desenvolupament:
  - **Incipient**
  - **Lent**
  - **Desigual**

CDL de Catalunya - 5 de febrer de 2009

### Dificultats metodològiques (1)

- Situació de desavantatge metodològic de la **comparació forense de textos escrits** respecte de la **comparació forense de la veu humana** - molt més desenvolupada gràcies a la fonètica i l'acústica forenses:
  - la comparació forense de veu ja compta amb informació relativa a la distribució poblacional sobre determinats paràmetres idiosincràtics de la veu, sobretot en anglès.
  - la comparació forense de textos escrits no compta amb aquesta informació de distribució poblacional respecte de marques escrites d'autoria.
- Aquesta informació és el que els lingüistes forenses denominem el **base-rate knowledge** (Grant 2007).

CDL de Catalunya - 5 de febrer de 2009

### Dificultats metodològiques (2)

- Naturalesa de les dades i la falsa comparació de les mostres lingüístiques (orals però, sobretot, escrites) amb l'empremta dactilar o amb l'ADN:
  - En l'**empremta dactilar** i l'**ADN**, cada mostra és idèntica i exhaustiva (és a dir, conté tota la informació relativa a un individu).
  - En canvi, cada **mostra lingüística** només facilita informació molt parcial de l'idiòlecte del seu creador.

CDL de Catalunya - 5 de febrer de 2009

### Dificultats metodològiques (3)

- Per tant, identificar a un parlant específic o atribuir l'autoria d'un text escrit a un escriptor concret, a partir de la comparació d'una o diverses mostres lingüístiques **dubitades** (d'un sospitós) amb mostres **indubitades** d'aquest parlant o escriptor
  - es presenta com a una tasca impossible,
  - ja que per poder-ho fer de forma fiable
    - s'hauria de poder comptar amb bancs de dades constituïts per mostres lingüístiques de cada individu,
    - en tots i cadascuns dels contextos, gèneres textuais i registres en què es pot realitzar i expressar el seu idiolecte,
    - i aquest fet comportaria l'existència de mostres representatives de milions d'idiolectes.
- Aquesta dificultat en l'obtenció del mostreig desitjable, com acostuma a passar en casos reals, dificulta també l'obtenció del **base-rate knowledge**.

CDL de Catalunya - 5 de febrer de 2009

### Casos reals de lingüística forense *in vivo*

CDL de Catalunya - 5 de febrer de 2009

### 1. El cas Bieber

Comparació forense de veu  
1 mostra dubitada

+  
Diverses mostres indubitades  
Un únic autor



### Agraïments

Peter French  
Philip Harrison  
[www.jpfrrench.com](http://www.jpfrrench.com)  
JP French Associates  
Registrats a Anglaterra i Gal·les  
University of York.  
Department of Language and Linguistic Science

CDL de Catalunya - 5 de febrer de 2009

### Context del cas (1)

- Desembre 2003, Leeds, Anglaterra.
- La policia de tràfic identifica un cotxe robat.
- El conductor és detingut i se'l situa a la part de darrera del cotxe policial.
  - Dona un nom fals: Nathan Coleman.
- Conversa gravada a través de la càmera de vídeo situada a la part de davant del cotxe policial.



CDL de Catalunya - 5 de febrer de 2009



### Context del cas (2)

- Després de la conversa gravada a l'interior del cotxe policial:
  - 'Coleman' treu una pistola, dispara i mata al policia Ian Broadhurst.
  - Intenta matar a dos policies més.
  - Finalment escapa.



CDL de Catalunya - 5 de febrer de 2009



### Context del cas (3)

CDL de Catalunya - 5 de febrer de 2009

- Unes setmanes més tard se'l torna a detenir en un hotel de Newcastle.
- Disfressat i amb una pistola amagada.
- No vol parlar amb la policia si no és per negar la seva participació en l'assassinat del policia Ian Broadhurst.



### Anàlisi de veu (1)

CDL de Catalunya - 5 de febrer de 2009

- S'havia de demostrar que la persona que havien detingut aquesta segona vegada era la mateixa que havien enregistrat dins el cotxe policial uns minuts abans d'assassinar el policia Broadhurst.
- La policia disposava d'una mostra criminal i necessitava una mostra, o diverses mostres, de referència comparativa.
- 'Coleman' era un jugador d'apostes i feia servir molt el servei de les agències d'apostes telefòniques que graven les trucades d'entrada.
- Les mostres de referència van ser aquestes trucades.



### Mostres de referència (trucades d'apostes)



Hey mate give me an update on the rugby union um the um Munster versus Trevis(o)- T X thirty five thirty one five. And college basketball uh Florida versus Louisville. Could you tell me how much time there is left please? Florida and Louisville. **OK mate, cheers.**

Hey mate. College basketball. Oregon. T X thirty five thirty one five. (that's ?) uh just gimme an update how much time's left. So Oregon... you know... **you guys are betting on it.** Seventy fifty two. Cheers mate.

Hello mate. Could you- can you give me an update on the uh **Leicester versus Gwent** rugby union score? T X thirty five thirty one five. Right...

### Mostra criminal (vídeo interior cotxe policial)



Right. Yeah, he mentioned ... Right. Mm actually I don't have my driver's licence. From Leeds yeah. Canada.

No, that's what I thought it was. F- f- I felt it. **I thought I had it.** (?) no I did not steal the car. No, I **don't have it, mate.** Don't have it, mate.

Are you sure it's coming out stolen? Did it come out? Wh- wh- what do y- what do you go by? (Yeah). Where was it stolen? What's he need now? Can you roll this window down a bit, I'm just claustrophobic. I'm on medication for it. Pardon? Can you let me out for a sec, I've got- I'm on medication for claustrophobia and anxiety. **Pardon?** Just- just the window down or something. What we waiting on? What we waiting on? Got to wait for vehicle recovery?

### Anàlisi de veu (2)

- El laboratori JP French Associates va analitzar diversos paràmetres de les veus dels dos conjunts de gravacions:
  - consonants i vocals
  - qualitat de la veu
  - entonació
  - lèxic
  - gramàtica
  - altres



	Mostra criminal (vídeo)	Mostra de referència (apostes)
<b>rhoticity</b> 	pardon	Leicester <b>r</b> versus Gwent
<b>'t' = [d]</b> 	thought <b>t</b> I had it	you guys are be <b>t</b> ting on it
<b>mate</b> ([e:]) 	don't have it <b>mate</b>	ok <b>mate</b> , cheers
<b>[o:]</b> 	<b>don't</b> have it mate	Springwood <b>Road</b>

## Anàlisi de veu (3)

DL de Catalunya - 5 de febrer de 2009

- A partir de proves auditives i acústiques es va poder observar que tant la mostra criminal com les mostres de referència presentaven una barreja molt distintiva de trets de:
  - l'anglès americà de Florida.
  - l'anglès de Yorkshire.
- La conclusió de JP French Associates va ser:
 

**Una veu molt consistent i excepcionalment distintiva**



## Anàlisi de veu(4)

DL de Catalunya - 5 de febrer de 2009

- Davant d'aquesta evidència "Coleman" va inventar-se una nova coartada argumentant que:
  - Es trobava dins del cotxe robat amb un altre home que tenia una veu semblant.
  - Que l'altre home era l'assassí.
  - Que no podia revelar el nom de l'assassí per no posar en perill la seva pròpia vida.



## Anàlisi de veu (5)

DL de Catalunya - 5 de febrer de 2009

- La investigació posterior també va permetre aclarir altres circumstàncies del cas i explicar el comportament de "Coleman":
- El seu nom real era David Bieber
  - Un ciutadà americà de Florida.
  - Buscat per diversos intents d'assassinat i de tràfic de drogues als EE.UU.
  - Amb el risc afegit de ser condemnat al Regne Unit per diversos delictes.



## Anàlisi de la veu (6)

DL de Catalunya - 5 de febrer de 2009

- Durant el judici, el pèrit lingüista Peter French va concloure que era:
 

**"Una de les veus més distintives que havia analitzat durant els seus 20 anys de dedicació a la fonètica forense".**
- A "Coleman" se'l va declarar culpable i se'l va condemnar a cadena perpetua.
- L'evidència lingüística va ser essencial i decididament el seu idiolecte característic el va delatar.



CDL de Catalunya - 5 de febrer de 2009

## 2. El cas dels missatges de text (sms)

Comparació forense de text  
Diverses mostres dubitades  
+  
Diverses mostres indubitades  
Dos possibles autors

"Criminals beware ... texting may lead the police to your door" The Scotsman 2008

## Agraïments

DL de Catalunya - 5 de febrer de 2009

- Malcolm Coulthard
  - [http://www.youtube.com/watch?v=4z6Krsjw\\_c84](http://www.youtube.com/watch?v=4z6Krsjw_c84)
- Tim Grant
  - <http://news.bbc.co.uk/2/hi/science/nature/7600769.stm>
- Centre for Forensic Linguistics  
Aston University
  - <http://www.forensiclinguistics.net>

## Identificació d'escriptors Missatges de text (sms)

CDL de Catalunya - 5 de febrer de 2009

- Ha quedat demostrada la utilitat de l'anàlisi lingüística de sms's (text messaging) en la resolució de diversos casos d'assassinat a Anglaterra.
- Les conclusions referides a casos reals i la recerca que comporta cada cas de lingüística forense evidencial és que es poden distingir escriptors per les seleccions lingüístiques que fan al escriure.
- Context investigador:
  - es reben sms's enviats des d'un mòbil després que el/la propietari d'aquest mòbil desaparegui.
  - es descobreix posteriorment que la persona desapareguda ha estat assassinada.
  - i no obstant això, se segueixen enviant i rebent missatges des del mòbil del desaparegut/da.

## El cas d'assassinat de Jenny Nicholl (19 años) Yorkshire - 2005

CDL de Catalunya - 5 de febrer de 2009

## Textos dubitats (1)

CDL de Catalunya - 5 de febrer de 2009

### sms dubitat 1:

"hi jen tell jak **i am** ok know ever 1s gona b mad tell them **i am** sorry.living in scotland wiv my boyfriend.shitting **meself** dads gona kill me mum dont give a shite.hope nik didnt grass me up.keeping **phone** of.tell dad car jumps out of gear and stalls put it back in auction.tell him **i am** sorry"

(Enviat a Jennifer Whelan el 9 de juliol de 2005)

## Textos dubitats (2)

CDL de Catalunya - 5 de febrer de 2009

### sms dubitat 2:

"Thought u wer grassing me up.mite b in trub wiv **me** dad told mum i was lving didnt giv a shit.  
**been2 kessick** camping was great. **ave2 go cya**"

(Enviat a Nicola Gosnold el 9 de juliol de 2005)

## Textos dubitats (3)

CDL de Catalunya - 5 de febrer de 2009

### sms dubitat 3:

"Y do u h8 me i know mum does.told her i was goin. **i aint** cumin back and the pigs wont find me. **i am** happy living up here.every1 h8s me in rich only m8 i got is jak.txt u couple wks tell pigs **i am** nearly 20 **aint** cumin back they can shite off She got me in this shit its her fault not mine get blame 4evrything. **i am** sorry ok just **had 2 lve** shes a bitch no food in and always searching me room eating me sweets.  
**ave2 go** ok **i am** very sorry x"

(Enviat a Brian Nicholl, pare de Jenny, el 14 de juliol de 2005)

## Textos indubitats (3/11)

CDL de Catalunya - 5 de febrer de 2009

### sms indubitat 1:

"Yeah shud b gud. I just **have2get** my finga out and do anotha tape"

### sms indubitat 2:

"Only just turned my **fone**. havent lied bout anything"

### sms indubitat 3:

"**im** tierd of defending **myself**. Theres no point. Bye.

**Taula 1**  
**Contrast de marques en els textos**  
**indubitats i dubitats**

Textos indubitats (Jenny)	Textos dubitats
<b>my</b> (my dad, myself)	<b>Me</b> (me dad, meself)
<b>cy</b> (por see you)	<b>cya</b>
<b>Fone</b>	<b>phone</b>
<b>im</b> (por I am, I'm)	<b>i am</b>
<b>have2get</b>	<b>been2 kessick</b>

CDL de Catalunya - 5 de febrer de 2009

**Taula 2**  
**Corpus de sms's**  
**(Tim Grant – CFL – Aston University)**

<b>Missatges de sms</b>	<b>5000</b>
Escriptors (10/persona)	<b>500</b>
Interactius	<b>80</b>
Independents	<b>420</b>

CDL de Catalunya - 5 de febrer de 2009

### **Conclusions** **Estudi corpus sms's (Grant)**

***Els missatges intra-escriptor són més similars entre ells que els missatges interactius.***

***Els missatges interactius són més similars que els missatges produïts independentment.***

CDL de Catalunya - 5 de febrer de 2009

### **Una antiga premissa lingüística**

- El llenguatge està governat per regles (“rule-governed nature of language”).
- Regles de selecció de paraules i d'altres paraules per acompanyar les primeres.

CDL de Catalunya - 5 de febrer de 2009

### **Conclusions** **Complementarietat inter-evidencial**

- Poques dades.
- Impossibilitat d'atribuir sms's al principal sospitós, l'oncle de Jenny: David Hodgson, per manca de textos indubitats d'aquest.
- Possibilitat d'excloure Jenny Nicholl com autora dels sms's dubitats.
- Complementarietat inter-evidencial:
  - Quan van detenir David Hodgson, ell tenia el mòbil de Jenny.
  - Companyia telefònica (trucades).
  - Empresa de lloguer de cotxes (quilometratge).

CDL de Catalunya - 5 de febrer de 2009

### **Conclusions finals**

- El Ministeri Fiscal va plantejar al jurat que:
  - aquesta evidència lingüística i no lingüística no podia ser fruit d'una simple coincidència.
  - l'acusat havia intentat imitar l'estil idiolectal de Jenny Nicholl però havien surat les seves marques lingüístiques, que el van delatar.
- La sentència va ser condemnatòria.
- La paradoxa per a l'acusat, que compleix una condemna de 18 anys, podria ser la següent:
  - de no haver enviat aquests missatges de text, molt possiblement el seu crim no hauria estat descobert.
  - El va delatar el seu idiolecte en escriure missatges de text, encara que intentés imitar l'estil de la seva víctima.

CDL de Catalunya - 5 de febrer de 2009



### 3. El cas de l'extorsió per correu electrònic (email)

Comparació forense de text  
Diverses mostres dubitades  
+  
Diverses mostres indubitades  
Mateix autor?

## El cas

- Empresa del sector turístic de Catalunya és acusada d'acomiadament improcedent per un empleat que els havia intentat extorsionar.
- Aquesta extorsió està expressada en un dels textos dubitats (@mails,TD), l'autoria dels quals aquest treballador va negar més endavant i que suposadament havia enviat a l'empresa durant un període de temps en què també va enviar els textos indubitats (faxes,TI), l'autoria dels quals no va negar en cap moment.
- Ambdós tipus d'escrits van ser suposadament enviats des de la sucursal francesa que l'empresa havia obert al nord de França.

## Proves lingüístiques del cas

	Textos
<b>TD@</b>	doc01 doc02 doc03 doc04
<b>TIfax</b>	doc05 doc06 doc07 doc08

## Anàlisi qualitativa

## Temporalitat i cronologia

- 1. doc01 (TD@)**  
'(...) y más de **6.000 litros de gasoil** de calefacción'.  
(22 de setembre de 2003).
- 2. doc05 (TIfax)**  
'(...) llegando a consumir **6.000 litros de gasoil**' (...).  
(28 de setembre de 2003)

## Errors ortogràfics

- 3. doc06 ((TIfax)**  
'(...) una de las personas que más **a trabajado** para la XXX'.
- 4. doc03 (TD@)**  
'(...) el cual siempre me **a demostrado** su interés (...)'.  
'...)

## Errors gramaticals

### 5. doc07 (Tlfax)

(...) 'al cual la empresa **le han pagado**'.

### 6. doc01 (TD@)

(...) 'no creo que **hayan** muchas personas'.

## Contacte de llengües

- i) La caiguda de la preposició 'de' en contextos on el seu ús és obligatori en espanyol estàndard, com per exemple en expressions com '**en el caso ^ que**' (doc05 Tlfax), '**se dio cuenta ^ que tenía** en sus manos' (doc02 TD@).
- ii) La caiguda de la preposició 'a', on el seu ús és obligatori en espanyol estàndard, per indicar la funció sintàctica del complement indirecte, com per exemple: '**para atender ^ los clientes**' (doc08 Tlfax).

## La riquesa lèxica

## Anàlisi contrastiva

Taula 3

Textos	Legòmena	Dislegòmena	Ràtio T/U	Riquesa Lèxica
TD1@	134	10	1,26	389
TD2@	257	38	1,34	332
TD3@	202	28	1,23	403
TD4@	194	25	1,33	357
TI5fax	223	28	1,31	369
TI6fax	143	14	1,23	451
TI7fax	257	44	1,49	274
TI8fax	257	36	1,38	324

## Lingüística de corpus

## P1PS - Lingüística de corpus

### 7. (Tlfax)

'(...) de momento **yo** continuo con las funciones (...)'.  
'(...) de la que no soy **yo** responsable, sólo bastaba decírmelo'.

### 8. doc02 (TD@)

'(...) de la que no soy **yo** responsable, sólo bastaba decírmelo'.

**P1PS - Ràtios****Taula 4**

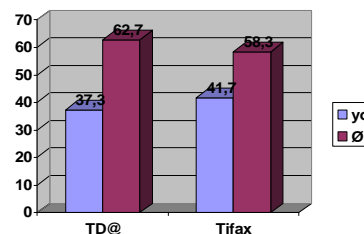
Distribució de yo i Ø (%) en @mails i faxes

	yo	Ø
<b>TD@</b>	37,3	62,7
<b>Tifax</b>	41,7	58,3

CDL de Catalunya - 5 de febrer de 2009

**P1PS - Ràtios****Gràfic 1**

Distribució de yo/ Ø - TD@ i Tifax



CDL de Catalunya - 5 de febrer de 2009

**Pronom relatiu - Lingüística de corpus****9. doc05 (Tifax)**

'(...) antes de la baja por enfermedad del Sr. XXX, **el cual** tenía que estar allí durante 1 o 2 días'.

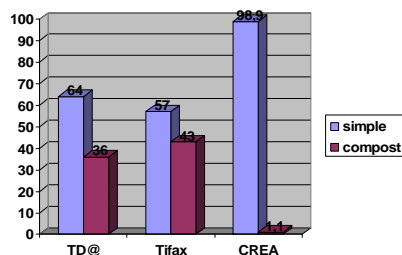
**10. doc03 (TD@)**

'Acabo de leer tu mail, **el cual** te agradezco, pero no me subestimes'.

CDL de Catalunya - 5 de febrer de 2009

**Pronom relatiu - Ràtios****Gràfic 2**

Distribució simple/compost - TD@, Tifax i CREA



CDL de Catalunya - 5 de febrer de 2009

**Seqüències d'etiquetes morfo-sintàctiques****Anàlisi quantitativa**

CDL de Catalunya - 5 de febrer de 2009

**Procediment analític**

- Selecció
- Pre-processament
- Processament
- Etiquetatge morfosintàctic
- Variables d'anàlisi
- Extracció de dades
- Anàlisi estadística

CDL de Catalunya - 5 de febrer de 2009

Delatats pel llenguatge!

## Seqüències d'etiquetes morfo-sintàctiques

[...]*el soroll de la mar gronxava les migdiades dels barcelonins* [...].

<	el	el soroll	de	la	mar	gronxav	les	midriad	de	los	barcelonese	>
...	NP	NP	PP	NP	NP	VP	NP	NP	P	NP	NP	...

○ bigrams  
○ trigrams

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA  
 UNIVERSITAT POMPEU FABRA

67

Delatats pel llenguatge!

## Classificació de textos Bigrams

Canonical Discriminant Functions

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA  
 UNIVERSITAT POMPEU FABRA

68

Delatats pel llenguatge!

## Classificació de textos Trigrams

Canonical Discriminant Functions

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA  
 UNIVERSITAT POMPEU FABRA

69

Delatats pel llenguatge!

## La tasca dels lingüistes forenses

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA  
 UNIVERSITAT POMPEU FABRA

70

Delatats pel llenguatge!

## Context general

CDL de Catalunya - 5 de febrer de 2009

A. Se situa en un context en el qual els detectius del llenguatge

- gairebé mai no han d'assenyalar a un autor entre milers de possibles candidats, basant-se únicament en proves lingüístiques,
- sinó que amb freqüència han de seleccionar a un únic autor entre un nombre limitat de candidats, sovint menys d'una dotzena i en molts casos només dos.

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA  
 UNIVERSITAT POMPEU FABRA

71

Delatats pel llenguatge!

## La tasca dels lingüistes forenses (1)

CDL de Catalunya - 5 de febrer de 2009

B. Activitats i tasques:

- 1. **Saber observar aquelles dades lingüístiques** en un text o conjunt de textos orals o escrits, l'autoria dels quals es desconeix o qüestiona,
  - a vegades per ajudar al jutge a determinar qui n'és l'autor, entre diversos possibles sospitosos o candidats;
  - altres vegades per ajudar al jutge a atribuir els textos dubitats a un únic autor.

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA  
 UNIVERSITAT POMPEU FABRA

72

**La tasca dels lingüistes forenses (2)**

**2. Analitzar aquestes dades amb absoluta independència**, és a dir, com a professional del llenguatge.

D'aquesta exigència se'n deriven dues conseqüències:

1. És preferible que el pèrit judicial lingüista actuï nomenat pel tribunal i no de part.
2. També seria desitjable que, tant en situacions exculpatòries com inculpatòries, els pèrits fossin independents.

CDL de Catalunya - 5 de febrer de 2009

**La tasca dels lingüistes forenses (3)**

**3. Aplicar els mètodes i tècniques adequats** per tal d'assegurar la validesa i fiabilitat dels resultats de les anàlisis que comporta el peritatge:

- realitzant experiments amb textos del món real (literaris, periodístics, i altres) per contrastar aquests resultats amb els que s'obtenen en l'anàlisi de textos de casos reals,
- definint acuradament l'univers de textos indubitats.
- aplicant tècniques estadístiques per establir la significació,
- i fent ús també de la lingüística de corpus.

CDL de Catalunya - 5 de febrer de 2009

**La tasca dels lingüistes forenses (4)**

**4. Saber seleccionar l'evidència més rellevant** que es deriva de l'aplicació dels mètodes qualitatiu i quantitatiu, és a dir, les dades lingüístiques.

CDL de Catalunya - 5 de febrer de 2009

**La tasca dels lingüistes forenses (5)**

**5. Saber documentar aquesta evidència de forma comprensible per al jutge, el tribunal o el jurat popular.**

- No usar taules numèriques; sempre gràfics i textos marcats.
- Les sentències incorporen el 80% del que un perito ha inclòs en el seu dictamen pericial.
- Els pèrits lingüistes actuen com a tals per donar suport al jutge o tribunal.

CDL de Catalunya - 5 de febrer de 2009

**La tasca dels lingüistes forenses (6)**

**6. Usar escales de probabilitat d'opinió qualificada**, a partir de la valoració de totes les anàlisis realitzades i les tècniques emprades.

CDL de Catalunya - 5 de febrer de 2009

**Escala d'opinió**

- Que **SI** es pot atribuir l'autoria de... a l'autor X amb un grau de probabilitat
 

5 -----	Molt alt
4 -----	Bastant alt
3 -----	Alt
2 -----	Baix
1 -----	Molt baix
0 -----	Possible
- Que **NO** es pot atribuir l'autoria de ... a l'autor X amb un grau de probabilitat
 

-1 -----	Molt baix
-2 -----	Baix
-3 -----	Alt
-4 -----	Bastant alt
-5 -----	Molt alt

CDL de Catalunya - 5 de febrer de 2009

## Idiolectometria

aplicada a la Lingüística Forense



CDL de Catalunya - 5 de febrer de 2009

- *Idiolectometria aplicada a la Lingüística Forense*  
<http://www.iula.upf.edu/rec/ideolec>
- *Idiolectometria forense i Índex de Similitud Idiolectal (ISI)*
- EXPLORA-INGENIO HUM2007-29140-E
- PLAN NACIONAL FFI2008-03583-FILO
- 2007-2008
- 2008-2010

## Conclusions

CDL de Catalunya - 5 de febrer de 2009

- **Compte!**
  - Hi ha detectius del llenguatge.
  - El llenguatge ens delata!

## Moltes gràcies!

CDL de Catalunya - 5 de febrer de 2009