

La labor del lingüista forense en la determinación/atribución de autoría

INGEFOR-08 6 de octubre de 2008

La labor del lingüista forense en la determinación/atribución de autoría

M. Teresa Turell

Directora de ForensicLab
Institut Universitari de Lingüística Aplicada
Universitat Pompeu Fabra
www.iula.upf.edu/forensiclab

Consultora asociada de FLAIR
Forensic Linguistic Advice, Investigation and Research
Centre for Forensic Linguistics
Aston University
www.forensiclinguistics.net/flair.html

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

1

La labor del lingüista forense en la determinación/atribución de autoría

INGEFOR-08 6 de octubre de 2008

Guión

- 1. Introducción.
- 2. Objetivos, datos y métodos de la LF y la D/AFA.
- 3. La labor del lingüista forense en D/AFA.
- 4. Dificultades.
- 5. La labor del lingüista en D/AFA.
 - 5.1 Determinación de autoría: sms's dubitados.
 - 5.2 Atribución de autoría: @mails dubitados.
- 6. En mi opinión...
- 7. Perspectivas de trabajo en D/AFA.

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

2

La labor del lingüista forense en la determinación/atribución de autoría

INGEFOR-08 6 de octubre de 2008

1. Introducción

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

3

La labor del lingüista forense en la determinación/atribución de autoría

INGEFOR-08 6 de octubre de 2008

Ámbitos de estudio, investigación y peritaje de la LF

- **Lenguaje de la Ley**
 - Language of the Law
- **Lenguaje Judicial**
 - Language of the Legal Process
- **Lenguaje Probatorio o Evidencial**
 - Language as Evidence

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

4

La labor del lingüista forense en la determinación/atribución de autoría

INGEFOR-08 6 de octubre de 2008

Determinación/Atribución Forense de Autoría (D/AFA)

- Idiolecto oral y escrito de un individuo:
 - rasgos lingüísticos que pueden revelar su género, edad, profesión, nivel educativo, religión, filiación política, entre otros factores sociales, y también
 - claves para la determinación del origen geográfico, la etnicidad o la raza de un individuo (Smith and Shuy 2002), de su procedencia no nativa en el uso de una segunda lengua o una lengua extranjera, e incluso de su condición de bilingüe.
- Esta evidencia lingüística puede ser usada en contextos de determinación/atribución de autoría y puede ayudar al juez o al tribunal a entender lo que dice un texto y quien es su autor (Coulthard 2004, 2005a, 2005b).

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

5

La labor del lingüista forense en la determinación/atribución de autoría

INGEFOR-08 6 de octubre de 2008

Premisas

- Existe mucha más variación inter-escriptor que intra-escriptor.
- El idiolecto escrito de un individuo se mantiene bastante constante en géneros textuales diferentes
- Dicho idiolecto cambia poco a lo largo del tiempo.

ForensicLab INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

6

2. Objetivos, datos y métodos de la LF y la D/AFA

Objetivos

- **Determinación de autoría**
 - Establecer entre varios candidatos, a cuál de ellos corresponde(n) el/los texto(s) escrito(s) dubitado(s) [es decir, cuya autoría presenta duda].
- **Atribución de autoría**
 - Establecer si un(os) texto(s) escrito(s) dubitado(s) corresponde(n) a un escritor concreto.
- **Perfil de autoría escrita**
 - Construir el perfil de un autor de un texto escrito (o varios textos escritos) desconocido.

Datos

- Datos lingüísticos.
- Objeto de estudio de la LF:
 - lenguaje y las lenguas en sus múltiples manifestaciones.
 - en D/AFA, datos lingüísticos escritos.
- No hay que confundirlos con datos estadísticos, ni con datos jurídicos.
- El Perito Judicial Lingüista describe lo que observa, pero no juzga ni mucho menos condena.

Métodos y Técnicas

- Establecer la significación estadística de los resultados de nuestros análisis.
- Asegurar en nuestros dictámenes y ante el juez la máxima fiabilidad y validez de los mismos.
- Ante la imposibilidad de cuantificar y aplicar pruebas estadísticas debe plantearse el carácter **complementario** de diversas técnicas de análisis lingüístico,
 - a) en la prueba pericial misma, lo que denomino la **complementariedad intra-evidencial**, y
 - b) entre diferentes pruebas de diferentes disciplinas, lo que denomino la **complementariedad inter-evidencial**.

3. La labor del lingüista en D/AFA

La labor del lingüista en D/AFA (1)

- A. Se sitúa en un contexto en el que el lingüista detective
- casi nunca tiene que señalar a un autor entre miles de candidatos basándose únicamente en pruebas lingüísticas,
 - sino que con frecuencia tiene que seleccionar a un único autor entre un número limitado de candidatos, a menudo menos de una docena y en muchos casos solo dos.

La labor del lingüista en D/AFA (2)

B. Y por tanto consiste en diversas actividades y tareas:

- 1. **Saber observar aquellos datos lingüísticos**, en este caso, marcas de autoría, en un texto o conjunto de textos escritos cuya autoría se desconoce o se cuestiona,
 - a veces para determinar quién es el autor entre varios posibles autores;
 - otras para atribuir los textos dubitados a un único autor.

INGEFOR-08 6 de octubre de 2008

La labor del lingüista en D/AFA (3)

2. **Analizarlos con entera independencia**, es decir, como profesional del lenguaje.

Dos consecuencias se derivan de esta exigencia:

1. Es preferible, como está ocurriendo cada vez con más frecuencia en casos civiles y mercantiles, que se actúe nombrado por el tribunal y no por parte.
2. También sería deseable que, por supuesto en situaciones exculpatorias, pero incluso en situaciones inculpatorias, los peritos fueran independientes,

INGEFOR-08 6 de octubre de 2008

La labor del lingüista en D/AFA (4)

3. **Dotarse de los métodos y técnicas** para asegurar la validez y fiabilidad de los resultados:

- realizando experimentos con textos del mundo real para contrastar esos resultados con lo observado en los textos objetos de análisis,
- aplicando técnicas estadísticas para establecer la significación,
- haciendo uso de la lingüística de corpus.

INGEFOR-08 6 de octubre de 2008

La labor del lingüista en D/AFA (5)

4. **Saber seleccionar la evidencia más relevante** que se deriva de la aplicación de los métodos cualitativos y cuantitativos,

- pero evitando que el bosque no nos deje ver los árboles, y
- los árboles son los datos lingüísticos.

INGEFOR-08 6 de octubre de 2008

La labor del lingüista en D/AFA (6)

5. **Saber documentar esta evidencia de forma comprensible para el juez, el tribunal o el jurado.**

- Por ejemplo, mejor no usar tablas numéricas; siempre gráficos y textos marcados.
- Las sentencias incorporan el 80% de lo que un perito ha incluido en su dictamen pericial.
- Los peritos lingüistas actuamos para asistir y apoyar al juez o al tribunal.

INGEFOR-08 6 de octubre de 2008

La labor del lingüista en D/AFA (7)

6. **Usar escalas de probabilidad de opinión cualificada**, a partir de la valoración de todos los análisis realizados y las técnicas usadas.

INGEFOR-08 6 de octubre de 2008

4. Dificultades

Dificultades de extensión Posibles razones (1)

- Naturaleza de nuestro sistema judicial (Pérez 2005), basado en el Civil Law.
- Otra razón estriba en el hecho de que no siempre se pide una segunda prueba cuando ya existe un dictamen.

Dificultades de extensión Posibles razones (2)

- Motivos para esta escasa presencia de los peritos lingüistas en nuestras salas de vistas:
 - Historia y la idiosincrasia de nuestra cultura jurídica.
 - Desconocimiento por parte de los actores judiciales de la actividad que puede llevar a cabo el lingüista forense.

Dificultades en el desarrollo científico y disciplinar (1)

- Situación actual de estudio, investigación y peritaje en D/AFA:
 - **Incipiente**
 - **desarrollo lento**

Dificultades en el desarrollo científico y disciplinar (2)

- Posición relativa con respecto a la Fonética y Acústica Forenses, ya que
 - la FAF cuenta con información sobre la distribución poblacional respecto a determinados parámetros idiosincrásicos de la voz,
 - la D/AFA no cuenta aún con esta distribución poblacional respecto a marcas escritas de autoría.
- Es lo que denominamos el *base-rate knowledge* (Grant 2007).

Dificultades en el desarrollo científico y disciplinar (3)

- Naturaleza de los datos:
 - En la huella dactilar física cada muestra es idéntica y exhaustiva.
 - Toda muestra lingüística facilita sólo información muy parcial acerca del idiolecto de su creador.
- Tarea imposible: enormes bancos de datos constituidos por muestras lingüísticas escritas representativas de millones de idiolectos.
- Esta dificultad en la obtención de muestreo dificulta también la obtención de ese deseable *base-rate knowledge* en D/AFA.

5.1 Determinación de autoría

Determinación de autoría (1) sms's (text messaging)

- Se puede distinguir a los escritores por las selecciones lingüísticas que hacen al escribir.
- Ha quedado demostrada la utilidad del análisis lingüístico de sms's (text messaging) en la resolución de casos de asesinato en Inglaterra.
- Contexto investigador:
 - Se reciben sms's mandados desde un móvil después de que haya desaparecido el dueño/la dueña de ese móvil.
 - Se descubre que ese desaparecido ha sido asesinado.
 - Se siguen enviando y recibiendo mensajes desde el móvil de esa persona.

Determinación de autoría (2) Asesinato de Jenny Nicholl (19 años) Yorkshire - 2005

- Agradecimientos:
 - Malcolm Coulthard
 - 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

Determinación de autoría (3) Textos dubitados

sms dubitado 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"

(Enviado a Jennifer Whelan el 9 de julio de 2005)

Determinación de autoría (4) Textos dubitados

sms dubitado 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**"

(Enviado a Nicola Gosnold el 9 de julio de 2005)

Determinación de autoría (5) Textos dubitados

sms dubitado 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"

(Enviado a Brian Nicholl, padre de Jenny, el 14 de julio de 2005)

Determinación de autoría (6) Textos indubitados (3/11)

- sms indubitado 1:
"Yeah shud b gud. I just **have2get** my finga out and do anotha tape"
- sms indubitado 2:
"Only just turned my **fone**. havent lied bout anything"
- sms indubitado 3:
"**im** tierd of defending **myself**. Theres no point. Bye.

INGEFOR-08 6 de octubre de 2008

Tabla 1 Contraste de marcas en los textos indubitados y dubitados

Textos indubitados (Jenny)	Textos dubitados
my (my dad, myself)	Me (me dad, meself)
cy (por see you)	cya
Fone	phone
im (por I am, I'm)	i am
have2get	been 2 kessick

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

Mensajes de sms	5000
Escritores (10/persona)	500
Interactivos	80
Independientes	420

Conclusiones Estudio corpus sms's (Grant)

Los mensajes intra-escritor son más similares que los mensajes interactivos

Los mensajes interactivos son más similares que los mensajes producidos independientemente

Una vieja premisa lingüística

- El lenguaje está gobernado por reglas ("rule-governed nature of language").
- Reglas de selección de palabras y de otras palabras para acompañar a las primeras.

INGEFOR-08 6 de octubre de 2008

Conclusiones finales sms's (text messaging)

- El Ministerio Fiscal planteó al jurado que toda esta evidencia no podía deberse a una simple coincidencia y la sentencia fue condenatoria.
- El corolario para el acusado, que cumple una condena de 18 años, podría ser el siguiente: quizás de no haber enviado esos sms's, su crimen no se hubiera descubierto.

INGEFOR-08 6 de octubre de 2008

Atribución de autoría (4) Errores gramaticales

INGEFOR-08 6 de octubre de 2008

- 5. **doc07** (Tlfax)
(...) 'al cual la empresa **le han** pagado'.
- 6. **doc01** (TD@)
(...) 'no creo que **hayan** muchas personas'.

Atribución de autoría (5) Contacto de lenguas

INGEFOR-08 6 de octubre de 2008

- i) La caída de la preposición 'de' en contextos donde su uso es obligatorio en español estándar, como por ejemplo en expresiones como 'en el caso ^ que' (doc05 Tlfax), 'se dio cuenta ^ que tenía en sus manos' (doc02 TD@).
- ii) La caída de la preposición 'a', donde su uso es obligatorio en español estándar, para indicar la función sintáctica del complemento indirecto, como por ejemplo: 'para atender ^ los clientes' (doc08 Tlfax).

Atribución de autoría (6) Secuencias de etiquetas morfosintácticas

INGEFOR-08 6 de octubre de 2008

- Selección.
- Pre-procesamiento.
- Etiquetaje morfosintáctico.
- Desambiguación y corrección de errores.
- Extracción de datos (Legolas 2.0.).
- Análisis estadístico.

Atribución de autoría (7) Bigrams/Trigrams

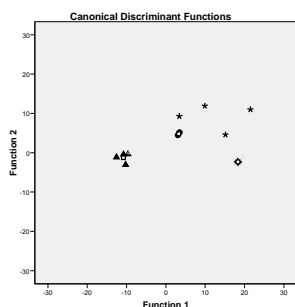
INGEFOR-08 6 de octubre de 2008

Secuencias de etiquetas morfosintácticas (SEM)

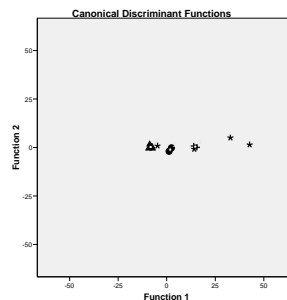
</s	el	fu	d	el	mar	ar	ull	tas	s	e	los	de	los	de	los	de	los	</s	
...
...

bigrams
trigrams

Atribución de autoría (8) Gráfico 1: Bigrams



Atribución de autoría (9) Gráfico 2: Trigrams



Atribución de autoría (10) P1PS - Lingüística de corpus

7. (Tifax)

'(...) 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'.

INGEFOR-08 6 de octubre de 2008

Atribución de autoría (11) P1PS - Ratios

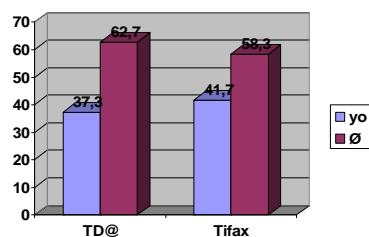
Tabla 4: Distribución de yo y Ø (%) en @mails y faxes

	yo	Ø
TD@	37,3	62,7
Tifax	41,7	58,3

INGEFOR-08 6 de octubre de 2008

Atribución de autoría (11 bis) P1PS - Ratios

Gráfico 3: Distribución de yo/Ø en TD@ y Tifax



INGEFOR-08 6 de octubre de 2008

Atribución de autoría (12) P. Relativo - Lingüística de corpus

9. doc05 (Tifax)

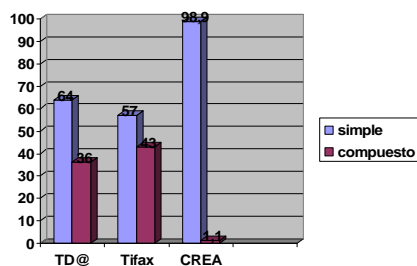
'(...) antes de la baja por enfermedad del Sr. XXX, **el cual** tenía que estar allí durante 1 o 2 días'.

10. doc03 (TD@)

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

Atribución de autoría (13) P. Relativo - Ratios

Gráfico 4: Distribución del relativo simple/compuesto español en los TD@, en los Tifax y en el CREA



INGEFOR-08 6 de octubre de 2008

6. En mi opinión...

Escala de probabilidad de opinión

INGEFOR-08 6 de octubre de 2008

- Que **SI** se puede atribuir la autoría de ... al autor X con un grado de probabilidad
 - 5 ----- Muy alta
 - 4 ----- Bastante alta
 - 3 ----- Alta
 - 2 ----- Baja
 - 1 ----- Muy baja
 - 0 ----- Posible
- Que **NO** se puede atribuir la autoría de ... al autor X con un grado de probabilidad
 - -1 ----- Muy baja
 - -2 ----- Baja
 - -3 ----- Alta
 - -4 ----- Bastante alta
 - -5 ----- Muy alta

7. Perspectivas de trabajo en D/AFA

INGEFOR-08 6 de octubre de 2008

Idiolectometría

aplicada a la Lingüística Forense



- Idiolectometría aplicada a la Lingüística Forense
- Idiolectometría forense e Índice de Similitud Idiolectal
- EXPLORA-INGENIO HUM2007-29140-E
- PLAN-NACIONAL FFI2008-03583-FILO
- 2007-2008
- 2008-2010

INGEFOR-08 6 de octubre de 2008

Muchas gracias