

LA TASCA DEL LINGÜISTA DETECTIU EN CASOS DE DETECCIÓ DE PLAGI I DETERMINACIÓ D'AUTORIA DE TEXTOS ESCRITS¹

M. Teresa Turell

ForensicLab, Institut Universitari de Lingüística Aplicada
Universitat Pompeu Fabra

1. INTRODUCCIÓ

En el nostre país, la lingüística forense és una disciplina molt poc coneguda encara, tant en el vessant conceptual com metodològic. En termes generals, aquesta disciplina tracta la interfície entre llenguatge i dret.

Al llarg de la seva curta història, la lingüística forense s'ha anat fonamentant en diversos models de la lingüística teoricodescriptiva general, i també en les premisses i generalitzacions que el desenvolupament d'aquesta disciplina ha permès proposar, principalment, per tal d'assolir una millor administració de la justícia. La consolidació de la lingüística forense actual es perfila en tres àrees d'especialització acadèmica i professional: l'àrea del llenguatge jurídic (*language of the law*), que abraça l'estudi del discurs legal, la comprensió lectora dels documents legals i la traducció jurídica, entre d'altres; l'àrea del llenguatge judicial (*language of the court and the legal process*), que inclou l'estudi del discurs de tots els actors del procés judicial (jutges, magistrats, testimonis, inculpats), dels interrogatoris judicials i policials, de la interpretació judicial i de suport en contextos multilingües i d'altres tipus de discurs del procediment judicial; i,

finalment, l'àrea del llenguatge probatori (*language as evidence*), que considera l'ús de l'evidència fonològica, morfològica, sintàctica, terminològica i discursiva en la comparació forense de veu i de textos escrits com a elements que poden ajudar en la identificació de veu, la creació de perfils lingüístics de textos orals o escrits, l'anàlisi de la imitació en la producció de textos amb finalitats criminals (notes de suïcidi), la determinació (diversos sospitosos) o atribució (un sol sospitós) d'autoria de textos escrits, la detecció de còpia d'idees i plagi lingüístic, els litigis de marques registrades, patents o etiquetatge de productes i, finalment, un conjunt de temes que tenen a veure amb la mediació lingüística forense, com ara la consideració de l'ambigüitat i alteració de textos oficials dins l'àmbit públic (Administració de l'Estat o comunitats autònomes) o de textos de l'àmbit privat (sobretot, empreses i companyies d'assegurances).

En el marc del llenguatge probatori els àmbits d'estudi de la detecció de plagi, d'una banda, i la determinació o atribució d'autoria, de l'altra, es poden considerar anversos d'una mateixa moneda en el sentit que si dos o més textos escrits presenten un grau elevat de similitud textual, s'ha de concloure que o bé han estat escrits pel mateix autor o bé han estat plagiat. Aquesta afirmació es pot fonamentar a

¹ Recerca portada a terme gràcies al projecte finançat pel Ministerio de Ciencia y Tecnología (FFI2008-03583/FILO; IP: Dra. M. Teresa Turell).

partir de la recerca lingüísticoforense ² que ha situat el límit de text idèntic que dos escriptors diferents poden expressar a l'atzar en sintagmes o frases de set o vuit paraules. Més enllà d'aquest límit, o bé els textos (o fragments de textos) idèntics són obra del mateix autor o bé un autor ha plagiat la producció escrita d'un altre. Amb tot, malgrat aquesta interrelació, tant la detecció de plagi com la determinació o atribució d'autoria presenten unes bases teòriques i metodològiques pròpies que són descrites en els apartats que segueixen.

No obstant això, abans de descriure en detall els conceptes, mètodes i aplicacions d'aquests dos àmbits d'estudi, és necessari remarcar el caràcter complementari de les ciències forenses, sobretot quan la naturalesa dels corpus lingüístics de casos forenses reals (pocs textos i molt curts) impedeix la quantificació i aplicació de proves estadístiques. En aquest sentit, he proposat (Turell 2010a: 9) el terme *complementarietat intraevidencial* per referir-me a l'ús de diverses proves en el marc de la mateixa disciplina, en aquest cas, la lingüística forense, i el de *complementarietat interevidencial* per referir-me a l'existència de diferents proves emprades per altres ciències forenses no lingüístiques. És important saber que, en la majoria de casos, tant civils com penals, és el jutge (o el tribunal) qui pren les decisions pertinents a la vista d'altres tipus d'evidència del cas, valora totes les anàlisis dutes a terme i les tècniques usades i pondera el pes específic que té cada prova

o tècnica i també, a vegades, cada variable o paràmetre d'una determinada prova.

2. LA DETECCIÓ DEL PLAGI

Aquest article fa una distinció entre la còpia d'idees i el plagi lingüístic. Hi pot haver còpia d'idees sense plagi lingüístic, però si es detecta plagi lingüístic, per la mateixa definició del signe lingüístic, també es produeix còpia d'idees. La detecció de plagi s'il·lustra en aquest treball a partir de diversos exemples de casos forenses reals on la còpia d'idees i el plagi lingüístic s'observen en obres literàries, obres científiques i entre les traduccions d'una obra original.

2.1. La naturalesa del plagi

La naturalesa del plagi és multidimensional pel nombre d'àmbits d'especialització afectats (literatura i assaig científic), els contextos i activitats en els quals ocorre (educació, traducció) i l'ús i instrumentalització que es fa de les xarxes digitals (per exemple, l'abast del plagi a Internet és doble: es pot plagiar directament d'una pàgina web o bé, metodològicament parlant, es pot emprar Internet, així com mètodes semi-automàtics (Grozea i Popescu 2010), per detectar la còpia d'idees i el plagi lingüístic).

La lingüística forense tracta el plagi com un fenomen complex i polivalent que mostra un cert solapament amb la determinació o atribució d'autoria (Johnson 1997, Grant i Baker 2001, Grant 2008) i amb l'estilística forense. Coulthard (2005: 261) es refereix al plagi com «un greu problema de determinació o atribució d'autoria» i el defineix com a «robatori, o ús no reconegut, d'un text creat per una altra persona».

Un objectiu teòric fonamental en l'estudi del plagi és l'establiment de les

² Basada en premisses de la lingüística i psicologia cognitives, amb referències generals com: Cifuentes Honrubia (1994), Schmid et al. (1996), Fauconnier (1997), Cuenca i Hilferty (1999), Inchaurrede i Vázquez (2000), Taylor (2002), Croft i Cruse (2004) i més específiques com Ochs (1979), Bereiter (1980), Kintsch (1988), Holyoak i Thagard (1997), Goa i García Albea (2000), Alamargot i Chanquoy (2001).

marques lingüístiques i els criteris que poden ser determinants en la detecció i descripció d'aquesta pràctica delictiva. Un criteri metodològic important és la consideració del nivell llindar de similitud textual entre textos, que serà decisiu per determinar si aquesta similitud és sospitosa.

2.2. Conceptes

El plagiat està directament relacionat amb el concepte de propietat intel·lectual, és a dir, amb la creació i amb tot allò que és creatiu. En la majoria de països s'estudia sota els epígrafs de propietat industrial i de *copyright*. La legislació espanyola distingeix entre *drets d'autor* i *copyright*. Aquests dos conceptes es refereixen a dues aproximacions diferents a la propietat intel·lectual literària i artística. El dret d'autor és un concepte que prové del dret continental (*civil law*, en anglès), particularment del dret francès; en canvi, l'ús del *copyright* deriva del dret anglosaxó (*common law*, en anglès). Els drets d'autor estan basats en el dret personal i no transferible que s'estableix en la relació entre l'autor i la seva creació. Es reconeix que una obra és l'expressió d'un individu i, per tant, queda protegida. En canvi, la protecció del *copyright* està estrictament restringida a una obra determinada i l'autor té uns drets específics que determinen com pot ser emprada aquesta obra. La protecció dels drets d'autor només cobreix l'expressió del contingut de l'obra, però no les idees.

El nou Codi penal aprovat el 1995 dedica tres articles relacionats amb la propietat intel·lectual (270, 271 i 272) ³ a partir dels quals un infractor dels drets de propietat intel·lectual és algú que, «de

manera total o parcial, reproduís, plagies, distribuís o comunicués públicament una obra literària, artística o científica, o la transformés, interpretés o executés artísticament, sense permís del corresponent titular del dret de propietat intel·lectual o dels seus hereus o persones delegades» i que s'aprofités de tercers o els perjudiqués. També, segons la Llei de propietat intel·lectual espanyola (LPI, articles 138 i 140), ⁴ l'autor plagiat tindria dret a una compensació per danys, tant moral com econòmica. D'altra banda, el concepte i terme de *copyright* especifica la protecció legal de la propietat intel·lectual, i el plagiat està íntimament relacionat amb el *copyright*. ⁵

Altres conceptes que poden ser útils en el marc del tractament del plagiat són: el coneixement de domini públic i les dades i els fets que poden ser verificats a partir de diverses fonts i són a l'abast d'un nombre considerable de persones. Si una dada forma part de l'entramat d'una disciplina i d'un model (lleis de moviment de Newton, teoria de l'evolució de Darwin, teorema de Pitàgores) no cal documentar-la; en canvi, si una dada o un fet formen part d'una nova teoria o model i s'ha de cercar en una font, aquesta font s'ha de documentar. El terme *dades* també és emprat per referir-se a fets, principis i estadístiques d'una determinada disciplina. Les dades es refereixen a resultats observables i verificables, i constitueixen la base d'una suposició o argumentació. Els fets i les dades com a tals no poden ser protegits legalment. Tot i que els fets i les dades com

³ <http://2ni2.com/juridico/penal/codigopenal.htm>

⁴ Per a més detalls sobre la LPI, vegeu <http://civil.udg.es/normacivil/estatal/real/Lpi.html>

⁴ Per a més detalls sobre la LPI, vegeu <http://civil.udg.es/normacivil/estatal/real/Lpi.html>

⁵ Per a més detalls sobre la legislació del *copyright* als Estats Units, vegeu <http://www.copyright.gov> i també <http://fairuse.stanford.edu/>

a tals no es poden protegir legalment, no significa que les fonts de les dades d'altres autors o investigadors no s'hagin d'esmentar.

2.3. Exemples de còpia d'idees i de plagi lingüístic

2.3.1. La còpia d'idees

Alguns exemples de còpia d'idees són:

- a) L'ús dels elements que estructuren una obra literària (novel·la, peça de teatre, poema, etc.): l'argument, els personatges, el temps, l'espai i les tècniques narratives.⁶
- b) En obres literàries: l'ús de totes, o gairebé totes, les figures retòriques sense citació explícita de l'autor original, fins i tot quan el vocabulari i la gramàtica emprats per expressar aquestes figures són diferents.⁷
- c) En traducció: la còpia d'una versió traduïda si aquesta traducció fa una contribució explícita, canviant de prosa a vers o bé transformant una obra clàssica en contemporània, o a l'inrevés.⁸
- d) En obres científiques: l'ús dels mateixos temes per descriure un període històric o la contribució d'un camp d'especialització.⁹
- e) En llibres de textos científics: la reproducció dels components essencials d'aquest tipus d'obres, com les activitats, preguntes i tècniques de laboratori,¹⁰ i també la reproducció d'una metodologia creativa.¹¹

⁶ Un exemple clar de còpia d'idees és el cas *Formoso v. Cela*. Vegeu Turell (2007 i 2008: 275-277).

⁷ Cas *Pujante v. Vázquez Montalbán*. Vegeu Turell (2004: 20-21 i 2008: 277).

⁸ Cas *Pujante v. Vázquez Montalbán*. Vegeu Turell (2004: 20 i 2008: 277-78).

⁹ Cas *XXX v. YYY* (convencions emprades per respectar els requisits de confidencialitat). Vegeu Turell (2008: 278).

¹⁰ Cas *Bruno v. Magister*. Vegeu Turell (2008: 278-79).

¹¹ Cas *XXX v. YYY*. Vegeu Turell (2008: 279-80).

2.3.2. El plagi lingüístic

D'acord amb Menasche (1977) i Roig (2006),¹² el plagi lingüístic (que, com ja s'ha dit, també comporta còpia d'idees) es produeix quan es donen les circumstàncies següents:

- a) Quan s'empren exactament les mateixes paraules i estructures gramaticals per descriure les idees pròpies (autoplagi) i les idees d'altres escriptors.
- b) Quan es produeix paràfrasi, és a dir, quan algú expressa amb les seves paraules les idees d'altres, però també usa de manera extensiva les paraules, sintagmes i oracions de l'obra original en la qual es basa.
- c) Quan s'usen diverses paraules i oracions sense citar la font, encara que se'n canviïn algunes.
- d) Quan la sintaxi original es manté i només es canvien paraules molt específiques per sinònims.
- e) Quan es reconeix l'autoria de l'obra original, però els canvis introduïts per desmarcar-se'n només impliquen una o dues paraules, l'ordre de parts de l'oració (WO), el canvi de veu (activa v. passiva) i/o el temps i l'aspecte verbals de les clàusules o de tot el text.¹³

2.4. La metodologia aplicada a casos forenses reals de plagi lingüístic i/o còpia d'idees

En les seccions que segueixen es presenten proves lingüístiques qualitatives i quantitatives que han resultat determinants i s'han utilitzat per: a) il·lustrar l'ús de marques i criteris lingüístics per demostrar

¹² Vegeu també:

<http://www.sscnet.ucla.edu/history/bartchy/classe/s/194a/98F/plagiarism.html> i

<http://facpub.stjohns.edu/~roigm/plagiarism/Index.html> respectivament.

¹³ *c. d. i e.* són tipus de paràfrasi. Vegeu Vila, Martí i Rodríguez (2011).

la direccionalitat del plagi lingüístic ¹⁴ i la còpia d'idees; *b*) considerar el grau de diferència o similitud textual per damunt del qual es pot establir si dos o més textos han estat plagiats; *c*) establir l'abast i la naturalesa del plagi, i *d*) concloure si les proves lingüístiques presentades poden permetre al tribunal l'establiment d'un cas *prima facie* ¹⁵

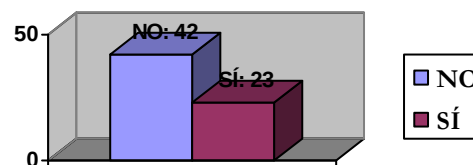
2.4.1. Mètodes i aplicacions forenses en casos de còpia d'idees

El cas *XXX v. YYY* il·lustra un tipus de còpia d'idees. Així, l'anàlisi qualitativa de tretze (13) seccions en les quals estava dividit un article dubitat (de l'autoria del qual es dubta) escrit per YYY per al *National Geographic* va demostrar que YYY havia emprat parcialment la mateixa font que usava XXX en la seva tesi doctoral. De fet, XXX va dedicar sis (6) dels seus set (7) capítols de la tesi al mateix tema, mentre que YYY va utilitzar-lo en set (7) de les tretze (13) seccions en les quals està dividit l'article del *National Geographic*. Aquesta anàlisi també va mostrar que es produïa còpia o reproducció d'idees en vint-i-tres (23) dels seixanta-cinc (65) temes considerats (35,5 %), comparant-t'ho amb els quaranta-dos (42) temes (64,5 %) que no es van copiar ni reproduir, tal com s'il·lustra al gràfic 1:

Gràfic 1. Còpia o reproducció d'idees

N = nombre de temes

XXX v. YYY



Font: Pròpia (Turell 2008: 278)

2.4.2. Mètodes i aplicacions forenses en casos de plagi lingüístic (i, per tant, també de còpia d'idees)

El cas *Bruño v. Magister* permet il·lustrar tota una sèrie de tècniques lingüísticotextuals d'aproximació al text dubtós (*Magister*) i al plagiat (*Bruño*) per establir la naturalesa i la direccionalitat del plagi.

a) La unitat i la coherència o cohesió (Halliday i Hasan 1976) d'un llibre (de text, en aquest cas), reflectides en les seccions, components metodològics i paràgrafs que constitueixen el contingut d'aquest text, poden esdevenir marques qualitatives que poden ajudar un tribunal a decidir quin text és l'original i quin és el text que s'ha escrit practicant el plagi.

En el llibre de text *Física y Química* de Bruño (2002) cada unitat es presenta de manera unificada, completa i coherent o cohesiva, i passa el mateix amb els components metodològics com les activitats, exemples, problemes, exercicis, pràctiques de laboratori, etc. Aquesta observació va permetre concloure que *Física y Química* de Bruño (2002) havia estat

¹⁴ Per a un enfocament automàtic de la direccionalitat del plagi, vegeu Grozea i Popescu (2010).

¹⁵ Un cas segons el qual les proves pericials són suficients per permetre al jutge o al tribunal emetre una sentència (per exemple, en el cas *Pujante v. Vázquez Montalbán*).

produït de manera independent i era, per tant, el text no dubtós (o indubitat). D'altra banda, la presentació que *Temario Magister* (2005) fa de cada unitat, i molt especialment de les activitats incloses en aquest text, s'havia produït de manera fragmentària i en molts casos una idea no conduïa a una altra de manera natural i progressiva d'acord amb el grau de dificultat conceptual.

Altres vegades *Temario Magister* (2005) (el text dubitat) inclou activitats completament il·lògiques i sense sentit a causa de l'ús de l'estratègia del retalla i enganxa (*cut & paste*), atès que, tot i que utilitza el mateix encapçalament d'una activitat de *Física y Química Bruño* (2002), les preguntes que segueixen formen part d'una altra activitat localitzada en una altra part del llibre plagiat, com es pot observar a la taula 1,¹⁷ on la similitud textual exacta es marca en cursiva.

Taula 1. Activitats a *Física y Química Bruño* i *Temario Magister*

<i>Física y Química Bruño</i> (2002)	<i>Temario Magister</i> (2005)
ACTIVITAT 23 (Unitat 2, pàg. 40)	ACTIVITAT 2 (Unitat 2, pàg. 11)
23. Desde la azotea de un rascacielos de 120 m de altura se lanza hacia abajo una pequeña bola con velocidad inicial de 20 m/s. Calcula: a) El tiempo que tarda en llegar al suelo. b) La velocidad que tiene en ese momento. Toma $g = 9,8 \text{ m/s}^2$.	2. Desde la azotea de un rascacielos de 120 m de altura se lanza hacia abajo una pequeña bola con velocidad inicial de 20 m/s. Calcula:
ACTIVITY 24 (Unit 2, page 49)	
24. Se lanza verticalmente hacia arriba un proyectil con velocidad de 200 m/s; al cabo de 4 segundos, se lanza otro proyectil con el mismo objeto. Calcula: a) La altura a la que se encuentran. b) El tiempo que tardan en encontrarse. c) La velocidad de cada proyectil en ese momento. Toma $g = 10 \text{ m/s}^2$.	a) La altura a la que se encuentran. b) El tiempo que tardan en encontrarse.

Font: Pròpia (Turell 2008, pàg. 283)

b) Paràmetres lingüístics¹⁶ i estratègies discursives

b.i. Inconsistència en l'estil referencial

En el llibre plagiat (o indubitat) *Física y Química Bruño* (2002), l'estil directe és

emprat mitjançant l'ús de l'imperatiu de segona persona singular en la formulació de les activitats, exercicis, tècniques, etc. i aquesta estratègia discursiva es manté al

¹⁶ Vegeu Stein i Meyer zu Eissen (2007), Stamatatos (2009) i Potthast et al. (2010) per a més detall sobre la detecció automàtic del plagí intrínsec.

¹⁷ Resultats obtinguts a partir de l'execució del programa CopyCatch (Llicència i Copyright©David Woolls, de CFL Development).

b) Paràmetres lingüístics ¹⁸ i estratègies discursives

b.i. Inconsistència en l'estil referencial

En el llibre plagiat (o indubitat) *Física y Química Bruño* (2002), l'estil directe és emprat mitjançant l'ús de l'imperatiu de segona persona singular en la formulació de les activitats, exercicis, tècniques, etc. i aquesta estratègia discursiva es manté al llarg de tot el llibre de text; en canvi, en el

text dubtós (o dubitat), *Temario Magister* (2005), s'observa l'ús de l'infinitiu per adreçar-se als estudiants i lectors, tot i que aquest ús és inconsistent, ja que a vegades també s'empra l'imperatiu (taula 2, diferències marcades en negreta), la qual cosa indica que a *Temario Magister* s'està plagiant lingüísticament el text original a partir de la tècnica de retalla i enganxa (*cut & paste*).

Taula 2. Estil referencial a *Física y Química Bruño* i *Temario Magister*

<i>Física y Química Bruño</i> (2002), pàg. 204	<i>Temario Magister</i> (2005), Unitat 11
ACTIVITATS 6, 7 i 10, 11	ACTIVITATS 1, 2 i 1, 2
	(pàg. 8)
6. Describe la experiencia de Geiger y Marsden que justifica el modelo atómico de Rutherford. 7. Explica brevemente en qué consiste el modelo de Rutherford.	1. Describir la experiencia de Geiger y Marsden que justifica el modelo atómico de Rutherford. 2. Explicar brevemente en qué consiste el modelo de Rutherford
10. El kriptón tiene seis isótopos cuyos números másicos son: 78, 80, 82, 83, 84 y 86. Consulta la Tabla Periódica y escribe la notación de esos isótopos indicando el número de protones, neutrones y electrones de cada uno. 11. ¿Cómo hallarías la masa atómica de un elemento conociendo las masas de sus isótopos y su abundancia relativa en tanto por ciento?	(pàg. 9) 1. El kriptón tiene seis isótopos cuyos números másicos son: 78, 80, 82, 83, 84 y 86. Consulta la Tabla Periódica y escribe la notación de esos isótopos indicando el número de protones, neutrones y electrones de cada uno. 2. ¿Cómo hallarías la masa atómica de un elemento conociendo las masas de sus isótopos y su abundancia relativa en tanto por ciento?

Font: Pròpia (Turell 2008: 285-286)

¹⁸ Vegeu Stein i Meyer zu Eissen (2007), Stamatatos (2009) i Potthast et al. (2010) per a més detall sobre la detecció automàtic del plagí intrínsec.

b.ii. Descontextualització

Temario Magister (2005) presenta exemples de descontextualització en el sentit que algunes activitats incloses a *Física y Química Bruño* (2002) són reproduïdes a *Temario Magister* (2005), però una part de la informació, «consultando la Tabla Periódica», és omesa (taula 3 en negreta).

Taula 3. Descontextualització a *Temario Magister*

<i>Física y Química Bruño</i> (2002), pàg. 194	<i>Temario Magister</i> (2005), (Unitat 11, pàg. 8 i 9)
ACTIVITAT 2	ACTIVITAT 1
2. Completa en tu cuaderno la siguiente Tabla 11.2 consultando la Tabla Periódica.	1. Completa la siguiente Tabla:

Font: Pròpia (Turell 2008: 286)

b.iii. Inversió en la gradació dels elements estructurals

La taula 4 il·lustra el fet que en el text plagiat, *Física y Química Bruño* (2002), les preguntes 3 i 4 de la pàgina 253 s'inclouen en aquest ordre, ja que la pregunta 3 requereix, per part de l'alumne, la integració de la noció d'entalpia dins l'equació química d'una *reacción exotérmica*, i la pregunta 4 li demana que pensi en dos exemples de reacció *exotérmica* i *endotérmica*, movent-se del que és particular a una generalització; en canvi, en el text dubtós, *Temario Magister* (2005), aquestes dues preguntes es presenten a l'inrevés i el resultat comporta una incongruència conceptual (taula 4).

Taula 4. Inversió de temes a *Temario Magister*

<i>Física y Química Bruño</i> (2002), pàg. 253	<i>Temario Magister</i> (2005), (Unitat 13, pàg. 10)
PREGUNTES 3 i 4	PREGUNTES 1 i 4
3. A veces se pone en uno de los miembros de la ecuación química de una reacción la entalpía de reacción. ¿En qué miembro pondrías la entalpía si la reacción fuera exotérmica?	1. Escribe dos ejemplos de reacciones exotérmicas y otros dos de reacciones endotérmicas, indicando cómo se escribe la entalpía.
4. Escribe dos ejemplos de reacciones exotérmicas y otros dos de reacciones endotérmicas, indicando cómo se escribe la entalpía.	4. A veces se pone en uno de los miembros de la ecuación química de una reacción la entalpía de reacción. ¿En qué miembro pondrías la entalpía si la reacción fuera exotérmica?

2.4.3. El nivell llindar i altres mesures del plagi lingüístic

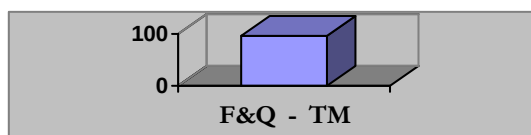
CopyCatch, un dels diversos programes de concordances existents per detectar el plagi lingüístic, permet calcular el nivell llindar de similitud textual a partir del qual aquesta similitud es converteix en sospitosa. Aquest programa incorpora diverses mesures (nivell llindar de vocabulari coincident, paraules compartides una sola vegada (*hapax legòmena*), paraules compartides dues vegades (*hapax dislegòmena*), vocabulari exclusiu i sintagmes compartits una sola vegada),¹⁹ que han estat útils per detectar diversos tipus de plagi. Per determinar el nivell llindar de vocabulari coincident, es compta amb resultats empírics que suggereixen que «fins a un 35 % de similitud es considera un percentatge normal, i que un 50 % no és inusual, mentre que com més per damunt del 50 % es trobi aquest nivell llindar més

¹⁹ Per raons d'espai, en aquest article només és possible il·lustrar el nivell llindar de vocabulari coincident, tot i que en un cas forense real s'haurien d'aplicar totes aquestes mesures, ja que cap en solitari és suficientment discriminatòria per arribar a resultats conclouents.

possibilitats hi ha que aquesta xifra indiqui» (Turell 2004: 8) que els textos que s'estan considerant no hagin estat produïts independentment sinó que un hagi estat un plagi de l'altre.

El gràfic 2 mostra que el nivell llindar de vocabulari coincident (96 %) entre les activitats que apareixen en el llibre de text indubitat *Física y Química Bruño* (2002) i *Temario Magister* (2005) és molt alt, resultat que indica que les activitats han estat reproduïdes textualment en el text dubtós.

Gràfic 2. Vocabulari coincident (F&Q-TM)
Activitats



Font: Pròpia (Turell 2008: 288)

3. LA DETERMINACIÓ I ATRIBUCIÓ D'AUTORIA

3.1. El marc conceptual de la determinació o atribució d'autoria

La determinació o atribució d'autoria de textos escrits pressuposa establir candidats a autor i suposa una tasca de classificació de trets identificats i atribuïbles a un autor o a un altre (Chaski 2001, Grant i Baker 2001, Kredens 2001, Love 2002). En els darrers anys, i sens dubte per la influència de la comparació forense de veu aplicada a la identificació de la veu humana, s'ha estès el terme i concepte de comparació forense de textos escrits, que pot conduir a una determinació o atribució d'autoria més rigorosa i fiable (Turell 2010a). No obstant això, jurídicament parlant, la determinació de l'autor d'un o diversos textos entre diferents possibles autors o l'atribució d'un

o diversos textos a un únic autor és una activitat que pertoca al jutge o tribunal, ja que els lingüistes forenses només aporten proves pericials lingüístiques que poden ajudar a aquestes instàncies judicials a dictar una sentència. També s'emptra el terme *identificació d'autoria* (Solan i Tiersma 2004) o *reconeixement d'autoria* (Hänlein 1999). D'altra banda, Grant (2007 i 2008) introdueix un terme més metodològic (anàlisi d'autoria) en parlar d'anàlisi d'autoria com l'intent «d'establir informació desconeguda o poc documentada sobre els orígens d'un text a partir de trets i paràmetres lingüístics interns». ²⁰

L'anàlisi forense de l'autoria de textos escrits comporta dos contextos de recerca i anàlisi:

- a) La determinació d'autoria implica establir entre diversos candidats a qui correspon el text escrit dubtós (o dubitat, és a dir, l'autoria del qual presenta dubte)
- b) L'atribució d'autoria implica establir si un text escrit dubtós es pot atribuir o no a un únic escriptor sospitós específic.

D'altra banda, aquesta anàlisi està restringida per les característiques usals dels textos forenses de casos reals. Usualment, aquests textos són missatges curts (anònims, correus electrònics, faxes, notes de suïcidi i d'altres). Els trets comuns que els caracteritzen tenen a veure amb el fet que són incidentals i espontanis, escrits per a una audiència limitada, produïts i situats dins d'un temps limitat i determinat i, molt freqüentment, són textos emotius.

²⁰ «Attempting to establish unknown or weakly-supported information about the origins of a text from internal linguistic evidence.» Tim Grant, doctorat en Ciències del Llenguatge i Lingüística Aplicada de l'UIULA (UPF), 2007.

3.2. Mètodes i tècniques aplicats a un cas real d'atribució forense d'autoria

Ara com ara no existeix un únic mètode o tècnica que es pugui emprar en l'anàlisi forense ni en les actuacions dels lingüistes com a pèrits judicials. Els conceptes de complementarietat intraevidencial i interevidencial són útils per entendre que la determinació o atribució forense d'autoria es troba lluny del que han assolit la fonètica i l'acústica forenses, pel que fa a la fiabilitat de les tècniques i mètodes emprats en la comparació forense de veu i a la formulació de les conclusions en les proves pericials sobre la identificació de la veu dels locutors.

La complementarietat intraevidencial comporta l'ús de diversos mètodes, tècniques i proves de la metodologia forense lingüística: enfocaments analítics textuals qualitatius, lingüística de corpus, ràtios tipus o unitat, hapax legòmena, hapax dislegòmena, densitat i riquesa lèxiques, anàlisi discriminant lineal (ADL) aplicat a l'estudi de les seqüències de categories lingüístiques (bigrames, trigrammes, etc.) com a marques d'autoria (Spasova 2009, Turell 2010b), relació de versemblança (LR, en anglès) (Turell 2010a), entre d'altres; evidència lingüística que seria complementada per altres tipus de proves provinents d'altres ciències forenses, és a dir, a partir de la complementarietat interevidencial.

Una bona part de la feina dels lingüistes forenses en casos de determinació o atribució d'autoria té a veure, com ja s'ha dit més amunt, amb una tasca classificatòria per determinar els trets lingüístics idiosincràtics que fan de l'estil idiolectal d'un individu únic i irrepetible, a partir de la comparació forense intraescriptor i interescriptor.

Un cas forense real d'extorsió per correu electrònic (cas *SEHRS v. ZZZ*, 2004)²¹ permet il·lustrar algunes d'aquestes tècniques lingüísticotextuals d'aproximació al text (o textos) dubtós, i també els dos tipus de complementarietat als quals s'ha al·ludit més amunt.

L'any 2004, una empresa del sector turístic de Catalunya fou acusada d'acomiadament improcedent per un treballador que l'havia intentat extorsionar. Aquesta extorsió està expressada en un dels textos escrits en espanyol en format electrònic (textos dubtosos o dubitats, TD@), però el treballador que suposadament va enviar aquest text a l'empresa i que durant la mateixa època escrivia cartes en format fax (textos no dubtosos o indubitats, TIfax), en va negar l'autoria. Ambdós tipus d'escrits van ser suposadament enviats des de la sucursal francesa que havia obert l'empresa al nord de França. El corpus d'anàlisi d'aquest cas està constituït per quatre (4) correus electrònics dubitats i quatre (4) faxos indubitats, amb una mitjana de 40-80 línies i entre 400-1000 paraules. Se suposa que aquests textos van ser produïts, en el cas dels TD@, per algú originari d'una zona castellanoparlant (província de Terol), però que havia viscut en una zona catalanoparlant (Catalunya) durant més de trenta anys, abans de ser traslladat a França per l'empresa demandada.

Per tal d'observar l'estil dels textos indubitats (TIfax) i els textos dubitats (TD@) d'aquest cas es van emprar tres apropaments analítics: una anàlisi qualitativa textual, una anàlisi de variables gramaticals i sociolingüístiques basada en la lingüística de corpus, i una anàlisi

²¹ Vegeu sentència número 115/04 del Jutjat del Social número 1 de Tarragona d'11 de febrer de 2004.

estadística de seqüències de categories lingüístiques.²²

L'anàlisi qualitativa es va basar en la consideració de similituds flagrants entre els faxos indubitats i els correus electrònics dubitats a partir de: *a)* incongruències temporals i cronològiques; *b)* errors ortogràfics, és a dir, no accentuació en espanyol del morfema *-ía* de la 1a i 3a persona de l'imperfet en la 2a i la 3a conjugació, i ús de l'auxiliar *haber* sense *h*. Errors, molt possiblement motivats pel contacte entre el català i l'espanyol, en el primer cas, i per un nivell educatiu baix, en el segon; errors gramaticals, és a dir, no concordança normativa entre subjecte i verb (*no creo que hayan muchas personas*); queisme (*en el caso ^ que*), que s'explica per la influència del català; dequeisme (*no me extrañaría nada de que este rumor sea verdad*), fenomen lingüístic explicat per la hipercorrecció i, finalment, *c)* característiques estilístiques, estructurals i textuals del format dels textos.

Certament, aquesta evidència qualitativa en solitari no permet atribuir l'autoria dels textos dubitats al suposat autor d'extorsió per mitjà d'un dels quatre correus electrònics; no obstant això, sí que permet afirmar que existeix una probabilitat (no estadística, sinó d'opinió) bastant alta que els dos conjunts de textos no hagin estat produïts independentment per dos autors diferents, sinó que hagin estat escrits per un mateix autor. Òbviament, aquests resultats derivats de l'observació qualitativa han de ser corroborats per altres tipus d'evidències lingüístiques.

La lingüística de corpus i el marc d'anàlisi basat en trets i paràmetres marcats

²² Per tal de descriure la variació intraescriptor i interescriptor amb més fiabilitat es van emprar correus electrònics anònims d'un altre cas forense real, que constitueix el corpus de control.

(*markedness*)²³ i prominents (*saliency*)²⁴ han servit per estudiar la correlació entre l'evidència gramatical i la sociolingüística aplicada a dues variables lingüístiques: el pronom personal de 1a persona en espanyol amb dues variants de realització (*yo/Ø*)²⁵ i el pronom relatiu, també en espanyol, amb dues variants de realització (simple: *que* /compost: *el cual* i les seves diferents realitzacions). Es parteix de la proposta de Jakobson i Halle (1956) que la forma marcada comporta informació més precisa, específica i addicional que la forma no marcada.

El concepte de prominència que és més útil en estudis forenses de l'estil idiolectal d'un escriptor deriva de l'enfocament combinat entre l'anàlisi del discurs i la lingüística de corpus (Butler 1998, Abecassis 2002). En aquest context la prominència es refereix a totes les paraules que destaquen estadísticament quan dos subcorpus es comparen o quan un subcorpus es compara amb la totalitat d'un corpus.

L'interès d'aquesta tècnica, basada en aquests dos conceptes i en l'aplicació de la lingüística de corpus, rau en la comparació

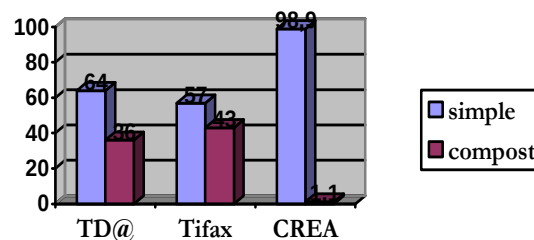
²³ Per a més informació sobre el tractament de la noció de *markedness* aplicat per diverses escoles de la historiografia lingüística del segle XX, vegeu Greenberg (1966), Chomsky i Halle (1968), Chomsky (1981), Kean (1981), Williams (1981), Viel (1984), Berwick (1985), Hyams (1987), Steriade (1987), Wexler i Manzini (1987), Myers-Scotton (1998), Dresher i Rice (2002), Aissen (1999 i 2003), Hume (2004).

²⁴ Vegeu Quirk et al. (1972) i Arnold i Wasow (1996) per comprendre com la prominència en el discurs afecta l'ordre dels constituents d'una oració i altres processos gramaticals, i Gordon, Grosz i Gilliom (1993) per entendre com aquesta prominència afecta la coherència discursiva. Giora (2003) i Keszkes (2001 i 2004) consideren la noció de prominència per explicar l'accessibilitat del referent i el processament del llenguatge.

²⁵ Considerant únicament les ocurrències on la variant *yo* és redundant, i no emfàtica ni gramaticalment obligatòria.

de la freqüència d'ús d'una determinada variable entre els corpus d'anàlisi i un corpus de referència general ²⁶ per establir-ne la idiosincràsia. El gràfic 3 indica que la pauta freqüencial d'ús dels relatius simple i compost, tant en els textos dubitats (TD@) com en els no dubitats (Tifax), és molt similar i mostra una freqüència alta de la forma simple en el corpus CREA. Aquests resultats també indiquen que, tot i que la freqüència d'ús del relatiu *que* en els dos conjunts de textos objecte d'anàlisi continua sent alta, la freqüència del compost *el qual* en espanyol (en totes les possibles realitzacions) és remarcable, sobretot si es compara amb la distribució observada en el corpus CREA, amb un 98,9 % d'ocurrències de *que* i només un 1,1 % de *el qual*. ²⁷ En aquest sentit, doncs, la freqüència d'ús del relatiu compost es converteix en una marca d'autoria, sociolingüísticament motivada pel contacte de llengües. ²⁸

Gràfic 3. Distribució del relatiu simple/compost en espanyol



Font: Pròpia (Turell 2010b: 232)

En el cas de la variable *yo/Ø*, aquesta tècnica no s'ha pogut aplicar a causa d'una limitació metodològica derivada de la lingüística computacional i del processament del llenguatge natural, en el sentit que els corpus existents etiquetats per a l'espanyol no han estat etiquetats per als conjunts buits (o sigui, els zeros (Ø)) i, per tant, no ha estat possible portar a terme una comparació de la distribució freqüencial d'aquesta variable entre els corpus objecte d'estudi i el corpus de referència CREA.

S'han d'esmentar altres tècniques quantitatives emprades també en la determinació o atribució d'autoria de textos escrits, com l'anàlisi de la riquesa lèxica, l'hapax legòmena (paraules que ocorren una sola vegada en un text) i hapax dislegòmena (paraules que ocorren dues vegades en un text), i la ràtio entre tipus (*types*) de categories i unitats d'ocurrència de cadascuna d'aquestes categories (*tokens*). En el cas dels textos TD@ i Tifax, aquesta tècnica també va aportar resultats molt concloents en el sentit que la riquesa lèxica, i sobretot les ràtio tipus o unitat, eren molt similar entre els textos dels dos conjunts analitzats.

En canvi, l'aplicació de l'anàlisi multivariant, que analitza el comportament d'una variable lingüística correlacionada

²⁶ En aquest cas forense s'ha emprat el Corpus de Referència del Español Actual (CREA) dins del corpus general de la Real Academia Española (RAE).

²⁷ En canvi, en l'ús general i usual del català, tot i que la freqüència d'ús del relatiu simple és alta, l'ús del relatiu compost està també molt estès.

²⁸ Per a més informació sobre marques de contacte de llengües que poden ajudar a identificar la pertinença de parlants o escriptors a una determinada varietat regional de l'espanyol, vegeu Payrató (1985), Silva-Corvalán (1995), Roca i Jensen (1996), Gómez Torrego (1997), Wesch (1997), Bosque i Demonte (1999), Powtoski i Cameron (2007), Blas Arroyo (2007).

amb altres variables lingüístiques internes i altres factors externs —que podria haver estat rellevant per establir pautes similars de comportament de determinades variables en els conjunts de textos dubitats i indubitats— no va resultar satisfactòria, ni en aquest cas ni en la majoria de casos forenses reals perquè en general els corpus són molt reduïts i els textos molt curts, amb la qual cosa els resultats s'han de considerar de manera temptativa.

La complementarietat inter-evidencial en aquest cas la va aportar la informàtica, ja que la part demandant, és a dir, el treballador acomiadat per l'empresa per extorsió, argumentava que algú de l'empresa des d'Espanya havia manipulat el seu ordinador i havia enviat els correus electrònics. No obstant això, el perit informàtic va poder demostrar que els correus s'havien enviat des de França a partir del domini @wanadoo.fr.

3. CONCLUSIONS

En aquest article he tractat de mostrar que la comparació forense de textos escrits que pot conduir a una detecció de plagi i a una determinació o atribució d'autoria més fiables comporta una activitat d'observació de l'anvers i el revers de la mateixa moneda, atès que si dos o més textos escrits presenten un grau elevat de similitud textual, s'ha de concloure que o bé han estat plagiats o bé han estat escrits pel mateix autor.

S'ha vist que en ambdues activitats es tracta d'establir el nivell llindar de similitud textual per damunt del qual es pot considerar que aquesta similitud és sospitosa, ja sigui per concloure que els textos comparats han estat plagiats o per determinar si han estat escrits per un mateix autor, o no.

D'una banda, s'ha il·lustrat que en ambdós àmbits d'anàlisi lingüísticoforense es produeix un apropament als textos molt similar, en el sentit que tant la lingüística de corpus com l'anàlisi de la riquesa lèxica poden ser aplicades de manera fiable i concloent. Particularment, en el cas de la detecció de plagi, a part de l'establiment del nivell llindar, s'han aportat altres enfocaments i tècniques, tant qualitatives com quantitatives, per detectar-lo, i també s'ha fet una descripció de les estratègies lingüísticodiscursives que empren els plagiaris.

En el cas de la determinació o atribució d'autoria també s'han emprat tècniques derivades de l'anàlisi textual i de la lingüística de corpus; a més a més, s'ha il·lustrat l'ús de seqüències de categories lingüístiques com a marques d'autoria, aplicant l'anàlisi discriminant lineal que permet l'establiment del potencial discriminant d'aquestes marques en termes de significació estadística.

Totes aquestes aproximacions als textos ajuden el lingüista detectiu a observar marques que un mateix escriptor sospitosos usa en l'estil idiolectal i que també ocorren en un determinat text dubtós o dubitat.

En el camp de la comparació forense de textos escrits sobresurten dues qüestions fonamentals, que no han estat plantejades en el cos de l'article: d'una banda, la necessitat de poder comptar amb una definició fiable de l'univers de textos indubitats, i de l'altra, la constatació que els estudis de determinació o atribució d'autoria estan molt endarrerits si es comparen amb la metodologia de la fonètica i l'acústica forenses usada en comparació forense de veu.

Pel que fa al primer punt, en casos forenses reals, quan un anònim arriba al laboratori i es defineix l'existència d'un o

dos sospitosos, no sempre és fàcil poder comptar amb un univers fiable de textos indubitats d'aquests sospitosos, sense el qual és impossible fer comparacions fiables de classificació i discriminació de textos i d'autors. Un problema afegit es planteja per la inexistència de corpus de referència etiquetats en la majoria de llengües. Aquests corpus etiquetats poden ajudar el lingüista a observar quan el comportament d'una determinada variable lingüística en un text dubtós és rar, però freqüent, en relació amb les pautes que s'observen en l'ús del llenguatge en el món real.

Pel que fa a l'endarreriment metodològic de la comparació forense de textos escrits i el clar desavantatge en el qual es troba respecte de la comparació forense de veu, l'únic camí d'actuació viable actualment és l'ús de diverses tècniques lingüístiques de manera complementària, ja que no hi ha un únic mètode fiable des del punt de vista de la significació estadística. I, sobretot, es tracta de poder assolir una distribució poblacional (*base rate knowledge*, en anglès) per a determinades marques i variables escrites d'autoria. Per exemple, en comparació forense de veu i per a algunes llengües, se sap que el valor mitjà de la F0 (freqüència fonamental) dels homes és entre els 120 i 130 Hz i el de les dones entre els 180 i 220 Hz.²⁹ En tot cas, aquest endarreriment metodològic a què he al·ludit anteriorment, només es podrà anar superant en la mesura que s'obtingui el nombre màxim de mostres del nombre màxim d'escriptors possible.

La idiolectometria forense està aplicant nous enfocaments que poden oferir perspectives futures. Aquests enfocaments comporten, entre d'altres, la possibilitat de

mesurar no només les diferències lingüístiques entre diversos estils idiolectals, sinó també la distància idiolectal entre cada individu, per obtenir un índex de similitud idiolectal. Aquest índex permetria als lingüistes forenses a) comparar diverses mostres lingüístiques, b) calcular la distància entre aquestes mostres lingüístiques i c) establir quin grau de similitud idiolectal és necessària per poder concloure que dues mostres lingüístiques han estat escrites, o no, per la mateixa persona.

4. REFERÈNCIES BIBLIOGRÀFIQUES

ABECASSIS, M. (2002). «Saliency and frequency in a corpus of 1930's French films» *California Linguistic Notes*, 27, 2. <http://hss.fullerton.edu/linguistics/CLN/fal02/abecassis-saliency.pdf>

AISSIN, J. (1999). «Markedness and subject choice in Optimality Theory». *Natural Language and Linguistic Theory*, 17, 673-711.

AISSIN, J. (2003). «Differential object marking: Iconicity vs. Economy». *Natural Language and Linguistic Theory*, 21, 435-448.

ALAMARGOT, D. i CHANQOY, L. (2001). *Through the models of writing: Studies in writing*, 9. Amsterdam: Kluwer Academic Publishers.

ARNOLD, J. i WASOW, T. (1996). «The effect of production constraints on particle movement and dative alternation». Pòster presentat a CUNY sobre *Human Sentence Processing*. New York.

BALL, M. J. i RAHILLY, J. (1999). *Phonetics. The Science of Speech*. London: Arnold Publishers.

BEREITER, C. (1980). «Development in Writing». GREGG L. i Steinberg, E. (coord.), *Cognitive Processes in Writing*. New York: Lawrence Erlbaum, 3-30.

²⁹ Vegeu Ball i Rahilly (1999: 32) i Garn-Nunn i Lynn (2004: 19).

- BERWICK, R.C. (1985). *The acquisition of syntactic knowledge*. Cambridge, Massachusetts: MIT Press.
- BLAS ARROYO, J. L. (2007). «El contacto de lenguas como factor de retención en procesos de variación y cambio lingüístico: Datos sobre el Español en una comunidad bilingüe peninsular». *Spanish in Context*, 4, 2, 263-291.
- BOSQUE, I. i DEMONTE, V. (1999). *Gramática descriptiva de la lengua española*. Madrid: Espasa-Calpe.
- BUTLER, C. (1998). «Collocational frameworks in Spanish». *International Journal of Corpus Linguistics*, 3, 1-32.
- CIFUENTES HONRUBIA, J. L. (1994). *Gramática cognitiva. Fundamentos críticos*, Madrid: Eudema.
- CHASKI, C. E. (2001). «Empirical evaluations of language-based author identification techniques». *Forensic Linguistics. The International Journal of Speech, Language and the Law*, 8, 1, 1-65.
- CHOMSKY, N. (1981). *Lectures on Government and Binding* (Studies in Generative Grammar, 9). Dordrecht: Foris.
- CHOMSKY, N. i HALLE, M. (1968). *The sound pattern of English*. New York: Harper and Row.
- COULTHARD, R. M. (2005). «The linguist as expert witness». *Linguistics and the Human Sciences*, 1, 1, 39-58.
- CROFT, W. i CRUSE, D. A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- CUENCA, M. J. i HILFERTY, J. (1999). *Introducción a la lingüística cognitiva*. Barcelona: Ariel.
- DRESHER, B. E. i RICE, K. (2002). *Markedness and the Contrastive Hierarchy in Phonology*.
<http://www.chass.utoronto.ca/~contrast/>
- FAUCONNIER, G. (1997). *Mappings in Thought and Language*. Cambridge: Cambridge University Press.
- GARN-NUNN, P. G. i LYNN, J. M. (2004). *Calvert's Descriptive Phonetics*. 3a ed. New York: Thieme.
- GÓMEZ TORREGO, L. (1997). *Gramática didáctica del español*. Madrid: SM.
- GRANT, T. (2007). «Quantifying evidence in forensic authorship analysis». *The International Journal of Speech, Language and the Law*, 14, 1, 1-25.
- GRANT, T. (2008). «Approaching questions in forensic authorship analysis». GIBBONS, J. i TURELL, M. T. (eds.), *Dimensions of Forensic Linguistics*. Amsterdam/Philadelphia: John Benjamins, 215- 229.
- GRANT, T. i BAKER, K. (2001). «Identifying reliable, valid markers of authorship: a response to Chaski». *Forensic Linguistics. The International Journal of Speech, Language and the Law*, 8, 1, 66-79.
- GREENBERG, J. H. (ed.) (1966). *Universals of language*. 2a ed. Cambridge, Massachusetts: MIT Press.
- GIORA, R. (2003). *On our mind: salience, context and figurative language*. Oxford University Press.
- GORDON, P. C., GROSZ, B. J. i GILLIOM, L. A. (1993). «Pronouns, names, and the centering of attention in discourse». *Cognitive Science*, 17, 3, 311-48.
- GROZEA, C. i POPESCU, M. (2010). «Who's the thief? Automatic Detection of the Direction of Plagiarism». *CICLing*, 700-710.
- HALLIDAY, M. A. K i HASAN, R. (1976). *Cohesion in English*. London.
- HÄNLEIN, H. (1999). *Studies in Authorship Recognition: A Corpus-based Approach*. Frankfurt: Peter Lang.
- HOLYOAK, K.J. i THAGARD, P. (1997). «The analogical mind». *American Psychologist*, 52, 1, 35-44.

- HUME, E. (2004). «Deconstructing markedness: A predictability-based approach». *Berkeley Linguistics Society: Proceedings of the Annual Meeting 2004*, 30, 182–198. Department of Linguistics, University of California, Berkeley. http://www.ling.ohio-state.edu/~ehume/papers/Hume_markedness_BLS30.pdf
- HYAMS, N. M. (1987). *Language acquisition and the theory of parameters*. Dordrecht: Reidel.
- INCHAURRALDE, C. i VÁZQUEZ, I. (2000). *Una introducción cognitiva al lenguaje y la lingüística*. Zaragoza: Mira.
- JAKOBSON, R. i HALLE, M. (1956). *Fundamentals of language*. The Hague: Mouton.
- JOHNSON, A. (1997). «Textual kidnapping: a case of plagiarism among three students texts?». *Forensic Linguistics*, 4, 2, 210-225.
- KEAN, M. L. (1981). «On a theory of markedness». BELLETTI, A. et al. (eds.), *Theory of markedness and generative grammar: Proceedings of the 1979 GLOW Conference*. Pisa: Scuola Normale Superiore, 559-604.
- KESCKES, I. (2001). «The 'graded salience hypothesis' in second language acquisition». NIEMEIER, S. i PUETZ, M. (eds.), *Applied cognitive linguistics*. Berlin: Mouton de Gruyter, 249–71.
- KESCKES, I. (2004). «The role of salience in processing pragmatic units». *Acta Linguistica Hungarica*, 51, 309–324.
- KINTSCH, W. (1988). «The role of knowledge in discourse processing: A construction-integration model». *Psychological Review*, 95, 163-182.
- KREDENS, K. (2001). «Language Corpora in Forensic Linguistics». LEWANDOWSKA-TOMASZCZYK, B. i MELIA, P. J. (eds.), *PALC'99. Practical Applications in Language Corpora*. Main: Peter Lang.
- LOVE, H. (2002). *Attributing authorship: An introduction*. Cambridge: Cambridge University Press.
- MENASCHE, L. (1977). *Writing a Research Paper*. Ann Arbor: The University of Michigan Press.
- MYERS-SCOTTON, C. (1998). *Codes and Consequences. Choosing Linguistic Varieties*. Oxford: Oxford University Press.
- OCHS, E. (1979). «Planned and unplanned discourse». GIVÓN, T. (ed.), *Discourse and Syntax*, V. New York: Academic Press, 51–80.
- PAYRATÓ, L. (1985). *La interferència lingüística. Comentaris i exemples català-castellà*. Barcelona: Curial / Abadia de Montserrat.
- POTTHAST, M., BARRÓN-CEDENO, A., STEIN, B. i ROSSO, P. (2010). «An Evaluation Framework for Plagiarism Detection». Estudi presentat a Proceedings of the 23rd International Conference on Computational Linguistics, COLING. Beijing: China, 997-1005.
- POWTOSKI, K. i CAMERON, R. (2007). *Spanish in contact: Policy, Social and Linguistic Inquiries*. Amsterdam/Philadelphia: John Benjamins.
- QUIRK, R., GREENBAUM, S., LEECH, G. i SVARTVIK, J. (1972). *A Comprehensive Grammar of the English Language*. London i New York: Longman.
- ROCA, A. i JENSEN, J. B. (1996). *Spanish in contact: Issues in Bilingualism*. Somerville, MA: Cascadilla Press.
- ROIG, M. (2006). *Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing*. 1a versió en línia: setembre 2003. Versió revisada en línia: agost 2006. <http://facpub.stjohns.edu/~roigm/plagiarism/>
- SCHMID, H. J. et al. (1996). *An Introduction to Cognitive Linguistics*. Nova York: Longman.
- SILVA-CORVALÁN, C. (1995). *Spanish in Four Continents: Studies in language contact and*

bilingualism. Washington, D.C.: Georgetown University Press.

SOLAN, L. i TIERSMA, P. (2004). *Speaking of Crime: The Language of Criminal Justice*. Chicago: University of Chicago Press.

SPASSOVA, M. S. (2009). El potencial discriminatorio de las secuencias de categorías gramaticales en la atribución forense de autoría de textos en español. Tesis doctoral no publicada. Barcelona: Universitat Pompeu Fabra. <http://www.tdx.cat/handle/10803/7512;jsessionid=453C659BABA9E765246B6E798BEDF543.tdx1>

STEIN, B. i MEYER ZU EISSEN, S. (2007). «Intrinsic Plagiarism Analysis with Meta Learning». *PAN*. Amsterdam: The Netherlands, 45-50.

STAMATATOS, E. (2009). «Intrinsic Plagiarism Detection Using Character ngram Profiles». *PAN*, Donostia-San Sebastian: Spain.

STERIADE, D. (1987). «Redundant values». *Chicago Linguistic Society*, 23, 239-362.

TAYLOR, J. R. (2002). *Cognitive Grammar*. Oxford: Oxford University Press.

TURELL, M. T. (2004). «Textual kidnapping revisited: the case of plagiarism in literary translation». *The International Journal of Speech, language and the Law. Forensic Linguistics*, 11, 1, 1-26.

TURELL, M. T. (2007). «Plagio y traducción literaria». *Vasos Comunicantes*, 37, 1, 43-54.

TURELL, M. T. (2008) «Plagiarism». GIBBONS, J. i TURELL, M. T. (eds.), *Dimensions of Forensic Linguistics*. John Benjamins: Amsterdam/Philadelphia, 265-299.

TURELL, M. T. (2010a). *Los retos de la lingüística forense en el siglo XXI. In Memoriam Enrique Alcaraz Varó*. Alacant: Departament de Filologia Anglesa, Universitat d'Alacant.

TURELL, M. T. (2010b). «The use of

textual, grammatical and sociolinguistic evidence in forensic text comparison». *The International Journal of Speech, Language and the Law*, 11, 1, 211-250.

VIEL, M. (1984). *La notion de 'marque' chez Trubetzkoy et Jakobson: Un épisode dans l'histoire de la pensée structurale*. Paris: Champion.

VILA, M., MARTÍ, M. A. i RODRÍGUEZ, H. (2011). «Paraphrase Concept and Typology. A Linguistically Based and Computationally Oriented Approach». *Procesamiento del Lenguaje Natural*, 46, 83-90.

WESCH, A. (1997). «El castellano hablado en Barcelona y el influjo del catalán. Esbozo de un programa de investigación». *Verba*, 24, 287-312.

WEXLER, K. i MANZINI, M. R. (1987). «Parameters and learnability in binding theory». ROEPER, J. i WILLIAMS, E. (eds.), *Parameter setting*. Dordrecht: Reidel, 41-89.

WILLIAMS, E. (1981). «Language acquisition, markedness and phrase structure». TAVAKOLIAN, S. L. (ed.), *Language acquisition and linguistic theory*. Cambridge, Massachusetts: MIT Press, 8-34.

5. CASOS CITATS

Bruño v. Magister. 2006. Plagi.

Formoso v. Cela. 1999. Suposat plagi.

Pujante v. Vázquez Montalbán. 1993. Plagi. entre traduccions.

SEHRS v. ZZZ. 2004. Autoria.

XXX v. YYY. 2007. Plagi.