

LA LINGÜÍSTICA FORENSE: LA LLENGUA COM A EVIDÈNCIA¹

Jordi Cicres*

Núria Gavaldá**

Resum

En aquest article presentem les principals aplicacions de la branca evidencial o probatòria de la lingüística forense, la disciplina que estudia l'espai d'interfície entre el llenguatge i el dret. A continuació, discutim el concepte d'estil idiolectal, entès com el conjunt d'eleccions lingüístiques individuals, i el paper que té en la pràctica forense. A través de tres exemples, centrats en la identificació de parlants, l'atribució d'autoria de textos i l'anàlisi del plagí, exposem breument les tècniques i variables d'anàlisi utilitzades per la disciplina. Finalment, discutim breument els aspectes ètics i els reptes de futur de la lingüística forense.

Paraules clau: lingüística forense; estil idiolectal; plagí; identificació de parlants; atribució d'autoria.

FORENSIC LINGUISTICS: LANGUAGE AS EVIDENCE

Abstract

In this article we introduce the main applications of the evidential branch of forensic linguistics, the discipline that studies the interface between language and law. We then discuss the concept of idiolectal style, understood as a set of individual linguistic choices, and its role in forensic practice. We offer three examples, focusing on speaker identification, authorship attribution and plagiarism analysis, to briefly explain the analytic techniques and variables used in the discipline. Finally, we briefly discuss the ethical aspects and future challenges of forensic linguistics.

Key words

Forensic linguistics; idiolectal style; plagiarism; speaker identification; authorship attribution.

¹ Finançat pel MINECO (Ref. FFI2012-34601).

*Jordi Cicres, Universitat de Girona, pl. de Sant Domènec, 9. 17071 Girona. jordi.cicres@upf.edu

**Núria Gavaldá, ForensicLab - Laboratori de Lingüística Forense. Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra, c. Roc Boronat, 138. 08018 Barcelona.

Article rebut el 28.10.2013. Avaluació cega: 29.11.2013. Data d'acceptació de la versió final: 23.12.2013

Citació recomanada: CICRES, Jordi; GAVALDÀ, Núria. «La lingüística forense: la llengua com a evidència», *Revista de Llengua i Dret*, núm. 61, 2014, p.60-71, DOI: [10.2436/20.8030.02.54](https://doi.org/10.2436/20.8030.02.54)

Sumari

1 La lingüística forense

2 Per què l'ús de la llengua ens fa diferents?

3 Exemple 1. Com es pot identificar una persona per la veu?

4 Exemple 2. Com es pot identificar una persona pels seus escrits?

5 Exemple 3. Com podem aportar indicis de plagi?

6 Conclusions i reptes de futur

Bibliografia

1 La lingüística forense

En un món cada vegada més complex, la interdisciplinarietat es fa necessària per resoldre nous problemes. En aquest sentit, la lingüística forense és un exemple de síntesi entre dues disciplines diferents: la lingüística i el dret. Gibbons i Turell (2008) distingeixen tres àmbits de treball de la lingüística forense: el llenguatge jurídic (*the language of the law*), el llenguatge judicial (*the language of the court*) i el llenguatge evidencial (*language as evidence*). És cert que els dos primers àmbits, centrats a estudiar el llenguatge utilitzat en lleis, normes, contractes, pòlisses, etc., d'una banda, i de l'altra l'ús del llenguatge dels actors dels procediments judicials (jutges, advocats, policies, testimonis, pèrits, etc.) tenen ja una llarga trajectòria, però l'àmbit evidencial ha tingut un desenvolupament molt més tardà. De fet, es considera que aquesta branca evidencial s'inicia com a disciplina acadèmica l'any 1968 al Regne Unit, a partir d'un estudi del lingüista Jan Svartvik, titulat «The Evans Statements: A case for Forensic Linguistics», en el qual realitza una anàlisi de les suposades declaracions fetes davant la policia de Timothy Evans, acusat d'assassinar la seva filla. Mitjançant l'anàlisi descriptiva i estadística, va demostrar que les declaracions havien estat manipulades per la policia. Malauradament, però, Evans va ser declarat culpable i executat divuit anys abans que Svartvik fes el seu treball. Però són molts els casos en què l'anàlisi lingüística ha estat útil en un procediment judicial.

L'ús del coneixement lingüístic en els procediments judicials pot ser clau en una tipologia molt diversa de situacions. Les més habituals, sens dubte, tenen a veure amb la identificació de parlants, l'atribució de l'autoria d'un text escrit, i amb l'anàlisi lingüística de possibles casos de plagi. En qualsevol cas, però, no són els pèrits els professionals qui *identifiquen* parlants o escriptors, o determinen l'existència de plagi, sinó que aquesta tasca correspon als jutges, tribunals o jurats tenint en compte les aportacions dels pèrits. Però les situacions en què l'expertesa d'un lingüista pot ser necessària van més enllà, i inclouen entre d'altres aspectes la transcripció i anàlisi de la integritat d'una gravació; l'anàlisi lingüística de marques registrades i patents; la traducció i interpretació en contextos judicials; la construcció de perfils lingüístics; l'aclariment del significat de textos en disputa (comercials, legals, d'amenaçes); l'anàlisi dels interrogatoris (quines instruccions es donen, quines estratègies lingüístiques utilitzen els interrogadors i els interrogats); la comprensió de textos jurídics (ordres de restricció, drets dels detinguts); l'anàlisi de declaracions de testimonis per determinar-ne l'autenticitat (denúncies, trucades als serveis d'emergència), i la determinació de l'origen geogràfic dels sol·licitants d'asil mitjançant l'anàlisi de la llengua o llengües que parlen.

Així, els lingüistes forenses que intervenen com a pèrits ho poden fer amb dues finalitats: investigatives i evidencials. Les primeres tenen a veure amb el desenvolupament d'investigacions judicials o policials. Típicament, es tracta d'aportar informacions com ara l'origen dialectal, el sexe, el nivell socioeducatiu de l'autor d'un text escrit o oral (és a dir, l'elaboració d'un perfil lingüístic), o bé altres tasques com ara la neteja i transcripció de gravacions. Aquestes informacions serveixen per delimitar la investigació o aportar-hi dades d'origen lingüístic. Al seu torn, les finalitats evidencials consisteixen a aportar evidència lingüística durant el procés. Aquestes evidències tenen a veure majoritàriament amb la comparació de textos orals o escrits amb l'objectiu d'assistir el jutge o tribunal en la determinació de l'existència de plagi, o bé en la determinació o atribució de l'autoria d'un text escrit o d'una gravació oral.

2 Per què l'ús de la llengua ens fa diferents?

La lingüística forense està estretament lligada a l'estudi de la variació lingüística que va introduir l'escola laboviana dins del camp de la sociolingüística. Segons aquesta perspectiva, tota producció lingüística és variable, i aquesta variació està estructurada i estratificada per factors lingüístics, socials i estilístics. Tradicionalment, els estudis sociolingüístics s'han centrat en l'estudi de la variació en comunitats lingüístiques o comunitats de parla –és a dir, la variació en el grup. Però més recentment, alguns estudis sociolingüístics s'han centrat en el paper que la variació individual té en la variació en grup, una perspectiva clau per a la lingüística forense, que vol demostrar que cada parlant té un ús distintiu de la llengua que el separa de la resta d'usuaris de la mateixa llengua. Segons Johnstone (1996: 16), «a large residue of inter-speaker variability is left even once differences in age, socioeconomic class, region, rurality, gender and contextual style are

carefully considered». En aquest sentit, diversos estudis dins la sociolingüística demostren que fins i tot dins de comunitats que comparteixen la mateixa varietat lingüística i el mateix context social, emergeixen diferències significatives entre subjectes que només poden explicar-se com a tries idiosincràtiques (Milroy i Milroy 1977, 1978; Ferguson 1979; Payne 1980). Encara que aquests estudis no estan relacionats amb l'aplicació de l'anàlisi de la llengua a contextos forenses, influeixen en gran mesura el concepte d'idiòlecte que sorgeix dins del camp de la lingüística forense. Un dels primers autors que va parlar sobre la idea d'idiòlecte des d'aquesta perspectiva és Baldwin (1979: 231), qui el defineix com la forma única d'un dialecte que tipifica un individu particular i que està constituït per les seves preferències lingüístiques com ara l'ús de certes paraules, la pronúncia de certs sons o patrons d'entonació, etc. Ja al segle XXI, altres autors com McMenamin (2002), Coulthard (2004), Coulthard i Johnson (2007) i Turell (2010) també han adoptat el concepte d'idiòlecte en la seva recerca i peritatge lingüístic. Segons Coulthard:

The linguist approaches the problem of questioned authorship from the theoretical position that every native speaker has their own distinct and individual version of the language they speak and write, their own *idiolect*, and the assumption that this *idiolect* will manifest itself through distinctive and idiosyncratic choices in texts [...]. Thus, whereas in principle any speaker/writer can use any word at any time, speakers in fact tend to make typical and individuating co-selections of preferred words. (2004: 431-432).

No obstant això, la variació lingüística també afecta el nivell idiòlectal, i per tant, tots aquells trets que caracteritzen l'ús de la llengua d'un individu també mostren variació. És a dir, no només hi ha la variació entre diferents subjectes —variació interpersonal—, sinó que també hi ha la variació dins d'un mateix individu —variació intrapersonal— que pot dependre de diversos factors com per exemple el context comunicatiu, l'estat d'ànim o fins i tot el pas del temps. De fet, la diferència més important entre l'evidència lingüística i altres tipus d'evidència com per exemple l'empremta dactilar o l'ADN és precisament aquesta variació intrapersonal, i el fet que mostres d'un mateix individu difereixen en certa mesura. Però tot i aquesta variació intrapersonal, és possible arribar a conclusions fiables sobre la mateixa o diferent procedència de mostres lingüístiques. Diversos estudis recents han demostrat que la variació interpersonal és generalment més alta que la variació intrapersonal (Cicres 2007a; Bel *et al.* 2012; Gavaldà 2013; Turell i Gavaldà 2013). És a dir, l'estil idiòlectal d'un individu, malgrat mostrar variació, és relativament estable, i la tasca principal del lingüista forense a l'hora de donar elements tècnics de decisió al jutge, tribunal o jurat perquè determini si les mostres dubitada i indubitada han estat produïdes per la mateixa o per diferents persones és identificar els paràmetres que poden ser la font de variació interpersonal i determinar el nivell de variació intrapersonal que aquests paràmetres poden mostrar (Nolan 1997).

3 Exemple 1. Com es pot identificar una persona per la veu?

Qualsevol acte de parla implica dos tipus d'informacions complementàries i independents, relacionades amb el *contingut* del missatge, d'una banda, i amb la forma acústica amb què es transmet (el *continent*), de l'altra. El contingut es vehicula lingüísticament a través d'opcions determinades per factors diversos, com l'origen dialectal del parlant, la influència d'altres llengües o el registre utilitzat, entre d'altres. Aquest contingut cal analitzar-lo amb les tècniques pròpies de les diferents branques de la lingüística: fonologia, morfologia, sintaxi, semàntica, pragmàtica i anàlisi del discurs. El continent, en canvi, ve determinat tant per la morfologia com per la posició i moviments dels òrgans articuladoris, que configuren cavitats supraglòtiques diferents en forma i volum que fan que el so provinent de la laringe s'amplifiqui i s'organitzi freqüencialment d'una o altra manera. Tant la morfologia dels òrgans articuladoris com —en menor grau— els moviments per produir els diferents sons de la parla presenten diferències entre els parlants, cosa que provoca que les seves produccions acústiques siguin també diferents.

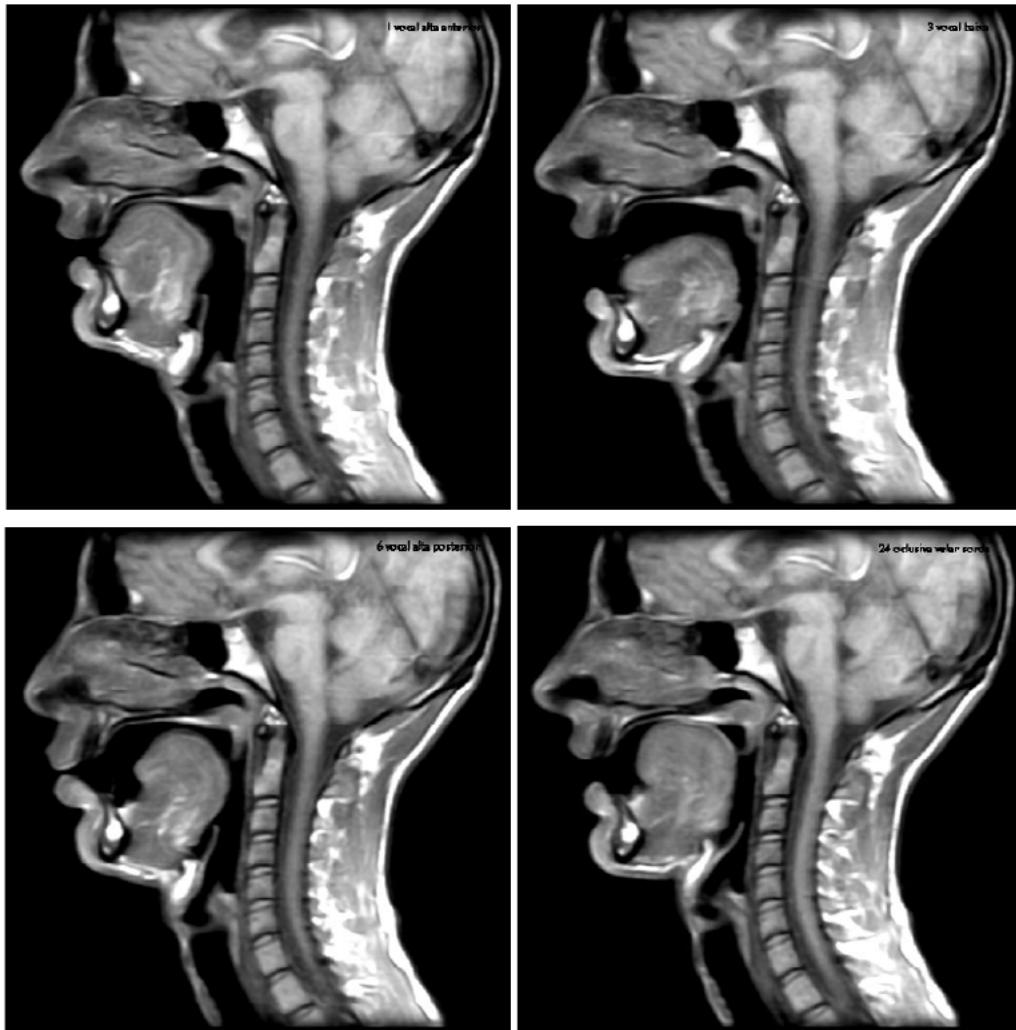


Figura 1. Imatges de ressonància magnètica que pertanyen a un mateix parlant amb configuracions articulatòries diferents ([i], [a], [u] i [k]).

Per tant, a l'hora de fer una comparació forense de veus, cal tenir en compte aquests dos components (Delgado 2001; Rose 2002). Un exemple que demostra la importància d'aquest enfocament combinat és el cas d'Òscar Sánchez, seguit amb molt interès pels mitjans de comunicació², víctima d'una suplantació d'identitat i condemnat sobre la base d'un peritatge de veu erroni elaborat pel pèrit enginyer italià Roberto Porto que no havia tingut en compte l'anàlisi lingüística i que incloïa errors significatius d'anàlisi de les dades acústiques i de la interpretació dels resultats.

Van ser necessaris fins a quatre nous peritatges per esmenar el primer, un dels quals fou realitzat per investigadors del Laboratori de Lingüística Forense (ForensicLab), de l'Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra. En aquest informe es van fer evidents diferències significatives en les mostres de veu dubitades (un conjunt de gravacions telefòniques interceptades per la policia, i atribuïdes inicialment a Òscar Sánchez) respecte de la mostra indubitada (efectivament d'Òscar Sánchez) en tots els nivells.

En la comparació forense de veus, cal analitzar diversos paràmetres acústics que tenen una relació directa amb la forma de les cavitats (nasal, faríngia i oral) de ressonància del tracte vocal, d'una banda, i amb la freqüència de vibració de les cordes vocals (que al seu torn, també depèn tant de factors fisiològics com el grau d'elasticitat i dimensions de les cordes vocals com de factors intensionals mitjançant la modificació

² Per exemple: «[Òscar Sánchez ya está en la calle](#)». «[Òscar Sánchez torna a Catalunya després de 20 mesos en una presó italiana per una suplantació d'identitat](#)»; «[Itàlia revisa hoy la condena de Óscar con la fiscalía a favor de la inocencia](#)».

del grau de tensió i de la pressió subglotal), de l'altra. Aquests paràmetres, per tant, es veuen reflectits en les propietats acústiques dels diferents segments fonètics (sons vocàlics i consonàntics) i en els aspectes suprasegmentals (freqüència fonamental, entonació, tempo i ritme).

A l'hora de definir els paràmetres d'anàlisi ideals, cal tenir en compte una sèrie de requisits (Nolan 1983: 11): que presentin una alta variabilitat entre parlants diferents i una variabilitat baixa en les diverses realitzacions d'un mateix parlant; que siguin resistents als intents d'emascament i d'imitació; que tinguin una alta freqüència d'ocurrència en les mostres d'anàlisi; que siguin robusts en la transmissió (és a dir, que la transmissió no n'alteri o elimini les propietats acústiques), i que siguin relativament fàcils d'extreure i d'analitzar. A més, Rose (2002: 52) hi afegeix que cada paràmetre sigui independent dels altres. Malgrat això, les característiques de moltes de les gravacions en el context forense dificulten la tasca, ja que poden ser molt breus, de baixa qualitat, passades pel filtre telefònic (de manera que les freqüències inferiors als 400 Hz i superiors als 3500 Hz no es transmeten), o una combinació de les tres. Per tant, la definició de les variables analitzables dependrà en cada cas de les característiques de les gravacions. Sigui com sigui, les més habituals (segons Gold i French 2011) són estudiar l'estructura espectral de les vocals (amb les transicions dels formants i distàncies interformàntiques), nasals, laterals i fricatives, el *Voice Onset Time* de les oclusives, les propietats de les consonants vibrants i la tipologia i propietats acústiques multidimensionals de les pauses dubitatives (Rose 2002; Cicres 2007b). A partir de l'anàlisi de la freqüència de vibració de les cordes vocals podem analitzar també l'entonació i l'alineació tonal (Baldwin i French 1990; Nolan 2002; Zetterholm 2002; Eriksson i Wretling 1997; Cicres i Turell 2005; Cicres 2007a).

Aquest conjunt de dades, obtingudes tant de les gravacions dubitades com de les indubitades, cal posar-les en relació amb les dades obtingudes de corpus de referència.

Amb l'anàlisi acústica de les gravacions es van observar diferències significatives en la majoria de paràmetres analitzats. En destacaven, però, de manera especial la producció del so fricatiu [s], molt més agut i definit en les gravacions indubitades que en les dubitades, i l'estructura de formants dels sons vocàlics, tant de les vocals lèxiques (les que formen part de les paraules del discurs) com la de les pauses plenes (els sons dubitatius, allargats, que s'utilitzen per omplir silencis).

Però l'aspecte més rellevant en aquest cas, totalment negligit en el dictamen del perit Porto, és l'anàlisi lingüística, que va evidenciar que els parlants de les gravacions dubitades i indubitada utilitzen variants dialectals diferents en cada cas: Òscar Sánchez utilitza una variant peninsular, mentre que el parlant de les gravacions dubitades és parlant d'una variant sud-americana (concretament, del dialecte uruguaià). Entre altres evidències, destaquen: l'ús del *seseo* (fenomen que consisteix a pronunciar el fonema fricatiu sord interdental com una [s]), present en les gravacions dubitades, però no en la indubitada; la pèrdua del moment oclusiu en l'articulació del grup ortogràfic *ch*, de manera que pronuncia el mot *ochó* com a [ofo]; o la neutralització de /r/ i /l/ en posició implosiva (de manera que pronuncia com a [r] la *l* present en mots com *algo* o *alguien*).

Aquest cas és un exemple clar de la importància d'una bona formació lingüística (general, especialitzada en el seu vessant forense en els àmbits fonètic i fonològic, i de la llengua o llengües dels textos o gravacions que cal analitzar) per a realitzar peritatges lingüístics. De fet, per aquest motiu, el codi ètic de l'Associació Internacional de Fonètica i Acústica Forenses (IAFPA, en les seves sigles en anglès) estableix que «els membres han de tenir especial precaució de realitzar l'anàlisi forense sigui del tipus que sigui en els enregistraments que contenen discursos en llengües en les quals no en són parlants nadius»³.

3 [International Association for Forensic Phonetics and Acoustics](#).

4 Exemple 2. Com es pot identificar una persona pels seus escrits?

Tots els escriptors poden fer servir les paraules que vulguin quan vulguin, però dins de les restriccions imposades per cada gènere textual (cartes, correus electrònics, faxes, textos literaris...) tendeixen a fer una sèrie de seleccions particulars de recursos lingüístics que són idiosincràtiques i que determinen el seu estil idiolectal. Una comparació forense de textos escrits fiable i rigorosa ha de combinar anàlisis qualitatives i quantitatives (o estadístiques) que confirmen la significança darrere les similituds o diferències establertes en l'anàlisi qualitativa.

Les tècniques que es poden utilitzar de manera combinada en una anàlisi forense de textos i que han demostrat detectar marcadors d'autoria fiables són molt diverses. D'una banda, la perspectiva estilística analitza els escrits qualitativament i identifica aquells trets consistents dins del mateix escriptor i distintius en escriptors diferents com poden ser els errors gramaticals o ortogràfics i altres elements idiosincràtics (McMenamin 2002; Grant 2010). Per això, és convenient tenir diferents escrits dels mateixos escriptors dels quals es vol realitzar l'anàlisi de cara a demostrar que la variació dins del mateix autor és sempre més baixa que la variació entre escriptors diferents. D'altra banda, la perspectiva estilomètrica consisteix en una quantificació mitjançant diversos mètodes computacionals i estadístics de diversos marcadors d'autoria com poden ser la llargada de les paraules, la llargada de les frases, la freqüència de paraules funcionals i de paraules lèxiques, la distribució de paraules amb diferents longituds, la densitat lèxica, o la riquesa lèxica (p.e. Burrows 1987; Holmes 1994; Wools i Coulthard 1998; Spassova i Turell 2007; Grant 2007). Finalment, també s'ha posat molt d'èmfasi en l'estudi i la identificació de trets referents a l'estil idiolectal (Chaski 2001; Grant i Baker 2001; Turell i Gavaldà 2013) i també en la importància de l'ús de corpus de referència per establir la prominència de les marques analitzades respecte de la producció general de la població rellevant (Coulthard 1994; Turell 2010).

El cas que s'exposa a continuació (citat a Turell 2010) és un exemple d'anàlisi forense d'autoria que combina una gran part dels mètodes qualitius i quantitius que s'acaben d'exposar. El cas en qüestió es va dur a terme al ForensicLab l'any 2003.

En aquest any, una empresa del sector turístic a Catalunya estava passant per un moment delicat econòmicament i corrien rumors per l'empresa que volien reduir la plantilla. Un dels treballadors, mogut per la por a ser acomiadat, va decidir enviar una sèrie de correus electrònics extorsionant l'empresa. Els directius de l'empresa sabien qui era el treballador que els estava extorsionant, ja que van esbrinar que els correus enviats van sortir del seu ordinador i van decidir acomiadar-lo. Com a resposta, el treballador va obrir una causa contra l'empresa per acomiadament improcedent. Va negar haver enviat els correus extorsionadors i va al·legar que alguna altra persona hauria intervingut l'ordinador i els hauria enviat. Així doncs, la part demandada va demanar una perícia lingüística amb l'objectiu de comparar els textos dubitats (els correus) amb un conjunt de textos indubitats (que, en aquest cas, eren una sèrie de faxes que el treballador va enviar a l'empresa durant la mateixa època), i aportar evidència científica sobre la possibilitat que haguessin estat escrits pel mateix autor. Tant els quatre correus com els quatre faxes analitzats es van escriure en castellà i tenien entre 400 i 1000 paraules.

El gènere textual sovint determina algunes de les variables lingüístiques que s'analitzaran, però tot i així, sempre hi ha espai per a preferències estilístiques de cada autor, i per tant, per a seleccions idiolectals. En aquest cas, tractant-se de correus electrònics i de faxes, es van analitzar les fórmules de salutació i comiat i la manera d'adreçar-se a l'interlocutor. En aquest sentit, l'anàlisi va demostrar que els dos conjunts de textos, els dubitats i els indubitats, mostraven les mateixes fórmules de salutació: *Hola César* o *Buenas tardes César*, i també la mateixa fórmula de comiat: *Recibe un cordial saludo*. A més a més d'aquestes fórmules relatives al gènere textual, també es va poder observar unes expressions molt similars que apareixien en ambdós conjunts de textos (taula 1).

Taula 1: expressions comunes als textos dubitats i indubitats. (Font: Turell 2010: 223)

Textos dubitats	Textos indubitats
...tú me has merecido el mayor de los respetos...	...siempre has merecido el mayor de mis respetos...
...lo echan a la calle como a un perro...	...me han echado a la calle como a un perro...
...por profesionalidad, los clientes siguen siendo atendidos...	...sigo atendiendo los grupos... por cuestión de profesionalidad...
...llegaría el día en que pondríamos las cartas sobre la mesa...	...pondremos todas las cartas sobre la mesa...

També es van analitzar els errors ortogràfics: els dos conjunts de textos mostraven errors d'accentuació a les mateixes paraules, com per exemple *tú* (pronomen personal), *mi* (pronomen personal) i el morfema de primera i tercera persona de l'indicatiu imperfecte en paraules com *tenía* o *sería*, i també errors gramaticals com l'omissió de la <h> al verb *haber* en 3a persona del singular en formes verbals compostes, com per exemple *(h)a trabajado* o *(h)a demostrado*.

Una de les característiques més significatives que es va observar en els dos conjunts de textos va ser la clara influència de la llengua catalana en l'escriptura en castellà en l'escriptor, qui, casualment, no era català d'origen, sinó que provenia d'un poble d'Aragó on no es parla català i que feia més de trenta anys que vivia a Catalunya. Així doncs, tots dos conjunts de textos mostraven el procés de queisme, en expressions com *se dio cuenta Ø que tenía en sus manos, en el caso Ø que*, i també la caiguda de la preposició *a* davant d'objecte directe com *para atender Ø los clientes*.

Finalment, aquest escriptor també mostrava una freqüència d'ús del pronomen relatiu compost *el cual* (amb totes les formes flexionades) molt més elevada del que és usual en castellà. De fet, consultat el corpus CREA per veure quina freqüència d'ús és habitual en la població de referència, s'observa que el relatiu compost només s'utilitza en un 2% de les ocasions, mentre que el relatiu *que* apareix en el 98% dels casos. Contràriament, el demandant (en els faxes, indubitats) utilitzava el relatiu simple *que* el 52% de les ocasions, i el compost (*el cual*) en el 48% dels casos. Aquesta freqüència d'ús era molt similar també en els textos dubitats.

Totes aquestes marques d'autoria, juntament amb anàlisis quantitatives referents a l'estructura del text, la freqüència de paraules funcionals i lèxiques, la riquesa lèxica i les seqüències de categories gramaticals mitjançant la tècnica dels n-grames, van portar a la conclusió que els textos exhibien una similitud molt alta entre si, i que la probabilitat que haguessin estat escrits pel mateix escriptor era molt alta. El dictamen que es va presentar va ajudar a decidir el cas a favor de l'empresa acusada.

5 Exemple 3. Com podem aportar indicis de plagi?

El plagi consisteix en l'apropiació d'idees i paraules escrites per una altra persona i fer-les passar per pròpies. Així, podem establir dos tipus de plagi: el plagi d'idees, que es produeix quan una persona utilitza idees d'una altra sense fer-ne cap reconeixement explícit, o bé el plagi lingüístic, que succeeix quan, a més de copiar les idees, utilitza les mateixes paraules o les mateixes estructures lingüístiques que l'autor del text original. Aquesta pràctica constitueix un delicte contra la propietat intel·lectual, i en litigis d'aquest tipus, el perit lingüista pot ajudar a establir el nivell de similitud textual entre dos o més textos i establir la probabilitat que no hagin estat produïts independentment, i per tant, que hi hagi hagut plagi lingüístic.

La primera pregunta de recerca que es fa el perit lingüista davant d'un possible cas de plagi lingüístic és la direccionalitat del possible plagi, és a dir, quin dels dos textos s'ha produït originàriament i, per tant, independentment de l'altre (el plagiat), i quin ha utilitzat d'alguna manera el text original (el plagiar). La majoria de vegades, si es tracta de textos publicats, la data de publicació sol indicar quin s'ha escrit abans

(tot i que no necessàriament ha de ser una dada concluent). Per això, també s'han de tenir en compte una sèrie de característiques lingüístiques, principalment relacionades amb la unitat, la coherència i la cohesió dels textos (Turell 2008).

Un cop establerta la direccionalitat, s'ha de determinar el nivell de similitud entre els textos, les marques d'autoria que han estat copiades i les tècniques emprades per copiar (i, si és el cas, intentar dissimular la còpia). Des del punt de vista qualitatiu, es poden detectar estratègies com la paràfrasi, és a dir, expressar la mateixa idea però canviant-ne algunes paraules i estructures lingüístiques, o substituir alguns elements lèxics per sinònims mantenint l'estructura sintàctica de l'original. A part d'una anàlisi qualitativa, els lingüistes forenses també tenen a la seva disposició eines informàtiques, com el programa informàtic *CopyCatch* (dissenyat per David Wools de CFL Software Development), que permet establir quantitativament el nivell d'igualtat de similitud a partir del qual és poc probable que dos textos hagin estat produïts de manera independent. Un cop aplicat el filtre de paraules que no s'han de tenir en compte (com les paraules funcionals, o les paraules específiques al tipus de textos que s'estan tractant), aquest programa incorpora diverses mesures quantitatives com els percentatges de vocabulari coincident, de paraules i de frases compartides un sol cop i el de vocabulari únic (paraules que apareixen només en un dels textos). Com més alt sigui el percentatge en els dos primers paràmetres, més alta és la probabilitat que els textos no hagin estat produïts independentment. Contràriament, el percentatge de vocabulari únic sol ser baix si els textos no són independents, és a dir, que un parteix de l'altre.

Un exemple de plagi lingüístic per al qual un lingüista forense va ser consultat i va poder aportar informació molt significativa que va decantar el jutge cap a la conclusió que sí que hi havia hagut plagi lingüístic va ser un cas que concernia dos llibres de text de física i química. El text A (el produït i publicat primer) era un llibre de batxillerat, i el text B (el plagiari) era un llibre per a oposicions a professorat de secundària. Els continguts que els dos textos presentaven eren quasi idèntics i estaven disposats en el mateix ordre. Però la similitud més manifesta es trobava en els exercicis, les qüestions, els exemples i les tècniques de treball científic. Molts d'aquests elements estaven redactats exactament iguals, i d'altres estaven fets amb la tècnica del «retalla i enganxa», amb l'encapçalament d'una activitat i les preguntes associades a una segona activitat, per la qual cosa moltes resultaven incoherents. En altres ocasions, alguns dels exercicis apareixien en el text A amb un ordre seqüencial, en el sentit que era necessari realitzar el primer exercici per a poder entendre i fer el segon; en el text B, en canvi, apareixien amb un ordre invertit, amb la consegüent dificultat per comprendre'ls. Un altre recurs tenia a veure amb l'estil referencial, és a dir, amb la manera com el text s'adreça al lector. El text A utilitzava sistemàticament la forma d'imperatiu de segona persona singular, com per exemple *analiza* o *redacta*. En canvi, el text B mostrava una inconsistència en l'estil referencial, ja que usava de manera no sistemàtica tant l'imperatiu com l'infinitiu. Finalment, un altre tret que es va observar en l'anàlisi qualitativa del text B va ser la supressió de parts del discurs que donava lloc a incongruències conceptuals, ja que hi havia conceptes i explicacions tretes de context que sovint eren difícils d'entendre. De manera similar, també es produïa la descontextualització d'algunes activitats, suprimint una part de la descripció de l'activitat que dificultava la comprensió de l'objectiu de l'exercici.

L'anàlisi qualitativa mitjançant el programa *CopyCatch* va establir que una majoria (un 66%) de les activitats que apareixien al text B també apareixien al text A, mentre que només un 12% de les activitats del text A apareixien al text B, la qual cosa indicava que el text A era més original i innovador, i el text B havia copiat una gran part de les activitats del text A. De manera similar, el percentatge de vocabulari coincident de les activitats mostrava un grau de coincidència del 96%, i un 98% en les qüestions. Per tant, hi havia evidència suficient per determinar que la gran majoria d'activitats i qüestions que constaven al text A havien estat reproduïdes textualment al text B. Amb tots aquests indicis lingüístics, el jutge encarregat del cas va determinar que hi havia hagut plagi i va fallar a favor de la part demandant.

6 Conclusions i reptes de futur

En aquest article hem vist, a través de tres exemples, com la branca evidencial o probatòria de la lingüística forense pot tenir un paper clau en alguns procediments judicials. Els àmbits en què els lingüistes poden actuar com a pèrits judicials són molt diversos, i tenen a veure amb la comparació forense de veus o textos escrits (que aporti informació científica encaminada a una eventual identificació del parlant o l'autor dels textos per part del jutge o tribunal), l'anàlisi del plagi, la transcripció forense de gravacions, l'anàlisi de la integritat de les gravacions en col·laboració amb enginyers per detectar-ne possibles manipulacions, etc., però en tots els casos cal un compromís ètic i uns coneixements metodològics específics.

En els darrers anys i des de moltes institucions vinculades en major o menor grau amb la recerca s'estan fent passos molt significatius per millorar les tècniques d'anàlisi i incrementar el rigor científic en els peritatges. Bàsicament, s'estan fent progressos molt importants des de tres àmbits diferents: en primer lloc, des de la lingüística aplicada. La lingüística forense té cada any més presència en les universitats d'arreu del món en la recerca i en la formació especialitzada. Altres institucions de recerca, com els laboratoris d'acústica forense de diversos cossos policials també treballen amb lingüistes. En segon lloc, des de l'àmbit de l'enginyeria (i concretament des de l'enginyeria lingüística i l'enginyeria acústica) s'estan fent progressos amb tècniques computacionals tant en la identificació de parlants com en la determinació d'autoria o detecció de plagi lingüístic. I en darrer lloc, des de l'àmbit del dret, els juristes estan treballant en els aspectes relacionats amb l'ètica i els estàndards en els testimonis pericials (Sanders 2007; Butters 2012; Solan 2013) amb l'objectiu de millorar les garanties que els peritatges compleixin els requisits legals d'admissibilitat.

Malgrat tot, hi ha encara tota una sèrie de reptes de futur que passen inicialment per una doble via: d'una banda, cal continuar millorant les tècniques d'anàlisi per obtenir resultats cada cop més fiables. Això implica, per exemple, dissenyar noves metodologies d'anàlisi i poder disposar de més coneixement sobre l'ús lingüístic de més grups poblacionals, de manera que la construcció i explotació de corpus cada vegada més grans i més representatius és un dels reptes més immediats que cal afrontar. Amb aquest coneixement es podran generalitzar tècniques com l'aplicació dels entorns bayesians per a l'anàlisi de tot tipus de texts (orals i escrits). En aquest sentit, cal reforçar el treball conjunt entre lingüistes i enginyers. D'altra banda, cal adaptar-se als canvis en la manera d'utilitzar la llengua que es van popularitzant cada vegada més ràpidament de la mà de les noves tecnologies (per exemple, l'ús dels missatges d'SMS va minvant ràpidament a favor d'altres sistemes de missatgeria instantània com els tuits o missatges de Whatsapp).

Bibliografia

- BALDWIN, John. «Phonetics and speaker identification». *Medicine, Science and the Law*, 19, 1979, p. 231–232.
- BALDWIN, John; French, Peter. *Forensic Phonetics*. Londres: Pinter, 1990.
- BEL, Núria, QUERALT; Sheila, SPASSOVA; Maria; TURELL, M. Teresa. «The use of sequences of linguistic categories in forensic written text comparison revisited». A: *Proceedings of the International Association of Forensic Linguists' Tenth Biennial Conference*, 2012, p. 192-209.
- BURROWS, John. *Computation into Criticism: a study of Jane Austen's novels and an experiment in method*. Oxford: Clarendon Press, 1987.
- BUTTERS, Ronald. «Retiring President's closing address: ethics, best practices, and standards». A: S. Tomblin, N. MacLeod; R. Sousa-Silva; M. Coulthard (ed.), *Proceedings of the Tenth International Association of Forensic Linguists' Biennial Conference*. Birmingham: Aston University, 2012, p. 351–361.

CHASKI, Carol. «Empirical evaluations of language-based author identification techniques». *Forensic Linguistics. The International Journal of Speech, Language and the Law*, 8, 1, 2001, p. 1-65.

CICRES, Jordi. [*Aplicació de l'Anàlisi de l'entonació i de l'alienació tonal a la identificació de parlants en fonètica forense*](#) [tesi doctoral en línia]. Barcelona: Universitat Pompeu Fabra, 2007a.

CICRES, Jordi. «Análisis discriminante de un conjunto de parámetros fonético-acústicos de las pausas llenas para identificar hablantes». *Síntesis tecnológica*, 3, 2, 2007b, p.87-96.

CICRES, Jordi; TURELL, M. Teresa. «Short and long-term variation in intonation patterns: a preliminary study for speaker identification», comunicació presentada al *7th Biennial Conference on Forensic Linguistics/ Language and Law*, Cardiff University, Cardiff, 2005.

COULTHARD, Malcolm. «Author identification, idiolect, and linguistic uniqueness». *Applied Linguistics*, 25, 4, 2004, p. 431-447.

COULTHARD, Malcolm; JOHNSON, Alison. *An introduction to forensic linguistics*, Londres i Nova York: Routledge, 2007.

DELGADO, Carlos. [*La identificación de locutores en el ámbito forense*](#) [tesi doctoral en línia]. Madrid: Facultad de Ciencias de la Información, Universidad Complutense de Madrid, 2001.

ERIKSSON, Anders; WRETTLING, Pär. «How flexible is the human voice? – A case study of mimicry». *Proceedings Eurospeech '97*, 2, Rhodes, 1997, p. 1043–1046.

FERGUSON, Charles. «Phonology as an Individual Access System: Some Data from Language Acquisition». A: Ch. Fillmore, D. Kempler i W. Want (ed.), *Individual differences in language ability and language behaviour*, Nova York, San Francisco, Londres: Academic Press, 1979, p. 189-201.

GAVALDÀ, Núria. [*Index of Idiolectal Similitude for the Phonological Module of English applied to Forensic Speech Comparison*](#) [tesi doctoral en línia]. Barcelona: Universitat Pompeu Fabra, 2013.

GIBBONS, John; TURELL, M. Teresa. *Dimensions of Forensic Linguistics*. Amsterdam/Philadelphia: John Benjamins, 2008.

GOLD, Erica; FRENCH, Peter «International practices in forensic speaker comparison». *The International Journal of Speech, Language and the Law*, 18, 2, 2011, p. 293-307.

GRANT, Timothy; BAKER, Kevin. «Identifying reliable, valid markers of authorship: a response to Chaski». *Forensic Linguistics. The International Journal of Speech, Language and the Law*, 8, 1, 2001, p. 66-79.

GRANT, Timothy. «Quantifying evidence in forensic authorship analysis». *The International Journal of Speech, Language and the Law*, 14, 1, 2007, p. 1-25.

GRANT, Timothy. «TXT 4N6: Idiolect free authorship analysis?», dins M. Coulthard i A. Johnson (ed.), *The Routledge Handbook of Forensic Linguistics*, Londres: Routledge, 2010, p. 508-522.

HOLMES, David. «Authorship attribution». *Computers and the Humanities*, 28, 2, 1994, p. 87-106.

JOHNSTONE, Barbara. *The Linguistic Individual: Self-Expression in Language and Linguistics*. Oxford: Oxford University Press, 1996.

McMENAMIN, Gerald. *Forensic Linguistics. Advances in Forensic Stylistics*. CRC Press, 2002.

MILROY, James; MILROY, Lesley. «Speech community and language variety in Belfast». *Report to the Social Science Research Council*, Londres, 1977.

MILROY, James; MILROY, Lesley. «Belfast: change and variation in an urban vernacular». A: P. Trudgill (ed.), *Sociolinguistic patterns in British English*. Londres: Eduard Arnold, 1978, p. 19-36.

NOLAN, Francis. *The Phonetic bases of speaker recognition*. Cambridge: CUP, 1983.

NOLAN, Francis. «Speaker recognition and forensic phonetics». A: W. Hardcastle i J. Laver (ed.), *The Handbook of Phonetic Sciences*. Oxford: Wiley-Blackwell, 1997, p. 744-767.

NOLAN, Francis. «Intonation in speaker identification: an experiment on pitch alignment features». *Forensic Linguistics*, 9, 1, 2002, p. 1-21.

PAYNE, Arvilla. «Factors controlling the acquisition of the Philadelphia dialect by out-of-state children». A: W. Labov (ed.), *Locating language in time and space*. Nova York: Academic Press, 1980, p. 143-178.

ROSE, Phillip. *Forensic speaker identification*. Londres: Taylor & Francis, 2002.

SANDERS, Joseph. «Expert Witness Ethics». *Fordham Law Review*, 76, 3, 2007, p. 1539-1584.

SOLAN, Lawrence. «Intuition versus Algorithm: The Case of Forensic Authorship Attribution». *Brooklyn Journal of Law and Policy*, 21, 551, 2013, p. 551-576.

SPASSOVA, Maria; TURELL, M. Teresa. «The use of morpho-syntactically annotated tag sequences as forensic markers of authorship attribution». A: M.T. Turell, M.S. Spassova i J. Cicres (ed.), *Proceedings of the Second European IAFL Conference on Forensic Linguistics, Language and the Law*. Barcelona: Publicacions de l'ÍULA, 2007, p. 229-237.

TURELL, M. Teresa. «Plagiarism». A: M. T. Turell i J. Gibbons (ed.) *Dimensions of Forensic Linguistics*. Amsterdam/ Philadelphia: John Benjamins, 2008, p. 265-299.

TURELL, M. Teresa. «The use of textual, grammatical and sociolinguistic evidence in forensic text comparison». *The International Journal of Speech, Language and the Law*, 17, 2, 2010, p. 211-250.

TURELL, M Teresa; GAVALDÀ, Núria. «Towards an Index of Idiolectal Similitude (or Distance) in Forensic Authorship Analysis». *Journal of Law and Policy*, núm. 21, 2013, p. 495-514.

WOOLS, David; COULTHARD, Malcolm. «Tools for trade». *Forensic Linguistics. The International Journal of Speech, Language and the Law*, 5, 1, 1998, p. 33-57.

ZETTERHOLM, Elisabeth. «Intonation Pattern and Duration Differences in Imitated Speech». *Proceedings of Speech Prosody 2002*, 2002, p. 731-734, <http://www.lpl.univ-aix.fr/sp2002/pdf/zetterholm.pdf>