

Podem elaborar perfils sociolingüístics forenses a partir de corpus discursius d'àmbit escolar?

Sheila Queralt i Roser Giménez

Laboratorio SQ—Lingüistas Forenses i CUSC—Centre de Recerca en Sociolingüística i Comunicació, Universitat de Barcelona

El discurs d'un individu ens pot aportar informació sociolingüística sobre ell o ella: quants anys té? És noi o noia? D'on és? Quina és la seva llengua inicial? Fins i tot, és possible definir un conjunt de trets lingüístics propis que caracteritzen aquest individu i el diferencien de la resta? En aquest article s'analitza amb finalitats forenses el corpus discursiu del projecte de recerca educativa RESOL, amb l'objectiu de determinar variables que mostrin diferències per edat, sexe, perfil geogràfic i individu.

mots clau: lingüística forense, sociolingüística, perfils sociolingüístics

La lingüística forense és el punt de trobada entre la llengua i el dret. És una ciència multidisciplinària que es forneix d'un ampli ventall de camps de coneixement, com el dret, l'enginyeria, l'estadística, la lingüística computacional, la lingüística teòrica o la lingüística descriptiva i, especialment, la sociolingüística.

Entre moltes altres tasques, els lingüistes forenses elaboren perfils sociolingüístics a partir de mostres orals o escrites, que poden usar-se en investigacions privades per reduir el nombre de sospitosos i en investigacions policials per reconduir la cerca de sospitosos. L'elaboració d'aquests perfils consisteix en l'anàlisi de textos o de gravacions d'autoria desconeguda per tal d'identificar-hi trets lingüístics relacionats amb característiques sociolingüístiques de l'autor. Alguns dels trets socioindividuals dels parlants que es poden determinar a partir

de l'ús de la llengua són la professió, l'edat, el sexe, el nivell educatiu, la religió, la ideologia política, la llengua inicial o el contacte de llengües.

En investigacions encara en marxa, s'utilitza el corpus del projecte RESOL de la Universitat de Barcelona per identificar característiques lingüístiques concretes que puguin ser útils a l'hora d'elaborar perfils sociolingüístics en contextos forenses. El corpus RESOL és un conjunt de mostres de llengua oral escolar, recollides longitudinalment entre 2006 i 2010 en diversos centres educatius de territoris de parla catalana. La part del corpus que s'utilitza en les investigacions que tractem aquí inclou una de les dues proves experimentals de competència lingüística que es van dur a terme, l'objectiu original de la qual era mesurar les capacitats interactives dels joves en un con-

text de negociació interpersonal. Concretament, s'utilitza material recollit quan els participants cursaven 1r i 4t d'ESO a dues de les poblacions estudiades —Manlleu i Mataró.

L'objectiu forense que es persegueix en les noves investigacions és identificar variables lingüístiques que permetin classificar les mostres segons l'edat —12 anys a 1r i 16 anys a 4t—, el sexe, l'àrea geogràfica —una de les dues ciutats— i l'estil idiolectal dels individus.

El concepte d'*estil idiolectal* (Turell 2010) es basa en el coneixement científic segons el qual la producció lingüística d'un individu, ja sigui oral o escrita, es correlaciona amb trets sociointividuals i sociocol·lectius del parlant, de manera que cada individu mostra un estil únic. A més, les característiques lingüístiques de cada producció són irrepetibles i generalment, sobretot en contextos informals, els parlants paren molta atenció a la manera com utilitzen la llengua, cosa que indica que no són conscients de com aquestes relacions entre característiques lingüístiques i trets individuals es poden usar per lligar-los amb les seves produccions. De manera semblant, aquests trets també passen desapercebuts a altres individus que tinguin la intenció de plagiar, imitar o copiar mostres orals o escrites alienes.

Tant la lingüística forense en sentit ampli com l'elaboració de perfils sociolingüístics beuen de diferents subdisciplines de la sociolingüística, entre les quals destaca la sociolingüística variacionista, que sosté que la variació és inherent a totes les llengües i a tots els nivells lingüístics i que, a més, no és casual, sinó que està subordinada a factors tant lingüístics (interns)

com socials (externs). Des d'aquesta perspectiva, les comunitats de parla són heterogènies, però estan estructurades de forma sistemàtica.

La variació és un dels conceptes clau en lingüística forense. Concretament, es parla de la variació interparlant i de la variació intraparlant. La premissa que fa possible l'elaboració de perfils i altres tasques en lingüística forense és la que afirma que la variació present a les mostres lingüístiques d'un individu és menor que la variació existent entre les mostres de diversos individus. En aquests termes, l'estil idiolectal es pot concebre com el ventall de possibles eleccions lingüístiques que respon a la variació intraparlant d'un individu.

Per assolir l'objectiu forense de la primera de les investigacions en marxa sobre el corpus RESOL s'analitzen diverses variables de nivell lèxic, morfosintàctic i pragmàtic. Les investigacions parteixen d'hipòtesis sobre quines variables lingüístiques podran relacionar-se amb l'edat, el sexe, l'àrea geogràfica i l'estil idiolectal dels parlants.

Segons la bibliografia existent s'espera que les variables lingüístiques que mostrin resultats més diferents entre els dos grups d'edat siguin aquelles relacionades amb l'augment d'habilitats lèxiques i sintàctiques amb el temps. Per exemple, d'una banda, l'increment de la fluïdesa lèxica, un augment de l'ús de substantius abstractes, de termes polisèmics o d'adverbis de probabilitat i, de l'altra, un augment de l'ús d'oracions passives, subordinades, coordinades o relatives, així com un increment del nombre de paraules per torn.

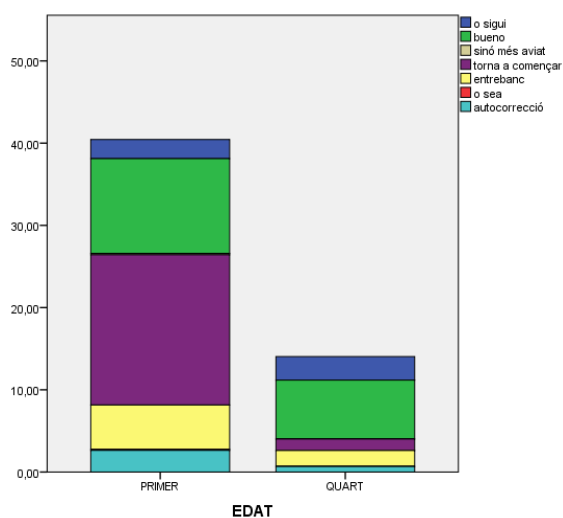
Quant a les variables que s'espera que resultin més útils per identificar el sexe dels individus, la bibliografia apunta a un ús més elevat de lèxic no estàndard per part dels nois i també a un cert avantatge en l'adquisició lingüística de les noies respecte els nois, que podria veure's en l'ús de marcadors interaccionals, la fluïdesa lèxica o la complexitat i longitud oracional. A més, també serà especialment interessant per a diverses aplicacions forenses observar si hi ha diferències entre els sexes quant a la selecció i l'ordre d'adjectius a l'hora de fer descripcions.

Pel que fa a les possibles variables relacionades amb diferències individuals, cal tenir en compte la gran diversitat de factors que podrien explicar-les. Per exemple, podem observar diferències individuals en la competència o actituds lingüístiques, en el grau de familiaritat amb diferents situacions comunicatives, o en preferències personals, entre molts altres factors no lingüístics. Algunes de les variables lingüístiques que podrien relacionar-se amb aquests factors inclouen l'ús de lèxic no estàndard, preferències per certs tipus d'adverbis, l'ús de marcadors discursius i pragmàtics, el nombre de paraules per torn, les preferències en l'ús de formes sinònimes —com *tindre* o *tenir*, *dins* o *dintre*, etc.— o la posició dels pronoms febles abans o després del verb.

Tot i que encara no disposem de resultats definitius, les anàlisis preliminars mostren certes tendències. Per exemple, en primer lloc, sembla que la variable «lapsus», és a dir, com assenyalen l'explicació d'un mot que en el

aquell moment no són capaços de produir, podria tenir correlacions amb l'edat dels participants. Com mostra el *Gràfic 1*, els parlants de 16 anys tenen menys ocurrencies d'aquesta variable —segurament, perquè tenen un ventall de vocabulari més ampli que els menors de 12 anys— i, a més a més, també s'observa que mantenen la preferència pel marcador *bueno* però disminueix considerablement l'estratègia de tornar a començar el segment discursiu que estaven produint.

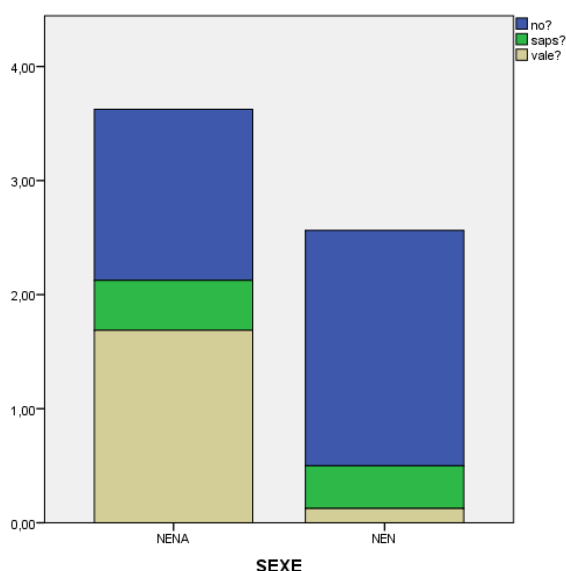
Gràfic 1. Variable «lapsus» segons l'edat del menor



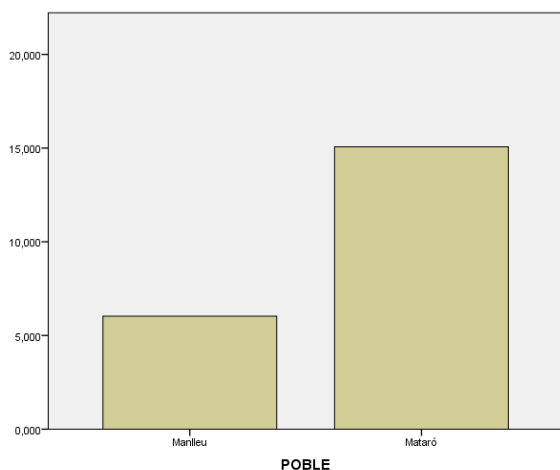
Quant al sexe dels parlants, sembla que afecta la variable «marcadors discursius de control». El *Gràfic 2* mostra la tendència dels nois a utilitzar amb menys freqüència els marcadors de control. A més a més, els nens mostren una alta preferència pel marcador *no?* mentre que les nenes usen amb freqüències similars els marcadors *no?* i *vale?*

Els resultats preliminars també semblen apuntar que variables com l'ús de mots híbrids com *llàmpara* o *escaleres* o préstecs del castellà com *jaula* o *lechuza* podrien interactuar amb l'àrea geogràfica. Els parlants de Manlleu solen utilitzar menys mots híbrids que els parlants de Mataró (Gràfic 3). Aquests resultats podrien explicar-se per diferències en el domini i l'ús del castellà i del català entre els parlants d'una i altra població.

Gràfic 2. Variable «marcador de control» segons el sexe del menor

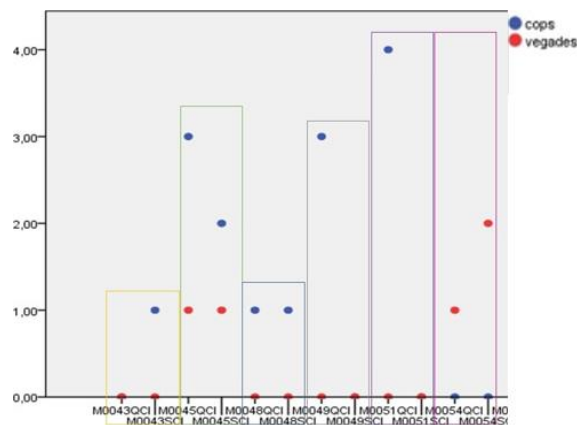


Gràfic 3. Variable «mots híbrids o préstecs» segons el perfil geogràfic del menor



Pel que fa a l'idiòlecte, la variable que representa l'opció del parlant entre l'ús de *cops* o *vegades* podria servir per identificar certs individus, segons les anàlisis preliminars. Per exemple, el Gràfic 4 mostra els individus que pronuncien una de les dues paraules i la freqüència en què l'utilitzen a la mostra de 1r i a la mostra de 4t d'ESO. En aquest gràfic és molt destacable el parlant M0054 —últim parlant, representat destacat en lila fluix— perquè mostra una tendència diferent a la resta. Aquest parlant mai utilitza *cops*. També és destacable el parlant M0045 —segon parlant, destacat en verd— perquè utilitza ambdues opcions, cosa que no fa cap altre parlant analitzat fins al moment. També es destacable el parlant M0048 —tercer parlant, destacat en blau— el qual presenta la mateixa distribució en la mostra de primer i de quart. En general, es pot observar que aquesta variable no mostra tendències globals, sinó que mostra una tendència pròpia per a cada parlant i, per tant, forma part dels trets que podrien caracteritzar l'estil idiolectal d'un parlant concret.

Gràfic 4. Variable «cops/vegades» per individu



Aquests resultats preliminars mostren com les noves investigacions poden tenir un gran nombre d'aplicacions en contextos forenses, ja que contribueixen directament a l'elaboració de perfils sociolingüístics d'adolescents en català i a la millora de la metodologia empleada en entrevistes forenses a menors. A més, les variables identificades en la primera investigació poden utilitzar-se, entre altres, per contribuir a investigacions de ciberassetja-

ment, a la suplantació policial d'identitats, a la comunicació en entrevistes amb menors o a l'anàlisi del discurs de menors amb l'objectiu de detectar possibles manipulacions.

Referències

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

Com citar aquest apunt

QUERALT, Sheila; GIMÉNEZ, Roser (2019). «Podem elaborar perfils sociolingüístics forenses a partir de corpus discursius d'àmbit escolar?». *Apunts de sociolingüística i política lingüística*, 5, 1-5

Editors

Xavier Mas Craviotto i Avel·lí Flors-Mas

Amb la col·laboració de:



Generalitat de Catalunya
**Departament
de Cultura**

Contacte

CUSC-Centre de Recerca en Sociolingüística i Comunicació, Universitat de Barcelona

Adreça: Gran Via de les Corts Catalanes, 585. 08007 Barcelona

Telèfon: +34 93 403 7065

Adreça-e: cusc@ub.edu

Web: <http://www.ub.edu/cusc>

Twitter: @CuscUB

Facebook: @cuscub