

ELS PROJECTES DE CODIFICACIÓ DE LA LLENGUA DE SIGNES CATALANA (LSC)

Josep QUER
Institut d'Estudis Catalans
ICREA-Universitat Pompeu Fabra

1. LA LLENGUA DE SIGNES CATALANA (LSC) A L'IEC

Des de fa una desena d'anys, la Secció Filològica de l'Institut d'Estudis Catalans ha incorporat de manera progressiva, però decidida, la llengua de signes catalana als àmbits on desenvolupa la seva activitat acadèmica i científica. Com és habitual amb les llengües de modalitat visuogestual, no ha estat fins fa relativament poc que s'han començat a percebre com a llengües naturals en el sentit ple del terme tant per part de la nostra societat com dels professionals de l'estudi de la llengua, tot superant l'estigma de la discapacitat que les ha minoritzat i les ha fet invisibles al llarg dels segles. Hi han contribuït, d'una banda, els avenços en el coneixement de les propietats lingüístiques d'aquest tipus de llengües durant el darrer mig segle, i, de l'altra, les reivindicacions dels seus drets lingüístics i cívics per part del moviment associatiu de persones sordes signants. Catalunya ha estat un bon exemple d'aquesta confluència d'objectius d'actors acadèmics i socials per tal d'aconseguir fites compartides com el reconeixement de l'estatus de llengua i dels drets dels seus usuaris. La Secció Filològica ha tingut clar des d'un bon començament que la seva activitat en l'àmbit de l'LSC havia de comptar amb la col·laboració de la Federació de Persones Sordes de Catalunya (FESOCA), i això ha quedat reflectit en diverses ocasions com la signatura de convenis, l'organització conjunta dels Seminaris LSC a la seu de l'IEC o la compartició de materials preexistents sobre la llengua.

La primera acció de visibilització del coneixement sobre l'LSC va consistir en la creació del *Portal de la llengua de signes catalana* de l'IEC, presentat el 2012 en versió catalana, castellana i anglesa (<http://blogs.iec.cat/lsc/>). L'objectiu d'aquest lloc web és posar a disposició d'usuaris i investigadors el coneixement sobre la llengua, perquè n'esdevingui el referent pel que fa als continguts lingüístics. En una primera fase s'hi van publicar els següents components: la *Gramàtica bàsica LSC* (Quer *et al.* 2005) en llengua de signes catalana, català, castellà i anglès, una base de dades bibliogràfica sobre l'LSC, l'alfabet dactilològic català, un recull de textos legislatius rellevants, un llistat

d'enllaços d'interès, i la presentació del projecte de constitució del corpus de l'LSC. La perspectiva és d'anar incorporant-hi materials en la mesura que estiguin disponibles per a la publicació en línia i, per tant, s'ampliarà amb els projectes de codificació que l'Institut està duent a terme i que presentarem en els apartats següents.

Des del punt de vista del reconeixement legal de l'LSC, l'Institut va ser una de les entitats convocades per la comissió de les Corts Espanyoles encarregada de tramitar la que esdevindria la llei 27/2007 de reconeixement de les llengües de signes espanyola i catalana. A més, va estar implicada directament en la redacció del text de la llei 17/2010, del 3 de juny, de la llengua de signes catalana, promulgada pel Parlament de Catalunya, que la reconeix com a llengua pròpia de Catalunya i en defensa el coneixement, la recerca, la difusió i l'ús en els diversos àmbits públics. Dins aquest nou marc legislatiu, l'IEC adquireix el paper d'autoritat normativa en el procés de normalització de l'LSC, així com d'entitat vertebradora de la recerca en aquest àmbit.

Tot i que la recerca en LSC ha crescut significativament en els darrers anys, el coneixement que en tenim és encara parcial i fragmentari, i els materials d'estudi i ensenyament disponibles són d'abast i qualitat variables. El fet que el punt de partida no és ni de lluny comparable al del català o de llengües semblants quan iniciaven una etapa de normalització impedeix engregar de manera immediata un procés de normativització de la llengua i s'imposa contribuir a consolidar i ampliar la descripció i codificació de la llengua primer de tot. Només un cop puguem disposar d'aquesta base sòlida serà raonable abordar l'establiment d'un estàndard i d'una norma per a l'LSC. Tanmateix, com que les necessitats dels usuaris no es poden deixar de banda fins que s'hagi completat aquest procés, l'IEC també s'ha proposat donar-hi resposta a mig termini. Els dos tipus d'iniciativa es concreten en dos projectes de naturalesa diferent, però estretament vinculades i coordinades: el projecte de constitució del corpus de l'LSC, i el projecte de lèxic i gramàtica.¹ A continuació s'exposaran les seves característiques i el seu desenvolupament.

2. EL PROJECTE DE CONSTITUCIÓ DE CORPUS DE L'LSC

Una de les mancances bàsiques a l'hora d'abordar la descripció de l'LSC és que no tenim una visió global de la variació que es dona dins de la llengua, i això és en part degut al fet que no en tenim registres —i molt menys una sistematització d'aquests registres— anteriors a la generalització de les tècniques d'enregistrament audiovisual com el vídeo. Alhora, el fet que la comunitat de signants més gran a Catalunya es trobi a Barcelona i que la majoria de materials es produeixin en aquesta ciutat ha determinat que sigui la va-

1. El projecte de constitució del corpus de l'LSC està coordinat per la Dra. Gemma Barberà i compta amb l'expert d'LSC Santiago Frigola. El projecte de lèxic i gramàtica està coordinat per la Dra. Eulàlia Ribera. Com a director dels projectes de codificació de l'LSC de l'IEC, aprofito l'oportunitat per agrair-los públicament la seva dedicació i la seva expertesa, sense les quals aquests projectes no es podrien dur a terme.

riant més coneguda i més influent dins el territori.² D'altra banda, també desconeixem la variació que existeix entre les diferents generacions de signants, que és considerable com a conseqüència dels canvis en l'escolarització de les persones sordes en les últimes dècades (que ha passat de les escoles de sords, de vegades amb internats, al model d'integració en aules normals) i en els hàbits de socialització de la comunitat de signants (molt centrats en la vida associativa fa uns quants anys i més influïts ara per les possibilitats de comunicació a distància en temps real i d'accés a la informació a través d'Internet). A més, l'escolarització en centres específics per a sords i segregats per a nens i per a nenes també ha creat un factor de variació en relació amb les escoles i amb el gènere de determinades generacions de signants que mai ha estat documentat adequadament. A més, com que es tracta d'una llengua que s'ha transmès tradicionalment de manera presencial i en contextos de comunicació informals, no compta amb un repertori complet de registres d'ús com altres llengües que ocupen tots els àmbits públics de la vida social. Tot i això, segons de quin gènere discursiu es tracti sí que es poden identificar diferències.

Davant aquestes mancances, era ben clar que calia obtenir una instantània global de la llengua tal com s'usa actualment a tot el territori i la manera de fer-ho era creant un corpus representatiu de discursos signats que recollissin una mostra significativa de la variació que es dona en els diversos paràmetres. Amb aquesta finalitat l'IEC va presentar el projecte de constitució del corpus de l'LSC a la Direcció General de Política Lingüística, que va aconseguir un finançament inicial de l'Obra Social «la Caixa» l'any 2012-2013. Aquest finançament va permetre dur a terme un projecte pilot i ha continuat sense interrupció fins al moment actual. No és el primer projecte de corpus de llengua de signes que es desenvolupa, i per tant hem pogut comptar amb l'expertesa derivada de l'experiència acumulada en altres projectes completats o en curs, com els de les llengües de signes dels Països Baixos, de la Bèlgica francòfona, britànica, alemanya, italiana o australiana. La informació compartida sobre aquests precedents ha estat molt valuosa a l'hora de dissenyar el projecte de corpus de l'LSC.³

Tal com apuntàvem més amunt, l'objectiu del projecte és establir les bases d'un corpus de referència de l'LSC que ofereixi una primera caracterització de la variació existent dins el seu domini d'ús, Catalunya. A fi de comptar amb dades comparables, tant des del punt de vista lingüístic com tècnic, es va decidir d'enregistrar-les totes seguint uns criteris unificats, tal com s'ha fet en altres projectes de corpus de llengües de signes. Els paràmetres de variació que es planteja de recollir són els següents:

- a) Variació geogràfica: es recullen dades de 6 poblacions o àrees representatives del territori: Terrassa, Girona (Palafrugell i Blanes), Lleida, Barcelona, Reus i Vic/Manresa.
- b) Variació intergeneracional: les dades cobreixen tres franges d'edat (signants joves 18-30, de mitjana edat 32-50 i d'edat avançada 51-80).
- c) Variació de discurs: els discursos recollits són o bé induïts o semiespontanis, a partir de tasques de diverses característiques que es proposen als signants.

2. Penseu, per exemple, en l'impacte que té des de fa anys el canal de televisió en LSC per Internet Webvisual.tv sobre tota la comunitat de signants, pel que representa com a via d'accés en llengua de signes a informacions d'actualitat i de caràcter general.

3. Per a més referències sobre el projecte pilot del corpus de l'LSC, vegeu Barberà *et al.* (2016).

d) Variació de gènere: cada parella de signants enregistrats està formada per una dona i un home.

Per a cada localització d'enregistrament, s'escullen tres parelles de signants de les tres franges d'edat previstes i dels dos sexes. Per tant, la versió final d'aquesta fase de constitució del corpus comptarà amb les dades de 42 signants (a Barcelona se'n recolliran de dues associacions). Atesa la gran variació en l'edat d'adquisició de la llengua de signes entre els signants, se seleccionen signants nadius o quasi nadius, és a dir, que hagin adquirit l'LSC durant els primers anys de vida i l'hagin usat com a llengua principal al llarg de la seva vida. Els contactes amb els signants i la seva selecció sempre es vehiculen a través de l'associació de persones sordes de la localitat on es volen recollir les dades, i més concretament mitjançant una persona de contacte que fa de mediadora i que en alguns casos condueix les sessions d'enregistrament. Aquestes tenen lloc generalment a les mateixes associacions, per tal d'evitar un entorn estrany als signants que pugui influir negativament en la naturalitat de les dades. Als signants que mostren interès a participar en el projecte, se'ls explica en què consisteix i quina serà la seva participació. Així mateix, se'ls demana autorització per a la utilització d'aquestes dades dins el projecte de corpus i per a la seva publicació en línia. Amb això ens plantejem complir amb els requisits ètics estàndards en la recerca amb subjectes humans i amb el tractament de les dades. Hem de tenir present que, per la naturalesa visual del discurs signat, no es pot deslligar la identitat del signant de les dades que produeix, encara que no es vinculin a les seves dades personals identificadores.

Tal com dèiem més amunt, la recollida de dades es planteja reflectir les possibles diferències segons el gènere discursiu produït. Amb aquesta finalitat es van dissenyar vuit activitats per facilitar la producció signada, que s'utilitzen a totes les sessions d'enregistrament. Les activitats són les següents: presentació individual i explicació de la motivació del signe característic personal (*signom*); explicació d'una història en vinyetes; narració d'un fragment de dibuixos animats; narració d'una anècdota personal vinculada al fet de ser sord; jocs vinculats als camps semàntics dels aliments, de les parts del cos humà i dels colors; narració d'un vídeo mut, i debat sobre un tema d'interès. Les sessions es duen a terme en parelles d'un home i una dona de la mateixa franja d'edat, que seuen gairebé un enfront de l'altre per a crear una situació d'interacció més o menys natural. El conductor de la sessió sempre és un signant sord (l'expert sord del projecte o el mediador local) per a evitar qualsevol interferència de la llengua oral i se situa davant de l'escenari de filmació. El seu paper és el de facilitar la realització de les tasques amb la mínima intervenció. La sessió s'enregistra amb dues càmeres d'alta definició que permeten extreure un pla mitjà i un pla de detall de cada signant, la qual cosa serveix per a diferents nivells de detall en l'anotació posterior. Les sessions duren una hora i mitja aproximadament i els participants reben una petita remuneració per la seva feina de col·laboració.

Un cop recollides les dades, es passa al procés d'edició i anotació. L'anotació és l'etapa més exigent pel que fa a temps i recursos humans. L'eina que s'utilitza és el programa ELAN (Eudico Linguistic Annotator),⁴ desenvolupat al *Max Planck Institute for Psycholinguistics* de Nimega, que permet alinear de manera sincrònica les anotacions

4. <https://tla.mpi.nl/tools/tla-tools/elan/>.

amb el vídeo que conté les dades. Tot i que no va ser dissenyat específicament per a l'anotació de dades de llengües de signes, ha esdevingut l'eina més emprada en aquesta àrea, donada la seva flexibilitat i facilitat d'ús. A més, la interfície d'usuari està disponible en català i castellà, la qual cosa permet que sigui emprada per anotadors que no dominen l'anglès. Un altre avantatge important és que permet fer cerques dins les dades anotades, i els resultats d'aquestes cerques sempre estan alineats amb les dades reals.

Les tasques d'anotació les duen a terme signants nadius d'LSC que, o bé ja tenen experiència en recerca lingüística, o bé han fet sessions d'entrenament. Es tracta d'una feina laboriosa, lenta i complexa, i és la part del projecte que suposa la despesa més important del pressupost. Dins el projecte pilot es va dur a terme una prova per a comparar les dues opcions de fer una anotació complexa, on s'anotaven, per exemple, tots els trets no manuals, o fer-ne una de simple que recollia el contingut lèxic executat per cada mà, i una traducció. Amb els recursos disponibles vam arribar a la conclusió clara que només es podia aspirar a oferir una anotació simple. Evidentment, aquesta es podrà completar en el futur en la mesura que es pugui finançar, o que d'altres investigadors la duïguin a terme i estiguin disposats a publicar-la.

Com en l'anotació de qualsevol corpus, és important revisar la coherència entre els resultats produïts per diferents anotadors: tot i que s'han establert uns criteris clars des d'un bon començament, de vegades les dades presenten problemes nous que no s'havien previst a l'avançada i que l'anotador resol segons el seu propi criteri. Per aquest motiu es fa sempre una revisió sistemàtica de les anotacions per part d'un col·laborador que n'examina la coherència global. Existeix, però, un problema de fons específic per a la llengua de signes: no existeix una glossa⁵ estàndard establerta per a cada signe, perquè no disposem d'un diccionari de referència i moltes vegades s'usen glosses diferents per a un mateix signe, o no està clar el criteri per a diferenciar signes relacionats per la forma com a variants d'un de sol o com a signes diferents. Un mètode per a compensar aquestes mancances és l'elaboració d'un vocabulari controlat, és a dir, una llista que conté totes les glosses que es van introduint i que els anotadors poden consultar per a assegurar-se de la coherència de les anotacions que ells mateixos introdueixen. Aquest és un aspecte decisiu que connecta de manera directa amb el projecte de base de dades lèxica al qual tornarem més avall.

Fins ara s'han recollit dades de tres localitzacions diferents, i tenim previst recollir les de les tres restants aquest mateix any 2016. Tot i que les tasques d'anotació requeriran aproximadament tres anys més perquè es puguin completar, està previst de posar a disposició del públic les dades recollides (sense anotacions) al llarg del curs 2016-17. Amb això no només es visibilitzarà el progrés del projecte de corpus, sinó que a més les dades podran començar a ser utilitzades per a finalitats que no requereixin anotació (usos didàctics, etc.).

5. La glossa és un mot de la llengua oral que tradueix el significat més prominent d'aquell signe.

3. EL PROJECTE DE BASE DE DADES LÈXICA I DE GRAMÀTICA DE L'LSC

Davant la necessitat de posar materials a l'abast dels usuaris, educadors i investigadors de l'LSC dins un termini més curt, l'IEC es va plantejar la possibilitat d'engegar un projecte de documentació lèxica i gramatical de la llengua que donés resposta a les necessitats més immediates. Amb aquest objectiu va ampliar l'abast del projecte de codificació de la llengua de signes catalana amb un d'específic sobre lèxic i gramàtica. Aquest nou projecte compta igualment amb finançament addicional de l'Obra Social «la Caixa» a través de la Direcció General de Política Lingüística de la Generalitat de Catalunya, i s'ha pogut iniciar a començament de 2016.

L'objectiu inicial d'aquest nou projecte és desenvolupar una base de dades lèxica que serveixi de base sòlida per al desenvolupament del que serà el lèxic normatiu de l'LSC. En una primera fase, es recollirà el lèxic bàsic de la llengua a partir de les obres lexicogràfiques publicades, tant en format paper (p. ex. Perelló / Frigola, 1997) com digital (p. ex. ILLESCAT 2004), i dels reculls existents en línia (p. ex. *Mira què dic*⁶). Aquests materials són de naturalesa i cobertura molt diversa i per tant cal fer-ne un buidatge crític. Tot i que la majoria recullen les variants lèxiques de la varietat de Barcelona, cal tenir en compte que Martín / Alvarado (2004), per exemple, reflecteixen la varietat de l'LSC de Ponent (Lleida). D'altra banda, existeixen alguns materials específics creats com a propostes neològiques en àrees concretes del coneixement (DOMAD 2002a, b), i per tant s'haurà d'avaluar fins a quin punt aquestes propostes han estat acceptades per la comunitat d'usuaris i han passat a formar part del lèxic de l'LSC.

Ahora, el desenvolupament del projecte de corpus, com dèiem abans, també genera un llistat dels ítems lèxics que apareixen en els discursos signats que el constitueixen i que anomenem vocabulari controlat. És especialment rellevant intentar connectar les dades lèxiques dels dos projectes (vg. Navarrete, 2005): d'una banda, perquè la base de dades lèxica ajuda a enriquir el vocabulari controlat i a assegurar-ne la coherència al llarg de tot el procés d' anotació de les dades; de l'altra, perquè la vinculació de les dues bases de dades permetrà accedir a exemples reals d'ús dels ítems lèxics dins el corpus, de manera paral·lela a com s'obtenen les concordances dins un corpus de llengua oral. Si la informació del projecte de corpus i de la base de dades lèxica està interconnectada, es recolliran a la base de dades totes les variants lèxiques que hagin aparegut en els discursos recopilats. Aquesta connexió entre els dos components representa un repte de disseny. Només ha estat implementat en part en alguns dels projectes d'altres llengües de signes com el del de la llengua de signes de la Bèlgica francòfona (LSFB) o l'australiana (AUSLAN) i serà per tant una aportació important en aquesta àrea de la lingüística descriptiva i de corpus.

Un dels aspectes centrals en el desenvolupament de la base de dades lèxica és determinar quines categories d'informació es codificaran per a cada lema, però també com s'organitzarà aquesta informació i com es podran realitzar les cerques. Gairebé sense excepció, els reculls lèxics d'LSC estan organitzats com a vocabularis bilingües LSC-ca-

6. <http://www.edu365.cat/mqdl/>.

talà/castellà, i la cerca normalment ha de passar per la glossa o el mot de la llengua oral que tradueix el significat més destacat del signe. Aquesta és una barrera que desapareix en una estructura de base de dades, atès que en principi la cerca d'informació avançada s'haurà de poder fer a partir de qualsevol dels trets codificats per a cada lema. La interfície de consulta estarà disponible en LSC, català, castellà i anglès.

Cada lema contindrà la informació següent (cf. Ribera, 2007; Barberà / Ribera, 2010):

- Forma del signe en format vídeo i en format *gif*
- Glossa (en català, castellà i anglès)
- Fonts on es troba documentat el signe
- Definició de l'ítem que apareix en el lema, en LSC i en català
- Característiques semàntiques: accepcions, aparició en expressions fixes, etc.
- Categoria gramatical
- Característiques fonològiques: paràmetres formatius, al·lòfons
- Característiques sintàctiques
- Característiques pragmàtiques
- Enllaços a altres lemes:
 - variants diatòpiques, diacròniques, etc.
 - mateixa arrel en signes compostos o derivats; sinònims i antònims
- Enllaç a la base de dades lèxica

En els apartats rellevants (característiques semàntiques, sintàctiques, pragmàtiques, etc.) s'inclouran exemples complets que n'il·lustrin la propietat en un context.

Un cop s'hagi completat la primera versió de la base de dades lèxica es farà pública al portal de l'LSC de l'IEC. S'establirà un canal de comunicació en línia, per tal que els usuaris de la base de dades puguin aportar comentaris i fer suggeriments en relació amb la informació codificada. D'aquesta manera es pretén que la mateixa comunitat d'usuaris pugui contribuir a millorar i completar amb informació que altrament només es podria recollir a través del treball de camp. Pensem, per exemple, en variants antigues de signes que no hagin estat recollides, o en formes que s'usin en determinades parts del territori.

A banda del projecte més immediat de base de dades lèxica, aquest projecte també es proposa completar i actualitzar els continguts gramaticals de la *Gramàtica bàsica LSC* (Quer *et al.* 2005), disponible en línia al portal de l'LSC de l'IEC del 2005. Com que l'equip de recerca *LSC Lab* de la Universitat Pompeu Fabra desenvoluparà una gramàtica de referència de l'LSC dins el projecte SIGN-HUB del programa Horizon2020 finançat per la Comissió Europea, se n'aprofitaran els resultats per a completar i actualitzar la descripció gramatical de la llengua al portal quan estiguin a l'abast. Mentrestant, però, s'incorporaran de manera més immediata els materials de pràctica de la *Gramàtica bàsica LSC* que van ser desenvolupats per FESOCA fa uns quants anys i no han estat mai a disposició del públic. D'aquesta manera s'ampliarà el volum de materials d'aprenentatge, i específicament en el domini on trobem les mancances més grans: la gramàtica.

4. CONCLUSIÓ

És clar que l'LSC s'enfronta a una etapa intensiva de descripció i codificació, que només s'explica per la falta d'atenció rebuda durant dècades per part dels estudiosos de la llengua, de les administracions i de la societat en general. De totes maneres, amb aquestes iniciatives de l'IEC i la recerca que es du a terme al grup d'investigació vinculats a les universitats, sembla factible encarar-la amb optimisme. Creiem que, un cop completats aquests dos projectes de codificació, estarem en unes condicions òptimes per a abordar el procés d'establiment d'una norma i d'una variant estàndard. Mentrestant, l'IEC vetllarà perquè els continguts i materials necessaris per a seguir aprenent, ensenyant i investigant sobre aquesta llengua siguin de la màxima qualitat i estiguin a l'abast de tota la població, començant pels seus mateixos usuaris.

REFERÈNCIES

- BARBERÀ, G. / RIBERA, E. (2010): «Lexicografia de les llengües de signes i l'estat de la lexicografia en la LSC». Dins MARTÍ, J. / MESTRES, J. M. (ed.): *Les llengües de signes com a llengües minoritàries: perspectives lingüístiques, socials i polítiques. Actes del seminari del CUIMPB-CEL 2008*, p. 207-237. Barcelona: Institut d'Estudis Catalans.
- BARBERÀ, G. / QUER, J. / FRIGOLA, S. (2016): «Primers passos cap a la documentació de discurs signat. El projecte pilot de constitució del corpus de la llengua de signes catalana». *Treballs de Sociolingüística Catalana* 25, p.287-302.
- DOMAD (2002a): *Vocabulari específic de la llengua de signes catalana (LSC). Àrees de coneixement del medi natural i ciències de la naturalesa*. Barcelona: FESOCA.
- DOMAD (2002b): *Vocabulari específic de la llengua de signes catalana (LSC). Àrees de coneixement del medi social i cultural i ciències socials*. Barcelona: FESOCA.
- ILLESCAT (2004): *Dilscat, el diccionari bàsic de la llengua de signes catalana*. Barcelona: Illescat.
- MARTÍN, I. / ALVARADO, M. (2004): *Diccionari Temàtic de Llenguatge de Signes Català*. Lleida.
- NAVARRETE, A. (2015): *Disseny d'una base de dades en línia associada al corpus de l'LSC a través de l'eina Signbank*. Treball de final de grau, Universitat Pompeu Fabra.
- PERELLÓ, J. / FRIGOLA, J. (1998): *Lenguaje de signos manuales*. Madrid: CIE Inversiones Editoriales Dossat.
- QUER, J. (direcció científica) / RONDONI, E. M. (direcció tècnica) / GRIN (2005): *Gramàtica bàsica de la llengua de signes catalana*. Barcelona: DOMAD.
- RIBERA LLONC, E. (2007): *Projecte de base de dades lexicogràfica per a la LSC (llengua de signes catalana)*. Treball de recerca dins el programa de doctorat Ciència cognitiva i llenguatge.