

Lexicografia de les llengües de signes i l'estat de la lexicografia en la LSC

Gemma Barberà

Universitat Pompeu Fabra

Eulàlia Ribera

Universitat Autònoma de Barcelona

1. Introducció

Actualment el reconeixement de les llengües de signes (LS) com a llengües naturals és un fet inqüestionable, però fa relativament poc temps que això és així. De fet, no va ser fins a la meitat del segle XX, que alguns lingüistes van publicar les primeres obres, en què defensaven el reconeixement de les LS basant-se en principis lingüístics (Tervoort, 1953; Stokoe, 1960). Els fonaments d'aquestes primeres obres van suposar l'inici d'un canvi de perspectiva que posteriorment va topar amb moltíssims obstacles per ser acceptada i generalitzada entre els especialistes. La forta tendència oral, no solament la dels estats que s'han resistit a implementar polítiques socials i educatives basades en aquest reconeixement, sinó també la de bona majoria dels professionals o de les mateixes famílies —en la major part dels casos, pares oients amb fills sords—, va fer que la recerca lingüística de les LS quedés estancada i molt endarrerida respecte dels estudis que es portaven a terme en les llengües orals (LO).

Aquesta situació va afectar també els estudis dins del camp lexicogràfic. La tendència oralista va fer desestimar el primer treball en lexicologia del signe de l'ASL (llengua de signes americana). Stokoe, Casterline i Groneberg van publicar el 1965 el primer diccionari en LS basat en principis lingüístics. Les entrades es presentaven per mitjà d'un sistema de notació desenvolupat per a la descripció dels signes, que els descrivien en relació amb la localització, la configuració i el moviment de les mans, és a dir, mitjançant els paràmetres fonològics de la modalitat signada. A conseqüència del context social, cultural i polític del moment, així com pel seu format, que el feia poc accessible a les persones no lingüistes pel nombre limitat d'il·lustracions i per les anotacions de les entrades, aquesta via d'exploració no va tenir èxit i no es va aprofitar per

a la confecció d'altres diccionaris. La major part de diccionaris en LS que es van publicar durant aquest període i fins pràcticament als anys vuitanta consistien en llistes de signes utilitzats com a equivalents de les paraules d'una LO, ordenats d'acord amb els principis d'aquella llengua; és a dir, per ordre alfabètic. A més, en aquestes publicacions normalment no hi havia cap referència sobre el significat dels signes, perquè es donava per entès que aquest era el mateix que el de la traducció de la LO que acompanyava cada signe. I és que el principal objectiu del diccionari consistia a donar als pares oients i als professors un llibre de referència que els facilités la comunicació amb els nens sords i, per tant, es confeccionaven llistes de paraules i s'identificava un signe amb aquelles paraules.

Aquest sistema no descrivia el component lèxic característic i propi de les LS, és a dir, deixava de banda la matriu semàntica, morfològica, sintàctica i pragmàtica pròpia de cada signe, amb la qual cosa ometia gran part de la informació que el parlant té a la ment com a part del seu coneixement d'un signe.

Per exemple, en el cas de l'elaboració del diccionari de la BSL (llengua de signes britànica), Brien explica en la introducció que a l'inici de la confecció, abans que el primer responsable de l'edició, Allan Hayhurst —desaparegut l'any 1981—, topés amb el treball de Stokoe cap al 1975 i introduís modificacions importants en la perspectiva i l'elaboració del diccionari, quan es considerava que no hi havia cap correspondència entre la LO i el signe, o bé s'esborrava el signe o bé se'n creava algun prenent com a referent el Gestuno.¹ D'altra banda, s'utilitzaven fotografies per a il·lustrar els signes i s'hi afegien dibuixos per a indicar els moviments involucrats. Cada fotografia anava acompanyada d'una descripció escrita de com es produïa, juntament amb les paraules angleses més comunes associades amb el signe.

En definitiva, no va ser fins als anys vuitanta que van veure's publicacions de diccionaris de LS organitzats d'acord amb els principis similars als defensats per Stokoe i els seus col·legues feia vint anys. Les noves tecnologies hi han ajudat molt, i això ha fet que actualment en LS puguem trobar diccionaris impresos i recursos lexicogràfics multimèdia, encara que la recerca i la varietat de materials lexicogràfics que s'han fet en aquesta modalitat queda molt lluny de la que podem trobar en la modalitat oral.

A continuació veurem amb quines dificultats es troben els lexicògrafs que volen elaborar qualsevol material lexicogràfic d'una llengua. En el primer apartat exposarem les dificultats comunes a totes les llengües naturals, veurem les especificitats de

1. Recull de signes internacionals publicat per la Federació Mundial de Sords (1975) per al contacte entre signants de diferents llengües de signes (Brien, 1992).

la modalitat visual i, posteriorment, parlarem de les característiques de les comunitats sordes. En el segon apartat, parlarem dels materials lexicogràfics que s'han editat en LS i, específicament, en LSC. Finalment, en el tercer apartat, exposarem una proposta de base de dades lexicogràfica (BD d'ara endavant), amb una mostra de la informació que aquesta base de dades podria recollir per confeccionar materials lingüístics tan importants com un diccionari normatiu descriptiu de la LSC.

1.1. *Especificitats de les llengües de signes*

Dur a terme la descripció del component lèxic d'una llengua no és en absolut, fàcil. D'entrada, el lexicògraf ha de tenir en compte tots els aspectes següents:

— *La representació abstracta que figura com a lema.* Una unitat lèxica d'una llengua pot adoptar diverses formes en funció del context sintàctic, la categoria morfològica, la semàntica o altres aspectes. Aquestes variacions no modifiquen el significat del signe, però cal despullar el signe d'aquests aspectes flexius per tal de trobar una forma que les representi totes. Aquesta representació abstracta d'una unitat lèxica s'anomena *lema*.

— *Les entrades i les subentrades.* Aquest apartat implica el tractament de l'homonímia i de la polisèmia. Quan dos signes tenen la mateixa forma però un significat totalment independent i diferent l'un de l'altre diem que són mots homònims. Si, en canvi, un signe té diferents significats relacionats entre ells diem que es tracta d'un signe polisèmic. Cal decidir quins signes es tracten com a homònims i quins com a polisèmics.

— *La microestructura.* Les unitats lèxiques d'una llengua porten moltíssima informació associada. Aquesta informació és la que fa que la puguem inserir en una combinació de paraules, o bé que no ho puguem fer, ja sigui perquè no pertany a la categoria adequada, perquè no forma part del registre adequat, etc. Depenent dels objectius i dels destinataris del material lexicogràfic que s'hagi d'elaborar, el lexicògraf ha d'escollir la informació que hi recull.

A més d'aquests aspectes comuns que cal tenir en compte en qualsevol treball lexicogràfic, sigui de la modalitat que sigui, també hem de tenir presents les diferències que trobem entre les LS i les LO, que són les següents:

— *La forma del senyal que emeten.* Les LO emeten un senyal acústic mentre que les LS emeten un senyal òptic. Això vol dir que els canals d'emissió i recepció de la llengua són diferents en una i altra modalitat.

— *Els paràmetres fonològics i la seva interpretació fonètica.* Com que utilitzen òrgans articuladoris diferents, la interpretació fonològica i els paràmetres fonològics són dife-

rents. En el cas de la modalitat oral, la interpretació fonètica es basa i es materialitza en els òrgans articuladoris vocals, mentre que en el cas de la modalitat visual, s'hi utilitzen les mans, la cara, el cap i el tors.

— *La simultaneïtat enfront de la seqüencialitat.* Encara que totes dues modalitats tenen components fonològics que es realitzen simultàniament i components que es realitzen seqüencialment, la modalitat visual té un percentatge de simultaneïtat més elevat que el que trobem en les LO.

— *L'ús de l'espai.* La modalitat visual permet l'expressió en tres dimensions, i això comporta un ús més complex i acurat de l'espai. A més, la ubicació variable dels signes respecte al cos del signant pot comportar canvis de significat. Per exemple, el signe *operar* pot articular-se a diferents parts del cos segons quin sigui l'òrgan que ha patit l'operació.

— *La iconicitat.* El grau d'iconicitat és més elevat en les LS que en les LO. De fet, la trobem també a un nivell sublèxic, ja que els paràmetres formatius poden associar-se a un significat determinat. Un exemple molt clar és l'associació entre la localització prop del front i certs elements lèxics relacionats amb processos mentals. Així, signes com *pensar*, *imaginar*, *idea* o *reflexionar* tenen el lloc d'articulació prop del front o bé hi inicien el moviment.

Quan l'objectiu és elaborar un material lexicogràfic en la modalitat visual, cal tenir en compte aquestes diferències i buscar una manera de recollir i presentar les dades adequada a la modalitat. Algunes de les dificultats amb què s'han trobat els lexicògrafs que han començat a treballar amb les LS són les següents:

— *La segmentació de les unitats lèxiques.* Quan un enunciat sencer es realitza mitjançant un sol signe, costa de saber quantes unitats lèxiques té aquell signe.

— *L'absència d'un codi escrit unificat.* En les llengües de signes no hi ha un codi escrit unificat. Aquesta és una dificultat afegida, ja que l'escriptura és la font i el mitjà principal d'elaboració d'aquest tipus de materials en LO. El material escrit ajuda a segmentar les paraules i a tractar informàticament les dades, de manera que, tot i que de vegades una unitat lèxica pot consistir en un grup de paraules—la qual cosa dona no poques dificultats als lexicògrafs— la linealitat i l'organització de l'escriptura ofereix una facilitat que en les LS no trobem.

— *La representació de les unitats lèxiques.* Habitualment es fa mitjançant la glossa, que és la reducció del significat del signe a una paraula de la LO més propera. Cal buscar noves vies de representació més adequades a la modalitat visual, com la imatge. Un dibuix o una fotografia pot representar el signe més fidelment que una glossa, per exemple.

— *La descripció de les unitats lèxiques.* En la modalitat oral les definicions s'ofereixen en format text. Les LS no tenen codi escrit, per la qual cosa no es pot oferir la informació en format text si no és recurrent a la LO més propera. Serà necessari utilitzar l'escriptura per fer possible el processament de les dades per part de la BD, però és evident que cal buscar noves vies de presentació a l'usuari adequades a la modalitat visual. Una possible solució és fer les descripcions en format vídeo.

— *La cerca dels ítems lèxics.* Actualment el sistema de cerca més estès és l'alfabètic. Aquest és el sistema més utilitzat també en les LO indoeuropees. Però en el cas de la modalitat visual, torna a fer-la dependent de la glossa i, per tant, de la LO més propera. Cal, doncs, buscar un sistema alternatiu més adequat a la modalitat visual, com pot ser el sistema que presentarem en aquest article, basat en els paràmetres fonològics dels signes.

La joventut dels treballs lexicogràfics en LS fa que molts d'aquests aspectes encara s'estiguin explorant. S'arrossega el voluminós bagatge de les LO, amb el seu sistema d'organització i de presentació, i això fa que resulti molt difícil trobar sistemes més adequats per al recull lexicogràfic de la modalitat visual. La introducció de les noves tecnologies ha significat un salt molt important en aquests tipus de treball perquè, tal com veurem en els apartats següents, permeten obtenir uns criteris visuals per a recollir i organitzar les dades.

1.2. Particularitats de la comunitat sorda

Entenem per *comunitat sorda* aquella «comunitat de parla signant que engloba un conjunt d'individus que comparteixen una llengua i uns valors sociolingüístics comuns, com regles d'ús i actituds lingüístiques, així com també una ideologia i una actitud vers aquests usos lingüístics» (Gras, 2006, p. 5). Tot i així, l'edat de la primera exposició de la llengua, la llengua materna de casa, la varietat social i regional i l'escolarització són elements que contribueixen a donar una gran varietat de color i heterogeneïtat a aquesta comunitat.

Un aspecte comú a les comunitats sordes és que la gran majoria dels seus membres van entrar en contacte amb la llengua durant l'escolarització, ja que gran part dels infants sords procedeixen de famílies oients (es calcula que entre el 90-95 %). Segons els estudis estadístics de Gras (2006), per al conjunt de l'Estat espanyol el 73 % dels informants que van servir per a elaborar les estadístiques del seu treball van entrar en contacte amb la LS en l'etapa infantil, mentre que només el 9,6 % ho van

fer a casa i, per tant, eren nadius de la LS. A més, cal tenir en compte també que estem parlant d'una llengua que va estar prohibida durant molt de temps i relegada, per tant, a l'ús familiar. Les distàncies territorials i les polítiques educatives aplicades en diferents èpoques també han fet que la llengua evolucionés de manera diferent a cada regió, i això fa que ens trobem davant d'una varietat regional molt poc homogènia.

Tots aquests aspectes cal tenir-los en compte a l'hora d'engegar qualsevol estudi lingüístic de la llengua. Coneixedors d'aquest fet i conscients que la llengua és un mosaic representatiu de tota una regió, els equips de recerca vinculats a les LS han intentat dur a terme diferents mesures. Un cas molt clar és l'efectiva solució a què han arribat des de l'Institut de la llengua de signes alemanya (DGS) i la comunicació dels sords de la Universitat d'Hamburg.² Aquest equip treballa des de fa quinze anys en un projecte de creació de corpus que ha permès el desenvolupament d'un diccionari electrònic basat en aquest corpus, una col·lecció de vocabulari bàsic i un diccionari de gramàtica. Conscients de la realitat heterogènia de la llengua, han creat un grup d'experts format per unes vint persones sordes que treballen d'assessors de la llengua. En casos de dubtes sobre l'estatus lèxic d'un signe, qüestions de canvi en la llengua, diferents significats o informació regional dels signes, recorren a aquest grup d'experts, que ha estat format prèviament en qüestions lingüístiques i són representants directes de la comunitat sorda d'arreu d'Alemanya. Així, qüestions relacionades amb la selecció de lemes i amb la bona formació de construccions gramaticals són constantment validades per part d'aquest grup (Prillwitz *et al.*, 2008). A vegades també actuen de contacte per aconseguir que altres persones sordes verifiquin dades en cas que sigui necessari. Tot i així, com que aquest grup d'experts és reduït, també han establert un sistema de votació públic que permet que la comunitat lingüística alemanya en general participi directament en la construcció del projecte. Així, els que ho desitgin poden votar i aportar variants regionals que en un primer moment no apareixen al corpus de dades.

Qualsevol iniciativa de creació de material, d'estudis lexicogràfics o de projectes de corpus de les LS ha de treballar des d'aquesta òptica, tot fent que la comunitat lingüística participi activament en el procés de realització dels projectes de la llengua i obtingui, així, una mostra de la llengua al més representativa possible.

2. Vegeu <http://www.sign-lang.uni-hamburg.de/>.

2. Material lexicogràfic en llengua de signes

Una base de dades (BD) lexicogràfica és un recull de lemes d'una llengua natural. Les BD lexicogràfiques creades amb l'ajuda de la informàtica amplien enormement les possibilitats d'emmagatzemar informació sobre el lèxic, i l'agilitat en el moment de fer recerca permet que puguin utilitzar-se no només per als usos habituals dels diccionaris, sinó també per a efectuar cerques lingüístiques sobre aspectes concrets de la llengua estudiada, per exemple sobre variació sociolingüística, fonologia o altres aspectes sintàctics o semàntics. A més, l'avenç en les noves tecnologies ha beneficiat en gran part els materials en LS, ja que ha permès visualitzar la llengua i integrar l'aspecte visual en la presentació de les entrades.

Dins dels recursos multimèdia és important fer la distinció entre el que entenem per *diccionari* i per *base de dades lexicogràfica*, perquè a vegades la frontera entre ambdós costa de distingir. Els diccionaris són el producte final que arriba a l'usuari per buscar-hi informació sobre una llengua natural, mentre que les BD són aquells reculls de tot el material lexicogràfic que duu a terme un lexicògraf per permetre que quan s'elabori un diccionari ofereixi les millors garanties de recerca per a l'usuari. Això implica la inclusió d'un nombre de camps molt més elevat del que després es veuran en el producte final, per exemple la vinculació amb el corpus i els informants, la transcripció en sistemes de notació coneguts només en els cercles investigadors o bé la inclusió de camps amb etiquetes semàntiques que permetin la recerca per aquesta via. De fet, els recursos multimèdia permeten emmagatzemar molta més informació que els diccionaris impresos, i això fa que el domini lexicogràfic sigui molt més ampli.

La major part dels diccionaris publicats des del 1965 no han estat sobre els signes utilitzats en les LS, sinó sobre signes utilitzats com a suplementes de les paraules d'una LO, ordenats, per tant, d'acord amb els principis de la LO. Normalment en aquestes publicacions no hi ha referència al significat dels signes, sinó que es dona per entès que aquest és el mateix que el de la traducció de la LO que l'acompanya. La motivació d'aquesta dependència cap a la LO, com ja hem comentat, prové del fet que en aquell moment el principal objectiu del diccionari era donar als pares oients i als professors un llibre de referència que els facilités la comunicació amb els nens sords.

Tot i les divergències en l'organització de la informació dels diferents diccionaris, els diccionaris impresos tenen en comú que la informació que es dona sobre el signe és bàsicament sobre la manera com s'executa i hi apareix també la glossa o paraula equivalent en la LO corresponent.

2.1. *Diccionaris impresos de llengües de signes*

La varietat de diccionaris en LS no és tan espectacular com en el cas de les LO, però també hi trobem diferències en l'organització dels signes. En aquest apartat presentem breument dos diccionaris impresos de LS que són rellevants per la presentació de les entrades lèxiques.

a) *Dictionary of British Sign Language - English* (1992), de David Brien i Mary Brennan

En aquesta obra les entrades s'han organitzat des d'un punt de vista semasiològic i s'ha fet servir un sistema visual de classificació utilitzant les parts del signe transcrites mitjançant un sistema de transcripció propi àmpliament conegut pels investigadors del grup. Segons explica Brien a la introducció del diccionari, els signes que s'hi recullen s'agrupen d'acord amb paràmetres fonològics. Això vol dir que hi trobem diferents grups de símbols classificats d'acord amb els aspectes següents:

- la configuració de la mà
- la localització del signe
- l'orientació de la mà i dels dits
- el moviment de les mans

Cadascun d'aquests grans grups es divideix en subgrups que ajuden en la cerca dels signes, encara que per accedir-hi fàcilment cal conèixer el sistema de transcripció descrit en les pàgines precedents del recull i decidir primer si el signe que es busca és de configuració simple o complexa (és a dir, si es fa amb una sola mà o amb totes dues), i a partir d'aquí anar escollint els subgrups de signes segons les seves característiques fonològiques. Les entrades consten de la descripció de l'execució del signe i de la traducció en anglès.

b) *La langue des signes, dictionnaire bilingue LSF / français* (1997), de M. Girod *et al.*

Aquesta obra organitza les entrades onomasiològicament, és a dir, per temes i subtemes que es troben llistats en un índex. Aquest és el criteri utilitzat a la major part de diccionaris d'altres LS. D'acord amb les explicacions de l'editor en la introducció, s'ha fet així perquè el diccionari anava dirigit a estudiants, i consideraven important que aquests poguessin veure diferències fonològiques subtils entre signes d'un mateix camp semàntic. Tampoc no s'hi defineix el significat del signe, sinó que mitjançant dibuixos se'n descriu l'execució i es posa l'equivalent glossat en francès.

D'altra banda, també conté un índex de classificació dels signes segons la configuració, però per trobar-hi un signe en primer lloc cal veure quina és la configuració principal del signe mitjançant la traducció; en segon lloc, cal fixar-se en la localització (si es troba al cap, al cos o davant del tors), i, finalment, cal mirar la llista alfabètica de les traduccions, on hi ha la numeració de la pàgina i el número de vinyeta dins la pàgina.

2.2. *Diccionaris impresos en llengua de signes catalana*

a) *Lenguaje de signos manuales* (1987), de Jorge Perelló i Juan Frigola

En la segona edició, del 1998, aquesta obra pionera consta de 4.500 signes, i se n'hi inclouen de les diferents comarques catalanes. Està organitzat temàticament i cada entrada consta del dibuix del signe, més la glossa en castellà i en català i un codi d'execució del signe, la interpretació del qual es pot trobar en els primers capítols. Una altra característica d'aquest diccionari és que té capítols dedicats a determinats aspectes gramaticals, com l'article, els substantius, els verbs, els adverbis, els adjectius, els pronoms, les preposicions, les conjuncions, les interjeccions i la construcció de frases. Al pròleg s'especifica que el seu objectiu no és ser un llibre de gramàtica, sinó un llibre de reforç per als aprenents de la LS, per la qual cosa dona quatre directrius pràctiques i molt generals d'alguns aspectes de la llengua.

b) *Diccionari temàtic bàsic de llengua de signes catalana de ponent* (1996), d'Ignasi Martín i de Marisa Alvarado

Aquest diccionari, amb 3.900 entrades en la seva segona edició, està organitzat temàticament, de manera semblant al de la LSF que hem comentat, però sense índex per configuració o altres paràmetres fonològics del signe. Les entrades es presenten a partir d'uns dibuixos senzills i clars, on el moviment apareix en una segona icona del dibuix. L'objectiu d'aquest és proporcionar recursos i documentació ordenada per a complementar l'aprenentatge de la LS, per la qual cosa conté un índex alfabètic de termes en llengua catalana i un altre en llengua castellana. La segona edició d'aquesta obra es va fer el 2004 i duia el nom de *Diccionari temàtic de llenguatge de signes català* (DTLSC).

2.3. *Recursos multimèdia en llengües de signes*

Una de les millores més importants dels recursos multimèdia aplicats a la LS és l'increïble augment de les capacitats d'emmagatzematge dels ordinadors, així com el desenvolupament de la tecnologia del vídeo digital, fet que ha permès relacionar un vídeo amb cada entrada. Això també ha comportat, però, que els sistemes de transcripció de les LS no siguin tan rellevants per a la identificació del signe, ja que el vídeo permet visualitzar-lo amb 3D. En aquest sentit, no volem dir que puguem prescindir-ne, ja que són necessaris, per exemple, per als sistemes de traducció, sinó que l'ús del vídeo ha provocat que els sistemes de transcripció i notació siguin menys rellevants en la presentació de les entrades (Johnston, 2001).

Actualment, el fet que es puguin utilitzar canals multimèdia fa que tant els diccionaris com les BD lexicogràfiques s'adaptin perfectament a la modalitat visual de les LS i que els criteris lingüístics aplicables puguin ser molt més propers a aquesta modalitat. En els darrers anys han sorgit força grups de recerca dedicats al treball lexicogràfic en diverses LS, que treballen amb BD. Alguns han presentat productes finals en format CD-ROM i DVD, mentre que altres treballen a la xarxa telemàtica, de manera que es beneficien de l'actualització immediata que permet aquesta via.

2.3.1. Bases de dades de llengües de signes

a) BD de l'Auslan (llengua de signes australiana)

La base de dades de la llengua de signes australiana és un exemple de feina lexicogràfica desenvolupada des de fa temps per a una LS concreta. Inclou informació fonològica, gramatical, semàntica, així com les glosses i les definicions dels signes. La informació visual es presenta amb gràfics, vídeos i transcripcions (Johnston, 2001). El fruit final d'aquesta BD és el material en línia que podem consultar a <http://www.auslan.org.au>. Aquí hi podem fer cerques a partir de la traducció en LO (en aquest cas, en anglès) i hi apareix el signe, amb les variants introduïdes, la definició, la categoria gramatical i la zona d'ús del signe. A més, la pàgina inclou uns apartats on els usuaris poden proposar signes que no hi apareguin i donar *feedback* sobre els signes ja introduïts.

b) BD de l'ÖGS (llengua de signes austríaca)

Recentment ha estat fet pública al web la BD de l'ÖGS. Desenvolupada a la Universitat de Klagenfurt, es pot consultar en línia a <http://ledasila.uni-klu.ac.at>. Aquesta BD té dues particularitats. D'una banda, és una plataforma oberta on és possible introduir-hi noves categories o valors, sempre que es tingui l'estatus d'administrador. S'ha dissenyat de manera que no hi hagi decisions fixes des de l'inici del projecte que impedeixin la introducció de noves categories a mesura que la recerca avança (Krammer *et al.*, 2001). A més, no és específica per a una LS en particular i pot servir com a plataforma per documentar altres LS, a través de l'aplicació web Leda Sila (Lexical Database for Sign Languages). Actualment, la pàgina web es pot consultar en anglès i alemany, però l'equip està treballant per incloure la finestra d'ajuda en ÖGS i en sistema de signes internacional. Les cerques dins la BD es poden fer a partir de les opcions següents: semàntica, regió, tipus de signe, lloc d'articulació, configuració i àmbit. Fins avui, Leda Sila només s'ha utilitzat per a la descripció de l'ÖGS i conté catorze mil vídeos de signes, incloent-hi les variants regionals (Krammer *et al.*, 2008).

c) BD de la NGT (llengua de signes holandesa)

La BD de la NGT és un dels projectes que es desenvolupen dins el Centre per a la Llengua de Signes Holandesa (NGc), fundat l'any 1992. Des del 2006, aquest centre s'ha convertit en l'institut nacional de lexicografia de la NGT, i s'hi han desenvolupat projectes sobre materials d'aprenentatge de la llengua, cursos educatius, una base de dades i diccionaris en DVD i en línia. La BD en línia <http://www.gebarencentrum.nl/gebarenwoordenboek.asp> conté actualment uns 14.000 signes. Aquesta BD també ha estat l'eina d'ús en la recollida de signes per al projecte de planificació de la llengua.

d) BD de la DSGS (llengua de signes suïssogermana)

Aquesta BD multimèdia ha servit com a base per a posteriors diccionaris i materials d'ensenyament, i també és una eina per a la investigació lingüística. Utilitza vídeos de demostració, fotogrames del vídeo, dibuixos, anotació fonètica amb HamNoSys i una petita part amb signescriptura per a la representació gràfica dels signes. A més, inclou informació de variants generacionals, així com l'etimologia dels signes que es coneix (Boyes Braem, 2001).

e) SignPhon

En aquest últim subapartat esmentem una base de dades amb objectius específics. Es tracta del SignPhon, una BD fonètica i fonològica dissenyada per a qualsevol LS. Com qualsevol altra BD, el SignPhon va ser dissenyat per a trobar patrons, regularitats distribucionals i freqüències d'un domini concret, amb la finalitat d'utilitzar aquesta informació per a construir una teoria. En aquest cas, l'objectiu era clarament realitzar estudis fonètics i fonològics de les LS i emprar les dades del SignPhon per a comprovar certes hipòtesis, o bé per desenvolupar-ne d'altres. En tots els inicis de les construccions de BD no es té un model comprensiu del domini. En començar a codificar les primeres dades cal ja saber per a quin aspecte serà útil la BD, és a dir, ja s'ha d'haver determinat la categorització rellevant de les dades. Per tant, el sistema de codificació ideal és, de fet, la millor teoria del domini. Ens trobem davant del que Crasborn *et al.* (2001) anomenen «la paradoxa de la base de dades» i és que sembla que qualsevol sistema de codificació serà una aproximació del primer ideal que es tingui.

2.3.2. Aplicacions de les bases de dades

L'aplicació més usual de tota base de dades és la creació d'un diccionari de la llengua. Aquest diccionari, però, com ja hem comentat, només pot incloure una part de tot el coneixement de la llengua contingut a la BD.

a) Diccionari de la DSL (llengua de signes danesa)

Aquesta obra conté aproximadament dos mil signes de la DSL, amb exemples i referències a sinònims i homònims. Actualment és accessible a <http://www.tegnsprog.dk> i hi podem fer cerques de la DSL a partir de les característiques manuals dels signes o bé a partir de la paraula en danès oral. En aquesta obra, els criteris que s'han seguit per a la selecció de les entrades són fonològics i semàntics. Així, dos signes que difereixen en significat esdevenen dues entrades diferents, i passa el mateix amb signes que tenen algun paràmetre que els diferenciï. El problema apareix en establir la frontera entre entrades que difereixen per trets fonològics i variants del mateix signe. En aquest material es tracten com a variants els signes que es diferencien per no més d'un paràmetre. Els signes que varien en dos paràmetres són tractats com a sinònims i tenen entrades diferents (Troelsgard i Kristoffersen, 2008).

b) Diccionari de la VGT (llengua de signes flamenca)

Aquest material el podem trobar en línia a <http://gebaren.ugent.be/>. Des d'aquest enllaç podem fer les cerques a partir de la paraula en flamenc i, al mateix temps que visualitzem el vídeo amb el signe, podem veure l'equivalent amb signescriptura. Tot i així, des de l'enllaç <http://gebaren.ugent.be/visueelzoeken.php> podem cercar signes per configuració i punt de contacte.

c) El projecte bilingüe SignBase

Aquest projecte va tenir unes aplicacions molt concretes. SignBase és una obra que presenta la informació en LO i LS, i conté informació morfèmica que inclou la unitat més petita de significat de cada signe (és a dir, té en compte el lèxic productiu de les LS, com ara els classificadors). A més, les entrades es presenten amb fotogrames, vídeos i s'hi inclou també la definició del terme (Schermer *et al.*, 2001). Aquest projecte va fer sortir a la llum dues aplicacions: el *Diccionari BSL - anglès* de la cultura i la comunitat sorda, i el CD-ROM de la naturalesa i el medi en NGT - holandès (material d'educació bilingüe). D'una banda, en el *Diccionari BSL* es poden fer cerques per la paraula oral o pel signe (a partir de la configuració), també mitjançant la informació gramatical, és a dir, per categories lèxiques. De l'altra, el CD-ROM de la naturalesa i el medi en NGT és, en canvi, una aplicació més específica que presenta conceptes i explicacions dels signes. No inclou una descripció gramatical explícita, però sí que la presenta indirectament amb els exemples, ja que el signe *beure*, per exemple, varia de configuració segons si el subjecte és un animal o un altre.

d) Diccionari de la FinSL (llengua de signes finesa)

Aquest diccionari, com ha ocorregut en altres casos, va ser el resultat del treball lexicogràfic realitzat en el desenvolupament de la BD de la llengua de signes finesa (Savolainen, 2001).

2.3.3. Aplicacions lexicogràfiques i projectes de corpus

En els darrers anys la creença que els treballs lexicogràfics s'han de basar en projectes de corpus s'ha estès àmpliament. A més, la selecció de les entrades d'un corpus requereix la construcció paral·lela d'un lèxic. Per aquesta raó, alguns dels treballs lexi-

cogràfics que es fan en diverses LS van units a la recollida de dades i posterior anotació, ja que ambdues feines s'enriqueixen amb l'altra. L'avantatge de disposar d'un corpus és que les decisions per a la identificació i la selecció dels signes es poden atribuir a les dades. El desavantatge és que un corpus és només un exemple limitat de l'ús de la llengua, que a vegades provoca que en quedin exclosos altres signes apropiats (Konrad *et al.*, en premsa).

a) Diccionaris de termes especialitzats

A la Universitat d'Hamburg s'han desenvolupat diccionaris de terminologia especialitzada amb temàtica diversa, que va des de la terminologia computacional, la psicologia, la fusteria i l'economia domèstica fins al treball social i la salut. Aquests reculls de termes tècnics es poden consultar lliurement al web <http://www.sign-lang.uni-hamburg.de/Projectseng/SLDicts.html>.

Una característica important de la LS és que els signes manuals interaccionen amb la vocalització per expressar el significat lèxic de l'element. Quan diem *vocalització* ens referim al «component parlat», que és el moviment de la boca que forma part del signe i que té el seu origen en el gest de pronunciació de la paraula en LO (Quer *et al.*, 2005). Aquest component parlat és de gran rellevància en els àmbits de vocabularis especialitzats, ja que per a les LS la situació és especial. Generalment, podem trobar pocs sords en els camps més especialitzats, i només uns contextos determinats i reduïts on els subjectes sords utilitzen els signes especialitzats. Sovint, però, la realitat és que aquests professionals sords utilitzen la LO escrita per a la comunicació i la documentació d'aquests termes. Per aquest motiu, en la terminologia especialitzada en signes trobem una gran influència de la LO que apareix reflectida en el component parlat del signe.

En aquest punt pren importància la diferència d'introducció de les entrades a partir de *tokens* i *types*. Els *types* es refereixen a la manera de citar els termes. Els *tokens* són la variació o la diferent realització del signe en un context determinat (König *et al.*, 2008). Pel que fa als termes especialitzats, la LS sovint utilitza el mateix signe manual amb un component parlat diferent. En LSC, per exemple, el signe manual *mentida* pot anar acompanyat amb l'oralització de la paraula *virtual*, de manera que es canvia així el significat del terme. Ens trobem amb dues instàncies del mateix signe, ja que la variant del significat *virtual* és un *token* del mateix *type*. Identificar aquestes dues formes com a dos lexemes diferenciats seria enganyós, perquè es tracta de la mateixa forma manual combinada amb la paraula de la LO.

b) L'ILex, l'eina que enllaça corpus i BD

L'ILex³ és una eina destinada a la feina lexicogràfica i de corpus, que permet treballar paral·lelament en la construcció d'un corpus i d'un lèxic relacionats. Amb el precedent del GlossLexer, una BD lèxica, i el SyncWRITER, els mateixos autors de la Universitat d'Hamburg han creat l'ILex (nom que prové de «lèxic integrat»). Es tracta d'una eina que modela els processos de transcripció dels corpus que garanteix l'enllaç entre *token* i *type* (Hanke i Storz, 2008). Les transcripcions consisteixen en *tokens*, és a dir, signes, que es refereixen als seus tipus respectius, fet que té una relació directa amb el procés de lematització. Això permet la creació de corpus amb molts anotadors al darrere i garanteix la coherència *intra* i *inter* anotadors. L'ILex no només s'utilitza per a feines lexicogràfiques i d'anàlisi del discurs, sinó també per a generar seqüències de transcripcions de la BD a través d'un avatar signant.

c) Altres projectes de corpus fets amb eines específiques

El MediaTagger és una eina d'anotació multimodal. Permet observar les ocurrencies d'un signe concret dins del context. S'ha emprat per a l'estudi de la llengua de signes de Nicaragua (*idioma de señas nicaraquíense*, ISN). L'objectiu era analitzar i comparar enunciats emesos per signants de les dues cohorts (grups de persones que van entrar a l'escola en diferents moments). L'eina els ha servit per a l'anotació i la posterior anàlisi comparativa de les descripcions que feien els dos grups (Senghas, 2001).

El SignStream⁴ forma part del projecte de recerca lingüística en ASL (llengua de signes americana) de la Universitat de Boston. Es tracta d'una aplicació que proporciona un entorn únic on es poden veure i analitzar les dades anotades, i també s'hi poden fer cerques, de manera que és molt útil per a la recerca lingüística en LS i per al component gestual de les LO. Permet fer cerques complexes mitjançant unes finestres que mostren els camps i els valors disponibles en l'arxiu seleccionat. Aquesta aplicació ha permès la creació d'un corpus anotat en ASL (Neidle, 2001).

El Berkeley Transcription System (BTS) és una eina que transcriu els elements al nivell morfèmic i que es va aplicar a un projecte sobre l'adquisició de les LS. El sistema es va desenvolupar utilitzant dades d'interaccions adults-infants en ASL i en NGT (Slobin *et al.*, 2001).

3. Vegeu <http://www.sign-lang.uni-hamburg.de/illex>.

4. Vegeu <http://www.bu.edu/asl/rp/SignStream/index.html>.

En darrer lloc, l'ELAN⁵ és una eina d'anotació de codi obert que funciona amb diferents plataformes i que ha estat emprada en el corpus ECHO,⁶ d'accés públic i multilingüe, que té la voluntat d'establir estàndards d'anotació i de facilitar l'intercanvi de dades.

Finalment, un darrer punt que ens agradaria destacar és el protagonisme que pren la part gestual de la llengua quan treballem amb corpus anotat. Sovint en un discurs signat trobem ítems pròpiament lingüístics i d'altres que els podem catalogar de gest mim. Aquest gest mim és molt freqüent en dades espontànies. Quan treballem amb una BD que està relacionada amb un corpus, ens trobem amb la dificultat d'haver de diferenciar d'alguna manera tot allò que és lingüístic de tot allò que no ho és. Evidentment, el lexicògraf treballa amb els elements lingüístics, però a vegades és difícil obviar una part de la llengua difícil de catalogar, però que per reduïda que sigui, existeix. Johnston (2008) proposa transcriure aquests signes a partir de la traducció que en faríem. És a dir, escriure «g: *que-babau-que-sóc*», en lloc de «g: *picar-el-palmell-al-front*». La forma del gest no cal que quedi plasmada en la tira de les glosses del signe, ja que aquesta es pot apreciar o bé en el vídeo relacionat o bé en la tira dedicada als paràmetres fonològics. Evidentment, tot i que aquesta part de la llengua no estaria inclosa a la BD, sí que se'n tindria constància en el corpus.

2.3.4. Recursos multimèdia de la LSC

a) *Dilscat: Diccionari bàsic de la llengua de signes catalana*

Un exemple clar de les possibilitats tècniques de les noves tecnologies aplicades a la LSC que tenim de moment és el *Dilscat: Diccionari bàsic de la llengua de signes catalana*, editat per ILLESCAT el 2004. Aquest diccionari, que conté mil cinc-cents vídeos, es presenta en format CD-ROM i això permet apreciar les característiques intrínseques pròpies de la modalitat de la llengua. Aquest diccionari permet dos tipus de cerques: la temàtica i l'alfabètica (segons la glossa), i està presentat en català, castellà, anglès i portuguès.

5. Vegeu <http://www.lat-mpi.eu/tools/elan>.

6. Vegeu <http://www.let.kun.nl/sign-lang/echo>.

b) *Mira què dic*

*Mira què dic*⁷ és un diccionari bàsic en LSC i disponible a través d'Internet realitzat per Montse Quijo, Montserrat Puigdurí i Pilar Viana. Està elaborat pel centre CEE Les Aigües, de Mataró, amb l'objectiu de dotar d'una eina lingüística els professors i pares d'alumnes amb discapacitat i millorar la comunicació de persones amb dificultats. Conté un nombre molt limitat de paraules (unes cinc-centes), les justes per a poder-se comunicar amb aquests alumnes, i proposa dues maneres d'efectuar la cerca: mitjançant un índex alfabètic de la glossa i mitjançant un índex temàtic. Aquest material conté vídeos que mostren com es fa el signe, i també dibuixos que identifiquen els signes, extrets del *Diccionari temàtic en llengua de signes catalana*.

c) Terminologia en LSC de l'àmbit de llengua

Des del Departament d'Educació de la Generalitat de Catalunya, un grup de docents vinculats a programes d'ensenyament bilingüe va crear el 2006 un glossari⁸ consultable en línia de terminologia específica de signes relacionats amb l'àrea de llengua, amb la finalitat d'estandarditzar els signes per potenciar l'ensenyament i l'aprenentatge de la llengua catalana i castellana a través de la LSC. Aquest glossari, coordinat per Marta Vinardell i Marga Cirera, es va crear a partir d'uns tallers amb diferents perfils de professionals que treballaven a partir d'un lèxic ja analitzat de termes relacionats amb la llengua i el llenguatge, la comunicació, la lingüística, les parts de l'oració i del text i les tipologies textuals. El resultat va ser un CD-ROM que es va repartir entre el professorat que treballa amb alumnat sord des de la vessant educativa bilingüe (LSC i llengua catalana/castellana).

3. Proposta de base de dades per a la LSC

La proposta de BD que presentem a continuació va ser presentada com a treball de recerca dins del curs de doctorat interuniversitari de ciència cognitiva i llenguatge, i va ser dirigit i tutoritzat per Josep Quer i M. Teresa Espinal (Ribera, 2007). El projecte consisteix en el disseny d'una BD que permeti emmagatzemar i consultar informació lingüística sobre les unitats lèxiques de la LSC i té tres objectius bàsics:

7. Vegeu <http://www.xtec.net/dnee/act/miraquedic>.

8. Vegeu http://www.xtec.net/dnee/lsc/terminologia_llengua/index.htm.

— Crear una eina de consulta sobre les unitats lèxiques de la LSC per a qualsevol tipus d'usuari.

— Crear una eina d'investigació lingüística.

— Crear una BD per a l'elaboració de materials lexicogràfics.

El format digital permet emmagatzemar moltes més dades que el format imprès, i també permet organitzar els tres objectius dins d'una BD adaptada a la modalitat visual. Per fer-la possible hem tingut en compte les aportacions de les investigacions existents sobre LS i sobre BD en LS, les quals ens han ajudat a establir criteris de selecció i d'organització de la informació. A continuació veurem com està organitzada. Els estudis sobre la fonologia del signe ens han servit per a organitzar el sistema de cerca de les unitats lèxiques d'acord amb les característiques de la modalitat visual. Dins l'apartat de la morfologia veurem que també hi ha diferències entre les LS i les LO que haurem de tenir en compte. Les dificultats que trobem en la semàntica i la pragmàtica són les pròpies de qualsevol llengua natural, per la qual cosa en aquests apartats tractarem els criteris que hem establert per al disseny d'aquesta BD en concret.

Tal com hem dit anteriorment, les diferències de modalitat fan necessària la recerca de nous sistemes d'organització de les unitats lèxiques. En els sistemes utilitzats fins ara s'ha intentat adaptar la modalitat visual als sistemes de cerca existents, que permeten una organització alfabètica o temàtica. Però aquest tipus de cerca fa que la modalitat visual sigui una modalitat dependent de la glossa.

En aquest treball proposem un nou tipus d'organització per a la LSC basada en els paràmetres fonològics de la modalitat visual. En altres LS s'ha iniciat, ja fa un temps, la investigació en aquesta direcció perquè les noves tecnologies ens han facilitat aquest nou punt de vista. El sistema permet la cerca de les unitats lèxiques mitjançant els paràmetres fonològics de la LSC, a banda de la tradicional cerca mitjançant la glossa, més còmoda potser per a altres usuaris no familiaritzats amb la LS. Aquests paràmetres són els següents: la configuració (forma que adopten les mans i els dits mentre s'articula el signe), la localització (lloc on s'articula), el moviment (acció de les mans) i l'orientació (direcció del palmell de la mà).

Aquest sistema de cerca està pensat per a aplicar-se en BD digitals. Tal com podem veure en la figura 1, és un sistema de menús que condueixen a submenús, els quals de mica en mica, mitjançant la tria de paràmetres, condueixen l'usuari fins al signe que busca. La cerca més detallada comença amb el menú de la configuració. En aquesta primera pantalla hi apareixen totes les configuracions rellevants identificades en LSC.

L'usuari haurà de prémer la icona de la configuració⁹ que consideri que té el signe que busca. Aleshores apareixerà un menú en què l'usuari haurà de dir si el signe que busca es realitza amb dues o amb una sola mà. Si és amb dues mans, prèviament al pas següent, haurà de seleccionar el tipus de moviment de les dues mans; si és altern o és simultani.



FIGURA 1. Sistema de menús

Un cop escollida l'opció anterior, apareixerà la localització. L'usuari haurà de dir si el signe s'articula al cap, al tronc, al braç o bé a la mà no dominant, d'acord amb la divisió de l'espai signic que fan Sandler i Lillo-Martin (2006). Un cop dins l'àrea major del cos on s'articula el signe, l'usuari podrà dir si aquella configuració es produeix a l'inici, al mig o al final del signe. Aquest pas és important de cara als signes compostos.

Finalment li apareixerà en pantalla una imatge de tots els signes que compleixen els requisits demanats. Com que podrien ser molts, també cal buscar un ordre visual per a aquests signes. Per ordenar-los, hem fet servir el concepte d'*arranjament*, utilitzat per Sandler i Lillo-Martin (2006), el qual fa referència a la posició concreta de l'articulació del signe respecte a l'àrea major en què es troba. Aquest ordre pot ser el següent: dins d'una àrea major del cos, primer els situats en la zona alta, després en la zona mitjana i finalment els de la zona baixa. Si dos signes coincideixen en una d'aquestes zones, primer els que estan en el mateix costat que la mà que signa (ipsilateral) i després els que s'articulen a l'altre costat (contralateral). Finalment, si dos signes coincideixen en aquests punts esmentats, primer els que tenen contacte, després els que es troben a una distància mitjana (proximal) i finalment els que estan en una zona allunyada (distal).

Aquest sistema permet un nivell de concreció del signe bastant alt. A més, en qualsevol dels nivells de selecció es pot fer un accés directe a la llista amb els elements seleccionats fins aleshores, la qual cosa pot ser d'utilitat en cas que l'usuari desconegui més trets distintius del signe. A banda d'aquesta via de cerca principal que s'inicia

9. El quadre de configuracions de la figura 1 correspon a la LSQ (*langue des signes quebecoise*) i les imatges s'han tret de Martin i Alvarado, 2004 [1996].

amb la configuració del signe, però que després inclou els altres paràmetres fonològics dels signes de la LSC, una base de dades en format informàtic pot permetre llistar tots els signes que compleixen un requisit concret com és un moviment, una localització i una orientació determinada. Aquesta opció s'aconsegueix posant a la pàgina principal icones que permetin la cerca directa com ara:

— Cerca per moviment: apareixerà un desplegable amb els moviments més rellevants en LSC.

— Cerca per localització: apareixerà un desplegable amb àrees majors del cos que després es pot detallar amb l'arranjament.

— Cerca per orientació: no tots els autors la consideren un paràmetre fonològic independent, ja que alguns la inclouen dins la configuració, però pot afegir-se com a possibilitat de cerca, llistant també les orientacions més rellevants en LSC.

Per a fer possible aquest sistema de cerca caldrà determinar en primer lloc quines són les configuracions, els moviments, les localitzacions i les orientacions distintives en LSC i afegir aquesta informació detallada en l'entrada del signe. Un cop trobat el signe que busquem podrem accedir a la informació que té associada seleccionant-lo amb el ratolí.

Una de les tasques més importants del lexicògraf és decidir la informació que inclourà dins del treball lexicogràfic que vol elaborar. Cada unitat lèxica d'una llengua conté nombrosa informació associada. Aquesta informació permet que la unitat lèxica pugui combinar-se o no amb altres unitats de la llengua, de manera que el parlant pugui expressar conceptes i intencions mitjançant la unió d'aquestes unitats lèxiques en estructures ben formades. El lexicògraf haurà de seleccionar-la en funció dels objectius del material que vulgui elaborar i també dels usuaris a qui vagi dirigit aquest material.

D'altra banda, un cop escollida la informació que es recollirà en l'entrada lèxica, cal organitzar-la. El format informàtic permet crear un formulari d'entrada per a cada signe i diferents accessos per a cada un dels aspectes recollits. Mitjançant el sistema que proposem, que podeu veure a la figura 2, l'usuari podrà accedir ràpidament al que desitja saber, fent clic directament sobre la icona que conté el que li interessa. D'aquesta manera s'evitarà acumular tota la informació en una sola pantalla, cosa que podria fer la consulta molt feixuga.

A continuació aprofundirem en alguns dels apartats que apareixen a la figura 2. Per raons d'extensió hem escollit una mostra de l'ampli ventall d'aspectes que podrien recollir-se en un treball lexicogràfic.

DUTXA

- Estructura Eventiva i argumental
- Categoria léxica i selecció categorial
- Estructura morfològica
- Fonologia
- Definició
- Paradigma flexiu
- Construccions amb classificadors
- Sinònims
- Antònims
- Variants

FIGURA 2. Formulari d'entrada del signe

3.1. Informació morfosintàctica

La morfologia és aquella part de la lingüística que descriu l'estructura interna dels mots i els processos de creació de noves paraules en una llengua. La sintaxi, en canvi, és aquella àrea de la lingüística que estudia l'estructura de les oracions.

3.1.1. Aspectes morfològics dels signes

Els signes poden estar formats per un sol morfema o bé per la unió de diversos morfemes. En les LO, la combinació de morfemes per crear un lexema es realitza mitjançant la unió de radicals i afixos en seqüències lineals. En les LS, en canvi, hi ha dues maneres de combinar els morfemes (Sandler i Lillo-Martin, 2006):

- Mitjançant l'adjunció d'afixos.
- Afegint material fonològic als seus components.

En estudis que s'han fet en diverses LS, s'ha observat que és molt més productiva la segona opció que no pas la primera, la qual és més pròpia de les LO. La simultaneïtat en la llengua de signes fa possible que estructures monosil·làbiques siguin a la ve-

gada bimorfèmiques, i aquesta característica afecta tant la morfologia flexiva com la derivativa.

La morfologia flexiva és aquella que modifica la forma del mot per a adaptar-lo al context sintàctic on apareix, sense que aquest canvi formal impliqui un canvi en el significat lèxic. La morfologia flexiva que s'acostuma a recollir en els diccionaris sol ser la irregular, és a dir, la dels mots que no segueixen la derivació habitual de la llengua. En una BD lexicogràfica, com que no hi ha problemes d'espai, es podrien incloure en un apartat específic tant la morfologia flexiva regular com la irregular. D'altra banda, la morfologia derivativa forma part dels procediments que utilitzen les llengües per a crear noves paraules i, en conseqüència, les formes procedents d'aquest tipus de derivació es recolliran com a lemes o entrades independents.

Un dels fenòmens que trobem en LSC, igual que en altres LS com ara l'ASL (Sandler i Lillo-Martin, 2006), són parells de signes amb categoria diferent, nom/verb, que es distingeixen pel moviment. Es tracta de verbs que representen una activitat, mentre que els noms són concrets i estan semànticament relacionats amb l'activitat que denota el verb. Així, per exemple, en LSC hi ha parelles com ara *dutxar-se*/*dutxa* o bé *avió*/*volar-avió* que es distingeixen només pel moviment general de l'execució del signe. En el cas de *dutxar-se* l'arrel es reduplica i s'executa més ràpidament que en el cas de *dutxa*, que té un moviment únic més curt i tens. En l'exemple de *volar-avió* la trajectòria és més llarga i el moviment no és tens, mentre que en *avió* el moviment és curt i tens.

En aquests exemples el signe derivat es construeix mitjançant la modificació d'algun aspecte del signe d'origen, i el resultat és un signe amb un significat diferent, d'una altra categoria lèxica, que podria tenir dins la BD la seva pròpia entrada. Una descripció detallada d'aquests aspectes fonològics que produeixen canvis de significat, a més, podria ser molt valuosa a l'hora de fer recerca mitjançant consultes de selecció dins la BD lexicogràfica. Per aquest motiu, a més d'entrar aquests ítems com a dos signes independents, s'hi podrien descriure diferències fonològiques tals com si és curt o llarg respecte del signe d'origen, si és lent o ràpid, si reduplica l'arrel, etcètera.

Construccions amb classificadors

Una de les vies de derivació morfològica més productives en les LS són les construccions amb classificadors. D'aquest tipus de construccions també en trobem en algunes LO, però no és un fenomen tan habitual com ho és en les LS. Una cons-

trucció amb classificador consisteix en la unió d'un morfema arrel a un morfema classificador que fa que la unitat lèxica adquireixi una altra dimensió semàntica. En el cas de les LO, aquest morfema s'afegeix a l'arrel de manera seqüencial, mentre que en les LS és simultània. En ambdues modalitats, però, el resultat és una estructura morfològica composta, la qual sovint acaba lexicalitzant-se i formant una nova paraula.

La utilització de classificadors està sotmesa a restriccions. No tots els signes poden fer-se servir en aquest tipus de construccions. Per exemple, un signe com ara *pensar* no té construccions de classificadors associades, però en canvi sí que en té *menjar*, que adopta una forma o una altra segons si el que es menja és un entrepà, espaguetis, una pizza, pipes, etcètera.

Com que no es poden entrar dins la BD totes les formes que poden adoptar els verbs *menjar*, *beure* o *agafar*, perquè la llista podria ser immensa (encara que s'hi poden posar exemples), haurem de buscar la manera de donar a l'usuari aquesta informació sense saturar l'entrada. Una manera és fer constar si un signe té associades construccions amb classificadors o no i, en cas que en tingui, de quin tipus.

3.1.2. Aspectes sintàctics dels signes

Són els aspectes que determinen, per una banda, la ubicació de la unitat lèxica quan s'insereix dins d'una estructura sintàctica, i per l'altra, amb quins altres elements lingüístics pot combinar-se.

a) La categorització i la subcategorització

La categorització sintàctica defineix cada unitat lèxica respecte a la classe distribucional i funcional de què forma part. En català, per exemple, són categories sintàctiques els noms, els verbs, els adjectius, els adverbis, les preposicions i les conjuncions (Cabré, 1994). El fet de pertànyer a una categoria significa que l'ítem lèxic pertany necessàriament a subcategories d'ordre morfològic, les quals estan determinades per cada classe distribucional. Per exemple, i d'acord amb la classificació tradicional de Padden (1988), els verbs en LS es divideixen en els tipus següents:

— Verbs simples o invariables: aquells que no presenten concordança i que, per tant, no modifiquen la forma en funció dels arguments del subjecte i l'objecte (pensar, recordar, etc.).

— Verbs direccionals o de concordança gramatical: aquells que concorden amb el subjecte i l'objecte modificant paràmetres fonològics. *Ajudar, donar, enganyar, obligar*, etcètera.

— Verbs espacials o de concordança locativa: aquells que modifiquen la forma en funció dels arguments locatius del predicat, o de la localització dels seus arguments animats. *Posar, transportar*, etcètera.

D'altra banda, els verbs poden ser transitius o intransitius, per exemple. Dependent de si pertanyen a una subcategoria o a una altra hauran d'aparèixer a l'oració amb un complement o més d'un. Aquesta informació també podria recollir-se dins l'entrada, ja que això orientaria l'usuari sobre com ha de construir morfològicament i sintàcticament el signe.

b) Selecció categorial

D'altra banda, la subcategorització especifica les combinacions possibles de la unitat lèxica des de dos punts de vista: el de la selecció categorial (els ítems presenten restriccions respecte a les categories amb les quals poden combinar-se) i de la selecció semàntica (també presenten restriccions pel que fa al paper temàtic de les peces amb les quals es poden combinar).

Un exemple de selecció categorial és el fet que en LSC puguem dir *pa menjo* però no *blau menjo*. No podem dir *blau menjo* perquè el verb *menjar* selecciona un nom com a complement, no un adjectiu. Un exemple de selecció semàntica és que puguem dir, en LSC, *persona pensa* però no *cotxe pensa*. El verb *pensar* selecciona un agent animat que dugui a terme l'acció.

Aquest tipus d'informació va més enllà del que solen recollir els diccionaris normatius i per això l'esmentem breument. No obstant això, és interessant saber que se sol recollir en els diccionaris anomenats *combinatoris*.

c) Informació semàntica

És aquella que fa referència al significat del signe.¹⁰ És fonamental poder definir-lo. En els materials existents sovint s'ha utilitzat la glossa com a única definició del signe. La

10. Hi ha alguns aspectes de l'àrea de la semàntica que també podrien recollir-se en un material lexicogràfic com l'estructura eventiva i l'estructura argumental de la unitat lèxica, la selecció semàntica, o els sinònims i antònims. Per raons d'extensió no els tractarem en aquest treball.

glossa pot ser utilitzada en algunes ocasions, però mai no pot substituir la definició feta en LS, atès que en pot escapar el significat. D'entrada la glossa és un mot d'una altra llengua, i els significats no sempre són equivalents. També ens podem trobar que una única glossa pugui definir més d'un signe que en LS tenen diferents matisos, o bé que diverses glosses puguin servir per a un sol signe.

Per això la millor opció és definir els signes en la pròpia llengua, mitjançant altres signes, i tenint present, entre d'altres, els aspectes següents:

— Fer les descripcions amb signes d'ús comú i d'alta freqüència entre els signants per tal que siguin tan comprensibles com sigui possible per a la majoria de signants de la comunitat.

— També cal fer una extracció del significat de tots els exemples que es trobin en el corpus i ordenar les accepcions seguint alguns criteris definits. Un signe es pot utilitzar en molts contextos i tenir, en cada un d'aquests contextos, un matís diferent. No podem recollir-los tots perquè la llista d'accepcions seria massa llarga i l'usuari s'acabaria perdent. Cal, doncs, fer una síntesi de tots els significats que trobem. També cal establir un ordre de les accepcions sintetitzades. Un criteri d'ordre podria ser la freqüència amb què un signe s'utilitza amb aquell significat, o bé la posició del signe dins de la comunitat, és a dir, llistar primer el significat comú a un col·lectiu més gran de signants dins la comunitat. Un altre criteri que també s'utilitza sovint és el criteri lògic, que és aquell que llista primer els significats bàsics del mot i després els significats relacionats. En tot cas, hi ha d'haver una certa flexibilitat per a intentar trobar el criteri que s'adapta millor a cada entrada.

— S'han d'evitar les definicions circulars. Una definició circular és aquella on la definició d'una entrada A es basa en el concepte d'una entrada B i la definició de l'entrada B es basa en el concepte de l'entrada A. De vegades les definicions es fan mitjançant sinònims. Aquest pot ser un recurs útil, però s'ha de vigilar perquè es pot caure fàcilment en una definició circular en què un signe es defineix mitjançant un sinònim, i la definició d'aquest darrer remet altra vegada al primer.

Una tècnica que es fa servir sovint per a definir les unitats lèxiques d'una llengua és recórrer a ítems de la mateixa categoria lèxica, ja que d'aquesta manera s'està facilitant simultàniament informació sobre la seva ubicació dins l'estructura sintàctica.

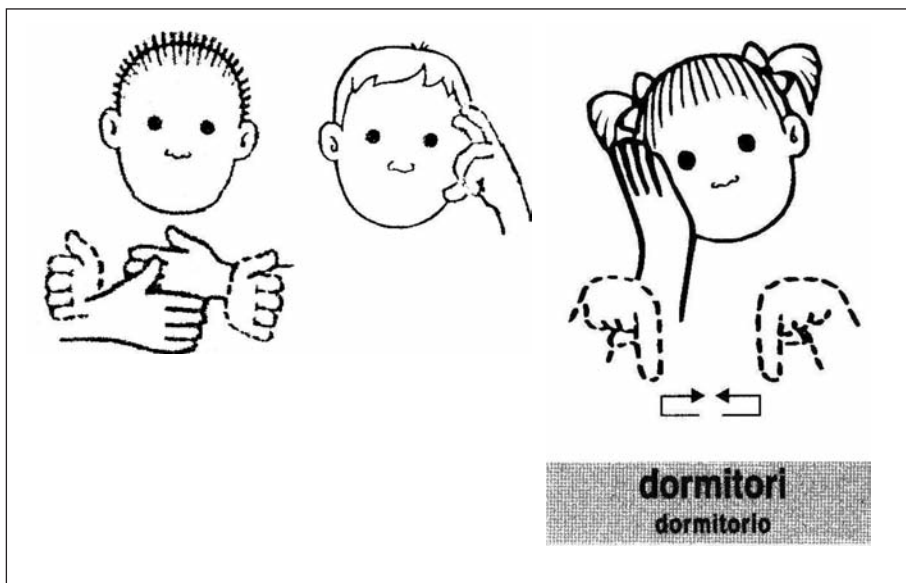
d) Informació pragmàtica

La informació pragmàtica és aquella relacionada amb l'ús de la peça lèxica en un context comunicatiu determinat. Les peces lèxiques formen part d'una xarxa associativa

basada en elements de l'experiència personal i col·lectiva. Això vol dir que els parlants interioritzen en quines situacions s'usen les paraules amb més o menys freqüència, com són les persones que les fan servir, quines reaccions provoquen quan s'utilitzen en un acte comunicatiu, etcètera.

En LSC, com en qualsevol llengua natural, podem trobar diferents tipus de variants:

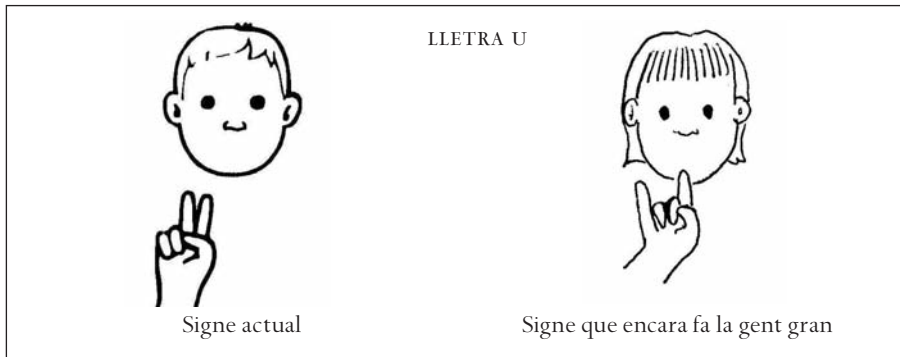
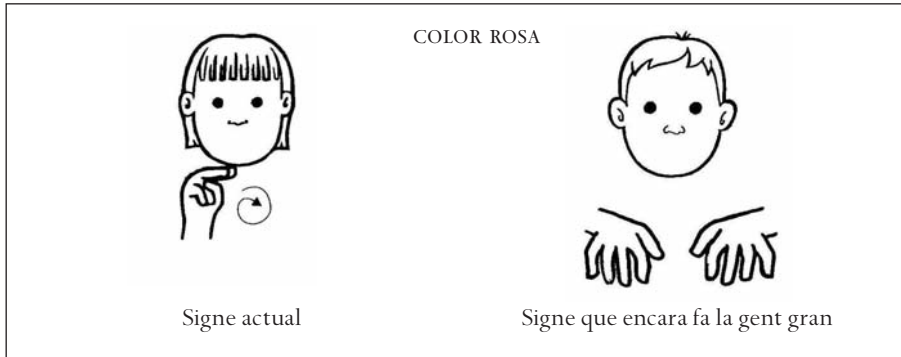
— *Variants diatòpiques*: ítems lèxics que s'usen en una zona geogràfica concreta o bé que presenten diferències d'ús, o d'execució, o d'interpretació segons la zona. Aquest és el cas de la paraula *dormitori*, que a la zona de Barcelona es diu d'una manera i a la zona de Lleida, d'una altra¹¹.



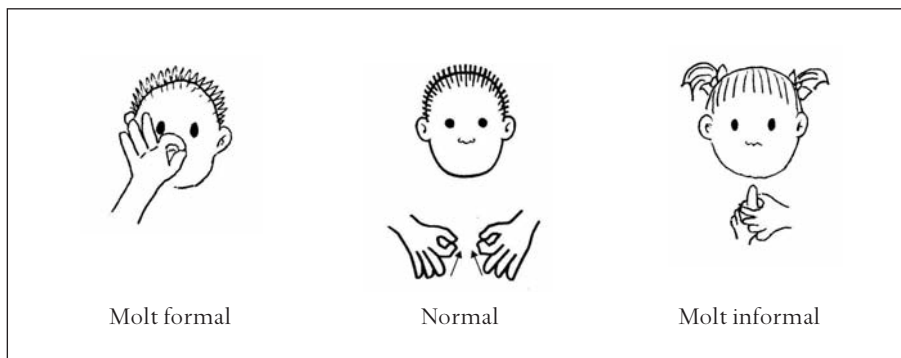
— *Variants diastràtiques*: ítems lèxics que utilitza un sector social específic de la comunitat sorda, per exemple els signes *color rosa* o bé *lletra u*.¹²

11. El primer dibuix està extret del material facilitat per LLESIG (Barcelona) per a l'aprenentatge de la llengua de signes, i el va fer Frank Vidiella (curs 2006-2007). El segon dibuix és del DTLSC (Martín i Alvarado, 2004), que conté signes de la zona de Lleida.

12. Dibuixos extrets del DTLSC (Martín i Alvarado, 2004).



— *Variants diafàsiques*: ítems lèxics que s'utilitzen en alguns registres concrets o signes que presenten diferències segons el registre on apareguin. Un exemple de variant diafàsica pot ser el signe *preservatiu*, que presenta tres variants segons la situació.



— *Variants diatècniques*: ítems lèxics propis d'una matèria especialitzada. El signe *dins*, per exemple, significa «matèria» quan fa referència a l'educació.



Definició 1: situació a l'interior d'un espai limitat.

Definició 2: (en educació) contingut, matèria, aspecte conceptual d'una assignatura, curs, etcètera.

Com que les BD permeten vincular els ítems, les variants poden vincular-se, de manera que *dins* la informació d'un signe es pugui accedir a les variants de diferents tipus existents en el domini lingüístic.

En aquest apartat hem intentat mostrar exemples del tipus d'informació que es podria incloure en una BD lexicogràfica. Tal com hem dit anteriorment, la informació que es reculli en el material lexicogràfic dependrà dels seus objectius i del tipus d'usuaris als quals vagi dirigit.

4. Conclusions

Les diferències de modalitat de les LS i les LO fan que el seu component lèxic no pugui descriure's mitjançant els mateixos paràmetres. Ambdues modalitats comparteixen principis lingüístics però divergeixen en els paràmetres que les configuren i en la manera com es transmet i es rep la competència gramatical que els parlants tenen interioritzada. Aquestes divergències han afectat els estudis sobre la lexicologia del signe, molt endarrerida respecte a la de la modalitat oral, i també els sistemes d'emmagatzematge dels ítems lèxics de les LS, ja que les vies tradicionals de recollida de dades s'adapten molt poc o gens a la modalitat visual.

Les diferències de modalitat fan que calgui buscar altres vies de descripció i d'emmagatzematge dels ítems lèxics de les LS, altres vies que vagin més enllà de la mera transcripció del significat a una glossa en la LO més propera. Tal com hem vist, les noves tecnologies ens posen a l'abast l'oportunitat d'oferir als signants una BD lexicogràfica molt més completa amb un ventall ampli d'informació sobre els signes que pot incloure, què signifiquen, amb quines altres peces es combinen i en quins contextos apareixen o s'utilitzen, entre altres aspectes.

En aquest article hem descrit l'estat de la qüestió en les LS en general. Ens hem centrat també en l'estat de la lexicografia en la LSC i hem esbossat les característiques que hauria de tenir una BD que tingués en compte les especificitats intrínseques de les llengües de signes. Esperem que aquest nou tipus d'organització de BD, basada en els paràmetres fonològics de la modalitat visual, pugui servir a futurs investigadors en la creació i aplicació de BD digitals.

Referències bibliogràfiques

- ALVARADO, M.; MARTÍN, I. (2004). *Diccionari temàtic de llenguatge de signes català*. Lleida.
- BOYES BRAEM, P. (2001). «A multimedia bilingual database for the lexicon of Swiss German Sign Language». A: *Sign Language and Linguistics*, vol. 4, núm. 1-2, p. 133-143.
- BRIEN, D. (1992). *Dictionary of British Sign Language / English*. London-Boston: Faber and Faber.
- CABRÉ, M. T. (1994). *A l'entorn de la paraula (I); Lexicologia general*. València: Universitat de València.
- CRASBORN, O. [et al.] (2001). «A phonological database for sign languages». A: *Sign Language and Linguistics*, vol. 4, núm. 1-2, p. 215-228.
- GIROD [et al.] (1997). *Dictionnaire LSF - français*. París.
- GRAS, V. (2006). *La Comunidad Sorda como comunidad lingüística: panorama sociolingüístico de las lenguas de signos en España*. [Tesi doctoral]. Barcelona: Universitat de Barcelona. Facultat de Filologia. Departament de Lingüística.
- HANKE, T.; STORZ, J. (2008) «ILex – A database tool integrating sign language corpus linguistics and sign language lexicography». A: *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, 26 de maig - 1 de juny de 2008. Marràqueix, Marroc: European Language Resources Association.
- JOHNSTON, T. (2001). «The lexical database of Auslan (Australian Sign Language)». A: *Sign Language and Linguistics*, vol. 4, núm. 1-2, p. 145-169.
- JOHNSTON, T. (2008). «Corpus linguistics & signed languages: no lemmata, no corpus». A: *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, 26 de maig - 1 de juny de 2008. Marràqueix, Marroc: European Language Resources Association.
- KÖNIG, L. [et al.] (2008). «Corpus-based Sign Dictionaries of Technical Terms — Dictionary Projects at the IDGS in Hamburg». A: *Proceedings of the Sixth International Con-*

- ference on Language Resources and Evaluation*, 26 de maig - 1 de juny de 2008. Marràqueix, Marroc: European Language Resources Association.
- KONRAD, R. [et al.] [Inèdit] «What's in a sign? Theoretical lessons from practical sign language lexicography». A: QUER, J. (ed.). *Signs of the time*. Hamburg: Signum (*Selected Papers from TISLR 8*).
- KRAMMER, K. [et al.] (2001). «The Klagenfurt database for sign language lexicons». A: *Sign Language and Linguistics*, vol. 4, núm. 1-2, p. 191-201.
- KRAMMER, K. [et al.] (2008). «The Klagenfurt lexicon database for sign languages as a web application: LedaSila, a free sign language database for international use». A: *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, 26 de maig - 1 de juny de 2008. Marràqueix, Marroc: European Language Resources Association.
- NEIDLE, C. (2001). «SignStream: A database tool for research on visual-gestural language». A: *Sign Language & Linguistics*, vol. 4, núm. 1-2, p. 203-214.
- PADDEN, C. (1988). *Interaction of Morphology and Syntax in American Sign Language*. Nova York: Garland. (Outstanding Dissertations in Linguistics) [Originalment publicada com a: *PhD dissertation*. San Diego: Universitat de Califòrnia, 1983].
- PERELLÓ, J.; FRIGOLA, J. (1998). *Lenguaje de signos manuales*. Madrid: CIE Inversiones Editoriales Dossat 2000.
- PRILLWITZ, S. [et al.] (2008). «DGS corpus project — Development of a corpus based electronic dictionary German Sign Language / German». A: *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, 26 de maig - 1 de juny de 2008. Marràqueix, Marroc: European Language Resources Association.
- QUER, J. [et al.] (2005). *Gramàtica bàsica LSC*. CD-ROM. Barcelona: Federació de Persones Sordes de Catalunya.
- RIBERA, E. (2007). *Projecte de base de dades lexicogràfica per a la LSC (llengua de signes catalana)*. [Trebball de recerca dins el programa de doctorat Ciència cognitiva i llenguatge. Director: Josep Quer. Tutora: Maria Teresa Espinal].
- SANDLER, W.; LILLO-MARTIN, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.
- SAVOLAINEN, L. (2001). «The database system used in the Finnish Sign Language Dictionary Project». A: *Sign Language and Linguistics*, vol. 4, núm. 1-2, p. 125-132.
- SCHERMER, T. [et al.] (2001). «Developing linguistic specifications for a sign language database: The development of Signbase». A: *Sign Language and Linguistics*, vol. 4, núm. 1-2, p. 253-274.
- SENGHAS, A. (2001). «Spatial and temporal coding of Nicaraguan Sign Language in-MediaTagger». A: *Sign Language & Linguistics*, vol. 4, núm. 1-2, p. 229-240.

- SLOBIN, D. [et al.] (2001). «Sign language transcription at the level of meaning components: The Berkeley Transcription System (BTS)». A: *Sign Language & Linguistics*, vol. 4, núm. 1-2, p. 63-104.
- STOKOE, W. (1960). *Sign Language structure: an outline of the visual communications systems of the American deaf*. Buffalo: University of Buffalo. (Studies in Linguistics: Occasional Papers.)
- STOKOE, W. [et al.] (1965). *A Dictionary of American Sign Language on Linguistic Principles*. Washington DC: Gallaudet College Press. [2a ed., 1976, per Linstok Press].
- TERVOORT B. (1953). *Strukturele Analyse van visueel taalgebruik binnen een groep dove kinderen*. Amsterdam: North Holland Publications.
- TROELSGARD, T.; KRISTOFFERSEN, J. H. (2008). «An electronic dictionary of Danish Sign Language». *Sign Languages: spinning and unraveling the past, present and future. TISLR 9, forty five papers and three posters from the 9th. Theoretical Issues in Sign Language Research Conference, Florianópolis, Brasil, desembre 2006*. QUADROS R. M. de (ed.). [En línia]. [Petrópolis, Rio de Janeiro, Brasil]. Editora Arara Azul. <<http://www.editora-arara-azul.com.br/EstudosSurdos.php>>.

Recursos web (consulta: juliol del 2008)

- <<http://gebaren.ugent.be>>
- <<http://ledasila.uni-klu.ac.at>>
- <<http://www.auslan.org.au>>
- <<http://www.gebarentrum.nl/gebarenwoordenboek.asp>>
- <<http://www.sign-lang.uni-hamburg.de>>
- <<http://www.sign-lang.uni-hamburg.de/Projectseng/SLDicts.html>>
- <<http://www.xtec.net/dnee/act/miraquedic>>
- <http://www.xtec.net/dnee/lsc/terminologia_llengua/index.htm>
- <www.tegnsprog.dk>