

## Instruments

**L'aportació de la Comissió Lingüística en el projecte del traductor automàtic entre el català i l'occità****Autor**

Aitor Carrera  
Universitat de Lleida

**L'article exposa la feina feta per una comissió de lingüistes procedents de tot el territori occità per fixar les formes i les estructures de l'occità estàndard que el traductor automàtic català-occità-català de la Generalitat de Catalunya ha de poder produir a partir d'un text català i ha de poder traduir també al català. S'hi exposen els precedents, els objectius, el funcionament de la Comissió Lingüística i les diverses solucions adoptades. Finalment es valora també el treball realitzat.**

**Precedents**

La segona llengua territorial del Principat de Catalunya és l'occità, que a l'Aran és anomenat popularment *aranès*. Les modalitats lingüístiques araneses són varietats del gascó, un dels sis dialectes de la llengua occitana. Tot i que se'n reconeix l'existència a través de lleis de segon ordre —i fins i tot a través d'alguna decisió regional recent—, i malgrat que l'occità és la segona llengua minoritzada de la Unió Europea en nombre de parlants (es calcula que cada dia és parlat per més de tres milions de persones), la llengua occitana no és oficial en la major part del seu territori, administrativament lligat a l'Estat francès. És oficial a l'Aran des de 1990 (any de la coneguda Lei d'Aran), i a tot el Principat d'ençà de la darrera reforma estatutària, i té un reconeixement a les valls occitanes del Piemont des de l'aparició de les *Norme in materia di tutela delle minoranze linguistiche storiche* de 1999. Aquesta anòmala situació legal de l'occità en bona part d'Occitània, juntament amb el fet que aquest país no ha format (gairebé) mai una unitat administrativa, és un dels factors que provoquen que l'occità no tingui una varietat estàndard coneguda i reconeguda pel conjunt de la població del seu país. La llengua occitana té encara una presència residual i purament folklòrica en els grans mitjans de comunicació de massa del seu territori, i a causa de les polítiques assimilatives de l'Hexàgon, bona part de la població occitana té —a diferència de fa cent anys— el francès com a llengua habitual.

Malgrat el que acabem de dir, la codificació de la llengua occitana no és un terreny precisament verge. Gairebé tot l'occitanisme modern admet la reforma de Loís Alibèrt, i en general s'accepta que hi ha una forma de referència d'occità de base llenguadociana anomenada *occitan referencial*, *occitan larg*, *occitan general* o *occitan estàndard*, al costat d'uns estàndards regionals convergents per a cada bloc dialectal (gascó, llemosí, alvernès, vivaroalpí, provençal) i fins per a algun territori amb una forta personalitat (com Niça). Les varietats referencials —la general i les regionals— de la llengua no han de substituir les formes dialectals o subdialectals de l'occità, sinó omplir un espai al qual fins ara l'occità no ha pogut accedir, o si més no, ha accedit només d'una manera més aviat limitada: el de la comunicació supradialectal.

## La necessitat d'una comissió de lingüistes

El dia 4 de maig de 2007, per iniciativa de la Secretaria de Política Lingüística de la Generalitat de Catalunya i amb el suport del Conselh Generau d'Aran, es presentava a Vielha una comissió de lingüistes encarregada de fixar les formes i les estructures de l'occità general que havien d'aparèixer en un traductor automàtic català-occità occità-català com a equivalents de les formes i de les estructures de la llengua catalana. Aquell mateix dia a la tarda, a la localitat aranesa de Betren, la Comissió Lingüística començava la primera d'una sèrie de sessions de treball que s'han anat repetint durant gairebé un any. A causa de la finalització del traductor automàtic, el dia 30 de març de 2008 es va aturar l'activitat de la Comissió, almenys fins que li arribi un nou encàrrec.<sup>1</sup>

La Comissió estava formada sobretot per lingüistes de prestigi, provinents de diferents regions occitanes de l'Estat francès, però també de les valls occitanes del Piemont (administrativament italianes) i de l'Aran. L'autor d'aquest document era l'encarregat de la direcció científica d'aquesta comissió, que actuava en paral·lel a una comissió «tècnica» en què hi havia representants polítics i institucionals de les diferents regions que donaven suport al projecte. Els membres de la Comissió Tècnica s'encarregaven, entre d'altres coses, d'estudiar la possible difusió que podia tenir la feina de la Comissió Lingüística, i les accions que se'n podien derivar.

## Composició de la Comissió

En la Comissió Lingüística hi era present gairebé tot el territori occità. Hi havia representants de Llenguadoc-Rosselló (Domergue Sumien i Claudi Balaguer), de Miègjorn-Pirenèus (Patrici Pojada i Patric Sauzet), d'Aquitània (Alain Viaut i Maurici Romieu), de Ròse-Aups (Joan Claudi Rixte i Joan Claudi Forêt), del Piemont (Dario Anghilante i Rosella Pellerino) i de l'Aran (Miquèu Segalàs i Jèp de Montoya). A més, en la Comissió hi havia també dos representants de l'Institut d'Estudis Occitans (IEO; Jacme Taupiac i Ives Lavalada) per la trajectòria històrica d'aquesta entitat en la promoció i defensa de la llengua occitana i, en un primer moment, en la seua codificació.

Uns quants membres —la majoria— de la Comissió Lingüística no només havien participat activament en experiències precedents de codificació de la llengua, sinó que són autors d'obres gramaticals o lexicogràfiques de referència sobre el conjunt de l'occità o sobre alguna de les seues varietats dialectals. El compromís de tots els membres de la Comissió va fer que gairebé tots participessin d'una manera activa en totes les jornades de treball malgrat les grans distàncies que en alguns casos separaven els seus domicilis de l'Aran, on s'han dut a terme totes les sessions de la Comissió. Aquells —pocs— que extraordinàriament no podien desplaçar-se fins a Vielha o Betren, trametien sovint per escrit a la direcció científica els seus comentaris o els seus suggeriments sobre algunes de les decisions presents en les actes de les sessions, amb l'objectiu que poguessin ser exposades a la resta de membres de la Comissió.

Remarquem que el fet que les regions occitanes designin representants per participar en una comissió encarregada de dur a terme tasques lingüístiques és un fet inèdit fins ara. És també un precedent, ja que aquest sistema de representació ha inspirat segurament el del darrer intent de crear una acadèmia de la llengua occitana per part de l'Institut d'Estudis Occitans i del Conselh Generau d'Aran, dues de les institucions que donaven suport a la tasca de la Comissió Lingüística impulsada per la Secretaria de Política Lingüística de la Generalitat de Catalunya.

La Comissió Lingüística reunia membres provinents de l'IEO i també alguns lingüistes designats per les regions que havien participat en el Conselh de la Lengua Occitana (CLO; d'entrada, el president i el vicepresident), la darrera temptativa d'aconseguir una autoritat normativa per a la

1 Vegeu l'article de Mònica Pereña «Dos nous parells de llengües al servei de traducció automàtica de la Generalitat de Catalunya: català-occità-català i castellà-occità-castellà», publicat en el número 42 d'aquesta revista.

llengua occitana, que feia diversos anys que vivia un període d'inactivitat pública. Durant el mes de desembre de 2007 aquesta entitat publicava un document en què es recollien i s'ordenaven les *Preconizacions* elaborades d'ençà de 1997, i en què el seu president valorava positivament el treball de la Comissió i el posava com a model: «Aquesta estructura es una bona prefiguración, çò'm par, de l'institucion que nos caldrà desenant far funcionar per contunhar l'òbra del CLO, que siá aquela institucion lo CLO meteís o una altra estructura».<sup>2</sup>

### Objectius i funcionament de la Comissió

La Comissió Lingüística ha realitzat un total de deu sessions de treball: 4 i 5 de maig de 2007; 14 i 15 de setembre de 2007; 1 i 2 de desembre de 2007; 2 i 3 de febrer de 2008; 29 i 30 de març de 2008. Les sessions han tingut una durada mínima de vuit hores i màxima de quinze, la qual cosa vol dir que els membres de la Comissió Lingüística han treballat conjuntament durant més de cinquanta hores, en les quals han pres tota mena de decisions de tipus lingüístic. Com ja hem assenyalat, la feina de la Comissió consistia sobretot —si deixem de banda la selecció d'un corpus textual inicial que serviria per veure els elements occitans més freqüents en textos representatius— a fixar les formes i les estructures occitanes que el traductor havia de produir a partir d'un text en català. Però la seua tasca no s'acabava aquí. La Comissió Lingüística també s'havia d'encarregar de fixar el conjunt de formes i d'estructures occitanes que el traductor automàtic havia de ser capaç d'interpretar, i per tant de traduir a la llengua catalana. Remarquem que aquesta qüestió és tan important com l'anterior si tenim en compte que la llengua occitana presenta una variació diatòpica molt més important que la llengua catalana.

La Comissió Lingüística ha fet tries gràfiques en casos en què en occità no hi havia una única solució acceptada per tothom. Ha dut a terme seleccions fonètiques, morfològiques i lexicals en situacions en què la diversitat de l'occità és notable i en què era imperatiu de trobar una solució única, perquè el traductor automàtic només pot generar un equivalent occità de cada element català, malgrat que hi pugui haver més d'una solució vàlida i admissible des d'un punt de vista normatiu. Ha determinat un bon nombre d'estructures sintàctiques i locucions en què el català i l'occità no coincideixen, i fins i tot ha examinat fets que separen l'occità general de l'occità de l'Aran malgrat que el seu objectiu principal no eren les formes del gascó parlat al Principat. La Comissió Lingüística ha parlat diverses vegades de l'aranès i ho ha recollit en els seus documents, sobretot amb l'objectiu d'acostar aquesta forma de gascó a la varietat occitana general del traductor automàtic, per evitar l'allunyament de l'aranès de la resta de l'occità, provocat pel fet que l'Aran pertany a un estat diferent de l'estat al qual pertany la major part d'Occitània.

La direcció científica dirigia les sessions de la Comissió, en fixava l'ordre del dia i determinava els afers que calia estudiar en cada trobada. També recollia les conclusions a les quals arribava la Comissió mitjançant actes i documents annexos —redactats intencionadament en aranès— i, després de la seua aprovació en la següent sessió, trametia aquests documents a l'empresa encarregada de realitzar el traductor automàtic per tal que hi introduís els acords de la Comissió Lingüística. La direcció científica, per altra banda, també assessorava tècnics d'aquella empresa entre sessió i sessió de la Comissió Lingüística. Aquest assessorament ha anat de la simple resolució de dubtes episòdics en el procés d'introducció de dades a la confecció de documents lingüístics en què es resumien les principals diferències sintàctiques entre el català i l'occità, passant per la recerca i la documentació bibliogràfiques sobre alguns punts lingüístics —sobretot sintàctics i lexicals— conflictius.

---

2. La revista electrònica *Lingüística Occitana* va dedicar el seu sisè número a publicar aquest document, *Preconizacions del Conselh de la Lengua Occitana* [<http://www.revistadoc.org>].

### La llengua del traductor. Els acords de la Comissió Lingüística

Ja hem advertit abans que l'occitanisme accepta d'una manera general que hi ha un model lingüístic de referència de base llenguadociana. El llenguadocià no només és la modalitat occitana més propera al català, sinó el dialecte central i el més usat en les publicacions actuals. És una varietat que evita certes originalitats —sobretot fonètiques— dels dialectes perifèrics, i al mateix temps és la modalitat més important des d'un punt de vista geogràfic, ja que és l'única que limita amb tots els altres dialectes. Per tant, la Comissió Lingüística —en la qual, recordem-ho, hi ha lingüistes de totes les zones dialectals d'Occitània— no va dubtar a proposar que les formes i estructures occitanes que havia de produir el traductor automàtic fossin coherents amb les que es proposen en els treballs publicats fins ara sobre la codificació de la llengua occitana, i més tenint en compte que ja existia una base de traducció automàtica entre el català i el gascó de l'Aran, la varietat occitana que es parla al Principat.

En el moment d'establir un corpus textual que servís de base per al futur traductor automàtic, la Comissió Lingüística va acordar que hi tinguessin una representació especial —sense deixar de banda els documents en altres varietats occitanes— els textos escrits en llenguadocià, i sobretot en aquest occità de referència de base llenguadociana que vol servir per a la comunicació supradialectal. Això es va materialitzar, per exemple, en la inclusió en el corpus des de textos de la *Gramatica occitana* de Loís Alibèrt fins a d'altres documents signats per alguns dels mateixos membres de la Comissió, passant per produccions occitanes de la xarxa o d'autors literaris llenguadocians presos sovint com a model per a l'elaboració de l'estàndard general occità (per exemple, com era previsible, Joan Bodon).

El llenguadocià que produeix el traductor automàtic parteix d'allò que en alguns treballs (per començar, en l'històric article de Bec de 1972),<sup>3</sup> ha estat anomenat *occitan mejan* o *occitan central*. Altrament dit: el llenguadocià que no és meridional, ja que les modalitats llenguadocianes del sud comparteixen unes quantes característiques amb el gascó —el dialecte occità més original— que no es repeteixen en altres parlars llenguadocians, i anuncien alguns trets presents en les llengües geogràficament peninsulars —com el català— que no fan acte de presència en l'occità septentrional, en provençal o —altre cop— en la resta del llenguadocià. Això fa, per exemple, que la traducció del català *llet* sigui *lach* i no pas *lait* (ni *lèit*); que *fet* sigui *fach* i no pas *fait* (ni *fèit*); que l'article definit masculí singular sigui *lo* i no pas *le*; o que el traductor produeixi un article partitiu en les traduccions occitanes (*vòli de pan* és l'equivalent occità de *vull pa*) que no s'utilitza en alguns parlars meridionals del Llenguadoc, que actuen per tant com el gascó o com el català. Que el traductor automàtic produeixi un occità que té com a base uns determinats parlars llenguadocians no vol dir que no es tinguin en compte la resta de varietats occitanes, ni tampoc que s'oblidin els altres dialectes occitans per determinar allò que ha de produir el traductor automàtic, i de fet per decidir allò que és propi de l'occità estàndard general.

Ja hem assenyalat que la Comissió Lingüística va decidir que el traductor reconegués formes pròpies d'altres dialectes occitans sempre que fos possible, sobretot si eren formes o estructures amb un pes territorial o simbòlic important, o almenys solucions típiques d'un dels grans dialectes occitans. Malgrat que —com ja hem avançat, també— caldria fixar d'una manera més sistemàtica els sinònims territorials de les formes que ara produeix el traductor automàtic, actualment aquesta eina ja sap interpretar certes formes i estructures que no són llenguadocianes i que podrien ser vistes com a pròpies dels estàndards regionals. La Comissió Lingüística ha proposat, per exemple, que es pugui traduir l'article plural invariable provençal *lei(s)* malgrat que el traductor automàtic produeixi *los* i *las* (i això consta justament en el primer acord de la Comissió Lingüística, del 4 de maig de 2007), i que caldria preveure la presència de la desinència de plural *-ei* en la resta de categories

3 Bec, Pèire. «Per una dinàmica novèla de la lenga de referéncia: dialectalitat de basa e diasistèma occitan». *Annales de l'Institut d'Etudes Occitanes*, 4e série, II, 6, 1972. P. 39-61.

nominals. Igualment, el traductor automàtic produeix la preposició *amb* com a equivalent del català *amb*, però reconeix les formes *emb*, *dab* o *damb*, i fins i tot formes amb *-e* promocionades per alguns lingüistes —tot i que no pel CLO— o usades per alguns autors, com *ambe* o *embe*. Tanmateix, insistim que queda molta feina per fer en l'àmbit de les característiques dels cinc dialectes perifèrics, i que en el futur seria bo que un parlant d'occità vivaroalpí, alvernès, llemosí, provençal o gascó —que no sigui aranès— pogués traduir al català textos escrits en el seu occità sense haver de recórrer a les solucions de l'estàndard general.

La Comissió Lingüística s'ha encarat sovint amb dificultats importants, ja que de vegades no hi ha una única solució que es pugui considerar estàndard o almenys millor que d'altres. Generalment, en aquesta mena de situacions en què hi havia diverses sortides completament vàlides, s'ha acabat donant prioritat a les solucions més operatives o funcionals des d'un punt de vista tècnic (hi han sortit guanyant, per exemple, les formes que evitaven les ambigüitats i les homonímies). Les decisions han estat preses d'una manera general per consens, i només en uns pocs casos ha calgut recórrer a la majoria. L'obligatorietat de traduir cada forma catalana per una única forma occitana ha fet que, per exemple, finalment hi hagi hagut més membres de la Comissió que s'han estimat més que l'equivalent occità de la forma catalana *illa* sigui *isla* i no pas *illa*. Tot i això, la Comissió Lingüística ha deixat constància del fet que probablement hi ha bones raons per triar qualsevol d'aquestes dues formes (acord 51 del 3 de febrer de 2008), de la qual cosa potser algú en podrà deduir que totes dues podrien ser considerades com a estàndards, o que almenys hi ha una part important dels principals lingüistes del país que estarien a favor d'aquesta opció o que preferirien la forma que ara no produeix el traductor. Un cas semblant s'ha produït en el moment de parlar dels descendents del grup *-RS-*, ja que ha calgut triar entre *borsa* i *bossa*, o entre *orsa* i *ossa*, posem per cas (acord 54 del 3 de febrer de 2008).

L'occità produït pel traductor automàtic és fidel a la norma. Dit d'una altra manera: escriu *amb* i *doas*, i no pas *ambe* i *doás*. Tot i això, la Comissió Lingüística ha proposat que el traductor automàtic reconegui formes que no són normatives o que són dubtosament normatives però que tenen un grau important de socialització, i sobretot una presència habitual en els documents occitans actuals i en els del corpus textual. Això no vol pas dir que se'n vulgui legitimar l'ús, ni tampoc que es vulgui sancionar certes formes com a normatives. Torna a ser una altra decisió funcional, que d'una banda evita les discussions estèrils sobre determinats fets polèmics (recordem que la Comissió no té el rol d'autoritat normativa) i que, d'una altra, busca que els usuaris occitans del traductor s'hi puguin reconèixer fins i tot en cas que hi introdueixin algunes solucions antinormatives o paranormatives amb una certa difusió. De fet, la Comissió Lingüística no podia assumir la responsabilitat de solucionar en un temps rècord tots els afers polèmics o problemàtics de la codificació de la llengua occitana, i el traductor automàtic no havia de ser tampoc l'estri prescriptiu de referència de la llengua occitana. Sense perdre de vista la norma, per tant, es fan concessions a formes o a representacions gràfiques que és probable que no apareguin en la majoria de les gramàtiques i els diccionaris occitans actuals, i que segurament tampoc no ho faran en els del futur. Això significa que el traductor genera la forma normativa *paur* i no pas la incorrecta *páur*, però que un usuari que escrigui *páur* obtindrà també *por* en català. El traductor automàtic produeix *astre* i reconeix *bonaür*, però alhora admet el francesisme *bonur*. La Comissió va decidir que el traductor acceptés la representació *çaquelà* al costat de la representació prioritària *ça que la* (acord 31 del 5 de maig de 2007), l'article provençal *li(s)* a més del normatiu *lei(s)* (acord 1 del 4 de maig de 2007) i fins i tot mots prefixats en què es fa un ús incorrecte del guionet (acord 5 del 2 de febrer de 2008). La Comissió Lingüística, per tant, va posar la utilitat del traductor automàtic per davant de la prescripció. Entenem que aquesta mesura, que ha fet avançar enormement el treball de la Comissió, si és anunciada com cal no perjudica la norma, ja que no afecta el procés de producció de textos en occità.

No seríem sincers si, malgrat tot el que hem dit, no assenyaléssim que hi ha hagut decisions de la Comissió Lingüística que podrien ser perfectament deliberacions d'una autoritat normativa. Hi ha hagut casos en què la Comissió ha modificat directament d'una manera intencionada algunes de les tries d'Alibèrt i ha estudiat certs afers que fins ara no havien estat exposats en cap dels documents

de les entitats o institucions que tenien vocació de ser el referent normatiu de la llengua occitana. D'aquesta manera, posem per cas, el traductor automàtic produeix *equipa* i *pàtria* malgrat que Alibèrt prefereix *esquipa* i *patria* (acord 53 del 30 de març de 2008; acord 79 del 3 de febrer de 2008). Aquesta mena de decisions intencionadament normatives o no del tot coincidents amb la norma han suscitat debats i reflexions molt productius, i algunes vegades han estat el resultat d'investigacions i de petites recerques d'una altíssima qualitat que la Comissió ha encarregat a algun dels seus membres (és el cas de *esquipa* i *equipa*, o el de la presència de *sos-*, *so-* i *sub-*, posem per cas; acord 49 del 30 de març de 2008). La futura autoritat normativa de la llengua occitana faria bé de tenir en compte algunes d'aquestes decisions, ja que potser podria aprofitar alguns dels treballs de la Comissió. Fins ara quasi no s'havia estudiat quin podia ser, posem per cas, l'equivalent del català *ramader* en l'occità estàndard general, i no hi havia hagut gaires oportunitats de discutir certs neologismes o certs termes especialitzats que ha estudiat la Comissió.

Algunes d'aquestes decisions, si no són del tot conscientment normatives, expressen almenys una certa inestabilitat de la norma actual que es manifesta en un posicionament majoritari dels membres de la Comissió a favor de solucions que fins ara no han estat etiquetades de normatives. El traductor produeix *arab* i no pas *arabi*, tot i que en el document del CLO publicat un mes abans de l'acord de la Comissió sobre aquestes formes (acord 80 del 3 de febrer de 2008), es donava prioritat a la forma *arabi*. No s'hi acabava de condemnar *arab* (que figurava en el text de l'acord que aquesta mateixa institució havia publicat uns anys abans sobre la famosa i polèmica *-e de sosten*), però la Comissió va decidir que tingués prioritat *arab* per davant de *arabi*, que en tot cas és una forma que pot ser utilitzada per qualsevol usuari amb la seguretat que serà traduïda al català. En casos de sinonímia, la preferència per una forma determinada no pot ser vista sistemàticament com el resultat d'una decisió normativitzadora. No només perquè certes decisions persegueixen únicament la funcionalitat, sinó perquè l'estandardització d'una llengua no implica que sempre hi hagi d'haver una única solució jeràrquicament superior a les altres. Alibèrt, per exemple, recollia les formes *comunitat* i *comunautat*, però el traductor automàtic, que únicament pot generar una forma occitana, produeix *comunitat* i reconeix *comunautat* perquè calia donar prioritat a un dels dos mots més que no pas perquè hi hagués una preferència raonada majoritària o unànime per una de les dues formes.

El resultat actual del traductor automàtic potser té certes limitacions, però la Comissió Lingüística va poder arribar a preveure fins i tot alguns petits detalls formals (per exemple, la manera d'enunciar les adreces postals; acord 65 del 2 de febrer de 2008), i fins va proposar d'incloure en el traductor una nombrosa representació de topònims catalans, occitans i internacionals en la seua forma occitana (que, per ser justos, cal dir que aprofita el treball que un membre de la Comissió ja havia dut a terme en el marc del CLO). Però cal remarcar sobretot l'abundant treball de comparació sintàctica entre el català i l'occità moderns duta a terme per la Comissió, amb resultats realment notables en alguns casos, i amb reflexions interessants sobre el procés de convergència que haurien de seguir els estàndards català i occità, contrari a la separació formal progressiva entre les dues llengües imposada per l'espanyol i pel francès. La recerca lingüística de la Comissió ha estat especialment laboriosa en l'àmbit de les locucions, molt més abundant que la que la direcció científica preveia inicialment, tot i que els resultats —com era d'esperar— encara són lluny de l'exhaustivitat.

### Previsions i advertències. Recapitulació

És impossible de preveure totes les equivalències entre dues llengües, i això fa que qualsevol traductor automàtic sigui per força imperfecte. D'altra banda, un traductor automàtic pot tenir en compte les freqüències de determinats elements o seqüències, fins i tot la seua condició sintàctica, però difícilment pot preveure fets estrictament semàntics. Això són limitacions generals de qualsevol traductor automàtic que no hem descobert nosaltres. Remarquem, per tant, alguns fets que



condicionen el resultat definitiu del traductor automàtic que sí que es refereixen exclusivament a una eina d'aquesta mena que tingui en compte la llengua occitana, i resumim al mateix temps alguns dels fets que hem assenyalat fins ara que cal tenir imperativament en compte abans de fer servir aquest traductor automàtic amb presència occitana:

a) En la llengua de destinació un traductor automàtic ha de produir un equivalent concret d'una forma o d'una estructura d'origen. La diversitat dialectal de l'occità i la situació actual del seu procés de codificació i d'estandardització fan que, a diferència del que pot passar en d'altres llengües, en aquesta hi hagi d'una manera habitual diverses formes i fins i tot diverses estructures que poden ser vistes com a estàndards. A això hi podem afegir l'existència d'estàndards regionals amb solucions específiques, o el fet que allò que és majoritari en el dialecte llenguadocià —que serveix de base per a l'occità estàndard general— no és a vegades majoritari en altres zones del domini.

b) La necessitat de seleccionar una única forma occitana en situacions en què cohabitaven diversos geosinònims amb una presència territorial similar, ha provocat que certes decisions de la Comissió Lingüística hagin estat preses en funció del grau d'operativitat que tenia una solució determinada en el marc del traductor automàtic. Hi ha seleccions de la Comissió que, més que ser una conseqüència de l'aplicació d'uns criteris de codificació lingüística (diasistematicitat de les solucions, historicitat, funcionalitat, autonomia, regularitat, coherència dels resultats amb els d'altres llengües...), es basen en la freqüència d'una determinada solució en el corpus o bé en la voluntat d'evitar ambigüitats o problemes d'interpretació per part del traductor automàtic. Això significa que hi ha acords de la Comissió Lingüística que poden ser catalogats de normativitzadors, ja que ben bé podrien pertànyer a una autoritat normativa des d'un punt de vista lingüístic, però que també n'hi ha d'altres en què no hi ha una prescripció, sinó una simple decisió presa en un context en què s'ha pensat que era prioritari que els usuaris es reconeguessin en la llengua produïda pel traductor automàtic.

c) La codificació de la llengua occitana encara és parcial, i la codificació del lèxic és especialment lluny de la normalitat. Això implica que els resultats d'un traductor automàtic que tingui en compte l'occità seran més limitats que si s'hagués tingut en compte llengües del tot consolidades, o fins i tot llengües minoritzades com el català, que es troba en un estadi de codificació més avançat que l'occità. Implica també que el procés d'introducció de dades en el traductor ha estat més feixuc i complicat que en varietats lingüístiques ben codificades, ja que ha calgut dur a terme una feina suplementària de codificació o de paracodificació. Per tant, ha estat necessari de superar tot tipus de problemes que no existirien en d'altres llengües: punts ortogràfics conflictius en què els lingüistes encara no han arribat a un acord, desavinences i contradiccions entre certes obres de referència, manca d'unanimitat en l'acceptabilitat de certes estructures o formes... La inexistència d'una codificació sòlida emparada per una autoritat normativa acceptada per tothom fa que la tasca de crear un traductor automàtic que tingui en compte la llengua occitana exigeixi més temps, més recerca i més documentació que si el traductor s'encarregués d'una varietat lingüística socialment normal.

d) Com a conseqüència d'aquesta estandardització parcial de la llengua occitana a què hem fet referència, potser hi haurà certs aspectes lingüístics de tipus estructural que la Comissió Lingüística no haurà estudiat, o que en tot cas haurà estudiat només parcialment. De fet, els continguts occitans d'aquest traductor automàtic necessitarien seguir un procés d'adaptació i revisió permanent, paral·lel als avanços lingüístics en el procés d'emergència i de renaixement de la llengua occitana. Una de les possibles ampliacions d'aquest traductor automàtic podria consistir, d'una altra banda, a incloure les formes i estructures més destacades i simbòliques de cada bloc dialectal occità, per tal d'augmentar el nombre d'usuaris que es poguessin reconèixer en el traductor. La Comissió ha fet en aquest àmbit una feina parcial.

e) En alguns casos la Comissió Lingüística ha comprovat que és difícil que es tradueixin adequadament certes estructures sintàctiques. És difícil que un sistema de traducció automàtica pugui preveure adequadament certs fets. En cas que hi arribi, aquesta previsió pot ser que no sigui rendible i que la seua presència provoqui més errors i més hipercorreccions que la seua absència. La Comissió Lingüística ha fet constar en les actes de les seues sessions alguns d'aquests fets sintàctics problemàtics: demostratius de represa i posició dels adjectius en el sintagma nominal (acord 64 del 2 de febrer de 2008); repetició de preposicions quan hi ha dos substantius o sintagmes nominals coordinats (acord 59 del 2 de febrer de 2008); ordre dels elements de la frase (acord 90 del 3 de febrer). Cal remarcar que en aquestes qüestions no cal veure-hi dèficits d'aquest traductor automàtic, sinó obstacles més aviat tècnics que per ara semblen difícils de superar en el marc de la traducció automàtica en general o en la traducció automàtica entre català i occità.

Pot semblar que tot el que acabem de dir minimitza els mèrits de la Comissió Lingüística, però entenem que el balanç de la seua actuació des del 4 de maig de 2007 fins al 30 de març de 2008 ha de ser per força positiu, tant des d'un punt de vista quantitatiu com des d'un punt de vista qualitatiu. La qualitat era previsible tenint en compte l'alta competència lingüística dels seus membres, però la quantitat queda lluny de les previsions inicials. Però l'existència d'aquesta comissió no només ha permès de dur a terme un volum important de treball, sinó que ha demostrat que és possible de crear un espai en el qual els lingüistes que realitzen investigacions sobre la llengua occitana puguin treballar conjuntament, independentment de les seues afinitats disciplinàries i personals, i de les seues variades opinions —previsibles en una situació de minorització lingüística— sobre alguns aspectes normatius. L'existència d'aquesta comissió hauria de servir com a estímul per a la creació definitiva d'una autoritat lingüística occitana, i per fer veure que aquest objectiu no és difícil si al darrere hi ha el suport imprescindible de les institucions públiques. Cal avançar amb celeritat en la codificació de l'occità. És imprescindible perquè aquesta llengua es pugui fixar objectius socials ambiciosos. Perquè reconquereixi els àmbits d'ús que ha perdut en el darrer segle.

