



## La localització de l'apli de missatgeria Telegram al sard: l'experiència de Sardware i una aplicació docent

Adrià Martín-Mor  
Grup Tradumàtica, UAB

### RESUM

Aquest article descriu l'experiència de localització col·laborativa del client de missatgeria Telegram a la llengua sarda. Atès que Telegram és un programa lliure que permet a l'usuari modificar i aplicar paquets de llengua externs, s'ha seguit l'exemple d'altres comunitats de parlants de llengües minoritzades com la catalana, de qui s'ha rebut assessorament. Després d'una introducció a la situació de la llengua sarda i d'una descripció de la configuració del projecte, s'explica el procés de localització amb una atenció especial a la qüestió lingüística. L'experiència també es presenta en aquest article des de la perspectiva de la docència de les tecnologies de la traducció.

**Paraules clau:** localització, aplicacions per a dispositius mòbils, llengua sarda, Telegram, llengües minoritzades, traducció col·laborativa

### ABSTRACT

This article describes a collaborative project localising the Telegram messaging client into Sardinian. Since Telegram is an open programme which allows users to make modifications and add external language packs, we followed the example of other minority language communities –as in the case of Catalan–, from which we received guidance. The article gives an introduction to the current situation of Sardinian, a description of the configuration of the project and explains the localising process with particular attention to the linguistic issues. This experience is also presented from the perspective of teaching translation technologies.

**Keywords:** localisation, apps for mobile devices, Sardinian language, Telegram, minoritised languages, crowdsourcing

### RESUMEN

Este artículo describe la experiencia de localización colaborativa del cliente de mensajería Telegram a la lengua sarda. Telegram es un programa de software libre que permite al usuario modificar y aplicar paquetes de lengua externos, lo cual ha hecho posible seguir el ejemplo de otras comunidades de hablantes de lenguas minorizadas como la catalana, de la que se ha recibido asesoramiento. Después de una introducción a la situación de la lengua sarda y de una descripción de la configuración del proyecto, se explica el proceso de localización con un especial hincapié en la cuestión lingüística. La experiencia también se presenta en este artículo desde la perspectiva de la docencia de las tecnologías de la traducción.

**Palabras clave:** localización, aplicaciones para dispositivos móviles, lengua sarda, Telegram, lenguas minorizadas, traducción colaborativa



## 1. Introducció

El sard és una llengua minoritzada en procés d'estandardització per a la qual no existeixen encara recursos bàsics com ara un diccionari normatiu o una gramàtica normativa. Aquest fet, però, no impedeix que grups d'usuaris s'auto-organitzin per a treballar en la localització de programari i en el desenvolupament de tecnologies per a la traducció. Aquest 2016, per exemple, s'han presentat la localització de Facebook al sard (Beccu i Martín-Mor, 2016) i el traductor automàtic Apertium ([www.apertium.org](http://www.apertium.org)) en la combinació italià-sard. Malgrat tot, la llengua sarda encara no té cap client de missatgeria per a mòbils intel·ligents. Per aquest motiu, des del grup d'usuaris de tecnologia en llengua sarda Sardware vam decidir començar la localització de les diverses aplis de Telegram al sard, seguint l'exemple i les recomanacions de Softcatalà (Mas i Montané, 2017).

Pel fet que es tracta de programari lliure, Telegram permet l'accés i la modificació del codi font. Aplicat al cas que presentem, això implica la possibilitat de localitzar els fitxers de llengua originals, i que qualsevol usuari pugui aplicar a la seva instal·lació de Telegram una localització personalitzada.

Com sol passar, doncs, el programari lliure representa una oportunitat per a les llengües minoritzades per poder fer-se un lloc en l'univers digital, condició indispensable per a la supervivència de les llengües minoritzades al segle XXI (Proyecto Trasno<sup>1</sup>, Softastur<sup>2</sup>).

## 2. La llengua sarda

La llengua sarda és una llengua romànica parlada en la major part de l'illa de Sardenya per més d'un milió de persones. Segons l'Atlas of the World's Languages in Danger de la Unesco, té aproximadament un milió tres-cents mil parlants,<sup>3</sup> mentre que segons l'Ethnologue en té menys d'un milió dos-cents mil.<sup>4</sup> Tal com es pot veure a la imatge següent, a l'illa cohabitaven amb el sard quatre llengües més (a banda de l'italià, l'única oficial): el lígur tabarquí (a les illes de Sant Pere i de Sant Antíoc), el català alguerès (al terme municipal de l'Alguer), el sassarès (al nord-oest de l'illa, sovint classificat com a parlar corso-toscà) i el gal·lurès (al nord-est de l'illa, considerat majoritàriament un dialecte cors o un parlar de transició entre el cors i el sard).<sup>5</sup>

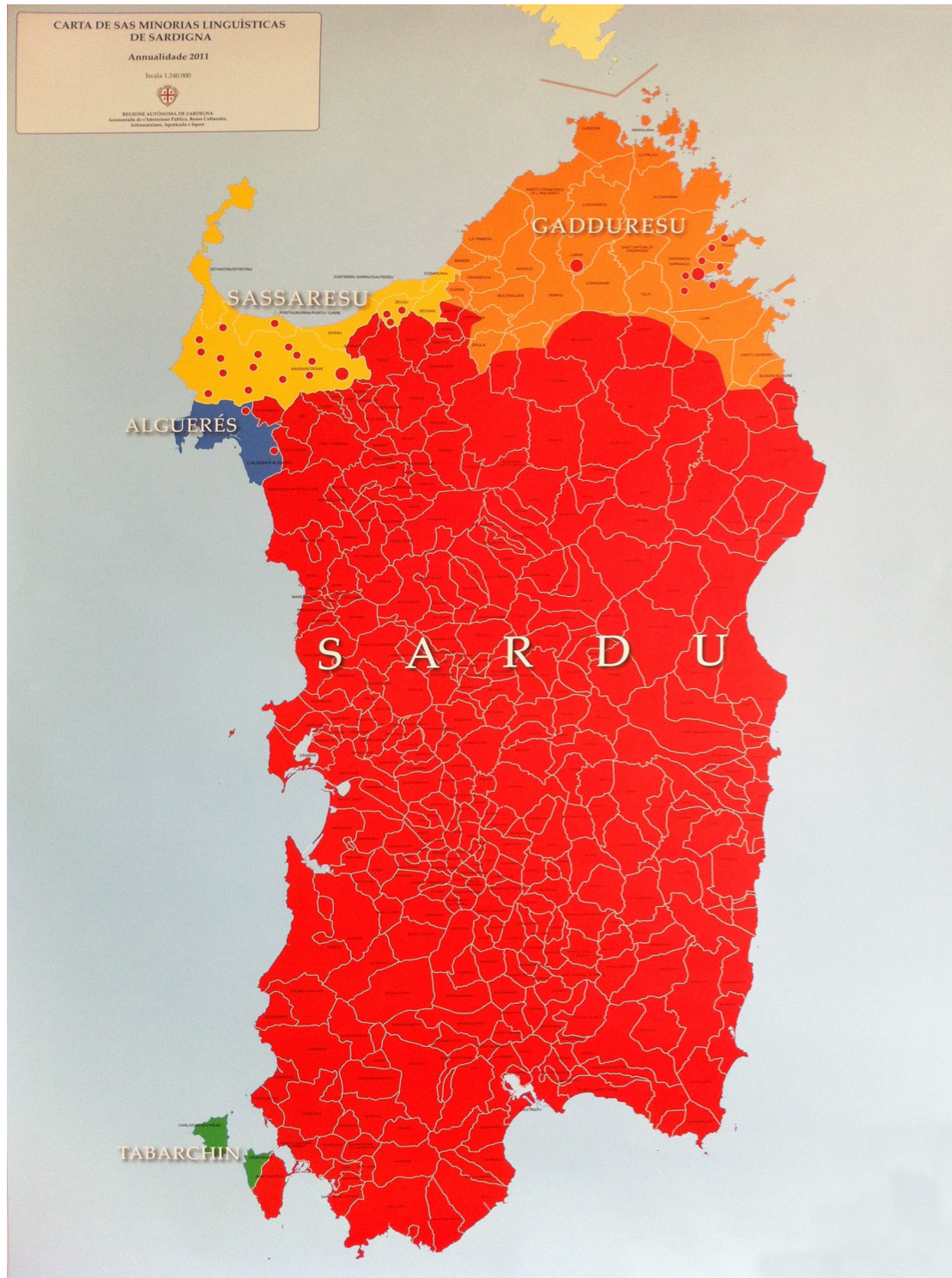
<sup>1</sup>[www.trasno.gal](http://www.trasno.gal).

<sup>2</sup>[www.softastur.org](http://www.softastur.org).

<sup>3</sup>Resultat de sumar els parlants del campidanès (900.000, <http://www.unesco.org/culture/languages-atlas/en/atlasmap/language-id-337.html>) i els del logudorès (400.000, <http://www.unesco.org/culture/languages-atlas/en/atlasmap/language-id-381.html>).

<sup>4</sup><https://www.ethnologue.com/language/srd>.

<sup>5</sup>A tall anecdòtic, al poble d'Isili (Càller) es conserva una llengua pràcticament extingida anomenada romanisku o arromanisca (de *rom*, 'gitano'). També formen part del patrimoni de l'illa les llengües que arriben durant el segle XX amb les migracions de les poblacions de la península (del Vènet i la Romania) i de l'Ístria, algunes d'elles establertes en zones del terme municipal de l'Alguer.



Imatge 1. Les llengües de Sardenya. Mapa editat pel govern sard

El sard, que ocupa la major part del territori, té una varietat interna molt marcada que a grans trets es pot classificar entre variants septentrionals (logudorès i nuorès) i meridionals (campidanès). Paradoxalment, l'immens patrimoni lingüístic de l'illa és també un dels factors



que alimenten les discussions sobre l'estandardització de la llengua. És simptomàtic, en aquest sentit, que els recursos citats anteriorment barregin els conceptes de *llengua* i de *dialecte*. L'Atlas de la Unesco, per exemple, no conté cap entrada anomenada *Sardinian*, sinó que conté l'entrada *Campidanese* i l'entrada *Logudorese*. L'Ethnologue, en canvi, sí que conté una entrada anomenada *Sardinian*, però hi afegeix el següent aclariment: «A macrolanguage [...]. Includes Campidanese Sardinian [sro], Gallurese Sardinian [sdn], Logudorese Sardinian [src], Sassarese Sardinian [sdc].»

Malgrat que els primers documents escrits en aquesta llengua es remunten als segles XIII i XIV (com ara els estatuts de la República de Sàsser o la Carta de Logu), actualment hi ha corrents que defensen vies diferents per a l'estandardització del sard. D'una banda, hi ha els qui defensen una doble norma per a l'escriptura. Aquest corrent, que segueix models per a la doble norma com ara el del noruec, té seguidors bàsicament al sud de l'illa, i el govern provincial de Càller hi ha donat suport per mitjà de la redacció de les *Arrègulas po ortografia, fonètica, morfologia e fueddàriu de sa Norma Campidanese de sa Lingua Sarda*. Aquest document és una proposta de norma per a la variant campidanese del sard i deixa clar que:

[s]a lingua sarda est *una lingua* a duas (macro) bariedades diatòpicas (diasistemas sintzillus etotu) e literàrias chi tenint – e depint tenni – sa matessi dinnidadi: su Campidanese e su Logudoresu. Duncas de imoi innantis espressadas che *lingua sarda* o *Sardu* [sic] inditant is duas bariedades o chistionis chi ddas pertocant. Intrendi in sa chistioni scientifica de custu traballu, de s'ora chi sa *lingua sarda est una, a duas bariedades, eus a fueddai de unu standard* (su de sa lingua sarda) *a duas normas* (sa norma de su Campidanese e sa norma de su Logudoresu). (Comitau Scientificu po sa Norma Campidanese de su Sardu Standard, 2009: 18).

Al fragment anterior, doncs, es poden observar elements propis de les variants meridionals, com ara l'article plural invariable *is* (en les variants septentrionals, *sos* per al masculí i *sas* per al femení), els substantius acabats en *-i* (*chistoni/-is* i no pas *chistione/-es*), els infinitius en *-i* (*tenni* en comptes de *tennere*), els pronoms de complement directe *ddus* i *ddas* (*lus* i *las*) i lèxic (*lingua/limba*; *imoi/como*).

L'altre corrent principal, el de la Limba Sarda Comuna (LSC), proposa un sol sistema d'escriptura per a totes les variants del sard. La LSC ha estat adoptada pel govern autònom sard (*regione*) i per algunes editorials (les editorials Papiros i Condaghes publiquen literatura original i traduïda en LSC), com també per revistes («Eja», «Làcanas») i blogs d'internet (Sa Gazeta, Sa limba sarda).

La coexistència d'aquestes dues propostes (i d'altres de més minoritàries) fa que qualsevol iniciativa que impliqui la llengua sarda exigeixi una reflexió prèvia al voltant dels models existents i una decisió conseqüent. Una solució digna de menció, pel que implica quant a autogestió, és la Viquipèdia sarda. En aquesta plataforma, els usuaris poden deixar constància de la variant en la qual han escrit un article per mitjà d'un requadre amb el nom de la variant i un logotip.



## Limba Sarda Comuna



Custu articulu est galu unu abbotzu. Lu podes modificare e fàghere mannu e pretzisu. Agiuda-si!

Sa **Limba Sarda Comuna (LSC)** est una norma de iscritura de sa **limba sarda**, pensada cun sa punna de trascrivere sas variantes faeddadas de sa limba sarda e adotada cun caràtere isperimentalmente in su 2006 dae sa **Regione Autònoma de Sardigna** pro s'iscritura de sos documentos ufitziales in essida, in co-ufitzialidade cun s'**italianu**.



Articulu in **LSC**



*Imatge 2. Entrada de la Viquipèdia sarda corresponent a la Limba Sarda Comuna. El missatge de la dreta indica que l'article està escrit en LSC*

Això, tanmateix, pot generar confusió en articles escrits en més d'una variant:

## Limba sarda

Su **sardu** est una de sas **limbas romanzas**, faeddada mescamente in su logu de orìgine, sa **Sardigna**, e dae sos disterrados.



Articulu in **logudoresu**



### Cuntenutu [cua]

1 **Datos generales**

2 **Istòria**



Articulu in **campidanesu**



*Imatge 3. L'entrada «Limba sarda» de la Viquipèdia, escrita en logudorès i en campidanès*

Val a dir que, tal com afirma Cheratzu (Mura i Viridis, 2015: 10), numèricament, la producció textual i literària és superior en LSC:

Sa cantidade de produzione literària de custas duas normas sunt però diferentes meda, cun unu raportu de nessi unu a chimbe, in favore de sa LSC. [...] A custu puntu su sèberu de impreare sa LSC fiat in pràtica obligadu.

### 3. La preparació del projecte

Telegram és un client lliure i multiplataforma de missatgeria instantània. Està disponible oficialment en vuit llengües (alemany, anglès, àrab, castellà, coreà, italià, neerlandès i portuguès) per a la majoria de plataformes. Amb tot, el programa permet l'aplicació de paquets de llengua externs per part dels usuaris. De fet, els desenvolupadors encoratgen les comunitats lingüístiques a localitzar el codi del programa per mitjà de la plataforma Transifex,<sup>6</sup> gratuïta per a projectes lliures.

Fins a la primavera del 2016, qualsevol usuari podia visitar la pàgina de Telegram al web de Transifex i descarregar-ne els fitxers de llengua originals. A partir d'aleshores, l'accés als fitxers de llengua va quedar restringit als usuaris registrats que formen part d'una comunitat de traductors. Avui, doncs, els usuaris que volen tenir accés als fitxers de llengua s'han de registrar a Transifex i sol·licitar l'accés a un dels grups de traducció. Els autors d'aquest

<sup>6</sup>Vegeu les preguntes més freqüents del lloc web de Telegram: <https://telegram.org/faq#q-can-i-trans-late-telegram>.



article, tal com han fet molts altres usuaris, han sol·licitat reiteradament l'accés a aquests grups sense rebre cap resposta.

Amb tot, el projecte va poder superar aquest contratemps gràcies a l'alternativa suggerida per Softcatalà. El robot de Telegram @TGlanguagesbot<sup>7</sup> descarrega periòdicament els paquets de llengua allotjats a Transifex en totes les llengües i els serveix als usuaris que els sol·liciten mitjançant el mateix Telegram. @tgstoreversions,<sup>8</sup> en canvi, és un canal de Telegram que informa els usuaris de quan es publica una nova versió dels clients de Telegram per als diversos sistemes operatius (Android, iOS, Windows Phone, PC, web i OS X). Combinant els dos sistemes, es pot suplir la impossibilitat d'accedir als fitxers de llengua de Telegram per mitjà de la plataforma oficial, Transifex, atès que cada cop que surt una nova versió els administradors de la localització al sard en som informats mitjançant el canal @tgstoreversions i podem descarregar-ne els fitxers amb el robot @TGlanguagesbot. Cal afegir que hi ha versions de Telegram per a altres sistemes, com ara iWatch i Ubuntu Touch (anomenat també Ubuntu Phone o uPhone). Aquestes versions, però, no es distribueixen amb el robot esmentat anteriorment sinó mitjançant altres sistemes, com per exemple, en el cas d'Ubuntu, per mitjà de la plataforma Launchpad.<sup>9</sup>

Malgrat tot, el fet de no poder formar part dels grups de localització oficials de Telegram a Transifex també implica que cal buscar un altre sistema per a la gestió i l'edició de traduccions. En primer lloc, pel fet que es tracta d'un projecte de localització col·laborativa, és imprescindible que els fitxers no siguin emmagatzemats en local i intercanviats entre els usuaris, atès el perill de profusió innecessària de fitxers i la necessitat de control de versions que això implica. Cal, doncs, un sistema col·laboratiu mitjançant el qual es puguin emmagatzemar els fitxers en un servidor i puguin ser localitzats directament al núvol.

A tal efecte, els autors d'aquest article es van decantar pel sistema de gestió i d'edició de traduccions (SGET) al núvol MateCat ([www.matecat.com](http://www.matecat.com)), programari amb llicència lliure resultat d'un projecte de recerca amb finançament de la Unió Europea. MateCat permet la creació de projectes i la traducció en línia fins i tot en modalitat anònima (sense registre), i ofereix la possibilitat d'accedir amb compte d'usuari per tal de dur a terme la gestió, per rebre notificacions de l'activitat d'altres usuaris, etc. Entre les especificitats del programa, cal fer ressaltar que funciona únicament amb el navegador Chromium o amb la seva versió privativa Chrome, i també el fet que s'ofereix en dues modalitats: com a servei web i com a programari instal·lable en un servidor propi (<http://www.matecat.com/open-source/>). En el cas del servei web, la disponibilitat de les llengües per a la creació dels projectes de traducció depèn dels administradors del servidor. Així, algunes llengües com el sard encara no estan disponibles a MateCat. Per a l'objectiu d'aquest projecte, doncs, es va crear un projecte amb una configuració de llengües que no es corresponia amb la real.<sup>10</sup>

<sup>7</sup><https://telegram.me/TGlanguagesbot>.

<sup>8</sup><https://telegram.me/tgstoreversions>.

<sup>9</sup> <https://translations.launchpad.net/telegram-app/v2/>.

<sup>10</sup>En el nostre cas, vam seleccionar occità com a llengua d'arribada del projecte, pel fet que, essent una llengua minoritzada com el sard, reduïem el risc que MateCat oferís coincidències provinents dels seus recursos propis (memòries de traducció, traductors automàtics). Paral·lelament, ens vam posar en contacte amb MateCat per demanar-los que afegissin el sard entre les llengües disponibles. En la configuració del projecte, vam desactivar l'enviament de les traduccions a la memòria de traducció de MateCat, per tal que la seva memòria anglès-occità no es veiés alterada per la presència de segments en sard.



MateCat lets you re-use repeated content better.  
It leverages MyMemory, Professional Machine Translation and internal fuzzy matches to make you save money and have a faster turnaround. [Find out more.](#)  
(\*\*Weighted words is how the industry counts words after discounting repetitions and matches against a desktop translation memory, like any other CAT tool would.)

Analysis complete Download Analysis Report

English US > Occitan Split in 2 Jobs Split

	Payable	Total	New	Repetition	Internal Matches (75%-99%)	TM Partial (50%-74%)	TM Partial (75%-84%)	TM Partial (85%-94%)	TM Partial (95%-99%)	TM 100%	Public TM 100%	TM 100% in context	Machine Translation
<b>Payable Rate Breakdown</b>			100%	30%	60%	100%	60%	60%	60%	30%	30%	0%	85%
564264 File details (4)	6,911	18,511	267	3,340	0	1,026	694	612	111	0	12,459	2	0
└ languages_{...}_en.strings	2,897	7,314	187	2,211	0	423	384	335	111	0	3,661	2	0
└ languages_tdes{...}_jop_en.strings	1,203	4,011	0	474	0	0	0	0	0	0	3,537	0	0
└ for_use_telegram_{...}_jngxml-48_en.xml	2,566	6,368	80	652	0	603	310	277	0	0	4,446	0	0
└ telegram_po_com_{...}_juntu.telegram.po	245	818	0	3	0	0	0	0	0	0	815	0	0

Imatge 4. Interfície de MateCat amb els projectes de traducció de Telegram

Com es pot veure a la imatge 4, el projecte de localització inclou quatre versions de Telegram: per a Android, aplicació d'escriptori, iOS i Ubuntu Phone. Es va descartar la localització de les altres versions pel fet que no permeten l'aplicació de paquets de llengua externs. Les últimes versions traduïdes (mostrades a la imatge) contenen 6.368 (Android), 4.011 (escriptori), 7.314 (iOS) i 818 paraules (Ubuntu Phone), que sumen al voltant de 18.500 paraules (en taronja a la imatge). El projecte tenia un 30 % de repeticions internes (3.340 paraules), la qual cosa feia imprescindible l'ús d'un SGET.<sup>11</sup> El projecte de traducció es va crear al juny del 2016.

#### 4. Una aplicació docent

Com s'ha dit anteriorment, el programari lliure representa una oportunitat d'accés a la tecnologia per a les llengües minoritzades, atès que permet l'accés al codi font i la seva modificació. Alhora, el programari lliure té un alt potencial formatiu, ja que és una gran font de recursos docents per a la traducció (Diaz Fouces, 2012: 163; Sandrini i García González, 2015). En el cas de Telegram, els autors d'aquest article l'han utilitzat en diversos cursos sobre tecnologies de la traducció, pels avantatges que suposa i que enumerem a continuació.

- Telegram, a més de ser programari lliure, és gratuït i té una comunitat d'usuaris força nombrosa. Això fa que els estudiants que encara no la tenen als seus dispositius se la puguin descarregar sense cap cost.
- És multiplataforma. Gràcies al fet que té versions per a la majoria de sistemes operatius (mòbils i de sobretaula), ofereix l'oportunitat d'analitzar tota una panoràmica de llenguatges de programació i de centrar-se en les especificitats de cada plataforma (XML per a Android, po per a Ubuntu, strings per a iOS, resx per a Windows, etc.).
- Permet accedir als fitxers de llengua originals des de l'interior mateix de la plataforma per mitjà de robots (vegeu l'apartat 3. La preparació del projecte).
- Permet l'aplicació de paquets de llengua externs per part de l'usuari sense cap restricció (tants paquets com es vulguin, sense prèvia aprovació per part d'administradors) ni de codificació de caràcters (en qualsevol alfabet).

<sup>11</sup>La cel·la de color verd, en canvi, mostra el nombre de paraules ponderades, és a dir, després de restar al total (a més de les repeticions) les coincidències parcials. La captura de pantalla està feta amb el projecte pràcticament avançat.



- Els fitxers de llengua contenen totes les característiques d'un projecte de localització: variables (com ara %1\$@), etiquetes (com ara <b> i </b>) i codificacions per a caràcters (\' per a l'apòstrof; \n per als salts de línia).

Tots aquests factors fan de la localització de Telegram un projecte docent engrescador i real amb el qual es pot dur a terme un cicle de localització complet, des de la descàrrega dels fitxers originals fins a l'aplicació de la traducció per part de l'usuari. El fet que els estudiants puguin veure el resultat de la traducció en pantalla al moment facilita la implicació amb el projecte i permet alhora introduir els alumnes a la tasca de verificació tècnica, és a dir, el procés de revisió del producte des del punt de vista lingüístic i funcional (Martín-Mor, Piqué i Sánchez-Gijón, 2016).

## 5. La localització al sard

Durant el procés de localització de Telegram a la llengua sarda, l'equip de traductors (que inclou nadius i no nadius)<sup>12</sup> ha hagut de prendre decisions lingüístiques, com en qualsevol projecte de localització. Tanmateix, la manca de recursos lingüístics per al sard, derivada del fet que és una llengua en procés d'estandardització, ha suposat un repte, especialment en l'àmbit de la terminologia. A grans trets, l'equip s'ha servit del Curretore regionale ortogràficu sardu (Cros, amb llicència lliure),<sup>13</sup> del document *Propostas isperimentales de unos cantos liniamentos grammaticales comunes de sa limba sarda* (conegut com a *liniamentos*)<sup>14</sup> i de consultes a parlants nadius i a experts (traductors, editors, etc.). Entre altres recursos consultats, hi ha també les traduccions d'Ubuntu al sard.<sup>15</sup> A continuació s'exposen algunes de les decisions preses.

Anglès (original)	Català (traducció de Softcatalà)	Sard
app	aplicació/app <sup>16</sup>	aplicatzione
broadcast list	llista de difusió	lista de difusione
forward	reenviar	torrare a imbiare
OK	D'acord	Andat bene
chat	xat	tzarrada
clipboard	porta-retalls	punta de billete
FAQ [Frequently Asked Questions]	PMF [preguntes més freqüents]	PF [pregontas fitianas]
sticker	adhesiu	adesivu

Taula 1. Terminologia de Telegram

<sup>12</sup>Els membres que són nadius s'han dedicat a la traducció i a la revisió, mentre que els no nadius s'han dedicat a la traducció (seguida de revisió per part de nadius) i a la gestió.

<sup>13</sup><http://www.sardegnaicultura.it/cds/cros-lsc/>.

<sup>14</sup>[http://www.sardegnaicultura.it/documenti/7\\_108\\_20090205130352.pdf](http://www.sardegnaicultura.it/documenti/7_108_20090205130352.pdf).

<sup>15</sup><https://wiki.ubuntu.com/Ubuntu-Sardu/>.

<sup>16</sup>Quant a la traducció de *app*, el Termcat ([http://www.termcat.cat/ca/Comentaris\\_Terminologics/Consultes\\_Terminologiques/183](http://www.termcat.cat/ca/Comentaris_Terminologics/Consultes_Terminologiques/183)) explica que admet tant *app* com *apli* al costat de *aplicació mòbil*.





Durant el projecte, també s'han tingut en compte obres de referència per a altres llengües —principalment el català—, com ara el cercador Termcat<sup>17</sup> i les directrius lingüístiques per a la localització de programari de Softcatalà (Softcatalà, 2010). En concret, es van tenir en compte les directrius següents.

### 5.1 Expressions innecessàries

Tal com s'afirma a les directrius de Softcatalà (2010: 11),

Els programadors en llengua anglesa tenen tendència a «humanitzar» el programari i la forma en què l'ordinador s'adreça a l'usuari. Això fa que, per exemple, sovint es disculpin quan, per alguna raó, no s'ha pogut realitzar alguna ordre sol·licitada, o bé que demanin a l'usuari que «si us plau» triï una opció. Aquestes expressions no les traduirem, ja que la manera catalana d'expressar el mateix concepte («em sap greu», «malauradament», etc.) ocupa més espai i no aporta cap mena d'informació addicional a la frase.

D'aquesta manera, tal com es pot veure a la taula següent, termes com ara *sorry* o *please* no han estat traduïts al sard.

Original	Traducció
Sorry. You can't access this channel because it violates App Store rules.	Non podes intrare a custu canale ca non respetat sas règulas de s'App Store.
Please help.	Agiuda-mi.

Taula 2. Expressions innecessàries

### 5.2 Veu activa versus veu passiva

Les directrius de Softcatalà (2010: 10) apunten que:

[S]'ha de vigilar amb les oracions passives angleses, ja que el català no en fa un ús tan habitual com l'anglès. Per tant, sempre intentarem convertir les frases a la forma activa. [...] Tingueu present, ahora, que habitualment cal modificar l'ordre dels components de la frase respecte de la llengua de partida de la traducció.

Així doncs, segments com per exemple «A new invite link has just been generated» s'han traduït al sard per «Generadu unu ligàmene de invitu nou», evitant, tal com recomanava Softcatalà, la veu passiva i reordenant els elements de la frase.

### 5.3 Expressions sintètiques

La guia d'estil de Softcatalà (2010: 14) afirma que:

Molts dels missatges en què l'anglès emprava expressions sintètiques (tant nominals com verbals) en català és preferible redactar-los de forma completa utilitzant oracions amb un verb conjugat. Aquest és el cas del gerundi i també de

<sup>17</sup>[www.termcat.cat](http://www.termcat.cat).



[...] l'infinitiu, [...] el substantiu, [...] el substantiu + adjectiu, [...] el substantiu + participi.

Per aquest motiu, el grup de traductors ha dut a terme les traduccions corresponents per als següents segments.

	Original	Traducció
gerundi	Checking username	Semus averguende su nùmene utente...
infinitiu	Terminate this session?	Nche cheres serrare custa sessione?
substantiu	No photos Error downloading image	Non b'at fotografias B'at àpidu una faddina in s'iscarrigamentu de s'imàgine
substantiu+participi	Two-Step verification enabled	Sa verifica de duas fases est ativa

Taula 3. Expressions sintètiques

#### 5.4 Temps verbal

En general, mantindrem el temps verbal de l'original, excepte en el cas del «past simple» anglès, que gairebé sempre caldrà traduir pel pretèrit indefinit o pel present d'indicatiu. Un exemple freqüent el trobem en l'ús de «could/couldn't» o «was able/unable to», que traduirem per «s'ha pogut/no s'ha pogut». Softcatalà (2010: 19).

Segments com ara «Unable to change privacy settings now, please wait» s'han traduït al sard com «Como non faghet a mudare sas cunfiguratziones de sa riservadesa, iseta». Malgrat que la directriu citada més amunt fa referència específicament al temps verbal, en aquest cas s'ha tingut en compte per a la traducció de l'estructura «unable to».

#### 5.5 Altres

Telegram fa servir abreviatures per a elements com els dies de la setmana i els mesos. En sard, a diferència de la majoria de llengües romàniques, hi ha mesos compostos per més d'una paraula:

Octubre: Santugaine

Novembre: Santandria

Tal com es pot veure a la imatge següent, aquests dos mesos s'han hagut d'abreviar formant un acrònim amb les inicials de cada paraula:



Aug	Aus
Sep	Cab
Oct	StGa
Nov	StAnd
Dec	Nad

Imatge 5. Abreviatures per als mesos de l'any

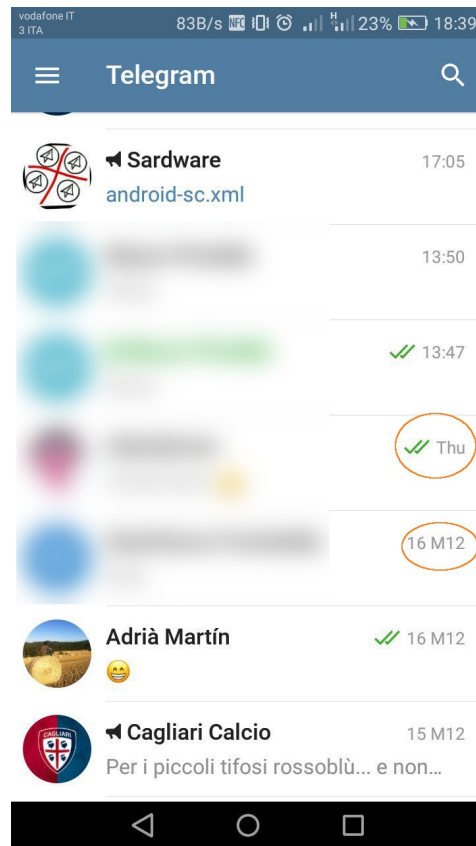
Creiem que aquesta proposta d'abreviatura permet identificar ràpidament el mes i no deixa lloc a ambigüitats. En el cas de mesos molt curts, com ara *maju*, es va decidir mantenir la forma plena, atès que l'estalvi de caràcters amb l'abreviatura era molt menor.

## 6. Control de qualitat, verificació tècnica i preparació de recursos

Un cop enllestida la traducció dels fitxers de llengua, i abans de la difusió pública dels paquets, es van dur a terme una sèrie de tasques associades a les fases esmentades al títol de l'apartat: control de qualitat, verificació tècnica i preparació de recursos.

En primer lloc, es van exportar les memòries dels quatre projectes de MateCat a TMX. Un cop descarregades, es van unir en un sol fitxer i s'hi va modificar la codificació de llengües, ja que MateCat havia assignat la combinació en>oc-FR atesa la configuració del projecte. Posteriorment, es va dur a terme un control de qualitat automatitzat sobre aquest fitxer amb els programes Heartsome TMX Editor i Okapi Checkmate. Els resultats sobre els quals es va parar una atenció especial van ser les repeticions i els segments antics (segments que havien estat modificats i que havien quedat a la memòria), a banda d'incoherències entre traduccions (segments de partida que tenien traduccions diferents o a l'inrevés). La memòria final, d'aproximadament 3220 segments i gairebé 20.800 paraules (en llengua d'arribada), va ser preparada per posar-la a disposició de la comunitat (vegeu més avall).

En segon lloc, es va dur a terme una verificació tècnica de la traducció. Aquesta fase va consistir en la instal·lació dels paquets de llengua en diferents dispositius. Malgrat que es fa difícil d'assegurar que absolutament totes les traduccions han estat verificades (la qual cosa implicaria visitar la totalitat de funcions del programa), es va poder detectar, almenys, un problema d'importància relativa, que anava més enllà de la correcció o l'adequació de les traduccions.



Imatge 6. Problemes detectats en la fase de verificació

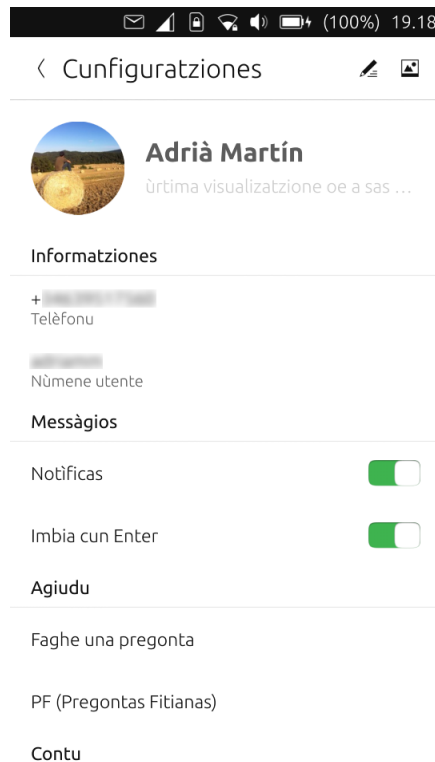
Tal com es pot observar a la imatge anterior, un cop aplicats els fitxers de llengua, alguns fragments de l'aplicació quedaven en anglès, com ara els noms de la setmana (a la captura anterior, «Thu», abreviació de *dijous*), mentre que d'altres apareixien amb una codificació alfanumèrica (a la captura, «16 M12», 'setze de desembre'). Gràcies un cop més a la col·laboració de Softcatalà, i especialment de Joan Montané, es va identificar la causa del problema. Telegram fa servir CLDR (Common Locale Data Repository), una base de dades lingüística utilitzada per diversos fabricants de maquinari que conté fragments habituals en diverses llengües. D'aquesta manera, els desenvolupadors de programari que necessiten elements habituals de les llengües a l'hora de desenvolupar programes poden fer-hi referència. Per mitjà de la inclusió d'un codi de llengua i de país al codi font, els programes «externalitzen» alguns segments habituals com ara els noms dels dies de la setmana o dels mesos de l'any. El sard no és present a CLDR, per la qual cosa l'ús del codi *sc* generava les errades de la imatge 6. Es va decidir, doncs, que es crearia un codi d'estat fictici (*sc-XS*) per tal que Telegram recuperés els fragments esmentats en anglès.

Per últim, es van preparar alguns recursos per tal de posar-los a disposició de la comunitat (vegeu més avall), entre els quals, unes instruccions per a l'usuari sobre la instal·lació de paquets de llengua, un full de càlcul amb els segments en anglès, català i sard alineats amb la corresponent línia de codi per a cada segment de les tres plataformes i la memòria de traducció.



## 7. La distribució del paquet de llengua

Seguint les recomanacions de Softcatalà (Mas i Montané, 2017), la distribució del paquet es fa per mitjà del canal @sardware de Telegram, <https://telegram.me/sardware>. Aquesta va ser l'opció triada per davant d'alternatives com ara la creació d'un robot, pel fet que una solució com aquesta requereix un cert grau de programació. El fet de recórrer a un canal, en canvi, permet tenir un registre dels usuaris subscrits i avisar-los cada cop que apareix una versió nova. Aquest mètode, que com a solució provisional pot resoldre algunes qüestions urgents, no acaba d'agradar ni als usuaris (Mas i Montané, 2017) ni als mateixos desenvolupadors de Telegram (del web —v. nota al peu 6—: «Yes, this *is* damn tricky, we'll simplify the process in the next versions.»).



*Imatge 7. Telegram per a Ubuntu Phone en sard*

## 8. Conclusions

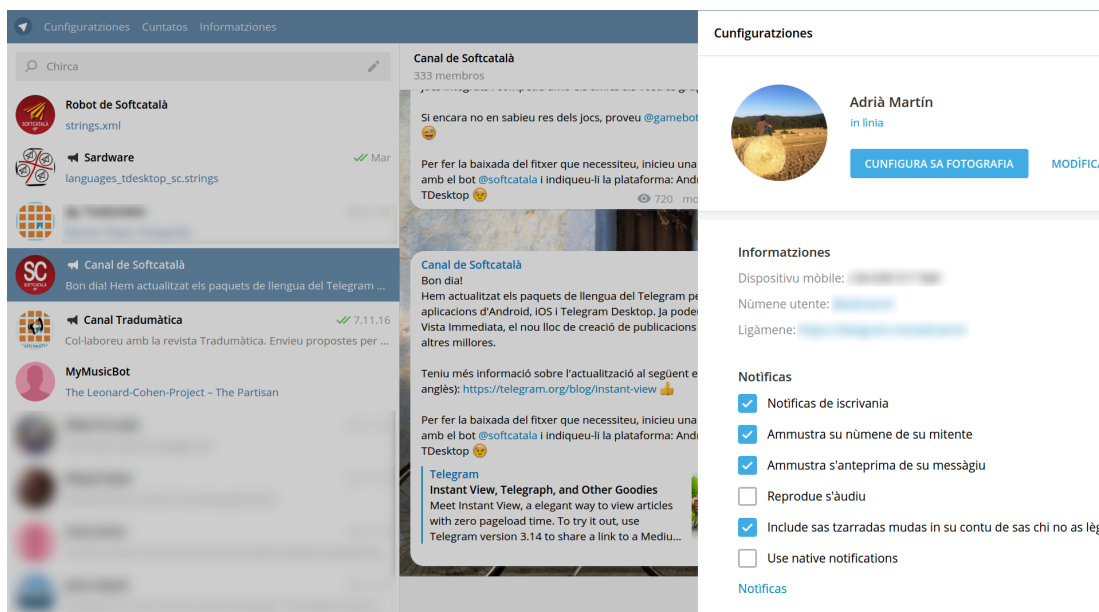
En aquest article hem presentat un projecte real de localització del client de missatgeria lliure Telegram a la llengua sarda. Aquest projecte pren com a referència experiències similars d'èxit per a altres comunitats de llengües minoritzades, com ara la catalana, i per tant contribueix a la diversitat lingüística digital. El fet que el sard sigui una llengua en procés d'estandardització implica que no es disposa de material normatiu de referència, com ara diccionaris o gramàtiques, i en el cas concret que hem presentat, d'un organisme normatiu de terminologia, bàsic per a l'àmbit tecnològic.

Cal tenir en compte que les traduccions han estat publicades recentment, i per tant en el moment de redactar aquest article els autors no tenen dades sobre l'acollida del projecte. En



qualsevol cas, atès que Telegram publica noves versions del codi del programa amb una certa regularitat (des del començament del projecte el juliol del 2016 fins al moment de la redacció d'aquest article, novembre del 2017, hi ha hagut almenys tres versions noves publicades), els autors consideren que un dels reptes més grans serà el manteniment del projecte de traducció. En aquest sentit, l'acollida per part de la societat sarda tindrà, ben segur, repercussions en el compromís de l'equip de traducció.

Quant a la publicació de la informació lingüística generada, està previst —a banda, tal com s'ha dit anteriorment, de la difusió per mitjà del canal @Sardware de Telegram (<http://telegram.me/sardware>)— publicar les memòries de traducció, els fitxers de llengua i glossaris terminològics al web <http://sardware.tradumatica.net>.



Imatge 8. Aplicació d'escriptori de Telegram en sard

Cal esmentar que, en l'última fase del projecte i just abans de publicar aquest article (desembre del 2016), el robot @TGLanguagesbot, que proporcionava accés alternatiu als fitxers de llengua originals de Telegram, va deixar de funcionar. Això suposa una nova dificultat per a accedir als fitxers originals i obre un interrogant sobre la manera com es podran dur a terme projectes com aquest en el futur.

Des del punt de vista del desenvolupament de programari i, fins i tot, de maquinari, aquest projecte ha permès constatar la necessitat de treballar per tal que el sard s'afegeixi a la base de dades CLDR, un procés relativament llarg però rendible des del punt de vista de l'optimització d'esforços en la traducció. Comunitats lingüístiques numèricament més petites, com ara l'asturiana, han seguit recentment aquesta via.<sup>18</sup> Aquest projecte també ha permès comprovar que alguns sistemes operatius encara no permeten fer servir tots els signes ortotipogràfics de la llengua sarda, com ara el punt volat (compartit amb el català). Aquest caràcter (·) és el recomanat per la proposta LSC per tal de separar els clítics de les formes verbals (*nara·mi*, 'digue'm'). El fet que el caràcter recomanat no estigui disponible en alguns

<sup>18</sup>Vegeu <http://noticies.softastur.org/ya-tamos-en-cldr.html>.



sistemes operatius pot desencoratjar alguns parlants i suposa una dificultat per a l'estandardització de la llengua.

Com a línies de futur, l'experiència podria ser replicada per a la resta de llengües de Sardenya. De fet, mentre que per a algunes de les llengües només caldria *localitzar* (en el sentit més estricte d'adaptar a les varietats locals) versions ja existents (com ara la catalana per a l'alguerès), la resta de llengües es podrien localitzar per mitjà de la coordinació amb grups d'usuaris afins (la comunitat corsa per al gal·lurès; la lígur per al tabarquí; la sarda per al sassarès). Cal no oblidar que absolutament totes les llengües pròpies de Sardenya (sard, gal·lurès, sassarès, català i lígur) estan catalogades per l'Atlas de la Unesco com a «definitely endangered». Des del punt de vista de la política lingüística institucional, doncs, apostar per un projecte com el que proposem tindria un simbolisme històric i demostraria un compromís real amb la preservació del patrimoni lingüístic de Sardenya.

### Bibliografia

- Beccu, A.; Martín-Mor, A. (2016). «Sa traduzione de Facebook in sardu». Tradumàtica, 14. Comitau Scientificu po sa Norma Campidanese de su Sardu Standard (2009). *Arrègulas po ortografia, fonètica, morfologia e fueddàriu de sa Norma Campidanese de sa Lingua Sarda*. Quartu S. Elena: Alfa.
- Diaz Fouces, O. (2012), «La naturaleza de las habilidades tecnológicas en la formación de traductores y el papel del software libre», dins Marcos Cánovas, Gemma Delgar, Lucrecia Keim, Sara Khan i Àngels Pinyana (eds.), *Challenges in languages and translation teaching in the web 2.0 era*. Granada: Comares, 159-167.
- Martín-Mor, A.; Piqué, R.; Sánchez-Gijón, P. (2016). *Tradumàtica: Tecnologies de la traducció*. Vic: Eumo.
- Mas, J.; Montané, J. (2017). Història del Telegram en català. Blog Tradumàtica: [www.blog.tradumatica.net](http://www.blog.tradumatica.net).
- Mura, R.; Viridis, M. (eds.) (2015). *Caratteri e strutture fonetiche, fonologiche e prosodiche della lingua sarda. Il sintetizzatore vocale SINTESA*. Càller: Condaghes.
- Sandrini, P., García González, M. (ed.). (2015). *Translation and openness*. Innsbruck: Innsbruck University Press. Disponible a [https://www.uibk.ac.at/iup/buch\\_pdfs/translation-and-openness/9783902936882.pdf](https://www.uibk.ac.at/iup/buch_pdfs/translation-and-openness/9783902936882.pdf). [últim accés: desembre del 2016].
- Softcatalà (2010). *Guia d'estil per a la traducció de productes informàtics*. Disponible a <https://www.softcatala.org/guia-estil-de-softcatala/>. [últim accés: desembre del 2016].

### Índex de taules

Taula 1. Terminologia de Telegram	8
Taula 2. Expressions innecessàries	9
Taula 3. Expressions sintètiques	10

### Índex d'imatges

Imatge 1. Les llengües de Sardenya. Mapa editat pel govern sard	3
Imatge 2. Entrada de la Viquipèdia sarda corresponent a la Limba Sarda Comuna. El missatge de la dreta indica que l'article està escrit en LSC	5
Imatge 3. L'entrada «Limba sarda» de la Viquipèdia, escrita en logudorès i en campidanès	5



---

Imatge 4. Interfície de MateCat amb els projectes de traducció de Telegram	7
Imatge 5. Abreviatures per als mesos de l'any	11
Imatge 6. Problemes detectats en la fase de verificació	12
Imatge 7. Telegram per a Ubuntu Phone en sard	13
Imatge 8. Aplicació d'escriptori de Telegram en sard	14