

Octubre 2023

APLICACIONS DE TRANSCRIPCIÓ I SUBTITULACIÓ EN CATALÀ: UN RECURS PER A LES PERSONES AMB SORDESA?

Anàlisi dels serveis en català
de les aplicacions i pàgines
web



L'OBSERVATORI
ACCESSIBILITAT I COMUNICACIÓ
ORAL DE LES PERSONES AMB SORDESA

ACAPPS
Famílies i persones amb sordesa



Aquest treball ha estat possible gràcies a:

La direcció de continguts, revisió i redacció final de David Prujà.

L'elaboració de tot el treball de camp, l'aplicació de la metodologia NER i la dedicació de moltes hores per part de Daniela Rodríguez.

El coneixement de la norma UNE de subtitulació i la metodologia NER i el seu consell a la Daniela de la Maria Riu.

El coneixement de la norma UNE de subtitulació i la metodologia NER i el seu consell a la Daniela de la Maria Riu.

El suport de **Softcatalà**, en especial de Jordi Mas, que ens ha generat una eina per comparar la transcripció original amb la que ens aportaven les diferents aplicacions, facilitant-nos així molt la feina.



En aquest document es glosa el resultat d'una prova de la funcionalitat de diferents aplicacions de transcripció d'àudios i vídeos al català, posant-les a prova amb diferents variants dialectals.

Es tracta d'una exploració amb la intenció d'aconseguir una opinió vers la utilitat d'aquestes aplicacions per garantir el dret a la informació i la comunicació de les persones amb sordesa.



Índex

01	Presentació.....	6
02	Metodologia.....	8
	Mesura de la qualitat de la transcripció.....	9
	Criteris de selecció de vídeos.....	10
	Criteris de selecció de fragments.....	10
	Selecció dels webs i aplicacions.....	11
03	Característiques dels vídeos.....	13
	Un jove pagès menorquí.....	14
	Refranys del món de la pagesia.....	14
	El català a Perpinyà.....	14
	Acte unitari “No als Parcs Eòlics”.....	14
	Inauguració ExpoEbre.....	14
	Premis Federació ACAPPS.....	14
	Evolució de l’alumnat amb sordesa a Catalunya.....	14
04	Informació de les funcionalitats de les aplicacions.....	15
	Amberscript.....	16
	Happy Scribe.....	18
	Omniscien.....	20
	Clideo.....	20
	FreeSubtitles.AI.....	22
	SoftCatalà.....	24
05	Quadre general resum de la funcionalita de les aplicacions.....	26
06	Valoració de la transcripció automàtica per vídeos.....	28
	Plenari de l’Ajuntament de Xàtiva.....	29
	Un jove pagès menorquí.....	29
	Refranys del món de la pagesia.....	29
	El català a Perpinyà.....	29
	Acte unitari “No als Parcs Eòlics”.....	29
	Inauguració ExpoEbre.....	29
	Premis Federació ACAPPS.....	30
	Evolució de l’alumnat amb sordesa a Catalunya.....	30
07	Anàlisi comparada de les aplicacions de transcripció i subtitulció.....	31
08	Resultats Aplicacions i Variants Dialectals.....	35
	Aplicacions.....	36
	Softcatalà:.....	36
	Freesubtitles:.....	36
	Clideo:.....	36
	Omiscien.....	36



Happy Scribe	36
Amberscript.....	36
Variants dialectals:.....	36
Valencià.....	36
Tortosí:.....	36
Català Oriental:.....	36
Rossellonès:.....	36
Català Central.....	36
09 . Conclusions	37

01 Presentació



Les persones amb sordesa que es comuniquen oralment tenen com una aliada estratègica bàsica la tecnologia. Ha estat gràcies a la millora de les pròtesis auditives des dels anys 90 – audiòfons i implants coclears bàsicament – que han pogut millorar substancialment la seva qualitat de vida i superar molts dels obstacles que la sordesa els posa en el seu projecte vital.

El desenvolupament de la Intel·ligència Artificial i les seves múltiples aplicacions sembla que ens sorprendrà amb canvis que avui ni ens imaginem.

Tanmateix, una de les coses que sembla que podrà concretar-se és la possibilitat d'aconseguir aplicacions que transcriguin textos amb una precisió a vegades sorprenent.

Aquesta funcionalitat és estratègica per a les persones amb sordesa, donat que com sempre insistim des de la Federació ACAPPS, les capacitats de les persones amb sordesa són diverses i en molts casos, per a poder accedir a la informació i a la comunicació són moltes les que necessiten subtítols en directe.

Aquest és un servei que avui han de fer les persones, com els intèrprets de tota mena d'idiomes, i no és possible la transcripció literal feta manualment, ja que parlem molt més de pressa del que es pot transcriure en temps real, de la mateixa manera que en molts contextos la interpretació en directe d'altres idiomes no és tampoc una traducció literal.

Això els desenvolupadors de software ho saben i cada vegada són més les aplicacions que apareixen que permeten subtítular vídeos, o transcriure'n els diàlegs de vídeos o d'àudios o fins i tot aplicacions que en temps real transcriuen els textos d'una conversa en el present.

Des de la Federació ACAPPS en el nostre treball per l'accessibilitat, hem estat amatents a aquesta realitat i hem pogut copsar com alguna d'aquestes aplicacions obtenen resultats sorprenentment bons en anglès, fins i tot en castellà i a vegades també en català tot i que no tant.

Quan hem posat a prova de manera no sistemàtica aquestes aplicacions hem pogut observar que depenent de l'idioma la seva

eficàcia és una o és una altra i que depenent també del context el resultat també canvia.

Cal tenir present que el que els desenvolupadors anomenen "motor" d'aquestes aplicacions no està pensat per a la transcripció del català, de manera que els cal encara un treball d'adaptació com a mínim al nostre idioma.

És per això que hem volgut generar un model de prova que ens permeti formar-nos una opinió fonamentada en fets objectius i poder traslladar aquest coneixement tan a les persones amb sordesa, per a que puguin saber quina aplicació els pot convenir més en cada context, a les persones que treballen desenvolupant aquestes eines, per a que tinguin presents les necessitats i particularitats de les persones amb sordesa i, finalment a totes aquelles empreses, institucions i entitats que s'interessen per aquesta tecnologia com un recurs per millorar l'accessibilitat de la comunicació oral de la seva activitat i instal·lacions.

Considerar aquestes aplicacions com una mesura d'accessibilitat substitutòria de la subtitulació en directe feta per persones requereix poder assegurar que en qualsevol context no generaran equívocs que poden resultar fatals. Contextos com el judicial, el mèdic o el científic són àrees en què resulta fàcil imaginar errors de transcripció que poden resultar fatals o si més no molt inconvenients.

No podem deixar d'agrair a les persones desenvolupadores de les empreses de software amb les que hem contactat, que en conèixer la nostra intenció han col·laborat amb nosaltres, facilitant-nos l'accés i fins i tot en alguns casos una formació per a poder fer una anàlisi completa de les seves funcionalitats.

Molt especialment a [Softcatalà](#), que ens han ajudat posant a la nostra disposició tot el seu coneixement tecnològic i generant una aplicació ad hoc que ens ha facilitat i automatitzat la feina d'identificar les diferències entre textos.

Serveixi aquest document com la primera aproximació (de les moltes que caldrà fer els propers anys) a una anàlisi en detall de diferents aplicacions amb vocació de servei.

02 Metodologia

Mesura de la qualitat de la transcripció

Aquest projecte comença a plantejar-se a l'entitat com una necessitat i és a mitjan 2022 quan comencem a explorar de quina manera podríem abordar una anàlisi pràctica i objectiva d'aquests recursos.

Som conscients que la limitació de recursos no ens ha permès abastar ni totes les aplicacions del mercat, ni totes les casuístiques que serien d'interès per a l'anàlisi. Tanmateix pensem que hem fet un treball rigorós i que pot ser molt útil als tres públics als que ens dirigim amb aquest treball: persones amb sordesa, desenvolupadores i responsables d'accessibilitat.

Volem però posar pel davant les limitacions amb les que hem abordat l'anàlisi, per tal que ningú no s'esperï allò que no hem fet:

1. **Coneixement.** Si bé a la Federació ACAPPS acumulem molts anys d'experiència en l'assistència a les persones amb sordesa per a l'accessibilitat de la comunicació oral, les noves tecnologies i avanços de la comunicació es desenvolupen cada vegada amb més rapidesa i sense pausa. Enfrontem aquest repte sense un coneixement específic de la tecnologia i ho fem en tant que persones usuàries i creiem que aquesta perspectiva és la que és més vàlida per resultar útils a les persones amb sordesa. No obstant això, hem elaborat l'anàlisi objectiva de la manera més rigorosa possible i hem valorat l'experiència que els tres públics als que va dirigida aquesta anàlisi poden trobar-hi.

2. **Caducitat i limitació de recursos.** Una de les principals pegues d'aquesta anàlisi és la velocitat a la que es belluga la tecnologia, de manera que la caducitat és molt propera a la seva publicació. L'esforç fet per l'entitat en l'aposta de temps i recursos per a la seva realització no és sostenible sense un suport públic específic. Ens agradaria pensar que aquest és el primer episodi d'una relació amb aquesta tecnologia que dona lloc a un document viu que pot renovar-se periòdicament, però estarem a expenses dels recursos públics i/o privats que puguem aconseguir.

3. **Mostra.** La selecció de vídeos s'ha fet tenint en compte una sèrie de criteris tècnics que hem considerat d'interès, però encara queden casos per analitzar. Ens hagués agradat

afegir en la nostra anàlisi la presència de vídeos en els que es barregi castellà i català en parlar, una situació molt comuna a casa nostra, o vídeos en els quals l'àudio tingui molta reverberació. Són casos que buscarem introduir en futures actualitzacions. La limitació també ha afectat al nombre d'aplicacions valorades. Hem prioritzat aquelles que complien un equilibri entre les que més ens han demandat persones amb sordesa, les que en una prova inicial ens han semblat més efectives en la transcripció i aquelles que o bé no són de pagament o bé en un primer contacte explicant les nostres motivacions, les empreses desenvolupadores ens han deixat una versió de demostració.

La generació d'una metodologia per analitzar objectivament l'efectivitat d'aquestes aplicacions quan les enfrontem amb el català ha estat el repte més ambiciós d'aquest treball. Li hem volgut dedicar tot el temps que ha calgut per poder repetir l'anàlisi de forma recurrent amb la mateixa metodologia i poder-ne apreciar l'evolució.

Així hem hagut de fer una mica de recerca i escoltar persones que en saben molt més que nosaltres i hem descobert que existeix una metodologia internacional que mesura quantitativament els errors que generen aquestes aplicacions, el mètode WER.

Aquesta metodologia, tot i que molt útil pensàvem que no ens permetria concloure massa, donat que un error de transcripció pot afectar molt o gens a la comprensió del que s'està transcrivint.

Per això hem volgut fer també una anàlisi qualitativa. Ambdues metodologies combinades ens permeten una visió més àmplia de la utilitat i capacitat de cadascuna d'aquestes aplicacions.

A continuació expliquem amb una mica de detall en què consisteixen aquests dos enfocaments que hem utilitzat per fer-ne la valoració:

- **Objectiu.** Per a poder valorar d'una manera objectiva el resultat de transcripció de les aplicacions hem utilitzat el mètode WER mitjançant una eina proporcionada per Softcatalà. El mètode WER ('word error rate') realitza un càlcul del nombre de substitucions, esborrats i insercions que han de fer-se sobre la transcripció automàtica perquè sigui idèntica a la transcripció manual de referència que s'ha fet per a cada vídeo. Com més s'acosta a 0 el



resultat, major és la qualitat de la transcripció de l'aplicació, ja que significa que cal fer menys modificacions perquè sigui igual a la transcripció manual.

- Subjectiu. Per a realitzar l'anàlisi subjectiva hem utilitzat el model NER recollit en la Norma UNE 153010 de 2012 per a la subtitulació per a persones amb sordesa. Amb aquest mètode es mesura la precisió de la transcripció tenint en compte el nombre de paraules i la gravetat dels errors.

- Es consideren errors greus "aquells que canvien el significat del discurs original, creant un significat nou que podria tenir sentit en aquest context". Cada error greu té valor de 1.
- Es consideren errors normals "aquells que, si bé no creen un nou significat, ometen una unitat d'informació del discurs original", aquells en els quals es detecta l'error perquè la frase perd sentit. Cada error normal té valor de 0,5.
- Es consideren errors lleus "aquells que permeten seguir la coherència del discurs original i fins i tot a vegades les seves paraules. Alguns casos de errors lleus de reconeixement són títols o majúscules incorrectes, petites insercions o omissions...". Cada error lleu té valor de 0,25.

Al total de paraules (incloent signes de puntuació com a paraula) se li resta el valor conjunt dels errors. Aquest resultat es divideix entre el nombre de paraules i es multiplica per cent. D'aquesta manera, el resultat indica **major precisió i qualitat com més proper al 100%**.

Exemple:

Una gota de petroli pot contaminar 25 litres d'aigua

Un gota de, petróli vol contaminar 35 litres d'algú

Nombre de paraules de la transcripció automàtica: 12

Errors lleus: $3 \times 0,25 = 0,75$

Errors normals: $2 \times 0,5 = 1$

Errors greus: $1 \times 1 = 1$

$0,75 + 1 + 1 = 2,75$

$[(12 - 2,75) / 12] \times 100 = (9,25 / 12) \times 100 = 77,08 \approx 77\%$ de precisió

En aquesta anàlisi subjectiva és possible que també es trobin paraules entre parèntesis escrites amb un altre color, i no subratllades, com (aquesta). Això significa omissió per part de la transcripció de l'aplicació d'alguna de les parts del vídeo. Aquestes paraules poden ser vermelles, grogues o verdes depenent de la gravetat, seguint els mateixos criteris ja esmentats.

És important aclarir, arribats a aquest punt, que l'anàlisi subjectiva pot variar segons la persona que el realitzi, arribant fins i tot a no considerar com a errors alguns canvis que es donen entre la transcripció de l'aplicació i la manual com, per exemple, l'ús optatiu de les comes. El que té major importància és que totes les transcripcions siguin valorades amb igualtat de criteri.

Criteris de selecció de vídeos

Per a assegurar la representació fidel de l'idioma, hem triat 8 vídeos de diferents zones geogràfiques, amb diferents dialectes i divergències lingüístiques del català. Per a cada vídeo s'han triat fragments que en conjunt tenen una durada aproximada de 2-3 minuts. Els vídeos utilitzats per a analitzar les aplicacions són:

- Plenari de l'Ajuntament de Xàtiva; València
- Acte unitari "No als Parcs Eòlics" de l'Alt Empordà; Català Central del nord.
- El català a Perpinyà; Rossellonès.
- Evolució de l'alumnat amb sordesa a Catalunya 25es jornades; Català Central
- Inauguració ExpoEbre; Tortosí.
- Premis Federació ACAPPS Commemoració 25è aniversari Jornades; Català Central.
- Refranys del món de la pagesia; Tortosí
- Un jove pagès menorquí fa de missatge; Català Oriental, Menorquí.

Criteris de selecció de fragments

Els vídeos seleccionats eren de durades diferents i hem hagut d'editar-los per afavorir el compliment de diferents criteris que hem considerat que aporten valor a la selecció per la diversitat de situacions lingüístiques i de captació del so a les que sotmetem aquestes aplicacions:



- Qualitat general del vídeo
- Qualitat general de l'àudio
- Canvis en el so:
 - Qualitat de l'àudio
 - Diversos interlocutors
- Soroll de fons:
 - Altres persones parlant (multituds)
 - Interrupcions d'altres persones
 - Animals
 - Música
 - Altres sorolls que afectin el so de la persona que parla
- Interrupcions
- Repetició de paraules
- Esment de sigles
- Esment de noms propis
- Elements que impedeixen la visualització i/o afecten el so:
 - Ús de mascareta
 - Micròfons mal regulats
- Ús de paraules complexes
- Pronunciació poc clara
- Rapidesa quan es parla

Selecció dels webs i aplicacions

Hem recollit una varietat d'aplicacions i webs de subtítulat i transcripció mitjançant consultes a les persones amb sordesa que ens han fet consultes i a persones tècniques de les entitats ACAPPS que estan en permanent contacte amb persones amb sordesa i han recaptat aquesta informació durant uns mesos.

Amb una primera recerca per comprovar quines d'aquestes aplicacions i webs són de pagament o gratuïtes i les de pagament quines estaven disposades a cedir una versió demostrativa per a la realització de l'anàlisi, ja en vam descartar unes quantes.

Finalment, amb una prova inicial hem pogut observar com algunes d'aquestes aplicacions encara estan lluny de poder oferir una qualitat suficient en català, de manera que la conjunció d'aquests criteris ens ha permès fer la selecció.

També hem tingut en compte la igualtat de condicions de les aplicacions en fer la selecció. Les aplicacions triades permeten introduir els vídeos seleccionats per a després descarregar la transcripció, és a dir, que reconeix l'àudio del vídeo. Analitzar altres aplicacions, com les de transcripció simultània, podria trencar aquesta igualtat de condicions, ja que en comptes de reconèixer l'àudio, estaria reconeixent la reproducció d'un àudio i no una veu natural, la qual cosa pot portar a una pitjor qualitat de transcripció produïda per la pròpia anàlisi i no per les capacitats de l'aplicació. En aquest cas existeix també una limitació legal: les aplicacions de transcripció simultània necessitarien escoltar el vídeo mentre aquest es reproduïx, amb la possibilitat que es gravin converses alienes, la qual cosa suposaria un delicte greu a la intimitat desenvolupat en l'article 197 del Codi Penal.

Les plataformes susceptibles a l'anàlisi són les que s'exposen a continuació. En negreta es troben les aplicacions que finalment han format part de l'estudi:

- Amberscript
- Happy Scribe
- Closed Caption Creator
- Kapwing
- Veed.io
- YellaUmbrella
- IOmniscien
- Autosubtitle
- Clideo
- Amazon transcribe
- Flixier
- Getsubly
- Sonix
- Freesubtitles AI
- Visualfy
- Softcatalà
- Whisper desktop app
- Webcaptioner
- Animaker
- Tactiq



Ens hagués agradat poder avaluar el comportament de l'aplicació de Visualfy, donat que és una aplicació que cada cop és més present a diversos espais públics i tenim referències positives per part de les persones de l'entitat. Ara bé, no ha estat possible, donat que aquesta aplicació genera subtítols al mòbil que no es poden recuperar després i no es pot fer una captura d'imatge perquè en aplicació de la llei de protecció de dades té cura que ningú no la pugui fer servir per espiar terceres persones.

Aquesta dificultat la trobarem amb diferents aplicacions en futures proves, de manera que ens caldrà definir metodologies diferents per a l'avaluació de la seva efectivitat.

03 Característiques dels vídeos

Plenari de l'Ajuntament de Xàtiva

La qualitat d'àudio i vídeo en aquest cas és molt bona. No obstant això, els interlocutors porten mascareta tot el temps. És un vídeo d'un plenari de l'Ajuntament de Xàtiva, al País Valencià, així que es parla amb variants valencianes pròpies del català occidental.

El vídeo original dura més d'una hora i n'hem editat un seguit de fragments, retallant el vídeo original.

Durada total del conjunt de fragments: 2:22

Un jove pagès menorquí

En aquest vídeo la qualitat d'imatge i d'àudio no són bones. Hi ha soroll constant. Es tracta d'un vídeo d'un jove de Menorca, per la qual cosa el dialecte que es parla és del bloc oriental de la llengua catalana i forma part del català balear, subdialecte menorquí.

El vídeo original dura gairebé deu minuts.

Durada total del conjunt de fragments: 2:06

Refransys del món de la pagesia

La qualitat d'imatge d'aquest vídeo és molt bona. La qualitat d'àudio també és bona, encara que a vegades hi ha sorolls de fons. El vídeo està gravat a Tivissa, un municipi de la província de Tarragona. El dialecte que es parla pertany al bloc occidental.

El vídeo original dura gairebé deu minuts.

Durada total del conjunt de fragments: 2:47

El català a Perpinyà

La qualitat d'àudio és bona, però la qualitat d'imatge no ho és tant. La pronunciació és clara encara que hi ha música i sons de fons que poden dificultar la comprensió. El vídeo està centrat en Perpinyà, una ciutat al sud de França, pertanyent el seu dialecte al bloc oriental, català septentrional.

La durada del vídeo original ronda els set minuts.

Durada total del conjunt de fragments: 2:10

Acte unitari "No als Parcs Eòlics"

El vídeo té molt bona qualitat, tant d'àudio com d'imatge. A vegades no es veu a la persona que parla, però la pronunciació és clara i no hi ha

sorolls que dificultin la comprensió. Es tracta d'un vídeo en el qual els interventors són residents de Capmany, al nord de Girona.

El vídeo original dura aproximadament sis minuts.

Durada total del conjunt de fragments: 2:07

Inauguració ExpoEbre

La qualitat d'imatge d'aquest vídeo no és dolenta, però tampoc massa bona. La qualitat d'àudio és bona en general, encara que hi ha canvis bruscos de volum. El dialecte és tortosí, del bloc occidental.

Durada total del conjunt de fragments: 2:54

Premis Federació ACAPPS

En aquest vídeo la qualitat del so és bona. Hi ha diversos interlocutors, però el dialecte en el que es parla és català central.

El vídeo original dura una mitja hora.

Durada total dels fragments seleccionats: 2:05

Evolució de l'alumnat amb sordesa a Catalunya

La qualitat d'imatge del vídeo no és bona, i l'àudio tampoc és massa bo, hi ha reverberació. El dialecte és català central.

La durada del vídeo original és aproximadament vint minuts.

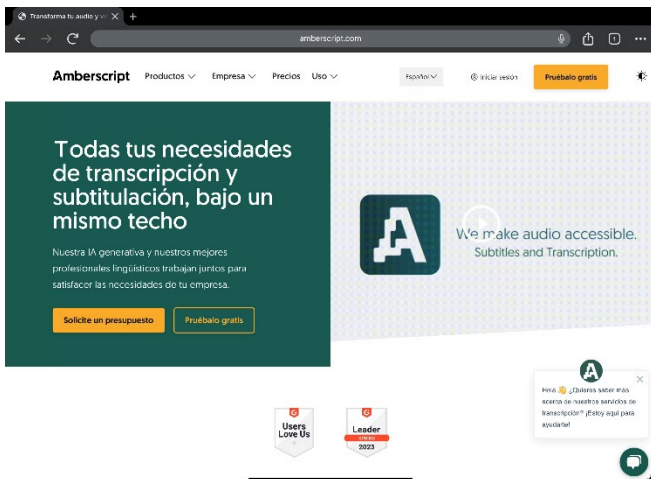
Durada total dels fragments seleccionats: 2:39

04. Informació de les funcionalitats de les aplicacions

Amberscript

subtitulació de persones amb sordesa UNE
153010:2012.

Figura 1. Aspecte web Amberscript



Captura de pantalla APP Amberscript

Amberscript és una empresa que ofereix, entre altres productes, transcripció automàtica, transcripció professional, subtítols automàtics, subtítols professionals i subtítols traduïts.

Admet arxius des del dispositiu, o mitjançant enllaços de Youtube o Google Drive. En pujar un arxiu s'ha de determinar l'idioma i el nombre de persones que parlen en ell. Després es tria el servei (transcripció o subtitulació) i si es vol que sigui realitzat de la mà d'un professional o de la màquina. També permet crear un diccionari que ajudi al programa de reconeixement.

Una vegada processat el vídeo, la transcripció pot exportar-se en diversos formats, incloent o excloent marques de temps i nombre de parlants:

- Word (.docx)
- JSON (.json)
- Text (.txt)
- Excel (.csv)

L'editor és bastant complert: permet editar el nombre i nom dels parlants, el contingut de la transcripció, té un corrector d'ortografia i permet subratllar. També es pot canviar la velocitat del vídeo mentre s'edita.

És una aplicació intuïtiva i fàcil d'utilitzar, amb passes clares i un tutorial d'editor fàcil d'entendre. La pàgina està disponible en castellà, però no en català.

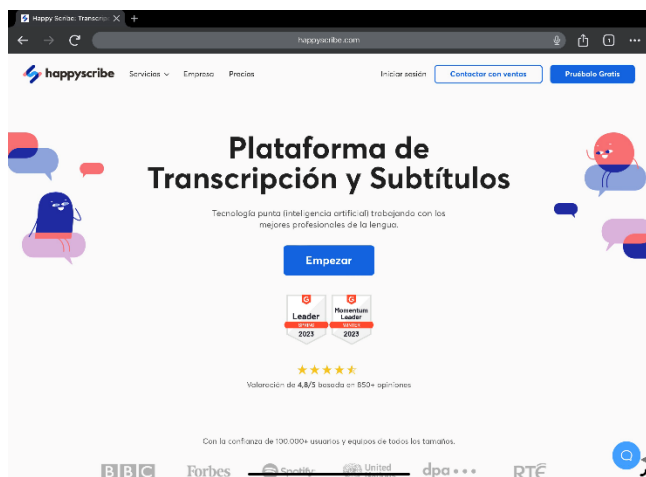
El fet que no permeti editar l'estètica dels subtítols impedeix adaptar-los a la norma de



<p>Serveis</p>	<ul style="list-style-type: none"> • Transcripció automàtica: SÍ • Transcripció professional: SÍ (de pagament) • Subtítols automàtics: SÍ • Subtítols professionals: SÍ (de pagament) • Subtítols traduïts: SÍ
<p>Principals formats d'exportació</p>	<ul style="list-style-type: none"> • Word (.docx): SÍ • JSON (.json): SÍ • Text (.txt): SÍ • Excel (.csv): SÍ • Subtitles (.srt): SÍ • Subtitles (WebVTT .vtt): SÍ
<p>Editor</p>	<ul style="list-style-type: none"> • Permet editar el nombre i nom dels parlants: SÍ • Permet editar el contingut de la transcripció: SÍ • Té corrector d'ortografia: SÍ • Permet Subratllar: SÍ • Permet editar la durada dels subtítols en pantalla: SÍ • Permet editar l'estètica dels subtítols: NO
<p>Idiomes</p>	<ul style="list-style-type: none"> • Pàgina en català: NO • Pàgina en castellà: SÍ • Servei en català: SÍ • Servei en castellà: SÍ
<p>Prova gratuïta</p>	<ul style="list-style-type: none"> • Crèdit gratuït de 10 minuts
<p>Versió de pagament</p>	<ul style="list-style-type: none"> • S'ha de demanar pressupost. Canvia depenent del tipus de servei i el nombre d'hores de vídeo.

Happy Scribe

Figura 2. Aspecte web Happy Scribe



Captura de Pantalla Happy Scribe

Happy Scribe és una eina de transcripció i subtítulat disponible en 62 idiomes.

Permet afegir arxius des del dispositiu, enllaços de Youtube o Google Drive, Dropbox, Box, Zoom i Wistia. Una vegada afegit el vídeo, s'han d'ajustar les referències de llenguatge. També permet afegir un vocabulari propi que ajudi al reconeixement. Finalment, s'ha de seleccionar si es vol un servei de subtítulat o transcripció professional o generat per la màquina.

Una vegada processat el vídeo, la transcripció pot exportar-se en diversos formats, incloent o excloent marques de temps i nombre de parlants:

- Excel (.csv)
- DaVinci Resolve Markers (.edl)
- JSON (.json)
- Maxqda (.txt)
- Final Cut Pro (.fcpxml)
- Premiere Pro Markers (.xml)
- Interactive Transcript (.html)
- Avid with on-screen timecodes (.txt)
- Avid DS Subtitles (.txt)
- Avid Markers (.txt)
- Facebook Subtitles (.srt)
- Subtitles (EBU-STL .stl)
- Subtitles (WebVTT .vtt)
- Subtitles (SubRip .srt)

- Word (.docx)
- PDF (.pdf)
- Text document (.txt)

L'editor permet editar el nombre i nom dels parlants, editar el contingut de la transcripció, buscar i reemplaçar, subratllar i convertir la transcripció en subtítulat. També es pot canviar la velocitat del vídeo mentre s'edita. Les marques de temps són editables i permet fer un seguiment de les parts que ja estan revisades. En l'editor de subtítols es pot editar la font (color, grandària, posició...), el fons de les lletres (negre, translúcid, sense fons, amb contorn, sense contorn...) i la posició (moure cap amunt o cap avall). També pot editar-se el nombre de línies que ocupen els subtítols, el màxim de caràcters per línia, l'espai entre les línies, els caràcters per segon, la durada mínima del subtítol... És un editor amb moltes opcions.

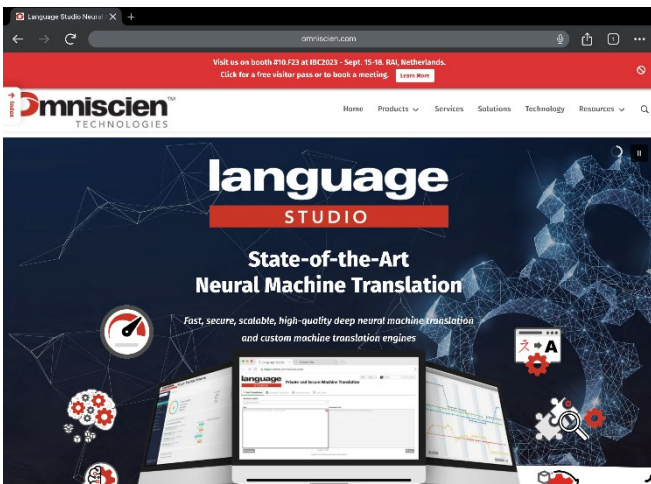
Happy Scribe és una aplicació que dona moltes facilitats. És intuïtiva, senzilla d'utilitzar i dona moltes eines d'edició i formats de descàrrega. La pàgina està disponible en castellà, no ho està en català i l'editor funciona en anglès.



<p>Serveis</p>	<ul style="list-style-type: none"> • Transcripció automàtica: SÍ • Transcripció professional: SÍ (de pagament) • Subtítols automàtics: SÍ • Subtítols professionals: SÍ (de pagament) • Subtítols traduïts: SÍ
<p>Principals formats d'exportació</p>	<ul style="list-style-type: none"> • Word (.docx): SÍ • JSON (.json): SÍ • Text (.txt): SÍ • Excel (.csv): SÍ • Subtitles (.srt): SÍ • Subtitles (WebVTT .vtt): SÍ
<p>Editor</p>	<ul style="list-style-type: none"> • Permet editar el nombre i nom dels parlants: SÍ • Permet editar el contingut de la transcripció: SÍ • Té corrector d'ortografia: NO • Permet Subratllar: SÍ • Permet editar la durada dels subtítols en pantalla: SÍ • Permet editar l'estètica dels subtítols: SÍ
<p>Idiomes</p>	<ul style="list-style-type: none"> • Pàgina en català: NO • Pàgina en castellà: SÍ • Servei en català: SÍ • Servei en castellà: SÍ
<p>Prova gratuïta</p>	<ul style="list-style-type: none"> • Crèdit gratuït per a subtitular i transcriure 1 vídeo
<p>Versió de pagament</p>	<ul style="list-style-type: none"> • Per hores de vídeo processat. 1 hora = 12€

Omniscien

Figura 3. Aspecte web Omniscien



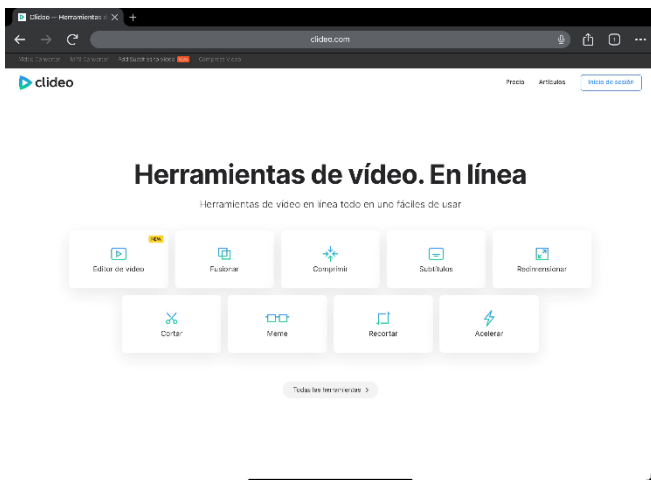
Captura de Pantalla Omniscien

Versió beta. No l'hem pogut provar personalment. No obstant això, els treballadors de Omniscien s'han mostrat molt col·laboratiu enfront del projecte d'anàlisi i ens han facilitat les transcripcions, de manera que els hem enviat els vídeos i ells els han processat en l'aplicació. Després d'analitzar una de les mostres vam decidir que, per la qualitat de la transcripció, mereixia estar en l'anàlisi. També vam tenir l'oportunitat de veure com serà l'aplicació i sembla molt completa, amb diverses opcions i idioma català disponible per a les transcripcions.

Tanmateix és important poder deixar clar que el procediment d'anàlisi ha estat diferent.

Clideo

Figura 4. Aspecte web Omniscien



Captura de Pantalla Omniscien

Clideo recull un conjunt d'eines dirigides a l'edició de vídeo en general, permetent redimensionar, tallar, accelerar o comprimir un vídeo. Entre moltes altres funcions, també compta amb una opció de subtítulat que es troba en versió Beta.

Una vegada processat l'arxiu, ens dona l'opció d'agregar manualment, pujar un arxiu .srt o generar els subtítols automàticament (aquesta última és la versió beta). Una vegada triada la generació automàtica, se selecciona l'idioma i es processa el vídeo. Una vegada processat, Clideo dona l'opció de descarregar la transcripció en un únic format (.srt).

L'editor permet canviar el contingut dels subtítols i modificar les marques de temps. També dona l'opció de canviar l'estil de subtítol, la posició, lletra, grandària, color...

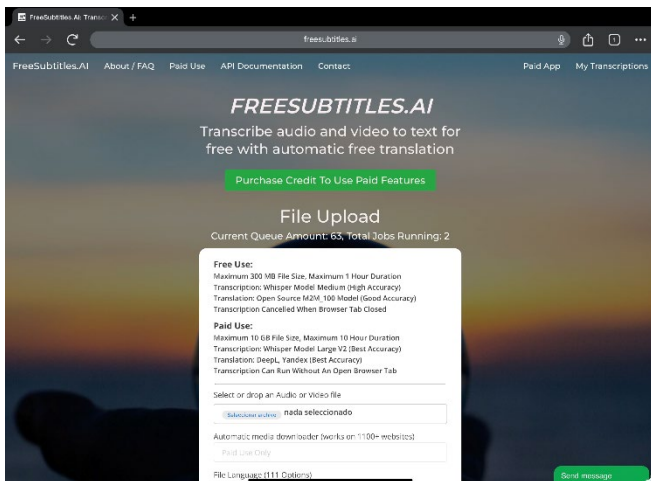
Les poques opcions de l'editor i la limitació de descàrrega a un sol format fan que no sigui l'aplicació més completa en comparació amb d'altres opcions que s'han analitzat. A més, la qualitat de la transcripció no és massa bona. S'ha de tenir en compte, no obstant això, que la transcripció automàtica segueix en fase de prova i que es tracta d'una aplicació més enfocada a l'edició general que a la subtítolació i transcripció de vídeos. La pàgina està disponible en castellà.



<p>Serveis</p>	<ul style="list-style-type: none"> • Transcripció automàtica: NO • Transcripció professional: NO • Subtítols automàtics: SÍ • Subtítols professionals: SÍ (de pagament) • Subtítols traduïts: SÍ
<p>Principals formats d'exportació</p>	<ul style="list-style-type: none"> • Word (.docx): NO • JSON (.json): NO • Text (.txt): NO • Excel (.csv): NO • Subtitles (.srt): SÍ • Subtitles (WebVTT .vtt): NO
<p>Editor</p>	<ul style="list-style-type: none"> • Permet editar el nombre i nom dels parlants: NO • Permet editar el contingut de la transcripció: SÍ • Té corrector d'ortografia: NO • Permet Subratllar: NO • Permet editar la durada dels subtítols en pantalla: SÍ • Permet editar l'estètica dels subtítols: SÍ
<p>Idiomes</p>	<ul style="list-style-type: none"> • Pàgina en català: NO • Pàgina en castellà: SÍ • Servei en català: SÍ • Servei en castellà: SÍ
<p>Prova gratuïta</p>	<ul style="list-style-type: none"> • Servei específic de subtitulació totalment gratuït
<p>Versió de pagament</p>	<ul style="list-style-type: none"> • (Per a altres funcions d'edició de vídeos). 9 dòlars al mes o 72 a l'any.

FreeSubtitles.AI

Figura 5. Aspecte web FreeSubtitles



Captura de Pantalla FreeSubtitles

Freesubtitles.ai és un web que ofereix una opció gratuïta molt àmplia: es poden pujar tots els vídeos que es vulgui amb un màxim de 300 MB per arxius i 1 hora de durada. Ofereix subtitulat i descàrrega de la transcripció, a més de traducció a un idioma diferent (entre 91 opcions, inclòs el català).

El primer que s'ha de fer és seleccionar un vídeo o àudio. Només permet introduir vídeos o àudios des del dispositiu, és a dir, no admet, per exemple, enllaços. Després es tria l'idioma entre 111 opcions. Després de pujar el vídeo, s'ha d'esperar en una cua perquè et toqui el torn de la transcripció (es mostra en pantalla la posició en la cua i un temps estimat d'espera). Una vegada processat el vídeo, permet descarregar la transcripció en els següents formats:

- Subtitles (.srt)
- Subtitles (WebVTT .vtt)
- Text document (.txt)

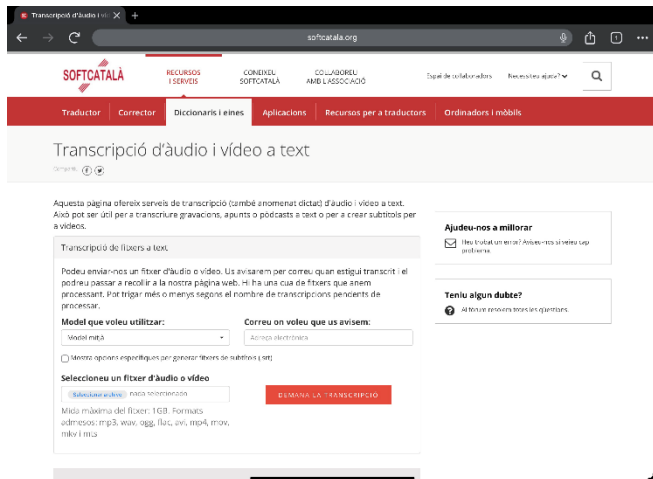
El web permet canviar la grandària i espaiat dels subtítols, però no el contingut d'aquests.

Les opcions d'edició del web són bastant limitades, però la transcripció no és de mala qualitat. A més, és la que té versió gratuïta més flexible. El procés és senzill i no requereix crear un perfil. La pàgina està disponible només en anglès.



<p>Serveis</p>	<ul style="list-style-type: none"> • Transcripció automàtica: SÍ • Transcripció professional: NO • Subtítols automàtics: SÍ • Subtítols professionals: NO • Subtítols traduïts: SÍ
<p>Principals formats d'exportació</p>	<ul style="list-style-type: none"> • Word (.docx): NO • JSON (.json): NO • Text (.txt): SÍ • Excel (.csv): NO • Subtitles (.srt): SÍ • Subtitles (WebVTT .vtt): SÍ
<p>Editor (No té editor)</p>	<ul style="list-style-type: none"> • Permet editar el nombre i nom dels parlants: NO • Permet editar el contingut de la transcripció: NO • Té corrector d'ortografia: NO • Permet Subratllar: NO • Permet editar la durada dels subtítols en pantalla: NO • Permet editar l'estètica dels subtítols: SÍ (limitat)
<p>Idiomes</p>	<ul style="list-style-type: none"> • Pàgina en català: NO • Pàgina en castellà: NO • Servei en català: SÍ • Servei en castellà: SÍ
<p>Prova gratuïta</p>	<ul style="list-style-type: none"> • Vídeos de 300 MB / 1 hora màxim
<p>Versió de pagament</p>	<ul style="list-style-type: none"> • 1 hora = 0,99 dòlars

Figura 5. Aspecte web SoftCatalà



Captura de Pantalla SoftCatalà

Softcatalà ofereix serveis de transcripció de vídeo i àudio a text. Admet fitxers mp3, wav, ogg, flac, avi, mp4, mov i mkv. Un cop seleccionat el vídeo, es dona el correu i es demana la transcripció. El que trigui a arribar al correu electrònic la transcripció dependrà de la longitud del vídeo i la quantitat de vídeos que tingui el web per processar. A partir d'aquí, permet descarregar la transcripció i els subtítols en:

- Text (.txt)
- Subtitles (.srt)
- JSON (.json)

L'editor permet modificar el contingut de la transcripció.

Aquest servei de Softcatalà és gratuït i, malgrat no tenir un editor de subtítols, sí que es pot fer alguns ajustos prèviament com, per exemple, el nombre de caràcters per línia. A més, ha ofert una transcripció de bona qualitat en els vídeos amb els quals hem provat. El procés és senzill i no requereix crear un perfil.



Serveis	<ul style="list-style-type: none"> • Transcripció automàtica: SÍ • Transcripció professional: NO • Subtítols automàtics: SÍ • Subtítols professionals: NO • Subtítols traduïts: NO
Principals formats d'exportació	<ul style="list-style-type: none"> • Word (.docx): NO • JSON (.json): SÍ • Text (.txt): SÍ • Excel (.csv): NO • Subtitles (.srt): SÍ • Subtitles (WebVTT .vtt): NO
Editor	<ul style="list-style-type: none"> • Permet editar el nombre i nom dels parlants: NO • Permet editar el contingut de la transcripció: SÍ • Té corrector d'ortografia: NO • Permet Subratllar: NO • Permet editar la durada dels subtítols en pantalla: SÍ • Permet editar l'estètica dels subtítols: NO
Idiomes	<ul style="list-style-type: none"> • Pàgina en català: SÍ • Pàgina en castellà: NO • Servei en català: SÍ • Servei en castellà: NO
Prova gratuïta	<ul style="list-style-type: none"> • Servei gratuït
Versió de pagament	<ul style="list-style-type: none"> • Servei gratuït

05. Quadre general resum de la funcionalita de les aplicacions



	Serveis	Editor	Idiomes	Preu
Amberscript	Transcripció automàtica Subtítols automàtics	Intuïtiu Senzill d'usar	Català Castellà	Demandar pressupost
Happy Scribe	Transcripció automàtica Subtítols automàtics	Molt complet Senzill d'usar	Català Castellà	1 hora = 12€
Omniscien	Transcripció automàtica Subtítols automàtics	No disponible sense contactar amb Omniscien	Català Castellà	En desenvolupament
Clideo	Subtítols automàtics	Complet	Català (beta) Castellà	1 mes = 9\$ 1 any = 72\$
FreeSubtitles AI	Transcripció automàtica Subtítols automàtics	No en té	Català Castellà	1 hora = 0,99\$ Menys = gratuït
Softcatalà	Transcripció automàtica Subtítols automàtics	No en té	Català	Gratuït

06. Valoració de la transcripció automàtica per vídeos



Plenari de l'Ajuntament de Xàtiva

APLICACIÓ	VALORACIÓ WER	VALORACIÓ NER
AMBERSRIPT	27.72%	97.58%
HAPPY SCRIBE	24.77%	98.01%
OMNISCIENT	25.66%	97.00%
CLIDEO	58.99%	88.09%
FREESUBTITLES AI	41.59%	97.05%
SOFTCATALÀ	31.26%	97.09%

El català a Perpinyà

APLICACIÓ	VALORACIÓ WER	VALORACIÓ NER
AMBERSRIPT	31.75%	95.05%
HAPPY SCRIBE	25.81%	98.15%
OMNISCIENT	50.74%	94.23%
CLIDEO	60.23%	88.96%
FREESUBTITLES AI	36.20%	96.86%
SOFTCATALÀ	26.11%	97.62%

Un jove pagès menorquí

APLICACIÓ	VALORACIÓ WER	VALORACIÓ NER
AMBERSRIPT	48.43%	90.81%
HAPPY SCRIBE	47.38%	96.40%
OMNISCIENT	68.64%	93.36%
CLIDEO	71.08%	88.19%
FREESUBTITLES AI	53.31%	96.93%
SOFTCATALÀ	45.99%	97.00%

Acte unitari "No als Parcs Eòlics"

APLICACIÓ	VALORACIÓ WER	VALORACIÓ NER
AMBERSRIPT	25.15%	97.58%
HAPPY SCRIBE	30.90%	98.09%
OMNISCIENT	100%	98.64%
CLIDEO	51.51%	95.99%
FREESUBTITLES AI	36.66%	98.71%
SOFTCATALÀ	29.39%	97.37%

Refranys del món de la pagesia

APLICACIÓ	VALORACIÓ WER	VALORACIÓ NER
AMBERSRIPT	43.92%	95.90%
HAPPY SCRIBE	47.80%	96.01%
OMNISCIENT	55.29%	96.15%
CLIDEO	62.79%	91.77%
FREESUBTITLES AI	45.99%	97.76%
SOFTCATALÀ	37.20%	98.20%

Inauguració ExpoEbre

APLICACIÓ	VALORACIÓ WER	VALORACIÓ NER
AMBERSRIPT	26.36%	97.36%
HAPPY SCRIBE	25.15%	97.93%
OMNISCIENT	34.27%	98.09%
CLIDEO	53.34%	93.54%
FREESUBTITLES AI	33.87%	97.80%
SOFTCATALÀ	37.52%	97.61%



Premis Federació ACAPPS

APLICACIÓ	VALORACIÓ WER	VALORACIÓ NER
AMBERSRIPT	28.00%	98.35%
HAPPY SCRIBE	28.00%	99.25%
OMNISCIENT	40.36%	98.73%
CLIDEO	60.36%	92.50%
FREESUBTITLES AI	33.45%	98.65
SOFTCATALÀ	23.63%	98.64%

Evolució de l'alumnat amb sordesa a Catalunya

APLICACIÓ	VALORACIÓ WER	VALORACIÓ NER
AMBERSRIPT	27.04%	95.49%
HAPPY SCRIBE	23.63%	99.06%
OMNISCIENT	37.72%	98.05%
CLIDEO	48.18%	94.32%
FREESUBTITLES AI	35.90%	98.36%
SOFTCATALÀ	28.63%	98.29%

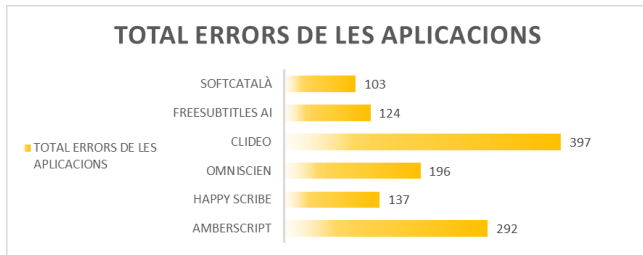
07. Anàlisi comparada de les aplicacions de transcripció i subtitulació



En aquest apartat podem valorar en conjunt el resultat i l'objectiu de tota l'anàlisi: quins resultats o valoracions ha tret cada aplicació de manera comparada amb la resta, de manera que podrem saber quines aplicacions funcionen millor en general i en concret també amb quines variants dialectals.

Comencem per una mirada general:

Gràfica 1. Aquest gràfic exposa els errors totals (lleus, normals i greus) dels 8 vídeos per a cada aplicació.

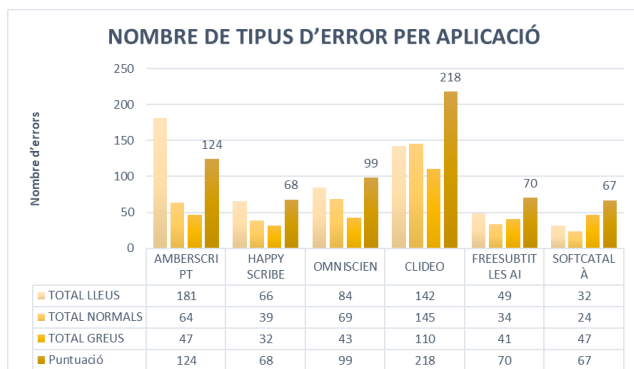


Elaboració pròpia

Podem veure que l'aplicació amb major nombre d'errors en les seves transcripcions és Clideo, seguida per Amberscript, i que les que menys errors tenen són Softcatalà i Freesubtitles AI.

No obstant això, una vegada vist el còmput total d'errors, que ens serveix per a l'avaluació quantitativa (WER) hem de preguntar-nos com es distribueixen els errors qualitativament a cada aplicació.

Gràfica 2. Aquest gràfic exposa la distribució dels errors, classificant-los segons les proves NER (lleus, normals i greus) per a cada aplicació.



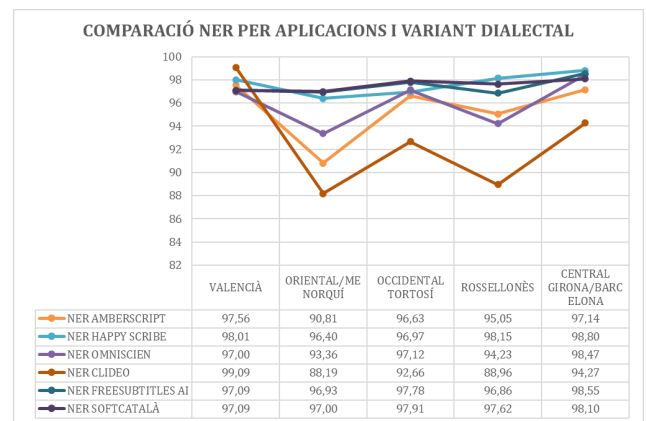
Elaboració pròpia

Ara podem veure que Clideo segueix en el primer lloc quant a errors normals i greus, però amberscript és l'aplicació amb major nombre d'errors lleus. També veiem que l'aplicació amb menor nombre d'errors greus és Happy Scribe, seguida per Omniscien, i que les aplicacions

amb menor nombre d'errors normals són Softcatalà i Freesubtitles AI respectivament.

La puntuació que obtenen les aplicacions fent servir la fórmula que pondera el valor dels errors (0'25 els lleus, 0'5 els normals i 1 els greus) ens mostra com en el conjunt de tots els vídeos, l'aplicació que millors resultats obté en la transcripció de vídeos en català és la de SoftCatalà seguida de Happy Scribe i ben a prop Freesubtitles AI, deixant les altres a força distància.

Gràfica 3. Aquesta gràfica compara els percentatges de les proves NER per a cada aplicació i variant dialectal analitzats.



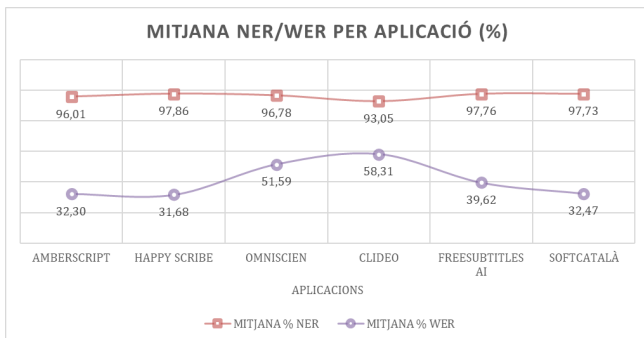
Elaboració pròpia

Podem veure que les aplicacions que es mostren més constants en les proves NER són Softcatalà i Freesubtitles AI, juntament amb Happy Scribe. La resta d'aplicacions mostren una major diferència en el NER segons el vídeo en el qual s'utilitzin. També es pot veure, per a cada vídeo, quina aplicació té un major percentatge en el NER (menor número o gravetat en els seus errors) i quina aplicació té un menor percentatge (major número o gravetat en els seus errors).

En general, les aplicacions segueixen una tònica més o menys constant dins de la seva pròpia línia, a excepció de les variants dialectals del català occidental/menorquí i rossellonès (vídeos "Un jove pagès menorquí" i "El català a Perpinyà", respectivament). La resta de variants (valencià, català occidental/tortosí i català central) mostren una major cohesió en totes les aplicacions.



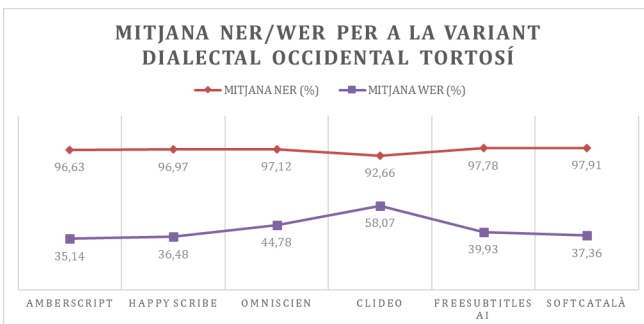
Gràfica 4. Aquesta gràfica compara les mitjanes de les proves WER i NER del conjunt de vídeos per a cada aplicació.



Elaboració pròpia

Quan comparem els resultats de les proves WER i NER, per al que s'ha de buscar l'aplicació amb menor percentatge WER i major percentatge NER. Ara veiem que l'opció òptima segons les mitjanes de tots els vídeos seria Happy Scribe. No obstant això, també convé tenir una visió més específica de cada vídeo per a saber si X aplicació és la més adequada per a X variant dialectal.

Gràfica 5. Aquesta gràfica mostra la mitjana de les proves NER i WER dels vídeos amb variant dialectal: català occidental/tortosí.

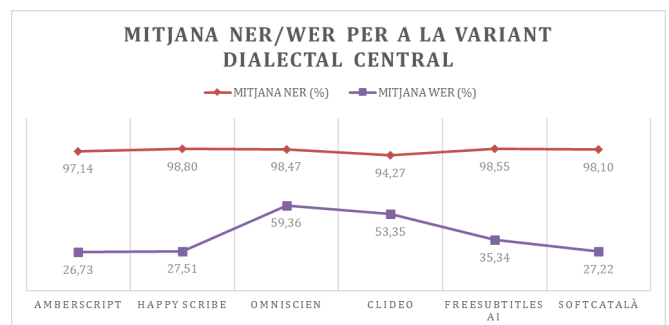


Elaboració pròpia

En aquest cas els vídeos utilitzats per a fer la mitjana són "Refrany" i "Expoebre", els dos amb bona qualitat d'àudio però amb alguns sorolls de fons que poden alterar-ho. En la gràfica, tant en la mitjana del NER com en la mitjana del WER, podem veure que les aplicacions estan molt igualades en general. El punt més alt del NER correspon a Softcatalà, mentre que el més òptim del WER, a AmberScript.

En el cas d'aquesta variant dialectal, al moment de triar quina aplicació usar, potser convé tenir en compte altres factors com la comoditat per a usar l'aplicació o web per la diversitat d'opcions que hi ha en cadascuna d'elles i dependent de les necessitats, ja que no hi ha una gran diferència entre una aplicació o una altra.

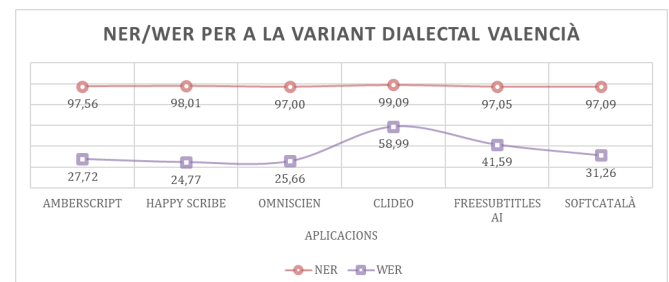
Gràfica 6. Aquesta gràfica mostra la mitjana de les proves NER i WER dels vídeos amb variant dialectal: català central.



Elaboració pròpia

En aquesta gràfica els vídeos utilitzats són "Acte unitari No als parcs eòlics", "Premis Federació ACAPPS" i "La situació de l'alumnat amb sordesa a Catalunya". El primer d'aquests vídeos té molta bona qualitat de vídeo i àudio, mentre que el segon té menor qualitat de vídeo i el tercer menor qualitat d'àudio. En aquest cas es pot veure una major variació entre les aplicacions, sobretot en la mitjana de les proves WER, encara que les proves NER es mantenen molt constants. El màxim del NER es troba en Happy Scribe, mentre que el mínim del WER en AmberScript.

Gràfica 7. Aquesta gràfica mostra les proves NER i WER dels vídeos amb variant dialectal: valencià.



Elaboració pròpia

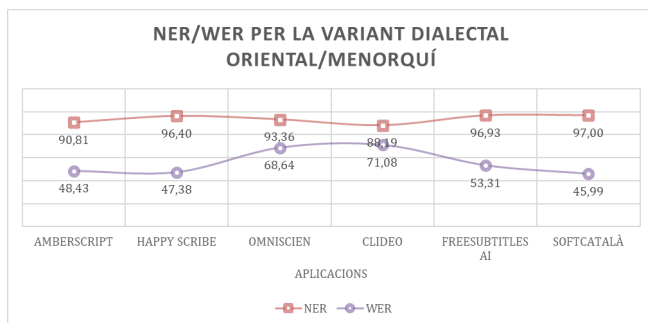
El vídeo que s'ha utilitzat és "Plenari de l'Ajuntament de Xàtiva", amb una qualitat d'àudio molt bona però elements com a màscares que poden afectar el reconeixement



de les paraules. En aquest cas Clideo és l'aplicació amb major percentatge en el NER, però és també la que té major percentatge en el WER. L'aplicació amb menor percentatge WER és Happy Scribe, sent la segona, no obstant això, a tenir major percentatge NER. De nou, el NER es manté constant per a totes les aplicacions, amb major variació en el WER.

El vídeo utilitzat per a aquesta gràfica és "El català a Perpinyà", que té bona qualitat d'àudio i pronunciació però molt soroll de fons. En aquest cas l'aplicació amb major percentatge NER i menor percentatge WER és Happy Scribe, seguida de prop per Softcatalà. Aquesta vegada la corba WER torna a veure's més pronunciada, encara que la línia dels percentatges NER es manté més plana.

Gràfica 8. Aquesta gràfica mostra les proves NER i WER dels vídeos amb variant dialectal: oriental/menorquí.



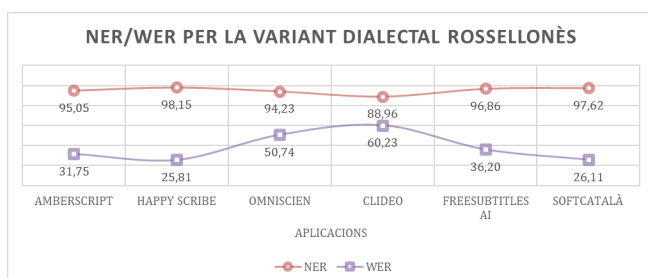
En aquest cas també es pot veure una menor igualtat entre les aplicacions en tots dos resultats.

Elaboració pròpia

El vídeo en aquest cas és "Un jove pagès menorquí". La qualitat d'àudio i vídeo no són bones i hi ha soroll constant. En aquest vídeo ja pot fer-se una diferenciació més clara de quina aplicació podria ser l'òptima. Softcatalà és l'aplicació amb menor percentatge WER i major percentatge NER. En els resultats del NER, Softcatalà va seguida per Freesubtitles AI, i en els resultats del WER per Happy Scribe.

El destacable d'aquesta gràfica és que ja hi ha una major diferència de resultats per aplicació tant en el NER com en el WER, deixant més clar quines aplicacions són més bones per al reconeixement d'aquesta variant dialectal i quines aplicacions poden resultar menys bones.

Gràfica 9. Aquesta gràfica mostra les proves NER i WER dels vídeos amb la variant dialectal: rossellonès.



08. Resultats Aplicacions i Variants Dialectals



	Softcatalà		Freesubtitles		Clideo	
	NER	WER	NER	WER	NER	WER
Valencià	97.09	31.26	97.05	41.59	99.09	58.99
Tortosí	97.91	37.36	97.78	39.93	92.66	58.07
Oriental	97.00	45.99	96.93	53.31	88.19	71.08
Rossellonès	97.62	26.11	96.86	36.20	88.96	60.23
Central	98.10	27.22	98.55	35.34	94.27	53.35

	Omniscien		Happy Scribe		Amberscript	
	NER	WER	NER	WER	NER	WER
Valencià	97.00	25.66	98.01	24.77	97.56	27.72
Tortosí	97.12	44.78	96.97	36.48	96.63	35.14
Oriental	93.36	68.64	96.40	47.38	90.31	48.43
Rossellonès	94.23	50.74	98.15	25.81	95.05	25.81
Central	98.47	59.36	98.80	27.51	97.14	26.73

Aplicacions

Softcatalà:

Manté una bona regularitat en la metodologia NER, aconseguint no baixar mai del 97%.

És l'aplicació que millor transcriu les variants dialectals del tortosí i del català oriental o balear.

Freesubtitles:

Tot i que no és l'aplicació que millor transcriu cap de les variants dialectals, aconsegueix puntuacions molt bones i una mitjana que en mètode NER no baixa de 96.86%.

Clideo:

Aconsegueix la millor puntuació en la transcripció del valencià en metodologia NER, però té valors en més d'una variant dialectal que baixen del 90%.

En metodologia WER obté valors molt elevats d'errades.

Omiscien

No aconsegueix esdevenir la millor aplicació transcrivint cap variant dialectal i la seva dispersió en mètode NER és elevada baixant en alguna variant del 90% i pujant en metodologia WER de 70 punts.

Happy Scribe

Aconsegueix esdevenir l'aplicació que millor transcriu el rossellonès i el català central i no baixa de 96.40 en mètode NER i no supera els 50 punts WER.

Amberscript

Amb una dispersió de resultats important en la metodologia NER, tot i no superar els 50 punts WER, no aconsegueix la primera posició a cap variant dialectal.

Variants dialectals:

Valencià

És una variant dialectal que sembla resultar de fàcil transcripció per al conjunt de les aplicacions, doncs cap d'elles baixa del 97% en mètode NER, tot i que en el WER els resultats són més dispersos.

Tortosí:

Excepte per l'aplicació Clideo també obté una bona regularitat en la qualitat de la transcripció del conjunt de les aplicacions, en què descomptant aquesta aplicació la menor puntuació NER que obté és del 96.63%.

Català Oriental:

Aquesta varietat dialectal no resulta tan fàcil per al conjunt de les aplicacions, aconseguint en algunes uns resultats relativament acceptables, superiors al 96%, però al mateix temps en unes altres quedant al voltant del 90%.

L'aplicació amb què millor es comporta és la de Softcatalà.

Rossellonès:

En una posició similar a la del català oriental, les diferents aplicacions aconsegueixen resultats molt diversos, amb menor dispersió que el cas del català oriental, però igualment significatius entre 98.15 i 88.96.

L'aplicació que millor la transcriu és Happy Scribe.

Català Central

És la variant dialectal en què més aplicacions aconsegueixen valoracions NER superiors al 97% (5 de 6) i la pitjor puntuació és del 94.27 %.

L'aplicació que millors resultats aconsegueix en la transcripció és Happy Scribe.

09. Conclusions



La primera qüestió a destacar és que una anàlisi d'aquestes característiques requereix una inversió de temps que fa difícil poder-la repetir sense comptar amb un finançament específic a l'activitat. Han estat més de 200 hores de recerca i elaboració de les metodologies i més de 300 hores de realització de les proves i redacció de l'informe. **En total un treball de 500 hores que es concreta en aquest informe.**

Una inversió de temps que hem considerat important de fer donades les expectatives que moltes persones amb sordesa tenen posades en aquestes aplicacions, així com també moltes empreses i institucions de cara a utilitzar-les per millorar l'accessibilitat de les seves instal·lacions i activitats.

En segon lloc, convé destacar que, tot i que aquestes aplicacions generen errades en la transcripció, tenen uns nivells sorprenents d'efectivitat. Sembla que **poden ser unes aplicacions perfectament vàlides per esdevenir un recurs d'accessibilitat allà on mai no posaríem una persona subtitulant**, com en un mostrador d'informació o en un espai on l'intercanvi d'informació no és determinant i les persones que han de prendre la paraula poden comprovar la precisió de la transcripció mentre parlen per assegurar la precisió de les dades que han de traslladar.

Malgrat això, generen **encara prou errors greus, dels que modifiquen el significat del que transcriuen** i dels que passen desapercibuts en el context, de manera que no resulten encara un recurs recomanable per a l'accessibilitat en contextos que requereixen precisió.

Pensem en Justícia, Salut, Ciència i Tecnologia com **els principals espais en què no resulten una mesura recomanable per substituir una subtitulació professional**. Fins i tot, en alguns espais de la docència poden resultar una eina que porti a equívocs, depenent segurament de l'aportació d'argot propi de la matèria en què s'utilitzi.

Cal tenir present que **no hem posat a prova aquestes aplicacions en situacions d'estrès de la qualitat de l'àudio o en situacions de dicció confusa, o en situacions de barreja de català i castellà**, o fins i tot d'anglès... D'aquesta manera cal prendre les dades que aporta aquest treball amb cautela.

Però tot i aquesta cautela, el resultat de l'anàlisi ens permet tenir una visió optimista del futur per

a les persones amb sordesa, doncs sembla que **la tecnologia continua oferint cada cop més i millors solucions per assegurar el dret a la informació i a la comunicació** a les persones amb sordesa.



hola@federacioacapps.org

**Famílies i persones amb sordesa que parlem i
treballem la inserció laboral, l'educació inclusiva,
l'accessibilitat a la comunicació**

C/Providència, núm. 42, 4t
08024 Barcelona



L'OBSERVATORI
ACCESSIBILITAT I COMUNICACIÓ
ORAL DE LES PERSONES AMB SORDESA