

ICA-PAAG. Guies breus

8

Títol:	La transcripció com a document d'arxiu
Autora:	Sandra Orihuela



International Council on Archives
Conseil International des Archives



Photographic and Audiovisual Archives Working Group
Groupe Archives Photographiques et Audiovisuelles

ICA – Photographic and Audiovisual Archives Group (PAAG)
Direcció: David Iglésias Franch
Autora: Sandra Orihuela
Publicació: Maig 2022

1. Introducció

Quan es parla de transcripcions en un arxiu, generalment es pensa en la transcripció que converteix un text manuscrit més o menys antic en un altre text adaptat a un sistema de notació més actual i intel·ligible. Les transcripcions que es tractaran en aquesta guia, però, són les que es poden fer a partir de documents sonors i audiovisuals: aquelles que copien en un text escrit allò que prèviament s'ha expressat oralment. Els reptes que presenta aquest tipus de transcripció tenen a veure amb el pas de la llengua oral a la llengua escrita, amb el canvi de suport i codi en documents en què tant la forma com el contingut són rellevants. Aquest pas d'una llengua a una altra pot plantejar diverses preguntes:

- Què és exactament una transcripció i quina importància té en un arxiu?
- Es pot considerar només una còpia d'un document original?
- Hi ha una manera universal i correcta de fer una transcripció?
- Com es poden traslladar a l'escriptura aquells elements propis de l'oralitat?
- La intervenció com a transcriptor pot tenir algun efecte en el document final?
- És possible substituir la transcripció manual per una transcripció automàtica?

A continuació, es donarà resposta a aquestes preguntes, analitzant què és una transcripció i què implica realitzar-la en el context d'una institució arxivística, quines pautes generals s'han de tenir en compte per desenvolupar-la i quines possibilitats obre l'ús de la tecnologia pel que fa a la transcripció automàtica.

2. Transcripció i les seves característiques

En general, es pot dir que la transcripció és un document, encara que no necessàriament d'arxiu, que és una còpia escrita d'un altre text o d'un discurs parlat. Aquesta còpia del discurs parlat pot estar feta a partir d'un enregistrament oral partint de notes preses en el moment en què s'escolta el discurs o en què es fa aquest mateix enregistrament. A més, la transcripció es pot fer utilitzant un sistema de caràcters o d'escriptura particular que permeti representar certs elements per als quals el sistema d'escriptura tradicional no és suficient, com podria ser el cas dels elements fonètics.

La transcripció com a document d'arxiu, però, introdueix altres qüestions a l'hora de tractar el document perquè l'ús que se'n fa és molt particular: el valor de la transcripció en aquest context rau en la seva utilització com a eina per part dels investigadors. Tot i que per a l'arxivística el document que té valor de prova i d'evidència és el document original, a la pràctica la transcripció és el principal document de consulta i, per tant, el que produeix realitat. En aquest sentit, és important que en fer les transcripcions es tingui en compte la voluntat de reflectir fidelment allò que ha succeït a l'enregistrament original.

Es pot considerar que la transcripció com a document d'arxiu:

- Ajuda els investigadors a avaluar de forma més ràpida si el document els resulta útil i rellevant, perquè requereix menys temps llegir un text que veure i escoltar una gravació.
- Facilita la conservació del suport original, que tendeix a ser inestable. En el cas dels documents més antics, per exemple, evita el desgast de les cintes.
- Facilita també la preservació de la informació, que està continguda en suports condicionats pels canvis constants a la tecnologia i que, per tant, són més fràgils.
- Facilita l'accés a aquesta informació perquè permet aclarir les ambigüitats i corregir els possibles errors del parlant (en dates o en localitzacions, per exemple). En aquest sentit, la transcripció pot ser més que una simple còpia, pot contenir noves capes d'informació que el document original no tenia.

En general, es considera que el document original és el document sonor i audiovisual, però la seva versió escrita pot ser molt més pràctica per estudiar i analitzar. Una transcripció corregida i indexada no només facilita la conservació, la preservació i l'accés, sinó que també pot suposar un valor afegit en aclarir els passatges ambigus i esmenar-ne els possibles errors en noms, llocs, dates, etc. Tot i això, cal tenir en compte que la transcripció també pot comportar algun inconvenient: principalment, la pèrdua d'una part de la informació present a l'enregistrament en no poder passar a escrit totes les característiques pròpies del discurs oral. Això és així perquè als documents sonors i audiovisuals no només trobem el llenguatge com a tal, sinó també altres elements presents en el context de l'enregistrament, com poden ser els sobreentesos entre els participants. La informació no està només en el llenguatge sinó també en les omissions, les mirades, l'entonació, les pauses i els gestos facials o corporals.

En definitiva, la informació que es desprèn de les fonts orals és rellevant però la manera com el parlant decideix fer-ho també ho és. La transcripció ha de ser capaç de passar, en la mesura que sigui possible, aquestes característiques del llenguatge oral i gestual al llenguatge escrit, que té un codi propi que resulta insuficient per recrear tot el que hi ha present a l'enregistrament original. En adquirir la forma escrita, el llenguatge perd l'espontaneïtat i ha de seguir unes regles pròpies. Un bon exemple d'això és el que passa amb les variacions a l'entonació, que en un text queden reduïdes a canvis en la puntuació – comes, punts, guions –, que no solen ser suficients per expressar la totalitat dels canvis de to.

Tanmateix, no es pot parlar d'una única manera de fer una transcripció: cada institució ha de decidir què és el que més interessa recollir per escrit del document sonor i audiovisual i, a partir d'aquí, escollir quin sistema de notació es vol utilitzar i quina informació es conservarà.

A grans trets, es pot parlar de tres maneres principals de fer transcripcions:

- 1) Una transcripció condensada, en què s'eliminen els elements propis de l'oralitat, com els dubtes, els titubejos i els falsos inicis del parlant. En aquest tipus de transcripció, l'atenció està posada principalment en la informació que es desprèn del discurs parlat. Aquesta és l'opció que se sol utilitzar a la publicació.

PORTELLI: How old was your daughter when she got married, the one you said had eleven children?

MOSSIE JOHNSON: Seventeen.

PORTELLI: So how many grandchildren, great-grandchildren, do you have?

JOHNSON: I'll have to get a pencil and a paper and figure out all night. I know I've got fourteen grandsons, and I don't know how many granddaughters. And I've got some great-grandchildren, and I think I've got three or four great-great-grandchildren. Lots of them I've not even seen, they live some ways away from here. Well, that's what the Bible said, it says fill the world up, so we filled the world up, we helped fill it up. But we had a good life.

Font: Portelli, Alessando. *They Say in Harlan County. An Oral History*. New York: Oxford University Press, Inc., 2011, p. 39.

- 2) Una transcripció que tracti de representar la complexitat del llenguatge oral mitjançant un sistema de notació que indiqui, gràcies a l'ús de diferents símbols, aspectes particulars com les diferències d'entonació, les emfatitzacions i les pauses. Aquesta forma és la que se sol utilitzar en disciplines com l'anàlisi de la conversa, en què es considera molt important com es diuen les coses i en què la transcripció respon a unes necessitats científiques molt determinades. Els símbols que es fan servir són, per tant, molts i molt complexos i no solen ser fàcils d'interpretar per a algú que no participi en aquestes disciplines. A més, com en qualsevol transcripció, el significat dels símbols no és universal, fet que encara dificulta més la comprensió.

Dean:	I ↑don't kno:w thè (·) full extent ↓'v it.↓ (0.7)
Dean:	°↓Uh:::eh° (0.9)
Nixon:	°I don'noo° 'bout anything else exch ^h e[pt
Dean: →	[I don't either in I: °w'd (h)als(h)o
	→ hhate tuh learn [some a'] these thi]ngs. ·hh-hh-hh·hh
Nixon:	[W e11] y a : h]
	(0.2)
Dean:	So ↑That's,hhhh that's <u>that</u> situation.

Font: Jefferson, Gail. «Glossary of transcript symbols with an introduction» a: *Conversation analysis. Studies from the first generation*. Amsterdam: John Benjamins Publishing Company, 2004, p. 15.

- 3) Una transcripció que combini les dues opcions anteriors, en què s'eliminin alguns dels elements orals, però se'n conserven d'altres per donar una idea de la "textura" del discurs parlat i de les particularitats de la persona gravada. En aquest tipus de transcripció s'utilitzen alguns símbols que no solen estar presents als textos més comuns però que sí que tenen certa tradició escrita i poden ser fàcilment reconeixibles pel lector. Aquesta opció sol ser la més recomanada per als arxius perquè permet incloure tant aspectes que tenen a veure amb el contingut com aquells aspectes de manera que poden resultar rellevants, sense arribar a la complexitat de la segona opció.

27		JUDGE CARTER: I was born in
28	00:02:20	Caryville, Florida. And when I was
29		very young, after my mother birthed
30		me, I was part of migration of family
31		out of the South to the North in the
32		Great Migration. My family was born
33		free, we were—I think maybe it was 60
34		years after the Civil War and so we
35		were a free black coming out of the
36		South in the Great Migration.
37		INTERVIEWER: And what year was that?
38	00:03:20	JUDGE CARTER: What?
39		INTERVIEWER: What year were you
40		born?
41		JUDGE CARTER: I was born in 1917.

Font: Carter, Robert L., Interviewee, Patricia Sullivan, and U.S Civil Rights History Project.
Robert L. Carter oral history interview conducted by Patricia Sullivan in New York, New York. 2010, p. 3.

Burnett:	Of course, in the 1950s, if you're talking about manufacturing, there are a lot of unionized places, and so there isn't a lot of surplus sort of left over. Then taxation rates are very high. So it might be a different picture, arguably, in 2018, but we'd have to get someone else to look at that again.
02-00:33:55 Tolley:	Yes, right.
Burnett:	That's another generation's job, right?
02-00:34:00 Tolley:	I think so. [Chuckles]
Burnett:	But I'm just getting a sense of the circulation of ideas. This emerging Chicago identity for that generation was a sense of being prepared to be surprised by the data, being prepared to be surprised by the empirical results of your research and testing assumptions or testing theory about how the economy works. That's something that you're exploring in your research, and that's something you took away from your dissertation.
02-00:34:40 Tolley:	Well, and from that whole experience. Schultz was an empiricist. I know I was—. He was running a workshop. [interview interruption]

Font: George S. Tolley, *George S. Tolley: From Agricultural to Resource, Urban, and Health Economics at the University of Chicago* conducted by Paul Burnett in 2018, Oral History Center, The Bancroft Library, University of California, Berkeley, 2020, p. 44.

En realitat, no importa si s'opta per una transcripció que elimini certs aspectes, com falsos començaments de frase, o si, per contra, es prefereix una transcripció que respecti al màxim possible tots els elements presents al document original. L'objectiu final és obtenir un text escrit que permeti entendre clarament el que ha passat i ha quedat recollit a l'enregistrament. El document resultant ha de ser alhora fidel i comprensible i ha de respondre a les necessitats de la institució. Sigui quina sigui la perspectiva que es prengui com a punt de partida, la transcripció ha de representar allò que s'escolta, o allò que s'escolta i es veu, de forma clara i amb respecte al que es troba al document sonor o audiovisual. Les decisions respecte a allò que roman o allò que s'elimina es veuen determinades per l'ús que se la donarà al document i pel plantejament que cada projecte

pren com a punt de partida. Si la institució considera que el més important és la informació que es desprèn del discurs, potser no cal fer una transcripció que reculli la majoria dels elements del llenguatge oral; si l'arxiu està més interessat a recrear també l'ambient on s'ha fet l'enregistrament i les particularitats lingüístiques dels parlants, la transcripció haurà de ser més detallada.

Totes aquestes decisions es poden considerar intervencions interpretatives del transcriptor que, per tant, tenen un efecte al document final. Cada decisió respecte a allò que es conserva o allò que s'elimina o respecte a quin símbol s'utilitza és una intervenció interpretativa, és a dir, una forma d'influir entre el document original i l'investigador que el consulta posteriorment. Aquestes intervencions es poden resumir com:

- Correccions i aclariments respecte a possibles errors a llocs, dates, noms, etc.
- La inclusió de certs elements implícits al diàleg que no queden fixats pel llenguatge, com els ja esmentats sobreentesos.
- La inclusió de gestos facials i corporals.
- L'eliminació o no de repeticions i titubejos.
- L'eliminació o no de les particularitats dialectals.
- L'ús dels signes de puntuació i la separació en paràgrafs.

La transcripció, per tant, no és neutra, sempre té un punt de vista, una perspectiva que interpreta i construeix el relat que finalment es llegirà i que està basada en decisions que, moltes vegades, es poden arribar a prendre de forma inconscient. La clau és actuar amb responsabilitat abans aquestes decisions interpretatives, que són inevitables, però també necessàries. En aquest sentit, és recomanable definir des del principi quina metodologia seguirà la institució per fer aquestes transcripcions i ser transparent amb l'usuari respecte a quines pautes se segueixen.

3. Pautes generals per al creació d'una guia de transcripció

Com ja s'ha comentat, no hi ha una forma universal i adequada de fer les transcripcions, encara que sí que hi ha guies que proposen una manera determinada i concreta. De la consulta d'algunes d'aquestes guies, se'n poden extreure unes pautes generals que permetin a cada institució desenvolupar el seu propi manual intern, a partir dels objectius i de les necessitats que cada arxiu hagi detectat.

El primer pas, per tant, és decidir en quina informació es té interès i què es mantindrà a la versió escrita de l'enregistrament: és a dir, decidir si només es vol reflectir el contingut de les paraules o si, al contrari, es vol també recrear els diferents elements que apareixen al document original (com el llenguatge no verbal, els gestos facials i corporals o anotacions sobre l'ambient en què s'ha desenvolupat la gravació).

A partir d'aquí, cal tenir en compte uns principis metodològics mínims o criteris bàsics:

- **L'elaboració de categories:** cal definir les categories que es transcriuran i assignar-los un símbol o grafia que correspondrà només a aquest cas. Per exemple, si es decideix que els comentaris del transcriptor (com les correccions) s'indicaran entre claudàtors, no s'hauria d'utilitzar el mateix símbol per a altres anotacions. El que és important no és quin símbol s'utilitza, sinó que aquest tingui un significat únic i clar.

- **Llegibilitat i economia:** la transcripció ha de ser llegible i, per tant, utilitzar grafies, símbols i altres recursos que siguin fàcils d'interpretar. És a dir, cal procurar no utilitzar una quantitat massa gran de símbols i utilitzar aquells amb què les persones ja estan familiaritzades perquè ja tenen una tradició escrita rellevant: cursiva, negreta, subratllat, parèntesis, claudàtors, etc.
- **Accessibilitat i robustesa:** els símbols han de ser fàcils d'identificar, aprendre i utilitzar. Tant la persona transscriptora com l'investigador han de ser capaços d'aprendre o intuir el significat dels símbols. El transcriptor, a més, ho ha de tenir fàcil per utilitzar-los: és a dir, els recursos han de ser fàcilment accessibles des del teclat de qualsevol ordinador.
- **Tractament computacional:** és aconsellable que la transcripció faciliti les cerques informatitzades. Això pot ser més o menys difícil de respectar depenent de la quantitat de particularitats de la parla que es vulguin afegir a la transcripció. Si es decideix que es recrearan gràficament variacions en la pronunciació, per exemple, es dificultarà la cerca automàtica: hi haurà certes paraules que, sent en realitat les mateixes, no es detectaran perquè la seva forma estarà expressada de manera diferent.

Tenint en compte aquests criteris bàsics, es pot passar a definir el guió amb els elements principals que ha de tenir la transcripció. Aquests elements es basen en diferents aspectes de format, edició, revisió i estil i formen part d'aquestes decisions interpretatives que cal prendre.

Format:

- Escollir tipus i mida de lletra, marges de pàgina, sagnies i espaiats. Generalment es proposa l'ús de Times New Roman, mida 12, amb doble interlineat i amb marges de pàgina d'uns 2,5 centímetres. A la pràctica, es pot utilitzar qualsevol font que sigui fàcilment llegible i un interlineat i uns marges que també resultin visualment agradables i facilitin la lectura.
- Definir què cal incloure a l'encapçalament de la transcripció: nom del projecte i de la institució que el realitza, nom complet de l'entrevistat i de l'entrevistador, lloc de l'entrevista, etc.
- Decidir com s'indiquen els canvis de cinta o les interrupcions per motius tècnics a l'enregistrament.
- Decidir si s'inclou un informe que acompanyi la transcripció i definir els elements que cal afegir-hi: comentaris respecte a l'ambient en què s'ha realitzat la gravació, apunts sobre particularitats de la parla, informació general sobre els participants, justificació del mètode de transcripció utilitzat.

Edició i revisió:

- Decidir com s'indiquen les correccions i d'altres anotacions del transcriptor: mitjançant símbols (corxets o parèntesis, per exemple) o en notes a part, a peu de pàgina o al final de la transcripció. Normalment s'escull una o altra opcions depenent de l'extensió del comentari del transcriptor: si es tracta només de completar un nom o corregir una data se sol anotar al mateix text de la transcripció; si, per contra, es tracta d'un aclariment una mica més extens, es pot anotar a peu de pàgina o al final de la transcripció.
- Decidir si s'indiquen els elements implícits al diàleg que no queden fixats pel llenguatge, com els sobreentesos o la ironia. En cas que es decideixi incloure'ls, definir com es farà.
- Decidir si s'indiquen els elements propis de l'oralitat: els canvis a l'entonació, les pauses, els gestos facials i corporals, etc. És important tenir en compte que alguns d'aquests aspectes només es poden anotar en el cas dels documents audiovisuals, on el so està acompanyat de la imatge.
- Si es decideix incloure'ls, cal definir també com es farà: si s'hi inclouen tots o no i quins símbols s'utilitzen. Una mostra d'això podria ser l'ús d'afegitons per part del participant en l'enregistrament: depenent del tipus de transcripció es pot decidir eliminar per complet aquests afegitons, mantenir-les totes (encara que això podria dificultar la lectura) o mantenir-ne només alguna (una de cada quatre, per exemple) per donar una idea de certes particularitats de la parla.
- Decidir si s'hi inclouen o s'eliminen els falsos inicis, els titubejos i les repeticions. En cas que s'inclouguin, definir en quin grau: és a dir, si es mantenen tots o només una part. Hi ha institucions que prefereixen eliminar-los del tot perquè entorpeixen la lectura; altres decideixen mantenir aquells que aporten certa informació, com podria ser un cert nerviosisme o dificultat per parlar d'alguns temes determinats.
- Definir com s'indiquen els fragments poc clars o intel·ligibles.
- Decidir si s'hi inclouen particularitats dialectals o de pronunciació. En cas que es vulguin incloure, establir uns límits clars i la forma que s'utilitzarà: per exemple, no és possible transcriure totes les variacions de pronunciació, però sí les més cridaneres o significatives; d'altra banda, certes particularitats dialectals es poden anotar, incloent-hi el significat per assegurar-ne la correcta interpretació.
- Decidir si s'inclouen o no certs aspectes del context de l'enregistrament: els sorolls de fons, els cops, les interrupcions brusques, etc. De la mateixa manera que amb els falsos inicis i els titubejos, aquests elements es poden eliminar completament si es considera que no aporten res, però alguns projectes els inclouen amb la voluntat de reflectir el més fidelment possible el context en què s'ha desenvolupat l'enregistrament.

Estil:

- Definir clarament quins símbols s'utilitzaran i què signifiquen: negreta, cursiva, subratllat, guions, comes, punts, parèntesis, claudàtors o qualsevol altre recurs que es vulgui utilitzar. Assignar a cada símbol una categoria determinada i, per tant, un significat únic.
- Establir uns principis mínims respecte a la separació en paràgrafs: si es fa o no i, en el cas que sí, quines pautes se segueixen per fer-los servir.
- Definir els criteris per a les notes a peu de pàgina i per a les notes al final de la transcripció.
- Definir com es fan servir les abreviatures i els acrònims.
- Definir com indicar les lletres: en el cas de lletres, per exemple.
- Definir com indicar els números: números romans, dates o hores.
- Decidir si es fan servir símbols o paraules per indicar percentatges, divises, etc.
- Definir com indicar les cites i les paràfrasis que els parlants puguin fer, respecte a obres o respecte a allò que altres persones han dit.

4. Altres guies

Les pautes anteriors s'han recollit a partir de la comparació de guies de transcripció ja existents, fetes des de la perspectiva de la història oral, i a partir d'altres textos plantejats des de la lingüística i des d'altres ciències socials.

Tots els articles consultats estan detallats a la bibliografia, però hi ha dos manuals principals que val la pena comentar més detalladament: la *Style Guide: A Quick Reference for Edit History Transcription Guide*, del Baylor University Institute for Oral History (BUIOH) i la *Oral History Transcription Guide* del Columbia University Center for Oral History Research (CCOHR). Totes dues guies són recents (2018) i són les que se solen recomanar des de les associacions d'història oral d parla anglesa ja que constitueixen una bona introducció a les transcripcions.

El manual del Baylor University Institute for Oral History (BUIOH) està estructurat de manera força simple: està compost d'una breu introducció, una nota respecte al format, regles generals d'ortografia i correcció de la transcripció i la guia d'estil ordenada alfabèticament. En canvi, la guia del Columbia University Center for Oral History Research (CCOHR) divideix la informació de manera més detallada, a més d'incloure a l'inici l'índex dels continguts, que facilita la cerca, i apèndixs amb plantilles i exemples.

Les dues guies parlen pràcticament del mateix, però la del CCOHR ho fa desglossant més els temes: introducció, procés, format, revisió, edició, estil, apèndixs i bibliografia. Encara que es pot considerar que la guia del CCOHR és més completa, en general, totes dues coincideixen en la majoria de pautes per fer les transcripcions, tant en la forma com en el contingut. És a dir, les dues institucions utilitzen, en gran part, els mateixos símbols per indicar-ne els diferents aspectes i també segueixen uns principis similars respecte a què eliminar i què no.

Com hem comentat, totes dues són una bona introducció i poden ser un bon punt de partida a l'hora de desenvolupar una guia pròpia sense haver de començar de zero. Les dues guies són bones opcions a partir de les quals cal fer adaptacions pròpies, especialment si s'opta per fer transcripcions que quedin a mig camí entre la senzillesa de la publicació i la complexitat de l'anàlisi del discurs.

Tot i això, és important tenir en compte que aquestes guies són en anglès i que, per tant, molts dels consells que ofereixen fan referència a les ambigüitats i els problemes que pot donar aquest idioma. En cas de fer transcripcions en altres llengües, caldrà adaptar les pautes a les particularitats concretes del nou idioma.

5. Transcripció automàtica

Fer una transcripció des de zero pot ser un procés llarg i costós i, per aquest motiu, en els darrers anys s'està plantejant la possible utilització de la transcripció automàtica com a substitutiu del manual, per tal de poder disminuir el temps que el transcriptor li té que dedicar a cada gravació.

La transcripció automàtica és possible gràcies a la tecnologia de reconeixement de veu, o *automatic speech recognition* (ASR), que permet que un programa informàtic processi la veu humana i la tradueixi a un text escrit. Perquè aquest processament sigui possible, el programari ha de ser capaç de reconèixer els diferents aspectes del llenguatge, com la gramàtica, la sintaxi i l'estructura, així com els elements propis de les senyals de veu i la seva composició. El sistema de ASR, per tant, no només capta el que es diu, sinó que ho interpreta i ho transcriu: divideix els sons en parts que pot reconèixer, analitzar i oferir la paraula escrita que es correspon amb el que s'ha analitzat. Aquesta tecnologia, que només és possible desenvolupar unint diferents disciplines (com la lingüística, les matemàtiques i l'estadística), es programa i es millora a mesura que se li proporciona més dades i es perfeccionen els algorismes que utilitza per analitzar el contingut. És dir, quants més dades anteriors se li han introduït més probabilitats de fer suposicions correctes en relació als patrons de la parla o a les paraules relacionades. La tecnologia de ASR, per tant, no només analitza el que pot escoltar, sinó que també proporciona solucions per a fragments que no s'han captat correctament, basant-se en algorismes i dades anteriors que ja coneix. És per això que el sistema es pot millorar i adaptar: per exemple, es pot augmentar la seva precisió si hi ha paraules sobre un tema específic que apareix amb freqüència.

Els sistemes d'ASR funcionen gràcies a diferents components que permeten dividir les tasques que té que desenvolupar: reconèixer l'entrada de veu, extreure les seves característiques, descodificar i donar una sortida a la transcripció utilitzant les paraules adequades. La decodificació és possible gràcies a models acústics, diccionaris de pronunciació i models lingüístics dels idiomes a transcriure, que s'han introduït prèviament per tal que el programari pugui reconèixer-los. Per altra part, els programaris en el camp de la intel·ligència artificial no treballen només reconeixent dades sinó també mitjançant l'«aprenentatge supervisat», que implica que el programa és capaç de donar una sortida determinada a una entrada que mai havia vist, o escoltat, perquè sí que s'ha trobat amb altres models de parells d'entrada-sortida semblants.

El problema en la transcripció automàtica és que el llenguatge és molt variable, fins i tot en el cas de persones que parlen un mateix idioma, i això pot augmentar la taxa d'error, o *word error rate* (WER), que s'utilitza per avaluar la seva precisió. Les particularitats en la pronunciació o l'accent, així com el to, el volum, la velocitat i els possibles sorolls de fons són determinants i poden provocar que aquesta taxa d'error augmenti considerablement. En el cas dels sorolls de fons, es pot «entrenar» al programari perquè sigui capaç de discernir que no formen part del llenguatge humà i els filtre o elimini. En el cas de les particularitats de la parla, implica un gran esforç per part de les persones que programen el sistema d'ASR, que han de recollir la major varietat possible de parles perquè pugui reconèixer-les i aprendre'n.

El perfeccionament dels sistemes d'ASR depèn en gran mesura de la quantitat i la qualitat de les dades que se li puguin proporcionar inicialment perquè pugui minimitzar els errors. En aquest sentit, els avenços en aquests tipus de programari estan molt determinats per aquelles llengües que tenen un nombre més gran de parlants o que es troben més presents en tots els entorns. Les llengües minoritàries, per tant, es veuen relegades a un segon pla i els progressos respecte a aquestes depenen dels diferents grups de recerca que s'interessen per oferir solucions i disminuir la taxa d'error de certs idiomes. El desenvolupament d'aquest tipus de tecnologia, en conseqüència, és desigual i és possible o no en funció de la llengua de què es parteixi, no perquè sigui tècnicament més difícil sinó perquè el desenvolupament de la tecnologia sol estar determinat pel mercat.

Com ja s'ha comentat, els sistemes d'ASR se solen veure limitats per certs inconvenients que, en el cas dels documents sonors i audiovisuals d'un arxiu, se solen veure agreujats:

- La qualitat general de les gravacions, que pot ser molt variable perquè l'origen dels documents també és variable.
- Possibles sorolls de fons, que, per la mateixa raó, són més o menys presents segons on s'ha fet l'enregistrament i amb quins mitjans.
- Les particularitats en la pronunciació i les particularitats dialectals, augmentades per la gran varietat de persones que participen en documents sonors i audiovisuals.
- El to de veu, el volum i la velocitat, que també són molt personals i depenen de cada parlant.
- Informació de context, que el sistema ASR no pot reconèixer i que pot ser rellevant per a la correcta interpretació de determinats passatges.

Per fer front a aquests problemes, normalment es proposa personalitzar i adaptar un sistema d'ASR a les necessitats de la institució i dels seus documents.

- Si es tracta de millorar la qualitat de l'àudio, hi ha dues possibilitats d'adaptació del model acústic: 1) modificar el senyal d'enregistrament de manera que es redueixi i compensi el soroll (*Speech Enhancement*); 2) entrenar el sistema mostrant-li àudios amb una gran varietat de sorolls, perquè pugui reconèixer-los en altres futurs àudios i no considerar-los com una part analitzable de la parla (*Multi-Condition Training*).

- Si es tracta de resoldre problemes lingüístics, cal centrar-se en l'adaptació del model lingüístic i la formació del sistema ASR mitjançant l'ús d'àudios amb una gran varietat de tipus de parlants que mostrin diferències en: sexe, edat, lloc de procedència, etc.

Perquè el sistema sigui precís, és necessari que la formació es faci a partir de la transcripció manual d'un gran nombre d'hores d'enregistrament. És a dir, s'han de dedicar molts recursos humans i econòmics a una tasca que no sempre dona resultats satisfactoris. Això és així perquè els resultats varien molt en funció del tipus de document que es vulgui transcriure: si es tracta de transcriure vídeos i àudios d'un canal de televisió, per exemple, és més probable que la personalització del sistema ASR doni bons resultats, perquè els parlants tendeixen a parlar en una llengua més estàndard; D'altra banda, si es pretén transcriure automàticament documents de disciplines com la història oral (com entrevistes o històries de vida), la precisió del sistema ASR es reduirà i caldrà fer moltes més correccions manuals.

En termes generals, la transcripció automàtica encara està en procés de desenvolupament. Els diferents factors acústics i lingüístics impedeixen que la taxa d'error es redueixi al màxim: el soroll, les converses superposades, la diversitat de pronunciacions o la varietat de dialectes són reptes que encara s'estan abordant. D'altra banda, hi ha certs elements presents en les gravacions que només pot ser registrats per un transcriptor i que no es tenen en compte en els sistemes ASR: aspectes relacionats amb l'entorn en què s'ha produït l'enregistrament, comentaris sobre l'entonació, gestos facials i corporals o aquells moments en què cal la interpretació d'una persona capaç de contextualitzar. Malgrat això, la necessitat d'aquests últims elements està molt determinada pel tipus de transcripció que es vol dur a terme.

De la mateixa manera que amb la guia de transcripció manual, cada institució ha de valorar, en funció del tipus de documents que custòdia, quins són els avantatges i els inconvenients de la transcripció automàtica i si li pot resultar realment útil. En el cas que es vulgui una transcripció merament informativa, en què només constin les paraules dites, la transcripció automàtica pot ser útil una vegada s'ha entrenat adequadament al programari: es pot utilitzar com a primera aproximació i corregir després manualment els errors que hi hagi pogut tenir. D'altra banda, també cal valorar si els esforços i els recursos, econòmics i humans, que requereix la personalització o adaptació dels sistemes d'ASR valen realment la pena en cada cas concret. L'entrenament del programari només és possible gràcies a una gran quantitat de dades que pot ser útil en col·leccions amb hores d'enregistrament abundants, però si la institució posseeix pocs documents, potser el més adequat sigui la transcripció manual.

En conclusió, la substitució completa de la transcripció manual per una transcripció automàtica és una cosa en què s'està treballant. Cada institució ha de valorar, en funció de la seva realitat i les seves necessitats, si val la pena dedicar temps i recursos a un tipus de tecnologia que no és del tot precís i que encara està en desenvolupament.

Bibliografia

ALLEN, S. E. «Resisting the Editorial Ego: Editing Oral History», a *Oral History Review* 10 (1982): 33-45.

BISSCHOP, Chantal. «Self-reflection: academic world versus cultural heritage». Ponència presentada al "XVIII Congreso de la IOHA: Poder y democracia: las múltiples voces de la historia oral", Barcelona, 9-12 de juliol de 2014.

MOORE, Kate. «Perversión de la palabra: la función de las transcripciones en la historia oral», en *Historia, antropología y fuentes orales*, n. 18 (1997): 13-24.

MOSS, William W. y Peter C. Mazikana. *Los archivos, la historia y la tradición orales: un estudio del RAMP*. París: Unesco, 1986.

PORTELLI, Alessandro. «Living Voices: The Oral History Interview as Dialogue and Experience», a *The Oral History Review* (2018): 1-10.

PORTELLI, Alessandro. «The Peculiarities of Oral History», a *History Workshop*, no. 12 (1981): 96-107.

SHOPES, Linda. «Transcribing Oral History in the Digital Age», en *Oral History in the Digital Age* (eds. Doug Boyd, Steve Cohen, Brad Rakerd, y Dean Rehberger). Washington, D.C.: Institute of Museum and Library Services, 2012. <http://ohda.matrix.msu.edu/2012/06/transcribing-oral-history-in-the-digital-age/> [Consulta: abril 2021]

THOMPSON, Paul. *The Voice of the Past*, Oxford: Oxford University Press, 2000.

WILLEMS, Dominique. «Lenguaje escrito y lenguaje oral», a *Historia y Fuente Oral* n. 1 (1989): 97-105.

WILMSEN, Carl. «For the record: Editing and the production of meaning in oral history», a *Oral History Review* 28, n. 1 (2001): 65-85.

Comparació de guies de transcripció i directrius generals

Baylor University Institute for Oral History (BUIOH). *Style Guide: A Quick Reference for Editing Oral History Transcripts*, 2018. <https://www.baylor.edu/old/2021-02/oralhistory/doc.php/14142.pdf> [Consulta: març 2021]

BASSI FOLLARI, Javier Ernesto. «El código de transcripción de Gail Jefferson: adaptación para las ciencias sociales» en *Quaderns de psicologia* 17, n. 1 (2015): 39-62. <http://dx.doi.org/10.5565/rev/qpsicologia.1252>

BLADAS MARTÍ, Òscar. *Manual de transcripció del discurs oral: materials de treball*, Barcelona: Universitat de Barcelona, Secció de Lingüística Catalana, Departament de Filologia Catalana, 2009. <http://hdl.handle.net/2445/106301>

Columbia University Center for Oral History Research (CCOHR). *Oral History Transcription Guide*, 2018.

<https://static1.squarespace.com/static/575a10ba27d4bd5d7300a207/t/5a5f95a2652dea3ce32aacc/1516213672572/CCOHR+Transcript+Style+Guide+2018.pdf> [Consulta: març 2021]

ÚBEDA QUERALT, Lluís. «El tratamiento archivístico y documental de las fuentes orales» *Història Oral, revista de la Assoc. Brasileira de Història Oral* 2, n. 2 (2004) 77-91.

Transcripció automàtica

REF, Michael, Joachim Köhler y Almut Leh. «Improved transcription and indexing of oral history interviews for digital humanities research», a *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Paris: European Language Resources Association (ELRA), 2018, 3124-3131.

REF, Michael, Oliver Walter, Christoph Schmidt, Sven Behnke, Joachim Köhler. «Multi-Stage Cross-Lingual Acoustic Model Adaption for Robust Speech Recognition in Real-World Applications - A Case Study on German Oral History Interviews», a *Proceedings of the 12th Language Resources and Evaluation Conference*. Paris: European Language Resources Association (ELRA), 2020, 6354-6362.

GÓMEZ, Lluís, Marçal Rusiñol, Ali Furkan Biten y Dimostenis Karatzas. «Subtitulació automàtica d'imatges. Estat de l'art i limitacions en el context arxivístic». Ponència presentada a Imatge i Recerca: 15es Jornades Antoni Varés, Girona, 21-24 de novembre, 2018.

GORISCH, Jan, Michael Gref y Thomas Schmidt. «Using Automatic Speech Recognition in Spoken Corpus Curation» a *Proceedings of the 12th Language Resources and Evaluation Conference*. Paris: European Language Resources Association (ELRA), 2020, 6423-6428.

OARD, Doug. "Can Automatic Speech Recognition Replace Manual Transcription?" a *Oral History in the Digital Age* (eds. Doug Boyd, Steve Cohen, Brad Rakerd, y Dean Rehberger). Washington, D.C.: Institute of Museum and Library Services, 2012. <http://ohda.matrix.msu.edu/2012/06/automatic-speech-recognition> [Consulta: abril 2021]