

SOFTWARE DI TRASCRIZIONE FORENSE: UNA PERICOLOSA DELEGA DI POTERE

Iacopo Benevieri



1. La prova in una sillaba.

Nel corso degli interrogatori di garanzia nell'inchiesta che ha recentemente coinvolto il presidente della Regione Liguria uno degli indagati avrebbe fornito una risposta recentemente caduta al centro dell'attenzione mediatica. In particolare, dalla trascrizione della fonoregistrazione del suo interrogatorio risulterebbe che il predetto avrebbe reso la seguente dichiarazione: il presidente della Regione *"ci chiedeva finanziamenti illeciti"* [\[1\]](#).

Tale trascrizione è stata contestata dallo stesso dichiarante, il quale ha precisato di ricordare *“con assoluta certezza di aver pronunciato le parole ‘finanziamenti leciti’ ”*. A seguito di tale contestazione, il pubblico ministero avrebbe formulato alla Giudice per le Indagini Preliminari una richiesta di disporre perizia fonica, presumibilmente finalizzata ad accertare se l'indagato ebbe o meno a pronunciare la sillaba “il” come prefisso della parola “leciti”, ovvero come continuum fonico tra la parola “finanziamenti” e la parola “leciti”.

La questione appare ancor più rilevante se si considera come la trascrizione in esame non sarebbe stata condotta da una persona fisica bensì da un software specificatamente programmato per convertire l'oralità del parlato in trascrizioni. Non parrebbe esser stato d'ausilio a fugare i dubbi neanche il verbale redatto in forma riassuntiva dal cancelliere, nel quale non si fa menzione del termine “finanziamenti”, né “leciti” né “illeciti”.

A prescindere dal diverso (rilevante) riflesso probatorio che nella fattispecie concreta assumerebbe la particella “il” nell'una e nell'altra ipotesi, la vicenda consente di riflettere nuovamente su una questione spesso trascurata, quella relativa all'assenza di idonee garanzie nella trascrizione forense del parlato, sia esso captato nel corso di una intercettazione, sia esso registrato durante un esame dibattimentale, un interrogatorio o un'assunzione di sommarie informazioni.

2. La trascrizione umana tra aspettative e proiezioni.

Nell'ipotesi in cui la trascrizione in esame fosse stata effettuata da una persona fisica, numerosi sono i fenomeni che avrebbero potuto indurre la percezione della parola “illeciti” anziché “leciti”. Ritenere che il soggetto che ascolta un dialogo cui non partecipa sia “terzo estraneo”, impermeabile dunque al processo comunicativo che ascolta, è convincimento smentito dagli studi di linguistica forense, che hanno sottolineato al contrario il ruolo di partecipazione attiva dell'ascoltatore, direttamente inserito o meno nell'interazione. Una delle più chiare evidenze di tale ruolo attivo di chi trascrive è costituita dal complesso fenomeno della previsione dell'ascolto.

Poiché il parlato è solitamente “ipoarticolato”, vale a dire caratterizzato da fenomeni di coarticolazione fonica, da una maggiore fluidità esecutiva, con frequenti eclissi di intere sillabe di una parola (come potrebbe esser avvenuto nella percezione e trascrizione delle parole “finanziamenti(l)leciti”), ebbene chi trascrive si troverà spesso a dover percepire segnali vocali incompleti con la conseguenza che tenderà a far ricorso ad altri elementi, soprattutto di tipo extralinguistico, e ad attuare aspettative e previsioni.

Nella percezione di una frase interviene un calcolo di previsione da parte di chi ascolta, il quale, basandosi sui dati in proprio possesso (per esempio, conoscenza degli interlocutori, del loro ruolo processuale, del contesto nel quale l'interazione è inserita, del tema oggetto del dialogo e, quindi, dell'oggetto del processo o delle indagini), nonché sull'ampiezza del proprio vocabolario, formula un elenco di possibili parole o frasi che potranno seguire quella percepita[2].

Si può quindi affermare come l'ascoltatore focalizzerà la propria attenzione sul dato acustico, attivando al contempo le proprie personali conoscenze, le proprie previsioni, così da ricostruire le parti di segnale vocale non sufficientemente chiare. E' questa la ragione per la quale sarebbe opportuno che chi trascrive non fosse a conoscenza dell'oggetto del procedimento o dell'atto processuale in corso, per evitare proprio la proiezione di personali attese sull'ascolto[3].

Garanzie maggiori, tuttavia, non vengono fornite nell'ipotesi di trascrizione forense effettuata tramite software di riconoscimento automatico del parlato.

3. La trascrizione automatica: una pericolosa delega di potere

Software di riconoscimento vocale e di trascrizione del linguaggio naturale sono ormai frequentemente utilizzati nell'ambito della ricerca sulla linguistica applicata (cd. ASR, "Automatic Speech Recognition"). Questo tipo di ricerca richiede spesso la trascrizione di dati parlati per procedere a una loro analisi. I continui progressi nell'elaborazione di questi applicativi stanno rendendo la trascrizione automatica dei testi parlati una possibilità interessante per chi conduce ricerca. Tale prassi si sta diffondendo anche nelle trascrizioni forensi, suscitando più di qualche perplessità sull'accuratezza dei risultati.

Recenti studi hanno evidenziato come i software più aggiornati, pur accedendo a modelli linguistici estremamente ampi, non forniscano idonee garanzie di un'esatta trascrizione del parlato. E' stato rilevato come la qualità della trascrizione possa dipendere, innanzitutto, dal diverso formato del file audio immesso nel sistema (wav o mp3), con significative differenze trascrittive prodotte dallo stesso software[4]. In generale, è stato segnalato un potenziale tasso di errore molto elevato, ritenuto "*del tutto inappropriato per i contesti forensi*" ("*not appropriate for forensic contexts*") in quanto "*una trascrizione in cui solo il 50% è corretto non è utilizzabile*" ("*a transcription in which only 50% is correct is not useable*").

Inoltre la mancanza di informazioni e di piena trasparenza sull'esatta architettura di tali software, e la

conseguente mancanza di certezza sulle cause dei diversi livelli di prestazioni, è un'altra ragione per cui i sistemi ASR non sono attualmente ritenuti adatti all'ambito forense[5].

A ciò si aggiunga che problemi relativi al riconoscimento di elementi lessicali particolarmente infrequenti, al rumore di fondo, alla distanza variabile dei parlanti dal microfono, alla sovrapposizione del parlato condizionano il corretto riconoscimento del parlato[6]. In particolare tali sistemi sembrano soffrire del medesimo fenomeno evidenziato in relazione al trascrittore umano, quello cioè della previsione delle parole: molti applicativi ASR, infatti, non hanno restituito prestazioni affidabili nella trascrizione di parole a bassa prevedibilità ("*low predictability items*"). Sotto questo profilo sarebbe interessante comprendere se la parola "leciti", affiancata a "finanziamenti", è attestata dall' algoritmo come parola a bassa prevedibilità rispetto alla parola "illeciti", con la conseguenza che la prima opzione risulterebbe ad alto rischio di essere scartata. La trascrizione che viene effettuata, dunque, si baserebbe su un modello probabilistico di ciò che è stato detto.

D'altronde, com'è stato osservato[7], uno dei sistemi più sofisticati di riconoscimento vocale e di trascrizione si basa su 680.000 ore di audio, tratte da programmi televisivi, udienze di tribunali e dibattiti politici, già accompagnate dalle trascrizioni ufficiali così da consentire al sistema di procedere a un lavoro di apprendimento migliore nel trovare regole comuni applicabili. Fornendo al sistema migliaia di registrazioni in cui persone diverse, in luoghi diversi, pronunciano, per esempio, una determinata parola, il sistema crea un modello probabilistico, con la conseguenza che *"questi sistemi possono essere influenzati negativamente da un pregiudizio culturale"*, determinato dall' *"insieme di dati di TV, podcast e notiziari utilizzati per addestrare il sistema"*. Di fatto, se nei dati vocali inseriti la parola "finanziamenti" è associata, nella maggioranza dei casi, a condotte illecite, il sistema riconoscerà come più probabile l'aggettivo "illeciti" rispetto a "leciti".

Per tale motivo è stato ritenuto auspicabile che la trascrizione automatica venga sempre sottoposta a revisione da parte del trascrittore umano, in assenza del quale si corre il rischio di concedere una pericolosa delega a tali sistemi nella rappresentazione del dato orale[8].

4. Conclusioni: l'urgente necessità di un intervento normativo

Possiamo affermare che la trascrizione è un'attività che costituisce l'esito di un' articolata analisi, per la quale sono necessarie competenze specialistiche altamente qualificate nell'ambito di più discipline (quali, per citarne solo alcune, l'analisi del suono, la fonetica articolatoria, la sociofonetica, la dialettologia, la sociolinguistica). D'altronde, poiché la trascrizione è un atto di interpretazione e di rappresentazione, è

sempre anche un atto di potere. Da ciò non può che derivare la necessità di un'attenta e urgente regolamentazione della materia, sia per quanto attiene alla istituzione di un albo nazionale dei trascrittori e alla previsione dei titoli e dei percorsi formativi necessari per accedervi, sia per quanto attiene alla elaborazione di raccomandazioni operative per una trascrizione che risponda a criteri di garanzia e attendibilità, sia per quanto attiene alla regolamentazione dell'uso degli applicativi di trascrizione automatica.

Se, infatti, la trascrizione costituisce l'espressione di una forma di "potere" nella misura in cui è atto di interpretazione e di rappresentazione, non può che concludersi come l'assenza di una qualsiasi forma di regolamentazione e la supina consegna di tale attività a sistemi "intelligenti" rischia di sottrarre l'esercizio di tale potere a qualsiasi garanzia.

[1]

https://www.corriere.it/cronache/24_maggio_26/caso-toti-finanziamenti-illeciti-no-leciti-audio-contestato-spi-nelli-jr-in-aula-cf63373d-f892-44f3-8e45-0148cf067x1k.shtml.

[2] L. Romito, *La misura dell'intelligibilità e il rapporto segnale-rumore. Atti del convegno "AISV (Associazione Italiana di Scienze della Voce)"*, Padova, 2004, p. 539 ss.

[3] F. Orletti-L. Mariottini, *Forensic Comunication in Theory and Practice, New Castle upon Tyne, Cambridge Scholars Publishing*, 2017, p. 17.

[4] D. Loakes, *Automatic speech recognition and the transcription of indistinct forensic audio: how do the new generation of systems fare?* *Frontiers in communications*, 14.2.2024; Harrington, L., and V. Hughes, "Automatic speech recognition: system variability within a sociolinguistically homogeneous group of speakers" in *Proceedings of the 20th International Congress of Phonetic Sciences Guarant International*, 2023.

[5] D. Loakes, *cit.*

[6] O'Shaughnessy, D., *Trends and developments in automatic speech recognition research*, in *Computer Speech & Language*, vol. 83, 2024.

[7] D. Turner, *Automated transcription and some risks of machine interpretation*, in Quirkos, 25.4.2023.

[8] D. Loakes, *Does automatic speech recognition (ASR) have a role in the transcription of indistinct covert recordings for forensic purposes?* in *Frontiers in communications*, 2022.