Automatic Speech Recognition: appunti sul Tutorial CLARIN (Arezzo, 2019)

Carlotta De Sanctis, Jessica Matteo, Marcello Nuccio, Chiara Paris, Caterina Pesce, Chiara Scarselletti, Patrick Urru

Premessa

Fra il 14 e il 16 febbraio 2019 si è tenuto ad Arezzo il quindicesimo convegno nazionale dell'Associazione Italiana Scienze della Voce (AISV) dall'interessante titolo *Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale*. I lavori del convegno hanno visto la partecipazione di studiosi di diverse discipline, provenienti da tutta Europa. È stata l'occasione per sperimentare un incontro e un confronto tra linguisti, ingegneri, informatici, storici, archivisti, antropologi che lavorano, ognuno nel proprio campo, con le fonti sonore. Gli interventi delle prime due giornate si sono focalizzati sull'analisi delle fonti orali dal punto di vista di discipline differenti, in particolare la linguistica e la storia orale. Nell'ultima mattinata, invece, in una tavola rotonda che univa tutte le anime del convegno, con i presidenti di associazioni e istituti, è stata affrontata da diversi punti di vista la questione della conservazione delle fonti sonore, di particolare interesse per AISO.

Questo breve resoconto raccoglie gli spunti principali del proficuo dialogo e confronto tra i partecipanti della nostra associazione al convegno, iniziato durante le giornate di studio e proseguito anche successivamente.

CLARIN e la trascrizione automatica

AISO ha preso parte a un *tutorial* organizzato da CLARIN, un'infrastruttura di ricerca internazionale (*European Research Infrastructure for Language Resources and Technology*) che si occupa della realizzazione e della condivisione di software *open source*, che aveva come scopo quello di fornire alcuni strumenti tecnici per la trascrizione automatica delle interviste. Il *tutorial*, condotto da Christoph Draxler, ingegnere informatico dell'Università di Monaco, è stato importante per iniziare a riflettere e porsi delle domande, ancora aperte, sulla concreta possibilità per gli oralisti di utilizzare strumenti informatici nella storia orale e, in particolare, su questo particolare *step* del processo, che è stato oggetto anche di uno specifico intervento alla recente scuola AISO milanese.

Il *tutorial* è stato strutturato in sei brevi sezioni, a cui purtroppo non è stata affiancata un'esercitazione pratica ed è stato questo forse il principale punto di debolezza della lezione:

1. registrazione dell'intervista con il programma open source SpeechRecorder https://www.bas.uni-muenchen.de/forschung/Bas/software/speechrecorder/

2. trascrizione automatica tramite ASR https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/ASR
3. trascrizione manuale usando OCRA https://www.phonetik.uni-muenchen.de/apps/octra/octra/login
4. segmentazione automatica con WebMAUS https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic
5. multifunzionale Pipeline https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Pipeline
6. analisi dei dati sonori attraverso Emu http://ips-lmu.github.io/EMU.html

Le caratteristiche che un file audio deve avere per essere tradotto in parole dai servizi BAS sono:

- una durata di 10 minuti, un formato di file non compresso (preferendo per esempio WAVE a MP3)
- assenza di disturbi ambientali nel luogo di registrazione
- una lingua supportata dal software di trascrizione (nel caso di quella italiana non è disponibile nessuna varietà dialettale).

La buona qualità della registrazione sonora è data principalmente dal tipo di strumentazione messa in campo: registratore Zoom H4, microfoni esterni per ogni parlante, computer per il controllo dei volumi.

Considerate queste condizioni, il software di trascrizione automatica in che misura può essere utile allo storico orale?

L'ambientazione ideale proposta dal tutorial sembra cozzare con la componente più "intuitiva" (o volendo "militante") della storia orale come pratica di ricerca. L'idea di dover ingessare il momento dell'intervista in un prontuario di norme da seguire genera chiaramente delle perplessità. Entra in contrasto con il principio di valorizzazione (e preservazione) dell'incontro tra intervistato e intervistatore inteso come momento dialogico, tanto che, ad esempio, il software non riconosce la sovrapposizione delle diverse voci e va persa quindi la dimensione del dialogo. D'altro canto, però, è stato messo in luce come, in alcuni casi, la presenza di strumenti tecnologici e l'utilizzo di maggiori accortezze nell'allestimento del set dell'intervista possono rivelarsi utili ad aumentare "la serietà" di ciò che si sta facendo, conferendo maggiore professionalità al lavoro. Tuttavia, va sottolineato che gli storici orali già da tempo si pongono il problema della misura in cui la presenza di strumenti più impattanti del semplice registratore (per esempio la videocamera) compromettano la spontaneità dell'intervista.

Altro limite emerso: l'impossibilità di trascrivere i dialetti. Questo potrebbe però essere solamente un limite attuale dei servizi offerti da BAS, in quanto la comunità scientifica è stata chiamata a rispondere con una collaborazione volta ad aggiungere altre varietà linguistiche. Ognuno di noi potrebbe infatti raccogliere interviste di qualità, nella varietà dialettale di propria competenza, e prestare delle trascrizioni letterali e

fonetiche, affinché quella chiave linguistica sia aggiunta a quelle già disponibili.

Un ulteriore punto debole riguarda la destinazione e l'uso delle registrazioni inserite nella piattaforma di cui CLARIN si serve per trascrivere: non sembra chiaro infatti se e in che misura le interviste sono protette e se il contenuto può essere accessibile e usato anche da terzi. Alla luce delle "buone pratiche" dell'Aiso, che attribuiscono al ricercatore la responsabilità della tutela del testimone e delle sue parole, questa mancanza di chiarezza risulta problematica e limitante.

Esperimento autogestito

Alla lezione è seguito poi un esperimento autogestito volto a contravvenire alla maggior parte dei consigli operativi di Christoph Draxler, al fine di comprendere fino a che punto un'intervista svolta in un ambiente meno "neutro" da un punto di vista sonoro potesse essere supportata dal sistema. A pochi metri da una sala gremita di gente (quindi in un ambiente per nulla ovattato e piuttosto caotico) abbiamo registrato un file audio di pochi minuti con un registratore di qualità medio bassa, senza microfoni, senza controllo dei volumi e con l'unica accortezza di parlare correttamente in italiano. Abbiamo simulato un dialogo a due voci analogamente a quanto succede nel corso di una qualsiasi intervista. Dopo alcuni tentativi non andati a buon fine e almeno cinque minuti di elaborazione (per un file di circa due minuti), il software ASR ci ha dato risposta affermativa: la trascrizione era avvenuta con successo.
Il risultato ha assunto la forma di un flusso indistinto di parole, non differenziate per ciascun parlante, ma, sorprendentemente, piuttosto fedele alle parole pronunciate; la trascrizione risultava sbagliata in particolare nei momenti in cui le voci andavano a sovrapporsi, caso in cui eliminava completamente il dialogo.

Considerazioni:

- È possibile utilizzare anche un normale registratore di fascia medio bassa. Forse non è così necessario investire in una strumentazione tecnologica di alto livello rischiando anche di compromettere l'informalità dell'intervista.
- Il problema della lingua: bisogna necessariamente raccogliere file audio in una delle lingue supportate dal sistema e ad oggi le varianti dialettali sono pochissime. Questo limita enormemente l'utilizzo del software per interviste più marcatamente connotate dal dialetto.
- La necessità di segmentare i file più pesanti in microframmenti di non più di 10 minuti rende il procedimento molto macchinoso.
- L'autenticazione del singolo utente deve essere supportata da istituzioni già iscritte all'infrastruttura CLARIN, dunque per poter utilizzare ASR bisogna essere muniti di credenziali di accesso. Al momento dell'esperimento abbiamo potuto utilizzare le nostre personali credenziali di Ca' Foscari; per coloro che non hanno legami con enti partenariati è possibile fare direttamente a CLARIN una richiesta individuale per ottenere delle credenziali di accesso, ma non è chiaro quanto questo canale sia

accessibile e di quanto tempo necessiti. (Forse Aiso potrebbe munirsi di credenziali specifiche di accesso da fornire agli associati?)

- La presenza e il controllo dello storico orale è ineludibile. Le trascrizioni automatiche assumono la forma di flussi indistinti di parole, non sempre corrispondenti a quelle effettivamente pronunciate: è necessaria quindi una correzione post trascrizione automatica, la differenziazione dei soggetti parlanti e l'aggiunta della punteggiatura.
- Resta una grande perplessità rispetto alla conservazione dei file caricati sul software: non è dato sapere in che misura i providers (CLST, IBM, Google, EML), a cui CLARIN si appoggia per la processazione dei file, trattengono e utilizzano le informazioni contenute nei file caricati. Questo evidentemente entra in contraddizione con le normative sulla privacy e l'imperativo di tutelare i dati sensibili.

Propositi

ASR non è l'unica possibilità esistente oggi per ottenere delle trascrizioni automatiche. La mattina di sabato 16 febbraio, nell'ultima sessione di interventi abbiamo ascoltato la relazione di Maria Palmerini sui servizi offerti dall'azienda Cedat 85: una società privata che sviluppa tecnologia per tradurre l'audio in testo (http://www.cedat85.com/) e si occupa principalmente di svolgere commissioni di trascrizioni automatiche. Cedat 85 però offre servizi a pagamento e si rivolge esclusivamente ad una clientela, a differenza di CLARIN che invece è un'infrastruttura europea volta alla creazione di strumenti gratuiti disponibili per la comunità dei ricercatori.

Con CLARIN, ciascuno di noi, autonomamente, può prendere dimestichezza con il software: valutare il tempo necessario a percorrere tutta la filiera dei passaggi, dalla segmentazione dei file più pesanti alla trascrizione di ogni intervista. Il sistema è tuttavia in fase di crescita e sperimentazione. In questa direzione, un altro possibile intervento (più lungimirante) potrebbe essere quello di impegnarsi a collaborare per aggiungere nuove chiavi linguistiche: raccogliere interviste di buona qualità e, lavorando in équipe con dei linguisti, fornire a BAS delle trascrizioni fonetiche per varianti dialettali (il che sarebbe anche in linea con il senso dell'iniziativa promossa da AISV, ossia di creare dei ponti di collegamento tra le comunità di pratica dei linguisti e quella degli storici orali). Ad oggi, quindi, il contributo che gli storici orali potrebbero dare a CLARIN non consiste tanto nell'uso dello strumento, quanto piuttosto nell'impegno attivo per migliorarlo e metterlo a punto.