**Processo di codifica: scelte adottate**

**<teiHeader>:**

All’interno del <teiHeader> ho incluso i seguenti elementi, nell’ordine presentato:

* <fileDesc>, contenente tutte le informazioni sul file elettronico. Al suo interno, oltre agli elementi figli obbligatori <titleStmt>, <publicationStmt> e <sourceDesc>, ho deciso di inserire anche gli elementi <editionStmt>, per raggruppare tutte le informazioni sull’edizione digitale che verrà realizzata, <extent>, per specificare, anche solo in maniera approssimativa, le dimensioni delle registrazioni a partire dalle quali il testo è stato trascritto, e <notesStmt>, in cui ho inserito alcune informazioni aggiuntive sul progetto. Per quanto riguarda il contenuto di <publicationStmt>, esso non è ancora stato completamente definito (avrei infatti bisogno di consultarmi con lei per sapere cosa poter eventualmente inserire alla voce <distributor> e/o <authority>). All’interno dell’elemento <sourceDesc>, essendo appunto il testo tratto da alcune registrazioni, ho inserito l’elemento <recordingStmt> (modulo ‘spoken’, capitolo 8, ‘Transcription of speech’) e poi 4 elementi <recording>, uno per ciascuna registrazione, accompagnati da tre attributi: @type, @dur e @xml:id. Dentro ogni elemento <recording> sono stati infine inseriti un elemento <p> per fornire una breve descrizione in prosa del file audio, e un elemento <date> con la data della registrazione.
* <encodingDesc>, con tutte le informazioni sul progetto e sulle scelte editoriali adottate (solo l’elemento <tagsDecl> è ancora da definire)
* <profileDesc>, in cui, oltre alla data di creazione e ai linguaggi presenti all’interno del testo, ho inserito un elemento <textClass>, contenente a sua volta un elemento <keywords> per specificare una serie di parole chiave indicative sul tema affrontato all’interno del testo, e poi gli elementi <particDesc> e <settingDesc> (modulo ‘corpus’, capitolo 15, ‘Language Corpora’), il primo per la descrizione accurata dei partecipanti all’intervista e il secondo per quella dell’ambiente in cui essa è stata svolta.

**<text>:**

All’interno dell’elemento <body>, figlio dell’elemento <text>, ho pensato di inserire 4 elementi <div>, ognuno corrispondente a una delle registrazioni (al momento gli elementi <div> sono solo due perché ho codificato solo le prime due registrazioni). Ogni <div> punta alla relativa registrazione attraverso l’attributo @decls di cui è provvisto.

Per la codifica dei vari fenomeni, ho fatto principalmente riferimento al modulo ‘spoken’ e agli elementi da esso messi a disposizione, nello specifico:

* Ho suddiviso la conversazione in *utterances*, ognuna marcata da un elemento <u>. Non essendoci una regola specifica per stabilire quando un *utterance* termini e quando ne inizi una nuova, ho solitamente preso come punti di riferimento pause più lunghe della norma o cambi di interlocutore; nei casi più complessi, come indicato nelle linee guida TEI, per la loro suddivisione ho paragonato le *utterances* ai paragrafi di un testo scritto. Inoltre, ho sempre inserito all’interno di ogni elemento <u> un attributo @who per specificare chi avesse pronunciato la frase e, quando era possibile, anche un attributo @trans, per indicare il modo in cui le varie battute si succedevano le une alle altre.
* Ho marcato tutti i fenomeni vocali ma non lessicali - come tosse, starnuti e suoni non lessicali come ‘ehhh’, ‘ahhh’, ‘mmmh’, ‘uhm’ - con l’elemento <vocal> immediatamente seguito dall’elemento <desc>, così da fornire una descrizione più dettagliata del fenomeno esaminato.
* Tutti i fenomeni accidentali - come i rumori di sottofondo - sono stati marcati dall’apposito elemento <incident> anch’esso seguito dall’elemento <desc> per una descrizione più precisa e approfondita
* I cambi di intonazione, voce, volume, tempo e ritmo sono stati marcati attraverso l’elemento vuoto <shift/> sempre provvisto dei due attributi @feature e @new
* Ho marcato tutte le pause con l’elemento vuoto <pause/>, sempre accompagnato dall’attributo @type per specificare la durata approssimativa della pausa stessa
* Poiché molte volte le frasi delle due interlocutrici si sovrapponevano, oppure quello che diceva l’una interrompeva ciò che stava dicendo l’altra, o, ancora, alcuni rumori di sottofondo duravano per parecchio tempo, nei casi che lo richiedevano ho fatto uso anche dell’elemento vuoto <anchor/>. Ho inserito, nello specifico, l’elemento <anchor/> all’interno delle varie *utternaces* fornendogli di volta in volta un @xml:id. Successivamente inserivo, a seconda del caso, un elemento <u> o un elemento <incident> con un attributo @start che puntasse all’elemento <anchor/> grazie al suo @xml:id (nei casi in cui era possibile, gli elementi <u> e <incident> sono stati provvisti anche di un attributo @end, così da specificare il punto preciso in cui la frase o il rumore terminava). In questo modo mi è stato possibile sincronizzare al meglio il susseguirsi delle varie battute, anche quando queste tendevano a essere pronunciate contemporaneamente.
* Oltre a questi elementi appartenenti al modulo ‘spoken’, ho utilizzato anche altri elementi più comuni e diffusi in molti altri testi, come:
  + <emph>, per marcare parole pronunciate in maniera particolare e con più enfasi
  + <foreign>, per le parole straniere (sempre accompagnato infatti dall’attributo @xml:lang)
  + <persName> seguito eventualmente da <forename> e/o <surname> per marcare i nomi di persona (sempre accompagnato da un attributo @ref per poter poi puntare al corrispondente elemento <person> che inserirò al termine della codifica in un’apposita <listPerson>)
  + <placeName>, per la codifica dei nomi di luogo (anch’esso accompagnato da un attributo @ref per puntare poi all’elemento <place> corrispondente contenuto all’interno di una <listPlace> che inserirò in appendice)
  + <choice> con gli elementi <sic> e <corr>, per la codifica degli errori presenti accompagnati dalla corrispondente forma corretta, oppure con gli elementi <orig> e <reg>, per marcare le forme dialettali accompagnate dalla corrispondente forma standard
  + <gap/>, per la codifica dei passi incomprensibili o inudibili
  + <unclear>, per la codifica dei passaggi poco chiari (sempre accompagnato dai due attributi @cert e @resp)
  + <cit>, per la codifica delle citazioni presenti all’interno del testo
  + <quote>, per marcare i discorsi diretti riportati nel testo
  + <title>, utilizzato per marcare i titoli di alcune opere letterarie citate nel testo