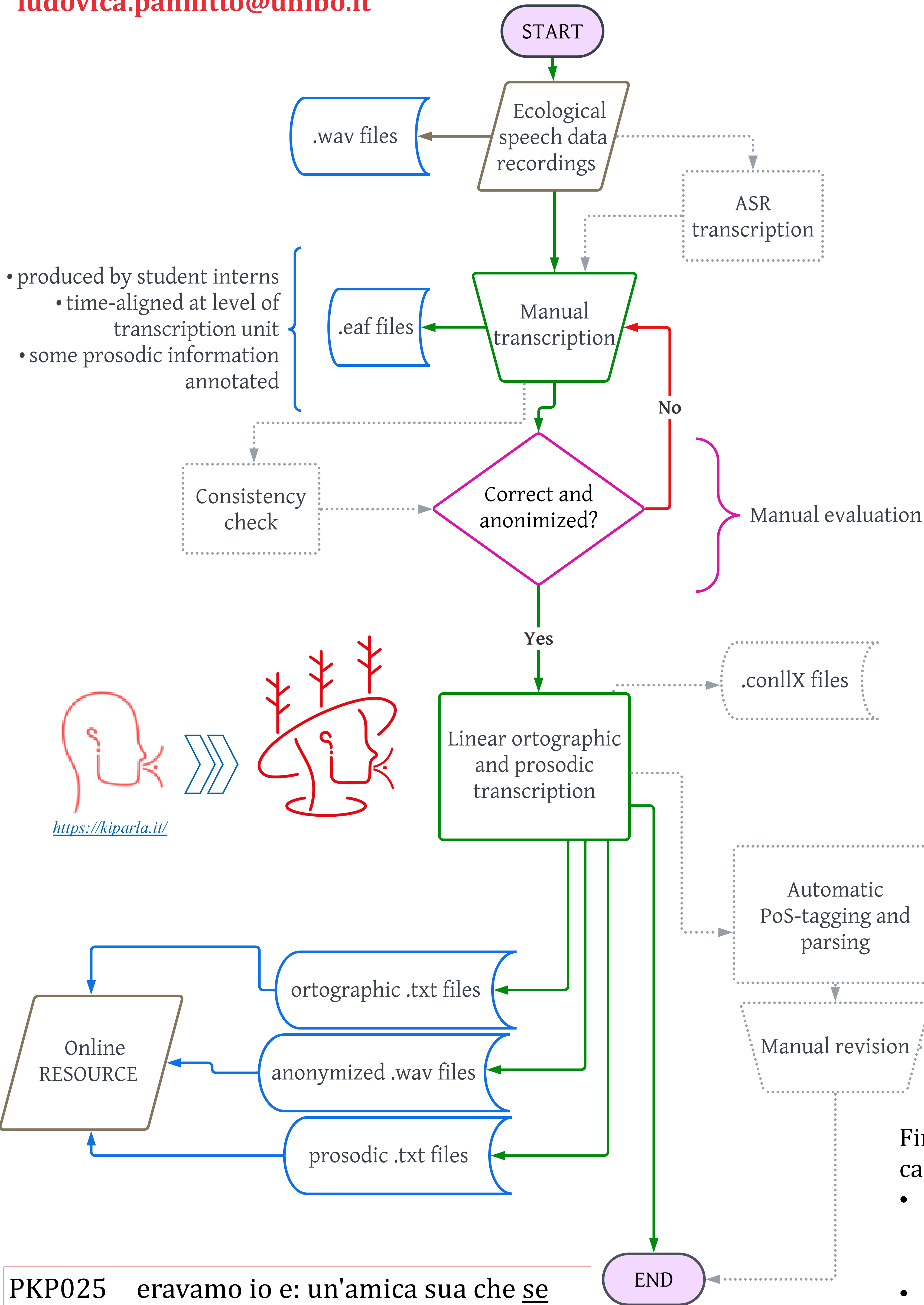


The KIPARLA forest of spoken Italian(s)

Ludovica Pannitto

Experimental Laboratory - University of Bologna

ludovica.pannitto@unibo.it



- produced by student interns
- time-aligned at level of transcription unit
- some prosodic information annotated

Overview of UD Spoken Language Treebanks

	Beja	Cantonese	Chinese	Chukchi	ParisStories	Rhapsodie	Frisian-Dutch	Naja	Komi-Zyrian	Norwegian	Slovenian	Turkish-German	KIPARLA forest
Sound file ID	yes	no	no	yes	yes	no	no	no	yes	no	no	no	yes
Text-sound alignment	yes	no	no	yes	no	no	no	no	yes	no	no	no	yes
Speaker ID	no	no	no	no	yes	yes	yes	no	yes	yes	no	no	yes
Language variety	no	no	no	no	no	no	yes	yes	no	yes	no	yes	yes
Standard orthography	no	no	yes	yes	yes	yes	yes	no	no	yes	yes	yes	yes
Capitalization	no	no	no	yes	no	no	no	yes	no	no	no	yes	no
Pronunciation	yes	no	no	yes	no	no	no	no	no	no	yes	no	no
Speaker overlap	no	no	no	no	no	yes	no	no	no	no	yes	no	yes
Punctuation	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	no	yes	no
Incomplete words	no	no	no	yes	yes	yes	no	no	yes	yes	yes	yes	yes
Fillers	no	no	no	no	yes	yes	yes	no	yes	yes	yes	yes	yes
Silent pauses	yes	no	no	no	no	no	no	no	yes	yes	yes	no	yes
Incidents	no	no	no	no	no	no	no	no	no	no	yes	no	yes

Adapted from Dobrovolicj [2022]

First task on Part-of-Speech tagging at EVALITA 2020 evaluation campaign (KIPoS, Bosco et al. [2020])

- The dataset provided for the KIPoS task included 30k tokens whose tagging was **manually revised** after processing with UDPipe;
- The task highlights difficulties in morphologically and syntactically tagging unconstrained speech contexts. This includes identifying **mode-specific phenomena** and handling **non-standard and creative linguistic forms**;
- **Specific PoS labels** were introduced for tokens from Italo-Romance dialects and languages other than Italian (labeled as DIA and LIA, respectively).

PKP025 eravamo io e: un'amica sua che se chia:ma: **mar~** odd[**io**]
there were me a:nd a friend of hers her na:me is mar~ oh my god
 PKP026 [**que**]lla magrolina?
[the] thin one?

PKP094 però lì quando ha detto così dico **xx** cazzo (questi) stanno male
but there when they said so I mean xx shit these people are mad
 PKP094 **m:h** dico **ehm** no no era così: te **vola~**, babbo diceva lascia stare **~ascia**
 stare
m:h I mean ehm no no it was like thi:s it fl~, papa said leave it be ~eave it be

PKP025 #o n'je fo niente ((ride))
#oh I'm not doing anything ((laughs))