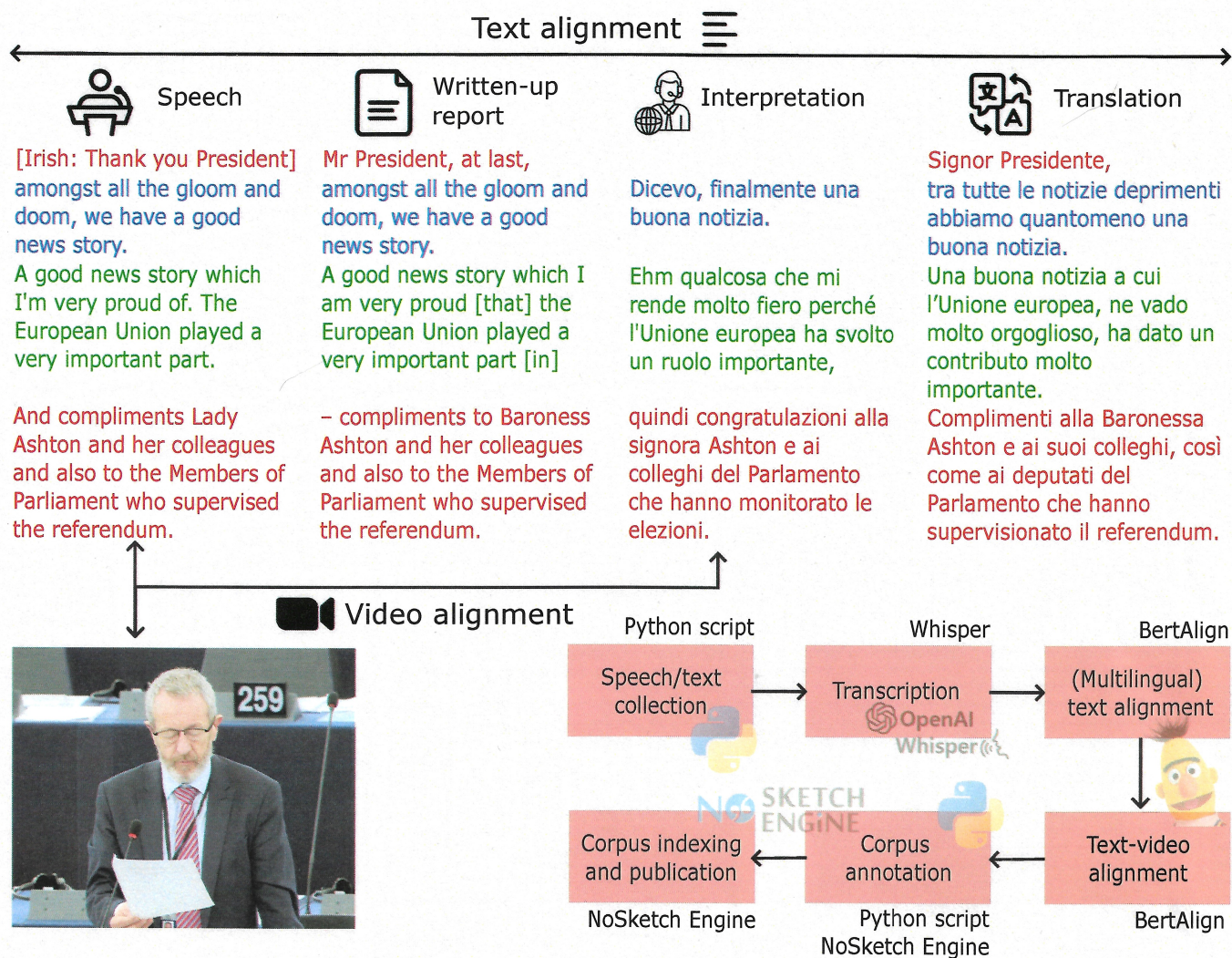


Constructing a Multimodal, Multilingual Translation and Interpreting Corpus: A Modular Pipeline and an Evaluation of ASR for Verbatim Transcription



A. Fedotova, A. Ferraresi, M. Miličević Petrović and **A. Barrón-Cedeño**
DIT, Università di Bologna, Forlì, Italy
[alice.fedotova2, adriano.ferraresi, maja.milicevic2, a.barron]@unibo.it

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI FORLÌ



OpenAI
Whisper WER

Model	English	Italian	French	Slovenian
Small	0.212	0.219	0.162	0.463
Small-FT	0.180	0.201	-	-
Medium	0.196	0.173	0.213	0.327
Large-v2	0.194	0.152	0.118	0.262

Speech	English	Italian
Native	0.104	0.131
Non-native	0.110	-
Interpreted	0.222	0.188



<https://corpora.dipintra.it/eptic/>

Tokens by language
Sources Targets

Language	Spoken	Written	Interpr.	Transl.
English	43,138	41,047	55,109	58,651
French	35,648	34,063	31,935	35,566
Italian	21,208	20,646	27,329	31,816
Polish	9,458	9,193	-	-
Slovene	-	-	19,717	22,476
German	-	-	18,258	19,822
Finnish	-	-	11,624	12,045