



ReTeRom proiectul 1: CoBiLiRo (Corpus Bimodal de Limbă Română)

Echipa:

CS III dr. Anca Bibiri, drd. Șerban Boghiu, prof. dr. Dan Cristea, ec. Lucian Gâdioi, CS II dr. Daniela Gîfu, CS III dr. Mihaela Onofrei, st. master Felix Cristian Pericică, lector dr. Ionuț Pistol, drd. Andrei Scutelnicu, conf. dr. Diana Trandabăț

Universitatea „Alexandru Ioan Cuza” din Iași


Proiectul CoBiLiRo

- etapa a treia -

2020

Obiective:

- ▶ noile resurse sunt disponibile pe platforma Cobiliro;
- ▶ completarea metadatelor asociate, alinieri și adnotări manuale și semiautomate ale corpusului voce/text;
- ▶ statistici ale corpusului bimodal;
- ▶ valorificarea lor atât în cadrul platformei (statistici și instrumente integrate) precum și în afara platformei, propunând o serie de proiecte ce utilizează tehnologiile dezvoltate în cadrul proiectelor partenere;
- ▶ activitate de diseminare: participarea la manifestări științifice și prezența/diseminarea în mass-media.



Activitatea 3.1: Augmentarea corpusului prin noi înregistrări vocale care dublează texte existente în corpusul CoRoLa

Anca-Diana Bibiri

Mihaela Onofrei

Dan Cristea

Universitatea „Alexandru Ioan Cuza” din Iași

Augmentarea corpusului vocal prin noi înregistrări

► **Indicatori de realizare:**

Corpusul voce/text este augmentat cu noi înregistrări care dubleză texte depuse în corpusul CoRoLa. Colecția este accesibilă prin infrastructura creată.

Augmentarea corpusului vocal prin noi înregistrări

- ▶ Resursele încărcate pe platforma CoBiLiRo însumează 28:09:44 de înregistrări sonore, cu transcrierile textuale corespunzătoare (etapa a III-a - 2020).

Repartizarea pe colecții:

- ▶ *Ghici cine vine la cină* - 10:23:28;
- ▶ *Alma Mater Iasiensis* - 12:10:16;
- ▶ *SoRoEs* - 2:49:20;
- ▶ *Povești*: 2:46:40.

Augmentarea corpusului vocal prin noi înregistrări

Transcrierile sunt neuniforme în privința apropierii de text:

- ▶ unele prelucrate deja de editori (pentru îndepărtarea unor repetiții, pauze, bâlbâieli, inerente vorbirii libere) în vederea tipăririi textelor (colecțiile de interviuri);
- ▶ altele urmărind destul de aproape idiosincraziile din vorbirea liberă;
- ▶ o altă categorie: înregistrări vocale care au rezultat ulterior existenței textelor corespunzătoare, prin citirea cu glas tare a textelor (*read speech*), acestea urmărind îndeaproape textul.

Augmentarea corpusului vocal prin noi înregistrări

- ▶ calitatea înregistrărilor sonore este diversă: de la înregistrări realizate în condiții profesionale (microfon performant, cameră anecoidă), până la înregistrări în condiții naturale (pe stradă, cu zgomot sau muzică în surdină etc.);
- ▶ eterogenitatea resurselor bimodale le asigură utilitatea experimentelor de antrenare a tehnologiei de recunoaștere a vocii, cât și în cea de sinteză vocală, resurse care să reproducă o cât mai diversă stare de lucruri.

Augmentarea corpusului vocal prin noi înregistrări

Colecțiile încărcate pe platforma CoBiLiRo respectă formatul standardizat al fișierelor vorbire-text relativ la metadata (care conțin informații header și alinieri).

Augmentarea corpusului vocal prin noi înregistrări

Exemplu I: colecția *Alma Mater Iassiensis*

```
<?xml version="1.0" encoding="utf-8"?>
```

```
<tei>
```

```
<teiHeader Description="" Collection="Iacob" Keywords="" Language="ro"  
Contributor="serban.boghiu@gmail.com" Distribution="-"> <speechSection  
SpeechCreator="UAIC" AcousticMedia="" Duration="55:01" SamplingFrequency="0"  
Resolution="0" RecordDate="" RecordTime="" Equipment="" Broadcast=""  
SpeechFile="" SpeechFileType="3" SpeechSegmentation="0"> <speaker  
SpeakerName="Alexandru Ungureanu, Constantin Rusu, Ovidiu Gabriel Iancu,  
Gheorghe Iacob" SpeakerAccent="" Gender="Male" Age="0" /> </speechSection>
```

Augmentarea corpusului vocal prin noi înregistrări

- ▶ Exemplul al II-lea: colecția *SoRoEs*

```
<?xml version="1.0" encoding="utf-8"?>
```

```
<tei>
```

```
<teiHeader Description="" Collection="SoRoEs" Keywords="București" Language="ro"  
Contributor="DSU-ICI-UAIC" Distribution="-"> <speechSection SpeechCreator="  
DSU-ICI-UAIC " AcousticMedia="în spațiu liniștit" Duration="4:00"  
SamplingFrequency="16" Resolution="0" RecordDate="25iunie 2016"  
RecordTime="09.00" Equipment="Professional Marantz recorder" Broadcast=""  
SpeechFile="wav" SpeechFileType=„start-stop" SpeechSegmentation="0"> <speaker  
SpeakerName=„9B3c" SpeakerAccent=„muntenesc" Gender=„Female" Age=„45" />  
</speechSection>
```

Concluzii

- ▶ Resurselor încărcate pe platforma CoBiLiRo le-au fost asignate metadatele, alinierea și adnotările corespunzătoare, agreeate în cadrul consorțiului ReTeRom;
- ▶ Platforma găzduiește inventarul colecțiilor de date lingvistice românești disponibile în cadrul consorțiului.



Activitatea A3.3: Realizarea de statistici privind corpusul bimodal voce/text

Ionuț Pistol, Andrei Scutelnicu, Șerban Boghiu, Felix-
Cristian Pericică

Rezumatul activității

Activitatea are ca obiectiv dezvoltarea platformei CoBiLiRo prin oferirea unor statistici privind resursele disponibile, atât la nivel individual cât și la nivelul întregului corpus.

Tehnologiile relevante

- Python Web Api -> Flask & Flask Restful
- Spacy - ro_core_news_sm -> Teprolin

Statistici la nivelul întregii colecții de resurse

Filtrele

Se obțin:

- resursele care corespund filtrelor aplicate
- numărul de resurse adiacente

Statistici la nivelul întregii colecții de resurse

Filtrele

- Pe tipul de segmentare al fișierelor audio:
 - “File”
 - “Start-Stop”
 - “File-Start-Stop”

Statistici la nivelul întregii colecții de resurse

Filtrele

- Pe tipul de fișier audio disponibil:
 - “wav” (Waveform Audio File Format)
 - “mp3” (MPEG Audio Layer 3)

Statistici la nivelul întregii colecții de resurse

Filtrele

- Pe nivelul de adnotare:
 - “Propoziție”
 - “Cuvânt”
 - “Fonem”
 - “Prozodic”

Statistici la nivelul întregii colecții de resurse

De asemenea este calculată și durata totală a fișierelor audio de fiecare tip, folosind *ffprobe* - un multimedia stream analyzer.

Statistici la nivelul fișierelor text

Fișierele text asociate resurselor sunt prelucrate și procesate pentru a obține rând pe rând următoarele componente de statistică:

- **Densitatea lexicală a textului transcris asociat resursei** - redată prin formula *nr. de cuvinte lexicale / nr total de cuvinte*;

Statistici la nivelul fișierelor text

- **raportul type/token** - *nr. cuvintelor care apar o dată / nr. total de cuvinte * 100;*

numărul de tokeni;

numărul de lemme - o lemmă o singură dată;

Statistici la nivelul fișierelor text

- **Numărul formelor flexionare distincte;**
- **Media formelor flexionare distincte - *nr. total de forme flexionare distincte / nr. total de tokeni;***
- **Media de lemme din text - *nr. de lemme distincte / nr. total de tokeni;***

Statistici la nivelul fișierelor text

- numărul de lemme care apar de mai multe ori;
- raportul de lemme care apar de mai multe ori - *nr. de lemme care apar de mai multe ori / nr. total de lemme luate în considerare o singură dată * 100;*

Statistici la nivelul fișierelor text

În abordarea prelucrărilor lingvistice ulterioare se va avea în vedere și inventarierea lemmelor unicate, urmând ca pe interfață să fie afișată o nouă componentă - *hapax legomenon*.

Toate fișierele sunt curățate în prealabil de semnele de punctuație și eventualele spații și/sau simboluri nejustificate.

Statistici la nivelul fișierelor audio

- *Background Job Worker* (biblioteca *HangFire*) - calculează zilnic durata totală a fișierelor încărcate pe platformă;
- analizatorul de stream-uri *ffprobe* - extragerea duratei fiecărui fișier audio.

Statistici la nivelul alinierilor

Pentru alinierea la nivel de cuvânt, cele mai relevante în majoritatea aplicațiilor care valorifică alinieri (antrenarea *text-to-speech* și *speech-to-text*, aliniatoare automate), am pregătit un set de statistici ce pot fi generate automat.

Scopul principal: oferirea unei priviri de ansamblu asupra calității alinierilor disponibile.

Statistici la nivelul alinierilor

Alinierile au fost realizate cu ajutorul componentei de analiză pusă la dispoziție de TADARAV:

File	AutoText	AlignedText	LongestUnmatch	StartsAt	SingleUnmatched
Alma-Mater-Iasiensi-24-Facultatea-de-Drept	5379	3761	44	30.51	193
Alma-Mater-Iasiensis-1-Aniversarile-Universitatii	7078	4400	121	2944.26	248
Alma-Mater-Iasiensis-11-Prof-Univ-Dr-Andrei-Margha	7072	4947	65	2688.87	266
Alma-Mater-Iasiensis-12-Facultatea-de-Fizica	5232	3312	52	3282.42	166
Alma-Mater-Iasiensis-13-Universitatea-din-Iasi-prezent-si-perspective	7006	5220	61	2109.15	281
Alma-Mater-Iasiensis-14-Universitatea-din-Iasi-Dimensiunea-culturala	6301	4558	41	2090.97	265

Statistici la nivelul alinierilor

Observații:

- procentul cuvintelor aliniate corect este de aproximativ 70%;
- în fișierul al doilea există o zonă lungă de cuvinte nealinate (121 cuvinte) ce poate indica prezența în transcrierea manuală a unui fragment care nu se regăsește și în înregistrarea audio;
- în fișierul 5 există un număr relativ mare de cuvinte izolate nealinate (281) ceea ce ar putea indica o calitate scăzută a semnalului audio sau un accent/prozodie nestandard a vorbitorului.


Concluzii

Oferirea unor statistici relevante contribuie la:

- mărirea încrederii unor potențiali noi contribuitori;
- potențialul resurselor de a-și găsi utilitatea în diverse contexte.

Concluzii

Statisticile deja incluse pe platformă precum și cele care vor fi incluse până în aprilie 2021 promet să asigure acest avantaj pentru colecția de resurse bimodale, principalul rezultat al proiectului CoBiLiRO.



Activitatea 3.4:
*Proiectare de aplicații de exploatare a
corpusului bimodal și a tehnologiilor de
prelucrări textuale și voce, create în
proiectele P2, P3, P4*

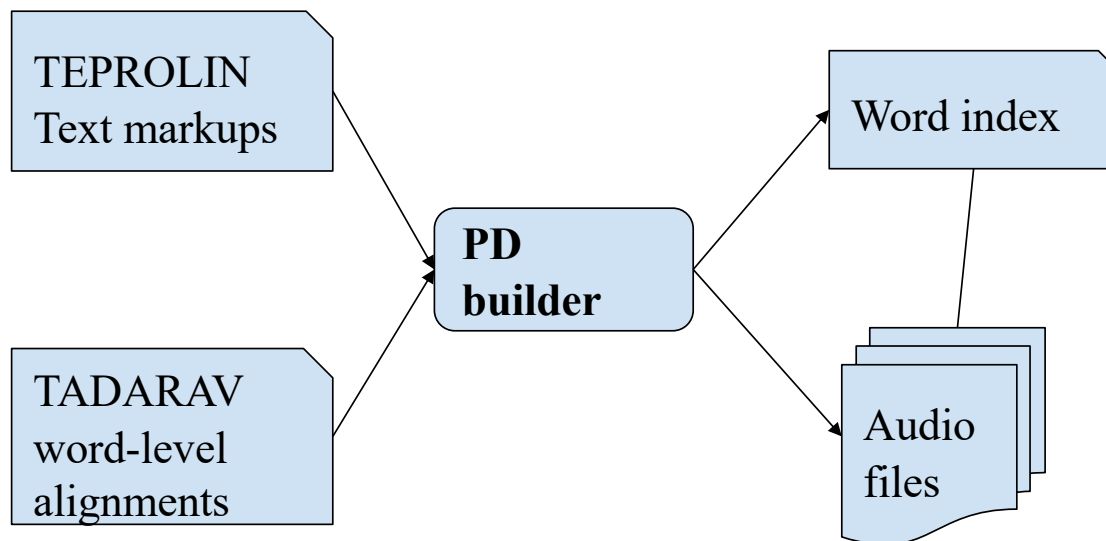
Ionuț Pistol, Andrei Scutelnicu, Diana Trandabăț, Daniela Gîfu, Șerban Boghiu

Activitatea 3.4 - Proiectare de aplicații de exploatare a corpusului bimodal și a tehnologiilor de prelucrări textuale și voce, create în proiectele P2, P3, P4

- P2 (TEPROLIN), tehnologii de prelucrare a textelor, stadiul curent descris în raportul activității 3.7 (Definitivarea, testarea, validarea și împachetarea într-o soluție „*ready-to-use*” a platformei integrate și configurabile de prelucrare a textelor în limba română.)
- P3 (TADARAV), transcrierea și alinierea automată a semnalului vocal, stadiul curent descris în raportul activității 3.10 (Îmbunătățirea soluției de filtrare și aliniere a transcrierilor aproximative cu semnalul de vorbire)
- P4 (SINTERO), generarea semnalului vocal, stadiul curent descris în raportul activității 3.15 (Dezvoltarea unei noi tehnologii pentru adaptarea vocii sintetice la stilul și expresivitatea unui nou vorbitor) și 3.17 (Integrare tehnologie nouă și demonstrare în realizarea interfețelor om-mașină pentru sinteza text-vorbire).
- Alte tehnologii pentru limba română, cum ar fi cele disponibile la <http://nlptools.info.ro> și <https://www.racai.ro/en/tools/text/>

Suport pentru învățarea limbii române - *PD builder (I)*

Obiectiv: crearea unui dicționar de pronunții pentru limba română



Word index

- Plecând de la textele adnotate de TEPROLIN, primul pas este indexarea intrărilor. Ca intrare considerăm atât forme de bază cât și eventualele forme flexionate disponibile.
- Corespondența între cuvântul de bază și formele flexionate se face pe baza lemei identificate de TEPROLIN.
- În cazul în care o intrare apare de mai multe ori în colecție, toate aparițiile sunt indexate. Acest lucru permite atât descoperirea unei pronunții standard cât și referirea la marcasele disponibile pe documentul sursă.

Suport pentru învățarea limbii române - *PD builder (II)*

Audio files

- Pentru pronunția unui cuvânt, în momentul adăugării unei intrări sau referințe în index aplicația extrage din fișierul audio corespunzător sursei zona aliniată cu acea intrare conform marcajelor TADARAV. În cazul în care există mai multe pronunții pentru un cuvânt sunt extrase și indexate toate aceste variante.
- Un mecanism propus pentru această aplicație, posibil cu suportul tehnologiilor SINTERO, este construirea sau identificarea unei pronunții standard pentru acel cuvânt. Formele de undă corespunzătoare pronunției aceluiași cuvânt ar trebui să fie similare, o medie a lor ar putea corespunde fie unei înregistrări disponibile, fie s-ar putea folosi pentru generarea unei înregistrări noi ce va fi asociată cu acel cuvânt.
- O metodă de clusterizare ar putea fi folosită pe setul de înregistrări, în cazul în care media este diferită semnificativ față de pronunțiile disponibile, pentru a descoperi eventuale alternative de pronunție. Astfel, pentru un cuvânt aplicația ar putea semnaliza și oferi două sau mai multe alternative de pronunție.

În lucru, primul prototip estimat Q2 2021

Analiza corpusurilor bimodale (I)

Obiectiv: Identificarea și analiza automată a erorilor de aliniere din corpusurile bimodale

Erorile în alinierea produse de aceste tehnologii au, în general, una din următoarele cauze:

- diferențe între transcrierea text și conținutul înregistrării;
 - calitatea scăzută a înregistrării;
 - erori ale aplicației de aliniere.
-
- Primele două surse de erori sunt independente de soluția de aliniere și influențează semnificativ calitatea potențială a alinierii produse precum și calitatea corpusului bimodal.
 - O parte din posibilele erori semnalate ar putea fi corectate, lucru care ar mări semnificativ calitatea alinierii automate dar și a resursei bimodale în general.
 - O mare parte din soluțiile propuse mai jos necesită disponibilitatea cel puțin a unei metode de aliniere automată capabilă să ofere și scoruri de încredere pentru alinierea produse.

Analiza corpurilor bimodale (II)

Diferențele între transcrierea text și conținutul înregistrării pot fi de mai multe tipuri:

- *Fragmente adiționale prezente în text sau în înregistrare.* Descoperirea automată a acestei situații poate fi făcută cu ajutorul unor soluții de aliniere automată ce ar putea să descopere zone din text sau audio care nu conțin alinieri precise.
- *Transcrieri problematice.* Există situații prezente în text ce sunt dificil de transpus în audio, în special la o citire care nu are ca obiectiv transpunerea identică a conținutului textual (prescurtările, abrevierile și acronimele, numerele și datele calendaristice). Descoperirea automată a acestor potențiale erori poate fi făcută de multe ori analizând doar fișierul text, dar dacă dispunem și de o aliniere produsă automat putem indica cu precizie zona echivalentă din audio unde eroarea este prezentă.
- *Particularități ale pronunției și prozodiei vorbitorului pe fișierul audio.* Cum aliniatoarele automate se bazează pe recunoașterea fonemelor și a granițelor (pauzelor) dintre ele, un vorbitor care nu are o dicție suficient de bună poate face aceste elemente mai dificil de descoperit. Descoperirea poate fi făcută automat prin observarea unor erori comune în alinierea anumitor cuvinte sau fragmente.

Analiza corpurilor bimodale (III)

- Calitatea scăzută a înregistrării poate fi descoperită automat prin parametri precum: rata de eșantionare (*sample rate*), dimensiunea compresiei (*bit rate*), calitatea compresiei (*bit depth*). Ei pot fi detectați automat și pot indica o calitate generală a înregistrării.
- Pentru determinarea unor zone specifice în care calitatea semnalului audio e scăzută (zgomot de fundal, semnal slab), se poate apela din nou la un aliniator automat care ar indica o zonă în care alinierea e problematică fără a exista potențialul unor diferențe între text și audio (durate estimate aproximativ egale, textul nu conține situații problematice, text similar aliniat corect în altă parte a resursei). Analiza formei de undă a semnalului audio poate indica prezența unor voci adiționale ce perturbă calitatea înregistrării.

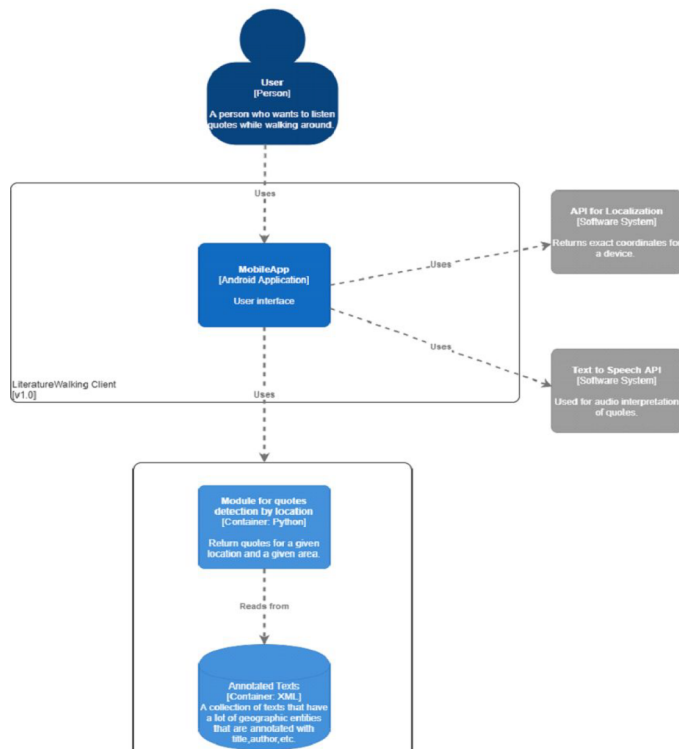
Analiza corpusurilor bimodale (IV)

File	AutoText	AlignedText	LongestUnmatch	StartsAt	SingleUnmatch	Observations			
Alma-Mater-Iasiensi-24-Facultatea-de-Drept	5379	3761	44	30.51	193				
Alma-Mater-Iasiensi-1-Aniversarie-Universitatii	7078	4400	121	2944.26	248				
Alma-Mater-Iasiensi-11-Prof-Univ-Dr-Andrei-Marga	7072	4947	65	2688.87	266				
Alma-Mater-Iasiensi-12-Facultatea-de-Fizica	5232	3312	52	3282.42	166				
Alma-Mater-Iasiensi-13-Universitatea-din-Iasi-prezent-si-perspective	7006	5220	61	2109.15	281	Bad transcription			
Alma-Mater-Iasiensi-14-Universitatea-din-Iasi-Dimensiunea-culturala	6301	4558	41	2090.97	265	Bad transcription			
Alma-Mater-Iasiensi-18-Viata-Universitatii	6405	3977	152	0	224				
Alma-Mater-Iasiensi-20-Facultatea-de-Geografie-si-Geologie	6775	4787	36	113.31	228				
Alma-Mater-Iasiensi-21-Historia-Magistra-vitae	6819	4934	38	1438.17	279	Bad transcription			
Alma-Mater-Iasiensi-22-Filosofia-la-Universitatea-din-Iasi	6488	4758	86	0	258				
Alma-Mater-Iasiensi-23-Manifestarile-Jubileului	6096	4518	52	46.89	193				
Alma-Mater-Iasiensi-25-Facultatea-de-Economie-si-Administrarea-Afacenilor	6075	4289	77	37.53	220				
Alma-Mater-Iasiensi-26-Facultatea-de-Informatica	5785	3921	66	44.67	226				
Alma-Mater-Iasiensi-27-Facultatea-de-Teologie-Romano-Catolica	7073	4937	153	0	275				
Alma-Mater-Iasiensi-28-Facultati-reinfiintate	6200	4205	40	0	214				
Alma-Mater-Iasiensi-4-Facultatea-de-Teologie-Ortodoxa	6758	4787	118	0	258				
Alma-Mater-Iasiensi-5-Institutia-Rectoratului-despre-rectorii-din-alte-vremuri	7999	4574	164	0	253				
Alma-Mater-Iasiensi-6-Universitatea-din-Iasi-in-relatiile-internationale	6562	4086	72	0	194				
Alma-Mater-Iasiensi-8-Prof-Univ-Dr-Vasile-Isan	5490	3939	63	0	203				
Alma-Mater-Iasiensi-9-Seminarul-Myller	5572	3452	59	0	180				
BarbulInterviu1	6369	2722	48	2670.93	135				
BarbulInterviu2	8023	3695	105	4397.55	236				
BarbulInterviu3	6521	2762	62	2955.03	187				
BarbulInterviu4	7119	2811	200	6839.19	177				
Ghici-Cine-mai-Vine-La-Cina-Constantin-Milica	5932	3368	116	1476.21	190				
Ghici-cine-mai-vine-la-cina-Cornelia-Gheorghiu	4484	1678	322	2709.12	76	Poor sound quality			
Ghici-Cine-Mai-Vine-La-Cina-Dan-Hatmanu	4915	2576	55	449.55	145				
Ghici-Cine-Mai-Vine-la-Cina-Emil-Brumaru	4110	1834	124	736.86	93				
Ghici-cine-mai-vine-la-cina-Grigore-IIsei	5887	4432	52	538.17	267	Bad transcription			
Ghici-cine-mai-vine-la-cina-Horia-Zillieru	4819	3140	57	2818.38	180				
Ghici-cine-mai-vine-la-cina-Ion-Atin	3971	1999	46	1757.61	107				

În lucru, primul prototip Q2 2020, o primă versiune completă (cu interfață) și includerea în platforma CoBiLiRo Q1 2021

I listen to my speaking agent reading book fragments as I walk by (I)

Obiectiv: oferirea de informații suplimentare privind locațiile vizitate și sugestii de lectură



- Aplicația are la bază o colecție de texte care abundă în entități geografice, marcate XML explicit, textele fiind însoțite de metadata care descriu minimum: autorul și titlul cărții, anul de apariție și editura.
- Instalată pe un dispozitiv mobil, ea va semnala proximitatea telefonului față de locațiile menționate în texte și va citi acele fragmente care includ mențiunile respective.

I listen to my speaking agent reading book fragments as I walk by (II)

- Un prim pas în realizarea proiectului propus este adnotarea XML a unui document, pentru marcarea entităților ce desemnează locuri, instituții, monumente istorice etc., alături de crearea metadatelor care specifică informații referitoare la cărțile în care apar aceste entități, cu o marcăre cât mai corectă și completă a acestora.
- Chestiunea identificării și a clasificării entităților cu nume, în implementarea experimentală realizată cu studenți, s-a făcut prin învățarea unor șabloane, capabile apoi să identifice și să clasifice chiar și entități care nu apar în textele adnotate. Metodele de clasificare a entităților sunt capabile să le asigneze acestora trei categorii: persoane, organizații și teritorii.
- Primele soluții aduse pentru problema recunoașterii entităților cu nume s-au bazat pe aplicarea unor șabloane, reguli sau automate finite, în general create manual. Această abordare presupunea expertiză umană pentru elaborarea șabloanelor, iar pentru că șabloanele create manual nu puteau acoperi toate cazurile de entități prezente în corpusuri mai mari, sistemele ulterioare au încercat să învețe automat aceste șabloane din corpusuri adnotate, folosind diverse tipuri de reguli, transductoare sau automate finite.

În lucru, primul prototip Q2 2020

Sistem de sinteză text-vorbire (TTS) și clonarea vocii în limba română cu metoda învățării prin transfer (I)

Obiectiv: sistem ce permite sinteza vocală plecând de la o voce cunoscută (clonarea vocală)

- Sistemul presupune un modul de sinteză din text în spectogramă mel (sintetizator) și unul de generare a vorbirii din spectograma mel (vocoder). Ca sintetizator s-a ales Tacotron 2, antrenat inițial pe setul de date LJ Speech, iar ca vocoder - WaveGlow, antrenat inițial pentru limba engleză. S-au ales parametri audio identici pentru ambele modele, corespunzători cu parametrii LJ Speech.
- Seturile de date în limba română au fost preprocesate, iar ambele modele au fost antrenate pe acestea. Sintetizatorul a fost antrenat pe un set de date aflat pe platforma RETEROM, corespunzător unui singur vorbitor, în timp ce vocoderul a fost antrenat pe un set de date cu mai mulți vorbitori.
- Sistemul TTS poate fi conceptualizat ca un sistem de clonare a vocii unei persoane, însă sarcinile diferă prin accentul pus pe identitatea vorbitorului: în cazul TTS general, interesul este de a genera o voce cu caracteristici naturale cu cât mai puține greșeli de sinteză, în timp ce sarcina de clonare este de a genera vocea unei persoane specifice necunoscute, folosind cât mai puține mostre de vorbire ale acesteia.

Sistem de sinteză text-vorbire (TTS) și clonarea vocii în limba română cu metoda învățării prin transfer (II)

- Sintetizatorul Tacotron 2 folosit s-a dovedit eficient în procesul de aliniere, folosind un mecanism de atenție care grăbește procesul, permițând modelului să fie mult mai adaptabil. S-a folosit implementarea publicată de Nvidia (*GitHub - NVIDIA/tacotron2*), bazată pe modele preantrenate în limba engleză pentru sintetizator, cât și pentru două vocodere (WaveGlow și WaveNet), toate componentele bucurându-se de încorporarea tehnologiilor Apex (*Apex 0.1.0 documentation*).
- Arhitectura modelului se bazează pe o rețea recurentă, care prezice spectrograme mel din text. Tacotron 2 este antrenat prin folosirea de perechi text-spectrogramă mel. *Spectrogramele mel* sunt un tip de spectrogramă obținute prin aplicarea unei transformări non-liniare axei de frecvență în cadrul transformării STFT (*Short-time Fourier Transform*) a formelor de undă, fiind inspirată de urechea umană, filtrând anumite frecvențe pentru a permite o reprezentare mai redusă, dar cu efect exagerarea frecvențele joase, critice pentru inteligibilitatea vorbirii, cât și reducerea frecvențelor înalte, care sunt dominate de consoane fricative (z, s, f) și care, în general, nu sunt necesare pentru modelarea vorbirii.

În lucru, primul prototip Q2 2020

Asistent inteligent al ședințelor online

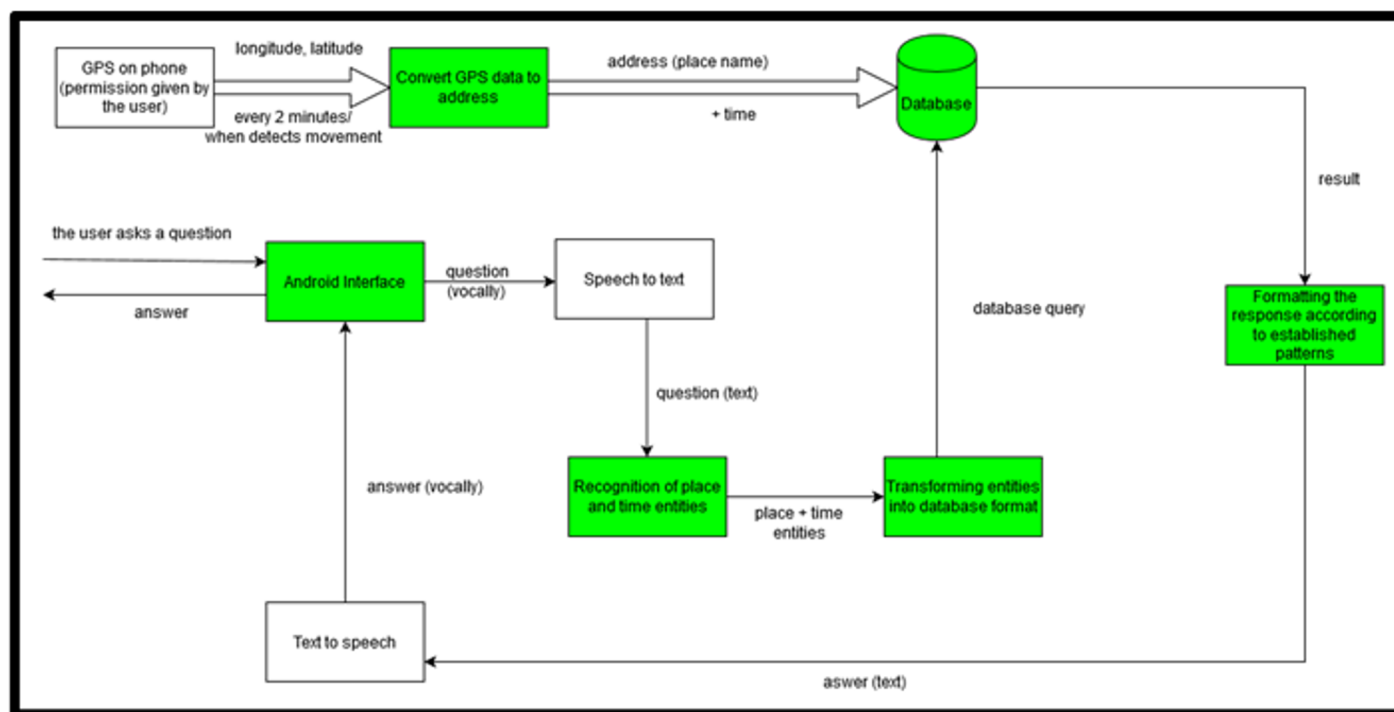
Obiectiv: extragerea de informații și generarea de rapoarte din discuțiile purtate și din mesajele schimbate pe parcursul interacțiunilor online

- extragerea de informații punctuale, doar pe anumite segmente din conferință (de exemplu pe durata apăsării unui buton) ori doar din intervențiile unor anumiți vorbitori, ceea ce ar elimina necesitatea ca utilizatorii să-și ia notițe în timpul conferinței, ceea ce le abate atenția de la discuții, putând duce chiar la omiterea unor informații esențiale;
- generarea automată a proceselor verbale, care ar trebui doar editate post-conferință;
- interogarea minutei ori a procesului verbal generat, pe bază de cuvinte cheie;
- căutarea în înregistrarea sonoră, a unor secvențe în care s-a menționat un anumit cuvânt cheie sau o anumită entitate, concept etc., cu reproducerea contextului vocal de apariție a lor.

În lucru, primul prototip Q4 2021

TRACKING ASSISTANT – Asistent inteligent pentru identificarea traseelor efectuate în decursul zilei (I)

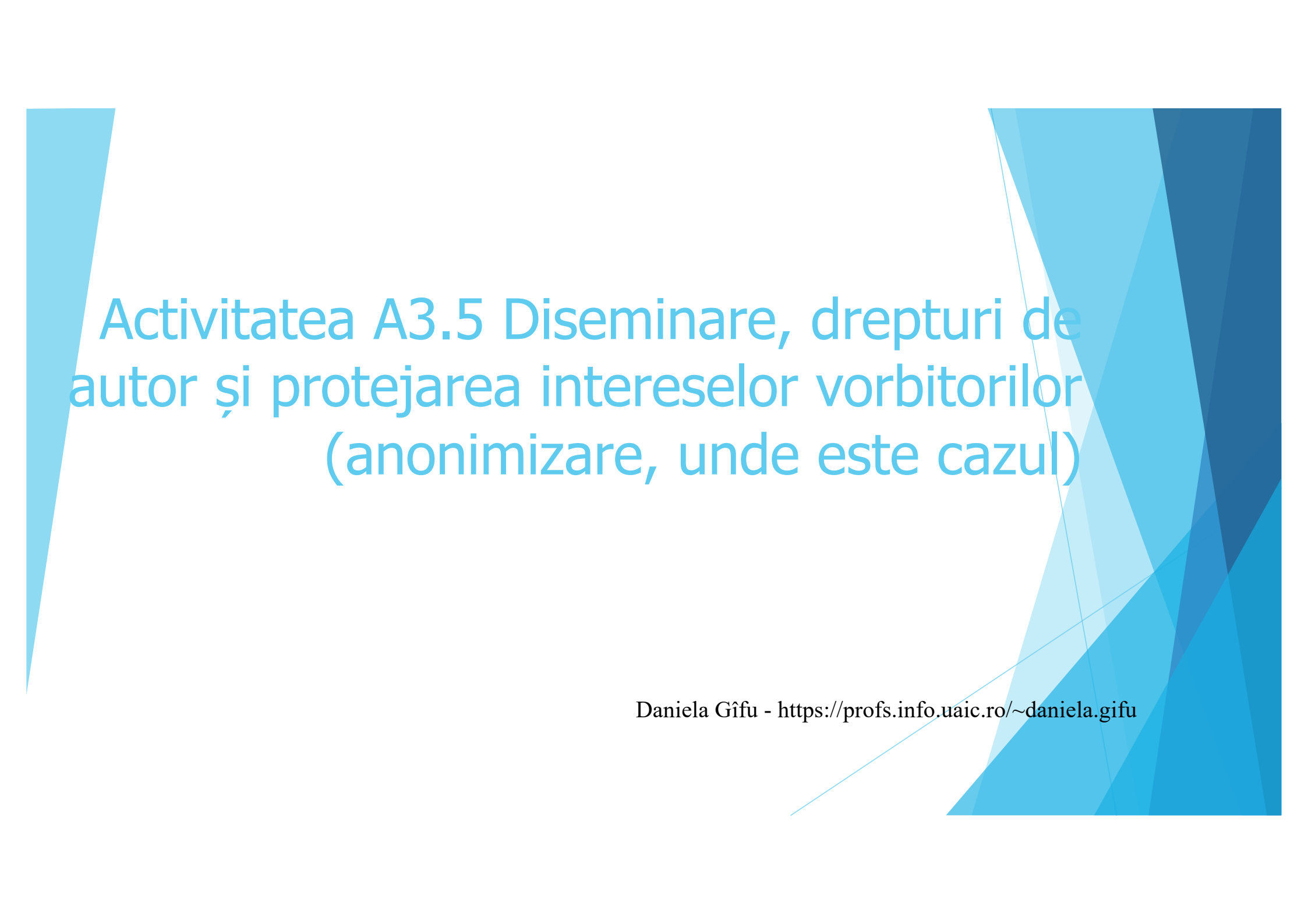
Obiectiv: asistarea persoanele care suferă de boala Alzheimer în a-și aminti traseul pe care l-au efectuat în timpul zilei printr-o interfață ușor de folosit, în limbaj natural



TRACKING ASSISTANT – Asistent inteligent pentru identificarea traseelor efectuate în decursul zilei (II)

- Primul modul se preocupă cu identificarea locației și stocarea sa în baza de date. Aplicația accesează modulul GPS o dată la fiecare 2 minute sau atunci când este detectată mișcare. Locațiile sunt stocate într-o bază de date împreună atât cu coordonatele GPS.
- Interogarea bazei de date se face în limbaj natural. Pentru aceasta, întrebarea este transformată în text folosind modulul de speech-to-text dezvoltat în cadrul proiectului component TADARAV. Ulterior, întrebarea este parsată pentru a fi extrase informații temporale.
- Următorul pas este identificarea referințelor la locații din întrebare, folosindu-se un gazetter, alături de șabloane de identificare a numelor de locații.
- După identificarea răspunsului în baza de date, aplicația folosește alt set de șabloane pentru a formula răspunsul. Șabloanele se bazează mult de cuvintele cheie din întrebare.
- Ultimul pas este transformarea textului în voce folosind API-ul aplicației dezvoltate în cadrul proiectului SINTERO, înainte de a-i fi oferit utilizatorului răspunsul.

În lucru, primul prototip Q2 2020



Activitatea A3.5 Diseminare, drepturi de autor și protejarea intereselor vorbitorilor (anonimizare, unde este cazul)

Daniela Gîfu - <https://profs.info.uaic.ro/~daniela.gifu>

Structură

ReTeRom

ReTeRom Project

- COBILIRO
- TEPROLIN
- TADARAV
- SINTERO

1. Introducere
2. Diseminarea corpusului bimodal, valorizare și utilizare
3. Comunicări & publicații
4. Concluzie

This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI – UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0818 - 73/2018 , within PNCDF III

DESCRIPTION OF COBILIRO

Name: Multi-level Annotated Bimodal Corpus for Romanian (COBILIRO).



The overall objective of this project is to create a thesaurus with audio and textual resources, annotated at different acoustic and linguistic levels, which is to become the most important reference for this type of resource for the Romanian language.



Applicability: The audio data and annotated text will be the foundation for the development of human machine interfaces technologies in natural language in Romanian: see Project 2 (TEPROLIN), Project 3 (TADARAV) and Project 4 (SINTERO).



Activities will consider: careful inventory of existing bimodal resources at project's partners; harmonization of representation formats, annotations and metadata; designing, building and testing the infrastructure hosting the resources; developing processing and access tools for the consortium; augmenting the voice-text corpus, completing it with metadata, alignments and annotations; conducting statistical studies on the corpus, exploitation for research and production, and wide dissemination of the bimodal corpus, valorization and use of type A1, A2 and B checks.

Introducere

Diseminare_Agenda ReTeRom - prezentarea rezultatelor obținute în anul 2020 în cadrul unor evenimente științifice cu impact major, totodată respectarea drepturilor de autor și protejarea intereselor vorbitorilor.

Instituții implicate: COBILIRO_UAIC, TEPROLIN_UTCN, TADARAV_UPB, SINTERO_ICIA

Indicatori: minimum 3 articole la conferințe și jurnale cotate internațional.

Diseminarea corpusului bimodal, valorizare și utilizare

- | -

Responsabil UAIC: Daniela Gîfu

Responsabil UPB: C. Burileanu, D. Burileanu,
Horia Cucu

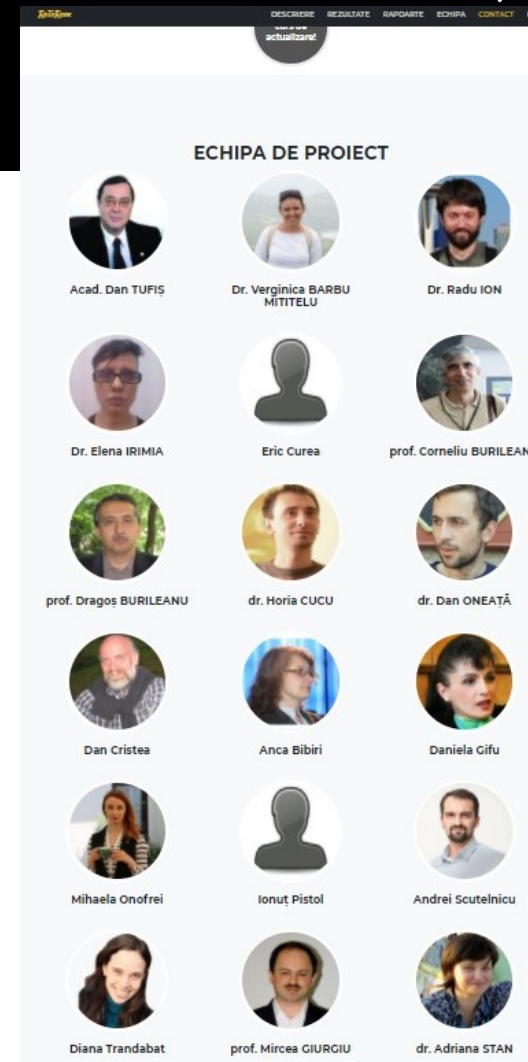
Responsabil UTCN: Mircea Giurgiu, Adriana Stan

Responsabil ICIA: Verginica Mititelu, Dan Tufiș

Actualizarea paginii web *ReTeROM*

- actualizarea cu informații a site-ului

<http://www.racai.ro/p/reterom/>



Diseminarea corpusului bimodal, valorizare și utilizare

- || -

Proiectare paginii web de control automat al prozodiei:
- demonstrator_versiune a modului de control automat al prozodiei:

https://speech.utcluj.ro/sintero/prosody_examples

1. Exemple voci naturale - stiluri de exprimare
2. Exemple voci sintetizate - modificarea tipului propoziției
3. Exemple voci sintetizate - modificarea proporțională a duratei vocalelor
4. Exemple voci sintetizate - adaptarea vorbitului la stilul jurnalistic
5. Exemple voci sintetizate - adaptarea vocii neutre la stilul narativ

The screenshot displays the SINTERO web interface, titled "Exemple audio - voci expresive". It features a navigation menu with "SINTERO" and a smiley icon. Below the title, a subtitle reads "Exemple audio din cadrul acestei pagini prezintă rezultatele primei versiuni a modului de control automat al prozodiei din cadrul sistemului de sinteză text-vorbire în limba română". The main content is organized into three sections, each with a header and a list of audio examples:

- Exemple voci naturale - stiluri de exprimare**: Includes "Stil neutru" (0:00 / 0:03), "Stil jurnalistic" (0:00 / 0:02), and "Stil narativ" (0:00 / 0:09).
- Exemple voci sintetizate - modificarea tipului propoziției**: Includes "Propoziție declarativă" (0:00 / 0:02), "Propoziție interogativă" (0:00 / 0:02), and "Propoziție exclamativă" (0:00 / 0:01).
- Exemple voci sintetizate - modificarea proporțională a duratei vocalelor**: Includes "Durată standard" (0:00 / 0:03) and "Durată scurtă" (0:00 / 0:01).

Each audio example is accompanied by a play button, a progress bar, and a volume icon.

Diseminarea corpusului bimodal, valorizare și utilizare

- III -

Proiectare paginii web demonstrator de sinteză text
vorbire

- demonstrarea sintezei folosește rețele DNN
de tip Tacotron-GST:

<https://speech.utcluj.ro/sintero/dnn-samples/>



SINTERO DNN-BASED TTS SAMPLES

Audio Samples

The following samples are generated with a DNN-based architecture similar to Tacotron-GST. The architecture also includes speaker ids in the training step. At synthesis time, the speaker id is also provided.

The speakers are a subset from the SWARA Corpus.

	Natural	Sample 1	Sample 2
Speaker 1			
Speaker 2			
Speaker 3			
Speaker 4			
Speaker 5			
Speaker 6			
Speaker 7			
Speaker 8			

Comunicări & publicații

UAIC – 4 articole științifice din care 3 ISI Web of Science (LREC_IWLTP & KES);
UPB – 5 articole științifice din care 2 ISI Web of Science (LREC & TSP);
UTCN – 5 articole științifice din care 1 ISI Web of Science (KES);
ICIA – 4 articole științifice din care 3 ISI Web of Science (LREC_IWLTP);

Concluzii

ReTeROM – se bucură de asigurarea promovării și vizibilității prin intermediul platformei - <http://dev.racai.ro/ReTeRom/> - compusă dintr-un “landing zone” și patru componente: “DESCRIERE”, “RAPOARTE”, “ECHIPA” și “CONTACT”.

Regăsim aici cele 4 subproiecte componente:

CoBiLiRo (Corpus bimodal pentru limba română adnotat pe multiple niveluri); **TEPROLIN** (Tehnologii pentru procesarea limbajului natural – text);

TADARAV (Tehnologii pentru adnotarea automată a datelor audio și pentru realizarea interfețelor de recunoaștere automată a vorbirii);

SINTERO (Tehnologii de realizare a interfețelor om-mașină pentru sinteza text-vorbire cu expresivitate).

Felicitări echipei

ReTeROM!