

Journée ETAPE

6 février 2012

Overview of the ETAPE campaign

Guillaume Gravier, Gilles Adda

<http://www.afcp-parole.org/eval/etape.html>



A special interest
group of



Many thanks to Gérard Chollet for arranging the meeting in Telecom ParisTech.

Why are we here today?

- Give an overview of the campaign
 - Tasks and evaluation protocols
 - Data: Annotations & Enrichment
- Get feedback from participants
- Get people to know each other

- Finalize evaluation plan
- Finalize calendar

Today's program

1. Introduction et campaign presentation (AFCP)
2. Data and annotation process (ELDA)
3. Metrics and scoring software (LNE)
4. Named entity annotation in Quaero (Sophie Rosset)
5. Phonetic alignements: first steps (LPP)
6. Syntax tree-banking: first steps (LLF?)

Today's program

1. Introduction et campaign presentation (AFCP)
2. Data and annotation process (ELDA)
3. Metrics and scoring software (LNE)
4. Named entity annotation in Quaero (Sophie Rosset)
5. Phonetic alignements: first steps (LPP)
6. Syntax tree-banking: first steps (LLF?)

ETAPE tasks

- **Segmentation tasks (S)**
 - Multiple speaker speech (SES-2)
 - Speaker diarization (SRL & SRL-X)
- **Transcription tasks (T)**
 - No constraints whatsoever
- **Information extraction tasks (E)**
 - Named entity detection (NE-ref & NE-asr)

Who signed up?

- CRIM, Eurecom, IRISA, IRIT, LIA, LIMSI, LINA, LI Tours, LIUM, LORIA, Telecom ParisTech
- Orange Labs, Synapse

What data is out there?

- **ETAPE data**

- 25h/8h/8h of radio data from DGA and TV data from ELDA (more details to come)

- **SES-2**

- not much except ETAPE data

- **TRS**

- ETAPE + ESTER (~200h) + EPAC (~100h)

- **NE**

- ETAPE + ESTER/Quaero reannotation

Questions for today

- From IRIT (on SES-2):
 - Detectin task or segmentation task ?
 - Conformity of annotations to the guideline
 - Submission format: 1 event or 4 events
 - Correctness of .trs files (path to DTD hard wired for TV shows)
 - Evaluation metrics: given the difficulty of the annotation, should we find the event or the region
 - Choice of regions included/excluded for scoring
- From EURECOM (on NE):
 - Submission format for NE?

How do we proceed?

- Organizer's presentations
 - Must be interactive ...
 - ... but we need time for everybody!
- Skype conf in the afternoon if possible
- Outcome of the meeting
 - minutes
 - Final version of the evaluation plan
 - Final calendar

Calendar

March 21: Distribution of test data for tasks S and T

March 28: Submission of runs for tasks S and T

April 2: Distribution of transcripts of task EN

start fo adjudication of tasks S and T

April 10: Submission of runs for task EN

April 13: Start of adjudication for task EN

April 16: End of adjudication for tasks S and T

April 27: End of adjudication for task EN

May 3: Release of official results

May 10-11: workshop

Open issues for discussion

- Will speaker attributed scoring be mandatory?
- Can we allow multiple submissions without identification of a primary one?
- Would AED be appropriate for SES-2?
- NE run on all ASR output made available?