

CAMOMILE



©Fir0002/Flagstaffotos

CHIST-ERA Project Seminar 2013
Brussels, March 26-27, 2013



presented by C. Barras



Centre de Recherche Public
Gabriel Lippmann

CAMOMILE project



- Collaborative Annotation of multi-MOdal, Multi-Lingual and multi-mEdia documents
- Call 2011 D2K, 42 months project
 - Oct. 1, 2012 to March 31, 2016
- 6 participants from 4 countries
 - France (3 CNRS labs: LIMSI, IMMI and LIG)
 - Turkey (ITU)
 - Spain (UPC)
 - Luxemburg (CRP-GL)

Scientific Background



- Large amount of heterogeneous, human-generated data available on the web
 - collection and analysis for research (social sciences, linguistics, economics, artificial intelligence & computer science)
- Focus on 3M data (multimodal, multimedia, multilingual)
 - share an object of study between scientific domains
- Statistical approaches rely on large amount of annotated data
 - human annotation costly
 - several modalities with different levels of abstraction

Scientific Background

The REPERE Challenge



- Multimedia challenge for person recognition in audiovisual broadcasts
 - Organized by French DGA and ANR, evaluation by LNE, 3 consortiums
- Fusion from different modalities
 - *who is seen* (face recognition)
 - *who speaks when* (speaker diarization)
 - *whose name is written* (OCR on overlay texts)
 - *whose name is pronounced* (speech to text + named entity detection)
- Data collection and reference annotation by ELDA
 - audio (Transcriber) and video (Viper)

Scientific Background

Annotation tools



Video annotation for REPERE with Viper

ViPER: Ground Truth Editor

File Edit View Media Timeline Scripts Window Help

file:/D:/Matthieu/Repere/DD0_v2/BFMTV_PlaneteShowbiz_2011-05-12_064500.MPG

☒ TEXTE
 ☒ PERSONNE
 ☒ PETITES_TETES
 ☒ File

P	ID	*NOM	*TETE	*APPARI...	*DISPARI...	*ORIENTA...	*OCCLU...	*SYNTHET...	*ATTRI...	*LUNE...	*COI...
☐	0	Louise_EKLAND	(262 99)(252 120)(2...	11069	11227	face	aucune	false	false	false	false
☒	1	Christophe_DELAY	(190 100)(182 121)(...	11228	11396	face	aucune	false	false	false	false
☒	2	Pascale_DE_LA_TOUR_DU_PIN	(465 113)(467 178)(...	11228	11396	face	aucune	false	false	false	false

Create Delete Duplicate

(1 / 16) 11350 f / 23996 frames Mark

TEXTE ☒
 PERSONNE ☒
 0 ☐
 1 ☐
 2 ☐
 PETITES_TETES ☒

Scientific Background Annotation tools



Audio annotation for REPERE with Transcriber

The screenshot shows the Transcriber 1.5.2 application window. The menu bar includes 'Fichier', 'Edition', 'Signal', 'Segmentation', 'Options', and 'Aide'. The main text area contains a transcript of a conversation. The transcript is divided into two sections: 'Arnaud ARDOIN' and 'Pascale BOISTARD'. The transcript text is as follows:

Arnaud ARDOIN

- &ouais, [dis=rep-]on a com(pris)[-dis=rep] on a compris votre message, parce que il y a triple message, [dis=hes-Jeuh[-dis=hes] désillusion, [r] [dis=hes-Jeuh hum[-dis=hes] incompris, vous les avez rencontrés. juste, <pers=Pascale_BOISTARD>Pascale Boistard</pers>, vous êtes aux premières loges. [r]
- [dis=hes-Jeuh[-dis=hes] question 1 et question 2 : quel est, selon vous, le statut qu'il faut donner [dis=rep-]à ces[-dis=rep] à ces hommes,
- [r] et quelle est aujourd'hui leur condition de vie ? je m'excuse de vous poser la question, alors que vous êtes adjointe [dis=hes-Jeuh[-dis=hes] à la ville [dis=rep-]de[-dis=rep] de Paris, mais [r]
- est-ce qu'on est en train de refabriquer un petit Sangatte, [b-]en quelque sorte, où[-b] on accueille ces hommes, avec des associations, [r] leur permettant de manger, [dis=rep-]de[-dis=rep] de s'installer et donc [dis=rep-]de[-dis=rep] de créer un point, comme on dit, un point d'ancrage ?

Pascale BOISTARD

- [r] [dis=hes-Jeuh[-dis=hes] je ne vois pas comment aurait pu faire la mairie de Paris [dis=hes-Jeuh[-dis=hes] autrement que ce qu'elle a fait aujourd'hui.

Below the transcript, there is a control bar with play, stop, and other audio controls. The file name 'LCP_CaVousRegarde_2011-05-09_235900.tr' and 'LCP_CaVousRegarde_2011-05-09_235900.MPG.wav' are displayed. Below the control bar is a waveform of the audio signal. At the bottom, there is a timeline with a cursor at 34:44.32. The timeline shows segments for 'Adrien...' and 'A' (Arnaud ARDOIN) and 'Pascale BOISTARD'. The segments are color-coded: orange for 'Adrien...', purple for 'A', and light blue for 'Pascale BOISTARD'. The transcript text is aligned with the timeline segments.



Key challenges and objectives

- Collaborative annotation framework
 - remote annotations combined locally
- Annotation of 3M data
 - unified framework for the different modalities
- Combine automatic and manual annotations
 - output automatic treatments as help for manual annotation
- Prototype of an experimental platform



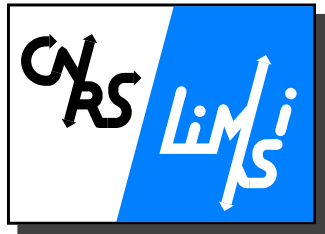
Aims for Collaborative Annotation

- Case study
 - ▶ Address 3M annotation and multimodal person identification
 - ▶ Build on the experience gained in the REPERE challenge
- Starting from mono-modal manual annotation tools
 - ▶ Speech: Transcriber, LDC XTrans...
 - ▶ Video: ANVIL, VIPER,...
 - ▶ NLP
 - ▶ Task oriented



Aims for Collaborative Annotation

- Go to multilingual / multimodal annotation
 - ▶ Annotation for various modalities
 - ▶ Distributed environment
 - ▶ Extension of REPERE (person identification)
 - ▶ New domains (TV fictions), new languages (Catalan, Turkish)
 - ▶ Collaborative annotation in parallel / in sequence / through crowdsourcing
- Coordination with international initiatives
 - ▶ ELDA, LDC, META-SHARE, ISO, OAC...
 - ▶ Through an international advisory group



Consortium CNRS - LIMSI (Orsay)



- Spoken Language Processing group
 - ▶ 18 academic + support staff, 8 contractuels and post-doct, 12 PhD
- Research topics
 - ▶ Speaker characterization in a multimodal context
 - ▶ Affective and social dimensions of spoken interactions
 - ▶ Perception and automatic processing of variation in speech
 - ▶ Robust analysis of spoken language and dialog systems
 - ▶ Translation and machine learning
 - ▶ Speech recognition
 - ▶ Language resources
- Participation to QUAERO program and REPERE challenge



- Oldest technical university in Turkey (1773)
 - ▶ The first faculty on computer and informatics in Turkey
 - ▶ Accepts students from top 1% of central university entrance exam
- Involved faculty members:
 - ▶ Asst. Prof. Hazım Kemal Ekenel (Principal Investigator –Computer Vision)
 - ▶ Assoc. Prof. Zehra Çataltepe (Machine Learning)
 - ▶ Asst. Prof. Cenk Demiroğlu (Speech Processing)





- Involvement in the project

- ▶ WP Leader -WP2: Multimodal Person Annotation Components

- ▶ Who is seen
 - ▶ Who is speaking
 - ▶ Combined decision

- ▶ WP3: Automatic Processing of Annotation

- ▶ T3.3 New data selection methods for active learning strategies –Active learning for multimodal and heterogeneous systems
 - ▶ T3.4 Evaluation for the multimodal person retrieval task

- ▶ WP4: Collaborative Annotation

- ▶ T4.2 Corpus selection and annotation
 - ▶ T4.3 Annotation guidelines



- 3 UPC labs

- ▶ Image and video processing group: 11 staff + 1 research assistant + 10 PhD
- ▶ Speech processing: 10 staff + 10 PhD students
- ▶ Natural Language processing: 9 staff + 8 PhD students

- Research lines

- ▶ Video Indexing: Region-based semantic annotation, Face ID, Generic object detection, QbE, Text extraction and recognition
- ▶ Hierarchical region-based image representation: Interactive segmentation, Video Coding, Object Detection
- ▶ 3D scene analysis and coding: Multi-camera reconstruction, body analysis, Tracking, 3DTV coding
- ▶ Speaker, language, and emotions recognition, Event detection (audio), Diarization
- ▶ Multilingual and robust speech recognition
- ▶ Multilingual text-to-speech conversion
- ▶ Speech to Speech Translation
- ▶ Natural Language Processing



Consortium CNRS - LIG (Grenoble)



- 165 academic and research staff, 260 doctoral and post-doctoral students, 65 support staff
- Two research teams involved (G. Quénot, L. Besacier)
 - GETALP: Study Group for Speech and Natural Language Translation/Processing
 - MRIM: Multimedia Information Modeling and Retrieval Group
- **Related research topics** (QUAERO consortium & REPERE challenge)
 - Multimodal person recognition in TV broadcast
 - Multimedia semantic indexing
 - Corpus annotation (Collaborative annotation)
 - Evaluation (TREC Vid, REPERE)
 - Active learning



Consortium: CNRS - IMMI (Orsay)

Institute for Multilingual and Multimedia Information



- French-German Institute
 - ▶ Joint venture between LIMSI-CNRS (Paris) + RWTH (Aachen) + KIT (Karlsruhe) (IMMI-Labs)
 - ▶ Director : J. Mariani, Associate Director : A. Waibel
 - ▶ Scientific Directors : J.L. Gauvain, H. Ney
 - ▶ Launched in the framework of QUAERO program
- 10 people at IMMI / 160 people in IMMI-Labs on MMI processing
- Participation in EC projects FLaReNet and META-NET
- Contributions to CAMOMILE
 - ▶ Annotation guidelines (ISO, META-DATA, NE)
 - ▶ Institutional guidance (META-SHARE, International)



- Department Informatics, Systems & Collaboration (ISC)
 - ▶ Applied Research and Technology Transfer: funding from research programs (international, national) & contract research (private, public)
 - ▶ Specialized in Visual Analytics: Data visualization, Data processing, Data provisioning, Software tools delivery
- In Camomile Consortium
 - ▶ Expert for a visualization platform that focuses on interpretation, interaction and heterogeneous data
 - ▶ Task leader for Cooperative Annotation Monitoring (WP4)
 - ▶ Apply the know-how of generic tools developed at ISC: Calluna (advanced visualization) & Cadral (data mining)

Workplan



- WP1 Management
- WP2 Multimodal Person Annotation components
- WP3 Automatic Processing for Annotation
- WP4 Collaborative Annotation



1. Management

- Internal meetings

- ▶ Kick-off meeting (Feb. 1, 2013)
 - ▶ use-case driven approach
- ▶ Several visio/audio conferences

- Web

- ▶ Wiki and mailing list in place for internal exchanges
- ▶ CMS for internet public site
 - ▶ waiting funding for a server dedicated to the project

- Consortium agreement

- ▶ Final version circulating before signature



1. Management

- Calendar issues
 - ▶ Spanish call for the project not yet issued
 - ▶ impact mostly UPC (only involved in WP2)
 - ▶ Notification sent by ANR on March, 15. to French partners
 - ▶ LIG and LIMSI hired a PhD and a post-doctorant in advance
 - ▶ other expenses delayed
- Schedule of deliverables slightly updated
 - ▶ take into account synchronization issues between partners



2. Multimodal Person Annotation

- Development of advanced components
 - ▶ Visual Person Annotation
 - ▶ locate and identify people in videos through robust face identification
 - ▶ Speaker Annotation in Audio
 - ▶ extend speaker diarization to PLDA and i-vector approach
 - ▶ Fusion (early vs. late)
- Assessment
 - ▶ On REPERE dataset
 - ▶ On CAMOMILE corpus (Turkish, Catalan)

2. Multimodal Person Annotation



- Work in progress for the first 6 months
 - ▶ Work on developing a robust face detection and tracking system for videos.
 - ▶ Research on advanced region-based shape descriptors for face representation.
 - ▶ Developed an i-vector based speaker verification system.
 - ▶ Classifier combination with kernelized eigenclassifiers
 - ▶ Acquired TV news broadcast recordings from a national TV channel.



3. Automatic Processing for Annotation

- Machine-assisted annotation of multilingual / multimodal data...
 - Active learning, semi-supervised learning, etc.
- ... but on a real scenario where annotations are partial and noisy
- If you cannot measure it, then you cannot improve it !
 - Metrics / tools to evaluate annotation quality (task-based) and productivity gains
- Efficient integration of annotators feedback
 - Iterative training / reinforcement learning



3. Automatic Processing for Annotation

- Use cases

- ▶ Active learning for multimodal annotation & multimodal fusion
- ▶ Active cleaning
- ▶ Resolve annotation conflicts with additional annotation steps
- ▶ Iterative model adaptation using continuously annotated data-sets

- Tasks

- ▶ MMM annotation platform dedicated to semi-supervised scenarios at a large scale
- ▶ Retraining - adaptation techniques and back-ends for the system supervisor
- ▶ New data selection methods for active learning strategies
- ▶ Evaluation for the multimodal person retrieval task



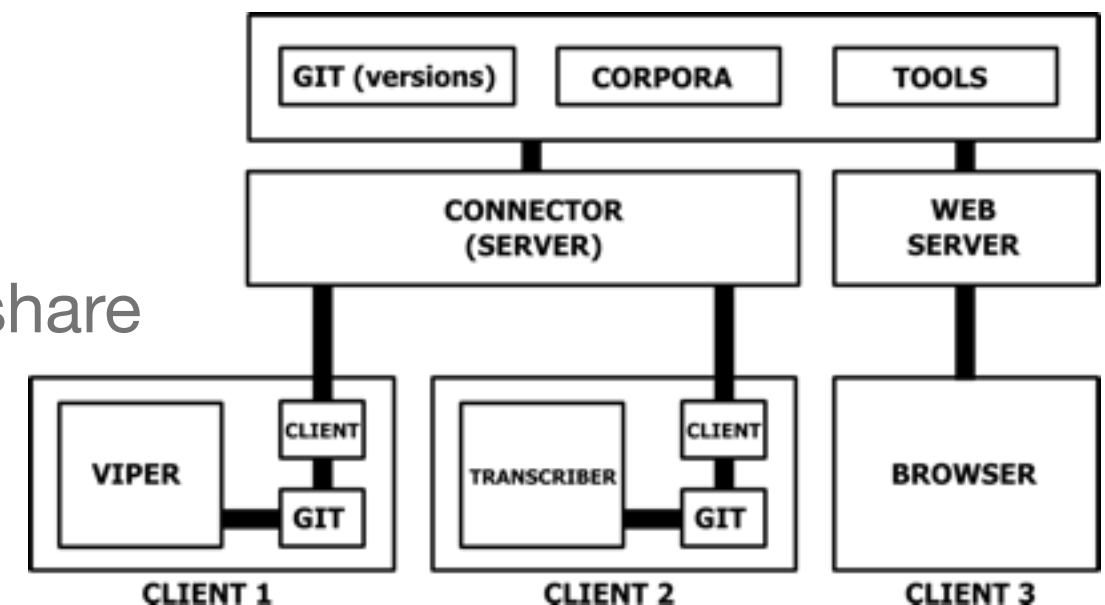
4. Collaborative Annotation

- Framework for collaborative annotation

- ▶ Analysis of existing annotation tools
- ▶ Specification of collaborative framework
- ▶ Interoperability through web services to share and perform collaborative annotations

- Corpus selection and annotation

- ▶ Representative corpus selected according to interest for WP2 use cases, increasing diversity / domain and languages
- ▶ Attention paid to IPR
- ▶ All partners participate in data annotation : real size test of collaborative annotation framework
- ▶ Improvement of data annotation collaborative framework





4. Collaborative Annotation

- Annotation guidelines
 - ▶ Definition of annotation guides within a collaborative system
 - ▶ Specification of various layers of annotation, in terms of semantic content and formal structure
 - ▶ Conformity automatically checked by program
 - ▶ Metadata based on META-SHARE
 - ▶ Named Entities in a multilingual/multicultural environment
 - ▶ Attention to the granularity of annotation in order to lock/unlock small annotation chunks for people working in parallel



4. Collaborative Annotation

- Institutional guidance

- ▶ Link with ELRA, LDC, META-SHARE, ISO, OAC,...
- ▶ International Advisory Group
- ▶ Tools and corpora should match needs and requirements of scientific community
- ▶ Presentation of annotation guidelines to potential institutional users
- ▶ Metadata presented to META-SHARE (META-DATA)
- ▶ Framework presented to META-SHARE for distribution / collaborative resource production / validation / evaluation
- ▶ Ethical dimension of LR production, including crowdsourcing



4. Collaborative Annotation

- Cooperative Annotation Monitoring
 - ▶ Design of an interactive web-based User Interface
 - ▶ Based on advances in cooperative visualization research
 - ▶ Overviews and details on demand about history of contributions to annotation of MM/ML data
 - ▶ Various filtering features based on visual query
 - ▶ Interface designed by CRP-GL



Annotation Platform / Specifications

- Input data
 - ▶ concentrate on evaluation scenarios of REPERE data
 - ▶ Interact with suggestions of elements to be annotated
- Support a use-case based approach
 - ▶ consider autonomous (naïve viewpoint) and lab users (evaluation viewpoint: datasets + algorithms)
 - ▶ feedback processing with interaction means
 - ▶ handling of conflictions situations
 - ▶ visualize, decide... who plays the moderator ?



Annotation Platform / Specifications

- Collaborative tool

- ▶ ability to share & monitor annotations
- ▶ web-based tool
- ▶ authentication system
- ▶ full historization & past versions recovery
- ▶ add-on: chat, tagging mechanism...

- Tests

- ▶ incremental tests on sample data
- ▶ extensive validation in common with other work packages

Expected Results



- Immediate impact
 - ▶ Improvement of automatic systems for Person Identification
 - ▶ scientific publications
 - ▶ Web services as interface to legacy annotation tool
 - ▶ released under open source licenses
 - ▶ Corpora of video annotations in additional languages (Turkish, Catalan)
- Mid- to long-term
 - ▶ Build on a first prototype of collaborative annotation framework
 - ▶ Multidisciplinary collaborative annotations
 - ▶ social sciences and computer sciences