Project-Team EXPRESSION

Expressiveness in Human Centered Data/Media

Vannes-Lannion-Lorient

Activity Report

2014

Contents

Contents

1 Team

1.1 Composition

Head of the team
Pierre-François Marteau, Professor, Université de Bretagne Sud

Administrative assistant
Sylviane Boisadan

Université de Bretagne Sud
Nicolas Béchet, Assistant Professor
Giuseppe Bério, Professor
Sylvie Gibet, Professor
Gildas Ménier, Assistant Professor
Jeanne Villaneau, Assistant Professor

Université de Rennes1
Nelly Barbot, Assistant Professor
Arnaud Delhay, Assistant Professor
Gwénolé Lecorvé, Assistant Professor
Damien Lolive, Assistant Professor

Associate members
Vincent Barreaud, Assistant Professor
Jean-François Kamp, Assistant Professor
Farida Said, Assistant Professor
Jonathan Chevelu, IGR, 09/2012, ANR Phorevox
Ludovic Hamon, post-doc, 09/2012, projet SIGN3D
Thibault Le-Naour, ATER, 10/2013, Université de Bretagne Sud
Caroline Larboulette, Associate researcher

PhD students
Clément Reverdy, Université de Bretagne Sud, CDE+Labo, first year
Marc Dupont, Université de Bretagne Sud, Thèse CIFRE Thales, first year
Pamela Carreno, Université de Bretagne Sud, ANR INGREDIBLE, second year

Lei Chen, Université de Bretagne Sud/Univ. McGill, ARED, second year
David Guennec, Université Rennes 1, third year
Hai Hieu Vu, Université de Bretagne Sud, fourth year
Raheel Qader, Université de Rennes 1, second year
Guyao Ke, Université de Bretagne Sud, fifth year

Master students

Clément Reverdy, Université de Bretagne Sud

1.2 Evolution of the staff

The permanent staff has been stable during 2014. Regarding the associate members, Ludovic Hamon has been Hired assistant professor at the UniversitÃ© du maine at Laval in September 2014. The number of PhD students is slightly increasing : Guiyao Ke as defended his PhD in February 2014, and Raheel Kader and Clement Reverdy have been hired as PhD students at IRISA, in October 2014 at UniversitÃ© de Rennes 1 (Lannion) and UniversitÃ© de Bretagne Sud (Vannes) respectively.

2 Overall Objectives

Expressivity or expressiveness are terms which are often used in a number of domains. In biology, they relate to genetics and phenotypes, whereas in computer science, expressivity of programming languages refers to the ability to formalize a wide range of concepts. When it comes to human expressivity, we will consider the following reading: expressivity is the way a human being conveys emotion, style or intention. Considering this definition, the Expression team focuses on studying human language data conveyed by different media: gesture, speech and text. Such data exhibit an intrinsic complexity characterized by the intrication of multidimensional and sequential features. Furthermore, these features may not belong to the same representation levels - basically, some features may be symbolic (e.g., words, phonemes, etc.) whereas others are digital (e.g., positions, angles, sound samples) - and sequentiality may result from temporality (e.g., signals).

Within this complexity, human language data embed latent structural patterns on which meaning is constructed and from which expressiveness and communication arise. Apprehending this expressiveness, and more generally variability, in multidimensional time series, sequential data and linguistic structures is the main proposed agenda of Expression. This main purpose comes to study problems for representing and characterizing heterogeneity, variability and expressivity, especially for pattern identification and categorization.

The proposed research project targets the exploration and (re)characterization of data processing models in three mediated contexts:

1. Expressive gesture analysis, synthesis and recognition,

2. Expressive speech analysis and synthesis,

3. Expression in text and language.

3 New Results

3.1 Main events

2014 has been the year of the finalization of the Expression Team Project [1] and its submission to the internal and external reviewing project by the end of the year. Comparing with 2013, the human resources have been globally unchanged except for the arrival of two PhD students and the defense of one PhD. No new project has been granted to the Team during this period. 2014 has been the end for the Sign3D project, funded by the "Investissements d'Avenir" initiative.

3.2 New Results by Key Issues

In accordance with the Team Project, the main outcomes for 2014 are listed into the following key issues items defined for the team:

Corpus annotation:

1. **Reliable metrics for multi-coders ordinal annotations**: The question of data reliability is of first importance to assess the quality of manually annotated corpora. Although Cohen's κ is the prevailing reliability measure used in NLP, alternative statistics have been proposed. We made an experimental study in [7] with four measures (Cohen's κ, Scott's π, binary and weighted Krippendorff's α) on three tasks: emotion, opinion and co-reference annotation. The reported studies investigate the factors of influence (annotator bias, category prevalence, number of coders, number of categories) that should affect reliability estimation. Conducted experiments show that the use of a weighted measure restricts this influence on ordinal annotations. They suggest that weighted α is the most reliable metrics for such an annotation scheme.

Data acquisition:

1. While human communication involves rich, complex and expressive gestures, available corpora of captured motions used for the animation of virtual characters contain relatively simple actions, ranging from locomotion to every day life emotions. We have created a novel corpus of expressive and meaningful gestures. We focused in particular on body movements and gestures involved in physical theater. A methodology for designing a corpus of full-body theatrical gestures has been proposed on the basis of magician tricks enriched with emotional content. Three sequences of magic tricks composed of 17 isolated gestures have been selected. Through the elicitation of different emotions in an actor's performance, we can increase the diversity and expressive richness of the proposed scenario and corpus. A set of four emotional states were chosen: happy, sad, stressed, and relaxed. A fifth state, neutral, was added to categorize the performances in which

[1] https://www-expression.irisa.fr/files/2011/07/expression_team_submitted.pdf

no emotion was intended. A Qualisys motion capture system composed of eight Oqus400 cameras was used for the recording of the movements. Each sequence (magic trick) was recorded twice per emotional state. Currently, our database contains the recorded movements of two non professional actors (one man and one woman). For each actor, one hundred and ten motion capture files were produced that corresponds to approximately one hour of recordings. We also validated the constructed corpus of theatrical gestures and expressive sequences of actions through several perceptual studies focusing on the complexity of the produced movements as well as the recognizability of the emotions produced [9]. This database is currently used within the ANR-Ingredible project.

2. Two small comparable corpora (English/French) have been produced to evaluate jointly the constructive methodology proposed in Guiyao Ke's PhD [1] to automatized the construction of comparable corpora and to evaluate the co-clustering and co-classification of comparable documents [17]. These Corpora are available at [2] and [3]

Multi-level representations:

1. **Easy, fast and consistent processing of large sequential annotated data collections**: The development of new methods for given speech and natural language processing tasks usually consists in annotating large corpora of data before applying machine learning techniques to train models or to extract information. Beyond scientific aspects, creating and managing such annotated data sets is a recurrent problem. While using human annotators is obviously expensive in time and money, relying on automatic annotation processes is not a simple solution neither. Typically, the high diversity of annotation tools and of data formats, as well as the lack of efficient middleware to interface them all together, make such processes very complex and painful to design. To circumvent this problem, the team has developped the toolkit Roots since several years. This toolkit is designed to efficiently handle massive complex sequential data and to allow quick and light prototyping, as this is often required for research purposes. Roots The toolkit has been released in open source in 2014 (`http://roots-toolkit.gforge.inria.fr`) and presented in two conferences [11, 10].

Data mining and knowledge extraction

1. **Pattern representation for ontology learning**. Ontology learning is a process for automating as much as possible ontology building from available sources. Textual resources are among the most important source types for ontology learning. Ontology learning from textual resources is a difficult task, facing to several difficulties, some of them being quite fundamental. In [23], we introduce a language for representing *lexico-syntactic patterns* for generating ontology content based on textual resources. We show how the language can be used to overcome some identified difficulties focusing on extraction of semantic relationships (other than taxonomical relationships). We also provide some details about the language implementation in *Text2Onto* tool.

[2] http://people.irisa.fr/Pierre-Francois.Marteau/Corpora/RSS_7classes.zip
[3] http://people.irisa.fr/Pierre-Francois.Marteau/Corpora/Wikipedia_21classes.zip

2. **Ontology validation process patterns**. Validation of ontologies is a critical open issue. Validation is a process fundamentally driven by an "ontology evaluation", often referred to as "quality evaluation". In the previous work, we have proposed a *standard typology* of problems impacting (negatively) on the quality of one ontology (shortly named *quality problems*). In [15], we report an experience, based on 2 ontologies learned from textual resources, showing how our previous work can be practically deployed for validation purpose. The analysis of experience results provides an earlier approach for ontology validation process patterns. These patterns are based on the one side, *dependencies* between problems, and, on the other side, on *potential correlations* between problems. Dependencies are order relationships between problems, establishing in which order problems have to be checked out. Dependencies are mostly based on well-known definitions and results, especially for logical frameworks. Correlations are relationships between problems, establishing co-occurrences of those problems within the reported experience. Correlations are mostly empirical and experience-driven, possibly meaningful for some types of text or domains.

3. **Time elastic positive definite kernels for time series datamining [5]**: We have proposed some extensions to the work on time elastic kernels dedicated to string or time series global alignment based on the aggregation of scores obtained by local alignments. The extensions that we propose allow for constructing, from classical recursive definition of elastic distances, recursive edit distance (or time-warp) kernels that are positive definite if some sufficient conditions are satisfied. The sufficient conditions we end-up with are original and weaker than those proposed in earlier works, although a recursive regularizing term is required to get the proof of the positive definiteness as a direct consequence of the Haussler's convolution theorem. Furthermore, the positive definiteness is maintained when a symmetric corridor is used to reduce the search space, and thus the algorithmic complexity, which is quadratic in the worse case. The classification experiment we conducted on three classical time warp distances (two of which being metrics), using Support Vector Machine classifier, leads to the conclusion that, when the pairwise distance matrix obtained from the training data is far from definiteness, the positive definite recursive elastic kernels outperform in general the distance substituting kernels for several classical elastic distances we have tested.

4. **Co-clustering of bilingual thematic data**: We have developed a methodology to assist the construction of bilingual thematic comparable corpora by means of co-clustering bilingual documents collected from raw sources such as the Web [17]. The proposed approach is based on a quantitative comparability measure [16] and a co-clustering approach which allow to mix similarity measures existing in each of the two linguistic spaces with a "thematic" comparability measure that defines a mapping between these two spaces. With the improvement of the co-clustering (k-medoids) performance we get, we use a comparability threshold and a manual verification to ensure the good and robust alignment of co-clusters (co-medoids). Finally, from any available raw corpus, we enrich the aligned clusters in order to provide "thematic" comparable corpora of good quality and controlled size. On a case study that exploit raw web data, we show that this approach scales reasonably well and is quite suited for the construction of thematic comparable

corpora of good quality.

5. **Sentence Similarity by combining Explicit Semantic Analysis and overlapping n-grams**: We have proposed a similarity measure between sentences which combines a knowledge-based measure, that is a lighter version of ESA (Explicit Semantic Analysis), and a distributional measure, Rouge [24]. We used this hybrid measure with two French domain-orientated corpora collected from the Web and we compared its similarity scores to those of human judges. In both domains, ESA and Rouge perform better when they are mixed than they do individually. Besides, using the whole Wikipedia base in ESA did not prove necessary since the best results were obtained with a low number of well selected concepts.

6. **Extraction and validation of induced syntactic relations**. Such relations are extracted by using a syntactic parser and a semantic closeness measure. Then, their validation is based on two different techniques: A Web Validation system on one part, then a Semantic-Vector based approach, and finally different combinations of both techniques in order to rank induced Verb-Object relations. The Semantic Vector approach is a Roget-based method which computes a syntactic relation as a vector. Web Validation uses a search engine to determine the relevance of a syntactic relation according to its popularity. An experimental protocol was set up to judge automatically the relevance of the sorted induced relations and experiments show that combination of both techniques perform better. This work was published in [2].

7. **Analogicial proportions in formal lattices and word lattices**. In collaboration with Laurent Miclet (Dyliss IRISA) and Henri Prade (IRIT), we have studied analogical proportions in the general setting of lattices. Analogical proportions are statements involving four entities, of the form 'A is to B as C is to D', and play an important role in analogical reasoning. They can turn out to be an original approach to classification in machine learning and their use shows promising in the task of language translation. Analogical proportions hold between examples described by vectors of Boolean features, i.e. between items of the same nature. However, statements having the form of analogical proportions often involve both objects and attributes. This has led us to the question of relating formal concept analysis and analogical proportions. We have proposed a definition of analogical proportion which is suitable for a non-distributive lattice structure as the one underlying formal concepts. We have studied the possibility to extract, from a concept lattice, proportional analogies pairing two (object, attribute)-pairs. These works, supported by the CNRS PEPS APA 2013-2015, have been pusblished in [22].

   In a collaboration with Laurent Miclet and Baptiste Jeudy (Laboratory Hubert Curien at the University of Saint Etienne), we have also defined the locally maximal subwords and locally minimal superwords common to a finite set of words. We have defined the corresponding sets of alignments and a partial order relation between such sets of alignments. We have shown that the constructed family of sets of alignments has the lattice structure. The study of analogical proportion in lattices gives hints to use this structure as a machine learning basis, aiming at inducing a generalization of the set of words. This work has been the subject of a book chapter [6].

Generation:

1. **Phonology modeling for expressive speech synthesis**: Expressive speech processing is an important scientific problem as expressivity introduces a lot of variability into speech. This variability leads to a degradation of speech application performances. In the phonological side of speech, pronunciation variants and of disfluencies are the two main phenomena impacted by expressiveness. The Ph.D. thesis of Raheel Qader has started on 1st, January 2014 on this topic. His focus is on phonology modeling for speech synthesis, but links also exist with automatic speech recognition because expressivity modelling in phonology is a cross-domain problem. A bibliographical review of the state of the art in expressivity and phonology modelling has been published as a research report [26]. Current work is focused on predicting prounciation variants to reflect a spontaneous speaking style.

2. **Adaptive statistical utterance phonetization for French**: Traditional utterance phonetization methods concatenate pronunciations of uncontextualized constituent words. This approach is too weak for some languages, like French, where transitions between words imply pronunciation modifications. Moreover, it makes it difficult to consider global pronunciation strategies, for instance to model a specific speaker or a specific accent. To overcome these problems, we have presented a new original phonetization approach for French to generate pronunciation variants of utterances. This approach offers a statistical and highly adaptive framework by relying on conditional random fields and weighted finite state transducers. The approach is evaluated on a corpus of isolated words and a corpus of spoken utterances. This work has been submitted to the conference ICASSP 2015 (results announced in January 2015).

3. **Reconstruction of marker trajectories from motion captured data**: A new method for reconstruction of marker trajectories which arise from motion capture. Assuming that there is a spatial and temporal correlation between the displacement of a marker and its neighbors, we propose to use the discrete Laplacian operator to reconstruct the missing data. Each posture of the motion is described by a graph whose vertices are characterized by differential information and some edges are associated with distance constraints. Differential information is used to preserve the spatial relationships between markers while distance constraints enable to preserve the length of some edges. We show through a set of experiments that The proposed method provides realistic reconstruction despite strong perturbations of the original signal [4].

Speech synthesis algorithms:

1. **Unit Selection Cost Function Exploration Using an A* based Text-to-Speech System:** Common corpus-based speech synthesis systems use the Viterbi algorithm to drive the unit selection step. In [14], we propose a system relying on the A^* algorithm, a general pathfinding strategy developing a graph (rather than a lattice). The implementation we propose uses state of the art techniques and is built as an experimental platform providing easy obtainment of the N-best paths as well as evaluation of new selection strategies. A perceptual evaluation assessing the overall quality of the system

output and the ranking of the paths returned by unit selection has also been conducted. Following this work, we have investigated a variant of the system by integrating a different set of filters and a penalty on non "sandwich units" [13]. The best strategy achieves a MOS score of 3.29 ($\pm$0.18). More interesting, the proposed system enables an in-depth analysis of unit selection.

Adapting prosodic models for expressive speech generation:

1. **Prosodic chunking algorithm for dictation with the use of speech synthesis:**
   In the context of the ANR project Phorevox, we have proposed an algorithm that automatically segment a text in prosodic chunks for dictation by conforming to the rules and procedures used in real settings to dictate a text to primary school children. A better understanding and modeling of these rules and procedures is crucial to develop robust automatic tools that could be used in autonomy by children to improve their spelling skills through dictation with the use of speech synthesis. The different steps used to derive the prosodic chunks from a given text will be explained through concrete examples. The proposal made here relies on the analysis of a corpus of 10 dictations given to children in French and French Canadian elementary schools. The phrasing observed in the data is described. It is thus simplified in order to develop an algorithm that automatically generates prosodic chunks from texts. This work has been published in [18, 19]. Moreover, a perceptual evaluation has been conducted as questionnaires that allow getting qualitative information on three distinct points concerning the algorithm and the speech synthesis system : (i) accuracy of the segmentation procedure, in particular regarding the location of the prosodic boundaries, (ii) size of the automatically calculated chunks, and (iii) accuracy and intelligibility of the speech synthesis system used, particular attention being given to the speech rate and the intonation patterns. Since we wanted to get feedback from potential users of the pedagogical tool we are currently developing, we designed a distinct questionnaire for two different populations (pupils and teachers). The results has been published in a conference paper [25]. To further investigate the behavior of the chunking algorithm, we compared its output to real teachers' chunking in [12]. Thus, we have conducted a corpus study of 10 dictations annotated or uttered by 13 teachers or speech therapists (10 and 3 respectively).

2. **Adaptation of prosodic models for different speech styles:** This work is a preliminary study whose main aim is to characterize four distinct speaking styles according to a limited set of prosodic features, including the length of prosodic phrases (AP and IP), the distribution of stressed syllables, pitch register span, the duration of silent pauses, etc. The analysis was performed using semi-automatic procedures on a corpus consisting of 30 minutes of speech per style. The study focuses on four styles, all of which are "overtly addressed to a given audience", but differ as to the nature of the audience (adults vs. children) and the desired impact of the address ("importance of being understood and convincing, or not"). Data analysis reveals that (a) dictation (addressed to children) and political speeches (addressed to adults) are different to the two other speaking styles (reading of novels and fairy tales) with respect to a specific set of prosodic cues; while (b) the speeches addressed to children differ from the ones addressed to adults, with respect

to another set of prosodic cues (especially pitch register span). These results have an interesting practical application: refining the design of pre-processing prosodic modules in a text-to-speech system, in order to improve the expressivity of synthesized speech. This work has been conducted in collaboration with the LLF (Laboratoire de Linguistique Formelle, UniversitÃ© Paris-Diderot) and Valibel Centre, Institute for Language & Communication (University of Louvain, Belgium) and published in [8].

Use cases and evaluation:

1. In the field of gestural action recognition, many studies have focused on dimensionality reduction along the spatial axis, to reduce both the variability of gestural sequences expressed in the reduced space, and the computational complexity of their processing. It is noticeable that very few of these methods have explicitly addressed the dimensionality reduction along the time axis. This is however a major issue with regard to the use of elastic distances characterized by a quadratic complexity. To partially fill this apparent gap, we have developed an approach based on temporal down-sampling associated to elastic kernel machine learning [20], [5], [21]. We experimentally show, on two data sets that are widely referenced in the domain of human gesture recognition, and very different in terms of quality of motion capture, that it is possible to significantly reduce the number of skeleton frames while maintaining a good recognition rate. The method proves to give satisfactory results at a level currently reached by state-of- the-art methods on these data sets. The computational complexity reduction makes this approach eligible for real-time applications.

Linguistic corpus design and corpus creation process:

1. Building a voice for TTS purposes generally relies on a script extracted from a big text corpus while optimizing the coverage of linguistic and phonologic events supposedly related to voice acoustic quality. While the reduction algorithms efficiency has been widely treated, the effective impact of the reduction and the choice of features to cover remains an open problem. We have simulated the voice creation process by taking two recorded corpora to evaluate the impact of their reduction on a TTS system output. The reduced TTS system output has been compared, using MOS and MUSHRA subjective evaluations, to TTS systems that use the full corpus or a randomly built one. This study quantifies the loss of quality induced by a corpus reduction. Moreover the random and the reduced corpora achieve a similar performance which is confirmed by objective measures. This result may indicate that covering very rare units causes an additional cost when selecting phrase-sized units. It enables new covering strategies for larger reductions that minimize quality loss. This work has been submitted to the SLT (IEEE Spoken Language Technology conference).

2. Linguistic corpus design is a critical concern for building rich annotated corpora useful in different domains of applications. For example, speech technologies such as ASR (Automatic Speech Recognition) or TTS (Text-to-Speech) need a huge amount of speech data to train data-driven model or to produce synthetic speech. Collecting data is always related to costs (recording speech, verifying annotations, etc), and as a rule of thumb,

the more data you gather, the more costly your application will be. Within this context, we have studied solutions to reduce the amount of linguistic text content while maintaining a sufficient level of linguistic richness required by a model or an application. This problem can be formalized as a Set Covering Problem (SCP) and we evaluate two algorithmic heuristics applied to design large text corpora in English and French for covering phonological information or Part of Speech labels. The first considered algorithm is a standard greedy solution with an agglomerative/spitting strategy and we propose a second algorithm based on Lagrangian relaxation. The latter approach provides a lower bound to the cost of each covering solution. This lower bound can be used as a metric to evaluate the quality of a reduced corpus whatever the algorithm applied. Experiments show that a suboptimal algorithm like a Greedy achieves good results; the cost of its solutions is not so far from the lower bound (about 4.35% for 3-phoneme coverings). Usually constraints in SCP are binary, we have proposed a generalization where the constraints on each covering feature can be multi-valued. This work has been submitted to MIT Press Computational Linguictics Journal.

3.3 Defended PhDs

Dr. Guiyao KE successfully defended his PhD thesis "Comparability measures for assisted construction thematic bilingual comparable corpus" at Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), Université de Bretagne Sud, Vannes, February 26, 2014. Reviewers have been Prof. Eric GAUSSIER (Université Joseph Fourier, LIG / Reviewer), Prof. Emmanuel MORIN (Université de Nantes, LINA / Reviewer), Prof. Geoffrey WILLIAMS (Université de Bretagne Sud, LICORN), Prof. Pierre-françois MARTEAU (Université de Bretagne Sud, IRISA) and Prof. Pascale SEBILLOT (INSA de Rennes, IRISA) who chaired the evaluation committee.

3.4 On going PhDs

1. Marc Dupont has completed his first year of research addressing gesture recognition using a data glove in the scope of controlling a moving robot in an adverse environment. His main achievements are mostly related to sensors testing and dedicated signal processing. He has also initiated an experimental protocol to optimize the gesture vocabulary in order to maximize the recognition rate.

2. David Guennec has completed his second year of research addressing the improvement of cost functions for unit selection speech synthesis. His main achievements concern the use of an A* algorithm and its comparison to the classical Viterbi algorithm and the introduction of a a new target cost to constrain the duration of phonemes. For this last point, duration are predicted using a neural network and used to constrain the selection process. The results he has obtained are encouraging.

3. Raheel Qader has completed his first year of research addressing phonology modeling for expressive speech synthesis. His main achievements are a review of the state of the

art, the analysis of a spontaneous speech English corpus and first experiments towards a pronunciation variant predictive model for speaker and style adaptation.

4. Pamela Carreno has completed her second year of research addressing expressive gesture synthesis. In the context of the ANR project Ingredible, she has proposed a methodology for building a corpus of full-body theatrical gestures on the basis of magician tricks enriched with emotional content. The constructed corpus and expressive sequences of actions have been validated through several perceptual studies focusing on the complexity of the produced movements as well as the recognizability of the produced emotions [9]. She also proposed a complete review of machine learning methods for motion synthesis [3], and developed an inverse kinematic algorithm and a classification approach to validate the synthesis.

5. Lei Chen has completed her second year of research addressing the analysis of expressive percussive motion. Using a database captured at McGill University (IDMIL lab.), she proposed a set of descriptors that enable to establish a ranking between expert percussionists.

6. Clément Reverdy has started his PhD. His research addresses the problem of facial expression synthesis in the context of sign language.

4 Software

4.1 SGN

Participants: Ludovic Hamon, Sylvie Gibet,.

In the context of the project *Investissement d'avenir* SIGN3D which ended in October 2014, we have completed a whole motion-capture-driven synthesis pipeline dedicated to the editing of gestures in French Sign Language (LSF) and the generation of gestures that can be visualized through a 3D virtual signer (library SGN). SGN is a set of software tools dedicated to a whole motion-capture-driven synthesis pipeline for the editing of gestures in French Sign Language (LSF) and the generation of gestures that can be visualized through a 3D virtual signer. The SGN is derived form the SMR library. It composed of 4 sub-libraries:

1. Sgn_Core. Sgn_Core is a library dedicated to the recording and the processing of one or several motions and one or several annotations from files formatted as "fbx", "bvh" and "eaf" files.

2. Sgn_DB. Sgn_DB is a library dedicated to the storing, the indexing and the retrieval of whole or part of captured and annotated motions, extracted from a heterogeneous and semi-structured database management system.

3. Sgn_Edit. Sgn_Edit is a library dedicated to the creation, the editing and the visualization of a motion obtained by the concatenative synthesis of a movement from elementary movements.

4. Sgn_IHM. Sgn_IHM is a library dedicated to the visualization of the motions (one or several articulations) in a 3D virtual environment.

4.2 ROOTS

Participants: Nelly Barbot, Vincent Barreaud, Jonathan Chevelu, Arnaud Delhay, Sébastien Le Maguer, Gwénolé Lecorvé, Damien Lolive.

The development of new methods for given speech and natural language processing tasks usually faces, beyond scientific aspects, various technical and practical data management problems. Indeed, the sets of required annotated features and their desired distribution in the training data are rarely the same for two different tasks, and many dedicated systems or expert resources use different file formats, time scales, or alphabets of tags.

In this context, ROOTS, stemming for Rich Object Oriented Transcription System, is an open source toolkit dedicated to annotated sequential data generation, management and processing, especially in the field of speech and language processing. It works as a consistent middleware between dedicated data processing or annotation tools by offering a consistent view of various annotation levels and synchronizing them. Doing so, ROOTS ensures a clear separation between description and treatment. Theoretical aspects of multilevel annotation synchronization have previously been published in [BBB+11] while a prototype had been presented and applied to an audiobook annotation task in [BCLML12].

As summarized in Figure 1, data are organized hierarchically in Roots, starting from fine grain information in items and moving to macroscopic representations as corpora. As a fundamental concept, data in ROOTS is modeled as sequences of items. These items can be of many types, e.g., words, graphemes, named entity classes, signal segments, etc., and can thus represent various annotation levels of the same data. Correspondences between items from different sequences are then defined as algebraic relations, leading to a graph where nodes are items and edges are derived from relations. Then, interrelated sequences are gathered into utterances. According to the application domain, utterances can refer to sentences, breath groups, or any relevant unit. A part of the recent work on ROOTS has focused on extending this hierarchization of data to easily handle large collections of data. Hence, the notion of corpus has been defined as a list of utterances or, recursively, as a list of subcorpora (called chunks), for instance to represent a chapter as a list of paragraphs. Besides chunks, corpora can also be partitioned "horizontally" into layers which gather annotations from a same field. The following operations are allowed for each data hierarchization level:

[BBB+11] N. BARBOT, V. BARREAUD, O. BOËFFARD, L. CHARONNAT, A. DELHAY, S. LE MAGUER, D. LOLIVE, "Towards a Versatile Multi-Layered Description of Speech Corpora Using Algebraic Relations", *in : Conference of the International Speech Communication Association (Interspeech)*, p. 1501–1504, Florence, Italie, 2011.

[BCLML12] O. BOËFFARD, L. CHARONNAT, S. LE MAGUER, D. LOLIVE, "Towards Fully Automatic Annotation of Audio Books for TTS", *in : Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*, European Language Resources Association (ELRA), Istanbul, Turkey, may 2012.

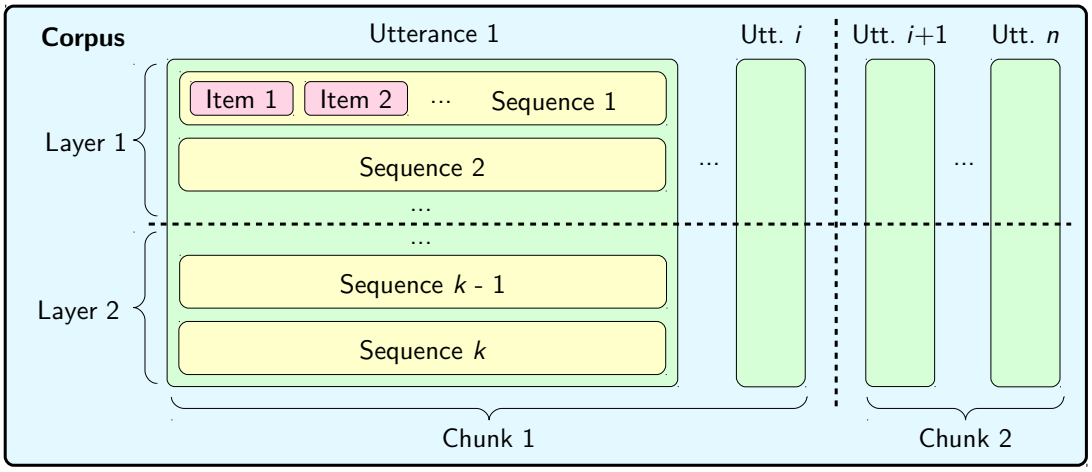

Figure 1: Hierarchical organization of data in ROOTS.

- Item: get/set the content/characteristics; get other items in relation; dump[4].
- Sequence: add/remove/get/update items; dump,
- Relation: get items related to another; link or unlink items; dump,
- Utterance: add/remove/get/update sequences; add/remove/get/update direct or composed relations; dump,
- Corpus: add/remove/get/update an utterance; add/remove chunks/layers; load/save; dump.

ROOTS is made of a core library and of a collection of utility scripts. All functionalities are accessible through a rich API either in C++ or in Perl. Recently, this API has greatly evolved and to ease building ROOTS corpora based on this API (e.g., with the notion of corpus), and accessing information in flexible and intuitive manners. Extra developments have also led to the following improvements: new wrapping scripts for basic corpus processing operations (merge, split, search) have been written and a LaTeX/PGF graphical output mechanism has been added in order to expertise and analyse the content of annotated utterances. This visualization functionality has been developed during the 3-month summer internship of Andrei Zene, a Romanian B.Sc. student.

The toolkit ROOTS is original compared to other related tools. Among them, GATE [CB02] proposes a framework to develop NLP pipelines but does not provide facilities to switch between GATE bundled processing components and external tools. More recently, the NITE XML Toolkit, or NXT, proposes a generic data organization model able to represent

[4]Dump refers to input/output operations in raw text, XML and JSON formats.

[CB02] D. CUNNINGHAM, H.AND MAYNARD, V. BONTCHEVA, K.AND TABLAN, "GATE: an architecture for development of robust HLT applications", *in : Proceedings of the Annual Meeting of the ACL*, p. 168–175, 2002.

large multimodal corpora with a wide range of annotation types [CEHK05,CCB+10]. Whereas NXT considers corpora as databases from which data is accessed through a query language, ROOTS lets the user browse data as he sees fit. In a more general approach, UIMA [FL04,FLG+06] proposes software engineering standards for unstructured data management, including annotation and processing. UIMA is technically too advanced for fast and light prototyping. It is rather devoted to industrial developments. In the end, ROOTS is closer to work done within the TTS system Festival [BTCC02]. This system relies on a formalism called HRG, standing for Heterogenous Relation Graphs, which offers a unique representation of different information levels involved in the TTS system [TBC01]. Still, our tool is different from HRG in the sense that the latter is part of the TTS system Festival whereas ROOTS is completely autonomous. Moreover, ROOTS comes along with a true application programming interface (API), in C++ and Perl for the moment.

As a result of recent improvements, ROOTS is now in use in most of the software developed for speech processing, namely the corpus-based speech synthesizer, corpus generation/analysis tools or the phonetizer. Moreover, ROOTS serves as a basis for corpus generation and information extraction for the ANR Phorevox project. For instance, we have built a corpus containing 1000 free e-books which is planned to be proposed to the community. Finally, ROOTS has been registered in 2013 at the Program Protection Agency (*Agence pour la Protection des Programmes*, APP) and publicly released under the terms of LGLP licence on `http://roots-toolkit.gforge.inria.fr`. A paper has been published in the main international language resource conference to let the community know about this release [11].

4.3 Web based listening test system

Participants: Vincent Barreaud, Arnaud Delhay, Sébastien Le Maguer, Damien Lolive.

The listening test platform is developed by the team, especially to evaluate speech synthesis models. This platform has been developed to propose to the community a ready to use tool to conduct listening tests under various conditions. Our main goals were to make the configuration of the tests as simple and flexible as possible, to simplify the recruiting of the testees and, of

[CEHK05] J. CARLETTA, S. EVERT, U. HEID, J. KILGOUR, "The NITE XML Toolkit: Data Model and Query Language", *Language Resources and Evaluation 39*, 4, 2005, p. 313–334.

[CCB+10] S. CALHOUN, J. CARLETTA, J. M. BRENIER, N. MAYO, D. JURAFSKY, M. STEEDMAN, D. BEAVER, "The NXT-format Switchboard Corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue", *Language Resources and Evaluation 44*, 4, 2010, p. 387–419.

[FL04] D. FERRUCCI, A. LALLY, "UIMA: an architectural approach to unstructured information processing in the corporate research environment", *Natural Language Engineering 10*, 3-4, 2004, p. 327–348.

[FLG+06] D. FERRUCCI, A. LALLY, D. GRUHL, E. EPSTEIN, M. SCHOR, J. W. MURDOCK, A. FRENKIEL, E. W. BROWN, T. HAMPP, Y. DOGANATA *et al.*, "Towards an interoperability standard for text and multi-modal analytics", 2006.

[BTCC02] A. W. BLACK, P. TAYLOR, R. CALEY, R. CLARK, "The Festival speech synthesis system", *research report*, University of Edinburgh, 2002.

[TBC01] P. TAYLOR, A. W. BLACK, R. CALEY, "Heterogeneous relation graphs as a formalism for representing linguistic information", *Speech communication 33*, 2001, p. 153–174.

course, to keep track of the results using a relational database.

The most widely used listening tests used in the speech processing community are available (AB-BA, ABX, MOS, MUSHRA, etc.).

This software is currently implemented in PHP and integrated in the Symfony2 framework with Doctrine as database manager and Twig templates. This configuration makes the platform accessible from a wide variety of browsers.

The platform is designed to enable researchers to build wide tests available through the web. The main functionalities provided are as follows:

- Users are given roles, which give them privileges,
- Users answer test during a trial which can be interrupted and resumed later,
- Users give information on their listening conditions at each trial beginning,
- Tests are imported from Zip archives that contain a XML configuration file and the stimulii,
- Users can be imported from a XML configuration file.
- A tester can monitor his test and discard results of a testee on the basis of its statistical behavior.
- The platform is open-source (under AGPLv3 Licence).

4.4 Automatic segmentation system

Participants: Damien Lolive.

The automatic segmentation system consists of a set of scripts aligning the phonetic transcription of a text with its acoustic signal counterpart. The system is made of two parts: the first one generates a phonetic graph including phonological variants (pauses, liaisons, schwas,...), the second one, based on HMM modeling, performs a Viterbi decoding determining the most likely phonetic sequence and its alignment with the acoustic signal.

To be efficient, the system should be applied to texts that have been manually checked (compliance with the recording, spelling, syntax) and annotated. The annotation stage consists in adding tags indicating excepts in foreign language, non standard pronunciation and noises (breathing, laughter, coughing, sniffing, snorting, etc.). It is also possible to improve the decoding performances by adding a list of phonetization of proper names and foreign pronunciations.

4.5 Corpus-based Text-to-Speech System

Participants: Nelly Barbot, Jonathan Chevelu, Arnaud Delhay, David Guennec, Damien Lolive.

For research purposes we developed a whole text-to-speech system designed to be flexible. The system, implemented in C++, intensively use templates and inheritance, thus providing the following benefits:

- the algorithm used for unit selection can be easily changed. For instance, we implemented both $A^\star$ and Beam-search simply by using subclassing and without changing the heart of the system.
- cost functions can also be changed the same way which provides a simple way to experiment new functions.

Moreover the system implements state of the art technique to achieve good performance while manipulated large speech corpora such as hash tables and pre-selection filters. To achieve this, each phone in the corpus is given a binary key which enables A^* to take or reject the unit. Thus, the key contains phonetic, linguistic and prosodic information. Binary masks are used to get access only to the desired information during runtime.

The pre-selection filters are integrated to the hash functions used to access the units in the corpus in order to reduce the number of candidates explored. For the moment, the whole set of filters is the following:

1. Is the unit a Non Speech Sound ?
2. Is the phone in the onset of the syllable?
3. Is the phone in the coda of the syllable?
4. Is the phone in the last syllable of its breath group?
5. Is the current syllable in word end?
6. Is the current syllable in word beginning?

Concretely, the pre-selection filters are relaxed one by one, starting from the end of the list, if no unit corresponding to the current set is found. One drawback is that we can explore candidates far from the target features we want, thus risking to produce artefacts but this backtracking mechanism insures to find a unit and to produce a solution. The priority order of the filters is the one given above.

Finally, high level features are also available to get, for example, the best path or the N-best paths, with a detailed output of the cost values.

Some developments are currently undertaken to provide more features and pre-selection filters and also to improve flexibility of the system to gain a fine control over prosody. This last objective is linked to the main objectives of the team to control expressivity during synthesis.

4.6 Recording Studio

Participants: Nelly Barbot, Vincent Barreaud, Damien Lolive.

A main goal of the Expression project consists in developing high quality voice synthesis. Our research activities use speech corpora as a raw material to train statistical models. A good speech corpus quality relies on a consistent speech flow (the actor does not change his speaking style during a session) recorded in a consistent (and quiet) acoustic environment. In order to expand our research scope, it is often interesting to vary the speech style (dialogs, mood,

accent, etc.) as well as the language style. Unfortunately, such corpora are hard to obtain and generally do not meet specific experimental requirements. To deal with these constraints, speech resources need to be recorded and controlled by our own protocols.

4.6.1 Hardware architecture

The funding of this recording studio comes from MOB-ITS (CPER, 2007-2013). The MOB-ITS platform (Mobile and interactive access to data) is a joint project of IRISA teams in Lannion (IUT and ENSSAT). This contract is part of the support to the "Pôle de compétitivité Images & Réseaux".

This recording studio consists in two rooms: an isolation booth and control room.

The isolation booth can fit three persons. It is designed to attenuate the noises of 50dB and is equipped with two recording sets. A recording set consists in a high quality microphone (Neumann U87AI), a high quality closed head set (Beyer DT 880 250ohms), a monitor and a webcam.

The control room is equipped with two audio networks, a video network and computer network. The first audio network is a high quality digital recording line going from the isolation booth microphones to a digital sound card through a preamplifier (Avalon Design AD2022), an equalizer (Neve 8803 Dual Channel) and finally an analogic/digital converter (Lynx Aurora 8). The digital sound is edited with a logical sampling table (Avid Pro Tools).

In addition to the signal issued by the isolation room, the digital audio network can record the signals from an Electro-Gloto Graph (EGG) that capture the glottal activity of the actor. This activity is used to induce the F0 (first formant) trajectory which is the main indicator of the prosody. This activity must be digitalized and recorded along with the audio activity in order to reduce the latency between the two signal.

The second audio network is for control purpose and is fully analogic. It is used by the operator to control the quality of the recorded sound, the consistency of the actor, the accuracy of the transcription. An actor can receive audio feedback of his own voice, disturbing stimulii (music, other voices, their own delayed voice) or directions from the operator through this audio line. This network consists in four Neumann KH 120 loud-speakers (two in the booth, two in the control room), a head set amplifier (ART headamp 6 pro) and an analogic sampling table (Yamaha MG206C). The computer network stores the recording sessions scenarii and prompt the actor.

The video network switches the video output (computers, webcam) to screens installed in the isolation booth (for prompting) and the control room (for monitoring).

4.6.2 Software architecture

Actors in the isolation booth must be prompted to utter speech with various indications (mood, intonation, speed, accent, role, ...). The prompt must be presented on the simplest interface, for instance a lcd screen or a tablet. The latest developments on the recording studio consist in a software implemented on the computer at the end of the digital audio network that record sound files, segment them and link them to the transcription. This software is controlled by the operator who checks that the actors actually uttered the prompted sentence and the quality

of the recording. Thus, the operator can possibly reject (in fact, annotate) a file and prompt the actors again with the discarded sentence.

The digital sound card used for recording only offers microsoft drivers. Consequently, this software has been developed with the Windows Audio and Sound API (WASAPI). The main difficulty resides in the simultaneous recording of two distinct channels (audio and EGG) without any jitter between the two signals.

5 Contracts and Grants with Industry

5.1 SIGN3D

Participants: Sylvie Gibet, Ludovic Hamon

.

The *SIGN3D* project was funded by the French Ministry of Industry (Program "Investissements d'avenir": usages, services and innovative contents). Three partners were participating to this project: two companies (Mocaplab: team leader, and Websourd), and one academic laboratory (IRISA-Expression). The project *SIGN3D* aimed at creating a range of innovative tools for the recording and the editing of captured motion of French Sign Language (LSF) content. The challenge was to design a complete workflow from the movement capture (including body and hand movements, facial expressions and gaze direction) to the restitution using concatenate synthesis applied on a 3D virtual signer. The project ended in October 2014. It enabled the collaboration between experts in the creation of a linguistic sign language corpus (Websourd) and the capturing of a high definition full-body mocap corpus (including synchronized channels such as facial expression, hand movements, gaze direction).
`http://sign3d.websourd.org/`

5.2 INGREDIBLE

Participants: Pamela Carreño, Sylvie Gibet, Pierre-François Marteau.

The *INGREDIBLE* project project is funded by the French Research Agency (program ANR CONTINT. The partners are: LabSTICC (team leader), LIMSI-CNRS, IRISA, Virtualys, Final users (DEREZO, Brest; STAPS lab., Orsay.

The goal of the INGREDIBLE project is to propose a set of scientific innovations in the domain of human/virtual agent interaction. The project aims to model and animate an autonomous virtual character whose bodily affective behavior is linked to the behavior of a human actor. The outcomes of the project are:

- The creation of different motion corpora (fitness, interactive video games motions, theatrical gestures);
- The development of algorithms for affect recognition from expressive features;
- The development of a behavioral system linked to the recognition and synthesis modules;

- The development of synthesis algorithms for animating a virtual character in interactive situations.

5.3 PHOREVOX

Participants: Nelly Barbot, Jonathan Chevelu, Damien Lolive.

Expression is leader of a ANR CONTINT project named Phorevox and accepted on the 16th december, 2011. Phorevox aims to promote the use of high quality speech synthesis to help in learning french. The consortium is made of IRISA/Expression, Voxygen (France Telecom spin off), CREAD (didactic and tests), LLF Paris VII (linguistics) and Zeugmo (web platform).

The project has official begun the 1st of june 2012 and finished on 30th of november 2014. During the first semester of the project, the main means deployed for the project has concerned the organisation of the project as well as the deployment of collaborative tools. We also have recruited a research engineer for the whole length of the project. His role is to work on the problematics concerning the IRISA contribution, to assure the project coordination and the technical support on collaborative tools. One of the first consortium actions was to build up tools simplifying exchanges between the group members and to improve project visibility.

Practically, some difficulties to recruit an engineer have postponed the real technical start of the project. This delay has been taken into account to postpone the end of the project too (6 months accepted by ANR).

At the end of November 2013, a mid-term presentation has been done in front of the ANR committee. During this presentation, we have shown a demonstration video of the prototype with several exercices (dictation, phonological opposition, word segmentation) including an adapted speech synthesis voice.

During the last year of the project, we have now carried on the development of the project with some experiments in real conditions to assess platform ergonomy and to have the first feedbacks concerning the suitability of the approach to help children learn phonology. The main outcomes of the project are that an adapted speech synthesis system is applicable in a learning context, the use of the oral modality helps improving writing skills. To achieve these results, we have developed a platform integrating a dynamic vocal feedback to students, exercices adapted to language learning for pupils (e.g. dictation) and reports targeted at teachers.

The main perspectives of the project concern language learning in the case of foreign learners or native adults wanting to improve their writing skills. On this topic, we have submitted a new ANR project.

6 Other Grants and Activities

6.1 International Collaborations

- **Horizon 2020 project proposal:** The team collaborated in writting a European project proposal with Orange Labs since October 2013. This proposal has been submitted as part of the ICT 2014 call of the Horizon 2020 Framework Programme of the

European Commission. Precisely, the consortium gathered the following partners (and countries): Athens International Airport (Greece), Athens University of Economics and Business (Greece), CEA (France), Fondazione Bruno Kessler (Italy), GeoMobile (Germany), Idiap Research Institute (Switzerland), and Orange Labs (France). The addressed topic is "multimodal and natural computer interaction" and focuses on smart conversational virtual assistants in international transport hubs and complex environments with an emphasis on multimodality and multilinguality. The EXPRESSION team would have addressed the speech synthesis issues within this project. Unfortunately, the project has not been accepted.

6.2 National Collaborations

- **Hybride ANR Project** Participant: Nicolas Béchet. The Hybride Research Project aims at developing new methods and tools for supporting knowledge discovery from textual data by combining methods from Natural Language Processing (NLP) and Knowledge Discovery in Databases (KDD). The consortium is made of INRIA/LORIA, GREYC, MoDyCo, Inserm (IRISA is associated to the GREYC in this project). The coordinator is Yannick Toussaint from INRIA/LORIA.

- **ANIMITEX CNRS MASTODONS Project** Participant: Nicolas Béchet. The ANIMITEX Project aims is to exploit the massive and heterogeneous textual data to provide crucial information in order to complete the analysis of satellite images. The consortium is made of LIRMM, TETIS, ICUBE, GREYC, LIUPPA, IRISA. The coordinator is Mathieu Roche, Cirad/TETIS.

7 Dissemination

7.1 Involvement in the Scientific Community

- Pierre-François Marteau served as a reviewer in international journals (IEEE TPAMI, IEEE TNNLS, IEEE TKDE, PRL). He serves as an expert for French Ministry of Research (CIR/JEI) and ANRT (CIFRE). He was member of a thesis committee at Nantes University, LINA. He is member of the Strategic Orientation Committee at IRISA and member of the scientific committee at Université de Bretagne Sud.

- Sylvie Gibet serves as a reviewer for the Journal of Universal Access in the Information Society, the IEEE Transactions on Affective Computing. She has served as a reviewer for the first and second Workshop on Motion Computing (MOCO), and for the Workshop on Dance Notations and Robot Motion (LAAS, Toulouse, November 2014, France). She has been a reviewer for the HDR thesis of Lionel Reveret (INRIA-Morpheo, May 2014), and an examiner for the HDR thesis of Frederic Bevilacqua (IRCAM, November 2014).

- Arnaud Delhay is an elected member of the 'Commission Recherche' (Research comittee) of the IUT of Lannion.

- Damien Lolive is an elected member of the 'Conseil Scientifique' (Scientific council) of ENSSAT, Lannion. He also serves as a reviewer for the IEEE Transactions on Speech and Language processing, for the *Traitement Automatique des Langues* journal, for the International conference of the International Speech Communication Association (Interspeech) and for the "*Journées d'Études sur la Parole*".

- Gwénolé Lecorvé is an elected member of the laboratory council of IRISA. He also serves as a reviewer for the International conference of the International Speech Communication Association (Interspeech), and for the "*Journées d'Études sur la Parole*".

- Caroline Larboulette is a member of various program committees for international conferences (CAe/EXPRESSIVE 2014, CASA 2014), a member of the editorial review board of the international journal of computer graphics and creative interfaces (IJCICG) and serves as a reviewer for various journals (Computer & Graphics, TVCG, CAVW). She is a member of the ACM SIGGRAPH Small Conferences Committee that attributes the ACM SIGGRAPH labels to conferences and supervises the budget of conferences sponsored by ACM SIGGRAPH. She also served on the steering committee of the ACM SIGGRAPH / Eurographics Symposium on Computer Animation from 2010 to 2014. She is also vice-president of the ACM SIGGRAPH Madrid Professional Chapter that she created in 2007.

7.2 Teaching

- Arnaud Delhay teaches databases and web programming (server- and client-side) in Licence levels at IUT of Lannion, calculability and computational complexity of problems in Master level and web server-side programming in Licence level at École National Supérieure des Sciences Appliquées et de Technologie (ENSSAT).

- Nicolas Béchet teaches various computer sciences courses at the Statistique et Informatique D'ecisionnelle department of IUT Vannes.

- Sylvie Gibet teaches the following computer science courses at the faculty of sciences, Université de Bretagne Sud: functional programming and algorithmic in License level, and multimedia digital processing and computer animation in Master level (master WMR).

- Jean-François Kamp teaches human-computer interaction, programming at the computer science department of IUT Vannes. He is responsible for student internships.

- Gwénolé Lecorvé teaches the following computer science courses at École National Supérieure des Sciences Appliquées et de Technologie (ENSSAT): Java, graphical interfaces, C++; distributed algorithmics; and artificial intelligence in Master level.

- Damien Lolive teaches the following computer science courses at École National Supérieure des Sciences Appliquées et de Technologie (ENSSAT): object-oriented programming in Licence Level, and XML, compilers architecture and formal languages theory in Master Level.

- Pierre-François Marteau teaches programming languages, logics, introduction to cryptography and information retrieval courses in computer sciences License and Master levels, mostly at Ecole Nationale Supérieure de Bretagne Sud.
- Gildas Ménier various computer sciences courses at the faculty of sciences, Université de Bretagne Sud
- Jeanne Villaneau teaches various computer sciences courses at Ecole Nationale Supérieure de Bretagne Sud.

7.3 Conferences, workshops and meetings, invitations

- Pierre-Francois Marteau has given an invited talk entitled "From time elastic distances to time elastic kernels" during the AMA (dAta analysis, Modeling and mAchine learning) Team seminar at Laboratoire d'Informatique de Grenoble (LIG), universitÃ© de Grenoble Alpes.

7.4 Graduate Student and Student intern

8 Bibliography

Doctoral dissertations and "Habilitation" theses

[1] G. Ke, *Comparability measures for the assisted construction of bilingual thematic comparable corpora*, Theses, Université de Bretagne Sud, February 2014, `https://tel.archives-ouvertes.fr/tel-00997837`.

Articles in referred journals and book chapters

[2] N. Béchet, J. Chauché, V. Prince, M. Roche, "How to Combine Text-Mining Methods to Validate Induced Verb-Object Relations?", *Computer Science and Information Systems 11*, 1, 2014, p. 133–155, `http://hal-lirmm.ccsd.cnrs.fr/lirmm-01054918`.

[3] P. Carreno, S. Gibet, P.-F. Marteau, "Synthèse de mouvements humains par des méthodes basées apprentissage : un état de l'art", *Revue Electronique Francophone d'Informatique Graphique Vol. 8*, N° 1, 2014, p. 1–19.

[4] T. Le Naour, N. Courty, S. Gibet, "Utilisation des relations spatiales pour la reconstruction de trajectoires de marqueurs issues de la capture de mouvement", *Revue Electronique Francophone d'Informatique Graphique Vol. 8*, N° 2, 2014, p. 55–65.

[5] P.-F. Marteau, S. Gibet, "On Recursive Edit Distance Kernels with Application to Time Series Classification", *IEEE Transactions on Neural Networks and Learning Systems*, June 2014, p. 1–14, 14 pages, `https://hal.archives-ouvertes.fr/hal-00486916`.

[6] L. Miclet, N. Barbot, B. Jeudy, "Analogical Proportions in a Lattice of Sets of Alignments Built on the Common Subwords in a Finite Language", *in : Computational Approaches to Analogical Reasoning: Current Trends, Studies in Computational Intelligence*, H. Prade and G. Richard (editors), Springer-Verlag Berlin Heidelberg, 2014, p. 245–260, `https://hal.inria.fr/hal-00974656`.

Publications in Conferences and Workshops

[7] J.-Y. Antoine, J. Villaneau, A. Lefeuvre, "Weighted Krippendorff's alpha is a more reliable metrics for multi-coders ordinal annotations: experimental studies on emotion, opinion and coreference annotation.", *in : The 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, A. for Computational Linguistics (editor), p. E14–1058, Gothenburg, Sweden, April 2014, https://hal.archives-ouvertes.fr/hal-01001811.

[8] M. Avanzi, G. Christodoulides, E. Delais-Roussarie, N. Barbot, D. Lolive, "Towards the Adaptation of Prosodic Models for Expressive Text-To-Speech Synthesis", *in : Interspeech*, ISCA, Singapore, Singapore, September 2014, https://hal.inria.fr/hal-01133316.

[9] P. Carreno, S. Gibet, C. Larboulette, P.-F. Marteau, "Corpus Creation and Perceptual Evaluation of Expressive Theatrical Gestures", *in : Intelligent Virtual Agents*, Boston, United States, August 2014, https://hal.archives-ouvertes.fr/hal-01165575.

[10] J. Chevelu, G. Lecorvé, D. Lolive, "ROOTS : un outil pour manipuler facilement, efficacement et avec cohérence des corpus annotés de séquences", *in : Journées d'Etude sur la Parole (JEP)*, Le Mans, France, June 2014, https://hal.inria.fr/hal-00975897.

[11] J. Chevelu, G. Lecorvé, D. Lolive, "ROOTS: a toolkit for easy, fast and consistent processing of large sequential annotated data collections", *in : Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland, May 2014, https://hal.inria.fr/hal-00974628.

[12] E. Delais-Roussarie, D. Lolive, H. Yoo, N. Barbot, O. Rosec, "Adapting prosodic chunking algorithm and synthesis system to specific style", *in : Interspeech*, ISCA, Singapore, Singapore, September 2014, https://hal.inria.fr/hal-01133319.

[13] D. Guennec, D. Lolive, "Unit Selection Cost Function Exploration Using an A* based Text-to-Speech System", *in : International Conference on Text, Speech and Dialogue (TSD)*, Brno, Czech Republic, September 2014, https://hal.inria.fr/hal-01133321.

[14] D. Guennec, D. Lolive, "Utilisation d'un algorithme A* pour l'analyse de la sélection d'unité en synthèse de la parole", *in : JEP - 30ème édition des Journées d'Etudes sur la Parole*, Le Mans, France, June 2014, https://hal.inria.fr/hal-00977468.

[15] M. Harzallah, G. Berio, T. Gherasim, P. Kuntz-Cosperec, "Ontology Quality Problems - An Experience with Automatically Generated Ontologies ", *in : Knowledge Engineering and Ontology Development (KEOD)*, Rome, Italy, October 2014, https://hal.archives-ouvertes.fr/hal-01165230.

[16] G. Ke, P.-F. Marteau, G. Ménier, "Variations on quantitative comparability measures and their evaluations on synthetic French-English comparable corpora", *in : LREC 2014, the 9th edition of the Language Resources and Evaluation Conference*, p. pp, Reykjavik, Iceland, May 2014, https://hal.archives-ouvertes.fr/hal-00995294.

[17] G. Ke, P.-F. Marteau, "Co-clustering of bilingual datasets as a mean for assisting the construction of thematic bilingual comparable corpora", *in : The 9th edition of the Language Resources and Evaluation Conference, LREC 2014*, p. pp, Reykjavik, Iceland, May 2014, https://hal.archives-ouvertes.fr/hal-00995297.

[18] S. Le Maguer, E. Delais-Roussarie, N. Barbot, M. Avanzi, O. Rosec, D. Lolive, "Algorithme de découpages en groupes prosodiques pour la dictée par l'usage de synthèse vocale", *in : Journées d'études sur la parole*, Le Mans, France, June 2014, https://hal.archives-ouvertes.fr/hal-01137707.

[19] S. Le Maguer, E. Delais-Roussarie, N. Barbot, M. Avanzi, O. Rosec, D. Lolive, "Prosodic chunking algorithm for dictation with the use of speech synthesis", *in : Proc. of Speech Prosody*, Dublin, Ireland, May 2014, https://hal.inria.fr/hal-00973866.

[20] P.-F. Marteau, S. Gibet, C. Reverdy, "Down-Sampling coupled to Elastic Kernel Machines for Efficient Recognition of Isolated Gestures", *in : ICPR 2014, International Conference on Pattern Recognition*, IAPR (editor), IEEE, p. pp, Stockholm, Sweden, August 2014, https://hal.archives-ouvertes.fr/hal-00995279.

[21] P.-F. Marteau, S. Gibet, C. Reverdy, "Sous échantillonnage et machine à noyaux élastiques pour la classification de données de mouvement capturé", *in : 14 emes journées Extraction et Gestion des Connaissances, EGC 2014*, *RNTI*, *E-26*, RNTI, p. 179–190, Rennes, France, January 2014, https://hal.archives-ouvertes.fr/hal-01003033.

[22] L. Miclet, N. Barbot, H. Prade, "From analogical proportions in lattices to proportional analogies in formal concepts", *in : ECAI - 21th European Conference on Artificial Intelligence*, Prague, Czech Republic, August 2014, https://hal.inria.fr/hal-01000314.

[23] G. Sami, N. Béchet, G. Berio, "Ontologies from Textual Resources: A Pattern Based Improvement Using Deep Linguistic Information", *in : Workshop on on Ontology and Semantic Web Patterns (WOP)*, *Proceedings of Workshop on on Ontology and Semantic Web Patterns (WOP)*, *1302*, CEUR-WS.org, p. 14–25, Riva del garda, Italy, October 2014, https://hal.archives-ouvertes.fr/hal-01118948.

[24] H.-H. Vu, J. Villaneau, F. Saïd, P.-F. Marteau, "Sentence Similarity by Combining Explicit Semantic Analysis and Overlapping N-Grams", *in : Text, Speech and Dialogue*, Springer (editor), *8655*, p. 201–208, Brno, Czech Republic, September 2014, https://hal.archives-ouvertes.fr/hal-01066170.

[25] H. Yoo, S. Le Maguer, E. Delais-Roussarie, N. Barbot, D. Lolive, "Evaluation d'un algorithme de chunking appliqué à la dictée", *in : JEP - 30ème édition des Journées d'Etudes sur la Parole*, Le Mans, France, June 2014, https://hal.inria.fr/hal-00977445.

Internal Reports

[26] R. Qader, G. Lecorvé, D. Lolive, P. Sébillot, "Phonology Modelling for Expressive Speech Synthesis: a Review", *Research Report number PI-2020*, University of Rennes 1, July 2014, https://hal.inria.fr/hal-01021911.