



MARIE SKLODOWSKA-CURIE ACTIONS

Co-funding of regional, national and international programmes (COFUND)

DOC2AMU PROJECT 2017 CALL FOR APPLICATIONS

Language Underpinnings of Conversational Interpersonal Dynamics (LUCID)

1. DESCRIPTION OF THE PHD THESIS PROJECT

1.1 OBJECTIVES OF THE PROJECT BASED ON THE CURRENT STATE OF THE ART

This PhD thesis subject aims at understanding how inter-individual relationships are established, managed and ultimately evolve during conversations. Spoken social interaction, such as every day conversations and task-oriented dialogues, remains the most frequent natural language usage. While every day conversations have been described as a crucial activity of human kind constituting a root of human sociality (Einfeld & Levinson, 2006), task-oriented dialogues are crucial in human language technologies.

Interactional profiles (“leader”, “follower”...) in terms of social roles either provided by a task (Pardo et al, 2010) or emerging from it (Valente & Vinciarelli, 2011) have been now investigated, and recently in a unsupervised fashion on *big data* sets (Li et al, 2017). However, little is known about the properties of conversational dyads and the interpersonal relationships dynamics in conversation. Some studies have exhibited general patterns such as convergence (Manson et al, 2013) but there has been no attempt to propose a classification of such patterns systematically. One of our starting points is that one should not neglect linguistics as a source of knowledge about language, the most basic tool to exchange information, act together and build social relationships. Moreover, linguistic knowledge of language can be exploited to set-up metrics capturing the properties of conversations and the situations they occur in, of individuals, and finally of interpersonal relationships and their dynamics. Finally, natural language processing techniques provide methodological knowledge for the scientific exploration of language through *Big Data* sets.

This project proposes to use data science on large conversational datasets to identify the linguistic traces of interpersonal relationships. These relationships can have a static

nature (*conviviality, confrontation...*) or a dynamic one (*convergence, divergence, ignorance...*). This work is based on two kinds of *Big Data* resources. On the one hand, the PhD candidate will opportunistically exploit existing large social interaction corpora in English, French and Mandarin (several million words for each language coming from *Language Data Consortium* --English and Mandarin--, *Academica Sinica* --Mandarin-- and *Ortolang* repositories). On the other hand, the customer-relationship task-oriented dialogue dataset produced and used by our private sector partner will provide an industry-focused counterpart to the other more conversational dataset. This approach will guarantee to include enough resources as well as provide more foundational impact of the results in language sciences and human language technologies. The PhD student will establish the relevant metrics to characterize interactions (formality, time pressure...), interactants (according to their interactional profile such as "attractor", "sloppy", "aligner", "neutral"...), and dyads ("converging", "aligned"...). The metrics investigated will include paralinguistic (voice parameters in our case) and linguistic features (from lexicon to discourse and interactional structure).

References

- Enfield, Nicholas J., and Stephen C. Levinson. "Roots of human sociality." *New York: Berg* (2006).
- Li, Yanxiong, et al. (2017) "Unsupervised classification of speaker roles in multi-participant conversational speech." *Computer Speech & Language* 42 (2017): 81-99.
- Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, 34(6), 419-426.
- Pardo, Jennifer S., Isabel Cajori Jay, and Robert M. Krauss. "Conversational role influences speech imitation." *Attention, Perception, & Psychophysics* 72.8 (2010): 2254-2264.
- Valente, F., & Vinciarelli, A. (2011). Language-Independent Socio-Emotional Role Recognition in the AMI Meetings Corpus. In *INTERSPEECH* (pp. 3077-3080).

1.2 METHODOLOGY

The project combines three main methodologies :

- Linguistic Analysis of Conversational Data
- Natural Language and Speech Processing of *Big Data* sets
- Machine Learning (Supervised and Unsupervised)

Linguistic Analysis of Conversational Data is required to understand the nature of the features being extracted for the data analysis as well as for the error analysis of the intermediate systems. More precisely we have to be able to propose and adapt the features (in particular phonetic and acoustic but also lexical or syntactic for example) and how precisely we can extract them from both transcribed and untranscribed datasets. In a later stage, based on the output of our clustering experiments but also on linguistic expertise in conversations, discrete categories will be determined. Finally systematic evaluation and qualitative error analysis are required at each processing phase.

Natural Language and Speech Processing (NLSP) are required to efficiently extract the features and prepare the data set for the data analysis phase. Two main NLSP pipelines have to be created : (i) one for the transcribed data that can rely on orthographic transcription, (ii) one for the untranscribed data that will rely more on acoustic features

but on which we will also run *Automatic Speech Recognition* in order to be able to extract, noisy but potentially useful, linguistic features.

Machine Learning will be used in the two classical flavors. Clustering (unsupervised machine learning) will be apply to our *big data* processed by our NLP pipelines and guided by linguistic considerations. This step will allow to understand better the space in which dyads and conversations can be represented and therefore constitute an important step in the identifications of reliable categories in this space. In a second time, a classifier (supervised machine learning) will be trained to identify the categories determined as explained above. Depending on the structure of the data manipulated at this point, this step will either be performed in a classic supervised way or through an active learning approach. Moreover two classification modes of the system can be distinguished, a real time version that will only use information available in real time during the interaction (and therefore usable for monitoring applications) and an off line mode that will process a conversation as a completed object.

1.3 WORK PLAN

During the first phase, the PhD student will conduct a careful review of linguistic and social signal processing literature, and perform in parallel some smaller scale pilot studies and evaluations in order to be as precise and efficient as possible in terms of relevant features to extract. The second phase will be devoted to unsupervised analysis of the different *Big Data* corpora. Outputs, such as clusters of interactions, participants or dyads, will be partially evaluated in collaboration with our international networks and with the private sector partner. The results will be a categorization of social interaction grounded on Big Conversational Data analysis. The third phase will be devoted to the development of a classification system to predict the category of individual interactions.

Activities	M0-6	M6-9	M9-15	M15-18	M18-21	M21-27	M27-30	M30-33	M33-36
Reading	Litterature Review				Follow-up				Follow-up
Analysis	Pilot Studies		Error Analysis		Pilot Studies	Error Analysis		Error Analysis	
Processing	Pre-processing	Clustering			Evaluation	Classification		Evaluation	
Writing		Posters	Workshops		Conferences and Journals		Manuscript		

Table : Program of the PhD (Blue : Exploration ; Yellow:Analysis ; Green:Production)

SUPERVISORS AND RESEARCH GROUPS DESCRIPTION

There is not other funds for this specific research. However it should be noted this project is related to the CoBra (Conversational Brains) network creation. The main difference of LUCID is the absence of brain-related study and the direct integration of local industrial partner.

2. 3I DIMENSIONS AND OTHER ASPECTS OF THE PROJECT

INTERDISCIPLINARY DIMENSION

This Phd is **interdisciplinary** through the deep integration of knowledge and methods coming from Humanities (mainly Linguistics and Social Psychology) and Computer Science (machine learning, data mining). It contributes to the emergence of a transdisciplinary framework.

In terms of supervision task division, Laurent Prévot provides the expertise on the linguistic analysis aspects, the research questions as well as the evaluation and the error analysis of the systems. Benoît Favre brings its expertise on all the machine learning aspects but also on speech processing. Both are experts of Natural Language Processing and both are at least familiar with all the aspects of the project ensuring a good coherence in terms of supervision. The labs contribution are aligned with the co-supervisors competences. LPL will provide the perfect linguistic environment for performing this work thanks, in particular, to a focus on empirical and in particular big corpus linguistics. LIF provides a strong environment both in terms of Natural Language Processing and Machine Learning.

2.2 INTERSECTORAL DIMENSION:

The industrial partner is Digifrance, a SME located in St-Remy-de-Provence specialized in call-centres for customer relations and emergency services. They process each year several millions of task-oriented conversations. This testbed will be an occasion to validate the models developed by the PhD candidate outside of lab conditions, and it will be an opportunity for the company to explore analytics components for leveraging interpersonal dynamics and eventually integrate them in their products. The industrial LDCpartner will be involved in a PhD steering committee every 6 months and help the PhD candidate set-up experiments on their customer data.

The project is **intersectorial** both from a very concrete perspective with the implication of a local company specialized in call-centres for customer relations and emergency services. Its intersectoriality can also be evaluated through the composition of the research community involving a large number of professionals from large IT companies and start-ups in the main associations of the field and in the scientific committees of its main conferences.

INTERNATIONAL DIMENSION:

This project is **international** by its research object: we are working simultaneously on corpora of languages spoken through the world (and actually recorded in Europe, US and Asia). The overall field in which the project locates, speech and language processing, is a very active internationalized domain as illustrated by the size and the nature of its major conferences (Interspeech, ACL...) that gather about 1000 researchers from all over the world. The target audiences for the productions of the PhD applicant are the top-tier international conferences in natural and language processing. Finally, the recruitment of the PhD candidate will be international by advertising the grant on all the major channels available.

3. RECENT PUBLICATIONS

- Prévot, L., J. Gorisch, and R. Bertrand (2016). A CUP of CoFee: A large Collection of feedback Utterances Provided with communicative function annotations. In: *Proceedings of 10th Language Resources and Evaluation Conference*. Portoroz.
- Prévot, L., Tseng, S. C., Peshkov, K., & Chen, A. C. H. (2015). Processing Units in Conversation: A Comparative Study of French and Mandarin Data. *Language and Linguistics*, 16(1), 69-92.
- Prévot, L., J. Gorisch, and S. Mukherjee (2015). Annotation and Classification of French Feedback Communicative Functions. In: *Proceedings of Pacific Asia Conference in Language, Information and Computation*. Shanghai.
- Balamurali A R, F. Béchet, B. Favre, "CallAn: A Tool to Analyze Call Center Conversations", International Workshop on Spoken Dialogue Systems (IWSDS), 2016.
- Frederic Bechet, Alexis Nasr, Benoit Favre, "Adapting dependency parsing to spontaneous speech for open domain spoken language understanding", Interspeech, Singapore, 2014
- Benoit Favre, Dilek Hakkani-Tür, "Phrase-level and Word-level Strategies for Detecting Appositions in Speech", Interspeech, Brighton (UK), 2009
- Sebastien Delecraz, Frederic Bechet, Benoit Favre, Mickael Rouvier, "[Fusion d'espaces de représentations multimodaux pour la reconnaissance du rôle du locuteur dans des documents télévisuels](#)", Actes de la conférence JEP 2016, 2016.

4. EXPECTED PROFILE OF THE CANDIDATE

We are looking for a candidate coming from either a linguistic or a computer science background. All the tasks of the project can be adapted to these two profiles. However, there are a few requirements for both profiles :

Linguistic Background Requirements

- Experience with Statistics or Quantitative Corpus Linguistics

- Experience in handling speech (Phonetics)

Computer Science Background Requirements

- Some experience with Machine Learning
- Some experience with text and speech processing

Interests required (as much as possible):

- Language structure, Social interaction, Cross-linguistic comparison
- Dialogue / Conversational corpora, Multilingual corpora
- Big Data, Data science, Digital Humanities

5. SUPERVISORS' PROFILES

Laurent Prévot is a full professor in Language Sciences at Aix Marseille Université and a research fellow at the Speech and Language laboratory (Laboratoire Parole et Langage). His original interests in linguistics are the semantic/pragmatic interface (or the study of linguistic meaning and of its usage for communicative activities) and natural language processing. Recently, he has been working on semantic and interactional analysis of spontaneous conversations. Since his arrival (2008) at Aix Marseille Université, he has coordinated an Erasmus Mundus Action 2 mobility program between Europe and South East Asia (2010-2015), an ANR « young researcher » project (2012-2015), two international collaboration programs (with Hong Kong and Taiwan) and has participated in many other funded researcher projects. He is a member of the Brain and Language Research Institute (BLRI), the Institute Language, Communication and the Brain (ILCB), the EQUIPEX Open Resources and TOOLS for LANGuage (Ortolang) and of the Institut Universitaire de France (IUF). He has also participated to 6 other ANR projects. He is co-author of more than 80 publications and a frequent reviewer for many conferences (ACL, COLING...) and journals (Lingua, Language Resources and Evaluation...).

Laurent Prévot obtained his « habilitation » in 2014 and co-supervises two PhD students. One of them will defend his PhD in Fall 2017 (5 years) and the other one in Fall 2017 (4 years). He has directed 4 master students both in Natural Language Processing and Linguistics and currently supervises another.

Benoit Favre is an Associate Professor at Aix-Marseille Université, where he was hired in 2010. He is a member of the natural language processing group at the Fundamental Computer Science (LIF) lab in Marseille. He obtained his thesis, titled "Speech summarization for efficient access to audio databases" from University of Avignon in France in 2007. He was a postdoc at UC Berkeley until 2009 and at University of Le Mans, in France until 2010. His research topic is automatic understanding of natural language with a focus on non-canonical language such as spontaneous speech and social media. He is also interested in multimedia understanding and multimodal fusion. He has participated to seven ANR projects, one A*midex transfer project, an EU FP7 project, and a DARPA project. He is a member of the *Brain and Language Research Institute* (BLRI). He is co-author of more than 80 publications, member of IEEE SPS, IEEE SLTC, ISCA, AFCEP and