Post-doctoral position Formal and statistical modeling of dialogue
application to pathological data

One postdoc position (1 year) in Natural Language Processing / Machine Learning is open in the Semagramme and SyNaLP team at LORIA and at ATILF. This position will be funded by the impact projet OLKI (Open Language and Knowledge for Citizens) from Université de Lorraine.

Informations

— Starting date : fall 2019
— Duration : 1 year
— Deadline for application : September 09th, 2109 (Interviews end of September)
— Location : Nancy [1], France
— Salary : around 2,000 euros per month net income
— Informal inquiries can be sent by email to Maxime Amblard (maxime.amblard@loria.fr) and Chloé Braud (chloe.braud@loria.fr).
— Application : CV, motivation letter, PhD evaluation, master TOR and support letter(s)

Supervisors

— Maxime Amblard, MCF HDR Univ. de Lorraine, Loria UMR 7503 - équipe Sémagramme
maxime.amblard@univ-lorraine.fr
— Chloé Braud, CR CNRS, Loria UMR 7503 - équipe Synalp
chloe.braud@loria.fr
— Michel Musiol, Pr Univ. de Lorraine, ATILF UMR 7118
michel.musiol@univ-lorraine.fr

Scientific environment

The project is proposed as part of a new collaboration between the supervisors. However, it is fully integrated into the SLAM project on the formal modelling of interviews of schizophrenic patients. One of the main members of this project will be involved in this post-doctoral work, Manuel Rebuschi, MCF HDR Université de Lorraine, AHP-PReST UMR 7117 (manuel.rebuschi@univ-lorraine.fr).

Note that the post-doc will be integrated in the INRIA Exploratory Action ODiM, led by Maxime Amblard.

1. https://www.nancy.fr/nancy-in-english/discover/living-in-nancy-1218.html

Keywords
NLP, Discourse and Dialogue, machine learning, logic, corpora, natural language, pathology

Scientific project

Modeling interaction is a crucial step for Natural Language Processing (NLP), which requires the development of automatic tools able to simulate these exchanges. A typical example is chatbots and all the services based on them. But Dialogue Models face two types of difficulties.

(I) The first issue concerns the availability of resources and models that can analyze and process dialogues. Modelling dialogues is very hard, in particular because conversations highlight particular uses such as the relationship between questions and answers. Recent models show that this issue is beyond semantics understanding, and even more so beyond discourse. However, it shares with the latter the involvement of formal methods (based on the principle of compositionality), of distributional semantics [7], and of machine learning. Most of formalisms for natural languages are based, in one way or another, on a notion of state change that is used to model the dynamic phenomena. As a consequence, it is difficult to use standard tools of mathematical logic at the level of the discourse interpretation. A type-theoretic way of rebuilding DRT (and its variants [8,9]) and dynamic logic [5] has been proposed. This proposal, which is based on Church's simple theory of type [3], takes advantage of the notion of continuation in order to allow quantifiers to dynamically extend their scopes. In particular, these continuations could be adapted to the dynamics of dialogue, the semantics and pragmatics.

Another perspective is to use Machine Learning approaches in order to identify dialogical relations and dialogical interactions. Generally speaking discursive analysis aims at building a structure representing the semantic links between sentences.

Automatically deriving this structure represents a challenge especially when the links are implicit, as proven for monologues [12]. In dialogue, an additional difficulty is that we need semantic links between speech acts. Annotated data are very scarce for dialogues, with one corpus developed for English [2]. Consequently, only a few automatic systems exist [1,10] with performance probably limited by the few amount of data. Semi-supervised or transfer learning methods could help to identify such kind of relations, based on data existing for monologues and on constraints identified through the formal modeling. More annotation is still needed at least for evaluating the models.

(II) The second is that dialogue models must be coordinated with pragmatic inferences at a higher level. In this case, we can refer to linguistic models of dialogue such as [6], or to models that capture conceptual links, such as in TTR [4]. While speech models provide important information, dialogue makes it possible to share information in a more sophisticated way. The solution must take into account the background of all speakers, as well as how they have common ground. To this end, the candidate will propose a dialogue model to structure the different necessary linguistic informations for interaction. This model will be implemented in a tool that finely manages interaction through formal and learning strategies. As a result, the development of multilingual annotated resources will be integrated into the post-doc work.

In particular, we propose to apply these approaches to the SLAM project corpus. The SLAM project links thought and language disorders. This corpus consists of transcriptions

of interviews with schizophrenic patients. Exchanges with these patients contain interactions that are difficult to interpret semantically and interactively. These sequences have been defined as characteristic in [11]. After a detailed formal study of these interactions, we wish to automatically experiment these exchanges to highlight diagnostic clues on the one hand and the automatic characterization of the manifestation of the dysfunction on the other.

Candidate skill

The post-doctoral fellow must have a PhD in Computational Linguistics/NLP, Computer Science or related fields, with a good programming skills. He/She must be fluent in English and have demonstrated its ability to publish at the highest international level. Note that Knowledge of French is NOT a requirement.

Supervision of students is possible, if wanted.

Supervisors

Maxime Amblard, MCF HDR, is a specialist in computational linguistics. He is interested in the use of logical and formal tools to model the semantics and pragmatics of natural language. His recent research focuses on a corpus of pathological uses of the language and dialogue's representation. http://members.loria.fr/mamblard/

Chloé Braud, CR CNRS, is a computer linguist specializing in discourse parsing. The strategies she implements are based on machine learning for domain adaptation tasks in a multilingual, cross-domain context. https://chloebt.github.io/

Michel Musiol, PR is a psycholinguist. He is interested in the analysis of the manifestation of thought disorders through language. He studied the particularities of interviews with shizophrenic patients. http://michel.musiol.free.fr/

The skills of the different supervisors are complementary for the subject, both on the formal, digital and psycholinguistic aspects.

Bibliography

[1] Afantenos, Stergos and Kow, Eric and Asher, Nicholas and Perret, Jérémy (2015). Discourse parsing for multi-party chat dialogues, Proceedings of ACL.

[2] Asher, N and Lascarides, A and Lemon, O and Guhe, M and Rieser, V and Muller, P and Afantenos, S and Benamara, F and Vieu, L and Denis, P (2012). Modelling strategic conversation : The STAC project, Proceedings of the Workshop on the Semantics and Pragmatics of Dialogue.

[3] Church, A. (1940). A formulation of the simple theory of types. Journal of Symbolic Logic, 5 :56–68.

[4] Cooper, R., & Ginzburg, J. (2015). TTR for Natural Language Semantics 2.

[5] de Groote, P. (2006). Towards a Montagovian account of dynamics. In M. Gibson and J. Howell, editors, Proceedings of Semantics and Linguistic Theory XVI. Cornell University, Ithaca, NY.

[6] Ginzburg, J. (2016). Semantics of dialogue. In M. Aloni & P. Dekker (Eds.), The Cambridge Handbook of Formal Semantics. Cambridge University Press.

[7] Harris, Z. S. (1954). Distributional structure. Word, 10(2-3), 146-162.

[8] Kamp, H. and U. Reyle (1993). From Discourse to Logic. Kluwer Academic Publishers, Dordrecht.

[9] van Eijck, J. and H. Kamp (1997). Representing Discourse in Context. In J. van Benthem and A. ter Meulen (eds.), Handbook of Logic and Language. Elsevier.

[10] Rebuschi, M., Amblard, M. & Musiol, M. (2014). Using SDRT to analyze pathological conversations. Logicality, rationality and pragmatic deviances. In M. Rebuschi et al. (Eds). Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics. (Dialogue, Rationality, and Formalism). Springer Cham Heidelberg New York, Ch 15, 345-371.

[11] Shi, Zhouxing and Huang, Minlie (2018). A Deep Sequential Model for Discourse Parsing on Multi-Party Dialogues, arXiv preprint arXiv :1812.00176.

[12] Xue, Nianwen and Ng, Hwee Tou and Pradhan, Sameer and Rutherford, Attapol and Webber, Bonnie and Wang, Chuan and Wang, Hongmin (2016). Conll 2016 shared task on multilingual shallow discourse parsing, Proceedings of the CoNLL-16 shared task.