Call for thesis candidates on the ANR project SUMM-RE
Distant Supervision for Incremental Discourse Parsing

Context
In the context of the ANR funded project SUMM-RE, piloted by the research and development team at LINAGORA and with the participation of the academic labs LPL (Laboratoire de Parole et Langage), LIX (Laboratoire d'informatique de l'École polytechnique) and IRIT (Institut de Recherche en Informatique de Toulouse), we are seeking a Ph.D. candidate to pursue a doctorate on the topic of incremental models for discourse parsing. The thesis will be conducted at IRIT within the Melodi team ([https://www.irit.fr/)](https://www.irit.fr/), which specializes in natural language processing and knowledge representation, in conjunction with the R&D team at LINAGORA Labs (https://research.linagora.com/).

Subject
The work in this thesis will improve computational models for learning discourse structures using low resource learning methods, which if successful can recover deep semantic and pragmatic information from text without the need to build large, expertly annotated corpora that require time-consuming and costly labor. While the research team associated with this dissertation has experience with these methods on text and has had success with them in building discourse structures for chat (Badene et al., 2019), we intend in the thesis work to apply them to spoken conversation, which poses both new challenges and possibly computational advantages.

As this thesis is part of a larger industrial and academic partnership, we intend the application of sophisticated NLP tools to conversation to have downstream commercial applications. In particular, as a part of the SUMM-RE project, whose objective is to use discourse structure to improve algorithms for automatic summarization, we will look at how the information we extract about discourse structure can help with the summarization of multi-party conversations, such as those that take place during business meetings. On a more general level, the models resulting from this research project will have an array of implications for developing more capable conversational, automated office assistants.

Profile
- The applicant must hold a Master (or equivalent) in Computer Science, Natural Language Processing or Applied Mathematics.
- Knowledge or experience in one of the following fields: Machine Learning, Natural Language Processing.

Details
Location: team Melodi, laboratoire IRIT, Toulouse and LINAGORA
Supervisors: Nicholas Asher (IRIT), Julie Hunter (LINAGORA)
PhD contract: 3 years
Starting: October 2021
Salary: 1875E / month (gross)

Applications should be sent to asher@irit.fr and jhunter@linagora.com, and should include: a CV detailing your education and experience, especially in research; a cover letter specifying your research interests; Master's grades; and if possible, references we could contact for recommendation.

We will examine applications as they arrive, and until the position is filled.