Fiche de projet tutoré / Project form

Thé ou café ? Oui. / Coffee or tea? Yes.

Encadrement / Supervisors
Sémagramme, Loria (Inria Nancy Grand-Est)
1. Maxime Amblard, Sémagramme — Loria maxime.amblard@univ-lorraine.fr
2. Maria Boritchev, Sémagramme — Loria maria.boritchev@univ-lorraine.fr

Description / Description
1. projet global/global project

Studying dialogical interactions is a major issue in natural language processing, since dialogues compose the basis of human communication. Addressing this problem is a complex task linking approaches from fields such as semantics, pragmatics, and more generally logic and cognition.

This project is part of a more general study of dialogical interactions conducted by the SÉMAGRAMME team (Loria, Inria Nancy Grand Est). It will focus on a systematic study of oral transcriptions of dialogues through the scope of an annotation schema. Work corpora will be chosen depending on the native languages of the participants of the project. The main task is to update, extend and automate the usage of an annotation scheme of dialogical interactions, starting from the question-answer relationship, in a crosslingual perspective.

2. biblio. UE 705 (semestre 7)

The aim of the bibliographic part of this project is for the participants first to get familiar with the general context of the study: [1], [2], [3] (the latter two are publications based on previous years editions of this project). This project is a follow-up on [4] and [5].

3. réalisation. UE 805 (semestre 8)

In order to extend the gold corpus in English (subset of the SCoSE corpus), we will annotate manually several (exact amount to be determined) dialogues. Then, to be able to automatically produce the annotation, it will be interesting to investigate lexical features such as punctuation marks for automatic spotting of questions, as well as semantic features such as word usage in order to determine the span of an answer (as questions can be follow-ups of previous ones). The most important part of this study will be the extension and comparison between the several considered languages.

**Informations diverses : matériel nécessaire, contexte de réalisation /
Various information: material, context of realization**

Most of the corpora are freely available online. Excerpts from the SLAM corpus will be handed to the students. The DinG corpus (French, non-native English) is under construction and will be freely available online/handed to the students, who will have the opportunity to help its creation.

The steps below can be followed:
• getting familiar with corpus annotation
• annotating relevant fragments of the corpora
• comparing the results obtained after annotating corpora in different languages
• (proposing and implementing a questions-and-answer pairs searching algorithm
• implementing a machine learning solution)
• running of experiments, comparing results

Livrables et échéancier / Deliverable and schedule

- October - January: bibliography, taking part in DinG's constitution, getting familiar with the existing annotation schema
- February: manual annotation of the corpora, updating of the annotation schema
- March: testing
- April - May: analysis of the results, report writing

Bibliographie /References (max. 4-5)
[*il ne s'agit pas de la bibliographie complète qui sera fournie aux étudiants au début du projet mais d'une bibliographie indicative pour aider à cerner le sujet*]

1. Maria Boritchev and Maxime Amblard. 2018. Coffee or tea? Yes. SEMDIAL 2018 (AixDial) - The 22nd workshop on the Semantics and Pragmatics of Dialogue.
2. Maria Andrea Cruz Blandon, Gosse Minnema, Aria Nourbakhsh, Maria Bortichev, and Maxime Amblard. 2019. Toward dialogue modeling: A semantic annotation scheme for questions and answers. In *The 13th Linguistic Annotation Workshop (The LAW XIII)*.
3. Maxime Amblard, Maria Boritchev, Marta Carletti, Léa Dieudonat, Yiting Tsai. 2019. A Taxonomy of Real-Life Questions and Answers in Dialogue. SEMDIAL 2019 (LondonLogue) - The 23nd workshop on the Semantics and Pragmatics of Dialogue.
4. María Andrea Cruz Blandón, Gosse Minnema, Aria Nourbakhsh. What's The Answer: Dialogue Annotation, May 2018, Supervised Project Report.
5. Marta Carletti, Léa Dieudonat, Yiting Tsai. Where's The Answer: Dialogue Annotation, May 2019, Supervised Project Report.