

Fiche de projet tutoré / Project form

What are you saying? Dialogue act annotation

Encadrement / Supervisors

Sémagramme, Loria (Inria Nancy Grand-Est)

1. Maria Boritchev, Sémagramme – Loria maria.boritchev@univ-lorraine.fr
2. Maxime Amblard, Sémagramme – Loria maxime.amblard@univ-lorraine.fr
3. Chuyuan Li, Sémagramme - Loria chuyuan.li@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 focus on a systematic study of oral transcriptions of dialogues through the scope of several annotation schema. The main task is to update, extend and automate the usage of two annotation schema of dialogical interactions, one resulting from previous years supervised projects [3], [4] and a standard one [5].

2. biblio. UE 705 (semestre 7)

The aim of the bibliographic part of this project is for the participants first to research the state of the art of dialogue parsers. The second part of the bibliography has to investigate dialogue annotations norms (such as DIT++, <https://dit.uvt.nl/>, ISO 24617-2, [5]) and usage, along with an exploration of the existing annotated dialogue corpora.

This project's annotation schema follow up on [1] and [2].

3. réalisation. UE 805 (semestre 8)

The main goal of this project is the development of a dialogue parser. Based on a transcription of an oral dialogue, this parser would produce dialogue acts annotations. It is to be developed and tested on corpora such as DinG (oral, French), STAC (chat data, English), and several English speaking oral corpora (see <https://linguistics.stanford.edu/resources/corpora/corpus-inventory>).

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

Most of the corpora are freely available online. The DinG corpus (French, oral) 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
- Annotation of relevant fragments of the corpora
- Comparison of the results of annotations with different annotation schema
- Comparison of different state of the art dialogue parsers
- Automation of dialogue annotation

Livrables et échéancier / Deliverable and schedule

- October - January: bibliography, getting familiar with the existing annotation schema, state of the art for dialogue parsing
- February: manual annotation of the corpora, updating of the annotation schema, dialogue parsing tests
- March: development of the dialogue parser, 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 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)*.
2. 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 23rd workshop on the Semantics and Pragmatics of Dialogue.
3. María Andrea Cruz Blandón, Gosse Minnema, Aria Nourbakhsh. What's The Answer: Dialogue Annotation, May 2018, Supervised Project Report.
4. Marta Carletti, Léa Dieudonat, Yiting Tsai. Where's The Answer: Dialogue Annotation, May 2019, Supervised Project Report.
5. Harry Bunt. 2009. The DIT++ taxonomy for functional dialogue markup. In *AAMAS 2009 Workshop, Towards a Standard Markup Language for Embodied Dialogue Acts*.