

# Projet ANR PRICE MINERAL



## Multi dimensional representation for conversation modeling



*Benoît Favre*  
Frédéric Béchet  
Abdellah Fourtassi

*Camille Guinaudeau*  
Sophie Rosset  
Cyril Grouin  
Sahar Ghannay

*Géraldine Damnati*  
Gwénolé Lecorvé  
Anastasia Shimorina  
ANR-24-CE23-4826

# Conversational speech modeling

Human interactions that mainly occur through spoken conversations remain a challenge for spoken language processing models

Conversational speech raises critical issues



# Conversational speech modeling

Human interactions that mainly occur through spoken conversations remain a challenge for spoken language processing models

Conversational speech raises critical issues

Non-verbal information  
can be essential to  
understand a message



# Conversational speech modeling

Human interactions that mainly occur through spoken conversations remain a challenge for spoken language processing models

Conversational speech raises critical issues

Non-verbal information can be essential to understand a message



Semantic grounding between participants is essential for communicating with each other

# Conversational speech modeling

Human interactions that mainly occur through spoken conversations remain a challenge for spoken language processing models

Conversational speech raises critical issues

Non-verbal information can be essential to understand a message



Semantic grounding between participants is essential for communicating with each other

Visual grounding between participants is often needed during a conversation

# Objectives

- Train a multimodal conversation representation model for **communicative acts**
- Study **communicative structures**
- Exploit a panel of ecological conversational datasets from various domains to evaluate the quality of the **generated scripts** and their impact on relevant **use cases**



*[Amy and Sheldon are speaking to Amy's mother over a video computer connection on a laptop]*

**Mrs. Fowler :** It's nice to meet you too, Sheldon. I honestly didn't believe Amy when she told me she had a boyfriend. *(to Sheldon Cooper)*

**Sheldon Cooper :** I assure you I am quite real. And I'm having regular intercourse with your daughter. *(to Mrs. Fowler)*

**Mrs. Fowler :** *[in a surprised tone]* What ?

**Sheldon Cooper :** Oh, yes. We're like wild animals in heat. It's a wonder neither of us has been hurt. *(to Mrs. Fowler)*

**Mrs. Fowler :** *[scared]* Amy, what is he saying ? *(to Amy Farrah Fowler)*

**Amy Farrah Fowler :** You wanted me to have a boyfriend, mother. Well, here he is. *(to Mrs. Fowler)*

*[Sheldon waves at the computer screen, while Mrs. Fowler nervously waves back]*

# Research questions

- How to group low-level descriptors into communication acts with the same underlying intention, particularly considering the various modalities and temporalities of the descriptors ?
- What types of relationships between communication acts need to be captured to reflect the structure of a conversation ?
- Which representation of conversation structure is most conducive to tasks involving automatic comprehension of a conversation ?

General questions about  
conversation analysis and  
structuring

# Research questions

- What are the dimensions of interest for representing communication acts for script generation ?
- Does the structure in communication acts contribute to the explainability of the generation process ?
- Can fiction training corpora contribute to producing reproducible results in real-world conditions ?
- To what extent are scripts a useful representation format for solving different practical cases ?

Specificities for  
script generation

Adequacy with  
the use cases

# Impacts

## Scientific

- Enhancing models to better understand and interpret the complexity of human conversation represent a notable leap forward in the field beyond the capabilities of current multimodal LLMs
- Script generation can:
  - contribute to linguistics, psychology, and cognitive science by providing empirical data on how information is structured and conveyed
  - enhances the model's explainability

## Societal

- Means of accessing spoken content for individuals with hearing
- Mimicking and understanding human conversational patterns can enhance the usability and accessibility of technology