

# Emile Chapuis

I am a postdoctoral researcher working with Caio Corro and Nadi Tomeh in the Spoken Language Processing group at LISN (ex LIMSI).

## Skills and Tools

- **Research experience:** Deep learning, Natural Language Processing, Image Processing, Statistics.
- **Programming languages:** Python, JAVA, R, Matlab, Javascript, C++, C.
- **Software / Libraries / Tools / Technologie:**
  - Scientific programming: NumPy.
  - Data science: scikit-learn, pandas.
  - Deep learning: **PyTorch**, Tensorflow, Keras.
  - Miscellaneous: Unix, Git, SQL, NoSQL (Cassandra).

## Experience

- 2022–present **Postdoctoral Researcher**, *LISN/CNRS*, Saclay.  
Subjects: Cross-lingual syntactic parsing and weakly supervised learning.  
Advisors: *Caio Corro* and *Nadi Tohem*
- 2018–2021 **Phd Student**, *IP Paris/Télécom Paris*, Saclay.  
Subjects: Representation learning for spoken dialog.  
Advisors: *Chloé Clavel* and *Mathieu Labeau*.
- 2018 **Data Scientist**, *L'Oreal*, Paris.  
Computer Vision with deep learning applied to Marteking.
- 2017 **Machine Learning Research Intern**, *Télécom Paris/Idemia*, Paris.  
Subjects: Structured prediction on images using convolutional neural network and operator random fourier features.  
Advisors: *Florence D'Alché Buc*.
- 2016 **Statistics Research Intern**, *MAIAGE-INRA*, Jouy-en-Josas.  
Subjects: Statistical analysis of metaproteomics data within the influence of the intestinal flora state on obesity.  
Advisors: *Sandra Plancade*, *Sylvie Huet* and Catherine Juste.
- 2013-2015 **Full stack developer**, *Nudge-apm*, Paris.  
Developping new features (in JAVA) for the Nudge APM plateform (now part of *Atakama technologies*).
- 2013 **Computer scientist intern**, *Nudge-apm*, Paris.  
Designing the Nudge APM plateform backend with NoSQL database (Cassandra).

## Education

- 2018–2021 **PhD, Computer Science**, *IP Paris/Télécom Paris*, Saclay.

Phd entitled: Neural methods for spoken dialogue understanding.

- 2016-2017 **Msc, Data Science**, *Polytechnique*, Saclay.
- 2015-2016 **Msc, Applied Mathematics**, *Paris-Sud University*, Orsay.
- 2010-2013 **Engineer in Computer Science**, *ENSIMAG*, Grenoble.
- 2008-2010 **Preparatory School**, *Lycée Chateaubriand*, Rennes.

## Publications

### Conference

- 2021 **Chapuis, Emile\***, Pierre Colombo\*, Matthieu Labeau, and Chloé Clavel. *Code-switched inspired losses for spoken dialog representations*. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021.
- 2021 Pierre Colombo, **Chapuis Emile**, Matthieu Labeau, and Chloé Clavel. *Improving Multimodal fusion via Mutual Dependency Maximisation*. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021.
- 2020 **Chapuis Emile\***, Colombo Pierre\*, Matteo Manica, Matthieu Labeau, and Chloé Clavel. *Hierarchical Pre-training for Sequence Labelling in Spoken Dialog*. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020.
- 2020 Pierre Colombo\*, **Emile Chapuis\***, Matteo Manica, Emmanuel Vignon, Giovanna Varni, and Chloe Clavel. *Guiding Attention in Sequence-to-Sequence Models for Dialogue Act Prediction*. Association for the Advancement of Artificial Intelligence (AAAI), 2020.

### Workshop

- 2020 **Chapuis Emile\***, Pierre Colombo\*, Matteo Manica, Giovanna Varni, Emmanuel Vignon, and Chloé Clavel. Guider l'attention dans les modèles de séquence à séquence pour la prédiction des actes de dialogue. In *Workshop sur les Affects, Compagnons artificiels et Interactions*, 2020.

### Preprint

- 2021 **Emile Chapuis** and 30 other co authors. NI-augmenter: A framework for task-sensitive natural language augmentation. *CoRR*, volume abs/2112.02721, 2021.

### Work in progress

Caio Corro & **Emile Chapuis**. Weakly supervised fenchel-young losses: a framework for constructing loss functions for learning from partial labels.

## Research Activities

### Scientific Advisor

- 2020 Unsupervised text clustering with interpretability, IBM
- 2019 Multimodal Emotion Recognition, Pôle Emploi

### Teaching Assisant (140 hours)

- Statistical Learning and Data Mining
- Applied Statistical Learning
- Statistics: Linear Models
- Structured Data : Learning, prediction, dependency and testing
- Machine Learning for Text Mining
- Natural Language Processing
- Basic Machine Learning
- Basic Analysis/ Basic Algebra/ Analysis/ Probabilit

## Opensource Contributor

- Contribution to *Datasets*: *SILICONE* (10 datasets) and *MIAM* (5 datasets).
- *NL-Augmenter*: A Framework for Task-Sensitive Natural Language Augmentation.

## Services

- Reviewer for ACL, EMNLP, NAACL, AAAI.
- Board member of the *ATALA*, a French association that promotes and animates research in natural language processing.

## Summer & winter school

2021 Advanced Language Processing School, *ALPS*.

---

## Miscellaneous

**Languages** French (native speaker), English (business fluent), Spanish (notions), Italian (notions), Breton (notions).

**Sports** Rugby, Surf, Swimming.

**Music** Guitar, Tin Whistle.

**Hobbies** Movies, Literature, Cooking.