

# Gaël Guibon

PhD. in Computer Science and Natural Language Processing

---

54 rue coriolis 75012 Paris  
☎ +33 6.43.59.29.31      ✉ gael.guibon@gmail.com  
<https://gguibon.github.io/>

## PROFILE

Natural Language Processing, Machine Learning, Information Extraction, Computational Linguist, Emotion Recognition.

## EXPERIENCE

- Postdoc Researcher in Computer Science** July 2020 - July 2021  
Deep learning for emotion recognition in chat messages.  
Team S2A in LTCI laboratory of Télécom Paris. Team Plateau IA in Research & Innovation dpt. of SNCF.
- Postdoc Researcher in Computer Science** June 2019 – June 2020  
Lexical Evolution of French.  
Inria, team-project ALMAnaCH, Paris
- PhD in Computer Science** March 2016 - May 2019.  
Title : Automatic and Adaptive Emoji Recommendation.  
CIFRE (mix academic and company) between the laboratory LIS-CNRS UMR 7020 (previously named LSIS-CNRS UMR 7296) and Caléa Solutions company, Marseille
- Short-term Research Engineer at ENS** February 2016  
CNRS-LaTTiCe, École Nationale Supérieure (ENS), Montrouge  
— API and web application for Old French dependency parsing
- Engineer at CNRS** October 2014 - October 2015  
CNRS-INIST, Institut Nationale de l'Information Scientifique et Technique, Nancy  
— Data pre-treatments using machine learning for terminology extraction  
— Collaborating on TermSuite adaptation to data  
— Workshop animation to UniteX introduction
- Research Laboratory Internship** Summer 2014  
CNRS-LIGM, Institut d'électronique et d'informatique Gaspard-Monge, Marne-la-Vallée  
— Creation of a dependency parser for Old French.  
— Dependency grammars usage.
- Research Laboratory Internship** Spring 2014  
CNRS-LaTTiCe, Université de Paris III Sorbonne-Nouvelle, Montrouge  
— Creation of a Part-of-Speech tagger for Old French.  
— Using Conditional Random Fields.

## PUBLICATIONS

### Journal

Guibon, Labeau, Lefeuvre, Clavel 22 : Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks. Elsevier, Software Impacts 2022, volume 12.

### International Conferences

Guibon, Labeau, Lefeuvre, Clavel 22 : EZCAT : an Easy Conversation Annotation Tool. In 13th Language Resources and Evaluation Conference (LREC 2022 - **rang C**).

Guibon, Labeau, Flamein, Lefeuvre, Clavel 21 : Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks. In the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021 - **rang A**).

Guibon, Labeau, Flamein, Lefeuvre, Clavel 21 : Meta-learning for Classifying Previously Unseen Data Source into Previously Unseen Emotional Categories. In the 1st Workshop on Meta Learning and Its Applications to Natural Language Processing (MetaNLP @ ACL 2021).

Guibon, Sagot 20 : OFrLex : A Computational Morphological and Syntactical Lexicon for Old French. In 12th Language Resources and Evaluation Conference (LREC 2020 - **rang C**).

Guibon, Courtin, Gerdes, Guillaume 20 : When Collaborative Treebank Curation Meets Graph Grammars. In 12th Language Resources and Evaluation Conference (LREC 2020 - **rang C**).

Guibon, Ermakova, Seffih, Firsov, Le Noé-Bienvenu 19 : Multilingual Fake News Detection with Satire on Vaccination Topic. In 20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLING 2019 - **rang B**).

*Espinasse, Fournier, Chifu, Guibon, Azcurra and Macé 19* : On the Use of Dependencies in Relation Classification of Text with Deep Learning. In 19th International Conference on Computational Linguistics and Intelligent Text Processing (CICLING 2019 - **rang B**).

*Guibon, Ochs, Bellot 18* : Emoji recommendation in private instant messages. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (pp. 1821-1823) (SAC 2018 - **rang B**). **Best Poster Award for SONAMA track**  
*Guibon, Ochs, Bellot 18* : LIS at SemEval-2018 Task 2 : Mixing Word Embeddings and Bag of Features for Multilingual Emoji Prediction. In SemEVAL 2018.

*Guibon, Ochs, Bellot 18* : From Emoji Usage to Categorical Emoji Prediction. In 19th International Conference on Computational Linguistics and Intelligent Text Processing (CICLING 2018 - **rang B**), Hanoi. **Best Verifiability, Reproducibility, and Working Description Award**

*Guibon, Tellier, Constant, Prevost, Gerdes 15* : Searching for Discriminative Metadata on Heterogenous Corpora, 14th Treebank and Language Theory (TLT), Warsaw

*Guibon, Tellier, Constant, Prévost, Gerdes 14* : Parsing Poorly Standardized Language Dependency on Old French, 13th Treebank and Language Theory (TLT), Tübingen.

## National Conferences

*Guibon, Ochs, Bellot 18* : De l'impact des emojis sur les sentiments à une recommandation pertinente. Workshop Affect Compagnon Artificiel Interaction (WACAI 2018), Ile de porquerolles .

*Guibon, Ochs, Bellot 17* : Prédiction automatique d'emojis sentimentaux, 14<sup>ème</sup> Conférence en Recherche d'Information et Application (CORIA), Marseille.

*Guibon, Ochs, Bellot 16* : From Emojis to Sentiment Analysis, Workshop Affect Compagnon Artificiel Interaction (WACAI 2016), Brest.

*Guibon, Tellier, Prévost, Constant, Gerdes 15* : Analyse syntaxique de l'ancien français : quelles propriétés de la langue influent le plus sur la qualité de l'apprentissage ?, 22<sup>ème</sup> Traitement Automatique des Langues Naturelles (TALN), Caen.

## Other Publications

*Guibon, Ochs, Bellot 17* : Une plateforme de recommandation automatique d'emojis. Traitement Automatique du Langage Naturel (TALN), Orléans. (Démo)

*Guibon 16* : SOFA : Une plateforme d'analyse syntaxique en ligne pour l'ancien français, 23<sup>ème</sup> Traitement Automatique du Langage Naturel (TALN), Paris. (Démo)

## Seminars / invited talks

**Inria's Developer meetup** (February 22th 2022, and October 29th 2019) : *VueJS and Interactive slides* <https://sed-paris.gitlabpages.inria.fr/developer-meetups/>

**Talk at Paris-Nanterre University** (October 26th 2021) : *De l'enrichissement de données diachroniques à l'exploitation de données privées*

**Talk at LTCI laboratory from Télécom Paris** (September 30th 2021) on few-shot learning in conversational context

**Talk at PhD students' seminar from LIS-CNRS laboratory** (march 2018) : Apprentissage non-supervisé pour l'orientation de la prédiction multi-classe d'emojis (FaceEmojis)

**Talk at the LaTTiCe-CNRS laboratory** (February 8th 2016) : Exploration de corpus d'ancien français : méta-données les plus discriminantes et influence sur la qualité l'apprentissage

## Scientific Popularization

**Scientific Popularization Article on Meta-Learning.** *Le méta-apprentissage fait ses premiers pas en traitement automatique du langage.* – Data Analytics Post – par Pierre Vandeginste.

<https://dataanalyticspost.com/meta-apprentissage-traitement-automatique-langage/>

## TEACHING

**Natural Language Processing** (IA 717, *lab project creation and supervision.*) <https://www.telecom-paris.fr/fr/masteres-specialises/formation-intelligence-artificielle> **Machine Learning for Natural Language Processing** (SE364, *original content for labs, lab assistant supervisor. Projects and courses in English.*) <https://nlp-ensae.github.io/>

ENSAE 2022 : 9H TD

**Machine Learning for Natural Language Processing** (SE364, *original content for labs and 3 courses. Projects and courses in English.*) <https://nlp-ensae.github.io/>

ENSAE 2021 : 4H30 CM et 9H TD

**Introduction to Programming and Python** (*Original Content. Courses in English*) <https://github.com/gguibon/essec-python-1>

ESSEC Business School 2020 : Global BBA 2nd year : 25H CM et TD. Students' Evaluation : 4/5 (+0.68 compared to 2019)

**Machine Learning for Natural Language Processing** (SE364, *original content for the 3rd lab and projects, courses in English*) <https://nlp-ensae.github.io/>

ENSAE 2020 : 4H30 TD

**Introduction to Python** (*original content*) <https://github.com/gguibon/essec-python-1>

ESSEC : Global BBA 2nd year : 25h

**JavaEE** (*original content*) <https://github.com/gguibon/CoursesJavaEE>

Aix-Marseille University 2018-2019 : Master 2 Computer Science : 59h

**Introduction to Deep Learning**

Aix-Marseille University 2017-2018 : LSIS-CNRS : Master 2 Research DiMaG : 8h

**JavaEE** (*Contenu original*) <https://github.com/gguibon/CoursesJavaEE>

Aix-Marseille University 2017-2018 : Master 2 Computer Science : 64h

**Introduction to Python** (*original content*) <https://github.com/gguibon/CoursesPython>

Aix-Marseille University 2017-2018 : Master 2 Computer Science : 8h

**Databases in Java** (*original content*) <https://github.com/gguibon/teaching-isi-projet>

Aix-Marseille University 2016-2017 : Master 1 Computer Science : 20h

## ADMINISTRATION ET RESPONSIBILITIES

### Roles

**ANR Project REVITALISE (2022–2025)** resp. Magalie Ochs (LIS-CNRS). With Chloé Clavel (Télécom Paris) and Béatrice Biancardi (Télécom Paris).

**ANR Project Autogramm (2022–2025)** resp. Sylvain Kahane <https://autogramm.github.io/>

**Supervision of an Annotator.** Aurore Lessieux (1 month).

**Board Member of ATALA (French NLP National Association).** 2020 - now

**Supervision of an intern (master 2 reseach in Computer Science).** René Azcurra.

Supervised with Bernard Espinasse, Sébastien Fournier and Adrian Chifu (defended on june 18th 2018 in Marseille)

**Vice-president and secretary of PhD and PhD Students from the LIS laboratory (A2DL)** . (2016–2018)

### Conference Organization

**SyntaxFest** (4 events : Quasy, Depling, TLT et UDWorkshop), Paris *26–30 August 2019* :

Member of the organization committee.

**Workshop Affect Compagnon Artificiel Interaction (WACAI 2018)** , Porquerolles *13–15 June 2018* :

Membre du comité d'organisation et responsable du site web.

**14th PhD Student Days from LSIS laboratory (JDL6)** , Carry-le-Rouet *22–23 May 2017* :

Member of the organization committee.

**Conference on Information Retrieval and Applications (CORIA 2017)** , Marseille *29–31 March 2017* :

Member of the organization committee and webmaster.

### Hackathons and Reading Group Organization

**Working Group ACAI** Saint-Etienne

22 june 2022 : member of the program committee.

**Reading Group ACAI** Virtual *2022, every month* : organizer

**HackaTAL 2018** Rennes *14–15 may 2018* : member of the organization committee.

**HackaTAL 2017** Orléans *24–26 june 2017* : member of the organization committee and in charge of the automatic summarization task (ideas proposition, setup, data collection through *web scrapping* et automatic teams evaluation)

### Reviews

**Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2022)** 27th june - 1st july, Avignon.

**Applied Intelligence** march 2022 issue

**Empirical Methods in Natural Language Processing (EMNLP 2021)** 7-11 november, Punta Cana and online.

**Numero 62-2 of French NLP journal "Traitement Automatique des Langues" (revue TAL)** On now NLP applications. 2021

**IEEE Transactions on Affective Computing (TAFFC)** 2021 Issue.

**European Chapter of the Association for Computational Linguistics (EACL 2021)** 19-23 april 2021, online.

**Workshop Affect Compagnon Artificiel Interaction (WACAI 2021)** 13-15 october 2021, Île d'Oléron.

**Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2021)** 28th june - 2nd july, Lille.

**CONFérence en Recherche d'Information et Applications (CORIA 2021)** Grenoble and/or online, 14-16 april 2021.

**Empirical Methods in Natural Language Processing (EMNLP 2020)** 16-20 november, online.

**Rencontres Jeunes Chercheur(euses) (RJC) CORIA/TALN 2018.**

**Third International KEYSTONE Conference (IKC2017)** Gdańsk Poland, 11-12 September 2017.

## RECENT SUPERVISORS

**Chloé Clavel** : Current postdoc. Professor. [chloe.clavel@telecom-paris.fr](mailto:chloe.clavel@telecom-paris.fr)

**Matthieu Labeau** : Current postdoc. Assistant Professor. [matthieu.labeau@telecom-paris.fr](mailto:matthieu.labeau@telecom-paris.fr)

**Benoît Sagot** : Scientific supervisor for the post-doctoral position in the PROFITEROLE project. Researcher. Contact : [benoit.sagot@inria.fr](mailto:benoit.sagot@inria.fr)

**Patrice Bellot** : Thesis supervisor. Professor. Contact : [patrice.bellot@lis-lab.fr](mailto:patrice.bellot@lis-lab.fr)

**Magalie Ochs** : Thesis co-supervisor. Assistant Professor (HDR). Contact : [magalie.ochs@lis-lab.fr](mailto:magalie.ochs@lis-lab.fr)

## FORMATION

**PhD in Computer Science**, Natural Language Processing, Information Retrieval.  
Aix-Marseille University. Laboratory of Computer Science and Systems (LIS-CNRS UMR 7020) and Caléa Solutions.  
2016-2019

**Master of Engineering in Computational Linguistics**, Linguistics and Computer Science, NATural Language Processing. Université Paris III Sorbonne-Nouvelle, Obtained in Septembre 2014

**Master 1 of Engineering of Language for Information Management (Ingénierie de la Langue pour la Gestion Intelligente de l'Information)**, Linguistics and Computer Science, Natural Language Processing. Université Paris 4 Sorbonne, 2011

**Bachelor in Language and Computer Science (Computational Linguistics)**, Natural Language Processing, *Diploma C2i*, Databases, etc. Université Paris 4 Sorbonne, 2010

**High-School Diploma on Litterature**, Specialized in cinema and audiovisual. High-School Paul Valéry, 2007  
Specialization : Text Mining, Natural Language Processing

## SKILLS

### Linguistics

Linguistics and Textual Linguistics, Stylistics, Rethorics, Coprus Linguistics, Computational Linguistics, Socio-linguistics, Formal Grammar, Discourse and Lexicometrics, Language Diversity and Typology, Semantics

### Computer Science

*Languages & Softwares* : Python, Java, JavaEE, Javascript, JavaFX, LateX, UniteX, Shell, DTD, XQuery, XPath, SQL, XSL, XML, HTML

*Operating Systems* : Unix, Windows.

*Softwares/Frameworks* : PyTorch, Keras, SciKitLearn, VueJS, SpringBoot, Tensorflow, AngularJS, Wapiti, Weka, Mate, MaltParser, UIMA, Unitex, Protege, Praat

*IDEs* : Visual Studio Code (Python, JS, VueJs), Eclipse (Java, JavaEE), Android Studio / IntelliJ

## SOFTWARES

**Git Repository** : <https://github.com/gguibon>

**EZCAT** (2022) Customizable interface for an easy annotation of conversation. (Evolution and generalization of the POC Annotator app) : <https://github.com/gguibon/ezcat>

**ProtoSeq** (2021) Source Code for *Prototypical Networks* (few-shot learning) adapted to sequence labeling for emotion recognition in conversation (EMNLP 2021) : <https://github.com/gguibon/ProtoSeq>

**MetaLearningDemo and source code** (2021) Demo dedicated to the present the work published at the MetaNLP workshop in ACL 2021, to be hosted soon in HiParis (<https://www.hi-paris.fr/>) and presented as a poster session. <https://github.com/gguibon/metalearning-emotion-datasource>.

**POC Annotator** (2020) Interface for Messages' Emotion Annotation in Conversation : <https://gguibon.github.io/poc/>

**LEMA : Lexicon and Enrichments Modification Application** : <https://lema.inria.fr:8888/>

**Arborator-Grew** (2020) Online visualization, modification and creation of treebanks (with Kim Gerdes, Bruno Guillaume, Marine Courtin, Kirian Guiller, Lila Kim and Sylvain Kahane.) : <https://arborator.github.io/>

**EmoReco** : EMOtional EMOji RECOmmendation system : <http://lsis-mood-emoji.lsis.org:8080/>

**FaceEmojis** : Emoji Embeddings and Clusters of Face Emojis : <https://github.com/gguibon/FaceEmojis>

**AirBnB Scrapper** used to create the corpus for participants of the NLP hackathon HackaTAL 2017 : <https://github.com/gguibon/airbnb-scrapper>

**Human Resource Management Tool** used at the LSIS laboratory for more than one year : <https://github.com/gguibon/LSIS-AppGestion>

**SOFA** : Online Demo for Old French dependency parsing : <https://github.com/gguibon/sofa>

**Interactive Tool for visualizing, browsing and comparing treebanks** : <https://github.com/gguibon/LAKME-SRCMF>

**API for dependency parsing and PoS tagging** for Old French : <https://github.com/gguibon/SRCMF-nlp-api>

**Involvements** : Update and optimization (computation time) of the dependency parser Echo : <https://github.com/gguibon/echo>

## LANGUAGES

*French* : Native Speaker *English* : Spoken, read, and written fluently ;

*Spanish* : Average