

☎ 06.43.59.29.31

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

19.11.1987 à Paris

PROFIL

Traitement automatique du langage naturel, apprentissage automatique, extraction d'information, linguiste informaticien.
Analyse d'émotions.

EXPERIENCE

- Post-doc en informatique** Juillet 2020 - juillet 2021
Apprentissage profond pour l'analyse d'émotions dans des données de chat.
Équipe S2A au LTCI de Télécom Paris et équipe Plateau IA au département Innovation & Recherche de la SNCF.
- Post-doc en informatique** juin 2019 – juin 2020
Évolution lexicale du français.
Inria, équipe-projet ALMAnaCH, Paris
- Doctorat en informatique** mars 2016 - mai 2019.
Sujet : Recommandation automatique et adaptative d'emojis.
CIFRE entre LIS-CNRS UMR 7020 (anciennement LSIS-CNRS UMR 7296) et Caléa Solutions, Marseille
- Vacataire à l'ENS** février 2016
CNRS-LaTTiCe, Ecole Nationale Supérieure (ENS), Montrouge
— API et application en ligne d'analyse syntaxique de l'ancien français
- Ingénieur d'étude au CNRS** Octobre 2014 à octobre 2015
CNRS-INIST, Institut Nationale de l'Information Scientifique et Technique, Nancy
— Pré-traitements des données par apprentissage automatique pour l'extraction terminologique
— Contribution à l'adaptation de TermSuite aux données
— Animation d'un atelier interne d'introduction à UniteX
- Stage en laboratoire de Recherche** Été 2014
CNRS-LIGM, Institut d'électronique et d'informatique Gaspard-Monge, Marne-la-Vallée
— Mise en place d'un analyseur syntaxique pour l'ancien français.
— Utilisation des grammaires de dépendance.
- Stage en laboratoire de Recherche** Printemps 2014
CNRS-LaTTiCe, Université de Paris III Sorbonne-Nouvelle, Montrouge
— Création d'un étiqueteur Part-Of-Speech pour l'ancien français.
— Utilisation de l'algorithme statistique des Conditional Random Fields.

PUBLICATIONS

Communications dans des conférences internationales

Guibon, Labeau, Flamein, Lefevre, 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, Lefevre, 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), Hanoï. **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.

Communications dans des conférences nationales

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^{ème} 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^{ème} Traitement Automatique des Langues Naturelles (TALN), Caen.

Autres 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^{ème} Traitement Automatique du Langage Naturel (TALN), Paris. (Démo)

Séminaires

Séminaire au laboratoire MoDyCo-CNRS (prévu pour octobre 2021) : *titre à venir (sur l'apprentissage frugal)*

Séminaire au LTCI (30 septembre 2021) sur l'apprentissage frugal en conversation

Séminaire de doctorants au LIS-CNRS (mars 2018) : Apprentissage non-supervisé pour l'orientation de la prédiction multi-classe d'emojis (FaceEmojis)

Séminaire au LaTTiCe-CNRS (8 février 2016) : Exploration de corpus d'ancien français : méta-données les plus discriminantes et influence sur la qualité l'apprentissage

Vulgarisation

Article de vulgarisation sur le méta-apprentissage. *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/>

ENSEIGNEMENTS

Machine Learning for Natural Language Processing (SE364, *contenu original pour les TDs et 3 CM. Projets et cours en anglais.*) <https://nlp-ensae.github.io/>

Vacation ENSAE 2021 : 4H30 CM et 9H TD

Introduction to Programming and Python (*Contenu original et en anglais*) <https://github.com/gguibon/essec-python-1>

Vacation ESSEC Business School 2020 : Global BBA deuxième année : 25H CM et TD. Évaluation du cours par les étudiants : 4/5 (+0.68 par rapport à 2019)

Machine Learning for Natural Language Processing (SE364, *contenu original pour le TD3 et projets, cours en anglais*) <https://nlp-ensae.github.io/>

Vacation ENSAE 2020 : 4H30 TD

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

ESSEC : Global BBA deuxième année : 25h

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

Vacation Aix-Marseille Université 2018-2019 : Master 2 Informatique : 59h

Introduction à l'apprentissage profond

Vacation Aix-Marseille Université 2017-2018 : LSIS-CNRS : Master 2 Recherche DiMaG : 8h

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

Vacation Aix-Marseille Université 2017-2018 : Master 2 Informatique : 64h

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

Vacation Aix-Marseille Université 2017-2018 : Master 2 Informatique : 8h

ADMINISTRATION ET AUTRES RESPONSABILITÉS

Rôles

- Projet ANR en soumission.** avec Magalie Ochs (LIS-CNRS), Chloé Clavel (Télécom Paris) et Béatrice Biancardi (Télécom Paris).
- Encadrement d'une annotatrice.** Aurore Lessieux (1 mois).
- Membre du conseil d'administration de l'ATALA.** 2020 - présent
- Co-encadrement d'un stagiaire de master 2 recherche en informatique.** René Azcurra.
Encadré avec Bernard Espinasse, Sébastien Fournier et Adrian Chifu (soutenu le 18 juin 2018 à Marseille)
- Vice-président et secrétaire de l'association des docteurs et doctorants du LIS (A2DL)** . (2016–2018)

Organisation de conférences

- SyntaxFest** (4 événements : Quasy, Depling, TLT et UDWorkshop), Paris *26–30 Août 2019* :
Membre du comité d'organisation.
- Workshop Affect Compagnon Artificiel Interaction (WACAI 2018)** , Porquerolles *13–15 Juin 2018* :
Membre du comité d'organisation et responsable du site web.
- 14èmes Journées des doctorants du LSIS (JDL6)** , Carry-le-Rouet *22–23 Mai 2017* :
Membre du comité d'organisation.
- Conférence en Recherche d'Information et Application (CORIA 2017)** , Marseille *29–31 Mars 2017* :
Membre du comité d'organisation et responsable du site web.

Organisation de Hackathons

- HackaTAL 2018** Rennes *14–15 mai 2018* : membre du comité d'organisation.
- HackaTAL 2017** Orléans *24–26 juin 2017* : Comité d'organisation et responsable de la tâche de résumés automatiques (proposition, mise en place, collecte des données par *web scrapping* et évaluations automatique des équipes)

Relectures

- Empirical Methods in Natural Language Processing (EMNLP 2021)** 7-11 novembre, Punta Cana.
- Numéro 62-2 de la revue "Traitement Automatique des Langues" (revue TAL)** Sur les nouvelles applications du TAL. 2021
- IEEE Transactions on Affective Computing (TAFFC)** Journal de 2021.
- European Chapter of the Association for Computational Linguistics (EACL 2021)** 19-23 avril 2021, en ligne.
- Workshop Affect Compagnon Artificiel Interaction (WACAI 2021)** 13-15 octobre 2021, Île d'Oléron.
- Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2021)** 28 juin au 2 juillet, Lille.
- CONFérence en Recherche d'Information et Applications (CORIA 2021)** Grenoble ou en ligne, 14-16 avril 2021.
- Empirical Methods in Natural Language Processing (EMNLP 2020)** 16-20 novembre, en ligne.
- Rencontres Jeunes Chercheur(euses) (RJC)** CORIA/TALN 2018.
- Third International KEYSTONE Conference (IKC2017)** Gdańsk Poland, 11-12 September 2017.

RESPONSABLES RÉCENTS

Benoît Sagot

Responsable scientifique pour le post-doctorat du projet PROFITEROLE. Chargé de recherche. Contact : benoit.sagot@inria.fr

Patrice Bellot

Encadrant de thèse. Professeur des universités. Contact : patrice.bellot@lis-lab.fr

Magalie Ochs

Co-encadrante de thèse. Maîtresse de conférence (HDR). Contact : magalie.ochs@lis-lab.fr

FORMATION

Doctorat en Informatique, Traitement Automatique des Langues (TAL), Recherche d'Information (RI).

Aix-Marseille Université. Laboratoire d'Informatique et Système (LIS-CNRS UMR 7020) and Caléa Solutions. 2016-2019

Master d'Ingénierie Linguistique, Linguistique Informatique, Traitement Automatique des Langues. Université Paris III Sorbonne-Nouvelle, Validation en Septembre 2014

Maîtrise d'Ingénierie de la Langue pour la Gestion Intelligente de l'Information, Linguistique Informatique, Traitement Automatique des Langues. Université Paris 4 Sorbonne, 2011

Licence Langue Française et Techniques Informatiques, Traitement Automatique des Langues, *Diplôme C2i Niveau 1*, Bases de données, etc. Université Paris 4 Sorbonne, 2010

COMPÉTENCES

Linguistique

Linguistique générale et textuelle, Stylistique, Rhétorique, Linguistique de corpus, Linguistique Computationnelle, Sociolinguistique, Grammaires formelles, Analyse du discours et lexicométrie, Diversité des langues et typologie, Sémantique

Informatique

Langages & Logiciels : Java, Python, JavaEE, Javascript, JavaFX, LateX, UniteX, Shell, DTD, XQuery, XPath, SQL, XSL, XML, HTML

Systèmes d'exploitations : Unix, Windows.

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

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

LOGICIELS

Dépôt Git : <https://github.com/gguibon>

ProtoSeq (2021) Code des *Prototypical Networks* (apprentissage frugal) adaptés à l'étiquetage de séquence pour la détection d'émotion en conversation (publié à EMNLP 2021) : <https://github.com/gguibon/ProtoSeq>

MetaLearningDemo et code source (2021) Démonstrateur des travaux publiés au workshop MetaNLP d'ACL 2021, à paraître dans HiParis (<https://www.hi-paris.fr/>) et présenté en session poster. <https://github.com/gguibon/metalearning-emotion-datasource>.

POC Annotator (2020) Annotation en émotion des messages de conversations : <https://gguibon.github.io/poc/>

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

Arborator-Grew (2020) Visualisation, édition et création d'arbres syntaxiques en ligne (avec Kim Gerdes, Bruno Guillaume, Marine Courtin, Kirian Guiller, Lila Kim et 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>

Scrapper de données Airbnb utilisé pour la constitution du corpus de l'HackaTAL 2017 : <https://github.com/gguibon/airbnb-scrapper>

Outil de gestion de personnel utilisé au LSIS pendant plus d'un an : <https://github.com/gguibon/LSIS-AppGestion>

SOFA : Démo en ligne d'analyse syntaxique de l'ancien français : <https://github.com/gguibon/sofa>

Outil interactif de visualisation, de parcours et de comparaison d'un corpus arboré : <https://github.com/gguibon/LAKME-SRCMF-visu>

API d'analyse syntaxique pour l'ancien français : <https://github.com/gguibon/SRCMF-nlp-api>

Participations : Mise à jour et optimisation du temps de calcul pour l'analyseur syntaxique Echo : <https://github.com/gguibon/echo>

LANGUES

Anglais : Parlé, lu et écrit couramment ;

Espagnol : Moyen