

# Jocelyn AZNAR

*A versatile perspective involving Corpus Linguistics, Natural Language Processing and Language Documentation*

Date of birth May 1986 the 14<sup>th</sup>  
Thomasstraße 67  
12053 Berlin  
☎ 49-157-33244-963  
✉ [contact@jocelynaznar.eu](mailto:contact@jocelynaznar.eu)  
🌐 <http://jocelynaznar.eu>



I am currently a post-doctoral fellow at ZAS (Leibniz-Zentrum Allgemeine Sprachwissenschaft<sup>1</sup>). My **methodologies** are inspired by Corpus Linguistics, Ethnolinguistics, Variational linguistics and Language Documentation and Description. The data I work with, **oral texts**, result from **field investigations and recordings**, which are used to develop language resources for the local and the research communities as well as language studies. **Annotations** are created on the basis of these interviews and finally the analyses are carried out qualitatively, helped by annotation tools like ELAN and quantitatively with programming language (Python, Pandas) and data processing frameworks for reproducible analysis (Jupyter).

## University curriculum

- 2020-22 **Postdoctoral researcher**, ZAS-QUEST, Berlin, QUEST (Quality - Established) is an initiative to develop and promote curation criteria and quality standards for audiovisual annotated language data. Within this project, I first contribute to a quality assurance framework, RefCo for language documentation corpora to maximise their potential for secondary use in cross-linguistic studies. Then, I reuse existing language documentation corpora thanks to a cooperation with DoReCo to produce cross-linguistics study on direct speech..
- 2019 **PhD in Language Sciences**, CREDO – EHESS, Marseille, Supervisors : Núria Gala, Professor of Universities – AMU, Laurent Dousset, Director of Study CREDO – EHESS.  
*Narrer un nabol : la production des textes nisvais en fonction de l'âge du locuteur et de la situation d'énonciation, Malekula, Vanuatu.* ; "Narrating a nabol : the production of Nisvai texts according to age and the situation of enunciation, Malekula, Vanuatu"  
Methodologies: corpus linguistics, variational linguistics, computational linguistics, ethnolinguistics, textual linguistics, participant observation, documentation and linguistic description.
- 2010–2012 **Master's Degree in Language Science, majoring in Language Technology**, Aix-Marseille University, Marseille.  
Topics studied: Programming, Algorithmics, Web Development, Oral Processing, Contemporary Grammars, Research Methodology, Statistics, Automatic Language Processing.
- 2009–2010 **Language Sciences**, University of Westminster, London.  
Exchange with the University of Westminster as part of the international exchange programme ERASMUS. Topics studied: History of the English Language, Language and Gender, Linguistics Applied to Literature, Psycholinguistics, Language Acquisition, Generative Grammar.
- 2008–2010 **Language Sciences License**, University of Provence, Aix-en-Provence.  
Topics studied : Phonetics, Phonology, Morphology, Syntax, Brain Function, Linguistic Theories, Semantics, Perl Programming.
- 2004–2008 **License in Translation and Multilingual Communication**, CFTTR, Rennes.  
Topics studied: Technical Translation, Technical Writing, Project Management, Computer-Aided Translation, Proof-reading.

## Publications

- Submitted **Article in journal**, in *Language Documentation & Conservation*.  
*Tree Visualization for text corpora with ArbresEtStats*
- In preparation **Chapter in edited volume**, Language Science Press, Spronck, Stef, Daniela Casartelli (dir).  
*Silent Gaps as a Distinctive Feature of the Nisvai Direct Speech Sequence A Corpus-based Description of reported speech within Nisvai Narratives, Vanuatu*
- In preparation **Chapter in edited volume**, in Lambert Lucas, Luca Greco and Sandra Nossik (dir), Études linguistiques et textuelles.  
*Le régime de vérité nisvaie: L'emploi des noms propres en fonction de la classe d'âge de l'orateur (Malekula, Vanuatu)* ; in "La narration : du discours à la multimodalité"
- Accepted **Revue TAL**.  
Review of "*Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics*", by Emily M. Bender et Alex Lascarides.

- 2020 **LIFT: Linguistique Informatique, Formelle et de Terrain**, *In proceedings*, Montrouge, France..  
"D'un corpus à l'autre : D'une étude reproductible et portable du discours direct nisvai à la comparaison linguistique"  
; *From one corpus to another: From a reproducible and portable study of nisvai direct speech to linguistic comparison*.
- 2020 **LIFT: Linguistique Informatique, Formelle et de Terrain**, *In proceedings*, Montrouge, France., With Frank Seifart.  
*RefCo: An initiative to develop a set of quality criteria for fieldwork corpora*
- 2020 **LREC: International conference on Language Resources and Evaluation**, *in Proceedings*, Marseille, Aznar, J. & Gala, N.  
*The Nisvai Corpus of Oral Narrative Practices from Malekula (Vanuatu) and its Associated Language Resources*.
- 2016 **Report for Huma-num**, CNRS, Marseille, Aznar, Jocelyn.  
"Protocol for the implementation of perennial archiving: Reflections and proposals for the implementation of ODSAS data archiving on the PAC." *Protocole d'implémentation de l'archivage pérenne : Réflexions et propositions pour la mise en place de l'archivage des données d'ODSAS sur la PAC*.

---

## Fieldworks

- 2015 **Participant Observations and Discussions**, CREDO – EHESS, Vanuatu, Malekula, Discussions and  
July to Participant observations on the previously developed resources, annotation of the Nisvai narrative records..  
October
- August 2013 **Studies of the Nisvai language and community**, CREDO – EHESS, Vanuatu, Malekula.  
to February Elements of oral literature and taxonomy. Further description of linguistic structures (phonology, morphology)  
2014
- 2011 **Meeting with the Nisvai community**, LIF-TALEP – CREDO, Vanuatu, Malekula.  
September– First Meeting with the Nisvai community. Elements of description of the phonology and morphology of the Nisvai  
October language.

---

## Language Resources and Software

- 2020 **ArbresEtStats**, Marseille, Aznar, Jocelyn.  
ArbresEtStats: a Python program to assist the phonological and characters' distribution studies of a text or a corpus through the visualization of its forms and statistics in trees. <https://hal.archives-ouvertes.fr/hal-02512985/>
- 2016 **Web-based interface for viewing and listening to Nisvai narrative practices**, [http://jocelynaznar.eu/lexique\\_nisvai/form\\_visualisation.php](http://jocelynaznar.eu/lexique_nisvai/form_visualisation.php).  
For the exhibition *Lokol Stael*, I developed an interface that emphasizes the text organization of Nisvai oral histories and the importance of the voice for the interpretation of texts.
- 2015 **Web interface to access the Nisvai corpus' annotations**, [http://jocelynaznar.eu/lexique\\_nisvai/form\\_lexique.html](http://jocelynaznar.eu/lexique_nisvai/form_lexique.html).  
Development a Web interface to consult the Nisvai narrative text corpus: dynamic access (Ajax) lexicon, annotations and translation. Linguistic information is provided with their contexts of occurrence.
- 2015 **Nisvai Bilingual Booklet**, [http://jocelynaznar.eu/livrets/recueil\\_v1\\_2015.pdf](http://jocelynaznar.eu/livrets/recueil_v1_2015.pdf).  
This bilingual booklet, the text on the left in Nisvai, the text on the right in French, about 80 pages long and containing 14 nisvais stories or songs recorded at the fields in Malekula, Vanuatu.
- 2015 **Nisvai–French bilingual lexicon**, [http://jocelynaznar.eu/livrets/lexique\\_v1\\_2015.pdf](http://jocelynaznar.eu/livrets/lexique_v1_2015.pdf).  
Lexicon from the corpus of annotated narrative texts. The lexicon proposes entries in Nisvai then in French, for each of its entries, from one to three contextual and referenced illustrations of the use of the term.
- 2012-2021 **Online Archive of Nisvai narrative corpus and fieldwork collection**, Ortolang: <https://www.ortolang.fr/market/corpora/sldr000783/v4>.  
The collection of recordings and annotations resulting from my fieldwork in Vanuatu was first archived on the online archive SLDR. Ortolang then welcomed the archive. The metadata was then updated according to Ortolang's expectations and the evolution of the collection.

---

## Languages

French	Mother Tongue	
English	C2	<i>Studied for years and practised at the university in London</i>
Castillan	B1	<i>Studied for years, little practice</i>
Nisvai	B2	<i>Studied during the fieldwork in Vanuatu</i>
Bislama	C1	<i>Studied during the fieldwork in Vanuatu</i>
German	A1	<i>Studying it in Berlin</i>

I also have basic notions and practice of Italian, Lithuanian and Chinese Mandarin.

---

## Computer Skills

Programming	Python, Php, Javascript	Web	xHTML, CSS, Apache, MySQL
Offices	LaTeX, LibreOffice, Microsoft Office	Misc.	Regular Expressions, XML
Bibliography	Bibtex, Zotero	Data Sciences	Pandas, Jupiter, Statsmodels, Seaborn, Plotly
Annotations	Elan, Transcriber	Documentation	Flex

## Teachings

- 2021 **Lecturer**, *Humboldt Universität*, Berlin, Master degree.  
*Empirical Corpus: From Fieldwork to Data*, 2 hours a week for the Summer Semester.
- 2019 **Intervention in Marie Durand's Pacific Anthropology course**, *Aix-Marseille University*, Aix-en-Provence.  
Nisvai narrative practices, Malekula, Vanuatu: An ethnolinguistic study of "nabol" according to the situation and the speaker
- 2015 **Intervention in Claudia Coppola's Anthropology course, in collaboration with Alice Servy, CREDO – EHESS, Baumettes' prison**, Marseille.  
Field work in Vanuatu : Contextualizing oral literature or "violences" against women, 3 hours
- 2014 **Lecturer in the Department of Language Science**, *Aix-Marseille University*, Marseille.  
CM and TD of Domains of Linguistics 28 hours (SCLA01), TD of Morphology and Syntax 16 hours (SCLC02)
- 2010 **French language courses**, *Beaux-Arts*, Kiev, Ukraine.  
Interventions, for a total of 16h, in the course of French from Professor Victoria Skrybka.

## Academic Communications

- April 2021 **Online Presentation**, *Symposium Education, Cultures, Identity The Ministry of Education and Training*, Port-Vila, the National University of Vanuatu & the Vanuatu Kultural Senta.  
*Le traitement informatisé dans la recherche coopérative: Articuler une recherche ethnolinguistique à l'élaboration de ressources langagières à Malekula, Vanuatu*
- March 2021 **Online Presentation**, *Panel Responsible Documentation?*, ASA2021: Responsibility, University of St Andrews.  
*Getting the speakers' voices into the language resources: Using annotated narratives to build critical language resources for the Nisvai community, Vanuatu*
- February 2021 **From dialogue to grammar - Language emerging from human sociality: the case of speech representation**, *Data Workshop on Reported Speech*, University of Helsinki.  
"The evaluation of narratives by Nisvais speakers: the role of proper names and the speaker's age group (Malekula, Vanuatu)" ; *Silent gaps as a distinctive feature of Direct Speech: a Corpus study of Nisvai Narratives, Vanuatu (virtual)*.
- January 2020 **Intervention in the LPL Seminar**, *Language and Speech Laboratory*, Aix-en-Provence.  
"Telling a Nabol: The production of texts nisvais according to age and enunciation situation, Malekula, Vanuatu" ; *Narrer une nabol : La production des textes nisvais en fonction de l'âge et de la situation d'énonciation, Malekula, Vanuatu*
- November 2019 **Presentation of a poster**, *LIFT 2019*, Orleans.  
"Tree visualization to assist in the study of nisvai sounds." ; *Visualisation en arbres pour assister l'étude des sons du nisvai*.
- February 2018 **Intervention in the EHESS Seminar: The Sound of Human Sciences**, *La Vieille Charité*, Marseille.  
"Transcribing the Sounds of Languages." ; *La transcription des sons des langues*.
- March 2018 **Intervention in the EHESS doctoral student's workshop**, *La Vieille Charité*, Marseille.  
Workshop on bibliographic data management with Zotero.
- November 2017 **Intervention**, *European Society for Oceanists*, Munich, Germany.  
With Alice Fromontel, *Humans and animals on display: A comparative study of the relationship between human and animal characters in Wallisian and Nisvai oral narratives*
- March 2017 **Intervention in the seminar *Actualité de la recherche en Océanie*, "Research news in Oceania"**, CREDO, Marseille.  
"Poetical sequences within Nisvai narrative sequences." ; *Les séquences poétiques dans les pratiques narratives nisvaises*.
- March 2016 **Intervention in the TALEP seminar**, *LIF-TALEP*, Marseille.  
"The corpus of narrative texts nisvais: methodology of a multi-purpose documentary process for an undescribed oral language, southeast of Malekula, Vanuatu." ; *Le corpus de textes narratifs nisvais: méthodologie d'un processus documentaire polyvalent pour une langue orale non décrite, au sud-est de Malékula, Vanuatu*.
- July 2015 **International Conference**, Paris, TRELA2015: Research Areas in Applied Linguistics.  
"From Linguistic Data Collection to Resource Creation : Coordinating the issues of documentation of a language, nisvai (Vanuatu)" ; *De la collecte de données linguistiques à la création de ressources : coordonner les enjeux de la documentation d'une langue, le nisvai (Vanuatu)*

- May 2015 **Study Day**, *Marseille*, CREDO: Languages in Context.  
 "Between tales and discourse: one story, two texts to introduce Nisvai oral literature, Vanuatu" ; *Entre conte et discours : une histoire, deux textes pour présenter la littérature orale nisvai, Vanuatu*
- 2014 **Participating in Maëlle Calandra's Social Science Writing Seminar**, *EHESS*, Marseille.  
 "Introduction to Computerized Collaborative Writing". Intervention in the writing workshop for EHESS Masters students
- 2013 **Participation in the writing seminar in social sciences at Fleur Beauvieux**, *EHESS*, Marseille.  
 "Introduction to Computerized Collaborative Writing". Intervention within the writing workshop for EHESS Master students
- 2012 **LIF-TALEP Seminar**, *LIF-TALEP*, Marseille.  
 "The documentation of nisvai: a still undescribed language" ; *La documentation du nisvai : une langue encore non décrite*

---

## Organization of scientific events

- 21 to 25 May **Member of the organizing committee of the colloquium "Doing Silence: experiences, materialities and powers"** *Faire Silence: expériences, matérialités, pouvoirs*, *La Vieille Charité*, Marseille, CNRS, EHESS, Aix-Marseille University.  
 2019  
 This international and interdisciplinary colloquium considered silence in its practical dimension, as an object, a conduct, an aesthetic or political hold. By welcoming and confronting the points of view of several academic disciplines - anthropology, science studies, history, linguistics, literature, museology, musicology, sociology, information and communication sciences - and artistic practices - music, slam, poetry, cinema, sound - , this colloquium studied how the existence of silence manifests itself through the material, symbolic and political modalities of silence.  
<https://fairesilence.sciencesconf.org/>
- 2018-19 **Co-organization of the EHESS seminar: "The Sound of Human Sciences"** *Le son des sciences humaines*, *La Vieille Charité*, Marseille.  
 For the second year, the seminar alternates between collective discussions of theoretical perspectives and presentations of current research work. One session per month from October to June. Organized by Christelle Rabier, with Laure Mouchart and Luigia Parlati. <https://enseignements-2018.ehess.fr/2018/ue/2118/>
- 2017-18 **Co-organization of the EHESS seminar: "The Sound of Human Sciences"** *Le son des sciences humaines*, *La Vieille Charité*, Marseille.  
 This seminar questions the historical and political links between the social sciences and sound, as well as debating their contemporary renewal through the arts. One session per month from October to June. Organized by Christelle Rabier, with Laure Mouchart and Luigia Parlati.
- 2017-18 **Co-organization of the EHESS doctoral student's workshop**, *La Vieille Charité*, Marseille.  
 One session per month from October to June, with Marjolaine Martin, Léa Renucci and Solenn Thircuir.
- 2014-15 **Coordinator of the inter-laboratory anthropology seminar** *Plus belle la mort: comment les morts façonnent les sociétés*, *CREDO*, *Idemec*, *IrAsia*, *Centre Norbert Elias*, *Imaf*, Aix-en-Provence Marseille.  
 Organization and management of the different stages of to the two-day seminar. Definition of the planning, preparation of the topic, coordination of the various laboratories, invitation of the researchers.
- 2014 **JEP-TALN-RECITAL**, *Laboratoire Parole Langage*, Marseille.  
 Participation in the organization of the international conference *Traitement Automatique des Langues Naturelles* : "Natural Language Processing".
- 2013-14 **Participation in the organization of the interlaboratory seminar** *Prise et emprise des drogues : les sociétés observées dans leurs substances* : "Taking and control of drugs: the societies observed in their substances", *CREDO*, *Idemec*, *IrAsia*, *Centre Norbert Elias*, *Imaf*, Aix-en-Provence Marseille.  
 Participation in the preparation of the topic, invitation of the researchers and organization of the two days' seminar.
- 2013 **Co-organizing the seminar** *Problèmes de santé en région Pacifique Sud* : "Health issues in South Pacific", *CREDO*, Marseille.  
 With Maëlle Calandra, Phd in Anthropology. Intervenant: Professor Fabrice Simon, doctor and researcher at the military hospital of Laveran.

---

## Scientific Outreach and Miscellaneous Communications

- April 2020 **Online Article**, *LinuxFR*, <https://linuxfr.org/news/comparaison-de-deux-outils-de-gestion-de-seminaires>  
 "Comparison of Two Webinar Management Tools: BigBlueButton and OpenMeetings" ; *Comparaison de deux outils de gestion de séminaires en ligne : BigBlueButton et OpenMeetings*
- 2019 **Online Article**, *LinuxFR*, <https://linuxfr.org/news/python-pour-les-sciences-une-presentation>.  
 "Python for Science, a presentation" ; *Python pour les sciences, une présentation*.

- 2018 **EN.QUETE! Anthropology Festival**, *la Vieille Charite*, MAAOA, CNE, CREDO, EHESS, IMAF and le Tamis.  
*La cuisine de l'enquête* : The concept was to present how a researcher's work is developed and how it relates to the object of their studies. Using a collected object or an original document, the researchers presented their work. Presentations were open to the general public and lasted 5 to 10 minutes.
- 2017 **Audio and photographic Exhibition**, *Café Villageois de Lauris*, With Raphaël Gravagna and Maëlle Calandra.  
 "Lokol Stael: an ethnographic point of view of the everyday life in Vanuatu (Malekula and Tongoa)."; "*Lokol stael*": *regards ethnographiques sur le quotidien au Vanuatu (Malekula et Tongoa)*
- 2016 **Audio and photographic Exhibition**, *Aix-Marseille Université, espace Fernand Pouillon*, Marseille, With Raphaël Gravagna and Maëlle Calandra.  
 "*Lokol stael*": *regards ethnographiques sur le quotidien au Vanuatu (Malekula et Tongoa)*
- 2016 **Narrative Discussion**, *Alcazar Library*, A Story of Stories, Marseille.  
 "A Story of Stories: An Encounter with of people and their stories, in Vanuatu, in the South-East of Malekula."
- 2015 **Oral presentation**, *Lycée Louis Antoine de Bouguinville*, Port Vila, Vanuatu.  
 "Study of the Oral Literature of the Nisvai community, southeast of Malekula: Presentation of a linguistic research"
- 2014 **"Science Fair"**, *Fête de la science, CREDO-IRASIA-MAP*, Marseille.  
 Presentation to school pupils and college students about the workings of a research laboratory and the various activities of a researcher in the humanities.
- 2013 **Oral presentation**, *Lycée Louis Antoine de Bouguinville*, Port Vila, Vanuatu.  
 Presentation given to 14th graders, with Bergmans Iati and Maëlle Calandra: "What is Linguistics? An Introduction to Linguistics at the University."
- March 2013 **Online Article**, *LinuxFR*, <https://linuxfr.org>.  
 "Release of Ardour 3"; *Sortie d'Ardour 3*, a free software to mix, edit and compose music and recordings.
- avril 2008 **Online Article**, *LinuxFR*, <https://linuxfr.org/news/le-libre-et-les-memoires-de-traduction>.  
 "Free Software and Translation Memories"; *Le libre et les mémoires de traduction*
- April 2008 **Speech Presentation**, *Grafotech Open Source Software Forum*, Rennes.  
 "Using of Open Source Software to Assist Translation"; *Utilisation des logiciels libres d'aide à la traduction*

---

## Other Professional Experience

- 2019 **Self-entrepreneur**, Marseille.  
 Self-enterprise providing services as a research engineer for projects involving data processing and production and humanities problematic.
- 2016 – 2017 **Teacher's Assistant in Mathematics**, *Lycée professionnel René Caillié*, Marseille.  
 Personalized and individual follow-up of students of different levels of education and courses. Work with newcomers learning French as a foreign language, on vocabulary and basic notions of mathematics.
- 2016 **Study Engineer**, *Consortium des archives des ethnologues, CNRS*, Marseille, <http://ethnologues.huma-num.fr>.  
 Development of a perennial archive protocol for ODSAS.net, an online archive. Writing a report explaining the issues surrounding digital perennial archiving, the current practices and describing the archiving protocol proposal.
- 2015 **Web Developer**, *European Consortium for Pacific Studies (ECOPAS)* [www.ecopas.info](http://www.ecopas.info) & <http://pacific-studies.net>, Marseille.  
 Development of modules for the ECOPAS project website to facilitate the consultation of the database. A first module allows the visualization of researchers as a cloud of keywords. A second module improved the query results from users. ECOPAS is a research programme funded by the European Union.
- 2013 **Web Developer**, *ECOPAS* [www.ecopas.info](http://www.ecopas.info) & <http://pacific-studies.net>, Marseille.  
 Development of a Web interface (PHP, HTML, Javascript/Jquery) and a database (MySQL) in order to develop a skill based network for social scientists and humanities researchers operating in the Pacific.
- 2007–2008 **Internships in technical writing and translation**, *OpenTrust, CRAIe*, Paris.  
 Writing the user and administrator guides of the company's software. Creation of a multilingual dictionary in the fields of electronics and tourism.

---

## Centres of Interest

- Music Practice of several instruments: piano, percussion, didgeridoo, and composition of acousmatic music.
- Free/Libre software Participates in the development of free software and works (Wikipedia, LinuxFr, Fedora).
- Yoga & Dance Regular practice of yoga and contemporary dance.