

Personal information

E-mail benoit.sagot@inria.fr

Website <http://pauillac.inria.fr/~sagot/>

Professional experience

Research

- Since 2017 **Head of Inria Paris' ALMA_nCH project-team** (Automatic Language Modelling and Analysis & Computational Humanities), *Inria Paris' research team specialised in Natural Language Processing (NLP) and Digital Humanities (DH)*, Paris, France.
Directeur de Recherche Inria (2020–), formerly Chargé de Recherche Inria (until 2020)
Holder of a chair in the PRAIRIE research institute (2019–2024)
Holder of the annual chair on Informatics and Digital Sciences at the Collège de France (2023–2024)
- 2014–2016 **Head of Inria Paris' ALPAGE project-team** (Analyse Linguistique Profonde à Grande Échelle), *joint research team specialised in NLP involving Inria and Université Paris-Diderot*, Paris and Rocquencourt, France.
Chargé de Recherche Inria (1st class)
- 2007–2013 **Researcher in Inria Paris' ALPAGE project-team**, Paris and Rocquencourt, France.
Chargé de Recherche Inria (2nd then 1st class)
- 2006–2007 **Researcher in Inria Futurs' SIGNES project-team** (Signes linguistiques, grammaire et sens: algorithmique logique de la langue), *an Inria research team within the UMR LabRI*, Bordeaux, France.
Chargé de Recherche Inria (2nd class)
- 2002–2006 **Ingénieur du Corps des Télécommunications seconded at Inria, Atoll** (*Atelier d'outils logiciels pour le langage naturel*), *Inria research team*, Rocquencourt, France.
Seconded to prepare my PhD, supervised by Laurence Danlos (Université Paris-Diderot)
- 2002 (4 months) **DEA (Master's) work placement**, *Institut Jean-Nicod*, Paris, France.
Internship in cognitive linguistics: theoretical and comparative study of several spatial relations, under the supervision of Richard Carter (Institut Jean Nicod, EHESS)
- 2001 (6 months) **Research placement**, *IBM France*, Paris, France.
Work placement in NLP: development of a natural language understanding application, under the supervision of Claire Waast-Richard
- 2000 (4 months) **Research placement**, *Johns Hopkins University*, Baltimore, USA.
Work placement in astrophysics within the scientific team of NASA's FUSE probe (UV astronomy)

Teaching

- 2023–2024 **General audience**, *Collège de France*, Paris, France.
Inaugural lecture followed by 8 lectures, each followed by an invited lecture
- Since 2017 **Graduate (5th year)**, *ENS Cachan*, Paris, France.
Course in NLP in the Mathematics, Vision and Machine Learning curriculum (24hrs/yr, with Emmanuel Dupoux)

- 2009–2015 **Graduate (5th year)**, *Université Paris-Diderot*, Paris, France.
Course in NLP in the computational linguistic curriculum (24hrs/yr)
- 2010–2011 **Undergraduate (3rd year)**, *Université Paris-Diderot*, Paris, France.
Introductory course to NLP in the computer science curriculum (24hrs/yr)
- 1998–2000 **“Classes préparatoires” (2nd year)**, *École Sainte-Geneviève*, Versailles, France.
Weekly oral exams in mathematics in the highly selective course (option “MP*”) in preparation for the competitive admission examinations to French “Grandes Écoles d’Ingénieur”

Degrees

- 2018 **Habilitation à diriger les recherches in Computer Science**, *Sorbonne Université*, Paris, France,
“*Informatiser le lexique: Modélisation, développement et exploitation de lexiques morphologiques, syntaxiques et sémantiques*” [Computerising the lexicon: Modelling, development and use of morphological, syntactic and semantic lexicons],
Mentor: Laurent Romary (Inria).
Other committee members: Philippe Blache (CNRS, reviewer), James P. Blevins (University of Cambridge, reviewer), Christiane D. Fellbaum (Princeton University), Ludovic Denoyer (Sorbonne Université), Anna Korhonen (University of Cambridge), Gertjan Van Noord (University of Groningen)
- 2006 **PhD in computer science**, *Université Paris-Diderot*, Paris, France,
“*Analyse automatique du français: lexiques, formalismes, analyseurs*” [Automatic analysis of French: lexicons, formalisms, parsers], *Mention très honorable avec les félicitations du jury* [highest distinction]
Supervision: Laurence Danlos (Full professor, Université Paris-Diderot); Co-supervision: Éric Villemonte de La Clergerie (Chargé de Recherches, Inria).
Other committee members: Philippe Blache (CNRS, reviewer), Gérard Huet (Inria, reviewer), John Carroll (University of Sussex), Pierre Boullier (Inria, invited)
- 2002 **DEA (Master 2) in artificial intelligence**, *Université Pierre et Marie Curie*, Paris, France, DEA IARFA (Artificial Intelligence, Pattern Recognition and Applications).
- 2002 **Engineering degree from Télécom ParisTech as a member of the Corps des Télécommunications**, *Télécom ParisTech*, Paris, France, Specialisation in computer science, algorithmics and NLP.
- 2000 **Engineering degree from École polytechnique**, *École polytechnique*, Palaiseau, France, Specialisation in theoretical physics.

Research activities

Responsibilities within research projects

- | | | |
|-----------|--|--|
| 2022– | National ANR project CulturIA | Local head for ALMA _{na} CH
<i>Towards a cultural history of AI</i> |
| 2021–2022 | BigScience initiative | Co-leader of the tokenisation working group
<i>Development of a very large open-source multilingual generative language model</i> |
| 2019–2024 | PRAIRIE (PaRis AI Research InstitutE) | Chair (2019–2024)
<i>Research Institute on Artificial intelligence headed by Isabelle Ryl</i> |

2011–2019	LabEx EFL (Empirical Foundations of Linguistics)	Head (2011–2015) then deputy head (since 2015) of one of the 7 research strands (strand 6 “Language Resources”), member (2011–2015) then deputy member (since 2015) of the Executive Board, head of several research operations. <i>LabEx headed by Jacqueline Vaissière (Université Paris 3, 2011–2014) then Christian Puech (Université Paris 3, since 2014)</i>
2016–	National ANR project Profiterole	Workpackage leader <i>Development of language resources and NLP tools for various stages of French with a focus on Old and Middle French, in order to model the diachronic evolution of the language</i>
2015–2017	National FUI project VerDI	Local head for ALPAGE then ALMAnaCH <i>Automatic identification of information concealment, especially in online journalistic texts</i>
2012–2016	National ANR project ASFALDA	Task co-leader <i>Development of a French FrameNet (lexicon, annotated corpus) and associated tools. PI: Marie Candito (ALPAGE)</i>
2011–2015	National FUI project PACTE	Local head for ALPAGE <i>Large-scale automatic correction of OCR outputs. Company leading the project: Numen Digital</i>
2010–2013	National ANR project EDyLex	Project leader (PI) <i>Dynamic extension of lexical resources</i>
2010–2012	Bilateral PROTEUS project (France-Slovenia)	Co-leader <i>Development and exploitation of parallel and comparable French-Slovene corpora for wordnet development. Other co-leader: Mojca Schalmberger Brezar (University of Ljubljana)</i>
2009–2011	National ANR project SEQUOIA	Local head for ALPAGE <i>Probabilistic parsing for French. PI: Alexis Nasr (Université de la Méditerranée)</i>
Participation in research projects		
2017–2023	National ANR grant BASNAGE	<i>Digitalisation and automatic exploitation of Basnage’s 1701 Dictionnaire Universel. PI: Laurent Romary (Inria)</i>
2016–2022	National ANR grant PARSITI	<i>Parsing and machine translation of noisy textual data as found on social media and more generally on the web. PI: Djamel Seddah (Université Paris-Sorbonne)</i>
2016–2021	Bilateral ANR-NSF project MCM-NL (France/USA)	<i>Investigation of correlations between neuroimaging data (fMRI, EEG) and automatic parsers, based on data from the Petit Prince read in both English and French. PI: John Hale (Cornell University)</i>
2015–2021	National ANR project SoSweet	<i>Parsing noisy data from Twitter for the joint sociolinguistic and graph-based analysis of the network of user relations. PI: Jean-Philippe Magué (ENS Lyon)</i>
2012–2018	Corpus Écrits then CORLI consortiums within the Huma-Num infrastructure	Member of the Board, participation in several research operations. In particular, participation in the CoMéRé initiative led by Thierry Chanier (Université Blaise-Pascal) on computer-mediated content Consortiums led by the Institut de Linguistique Française, under the supervision of Franck Neveu (Université Paris-Sorbonne)

2012–2014	National IARPA “BABEL Program” (USA)	External expert <i>Fast and language-independent development of information retrieval systems for speech data. Expertise for the unsupervised morphological analysis task (project Lorelei) led by Owen Rambow and Nizar Habash (Columbia University)</i>
2009–2011	Bilateral ANR-DFG project PerGram (France/Germany)	<i>Development of linguistic descriptions and resources (lexicon, grammar) for Persian. PIs: Pollet Samvelian (Université Paris 3) and Stefan Müller (Freie Universität Berlin)</i>
2007–2009	National ANR project PASSAGE	<i>Automatic development and exploitation of large-scale treebanks. PI: Éric de La Clergerie (ALPAGE)</i>
2005–2007	National ILF project LexSynt	<i>Syntactic lexicons for French. PI: Sylvain Kahane (Université Paris X)</i>
2004, 2007	National Technolangue project EASy	<i>National French parser evaluation campaign. Participant with the SxLFG parser, and indirect participant with the Lefff and the SxPipe shallow processing chain, use by both the SxLFG and the FRMG parsers)</i>

Research visits

2013	Columbia University, USA	<i>One-week visit as part of a collaboration with Owen Rambow and Nizar Habash</i> Formal and computational morphology, within the IARPA project “BABEL” (see above)
2008–2012	University of Ljubljana, Slovenia	<i>Several one-week visits as part of a collaboration with Darja Fišer</i> Development of wordnets for French and Slovene, partly within the above-mentioned bilateral PROTEUS project
2007–2011	University of Vigo, Spain, and University of Nice, France	<i>Three short stays</i> Development of lexical resources for French, Spanish and Galician, in relation with the Galician “Victoria” project
2009	University of Padova, Italy	<i>One-week stay as part of a collaboration with Giorgio Satta</i> Formal properties of mildly context-sensitive languages
2007	Polish Academy of Sciences, Warsaw, Poland	<i>Two-month stay</i> Computational morphology of Polish, development of morphological lexicons, improvement of the Polish national corpus

Supervision

PhD supervision

2023–	You Zuo	<i>Neural approaches to patent processing</i> PhD thesis co-funded by the start-up company qatent <i>Other supervisor: Kim Gerdes (qatent)</i>
2023–	Wissam Antoun	<i>Automatic detection of language-model-generated content</i> <i>Main supervisor: Djamé Seddah (ALMAAnCH)</i>
2021–	Nathan Godey	<i>Adversarial approaches to neural language modelling</i> PhD funded by B. Sagot’s PRAIRIE chair <i>Other supervisor: Éric Villemonte de La Clergerie (ALMAAnCH)</i>

2021–	Lydia Nishimwe	<i>Robust Neural Machine Translation</i> PhD funded by R. Bawden's PRAIRIE chair <i>Other supervisor: Rachel Bawden (ALMAAnaCH)</i>
2021–	Roman Castagné	<i>Linguistically informed neural language modelling</i> PhD funded by B. Sagot's PRAIRIE chair <i>Other supervisor: Éric Villemonte de La Clergerie (ALMAAnaCH)</i>
2021–	Matthieu Futeral-Peter	<i>Text-image multimodal models</i> PhD jointly funded by Inria, by B. Sagot's PRAIRIE chair and by the WILLOW team at Inria Paris <i>PhD director: Ivan Laptev (WILLOW). Other supervisors: Rachel Bawden (ALMAAnaCH), Cordelia Schmid (WILLOW)</i>
2021–	Paul-Ambroise Duquenne	<i>Study of vector spaces for sentence representation</i> PhD thesis co-funded by the META Artificial Intelligence Research lab <i>Other supervisor: Holger Schwenk (META AI Paris)</i>
2021–	Tu Anh Nguyen	<i>Unsupervised acquisition of linguistic representations from speech (audio) data</i> PhD thesis co-funded by the META Artificial Intelligence Research lab <i>Other supervisor: Emmanuel Dupoux (META AI Paris)</i>
2019–	Robin Algayres	<i>Unsupervised segmentation and lexicon extraction from raw speech data</i> State-funded PhD <i>PhD director: Emmanuel Dupoux (CoML).</i>
2019–2022	Clémentine Fourier	<i>Neural models of language evolution</i> Inria-funded PhD Defended on 26th September 2022
2018–2022	Benjamin Muller	<i>How Can We Make Language Models Better at Handling the Diversity and Variability of Natural Languages ?</i> PhD thesis funded by the SoSweet and ParSiTi national (ANR) projects <i>Other supervisor: Djamé Seddah (ALMAAnaCH)</i> Defended on 17th November 2022
2018–2022	Pedro Javier Ortiz Suárez	<i>Neural approaches for the automatic analysis of and information extraction from legacy dictionaries</i> PhD thesis funded by the BASNUM national (ANR) project <i>PhD director: Laurent Romary (ALMAAnaCH)</i> Defended on 27th June 2022
2018–2021	Louis Martin	<i>Automatic text simplification</i> PhD thesis co-funded by the Facebook Artificial Intelligence Research lab <i>Other supervisors: Éric Villemonte de La Clergerie (ALMAAnaCH), Antoine Bordes (FAIR Paris, Facebook)</i> Defended on 27th October 2021
2012–2015	Marion Baranes	<i>Normalisation of noisy text for opinion mining</i> PhD thesis co-funded by the viavoo company <i>PhD director: Laurence Danlos (ALPAGE).</i> Defended on 23rd October 2015, with honours
2011–2015	Valérie Hanoka	<i>Semi-automatic construction and extension of multilingual lexical networks</i> PhD thesis co-funded by the Verbatim Analysis company <i>PhD director: Laurence Danlos (ALPAGE).</i> Defended on 6th July 2015, with honours
2010-2013	Pierre Magistry	<i>Unsupervised Word Segmentation and Wordhood Assessment. The case for Mandarin Chinese</i> State-funded PhD <i>PhD director: Sylvain Kahane (Université Paris X); other co-supervisor: Marie-Claude Paris (Université Paris-Diderot).</i> Defended on 19th December 2013, with the highest honours

2009–2013 **Rosa Stern** *Automatic identification of named entities to enrich textual data*
PhD thesis co-funded by the Agence France-Pressé
PhD director: Laurence Danlos (ALPAGE).
Defended on 28th June 2013, with honours

Post-doc supervision

2019–2020 **Gaël Guibon** *Modelling the Old French lexicon and its diachronic evolution*
National (ANR) project Profiterole

2016–2017 **Hector Martínez Alonso** *Neural part-of-speech tagging, Automatic identification of omissions in journalistic texts*
ANR-RAPID project VerDI

2013–2015 **Kata Gábor** *Identification and correction of domain-specific named entities in OCR outputs*
FUI project PACTE

2012–2013 **Damien Nouvel** *Dynamic extension of lexical resources using dynamic corpora (news wires)*
National (ANR) project EDyLex

2012–2013 **Yves Scherrer** *Lexicon development and part-of-speech tagging for non-resourced languages based on closely related resources languages*
LabEx (excellency cluster) EFL

2010–2011 **Yayoi Nakamura–Delloye** *Automatic extraction of relations between named entities*
National (ANR) project EDyLex

2010–2011 **Marianna Apidianaki** *Automatic extension of semantic lexicons (wordnets)*
National (ANR) project EDyLex

2009–2011 **Sattisvar Tandabany** *Algorithmics of non-contextual probabilistic parsing*
National (ANR) project Sequoia

Research engineer supervision

2022– **Anna Chepaikina** *Automatic generation of wine tasting notes*
Bilateral partnership with the Winespace start-up company

2022– **Wissam Antoun** *Hate speech detection in social media texts. Main supervisor: Djamé Seddah*
Funded jointly by B. Sagot's PRAIRIE chair and the European project CounteR

2022– **Rua Ismail** *Language identification for large-scale multilingual raw corpus development*
Funded by B. Sagot's PRAIRIE chair

2021–2022 **You Zuo** *Fine-grained patent classification. Co-supervised with Kim Gerdes*
Bilateral partnership with INPI, the French patent office

2021 **Thomas Wang** *Neural language models with linear attention*
Funded by B. Sagot's PRAIRIE chair

2021 **Arij Riabi** *NLP for low-resource, non-standardised language varieties, especially North-African dialectal Arabic written in Latin script. Main supervisor: Djamé Seddah*
Funded jointly by B. Sagot's PRAIRIE chair and the European project CounteR

2021– **Julien Abadji** *Large-scale multilingual corpus development and extension (OSCAR corpus). Co-supervised with Pedro Ortiz Suarez*
Funded by B. Sagot's PRAIRIE chair

2021–2022 **Thibault Charmet** *Similarity detection in Cour de cassation rulings. Co-supervised with Rachel Bawden*
Government-funded collaboration (LabIA)

Master 2 dissertation supervision

2021–2022	Camille Rey	<i>Inalco master in NLP. Co-supervised with Rachel Bawden. Améliorer la désambiguïstation lexicale en traduction automatique neuronale [Improving lexical disambiguation in neural machine translation].</i>
2021	Roman Castagné	<i>“Mathématiques, Vision, Apprentissage” master at École Normale Supérieure (MVA). Quelle tokenisation pour les modèles de langue ? [Which tokenisation for language models?]</i>
2021	Matthieu Futeral-Peter	<i>“Mathématiques, Vision, Apprentissage” master at École Normale Supérieure (MVA). Co-supervised with Cordelia Schmid (WILLOW), Rachel Bawden (ALMAAnaCH) and Ivan Laptev (WILLOW). Exploration of multilingual and multimodal word embeddings</i>
2019	Charlotte Rochereau	<i>Cognitive Science master at École Normale Supérieure (CogMaster). Co-supervised with Emmanuel Dupoux (CoML). Comparing human judgments and neural language models’ assessment of argument order in German.</i>
2014	Sacha Beniamine	<i>Computational linguistics Masters at Université Paris-Diderot. Towards a linguistically motivated treatment of French multi-token units in constituency parsing</i>

Member of PhD and Habilitation committees

2022	Léo Laugier	<i>Committee president PhD in computer science, Télécom Paris, France Analysis and Control of Online Interactions through Neural Natural Language Processing PhD supervisors: Thomas Bonald (Télécom Paris), Lucas Dixon (Google)</i>
2021	Hicham Boukkouri	EI <i>Reviewer PhD in computer science, Université Paris-Saclay, France Domain Adaptation of Word Embeddings Through the Exploitation of In-domain Corpora and Knowledge Bases PhD supervisors: Pierre Zweigenbaum, Olivier Ferret, Thomas Lavergne</i>
2021	Adelle Abdallah	<i>Reviewer PhD in computer science, Université Paris 8 Vincennes à Saint-Denis, France & Université Libanaise, Lebanon Catégorisation sémantique et information grammaticale en arabe PhD supervisors: Gilles Bernard, Mohammad Hajjar</i>
2020	Silvia Mendéz	García <i>Reviewer PhD in computer science, Universidad de Vigo, Spain Contribution to Natural Language Generation for Spanish PhD supervisors: Enrique Costa Montenegro, Milagros Fernández Gavilanes</i>
2020	Alice Millour	<i>Reviewer (“rapporteur”) PhD in computer science, Sorbonne Université, France Myriadisation de ressources linguistiques pour le traitement automatique de langues non standardisées PhD supervisors: Claude Montacié (Sorbonne Université) and Karën Fort (Sorbonne Université)</i>
2020	Jack Bowers	<i>Committee member Language Documentation and Standards in Digital Humanities : TEI and the documentation of Mixtepec-Mixtec PhD in linguistics, EPHE, France PhD supervisor: Laurent Romary (Inria)</i>
2020	Jacobo Abitbol	Levy <i>Reviewer (“rapporteur”) PhD in computer science, Université de Lyon, France Computational detection of socioeconomic inequalities PhD supervisors: Éric Fleury (Inria) and Márton Karsai (Central European University & Inria)</i>

- 2019 **Hazem Al Saied** *Committee member*
PhD in computer science, Université de Lorraine, France
Analyse automatique par transitions pour l'identification des expressions polylexicales
PhD supervisors: Matthieu Constant (Université de Lorraine), Marie Candito (Université Paris-Diderot)
- 2019 **Tamara Álvarez López** *Reviewer ("rapporteur")*
PhD in computer science, University of Vigo, Spain
Análisis de Opinión en Redes Sociales mediante Técnicas de Procesamiento del Lenguaje Natural
PhD supervisors: Enrique Costa Montenegro and Milagros Fernández Gavilanes (University of Vigo)
- 2019 **Nourredine Alliane** *Reviewer ("rapporteur")*
PhD in computer science, Université Paris 8, France
Evaluation des représentations vectorielles de mots
PhD supervisor: Gilles Bernard (Université Paris 8)
- 2018 **Sébastien Delecraz** *Reviewer ("rapporteur")*
PhD in computer science, Aix-Marseille Université, France
Approches jointes texte/image pour la compréhension multimodale de documents
PhD supervisors: Frédéric Béchet, Alexis Nasr (Aix-Marseille Université)
- 2018 **Kim Gerdes** *Committee president*
Same same but different: Paradigms in syntax
Habilitation in linguistics, Université Paris X Nanterre, France
Habilitation mentor: Sylvain Kahane (Université Paris X Nanterre)
- 2015 **Wajdi Zaghouni** *Committee member*
Le développement de corpus annotés pour la langue arabe
PhD in computer science, Université Paris X Nanterre, France
PhD supervisor: Sylvain Kahane (Université Paris X Nanterre)
- 2014 **Édouard Grave** *Committee member*
A Markovian approach to distributional semantics
PhD in computer science, Université Pierre et Marie Curie, France
PhD supervisors: Francis Bach (Inria) and Guillaume Obozinski (École Nationale des Ponts et Chaussées)
- 2011 **Rania Voskaki** *Committee member*
Le lexique-grammaire des verbes du grec moderne : constructions transitives non locatives à un complément d'objet direct
PhD in computational linguistics, Université Paris-Est Marne-la-Vallée, France
PhD supervisor: Tita Kyriakopoulou
- 2010 **Claire Mouton** *Committee member*
Ressources et méthodes semi-supervisées pour l'analyse sémantique de texte en français
PhD in computer science, Université Paris Sud, France
PhD supervisor: Anne Vilnat (Université Paris Sud). Co-supervisor: Gaël de Chalendar (CEA)
- 2010 **Lionel Nicolas** *Committee member*
Efficient production of linguistic resources : the Victoria project
PhD in computer science, Université de Nice, France
PhD supervisor: Jacques Farré (Université de Nice)
- 2009 **Juan Otero Pombo** *Reviewer*
Análisis léxico robusto
PhD in computer science, Universidade de Vigo, Spain
PhD supervisors: Manuel Vilares Ferro (Universidade de Vigo) and Jorge Graña Gil (Universidade de A Coruña)

2008 **Laurence Delort** *Committee member*
L'expression des relations temporelles dans le discours : explication et formalisation de quelques emplois de "avant"
PhD in linguistics, Université Paris Diderot
PhD supervisor: Laurence Danlos

Community activities

2019– **Standardisation** *Project leader on the standardisation of inflectional morphology within the Working Group 4 (Lexical resources) of the Subcommittee 4 (Language resource management) of the Technical Committee 37 (Language and terminology)*

2018 **Invited co-editor** *Co-editor, with Olivier Bonami, of the special issue of the Morphology journal on computational approaches to morphology*

2012 **Invited co-editor** *Co-editor, with Núria Bel, of the special issue of the Traitement Automatique des Langues journal on language resources*

2016– **Responsibility within Inria** *Member of the Executive Board of the Inria Paris research centre*

2011–2019 **Responsibility within Inria** *Member of the working group on international relations of the Inria Scientific and Technological Committee*

2015– **Scientific Board** *Member of the Scientific Council of the EquipEx Ortolang (French national infrastructure for language resources)*

2012– **Scientific Board** *Member of the Executive Committee of the Corpus Écrits then CORLI consortiums, within the TGIR (very large research infrastructure) Huma-Num*

2011–2019 **Scientific Board** *Member of the Restricted Scientific Council (RSC) of the LabEx (excellency cluster) EFL, former head of the research strand 6 Language Ressources — 2011–2015, then deputy head and alternate member of the RSC*

2016 **Workshop co-chair** *Workshop on "Computational methods for descriptive and theoretical morphology" co-located with the 2016 edition of the International Morphology Meeting (Vienna, Austria). Co-organised with Olivier Bonami*

2011 **Workshop chair** *WoLeR, International Workshop on Lexical Resources, co-located with the ESSLLI 2011 summer school (Ljubljana, Slovenia)*

2016 **Expertise** *Expert for the French Research Evaluation Agency HCERES, member of the evaluation committee for the ATILF laboratory*

2011– **Expertise** *Expert for the French national funding agency ANR, in charge of the evaluation of research proposals in several fields (humanities and social sciences, information technology, generic)*

2005– **Programme and reviewing committees** *Member of programme or reviewing committees for a number of journals and conferences such as ACL, EMNLP, EACL, CoLing, IJCNLP, Computational Linguistics, Language Resources and Evaluation, Natural Language Engineering, Journal of Language Modelling and Traitement Automatique des Langues*

2017– **Learned society** *Elected member of the Board, former Deputy Treasurer and current Administrator of the Société de Linguistique de Paris*

2005–2016 **Learned society** *Former elected member of the Board (2007-2016) and Secretary (2010-2013) of ATALA, the French learned society for NLP*

2014–2016 **Learned society** *Member of the Permanent Committee of the TALN and RECITAL conferences, representing the Board of ATALA*

Languages

Native	French
Fluent	English
Intermediate	German
Conversational	Slovak

Publications

My full publication list can be accessed on the HAL platform (<http://hal.inria.fr>) and on my web page (<http://pauillac.inria.fr/~sagot/index.html#allpubs>).

My publication list includes:

- 5 book chapters,
- 13 articles in international journals, including three articles in *Language Resources and Evaluation*, two in *Linguisticæ Investigationes*, one in the *Transactions of the Association for Computational Linguistics*, one in the *Journal of Language Modelling*, one in *Indogermanische Forschungen*, and one in the *Münchener Studien zur Sprachwissenschaft*,
- 7 articles in French language journals, including four in *Traitement Automatique des Langues*,
- 2 co-edited journal issues for the *Traitement Automatique des Langues* (2011) and *Morphology* (2018) journal,
- over 120 articles in the proceedings of international peer-reviewed conferences, including four papers presented at ACL and two presented at CoLing.
- over 40 articles in the proceedings of national peer-reviewed conferences, mostly in France but also in Spain and Slovenia