

Rachel Bawden

Inria Researcher in Natural Language Processing

Machine Translation (MT), Context-aware NLP and MT, NLP, Low-Resource NMT

Personal details

Date of birth	04/10/1988	Nationality	British, naturalised French
E-mail	rachel.bawden@inria.fr	Website	https://rbawden.github.io

Education and degrees

- 2019 **French lecturer qualifications**, *CNU*, Sections 7 (Linguistics) and 27 (Informatics).
- 29/11/2018 **PhD in computer science**, *Université Paris-Sud*, Going beyond the sentence: Contextual Machine Translation of Dialogue.
- Supervisors: Sophie Rosset (LIMSI, CNRS) and Thomas Lavergne (LIMSI, Univ. Paris-Sud)
 - President: Nicolas Sabouret (LIMSI, Univ. Paris-Sud)
 - Reviewers: Jörg Tiedemann (University of Helsinki) and Loïc Barrault (Université du Mans)
 - Examiners: Lucia Specia (University of Sheffield and Imperial College) and Andrei Popescu-Belis (Haute École D'Ingénierie et de Gestion du Canton de Vaud)
- 2013–2015 **Master's degree in Computational Linguistics**, *Université Paris-Diderot*, First class ("mention très bien"). Machine learning, syntactic parsing, computational semantics, probabilities for NLP, machine translation, formal languages.
- 2012–2013 **Master's degree (first year) in Linguistic Engineering**, *Université Sorbonne Nouvelle*, First class ("mention très bien"). Corpus processing, formal grammars, lexicography, computational phonetics.
- 2007–2011 **Bachelor of Arts in French and Linguistics**, *Oxford University*, First Class. Phonetics/phonology, psycholinguistics, general linguistics, syntax, historical linguistics, sociolinguistics, French linguistics.
- 2005–2007 **A-levels**, *The Latymer School, Edmonton, London*, French, English, Biology, Geography, four grade "A"s.

Professional experience

Research

- 2020–present **Researcher (chargée de recherches)**, *Inria, Paris in the ALMAAnaCH project-team*, Natural Language Processing and Machine Translation.
- 2018–2020 **Research associate**, *ILCC, University of Edinburgh*, Neural machine translation for low resourced languages, Working with Alexandra Birch.
- 2015–2018 **PhD in computer science**, *LIMSI, CNRS, Univ. Paris-Sud*, Going beyond the sentence: Contextual Machine Translation of Dialogue, Supervised by Sophie Rosset and Thomas Lavergne.
- May–Aug 2017 **Research visit**, *ILCC, University of Edinburgh*, Neural machine translation for discourse. Collaboration with Alexandra Birch, Barry Haddow and Rico Sennrich.
- Feb–Aug 2015 **Master's research placement**, *Alpage (Inria & Université Paris-Diderot)*, Boosting for model selection for statistical syntactic parsing. Supervised by Benoit Crabbé.

- May–Aug 2014 **Master's research placement**, *LaTTiCe (CNRS & Université Sorbonne Nouvelle) and Télécom ParisTech*, Modelling communicative acts for an embodied conversational agent. Supervised by Frédéric Landragin and Chloé Clavel.
- June–Aug 2013 **Master's research placement**, *MoDyCo Laboratory (CNRS & Université Paris-Ouest)*, Development and semi-automatic correction of a dependency treebank of spoken French. Supervised by Sylvain Kahane.
- 2013 **Translation proof-reading and revision**, *Lucien Tesnière, Elements of structural syntax, published in French in 1959, translated by Timothy Osborne and Sylvain Kahane, John Benjamins, Amsterdam, 2015.*
- March–June 2013 **Research placement**, *MoDyCo Laboratory (CNRS & Université Paris-Ouest)*, Annotation of a treebank of spoken French for syntactic dependencies.

Participation in research projects

- 2019–2020 **UKRI fellowship**, *MTStretch: Low Resource Machine Translation*, Held by Alexandra Birch, University of Edinburgh, UK.
<https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/S001271/1>
- 2019–2020 **EU project work package leader**, *GoURMET (Global Under-Resourced Media Translation)*, Leader of the work package on morphological modelling, University of Edinburgh, UK, in collaboration with the University of Alicante, the University of Amsterdam, the BBC and Deutsche Weller.
<https://gourmet-project.eu>

Teaching

- 2016–17, 2017–18 **Introduction to Natural Language Processing**, *4th year*, Polytech Paris-Sud, Course creator and supervisor: lectures, tutorials and practical classes (30hrs/year).
- 2017–18 **Algorithmics and C++**, *2nd year*, Polytech Paris-Sud, Practical classes (10hrs).
- 2015–16, 2016–17 **Databases**, *3rd year*, Polytech Paris-Sud, Tutorials and classes (24hrs/year).
- 2015–16, 2016–17 **Algorithmics/C**, *3rd year*, Polytech Paris-Sud, Tutorials and classes (27hrs+28hrs).
- 2015–16 **Tutoring in computer science**, *3rd year*, Polytech Paris-Sud (12hrs).
- 2009–2010 **English language assistant**, *4 primary schools in Le Puy-en-Velay, France.*

Supervision

- Sep–present **Quentin Burthier, Master 2 student from ENS Paris-Saclay**, *Machine Translation of Noisy Texts*, Co-supervised with Djamé Seddah.
- Sep–present **Clémentine Fourier, Inria-funded PhD**, *Neural models of language evolution*, Co-supervised with Benoît Sagot.
- Feb–June 2020 **Farid Arthaud, visiting Master 1 student from ENS, Paris**, *Continuous learning for Neural Machine Translation from Human Post-edits*, School of Informatics, Edinburgh, UK, Co-supervised with Alexandra Birch.
- March–Aug 2020 **Ashwani Tanwar, MSC thesis (to be confirmed)**, *Improving Low-Resource Neural Machine Translation of Related Languages by Transfer Learning*, School of Informatics, Edinburgh, UK, Co-supervised with Alexandra Birch.
- March–Aug 2019 **Radina Dobрева, MSC thesis**, *Integrating document structure information into Neural Machine Translation using cache-based models*, School of Informatics, Edinburgh, UK, Co-supervised with Annie Louis and Bonnie Webber.
- March–Aug 2019 **Jie Zhou, MSC thesis**, *Exploiting Predictable Document Substructure in Neural Machine Translation*, School of Informatics, Edinburgh, UK, Co-supervised with Annie Louis and Bonnie Webber.

Examination Committees

Sep 2020 **Examiner**, *Kirian Guiller*, Master 2, Université Sorbonne Nouvelle, Paris, France, Automatic syntactic parsing of the Nigerian pidgin-creole with help from a transformer (BERT): Methods and Results.

Community activities

2020 **Co-president of the 2020 ATALA thesis prize.**

2020–2021 **Member of an editorial board**, *The Northern European Journal of Language Technology (NEJLT)*.

2017–present **Programme committees for conferences and workshops**, LREC'17, ACL-SRW'17, EMNLP'18, NAACL-SRW'18-19, ACL'18*, '19 & '20, NAACL'19 RECITAL'19, WMT'19, MT Summit '19, EAMT'20, WMT'20, EMNLP'20* (*=Recognised as an "outstanding reviewer").

2016 **Jury member of the Doctoral Prize**, *The Doctoral School of Science and Communication Technology (STIC)*, Université Paris-Saclay.

2016 **Co-organiser of the PhD Day**, *LIMSI, CNRS*.

2016–2017 **PhD representative for the laboratory board**, *LIMSI, CNRS*.

Awards

2019 **ATALA PhD prize**, *Association pour le Traitement Automatique des Langues (French association for Natural Language Processing)*.

Additional training

50hrs Lisbon Machine Learning Summer School (LxMLS) 2016

Computer skills

Languages **Python, C++, Bash**, C, Perl, Java, OCaml, Praat, SQL

Web technologies HTML, CSS, XML, XSL

Languages

Native speaker English

Bilingual French

Interests

Piano Grade 8, 2006, Guildhall Exams

Saxophone alto Grade 8, 2005, Associated Board

Publications

PhD thesis

[1] Rachel Bawden. *Going beyond the sentence: Contextual Machine Translation of Dialogue*. PhD thesis, LIMSI, CNRS, Université Paris-Sud, Université Paris-Saclay, Orsay, France, 2018.

Journal articles

- [2] Rachel Bawden, Eric Bilinski, Thomas Lavergne, and Sophie Rosset. [DiaBLa: A Corpus of Bilingual Spontaneous Written Dialogues for Machine Translation](#). *Language Resources and Evaluation*, 2020.
- [3] Rachel Bawden, Chloé Clavel, and Frédéric Landragin. [Towards the generation of dialogue acts in socio-affective ECAs: a corpus-based prosodic analysis](#). *Language Resources and Evaluation*, 4(2):821–838, 2016.

Book chapters

- [4] Sylvain Kahane, Kim Gerdes, and Rachel Bawden. Microsyntactic annotation. In Anne Lacheret-Dujour, Sylvain Kahane, and Paola Pietrandrea, editors, *Rhapsodie – A Prosodic and Syntactic Treebank for Spoken French*, chapter 4. John Benjamins, Amsterdam, 2019.
- [5] Kim Gerdes, Sylvain Kahane, Rachel Bawden, Julie Belião, Eric de la Clergerie, and Ilaine Wang. Annotation tools for syntax. In Anne Lacheret-Dujour, Sylvain Kahane, and Paola Pietrandrea, editors, *Rhapsodie – A Prosodic and Syntactic Treebank for Spoken French*, chapter 7. John Benjamins, Amsterdam, 2019.
- [6] Anne Lacheret-Dujour, Sylvain Kahane, Rachel Bawden, Serge Fleury, and Ilaine Wang. Exploration of the Rhapsodie corpus: Data structure, formats and query tools. In Anne Lacheret-Dujour, Sylvain Kahane, and Paola Pietrandrea, editors, *Rhapsodie – A Prosodic and Syntactic Treebank for Spoken French*, chapter 15. John Benjamins, Amsterdam, 2019.

Conference papers

- [7] Farid Arthaud, Rachel Bawden, and Alexandra Birch. Few-shot learning through contextual data augmentation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, **EACL'21**, Online, 2021.
- [8] Rachel Bawden, Biao Zhang, Lisa Yankovskaya, Andre Tättar, and Matt Post. [A Study in Improving BLEU Reference Coverage with Diverse Automatic Paraphrasing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, **EMNLP'20**, Online, 2020.
- [9] Rachel Bawden, Biao Zhang, Andre Tättar, and Matt Post. [ParBLEU: Augmenting Metrics with Automatic Paraphrases for the WMT'20 Metrics Shared Task](#). In *Proceedings of the 5th Conference on Machine Translation*, WMT'20, Online, 2020.
- [10] Nikita Moghe, Christian Hardmeier, and Rachel Bawden. [The University of Edinburgh-Uppsala University's Submission to the WMT 2020 Chat Translation Task](#). In *Proceedings of the 5th Conference on Machine Translation*, WMT'20, Online, 2020.
- [11] Rachel Bawden, Alexandra Birch, Radina Dobрева, Arturo Oncevay, Antonio Valerio Miceli Barone, and Philip Williams. [The University of Edinburgh's English-Tamil and English-Inuktitut Submissions to the WMT20 News Translation Task](#). In *Proceedings of the 5th Conference on Machine Translation*, WMT'20, Online, 2020.
- [12] Rachel Bawden, Giorgio Maria Di Nunzio, Cristian Grozea, Iñigo Jauregi Unanue, Antonio Jimeno Yepes, Nancy Mah, David Martinez, Aurélie Névéol, Mariana Neves, Maite Oronoz, Olatz Perez de Viñaspre, Massimo Piccardi, Roland Roller, Amy Siu, Philippe Thomas, Federica Vezzani, Maika Vicente Navarro, Dina Wiemann, and Lana Yeganova. Findings of the WMT 2020 Biomedical Translation Shared Task: Basque, Italian and Russian as New Additional Languages. In *Proceedings of the 5th Conference on Machine Translation*, WMT'20, Online, 2020.
- [13] António Lopes, M. Amin Farajian, Rachel Bawden, Michael Zhang, and André T. Martins. [Document-level Neural MT: A Systematic Comparison](#). In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, Lisbon, Portugal, 2020.
- [14] Susie Coleman, Andrew Secker, Rachel Bawden, Barry Haddow, and Alexandra Birch. [Architecture of a Scalable, Secure and Resilient Translation Platform for Multilingual News Media](#). In *Proceedings of the 1st International Workshop on Language Technology Platforms*, pages 16–21, Marseille, France, May 2020. European Language Resources Association.

- [15] Radina Dobrev, Jie Zhou, and Rachel Bawden. [Document Sub-structure in Neural Machine Translation](#). In *Proceedings of the 12th Language Resources and Evaluation Conference, LREC'20*, Marseille, France, 2020.
- [16] Rachel Bawden, Nikolay Bogoychev, Ulrich Germann, Roman Grundkiewicz, Faheem Kirefu, Antonio Valerio Miceli Barone, and Alexandra Birch. [The University of Edinburgh's Submissions to the WMT19 News Translation Task](#). In *Proceedings of the 4th Conference on Machine Translation, WMT'19*, pages 103–115, Florence, Italy, 2019.
- [17] Rachel Bawden, Kevin Bretonnel Cohen, Cristian Grozea, Antonio Jimeno Yepes, Madeleine Kittner, Martin Krallinger, Nancy Mah, Aurelie Neveol, Mariana Neves, Felipe Soares, Amy Siu, Karin Verspoor, and Maika Vicente Navarro. [Findings of the WMT 2019 Biomedical Translation Shared Task: Evaluation for MEDLINE Abstracts and Biomedical Terminologies](#). In *Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2)*, WMT'19, pages 29–53, Florence, Italy, 2019.
- [18] Alexandra Birch, Barry Haddow, Ivan Tito, Antonio Valerio Miceli Barone, Rachel Bawden, Felipe Sánchez-Martínez, Mikel L. Forcada, Miquel Esplà-Gomis, Víctor Sánchez-Cartagena, Juan Antonio Pérez-Ortiz, Wilker Aziz, Andrew Secker, and Peggy van der Kreeft. [Global Under-Resourced Media Translation \(GoURMET\)](#). In *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, pages 122–122, Dublin, Ireland, 2019.
- [19] Rachel Bawden, Rico Sennrich, Alexandra Birch, and Barry Haddow. [Evaluating Discourse Phenomena in Neural Machine Translation](#). In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL'18*, pages 1304–1313, New Orleans, USA, 2018.
- [20] Rachel Bawden, Thomas Lavergne, and Sophie Rosset. [Detecting context-dependent sentences in parallel corpora](#). In *Proceedings of the 25th Conférence sur le Traitement Automatique des Langues Naturelles, TALN'18*, pages 393–400, Rennes, France, 2018.
- [21] Rachel Bawden. [Machine Translation, it's a question of style, innit? The case of English tag questions](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP'17*, pages 2497–2502, Copenhagen, Denmark, 2017.
- [22] Rachel Bawden. [Machine Translation of Speech-Like Texts: Strategies for the Inclusion of Context](#). In *Proceedings of the 19th REcontres jeunes Chercheurs en Informatique pour le TAL, RECITAL'17*, pages 1–14, Orléans, France, 2017.
- [23] Rachel Bawden and Benoit Crabbé. [Boosting for Efficient Model Selection for Syntactic Parsing](#). In *Proceedings of the 26th International Conference on Computational Linguistics, COLING'16*, pages 1–11, Osaka, Japan, 2016.
- [24] Rachel Bawden. [Cross-lingual Pronoun Prediction with Linguistically Informed Features](#). In *Proceedings of the 1st Conference on Machine Translation, WMT'16*, Berlin, Germany, 2016.
- [25] Rachel Bawden, Guillaume Wisniewski, and Hélène Maynard. [Investigating gender adaptation for speech translation](#). In *Proceedings of the 23rd Conférence sur le Traitement Automatique des Langues Naturelles, TALN'16*, Paris, France, 2016.
- [26] Rachel Bawden, Marie-Amélie Botalla, Kim Gerdes, and Sylvain Kahane. [Correcting and Validating Syntactic Dependency in the Spoken French Treebank](#). In *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC'14*, pages 2320–2325, Reykjavik, Iceland, 2014.

Other documents

- [27] Rachel Bawden. Boosting for model selection in syntactic parsing. Master's thesis, Inria (Alpage), Paris, France, 2015. Supervision by Benoît Crabbé.
- [28] Sylvain Kahane, Kim Gerdes, Pietrandrea Paola, Benzitoun Christophe, Rachel Bawden, Marie-Amélie Botalla, and Adèle Désoyer. [Protocole de codage microsyntaxique](#), 2013. Technical document. Translated into English by Rachel Bawden as "Protocol for micro-syntactic coding".
- [29] Lucien Tesnière. *Elements of Structural Syntax*, 2015. Rereading and revision (first part) by Rachel Bawden.

Oral communications and invited talks

- 2020 Traitement Automatique des Langues (Natural Language Processing), *Présentation invitée*, Rendez-vous des Jeunes Mathématiciennes et Informaticiennes, Inria Paris, France.
- 2019 Contextual Machine Translation of Dialogue: Strategies and Evaluation, *Invited talk*, Amazon, Berlin, Germany.
- 2019 Contextual Machine Translation of Dialogue: Strategies and Evaluation, *Invited talk*, Naver labs, Grenoble, France.
- 2019 Going beyond the sentence: Contextual Machine Translation of Dialogue, *ATALA PhD prize talk*, (in French), Toulouse, France.
- 2019 Discourse and Context in Machine Translation, *Invited talk*, MSC Course in Informatics: Natural Language Understanding, Generation, and Machine Translation, Edinburgh, UK.
- 2018 Mediating Bilingual Human Dialogues with Machine Translation, *SICSA Workshop on Conversational AI*, Heriot Watt University, Edinburgh, UK.