Rachel Bawden

May-Aug 2017

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, <i>CNU</i> , Sections 7 (Linguistics) and 27 (Informatics).		
	PhD in computer science , <i>Université Paris-Sud</i> , Going beyond the sentence: Contextual Machine Translation of Dialogue.		
29/11/2018			
	- 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 , <i>Oxford University</i> , First Class. Phonetics/phonology, psycholinguistics, general linguistics, syntax, historical linguistics, sociolinguistics, French linguistics.		
2005–2007	A-levels , <i>The Latymer School, Edit</i> Geography, four grade "A"s.	monton, Lond	don, French, English, Biology,
	Professional experience		
	Research		
2020-present	Researcher (chargée de recherches Natural Language Processing and Mad	chine Translat	tion
	Holder of a "springboard" chair in the		
2018–2020	Research associate , <i>ILCC</i> , <i>University of Edinburgh</i> , Neural machine translation for low resourced languages, Working with Alexandra Birch.		
2015–2018	PhD in computer science , <i>LIMSI</i> , sentence: Contextual Machine Transla and Thomas Lavergne.		9 9

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.
 - 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–Aug 2013 Master's Research placement, MoDyCo Laboratory (CNRS & Université Paris-Ouest), Annotation, development and semi-automatic correction of a dependency treebank of spoken French (Rhapsodie). Supervised by Sylvain Kahane.

Research projects

- 2021–2023 **DadaNMT: Domain ADAptation for Neural Machine Translation**, *Projet Emergence*, Sorbonne Université, **Project PI**.
- 2021–2024 **PRAIRIE** (PaRis Al Research InstitutE), Research Institute on Artificial intelligence headed by Isabelle Ryl, Holder of a "springboard" chair, https://prairie-institute.fr.
- 2021–2024 **CounteR**, European H2020 project, https://prairie-institute.fr.
- 2019–2020 MTStretch: Low Resource Machine Translation, UKRI fellowship, held by Alexandra Birch, University of Edinburgh, UK, Funded postdoc position. https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/S001271/1
- 2019–2020 **GoURMET (Global Under-Resourced Media Translation)**, European H2020 project, 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.

Supervision

PhD students

- Oct 2021-present Lydia Nishimwe, PR[AI]RIE-funded PhD, Robust Neural Machine Translation, Co-supervised with Benoît Sagot.
- Oct 2021-present Matthieu Futeral-Peter, PhD funded by Inria and PR[AI]RIE, Multimodal Neural Machine Translation, Co-supervised with Ivan Laptev, Cordelia Schmid and Benoît Sagot.
 - Sep 2020–Sep Clémentine Fourrier, Inria-funded PhD, Neural models of language evolution, Co-2022 supervised with Benoît Sagot.

Interns and engineers

- Feb-June 2022 **Jesujoba Alabi, Engineer**, *Domain adaptation for machine translation(DadaNMT project)*.
- Sep 2021–June Camille Rey, Master 2 student from INALCO, Contrastive training of machine translation models for lexical disambiguation, Co-supervised with Benoît Sagot.
- June-Aug 2021 **Sonal Sannigrahi, BSC student from École Polytechnique**, *Investigating the effect of input representations on language sharing in multilingual models.*
- May-Oct 2021 Matthieu Futeral-Peter, Master 2 student from ENSAE and ENS Paris-Saclay, Exploration of multilingual and multimodal word embeddings, Co-supervised with Benoît Sagot, Cordelia Schmid and Ivan Laptev.

- Feb 2021–Jan **Thibault Charmet, Engineer**, Automatic tools for improving jurisprudence consistency, Co-supervised with Benoît Sagot and in collaboration with the Cour de Cassation.
- Sep 2020–March Quentin Burthier, Master 2 student from ENS Paris-Saclay, Machine Translation of Noisy Texts, Co-supervised with Djamé Seddah.
 - 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**, *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 Dobreva, 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 Commitees

- July 2022 **Examiner**, *Baptiste Rozière*, *PhD*, Université Paris Dauphine-PSL, France, Traduction Non Supervisée de Langages de Programmation.
- July 2022 **Examiner**, Antoine Simoulin, PhD, Université Paris Cité, France, Sentence embeddings and their relation with sentence structures.
- Sep 2020 **Examiner**, *Kirian Guiller*, *Master 2*, Université Sorbonne Nouvelle, Paris, France, Automatic syntactic parsing of the Nigerian pidgin-creole with help form 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, ACL-SRW EMNLP, NAACL-SRW, 21, ACL, NAACL' RECITAL, WMT, MT Summit, EAMT, EACL demos, BSL.
 - 2016 **Jury member of the Doctoral Prize**, The Doctoral School of Science and Communcation Technology (STIC), Université Paris-Saclay.
 - 2016 Co-organiser of the PhD Day, LIMSI, CNRS.
 - 2016–2017 PhD representative for the laboratory board, LIMSI, CNRS.

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.

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] Simon Gabay, Rachel Bawden, Philippe Gambette, Jonathan Poinhos, Eleni Kogkitsidou, and Benoît Sagot. Le changement linguistique au xviie s.: nouvelles approches scriptométriques. SHS Web of Conferences, 138, 2022.
- [4] 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

- [5] 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.
- [6] 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.
- [7] 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

- [8] Rachel Bawden, Jonathan Poinhos, Eleni Kogkitsidou, Philippe Gambette, Benoît Sagot, and Simon Gabay. Automatic Normalisation of Early Modern French. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages 3354–3366, Marseille, France, 2022.
- [9] Simon Gabay, Pedro Ortiz Suarez, Alexandre Bartz, Alix Chagué, Rachel Bawden, Philippe Gambette, and Benoît Sagot. From FreEM to D'AlemBERT: a Large Corpus and a Language Model for Early Modern French. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages 3367–3374, Marseille, France, 2022.
- [10] Simon Gabay, Pedro Ortiz Suarez, Rachel Bawden, Alexandre Bartz, Philippe Gambette, and Benoît Sagot. Le projet FREEM: ressources, outils et enjeux pour l'étude du français d'ancien régime (the F RE EM project: Resources, tools and challenges for the study of ancien régime French). In Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1: conférence principale, pages 154–165, Avignon, France, 2022. ATALA.
- [11] Thibault Charmet, Inès Cherichi, Matthieu Allain, Urszula Czerwinska, Amaury Fouret, Benoît Sagot, and Rachel Bawden. Complex Labelling and Similarity Prediction in Legal Texts: Automatic Analysis of France's Court of Cassation Rulings. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages 4754–4766, Marseille, France, 2022.
- [12] Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M. Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal V. Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Févry, Jason Alan Fries, Ryan Teehan, Stella Biderman, Leo Gao, Tali Bers, Thomas Wolf, and Alexander M. Rush. Multitask prompted training enables zero-shot task generalization. In Proceedings of the 10th International Conference on Learning Representations, Online, 2022.
- [13] Alexandre Bartz, Juliette Janes, Laurent Romary, Philippe Gambette, Rachel Bawden, Pedro Ortiz Suarez, Benoît Sagot, and Simon Gabay. Expanding the content model of annotationBlock. In *Next Gen TEI*, 2021 TEI Conference and Members' Meeting, Virtual, United States, 2021.
- [14] Lana Yeganova, Dina Wiemann, Mariana Neves, Federica Vezzani, Amy Siu, Inigo Jauregi Unanue, Maite Oronoz, Nancy Mah, Aurélie Névéol, David Martinez, Rachel Bawden, Giorgio Maria Di Nunzio, Roland Roller, Philippe Thomas, Cristian Grozea, Olatz Perez-de Viñaspre, Maika Vicente Navarro, and Antonio Jimeno Yepes. Findings of the WMT 2021 biomedical translation shared task: Summaries of animal experiments as new test set. In Proceedings of the Sixth Conference on Machine Translation, pages 664–683, Online, 2021.
- [15] Clémentine Fourrier, Rachel Bawden, and Benoît Sagot. Can Cognate Prediction Be Modelled as a Low-Resource Machine Translation Task? In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, ACL Findings'21, pages 847–861, Online, 2021.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.

- [20] Rachel Bawden, Alexandra Birch, Radina Dobreva, 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.
- [21] 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.
- [22] 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.
- [23] 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, 2020. European Language Resources Association.
- [24] Radina Dobreva, 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] Rachel Bawden. Machine Translation of Speech-Like Texts: Strategies for the Inclusion of Context. In *Proceedings of the 19th REncontres jeunes Chercheurs en Informatique pour le TAL*, RECITAL'17, pages 1–14, Orléans, France, 2017.
- [32] 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.

- [33] Rachel Bawden. Cross-lingual Pronoun Prediction with Linguistically Informed Features. In *Proceedings of the 1st Conference on Machine Translation*, WMT'16, Berlin, Germany, 2016.
- [34] 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.
- [35] 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

- [36] Rachel Bawden. Boosting for model selection in syntactic parsing. Master's thesis, Inria (Alpage), Paris, France, 2015. Supervision by Benoît Crabbé.
- [37] 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".
- [38] Lucien Tesnière. *Elements of Structural Syntax*, 2015. Rereading and revision (first part) by Rachel Bawden.
 - Oral communications and invited talks
- 2021 Natural Language Processing: Applications and Neural Modelling, *Speaker*, Mathematical Summer in Paris, Online.
- 2020 Traitement Automatique des Langues (Natural Language Processing), *Invited Talk*, Rendezvous 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.