

Frédéric Blain

Research Associate in Machine Translation
Department of Computer Science, University of Sheffield (UK)

Personal Details

Date of birth – 13 September 1986
Website – fredblain.org
LinkedIn – fredblain
Twitter – @fblain

Nationality – French
GPG key – FC7C3BC0

Contact : f.blain [at] sheffield [dot] ac.uk

Education & Qualifications

2009 - 2013	Ph.D. in Computer Science Computer science Laboratory of the University of Le Mans (LIUM) Le Mans University (France) Thesis « <i>Evolutionary translation models</i> » Supervisors Prof. Holger Schwenk – Facebook, LIUM Dr. Jean Senellart – SYSTRAN International
2007 - 2009	Master of Research Degree in Computer Science Le Mans University (France) Domain Human-computer communication and educational engineering Dissertation « <i>Creation of bilingual corpus from comparable resources</i> » Supervisors Prof. Holger Schwenk – Facebook, LIUM Dr. Jean Senellart – SYSTRAN International
2004 - 2007	Bachelor's Degree in Computer Science Le Mans University (France)

Current & Previous Appointments

2015 –	Research Associate in Machine Translation at the University of Sheffield Department of Computer Science – NLP Research Lab Managed by Pr Lucia Specia
2013 – 2014	Postdoc Researcher in Machine Translation at the University of Le Mans Department of Computer Science – Language and Speech Technology group Managed by Pr Holger Schwenk
2009 – 2012	Research engineer SYSTRAN SA – Paris (France)

Professional and External Standing

RESEARCH INTERESTS

Natural Language Processing with focus on Machine Translation to enrich end users' experience by the way of Quality Estimation, Evaluation, Continuous Adaptation and work on Post-Editing.

RESEARCH GRANTS

Date	Sponsor	Title	Funding as PI
2017 - 2018	Amazon	Predicting Relevance and Quality of Machine Translation for Product Reviews	\$ 68,000
2017 - 2018	EAMT	Predicting Relevance and Quality of News Translation	€ 9,800

RESEARCH PROJECTS

Quality Translation 21 (QT21) – (April 2015 - January 2018)

QT21 is a machine translation project which has received funding from the European Union's Horizon 2020 Research and Innovation programm. It tackles solutions for morphologically riched and under-resourced european languages suffering from poor translation quality.

MateCAT – (October 2011 - October 2014)

European project led by the Bruno Kessler Foundation (FBK), and conducted with the Computer Science laboratory of Le Mans University (LIUM), The University of Edinburgh and Translated Srl. For professional translators, it aimed at reducing the post-editing cost through the use of an optimized web-based CAT tool. To improve the user's productivity, the project partners have worked on in-domain adaptation, project adaptation, automatic quality estimation and both on-line and incremental adaptation from user feedback. MateCAT nowadays is used by thousands of professional translators to deliver translations in more than 100 languages to 10,000 active users all over the world.

COSMAT – (October 2009 - October 2012)

Led by the LIUM, working with SYSTRAN and the INRIA, the project aimed at providing a collaborative translation service of scientific documents to the scientific community. The result of this project was planned to be hosted on the HAL, an open archive where authors can deposit scholarly documents from all academic fields. Independently of the characteristics bound to scientific documents (domain adaptation, entities recognition, etc.), the collaborative aspect of this project relied on both translated and reviewed versions of the scientific documents (PhD thesis, article, etc.) which are used to improve the quality of the machine translation system through an analysis based on post-editing.

PROGRAM COMMITTEES & REVIEWING

Program Committee Member for Conferences : ACL long/short (2016, 17, 18); COLING (2016, 18); CoNLL (2017); EACL long/short (2017); EAMT long (2017); EMNLP (2015, 17, 18); IJCNLP long/short (2017); LREC (2016, 18); MT-Summit (2017); NAACL (2015, 18)

Program Committee Member for Workshops : WMT (2015, 16, 17, 18); IAMT (2014); SemEval (2016); QEAPE (2018); NLPOSS (2018)

PUBLICATIONS

Refereed Conference Papers – in Print/Press

IVE, J., **BLAIN, F.**, SPECIA, L. « deepQuest : A Framework for Neural-based Quality Estimation ». *Proceedings of COLING 2018, the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, August 20-26, 2018.

CHATTERJEE, R., NEGRI, M., TURCHI, M., **BLAIN, F.**, SPECIA, L. « Combining Quality Estimation and Automatic Post-editing to Enhance Machine Translation output ». *Proceedings of the 13th Biennial Conference of the Association for Machine Translation in the America*, Boston, USA, March 2018.

BLAIN, F., SPECIA, L., MADHYASTHA, P. « Exploring Hypotheses Spaces in Neural Machine Translation ». *Proceedings of the Machine Translation Summit XVI*, Nagoya, Japan, September 2017.

SPECIA, L., HARRIS, K., **BLAIN, F.**, BURCHARDT, A., MACKETANZ, V., SKADIPA, I., NEGRI, M., TURCHI, M. « Translation Quality and Productivity : A Study on Rich Morphology Languages ». *Proceedings of the Machine Translation Summit XVI*, Nagoya, Japan, September 2017.

BLAIN, F., SCARTON, C., SPECIA, L. « Bilingual Embeddings for Quality Estimation ». *Proceedings of the Second Conference on Machine Translation, Volume 2 : Shared Task Papers*, Copenhagen, Denmark, September 2017.

CHATTERJEE, R., NEGRI, M., TURCHI, M., FEDERICO, M., SPECIA, L., **BLAIN, F.** « ». *Proceedings of the Second Conference on Machine Translation, Volume 1 : Research Papers*, Copenhagen, Denmark, September 2017.

PETER, J-T., NEY, H., BOJAR, O., PHAM, N-Q., NIEHUES, J., WAIBEL, A., BURLLOT, F., YVON, F., PINNIS, M., SICS, V., BASTINGS, J., RIOS, M., AZIZ, W., WILLIAMS, P., **BLAIN, F.**, SPECIA, L. « The QT21 Combined Machine Translation System for English to Latvian ». *Proceedings of the Second Conference on Machine Translation, Volume 2 : Shared Task Papers*, Copenhagen, Denmark, September 2017.

LOGACHEVA, V., **BLAIN, F.**, SPECIA, L. « USFD's Phrase-level Quality Estimation Systems ». *Proceedings of the First Conference on Machine Translation (WMT)*, Berlin, Germany, August 2016.

BLAIN, F., SONG, X., SPECIA, L. « Sheffield Systems for the English-Romanian WMT Translation Task ». *Proceedings of the First Conference on Machine Translation (WMT)*, Berlin, Germany, August 2016.

PETER, J-T., ALKHOULI, T., NEY, H., HUCK, M., BRAUNE, F., FRASER, A., TAMCHYNA, A., BOJAR, O., HADDOW, B., SENNRICH, R., **BLAIN, F.**, SPECIA, ET AL. « The QT21/HimL Combined Machine Translation System ». *Proceedings of the First Conference on Machine Translation (WMT)*, Berlin, Germany, August 2016.

AKER, A., **BLAIN, F.**, DUQUE, A., FOMICHEVA, M., SEVA, J., SHAH, K. « USFD at SemEval-2016 Task 1 : Putting different State-of-the-Arts into a Box ». *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, California, June 2016.

BLAIN, F., LOGACHEVA, V., SPECIA, L. « Phrase-Level Segmentation and Labelling of Machine Translation Errors ». *Proceedings of the 10th Edition of Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia, May 2016.

SHAH K., LOGACHEVA, V., PAETZOLD, G., **BLAIN, F.**, BECK, D., BOUGARES, F., SPECIA, L. « SHEF-NN : Translation Quality Estimation with Neural Networks ». *Proceedings of the Tenth Workshop on Statistical Machine Translation (WMT)*, Lisbon, Portugal, September 2015.

BLAIN, F., BOUGARES, F., HAZEM, A., BARRAULT, L., SCHWENK, H. « Continuous Adaptation to User Feedback for Statistical Machine Translation ». *North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015)*, Denver, Colorado, USA, June 2015.

BLAIN, F., HAZEM, A., BOUGARES, F., BARRAULT, L., SCHWENK, H. « Project adaptation over several days ». *Translation in Transition 2015*, Gernersheim, Germany, January 2015.

FEDERICO, M., BERTOLDI, M., CETTOLO, M., NEGRI, M., TURCHI, M., TROMBETTI, M., CATTELAN, A., FARINA, A., LUPINETTI, D., MARTINES, A., MASSIDDA, A., SCHWENK, H., BARRAULT, L., **BLAIN, F.**, KOEHN, P., BUCK, C., GERMANN, U. « The MateCat tool ». *Proceedings of the 25th International Conference on Computational Linguistics (COLING'14)*, Dublin, Ireland, August 2014.

BLAIN, F. « Projet COSMAT : traduction automatique de contenus scientifiques pour l'anglais et le français ». *Workshop "Les Rencontres du Numérique de l'ANR" by the French National Research Agency*, Paris, France, April 2013.

BLAIN, F., SCHWENK, H., SENELLART, J. « Incremental Adaptation Using Translation Information and Post-Editing Analysis ». *Proceedings of the Eighth International Conference on Language Resources and Evaluation (IWSLT'12)*, Hong-Kong, Chine, December 2012.

LAMBERT, P., SCHWENK, H., **BLAIN, F.** « Automatic Translation of Scientific Documents in the HAL Archive ». *Proceedings of the Eight International Conference on Language Resources and Evaluation*, Istanbul, Turquie, May 2012.

LAMBERT, P., SENELLART, J., ROMARY, L., SCHWENK, H., ZIPSER F., LOPEZ, P., **BLAIN, F.** « Collaborative Machine Translation Service for Scientific texts ». *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, France, April 2012.

BLAIN, F., SENELLART, J., SCHWENK, H., PLITT, M., ROTURIER, J. « Qualitative Analysis of Post-Editing for High Quality Machine Translation ». *Proceedings of the 13th Machine Translation Summit 2011*, Xiamen, Chine, September 2011.

Non-Refereed Abstracts, Reports & other Publications – in Print

BOJAR, O., VARIŠ, D., LIU, Q., NEGRI, M., TURCHI, M., NIEHUES, J., SPECIA, L., **BLAIN, F.** « Human-informed Continuous Learning ». *QT21 deliverable*, 2018.

BURCHARDT, A., BOJAR, O., GRAHAM, Y., HARRIS, K., LIU, Q., MA, Q., MARHEINECKE, K., SPECIA, L., **BLAIN, F.**, SKADINA, I., PINNIS, M., TURCHI, M., MACKETANZ, V., PETER, J-T., VARIŠ, D., WILLIAMS, P. « Quality Estimation Metrics and Analysis of 2 nd Annot. Round and Error Profiles ». *QT21 deliverable*, 2018.

NIEHUES, J., HA, T-L., BURLLOT, F., YVON, F., PETER, J-T., **BLAIN, F.**, BOJAR, O., SKADINA, I., DAIBER, J., SIMA'AN, K., VALERIO MICELI BARONE, A., SENNRICH, R., WILLIAMS, P., KIM, Y., SCHAMPER, J., ALKHOULI, T., ESPAÑA-BONET, C. « Final Report on Under-Resourced Languages ». *QT21 deliverable*, 2018.

BURCHARDT, A., **BLAIN, F.**, BOJAR, O., DEHDARI, J., GRAHAM, Y., GÖRÖG, A., HEIGOLD, G., LIU, Q., MA, Q., SPECIA, L., SKADINA, I., PINNIS, M., TURCHI, M., MACKETANZ, V., PETER, J-T., WILLIAMS, P. « Evaluation Metrics and Analysis of First Annotation Round ». *QT21 deliverable*, 2017.

NIEHUES, J., BURLLOT, F., PETER, J-T., **BLAIN, F.**, BOJAR, O., SKADINA, I., SIMA'AN, K., WILLIAMS, P. « Intermediate Report on Under-resourced languages ». *QT21 deliverable*, 2017.

BLAIN, F., BURCHARDT, A., BOJAR, O., DUGAST, C., GRAHAM, Y., HARRIS, K., HUCK, M., LOMMEL, A., NEGRI, M., NIEHUES, J., SPECIA, L., THORSTEN, J-P., TURCHI, M., YVON, F., BRANDON, L., CORNELIUS, E., LIU, H., MELBY, A. « Periodic Report M1-M18 ». *QT21 deliverable*, 2016.

BERTOLDI, **BLAIN, F.** « Second Report on User-adaptive MT ». *MateCAT deliverable*, 2014.

BERTOLDI, N., TURCHI, M., GERMANN, U., **BLAIN, F.**, SCHWENK, H., CATTELAN, A. « Open Source Distribution ». *Third report on field and lab tests*, 2014.

BERTOLDI, N., TURCHI, M., GERMANN, U., **BLAIN, F.**, SCHWENK, H., ROUSSEAU, A., CATTELAN, A. « Open Source Distribution ». *MateCAT deliverable*, 2014.

Non-Refereed Abstracts, Reports & other Publications – in Press

BLAIN, F. « COSMAT : traduction automatique de contenus scientifiques pour l'anglais et le français ». *Workshop "Les Rencontres du Numérique de l'ANR" by the French National Research Agency*, Paris, France, 2012.

BLAIN, F. « Learn from Post-Editing for High Quality Machine Translation ». *Forum "Jeune Recherche" by the Ph.D. school (finalist of the poster competition)*, Le Mans, France, November 2011.

BLAIN, F. « Post-Editing Analysis for High Quality Machine Translation ». *Young researchers away day*, Nantes, France, April 2011.

BLAIN, F. « Modèle de traduction évolutif ». « *R&T PME* » *workshop by the Delegation General pour l'Armement (French Defense Agency)*, Issy-Les-Moulineaux, France, April 2011.

Contribution to open source projects

QUEST++ – An open source toolkit for pipelined Translation Quality Estimation

MOSES – An open source Statistical Machine Translation system

MateCAT – An open source Computer Assisted Translation tool

Technical skills

Programming	Python, Perl, Shell scripting (sh/bash/csh), C/C++
Systems	GNU/Linux, Unix
Versioning	Git, SVN, CVS
Languages	English, native French

References

On request.

Sheffield, July 2018.