Alberto Poncelas Rodriguez

aponcelas2@hotmail.com
https://alberto-poncelas.github.io
https://www.linkedin.com/in/albertoponcelas/

Work Experience

May 2021 – Present
- **Research Scientist** in Rakuten, Singapore.
  Research in the area of Machine Translation.

Oct 2019 – Apr 2021
- **Postdoctoral Researcher** in ADAPT Centre, Ireland.
  Research in the area of Machine Translation and Natural Language Processing. Investigation of indirect translation and sentiment classification. Technologies: Python, Pandas, Numpy, Spacy, Keras.

Oct 2015 – Sep 2019
- **PhD Researcher** in ADAPT Centre, Ireland.
  Research in the area of Machine Translation and Natural Language Processing. Use of data-selection techniques and synthetic data generation for improving machine-translation. Technologies: OpenNMT, Python, Scala, C language.

Aug 2013 – Sep 2015
- **Big Data Engineer** in BEEVA, Spain
  Projects involved: (i) Development of a tool for calculating the risks involved in the granting of loans;(ii) Data exploration using Spark and Scala; (iii) A client segmentation analysis using data mining techniques. Technologies: R and Hive; and (iv) Development of a tool based in intelligent agent ecosystem. Technologies: node.js and Neo4j.

Jun 2012 – Jul 2012
- **Software Developer** in LKS, Spain
  Use of Oracle PL/SQL, Oracle Forms and Oracle Reports technologies

May 2011 – Aug 2011
- **Web Developer** in Sector04 Consultoría Y Soluciones Informáticas S.L.
  Web application development using the following technologies: PHP (CodeIgniter framework), HTML, jQuery, AJAX, CSS, MySQL

Education

2015 – 2019
- **Ph.D., Machine Translation** in Dublin City University, Ireland.
  Research in the area of Machine Translation.
  Thesis title: *Improving Transductive Data Selection Algorithms for Machine Translation*

2012 – 2013
- **M.Sc., Machine Learning** in University of the Basque Country, Spain.
  Thesis title: *Preprocess and data analysis techniques for affymetrix DNA microarrays using bioconductor: a case study in Alzheimer disease*

2009 – 2010
- **Erasmus study year** in École pour l'informatique et les techniques avancées (EPITA), France

2004 – 2011
- **B.Sc. Computer Science** in University of the Basque Country, Spain.

Certificates and Courses

2016 • **Machine learning** Coursera
https://www.coursera.org/account/accomplishments/certificate/5XKFZ85CN8U8

2014 • **Data Science Specialization** Coursera
https://www.coursera.org/account/accomplishments/specialization/certificate/3QEOAonkEe

Teaching Experience

2021 • **Teaching** in Dublin City University.
Machine Learning

2020 • **Teaching** in Dublin City University.
Machine Learning
Statistical Machine Translation

• **Teaching** in Trinity College Dublin.
Interlingual Technologies: Machine Translation lecture.

2019 • **Teaching Assistant** in Dublin City University.
Statistical Machine Translation: teaching, coding/evaluation of lab exercises and grading.

2006 • **Teaching Assistant** in University of the Basque Country.
Data Structures and Algorithms: teaching, coding/evaluation of lab exercises and grading.

2005–2008 • **Teaching Assistant** in University of the Basque Country.
Programming I: teaching, coding/evaluation of lab exercises and grading.

Research Publications

1 Popović, M., **Poncelas**, **A.**, Bakarić, M. B., & Way, A. (2021). On Machine Translation of User Reviews. In *Proceedings of recent advances in natural language processing (RANLP)* (pp. 1113–1122).

2 **Poncelas**, **A.**, Aboomar, M., Buts, J., Hadley, J., & Way, A. (2020). A Tool for Facilitating OCR Postediting in Historical Documents. In *Workshop on Language Technologies for Historical and Ancient Languages, LT4HALA* (pp. 47–51). Marseille, France.

3 **Poncelas**, **A.**, Buts, J., Hadley, J., & Way, A. (2020). Using Multiple Subwords for Improving English-Esperanto Automated Literary Translation Quality. In *Proceedings of the 3rd workshop on technologies for mt of low resource languages* (pp. 108–117). Suzhou, China.

4 **Poncelas**, **A.**, de Buy Wenniger, G. M., & Way, A. (2020). Improved feature decay algorithms for statistical machine translation. *Natural Language Engineering*, 1–21.

5 **Poncelas**, **A.**, Lohar, P., Way, A., & Hadley, J. (2020). The Impact of Indirect Machine Translation on Sentiment Classification. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas, AMTA* (pp. 78–88). Orlando, Florida.

6 **Poncelas**, **A.**, Pidchamook, W., Liu, C.-H., Hadley, J., & Way, A. (2020). Multiple Segmentations of Thai Sentences for Neural Machine Translation. In *Proceedings of The 1st Joint Spoken Language Technologies for Under-resourced languages and Collaboration and Computing for Under-Resourced Languages Workshop, SLTU-CCURL* (pp. 240–244). Marseille, France.

7 Popović, M. & **Poncelas**, **A.** (2020a). Extracting correctly aligned segments from unclean parallel data using character n-gram matching. In *Konferenca Jezikovne tehnologije in digitalna humanistika* (pp. 74–80). Ljubljana, Slovenia.

8 Popović, M. & **Poncelas**, **A.** (2020b). Neural Machine Translation between similar South-Slavic languages. In *Proceedings of the Fifth Conference on Machine Translation (WMT 2020)* (pp. 430–436).

9 Popović, M., **Poncelas**, **A.**, Way, A., & Bakarić, M. B. (2020). Neural Machine Translation for translating into Croatian and Serbian. In *Proceedings of the Seventh Workshop on NLP for Similar Languages, Varieties and Dialects* (pp. 102–113). Barcelona, Spain.

10 Soto, X., Shterionov, D., **Poncelas**, **A.**, & Way, A. (2020). Selecting Backtranslated Data from Multiple Sources for Improved Neural Machine Translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL* (pp. 3898–3908). Seattle, USA.

11 Way, A., Haque, R., Xie, G., Gaspari, F., Popović, M., & **Poncelas**, **A.** (2020a). Facilitating access to multilingual covid-19 information via neural machine translation. *arXiv preprint arXiv:2005.00283*.

12 Way, A., Haque, R., Xie, G., Gaspari, F., Popović, M., & **Poncelas**, **A.** (2020b). Rapid Development of Competitive Translation Engines for Access to Multilingual COVID-19 Information. *7*(2), 1–21.

13 **Poncelas**, **A.** (2019). *Improving transductive data selection algorithms for machine translation* (Doctoral dissertation, Dublin City University).

14 **Poncelas**, **A.**, de Buy Wenniger, G. M., & Way, A. (2019a). Adaptation of machine translation models with back-translated data using transductive data selection methods. In *20th international conference on computational linguistics and intelligent text processing, cicling*. La Rochelle, France.

15 **Poncelas**, **A.**, de Buy Wenniger, G. M., & Way, A. (2019b). Transductive data-selection algorithms for fine-tuning neural machine translation. In *Proceedings of the 8th workshop on patent and scientific literature translation* (pp. 13–23). Dublin, Ireland.

16 **Poncelas**, **A.**, Popović, M., Shterionov, D., de Buy Wenniger, G. M., & Way, A. (2019). Combining SMT and NMT back-translated data for efficient NMT. In *Proceedings of recent advances in natural language processing (RANLP)* (pp. 922–931). Varna, Bulgaria.

17 **Poncelas**, **A.**, Sarasola, K., Dowling, M., Way, A., Labaka, G., & Alegria, I. (2019). Adapting NMT to caption translation in Wikimedia Commons for low-resource languages. *Procesamiento del Lenguaje Natural*, *63*, 33–40.

18 **Poncelas**, **A.** & Way, A. (2019). Selecting Artificially-Generated Sentences for Fine-Tuning Neural Machine Translation. In *Proceedings of the 12th international conference on natural language generation* (pp. 219–228). Tokyo, Japan.

19 Vanmassenhove, E., Moryossef, A., **Poncelas**, **A.**, Way, A., & Shterionov, D. (2019). ABI Neural Ensemble Model for Gender Prediction Adapt Bar-Ilan Submission for the CLIN29 Shared Task on Gender Prediction. In *Computational linguistics of the netherlands clin29*. Groningen, The Netherlands.

20 Dowling, M., Lynn, T., **Poncelas**, **A.**, & Way, A. (2018). SMT versus NMT: Preliminary comparisons for Irish. In *Technologies for mt of low resource languages (loresmt 2018)* (pp. 12–20). Boston, USA.

21 **Poncelas, A.**, de Buy Wenniger, G. M., & Way, A. (2018). Data selection with feature decay algorithms using an approximated target side. In *15th international workshop on spoken language translation (iwslt 2018)* (pp. 173–180). Bruges, Belgium.

22 **Poncelas, A.**, Maillette de Buy Wenniger, G., & Way, A. (2018). Feature decay algorithms for neural machine translation. In *Proceedings of the 21st annual conference of the european association for machine translation* (pp. 239–248). Alicante, Spain.

23 **Poncelas, A.**, Shterionov, D., Way, A., de Buy Wenniger, G. M., & Passban, P. (2018). Investigating backtranslation in neural machine translation. In *21st annual conference of the european association for machine translation* (pp. 249–258). Alicante, Spain.

24 **Poncelas, A.**, Way, A., & Sarasola, K. (2018). The ADAPT System Description for the IWSLT 2018 Basque to English Translation Task. In *International workshop on spoken language translation* (pp. 72–82). Bruges, Belgium.

25 Silva, C. C., Liu, C.-H., **Poncelas, A.**, & Way, A. (2018). Extracting in-domain training corpora for neural machine translation using data selection methods. In *Proceedings of the third conference on machine translation: research papers* (pp. 224–231). Brussels, Belgium.

26 Dzendzik, D., **Poncelas, A.**, Vogel, C., & Liu, Q. (2017). ADAPT Centre Cone Team at IJCNLP-2017 Task 5: A Similarity-Based Logistic Regression Approach to Multi-choice Question Answering in an Examinations Shared Task. In *Proceedings of the ijcnlp 2017, shared tasks* (pp. 67–72). Taipei, Taiwan.

27 Liu, C.-H., Groves, D., Hayakawa, A., **Poncelas, A.**, & Liu, Q. (2017). Understanding Meanings in Multilingual Customer Feedback. In *Proceedings of First Workshop on Social Media and User Generated Content Machine Translation (Social MT 2017)*. Prague, Czech Republic.

28 Liu, C.-H., Moriya, Y., **Poncelas, A.**, & Groves, D. (2017). IJCNLP-2017 Task 4: Customer Feedback Analysis. In *Proceedings of the IJCNLP 2017, Shared Tasks* (pp. 26–33). Taipei, Taiwan.

29 **Poncelas, A.**, Maillette de Buy Wenniger, G., & Way, A. (2017). Applying n-gram alignment entropy to improve feature decay algorithms. *The Prague Bulletin of Mathematical Linguistics*, *108*(1), 245–256.

30 **Poncelas, A.**, Way, A., & Toral, A. (2016). Extending feature decay algorithms using alignment entropy. In *International workshop on future and emerging trends in language technology, fetlt* (pp. 170–182). Springer. Seville, Spain.

31 **Poncelas, A.** (2013). *Preprocess and data analysis techniques for affymetrix dna microarrays using bioconductor: a case study in alzheimer disease* (Master's thesis, University of the Basque Country).

Presentations

LoResMT 2020 • **Using Multiple Subwords for Improving English-Esperanto Automated Literary Translation Quality** Online

AMTA 2020 • **The Impact of Indirect Machine Translation on Sentiment Classification** Online

Mots/Machines 2020 • **The sentiment classification on indirect translation** (in French) Brest, France

PSLT 2019 • **Transductive Data-Selection Algorithms for Fine-Tuning Neural Machine Translation** Dublin, Ireland

Presentations (continued)

RANLP 2019 • **Combining SMT and NMT Back-Translated Data for Efficient NMT** Varna, Bulgaria

CICLING 2019 • **Adaptation of Machine Translation Models with Back-translated Data using Transductive Data Selection Methods** La Rochelle, France

SocialMT 2017 • **Understanding Meanings in Multilingual Customer Feedback** Prague, Czech Republic

IJCNLP 2017 • **IJCNLP-2017 task 4: Customer feedback analysis Workshop presentation** Taipei, Taiwan

• **A Similarity-Based Logistic Regression Approach to Multi-choice Question Answering in an Examinations Shared Task** Taipei, Taiwan

FETLT 2016 • **Extending feature decay algorithms using alignment entropy** Sevilla, Spain

Miscellaneous

Professional services

2021 • **European Chapter of the Association for Computational Linguistics (EACL)** Reviewer

• **LoResMT 2021 (MTSummit 2021)** Reviewer

2020 • **LoResMT (AACL 2020)** Reviewer

• **Post-Editing in Modern-Day Translation (PEMDT1)** Reviewer

• **Transactions on Asian and Low-Resource Language Information Processing (TALLIP)** Reviewer

• **Machine Translation journal** Reviewer

• **ACL 2020** Reviewer

2019 • **MTSummit 2019** PC member / Proceedings chair

• **LoResMT (MTSummit 2019)** PC member / Reviewer

• **RANLP 2019** PC member / Reviewer

• **Language Resources and Evaluation (LREV)** Reviewer

2018 • **LoResMT (AMTA 2018)** Reviewer

2017 • **Cupral Workshop (IJCNLP 2017)** Reviewer

• **Customer Feedback Analysis Shared Task (IJCNLP 2017)** Co-organizer / Reviewer

2015 • **DL4MT Winter School: Deep Learning for Machine Translation** volunteer

Grants and Awards

2018 • President's Awards for Engagement (nominated).

2017 • School of computing travel grant

• "Multi-choice Question Answering in Examinations" (IJCNLP 2017) Shared task winner

Miscellaneous (continued)

2015 – 2019 • Science Foundation Ireland (SFI). PhD funded by SFI through the ADAPT Centre.

2009 – 2010 • Erasmus Scholarship.

Scientific outreach

2010 • **Rapid Development of Competitive Translation Engines for Access to Multilingual COVID-19 Information**

2017 – 2018 • **My Thesis in 3 in French**

2016 – 2019 • **DCU Research Day**

Volunteering Experience

2017 – 2019 • **Coordinator** All Ireland Linguistics Olympiad (AILO) and International Linguistics Olympiad (2017)

2018 • **Spanish Teacher**

2016 • **Intercultural Ambassador.**

Languages

- Native: Spanish
- Proficiency level: English
- Intermediate level: French, Basque
- Basic level: Japanese, Serbian, German, Italian