

Leon Derczynski

Phone: +45 5157 4948
E-mail: leonderczynski@gmail.com
URL: <http://www.derczynski.com/>

Born: Sussex, UK
Nationality: British

Areas of specialisation

AI: Natural Language Processing
Information Extraction · Social Media · Spatio-Temporal Language

Current position

Research Fellow, Department of Computer Science, University of Sheffield

Research projects

- 2016- [Comrades](#): Community Resilience and Social Innovation during Crises (EC H2020)
- 2014-2017 [PHEME](#): Establishing veracity of claims on the web - “a lie detector for social media.” (EC FP7)
- 2012-2016 [uComp](#): Closing the AI gap using human computation. (EC FP7/CHIST-ERA)
- 2012-2013 [TrendMiner](#): Extracting and predicting trends from social media. (EC FP7)

Grants

- 2016-2018 [Comrades](#): Community Resilience and Social Innovation during Crises. Co-Investigator. (EC €2.0M)
- 2014-2016 [PHEME](#): Veracity and trust in social media. Co-author, named investigator; scientific co-ordinator. (EC €4.3M)
- 2008-2012 Enhanced doctoral training grant. (UK EPSRC, £60k)

Honours & awards

- 2016 NVIDIA Hardware Grant.
- 2016 Engineering Researcher Society’s Development Opportunity award (University of Sheffield).
- 2015 5th place at TRSAC in the 4K code category.
- 2015 WEBIST best student paper, with co-supervised student Martin Leginus.
- 2014 University of Sheffield Exceptional Contribution Award (given to top 4% of researchers).
- 2013 University of Sheffield Exceptional Contribution Award (given to top 4% of researchers).
- 2013 ACM SIGWEB Ted Nelson best newcomer shortlisted.
- 2013 RANLP best young researcher nominated.
- 2008 [Douglas Lewin prize](#) for best final-year performance – Computer Science, University of Sheffield.
- 2006, 2007 Best English-language Teacher – Nanjing University of Aeronautics and Astronautics.

Appointments held

2016-

[University of Sheffield](#), UK. **Research Fellow.**

This permanent research position, at the level of lecturer (UK)/assistant professor (US), began May 2016, and is my current main employment.

In this position, I work in the EC COMRADES project, of which I am a co-investigator. This project address real-time identification and geolocation of crisis needs. As we developed powerful veracity analysis tools in both PHEME and COMRADES, working in a real-time pipeline that identifies claims and the dialogue around them, clusters them, and returns realtime streaming veracity updates, fed into dashboards. My work focuses on streaming event clustering and the underlying social media processing required, as well as spatio-temporal annotation and real-time reconstruction of emerging events expressed in data streams.

The University of Sheffield is a competitive UK university (QS world rank #82), with a very strong Department of Computer Science. This Department was ranked fifth overall for research in the UK in the most recent national research evaluation exercise, outranking (for example) Oxford and Cambridge. Its home faculty, the Faculty of Engineering, has the highest engineering faculty income in the UK. The Department is also home to the largest concentration of Artificial Intelligence research in England, with world-leading groups in Machine Learning, Robotics, Speech and Hearing, and Natural Language Processing. My home research group (NLP) here is around forty people, with four active full professors. This has made Sheffield a strong place to base and grow my skill set.

2017

[University of California San Diego](#), USA. **Visiting Professor.**

This half-year long stay at the Computer Science department, with Julian McAuley in the Artificial Intelligence group, focuses on applying deep learning methods to text, in the problem domains of community detection and of analysis of variations across dimensions of context such as time, place, and group.

2016-2017

[Innopolis University](#), Tatarstan, Russia. **Visiting Faculty.**

Advised on advanced machine learning and natural language processing over a two semester period, delivering a challenging full-length course to final-year masters' and to bachelors' students. Innopolis is a novel and ambitious private university operating outside of the state, funded by industrial partners, where students pay no fees and the only means of admission is excellent academic performance. This is Russia's only university operating in English.

2012-2016

[University of Sheffield](#), UK. **Postdoctoral Research Associate.**

This post-doctoral research position was at the department ranked 5th in the UK for computer science research. Work focused on NLP for social media, and as a member of the [GATE](#) group. The post led to strong research outputs. My paper at Hypertext '13 was one of three nominated and shortlisted for the ACM SIGWEB Ted Nelson Award, and my EDBT paper was one of only three chosen to start that conference's Vision Track the same year. [TwitIE](#), a social media NLP pipeline, was also developed in this position, and represents the first practical tool of its kind. The toolkit is already in use at Intel and the Swedish Defence Research Agency, yielding immediate research impact; the part-of-speech tagger it includes is used by CSIRO, the Australian disaster response body, for detecting and responding to bush fires through latent social media signals.

More recently, work has focused on better computational representations of text, leading to a re-factoring of the powerful Brown Clustering algorithm, which was competitive at the state-of-the-art despite not having been altered since 1992. The resulting generalized implementation of this algorithm not only runs degrees of magnitude faster than prior software but also enables rapid fea-

ture extraction and exploration of the latent distributional structures in any sequence-based data.

Voluntary student supervision work during this post has led to publications on expression of time in language with a Masters' student from the University of Bologna; prize-winning work in information access and entity agglomeration with a PhD student from Aalborg University; a Danish-language toolkit with a PhD student at Aarhus University; and successful collaboration addressing social media overload with a PhD student within the Sheffield group.

In this role I co-authored and led the PHEME project as Scientific co-ordinator, as Sheffield co-ordinated and I co-authored this EC grant. Through it, we generated many best papers, with consistent scientific excellence, the project rated the top grade of "Excellent" by the Commission.

2012- [Aarhus University](#), Denmark. **Visiting Scholar.**

Collaborations at AU have focused on spatio-temporal data extraction (under Christian S. Jensen) and machine learning (under Ira Assent). An early paper from this was published at EDBT and was one of only three selected to start the vision track at this highly-regarded data management conference. Subsequent work with visiting students has been nominated for awards at RANLP. The most recent output from AU, published at AAAI, focuses on clustering and unsupervised NLP, working on the interface between distributional representations and data mining. This has led to, for example, to work on NUMA and GPU optimisation of representation learning from text.

2008 [Illusive Creations Development](#), Calgary, Canada. **Head of development.**

Worked to provide business development and technical management for a portfolio of enterprises, including property management and sales, hobbyists, and human resource management. A high-responsibility position at a multi-million dollar company.

Of note during my time at Illusive Creations, projects I led included: launching a national Canadian property sales site which generated leads for over CAD\$40M of properties per month; designing and launching a new high-resilience web hosting platform from scratch; and through marketing actions, doubling traffic to a client's hobby sites to 8 million monthly pageviews.

2006-2007 [Nanjing University of Aeronautics and Astronautics](#), Nanjing, PRC. **Teacher of specialist English.**

Developed and delivered language courses for university students, in business, engineering, economics and computer science. Full-time job with 40 contact weeks / year each having 16-20 contact hours. The job also offered ample opportunities for travel and to pick up casual Chinese.

While at NUAA, I organised an English-language letter writing campaign with the students to raise money for a local "hope school" (a country school for disadvantaged Chinese children). During this, the students engaged hundreds of businesses across the world and raised a significant amount for equipment and scholarships at this school, receiving some local press coverage.

2003-2006 [Summit Media](#), Brough, UK; Prague, Czech Republic. **Head of R&D.**

Commercial research and development work for a UK top-10 new media marketing agency. This included research in information retrieval and NLP, to aid the creation of cutting-edge tools and adapting emerging technologies to clients' best advantage, as well as managing a team of twenty developers in Prague. This post involved close collaboration with Google and Yahoo!.

The [SummitFEED](#) product that I proposed, designed, developed and launched at Summit Media automatically creates, manages and adjusts product advertisements and placement according to given return-on-interest targets, and is still in use today.

Education

UNIVERSITY

- 2013 **PHD in Computer Science**
University of Sheffield; “Determining the Types of Temporal Relations in Discourse”.
Advisor: Robert Gaizauskas. Examiners: Mark Steedman, Yorick Wilks.
- 2008 **MCOMP in Computer Science**
University of Sheffield; Graduated equiv. *dux litterarum*, best in year
(my mean final year mark was 74.1%; the UK boundary for a first-class grade is 70%).
Advisors: Mark A. Greenwood, Robert Gaizauskas.
- 2000-2003,
2005-2006 **BENG in Software Engineering**
University of Sheffield; Part of integrated master’s.
Final year dissertation on “Machine Learning Techniques for Document Selection”
(scored 81%; the boundary for a first-class grade is 70%).
Advisor: Amanda Sharkey. Examiner: Roger K. Moore.

SECONDARY

- 1991-1998 **UK GCSEs, A-levels**
[Exeter School](#). Won an [assisted place](#) based on academic assessment. Directed and worked in a successful Young Enterprise company during sixth form. Four full A-levels and nine GCSEs total; scored 8th best in the country for Electronics. Left to a [Year in Industry](#) placement after completing this education one year ahead.

TRAINING

- 2013 LxMLS Machine Learning School – *Lisbon, Portugal*
2011 GATE: Text Mining summer training course – *Sheffield, UK*
2010 European Summer School in Logic, Language and Information – *Copenhagen, Denmark*
2009 European Summer School in Logic, Language and Information – *Bordeaux, France*
2009 Midlands Graduate School in Computer Science – *Leicester, UK*
1999 National Examining Board in Supervisory Management – *Pershore, UK*

NATURAL LANGUAGES

English (British, RP, native)
Danish (intermediate proficiency - fint, det går nok)
Chinese (working basic conversational - 马马虎虎)
French (basic - pas mieux qu’un rosbif typique)

PROGRAMMING & OTHER COMPUTER SKILLS

Programming languages (Python, Java)
Science packages (R, SciPy, Keras)
Language Processing Frameworks (GATE – certified, committer; NLTK – project committer)
My profile on [GitHub](#)

VOLUNTEER WORK

- 2017 San Diego Park River Foundation – conservancy work for the San Diego river. While I was there, SDPRF won California Non-Profit of the year, and its weekly River Rescue event was recognized by the City of San Diego; July 18 became River Rescue Day.
- 2017- Dungeons & Dragons San Diego / Aarhus – I'll always volunteer to DM.

Teaching & talks

HIGHER EDUCATION COURSES

- 2016, 2017 **Innopolis University, Russia**
Natural Language Processing and Machine Learning
This course gives a tour of data-intensive natural language processing (NLP) and machine learning (ML), worth 9 ECTS. Taught over four weeks, each week having three six-hour days, for a total of 72 hours contact. This was an intensive, interactive, pre-PhD level course.
As AI is in the press these days achieving many exciting things, each week, we took a problem or that had received coverage and attempted to replicate it. Sessions were run with taught material, frequent examples of [real-world problems language technology faces](#), and IPython notebooks that students solved collaboratively. Example assignments included building a self-driving car, creating a traffic light recogniser, social media text tagging, and extracting sociolinguistic characteristics like age and gender from user text.
Topics covered included various takes on language modelling, deep neural nets, handy machine learning techniques (Rademacher, dealing with imbalanced classes, tuning SGD optimisation), reinforcement learning, statistical & neural MT, dialogue systems, and social media processing. The style is to set students up for either PhD-level research or going straight into a competitive startup.
- 2016 **Aarhus Universitet, Denmark**
[Machine Learning for Natural Language Processing](#) – 4 ECTS (guest lecturer).
This course involved four separate weeks, each having their own assignment. The last assignment is decided by the students, to get the most engagement out of them and so hopefully their best work. Attended by students from multiple faculties, the course was a success.
- 2010, 2011 **University of Sheffield, UK**
Teaching Assistant (60 students, weekly tuition and exercises, five hours/week)
This involved developing course material, delivering tutorials every week, responding to student feedback, and marking. I assisted on the course for two years, which enabled iterative growth. In the second year, we changed assessment to be a short weekly programming assignment, which was received well by the students.
MSc Programme in Computer Science – COM4250: Natural Language Processing – 5 ECTS
Postgraduate student representative for two years; disciplinary board student member.
- 2006-2007 **Nanjing University of Aeronautics and Astronautics, PRC**
Teacher (250 students in six sets, 20 contact hours/week)
Supporting various Engineering and Science Bachelor programmes, training students in application-specific English language. Responsibilities included developing and delivering course material, marking, and feedback, over a full academic year.
English Language and *Specialised English for engineering and business*

SHORT COURSES TAUGHT

- 2017 Networks and User-generated Content (with [Matteo Magnani](#)) – *ESSLLI, University of Toulouse, France*
- 2016 [Language and Networks](#) (guest lecturer) – *Uppsala Universitet, Sweden*
- 2016 Using GATE for social media analysis – *GATE Summer School, UK*
- 2015 Using GATE for social media analysis – *GATE Summer School, UK*
- 2014 Mining social media with GATE – *GATE Summer School, UK*
- 2014 [Network Science and online Social Network Analysis](#) (guest lecturer) – *Uppsala Universitet, Sweden*
- 2014 Human Information Networks (with [Matteo Magnani](#)) – *ESSLLI, University of Tübingen, Germany*

- 2013 [NLP for Social Media](#) (with [Kalina Bontcheva](#)) – *Doctoral school, University of Szeged, Hungary*
 2013 [TwitIE: Open Text Engineering for Twitter for GATE](#) – *GATE Summer School, UK*
 2012 [Mining Twitter, Facebook, and other social media with GATE: Basics, Tools, and Applications](#) – *GATE Summer School, UK*

TUTORIALS GIVEN

- 2016 [Humanities Programming Python](#) – *Institute for Communication and Culture, AU, Denmark*
 2016 [Social Media: Practical Challenges and Opportunities](#) – *Digital Humanities summer school, Aarhus, Denmark*
 2015 [NLP for Social Media](#) – *RANLP, Hissarya, Bulgaria*
 2015 [Practical Annotation and Processing of Social Media with GATE](#) – *ESWC, Portorož, Slovenia*
 2014 [Practical Social Media Analysis: finding utility in trivia](#) (with [Diana Maynard](#)) – *LREC, Reykjavík, Iceland*
 2014 [Natural Language Processing for Social Media](#) (with [Kalina Bontcheva](#)) – *EACL, Chalmers University of Technology, Sweden*

INVITED WORKSHOPS

- 2017 [Social AI: Inequality and Discrimination](#) – *UCI Donald Bren School of Information and Computer Sciences, California*
 2014 [How to analyse social media content](#) – *Universität der Bundeswehr München, Germany*

KEYNOTES

- 2016 [Why scientists can't teach AI to pass intelligence tests](#) ([Summit on Human Problem Solving and Artificial Intelligence](#)) – *Sandbjerg, Denmark*
 2015 [Handling and Mining Linguistic Variation in UGC](#) ([LT₄VarDial](#)) – *RANLP, Hissarya, Bulgaria*

INVITED TALKS AND LECTURES

- 2016 [Why AI needs the crowd and your social media](#) – *Aarhus University Datalab, Denmark*
 2016 [How we write online, and what it says about us](#) – *Massive Cultural Data, Aarhus, Denmark*
 2016 [Processing and Profiling Social Media](#) – *Innopolis University, Russian Federation*
 2016 [Broad Twitter Corpus: A Diverse Named Entity Recognition Resource](#) – *University of Sheffield, UK*
 2016 [Sampling human discourse and activity](#) – *Interacting Minds Centre, Aarhus University, Denmark*
 2015 [Frontiers in named entity recognition and linking](#) – *Cambridge University, UK*
 2014 [NLP and Information Access in Social Media](#) – *Aalborg University, Denmark*
 2013 [Mining social media with linked open data, entity recognition, and event extraction](#) – *DEOS, Oxford University, UK*
 2012 [Building Stories: Determining temporal orderings from natural language](#) – *AU, Denmark*
 2011 [Introduction to the current state of the art in Word Sense Disambiguation](#) – *University of Sheffield, UK*
 2010 [Panelist on "Temporal and Spatial Annotation"](#) – *LREC, Valletta, Malta*
 2009 [Sequencing of Events & their Durations based on Event Descriptions](#) – *University of Sheffield, UK*
 2006 [Challenges of Migrating to Agile Methodologies](#) – *University of Sheffield, UK*

STUDENT CO-SUPERVISION & TUTELAGE

- 2016- [Rehab Qarout](#) – PhD student, University of Sheffield
 2015 [Manuel Ciosici](#) – Masters' student, Aarhus University; completed his masters' with the top possible mark (12/12)

- 2015 [Michał Łukasik](#) – PhD student, University of Sheffield
- 2014 [Martin Leginus](#) – PhD student, Aalborg University; best student paper at WEBIST'15
- 2013 [Matteo Brucato](#) – Masters' student, University of Bologna
- 2013 [Dominic Rout](#) – PhD student, University of Sheffield

Publications

Publication profile on [Google Scholar](#).

BOOKS

- 2017 **Temporal Information Processing in Language**
Pustejovsky, J., Derczynski, L., Verhagen, M. – *forthcoming*
O'Reilly
- 2017 **[Automatically Ordering Events and Times in Text](#)**
Derczynski, L. *Studies in Computational Intelligence*
Springer

JOURNAL ARTICLES

- 2017 **Generalisation in Named Entity Recognition: A Quantitative Analysis**
Augenstein I., Derczynski L., Bontcheva K.
Computer Speech and Language
- 2016 **Novel psychoactive substances: an investigation of temporal trends in social media and electronic health records**
Kolliakou, A., Ball, M., Derczynski, L., Chandran, D., Gkotsis, G., Deluca, P., Jackson, R., Shetty, H., Stewart, R.
European Psychiatry
- 2015 **Analysis of Named Entity Recognition and Linking for Tweets**
Derczynski, L., Maynard, D., Rizzo, G., van Erp, M., Gorrell, G., Troncy, R., Petrak, J., Bontcheva, K.
Information Processing & Management (IPM)
- 2015 **Time and Information Retrieval: Introduction to the Special Issue**
Derczynski, L., Strötgen, J., Campos, R., Alonso, O.
Information Processing & Management (IPM)

THESES AND DISSERTATIONS

- 2013 **Determining the Types of Temporal Relations in Discourse**
Derczynski, L.,
Doctoral Thesis, University of Sheffield
Springer later approached me and we adapted and updated this to a book.
- 2008 **Question Answering Against Very Large Text Collections**
Derczynski, L.,
Masters' Report, University of Sheffield
- 2006 **Machine Learning Techniques for Document Selection**
Derczynski, L.,
Bachelor's Dissertation

PEER-REVIEWED CONFERENCE PAPERS

- 2017 **Simple Open Stance Classification for Rumour Analysis - [RANLP](#)**
Aker, A., Derczynski, L., Bontcheva, K.
Proceedings of the International Conference on Recent Advances in Natural Language Processing
- 2016 **Broad Twitter Corpus: A Diverse Named Entity Recognition Resource - [COLING](#)**
Derczynski, L., Bontcheva, K., Roberts, I.
Proceedings of the 26th International Conference on Computational Linguistics
- 2016 **Representation and Learning of Temporal Relations - [COLING](#)**
Derczynski, L.
Proceedings of the 26th International Conference on Computational Linguistics
- 2016 **Complementarity, F-score, and NLP Evaluation - [LREC](#)**
Derczynski, L.
Proceedings of the 10th International Conference on Language Resources and Evaluation
- 2016 **GATE-Time: Extraction of Temporal Expressions and Events - [LREC](#)**
Derczynski, L., Strötgen, J., Maynard, D., Greenwood, M.A., Jung, M.
Proceedings of the 10th International Conference on Language Resources and Evaluation
- 2016 **Generalised Brown Clustering and Roll-Up Feature Generation - [AAAI](#) full paper, talk**
Derczynski, L., Chester, S.
Proceedings of the 30th AAAI Conference on Artificial Intelligence - acceptance rate 26%
- 2015 **Enhanced Information Access to Social Streams through Word Clouds with Entity Grouping - [WEBIST](#) full paper**
Leginus, M., Derczynski, L., Dolog, P.
Proceedings of the 11th International Conference on Web Information Systems and Technologies - acceptance rate 14%, Best Student Paper
- 2015 **Tune Your Brown Clustering, Please - [RANLP](#) poster**
Derczynski, L., Chester, S., Bøgh, K.S.
Proceedings of the International Conference on Recent Advances in Natural Language Processing
- 2015 **Temporal Relation Classification using a Model of Tense and Aspect - [RANLP](#) poster**
Derczynski, L., Gaizauskas, R.
Proceedings of the International Conference on Recent Advances in Natural Language Processing
- 2015 **Efficient named entity annotation through pre-empting - [RANLP](#) talk**
Derczynski, L., Bontcheva, K.
Proceedings of the International Conference on Recent Advances in Natural Language Processing
- 2014 **Passive-Aggressive Sequence Labeling with Discriminative Post-Editing for Recognising Person Entities in Tweets - [EACL](#) talk**
Derczynski, L., Bontcheva, K.
Proceedings of the conference of the European Chap. of the Association for Computational Linguistics - acceptance rate 23%

- 2014 **DKIE: Open Source Information Extraction for Danish** - [EACL](#) demo
Derczynski, L., Field, C.V., Bøgh, K.S.
Proceedings of the conference of the European Chap. of the Association for Computational Linguistics
- 2014 **The GATE Crowdsourcing Plugin: Crowdsourcing Annotated Corpora Made Easy** - [EACL](#) demo
Bontcheva, K., Roberts, I., Derczynski, L.
Proceedings of the conference of the European Chap. of the Association for Computational Linguistics
- 2014 **Corpus Annotation through Crowdsourcing: Towards Best Practice Guidelines** - [LREC](#) talk
Sabou, M., Bontcheva, K., Derczynski, L. and Scharl, A.
Proceedings of the 9th International Conference on Language Resources and Evaluation
- 2013 **Information Retrieval for Temporal Bounding** - [ICTIR](#) poster
Derczynski, L. and Gaizauskas, R.,
Proceedings of the 4th International Conference on the Theory of Information Retrieval, ACM SIGIR - acceptance rate 22%
- 2013 **Temporal Signals Help Label Temporal Relations** - [ACL](#) short
Derczynski, L. and Gaizauskas, R.,
Proceedings of the meeting of the Association for Computational Linguistics - acceptance rate 24%
- 2013 **Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data** - [RANLP](#) talk
Derczynski, L., Ritter, A., Clark, S., Bontcheva, K.,
Proceedings of the International Conference on Recent Advances in Natural Language Processing - long, acceptance rate 11.8%
- 2013 **Recognising and Interpreting Named Temporal Expressions** - [RANLP](#) talk
Brucato, M., Derczynski, L., Llorens, H., Bontcheva, K., Jensen, C.S.,
Proceedings of the International Conference on Recent Advances in Natural Language Processing - long, acceptance rate 11.8%
- 2013 **TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text** - [RANLP](#) talk
Bontcheva, K., Derczynski, L., Funk, A., Greenwood, M.A., Maynard, D., Aswani, N.,
Proceedings of the International Conference on Recent Advances in Natural Language Processing - long, acceptance rate 11.8%
- 2013 **Microblog-Genre Noise and Impact on Semantic Annotation Accuracy** - [ACM Hypertext](#) talk
Derczynski, L., Maynard, D., Aswani, N. and Bontcheva, K.,
Proceedings of the ACM Conference on Hypertext and Social Media - acceptance rate 16.7% (paper shortlisted for ACM SIGWEB Ted Nelson award)
- 2013 **Empirical Validation of Reichenbach's Tense Framework** - [IWCS](#) talk
Derczynski, L. and Gaizauskas, R.,
Proceedings of the International Conference on Computational Semantics
- 2013 **Towards Context-Aware Search and Analysis on Social Media Data** - [EDBT](#) talk
Derczynski, L., Yang, B. and Jensen, C.S.,
Proceedings of the International Conference on Extending Database Technology, ACM - track acceptance rate 13.6%

- 2012 **TIMEN: an Open Temporal Expression Normalization Resource** - [LREC](#) talk
Llorens, H., Derczynski, L., Gaizauskas, R. and Saquete, E.,
Proceedings of the 8th International Conference on Language Resources and Evaluation
- 2012 **Massively increasing TIMEX₃ resources: a transduction approach** - [LREC](#) poster
Derczynski, L., Llorens, Hector and Saquete, E.,
Proceedings of the 8th International Conference on Language Resources and Evaluation
- 2010 **Analysing Temporally Annotated Corpora with CAVaT** - [LREC](#) talk
Derczynski, L. and Gaizauskas, R.,
Proceedings of the 7th International Conference on Language Resources and Evaluation
- BOOK CHAPTERS
- 2016 **Entity Grouping for Accessing Social Streams via Word Clouds**
Leginus, M., Derczynski, L., Dolog, P.,
Lecture Notes in Business Information Processing
- 2016 **Extracting Information from Social Media with GATE**
Bontcheva, K. and Derczynski, L.,
Working with Text: Tools, Techniques and Approaches for Text Mining - Emma Tonkin, Gregory Tourte, eds.
- 2015 **Crowdsourcing Named Entity Recognition and Entity Linking Corpora**
Bontcheva, K., Derczynski, L. and Roberts, I.,
The Handbook of Linguistic Annotation - Nancy Ide, James Pustejovsky, eds.
- PEER-REVIEWED WORKSHOP PAPERS
- 2017 **Results of the WNUT₂₀₁₇ Shared Task on Novel and Emerging Entity Recognition** - [W-NUT](#) talk
Derczynski, L., Nichols, E., van Erp, M., Limsopatham, N.
Proceedings of the 3rd Workshop on Noisy User-Generated Text (W-NUT 2017), at EMNLP
- 2017 **SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours** - [SemEval](#) talk
Derczynski, L., Bontcheva, K., Liakata, M., Zubiaga, A., Procter, R., Wong Sak Hoi, G.
Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval 2017), at ACL
- 2016 **Twitter Geolocation Prediction Shared Task of the 2016 Workshop on Noisy User-generated Text** - [W-NUT](#) talk
Han, B., Rahimi, A., Derczynski, L., Baldwin, T.
Proceedings of the 2nd International Workshop on Noisy, User-generated Text (W-NUT 2016), at COLING

- 2016 **SemEval-2016 Task 12: Clinical TempEval - [SemEval](#) talk**
Bethard, S., Savova, G., Chen, W-T., Derczynski, L., Pustejovsky, J., Verhagen, M.
Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016), at NAACL
- 2015 **SemEval-2015 Task 6: Clinical TempEval - [SemEval](#) talk**
Bethard, S., Derczynski, L., Savova, G., Pustejovsky, J., Verhagen, M.
Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), at NAACL
- 2015 **UFPRSheffield: Contrasting Rule-based and Support Vector Machine Approaches to Time Expression Identification in Clinical TempEval - [SemEval](#) poster**
Tissot, H., Gorrell, G., Roberts, A., Derczynski, L., Didonet del Fabro, M.
Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), at NAACL
- 2015 **Swiss-Chocolate: Combining Flipout Regularization and Random Forest with Artificially Built Subsystems to Boost Text-Classification for Sentiment - [SemEval](#) poster**
Uzdilli, F., Jaggi, M., Egger, D., Julmy, P., Derczynski, L., Cieliebak, M.
Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), at NAACL
- 2015 **Analysis of Temporal Expressions Annotated in Clinical Notes - [ISA](#) talk**
Tissot, H., Roberts, A., Derczynski, L., Gorrell, G., Didonet del Fabro, M.
Proceedings of the 11th Joint ACL - ISO Workshop on Interoperable Semantic Annotation (ISA), at IWCS
- 2015 **USFD: Twitter NER with Drift Compensation and Linked Data - [W-NUT](#)**
Derczynski, L., Augenstein, I., Bontcheva, K.
Proceedings of the Workshop on Noisy User-generated Text, at ACL
- 2014 **PHEME: Veracity in Digital Social Networks - [UMAPProS](#) talk**
Derczynski, L., Bontcheva, K.
Proceedings of the UMAP Project Synergy (UMAP ProS) Workshop, at UMAP
- 2014 **Spatio-temporal grounding of claims made on the web, in PHEME - [ISA](#) talk**
Derczynski, L., Bontcheva, K.
Proceedings of the 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation, at LREC
- 2013 **SemEval-2013 Task 1: TempEval-3: Evaluating Events, Time Expressions, and Temporal Relations - [SemEval](#) talk**
UzZaman, N., Llorens, H., Derczynski, L., Allen, J.F., Verhagen, M., Pustejovsky, J.,
Proceedings of the 7th International Workshop on Semantic Evaluation, at NAACL
- 2012 **Applying ISO-Space to Healthcare Facility Design Evaluation Reports - [ISA](#) talk**
Gaizauskas, R., Barker, E., Chang, C.L., Derczynski, L., Phiri, M., and Peng, C.Z.,
Proceedings of the 7th Joint ACL - ISO Workshop on Interoperable Semantic Annotation, at LREC
- 2011 **An Annotation Scheme for Reichenbach's Verbal Tense Structure - [ISA](#) talk**
Derczynski, L. and Gaizauskas, R.,
Proceedings of the 6th Joint ACL - ISO Workshop on Interoperable Semantic Annotation, at IWCS

- 2010 **USFD2: Annotating Temporal Expressions and TLINKs for TempEval-2** - [SemEval](#) poster
Derczynski, L. and Gaizauskas, R.,
Proceedings of the 5th Workshop on Semantic Evaluations, at ACL
- 2008 **A Data Driven Approach to Query Expansion in Question Answering** - [IR4QA](#) talk
Derczynski, L., Greenwood, M.A., Gaizauskas, R. and Wang, J.,
Proceedings of the 2nd Workshop on Information Retrieval for Question Answering, at CoLing
- PEER-REVIEWED CONFERENCE ABSTRACTS
- 2016 **Language as a reflection of mental time travel** - [Travelling in time](#) abstract, poster
Derczynski, L.
Conference on Travelling in time: The construction of past and future events across domains
- 2011 **A Corpus-based Study of Temporal Signals** - [Corpus Linguistics](#) talk
Derczynski, L. and Gaizauskas, R.,
Proceedings of the Corpus Linguistics Conference
- NON-REVIEWED PAPERS AND TECHNICAL VOLUMES
- 2016 **Desiderata for Vector-Space Word Representations**
Derczynski, L.
arXiv:1608.02094
- 2015 **Political Futures Tracker: Technical report**
Maynard, D., Roberts, I., Greenwood, M.A., Derczynski, L., Bontcheva, K.,
[Nesta working paper 15/12](#)
- 2014 **Clinical TempEval**
Bethard, S., Derczynski, L., Pustejovsky, J., Verhagen, M.
arXiv:1403.4928
- 2013 **TimeML-strict: clarifying temporal annotation**
Derczynski, L., Llorens, H., UzZaman, N.,
arXiv:1304.7289
- 2012- **Developing Language Processing Components with GATE Version 8 (a User Guide)**
Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, A., Gorrell, G., Funk, A., Roberts, I., Damjanovic, D., Heitz, T., Greenwood, M.A., Saggion, H., Petrak, J., Li, Y., Peters, W., Derczynski, L.
The University of Sheffield
- 2012 **TempEval-3: Evaluating Events, Time Expressions, and Temporal Relations**
UzZaman, N., Llorens, H., Allen, J.F., Derczynski, L., Verhagen, M. and Pustejovsky, J.,
arXiv:1206.5333v1

2011 **USFD at KBP 2011: Entity Linking, Slot Filling and Temporal Bounding** - [TAC](#)
Burman, J., Kannan, K., Alhelbawy, A., Derczynski, L. and Gaizauskas, R.,
Proceedings of the Text Analysis Conference

BLOG POSTS AND EDITORIALS

2017 **A Random Walk Through EMNLP 2017** - [Approximately Correct](#)

2017 **Pathology of a fake news article** - [medium](#)
This was the top-ranked article for the May Day weekend, headlining the front page of medium.com.

Service to the profession

PROFESSIONAL BODIES & ADMINISTRATIVE

Current

Member of the [British Computer Society](#).
Member of the [Association for Computational Linguistics](#).
Member of the [Association for the Advancement of Artificial Intelligence](#).
Member of the [AU Datalab](#).

Past

2011-2013 Participant in [ISO/TC37/SC4](#) Language Annotation Standards working group.
2008-2013 Language Resource co-ordinator for [Institute for Language, Speech and Hearing](#).
2010-2011 Participant in [OSEPA](#) (Open Source in European Public Administration).

CHALLENGES AND EVALUATIONS

2016-2017 [RumourEval 2017](#) – Lead task organiser.
2016 [W-NUT₂](#) – Spatial task organiser.
2015-2016 [Clinical TempEval 2016](#) – Task organiser.
2013-2015 [Clinical TempEval 2015](#) – Task organiser.
2011-2013 [TempEval-3](#) – Task organiser.

OPEN-SOURCE DEVELOPMENT PROJECTS

2012- [GATE](#) – General Architecture for Text Engineering. Contributions include social media pipeline.
2011- [TIMEN](#) – Open source collaborative temporal normalisation tool.
2010- [NLTK](#) – Natural Language ToolKit. Contributions include SVM, machine learning framework.

ADVISORY

Program Chair for:

2018 [COLING](#) - International Conference on Computational Linguistics, Santa Fe. With [Emily Bender](#).

Journal Editor for:

2014-2015 [IPM](#) - Lead guest editor for a special issue on [Time and Information Retrieval](#)

Editorial board member of:

2012- [LangSci](#) - Language Science Press
2016-2017 [IPM](#) - Information Processing & Management

Workshop chair for:

2017 [W-NUT₃](#) - General chair for 3rd Workshop for Noisy User-Generated Text, at [EMNLP](#)

Area chair for:

2018 [SemEval](#) - International Workshop on Semantic Evaluation
2016 [SemEval](#) - International Workshop on Semantic Evaluation, at [ACL](#)
2015 [SemEval](#) - International Workshop on Semantic Evaluation, at [NAACL](#)

Workshop organiser for:

2016 [W-NUT2](#) - 2nd Workshop for Noisy User-Generated Text, at [COLING](#)

Session chair at:

2016 [COLING](#) - Computational Linguistics, Applications II

2015 [Digifeet15](#) - Preparing and analyzing large social media datasets, morning sessions and conclusion

2015 [RANLP](#) - Recent Advances in Natural Language Processing, NLP for Internet and Social Media

2015 [ESWC](#) - Extended Semantic Web Conference, Opening NLP track

Scientific committee / conference reviewer for:

2017 [AAAI](#) - AAAI Conference on Artificial Intelligence

2017 [ACL](#) - Annual meeting of the Association for Computational Linguistics

2017 [EACL](#) - Conference of the European Chapter of the Association for Computational Linguistics

2017 [EMNLP](#) - Conference on Empirical Methods in Natural Language Processing

2017 [ISWC](#) - International Semantic Web Conference

2016 [AAAI](#) - AAAI Conference on Artificial Intelligence

2016 [ACL](#) - Annual meeting of the Association for Computational Linguistics

2016 [COLING](#) - International Conference on Computational Linguistics

2016 [CNews](#) - Workshop on Computing News Storylines, at EMNLP

2016 [EMNLP](#) - Conference on Empirical Methods in Natural Language Processing

2016 [LREC](#) - International Conference on Language Resources and Evaluation

2016 [SNOW](#) - Workshop on Social News on the Web, at WWW

2016 [SocialNLP](#) - Workshop on Natural Language Processing for Social Media

2016 [SemEval](#) - International Workshop on Semantic Evaluation, at NAACL

2015 [ACL](#) - Annual meeting of the Association for Computational Linguistics

2015 [CNews](#) - Workshop on Computing News Storylines, at ACL

2015 [EMNLP](#) - Conference on Empirical Methods in Natural Language Processing

2015 [ESWC](#) - Extended Semantic Web Conference

2015 [IWCS](#) - International Conference on Computational Semantics

2015 [MASS](#) - Mobility Analytics from Spatial and Social Data workshop, at ICDM

2015 [#Microposts 2015](#) - #Microposts Workshop, at WWW

2015 [NAACL](#) - Conf. of the North American Chapter of the Assoc. for Computational Linguistics

2015 [NAACL SRW](#) - NAACL Student Research Workshop

2015 [NLPIT](#) - Natural Language Processing for Informal Text, at ICWE

2015 [RDSM](#) - Rumours and Deception in Social Media workshop, at WWW

2015 [SemEval](#) - International Workshop on Semantic Evaluation, at NAACL

2015 [W-NUT](#) - Workshop on Noisy User-generated Text, at ACL

2015 [WTI@SIMBig](#) - International Symposium on Information Management and Big Data

2014 [ACL](#) - Annual meeting of the Association for Computational Linguistics

2014 [ACL Demo](#) - ACL Demo session

2014 [ACL SRW](#) - ACL Student Research Workshop

2014 [AHA!](#) - Workshop on Information Discovery in Text, at CoLing

2014 [EACL](#) - Conference of the European Chapter of the Association for Computational Linguistics

2014 [EMNLP](#) - Conference on Empirical Methods in Natural Language Processing

2014 [ESWC](#) - Extended Semantic Web Conference

- 2014 [IDEAS](#) - International Database Engineering & Applications Symposium
 2014 [LREC](#) - International Conference on Language Resources and Evaluation
 2014 [#Microposts](#) - Making Sense of Microposts Workshop, at WWW
 2014 [OIAF4HLT](#) - Joint Workshop on Open Infrastructures for HLT Resource Processing and Development and Text Analysis Frameworks, at CoLing
 2014 [WASSA](#) - Workshop on Comp. Approaches to Subjectivity, Sentiment & Social Media Analysis, at ACL
- 2013 [AAAI](#) - Association for the Advancement of Artificial Intelligence Conference
 2013 [ACL](#) - Annual meeting of the Association for Computational Linguistics
 2013 [CISIS](#) - Computational Intelligence in Security for Information Systems
 2013 [EMNLP](#) - Conference on Empirical Methods on Natural Language Processing
 2013 [ISWC](#) - International Semantic Web Conference
 2013 [NAACL-HLT](#) - Conf. of the North American Chapter of the Assoc. for Computational Linguistics
 2013 [RANLP](#) - Recent Advances in Natural Language Processing
 2013 [SemEval](#) - International Workshop on Semantic Evaluation, at NAACL
 2013 [WebSci](#) - ACM Conference on Web Science
- 2012 [ACL](#) - Annual meeting of the Association for Computational Linguistics
 2012 [CoLing](#) - International Conference on Computational Linguistics
 2012 [GEOCROWD](#) - Workshop on Crowdsourced and Volunteered Geographic Info., at SIGSPATIAL
 2012 [LREC](#) - International Conference on Language Resources and Evaluation
 2012 [*SEM](#) - Joint Conference on Lexical and Computational Semantics
- 2011 [ACL](#) - Annual meeting of the Association for Computational Linguistics
 2011 [EMNLP](#) - Conference on Empirical Methods on Natural Language Processing
- 2010 [IceTAL](#) - International Conference on Natural Language Processing
- 2009 [IJCAI](#) - International Joint Conference on Artificial Intelligence

Journal reviewer for:

- [CL](#) - Computational Linguistics
[TOIS](#) - ACM Transactions on Information Systems
[TIST](#) - ACM Transactions on Intelligent Systems and Technology
[IPM](#) - Information Processing and Management
[CSL](#) - Computer Speech and Language
[JNLE](#) - Journal of Natural Language Engineering
[SWJ](#) - Semantic Web Journal
[IS](#) - Information Sciences
[JBI](#) - Journal of Biomedical Informatics
[JLRE](#) - Journal of Language Resources and Evaluation
[JoLLI](#) - Journal of Logic, Language and Information
[LJ-IGPL](#) - Logic Journal of the Interest Group in Pure and Applied Logics
[CAI](#) - Computing And Informatics
[IJICT](#) - International Journal of Information and Communication Technology
[TAL](#) - Traitement Automatique des Langues

Last updated: October 26, 2017
Typeset in X_YT_X
<http://derczynski.com/sheffield/cv.pdf>