THE UNIVERSITY OF SHEFFIELD

CURRICULUM VITAE

ROBERT JOHN GAIZAUSKAS

May, 2024

CURRICULUM VITAE: ROBERT JOHN GAIZAUSKAS
SUMMARY

Research

- Internationally recognised researcher in computational natural language processing, specifically text mining, summarisation, machine translation and dialogue.
- Principal or co-investigator on 29 research grants and contracts since appointment in 1993 worth $\sim$ £27.5 million in total of which $\sim$ £8 million has or will come to Sheffield.
- Author of 2 edited books, 27 refereed journal papers, 100 refereed conference papers and 63 refereed workshops papers, amongst other publications, since 1993.
- Google Scholar H-index of 56 with 14040 citations (08/05/24).

Teaching

- Lecturer on 8 undergraduate/postgraduate courses, at all levels, since appointment.
- Responsible for the introduction of two new core modules in the taught curriculum.
- Involved in supervising over 30 PhD students since appointment.
- Co-applicant and co-director, UKRI Centre for Doctoral Training in Speech and Language Technologies (SLT) and Their Applications (2019-2027), which will fund and train 60 PhD students (UKRI award + matched industry funding of $\sim$ £8 million).
- External examiner for PhD theses and MSc courses at major UK universities and abroad.

Administration

- As REF co-ordinator, co-responsible for preparing the Department's 2014 REF return.
- As Director of Research and Deputy Head of Department, co-responsible for academic staff recruitment, monitoring Departmental research profile, and deputising for Head.
- As Examinations Officer, improved Departmental examinations procedures by streamlining board meetings and establishing intranet procedures.
- As Projects Officer, responsible for introduction of a more efficient, transparent and fair method for assigning projects to 3rd year and MSc students.

Professional Activities

- Appointed member of the EPSRC Peer Review College (2000-present).
- Invited expert in EC Human Language Technology programme consultations.
- Invited member of programme committee for multiple international conferences.
- Reviewer for all the major conferences/journals in my research area.
- Guest editor for special issues of two leading journals in my research area.
- Invited speaker at over 40 research seminars at universities, research institutes and companies in the UK and abroad since appointment in 1993.

CURRICULUM VITAE
ROBERT JOHN GAIZAUSKAS

Contents

ROBERT JOHN GAIZAUSKAS
CURRICULUM VITAE

1   Personal Details

Surname Gaizauskas

Forenames Robert John

Date of Birth February 11, 1955

Citizenship Canadian and British

Languages English; French (working knowledge); Russian (some)

1.1   Education

D.Phil. School of Cognitive and Computing Sciences, University of Sussex, 1992
Thesis Title: Deriving Answers to Logical Queries by Answer Composition:
A New Approach to Deductive Question Answering

M.A. Cognitive Studies, University of Sussex, 1986
Dissertation Title: Model Theory and Knowledge Representation

Information Processing Certificate, Algonquin College, Ottawa, 1981

M.A. Philosophy, Carleton University, Ottawa, 1978 (with Distinction)
Thesis Title: Contradictions and Incompleteness:
A Re-examination of Godel's Result

Honours B.A. Philosophy, Carleton University, Ottawa, 1975 (First Class Honours)

Senior Matriculation, Nepean High School, Ottawa, 1972 (First Class Honours)

1.2   Awards and Scholarships

1985-89 Doctoral Fellowship, Social Sciences and Humanities Research Council of Canada

1985-88 Overseas Research Student Award, British Council

1975-77 Carleton University Graduate Scholarship

1972 Ontario Scholarship

1.3   Academic Work Experience

2002-present Professor (Computer Science), University of Sheffield
> See next entry

1999-2002 Reader (Computer Science), University of Sheffield
> As next entry.

1993-1999 Lecturer (Computer Science), University of Sheffield
> Research responsibilities have included actively publishing and applying for and managing research grants and contracts (details below).
>
> Teaching duties have included lecturing undergraduate courses, individual and group tutoring, final year project supervision and MSc and PhD supervision (details below).
>
> Departmental administrative responsibilities have included being Departmental Director of Research and Innovation, Head of the NLP Research Group, Examinations Officer, Director of Publicity, Deputy Director Research and Innovation, Deputy Admissions Tutor, MSc and 3rd Year Projects Officer, World Wide Web director and sitting on various departmental committees, including an MSc curriculum redesign committee.

1990-93 Research Fellow (AI), University of Sussex, Brighton
> Worked on two related applied natural language (NL) research projects. The first was the POrtable Extendable Traffic Information Collator (POETIC) project, funded by the UK Department of Trade and Industry and the Science and Engineering Research Council and carried out in collaboration with several UK industrial organisations. Responsibilities included designing, implementing and testing software to extract information from police incident logs, writing and presenting research papers, and liaising with users and industrial collaborators.
>
> The second was participating in an international NL system evaluation programme, the Fifth Message Understanding System Evaluation, sponsored by the U.S. Advanced Research Projects Agency. Responsibilities included obtaining the necessary funding for participation and modifying the existing Sussex NL system to take part.

1989 Lecturer (AI), University of Sussex, Brighton
> Lectured a senior undergraduate course in Advanced AI and an MSc course in Prolog Programming. Tutored an introductory course in AI. Supervised MSc dissertations.

1978-81 Lecturer (Philosophy - part-time), Carleton University, Ottawa
> Lectured undergraduate courses in ethics and formal logic.

1977 Researcher, Atomic Energy Control Board of Canada, Ottawa
> Researched and wrote a report on the ethical question of responsibility to future generations in the context of nuclear waste management.

1975-78 Teaching Assistant, Carleton University, Ottawa
> Assisted philosophy lecturers by leading undergraduate tutorial sessions and grading essays.

1.4   Industrial Work Experience

1987-88 Consultant, Cognitive Applications Ltd., Brighton, U.K.
Contracted by a Government Project Monitoring Officer to summarise technical reports and meetings of an Alvey Project Group on Logic Programming whose participants were ICL, Imperial College, Bradford University and University of Edinburgh.

1984-85 Software Engineer, Fulcrum Technologies Inc., Ottawa, Canada
Contributed to software product design and implementation for an extremely successful small firm which specialised in full text retrieval systems. Designed and programmed a file access and indexing sub-system in C. Wrote significant portions of a Programmer's Interface Manual sold with the software.

1983-84 Programmer/Analyst, Domus Software Ltd., Ottawa, Canada
As a software consultant, contributed to the analysis, design and implementation of a variety of software systems, including a system for administering social housing programs for the Canada Housing and Mortgage Corporation, and systems for forms management and space inventory management for Supply and Services Canada. Responsibilities included eliciting system requirements from clients, using structured system design techniques to produce design and functional specification documents, writing code and supervising programming activities of others, designing test plans and writing user documentation.

1981-83 Programmer, Mobius Software Ltd. (Nabu Corp.), Ottawa, Canada
Nabu Corporation's mission was to deliver application software and text information over broadband cable networks into customised microcomputers in customers' homes. Worked on the design and implementation of software for packaging the application software for broadcast over the cable network. Responsibilities involved testing an early version of the system, implementing revisions and enhancements, and collaborating in producing a general design and functional specification for a completely redesigned system.

2 Research

2.1 Research Areas

My research interests lie in developing computational techniques to access and manipulate information as expressed in natural language, particularly in written form. These techniques may then be utilised in applications such as document retrieval, text indexing, information extraction, text summarisation, question answering, dialogue processing, machine translation, natural language generation and so forth. I also view the empirical validation and analysis of language processing techniques through quantitative evaluation in 'real' applications on significant amounts of 'real' text as a central component of my research methodology and an interesting research topic in its own right.

Information Extraction

Information extraction (IE) is the activity of extracting information about predefined classes of entities and relationships from natural language texts and placing this information into a structured representation or 'template'. Once extracted this information can be used for a variety of purposes, including text indexing, database construction and querying, data mining, and summarisation. Stimulated by the methodology of quantitative system evaluation pioneered by the U.S. DARPA Message Understanding Conference (MUC) open competitions, 1987-1998 this technology is now beginning to emerge into applications which are of significant interest to commercial, governmental and research organisations, where rapid access to specific information which confers competitive advantage is essential.

Within the context of these applications my interests are to develop and test computational language processing techniques. In particular I am interested in how conceptual and domain modelling can be used to support language processing activities such as parsing, semantic interpretation and coreference resolution, and I have published in these areas. I am also interesting in how to model and to extract temporal and spatial information for natural language texts.

Question Answering

There has been a growing wave of interest in "open domain question answering", to push information access technology beyond the document retrieval capabilities of current search engines, towards genuine ability to answer questions in any area given large text collections containing the answers. I have led a group from Sheffield to enter an international question answering competition, the first group in the UK to do so, and have had a PhD student work in the area as well as a research grant, the Electronic Cub Reporter project, from the EPRSC, together with Sheffield University Journalism Department and the Press Association, to explore this technology by building tools to assist journalists in answering questions when writing background pieces to breaking news stories.

Text Similarity

When journalists reuse news agency copy, a new text is created which is similar to but different from the source. Whether we can measure this difference was the focal question of the

METER (EPSRC ROPA grant) project and of significant practical concern to the Press Association, a collaborator on the project. A related question, in the general area of text similarity, is whether documents on similar topics can be automatically placed into a general-to-specific ordering to help, for example, in displaying search engine results to users. This was addressed in the EC-funded CLARITY project (joint with Information Studies), which aims to build a cross-language information retrieval system to help multi-language speakers in dealing with multiple language document collections.

Machine Translation

Modern machine translation (MT) is dominated by statistical approaches. However these approaches require very large amounts *parallel* text (i.e. texts that are translations of each other) in the source and target languages to work effectively. The required volumes of text are simply not available for less widely spoken languages (e.g. many of the languages of the EU) or in narrow domains. As a consequence researchers are exploring whether *comparable* texts, texts on the same topic which are not translations, such as concurrent news reports on the same news event, can be exploited by statistical MT. I am working in this area with colleagues in an EU consortium, where we are using our experience with IE techniques and with news texts to bring a novel approach to the exploitation of comparable corpora for MT.

Software Infrastructure for Natural Language Engineering

In 1993 I proposed that a general software infrastructure to support text engineering research was necessary to promote the sharing and reuse of algorithmic and data resources in the language engineering community. This proposal has lead, through two EPSRC-funded projects on which I have been co-investigator to the creation of GATE, the General Architecture for Text Engineering. GATE has served as a platform for my other text engineering research, and has also been used at other institutions in the UK, in Europe, and around the world, and was recently selected by DARPA in the U.S. to act as a final demonstrator platform at the end of its six year TIPSTER text programme. While less involved in this work than I once was, I retain an interest in offering web service or cloud-based access to language processing tools, for example biomedical terminology tools or terminology translation tools.

Future Plans

My research plan is to continue working primarily on information extraction,and on question answering, multidocument summarisation and temporal and spatial information processing, in a multilingual context. For IE the big challenges are to continue improving the accuracy of the underlying technology and to make systems which can be customised to specific domains by end users rather than by specialists. For question answering, the big challenges once again are to improve accuracy of the underlying technology and to develop application settings in which imperfect technology can still be used to the benefit of information seekers. I see this technology being developed both in the setting of journalistic practice, where I have an ideal collaboration opportunity with the Journalism department and their links to the Press Association. Of particular interest are questions of how interactive or dialogue-based information access tools can be developed so that information seekers can explore a topic through iterative queries where the system remembers what the user has seen

and has a model of the user's task, so that progressively richer reelvant and non-redundant content can be delivered to the user.

As the written expression of human culture becomes universally available via electronic text, there is no doubt that language technology will play an increasing pivotal role in defining how we interact with this material. My aim is to be at the forefront of these developments.

2.2   Research Grants and Contracts

Key

BBSRC – U.K. Biotechnology and Biological Sciences Research Council

CEC – Commission of the European Communities

CHIST-ERA – European Coordinated Research on Long-term Challenges in Information and Communication Sciences & Technologies ERA-Net

DARPA – U.S. Defense Advanced Projects Research Agency

DTI – U.K. Department of Trade and Industry

EPSRC – U.K. Engineering and Physical Sciences Research Council

MRC - UK Medical Research Council

ROPA – "Realising Our Potential Awards"

(For project descriptions see:
http://staffwww.dcs.shef.ac.uk/people/R.Gaizauskas/

University of Sheffield Centre for Machine Intelligence AI Research Engineering Project: "Text Correction for Historical Documents'. July 2023-??. Investigators: R.Gaizauskas (PI), M. Pidd, V. Vitale, R. Shoemaker Value: Support for one AI Research Engineer .

University of Sheffield Knowledge Exchange Fund: "A Multimodal Speech and Graphical Interface for Hands-free Data Capture and Querying in MRO". July 2019-March 2020. Grantholders: R.Gaizauskas (PI), E. Barker, J.P. Barker, N. Ma. Value: £85,009 .

EPRSC Grant: "Investigating Spoken Dialogue to Support Manufacturing Processes". April, 2017 – March 2018. Grantholders: R.Gaizauskas (PI), J. Law and E. Barker. Value: £63,502.

CEC 7th Framework Grant: "SENSEI: Making Sense of Human-Human Communication". November, 2013 – October 2016. Grantholders: R.Gaizauskas (PI), M. Hepple and J.Foster. Total Value: Euro ~2.65M of which Sheffield (Euro 546,466)   £428,282.

CHIST-ERA Grant: "Visual Sense (ViSen): Tagging visual data with semantic descriptions". January, 2013 – December 2015. Grantholders: R.Gaizauskas (PI) and P. Clough. Total Value: Euro ~1.3M of which Sheffield: £310,677.

EPSRC Grant: "Language Processing for Literature Based Discovery in Medicine". June, 2012 – May, 2015. Grantholders: Mark Stevenson (PI), R. Gaizauskas, O. Bandmann, S. Danson. Value: £293,127.

CEC 7th Framework Grant: "TAAS: Terminology as a Service". June, 2012 – May 2014. Sheffield grantholders: R. Gaizauskas (PI). Total value: Euro 1,820,000 of which Sheffield (Euro 335,040):   £267,477 .

CEC 7th Framework Grant: "Accurat: Analysis and Evaluation of Comparable Corpora for Under Resourced Areas of Machine Translation". January, 2010 – June 2012. Sheffield

grantholders: R. Gaizauskas (PI) and Mark Sanderson. Total value: Euro 3,100,000 of which Sheffield (Euro 428,000):   £385, 840.

CEC 6th Framework Grant: "Tripod: TRI-Partitite multimedia Object Description". January, 2007 – December 2009. Sheffield grantholders: Mark Sanderson (PI) and R. Gaizauskas. Total value: Euro 3,150,000 of which Sheffield (Euro 554,973):   £372,300.

EPSRC Grant: "Lexical Disambiguation for the Biomedical Domain". December, 2006 – November, 2009. Grantholders: Mark Stevenson (PI) and R. Gaizauskas.
Value: £239,921.

EPSRC Grant: "CRONOPATH: Timeline and Named Entity Extraction for Hyperlink Corpora". July, 2005 – December, 2007. Grantholders: Yorick Wilks (PI) and R. Gaizauskas.
Value: £294,632.

MRC Grant: "CLEF-Services". January 2005 – December 2007. Grantholders: A. Rector (PI), I. Buchan, C. Goble, M. Elliot, N. Zhang, (Manchester), R. Gaizauskas and M. Hepple (Sheffield), D. Ingram, D. Kalra, J. MacKay, M-A. Sasse (UCL), D. Scott (Brighton), J. Milan (Royal Marsden). Total value: £2,428,000 of which Sheffield: £381,000.

EPSRC Platform Grant: "The Epitheliome: Computational Modelling of Epithetial Tissue". August, 2004 – July, 2009. Grantholders: R. Smallwood (PI), W.M.L. Holcombe, E. Patterson, R. Gaizauskas, D.R. Hose, S. MacNeil, and R.H. Clayton (Sheffield), J. Southgate (York). Total value: £2,039,320. of which Sheffield   £1,596,474

EPSRC Grant: "Real-Time Text Mining for the Biomedical Literature: A Collaboration between DiscoveryNet and myGrid". October 2004 – September 2005. Grantholders: Y. Guo (Imperial) and R. Gaizauskas (Sheffield). Total value: £133,352 of which Sheffield:
£56,588

CEC 6th Framework Grant: "VIKEF: Virtual Information and Knowledge Environment Framework". April 2004 - March 2007. Sixteen partner FW6 Integrated Project. R. Gaizauskas (Sheffield grantholder). Total value: Euro 5,560,000 (£3,715,748) of which Sheffield: Euro 299,296 =   £200,020

EPSRC Grant: "Electronic Cub Reporter: Automatically Gathering and Collating Background Information from Digital Text". January 2003 - December 2005. R. Gaizauskas (PI), J. Foster, Y. Wilks. Total value: £307,972

MRC Grant: "CLEF: Clinical E-Science Framework". October 2002 - September 2005. Grantholders: A. Rector (PI), H. Heathfield, C. Taylor (Manchester), R. Gaizauskas and M. Hepple (Sheffield), D. Ingram and D. Kalra (UCL), D. Scott and R. Power (Brighton), J. Milan (Royal Marsden), D. Demeter (Judge Institute, Cambridge). Total value: £1,683,343 of which Sheffield: £280,725

EPSRC Grant: "MyGrid: Directly Supporting the E-Scientist". October 2001 - September 2004. Grantholders: C. Goble (Manchester – PI), T. Rodden (Nottingham), P. Watson (Newcastle), L. Moreau (Southampton), R. Gaizauskas (Sheffield). Total value: £3,275,014 of which Sheffield: £320,206

CEC 5th Framework Grant: "Clarity: Cross Lanuage Information Retrieval and Organisation of Text and Audio Document". February 1, 2001 - January 31, 2004. Contract obtained by R. Gaizauskas (PI) and M. Sanderson. Total value: Euro 1,775,000 ≈ £1,070,000 of which Sheffield: Euro 584,089 ≈   £351,952

EPSRC Grant: "EMILLE: Enabling Minority Language Engineering". June 2001 - September 2003. Grantholders: T. McEnery (PI), University of Lancaster, and R. Gaizauskas. Total value £241,000 of which Sheffield: £35,859

Glaxo-Wellcome Research: "TRESTLE: Text Retrieval, Extraction and Summarisation Technologies for Large Enterprises". June 1, 1999 - May 31, 2001. Contract obtained by R. Gaizauskas (PI) and P. Willett. Value: £303,000

EPRSC ROPA Grant "Measuring Text Reuse". April 1, 1999 – April 2001. Grantholder: R. Gaizauskas. Value: £181,000

EPSRC Grant "GATE 2: Software Infrastructure for Language Processing R & D". March 1999 - March 2001. Grantholders: Y. Wilks (PI) and R. Gaizauskas.   Value: £197,000

EPSRC Grant "Structured Transcription of Broadcast Speech". December 1998 - November 2001. Grantholders: S. Renals (PI) and R. Gaizauskas. Value: £238,000

CEC 4th Framework Grant "ELSE: Evaluation in Language and Speech Engineering". December, 1997 – May, 1999. Grantholders: R. Gaizauskas. Value: 26,000 ecu ≈   £16905

Glaxo-Wellcome Research/Elsevier Science: "EMPathIE: Automatic Extraction of Information concerning Enzymes and Metabolic Pathways from Scientific Journals". Contract obtained by R. Gaizauskas (PI) and P. Willett. Value: £60,000

BBSRC/EPRSC BioInformatics Programme "Use of Text Extraction Techniques to Create a Database of Protein Active Sites". October, 1997 – October 2000. Grantholders: R. Gaizauskas (PI), Y. Wilks, P. Artymiuk and P. Willett. Value: £142,000

CEC 4th Framework Grant "ECRAN – Extraction of Content: Research at Near Market" October, 1994 - September 1997. Grantholders: Y. Wilks (PI) and R. Gaizauskas. Value: £300,000

EPSRC Grant "Large Scale Information Extraction from Text", October, 1994 - September 1997. Grantholders: Y. Wilks (PI) and R. Gaizauskas. Value: £300,000

DTI Funding for Participation in the Sixth Message Understanding System Evaluation and Message Understanding Conference (MUC-6), September, 1994 - May, 1995. Grantholders: R. Gaizauskas. Value: £10,000

DARPA Funding for Participation in the Fifth Message Understanding System Evaluation and Message Understanding Conference (MUC-5), March - August, 1993. Grantholders: R. Gaizauskas, R. Evans, and L. Cahill. Value: US $20,000 ≈   £15,000

2.3   Publications (Summary)

Books and edited collections: 2 – see section 6.1.

Articles in refereed journals: 27 – see section 6.2.

Papers in refereed conference proceedings: 100 – see section 6.3.

Papers in refereed workshop proceedings: 63 – see section 6.4.

Posters (short papers) in refereed conference proceedings: 4 – see section 6.5.

Chapters in edited collections: 8 – see section 6.6.

Papers in unrefereed conference and workshop proceedings: 14 – see section 6.7.

Technical reports: 18 – see section 6.8.

Book reviews: 4 – see section 6.10.

3 Teaching

3.1 Most Recent Undergraduate and Post Graduate Teaching

Lecturing

I am involved preparing and presenting lectures, setting and assessing practical work, and setting and assessing examinations.

Introduction to Machines and Intelligence. 1st year Computer Science module, Sheffield University: 350 students (2023-24); taught 2013-2024

Text Processing. 3rd year, 4th year and MSc Computer Science module, Sheffield University; 150 students (2023-24); taught 2001-2003; 2016-2024

Tutoring

I do/have done small group personal and academic tutoring for

1st and 2nd year Computer Science/Software Engineering/Artificial Intelligence courses at Sheffield (1993-)

Taught MSc courses Sheffield (1993-)

Project Supervision

I have supervised:

3rd year projects – 2-4 annually since 1993 at Sheffield in Computer Science, Artificial Intelligence, Software Engineering.

4th year/Advanced MSc Darwin Research Project (2006-2008) – a group advanced research project competitively selected by students from a pool of topics proposed by potential supervisors

Msc dissertations – 2-5 annually since 1993 at Sheffield in Software Systems Technology, Advanced Software Engineering and Information Systems, Advanced Computer Science, Speech and Language Technology

3.2 Previous Undergraduate and Post Graduate Teaching

Lecturing

I have been sole or principal lecturer on the following courses. In each case this has involved preparing and presenting lectures, setting and assessing practical work, and setting and assessing examinations.

Experiencing Genesys. 3rd year Information Technology and Business Management module, Sheffield University: 10-15 students; 2012-2013.

Introduction to Machine Learning. 3rd year and MSc Computer Science module, Sheffield University: 20-70 students; 2001-2010.

Advanced Research Topic: Relation Extraction from Text. MSc in Human Language Technology; 4 students; 2008.

Advanced Research Topic: Current Topics in NLP. MSc in Human Language Technology; 4 students; 2008.

Machines and Languages (automata and formal language theory). Core 2nd year Computer Science module, Sheffield University; 70-100 students; 1993-2005.

Natural Language Processing. 3rd year Computer Science module, Sheffield University: 18-50 students; 1994-2014.

Formal Methods in Software Engineering (Z). MSc in Software Systems Technology module, Sheffield University; 60-70 students; 1994-95.

Practical Computing. 1st year Computer Science module for subsidiary students, Sheffield University; 200 students; 1993.

Advanced Artificial Intelligence (automated reasoning). 3rd year Cognitive Science module, Sussex University; 25 students; 1989.

Prolog Programming. MSc in Knowledge-based Systems module, Sussex University; 30 students; 1989.

Formal Logic. 3rd year Philosophy module, Carleton University; 25 students; 1978-79.

Tutoring

As in current teaching above, plus tutoring for 1st year Introduction to Computers and AI module at Sussex (1989).

Project Supervision

As in current teaching above, plus supervising 4 MSc dissertations in 1989 at Sussex in Knowledge-based Systems.

3.3   Research Supervision

PhD thesis supervision:

- primary supervisor for 7 PhDs in progress and for 14 successfully completed PhDs completed PhDs

- secondary supervisor for over 30 current and successfully completed PhDs

3.4   Teaching Innovation and Development

- In 2013-14 I completely redesigned the content and assessment of the 1st semester of the 1st year Machines and Intelligence.

- In 2009 I bid for and was awarded an Engineering Faculty Learning and Teaching grant to support the development of a set of novel lab classes for my 3rd year/MSc Natural Language Processing module.

- I have proposed, had adopted into the syllabus, developed and delivered two completely new 3rd year modules: *Text Processing* and *Introduction to Machine Learning*

- I have completely redesigned the NLP and Theory of Machines modules I have taught at Sheffield (my materials for the later are still being used 3 years after I turned over the course).

- I have participated in the creation and specification of an AI degree course at Sheffield.

- I served on a curriculum redesign committee for the Sheffield CS department Masters degrees which rationalised the degrees offerred and their contents.

3.5 External Examining

PhD thesis examiner for Universities of Cambridge, Dundee, Durham, Edinburgh, Helsinki, Lancaster, Leeds, London (Kings, Birkbeck), Manchester, Middlesex, the Open University, Stockholm, Surrey, Sussex, Technical University of Catalonia, Barcelona and UNED, Madrid (30+ PhDs).

MPhil or MSc by research thesis examiner for UMIST, University College (Dublin), and Nanyang Technological University, Singapore.

MSc Course Examiner for UMIST, (MSc in Machine Translation and MSc in Natural Language Processing), 2001-2003.

MSc Course Examiner for University of Sussex, (MSc in Intelligent Systems), 2009-2013.

Tripos Part III and MPhil Course Examiner in Computer Science for the University of Cambridge, 2013-2015.

4   Administration

4.1   Current Administrative Activities

Co-Director UKRI Centre for Doctoral Training (CDT) in Speech and Language Technologies and Their Applications. Duties, shared with the other co-director and the Centre manager, include: (1) recruiting new PhD candidates (advertising, interviewing, selecting); (2) designing, reviewing and delivering the CDT training programme across 4 years of the PhD cycle for each of 5 cohorts of students, including setting and marking assignments for the PGDip progamme and participating in annual exam board to determine progression; (3) identifying and engaging industrial partners to sponsor PhD studentships in the CDT; (4) organising and delivering an annual CDT conference; (5) participating in the day-to-day management of the CDT and in other management bodies such as the CDT content planning group and the CDT advisory board; (6) complying with the reporting requirements laid down by our funders, UKRI; (7) managing the CDT budget.

4.2   Previous Administrative Activities

Departmental Director of Research and Innovation (2014-2018) Duties include organizing and chairing the Department's Research Committee, membership on the Department's executive and strategy committees, membership on Faculty Research and Innovation Committee (FRICOM), attendance at relevant Faculty and University level meetings as necessary, preparation of Departmental research strategy plans and research reports for the University and outside world, maintaining a record of the Department's research profile (grants, publications, progress of graduate students), organization of the annual Departmental Research Retreat, participation in academic staff recruitment (long and short listing, interviewing, appointment panels).

Head of the Natural Language Processing (NLP) Research Group. Duties include offerring leadership and vision for the NLP group, the largest research group in the Department, acting as primary point of contact for external bodies interested in the research group activities and managing the external face of the group, representing the group in the departmental Research Committee, chairing regular research group meetings.

REF Co-ordinator (2012-2013) Duties included co-authoring and editing the Departmental REF return documents with HoD and DDRI, interviewing all member of department and helping them select their best publications and author their 100 word statements about them, interviewing members of the department contributing impact case studies and helping them author these, attending numerous Engineering Faculty REF meetings and critically reading and commenting on returns from other departments.

Member University REF Equality Advisory Panel (2012-2013) Duties included considering, together with other members of the panel, the cases of individual academics seeking a reduction in the normal quota of papers to be returned for REF due to personal circumstances and deciding, by interpreting HEFCE guidance, whether to recommend accepting the request for reduction or not.

Director of Research and Deputy Head of Department (2008-2011) Duties included organizing and chairing the Department's Research Committee, membership on the Department's executive and strategy committees, representing the Department at Faculty and University level meetings as necessary, preparftion of Departmental research strategy plans and research reports for the University and outside world, maintaining a record of the Department's research profile (grants, publications, progress of graduate students), organization of the annual Departmental Research Retreat, participation in academic staff recruitment (long and short listing, interviewing, appointment panels).

Member Departmental Staffing Committee (by election) (2008-2011). Duties included considering cases for promotions and sabbaticals and acting as the Departmental Staff Review and Development Review Panel.

Year 2 Tutor (2007 - 2008) Duties included monitoring second year modules, espccially assessment loads and scheduling, acting as fallback personal tutor for all second years, attending teaching committee and staff-student liaison committee meetings, particpating in the curriculum review process.

Examinations Officer (2001 - 2007) Duties included overseeing the examinations process, chairing examination board meetings, attending faculty examinations meetings.

Publicity Officer (2000 - 2002) Duties included co-ordinating departmental publicity efforts, monitoring publicity materials to ensure they are up-to-date.

Engineering Faculty Board Representative (1999-2001) Duties included representing the Department at Engineering Faculty Board meetings.

3rd Year/MSc Projects Officer (1999-2000) Duties included overseeing the process of allocating 3rd year and MSc projects to students.

Deputy Research Director (1996-1999) Duties included attending Departmental Research Committee meetings and taking over for the Departmental Director of Research when absent.

Deputy Admissions Tutor (1993-1995) Duties included handling UCAS forms for overseas and foundation year students, and standing in for/assisting the Admissions Tutor in a variety of ways.

Web Officer (1993-1995) Duties included initiating, designing and overseeing the implementation of the first Departmental website.

4.3 Innovations in Administration

- Recommended and oversaw the creation of a Departmental Impact Register, implemented as a Google Form, for members of the Department to record instances of non-academic "impact", as defined in REF 2013, to ease the collection of data for REF 2013, but also to anticipate and reduce the last minute effort of collecting of data for future REF exercises.

- Proposed and implemented various enhancements to the Departmental examinations procedures, including a streamlined internal exams board procedure, web-based annual calendar of exams related activities and web-based delivery of marks to students.

- Responsible for introduction of a more efficient, transparent and fair method for assigning projects to 3rd year and MSc students.

- Responsible for the movement of the MSc project allocation and management to the Web.

- Responsible for overseeing the process of creating the first Departmental website, and suggestions that much Departmental routine business (forms distribution, publicity, etc) could be moved to the Web.

5   Professional Activities

5.1   Grant Reviewing

Member of the Engineering and Physical Sciences Research Council (EPSRC) Peer Review College, 2000-present.

Member Biotechnology and Biological Sciences Research Council (BBSRC) Bioinformatics E-Science Panel, 2001-2002.

Grant proposal reviewer for the EPSRC, 1996 - present.

Grant proposal reviewer for the BBRSC, 1999 - present.

Grant proposal evaluator for the EC Framework V, Information, Society and Technology (IST) programme.

5.2   Project Reviewing/Evaluation

Member of the Economic and Social Research Council (ESRC) Evaluation Panel of the Human Communication Research Centre (HCRC), Universities of Edinburgh and Glasgow, 1997.

Project Review Auditor within European Commission Language Research and Engineering programme, 1995 and 1996.

5.3   Journal/Conference Reviewing

Member of the editorial board of the journal *Corpora*, 2005-2010.

Reviewer for the journals of *Computational Linguistics*, *Transactions of the Association for Computational Linguistics*, *Natural Language Engineering*, *Language Resources and Evaluation*, *Traitement Automatique des Langues*, *Artificial Intelligence Research*, *BioInformatics*, *Biomedical Informatics*, *BMC Bioinformatics*, *Medical Informatics*, *Computers and the Humanities*, *Literary and Linguistic Computing*, *IEEE Computer Magazine*, *Lingvisticae Investigationes*, *ACM Transactions on Asian Language Information Processing*,*Information Research: an international electronic journal*.

Reviewer for the Association for Computational Linguistics Conference (2000-2014); the Applied Natural Language Processing Conference (2000); Conference on Empirical Methods in Natural Language Processing (EMNLP) (2003, 2008); the European Chapter of the Association for Computational Linguistics (EACL) (1999,2001,2006); the European Conference on Information Retrieval (ECIR), (2004); the International Joint Conference on Artificial Intelligence (IJCAI) (1999); European Conference on Artificial Intelligence (ECAI) (2002, 2012); the International Conference on Computational Linguistics (COLING) (1994, 1996, 1998, 2000, 2002,2006,2008, 2012); the 8th IEEE International Conference on Tools with Artificial Intelligence (ICTAI96) (1996); the International Conference on Language Resources and Evaluation (LREC) (1998,2000,2002,2008, 2010, 2012, 2014); the North American Chapter of the Association for Computational Linguistics and Human Language Technologies conference (NAACL-HLT) (2006,2009);

Recent Advances in Natural Language Processing (RANLP) (2005,2007,2009); the ACM Special Interest Group on Information Retrieval (SIGIR) (2001,2002,2003,2009), Symposium on Semantic Mining in Biomedicine (2005,2006), amongst others.

Area Chair for: the 2008 International Conference on Computational Linguistics (COLING) for the area: "Information Extraction and Question Answering"; the 2010 the Association for Computational Linguistics Conference for the area: "Text Mining and NLP on Noisy Unstructured Text".

Conference programme committee member for the 22nd International Conference on Computational Linguistics (COLING 2008), 26th European Conference on Information Retrieval (ECIR04), the 25th ACM Special Interest Group on Information Retrieval (SIGIR02), First Human Language Technology Conference (HLT01), 9th and 11th Conferences of the European Chapter of the Association for Computational Linguistics (EACL'99 and EACL-06)) and the DARPA Seventh Message Understanding Conference (MUC-7).

5.4   Conference, Workshop and Tutorial Organisation

Co-proposer, organiser, and programme committee member of the TempEval track in the 4th International Workshop on Semantic Evaluations held in conjunction with the 45th Annual Meeting of the Association for Computational Linguistics (ACL), Prague, June, 2007.

Organiser and programme committee member of the Workshop on Ontologies and Text Processing held in conjunction with the 4th European Conference on Computational Biology (ECCB05), Madrid, September, 2005.

Co-proposer, organiser, and programme committee member of the Workshop on Information Retrieval for Question Answering held in conjunction with the 27th ACM Special Interest Group on Information Retrieval Conference (SIGIR04), Sheffield, July, 2004.

Co-proposer, organiser, and programme committee member of the Workshop on Annotation Standards for Temporal Information in Natural Language held in conjunction with the Third International Conference on Language Resources and Evaluation (LREC02), Las Palmas, Spain, May, 2002.

Co-proposer, organiser, and programme committee member of the Workshop on the Evaluation of Parsing Systems held in conjunction with the First International Conference on Language Resources and Evaluation (LREC98), Granada, Spain, May, 1998.

Tutorial Lecturer for a course on Information Extraction at the 9th European Summer School in Logic, Language and Information (ESSLLI97) at Aix-en-Provence, France, August, 1997.

Seminar Co-Chair (with Y. Wilks) for a two-day commercial seminar *Natural Language Processing: Extracting Information for Business Needs* organised by Unicom Seminars Ltd., London, 1997.

Workshop Chair and Organiser for the U.K. Speech and Language Technology (SALT) Club Workshop *Evaluation in Speech and Language Technology*, Sheffield, June 1997.

Workshop/Tutorial Chair for the Tenth Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB-95), Sheffield, April 1995.

5.5   Invited Talks

I have delivered over 40 invited presentations (exclusive of conference/workshop presentations) since 1993 to universities, research institutes and companies in the UK, Europe, the US, and Canada, including the Universities of Aberdeen, Brighton, Edinburgh, Essex, Glasgow, London (QMC), Oxford, Manchester, Sussex, York, Nancy (in French), New Mexico State, Montreal, Toronto, Dublin City University, Glaxo-Wellcome Research and Sharp Research Laboratories, and the National Research Council of Canada.

Keynote Presentations include:

- *Generating Image Captions using Topic Focused Multi-document Summarization* presented at The 22nd International Conference on Computational Linguistics Workshop on Multi-source, Multilingual Information Extraction and Summarization (MMIES-2), held in Manchester, UK in August 2008.

- *Evaluating Language Processing Applications and Components* presented at the 6th International Workshop on Computational Processing of the Portuguese Language, PROPOR 2003, held in Faro, Portugal in June 2003.

5.6   Public Engagement

I have given talks to the Café Scientifique in Sheffield (2009) and York (2010).

5.7   Membership on Professional Committees

Member Arts and Humanities Research Council e-Science Initiative Steering Committee, 2006-2007.

Member of the US NIST/TREC Question Answering Roadmap Committee, planning the future of the TREC Question Answering Track Evaluations, 2000.

Member European Commission Expert Advisory Group on Language Engineering Standards (EAGLES) subgroup on Lexical Semantics, 1997-98.

Elected national Academic Representative for Language on the UK Speech and Language Technology (SALT) Club Steering Committee, 1996-1998.

Rapporteur for European Commission Study Group on Assessment and Evaluation in Language Engineering, 1995.

5.8   Membership in Professional Organisations

Member of the Association for Computational Linguistics (ACL)

Member of the American Association for Artificial Intelligence (AAAI)

6   Publications

6.1   Books

Skadiņa, I., Gaizauskas, R., Babych, B., Ljubešić, N., Tufiş, D. and Tufiş, D. (ed). Using Comparable Corpora for Under-Resourced Areas of Machine Translation. Springer Cham, 2019.

Mani, I., Pustejovsky, J. and Gaizauskas, R. (ed). The Language of Time: A Reader. Oxford University Press, 2005.

6.2   Articles in Refereed Journals

Tait, J., Gaizauskas, R. and Bontcheva, K. Obituary: Yorick Wilks. *Computational Linguistics*, 49(3), 767-772, 2023.

Tang, Y., Wang, J.K., Wang, X., Gao, B., Dellandréa, E., Gaizauskas, R. and Chen, L. Visual and semantic knowledge transfer for large scale semi-supervised object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, November, 2017.

Gaizauskas, R., Paramita, M., Barker, E., Barker, E., Aker, A. and Pahisa, M. Extracting bilingual terms from the Web. *Terminology*, 21(2), 205-236, 2015.

Preiss, J., Stevenson, M. and Gaizauskas, R. Exploring Relation Types for Literature-based Discovery. *Journal of the American Medical Informatics Association*, 22(5), 987-92, 2015.

Aker, A. and Gaizauskas, R. Generating descriptive multi-document summaries of geo-located entities using entity type models. *Journal of the Association for Information Science and Technology*, 66(4), 721-738, 2015.

Di Fabbrizio , G., Aker, A. and Gaizauskas, R. Summarizing Online Reviews Using Aspect Rating Distributions and Language Modeling. *IEEE Intelligent Systems*, 28(3), 28–37, May, 2013.

Aker, A., Plaza, L., Lloret, E. and Gaizauskas, R. J. Do humans have conceptual models about geographic objects? A user study. *Journal of the Association for Information Science and Technology*, 64(4), 689–700, 2013.

Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Moszkowicz, J. and Pustejovsky, J. The TempEval Challenge: Identifying Temporal Relations in Text. *Journal of Language Resources and Evaluation*, 43(2), 161-179, 2009.

Roberts, A., Gaizauskas, R., Hepple, M., Demetriou, G., Guo, Y., Roberts, I. and Setzer, A. Building a semantically annotated corpus of clinical texts. *Journal of Biomedical Informatics*, 42(5), 950-966, 2009.

Roberts, A., Gaizauskas, R., Hepple, M. and Guo, Y. Mining clinical relationships from patient narratives. *BMC Bioinformatics*, 9(Suppl 11), S3, 2008.

Stevenson, M., Guo, Y., Gaizauskas, R. and Martinez, D. Disambiguation of Biomedical Text using Diverse Sources of Information. *BMC Bioinformatics*, 9(Suppl 11), S7, 2008.

Setzer, A., Gaizauskas, R. and Hepple, M. The Role of Inference in the Temporal Annotation and Analysis of Text. *Journal of Language Resources and Evaluation*, 39(2-3), 243-265, 2006.

Davis, N., Demetriou, G., Gaizauskas, R., Guo, Y. and Roberts, I. Web Service Architectures for Text Mining: An Exploration of the Issues via an E-Science Demonstrator. *International Journal of Web Services Research, Vol. 3, Issue 4, 2006*, 3(4), 2006.

Harkema, H., Roberts, I., Gaizauskas, R. and Hepple, M. A Web Service for Biomedical Term Look-up. *Comparative and Functional Genomics*, 6(1-2), 86-93, 2005.

Baker, P., Bontcheva, K., Cunningham, H., Gaizauskas, R., Hamza, O., Hardie, A., Jayaram, B.D., Leisher, M., McEnery, A.M., Maynard, D., Tablan, V., Ursu, C. and Xiao, Z. Corpus linguistics and South Asian languages: Corpus creation and tool development. *Literary and Linguistic Computing*, 19(4), 2004.

Gaizauskas, R., Demetriou, G., Artymiuk, P. and Willett, P. Protein Structures and Information Extraction from Biological Texts: The PASTA System. *Journal of Bioinformatics*, 19(1), 135-143, 2003.

Harmain, H. M. and Gaizauskas, R. CM-Builder: A Natural Language-based CASE Tool. *Journal of Automated Software Engineering*, 10, 157-181, 2003.

Hirschman, L. and Gaizauskas, R. Natural Language Question Answering: The View from Here. *Journal of Natural Language Engineering*, 7(4), 275-300, 2001.

Gaizauskas, R., Rodgers, P. J. and Humphreys, K. Visual Tools for Natural Language Processing. *Journal of Visual Languages and Computing*, 12(4), 351-373, 2001. Online journal version at: `http://www.idealibrary.com/links/toc/jvlc/12/4/0`.

Humphreys, K., Demetriou, G. and Gaizauskas, R. Bioinformatics Applications of Information Extraction from Journal Articles. *Journal of Information Science*, 26(2), 75-85, 2000.

McEnery, A.M., Baker, P., Gaizauskas, R. and Cunningham, H. EMILLE: Building a Corpus of South Asian Languages. *Vivek, A Quarterly in Artificial Intelligence*, 13(3), 23-32, 2000.

Azzam, S., Humphreys, K., Gaizauskas, R. and Wilks, Y. Using a Language Independent Domain Model for Multilingual Information Extraction. *Journal of Applied Artificial Intelligence*, 13(7), 705-724, 1999.

Krotov, A., Hepple, M., Gaizauskas, R. and Wilks, Y. Evaluating Two Methods for Treebank Grammar Compaction. *Journal of Natural Language Engineering*, 5(4), 377-394, 1999.

Gaizauskas, R. and Wilks, Y. Information Extraction: Beyond Document Retrieval. *Journal of Documentation*, 54(1), 70-105, 1998.

Gaizauskas, R. Evaluation in language and speech technology. *Journal of Computer Speech and Language*, 12(3), 249-262, 1998.

Gaizauskas, R. and Humphreys, K. Using a Semantic Network for Information Extraction. *Journal of Natural Language Engineering*, 3(2/3), 147-169, 1997.

Evans, R., Gaizauskas, R., Cahill, L.J., Walker, J., Richardson, J. and Dixon, A. POETIC: A System for Gathering and Disseminating Traffic Information. *Journal of Natural Language Engineering*, 1(4), 363-387, 1995.

6.3 Papers in Refereed Conference Proceedings

Booth, C., Thomas, A. and Gaizauskas, R. BLN600: A Parallel Corpus of Machine/Human Transcribed Nineteenth Century Newspaper Texts. In *Proceedings of the Language Resources and Evaluation Conference*, European Language Resources Association, Turin, Italy, May, 2024.

Barker, E., Barker, J., Gaizauskas, R., Ma, N. and Lestari Paramita, M. SNuC: The Sheffield Numbers Spoken Language Corpus. In *Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022)*, 2022.

Booth, C., Shoemaker, R. and Gaizauskas, R. A Language Modelling Approach to Quality Assessment of OCR'ed Historical Text. In *Proceedings of the Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, 5859–5864, June, 2022.

Funk, A., Aker, A., Barker, E., Lestari Paramita, M., Hepple, M. and Gaizauskas, R. The SENSEI overview of newspaper readers' comments. In *Proceedings of the Thirty-Ninth European Conference on Information Retrieval (ECIR'17)*, 758-761, April, 2017.

Lestari Paramita, M., Clough, P., Aker, A. and Gaizauskas, R. Using section headings to compute cross-lingual similarity of wikipedia articles. In *Proceedings of the Thirty-Ninth European Conference on Information Retrieval (ECIR'17)*, 633-639, April, 2017.

Wang, J. and Gaizauskas, R. Cross-validating Image Description Datasets and Evaluation Metrics. In Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Thierry Declerck and Marko Grobelnik and Bente Maegaard and Joseph Mariani and Asuncion Moreno and Jan Odijk and Stelios Piperidis (ed.) *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, European Language Resources Association (ELRA), Paris, France, 3059–3066, may, 2016.

Tang, Y., Wang, J., Gao, B., Dellandrea, E., Gaizauskas, R. and Chen, L. Large Scale Semi-supervised Object Detection using Visual and Semantic Knowledge Transfer. In *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition (CVPR 2016)*, June, 2016.

Barker, E., Paramita, M., Funk, A., Kurtic, E., Aker, A., Foster, J., Hepple, M. and Gaizauskas, R. What's the Issue Here?: Task-based Evaluation of Reader Comment Summarization Systems. In Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Thierry Declerck and Marko Grobelnik and Bente Maegaard and Joseph Mariani and Asuncion Moreno and Jan Odijk and Stelios Piperidis (ed.) *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, European Language Resources Association (ELRA), Paris, France, may, 2016.

Aker, A., Kurtic, E., Rajendran, B. A., Lestari Paramita, M., Barker, E., Hepple, M. and Gaizauskas, R. A Graph-based Approach to Topic Clustering for Online Comments to News. In *Proceedings of the 38th European Conference on Information Retrieval (ECIR 2016)*, 2016.

Funk, A., Gaizauskas, R. and Favre, B. A Document Repository for Social Media and Speech Conversations. In Nicoletta Calzolari (Conference Chair) and Khalid Choukri and

Thierry Declerck and Marko Grobelnik and Bente Maegaard and Joseph Mariani and Asuncion Moreno and Jan Odijk and Stelios Piperidis (ed.) *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, European Language Resources Association (ELRA), Paris, France, may, 2016.

Wang, J. and Gaizauskas, R. Don't Mention the Shoe! A Learning to Rank Approach to Content Selection for Image Description Generation. In *Proceedings of the 9th International Natural Language Generation Conference (INLG16)*, 2016.

Aker, A., Lestari Paramita, M., Kurtic, E., Funk, A., Barker, E., Hepple, M. and Gaizauskas, R. Automatic label generation for news comment clusters. In *Proceedings of the 9th International Natural Language Generation Conference (INLG16)*, 2016.

Barker, E., Paramita, M., Aker, A., Kurtic, E., Hepple, M. and Gaizauskas, R. The SENSEI Annotated Corpus: Human Summaries of Reader Comment Conversations in On-line News. In *Proceedings of the 17th Annual SIGdial Meeting on Discourse and Dialogue (SIG-DIAL 2016)*, 2016.

Ramisa, A., Wang, J., Lu, Y., Dellandrea, E., Moreno-Noguer, F. and Gaizauskas, R. Combining Geometric, Textual and Visual Features for Predicting Prepositions in Image Descriptions. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Lisbon, Portugal, 214–220, September, 2015.

Aker, A., Kurtic, E., Hepple, M., Gaizauskas, R. and Di Fabbrizio, G. Comment-to-Article Linking in the Online News Domain. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2015)*, 245-249, 2015.

Derczynski, L. and Gaizauskas, R. Temporal Relation Classification using a Model of Tense and Aspect. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2015*, Hissar, Bulgaria, 118-122, 2015.

Di Fabbrizio, G., Stent, A. and Gaizauskas, R. A Hybrid Approach to Multi-document Summarization of Opinions in Reviews. In *Proceedings of the 8th International Natural Language Generation Conference (INLG)*, Association for Computational Linguistics, Philadelphia, Pennsylvania, U.S.A., 54–63, June, 2014.

Aker, A., Lestari Paramita, M., Pinnis, M. and Gaizauskas, R. J. Bilingual dictionaries for all EU languages. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 26-31, 2014.*, 2839–2845, 2014.

Alhelbawy, A. and Gaizauskas, R. Collective Named Entity Disambiguation using Graph Ranking and Clique Partitioning Approaches. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 1544–1555, August, 2014.

Alhelbawy, A. and Gaizauskas, R. Graph Ranking for Collective Named Entity Disambiguation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Association for Computational Linguistics, Baltimore, Maryland, 75–80, June, 2014.

Derczynski, L. and Gaizauskas, R. Temporal Signals Help Label Temporal Relations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Association for Computational Linguistics, Sofia, Bulgaria, 645–650, August, 2013.

Derczynski , L. and Gaizauskas, R. Empirical Validation of Reichenbach's Tense Framework. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, 71–82, 2013.

Aker, A., Paramita, M. and Gaizauskas, R. Extracting bilingual terminologies from comparable corpora. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Sofia, Bulgaria, 402–411, August, 2013.

Alhelbawy, A. and Gaizauskas, R. Named Entity Disambiguation Using HMMs. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 03*, WI-IAT '13, IEEE Computer Society, Washington, DC, USA, 159–162, 2013.

Derczynski, L. and Gaizauskas, R. Information Retrieval for Temporal Bounding. In *Proceedings of the 2013 Conference on the Theory of Information Retrieval*, ICTIR '13, ACM, New York, NY, USA, 29:129–29:130, 2013.

Skadiņa, I., Aker, A., Mastropavlos, N., Su, F., Tufiņ, D., Verlic, M., Vasiļjevs, A., Babych, B., Clough, P., Gaizauskas, R., Glaros, N., Lestari Paramita, M. and Pinnis, M. Collecting and Using Comparable Corpora for Statistical Machine Translation. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May, 2012.

Barker, E. and Gaizauskas, R. Assessing the Comparability of News Texts. In N. Calzolari (Conference Chair) and K. Choukri and T. Declerck and M. Uğur Doğan and B. Maegaard and J. Mariani and J. Odijk and S. Piperidis (ed.) *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), Istanbul, Turkey, may, 2012.

Aker, A., Fan, X., Sanderson, M. and Gaizauskas, R. Investigating summarization techniques for geo-tagged image indexing. In *Proceedings of the 34th European Conference on Information Retrieval (ECIR12)*, Barcelona, apr, 2012.

Llorens, H., Derczynski, L., Gaizauskas, R. and Saquete, E. TIMEN: An Open Temporal Expression Normalisation Resource. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May, 2012.

Aker, A., Kanoulas, E. and Gaizauskas, R. A light way to collect comparable corpora from the Web. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC2012)*, Istanbul, may, 2012.

Aker, A., Feng, Y. and Gaizauskas, R. Automatic bilingual phrase extraction from comparable corpora. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, Mumbai, dec, 2012.

Lestari Paramita, M., Clough, P., Aker, A. and Gaizauskas, R. Correlation between Similarity Measures for Inter-Language Linked Wikipedia Articles. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May, 2012.

Alhelbawy, A. and Gaizauskas, R. Named Entity Based Document Similarity with SVM-Based Re-ranking for Entity Linking. In *Proceedings of First International Conference on Advanced Machine Learning Technologies and Applications (AMLTA12)*, Mumbai, dec, 2012.

Llorens, H., Saquete, E., Saquete, E. and Gaizauskas, R. Time-Surfer: Time-based Graphical Access to Document Content. In Advances in Information Retrieval (Clough, Paul and Foley, Colum and Gurrin, Cathal and Jones, Gareth and Kraaij, Wessel and Lee, Hyowon and Mudoch, Vanessa,ed.), Springer Berlin / Heidelberg, 767-771, 2011.

Derczynski, L. and Gaizauskas, R. A Corpus-based Study of Temporal Signals. In Nicholas Groom and Oliver Mason (ed.) *Proceedings of Corpus Linguistics 2011 (CL2011)*, Birmingham, UK, July, 2011.

Aswani, N. and Gaizauskas, R. English-Hindi Transliteration using Multiple Similarity Metrics. In Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Bente Maegaard and Joseph Mariani and Jan Odijk and Stelios Piperidis and Mike Rosner and Daniel Tapias (ed.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta, may, 2010.

Aker, A., Cohn, T. and Gaizauskas, R. Multi-document summarization using A* search and discriminative training. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, 482–491, 2010.

Catizone, R., Dingli, A. and Gaizauskas, R. Using Dialogue Corpora to Extend Information Extraction Patterns for Natural Language Understanding of Dialogue. In Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Bente Maegaard and Joseph Mariani and Jan Odijk and Stelios Piperidis and Mike Rosner and Daniel Tapias (ed.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta, may, 2010.

Aker, A. and Gaizauskas, R. Model Summaries for Location-related Images. In Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Bente Maegaard and Joseph Mariani and Jan Odijk and Stelios Piperidis and Mike Rosner and Daniel Tapias (ed.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta, may, 2010.

Derczynski, L. and Gaizauskas, R. Using Signals to Improve Automatic Classification of Temporal Relations. In *Proceedings of the 15th ESSLLI Student Session*, 2010.

Fan, X., Aker, A., Tomko, M., Smart, P., Sanderson, M. and Gaizauskas, R. Automatic Image Captioning From the Web For GPS Photographs. In *Proceedings of the 11th ACM SIGMM International Conference on Multimedia Information Retrieval (MIR 2010)*, 2010.

Aswani, N. and Gaizauskas, R. Developing Morphological Analysers for South Asian Languages: Experimenting with the Hindi and Gujarati Languages. In Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Bente Maegaard and Joseph Mariani and Jan Odijk and Stelios Piperidis and Mike Rosner and Daniel Tapias (ed.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta, may, 2010.

Skadiņa, I., Aker, A., Giouli, V., Tufiş, D., Gaizauskas, R., Mieriņa, M. and Mastropavlos, N. A Collection of Comparable Corpora for Under-resourced Languages. In *Proceeding of the 2010 conference on Human Language Technologies – The Baltic Perspective: Proceedings of the Fourth International Conference Baltic HLT 2010*, IOS Press, Amsterdam, The Netherlands, The Netherlands, 161–168, 2010.

Derczynski, L. and Gaizauskas, R. Analysing Temporally Annotated Corpora with CAVaT. In Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Bente Maegaard and Joseph Mariani and Jan Odijk and Stelios Piperidis and Mike Rosner and Daniel Tapias (ed.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta, may, 2010.

Aker, A. and Gaizauskas, R. Generating image descriptions using dependency relational patterns. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, 1250–1258, 2010.

Aker, A. and Gaizauskas, R. Summary Generation for Toponym-Referenced Images using Object Type Language Models. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP-2009)*, 2009.

Stevenson, M., Alamri, A., Guo, Y. and Gaizauskas, R. A Corpus of Biomedical Abbreviations. In *Proceedings of Corpus Linguistics 2009*, Liverpool, UK, 2009.

Demetriou , G., Gaizauskas, R., Sun, H. and Roberts, A. ANNALIST: ANNotation ALIgnment and Scoring Tool. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, 2008.

Roberts, A., Gaizauskas, R., Hepple, M. and Guo, Y. Combining terminology resources and statistical methods for entity recognition: an evaluation. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, 2008.

Stevenson, M., Guo, Y. and Gaizauskas, R. Acquiring Sense Tagged Examples using Relevance Feedback. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING2008)*, Manchester, August, 2008.

Roberts, A., Gaizauskas, R., Hepple, M., Davis, N., Demetriou, G., Guo, Y., Kola, J., Roberts, I., Setzer, A., Tapuria, A. and Wheeldin, B. The CLEF Corpus: Semantic Annotation of Clinical Text. In J. M. Teich and J. Suermondt and G. Hripcsak (ed.) *American Medical Informatics Association 2007 Proceedings. Biomedical and Health Informatics: From Foundations to Applications to Policy*, Chicago, IL, USA, 625–629, November, 2007.

Saggion, H. and Gaizauskas, R. Experiments on Passage Selection and Answer Identification for Question Answering. In *Lecture Notes in Artificial Intelligence (FinTAL 2006 Vol. 4139)*, Springer Verlag, August, 2006.

Davis, N., Harkema, H., Gaizauskas, R., Guo, Y.K., Ghanem, M., Barnwell, T., Guo, Y. and Ratcliffe, J. Three Approaches to GO-Tagging Biomedical Abstracts. In *Proceedings of the Second International Symposium on Semantic Mining in Biomedicine (SMBM06)*, Jena, April, 2006.

Barker, E., Higashinaka, R., Mairesse, F., Gaizauskas, R., Walker, M. and Foster, J. Simulating Cub Reporter Dialogues: The collection of naturalistic human-human dialogues for information access to text archives. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, May, 2006.

Saggion, H. and Gaizauskas, R. Language Resources for Background Gathering. In *Proceedings of Language, Resources and Evaluation (LREC 2006)*, Genoa, Italy, May, 2006.

Gaizauskas, R., Harkema, H., Harkema, H. and Setzer, A. Task-Oriented Extraction of Temporal Information: The Case of Clinical Narratives. In *Proceedings of the 13th International Symposium on Temporal Representation and Reasoning (TIME2006)*, 188-195, 2006.

Ghanem, M., Curcin, V., Guo, Y., Davis, N., Gaizauskas, R., Guo, Y.K., Harkema, H., Roberts, I. and Ratcliffe, J. GoTag: A case study in using a shared UK e-Science infrastructure for the Automatic Annotation of Medline Documents. In S.J. Cox (ed.) *Proceedings of the 4th UK e-Science All Hands Meeting*, Nottingham, UK, 2005. Available at: http://www.all-hands.org.uk/2005/proceedings/.

Saggion, H., Barker, E., Gaizauskas, R. and Foster, J. Integrating NLP Tools to Support Information Access to News Archives. In *Proceedings of the Fifth International Conference on Recent Advances in Natural Language Processing RANLP-2005*, Borovets, Bulgaria, September, 2005.

Harkema, H., Roberts, I., Gaizauskas, R. and Hepple, M. Information Extraction from Clinical Records. In S.J. Cox (ed.) *Proceedings of the 4th UK e-Science All Hands Meeting*, Nottingham, UK, 2005. Available at: http://www.allhands.org.uk/2005/proceedings/.

Harkema, H., Setzer, A., Gaizauskas, R., Gaizauskas, R., Power, R. and Rogers, J. Mining and Modelling Temporal Clinical Data. In S.J. Cox (ed.) *Proceedings of the 4th UK e-Science All Hands Meeting*, Nottingham, UK, 2005. Available at: http://www.all-hands.org.uk/2005/proceedings/.

Gaizauskas, R., Davis, N., Demetriou, G., Guo, Y. and Roberts, I. Integrating Text Mining Services into Distributed Bioinformatics Workflows: A Web Services Implementation. In *Proceedings of the IEEE International Conference on Services Computing (SCC 2004)*, Shanghai, 2004.

Saggion, H. and Gaizauskas, R. Mining on-line sources for definition knowledge. In *Proceedings of the 17th International FLAIRS Conference*, Miami Beach, 45-52, May, 2004.

Harkema, H., Gaizauskas, R., Hepple, M., Davis, N., Guo, Y., Roberts, A. and Roberts, I. A Large-Scale Resource for Storing and Recognizing Technical Terminology. In *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC2004)*, Lisbon, May, 2004.

Mitchell, B. and Gaizauskas, R. A Labelled Corpus for Prepositional Phrase Attachment. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal, May, 2004.

Gaizauskas, R., Davis, N., Demetriou, G., Guo, Y. and Roberts, I. Integrating Biomedical Text Mining Services into a Distributed Workflow Environment. In S. Cox (ed.) *Proceedings of UK e-Science All Hands Meeting 2004, Nottingham, UK*, 2004. Available at: http://www.allhands.org.uk/proceedings/.

Roberts, I. and Gaizauskas, R. Evaluating Passage Retrieval Approaches for Question Answering. In S. McDonald and J. Tait (ed.) *Advances in Information Retrieval: Proceedings of the 26th European Conference on Information Retrieval (ECIR'04)*, Lecture Notes in Computer Science, Vol. 2997, Springer, Sunderland, 72-84, April, 2004.

Demetriou , G., Skadina, I., Keskustalo, H., Karlgren, J., Deksne, D., Petrelli, D., Hansen, P., Gaizauskas, R. and Sanderson, M. Cross-Lingual Document Retrieval, Categorisation and Navigation Basedon Distributed Services. In *Proceedings of the First Baltic Conference Human LanguageTechnologies – the Baltic Perspective*, Riga, 2004.

Gaizauskas, R., Burnard, L., Clough, P. and Piao, S. Using the XARA XML-Aware Corpus Query Tool to Investigate the METER Corpus. In *Proceedings of Corpus Linguistics 2003*, Lancaster, 647-656, March, 2003.

Pustejovsky, J., Hanks, P., Saurí, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L. and Lazo, M. The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics 2003*, Lancaster, 647-656, March, 2003.

Gaizauskas, R., Hepple, M., Davis, N., Guo, Y., Harkema, H., Roberts, A. and Roberts, I. AMBIT: Acquiring Medical and Biological Information from Text. In S. Cox (ed.) *Proceedings of UK e-Science All Hands Meeting 2003, Nottingham, UK*, 2003. Available at: http://www.nesc.ac.uk/events/ahm2003/AHMCD/.

Clough, P., Gaizauskas, R. and Piao, S.L. Building and annotating a corpus for the study of journalistic text reuse. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, 1678-1691, May, 2002.

Clough, P., Gaizauskas, R., Piao, S. and Wilks, Y. METER: MEasuring TExt Reuse. In *Proceedings of the 40th Anniversary Meeting for the Association for Computational Linguistics (ACL-02)*, Philadelphia, 152-159, July, 2002.

Mitchell, B. and Gaizauskas, R. A Comparison of Machine Learning Algorithms for Prepositional Phrase Attachment. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, Spain, 2082-2087, May, 2002.

Ainsworth, S. E., Clarke, D. and Gaizauskas, R. J. Using edit distance algorithms to compare alternative approaches to ITS authoring. In S. A. Cerri and G. Gouardères and F. Paraguaçu (ed.) *Intelligent Tutoring Systems*, Lecture Notes in Computer Science, No. 2363, Springer-Verlag, 873-882, 2002.

Gaizauskas, R., Foster, J., Wilks, Y., Arundel, J., Clough, P. and Piao, S. The Meter Corpus: A Corpus for Analysing Journalistic Text Reuse. In *Proceedings of Corpus Linguistics 2001 Conference*, 214-223, 2001.

Scott, S. and Gaizauskas, R. QA-LaSIE: A Natural Language Question Answering System. In E. Stroulia and S. Matwin (ed.) *Advances in Artificial Intelligence: Fourteenth Biennial Conference of the Canadian Society for Computational Studies of Intelligence*, Springer, Berlin, 172-182, June, 2001.

Gaizauskas, R., Herring, P., Oakes, M., Beaulieu, M., Willett, P., Fowkes, F. and Jonsson, A. Intelligent Access to Text: Integrating Information Extraction Technology into Text Browsers. In *Proceedings of the Human Language Technology Conference (HLT2001)*, San Diego, 189-193, 2001.

Demetriou, G. and Gaizauskas, R. Automatically Augmenting Terminological Lexicons from Untagged Text. In *Proceedings of the Second International Conference on Language Resources and Evaluation*, Athens, 861-868, 2000.

Harmain, H.M. and Gaizauskas, R. CM-Builder: An Automated NL-based CASE Tool. In *Proceedings of the 15th IEEE International Conference on Automated Software Engineering (ASE'2000)*, Grenoble, 45-53, 2000.

Gaizauskas, R. and Humphreys, K. A Combined IR/NLP Approach to Question Answering Against Large Text Collection. In *Proceedings of RIAO 2000: Content-Based Multimedia Information Access*, Paris, 1288-1304, 2000.

Stevenson, M. and Gaizauskas, R. Using Corpus-derived Name Lists for Named Entity Recognition. In *Proceedings of the Sixth Conference on Applied Natural Language Processing and First Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, 84-89, 2000.

Setzer, A. and Gaizauskas, R. Annotating Events and Temporal Information in Newswire Texts. In *Proceedings of the Second International Conference on Language Resources and Evaluation*, Athens, 1287-1294, 2000.

Stevenson, M. and Gaizauskas, R. Experiments on Sentence Boundary Detection. In *Proceedings of the Sixth Conference on Applied Natural Language Processing and First Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, 84-89, 2000.

Krotov, A., Hepple, M., Gaizauskas, R. and Wilks, Y. Compacting the Penn Treebank Grammar. In *Proceedings of the Joint 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computation Linguistics (COLING-ACL'98)*, Montreal, 699-703, 1998.

Azzam, S., Humphreys, K. and Gaizauskas, R. Evaluating a Focus-based Approach to Anaphora Resolution. In *Proceedings of the Joint 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98)*, Montreal, 74-78, 1998.

Azzam, S., Humphreys, K. and Gaizauskas, R. Extending a Simple Coreference Algorithm with a Focusing Mechanism. In *Proceedings of the Second Discourse, Anaphora and Reference Resolution Colloquium (DAARC2)*, Lancaster, UK, 1998.

Gaizauskas, R., Hepple, M. and Huyck, C. A Scheme for Comparative Evaluation of Diverse Parsing Systems. In *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC'98)*, Granada, 143-149, 1998.

Rodgers, P., Gaizauskas, R., Humphreys, K. and Cunningham, H. Visual Execution and Data Visualisation in Natural Language Processing. In *Proceedings of the IEEE Conference on Visual Languages (VL'97)*, Capri, Italy, 1997.

Cunningham, H., Humphreys, K., Gaizauskas, R. and Wilks, Y. Software Infrastructure for Natural Language Processing. In *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP-97)*, 1997.

Gaizauskas, R. and Robertson, A.M. Coupling Information Retrieval and Information Extraction: A New Text Technology for Gathering Information from the Web. In *Proceedings of the 5th Computer-Assisted Information Searching on Internet Conference (RIAO'97)*, Montreal, 356-370, 1997.

Robertson, A.M. and Gaizauskas, R. On the Marriage of Information Retrieval and Information Extraction. In *Proceedings of BCS IRSG 19th Annual Colloquium on Information Retrieval Research*, Aberdeen, Scotland, 60-67, 1997.

Wakao, T., Gaizauskas, R. and Wilks, Y. Evaluation of an Algorithm for the Recognition and Classification of Proper Names. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING96)*, Copenhagen, 418-423, 1996.

Gaizauskas, R. and Humphreys, K. XI: A Simple Prolog-based Language for Cross-Classification and Inheritance. In *Proceedings of the 7th International Conference on Artificial Intelligence: Methodology, Systems, Applications (AIMSA96)*, Sozopol, Bulgaria, 86-95, 1996.

Gaizauskas, R., Cunningham, H., Wilks, Y., Rodgers, P. and Humphreys, K. GATE – an Environment to Support Research and Development in Natural Language Engineering. In *Proceedings of the 8th IEEE International Conference on Tools with Artificial Intelligence (ICTAI-96)*, Toulouse, France, 58-66, October, 1996.

Cunningham, H., Wilks, Y. and Gaizauskas, R. New Methods, Current Trends and Software Infrastructure for NLP. In *Proceedings of the Conference on New Methods in Natural Language Processing (NeMLaP-2)*, Turkey, September, 1996.

Gaizauskas, R. and Humphreys, K. Quantitative Evaluation of Coreference Algorithms in an Information Extraction System. In *DAARC96 - Discourse Anaphora and Anaphor Resolution Colloquium*, Lancaster University, 1996.

Krotov, A., Gaizauskas, R. and Wilks, Y. Acquiring a Stochastic Context-Free Grammar from the Penn Treebank. In *Proceedings of the Third International Conference of the Cognitive Science of Natural Language Processing (CSNLP94)*, Dublin, 79-86, July, 1994.

Dixon, A., Evans, R., Walker, J., Cahill, L., Gaizauskas, R. and Higginson, J. POETIC: A System for Automatic Detection and Broadcasting of Road Congestion Information. In *Proceedings of the First World Congress on Applications of Transport Telematics and Intelligent Vehicle-Highway Systems*, Paris, 2547-2554, November and December, 1994.

Evans, R., Gaizauskas, R. and Cahill, L. MUCing about in Pop: Message Understanding in five languages!. In I.M. Graham (ed.) *Applications and Innovations in Expert Systems*, BHR Group and BCS SGES, Cambridge, 295-308, 1993.

Gaizauskas, R. Deriving Answers to Logical Queries via Answer Composition. In Wiggins, G.A. and Mellish, C. and Duncan, T. (ed.) *3rd UK Annual Conference on Logic Programming, Edinburgh 1991*, Springer-Verlag, London, 112-134, 1992.

6.4   Papers in Refereed Workshop Proceedings

Thomas, A., Gaizauskas, R., Booth, C. and Lu, H. Leveraging LLMs for Post-OCR Correction of Historical Newspapers. In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages – LT4HALA 2024*, Turin, Italy, 2024.

Alrashid, T. and Gaizauskas, R. ScANT: A Small Corpus of Scene-Annotated Narrative Texts. In *Proceedings of the 6th workshop on narrative extraction from texts (Text2Story 2023)*, 143-149, 2023.

Aldihan, H., Gaizauskas, R. and Fitzmaurice, S. A Pilot Study on the Collection and Computational Analysis of Linguistic Differences Amongst Men and Women in a Kuwaiti Arabic WhatsApp Dataset. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 372–380, Dec, 2022.

Clayton, J. and Gaizauskas, R. Predicting the Presence of Reasoning Markers in Argumentative Text. In *Proceedings of the 9th Workshop on Argument Mining*, International Conference on Computational Linguistics, Online and in Gyeongju, Republic of Korea, 137–142, Oct, 2022.

Alrashid, T. and Gaizauskas, R. A Pilot Study on Annotating Scenes in Narrative Text using SceneML. In *Proceedings of the 4th international workshop on narrative extraction from texts (Text2Story 2021)*, 7-14, 2021.

Papazoglou, V. and Gaizauskas, R. Using Listeners' Interpretations in Topic Classification of Song Lyrics. In *Proceedings of the 2nd Workshop on NLP for Music and Spoken Audio (NLP4MusA21)*, 22-26, 2021.

, and Gaizauskas, R. A Framework for Annotating Co-working Dialogues in Complex Task Settings. In *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue*, SEMDIAL, London, United Kingdom, 70-78, sep, 2019.

Alrashid, T. and Gaizauskas, R. Annotating and Recognising Visually Descriptive Language. In *Proceedings of the 15th Workshop on Interoperable Semantic Annotation (ISA-15)*, Gothenburg, Sweden, 2019.

Gaizauskas, R. and Alrashid, T. SceneML: A Proposal for Annotating Scenes in Narrative Text. In *Proceedings of the 15th Workshop on Interoperable Semantic Annotation (ISA-15)*, Gothenburg, Sweden, 2019.

Barker, E. and Barker, E. Summarizing Multi-Party Argumentative Conversations in Reader Comment on News. In *Proceedings of the ACL 2016 3rd Workshop on Argument Mining*, aug, 2016.

Riccardi, G., Bechet, F., Danieli, M., Favre, B., Gaizauskas, R., Kruschwitz, U. and Poesio, M. The SENSEI Project: Making Sense of Human Conversations. In Future and Emerging Language Technologies, First International Workshop (FETLT 2015), Seville, Spain, November 19-20, 2015, Revised Selected Papers (,ed.), Springer, 10-33, 2016.

Wang, J. and Gaizauskas, R. Generating Image Descriptions with Gold Standard Visual Inputs: Motivation, Evaluation and Baselines. In *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*, Association for Computational Linguistics, Brighton, UK, 117–126, September, 2015.

Gilbert, A., Piras, L., Wang, J., Yan, F., Dellandrea, E., Gaizauskas, R., Villegas, M. and Mikolajczyk, K. Overview of the ImageCLEF 2015 Scalable Image Annotation, Localization and Sentence Generation task. In *CLEF2015 Working Notes*, CEUR Workshop Proceedings, CEUR-WS.org, Toulouse, France, September 8-11, 2015.

Gaizauskas , R., Gaizauskas , R. and Gaizauskas , R. Defining Visually Descriptive Language. In *Proceedings of the Fourth Workshop on Vision and Language*, Association for Computational Linguistics, Lisbon, Portugal, 10–17, September, 2015.

Wang, J., Yan, F., Aker, A. and Gaizauskas, R. A Poodle or a Dog? Evaluating Automatic Image Annotation Using Human Descriptions at Different Levels of Granularity. In *Proceedings of the 3rd Workshop on Vision and Language (VL'14)*, Dublin, 2014.

Gaizauskas, R., Gaizauskas, R. and Aker, A. Assigning Terms to Domains by Document Classification. In *Proceedings of the 4th International Workshop on Computational Terminology (Computerm)*, Dublin, 2014.

Aker, A., Cohn, T. and Gaizauskas, R. Redundancy reduction for multi-document summaries using A* search and discriminative training. In *Proceedings of the SEPLN 2012 Workshop on Automatic Text Summarization of the Future*, Spain, 2012.

Gaizauskas, R., Barker, E., Chang, C-L., Derczynski, L., Phiri, M. and Peng, C. Applying ISO-Space to Healthcare Facility Design Evaluation Reports. In *Proceedings of the Joint ISA-7 Workshop on Interoperable Semantic Annotation, SRSL-3 Workshop on Semantic Representation for Spoken Language and I2MRT Workshop on Multimodal Resources and Tools*, Istanbul, 13-20, 2012.

Derczynski, L. and Gaizauskas, R. An Annotation Scheme for Reichenbach's Verbal Tense Structure. In *Proceedings of the 6th Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, 10-17, 2011.

Di Fabbrizio, G., Aker, A. and Gaizauskas, R. STARLET: Multi-document Summarization of Service and Product Reviews with Balanced Rating Distributions. In *Proceedings of the 2011 IEEE International Conference on Data Mining (ICDM) Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction (SENTIRE)*, Vancouver, Canada, december, 2011.

Derczynski, L. and Gaizauskas, R. USFD2: Annotating temporal expresions and TLINKs for TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, Association for Computational Linguistics, Stroudsburg, PA, USA, 337–340, 2010.

Skadiņa, I., Vasiļjevs, A., Skadiņš, R., Gaizauskas, R., Tufiş, D. and Gornostay, T. Analysis and Evaluation of Comparable Corpora for Under Resourced Areas of Machine Translation. In *Proceedings of the 3rd Workshop on Building and Using Comparable Corpora (BUCC 2010)*, No. May, ELRA, 6–14, 2010.

Stevenson, M., Guo, Y., Alamri, A. and Gaizauskas, R. Disambiguation of Biomedical Abbreviations. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT) workshop "BioNLP 2009"*, Boulder, CO., 2009.

Roberts, A., Gaizauskas, R., Hepple, M., Demetriou, G., Guo, Y. and Setzer, A. Semantic Annotation of Clinical Text: The CLEF Corpus. In *Proceedings of the LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining*, Marrakech, 19-26, May, 2008.

Stevenson, M., Guo, Y. and Gaizauskas, R. Knowledge Sources for Word Sense Disambiguation of Biomedical Text. In *Proceedings of the ACL Workshop on Current Trends in Biomedical Natural Language Processing*, Association for Computational Linguistics, Columbus, Ohio, 80–87, June, 2008.

Shaw, R., Solway, B., Gaizauskas, R. and Greenwood, M. A. Evaluation of Automatically Reformulated Questions in Question Series. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008) Workshop on Information Retrieval for Question Answering (IR4QA08)*, 2008.

Roberts, A., Gaizauskas, R. and Hepple, M. Extracting Clinical Relationships from Patient Narratives. In *Proceedings of the ACL Workshop on Current Trends in Biomedical Natural Language Processing*, Association for Computational Linguistics, Columbus, Ohio, 10–18, June, 2008.

Derczynski, L., Wang, J., Gaizauskas, R. and Greenwood, M. A. A Data Driven Approach to Query Expansion in Question Answering. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008) Workshop on Information Retrieval for Question Answering (IR4QA08)*, 34-41, 2008.

Barker, E., Polifroni, J., Walker, M. and Gaizauskas, R. Angle Seeking as a Scenario for Task-Based Evaluation of Information Access Technologies. In *Proceedings of the 30th annual European Conference on Information Retrieval (ECIR08) Workshop on Novel Methodologies for Evaluation in Information Retrieval*, Glasgow, March, 2008.

Aker, A. and Gaizauskas, R. Evaluating automatically generated user-focused multi-document summaries for geo-referenced images. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008) Workshop on Multi-source, Multilingual Information Extraction and Summarization (MMIES2)*, 2008.

Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G. and Pustejovsky, J. SemEval-2007 Task 15: TempEval Temporal Relation Identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, Prague, 75-80, 2007.

Hepple, M., Setzer, A. and Gaizauskas, R. USFD: Preliminary Exploration of Features and Classifiers for the TempEval-2007 Task. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, Prague, 438-441, 2007.

Guo, Y., Gaizauskas, R., Roberts, I., Demetriou, G. and Hepple, M. Identifying Personal Health Information Using Support Vector Machines. In *Proceedings of the AMIA 2006 Workshop on Challenges in Natural Language Processing for Clinical Data*, Washington, November, 2006.

Aswani, N. and Gaizauskas, R. Aligning words in English-Hindi parallel corpora. In *Proceedings of the ACL 2005 Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond)*, Ann Arbor, Michigan, 2005.

Saggion, H. and Gaizauskas, R. Experiments on Statistical and Pattern-based Biographical Summarization. In *Lecture Notes in Artificial Intelligence (EPIA 2005 Vol. 3808)*, Springer Verlag, December, 2005.

Gaizauskas, R., Hepple, M., Saggion, H., Saggion, H. and Humphreys, K. SUPPLE: A Practical Parser for Natural Language Engineering Applications. In *Proceedings of the 9th International Workshop on Parsing Technologies (IWPT2005)*, Vancouver, 2005.

Aswani, N. and Gaizauskas, R. A hybrid approach to aligning sentences and words in English-Hindi parallel corpora. In *Proceedings of the ACL 2005 Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond)*, Ann Arbor, Michigan, 2005.

Barker, E. and Gaizauskas, R. Evaluating Cub Reporter: proposals for extrinsic evaluation of journalists using language technologies to access a news archive in research. In A. Bailey and I. Ruthven and L. Azzopardi (ed.) *Proceedings of the Workshop on Evaluating User Studies in Information Access, 5th International Conference on Conceptions of Library and Information SCience,(COLIS 2005)*, 2005.

Roberts, I., Harkema, H., Gaizauskas, R. and Gaizauskas, R. A Web Service for Biomedical Term Look-Up. In *Proceedings of 12th International Conference on Intelligent Systems for Molecular Biology (ISMB2004) BioLink Workshop*, Glasgow, July, 2004.

Saggion, H. and Gaizauskas, R. Multi-Document Summarization by Cluster/Profile Relevance and Redundancy Removal. In *Proceedings of the HLT/NAACL Document Understanding Workshop (DUC 2004)*, Boston, May, 2004. Available at http://duc.nist.gov/pubs/-2004papers/usheffield.saggion.pdf.

Harkema, H., Gaizauskas, R., Hepple, M., Roberts, A., Roberts, I., Davis, N. and Guo, Y. A Large Scale Terminology Resource for Biomedical Text Processing. In *Proceedings of the NAACL/HLT 2004 Workshop on Linking Biological Literature, Ontologies and Databases: Tools for Users*, Boston, May, 2004.

Saggion, H., Gaizauskas, R., Hepple, M., Roberts, I. and Greenwood, M. A. Exploring the Performance of Boolean Retrieval Strategies For Open Domain Question Answering. In *Proceedings of the SIGIR 2004 Workshop on Information Retrieval for Question Answering*, Sheffield, July, 2004.

Greenwood, M. and Gaizauskas, R. Using a Named Entity Tagger to Generalise Surface Matching Text Patterns for Question Answering. In *Proceedings of the European Conference for Computational Linguistics (EACL03) Workshop on Natural Language Processing for Question Answering*, Budapest, 29-34, April, 2003.

Setzer, A., Gaizauskas, R. and Hepple, M. Using Semantic Inference for Temporal Annotation Comparison. In *Proceedings of the Fourth International Workshop on Inference in Computational Semantics (ICoS-4)*, Nancy, France, September, 2003.

Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A. and Katz, G. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, Tilburg, January, 2003.

Setzer, A. and Gaizauskas, R. On the Importance of Annotating Temporal Event-Event Relations in Text. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC2002) Workshop on Annotation Standards for Temporal Information in Natural Language*, Las Palmas de Gran Canaria, 52-60, May, 2002.

Demetriou, G. and Gaizauskas, R. Utilizing Text Mining Results: The PastaWeb System. In *Proceedings of the Association for Computational Linguistics Workshop on Natural Language Processing in the Biomedical Domain*, Philadelphia, 77-84, July, 2002.

Setzer, A. and Gaizauskas, R. A Pilot Study on Annotating Temporal Relations in Text. In *Proceedings of the ACL01 Workshop on Temporal and Spatial Information Processing*, Toulouse, France, July, 2001.

Setzer, A. and Gaizauskas, R. Building a Temporally Annotated Corpus for Information Extraction. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC2000) Workshop on Information Extraction Meets Corpus Linguistics*, Athens, 9-14, 2000.

Stevenson, M. and Gaizauskas, R. Improving Named Entity Recognition Using Annotated Corpora. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC2000) Workshop on Information Extraction Meets Corpus Linguistics*, Athens, 24-30, 2000.

Gaizauskas, R., Demetriou, G. and Humphreys, K. Term Recognition and Classification in Biological Science Journal Articles. In *Proceedings of the Computational Terminology for Medical and Biological Applications Workshop of the 2nd International Conference on Natural Language Processing (NLP2000)*, Patras, Greece, 37-44, June, 2000.

Humphreys, K., Demetriou, G. and Gaizauskas, R. Two Applications of Information Extraction to Biological Science Journal Articles: Enzyme Interactions and Protein Structures. In *Proceedings of the Workshop on Natural Language Processing for Biology, held at the Pacific Symposium on Biocomputing (PSB2000)*, Hawaii, USA, January, 2000.

Humphreys, K., Demetriou, G. and Gaizauskas, R. Information Extraction from Biological Science Journal Articles. In *Proceedings of the IJCAI'99 Workshop on Text Mining*, Stockholm, 50-51, August, 1999.

Cunninghmam, H., Gaizauskas, R., Humphreys, K. and Wilks, Y. Experience with a Language Engineering Architecture: 3 Years of GATE. In *Proceedings of the AISB'99 Workshop on Reference Architectures and Data Standards for NLP*, Edinburgh, April, 1999.

Azzam, S., Humphreys, K. and Gaizauskas, R. Using Coreference Chains for Text Summarization. In *Proceedings of the ACL'99 Workshop on Coreference and its Applications*, Baltimore, June, 1999.

Renals, S., Gotoh, Y., Gaizauskas, R. and Stevenson, M. Baseline IE-NE Experiments Using the SPRACH/LASIE System. In *Proceedings of the DARPA Broadcast News Workshop*, Herndon, VA, February, 1999.

Gaizauskas, R., Hepple, M. and Huyck, C. Modifying Existing Annotated Corpora for General Comparative Evaluation of Parsing. In *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC'98) Workshop on Evaluation of Parsing Systems*, Granada, 1998.

Azzam, S., Humphreys, K. and Gaizauskas, R. Using a Language Independent Domain Model for Multilingual Information Extraction. In *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC'98) Workshop on Linguistic Coreference*, Granada, 1998.

Azzam, S., Humphreys, K., Gaizauskas, R., Cunningham, H. and Wilks, Y. Using a Language Independent Domain Model for Multilingual Information Extraction. In Costas Spryrodopoulos (ed.) *Proceedings of the IJCAI-97 Workshop on Multilinguality in the Software Industry: the AI Contribution (MULSAIC-97)*, Nagoya, Japan, 1997.

Humphreys, K., Gaizauskas, R. and Azzam, S. Event Coreference for Information Extraction. In *Proceedings of the ACL-97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, Madrid, Spain, 1997.

Gaizauskas, R. and Humphreys, K. Using Verb Semantic Role Information to Extend Partial Parses via a Co-reference Mechanism. In J. Carroll (ed.) *Proceedings of the Workshop on Robust Parsing*, Prague, Czech Republic, 103-113, August, 1996.

Cunningham, H., Wilks, Y. and Gaizauskas, R. Software Infrastructure for Language Engineering. In *Proceedings of the AISB Workshop on Language Engineering for Document Analysis and Recognition*, Brighton, UK, 1996.

Evans, R., Gaizauskas, R. and Hartley, A.F. POETIC – The Portable Extendable Traffic Information Collator. In H. Jamsa (ed.) *OECD Workshop on Knowledge-Based Expert Systems in Transportation*, 171-184, 1990.

6.5   Refereed Poster Papers

Aker, A. and Gaizauskas, R. Understanding the types of information humans associate with geographic objects. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, ACM, New York, NY, USA, 1929–1932, 2011.

Oakes, M., Gaizauskas, R., Fowkes, H., Jonsson, A., Wan, V. and Beaulieu, M. Comparison Between a Method Based on the Chi-Square Test and a Support Vector Machine for Document Classification. In *Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR01)*, New Orleans, 2001.

Azzam, S., Humphreys, K., Gaizauskas, R., Cunningham, H. and Wilks, Y. A Design for a Multilingual Information Extraction System. In *IJCAI-97 Poster Session Abstracts*, Nagoya, Japan, 11, 1997.

Cunningham, H., Wilks, Y. and Gaizauskas, R. GATE – a General Architecture for Text Engineering. In *Proceedings of the 16th International Conference on Computational Linguistics COLING96*, Copenhagen, 1057-1060, 1996.

6.6   Chapters in Edited Collections

Aker, A., Plaza, L., Lloret, E. and Gaizauskas, R. Multi-Document Summarization Techniques for Generating Image Descriptions: A Comparative Analysis. In Multi-source, Multilingual Information Extraction and Summarization (T. Poibeau and H. Saggion and J. Piskorski and R. Yangarber,ed.), Springer Berlin Heidelberg, 299-320, 2013.

Di Fabbrizio, G., Stent, A. J. and Gaizauskas, R. Summarizing Opinion-Related Information for Mobile Devices. In Mobile Speech and Advanced Natural Language Solutions (,ed.), Springer, 2013.

, and , Corpora and Text Reuse. In Corpus Linguistics: An International Handbook (Lüdeling, Anke and Kytö, Merja,ed.), Mouton de Gruyter, 2009. Chapter 59.

Gaizauskas, R., Saggion, H. and Barker, E. Information Access and Natural Language Processing: A Stimulating Dialogue. In Words and Intelligence II: Essays in Honor of Yorick Wilks (K. Ahmad and C.Brewster and M. Stevenson,ed.), Springer, 2007.

Gaizauskas, R. and Barker, E. J. Mice from a Mountain: Reflections on Current Issues in Evaluation of Written Language Technology. In Charting a New Course: Natural Language Processing and Information Retrieval (J. I. Tait,ed.), Springer, 195-238, 2005.

Gaizauskas, R. and Humphreys, K. Quantitative Evaluation of Coreference Algorithms in an Information Extraction System. In Corpus-based and Computational Approaches to Discourse Anaphora (S. Botley and T. McEnery,ed.), John Benjamins(Amsterdam), 145-169, 2000.

Wilks, Y. and Gaizauskas, R. LaSIE Jumps the GATE. In Natural Language Information Retrieval (T. Strazlkowski,ed.), Kluwer Academic Publishers(Dordrecht), 197-214, 1999.

Gaizauskas, R., Humphreys, K., Azzam, S. and Wilks, Y. Concepticons *vs.* Lexicons: An Architecture for Multilingual Information Extraction. In M.T. Pazienza (ed.) *Proceedings of the Summer School on Information Extraction (SCIE-97)*, LNCS/LNAI, No. 1299, Springer-Verlag, 28-43, 1997.

6.7   Unrefereed Conference and Workshop Proceedings

Aker, A., Celli, F., Kurtic, E., Hepple, M. and Gaizauskas, R. Sheffield-Trento System for Comment-to-Article Linking and Argument Structure Annotation in the Online News Domain. In *MultiLing 2015 Proceedings Addendum*, 2015.

Burman, A., Jayapal, A., Kannan, S., Kavilikatta, M., Alhelbawy, A., Derczynski, L. and Gaizauskas, R. USFD at KBP 2011: Entity Linking, Slot Filling and Temporal Bounding. In *Proceedings of the 4th Text Analysis Conference (TAC 2011)*, 2011.

Yu, J., Mujgond, O. and Gaizauskas, R. The University of Sheffield System at TAC KBP 2010. In *Proceedings of the Third Text Analysis Conference (TAC 2010)*, Gaithersburg, Maryland, USA, 418-423, November, 2010.

Gaizauskas, R., Greenwood, M., Hepple, M., Harkema, H., Saggion, H. and Sanka, A. The University of Sheffield's TREC 2005 Q&A Experiments. In *Proceedings of the 14th Text REtrieval Conference (TREC 2005)*, Gaithersburg, USA, November, 2005.

Gaizauskas, R., Greenwood, M., Hepple, M., Roberts, I. and Saggion, H. The University of Sheffield's TREC 2004 Q&A Experiments. In *Proceedings of the 13th Text REtrieval Conference (TREC 2004)*, Gaithersburg, USA, November, 2004.

Gaizauskas, R., Greenwood, M., Hepple, M., Roberts, I., Saggion, H. and Sargaison, M. The University of Sheffield's TREC 2003 Q&A Experiments. In *Proceedings of the 12th Text REtrieval Conference (TREC 2003)*, Gaithersburg, USA, November, 2003.

Greenwood, M., Roberts, I. and Gaizauskas, R. The University of Sheffield, TREC Q & A System. In *Proceedings of the Eleventh Text Retrieval Conference (TREC-2002)*, NIST Special Publications, No. 500-251, 2002.

Scott, S. and Gaizauskas, R. The University of Sheffield, TREC9 Q & A System. In *Proceedings of the Ninth Text Retrieval Conference (TREC9)*, NIST Special Publications, No. 500-249, 2000.

Humphreys, K., Gaizauskas, R., Gaizauskas, R. and Gaizauskas, R. University of Sheffield TREC-8 Q&A system. In *Proceedings of the Eighth Text Retrieval Conference (TREC-8)*, Gaithersburg, November, 1999. Available at http://trec.nist.gov/pubs/trec8/t8_proceedings.html.

Humphreys, K., Gaizauskas, R., Azzam, S., Huyck, C., Mitchell, B., Cunningham, H. and Wilks, Y. University of Sheffield: Description of the LaSIE-II System as used for MUC-7. In *Proceedings of the Seventh Message Understanding Conference*, 84-89, 1998. Available at http://www.itl.nist.gov/iad/894.02/related_projects/muc/proceedings/muc_7_toc.html.

Takemoto, Y., Wakao, T., Yamada, H., Gaizauskas, R. and Wilks, Y. Description of the NEC/Sheffield System as used for MET Japanese. In *Proceedings of the Tipster Phase II 24 Month Workshop*, Morgan Kaufmann, Tysons Corner, VA, 1996.

Cunningham, H., Humphreys, K., Gaizauskas, R. and Wilks, Y. TIPSTER-Compatible Projects at Sheffield. In *Proceedings of the Tipster Phase II 24 Month Workshop*, Morgan Kaufmann, Tysons Corner, VA, 1996.

Gaizauskas, R., Wakao, T., Humphreys, K., Cunningham, H. and Wilks, Y. University of Sheffield: Description of the LaSIE system as used for MUC-6.. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, Morgan Kaufmann, 207-220, 1995.

Gaizauskas, R., Cahill, L.J. and Evans, R. Sussex University: Description of the Sussex System Used for MUC-5. In *Proceedings of the Fifth Message Undersanding Conference (MUC-5)*, Morgan Kaufmann, 321-335, 1993.

6.8   Technical Reports

Roberts, I. and Gaizauskas, R. Evaluating Passage Retrieval Approaches for Question Answering. Technical Report CS-03-06, University of Sheffield, 2003.

Humphreys, K., Gaizauskas, R. and Cunningham, H. LaSIE Technical Specifications. Technical report $CS - 00 - 09$, Department of Computer Science, University of Sheffield, 2000.

Gaizauskas, R., Rodgers, P. and Humphreys, K. Visual Natural Language Processing. Technical report $CS - 99 - 02$, Department of Computer Science, University of Sheffield, 1999.

Harmain, H.M. and Gaizauskas, R. Natural Language Processing and Object-Oriented Analysis. Technical report $CS - 98 - 08$, Department of Computer Science, University of Sheffield, 1998.

Gaizauskas, R. and Wilks, Y. Information Extraction: Beyond Document Retrieval. Technical report $CS - 97 - 10$, Department of Computer Science, University of Sheffield, 1997.

Gaizauskas, R. and Humphreys, K. Quantitative Evaluation of Coreference Algorithms in an Information Extraction System. Technical report $CS - 97 - 19$, Department of Computer Science, University of Sheffield, 1997.

Krotov, A., Hepple, M., Gaizauskas, R. and Wilks, Y. Compacting the Penn Treebank Grammar. Technical report $CS - 97 - 04$, Department of Computer Science, University of Sheffield, 1997.

Gaizauskas, R. and Humphreys, K. Using a Semantic Network for Information Extraction. Technical Report CS-97-03, Department of Computer Science, University of Sheffield, 1997.

Gaizauskas, R., Rodgers, P., Cunningham, H. and Humphreys, K. GATE User Guide. , 1996. Available at `http://www.dcs.shef.ac.uk/research/groups/nlp/gate`.

Gaizauskas, R. XI: A Knowledge Representation Language Based on Cross-Classification and Inheritance. Technical Report CS-95-24, Department of Computer Science, University of Sheffield, 1995.

Gaizauskas, R. Investigations into the Grammar Underlying the Penn Treebank II. Technical Report CS-95-25, Department of Computer Science, Univeristy of Sheffield, 1995.

Crouch, R., Gaizauskas, R. and Netter, K. Report of the Study Group on Assessment and Evaluation. Technical Report , , 1995. EAGLES Project Technical Report. Available as `http://arxiv.org/abs/cmp-lg/9601003`.

Gaizauskas, R., Wakao, T., Humphreys, K., Cunningham, H. and Wilks, Y. LaSIE – A Large-Scale Information Extraction System. Technical Report CS – 95 – 27, Department of Computer Science, University of Sheffield, 1995.

Cahill, L.J., Gaizauskas, R. and Evans, R. POETIC: A Fully-Implemented NL System for Understanding Traffic Reports. In *Fully-Implemented Natural Language Understanding Systems: Proceedings of the Trento Workshop*, Heidelberg, 86-99, March, 1992.

Mellish, C., Allport, D., Hartley, A.F., Evans, R., Cahill, L.J., Gaizauskas, R. and Walker, J. The TIC Message Analyser. Cognitive Science Research Paper #225, University of Sussex, 1992.

Gaizauskas, R. Deriving Answers to Logical Queries by Answer Composition. Cognitive Science Research Paper 214, University of Sussex, 1992.

Gaizauskas, R. A Theory of Answers for First Order Deductive Database Systems. Cognitive Science Research Paper 128, University of Sussex, 1988.

Gaizauskas, R. A Philosophical Examination of Our Responsibilities to Future Generations. Technical Report 1159, Atomic Energy Control Board of Canada, 1977.

6.9   Journals/Proceedings Edited

Hirschman, L. and Gaizauskas, R. (ed). Journal of Natural Language Engineering, Special Issue on Question Answering. L. Hirschman and R. Gaizauskas (editor), 2001. Guest editors.

Gaizauskas, R. (ed). Journal of Computer Speech and Language, Special Issue on Evaluation. , 1998. Guest editor.

Gaizauskas, R. (ed). Proceedings of the UK Speech and Language Technology (SALT) Club Workshop on Evaluation in Speech and Language Technology. R. Gaizauskas (editor), June, 1997.

6.10   Book Reviews

Gaizauskas, R. A Review of *Recent Advances in Terminology* edited by D. Bourigault, C. Jacquemin and M-C. LHomme. *Computational Linguistics*, 29(2), 328-332, 2003.

Gaizauskas, R. A Review of *Evaluating Natural Language Processing Systems: An Analysis and Review* by Karen Sparck Jones and Julia Galliers. *Journal of Natural Language Engineering*, 4(2), 1998.

Gaizauskas, R. A Review of *Truth and Modality for Knowledge Representation* by Raymond Turner. *AISB Quarterly, Spring edition*(79), 38-39, 1992.

Gaizauskas, R. A Review of *Common Lispcraft* by Robert Wilensky. *AISB Quarterly, Spring edition*(61), 1987.