What do we do and why

- Natural Language Processing (NLP) is an interdisciplinary field that uses computational methods
  - To investigate how human language functions (science)
  - To develop practical applications involving the manipulation of language (engineering)

- For historic reasons, NLP concentrates on language in written form

- Why study language? Importance of language hard to underestimate. It is:
  - A distinguishing feature of human beings, key medium of social interaction, basis of collective memory/culture/commerce

- NLP group (founded: 1993) central areas of current interest:
  - Information access: text mining/analytics, semantic annotation, question answering, summarisation, clustering/classification
    * To/from: news, social media, scientific papers, clinical records, …
  - Language technology integration platforms: General Architecture for Text Engineering (GATE)
  - Machine Translation: new models of statistical MT, evaluation of MT systems, mining on-line resources for training data for SMT systems
  - Foundational Topics: word sense disambiguation, parsing, semantics of time and space, machine learning and NLP

Who we are

Academics (6)

Prof. Rob Gaizauskas (Head of Group)   Dr. Mark Hepple   Dr. Lucia Specia   Dr. Mark Stevenson

Dr. Andreas Vlachos (09/15)   Prof. Yorick Wilks (emeritus)

Senior Research Scientists (2)

Dr. Kalina Bontcheva   Prof. Hamish Cunningham

Research Associates / Fellows* (17/5*)

Dr. Ahmet Aker*   Isabelle Augenstein   Dr. Emma Barker   Dr. Frederic Blain

Dr. Leon Derczynski   Dr. Adam Funk   Dr. Genevieve Gorrell   Dr. Mark Greenwood

Dr. Emina Kurtic   Vavara Logacheva   Dr. Diana Maynard*   Monica Paramita

Dr. Wim Peters*   Johann Petrak*   Srijith Prabhakaran Nain   Dr. Judita Preiss

Dr. Angus Roberts*   Ian Roberts   Dominic Rout   Carolina Scarton

Dr. Josiah Wang   Dr. Kashif Shah

Research Students (22)

Abdulaziz Alamri   Nada Al-Mhabis   Faisal Alvi   Marcelo Amanco

Daniel Beck   Zsolt Bitvai   Rachel Cotterill   Nigel Dewdney

Ignatius Ezeani   Douwe Gelling   Michal Lukasik   Maryhilda McClure

Ikechukwu E Onyenwe   Gustavo Paetzold   Adam Poulston   Roland Roller

Nattapong Sanchan   Yunita Sari   Karin Sim Smith   Xingyi Song

David Steele   Konstantin Yershov

Admin Support (3)

Lucy Moffatt   Joanne Suter   Alice Tucker

13 May 2015   Presentation to IAB

Who we are

Academics (6)

Prof. Rob Gaizauskas (Head of Group)   Dr. Mark Hepple   Dr. Lucia Specia   Dr. Mark Stevenson

Dr. Andreas Vlachos (09/15)   Prof. Yorick Wilks (emeritus)

Senior Research Scientists (2)

Dr. Kalina Bontcheva   Prof. Hamish Cunningham

Research Associates / Fellows* (17/5*)

Dr. Ahmet Aker*   Isabell...   ...ederic Blain

Dr. Leon Derczynski   ...   ...r. Mark Greenwood

Dr. Emina Kurtic   Monica Paramita

Dr. Wim Peters*   ...aran Nain   Dr. Judita Preiss

Dr. Angus Roberts*   ...Rout   Carolina Scarton

Dr. Josiah Wang...

Research S...

Abdulaziz Alamr...   ...whabis   Faisal Alvi   Marcelo Amanco

Daniel Beck   ...t Bitvai   Rachel Cotterill   Nigel Dewdney

Ignatius Ezeani   Douwe Gelling   Michal Lukasik   Maryhilda McClure

Ikechukwu E Onyenw...   Gustavo Paetzold   Adam Poulston   Roland Roller

Nattapong Sanchan   Yunita Sari   Karin Sim Smith   Xingyi Song

David Steele   Konstantin Yershov

Admin Support (3)

Lucy Moffatt   Joanne Suter   Alice Tucker

13 May 2015

One of the largest Language Processing/Computational Linguistics Groups in Europe

Selected Current Research Projects

Information access

- DecarboNET: A Decarbonisation Platform for Citizen Empowerment and Translating Collective Awareness into Behavioural Change (Bontcheva, Cunningham; EU FP7)
  - Industrial/NGO partners: WWF, Switzerland; Green Energy Options, Cambridge

- SENSEI: Mining information and sentiment from conversational data from call centres/reader comment in on-line news (Gaizauskas, Hepple; EU FP7)
  - Industrial partners: Teleperformance, Italy; Websys, Barcelona; *TheGuardian*

- ForgetIT: Concise Preservation by combining Managed Forgetting and Contextualized Remembering (Cunningham; EU FP7)
  - Industrial partners: dkd Internet Service GmbH; Eurix Srl; IBM Israel; Turk Telekom

Selected Current Research Projects

Information access

- KConnect: Khresmoi Multilingual Medical Text Analysis, Search and Machine Translation Connected in a Thriving Data-Value Chain (Roberts, Bontcheva; EU H2020)

  - Industrial partners: Precognox Informatikai KFT, Hungary; Findwise AB, Sweden; Ontotext AD, Bulgaria; Trip Database Ltd

- PHEME: Computing Veracity Across Media, Languages, and Social Networks (Bontcheva, Hepple; EU FP7)

  - Industrial partners: Ontotext AD, Bulgaria; ATOS, Spain, SA; iHub, Nairobi; SwissInf.ch, Switzerland

- SenseMaking: Information Processing and Sensemaking: An Exploratory Search System for Document Collections (Stevenson; DSTL)

- Individual Profiling using Text Analysis (Stevenson; US DoD)

Selected Current Research Projects

Language technology integration platforms

- MLi: Towards a MultiLingual Data Services infrastructure (Cunningham; EU FP7)
  - Industrial partners: INMARK, Madrid; ATOS, Spain; ESTeam AB; Tilde, SIA

- OpenMinTed: Open Mining INfrastructure for TExt and Data (Roberts, Bontcheva; EU H2020)
  - Industrial/NGO partners: Agro Know I.K.E, Greek Research And Technology Network S.A., Frontiers Media SA

Selected Current Research Projects

Machine Translation

- Cracker: Cracking the Language Barrier: Coordination, Evaluation and Resources for European MT Research (Specia; EU H2020)

- Quality Translation 21 (Specia; EU H2020)
  - Industrial partners: TAUS BV, text&form GmbH, Tilde Sia

- EXPERT: EXPloiting Empirical appRoaches to Translation (Specia, EU FP7 Marie Curie ITN)

- MoDiST: Modelling Discourse in Statistical Machine Translation (Specia; EPSRC)

Selected Current Research Projects

Foundational Topics

- ViSEN: Generating image descriptions using computer vision analyses and large-scale text corpus analysis (Gaizauskas, Clough: EPSRC/CHIST-ERA)

- Distinguishing Common and Proper Nouns (Stevenson; Google)

- uComp: Embedded Human Computation for Knowledge Extraction and Evaluation (Peters, Bontcheva, Hepple; EPSRC/CHIST-ERA)

- KDisc: Language Processing for Literature Based Discovery in Medicine (Stevenson, Gaizauskas with Oncology & Neuroscience, EPSRC)

http://nlp.shef.ac.uk/projects.html

What we can offer to industry?

- **Expert know-how --** in all areas of language processing, but especially:
  - Text Analytics
    - ∗ Over: Social media; news; intranet data; clinical records; scientific and technical literature
  - Machine Translation
  - Language Engineering
    - ∗ Construction of robust, multilingual solutions for real world data and and real world platforms
- **Open Source Software**
  - GATE (General Architecture for Text Engineering)
    - ∗ an integrated development environment for language engineering solutions
  - Many other tools produced via funded research projects

Industrial Collaborators Past & Present

- Alinari
- Ansys
- Agro Know I.K.E
- Asplan Viak
- ATOS SA
- BBC
- CNN
- CRIC
- CTI
- Dassault Aviation
- Financial Times
- Findwise AB
- Fizzback
- Foodity
- France Telecom SA
- Frontiers Media SA
- Greek Research and Technology Network S.A.

- Green Energy Options Ltd
- GlaxoSmithKline
- Hardick Fintrade
- Health on the Net
- I-HUB Ltd.
- In & Out S.p.A. a socio unico
- Innovantage
- i.sieve
- Internet Memory Research,
- Loquendo SPA
- MDR partners
- Ontoprise GMBH
- OntoText AD
- Philips
- Precognox Informatikai KFT
- Quinary Spa
- Schweizerische Radio-und Fernsehgesellschaft Association

- Siemens
- SingularLogic
- Starlab
- TAUS BV
- Telefonica
- text & form GmbH
- Text Mining Solutions
- TheGuardian
- The Press Association
- TheWriter
- Tilde, SIA
- Trip Database Ltd
- Vodafone
- Websays SL
- Xerox
- Yahoo