

Director's Column



Prof Josef van Genabith
Director, CNGL

"Welcome to the Q2 2011 edition of CNGL News, the quarterly newsletter of the Centre for Next Generation Localisation (CNGL).

The second three months of 2011 have witnessed some significant events at CNGL, most notably the first successful PhD thesis defence for the Centre in June. Congratulations to Yifan He for making CNGL history! Yifan's achievement was emulated by Rejanwul Haque and Sergio Penkale last week. With many CNGL scholars now in the advanced stages of their PhD research, we look forward to many more PhD defences soon."



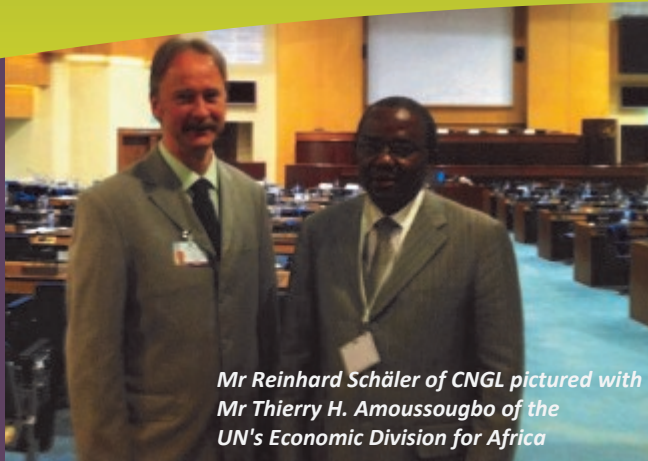
Yifan He



Rejunwal Haque



Sergio Penkale



Mr Reinhard Schäler of CNGL pictured with Mr Thierry H. Amoussougbo of the UN's Economic Division for Africa

this issue

- Headline News **P.1**
- Partnerships & Commercialisation **P.2**
- Education & Outreach **P.3**
- Research Track Updates **P.4 - P.5**
- CNGL People **P.6 - P.7**
- Conferences & Publications **P.8 - P.10**
- Affiliated Projects **P.11**
- Upcoming Events **P.12**

Taking Localisation to a Global Stage!

Reinhard Schäler addresses United Nations

In recognition of his work in spearheading attempts to bridge the global "digital divide", Mr Reinhard Schäler of CNGL at University of Limerick was invited to address the United Nations Economic and Social Council in May.

Reinhard spoke at the Second Meeting of the Committee on Development Information, Science & Technology (CODIST-II) of the UN's Economic Commission for Africa, which took place from 2-5 May in Addis Ababa, Ethiopia. Reinhard addressed more than 1,000 senior officials and experts from African member states on the



theme 'Localisation as an Industrial Strategy'.

Speaking from the CODIST-II meeting in Addis Ababa, he said "Being invited to address the United Nations Economic Commission for Africa at CODIST-II is an honour for me and it highlights the international standing of Centre for Next Generation Localisation and spin-off not-for-profit organisation, The Rosetta Foundation. My presence here reflects both Ireland's leadership position in localisation and the economic and social impact of investment in fourth-level research in Ireland."

[Read full story](#)

Sharon O'Brien at European Commission

Dr Sharon O'Brien, researcher in Translation Studies with CNGL at DCU SALIS, presented to some of Europe's leading translators at the invitation of the Directorate-General for Translation of the European Commission in April.

Sharon's talk, on the theme of post-editing machine translation output, was given at the 'Translators and Machine Translation' conference hosted by DGT in Brussels, Belgium. 80 translators from the

Commission, Parliament and Council attended, while others joined by video link from Luxembourg.



EUROPEAN COMMISSION

Sharon was one of only two external speakers invited to present at the conference — a fact that reflects, yet again, the high esteem in which CNGL is held by the international translation and localisation communities.

[Read full story](#)

APPLIED LANGUAGE SOLUTIONS



CNGL Partner Applied Language Solutions set to turn MT theory into reality with growing division

CNGL's Commercial Development Manager, Steve Gotz, added recently to

our set of industry collaboration case studies. In the spotlight this time was CNGL industry partner, Dai Nippon Printing. Why not check out the case study for yourself!

[View the case study](#) (pdf - 327kb)

Following his recent arrival at language services specialist, Applied Language Solutions, Professor Andy Way, computational linguistic expert and president of the European Association for Machine Translation, highlights some of the developments in machine translation technology taking place at the business: "ALS has been investing heavily in its machine translation research and development expertise, most recently through the additional recruitment of three Postdoctoral researchers from my MT group at CNGL to join the language technology team. They have been brought in

to use their in-depth knowledge of statistical MT technology to lead the company's development of customised, real-time MT engines.

"The growth of the team gives ALS the expertise and the research capability to truly become a leading authority on machine translation and will help to maintain its position at the cutting edge of technology developments.

"Having worked with ALS over the last few months in a

consultancy role it quickly became obvious that the company really knows what it wants to achieve in terms of a high quality machine translation offer, in a space

"The growth of the team gives ALS the expertise and the research capability to truly become a leading authority on MT"

where there is still a lot of debate. I am excited about what we are going to be able to develop in a short space of time with the fantastic team we have in place and funding from the EU."



VistaTEC hosts Silicon Valley Localization Unconference

CNGL industry partner VistaTEC presented its latest Localisation Unconference on 29 April 2011 in Silicon Valley.

VistaTEC partnered with Salesforce.com to bring together industry heavyweights for open discussions ranging from Localization Quality, Machine Translation, and Community Translations, to "What makes a good Localization Partner". Participating companies included Cisco, Apple, Salesforce.com, PayPal, and Adobe, to name a few.

Participants learned best practices from their peers, as well as discussions of where they'd like to see the industry in the future. Innovation, process, and the ability to measure success were main points across each of the topics. With over 60 participants, ideas flowed and solutions were created.

"By coming together in this non-competitive environment, participants felt empowered to share ideas and knowledge in an uncommon way", said Scott Schwalbach, Director of Global Solutions at VistaTEC.

CNGL and IIIT Hyderabad pledge to erode English—Indian Language Barrier

Pioneers at CNGL and the Language Technologies Research Centre (LTRC) at the International Institute of Information Technology Hyderabad (IIIT Hyderabad) will collaborate to build accurate translation systems from English to Indian languages.

Researchers at the institutions, which are two of the world's major language technology research centres, have committed their joint efforts to enhance the quality of translation from English to

the major languages of India.

This new collaboration on English-Indian translation systems will continue the mutually-beneficial partnership on localisation R&D between Ireland and India – two significant players on the localisation world stage. The project will focus on syntax-driven approaches for statistical machine translation into Indian languages.



IIIT, HYDERABAD

In addition to core funding from Science Foundation Ireland, the work will be supported by the Government of India through its India-Ireland Cooperative Science Programme and by CNGL academic partner Dublin City University through its DCU India Fund.

"This collaboration will continue the mutually-beneficial partnership on localisation R&D between Ireland and India"

Education and Outreach

LRC Summer School is a sell-out success!

The 2011 LRC Summer School in June provided unique insights into the exciting world of games localisation



The 2011 LRC Internationalisation and Localisation Summer School took place at the University of Limerick from 30 May to 2 June 2011. Computer and Video Game Localisation was the theme of the event and, over the four days, numerous industry experts contributed their time and expertise to ensure that the Summer School was a great success.

This special event, supported by CNGL, provided 38 eager participants with a unique opportunity to experience a different side of the localisation industry and gain

insight into one of the most exciting and fastest growing areas of the localisation industry: games localisation.

Representatives from Big Fish Games, Enzyme Testing Labs, Guerilla Translations, Keywords International, Language Automation Inc., Microsoft Games Studios, Mi'pu'mi games and OnLegends gave talks and workshops that were both interesting and engaging.

These speakers gave participants an insider view of the processes behind the scenes

of the game industry. They also demonstrated how the different elements of the localisation process are carried out at some of the most innovative companies in the industry.

"Participants got an insider view on the processes behind the scenes of the game industry"

The LRC and CNGL thank both the speakers and participants of the Summer School for making it such an interesting and exciting event. Special thanks go to Kilgray Translation Technologies and PopCap Games for their support of the Summer School.

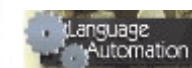
[View Presentations](#)



Microsoft



ONLEGENDS
Turning Local Into Global



Team Ireland prepares for International Linguistics Olympiad

Four of Ireland's top young linguists are gearing up to represent their country at the International Linguistics Olympiad, which takes place later this month in Pittsburgh, Pennsylvania.

Alec Fair and Medbh Campbell (both of Methodist College Belfast), Niamh Dhondt (Loreto Kilkenny) and Declan Manning (Ballinacollig Community School) will test their wits against some of the world's toughest puzzles in language and linguistics at Carnegie Mellon University from 24-30 July 2011. The

students will compete in individual and team rounds against budding linguists from around the world in the hope of being crowned World Champions.

The four members of Team Ireland were selected due to their strong performance in the national finals of All Ireland Linguistics Olympiad (AILO), which is coordinated by CNGL. The team's kit will feature the new AILO logo (seen right).

www.cngl.ie/ailo



Medbh Campbell of Team Ireland pictured with AILO Coordinator, Harold Somers

Introducing the new AILO logo...



UNDERGRADUATE INTERNSHIP PROGRAMME 2011

Recruitment for CNGL's summer 2011 Undergraduate Internship Programme is now at an advanced stage, with a number of interns already in place. A total of 13 internships are available.

The primary aim of this programme is to offer undergraduate students the opportunity to participate in and contribute to exciting research projects at CNGL. The goal is to inspire these students to take the first step on a path to a research career.

The internship programme runs annually, offering paid 8-week undergraduate internships based in our member institutions and industry partners. We're looking forward to working with the talented new recruits over the summer!

<http://www.cngl.ie/internships.html>

LOCALISATION CAREERS BROCHURE

CNGL is delighted to report that its new localisation careers brochure will be published the week commencing 4 July. This publication aims to inform secondary school students, teachers and guidance counsellors about the exciting careers available in localisation and related areas. The brochure features profiles from industry professionals, insights into the variety of job roles and industries that rely on localisation, and details of relevant third level courses. It will be distributed to secondary schools across the country in September. Many thanks to all who contributed!

EVALUATION GROUP

First of all, a big thanks to all CNGL researchers for responding so completely to the recent requests for Evaluation Metrics and User Studies data.

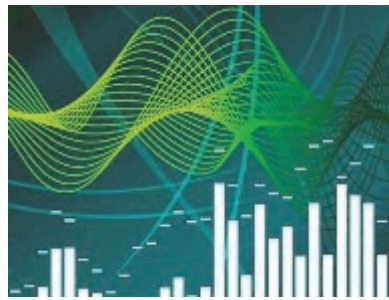
The Evaluation Metrics Taxonomy Development study is now complete, and the results will be used to drive the Demonstrator Evaluation process. The data gathered is also available on the Evaluation Group wiki and is a useful reference for metrics and their usage across CNGL. Indeed, the Evaluation Group wiki is currently ranked third most viewed CNGL wiki page, with 1,210 views to date!

The User Studies Survey is now complete, the aim being to assess the extent to which user studies were being used to inform and/or evaluate CNGL Researchers' work. There have been 16 individual studies completed to date, involving 369 participants, 223 external to CNGL, 146 internal.

From a publications perspective, a number of cross-strand evaluation papers are currently being prepared for submission by the group, based on the findings of the recent Evaluation Group studies.

- Dr. Ian R O'Keefe
(UL-LOC) Chair

CNGL Research Track Updates



As ILT prepares for the departure of Track Leader, Andy Way, the track is celebrating the considerable achievements of its researchers this quarter.

Yifan He (ILT1) became the first CNGL student to complete his PhD defence in June. Yifan is currently on internship with Adobe in California. Rejwanul Haque (ILT1) and Segio Penkale (ILT1) defended their theses last week.

Sandipan Dandapat (ILT1) has successfully transferred from PhD-Track to PhD and he has commenced his three-month internship at Microsoft Research in Redmond, Washington. Johann Roturier of Symantec has

Integrated Language Technologies (ILT)

Three successful PhD defences in June have rounded off a highly productive quarter for ILT

been approved as co-supervisor of Pratyush Banerjee, PhD candidate with ILT1.

The quality and volume of work produced by ILT is evident in the track's 8 papers presented at EAMT-11.

Efforts continue to find a replacement for Track Leader, Andy Way, who is moving to CNGL partner Applied Language Solutions (ALS). Andy, Páraic Sheridan (CNGL Operations Director) and Stephen Roantree (CNGL IP Manager) visited ALS in the UK this quarter to discuss opportunities for

future collaboration.

Andy Way and Harold Somers represented DCU at the Intensive Summer School in Machine Translation in Chelyabinsk, Russia in May.

"The quality of work

produced by ILT is evident in

the track's 8 papers

presented at EMAT-11"

Funding from DCU's Ireland-India Fund and the Government of India's India-Ireland Cooperative Science Programme will facilitate collaboration of CNGL with IIIT Hyderabad on English-Indian translation systems.

Next Generation Localisation (LOC)

This quarter has seen commercial developments, new faces, and further work on the SOLAS platform

Development of the Service Oriented Localisation Architecture Solution (SOLAS) continues, and discussion with industrial partners on component integration is on-going.

Lamine Aouad (LOC3 Postdoctoral Researcher) has left CNGL and is replaced by Eoin Ó Conchúir, who was previously Web Developer with LOC2.

LOC is successfully advocating for XLIFF as a pivotal standard on international forums. It has secured

funding for LT standardisation within W3C.

David Filip, a world-leading expert on localisation standards, is now in situ as a postdoctoral researcher at UL. David presented a talk on XLIFF 2.0 development and LT standardisation areas at the MultilingualWeb workshop in Pisa in April.

The Evaluation Group and User Studies Group, which LOC coordinates, are making a large



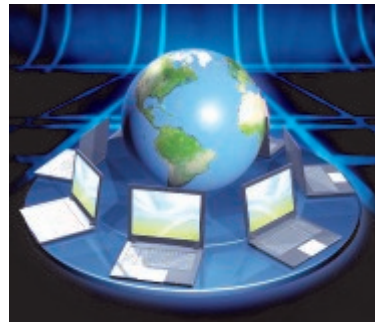
contribution to CNGL, both in terms of data gathering and in nurturing community spirit.

On the commercialisation front, LOC has filed an invention disclosure with UL's Technology Transfer Office and a software disclosure with CNGL's IP Committee for Translation eXchange. CNGL spin-out activity The Rosetta Foundation is to hire three interns.

"LOC is advocating for XLIFF as pivotal standard on international forums"

Systems Framework (SF)

Publication success, PhD transfers and joint CNGL-industry funding proposals have been the focus for SF in Q1



The strong international focus of SF research is emphasised by its on-going collaboration with the National Institute of Informatics (NII) in Tokyo on Natural Language Generation in virtual environments. SF members are also working closely with University of Ulster on preparation of a funding proposal for Health Assistance through Multimodal Interaction and Entertainment (HAMITE).

Further progress in funding

diversification includes submission of 2 FP7 SME proposals and the finalising of negotiations with KnowHoIEM for FP7 (value to TCD €279k). Negotiations continue with two further partners on FP7 calls (total value to TCD of €645k).

Industry collaboration this quarter includes engagement with Indysis and Tesco (the latter as part of the HAMITE proposal).

Ilana Rozanes (SF1) has passed

her PhD confirmation hearing and is interviewing medical interpreters in Irish and US hospitals.

“Industry collaboration includes engagement with Indysis and Tesco”

It has been a busy time on the publications front, with 2 papers

accepted for publication, 1 paper accepted for presentation, and 4 papers submitted for upcoming conferences.



Digital Content Management (DCM)

Completion of the Personalised Multilingual Information Retrieval Demonstrators, conference hosting, and publication success were the major themes of Q1 for DCM

Quarter 2 has seen the DCM track finalise development of the Personalised Multilingual IR Demonstrators. Scalability testing of a second prototype of the SOCOM framework has also been completed. Collaboration with Symantec and Microsoft on multilingual and cross-lingual personalisation has continued.

DCM hosted a number of international events this quarter. These include the IEEE Integrated Management Symposium 2011 hosted in May by TCD. The symposium was co-chaired by

DCM Track Leader Vincent Wade. (Patrons Chair: Dave Lewis; Local Chair: Declan O’Sullivan). In April Gareth Jones served as Co-Chair of ECIR 2011 in Dublin.

Séamus Lawless and Johannes Leveling served as Workshop Chairs at the Multilingual Personalisation Workshop at ACM Hypertext 2011 in Eindhoven in June. Vincent Wade delivered the keynote presentation at Multilingual Personalisation. An Evaluation Track proposal by Debasis

Ganguly, Johannes Leveling and Gareth Jones has been accepted for FIRE2011.

Publication continues apace with 7 papers submitted for publication, 13 papers accepted for publication, 6 papers published and 1 edited proceedings published.

DCM PhD scholars are making good progress, with Bo Fu of DCM2 completing the first draft of her PhD thesis.

The Meta-data group has been busy building semantic models of CNGL from the complete set of seventeen demonstrators showcased at Microsoft in 2010.

These models describe the CNGL Demonstrators as services, content and content type properties. They have been produced using an interview approach with each of the demonstrator team leads. During each interview the demo system in question was compared against seed models of services and content. The similarities and differences of the systems were recorded and fed back into the models, building up a more accurate semantic meta-model of the CNGL piece by piece.

Results from the interviews are currently being prepared for publication, along with the V1 semantic models. Initial analysis has shown that the participants understood the purpose, process, rationale and approach taken in the experiment. It was necessary to extend upon the seed models for several components, and this was achieved using an iterative approach incorporating each demo team’s input to the models.

This modelling process will, in future, be extended to new iterations of the Demonstrator components. This will involve assessing the semantic evolution of the meta-data in the systems, to include new meta-data and track changes in its current representation.



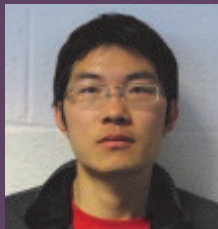
Pawel Plesniak
Graphic Designer



Sophie Mataboro
Centre Administrator



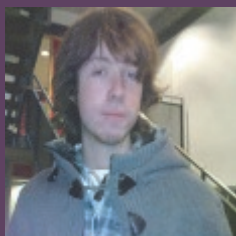
Zhaopeng Tu
(Chinese Academy of Sciences)
Visiting Student, Euromatrix+



Jun Xie
(Chinese Academy of Sciences)
Visiting Student, Euromatrix+



Daqi Zheng
(Chinese Academy of Sciences)
Visiting Student, Euromatrix+



Kieran O'Brien
Undergraduate Intern
TCD

CNGL People



Featured Researcher:

SOLOMON GZAW (ሰለሞን ግዛው)

Solomon Gzaw, PhD scholar with LOC3 at UL, outlines his research priorities and explains some of the challenges facing his native tongue, Amharic...

"I am doing my PhD on the "Taxonomy of Personalisation for generating Personalised Content" with Dr Jim Buckley in the Localisation Research Centre (LRC) at the University of Limerick (UL).

I am originally from Ethiopia, a country located in the Horn of Africa and the second-most populous nation in Africa, with over 85.2 million people. Amharic (አማርኛ amareñña) the official working language of Ethiopia is a Semitic language and is the second most-spoken Semitic language in the world, after Arabic. It is written using Amharic Fidel, ፊደል ("alphabet", "letter," or "character").

Amharic: An Under-resourced Language

There is no other foreign language spoken in Ethiopia as a first language. English is used as a medium of instruction in high schools



and universities; however, most of the population have difficulties using English or other foreign languages. It is widely recognised that, without embracing information and communication technologies (ICT), the development gap between Africa, Asia and the rest of the world will continue to expand.

Even though *“Without embracing ICT, the development gap between Africa, Asia and the rest of the world will continue to expand”* Ethiopia that speaks

Amharic is far greater than that of some European countries, the majority of publishers and service providers haven't shown interest in localising into languages such as Amharic. Under-resourced languages such as Amharic are viewed as not being viable to the market and as such do not generate enough Return on Investment (ROI). However, recently some organisations such as Google and Microsoft etc. have started considering these languages. As a result of this, I have become an Amharic language specialist in Google (part-time).

A Focus on Personalisation

Conscious that language translation is not always enough for optimal understanding/ dissemination of information with respect to locale, I become interested in personalisation.

My PhD research will attempt to design a ranked taxonomy of personalisation

issues for characterising and generating personalised content in a customer care scenario. The work will concentrate on empirically identifying and ranking the attributes that need to be considered in an inductive fashion from technical support forums. These may include cultural preferences, social norms, experiences, emotions, motivation, and expectations.

Being part of the CNGL has given me the opportunity to observe what is happening in the world in Localisation. I feel blessed to be able to learn and be part of the most important research centre in localisation."

Featured Researcher:

YIFAN HE

Yifan He made history in June, when he became the first CNGL student to complete his PhD viva. Here, he explains that industrial collaboration has been central to his research...



"I recently passed my PhD defense at Centre for Next Generation Localisation (CNGL), School of Computing, Dublin City University, under the supervision of Prof. Andy Way and Prof. Josef van Genabith. My PhD thesis focuses on techniques that integrate Statistical Machine Translation (SMT) and Translation Memory (TM). During my PhD study, I also worked on automatic evaluation methods for MT and MT tuning.

I am currently doing an internship at Adobe Systems in San Jose,

supervised by Dr. Ray Flournoy, working on enterprise applications of MT and low-resource MT. After that, I will return to DCU as a post-doctoral researcher for the European-MatrixPlus project.

I believe that the unique academic-industrial alliance is essential for the research reported in my PhD thesis to happen: the need, the data,

and the support received from the industrial partners of CNGL (especially Symantec) all helped to shape this thesis; at the same time, it is the vision, the experience, and the encouragement from professors Andy and Josef, and the excellent colleagues such as Yanjun Ma that have guided me to complete this research."

"The unique academic-industrial alliance was essential to my PhD thesis"

Where are they now? PRASENJIT MAJUMDAR



"During my time at CNGL I was a member of the DCM2 team at University College Dublin. I moved on from CNGL in late 2010 to take up my current position as Assistant

Professor of DAIICT (www.daiict.ac.in). I co-coordinate the FIRE (www.isical.ac.in/~fire) workshop and am the founder secretary of Information Retrieval Society of India (www.irs.res.in). I received a research grant of approximately USD 160,000 from the Department of Information Technology, Government of India, for the FIRE evaluation

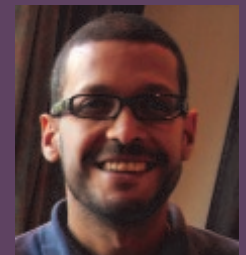
campaign. The project started from October 2010 and will continue till 2013 September. I served as guest editor for a special issue on Indian Language Information Retrieval in ACM TALIP (volume 9 issue 3 and 4). CNGL was a very useful exposure for me. Some of my fondest memories are of my CNGL days and I wish to collaborate again in the future."

Staff News



It's a boy!

Congratulations to Joss Moorkens, PhD scholar based at DCU, on the birth of his beautiful son, Evan Þór Moorkens, in April 2011.



Bon voyage!

Best wishes to Lamine Aouad, LOC Research Fellow, who departed from UL recently.



Good luck!

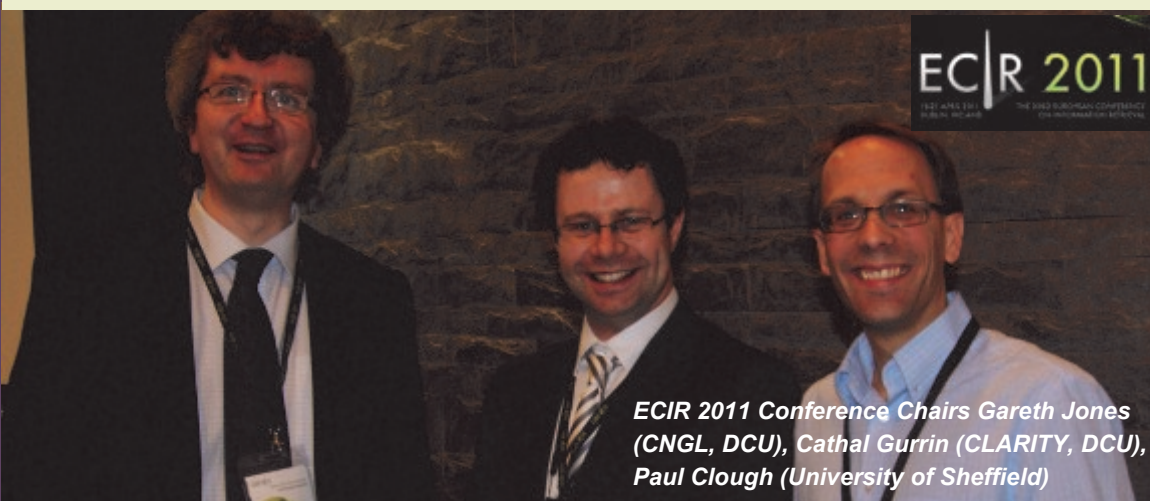
Best of luck to Séamus Lawless, DCM Postdoctoral Researcher at TCD, on his upcoming nuptials. We wish Shay and his fiancée Pamela McDermott a long and happy marriage!

Download copies of our publications at:
www.cngl.ie/researchpub.html

- Morera, A. and L. Auouad (2011). *Towards an intelligent localisation workflow management system.*
- Morado Vázquez, L., Torres del Rey, J. (2011). *The relevance of metadata during the localisation process – an experiment.*
- Morado Vázquez, L. (2011). *Defining localisation and its relation with translation studies.*
- Van der Sluis, I., Luz, S. (2011). *Issues in translating and producing Japanese referring expressions for dialogues.*
- Javed, M. and Abgaz, Y.M. and Pahl, C. (2011). *A Layered Framework for Pattern-Based Ontology Evolution.*
- Nishio, N., Anastasiou, D., O’Keeffe, I., Tarquini, G., Collins, J.J., Schäler, R. (2011). *Accessing Multilingual Public Service Information through Localisation.*
- Maldonado-Guerra, Alfredo, and Emms, Martin (2011). *Measuring the compositionality of collocations via word co-occurrence vectors: Shared task system description.*
- Melike Sah and Vincent Wade (2011). *Automatic Mining of Cognitive Metadata using Fuzzy Inference.*
- Ben Steichen, Alexander O’Connor and Vincent Wade (2011). *Personalisation in the wild: providing personalisation across semantic, social and open-web resources.*
- Ghorab, M. Rami, Zhou, Dong, Steichen, Ben, and Wade, Vincent (2011). *Towards Multilingual User Models for Personalized Multilingual Information Retrieval.*
- D. Ganguly, J. Leveling and G.J.F. Jones (2011). *Query Expansion for Language Modeling using Sentence Similarities.*
- Debasis Ganguly, Johannes Leveling, Wei Li, Gareth Jones (2011). *Towards Evaluation of Personalized and Collaborative Information Retrieval.*
- Ankit Srivastava, Yanjun Ma, and Andy Way (2011). *Oracle-based Training for Phrase-based Statistical Machine Translation.*
- Hala Almaghout, Jie Jiang, Andy Way (2011). *CCG Contextual labels in Hierarchical Phrase-Based SMT.*
- Antonio Toral, Federico Gaspari, Sudip Kumar Naskar, and Andy Way (2011). *Comparative evaluation of research vs. online MT systems.*

Conferences and Publications

CNGL co-hosts ECIR 2011: The 33rd European Conference on Information Retrieval



ECIR 2011 Conference Chairs Gareth Jones (CNGL, DCU), Cathal Gurrin (CLARITY, DCU), Paul Clough (University of Sheffield)

CNGL academic partner, DCU, co-hosted ECIR 2011: The 33rd European Conference in Information Retrieval on 18-21 April at the Radisson Blu Hotel in Dublin city centre. There was a strong CNGL presence in the conference line-up.

ECIR is the main annual European forum for the presentation of new research results in the field of Information Retrieval, and is one of the leading international meetings for information retrieval attracting delegates from across the world. In 2011 the conference was organised by DCU in co-operation with the University of Sheffield, supported by BCS IRSG and ACM SIGIR.

Dr Gareth Jones of CNGL at DCU was Programme Co-Chair, while

Dr Johannes Leveling (DCU) and Dr Dong Zhou (TCD) served on the Programme Committee.

The first day of the conference was held at the Guinness Storehouse. Daytime activities were devoted to a set of specialist tutorials and workshops, followed by a Welcome Reception for all delegates in the Storehouse’s Gravity Bar with its spectacular view over Dublin. Delegates got to see more of the Dublin region at the conference banquet held at the Village at Lyons in Kildare.

The conference consisted of a three keynote presentations, oral paper presentations and an evening poster session.

CNGL representation at the poster session included a poster by Walid Magdy (DCU) and Gareth Jones asking ‘Should MT Systems be Used as Black Boxes in CLIR?’, a second poster by Walid and Gareth presenting joint work with Patrice Lopez of INRIA - Humboldt University, Germany on ‘Simple vs. Sophisticated Approaches for Patent Prior-Art Search’ and a poster by Johannes Leveling presenting joint

work with Giorgio Maria Di Nunzio of University of Padua, Italy and Thomas Mandl of University of Hildesheim, Germany on ‘Multilingual Log Analysis: LogCLEF’.

CNGL was also well represented in the workshop programme. Liadh Kelly (DCU) was Programme Co-Chair for the EPS 2011: Evaluating Personal Search Workshop. CNGL also presented work at the Workshop on Information Retrieval Over Query Sessions. Debasis Ganguly (DCU) with Johannes and Gareth presented a paper on ‘Automatic Generation of Query Sessions using Text Segmentation’. Johannes and Gareth also gave a paper on the theme ‘Same Query - Different Results? A Study of Repeat Queries in Search Sessions’.

The conference closing session featured an invitation to delegates to come back for Dublin’s next major involvement in information retrieval conferences when we welcome SIGIR, the premiere international conference in information retrieval, to Dublin in 2013.

For further information on ECIR, 2011, see www.ecir2011.dcu.ie.

CNGL at LocWorld 2011

CNGL permeates conference and workshop agenda at Localization World Barcelona, 14-16 June 2011

The reach and influence of CNGL was evident last month as the Centre's academic and industry partners played a central role in Localization World Barcelona, one of the largest conferences in the localisation professional's calendar. The conference theme was 'Innovation'.

Sharon O'Brien of CNGL at DCU co-presented a pre-conference workshop on 'Best Practices in Post Editing' with Rahzeb Choudhury of TAUS. The half-day workshop provided an overview of machine translation post editing, good practice guidelines, and recent research and developments in post editing. Also delivering pre-conference workshops were Gráinne Maycock of VistaTEC and Derek Coffey of Welocalize.

Day One saw Sophie Hurst of CNGL partner SDL on the panel for 'Global Marketing Communications: Bringing Order to Chaos'. Iris Orriss of Microsoft and Allan Hall of SDL were 'Bridging the Gap Between the



Client and Language Service Provider', and Mark Flanagan of VistaTEC was a panellist for 'Beyond Marketing: Driving Brand Integrity and Market Share across Cultures, Languages and Locales'.

David Filip of LRC/CNGL was a member of the panel that asked 'Standards. We Know Where We've Been, But Where are We Going?' With the recent demise of

LISA, this session highlighted that there are critical activities necessary for the adoption, promotion, enforcement, enhancement and auditing of standards to ensure their utility and worth.

Day Two opened with Yvan Hennecart of SDL addressing the top three things to master in a localization strategy - 'People, Process and Technology.' Salim Roukos and Saroj Vohra of IBM and Smith Yewell of Welocalize opened the TAUS 2011 Series of MT Debates. Pedro Gomez of Microsoft presented 'Localization by the Numbers', and IBM presented valuable insights into the value of post-editing and MT experiences at their company.

In 'Innovation Bottom Up', Fred Hollowood and Martin Roche of Symantec shared their experiences with their firm's in-house innovation programme.

In the exhibition area, CNGL's booth attracted a large number of visitors and potential collaborators. A representative from IDA joined Operations Director Páraic Sheridan, IP Manager Stephen Roantree, and Commercial Development Manager Steve Gotz at the CNGL booth.



PUBLICATIONS

- **Morrissey, Sara (2011).** *Assessing Three Representation Methods for Sign Language Machine Translation and Evaluation.*
- **Fu B., Brennan R., O'Sullivan D (2011).** *Using Pseudo Feedback to Improve Cross-Lingual Ontology Mapping.*
- **Alexandru Ceausu, John Tinsley, Jian Zhang and Andy Way (2011).** *Experiments on Domain Adaptation for Patent Machine Translation in the PLuTO project.*
- **Sarah Ebling, Andy Way, Martin Volk, and Sundip Kumar Naskar (2011).** *Combining semantic and syntactic generalization in example-based machine translation.*
- **Way, Andy and Mary Hearne (2011).** *On the Role of Translations in State-of-the-Art Statistical Machine Translation.*
- **Christos Tsarouchis, Declan O'Sullivan and David Lewis (2011).** *A study in the expressiveness of semantically different policy modelling schemes.*
- **N. Campbell, J. Kane (2011).** *Processing 'yup' and other short utterances in interactive speech.*
- **Tsuyoshi Okita, Alexandru Ceausu, and Andy Way (2011).** *Statistical Machine Translation with Factored Translation Model: MWEs, Separation of Affixes, and Others.*
- **Tsuyoshi Okita and Andy Way (2011).** *Given Bilingual Terminology in Statistical Machine Translation: MWE-sensitive Word Alignment and Hierarchical Pitman-Yor Process-based Translation Model Smoothing.*
- **Van der Sluis, I., C. Mellish and G. Doherty (2011).** *Affective Text: Generation Strategies and Emotion Measurement Issues.*
- **Veale, T. and Li, G. (2011).** *Creative Introspection and Knowledge Acquisition: Learning about the world thru introspective questions and exploratory metaphors.*
- **Veale, T. (2011).** *Creative Language Retrieval: A Robust Hybrid of Information Retrieval and Linguistic Creativity.*
- **Magdy W., P. Lopez, and G. J. F. Jones (2011).** *Simple vs. Sophisticated Approaches for Patent Prior-Art Search.*
- **Debasis Ganguly, Johannes Leveling, and Gareth Jones (2011).** *Automatic Generation of Query Sessions using Text Segmentation.*

Sharon O'Brien of CNGL co-presents pre-conference workshop on Best Practices for Post Editing



CNGL academic partner institution, Trinity College Dublin (TCD), is this week (4-8 July 2011) hosting a five-day summer school, 'From Metadata to Linked Data'. The event, which is run jointly with the Digital Humanities Observatory (DHO), is a joint Irish Research Council for Humanities and Social Sciences (IRCHSS) COST (Interedition) training school.

The Training School will feature seminars and hands-on workshops by scholars in the field including Owen Conlan, Jennifer Edmond, and Séamus Lawless of TCD. [Read More](#)

CNGL SUPPORTS PHMR 2011

Researchers from CNGL's Digital Content Management track organised the First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011), which was held in Eindhoven on 6 June 2011. The workshop preceded ACM Hypertext 2011. Séamus Lawless of Trinity College Dublin and Johannes Leveling of DCU School of Computing were programme chairs at the workshop, while Alex O'Connor and Dong Zhou, both of TCD, served on the programme committee.

The PMHR workshop promoted the exchange of ideas between researchers working on hypertext and adaptive hypertexts, cross-lingual information retrieval, personalised search, personalisation for Web and hypertexts, and recommender systems. [Read More](#)



LRC Announces Annual Conference Co-location with Multilingual Web Workshop

The Localisation Research Centre (LRC) and the Multilingual Web project have announced that the third workshop in the Multilingual Web series will be co-located with the 16th Annual Internationalisation and Localisation Conference organised by the LRC in association with CNGL.

This joint event will run from 21-23 September 2011 at the Castletroy Park Hotel in Limerick, Ireland and aims to

bring together the world's leading experts in localisation, internationalisation and Multilingual Web design and conceptualisation for a series of presentations, talks and workshops. The event will give attendees the opportunity to learn about the latest advances and developments in the field of localisation, as well as provide them with the opportunity to participate and have their say in

workshops that will contribute towards the future of the Multilingual Web.

This year's LRC Conference will be the 16th edition of the popular annual event. The theme will be

Vertical versus Virtual - 21st Century Localisation Strategies.

Registration is now open.

Further information is available at www.localisation.ie and www.multilingualweb.eu

Resounding Success for CNGL at EAMT-

Such was the success of CNGL in the call for papers for EAMT-2011 that almost one in four papers at the prestigious conference was presented by CNGL. This tremendous achievement reinforces CNGL's position as a world leader in machine translation research.

EAMT-2011, the 15th annual conference of The European Association for Machine Translation, was held in Leuven, Belgium on 30-31 May 2011.

EAMT-2011 brought together all those – developers, researchers, users, professional translators, managers, marketing experts –

who have a stake in the vision of an information world in which language issues become transparent to the information consumer. The conference provided a forum for users to share their experiences, developers to describe what is happening in the internet marketplace, researchers to present new capabilities, and visionaries to describe the future.

The conference's User track featured two oral presentations from CNGL and one poster presentation. The Research track included an incredible four CNGL presentations and two posters.

CNGL also played a central role on the Programme Committee for EAMT-2011. Declan Groves, Jie Jian, Yanjun Ma, John Tinsley, Andy Way and Pavel Pecina, all of CNGL at DCU, served on the Research Track of the Programme Committee.

On the User Track Programme Committee, CNGL was represented by Fred Hollowood of industry partner Symantec, Dag Schmidke of industry partner Microsoft, and Andrew Bredenkamp of acrolinx, who is a member of CNGL's Scientific Advisory Board.

Well done to all concerned!



Affiliated Projects

In the spotlight:



META-NET

META-NET is a Network of Excellence dedicated to fostering the technological foundations of a multilingual European information society. CNGL is a founding member of the Network of Excellence, along with 12 other partners from around Europe.

In order to represent every EU country and official European language with a node in the network in each locality, META-NET has grown beyond the 13 founding members. Now, through cooperation with three other EU funded projects, the META-NET Network of Excellence consists of 47 members in 31 countries across the EU and greater Europe. In addition to this, META-NET has collaboration agreements with many other European projects

including CLARIN, FLaReNeT, PLuTO, Panacea and many more (www.meta-net.eu/collaborations).

META-NET's goals are to build a strong community around language technology encompassing all stakeholders including research, industry, users and language communities themselves; to enable greater interaction and sharing of information and resources within this community by providing an open resource exchange infrastructure called META-SHARE; and through innovative core research and community engagement, to build bridges with neighbouring fields of interest for mutual benefit.

Now in its second year of work,

CNGL is a founding partner of the META-NET Network of Excellence, which recently presented a Strategic Research Agenda for Multilingual Europe

META-NET is pushing ahead with the publication of many important reports on the state of the art in Language Technology and multilingualism in Europe.

A Strategic Research Agenda for Multilingual Europe was presented at the recent META-FORUM in Budapest, along with the publication of 31 Language Whitepapers covering every official EU language and several others as well as regional languages.

All of these publications, and more information, are available online from the project website www.meta-net.eu. For more information or to get involved contact John Judge at DCU.

META-VISION: Building a community with a shared vision and strategic research agenda

META-SHARE: Building an open resource exchange infrastructure

META-RESEARCH: Building bridges to neighbouring technology fields

REVIEW SUCCESS FOR AFFILIATED PROJECTS

Q2 2011 saw five CNGL-affiliated projects successfully complete their Year 1 reviews. Congrats to all involved with PLuTO, CoSyne, META-NET, Panacea, and EuroMatrix+

CNGL-AFFILIATED PROJECTS

Funding diversification is a major objective of CNGL. Below is a summary of CNGL-affiliated projects.

- EuroMatrixPlus
www.euromatrixplus.net
- Panacea
www.panacea-lr.eu
- CoSyne
www.cosyne.eu
- PLuTO
www.pluto-patenttranslation.eu
- META-NET
www.meta-net.eu
- Percolate
www.percolate.ie
- Khresmoi
www.khresmoi.eu
- AXES
www.axes-project.eu
- PetaMedia
www.petamedia.eu
- IM Real
www.imreal-project.eu
- AMAS
<http://kdeg.cs.tcd.ie/amaz>
- Multilingual Web ICT-PSP
www.multilingualweb.eu

Pavel Pecina wins Best Book Award



Congratulations to Dr Pavel Pecina of ILT at DCU for receiving the "Best book of MFF UK" award for 2011 from the Faculty of Mathematics and Physics at Charles University in Prague, Czech Republic. Pavel's book, which was based on his PhD thesis and published in 2009, is titled 'Lexical Association Measures (Collocation Extraction)'.

Upcoming Events

- **SFI Site Review**
11-12 July 2011, The Helix, Dublin City University, Ireland
- **From Metadata to Linked Data Summer School**
4-8 July 2011, Trinity College Dublin, Ireland [Further Details](#)
- **International Linguistics Olympiad (IOL)**
25-30 July, Carnegie Mellon University, Pennsylvania, USA [Further Details](#)
- **16th Annual LRC Conference: Vertical versus Virtual - 21st Century Localisation Strategies & 3rd Multilingual Web Workshop**
21-23 September 2011, Carlton Castletroy Park Hotel, Limerick, Ireland [Further Details](#)
- **2nd XLIFF Symposium**
28 September 2011, Warsaw, Poland [Further Details](#)
- **NCLT/CNGL Seminar Series 2011 (Every Wednesday)**

Commercialisation Factsheets now available!



CNGL's Commercial Development Manager has produced a series of technical summary sheets, which outline the features, benefits, technology status and IP status of CNGL technologies that are ripe for commercialisation.

[Download the summary sheets \(pdf-6.32MB\)](#)

Have you seen our new CNGL brochure?



In preparation for LocWorld 2011 in June, CNGL produced a new promotional brochure. The 6-page publication provides an overview of CNGL's main research priorities and opportunities for collaboration.

[Download your copy today \(pdf - 861KB\)](#)



The Centre for Next Generation Localisation (CNGL) is a dynamic academia-industry partnership that conducts ground-breaking research into next-generation localisation (i.e. the process of adapting digital content to culture, location and linguistic environment) by addressing key challenges in information volume, access and personalisation.



Got an idea for a news article? Perhaps you'd like to contribute?
Contact our editor, Laura Grehan, at Email: lgrehan@computing.dcu.ie or Tel: +353 (0)1 7006705