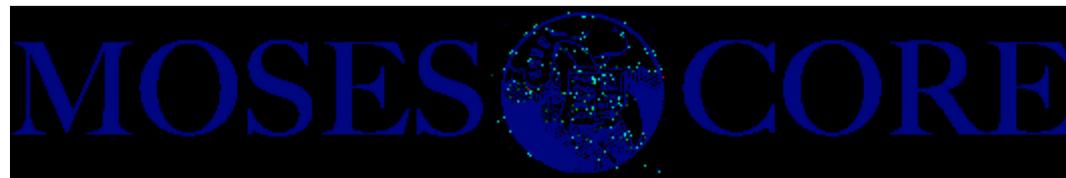# MT R&D in academia and industry: observations from the trenches

*Prof. Andy Way*
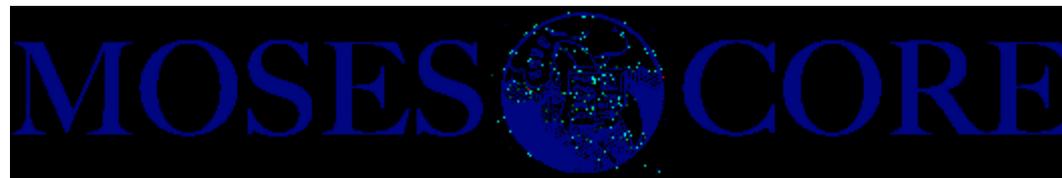
Director of Language Technology
Applied Language Solutions

andy.way@appliedlanguage.com

# **Currently**…

- On leave as Prof. Computer Science @ DCU

- Director of Language Technology @ ALS

- EAMT President 2009—to date (eamt.org)

- IAMT President 2011—13 (ending @ MT Summit)

- Editor of the *Machine Translation* journal

# EUROPEAN ASSOCIATION FOR MACHINE TRANSLATION

Home

## Welcome to the Web site of the EAMT

### What is the EAMT?

The European Association for Machine Translation (EAMT) is an organization that serves the growing community of people interested in MT and translation tools, including users, developers, and researchers of this increasingly viable technology.

### IAMT

The EAMT is one of three regional associations of the International Association for Machine Translation (IAMT). Its sister organizations are the Association for Machine Translation in the Americas (AMTA) and the Asian-Pacific Association for Machine Translation (AAMT).

### Activities

Together with the AMTA and the AAMT, the EAMT organizes workshops and conferences, such as the bi-annual MT Summit, and the annual EAMT conferences or workhops. Under the auspices of the IAMT, it also compiles listings of companies and products which are distributed free or at nominal cost to its members. The EAMT also maintains a mailing list, mt-list@eamt.org, as a public forum for the discussion of translation technology.

### Non-profit

## LATEST NEWS

**AMTA 2012: Workshop on Monolingual Machine Translation (MONOMT 2012)**

The **CFP for Workshop on Monolingual Machine Translation (MONOMT 2012)** to be held at the **AMTA-2012** in San Diego, California **28 Oct - 1 Nov 2012** [more]

**EAMT Best Thesis Award 2012**

The **European Association for Machine Translation**

# International Association for Machine Translation

**IAMT**

The EAMT is one of three regional associations of the International Association for Machine Translation (IAMT). Its sister organizations are the Association for Machine Translation in the Americas (AMTA) and the Asian-Pacific Association for Machine Translation (AAMT).

---

### IAMT Council 2009-2011

**President (also president EAMT)**
Andy Way
andy...@appliedlanguage.com

**President-Elect (also President AMTA)**
Alon Lavie
pres...@amtaweb.org

**Secretary (from AMTA)**
Nizar Habash
hab...@cs.columbia.edu

**Treasurer (from AMTA)**
Michelle Vanni

**President AAMT (ex officio)**
Hitoshi Isahara
isah...@nict.go.jp

### Other members:

Jun'ichi Tsujii *(AAMT Director)*
tsu...@is.s.u-tokyo.ac.jp

Shoichi Yokoyama *(AAMT Director)*

---

**LATEST NEWS**

**AMTA 2012: Workshop on Monolingual Machine Translation (MONOMT 2012)**

The **CFP for Workshop on Monolingual Machine Translation (MONOMT 2012)** to be held at the **AMTA-2012** in San Diego, California **28 Oct - 1 Nov 2012** [more]

---

**EAMT Best Thesis Award 2012**

The **European Association for Machine Translation** invites entries for its **fourth EAMT Best Thesis Award** for a PhD or equivalent thesis on a topic related to machine translation.[more]

---
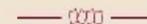
**Languages & The Media**

Languages & The Media: 9th International Conference & Exhibition on Language Transfer in Audiovisual

# Machine Translation Summit XIV, Nice, France
## 2 - 6 September 2013

The 14th MT Summit will take place in Nice, France at the ACROPOLIS conference centre.

Pre-Conference Workshops & Tutorials: Sept. 2 - 3, 2013
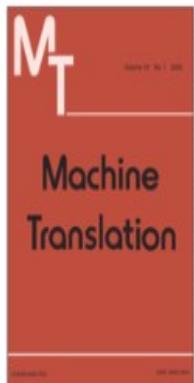
Main Conference: Sept. 4-6, 2013

More information about the conference is coming up here in due time.

*(contact: org@mtsummit2013.info)*

http://www.mtsummit2013.info/

# » Artificial Intelligence

Home > Computer Science > Artificial Intelligence

SUBDISCIPLINES | JOURNALS | BOOKS | SERIES | TEXTBOOKS | REFERENCE WORKS

## Machine Translation

Editor-in-Chief: Andy Way

ISSN: 0922-6567 (print version)
ISSN: 1573-0573 (electronic version)

Journal no. 10590

Read online

▸ ≔    RECOMMEND TO LIBRARIAN

f Like  ‹17    ▾ Tweet ‹2    g +1  1

ABOUT THIS JOURNAL    EDITORIAL BOARD    CALL FOR PAPERS

OPEN ACCESS    **Springer Open Choice**
Your Research. Your Choice.    CLICK

▹ Covers all branches of computational linguistics and language engineering, especially those that incorporate a multilingual aspect.
▹ Examines theoretical, descriptive as well as computational aspects.
▹ Features coverage of a wide range of topics, from multilingual text composition and generation to the history of machine translation.

### READ THIS JOURNAL ON SPRINGERLINK

Online First Articles

Current Issue

Free: Most Downloaded Articles

### FOR AUTHORS AND EDITORS

Aims and Scope

Submit Online

Open Choice - Your Way to Open Access

Instructions for authors

Language

### SERVICES FOR THE JOURNAL

Contacts

Product Flyer

Order back issues

Article Reprints

Bulk Orders

### ALERTS FOR THIS JOURNAL

Get the table of contents of every new issue published in

# Background

- 1982—86: BA French, German, Linguistics

- 1987—88: MSc Computer Science (Cog. Sci)

- 1988—1991: Eurotra MT Project
  - German-to-English Transfer rules
  - English Synthesis rules
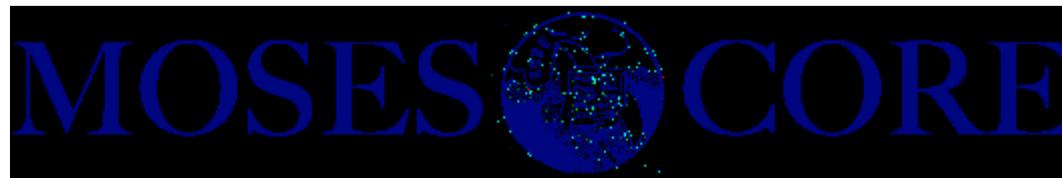
# Middle ground

- Started up BSc in Applied Computational Linguistics (1991—2006)

- PhD in MT (Distance Learning) 1996—2001

- 2006—11: Teaching on BSc in Computing

- Built up NCLT, then MT group
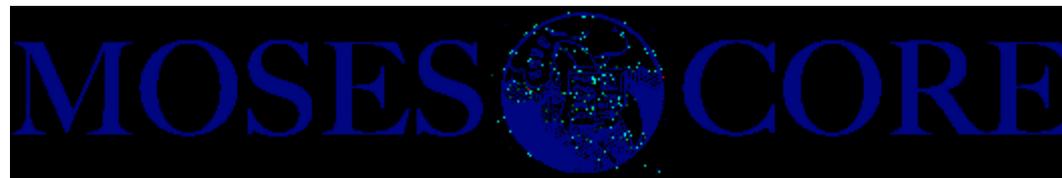
# Machine Translation @ the NCLT & the CNGL

## Dublin City University, Ireland

Welcome to the website of the Machine Translation (MT) group in the **School of Computing** at **DCU**. The research that we carry out is undertaken in the **National Centre for Language Technology (NCLT)**, and the **Centre for Next Generation Localisation (CNGL)**. We carry out a wide range of research in MT in the NCLT and CNGL, including:

- Integrating Syntax & Semantics into Phrase-Based Statistical Machine Translation
- Hybrid Statistical & Example-Based Machine Translation
- Tree-Based Machine Translation
- Sub-sentential Alignment for Machine Translation
- Spoken Language Translation
- Probabilistic Transfer-Based Machine Translation
- Evaluation in Machine Translation
- Controlled Language & Machine Translation
- Human Factors in Machine Translation
- Quality Estimation in Machine Translation

I'm delighted to announce that we have been successful in a number of FP7 projects, including:

- **EuroMatrix+**: Bringing Machine Translation for European Languages to the User
- **Panacea**: Platform for Automatic, Normalized Annotation and Cost- Effective Acquisition of Language Resources for Human Language Technologies
- CoSyne: Multi-Lingual Content Synchronisation with WIKIS
- PLuTO: Patent Language Translations Online
- **T4ME**: Technologies for the Multilingual European Information Society

See our **Projects** section for details on each individual project. These projects will bring in another 11 staff to our already large, well-established team. Some of our recent recruits include:

- Dr. Raphael Rubino
- Dr. Xiaofeng Wu
- Dr. Junhui Li
- Rasoul Samad Zadeh Kaljahi

See our **People** section for further information. Once all the new staff are in place, our MT team here will constitute 33 people. Other project proposals are currently under review.

We'll continue to get involved in large-scale MT evaluation campaigns (e.g. we've recently participated in IWSLT, NTCIR and MEDAR campaigns, with WMT and MetricsMATR evaluation teams just been put together).

Note also that our determination to make some of our software available in the public domain has resulted in the release of **Marclator**, a standalone EBMT system which we use as a wrapper around Moses in our successful MT-evaluation campaigns. We hope you try it out, and get back to us with comments! This is intended to be the first release of MT software from our group, with the ultimate intention to release the full MaTrEx system.

We have formed links with some of the leading groups in the fields of MT , where we carry out joint research, host research visits and submit joint research proposals. We would be delighted to become actively involved with more research groups, so contact us if interested.

Members of the team play a prominent role in the **European Association for MT** and the **International Association for MT**. We urge you to **join** your local association! Furthermore, members of the group are leading representatives on the Editorial Board of the journal **Machine Translation**. Again, we would encourage you to submit papers to the journal, or review books that are of interest to the readership of the journal.

Thanks for accessing our site, and let us know if it did or did not meet your expectations -- if there's something wrong, or missing, we can only fix it if you let us know!

-- *Andy Way, February 2010*

---

### Recent News

**NEW!** July 3rd, 2012: **Sandipan Dandapat** (supervised by **Prof. Andy Way** and **Dr. Sara Morrissey**) succesfully defended his Phd Thesis titled *Mitigating the Problems of SMT using EBMT*.

**NEW!** June 25th, 2012: CNGL researcher **Dr. Tsuyoshi Okita** of DCU is to co-organise the AMTA 2012 Workshop on Monolingual Machine Translation (**MONOMT 2012**), which takes place on 1st November in San Diego, USA.

**Older News**

# PhD Students

- 2004: Aoife Cahill: *Parsing with Automatically Acquired, Wide-Coverage, Robust, Probabilistic LFG Approximations* (jointly, with Josef Van Genabith)

- 2005: Nano Gough: *Example-Based Machine Translation Using the Marker Hypothesis*

- 2005: Mary Hearne: *Data-Oriented Models of Parsing and Translation*

- 2005: Ruth O'Donovan: *Automatic Extraction of Large-Scale Multilingual Lexical Resources* (jointly, with Josef Van Genabith)

- 2006: Mick Burke: *Automatic Treebank Annotation for the Aquisition of LFG Resources* (jointly, with Josef Van Genabith)

- 2007: Declan Groves: *Hybrid Data-Driven Models of Machine Translation*

- 2007: Bart Mellebeek: *TransBooster: Black Box Optimisation of Machine Translation Systems* (jointly, with Josef Van Genabith)

- 2008: Karolina Owczarzak: *A Novel Dependency-Based Evaluation Metric for Machine Translation* (jointly, with Josef Van Genabith)

- 2008: Sara Morrissey: *Data-Driven Machine Translation for Sign Languages*

- 2009:  Hany Hassan: *Lexical Syntax for Statistical Machine Translation* (jointly, with Khalil Sima'an)

- 2009:  Ventsislav Zhechev: *Automatic Generation of Parallel Treebanks: An Efficient Unsupervised System*

- 2009:  Yanjun Ma: *Constrained Models of Word Alignment for Statistical Machine Translation*

- 2010:  John Tinsley: *Resourcing Machine Translation with Parallel Treebanks*

- 2011:  Yifan He: *The Integration of Machine Translation and Translation Memory* (jointly, with Josef Van Genabith)

- 2011:  Rejwanul Haque: *Integrating Source-Language Context into Log-Linear Models of Statistical Machine Translation*

- 2011:  Sergio Penkale: *Incorporating Translation Quality-Oriented Features into Log-Linear Models of Machine Translation*

- 2012:  Tsuyoshi Okita: *Word Alignment and Smoothing Method in Statistical Machine Translation: Noise, Prior Knowledge and Overfitting*

- 2012:  Sandipan Dandapat: *Mitigating the Problems of SMT using EBMT*
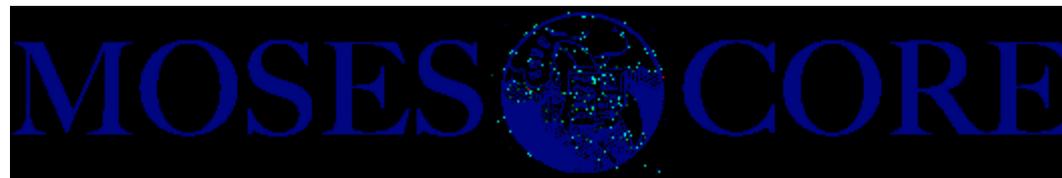
# MSc Students

# Foreground

- Traslán Teoranta: 2004—2010 (bought by ALS)

- 2010—2011: Consultant for ALS LT

- 2011—2014: 3 years' leave-of-absence from DCU, Director of LT @ ALS
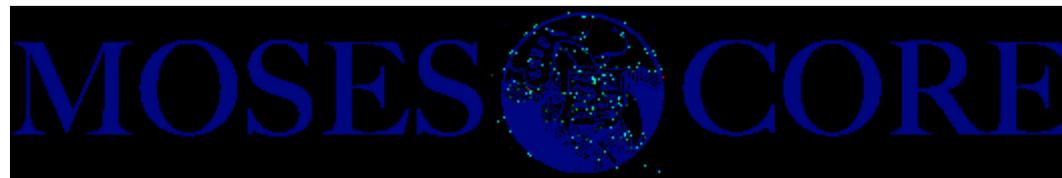
# MT at DCU/CNGL

- Data Cleaning (Okita et al, 2010; Pecina et al., 2011)
- Handling Markup (Du et al., 2010)
- Word Alignment (Ma et al., 2009; Ma & Way, 2010)
- Phrase Extraction (Haque et al., 2010)
- Tuning via MERT (He & Way, 2009a)
- Language Modelling (Okita & Way, 2010)
- Decoding (Srivastava et al., 2011)
- MT Evaluation (Galron et al., 2009; He & Way, 2009b; Naskar et al., 2011)
- TM and MT Combination (He et al. 2010a/b, 2011; Ma et al. 2010)
- Phrase-/Tree-Based SMT (Almaghout et al., 2010, 2011; Haque et al., 2011; Jiang et al., 2010, 2011; Penkale et al., 2010)

# MT at DCU/CNGL

- ✔ excellent quality students

- ✔ good group dynamic, despite the very large size
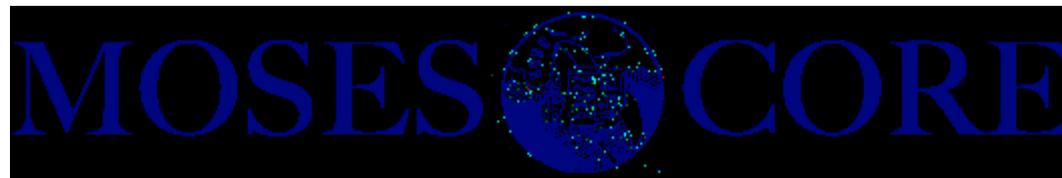
# MT at DCU/CNGL

- Limited view of the client's needs (I'd say '*no* view' if we hadn't been in the CNGL!)
- Limited urgency (but *much* better than in other universities, cf. PhD completion time)
- Code-branching given different funding agencies (SFI, EI, EU …)
- Huge group size: (i) MT group in DCU = 20 people; (ii) ILT group = 40; (iii) five FP7 project teams
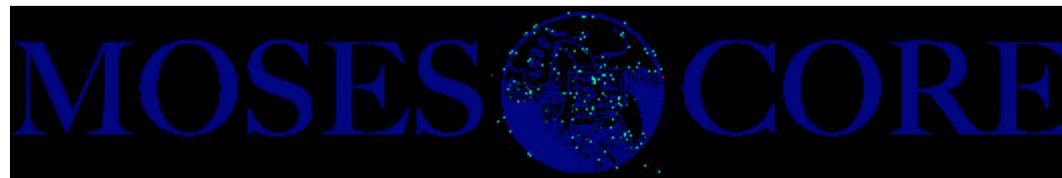
# MT at DCU/CNGL

- € vs £, high taxation

- Research management vs state-of-the-art MT research

- Researchers not supported sufficiently

- Uncertainty of funding situation (in IRL & EU)

- Many good students and postdocs left

- Professors leaving in droves
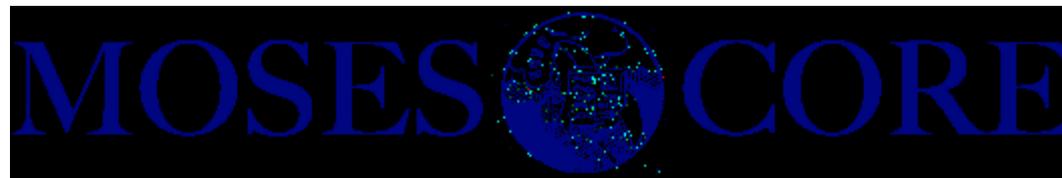
# MT at ALS

- ✓ Exciting opportunity (after 25 years in academia)
- ✓ Move from IRL to UK not 'too' disruptive!
- ✓ Really *wanted* by ALS!
- ✓ Major impact very quickly, SmartMATE, customized engine builds, haven't lost a trial, BLEU scores very high (cf. academia!)
- ✓ *Real* problems, our team very customer-facing …
- ✓ Very low admin overhead, meetings run on-time!
- ✓ 3-year leave-of-absence *huge* safety net!
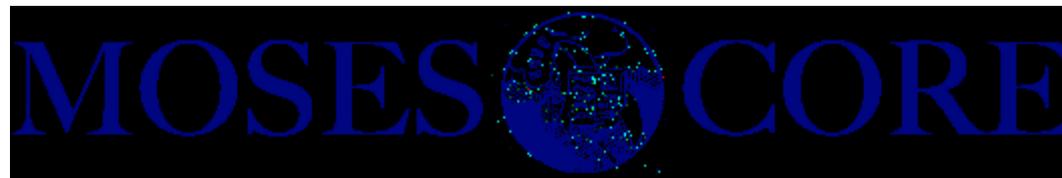
# MT at ALS

- 2x User papers @ AMTA-12 (DUDU, DCU SMT Training)
- 2x Workshop papers @ AMTA-12 (eBay, Spil Games)
- FP7 project papers @ EAMT-12
- 2x Research papers @ EAMT-12
- Research paper @ EACL-12 Workshop
- 3x papers in *Machine Translation*
- 5x Research Papers @ MT Summit-11
- Presentations @ TAUS, GALA, LRC …

- R&D platforms like SmartMATE
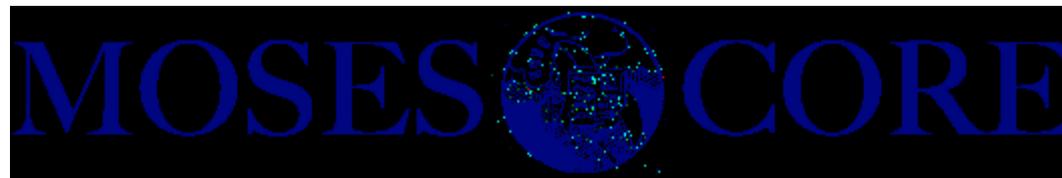
# SmartMATE

*Self-serve translation platform*

Your one stop, self-serve translation technology

**Register here** for your free 30 day trial

*introducing* SmartMATE

## Welcome to SmartMATE, a language translation engine

### Welcome to SmartMATE

SmartMATE is a revolutionary, one-stop, self-serve machine assisted translation platform.

It's the only system that offers integration of translation memories, glossary management, self-serve personalized machine translation engine build and access to a concurrent user editing suite...all in one easy to use translation platform.

The various elements can be used in isolation or as a combined full self-serve translation tool depending on your needs as a translator, language company, SME or complex global enterprise. We've also designed a pricing model that means SmartMATE is **accessible** to everyone. **It's also fully hosted, affordable and secure.**

Sounds too good to be true? Well why not try it for yourself and then give us your honest feedback in a quick **survey**. We look forward to reading your opinions. Click here to complete the survey now.
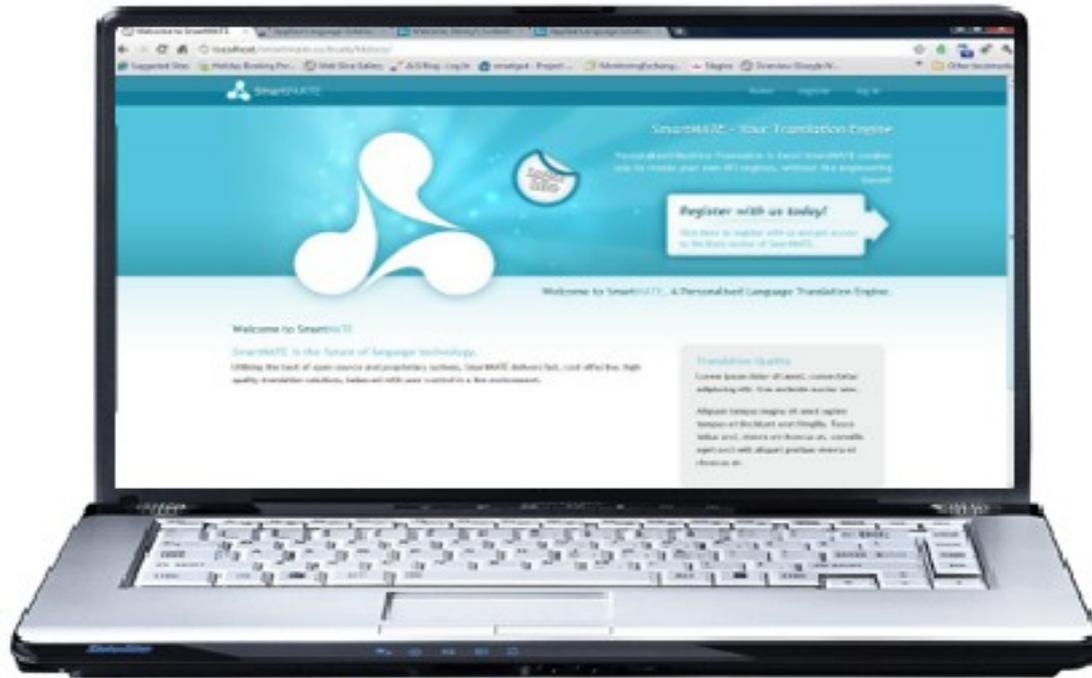
### In the meantime...

### Build your MT Engine - Quick Start

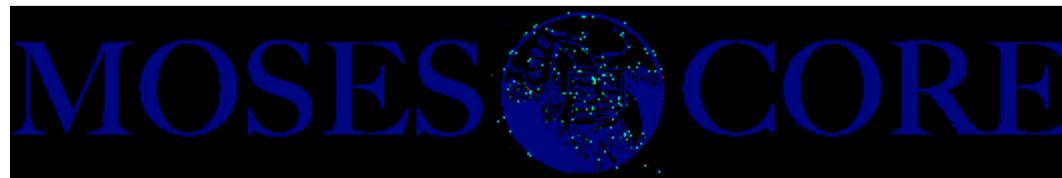### A guide to the SmartMATE editor suite

# SmartMATE 30-Day Free Trial

www.smartmate.co

# MT at ALS

➢ Different company following Capita buyout

➢ Bad PR from Interpreting contract (& boycotting talks)

➢ Salary discussions (cf. publicly available salary scales in IRL)

➢ Check licences! (research vs. commercialisation)

# Final Thoughts: MT at DCU

- ✔ New Group Leader: Prof Qun Liu (CAS, Beijing)

- ✔ CNGL II Funding very likely

- ✔ Research Fellow opportunities

- ✔ Crucial Post-doctoral researchers upgraded

- ✔ Good opportunity for team members to work with people other than me!

- ✔ Hopeful that the group will keep going!
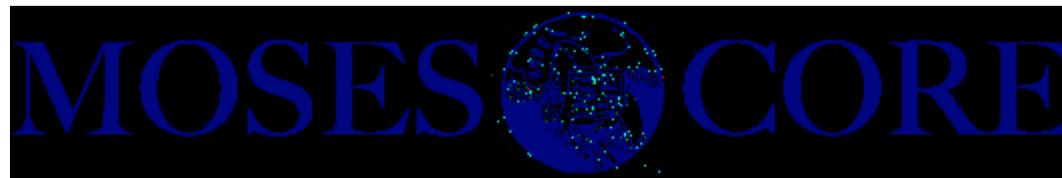
# Final Thoughts: MT at ALS

- ✔ New management in place

- ✔ No financial worries

- ✔ Rebranding

- ✔ Clear strategy and support for MT

- ✔ New recruits soon
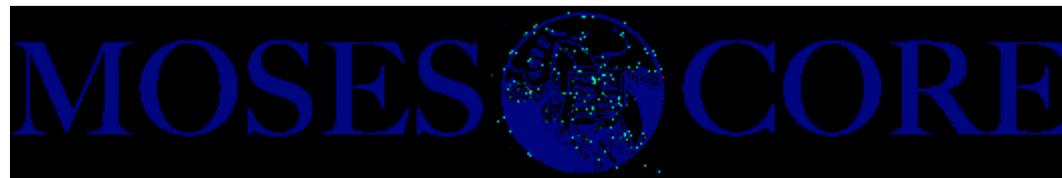
## Senior Director Machine Translation eBay

Job Location: San Jose, CA

eBay is a global leader in enabling commerce across the world and the leading mobile commerce platform. We want to step that up and make it surprisingly easy for people to buy across borders, even if they don't speak a common language. Towards that purpose we are building a world class Machine Translation team at eBay. We are looking for a senior leader with a track record of establishing and leading research and development teams in the machine translation area.

- Do you have a strong scientific background in the statistical and linguistic methods of machine translation?

- Have you tested models, grammars, and launched applications of machine translation and language science in the real world?

- Are you ready to build the ultimate translation engine for major world languages to enable commerce across the globe?

- Are you a senior leader who can build a team to determine the future of international commerce?

We can offer you a chance of making a difference in the lives of millions of people buying internationally and help make research ideas a reality. eBay is committed to this area as a strategic focus and are looking for someone to turn this vision into reality. If you are excited about this opportunity we would love to hear from you.

Here is a first list of your job responsibilities:

- Recruit and build a world class team of experts in machine translation

- Define a research and development roadmap that aggressively tests machine translation capabilities in the context of cross border commerce

- Establish relationships with other senior leaders in eBay Marketplaces to ensure alignment between day to day work and long term vision as it relates to machine translation

- Educate and set expectations within eBay to have clear understanding of what value machine translation can deliver for our users today and how you and your team will improve on those capabilities

- Ensure that the machine translation capabilities are integrated into the user experience in a seamless and intuitive way

- Build an engineering team that designs, integrates, and implements all required parts for a complete machine translation service solution

### Job Requirements

Here is a list of the experience and background you should have to make you successful in this role in eBay:

- Proven ability to lead a team of machine translation researchers, computational linguists, and software engineers
- Deep understanding of latest scientific methods and developments in machine translation
- Proven track record of taking concepts from the lab into a real life, commercial environment
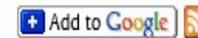
**If you get the chance to switch from academia to industry (even for a short time), do it!**
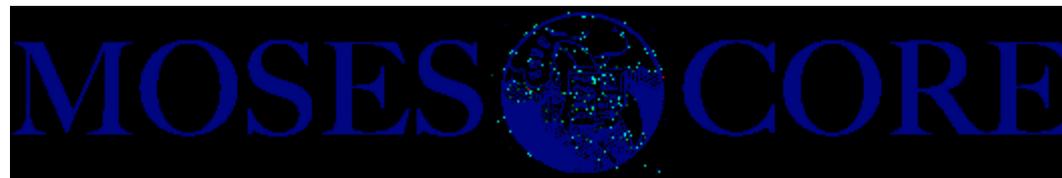
**Thanks for listening!**

**andy.way@appliedlanguage.com**