

---

# Moses Statistical Machine Translation System

Philipp Koehn, University of Edinburgh

pkoehn@inf.ed.ac.uk

<http://www.statmt.org/moses/>

Talk given by Steve Renals



# What is Moses?



- Statistical machine translation system:  
Provide translated text → build MT system
- Developed with European funding
- Open source (liberal LGPL license)

**Main developer:** University of Edinburgh  
Significant contributions from  
FBK-IRST (Italy), Charles University (Czech Republic),  
and many other groups in Europe and world-wide

# Widely Used

- **Web site** gets **7000 visits** and **20,000 page views** per month
- **Mailing list** distributes **100 emails** per month to **600 members**
- **Academic** users
  - de-facto benchmark for new MT methods
  - Moses paper most cited paper in 2010 ACL conferences
  - starting point for most new research groups
  - about half of evaluation campaign (IWSLT, WMT) submissions use Moses
- **Commercial** users
  - Moses "re-sellers" build customized systems
  - language services companies use Moses in-house



# Web Site

<http://www.statmt.org/moses/>

- **Download**

- compiled binaries for Unix and Windows
- current source code from SVN repository

- **Documentation**

- introduction to statistical MT methods
- step-by-step tutorial on training, decoding, factored models, syntax models
- step-by-step instructions on how to build a baseline system
- descriptions of all features
- code documentation
- mailing lists for users and developers

# Online Demo

<http://demo.statmt.org/>

## Moses Machine Translation Demo



### Source:

Au Congrès américain, Nicolas Sarkozy reçoit une "standing ovation"  
Le président français a souligné que la France est "l'amie des Etats-Unis", tournant ainsi la  
page de la brouille entre les deux pays, avant d'évoquer les grands dossiers internationaux.

French-English (Europarl) ▼

Show Debug Output  Show Alignment

Translate

### Translation:

The American Congress, Mr Sarkozy receives a standing ovation 'station' The French President has pointed out that France is the 'friend of the United States', thereby turning the page of the dispute between the two countries, before mentioning the major international issues.

# Who will do MT Research?

- If MT research requires the development of **many resources**
  - who will be able to do relevant research?
  - who will be able to deploy the technology?
- A **few** big labs?

The Google logo, consisting of the word "Google" in its characteristic multi-colored font.The IBM logo, consisting of the letters "IBM" in a blue, striped font.The Microsoft logo, featuring the four-pane Windows logo above the word "Microsoft" in a bold, black font.

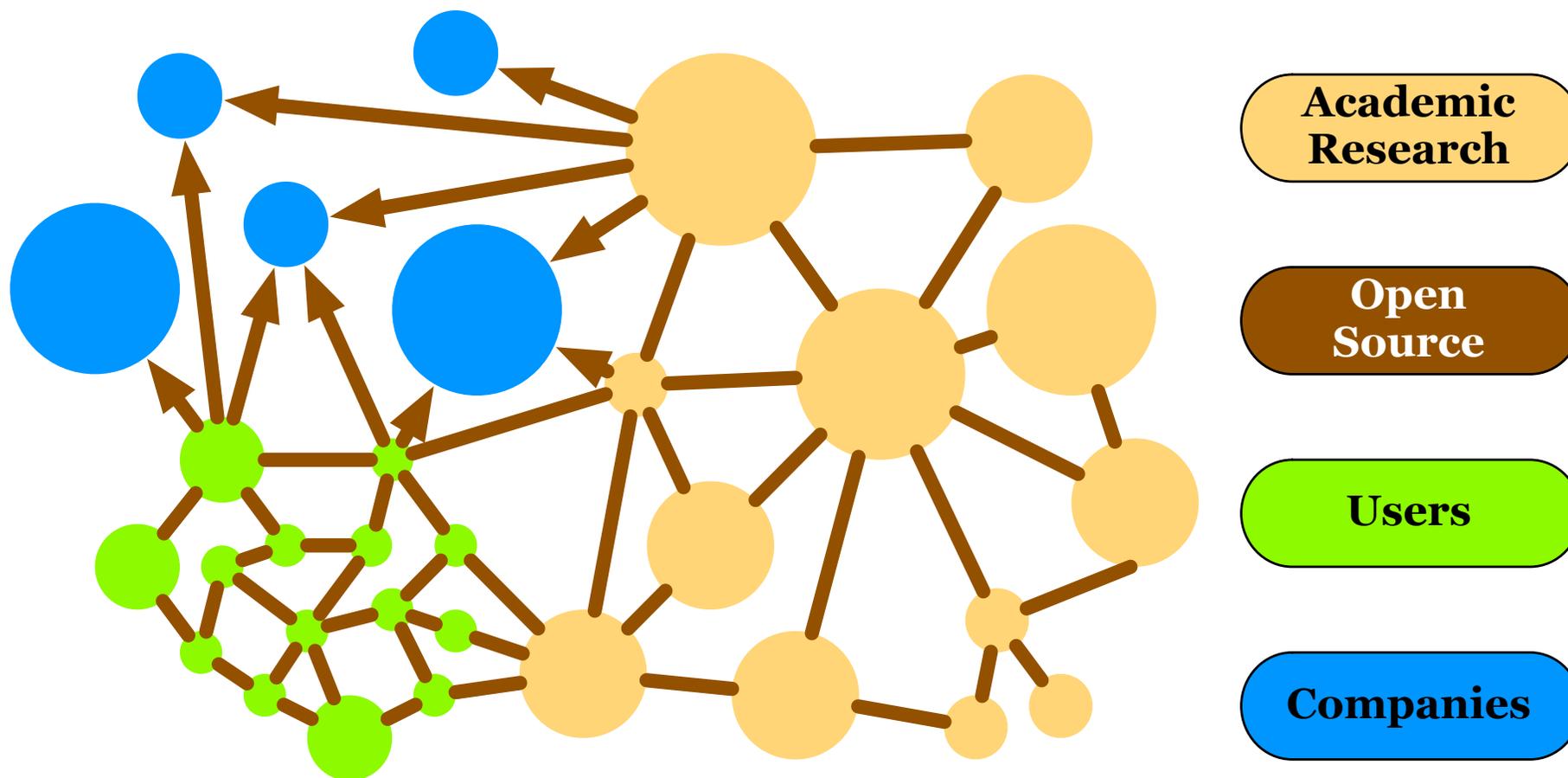
- ... or a **broad network** of academic and commercial institutions?



# MT is diverse

- Many different **stakeholders**
  - academic researchers
  - commercial developers
  - multi-lingual or trans-lingual content providers
  - end users of online translation services
  - human translation service providers
- Many different **language pairs**
  - few languages with rich resources: English, Spanish, German, Chinese, ...
  - many second tier languages: Czech, Danish, Greek, ...
  - many under-resourced languages: Gaelic, Basque, ...

# Open Source Connects Stakeholders





# Involving Research Community

- Moses development is centered in the research community
  - research projects directly extend system
  - contributions from academics across the world
- Sixth MT Marathon: **5-10 September 2011, Trento, Italy**  
<http://mtmarathon2011.fbk.eu/>
  - one-week **intense class** with hands-on experience
  - **research showcase** with talks from leading researchers
  - **open source** convention
  - work on **open source projects**

# Involving User Community

- TAUS organized many **outreach events** to assess user needs
- **"How-To" tutorials** at industry events
- **Ongoing projects** to address needs by industry
  - LetsMT! project (service to build customized systems)
  - Moses for Localization project
  - Do Moses Yourself project
  - Moses for Mere Mortals project
- **MosesCore: FP7 support action currently under negotiation**

# Future Directions

- **Technology**
  - statistical models that incorporate **syntax and semantics** ("hybrid")
  - integration with **speech** technologies
  - online **adaptable** to domains, tasks, users, time, etc.
  - **EU-BRIDGE: FP7 IP currently under negotiation**
- Integration in **Computer Aided Translation** tools
  - integration of **translation memory** and statistical machine translation
  - novel types of assistance for users
  - learning from user interaction
  - **CASMACAT and MateCat: FP7 STREPs currently under negotiation**

**Thank you**

Questions?



pkoehn@inf.ed.ac.uk  
<http://www.statmt.org/moses/>