



Directorate-General for Translation

Is machine translation ripe for EU translators?

Josep Bonet
Head of the IT Unit
Paris, 02-12-2010

EUROPEAN
COMMISSION



Who talked about hype?

- Maybe the wrong question
- Are **translation consumers** ripe for MT?
- We promised the moon ...
- Now they've seen Google and they believe
- We have to offer them a nice planet to live in



And the translators?

- Change in attitude
- What was despised yesteryear is asked for now
- A noisy minority ... will lead the others
- They are already putting high pressure on us!



The approach to MT

- The importance of directions
 - MT, what for?
 - When a means becomes an aim
 - In any case, it's circular
-
- Now, has everybody realised that the profession has to change?



How good is good?

- We are on a high end
 - Niche market based on quality
 - Full human quality is sought for
 - And nevertheless...
-
- Do we serve the multilingual needs of the EU?



Translation @ EC

Directorate-General for Translation

- Staff:
1750 linguists and 600 support
- Production (M pages):
0.9 (1992) 1.2 (2004) 1.8 (2008)

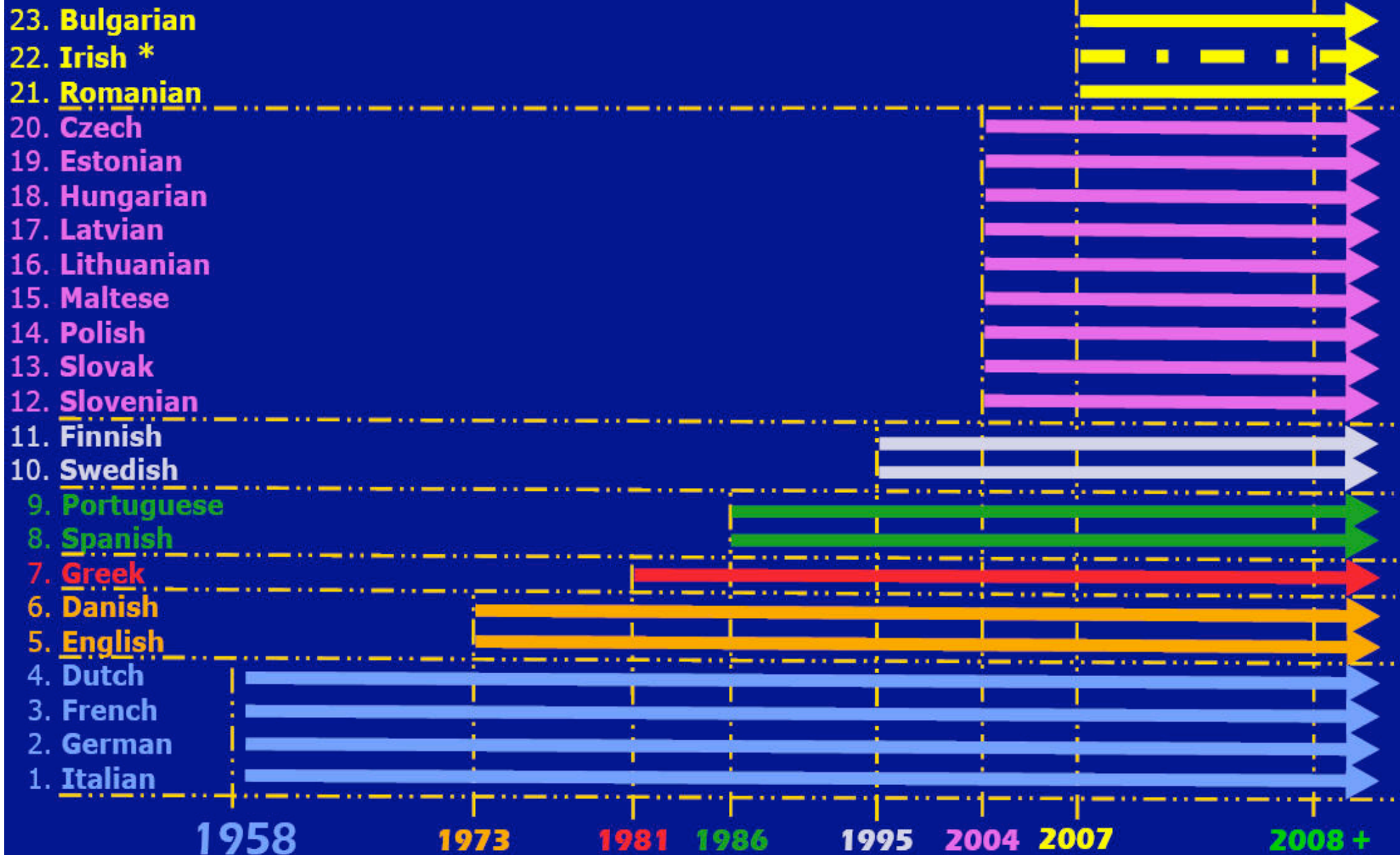
BUT

to make europa.eu fully multilingual

- translate almost 6.8 million documents
- 8,500 translators working full-time for one year
- not feasible if not using new technologies like
MT



Languages from and into which we translate



* **Irish:** 'Treaty Language' from 1973 onwards, official language status since 2007.

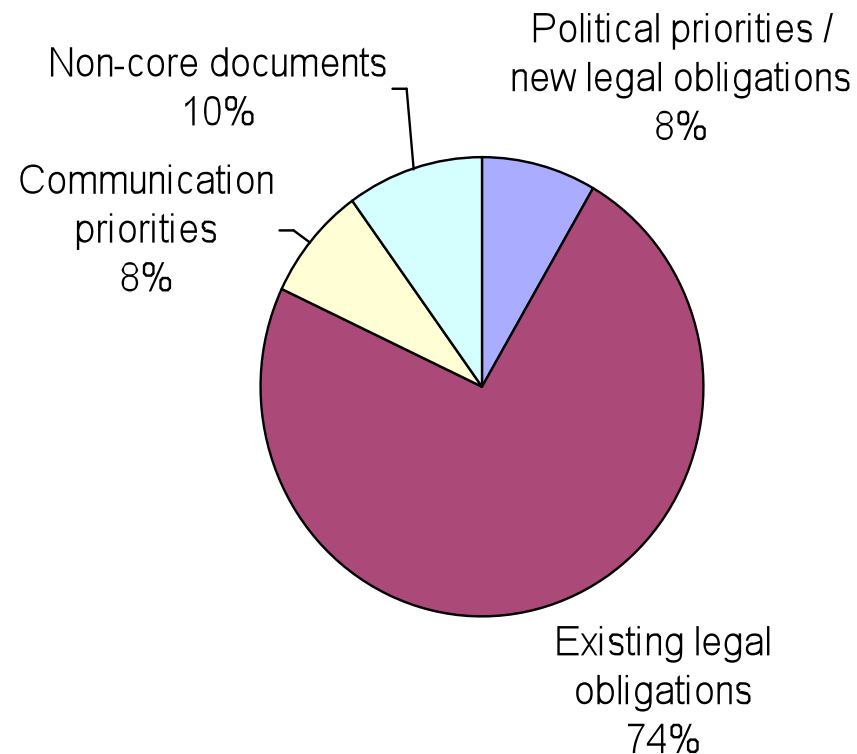
What do we translate?

- Legal acts and preparatory documents
- Commission decisions and communications
- Publications
- Correspondence
- Speeches, minutes
- Reports, working documents
- Web pages



What do we translate?

Document types at the European Commission



Why?

- **Council Regulation No 1/58**
 - Regulations and other documents of general application shall be drafted in the official languages.
- **Treaty establishing the European Community and the Lisbon Treaty**
 - Citizens have a right to address the official EU bodies in any of the EU's official languages and to receive a reply in that language.





European Parliament

760



European Commission

1750



Council of the EU

650



Translation Centre

110

**EU is more than EC
Here's the whole
picture!**



Committee of the Regions



Court of Auditors

100

Court of Justice

620



**Committee of the Regions
and
European Economic and
Social Committee**

350

350

We use tools

- Translation memories
 - Terminology tools
 - Documentary databases
 - Virtual libraries
 - Electronic dictionaries
 - and ECMT
-
- But significant grey zones remain



The present: ECMT service

- rule-based machine translation
- developed since 1975
- 28 language pairs available (ten languages)
- since 2006 no significant work on any pair
- use (requests x 10⁶): 1.5 (2006); over 2.5 (2009)
- used by
 - **EU institutions for gisting**
 - **Online services and information systems for raw translation**
 - **DGT as a CAT tool**



The future: MT@EC service

Policy

Commission Communication on "Multilingualism" 2008:
"human and automatic translation is an important part of multilingualism policy"

Facts

- ECMT is costly to develop
- Data-driven systems are cheap and quick to develop... if you have the data

Language Technology Watch

- Market and research observation
- Tests of commercial and non commercial tools and MT systems



MT@EC Needs – resources - action

MT@EC strategy

- Adopted in June 2009 by DGT
- Task Force created *November 2009*

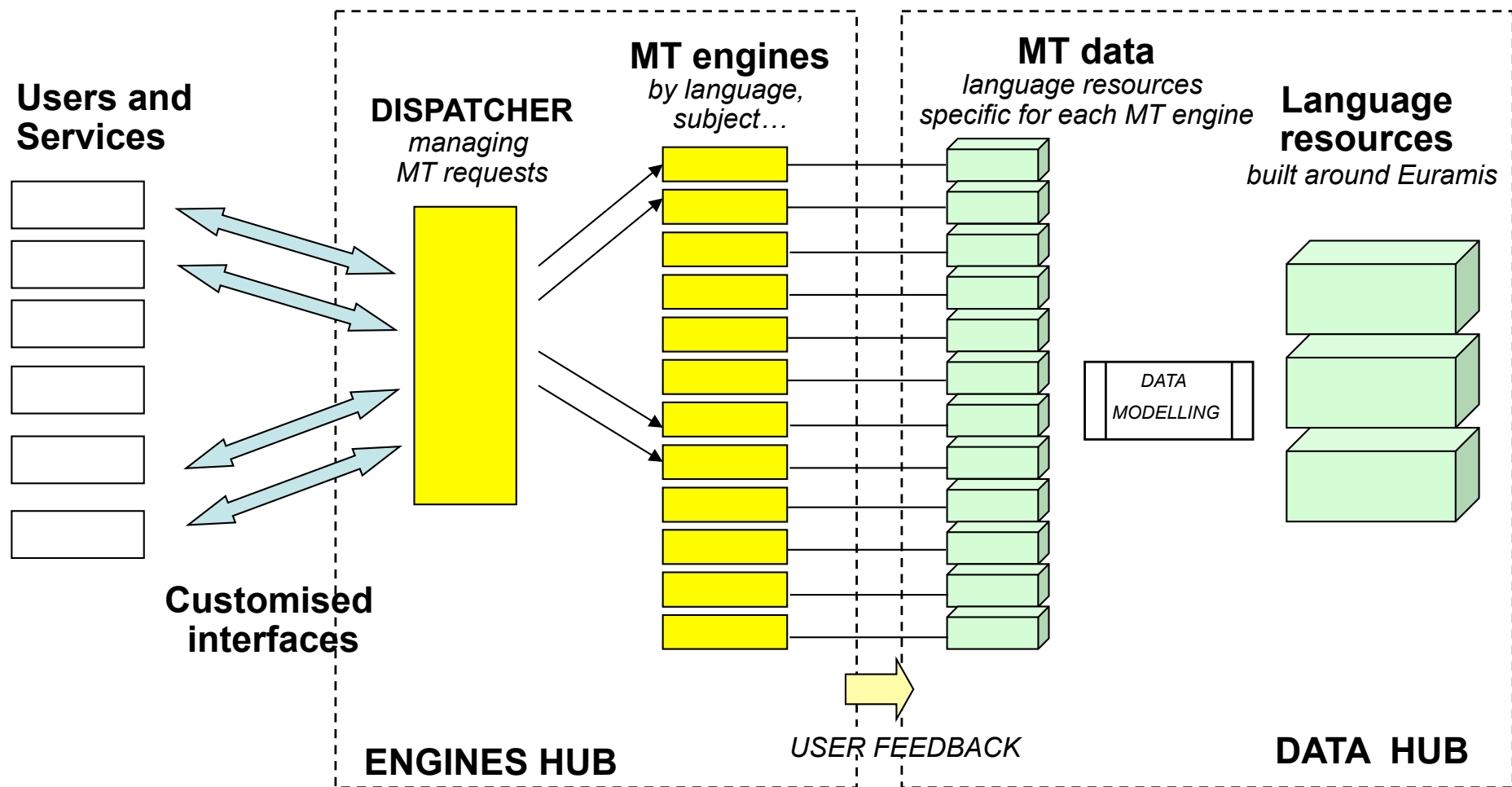
Task Force results *April 2010*

- MT@EC is necessary for the Commission
(trust, confidentiality, continuity)
- Data-driven systems: a major technological breakthrough
- User requirements have been collected
- An outline of an “architecture” has been elaborated
(flexible, sustainable, ensuring technological independence)
- Recommendations on organisational and financial arrangements



Machine Translation Service

Outline of the proposed MT@EC architecture



Machine Translation Service

A number of projects within an “MT@EC programme”

“MT Engines - baseline” project **(EC)**

IT infrastructure for the core of the “MT Engines Hub”

“MT data management hub” projects **(DGT)**

Language resources (LR) underlying the MT system

“Customised MT solutions” projects **(clients)**

“Client” requesting development of (examples) :

- a domain specific MT engine
- a specific interface to external services



Exodus

- Internal DGT experimentation with Moses toolkit
- Using Euramis (internal) TM data
- With temporary redeployment of existing ICT and human resources
- With the active contribution of :
 - the DGT's Portuguese department
 - the EuromatrixPlus project
 - the Translation DG of the European Parliament



What was done

- Corpus preparation and cleaning
- Development of an EN->PT engine
- Human evaluation by the PT LD (more than 30 translators involved)

What has not been done

(due to time and resource limitations)

- No iterative process for improving corpus quality.
- No incremental updates of translation and language models
- No engineering interventions



Exodus

First conclusions

- Quality evaluation of MT output for EN->PT results very encouraging
- Dedicated analysis on IT engineering work required for production ready system for all EU languages
- Quality of data cleaning and preparation: the main "comparative" advantage of DGT

Note:

More Exodus pairs are currently being evaluated by the European Parliament, who also submitted an Exodus pair (EN-to-FR) to the WMT 2010 competition



Next: putting pieces together

MT Action plan *June 2010*

- Action line 1: MT data
started with: internal translation memories
challenge: prepared for optimising all kinds of data for MT



Next: putting pieces together (2)

MT Action plan *June 2010*

- Action line 2: MT engines
started with: open source tools
challenge: compare alternative systems (both commercial and non-commercial) in terms of *quality of output, price (total cost of ownership), feasibility, language coverage*



Next: putting pieces together (3)

MT Action plan *June 2010*

- Action line 3: MT service
started with: prototype of architecture according to TF
challenge: flexible and sustainable implementation and governance of MT service

In parallel EC is preparing to continuously update the DGT Multilingual Translation Memory of the *Acquis Communautaire* (DGT-TM)



And the question is ...

- Are the EU translators ripe for MT?



