

---

# Machine translation in the European Commission

**Josep Bonet**  
**LangTech 2008**  
**Rome, 28/02/2008**

# Stating the obvious

---

**DGT means DG for Translation**

**Machine translation is about Translation**

**Therefore DGT is concerned with machine translation**

**Eh... Maybe, but not sure**

# The plan

---

**The infancy (how it all started)**

The first maturity (translators get focus)

The middle age crisis (I don't get no satisfaction)

The second maturity (will this time be for real?)

# How it all started

---

**When? In the mid-70s**

**By whom? DG XIII, Telecommunications, Information Market,  
and Exploitation of Research**

**Why? To overcome language barriers**

**Why? To support the development of enabling technologies**

**What? Rule-based system for EN=>FR and FR=>EN**

## How it all started (cont'd)

---

**Initial systems improved**

**Dictionaries enlarged and completed**

**Rules added and improved**

**New systems: German, Spanish, Italian, etc.**

**Joint venture DG XIII – SdT**

# The second way

---

## Eurotra

**Big research project** 👍

**Many teams in most of Member States** 👍

**Innovative approach** 👍

**Too heavily dependent on computing power** 👎

**No working MT system delivered** 👎

**Big boost to language industries in Europe** 👍

# The plan

---

The infancy (how it all started)

**The first maturity (translators get focus)**

The middle age crisis (I don't get no satisfaction)

The second maturity (will this time be for real?)

# Translators get focus

---

**(< 1990) Translator equals monk or if I had a dictaphone**

**Terminology and documentation go digital**

**Typing kind of goes digital**

**Translators scribble or dictate**

**Translation is an art and the translator the artist**



## Translators get focus (cont'd)

**(> 1990) PCs become pervasive**

**Translation is (almost) fully electronic**

**Tools for translators get the focus**

**Art versus consistency. A new view on quality**

**Time is gold. And gold matters**

## Translators get focus (cont'd)

**New orientation in the '90s**

**EC to finance research projects, not to develop**

**Help the industry become strong**

**Stress on pre-competitive research**

**End-user products out of scope**

## Translators get focus (cont'd)

**(Mid-'90s) Development of MT transferred to DGT**

**Development of further language pairs**

**Improvement of dictionaries and lexical routines**

**Total number of pairs available: 26**

**English and French as source or target language**

**Coupled (to and fro) with DE, IT, ES, NL, PT, EL, SV, DA**

# Translators get focus (cont'd)

**Varying quality**

**From rather acceptable**

**To rather unacceptable (prototype)**

**For best output, less important text, easier input and more committed community:**

**MT with light and quick post-processing**

**Keep you eyes on the client**

**Tam-tam and Tchernobyl**

# How do you get MT in the EC?

- 1) E-mail > phased out
- 2) Euramis: integrator of language services
- 3) Web interface
- 4) Automated pre-processing

**Methods 2 and 4 only for translators**

**Methods 1 and 3 also for all EC staff**

# The plan

---

The infancy (how it all started)

The first maturity (translators get focus)

**The middle age crisis (I don't get no satisfaction)**

The second maturity (will this time be for real?)

# I don't get no satisfaction

---

**Creating ex nihilo or re-using?**

**The green era: recycle rules**

**Europe enlarges, so what?**

**23 languages > great impact**

**Redefinition of translation business**

**Expandability of systems**

**The corpus is king**

**TM won the battle against MT**

# I do get satisfaction, by the way

**There is life beyond translation**

**Who are the big users of MT in EC?**

**Are they satisfied?**



# The plan

---

The infancy (how it all started)

The first maturity (translators get focus)

The middle age crisis (I don't get no satisfaction)

**The second maturity (will this time be for real?)**

# Will this time be for real?

---

**Fully Automatic High Quality Translation is out**

**Fully Automated Usable Translation is in**

**Corpus-based developments and global markets**

**Quality is in the eye of the reader**

**The user finally gets the focus**

# Caring about the user

---

**Translator: prefers quality, even if incomplete**

**End-user: requests all the information, even if little readable**

**Don't overlook tolerance thresholds**

**Language is extremely personal**

**It involves deep feelings: sensitive issue**

**MT should be seen as an added service, not a denial of service**

# Let's start wrapping up

---

**Availability of big corpora has boosted developments**

**DGT has released its corpus of aligned EU legislation in force in 22 languages**

**Exponential growth of demand for quick and cheap translation**

**User satisfaction with MT solutions**

**Resignation? Rather imperfect than non existing**

**The case of Microsoft's Knowledge Base**

# Let's start wrapping up

---

**Concentration on major languages**

**Most European languages have few speakers**

**MT remains mostly intelligence-related**

---

Thank you