MT@EC

Final Multilingual Web workshop
Luxembourg 15-16 March 2012

Spyridon Pilos
Language applications
DG Translation of the European Commission
A new machine translation service for the European Commission

- Necessary
- Based on data-driven MT technology
- Making best use of Commission language resources
- Making best use of internal linguistic expertise (1700 translators for 23 languages)
- Open and flexible
- Ensuring technological independence
- Being built by DG Translation
- Started: summer 2010
- Deploy: summer 2013
Multilingual web = Multilingual content

\[ MLW = MLC \]

\[ MLC = C \text{ available in } L_1, L_2, L_3, \ldots, L_n \]

How it is done:
- Author produces \( C \) in one language
- Translators transfer the \( C \) to other languages
- Publisher gets \( C \) in all languages and publishes it
Multilingual web = Multilingual content

(MLW = MLC)

Author
L₁

Translate
L₂, L₃

Translate
L₄

Translate
L₅

Translate
Ln

Publish
L₁, L₂, L₃, ..., Ln

Publish
different
L₁, L₂, L₃, ..., Ln

MLWeb workshop

MT@EC
Multilingual web = Multilingual content

\[ \text{MLW} = \text{MLC} \]
Multilingual web = Multilingual content  
\( MLW = ML_C \)

\( ML_C = C \) available in \( L_1, L_2, L_3, \ldots, L_n \)

How it could be done:

- Publisher prepared for \( C \) in all languages (“placeholders”)
- Author produces \( C \) in one language “ready for publication”
- Translators produce \( C \) in other languages “ready for publication”
Language applications: example MT
Getting multilingual data from the web

- europa.eu has lots of MLC
- We tried with web site 1: not that difficult
  
  BUT
  
  ...when going to web site 2, we need to analyse the way that site manages linguistic versions

  ... web site n has yet another way of managing linguistic versions

- ... and why not, since there is no standard to follow?
Giving multilingual data to the web

**DGT-TM**
- Translation memories from the Official Journal of the EU
- uses tmx
- yearly updates (to come): “do not change, so that there is continuity”

**DGT-Acquis**
- New resource (to come)
- Parallel text from the Official Journal of the EU
- First version: “do not change too much compared to JRC-Acquis”
- establish “our” standard way of working (which we make public)
- yearly updates (to come): “do not change, so that there is continuity”
Conclusion

• Need some order for multilingual content on the web
  • “Standard” data models?
  • “Standard” structure for data “storage” or “publication”?
  • Standard?
• MW consortium (and MW-LT) are meant to:
  • propose a feasible approach
  • demonstrate the benefit for all
• Commission DG for Translation: founding member of “Language interoperability portfolio” (Linport)
Thank you for your attention