

MyMemory

Creating the world's largest Translation Memory

Marco Trombetti
Translated
ASLIB, London
November 2009

<http://mymemory.translated.net>

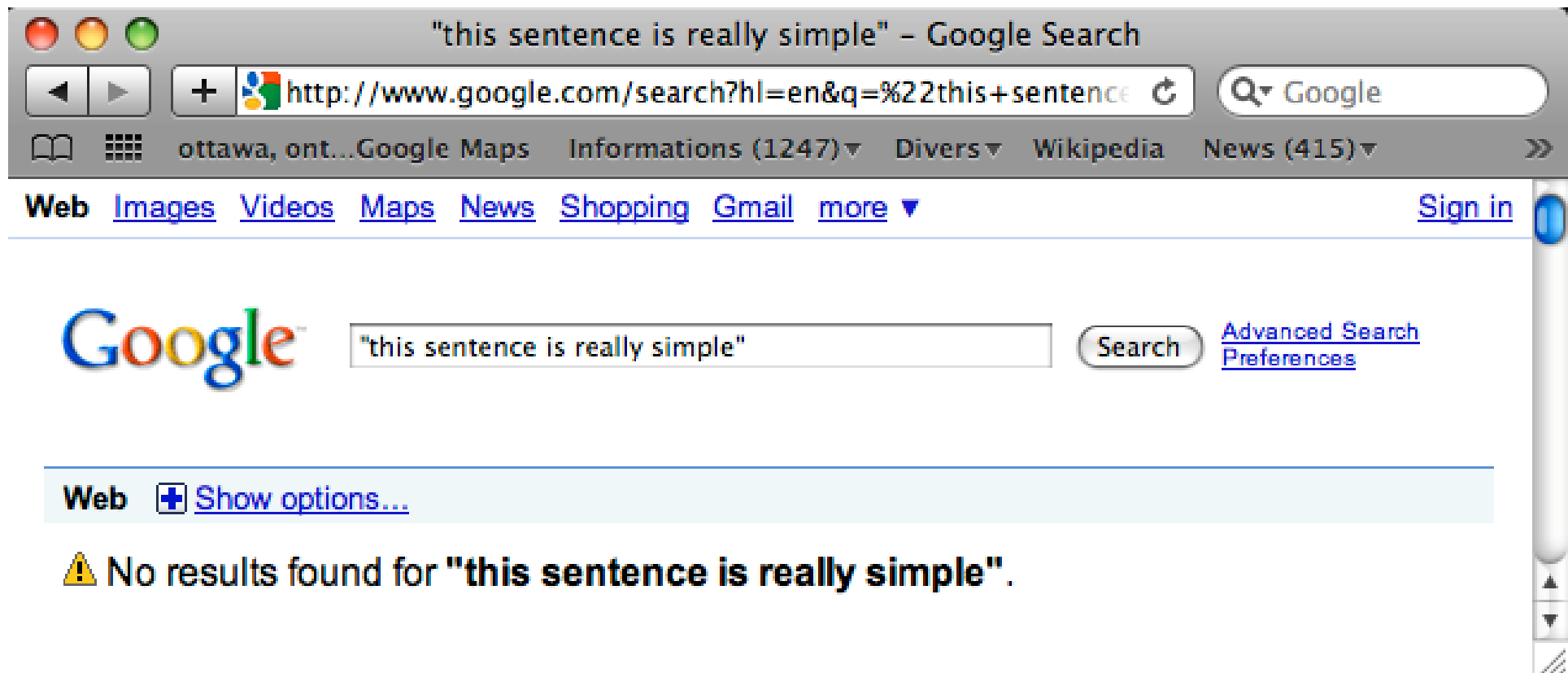
Translated

Language Service Provider

Human Translation in 80 languages, 9000 customers,
100K translations delivered.

Core Technology

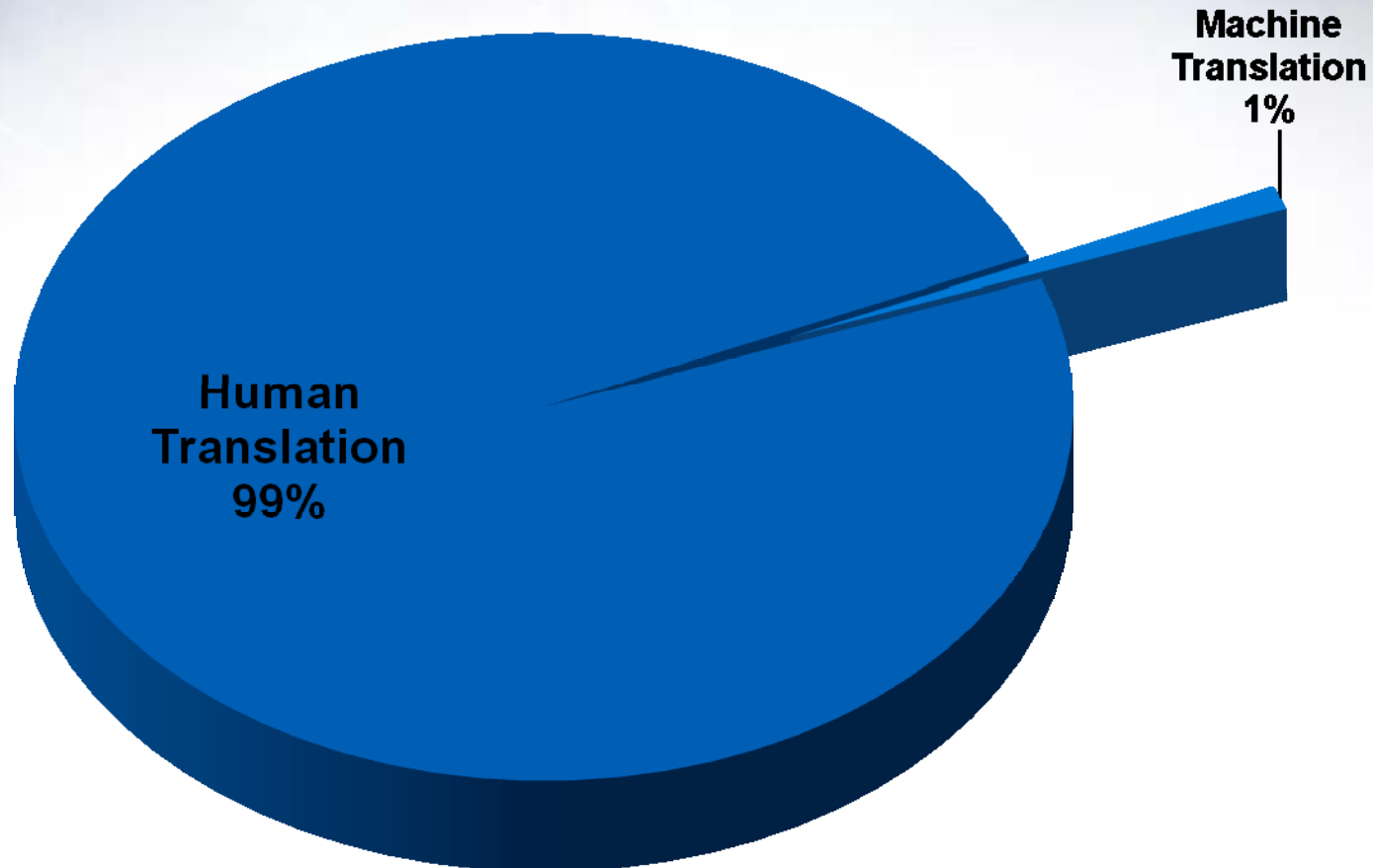
A fully automated translation workflow.
Customers can access 35,000 translators
through a Web Service.



Translation Memories fail!

Prob(10-word-sentence) is too small

Why is MT not focused towards helping HT?



Source: Allied Business Intelligence

My wish-list for the perfect TM Platform

MyMemory

Billions of segments, but still FAST



Web 2.0, Wiki-like contributions



One-click upload/download of TM



Adaptive matching, a priori probability



Propagate changes in real-time



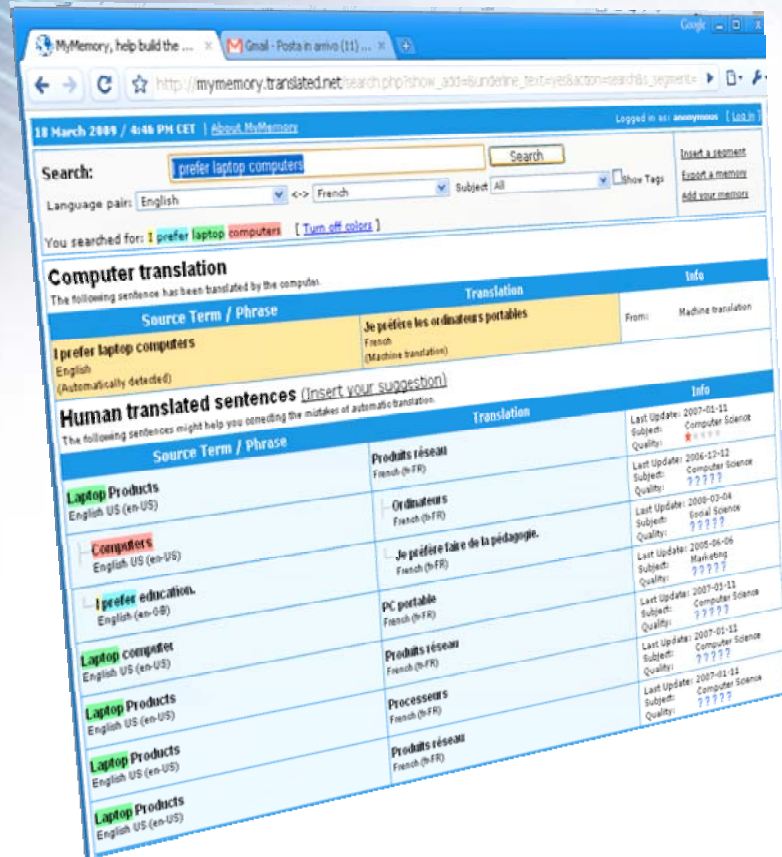
Fully integrated statistical MT



Open standard and CAT support



The Solution



Free,
Collaborative
Web Translation
Memory

Barriers to contribution



Privacy and IP



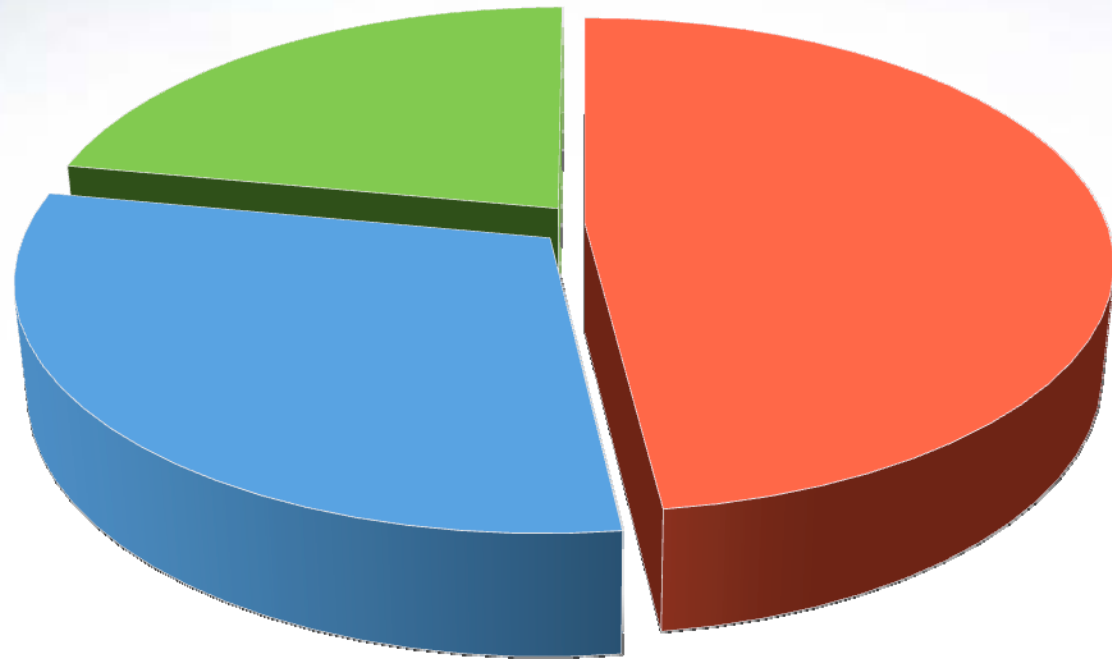
Time vs Value

Will people contribute?

Yes! 100 words per second for 1,5 years

5 Billion words

Sources



■ Web ■ Translators ■ Customers

Benefits

LSP & Clients

- ✔ More control
- ✔ Free contribution from professionals & the crowd
- ✔ Quick reference and QA

Translators

- ✔ Quick and relevant reference
- ✔ Improved organisation and usability
- ✔ Free memories

Technology Challenges

- ✓ **Scalable Real Time Indexing**
- ✓ **Adaptive Ranking**
- ✓ **Cost-Effective Scalability**

Generating the interest



Pre-Populate



Make it easy to contribute



Make it free

Protecting IP

- ✓ **Private memories**
- ✓ **Hide proper nouns**
- ✓ **One-click to report abuse**
- ✓ **Robots.txt**

Quality

- ✓ **Feedback from the crowd**
- ✓ **QA correlation**
- ✓ **Automatic QA**

Is 5 Billion words is enough?

No.

Professional translators produce

**Tens of billions of
words per year**

Next Steps

More data & technology

- ✔ More data extraction from the web
- ✔ Real time SMT training
- ✔ Every CAT tool connected

QA crowd platform

- ✔ Fast and cost-effective MT evaluation
- ✔ Parallel corpus evaluation and clean-up