

MyMemory

Creating the world's largest Translation Memory

Marco Trombetti

Translated

MT Marathon

Trento, Sep 2011

<http://mymemory.translated.net>

Translated

Language Service Provider

Human translation in 80 languages, 15,000 customers, 170,000 translations delivered.

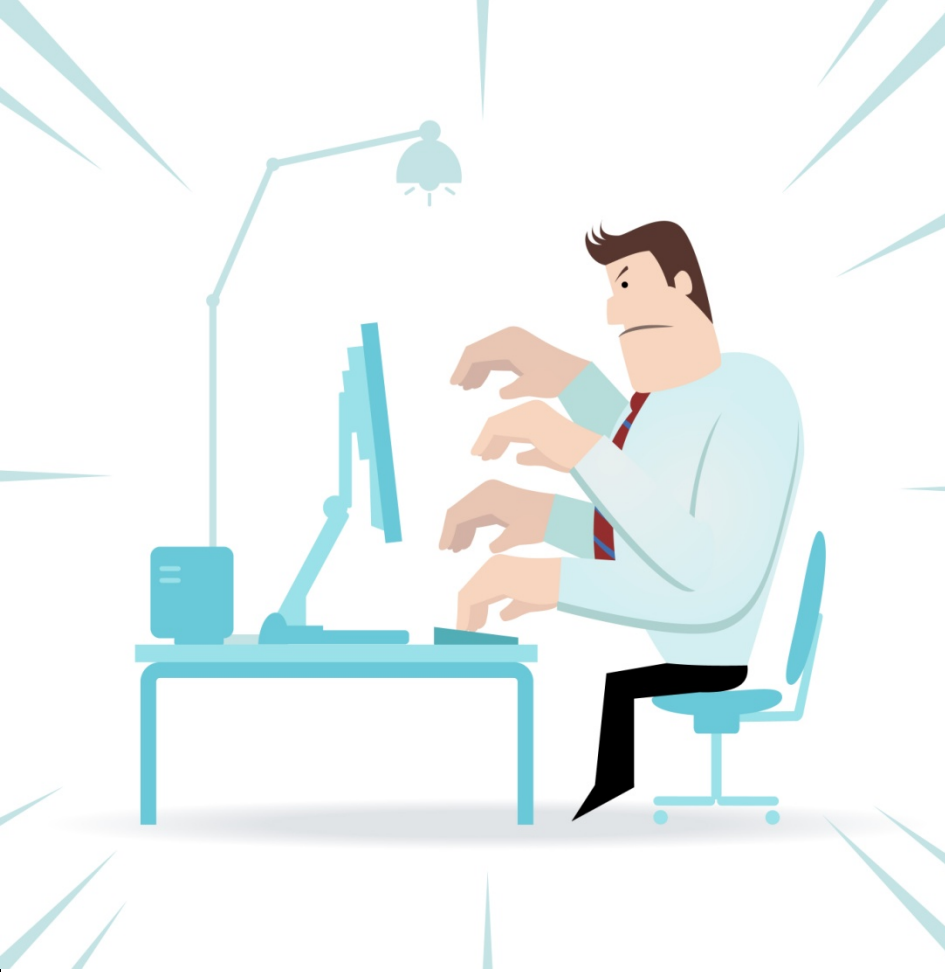
Core Technology

A fully automated translation workflow.
Customers and LSPs can access 45,000 translators through a web form or an API.



2

Challenges



1.

Help Translators

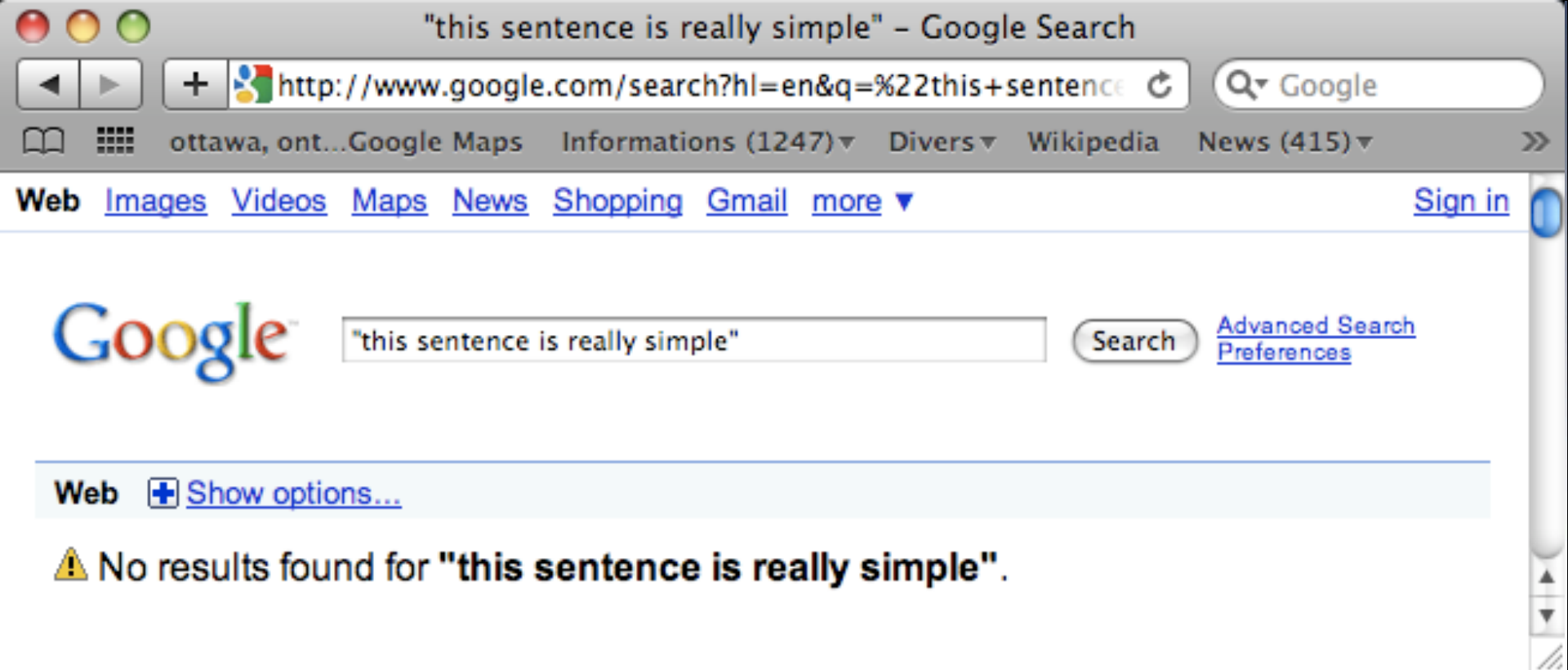


2.

Cut out the middle man

The background features a gradient of blue and white. In the upper left, there are numerous thin, bright blue lines radiating outwards, creating a sense of motion or light trails. A large, soft, white, cloud-like or smoke-like shape is positioned in the center, partially obscuring the blue background. The overall aesthetic is clean, modern, and high-tech.

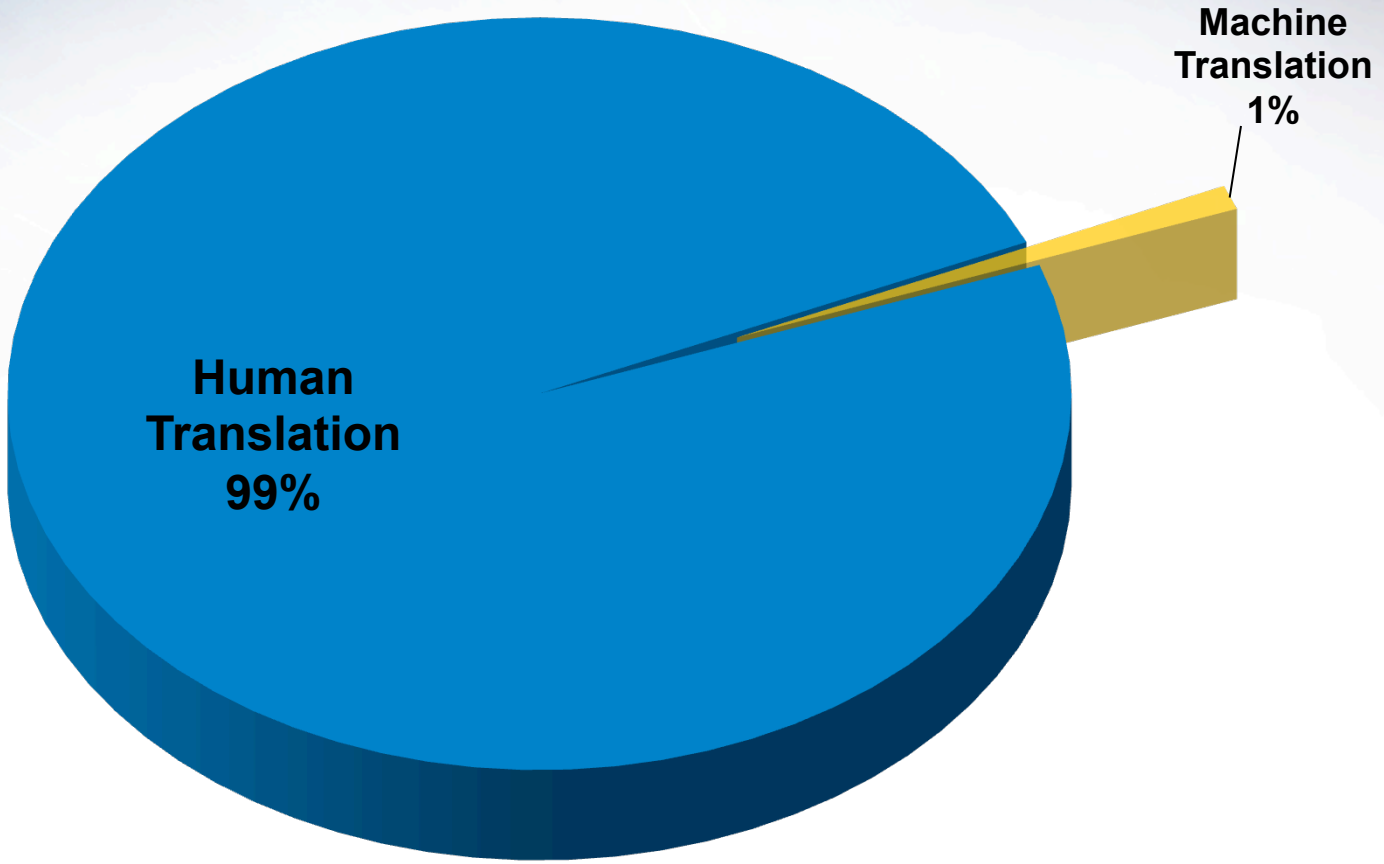
How?



Translation Memories fail!

Prob(10-word-sentence) is too small

Machine Translation Market



Source: Allied Business Intelligence

My wish-list for the perfect TM Platform

MyMemory

Search across all memories, and still be FAST



Web 2.0, Wiki-like contributions



One-click to download a TM



Adaptive Matching, A priori probability



Propagate changes in real-time



Fully integrated Statistical MT



Open standard and CAT support



Barriers to contribution



Privacy and IP



Time vs Value

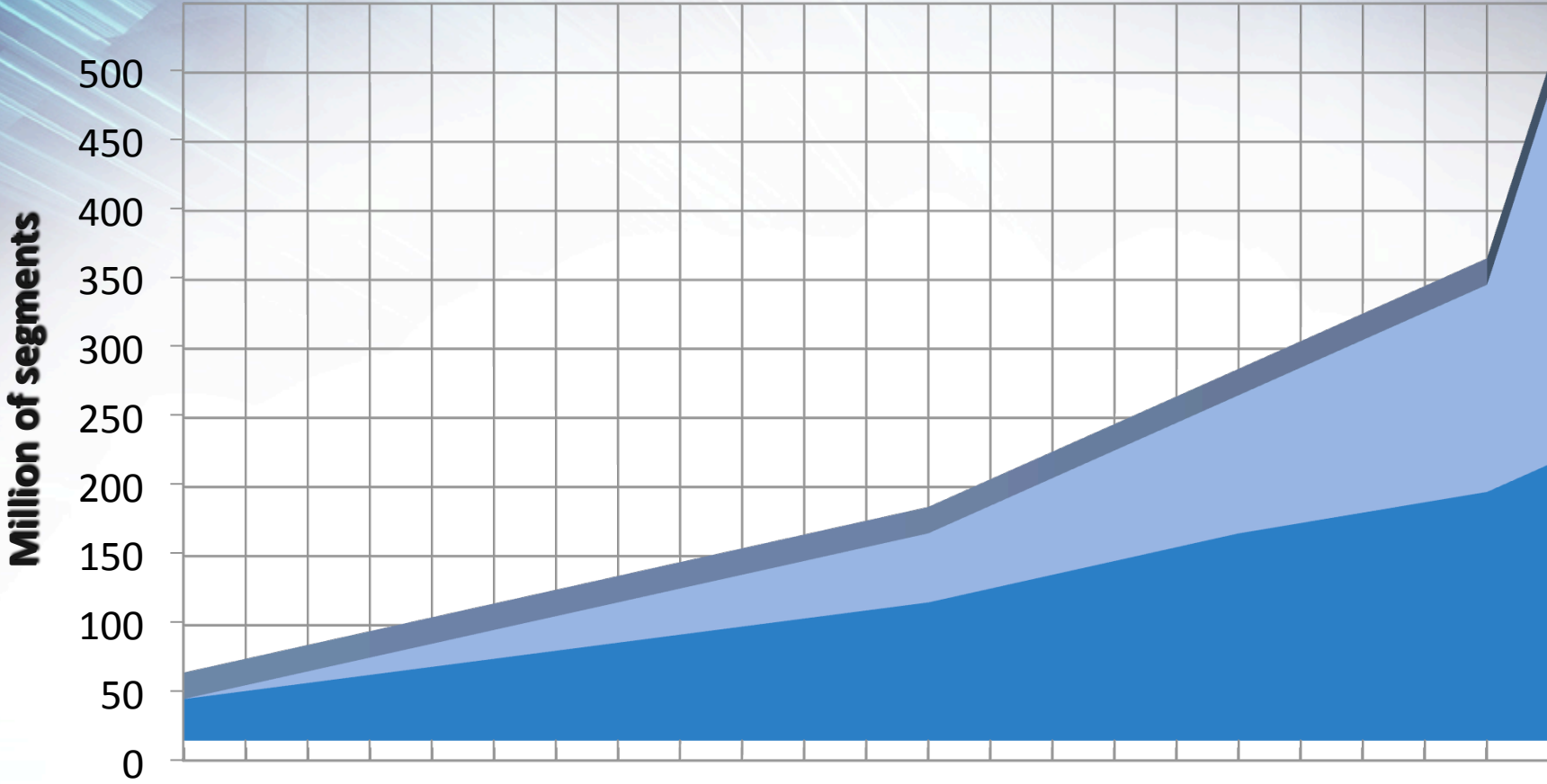
Will people contribute?

Yes!

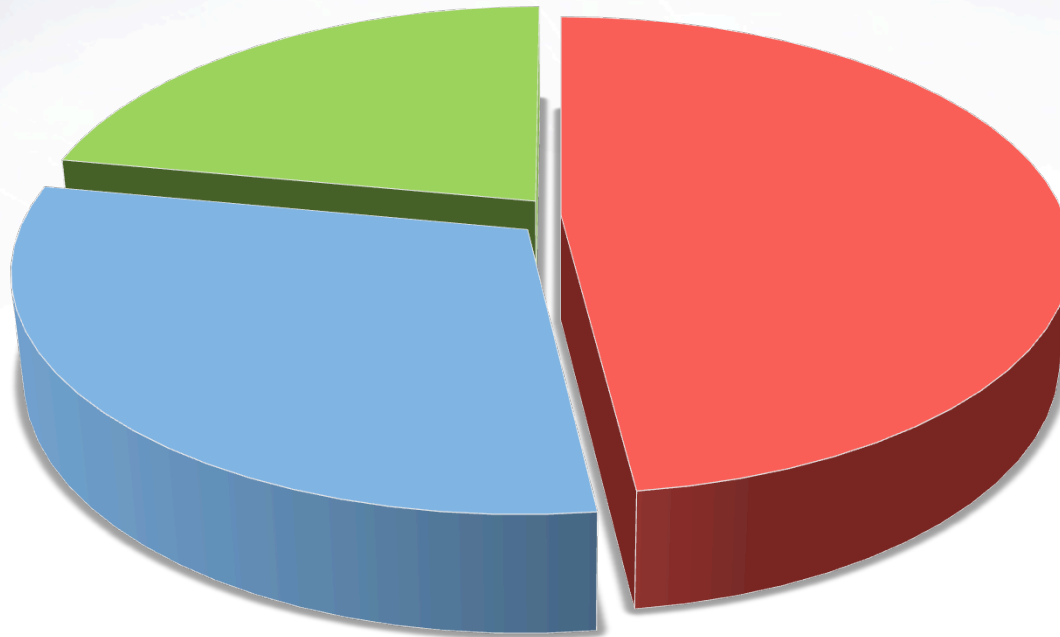
100 words per second during the last year

5 billion words

Contribution Trend



Sources

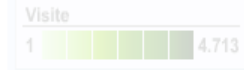


■ Web ■ Translators ■ Customers

How many users yesterday?

74.399 Unique visitors

960.000 Queries



31.215 visite provenienti da 164 Paesi/zone

Livello dettagli: Città | Paese/zona | Regione del subcontinente | Continente Dimensione: Nessuna

Uso del sito Impostazione obiettivo 1

Visualizzazioni:

Visite 31.215 % del totale del sito: 100,00%	Pagine/Visita 2,69 Media sito: 2,69 (0,00%)	Tempo medio sul sito 00:03:34 Media sito: 00:03:34 (0,00%)	% visite nuove 75,08% Media sito: 75,07% (> 0,00%)	Frequenza di rimbalzo 64,47% Media sito: 64,47% (0,00%)
-----------------------------------------------------------	----------------------------------------------------------	-------------------------------------------------------------------------	-----------------------------------------------------------------	----------------------------------------------------------------------

Livello dettagli: Paese/zona

Visite ↓

Pagine/Visita

Tempo medio sul sito

% visite nuove

Frequenza di rimbalzo

Benefits

LSPs & Clients

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

Translators

- ✔ Quick, relevant reference
- ✔ Improved organization and usability
- ✔ Free memories

Technology Challenges

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

Generating the interest



Pre-Populate



Make contributing easy



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 enough?

No.

The language industry produces
**tens of billions of
words per year**

Next Step

**Every CAT tool
connected**

Brave



Live Demo

<http://mymemory.translated.net>