

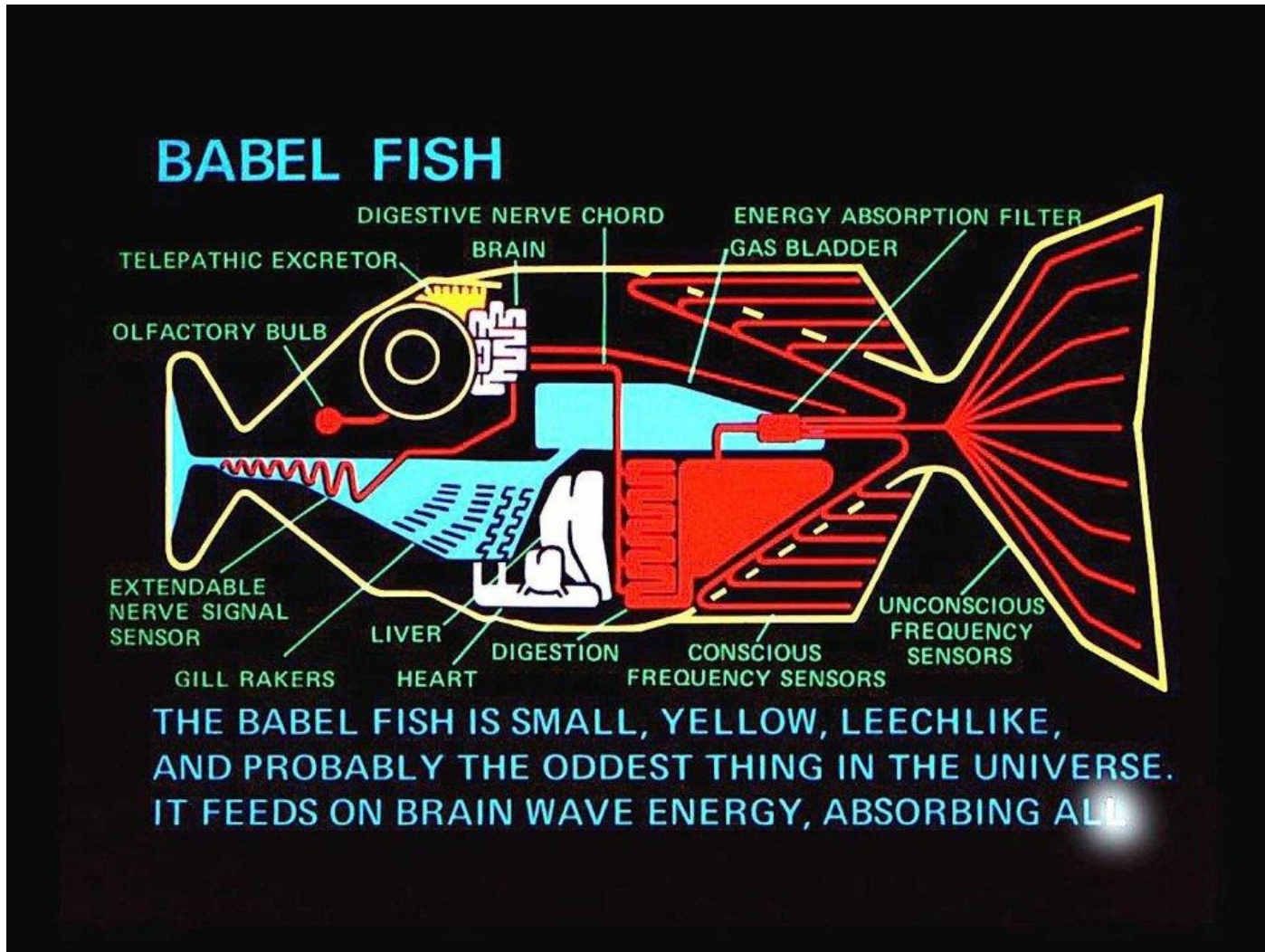


Translation technology:  
Training students to become tomorrow's  
translators

Mark Shuttleworth

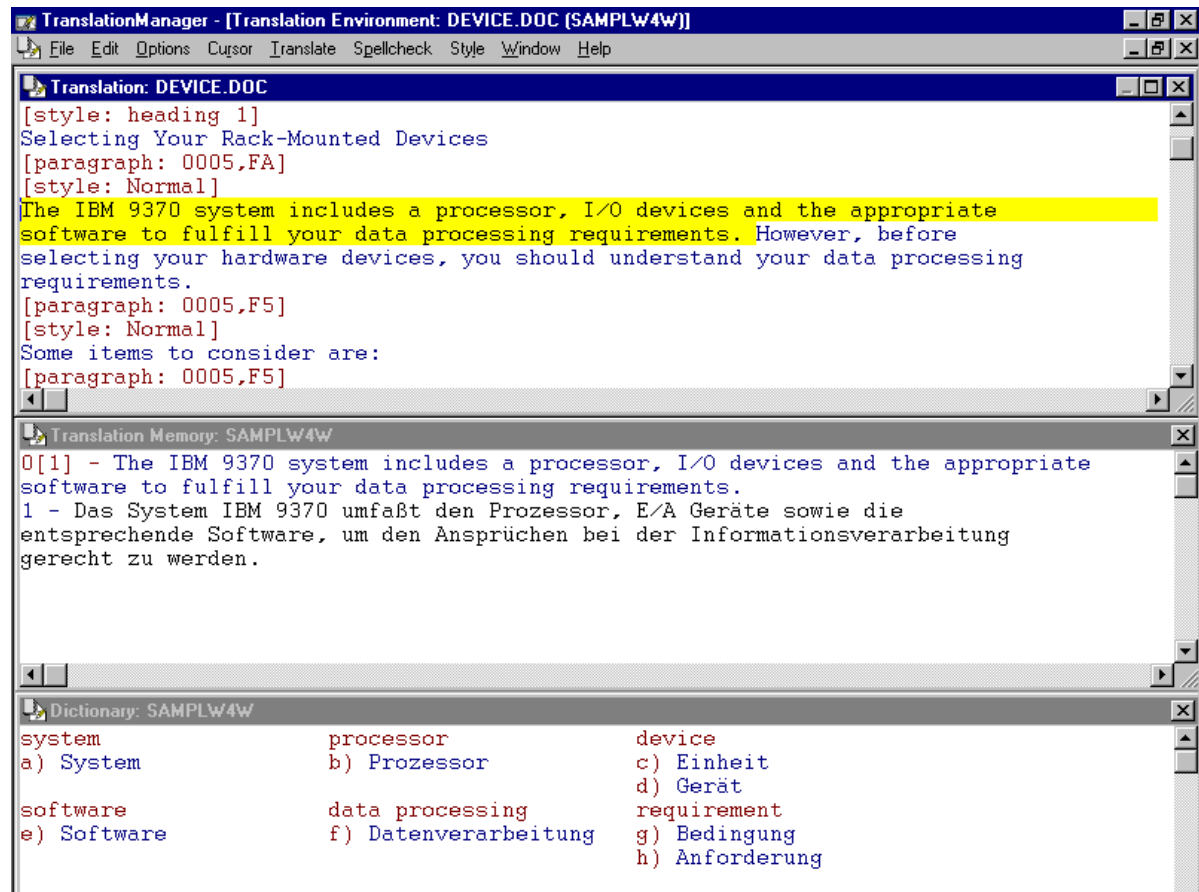


# An early vision of translation technology



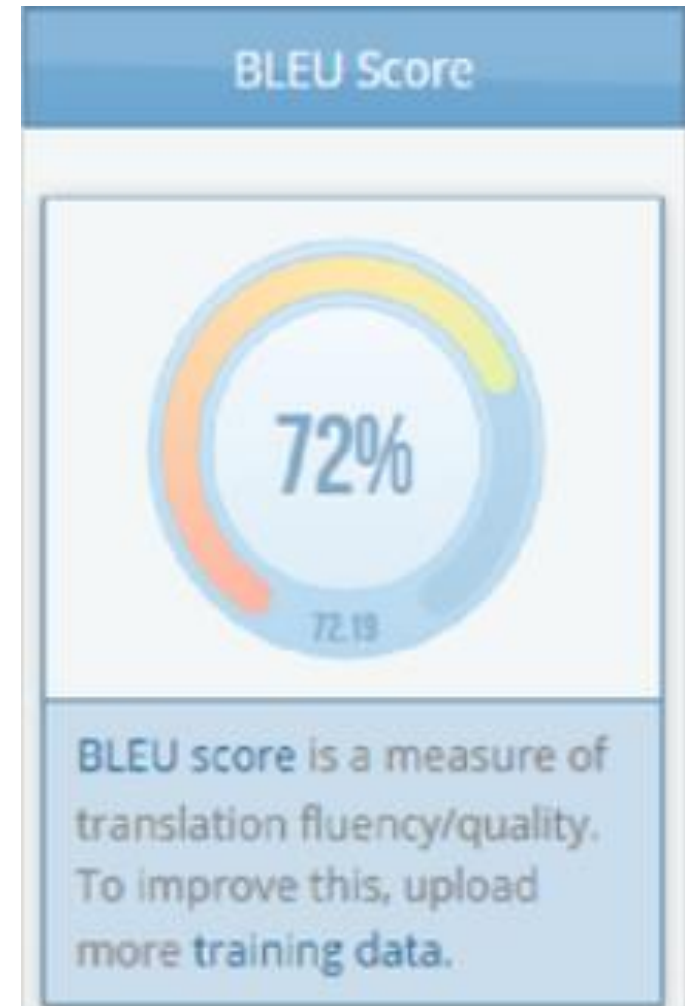
# From IBM TranslationMananger in 1996...

- ▶ one of the first translation memory tools
- ▶ good but basic functionality
- ▶ in one sense, the fundamentals of TM are still very much the same



## ... to KantanMT ...

- ▶ for companies and individual translators:
- ▶ create your own MT engine



# ... and Lilt in 2018

- ▶ interactive, adaptive machine translation
- ▶ learns continuously from previous jobs
- ▶ adapts its translation as you type



1

Les rivières isolées s'écoulent vers la mer,

The isolated rivers flow to the sea,

Next Word Alternatives **Isolated** Rivers

2

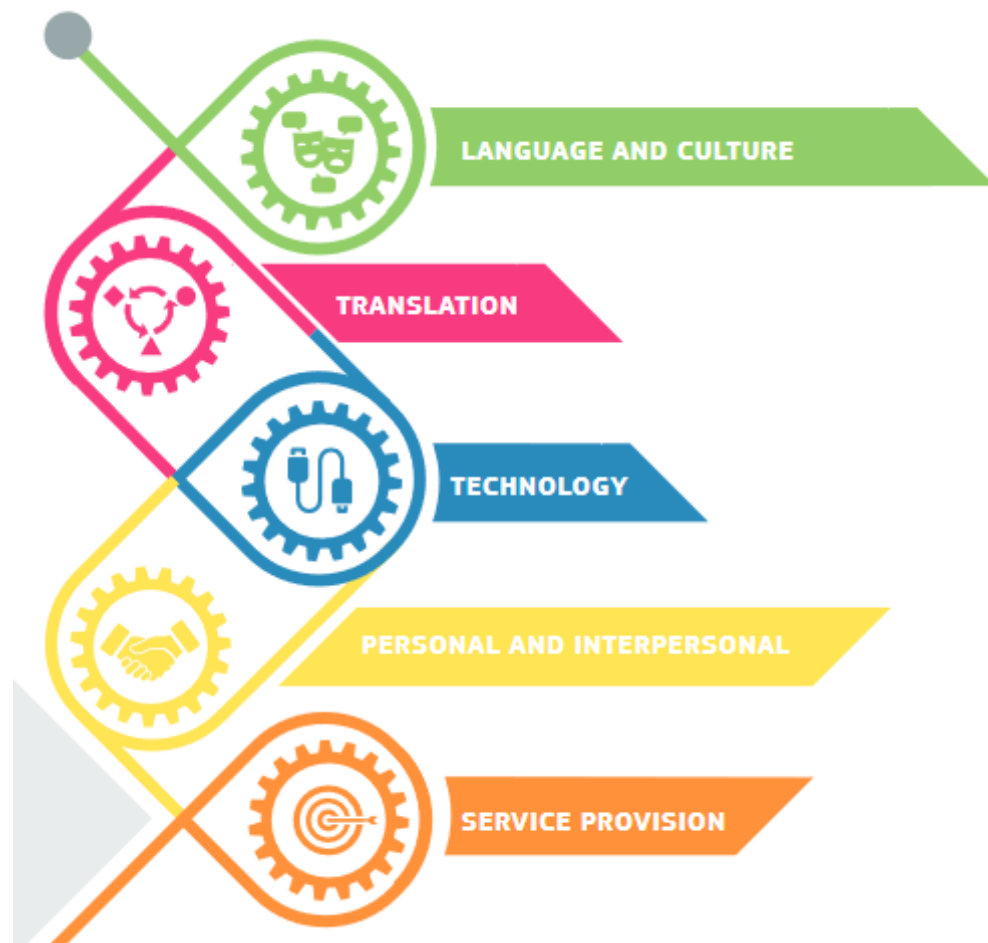
Les rivières isolées soupirent attends moi, attends moi

Rivers isolated sigh look me, look me

Next Word Alternative **The**

# Technology skills fit into a larger picture

- ▶ new benchmarks – e.g. the European Masters in Translation competency framework
- ▶ comprehensive, up-to-date training for translators ...
- ▶ ... and also for project managers, terminologists, etc.



# Translation technology training: a possible blueprint

# Training developed at UCL

- ▶ Understanding and using Translation Technology I and II
- ▶ optional modules on the MA in Translation Theory and Practice, taught between 2014 and 2017
- ▶ mixed language group
- ▶ weekly two-hour session in computer cluster



# Aims

---

- ▶ prepare students for certain roles in the translation industry
- ▶ equip them to use all types of translation technology and for the future
- ▶ ensure theoretical understanding of the technology

# The learning curve

---

1. terminology work
2. translation memory
3. machine translation (+ post-editing)
4. parallel text acquisition
5. building your own MT system
6. tool & system evaluation

# Other skills

---

- ▶ Office – e.g. use of Excel
- ▶ working with PDFs
- ▶ working with different file types
- ▶ importing and exporting: TMX

# What I'd also like to cover, given time

---

- ▶ fourth TM tool (e.g. a newer tool such as MateCat)
- ▶ the Mac
- ▶ XML
- ▶ more on exploiting corpora



# Industry partnerships

---

- ▶ SDL
- ▶ Kilgray
- ▶ KantanMT
- ▶ Screen Systems
- ▶ Lilt
- ▶ + friendly relations with others

# How the modules work

---

- aim to reflect professional realities
- use of software to improve speed and consistency
- learning by doing: task- and problem-driven
- stress on understanding, not skill with a particular system

# Key emphases

---

- ▶ including both theory and practice
- ▶ using the software for practical translation tasks
- ▶ encouraging collaboration and teamwork
- ▶ taking a task-based approach to learning
- ▶ solving problems
- ▶ collaborating within the institution

# Problems

---

- ▶ IT ability range
- ▶ TM don't have time to grow
- ▶ standardised direction of translation?
- ▶ TM vs. MT
- ▶ including both theory and practice
- ▶ not enough time!

# The technology covered

# Terminology acquisition

---

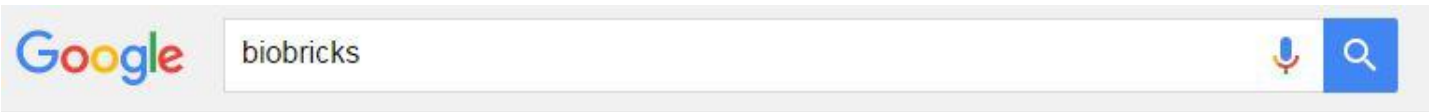
- ▶ use of existing resources:
  - ▶ EU, UN, etc.
- ▶ active search for relevant glossaries

Search by:  Language  Subject  Title

Language(s):		Chinese	English
Subject:			
<input type="button" value="GO!"/>			

- ▶ also e.g. Proz.com

# Terminology acquisition



All Images Videos News Maps More Search tools

Any country Custom Any time All results Clear

## Images for biobricks

Report images



More images for biobricks

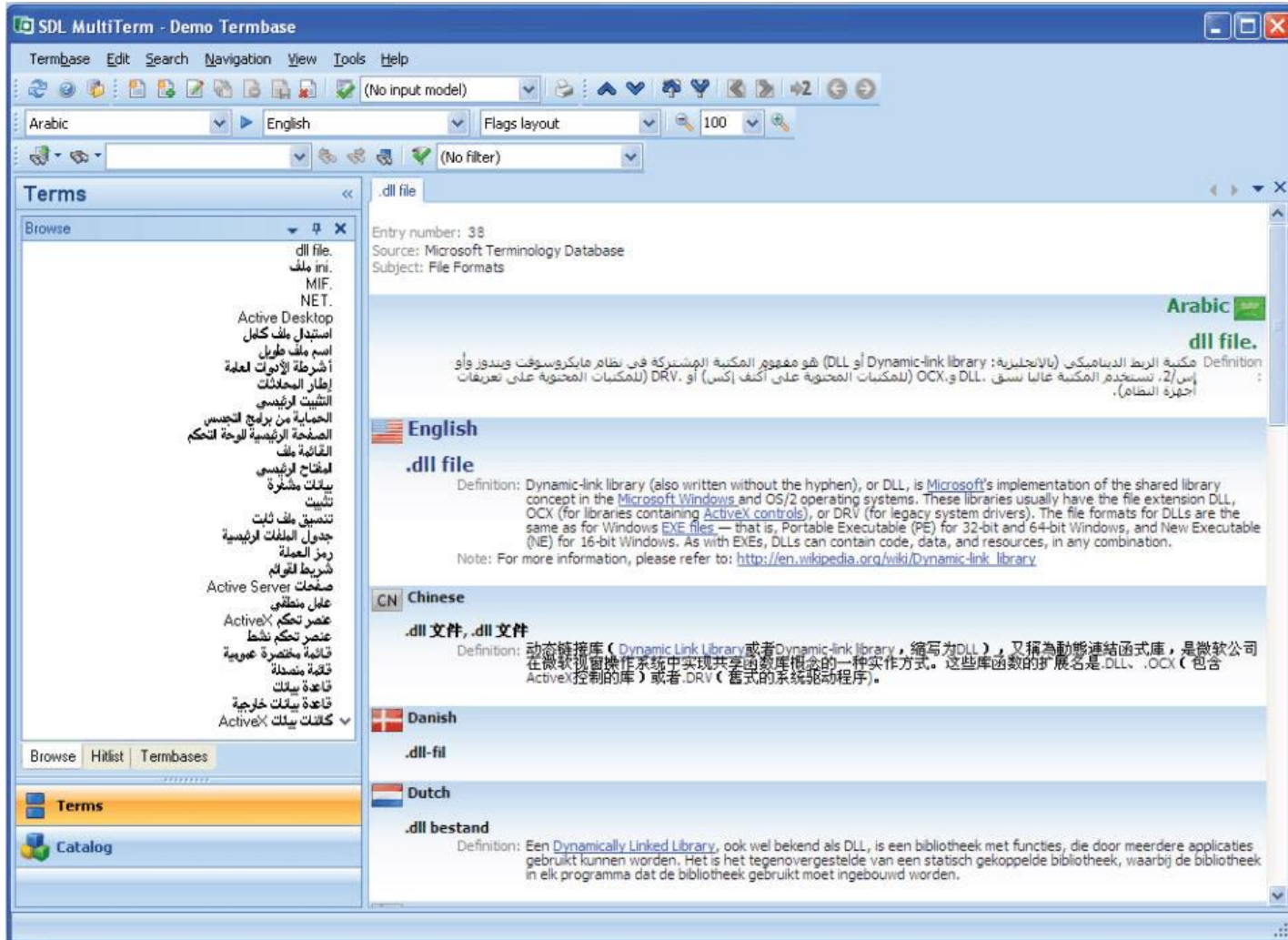
## BioBricks: as partes biológicas padrão da biologia sintética ...

scienceblogs.com.br/.../[biobricks](#)-as-partes-biologicas-padrao-da-bi... [▼ Translate this page](#)  
26 Jan 2011 - A inovação principal dos BioBricks é a padronização do modo de montagem dessas partes padrão: de maneira que a montagem de dois ...

## BioBricks | SynbioBrasil - ScienceBlogs Brasil

scienceblogs.com.br/synbiobrasil/category/[biobricks](#)/ [▼ Translate this page](#)  
8 Mar 2014 - Confira Adventures in Synthetic Biology e aprenda de maneira didática o que são biobrick, PoPs, entre outros conceitos básicos.

# Terminology utilities



# Terminology extraction

- ▶ candidate terms can be extracted semi-automatically from text corpora
- ▶ e.g. Polish multi-word items of Wikipedia terminology
- ▶ demo version of Sketch Engine generally sufficient for this
- ▶ bilingual extraction also possible in principle

Multi-word		Score	F	RefF
<input type="checkbox"/> zawartość wikipedii	<input type="checkbox"/> W	1,478.02	<u>6</u>	<u>0</u>
<input type="checkbox"/> liczba artykułów	<input type="checkbox"/> W	1,311.19	<u>6</u>	<u>0</u>
<input type="checkbox"/> neutralny punkt widzenia	<input type="checkbox"/> W	1,203.14	<u>5</u>	<u>0</u>
<input type="checkbox"/> neutralny punkt	<input type="checkbox"/> W	1,193.87	<u>5</u>	<u>0</u>
<input type="checkbox"/> wszystkie edycje	<input type="checkbox"/> W	1,002.62	<u>5</u>	<u>0</u>
<input type="checkbox"/> wolna licencja	<input type="checkbox"/> W	840.33	<u>4</u>	<u>0</u>
<input type="checkbox"/> angielska Wikipedia	<input type="checkbox"/> W	743.94	<u>3</u>	<u>0</u>
<input type="checkbox"/> treść wikipedii	<input type="checkbox"/> W	739.50	<u>3</u>	<u>0</u>
<input type="checkbox"/> wszystkie wersje	<input type="checkbox"/> W	636.81	<u>5</u>	<u>0</u>
<input type="checkbox"/> dyskusyjna Nupedii	<input type="checkbox"/> W	496.29	<u>2</u>	<u>0</u>
<input type="checkbox"/> angielskojęzyczna Wikipedia	<input type="checkbox"/> W	496.29	<u>2</u>	<u>0</u>

# (Fairly) complete list of TM tools

---



# Translation memory

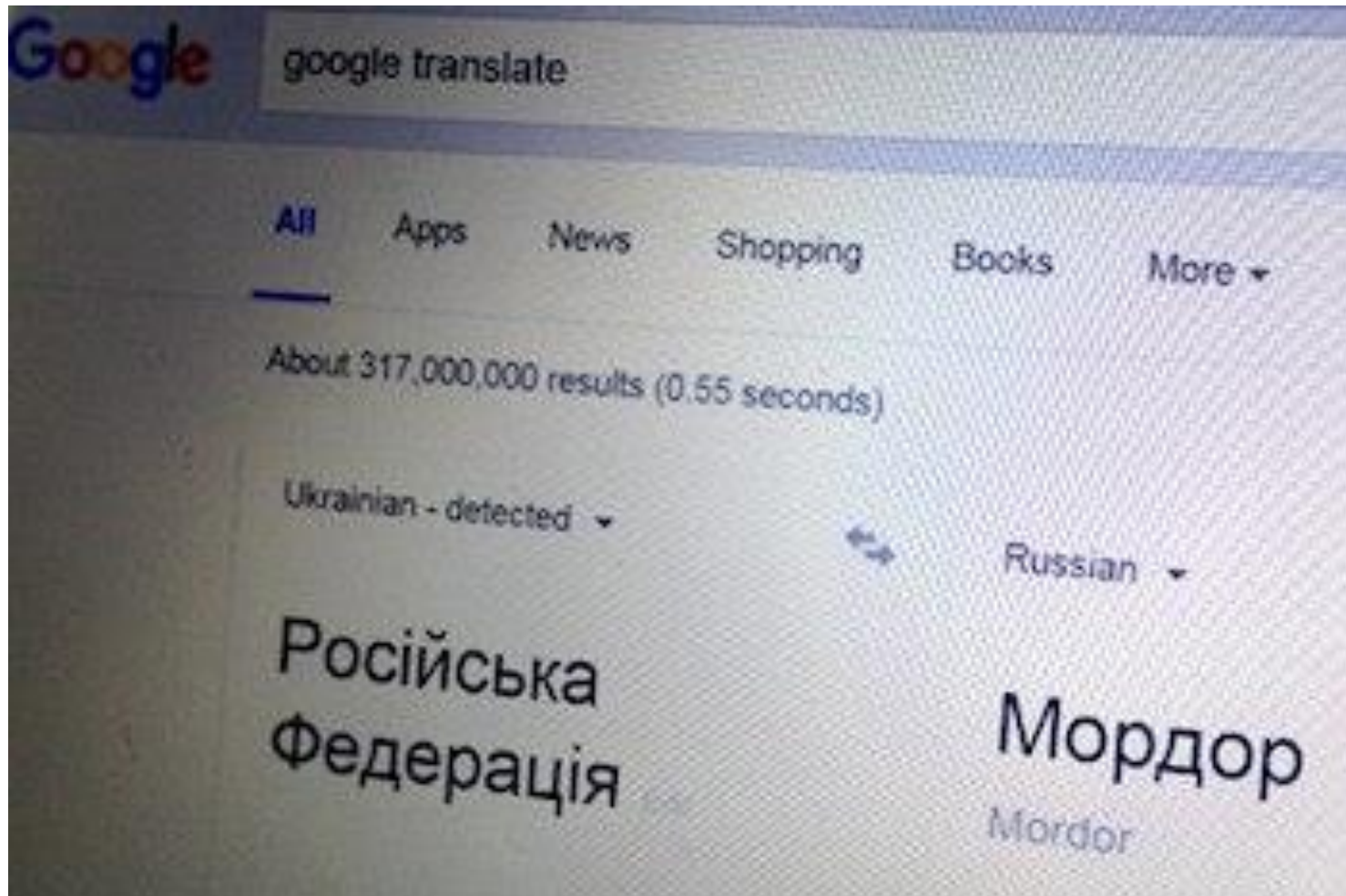
---

- ▶ different systems have the same basic functionality
- ▶ students should be:
  - ▶ competent in 3 tools covered: Wordfast Anywhere, memoQ, SDL Trados Studio
  - ▶ able to learn (or evaluate) a new tool quickly and efficiently



# Machine translation

---



# Machine Translation

- ▶ Google Translate; Google as paid-for service
- ▶ post-editing
- ▶ Google Translator Toolkit
- ▶ how statistical MT works
- ▶ MT within a TM tool

Choosing a hotel for a first trip to Paris was more than a bit overwhelming. Once the neighborhood of Montmartre was the pick, the Mercure seemed a good fit. And it was. Hotel is quiet, the room on the 9th floor is modern and well appointed. Spectacular view of the Eiffel Tower at night. The neighborhood is great for walking and the staff especially Birmunther very helpful to 2 non French speakers. I would choose this hotel again

Wybór hotelu na pierwszą wycieczkę do Paryża był nieco przytłaczający. Gdy tylko okolica Montmartre była wyborem, hotel Mercure wyglądał na dobrze dopasowanego. I to było. Hotel jest cichy, pokój na 9. piętrze jest nowoczesny i dobrze wyposażony. Spektakularny widok na Wieżę Eiffla w nocy. Okolice jest świetna na spacer, a personel, zwłaszcza Birmunther, jest bardzo pomocny dla 2 osób spoza Francji. Wybrałbym ten hotel ponownie

Okolice jest świetna na spacer, a personel, zwłaszcza Birmunther, jest bardzo pomocny dla 2 osób spoza Francji.  
Okolice jest doskonałym miejscem na spacer, a personel bardzo pomocny zwłaszcza Birmunther 2 non frankofonów.

Improve this translation

ocny dla 2 osób spoza Francji.

Use

# Parallel text acquisition

Translated.net  
My Memory

Logged in as: **markshut**

Upload the document to translate, we will give you a TMX.  
Please [email us](#) if the TMX has problems.

**Get a memory for my document** | **Export a memory** | **History**

Choose a name:

File to upload:   .html, .doc, .tmx, .txt or .zip

Choose Languages:

Choose Subject:

Use only your own memories **i**

Use Statistical Machine Translation for no matches **i**

**Create memory**

# Parallel text acquisition

Translated.net  
My Memory

Logged in as: markshut

Upload the document to translate, we will give you a TMX.  
Please [email us](#) if the TMX has problems.

Get a memory for my document | Export a memory | **History**

Name	Source	Target	Subject	Segments	Date	Status	Actions
Unnamed	German	English	Science	1276	17 Feb 2015	Ready <a href="#">View stats</a>	<a href="#">Download</a>   <a href="#">Delete</a>

# ORPUS ... the open parallel corpus

---

corpus	doc's	sent's	src tokens	trg tokens	XCES/XML	raw	TMX	Moses
<b>GNOME</b>	1873	0.7M	4.0M	4.6M	[ xces en pt_BR ]	[ en pt_BR ]	[ tmx ]	[ mooses ]
<b>KDE4</b>	2163	0.2M	2.4M	2.6M	[ xces en pt_BR ]	[ en pt_BR ]	[ tmx ]	[ mooses ]
<b>Ubuntu</b>	452	0.2M	0.8M	0.7M	[ xces en pt_BR ]	[ en pt_BR ]	[ tmx ]	[ mooses ]
<b>PHP</b>	3281	40.4k	0.5M	0.2M	[ xces en pt_BR ]	[ en pt_BR ]	[ tmx ]	[ mooses ]
<i>total</i>	<b>7769</b>	<b>1.1M</b>	<b>7.8M</b>	<b>8.0M</b>	<b>1.1M</b>		<b>0.2M</b>	<b>0.3M</b>

(<http://opus.nlpl.eu/>)

# MT (and TM) evaluation

---

- ▶ a hands-on introduction to human-based MT evaluation...
- ▶ ... and automatic measures
- ▶ the latter focuses on the BLEU metric
- ▶ this will be used later on, in the KantanMT project



# The Museum project

# UCL's Grant Museum of Zoology

- ▶ only university zoological museum in London
- ▶ one of three UCL museums
- ▶ houses 68,000 specimens



# The Grant Museum project

## TOP TEN OBJECTS

Grant Museum of Zoology  
self-guided tour



### Quagga skeleton

\* Extinct and Endangered\* display  
near museum entrance

There are only seven known examples of this South African zebra, extinct since 1883, making this the rarest skeleton in the world.



### Thylacines

\* Extinct and Endangered\* display  
near museum entrance

These dog-like carnivorous marsupials were deliberately driven to extinction in 1936. The Museum houses extremely rare preserved dissections and skeletons.



### Dodo bones

\* Extinct and Endangered\* display  
near museum entrance

These icons of extinction had disappeared by 1681. The Grant Museum has a large assemblage of bones.



### Giant deer

By the museum entrance

This ice-age wonder went extinct about 7000 years ago. Its antlers were almost twice as wide as its height.



### Blaschka glass models of invertebrates

Case 14, near the museum door

A famous collection of anatomically perfect models of species which were hard to preserve, made by renowned 19th century Czech jewellers.

For more information about these objects visit  
[ucl.ac.uk/museums/zoology/about/collections/objects](http://ucl.ac.uk/museums/zoology/about/collections/objects)



### The Micrarium: A place for tiny things

Opposite Case 5

A beautiful back-lit cave of 2300 microscope slides giving a glimpse of the vast diversity of animal life, nearly all of which is minute.



### Jar of preserved moles

Case 12, shelf 3 near the museum door

One of the most bizarre objects in the Museum's collection is a large jar crammed full of whole preserved moles.



### The brain collection

Case 13, shelf 3 near the museum door

An unusual comparative anatomy collection displaying the differences between the brains of a number of mammals and a reptile.



### Rock python skeleton

Case 16. Back, right, of the museum

This intricate 5m long snake skeleton is unusually displayed wrapped around a branch. The animal lived at London Zoo.



### The Negus collection of bisected heads

Cases 21 and 94

These preserved heads were prepared to compare the anatomy of animal throats, including a chimp, seal, orang-utan, red panda and alligator.

### visitor information

Grant Museum of Zoology,  
UCL  
Rocefeller Building,  
21 University Street,  
London WC1E 6DE  
Monday-Saturday 1-5pm  
+44 (0)20 3108 2052  
[zoology.museum@ucl.ac.uk](mailto:zoology.museum@ucl.ac.uk)

@GrantMuseum

[ucl.ac.uk/museums/zoology](http://ucl.ac.uk/museums/zoology)

# The Grant Museum project

---

- ▶ group divided into teams
- ▶ each person given a role: translator, terminologist or project manager
- ▶ the task: translate the 'Top Ten Objects' document + other materials into the team's 1-2 languages against a strict deadline
- ▶ team files and individual reports submitted for assessment

# The Grant Museum project (Spanish)

**DIEZ OBJETOS DESTACADOS**

**Museo Grant de Zoología**  
tour autoguiado

Para más información acerca de estos objetos, visite:  
[ucl.ac.uk/museums/zoology/about/collections/objects](http://ucl.ac.uk/museums/zoology/about/collections/objects)

**Esqueleto de quagga**  
Vitrina 'Extinguidos y en Peligro de Extinción', junto a la puerta del museo  
Existen únicamente siete ejemplares conocidos de esta cebra de Sudáfrica que se extinguió en 1883. Esto lo convierte en el esqueleto más escaso del mundo.

**Tilacino**  
Vitrina 'Extinguidos y en Peligro de Extinción', junto a la puerta del museo  
Estos marsupiales carnívoros con aspecto de perro fueron exterminados deliberadamente en 1936. El museo alberga disecciones y esqueletos muy poco conocidos que han sido preservados.

**Huesos de dodo**  
Vitrina 'Extinguidos y en Peligro de Extinción', junto a la puerta del museo  
Estos símbolos de la extinción desaparecieron en 1681. El Museo Grant tiene una gran colección de huesos.

**Ciervo gigante**  
Junto a la entrada del museo  
Esta maravilla de la Edad de Hielo se extinguió hace unos 7.000 años. La longitud de sus astas casi doblaba la altura del animal.

**Modelos de invertebrados en cristal de Blaschka**  
Vitrina 14, junto a la puerta del museo  
Una famosa colección de modelos de especies anatómicamente perfectas, difíciles de preservar y realizada por un grupo de prestigiosos joyeros checos del siglo XIX.

**El «Micranium», un lugar para cosas diminutas**  
Lado opuesto, vitrina 5  
Una bella gruta con 2.300 platinas de microscopio que permiten echar un vistazo a la gran diversidad de la vida animal, de la cual la enorme mayoría de la cual es diminuta.

**Frasco con topes conversados**  
Vitrina 12, estante 3, junto a la puerta del museo  
Uno de los objetos más extraños de la colección del museo, se trata de un frasco repleto de topes conservados íntegramente.

**Colección de cerebros**  
Vitrina 13, estante 3, junto a la puerta del museo  
Colección anatómica comparativa poco común que exhibe las diferencias entre los cerebros de una serie de mamíferos y reptiles.

**Esqueleto de piñón de Seba**  
Vitrina 16, al fondo a la derecha del museo  
Este complejo esqueleto de serpiente, que cuenta con cinco metros de largo, se exhibe en una posición poco habitual, enroscado en una rama. El animal vivía en el zoo de Londres.

**Colección Negus de cabezas diseccionadas**  
Vitrinas 21 y 94  
Esta colección de cabezas conservadas fue preparada para comparar la anatomía de las gargantas de los animales, e incluye cabezas de chimpancé, foca, orangután, panda rojo y caimán.

**Información para visitantes**  
Museo Grant de Zoología, UCL  
Rockefeller Building, 21  
University Street, London  
WC1E 6DE  
Lunes-Sábado, 13.00-17.00  
+44 (0)20 3108 2052  
[zoology.museum@ucl.ac.uk](mailto:zoology.museum@ucl.ac.uk)

[ucl.ac.uk/museums/zoology](http://ucl.ac.uk/museums/zoology)

@GrantMuseum

# The Grant Museum project (Chinese)

UCL

## 十大藏品

格兰特动物学博物馆  
自助导览

更多藏品信息请点击：  
[ucl.ac.uk/museums/zoology/about/collections/objects](http://ucl.ac.uk/museums/zoology/about/collections/objects)

---



**鹿头骨架**  
“灭绝与濒危”动物展  
近博物馆入口  
斑羚已于1883年灭绝，全世界范围内仅存七具骨架。该藏品极其稀有。



**袋象**  
“灭绝与濒危”动物展  
近博物馆入口  
这些外观似狗的食肉有袋类动物已于1936年因人类活动故意灭绝。本馆藏有珍稀的袋象解剖标本和骨架。



**渡渡鸟骨**  
“灭绝与濒危”动物展  
近博物馆入口  
渡渡鸟灭绝于1681年，是最为知名的灭绝物种之一。馆藏有大量渡渡鸟骨头。



**大角鹿**  
博物馆入口处  
大角鹿生活在冰河时期，于约7000年前灭绝。该物种的鹿角约为其身高两倍。



**无脊椎动物布拉施卡玻璃模型**  
14号展柜，近博物馆大门  
一些在19世纪由知名捷克珠宝师制作的无脊椎动物标本，完美地保存了其生理结构，制作工艺非常复杂。



**迷你馆：小小动物集合处**  
5号展柜对面  
在美丽的展厅内背景灯的映照下，2300多张显微照片让您惊叹动物界的多样性。照片拍摄的几乎都是体型很小的动物。



**罐鼠罐**  
12号展柜，3号架，近博物馆大门  
本馆藏奇特的收藏之一就是两两一罐完整保存的鼠眼。



**大脑收藏**  
13号展柜，3号架，近博物馆大门  
一个特别的比较解剖学展览，让您一窥哺乳动物大脑和爬行动物大脑的区别。



**岩蜥骨架**  
16号展柜，馆内深处靠右  
骨架长约五米。特别之处在于，该骨架盘绕在树枝上。骨架的主人曾居住在伦敦动物园。



**尼格斯羊剖头颅展览**  
21号展柜，94号展柜  
展品包括一只黑猩猩、一只海狗、一只猩猩、一只小熊猫和一只短吻鳄对羊剖开的头颅。这些头颅最早是为比较这些动物的眼眶结构而准备的。


[ucl.ac.uk/museums/zoology](http://ucl.ac.uk/museums/zoology)

**参观信息：**  
Grant Museum of Zoology,  
UCL  
Rookefeller Building,  
21 University Street,  
London WC1E 6DE  
周一至周五下午 1-5 点  
+44 (0)20 3108 2052  
[zoology.museum@ucl.ac.uk](mailto:zoology.museum@ucl.ac.uk)

@GrantMuseum

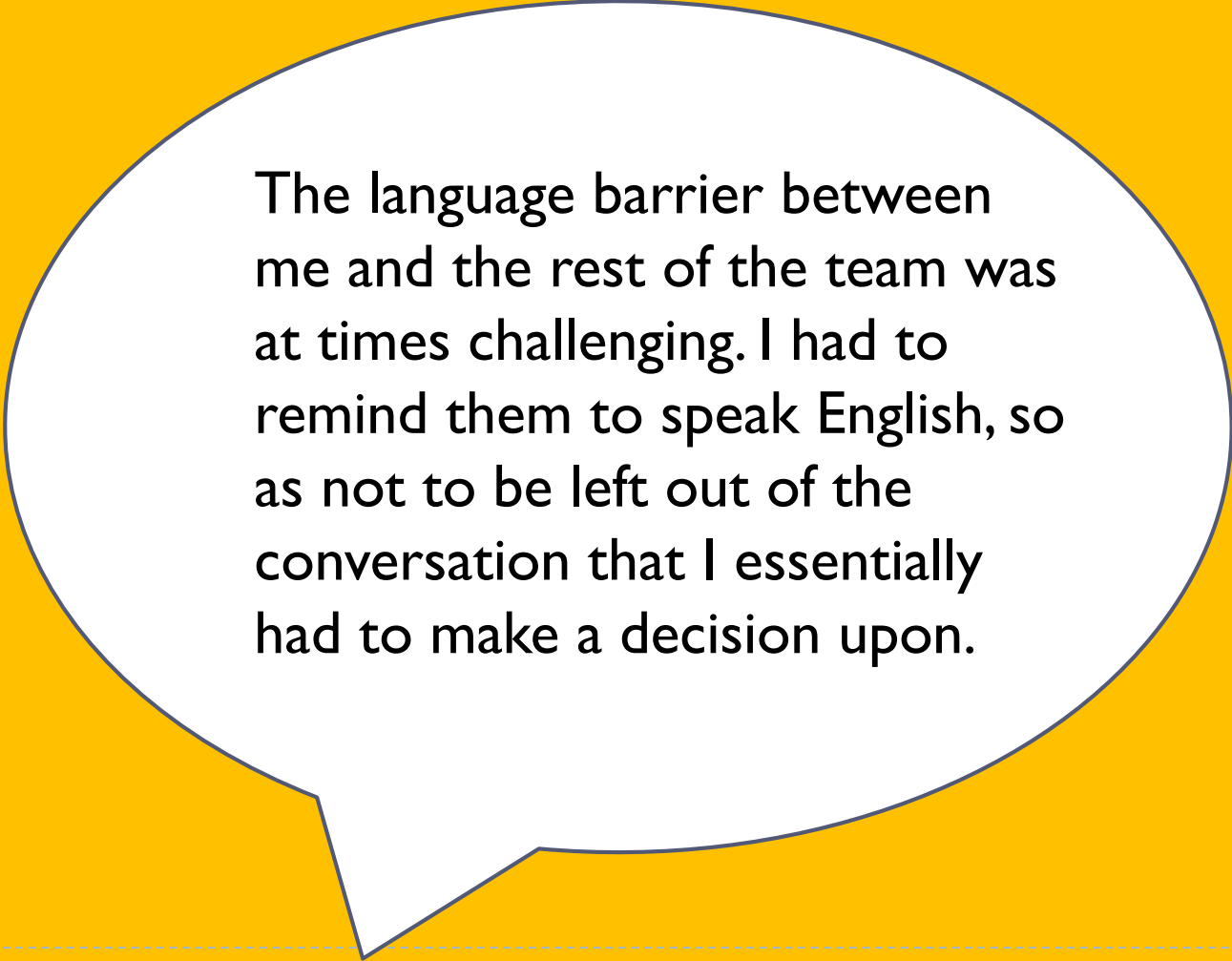
# What the long-suffering PMs said

---




All in all, our project proceeded very well and I am proud of our work and my teammates.

---



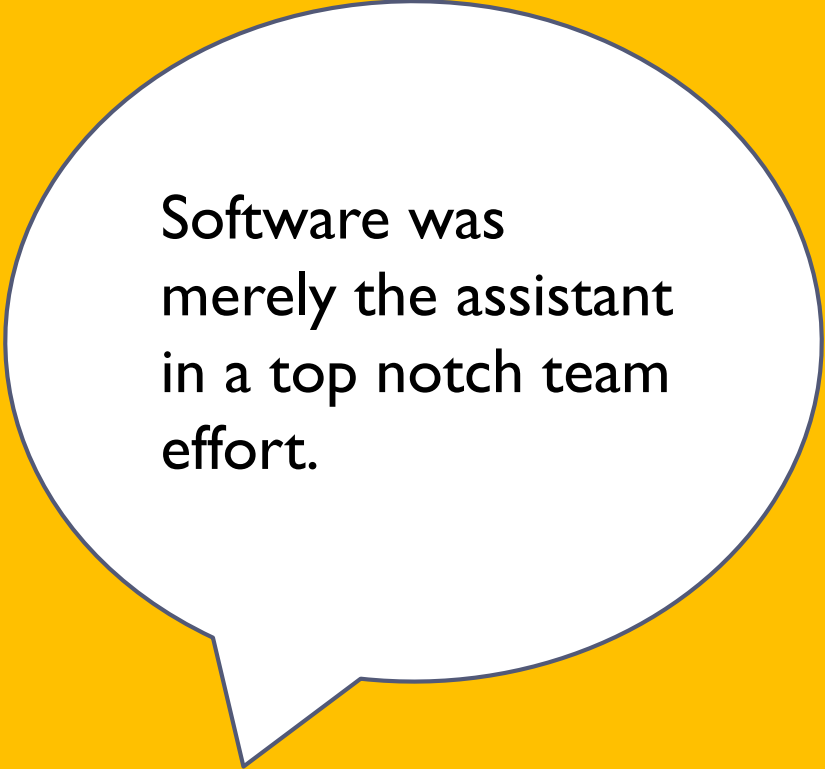
The language barrier between me and the rest of the team was at times challenging. I had to remind them to speak English, so as not to be left out of the conversation that I essentially had to make a decision upon.

---



Conclusively, we made the best of the resources available to us; both our own personal expertise and the software at hand. What carried us through though was communication.


---



Software was  
merely the assistant  
in a top notch team  
effort.

It was a bit stressful to see that other groups quickly abandoned the baffling server and shifted to other online services to exchange files and got head starts, but I decided that we should still spend some extra time on our exploration with the server since we still had enough time left for starting over again. Though there were some twists and turns, I managed to settle down the workflow.

---



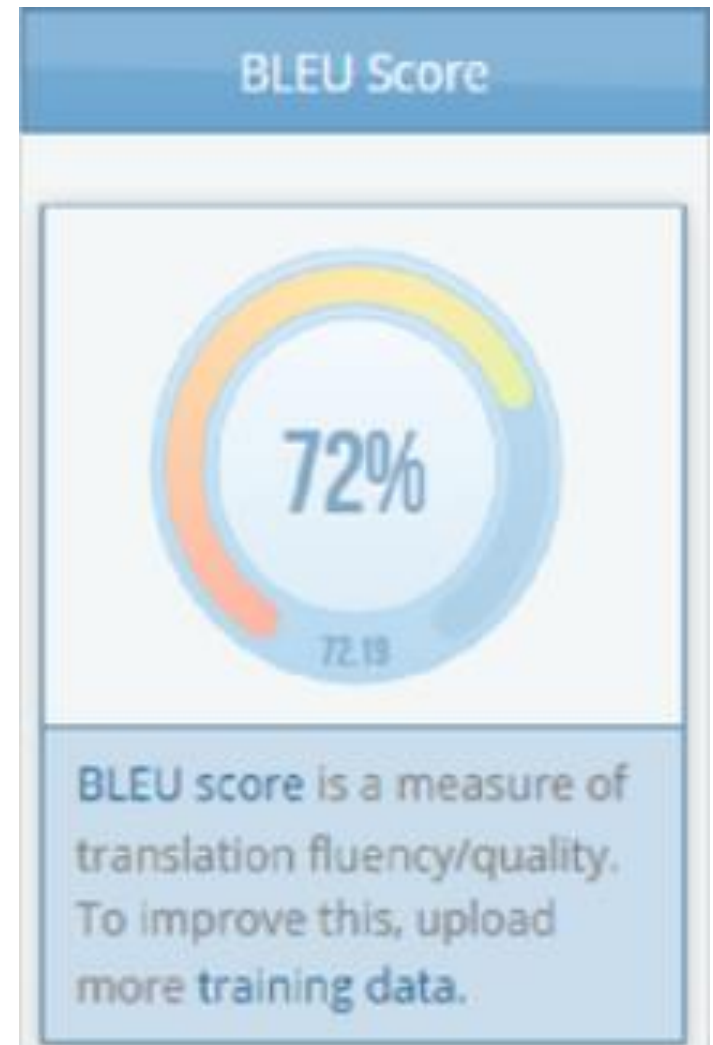
My group members not only shared with me their discoveries, but also unswervingly supported my decisions. I want to extend my greatest gratitude to them at the end of this report.

# The KantanMT project

# KantanMT

## ► The task:

1. choose a subject area and select a test text
2. create an MT engine and choose a 'stock engine'
3. try to improve the quality of the engine incrementally by adding resources to it
4. track the BLEU score
5. write reflectively on your experiences



# What resources one student uploaded

Name	Source	Format	Size	TUs
MONOLINGUAL_WIKI (x12)	Wikipedia articles (x12)	DOCX	424 KB	-
TB_MicrosoftTerm...	<a href="#">Microsoft.com</a>	TBX	13.7 MB	-
TB_IATE_Information...	<a href="#">Multifarious</a>	TBX	23 MB	-
Personal_TM	PC	TMX	219 MB	355,872
TM_Translated.net...	<a href="#">Translated.net</a>	TMX	108 KB	208
TM_Microsoft	<a href="#">TTMEM</a>	TMX	8.66 MB	21,160



# What resources one student uploaded

---

TM_IATE(sample)	<a href="#">ITMEM</a> (sample)	TMX	443 KB	1,000
TM_Gnome	<a href="#">OPUS</a>	TMX	596 KB	2,262
TM_DGT (x11)	<a href="#">DGT</a>	TMX	1.66 GB	2,249,419
TM_Tatoeba	<a href="#">OPUS</a>	TMX	60.4 MB	136,192
TM_UN_Corpora	<a href="#">UN CORPORA</a>	TMX	52.8 MB	52,909
<b>TOTAL (TM)</b>	-	<b>TMX</b>	<b>≈ 2 GB</b>	<b>2,819,022</b>





# What the team was involved in

---

- ▶ investigating features of existing tools
- ▶ wiki writing
- ▶ blogging
- ▶ brainstorming ideas
- ▶ thinking about design issues



# The app

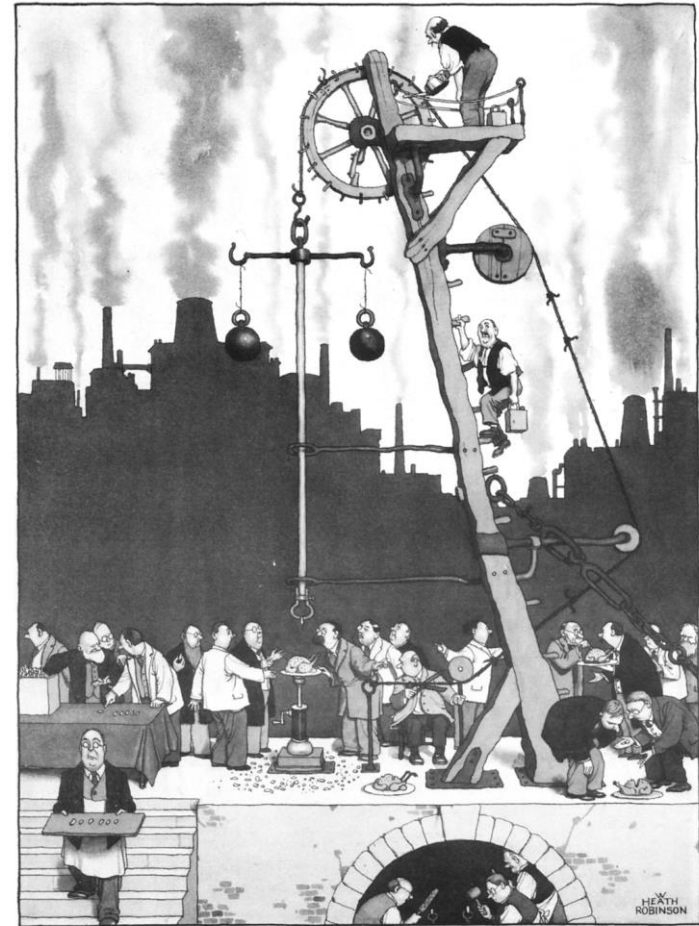
---

- ▶ mono- and bilingual glossary searching
- ▶ 64 languages
- ▶ first version for Android

ar	hy
Arabiese	armenian
عربية	أرمينية
arab	erməni
арабски	армянски
арабски	арменски
আরবি	আমেনিয়ান
àrab	armeni
arabština	arménský
Arabeg	Armeneg
arabisk	armensk
Arabisch	Armenisch
Αραβικά	Αρμένικα

# Has our training kept pace with reality?

- ▶ the modules continue to be based on translation memory, in its more recent forms
- ▶ however, students also need to know how to:
  - ▶ create and capture content
  - ▶ absorb a range of different types of technology
  - ▶ adopt a flexible approach to technology
  - ▶ learn new tools quickly and confidently
  - ▶ know when not to use technology



Testing Artificial Teeth

# Destinations

---

- ▶ freelance translators
- ▶ translator at FIFA.com
- ▶ translator at KPMG
- ▶ translator in Prime Minister's office
- ▶ translation/localisation project managers
- ▶ machine translation language specialist
- ▶ lecturer
- ▶ translator at SDL
- ▶ translation company owner
- ▶ trainer at SDL



And finally...

# This Earpiece Translator Claims To Let You Talk To People In A Foreign Language

---


- ▶ Douglas Adams' Babel fish: is the future already here?



# 'The Pilot' by Waverly Labs

---



A collection of broken oyster shells and pieces of agate on a concrete surface. The shells are dark and some show the interior. The agate pieces are translucent and have various colors like brown, orange, and grey. The background is a rough, grey concrete wall.

Dziękuję  
bardzo!  
Ačiū!  
Thank you!



Mark Shuttleworth



[m.shuttleworth@ucl.ac.uk](mailto:m.shuttleworth@ucl.ac.uk)



[ucl.academia.edu/MarkShuttleworth](http://ucl.academia.edu/MarkShuttleworth)



[www.termseeker.wordpress.com](http://www.termseeker.wordpress.com)



[www.facebook.com/termseeker](http://www.facebook.com/termseeker)



# Photo credits

---

- ▶ Pebble stack (slide 13): © Zzubnik – Own work, CC BY 3.0, <https://commons.wikimedia.org/w/index.php?curid=4086108>
- ▶ Mr pipo Learning and forgetting (slide 23): © Nevit Dilmen [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>) or GFDL (<http://www.gnu.org/copyleft/fdl.html>)], via Wikimedia Commons, [https://commons.wikimedia.org/wiki/File%3AMr\\_pipo\\_Learning\\_and\\_forgetting.svg](https://commons.wikimedia.org/wiki/File%3AMr_pipo_Learning_and_forgetting.svg)
- ▶ Heath Robinson invention (slide 50): © Por Amada44 (From the Book: William Heath Robinson Inventions) [Public domain], undefined ([https://commons.wikimedia.org/wiki/File%3AWilliam\\_Heath\\_Robinson\\_Inventions\\_-\\_Page\\_059.png](https://commons.wikimedia.org/wiki/File%3AWilliam_Heath_Robinson_Inventions_-_Page_059.png))
- ▶ Slides 8, 46: © UCL Media Services - University College London
- ▶ Slides 1, 9-12, 14-16, 29, 55: © Mark Shuttleworth