



Language resources for Public Translation Services

Practical uses of data sharing - Tapadóir

- 
1. Overview of technology used by Irish-language translators
 2. Importance of consistency in terminology
 3. Translation memories
 4. Machine translation
 5. Practical experience



Technology used by Irish-language translators



On average, a professional translator...

- Translates 2500-3000 per day
- Proofreads 1000 words per hour
- Post-edits ????? words per day

Various types of technology used by translators

- Online dictionaries
- Online terminology databases
- Searchable websites
- Translation memories
- Machine translation

Online dictionaries



☰ 🔍 Type a word or phrase

Similar words: Cailceach · Cipireach · Doiminiceach · Peirseach · Seiceach

FGB FB EID GR 🔊

imirceach ★ 📄

- ADJECTIVE
- 1st DECLENSION

Singular

NOMINATIVE
imirceach (MASC.)
imirceach (FEM.)

GENITIVE
imircigh (MASC.)
imircí (FEM.)

Plural

NOMINATIVE
imirceacha

Online terminology databases



téarma.ie

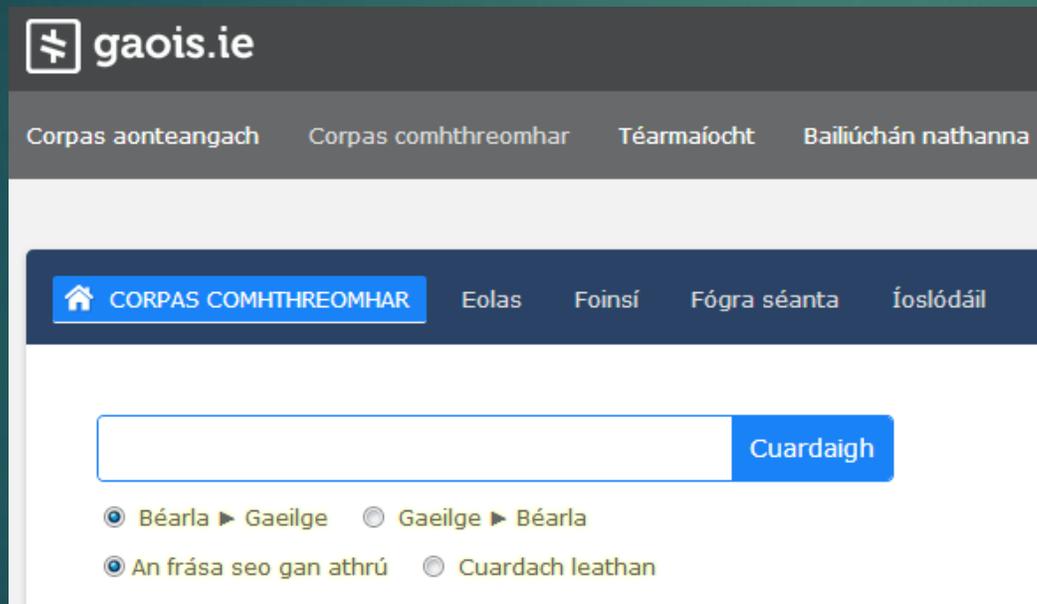


EUR-Lex
Access to European Union law



Iate InterActive Terminology
for Europe

Searchable websites



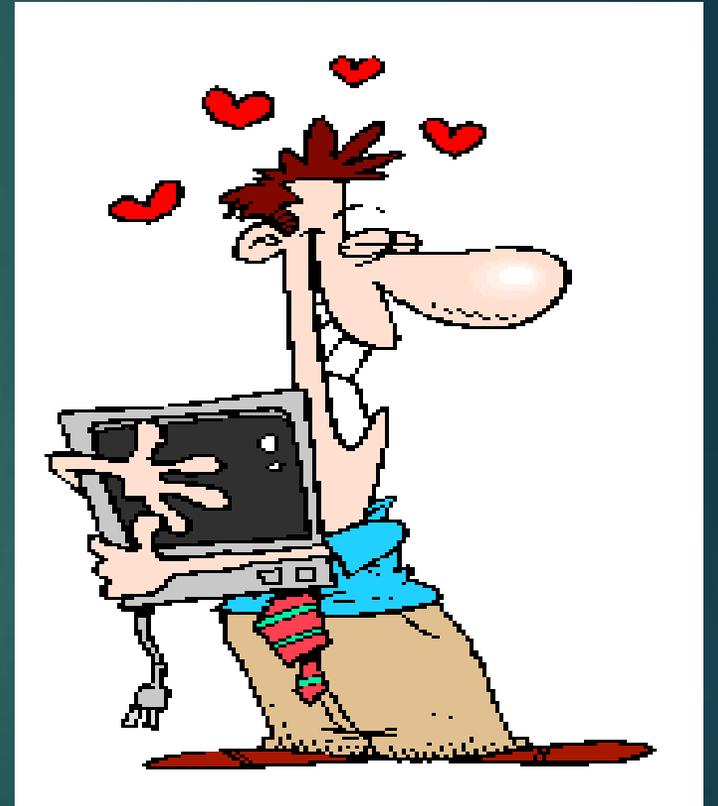
The screenshot shows the gaois.ie website interface. At the top left is the logo and domain name "gaois.ie". Below it is a navigation bar with four items: "Corpas aonteachach", "Corpas comhthreomhar", "Téarmaíocht", and "Bailiúchán nathanna". A secondary navigation bar contains a home icon, "CORPAS COMHTHREOMHAR", "Eolas", "Foinsí", "Fógra séanta", and "Íoslódáil". The main content area features a search input field with a blue "Cuardaigh" button to its right. Below the search field are two rows of radio button options: the first row has "Béarla ► Gaeilge" (selected) and "Gaeilge ► Béarla"; the second row has "An frása seo gan athrú" (selected) and "Cuardach leathan".



Translation memories



WORDFAST

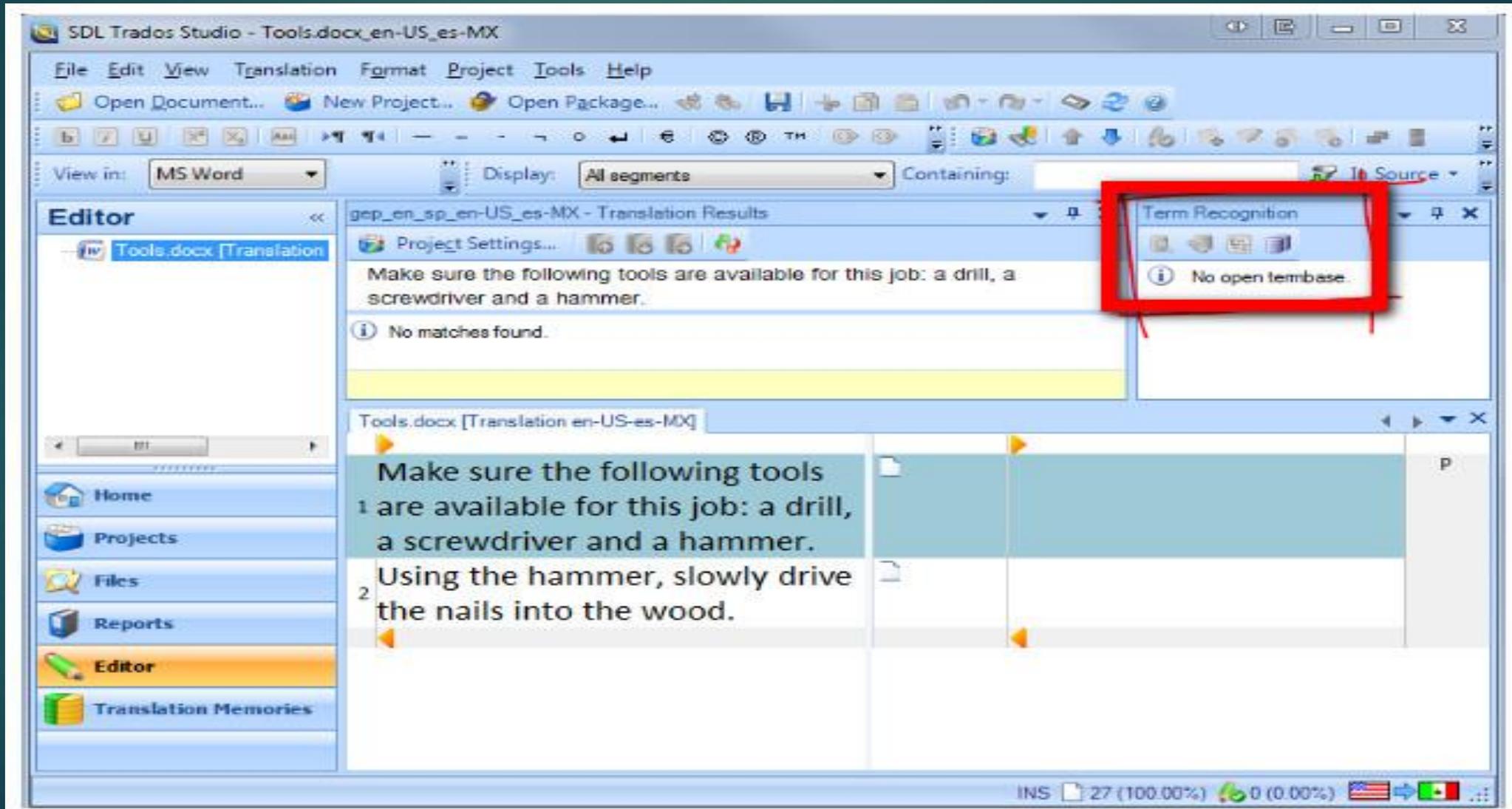




Importance of consistency in terminology

- 
- Adherence to house style, especially when it comes to legislation (e.g. beartas/polasaí; roinn/alt)
 - If the translator has access to a termbase, the editor will have much less work to do
 - Termbase in Trados

Termbase



Translation memories

Basic features

- **Leverage** of previous translations
- **Analysis** for quoting, planning and keeping track of progress
- **Concordance** for sub-segment searches
- **Maintenance** to perform global changes, import/export content, etc.
- **Quality Assessment**: terminology control, typos, etc.

Editor

PR Peace Proms - Gaeilg

Grúthaistriúcháin an Aonaid, Iconic Plugin - Translation Results

Project Settings

Speaking at the announcement of the Peace Proms, Minister Heather Humphreys said:

1	Speaking at the launch announcement of the Peace Proms, Minister Heather Humphreys said:	71%	Agus í ag labhairt ag seoladh an leabhair inniu, dúirt an tAire Mhic Unfraidh:
2	Speaking at the announcement of the Peace Proms, Minister Heather Humphreys said:	AT	Ag labhairt di nuair a fógraíodh an peace Proms, dúirt an tAire Heather Mhic Unfraidh:

Grúthaistriúcháin an Aonaid 12/8/2014 5:38:56 PM TSR\virabhataigh

Grúthaistriúcháin an Aonaid, Iconic Plugin - Transl... Grúthaistriúcháin an Aonaid, Iconic Plugin - Conco... Comments TQAs (0) Messages

Term Recognition

speaking
speaking
announcement
announcement
peace
peace
prom
prom
PROM
programmable read-only memory
PROM
Minister Humphreys

Term Recognition Termbase Search

- Welcome
- Projects
- Files
- Reports
- Editor
- Translation Memories

PR Peace Proms - Gaeilge.docx.sdbdiff [Translation]

7	The young performers, who range in age from 10 to 19, have been drawn from across the island of Ireland, and are all exceptional and award-winning artists and performers.	tagairt na taibheoirí óga, aca idir 10 agus 19 bliana d'aois, as gach cearn d'oileán na hÉireann, agus is ealaíontóirí agus taibheoirí den scoth, a bhfuil gradaim buaite acu, gach duine díobh.	
8	Speaking at the announcement of the Peace Proms, Minister Heather Humphreys said:		P
9	"This promises to be an extremely exciting and uplifting event; the ideal start to our year of commemorations in 2016.		P
10	The Peace Proms will be a wonderful occasion, bringing together talented and committed young musicians from across the island of Ireland.		
11	I would like to thank all of the young performers who will be taking part; through their music and song they will set		

Translation Details:
Status: Draft
Origin: Interactive
Score: 0%

How it works – leveraging translation memories

- Retains previously translated texts
- Recognises similar sentences, including those which are only partially translated, and provides the translation
- Segment-level fuzzy match
- Highlighting of differences
- TM match information (who, when, ...?)

Fuzzy match

- Context match
- 100%
- 95-99%
- 85-94%
- 75-84%
- New words
- Repetitions

Fuzzy match

- Percentage/threshold can be selected and set
- If the match is above the selected threshold, the sentence is edited. If it is below that threshold, the sentence is translated from scratch.

Management of translation memories

- It is important that the translation memory is managed so that all translators working on a particular project are using the same terminology
- One translation memory file can be shared by multiple translators
- Translation memory files can be merged

Machine translation

Machine translation / Tapadóir

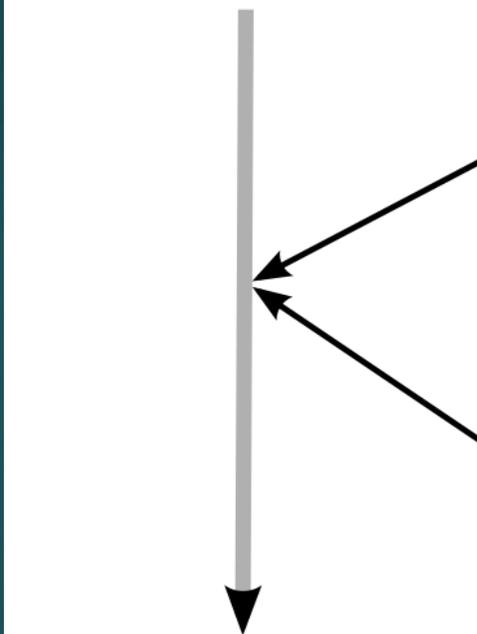
- It is standard international practice to have a tailored machine translation system which caters to specific user needs
- Tapadóir is an example of that type of tailored system, trained for the particular type of texts used in the Department (doesn't work as well for exceptional projects, e.g. Ireland 2016)
- MTPE can improve productivity even when the fastest translators work with content which allow for higher than usual translation throughputs

Machine Translation / Tapadóir

- Not to be confused with Google Translate
- Tapadóir is on a secure server and can only be accessed by authorised translators/editors

How statistical machine translation works

似乎格式有問題



English output

translation model

language model

parallel corpus

网站资讯分析网数
据显示的主域名为
全世界访问量最高
的站点除此之外搜
索在其他国家或地
区域名下的多个站
点等等及旗下的等

The corporation has been estim
to run more than one million pag
in data centers around the world
to process over one billion search
requests and about twenty-four i
of user-generated data each dat
December 2012 Alexa listed as

monolingual corpus

started functioning in 1928 and established the tradition of
large exhibitions and trade fairs held in Brno, and nowadays
also ranks among the sights of the city. Brno is also
known for hosting big motorbike and other races on the
Masaryk Circuit, a tradition established in 1930 in which
the Road Racing World Championship Grand Prix is
one of the most prestigious races. Another notable cultural
tradition is an international fireworks competition.

Tapadóir

Project Settings Configuration | Clipboard | File Actions | Formatting | QuickInsert | Translation Memory | Terminology | Segment Actions | Navigation

Editor | PR Peace Proms - Gaeilg

Gnáthairíocháin an Aonaid - Translation Results

Project Settings | Speaking at the announcement of the Peace Proms. Minister Heather Humphreys said:

1 Speaking at the announcement of the Peace Proms, Minister Heather Humphreys said: 1% Agus í ag labhairt ag seoladh an leabhair inniu, dúirt an tAire Mhic Unfraidh:

2 Speaking at the announcement of the Peace Proms, Minister Heather Humphreys said: AT Ag labhairt di nuair a fógraíodh an peace Proms, dúirt an tAire Heather Mhic Unfraidh:

Gnáthairíocháin an Aonaid 12/8/2014 5:38:56 PM - TSR/vrabbhataighe

PR Peace Proms - Gaeilge.docx.sdxdiff [Translation]

7 The young performers, who range in age from 10 to 19, have been drawn from across the island of Ireland, and are all exceptional and award-winning artists and performers.

8 Speaking at the announcement of the Peace Proms, Minister Heather Humphreys said:

9 "This promises to be an extremely exciting and uplifting event; the ideal start to our year of commemorations in 2016.

10 The Peace Proms will be a wonderful occasion, bringing together talented and committed young musicians from across the island of Ireland.

11 I would like to thank all of the young performers who will be taking part; through their music and song they will set the perfect tone as kick off what is set to be a very special year."

Translation Details: Status: Draft Origin: Interactive Score: 0%

All segments INS 53.29% 32.06% 14.65% Chars: 0 0/2630

Tapadóir

Translations provided by Tapadóir are used

- to edit a fuzzy match

or

- as an alternative to the translation provided by the translation memory

Tapadóir

- All of the work is done in one window
- No need to copy and paste from one window to another
- = saves time (thereby reducing costs)



Uptake of the system / productivity

Previous practice in the Department of Culture, Heritage and the Gaeltacht

- All content fully **human**-translated
- Computer Assisted Translation tool: **SDL Trados**
- Many **repetitive** translation tasks
- **No Machine Translation**

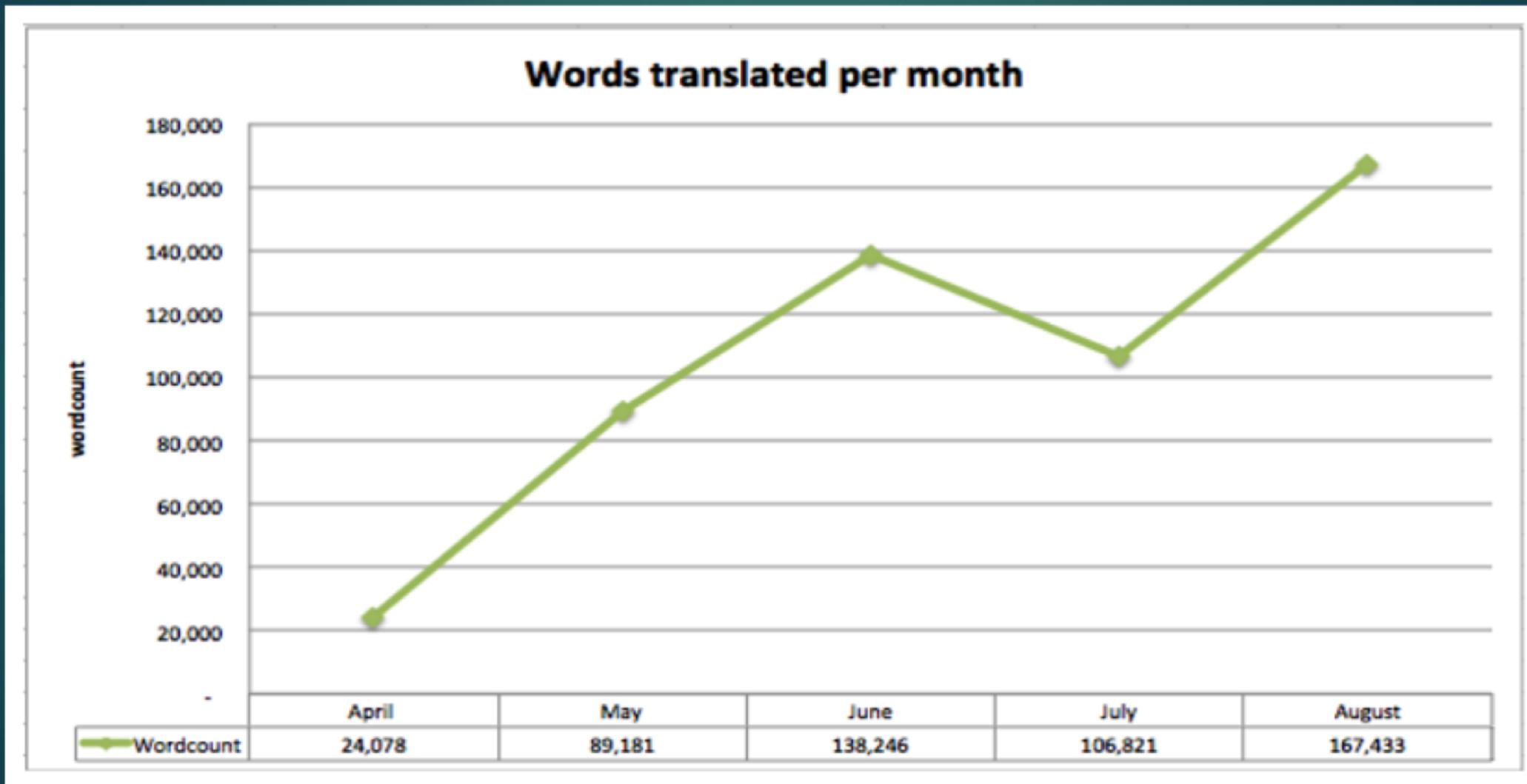
Early stages

- ▶ **Initial meetings** between ADAPT and senior staff at DCHG
- ▶ DCHG **recognised potential** for investing in language technology, specifically Machine Translation
- ▶ Potential **cost-savings** apparent
- ▶ Opportunity for ADAPT to carrying out **research** into Irish **SMT**

Use of Tapadóir in the Department

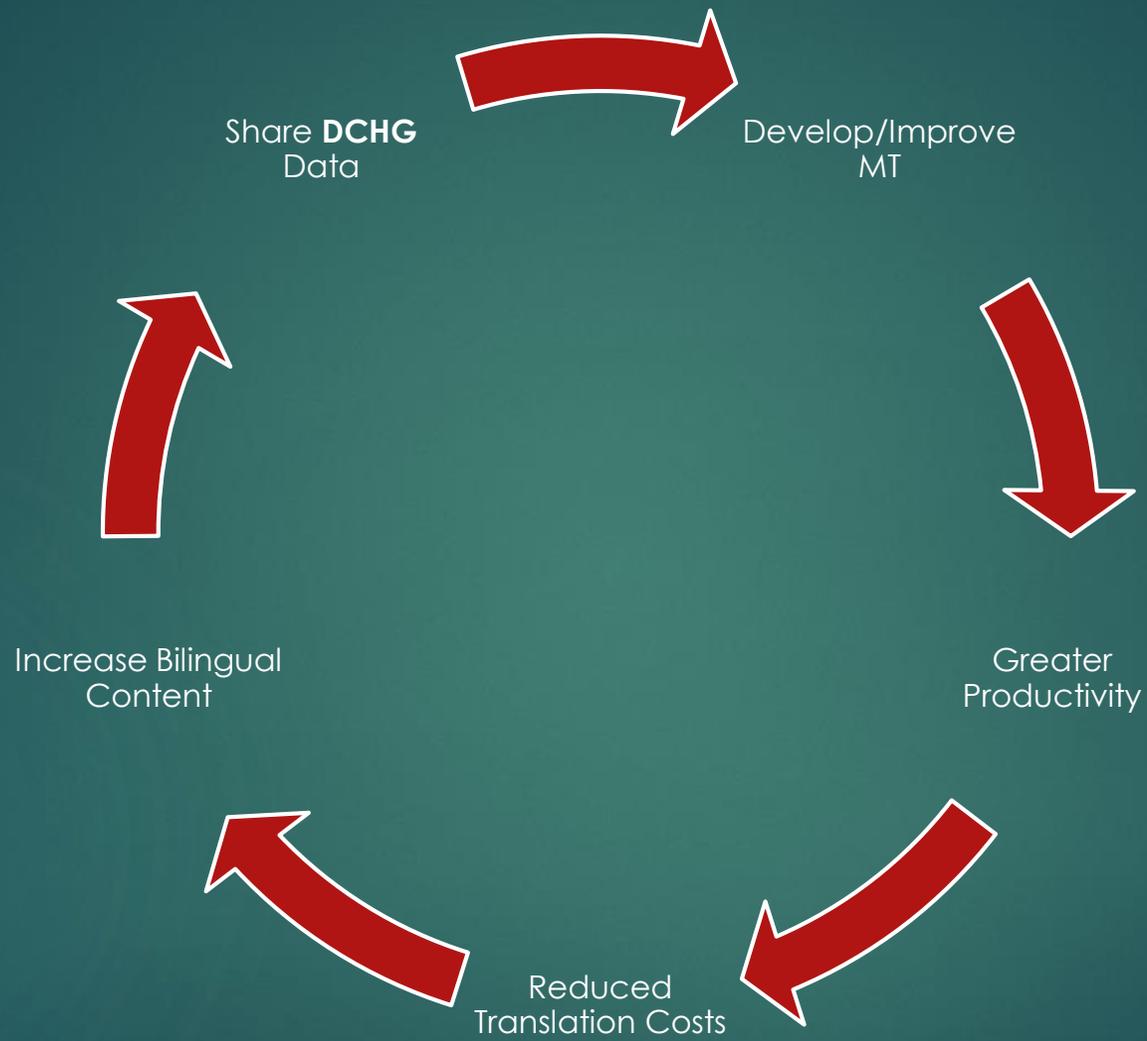
- Dept translators
- Monitoring of the translation rate
- Slow at first; results to be seen after a while
- The time it takes to get used to a new method of working

Uptake of Tapadóir



Post-editing and tweaking

- The Department's translators could communicate issues to the team in DCU, who would then tweak the system in order to improve it
- Studies on post-editing have shown that MTPE can improve productivity even when the fastest translators work with content which allow for higher than usual translation throughputs
- Studies needed on post-editing in Irish



Take-aways

- ▶ Tapadóir is a perfect example of how data sharing benefits users
- ▶ Share your data! 😊



Go raibh maith agaibh!
Thank you!