

# MACHINE TRANSLATION

---

a guide to onboarding

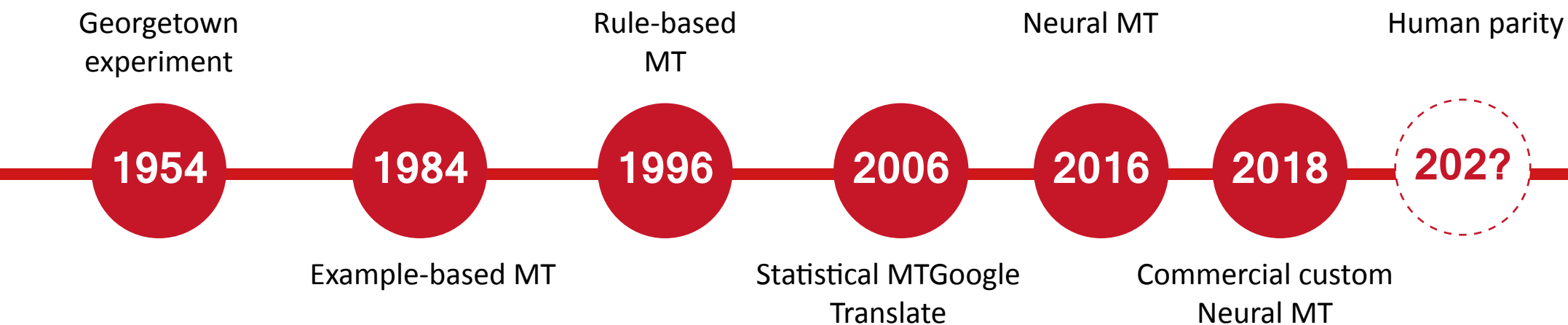
# Machine translation is finally good enough

---

It can efficiently translate chat, social media updates, support exchanges, email in foreign languages. The latest generation of MT systems is based on neural networks. With training in specific terminology, such as medical or legal, they can produce surprisingly good results.

Thanks to improvements in quality, and to ease of use MT engines are widely used. They churn about a trillion words a day, more than all human translators on the planet can do in a year!

# MT timeline



# Uses for machine translation

## 01

### Boost human translators

Professional translators can work faster when aided by good MT. Human post-edit raw machine output in a professional translation tool. This results in 5% to 54% faster turn-around and lower cost, depending of the quality of the engine and the type of the text.

Beware of bad quality MT in this scenario, which is detrimental to human effort.

## 02

### Translate massive volumes

In this scenario, there is no human editing, readers get to see raw machine output, regardless of quality.

Consider user reviews for products and hotels, email exchanges and chat, huge support knowledgebases, shop e-commerce catalogues.

A typical scenario involves millions of words of content, something unfeasible for human translation.

## 03

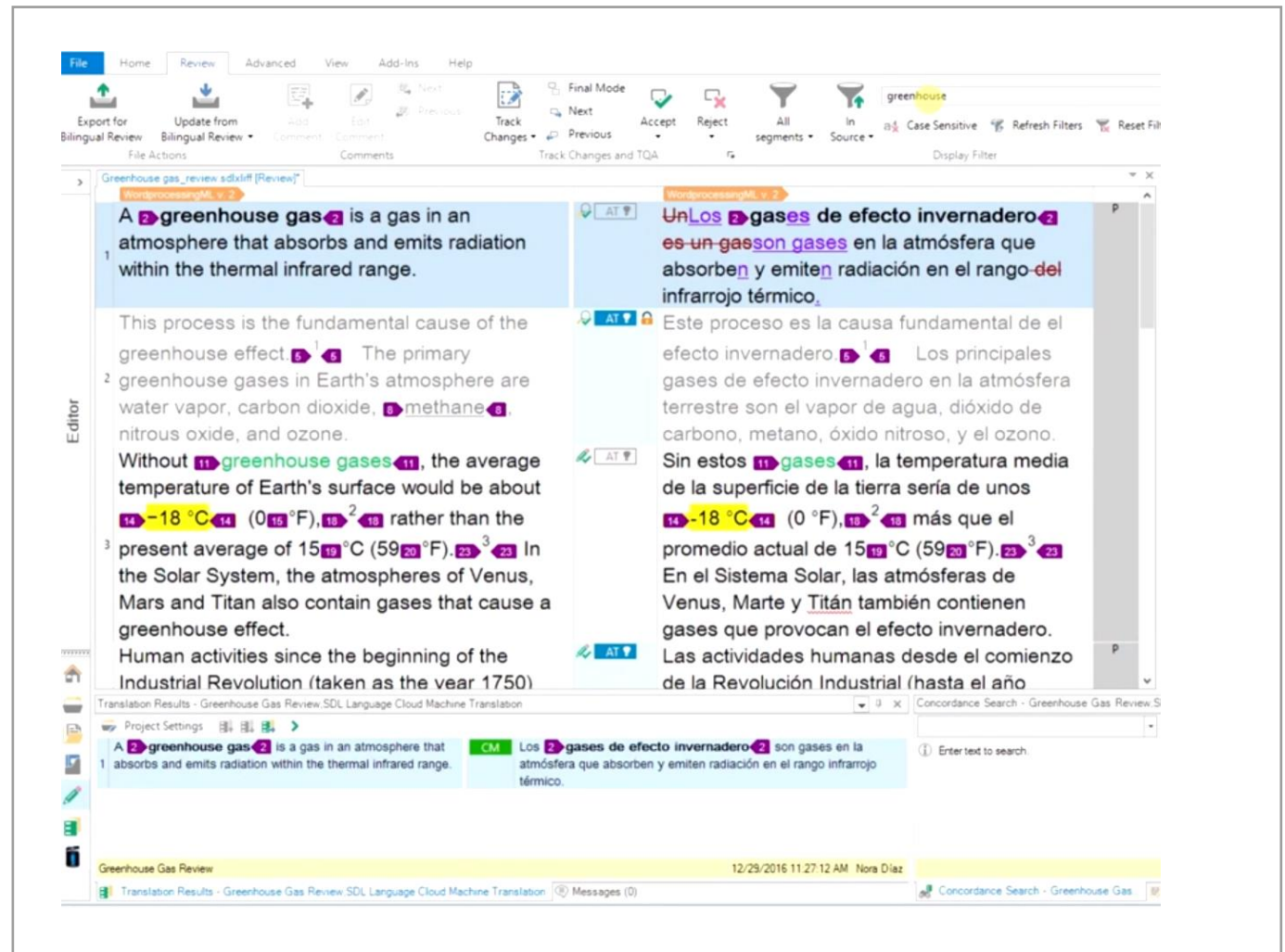
### Multilingual search for data

Data comes in many languages. In conjunction with keyword search and entity recognition technology, MT helps to find relevant information from non-English sources.

New laws, patents, biomedical research papers, news and social media updates surge across Europe, Asia and the Americas. From sentiment analysis to legal discovery and litigation, machine translation enables search to get to the right data.

# Example

Post-editing machine translation output in SDL Trados



# Example

TripAdvisor integrated  
Google Translate for hotel  
reviews



## Great Hospitality, Nice Location

[Google Traduction](#)

«We stayed for 4 nights in December 2018 at the Corpo Santo Historical Hotel in Lisbon. The location is perfect and the services are sophisticated. We were enjoying complimentary sweets, drinks, snacks which offered everyday at the lobby. We also joined the free walking tour. We t...»

[Plus](#) ▼

**Date du séjour :** décembre 2018

[Utile](#)

[Republier](#)

[Partager](#)



Réponse de Nuno\_Goncalo, Guest Service de Corpo Santo Lisbon

A répondu hier

[Google Traduction](#)

Dear okanaha We are delighted that you liked Corpo Santo Lisbon Historical Hotel and consider that this is the place to be during your time here in Lisbon. We agree that the location is just perfect for tranquility and practicality. Thank you for such a lovely comment on the staff and the personal attention you felt during

[Plus](#) ▼

# Example

Microsoft has machine-translated their help documentation for Skype for Business with a link to the English original

## Веб-планировщик Skype для бизнеса

*Lync 2013 для Office 365, Office для бизнеса, Дополнительно...*

**Примечание:** Мы стараемся как можно оперативнее обеспечивать вас актуальными справочными материалами на вашем языке. Эта страница переведена автоматически, поэтому ее текст может содержать неточности и грамматические ошибки. Для нас важно, чтобы эта статья была вам полезна. Просим вас уделить пару секунд и сообщить, помогла ли она вам, с помощью кнопок внизу страницы. Для удобства также приводим [ссылку на оригинал \(на английском языке\)](#).

Skype для бизнеса Веб-планировщик представляет собой веб-приложение, можно использовать для создания Skype собраний, если у вас нет Outlook, или еще не установили браузер с поддержкой Outlook Web App. Skype для бизнеса веб-планировщик можно создавать новые собрания, изменение существующих собраний и отправка приглашений с предпочитаемой почтовой программе.

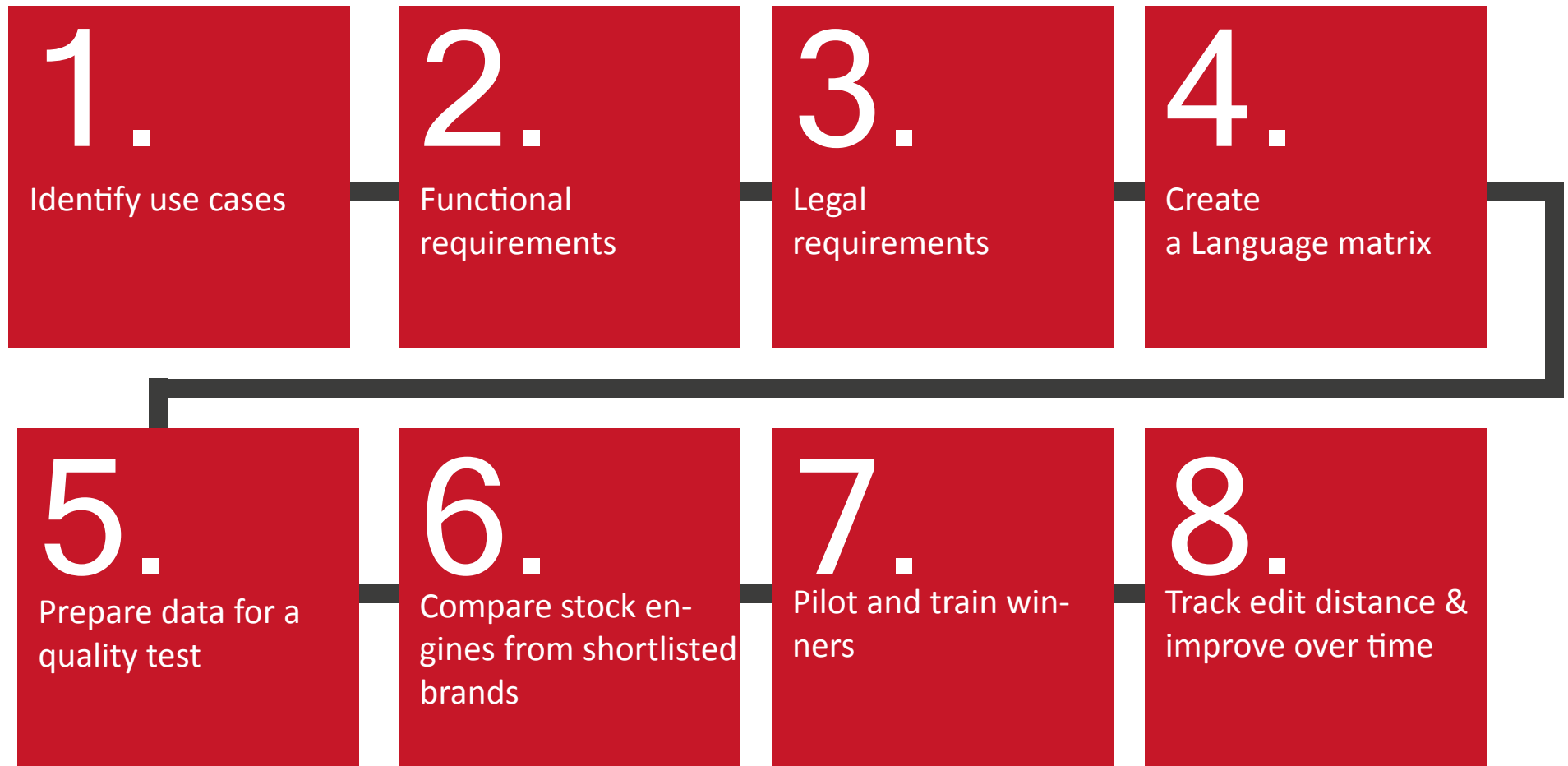
Веб-адрес для веб-планировщика Skype для бизнеса зависит от вашей организации. Если вы не знаете адрес веб-планировщика или не уверены, выполните приведенные ниже действия.

1. Откройте веб-планировщика Office 365 по адресу <https://sched.lync.com> и попробуйте выполнить вход с помощью рабочей или учебной учетной записи.

# Choosing MT providers



# MT Request for Proposal Stages



# 50+ brands available

## Cloud API from tech giants



Amazon Translate



Baidu



Google



IBM Watson Translator



Microsoft



Yandex

## Cloud API - Industry players



DeepL



GlobalLink AI



KantanMT.com

Kantan MT



Lilt



GTCOM

GTCOM



株式会社ロゼッタ

Rozetta Jukkou

## Dedicated vendors with OnPrem



Omniscien MT

Tilde MT



Modern MT



Prompt



Iconic Translation Engines

Systran



Globalese



SDL

## Other

Slate Desktop



Lingua Custodia Verto

Textra MT



Lucy LT

Reverso by Softissimo

Sunda MT

# Factors influencing choice

Vendor Type	Example
Tech giants:	Amazon, Google, Microsoft, IBM
Dedicated MT vendors:	Omniscien, Globalese, Kantant, Systran
LSPs:	Lingosoft, Pangeanic, Moravia, Capita, GTCOM, Modern MT, etc.

Strategy
Single-engine
Single-engine with training
Multi-engine untrained
Multi-engine with training
Multi-engine via middleware

Type	Difficulty to train
Stock/Generic	-
Domain Models	-
Adaptive	Easy
Domain Adaptive (new)	Medium
Custom On Prem	Hard
NMT toolkit	NMT toolkit

Deployment	
Cloud	Per 1 mil characters..
Hosted/«Private Cloud»	License + usage
On-premise/Server	License, hardware, electric bill

MT quality is very different from language to language and from subject matter area to another. That's why it makes sense to use multiple brands of MT at the same time, and periodically look at new offerings on the market. **This is called a multi-engine strategy.**

**With engine training,** quality can be further improve, but this process takes time, expertise and investments. The results are not guaranteed. The latest type of MT, Domain-Adaptive, makes training easier and can produce results with a relatively small sample of just 10,000 segments of translation memory. It will still take a specialist to train and maintain engines.

**Cloud API MT** technologies are easy to use and are relatively inexpensive with prices from \$100 to \$4 per million characters. On Prem servers are more expensive at launch, with prices in tens of thousands of USD, but they do not have usage limits and keep the data on the customer's computers. They come into play when confidentiality is important, and when the engine is expected to translate millions of characters every day.

**MT vendor type** is important for support: small customers are not likely to get a personalized manager with a tech giant. They can, however, become very important for smaller MT providers.

# Typical Requirements

## Hard requirements

1. Language combination support
2. Security
  1. Deployment model
  2. Server locations
  3. No trace availability (yes, no, opt-in)
  4. SSL encryption strength
3. Total cost of ownership
4. Volume requirements
5. Robust API/Infrastructure
6. File format support
7. Anonymization/redactions for sensitive data

## Good to have

8. Terminology & glossary support (i.e. IBM, Amazon, MS support glossaries)
9. Formatting fidelity
- 10 Connectors to user systems (Outlook, Skype, CMS, PIM)
11. User portal
12. Management dashboard with reporting on quality
13. Support 24/7
14. Can I train via API (re-upload content right back in)
15. Can I train multiple engines and select the best inside
16. Can I upload an edited doc into the system to get TER?
17. Open source?
18. What does it do for source text

## Legal

19. Does the provider collect and store my data? What do they do with it?
20. Can I upload my data and my customer data to their servers and comply with GDPR?
21. Are they a potential competitor to my company in any areas?
22. What happens if the MT provider exposes my data?
23. What happens to my business if the MT provider's system goes down?
- 24 Can the MT provider be acquired by my competitor? If yes, what are the implications?

# A Template Language/Project Matrix

From> To v	English	French	German	Chinese
English	X	Marketing	Marketing	Legal
French	Medical Legal Marketing	X		Marketing
German	Medical Legal Marketing	Marketing	X	Marketing
Chinese	Medical Legal Marketing	Legal Marketing	Marketing	X

---

Design a similar matrix, mapping subject matter areas for each language combination needed. Prepare a set of test data for each cell in the matrix. Then run a test on 3-10 engines to evaluate which ones perform the best with your data.

# A Template Language/Project Matrix

## Automated metrics

BLEU

TER

LEPOR..

Memsorce MT QE (as you go)

First evaluation round

If one engine is clearly superior for your data, it's possible to stay single-engine. Otherwise, assume a multi-engine model

Train engines for content that scores low with all engines.

Dedicated MT companies and LSPs provide training services

New wave of domain-adaptive MT makes it easy to train (upload a tmx, launch, test)

## Human evaluation

LSPs

OHT, Logrus IT, Welocalize..

Second evaluation round

---

Needs to use multiple evaluators, blind, needs a statistical model, needs a tool.

# Post-pilot: Useful KPIs

Metric	Details	What is tracked
Edit distance per engine	Keystrokes or minutes needed to improve 1000 words of text to human-level	Cost of editing
User behavior	1) Surveys built into the product: «was this page useful?» 2) Translated page traffic and view time compared to the original	User experience
Percentage of 100% matches	Ratio of phrases and translation segments not changed at all by the human editor. Is it closer to 5% or 50%	MT quality and leverage



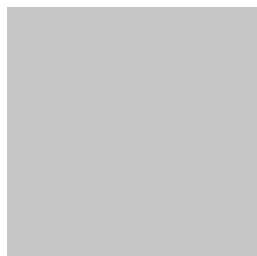
# Use MT specialists at Effectiff

Evaluate and choose MT for your content

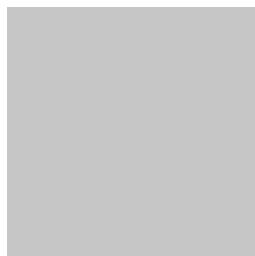
Post-edit the output

Find new use cases

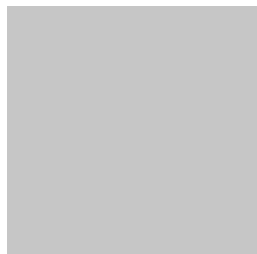
Train MT for terminology



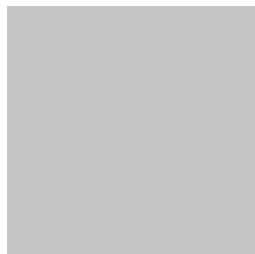
Simon Cook  
Draper Esprit



Simon Cook  
Draper Esprit



Simon Cook  
Draper Esprit



Simon Cook  
Draper Esprit

## Effectiff Translation Company

15 years of outstanding service  
in the language industry

Our network of 1,000 interpreters  
and translators in 50 languages

500 core clients worldwide

Long-term collaboration with small,  
mid-sized and major companies

Fully automated workflow: The usage  
of TMs (translation memories) allows  
us to achieve best results in less time.

SmartCAT Translation Automation Platform

Glossary Creation: TB (terminology base)  
creation allows to reduce costs for  
editing by 30%