

# How to Unlock NMT?



1,271 views | Nov 8, 2019, 08:15am

# The Death Of Translation Memory



**Gabriel Fairman** Forbes Councils Member  
**Forbes Technology Council** COUNCIL POST | Paid Program

Innovation

Tools that are already available can offer translators better feeds than translation memory would. However, the industry lags. It still doesn't know how to **merge translation memories and ATNMT into a single paradigm**. As an

# Topics



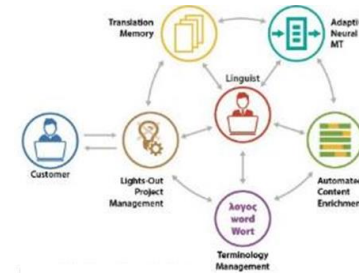
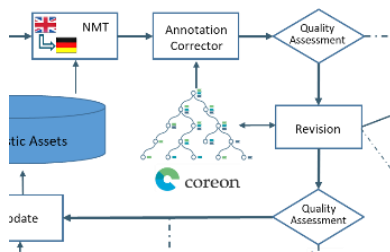
NMT disruptes translation

Industry getting it wrong

NMT Workflow sunsets CAT

Multilingual Knowledge System

New roles for linguists



# Disruption



image credit: mindjet.com

# “Nearly Indistinguishable From Human Translation”—Google Claims Breakthrough

by Florian Faes on September 27, 2016



The Google machine translation coming out of the 2016 entity...

## Intelligent Machines

# Google’s New Service Translates Languages Almost as Well as Humans Can

Chris Wendt @Tian500

Follow

Replying to @Jeromobot @tarfandy @allisonlinn

Thanks. It is not a career path I can reasonably recommend to a young person. Not for that language pair and without a really deep specialization. We can debate the human parity claim endlessly. There is no doubt about the fact that machines are getting better at it every day.

5:55 PM - 15 Mar 2018

Microsoft | The AI Blog | The Official Microsoft Blog | Microsoft On the Issues

## Microsoft Research Achieves Human Parity For Chinese English Translation

Written by Sue Gee

Wednesday, 21 March 2018

Researchers in Microsoft’s labs in Beijing and in Redmond and Washington have developed an AI machine translation system that can translate with the same accuracy as a human from Chinese to English.

jump in efforts to

y Tom Sin



# Microsoft reaches a historic milestone, using AI to match human performance in translating news from Chinese to English

March 14, 2018 | Allison Linn



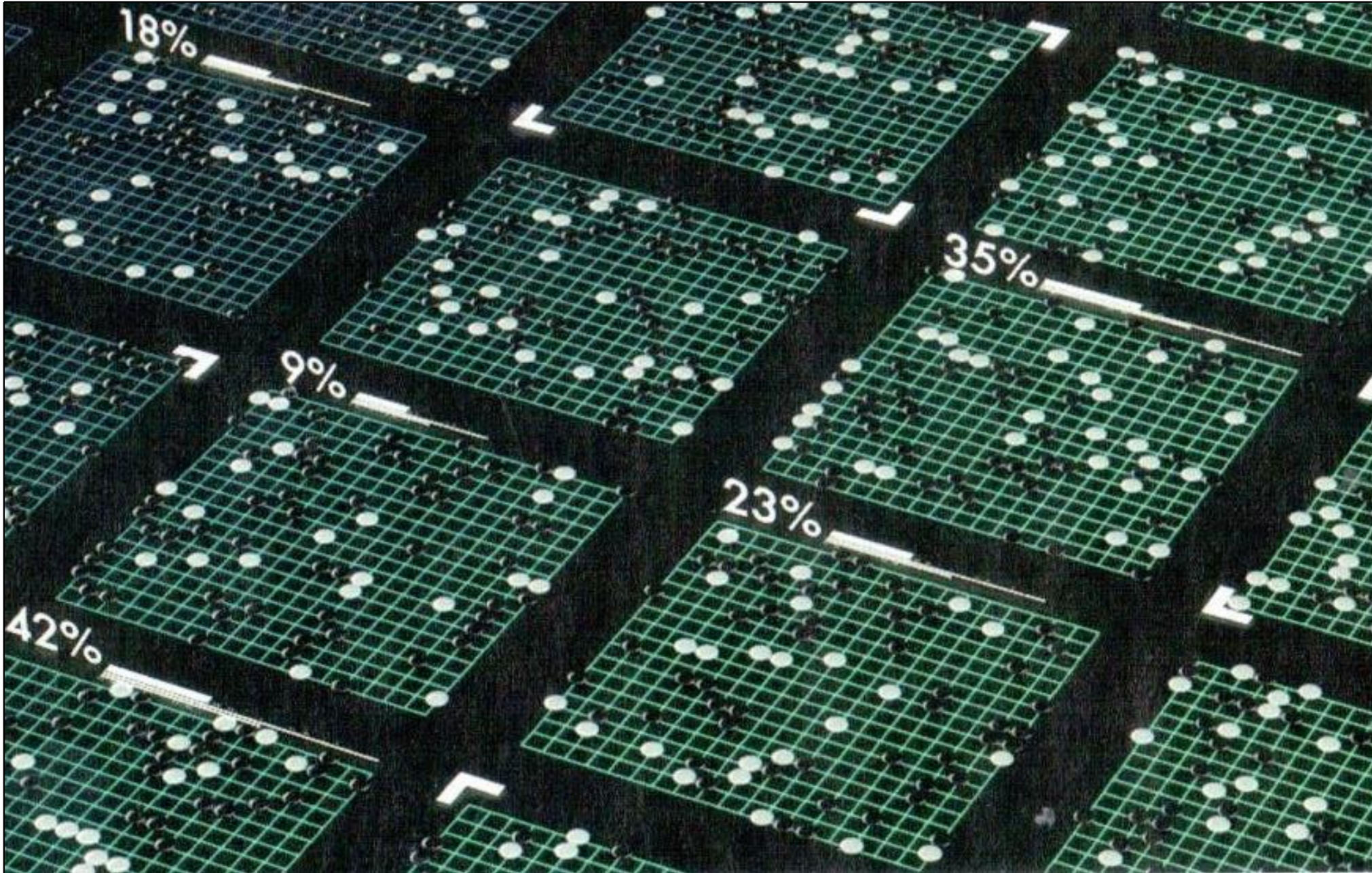
Jagwar [grrr...] @terakatorotany · 29 Aug 2017

I've just realised that the quality of #DeepL machine translations is mind-blowingly good! 🤖 Holy sh\*t, the future is now! #FutureIsNow

# SDL Cracks Russian to English Neural Machine Translation

Global Enterprises to Capitalize on Near Perfect Russian to English Machine Translation as SDL Sets New Industry Standard

# Human Parity

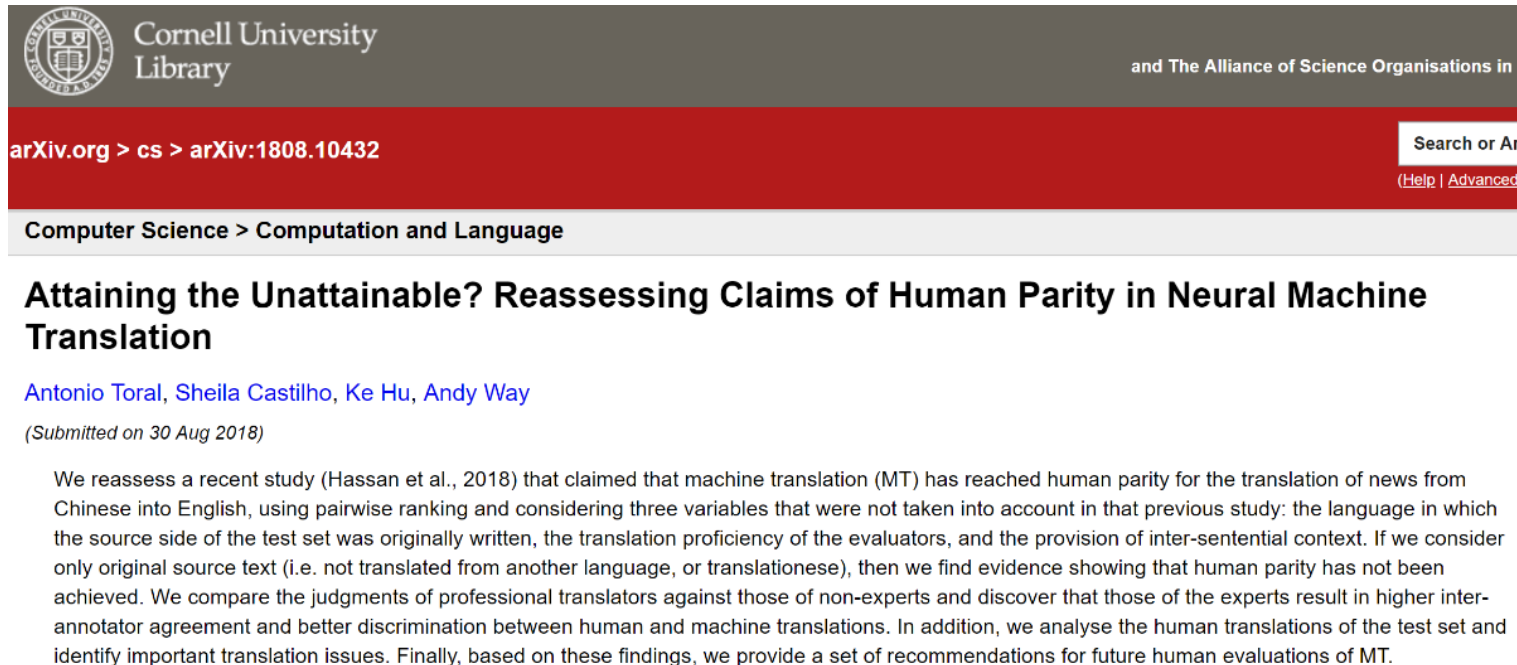


# Intuition

# Why now?



“...evidence showing that human parity has not been achieved.”



The screenshot shows the top portion of an arXiv paper page. At the top left is the Cornell University Library logo. The breadcrumb trail reads 'arXiv.org > cs > arXiv:1808.10432'. The paper title is 'Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation' by Antonio Toral, Sheila Castilho, Ke Hu, and Andy Way, submitted on 30 Aug 2018. The abstract begins with: 'We reassess a recent study (Hassan et al., 2018) that claimed that machine translation (MT) has reached human parity for the translation of news from Chinese into English, using pairwise ranking and considering three variables that were not taken into account in that previous study: the language in which the source side of the test set was originally written, the translation proficiency of the evaluators, and the provision of inter-sentential context. If we consider only original source text (i.e. not translated from another language, or translationese), then we find evidence showing that human parity has not been achieved. We compare the judgments of professional translators against those of non-experts and discover that those of the experts result in higher inter-annotator agreement and better discrimination between human and machine translations. In addition, we analyse the human translations of the test set and identify important translation issues. Finally, based on these findings, we provide a set of recommendations for future human evaluations of MT.'

1. originally written in source language
2. take the whole document into account
3. evaluations be done by professional translators

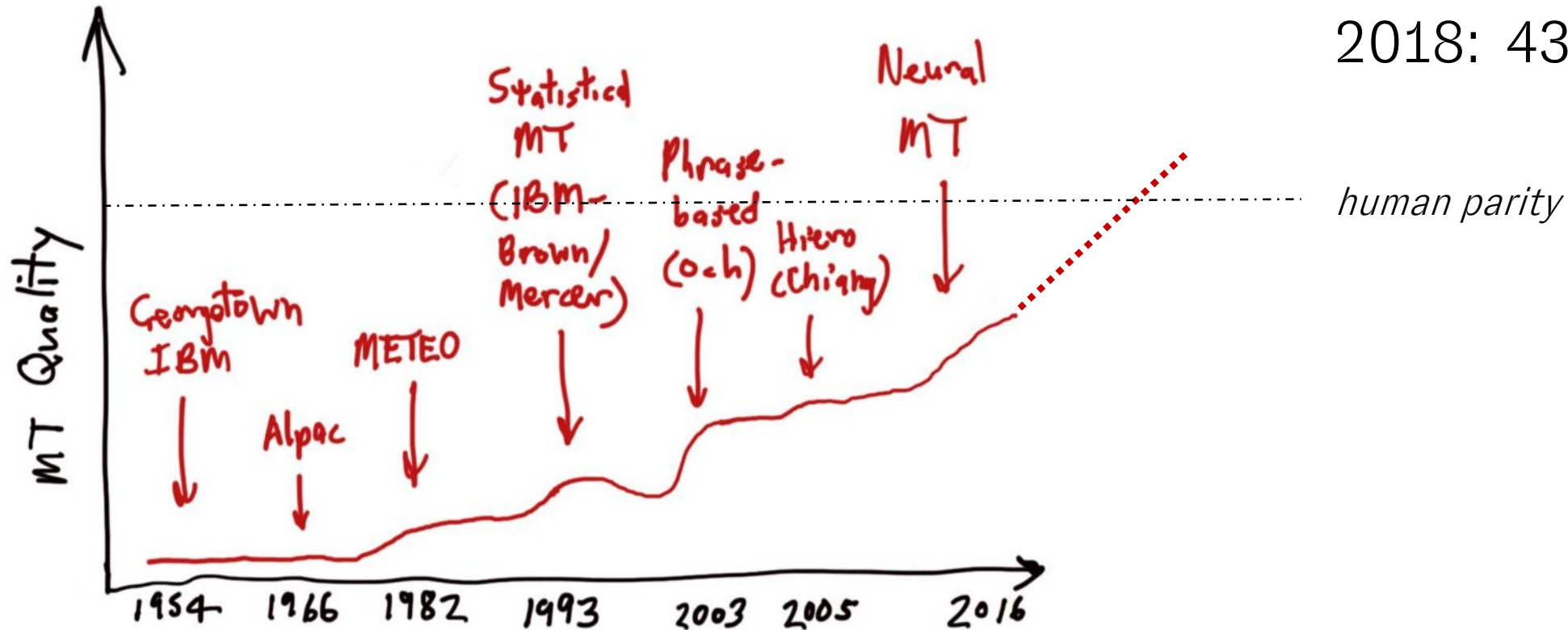
4. time and money?





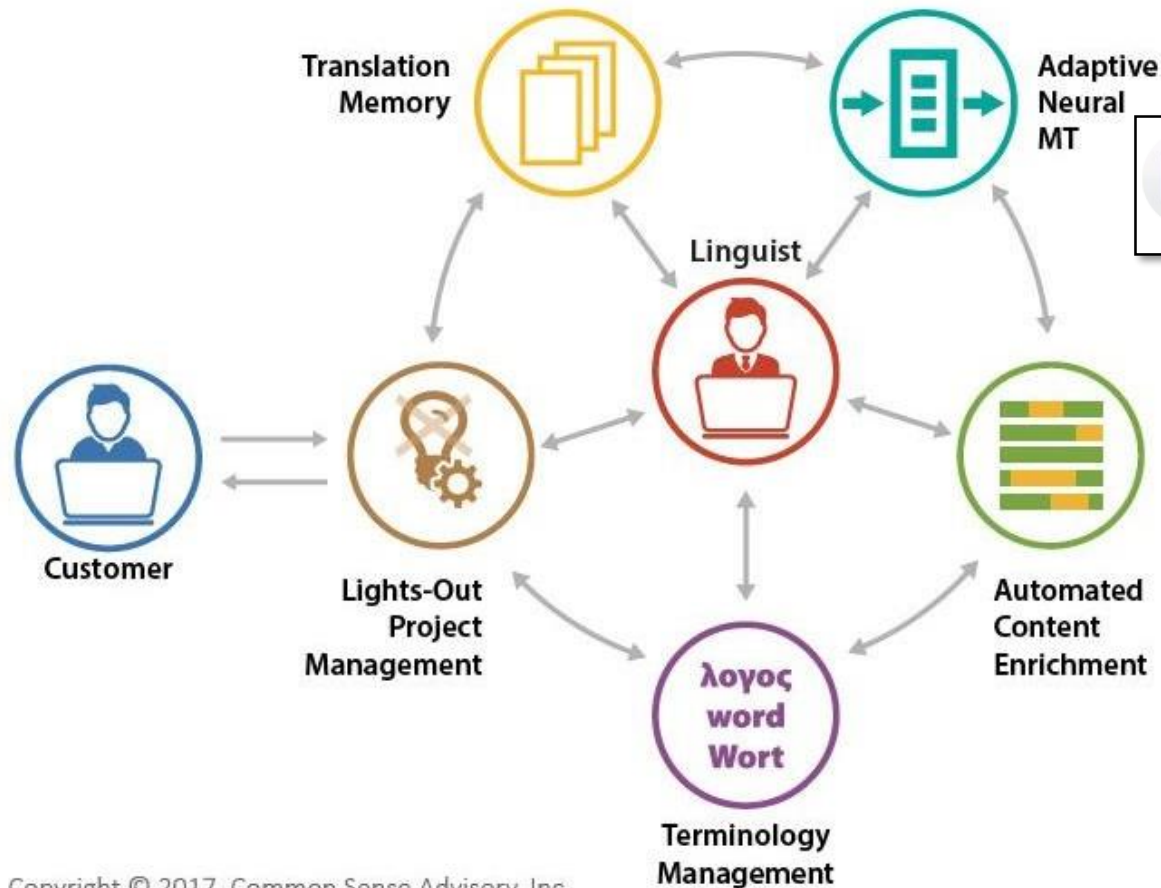
# NMT Crossing the Rubicon

Progress in MT




2017: 199 papers  
2018: 430 papers

# Augmented Translation?



Copyright © 2017, Common Sense Advisory, Inc.

 **Lilt** @LiltHQ · 18 Feb 2017  
Adaptive MT from **Lilt** one of four fundamental technologies for the "augmented translator" #xl8 @CSA\_Research buff.ly/2ITm5j2

**Jeromobot** @Jeromobot  
Replying to @ProfessorOlsen @adrechsel and 3 others  
Yes, I think this is true. The positive aspect of "augmented translation" is that the translator is the agent who is being supplied with data by technology as s/he drives the process. Post-editing of MT would therefore not be augmented translation.

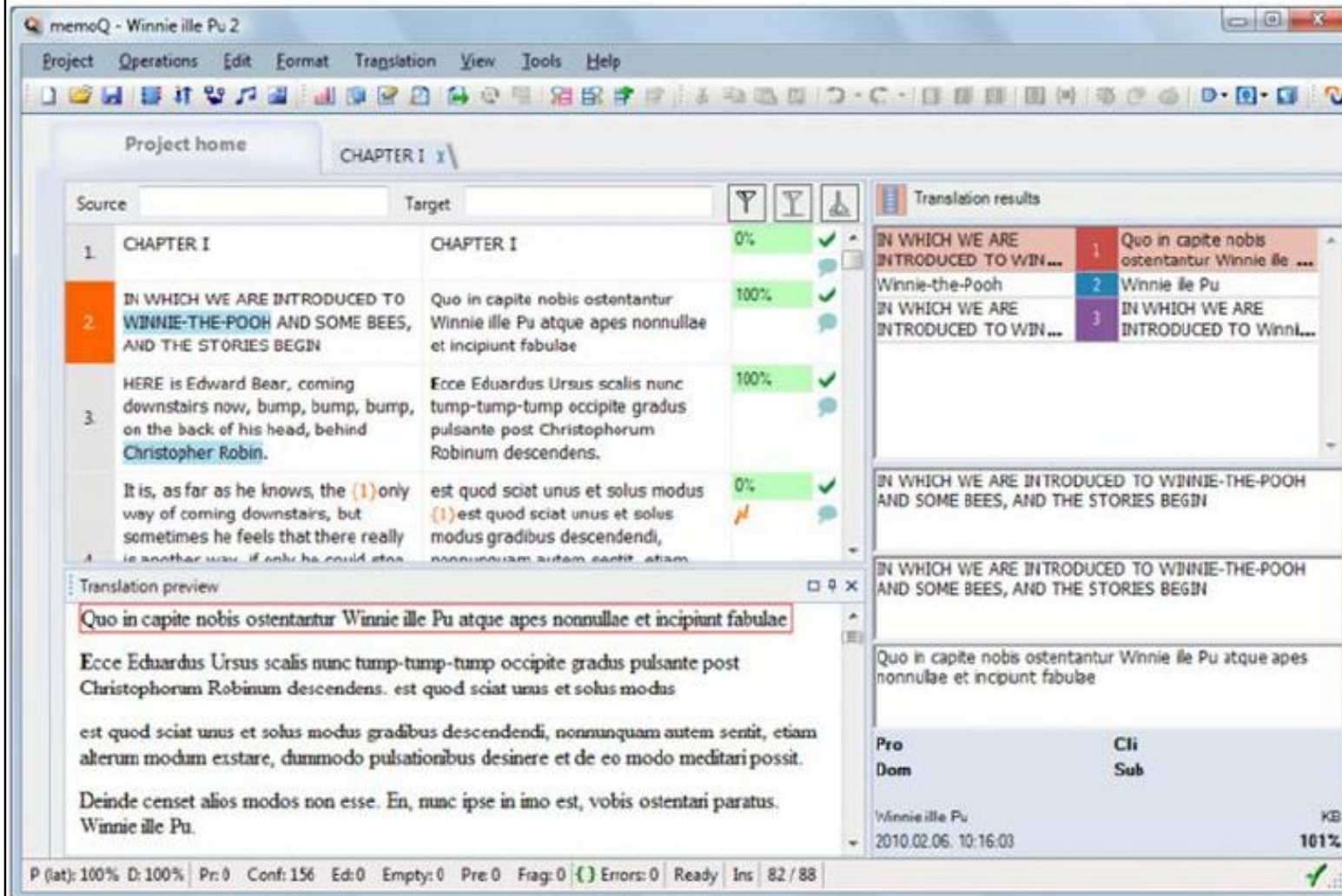
 **Véronique Özkaya** @vozkaya Following  
With NMT, the future is happening faster. Augmented translation at our door.  
[#SlatorConSE18](#)

# Post Editing?



1. read source
2. indentify critical passages
3. read MT translation
4. match to source
5. reconsile terms, memory, head
6. fix
7. check quality
8. repeat

# Fat, Complex Clients



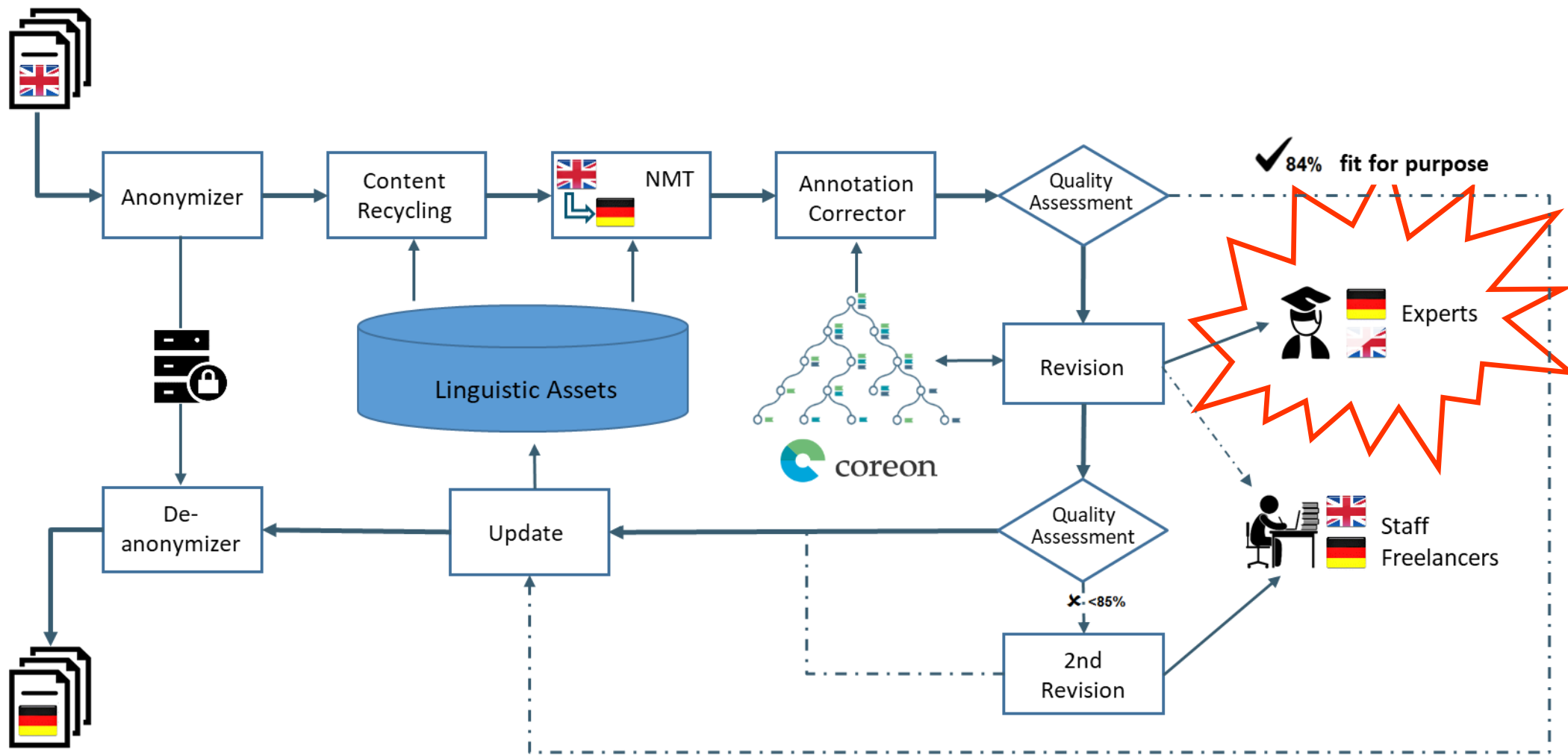
# Meanwhile the MT Community



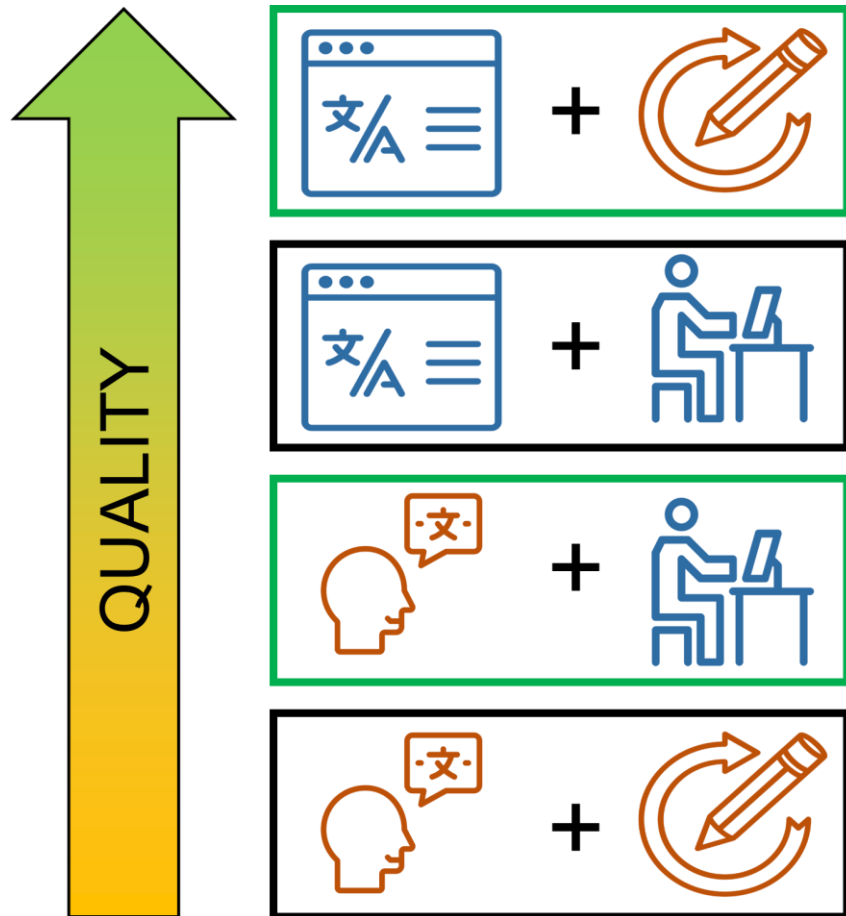
- Everyone is using the same OSS engines
- NMT engines are a black box
- Clean data is everything
- Focus on quality estimation (QE)
- Hard to integrate Translation Memories

**“Non-MT users are the non-CAT users of the early 2000 (a.k.a. dinosaurs).”**

# New Workflow, Actors,



# A Real, Tricky Test



NMT + revision

“What is the optimal translation workflow?”  
Joke Daems, JIAMCATT, 14 May 2019

# Established Players cannot Disrupt





# The Future Gets Funded!

AI

Unbabel raises  
human translation

PAUL SAWERS @PSAWERS



Above: Founders: Vasco Pedro, Hugo Silva, Bruno Prezado, João Graça.

Google Translator Toolkit will not be available after December 4, 2019. L

## Google Translator Toolkit is shut

### Goodbye, Google Translator Toolkit

Google Translator Toolkit launched over a decade ago to help our users, to share translations. When we first launched, there were few web-based options, but now there are [many great tools](#) available, including [Google Translate](#), which will continue to be available and is unaffected by this. As a result, we've seen declining usage for Translator Toolkit over the past few years. So now, after many years and billions of words translated, we're saying goodbye to Translator Toolkit. A warm thank you to our users around the world.

### German LSP lengoo Raises EUR 6.0m Series A from Redalpine, Angel Investors

3 Weeks Ago · M&A And Funding · By Esther Bond On September 10, 2019



Germany-based language service provider (LSP) lengoo has raised EUR 6m (USD 6.6m) in Series A funding. The funding round, completed in August 2019, was led by Zurich-based venture capital firm Redalpine, with other follow-on investors and angels participating.

Lengoo develops a translation workflow that combines neural machine translation (NMT) with human post-

editing, resembling the operational models of tech-forward LSPs Lilt and Unbabel.

### Benchmark investiert in DeepL

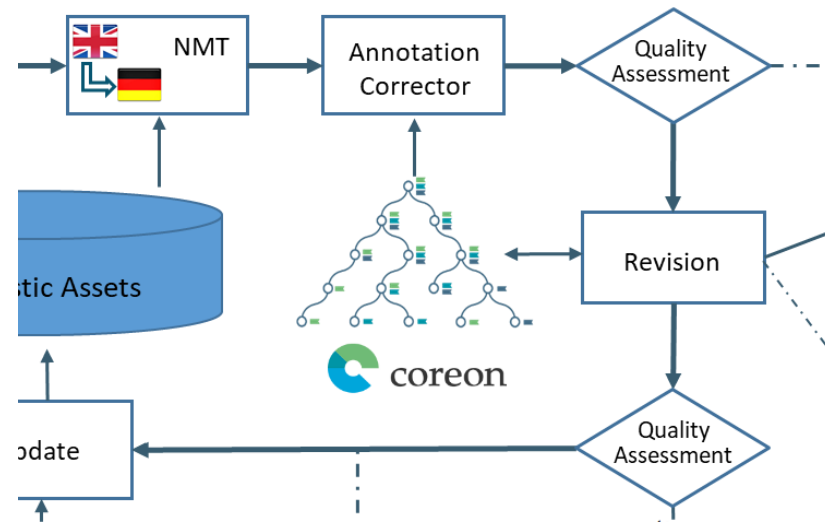
+++ **Benchmark Capital** investiert einen ungenannten Betrag – sicherlich aber einen zweistelligen Millionenbetrag – in den Übersetzungsdienst **DeepL**. Wie die FAZ berichtet, hält der bekannte Geldgeber nun 13,6 % am Kölner Unternehmen. In Sachen maschinelle Übersetzungen kommt weltweit gerade niemand an DeepL vorbei. Seit 2009 kümmert sich der Kölner Hidden Champion um Übersetzungen aller Art – zunächst mit Linguee, einer Internet-

# NMT Workflow Questions



- ¿ Content to recycle?
- ¿ Translations to train the NMT?
- ¿ NMT engine to pick?
- ¿ Terms correct or wrong?
- ¿ Style rules to apply?
- ¿ Quality fit-for-purpose?
- ¿ Best experts?
- ¿ If busy, second best?

# Key Enabler Multilingual Knowledge System



An MKS classifies linguistic assets, human resources, workflow rules, and projects in a unified system which is expandable, dynamic, and provides fallback paths.

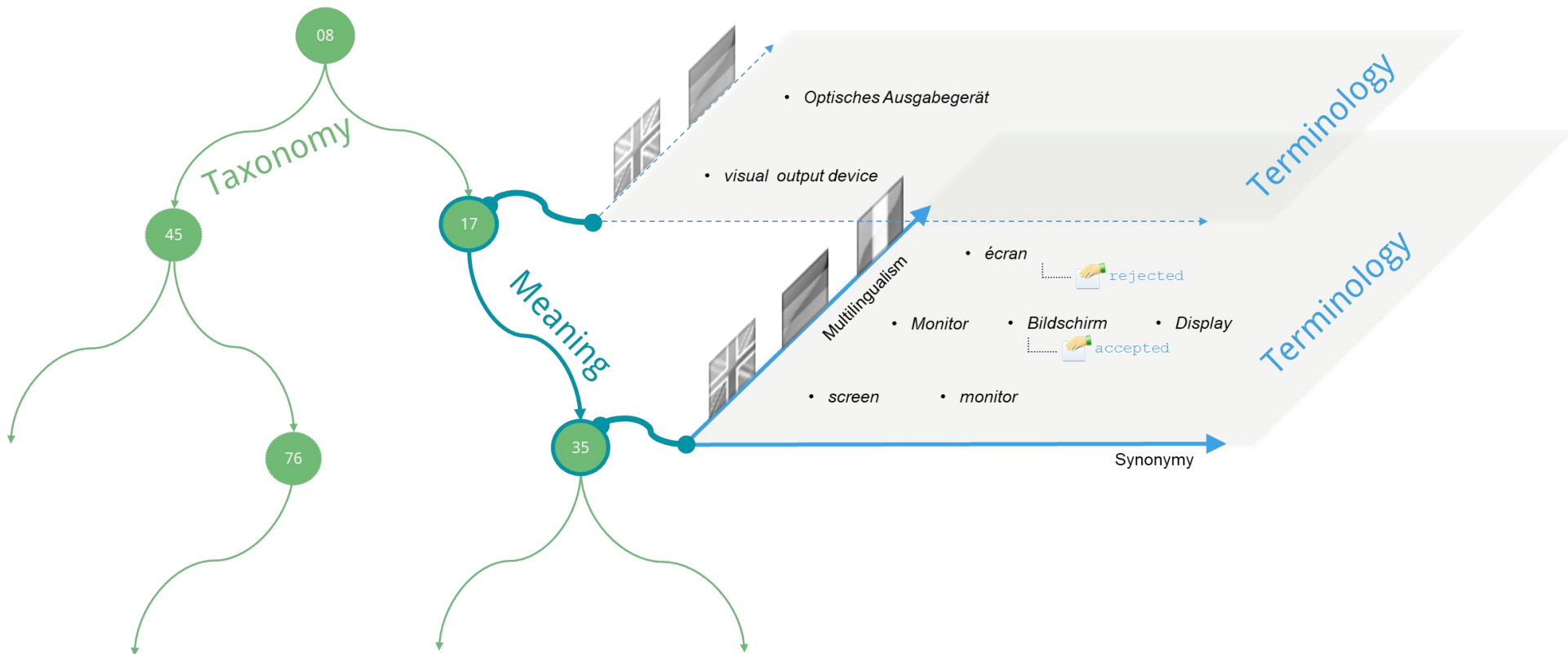
# Multilingual Knowledge System

1: Taxonomy: ↘

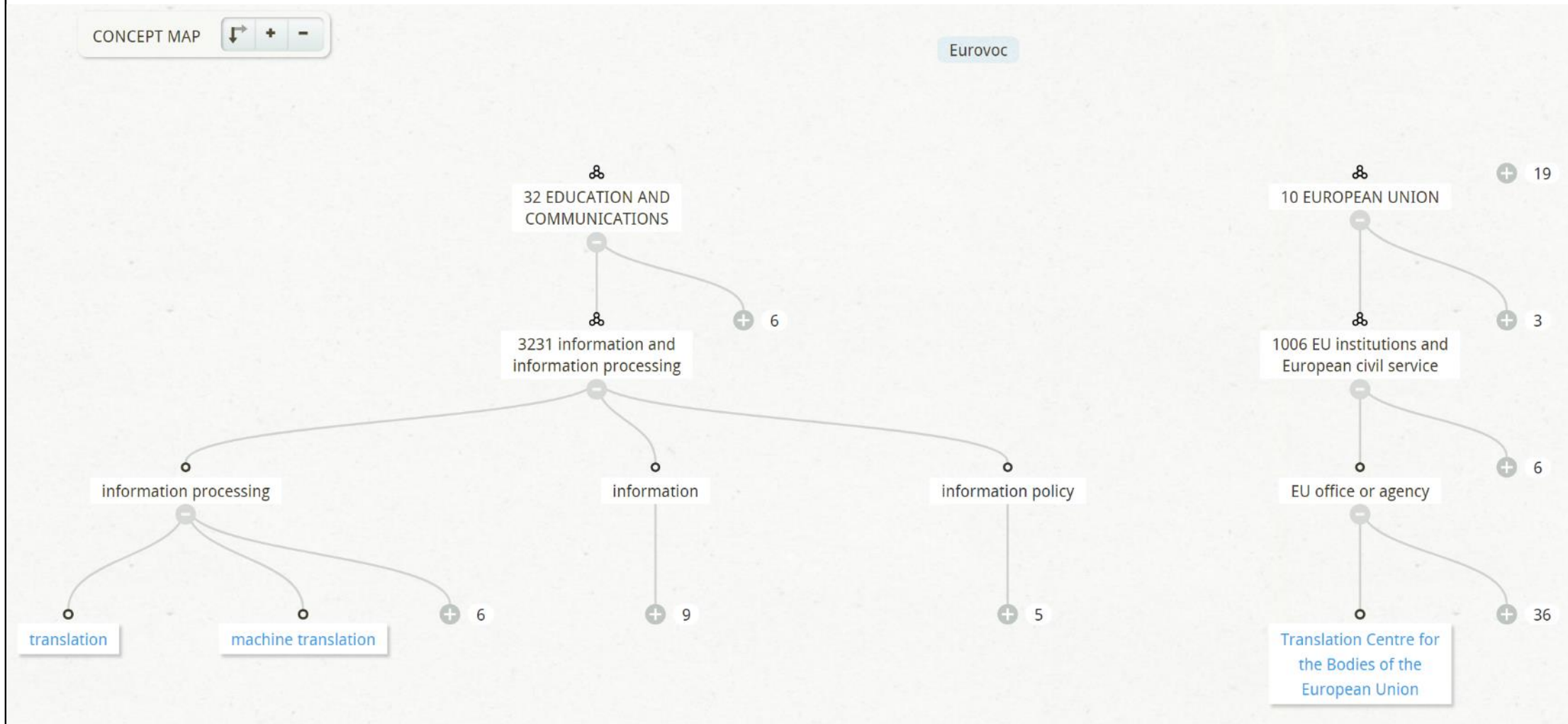
2: Synonymy: →

3: Multilingualism: ↗

4: Control: ↵



# Multilingual Knowledge System



# Leverage Existing Assets

Multilingual Terminology List



Knowledge Graphs



*Structure  
semantically*

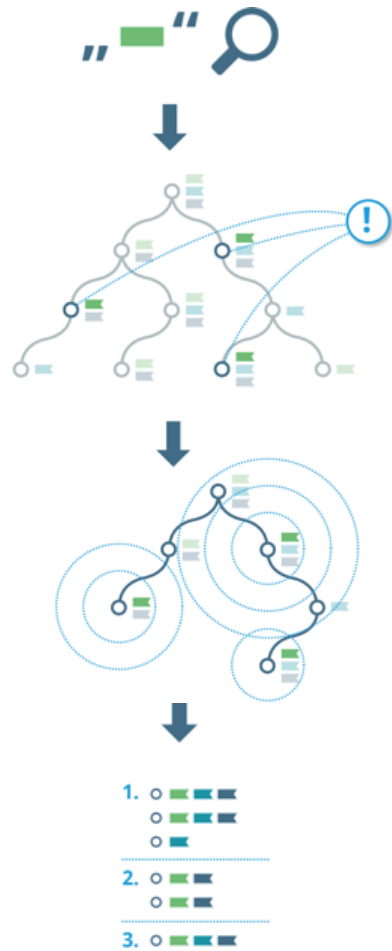


*Fill-up with languages,  
variants, control*



**Multilingual Knowledge System**

# From Strings to Meaning



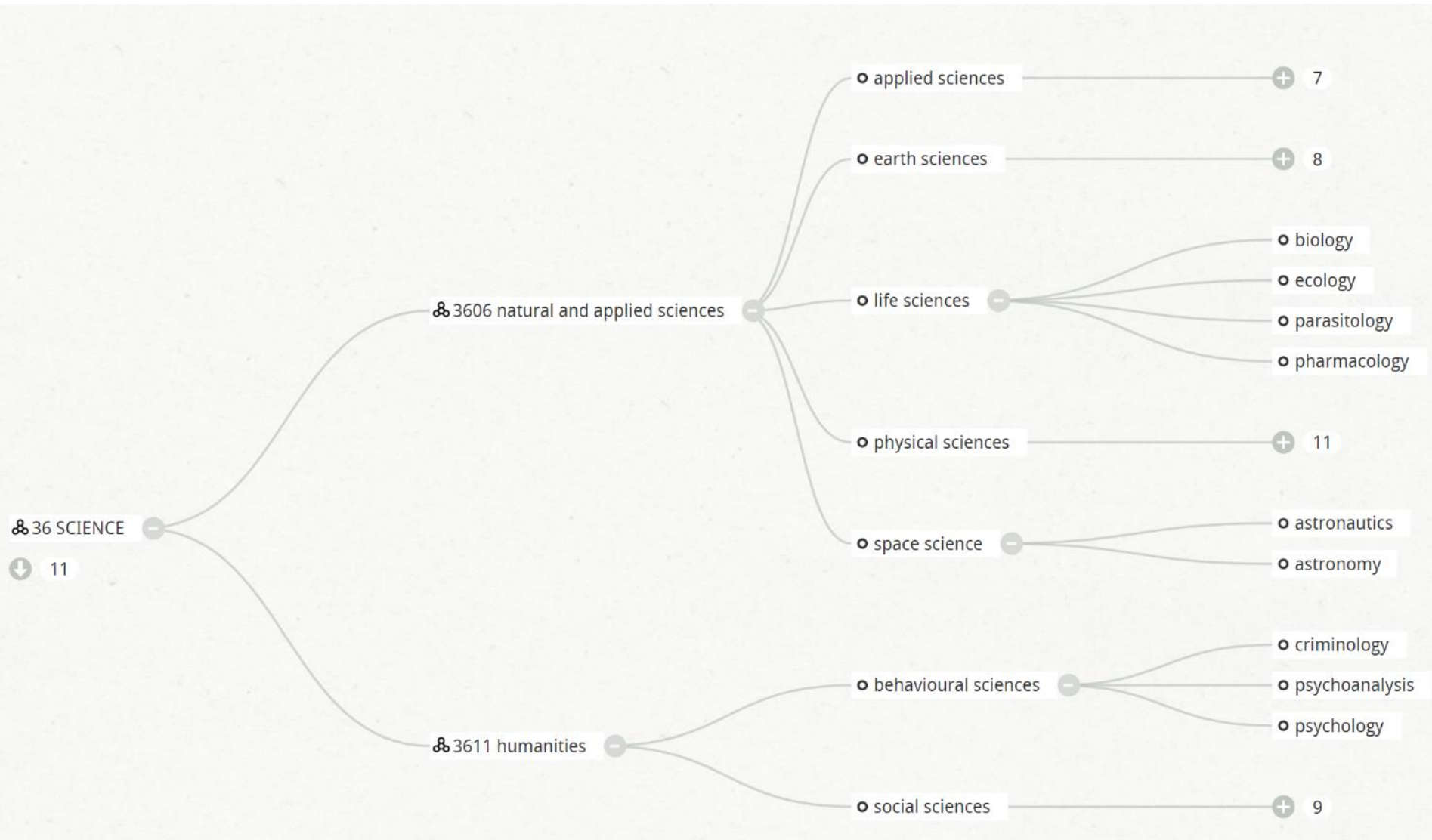
Find terms in text using Advanced Linguistic Search.

Locate them in the knowledge graph.

Create domain heat maps.

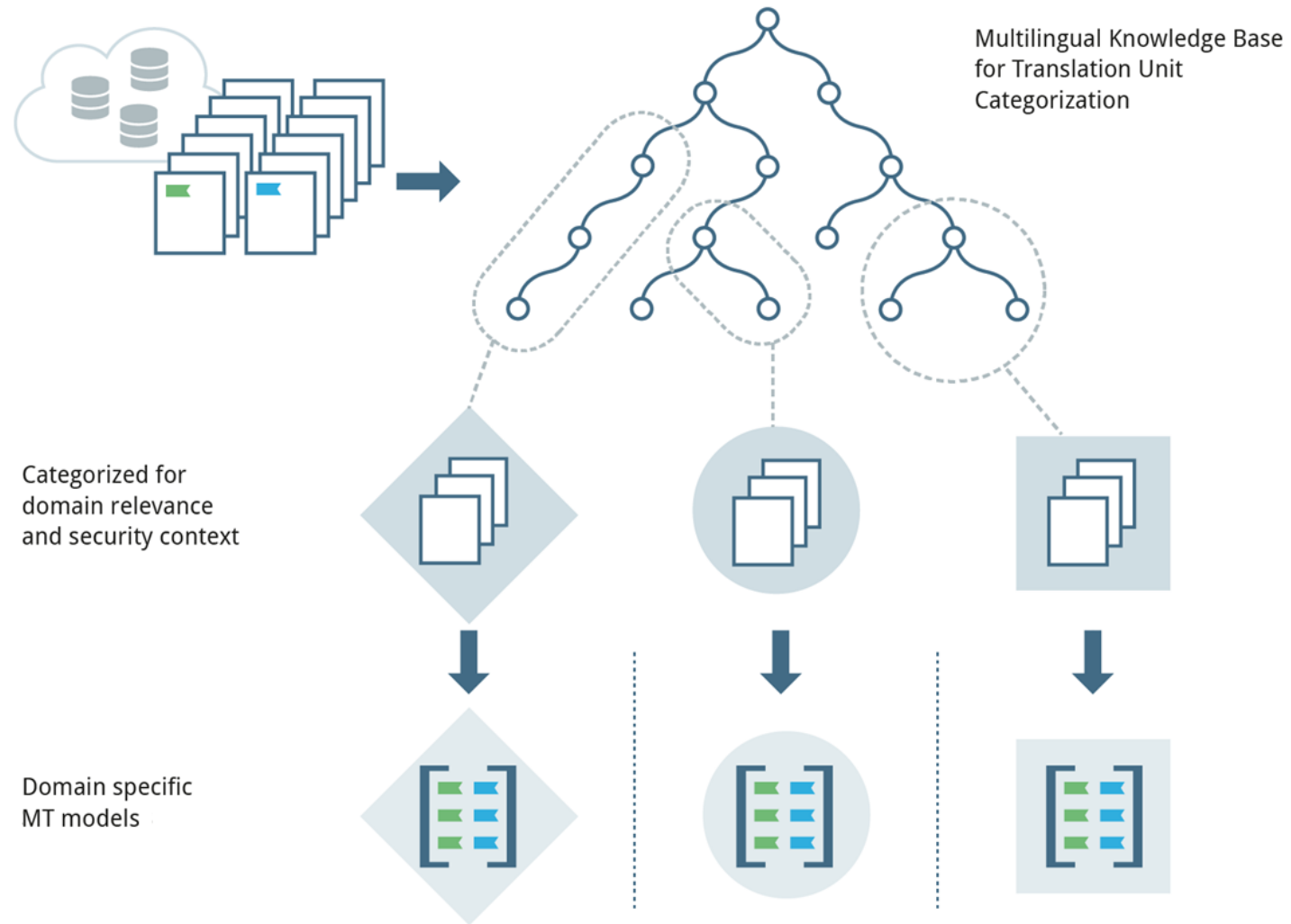
Link to other datasets (HR, DMS, QA rules).

# Classification with a Multilingual Taxonomy

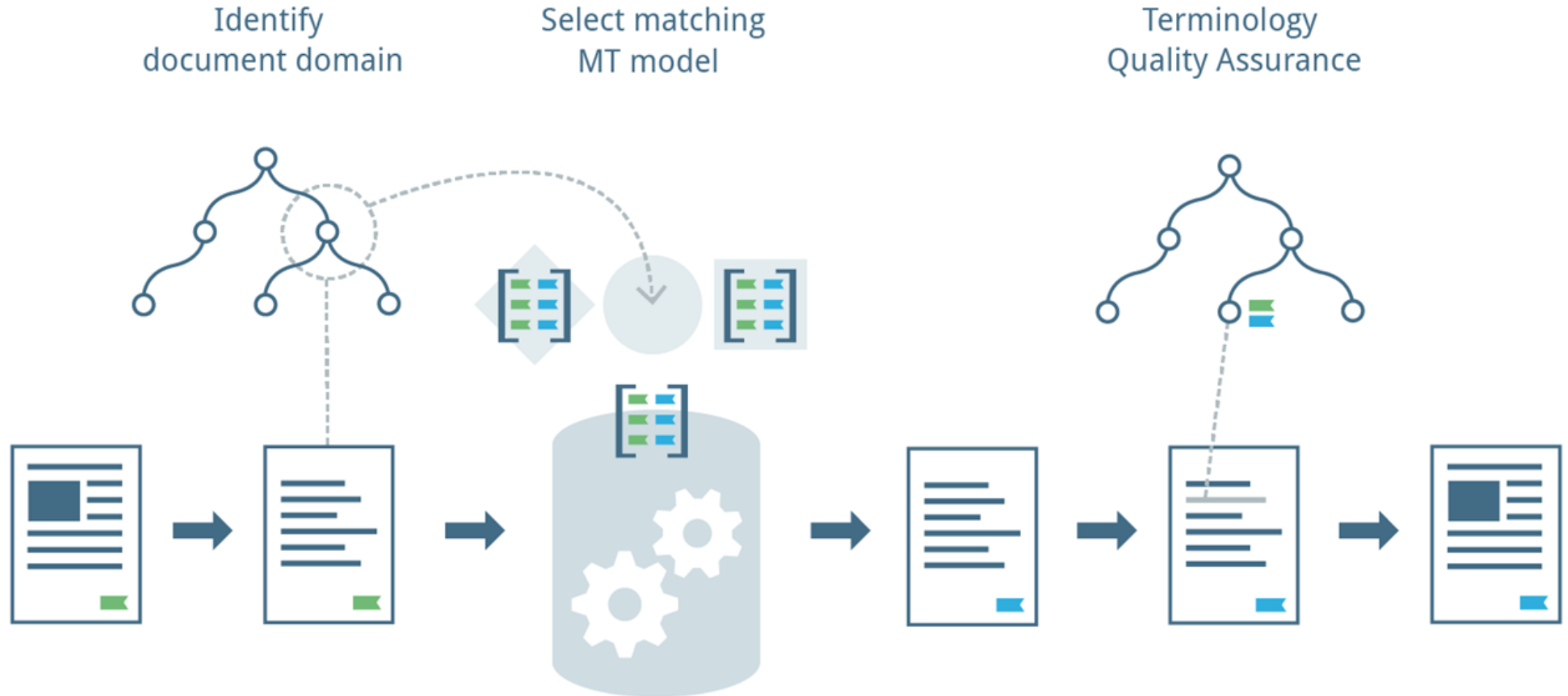




# Knowledge-based MT Training



# Knowledge-based MT Workflow

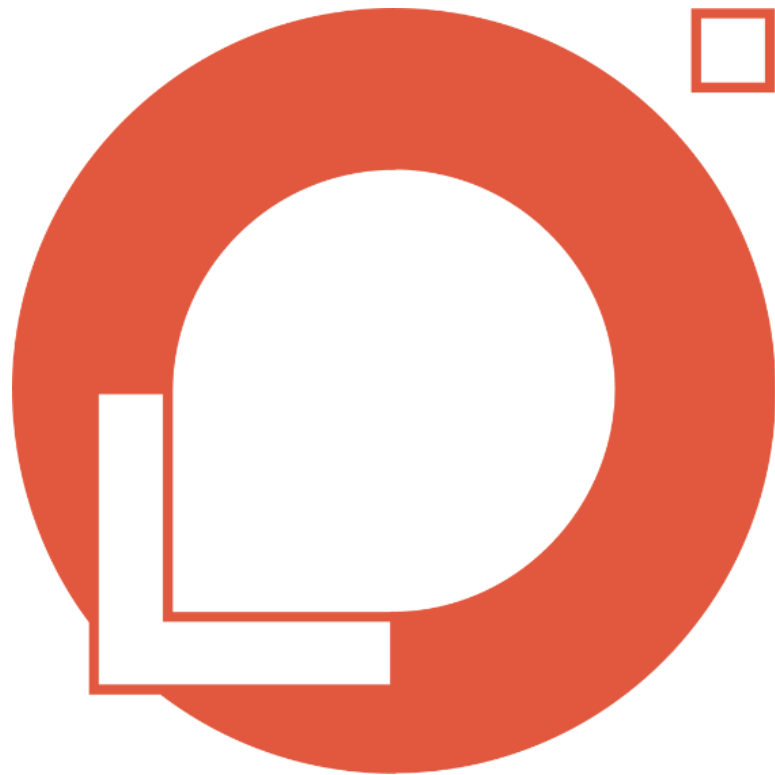


# New Roles for Linguists



- ✓ Revisor
- ✓ Multilingual Knowledge Manager
- ✓ Linguistic Assets Curator
- ✓ Translation Workflow Engineer

# Where do we go from here?



- welcome change
- elevate terminology into knowledge
- define translation quality
- triage workload
- identify low-hanging fruits
- run a pilot
- learn and deploy

# Ask me two Questions!



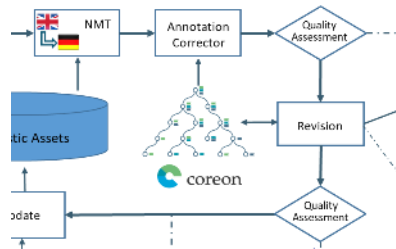
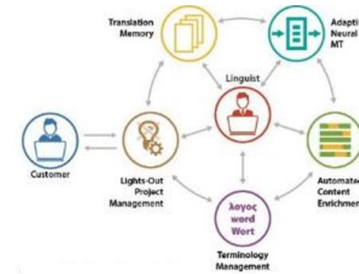
NMT disruptes translation

Industry getting it wrong

NMT Workflow sunsets CAT

Multilingual Knowledge System

New roles for linguists





@JochenHummel • [www.semiox.com](http://www.semiox.com)