

# Semantix

LT Innovate June 25<sup>th</sup>



# MEMSOURCE



# Deriving Business Value from Predicting MT Quality

Semantix

 MEMSOURCE

# Outline

- Intro
- Industry trends
- Memsorce background on MTQE
- Pilot
- Questions





# Who we are

## **Nils Adie**

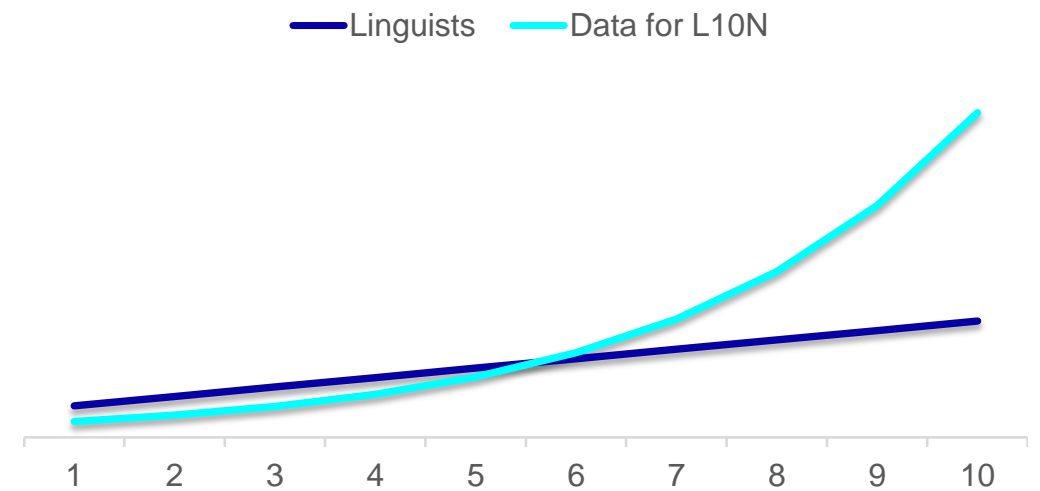
- Director of Technology, Semantix
- In Localization Technology since 2007
- Interest in Users getting the most out of Tech
- Data driven decisions and process improvement

## **Andrea Tabacchi**

- Lead Solution Architect, Memsource
- Technology Enthusiast
- 14 years in Localization
- Working with MT the last 10 years

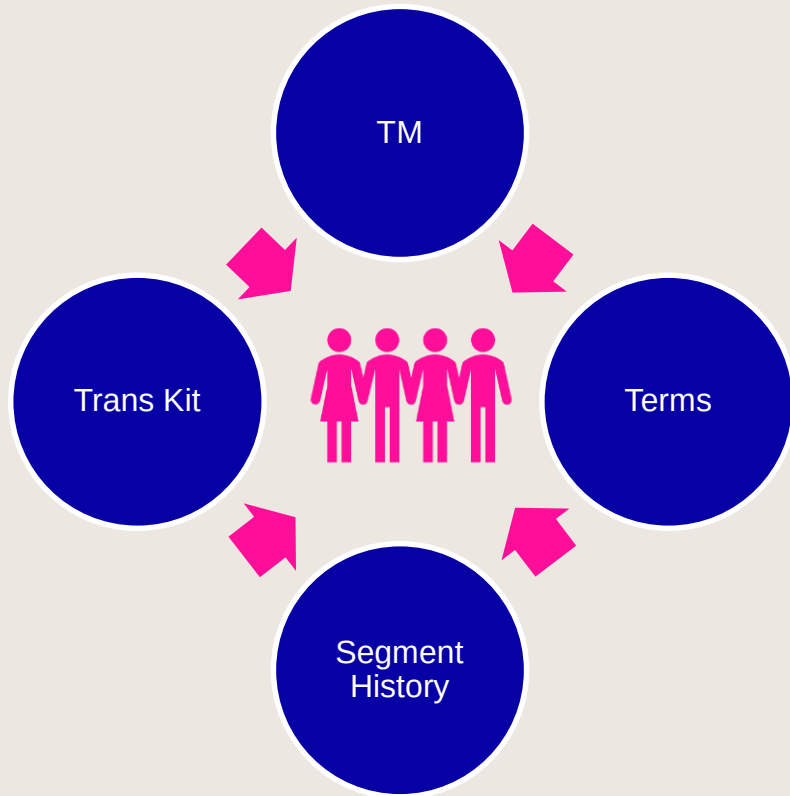
# Industry Trends

- Exponential growth of Data
- Sustainability
- Artisanal versus Assembly line style localization
- Killing the Linguist Centric Localization Setup
- Machine Translation & AI Evolution



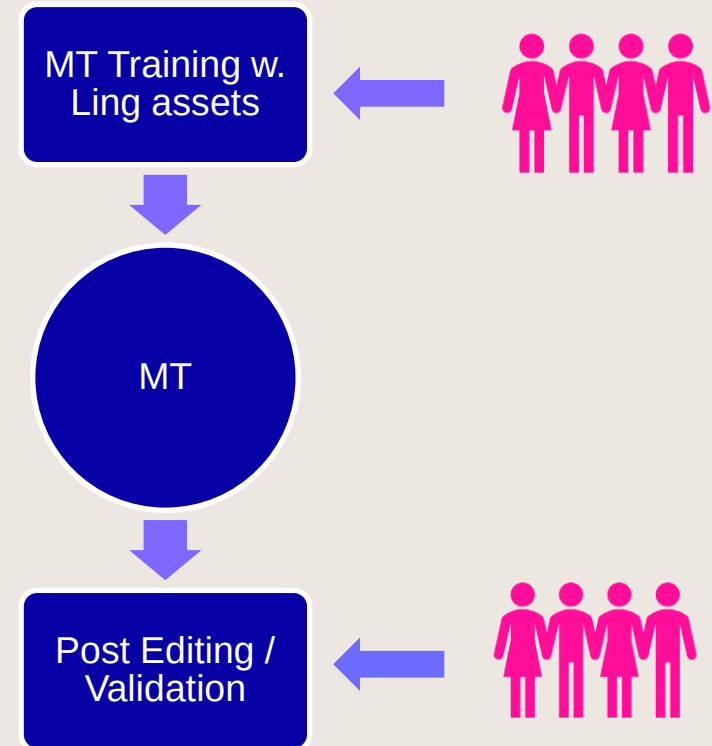
# Localization Centre of Gravity

## Linguist Centric



Becomes

## MT Centric (Linguist Enhanced)



# MACHINE TRANSLATION EVOLUTION

10 years ago

## **IT DOESN'T REALLY WORK**

Because of lack of training data, technology, quality of source content, post-editors availability

5 years ago

## **IT MIGHT ACTUALLY WORK**

But change is hard (and procrastinating is human)

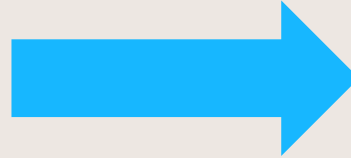
Now

## **IT WORKS**

How can I get the best out of it?

# COMMON OBSTACLES IN MT IMPLEMENTATION

- Complexity of use and predictability of results
- Number of available technologies on the market
- Undefined effort estimation and pricing models



# AND HOW WE SOLVED THEM

- Offer connections to as many technologies as possible
- Simplify the selection of the best tech
- Develop ways to estimate the effort (MTQE) prior to PE or verify it afterwards (edit distance report and time-tracking)



# MEMSOURCE'S AI JOURNEY

2016 – What's all this hype about AI? Studied, went to conferences

2017 – Can we do something with it? Yes we can, but we need a team (and an infrastructure)

2018 – We released NTs and MTQE and we have now a clear vision: **building an intelligent TMS!**

2019 – Can't share right now 😊

# MACHINE TRANSLATION QUALITY ESTIMATION (MTQE)

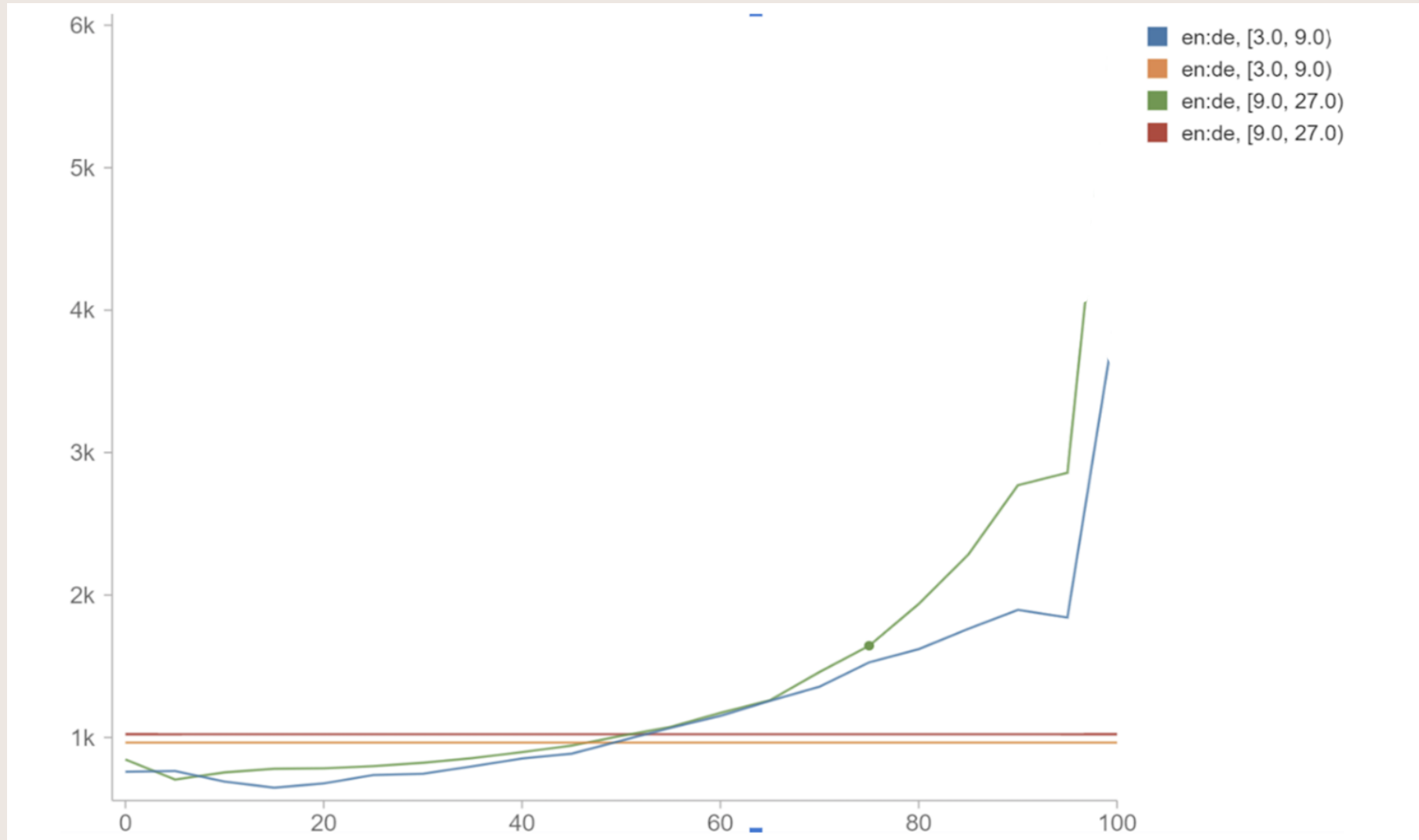
Apply a confidence score to MT outputs

- Based on a proprietary modification of chrF3
- Utilizes deep neural networks
- Trained historical data
- Support for 71 language pairs
- MT engine independent

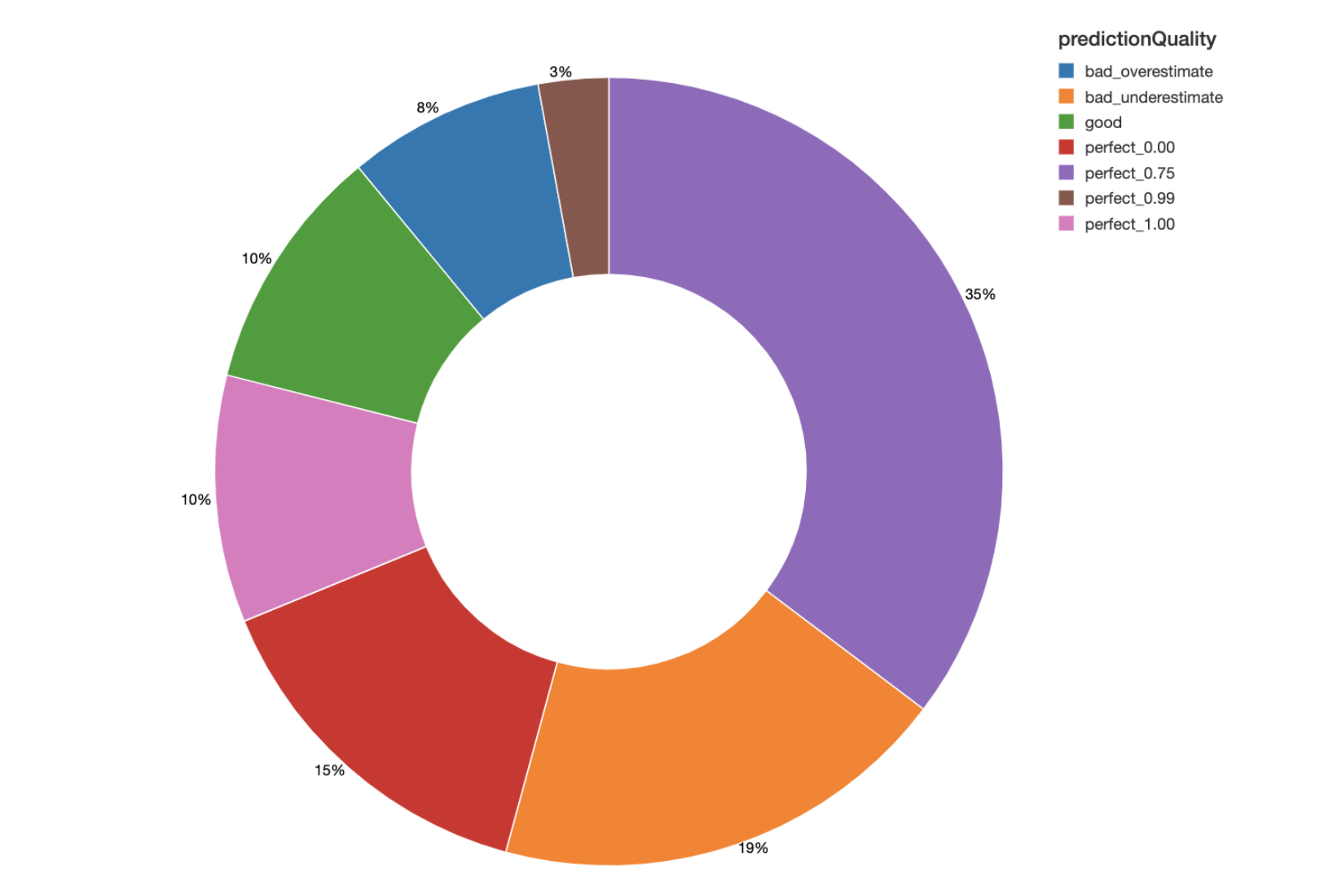
## Goals

- Better project scoping
- Better post-editing

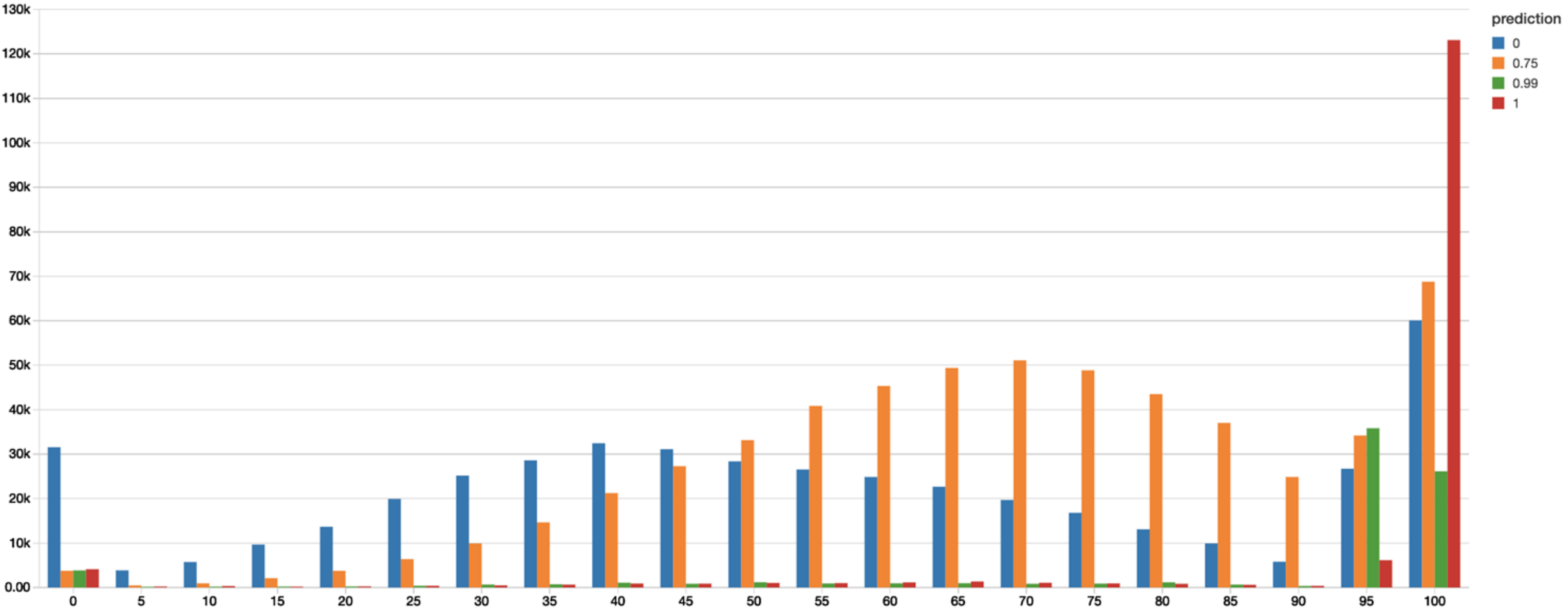
# MACHINE TRANSLATION QUALITY ESTIMATION (MTQE)



# MACHINE TRANSLATION QUALITY ESTIMATION (MTQE)



# MACHINE TRANSLATION QUALITY ESTIMATION (MTQE)



# The Pilot

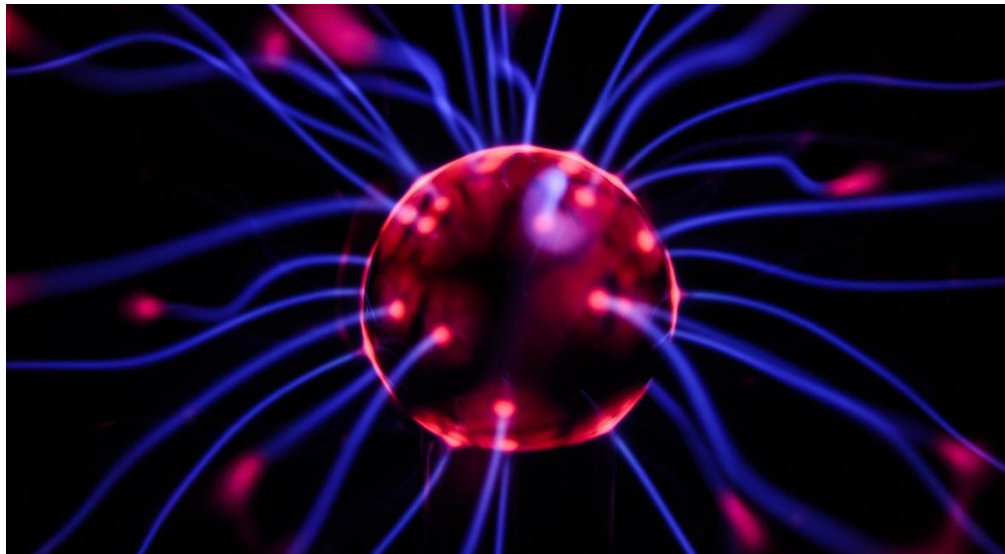
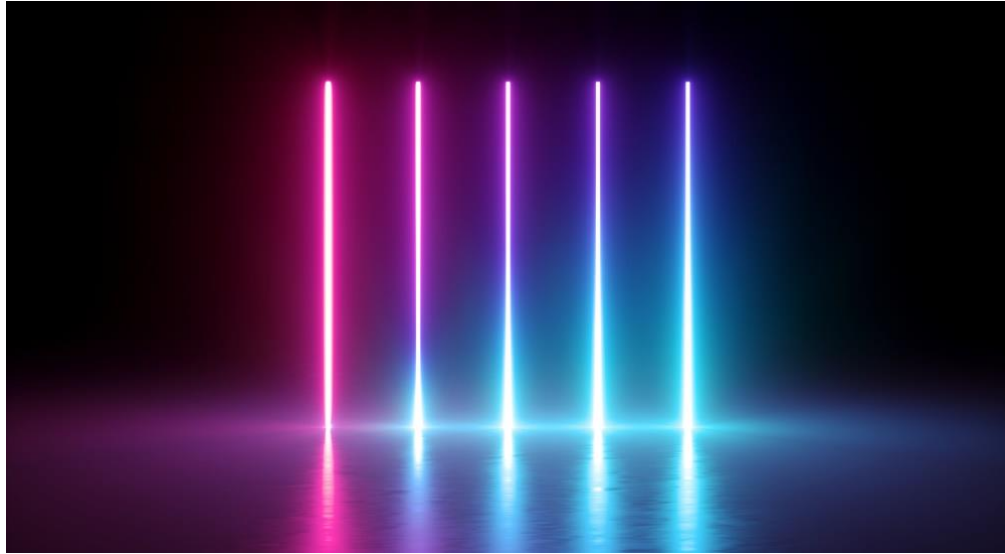
## Background

In a study in Q4 2018 we saw a 50% increase in the net word count drop when deploying MTQE. This underpinned the decision to deploy across the board at Semantix. Did we do the right thing?

## Questions we asked:

- Did the data support our decision?
- What is the MTQE predictive accuracy?
- Can we find a measurable effect on linguist productivity?





# The Pilot: Method

On a subset of our Memsorce clouds Semantix captured aggregate net wordcount data for Jan-May across mainly Scandinavian language combinations.

For 100 projects:

- we ran a separate analysis for each project to be able to separate MTQE predictions from MT post editing results.
- On segment level we extracted linguist editing time, MT score, TM score, and segment length

# The Pilot: Findings

## Question:

Did the data support our decision?

## Results:

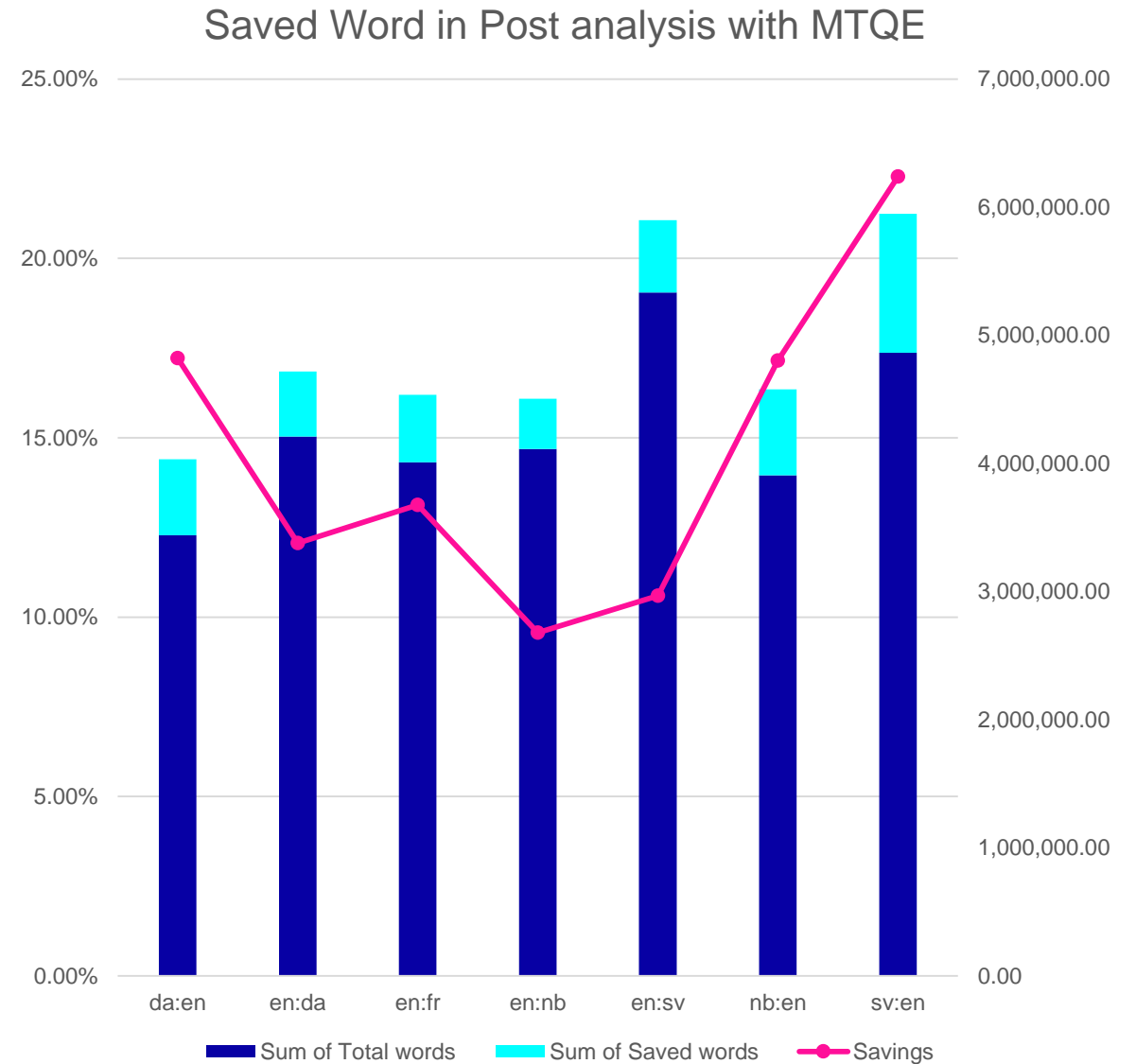
Variations, but average of 15% savings.

## Conclusion:

Yes, in line with our predictions from the previous study.

## Bonus:

Also see smaller but significant savings in unexpected language pairs.





# The Pilot: Findings

## Question:

What is the MTQE predictive accuracy?

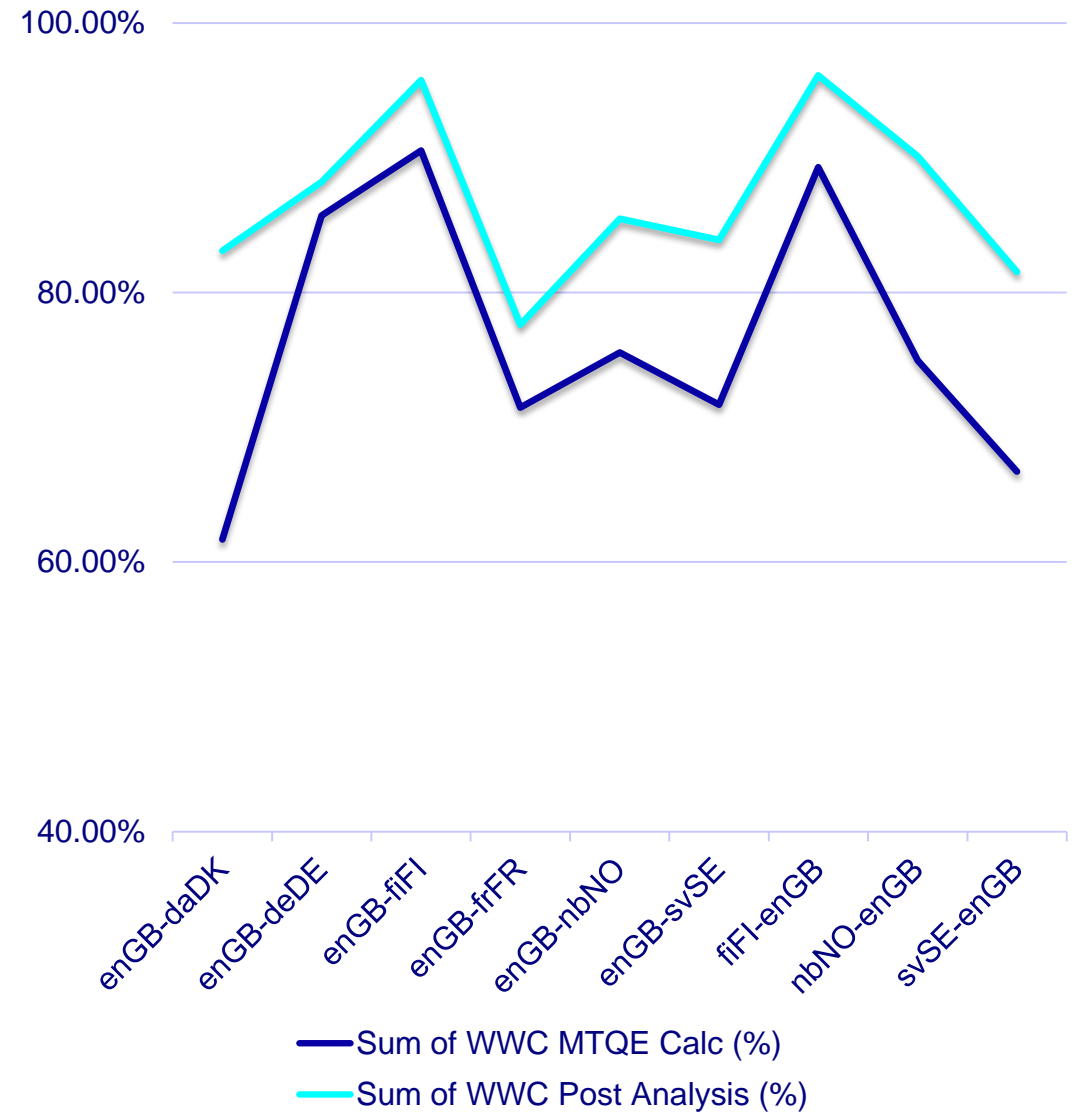
## Results:

MTQE savings predictions are overly optimistic, but there is a correlation with the actual outcome.

## Conclusion:

We can work with this. By tweaking our net rate schemes we can adjust for the outcomes (and have found a pattern that works).

## How outcome compares to prediction



# The Pilot: Findings

## Question:

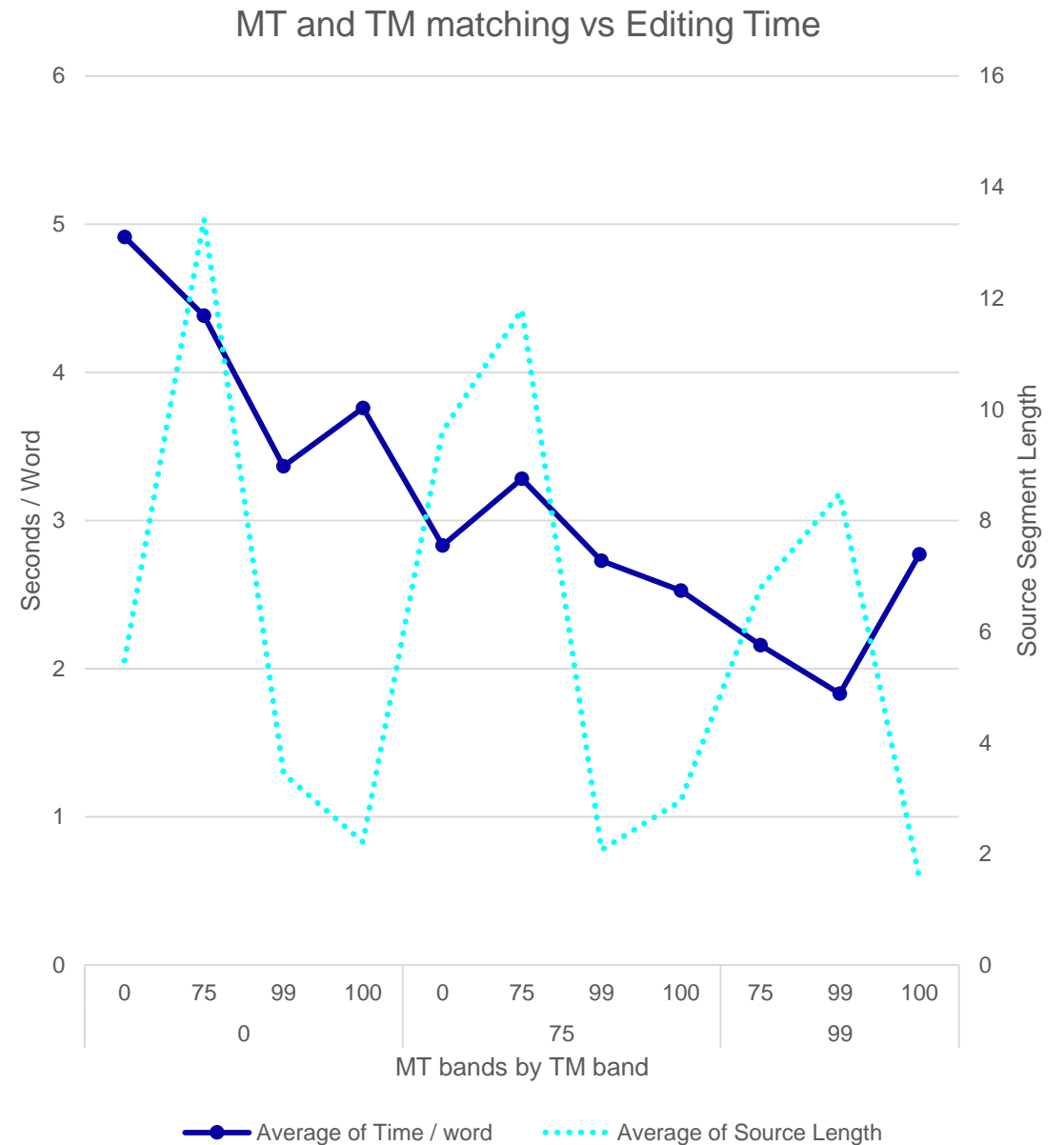
Can we find a measurable effect on linguist productivity?

## Assumption:

Linguist relies on highest of MT and TM score

## Results:

Largely Inconclusive



Questions?