

# LANG.TEC

## SEMANTIC TEXT PROCESSING

---

# LANG.TEC

---

## SEMANTIC TEXT PROCESSING

## Presentation Outline

1. Who we are
2. Business areas
3. Clients
4. Introducing QUEST – the MT Quality Estimator
5. Comparing translation modes (HT, PEMT, QUEST)
6. Demo
7. Model architecture
8. Parting words

# Who we are

**LANG.TEC**

SEMANTIC TEXT PROCESSING

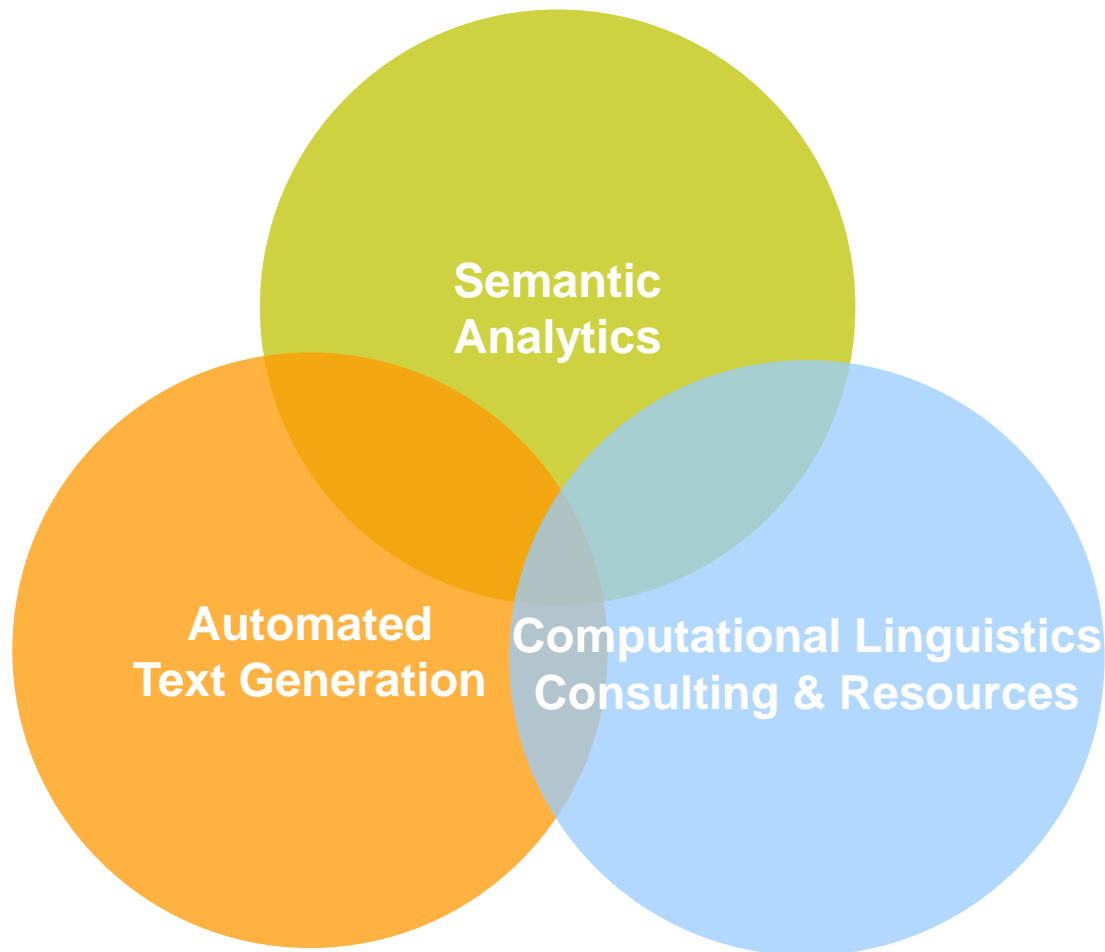
## LangTec:

- **10 years in business, 15 employees**
- **We've been working on MTQE since 2018**
- **QUEST is our second fully functioning QE model and a significant advance over its predecessor**

# Business areas

# LANG.TEC

SEMANTIC TEXT PROCESSING



# Clients

## 1. Semantic Analytics

# LANG.TEC

## SEMANTIC TEXT PROCESSING



# Clients

## 2. Automated Text Generation

# LANG.TEC

SEMANTIC TEXT PROCESSING

TextWriter Client



NLG Consulting



Bespoke Software



# Clients

## 3. Computational Linguistics

# LANG.TEC

## SEMANTIC TEXT PROCESSING

### Localisation



### Resources



### Linguistic Analysis





---

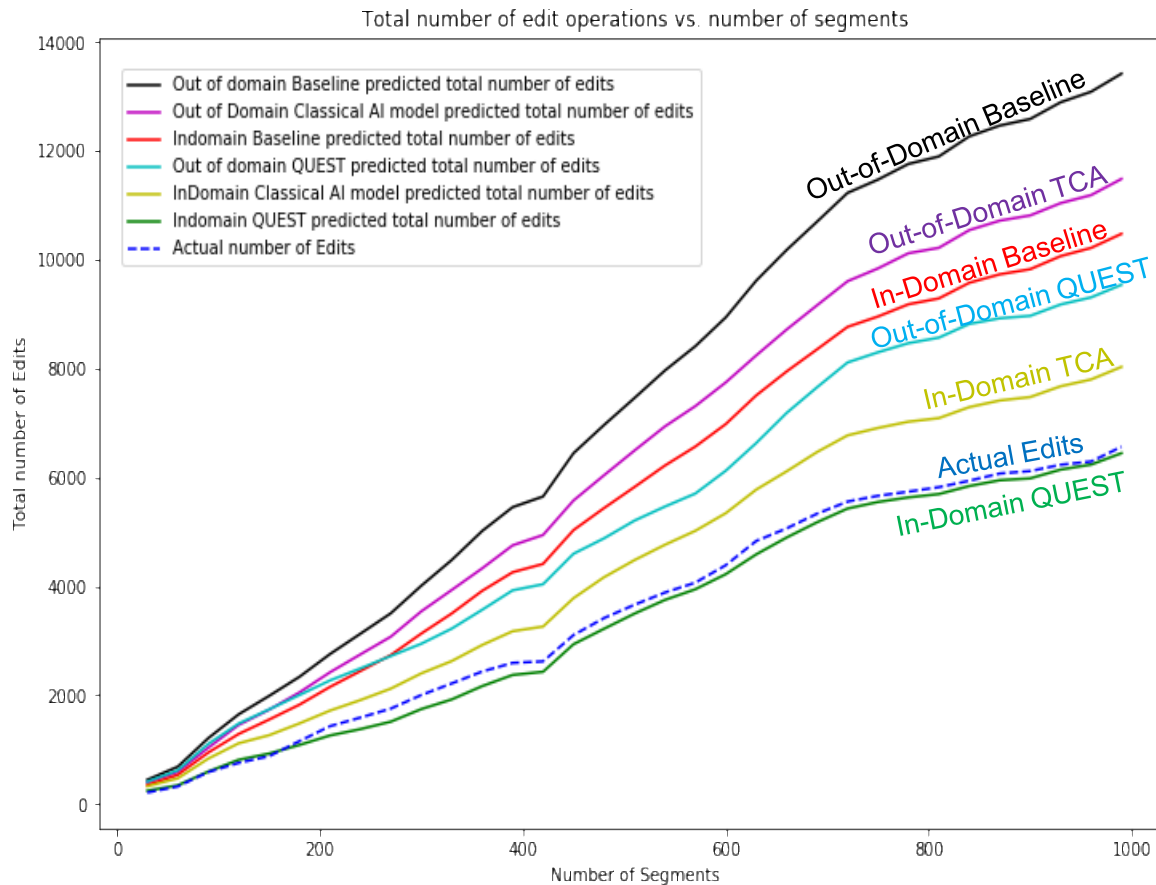
# Introducing QUEST – MT Quality Estimator

# Overview

1. Our in-domain QUEST model is capable of predicting post-editing effort with unprecedented accuracy: only 2% error
2. QUEST can be used with any language pair, any machine translation engine and for any domain.
3. QUEST can also be used out-of-domain still providing superior prediction accuracy compared with experience-based estimations and other machine learning models
4. QUEST's prediction accuracy enables translation service providers to achieve much more competitive pricing and deadline-setting.

## In-Domain QUEST Model

### SEMANTIC TEXT PROCESSING



- When historical domain data are available we can also train in-domain models.
- The in-domain baseline beats the out-of-domain baseline.
- The in-domain baseline also beats the out-of-domain TCA.
- The out-of-domain QUEST model beats all baselines and the out-of-domain TCA.
- The in-domain QUEST model beats all baselines and all TCA models, very closely approximating the actual post-editing effort.

## Comparing approaches

### SEMANTIC TEXT PROCESSING

	Manual Translation (HT)	MT + Post-Editing (In-domain baseline)	MT + Post-Editing + QUEST (out-of-domain)	MT + post-editing + QUEST (in-domain)	Actual
Word Count	22,635	22,635	22,635	22,635	22,635
Predicted Post-Editing Operations	0	10,474	9,538	<b>6,443</b>	6,561
Prediction Error	0%	60%	45%	<b>2%</b>	0%
Total Turnover Effort	56.59 hrs. 7.07 person days	48.00 hrs. 6.00 person days	43.72 person hrs. 5.46 person days	<b>29.53 hrs. 3.69 person days</b>	30.07 hrs 3.76 person days
Fraction of words to be processed	100%	46%	42%	<b>28%</b>	29%
Per-Word Rate	EUR 0.09	EUR 0.05	EUR 0.04	<b>EUR 0.03</b>	EUR 0.03
Total Cost	EUR 2,037	EUR 1,132	EUR 905	<b>EUR 679</b>	EUR 679

# Demo

# LANG.TEC

## SEMANTIC TEXT PROCESSING

### QUEST LIVE-DEMO

Submit a text and get a machine learning-based prediction of the post-editing effort that will be needed to correct its machine translation.

Currently supported are German texts to be translated into English. .tmx files may also contain English machine translations in addition to the German source segments. Our models perform most reliably in the domain of 'technical manuals'. For this demo, we recommend submitting texts that contain no more than 5000 segments.

### Select Model Configuration

Select Your Prediction Model for Quality Estimation:

Deep Learning Model with with 17.8 ▾

Select Language Pair:

DE --> EN ▾

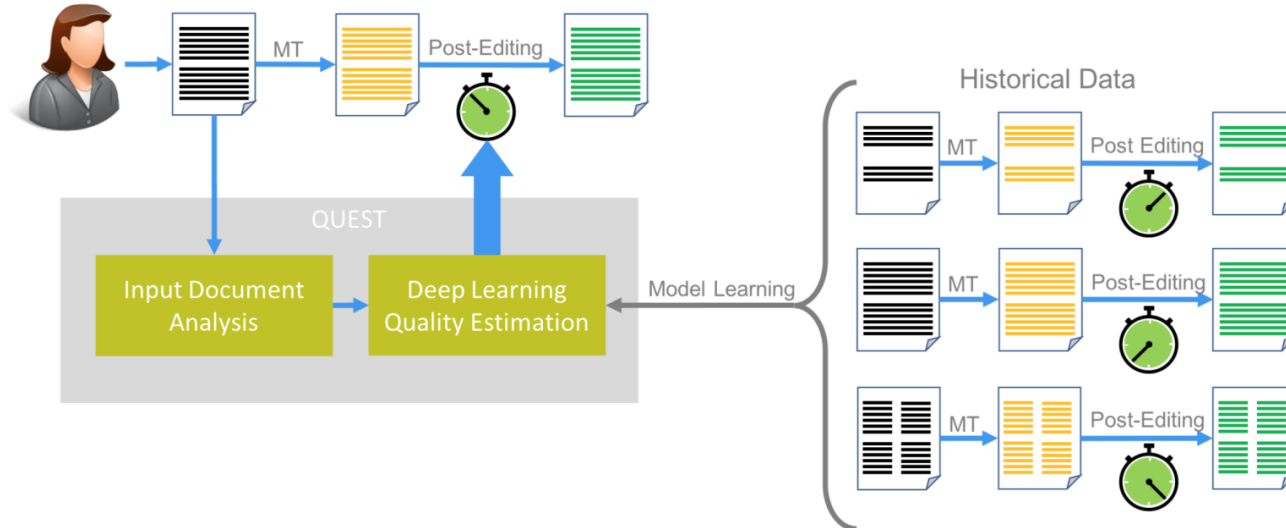
### Upload File

[Choose a file](#)  
(.tmx & .txt only)

### Input Text

## Different PE effort estimations

# Model architecture



## QUEST's usage contexts

1. Post-Edit Effort Prediction
2. MT-Recommendation Tool
3. MT-Profiling Tool
4. MT Quality Document Sorting
5. MT Quality Threshold Plug-In

## Parting words

- QUEST can be used on-premise or in the cloud and custom models can be built within hours.
- Our machine learning model also scales to any language pair
- QUEST works as solution for leveraging MT for lesser-resourced languages
- LangTec can provide a free customized trial QUEST model for interested parties



---

# Thank you!

# LANG.TEC

---

SEMANTIC TEXT PROCESSING

LangTec

Dr. Christopher Reid  
Associate Product Owner QUEST

Bundesstraße 28 a  
20146 Hamburg  
Germany

Phone: +49 . 40 . 180 610 98

E-mail: [christopher.reid@langtec.de](mailto:christopher.reid@langtec.de)

Web: [www.langtec.de](http://www.langtec.de)

---

# LANG.TEC

---

## SEMANTIC TEXT PROCESSING