www.cdt.europa.eu

TRANSLATION CENTRE
FOR THE BODIES OF THE EUROPEAN UNION

Machine Translation Quality Monitoring at the CdT

Luxembourg, 24 November 2022

Daniel Marín & Mauro Bubnic

SUBTITLING

MODIFICATION
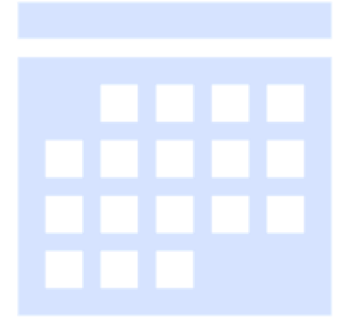
TRANSLATION

EDITING

LANGUAGE CONSULTANCY

REVISION

TERMINOLOGY

Background

2018-2020
Transformation plan
Cooperation program (EUIPO)

2020-2021
MT/PE as part of the translation workflow

2021-2022
Light post-editing service

2023-
Extend PE
+ Multi-engine

MT-based services at the CdT

FPE Full Post-Editing

- Internal working method
- Four-eyes principle (post-editor + revisor)
- Quality equivalent to that of human translation

LPE Light Post-Editing

- Fully-fledged service
- Two-eyes principle (post-editor)
- Accurate translation, but without stylistic rework

AT Automatic Translation

- Fully automatic service
- Translation memories + machine translation + layout

CdT's MT system

NICE: Neural Integrated Custom Engines

Daniel Marín Buj[1], **Daniel Ibáñez García**[1], **Zuzanna Parcheta**[1], **Francisco Casacuberta**[2]

daniel.marin@cdt.europa.eu
{daniel.ibanez,zuzanna.parcheta}@ext.cdt.europa.eu
fcn@prhlt.upv.es

[1] Translation Centre for the Bodies of the European Union
[2] PRHLT Research Center, Universitat Politecnica de València

- Internal implementation using the **open-source toolkit OpenNMT-tf**[1] for NMT.
- All models are **customised and fine-tuned** for the post-editing use case.
- The first models were built in the **intellectual property** and **public health** fields.
- All engines are **evaluated by translators before deployment** in production.

[1] Klein, G. et al, 2017. Opennmt: Open-source toolkit for neural machine translation. arXiv preprint arXiv:1701.02810.

Monitoring objectives

Analyse the actual effort made by post-editors.

Maintain and improve internal MT systems.

Define long-term strategies for MT and PE.

Metric

The main metric used is the **normalised edit distance (Levenshtein):**

The minimum number of single-character edits required to change one sequence into another.

Similarity
- **A ratio reflecting the similarity between two sentences:**
  - 100% means that the two sentences are identical.
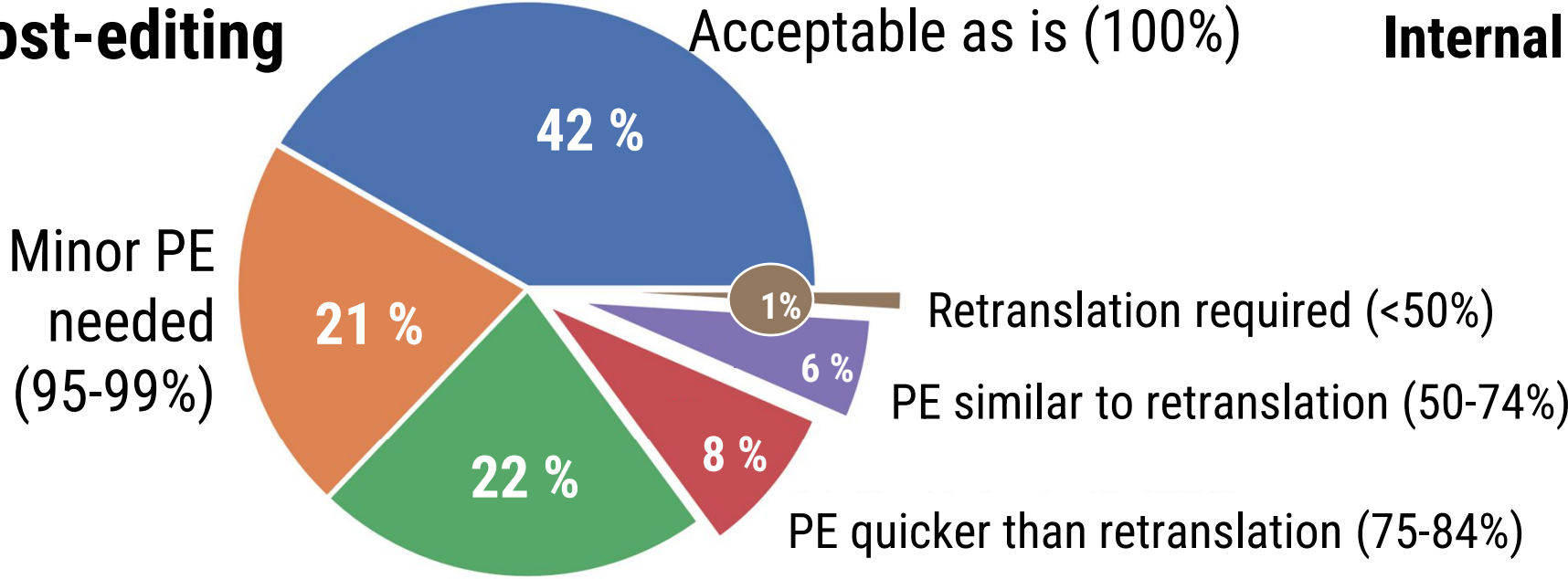 - 0% means that the two sentences have nothing in common.

Effort
- **The effort is computed as the inverse of the similarity:**
  - 0% means no editing.
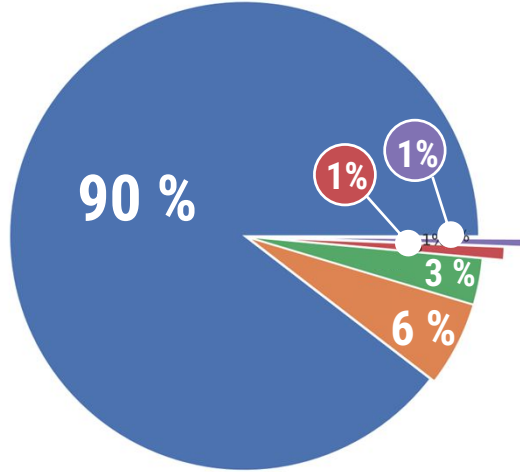 - 100% means full retranslation.

Conclusions

Overall MT quality is fit for post-editing as initial evaluations anticipated.
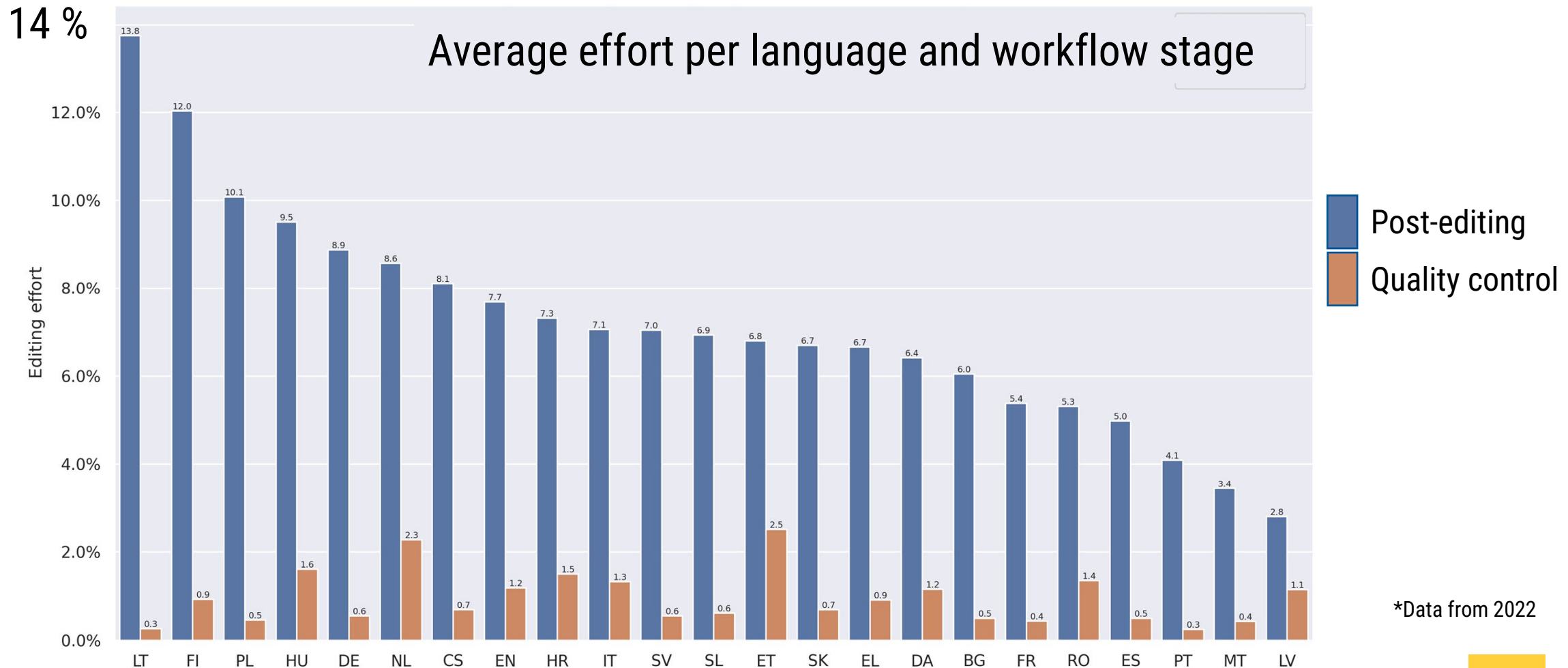
Post-editing

Internal revision

Acceptable as is (100%)

42 %

Minor PE needed (95-99%)

21 %

22 %

8 %

1%

6 %

Retranslation required (<50%)

PE similar to retranslation (50-74%)

PE quicker than retranslation (75-84%)

Some PE needed (85-94%)

90 %

1%

1%

3 %

6 %

*Data from 2022

TRANSLATION CENTRE FOR THE BODIES OF THE EUROPEAN UNION

Language Technology Group

Conclusions

14 %

Average effort per language and workflow stage

Editing effort

- Post-editing
- Quality control

*Data from 2022

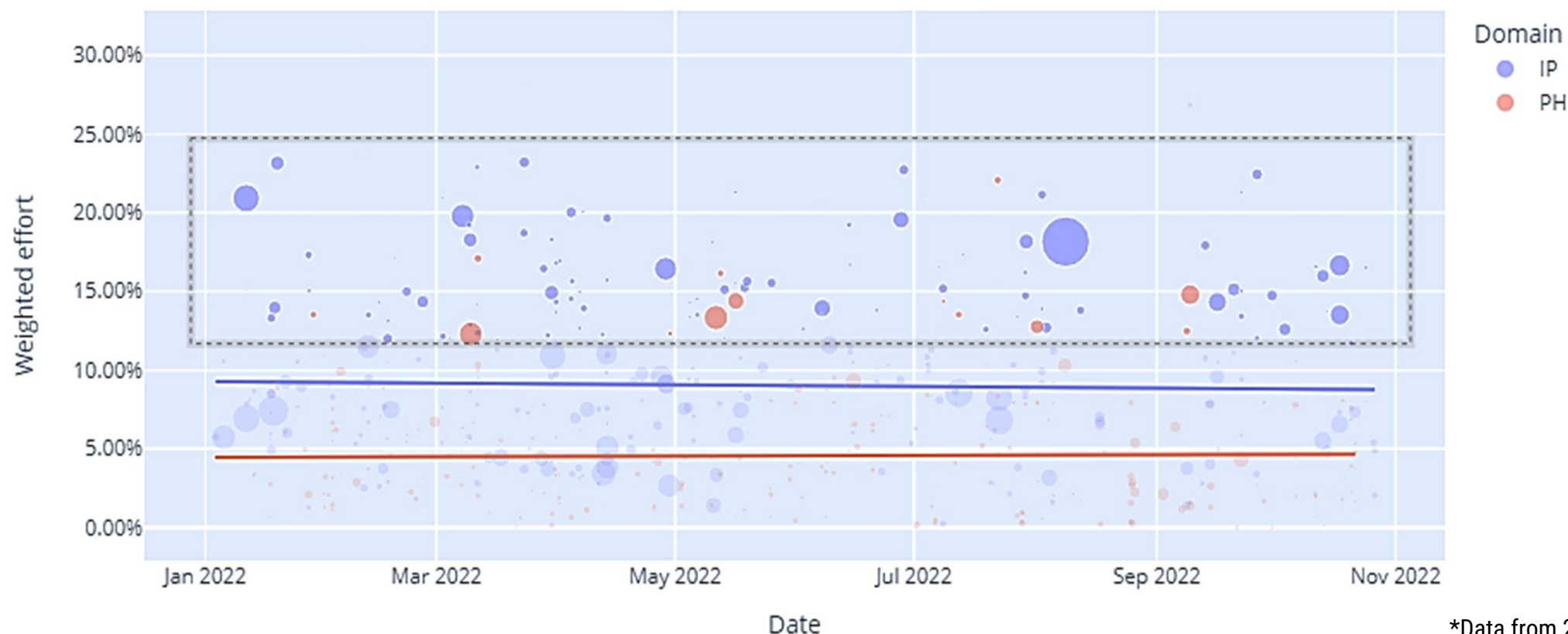| Language	Post-editing	Quality control
LT	13.8	0.3
FI	12.0	0.9
PL	10.1	0.5
HU	9.5	1.6
DE	8.9	0.6
NL	8.6	2.3
CS	8.1	0.7
EN	7.7	1.2
HR	7.3	1.5
IT	7.1	1.3
SV	7.0	0.6
SL	6.9	0.6
ET	6.8	2.5
SK	6.7	0.7
EL	6.7	0.9
DA	6.4	1.2
BG	6.0	0.5
FR	5.4	0.4
RO	5.3	1.4
ES	5.0	0.5
PT	4.1	0.3
MT	3.4	0.4
LV	2.8	1.1

Conclusions

Mapping clients and engines does not always account for the actual domain of the documents.
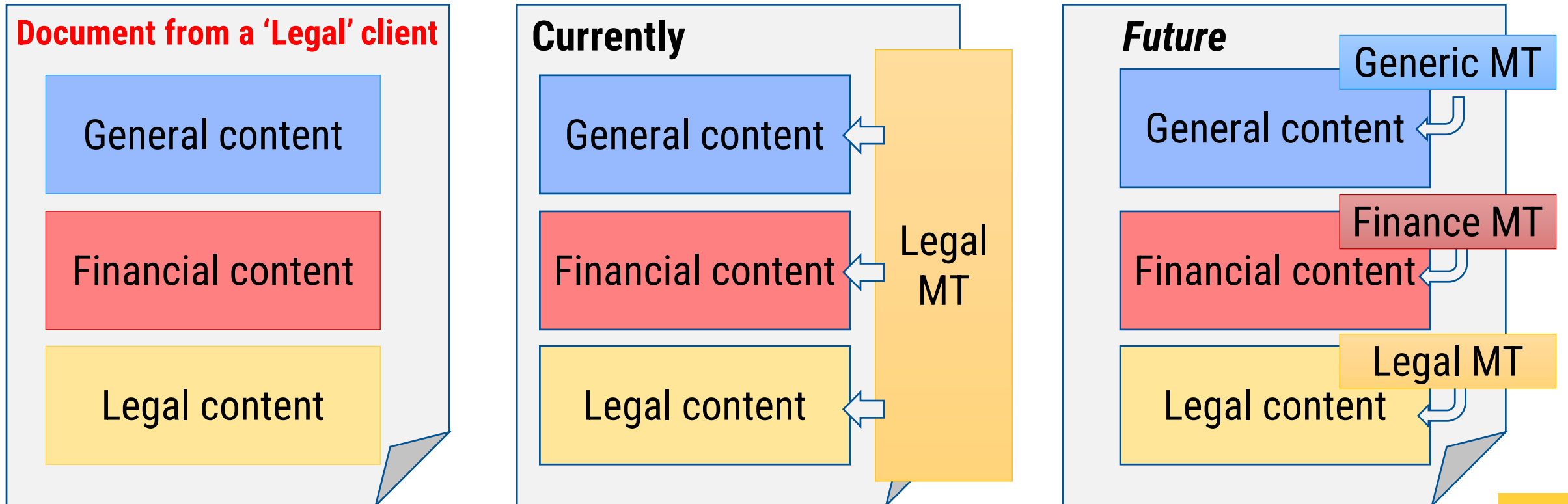
Effort by request

*Data from 2022

Multi-engine strategy

- Supported by quality monitoring and users' feedback.
- More dynamic and adapted approach for MT-based services.
- Applying domain adaptation at sentence level in real time.

Document from a 'Legal' client

General content

Financial content

Legal content

Currently

General content

Financial content

Legal content

Legal MT

Future

Generic MT

General content

Finance MT

Financial content

Legal MT

Legal content

www.cdt.europa.eu

TRANSLATION CENTRE
FOR THE BODIES OF THE EUROPEAN UNION

Thanks!

Daniel MARÍN / Mauro BUBNIC

daniel.marin@cdt.europa.eu
mauro.bubnic@cdt.europa.eu

cdt.europa.eu

translationcentre

eutranslationcentre

@EU_CdT

SUBTITLING

MODIFICATION

TRANSLATION

EDITING

LANGUAGE CONSULTANCY

REVISION

TERMINOLOGY