# Building
# A Multi-Purpose
# MT Portfolio

AMTA VIRTUAL 2020

INTENTO

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*
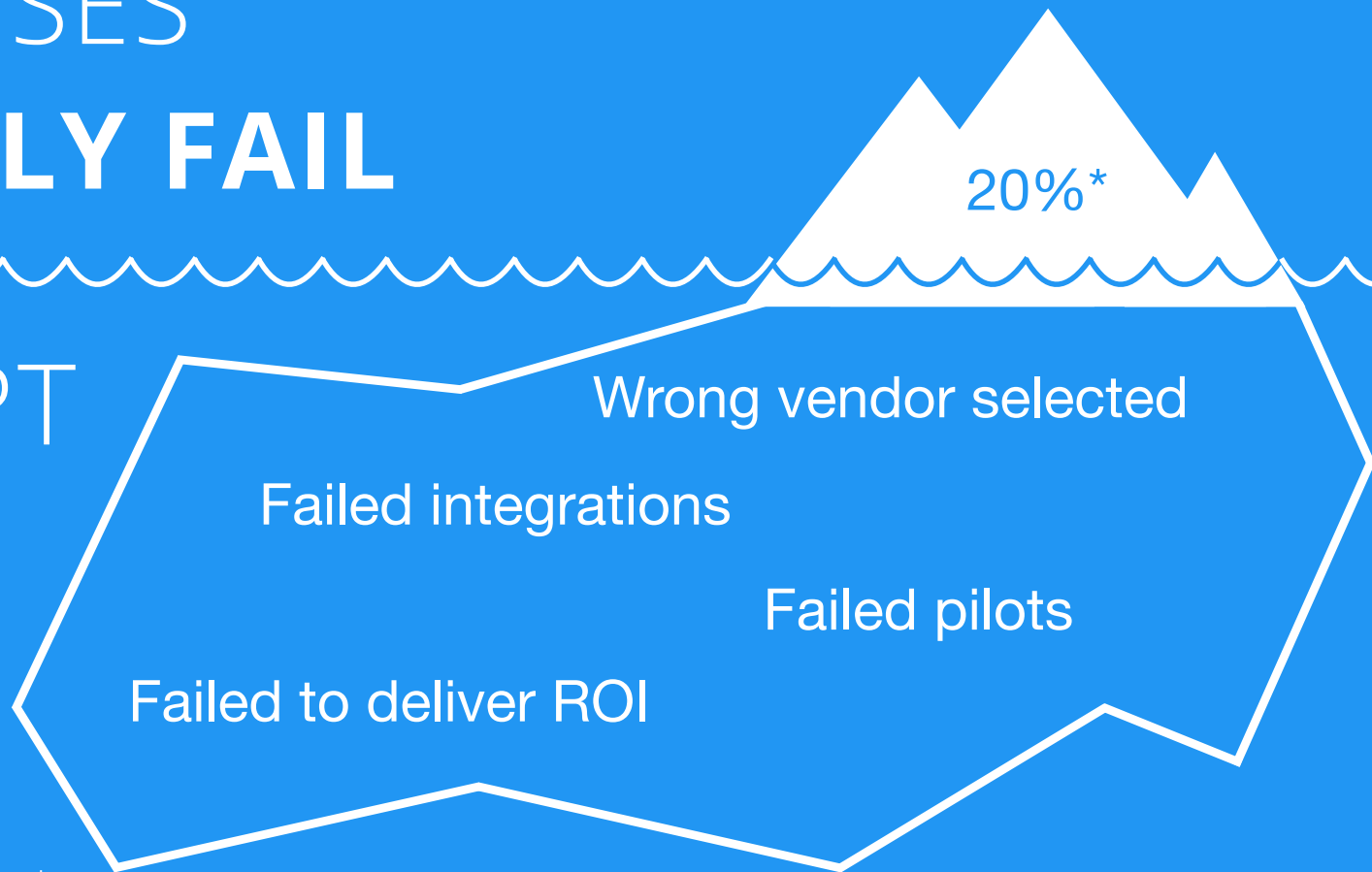
*Page 309*

# AGENDA

Multi-Purpose MT?

—

MT usage scenarios and requirements

—

Case Study 1: **Entity Protection**

—

Case Study 2: **Custom Terminology**

—

Case Study 3: **Tone of Voice**

—

Key Takeaways

2

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 310*

# MULTI-PURPOSE MT?

3

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 311*

# ENTERPRISES
## MASSIVELY FAIL

20%*

## TO ADOPT
## AI

Wrong vendor selected
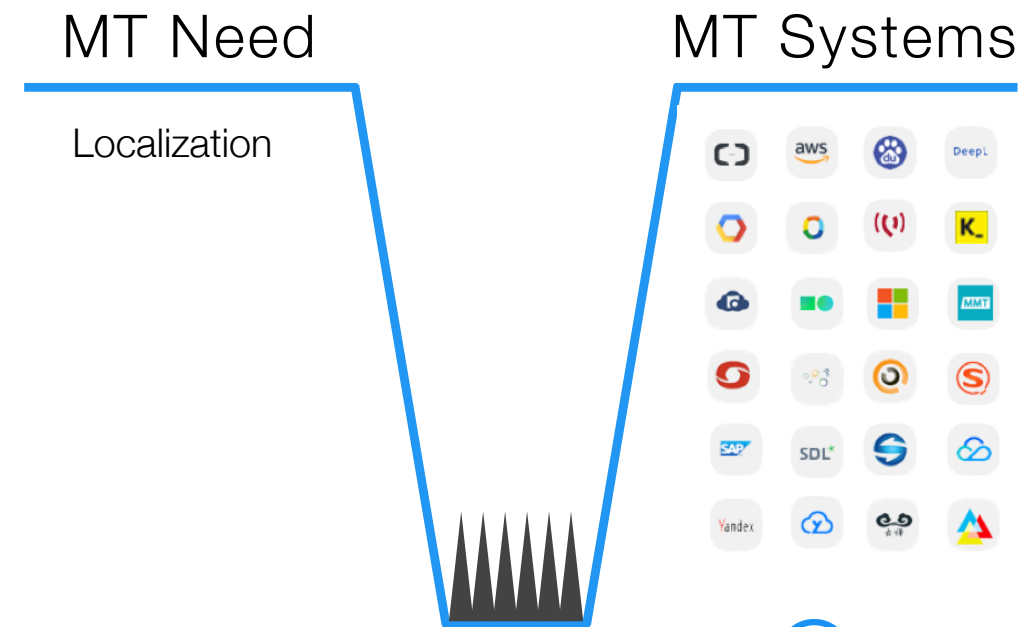
Failed integrations

Failed pilots

Failed to deliver ROI

* Share of US companies with successful AI deployment
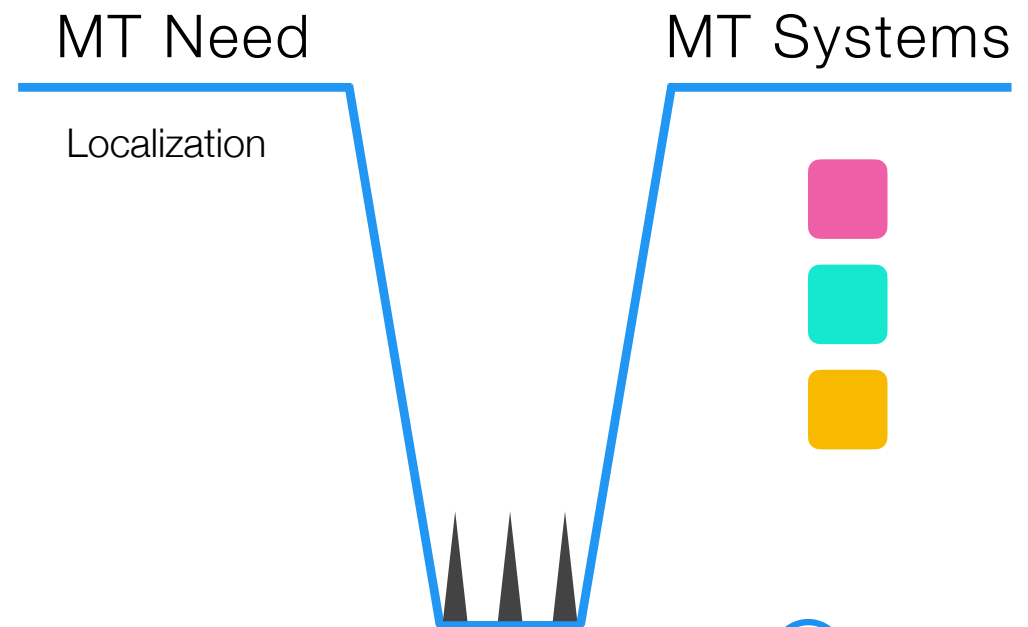  (*Deloitte State of Cognitive Survey 2017*)

© Intento, Inc.  /  September 2020

4

**INTENTO**

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page  312*

# BRIDGING THE GAP BETWEEN
## MT CAPABILITIES AND ADOPTION

MT Procurement

MT Need                                      MT Systems

Localization

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page  313

# BRIDGING THE GAP BETWEEN
MT CAPABILITIES AND ADOPTION

MT Procurement

MT Need

MT Systems

Localization

6

Intento

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 314

# BRIDGING THE GAP BETWEEN
## MT CAPABILITIES AND ADOPTION

MT Procurement

—

MT Curation

MT Need        MT Systems

Localization

7

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 315*

# BRIDGING THE GAP BETWEEN
## MT CAPABILITIES AND ADOPTION

MT Procurement

—

MT Curation

—

Multi-Engine MT

MT Need                                    MT Systems

Localization

Intento
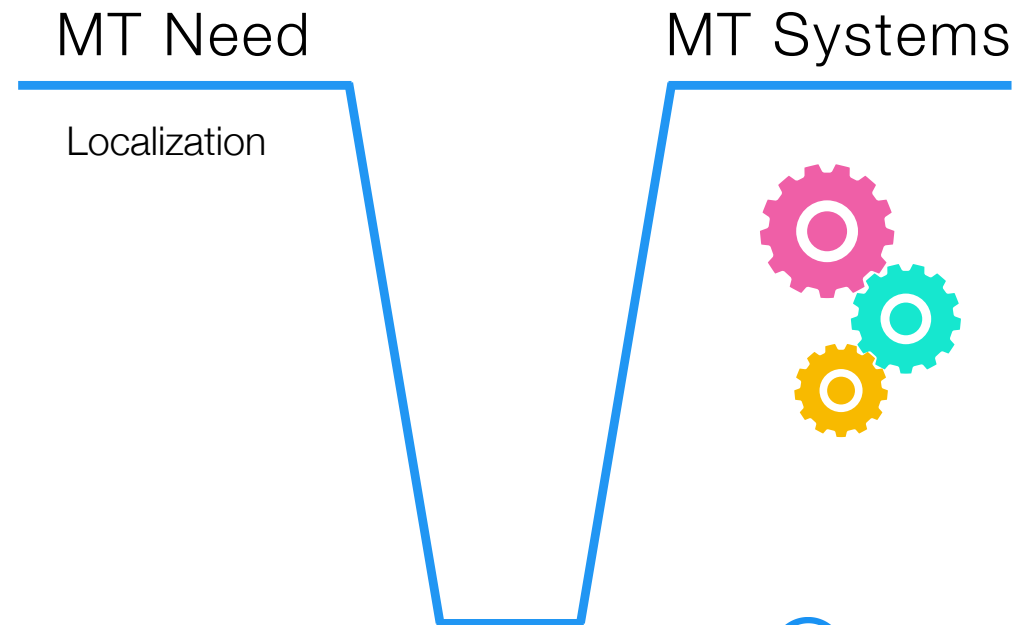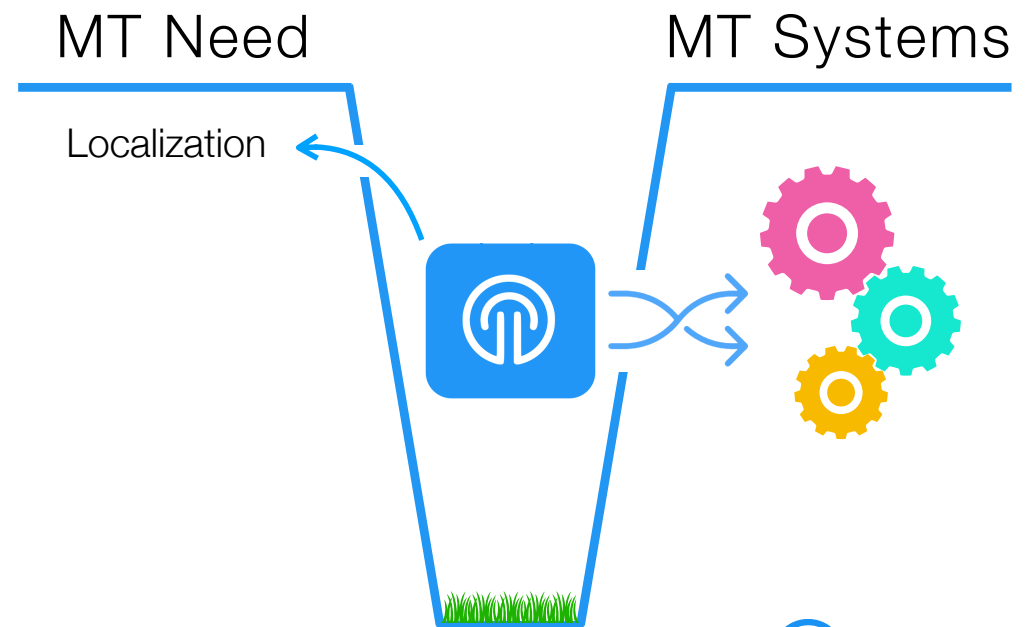
# BRIDGING THE GAP BETWEEN
## MT CAPABILITIES AND ADOPTION

MT Procurement
—
MT Curation
—
Multi-Engine MT
—
Multi-Purpose MT

MT Need                              MT Systems

Localization

Customer Service

Office Productivity

Global Community

9

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 317*
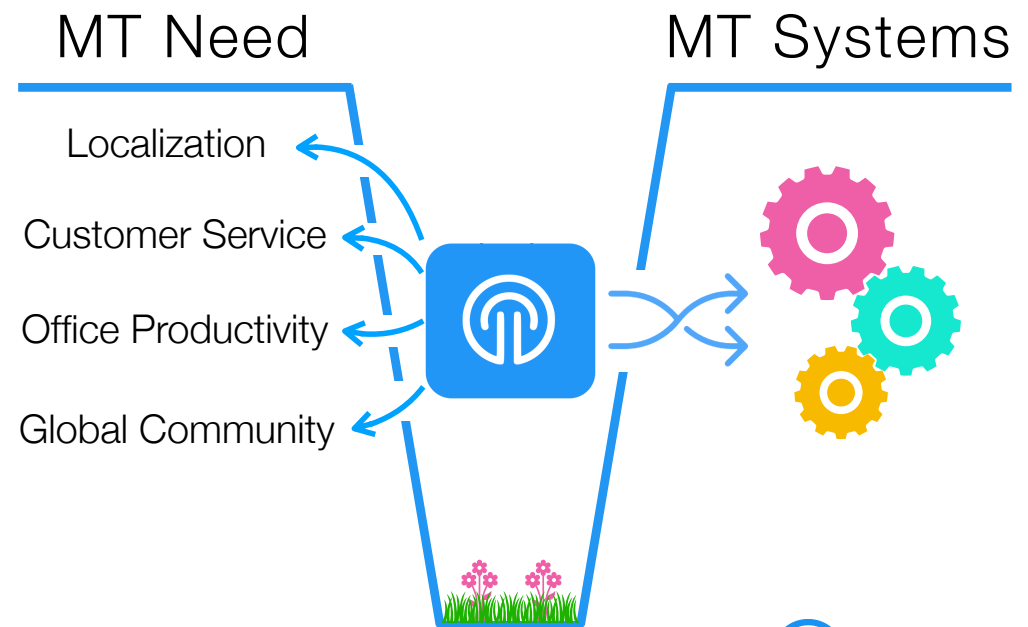
# MULTI-PURPOSE MT

Instant ROI on the investments already made

—

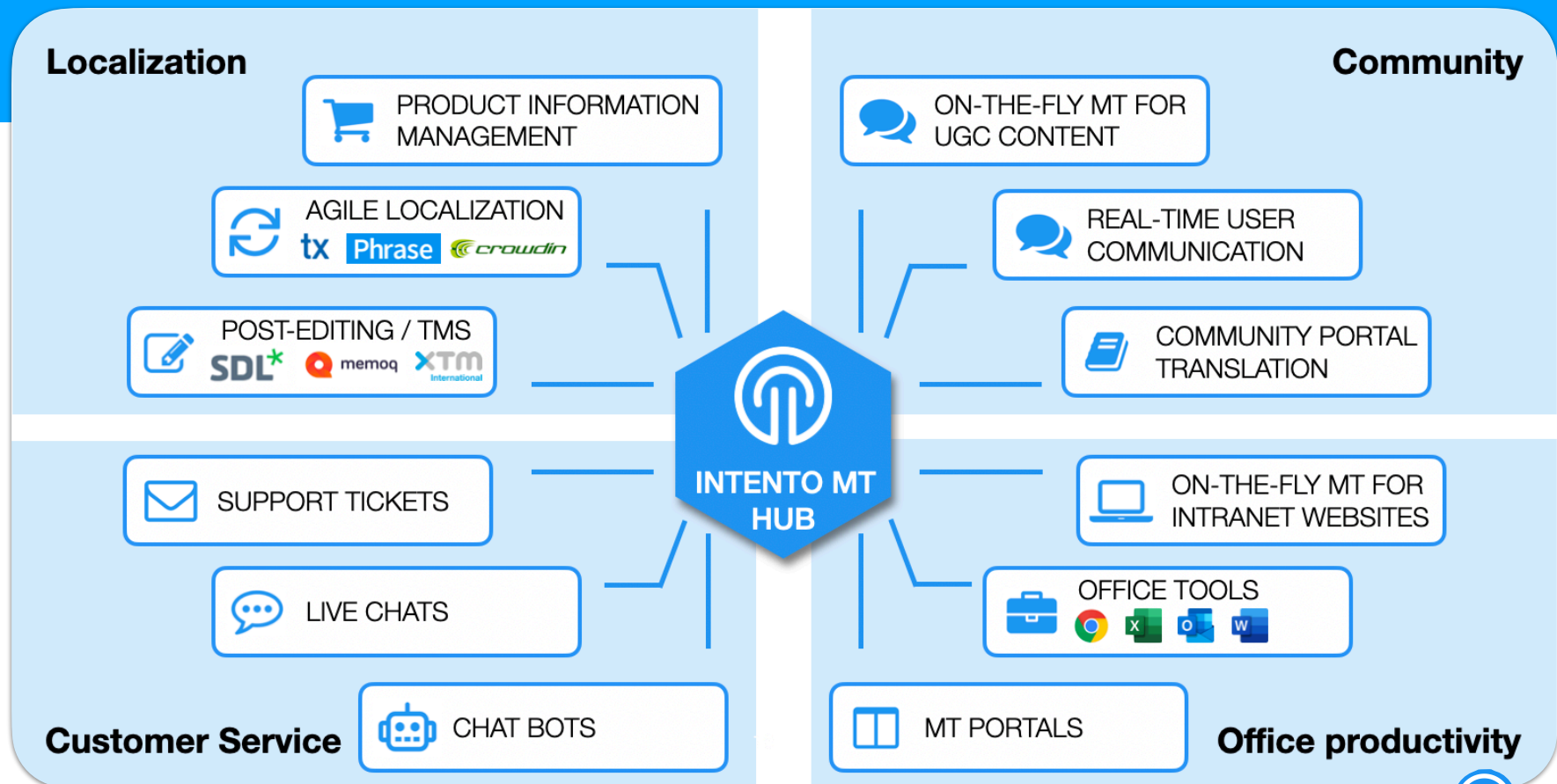Combining resources of multiple stakeholders to benefit everyone

—

MT Requirements beyond the objective linguistic quality

—

Optimizing for features may compromise the quality

10

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page  318

# MT USAGE SCENARIOS AND REQUIREMENTS

11

**Intento**

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 319*

# MULTI-PURPOSE MT

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 320*

# MULTI-PURPOSE MT
## REQUIREMENTS BEYOND QUALITY

large text translation

—

batch translation

—

latency and jitter

—

tolerance to bad source

—

language detection

—

tag support

multilingual source

—

profanity control

—

metadata protection

—

entity protection

—

custom terminology

—

tone of voice consistency

13

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 321*

# ADDITIONAL CHALLENGES
## WITH SPECIFIC COMBINATIONS

large text translation + HTML support

—

source language detection + multilingual source

—

…

Intento

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 322

# MT REQUIREMENTS MATRIX
## EVERY CASE HAS ITS OWN NEEDS

| | large text translation | batch translation | latency and jitter | tolerance to bad source | language detection | tag support | multilingual source | profanity control | metadata protection | entity protection | custom terminology | tone of voice control |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Post-editing / TMS | | ● | | | ● | | | | | ● | | |
| Support tickets | ● | | | | ● | | ● | ● | ● | ● | ● | |
| Live chats | | | ● | ● | ● | | | ● | | ● | ● | ● |
| Chatbots | | | ● | ● | ● | ● | | ● | | ● | ● | ● |
| On-the-fly UGC | | ● | ● | ● | ● | ● | ● | ● | | ● | | ● |
| Real-time communication | | | ● | ● | | | | | | ● | | ● |
| Knowledge bases | ● | | | | ● | | | ● | | ● | | |

Intento

15

# MT REQUIREMENTS MATRIX
## SAMPLE

**ALSO**

**different for inbound and outbound…**

| | large text translation | batch translation | latency and jitter | tolerance to bad source | language detection | tag support | multilingual source | profanity control | metadata protection | entity protection | custom terminology | tone of voice control |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Post-editing / TMS | | ● | | | ● | | | | | | ● | |
| Support tickets | ● | | | | ● | | ● | ● | ● | ● | ● | |
| Live chats | | | ● | ● | ● | | | ● | | ● | ● | ● |
| Chatbots | | | ● | ● | ● | ● | | ● | | ● | ● | ● |
| On-the-fly UGC | | | ● | ● | ● | ● | ● | ● | | | ● | ● |
| Real-time communication | | | ● | ● | | | | | | ● | | ● |
| Knowledge bases | ● | | | | ● | | | | ● | | ● | |

16

16

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 324*

# MT REQUIREMENTS SUPPORT
## BY POPULAR MT ENGINES

● **supported**

○ **support or its quality depends on the language pair / model**

| | large text translation | batch translation | latency and jitter | tolerance to bad source | language detection | tag support | multilingual source | profanity control | metadata protection | entity protection | custom terminology | tone of voice control |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Amazon Translate | ● | | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ● | |
| Google Translate Advanced | ● | ● | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ● | |
| DeepL Pro API | ● | ● | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | | ● |
| IBM Watson Translator | ● | ● | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ● | |
| Microsoft Text Translator | | ● | ○ | ○ | ○ | ○ | ○ | ● | ○ | ○ | ● | |
| ModernMT | | ● | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | | |
| Systran PNMT | ● | ● | ○ | ○ | ○ | ○ | ○ | | ○ | ○ | ● | |

17

17

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 325

# CASE STUDY 1:
# ENTITY PROTECTION

18

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 326*

# ENTITY PROTECTION
## SOME SAMPLES

Simplest cases: protecting email, URLs, phone numbers, file paths

—

Crucial for Customer Service

—

Easily broken by MT

| Source text (English) | Machine Translation |
|---|---|
| I just want to let you know about a spam mail I have received on Friday - it's in D:\Drv\Prt\Epson\Universal driver x64\ABC6\eeecu120m.inf | Я просто хочу уведомить вас о спаме, который я получил в пятницу - он здесь D:\Drv\Prt\Epson\**Универсальный драйвер** x64\ABC6\eeecu120m.inf |
| It has been Ivan Mitrich (ASAP, email.some+plus@example.com.tr) from Belgrad, but in the future it will be me. | Bio je to Ivan Mitrich (ASAP, email.some+plus@**ekample**.com.tr) iz Belgrada, ali u budućnosti to ću biti ja. |
| Would you like to help with a new phone for the ABC department - (772) 194 59 65 ext 4406/4408). | Desideri aiutarti con un nuovo telefono per il dipartimento ABC - (772) 194 59 65 ext **406**/4408). |
| You must submit such a request via ABC-portal, attached link: www.example.com/en/submit | Deve enviar o pedido de tal atraves do ABC-portal, link anexo: www.example.com/**pt**/submit |

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
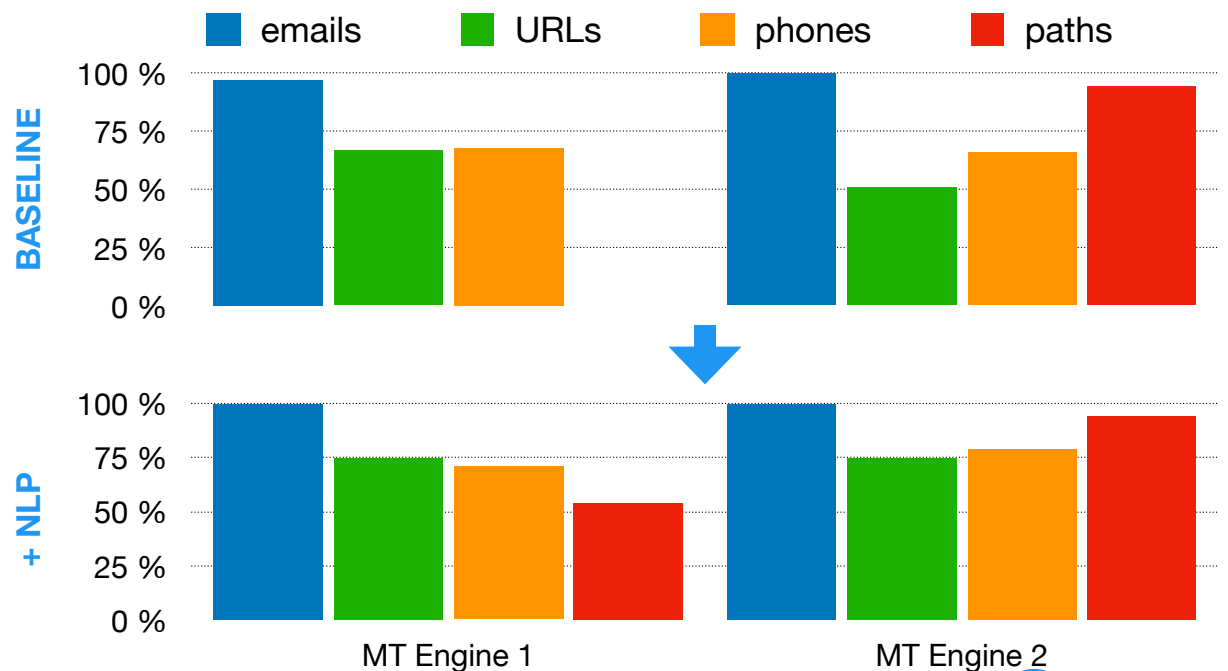*October 6 – 9, 2020, Volume 2: MT User Track*

*Page  327*

# ENTITY PROTECTION
## EXPERIMENTAL RESULTS

Selecting the MT based on the default entity protection may compromise the quality

What if we enforce protection via MT-agnostic NLP?

### ENTITY PROTECTION IN TWO STOCK ENGINES
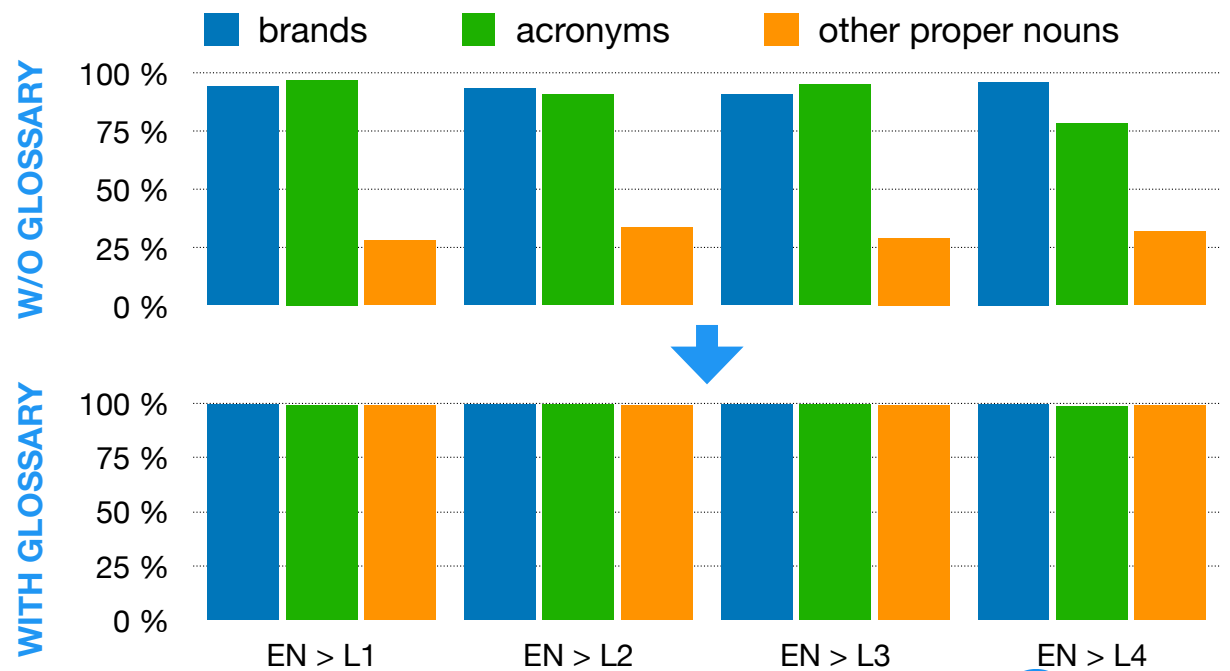


Legend: emails · URLs · phones · paths

BASELINE / + NLP — MT Engine 1, MT Engine 2

20

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 328

# CASE STUDY 2:
# CUSTOM TERMINOLOGY

21

Intento

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 329

# CUSTOM TERMINOLOGY
## IMPROVES FIDELITY

Simplest cases: enforcing acronyms, brand names and other proper nouns.

—

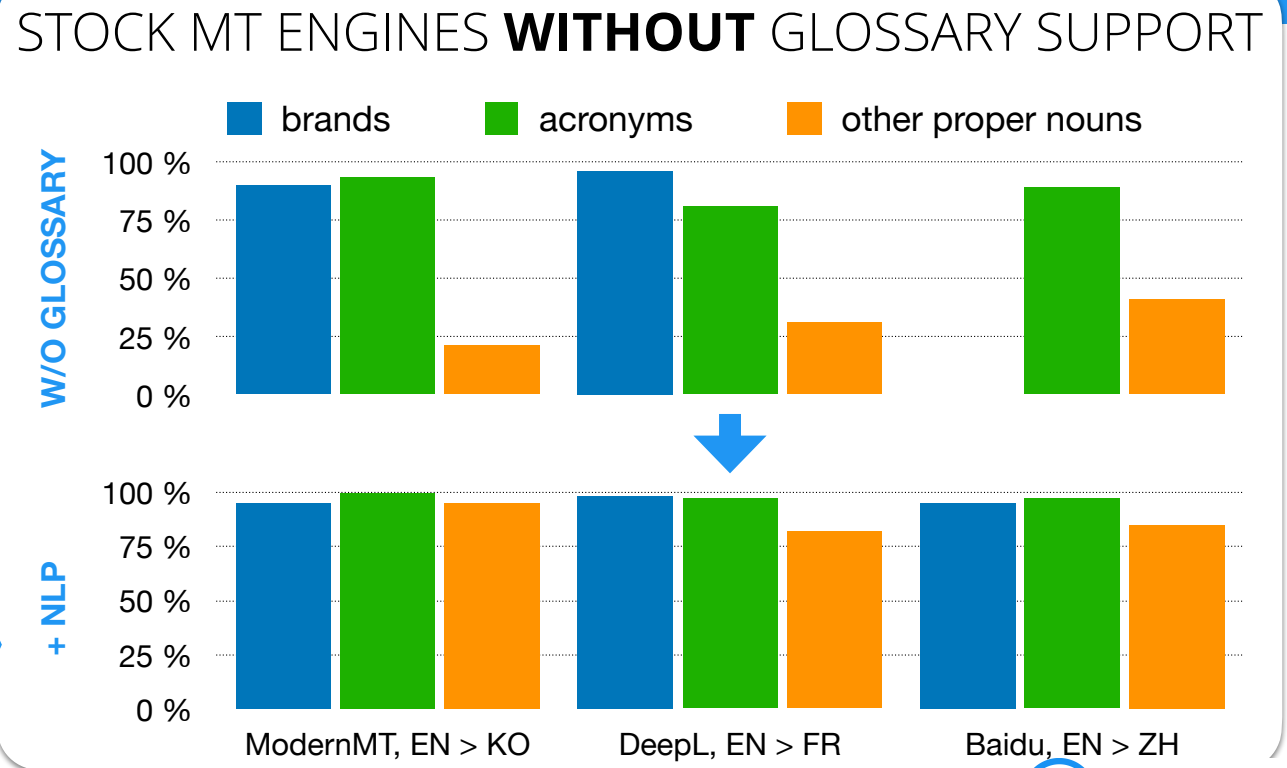Without Custom Terminology support, NMT easily breaks them.



GOOGLE TRANSLATE ADVANCED (WITH GLOSSARY SUPPORT)

- brands
- acronyms
- other proper nouns

W/O GLOSSARY

WITH GLOSSARY

EN > L1   EN > L2   EN > L3   EN > L4

22

Intento

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 330

# CUSTOM TERMINOLOGY
## IMPROVES FIDELITY

Selecting the MT engine by the custom terminology support may compromise MT Quality

MT-agnostic glossary on a top of NMT

### STOCK MT ENGINES **WITHOUT** GLOSSARY SUPPORT

Legend: ■ brands ■ acronyms ■ other proper nouns

W/O GLOSSARY (100%, 75%, 50%, 25%, 0%)

+ NLP (100%, 75%, 50%, 25%, 0%)

ModernMT, EN > KO    DeepL, EN > FR    Baidu, EN > ZH

23

Intento

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 331

# CASE STUDY 3:
# TONE OF VOICE CONTROL

24

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 332*

# TONE OF VOICE CONTROL
## SAMPLES FROM SUPPORT CHATS

Formal vs. Informal

—

Crucial for Live Chats

—

Baseline MT engines are not consistent

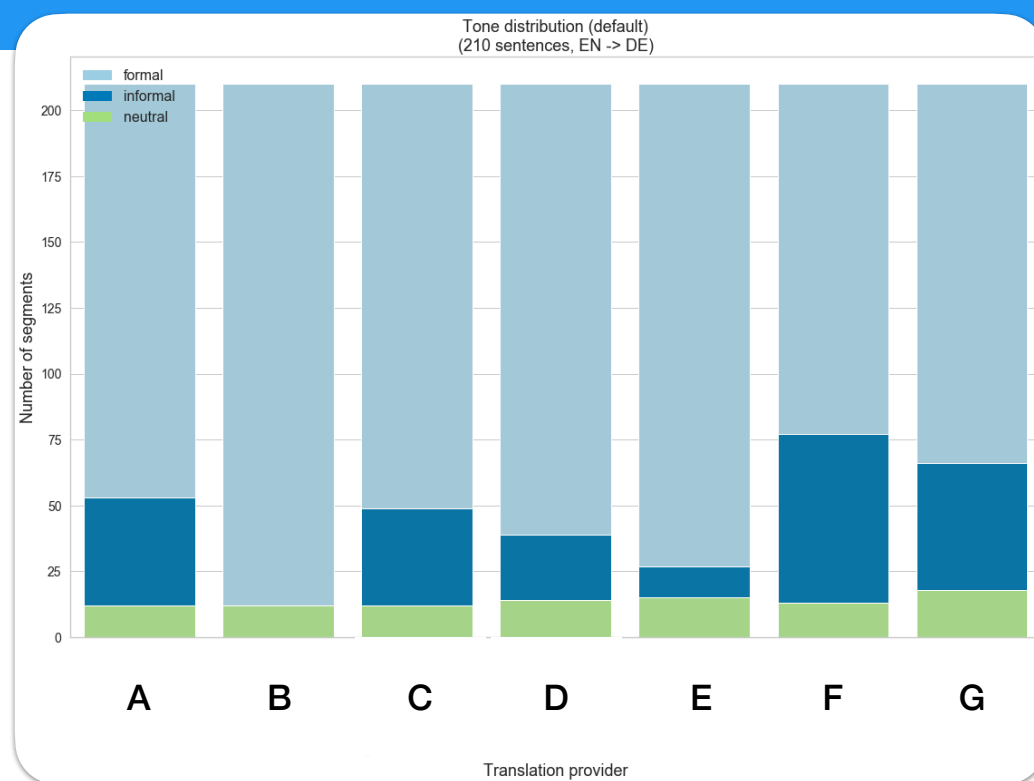| Source text (English) | Machine Translation (German) | COMMENT |
|---|---|---|
| Can you share your screen? | Können Sie Ihren Bildschirm freigeben? | FORMAL |
| Could you help me? | Kannst du mir helfen? | INFORMAL |
| Make sure you report any of these issues. | Stellen Sie sicher, dass Sie eines dieser Probleme melden. | FORMAL |
| Can you give an example? | Kannst du ein Beispiel geben? | INFORMAL |

25

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 - 9, 2020, Volume 2: MT User Track*

*Page 333*

# TONE OF VOICE CONTROL
## DEFAULT MT OUTPUT

English to German

—

210 segments

—

stock models



Tone distribution (default)
(210 sentences, EN -> DE)

26

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 334*

# TONE OF VOICE CONTROL
## HOW TO MAKE IT INFORMAL?

**Option 1:** Use DeepL with formality=less (99.5% accuracy)

—

**Option 2:** Generate synthetic training data, hoping translations become more informal

—

**Option 3:** MT-agnostic NLP

What if you need a custom model and terminology, or another MT has better linguistic quality for you?

Expensive and time-consuming, also introduces bias into the model

Works to a certain extent, provides a wider choice of MT engines

27

Intento

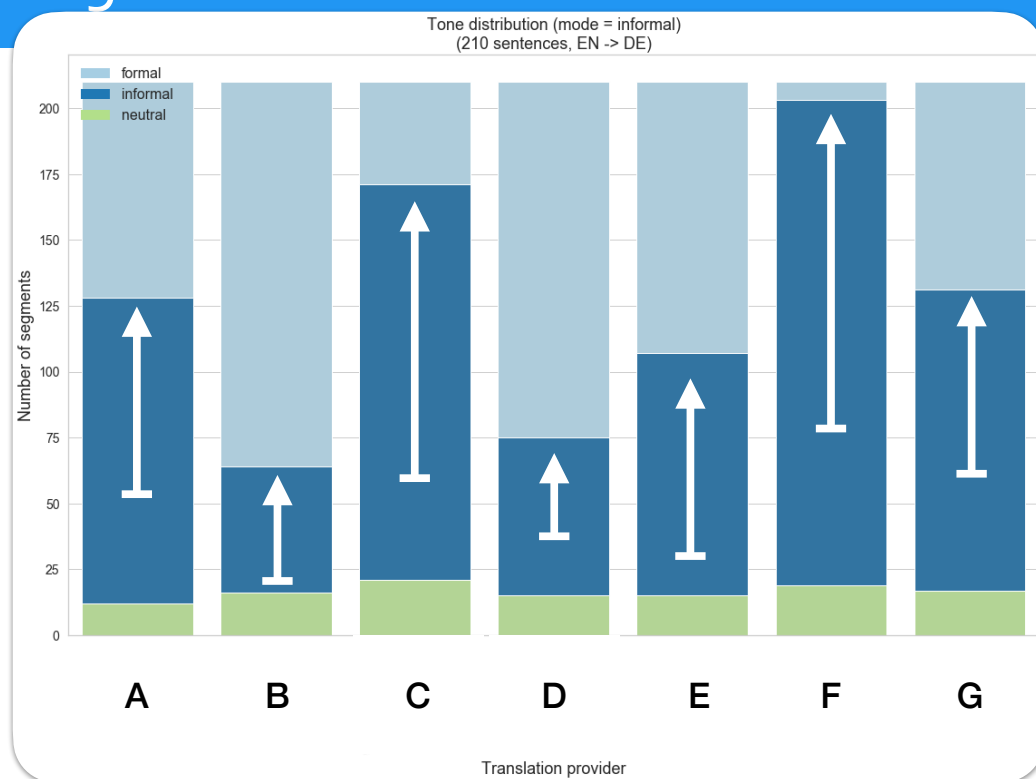*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 335*

# TONE OF VOICE CONTROL
## MT-AGNOSTIC ADJUSTMENT

English to German

—

210 segments

—

stock models

let's make it more
**INFORMAL**

Tone distribution (mode = informal)
(210 sentences, EN -> DE)

legend:
- formal
- informal
- neutral

Number of segments

A  B  C  D  E  F  G

Translation provider

28

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page  336*

# TONE OF VOICE CONTROL
## MT-AGNOSTIC ADJUSTMENT

English to German

—

210 segments

—
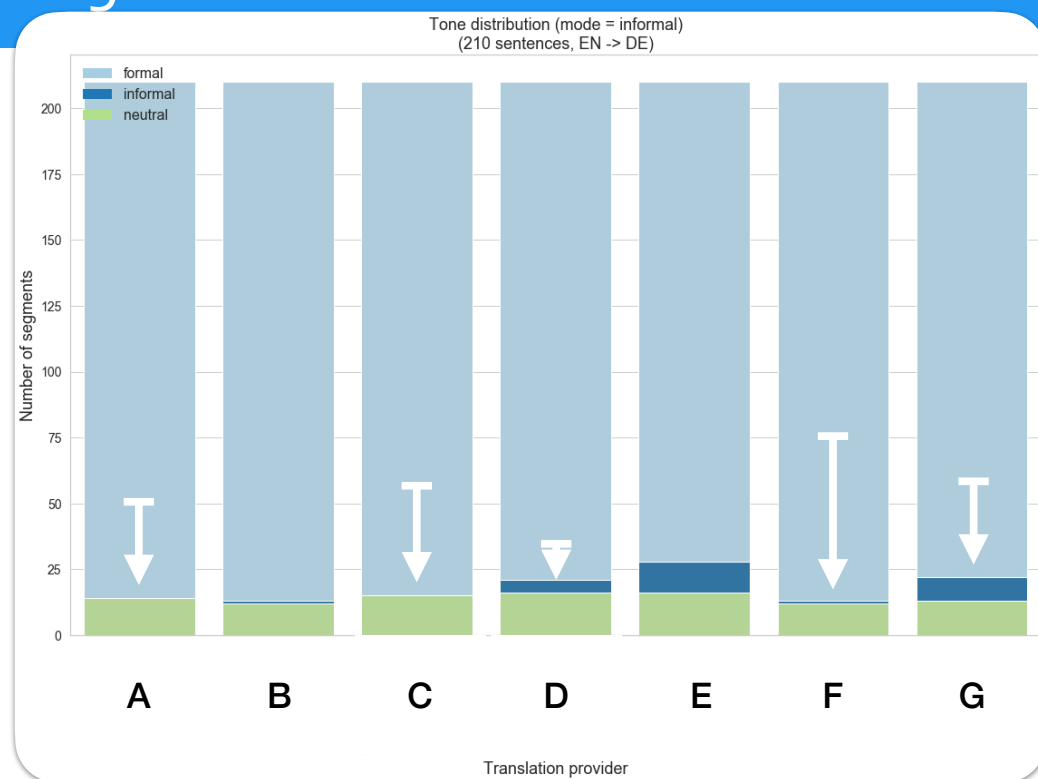
stock models

let's make it more **FORMAL**



Tone distribution (mode = informal)
(210 sentences, EN -> DE)

29

# KEY TAKEAWAYS

Multi-Purpose MT brings an instant ROI on the MT investment already made.

—

Different use cases impose multiple requirements beyond the linguistic quality.

—

Meeting the requirements takes either the right MT engine choice or clever engineering.

—

We do both, by implementing MT-agnostic fine tuning algorithms to avoid compromising the MT quality.

Intento

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 338*

**AMTA VIRTUAL 2020**

# THANKS!
## ks@inten.to

**INTENTO**
https://inten.to

Konstantin Savenkov, CEO
ks@inten.to
2150 Shattuck Ave
Berkeley CA 94705

31

*Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*
*October 6 – 9, 2020, Volume 2: MT User Track*

*Page 339*