



universität
wien

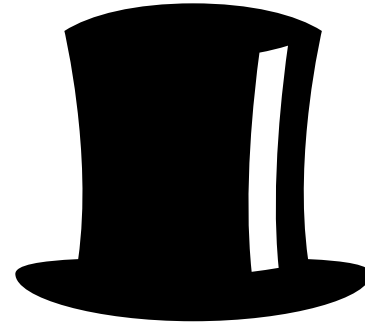
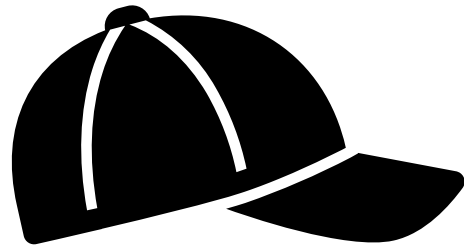
Post-editing for translators – the theory and practice

Claudia Wiesinger

**Human and Artificial Intelligence in
Translation (HAITrans)**

University of Vienna

Two hats



Post-editing machine translation: two sides of the same coin?



theory

practice

slido



**What do you associate with the term
"post-editing"?**

① Start presenting to display the poll results on this slide.
CIOL Translators Day 2024 - Claudia Wiesinger



Three common assumptions

- 1. Machine translation (MT)** is or will soon be just as good as human translators

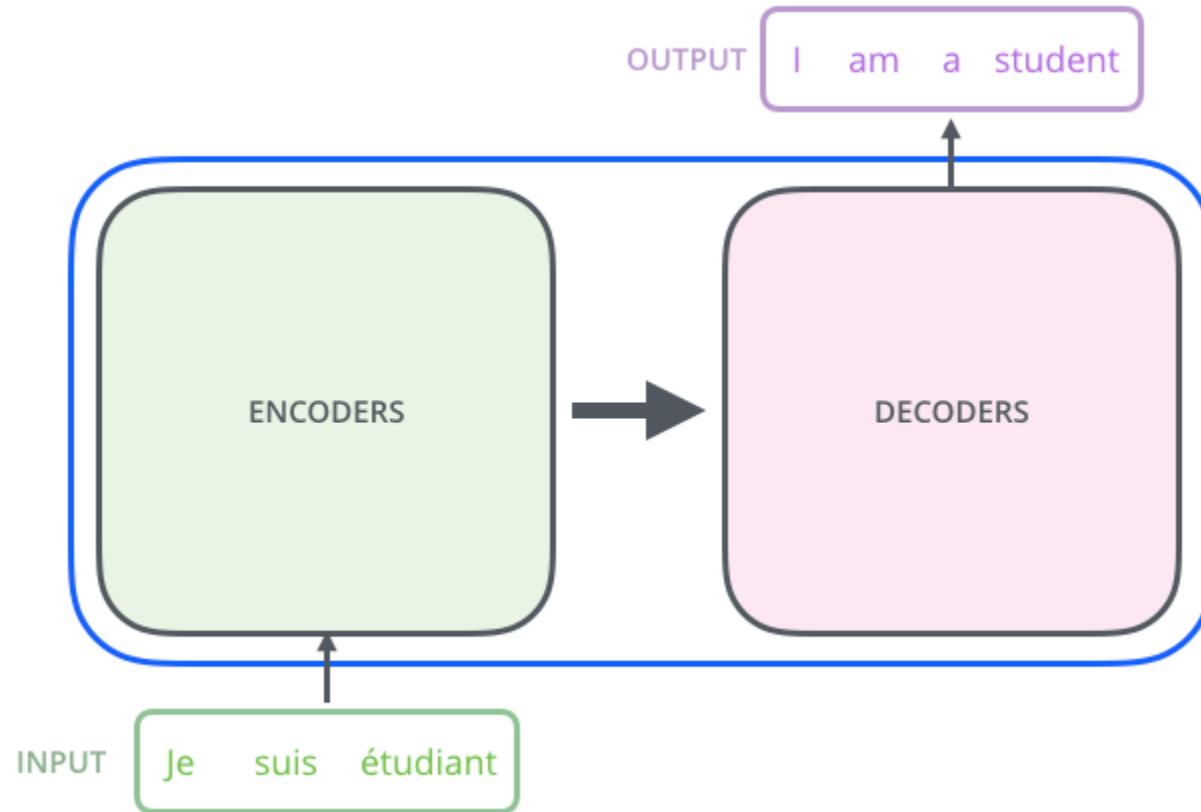
Will Machine Learning AI Make Human Translators An Endangered Species?



Bernard Marr Contributor 
Enterprise & Cloud

-  Translating between human languages is something which artificial intelligence – specifically machine learning – has proven to be very competent at.
-  So much so that the CEO of one of the world's largest employers of human translators has warned that many of them should be facing up to the stark reality of losing their job to a machine.

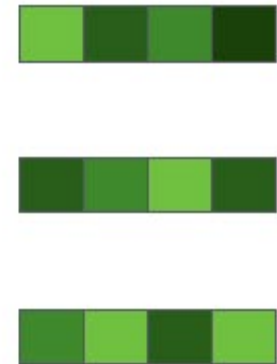


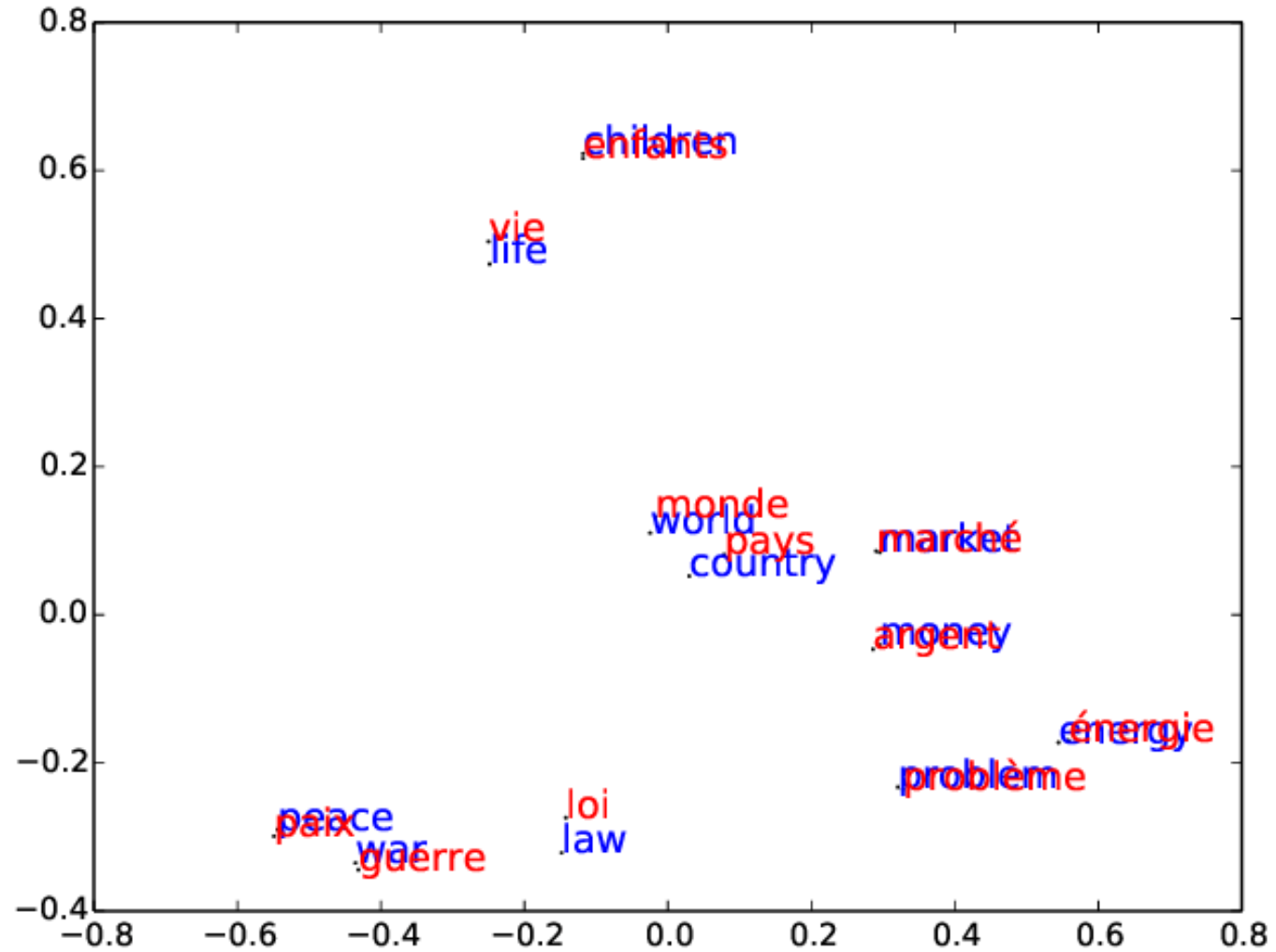


Input

Je
suis
étudiant

0.901	-0.651	-0.194	-0.822
-0.351	0.123	0.435	-0.200
0.081	0.458	-0.400	0.480







Input Sentence

ich glaube aber auch, er ist clever genug um seine Aussagen vage genug zu halten, so dass sie auf verschiedene Art und Weise interpretiert werden können.

Output Word Predictions

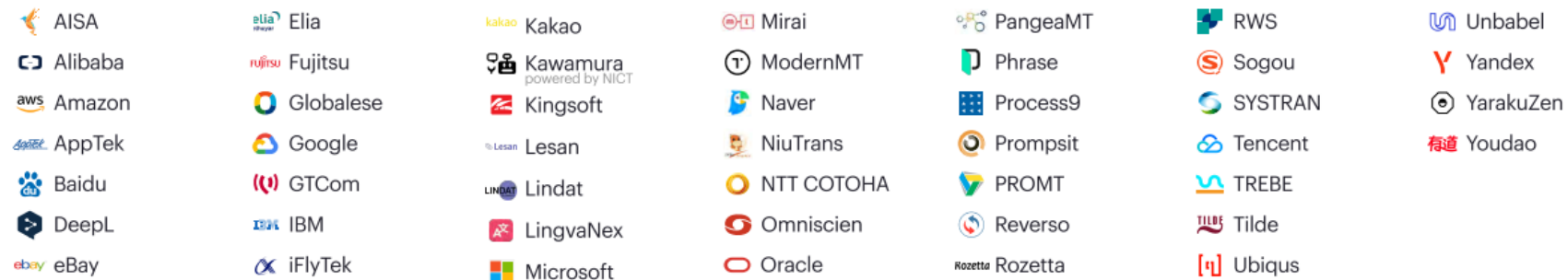
Best		Alternatives
but	(42.1%)	<i>however (25.3%), I (20.4%), yet (1.9%), and (0.8%), nor (0.8%), ...</i>
I	(80.4%)	<i>also (6.0%), , (4.7%), it (1.2%), in (0.7%), nor (0.5%), he (0.4%), ...</i>
also	(85.2%)	<i>think (4.2%), do (3.1%), believe (2.9%), , (0.8%), too (0.5%), ...</i>
believe	(68.4%)	<i>think (28.6%), feel (1.6%), do (0.8%), ...</i>
he	(90.4%)	<i>that (6.7%), it (2.2%), him (0.2%), ...</i>
is	(74.7%)	<i>'s (24.4%), has (0.3%), was (0.1%), ...</i>
clever	(99.1%)	<i>smart (0.6%), ...</i>
enough	(99.9%)	
to	(95.5%)	<i>about (1.2%), for (1.1%), in (1.0%), of (0.3%), around (0.1%), ...</i>
keep	(69.8%)	<i>maintain (4.5%), hold (4.4%), be (4.2%), have (1.1%), make (1.0%), ...</i>
his	(86.2%)	<i>its (2.1%), statements (1.5%), what (1.0%), out (0.6%), the (0.6%), ...</i>
statements	(91.9%)	<i>testimony (1.5%), messages (0.7%), comments (0.6%), ...</i>
vague	(96.2%)	<i>v@@ (1.2%), in (0.6%), ambiguous (0.3%), ...</i>
enough	(98.9%)	<i>and (0.2%), ...</i>
so	(51.1%)	<i>, (44.3%), to (1.2%), in (0.6%), and (0.5%), just (0.2%), that (0.2%), ...</i>
they	(55.2%)	<i>that (35.3%), it (2.5%), can (1.6%), you (0.8%), we (0.4%), to (0.3%), ...</i>
can	(93.2%)	<i>may (2.7%), could (1.6%), are (0.8%), will (0.6%), might (0.5%), ...</i>
be	(98.4%)	<i>have (0.3%), interpret (0.2%), get (0.2%), ...</i>
interpreted	(99.1%)	<i>interpre@@ (0.1%), constru@@ (0.1%), ...</i>
in	(96.5%)	<i>on (0.9%), differently (0.5%), as (0.3%), to (0.2%), for (0.2%), by (0.1%), ...</i>
different	(41.5%)	<i>a (25.2%), various (22.7%), several (3.6%), ways (2.4%), some (1.7%), ...</i>
ways	(99.3%)	<i>way (0.2%), manner (0.2%), ...</i>
.	(99.2%)	<i></s> (0.2%), , (0.1%), ...</i>
</s>	(100.0%)	

There's no “best” MT system

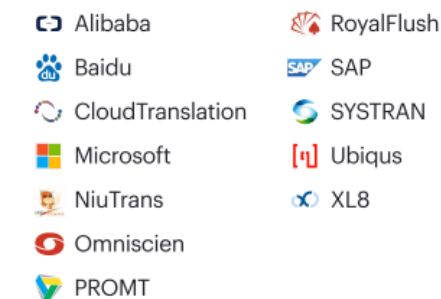
MT performance depends on how similar your data is to the data used to train the vendors' models, as well as their algorithms.

1.1 Machine Translation Landscape

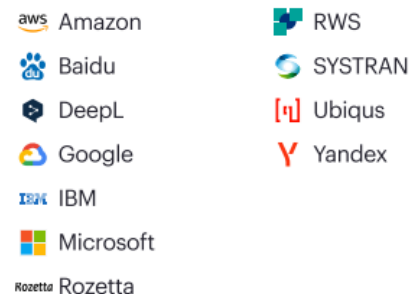
Generic stock models



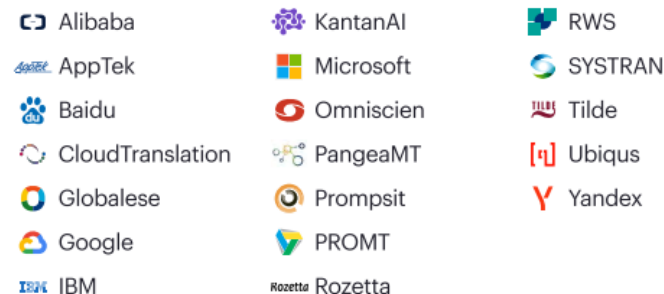
Vertical Stock Models



Custom terminology support



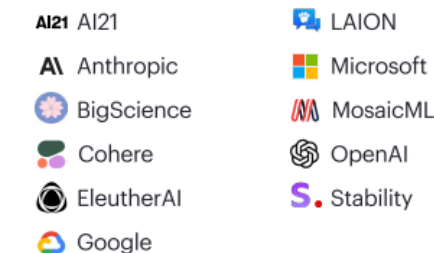
Static domain adaptation



Dynamic domain adaptation



Large Language Models



Standalone commercial products with an API. All product names, trademarks and registered trademarks are property of their respective owners. All company, product and service names used in this material are for identification purposes only. Use of these names, trademarks and brands does not imply endorsement.

Proceed with...



CAUTION!

Adequacy vs. fluency in neural MT

“Compared to statistical machine translation, neural machine translation may be **less adequate**, even if **more fluent**. In other words, the translation may **diverge from the input** in various ways, such as not translating part of the sentence or generating unrelated output words.”

Things to look out for – Mistranslations



Things to look out for – Hallucinations

the Ministry of Energy, Natural Resources, Environment and Mines (MiAmbiente), the Forest Conservation Institute (ICF), the Institute of Property (IP), the National Agrarian Institute (INA), the Revenue Administration Service (SAR), the Customs Revenue Subdirectorate (DARA), the Association of Honduran Municipalities (AMHON), the Honduran Social Security Institute, (IHSS), the Ministry of Labour and Social Security (STSS), the Ministry of Agriculture and Livestock (SAG), the National Directorate of Indigenous and Afro Honduran Peoples (DINAFROH), the Institute for Access to Public Information (IAIP), the Public Prosecutor's Office (MP), the Security Ministry (SSN, representing the National Preventive Police (PNP)), the National Defence Ministry (SEDENA, representing the armed forces of Honduras (FFAA)), the Attorney General's Office (PGR), the Central Bank of Honduras (BCH), the National School of Forest Sciences (ESNACIFOR) and other academic institutions concerned with environmental and forestry issues.

Ministerul Energiei, Resurselor Naturale, Mediului și Minelor (MICF), Institutul pentru Conservarea Naturii (ICF), Institutul National Agrare (IP), Serviciul de administrare fiscală (SAR), Subdirecția de Administrare Fiscală (AMHON), Ministerul Muncii și Securității Sociale (SAG), Ministerul Muncii și Securității Sociale (SAG), Ministerul Muncii și Securității Sociale (SAG), Ministerul Muncii și Securității Sociale (SAG), Ministerul Muncii și Securității Sociale (SAG), Ministerul Muncii și Securității Sociale (SAG), Ministerul Muncii și Securității Sociale (SSN), Ministerul Muncii și Securității Sociale (SSN, reprezentant al forțelor armate ale Honduras — FFAA), Biroul Procurorului General (PGR), Central Bank of Honduras (CSIB), Școala Națională de Științe Forestiere (ESNAF), Școala Națională de Științe Forestiere (ESNAF) și alte instituții academice care se ocupă de probleme de mediu și silvice.

Things to look out for – Distortions



STIFT KLOSTER NEUBURG DE Q MENU

Willkommen im Stift Klosterneuburg


AUSSTELLUNG

Was leid tut

Ab 1. Juli 2020
Sala terrena Galerie, Stift Klosterneuburg

Aufgrund der von der **Bundesregierung** erlassenen aktuellen COVID-19-Maßnahmenverordnung bleibt das Museum von 3. bis 30. November geschlossen. Wir bitten um Ihr Verständnis und freuen uns auf ein Wiedersehen!

GERMAN ENGLISH X



ft-klosterneuburg.at 68

STIFT KLOSTER NEUBURG DE Q MENU

Welcome to Klosterneuburg Abbey

EXHIBITION

What sorry

From July 1, 2020
Sala terrena Galerie, Klosterneuburg Abbey

Due to the current COVID-19 regulation issued by the **German government**, the museum will be closed from November 3 to 30. We ask for your understanding and look forward to seeing you again!

GERMAN ENGLISH X

Things to look out for – Loss of coherence

Text

Bilder

Dokumente

Websites

Englisch – erkannt Deutsch Englisch Französisch

Images must not contradict the landing page and must be of high resolution and quality and aesthetically pleasing. Images must not:

- Be individual or multiple product images on a solid or transparent background.
- Be an individual brand logo or combination of logos.
- Be one of your selected product images.
- Contain crowded, excessive, poorly cropped or illegible elements.
- Contain additional text other than text naturally present in the image (such as on product packaging).
- Contain letterbox or pillarbox formats.



531 / 5.000

Deutsch Englisch Französisch

Bilder dürfen nicht im Widerspruch zur Landingpage stehen und müssen eine hohe Auflösung, Qualität und ein ästhetisches Aussehen

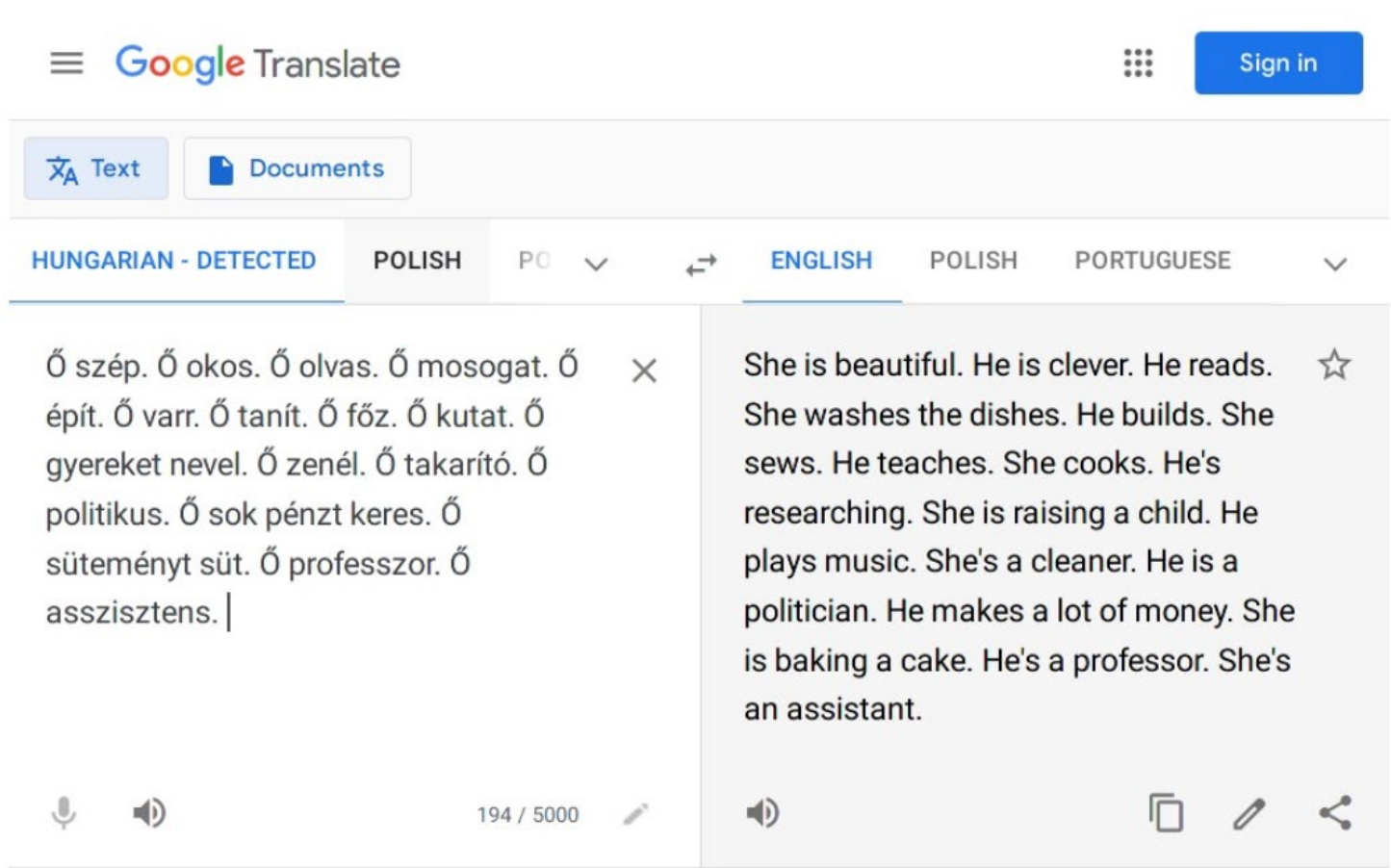
haben. Bilder dürfen nicht:

- Seien Sie einzelne oder mehrere Produktbilder auf einem einfarbigen oder transparenten Hintergrund.
- Seien Sie ein einzelnes Markenlogo oder eine Kombination von Logos.
- Seien Sie eines Ihrer ausgewählten Produktbilder.
- überfüllte, übertriebene, schlecht beschnittene oder unleserliche Elemente enthalten.
- Zusätzlichen Text enthalten, der nicht natürlich im Bild vorhanden ist (z. B. auf der Produktverpackung).
- Enthalten Letterbox- oder Pillarbox-Formate.



Feedback geben

Things to look out for – Gender bias



The screenshot shows the Google Translate interface. The source language is detected as HUNGARIAN, and the target language is set to ENGLISH. The source text is: "Ő szép. Ő okos. Ő olvas. Ő mosogat. Ő épít. Ő varr. Ő tanít. Ő főz. Ő kutat. Ő gyereket nevel. Ő zenél. Ő takarító. Ő politikus. Ő sok pénzt keres. Ő süteményt süt. Ő professzor. Ő asszisztens." The translated text is: "She is beautiful. He is clever. He reads. She washes the dishes. He builds. She sews. He teaches. She cooks. He's researching. She is raising a child. He plays music. She's a cleaner. He is a politician. He makes a lot of money. She is baking a cake. He's a professor. She's an assistant." The translation consistently uses "she" for female-coded activities and "he" for male-coded activities, illustrating gender bias.

With many thanks to
Alexandru-Valentin Timaru

Things to look out for – Inaccuracies (names)

Iryna Skliar

Iryna S**g**liar

James R Elliot

James R Elliott

He, H. and Ye, Z.

Er, H. and **Ya**, Z.

Roland W Force

Roland W **Macht**

Roland W **Kraft**

Roland W **Streitkräfte**;

Michael P Lesser

Michael **S.** Lesser



Tosepu, Ramadhan

Tosepu, **RAMADAN**

Fredrik Palm

Friedrich **von Schiller**

Things to look out for – Inaccuracies (numbers)

1674364114		Secs/Word: 06" PEE: 0%	Segment status APPROVED
Source	Gefördert wurden rund 7.800 Pkws mit reinem Elektroantrieb, knapp 1.100 Plug-In-Hybride und über 1.500 E-Motorräder und E-Mopeds.		Words: 17
Suggestion			
Translation	 Around 7.800 have been supported with electric cars, almost 100 Plug-In-Hybride and more than 1,500 E-Motorräder and E-Mopeds.		TTE: 24"
Revision	 Around 7.800 cars have been supported with electric cars drive, almost 100 Plug-In-Hybride and more than 1,500 E-Me-motorräderbikes and E-Me-mopeds.		TTE: 01'34"
QA	Human (1)	Terminology and translation consistency: [Neutral]	

eTranslation

commission.europa.eu/resources-partners/etranslation_en

eTranslation

The European Commission's Machine Translation system

PAGE CONTENTS

- [Translate online](#)
- [Product features](#)
- [Raw quality](#)
- [Access](#)
- [Privacy statement](#)
- [How to use it](#)
- [Help / contact](#)
- [Related links](#)

Translate online

eTranslation is a cutting-edge neural machine translation service provided by the European Commission. It was officially launched on 15 November 2017, superseding the earlier statistical system, MT@EC.

The [web page](#) is free to use for EU institutions, public administrations, universities, EU freelance translators, SMEs, European NGOs and projects financed by the [Digital Europe Programme](#), located in an EU country, Iceland, Norway, Liechtenstein or Ukraine. **Registration is required** (see **Access** below). eTranslation also offers a web service API for machine-to-machine connections from eligible parties.

Product features

- High security - all data processed by the system stays within the Commission's firewalls and can't be seen by outsiders
- Translates from and into any [official EU language](#) as well as Arabic, Chinese, Icelandic, Japanese, Norwegian, Russian, Turkish and Ukrainian
- Offers several styles of translation ("domains") in the fields of EU Formal Language, General Text, Finance and a few more
- Supports all common office formats, including pdf
- Can provide tmx or xliif files for professional translators

Raw quality

eTranslation produces raw machine translations. Use it to get the gist of a text or as the starting point for a human-quality translation. If you need a perfectly accurate, high-quality translation, the text still needs to be revised by a skilled professional translator.

Three common assumptions

1. Machine translation (MT) is or will soon be just as good as human translators
2. Machine translation is the same as translation

Syntactic strategies

G1	Literal Translation
G2	Loan
G3	Word structure change
G4	Unit shift
G5	Phrase structure change
G6a	Clause structure change
G6b	Clause structure change
G7	Sentence structure change
G9	Modality shift
G10	Scheme change
G11	Addition
G12	Omission
G13	Punctuation change

Semantic strategies

S1	Semantically Equivalent
S2	Antonymy
S3	Hyponymy
S4	Converses
S5	Abstraction change
S6	Distribution change
S7	Emphasis change
S8	Paraphrase
S9	Trope change
S10	Cohesion change

Pragmatic strategies

Pr1a	Domestication
Pr1b	Foreignization
Pr2a	Explicitation
Pr2b	Implication
Pr3a	Information addition
Pr3b	Information omission
Pr4	Interpersonal change
Pr5	Illocutionary change
Pr6	Coherence change
Pr9	Transediting
Pr11	No change
Pr12	Domain adaptation
Pr13	External Information adaptation

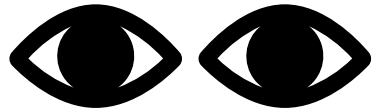
Three common assumptions

1. Machine translation (MT) is or will soon be just as good as human translators
2. Machine translation is the same as translation
3. **Post-editing** is a universally understood concept & set of practices

Post-editing is...

...the editing and correcting machine translation output (ISO 18587:2017)

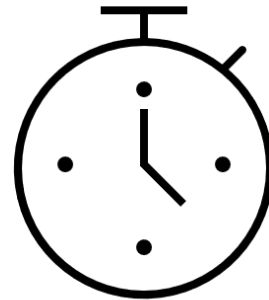
- **Full** post-editing or **light** post-editing?
- How is post-editing different from **revision**?



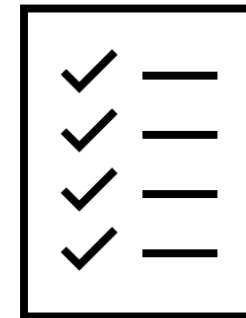
Post-editing skills



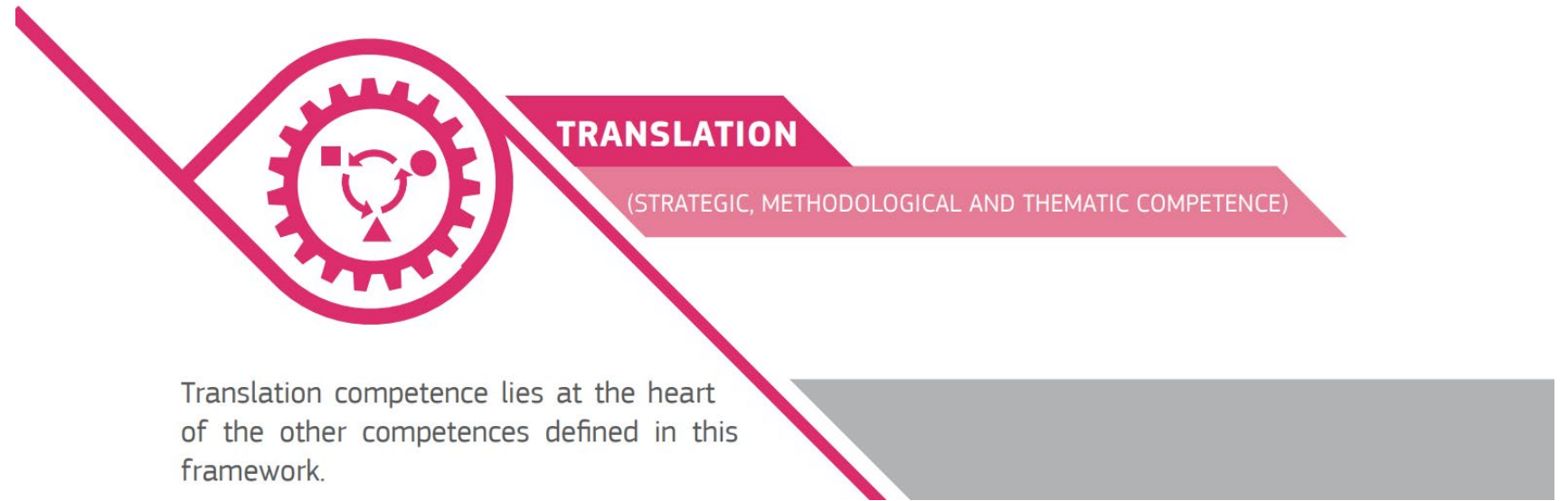
Identifying errors



Reasonable speed



Guidelines

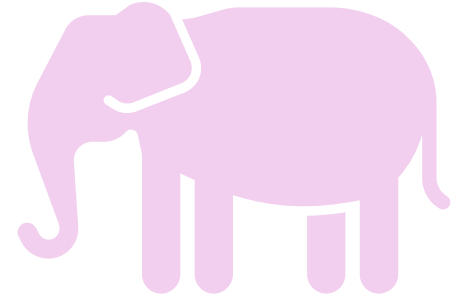


14

Post-edit MT output using style guides and terminology glossaries to maintain quality standards in MT-enhanced translation projects

A disclaimer

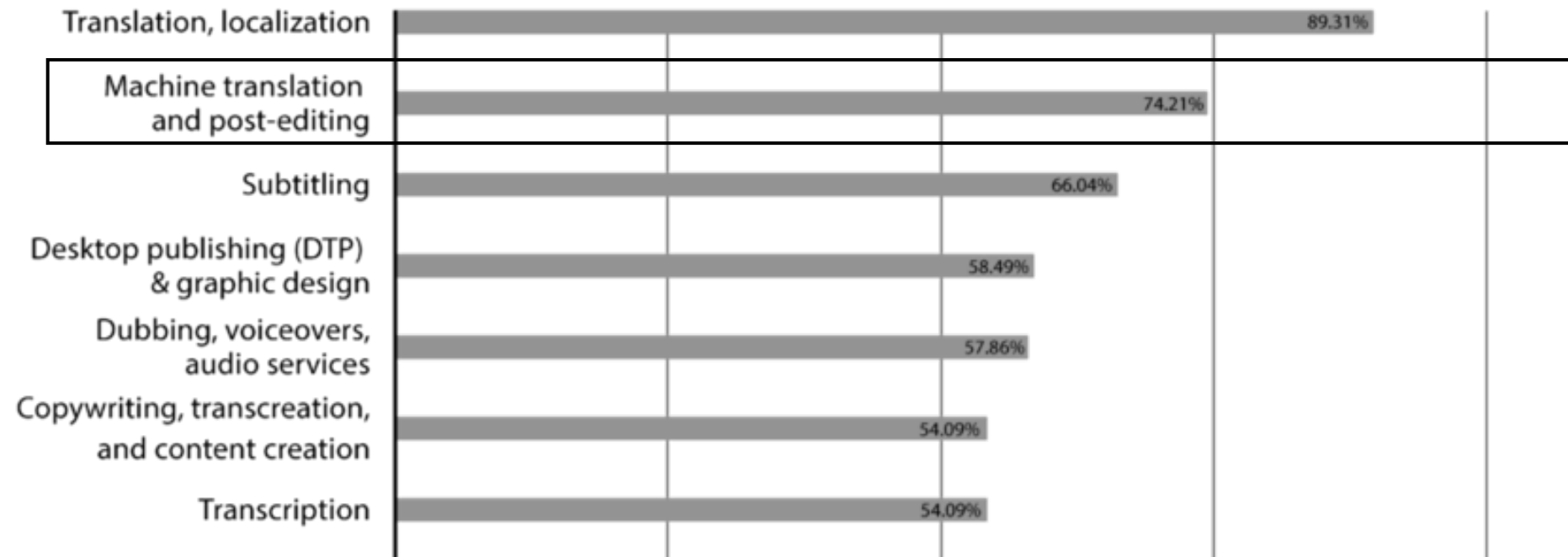




The priming effect and post-editeese

“we have found that [post-edited texts] tend to be **simpler** and more normalised and to have a **higher degree of interference** from the source text than [human translations]. This seems to be caused because these characteristics are already present in the MT outputs that are the starting point of the [post-edited texts]”

Machine translation and post-editing are popular services



% of language service providers offering this service

And yet...

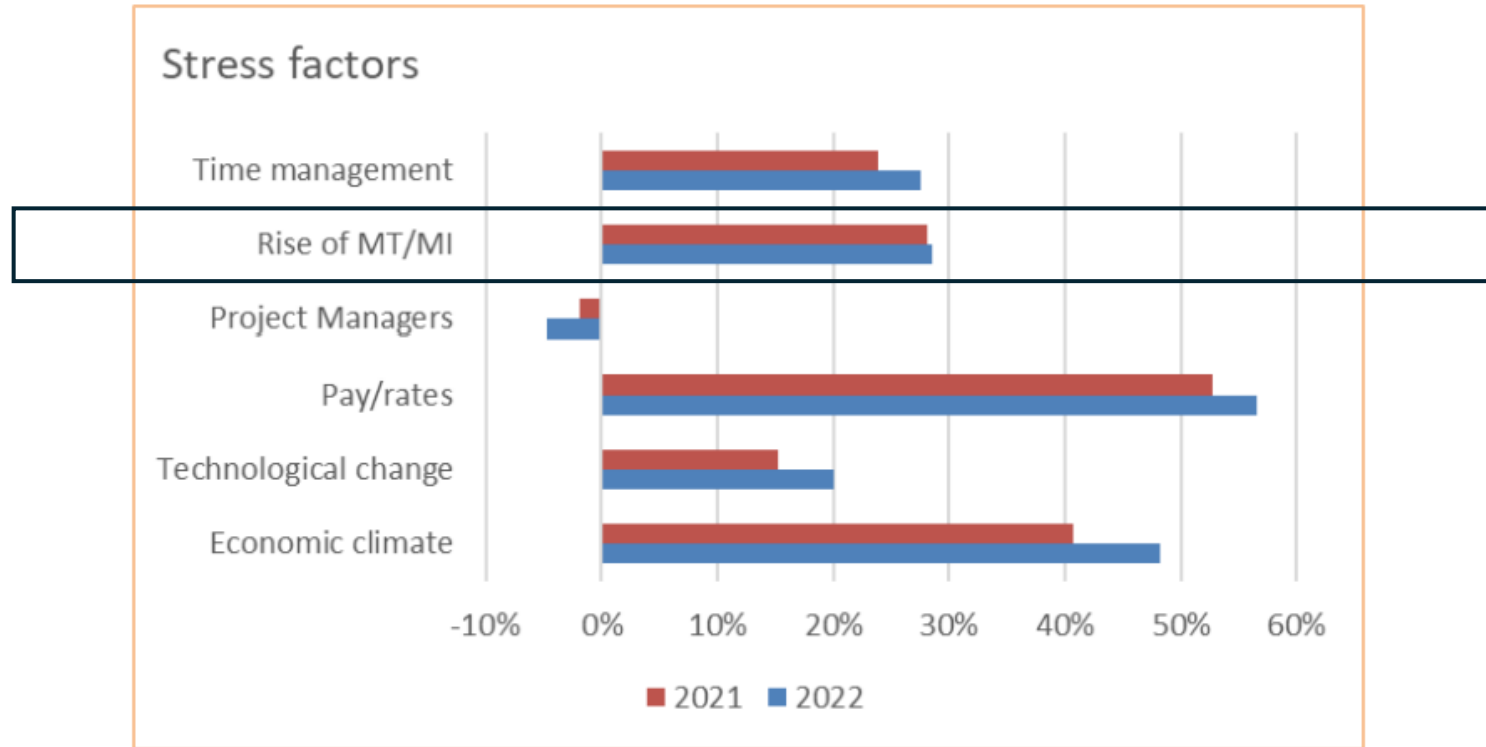


Figure 37 - Stress factors independent professionals

And yet...

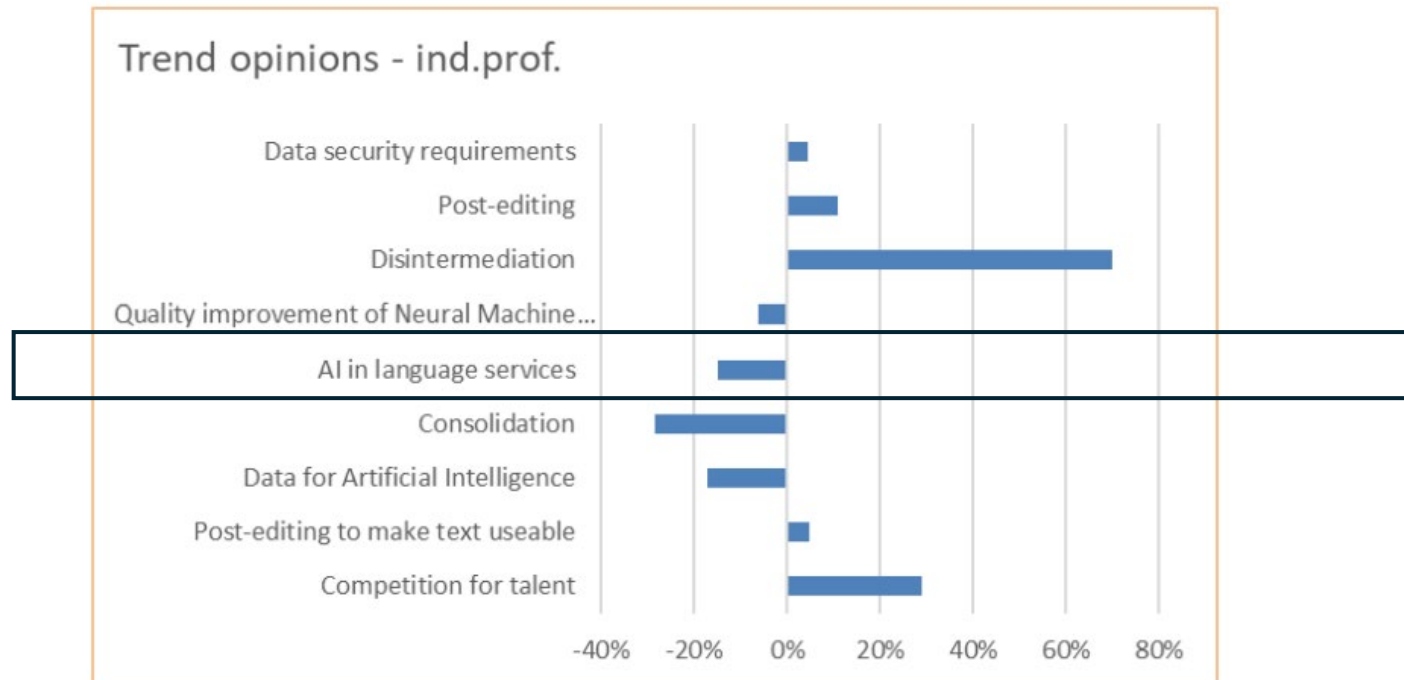


Figure 36 - Trend opinions - independent professionals

Workshop



MT post-editing skills workshop (English as the target language)

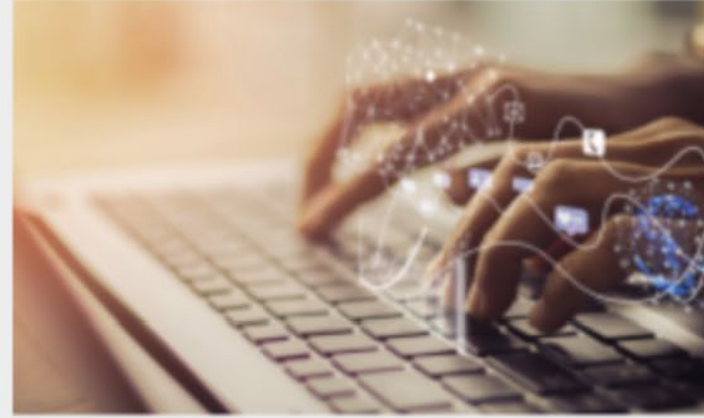
Friday 26 April 2024

Delivered by Prof. Dragos Ciobanu and HAITrans

9am - 12pm UK time

[Click here to find out more](#)

Workshop



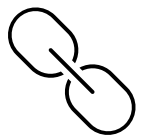
MT post-editing skills workshop (English as the source language)

Friday 26 April 2024

Delivered by Prof. Dragos Ciobanu and HAITrans

1pm - 4pm UK time

[Click here to find out more](#)



<https://www.ciol.org.uk/online-workshops-focus-sessions>

By way of conclusion

- MT has got big backers and is constantly marketed as “the future of translation”
- There are a lot of misconceptions when it comes to MT
- MT can be genuinely useful (and occasionally catastrophic)
- Post-editing has its own ISO standard (ISO 18587:2017)
- MT/AI is on the rise, but perceived as a threat by many language professionals

Passing the mic to you



A decorative background on the left side of the slide featuring a dense pattern of tropical leaves in various shades of blue, teal, and green. The leaves have different textures, including solid colors, white outlines, and patterns of dots or stripes.

Many thanks!

Claudia Wiesinger

claudia.wiesinger@univie.ac.at

HAITrans research group

<https://haitrans.univie.ac.at/>