

POSTĘDYCJA – FAKTY I MITY

MARTA BARTNICKA



CZYM JEST POSTEDYCJA?

Według ISO 18587:

post-edit

edit and correct machine translation output

full post-editing

process of post-editing to obtain a product comparable to a product obtained by human translation

Według słowniczka z Localization.pl:

Postedycja: *poprawianie przez tłumacza odpowiedzi maszynowych, czyli propozycji tłumaczenia wygenerowanych przez tłumaczenie maszynowe. Celem postedycji jest sprawdzenie, czy tłumaczenie oddaje sens oryginału, oraz poprawienie stylu i zgodności segmentu z resztą tekstu.*

1. Do postedycji nie trzeba patrzeć w oryginał, wystarczy wygładzić tłumaczenie.
2. Postedycja wyklucza ochronę praw autorskich, danych osobowych oraz *ius primae noctis*.
3. Jakość MT dla tekstów prawnych przekracza jakość tłumaczenia profesjonalnego.
4. Każdy tłumacz musi nauczyć się postedycji, żeby nie wypaść z obiegu.



1. DO POSTEDYCJI NIE TRZEBA PATRZEĆ W ORYGINAŁ, WYSTARCZY WYGŁADZIĆ TŁUMACZENIE? NIE!

- Dobra jakość usypia czujność (dużo czytania, mało poprawiania)
- Silniki MT nie są przewidywalne
- MT może wymyślać słowa, jąkać się, pomijać/dodawać wyrazy
- Kontekst jest ograniczony do zdania/frazy (ew. akapitu)
- Jeśli w korpusie nie ma odpowiedniej terminologii, będzie użyta błędna
- **Maszyna nadal nic nie rozumie!**

“The poisoned cookie now even tastes good”
(Alex Waibel, Carnegie Mellon University / Karlsruhe Institute of Technology)

For the test to see if the study drug has been absorbed, your child will have a blood sample taken before the morning dose of study drug, as well as a second blood sample 2.5 hours after he or she has been given the study drug.

Aby sprawdzić, czy badany lek został wchłonięty, dziecko otrzyma próbkę krwi pobraną przed poranną dawką badanego leku, jak również drugą próbkę krwi 2,5 godziny po podaniu badanego leku.

MECHANIKA

- **Neural Machine Translation** wykorzystuje sztuczną sieć neuronową do przewidywania prawdopodobieństwa wystąpienia sekwencji słów, zwykle modelując całe zdania w jednym zintegrowanym modelu.
- **Korpus** to dane dwujęzyczne, na których maszyna się uczy – najlepsze są dobrej jakości pamięci tłumaczeniowe.
- **Trening** to mechanizm uczenia maszyny, najlepiej powtarzany co jakiś czas (kiedy są dostępne nowe korpusy), oparty na korpusie lub dodatkowo glosariuszu (+DNT).

WARSZTAT POSTEDYCJI NMT

1. Zaczynij od przeczytania i zrozumienia oryginału
2. Szybko (w 3s 😊) oceń, czy odpowiedź MT oddaje sens oryginału...
3. ...lub czy da się ją niewielkim wysiłkiem poprawić
4. Zwróć uwagę na:
 - a) pominięcia
 - b) przetłumaczone nazwy własne, nazwiska itd.
 - c) pozorną zgodność gramatyczną
5. Sprawdzaj spójność tekstu (terminologiczną, stylistyczną) – NMT nie umie!

2. POSTĘPYCJA WYKLUCZA OCHRONĘ PRAW AUTORSKICH, DANYCH OSOBOWYCH ITD.? TO ZALEŻY.

- Każdy dostawca MT ma własne zasady gromadzenia i przechowywania treści
 - Czytamy warunki (Terms and Conditions)
 - Są różnice między dostępem z przeglądarki (bezpłatnie) i z CAT-a (odpłatnie)
 - Polityka dostawcy MT trochę zależy od tego, skąd jest (UE, USA, Chiny, Rosja)
- Na co Klient wyraża zgodę w zakresie korzystania przez Tłumaczy z MT?
- *Który wywiad ma gromadzić dane: USA, chiński czy rosyjski? 😊*

PORÓWNANIE DOSTAWCÓW MT

- <https://modelfront.com/compare>
- <https://try.inten.to/machine-translation-report-2021>

3. JAKOŚĆ MT DLA TEKSTÓW PRAWNYCH PRZEKRACZA JAKOŚĆ TŁUMACZENIA PROFESJONALNEGO? RACZEJ NIE.

- Dokumenty, na których są wykonywane porównania, to zwykle materiały prasowe, teksty o turystyce i t.p. michałki niekrytyczne dla odbiorcy
- Oceny nie zawsze obejmują cały dokument – raczej poszczególne zdania
- EAMT 2020: “Reassessing Claims of Human Parity and Super-Human Performance in Machine Translation at WMT 2019” Antonio Toral
<https://arxiv.org/abs/2005.05738>

DOBRE PRAKTYKI PORÓWNYWANIA TŁUMACZY Z MASZYNA

- <https://www.jair.org/index.php/jair/article/view/11371>
 - kto tłumaczył (amatorzy, zawodowcy)
 - kto porównuje (amatorzy*, zawodowcy)
 - jaki jest tekst źródłowy (z MT, z tłumaczenia, oryginalny)
 - czy uwzględniamy kontekst (cały dokument)

A JEDNAK MT BYWA LEPSZE...?

<https://www.nature.com/articles/s41467-020-18073-9>

“In a context-aware blind evaluation by human judges, CUBBITT significantly outperformed professional-agency English-to-Czech news translation in preserving text meaning (translation adequacy). While human translation is still rated as more fluent, CUBBITT is shown to be substantially more fluent than previous state-of-the-art systems. Moreover, most participants of a Translation Turing test struggle to distinguish CUBBITT translations from human translations.”

“Inspired by a recent discussion of the translation proficiency of evaluators²², we recruited two groups of evaluators: six professional translators (native in the target language) and seven non-professionals (with excellent command of the source language and native in the target language). An additional exploratory group of three translation theoreticians was also recruited. In total, 15 out of the 16 evaluators passed a quality control check, giving 7824 sentence-level scores on 53 documents in total.”

4. KAŻDY TŁUMACZ MUSI NAUCZYĆ SIĘ POSTEDYCJI, ŻEBY NIE WYPAŚĆ Z OBIEGU? NIE, ALE WARTO SPRÓBOWAĆ.

- Większość postedycji odbywa się w zwykłym środowisku pracy tłumacza w jednym z narzędzi CAT (Trados, memoQ, Memsource, Smartling, ...)
- Można wypróbować MT przez darmowy okres próbny (DeepL, ModernMT, Tilde) lub w darmowym limicie słów (Google, Bing - Microsoft, AWS - Amazon)
- Warto ocenić, czy MT pomaga w tłumaczeniu:
 1. Czy z MT robię więcej słów na godzinę (dzień, tydzień) niż bez niego?
 2. Czy korekta/redakcja/weryfikator widzi różnicę?
 3. Jak się czuję z tłumaczeniem, gdzie do każdego segmentu mam podpowiedź?

MEMOQ I TRADOS

- Funkcją pre-translate można wstawić podpowiedzi z MT do całego projektu lub wybranych dokumentów.
- Podpowiedzi z 1 lub więcej maszyn mogą być wyświetlane dla kolejnych segmentów tak jak podpowiedzi z pamięci.
- Silnik MT może też służyć do poprawiania podpowiedzi przybliżonych (MatchPatch/MatchRepair).
- Narzędzie CAT może generować raporty z wydajności tłumaczenia lub z % zmian w tłumaczeniu względem podpowiedzi.

DOPASOWANIE MT

MT trenowane na naszych korpusach (pamięciach)

- ModernMT, Tilde, Amazon, Google, RWS (SDL)

MT uczące się z poprawek postedytora

- ModernMT, Tilde

MT z naszym glosariuszem

- DeepL, Google, ...

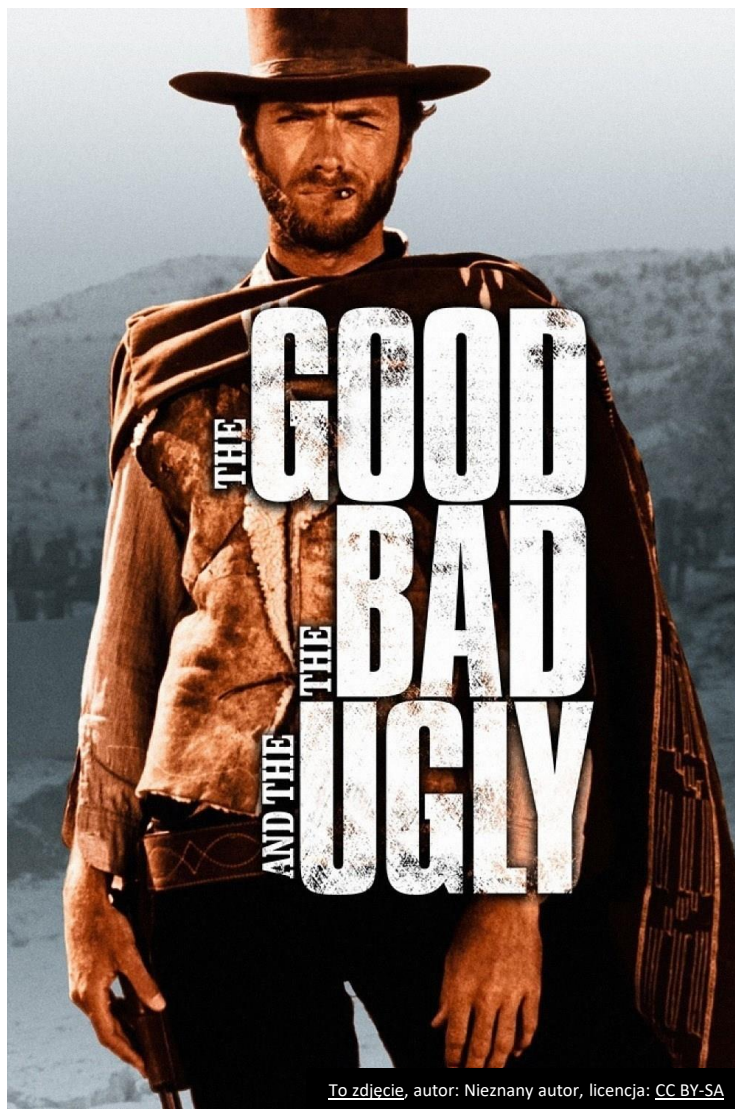
Wybór języka formalnego/nieformalnego lub dziedziny

- DeepL, eTranslation

A JAK NIE POSTĘDYCJA, TO CO?

- Transkrecja
- Tłumaczenia specjalistyczne w nietypowych parach językowych
- Dokumentacja (technical writing)

Tsukihime remake with DeepL is honestly probably better than anything some really bad western localizers will come up with.



To zdjęcie, autor: Nieznany autor, licencja: [CC BY-SA](#)






Autorka bredzi, DeepL bredzi, a ja wiem, co z tym zrobić.

Klient chce obniżyć stawkę o 40%, bo to postędyca, a nam jego MT spowalnia pracę tłumaczy o 30%.

DOBRA PRAKTYKA: NAJPIERW PAMIĘĆ, POTEM MT

When no match found

- Leave target segment empty
- Copy source to target
- Apply automated translation

Name	Enabled	Languages	Lookup	Penalty	Concordance	Update
 Juan English to Polish.sdltm	<input checked="" type="checkbox"/>	 → 	<input checked="" type="checkbox"/>	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
 DeepL Translation provider	<input checked="" type="checkbox"/>	n/a	<input checked="" type="checkbox"/>	0	<input type="checkbox"/>	<input type="checkbox"/>
 ModernMT Adaptive Neural Machine Translation	<input checked="" type="checkbox"/>	n/a	<input checked="" type="checkbox"/>	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>

DOBRA PRAKTYKA: STAWKA ZA PE PROPORCJONALNA DO ZWIĘKSZENIA WYDAJNOŚCI PRACY

„model Acolad” 2020.eamt-1.42.pdf (aclweb.org)

- Upust PE w stosunku do ceny bazowej tłumaczenia musi być związany ze wzrostem wydajności PE w porównaniu z tłumaczeniem bez MT
- Pomiar dokonywany jest dla każdej pary językowej i dziedziny, przez doświadczonych tłumaczy, na dużej próbie tekstu (5000 słów)
- Upust PE nie może być większy niż wzrost produktywności

ZŁA PRAKTYKA: JEDEN SILNIK MT DO WSZYSTKIEGO

Original Sentence
MEP pressing anvil
New Version
Machine Translation
75 Posel do PE naciskający na kowadło



ZŁA PRAKTYKA: LIGHT POST-EDITING

light post-editing

process of post-editing to obtain a merely comprehensible text without any attempt to produce a product comparable to a product obtained by human translation

BRZYDKA PRAKTYKA: UDAWANIE

- Klient zleca PE i udaje, że to redakcja (daje stawkę jak za redakcję)

„NMT post-editing involves less editing than TM segments, but this editing appears to take more time, with the consequence that NMT post-editing does not seem to improve productivity as may have been expected. This might be due to the fact that NMT segments show a higher variability in terms of quality and time invested in post-editing than TM segments that are ‘more similar’ on average.” https://www.researchgate.net/publication/332240015_Post-editing_neural_machine_translation_versus_translation_memory_segments

- Tłumacz przepuszcza tekst przez MT i udaje, że to tłumaczenie z korektą

BRZYDKA PRAKTYKA: „PUSZCZANIE NA PRODUKCJĘ”

Otrzymaj 5€ zniżki na pierwsze zamówienie na Siroko.com! Subskrybuj nasz newsletter, aby otrzymywać nasze promocje:

Podaj swój adres e-mail

Kobieta Człowiek

WYŚLIJ

OKULARY SPORTOWE ▾

OKULARY ▾

SIROKO ▾

REDAKCYJNY ▾



WIĘCEJ O MT I PE?

Neural MT

- <https://towardsdatascience.com/introduction-to-machine-translation-9cb0e93e7cb>
- <https://www.coursera.org/learn/machinetranslation>

Historia MT

- <https://medium.com/free-code-camp/a-history-of-machine-translation-from-the-cold-war-to-deep-learning-f1d335ce8b5>
- <https://www.youtube.com/channel/UCz1aeF3nNrKFgDnTEB3uS9w>

Szkolenia z MT i PE

- <https://machinetranslation.pl/>