Nr rejestracyjny: 2025/57/N/HS2/04961; Kierownik projektu: mgr inż. Maciej Rapacz

Od strategii tłumacza do tendencyjności algorytmu. Ilościowy model analizy przekładu

Współczesne systemy sztucznej inteligencji (AI) pozwalają każdemu z niespotykaną dotąd łatwością generować tłumaczenia. Powszechna dostępność tych narzędzi rodzi pytania dotyczące charakterystyki tekstów generowanych maszynowo, zwłaszcza w przypadku dzieł o dużej wartości kulturowej i interpretacyjnej. W przeszłości opierano się na przekładach instytucjonalnych, zgodnych z określoną tradycją interpretacyjną; dziś każdy może wygenerować nowy przekład. Stwarza to niepewność co do *obiektywności* modelu: którą tradycję będzie naśladował? Czy będzie robił to spójnie?

Badaniami przekładów zajmują się dwie dziedziny: informatyka i translatoryka. Informatyczne badania nad przekładem koncentrują się na metrykach, które raczej oceniają *poprawność* stylu zamiast go *opisywać*. Z kolei translatoryka oferuje ramy teoretyczne do takiej charakterystyki, postrzegając każdy przekład jako adaptację w określonym celu i dla określonego odbiorcy. Jednak metody ilościowe stosowane w przekladoznawstwie wymagają gromadzenia dużych zbiorów tekstów w celu statystycznej analizy decyzji tłumacza, co wyklucza badanie na poziomie zdania. Ponadto fundamenty teorii przekładoznawstwa, takie jak koncepcja strategii przekładu, zostały opracowane przy założeniu istnienia tłumacza (człowieka), który podejmuje decyzje świadomie i w sposób konsekwentny. Tymczasem w przypadku AI wynik działania modelu nie wynika ze świadomych decyzji, lecz ze statystycznego rozkładu danych treningowych i elementu losowości w doborze słów. To, co u człowieka interpretujemy jako *strategię*, w przypadku modelu jest raczej jego wyuczoną *tendencyjnością* (ang. bias). Istnieje więc wyraźna potrzeba stworzenia metod, które tę tendencyjność nie ocenią, lecz opiszą. Metody te powinny jednocześnie umożliwiać analizę stylu na poziomie pojedynczych zdań, a przy tym być na tyle skalowalne, by pozwalać na badanie spójności tej tendencji w całych dziełach.

W odpowiedzi na tę potrzebę proponujemy dwie nowe metody obliczeniowe do charakteryzowania stylistycznych i semantycznych tendencji w tłumaczeniach maszynowych. W celu ich walidacji zbudujemy nowy, wielojęzyczny korpus tłumaczeń Nowego Testamentu – tekstu o najbogatszej na świecie historii przekładu, którego liczne wersje często wyraźnie wskazują na tradycję wyznaniową czy docelowego odbiorcę. Duża gęstość interpretacyjna powoduje, że wybór konkretnego słowa może odzwierciedlać dane stanowisko teologiczne, co czyni ten tekst idealnym materiałem do badania tendencyjności w szerokim zakresie. Zgromadzimy ponad 600 tekstów z kilkunastu stron internetowych w pięciu językach: angielskim, hiszpańskim, francuskim, polskim i włoskim. Teksty te zostaną wzbogacone o metadane, takie jak tradycja wyznaniowa i rok publikacji, co pozwoli na analizę porównawczą w czasie, między tradycjami i w obrębie różnych języków.

Po zbudowaniu korpusu przystąpimy do walidacji naszej pierwszej metody, która wykorzystuje tłumaczenie interlinearne jako obliczeniowy punkt odniesienia do charakterystyki stylu tłumaczenia. Tekst interlinearny to przekład słowo w słowo, który zachowuje strukturę tekstu źródłowego, a jednocześnie używa wyrazów z języka docelowego. Nasza hipoteza zakłada, że różnica między wektorowymi reprezentacjami tekstu interlinearnego i badanego tłumaczenia (nazywana przez nas *wektorem interwencji*) będzie reprezentować wybory stylistyczne i semantyczne dokonane przez tłumacza. Sprawdzimy to, badając, czy algorytmy klastrujące zgrupują wspomniane wektory zgodnie ze znanymi metadanymi (tradycja religijna, wiek powstania) oraz dekodując same wektory, aby przełożyć dane matematyczne na terminy opisowe i zidentyfikować konkretne tendencje stylistyczne, takie jak np. archaizacja.

Za pomocą drugiej metody zbadamy iteracyjne tłumaczenie zwrotne (round-trip translation) jako technikę wzmacniania ukrytych tendencji modelu. Nasza hipoteza zakłada, że cykliczne tłumaczenie tekstu z języka A na język B, a następnie z B na A działa jak sprzężenie zwrotne. Każdy cykl tłumaczenia może subtelnie wzmacniać preferencje stylistyczne modelu, czyniąc je wyraźniejszymi i łatwiejszymi w obserwacji. Przeprowadzimy dwa typy eksperymentów: w jednym rozpoczniemy ten proces od oryginalnego greckiego tekstu, natomiast w drugim od istniejących tłumaczeń ludzkich, aby sprawdzić, czy tendencje modelu można ukierunkować na konkretną tradycję.

W rezultacie projekt dostarczy dwie metody pomiaru charakterystyki tłumaczeń oraz wielojęzyczny korpus wspierający te badania. Naszym celem jest wypełnienie luki metodologicznej między informatyką a przekładoznawstwem poprzez stworzenie nowych sposobów badania przekładu na skalę, która wcześniej była trudna do osiągnięcia. Docelowo przyczyni się to do lepszego zrozumienia technologii AI, z których na co dzień korzysta coraz więcej osób. Choć projekt koncentruje się na tłumaczeniach maszynowych, mamy nadzieję, że jego wyniki przyczynią się również do badań nad przekładem ludzkim.