



INSTYTUT PODSTAW INFORMATYKI  
POLSKIEJ AKADEMII NAUK  
ul. Jana Kazimierza 5, 01-248 Warszawa

tel.: +48 22 380 05 00  
fax.: +48 22 380 05 10

e-mail: [ipi@ipipan.waw.pl](mailto:ipi@ipipan.waw.pl)  
[www.ipipan.waw.pl](http://www.ipipan.waw.pl)

Dr hab. Maciej Ogrodniczuk, prof. IPI PAN  
Zakład Sztucznej Inteligencji  
Zespół Inżynierii Lingwistycznej

[maciej.ogrodniczuk@ipipan.waw.pl](mailto:maciej.ogrodniczuk@ipipan.waw.pl)  
tel. 22 380 05 63  
533 675 675

Warszawa, 8 czerwca 2021 r.

## RECENZJA ROZPRAWY DOKTORSKIEJ

**Tytuł rozprawy:** Wydajne algorytmy dla hybrydowego modelu neuronowego tłumaczenia maszynowego  
**Autor rozprawy:** mgr Tomasz Dwojak  
**Dyscyplina:** informatyka  
**Promotor rozprawy:** prof. dr hab. Krzysztof Jassem  
**Promotor pomocniczy:** dr Marcin Junczys-Dowmunt

Recenzja wykonana jest na zlecenie Rady naukowej dyscyplin matematyka i informatyka UAM zgodnie z jej uchwałą z 23 lutego 2021 r.

### Przedmiot rozprawy

Do oceny przedłożona została rozprawa w postaci zbioru sześciu jednotematycznych artykułów w języku angielskim opublikowanych w recenzowanych materiałach z konferencji o zasięgu międzynarodowym (pięć artykułów) oraz w recenzowanym czasopiśmie naukowym (jeden artykuł). Głównym tematem rozprawy jest porównanie dwóch stosowanych w ostatnim czasie podejść do tłumaczenia maszynowego języka naturalnego – metody statystycznej i neuronowej. W swoich badaniach Doktorant weryfikuje hipotezę o zasadności łączenia elementów tych obu podejść.

Tak postawione zagadnienie wymaga w pierwszej kolejności pracochłonnej i wnikliwej analizy istniejącego stanu technologii, przetestowania rozwiązań realizujących oba podejścia w różnych konfiguracjach, ale także uwzględnienia wielu aspektów ewaluacji: jakościowego, wydajnościowego czy coraz częściej podnoszonej kwestii zużycia zasobów podczas trenowania modeli i ich wykorzystania. Dodatkowo istniejące rozwiązania mogą wymagać pewnych usprawnień algorytmicznych.

### Treść rozprawy

#### Artykuł 1: *Is neural machine translation ready for deployment? A case study on 30 translation directions*

Artykuł porównuje jakość i szybkość działania statystycznych systemów tłumaczenia maszynowego opartych na frazach (ang. *Phrase-Based Statistical Machine Translation*) z systemami neuronowymi (ang. *Neural Machine Translation*) na materiale korpusu ONZ oraz jego podkorpusu zrównoleżonego dla 6 języków na poziomie zdań. Wynikiem tego porównania jest potwierdzenie dominacji systemów neuronowych nad statystycznymi pod względem jakości i szybkości tłumaczenia. Tekst kończy się konkluzją, że neuronowe systemy tłumaczenia maszynowego mogą z powodzeniem zastąpić wykorzystywane produkcyjnie systemy statystyczne.

Artykuł jest bardzo dobrym wprowadzeniem w temat rozprawy, zgodnie z przyjmowaną przeze mnie nieformalną zasadą, że w zbiorze prac co najmniej jedna powinna przedstawiać podstawy teoretyczne przeprowadzonych badań. Pierwszy tekst ze zbioru spełnia tę zasadę, jest klarowny i realizuje wszystkie stawiane sobie cele. Już w obecnej chwili jest pracą, na którą często powołują się inni autorzy w rozdziałach wprowadzających do swoich badań, przywołując stwierdzenie o wyższości neuronowych systemów tłumaczenia maszynowego nad statystycznymi. Potwierdza to także rosnąca liczba cytowań pracy (wzrost cytowań Google Scholar ze wskazanych w autoreferacie 149 do 167 od momentu złożenia autoreferatu 19 marca do 8 czerwca 2021 r.)

Za najważniejszy wkład autora w opisane prace zgodnie z deklaracją z autoreferatu uznaję współautorstwo (przyjmuję, że także koncepcyjne) systemu AmuNMT. Algorytm do selekcji podśłów zgodnie z tekstem wydaje się być wzorowany na pracach Jean et al. (*On Using Very Large Target Vocabulary for Neural Machine Translation*, 2015) oraz Mi et al. (*Vocabulary Manipulation for Neural Machine Translation*, 2016). Udział Doktoranta w pracach został oceniony na 20%.

## Artykuł 2: The AMU-UEDIN Submission to the WMT16 News Translation Task: Attention-based NMT Models as Feature Functions in Phrase-based SMT

Artykuł opisuje system zgłoszony do udziału w zadaniu ewaluacyjnym (ang. *shared task*) pn. *Machine Translation of News* na konferencji *1st Conference on Machine Translation*. Konferencja ta, mimo że opatrzona numerem pierwszym, miała już dziesięcioletnią historię jako warsztat *Workshop on Statistical Machine Translation* organizowany przy najważniejszych konferencjach środowiska przetwarzania języka naturalnego (takich jak NAACL, ACL, EACL czy EMNLP).

Zadaniem uczestników było przetłumaczenie zbioru testowego pomiędzy angielskim a jednym z sześciu języków testowych, którego jakość była następnie oceniana metodami ręcznymi i automatycznymi. Zgłoszony tekst dotyczył wyłącznie tłumaczenia z rosyjskiego na angielski i z angielskiego na rosyjski. Rozwiązanie opierało się na połączeniu modelu neuronowego z atencją z modelem statystycznym opartym na frazach.

Zaprezentowany system ustanowił najlepszy dotychczasowy wynik na świecie, poprawiając dotychczasowy o 0,8 pkt miary BLEU oraz okazał się najlepszy w ręcznej ocenie wyników dla trybu ograniczonego (ang. *constrained*), zakładającego korzystanie wyłącznie z dostarczonych danych. System tłumaczący z angielskiego na rosyjski okazał się gorszy jedynie od systemu NMT-4 Sennricha i Haddowa, ale także najlepszy w ewaluacji ręcznej.

Wkład Doktoranta w przedstawione badania oprócz prac implementacyjnych polegał na wspólnym opracowaniu sposobu integracji modelu neuronowego z atencją z modelem statystycznym oraz pracach implementacyjnych. Jego udział został oceniony na 35%. Liczba cytowań tekstu wg Google Scholar zwiększyła się od złożenia pracy z 44 do 51.

## Artykuł 3: Marian: Fast Neural Machine Translation in C++

Artykuł opisuje wyniki prac nad platformą programistyczną w języku C++ służącą do tworzenia neuronowych systemów tłumaczenia maszynowego. Na danych angielsko-niemieckich z konferencji WMT17 (*Second Conference on Machine Translation*) system stworzony na tej platformie i wykorzystujący sieć neuronową typu transformer uzyskał o 1,2 punktu lepsze wyniki niż zwycięski system z poprzedniego roku.

Tekst został zamieszczony w części *System Demonstrations*, więc zgodnie z wymaganiami formatu konferencji jest krótki (5 stron) i prezentuje raczej rozwiązanie techniczne niż badawcze. Taki charakter ma też wkład Doktoranta (niski – 10%, ale przy 12 współautorach), polegający głównie na pracach implementacyjnych. Można to jednak uznać za naturalną kolej rzeczy w przypadku opisu systemu, gdzie implementacja stanowi najważniejszą część pracy.

Bez wątplenia należy natomiast docenić użyteczność rozwiązania (wykorzystanego m.in. przez Microsoft i Światową Organizację Własności Intelektualnej) oraz oddźwięk, jakie rozwiązanie to zyskało także w świecie naukowym (artykuł jest najbardziej cytowaną pracą Doktoranta – wzrost liczby cytowań w wymienionym wyżej okresie ze 180 do 281).

Lista potwierdzeń merytorycznego i procentowego wkładu Doktoranta w powstanie pracy nie zawiera oświadczeń siedmiu autorów (Toma Neckermanna, Franka Seide, Ulricha Germanna, Alhama Fikri Aji, Nikolaya Bogoycheva, André F. T. Martinsa oraz Alexandry Birch, której oświadczenie dla innego tekstu zostało dołączone), co rozumiem jako wynikające z dużej liczby autorów (oświadczenia pięciu pierwszych autorów tekstu zostały dołączone).

#### Artykuł 4: Predicting Target Language CCG Supertags Improves Neural Machine Translation

Artykuł z materiałów konferencji *Second Conference on Machine Translation (WMT17)* jest odpowiedzią na problemy systemów tłumaczenia maszynowego ze składnią tekstu wynikowego (niewłaściwe formy skoordynowane, kwestia dołączania fraz przysłówkowych). Po refleksji nt. niewielkiej poprawy jakości wyników przy próbach wykorzystania struktur składniowych danych źródłowych, autorzy z powodzeniem uzyskują taką poprawę (nawet o 1,2 punktu BLEU dla tłumaczenia z rumuńskiego na angielski) po uzupełnieniu danych wynikowych o kategorie składniowe CCG.

Za najważniejszy wkład Doktoranta (raportowany na poziomie 20%) uznaję opracowanie planu eksperymentów oraz udział w opracowaniu algorytmów korzystających z informacji składniowej (według autoreferatu; potwierdzenia innych autorów zawierają nieco inny w wymowie tekst „*Research on possible model architectures with incorporated tags*”). Tekst zanotował wzrost cytowań z 58 do 72 według Google Scholar.

Lista potwierdzeń merytorycznego i procentowego wkładu Doktoranta w powstanie pracy nie zawiera oświadczeń dwóch autorów – Marcina Junczysa-Dowmunta (którego oświadczenia dla pozostałych czterech tekstów zostały dołączone) oraz Philippa Koehna. Przy dużej liczbie współautorów traktuję jednak ten fakt jako mało znaczące niedopatrzenie.

#### Artykuł 5: Fast Neural Machine Translation Implementation

Artykuł został zgłoszony do zadania ewaluacyjnego *Efficient NMT* na warsztacie *2nd Workshop on Neural Machine Translation and Generation (WNMT 2018)* i miał na celu poprawę wydajności dostępnego rozwiązania tłumaczeniowego.

Wkład Doktoranta w pracę uznaję za decydujący (podany udział: 50%), gdyż polegał na opracowaniu dwóch najważniejszych algorytmów usprawniających wydajność działania systemu AmuNMT.

W tabeli wymieniającej artykuły naukowe stanowiące osiągnięcie badawcze Doktoranta (s. 6–7) podana jest błędna (w moim rozumieniu sprawy) wartość punktowa pracy (200 pkt) podczas gdy warsztatu *2nd Workshop on Neural Machine Translation* nie ma w wykazie MEiN czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych (warsztatów przy konferencji w mojej opinii nie należy traktować jak konferencji).

#### Artykuł 6: Statistical versus neural machine translation – a case study for a medium size domain-specific bilingual corpus

Najobszerniejszy artykuł zestawu pochodzi ze specjalnego numeru czasopisma *Poznan Studies in Contemporary Linguistics* poświęconego polskiej lingwistyce komputerowej (pod moją redakcją) i dotyczy porównania jakości systemu statystycznego i neuronowego na materiale polsko-angielskiego korpusu dziedzinowego. Przeprowadzone eksperymenty wykazały, że model statystyczny może uzyskiwać wyniki porównywalne z modelem neuronowym.

W mojej opinii artykuł spełnia niepisany warunek, że Doktorant powinien być wiodącym autorem przynajmniej jednego artykułu, a jego wkład (potwierdzony przez współautorów) powinien być na poziomie co najmniej 50% – mimo że w autoreferacie podana jest inna niż faktyczna kolejność autorów. Raportowany udział Doktoranta w pracy to właśnie 50%, jego wkład uznaję jednak w tym przypadku za decydujący. Aktualna liczba cytowań tekstu to 2.

## Podsumowanie

Artykuły ułożone są w logiczną całość. Pierwszy stanowi swoiste wprowadzenie i punkt wyjścia do rozważań poprzez wykazanie, że neuronowe systemy tłumaczenia maszynowego przewyższyły pod względem jakości używane przez wiele lat systemy statystyczne, co wskazuje ścieżkę dla dalszych badań oraz skłania do przeprowadzenia eksperymentów z systemami hybrydowymi, łączącymi oba podejścia.

Ten ostatni temat podejmuje druga praca, integrująca model statystyczny oparty na frazach z modelem neuronowym z atencją. Trzecia praca, choć chronologicznie późniejsza niż czwarta, podejmuje temat stworzenia architektury do tworzenia neuronowych modeli tłumaczenia maszynowego, co jest o tyle ważne, że liczba rozwiązań i możliwych do zastosowania parametrów jest w tym przypadku olbrzymia.

Praca czwarta omawia usprawnienie procesu tłumaczenia maszynowego poprzez uzupełnienie danych wyjściowych o struktury składniowe, co ma pomóc w poprawie jakości wyników tłumaczenia. Artykuł piąty podejmuje problem wydajności modeli neuronowych i wykorzystuje stworzoną wcześniej architekturę do (zakończonych pomyślnie) eksperymentów w tym zakresie.

Artykuł szósty jest jedynym dotyczącym kwestii tłumaczenia maszynowego dla języka polskiego. Dodatkowo, w związku z uzyskanym przez system statystyczny wynikiem porównywalnym z wynikami systemu neuronowego przy znacznie szybszym działaniu, za to mniejszej płynności tekstu wyjściowego, tekst stanowi kłamrę podsumowującą prace nad systemami obu typów.

## Uwagi redakcyjne

Osiągnięciami badawczymi Doktoranta (p. autoreferat, rozdział 2, s. 5) są artykuły naukowe, nie są nimi natomiast cytowania, informację o których należałoby dołączyć do poprzedniego punktu.

Na s. 15 autoreferatu czytamy: „Przeprowadzona analiza potwierdziła aktualność postawionej hipotezy badawczej” (o zasadności podejścia łączącego metodę statystyczną i neuronową) – wyrażenie „aktualność hipotezy” wydaje mi się niezręczne – wolałbym, żeby hipoteza została potwierdzona albo obalona.

Na liście artykułów (strona vii pracy) tekst *Is neural machine translation ready for deployment? A case study on 30 translation directions* wydaje się mieć błędny numer warsztatu w odwołaniu bibliograficznym – na stronie warsztatu IWLST 2016 (<https://workshop2016.iwslt.org>) pojawia się informacja, że był to *13th International Workshop on Spoken Language Translation*, zaś w autoreferacie mamy *Ninth International Workshop on Spoken Language Translation*.

## Naukowa wartość rozprawy

Głównym wkładem Doktoranta jest porównanie podejścia statystycznego z neuronowym prowadzone w latach niezwykle intensywnego rozwoju metod uczenia głębokiego. W ramach swojej pracy Autor prowadził eksperymenty na najwyższym światowym poziomie, doprowadzając do rozwoju algorytmów w wykorzystywanych do dziś systemach tłumaczenia maszynowego.

Wszystkie prace Doktoranta mają wymiar praktyczny. Ich techniczny charakter mógłby stanowić zarzut, jednak w obecnym świecie informatycznym rozwiązania wysokiego poziomu nie ograniczające się do pojedynczego algorytmu, ale wymagające wielopłaszczyznowej optymalizacji, muszą opierać się na praktycznych testach zaproponowanych rozwiązań. Jest to tym ważniejsze w przypadku rozwiązań o tak dużym oddźwięku społecznym jak właśnie systemy tłumaczenia maszynowego, wykorzystywane dziś powszechnie i będące także przedmiotem zainteresowania wielu podmiotów komercyjnych, wystawiających swoje zespoły w tych samych konkursach, w których brał udział Doktorant.

Potencjalny zarzut mniejszej wagi artykułów konferencyjnych w porównaniu z pracami w czasopiśmie jest w tym przypadku chybiony: w świecie przetwarzania języka naturalnego to właśnie prace konferencyjne wyznaczają światowe trendy, przede wszystkim ze względu na bardzo szybką dostępność pracy w finalnym kształcie – już w momencie jej prezentacji. Co więcej, obecność prac na uznanych w środowisku naukowym warsztatach tematycznych oznacza, że były one recenzowane przez co najmniej dwóch niezależnych recenzentów zewnętrznych, specjalistów nie tylko w ogólnie rozumianej inżynierii lingwistycznej, ale właśnie w wąskiej dziedzinie tłumaczenia maszynowego.

Prace wchodzące w skład rozprawy prezentują niewątpliwie oryginalne rozwiązanie problemu naukowego, a ich polskich autorów należy uznać za jednych z najlepszych rodzimych specjalistów w tej dziedzinie, trzeba jednak postawić pytanie, jaki udział w ich powstaniu miał Doktorant. Nie jest on pierwszym ani samodzielnym autorem żadnego artykułu, a jego wkład w przedstawione prace nie przekracza 50% (w dwóch pracach; w pozostałych 35%, 20%, 20% i 10%). Wierząc, że Doktorant będzie miał okazję wyjaśnić ewentualne wątpliwości podczas publicznej obrony, tłumaczę ten fakt zespołowym trybem działania, coraz bardziej naturalnym w obecnych czasach, zwłaszcza w dziedzinie technicznej. Prace światowego poziomu nie mogą być tworzone przez samodzielnych badaczy, lecz coraz częściej przez grupy badawcze. Fakt funkcjonowania Doktoranta w tego rodzaju grupie, potwierdzony także jego udziałem w stażach naukowych w zagranicznych ośrodkach badawczych (w tym w ponad rocznym stażu na Uniwersytecie w Edynburgu – jednym z najlepszych uniwersytetów w Europie) oraz międzynarodowych grantach z programu Horyzont 2020, traktuję zatem jako pozytywny, przy założeniu, że udział Doktoranta nie ograniczał się w czterech pracach o jego najmniejszym wykazanym udziale procentowym do implementacji pomysłów pozostałych badaczy z zespołu (w szczególności promotora pomocniczego).

Fakt, że promotor lub promotor pomocniczy są współautorami prawie wszystkich (oprócz jednej) publikacji, które wchodzą w skład rozprawy doktorskiej, również traktuję jako pozytywny, spełniający warunek przygotowania pracy pod kierunkiem naukowym promotora i promotora pomocniczego.

## Konkluzja

Podsumowując stwierdzam, iż przedłożona mi do recenzji rozprawa, której autorem jest mgr Tomasz Dwojak, zawiera ważne osiągnięcia w obszarze przetwarzania języka naturalnego w zakresie tłumaczenia maszynowego. Doktorant wykazał się dużą wiedzą w tematyce rozprawy oraz znajomością metod badawczych. Przedstawiony zestaw artykułów cechuje spójność tematyczna w ramach weryfikacji hipotezy badawczej stawianej w autoreferacie.

Recenzowana praca spełnia wymagania ustawowo stawiane rozprawom doktorskim, zatem wnoszę o to, by mgr Tomasz Dwojak został dopuszczony do publicznej obrony.



Maciej Ogrodniczuk