

RECENZJA ROZPRAWY DOKTORSKIEJ

Mgr. Steven'a Jarosza

A Neural Network Approach to Modeling the Acoustic Effects of Phonological Neighborhood Density and Word Frequency in Polish Speech

przygotowanej pod kierunkiem

prof. dr. hab. Arkadiusza Rojczyka

oraz

prof. Margherity Dore

1. Uwagi wprowadzające

Przedłożona do recenzji rozprawa doktorska mgr. Stevena Jarosza pt. *A Neural Network Approach to Modeling the Acoustic Effects of Phonological Neighborhood Density and Word Frequency in Polish Speech* stanowi interesujące i ambitne studium z pogranicza fonetyki eksperymentalnej, psycholingwistyki oraz współczesnych badań nad sztuczną inteligencją. Autor analizuje problem wpływu gęstości sąsiedztwa fonologicznego (Phonological Neighborhood Density – PND) oraz frekwencji wyrazowej (Word Frequency–WF) na realizację akustyczną samogłosek w języku polskim, rozszerzając jednocześnie analizę o modele syntezy i rozpoznawania mowy oparte na architekturach neuronowych.

Rozprawa liczy 147 stron (w tym 116 stron rozprawy, 19 stron bibliografii oraz 5 stron załączników) i została podzielona na sześć rozdziałów, poprzedzonych wprowadzeniem oraz zakończonych obszerną bibliografią i aneksami. Konstrukcja pracy jest przejrzysta, logiczna i podporządkowana jasno sformułowanym pytaniom badawczym. Autor konsekwentnie prowadzi czytelnika od podstaw teoretycznych poprzez opis metodologii aż do prezentacji wyników i ich interpretacji.

Problematyka podjęta w rozprawie jest aktualna i istotna zarówno dla językoznawstwa polonistycznego, jak i dla współczesnych badań nad sztuczną inteligencją. Szczególne znaczenie ma fakt, iż większość wcześniejszych badań nad wpływem frekwencji leksykalnej i gęstości sąsiedztwa fonologicznego prowadzona była na materiale anglojęzycznym. Rozprawa mgr. Jarosza wypełnia zatem istotną lukę badawczą, dostarczając nowych danych dotyczących języka polskiego.

Jednocześnie praca posiada wyraźny wymiar interdyscyplinarny. Autor nie ogranicza się do klasycznej analizy fonetycznej, lecz podejmuje próbę odpowiedzi na pytanie, czy współczesne modele neuronowe odtwarzają mechanizmy obserwowane w ludzkiej produkcji i percepcji mowy. Tak postawiony problem należy uznać za nowatorski i ważny poznawczo.

2. Znaczenie problematyki badawczej

Wybór tematu rozprawy należy ocenić bardzo wysoko. Autor podejmuje zagadnienie znajdujące się obecnie w centrum zainteresowania zarówno psycholingwistyki, jak i lingwistyki obliczeniowej. Problemy związane z reprezentacją leksykalną, organizacją słownika umysłowego oraz wpływem częstotliwości użycia jednostek językowych na ich realizację fonetyczną należą do fundamentalnych pytań współczesnego językoznawstwa.

Szczególne znaczenie ma wykorzystanie kategorii sąsiedztwa fonologicznego. Koncepcja ta, rozwijana od prac Luce'a i Pisoniego, dostarczyła wielu istotnych ustaleń dotyczących percepcji i produkcji mowy. Dotychczas jednak jej zastosowanie do języka polskiego pozostawało ograniczone. Autor słusznie zauważa, że odmienna struktura fonologiczna polszczyzny może prowadzić do odmiennych wzorców organizacji leksykalnej, a tym samym wymaga odrębnych badań empirycznych.

Drugim ważnym aspektem rozprawy jest problem frekwencji wyrazowej. Autor podejmuje klasyczne pytanie o relacje między częstością użycia jednostek leksykalnych, a ich realizacją akustyczną, odwołując się zarówno do teorii redukcji artykulacyjnej, jak i modeli hiper- i hipoartykulacji Lindbloma. Szczególnie cenne jest zestawienie wyników badań prowadzonych dla języka angielskiego z rezultatami uzyskanymi dla polszczyzny.

Trzecim, najbardziej nowatorskim wymiarem pracy jest wykorzystanie współczesnych systemów sztucznej inteligencji. Autor analizuje nie tylko mowę ludzką, ale również zachowanie modeli Amazon Polly, ElevenLabs oraz wav2vec 2.0. Dzięki temu rozprawa wykracza poza tradycyjne ramy fonetyki eksperymentalnej i wpisuje się w dynamicznie rozwijający się nurt badań nad podobieństwami i różnicami między ludzkim przetwarzaniem języka a reprezentacjami tworzonymi przez sieci neuronowe.

3. Ocena podstaw teoretycznych

Część teoretyczna rozprawy została opracowana na wysokim poziomie merytorycznym. Autor wykazuje bardzo dobrą znajomość literatury przedmiotu, obejmującej zarówno klasyczne opracowania psycholingwistyczne, jak i najnowsze publikacje z zakresu sztucznej inteligencji.

Szczególnie pozytywnie należy ocenić omówienie modeli organizacji słownika mentalnego. Autor prezentuje teorię Neighborhood Activation Model, koncepcję konkurencji leksykalnej oraz współczesne podejścia dynamiczne, wskazując ich potencjalne implikacje dla badań fonetycznych. Nie ogranicza się przy tym do referowania stanowisk, lecz podejmuje próbę ich krytycznego porównania.

Warto podkreślić również bardzo dobre przedstawienie teorii H&H Lindbloma. Autor przekonująco argumentuje, że relacje między frekwencją wyrazową, a stopniem artykulacyjnej redukcji mogą być interpretowane w perspektywie ekonomii wysiłku komunikacyjnego. Te rozważania stanowią ważny fundament dla późniejszej interpretacji wyników eksperymentalnych.

Na uznanie zasługuje również omówienie zagadnień związanych z nowoczesnymi architekturami neuronowymi. Autor prezentuje podstawowe zasady działania sieci głębokich, mechanizmów uwagi „attention mechanisms” oraz modeli transformacyjnych w sposób zrozumiały i jednocześnie zachowujący odpowiednią precyzję naukową.

Można jedynie zauważyć, że niektóre fragmenty rozdziału dotyczącego sztucznej inteligencji mają charakter bardziej podręcznikowy, niż bezpośrednio związany z głównym problemem rozprawy. W szczególności szczegółowe omówienie funkcji aktywacji, czy podstaw algebry liniowej mogłoby zostać nieco skrócone. Uwaga ta ma jednak charakter redakcyjny i nie wpływa na wysoką ocenę całej części teoretycznej.

4. Metodologia badań – ocena ogólna

Najmocniejszą stroną rozprawy jest, moim zdaniem, jej warstwa metodologiczna. Autor zaprojektował wieloetapowe badanie obejmujące zarówno produkcję mowy przez rodzimych użytkowników języka polskiego, jak i analizę współczesnych modeli sztucznej inteligencji. Tak szeroko zakrojony projekt badawczy wymagał znacznych kompetencji z zakresu fonetyki eksperymentalnej, programowania, analizy danych oraz statystyki.

Na szczególne uznanie zasługuje fakt, że Autor nie ograniczył się do jednego eksperymentu, lecz stworzył spójny program badawczy obejmujący trzy wzajemnie uzupełniające się obszary:

1. analizę mowy produkowanej przez ludzi,
2. analizę neuronowych systemów syntezy mowy,
3. analizę modeli automatycznego rozpoznawania mowy.

Takie podejście znacząco zwiększa wartość poznawczą rozprawy, umożliwiając porównanie mechanizmów obserwowanych w różnych systemach przetwarzania języka.

5. Dobór materiału eksperymentalnego

Materiał badawczy został przygotowany z dużą starannością. Autor wykorzystał zbiór 104 jednostek leksykalnych dobranych pod kątem frekwencji użycia oraz poziomu gęstości sąsiedztwa fonologicznego. Szczególnie istotne jest wykorzystanie zasobów JIwar, które stanowią obecnie jedno z najważniejszych narzędzi badań psycholingwistycznych dla języka polskiego.

Dobór bodźców uwzględniał konieczność kontrolowania szeregu zmiennych zakłócających, takich jak długość wyrazu, struktura sylabiczna czy pozycja analizowanych samogłosek. Dzięki temu możliwe było wyizolowanie wpływu interesujących Autora czynników.

Pozytywnie należy ocenić również decyzję o wykorzystaniu zarówno wyrazów o wysokiej, jak i niskiej frekwencji oraz różnym stopniu zagęszczenia sąsiedztwa fonologicznego. Pozwoliło to na przeprowadzenie analiz porównawczych i uniknięcie uproszczonych interpretacji.

6. Uczestnicy badań i procedura nagraniowa

Badanie przeprowadzono z udziałem dwudziestu sześciu rodzimych użytkowników języka polskiego. Liczebność próby należy uznać za wystarczającą dla badań fonetycznych tego typu, zwłaszcza, że Autor uzyskał bardzo dużą liczbę analizowanych realizacji dźwiękowych.

Procedura nagraniowa została opisana szczegółowo i spełnia standardy współczesnej fonetyki eksperymentalnej. Autor zadbał o standaryzację warunków rejestracji, eliminację czynników zakłócających oraz odpowiednią jakość materiału akustycznego.

Łączna liczba przeanalizowanych tokenów – obejmująca zarówno mowę ludzką, jak i syntetyczną – świadczy o dużej skali przedsięwzięcia badawczego i zwiększa wiarygodność uzyskanych rezultatów.

Na uwagę zasługuje również fakt, że Autor konsekwentnie dokumentuje wszystkie etapy przygotowania materiału badawczego, umożliwiając potencjalną replikację eksperymentu przez innych badaczy.

7. Metody analizy akustycznej

Szczególnie wysoko należy ocenić zastosowane procedury analizy akustycznej. Autor wykorzystał uznane narzędzia badawcze, w tym program PRAAT oraz bibliotekę Parselmouth, zapewniające wysoką precyzję pomiarów.

Istotnym elementem metodologii było zastosowanie normalizacji Lobanova, pozwalającej ograniczyć wpływ indywidualnych różnic anatomicznych między uczestnikami. Decyzję tę należy uznać za w pełni uzasadnioną i zgodną z najlepszymi praktykami współczesnej fonetyki eksperymentalnej.

Dodatkowo Autor wykorzystał transformację do skali Bark, umożliwiającą lepsze odzwierciedlenie właściwości percepcyjnych ludzkiego słuchu. Takie rozwiązanie świadczy o dużej świadomości metodologicznej i umiejętności integrowania perspektyw produkcyjnych i percepcyjnych.

Na szczególne uznanie zasługuje zastosowanie wielu komplementarnych miar akustycznych: analiz centroidowych, odległości euklidesowych między kategoriami samogłoskowymi, miar wewnątrzkategorialnych oraz analiz kierunków przesunięć artykulacyjnych. Dzięki temu interpretacja wyników nie opiera się na pojedynczym wskaźniku, lecz na spójnym systemie wzajemnie uzupełniających się analiz.

Równie pozytywnie należy ocenić wykorzystanie modeli efektów mieszanych, które obecnie stanowią standard analiz statystycznych w badaniach językoznawczych. Autor poprawnie uwzględnia zarówno efekty stałe, jak i zmienność indywidualną uczestników, co zwiększa wiarygodność wyprowadzanych wniosków.

8. Wyniki eksperymentu dotyczącego mowy rodzimych użytkowników języka polskiego

Najważniejszą część rozprawy stanowią wyniki badań empirycznych przeprowadzonych na materiale pochodzącym od rodzimych użytkowników języka polskiego. Autor podjął próbę odpowiedzi na pytanie, czy frekwencja wyrazowa oraz gęstość sąsiedztwa fonologicznego wpływają na realizację akustyczną samogłosek w sposób analogiczny do obserwowanego w języku angielskim.

Uzyskane wyniki wskazują, że zależności te rzeczywiście występują również w polszczyźnie, choć ich charakter nie zawsze jest identyczny z ustaleniami formułowanymi dla innych języków. Szczególnie interesujące są obserwacje dotyczące stopnia rozproszenia przestrzeni samogłoskowej oraz zmian odległości między kategoriami wokalicznymi w zależności od analizowanych zmiennych psycholingwistycznych.

Autor interpretuje uzyskane rezultaty w perspektywie teorii hiper- i hipoartykulacji Lindbloma. Wysoka frekwencja wyrazowa wiąże się bowiem z tendencją do ekonomizacji wysiłku artykulacyjnego, podczas gdy większa gęstość sąsiedztwa fonologicznego może prowadzić do

wzrostu precyzji realizacyjnej wynikającej z konieczności unikania niejednoznaczności komunikacyjnej. Interpretacja ta została przedstawiona w sposób wyważony i przekonujący.

Na szczególne uznanie zasługuje fakt, że Autor nie ogranicza się do prostych porównań średnich wartości formantowych, lecz analizuje całą strukturę przestrzeni samogłoskowej. Wykorzystanie analiz centroidowych, miar odległości euklidesowych oraz wskaźników rozproszenia wewnątrzkategorialnego pozwala uchwycić subtelne efekty, które pozostawałyby niewidoczne przy zastosowaniu tradycyjnych metod analizy.

Wyniki badań wskazują również, że relacje między frekwencją wyrazową, a gęstością sąsiedztwa fonologicznego mają charakter złożony i wielowymiarowy. Nie można ich sprowadzić do prostych zależności liniowych. Autor trafnie podkreśla konieczność uwzględniania współdziałania obu czynników oraz ich potencjalnego wpływu na organizację słownika mentalnego użytkowników języka.

Szczególnie cenne jest zachowanie ostrożności interpretacyjnej. Doktorant unika nadmiernych uogólnień, wskazując zarówno mocne strony uzyskanych rezultatów, jak i ich ograniczenia. Takie podejście świadczy o wysokiej kulturze metodologicznej oraz dojrzałości naukowej.

9. Analiza wyników uzyskanych dla modeli syntezy mowy

Drugim istotnym elementem rozprawy jest analiza dwóch współczesnych systemów syntezy mowy: Amazon Polly oraz ElevenLabs. Włączenie modeli generatywnych do badań fonetycznych należy uznać za rozwiązanie nowatorskie i niezwykle interesujące poznawczo.

Autor słusznie wychodzi z założenia, że skuteczność współczesnych modeli sztucznej inteligencji nie przesądza jeszcze o ich podobieństwie do ludzkich mechanizmów przetwarzania języka. Analiza subtelnych właściwości akustycznych może bowiem ujawnić istotne różnice pomiędzy systemami biologicznymi i sztucznymi.

Wyniki badań wskazują, że oba analizowane systemy wykazują częściową zdolność odtwarzania właściwości obserwowanych w ludzkiej mowie. Jednocześnie ujawniają się istotne ograniczenia dotyczące reprezentacji czynników psycholingwistycznych, takich jak frekwencja wyrazowa czy gęstość sąsiedztwa fonologicznego.

Na szczególne podkreślenie zasługuje porównawcza analiza obu modeli. Autor wykazuje, że różnice między nimi nie mają charakteru wyłącznie technicznego, lecz dotyczą również sposobu organizacji reprezentacji językowych. Wyniki te posiadają znaczenie wykraczające poza samą fonetykę eksperymentalną i mogą stanowić punkt wyjścia dla dalszych badań nad architekturami generatywnymi.

Interesujące są zwłaszcza obserwacje dotyczące stopnia naturalności realizacji samogłoskowych. Pomimo bardzo wysokiej jakości generowanej mowy, modele neuronowe nie zawsze odtwarzają wzorce artykulacyjne charakterystyczne dla rodzimych użytkowników języka. Oznacza to, że osiągnięcie naturalności percepcyjnej nie musi być równoznaczne z reprodukowaniem wszystkich mechanizmów leżących u podstaw ludzkiej produkcji mowy.

Autor trafnie wskazuje, iż współczesne systemy sztucznej inteligencji mogą wykorzystywać odmienne strategie optymalizacyjne niż człowiek. Z perspektywy praktycznej nie stanowi to problemu, jednak z punktu widzenia badań kognitywnych wymaga zachowania ostrożności przy formułowaniu analogii między systemami biologicznymi i sztucznymi.

Rozważania te należy uznać za jeden z najbardziej wartościowych elementów rozprawy.

10. Eksperyment dotyczący modelu wav2vec 2.0

Trzecim komponentem projektu badawczego jest analiza modelu automatycznego rozpoznawania mowy wav2vec2-large-xlsr-53-polish. Wybór tego rozwiązania należy ocenić pozytywnie, gdyż model ten reprezentuje współczesny standard w zakresie technologii ASR dla języka polskiego.

Autor nie ogranicza się do prostego pomiaru skuteczności rozpoznawania, lecz podejmuje znacznie ambitniejsze zadanie polegające na analizie *wpływu* zmiennych psycholingwistycznych na funkcjonowanie systemu. Taka perspektywa badawcza ma charakter nowatorski i wpisuje się w aktualne tendencje rozwojowe lingwistyki obliczeniowej.

Wyniki wskazują, że reprezentacje tworzone przez model neuronowy wykazują jedynie częściowe podobieństwo do mechanizmów funkcjonujących w ludzkim systemie językowym. Niektóre efekty związane z frekwencją wyrazową są reprodukowane stosunkowo dobrze, podczas gdy inne zależności pozostają słabo zaznaczone lub nie występują wcale.

Interpretacja tych rezultatów została przeprowadzona w sposób niezwykle kompetentny. Autor unika uproszczonych tez dotyczących „naśladowania” ludzkiego poznania przez sztuczną inteligencję, wskazując raczej na funkcjonalne podobieństwa i różnice wynikające z odmiennych zasad organizacji systemów biologicznych i obliczeniowych.

Na uwagę zasługuje również umiejętność integrowania wyników eksperymentalnych z szerokim kontekstem teoretycznym. Dyskusja dotycząca reprezentacji rozproszonych, mechanizmów uwagi „attention mechanisms” oraz wielowymiarowych przestrzeni embeddingowych świadczy o bardzo dobrym przygotowaniu Autora do prowadzenia badań interdyscyplinarnych.

11. Znaczenie uzyskanych rezultatów

Rezultaty przedstawione w rozprawie posiadają istotne znaczenie zarówno dla badań podstawowych, jak i potencjalnych zastosowań praktycznych.

Po pierwsze, praca wnosi nową wiedzę dotyczącą funkcjonowania języka polskiego. Uzyskane wyniki pozwalają lepiej zrozumieć relacje między organizacją słownika umysłowego, a realizacją fonetyczną jednostek leksykalnych.

Po drugie, rozprawa stanowi ważny wkład do badań porównawczych prowadzonych na różnych językach. Pokazuje ona, że część prawidłowości obserwowanych w języku angielskim ma charakter uniwersalny, podczas gdy inne zależą od właściwości konkretnego systemu językowego.

Po trzecie, wyniki badań nad modelami sztucznej inteligencji posiadają znaczenie dla rozwoju technologii przetwarzania mowy. Identyfikacja różnic między systemami neuronowymi a ludzką produkcją językową może przyczynić się do projektowania bardziej naturalnych modeli generatywnych.

Wreszcie, rozprawa otwiera nowe kierunki badań nad relacjami między lingwistyką eksperymentalną a sztuczną inteligencją. Stanowi przykład twórczego wykorzystania narzędzi obliczeniowych do rozwiązywania klasycznych problemów językoznawczych.

12. Uwagi krytyczne

Wysoka ocena rozprawy nie oznacza oczywiście całkowitego braku uwag krytycznych. Mają one jednak charakter dyskusyjny i nie wpływają na moją jednoznacznie pozytywną ocenę całości pracy.

Po pierwsze, można rozważyć dalsze zwiększenie liczebności próby badawczej w przyszłych projektach. Chociaż liczba uczestników jest wystarczająca dla analiz fonetycznych tego typu, szerszy materiał empiryczny umożliwiłby bardziej szczegółowe badanie różnic indywidualnych.

Po drugie, pewien niedosyt budzi ograniczenie analiz modeli generatywnych do dwóch komercyjnych systemów. Interesującym rozwinięciem badań byłoby uwzględnienie większej liczby modeli otwartoźródłowych, umożliwiających dokładniejszą kontrolę parametrów treningowych.

Po trzecie, część rozważań dotyczących przyszłości sztucznej inteligencji oraz perspektyw rozwoju systemów AGI wykracza niekiedy poza bezpośredni zakres badań empirycznych przedstawionych w rozprawie. Fragmenty te mają jednak charakter inspirujący i nie zakłócają zasadniczej argumentacji.

Po czwarte, niektóre partie teoretyczne mogłyby zostać nieco skrócone na rzecz jeszcze szerszej prezentacji wyników eksperymentalnych. Uwaga ta dotyczy zwłaszcza rozbudowanych wprowadzeń do architektur sieci neuronowych.

Podkreślam jednak wyraźnie, że wszystkie przedstawione uwagi mają charakter drugorzędny i nie wpływają na wysoką wartość naukową rozprawy.

13. Ocena końcowa

Przedłożona rozprawa stanowi dojrzałe, samodzielne i oryginalne osiągnięcie naukowe. Autor wykazał się bardzo dobrą znajomością literatury przedmiotu, wysokimi kompetencjami metodologicznymi oraz umiejętnością prowadzenia badań interdyscyplinarnych łączących językoznawstwo i sztuczną inteligencję.

Do najważniejszych osiągnięć Doktoranta zaliczam:

- podjęcie aktualnej i oryginalnej problematyki badawczej;
- wypełnienie istotnej luki dotyczącej badań nad językiem polskim;
- opracowanie spójnego i wieloetapowego projektu eksperymentalnego;
- poprawne zastosowanie nowoczesnych metod analizy akustycznej i statystycznej;
- twórcze wykorzystanie narzędzi sztucznej inteligencji w badaniach językoznawczych;
- ostrożną i kompetentną interpretację uzyskanych rezultatów.

Rozprawa wnosi nową wiedzę do dyscypliny językoznawstwo oraz posiada znaczący potencjał dalszego rozwoju badawczego.

14. Konkluzja

Stwierdzam, że rozprawa doktorska mgr. Stevena Jarosza pt. *A Neural Network Approach to Modeling the Acoustic Effects of Phonological Neighborhood Density and Word Frequency in Polish Speech* spełnia wymagania stawiane rozprawom doktorskim określone w obowiązujących przepisach prawa, tj. art. 187. p. 1-2 ustawy Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 roku (Dz.U.2024.1571).

oraz odpowiada wysokim standardom współczesnej, oryginalnej pracy naukowej.

Autor wykazał się umiejętnością samodzielnego prowadzenia badań naukowych, formułowania oryginalnych problemów badawczych, właściwego doboru metod badawczych oraz krytycznej interpretacji uzyskanych wyników. Rozprawa stanowi istotny wkład w rozwój badań nad fonetyką języka polskiego, psycholingwistyką i zastosowaniami sztucznej inteligencji w językoznawstwie.

W mojej ocenie praca w pełni spełnia wymogi stawiane rozprawom doktorskim i **zasługuje na przyjęcie oraz dopuszczenie Autora do dalszych etapów postępowania w sprawie nadania stopnia doktora.**

W związku z powyższym **wnoszę o przyjęcie rozprawy doktorskiej mgr. Stevena Jarosza bez zastrzeżeń.**



.....
dr hab. Tomasz Ciszewski, prof. UG