

Wrocław/Opole, 30 maja 2023 r.

prof. dr hab. Łukasz Grabowski

Instytut Językoznawstwa

Uniwersytet Opolski

Recenzja rozprawy doktorskiej Łukasza Wiłkonია pt. *Natural language processing and communication models in chosen crisis situations / Modele komunikacji oraz przetwarzanie języka naturalnego w wybranych sytuacjach kryzysowych* (ss. 362) przygotowanej pod kierunkiem prof. dr. hab. Rafała Molenckiego oraz dr. hab. inż. Krzysztofa Simińskiego, prof.

Pol. Śl.

Uwagi wstępne

Przedstawiona do recenzji rozprawa doktorska z pogranicza językoznawstwa i przetwarzania języka naturalnego (NLP), a także czerpiąca szeroko z dorobku psychologii, pragmatyki językowej, danologii i kryminalistyki, stanowi ważny wkład w dotychczasowe badania nad komunikacją w sytuacjach kryzysowych. Wyniki szeroko zakrojonej analizy zapisu negocjacji policyjnych z Grantem Sattaurem (z wykorzystaniem dla celów porównawczych większego korpusu transkrypcji negocjacji policyjnych podczas 51-dniowego oblężenia kompleksu Mount Carmel w okolicach miasta Waco w Teksasie) dostarczyły szczegółowego opisu cech językowych, przede wszystkim zdań nacechowanych pozytywnie i negatywnie, tonu/sentymentu wypowiedzi, językowych wyznaczników mowy nienawiści vs. języka obraźliwego, religijności i myśli samobójczych, przykładów metaforycznego użycia języka, a także językowych wyznaczników strategii negocjacyjnych, będących mieszanką zarówno tzw. technik miękkich, jak i twardych. Każda mikro-analiza została przeprowadzona z wykorzystaniem specjalnie dobranych danych treningowych oraz metod badawczych z obszaru NLP. Wyniki dysertacji torują drogę do dalszych badań nad językiem negocjacji, zarówno opisowych, jak i stosowanych, a Autor – zdając sobie sprawę z ograniczeń

metodologicznych i materiałowych przeprowadzonego badania – formułuje skonkretyzowane propozycje jego kontynuowania w przyszłości.

Ze względu na moją specjalność naukową w niniejszej recenzji ocenę dysertację głównie z perspektywy językoznawstwa korpusowego i pragmatyki językowej, jednak postaram się zwrócić także uwagę na jej mocniejsze i słabsze strony związane z warsztatem naukowym.

Struktura rozprawy

Recenzowana dysertacja liczy 362 strony; składa się ze streszczenia w języku polskim, wykazu licznych skrótowców wykorzystanych w pracy, wstępu, 9 rozdziałów (w tym czterech teoretycznych i jednego obszernego rozdziału metodologiczno-empirycznego), podsumowania, bibliografii (liczącej 64 strony!), a także załączników składających się na opis materiału badawczego.

Teoretyczna część pracy, obejmująca rozdziały 1-4, stanowi wprowadzenie do kluczowych zagadnień. I tak w rozdziale 1 odnajdujemy szczegółowe omówienie zagadnień z takich dziedzin, jak studia nad bezpieczeństwem, komunikologia i mediacja, a także negocjacje, ze szczególnym naciskiem na ich specyficzność w sytuacjach kryzysowych. Rozdział ten cechuje szczegółowe omówienie typologii technik i strategii negocjacyjnych osadzone w bogatej literaturze specjalistycznej. W rozdziale 2 Autor omawia najważniejsze zagadnienia związane z technikami i metodami przesłuchań stosowanymi przez organy ścigania. Osadzony w psychologii i językoznawstwie rozdział 3 przybliży typologie i kategorie emocji spotykane w literaturze specjalistycznej, językowe wykładniki emocji w bezpośredniej i zapośredniczonej komunikacji językowej, a także traktuje o wpływie emocji na proces i wynik negocjacji. Rozdział 4 kieruje uwagę czytelnika w stronę zagadnień ściśle językoznawczych, omawia kluczowe pojęcia z zakresu pragmatyki językowej (m.in. spójność logiczną i znaczeniową, zasadę współpracy, analizę konwersacyjną i zmianę ról w procesie komunikacji, teorię grzeczności, koncepcję aktów mowy) oraz krytycznej analizy dyskursu. Wspomniana część zasługuje na pochwałę ze względu na to, że Autor nie tylko syntetyzuje i definiuje wszystkie pojęcia, ale również precyzyjnie je operacjonalizuje dla potrzeb swojego badania.

Pierwszą część rozdziału 5 poświęcono metodologii automatycznej analizy tekstu. Autor dokładnie omawia wybór danych treningowych dla poszczególnych etapów badania: analizy sentymentu, identyfikacji emocji, analizy językowych wyznaczników toksyczności tekstu, treści samobójczych, mowy nienawiści i obraźliwego języka, sarkazmu, metaforyczności i perswazji. Następnie omówione zostały narzędzia przetwarzania języka, etapy przygotowania materiału badawczego (tokenizacja i lematyzacja danych językowych, opracowanie stop-listy, a także oczyszczanie i normalizacja danych językowych). Omówiono również narzędzia do automatycznej klasyfikacji tekstu i analizy eksploracyjnej. W podrozdziale 3 syntetycznie omówiono relewantne metody przetwarzania języka naturalnego, ze szczególnym naciskiem na tzw. zanurzenia słów (word2vec, doc2vec i GloVe), kończąc na dyskryminacyjnych metodach uczenia maszynowego (m.in. naiwny klasyfikator Bayesa, metoda wektorów nośnych (SVM) i regresja logistyczna) oraz uczenia głębokiego (począwszy od konwolucyjnych i rekurencyjnych sieci neuronowych po tzw. transformery i ich różne odmiany). Na podstawie tego szczegółowego omówienia w podrozdziale 5 Autor opisuje doprecyzowaną metodologię wykorzystaną w swoim badaniu, uzasadniając wybór modeli językowych i miar ich jakości.

Należy stwierdzić, że Autor syntetycznie i precyzyjnie omówił wszystkie wybrane problemy wykazując się doskonałą znajomością literatury specjalistycznej z różnych dziedzin nauki, w tym tej najnowszej, zarówno anglojęzycznej, jak i polskojęzycznej. Niewątpliwie na pochwałę zasługuje umiejętność syntezy informacji przez Autora, a także klarowność wyводу.

Ostatni podrozdział 5 to już przedstawienie szczegółowych wyników analiz klasyfikacyjnych wykonanych na materiale zapisów negocjacji policyjnych z Grantem Sattaurem, obejmujący – jak wspomniano wyżej – analizę sentymentu, identyfikację emocji, tekstowych wykładników toksyczności, sarkazmu, metaforycznego języka, perswazyjności, religijności, mowy nienawiści i obraźliwego języka oraz treści samobójczych.

Dysertację wieńczą konkluzje, gdzie Autor podkreśla również ograniczenia przeprowadzonego badania.

Poniżej zwrócę uwagę na wybrane przeze mnie cechy dysertacji, które świadczą o jej oryginalności i mocnych stronach, a także przedstawię swoje uwagi krytyczne, podkreślając zauważone niedociągnięcia i braki; skomentuję również językową stronę pracy.

Oryginalność recenzowanej dysertacji doktorskiej

W niniejszej recenzji chciałbym podkreślić kilka aspektów ocenianej rozprawy, które nadają jej twórczy i oryginalny charakter:

- Wyczerpująca, wszechstronna i przemyślana synteza teoretyczna w rozdziałach 1-4 odnosząca się do kluczowych pojęć i koncepcji wykorzystanych w dysertacji;
- Wyczerpujące i syntetyczne przedstawienie bogatego zestawu dostępnych danych treningowych do wykorzystania w różnych zadaniach NLP (Rozdział 5, podrozdział 1) oraz narzędzi i bibliotek do przetwarzania i analizy tekstów cyfrowych (podrozdział 2);
- Praca potwierdza umiejętności Autora nie tylko w zakresie syntezy wiedzy językoznawczej w celu zdefiniowania wykorzystanych w dysertacji pojęć, ale także ich operacjonalizacji dla potrzeb przeprowadzonego badania zapisów negocjacji policyjnych;
- Opracowanie dwóch reprezentatywnych korpusów wykorzystanych w badaniu, a także zaadoptowanie wybranych korpusów treningowych na potrzeby badania, które – jeśli zostaną udostępnione w domenie publicznej – mogą stać się cennym materiałem badawczym, umożliwiając nie tylko weryfikację uzyskanych wyników, ale również prowadzenie dalszych badań w przyszłości (zwłaszcza pod kątem dalszej weryfikacji optymalnych modeli językowych dla poszczególnych zadań NLP).
- Uzyskanie dobrych wyników (lepszych niż dla modelu referencyjnego) w zakresie wykrywania mowy nienawiści w tekstach (precyzja = 0.88, czułość = 0.91).
- Z perspektywy językoznawczej (humanistycznej) niewątpliwym walorem części empirycznej pracy są – prócz zróżnicowanych danych ilościowych – przykłady odpowiednio sklasyfikowanych zdań: czytelnik przekonuje się, że za gąszczem wykresów i metryk kryje się treść, która świadczy o specyfice języka negocjacji policyjnych (np. przykłady zdań ilustrujących myśli samobójcze lub depresyjne na str. 271).
- Autor jest świadomy ograniczeń przeprowadzonego badania: na str. 284 wyraźnie podkreśla, że analiza jest przeprowadzona na poziomie zdania (szerokość kontekstu przyjęta w badaniu) i pomija sygnały zarówno para-werbalne, jak i pozawerbalne. Autor zdaje sobie również sprawę (str. 285), że przygotowanie danych treningowych,

w zwłaszcza nadawanie etykiet różnym cechom językowym, jest wysoce subiektywne (np. przy przygotowaniu danych treningowych do wykrywania mowy nienawiści, obraźliwego języka, czy wyznaczników treści samobójczych) i może wpływać na wyniki uzyskane w badaniach (podkreśla się fakt, że znaczenie jest emergentne w określonym kontekście).

Uwagi do rozprawy doktorskiej

W trakcie lektury rozprawy zauważyłem nieliczne fragmenty (większość w rozdziale 5), w odniesieniu do których chciałbym wyrazić swoje uwagi. Większość z nich ma jedynie charakter polemiczny i nie rzutuje na pozytywną ocenę przedstawionej do recenzji rozprawy.

- W niektórych fragmentach pracy Autor dość pobieżnie omawia wykorzystanie niektórych koncepcji językoznawczych: np. na str. 215 czytamy jedynie o trzech – odnoszę wrażenie, że przypadkowo wybranych – przykładach wykorzystania n -gramów (sekwencji n znaków lub wyrazów) w badaniach językoznawczych, pomijając ich szerokie wykorzystanie w badaniach nad atrybucją autorstwa, przyswajaniem języka drugiego, badaniach literaturoznawczych itp. Podobną przypadkowość widać na str. 220 w sekcji 3.1.1: Autor w jednym zdaniu przytacza rys historyczny perceptronów jedno- i wielowarstwowych, po czym wspomina o jednym narzędziu (TextBlob) wykorzystującym perceptron jedno-warstwowy. Podobnie rzecz się ma w wypadku omówienia różnych metod regresji (str. 223), gdzie Autor nie przytacza żadnych badań językoznawczych wykorzystujących np. regresję liniową lub logistyczną.
- Bardzo rzadko, ale jednak, zdarzają się – jak mierniam – niezawinione błędy merytoryczne wynikające zapewne z przejęzyczeń, np. na str. 216 możemy zapoznać się z nieco zmienioną wersją hipotezy dystrybucyjnej Zelliga Harrisa („in which terms that occur in a similar sense have a similar meaning”), gdzie słowa, które występują w podobnym sensie mają podobne znaczenie (zapewne chodzi tutaj o występowanie w podobnym kontekście).
- Autor przedstawia (str. 217-218) liczne typologie wektorowych osadzeń/zanurzeń słów (*word embeddings*) w zależności od stopnia wytrenowania danych, czasu powstania względem opracowania modeli LSTM, kontekstowości. Odnosi się wrażenie, że sekcja 3.8 nagle się urywa: Autor nie wyjaśnia, a jedynie wzmiankuje

fakt, że tzw. *quick-thought vectors* są rozwinięciem wcześniejszej koncepcji *skip-thought vectors* opartej na skip-gramach. Niekiedy ma się również wrażenie, że Autor dosłownie zasypuje czytelnika szczegółową wiedzą (np. wyliczenie przykładów algorytmów wzmocnienia gradientowego na str. 221), choć nie wiadomo do końca, jak przekłada się ona na przeprowadzone badanie.

- Czasem brakuje precyzji w opisie: na str. 221 czytamy, że wnioskowanie bayesowskie pierwszy raz wykorzystano w pracy Mostellera i Wallace'a z 1964 r.: dobrze byłoby wspomnieć, że chodziło o studium atrybucji autorstwa zbioru felietonów pt. *Federalist papers* z 1787 r., gdzie wykazano, że obiecującą cechą różnicującą autorów są wyrazy synsemantyczne: rodzajniki, spójniki, przyimki, partykuły i niektóre zaimki osobowe. Na kolejnej stronie znajdujemy też ogólne stwierdzenie w rodzaju: „BERT focuses on semantics, and NB-SVM focuses on linguistics”, choć nie wiadomo, co się za tym kryje – podane przykłady wykorzystania omówionych rozwiązań zapewne dotyczą zadań klasyfikacyjnych. Reasumując: większa precyzja opisu i/lub odfiltrowanie informacji niebezpośrednio związanych z badaniem, jak np. krótkie rysy historyczne, np. o SVM czy regresji liniowej, byłoby korzyścią dla czytelnika.
- Opisująca metodologię sekcja 4 (w rozdziale 5) zyskałaby na jakości po dodaniu krótkiego wprowadzenia wyliczającego poszczególne etapy badania: czytelnik zostaje od razu „wrzucony” w szczegółowe parametry dzielenia danych na treningowe, testowe i walidacyjne.
- Z korzyścią dla czytelnika byłoby krótkie wyjaśnienie zalet wykorzystania w badaniu biblioteki ktrain, zwłaszcza tzw. *rapid model deployment* (Czy chodzi tylko o oszczędność czasu i mocy obliczeniowych?) oraz *cased version of BERT and XLNet* (Czy chodzi tylko o zachowanie pisowni małych i wielkich liter przekładające się na lepsze działanie modelu językowego?).
- W ocenie jakości modeli językowych istotny jest wybór odpowiednich metryk, które Autor definiuje na str. 236-238. Myślę, że o wiele większe walory poznawcze miałyby krótka informacja o celowości zastosowania każdej z metryk (tj. co ona w praktyce pokazuje/co dzięki niej widać/co zyskujemy, a co tracimy z pola widzenia), np. wysoka precyzja przekłada się na mniejszą liczbę danych sklasyfikowanych jako fałszywie pozytywne; wysoka czułość oznacza zminimalizowanie danych fałszywie negatywnych; wysoka dokładność przekłada się na poprawne zaklasyfikowanie jak największej ilości danych prawdziwie pozytywnych i negatywnych: wskaźnik F1

znajduje „złoty środek” między precyzją i czułością modelu i jest pewną alternatywą dla dokładności itp. Jest to o tyle istotne, że w poszczególnych etapach badania Autor kładzie nacisk na różne metryki, a także stosuje próg 0.9 (choć eksplicytnie nie napisano, czy ów próg dotyczy dokładności/*accuracy*, czy też innej metryki). Opis wykresów nr 4, 5, 6 (sekcje 5.1-5.3) nie precyzuje, jakiej metryki dotyczy próg 0.9; tabela nr 17 na str. 252 podpowiada, że najprawdopodobniej chodzi o dokładność. Natomiast tabela nr 20 na str. 256 może wprowadzać błąd: Autor pisze nad tabelą o porównaniu wskaźników F1, a w opisie tabeli czytamy o porównaniu dokładności modeli (chodzi o etap badania: wykrywanie metafor).

- W pracy zorientowanej językoznawczo dobrze byłoby umieścić pełniejsze interpretacje uzyskanych wyników, np. w wypadku wykrywania metafor, gdzie zidentyfikowano jedynie 37 takich zdań, czy można stwierdzić, że język negocjacji policyjnych jest raczej dosłowny i prosty, zdominowany przez przejrzyste i jasne komunikaty?
- Jeśli idzie o wykrywanie perswazji to dobrze byłoby wyjaśnić, dlaczego Autor podzielił dane treningowe i testowe (str. 259) w proporcjach 90%/10% (choć wcześniej – str. 231 – pisał o zwyczajowym podziale w proporcjach 80%/20%). Czy mogło to mieć wpływ na ocenę jakości modeli?
- Czy Autor – ze względu na ograniczenia i niepełne informacje (str. 271) o narzędziu Google służącym do analizy sentymentu – planuje zweryfikować uzyskane wyniki w oparciu o inne narzędzia, np. opracowane przez konsorcjum CLARIN-PL (Sentemo czy Multiemo) lub jeszcze inne?

Ocena strony językowej i edycyjnej rozprawy

Strona językowa i redakcyjną recenzowanej dysertacji jest wysoka – Autor posługuje się dojrzałą, zwięzłą, precyzyjną angielszczyzną na poziomie akademickim. Rozprawa jest również dobrze przygotowana pod względem edycyjnym. Z obowiązku recenzenckiego jestem zobligowany zwrócić uwagę na incydentalne niedociągnięcia:

str. 218 – skip-though vectors ...(zamiast: skip-through)

str. 223 – Chines origin ...(zamiast: Chinese)

str. 249 – covered... (zamiast: covered)

str. 273 – give class... (zamiast: given class)

str. 282 – It must be noted that toxic question classification **does not classifies** only questions (zamiast: does not classify)

W pracy zdarzają się okazjonalne usterki stylistyczne: np. str. 274: „The goal of this work was a description of the language used during crisis interventions through artificial intelligence and linguistics” (być może lepiej: “using research methods developed in the field of artificial intelligence and linguistics”).

Wnioski

Moim zdaniem rozprawa doktorska Łukasza Wiłkonja pt. *Natural language processing and communication models in chosen crisis situations*, przygotowana pod kierunkiem prof. prof. Rafała Molenckiego i Krzysztofa Simińskiego, spełnia wszystkie wymagania stawiane dysertacjom doktorskim. Praca wskazuje na wysoki poziom wiedzy teoretycznej i metodologicznej Autora, a także na umiejętność prowadzenia pracy naukowej na pograniczu językoznawstwa i informatyki (tj. w obszarze przetwarzania języka naturalnego). W związku z powyższym jednoznacznie wnioskuję o dopuszczenie Autora do dalszych etapów postępowania doktorskiego.



/prof. dr hab. Łukasz Grabowski/