



**Uniwersytet im. Adama Mickiewicza
w Poznaniu**
Zakład Lingwistyki Informatycznej i Sztucznej Inteligencji
PL-61-614 Poznań, ul. Umultowska 87, tel. +48-61-8295380, fax +48-61-8295315
Kierownik Zakładu: prof. zw. dr hab. Zygmunt Vetulani
<http://amu.edu.pl/~vetulani>
vetulani@amu.edu.pl

Recenzja rozprawy doktorskiej mgr Magdaleny Igras-Cybulskiej pt. "Analysis of non-linguistic content of speech signals"

Do oceny przedłożona została rozprawa w języku angielskim w postaci tekstu, którego część merytoryczna obejmuje 169 strony (czcionka Times 10, interlinia pojedyncza). Na tekst składa się osiem rozdziałów merytorycznych, bibliografia (342 pozycje) i dodatki (w tym zestawienie osiągnięć Doktorantki).

W rozprawie dotyczącej (teoretycznych) podstaw technologii mowy Autorka koncentruje się na warstwie paralingwistycznej i pozalingwistycznej sygnału mowy. W tym zakresie rozważania skierowane są na ekstrakcję informacji dotyczącej mówcy, w tym także informacji o jego stanie emocjonalnym i stosunku do wypowiedzianej w akcie mowy treści komunikatu. Podstawowa informacja lingwistyczna poznawalna w drodze analizy morfologicznej, składniowej, semantycznej i pragmatycznej nie obejmuje ważnych własności mówcy i nie informuje o sposobie realizacji akustycznej aktu mowy. Tym samym często nie wystarcza do pełnej interpretacji wypowiedzi. Uzupełnieniem jest informacja paralingwistyczna i poza-lingwistyczna (zawarta w tych elementach, których nie obejmuje zawartość słowna) dotycząca sposobu ekspresji treści (segmentowanie granic fraz, akcentowanie), jak też relacji wypowiadającego do treści i odbiorcy (emocje, ironia itp.). Posługując się metodami analizy ilościowej i jakościowej Doktorantka dąży do wyodrębnienia parametrów informujących o strukturze czasowej mowy i o indywidualnych cechach aktu mowy charakterystycznych dla mówcy w momencie wypowiedzi (np. w

powiązaniu ze stanem emocjonalnym). Badania prowadzone są na materiale empirycznym w postaci korpusów mowy (Rozdział 4).

W rozprawie można wyróżnić część wprowadzającą (rozdziały 1 do 5, od początku do str. 6) oraz część zasadniczą (rozdziały 6 i 7, str. od 61 do 149) zawierającą wyniki oryginalne, zaś Rozdział 8. stanowi jej podsumowanie. Dodatkowo, w Rozdziale 9., Doktorantka zestawiała swoje osiągnięcia naukowe. Całość zamyka obszerna bibliografia (str. 155-169). O ile Rozdział 1. (Introduction) jest standardowym wprowadzeniem bez treści merytorycznej, to kolejny Rozdział 2. zawiera ważne dla dalszej lektury omówienie pojęcia *informacji paralingwistycznej* w realizacji aktów mowy. Biorąc pod uwagę fakt, że granice pomiędzy tymi pojęciami nie są w literaturze dokładnie ustalone, Doktorantka precyzuje, że termin „paralingwistyczny” odnosi się do formy mowy i jego prozodii i jakości („*paralinguistic would refer to the form of speech and language, its prosody and quality*”), zaś termin „nie-lingwistyczny” do mówcy, jego cech i stanów przenoszonych przez sygnał mowy („*non-linguistic would refer to the information on the speaker, speaker's traits and states, conveyed in the speech signal*”) (str. 23). To rozróżnienie znajduje odbicie w konstrukcji głównej i najobszerniejszej części rozprawy, którą stanowi Rozdział 6. Jej część pierwsza (6.1) poświęcona jest zjawiskom paralingwistycznym, zaś część druga (6.2), mniej obszerna, wybranym zagadnieniom głosowej ekspresji emocji i stosunku do mówcy i treści wypowiedzi (por. uwagi dotyczące tytułu rozprawy w części formalnej recenzji). W dalszym ciągu tego rozdziału, Autorka uzasadnia tematykę badań przez wskazanie (2.3.) licznych obszarów zastosowań, którymi są: interfejsy człowiek-komputer, modelowanie mowy dla systemów automatycznego rozpoznawania sygnału mowy, NLP, językoznawstwo deskryptywne, biometria mówcy, synteza mowy, zastosowania biomedyczne. Część ta kończy się krótkim wprowadzeniem do komputerowej analizy sygnału mowy poprzez omówienie metod parametryzujących (wykorzystywanie określonych, powszechnie stosowanych parametrów), metod statystycznych, klasyfikujących (sieci neuronowe, drzewa decyzyjne, HMM, ...), metod oceny jakości wyników badań. Rozdział kolejny (trzeci) to obszerny przegląd literatury dotyczącej aktualnych badań odnośnie do poszczególnych kategorii paralingwistycznych będących przedmiotem badań Autorki, takich jak: pauzy, struktura segmentacyjna, ograniczniki zdań i fraz, akcent, jakość głosu, barwa. W odniesieniu do cech nie-

lingwistycznych omawiane są pozycje dotyczące stanów emocjonalnych, ironii, nie-lingwistycznych zachowań grzecznościowych. Niezależnie od oryginalnych wyników badań Autorki opisywanych w rozdziale 6, obszerny przegląd literatury przeprowadzony w Rozdziale 3. może być szeroko przydatny dla badaczy i inżynierów mowy.

Kolejne dwa rozdziały wprowadzają czytelnika bliżej w aparat narzędziowy pracy. W Rozdziale 4. omówione są bazy danych (korpusy mowy) zbudowane z nagrań i ich transkrypcji tekstowych wraz z metadanymi. Są to dane empiryczne o różnym charakterze, różnej wielkości i metodologii pozyskania. Różnią się także naturą materiału lingwistycznego: laboratoryjnie utworzony korpus mowy emocjonalnej, korpus nagrań z telefonu ratunkowego, nagrania z youtube, nagrania studentów emisji głosu (voice emission students), korpus próbek barwy dźwięku, korpus monologów (dla badań nad zjawiskiem pauzy). Doktorantka zdaje sobie sprawę z nierównego charakteru dostępnych materiałów empirycznych i niejednokrotnie daje temu wyraz w rozprawie. Część wprowadzającą uzupełnia krótki rozdział (5 str.) omawiający wybrane przez Doktorantkę metody analizy falkowej i falkowo-fourierowskiej. Rozdział ten jest bardzo techniczny i jego czytelność będzie ograniczona do grona specjalistów już obznajomionych z podstawami tych narzędzi.

Twórczą część pracy stanowi w pierwszym rzędzie Rozdział 6 (str. od 61 do 138) poświęcony zjawiskom paralingwistycznym i nie-lingwistycznym obserwowanym dla języka polskiego. Składa się on z dwóch podrozdziałów 6.1. i 6.2. Rozdział obszerniejszy (6.1) odnosi się do zjawisk paralingwistycznych, takich jak pauzy, struktura segmentalna sygnału mowy, granice zdań, akcent i barwa głosu. Część nie-lingwistyczna jest poświęcona identyfikowaniu cech emocji, ironii i elementów grzecznościowych w sygnale mowy. Badania nad pauzami, którym Autorka poświęciła najwięcej uwagi, prowadzone były w oparciu o nagrania o charakterze monologów o różnym stopniu integralności (mowy parlamentarne, czy teksty wypreparowane z wywiadów). Generalną konkluzją tych rozważań jest stwierdzenie, że wnoszą one istotną informację odnośnie do przyzwyczajęń językowych mówiącego, w tym informacje biometryczne pomocne w automatycznej identyfikacji mówcy. Autorka zwraca też uwagę na fakt, że automatyczne wykorzystywanie pauz akustycznych jest przydatne do polepszenia algorytmów przetwarzania języka, w szczególności systemów transkrypcji mowy na tekst i

parserów mowy (pauzy zawierają informacje składniowe podobnie jak znaki interpunkcyjne w kodzie pisanym). Swoje prace nad pauzami w sygnale mowy Autorka ocenia jako „obiecujące” i w podsumowaniu wskazuje na dalsze kierunki ich kontynuowania. Poza pauzą, jako możliwym markerze, Autorka w kolejnym rozdziale (6.1.2) bada akustyczne realizacje końca zdania oraz akcent. Tę część badań realizuje w oparciu o dwa korpusy nagranych zdań. Pierwszy z nich obejmuje prawie 100 000 głosek, zaś drugi dziesięciokrotnie mniej (10 000). Przyznając, że jakość dostępnych danych empirycznych nie jest najlepsza (str. 93), Autorka stwierdza jednak, że obserwowane parametry (czas trwania, energia i moc) na tyle wyróżniają mówców, iż można je uznać za przydatne w automatycznej identyfikacji mówców i syntezie mowy polskiej. Specjalną uwagę poświęca Autorka zjawiskom występującym na końcu zdania (6.1.3). Obserwuje ona korelację pomiędzy wydłużeniami głosek a ich pozycją przed przecinkiem lub kropką końca zdania. Przy stwierdzeniu tego zjawiska Autorka zaznacza jednak, że jego nasilenie jest indywidualnie zależne od mówcy (str. 97). Metody podobne do zastosowanych w badaniach nad długością głosek w końcówkach zdań zostały także zastosowane przez nią w studiach nad zmiennością energii i mocy (malejących przed końcem zdania).

Przeprowadzone badania podstawowe w zakresie zjawisk paralingwistycznych w obrębie końca zdania są podsumowane relacją z eksperymentów w zakresie automatycznego wykrywania końca zdania bądź frazy. Prace te można uważać za pionierskie i obiecujące, choć jak zaznacza Autorka, wymagają one dalszych studiów i eksperymentów na bardziej wymagającym materiale empirycznym (str. 110). Nie ulega wątpliwości, że wnioski, nawet wtedy gdy są spodziewane, Autorka wywodzi z bardzo rzetelnej analizy danych, odwołując się do wyrafinowanych metod statystycznych.

Prace nad akcentem (6.1.4) i barwą (6.1.5) dźwięków dopełniają palety zainteresowań Autorki w zakresie paralingwistyki. Zaznacza ona, iż uzyskane modele probabilistyczne były budowane dla oceny zjawisk znanych z literatury fonetycznej („to evaluate acoustically the phenomena known from phonetic literature”), jak np. istotny wzrost energii dla sylab akcentowanych. Stwierdzając ogólną zgodność poczynionych przez siebie obserwacji z obserwacjami innych badaczy, Doktorantka zaznacza jednak, że uważa problem akcentu w

języku polskim za ciągle otwarty i wymagający dalszych szczegółowych badań, np. z uwzględnieniem różnych rodzajów akcentu. W rozdziale 6.1.5 Autorkę interesują semantyczne aspekty rozróżnień dotyczących barwy. Podkreśla ona pilotażowy charakter przeprowadzonych prac. Zwraca przy tym uwagę, że zagadnienie barwy dźwięków, o ile jest często podnoszone w odniesieniu do brzmienia instrumentów muzycznych, to znacznie rzadziej jest przedmiotem badań nad dźwiękami mowy. Ta część rozprawy w zasadzie ogranicza się do zasygnalizowania zainteresowań Autorki i zapowiedzi prac szczegółowych.

W zakresie zainteresowań zjawiskami nie-lingwistycznymi rozprawa poświęcona jest trzem wybranym grupom o rozmaitym potencjale praktycznym, a mianowicie wykładnikom emocji, ironii i nie-lingwistycznym formom grzecznościowym. Jako materiał empiryczny do badań nad wyrażaniem emocji Autorka wykorzystuje korpus nagrań zebranych w korpusie rozmów telefonicznych z serwisem interwencyjnym (typu telefon 112). O ile dobór tego materiału jest atrakcyjny przez fakt spodziewanego dużego ładunku emocjonalnego, o tyle ten sam czynnik sprawia, że jest on obciążony przewagą emocji negatywnych nad pozytywnymi, która nie występuje w mowie neutralnej. Po wprowadzeniu w zagadnienia analizy emocji, Autorka omawia trzy rodzaje eksperymentów automatycznego wykrywania emocji, do oceny których zastosowano różne narzędzia matematyczne: analizę falkową, analizę falkowo-fourierowską i modele gaussowskie. W tej części rozprawy, pewien niedosyt budzić może brak wyraźnie wyciągniętych wniosków z przedstawionych eksperymentów.

Kolejnym z omawianych problemów nie-lingwistycznych jest automatyczna identyfikacja parametrów wokalnych ironii w języku polskim. W tym wypadku Autorka przeprowadziła interesujące studium porównawcze z innymi językami wskazując na liczne podobieństwa (angielski, chiński, francuski, niemiecki, włoski). Badania te mają jednak charakter obiecujących prac wstępnych między innymi ze względu na stosunkowo niską liczbę obserwacji wziętych pod uwagę w części analitycznej.

Ostatni z podjętych tematów nie-lingwistycznych dotyczył identyfikacji akustycznych korelatów dla wyrażania uprzejmości (politeness). W tym zakresie zostały przeprowadzone wstępne badania. Mimo zróżnicowania wyników pomiędzy grupami eksperymentalnymi,

szczególnie dla kanadyjskiego (*noa bene* Autorka pozostawia domyślności czytelnika, czy chodzi o kanadyjski angielski, francuski czy może oba?), udało się zaobserwować pewne jednakowe tendencje w obu przypadkach. Np. elementy sygnalizujące uprzejmość posiadały dłuższy czas ekspresji. W tym fragmencie pracy brak jest szczegółowych rozważań technicznych a konkluzje Autorki nie są całkowicie przejrzyste.

O ile Rozdział 6, najobszerniejszy i najistotniejszy w ocenie dorobku dobrze odzwierciedla zakres zainteresowań i na ogół bardzo rzetelnie prezentuje wyniki teoretyczne, o tyle „kropką nad i” jest wieńczący pracę Rozdział 7 poświęcony dwóm aplikacjom powstałym z udziałem Autorki. Jest to po pierwsze system budowy profilu klienta serwisu ratunkowego typu „telefon 112” na podstawie analizy paralingwistycznej. Drugą aplikacją jest system treningu wymowy adresowany do profesjonalistów, dla których używanie mowy jest istotnym elementem pracy. Autorka zwraca przy tym uwagę na możliwe wykorzystanie istotnych elementów tego systemu do wsparcia automatycznego rozpoznawania mowy. Aplikacje te są ważnym przyczynkiem do pozytywnej oceny dorobku z punktu widzenia przydatności w pracach inżynierskich wykorzystujących badania podstawowe w obszarze technologii mowy.

Uwagi redakcyjne

Stwierdzam, iż pod względem redakcyjnym, słabe strony są nieliczne i nie rzutują na zdecydowanie pozytywny bilans całości. Pod względem językowym praca jest nierówna. Większość tekstu napisana jest poprawną i zrozumiałą angielszczyzną, choć zdarzają się fragmenty odbiegające od dopuszczalnego standardu (nawiązki), jak np. pierwszy akapit Rozdziału 4, co może oznaczać przeklejenie z różnych tekstów Autorki (lub po prostu niestaranną redakcję i niedostateczną (auto)korektę). (Str. 46: „They are usually very valuable resources, because of the time- and labor-consuming process of preparing the annotation.”). Usterki redakcyjne, jak np. brak dokończenia zdania otwierającego rozdział 2.3. (po dwukropku winno być wyliczenie, zamiast tego występują nienumerowane podrozdziały) są jednak bardzo rzadkie.

Nie stwierdziłem usterek ortograficznych czy istotnych błędów interpunkcyjnych. Do oceny walorów stylu Autorki nie czuję się kompetentny, poprzestanę na stwierdzeniu, że tekst czyta się dobrze. Niezależnie od tego można by się zastanawiać nad zasadnością wyboru języka angielskiego jako redakcyjnego, biorąc pod uwagę, że w obecnym kształcie rozprawa nie ma cech monografii przeznaczonej do publikacji międzynarodowej. Z kolei redakcja polska byłaby przyczynkiem do ugruntowania się terminologii polskiej, co jest często – a niesłusznie – ignorowane przy ocenie przydatności rozpraw i monografii.

Podsumowanie

Za główne osiągnięcie rozprawy uważam szczegółowe i rzetelne studium wybranych elementów akustycznych mowy jako wskaźników treści para-lingwistycznej mowy oraz, w mniejszym zakresie, źródło informacji o wybranych stanach mówcy (emocja, nastawienie treści komunikatu mowy czy też do rozmówcy). Autorka ma ponadprzeciętną orientację w literaturze przedmiotu, a wykaz źródeł jest imponujący.

Autorka wykazuje się dużą wiedzą techniczną i erudycją z szeregu zagadnień szczegółowych, które były przedmiotem jej własnych badań. Nie sposób jednak nie dostrzec, że chęć przedstawienia całości prac Doktorantki w zakresie tytułowym sprawia, z konieczności, że rozprawa robi wrażenie nierównej, szczególnie tam gdzie mowa jest o problemach będących na etapach badań początkowych (np. w zakresie niektórych problemów poza-lingwistycznym).

Praca ma układ bardzo przemyślany, dojrzały i zorientowany na czytelnika. Obszerna (i bardzo przydatna) część wprowadzająca ma przemyślaną strukturę, niewątpliwie ułatwiającą czytelnikowi koncentrację na części zasadniczej zawierającej wyniki własne. Prześledzenie bardzo technicznych wywodów szczegółowych ułatwiają liczne zestawienia syntetyczne w postaci tabel, diagramów i wykresów.

Brakiem formalnym utrudniającym dokładne studiowanie rozprawy jest jednak brak indeksu rzeczowego. Na utrudnienie śledzenia przywoływanych źródeł ma wpływ także niefortunna konwencja odsyłania do bibliografii przez ograniczenie się do liczby porządkowej w nawiasie

kwadratowym. W publikacjach wymiaru rozprawy znacznie korzystniejszy jest format wskazujący na pierwszego autora i rok publikacji.

Osobną kwestią budzącą wątpliwości recenzenta jest sprawa tytułu rozprawy. Autorce brakuje konsekwencji w posługiwaniu się tytułowymi pojęciami pracy. Jest wszakże świadoma dyskusji metodologicznej odnośnie zakresu terminów „paralingwistyczny” i „poza-lingwistyczny”, czemu daje wyraz w Rozdziale 2., a układ pracy (por. Rozdział 6) wskazuje na to, że przychyliła się do poglądu, iż terminy te odnoszą się do wyraźnie rozłącznych obszarów. Temu pogładowi przeczy z kolei tytuł rozprawy, w którym nie ma odniesienia do zjawisk paralingwistycznych, które jednak są niewątpliwie głównym obiektem badań Doktorantki.

Z analiza zarówno rozprawy, jak i całości dorobku wskazuje, że recenzent wynosi przeświadczenie, Doktorantka jest rzetelnym badaczem, ma bardzo dobrze opanowany warsztat pracy oraz doskonałą orientację w literaturze. Niewątpliwie pod wieloma względami można uznać dorobek za wyróżniający. Rozległość udokumentowanych wyników świadczy o dużej pracowitości Doktorantki, doskonale rokującej o jej dalszej karierze naukowej.

Konkluzja

Biorąc pod uwagę całość poczynionych obserwacji stwierdzam, bez najmniejszych wątpliwości, iż przedłożona rozprawa spełnia warunki stawiane przez Ustawę o Stopniach i Tytułach Naukowych w stopniu w pełni umożliwiającym przejście do dalszych etapów przewodu doktorskiego. Wniosuję zatem o przyjęcie rozprawy oraz o dopuszczenie Doktorantki do dalszych etapów tegoż przewodu.

Poznań, dnia 31.01.2017


Prof. zw. dr hab. Zygmunt Vetulani