

prof. dr hab. inż. Krzysztof Marasek
Katedra Multimediów
PJATK
ul. Koszykowa 86
02-008 Warszawa

Warszawa, 18.08.2017

Recenzja

rozprawy doktorskiej mgr. Agaty Trawińskiej pt. „Charakterystyka akustyczna realizacji miękkiego odpowiednika bezdźwięcznego okluzywu welarnego w polszczyźnie ogólnej”

Promotor: prof. dr hab. Irena Sawicka

Niniejszą recenzję opracowano na zlecenie Dyrektora Instytutu Sławistyki PAN z dn. 22.06.2017.

1. Charakterystyka tematu, celu i tez badawczych rozprawy

Przedłożona do recenzji praca poświęcona jest opisowi i analizie fonetycznej segmentu /c/ w języku polskim (i innych językach słowiańskich), w szczególności przy wykorzystaniu metod analizy artykulacji i odniesieniu ich do oceny odsłuchowej (audytywnej). Autorka w swoich rozważaniach opiera się przede wszystkim na analizie zgromadzonych nagrań stosunkowo dużej populacji mówców, mniej korzystając z metod obrazowania artykulacji. Podejście takie wyjątkowo dobrze do mnie trafia, uważam bowiem badania korpusowe za o wiele bardziej interesujące i pozwalające na wyciąganie ogólniejszych wniosków.

Na str.63 rozprawy sformułowano główne twierdzenia i hipotezy pracy:

(t1) w odróżnieniu od spalatalizowanych kontoidów, np. [pʲ], dla [c] możliwa jest palatalna artykulacja wszystkich jego faz,

(t2) artykulacja [c] jest znacznie bardziej przednia niż opisują to współczesne gramatyki języka polskiego,

(t3) akustycznie kontoid [c] może być realizacją niejednorodną, tj. charakteryzującą się wyższą częstotliwością maksimum energii sygnału w obrębie afrykacji niż plosji; niejednorodność akustyczna nie musi w każdym przypadku oznaczać niejednorodności artykulacyjnej [c],

(t4) w strukturze zwarto-wybuchowego palatalnego kontoidu [c] zaznacza się komponent szumowy, wskazujący na zaawansowany proces afrykacji [c]

(w1) uwarunkowana geograficzno-dialektalnie wymowa mówcy modelowego mogła dać podstawę do przyjęcia postpalatum jako miejsca artykulacji [c] dla polszczyzny ogólnej,
(w2) postpalatalna charakterystyka [c] może dokumentować wcześniejsze stadia rozwoju kontoidów, powstałych na skutek aktywności tzw. polskiej palatalizacji welarnych, które w ciągu ostatnich dziesięcioleci uległy wyraźnemu uprzednieniu,
(w3) zachodzi związek między słabszą ekspresją palatalizacji welarnych a mediopalatalną charakterystyką /i/, zwłaszcza dla mówców niemających w swoim systemie palatalnego fonemu /c/,
(w4) tzw. jota w grupie *kie* jest percepcyjnym śladem biomechanicznie uwarunkowanego przejścia od bardziej przedniej artykulacji [c] do bardziej tylnej, relatywnie ujmując, artykulacji [ɛ], a jest ona wyraźniej percypowana w wymowie młodego pokolenia z uwagi na nasilającą się w XX w. tendencję do uprzedniania realizacji /c/, zwiększającą jednocześnie dystans między punktami docelowymi artykulacji /c/ i /ɛ/

Podstawowym celem pracy, zgodnie z jej tytułem, jest zatem opis akustyczny realizacji /c/ w języku polskim. Temat ten, choć poruszany w literaturze, na pewno nie został jeszcze w pełni opracowany, a wyniki mogą mieć istotne znaczenie dla fonetyki języka polskiego. Przyjęta w rozprawie metodologia, odwołująca się do artykulacyjnej interpretacji danych akustycznych, także dobrze rokuje – o ile oczywiście wszystkie przeprowadzone pomiary zostaną rzetelnie zinterpretowane.

Zagadnienie badawcze uważam zatem za dobrze postawione, aktualne i ważne naukowo, a przyjętą metodologię prac za prawidłową.

2. Zawartość rozprawy

Rozprawa składa się z 3 głównych części¹ (zagadnienia wstępne, część eksperymentalna, podsumowanie) podzielonych na 9 rozdziałów, spisów ilustracji i tabel, dodatku z ilustracjami dotyczącymi rozkładu artykulacji /c/ i /k/ przez poszczególnych mówców, bibliografii (z opisem źródeł nagrań), oraz nieujętego w spisie treści angielskiego streszczenia. Dokument liczy w sumie 593 strony (!). Stwierdzenie zatem, że recenzowana praca jest wyjątkowo obszerna, jest w tym wypadku eufemizmem. Praca jest monstrualna, o wiele przekraczając zwyczajowe rozmiary rozprawy doktorskiej. Rozumiem, że ambitnym celem Autorki było wnikliwe przedstawienie wszystkich aspektów opracowywanego zagadnienia, ale taka wielkość utrudnia prześledzenie wkładu Autorki w rozwój dziedziny i nieco rozmywa główny temat.

Zamieszczona bibliografia pracy liczy 20 stron, po ok. 16 pozycji na stronę – jest zatem niezwykle obszerna i wskazuje na erudycję Autorki, która w treści pracy, w przypisach, szeroko cytuje i dyskutuje zagadnienia badawcze przedstawione w analizowanych pracach. Pozycje bibliograficzne to szeroki przekrój książek, artykułów z czasopism i konferencji naukowych. W literaturze znalazłem też odwołania do sześciu prac Autorki, które dobrze

¹ Choć numeracja sugeruje podział na rozdziały, ze względu na ich obszerność uznałem, że odpowiadają one częściom pracy, a one z kolei dzielą się na rozdziały

korespondują z tematem recenzowanej rozprawy (jedna z prac jest tematycznie tożsama z rozprawą).

W pracy nie zamieszczono wykazu skrótów i oznaczeń, brakuje mi także indeksu pojęć – a to ze względu na dość rozbudowane wykorzystanie terminologii językoznawczej.

Struktura pracy jest logiczna. Praca rozpoczyna się wstępem (część pierwsza) przedstawiającym analizowany problem badawczy, postawione cele i tezy pracy, przegląd i wnioski wynikające ze studiów literatury przedmiotu oraz analizowany materiał badawczy. Następną część pracy ma charakter analityczno-teoretyczny. Zawarty w niej rozdział drugi stanowi wprowadzenie do metod analizy i reprezentacji mowy, segmentacji /c/ na poszczególne fazy, charakterystyki czasowo-częstotliwościowej tego dźwięku mowy i artykulacyjnej interpretacji pomiarów. Kolejny rozdział opisuje eksperymenty i opis artykulacyjny fonemu /c/ w polszczyźnie ogólnej i jej wariantach oraz w kontekstach /ε/, /s'/, /j/ oraz jego realizację w wybranych językach słowiańskich (słowacki, macedoński, słoweński). Pracę wieńczy rozbudowane podsumowanie, w którym przedstawiono wnioski z prowadzonych analiz oraz odniesiono się do postawionych we wstępie tez (twierdzeń). Twierdzenia te rozszerzono także na dialektalne warianty polszczyzny oraz wybrane języki słowiańskie.

Praca jest bardzo bogato ilustrowana – 235 rycin (plus kilkadziesiąt wykresów w Dodatku), 8 ilustracji i 81 tabel przedstawiają wyniki ogromnej ilości przeprowadzonych eksperymentów i pomiarów. Ogrom pracy wykonanej przez p. mgr A. Trawińską budzi mój szczerzy podziw!

3. Uwagi krytyczne i wskazówki dotyczące rozprawy

Tą część chciałbym rozpocząć od pochwał: przede wszystkim chciałbym podkreślić niezwykle staranność widoczną w przeprowadzonych eksperymentach i opracowanych wynikach – wszystko jest dokładnie opisane, zilustrowane i przeanalizowane. Podkreślić chciałbym też świetne przygotowanie części erudycyjnej, w której Autorka przedstawiła chyba całe światowe piśmiennictwo na ten temat, ale też wykazała głębokie zrozumienie literatury i zagadnień fonetyki akustycznej.

Jako informatyk (choć z dyplomami z fonetyki eksperymentalnej), nie czuję się w pełni kompetentny, aby dyskutować zagadnienia językoznawcze przedstawione w recenzowanej rozprawie. Moje uwagi i wskazówki dotyczyć będą przede wszystkim aspektów pomiarów i analizy sygnałów akustycznych, stanowiących jednak podstawę wniosków Autorki.

Zacznę jednak od uwag ogólnych:

- I. Zgodnie z tytułem praca dotyczy charakterystyki realizacji miękkiego odpowiednika bezdźwięcznego okluzywu welarnego w polszczyźnie ogólnej. W treści pracy, szczególnie w części eksperymentalnej, opisano i zrealizowano wiele dodatkowych prac dotyczących realizacji tego segmentu w wielu uwarunkowanych geograficznie odmianach polszczyzny oraz dla innych języków słowiańskich. Tak jak już wspominałem w p.2 recenzji, uważam tą część działań za zbyt rozbudowaną, nie odnoszącą się bezpośrednio do głównego tematu pracy, szczególnie w odniesieniu do stosunkowo niewielkiej bazy rozważanych języków słowiańskich. Wydaje się, że te rozważania są świetnym materiałem na dodatkową monografię, ale tu niepotrzebnie rozmywają główny wątek pracy.

- II. W całej rozprawie często używane jest określenie „fonem /c/”. Nie wchodząc w tej recenzji w poważne dyskusje na temat definicji fonemu, podejścia strukturalistycznego czy generatywnego, ja nadal uważam, że /c/ w języku polskim nie jest fonemem. W prowadzonych przeze mnie pracach dotyczących automatycznego rozpoznawania mowy istotne jest bowiem jasne rozróżnienie pomiędzy fonemem a alofonem. Ta uwaga ma tylko charakter porządkujący i nie deprecjonuje wniosków Autorki.

Przechodząc do uwag technicznych:

- III. Podstawowa metoda analizy sygnału mowy zastosowana w pracy polega, w uproszczeniu, na segmentacji sygnału i wyznaczeniu częstotliwości maksimumów obwiedni widma dla ploszki i następującego po niej komponentu szumowego (frykcji i ew. aspiracji) mierzonego dla czasu VOT (tj. do początku drgań fałd głosowych) oraz określeniu czasu trwania zwarcia i czasu trwania VOT (str.94). Wyznaczenie maksimumów obwiedni widma dla ploszki (Ryc. 6 i Ryc.7) wymaga zatem wyznaczenia rozkładu widmowego dla stosunkowo krótkiego segmentu sygnału (rzędu 20 ms). W pracy nie podano niestety częstotliwości próbkowania wykorzystanych nagrań, zwykle w badaniach fonetycznych waha się ona pomiędzy 16 a 48 kHz. Zakładając tę wyższą, otrzymujemy rozdzielczość FFT na poziomie 50 Hz². Taka będzie zatem też dokładność oszacowania maksimum obwiedni widma. To przekłada dokładność wyznaczania miejsca zwarcia (str. 84 rozprawy) na nieco poniżej 1 mm; przy próbkowaniu rzędu 16 kHz ta dokładność spadnie do ok. 4 mm. Wynika to z ograniczeń metody FFT; przedłużanie okna obserwacji sygnału nic nie pomoże – FFT uśrednia widmo sygnału, a przedłużenie okresu analizy poprzez dodanie zer powoduje tylko interpolację sygnału. Zapewne te ograniczenie jest szczególnie wyraźne w przypadku analizowanych nagrań historycznych, bądź ze względu na niższą jakość materiału nagraniowego, bądź ze względu na często stosowaną kompresję w przypadku internetowych źródeł nagrań. Warto zdać sobie sprawę z tych ograniczeń i pamiętać o nich przy porównaniach z palatogramem czy na wykresach trajektorii realizacji /c/.
- IV. Na str. 85 Autorka pisze „Maksima energii sygnału ploszki³ i komponentu szumowego wyznaczamy przy użyciu algorytmów szybkiej transformaty Fouriera i kodowania predykcją liniową (LPC) ...” co zilustrowano na rycinach. Kodowanie LPC niezbyt nadaje się do analizy rozważanych tu sygnałów impulsowych i szumowych, zakłada bowiem liniowy związek pomiędzy obserwowanymi i następującymi próbkami sygnału. Korzystanie z LPC, dobór ilości współczynników LPC wymaga dużego doświadczenia i może spowodować błędy. Jak widać jednak na Ryc. 7 Autorka nie zawsze wyznaczała maksima na podstawie obwiedni LPC – i dobrze.
- V. Mam wątpliwości co do dokładności i poprawności estymacji miejsca zwarcia i formowania się szczeliny tylko na podstawie hipotetycznego rezonansu w rurze jednostronnie zamkniętej. Proszę bowiem zwrócić uwagę, że dodatkowa przeszkodę stanowią tu zęby, także wargi mogą odgrywać jakąś rolę w przepływie powietrza

² (20 ms to 960 próbek dla próbkowania 48000 Hz, dzieląc częst. próbk. przez ilość próbek (długość okna) otrzymujemy rozdzielczość widma 50 Hz)

³ Oczywiście chodzi tu o maksima energii widma sygnału

(przekrój rury nie jest stały na całej długości). A rura ta jest krótka, przepływ powietrza szybki i taka interpretacja zjawisk akustycznych wydaje mi się dość powierzchowna, zresztą pewne wątpliwości można też zauważyć w cytowanej pracy Stevens, Keyser, gdzie przesunięcie zwarcia powoduje „przeskok” w widmie i kategorii artykułowanego segmentu. Także w momencie niepełnego zwarcia kanał głosowy położony w głębi (poniżej przewężenia) jest w pewien sposób pobudzany i może wpływać na charakterystykę akustyczną przedniego segmentu kanału głosowego - tylna część toru głosowego wówczas może wpływać na rozkład rezonansów w generowanym szumie trącym. Jak zresztą zauważa sama Autorka, także różnice w wielkości układu mownego i kształtu podniebienia mogą wpływać na rezonanse. Wydaje się, że idealnym uzupełnieniem pracy p. Trawińskiej byłoby przeprowadzenie serii eksperymentów z wykorzystaniem artykulacyjnego syntetyzatora mowy. Taki syntetyzator jest dostępny (<http://www.vocaltractlab.de>, autorstwa P. Burkharda) i można w nim precyzyjnie ustalić wszystkie wymagane parametry, a także prześledzić zmiany charakterystyki widmowej generowanego dźwięku (jak to zrobiła w rozprawie doktorskiej pani Lasarczyk⁴). To pozwoliłoby także na analizę błędów oszacowania miejsc zwarcia języka z podniebieniem i tworzenia się szczeliny oraz oszacować wpływ VOT.

- VI. W przypisie na str. 85 Autorka opisuje metodę wyznaczania równania locusa, co z kolei pozwala oszacować miejsca artykulacji segmentów zwartych. Z treści pracy nie wynika dlaczego nie zdecydowała się ona na wykorzystanie tej metody, a na pewno porównanie stosowanego w pracy oszacowania z tym wynikającym z równania locusa, byłoby interesującym uzupełnieniem i skutkowało by zwiększeniem wiarygodności otrzymanych wyników⁵. Wydaje się też, że w przypadku analizy nagrań historycznych, których jakość może być niska, metoda analizy locusa może być mniej podatna na zakłócenia (szum nagrań czy obcięcie pasma może bowiem przysłonić wysokoczęstotliwościowe komponenty plozji).
- VII. Jednym z zagadnień rozważanych w pracy jest tzw. jota. Także w tym wypadku odpowiednia synteza (lub resynteza) wraz z badaniem odsłuchowym mogłyby wspomóc siłę argumentacji. Badania percepcyjne mogłyby też dać odpowiedź na dość zasadnicze w kontekście tej rozprawy pytanie: czy na podstawie komponentu szumowego można percepcyjnie rozróżnić segment spalatalizowany od niepalatalnego? Liczne eksperymenty (np. resynteza) dowodzą, że percepcyjnie najistotniejsze jest ugięcie formantów (F2) samogłoski, a to ze względu choćby na głośność segmentu samogłoskowego często „maskującego” własności szumu.
- VIII. Coraz powszechniej przyjmuje się, że podstawowym segmentem przy planowaniu ruchów artykulacyjnych jest sylaba. Przy takim podejściu ruch artykulacyjny nie składa się z sekwencji pojedynczych konfiguracji toru głosowego, lecz stanowi pewną całość w planowaniu artykulacji sylaby (niektórzy uważają, że również w percepcji mowy). Zgodnie z taką teorią głoski nie mają swoich niezmiennych konfiguracji i ruchy artykulacyjne zależą od typu wymawianej sylaby. Warto byłoby także

⁴ http://www.coli.uni-saarland.de/groups/WB/Phonetics/contents/phonus-pdf/phonus17/phonus_17_lasarczyk.pdf

⁵ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2981116/>

rozważyć, czy położenie segmentu [c] w pobliżu samogłoski akcentowanej ma wpływ na jego artykulację (wydaje się, że może mieć).

- IX. Chciałbym podkreślić, że niezwykle spodobał mi się pomysł wykorzystania w przeprowadzonych analizach nagrań archiwalnych i dostępnych w sieci nagrań znanych i charakterystycznych mówców.

Dwie uwagi związane ze statystyką:

- X. W licznych tabelach Autorka podaje wartość średnią i odchylenie standardowe mierzonej wielkości (np. parametry akustyczne realizacji /c/). Wielkości te charakteryzują rozkład wartości, ale tylko przy założeniu normalności ich rozkładu. Mam wątpliwości, czy te rozkłady są normalne: dobrze byłoby przeprowadzić albo test normalności, albo skorzystać z innych charakterystyk (np. mediana, odstęp między kwantylami).
- XI. W pracy generalnie brakuje analiz istotności. Statystyka daje nam narzędzia, dzięki którym możemy stosunkowo łatwo zweryfikować hipotezy o statystycznej istotności różnic pomiędzy populacjami (poczynając od testów *t* Studenta, a kończąc na złożonej analizie wariancji).

4. Ocena strony redakcyjnej

Od strony edytorskiej praca jest bardzo dobra. Właściwie nie znalazłem w niej literówek, błędnych oznaczeń czy odnośników. Także stylistycznie praca jest na dobrym poziomie. Sam układ treści także nie budzi zastrzeżeń.

Drobne zastrzeżenia sprawdzają się do:

- momentami niejasnego dla mnie użycia nawiasów // i []. Przyjęło się, że fonemy prezentowane są w nawiasach //, a alofony w nawiasach []; nawet w tezach pracy stosowane są nawiasy [], co przy przyjęciu przez Autorkę /c/ jako fonemu, budzi we mnie pewne wątpliwości;
- w Tabelach 1 do 7 pojawiają się określenia V1, V2 bez podania znaczenia (oczywiście łatwo się domyśleć o co chodzi).

5. Wnioski końcowe recenzji

Moja ocena przedstawionej od recenzji rozprawy mgr Agaty Trawińskiej jest jednoznacznie pozytywna.

Praca jest dobra warsztatowo, doktorantka wykazała się w niej wiedzą z zakresu fonetyki akustycznej i językoznawstwa, a wskazane w recenzji pewne słabości w żadnej mierze nie deprecjonują osiągnięć badawczych Autorki.

Uważam, że rozprawa wykazuje zdolność doktorantki do samodzielnego rozwiązania złożonego problemu badawczego przy użyciu nowatorskich metod i własnych osiągnięć (korpusowa analiza segmentu /c/ dla różnych kontekstów wypowiedzi, wykazanie miejsca jej artykulacji, badania na różnych populacjach mówców).

W konkluzji stwierdzam, że opiniowana rozprawa spełnia wymogi stawiane przez ustawę o stopniach i tytułach naukowych dla prac doktorskich, a jej Autorka zasługuje na przyznanie stopnia naukowego doktora nauk humanistycznych. Stawiam więc wniosek o dopuszczenie tej rozprawy do publicznej obrony.

Stawiam też formalny wniosek o wyróżnienie przedłożonej rozprawy.

A handwritten signature in blue ink, consisting of several fluid, connected strokes. The signature is positioned in the upper right quadrant of the page.