

# A R T Y K U Ł Y I R O Z P R A W Y

Piotra Łobacz

(Uniwersytet im. Adama Mickiewicza,  
e-mail: plobacz@amu.edu.pl),

Anita Lorenc

(Uniwersytet Warszawski,  
e-mail: anita.lorenc@uw.edu.pl)

## ROZWAŻANIA NAD PODSTAWAMI NAUCZANIA NOWOCZESNEJ FONETYKI JĘZYKA POLSKIEGO<sup>1</sup>

### 1. MIĘDZY JĘZYKOZNAWSTWEM A TECHNOLOGIĄ MOWY

W obecnych czasach bardzo trudno jest w sposób jasny i zarazem prosty sformułować przedmiot fonetyki jako dziedziny względnie jednorodnej. Natura mowy jako obiektu badań wymusza coraz większą wielodyscyplinarność. Wydane na świecie w ostatnich latach prace z zakresu fonetyki ogólnej są obszernymi (500–1000-stronicowymi) opracowaniami o wieloaspektowej tematyce. Przykładem takiego podejścia jest *The Handbook of Phonetic Sciences*, wyd. II [Hardcastle, Laver, Gibbon (red.) 2010], czy starsza o kilkanaście lat monografia R.D. Kenta [1997]. Z jednej strony dotyczą one treści podstawowych, jak pojęcia z fizyki i matematyki, czy anatomii (system mowy i dźwięki mowy, teorie percepcyjne), a z drugiej aspektu stosowanego: komunikacji pomiędzy człowiekiem a komputerem, fonoskopii (tzw. fonetyki sądowniczej), wybranych zaburzeń procesu artykulacji, związków z innymi dziedzinami nauki, takimi jak psychiatria, sztuka i kultura oraz ewolucja mowy.

Pod koniec lat 90. ubiegłego wieku, w ramach europejskiego grantu badawczego *Erasmus* oraz programu edukacyjnego *Socrates*, dokonano m.in. szczegółowego przeglądu nauczania fonetyki w wielu ważnych ośrodkach Europy i wypracowano założenia wzorcowe do nauczania wiedzy z zakresu komunikacji za pomocą mowy (cykl publikacji z lat 1997–1999 nosi wspólny tytuł *The Landscape of Future Education*).<sup>2</sup> Nauki o łączności za pomocą mowy stanowią szereg dyscyplin, w których cen-

---

<sup>1</sup> Artykuł powstał w ramach projektu naukowo-badawczego nr 2012/05/E/HS2/03770 pt. „Współczesna wymowa polska. Badanie z wykorzystaniem trójwymiarowej artykulografii elektromagnetycznej”, realizowanego pod kierownictwem A. Lorenc. Projekt został sfinansowany ze środków Narodowego Centrum Nauki na podstawie decyzji nr DEC-2012/05/E/HS2/03770.

<sup>2</sup> To są trzy nieobszerne tomy. W pierwszym dokonano analizy europejskiej edukacji fonetyki, technologii mowy oraz terapii mowy i języka [Bloothoof i in. 1997], w drugim przedstawiono dla wszystkich trzech dziedzin propozycje nauczania [Bloothoof i in. 1998], a w trzecim zarekomendowano określonych autorów, przykładowe konspekty, odniesienia do stron internetowych wybranych

tralną pozycję zajmują fonetyka i technologia mowy. Obie dziedziny mają bezpośrednie powiązania z zaburzeniami mowy i języka, językoznawstwem komputerowym i przetwarzaniem języka naturalnego. Szczegółowa analiza sugestii programowych zawartych w tych tomach prowadzi do wniosku, że nie jest możliwe (lub nawet pożądane) ujęcie całej problematyki w ramach jednego przedmiotu nauczania. Nie to jest najistotniejsze, żeby propozycje zaproponowane pod koniec XX wieku wdrożyć i mniej lub więcej dokładnie zrealizować. Znacznie ważniejsza jest konstatacja, że współczesne narzędzia dydaktyczne w obrębie nauk fonetycznych wymagają reformy. Reformy takiej nie uda się jednak dobrze przeprowadzić bez kolejnej próby wyjaśnienia kilku podstawowych kwestii. Niniejszy tekst koncentruje się na dwóch takich prymarnych aspektach: wiedzy językowej i praktycznym wykorzystaniu technologii mowy.

Ponieważ w nowszych badaniach szczegółowych, w tym w opracowaniach wykonanych przez zawodowych logopedów, dostrzega się wyraźnie zastosowanie zintegrowanej wiedzy z zakresu wytwarzania mowy, natury fali akustycznej i jej odbioru słuchowego, to wpisując się w takie potrzeby, propozycja nowej koncepcji nauczania fonetyki powinna opierać się na wykorzystaniu wszystkich naukowych narzędzi (teorii i badań) dotyczących najściślejszych współzależności między tymi wszystkimi trzema aspektami; od licznych teorii kompensacyjnych w procesie artykulacji, po estymację i ekstrapolację procesów artykulacyjnych na falę głosową i odwrotnie: odtworzenie cech artykulacyjnych na podstawie oceny charakterystyki widma akustycznego, zawsze w odniesieniu do wrażeń słuchowych. [por. np. Perrier, Fuchs 2015]. Wymaga to jednak przypomnienia, przynajmniej kilku, nierozstrzygniętych kontrowersji koncepcyjnych.

Technologia mowy, w obecnych rozważaniach, niczego jej nie ujmując w zakresie budowania teorii, wykorzystywana jest głównie jako dziedzina stwarzająca określone narzędzia badawcze i dydaktyczne. Tak jak w zakresie fonetyki akustycznej, analiza zmiennego w czasie sygnału mowy jest standardowo stosowana od około 70 lat i jest całkowicie nieinwazyjną metodą, tak w zakresie szczegółowego badania przebiegu artykulacji większość wykorzystywanego instrumentarium była albo inwazyjna, albo wymagała bardzo starannej i czasochłonnej późniejszej obróbki uzyskanych wyników. Nowoczesna aparatura pozwala w znacznym zakresie ową niepożądaną inwazyjność ograniczyć, ponadto jest na tyle zintegrowana i wielomodalna, że bezpośrednio w czasie nagrań pozwala na uzyskanie zewnętrznego obrazu (wideo) szczegółowego przebiegu artykulacji, praktycznie z dowolnie wybraną dokładnością. Daje możliwość wiarygodnej ekstrapolacji funkcjonowania narządów nieobjętych bezpośrednim badaniem, jak również uzyskania ze zmagazynowanych danych dowolnego „obrazu” akustycznego [por. Mik i in. 2018]. Atrakcją nowoczesnych

---

europejskich ośrodków, istniejące wówczas programy multimedialne itp. [Blotthoof i in. 1999].

osiągnąć technologii mowy są interaktywne programy multimedialne, które pozwalają samodzielnie w czasie rzeczywistym analizować sygnał mowy, dokonywać jego automatycznej segmentacji itd. Wszystkie te narzędzia powinny być w codziennej dydaktyce fonetyki wykorzystywane, tym bardziej że nie wymagają one jakiegoś specjalnego czy żmudnego przygotowywania się do ich używania w codziennej praktyce zawodowej i dydaktycznej.

## 2. ROLA WYBRANYCH ASPEKTÓW WIEDZY JĘZYKOWEJ

Jak wspomniano powyżej, sygnalizując złożoną interdyscyplinarność nauk fonetycznych oraz rozwój jej poszczególnych dziedzin, istotne są ponowne reinterpretacje niektórych aspektów, z dzisiejszego punktu widzenia najważniejszych. Zacząć należy od kwestii podstawowych. Warto pamiętać, że podstawową motywacją wszystkich nauk fonetycznych jest zawsze baza lingwistyczna. Wiedza językowa jest pojęciem bardzo szerokim, wobec czego na potrzeby dydaktyki współczesnej fonetyki należy ją na początek ograniczyć do podstaw fonetyki ogólnej (teorii i znajomości wybranych systemów dźwiękowych języków świata), głównych teorii fonologicznych oraz podstawowych procesów intelektualnych w zakresie wytwarzania i odbioru mowy. Nie chodzi w tym wypadku o epatowanie odbiorców jakimiś wysublimowanymi analizami porównawczymi systemów fonologicznych wybranych języków, chociaż może warto zdawać sobie sprawę, że wiedza taka przydaje się między innymi w opiece logopedycznej, ponieważ rozszerza znacznie wiedzę o możliwościach ludzkiego aparatu artykulacyjnego i ugruntowuje przekonanie, że to, co w języku *x* jest normą, w języku *y* może stanowić już realizację wadliwą, zarówno w warstwie artykulacyjnej, jak i odbiorczej. Można zatem stwierdzić, że im lepsza znajomość systemów fonetycznych języków świata, tym lepsze możliwości badania i analizy różnorodnych zjawisk w zakresie dźwięków mowy rodzimego języka.

### 2.1. ZRÓŻNICOWANIE DŹWIĘKÓW JĘZYKÓW ŚWIATA

Przedstawione zostaną w tym miejscu tylko 3 przykłady dla ilustracji samej idei.

- (a) Świadomość, że z typologicznego punktu widzenia system fonologiczny języka polskiego jest zbliżony w warstwie segmentalnej do języka japońskiego,<sup>3</sup> interesujący jest głównie dla studentów japończyki. Może jednak stanowić także istotną refleksję, że w zakresie podobieństwa repertuaru dźwięków pokrewieństwo genetyczne języków nie zawsze jest istotne, podobnie jak istotna jest wiedza o braku

---

<sup>3</sup> Por. np. [Majewicz 1986].

podobieństwa obu wymienionych języków w zakresie cech suprasegmentalnych (szczególnie takich jak struktura sylaby, bardzo rygorystyczne poczucie rytmu oraz intonacja).

- (b) Zgola odmienny problem istnieje w zakresie wiadomości na temat północnobuszmeńskiego języka *!xū* (*kung*), uchodzącego za najstarszy żywy relikw językowy. Jego opisany system fonologiczny zawiera: 24 samogłoski, 22 dyftongi, 46 spółgłosek o płucnym bądź krtaniowym mechanizmie wytwarzania prądu powietrza oraz 48 mlasków.<sup>4</sup> Stanowi to łącznie 140 elementów. Lista dystynktywnych cech artykulacyjnych jest bardzo długa. Samo uzmysłowienie sobie, na czym polega artykulacja dźwięcznej sybilantnej (silnie szumowej), eiektywnej (krtaniowej) afrykaty jest trudne, lecz o wiele trudniejsze jest ewentualne wytworzenie takiej głoski w europejskim kręgu kulturowym, wyłącznie na podstawie opisu artykulacyjnego. Wiedza ta jest bardzo pouczająca, wskazuje bowiem na dużą elastyczność toru głosowego i uświadamia bardzo wysoki stopień zróżnicowania barw dźwięków mowy, które w naszym obszarze językowym i kulturowym są znacznie mniej zróżnicowane. Dowodem tego jest słaba rozróżnialność przez polskich studentów barwy mlasków, ograniczająca się wyłącznie do kryterium: mocniejszy sygnał (mlaski welarne) i odpowiednio słabszy (wszystkie pozostałe). Najbardziej klarownym podziałem samogłosek w *!xū* jest kryterium iloczynowe (opozycja fonologiczna: samogłoski krótkie i długie, podobnie jak w wielu językach europejskich). Dla obu typów, choć nie zawsze symetrycznie, występują np. samogłoski faryngalizowane obok samogłosek nazalizowanych, są systemowe samogłoski jednocześnie faryngalizowane i nazalizowane (o miejscu artykulacji środkowym i tylnym).<sup>5</sup>
- (c) Nieco ponad 100 lat trwają prace nad zestawieniem spójnego, jednolitego sposobu notacji bardzo bogatego zbioru dźwięków występującego w językach świata, zarówno w odniesieniu do spółgłosek, jak i samogłosek. W wypadku wokalizmu opracowany przez IPA<sup>6</sup> artykulacyjno-akustyczny czworobok samogłoskowy stanowi istotne kryterium klasyfikacyjne, dzięki któremu można scharakteryzować wszystkie dźwięki samogłoskowe dowolnego języka w zakresie dwóch, a częściowo nawet trzech podstawowych cech artykulacyjnych. Jednakże, jako twór teoretyczny, czworobok nie stanowi ścisłego odzwierciedlenia w żadnym konkretnym, dotychczas zbadanym, języku. W starszych opracowaniach napotkać można było przykład języka *ngwe* (*ngwag*) z grupy Benua (obszar Kongo w Kamerunie), najlepiej spełniający kryterium skrajnych – przednich i tylnych – elementów czworoboku, lecz nowsze prace [np. Ladefoged, Maddieson 1996]

<sup>4</sup> Por. [Maddieson 1984].

<sup>5</sup> Ibidem.

<sup>6</sup> International Phonetic Association.

wskazują na zjawisko centralizacji w tym języku. Wprowadzenie trójstopniowej opozycji fonologicznej (przednia / środkowa / tylna) nie jest oczywiście sprzeczne z ideą czworoboku, ale odbiega nieco od klasycznej interpretacji obecności w tym języku 4 samogłosek przednich niezaokrąglonych i 4 tylnych zaokrąglonych, uważanych niegdyś za najbardziej klasyczną konfigurację tych dźwięków w interpretacji IPA.

## 2.2. NIEJEDNOZNACZNE INTERPRETACJE FAKTÓW FONETYCZNYCH

Do dziś istnieje spory zbiór zjawisk, powodujący występowanie określonych kontrowersji. Oto kilka przykładów:

- (a) Współczesne instrumentalne badania fonetyczne bardzo często wykorzystują, czasami dość skomplikowane, czasami prostsze zasady odpowiedniości pomiędzy najważniejszymi mechanizmami fizjologicznymi wytwarzania dźwięków mowy a najbardziej istotnymi cechami widmowymi oraz wrażeniem słuchowym, jak np. usytuowanie języka w torze głosowym a wysokość częstotliwości (mierzona np. w hercach) maksimów widma w odniesieniu do barwy danej samogłoski. Wobec tego jednym z ważniejszych problemów fonetycznych jest znajdowanie czasowych segmentów o dużej niezmienności parametrów w dynamicznych procesach wytwarzania, przenoszenia i odbioru mowy. Poszukuje się w określonym przedziale czasu najściślejszych relacji pomiędzy najbardziej ustabilizowanym artykulacyjnym gestem, sygnałem akustycznym, charakteryzującym się najmniejszą prędkością zmian, a funkcjonalnie identycznym obiektem percepcji, np. prototypem kategorii (choćby w teorii magnesu artykulacyjnego dla dziecięcych samogłosek [Kuhl, np. 2007]<sup>7</sup> czy wyników testów percepcji kategoryjalnej spółgłosek [por. np. Mattingly 1985; Liberman 1996]). Takie poszukiwania są nie tylko zasadne, odnoszą się przecież do codziennej komunikacji, dla której są niezbędne, są także potrzebne. Pozwalają na podstawie tylko jednego aspektu analizy ekstrapolować kształt badanych obiektów z pozostałych dwóch punktów widzenia. A jednak w niektórych sytuacjach brak jest takich bezpośrednich relacji pomiędzy artykulacją, falą głosową a jej brzmieniem w interpretacji odbiorcy. Dowodzi tego rewizja tez motorycznej teorii percepcji mowy dokonana przez B. Galantucciego, C.A. Fowler i M.T. Turvey'a [2006], stwierdzająca, że z trzech głównych filarów tej teorii tylko dwa są zasadne (z pewnymi modyfikacjami).<sup>8</sup> Potocznym, charakterystycznym przykładem jest tzw. zjawisko kompensacji artykulacyjnej, często występujące jako doraźna pomoc w niedostatkach wymowy,

<sup>7</sup> Oryginalny tekst ukazał się w 2000 r. w: W.M.S. Gazzaniga (red.), *The new cognitive neurosciences*, New York.

<sup>8</sup> Odrzucono postulat pierwszy o specyficznym, wyłącznie dla mowy (tzw. *speech mode*), przetwarzaniu sygnałów w ośrodkowym układzie nerwowym.



nie wspominając już o funkcjonującej od ponad 60. lat konstatacji, potwierdzonej wieloma eksperymentami, że względnie mała zmiana w artykulacji może spowodować istotnie znaczącą różnicę przebiegu akustycznego i odwrotnie – mała zmiana w sygnale akustycznym może spowodować słuchowe wrażenie zupełnie innego funkcjonalnie dźwięku (por. klasyczne kontrowersje Lane'a [np. Lane 1970]).

- (b) Próby pogodzenia powyższych postaw wymagają zdania sobie sprawy, że w procesie komunikacji występują oba zjawiska, co musi prowadzić do formułowania przez badaczy bardzo wyważonych konkluzji. Przez kolejne lata przeprowadzono wiele badań, lecz szczególnie interesujące dla dalszych rozważań (por. p. 2.3.) są związki pomiędzy percepcją a produkcją [por. np. Galantucci, Fowler, Goldstein 2009]. Zasady odpowiedniości pomiędzy procesem wytwarzania mowy, jej kształtem akustycznym i przekazem słuchowym są zatem wyraźne, ale np. nie zawsze możliwe do opisanego za pomocą np. wspólnego czasowego segmentu fonetycznego, por. dalej podejście fonologii artykulacyjnej [np. Fowler, Galantucci 2005]. Podkreśla się jednak w założeniach teoretycznych tej koncepcji, że język nie spełniłby swojej podstawowej roli, gdyby nie założyć zgodności pomiędzy procesami produkcji i percepcji, co w znacznym stopniu wynika także z funkcjonowania ludzkiego umysłu.
- (c) Znacznym kłopotem w porównywaniu płaszczyzny artykulacyjnej i percepcyjnej jest zjawisko koartykulacji. J. Ohala [1981] stwierdził, że wzajemne zależności pomiędzy sąsiednimi dźwiękami funkcjonują jak zakłócenia i w sposób istotny modyfikują niezmiennosc gestów artykulacyjnych oraz zmniejszają stabilność fali głosowej. Ze słuchowego punktu widzenia mogą istotnie modyfikować barwę dźwięków wyciętych z wypowiedzi ciągłych. Dotyczy to szczególnie samogłosek. Test identyfikacji wyciętych sześciu samogłosek ustnych z krótkich polskich zdań, odsłuchanych przez 90-osobową grupę studentów, wykazał, że najwięcej poprawnych identyfikacji uzyskano dla samogłoski [i] 78%, dla [i] 60%, dla [u] 53%, ale dla [ɔ] 44%, dla [ɛ] 36%, a dla [a] już tylko 32%. 48% słuchaczy wskazało zamiast [ɛ] samogłoskę [i], 49% słuchaczy natomiast zidentyfikowało [a] jako [ɛ]. Rozbieżności te wynikają głównie z bardzo silnie działającego w języku polskim kontekstu palatalizacyjnego [por. Łobacz 1977].

### 2.3. KONTROWERSJA FONETYKA A FONOLOGIA

W pracach badawczych z zakresu polskiej logopedii często ich autorzy odwołują się do poziomu analizy *stricte* językoznawczej i deklarują, że ich badania eksperymentalne odnoszą się do poziomu fonetycznego, rozumianego przez nich jako badania realizacji konkretnych głosek w wypadku różnego typu zaburzeń wymowy, a kwestie fonologiczne zostawiają do późniejszego całościowego opracowania (metoda *a posteriori*). Równno-

częściej, by wykonać określone prace pomiarowe, diagnostykę, ocenę terapii itp., poszczególni badacze odwołują się do już istniejących opisów fonologicznych współczesnej polszczyzny (metoda *a priori*) i przechodzą do szczegółowych interpretacji poziomu niższego niż fonologiczny. Fraza „fonem X jest realizowany jako zbiór określonych głosek...” bywa w tego typu podejściu praktycznie jedynym odwołaniem się do fonologii. Świadomość, że jest to podejście niewystarczające znacząco rośnie, mimo że w dalszym ciągu wykorzystuje się najczęściej intuicyjne podejście do fonologii, a nie jakąkolwiek konkretną, zwłaszcza bardziej nowoczesną teorię postgeneratywną. Problem dotyczy także wielu opracowań dialektologicznych. Należy się poważnie zastanowić, czy takie podejście jest w pracy badawczej i terapeutycznej poprawne, wystarczające, a nawet zasadne.<sup>9</sup> Prawie dla każdego żywego języka naturalnego można bowiem, na podstawie wybranych teorii, przedstawić kilka nieidentycznych, ale wewnętrznie systemowo spójnych rozwiązań o różnej liczbie jednostek fonologicznych, np. w zakresie kontrowersji dotyczących samodzielności fonologicznej [j] oraz [w].<sup>10</sup>

Istnieją także prace, w których podjęto próby analizy fonologicznej i interpretacji systemu fonemów języka polskiego z punktu widzenia uzyskanych rezultatów. Czy w wypadku badań fonetycznych nad normą i patologią mowy podejście *a posteriori* jest w ogóle możliwe? Jeśli uwzględnia się fakty percepcyjne, nieograniczające się do dekodowania pojedynczych izolowanych dźwięków mowy, to na ogół wykorzystuje się nabytą wiedzę językową i słuchacz odwołuje się do jednostek abstrakcyjnych na poziomie rozpoznawania w ośrodkowym układzie nerwowym.

Od początku ukonstytuowania się fonologii jako odrębnej dziedziny wiedzy, ze zrozumiałych względów, powstawały liczne prace dotyczące związków pomiędzy fonetyką a fonologią. Próby teoretycznych rozważań nad związkami pomiędzy formami fonologicznymi jako zbiorem statycznych atrybutów: bez skali czasu, dyskretnych i nieuwarunkowanych kontekstowo z przebiegiem artykulacji jako procesem dynamicznym, częściowo nakrywających się gestów, czyli mocno zależnych od kontekstu, prowadzących do wytworzenia sygnału akustycznego o silnej zmienno-

<sup>9</sup> Często badacze stają przed dylematem, że bez fonologii nie da się dzisiaj wyjaśnić wszystkich problemów związanych z różnego typu deficytami wymowy, ale jednocześnie sami dostrzegają, że dotychczasowe rozstrzygnięcia fonologiczne nie zawsze są satysfakcjonujące [por. np. Ostapiuk 2013a; 2013b].

<sup>10</sup> W późnych latach 60. XX w. rozpatrywano status fonologiczny tzw. polskich samogłosek niesylabicznych, ostatnio, zgodnie z klasyfikacją IPA, nazywanych aproksymantami, odwołując się do prac z zakresu szeroko pojmowanej fonologicznej teorii strukturalnej [Łobacz 1970]. W 2016 roku ukazał się artykuł, dotyczący tych samych głosek, wykorzystujący teorię fonologiczną funkcjonalizmu aksjomatycznego, który zdecydowanie przesądza o przynależności analizowanych głosek do odpowiednich fonemów samogłoskowych /i/ oraz /u/ i wprowadza dla nich inne nazewnictwo [Wagiel 2016].

ści bez widocznych segmentów o rozciągłości głoski (dokładność segmentacji wynika z zastosowanej „ziarnistości” skali czasu), były jednym z podstawowych problemów naukowych.<sup>11</sup> J. Ohala [1997] stwierdził, że interpretacja relacji pomiędzy fonetyką a fonologią wykorzystuje osobiste doświadczenia, przekonania, a nawet uprzedzenia poszczególnych interpretatorów. Stąd pewnie wynika taka mnogość teorii fonologicznych, szczególnie w okresie postgeneratywnym. Nie ulega kwestii, że fonologia wykorzystuje wiedzę fonetyczną, ale także, co podkreśla J. Ohala [1997], jest wspierana przez psychologię, etnologię, socjologię itd. Należałoby jeszcze dodać, że często także przez logikę, bo wiele teorii fonologicznych w znacznym stopniu jest wynikiem procedur formalizacyjnych.

Od połowy lat 80. XX w. ukonstytuowała się kolejna inicjatywa na gruncie fonologii, którą nazwano fonologią artykulacyjną. Jej główne tezy przedstawiono w wielu pracach, z których przytacza się w tym miejscu jedynie wybrane.<sup>12</sup> W 1992 roku ukazał się zeszyt „Phonetica”, o podwójnej numeracji, poświęcony wyłącznie temu nurtowi fonologii. Twórcy tej teorii (C.P. Browman i L. Goldstein) przedstawili najpierw główne jej założenia,<sup>13</sup> po czym inni autorzy zaprezentowali teksty zarówno krytyczne, jak i wspierające ową koncepcję. Do dzisiaj znajduje ona swoich zwolenników, choć oczywiście jest stale modyfikowana [por. np. cytowaną już pracę Galantucci i in. 2009].

Z punktu widzenia najnowszych narzędzi badania artykulacji jako procesu, a nie wyekstrahowanych stanów, należy powrócić do koncepcji, że system głoskowy jest także abstrakcyjny. Nowoczesna fonetyka buduje własne teorie i wprowadza prawa uogólniające. Propozycja fonologii artykulacyjnej dotycząca abstrakcyjnego, artykulacyjnego gestu jest szczególnie atrakcyjna, zwłaszcza że również umożliwia holistyczne podejście do systemu głoskowego ze względu na ścisłe związki między artykulacją, obrazem fali głosowej a szczególnie procesami umysłowymi jej odbioru przez słuchacza. Jest to istotne dla fonetyków nauczających wymowy poszczególnych języków w obrębie szeroko rozumianej glotto-dydaktyki, jak również dla logopedów, kształconych przeciwieństwo umiejętności słuchania na poziomie fonetycznym (głoskowym) w trakcie zajęć z tzw. treningu słuchowego.

---

<sup>11</sup> Z licznych prac przeglądowych dotyczących związków fonetyki i fonologii przytacza się tu przykładowo tylko dwie publikacje: [Diehl (red.) 1991] oraz esej [Durand, Laks 2002].

<sup>12</sup> Por. np. [Diehl (red.) 1991]; [Fowler, Galantucci 2005].

<sup>13</sup> Obecny tekst nie rozpatruje szczegółowo wszystkich wątpliwości teoretycznych. Zwraca raczej uwagę na fakt, dla wielu zastosowań fonetyki istotny, że główne pomysły autorów mogą stanowić przydatną podbudowę uzyskiwanych wyników na podstawie dynamicznych analiz artykulacji.



### 3. ANALIZA DYNAMIKI GESTÓW ARTYKULACYJNYCH

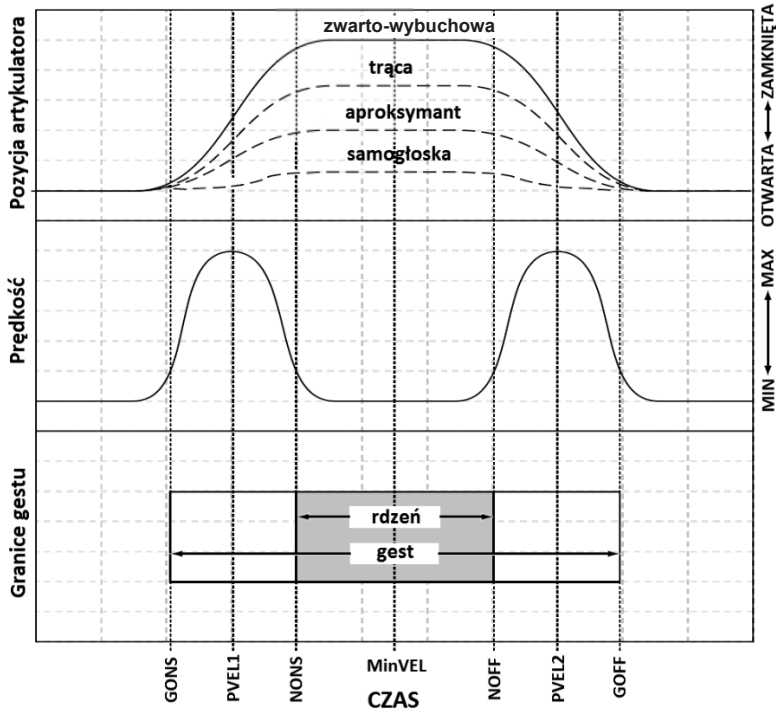
W świetle fonologii artykulacyjnej [por. Browman, Goldstein 1992] jednostki fonologiczne są modelowane na podstawie dynamicznych działań artykulacyjnych zwanych gestami (ang. *gestures*). Analiza dynamiki gestów artykulacyjnych wymaga przede wszystkim ustalenia ich składowych, relewantnych elementów oraz zdefiniowania sposobu wyznaczania granic pomiędzy nimi. Konieczne jest wskazanie artykulatora, którego ruch odgrywa główną rolę w tworzeniu docelowej artykulacji. Jest to tzw. artykulator krytyczny (ang. *critical articulator*) lub artykulator kluczowy (ang. *crucial articulator*) [Kim i in. 2014]. Wyznacza się go na podstawie danych dotyczących metrycznej zmienności gestów [Jackson, Singampalli 2008] lub ich metrycznego stopnia zbliżenia [Recasens i in. 1997], a także na podstawie teoretycznych modeli produkcji mowy z zakresu fonologii artykulacyjnej [Browman, Goldstein 1992] lub modeli statystycznych [por. Saltzman, Kelso 1987; Saltzman, Munhall 1989]. Mimo że liczba artykulatorów kluczowych dla każdego fonemu może być większa niż jeden, zazwyczaj wybiera się spośród nich taki, który pełni funkcję prymarną. Istotne punkty czasowe pozostałych, sekundarnych artykulatorów kluczowych mogą być ustalone na podstawie czasowej pozycji artykulatora prymarnego. Artykulatory krytyczne przyjmują pozycje o różnym stopniu otwarcia, począwszy od najbardziej otwartych (typowych dla samogłosek), aż po przewężenia typowe dla spółgłosek, najmniejsze dla zwartych. W obrębie dynamicznego gestu artykulacyjnego, tworzono go za pomocą ruchu artykulatora krytycznego, można wyznaczyć składowe granice zależne od jego prędkości. Omawiane zależności zaprezentowano na ryc. 1.

Profile prędkości w analizie dynamiki gestów artykulacyjnych wyznaczone są z wykorzystaniem specjalistycznego oprogramowania. Należy do niego m.in. aplikacja phoneEMAtool<sup>14</sup> [Mik, Lorenc 2015] umożliwiająca dynamiczną wizualizację trajektorii ruchu sensorów artykulografu<sup>15</sup> w osiach  $X$  (przód – tył) i  $Z$  (górze – dół) oraz analizę i ekstrakcję informacji związanych z położeniem poszczególnych sensorów w czasie we wszystkich osiach, z uwzględnieniem wychyleń kątowych  $\varphi$  i  $\theta$ . Ponadto oprogramowanie pozwala na obliczenie prędkości ruchu sensorów w czasie oraz na wyznaczenie dzięki temu jej minimów i maksimów. W celu zniwelowania niepożądanych zniekształceń zastosowano filtr Savitzky–Golay. Aby uzyskać bardziej precyzyjne wygładzanie, algorytm stosowano dwukrotnie.

<sup>14</sup> Autorzy większości analiz gestów artykulacyjnych prowadzonych na podstawie danych artykulograficznych używają funkcji FindGest, dostępnej w aplikacji MVIEW [Tiede 2005; 2010].

<sup>15</sup> Więcej na temat systemów EMA i artykulografii elektromagnetycznej w publikacjach A. Lorenc [por. np. 2016, 112–119].

**Ryc. 1. Granice gestów artykulacyjnych dla dźwięków mowy o różnym stopniu otwarcia (spółgłoski zwarto-wybuchowe, trąca, aproksymanty, samogłoski) identyfikowane na podstawie profili prędkości.**



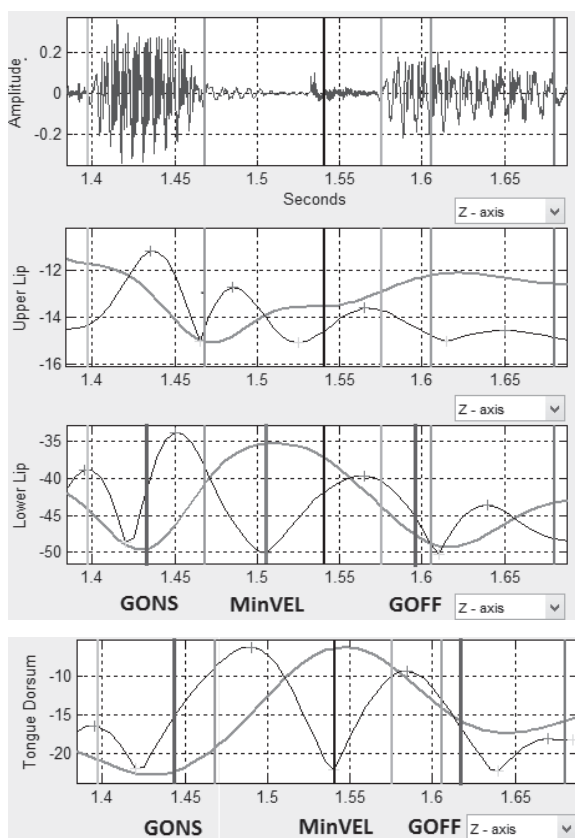
Oznaczenie: a) GONS (ang. *gesture onset*) – początek gestu artykulacyjnego (początkowy, 20-procentowy próg PVEL1 na rosnącym zboczach prędkości); b) PVEL1 (ang. *peak velocity*) – szczyt prędkości w początkowej części gestu; c) NONS (ang. *nucleus onset*) – początek rdzenia gestu artykulacyjnego (kolejny, końcowy, 20-procentowy próg PVEL1 na malejącym zboczach prędkości: początek przewężenia); d) NOFF (ang. *nucleus offset*) – koniec rdzenia gestu artykulacyjnego (początkowy, 20-procentowy próg PVEL2 na rosnącym zboczach prędkości: koniec przewężenia); e) MinVEL (ang. *minimum velocity*) – minimalna prędkość w obrębie rdzenia gestu artykulacyjnego (pomiędzy NONS a NOFF); f) PVEL2 (ang. *peak velocity*) – szczyt prędkości w końcowej części gestu; g) GOFF (ang. *gesture offset*) – koniec gestu artykulacyjnego (końcowy, 20-procentowy próg PVEL2 na malejącym zboczach prędkości).

Program phoneEMAtool [Mik, Lorenc 2015] pozwala na jednoczesne i zsynchronizowane przetwarzanie trzech typów danych: a) audio (\*.wav); b) EMA (\*.txt); c) segmentacji akustycznej (\*.TextGrid). Ryc. 2 poniżej przedstawia okno robocze programu phoneEMAtool podczas przykładowej analizy danych.

W analizie spalatalizowanej spółgłoski [p] w wyrazie *stapiarka* przyjęto, że funkcję kluczowego artykulatora prymarnego pełni dolna warga (c), której ruch prowadzi do uformowania zwarcia z wargą górną (b). Mak-

symalne przewężenie pomiędzy wymienionymi artykulatorami utworzone zostało w punkcie czasowym ok. 1,5 s oznaczonym na ryc. 2.(c) jako MinVEL (ang. *minimum velocity*), czyli w momencie, w którym dolna warga osiągnęła minimalną prędkość ruchu w obrębie rdzenia gestu artykulacyjnego i zarazem swoje maksymalne wzniesienie w osi Z (do góry, w kierunku górnej wargi). Porównując przebieg opisywanego gestu artykulacyjnego z oscylograficznym<sup>16</sup> zapisem fali głosowej (a), można łatwo zauważyć, iż dwuwargowe zwarcie artykulacyjne zostało uformowane mniej więcej w połowie odpowiadającego mu segmentu akustycznego.

**Ryc. 2. Oscylogram (a) i trajektorie ruchu sensorów artykulografu (b) górnej wargi, (c) dolnej wargi i (d) środkowej części języka w osi Z (górną – dół) sekwencji [apɨja] w wyrazie *stapiarka* (mówca AK\_m, plik 176).**



<sup>16</sup> Pionowe linie na oscylogramie to granice segmentacji akustycznej wprowadzonej ręcznie w programie Praat [Boersma, Weenick 2014].

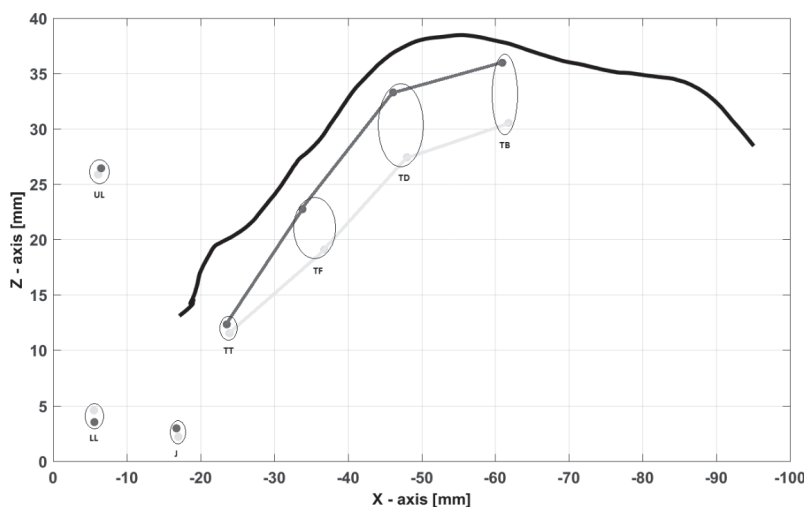
Zważywszy, iż opisywany przykład dotyczy spółgłoski spalatalizowanej, warto prześledzić również trajektorię ruchu artykulatora sekundarnego – środkowej części języka (e). Na ryc. 2.(e) za pomocą czarnej pionowej linii, pojawiającej się w tych samych punktach czasowych w pozostałych oknach, zaznaczono moment, w którym środkowa część języka osiąga swą minimalną prędkość, a zarazem maksymalne wzniesienie ku górze (podniebieniu). Porównując ten moment z punktem czasowym, w którym osiągnięte zostało maksymalne zbliżenie dolnej i górnej wargi (MinVEL na ryc. 2.(c), można stwierdzić, iż gest ten osiągnięty został już po uformowaniu artykulacyjnego zwarcia. Z kolei porównanie z obrazem fali akustycznej pozwala wnioskować, iż maksymalne wzniesienie środkowej części języka związane z gestem palatalizacyjnym pojawiało się już po akustycznym impulsie. Na ryc. 3. przedstawiono położenie sensorów artykulografu umieszczonych na ruchomych narządach mowy w odniesieniu do obrysu artykulatorów pasywnych w dwóch wyżej opisanych momentach. Pierwszy z nich (jasny kolor), a zarazem wcześniejszy, to uzyskanie maksymalnego zbliżenia przez dolną i górną wargę. Wyraźnie widać, iż masa języka jest już w tym czasie wzniesiona w kierunku podniebienia twardego, ale nie jest to jeszcze maksymalna wartość tego ruchu (ciemny kolor). Została ona osiągnięta z opóźnieniem wynoszącym 35 ms w stosunku do artykulacyjnego zwarcia, ale też, jak wcześniej opisano, już po realizacji akustycznego impulsu. Opisany przykład ilustruje tym samym asynchroniczną realizację miękkości spółgłoski wargowej.

Na ryc. 2. (c) i (e) zaznaczono również momenty odpowiadające początkowi (GONS – ang. *gesture onset*) i końcowi (GOFF – ang. *gesture offset*) artykulacyjnych gestów formowania zwarcia dolnej i górnej wargi (c) oraz wznoszenia środkowej części języka (e). Porównując je z przebiegiem fali akustycznej (a), można zauważyć, że powstają one i kończą się w czasie trwania przebiegów regularnych odpowiadających otaczającym samogłoskom.

#### 4. PODSUMOWANIE

Metodologia wypracowana przez fonologię artykulacyjną, w tym propozycja dynamicznej analizy gestów artykulacyjnych, umożliwiła rozstrzygnięcie rozmaitych kontrowersji fonetyczno-fonologicznych, związanych z nazalizacją, retrofleksją, rozróżnianiem artykulacji centralnych i lateralnych [por. Lorenc 2016] czy też ze zjawiskiem palatalizacji rozważanym w niniejszym artykule. Wszystkie wymienione zjawiska należą do najczęściej dyskutowanych od lat w polskich opracowaniach z zakresu fonetyki i fonologii.

**Ryc. 3. Obrys artykulatorów pasywnych i położenie sensorów górnej wargi (UL), dolnej wargi (LL), zuchwy (J), czubka języka (TT), przedniej (TF), środkowej (TD) i tylnej (TB) części języka w osi Z (góra – dół) w momencie osiągnięcia minimalnej prędkości przez sensor dolnej wargi (jasny kolor) i środkowej części języka (ciemny kolor) w wyrazie *stapiarka* (mówca AK\_m, plik 176).**



## Bibliografia

- G. Bloothoof, W. van Dommelen, C. Espain, Ph. Green, V. Hazan, M. Huckvale, E. Wigforss, 1997, *The landscape of future education, Speech Communications Sciences*, vol. 1.: *Analysis*, Utrecht.
- G. Bloothoof, W. van Dommelen, C. Espain, V. Hazan, M. Huckvale, E. Wigforss, 1998, *The landscape of future education, Speech Communication Sciences*, vol. 2.: *Proposals*, Utrecht.
- G. Bloothoof, W. van Dommelen, K. Fellbaum, V. Hazan, M. Huckvale, M. Leahy, E. Wigforss, 1999, *The landscape of future education, Speech Communication Sciences*, vol. 3.: *Recommendations*, Utrecht.
- C.P. Browman, L. Goldstein, 1992, *Articulatory Phonology. An Overview*, „Phonetica” 49, z. 3–4, s. 155–180.
- P. Boersma, D. Weenick, 2014, *Praat: Doing phonetics by computer*, wersja 4.3.14 (program komputerowy), <http://www.praat.org> [dostęp: 20.12.2017].
- R.L. Diehl (red.), 1991, *On the Relation between Phonetics and Phonology*, „Phonetica” 48, z. 2–4.
- R.L. Diehl (red.), 1992, *Articulatory Phonology*, „Phonetica” 49, z. 3–4.
- J. Durand, B. Lacks, 2002, *Phonology, Phonetics and Cognition an Overview* [w:] J. Durand, B. Lacks (red.), *Phonetics, Phonology and Cognition*, Oxford, s. 10–5.



- C.A. Fowler, B. Galantucci, 2005, *The Relation of speech Perception and Speech Production* [w:] D.B. Pisoni, R.E. Remez (red.), *The Handbook of Speech Perception*, Oxford, s. 633–652.
- B. Galantucci, C.A. Fowler, L. Goldstein, 2009, *Perceptuomotor compatibility effects in speech*. *Perception and Psychophysics*, 71, s. 1138–1149 (reprint pdf: <http://www.Haskins.yale.edu/HL1542/>).
- B. Galantucci, C.A. Fowler, M.T. Turvey, 2006, *The motor theory of speech perception reviewed*, „Psychonomic Bulletin and Review” 13 (3), s. 361–377.
- W.J. Hardcastle, J. Laver, F.E. Gibbon (red.), 2010, *The Handbook of Phonetics Sciences*, wyd. II, Oxford, Cambridge Massachusetts.
- P. Jackson, V. Singampalli, 2008, *Statistical identification of critical, dependent and redundant articulators*, „Journal of the Acoustical Society of America” 123, s. 3321.
- R.D. Kent, 1997, *The Speech Sciences*, San Diego–London.
- J. Kim, S. Lee, S. Narayanan, 2014, *Estimation of the movement trajectories of non-crucial articulators based on the detection of crucial moments and physiological constraints*, 15th Annual Conference of the International Speech Communication Association (INTERSPEECH) 2014, 14–18 September, Singapore, s. 163–168.
- P. Kuhl, 2007, *Język, umysł i mózg: doświadczenie zmienia percepcję* [w:] I. Kurcz (red.), *Psychologia języka dziecka*, Gdańsk, s. 34–62.
- P. Ladefoged, I. Maddieson, 1996, *The Sounds of the World Languages*, Oxford.
- H. Lane, 1970, *Production et perception de la parole: rapports et differences*, *Nouvelles perspectives en phonétique*, Bruxelles, s. 87–114.
- A.M. Liberman, I. Mattingly, 1985, *The motor theory revisited*, „Cognition” 21, s. 1–36.
- A. Lorenc, 2016, *Wymowa normatywna polskich samogłosek nosowych i spółgłoski bocznej*, Warszawa.
- P. Łobacz, 1970, *An acoustic analysis of the opposition /j:i/ and /w:u/ in present-day Polish* [w:] W. Jassem (red.), *Speech Analysis and Synthesis*, vol. 2, Warszawa, s. 119–134.
- P. Łobacz, 1979, *Perception of Polish vocalic segments compared to quasi-automatic recognition*, „Lingua Posnaniensis” XX, s. 97–107.
- I. Maddieson, 1984, *Patterns of sounds*, Cambridge–London–New York–New Rochelle–Melbourne–Sydney.
- A.F. Majewicz, 1986, *A contrastive analysis of Polish and Japanese phonetic systems with special reference to the teaching of Japanese in Poland*, Poznań.
- Ł. Mik, A. Lorenc, D. Król, R. Wielgat, R. Święciński, R. Jędryka, 2018, *Fusing the electromagnetic articulograph, high-speed video cameras and a 16-channel microphone array for speech analysis*, „Bulletin of the Polish Academy of Sciences Technical Sciences”, vol. 66 [w druku].
- Ł. Mik, A. Lorenc, 2015, phoneEMAtool [program komputerowy].
- J. Ohala, 1981, *Speech timing as a tool in phonology*, „Phonetica” 38, s. 204–212.
- J. Ohala, 1997, *The Relation between Phonetics and Phonology* [w:] W.J. Hardcastle, J. Lever (red.), *The Handbook of Phonetic Sciences*, Oxford–Cambridge Massachusetts, s. 674–694.
- B. Ostapiuk, 2013a, *Dyslalia. O badaniu jakości wymowy w logopedii*, Szczecin.

- B. Ostapiuk, 2013b, *Głoska jako dźwięk fonemowy w mowie oraz w logopedycznym badaniu* [w:] J. Panasiuk, T. Woźniak (red.), *Język. Człowiek. Społeczeństwo*, Lublin, s. 717–734.
- P. Perrier, S. Fuchs, 2015, *Motor Equivalence in Speech Production* [w:] M.A. Redford (red.), *The handbook of Speech production*, West Sussex, s. 225–247; <https://hal.archives-ouvertes.fr/hal-01205656> [dostęp od 29.08.2015].
- D. Recasens, M.D. Pallarès, J. Fontdevila, 1997, *A model of lingual coarticulation based on articulatory constraints*, „Journal of the Acoustical Society of America” 102(1), s. 544–561.
- E. Saltzman, J. Kelso, 1987, *Skilled actions: A task dynamic approach*, „Psychological Review” vol. 94, s. 84–106.
- E. Saltzman, K.A. Munhall, 1989, *Dynamical approach to gestural patterning in speech production*, „Ecological Psychology” vol. 1, s. 333–382.
- M. Tiede, 2005, *MVIEW: Software for visualization and analysis of concurrently recorded movement data* [program komputerowy].
- M. Tiede, 2010, *MVIEW: Multi-channel visualization application for displaying dynamic sensor movement* [program komputerowy].
- M. Wagił, 2016, *Pozycje nuklearne i peryferyjne a interpretacja fonologiczna aproksymantów [j, w] w języku polskim*, „Język Polski” t. XCVI, z. 1, s. 68–79.

### **Reflections on the principles of teaching modern phonetics of the Polish language**

#### Summary

Due to the complex interdisciplinary nature of modern phonetic sciences and the development of individual fields of phonetics, it is important to reinterpret some of its crucial aspects. There is still a large collection of phenomena related to the production and perception of speech, which results in the occurrence of certain ambiguous interpretations of phonetic facts and controversies between phonetics and phonology. On their basis, articulatory phonology has been established since the mid-1980s. This proposal, regarding the abstract articulatory gesture with the tools enabling the investigation of articulation as a process (EMA systems), is particularly attractive since it allows a holistic approach to the speech sounds system due to the close relationship between articulation, the image of the speech wave and in particular the mental processes of its reception by the listener. This is important for the phoneticians who teach the pronunciation of individual languages within broadly defined glottodidactics, as well as for speech therapists, who are trained in listening skills at the phonetic (speech sounds) level during classes in the so-called auditory training.

Adj. Monika Czarnecka