

Maciej Karpiński

Wydział Neofilologii

Uniwersytet im. Adama Mickiewicza

maciejk@amu.edu.pl

ORCID: <https://orcid.org/0000-0001-5532-6188>



Laboratorium i korpus MultiCo

Słowa kluczowe: humanistyka i nauki o sztuce, infrastruktura cyfrowa, cyfryzacja, analiza gestów, komunikacja niewerbalna, sztuczna inteligencja

Współczesne badania nad komunikacją interpersonalną, w tym językoznawcze, coraz częściej przyjmują perspektywę multimodalną, która oznacza uwzględnianie większej liczby modalności sensorycznych. Wykraczają one zatem poza akustyczny wymiar wypowiedzi ustnej (głos), obejmując również m.in. udział zmysłów wzroku (mimika, gestykulacja, zmiany pozycji ciała) czy dotyku. Stanowi to konieczny krok na drodze do zrozumienia i efektywnego modelowania mechanizmów komunikacji językowej, także w perspektywie ewolucyjnej i rozwojowej. Jakkolwiek taka koncepcja wydaje się intuicyjnie oczywista, wymaga ona szeregu zmian w podejściu do badań, począwszy od kwestii metodologicznych, aż po sferę techniczną czy nawet organizacyjną.

Gromadzenie danych do badań uwzględniających multimodalność procesu komunikacji stanowi wyzwanie techniczne – wymaga odpowiedniej aparatury, oprogramowania, procedur oraz bardzo dużych nakładów pracy specjalistów. Wśród wyzwań organizacyjnych wymienić można rekrutację uczestników nagrań, pozyskanie od nich odpowiednich zgód, zabezpieczenie ich interesów i dokonanie szeregu uzgodnień prawnych, związanych ze sposobami przechowywania i płatnego lub bezpłatnego udostępniania danych, np. w obszarze badań lub dydaktyki. Przekłada się to na niebagatelne koszty i czasochłonność prac. Nie wszyscy zainteresowani dysponują możliwościami pozyskania takiego materiału badawczego. A nawet jeśli takie posiadają, to jego samodzielne gromadzenie wyłącznie na własne potrzeby może się okazać nieekonomiczne, przecząc postulatowi wielokrotnego wykorzystywania tych samych danych do różnych celów badawczych. W środowisku naukowym, w tym również badawczo-rozwojowym, rośnie zatem zapotrzebowanie na zasoby, które pozwoliłyby analizować multimodalny proces komunikacji z różnych perspektyw teoretycznych, różnymi metodami i technikami badawczymi, pod kątem różnych celów poznawczych i zastosowawczych. Na przeciw takiemu zapotrzebowaniu wychodzą utworzone w ramach projektu [DARIAH-PL](#)¹ laboratorium badawcze oraz powstały w nim korpus [MultiCo](#).

Laboratorium wyposażono w aparaturę do wielokanałowej rejestracji dźwięku i obrazu, a także do rejestracji ruchu z wykorzystaniem bezmarkerowych systemów mocap (*motion capture*). Możliwa jest jednoczesna rejestracja zachowań komunikacyjnych kilku osób, choć system zoptymalizowano pod kątem rejestracji monologów i dialogów. Stacje robocze i odpowiednie oprogramowanie umożliwiają edycję, przetwarzanie i analizę plików z zarejestrowanymi zachowaniami komunikacyjnymi, w tym fonetyczno-akustyczną analizę warstwy dźwiękowej wypowiedzi oraz analizę ruchów

¹ Wszystkie odesłania do stron internetowych przedstawiają wersję aktualną w dn. 5.04.2024 r.

mówcy – zarówno z zastosowaniem specjalistycznego systemu analizy biomechanicznej, jak i wspomaganą komputerowo analizą eksperckiej. Na podstawie danych mocap możliwe jest tworzenie modeli ruchu oraz generowanie animacji. Dostępna w laboratorium aparatura może zostać elastycznie skonfigurowana do badań prowadzonych zarówno na miejscu, jak i do prac terenowych w obiektach zamkniętych i na wolnym powietrzu. Sercem laboratorium jest serwer wraz z powstałym w zespole laboratorium oprogramowaniem bazodanowym do gromadzenia, archiwizacji i udostępniania zasobów multimodalnych CorpusMINI. Oprogramowanie to pozwala nie tylko na efektywne zarządzanie archiwami i zapewnienie bezpiecznego przechowywania i dostępu, lecz również na organizację zespołowej pracy nad analizą materiałów.

Wizytówką możliwości laboratorium jest korpus MultiCo złożony z trzech głównych komponentów, zawierających nagrania monologów, dialogów oraz polilogów. Pierwszy podkorpus PERSWAZJA oparto na dostępnych publicznie nagraniach wystąpień sejmowych (Archiwum Sejmu RP, ok. 2,5 godziny) oraz nagraniach z krajowej edycji TEDx (YouTube, ok. 2,5 godziny). W obu przypadkach są to mowy o charakterze perswazyjnym, realizowane ze stosunkowo ograniczonym poziomem spontaniczności – z reguły wcześniej przygotowane. Podkorpus DIALOGI zarejestrowano w laboratorium z wykorzystaniem kamer tradycyjnych oraz mocap i dodatkowego omikrofonowania. Zawiera wysokiej jakości nagrania zadania dialogowego, polegającego na wspólnym budowaniu w wyobraźni wieży z klocków. Każdy z uczestników był rejestrowany przez dwie kamery konwencjonalne, odrębny system mocap oraz dwa mikrofony pojemnościowe (nagłówny i umieszczony na statywie). Realizację zadania przez dwadzieścia par nagrano w warunkach wzajemnej widoczności, przez kolejnych dwadzieścia – przy braku widoczności (między uczestnikami znajdowała się przestona). Każda sesja trwała około 5 minut, co łącznie daje około 3,5 godzin dialogów zapisanych w około 14 godzinach nagrań wideo, około 14 godzinach nagrań audio oraz 7 godzinach nagrań mocap. Podkorpus polilogowy SPORT zawiera ok. 5 godzin nagrań z telewizyjnego programu publicystycznego o tematyce sportowej, w którym występuje zwykle czworo uczestników (prowadzący i troje gości).

Ze względu na duże różnice w naturze zgromadzonych materiałów podkorpusy istotnie różnią się sposobem opisu. W przypadku korpusów opartych na gotowym materiale filmowym w szeregu sytuacji pełny opis gestykulacji nie był możliwy (kadrowanie, ruchy kamery). W przypadku korpusu polilogowego problem stanowiła również transkrypcja wypowiedzi, gdyż te mogły się pokrywać i nie być w pełni zrozumiałe. W przypadku podkorpusów PERSWAZJA i SPORT przeprowadzono ekspercką analizę podziału na frazy gestowe, natomiast w korpusie DIALOGI dane o ruchu uczestników pochodzą z systemu mocap. Wszystkie podkorpusy wyposażono w synchroniczną transkrypcję ortograficzną oraz fonematyczną, z podziałem na realizacje sylab i głosek (tabela 1).

Dzięki zespołowi ekspertów z wieloletnim doświadczeniem laboratorium może oferować wsparcie w gromadzeniu i analizie danych multimodalnych związanych z komunikacją językową innym badaczom, lecz również usługi komercyjne z obszaru badania zachowań komunikacyjnych, jakości i efektywności komunikacji oraz specyfiki jej przebiegu. Można w nim realizować ekspertyzy dotyczące wystąpień publicznych, negocjacji, przekazu informacji w warunkach biznesowych, w kontekście marketingu, edukacji, sztuki (teatr, taniec, muzyka i inne) i wielu innych.

Tabela 1. Poziomy anotacji materiału w korpusie MultiCo

Podkorpus	Media (liczba kanałów)			Transkrypcja mowy							Anotacja gestów
	audio	wideo	animowany szkielet	ort.	fon.	seg.	fraz.	wyr.	syl.	głoski	
PERSWAZJA (monologi)	1	1	--	+	+	+	+	+	+	+	Manualny podział na frazy gestowe
ZADANIE (dialogi)	4	4	2	+	+	+	+	+	+	+	Dane z systemu motion capture
SPORT (polilogi)	1	1	--	+	+	+	+	+	--	--	Manualny podział na frazy gestowe (wybrane fragmenty)

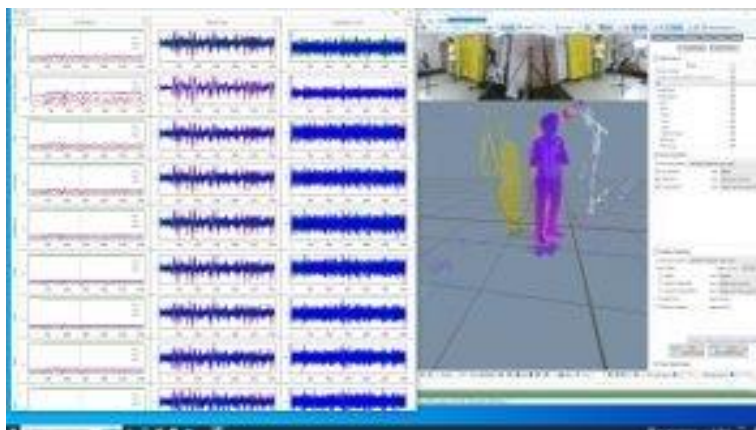
Źródło: Specyfikacja korpusu MultiCo (materiały własne).

Możliwa jest rejestracja sygnału mowy w różnych warunkach (kabina bezdechowa, inne rodzaje pomieszczeń, otwarta przestrzeń), z wykorzystaniem wielu typów mikrofonów i ich kombinacji (nagłownych, statywowych, o różnych charakterystykach kierunkowych i częstotliwościowych), z rejestracją dźwięku na ośmiu niezależnych kanałach (wielokanałowy interfejs audio współpracujący z programami Cubase Pro oraz WaveLab Pro). Nagrania są anotowane i analizowane głównie z wykorzystaniem programów Annotation Pro (oprogramowanie własne) oraz Praat (niezwykle popularny darmowy program do analizy sygnału mowy). Rejestrację wizji można realizować czterema kamerami o rozdzielczości 4k z możliwościami transmisji danych zarówno przez sieć bezprzewodową, jak i połączenia kablowe DNS oraz czterema kamerami HD przy transmisji przez połączenie kablowe HDMI. Do rejestracji i edycji materiału służy system oparty na oprogramowaniu BlackMagic, wielokanałowej karcie wideo oraz wydajnej karcie graficznej. Wspomagane komputerowo eksperckie analizy ruchów ciała lub mimiki odbywają się z wykorzystaniem bezpłatnego oprogramowania ELAN. Do rejestracji ruchu można wykorzystać kamery z detekcją głębi Kinect2 lub Kinect Azure współpracujące z oprogramowaniem iPi. Analizy danych o ruchu realizuje się m.in. z zastosowaniem oprogramowania iPi Biomech oraz darmowych programów statystycznych, jak np. R. Laboratorium dysponuje przenośnym systemem oświetlenia oraz zestawem ustrojów akustycznych, umożliwiającą adaptację akustyczną pomieszczeń odpowiadającą różnym zadaniom. Jest gotowe zarówno do realizacji zadań w ramach wspierania badań naukowych, podstawowych i zastosowawczych, jak i do prac komercyjnych na potrzeby podmiotów gospodarczych i indywidualnych zleceniodawców.



Fot. 1. System rejestracji wizji i fonii

Fot. Maciej Karpiński



Fot. 2. Oprogramowanie do biomechanicznej analizy ruchu (iPi BioMech)

Fot. Maciej Karpiński



Fot. 3. Pomieszczenie laboratoryjne do rejestracji zachowań komunikacyjnych w badaniach nad komunikacją multimodalną

Fot. Maciej Karpiński

Tekst powstał w ramach projektu POIR.04.02.00-00-D006/20 „Cyfrowa infrastruktura badawcza dla humanistyki i nauk o sztuce DARIAH-PL” realizowanego w latach 2021-2023.



Fundusze Europejskie
Inteligentny Rozwój



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

