KONRAD JUSZCZYK

TWORZENIE KORPUSÓW MULTIMODALNYCH DO BADANIA MOWY, JĘZYKA I KOMUNIKACJI MIĘDZYLUDZKIEJ

Tworzenie korpusów multimodalnych do badania mowy, języka i komunikacji międzyludzkiej

Biblioteka Metodologii Lingwistyki 1

Konrad Juszczyk

Tworzenie korpusów multimodalnych do badania mowy, języka i komunikacji międzyludzkiej

wydawnictwo

Poznań 2024

Projekt okładki:
Konrad Juszczyk

Grafika na okładce została wygenerowana przez DALL-E.

Recenzent:
dr hab. Włodzimierz Lapis, prof. UAM

Biblioteka Metodologii Lingwistyki 1
pod redakcją Włodzimierza Lapisa i Konrada Juszczyka

Wydanie I
Poznań 2024

ISBN 978-83-68006-05-6

DOI 10.48226/978-83-68006-05-6

Wydanie:

Wydawnictwo Rys
ul. Kolejowa 41
62-070 Dąbrówka
tel. 600 44 55 80
e-mail: tomasz.paluszynski@wydawnictworys.com
www.wydawnictworys.com

Spis treści

W tym miejscu autor dziękuje:

Katedrze Metodologii Lingwistyki Uniwersytetu im. Adama Mickiewicza w Poznaniu za dofinansowanie wydania niniejszego opracowania.

Członkom zespołu DiaGest – Agnieszce Czosce, Ewie Jarmołowicz-Nowikow, Maciejowi Karpińskiemu, Katarzynie Klessie, Wojciechowi Laskowskiemu, Zofii Malisz i Michałowi Szczyszkowi – za możliwość udziału w licznych projektach badawczych poświęconych komunikacji multimodalnej.

Członkom zespołu projektu zwanego w skrócie MULTIMET – Bożenie Grabowskiej, Ewie Jarmołowicz-Nowikow, Victorii Kamasie, Barbarze Konat, Agnieszce Marlińskiej, Bożenie Pieskiewicz i Michałowi Szczyszkowi za możliwość realizacji nagrań sesji coachingowych z udziałem 50 uczestników.

Kamilowi Ciecierskiemu, który napisał aplikację Kinemo.

Załączone w opracowaniu komendy ELAN-a oraz fragmenty kodu źródłowego w Pythonie i komendy w Bashu były testowane na następujących wersjach języków i bibliotek:

- ELAN: 6.7;
- Python: 3.12.0;
- Pympi-ling: 1.70.2;
- Bash: 5.2.15(1);
- FFmpeg: 6.0.1;
- Whisper: v20231106.

Dostęp do wszystkich podanych w książce linków do stron internetowych został sprawdzony w listopadzie 2023 roku.

Skrypty i kod podany w publikacji jest także udostępniony w postaci notesów Google Colab, których adresy są zalinkowane w nazwach notesów poniżej oraz w repozytorium github:

- Skrypty w bashu do generowania próbek filmów na postawie anotacji z pliku *.eaf* (opisane w sekcji 2.6.19 niżej) są pod nazwą EAF2CSV2MP4.ipynb.
- Skrypt do istalacji i uruchomienia usług *Whisper* (opisany w sekcji 3.2.5.3 niżej) jest dostępny pod nazwą Whisper.ipynb.
- Kod w Pythonie do przertwarzania anotacji w *Pympi-Ling* jest (opisany w sekcji 3.6 niżej) pod nazwą ELAN-PYMPI-LING.ipynb.

Źródła ilustracji:

XLR, RCA, JACK, VGA, DVI, DISPLAYPORT, HDMI, ETHERNET, USB-C, USB, SIECIOWY TYP C, WTYK DC, C7 lub ÓSEMKA, C5 lub MYSZKA MIKI, Trump i Obama, Mikrofon: image from pngtree.com, Osoba stojąca, Osoba siedząca: Image by Freepik, Kamera na statywie: Apple Keynote, kadry z badań Deba Roy'a.

Wprowadzenie

Celem niniejszego przewodnika jest przedstawienie procesu tworzenia korpusów multimodalnych na potrzeby badania mowy, języka i międzyludzkiej komunikacji multimodalnej. Praca składa się z pięciu rozdziałów. W pierwszym wyjaśniamy czemu służą korpusy multimodalne i opisujemy je na przykładach projektów realizowanych przez grupę DiaGest i autora. W tym rozdziale opisano także metody badań komunikacji multimodalnej. Drugi rozdział dotyczy tworzenia korpusu nagrań, czyli zbierania danych od uczestników, wyboru aparatury oraz archiwizacji i anonimizacji nagrań. Szczególnie dużo miejsca poświęcono przetwarzaniu nagrań filmowych przy pomocy pakietu darmowego *ffmpeg,* którego komendy łatwiej i szybciej wykonać na dowolnej liczbie plików niż klikać w programach z interfejsem graficznym. Trzeci rozdział odpowiada trzeciemu etapowi tworzenia korpusu – przygotowaniu transkrypcji i anotacji zachowań komunikacyjnych. Omówiono tutaj wybrane funkcje ELANa – najbardziej popularnego programu do badań komunikacji multimodalnej oraz sposoby przetwarzania plików z ELANa przy pomocy darmowej biblioteki *pympi-ling* w języku *Python.* Przedstawiono także narzędzia konsorcjum CLARIN do automatycznej analizy jednostek języka i wspomniano o systemach anotacji zachowań komunikacyjnych. Przewodnik zamykają dwa krótkie rozdziały, jeden dotyczy udostępniania danych, a drugi jest krótkim podsumowaniem przewodnika. W pracy pokazano jak prowadzimy badania, ale nie opisano oczywiście wszystkiego, co robimy ani też nie uważamy, że podane rozwiązania są najlepsze. Czytelników znających inne rozwiązania zapraszamy do kontaktu z autorem. Wierzymy, że prezentacja praktyk badawczych będzie wsparciem dla przyszłych korpusowych badań mowy, języka i multimodalnej komunikacji międzyludzkiej.

1. Istota korpusowych badań komunikacji

Tworzenie korpusu nagrań z udziałem ludzi do celów badawczych wymaga wielu zabiegów. Niniejsza publikacja jest zbiorem zaleceń dotyczących pozyskiwania i przetwarzania danych z nagrań dźwiękowych i wizualnych. Pomijając liczne szczegóły metodologiczne, którymi różnią się naukowe badania komunikacji, wspólnym celem nagrań naukowych jest zebranie wielu wystąpień podobnych zjawisk i zachowań ludzi. Dlatego chcąc zarejestrować powtarzalne zjawiska i zachowania, do nagrań zapraszamy wiele osób, ale każda z nich powinna wystąpić w tej samej roli, zmieścić się w tym samym kadrze oraz mówić i poruszać się w tych samych lub możliwie najbardziej zbliżonych okolicznościach. Badacz świadomie wybiera warunki nagrań, które kontroluje, bo zakłada wpływ co najmniej jednego warunku na przebieg nagrań i zadania, jakie mają nagrywani. Zmiana ustawienia urządzeń nagrywających, kadru, poruszanie kamerą czy mikrofonem, a także obecność (lub jej brak) innych osób w sali nagrań – wszystko to może oddziaływać na wyniki badań.

1.1. Projekty multimodalne

Multimodalne korpusy powstają rzadziej niż korpusy mowy lub tekstu. Charakterystykę takich korpusów przedstawiam na podstawie doświadczeń zebranych w projektach, w których uczestniczyłem jako badacz, oraz kilku wybranych publikacji (wymienionych w bibliografii), które przedstawiają multimodalną komunikację międzyludzką. W latach 2008-2017 byłem członkiem grupy DiaGest i uczestniczyłem w trzech projektach kierowanych przez prof. Macieja Karpińskiego (NARRACJE[1], ORIGAMI[2],

[1] Pełny tytuł projektu to „Komplementarność niewerbalnych i werbalnych składników wypowiedzi: Analiza porównawcza dzieci i dorosłych". Projekt był finansowany przez Ministerstwo Edukacji Narodowej RP (199/N-COST/2008/0) i realizowany w latach 2008-2009.

[2] Pełny tytuł projektu to „Interakcja werbalna i niewerbalna w dialogach zadaniowych. Modele multimodalnych aktów dialogowych". Projekt był finansowany przez Komitet Badań Naukowych MNiSW (MNiSW: N N104 010337) i realizowany w latach 2009-2010.

BORDERLAND[3]) i kierowałem dwoma projektami (MULTIMET[4] i KINEMO[5]). W projektach tych w sumie wzięło udział około 300 uczestników, nagraliśmy łącznie około 5500 minut audio i wideo, z czego przetranskrybowano i zanotowano około 2000 minut. Projekty omawiam skrótowo, by porównać cele badań, zadania uczestników, wielkości korpusów oraz zakres transkrypcji i anotacji. Wyniki analiz transkrypcji i anotacji zebranych w projektach są opisane w licznych publikacjach (Karpiński i in. 2008; Karpiński 2009b; Jarmołowicz-Nowikow 2009; Karpiński i Jarmołowicz-Nowikow 2010; Jarmołowicz-Nowikow i Karpiński 2011; Juszczyk 2011; Szczyszek 2013; Karpiński i in. 2018; Juszczyk 2017).

1.1.1. Cele badań

Wspólnym celem projektów była analiza zachowań komunikacyjnych w dialogach i monologach. Projekt zwany tutaj NARRACJAMI wzorowany był na badaniach opowiadania historii w zespole McNeilla (McNeill 1992; 2005), a celem projektu zwanego ORIGAMI było zebranie materiału multimodalnego z zadania z figurą składaną z papieru, w którym uczestnicy byli skłonieni do zarządzania dialogiem w dwóch warunkach: widzenia się i niewidzenia. W projekcie, którym kierowałem (zwanym w skrócie MULTIMET), celem badań były między innymi analizy multimodalnych wyrażeń metafor konceptualnych (Juszczyk 2017). Celem nagrań w projekcie KINEMO było zebranie testowych nagrań do aplikacji KINEMO, która rejestruje i rozpoznaje ruchy rąk przy pomocy czujnika ruchu Microsoft Kinect.

---

[3] Pełny tytuł projektu to „Język pogranicza – pogranicze języka. Parajęzykowe aspekty komunikacji". Projekt był finansowany przez Narodowy Program Rozwoju Humanistyki MNiSW (12H 13 0524 82) i realizowany w latach 2014-2017.

[4] Pełny tytuł projektu to „Multimodalne wyrażenia metafor konceptualnych a spójność i synchronia zachowań komunikacyjnych w dialogu". Projekt był finansowany przez Narodowe Centrum Nauki (2011/03/D/HS6/05993) i realizowany w latach 2012-2017.

[5] Aplikacja KINEMO powstała w ramach projektu FNP INTER, który był współfinansowany ze środków przyznanych przez Fundację na rzecz Nauki Polskiej na podstawie Umowy nr 126/UD/SKILLS.2013 i wykorzystania nagrody przyznanej w konkursie popularyzatorskim w ramach projektu SKILLS współfinansowanego z Europejskiego Funduszu Społecznego. Projekt był realizowany w latach 2014-2015.

1.1.2. Zadania uczestników

Opisywane projekty są przykładami różnych sposobów nakłaniania uczestników do interakcji, czyli elicytacji zachowań komunikacyjnych. Pierwszym z nich jest dialog zadaniowy (ORIGAMI i BORDERLAND), drugim – opowiadanie historii na podstawie wcześniej obejrzanego filmu animowanego (NARRACJE), a trzecim – dialog w formie sesji coachingowej (MULTIMET). Czwartym sposobem elicytacji jest eksperyment behawioralny, w którym mierzy się na przykład czas reakcji i inne zmienne dotyczące zachowania, ale w omawianych projektach nie stosowano psycholingwistycznych metod badań.

1.1.2.1. NARRACJE

Uczestnicy oglądali minutowy fragment filmu animowanego dla dzieci o kocie Sylwestrze i ptaszku Tweetym (Canary Row 1950). Następnie każdy uczestnik był proszony o zreferowanie wydarzeń pokazanych na filmie. W ten sposób elicytowano wypowiedzi multimodalne, czyli wielokanałowe: złożone zarówno z mowy, jak i ruchów rąk oraz całego ciała. Nagrania dorosłych przeprowadzono w studiu zaaranżowanym w budynku UAM. Natomiast dzieci nagrano w szkołach, czyli w miejscu im znanym, w którym mogły się czuć pewnie i bezpiecznie. Zgodę na nagrania dorośli wyrażają samodzielnie, ale w przypadku dzieci zgody uzyskano zarówno od samych nagrywanych, jak i od ich rodziców lub opiekunów oraz dyrekcji szkół (Karpiński i in. 2008).

1.1.2.2. ORIGAMI

Uczestnicy byli nagrywani w parach, w dwóch warunkach. Jedna osoba instruowała drugą, jak ma złożyć konstrukcję z papieru, spinaczy i zapałek. Instruujący miał przed sobą złożoną konstrukcję, a instruowany tylko elementy składowe. Warunki różniły się tym, że uczestnicy widzieli się nawzajem lub nie, czyli byli przedzieleni zasłoną. W ten sposób chciano sprawdzić, jak wzajemne widzenie się wpływało na różnice obserwowane w wypowiedziach i gestach

uczestników nagrań. Nagrywano ich czterema kamerami: po jednej na każdego z uczestników na wprost i z boku. Takie ustawienie kamer służyło zebraniu materiału pokazującego ruchy rąk z dwóch perspektyw, by ich opis był dokładniejszy.

1.1.2.3. MULTIMET

Uczestników zaproszono do udziału w dwóch sesjach coachingowych prowadzonych przez zawodowych coachów, którzy zadawali pytania dotyczące kariery zawodowej. W pytaniach coachowie powtarzali wybrane słowa i gesty uczestników, by nawiązywać do treści ich wypowiedzi. Uczestnicy nie byli uprzedzani przed nagraniem, że coachowie będą powtarzać ich słowa i gesty.

1.1.2.4. KINEMO

Uczestnicy byli nagrywani w parach, w których prowadzili rozmowy na dowolne tematy, lub brali udział w sesji coachingowej prowadzonej przez zawodowego coacha. Drugie zadanie polegało na wykonywaniu prostych ruchów rąk według instrukcji wyświetlanych na ekranie komputera. W rezultacie zarejestrowano podobne ruchy rąk pochodzące od różnych uczestników. Nagrania i anotacje posłużyły do optymalizacji algorytmów rozpoznawania ruchów rąk.

1.1.3. Wielkość korpusów

Korpusy powstałe w opisywanych projektach mają od kilkuset do kilku tysięcy minut nagrań dźwięku i obrazu podczas zachowań komunikacyjnych kilkudziesięciu uczestników w każdym z projektów (średnio 60 osób), a w sumie około 300 uczestników. Czas trwania jednego nagrania mieści się w przedziale od 5 minut w przypadku dialogów zadaniowych i narracji do prawie godzinnych nagrań sesji coachingowych w projekcie MULTIMET, ale średni czas trwania sesji coachingowych to 40 minut, a w projekcie KINEMO ograniczono go

do 25 minut. Zadania wykonywane w projekcie KINEMO zajmowały po kilka minut.

1.1.4. Transkrypcja i anotacja

Nie wszystkie nagrania zostały przetranskrybowane ortograficznie, bo zwykle około jedna piąta nagrań zgromadzonych w projekcie nie nadaje się do analiz z przyczyn technicznych. Transkrypcja ortograficzna i fonemiczna, czyli oparta na systemie SAMPA (Wells 2000; Demenko, Wypych i Baranowska 2003), została wykonana przez członków zespołów badawczych i studentów filologii polskiej. W transkrypcjach uwzględniono także jednostki parawerbalne, czyli śmiech, westchnienia, okrzyki (Szczyszek 2013). W projektach ORIGAMI i NARRACJE wypowiedzi zostały zanotowane w celu oznaczenia części mowy na warstwie leksykalnej, składni zdań na warstwie syntaktycznej oraz budowy wyrazów na warstwie słowotwórczej (Karpiński i in. 2008; Szczyszek 2013). Wypowiedzi w projektach ORIGAMI zostały zanotowane w celu klasyfikacji ruchów dialogowych w systemie zaproponowanym przez autorów pierwotnego zadania Map Task (Anderson i in. 1991). Transkrypcje ortograficzne i anotacje ruchów dialogowych i gestów wykonano ręcznie w różnych programach: MS WORD, MS EXCEL, ELAN (Auer i in. 2010), PRAAT (Boersma i Weenink 2006) i automatycznie w serwisie CLARIN WEBMAUS. Anotacja zachowań niewerbalnych była wykonywana w programie ELAN i w systemie klasyfikacji gestów McNeilla lub kodowania ruchów rąk NEUROGES (Lausberg 2013; 2019; Lausberg i Sloetjes 2015) w przypadku projektów MULTIMET i KINEMO.

1.2. Metody badań komunikacji multimodalnej

Podział badań wedle metodologii pozyskiwania i analizy danych wyznacza cztery główne typy: introspekcyjne, obserwacyjne (korpusowe), eksperymentalne, ankietowe. Omówimy krótko każdy z nich, uwzględniając problemy dotyczące stopnia zbliżenia badanych zjawisk do zjawisk występujących w rzeczywistości. Żadna metoda badawcza,

żadne narzędzie do analiz czy aparatura nie pozwala na pełen opis badanych zjawisk. W przypadku badań języka dyskusja nad empirycznością badań toczy się co najmniej od starożytności (Heinz 1983). Pytania, po czym poznać czy badania są empiryczne, jakie przykłady należy opisać, ile przykładów należy zebrać, ilu użytkowników przepytać czy nagrać, pozostają otwarte. Językoznawcy zajmują wobec tego problemu skrajne stanowiska. Generatywiści są przeciwni badaniom wypowiedzi użytkowników języka, a w ocenie zdań generowanych przez reguły gramatyki uniwersalnej posługują się intuicją idealnego mówcy-słuchacza (Chomsky 1965). Natomiast przedstawiciele kognitywnego językoznawstwa korpusowego przeciwnie – uważają, że przedmiotem badań są właśnie wypowiedzi rodzimych użytkowników języka w takiej postaci, w jakiej uda się je zapisać lub zarejestrować w korpusie (Glynn i Fischer 2010). Fabiszak i Konat uważają, że „korpusy zwiększają stopień pewności, z jakim można przyjmować wyniki badań prowadzonych w językoznawstwie kognitywnym. Według metodologii nauki, stopień pewności z jakim można przyjmować wyniki badań zależy od stopnia w jakim spełnione są warunki intersubiektywnej komunikowalności badań. W naukach empirycznych, ważnym kryterium naukowości jest intersubiektywna komunikowalność i intersubiektywna sprawdzalność prowadzonych badań" (Such i Szcześniak 2006; Nowak 1977; Fabiszak i Konat 2013: 138).

Wypowiedzi wyszukane w korpusie są jednak jedynie tymi, które są w korpusie poświadczone jako występujące w tekstach zebranych w danym korpusie, a nie wypowiedziami, które padają w użyciu języka. Jeśli natomiast nie znajdujemy w korpusie wypowiedzi, o której uważamy, że jest w danym języku prawdopodobna, to nie oznacza, że nie mogłaby wystąpić w użyciu języka (Fabiszak i Konat 2013). Badania korpusowe języka pozwalają na wyznaczenie częstości i innych cech jednostek języka, ale częstość w korpusie nie jest tożsama z częstością użycia. Zakłada się jednak, że częstość wyliczona na podstawie korpusów referencyjnych jest najbardziej zbliżona do częstości użycia, bo korpus referencyjny jest największy i stanowi wspólny punkt odniesienia dla interpretacji wyników z różnych badań. Twórcy korpusów są świadomi problemu reprezentatywności zawartych w nich danych językowych. Nie wiadomo, między innymi, w jakich proporcjach mają być reprezentowane poszczególne odmiany języka, by były to proporcje

reprezentatywne względem całości użycia języka, bo te proporcje nie są znane i są trudne do ustalenia (Mykowiecka 2007; Fabiszak i Konat 2013). Kolejnym problemem korpusów jest ich zrównoważenie, czyli „dbałość o taką budowę korpusu, by żaden składnik na żadnym z poziomów nie dominował nad innymi" (Przepiórkowski i in. 2012). Twórcy NKJP – Narodowego Korpusu Języka Polskiego – przyznają jednak, że reprezentatywność i zrównoważenie wykluczają się, więc korpus jest reprezentatywny i zrównoważony w pewnym, umownym stopniu. W jednym z największych korpusów języka angielskiego – *British National Corpus* – udział danych z nagrań to 10% ze 100 milionów słów całości korpusu. Podobnie przyjęto w największym korpusie języka polskiego – Narodowym Korpusie Języka Polskiego, gdzie 10% z 300 milionów słów stanowią dane mówione, czyli transkrypcje ortograficzne wywiadów i debat z programów telewizyjnych i radiowych oraz stenogramy posiedzeń sejmowych i sejmowych komisji śledczych, a także zapisy konwersacji i tekstów czytanych. W obu przypadkach jest to decyzja arbitralna twórców korpusów podjęta ze względu na koszty i komplikacje związane z pozyskiwaniem danych mówionych (Przepiórkowski i in. 2012). Wiadomo, że w komunikacji międzyludzkiej liczba wypowiedzi i ich twórców ilościowo przeważa nad liczbą publikowanych tekstów i ich autorów, bo każdy użytkownik języka coś mówi, ale nie każdy publikuje. Ponadto publikacja w mediach tradycyjnych jest kosztowna i ryzykowna, ale wraz z rozwojem nowych mediów zwiększa się liczba udostępnianych nagrań dźwiękowych i filmowych przeciętnych użytkowników jezyka. Jeśli przyjąć, że korpusy językowe mają reprezentować język w w zakresie mowy, czyli *parole* (Saussure 2002) lub wykonania (Chomsky 1957), bo nie mogą odzwierciedlać całego mentalnego czy społecznego systemu językowego – *langue* czy kompetencji językowej (Chomsky 1957), to korpusy mówione i multimodalne są na pewno bliższe mowie niż korpusy tekstowe.

W rezultacie językoznawcy nie mogą prowadzić badań takich jak socjolodzy, którzy dobierają próby badawcze reprezentatywne względem całej populacji, na przykład pod względem wieku czy wykształcenia badanych, na podstawie danych publikowanych przez Główny Urząd Statystyczny. W projektach o dużym budżecie zdarza się, że uwzględnia się kryteria doboru grupy badanych wedle proporcji podawanych

przez GUS. Na przykład w projekcie SENTIMENTI uczestników badań dobierano wedle proporcji podanych przez GUS. Proporcje dotyczyły wieku, płci i wykształcenia oraz miejsca zamieszkania 22500 badanych w Polsce. Celem badań było zebranie ocen 30 000 znaczeń wyrazów dla ośmiu emocji podstawowych i dwóch skal: polaryzacji i pobudzenia (Wierzba i in. 2021).

W przypadku zachowań komunikacyjnych sytuacja jest złożona. Tradycja badań komunikacji sięga starożytności, kiedy filozofowie greccy oraz gramatycy indyjscy zastanawiali się nad formą mowy, która miała być przekonująca, poprawna i piękna. Pierwsze zachowane pisma na ten temat są preskryptywne, czyli mówiące innym użytkownikom, jak mają mówić czy pisać, a nie opisujące cech ich mowy (Heinz 1983). We współczesnych publikacjach o komunikacji spotykamy się z opisami zachowań komunikacyjnych, w których nie wspomina się o poprawności, ale zakłada się, że są to zachowania spontaniczne, naturalne, typowe dla badanej społeczności językowej i kultury (Efron 1972; Kendon 2004; McNeill 1992; Klima, Bellugi i Battison 1979; Barre 2013; Lausberg 2013; Antas 2013; Załazińska 2006; Jarmołowicz-Nowikow 2019).

Korpusy nagrań komunikacji multimodalnej przedstawiają zachowania ludzi w sposób pełniejszy niż opisy tekstowe czy szkice (Antas 2013). Pozostają jednak problemy: jakie zachowania wybrać, ile próbek zebrać lub nagrać, ile osób zaprosić do badań jako uczestników – rodzimych użytkowników języka. Wątpliwości budzi także stopień spontaniczności *versus* sztuczności zachowań komunikacyjnych. Wierzymy, że badania komunikacji u osób publicznych (np. Calbris 2011; Załazińska 2006; Antas 2013) przedstawiają sposób ich występowania możliwie jak najwierniej, lecz nie wiemy, w jakim stopniu ich występy są podobne do zachowań innych ludzi. Nagrania „niepublicznych" użytkowników języka pokazują zachowania *quasi*-spontaniczne, ale ze względu na sposób doboru nagrywanych wnioski dotyczące powszechności czy typowości zachowań komunikacyjnych także są ryzykowne (Cienki 2016; Karpiński i in. 2018; Jarmołowicz-Nowikow 2019). Przedstawiane w takich badaniach dane ilościowe na temat wystąpienia poszczególnych kategorii zachowań komunikacyjnych, na przykład gestów, należy wówczas traktować jako charakterystykę danego korpusu, a nie dane dotyczące populacji.

1.2.1. Introspekcyjne

Introspekcją nazywamy postępowanie badacza, który analizuje własne stany psychiczne, przeżycia i wypowiedzi werbalne (Miłkowski 2003). Metoda ta wywodzi się z filozofii umysłu i psychologii poznawczej, gdzie zyskała uznanie wśród badaczy procesów poznawczych i jest stosowana w szeroko pojętej kognitywistyce. Źródłem danych dotyczących komunikacji, języka czy gestów dla introspekcjonizmu są własne przykłady autora i zaobserwowane przez niego zachowania komunikacyjne innych osób, jednakże autor takich badań niekoniecznie podaje, w jaki sposób dokładnie zebrał analizowane przykłady, czyli w jakich okolicznościach je obserwował: zarejestrował, nagrał czy tylko zapamiętał i opisał, a być może przykłady spreparował na potrzeby tworzenia swojej teorii. Autorzy takich badań zakładają, że przykłady zaobserwowane, zasłyszane lub spreparowane są możliwymi i akceptowalnymi przez użytkowników danego języka zachowaniami. Dane w tych badaniach są przedstawiane za pomocą szczegółowego opisu, który ma czytelnikom wystarczyć do odtworzenia lub wyobrażenia albo zauważenia podobnych przykładów, tak jak wedle Miłkowskiego możliwe jest odtwarzanie „doznań przez wykonywanie odpowiednich instrukcji" (Miłkowski 2003: 128). W opisie przykładów znajdujemy terminy, które badacz szczegółowo definiuje i tezy, które stara się rzetelnie uzasadnić. Do introspekcyjnych badań komunikacji międzyludzkiej zaliczymy takie, w których badacz nie podaje przykładów udokumentowanych tak szczegółowo, jak w badaniach obserwacyjnych (korpusowych), czyli nie informuje o tym, kiedy, gdzie ani u kogo zaobserwował opisywane komunikaty, nie posługuje się danymi ilościowymi i choć nie podaje danych o częstości ich występowania, to uważa te komunikaty za typowe dla użytkowników danego języka, członków danej kultury lub ludzi jako gatunku. Badania takie nie wymagają nagrań, więc ich autorzy prezentują przykłady za pomocą rysunków lub schematów.

Cienki zwraca uwagę, że badania introspekcyjne komunikacji multimodalnej są narażone na uprzedzenia poznawcze i językowe. Jeśli autor badań nie podaje danych o pochodzeniu przykładu, to może się okazać, że przykłady zachowań komunikacyjnych opisane w publikacji naukowej wcale nie występują zbyt często albo w ogóle nie występują

w spontanicznej konwersacji użytkowników języka w takiej postaci, jak to opisał autor badań opartych na introspekcji (Cienki 2016). Uznajemy, że są to badania introspekcyjne, w których badacz zebrał przykłady zachowań i przeanalizował je oraz sklasyfikował wedle własnych intuicji albo wedle intuicji, które uważa za przyjęte w danej społeczności kulturowej, choć nie wiemy, jak je ustalił. Efektem takich badań są dokumentacje zachowań komunikacyjnych w wybranych kontekstach i gestuariusze, czyli klasyfikacje gestów wedle nazw, cech ruchu, znaczeń i miejsc występowania (Morris 1994).

1.2.2. Korpusowe

Współczesne badania gestów zapoczątkował prawdopodobnie Efron, kiedy opublikował obserwacje gestów u Włochów i Żydów zamieszkujących w Nowym Jorku (Efron 1941). Większość, jeśli nie wszystkie, analizy gestów w publikacjach wydanych po 1941 roku jest oparta na nagraniach i korpusach multimodalnych (np.: Kendon 2010, McNeill 1992, Chui 2022, Li 2014, Antas 2013, Załazińska 2001 i 2006, Kiełbawska 2012, Jarmołowicz-Nowikow 2019). Omówimy krótko kilkanaście wybranych publikacji książkowych dotyczących gestów, gdyż są to systematyczne, szczegółowe analizy zachowań komunikacyjnych przeprowadzane na dużych korpusach tworzonych przez samych badaczy i ich współpracowników. Przykłady gestów opisywane przez Kendona zostały wybrane z kilkudziesięciu nagrań rozmów ludzi podczas posiłków, grze w karty, spotkań na straganach i wycieczek z przewodnikami turystycznymi we Włoszech, Wielkiej Brytanii i Stanach Zjednoczonych (Kendon 2010). McNeill opiera się na kilkudziesięciu nagraniach dzieci i dorosłych, którzy opowiadali o bajce dla dzieci - Tweety i Sylvester (Canary Row 1950) (McNeill 1992). Nagrania Kendona oraz McNeilla i ich współpracowników zostały przeprowadzone głównie w latach 90-tych XX wieku, kiedy dostępne były jedynie kamery na taśmy w formatach VHS, 16 mm, 8 mm, Hi8 i Mini DV, a później nagrania zostały przetworzone na postać cyfrową. Korpusy tworzone od początku XXI wieku zawierają nagrania cyfrowe. Badania niewerbalnych sturktur dialogu i schematów wyobrażeniowych zostały przeprowadzone na korpusie, który składa

się z fragmentów wywiadów i debat politycznych w języku polskim, emitowanych w telewizji w latach 1999-2003 i mających w sumie ponad 20 godzin (Załazińska 2001 i 2006; Antas 2013). Próbki nagrań polskich zostały dołączone do publikacji w postaci płyty CD (Załazińska 2006) oraz plików osadzonych w ebooku (Antas 2013). Zespół niemieckich badaczek gestów zebrał fragmenty konwersacji, debat politycznych, przemówień w parlamencie niemieckim i teleturniejów w ramach projektu ToGoG (Towards a Grammar of Gesture: Evolution, Brain, and Linguistic Structures) w latach 2004-2010. Analizy form gestów i ich powtórzeń (Bressem 2021) oraz gestów jako jednostek gramatyki kognitywnej (Ladewig 2020) są oparte na odpowiednio 30 i 20 godzin nagrań z projektu ToGoG . Gesty w konwersacjach w języku chińskim zostały wybrane z korpusu nagrań zwanego NCCU Corpus of Spoken Taiwan Mandarin, gdzie NCCU oznacza National Chengchi University na Taiwanie. Analizie poddano około 15 godzin nagrań obrazu z tego korpusu, ale udostępniono jedynie nagrania dźwiękowe, transkrypty i rysunki wybranych kadrów (Chui 2022). Analizy gestów podczas zabierania głosu i zmiany kolejek w mandaryńskim chińskim są wykonany na podstawie 15 godzin nagrań rozmów wśród członków rodzin i przyjaciół (Li 2014). Podsumowując, opisane korpusy multimodalne mają od kilku do kilkudziesięciu godzin, co pozwala sądzić, że jest to optymalna wielkość korpusu, w którym badacze znajdują wystarczająco zróżnicowany materiał badawczy.

Do badań korpusowych zaliczamy takie, w których autorzy informują o sposobie zbierania danych, czyli czasie, miejscu i warunkach, w jakich pozyskali opisywane przykłady. W badaniach korpusowych wyróżniamy takie, w których próbki filmowe i dźwiękowe zostały nagrane na potrzeby danego badania, czyli tutaj nazywane wewnętrznymi, oraz takie, w których wykorzystane zostały nagrania wcześniej opublikowane – tutaj nazywane zewnętrznymi. W przypadku tych drugich badacz określa kryteria wyboru próbek i podaje ich źródło, tak jak to czyni Antas (2013) czy Załazińska (2006) oraz Ladewig (2020) i Bressem (2021). Nagrania zewnętrzne, tj. wcześniej opublikowane, badacze znajdują w archiwach telewizji i na platformach multimedialnych dostępnych przez Internet oraz w repozytoriach danych. W odróżnieniu od nagrań wewnętrznych zewnętrzne są przetworzone, zmontowane bez udziału badacza i udostępnione widzom w taki sposób, by były

atrakcyjne, informatywne lub rozrywkowe. Badacz nie ma więc kontroli nad tym, co znalazło się w kadrze kamery, co zostało pominięte lub wycięte ani jak nagranie zostało zaplanowane. Nie wiemy też, w jakim stopniu występujące na nagraniach osoby są doświadczone w pracy w mediach, czy są zawodowcami, czy naturszczykami, czy zachowują się naturalnie i spontanicznie. Nie wiemy, czy znajdowane nagrania są efektem pierwszego podejścia, czy występujący nagrywali się tyle razy i tak długo, aż osiągnęli oczekiwany efekt, tak jak tworzy się filmy fabularne z aktorami lub statystami odgrywającymi role wedle scenariusza i poleceń reżysera.

Źródłem nagrań zewnętrznych są wywiady (Antas 2013), debaty (Załazińska 2006), przemówienia (Calbris 2011), wystąpienia aktorów i artystów przed widownią (Lecoq 2006) czy przesłuchania, sesje psychoterapeutyczne (Sikorski 2005; Barre 2013), a także nagrania z udziałem tłumaczy (Kiełbawska 2012) lub nauczycieli (Goldin-Meadow 2013). Zaletą nagrań zewnętrznych jest niski koszt pozyskania danych, a wadą – zróżnicowanie jakości. Nagrania wewnętrzne są potrzebne przede wszystkim wtedy, jeśli nie znajdujemy wśród nagrań opublikowanych sytuacji komunikacyjnych, które chcemy zbadać. Plusem nagrań wewnętrznych jest kontrola nad jakością nagrań, lecz są dużo bardziej kosztowne niż zewnętrzne.

Przykładami korpusów wewnętrznych są opisane przez (Kendon 2010; Jarmołowicz-Nowikow 2019). Osobnego opracowania wymagają kwestie kryteriów doboru próbek, systemów anotacji i procedur ustalania zgodności anotacji nagrań w korpusach multimodalnych.

1.2.3. Eksperymentalne

Zarówno w przypadku badań eksperymentalnych, jak i korpusowych stosowana jest analiza ilościowa, to znaczy statystyka deskryptywna i inferencyjna, o ile dane są wystarczająco liczne i zróżnicowane. Eksperymenty służą testowaniu hipotez wyprowadzonych z teorii komunikacji i badań obserwacyjnych. Tworzenie korpusów multimodalnych jest więc przydatne w fazie projektowania eksperymentów. Typowe badania eksperymentalne polegają na zaaranżowaniu sytuacji komunikacyjnych różniących się co najmniej jedną cechą – opisywanej

za pomocą zmiennej niezależnej, którą badacz manipuluje, bo zakłada, że ma ona wpływ na badane zjawisko – mierzone zmienną zależną. W badaniach behawioralnych eksperymenty służą do przyjęcia lub odrzucenia hipotezy wskazującej zależność przyczynowo-skutkową pomiędzy zmiennymi opisującymi zjawiska (w tym zachowania komunikacyjne lub cechy uczestników komunikacji), które wybiera badacz. Hipoteza, którą badacz przyjmuje w badaniach eksperymentalnych jest nazywana zerową, gdyż jest przeciwieństwem hipotezy głównej, czyli proponowanego wyjaśnienia badanych zależności. W hipotezie zerowej przyjmuje się, że zmienna niezależna nie ma żadnego wpływu na zmienną zależną. O odrzuceniu hipotezy zerowej decyduje wynik testu istotności statystycznej, który mówi o prawdopodobieństwie z jakim otrzymane różnice w danych są większe niż wynikałoby to z przypadku (Shaughnessy i in. 2007). W ten sposób badania spełniają warunek falsyfikacjonizmu, czyli pokazują w jakich warunkach wybrane przewidywanie sformułowane przez badacza należy odrzucić, a w jakich uznać za prawdopodobne wyjaśnienie badanego problemu (Popper 1959). Uczestnicy badań eksperymentalnych i nagrań do badań korpusowych otrzymują instrukcje do wykonania zadania (Karpiński i in. 2018), oceniają zachowania komunikacyjne lub starają się je naśladować albo biorą udział w wywiadach czy dyskusjach na wybrane tematy bądź są proszeni o krótkie wypowiedzi przed kamerą. W niektórych badaniach stosuje się specjalne metody elicytacji zachowań komunikacyjnych. McNeill opisuje nagrania dzieci i dorosłych, którzy referują wcześniej obejrzaną bajkę dla dzieci – Tweety i Sylvester (Canary Row 1950), jako eksperymenty (McNeill 1992). Związki ruchów rąk i jednostek mowy są badane przy założeniu, że niektóre ruchy rąk współwystępują z niektórymi jednostkami lub cechami mowy (zobacz przegląd metod elicytacji w Holler 2013). Wyniki badań eksperymentalnych są porównywane w ramach jednego badania, na przykład między dwiema grupami badanych, albo w ramach jednego systemu anotacji stosowanego w wielu badaniach. Przykładem takich badań są studia nad związkami zaburzeń psychicznych i cech ruchów rąk, które są obserwowane u dwóch lub więcej grup uczestników badań, a następnie anotowane w systemie NEUROGES (Lausberg 2013; Lausberg i Sloetjes 2015).

1.2.4. Ankietowe

Obok badań introspekcyjnych badacz, zamiast polegać na własnej intuicji, zbiera opinie na temat zachowań komunikacyjnych wśród innych ludzi, czyli przeprowadza ankietę na wzór badań socjologicznych. Zebrane w ten sposób dane informują go o tym, jak członkowie wybranej wspólnoty językowej lub kulturowej czy zawodowej postrzegają swoje albo cudze komunikaty. Autor badań ankietowych posługuje się danymi ilościowymi o odpowiedziach uzyskanych od badanych, podaje także ich cechy demograficzne i kryteria wyboru (dobór próby), a wnioski z opisu odpowiedzi z ankiet stara się uogólnić na całą populację. Przykładem badań ankietowych są studia gestów emblematycznych (Szczepaniak 2017), w których uczestnicy badań oceniali pokazywane im gesty i wyjaśniali ich znaczenia.

2. Nagrania do korpusu multimodalnego

Zachowania na nagraniach powinny być zbliżone do naturalnych. Nagrań w celach naukowych nie przerywamy ani nie powtarzamy, jeśli w trakcie okaże się, że uczestnik chciał się inaczej zachować czy coś zmienić, odmiennie wykonać zadanie lub coś dodać. Dlatego – z wyjątkiem wycinania fragmentów do dalszych analiz – nagrań w celach naukowych nie modyfikujemy, nie montujemy. Nagrania nadawane publicznie w telewizji, jeśli nie są na żywo, to zawsze są zmontowane. Usuwane są niekompletne wypowiedzi, nieostre kadry, niefortunne zachowania.

W trakcie badań do celów naukowych należy także zadbać o to, by uczestnicy nie mieli na sobie żadnej biżuterii czy innych części garderoby, które mogłyby wydawać niepotrzebne dźwięki lub ograniczać ich ruchy. Nie powinni mieć także żadnych przedmiotów, które nie są związane z celem nagrań i badań, czyli na przykład naczyń z napojami, bo mogą je potrącić; narzędzi do pisania, w tym papieru, jeśli elementem nagrań nie jest pisanie czy rysowanie; powinni wyłączyć telefony komórkowe lub zostawić je poza salą nagrań, bo zakłócą urządzenia.

Badania psychologów i socjologów wykazały, że obecność obserwatora, prowadzącego nagrania i/lub samego badacza ma wpływ na przebieg nagrań, ich uczestników i wynik badań. Dlatego zalecane jest nagrywanie w dwóch pomieszczeniach albo podzielenie dużej sali, by odseparować uczestników i aparaturę oraz prowadzących nagrania i badania.

Przed przystąpieniem do nagrań należy odpowiedzieć na następujące pytania:

1. Jaki jest cel badań i nagrań? Na jakie pytanie badawcze mają pomóc odpowiedzieć?
2. Co nagrywamy? Dźwięk, bo do badań potrzebny nam tekst, czy także obraz, bo musimy analizować zachowania niewerbalne, czyli ruchy rąk, ekspresje emocji na twarzy albo ruchy innych części ciała; czy też planujemy mierzyć zachowanie czujnikami i badać inne cechy uczestników przed i/lub po nagraniach za pomocą kwestionariuszy?
3. Ile osób musimy nagrać?
4. Jakie warunki mają spełniać nagrywane osoby?
5. Ile minut nagrania każdej osoby potrzebujemy?
6. Ile razy musimy nagrać każdą osobę?

7. Ile osób nagrywamy jednocześnie?
8. W jakiej sali przeprowadzimy nagrania? Jakie są akustyka i oświetlenie sali?
9. Jak rozmieścimy w sali aparaturę i jak ją oddzielimy od uczestników?
10. W przypadku nagrań dialogów: kto i w jaki sposób będzie prowadził dialog? Jeśli prowadzący dialog ma zadawać pytania, to warto ustalić listę możliwych pytań przed nagraniami, by były takie same lub podobne dla wszystkich uczestników.
11. W przypadku nagrań rozmów, w których obie osoby miałyby mieć równy wkład: jak poinstruować uczestników, by zachowywali się swobodnie i spontanicznie? Możemy ich uprzedzić, że mają zachować się tak, jakby mieli razem do omówienia wybrany lub narzucony przez nas temat. Nie wspominamy o tym, by zachowywali się swobodnie i spontanicznie, bo to niektórych wprowadza w zakłopotanie.
12. W przypadku zadań: jakie polecenie otrzymają uczestnicy nagrań zadaniowych? Przed każdym nagraniem należy upewnić się, że uczestnicy zrozumieli polecenia.
13. W przypadku badań eksperymentalnych: w jakimi warunkami nagrania będą się różnić? Opis warunków przygotowujemy przed nagraniami.
14. Ile czasu zajmą nam archiwizacja, analiza i anotacja nagrań?
15. Jak zaprosimy uczestników do nagrań?
16. Jakich zgód na nagrania i badania potrzebujemy?

Nagrania kilkudziesięciu osób zajmują zwykle kilka miesięcy, a archiwizacja od kilku dni do kilku tygodni; następnie analiza i anotacja mogą trwać od kilku miesięcy do kilku lat, zależnie od celów. Nie sposób prowadzić nagrań w pojedynkę; nawet jeśli mamy wieloletnie doświadczenie w badaniach i orientację w technologii, trudno zapanować jednocześnie nad ustawieniem aparatury, umówieniem uczestników i archiwizacją nagrań. Dlatego nagrania do celów badawczych są zwykle wykonywane przez zespoły badawcze liczące co najmniej trzy osoby, z czego co najmniej jedna sprawdza, czy nagrania są zgodne z celami badań, jedna jest zorientowana technicznie i jedna odpowiada za umawianie uczestników i uzyskiwanie zgód. Warto także prowadzić protokół nagrań, w którym zaznaczamy używane w danym dniu i nagraniu urządzenia, wpisujemy kody uczestników i upewniamy się, że nagranie zostało zarchiwizowane.

Przed nagraniami głównymi warto przeprowadzić nagrania próbne, w których sprawdzimy:

1. sprzęt do nagrań,
2. role i rutyny prowadzących nagrania,
3. role i zadania uczestników nagrań.

Jeśli chodzi o role i rutyny prowadzących, to są to czynności, które wykonują w trakcie nagrań, takie jak:

1. ustawienie sprzętu,
2. podłączenie sprzętu do prądu,
3. wprowadzenie uczestników do sali,
4. włączenie sprzętu,
5. nagranie,
6. wyłączenie sprzętu,
7. wyprowadzenie uczestników z sali nagrań,
8. skopiowanie nagrania na dysk.

2.1. Dane uczestników

Zbieramy tylko potrzebne dane – na przykład imię i nazwisko oraz telefon kontaktowy są konieczne na etapie znajdowania uczestników, ale w trakcie nagrań i analiz posługujemy się jedynie numerami lub kodami uczestników. Wszystkie inne dane pomijamy, bo na przykład nie potrzebujemy adresu zamieszkania, numeru dowodu osobistego ani numeru PESEL. Uprzedzamy badanych, że tych danych nie zbieramy, dzięki czemu unikamy podejrzeń o wyłudzanie danych chronionych i poufnych czy wrażliwych (w tym takich jak pochodzenie). Wyjątkiem są dane niezbędne do realizacji określonego celu badań – na przykład płeć, znajomość języków obcych, wykształcenie, zawód – i dane zbierane za pomocą kwestionariuszy.

2.1.1. Informacje o nagraniach i badaniach dla uczestników

Uprzedzanie uczestników o celu nagrań i badań ma wpływ na ich wynik, dlatego zaproszenie do nagrań nie może ujawniać rzeczywistego, ostatecznego celu nagrań ani badań. Podajemy cel pośredni,

na przykład zadanie do wykonania albo temat do rozmowy, lecz nie sugerujemy uczestnikom, jakie mają wykonać zadanie ani jak mają mówić, rozmawiać czy poruszać rękoma. Nie zdradzamy, jak poradzili sobie inni badani. Uprzedzamy, że uczestnik bierze udział w badaniu dobrowolnie i może się z niego w każdej chwili wycofać bez konieczności podawania przyczyny. Ewentualnie możemy wyjaśnić rzeczywisty cel badania po nagraniu (ang. *debriefing*), ale musimy także poprosić, by uczestnik nie ujawniał go innym osobom ani nie opowiadał innym o przebiegu badania. Przykład zaproszenia do badań pokazuje Tabela 1.

2.1.2. Zgoda na udostępnienie danych osobowych i wizerunku

Badani podpisują zgodę na udział w badaniach oraz udostępnienie swoich danych i wizerunku w celach badawczych innym członkom zespołu badawczego po anonimizacji. Zgoda musi uwzględniać RODO (ogólne rozporządzenie o ochronie danych[6]), czyli zawierać informacje o tym, kto, w jaki sposób i jak długo będzie przechowywał dane uczestników.

2.1.3. Zgoda komisji etycznej

Organizacje finansujące badania oraz niektóre czasopisma naukowe wymagają opisu warunków przeprowadzania badań z udziałem ludzi i zgody komisji etycznej. Zgodę na przeprowadzenie badań z udziałem ludzi powinna wydać komisja etyczna ds. badań naukowych z udziałem ludzi. Dotyczy to zwłaszcza badań eksperymentalnych z wykorzystaniem kwestionariuszy. Wniosek o zgodę składamy do komisji na uczelni, na której realizowane są badania.

---

[6] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych).

Tabela 1. Przykładowe zaproszenie do badań.

Droga Studentko, Drogi Studencie!

Zapraszamy Cię do wzięcia udziału w programie „Coaching dla Studenta, Absolwenta". Masz 25-35 lat, jesteś absolwentem/absolwentką lub niebawem kończysz studia? Zapraszamy do wzięcia udziału w sesji coachingowej, która pomoże Ci znaleźć własną ścieżkę kariery zawodowej!

Coaching to proces, który dzięki interakcji coacha i klienta pomaga klientowi w ustaleniu celów, optymalizacji środków ich osiągania oraz lepszym wykorzystaniu naturalnych umiejętności.

W projekcie „Coaching dla studenta i absolwenta" proponujemy Ci bezpłatną sesję coachingową dotyczącą Twojego rozwoju zawodowego i planowanej kariery. Jako absolwent, student ostatniego roku czy doktorant z pewnością masz wątpliwości, **jak potoczy się Twoja kariera zawodowa**. Coach może pomóc Ci w podjęciu decyzji, ułatwić określenie hierarchii celów i podnieść Twoją samoświadomość.

Jak wziąć udział w projekcie? Wystarczy wypełnić formularz na stronie: … (zakładka Zapisy).

Dlaczego udział w projekcie jest bezpłatny?

Udział w projekcie „Coaching dla studenta i absolwenta" jest darmowy dla wszystkich uczestników, bowiem jest częścią badań naukowych finansowanych przez Narodowe Centrum Nauki.

Kto może wziąć udział w projekcie?

Każdy urodzony między 1978 a 1988 rokiem, kto posiada dyplom studiów licencjackich i/lub jest studentem studiów uzupełniających magisterskich i/lub doktoranckich. W przypadku jednolitych studiów magisterskich pięcioletnich (prawo, psychologia, malarstwo) zapraszamy studentów wyłącznie ostatniego roku. Zapraszamy studentów i absolwentów wszystkich uczelni wyższych.

Z powodów formalnych w projekcie nie mogą wziąć udziału studenci ani absolwenci następujących wydziałów Uniwersytetu im. Adama Mickiewicza w Poznaniu: Wydziału Nauk Społecznych, Wydziału Anglistyki, Filologii Polskiej i Klasycznej oraz Wydziału Neofilologii.

Pytania, wątpliwości? Napisz lub zadzwoń: … tel. …

Do zobaczenia wkrótce!

Zespół CDA

2.2. Sala nagrań

Wybór odpowiedniego miejsca nagrywania dźwięku i obrazu jest istotny ze względu na zakłócenia, jakie występują w otoczeniu. Istotne, by sala nagrań znajdowała się daleko od głośnych źródeł dźwięku, takich jak ulica pełna samochodów lub sala czy korytarz wypełniony ludźmi. Izolację od dźwięku zapewniają grube mury budynku, w którym mieści się sala wybrana do nagrań i szczelne okna oraz maty wygłuszające zamontowane na ścianach wewnątrz sali. W trakcie nagrań w sali nie powinny znajdować się żadne inne urządzenia wydające dźwięk, czyli tykające zegary, telefony czy głośniki. Wielkość i kształt pomieszczenia mają znaczenie dla rozchodzenia się głosu, dlatego optymalne są duże sale wykładowe, wysokie na co najmniej 3 m. Mniejsze wywołują większy pogłos, bo dźwięk odbija się od ścian, dlatego zaleca się montowanie w nich mat tłumiących. Dostęp do sali powinny mieć tylko osoby zaangażowane w nagrania.

Dla nagrań obrazu istotne jest oświetlenie. Naturalne źródła światła są odpowiednie, ale zbyt intensywne. Zmienne oświetlenie powoduje zbyt jaskrawe barwy oraz cienie przechodzące przez salę, w ciągu dnia. Dlatego lepszym źródłem światła jest lampa o barwie zbliżonej do słonecznej (ok. 4000 K), która świeci bez przerw i nie przemieszcza się, więc nie wywołuje nagłych zmian natężenia światła ani ruchomych cieni. Jeśli nagrywamy w pomieszczeniu bez okien, to należy pamiętać, że oświetlenie montowane w miejscach do pracy ma barwę około 6000 K, czyli jest zimne i niebieskawe. Wówczas trzeba pamiętać o korekcie barwy w ustawieniach kamery. Nagrania za pomocą kamer czułych na fale podczerwieni mogą być realizowane w ciemności lub zaciemnieniu, z minimalnym oświetleniem, o ile nagrywani nie potrzebują więcej światła, by się widzieć.

Dodatkowo istotne jest tło nagrań, czyli kolor ścian za nagrywanymi osobami. Ściany sal są zwykle jasne, białe, beżowe lub żółte, a lepsze nagrania otrzymamy na tle w kolorze zielonym, który kontrastuje z kolorem skóry i ubrań. Nagrania na zielonym tle są łatwiejsze dla analiz obrazu i ruchu, gdyż algorytmy rozpoznawania obrazu i ruchu działają najlepiej na obrazach z wysokim kontrastem i rozpoznają obiekty, które wyróżniają się barwą na tle innych obiektów. W celu zapewnienia zielonego tła wieszamy na statywach dużą zieloną zasłonę (*green screen*)

za nagrywanymi osobami, odpowiednio daleko od nich (20-30 cm), by im nie przeszkadzała.

2.3. Nagrywanie

Do rejestracji dźwięku, obrazu i ruchu służą mikrofony i kamery oraz czujniki ruchu. Przed każdym nagraniem sprawdzamy w każdej nagrywarce, kamerze, czujniku i komputerze stan kart pamięci i wewnętrzne zegary oraz inne ustawienia. Sprawdzamy także poziomy czułości mikrofonów w nagrywarce i ustawienia obiektywu w kamerze (ostrość na nagrywane osoby i obiekty oraz odpowiednia ogniskowa i kadr), a także naświetlenie (niektóre kamery dostosowują parametry nagrania do rozpoznanego oświetlenia, a inne należy skonfigurować ręcznie). Ustawienia zapisujemy lub fotografujemy, by móc je przywrócić po awarii (np. po nagłym braku prądu). Każde nagranie powinno być wykonane z takimi samymi ustawieniami wszystkich urządzeń.

2.3.1. Aparatura do nagrań

Aparatura potrzebna do nagrań składa się z kilku urządzeń w wersji minimalnej (1-5) lub maksymalnej (wszystkie wymienione niżej):

1. kamera,
2. mikrofon w kamerze,
3. statyw do kamery,
4. zasilacz do kamery (wtyk),
5. przewód łączący zasilacz kamery i gniazdko elektryczne (ósemka),
6. mikrofon lub dwa,
7. nagrywarka, do której jest podłączony mikrofon lub dwa,
8. przewody łączące mikrofon (mikrofony) i nagrywarkę (XLR),
9. przewód łączący zasilacz nagrywarki i gniazdko elektryczne (ósemka),
10. statyw do mikrofonu,
11. przedłużacz z wieloma gniazdkami.

Dodatkowo przydatne są:

12. taśma naprawcza do przyklejenia przewodów na podłodze i oznaczenia miejsca dla uczestników,
13. nalepki na urządzeniach, zasilaczach i przewodach,
14. rzepy lub trytytki,
15. lista urządzeń,
16. informacja na drzwiach sali o nagraniach (UWAGA NAGRANIA, PROSIMY O CISZĘ).

Koszt aparatury do nagrań jest zależny od wybranych urządzeń, a tych na rynku jest po kilkaset modeli z każdego typu, więc unikamy podawania tutaj konkretnych cen. Można natomiast określić koszt aparatury na równy kwocie przeciętnego wynagrodzenia w Polsce lub wielokrotnie wyższy.

2.3.2. Mikrofony

Mikrofon jest urządzeniem, które rejestruje dźwięk, czyli przekształca energię akustyczną na elektryczną. Przy wyborze mikrofonu do nagrań głosu zwracamy uwagę na:

- przetwornik,
- pasmo przenoszenia,
- kierunkowość,
- złącze.

Osoby pracujące głosem, czyli lektorzy, dziennikarze, śpiewacy czy wokaliści, używają mikrofonów z przetwornikiem pojemnościowym, z szerokim pasmem przenoszenia, jednokierunkowych i przewodowych. Pojemnościowe przetworniki są lepsze od dynamicznych, bo mają większą czułość i szersze pasmo przenoszenia dźwięków. Pasmo przenoszenia dźwięku powinno być nieco szersze od pasma fal akustycznych odbieranych przez ludzkie ucho, które obejmuje częstotliwości od 16 Hz do 20 kHz u młodych ludzi. Z wiekiem wrażliwość na tony wysokie się zmniejsza. Im szersze pasmo przenoszenia dźwięku mikrofonu, tym pełniejszy, bardziej wyraźny zapis mowy. Mikrofony kierunkowe są skonstruowane tak, by rejestrować dźwięk z jednego źródła, na przykład jednego mówcy. Do rejestracji mowy dwóch osób zaleca się stosowanie dwóch mikrofonów, z których każdy

podłączony jest do innego kanału, na przykład lewego i prawego. Taki sposób nagrywania zapewnia czystość dźwięku i rozdzielenie głosów mówców, a to z kolei ułatwia transkrypcję i analizę nagrań. Mikrofony przewodowe są mniej zawodne od bezprzewodowych, choć te drugie są bardziej praktyczne. Złącze analogowe (XLR, jack lub minijack) lub cyfrowe (USB) zapewnia ciągłe, pozbawione zakłóceń nagranie, natomiast bezprzewodowe (radiowe, np. Bluetooth) bywa łatwo zakłócane i zrywane na skutek działania innych urządzeń radiowych takich jak telefony komórkowe. Dlatego w sali nagrań nie powinny się znajdować żadne inne urządzenia niż te, które służą nagraniom. Zarówno telefony komórkowe, jak i niektóre rodzaje świateł mają negatywny wpływ na rejestrację dźwięku, obrazu czy ruchu.

Mikrofon może zostać podłączony do urządzenia zapisującego dźwięk – nagrywarki, kamery, telefonu lub komputera. Zalecane jest podłączanie mikrofonu do nagrywarki, która działa niezależnie od kamery czy komputera i umożliwia oddzielne nagrywanie dźwięku i obrazu. Mikrofon umieszczamy w koszyku, by zminimalizować wpływ wibracji, a koszyk przykręcamy do statywu.

2.3.3. Nagrywarki dźwięku

Mikrofony muszą być podłączone do nagrywarki, komputera lub kamery. Najlepszą opcją jest nagrywarka, czyli urządzenie, do którego możemy podłączyć dwa lub więcej źródeł dźwięku (XLR, jack lub minijack) i zapisać dźwięk na nośniku pamięci. Dobre nagrywarki umożliwiają zapis w wielu formatach i kanałach, a także mają wyjście na słuchawki, którymi sprawdzamy jakość nagrania.

Jeśli nie mamy nagrywarki, to możemy mikrofony podłączyć do komputera, ale w tym przypadku należy zwrócić uwagę na jego wentylator. Szum wentylatorów komputera –zarówno stacjonarnego, jak i przenośnego – będzie słyszalny na nagraniu, a gwałtowne zmiany napięcia w komputerze mogą mieć negatywny wpływ na nagranie oraz działanie nagrywarek i innych czujników. Należy także zadbać o to, by komputer miał wyłączone funkcje samodzielnego usypiania i wyłączania się oraz by w tle nie działały inne aplikacje, zwłaszcza takie, które wydają dźwięki i powiadomienia.

Mikrofony podłączone do kamery pozwalają na rejestrację dźwięku bliżej mówców niż te wbudowane w kamerę. Jednakże dźwięk i obraz z jednej kamery zostanie zapisany w jednym pliku, co gwarantuje synchronizację dźwięku i obrazu. Jednocześnie, zależnie od modelu kamery, taki zapis może obniżać jakość dźwięku i obrazu..

2.3.4. Kamery

Kamery to urządzenia, które – podobnie jak aparaty fotograficzne – przekształcają fale świetlne na obrazy utrwalane na kliszy lub innym nośniku. Obrazy są składane w serie i przy odpowiedniej prędkości ich wyświetlania uzyskujemy film. Filmy nagrywane na kliszy mają wysoką jakość i mimo rozwoju cyfrowych technik rejestracji obrazu kamery analogowe są nadal stosowane w kinematografii. Pierwsze nagrania komunikacji multimodalnej w latach 70. i 80. XX wieku były realizowane za pomocą kamer analogowych (McNeill 1992; Kendon 1990) i analogowych aparatów fotograficznych (Klima, Bellugi i Battison 1979). Wysokiej jakości kamery analogowe są jednak kosztowne i trudne w obsłudze. Przed analizą nagranie analogowe musi zostać przekształcone (przechwycone) na postać cyfrową, co dodatkowo wydłuża proces archiwizacji. W kamerach analogowych obraz w trakcie nagrań jest widoczny tylko przez wizjer albo na dodatkowym ekranie. Dlatego w nagraniach do celów badawczych stosuje się kamery cyfrowe, które rejestrują obraz na kartach pamięci SD lub dysku SSD. Przy wyborze kamery zwracamy uwagę na:

- przetwornik,
- format obrazu i rozdzielczość,
- obiektyw,
- złącza i zapis.

Przetwornik w kamerze odpowiada za jakość obrazu i standardem jest CMOS lub MOS, czyli Complementary MOS (*metal oxide semiconductor*). Im większa matryca przetwornika, tym więcej światła kamera może rejestrować; zalecana wielkość to przekątna 1'', czyli 25 mm.

Typowy format obrazu to 4:3 (telewizyjny) lub 16:9 (kinowy), a zalecana rozdzielczość to co najmniej 1920 × 1080, czyli tzw. full HD. Im wyższa rozdzielczość, tym bardziej szczegółowy, ostry obraz i większa

objętość pliku. Dodatkowym parametrem formatu nagrań jest liczba klatek na sekundę; standardem jest liczba 24, 25 lub 30 klatek. Wyższa wartość pozwala na dokładniejszy zapis, który nawet po zmniejszeniu prędkości odtwarzania da wyraźny obraz, ale taki zapis wymaga z kolei więcej miejsca na nośniku pamięci oraz więcej mocy przetwornika. Dostępne są także tryby nagrań *slow-motion*, w których kamery rejestrują 60, 120 i 240 albo więcej klatek na sekundę, lecz nagrania takie wymagają wielokrotnie więcej pamięci i większej mocy procesora.

Obiektyw ma wpływ na ilość światła i szerokość kadru, a także na jakość obrazu (obraz w kamerze jest dwuwymiarowym modelem przestrzeni trójwymiarowej). Do nagrywania osób i obiektów z odległości 2 m obiektyw w kamerze powinien mieć wielkość przesłony f/2,8'' i wielkość ogniskowej 28, 35 lub 50 mm. Im większa przesłona, tym więcej światła wpada na matrycę i płytsza jest ostrość obrazu, a im większa ogniskowa, tym węższy kąt widzenia i większe przybliżenie osoby i obiektów. Obiektywy o większej ogniskowej (tzw. *zoom* albo *tele*) mają mniejsze przesłony, co obniża jakość nagrań, więc zaleca się stosowanie obiektywów szerokokątnych, czyli o około 75° kąta widzenia. Jeśli obiektyw ma zoom optyczny lub cyfrowy, to zaleca się stosowanie tylko optycznego albo bliższe ustawienie kamery, by zminimalizować zniekształcenia obrazu. Zoom cyfrowy można uzyskać przez zmniejszenie kadru po nagraniu.

Kamery cyfrowe mają własne ekrany do podglądu nagrania na żywo, ale niektóre mają także wyjście HDMI na zewnętrzny monitor lub komputer. Zapis w kamerach cyfrowych jest realizowany na kartach pamięci typu SD lub dyskach SSD. Niektóre kamery mają możliwość rejestracji filmu na dwóch nośnikach pamięci jednocześnie, by zminimalizować ryzyko utracenia nagrania w wyniku awarii jednego z nośników.

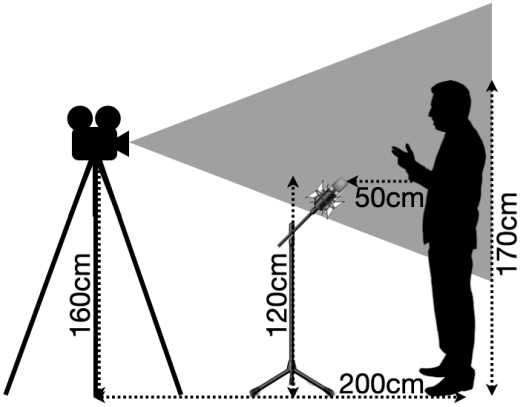
2.3.5. Statywy

Obraz w kadrze musi być stabilny, a mikrofony nie powinny się stykać z jakimkolwiek obiektem. Dlatego kamery i mikrofony montujemy na statywach, które ustawiamy mniej więcej na wysokości metra, a dokładniej – kamerę na wysokości 120 cm, a mikrofon niżej, by nie był widoczny w kamerze, na przykład na wysokości 100 cm. Kamery

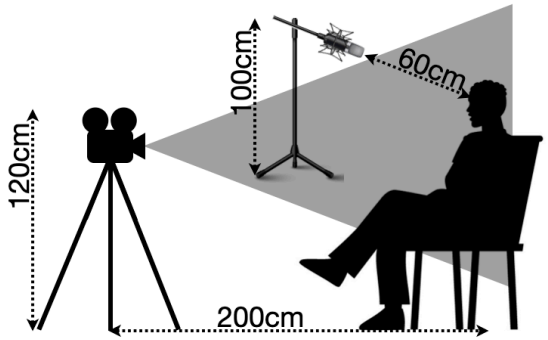
ustawiamy w odległości nie mniejszej niż 2 m od nagrywanych osób na wprost lub z boku albo po przekątnej (45° względem środkowej linii ciała, czyli linii przebiegającej przez czubek nosa i pępek), a mikrofony powinny być bliżej, więc umieszczamy je nie dalej niż metr od mówców. Jeśli nagrywamy większą liczbą kamer, to ustawiamy je tak, by się nawzajem nie widziały, czyli by w kadrze jednej nie była widoczna druga i odwrotnie. W kadrze kamery powinna się mieścić jedna osoba stojąca lub siedząca z rozsuniętymi na boki rękoma, czyli w pozycji litery T (*T pose*), od pasa w górę, do około 30 cm nad głową. Przykładowe ustawienia pokazano poniżej:

1. Grafika 1. Ustawienie kamery i mikrofonu z osobą stojącą.
2. Grafika 2. Ustawienie kamery i mikrofonu z osobą siedzącą.
3. Grafika 3. Ustawienie kamery i dwóch mikrofonów z dwiema osobami siedzącymi.
4. Grafika 4. Ustawienie dwóch kamer i dwóch mikrofonów z dwiema osobami siedzącymi.
5. Grafika 5. Ustawienie trzech kamer i dwóch mikrofonów z dwiema osobami siedzącymi.

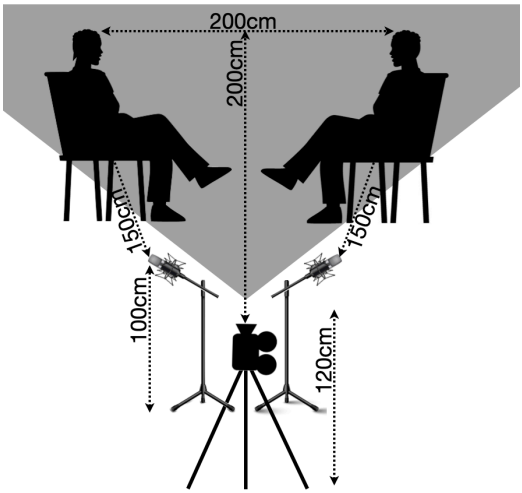
Dokładne odległości między nagrywanymi osobami oraz między statywami z kamerami i mikrofonami są zależne od kąta widzenia w kadrze. Należy pamiętać, by w kadrze kamery znajdowała się osoba nagrywana co najmniej od pasa w górę, a nie były widoczne mikrofony.



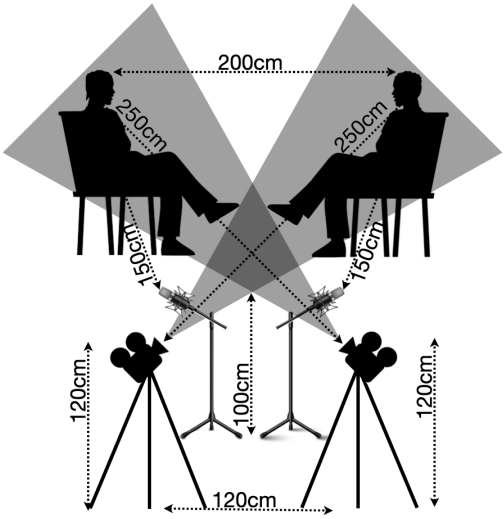
Grafika 1. Ustawienie kamery i mikrofonu z osobą stojącą.



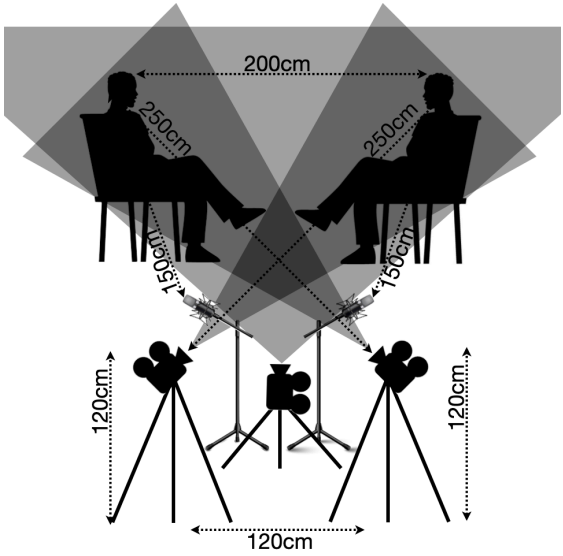
Grafika 2. Ustawienie kamery i mikrofonu z osobą siedzącą.



Grafika 3. Ustawienie kamery i dwóch mikrofonów z dwiema osobami siedzącymi.



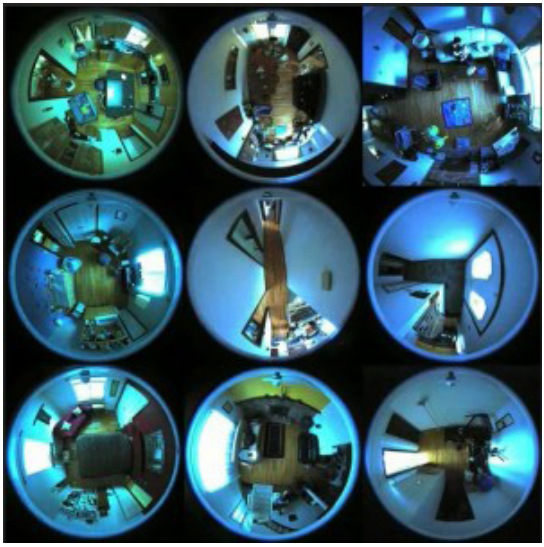
Grafika 4. Ustawienie dwóch kamer i dwóch mikrofonów z dwiema osobami siedzącymi.



Grafika 5. Ustawienie trzech kamer i dwóch mikrofonów z dwiema osobami siedzącymi.

Zwykle w celach badawczych nagrywa się kamerami, które w trakcie nagrań są nieruchome. Jeśli jednak potrzebne jest nagranie ruchu całych osób (a nie tylko rąk) czy obiektów, to możemy obracać kamerą po linii poziomej kadru, czyli horyzontalnie (*panning*). Dostępne są także statywy-stabilizatory, które umożliwiają utrzymanie stabilnego kadru przy jednoczesnym poruszaniu kamerą w powietrzu. Pozwala to na śledzenie osób lub obiektów, ale wymaga dodatkowej osoby, która trzyma kamerę na stabilizatorze trójosiowym, tzw. gimbalu.

Jeśli z kolei celem nagrań jest rejestracja ruchu na dużej przestrzeni, na przykład na całej sali, a dokładna rejestracja ruchów rąk nie jest potrzebna, to możemy zamontować kamerę na suficie i objąć całą salę z góry za pomocą szerokiego obiektywu. Takie rozwiązanie zostało zastosowane na przykład w badaniach Deba Roya (Grafika 6). Miały one na celu śledzenie ruchu domowników (Roya, jego żony, syna i opiekunki syna) i interakcji między nimi w nabywaniu języka syna badacza. W mieszkaniu umieszczono 9 kamer i nagrano 90 000 godzin interakcji (Roy 2009).



Grafika 6. Kadr zwany rybim okiem, kamera umieszczona na suficie pomieszczeń w mieszkaniu. Kadr pochodzi z nagrań Deba Roya.

2.3.6. Przewody

Każde urządzenie powinno być podłączone do źródła energii w gniazdku, bo wbudowane w nagrywarki czy kamery baterie mają różną pojemność i trudno zadbać o to, by zawsze były gotowe do pracy. Dodatkowym zabezpieczeniem zasilania są przedłużacze z bezpiecznikami i dodatkowymi bateriami, które umożliwiają pracę urządzeń na wypadek przerw w dostawie prądu. Nie należy podłączać zbyt wiele urządzeń do jednego przedłużacza, bo zwykle maksymalne obciążenie dla jednego przedłużacza to 2000 W. Przewody powinny być poprowadzone po podłodze i przyklejone do niej mocną taśmą (np. naprawczą) albo przykryte wykładziną, by nie były widoczne. O wiszące przewody łatwo się potknąć i przewrócić, co może się okazać niebezpieczne dla uczestników nagrań i kosztowne, jeśli dodatkowo przewrócimy statyw z urządzeniem.

2.3.7. Złącza

Warto znać powszechnie stosowane złącza do przesyłania dźwięku, obrazu i danych. Szczegółowe dane dotyczące protokołów transmisji danych pomijamy, ale istotne jest, by pamiętać, że:

- nie powinno się podłączać do urządzeń innych zasilaczy niż dostarczone przez producenta, bo można je spalić;
- nie powinno się łączyć przewodów w celu ich przedłużenia, bo można je przeciążyć;
- wszystkie przewody powinny być na podłodze, najlepiej przyklejone, by nikt się o nie nie przewrócił;
- nie wolno dotykać uszkodzonych przewodów podłączonych do sieci, by nie porazić się prądem;
- złącza wtykamy tam, gdzie pasują, więc trudno się pomylić i podłączyć USB do HDMI czy jacka do minijacka, ale warto je rozróżniać, by wiedzieć, jakie są potrzebne do konkretnej aparatury.

Choć każdy z przewodów jest oferowany w różnych długościach i dostępne są złączki pozwalające na ich przedłużenie, to im dłuższy przewód, tym mniejsza przepustowość łącza. Dlatego minimalne długości przewodów dostępnych w sprzedaży mieszczą się w przedziale 1-2 m. Dłuższe muszą być grubsze, a także ekranowane i zasilane

niezależnym źródłem prądu. Dlatego 5-metrowe i dłuższe przewody HDMI są bardzo drogie, ale zapewniają stabilny przekaz.

Dobrą praktyką jest oznaczanie przewodów, urządzeń, statywów i zasilaczy naklejkami, by w prosty sposób je odróżniać przy podłączaniu. Dodatkowo przydatne są rzepy lub plastikowe opaski zaciskowe (trytytki) do zaciskania zwiniętych przewodów i taśma naprawcza do przyklejania przewodów do podłogi.

Tabela 2. Oznaczenia i kształty wtyczek do przesyłania dźwięku, obrazu, danych i zasilania.

XLR		RCA	
JACK		VGA	
DVI		DISPLAYPORT	
HDMI		ETHERNET	
USB-C		USB	
SIECIOWE TYP C		WTYK DC	
C7 lub ÓSEMKA		C5 lub MYSZKA MIKI	

2.3.7.1. Złącza do przesyłania dźwięku

- XLR: *External Line Return* to złącze stosowane pomiędzy mikrofonami a nagrywarkami.
- RCA lub CHINCH: *Radio Corporation of America* to łącze analogowe do przesyłania dźwięku lub obrazu (trzy lub więcej złączy RCA). Popularne do łączenia głośników, wzmacniaczy i odtwarzaczy CD. Cyfrowym złączem do przesyłania dźwięku jest przewód optyczny.
- JACK (6,3 mm) lub MINIJACK (3,5 mm) to także łącza analogowe do przesyłania dźwięku, najczęściej stosowane do łączenia słuchawek i niektórych mikrofonów.

2.3.7.2. Złącza do przesyłania obrazu

Złącza do przesyłania obrazu łączą komputer (ewentualnie kamery) i monitor lub projektor. VGA i DVI to starsze rodzaje złączy, które nie występują w urządzeniach produkowanych po roku 2010, kiedy dominującymi stały się HDMI i Ethernet.

- VGA: *Video Graphics Array* to złącze analogowe, stąd mała rozdzielczość obrazu (2048 × 1536) i brak przenoszenia dźwięku.
- DVI: *Digital Visual Interface* to jedno z pierwszych złączy jednocześnie analogowych i cyfrowych, obsługuje większe niż VGA rozdzielczości (2560x1600@60Hz i 3840x2400@33Hz) oraz dźwięk.
- DISPLAYPORT i MINIDISPLAYPORT pełnią takie same funkcje jak złącza HDMI.
- HDMI: *High Definition Multimedia Interface* to najbardziej rozpowszechnione i uniwersalne złącze do przesyłania obrazu i dźwięku w wysokich rozdzielczościach (do 10240x4320@120 w wersji 2.1). HDMI występuje także w postaci mini- i micro-HDMI w komputerach stacjonarnych i przenośnych, w monitorach i projektorach oraz w niektórych kamerach, gdzie służy do podglądu.
- ETHERNET: złącze używane do przesyłania danych przez sieć LAN lub Internet; jest stosowane do łączenia kamer, monitorów, komputerów i projektorów, bo ma zbliżone do HDMI możliwości. Ze względu na formę jest nazywane skrętką.

2.3.7.3. Złącza do przesyłania danych

Złącza używane do przekazywania danych po nagraniu z kamer lub nagrywarek:

- USB-C: standard wprowadzony w 2014 roku, jest aktualnie najszybszym złączem do przesyłania danych, obrazu lub dźwięku jednocześnie w wysokich rozdzielczościach zarówno pomiędzy kamerą i komputerem, jak i zewnętrznym dyskiem czy monitorem.
- USB: *Universal Serial Bus* to dominujące złącze dla wszelkich urządzeń multimedialnych i innych podłączanych do komputera lub kamery czy nagrywarki i niektórych mikrofonów. Powszechne w wersji 2.0 ma ograniczoną do 480 Mbps przepustowość, co wystarczy do przesyłania danych pomiędzy kamerą a komputerem po nagraniu, a także do podglądu.

2.3.7.4. Złącza do przesyłania prądu

Złącza sieciowe służą do podłączenia urządzenia do gniazdka elektrycznego.

- SIECIOWE typu C lub E, czyli dwa bolce (C) z uziemieniem (E): nie sprawia kłopotów, o ile pamiętamy, że zwykle gniazdka mają moc ograniczoną do 2000 W lub 3600 W, a jej przekroczenie skutkuje wyłączeniem całej instalacji w sali czy budynku albo spaleniem (!) jednego z urządzeń.
- WTYK DC (prądu stałego): przewody łączące zasilacze i kamery lub komputery przenośne mają wtyki o różnej średnicy, bo służą do dostarczania prądu stałego o odmiennych parametrach. Bardzo istotne jest, by nie podłączać przewodu i zasilacza do urządzenia, które nie było dostarczone przez producenta, bo można spalić zasilacz lub urządzenie (!) i naprawa w ramach gwarancji nie będzie możliwa. Istnieją zamienniki zasilaczy różnych firm, ale należy zawsze dokładnie sprawdzić parametry wymagane przez producenta urządzenia, do którego podłączamy inny zasilacz.
- ÓSEMKA: wtyk zwany ze względu na swój kształt ósemką jest używany w niektórych zasilaczach do kamer i komputerów.

- MYSZKA MIKI: wtyk wygląda jak twarz Myszki Miki, ale jest mniej popularny w Polsce.

2.3.8. Synchronizacja nagrań

Jeśli w trakcie nagrania korzystamy z więcej niż jednego rejestratora, konieczna jest synchronizacja wszystkich urządzeń. W tym celu synchronizujemy wewnętrzne zegary nagrywarek, kamer i komputerów, włączamy wszystkie urządzenia jednocześnie albo w jak najkrótszych odstępach czasu i dajemy sygnał startowy po ich włączeniu: klaps, klaśnięcie lub puknięcie. Pliki zapisywane są wraz z informacją o czasie utworzenia i zmian, czyli zakończenia zapisu. Dlatego mając pliki z kilku urządzeń, które są synchronizowane, możemy łatwo przyciąć nagrania przed dalszą analizą. Odcinamy fragment od początku nagrania do sygnału startowego i fragment od sygnału końcowego do końca każdego nagrania. Sygnał końcowy może być taki sam jak startowy. Niektóre programy do przetwarzania dźwięku i obrazu mają funkcję automatycznej synchronizacji, która polega na rozpoznaniu sygnału startowego i odcięciu fragmentów początkowych i końcowych. Synchronizacja automatyczna pozwala na zaoszczędzenie wielu godzin pracy przed dalszą analizą, a odcięte fragmenty zawierają zakłócenia lub zjawiska, które nie powinny być dalej udostępniane. W celu jednoczesnego włączania nagrywania wielu urządzeń można zastosować piloty, bo niektóre modele kamer można obsługiwać tym samym pilotem, zwłaszcza jeśli jest to ten sam producent i model. Niektóre urządzenia można złączyć w jeden system i uruchamiać z komputera.

2.3.9. Nośniki pamięci

Analogowe nośniki dźwięku i obrazu są nadal stosowane w radiu, telewizji czy kinematografii, ale wysokiej jakości zapis na taśmie wymaga zaangażowania większej liczby specjalistów i jest bardziej kosztowny. Cyfrowe formaty zapisu są wygodniejsze, lecz w celu zachowania wysokiej jakości należy wybrać formaty o dużej częstości próbkowania i rozdzielczości, które zwykle zajmują więcej pamięci niż

formaty niskiej jakości. Dla dźwięku próbkowanie o jakości zbliżonej do zakresu dźwięków słyszanych przez ludzkie ucho to 44,1 kHz, które określane jest jako CD (*compact disc* był pierwszym formatem cyfrowym w postaci płyt). Próbkowanie 22 kHz lub 11 kHz zajmie mniej miejsca na nośniku pamięci, ale pozwoli na mniej dokładny zapis głosu. Podobnie rozdzielczość obrazu w formacie full HD (1920 × 1080 px) daje nam obraz wyraźny w szerokiej palecie barw, a w formatach niższych obraz jest zamazany, ale zajmie mniej miejsca na nośniku pamięci. Lepiej nagrać dźwięk i obraz w wysokiej jakości i w razie potrzeby skompresować nagranie, bo poprawianie nagrania w niskiej jakości przez podwyższenie próbkowania czy rozdzielczości jest technicznie możliwe, lecz trudne (dostępne od 2020 roku algorytmy sztucznej inteligencji firmy Adobe uzupełniają dźwięk i obraz w celu zwiększenia jakości i formatu). Standardowym nośnikiem pamięci w nagrywarkach i kamerach jest karta SD dostępna w wielu pojemnościach. Przy wyborze należy także zwrócić uwagę na prędkość zapisu, która dla kamer jest określana jako CLASS 10 lub wyższa, ale dla nagrywarek dźwiękowych nie jest istotna. Karty pamięci SD występują w kilku pojemnościach: 32, 64, 128, 256 GB. Do nagrań warto mieć co najmniej dwie czyste karty o pojemności minimum 32 GB. Na jednej karcie o pojemności 32 GB można nagrać do 4 godzinnych filmów o rozdzielczości 1080 p w formacie *.mp4* i z kompresją H264 lub około 53 godziny dźwięku o próbkowaniu 44,1 kHz, stereo.

2.3.10. Formaty plików

Format zapisu danych dźwiękowych lub filmowych w pliku jest realizowany przy użyciu kodeków (*codecs*). Standardem zapisu plików dźwiękowych jest *.wav* i jest to zapis bezstratny, bez kompresji, a więc wymaga więcej pamięci niż formaty zapisu z kompresją, takie jak popularny *.mp3*. Plik zapisany w formacie *.wav* łatwo przetworzyć na format *.mp3*, ale operacja odwrotna nie podwyższy jakości. Dlatego należy sprawdzić, jaki format jest domyślny w nagrywarce czy programie do zapisu. Zapis może być mono, czyli jednokanałowy, lub stereo, czyli dwukanałowy. Zapis dwukanałowy można przetworzyć na zapis jednokanałowy, czyli oba kanały zapisać na jednej ścieżce, na której

dźwięk z lewego i prawego kanału będzie identyczny, lub rozdzielić kanały i zapisać w osobnych plikach, co jest nazywane separacją kanałów. Formaty dźwiękowe różnią się także częstością próbkowania, co przekłada się na jakość nagrania i jest mierzone w kHz (kiloherz to 1000 drgań cząsteczek powietrza na sekundę). Optymalne do analizy nagrań są następujące parametry kodowania nagrań dźwiękowych:

- format bezstratny: *.wav*,
- częstość próbkowania: 44,1 kHz lub 48 kHz,
- liczba kanałów: 2 (stereo) dla dwóch mówców lub jeden (mono) dla jednego mówcy,
- liczba bitów kodowania: 8 lub 16, ewentualnie 24.

Parametry podane wyżej są także zalecane w poradniku twórców programu ELAN. Minuta nagrania dźwięku stereo w formacie *.wav* (bez kompresji) zajmuje około 10 MB, a nagranie mono będzie dwukrotnie mniejsze. Konwersja nagrania mono do formatu stratnego *.mp3* przy kompresji 320 kbps utworzy czterokrotnie mniejszy plik niż plik stereo bez kompresji.

W przypadku plików filmowych nagrania w formacie bez kompresji są określane jako *RAW* i oferowane w drogich, specjalistycznych kamerach kinematograficznych (np. *blackmagic* oraz najwyższych modelach telefonów komórkowych *Apple iPhone* czy *Google Pixel*). Pliki *RAW* są jednak wielokrotnie większe niż skompresowane i do ich przetwarzania potrzebne jest specjalistyczne oprogramowanie do montażu filmów kinowych czy telewizyjnych. Tańsze kamery rejestrują obraz z kompresją stratną w plikach o rozszerzeniu *.mts* (*Movie Transport Stream*), *.mov*, *.mp4*, *.mpg* i starszym *.avi*. Starsze kodeki MPEG (*Motion Picture Experts Group*) są wypierane przez nowsze H264 i H265, które zapewniają optymalne proporcje pomiędzy jakością a wielkością pliku i są stosowane do strumieniowania multimediów w Internecie na platformach Netflix, YouTube lub Facebook.

Jeśli nagrywamy dźwięk i obraz osobno, to do pliku filmowego możemy dodać ścieżkę dźwiękową, ale należy ją zsynchronizować z obrazem; a jeśli nagrywamy dźwięk i obraz razem do jednego pliku, to możliwe jest także oddzielenie dźwięku od filmu, jeżeli potrzebny jest plik dźwiękowy do transkrypcji. Możliwe jest także usunięcie ścieżki dźwiękowej z filmu, jeśli potrzebny jest niemy film do anotacji.

Formaty zapisu różnią się (1) rodzajem kodowania, (2) rozdzielczością w pikselach i (3) współczynnikiem kompresji. Wszystkie wspomniane parametry mają wpływ na jakość nagrania i wielkość pliku, dlatego przy wyborze formatu zwracamy uwagę na cel nagrań i dostępną pamięć. Optymalne do analizy nagrań są następujące parametry kodowania nagrań filmowych:

- rodzaj kodowania (kodek): H264, H265 lub MPEG (rozszerzenie pliku: *.mp4*, *.mpg* lub *.mov*, ewentualnie *.avi*, ale ten format sprawia kłopoty w ELAN-ie);
- format filmu: 1080 p lub 720 p;
- rozdzielczość (rozmiar klatki): 1920 × 1080 px lub 1280 × 720 px;
- liczba klatek na sekundę: 25.

Szczegółowe zalecenia dotyczące kodowania oraz kompresji filmów i nagrań dźwiękowych znajdziemy w poradniku twórców ELAN-a, gdzie podano minimalne i maksymalne parametry nagrań, które ELAN wyświetla. Jeśli ELAN nie widzi filmu, który wczytał, to należy ten film przekonwertować do formatu opisanego w poradniku.

Jedna sekunda filmu w formacie 1080 p zajmuje około 10 MB, a w formacie 720 p – około 6,6 MB[7]. Dokładny rozmiar różni się zależnie od modelu kamery i ustawień nagrywania oraz kompresji. Przykładowe nagranie jednej sekundy na małej, poręcznej kamerze Panasonic HC-X900 w formacie 1080 p zajmuje około 2,3 MB, a godzina zajmuje ponad 8 GB. Natomiast po kompresji kodekiem H264 (*libx264 -preset slow -profile:v high -level 4.2*) godzina filmu zajmie około 1,5 GB, czyli prawie sześciokrotnie mniej.

[7] [@:] Wynik wyliczono przy pomocy kalkulatora rozmiaru pliku video dostępnego na stronie https://www.digitalrebellion.com/webapps/videocalc

2.4. Archiwizacja

Nagrania z urządzeń należy archiwizować, by móc je później analizować i by ich nie stracić w trakcie analiz. Im więcej kopii zapasowych, tym mniejsze ryzyko utraty danych, stąd powtórzenie archiwizacji. Nagrania z kart pamięci przenosimy na dyski zewnętrzne, najlepiej dwa. Pliki z pierwszego dysku będą analizowane w następnych etapach badań, a pliki na drugim dysku będą potrzebne, jeśli w trakcie analiz zauważymy zmiany i zniekształcenia, które nie są pożądane. Archiwizację plików warto wykonywać ręcznie lub automatycznie na co najmniej dwóch dyskach wewnętrznych, zewnętrznych oraz sieciowych, aby w razie awarii jednego móc odzyskać dane z drugiego. Dodatkowym, choć czasochłonnym i kosztownym, rozwiązaniem jest zachowanie kopii na płycie CD (0,7 GB), DVD (4,7 GB) lub BD (9,4 GB); dostępne są także płyty o pojemności 25, 50 i 100 GB, ale wymagają specjalnych nagrywarek i odtwarzaczy, rzadko produkowanych (np. *LaCie d2 Blu-ray XL*).

2.4.1. Normalizacja nagrań dźwiękowych

Nagrania dźwiękowe wymagają normalizacji, ponieważ w ich trakcie pojawiają się znacznie głośniejsze fragmenty, szumy i stuki (pukanie, klaskanie, trzaskanie drzwiami, kichnięcia, kaszlnięcia). Fragmenty głośniejsze należy wyciąć, a następnie całość nagrania znormalizować, czyli uśrednić natężenie dźwięku. W rezultacie cichsze fragmenty staną się głośniejsze, a znacznie głośniejsze – cichsze. Przesłuchiwanie całych nagrań i wycinanie niepotrzebnych fragmentów jest czasochłonne, ale istnieją programy, które wykonują takie operacje automatycznie. Należy także pamiętać, by nie usuwać fragmentów występujących podczas wypowiedzi, którą planujemy transkrybować i analizować, bo ingerencja może sprawić, że wypowiedź będzie znacznie zniekształcona. Dodatkowo, jeśli planujemy analizy obrazu i zauważamy niepotrzebne fragmenty dźwiękowe, rezygnujemy z ich usuwania i wyciszamy je, aby materiał wizualny pozostawić bez zmian.

2.4.2. Porządkowanie plików nagrań

W trakcie porządkowania plików nagrań należy pamiętać, że mają nazwy i rozszerzenia, które można dowolnie zmieniać, niezależnie od ich zawartości; pliki mają swoje daty utworzenia i zmian, które mogą się zmienić przy kopiowaniu lub synchronizacji; pliki mogą się zgubić lub źle zapisać, tak że niemożliwe będzie odzyskanie danych; a także o tym, że pliki zawsze są w chaosie i jedynie systematyczne pilnowanie ich parametrów i miejsc zapisu pozwala na uporządkowanie. Dodatkowo należy pamiętać, że pliki można sortować ręcznie lub automatycznie za pomocą skryptów w Bashu lub Pythonie, dzięki czemu tworzenie katalogów i przenoszenie plików jest łatwiejsze i szybsze, o ile nie dochodzi do pomyłek. Do sortowania plików można zastosować tagi zamiast katalogów, co umożliwia utworzenie bazy plików w jednym katalogu, w której pliki identyfikujemy po tagach jako grupy plików z tego samego urządzenia czy dnia nagrań albo z udziałem tego samego uczestnika i tym podobne. Pomocne jest także wyszukiwanie plików według zapytań o datę nagrania czy kod uczestnika i tak dalej. Wyszukiwania można zachowywać.

Nagrywarki i kamery nazywają pliki samoczynnie, zgodnie z ustawieniami producenta. Jeśli można je zmienić, to warto uwzględnić w nazwie pliku numer nagrania oraz datę i godzinę jego rozpoczęcia. Każdy plik zapisywany na komputerze ma swoje daty utworzenia i ostatnich zmian, ale po kopiowaniu i kompresji daty nowych plików mogą się różnić od dat plików oryginalnych. Zwykle w nazwie pliku z nagrywarki czy kamery jest numer nagrania, który pozwala na rozpoznanie kolejności nagrań. Do nazw plików dodajemy informacje o projekcie i uczestnikach. Informacje o projekcie kodujemy w postaci akronimu albo jednowyrazowej nazwy, gdyż długie nazwy plików nie są praktyczne. Dane uczestników kodujemy poprzez nadanie im niepowtarzalnych numerów lub ciągów liter i cyfr. Kody projektu i uczestników oddzielamy dywizem (-) lub podkreśleniem (_). Należy pamiętać, że systemy operacyjne nie pozwalają, by w nazwach plików i folderów znajdowały się następujące znaki:

< , > , : , “ , ‘, `, ~, / , \ , | , ? , !, $, *, #, %, &, {, }, @, +, =.

Celem uniknięcia nieporozumień i kłopotów z synchronizacją oraz kopiowaniem plików należy stosować jedynie znaki alfabetu angielskiego i cyfry oraz dywizy (-) i podkreślniki (_). Przykładowy system nazywania plików nagrań składa się z numeru nagrania, myślnika, numeru uczestnika, myślnika i numeru sesji, jeśli uczestnik brał udział w więcej niż jednym nagraniu, czyli na przykład 07-07-S1, 11-07-S2 albo 13-09-S1. Ponieważ pliki dźwiękowe i filmowe różnią się formatem i rozszerzeniem, nie ma potrzeby ich odróżniania w nazwie pliku. Pliki możemy kopiować, przenosić, usuwać, synchronizować między katalogami lub komputerami, a także konwertować na inne formaty i kompresować, żeby zajmowały mniej miejsca, oraz szyfrować. Oto przykładowa tabela nagrań z danymi o uczestnikach, terminach, nazwach plików i tak dalej.

Tabela 3. Przykładowa tabela nagrań z danymi z badania.

KOD NAGRANIA	DATA NAGRANIA	CZAS TRWANIA NAGRANIA	KOD UCZESTNIKA	WARUNEK	NAZWA PLIKU
04	23.10.23	04:55	92	A	04-92-A.wav
05	24.10.23	05:32	43	B	05-43-B.wav
08	24.10.23	07:21	65	A	08-65-A.wav

2.4.3. Kompresja plików

Pliki kompresujemy, aby zajmowały mniej miejsca. Do kompresji używamy formatu *.zip*, który jest popularny, więc łatwo pliki rozpakować na każdym komputerze i bezstratny, czyli dane po dekompresji są identyczne jak dane oryginalne. Pliki kompresowane można zabezpieczyć hasłem.

2.4.4. Synchronizacja plików

Jeśli pracujemy na dwóch komputerach albo udostępniamy nagrania członkom zespołu, mamy dwie możliwości: (1) wysyłać pliki na nośnikach lub jako załączniki e-mailem albo (2) synchronizować pliki na dysku na serwerze w sieci, czyli w chmurze danych. Każde z tych rozwiązań wymaga uwagi, bo każde bywa zawodne i ma swoje ograniczenia. Należy uważać, komu udostępniamy pliki, bo w żadnym przypadku nie mamy kontroli nad tym, co się stanie z plikiem. Nie udostępniamy plików, jeśli nie mamy kopii zapasowej, ani nie udostępniamy plików osobom, które nie są członkami zespołu badawczego. Pliki umieszczone na nośniku danych mogą trafić w niepowołane ręce albo zginąć wraz tym nośnikiem bądź sam nośnik może ulec awarii. Pliki załączone do e-maila mogą zostać cofnięte i nie dotrą do odbiorcy, a przesyłanie plików większych niż 1 MB może zapchać skrzynkę pocztową, a większych niż 25 MB może okazać się w ogóle niemożliwe. Zaletą wysyłania plików jako załączników jest to, że odbiorca pobiera kopię pliku, a załącznik pozostaje niezmieniony na serwerze pocztowym, dopóki nie zostanie usunięty. Innymi słowy, wysyłając plik jako załącznik, tworzymy także co najmniej jego dwie kopie zapasowe. Jednakże wysyłanie i odbieranie wielu plików jako załączników jest kłopotliwe, a jeśli chodzi o pliki anotacji lub inne dane, które są tworzone przez wiele osób, zalecane jest oznaczanie wersji plików.

Synchronizacja plików na dysku sieciowym jest rozwiązaniem wyżej wspomnianych problemów z plikami jako załącznikami w e-maliach lub na nośnikach pamięci. Warunkiem jest wybór odpowiednio pojemnego dysku sieciowego i jednego programu do synchronizacji. Popularne serwisy firm GOOGLE (Google Drive) lub MICROSOFT (Microsoft OneDrive) czy DROPBOX udostępniają aplikacje do automatycznej synchronizacji katalogów i plików, aczkolwiek jest to ograniczone do dysków wewnętrznych komputera. Dane na dysku zewnętrznym należy archiwizować za pomocą innych aplikacji. Ważną zaletą synchronizacji automatycznej jest także automatyczne wersjonowanie plików i możliwość przywrócenia wersji pliku na przykład sprzed trzech miesięcy od daty utworzenia.

Alternatywnym rozwiązaniem jest umieszczenie plików na jednym komputerze, do którego dostęp użytkownicy mają poprzez pulpit zdalny

lub inną aplikację, taką jak TEAM VIEWER czy ANYDESK. Aplikacje te pozwalają także na pracę wielu osób na tych samych plikach, ale niekoniecznie jednoczesną. Wygodne jest natomiast organizowanie spotkań, na których jedna z osób udostępnia pliki ze swojego komputera i omawia je z innymi badaczami.

2.5. Anonimizacja

W celu ukrycia danych osobowych uczestników stosujemy kodowanie, czyli dane osobowe pozostają w ankietach, jeśli są potrzebne, a członkom zespołu badawczego udostępniamy pliki nagrań, których nazwy mają identyfikatory uczestników (np. numery). Dodatkowo możemy zaciemnić lub rozmazać twarze na filmach za pomocą programów do anonimizacji takich jak *deface,* co opisujemy w sekcji 2.6.16. Anonimizacja nagrań dźwiękowych polega na usuwaniu danych osobowych i wrażliwych – imion i nazwisk, miejsca pracy, wykształcenia czy innych wyrazów według listy. Do anonimizacji nagrań dźwiękowych można wykorzystać usługi CLARIN-WEBMAUS, które opisujemy w sekcji 3.2.5.

2.6. Przetwarzanie nagrań filmowych

Pliki nagrań są zapisem surowych danych, które możemy dowolnie przetwarzać. Należy przy tym pamiętać, aby notować operacje na plikach, zachować wersje surowe, kontrolować nazywanie plików i korzystać z programów, które pozwalają na seryjne operacje na plikach (*batch processing*). Poniżej przedstawimy kilka podstawowych i przydatnych operacji na plikach, które są dostępne przy pomocy darmowego oprogramowania FFmpeg. Pakiet FFmpeg pobieramy ze strony, w wersji dla systemów LINUX, MACOS, WINDOWS. Po pobraniu i instalacji uruchamiamy z linii poleceń. W celu demonstracji efektów przetwarzania filmu posłużymy się nagraniem spotkania Obamy i Trumpa, które zostało wybrane ze względu na licencję (domena publiczna) i wysoki kontrast.

2.6.1. Konwersja

Konwersja formatów zostanie przeprowadzona automatycznie. Jeśli wskażemy tylko nazwy plików wejściowych i wyjściowych z opcją -i (*input*), FFmpeg dopasuje kodowanie filmu do formatu podanego w rozszerzeniu nazwy pliku wyjściowego: *.mp4*, *.mov*, *.mpg*, *.avi*.

- Konwertuj *.mts* do *.mp4*:
  `ffmpeg -i input.mts output.mp4.`
- Skopiuj ścieżkę filmu i dźwięku:
  `fmpeg -i input.mov -acodec copy -vcodec copy output.avi.`
- Pomiń kodowanie, żeby zaoszczędzić czas: -c:v copy, czyli skopiuj ścieżkę filmu:
  `ffmpeg -i input.mts -c:v copy output.mp4`
  `-c to -codec, a v to video, w skrócie -c:v.`
- Usuń ścieżkę dźwiękową z filmu i zachowaj film w formacie *.mp4*: -an:
  `ffmpeg -i input.mts -an output.mp4.`
- Wskaż format: -f mp4:
  `ffmpeg -i input.mts -c:v -f mp4 output.mp4.`

2.6.2. Skracanie

Skróć film: -ss -t -i

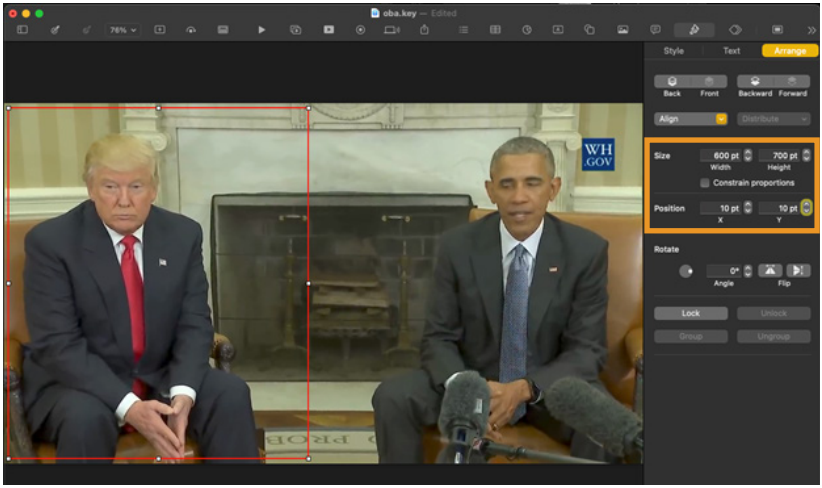
```
ffmpeg -i input.mp4 -ss 01:00 -t 01:00 -c:v copy
-c:a copy output.mp4
ffmpeg -i input.mp4 -ss 01:00 -to 02:00 -c:v copy
-c:a copy output.mp4
```

-ss to od; -to to do, a -t to czas trwania fragmentu do wycięcia z filmu.

2.6.3. Kadrowanie

Skadruj film: -vf „crop=w:h:x:y". Kadrowanie wymaga kilku kroków. Jeśli chcemy dokładnie wyciąć określony kadr z filmu, pomocny będzie program do prezentacji lub obróbki grafiki i następująca procedura:

1. Wygeneruj pierwsze ramki z filmów i sprawdź ich wielkość (np. 100):
   `fmpeg -i input.mp4 -vf "select=lt(n\,100)" -vsync vfr output_%03d.png.`
   Rozmiar ramki jest taki sam jak rozmiar wygenerowanych obrazów w formacie *.png*.
2. Utwórz prezentację o wielkości slajdów takiej samej jak ramka.
3. Wklej ramkę do prezentacji, czyli przenieś ramkę na listę slajdów.
4. Umieść ramkę w lewym górnym rogu slajdu, bez powiększania.
5. Ustaw miary na piksele lub punkty i liczenie od lewego górnego rogu, a nie od środka.
6. Narysuj prostokąt i zanotuj dane o jego szerokości, wysokości i położeniu, czyli *w*, *h*, *x*, *y*. Poniżej pokazujemy kadr z filmu z danymi w pomarańczowej ramce.
7. Wygeneruj w Excelu komendy ffmpeg i wykonaj w wierszu poleceń (*textjoin*).
   `ffmpeg -i input.mp4 -vf "crop=600:700:10:10" output.mp4.`
8. Sprawdź kadr. Popraw i powtórz.



Grafika 7. Kadrowanie filmu w programie do prezentacji Apple Keynote.

2.6.4. Zmiana rozdzielczości

Zmień rozdzielczość filmu, dodajemy -s i podajemy w pikselach:

```
ffmpeg -i input.mov -s 1280x720 output.mov
```

2.6.5. Rozdzielanie ścieżek

Wydziel ścieżkę dźwiękową:
Prosta konwersja z wideo na audio:

```
ffmpeg -i input.avi output.mp3
```

Konwersja z usunięciem wideo (-vn) i parametrami próbkowania (-ar), liczbą kanałów (-ac) i kompresją (-ab):

```
ffmpeg -i input.avi -vn -ar 44100 -ac 2 -ab 192 -f
mp3 output.mp3
```

2.6.6. Łączenie ścieżek

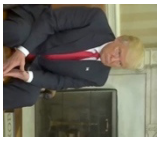
Połącz ścieżki dźwięku i obrazu:

```
ffmpeg -i input.wav -i input.avi output.mpg
```

2.6.7. Odwracanie i obracanie kadru

Odwróć i obróć kadr: komendy i przykłady przetworzonych kadrów podano w tabeli 4.

Tabela 4. Odwracanie i obracanie kadru za pomocą komend ffmpeg.

Odwróć pionowo (lustrzane odbicie w pionie):	`ffmpeg -i input.mp4 -vf vflip -c:a copy output.mp4`	
Odwróć pionowo (lustrzane odbicie w poziomie):	`ffmpeg -i input.mp4 -vf hflip -c:a copy output.mp4`	
Obróć o 90° niezgodnie z ruchem wskazówek zegara, czyli o 270° w prawo:	`ffmpeg -i input.mp4 -vf transpose=0 -c:a copy output.mp4`	
Obróć o 90° zgodnie z ruchem wskazówek zegara:	`ffmpeg -i input.mp4 -vf transpose=1 -c:a copy output.mp4`	
Obróć o 90° niezgodnie z ruchem wskazówek zegara i odwróć w pionie:	`ffmpeg -i input.mp4 -vf transpose=2 -c:a copy output.mp4`	
Obróć o 90° zgodnie z ruchem wskazówek zegara i odwróć w pionie:	`ffmpeg -i input.mp4 -vf transpose=3 -c:a copy output.mp4`	

2.6.8. Dodawanie znaków wodnych

Nałóż obraz na film:

```
ffmpeg -i input.mpg -vhook ,/usr/lib/vhook/watermark.
so -f watermark.png -m 1 -t 222222' -an output.mpg
```

2.6.9. Dodawanie tekstu

Nałóż nazwę pliku lub inny tekst na film:

```
ffmpeg -i input.mp4 -vf "drawtext=text='input.
mp4':x=10:y=10:fontsize=24:fontcolor=white" output.
mp4
```

2.6.10. Generowanie klatek

Wygeneruj klatki (obrazki) z filmu, czyli przetwórz na *.jpg* i ponumeruj (%d):

```
ffmpeg -i input.mpg %d.jpg
```

Określ liczbę generowanych klatek na sekundę: -r 2

```
ffmpeg -i input.mpg -r 2 -f image2 %d.png
```

2.6.11. Generowanie serii klatek

Wygeneruj klatki połączone w jeden obraz, na przykład 10 klatek, jedna po drugiej:

```
ffmpeg -i input.mp4 -frames 1 -vf "select=not(-
mod(n\,1)),tile=10x1" output.jpg
```

Grafika 8. Seria klatek z fragmentu filmu jako obraz.

2.6.12. Podawanie czasu trwania

Podaj długość filmu. Dane są dostępne przez ffprobe, czas trwania w sekundach poda:

```
ffprobe -v error -show_entries format=duration -of default=noprint_wrappers=1:nokey=1 input.mp4
```

Jeśli chcemy mieć czas trwania w formacie hh:mm:ss.sss, to przetwarzamy wynik. Krótki skrypt poda i zachowa do pliku o nazwie *m.txt* długość nagrań w plikach o rozszerzeniu *.mpg*:

Krótki skrypt poda i zachowa do pliku o nazwie *m.txt* długość plików o rozszerzeniu *.mpg*:

```
for m in *.mpg; do
    duration=$(ffprobe -v error -show_entries
    format=duration -of default=noprint_wrappers=1:nokey=1 $m)
    hours=$(bc <<< "$duration/3600")
    minutes=$(bc <<< "($duration%3600)/60")
    seconds=$(bc <<< "$duration%60")
    printf "$m %02d:%02d:%06.3f\n" $hours $minutes $seconds
    ; done > m.txt
```

2.6.13. Generowanie gifa

Wygeneruj gifa z filmu:

```
ffmpeg -i input.mp4 output.gif
```

2.6.14. Łączenie filmów w jeden film ciągły

Połącz próbki filmów jedna po drugiej:

1. Umieść wszystkie pliki, które chcesz połączyć, w jednym katalogu.
2. Przejdź do tego katalogu w terminalu (cd nazwakatalogu).
3. Wygeneruj listę plików jako S.txt:
   ```
   printf "file ,%s'\n" ./S*.mpg > S.txt.
   ```
4. Użyj tej listy w ffmpeg:
   ```
   ffmpeg -f concat -i S.txt -c copy output.mpg.
   ```

2.6.15. Łączenie filmów w jednym kadrze

Jeśli nagraliśmy jedną osobę dwiema kamerami i chcemy jednocześnie widzieć kadry z obu kamer w jednym kadrze, to pomocne jest połączenie filmów w poziomie. Filmy wejściowe powinny mieć taką samą rozdzielczość i długość. Zależnie od opcji efektem będzie film dwa razy szerszy (*hstack*) lub wyższy (*vstack*).

```
ffmpeg -i input1.mp4 -i input2.mp4 -filter_complex
"hstack=inputs=2" -map "" output.mp4
```

Grafika 9. Dwa filmy odwrócone w poziomie i złączone w jeden kadr, na środku można dostrzec linię dzielącą filmy.

2.6.16. Anonimizacja

Tożsamość i dane nagrywanych osób powinny być chronione, więc przed publikacją i prezentacją przykładów zamazujemy twarz, o ile celem badania nie jest mimika. Jeśli celem badania są ruchy rąk, to niektóre mogą zostać także zamazane, zwłaszcza jeśli będą wykonywane blisko twarzy. W takim przypadku najpierw anotujemy, a potem anonimizujemy. Do zamazania twarzy na filmach używamy programu Deface, który automatycznie znajduje i zamazuje twarze. Deface jest skryptem napisanym w Pythonie, a analiza opiera się na maszynowym uczeniu, więc jest automatyczny, ale zdarzają się pomyłki, na przykład konfiguracja palców może zostać rozpoznana jako twarz i rozmazana. Jeśli zauważymy taki błąd, a celem badań jest opis ruchu rąk i układu palców, anotację tego fragmentu przeprowadzimy na oryginalnym filmie. Proces zamazywania trwa bardzo długo, czasem nawet więcej niż film wejściowy. Efektem programu Deface jest zanonimizowany film pozbawiony dźwięku i z zamazanymi twarzami. Wyjściowy film jest zwykle o wiele mniejszy i jakościowo gorszy, ale jeśli zależy nam na jakości, to wskazujemy kodek, na przykład H264. Zanonimizować możemy także pliki obrazów. Szczególnie przydatne jest uruchomienie programu *deface* w pętli, na wszystkich plikach w wybranym katalogu. Jeśli filmy w sumie mają kilka godzin, to proces anonimizacji zajmie nawet kilkanaście godzin.

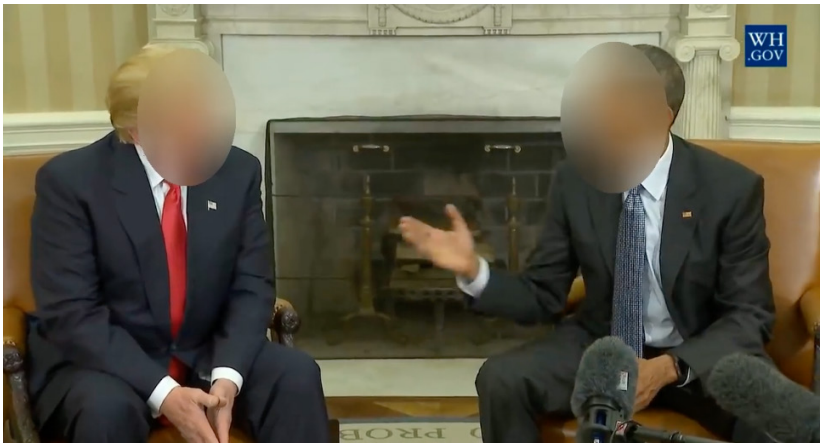
```
deface --ffmpeg-config ,{"codec": "libx264", "quali-
ty": 9}' input.mp4
```

Krótsza komenda anonimizacji z ustawieniami domyślnymi:

```
deface input.mp4
```

Jeśli nazwa pliku ma pozostać bez zmian, to podajemy:

-o input.mp4 albo w pętli -o $f, gdzie $f oznacza nazwę pliku wczytaną w pętli.



Grafika 10. Kadr filmu z zamazanymi twarzami obu osób rozpoznanych przez *program* deface.

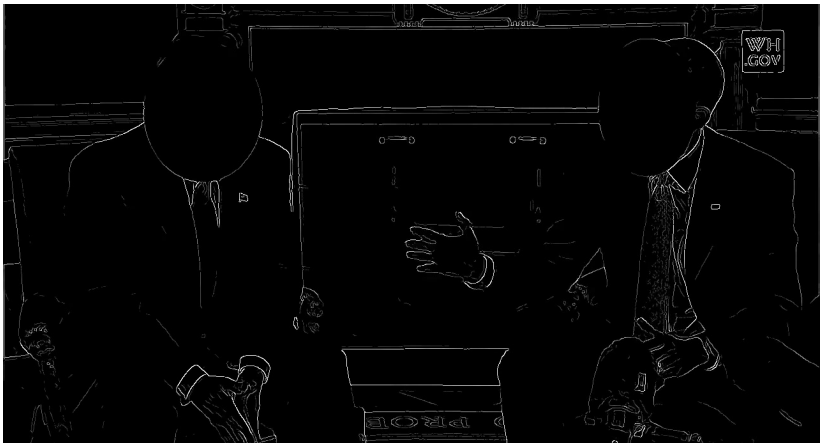
2.6.17. Wykrywanie krawędzi

Przefiltruj przez wykrywanie krawędzi obiektów i osób na filmie. Filtr *edge detect* jest używany do śledzenia ruchu na filmie i pozwala na częściowe ukrycie pól, których krawędzie zostały wykryte. Efektem działania filtru jest film monochromatyczny, czarno-biały:

```
ffmpeg -i input.mp4 -vf "edgedetect" output.mp4
```

lub niebiesko-biały:

```
ffmpeg -i input.mp4 -vf "edgedetect,colorbalance=bs=1.0" output_edge_colored.mp4
```



Grafika 11. Kadr filmu po nałożeniu filtru wyróżniającego krawędzie edge detect.

2.6.18. Dzielenie filmu na równe odcinki

Nagrania dźwiękowe dzielimy na próbki, jeśli musimy przeprowadzić transkrypcję automatyczną, która lepiej działa na mniejszych plikach, na przykład 10-minutowych. Nagrania filmowe dzielimy na potrzeby wyznaczania zgodności anotatorów, którą mierzymy dla ćwiartek (25%) albo dziesiątych części (10%) nagrań.

```
for i in {3..100}; do
    ffmpeg -i input.mp4 -ss $(($((i))*600)) -t
$(($((i+1))*600)) $i.mp4 ;
done;
```

Wersja komendy bez pętli:

```
ffmpeg -i input.wav segment -segment_time 10:00 out-
put%02d.wav
```

2.6.19. Dzielenie filmu na podstawie pliku *.eaf.*

Po wykonaniu anotacji przydatne jest zestawić fragmenty filmów, które mają takie same etykiety lub według innych kryteriów. Wówczas

pomocna będzie operacja, która podzieli film na fragmenty na podstawie pliku *.eaf*. Operacja składa się z dwóch skryptów. Pierwszy przetwarza dane z pliku *.eaf*, z którego pobierze nazwę próbki oraz początek i koniec próbki, i zachowuje je w pliku *.csv*.:

```
#!/bin/bash
eaf_file="$1"
    if [[ ! -f "$eaf_file" ]]; then
    echo "EAF file not found!"
exit 1
fi
output_csv="${eaf_file%.eaf}.csv"
# Print CSV header
echo  "\"TIER\",\"START_TIME\",\"END_TIME\",\"AN-
NOTATION\"" > "$output_csv"
# Get a list of unique tier IDs, sort them and then
remove duplicates
tiers=$(xmlstarlet sel -t -v "//TIER/@TIER_ID"
"$eaf_file" | sort | uniq)
# For each tier, extract the relevant details
IFS=$'\n'
for tier in $tiers; do
    xmlstarlet sel -t \
        -m "//TIER[@TIER_ID='$tier']/ANNOTATION/
        ALIGNABLE_ANNOTATION" \
        -v "concat('\"', '$tier', '\"', ',', '\"',
/ANNOTATION_DOCUMENT/TIME_ORDER/TIME_SLOT[@TIME_
SLOT_ID=current()/@TIME_SLOT_REF1]/@TIME_VALUE,
'\"', ',', '\"',
/ANNOTATION_DOCUMENT/TIME_ORDER/TIME_SLOT[@TIME_
SLOT_ID=current()/@TIME_SLOT_REF2]/@TIME_VALUE,
'\"', ',', '\"', ANNOTATION_VALUE, '\"')" \
        -n "$eaf_file" >> "$output_csv"
done
echo "Output saved to $output_csv"
```

Drugi skrypt wygeneruje próbki i złoży ich nazwy za pomocą danych z wcześniej utworzonego pliku *.csv*. Ten skrypt zamieni znaki zapytania na "question" i odstępy na podkreślniki (_) by uniknąć problemów przy dalszym przetwarzaniu plików, więc nazwa pliku na podstawie „S-LH-FOCUS,0,1018,in space" to „S-LH-FOCUS_0_1018_in_space__input.mp4".

```
#!/bin/bash
# Check for the correct number of arguments
if [ "$#" -ne 2 ]; then
    echo "Usage: $0 <input_video.mp4> <input.csv>"
    exit 1
fi

input_video="$1"
base_input_video=$(basename "$input_video" .mp4)
# extracting
base name without extension
csv_file="$2"

# Check if the files exist
if [[ ! -f "$input_video" || ! -f "$csv_file" ]];
then
    echo "Error: One of the files does not exist."
    exit 2
fi

# Skip the header line in the CSV
tail -n +2 "$csv_file" | while IFS=, read -r part1
start_ms end_ms
part4; do
   # Remove any surrounding quotes and potential
   carriage returns
    part1=$(echo $part1 | tr -d '"\r')
    start_ms=$(echo $start_ms | tr -d '"\r')
    end_ms=$(echo $end_ms | tr -d '"\r')
    part4=$(echo $part4 | tr -d '"\r')

    # Replace spaces with underscores and ? with
   'question'
    part1=${part1// /_}
    part4=${part4// /_}
    part4=${part4//\?/question}

    # Construct the filename using original ms format
    filename="${part1}_${start_ms}_${end_
   ms}_${part4}__${base_input_video}.mp4"
    # Print the ffmpeg command for debugging
```

```
    echo "Processing: $filename"
ffmpeg -i "$input_video" -ss "${start_ms}ms" -t "$(($end_
ms - $start_ms))ms" -c:v copy -c:a copy "$filename"
done
```

Aby uruchomić skrypty, należy zachować je w dwóch osobnych plikach o nazwach eaf2csv.sh i csv2spl.sh, a następnie uruchomić na pliku input.mp4 i input.eaf:

```
./eaf2csv.sh input.eaf
./csv2spl.sh input.mp4 input.csv
```

2.6.20. Przetwarzanie w pętli

W celu przetworzenia wielu plików tworzymy krótki skrypt z pętlą *for*, gdzie plik będzie oznaczony zmienną `file` i `$file`, a zakresem zmiennej będą wszystkie pliki o wybranym rozszerzeniu (formacie wejściowym), na przykład **.mts*. Pomiędzy `do` a `done` wpisujemy komendy, które mają być wykonane na kolejnych plikach wspomnianych w wyrażeniu pętli.

```
for file in *.mts; do
    echo "$file"
done
```

2.6.21. Przyśpieszanie przetwarzania

Przetwarzanie nagrań jest czasochłonne, więc jeśli mamy mocny procesor i dużo pamięci RAM, to możemy przyśpieszyć *ffmpeg* przez dodanie opcji -threads 8, gdzie 8 to liczba rdzeni procesora.

Podsumowując, chodzi o to, by film przekonwertować, skrócić, skadrować, usunąć ścieżkę dźwięku, zanonimizować, zanotować, podzielić, połączyć, podpisać, posortować w katalogach według sesji, przefiltrować by utworzyć korpusy próbek. Proponowany tok przetwarzania plików ma na celu stworzyć bazę próbek filmów z zachowaniami, które mają takie same etykiety, choć pochodzą od różnych uczestników nagrań. Próbki posortowane według etykiet (etykiety są w nazwach

próbek) są łączone w nowe filmy, na których możemy sprawdzić, czy widoczne jednostki badań, na przykład gesty o określonych cechach, zgadzają się z kryteriami systemu anotacji lub czy współwystępują z podobnymi jednostkami języka.

2.7. Podsumowanie

Poniższy schemat postępowania służy podsumowaniu wyżej opisanych etapów zbierania danych dla korpusu multimodalnego:

1. PLAN NAGRAŃ (kalendarz dostępności aparatury, sali lub studia, prowadzących nagrania i uczestników).
2. PRZYGOTOWANIE sali lub studia do nagrań (usunięcie zbędnych przedmiotów i urządzeń, sprawdzenie instalacji elektrycznej, sprawdzenie szczelności okien, sprawdzenie oświetlenia, ustalenie miejsc dla uczestników nagrań).
3. PRZYGOTOWANIE aparatury (rozmieszczenie aparatury ze statywami w sali, podłączenie i ukrycie przewodów, sprawdzenie stanu kart i dysków pamięci, sprawdzenie kadrów kamer i czułości mikrofonów).
4. USTALENIE sygnału startowego i końcowego do nagrań.
5. SPRAWDZENIE i synchronizacja czasu w każdym urządzeniu.
6. PRZED KAŻDYM NAGRANIEM:
   - 6.1. SPRAWDZENIE, czy nagrywane osoby mieszczą się w kadrze i czy ich głos jest słyszalny w słuchawkach podłączonych do nagrywarki;
   - 6.2. WŁĄCZENIE nagrywania we wszystkich urządzeniach jednocześnie lub w krótkich odstępach czasu, zawsze w tej samej kolejności, na przykład nagrywarka, kamery, czujniki;
   - 6.3. WYDANIE sygnału startowego;
   - 6.4. WYJŚCIE, jeśli aparatura i uczestnicy nagrań są w tej samej sali, prowadzący nie musi być obecny.
7. PO NAGRANIU:
   - 7.1. POWRÓT prowadzącego nagrania;
   - 7.2. WYDANIE sygnału końcowego;
   - 7.3. SKOPIOWANIE danych z kart pamięci na dysk zewnętrzny albo wymiana kart na czyste.

8. PO WSZYSTKICH NAGRANIACH:
   8.1. ANONIMIZACJA i ARCHIWIZACJA nagrań na dysku zewnętrznym lub sieciowym;
   8.2. UPORZĄDKOWANIE plików (nadawanie nazw i kodowanie danych uczestników i innych danych);
   8.3. UDOSTĘPNIANIE plików;
   8.4. ANALIZY i ANOTACJE.

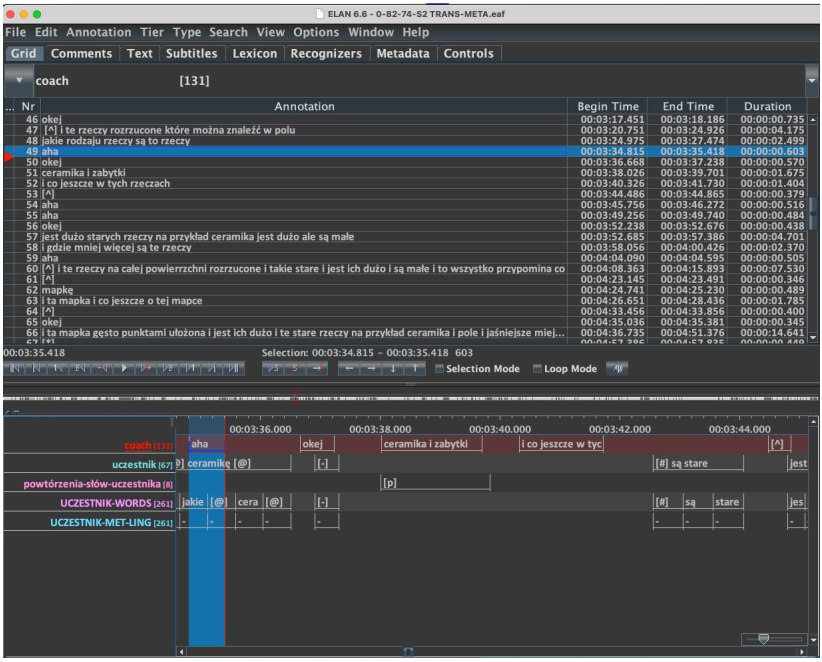
3. Transkrypcja i anotacja nagrań

Po zakończeniu fazy nagrań i porządkowania plików przechodzimy do opisu badanych zjawisk komunikacji multimodalnej, czyli transkrypcji mowy i anotacji pozostałych zachowań komunikacyjnych. Omówimy rodzaje i ogólne zasady transkrypcji ręcznej i automatycznej, a także typy anotacji.

3.1. Segmentacja

Przed przystąpieniem do transkrypcji wyznaczamy wypowiedzi, w których później wyznaczamy wyrazy. Wypowiedzią jest segment nagrania, który ma jeden kontur intonacyjny, czyli fraza intonacyjna, która w przybliżeniu przedstawia jedną myśl mówcy (Karpiński 2006; Szczyszek 2013). Zwykle wypowiedzi są oddzielane pauzami (wypełnionymi parawerbaliami lub nie), które mają od kilkuset milisekund (tysięcznych części sekundy) do kilku sekund (należy ustalić zakres trwania pauz dla danych nagrań) i są dodatkowym kryterium wydzielania wypowiedzi (Karpiński i Klessa 2021: 48-52). Wypowiedź niekoniecznie jest zdaniem, bo może być ciągiem wyrazów, z których żadne nie jest orzeczeniem. Wypowiedzi to zdania, wtrącenie, dopowiedzenia, urywki zdań, wykrzyknienia (Karpiński 2006; Szczyszek 2013). Wyznaczanie granic między wyrazami także jest problematyczne, zwłaszcza jeśli na nagraniu nie słyszymy całych wyrazów. Wówczas mamy dwie możliwości: dopasowujemy słyszane głoski do wyrazu, który występuje w słowniku danego języka, albo oznaczamy jako niekompletny, na przykład minusem w nawiasach kwadratowych (-). Wykonując transkrypcję, znamy słownictwo danego języka jako mówcy natywni języka polskiego, ale kiedy trafiamy na wyrazy niekompletne, niezrozumiałe i rzadkie, to odwołujemy się do wybranego słownika, w przypadku języka polskiego polecamy internetowy *wsjp.pl* czy *sjp.pwn.pl* albo starsze wydania papierowe (Szymczak i in. 1992). Grafika 12 pokazuje przykład transkrypcji z podziałem potoku mowy na wypowiedzi według konturów intonacji i dłuższych pauz (warstwy

coach i uczestnik) oraz podział wypowiedzi na wyrazy, które występują w słowniku danego języka (warstwa UCZESTNIK-WORDS).

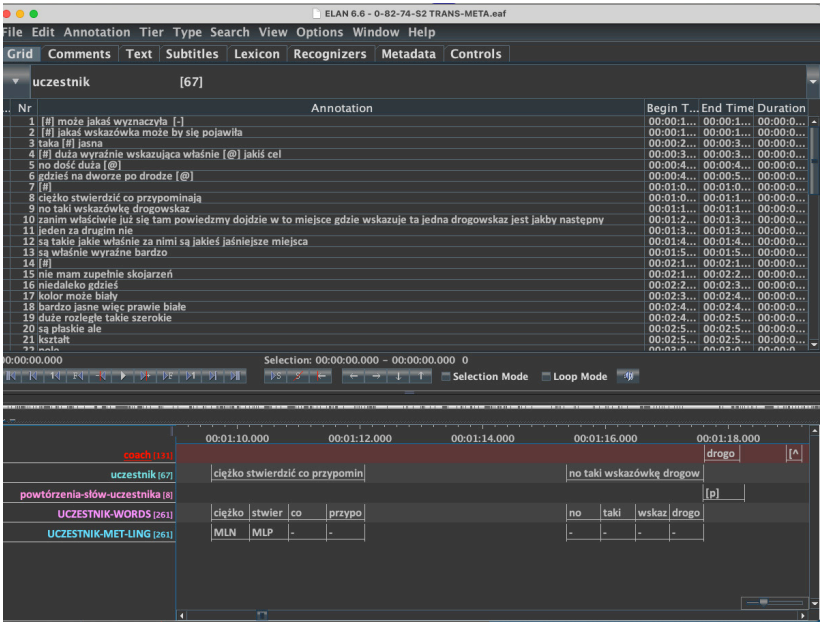


Grafika 12. Przykład podziału nagrania na wypowiedzi według konturów intonacji i dłuższych pauz (warstwy coach i uczestnik) oraz podział wypowiedzi na wyrazy, które występują w słowniku danego języka (warstwa UCZESTNIK-WORDS). Transkrypcja ręczna w ELAN-ie.

Transkrypcję wykonujemy w dowolnym edytorze tekstu, jeśli nie musimy oznaczać czasu wystąpienia ani trwania wypowiedzi. Wśród wielu programów, które umożliwiają transkrypcję z uwzględnieniem czasu wystąpienia wypowiedzi, najbardziej popularnym w zespołach badawczych jest ELAN[8]. Jego główną zaletą jest wielowarstwowość

[8] Nazwa ELAN to skrót od *EUDICO Linguistic Annotator*, gdzie EUDICO jest skrótem od *European Distributed Corpora Project.*

transkrypcji lub anotacji. Na osobnych warstwach umieszczamy wypowiedzi różnych osób na tym samym nagraniu, a także, jeśli to konieczne, różne jednostki języka: wypowiedzi, wyrazy, głoski, fonemy, oznaczenia intonacji, parawerbalia i inne. Rozdzielanie na warstwach pozwala na wyodrębnienie zjawisk na odmiennych poziomach języka i przeprowadzanie osobnych dla każdej warstwy analiz anotacji, takich jak analizy syntaktyczne, leksykalne, słowotwórcze, morfologiczne, pragmatyczne i inne. Grafika 13 pokazuje anotacje na wielu warstwach: dwóch mówców (warstwy nazwane coach i uczestnik) oraz anotacje słów uczestnika pod względem metaforyczności (UCZESTNIK--WORDS i UCZESTNIK-MET-LING).



Grafika 13. Anotacje na wielu warstwach w ELAN-ie.

Ręczna transkrypcja jest czasochłonna, ale ELAN oferuje pomocne moduły. Na przykład do segmentacji dostępny jest moduł wykrywania

segmentów ciszy (*silence detector*), który automatycznie wykrywa i oznacza pauzy niewypełnione w nagraniu dźwiękowym. Czułość i inne parametry tego narzędzia można zmodyfikować i dostosować do zebranych nagrań. Po segmentacji nagrania przechodzimy do transkrypcji.

3.2. Transkrypcja

Przepisywanie mowy ma na celu utrwalenie mowy w postaci tekstowej, która znacznie ułatwia analizę języka w językoznawstwie korpusowym i komputerowym (komputacyjnym), a także w stosowanym i opisowym (fonetyce, fonologii, logopedii, dialektologii, psycholingwistyce, akwizycji języka). Pierwszym etapem jest segmentacja, czyli podział potoku mowy na mniejsze jednostki: wypowiedzi, frazy, wyrazy, sylaby. Następnie decydujemy o zakresie i rodzaju transkrypcji. Rozróżniamy cztery typy transkrypcji ze względu na jednostkę i cel badań:

1. fonetyczną,
2. fonologiczną,
3. ortograficzną,
4. suprasegmentalną.

Obok transkrypcji wyróżnia się także transliterację, czyli przypisanie literom jednego alfabetu lub systemu pisma znaków z innego alfabetu albo systemu pisma. Transliteracja jest stosowana głównie w językoznawstwie porównawczym i dokumentacji języków (Karpiński i Klessa 2021).

3.2.1. Transkrypcja fonetyczna

Fonetyczna transkrypcja polega na przypisaniu znaków alfabetu fonetycznego do głosek, by zapisać wymowę, jaką słyszymy na nagraniu. Transkrypcja fonetyczna zależnie od celów badań ma różny stopień dokładności, na przykład do szczegółowego opisu wymowy w ramach badań logopedycznych czy fonetycznych stosuje się bardziej rozbudowany zestaw liter i znaków diakrytycznych niż w badaniach dla językoznawstwa korpusowego, kiedy potrzebna jest jedynie transkrypcja zgodna z ogólnymi zasadami wymowy w danym języku.

3.2.2. Transkrypcja fonologiczna

W transkrypcji fonologicznej, zwanej także fonemiczną, litery oznaczają fonemy w wybranym systemie badanego języka. W transkrypcji fonetycznej i fonologicznej używamy alfabetów fonetycznych, na przykład IPA (*International Phonetic Alphabet*) albo SPA (*Slavistic Phonetic Alphabet*). Istotną różnicą pomiędzy transkrypcją fonetyczną i fonologiczną a ortograficzną jest stosowanie pojedynczych liter dla pojedynczych głosek w przypadku transkrypcji fonetycznej lub fonemów w przypadku transkrypcji fonemicznej, podczas gdy w transkrypcji ortograficznej posługujemy się alfabetem niefonetycznym, w którym zdarza się, że niektóre głoski są oznaczane przez więcej niż jedną literę, albo odwrotnie – niektóre litery oznaczają więcej niż jedną głoskę, zależnie od kontekstu i zasad pisowni w danym języku.

W projekcie NARRACJE zastosowaliśmy szeroką transkrypcję fonetyczną, w której „transkrybujący starają się oddać w przybliżeniu słyszane segmenty wypowiedzi i unikać w miarę możliwości sugerowania się własną teoretyczną kompetencją fonologiczną mówcy języka polskiego. W konsekwencji, nie wyklucza się możliwości pojawienia się zapisów, które nie stanowią odzwierciedlenia poprawnej wymowy wyrazów języka polskiego, ani nawet odpowiadających jego systemowi fonologicznemu i fonotaktyce logatomów. Jest to możliwe szczególnie w przypadku różnego typu zaburzeń płynności wypowiedzi (zająknięcia, pauzy wypełnione)" (Karpiński i in. 2008: 85).

3.2.3. Transkrypcja ortograficzna

Transkrypcja ortograficzna polega na przypisaniu słyszanym głoskom liter alfabetu lub znaków systemu pisma dla badanego języka zgodnie z zasadami pisowni w tym języku. W przypadku języka polskiego w transkrypcji ortograficznej w niektórych badaniach rezygnuje się ze znaków interpunkcyjnych i zasad pisowni wielką literą, gdyż celem jest rekonstrukcja wypowiedzi, a nie tekstu w odmianie pisanej (Szczyszek 2013). Szczegółowe zasady transkrypcji w języku polskim zawarto między innymi w podręcznikach fonetyki i fonologii języka polskiego (Wiśniewski 2007; Ostaszewska 2008; Madelska i Witaszek-Samborska 2015).

3.2.4. Transkrypcja suprasegmentalna

Transkrypcję suprasegmentalną stosujemy, jeśli badanie dotyczy intonacji i zachowań parawerbalnych oraz specyficznych cech wymowy nagrywanych osób. Wówczas posługujemy się bardziej szczegółowym alfabetem fonetycznym albo systemem oznaczeń opracowanym przez badaczy konwersacji (*conversation analysis*). Parawerbaliami nazywamy odgłosy wydawane przez aparat artykulacyjny ludzi, które nie należą do systemu dźwiękowego badanego języka, czyli pauzy wypełnione śmiechem, chrząknięciem, westchnieniem, cmokaniem i tym podobne. W tabeli 5 podajemy listę oznaczeń parawerbaliów stosowaną w projektach opisujących konwersacje.

Do badań w językoznawstwie korpusowym i komputerowym, jeśli nie planujemy analizy wymowy, wystarczy transkrypcja ortograficzna. Niezależnie od wybranego typu transkrypcji należy pamiętać, że transkrypcja jest interpretacją mowy, czyli badacz, transkrybujący, podejmuje decyzję, jakimi literami oznaczyć słyszane głoski, ale to wcale nie oznacza, że transkrypcja jest dokładnie tym, co ktoś powiedział. Dlatego, podobnie jak w przypadku anotacji innych zachowań komunikacyjnych, konieczne są: kontrola przebiegu transkrypcji, ustalenie kryteriów transkrypcji i poprawienie przez bardziej doświadczonego badacza. Oto przykład zasad transkrypcji zastosowanych w projekcie nagrań sesji coachingowych (dialogów) (Juszczyk 2017), opracowanych przez Szczyszka (2013).

Tabela 5. Przykład wytycznych dotyczące transkrypcji.

Wytyczne do transkrypcji w projekcie MULTIMET

1. Transkrypcje wykonujemy w programie ELAN (aktualna wersja: http://tla.mpi.nl/tools/tla-tools/elan/download/).
2. Transkrypcje wykonujemy zgodnie z polską współczesną ortografią – standard ortograficzny zapisany w *Wielkim słowniku ortograficznym PWN z zasadami pisowni i interpunkcji*, redakcja naukowa oraz opracowanie zasad pisowni i interpunkcji polskiej prof. Edward Polański, wydanie III, poprawione i uzupełnione, Warszawa 2010.
3. Wielkie litery w nazwach własnych czy na początku zdania (wypowiedzenia) są nieistotne, więc pomijalne. Dlatego zamiast wielkich liter w nazwach własnych stosujemy małe litery.
4. Nie stosujemy znaków interpunkcyjnych.
5. Podstawową jednostką transkrypcji jest fraza intonacyjna. Transkrypcja jest przeprowadzana fraza po frazie. Frazy wyznaczane są konturem intonacyjnym (pomocą służą oscylogramy (*waveforms*) w ELANie). Fraza może mieć postać jednego wyrazu lub grupy wyrazów.
6. Zapisujemy wszystkie słyszalne i zrozumiałe wypowiedzi zarówno COACHA, jak i UCZESTNIKA (na osobnych warstwach).
7. Fragmenty niezrozumiałe zapisujemy w postaci: [-]. TO JEST DYWIZ!
8. Jeśli zrozumiały jest tylko fragment wyrazu tekstowego, a jego pozostała część jest niezrozumiała i niemożliwa do rekonstrukcji (nawet w kontekście), to transkrybujemy to, co jest zrozumiałe (stawiając kropkę po część zrozumiałej), a resztę oznacza się jako niezrozumiałe (w nawiasie kwadratowym zapisuje się [-]), np. *strukt.[-]*.
9. Jeśli zrozumiały jest tylko fragment wyrazu tekstowego, a jego pozostała część jest niezrozumiała, ale możliwa do zrekonstruowania, to transkrybujemy to, co jest zrozumiałe (stawiając kropkę po część zrozumiałej), a rekonstrukcję reszty wyrazu zapisujemy w nawiasie kwadratowym […]), np. *strukt.[uralizm]*.
10. W szablonie będą dwie główne warstwy transkrypcji: COACH, UCZESTNIK oraz jedna dodatkowa: POWTÓRZENIA.
11. Pauz cichych – nie oznaczamy (mogą w nich wystąpić szumy techniczne, przesuwanie krzeseł).
12. Pauzy wypełnione – czyli wszystkie pozasłowne elementy wypowiedzi artykułowane przez człowieka; wyodrębniono ich 5 rodzajów oraz zbiór otwarty – oznaczamy następującymi symbolami:

1. śmiech: [@],
2. jęki namysłu (np.: mmm, yyy, aaa): [#],
3. westchnienie: [$],
4. chrząknięcie: [%],
5. cmoknięcie: [&],
6. „uhm”, „mhm” w znaczeniu: „potwierdzam, że słucham; przytaknięcie”: [^],
7. inne (niewyodrębnione powyżej, ale artykułowane przez człowieka): [*].

3.2.5. Transkrypcja automatyczna

Doświadczeni transkrybujący potrafią przepisać nagranie na tekst w tempie 5 minut pracy na minutę nagrania, zatem transkrypcja godziny nagrania zajmie około 5 godzin, a całego korpusu liczącego kilkanaście godzin – kilka dni pracy. Transkrypcja fonetyczna czy fonologiczna wymaga więcej czasu niż ortograficzna. Transkrypcja jest nie tylko czasochłonna, ale także kosztowna, bo dokładna transkrypcja, zwłaszcza fonetyczna, wymaga znalezienia przeszkolonych specjalistów lub przeszkolenia transkrybentów.

Alternatywą jest transkrypcja automatyczna, czyli serwisy oparte na rozpoznawaniu mowy (*speech to text*, *speech recognition services*). Transkrypcja automatyczna trwa mniej więcej tak samo długo jak nagranie, którego dotyczy. W Internecie znajdziemy setki takich serwisów, płatnych i bezpłatnych, w większości jednak są to serwisy generujące transkrypcję ortograficzną z nagrań przemówień czy notatek głosowych i spotkań biznesowych. Natomiast do celów badawczych potrzebna jest transkrypcja dokładna. Omówimy krótko trzy serwisy: Whisper firmy OPENAI i Cloud speech-to-text firmy Google oraz CLARIN MOWA, które są bezpłatne, mają wysoką, potwierdzoną badaniami, dokładność transkrypcji i zostały opracowane przez zespoły badawcze. Serwisy do rozpoznawania mowy są oparte na maszynowym uczeniu i trenowaniu modeli języka na ogromnych korpusach wielu języków. Na przykład najnowszy, bo udostępniony w 2022 roku, serwis Whisper był trenowany na 680 000 godzin nagrań i ma dokładność na poziomie kilku procent błędów (*word error rate*) w takich językach, jak hiszpański, angielski, niemiecki czy polski, czyli zbliżoną do poziomu błędów w transkrypcji wykonanej przez człowieka (Radford i in. 2022). System rozpoznający mowę w języku polskim dostępny na *web-maus* był trenowany na korpusie CLARIN-PL Studio corpus, który liczy 305 000 słów[9]. Dla pracowników uczelni usługi Google i IBM są udostępniane bezpłatnie między innymi przez konsorcjum badawcze CLARIN na niemieckiej stronie *web-maus*. Główną zaletą web-maus jest generowanie pliku w for-

[9] [@:]https://clarin.phonetik.uni-muenchen.de/BASWebServices/services/runMAUSGetInventar?LANGUAGE=pol-PL

macie ELAN-a (*.eaf*), a także rozróżnianie i rozdzielanie mówców (*speaker diarization*) oraz transkrypcja fonetyczna i fonologiczna. Whisper i Google są wielojęzyczne i oferują transkrypcję dla 97 (Whisper) i dla ponad 300 (Google) języków, w tym także dla języka polskiego. Whisper jest także dostępny w ELAN-ie (Rodríguez i Cox 2023).

Przed przystąpieniem do transkrypcji automatycznej należy zapoznać się z instrukcją danego serwisu, by dowiedzieć się, w jakich formatach są przyjmowane pliki dźwiękowe. Zwykle akceptowany jest format plików 16 kHz, jednokanałowy (mono) i podzielony na krótsze fragmenty (np. 10-minutowe). Konwersję i podział przeprowadzimy za pomocą programów do analizy nagrań, takich jak Audacity lub FFmpeg.

3.2.5.1. CLARIN-PL mowa

Usługi CLARIN-PL mowa są darmowe i dostępne na stronie www. Obok rozpoznawania mowy (serwis zwany RECO) znajdziemy:

- ALIGN – narzędzie do dopasowywania tekstu do audio,
- DIA – narzędzie do rozpoznawania mówców (biometryka głosu),
- G2P – konwersję zapisu ortograficznego na fonetyczny (*grapheme-to-phoneme*),
- KWS – detekcję słów kluczowych (*keyword spotting*),
- VAD – detekcję mowy (*voice activity detection*).

Serwis CLARIN-PL mowa umożliwia stworzenie własnego korpusu nagrań i transkrypcji. Wszystkie wspomniane serwisy są dostępne także programistycznie, czyli z wiersza poleceń lub przez języki programowania Python czy Java.

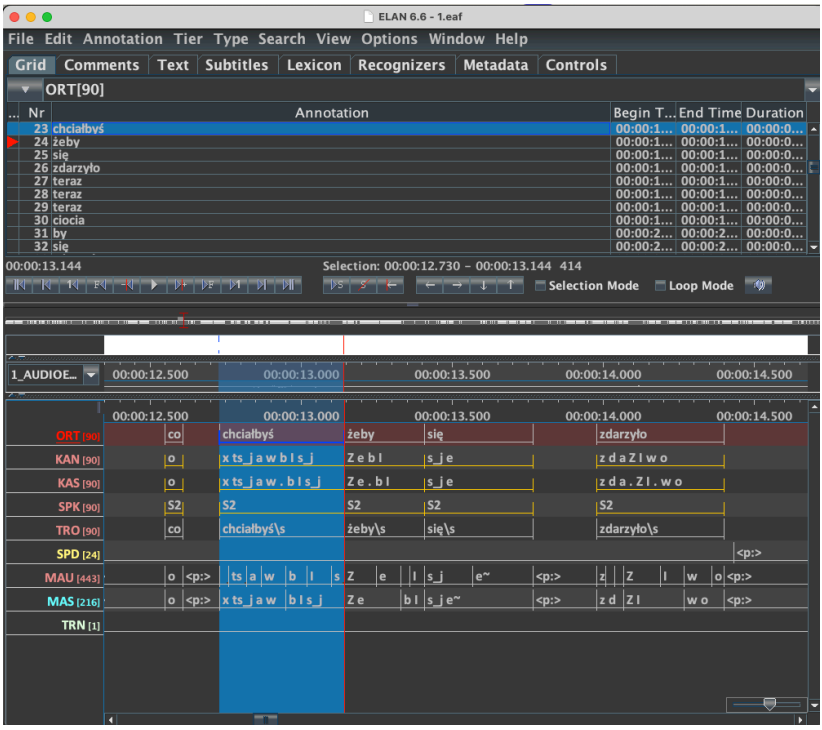
3.2.5.2. CLARIN BASWebServices

Usługi przetwarzania nagrań są dostępne także na niemieckim serwisie CLARIN (Schiel 1999). Rozpoznawanie mowy wymaga zalogowania się loginem i hasłem uniwersytetu, na przykład Adam Mickiewicz University. W celu przetranskrybowania nagrania należy wczytać pliki

dźwiękowe i wybrać potok przetwarzania (*pipeline*), na przykład: ASR > G2P > CHUNKER > MAUS > PHO2SYL:

- ASR – *automatic speech recognition*: automatyczne rozpoznawanie mowy,
- G2P – *grapheme-to-phoneme*: transliteracja z transkrypcji ortograficznej na fonemiczną,
- CHUNKER – *chunk pre-segmentation*: dzielenie wypowiedzi na segmenty,
- MAUS – *phonetic segmentation*: dzielenie mowy na głoski,
- PHO2SYL – *syllabification* (*phonemic and phonetic*): dzielenie mowy na sylaby.

Dostępna jest także usługa rozpoznawania mówców: SD = *speaker diarization* (*‚SpeakDiar'*). Następnie wybieramy format pliku wyjściowego, na przykład *.eaf* lub *.csv* i opcje dodatkowe. Jeśli w opcjach dodatkowych podamy adres e-mailowy, to po transkrypcji otrzymamy powiadomienie o zakończonej pracy i link do plików z transkrypcją do pobrania. Pliki zostaną po 24 godzinach automatycznie usunięte z serwerów CLARIN. Grafika 14 pokazuje przykład rezultatu automatycznej transkrypcji w serwisie CLARIN WEBMAUS z podziałem na wypowiedzi, wyrazy, sylaby, głoski zapisane w alfabecie X-SAMPA (Wells 2000; Demenko, Wypych i Baranowska 2003).



Grafika 14. Przykład podziału wypowiedzi na wyrazy oraz wyrazów na głoski i sylaby w ELAN-ie. Transkrypcja *automatyczna* uzyskana za pomocą narzędzia CLARIN WEBMAUS.

3.2.5.3. Whisper

Usługa Whisper jest nowa i zapewne będzie jeszcze wielokrotnie aktualizowana, więc niniejsza procedura transkrypcji nagrań zmieni się, prawdopodobnie powstaną aplikacje do transkrypcji, które nie będą wymagały używania wiersza poleceń. Aktualnie, pod koniec 2023 roku, mamy dostęp do Whispera głównie przez wiersz poleceń lub przez programowanie w Pythonie.

1. W celu wywołania Whispera (wersja dla Windowsa) musimy zainstalować Chocolatey, Conda, Python, ffmpeg, CUDA, Pytorch.

Na szczęście w Internecie znajdziemy skrypty, które wykonają instalację za nas[10].

```
# Copyright (C) 2023 TroubleChute (Wesley Pyburn)
# Licensed under the GNU General Public License v3.0 (the "License");
# you may not use this file except in compliance with the License.
# You may obtain a copy of the License at
#
    # https://www.gnu.org/licenses/gpl-3.0.en.html
    #
    # This program is free software: you can redistribute it and/or modify
    # it under the terms of the GNU General Public License as published by
    # the Free Software Foundation, either version 3 of the License, or
    # (at your option) any later version.
    #
    # This program is distributed in the hope that it will be useful,
    # but WITHOUT ANY WARRANTY; without even the implied warranty of
    # MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
    # GNU General Public License for more details.
    #
    # You should have received a copy of the GNU General Public License
    # along with this program. If not, see .
    #
    # -----------------------------------------
# This script:
# 1. Installs Chocolatey (for installing Python and FFMPEG) - https://chocolatey.org/install
# 2. Check if Conda or Python is installed. If neither: install Python using Choco (if Python not already detected)
```

[10] Skrypt wklejono na następnej stronie, a aktualizowana wersja jest dostępna na https://github.com/TCNOco/TcNo-TCHT/blob/main/PowerShell/AI/whisper.ps1.

```
# 3. Installs FFMPEG using Choco (if FFMPEG not
already detected)
# 4. Install CUDA using Choco (if CUDA not al-
ready detected)
# 5. Install Pytorch if not already installed, or
update. Installs either GPU version if CUDA found,
or CPU-only version
# 6. Verify that Whisper is installed. Reinstall
using another method if not.
# --------------------------------------

Write-Host "--------------------------------------
-------" -ForegroundColor Cyan
Write-Host "Welcome to TroubleChute's Whisper in-
staller!" -ForegroundColor Cyan
Write-Host "Whisper as well as all of its other
dependencies should now be installed..." -Fore-
groundColor Cyan
Write-Host "[Version 2023-06-06]" -ForegroundColor
Cyan
Write-Host "`nThis script is provided AS-IS without
warranty of any kind. See https://tc.ht/privacy &
https://tc.ht/terms."
Write-Host "Consider supporting these install
scripts: https://tc.ht/support" -ForegroundColor
Green
Write-Host "--------------------------------------
-------`n`n" -ForegroundColor Cyan

Set-Variable ProgressPreference SilentlyContinue
# Remove annoying yellow progress bars when doing
Invoke-WebRequest for this session

# 1. Install Chocolatey
Write-Host "`nInstalling Chocolatey..." -Fore-
groundColor Cyan
Set-ExecutionPolicy Bypass -Scope Process -Force;
[System.Net.ServicePointManager]::SecurityProtocol
= [System.Net.ServicePointManager]::SecurityProto-
col -bor 3072; iex ((New-Object System.Net.WebCli-
```

```
ent).DownloadString('https://community.chocolatey.
org/install.ps1'))

# Import function to reload without needing to re-
open Powershell
iex (irm refreshenv.tc.ht)

# 2. Check if Conda or Python is installed
# Check if Conda is installed
Import-FunctionIfNotExists -Command Get-UseConda
-ScriptUri "Get-Python.tc.ht"
# Check if Conda is installed
$condaFound = Get-UseConda -Name "Whisper" -EnvName
"whisper" -PythonVersion "3.10.11"
# Get Python command (eg. python, python3) & Check
for compatible version
if ($condaFound) {
   conda activate "whisper"
   $python = "python"
} else {
   $python = Get-Python -PythonRegex 'Python ([3].
[1][0-1].[6-9]|3.10.1[0-1])' -PythonRegexExpla-
nation "Python version is not between 3.10.6 and
3.10.11." -PythonInstallVersion "3.10.11" -Manu-
alInstallGuide "https://hub.tcno.co/ai/whisper/
install/"
    if ($python -eq "miniconda") {
    $python = "python"
    $condaFound = $true
}
}
# 3. Install FFMPEG with Choco if not already in-
stalled.
if (-not (Get-Command ffmpeg -ErrorAction Silent-
lyContinue)) {
   Write-Host "`nFFMPEG is not installed. Install-
   ing..." -ForegroundColor Cyan
   choco upgrade ffmpeg -y
   Update-SessionEnvironment
}
```

```
if (Get-Command ffmpeg -ErrorAction SilentlyContin-
ue) {
Write-Host "FFmpeg is installed." -ForegroundColor
Green
}
else {
  Write-Host "FFmpeg is not installed. Please add
FFMPEG to PATH (install ffmpeg) and run this script
again." -ForegroundColor Red
  Write-Host "Alternatively, follow this guide
for manual installation: https://hub.tcno.co/ai/
whisper/install/" -ForegroundColor Red
  Read-Host "Process can not continue. The program
will exit when you press Enter to continue..."
 Exit

iex (irm Import-RemoteFunction.tc.ht)
# 4. Install CUDA using Choco if not already in-
stalled.
if ((Get-CimInstance Win32_VideoController).Name
-like "*Nvidia*") {
  Import-FunctionIfNotExists -Command Install-Cu-
daAndcuDNN -ScriptUri "Install-Cuda.tc.ht"
  Install-CudaAndcuDNN -CudaVersion "11.8" -Cud-
nnOptional $true

# 5. Install Pytorch if not already installed,
or update.
  Write-Host "`nInstalling or updating PyTorch
(With GPU support)..." -ForegroundColor Cyan
  if ($condaFound){
  conda install pytorch torchvision torchaudio
pytorch-cuda=11.8 -c pytorch -c nvidia -y
} else {
  &$python -m pip install --upgrade torch torch-
vision torchaudio --index-url https://download.
pytorch.org/whl/cu118
  }
} else {
```

```
  Write-Host "Nvidia CUDA is not installed. Please
install the latest Nvidia CUDA Toolkit and run this
script again." -ForegroundColor Red
  Write-Host "For now the script will proceed with
installing CPU-only PyTorch. Whisper will still
run when it's done." -ForegroundColor Red

# 5. Install Pytorch if not already installed,
or update.
Write-Host "`nInstalling or updating PyTorch
(CPU-only)..." -ForegroundColor Cyan
if ($condaFound) {
  conda install pytorch torchvision torchaudio
cpuonly -c pytorch -y
  } else {
&$python -m pip install torch torchvision torchau-
dio
  }
}
Write-Host "`nInstalling or updating Whisper..."
-ForegroundColor Cyan
if ($condaFound) {
  # For some reason conda NEEDS to be deactivated
and reactivated to use pip reliably... Otherwise
python and pip are not found.
  conda deactivate
  #Open-Conda
  conda activate whisper
  pip install -U openai-whisper # Environment is
already active
} else {
  &$python -m pip install -U openai-whisper
  Update-SessionEnvironment
}
# 6. Verify that Whisper is installed. Reinstall
using another method if not.
if (Get-Command whisper -ErrorAction SilentlyCon-
tinue) {
  Write-Host "`n`nWhisper is installed!" -Fore-
groundColor Green
```

```
  Write-Host "You can now use `whisper --help` for
more information in this PowerShell window, CMD or
another program!" -ForegroundColor Green
}
else {
  Write-Host "Whisper is not installed, trying
again but this time installing from the openai/
whisper GitHub repo" -ForegroundColor Green

  if ($condaFound){
   pip install -U setuptools-rust
   pip install git+https://github.com/openai/
whisper.git
} else {
  &$python -m pip install -U setuptools-rust
  &$python -m pip install -U --no-deps --force-re-
install git+https://github.com/openai/whisper.git
}

  if (Get-Command whisper -ErrorAction Silently-
Continue) {
   Write-Host "`n`nWhisper is installed!" -Fore-
groundColor Green
   Write-Host "You can now use whisper --help for
more information in this PowerShell window, CMD or
another program!" -ForegroundColor Green
} else {
  Write-Host "`n`nWhisper is not installed. Please
follow this guide for manual installation: https://
hub.tcno.co/ai/whisper/install/" -ForegroundColor
Red
  Read-Host "Process can not continue. The program
will exit when you press Enter to continue..."
}
}
```

2. Pliki dźwiękowe przed transkrypcją dzielimy na mniejsze, bo wówczas transkrypcja będzie sprawniejsza i serwis nie ulegnie przepełnieniu.

```
ffmpeg -i input.wav segment -segment_time 10:00 out-
put%02d.wav
```

3. Pliki dźwiękowe konwertujemy do formatu 16 kHz, mono:
```
ffmpeg -i input.wav -ac 1 -ar 16000 output-16k.wav
```
a w pętli dla wszystkich plików w katalogu:
```
for f in *.wav; do ffmpeg -i $f -ac 1 -ar 16000
$f-16k.wav; done;
```
4. Whisper ma wiele parametrów i trybów pracy, więc ograniczymy się do komendy inicjującej transkrypcję nagrań w języku polskim. Do transkrypcji polskiego potrzebny jest duży model językowy *ggml-large.bin*.
```
set model = large curl - L
https://huggingface.co/ggerganov/whisper.cpp/re-
solve/main/ggml-%model%.bin -o models\ggml-%model
%.bin
```
5. Inicjujemy transkrypcję (parametr -l określa język nagrania do rozpoznania, pl – polski):
```
./main -l pl -m models/ggml-large.bin -f samples/
input.wav >output.txt
```
a w pętli dla wszystkich plików w katalogu:
```
for f in /ścieżka/16k/*.wav; do ./main -t 8 -l pl
-m models/ggml-large.bin -f $f -otxt -ps -pp -pc;
done;
```

Zamiast *ścieżka* należy wpisać ścieżkę do katalogu z plikami dźwiękowymi w formacie 16 kHz. W przypadku innej procedury instalacji *./main* zamieniamy na *whisper*. Dostępne są także usługi przekładu. Dostęp programistyczny do Whispera w Pythonie:

```
import whisper
# whisper has multiple models that you can load as
per size and requirements
model = whisper.load_model("small.en")
# path to the audio file you want to transcribe
PATH = "audio.mp3"
result = model.transcribe(PATH)
print(result)
```

3.2.6. Porównanie wyników działania serwisów do transkrypcji automatycznej

Porównywanie usług transkrypcji automatycznej jest trudne, bo wymaga dużych, zróżnicowanych korpusów i złożonych metod pomiaru. W tym opracowaniu ograniczymy się do pokazania efektu działania wyżej wspomnianych serwisów na krótkiej, minutowej próbce nagrania z udziałem dwóch osób mówiących w języku polsku. Tekst pochodzi z nagrań realizowanych w jednym z projektów autora. Czas trwania transkrypcji ortograficznej:

- CLARIN-PL: 74 sekundy;
- CLARIN-WEBMAUS: 25 sekund dla potoku ASR > G2P > CHUNKER;
- CLARIN-WEBMAUS: 95 sekund dla potoku ASR > G2P > CHUNKER > MAUS > SUBTITLES > PHO2SYL > SD (rezultat w pliku *.eaf* wraz z transkrypcją ortograficzną i fonetyczną);
- WHISPER: 58 sekund (whisper --language pl „1min.wav" --output_format txt > output.txt --word_timestamps False)[11].

Każde nagranie musiało być najpierw przesłane, a potem odebrane z serwera, co mogło nieznacznie wydłużyć działanie serwisów, zależnie od dostępności i szybkości połączenia. Drugą zmienną wpływającą na czas jest zlecone zadanie: na CLARIN-PL i Whisper można zlecić jedynie transkrypcję ortograficzną, a na CLARIN-WEBMAUS mamy wybór między transkrypcją fonemiczną wedle reguł konwersji G2P i ortograficzną. Niemniej jednak wyniki są jakościowo wyraźnie różne i najbliższe transkrypcji ręcznej są teksty z serwisów Whisper i CLARIN-WEBMAUS, natomiast CLARIN-PL pominął 2/3 słów.

[11] Mimo wyłączenia funkcji podawania czasu dla wypowiedzi Whisper zwraca tekst ze znacznikami czasowymi.

Tabela 5. Wynik działania automatycznej transkrypcji CLARIN-PL.

jej pełne życia wolne przestrzenie ale trochę zoo rosa ukierunkowaną ostatecznie bardzo bardzo bardzo długi czas rolę samo istnienie ja myślałam kierunków studiów i chciał się mieści w głowie ale razem z jednym ruchem można skojarzenia

Tabela 6. Wynik działania automatycznej transkrypcji CLARIN-WEBMAUS (po eksporcie z pliku .eaf).

Jakie są te fale wzburzone energiczne piękne takie pełne życia takie wolne o przestrzeń właśnie wolność może w ten sposób metale wolne to jeszcze o karierze zawodowej nieokiełznane trochę ale są dlatego w sumie że dopiero ostatnio się ostatecznie Powiedz miałam bardzo bardzo bardzo długi czas problem z sam określeniem się i w sumie tyle zawodów o jakie Ja myślałam litery kierunków studiów jakie chciałam podjąć nie będą zdecydowana na nic konkretnego to też się nie mieści w głowie ale teraz już wiem i właśnie z tym kierunku sobie powoli te moje mamy rozwijam albo lubię bym może może dlatego tak przychodzą te Tesco wodą

Tabela 7. Wynik działania automatycznej transkrypcji Whisper po usunięciu znaczników czasowych.

Jakie są te fale? Zburzone, energiczne, ale piękne takie pełne życia. Takie coś w tym stylu. Wolne, o przestrzeń właśnie wolność. Może w ten sposób. A? Te fale, wolne. Przestrzeń, co jeszcze o tym? O karierze zawodowej mojej? Nie okiełznane trochę te fale są, dlatego w sumie, że dopiero ostatnio się ukierunkowałam ostatecznie. Miałem bardzo, bardzo, bardzo długi czas, problem z samokreśleniem się i w sumie tyle zawodów, o jakich ja myślałam i tyle kierunków studiów, jakie chciałam podjąć. Tak naprawdę nie będą zdecydowane o nic konkretnego, to już się nie mieści w głowie. Ale teraz już wiem i właśnie w tym kierunku sobie powoli te moje żalne rozwijam. A w ogóle lubię może może dla tego mi przechodzą te skojarzenie w wodą.

Tabela 8. Transkrypcja ręczna tekstu, który poddaliśmy także transkrypcji automatycznej.

jakie są te fale wzburzone energiczne ale piękne tekie pełne życia takie coś w tym stylu wolne przestrzeń właśnie wolna może jeszcze może w ten sposób te fale wolne przestrzeń co jeszcze o karierze zawodowej mojej nieokiełznane trochę te fale są dlatego w sumie że dopiero ostatnio się ukierunkowałam ostatecznie powiedzmy miałam bardzo bardzo bardzo długi czas problem z samookreśleniem się i w sumie tyle zawodów o jakich ja myślałam i tyle kierunków studiów jakie chciałam podjąć tak naprawdę nie będąc zdecydowaną na nic konkretnego to aż się nie mieści w głowie ale teraz już wiem i właśnie w tym moim kierunkusobie powoli te żagle moje rozwijam a w ogóle uwielbiam morze może dlatego mi tak przychodzą te skojarzenia z wodą

3.2.7. Oznaczenia stosowane w transkrypcji

Tabela znaków stosowanych dla zapisu języka polskiego w dwóch alfabetach fonetycznych: międzynarodowym i slawistycznym znajduje się na stronie Jagodzińskiego i stronie projektu PolFon, a odpowiedniki w alfabecie SAMPA i X-SAMPA na stronach internetowych Wellsa. Tabela 9 pokazuje zestawienie znaków stosowanych w czterech alfabetach fonetycznych dla języka polskiego: SAMPA, X-SAMPA, slawistycznym i międzynarodowym alfabecie fonetycznym wraz z komentarzami Jagodzińskiego i Juszczyka.

Tabela 9. Zestawienie znaków stosowanych w czterech alfabetach onetycznych z przykładami zapisu i komentarzami.

nr	S	XS	AS	IPA	zapis AS	zapis IPA	zapis ort.
1.	a	a	a	a	/mama/	/mama/	*mama*
2.	e	E	e	ɛ	/eva/	/ɛva/	*Ewa*
3.	o	O	o	ɔ	/rok/	/rɔk/	*rok*
4.	i	i	i	i	/ližba/	/liʤ͡ʒba/	*liczba*
5.	I	1	y	ɨ	/być/	/bɨʨ͡ɕ/	*być*
6.	u	u	u	u	/muχa/	/muxa/	*mucha*
7.		a_~	ą	ã	/šąsa/	/ʃãsa/	*szansa*
8.	e~	E_~	ę	ɛ̃	/męśki/	/mɛ̃sci/	*męski*
9.	o~	O_~	ǫ	ɔ̃	/vǫśki/	/vɔ̃sci/	*wąski*
10.		i_~	į	ĩ	/įstyŋkt/	/ĩstɨŋkt/	*instynkt*
11.		1_~	ẏ	ɨ̃	/rẏštok/	/rɨ̃ʃtɔk/	*rynsztok*
12.		u_~	ų	ũ	/kųšt/	/kũʃt/	*kunszt*
13.	j	j	i̯	j	/moi̯e/	/mɔjɛ/	*moje*
14.	w	w	u̯	w	/u̯oŋka/	/wɔŋka/	*łąka*
15.		5	ł	ł	/łoŋka/	/łɔŋka/	*łąka*
16.	l	l	l	l	/lala/	/lala/	*lala*
17.	r	r	r	r	/rura/	/rura/	*rura*
18.	m	m	m	m	/mama/	/mama/	*mama*
19.		m_j	ḿ	mʲ	/ḿešek/	/mʲɛʃɛk/	*mieszek*

nr	S	XS	AS	IPA	zapis AS	zapis IPA	zapis ort.
20.	n	n	n	n	/noga/	/nɔga/	*noga*
21.	n’	n_j	ń	ɲ̟	/koń/	/kɔɲ̟/	*koń*
22.	N	N	ŋ	ŋ	/peŋkać/	/pɛŋkat͡ɕ/	*pękać*
23.	p	p	p	p	/pačka/	/pa͡ʧka/	*paczka*
24.	b	b	b	b	/baj̟ka/	/bajka/	*bajka*
25.	f	f	f	f	/faj̟ka/	/fajka/	*fajka*
26.	v	v	v	v	/voda/	/vɔda/	*woda*
27.		p_j	ṕ	pʲ	/ṕesek/	/pʲɛsɛk/	*piesek*
28.		b_j	b́	bʲ	/b́au̯y/	/bʲawɨ/	*biały*
29.		f_j	f́	fʲ	/traf́a/	/trafʲa/	*trafia*
30.		v_j	v́	vʲ	/v́eʒ́eć/	/vʲɛd͡ʑɛt͡ɕ/	*wiedzieć*
31.	t	t_d	t	t̪	/tata/	/tata/	*tata*
32.	d	d_d	d	d̪	/rudy/	/rudɨ/	*rudy*
33.	ts	ts	c	t͡s	/ulica/	/ulit͡sa/	*ulica*
34.	dz	dz	ʒ	d͡z	/ʒvon/	/d͡zvɔn/	*dzwon*
35.	s	s	s	s	/ser/	/sɛr/	*ser*
36.	z	z	z	z	/zupa/	/zupa/	*zupa*
37.	tS	tS	č	ʧ	/čarny/	/ʧarnɨ/	*czarny*
38.	dZ	dZ	ǯ	d͡ʒ	/ǯem/	/d͡ʒɛm/	*dżem*
39.	S	S	š	ʃ	/škou̯a/	/ʃkɔwa/	*szkoła*

nr	S	XS	AS	IPA	zapis AS	zapis IPA	zapis ort.
40.	Z	Z	ž	ʒ	/žyće/	/ʒɨtɕɛ/	*życie*
41.	ts’	ts\	ć	tɕ͡	/ćastko/	/tɕ͡astkɔ/	*ciastko*
42.	dz’	dz\	ʒ́	dʑ͡	/ʒ́ivny/	/dʑ͡ivnɨ/	*dziwny*
43.	s’	s\	ś	ɕ	/śiny/	/ɕinɨ/	*siny*
44.	z’	z\	ź	ʑ	/źarno/	/ʑarnɔ/	*ziarno*
45.		c	ḱ	c	/ḱedy/	/cɛdɨ/	*kiedy*
46.		J\	ǵ	ɟ	/ǵońć/	/ɟɔɲtɕ͡/	*giąć*
47.		C	χ́	ç	/χ́iǵena/	/çiɟɛna/	*higiena*
48.		j\	ɣ́	ʝ	/ɣ́iǵena/	/ʝiɟɛna/	*higiena*
49.	k	k	k	k	/kot/	/kɔt/	*kot*
50.	g	g	g	g	/gazeta/	/gazɛta/	*gazeta*
51.	x	x	χ	x	/χak/	/xak/	*hak*
52.			ɣ	ɣ	/ɣak/	/ɣak/	*hak*
53.		h\	h	ɦ	/hak/	/ɦak/	*hak*
54.		h_j	ḣ	ɦʲ	/ḣiǵena/	/ɦʲiɟɛna/	*higiena*
55.		{	ä	æ	[i̯äi̯ko]	[jæjkɔ]	*jajko*
56.		e	ė	e	[ʒ́ėń]	[dʑ͡eɲ]	*dzień*
57.		o	ȯ	o	[ćȯća]	[tɕ͡otɕ͡a]	*ciocia*
58.		}	ü	ʉ	[jüźo]	[jʉʑɔ]	*Józio*
59.		w_0	u̯̣	w̯̣	[umysu̯̣]	[umɨsw̯̣]	*umysł*

nr	S	XS	AS	IPA	zapis AS	zapis IPA	zapis ort.
60.		5_0	ł̩	ɫ̩	[umysł̩]	[umɨsɫ̩]	*umysł*
61.		l_0	l̩	l̩	[myśl̩]	[mɨɕlˌ]	*myśl*
62.		L	l′	ʎ	[l′ista]	[ʎista]	*lista*
63.		r_0	r̩	r̩	[ṕatr̩]	[pʲatrˌ]	*wiatr*
64.		r_j	r′	rʲ	[baler′ina]	[balɛrʲina]	*balerina*
65.		m_0	m̩	m̩	[ṕism̩]	[pʲismˌ]	*pism*
66.		n_0	n̩	n̩	[ṕosn̩ka]	[pʲɔsn̩ka]	*piosnka*
67.		n	n̩	n͡	[soǹček]	[sɔn͡ʧɛk]	*sączek*
68.		n_J_0	ń̩	ɲ̩	[ṕeśń̩]	[pʲɛɛɲˌ]	*pieśń*
69.		j_~	i̯	j̃	[paɹsḱi]	[paj̃sci]	*pański*
70.		N_j	ŋ́	ŋʲ	[veŋ́ǵel]	[vɛŋʲɟɛl]	*węgiel*
71.		w_~	u̯	w̃	[vouɹsḱi]	[vɔw̃sci]	*wąski*
72.		p_j	ṕ	pʲ	[ṕić]	[pʲitɕ]	*pić*
73.		b_j	b́	bʲ	[b́ić]	[bʲitɕ]	*bić*
74.		f_j	f́	fʲ	[traf́ić]	[trafʲitɕ]	*trafić*
75.		v_j	v́	vʲ	[v́ić]	[vʲitɕ]	*wić*
76.		t	t̩	t͡	[t̩šy]	[t͡ʃɨ]	*trzy*
77.		d	d̩	d͡	[d̩ževo]	[d͡ʒɛvɔ]	*drzewo*
78.		t_j	t′	tʲ	[t′ik]	[tʲik]	*tik*
79.		d_j	d′	dʲ	[d′iva]	[dʲiva]	*diwa*

nr	S	XS	AS	IPA	zapis AS	zapis IPA	zapis ort.
80.		ts_j	c′	tsʲ	[c′ito]	[tsʲitɔ]	*cito*
81.		dz_j	ʒ′	dzʲ	[goʒ′illa]	[gɔd͡ʒʲilla]	*Godzilla*
82.		s_j	s′	sʲ	[s′inus]	[sʲinus]	*sinus*
83.		z_j	z′	zʲ	[z′in]	[zʲin]	*zin*
84.		tS_j	č′	ʧʲ	[č′ipsy]	[ʧʲipsɨ]	*chipsy*
85.		dZ_j	ǯ′	ʤʲ	[ǯ′iu̯sy]	[ʤʲiw̃sɨ]	*dżinsy*
86.		S_j	š′	ʃʲ	[š′in]	[ʃʲin]	*szin*
87.		Z_j	ž′	ʒʲ	[rež′im]	[rɛʒʲim]	*reżim*
88.		h	ḥ	h	[druḥ]	[druh]	*druh*
89.		R	R	R	[RoveR]	[RɔvɛR]	*rower*

S: SAMPA wg Wellsa, SAMPA oznacza *Speech Assessment Methods Phonetic Alphabet.*
XS: X-SAMPA wg Wellsa oraz strony https://tools.lgm.cl/xsampa.html.
X-SAMPA oznacza *Extended Speech Assessment Methods Phonetic Alphabet.*
AS: alfabet slawistyczny wg Jagodzińskiego.
IPA: międzynarodowy alfabet fonetyczny (*International Phonetic Alphabet)*

Zapis fonemiczny jest podany w nawiasach pochyłych (//) od 1. do 54., a fonetyczny w kwadratowych ([ ]), od 55. do 89.

Oznaczenia XSAMPA:
_~ nosowe
_j miękka
_d zębowe

_0 bezdźwięczny
\ oznacza środkowojęzykowe, ale h\ w XSAMPA oznacza spółgłoskę gardłową, szczelinową, dźwięczną.
Łączna wymowa t i s jako zwartoszczelinowej nie jest oznaczana w XSAMPA.

Uwagi Jagodzińskiego odnośnie AS:
1. Znak /a/ w AS oznacza samogłoskę niską, środkową, ale w IPA brakuje osobnego oznaczenia tej samogłoski.
2.-3. Znaki /e/ i /o/ są prostsze niż [ė] i [ȯ] lub [ë] i [ö]
5. Znak /y/ w AS jest spójny z ortografią.
7.-12. Samogłoski nosowe są dyftongami, ale są oznaczane za pomocą jednego znaku.
13. Patrz 2.
14. Patrz 2. i 3 (5).
15. Spółgłoska zębowa, sonorna. Patrz 3 (5).
16. Spółgłoska zazębowa sonorna.
17. Patrz 4 (7).
19. Patrz 5 (15).
21. Także IPA /ɲ/, patrz 6 (16).
27.-30. Patrz 5 (15).
31.-36. Patrz 7 (21).
37.-40. Patrz 8 (55).
45.-46. Patrz 9 (60).
47.-48. Patrz 3. (5) i 9 (60).
51.-54. Patrz 3. (5)
55.-58. Samogłoski podniesione występują pomiędzy spółgłoskami miękkimi.
59. Bezdźwięczny allofon /u̯/.
60. Bezdźwięczny allofon /ɬ/.
61. Bezdźwięczny allofon /l/.
62. Allofon /l/, patrz 6 (16).
63. Bezdźwięczny allofon /r/.
64. Allofon /r/.
65. Bezdźwięczny allofon /m/.
66. Bezdźwięczny allofon /n/.
67. Patrz 7 (21).

68. Także IPA [n̥], allofon /ń/.
69. Allofon /ń/, AS także [ñ].
70. Allofon /ŋ/ przed /ḱ ǵ/.
71. Także AS [ũ], IPA [ũ̯]. Patrz 1.
72.-75. Patrz 5 (15).
76.-77. Patrz 7 (21).
78.-87. Patrz 10 (61).
88. Patrz 3. (5)

Uwagi Juszczyka:
72. Bardziej czytelne oznaczenie w AS to [p’].
73. Bardziej czytelne oznaczenie w AS to [b’].
74. Bardziej czytelne oznaczenie w AS to [f’].
75. Bardziej czytelne oznaczenie w AS to [v’].
76. Zębowe t, d i n na phoible oznaczono za pomocą t̪, d̪, n̪, więc znaki t, d i n oznaczają w IPA dziąsłowe (alveolar) realizacje tych fonemów.
77. Jagodziński podaje tutaj znak równości pod spółgłoską, co IPA numer 675 i oznacza alveolarized.
89. Wariantywna wymowa R, czyli drżąca języczkowa, jak we francuskim.
Prawidłowe wyświetlenie znaków zapewnia krój Lucida Grande.

Należy zwrócić uwagę, że SAMPA została opracowana dla transkrypcji fonemicznej, a rozszerzona X-SAMPA dla fonetycznej. Jednakże automatyczna konwersja zapisu ortograficznego do fonemicznego lub fonetycznego niekoniecznie jest oparta na nagraniach, a raczej na regułach konwersji mówiących o tym, jakie znaki alfabetu łacińskiego dla oznaczenia fonemów lub głosek należy podmienić na znaki alfabetu fonetycznego. Transkrypcja fonetyczna, która miałaby pokazywać wymowę na nagraniu może być wykonana jedynie na podstawie nagrań, tak jak to zostało opisane w jedynym jak dotąd słowniku wariantywności wymowy polskiej (Madelska 2005). Transkrypcja oparta na konwersji znaków pokazuje jedynie wymowę zgodną z przyjętymi dla danego języka zasadami wymowy poprawnej. Zasady takiej transkrypcji dla języka polskiego przedstawia między innymi (Śledziński 2022).

3.3. Anotacja jednostek języka

Korpus transkrypcji jest anotowany także na poziomie składni i morfologii czy pragmatyki (przykład opracowania gramatycznego nagrań z korpusów grupy DIAGEST: Szczyszek 2013). W projekcie NARRACJE w korpusie nagrań dzieci i dorosłych wyodrębniono następujące poziomy opisu w warstwie leksykalno-składniowej:

- wypowiedzenia (podział toku narracyjnego na wypowiedzenia),
- schematy syntaktyczne zdań niezłożonych,
- schematy syntaktyczne zdań złożonych,
- części zdania,
- predykaty,
- argumenty,
- części mowy,
- wyrazy słowotwórczo podzielne,
- wyzyskane typy formantów (Karpiński i in. 2008).

Szczegółowe kryteria anotacji gramatycznej są przedstawione w publikacjach Szczyszka (2013) i zespołu DiaGest (Karpiński i in. 2008), więc tutaj ich nie powtarzamy.

3.3.1. Automatyczna anotacja jednostek języka

Obok ręcznej anotacji dostępne są narzędzia do automatycznej anotacji za pomocą urządzeń do przetwarzania języka naturalnego. Omówimy zasady działania usług konsorcjum CLARIN-PL, bo są przystosowane do języka polskiego, otwarte i darmowe. Narzędzia mają różny stopień trudności obsługi, czyli wymagają podstawowej lub zaawansowanej znajomości technologii NLP i umiejętności programowania. Niektóre są przeznaczone tylko dla deweloperów i informatyków. Wśród usług CLARIN-PL znajdziemy zarówno proste w obsłudze serwisy do analizy gramatyki, jak i narzędzia do klasyfikacji tekstów. Te ostatnie są *de facto* potokami przetwarzania danych językowych złożonych z niżej wymienionych narzędzi (głównie na Parserze, Tagerze WCRFT, lemmatyzatorze Morfeusz). Poniżej podajemy metody automatycznej anotacji tekstu wraz z ich angielskimi odpowiednikami. Usługi są podzielone wedle etapu przetwarzania danych:

1. Formowanie (skąd uzyskać materiał tekstowy?) Mowa, Anonimizer, Paragraphs, Punctuator, Speller, Sympsell, Tokenizer, Txtclean, Wordifier, Korpusomat,DSpace, CLARIN Cloud, KonText, Inforex.
2. Opracowanie (jak przygotować materiał?) Inforex, Korpusomat, WSD.
   2.1. Oznaczanie struktur składniowych: *parser,* czyli „przetwornik" zdań; Parser, Spejd.
   2.2. Oznaczanie części mowy: *POS tagger,* czyli *parts of speech tagger;* Tager WCRFT, TaKIPI.
   2.3. Oznaczanie morfemów i cech gramatycznych oraz wyznaczanua formy podstawowej wyrazów w tekście: *lemmatizer*, czyli lematyzator: Morfeusz*;* W przypadku języków niefleksyjnych, takich jak angielski, stosuje się *stemming*, czyli wyznaczanie tematów słów, które są niezmienne dla wszystkich form (polska definicja za Mykowiecka 2007, 69).
   2.4. Ujednoznacznianie: *Word Sense Disambiguation,* narzędzie oparte na Słowosieci: WoSeDon.
3. Analiza (jakie informacje można uzyskać z materiału?) (*topic modelling, word2vec, Latent Dirichlet Allocation, Latent Semantic Analysis, tf-idf, keyword analysis*).
   3.1. Rozpoznawanie nazw własnych, numerów, wyrażeń temporalnych i innych informacji: LiNER2, TermoPL, wyrażeń przestrzennych Spatial.
   3.2. Klasyfikacja tematyczna WiKNN.
   3.3. Grupowanie tekstów Websty, LEM.
   3.4. Analiza autorstwa : Websty, LEM, Topic.
   3.5. Porównywanie i przeglądanie zawartości korpusów KonText, Korpusomat, Inforex, Chronocorpus, Federated Content Search, VLO.
   3.6. Analiza cech gramatycznych tekstu Websty, LEM, Topic, ComCorp.
   3.7. Wydobywanie fraz charakterystycznych, jednostek wielowyrazowych i terminologii (*Named Entity Recognition)*: TermoPL, MeWeX i innych informacji z tekstu: Websty, LEM, Topic, Spatial, NER.
   3.8. Wykrywanie wyrazów emotywnych, czyli *sentiment analysis*: dla krótkich fragmentów tekstu – Sentemo i dla dłuższych –

Wydźwięk, wielojęzyczne wyznaczanie wydźwięku emocjonalnego – Multiemo.

3.9. Wydobywanie grup tematycznych: Topic i wielojęzyczne modelowanie tematyczne: TopicML (oparte na bibliotekach Gensim, Mallet).

3.10. Modelowanie tematyczne krótkich tekstów: Shortextopic (oparte na bibliotece BERTopic).

3.11. Wielojęzyczny system analizy podobieństwa tekstów: WebSim.

3.12. Analiza podobieństwa tekstów: Websty, Verbs (oparte na SuperMatrix (Broda i Piasecki 2008), Gensim, Mallet) i wielojęzyczny system analizy podobieństwa tekstów WebStyMLi.

3.13. Klasyfikacja oparta na analizie wektorów tekstów Word2vec.

4. Dyskusja (jak interpretować informacje uzyskane z materiału?) narzędzia poza CLARIN-PL.

Wspólna procedura korzystania z usług na stronach CLARIN-PL jest następująca:

1. wybierz serwis;
2. wczytaj tekst (pliki o rozszerzeniach *.doc*, *.docx*, *.pptx*, *.xlsx*, *.odt*, *.pdf*, *.html*, *.rtf* zostaną automatycznie przekonwertowane do tekstu), rozmiar tekstu jest ograniczony do 1 048 576 znaków; w przypadku korpusów tekstów wczytaj plik skompresowany o rozszerzeniu *.zip*;
3. analizuj (czas ładowania jest zależny od wielkości załadowanych plików);
4. pobierz wynik w formacie *.xml* albo przejrzyj wynik z wizualizacją, wykresem lub schematem zdania.

Poniżej przedstawiamy przykłady analiz próbki tekstu z transkrypcji ortograficznej z dialogu ORIGAMI (Tabela 9) przy pomocy *parsera* (Tabela 10) i *tagera* (Tabela 14) oraz tabele tagów (Tabela 11, Tabela 12, Tabela 13).

Tabela 10. Przykład fragmentu tekstu transkrypcji ortograficznej z dialogu ORIGAMI (50 wyrazów).

dolna są dwie figury dolna jest w kształcie takiego jakby talerza czegoś takiego jak to tylko że do góry jeszcze z rowkami czyli musisz i zy jest do tego są do tego użyte te spinacze każdy na każdym rogu są spięte te zagięcia spinaczami no tak musi być to zamknięte

Próbkę tekstu, którą pokazuje Tabela 10, poddano analizie przy pomocy *parsera* i otrzymano wynik w formacie CoNLL. Tabela 11 prezentuje wynik po przekształceniu w tabelę. W kolejnych kolumnach mamy numer porządkowy wyrazu w tekście (1), wyraz w formie tekstowej („dolna"), wyraz w formie podstawowej (po lemmatyzacji) („dolny"), część mowy (adj), cechy gramatyczne (adj, sg|nom|f|pos), numer wyrazu w strukturze zdania (2) i nazwę części zdania (pd). Widzimy także wyrazy zignorowane, czyli nierozpoznane, takie jak (26: „zy").
Tabela 10. Przykład fragmentu tekstu po analizie za pomocą parsera w formacie CoNLL (po przekształceniu w tabelę).

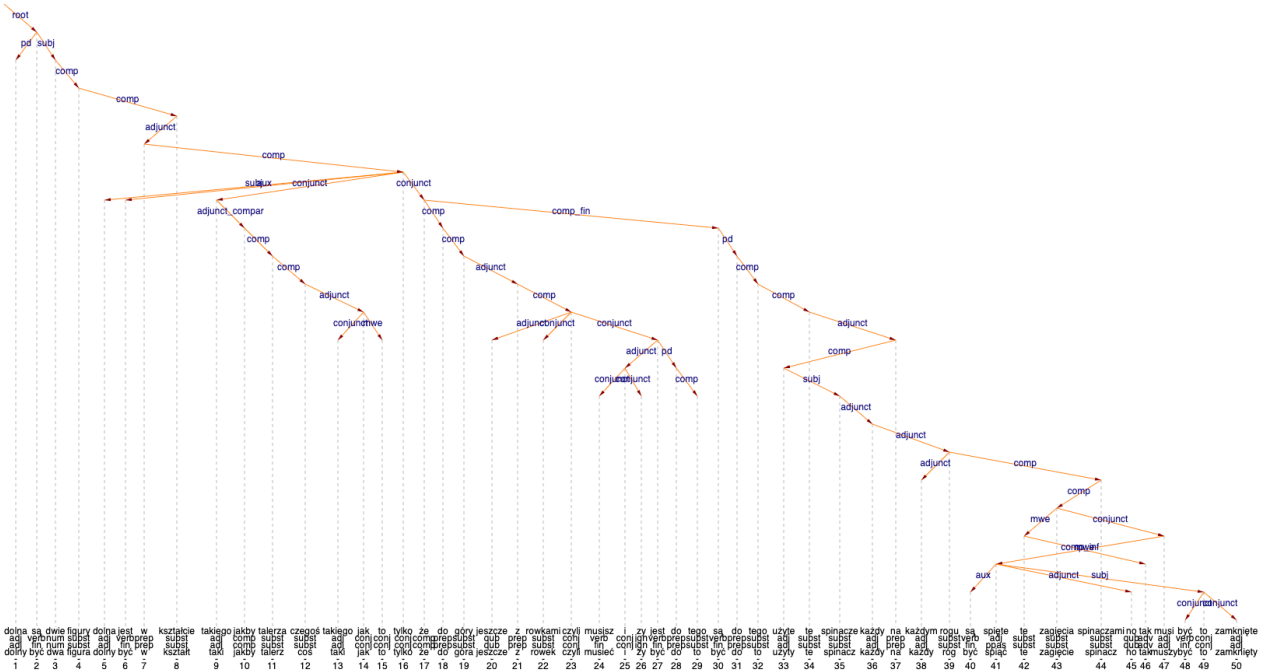
Tabela 11. Przykład fragmentu tekstu po analizie za pomocą parsera w formacie CoNLL (po przekształceniu w tabelę).

1	dolna	dolny	adj	adj	sg|nom|f|pos	2	pd
2	są	być	verb	fin	pl|ter|imperf	0	root
3	dwie	dwa	num	num	pl|nom|m1|rec	2	subj
4	figury	figura	subst	subst	sg|gen|f	3	comp
5	dolna	dolny	adj	adj	sg|nom|f|pos	6	pd
6	jest	być	verb	fin	sg|ter|imperf	4	adjunct_rc
7	w	w	prep	prep	acc|nwok	6	adjunct
8	kształcie	kształt	subst	subst	sg|loc|m3	4	comp
9	takiego	taki	adj	adj	sg|gen|m1|pos	16	conjunct
10	jakby	jakby	comp	comp	_	9	adjunct_compar
11	talerza	talerz	subst	subst	sg|gen|m3	10	comp
12	czegoś	coś	subst	subst	sg|gen|n	11	comp
13	takiego	taki	adj	adj	sg|gen|m1|pos	14	conjunct
14	jak	jak	conj	conj	_	12	adjunct
15	to	to	conj	conj	_	14	mwe
16	tylko	tylko	conj	conj	_	5	comp
17	że	że	comp	comp	_	16	conjunct
18	do	do	prep	prep	gen	17	comp

19	góry	góra	subst	subst	sg|gen|f	18	comp
20	jeszcze	jeszcze	qub	qub	_	23	adjunct
21	z	z	prep	prep	gen|nwok	19	adjunct
22	rowkami	rowek	subst	subst	pl|inst|m3	23	conjunct
23	czyli	czyli	conj	conj	_	21	comp
24	musisz	musieć	verb	fin	sg|sec|imperf	25	conjunct
25	i	i	conj	conj	_	27	adjunct
26	zy	zy	ign	ign	_	25	conjunct
27	jest	być	verb	fin	sg|ter|imperf	23	conjunct
28	do	do	prep	prep	gen	27	pd
29	tego	to	subst	subst	sg|gen|n	28	comp
30	są	być	verb	fin	pl|ter|imperf	17	comp_fin
31	do	do	prep	prep	gen	30	pd
32	tego	to	subst	subst	sg|gen|n	31	comp
33	użyte	użyty	adj	adj	sg|nom|n|pos	37	comp
34	te	te	subst	subst	sg|nom|n	32	comp
35	spinacze	spinacz	subst	subst	pl|nom|m3	33	subj
36	każdy	każdy	adj	adj	sg|nom|m1|pos	35	adjunct
37	na	na	prep	prep	acc	34	adjunct
38	każdym	każdy	adj	adj	sg|inst|m1|pos	39	adjunct

39	rogu	róg	subst	subst	sg|gen|m3	36	comp
40	są	być	verb	fin	pl|ter|imperf	41	aux
41	spięte	spiąć	adj	ppas	sg|nom|n|perf|aff	39	comp_inf
42	te	te	subst	subst	sg|nom|n	41	subj
43	zagięcia	zagięcie	subst	subst	sg|gen|n	42	comp
44	spinaczami	spinacz	subst	subst	pl|inst|m3	43	adjunct
45	no	no	qub	qub	_	48	adjunct
46	takt	takt	subst	subst	sg|nom|m3	44	comp
47	musi	muszy	adj	adj	pl|nom|m1|pos	46	adjunct
48	być	być	verb	inf	imperf	49	conjunct
49	to	to	conj	conj	_	47	comp_inf
50	zamknięte	zamknięty	adj	adj	sg|nom|n|pos	49	conjunct

Parser generuje także drzewo zdania w postaci grafu. Dane z formatu CCL i CoNLL możemy przetworzyć za pomocą skryptów w Pythonie. Tabela 12, Tabela 13 i Tabela 14 podają spis tagów stosowanych do oznaczania cech gramatycznych dla danych zwracanych przez usługi CLARIN-PL. Te same tagi są stosowane w Narodowym Korpusie Języka Polskiego (Przepiórkowski i in. 2012). Do wszystkich usług CLARIN jest także dostęp przez API w Pythonie, dzięki czemu możemy tworzyć potoki przetwarzania danych językowych, czyli łączyć usługi w ramach jednej aplikacji.



Grafika 15. Struktura zdania w postaci grafu otrzymana za pomocą parsera.

Tabela 12. Kategorie gramatyczne i ich oznaczenia stosowane w narzędziach do przetwarzania języka naturalnego (Przepiórkowski i in. 2012).

kategoria gramatyczna w języku angielskim	termin angielski	skrót	przykład polski	kategoria gramatyczna	termin polski
Number	singular	sg	oko	liczba	pojedyncza
Number	plural	pl	oczy	liczba	mnoga
Case	nominative	nom	woda	przypadek	mianownik
Case	genitive	gen	wody	przypadek	dopełniacz
Case	dative	dat	wodzie	przypadek	celownik
Case	accusative	acc	wodę	przypadek	biernik
Case	instrumental	inst	wodą	przypadek	narzędnik
Case	locative	loc	wodzie	przypadek	miejscownik
Case	vocative	voc	wodo	przypadek	wołacz
Gender	human masculine (virile)	m1	papież, kto, wujostwo	rodzaj	męski osobowy
Gender	animate masculine	m2	baranek, walc, babsztyl	rodzaj	męski zwierzęcy
Gender	inanimate masculine	m3	stół	rodzaj	męski rzeczowy

kategoria gramatyczna w języku angielskim	termin angielski	skrót	przykład polski	kategoria gramatyczna	termin polski
Gender	feminine	f	stuła	rodzaj	żeński
Gender	neuter	n	dziecko, okno, co, skrzypce, spodnie	rodzaj	nijaki
Person	first	pri	bredzę, my	osoba	pierwsza
Person	second	sec	bredzisz, wy	osoba	druga
Person	third	ter	bredzi, oni	osoba	trzecia
Degree	positive	pos	cudny	stopień	równy
Degree	comparative	com	cudniejszy	stopień	wyższy
Degree	superlative	sup	najcudniejszy	stopień	najwyższy
Aspect	imperfective	imperf	iść	aspekt	niedokonany
Aspect	perfective	perf	zajść	aspekt	dokonany
Negation	affirmative	aff	pisanie, czytanego	zanegowanie	niezanegowana
Negation	negative	neg	niepisanie, nieczytanego	zanegowanie	zanegowana
Accentability	accented (strong)	akc	jego, niego, tobie	akcentowość	główny
Accentability	non-accented (weak)	nakc	go, -ń, ci	akcentowość	akcentowana

kategoria gramatyczna w języku angielskim	termin angielski	skrót	przykład polski	kategoria gramatyczna	termin polski
Post-prepositionality	post-prepositional	praep	niego, -ń	poprzyimkowość	nieakcentowana
Post-prepositionality	non-post-prepositional	npraep	jego, go	poprzyimkowość	zneutralizowana
Accommodability	agreeing	congr	dwaj, pięcioma	związek	uzgadniająca
Accommodability	governing	rec	dwóch, dwu, pięciorgiem	związek	rządząca
Agglutination	non-agglutinative	nagl	niósł	aglutynacyjność	nieaglutynacyjny
Agglutination	agglutinative	agl	niosł-	aglutynacyjność	aglutynacyjny
Vocalicity	vocalic	wok	-em	wokaliczna	końcówka z samogłoską
Vocalicity	non-vocalic	nwok	-m	niewokaliczna	końcówka bez samogłoski
Fullstoppedness	with full stop	pun	tzn	kropkowalność	z następującą kropką
Fullstoppedness	without full stop	npun	wg	kropkowalność	bez następującej kropki

Tabela 13. Klasy gramatyczne i kategorie gramatyczne (Przepiórkowski i in. 2012).

	liczba	przypadek	rodzaj	przyrodz.	osoba	stopień	aspekt	zaneg.	akcent.	poprzyim.	akomod.	aglutyn.	wokal.	kropk.
rzeczownik	⊕	⊕	⊙	⊙										
rzeczownik deprecjatywny	⊙	⊕	⊙											
liczebnik główny	⊙	⊕	⊕								⊕			
przymiotnik	⊕	⊕	⊕			⊕								
przymiotnik przyprzym.														
przymiotnik poprzyim.		⊕												
przysłówek						⊕								
zaimek nietrzecioosobowy	⊙	⊕	⊕		⊙				⊕					
zaimek trzecioosobowy	⊕	⊕	⊕		⊙				⊕	⊕				
zaimek SIEBIE		⊕												
forma nieprzeszła	⊕				⊕		⊙							
forma przyszła BYĆ	⊕				⊕		⊙							
aglutynant BYĆ	⊕				⊕		⊙						⊕	
pseudoimiesłów	⊕		⊕				⊙					⊕		
rozkaźnik	⊕				⊕		⊙							
bezosobnik							⊙							
bezokolicznik							⊙							
im. przys. współczesny							⊙							
im. przys. uprzedni							⊙							
odsłownik	⊕	⊕	⊙				⊙	⊕						
im. przym. czynny	⊕	⊕	⊕				⊙	⊕						
im. przym. bierny	⊕	⊕	⊕				⊙	⊕						
winien	⊕		⊕				⊙							
predykatyw														
przyimek		⊙											⊕	
spójnik współrz.														
spójnik podrz.														
partykuła													⊕	
skrót														⊕
człon wyrażenia														
wykrzyknik														
znak interpunkcyjny														
ciało obce														

Tabela 14. Skróty nazw klas gramatycznych oraz ich formy hasłowe (Przepiórkowski i in. 2012).

fleksem	**skrót**	**forma podstawowa**	**przykład**
rzeczownik	*subst*	mianownik l. poj.	*doktor*
rzeczownik deprecjatywny	*depr*	mianownik l. poj. rzeczownika	*doktor*
liczebnik główny	*num*	mianownik rodz. m3	*pięć*, *dwa*
przymiotnik	*adj*	mianownik l. poj. rodzaju męskiego st. równego	*polski*
przymiotnik przyprzyim.	*adja*	mianownik l. poj. rodz. męskiego przymiotnika w st. równym	*polski*
przymiotnik poprzyim.	*adjp*	mianownik l. poj. rodz. męskiego przymiotnika w st. równym	*polski*
przysłówek	*adv*	forma stopnia równego	*dobrze*, *bardzo*
zaimek nietrzecioosobowy	*ppron12*	mianownik l. poj.	*ja*
zaimek trzecioosobowy	*ppron3*	mianownik l. poj.	*on*
zaimek SIEBIE	*siebie*	biernik	*siebie*
forma nieprzeszła	*fin*	bezokolicznik	*czytać*
forma przyszła BYĆ	*bedzie*	bezokolicznik	*być*
aglutynant BYĆ	*aglt*	bezokolicznik	*być*
pseudoimiesłów	*praet*	bezokolicznik	*czytać*
rozkaźnik	*impt*	bezokolicznik	*czytać*
bezosobnik	*imps*	bezokolicznik	*czytać*
bezokolicznik	*inf*	bezokolicznik	*czytać*
im. przys. współczesny	*pcon*	bezokolicznik	*czytać*
im. przys. uprzedni	*pant*	bezokolicznik	*czytać*
odsłownik	*ger*	bezokolicznik	*czytać*
im. przym. czynny	*pact*	bezokolicznik	*czytać*
im. przym. bierny	*ppas*	bezokolicznik	*czytać*
winien	*winien*	forma męska l. poj.	*winien*, *rad*

predykatyw	*pred*	jedyna forma fleksemu	*warto*
przyimek	*prep*	niewokaliczna forma fleksemu	*na, przez, w*
spójnik współrz.	*conj*	jedyna forma fleksemu	*oraz*
spójnik podrz.	*comp*	jedyna forma fleksemu	*że*
partykuła	*part*	jedyna forma fleksemu	*nie, -li, się*
skrót	*brev*	forma hasłowa rozwinięcia skrótu	*rok, i tak dalej*
człon wyrażenia	*frag*	jedyna forma fleksemu	*wskroś, dala*
wykrzyknik	*interj*	jedyna forma fleksemu	*laboga, pst*
znak interpunkcyjny	*interp*	jedyna forma fleksemu	*:, !, ?*
ciało obce	*xxx*	jedyna forma fleksemu	*wsio, revolutionibus*

Ten sam fragment tekstu po analizie za pomocą *tagera* jest zwracany w formacie CCL .xml, co pokazuje Tabela 15. Fragment tekstu po tagowaniu przy pomocy tagera CLARIN-PL w formacie CCL .xml. Fragment tekstu po tagowaniu przy pomocy tagera CLARIN-PL w formacie CCL .xml.

Tabela 15. Fragment tekstu po tagowaniu przy pomocy tagera CLARIN-PL w formacie CCL .xml.

```
<?xml version=”1.0” encoding=”UTF-8”?>
<!DOCTYPE chunkList SYSTEM „ccl.dtd”>
<chunkList>
<chunk id=”ch1” type=”p”>
<sentence id=”s1”>
<tok>
<orth>dolna</orth>
<lex disamb=”1”><base>dolny</base><ctag>adj:sg:nom:f:pos</ctag></lex>
</tok>
<tok>
<orth>są</orth>
<lex disamb=”1”><base>być</base><ctag>fin:pl:ter:imperf</ctag></lex>
</tok>
<tok>
<orth>dwie</orth>
<lex disamb=”1”><base>dwa</base><ctag>num:pl:nom:m1:rec</ctag></lex>
<lex disamb=”1”><base>dwa</base><ctag>num:pl:nom:m1:rec</ctag></lex>
<lex disamb=”1”><base>dwa</base><ctag>num:pl:nom:m1:rec</ctag></lex>
</tok>
<tok>
<orth>figury</orth>
<lex disamb=”1”><base>figura</base><ctag>subst:sg:gen:f</ctag></lex>
</tok>
<tok>
<orth>dolna</orth>
<lex disamb=”1”><base>dolny</base><ctag>adj:sg:nom:f:pos</ctag></lex>
</tok>
<tok>
<orth>jest</orth>
<lex disamb=”1”><base>być</base><ctag>fin:sg:ter:imperf</ctag></lex>
</tok>
<tok>
<orth>w</orth>
<lex disamb=”1”><base>w</base><ctag>prep:acc:nwok</ctag></lex>
</tok>
</sentence>
</chunk>
</chunkList>
```

Odpowiedniki wielojęzyczne wymienionych narzędzi CLARIN-PL i wiele innych znajdziemy także w bibliotekach do przetwarzania języka naturalnego w językach do analizy danych, czyli w Pythonie: *spacy*, *nltk*, *tmtoolkit*, *gensim*, *SparkNLP*, *transformers* lub *R: quanteda*, *topicmodels*, *korpus*, *text2vec*. Szczegółowe procedury korzystania z tych bibliotek są podane w podręcznikach do programowania (Deitel 2019; Deng 2014; Imai 2017; Imai i Bougher 2021; Jockers 2021; Osinga 2018; Lane, Howard i Hapke 2019; Van Atteveldt, Trilling i Calderón 2022).

Dla języków programowania przeznaczonych do tworzenia aplikacji dostępne są duże biblioteki NLP: *Treat* dla *Ruby*, *SparkNLP* dla *Scala*, *Prose* dla *Go*, *Natural Language* dla *Swift*, *ML.NET* dla C#. Przy wyborze biblioteki kierujemy się zakresem zadań z przetwarzania języka naturalnego i datą opublikowania ostatniej wersji (najnowsze są najlepsze) oraz wsparciem (forum, grupa dyskusyjna, podręczniki, dokumentacja). Tabela 16 przedstawia porównanie kilku wybranych bibliotek dla języka Python w zakresie zadań NLP. Tabela pochodzi z bloga jednej z firm świadczącej usługi NLP i została przełożona przez autora na język polski. Widzimy, że *spacy* jest bardziej wszechstronne od starszej biblioteki *nltk*, natomiast *CoreNLP* i *SparkNLP* pozwalają na wykonanie prawie wszystkich zadań. W tym miejscu należy jeszcze dodać, że do języka polskiego najlepiej przystosowana jest biblioteka *spacy*.

Tabela 16. Porównanie możliwości bibliotek w Pythonie.

zadanie	Spark NLP	spaCy	NLTK	CoreNLP
Wykrywanie zdań	TAK	TAK	TAK	TAK
Tokenizacja	TAK	TAK	TAK	TAK
Stemming	TAK	TAK	TAK	TAK
Lematyzacja	TAK	TAK	TAK	TAK
Tagowanie części mowy	TAK	TAK	TAK	TAK
Rozpoznawanie nazw, dat i miejsc	TAK	TAK	TAK	TAK
Analiza zależności gramatycznych	TAK	TAK	TAK	TAK
Dopasowywanie tekstu	TAK	TAK	NIE	TAK
Dopasowywanie daty	TAK	NIE	NIE	TAK
Dzielenie	TAK	TAK	TAK	TAK
Moduł sprawdzania pisowni	TAK	NIE	NIE	NIE
Analiza emocji	TAK	NIE	NIE	TAK

3.3.2. Anotacja pragmatyki

Korpusy multimodalne są także anotowane na poziomie pragmatyki. W projekcie NARRACJE na przykład wyróżniamy multimodalne akty dialogowe, a w ORIGAMI rekonstruujemy multimodalne modele mentalne znaczeń w instrukcyjnych aktach dialogowych. Akt dialogowy jest wielowarstwowy, bo jedna wypowiedź może realizować wiele funkcji dialogowych. W systemie DiaGest przyjęliśmy cztery wymiary i warstwy opisu (Karpiński i in. 2008; Karpiński 2009a):

1. *Task Action Control* (TAC): polecenia, ich objaśnienia, potwierdzenia ich zrozumienia oraz realizacji;
2. *Dialogue Flow Control* (DFC): przebieg dialogu związany z mechanizmem zabierania głosu;

3. *Information Transfer Management* (ITM): kierunek i charakter przepływu informacji;
4. *Approach Expression Marker* (AEM): nastawienie nadawcy do tematu wypowiedzi i rozmówcy.

System opisu multimodalnych aktów dialogowych wywodzi się z teorii aktów mowy (Austin 1993; Searle 1970) i taksonomii aktów dialogowych zespołu Bunta, zwanej *Dynamic Interpretation Theory (DIT)* (Bunt 2011). Charakterystyka działania za pomocą słów jest w multimodalnym akcie dialogowym uzupełniona o opis oddziaływania intonacji i zachowań niewerbalnych, a w szczególności gestów (Konat i Juszczyk 2015).

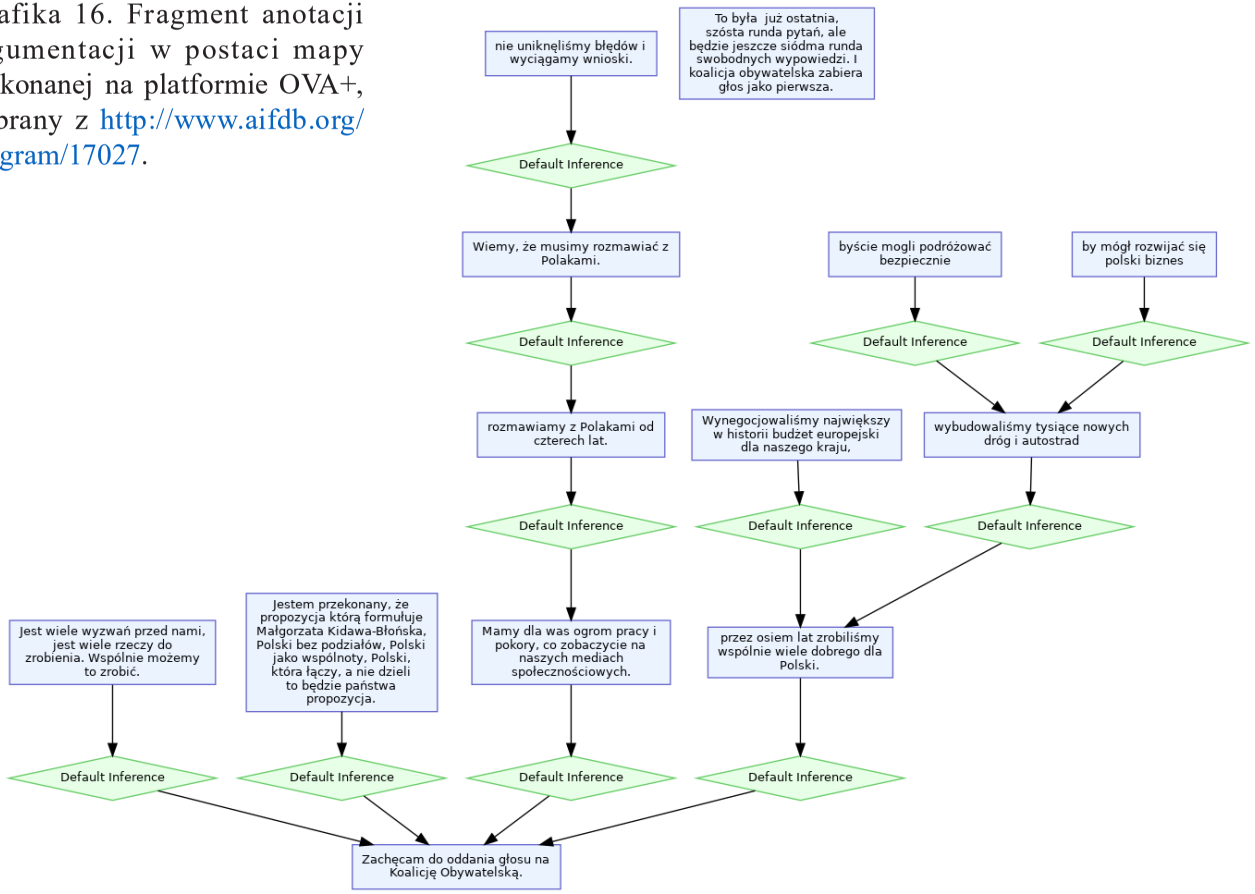
W przypadku sesji psychoterapeutycznych i coachingowych opis pragmatyczny dotyczy sposobu prowadzenia sesji przez psychoterapeutę lub coacha. Anotacja sesji służy wówczas do analizy przebiegu sesji zgodnie z wybranym scenariuszem czy protokołem (Kopp i Craw 1998; Tay 2013; Juszczyk 2017; Dunbar 2016; Stoltzfus 2012; Gupta i in. 2020). Anotowane są głównie wypowiedzi psychoterapeuty lub coacha jako kategorie pytań albo rodzaje interwencji, a rzadziej wypowiedzi uczestnika sesji jako etapy przechodzenia przez proces zmiany (Juszczyk 2017).

3.3.3. Anotacja argumentacji

W niektórych projektach celem analizy tekstu jest argumentacja, czyli rodzaje przesłanek, struktury argumentacyjne przesłanek i konkluzji, ocena wnioskowania czy poprawności argumentów i prawdziwości przesłanek (Szymanek 2012). W tym opracowaniu skupimy się jedynie na korpusach struktur argumentacyjnych. W języku potocznym, słowem „argument" określa się sąd wspierający inny sąd. W ramach Teorii Zakotwiczania Inferencji taki sąd nazwiemy przesłanką, natomiast argumentem jest „para zdań oznajmujących, z których jedno jest konkluzją (k), a drugie jest przesłanką jako odpowiedzią na pytanie DLACZEGO k?" (Budzyńska i Reed 2011). Innymi słowy przez argument rozumiemy tutaj całość składającą się z konkluzji i (co najmniej jednej) przesłanki. W Polsce korpusowe badania argumentacji, logosu,

etosu i patosu prowadzą między innymi zespoły Budzyńskiej i Konat (Budzyńska i in. 2015; Budzyńska, Konat i Koszowy 2016; Konat i in. 2016; Koszowy i in. 2022; Janier, Lawrence i Reed 2014). Anotacje argumentacji są wykonywane na internetowej platformie OVA+ i tam też są udostępniane korpusy anotacji. Przykładami korpusów na OVA+ są ręczne anotacje argumentacji dyskursu w polskich debatach parlamentarnych z 2019 roku w TVP i TVN (Juszczyk, Konat i Fabiszak 2022). Fragment korpusu anotacji argumentów pokazuje Grafika 16.

Grafika 16. Fragment anotacji argumentacji w postaci mapy wykonanej na platformie OVA+, pobrany z http://www.aifdb.org/diagram/17027.



3.3.4. Anotacja wyrażeń metaforycznych

Anotacje korpusowe dotyczą także wyrażeń wielowyrazowych, na przykład wyrażeń metafor konceptualnych. Stosowane w tym celu metody anotacji są oparte na założeniach teorii metafory konceptualnej (Lakoff i Johnson 2010; Lakoff 1992; 2011) i można je podzielić na dwa typy: introspekcyjne i korpusowe. Pierwsze są kontynuacją badań Lakoffa i Johnsona, którzy postępowali wedle następującej procedury (Valenzuela i Soriano 2005):

1. Wybierają przykłady wyrażeń metaforycznych.
2. Klasyfikują przykłady wyrażeń metaforycznych, czyli wyznaczają wspólne dla poszczególnych list wyrażeń metaforycznych odpowiednie metafory konceptualne.
3. Znajdują regularności dotyczące wyrażeń metaforycznych i metafor konceptualnych.
4. Proponują teorię języka, znaczenia i poznania w ramach językoznawstwa kognitywnego.

Opisana wyżej metoda jest popularna wśród badaczy metafor w wielu językach (np. Kövecses 2011; 2009; 2003; Idström, Piirainen i Falzett 2012), lecz przez niektórych jest uważana za wadliwą (Valenzuela i Soriano 2005; Gibbs 2011). Główny zarzut krytyków dotyczy postępowania badacza, który samodzielnie dobiera, przygotowuje przykłady wypowiedzi z wyrażeniami metaforycznymi i samodzielnie, często wedle nieopisanych kryteriów, wyznacza dla nich metaforę konceptualną (Gibbs 2011). Alternatywą jest *Metaphor Identification Procedure*, tj. procedura indentyfikacji wyrażeń metaforycznych i interpretacji tych wyrażeń przez wyznaczenie metafory konceptualnej. Obie procedury zostały opracowane przez zespoły Steena (Praggleja Group 2007; Steen 2009; Steen i in. 2010). Procedura składa się z czterech kroków:

1. Przeczytaj cały tekst-dyskurs, aby ustalić ogólny sposób rozumienia tego tekstu i jego ogólne znaczenie.
2. Ustal jednostki leksykalne w tekście-dyskursie.
   2.1. Dla każdej jednostki leksykalnej w tekście ustal jej znaczenie w kontekście, czyli jak odnosi się do bytu, relacji albo cechy w sytuacji przywołanej w tekście (znaczenie kontekstowe jednostki). Weź pod uwagę to, co występuje przed jednostką leksykalną i po niej.

2.2. Dla każdej jednostki leksykalnej określ, czy posiada bardziej podstawowe i współczesne znaczenie w innych kontekstach niż to, w którym wystąpiła. Znaczenia podstawowe zwykle są:
   2.2.1. bardziej konkretne (to, co wywołują, jest łatwiejsze do wyobrażenia, zobaczenia, usłyszenia, poczucia, powąchania i posmakowania);
   2.2.2. związane z działaniem za pomocą ciała (*bodily action*);
   2.2.3. bardziej precyzyjne (w przeciwieństwie do mętnych – *vague*);
   2.2.4. historycznie starsze.
3. Jeśli jednostka leksykalna ma bardziej podstawowe i współczesne znaczenie w kontekście innym niż to, w którym wystąpiła w badanym tekście ustal, czy znaczenie kontekstowe jednostki kontrastuje ze znaczeniem podstawowym i czy może być zrozumiane przez porównanie z nim.
4. Jeśli tak jest, to oznacz tę jednostkę leksykalną jako metaforyczną (przekład autora).

Procedura MIP została rozszerzona i dostosowana do wielu języków (Nacey i in. 2019), w tym do języka polskiego (Marhula i Rosiński 2014; 2019). Metaforyczne wyrażenia w języku polskim anotowaliśmy z wykorzystaniem modyfikowanej MIP w projekcie autora zwanym w skrócie MULTIMET (Juszczyk 2017). Główna modyfikacja MIP na potrzeby języka polskiego dotyczy włączenia do anotacji wyrazów tworzących wyrażenie metaforyczne i powiązanych składniowo z wyrazem w znaczeniu przenośnym, na przykład zamiast oznaczania samego wyrazu *twarda* w wyrażeniu *twarda zasada* oznaczamy także wyraz *zasada*. Dzięki takiej anotacji mamy możliwość wyznaczania metafor konceptualnych przy założeniu, że pierwszy wyraz w przykładowym wyrażeniu – *twarda* – odnosi się do domeny źródłowej, bo ma znaczenie konkretne, a drugi – *zasada* – do domeny docelowej, bo dotyczy problemu abstrakcyjnego. W ustalaniu znaczeń wyrazów korzystamy z dużego słownika języka polskiego, na przykład *Wielkiego słownika języka polskiego*, w którym przy wyrazie *twardy* znajdujemy definicję znaczenia oznaczonego tam jako pierwsze: „taki, którego trudno jest zgnieść, przeciąć lub odkształcić w inny sposób", a przy wyrazie *zasada*: „sposób postępowania w danych okolicznościach, w danej dziedzinie życia, odgórnie sankcjonowany przepisami prawa".

Na podstawie tych definicji możemy uznać, że *twardy* ma znaczenie konkretne, a *zasada* – abstrakcyjne, ale także zauważamy, że *twardy* odnośnie do *zasady* ma znaczenie kontekstowe zbliżone do znaczenia oznaczonego w WSJP jako czwarte: „taki, którego istnieniu ani treści nie można zaprzeczyć" i w tym znaczeniu według WSJP *twardy* dotyczy *dowodu.* Oczywiście rozróżnianie znaczeń jest zależne do założeń, jakie przyjęto w danym zespole anotatorów, więc w tym opracowaniu podajemy jedynie przykład anotacji wyrażeń metaforycznych z wykorzystaniem rozszerzonej procedury MIP i słownika WSJP. Nie zawsze w wyrażeniu metaforycznym pierwszy wyraz jest w znaczeniu konkretnym i przenośnym zarazem w kontekście drugiego wyrazu, który ma znaczenie abstrakcyjne ani nie zawsze znaczenie konkretne jest oznaczone w słowniku jako pierwsze. Jednakże oznaczanie całych wyrażeń metaforycznych pozwala na uzasadnienie, dlaczego dany wyraz jest w znaczeniu przenośnym, a w MIP byłby oznaczony jedynie jako wyraz związany z metaforą (*metaphorical lexical unit*). Szczegółowy opis tej wersji MIP znajduje się w publikacjach autora (Juszczyk i Kamasa 2016; Juszczyk 2017; Juszczyk, Konat i Fabiszak 2022). Podobne podejście zostało także później zastosowane w anotacjach przykładów z NKJP w projekcie *Cognitive and sociocultural analysis of metaphorical expressions in Polish texts (Cormetan)*, w którym dokonano także szczegółowej analizy składniowej wyrażeń metaforycznych (Hajnicz 2022).

MIP nie pozwala jednak na identyfikację metafory konceptualnej ani jej domen, dlatego zespół Steena opracował także procedurę interpretacji wyrażeń metaforycznych, której celem jest rzetelne, systematyczne wyznaczenie metafory konceptualnej (Steen 1999). Inne podejścia do korpusowych badań metafor konceptualnych prezentują badacze brytyjscy (Deignan i Semino 2010; Pitcher 2013; Deignan 2015; Cameron i Maslen 2010).

Do anotacji metafor nie potrzeba specjalistycznego oprogramowania, ale pomocne są specjalne słowniki semantyczne, na przykład USAS UCREL[12] (Piao i in. 2016), oraz platformy do współpracy anotatorów tekstu jak *e-Margin: A Collaborative Textual Annotation Tool* (Kehoe

[12] USAS to skrót od *Ucrel Semantic Analysis System,* gdzie UCREL to skrót od *University Centre for Computer Corpus Research on Language (UCREL) at Lancaster University* w Wielkiej Brytanii.

i Gee 2013), które wykorzystaliśmy w analizie argumentów i metafor używanych w dyskursie politycznym (Juszczyk, Konat i Fabiszak 2022). Anotacje w tym małym korpusie nie obejmują obrazu, ale materiał jest multimodalny, bo powstał na bazie transkryptów debat politycznych realizowanych na żywo w telewizji polskiej (TVP i TVN) i amerykańskiej. Zaletą *e-Margin* jest możliwość anotacji bezpośrednio na tekście za pomocą kolorowych oznaczeń, dodawania komentarzy i tagów oraz linków do słownika. Fragment tekstu anotowanego w *e-Margin* pokazuje Grafika 17 (zasłonięto jedynie dane anotatora). Niestety *e-Margin* nie pozwala na publikację anotacji, ale jest możliwość pobrania korpusu tekstów anotacji w formacie .xml w celu dalszej analizy w językach R lub Python.

Grafika 17. Fragment anotacji wyrażeń metaforycznych na platformie e-Margin.

3.4. Anotacja zachowań komunikacyjnych

Badania komunikacji międzyludzkiej wymagają anotacji, czyli oznaczania zachowań komunikacyjnych, które mają być przeznaczone do późniejszych analiz jakościowych i ilościowych. Podobnie jak w przypadku planowania nagrań wybór zachowań i systemu anotacji jest zależny od celów badań. Metody anotacji dzielimy na wybiórcze i całościowe, czyli albo anotujemy jedynie wybrane zachowania, albo wszystkie, jakie są zarejestrowane. W obu typach konieczne są kryteria anotacji, które służą do klasyfikacji zachowań. Najczęściej, obok transkrypcji, a w ELAN-ie pod transkrypcją, anotowane są ruchy rąk lub same gesty oraz ekspresje mimiczne emocji, a rzadziej ruchy głowy (Kousidis i in. 2013; Drewes i in. 2020), nóg czy całego ciała (Lausberg i in. 1988; Lausberg, Wietersheim i Feiereis 1996).

Kolejna decyzja, jaką podejmujemy w planowaniu anotacji, dotyczy formy nagrania: anotujemy filmy nieme i bez transkrypcji czy filmy z dźwiękiem z transkrypcją lub bez. W obu formach badań anotacja jest systematyczna, bo oznaczane są podobne do siebie zachowania komunikacyjne. Analizy takich anotacji są zarówno ilościowe, jak i jakościowe.

Pierwsza forma pozwala na kierowanie się kryteriami związanymi tylko z tym, co jest widoczne w kadrze filmu, a zatem pomijamy w trakcie takiej anotacji to, co mówią uczestnicy. Takie podejście jest stosowane głównie w badaniach behawioralnych, psychologicznych i neuropsychologicznych (Lausberg 2013; 2019), gdzie na pierwszy plan wysuwa się natura ruchu rąk lub innych części ciała w różnych sytuacjach, a analiza języka jest drugorzędna. Istotne są bowiem cechy ruchu zależne od procesów poznawczych i zaburzeń umysłowych. Uczestnicy takich badań reagują na określone polecenia słowne i wykonują zadania, na podstawie których są diagnozowani, lub biorą udział w sesjach psychoterapeutycznych, gdzie z kolei analizowana jest interakcja z psychoterapeutą.

Druga forma nagrań, czyli filmy z dźwiękiem z transkrypcją lub bez, są częściej stosowane w badaniach językoznawczych w celu dokumentacji sposobów wyrażania się członków danej wspólnoty kulturowej lub grupy wiekowej czy zawodowej. W niektórych językoznawczych badaniach komunikacji filmy są nieme na wstępnym etapie anotacji,

lecz wybrane próbki zachowań komunikacyjnych są uzupełniane tekstem wypowiedzi uczestników na etapie analizy i interpretacji. Próbki gestów, na przykład, są przedstawiane w publikacjach wraz ze słowami, przy których się pojawiają (Bressem 2021; Ladewig 2020; Jarmołowicz-Nowikow 2019; Antas 2013; Załazińska 2006).

3.4.1. Systemy anotacji ruchów rąk

Systemy anotacji zachowań komunikacyjnych najczęściej dotyczą szczegółowego opisu ruchów rąk, a zwłaszcza gestów, które są uważane za nośniki znaczeń (McNeill 1992; Kendon 2004). Badacze ruchów rąk opracowali kilkanaście systemów różniących się precyzją opisu, liczbą etykiet, procedurą uzgadniania anotacji i kontekstem zastosowań naukowych. Spróbujemy podzielić klasyfikacje i opisy ruchów rąk pod względem kontekstu badań:

- ruchy rąk jako zachowania niewerbalne:
  - klasyfikacja funkcji gestów (Ekman i Friesen 1969)[13];
  - KINEZYKA: ogólna typologia zachowań komunikacyjnych człowieka z systemem cech wzorowanym na fonologii (Birdwhistell 1970; Jolly 2000)
  - SYNERGOLOGIA: uniwersalny system opisu znaczeń ruchu całego ciała (Turchet 2009; 2006);
- ruchy rąk jako zachowania werbalne, czyli powiązane z systemem języka:
  - klasyfikacja funkcji gestów (McNeill 1992);
- ruchy rąk w powiązaniu z pragmatyką i prozodią, czyli systemy multimodalne:
  - *The MultiModal MultiDimensional (M3D) labeling system*[14] (Gibert i in. 2020; Rohrer, Delais-Roussarie i Prieto 2023);

[13] Klasyfikacje Ekmana i Friesena oraz McNeilla zawiera podobne etykiety i kryteria podziału, lecz twórcy pierwszej badali szeroko rozumianą komunikację niewerbalną i wyrażanie emocji, a twórca drugiej – system językowy. Dlatego w pierwszej wyróżnili autoadaptatory (gesty autodotyku), których nie wyróżnia McNeill.

[14] *Wielokanałowy i wielowymiarowy system anotacji* lub *Multimodalny i wielowymiarowy system anotacji* [przekład autora].

- *Prosodic and Gestural Entrainment in Conversational Interaction Across Diverse Languages*[15] *PAGE GAS (Gesture Annotation Scheme)*[16] (Karpiński, Jarmołowicz-Nowikow i Czoska 2015);
- *Linguistic Annotation System for Gestures (LASG)* (Bressem, Ladewig i Muller 2013);
- *Conversational Gesture Transcription system*[17] (CoGesT) (Trippel i in. 2004);
- *MuMin coding system*, gdzie *MuMin* to *MUltiModal Interfaces*[18] (Allwood i in. 2007);

- ruchy rąk opisywane asemantycznie w kategoriach geometrii ruchu (Martell 2002)[19];
- ruchy rąk osób niesłyszących i słyszących (Klima, Bellugi i Battison 1979; Kendon 2004; Stokoe 1960);
- ruchy rąk jako przedjęzykowe zachowania komunikacyjne:
  - w ewolucji człowieka (Kendon, Sebeok i Umiker-Sebeok 1981; Corballis 2013);
  - w rozwoju dzieci (Morgenstern i Goldin-Meadow 2022);
- ruchy rąk w kontekście różnic między naczelnymi – małpami a ludźmi (Tomasello i Call 2019);
- ruchy rąk jako zachowania zależne od funkcjonowania mózgu (Lausberg 2019);
- ruchy rąk i innych części ciała jako test diagnostyczny różnych chorób i zaburzeń: *BewegungsAnalyse Skalen und Test (BAST)*[20] (Lausberg, Wietersheim i Feiereis 1996).

Powyższa lista nie jest wyczerpująca, a podział ma jedynie cel dydaktyczny. Tutaj ograniczymy się do porównania dwóch systemów. Pierwszy – NEUROGES – dominuje we wcześniej wspomnianych badaniach behawioralnych i neurologicznych (Lausberg 2013; 2019), a drugi – LASG – w lingwistycznych (Bressem, Ladewig i Muller 2013).

[15] *Dostrojenie prozodyczne i gestykulacyjne w interakcji konwersacyjnej w różnych językach* [przekład autora].

[16] *Schemat Anotacji Gestów* [przekład autora].

[17] *System Transkrypcji Gestów Konwersacyjnych* [przekład autora].

[18] *System kodowania Interfejsów Multimodalnych* [przekład autora].

[19] W tym kontekście toczą się także badania na potrzeby automatycznej anotacji i rozpoznawania ruchów rąk, których charakterystykę w niniejszym opracowaniu pomijamy.

[20] *Skale i testy analizy ruchu* [przekład autora].

3.4.1.1. NEUROGES

Nazwa NEUROGES nawiązuje do celu badań behawioralnych, w których testowane hipotezy dotyczą zależności pomiędzy funkcjonowaniem i budową systemu nerwowego (NEURO z ang. *neuronal*) a rodzajami ruchów rąk (GES z ang. *gesture*). Szczegółowy opis systemu jest przedstawiony w dwóch książkach (Lausberg 2013; 2019). Jego autorami są prof. Hedda Lausberg z Niemieckiej Sportowej Szkoły Wyższej w Kolonii i Han Slojtes z Instytutu Maxa Plancka w Nijmengen. Pierwsza jest neuropsychiatrą i neuropsychologiem, a drugi jest współautorem programu do anotacji ELAN.

System składa się z 55 etykiet umieszczanych na 7 podstawowych typach warstw oraz 56 etykiet dodatkowych na 7 dodatkowych typach warstw. Anotowane są wszystkie ruchy obu rąk z podziałem na jedno- i dwuręczne. Filmy są oznaczane w ELAN-ie przy pomocy specjalnego szablonu ze wszystkimi etykietami i typami warstw. Kryteria anotacji są nazywane kinetycznymi, bo opisują tylko cechy ruchu rąk i ewentualnie nóg, więc anotowane filmy są pozbawione dźwięku i transkrypcji. Istotnym etapem anotacji w systemie NEUROGES jest procedura ustalania zgodności anotatorów dla wybranych próbek korpusu. Procedura składa się z 5 kroków:

1. Autorzy systemu zalecają wykonanie anotacji przez dwóch anotatorów dla jednej czwartej nagrań dla każdego uczestnika badań.
2. Następnie mierzy się zgodność anotacji.
3. Jeśli zgodność jest niska, anotatorzy porównują swoje anotacje i doprecyzowują kryteria.
4. Anotacje są poprawiane i ich zgodność jest znów mierzona.
5. Jeśli zgodność jest wysoka, anotatorzy przechodzą do anotacji pozostałych części nagrań.

Zgodność anotacji jest mierzona za pomocą dwóch miar: miarą pokrycia anotacji oraz miarą zgodności anotacji Cohena, zwaną też *kappą*. Miara pokrycia anotacji określa stopień, w jakim anotacje tych samych ruchów rąk, ale wykonane przez różnych anotatorów, pokrywają się czasowo na warstwie pierwszego typu – *Activation*. Oczekiwana wartość wyniku pomiaru pokrycia to co najmniej 80%, a im wyższa, tym lepsza. Anotacje z warstwy *Activation* są kopiowane do następnych warstw, więc wynik pomiaru pokrycia ma wpływ

na wynik pomiaru zgodności na pozostałych warstwach. Druga miara wskazuje prawdopodobieństwo, z jakim należy przyjąć, że anotacje nie są losowe. Do pomiaru zgodności etykiet brane są tylko te, których pokrycie jest większe niż 60%. Zgodność *kappa* Cohena jest określana dla każdej etykiety każdej warstwy oraz globalnie dla każdej warstwy, każdego nagrania i całego korpusu. Oczekiwane wartości zgodności mieszczą się w przedziale 0.4-0.9, zależnie od typu warstwy i etykiet (Lausberg i Slojtes 2015). Procedura pomiaru zgodności ma na celu zachowanie rzetelności i powtarzalności anotacji i badań. Wysoka zgodność i szczegółowe kryteria anotacji pozwalają innym badaczom na powtórzenie anotacji na tych samych lub innych próbkach nagrań i przy zachowaniu zbliżonych warunków badań – otrzymanie wysokiej zgodności anotacji oraz tych samych wyników badań. Wymogi wysokiej zgodności obejmują przede wszystkim badania eksperymentalne i diagnostyczne, a rzadziej językoznawcze. Samo mierzenie zgodności anotacji wykonujemy w ELAN-ie za pomocą wbudowanych w nim funkcji *Calculate Inter-annotator Reliability.* Algorytm mierzenia zgodności został pierwotnie opracowany w języku MATLAB (Holle i Rein 2015), a następnie zaimplementowany w języku Java w ELAN-ie, co ułatwia korzystnie z tych pomiarów. Algorytm jest modyfikacją miary Cohena Kappa, bo uwzględnia wspomniany wcześniej wskaźnik pokrycia anotacji i wyznacza zgodność *kappa* dla wielu, a nie tylko dwóch, jak u Cohena, kategorii (etykiet). Wyliczanie zgodności *kappa* w innych programach (np. w EXCELU, SPSS czy bibliotekach do analizy statystycznej Pythona *scikit-learn*) daje nieco inne wyniki, więc autorzy NEUROGES zalecają używanie w tym celu ELAN-a w najnowszej wersji (5.0 lub wyższa).

3.4.1.2. LASG

Drugi system anotacji nosi nazwę *Linguistic Annotation System for Gestures*, czyli System Lingwistycznej Anotacji Gestów. Anotowanymi jednostkami są gesty zdefiniowane zgodnie z założeniem Kendona i McNeilla jako najwyraźniejszy ruch ręki niosący znaczenie językowe (McNeill 1992; Kendon 2004). Podobnie jak NEUROGES system LASG składa się z bardzo wielu warstw (19) i etykiet (w sumie 231).

W odróżnieniu od NEUROGES w LASG anotacji podlegają jedynie wybrane ruchy rąk – gesty i niekoniecznie wszystkie gesty w całym korpusie, a jedynie te, które są istotne dla danych badań ze względu na ich fazy, formy i funkcje. Autorki systemu (Bressem, Ladewig i Muller 2013) opierają się na klasyfikacji podstawowych funkcji gestów według Kendona i McNeilla, którzy wyróżnili 5 kategorii:

- uderzeniowe (zwane też rytmicznymi, *beats*);
- wskazujące (zwane też deiktycznymi, *deictics*);
- obrazujące (zwane też ikonicznymi, *iconics*);
- metaforyczne (*metaphorics*);
- emblematyczne (*emblems*).

Liczne etykiety widoczne w szablonie LASG określają zarówno cechy ruchu rąk, konfiguracje palców, kształt dłoni, kierunki ruchu, fazy gestów, jak i funkcje gestów, pragmatyczne i semantyczne, oraz relacje z jednostkami mowy i części mowy dla wyrazów, które z badanymi gestami współwystępują. Odmienny od systemu NEUROGES jest w LASG sposób uzgadniania anotacji, bo w publikacjach nie znajdujemy wyników pomiaru zgodności. Procedura uzgadniania anotacji składa się z 3 etapów:

1. co najmniej dwóch anotatorów oznacza gesty części nagrań (np. 10% korpusu);
2. anotatorzy porównują anotacje i jeśli trafiają na różnice w oznaczonych etykietach, to doprecyzowują kryteria;
3. po rozstrzygnięciu anotacji przykładów spornych przechodzą do anotacji pozostałych fragmentów korpusu.

Językoznawcy badający gesty uważają takie postępowanie za rzetelne i, podobnie jak społeczność stosująca NEUROGES, zakładają, że tak przeprowadzone anotacje i badania są powtarzalne, bo kryteria anotacji są szczegółowo opisane w licznych publikacjach.

W niektórych projektach i publikacjach stosowane są systemy oparte na wspomnianej typologii gestów McNeilla czy Kendona (McNeill 1992; Kendon 2004), ale bez podawania wszystkich szczegółowych kryteriów klasyfikacji. Wówczas badacze przyjmują, że system anotacji jest otwarty i zależnie od celu badań i materiału mogą dodawać etykiety i kryteria, które pozwalają na precyzyjny opis obserwowanych zachowań. Takie podejście jest stosowane w badaniach dokumentacyjnych.

3.5. ELAN

W ostatnich dekadach powstało wiele programów do analizy nagrań mowy i filmów, ale większość z nich przestała być używana[21], a standardem wśród badaczy języka i komunikacji stał się ELAN. Program jest darmowy i aktualizowany, dostępny dla MS Windows, Linux i MacOS. Nie opisujemy tutaj wszystkich funkcji ELAN-a, bo instrukcja tego programu liczy ponad 500 stron, a wraz z aktualizacją jest uzupełniania, ale zwracamy uwagę na kilka wygodnych rozwiązań.

3.5.1. Tryby pracy programu

W ELAN-ie mamy 5 trybów pracy: *Annotation, Media Synchronization, Transcription, Segmentation* i *Interlinearilization mode.* Omówimy cztery z nich, bo są związane z anotacją ruchu. Zaczniemy od *Media Synchronization*, ponieważ jeśli mamy dwa nagrania, na przykład osobno dźwięk i film, albo dwa filmy z tej samej sesji nagrań, które różnią się czasem inicjacji nagrywania, to musimy je przed anotacją zsynchronizować. Najpierw na każdym nagraniu znajdujemy klatkę lub dźwięk (punkt startowy), które są identyczne i względem których będziemy synchronizować nagrania. Następnie wybieramy nagranie, które będzie punktem odniesienia dla pozostałych nagrań, a potem ustalamy różnice (*offset*) pomiędzy punktem startowym pierwszego nagrania a punktem startowym innych nagrań. Jeśli punkty się zgadzają, to klikamy *Aplly Current Offsets*. Na końcu wracamy do trybu *Annotation*, gdzie sprawdzamy, czy nagrania są zsynchronizowane. Prostsza metoda synchronizacji to odcinanie początków nagrań, które uznajemy za nieistotne, na przykład od początku do sygnału startowego. Dzięki tej metodzie pominiemy synchronizację w ELAN-ie, która przy wielu nagraniach zajmie bardzo dużo czasu. Pozostałe tryby są wykorzystywane do oznaczania ruchów etykietami lub do transkrypcji mowy. Tekst transkrypcji i tagi anotacji umieszczamy na osobnych warstwach jednego pliku *.eaf.* Anotacja fragmentów tekstu, ruchów rąk czy gestów albo innych zachowań składa się z 4 etapów:

[21] Wśród programów, które przestały być aktualizowane i używane, są między innymi Wavesurfer i Anvil.

1. segmentacja,
2. korekta segmentacji,
3. anotacja segmentów,
4. korekta anotacji.

3.5.1.1. Segmentacja

Segmentacja polega na wyznaczaniu granic czasowych dla oglądanych ruchów. W ELAN-ie ten etap wykonujemy w trybie *Segmentation (MENU > Options > Segmentation mode)*, w którym odtwarzamy film i naciskamy *enter* na początku i na końcu ruchu. W celu dokładniejszej segmentacji zwalniamy odtwarzanie filmu do 50% i ustawiamy spóźnienie reakcji na przycisk *enter* o 300 ms (tryb *Delayed*). W ten sposób niwelujemy błąd, jaki zwykle popełniamy, bo od zauważenia początku lub końca ruchu do naciśnięcia przycisku mija około 300 ms.

3.5.1.2. Korekta segmentacji

W drugim etapie poprawiamy segmentację w trybie *Annotation*. Przesunięcia segmentu anotacji dokonujemy przez zaznaczenie segmentu[22] anotacji i naciśnięcie przycisku *alt*, a potem przytrzymujemy lewy przycisk myszy, by oznaczyć segment anotacji zielonym kolorem i przesuwamy ten segment kursorem myszy w lewo lub w prawo. Granice segmentu anotacji zmieniamy podobnie: najeżdżamy na początek lub koniec segmentu anotacji i naciskamy *alt*, a potem znów, trzymając lewy przycisk myszy, przesuwamy granicę czasową segmentu anotacji w lewo lub w prawo, zależnie od obserwowanego ruchu na filmie. Gra-

[22] W niniejszym opracowaniu przyjęto, że polskim odpowiednikiem angielskiego *annotation* jest termin „anotacja". Jednakże liczba mnoga tego terminu to anotacje, a dopełniacz liczby pojedynczej i mnogiej to „anotacji". Dlatego przez anotacje należy rozumieć uporządkowany w czasie zbiór anotacji zapisany w pliku lub korpusie, czyli „plik anotacji" (plik pierwotnie oznacza także zbiór, np. dokumentów), a w przypadku pojedynczej anotacji używamy terminu „segment anotacji", który jest oznaczeniem fragmentu nagrania z określonym w ELAN-ie początkiem i końcem oraz etykietą. Początek i koniec nazywamy tutaj „granicami czasowymi segmentu anotacji".

fika 18 pokazuje podgląd korekty granic czasowych segmentu anotacji: wydłużenie i przesunięcie segmentu na inną warstwę. Możliwe jest także przesuwanie segmentu na tej samej warstwie.

Grafika 18. Korekta granic czasowych segmentu anotacji: po lewej widzimy zmianę długości segmentu, a po prawej zmianę pozycji segmentu anotacji, czyli przeniesienie na górną warstwę.

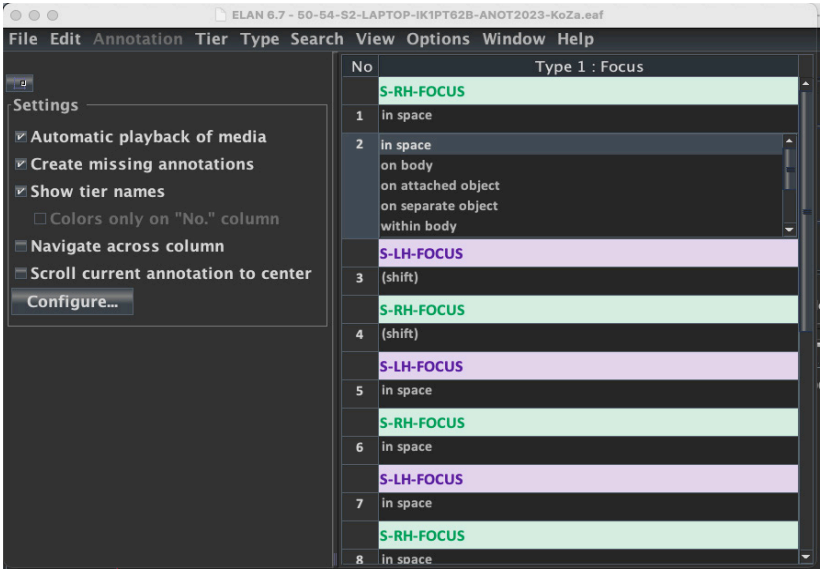
3.5.1.3. Anotacja segmentów

Trzeci etap anotacji przeprowadzamy także w trybie *Annotation*, w którym oglądamy wybrany segment i wybieramy odpowiednią dla niego etykietę na wybranej warstwie. Anotację tworzymy w trybie segmentacji albo anotacji. W celu utworzenia anotacji w trybie *Annotation* najeżdżamy na fragment warstwy, zaznaczamy segment, mając wciśnięty lewy przycisk myszy, następnie klikamy dwa razy lewym przyciskiem myszy, a potem wybieramy z menu odpowiednią etykietę (jeśli pracujemy w szablonie, przykład etykiet podanych w menu pokazuje Grafika 21) albo wpisujemy tekst lub etykietę. Inna metoda polega na zaznaczeniu segmentu i po wciśnięciu prawego przycisku myszy wybieramy *New annotation Here.*

3.5.1.4. Korekta anotacji i tryb transkrypcji

Ostatni etap to poprawki anotacji wedle kryteriów i konsultacji z innym doświadczonym annotatorem. Tryb *Transcription* jest także przydatny w anotacji, bo umożliwia szybkie przechodzenie z oznaczanego ruchu do następnego. Anotacje w tym trybie są widoczne w pionie, a film wyświetla się po lewej stronie ekranu. Klikamy na segment, fragment filmu oznaczony jako ten segment odtwarza się automatycznie, wybieramy odpowiednią etykietę i wciskamy *enter*,

kursor przechodzi do następnego segmentu i kolejny fragment filmu znów odtwarza się automatycznie. Do opisu poprawek przydatny jest także zakładka *Comments* opisana w sekcji 3.5.11. Grafika 19 pokazuje pracę w trybie transkrypcji.



Grafika 19. Tryb transkrypcji w ELANie (nie pokazano podglądu filmu, który byłby po lewej stronie).

3.5.2. Odtwarzanie fragmentu filmu w pętli

Zaznaczony segment odtwarzamy w pętli, aby zobaczyć kilka razy ten sam fragment. Pętlę aktywujemy przez zaznaczenie opcji *Loop Mode.*

3.5.3. Powiększanie fragmentu filmu, etykiet i segmentów anotacji

W trakcie anotacji możemy powiększyć (1) fragment filmu, (2) litery etykiet, (3) segmenty anotacji, (4) litery etykiet i warstw.

1. Klikamy prawym przyciskiem myszy na filmie i wybieramy *Zoom*, a potem procent powiększenia.
2. Klikamy prawym przyciskiem myszy na anotacjach i wybieramy *Font size*, a potem wielkość liter w punktach.
3. Klikamy prawym przyciskiem myszy na anotacjach i wybieramy *Zoom*, a potem procent powiększenia.
4. Litery etykiet i warstw powiększamy po naciśnięciu prawego przycisku myszy.

3.5.4. Szablony

Szablon to plik bez anotacji, który zawiera (1) typy warstw (*linguistic type*), (2) zestawy etykiet (*controlled vocabularies*) dla każdego typu warstwy, (3) przykładowe nazwy warstw. Plik anotacji utworzony na podstawie szablonu jest łatwiejszy w użyciu, gdyż zamiast wpisywać przy użyciu klawiatury etykiety za każdym razem, kiedy chcemy oznaczyć nimi dane zachowanie, wybieramy jedną z etykiet z menu, które rozwija się po utworzeniu anotacji na danej warstwie. Szablon ma rozszerzenie *.etf* i można go użyć przy tworzeniu nowego pliku anotacji lub narzucić szablon już utworzonemu plikowi anotacji przez wybranie w menu *File > Multiple File Processing... > Update transcriptions with Template...* Szablony mają rozszerzenie *.etf* i znajdziemy wiele szablonów przez wpisanie w wyszukiwarce *ELAN templates* lub *gesture annotation guide with ELAN template.*

3.5.5. Kopie zapasowe

Automatyczny zapis kopii zapasowych pozwala na zachowanie anotacji w tle i odzyskanie poprzednich wersji w przypadku utraty lub nadpisania pliku *.eaf* nowszą wersją. Zapis kopii aktywujemy w menu *FILE > Automatic backup* i wybieramy częstość zapisu. Plik kopii

zapasowej ma rozszerzenie *.001*, więc aby zobaczyć starsze anotacje w ELAN-ie, należy zmienić rozszerzenie na *.eaf*.

3.5.6. Monitor aktywności

Automatyczny zapis zmian anotacji w ELAN-ie służy do monitorowania pracy anotatora. Log zostanie zachowany w pliku *.txt*, z którego odczytamy, kiedy został otwarty dany plik i jaką anotację utworzono, usunięto lub zmieniono. Monitorowanie aktywujemy w menu *OPTIONS > Activity Monitoring*.

3.5.7. Skróty klawiaturowe

W ELAN-ie działają skróty klawiaturowe znane z innych programów (utwórz plik, otwórz plik, zapisz plik, zamknij plik, skopiuj tekst, wytnij tekst, wklej tekst, cofnij, powtórz, szukaj, drukuj) oraz skróty specyficzne dla pracy w tym programie. Wszystkie je możemy podejrzeć po wybraniu z menu *View > Shortcuts*, a zmieniać po wybraniu *Edit > Preferences > Edit Shortcuts*. Polecamy ustawienie skrótu . dla przesuwania okienka anotacji naprzód, bo nad kropką na klawiaturze jest > , a , dla cofania okienka anotacji, bo nad przecinkiem jest < i / dla menu etykiet. Dzięki temu można łatwo przechodzić z anotacji na anotację i wybierać odpowiednie dla obserwowanych ruchów etykiety przyciskiem do niej przypisanym, a potem przesuwamy się do następnej anotacji, naciskając przycisk stawiania kropki. Tabela 17 podaje proponowane skróty klawiszowe dla systemu MACOS, a w przypadku systemu Windows można zamienić przycisk CMD na inny.

Tabela 17. Propozycja skrótów klawiaturowych w ELAN-ie.

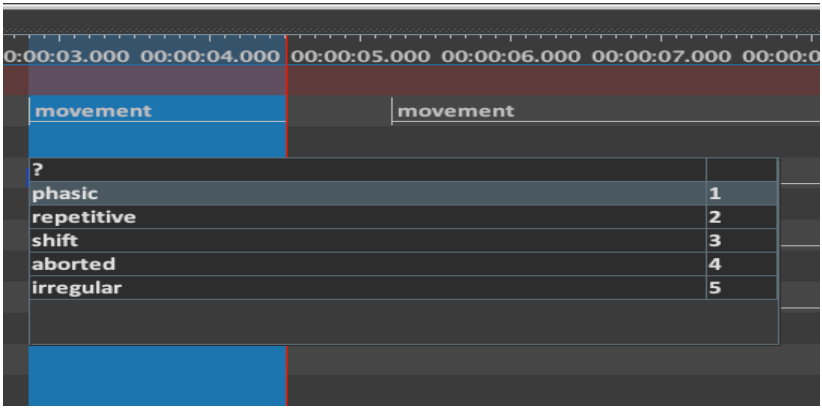
Play/pause	CTRL+spacja lub sam shift
Go to next annotation	.
Modify annotation value	/
Go to prev annotation	,
new annotation here	CTRL + SHIFT + N
split annotation	CTRL + SHIFT + S
merge with next annotation	CTRL + SHIFT + >
merge with prev annotation	CTRL + SHIFT + <
duplicate annotation	CMD + D
delete annotation	CTRL + D
copy tier	CTRL + ALT + CMD + C
delete tier	CTRL + ALT + CMD + D
remove annotation values	CTRL + ALT + CMD + R
label and number annotations	CTRL + ALT + CMD + L
switch to annotation mode	CTRL + ALT + CMD + A
switch to segmentation mode	CTRL + ALT + CMD + S

3.5.8. Skróty klawiaturowe dla etykiet

Skróty klawiaturowe możemy także ustawić dla etykiet stosowanych w wybranym systemie anotacji. W tym celu wybieramy z menu *EDIT > Edit Controlled Vocabularies* i przy wybranej etykiecie klikamy *More Options...*, a w polu *Entry Shortcut Key* wpisujemy przycisk, który ma być skrótem klawiaturowym do danej etykiety. Każda etykieta w jednym zestawie CONTROLLED VOCABULARY musi mieć unikalny klawisz, ale klawisze mogą się powtarzać dla różnych CONTROLLED VOCABULARY. Po przypisaniu klawiszy do etykiet możemy wydrukować naklejki na klawisze i okleić nimi klawiaturę, by łatwiej było nauczyć się przypisanych klawiszy. Przykład oklejonej klawiatury pokazuje Grafika 20.

Grafika 20. Przykład klawiatury oklejonej kolorowymi etykietami ze skrótami w systemie NEUROGES (projekt autora).

Inna możliwość to wpisanie klawiszy skrótów w pole *Entry description*. Wówczas przypisany do etykiety klawisz zobaczymy obok etykiety w liście etykiet przy anotacji, po kliknięciu w wybrany segment anotacji, co obrazuje Grafika 21. Cyfry przypisane do etykiet w ELAN-ie.

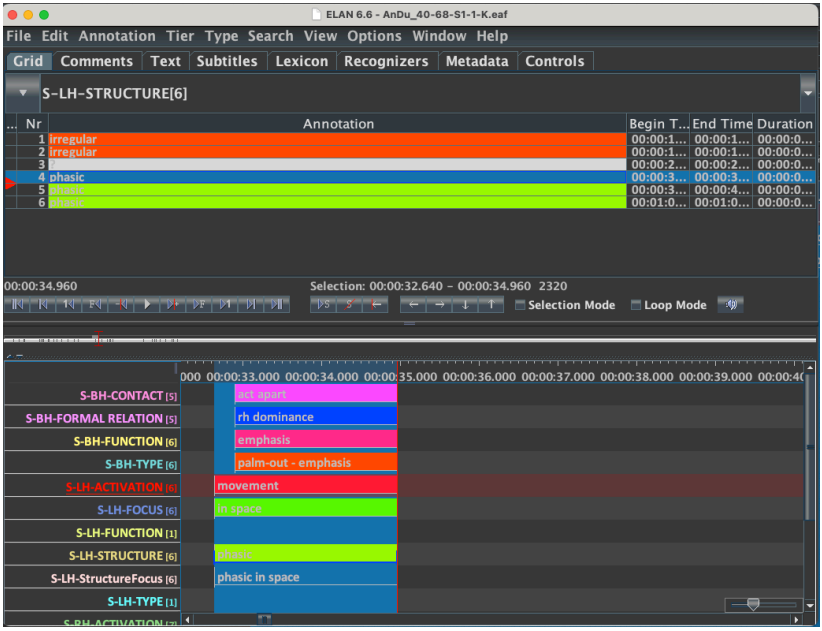


Grafika 21. Cyfry przypisane do etykiet w ELAN-ie.

3.5.9. Kolorowe etykiety

Dla etykiet stosowanych w wybranym systemie anotacji możemy także wybrać kolor. W tym celu wybieramy z menu *EDIT > Edit Con-*

trolled Vocabularies i przy wybranej etykiecie klikamy *More Options...*, a potem *Browse...* i wybieramy kolor, który ma być przypisany do danej etykiety. Należy pamiętać, że dane o skrótach klawiaturowych i kolorach etykiet ELAN zapisuje w pliku preferencji *.pref* dla aktualnie otwartego pliku *.eaf.* Jeśli chcemy skorzystać z tych ustawień w innym pliku *.eaf*, to należy skopiować plik *.pref* i zmienić jego nazwę na nazwę innego pliku *.eaf*. Grafika 22 pokazuje widok kolorowych etykiet w trybie anotacji.



Grafika 22. Widok kolorowych etykiet w ELAN-ie (projekt autora).

3.5.10. Przeszukiwanie anotacji

Szczegółowy opis trybów przeszukiwania znajdziemy w instrukcji ELAN-a, natomiast tutaj zwracamy uwagę na *Find and Replace*, czyli możliwość automatycznej korekty za pomocą znajdź i zamień, oraz *Structured Search*, czyli szukania strukturalnego, w którym możemy

formułować zapytania dotyczące wybranych etykiet na wybranych warstwach oraz sprawdzać relacje czasowe anotacji. Przed przeszukiwaniem warto zdefiniować zakres plików, które mają być przeszukiwane (*Define Domain*), dzięki czemu zapiszemy zestaw plików jako zestaw, do którego możemy wracać. Jeśli wybierzemy katalog, w którym zebraliśmy pliki anotacji, to ELAN uwzględni przy przeszukiwaniu także pliki w podkatalogach tego katalogu. W przeszukiwaniu mamy także możliwość tworzenia zapytań w wyrażeniach regularnych (*ELAN RegEx*), co jest przydatne do wyszukiwania wzorców.

3.5.11. Komentowanie anotacji

Niektóre zachowania komunikacyjne, mimo szczegółowych kryteriów anotacji, wymagają komentarza. W ELAN-ie komentarz można wpisać w zakładce *Comments* i wówczas zostanie on przypisany do wybranej anotacji. Komentarze można przeszukiwać i udostępniać innym badaczom we wspólnym katalogu, więc jest to funkcja bardzo przydatna przy poprawianiu i konsultowaniu anotacji. Grafika 23 pokazuje widok komentarzy do anotacji w ELAN-ie.

Grafika 23. Widok komentarzy do anotacji w ELAN-ie.

3.5.12. Operacje na warstwach

Tworzenie i zmiany na warstwach przeprowadzamy za pomocą menu TIER. Typ warstwy ustawiamy po wybraniu *Change Tier Attributes...*, a typy tworzymy lub zmieniamy ich parametry w menu głównym *Type*. Do każdego typu warstwy w szablonie jest utworzona lista etykiet (*controlled vocaubularies*). Podgląd list etykiet i ich parametrów znaj-

dziemy w menu głównym: *Edit > Edit Controlled Vocabularies*. Opcje zmian na wielu anotacjach w wielu plikach jednocześnie są w ELAN-ie dostępne w menu głównym: *File > Multiple File Processing.*

3.5.13. Problemy i rozwiązania

Dodatkowo warto znać rozwiązania problemów w ELAN-ie:

- ELAN się zacina przy dużych plikach *.eaf*, długich filmach. Czasem pomaga reinstalacja ELAN-a, zmiana dekodera filmowego i formatu filmów, restart komputera czy chwila przerwy w pracy, ale niekiedy nie pomaga nic, więc pozostaje jedynie cierpliwie klikać.
- Na wszelki wypadek warto aktualizować silnik ELAN-a – Javę.
- Jeśli nie widać filmu w oknie ELAN-a lub nie odtwarza się on płynnie, to należy sprawdzić, jaki plik jest zalinkowany: *Edit > Linked files...*
- Jeśli otwieranie pliku *.eaf* prowadzi do zawieszenia aplikacji, to być może plik jest za duży, czyli ma za dużo warstw albo zawiera za dużo anotacji dla długiego filmu. Wówczas należy rozważyć podział pliku na krótsze fragmenty i anotować na mniejszej liczbie warstw. Jeśli otwarcie pliku w ELAN-ie okazuje się niemożliwe, możemy spróbować podzielić plik za pomocą biblioteki *pympi-ling* w Pythonie, co opisujemy poniżej.

3.6. Przetwarzanie anotacji z ELAN-a w Pythonie

ELAN pozwala na tworzenie anotacji dla wielu plików na wielu warstwach, co znacznie ułatwia kompletowanie korpusu danych. Dostępne są także operacje na wielu plikach na raz (*File > Multiple File Processing*), a pozostałe możemy przeprowadzić za pomocą biblioteki *pympi-ling* w Pythonie. Poniżej podajemy skrypty, które zostały opracowane na potrzeby porządkowania plików anotacji, warstw i samych anotacji. Poza operacjami prostymi, takimi jak usuwanie czy kopiowanie warstw, biblioteka *pympi-ling* umożliwia korekty anotacji wedle reguł, dodawanie informacji do plików anotacji o uczestnikach i anotatorach, porządkowanie komentarzy do anotacji oraz ustawianie

kolorów i skrótów klawiaturowych do etykiet na podstawie listy pobieranej z pliku *.csv*. Omówimy wybrane funkcje biblioteki.

3.6.1. Instalacja i import

Pierwszy krok to instalacja biblioteki w wierszu poleceń:
pip install pympi-ling
Następnie wczytujemy bibliotekę `pympi` w Pythonie:

```
import pympi
```

3.6.2. Wczytanie pliku *.eaf*

Przed każdą operacją należy wczytać plik *.eaf*, dlatego dla wygody określamy katalog roboczy:

```
import os
eafpath = "plikiZelana"
os.chdir(eafpath)
```

Zamiast `plikiZelana` należy wpisać katalog, w którym mamy zachowane pliki z ELAN-a.

Wczytujemy plik *.eaf* jako obiekt do klasy *pympi.Eaf*:

```
eaffile = pympi.Eaf(eaf_file)
```

Zamiast `eaf_file` wpisujemy nazwę pliku *.eaf*.

3.6.3. Przykładowe operacje w *pympi-ling*

Podaj listę warstw w pliku:

```
listier = eaffile.get_tier_names()
```

Dodaj warstwę:

```
eaffile.add_tier(tiername, ling= tiertype)
```

Zamiast `tiername` wpisujemy nazwę warstwy, a zamiast `tiertype` – nazwę typu dla warstwy. Typ musi być utworzony wcześniej w pliku *.eaf.*

Usuń warstwę:

```
eaffile.remove_tier(oldname)
```
Zamiast `oldname` wpisujemy nazwę warstwy, którą chcemy usunąć.

Zmień nazwę warstwy:
```
eaffile.rename_tier(oldname, newname)
```
Zamiast `oldname` wpisujemy nazwę warstwy, której nazwę zmieniamy, a zamiast `newname` – nową nazwę warstwy.

Połącz dwie warstwy w jedną:
```
eaffile.merge_tiers(tier1,tier2, safe=True)
```
Zamiast `tier1` i `tier2` wpisujemy nazwy warstw, które chcemy połączyć.

Podaj listę anotacji na wybranej warstwie:
```
annotations = eaffile.get_annotation_data_for_
tier(tier)
```
Zamiast `tier` wpisujemy nazwę warstwy, z której chcemy uzyskać listę anotacji.

Usuń wszystkie anotacje na wybranej warstwie:
```
eaffile.remove_all_annotations_from_tier(tier)
```
Zamiast `tier` wpisujemy nazwę warstwy, z której chcemy usunąć anotacje.

Dodaj anotację na wybranej warstwie:
```
eaffile.add_annotation(tier,start,stop, value= text)
```
Zamiast tier wpisujemy nazwę warstwy, na której chcemy dodać anotacje złożone z początku – start, końca – start i etykiety – text.

3.6.4. Zachowywanie zmian w pliku *.eaf*

Zmiany zachowujemy, wywołując:
```
eaffile.to_file(eaf_file)
```

3.6.5. Wywołanie poleceń pympi w formie funkcji

Każdą z wyżej wymienionych funkcji możemy włączyć w nową, która wczyta plik *.eaf* i zwróci przerobiony plik, na przykład:

```
def ADDtier(eaf_file, tiername, tiertype):
    eaffile = pympi.Eaf(eaf_file)
    eaffile.add_tier(tiername, ling= tiertype)
    eaffile.to_file(eaf_file)
```

3.6.6. Kopiowanie warstwy

Skopiuj warstwę w pliku *.eaf*, czyli zduplikuj warstwę. Do tej operacji potrzebna jest nowa funkcja:

```
def DUPtier(eaf_file, oldname, newname, tiertype,
parentier):
    eaffile = pympi.Eaf(eaf_file)
    annotations = eaffile.get_annotation_data_for_
tier(oldname)
eaffile.add_tier(newname, ling=tiertype,
parent=parentier)
    for s in range(0,len(annotations)):
          (start, stop, text) = annotations
          eaffile.add_annotation(newname,start,stop,val-
ue=text)
eaffile.to_file(eaf_file)
del (eaffile,annotations)
```

Zamiast `eaf_file` wpisujemy nazwę pliku, w którym chcemy powtórzyć warstwę o nazwie `oldname`, a nowa warstwa będzie miała nazwę `newname` i typ `tiertype` oraz warstwę nadrzędną `parentier`.

Albo w innej wersji:

```
def DUPtier2(eaf_file, oldname, newname):
    annofile = pympi.Elan.Eaf(eaf_file)
    tempfile = pympi.Elan.Eaf(pympi.__init__(" emp-
ty.eaf" ))
    annofile.copy_tier(tempfile,oldname)
    annofile.remove_tier(oldname)
    tempfile.rename_tier(oldname,newname)
    tempfile.copy_tier(annofile,newname)
    annofile.to_file(eaf_file)
```

3.6.7. Łączenie warstw anotacji

Połącz anotacje z dwóch plików w jeden, czyli funkcja dostępna w ELAN-ie w menu *File > Merge transcriptions...* Łączenie anotacji w jeden plik przydaje się, jeśli chcemy porównać anotacje dwóch anotatorów dla jednego nagrania. Pierwsza wersja tej funkcji dodaje warstwy anotacji z jednego pliku do drugiego i zachowuje jako nowy plik.

```
def MEReafs(eaf_file1, eaf_file2):
    eaffile1 = pympi.Eaf(eaf_file1)
    eaffile2 = pympi.Eaf(eaf_file2)
    tierlist = eaffile1.get_tier_names()
    for tier in tierlist:
          annotations =
    eaffile1.get_annotation_data_for_tier(tier)
          eaffile2.add_tier(tier)
          for s in range(0,len(annotations)):
                (start, stop, text) = annotations
                eaffile2.add_annota-
tion(tier,start,stop,value=text)
          eaffile2.to_file(eaf_file2)
    del(eaffile1,eaffile2,annotations)
```

Zamiast `eaf_file1, eaf_file2` wpisujemy nazwy łączonych plików. Jednakże jeśli w dwóch plikach *.eaf* mamy takie same nazwy warstwy, to musimy je zmienić przed kopiowaniem do jednego pliku, bo ELAN nie pozwala na dwie warstwy o takich samych nazwach w jednym pliku. Dlatego druga funkcja uwzględnia zmiany nazwy warstw na nowy format złożony z nazwy warstwy i ostatnich trzech liter z nazwy pliku do oznaczenia anotatora. Załóżmy, że mamy dwa pliki anotacji dla tego samego nagrania wykonane przez różnych anotatorów: 81-60-S5-4-K-Mly.eaf i 81-60-S5-4-K-Occ.eaf, gdzie Mly i Occ oznaczają właśnie tych anotatorów. Łączymy oba pliki w nowy plik o nazwie 81-60-S5-4-K-Occ-Mly.eaf za pomocą funkcji:

```
def MEReafs(eaf_file1, eaf_file2, new_file):
   import pympi
   import os
   eaffile1 = pympi.Eaf(eaf_file1)
   eaffile2 = pympi.Eaf(eaf_file2)
   eaffile_new = pympi.Eaf() # Create a new Eaf file
   # Extracting the last three characters of the file
   names without extension
   file1_name_suffix =
os.path.splitext(os.path.basename(eaf_file1))
   file2_name_suffix =
os.path.splitext(os.path.basename(eaf_file2))
   # Process the first file
   tierlist1 = eaffile1.get_tier_names()
   for tier in tierlist1:
   new_tier_name = f"{tier}_{file1_name_suffix}"
   annotations =
eaffile1.get_annotation_data_for_tier(tier)
   eaffile_new.add_tier(new_tier_name)
   for start, stop, text in annotations:
   eaffile_new.add_annotation(new_tier_name, start,
stop, value=text)
   # Process the second file
   tierlist2 =
eaffile2.get_tier_names()
   for tier in tierlist2:
   new_tier_name = f"{tier}_{file2_name_suffix}"
   annotations = eaffile2.get_annotation_data_for_
   tier(tier)
   eaffile_new.add_tier(new_tier_name)
   for start, stop, text in annotations:
   eaffile_new.add_annotation(new_tier_name, start,
stop, value=text)
   # Save the new file
   eaffile_new.to_file(new_file)
   del(eaffile1, eaffile2, eaffile_new)
```

Zamiast `eaf_file1, eaf_file2` wpisujemy nazwy łączonych plików, a `new_file` to nazwa nowego pliku.

3.6.8. Zmiana typu warstwy

Zmień typ warstwy:

```
def SETtype(eaf_file, tiername, tiertype):
   eaffile = pympi.Elan.Eaf(eaf_file)
   tierlist = eaffile.get_tier_names()
   if tiername in tierlist:
       annotations =
eaffile.get_annotation_data_for_tier(tiername)
   eaffile.remove_tier(tiername)
   eaffile.add_tier(tiername, ling=tiertype)
   for s in range(0, len(annotations)):
       (start, stop, text)=annotations
       eaffile.add_annotation(tiername, start, stop,
   value=text)
   eaffile.to_file(eaf_file)
del (eaffile, tiername, tiertype, tierlist)
```

Zamiast `eaf_file` wpisujemy nazwę pliku *.eaf*, a zamiast `tier-name` – nazwę warstwy i zamiast `tiertype` – nazwę typu.

3.6.9. Zamiany tekstu w anotacji

Znajdź i zamień w anotacjach lub transkrypcjach na wybranej warstwie:

```
def findANDrepl(eaf_file,tier,findtext,repltext):
   eaffile = pympi.Eaf(eaf_file)
   annotations = eaf_file.get_annotation_data_for_
   tier(tier)
   eaffile.remove_all_annotations_from_tier(tier)
   for s in range(0,len(annotations)):
       (start, stop, text) = annotations
       text = re.sub(findtext, repltext, text)
       eaffile.add_annotation(tier,start,stop, value=-
   text)
   eaffile.to_file(eaf_file)
```

Zamiast `eaf_file` wpisujemy nazwę pliku *.eaf*, a zamiast `tier` – warstwy i `findtext` – szukany tekst, a `repltext` – tekst do zamiany w miejsce `findtext`. W miejscu *re.sub* można podstawić dowolne wyrażenie regularne, wedle którego chcemy dokonać korekty. W ten sposób możemy oczyścić transkrypcję z numerów, znaków niealfabetycznych albo zamienić wielkie litery na małe i tym podobne. Funkcja jest dostępna w ELAN-ie w menu *Search > Find (and Replace)*.

3.6.10. Tokenizacja anotacji

Rozdziel anotacje lub transkrypcje na anotacje, gdzie spacja jest separatorem. Funkcja jest dostępna w ELAN-ie w menu *Tier > Tokenize tier*:

```
def TOKtier(eaf_file, oldname, newname):
   transeaf = pympi.Eaf ( eaf_file )
   aucz = transeaf.get_annotation_data_for_tier(old-
   name)
   transeaf.add_tier(newname)
   for s in range ( 0 , len ( aucz ) ):
       (start , stop , text) = aucz
       wordList = re.sub ( "",” “ , text ).split ( )
       licz = len(wordList)
       if licz!=0:
             leng = stop-start
             word = leng/licz
             w = 0
             for wyr in wordList:
                   starw = start+(word*w)
                   starw = round ( starw )
                   w = w + 1 # początek anotacji
                   stopw = start+(word*w) #koniec
                   stopw = round ( stopw )
                   transeaf.add_annotation (newname ,
   starw , stopw , value = wyr )
   transeaf.to_file(eaf_file)
```

Zamiast `eaf_file` wpisujemy nazwę pliku *.eaf*, a `oldname` i `newname` to nazwy warstw. Efekt tokenizacji pokazuje Grafika 17, gdzie segmenty anotacji na warstwie „UCZESTNIK-WORDS" to słowa wydzielone z wypowiedzi z warstwy „uczestnik". Taki rodzaj tokenizacji jest prostym podziałem tekstu na jednostki wyrazowe oddzielane spacjami (pauzami), natomiast tokenizacja w ramach przetwarzania języka naturalnego wydziela także jednostki wielowyrazowe jako segmenty, na przykład nazwy typu *Nowy Tomyśl* jako jeden token.

Bardziej złożone operacje pozwalają na uzupełnienie danych o uczestnikach lub anotatorach na podstawie pliku *.csv* albo na ustawienie skrótów i kolorów etykiet także z pliku *.csv*; wyznaczenie fragmentów anotacji, w których współwystępują określone słowa i gesty. Możliwe jest także zdefiniowanie zestawu reguł anotacji, wedle których chcielibyśmy uporządkować anotacje w ramach wybranego systemu. Reguły dotyczyłyby usuwania, dodawania lub zmieniania anotacji na określonych warstwach zależnie od anotacji na innych warstwach. Nie dyskutujemy ich tutaj, gdyż wymagałoby to szczegółowego omówienia wybranego systemu anotacji, na przykład NEUROGES (Lausberg 2019).

3.7. Anotacje ruchów pozostałych części ciała

Anotacje mimiki i ruchów głowy są rzadziej wykonywane, ale warto wspomnieć o systemie *Facial Action Coding System* (P. Ekman i Friesen 1975) czy mało znanej wśród naukowców *Synergologii* (Turchet 2009). Pierwszy system składa się z 48 etykiet oznaczających ruchy mięśni twarzy, które w określonych przez autorów FACS kombinacjach wyrażają emocje podstawowe. Drugi system, choć nie cieszy się uznaniem w społeczności naukowców, o czym świadczy brak cytowań Turchet w publikacjach naukowych, pozwala opisać ruchy całego ciała: głowy, oczu, ust, rąk, tułowia i nóg, które w toku dalszej analizy synergologicznej są interpretowane semantycznie i pragmatycznie. Anotację ruchów głowy, nóg i tułowia można także przeprowadzić w rozszerzonej wersji systemu NEUROGES (Lausberg 2019).

Istnieją także programy do automatycznej anotacji mimiki, gdyż zachowania mimiczne są między innymi wyrazami emocji, a te z kolei

są badane w celach nie tylko naukowych (*Hume, Imotions*), ale także marketingowych (*Affectiva*). Badaniu podlegają reakcje widzów na pokazywane im bodźce: słowa, obrazy, fragmenty filmów, reklamy.

Systemów anotacji ruchów rąk, które są używane w językach migowych i miganych, nie omawiamy, ale zainteresowanych odsyłamy do baz znaków migowych (*SignBank*), gdzie znajdziemy przykłady w wielu językach. Niektóre z etykiet używanych do opisu migów (głównie konfiguracje palców, kształt dłoni i typy trajektorii oraz określanie osi ruchu) są także używane w systemach anotacji ruchów rąk języka mówionego. Przykładem systemu kodowania znaków języka migowego jest na przykład *HamNoSys* (Hanke 2004), czyli Hamburski System Notacji Języków Migowych (*Hamburg Sign Language Notation System*), tj. wizualno-przestrzennego odpowiednika międzynarodowego alfabetu fonetycznego (*International Phonetic Alphabet*, IPA). Korpusowe badania dotyczące polskiego języka migowego kodowanego w tym systemie są prowadzone w Pracowni Lingwistyki Migowej Uniwersytetu Warszawskiego (Rutkowski i in. 2014).

3.8. Automatyczne anotacje ruchu

Wraz z rozwojem sztucznej inteligencji w ostatniej dekadzie pojawiły się liczne biblioteki rozpoznawania obrazu na filmie do automatycznej anotacji ruchu całego ciała z wyróżnieniem ponad 20 stawów, dzięki czemu możliwe jest śledzenie ruchu rąk, każdego palca, a także ruchów głowy i mimiki nie tylko jednej, ale wielu osób jednocześnie. Szczególnie wymienić trzeba: MediaPipe, OpenPose, Kinectics toolkit oraz envisionbox, pakiet łączący te biblioteki do przetwarzania obrazu w celach badawczych. Wspomniane biblioteki pozwalają na automatyczną anotację ruchu na podstawie nagrania zwykłą, dowolną kamerą albo czujnikami ruchu na podczerwień takimi jak Microsoft Kinect. Rozpoznawanie funkcji ruchów rąk i gestów wciąż się rozwija i znajduje więcej zastosowań. W najnowszych publikacjach wspomnianych zespołów zwraca się uwagę na konieczność publikowania korpusów i kodu, by zwiększyć powtarzalność badań i prawdopodobieństwo uzyskiwania zbliżonych wyników badań.

3.9. Inne programy do anotacji i analizy korpusów multimodalnych

Istnieją także inne specjalistyczne programy do anotacji i analizy korpusów multimodalnych, które tutaj jedynie wymienimy, a ich opis jest w publikacjach:

- SPPAS (*Automatic Annotation and Analyses of Speech*): aplikacja w *Pythonie* do porządkowania i analizowania anotacji z ELANa o zakresie funkcji zbliżonym do *Pympi-Ling* (Bigi 2012).
- ANNOTATION PRO (*A Desktop Module for Automatic Segmentation and Transcription)*: program do transkrypcji, anotacji i analizy mowy i ruchu o zakresie funkcji zbliżonym do ELANa (Klessa 2016).
- MEA (*Motion Energy Analysis*): analizuje przemieszczenie pikseli w wybranych obszarach kadru filmu, stosowany do pomiaru dynamiki zachowań symetrycznych i synchronicznych pomiędzy uczestnikami sesji psychoterapeutycznych (Ramsayer 2020).
- THEME: statystyczna analiza złożonych wzorców czasowych w danych z różnych źródeł i badań, od ruchu zwierząt i ludzi do aktywności neuronów w mózgu (Anolli i in. 2005).

4. Udostępnianie korpusu nagrań i anotacji

Publikowanie danych umożliwia badaczom zapoznanie się z materiałem i porównanie badań. W publikacjach na temat komunikacji multimodalnej znajdujemy cztery formy danych:

- opisy gestów (McNeill 1992 oraz większość innych książek i artykułów o gestach);
- szkice osób i układów rąk (Calbris 2011; Bressem 2021; Ladewig 2020; Chui 2022; Li 2014 i wiele innych);
- klatki z filmów (na przykład Załazińska 2006; Klima, Bellugi i Battison 1979);
- filmy na płycie CD (Załazińska 2006; Kiełbawska 2012) lub osadzone w pliku PDF (Antas 2013).

Niektóre czasopisma naukowe (np. „Language and Cognition") wymagają, by dane opisywane w publikacji były w otwartym dostępie i aby każdy mógł je pobrać i obejrzeć. Dlatego dane korpusowe są udostępniane innym badaczom w repozytoriach danych w Internecie. Czasem do danych dołączane są fragmenty kodu, które pozwalają na powtórzenie analiz opisanych w artykule. Przy upublicznianiu danych korpusowych należy wziąć pod uwagę: (1) zgodę na udostępnienie danych i nagrań, (2) rodzaj i zakres publikowanych danych, (3) miejsce publikacji, (4) licencję, (5) ograniczenia dostępu, (6) metadane.

1. Zgodę na publikację danych otrzymujemy od uczestników badań i uczelni, na której są realizowane badania, oraz instytucji finansującej badania, jeśli to konieczne.
2. Wśród rodzajów danych publikowanych jako korpusów wyróżniamy:
   2.1. nagrania dźwiękowe i nagrania filmowe;
   2.2. transkrypcje (teksty jako *.txt* lub pliki *.eaf* lub w innym formacie);
   2.3. anotacje (tagi jako *.csv* lub pliki *.eaf* lub w innym formacie);
   2.4. wyniki badań ankietowych i kwestionariuszowych (dodatkowo można opublikować ankietę lub kwestionariusz, jeśli mamy do tego prawo);
   2.5. wyniki badań eksperymentalnych (dodatkowo można opublikować opis i listę bodźców, jeśli mamy do tego prawo);
   2.6. kod do analizy danych wraz ze wskazaniem bibliotek i ich wersji dla danego kodu.

3. Repozytoria i bazy danych utrzymywane są przez uczelnie lub organizacje naukowe oraz firmy, takie jak Microsoft czy HuggingFace, Kaggle, na przykład:
   3.1. korpusy badawcze, czyli nagrania z transkrypcjami i anotacjami, raporty z badań:
      3.1.1. *The Language Archive* na *Max Planck Institute for Psycholinguistics in Nijmegen*;
      3.1.2. *Open Science Foundation*: repozytorium służące otwartemu dostępowi do danych;*nieniu od CLARIN-PL* i *CLARIN-EU:* repozytorium przede wszystkim danych językowych;
      3.1.3. *Endangered Languages Project*: projekt służący zachowaniu języków zagrożonych;
      3.1.4. *DOBES*: dokumentacja zagrożonych języków;
      3.1.5. *Talk Bank*: baza konwersacji z nagraniami audio i wideo, dzieci i dorosłych;
      3.1.6. *CORLI: Consortium HN CORpus, Langues et Interactions*: dokumentacja języków;
      3.1.7. *OrtoLang*: platforma narzędzi i zasobów językowych przystosowanych do analiz języka francuskiego;
      3.1.8. *The International Distributed Little Red Hen Lab*, czyli *Międzynarodowe rozproszone laboratorium Red Hen*: korpusy multimodalne na podstawie nagrań telewizyjnych i internetowych oraz narzędzia do ich automatycznej analizy;
   3.2. zbiory danych do trenowania rozpoznawania różnych danych oraz kody źródłowe:
      3.2.1. *HuggingFace*,
      3.2.2. *Kaggle*,
      3.2.3. *GitHub*,
      3.2.4. *Roboflow*,
      3.2.5. *Papers with code*,
      3.2.6. *Sketch Engine*.
4. Jeśli uzyskaliśmy zgodę na publikowanie nagrań i anotacji, możemy wybrać rodzaj licencji. Na tym etapie pomocne będą narzędzia dostępne na stronach organizacji *Creative Commons* lub *Open Science Foundation*.

5. Przy udostępnianiu danych możemy ograniczyć dostęp do osób, które zarejestrują się w bazie danych i zastrzec, że udzielimy dostępu na życzenie przesłane na adres e-mailowy.
6. Metadane to informacje opisujące korpus multimodalny, popularny i polecany format metadanych dla korpusów językowych został opracowany przez konsorcjum TEI: *Text Encoding Initiative*, a dla multimodalnych CMDI: *The Component Metadata Infrastructure* (Freigang i Bergmann 2013).

Podane wyżej repozytoria danych pozwalają także na pobieranie zapisanych tam korpusów i tworzenie korpusów zewnętrznych, czyli na podstawie już istniejących danych.

5. Podsumowanie

W tym krótkim przeglądzie zaleceń i procedur dla badaczy komunikacji multimodalnej skupiłem się na kwestiach technicznych tworzenia korpusu nagrań, transkrypcji, anotacji i ich publikacji. Podzieliliśmy się doświadczeniem zdobytym w kilku projektach oraz dokonaliśmy przeglądu praktyk związanych z opisywaniem korpusów multimodalnych w wybranych publikacjach. Żaden podręcznik ani przewodnik nie jest wyczerpujący, ale czytelnik może swoją wiedzę uzupełnić, korzystając z innych publikacji (Karpiński i Klessa 2021; Kipp 2009; Berez-Kroeker i in. 2022; Thiran, Bourlard i Marques 2010).

Bibliografia

Allwood, Jens, Loredana Cerrato, Kristiina Jokinen, Costanza Navarretta i Patrizia Paggio. 2007. „The MUMIN Coding Scheme for the Annotation of Feedback, Turn Management and Sequencing Phenomena". *Language Resources and Evaluation* 41 (3–4): 273–87. https://doi.org/10.1007/s10579-007-9061-5.

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard i in. 1991. „The Hcrc Map Task Corpus". *Language and Speech* 34 (4): 351–66. https://doi.org/10.1177/002383099103400404.

Anolli, Luigi, Starkey Duncan Jr, i Magnus S. Magnusson, red. 2005. *The Hidden Structure of Interaction: From Neurons to Culture Patterns: 7*. Amsterdam ; Washington, DC: IOS Press.

Antas, Jolanta. 2013. *Semantyczność ciała. Gesty jako znaki myślenia*. Łódź: Primum Verbum.

Auer, Eric, Albert Russel, Han Sloetjes, Peter Wittenburg, Oliver Schreer, S. Masnieri, Daniel Schneider i Sebastian Tschöpel. 2010. „ELAN as flexible annotation framework for sound and image processing detectors". W , 890–93. European Language Resources Association (ELRA). https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_442676_4.

Austin, John Langshaw. 1993. *Mówienie i poznawanie: rozprawy i wykłady filozoficzne*. Tłum. Jan Woleński i Bohdan Chwedeńczuk. Biblioteka Współczesnych Filozofów. Warszawa: Wydaw. Naukowe PWN.

Barre, Frances La. 2013. *On Moving and Being Moved: Nonverbal Behavior in Clinical Practice*. New York: Routledge.

Berez-Kroeker, Andrea L., Bradley James McDonnell, Eve Koller i Lauren B. Collister, red. 2022. *The Open Handbook of Linguistic Data Management*. Open Handbooks in Linguistics Series. Cambridge, Massachusetts: The MIT Press.

Bigi, Brigitte. 2012. „SPPAS: a tool for the phonetic segmentation of speech". W *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Red. Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, i Stelios Piperidis, 1748–55. Istanbul, Turkey: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/1116_Paper.pdf.

Birdwhistell, R. 1970. *Kinesics and context: essays on body motion communication*. Philadelphia: University of Pennsylvania Press.
Boersma, Paul i David Weenink. 2006. „Praat: doing phonetics by computer". https://www.fon.hum.uva.nl/praat/.
Bressem, Jana. 2021. „Repetitions in Gesture: A Cognitive-Linguistic and Usage-Based Perspective". W *Repetitions in Gesture*. De Gruyter Mouton. https://doi.org/10.1515/9783110697902.
Bressem, Jana, Silva Ladewig, i Cornelia Müller. 2014. „Linguistic Annotation System for Gestures (LASG)". W *Body – Language – Communication*. Red. Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, Jana Bressem, i David McNeill, 1098–1124. Berlin ; Boston: Walter de Gruyter GmbH & Co KG.
Broda, Bartosz i Maciej Piasecki. 2008. „SuperMatrix: a General Tool for Lexical Semantic Knowledge Acquisition". W *Speech and Language Technology*, 11:239–54. Polish Phonetics Assocation.
Budzyńska, Katarzyna, Mathilde Janier, Juyeon Kang, Barbara Konat, Chris Reed, Patrick Saint-Dizier, Manfred Stede i Olena Yaskorska. 2015. „Automatically identifying transitions between locutions in dialogue", czerwiec, 1–18.
Budzyńska, Katarzyna, Barbara Konat i Marcin Koszowy. 2016. „Korpusowe metody badania logosu i etosu". *Zagadnienia Naukoznawstwa*, nr No 3. https://doi.org/10.24425/118018.
Budzyńska, Katarzyna i Chris Reed. 2011. „Whence inference? Technical report."
Bunt, Harry. 2011. „Multifunctionality in Dialogue". *Computer Speech & Language* 25 (2): 222–45. https://doi.org/10.1016/j.csl.2010.04.006.
Calbris, Geneviève. 2011. *Elements of Meaning in Gesture*. John Benjamins Publishing Company. https://doi.org/10.1075/gs.5.
Cameron, Lynne i Robert Maslen. 2010. *Metaphor Analysis: Research Practice in Applied Linguistics, Social Sciences and the Humanities*. Equinox.
Chomsky, Noam. 1957. „Recenzja z Verbal Behaviour B.F. Skinnera". W *Lingwistyka a filozofia*. Red. B. Stanosz. Warszawa: Państwowe Wydawnictwo Naukowe.
———. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
Chui, Kawai. 2022. *Language and gesture in Chinese conversation: Bǐshǒushuōhuà*. Routledge studies in Chinese discourse analysis. London ; New York: Routledge, Taylor & Francis Group.
Cienki, Alan. 1998. „Metaphoric gestures and some of their relations to verbal metaphoric expressions". *Discourse and cognition: Bridging the gap*, 189–204.

———. 2016. „Cognitive Linguistics, Gesture Studies, and Multimodal Communication”. *Cognitive Linguistics* 27 (4): 603–18. https://doi.org/10.1515/cog-2016-0063.

Corballis, M. C. 2013. „Gesture as Precursor to Speech in Evolution”. W *Body – Language – Communication*, I:466–79. Berlin ; Boston: Walter de Gruyter.

Deignan, Alice. 2015. „MIP, the corpus and dictionaries”. *Metaphor and the Social World* 5 (1): 145–54. https://doi.org/10.1075/msw.5.1.09dei.

Deignan, Alice i Elena Semino. 2010. „Corpus techniques for metaphor analysis”. W *Metaphor Analysis*, 161–79. Equinox.

Deitel, Paul. 2019. *Python for Programmers: With Big Data and Artificial Intelligence Case Studies*. Boston, MA: Pearson.

Demenko, Grażyna, Maciej Wypych i Emilia Baranowska. 2003. „Polphone – (grapheme-to-phoneme conversion) Implementation of grapheme-to-phoneme rules and extended SAMPA alphabet in Polish text-to-speech synthesis.” *Speech and Language Technology* 17 (7).

Deng, Li. 2014. „Deep Learning: Methods and Applications”. *Foundations and Trends® in Signal Processing* 7 (3–4): 197–387. https://doi.org/10.1561/2000000039.

Drewes, V, N Neumann i Konstantinidis i I Helmich. 2020. „Spontaneous Head Movements Characterize Losing Athletes during Competition”. *International Journal of Sports Science & Coaching* 15 (5–6): 669–76. https://doi.org/10.1177/1747954120934598.

Dunbar, Angela. 2016. *Clean Coaching: The Insider Guide to Making Change Happen*. London: Taylor & Francis.

Efron, David. 1972. *Gesture, Race and Culture: A Tentative Study of the Spatio-Temporal and „Linguistic” Aspects of the Gestural Behavior of Eastern Jews and Southern Italians in New York City, Living Under Similar as Well as Different Environmental Conditions*. Mouton.

Ekman, Paul i Friesen. 1975. *Unmasking the Face: a Guide to Recognizing Emotions from Facial Clues*. Prentice Hall, Englewood.

Ekman, Paul i Wallace V. Friesen. 1969. „The repertoire of nonverbal behavior: Categories, origins, usage, and coding”. *Semiotica* 1 (1): 49–98.

Fabiszak, Małgorzata i Barbara Konat. 2013. „Zastosowanie korpusów językowych w językoznawstwie kognitywnym”. W *Metodologie językoznawstwa. Ewolucja języka Ewolucja teorii językoznawczych*. Red. Piotr Stalmaszczyk, 1:131–42. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Freigang, F. i K. Bergmann. 2013. „Towards Metadata Descriptions for Multimodal Corpora of Natural Communication Data”. W *Proceedings of the Workshop on Multimodal Corpora 2013: Multimodal Corpora: Beyond Audio and Video*. Red. J. Edlund, D. Heylen i P. Paggio.

Freleng, Friz. 1950. Canary Row. Animation, Short, Adventure. Warner Bros.
Gibbs, Raymond W. 2011. „Evaluating Conceptual Metaphor Theory". *Discourse Processes* 48 (8): 529–62. https://doi.org/10.1080/01638533X.2011.606103.
Gibert, Núria Esteve, PeiLin Ren, Patrick Louis Rohrer, Júlia Florit-Pons, Ulya Tütüncübasi, Ingrid Vilà-Giménez, Stefanie Shattuck-Hufnagel i Pilar Prieto. 2020. „The MultiModal MultiDimensional (M3D) labeling system". https://doi.org/10.17605/OSF.IO/ANKDX.
Glynn, Dylan i Kerstin Fischer, red. 2010. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. De Gruyter Mouton. https://doi.org/10.1515/9783110226423.
Goldin-Meadow, Susan. 2013. „How Our Gestures Help Us Learn". W *Body – Language – Communication*, I:792–803. Berlin ; Boston: Walter de Gruyter.
Gupta, Itika, Barbara Di Eugenio, Brian Ziebart, Aiswarya Baiju, Bing Liu, Ben Gerber, Lisa Sharp, Nadia Nabulsi i Mary Smart. 2020. „Human-Human Health Coaching via Text Messages: Corpus, Annotation, and Analysis". W *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Red. Olivier Pietquin, Smaranda Muresan, Vivian Chen, Casey Kennington, David Vandyke, Nina Dethlefs, Koji Inoue, Erik Ekstedt i Stefan Ultes, 246–56. 1st virtual meeting: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.sigdial-1.30.
Hajnicz, Elżbieta. 2022. „Annotation of metaphorical expressions in the Basic Corpus of Polish Metaphors". W *Proceedings of the Language Resources and Evaluation Conference*, 5648–53. Marseille, France: European Language Resources Association. https://aclanthology.org/2022.lrec-1.606.
Hanke, Thomas. 2004. „HamNoSys – representing sign language data in language resources and language processing contexts." W *LREC 2004, Workshop proceedings : Representation and processing of sign languages*. Red. Oliver Streiter i Vettori Chiara, 1–6. ELRA: Paris.
Heinz, Adam. 1983. *Dzieje językoznawstwa w zarysie*. Wyd. 2. Warszawa: Państwowe Wydaw. Naukowe.
Holle, Henning i Robert Rein. 2015. „EasyDIAg: A Tool for Easy Determination of Interrater Agreement". *Behavior Research Methods* 47 (3): 837–47. https://doi.org/10.3758/s13428-014-0506-7.
Holler, Judith. 2013. „Experimental methods in co-speech gesture research". W *Body – language – communication: an international handbook on multimodality in human interaction*. Red. Cornelia Müller, Alan J. Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill, i Sedinha Tessendorf, 837–56. Handbooks of linguistics and communication science, 38.1. Berlin ; Boston: De Gruyter Mouton.

Idström, Anna, Elisabeth Piirainen i Tiber Falzett, red. 2012. *Endangered metaphors*. Cognitive linguistic studies in cultural contexts, v. 2. Amsterdam ; Philadelphia: John Benjamins Pub. Co.

Imai, Kosuke. 2017. *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press.

Imai, Kosuke i Lori D. Bougher. 2021. *Quantitative Social Science: An Introduction in Stata*. Princeton: Princeton University Press.

Janier, Mathilde, John Lawrence i Chris Reed. 2014. „OVA+: An argument analysis interface". W *Proceedings of the 5th International Conference on Computational Models of Argument (COMMA'14)*, 463–64.

Jarmołowicz-Nowikow, Ewa. 2009. „Polish Children's Gesticulation in Narrating (Re-telling) a Cartoon". W *Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions*. Red. Anna Esposito i Robert Vích, 5641:239–47. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03320-9_23.

———. 2019. *Intencjonalność komunikacyjna gestów wskazujących*. Językoznawstwo Stosowane – Uniwersytet im. Adama Mickiewicza w Poznaniu 30. Poznań: Wydawnictwo Naukowe UAM.

Jarmołowicz-Nowikow, Ewa i Maciej Karpiński. 2011. „Communicative intentions behind pointing gestures in task-oriented dialogues". W *Proceedings | GESPIN*. http://gespin.amu.edu.pl/?q=node/66.

Jockers, Matthew L. 2021. *Text Analysis with R: For Students of Literature.* Springer International Publishing AG.

Jolly, Stephen. 2000. „Understanding Body Language: Birdwhistell's Theory of Kinesics". *Corporate Communications: An International Journal* 5 (3): 133–39. https://doi.org/10.1108/13563280010377518.

Juszczyk, Konrad. 2011. „Multimodalny model mentalny znaczeń w instrukcyjnych aktach dialogowych". *Studia z kognitywistyki i filozofii umysłu* 5 (1): 39–58.

———. 2017. *Mądrość metafory: komunikacyjny model metafory kognitywnej*. Poznań: Coachspace.

Juszczyk, Konrad i Victoria Kamasa. 2016. „Ku metodzie identyfikacji wyrażeń metaforycznych dla polszczyzny na przykładzie rozmów o karierze zawodowej." W *Język a komunikacja*, 37:177–86. TERTIUM.

Juszczyk, Konrad, Barbara Konat i Małgorzata Fabiszak. 2022. „Speakers Who Metaphorize Together – Argue Together: Interaction between Metaphors and Arguments as a Dynamic Discourse Phenomenon", kwiecień. https://doi.org/10.1075/msw.21016.jus.

Karpiński, Maciej. 2006. *Struktura I Intonacja Polskiego Dialogu Zadaniowego*. Seria Językoznawstwo / Uniwersytet im. Adama Mickiewicza

w Poznaniu, nr 26. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

———. 2009a. „From Speech and Gestures to Dialogue Acts". W *Multimodal Signals: Cognitive and Algorithmic Issues*. Red. Anna Esposito, Amir Hussain, Maria Marinaro i Raffaele Martone, 5398:164–69. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-00525-1_16.

———. 2009b. „Preliminary Prosodic and Gestural Characteristics of Instructing Acts in Polish Task-Oriented Dialogues". W *Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions*. Red. A. Esposito, A. Hussain, M. Marinaro i R. Martone, 227–38. LNAIS 641. Berlin- Heidelberg: Springer-Verlag.

Karpiński, Maciej, Agnieszka Czoska, Ewa Jarmołowicz-Nowikow, Konrad Juszczyk i Katarzyna Klessa. 2018. „Aspects of gestural alignment in task-oriented dialogues". *Cognitive Studies | Études cognitives*, nr 18 (grudzień). https://doi.org/10.11649/cs.1640.

Karpiński, Maciej, Ewa Jarmołowicz-Nowikow, Konrad Juszczyk, Zofia Malisz i Michał Szczyszek. 2008. „Rejestracja, transkrypcja i tagowanie mowy oraz gestów w narracji dzieci i dorosłych". *Investigationes Linguisticae* 16: 83–98.

Karpiński, Maciej i Ewa Jarmołowicz-Nowikow. 2010. „Prosodic and Gestural Features of Phrase-Internal Disfluencies in Polish Spontaneous Utterances". W *Proceedings of Speech Prosody 2010*. Chicago.

Karpiński, Maciej, Ewa Jarmołowicz-Nowikow i Agnieszka Czoska. 2015. „Gesture annotation scheme development and application for entrainment analysis in task-oriented dialogues in diverse cultures." W *Proceedings of GESPIN 2015 Conference*, 161–66. Nantes, France.

Karpiński, Maciej i Katarzyna Klessa. 2021. *Linguist in the Field: A Practical Guide to Speech Data Collection, Processing, and Management*. 1. wyd. PL: Wydawnictwo Rys Tomasz Paluszyński. https://doi.org/10.48226/978-83-66666-89-4.

Kehoe, Andrew i M. Gee. 2013. „eMargin: A Collaborative Textual Annotation Tool". Undefined. 2013.

Kendon, Adam. 1990. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. CUP Archive.

———. 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press.

Kendon, Adam, Thomas A. Sebeok i Jean Umiker-Sebeok. 1981. *Nonverbal Communication, Interaction, and Gesture: Selections from SEMIOTICA*. de Gruyter Mouton.

Kiełbawska, Amelia. 2012. *Funkcje komunikacji niewerbalnej w interakcji mówcy i tłumacza*. Kraków: Universitas.
Kipp, Mike, red. 2009. *Multimodal corpora: from models of natural interaction to systems and applications*. Lecture notes in computer science 5509. Berlin ; New York: Springer.
Klessa, Katarzyna. 2016. *Annotation Pro: enhancing analyses of linguistic and paralinguistic features in speech*. Poznań: Wydział Neofilologii UAM.
Klima, Edward S., Ursula Bellugi i Robbin Battison. 1979. *The Signs of Language*. Cambridge, Mass. ; London: Harvard University Press.
Konat, Barbara i Konrad Juszczyk. 2015. „Multimodal communication in career coaching sessions: lexical and gestural corpus study". W *Empirical Methods in Language Studies*. T. 37. Lodz Studies in Language.
Konat, Barbara, John Lawrence, John Park, Katarzyna Budzynska i Chris Reed. 2016. „A corpus of argument networks: Using graph properties to analyse divisive issues." W *LREC 2016*, 3899–3906. ELRA.
Kopp, Richard R. i Michael Jay Craw. 1998. „Metaphoric Language, Metaphoric Cognition, and Cognitive Therapy." *Psychotherapy: Theory, Research, Practice, Training* 35 (3): 306–11. https://doi.org/10.1037/h0087795.
Koszowy, M., S. Oswald, Katarzyna Budzynska, Barbara Konat i P. Gygax. 2022. „A Pragmatic Account of Rephrase in Argumentation: Linguistic and Cognitive Evidence". *Informal Logic* 42 (1): 49–82.
Kousidis, Spyridon, Zofia Malisz, Petra Wagner i David Schlangen. 2013. „Exploring annotation of head gesture forms in spontaneous human interaction."
Kövecses, Zoltán. 2003. *Metaphor and Emotion: Language, Culture, and Body in Human Feeling*.
———. 2009. *Metaphor: A Practical Introduction*. Oxford University Press.
———. 2011. *Język, umysł, kultura: praktyczne wprowadzenie*. Tłum. Anna Kowalcze-Pawlik i Magdalena Buchta. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych Universitas.
Ladewig, Silva H. 2020. „Integrating Gestures". *Integrating Gestures*. https://doi.org/10.1515/9783110668568.
Lakoff, George. 1992. „The Contemporary Theory of Metaphor". W *Metaphor and Thought*. Cambridge University Press.
———. 2011. *Kobiety, ogień i rzeczy niebezpieczne: co kategorie mówią nam o umyśle?* Tłum. Anna Skucińska, Elżbieta Tabakowska, Magdalena Buchta i Agnieszka Kotarba. Językoznawstwo Kognitywne, t. 12. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych Universitas.
Lakoff, George i Mark Johnson. 2010. *Metafory w naszym życiu*. Tłum. Tomasz Krzeszowski. Warszawa: Wydawnictwo Aletheia.

Lane, Hobson, Cole Howard i Hannes Max Hapke. 2019. *Natural language processing in action: understanding, analyzing, and generating text with Python*. Shelter Island, NY: Manning Publications Co.

Lausberg, Hedda, red. 2013. *Understanding body movement: a guide to empirical research on nonverbal behaviour: with an introduction to the NEUROGES coding system*. Frankfurt am Main ; New York: PL Academic Research.

———. 2019. *The NEUROGES® analysis system for nonverbal behavior and gesture: the complete research coding manual including an interactive video learning tool and coding template*. Berlin: Peter Lang Internationaler Verlag der Wissenschaften.

Lausberg, Hedda i Han Sloetjes. 2015. „The Revised NEUROGES–ELAN System: An Objective and Reliable Interdisciplinary Analysis Tool for Nonverbal Behavior and Gesture". *Behavior Research Methods*, październik. https://doi.org/10.3758/s13428-015-0622-z.

Lausberg, Hedda, J. Wietersheim i H. Feiereis. 1996. „Movement Behaviour of Patients with Eating Disorders Ans Inflammatory Bowel Disease. A Controlled Study". *Psychotherapy and Psychosomatics* 65: 272–76.

Lausberg, Hedda, J. Wietersheim, E. Wilke i H. Feiereis. 1988. „Bewegungsbeschreibung psychosomatischer Patienten in der Tanztherapie". *Psychotherapie, Psychosomatik, Medizinische Psychologie* 38: 259–64.

Lecoq, Jacques. 2006. *Theatre of Movement and Gesture*. Routledge.

Li, Xiaoting. 2014. *Multimodality, Interaction and Turn-Taking in Mandarin Conversation*. T. 3. Studies in Chinese Language and Discourse. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/scld.3.

Madelska, Liliana i Małgorzata Witaszek-Samborska. 2015. *Zapis fonetyczny: zbiór ćwiczeń*. Wyd. 7. popr. i uzup. Poznań: Wydawnictwo Naukowe im. Adama Mickiewicza.

Madelska, Liliana. 2005. *Słownik Wariantywności Fonetycznej Współczesnej Polszczyzny*. Kraków: Collegium Columbinum.

Marhula, Joanna i Maciej Rosiński. 2014. „Identifying metaphor in spoken discourse: Insights from applying MIPVU to radio talk data". *Studia Anglica Resoviensia*, nr 85: 32–43.

———. 2019. „Chapter 9. Linguistic Metaphor Identification in Polish". W *Converging Evidence in Language and Communication Research*. Red. Susan Nacey, Aletta G. Dorst, Tina Krennmayr i W. Gudrun Reijnierse, 22:184–202. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/celcr.22.09mar.

Martell, Craig. 2002. „FORM: An Extensible, Kinematically-based Gesture Annotation Scheme." W *Proceedings of the Third International Confer-*

ence on Language Resources and Evaluation (LREC'02). Red. Manuel González Rodríguez i Carmen Paz Suarez Araujo. Las Palmas, Canary Islands – Spain: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2002/pdf/304.pdf.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

———. 2005. *Gesture and thought*. University of Chicago Press.

Miłkowski, Marcin. 2003. „Heterofenomenologia i introspekcja. O możliwości poznania przeżyć świadomych". *Przegląd filozoficzny i literacki* 4 (6): 111–29.

Morgenstern, Aliyah i Susan Goldin-Meadow, red. 2022. *Gesture in language: development across the lifespan*. Language and the human lifespan. Washington , DC: American Psychological Association.

Morris, Desmond. 1994. *Bodytalk: a world guide to gestures*. London: Jonathan Cape.

Mykowiecka, Agnieszka. 2007. *Inżynieria lingwistyczna: komputerowe przetwarzanie tekstów w języku naturalnym*. Podręczniki Akademickie / Polsko-Japońska Wyższa Szkoła Technik Komputerowych, t. 27. Warszawa: Wydawnictwo Polsko-Japońskiej Wyższej Szkoły Technik Komputerowych.

Nacey, Susan, Aletta G. Dorst, Tina Krennmayr i W. Gudrun Reijnierse, red. 2019. *Metaphor Identification in Multiple Languages: MIPVU around the World*. T. 22. Converging Evidence in Language and Communication Research. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/celcr.22.

Nowak, Leszek. 1977. *Wstęp do Idealizacyjnej Teorii Nauki. Wybrane Zagadnienia Metodologii Nauk, Naukoznawstwa i Informacji Naukowej*. Warszawa: Państwowe Wydaw. Naukowe.

Osinga, Douwe. 2018. *Deep Learning Cookbook: Practical Recipes to Get Started Quickly*.

Ostaszewska, Danuta. 2008. *Fonetyka I Fonologia Współczesnego Języka Polskiego*. Wyd. 2, 3 dodr. Warszawa: Wydawnictwo Naukowe PWN.

Piao, Scott, Paul Rayson, Dawn Archer, Francesca Bianchi, Carmen Dayrell, Mahmoud El-Haj, Ricardo-María Jiménez i in. 2016. „Lexical Coverage Evaluation of Large-Scale Multilingual Semantic Lexicons for Twelve Languages". W *In Proceedings of the 10th Edition of the Language Resources and Evaluation Conference (LREC2016*, 2614–19. Portoroz, Slovenia.

Pitcher, Rod. 2013. „Using Metaphor Analysis: MIP and Beyond". *The Qualitative Report* Volume 18 (Article 68): 1–8.

Popper, K. R. 1959. *The logic of scientific discovery*. University Press.
Pragglejaz Group. 2007. „MIP: A Method for Identifying Metaphorically Used Words in Discourse". *Metaphor and Symbol* 22 (1): 1–39.
Przepiórkowski, Adam, Mirosław Bańko, Rafał Górski i Barbara Lewandowska-Tomaszczyk, red. 2012. *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN.
Radford, Alec, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey i Ilya Sutskever. 2022. „Robust Speech Recognition via Large-Scale Weak Supervision".
Ramseyer, Fabian T. 2020. „Motion Energy Analysis (MEA): A Primer on the Assessment of Motion from Video." *Journal of Counseling Psychology* 67 (4): 536–49. https://doi.org/10.1037/cou0000407.
Rodríguez, Lorena Martín, i Christopher Cox. 2023. „Speech-to-text recognition for multilingual spoken data in language documentation". W Proceedings of the Sixth Workshop on the Use of Computational Methods in the Study of Endangered Languages, zredagowane przez Atticus Harrigan, Aditi Chaudhary, Shruti Rijhwani, Sarah Moeller, Antti Arppe, Alexis Palmer, Ryan Henke, i Daisy Rosenblum, 117–23. Remote: Association for Computational Linguistics. https://aclanthology.org/2023.computel-1.17.
Rohrer, Patrick Louis, Elisabeth Delais-Roussarie i Pilar Prieto. 2023. „Visualizing Prosodic Structure: Manual Gestures as Highlighters of Prosodic Heads and Edges in English Academic Discourses". *Lingua* 293 (październik): 103583. https://doi.org/10.1016/j.lingua.2023.103583.
Roy, Deb. 2009. „New Horizons in the Study of Child Language Acquisition". W *Interspeech 2009*, 13–20. ISCA. https://doi.org/10.21437/Interspeech.2009-3.
Rutkowski, Paweł, Sylwia Łozińska, Uniwersytet Warszawski i Wydział Polonistyki. 2014. *Lingwistyka przestrzeni i ruchu: komunikacja migowa a metody korpusowe*. Warszawa: wydano nakładem Wydziału Polonistyki Uniwersytetu Warszawskiego.
Saussure, Ferdinand de. 2002. *Kurs Językoznawstwa Ogólnego*. Wyd. 3. Warszawa: Wydaw. Naukowe PWN.
Schiel, Florian. 1999. „Automatic Phonetic Transcription of Non-Prompted Speech". https://doi.org/10.5282/UBM/EPUB.13682.
Searle, John R. 1970. *Speech Acts: An Essay in the Philosophy of Language*. New Ed edition. Cambridge: Cambridge University Press.
Shaughnessy, John J, Jeanne S Zechmeister i Eugene B Zechmeister. 2007. *Metody badawcze w psychologii*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
Sikorski, Wiesław. 2005. *Gesty zamiast słów: psychologia i trening komunikacji niewerbalnej*. Kraków: Oficyna Wydawnicza „Impuls".

Steen, Gerard J. 1999. „From Linguistic to Conceptual Metaphor in Five Steps". W *Current Issues in Linguistic Theory*. Red. Raymond W. Gibbs i Gerard J. Steen, 175:57. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.175.05ste.

———. 2009. *Finding Metaphor in Grammar and Usage: A Methodological Analysis of Theory and Research*. Paperback ed. Converging Evidence in Language and Communication Research 10. Amsterdam: Benjamins.

Steen, Gerard J., Aletta G. Dorst, J. Berenike Herrmann, Anna Kaal, Tina Krennmayr i Trijntje Pasma. 2010. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. John Benjamins Publishing.

Stokoe, William C. 1960. „Sign Language Structure: An Outline of the Visual Communication System of the American Deaf." W *Studies in Linguistics, Occasional Papers*. T. 8. New York: University of Buffalo.

Stoltzfus, Tony. 2012. *Sztuka zadawania pytań w coachingu: jak opanować najważniejszą umiejętność coacha?* Tłum. Bożena Olechnowicz. Wrocław: Aetos Media.

Such, Jan i Małgorzata Szcześniak. 2006. *Filozofia nauki*. Wyd. 5. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

Szczepaniak, Agnieszka. 2017. *Gesty emblematyczne w międzykulturowej komunikacji niewerbalnej: polsko-grecko-brytyjskie studium porównawcze i gestownik*. Kaliskie Towarzystwo Przyjaciół Nauk.

Szczyszek, Michał. 2013. *O porozumiewaniu się międzyludzkim: wariant mówiony języka polskiej wspólnoty komunikatywnej : słowotwórstwo, leksyka, składnia*. Poznań: Wydawnictwo Rys.

Szymanek, Krzysztof. 2012. *Sztuka argumentacji: słownik terminologiczny*. Wyd. 2, 5 dodr. Warszawa: Wydawnictwo Naukowe PWN.

Szymczak, Mieczysław, Stanisław Bik, Celina Szkiłądź i Halina Chociłowska, red. 1992. *Słownik języka polskiego. T. 1-3*. Wyd. 7 zm. i popr. Warszawa: Wydaw. Naukowe PWN.

Śledziński, Daniel. 2022. *Wielowarstwowy Model transkrypcji tekstu w Języku Polskim.* Wyd. I. ed. Poznań: Wydawnictwo Rys. https://doi.org/10.48226/978-83-67287-28-9.

Tay, Dennis. 2013. *Metaphor in Psychotherapy: A Descriptive and Prescriptive Analysis*. John Benjamins Publishing.

Thiran, Jean-Philippe, Hervé Bourlard i Ferran Marques. 2010. *Multimodal Signal Processing: Theory and Applications for Human-Computer Interaction*. EURASIP and Academic Press Series in Signal and Image Processing. Oxford, UK Boston: Academic Press.

Tomasello, Michael i Josep Call. 2019. „Thirty Years of Great Ape Gestures". *Animal Cognition* 22 (4): 461–69. https://doi.org/10.1007/s10071-018-1167-1.

Trippel, Thorsten, Dafydd Gibbon, Alexandra Thies, Jan-Torsten Milde, Karin Looks, Benjamin Hell i Ulrike Gut. 2004. „CoGesT: a Formal Transcription System for Conversational Gesture". W *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. Red. Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa i Raquel Silva. Lisbon, Portugal: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2004/pdf/650.pdf.

Turchet, Philippe. 2006. *Mowa ciała: zrozumieć człowieka po jego gestach*. Tłum. Elżbieta Ptaszyńska-Sadowska. Warszawa: Klub dla Ciebie.

———. 2009. *Le langage universel du corps*. les Éd. de l'Homme.

Valenzuela, Javier i Cristina Soriano. 2005. „Cognitive Metaphor and Empirical Methods". *Barcelona Language and Literature Studies.*

Van Atteveldt, Wouter, Damian Trilling i Carlos Arcíla Calderón. 2022. *Computational Analysis of Communication*. Wiley Blackwell.

Wells, John. 2000. „Computer-coding the IPA: a proposed extension of SAMPA". *UCL Phonetics and Linguistics. University College London.* (blog). https://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm.

Wierzba, Małgorzata, Monika Riegel, Jan Kocoń, Piotr Miłkowski, Arkadiusz Janz, Katarzyna Klessa, Konrad Juszczyk i in. 2021. „Emotion norms for 6000 Polish word meanings with a direct mapping to the Polish wordnet". *Behavior Research Methods* 54 (5): 2146–61. https://doi.org/10.3758/s13428-021-01697-0.

Wiśniewski, Marek. 2007. *Zarys Fonetyki I Fonologii Współczesnego Języka Polskiego: (skrypt Dla Studentów Filologii Polskiej)*. Wyd. 5. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika.

Załazińska, Aneta. 2001. *Schematy Myśli Wyrażane W Gestach : Gesty Metaforyczne Obrazujące Abstrakcyjne Relacje I Zasoby Podmiotu Mówiącego*. Kraków: Universitas.

Załazińska, Aneta. 2006. *Niewerbalna Struktura Dialogu: W Poszukiwaniu Polskich Wzorców Narracyjnych I Interakcyjnych Zachowań Komunikacyjnych*. Pragmatyka i Semantyka Mowy 3. Kraków: TAiWPN UNIVERSITAS.

Spis tabel

Spis grafik