

Piotr Mostowski

## Korpus polskiego języka migowego — zarys projektu

### 1. Wstęp

Budowa korpusu polskiego języka migowego rozpoczęła się w 2010 roku. Inicjatywa ta była następstwem dwóch wydarzeń. W strukturze organizacyjnej Wydziału Polonistyki Uniwersytetu Warszawskiego wyodrębniono wówczas Pracownię Lingwistyki Migowej (PLM), pierwszą w Polsce jednostkę naukową zajmującą się badaniami nad komunikacją wizualno-przestrzenną. W tym też roku Fundacja na rzecz Nauki Polskiej przyznała kierownikowi PLM dr. Pawłowi Rutkowskiemu trzyletni grant w ramach programu Focus: „Kategoryzacja gramatyczna przez przestrzeń i ruch w polskim języku migowym”, który stał się punktem wyjścia do stworzenia korpusu tekstów służącego dalszej analizie językoznawczej. W niniejszym rozdziale opisane zostały kolejne etapy realizacji tego przedsięwzięcia, od wstępnych założeń metodologicznych aż po stan obecny. Każde zagadnienie zostało przedstawione w szerszej perspektywie uwzględniającej liczne problemy towarzyszące pracom nad korpusami (nie tylko języka migowego).

### 2. Założenia wstępne

Pierwszym etapem projektu było dokładne zaplanowanie struktury korpusu. Jak pokazują przykłady z innych krajów, jest to fundament dalszych prac. Należało przy tym uwzględnić ograniczenia dotyczące lokalnej sytuacji — budżetu, zaplecza sprzętowego, sytuacji środowiska głuchych w Polsce (brak ustawy o polskim języku migowym, dominacja systemu językowomigowego, niski poziom edukacji). Nie zakładano wielkości korpusu porównywalnej do korpusów europejskich. Od początku jednak uwzględniano możliwości

dalszej rozbudowy projektu tak, aby osiągnąć zadowalające wyniki. Logiczna wydała się więc etapowa realizacja pracy, analogicznie np. do korpusu NGT (*Nederlandse Gebarentaal* — niderlandzki język migowy). Najważniejsze było jednak prawidłowe opracowanie architektury korpusu, systemu wyszukiwania informatorów, procedur elicytacyjnych, a także określenie zasad przechowywania materiału, wytycznych do anotacji i publikowania treści. Od przyjętej w tym momencie metodologii zależało, czy w przyszłości będzie można płynnie powiększać zgromadzony materiał. Już na najwcześniejszym etapie prac należało zatem przede wszystkim odpowiedzieć na pytanie o przeznaczenie zebranego materiału.

Podstawowym celem korpusu było dostarczenie materiału do badań na potrzeby grantu — sporządzenie szczegółowego opisu najbardziej charakterystycznych aspektów gramatyki polskiego języka migowego, a po drugie — zanalizowanie tych danych z perspektywy językoznawstwa teoretycznego i typologicznego. Nie można jednak ograniczać przyszłych zastosowań korpusu tylko do źródła materiału lingwistycznego. Kolejnym rezultatem powinien być słownik dwujęzyczny (PJM-polszczyzna i polszczyzna-PJM), którego brak zauważa wiele osób zajmujących się polskim językiem migowym. Inne możliwości wykorzystania są takie same, jak w wypadku zagranicznych korpusów migowych — pomoc w nauczaniu języka migowego jako obcego, teksty do szkoleń tłumaczy (zob. rozdział *Rola materiałów korpusowych w dydaktyce przekładu języka migowego* autorstwa Aleksandry Kalaty-Zawłockiej w niniejszym tomie) oraz archiwum kultury Głuchych (przy założeniu powszechnego dostępu online). Ze względu na planowaną szeroką użyteczność najwłaściwsze wydawało się wzorowanie na procedurach wykorzystanych w korpusie niemieckiego języka migowego (DGS, niem. *Deutsche Gebärdensprache*), przy stosowaniu zaleceń standaryzacji spisanych przez zespół Sign Linguistics Corpora Network (Crasborn, 2010) oraz odwoływaniu się do dokonań innych ośrodków badawczych.

W dalszej części niniejszego rozdziału opisano szczegółową procedurę tworzenia korpusu polskiego języka migowego. Poszczególne etapy prac przedstawione są w kolejności od wyszukania informatora do publikacji gotowego materiału. Nie jest to porządek chronologiczny (w którym najpierw powinno być opisane przygotowanie studia i materiałów elicytacyjnych, testy, a dopiero potem poszukiwanie ochotników), ale został on wybrany ze względu na swoją czytelność i logikę. Pominięto detale techniczne (np. modele kamer), które, chociaż bardzo ważne, nie są istotne z obranej perspektywy. Myślą przewodnią niech będą słowa J. Sinclaira (2005): ważne jest, aby unikać perfekcjonizmu w budowaniu korpusu. Jest to niesprecyzowane zagadnienie i nikt nadal nie wie, jak powinien wyglądać korpus idealny.

### 3. Charakterystyka korpusu polskiego języka migowego

Zgodnie z klasyfikacjami obowiązującymi w lingwistyce korpusowej korpus polskiego języka migowego jest (Waliński, 2005):

- korpusem ogólnym (odzworowanie podstawowego, powszechnego poziomu języka z pominięciem specjalistycznych żargonów, dialektów),
- korpusem zawierającym teksty w całości (teksty w całościowym ich brzmieniu),
- zrównoważonym korpusem referencyjnym (jak najwierniej odwzorowuje naturalne struktury w języku przy zachowaniu ich proporcji),
- korpusem jednojęzycznym (tylko polski język migowy),
- korpusem anotowanym,
- korpusem synchronicznym (teksty współczesne, z jednego odcinka czasowego).

### 4. Informatorzy

Nie wiadomo dokładnie, ile jest w Polsce osób niesłyszących. Wynika to z bardzo pojemnego znaczenia kategorii „niesłyszący” — trudno określić, jaki ubytek słuchu kwalifikuje do tego miana. Ze względu na to, jaką definicję głuchoty i osoby niesłyszącej przyjąć, uzyska się inną statystykę. Jeszcze trudniejsza jest odpowiedź na pytanie, ile osób zna polski język migowy. Po pierwsze, nie tylko głusi są jego użytkownikami, lecz także słyszące dzieci głuchych rodziców i nieliczni słyszący niespokrewnieni z głuchymi. Po drugie, dużo osób korzysta z systemu miganego albo z pidżynu łączącego elementy SJM (systemu językowo-migowego) i PJM.

Korpus powinien zawierać tekst migowy jak najbliższy „naturalnemu” polskiemu językowi migowemu. Mimo iż jest to założenie, którego zdaniem wielu twórców korpusów nie można spełnić w całości, należy dążyć do pozyskania danych jak najbardziej spontanicznych. Jak pokazują doświadczenia innych zespołów, kluczowa jest rola głuchego moderatora, który komunikuje się z uczestnikami sesji w naturalnym języku migowym (a tym samym narzuca niejako ten język jako obowiązujący). Bardzo ważny jest także właściwy materiał elicytacyjny, w którym nie występują elementy polszczyzny.

Pozyskiwanie informatorów w warunkach polskich odbywało się analogicznie do poszukiwań holenderskich. Ze względów finansowych niewykonalne było specjalistyczne szkolenie osób kontaktowych (zob. korpus DGS), które

najpierw wyszukałyby ochotników w społeczności lokalnej, a potem zrealizowały nagranie w danym miejscu. Skorzystano więc z ogłoszeń na portalach związanych ze społecznością głuchych. Głusi biorący udział w nagraniu informowali też o przedsięwzięciu swoich znajomych. System taki zadziałał w innych krajach realizujących projekty korpusowe — z biegiem czasu problemem nie było znajdowanie chętnych, ale konieczność odmawiania coraz większej ilości ochotników.

Starano się zachować odpowiednie proporcje dotyczące:

- płci uczestników (jak najbliższa stosunkowi 1:1),
- pochodzenia (informatorzy pozyskiwani byli z kilku regionów Polskich — ze względu na historię polskiego języka migowego),
- wieku (ta sama liczba osób w każdej z kategorii wiekowych: 18–30, 30–45, 45–60, 60+).

Przy wyborze informatorów zwracano uwagę na cechy opisane przez Fabisiak (2010, s. 342). Poszukiwano więc osób:

- będących rodzimymi użytkownikami języka, którzy nabyli go w procesie naturalnej akwizycji,
- nastawionych pozytywnie do języka migowego i akceptujących swoją głuchotę,
- należących do społeczności głuchych, utrzymujących kontakt z innymi niesłyszącymi, najchętniej także działających aktywnie w środowisku,
- reprezentujących przeciętny poziom kompetencji językowej w języku polskim,
- zróżnicowanych pod względem wieku i wykształcenia oraz pochodzenia.

W początkowej fazie projektu nagrano 60 osób. Ochotnicy dobrani w pary ze względu na zbliżony wiek i pochodzenie byli zapraszani w dogodnych terminach do studia nagraniowego. Przed sesją każda osoba musiała wypełnić kwestionariusz zawierający pytania o poziom znajomości PJM, historię edukacji, kontakty z innymi niesłyszącymi itp. Koniecznym warunkiem było podpisanie zgody na upublicznienie swojego wizerunku. Zgodnie z art. 23 kodeksu cywilnego (Dz.U. z dnia 18 maja 1964 r.) wizerunek człowieka jest zaliczany do jego dóbr osobistych i podlega ochronie:

Art. 23. Dobra osobiste człowieka, jak w szczególności zdrowie, wolność, cześć, swoboda sumienia, nazwisko lub pseudonim, [...] pozostają pod ochroną prawa cywilnego niezależnie od ochrony przewidzianej w innych przepisach.

Prawo do wizerunku jest niejako oddzielne od innych dóbr chronionych prawem — ochrona przysługuje w przypadku bezprawnego posłużenia się wizerunkiem przez osobę trzecią, nawet jeśli nie doszło do naruszenia innych dóbr.

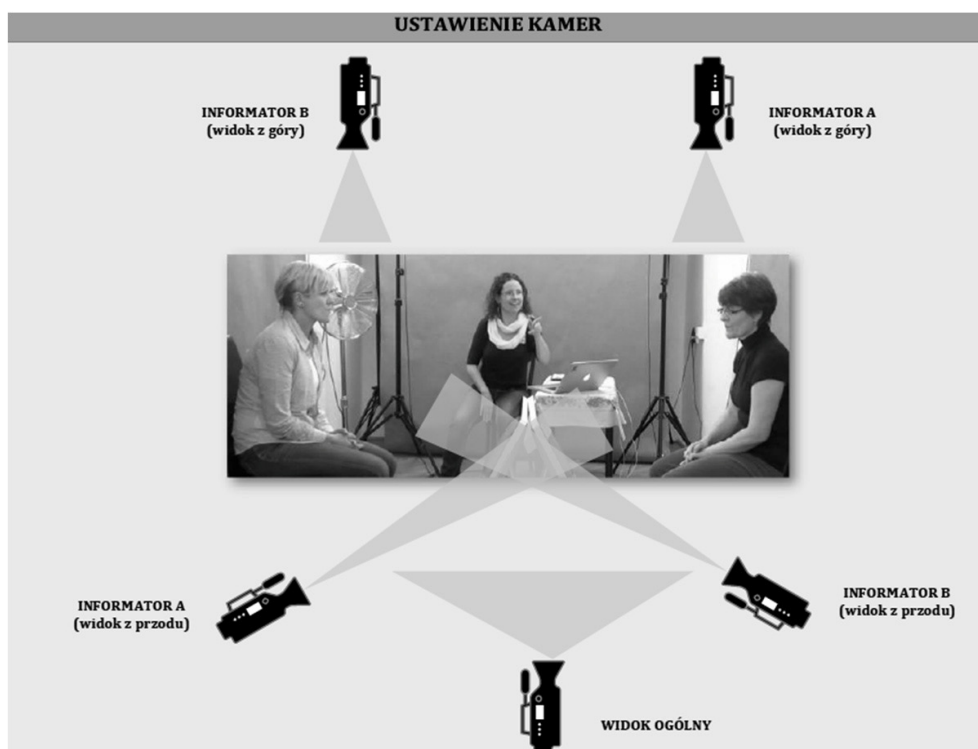
Artykuł 81 ust. 1 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz.U. z 2000 r., Nr 80, poz. 904 ze zm.) [...] stanowi, iż rozpowszechnianie wizerunku wymaga zezwolenia osoby na nim przedstawionej. [...] Zezwolenie na rozpowszechnianie wizerunku nie wymaga pisemnej zgody osoby na nim przedstawionej. Jednakże nie mogą zaistnieć jakiegokolwiek wątpliwości, iż zgoda taka została udzielona, a ponadto, zgoda taka nie może być ogólna. Zezwolenie powinno więc dotyczyć zarówno czasu, jak i miejsca rozpowszechniania wizerunku oraz formy, kontekstu i celu jego prezentacji. (Przegaliński, 2006)

Osoba nagrywana była więc szczegółowo poinformowana o tym, czym jest korpus, jak wygląda sesja nagraniowa i co dalej stanie się z klipami wideo. Konieczne były dwie wersje językowe takiego wyjaśnienia: w języku polskim i w PJM. Następnie uczestnik proszony był o podpisanie zezwolenia na wykorzystanie wizerunku.

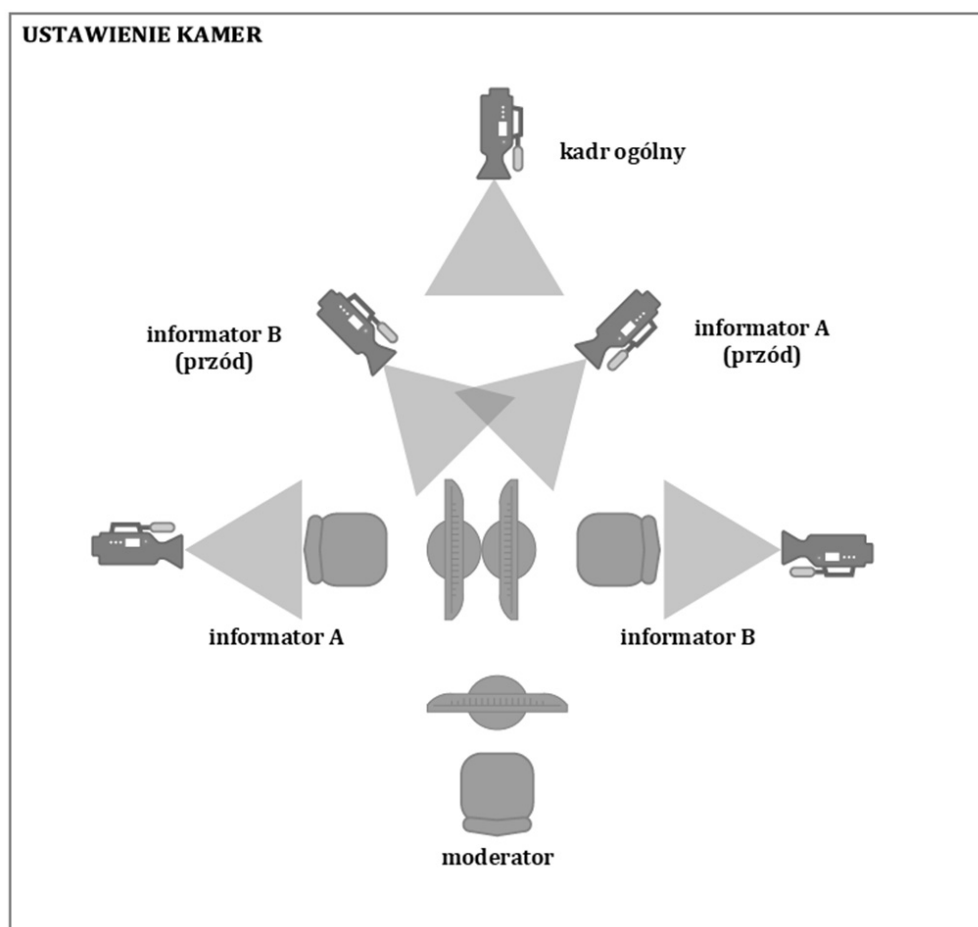
## 5. Studio nagraniowe

Główne studio nagraniowe znajdowało się w Warszawie, gdzie mieści się siedziba Pracowni Lingwistyki Migowej. Studio mobilne umożliwiała z kolei zbieranie danych podczas sesji wyjazdowych.

Do nagrań wykorzystano zestaw analogiczny do innych projektów korpusowych. Pięć kamer rozstawionych było w następujący sposób: 2 kamery skierowane *en face* na informatorów, 2 kamery podsufitowe nagrywające informatorów z góry oraz 1 kamera, która obejmowała wszystkie postaci biorące udział w sesji. Był to sprzęt rejestrujący obraz w jakości HD (1080p). Dzięki temu możliwe było np. późniejsze zbliżanie twarzy w celu precyzyjnego odczytania mimiki. Zakupiono również zestaw świateł studyjnych i teł. Zrezygnowano jednak z kamer stereoskopowych i zestawu *motion capture*.



Rys. 1. Studio nagraniowe



Rys. 2. Studio nagraniowe

## 6. Elicytacja

Elicytacja jest jednym z najważniejszych etapów tworzenia korpusu — od jego przygotowania zależy, w jakiej mierze zebrane teksty będą zbliżone do naturalnego języka migowego. Środki mające zapewnić spontaniczność wypowiedzi migowych to m.in.: obecność wyłącznie osób głuchych podczas nagrania, jak najmniejsza ilość tekstu w zadaniach, wskazówki do zadań nagrane w polskim języku migowym. Obecny podczas sesji głuchy moderator powinien ograniczyć do minimum swoje ingerencje w to, co dzieje się w studiu. Jego zadanie to kontrolowanie czasu, wyświetlanie kolejnych poleceń i reagowanie w wyjątkowych sytuacjach (np. kiedy uczestnik mimo obejrzanego wprowadzenia nadal nie rozumie zadania albo ma dodatkowe pytania).

Zadania wyświetlano na monitorach umieszczonych przed każdym z uczestników — przebieg nagrania sterowany zdalnie z komputera moderatora. Zestaw materiałów elicytacyjnych powstał jako kompilacja elementów używanych w innych ośrodkach badawczych.

Kolejność zadań elicytacyjnych:

1. Informacje organizacyjne. Nagranie informujące, czym jest korpus i co się stanie z danymi. Wyjaśnienie, jak będzie przebiegała sesja i jaka jest rola moderatora.
2. Zapoznanie się. Uczestnicy na początku przedstawiają się i podają swój znak migowy oraz jego pochodzenie. Przydomek migowy służący do identyfikacji danej osoby w społeczności głuchych jest bardzo ważnym elementem kultury. Zebranie kilkudziesięciu przykładów wraz z ich opisem to doskonały materiał do badań antroponimicznych i etymologicznych. Największą wartość klipy te będą miały dla dokumentacji kultury Głuchych i zachowania jej ciągłości.
3. Dowcip. Informatorzy opowiadają dowcip, który mieli przygotować w domu. Jest to część rozgrzewki — uczestnicy mogą się pewnie poczuć i rozluźnić, co wpływa na dalszy przebieg sesji. Oprócz tego celem jest archiwizacja tekstów kultury Głuchych.
4. Dyskusja. Uczestnicy proszeni są o opowiedzenie swoich doświadczeń dotyczących funkcjonowania w społeczności niesłyszących (np. pobytu w szkole z internatem, udziału w ważnych wydarzeniach). Jest to kolejne zadanie nastawione bardziej na kulturę Głuchych niż na lingwistykę. Identyczny punkt pojawia się w korpusach innych krajów, dlatego zebrany materiał może posłużyć do prac porównawczych.
5. Zadanie kalendarzowe. Punkt zaczerpnięty z procedur elicytacyjnych stosowanych przez zespół Instytutu Niemieckiego Języka

Migowego i Komunikacji Niesłyszących (niem. *Institut für Deutsche Gebärdensprache und Kommunikation Gehörloser*, IDGS) z Uniwersytetu w Hamburgu (por. rozdział *Badania korpusowe a lingwistyka migowa* w niniejszym tomie). Uczestnikom zostają wyświetlone ich plany tygodnia, następnie muszą ustalić dogodny termin na spotkanie. Zamierzeniem jest wywołanie wyrażen charakterystycznych dla negocjacji, określeń temporalnych oraz nazw czynności.

- 6a. Historyjka *Gdzie jesteś, żabo?* (*Frog, where are you?*, autor: Meyer). Informator A ogląda krótką historyjkę obrazkową o chłopcu poszukującym zaginionej żaby, a następnie opowiada ją partnerowi. Historyjka była wykorzystywana w wielu projektach lingwistyki migowej. Na jej podstawie badano m.in. użycie klasyfikatorów, anafory i deiktyczność. Zarazem jest to doskonały materiał do badań międzyjęzykowych — nie tylko międzymigowych, lecz także międzymodalnych, ponieważ historyjka *Gdzie jesteś, żabo?* była wykorzystywana także do badań języków fonicznych (głównie ich odmian dziecięcych).
- 6b. Kreskówka *Tweety i Sylwester* (odcinek *Canary row*). Zadanie analogiczne do 6a, tym razem informator B opowiada treść obejrzanej animacji. Ten sam odcinek był wykorzystywany w innych analizach językoznawczych — jest cennym materiałem m.in. w kontekście badań porównujących funkcjonowanie klasyfikatorów w różnych językach migowych.
7. Dyskusja. Tym razem rozmowa ma dotyczyć jednego z zaproponowanych kontrowersyjnych tematów dotyczących Głuchych i języka migowego, takich jak przyszłość języka migowego lub powszechna implantacja. Celem jest sprowokowanie konwersacji nacechowanej emocjonalnie.
8. Swobodna rozmowa. Punkt zaczerpnięty z procedur elicytacyjnych stosowanych w Hamburgu. Moderator opuszcza pomieszczenie na 15 minut, poinformowawszy uczestników o włączonych kamerach. Nie ma żadnych innych bodźców i wytycznych. Zamierzeniem jest zarejestrowanie swobodnej konwersacji na dowolne tematy.
9. Znaki izolowane. Prezentacja kilkunastu zdjęć/obrazków (ewentualnie polskiego słowa w celu sprecyzowania pojęcia). Uczestnicy są proszeni o zamiganie znaku odnoszącego się do danego przedmiotu/pojęcia. Celem tego zadania jest zgromadzenie materiału do badań leksyko-graficznych, w tym do opracowań na temat regionalizmów w PJM.
10. Komiksy. Informator ma za zadanie opowiedzieć partnerowi historię z krótkiego komiksu. Wykorzystano paski komiksowe z Myszka Miki i Kaczorem Donaldem — w żadnym nie pojawiają się słowa. Pozyskany materiał pozwala analizować struktury narracyjne, a także użycie klasyfikatorów i przestrzeni topograficznej.



11. Klipy. Informator ma za zadanie zrelacjonować partnerowi kilkusekundowy klip wideo. Materiał elicytacyjny dobrano w celu wywołania struktur opisujących czynności i relacje przestrzenne.
12. Znaki. Zadanie zaczerpnięte z procedur elicytacyjnych stosowanych w Hamburgu. Prezentowane są różne nietypowe znaki informacyjne (w większości odnoszące się do zakazów lub nakazów). Uczestnicy muszą zdecydować, jakie jest ich znaczenie. Celem zadania jest zebranie materiału do badań nad negacją oraz konstrukcjami wyrażającymi zakaz lub powinność.
13. Ważne wydarzenia z przeszłości. Zadanie wzorowane na analogicznym poleceniu z procedury elicytacyjnej stosowanej w Hamburgu. Wyświetlane są plansze ze zdjęciami powiązаныmi z ważnymi wydarzeniami historycznymi. Informatorzy mają za zadanie opowiedzieć partnerowi, co robili, gdy dane wydarzenie miało miejsce. Ze względu na zróżnicowany wiek uczestników należy wybrać zdarzenia, które na pewno będą przez nich pamiętane. Przykłady z historii najnowszej to: katastrofa samolotu prezydenckiego w Smoleńsku, śmierć Jana Pawła II, atak na World Trade Center. Do wydarzeń odleglejszych należą m.in.: upadek muru berlińskiego, ogłoszenie stanu wojennego, wybór papieża Polaka itp. Pozyskane nagrania pozwalają analizować struktury gramatyczne służące wyrażaniu czasu przeszłego.
- 14a. Film *Pear story*. Punkt analogiczny do 6a–6b. Tym razem informator A ogląda krótki film, którego fabułę ma zreferować partnerowi. Film wykorzystywany jest przez wielu badaczy analizujących kognitywne, lingwistyczne i kulturowe aspekty narracji w różnych językach fonicznych (Chafe, 1980). Głównym argumentem za skorzystaniem z tego akurat klipu są badania międzyjęzykowe (i międzymodalnościowe).
- 14b. Fragment filmu *Brzdąc* (1921). Wprowadzenie do scenariusza sesji nagraniowej filmu Charliego Chaplina uzupełniło materiały elicytacyjne o bodźce o charakterze komediowym. W projektach korpusowych realizowanych w Holandii i Niemczech użyto w tym celu fragmentów sitcomów tworzonych przez Głuchych. W Polsce materiały takie nie istnieją.
15. Plansze tematyczne. Zadanie wzorowane na procedurze elicytacyjnej stosowanej w Hamburgu. Wyświetlane są plansze z kolażem ilustracji dotyczących różnych dziedzin życia (np. polityka, czas wolny). Uczestnicy opowiadają o swoich skojarzeniach z danym tematem. Celem zadania jest zgromadzenie słownictwa z poszczególnych dziedzin do wykorzystania w pracach słownikarskich.
16. Instrukcja. Punkt zaczerpnięty z procedur elicytacyjnych stosowanych w Hamburgu. Jeden z informatorów jest proszony o przedstawienie

krok po kroku, jak wykonuje się daną czynność, np. zakup biletu lotniczego przez Internet, przygotowanie potrawy. Pozwala to zgromadzić materiał ilustrujący konstrukcje sekwencyjne i inne środki językowe służące strukturyzacji wypowiedzi.

17. Specjalności regionalne. Zadanie inspirowane analogicznym poleceniem z procedury elicytacyjnej stosowanej w Hamburgu. Uczestnicy są proszeni o opowiedzenie partnerowi o specjalnościach swojego regionu (potrawach, zabytkach, zwyczajach, wydarzeniach kulturalnych itp.). Pozwala to na pozyskanie danych językowych przydatnych w badaniach dotyczących regionalizmów oraz mechanizmów nazwotwórczych.
18. Geografia. Głusi mają podać migowe odpowiedniki wyświetlonych nazw geograficznych. Zadanie wprowadzone na potrzeby stworzenia słownika polskiego języka migowego.
19. Opowiadanie filmu *Signs*. Zadanie zaczerpnięte z procedury elicytacyjnej stosowanej w Hamburgu. Uczestnicy oglądają krótki film zmontowany na potrzeby projektu na podstawie reklamy napoju Schweppes, a następnie rozmawiają na ten temat. Zadanie opcjonalne — pomijane w wypadku braku czasu. Celem tego punktu jest pozyskanie słownictwa dotyczącego sfery emocjonalnej człowieka oraz konstrukcji służących wyrażaniu przypuszczeń.
20. Nowe i stare znaki. Zadanie inspirowane procedurą elicytacyjną stosowaną w Hamburgu. Rozmowa na temat starych znaków, które wyszły już z użycia oraz znaków używanych głównie przez młodzież. Brak materiału wywołującego — swobodna dyskusja. Zgromadzony materiał może zostać wykorzystany w badań diachronicznych i socjolingwistycznych. Stanowi też dokumentację kultury Głuchych.
21. Opowiadanie historyjki zamiganej w PJM. Uczestnik powtarza partnerowi obejrzaną opowieść (zamiganą przez natywnego użytkownika PJM). Zadanie pozwala prześledzić mechanizmy towarzyszące odtwarzaniu tekstu migowego przez inną osobę migającą.
22. Mapa. Na wyświetlonej mapie informator A ma zaznaczoną pewną trasę oraz mijane obiekty (np. poczta, kościół, rynek). Proszony jest o przekazanie tych informacji partnerowi. Zadaniem informatora B jest zaznaczenie poszczególnych obiektów na swojej mapie. W ten sposób wywoływane są dane ilustrujące sposoby wyrażania relacji przestrzennych oraz leksyka dotycząca przestrzeni miejskiej.
23. *Baranek Shaun*. Informatorzy opowiadają sobie nawzajem obejrzany odcinek kreskówki *Baranek Shaun*. Pozyskiwane dane językowe to m.in. konstrukcje klasyfikatorowe i złożone struktury narracyjne.

24. Ewaluacja. Na koniec spotkania informatorzy są poproszeni o podzielenie się swoimi uwagami i opiniami dotyczącymi poszczególnych zadań. Mogą też dowolnie skomentować całe przedsięwzięcie. Uzyskane informacje zwrotne pomagają ulepszyć procedurę elicytacyjną. Nagrania z tego fragmentu sesji nie będą włączone do korpusu.

Zadania 21–22 są zaczerpnięte z projektu Dicta-Sign (opisanego w tekście Piotra Mostowskiego *Badania korpusowe a lingwistyka migowa* w niniejszym tomie). Ich wykorzystanie w korpusie PJM jest wskazane ze względu na perspektywę współpracy międzynarodowej oraz badań porównawczych.

Przed rozpoczęciem właściwych nagrań wszystkie zadania zostały przetestowane przez niesłyszących członków zespołu Pracowni Lingwistyki Migowej. Na tej podstawie ustalono czas trwania każdego z nich oraz zaplanowano przerwy. Przeprowadzono dwojakie testy:

- wstępne (test każdego zadania osobno) — Czy wszystkie materiały wyświetlają się prawidłowo? Czy nagrane instrukcje są czytelne i zrozumiałe? Czy udzielanie odpowiedzi na pytania nie jest kłopotliwe? Czy zadania są zrozumiałe bez względu na poziom wykształcenia uczestników? Ile trwa realizacja zadania? Czy udaje się pozyskiwać oczekiwany materiał lingwistyczny?
- końcowe (test całej sesji) — Ile trwa realizacja zadania w trakcie sesji? Ile trwa cała sesja? Czy zauważalna jest interferencja pomiędzy kolejnymi zadaniami? Czy komunikacja z moderatorem przebiega bezproblemowo? Czy przerwy zaplanowane są we właściwych momentach?

Na podstawie testów usprawniono procedurę elicytacyjną przez zmodyfikowanie kolejności zadań oraz ich zawartości. Dodatkową pomocą były opinie ewaluacyjne właściwych informatorów (patrz pkt 24. procedury elicytacyjnej). Oprogramowaniem do kontrolowania całej sesji jest *Session Director* udostępniony zespołowi PLM przez hamburski IDGS.

## 7. Wstępna obróbka i przechowywanie nagrań

Do przechowywania tak obszernego zbioru danych wykorzystano infrastrukturę informatyczną Uniwersytetu Warszawskiego. Nagrania z pięciu kamer po skopiowaniu zostały umieszczane na serwerach uniwersyteckich. Dodatkowo wykonano kopię zapasową całego materiału. Na potrzeby prac anotacyjnych pliki z nagraniami w jakości HD zostały skompresowane.

Usunięto z nich także zbędne fragmenty (np. nagrania rejestrowane w trakcie przerw). Nieobrobione dane z kamer nie zostały jednak skasowane — jakość HD jest niezbędna przy tworzeniu materiałów audiowizualnych wykorzystujących dane korpusowe (np. prezentacji konferencyjnych lub materiałów szkoleniowych).

## 8. Publikacja i perspektywy

Korpus z anotacją\* powinien być (przynajmniej w części) dostępny w Internecie. Ze względu na czasochłonność procesu anotacyjnego nastąpi to najwcześniej po kilku latach od ukończenia zbierania materiałów. Do tego czasu opublikowane mogą zostać poszczególne sesje albo fragmenty nagrań (np. pojedyncze zadania elicytacyjne). Ewentualna publikacja nastąpi na licencji Creative Commons.

Udostępnienie materiału korpusowego w Internecie będzie z jednej strony zwieńczeniem pewnego etapu prac, z drugiej – powinno stanowić impuls do jeszcze szerszego wykorzystania zebranych danych. Naturalnym następstwem powstania korpusu będzie stworzenie pierwszego słownika PJM opartego w pełni na danych empirycznych. Równoległe musi postępować coraz dokładniejsza anotacja, dzięki której będzie można opracowywać m.in. materiały lektoratowe, teksty szkoleniowe dla tłumaczy, szczegółowe analizy gramatyczne.

W sferze pozalingwistycznej warto wspomnieć o roli korpusu w dokumentacji kultury Głuchych. Jest to skarbiec tekstów języka migowego, swoista biblioteka. Nie istnieje wersja pisana PJM, a zatem to właśnie nagrania wideo są jedyną skuteczną metodą przekazywania treści kulturowych kolejnym pokoleniom. Część zadań elicytacyjnych zorientowana jest właśnie na ten aspekt. W przyszłości powinny one stać się obiektem szczegółowego opisu kulturoznawczego. Jest to też argument przemawiający za pełnym dostępem do nagrań przez Internet — głusi z całej Polski powinni mieć możliwość czerpania z tych zasobów.

Jeśli korpus PJM podzieli losy podobnych korpusów zagranicznych, to w przyszłości posłuży rozwijaniu technologii rozpoznawania obrazu i automatycznej analizy języka migowego. W Polsce istnieje wiele ośrodków badawczych (w szczególności politechnicznych), które mogą wykazać zainteresowanie tym zagadnieniem.

---

\* Anotacja jest tak ważnym etapem w tworzeniu korpusu, że został jej poświęcony osobny rozdział niniejszego tomu (*Anotacja korpusu PJM* autorstwa Joanny Filipczak).

Pierwszy zamknięty etap budowy korpusu pokazał, jak wielki potencjał kryje się w tym przedsięwzięciu. Zasadne wydaje się stwierdzenie, że już teraz stało się ono kamieniem milowym w historii polskiej lingwistyki migowej.

## Bibliografia

- Chafe, Wallace, red. (1980). *The Pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood: Ablex.
- Crasborn, Onno (2010). The Sign Linguistics Corpora Network: Towards standards for signed language resources. W N. Calzolari i in. (red.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, ss. 457-460. Paryż: ELRA.
- Fabisiak, Sylwia (2010). Języki migowe a lingwistyka korpusowa, *Język Polski* 4/5, ss. 338–345.
- Przegaliński, Piotr (2006). *Ochrona wizerunku osoby fizycznej w prawie autorskim*. Pobrano z: <http://wiadomosci.mediaron.pl/artypul/prawo-raport,ochrona-wizerunku-osoby-fizycznej-w-prawie-autorskim,14921,2,1,1.html>. Dnia (01.03.2013).
- Sinclair, John (2005). Corpus and text — basic principles. W M. Wynne (red.), *Developing linguistic corpora: a guide to good practice*, ss. 1–16. Oxford: Oxbow Books.
- Waliński, Jacek (2005). Typologia korpusów oraz warsztat informatyczny lingwistyki korpusowej. W B. Lewandowska-Tomaszczyk (red.), *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.