**Joanna Filipczak**

# Anotacja korpusu PJM

## 1. Założenia

Korpus polskiego języka migowego (PJM) obejmuje ok. 300 godzin materiałów filmowych zawierających wypowiedzi głuchych informatorów zarejestrowanych podczas sesji nagraniowych. Taki materiał, aby mógł służyć analizom językoznawczym, musi zostać poddany obróbce: segmentacji oraz anotacji. Dzięki temu korpus staje się zbiorem szczególnego typu transkrybowanych tekstów, które mogą być otagowane, a w konsekwencji, przeszukiwalne.

Każda transkrypcja języka naturalnego może uwzględnić tylko wybrane jego aspekty. Badacz stosujący transkrypcję służącą przedstawieniu procesów morfologicznych zrezygnuje zapewne z zapisu intonacji, akcentów czy innych elementów z jego punktu widzenia nieistotnych dla opisywanego problemu. Zadaniem lingwisty jest podjęcie decyzji, które elementy powinno się w zapisie zachować, a które należy pominąć — nie tylko po to, by notacja służyła wyznaczonym celom badawczym, lecz także, by analizowane zjawiska językowe były wyraźnie widoczne dla potencjalnego odbiorcy.

W pracach nad językami fonicznymi badacz ma do dyspozycji różnorakie narzędzia transkrypcyjno-anotacyjne, np. powszechnie stosowane systemy transkrypcji fonetycznej, takie jak międzynarodowy alfabet fonetyczny (IPA — ang. *International Phonetic Alphabet*). Standardy takie nie wykształciły się jeszcze w ramach lingwistyki migowej. Poszczególni badacze stosują różne rozwiązania, choć trwają dyskusje nad koniecznością ich ujednolicania (Kato, 2008, Prillwitz i in., 1989, Pizzuto, Pietrandrea, 2001, Hoiting, Slobin, 2002). Dodatkowo, transkrypcja wypowiedzi migowych jest niedoskonała ze względu na jej statyczność oraz niemożność dokładnego przełożenia trójwymiarowego (nielinearnego) tekstu na zapis na papierze

(Morgan, 2003)*. Należy również podkreślić, że z uwagi na koszty nagrań i czasochłonność procesu anotacji korpusy migowe zazwyczaj nie mają określonego szczegółowego celu badawczego, a więc transkrypcja powinna zakładać jak najszersze możliwości późniejszego wykorzystania materiału do analizy językowej.

W ramach badań nad językami migowymi i powstającymi na świecie korpusami języków migowych wypracowano pewne zasady transkrypcji całych tekstów migowych, tak by mogły służyć badaniom językoznawczym oraz samej wymianie danych między badaczami. Powszechnie stosowaną zasadą transkrypcji jest tzw. transkrypcja glosowa oraz tagowanie (zestaw skrótów i symboli dołączanych do glos). Jednym z pierwszych zastosowań tego typu zapisu był anotowany, wielojęzyczny korpus języków migowych tworzony w ramach projektu ECHO (Crasborn i in., 2007). Jest to linearny sposób notacji tekstów migowych, w którym znaki zapisane są za pomocą glos z wybranego języka fonicznego w formie hasła słownikowego danego leksemu. Metoda ta jest powszechnie stosowana zarówno w anotacji korpusów migowych, jak i pracach badawczych; służy szybszej analizie danych językowych oraz umożliwia przeprowadzanie analiz porównawczych między językami migowymi (por. Johnston, 2010, Konrad, Langer, 2009).

Przy anotacji korpusu PJM uwzględniono podstawowe założenia transkrypcji glosowych stosowanych w badaniach międzynarodowych, ale dostosowano je do specyfiki badań realizowanych przez Pracownię Lingwistyki Migowej Uniwersytetu Warszawskiego. W niniejszym rozdziale przedstawiony jest opis procesu segmentacji nagranych danych (m.in. wyodrębnianie znaków z potoku mowy oraz problem delimitacji jednostek), a także zasady konstrukcji glos i ich podziału na poszczególne typy oraz informacje dotyczące użytych w korpusie skrótów i symboli. Osobny podrozdział dotyczy stosowanej w korpusie transkrypcji fonetycznej znaków, tzw. hamburskiej notacji dla języków migowych (HamNoSys), na końcu zaś przedstawiono krótki opis systemu tagowania zaglosowanych danych korpusowych.

## 2. Segmentacja

Filmy pochodzące z sesji nagraniowych, stanowiące surowy, nieopracowany materiał, zostają przekonwertowane na pliki z odpowiednim rozszerzeniem, wprowadzone na serwer, przeniesione do programu, w którym przeprowadza się

* Istnieje wiele propozycji zapisu narracji migowych, w których przestrzeń oraz ruch (szczególnie w konstrukcjach klasyfikatorowych) wymykają się możliwościom linearnej transkrypcji.

anotację, i podzielone na mniejsze części (segmenty), odpowiadające poszczególnym zadaniom elicytacyjnym (w trakcie sesji każdy z informatorów wykonuje 26 zadań). Tak przygotowany materiał zostaje przekazany do pierwszego etapu anotacji, czyli segmentacji filmu na poszczególne znaki migowe.

Głównym celem tego procesu jest wyodrębnienie jednostek leksykalnych z potoku mowy (migów), tak aby w późniejszej fazie anotacji można było im przypisać odpowiednią glosę. Segmentacja jest procesem żmudnym i pracochłonnym, który wymaga już na wstępie określenia pewnych założeń metodologicznych odnośnie do delimitacji jednostek. Przede wszystkim należy podjąć decyzję, co uznajemy za element komunikatu migowego, a następnie, na jakiej podstawie będziemy określać granice konkretnego znaku migowego.

W komunikacji migowej wykorzystywane jest całe ciało, które porusza się w trójwymiarze. Znaki migowe wykonywane są w tzw. przestrzeni migania — obejmuje ona samego migającego (fizycznie obecnego w tej przestrzeni) oraz obszar przed nim, wyznaczony mniej więcej od wysokości głowy do linii bioder. Ręce, dłonie, twarz migającego są artykulatorami, analogicznie do języka, warg, czy zębów w językach fonicznych.

W korpusie PJM segmentacji (wyodrębnieniu) podlegają przede wszystkim znaki manualne, a więc te, które wykonywane są za pomocą rąk. Podstawą wyróżnienia znaku staje się więc przede wszystkim wyjściowy układ dłoni, a nie przykładowo tzw. *mouthing* (ruchy warg odpowiadające artykulacji słów fonicznych — istotne przy późniejszym procesie lematyzacji). Nie uwzględnia się na tym etapie parametrów niemanualnych, takich jak mimika, ruchy głową, wychylenia ciała.

W trakcie segmentacji wyodrębnieniu podlegają również te elementy, które niekoniecznie muszą posiadać status językowy, takie jak gesty naturalne, gesty fatyczne (rozpoczynające lub kończące komunikat) oraz inne znaki o wątpliwym statusie komunikacyjnym (znaki przerwane, pomyłki). Wszystkie te elementy zostają wyodrębnione, tak by w następnym etapie anotacji można było zdecydować, jaka jest ich faktyczna funkcja w komunikacie językowym. Argumentem stojącym za tym rozwiązaniem (zważywszy na brak szczegółowych celów badawczych postawionych korpusowi) jest fakt, że anotacja tego typu elementów może okazać się podstawą do różnorodnych badań gestologicznych.

Jednym z największych problemów związanych z segmentacją jest właściwe wyodrębnienie jednostek z potoku wypowiedzi. Materiał językowy, z jakim mamy do czynienia w korpusie, mimo że w dużej mierze moderowany (ze względu na sam proces elicytacji danych), jest jednak przede wszystkim zbiorem spontanicznych wypowiedzi. Głównym celem tworzenia

korpusu jest zebranie tekstów jak najbardziej zbliżonych do naturalnych zachowań komunikacyjnych. Korpusowi migowemu bliżej zatem do korpusów mówionych języków fonicznych niż do korpusów tekstów pisanych. W takim zbiorze danych bardzo liczne są anakoluty, powtórzenia, pomyłki i pauzy. Znaki wykonywane są szybko i bez (sztucznej) dbałości o poprawność kształtu dłoni. Poszczególne migi są również często na tyle zmodyfikowane pod wpływem sąsiadujących elementów*, że właściwe ich wydzielenie wymaga kompetencji rodzimego użytkownika języka.

Wielu kłopotów nastręcza też ustalenie punktów początku i końca znaku. Segmentacja tekstów opiera się na tzw. wąskim rozumieniu struktury znaku, w którym jako punkt początkowy artykulacji bierze się pod uwagę układ dłoni w lokacji początkowej, a następnie analizuje się ruch-ścieżkę oraz lokację końcową — według modelu zaproponowanego przez Sandler (2008). Nie uwzględnia się przy tym ruchów przejściowych między znakami, tzw. transjentów, a przy anotacji zaznacza się jedynie dłuższe pauzy między poszczególnymi znakami.

Przy konstrukcjach dwuręcznych symultanicznych, w których każdy z artykulatorów wykonuje inny znak (o ile nie są to konstrukcje klasyfikatorowe), wyodrębnia się osobno znaki wykonywane ręką dominującą i niedominującą. Na takie rozwiązanie pozwala stosowany przy anotacji program iLEX, którego szczegółowe wykorzystanie zostało opisane w kolejnym rozdziale.

## 3. iLEX

iLex, czyli *integrated lexicon* (Hanke, Storz, 2008) jest programem zaprojektowanym specjalnie do anotacji tekstów języków migowych. Jego zaletą jest możliwość jednoczesnego dostępu (za pośrednictwem Internetu) wielu anotatorów do danych korpusowych umieszczonych na serwerze.

Program iLex wykorzystywany jest na wszystkich etapach opracowywania danych zebranych w korpusie PJM (segmentacja, lematyzacja, tagowanie, kontrola). Każdemu zadaniu elicytacyjnemu przypisany jest plik transkrypcji, który może zostać w odpowiednim oknie poddany segmentacji, glosowaniu i tagowaniu. Okno transkrypcji programu iLex widoczne jest na rys. 1.

---

* Zmiany układów dłoni pod wpływem sąsiadujących migów postrzegane są jako procesy fonologiczne (asymilacje fonologiczne).

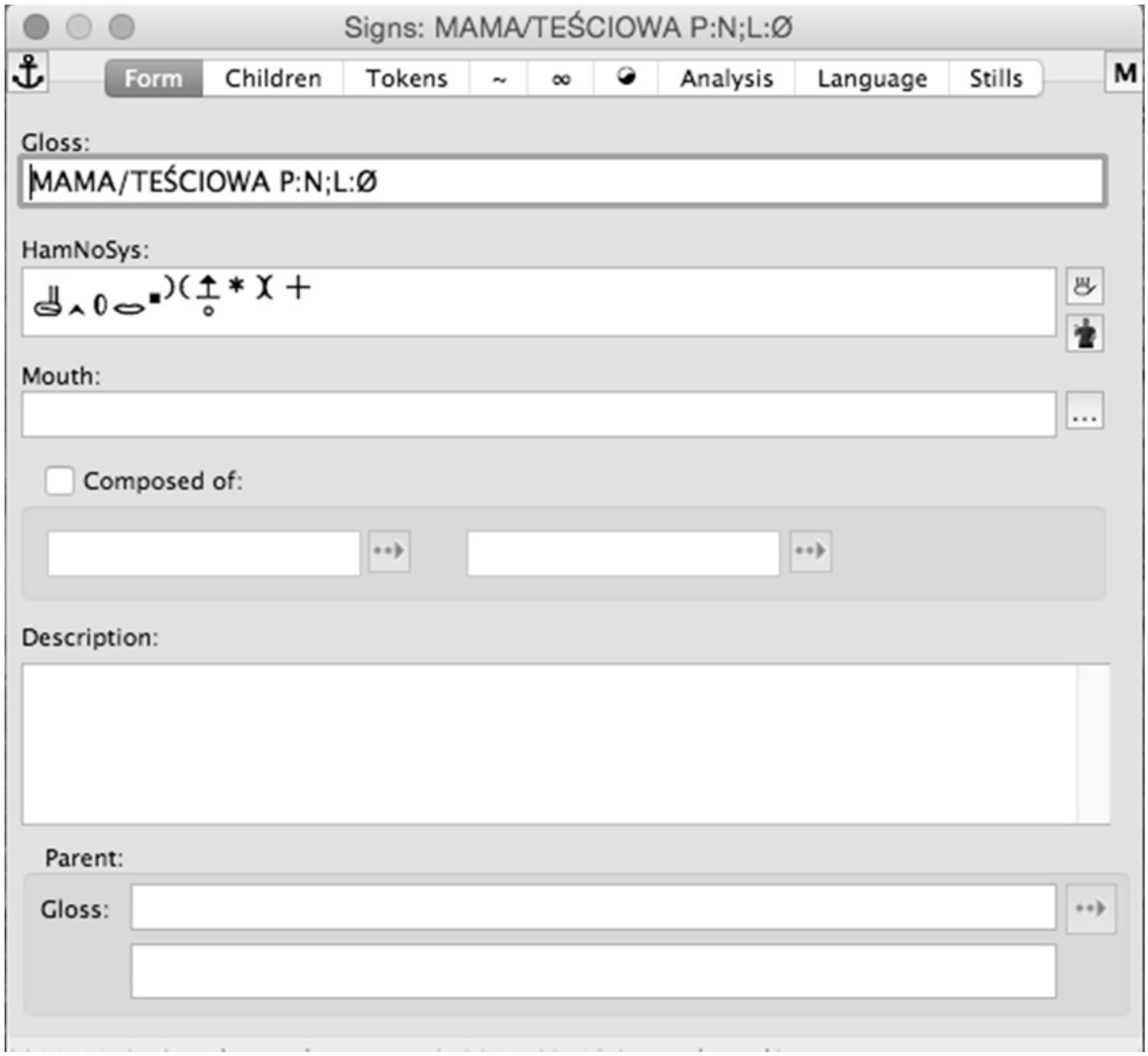

Rys. 1. Okno segmentacji w programie iLex z pociętym materiałem filmowym

Po lewej stronie widoczny jest anotowany film, po prawej — kolumny, którym przypisane są kolejno wartości: segmentacji, glosy dla ręki dominującej, glosy dla ręki niedominującej, transkrypcji HamNoSys, odstępstw od standardowej artykulacji, komentarzy, tłumaczenia, elementów niemanualnych, mimiki, tagów gramatycznych i innych. Najważniejszą informacją jest glosa dla ręki dominującej, czyli podstawowego artykulatora każdego znaku. Najczęściej dominacja którejś z rąk powiązana jest z ręcznością: osoby praworęczne zazwyczaj używają ręki prawej jako dominującej, leworęczne — lewej. W zależności od preferencji migającego kolumna dla ręki niedominującej będzie zawierać glosy dla ręki lewej lub prawej.

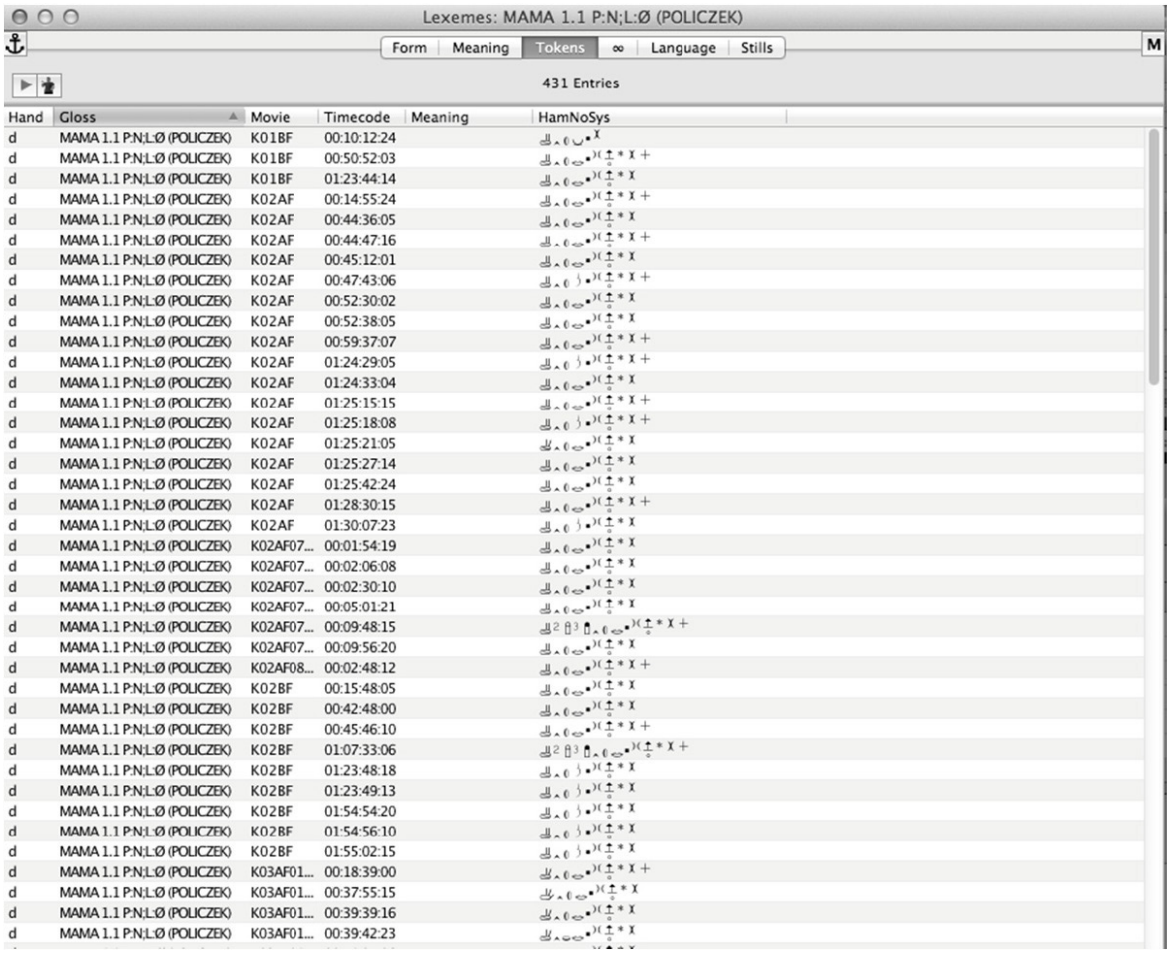
## 4. Lematyzacja

Lematyzacja polega na przypisaniu poszczególnym wystąpieniom tekstowym ich formy podstawowej. Dla języków migowych formą podstawową leksemu jest tzw. forma izolowana (ang. *citation form*), a więc znak wykonywany samodzielnie — nie jako część komunikatu językowego. Lematyzacja opiera się na przypisaniu do formy znaku występującej w tekście (tokenu)

odpowiedniej glosy, która służy przede wszystkim odróżnieniu danego leksemu od innych (ze względu na jego kształt) oraz oddaniu przybliżonego znaczenia (na rysunkach 2a i 2b zamieszczono przykładową glosę oraz jej wystąpienia).




Rys. 2a. Okno programu iLex pokazujące informacje o przykładowym haśle




Rys. 2b. Okno programu iLex pokazujące poszczególne wystąpienia znaku

W podrozdziale dotyczącym segmentacji wspomniano już o problemach rozróżnienia elementów, które stanowią część komunikatu językowego, i gestów lub elementów o niejasnym statusie językowym. W trakcie lematyzacji anotator ocenia dane wystąpienie pod względem jego statusu i podejmuje decyzję co do wyboru konkretnej glosy. W słowniku glos wyróżniono kilka grup nieznaków, dla których ustalono odpowiednie symbole. Spis symboli i ich funkcje zamieszczone zostały w tabeli 1.

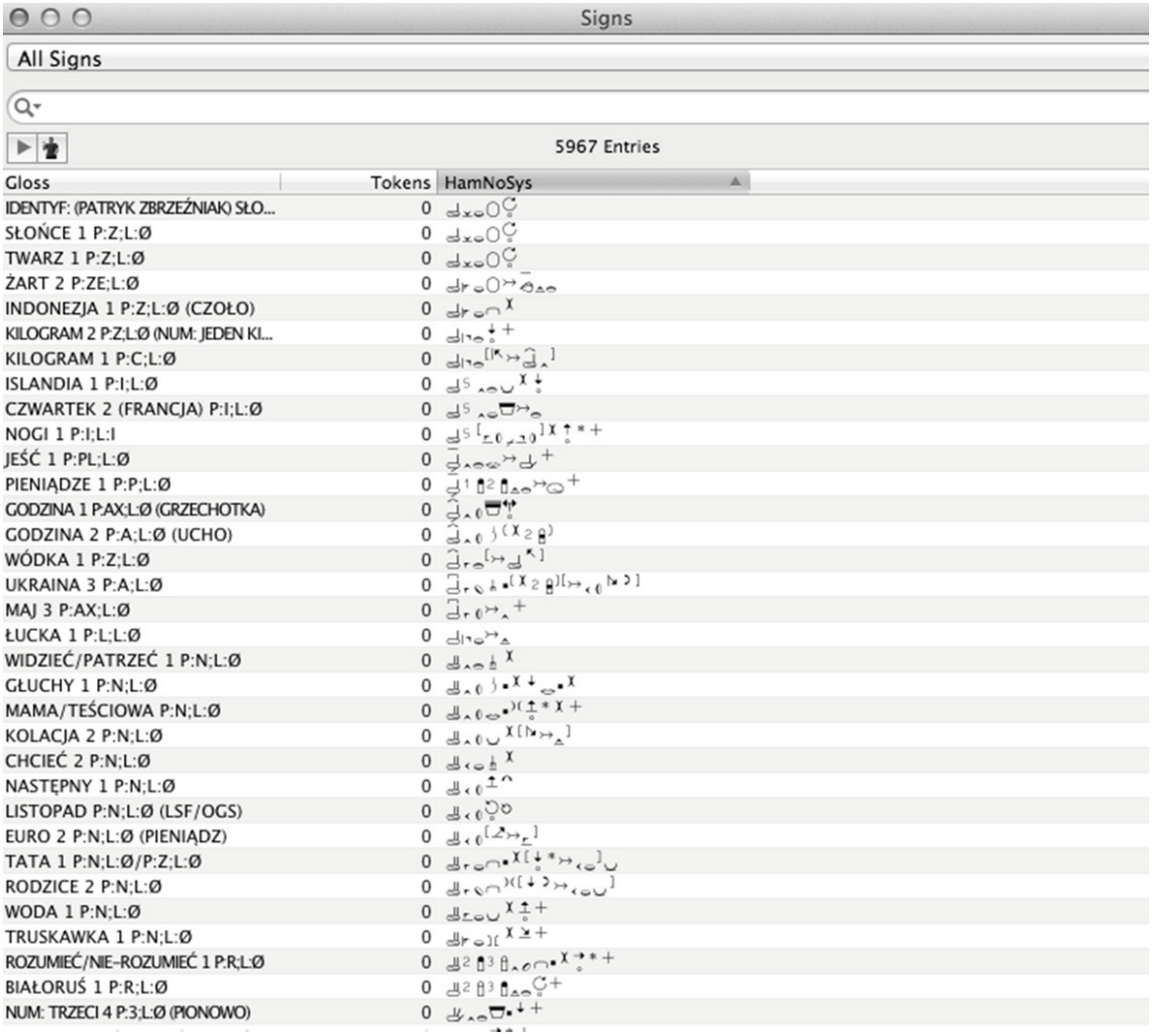
| SYMBOL | ZNACZENIE |
|---|---|
| ### | znaki urwane, niedokończone lub zupełnie niezrozumiałe, pomyłki |
| % | gest artykułowany otwartą dłonią w kształcie znaku daktylograficznego B, np. podniesienie rąk do góry na końcu wypowiedzi (ang. *palm-up*) |
| & | machnięce ręką (‘nieważne!’) |
| ^ | pauza w miganiu, przerwa między jednym a drugim znakiem |
| @ | rozpoczęcie, zakończenie migania, zwrócenie uwagi rozmówcy |

Tab. 1. Symbole stosowane w słowniku programu iLex

Na szczególną uwagę zasługuje wyróżnienie migów oznaczanych symbolem % (ang. *palm-up*), których status w komunikacji migowej jest zarówno gestowy, pragmatyczny (sygnał końca komunikatu, przekazanie roli rozmówcy), jak i metatekstowy (stanowiący komentarz do wypowiadanych treści co do ich pewności, źródła pochodzenia itp.).

Poza powyższymi grupami główną część słownika stanowią tzw. znaki manualne oznaczane za pomocą glos właściwych. Glosa przyjmuje formę hasłową leksemu z wybranego języka fonicznego (w tym przypadku języka polskiego) i zostaje zapisana za pomocą kapitalików. Znak odnoszący się do psa lematyzowany jest zatem jako PIES (sekwencja kapitalików oddzielona z obu stron spacjami), o ile nie stanowi elementu złożenia bądź inkorporacji.

Rysunek 3 przedstawia wyciąg ze słownika glos stanowiącego część oprogramowania iLex.




Rys. 3. Słownik glos w programie iLex

Warto zauważyć, że każdej glosie towarzyszą skróty literowe i symbole zgodne z następującymi schematami:

ZNAK P:X;L:X
ZNAK P:X;L:⃠

W wypadku znaku dwuręcznego po glosie właściwej występuje symbol układu dłoni dla ręki dominującej (P) oraz niedominującej (L) podany w postaci literowej (wstawiany w miejsce X), która nawiązuje do układów dłoni z alfabetu palcowego stosowanego w PJM. Jeśli znak jest jednoręczny, po skrócie dla ręki niedominującej występuje symbol ⃠ (oznaczający brak układu dłoni). Taka konwencja pozwala na zapis alternatywnych form danego znaku (zmodyfikowany układ ręki lub orientacja, znaki symetryczne/niesymetryczne).

Znaki różniące się kształtem, ale posiadające to samo lub zbliżone znaczenie, a więc i taką samą glosę, odróżniane są, poza skrótami literowymi, za pomocą cyfr arabskich.

Dla znaków, których najbliższym polskim odpowiednikiem nie jest pojedyncze słowo, lecz fraza lub wyrażenie, poszczególne słowa współtworzące glosę oddzielone są dywizami. Przykładem takich glos są:

BLOK-POZIOM P:C;L:C
JEDNA-GODZINA P:Z;L:Ø
NIE-SŁYSZEĆ P:E;L:E

W trakcie anotacji stwierdzono również, że należy wyodrębnić dodatkowe typy glos tworzące odrębne grupy w słowniku. Wśród nich wyróżniono tzw. znaki kulturowe, konstrukcje klasyfikatorowe, przydomki, wskaźniki (indeksy), gesty oraz literowania (patrz tabela 2).

Glosy te skonstruowane są na innej zasadzie niż glosy właściwe, ponieważ najpierw stosuje się w nich odpowiedni symbol (Q, $:KL,), a następnie literowy odpowiednik układu dłoni (X, dla znaków dwuręcznych w postaci X+X), a w nawiasie przybliżone znaczenie.

| SYMBOL | ZNACZENIE |
|---|---|
| Q:X (...) | znaki kulturowe |
| $:KL:X (...) | konstrukcje klasyfikatorowe |
| IDENTYF: (...) | identyfikatory, przydomki migowe |
| WSKAZ: X | znaki oznaczające wszystkie typy wskazywania, w różnych kierunkach |
| A.B.C.D. | wyrazy literowane w alfabecie palcowym |

Tab. 2. Skróty stosowane do oznaczania dodatkowych typów glos

Znaki kulturowe to znaki manualne, które nie posiadają oczywistego odpowiednika w języku polskim w postaci zleksykalizowanej formy hasłowej. Przekazanie znaczenia tego typu leksemów jest często możliwe jedynie za pomocą opisu lub podania konkretnych kontekstów użycia. Do tego typu glos będą należeć:

Q: 5+5 (BRAWO+DEAF)
Q: 5 (WU+GDYBY NIE TO/TO NIE TAK)
Q: E+E (NIEZRĘCZNA SYTUACJA)

Konstrukcje klasyfikatorowe to bardzo duża grupa struktur produktywnych, które za pomocą klasyfikatorów w kombinacji z innymi znakami wyrażają relacje przestrzenne, charakteryzują kształt i wymiar obiektów lub np. właściwości opisywanego ruchu. Przykłady glos klasyfikatorowych:

$:KL: 5+5 (PRZEBRAĆ+IŚĆ JAK DUCH)
$:KL: N+N (IŚĆ/BIEGAĆ STOPAMI/ZWIERZĘTA MAJĄCE 4 ŁAPKI)
$:KL: O+O (WISZĄCE MEDALE)

Przydomki to zbiór znaków, które traktowane są w słowniku jako nazwy własne, odnoszące się do konkretnych osób, dla których, zamiast literowania ich imion, stosuje się ustalone znaki*.

W razie potrzeby do każdej glosy dopisać można dodatkowe informacje związane ze znaczeniem czy artykulacją znaku. Słownik glos i stosowane w nim konwencje mają przede wszystkim służyć anotatorom oraz potencjalnym użytkownikom korpusu. Dlatego też sama procedura identyfikowania glos (np. przez nadawanie im numerów) nie byłaby wystarczająca — forma hasłowa w glosie powinna sugerować przynajmniej przybliżone znaczenie znaku oraz sposób jego artykulacji, tak by stosunkowo szybko można było odtworzyć transkrypcję, nawet bez dostępu do nagrań filmowych.

## 5. HamNoSys

Do każdej formy glosowej dołączona jest także transkrypcja fonetyczna znaku zapisana za pomocą systemu HamNoSys (hamburski system notacji

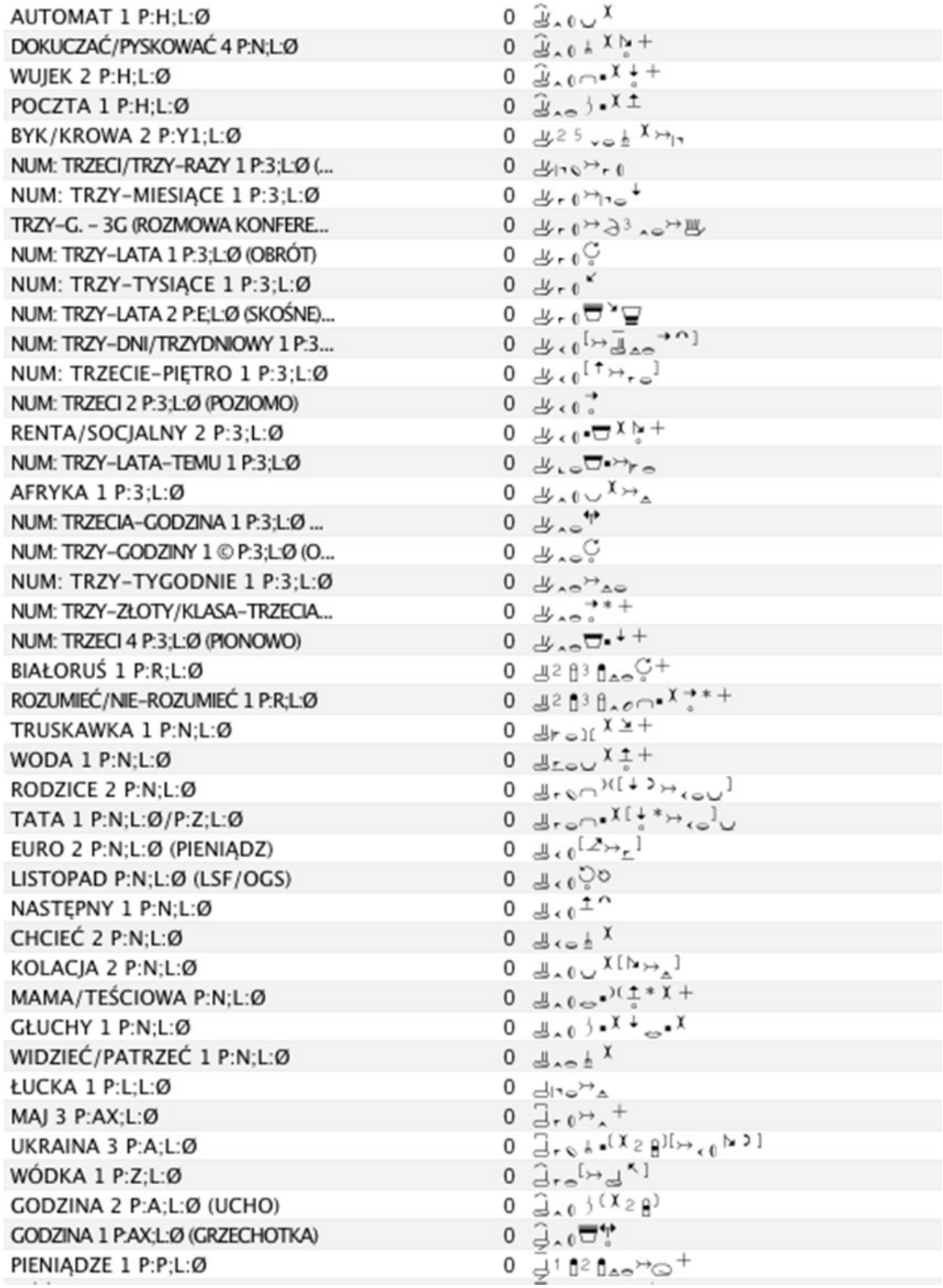
* Przydomki migowe są szczególną cechą kulturową w PJM. Każdy członek społeczności otrzymuje przydomek, który może być związany z jego wyglądem (znak OKULARY, DŁUGIE-WŁOSY, DUŻY-NOS), cechami charakteru (OTWARTY, RADOSNY), pracą, zainteresowaniami lub pochodzi od nazwiska (OSA od Osowska).

znaków dla języków migowych, ang. *Hamburg Sign Language Notation System*). Jest to migowy odpowiednik międzynarodowego alfabetu fonetycznego (IPA) stosowanego w transkrypcji języków fonicznych. Został stworzony w roku 1985 przez zespół badaczy z Uniwersytetu w Hamburgu (pod kierownictwem Thomasa Hankego) i od tego czasu jest systematycznie rozwijany, obecnie używana jest wersja 4.0 (Hanke, 2004).

System hamburski odwołuje się do koncepcji budowy fonetycznej znaku zaproponowanej przez Stokoe'ego (1960). Wyróżnił on trzy elementy (parametry), z których zbudowany jest znak migowy: układ dłoni (*hand configuration*), ruch (*movement*) oraz miejsce artykulacji (*location*). Układ dłoni obejmuje kształt dłoni, układ palców oraz orientację (pozycję wnętrza dłoni względem ciała migającego). Kategoria ruchu odnosi się do każdego typu ruchu rąk, dłoni i nadgarstków; miejsce artykulacji natomiast określa punkt w przestrzeni (lub na ciele), w którym znak jest wykonywany. Dystynktywną funkcję tych elementów mają potwierdzać pary minimalne znaków migowych, które różnią się tylko jednym parametrem (Stokoe, 1960, s. 14). Te same parametry uwzględnia transkrypcja hamburska, która przyjmuje postać zapisu symboli w sekwencji od lewej do prawej: układu palców, orientacji palców, orientacji wnętrza dłoni, lokalizacji oraz ruchu. Transkrypcja ta w zamyśle jest uniwersalna — może być wykorzystana do zapisu znaków dowolnego języka migowego, a więc przewiduje wszystkie możliwe układy dłoni, lokalizacje i rodzaje ruchu, przedstawiając je w postaci symboli i cyfr.

W projekcie korpusu PJM każdej glosie zamieszczonej w słowniku programu iLex przypisana jest transkrypcja fonetyczna znaku sporządzona w systemie HamNoSys. Zapis ten odnosi się do formy podstawowej leksemu, czyli tzw. formy izolowanej. Dla każdego wystąpienia (tokenu) można również dodać dodatkową transkrypcję fonetyczną (dzieje się tak w przypadku, gdy forma tekstowa nieznacznie różni się od formy wzorcowej). Przykładowe transkrypcje fonetyczne glos ilustruje rysunek 4.

System iLex pozwala — dzięki programowi *SiGML Service Player* — zwizualizować zapisaną transkrypcję za pomocą awatara. Funkcja ta jest niezwykle przydatna przy sprawdzeniu poprawności notacji (bez potrzeby odtwarzania filmu).




Rys. 4. Przykład transkrypcji HamNoSys w słowniku programu iLex

## 6. Tagowanie

Po segmentacji, zaglosowaniu oraz dodaniu transkrypcji fonetycznej następuje drugi etap, tzw. anotacja szczegółowa, w której zapisuje się wszystkie elementy niemanualne wykonywane podczas migania: ruchy głową, ruchy ciała, mimikę, zmiany kierunku patrzenia, ruch ust (ang. *mouthing*) oraz inne elementy istotne z punktu widzenia badacza (znaczenie znaku, informacje gramatyczne itp.). Trudno wyznaczyć wyraźną granicę między czystą deskrypcją (notowaniem wystąpienia danego elementu) a opisem jego funkcji w komunikacie. Dlatego też tagowanie jest ściśle powiązane z analizą językową nagranego materiału i realizowane najczęściej przy okazji prowadzenia szczegółowych badań lingwistycznych. W tabeli 3 zamieszczono przykładowe skróty stosowane przy tagowaniu materiału korpusowego.

| SYMBOL | ZNACZENIE |
|---|---|
| N, V, ADJ, Num | literowe skróty oznaczające kategorię gramatyczną: rzeczownik (ang. *noun*), czasownik (ang. *verb*), przymiotnik (ang. *adjective*), liczebnik (ang. *numeral*) |
| repet, redupl, DC | skróty odnoszące się do zjawisk morfosyntaktycznych: repetycji, reduplikacji, konstrukcji podwojonej (ang. *doubling construction*) |
| wz, gł, rshft | skróty oznaczające skierowanie wzroku/głowy/ciała w określony punkt przestrzeni oraz zmianę ról (ang. *role shift*) |
| pyt/neg/rozk | skróty oznaczające niemanualne wykładniki zdania pytajnego/ negacji/zdania rozkazującego |

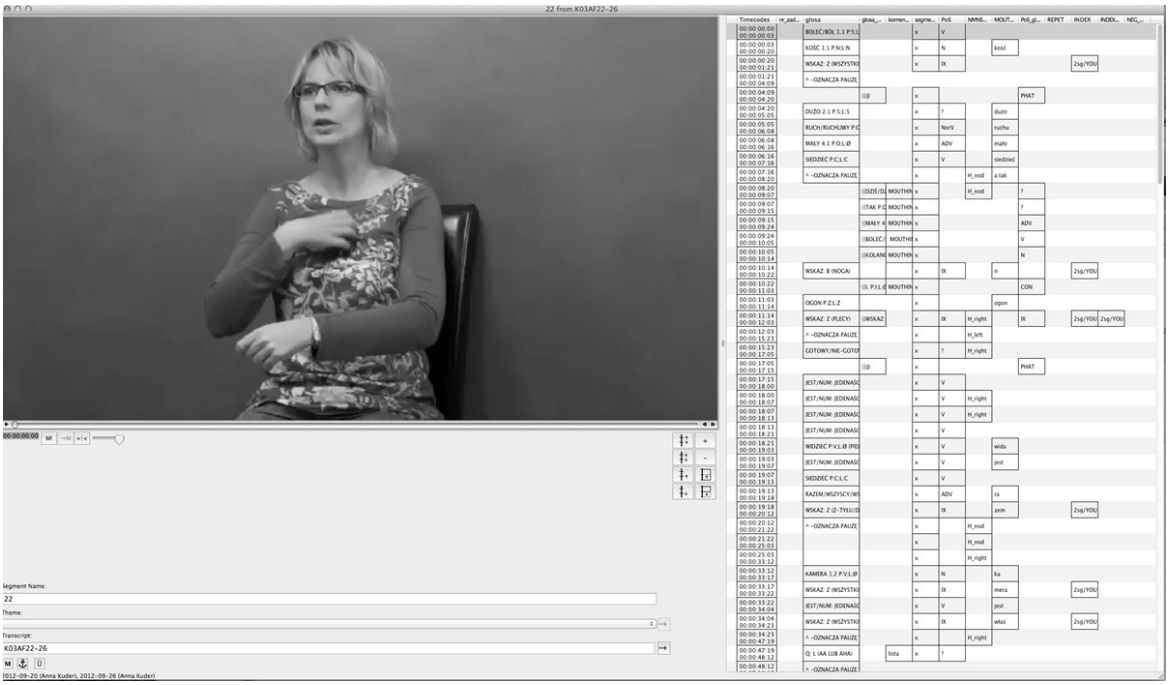
Tab. 3. Przykładowe skróty stosowane przy tagowaniu korpusu

Tagowanie glos koncentruje się na dwóch najważniejszych aspektach komunikatu migowego: elementach niemanualnych oraz użyciu przestrzeni. Najczęściej mamy do czynienia z sytuacją, w której na znak zostają nałożone elementy niemanualne, wykonywane symultanicznie z elementami manualnymi. Taki status mają chociażby niemanualne wykładniki pytajności, negacji, trybu rozkazującego, czy też elementy odpowiadające za modyfikację przysłówkową znaków o charakterze czasownikowym (np. PRACOWAĆ$_{[intens]}$, gdzie parametr $_{[intens]}$, będzie oznaczał, że czynność jest wykonywana z wysiłkiem: zmianie podlega sposób wykonywania ruchu oraz jednocześnie zostaje dodany komponent niemanualny — mimika kojarząca się z wysiłkiem fizycznym).

Element przestrzenny uwzględniają takie oznaczenia, jak np.: *wz* (wzrok), *gł* (głowa), *rshift* (ang. *role shift*, przesunięcie ciała, które sygnalizuje, że osoba migająca wciela się w czyjąś rolę, „cytuje" zachowania osoby, której

dotyczy wypowiedź). Pełnią one niejednokrotnie w tekście migowym funkcję składniową: uzgodnienia niemanualnego (dla subiektu i obiektu) czy wykładnika zdań złożonych.

Oznaczenia typu *repet*, *DC* i *redupl* odnoszą się do różnorakich powtórzeń pojawiających się w tekstach migowych: pełniących funkcję morfologiczną (np. służących tworzeniu liczby mnogiej lub sygnalizujących intensyfikację), składniową (jako tzw. *doubling constructions*, czyli konstrukcje podwojone z emfazą) czy pragmatyczną (jako wykładniki spójności tekstu lub funkcji fatycznej). Rysunek 5 przedstawia przykład tekstu migowego, który został już zaglosowany oraz otagowany.



Rys. 5. Przykładowa transkrypcja materiału zaanotowanego w programie iLex

## 7. Podsumowanie

Korpus języka migowego ma służyć analizie danych wizualno-przestrzennych, niestety ze względu na wiele uwarunkowań technicznych (m.in. niewystarczające możliwości komputerowej analizy znaków) oraz problemy z jednoznacznym wyodrębnianiem kategorii gramatycznych niemożliwe jest zastosowanie w jego anotacji programów do automatycznego przyporządkowywania poszczególnych migów do jednostki leksykalnej (leksemu) czy automatycznej analizy składniowej. Cały proces anotowania odbywa się więc przez ręczne znakowanie nagranego materiału. Dopracowane w szczegółach

zasady anotacji sprawiają, że zebrane dane językowe mogą stanowić podstawę badań lingwistycznych nad komunikacją migową w każdym jej aspekcie: składniowym, semantycznym oraz pragmatycznym.

## Bibliografia

Crasborn, Onno, Mesch, Johanna, Waters, Dafydd, Nonhebel, Annika, van der Kooij, Els, Woll, Bencie, Bergman, Brita (2007). Sharing sign langauge data online. Experiences from the ECHO project. *International Journal of Corpus Linguistics 12*(4), ss. 537–564.

Hanke, Thomas (2004). HamNoSys — representing sign language data in language resources and language processing contexts. W O. Streiter, C. Vettori (red.), *Proceedings of the workshop on the representation and processing of sign languages. 4th International Conference on Language Resources and Evaluation, LREC 2004, Lisbon*, ss. 1–6. Paryż: ELRA.

Hanke, Thomas, Storz Jakob (2008). iLex — A database tool for integrating sign language corpus linguistic and sign language lexicography. W Crasborn, Onno i in. (red.), *Proceedings of the 3rd Workshop on the Representation and Processing of Signed Languages: Construction and exploitation of sign language corpora. International Conference on Language Resources and Evaluation,* ss. 64–67. Paryż: ELRA.

Hoiting, Nini, Slobin, Dan (2002). Transcription as a tool for understanding: The Berkeley Transcription System for sign language research (BTS). W G. Morgan, B. Woll (red.), *Directions in sign language acquisition*, ss. 55–75. Amsterdam, Filadelfia: John Benjamins.

Johnston, Trevor (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics 15*(1), ss. 106–131.

Kato, Mihoko (2008). A study of notation and sign writing systems for the Deaf. *Intercultural Communication Studies XVII*(4), ss. 97–114.

Konrad, Reiner, Langer, Gabriele (2009). Synergies between transcription and lexical database building: The case of German Sign Language (DGS). W M. Mahlberg, V. González-Díaz, C. Smith (red.), *Proceedings of the Corpus Linguistics Conference* (*CL2009*). Liverpool: University of Liverpool.

Morgan, Gary (2003). Transcription of child sign language. *Deafness & Education International 5*(3), ss. 157–166.

Pizzuto, Elena, Pietrandrea, Paola (2001). The notation of signed texts: Open questions and indications for further research. *Sign Language & Linguistics 4*, ss. 29–45.

Prillwitz, Siegmund, Leven, Regina. Zienert, Heiko, Hanke, Thomas, Henning, Jan (1989). *HamNoSys: Version 2.0. Hamburg Notation System for sign languages. An introductory guide.* Hamburg: Signum.

Sandler, Wendy (2008). The syllable in sign language: Considering the other natural language modality. W B. Davis, K. Zajdo (red.), *Ontogeny and phylogeny of syllable organization, Festschrift in honor of Peter MacNeilage*, ss. 379–408. Nowy Jork: Taylor Francis.

Stokoe, William C. (1960). Sign language structure: An outline of the visual communication systems of the American deaf. *Studies in linguistics: Occasional papers 8*. Buffalo, NY: University of Buffalo.