

# Instrukcja korzystania z wyszukiwarki do Korpusu Tekstów Gwarowych ze Wsi Maćkowce na Podolu

26 stycznia 2021 r.

## Spis treści

1. Wprowadzenie.....	1
2. Dwie warstwy zapisu.....	2
2.1. Interpunkcja.....	3
3. Segmentacja.....	3
4. Zasady lematyzacji i anotacji morfosyntaktycznej.....	3
5. Interfejs wyszukiwarki MTAS .....	4
6. Język zapytań .....	4
6.1. Zapytania o segmenty.....	5
6.2. Zapytania z innymi atrybutami.....	6
6.3. Ograniczenie zapytania za pomocą atrybutów leksykalnych .....	7
6.3.1. Znaczenie i etymologia.....	8
7. Zestaw znaczników morfosyntaktycznych.....	10
7.1. Kategorie gramatyczne.....	10
7.2. Klasy gramatyczne .....	10
7.3. Zapytania o znaczniki morfosyntaktyczne .....	14
8. Wykaz skrótów .....	15
8.1. Skróty oznaczające etymologię leksemów:.....	15
8.2. Skróty zakresów tematycznych rozmowy:.....	16

## 1. Wprowadzenie

Niniejszy dokument powstał w oparciu o [Instrukcję korzystania z wyszukiwarki korpusowej w Korpusomacie](#), opracowaną przez Witolda Kierasia, a także o [Instrukcję korzystania z wyszukiwarki do Elektronicznego Korpusu Tekstów Polskich z XVII i XVIII w. \(do 1772 r.\)](#), opracowaną przez Włodzimierza Gruszczyńskiego i Renatę Bronikowską. Z kolei podstawą tych instrukcji była [Ściągawka do Narodowego Korpusu Języka Polskiego](#), czyli instrukcja użytkownika wyszukiwarki Poliqarp. Niniejsza instrukcja opisuje sposób użytkownika wyszukiwarki MTAS, niepowiązanej z Poliqarphem, ale wykorzystującej podobny język zapytań znany pod nazwą *Corpus Query Language* (CQL). *Ściągawka* została opracowana przez Adama Przepiórowskiego, a następnie poprawiona

i rozszerzona przez Jakuba Wilka i Aleksandra Buczyńskiego. Jej pełna wersja znajduje się w repozytorium wyszukiwarki Poliqarp. Modyfikacje wprowadzone do pierwotnej wersji instrukcji uwzględniają różnice w języku zapytań oraz specyfikę korzystania z korpusu gwarowego. Za zgodą wszystkich wyżej wymienionych autorów niniejsza wersja dokumentu zostaje udostępniona na zasadach licencji [Creative Commons BY-SA](#).

## 2. Dwie warstwy zapisu

Teksty w korpusie dostępne są w dwóch wariantach zapisu. Pierwszy z nich oddaje wszystkie cechy fonetyczne, ale wykorzystuje przede wszystkim litery i dwuznaki polskiego alfabetu. Uzupełnienie stanowią następujące znaki:

- é – dźwięk pośredni między *e* i *y*
- ò – dźwięk pośredni między *o* i *u*
- å – dźwięk pośredni między *a* i *o*
- ə – samogłoska o niejasnej barwie (tzw. *szwa*)
- ĭ – *i* nosowe
- ɔ – *o* nosowe
- ɣ – *y* nosowe
- ɯ – *u* nosowe
- ɸ – *u* niezgłoskotwórcze
- ɨ̃ – *i* niezgłoskotwórcze nosowe
- φ – *f* dwuwargowe
- η – *n* tylnojęzykowe.

Samogłoski pośrednie niewymienione powyżej zapisywane są za pomocą dwuznaków: *ey*, *ye* (dźwięki pośrednie między *e* i *y*), *ae*, *ea* (dźwięki pośrednie między *a* i *e*), *ao*, *oa* (dźwięki pośrednie między *a* i *o* – bliższe barwie samogłoski *a* lub samogłoski *o*).

Zapis fonetyczny oddaje następujące cechy wymowy:

- zmiany barwy samogłosek, np. *rynék*, *bydló*, *kuszule*, *zdumął*
- denazalizację dawnych samogłosek nosowych, np. *chodze*
- asynchroniczną realizację dawnych samogłosek nosowych, np. *domb*, *monk'i*, *w'ejksza*
- labializację i prejotację samogłosek, np. *ɯuczyl*, *jim'e*
- nosową wymowę niezgłoskotwórczego *i*, np. *koĩski*
- różną realizację spółgłosek *ś ź ć dź*, np. *pujdze* (spółgłoski miękkie), *wyjdz'e* (spółgłoski półmiękkie), *miēs''c''ie* (wymowa pośrednia)
- osłabioną miękkość innych spółgłosek miękkich, np. *fs\*e*
- relikty mazurzenia, np. *cy*
- rozróżnienie pomiędzy *ł* przedniojęzykowo-zębowym, np. *muwił*, a *u* niezgłoskotwórczym (takim jak w polszczyźnie ogólnej), np. *muwiu*
- dwuwargową wymowę głoski *f*, np. *sɸakowe*
- wydłużoną wymowę spółgłosek, np. *jeć:e* 'jedźcie'
- udźwięcznienia i ubezdźwięcznienia w śródgłosie i w wygłosie oraz międzywyrazowe, np. *glut*, *wracadź du sfojich*, *jag nu*
- inne zmiany w grupach spółgłoskowych i upodobnienia międzywyrazowe, np. *dże* 'gdzie', *kszcíł* 'chrzył', *ś sóła* 'z sióła'
- różne inne uproszczenia wymowy, np. *prendzy* 'prędzej'.

W tekście zaznaczony jest akcent inny niż paroksytoniczny, a także paroksytoniczny w takich kontekstach, które mogą budzić wątpliwości, np. *iszcz'e*, *d'o dom*. Spółgłoska *l* jest realizowana zawsze miętko, jednak dla uproszczenia procedury standaryzacji zapisu nie zostało to zaznaczone. Zatem zapis *las*, *rubla* należy czytać [l'as], [rubl'a].

Druga warstwa to zapis standaryzowany. Jego wprowadzenie miało ułatwić lematyzację korpusu za pomocą istniejących już narzędzi do automatycznej analizy tekstów polskich i tylko od tego były uzależnione decyzje o zakresie standaryzacji. W warstwie tej przywrócone zostały cechy polskiej ortografii (np. *muwiał* → *mówił*, *pujdże* → *pójdzie*, *domp* → *dąb*), standardowej wymowy (np. *rynék* → *rynek*, *bydló* → *bydło*, *kažo* → *każę*, *s'izd'eli* → *siedzieli*, *prendzy* → *prędeży*, *zdwi* → *drzwi*), a także wprowadzone zostały ogólnopolskie końcówki gramatyczne na miejsce gwarowych (np. *šostruf* → *sióstr*, *idz'ema* → *idziemy*). Tam, gdzie istniały różnice w stosunku do polszczyzny ogólnej spowodowane wyrównaniami morfologicznymi, przywrócona została standardowa postać rdzenia wyrazu (np. *b'ero* → *biorę*). Jednak większość zleksykalizowanych cech fonetycznych nie podlegała standaryzacji (np. w formach leksemu ABIT 'habit') lub też podlegała jej dopiero podczas lematyzacji (np. leksem CMENTARZ w warstwie standaryzowanej jest zapisany jako *smentarz*).

## 2.1. Interpunkcja

W części tekstów występuje standardowa polska interpunkcja, stosowana tak, by oddawała intonację mówiącego. W innych natomiast stosowany jest przede wszystkim znak | na oznaczenie pauzy. Intonacja inna niż neutralna zaznaczana jest znakami ? ! oraz ...

## 3. Segmentacja

Znaczniki morfosyntaktyczne, tzw. tagi, przypisane są segmentom (tokenom, w przybliżeniu słowom). Segmenty nie są dłuższe niż słowa ortograficzne (słowa „od spacji do spacji”). Podobnie jak w Narodowym Korpusie Języka Polskiego jako oddzielne segmenty traktowane są:

- formy aglutynacyjne leksemu BYĆ, np. [*słyszal*][*am*], [*tag*][*źmy*]
- partykuły *by* i *-ż(e)*, np. [*było*][*ż*], [*był*][*by*].

Zostały one zapisane oddzielnie, niezgodnie z polską ortografią, ponieważ w pierwotnej wersji korpusu nie był możliwy podział słów na odrębne segmenty. Spacja występuje także przed znakami interpunkcyjnymi.

## 4. Zasady lematyzacji i anotacji morfosyntaktycznej

Lematyzacja korpusu została przeprowadzona w pierwszym etapie prac, podczas przygotowania wersji stacjonarnej korpusu. Większa część segmentów została zlematyzowana przy użyciu słownika analizatora morfologicznego Morfeusz SLaT<sup>1</sup>. Słownik morfologiczny zawiera dla wielu form więcej niż jedną interpretację leksykalną i w takich sytuacjach do wyboru jednej z nich wykorzystano listę frekwencyjną, zbudowaną na podstawie korpusu *Słownika frekwencyjnego polszczyzny współczesnej*<sup>2</sup>. Dla wyrazów niewystępujących w słowniku Morfeusza sporządzony został osobny słownik morfologiczny. W obecnej wersji korpusu ta lematyzacja została zachowana.

Podczas przygotowania korpusu do publikacji w Internecie segmentom została przypisana informacja gramatyczna: znaczniki klas gramatycznych oraz kategorii gramatycznych (zob. p. 7). Tagowanie znacznikami morfosyntaktycznymi przeprowadzone zostało za pomocą tagera Concraft 2, opartego na analizatorze morfologicznym Morfeusz 2. Analizator zwraca wszystkie możliwe interpretacje dla danego słowa, a tager wybiera najbardziej prawdopodobną ze względu na swój model statystyczny. Interpretacje Morfeusza pochodzą ze *Słownika gramatycznego języka polskiego* (SGJP). Jeśli danego słowa nie da się w żaden sposób zinterpretować jako formy wyrazowej leksemu zanotowanego w SGJP (w korpusie tekstów z Maćkowiec dotyczy to przede wszystkim wyrazów dyferencjalnych), to Morfeusz nie zwróci żadnej interpretacji. Wówczas tager „zgaduje” znacznik morfosyntaktyczny,

<sup>1</sup> Zob. Marcin Woliński, *Morfeusz — a practical tool for the morphological analysis of Polish*, (w:) Mieczysław A. Kłopotek, Sławomir T. Wierchoń, Krzysztof Trojanowski (red.), *Intelligent Information Processing and Web Mining*, Advances in Soft Computing, Berlin: Springer-Verlag 2006, s. 503–512.

<sup>2</sup> *Słownik frekwencyjny polszczyzny współczesnej*, opr. I. Kurcz, A. Lewicki, J. Sambor, K. Szafran, J. Woronczak, red. Z. Saloni, t. 1—2, Kraków 1990; nowsza wersja korpusu obecnie dostępna jest na stronie <http://clip.ipipan.waw.pl/PL196x>.

czyli wybiera taki, który zgodnie z jego modelem jest najbardziej prawdopodobny. Skuteczność zgadywania jest w oczywisty sposób dużo niższa niż skuteczność wybierania spośród gotowych interpretacji z Morfeusza, więc znakowanie wyrazów dyferencjalnych zawiera więcej błędów niż tagowanie form leksemów ogólnopolskich.

Prezentowane przy każdym segmencie lematy i znaczniki morfosyntaktyczne pochodzą zatem z dwóch różnych źródeł: lematy z analizatora Morfeusz SIA T, uzupełnionego słownikiem morfologicznym wyrazów dyferencjalnych oraz prostego narzędzia ujednoznaczniającego, posługującego się listą frekwencyjną, a znaczniki morfosyntaktyczne z analizatora Morfeusz 2 i tagera Concraft 2. W rezultacie mogą pojawiać się niespójne interpretacje, np. segment *Boże* we fragmencie *na Boże narodzenie* ma interpretację [Bóg:adj:sg:acc:n:pos], a zatem został błędnie zlematyzowany jako rzeczownik BÓG, lecz otrzymał prawidłową interpretację gramatyczną jako przymiotnik w stopniu równym, w bierniku liczby pojedynczej rodzaju nijakiego. A jednak w skali całego korpusu pierwotna lematyzacja przyniosła znacznie lepsze rezultaty niż lematyzacja przeprowadzona tagerem Concraft 2, który nie ma słownika do interpretacji wyrazów dyferencjalnych; z kolei tagowanie morfosyntaktyczne tagerem Concraft 2 przyniosło całkiem niezłe rezultaty. Dlatego też zdecydowaliśmy się na prezentację w korpusie połączenia tych informacji.

## 5. Interfejs wyszukiwarki MTAS

Interfejs wyszukiwarki zawiera pasek wyszukiwania, w którym wpisuje się zapytanie skonstruowane zgodnie ze składnią zapytań programu MTAS (por. p. 6). Po prawej stronie paska znajdują się przyciski umożliwiające wpisanie liter, które nie występują w polskiej ortografii, ale mogą się pojawiać w zapisie fonetycznym przeszukiwanych tekstów.

Poniżej paska wyszukiwania znajdują się dwa przyciski dające dodatkowe możliwości pracy z wyszukiwarką. Przycisk KONSTRUKTOR ZAPYTAŃ daje dostęp do prostej nakładki na wyszukiwarkę, która ułatwia zdefiniowanie warunków określających cechy segmentu lub sekwencji segmentów występujących w zapytaniu, tj. postać segmentu w obu warstwach zapisu, formę hasłową, znaczenie, osobę mówiącą, etymologię, czy jest to fragment recytacji, wypowiedzi nie po polsku lub kwestionariusza gramatycznego, czy jest to jednowyrazowy wtret obcy, czy jest to nazwa własna (fragment nazwy własnej), czy jest to wymówienie niewyraźne bądź urwane, czy jest to komentarz osoby przepisującej tekst. Poszczególne warunki w obrębie segmentu mogą być łączone operatorami *oraz* (koniunkcja) i *lub* (alternatywa). Zbudowane za pomocą konstruktora zapytanie pojawi się w pasku wyszukiwania, dzięki czemu można zweryfikować jego poprawność. Przycisk ODRZUĆ SŁOWA EKSPLOATORA I KOMENTARZE spowoduje dodanie do zapytania ograniczenia, dzięki któremu wśród wyników wyszukiwania nie pojawią się segmenty wypowiedziane przez eksploratorki oraz będące komentarzami osoby przepisującej.

Wyniki wyszukiwania mogą wyświetlać się w zapisie fonetycznym lub standaryzowanym. Sposób prezentacji wyników wybieramy w oknie WARSTWA WYŚWIETLANIA. Na liście wyników wyszukiwana sekwencja segmentów opatrzonych znacznikami zostaje wyróżniona kolorem, a po nakierowaniu na nią kursora pokazuje się dymek z informacjami o danym segmencie. Po kliknięciu na wybrany wynik na dole strony pokazuje się szerszy kontekst. Niektóre segmenty zostały w nim wyróżnione następującymi stylami:

- komentarz – szare tło
- cytat obcy lub fragment dłuższej wypowiedzi nie po polsku – niebieska czcionka
- recytacja – kursywa
- kwestionariusz gramatyczny – żółte tło.

W oknie LICZBA WYNIKÓW NA STRONĘ można określić, ile wyszukiwanych pozycji ma być widocznych jednocześnie.

## 6. Język zapytań

Składnia zapytań w programie MTAS została oparta na języku zapytań o nazwie Corpus Query Language (CQL), wykorzystywanym w wielu innych tego typu programach, m.in. w programie

Sketch Engine, ale też w znanym z NKJP Poliqarpie. Należy jednak zwrócić uwagę na drobne różnice, ponieważ mogą one wpływać na poprawność formułowanych zapytań.

Niniejsza instrukcja nie uwzględnia wszystkich możliwości wyszukiwarki, częściowo ze względu na jej skrótowy charakter, a częściowo dlatego, że nie wszystkie zapytania będą miały sens w odniesieniu do korpusu gwarowego. Podstawowa dokumentacja wyszukiwarki znajduje się na jej [stronie internetowej](#).

## 6.1. Zapytania o segmenty

Podstawową jednostką wyszukiwaną w korpusie jest segment. Segmenty w zapytaniach są ograniczone nawiasami kwadratowymi, wewnątrz których można określać konkretne cechy, które segment ma spełniać. W najprostszym przypadku jest to kształt tekstowy (napis) w zapisie standaryzowanym. Do zapytań o tę postać standaryzowaną segmentu służy atrybut `stand`, można też jednak ograniczyć się do wpisania w oknie wyszukiwarki poszukiwanego słowa (lub słów). Zatem poniższe zapytanie o dwa sąsiadujące ze sobą segmenty:

```
[stand="wiem"] [stand="gdzie"]
```

można zadać również w prostszy sposób:

```
wiem gdzie
```

Domyślnie rozróżniana jest kasztowość (wielkość) liter, a zatem poniższe dwa zapytania dadzą różne wyniki:

```
wiem
```

```
Wiem
```

W zapytaniach o segmenty mogą wystąpić standardowe wyrażenia regularne wykorzystujące następujące znaki specjalne: `?`, `*`, `+`, `.`, `,`, `|`, `[`, `]`, `(`, `)` oraz liczby naturalne pisane cyframi arabskimi, np. `0` czy `21`. Ponieważ formalny opis wyrażeń regularnych wykracza poza ramy niniejszej instrukcji, ograniczymy się tutaj do kilku przykładów, które powinny pozwolić użytkownikowi na szybkie przyswojenie składni i znaczenia takich wyrażeń (przykłady w tym punkcie pochodzą ze [Ściągawki](#) do NKJP; tam też znajduje się więcej przykładów).

1. `[stand="(Ala|Ela)"]`

znak `|` oznacza alternatywę dwóch wyrażeń (całość należy dodatkowo ująć w nawiasy okrągłe), a zatem zapytanie to może zostać użyte do znalezienia wszystkich wystąpień segmentów *Ala* lub *Ela*,

2. `[stand="[AE]la"]`

nawiasy kwadratowe oznaczają alternatywę znaków, a zatem zapytanie to może zostać użyte do znalezienia tych segmentów, których pierwszy znak to *A* lub *E*, po którym następuje ciąg znaków postaci *la*, tj. zapytanie to jest równoważne poprzedniemu,

3. `[stand="beza?"]`

znak zapytania oznacza opcjonalność znaku (tutaj ostatniego *a*) lub ujętego w nawiasy okrągłe wyrażenia bezpośrednio poprzedzającego znak `?`, a zatem w wyniku zadania tego zapytania znalezione zostaną segmenty *bez* i *beza*,

4. `[stand="bez. "]`

kropka oznacza dowolny znak, a zatem wynikiem tego zapytania będą segmenty *beza*, *bezy*, *bezq* itp., ale nie *bez* czy *bezami*,

5. `[stand="bez. ?"]`

znak zapytania oznacza opcjonalność znaku lub ujętego w nawiasy okrągłe wyrażenia bezpośrednio przed nim, a zatem w wyniku zadania tego zapytania znalezione zostaną *bez*, *beza*, *bezy*, *beza* itp., ale nie *bezami*,

6. [stand="a\*by"]

gwiazdka oznacza dowolną liczbę wystąpień znaku lub wyrażenia bezpośrednio przed nią, a zatem zapytanie to może posłużyć do znalezienia segmentów składających się z dowolnej liczby liter *a*, po których następuje ciąg *by*, np. *by* (zero wystąpień *a*), *aby*, *aaaaby* itp.,

7. [stand="Ala.\*"]

segmenty zaczynające się na *Ala*, np. *Ala* i *Alabama*,

8. [stand=".\*al+"]

plus ma działanie podobne do gwiazdki i oznacza dowolną większą od zera liczbę wystąpień znaku lub wyrażenia bezpośrednio przed nim, a zatem wynikiem tego zapytania będzie znalezienie segmentów kończących się na *al*, *all*, *alll* itd., ale nie na *a*, np. *dal*, *robal* i *Gall*.

## 6.2. Zapytania z innymi atrybutami

Do wyszukiwania segmentów w zapisie fonetycznym służy atrybut *phon*, natomiast atrybut *base* pozwala wyszukać wszystkie wystąpienia leksemu według jego formy podstawowej. Zatem aby znaleźć segment *dz'es'* w jego zapisie fonetycznym, należy użyć następującego zapytania:

```
[phon="dz'es'"]
```

Wartością atrybutu *base* powinna być specyfikacja formy podstawowej (hasłowej), a zatem zapytanie [base="pisać"] może być użyte do znalezienia form typu *pisać*, *pisze*, *pisala*, *piszcie*, *pisanie*, *pisano*, *pisane* itp.

Podobnie jak w wypadku atrybutu *stand* wartościami atrybutów *phon* i *base* mogą być wyrażenia regularne, np.:

```
[phon="b[ye]ł.*"]
```

```
[base="w[ae]reta"]
```

Pierwsze zapytanie pozwoli wyszukać wszystkie formy liczby pojedynczej czasu przeszłego leksemu BYĆ ze standardową lub obniżoną wymową pierwszej samogłoski. Wynikiem drugiego zapytania będą wszystkie segmenty, których forma hasłowa ma postać WERETA lub WARETA.

Zapytania o różne atrybuty segmentów można łączyć. Na przykład, aby znaleźć wszystkie wystąpienia leksemu SIĘ z półmiękką realizacją początkowej spółgłoski, można zadać następujące zapytanie:

```
[base="się" & phon="s'.*"]
```

W powyższym zapytaniu operator *&* spełnia rolę logicznej koniunkcji (oznacza on, że oba warunki muszą być spełnione). Operator *|* spełniający rolę logicznej alternatywy (musi być spełniony co najmniej jeden warunek). Na przykład zapytanie:

```
[base="ładnie" | base="ładno"]
```

pozwoli odnaleźć wszystkie wystąpienia przysłówków ŁADNIE i ŁADNO; równoważne zapytaniu [base="ładnie|ładno"].

Aby lepiej zrozumieć różnice pomiędzy operatorami *&* i *|*, porównajmy następujące dwa zapytania:

```
[base="się" & phon="s'.*"]
```

```
[base="się" | phon="s'.*"]
```

W wyniku zadania pierwszego zapytania znalezione zostaną te segmenty, które są jednocześnie (koniunkcja) formą leksemu SIĘ i w zapisie fonetycznym rozpoczynają się od znaków *s'*, a więc np. *s'e*, *s'i*, *s'e*. W wyniku zadania drugiego zapytania znalezione natomiast zostaną zarówno wszystkie wystąpienia leksemu SIĘ, jak i wszystkie formy, które rozpoczynają się w zapisie fonetycznym od *s'*, a więc zarówno *śe*, *się*, jak i *s'e*, *s'imj`a*, *s'eli* itd.

Operator *!* oznacza, że wyniki wyszukiwania mają nie spełniać zapisanego po nim warunku. Zatem np. zapytanie:



```
[stand="ładn.*" & !base="ładno"]
```

oznacza, że mają zostać znalezione wszystkie formy zaczynające się w warstwie standaryzowanej od *ładn*, ale niebędące formami przysłówka ŁADNIE.

Specyfikacje pozycji w korpusie, ujęte w nawiasy kwadratowe, mogą zawierać dowolną liczbę warunków typu `atrybut="wartość"` (na przykład `stand="gdzieś"`) połączonych operatorami `&` i `|`, tak jak pokazują to powyższe przykłady. Możliwe jest także całkowite pominięcie jakichkolwiek warunków. Poniższe zapytanie mogłoby posłużyć do znalezienia wszystkich segmentów w korpusie.

```
[]
```

Taka „pusta” specyfikacja pozycji w korpusie, pasująca do dowolnego segmentu, może posłużyć na przykład do znalezienia dwóch form oddzielonych od siebie dowolnymi dwoma segmentami, np.:

```
[stand="się" ] [] [base="bać" ]
```

W wyniku tego zapytania zostaną znalezione ciągi takie jak *się mnie też bać* czy *się nie chce bać*.

Dla wielu zastosowań ciekawsza byłaby możliwość zapytania na przykład o formy oddalone od siebie o najwyżej pięć pozycji. MTAS umożliwia zadawanie takich pytań, gdyż pozwala na formułowanie wyrażeń regularnych także na poziomie pozycji korpusu. Na przykład zapytanie o formę leksemu BAĆ występująca dwie, trzy lub cztery pozycje dalej niż forma *się* może wyglądać następująco:

```
[stand="się" ] [] {2,4} [base="bać" ]
```

W wyniku tego zapytania zostaną znalezione ciągi uzyskane w wyniku poprzedniego zapytania, a także na przykład ciąg *się pani niczego nie boi*.

### 6.3. Ograniczenie zapytania za pomocą atrybutów leksykalnych

Segmenty w korpusie zostały opatrzone dodatkowymi informacjami, pomocnymi w badaniach nad leksyką. Mogą one posłużyć m.in. do oddzielenia wypowiedzi informatorów od pytań eksploratorek, do wyszukania wszystkich synonimów (opatrzonych identycznym znaczeniem), do wyszukania wyrazów ogólnopolskich, zapożyczonych, gwarowych itd. czy też do wykluczenia z badań fragmentów recytacji, nazw własnych itp.

Wszystkie segmenty poza komentarzami mają przypisane następujące informacje:

- osoba mówiąca
- znaczenie
- temat rozmowy
- czy jest to cytat obcy
- czy jest to fragment dłuższej wypowiedzi nie po polsku
- czy jest to fragment recytacji
- czy jest to fragment kwestionariusza gramatycznego
- czy jest to nazwa własna
- czy jest to wymówienie niewyraźne lub urwane (niepodlegające interpretacji).

Znaczenie i etymologię mają przypisane wszystkie segmenty poza komentarzami, wtrętami obcymi oraz wymówieniami niewyraźnymi; komentarze, segmenty niepolskie oraz wymówienia niewyraźne nie są opatrzone żadnymi znacznikami prócz `phon` i `stand`.

W celu wyszukania wszystkich wystąpień leksemu NAZYWAĆ wypowiedzianych przez informatorów (nie eksploratorki) należy wpisać następujące zapytanie:

```
[base="nazywać" & !inform="Eksplorator"]
```

Jeśli dodatkowo chcemy wykluczyć fragmenty wypowiedziane w ramach recytacji, należy wpisać:

```
[base="nazywać" & !inform="Eksplorator" & !recyt="tak"]
```

W celu wyszukania wszystkich wyrazów zapożyczonych, ale innych niż nazwy własne, należy wpisać następujące zapytanie:

```
[etym="Zap" & !nazwa="(onom|nazwaRel|nazwaIn)"]
```

Jeśli chcemy znaleźć wszystkie wystąpienia leksemów mających znaczenie ‘córka’, należy wpisać następujące zapytanie:

```
[def="córka"]
```

### 6.3.1. Znaczenie i etymologia

Dla wyrazów zgodnych z ogólnopolskimi znaczenie jest identyczne z formą podstawową, np. CHODZIĆ ‘chodzić’. Wyrazom dyferencjalnym (zapożyczonym, gwarowym, archaicznym, innowacyjnym, kontaminowanym) przypisany został odpowiednik ogólnopolski. Zwykle jest to odpowiednik jednowyrazowy (np. ABRAZOWANY ‘wykształcony’), a tylko czasem składa się z dwóch lub więcej wyrazów, zawsze oddzielonych podkreślnikiem (np. ADINOCZKA ‘samotna\_matka’). W przypadku większości homonimów lematowi przypisane są dwa (niekiedy trzy) znaczenia, rozdzielone średnikiem bez spacji (np. CZLEK ‘człowiek;mąż’). Wszystkie wystąpienia tego leksemu w korpusie otrzymały w ten sposób zapisane znaczenie. Tylko niekiedy, w toku ręcznych poprawek, znaczenia zostały rozdzielone (np. WPRZÓD 1. ‘wprzód’, 2. ‘do\_przodu’).

Z powodu przyjętego podziału na segmenty nie było możliwe przypisywanie znaczeń czasownikom z *się* jako całości. Arbitralnie znaczenia przypisane są części czasownika przed *się*, np. PISAĆ ‘mieć na nazwisko’ – choć znaczenie to ma cały leksem PISAĆ SIĘ.

Wyrazy przydzielone zostały pod względem etymologii do następujących kategorii: leksem ogólnopolski (P), polski gwarowy (G), polski archaiczny (Ar), polski archaiczny lub gwarowy (GAR; w wielu wypadkach nie da się tego ustalić), polski gwarowy ze wsparciem obcojęzycznym (Gw), polski archaiczny ze wsparciem obcojęzycznym (Arw), polski archaiczny lub gwarowy ze wsparciem obcojęzycznym (GARw), zapożyczony z języka ukraińskiego (U), zapożyczony z języka rosyjskiego (R), zapożyczony z języka ukraińskiego lub rosyjskiego (UR; tu znów często trudno jest o rozstrzygnięcie), innowacyjny (In), kontaminowany (Ko). Wśród rusycyzmów znajduje się grupa zapożyczonych przez medium języka ukraińskiego (RM). Wśród wyrazów innowacyjnych znajdują się innowacje na bazie wyrazów ukraińskich (InU), rosyjskich (InR), ukraińskich lub rosyjskich (InUR) oraz na bazie wyrazu archaicznego ze wsparciem obcojęzycznym (InArw). Homonimy otrzymały więcej niż jeden kwalifikator etymologiczny, np. leksem NAZWISKO ‘nazwisko;nazwa’ ma kwalifikator PGAR, gdyż w znaczeniu pierwszym jest wyrazem ogólnopolskim, a w drugim – wyrazem gwarowym lub archaicznym. Objasnienia wszystkich skrótów oznaczających etymologię oraz zakres tematyczny rozmowy znajdują się w p. 8.

Dla ułatwienia skrótów oznaczające etymologię zostały zebrane w kilka grup. Pierwsza to wszystkie wyrazy ogólnopolskie, także te, które tylko w pierwszym ze znaczeń są ogólnopolskie (czyli nie tylko wyrazy z kwalifikatorem P, ale też np. PGAR), druga to wszystkie wyrazy rodzime (Rodz), trzecia – zapożyczenia (Zap), czwarta – wszystkie wyrazy dyferencjalne (czyli nieogólnopolskie, Dyf), piąta – wszystkie wyrazy dyferencjalne rodzime (DyfP). Wyrazy o podwójnej lub potrójnej etymologii (homonimy) znajdują się zawsze tylko w jednej z tych grup, zgodnie z tym, które znaczenie zostało uznane za najczęstsze i wpisane na pierwszym miejscu. Zatem zapytanie:

```
[etym="Zap"]
```

jest równoważne zapytaniu:

```
[etym="(InR|InU|InUR|R|RM|U|UR|URKo|URP)"]
```

Zapytanie:

```
[etym="DyfP"]
```

jest równoważne zapytaniu:

```
[etym="(Ar|Arw|G|GAR|GARR|GARw|GP|GU|Gw|InArw)"]
```

a zapytanie:

```
[etym="Dyf"]
```



jest równoważne zapytaniu:

[etym="(Ar|Arw|G|GAR|GARR|GARw|GP|GU|Gw|In|InArw|InR|InU|InUR|Ko|R|RM|U|UR|URKo|URP) " ]

Wszystkie wyrazy, które mają znaczenie nierównobrzmiące z lematem lub też kwalifikator etymologiczny inny niż P znalazły się w [Słowniku](#).

Wykaz wszystkich atrybutów innych niż znaczniki morfosyntaktyczne oraz ich możliwych wartości prezentuje Tabela 1.

atrybut	kategoria	możliwe wartości
koment	komentarz	tak nie
niewyr	wymówienie niewyraźne, urwane i nierozpoznane	tak
inform	osoba mówiąca	AB24 AB33 AL23 AS20 BB26 Eksplorator FB28 FJ12 H40 HB27 KB25 KP23 KP27 KP30 KR31 ML45 MM17 S54 SP23 WB29 WL22 X
etym	etymologia wyrazu	Dyf DyfP Ogp Zap  Ar Arw G GAR GARR GARw GP GU Gw In InArw InR InU InUR Ko P PAr PArw PArwG PG PGAR PGArw PGw PIn PInUR PR PU PUIIn PUR R RM U UR URKo URP
nazwa	nazwy własne	onom nazwaRel nazwaIn
cytat	cytat ukraiński lub rosyjski (z obcą końcówką lub wskazany przez informatora)	cytGr cytInf
npol	fragmenty wypowiedziane nie po polsku	tak
recyt	fragmenty wierszy, piosenek, modlitw i innych tekstów nauczonych	tak
kwestGr	kwestionariusz gramatyczny	tak
temat	temat rozmowy	kultura gospodarstwo inne oSobie tekst kAntroponimia kChata kChoroby kCialo kCmentarz kCzas kHistoria kJarmark kKartofle kKosciol kLas kLen kMiejsceNaUbrania kMleko kNaczynia kObejscie kObuwie kOkreslenia kOswietlenie kOwady kPlug kPotrawy kPraca kPranie kPtactwo kRodzina kRzemieslnicy kSad kSasiednieMiejscowosci kSasiedzi kSiano kSpanie kStol kSwieta kTeren kUbranie kVaria kWPolu kWDomu kWierzenia kWoz kZCzegoJedli kZabijac kZboze kZupy kZwierzeta kZycie kLeksykalny

**Tabela 1.** Wykaz atrybutów wraz z możliwymi wartościami

## 7. Zestaw znaczników morfosyntaktycznych

Każdy znacznik morfosyntaktyczny jest ciągiem wartości rozdzielonych dwukropkami, np.: subst:sg:nom:m1 dla segmentu *chłopiec*. Pierwsza wartość, np. subst, określa klasę gramatyczną (por. p. 7.2), następne zaś, np. sg, nom i m1 wartości odpowiednich dla tej klasy kategorii gramatycznych (por. p. 7.1).

### 7.1. Kategorie gramatyczne

Tabela 2 przedstawia repertuar kategorii gramatycznych używanych w znakowaniu tekstów. Repertuar kategorii pochodzi z tagsetu [analizatora morfologicznego Morfeusz](#), który jest podobny do tagsetu NKJP, ale zawiera też kilka modyfikacji. Najważniejsza cecha tagsetu Morfeusza w stosunku do tagsetu NKJP to wprowadzenie obok kategorii rodzaju opcjonalnej tzw. kategorii przyrodzaju o wartościach pt, col i ncol — przysługują one wyłącznie klasom subst i num. W wypadku subst informują o łączliwości danego rzeczownika rodzaju nijakiego z liczebnikami głównymi (ncol, np. OKNO) lub zbiorowymi (col, np. DZIECKO) oraz o braku w paradygmacie rzeczownika form liczby pojedynczej (pt, np. SKRZYPCE). W wypadku klasy num wartość kategorii przyrodzaju informuje o tym, że dana forma jest zbiorowa (col, np. *dwoje*) lub niezbiorowa (ncol, np. *dwa*). Konsekwencją wprowadzenia kategorii przyrodzaju w klasie num było usunięcie z tagsetu klasy numcol. Inne różnice to m.in.:

- ograniczenie klasy przymiotników poprzyimkowych *adjp* wyłącznie do form dopełniacza (np. *(z) wolna*) i celownika (np. *(po) polsku*) dawnej odmiany niezłożonej przymiotników oraz dodanie im wartości przypadku;
- przemianowanie klas qub i burk odpowiednio na part i frag.

### 7.2. Klasy gramatyczne

W Elektronicznym Korpusie Tekstów Gwarowych z Maćkowiec na Podolu, podobnie jak w NKJP, klasy gramatyczne rozumiane są morfosyntaktycznie - są one oparte na pojęciu fleksemu, będącym pojęciem węższym od terminu leksem. O ile tradycyjnie ujmowane leksemy mogą zawierać formy, którym przysługują różne kategorie gramatyczne, o tyle fleksemy stanowią zbiory tylko tych form, które można scharakteryzować za pomocą tych samych kategorii gramatycznych<sup>3</sup>.

Tabela 3 zawiera przybliżoną charakterystykę morfosyntaktyczną wszystkich klas fleksyjnych przyjętych w niniejszym tagsecie. Symbol + oznacza, że dla danej klasy fleksyjnej dana kategoria gramatyczna jest morfologiczna (fleksemy należące do tej klasy „odmieniają się” przez tę kategorię), zaś symbol • oznacza, że dana kategoria jest słownikowa (wszystkie formy dowolnego fleksemu należące do tej klasy mają tę samą wartość tej kategorii, choć mogą to być różne wartości dla różnych fleksemów, jak w wypadku rodzaju rzeczowników).

Tabela 4 zawiera informacje o formach podstawowych dla poszczególnych klas fleksyjnych, a także skróty nazw klas fleksyjnych używane w korpusie.

---

<sup>3</sup> Przykładowo w obrębie leksemu czasownikowego można wyróżnić m.in. osobowe formy czasownika (odmienne przez osoby i liczby), bezokolicznik (nieodmienny), a w niektórych ujęciach także odśłownik (odmienny przez przypadki i liczby).

<b>Liczba:</b> (2 wartości)		
pojedyncza	<i>sg</i>	<i>oko</i>
mnoga	<i>pl</i>	<i>oczy</i>
<b>Przypadek:</b> (7 wartości)		
mianownik	<i>nom</i>	<i>woda</i>
dopełniacz	<i>gen</i>	<i>wody</i>
celownik	<i>dat</i>	<i>wodzie</i>
biernik	<i>acc</i>	<i>wodę</i>
narzędnik	<i>inst</i>	<i>wodą</i>
miejscownik	<i>loc</i>	<i>wodzie</i>
wołacz	<i>voc</i>	<i>wodo</i>
<b>Rodzaj:</b> (5 wartości)		
męski osobowy	<i>m1</i>	<i>papież, kto, wujostwo</i>
męski zwierzęcy	<i>m2</i>	<i>baranek, walc, babsztyl</i>
męski rzeczowy	<i>m3</i>	<i>stół</i>
żeński	<i>f</i>	<i>stula</i>
nijaki	<i>n</i>	<i>dziecko, okno, co, skrzypce, spodnie</i>
<b>Przyrodzaj:</b> (3 wartości)		
przymnogi	<i>pt</i>	<i>wujostwo, skrzypce, spodnie</i>
zbiorowy	<i>col</i>	<i>dziecko</i>
niezbiorowy	<i>ncol</i>	<i>okno</i>
<b>Osoba:</b> (3 wartości)		
pierwsza	<i>pri</i>	<i>bredzę</i>
druga	<i>sec</i>	<i>bredzisz</i>
trzecia	<i>ter</i>	<i>bredzi</i>
<b>Stopień:</b> (3 wartości)		
równy	<i>pos</i>	<i>cudny</i>
wyższy	<i>com</i>	<i>cudniejszy</i>
najwyższy	<i>sup</i>	<i>najcudniejszy</i>
<b>Aspekt:</b> (2 wartości)		
niedokonany	<i>imperf</i>	<i>iść</i>
dokonany	<i>perf</i>	<i>zajść</i>
<b>Zanegowanie:</b> (2 wartości)		
niezanegowana	<i>aff</i>	<i>pisanie, czytaniego</i>
zanegowana	<i>neg</i>	<i>niepisanie, nieczytaniego</i>
<b>Akcentowość:</b> (3 wartości)		
akcentowana	<i>akc</i>	<i>niego, jego, tobie</i>
nieakcentowana	<i>nakc</i>	<i>go, -ń, ci</i>
zneutralizowana	<i>neut</i>	<i>one, im, je</i>
<b>Poprzyimkowość:</b> (2 wartości)		
poprzyimkowa	<i>praep</i>	<i>niego, -ń</i>
niepoprzyimkowa	<i>npraep</i>	<i>jego, go</i>
<b>Akomodacyjność:</b> (2 wartości)		
uzgadniająca	<i>congr</i>	<i>dwaj, pięcioma</i>
rządząca	<i>rec</i>	<i>dwóch, dwu, pięciorgiem</i>
<b>Aglutynacyjność:</b> (2 wartości)		
nieaglutynacyjna	<i>nagl</i>	<i>niósł</i>
aglutynacyjna	<i>agl</i>	<i>niosł-</i>
<b>Wokaliczność:</b> (2 wartości)		
wokaliczna	<i>wok</i>	<i>-em</i>
niewokaliczna	<i>nwok</i>	<i>-m</i>
<b>Kropkwalność:</b> (2 wartości) <sup>4</sup>		
z następującą kropką	<i>pun</i>	<i>tzn</i>
bez następującej kropki	<i>npun</i>	<i>wg</i>

**Tabela 2.** Kategorie gramatyczne

<sup>4</sup> Rzecz jasna, w korpusie języka mówionego nie występują skróty, jednak pewna liczba segmentów otrzymała taką interpretację w wyniku błędów tagera.

	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	+				+		•							
pseudoimiesłów	+		+				•					+		
rozkaznik	+				+		•							
bezosobnik							•							
bezokolicznik							•							
odsłownik	+	+	•				•	+						
im. przysł. współczesny							•							
im. przysł. uprzedni							•							
im. przym. czynny	+	+	+			+	•	+						
im. przym. bierny	+	+	+			+	•	+						
aglutynant BYĆ	+				+		•						+	
winien	+		+				•							
przyimek		•											+	
spójnik współrzędny														
spójnik podrzędny														
partykuła													+	
skrót														+
człon wyrażenia														
wykrzyknik														
znak interpunkcyjny														
ciało obce														

Tabela 3. Klasy gramatyczne

<b>fleksem</b>	<b>skrót</b>	<b>forma podstawowa</b>	<b>przykład</b>
rzeczownik	<i>subst</i>	mianownik l. poj.	<i>doktor</i>
rzeczownik deprecjatywny	<i>depr</i>	mianownik l. poj. rzeczownika	<i>doktor</i>
liczebnik główny	<i>num</i>	mianownik rodz. m3	<i>pięć, dwa</i>
przymiotnik	<i>adj</i>	mianownik l. poj. rodzaju męskiego st. równego	<i>polski</i>
przymiotnik przyprzym.	<i>adja</i>	mianownik l. poj. rodz. męskiego przymiotnika w st. równym	<i>polski</i>
przymiotnik poprzym.	<i>adjp</i>	mianownik l. poj. rodz. męskiego przymiotnika w st. równym	<i>polski</i>
przysłówek	<i>adv</i>	forma stopnia równego	<i>dobrze, bardzo</i>
zaimek nietrzeciosobowy	<i>ppron12</i>	mianownik l. poj.	<i>ja</i>
zaimek trzeciosobowy	<i>ppron3</i>	mianownik l. poj.	<i>on</i>
zaimek SIEBIE	<i>siebie</i>	biernik	<i>siebie</i>
forma nieprzeszła	<i>fin</i>	bezokolicznik	<i>czytać</i>
forma przyszła BYĆ	<i>bedzie</i>	bezokolicznik	<i>być</i>
aglutynant BYĆ	<i>aglt</i>	bezokolicznik	<i>być</i>
pseudoimiesłów	<i>praet</i>	bezokolicznik	<i>czytać</i>
rozkaznik	<i>impt</i>	bezokolicznik	<i>czytać</i>
bezosobnik	<i>imps</i>	bezokolicznik	<i>czytać</i>
bezokolicznik	<i>inf</i>	bezokolicznik	<i>czytać</i>
im. przys. współczesny	<i>pcon</i>	bezokolicznik	<i>czytać</i>
im. przys. uprzedni	<i>pant</i>	bezokolicznik	<i>czytać</i>
odsłownik	<i>ger</i>	bezokolicznik	<i>czytać</i>
im. przym. czynny	<i>pact</i>	bezokolicznik	<i>czytać</i>
im. przym. bierny	<i>ppas</i>	bezokolicznik	<i>czytać</i>
winien	<i>winien</i>	forma męska l. poj.	<i>winien, rad</i>
predykatyw	<i>pred</i>	jedyna forma fleksemu	<i>warto</i>
przyimek	<i>prep</i>	niewokaliczna forma fleksemu	<i>na, przez, w</i>
spójnik współrz.	<i>conj</i>	jedyna forma fleksemu	<i>oraz</i>
spójnik podrz.	<i>comp</i>	jedyna forma fleksemu	<i>że</i>
partykuła	<i>part</i>	jedyna forma fleksemu	<i>nie, -li, się</i>
skrót <sup>5</sup>	<i>brev</i>	forma hasłowa rozwinięcia skrótu	<i>rok, i tak dalej</i>
człon wyrażenia	<i>frag</i>	jedyna forma fleksemu	<i>wskroś, dala</i>
wykrzyknik	<i>interj</i>	jedyna forma fleksemu	<i>laboga, pst</i>
znak interpunkcyjny	<i>interp</i>	jedyna forma fleksemu	<i>;, !, ?</i>
ciało obce	<i>xxx</i>	jedyna forma fleksemu	<i>osnownoje</i>

**Tabela 4.** Skróty nazw klas gramatycznych oraz ich formy hasłowe.

<sup>5</sup> Zob. przypis 3.

### 7.3. Zapytania o znaczniki morfosyntaktyczne

Wartościami atrybutu `pos` (ang. *part of speech* ‘część mowy’) są skróty nazw klas gramatycznych omówionych w p. 7.2 (por. tabela 3). Na przykład zapytanie o sekwencję dwóch form rzeczownikowych rozpoczynających się na *a* może być sformułowane w sposób następujący:

```
[pos="subst" & stand="a.*"]{2}
```

Podobnie jak to miało miejsce w wypadku specyfikacji form obu warstw tekstowych i form hasłowych, także specyfikacje klas gramatycznych mogą zawierać wyrażenia regularne. Na przykład, zważywszy na to, że zaimki osobowe należą do klasy zaimków trzecioosobowych *ppron3* i do klasy zaimków nietrzecioosobowych *ppron12*, poniższe zapytania mogą posłużyć do znalezienia dowolnych form dowolnych zaimków osobowych:

```
[pos="ppron12" | pos="ppron3"]
```

```
[pos="ppron12|ppron3"]
```

```
[pos="ppron(12|3)"]
```

```
[pos="ppron[123]+"]
```

```
[pos="ppron.+"]
```

W zapytaniach można określić nie tylko postać w zapisie fonetycznym lub standaryzowanym segmentu (za pomocą atrybutu `phon` lub `stand`), formę hasłową (za pomocą `base`) i klasę gramatyczną (za pomocą `pos`), lecz także wartości poszczególnych kategorii gramatycznych, np. przypadku czy rodzaju. Służą do tego następujące atrybuty (por. p. 7.1):

atrybut	kategoria	możliwe wartości
<code>number</code>	liczba	<code>sg pl</code>
<code>case</code>	przypadek	<code>nom gen dat acc inst loc voc</code>
<code>gender</code>	rodzaj	<code>m1 m2 m3 f n</code>
<code>subgender</code>	przyrodzaj	<code>col ncol pt</code>
<code>person</code>	osoba	<code>pri sec ter</code>
<code>degree</code>	stopień	<code>pos comp sup</code>
<code>aspect</code>	aspekt	<code>imperf perf</code>
<code>negation</code>	zanegowanie	<code>aff neg</code>
<code>accentability</code>	akcentowość	<code>akc nakc</code>
<code>post-prepositionality</code>	poprzyimkowość	<code>npraep praep</code>
<code>agglutination</code>	aglutynacyjność	<code>agl nagl</code>
<code>vocalicity</code>	wokaliczność	<code>nwok wok</code>
<code>fullstoppedness<sup>6</sup></code>	kropkowność	<code>pun npun</code>

A zatem możliwe jest zadanie na przykład następujących zapytań:

1. `[number="sg"]`

- znalezione zostaną wszystkie formy w liczbie pojedynczej,

2. `[pos="subst" & number="sg"]`

<sup>6</sup> Por. przypis 3.



- znalezione zostaną formy rzeczownikowe w liczbie pojedynczej,
- 3. [pos="subst" & !gender="f"]
- znalezione zostaną formy rzeczownikowe rodzaju męskiego lub nijakiego,
- 4. [number="sg" & case="nom|acc" & gender="m[123]"]
- znalezione zostaną pojedyncze mianownikowe lub biernikowe formy męskie.

O klasy gramatyczne i kategorie gramatyczne można także pytać łącznie, używając do tego atrybutu `tag`. Na przykład, aby znaleźć wszystkie rzeczowniki żeńskie w mianowniku o pojedynczej wartości liczby, można zadać następujące zapytanie:

```
[tag="subst:sg:nom:f"]
```

Wartości atrybutu `tag` mają postać `kl:kat1:kat2:...:katn`, gdzie `kl` to nazwa klasy gramatycznej, a `kati` to wartości kategorii przysługujących tej klasie w kolejności, w jakiej zostały podane w Tabeli 3. Jak w wypadku innych atrybutów, specyfikacja atrybutu `tag` może być zadana wyrażeniem regularnym, np.:

```
[tag=".*:sg:(nom|acc):m[123].*"]
```

Ponieważ nazwy wartości poszczególnych kategorii są rozłączne, można również stosować zbiorczą kategorię `feat` (ang. *feature* ‘cecha’) w zastępstwie każdej innej. Ujednoznaczenie dokona się przez odpowiednią wartość. Dlatego następujące dwa zapytania zwrócą te same wyniki:

- [pos="subst" & case="acc" & number="pl" & gender="f"]
- [pos="subst" & feat="acc" & feat="pl" & feat="f"]

## 8. Wykaz skrótów

### 8.1. Skróty oznaczające etymologię leksemów:

Ar – archaizm polski

Arw – archaizm polski ze wsparciem podobnie brzmiącego wyrazu ukraińskiego lub rosyjskiego

Dyf – leksem dyferencjalny (nieogólnopolski)

DyfP – leksem dyferencjalny rodzimy (wyraz gwarowy lub archaizm polski)

G – leksem gwarowy polski

GAr – leksem gwarowy lub archaizm polski

GArR – dwa homonimy: 1. dialektyzm lub archaizm polski oraz 2. rusycyzm

GArw – leksem gwarowy lub archaizm polski, posiadające wsparcie podobnie brzmiącego wyrazu ukraińskiego lub rosyjskiego

GP – dwa homonimy: 1. wyraz gwarowy polski oraz 2. wyraz ogólnopolski

GU – dwa homonimy: 1. wyraz gwarowy polski oraz 2. zapożyczenie ukraińskie

Gw – leksem gwarowy polski ze wsparciem podobnie brzmiącego wyrazu ukraińskiego lub rosyjskiego

In – innowacja leksykalna

InArw – innowacja leksykalna na bazie polskiego archaizmu posiadającego wsparcie podobnie brzmiącego wyrazu obcego

InR – innowacja leksykalna na bazie rusycyzmu

InU – innowacja leksykalna na bazie ukrainizmu

InUR – innowacja leksykalna na bazie zapożyczenia z języka ukraińskiego lub rosyjskiego

Ko – kontaminacja wyrazu rodzimego i zapożyczonego

Ogp – leksem ogólnopolski (należą tu zarówno wyrazy ogólnopolskie, które nie mają homonimów zapożyczonych, jak i te, które takie homonimy mają)

P – wyraz ogólnopolski, nieposiadający homonimów zapożyczonych ani gwarowych

PAr – dwa homonimy: 1. wyraz ogólnopolski oraz 2. archaizm polski

PArW – dwa homonimy: 1. wyraz ogólnopolski oraz 2. archaizm polski ze wsparciem podobnie brzmiącego wyrazu ukraińskiego lub rosyjskiego

PArWG – trzy homonimy: 1. wyraz ogólnopolski, 2. archaizm polski ze wsparciem podobnie brzmiącego wyrazu ukraińskiego lub rosyjskiego, 3. wyraz gwarowy polski

PG – dwa homonimy: 1. wyraz ogólnopolski oraz 2. wyraz gwarowy polski

PGAr – dwa homonimy: 1. wyraz ogólnopolski oraz 2. wyraz gwarowy lub archaizm polski

PGArW – dwa homonimy: 1. wyraz ogólnopolski oraz 2. wyraz gwarowy lub archaizm polski ze wsparciem podobnie brzmiącego wyrazu ukraińskiego lub rosyjskiego

PGW – dwa homonimy: 1. wyraz ogólnopolski oraz 2. wyraz gwarowy polski ze wsparciem podobnie brzmiącego wyrazu ukraińskiego lub rosyjskiego

PIn – dwa homonimy: 1. wyraz ogólnopolski oraz 2. innowacja leksykalna

PInUR – dwa homonimy: 1. wyraz ogólnopolski oraz 2. innowacja na bazie zapożyczenia z języka ukraińskiego lub rosyjskiego

PR – dwa homonimy: 1. wyraz ogólnopolski oraz 2. zapożyczenie z języka rosyjskiego

PU – dwa homonimy: 1. wyraz ogólnopolski oraz 2. zapożyczenie z języka ukraińskiego

PUIn – trzy homonimy: 1. wyraz ogólnopolski, 2. zapożyczenie z języka ukraińskiego oraz 3. innowacja leksykalna

PUR – dwa homonimy: 1. wyraz ogólnopolski oraz 2. zapożyczenie z języka ukraińskiego lub rosyjskiego

R – zapożyczenie z języka rosyjskiego

RM – zapożyczenie z języka rosyjskiego, którego fonetyka wskazuje na pośrednictwo ukraińskie

U – zapożyczenie z języka ukraińskiego

UR – zapożyczenie z języka ukraińskiego lub rosyjskiego

URKo – kontaminacja zapożyczenia z języka ukraińskiego i rosyjskiego

URP – dwa homonimy: 1. zapożyczenie z języka ukraińskiego lub rosyjskiego oraz 2. wyraz ogólnopolski

Zap – zapożyczenie

## **8.2. Skróty zakresów tematycznych rozmowy:**

kLeksykalny – kwestionariusz A. Wieczorek do badania wybranych leksemów

oSobie – życiorys informatora

tekst – tekst ciągły

**Skróty zakresów tematycznych rozmów prowadzonych na podstawie *Kwestionariusza do badań polskich gwar kresowych J. Riegera* (komputeropis):**

kAntroponimia – antroponimia

kChata – gospodarstwo, chata

kChoroby – choroby i medycyna ludowa

kCialo – części ciała, fizjologia

kCmentarz – cmentarz

kCzas – czas, zjawiska meteorologiczne

kHistoria – z historii

kJarmark – jarmarki

kKartofle – kartofle  
kKosciol – kościół, w kościele  
kLas – las, drzewa, krzaki  
kLen – praca ze lnem  
kMiejsceNaUbrania – miejsce na ubrania  
kMleko – przetwory mleczne  
kNaczynia – naczynia kuchenne  
kObejscie – obejście  
kObuwie – obuwie  
kOkreslenia – określenia ludzi według cech fizycznych i psychicznych  
kOswietlenie – oświetlenie dawniej  
kOwady – owady  
kPlug – pług i jego części  
kPotrawy – potrawy i ich przyrządzenie  
kPraca – praca w domu i koło domu  
kPranie – pranie, toaleta  
kPtactwo – ptactwo domowe  
kRodzina – człowiek, rodzina  
kRzemieslnicy – rzemieślnicy i wyroby rzemieślnicze  
kSad – sad i ogród  
kSasiednieMiejscowosci – sąsiednie miejscowości  
kSasiedzi – sąsiedzi  
kSiano – siano  
kSpanie – spanie  
kStol – stół, ławy  
kSwieta – święta kościelne  
kTeren – ukształtowanie terenu  
kUbranie – ubranie  
kVaria – varia  
kWPolu – praca w polu  
kWDomu – w domu  
kWierzenia – wierzenia i zabobony  
kWoz – wóz i jego części  
kZCzegoJedli – z czego i co jedli  
kZabijac – zabijać zwierzęta domowe  
kZboze – zboże  
kZupy – zupy  
kZycie – życie człowieka od kołyski aż do śmierci  
kZwierzeta – zwierzęta domowe

gospodarstwo – zbiorczy skrót dla tematów: kChata, kKartofle, kLas, kLen, kMiejsceNaUbrania, kMleko, kNaczynia, kObejscie, kObuwie, kOswietlenie, kOwady, kPlug, kPotrawy, kPraca, kPranie, kPtactwo, kSad, kSiano, kSpanie, kStol, kUbranie, kWDomu, kWPolu, kWoz, kZCzegoJedli, kZabijac, kZboze, kZupy, kZwierzeta  
kultura – zbiorczy skrót dla tematów: kCmentarz, kHistoria, kKosciol, kSwieta, oSobie, tekst  
inne – zbiorczy skrót dla tematów: kJarmark, kRzemieslnicy

### **8.3. Inne skróty:**

cytGr – cytat z języka ukraińskiego lub rosyjskiego z obcą końcówką gramatyczną

cytInf – cytat z języka ukraińskiego lub rosyjskiego wskazany przez informatora

nazwaIn – inna nazwa własna

nazwaRel – nazwa własna związana z praktykami religijnymi

onom – nazwa własna osobowa lub miejscowa

Opracowanie: Aleksandra Wiczorek