

Instrukcja znakowania tekstów polskich jednostkami leksykalnymi *Słowsieci* za pośrednictwem programu **WebAnno**

Elżbieta Hajnicz i Małgorzata Marciniak

Spis treści

1	Informacje wstępne	2
2	Korekta znakowania morfoskładniowego	3
2.1	Skróty i skrótowce	4
2.2	Gerundia i imiesłowy	5
2.3	Nazwy własne	5
3	Podstawowa anotacja semantyczna	5
3.1	<i>Słowsiec</i>	5
3.2	Właściwa procedura	9
3.3	Gerundia i imiesłowy przymiotnikowe	9
4	Nazwy własne, skrótowce i wtrącenia obcojęzyczne	10
4.1	Skróty	11
5	Dwa poziomy znakowania	11
5.1	Przymiotniki i przysłówki stopniowane nieregularnie	12
5.2	Pluralia tantum	12
6	Lematy wieloczłonowe	13
6.1	Współwystępowanie czasowników, gerundiów i imiesłowów z cząstką <i>się</i>	13
6.2	Standardowe leматы wieloczłonowe	14
6.2.1	Przysłówki zaczynające się od przyimka	15
6.2.2	Cząstka <i>sobie</i> w lemacie czasownika	16
6.3	Segmentacja	16
6.4	Wieloczłonowe nazwy własne i zwroty obcojęzyczne	17
6.4.1	Burkinostki	17
6.5	Tytuły	18
7	Segmenty niepodlegające znakowaniu	18
7.1	Rzeczowniki w wyrażeniach przyimkowych	18
7.2	Fragmenty retoryczne	19
8	Segmenty specjalnej troski	19
8.1	Zaimki osobowe	19
8.2	Zaimki pytajno-zależne i względne	20
8.3	Znakowanie formy <i>to</i>	20
8.4	Znaczenia czasownika <i>być</i>	22
8.5	Znaczenia czasownika <i>mieć</i>	24

1 Informacje wstępne

Zadanie polega na przypisaniu wydzielonym już segmentom rzeczownikowym, czasownikowym, przymiotnikowym i przysłówkowym w tekście identyfikatorów (numerów) znaczeń pochodzących ze *Słownosieci* (plWordnet). Znakowanie odbywa się przy użyciu programu **WebAnno** za pośrednictwem przeglądarki internetowej. Tak więc, w przeciwieństwie do etapu, w którym znakowanie odbywało się za pomocą arkusza kalkulacyjnego Excel, część pracy program wykona za nas.

Znakowane próbki zostaną podzielone na trzy zestawy (zwane projektami), por. *Praktyczna instrukcja znakowania*¹ sekcja 1, w zależności od pochodzenia danych i stopnia zaawansowania znakowania:

1. PCC-metaph: próbki z korpusu koreferencyjnego z gotowym poziomem anotacji semantycznej (ok. 350 próbek), w których znakujemy wyłącznie wyrażenia metaforyczne;
2. PCC-sem+metaph: próbki z korpusu koreferencyjnego, w którym trzeba będzie wykonać znakowanie na poziomie semantycznym i metafor (ok. 350 próbek);
3. Składnica: próbki z fragmentu milionowego NKJP, dla których część zdań jest oznakowana na poziomie semantycznym (ok. 2000 próbek).

Program rozdziela znakowanie na warstwy. Będziemy tu mieli następujące warstwy:

- I. warstwę tekstową (warstwę form wyrazowych),
- II. warstwę lematów,
- III. warstwę części mowy,
- IV. warstwę korekty poziomu morfoskładniowego (Correction),
- V. warstwę znakowania semantycznego (Lex_Semantics oraz Lex_Part),
- VI. warstwę znakowania metaforycznego (Metaphor oraz Met_Part).

Pierwsze trzy warstwy pochodzą z NKJP i nie mogą być zmieniane, dlatego nie widać ich nazw. Czwarta warstwa pojawiać się będzie jedynie dla dwóch pierwszych zestawów, gdyż PCC jest podkorpusem pełnego NKJP znakowanym automatycznie na poziomie morfoskładniowym, czyli jest to znakowanie obarczone błędami. Dla PCC-metaph korekta już została dokonana i nie może być zmieniana, podobnie jak warstwa anotacji semantycznej.

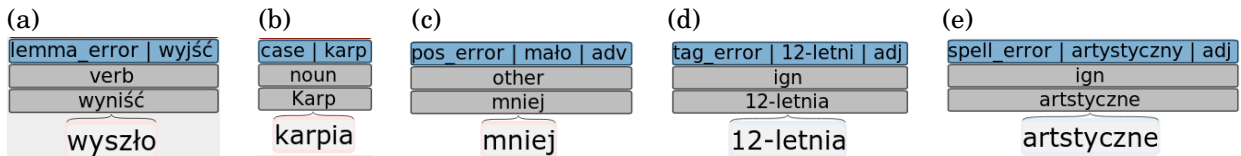
Niniejsza instrukcja dotyczy zasad tworzenia warstwy IV i V, przede wszystkim pod kątem znakowania zestawu PCC-sem+metaph, ale też ponieważ i zestawu Składnica (tylko warstwa V).

Warstwy V i VI składają się z dwóch faktycznych warstw każda. Jednak warstwy identyfikowane jako Lex_Part i Met_Part służą jedynie do podczepiania elementów wyrażen wielocłonowych i jako takie nie mają samodzielnego charakteru, por. *Praktyczna instrukcja znakowania* sekcja 4.2.

Znakowaniu semantycznemu podlegają segmenty należące do następujących części mowy:

- przymiotniki adj,
- przysłówki adv,
- rzeczowniki noun,
- czasowniki verb,
- imiesłowy (ppas i pact) oraz odsłowniki ger,
- burkinostki burk,

¹Dokładniej: *Praktyczna instrukcja znakowania poziomu leksykalno-semantycznego i metaforycznego za pośrednictwem programu WebAnno*



Rysunek 1: Korekta błędów różnego typu – widok w programie WebAnno

- skróty brev (pod warunkiem, że są to skróty wyrazów reprezentujących wyliczone powyżej części mowy).

Dla uproszczenia wszystkie pozostałe części mowy mają oznaczenie *other* (poza znakami przestankowymi *interp* oraz wyrazami nieznanymi *ign* – te ostatnie zazwyczaj wymagają korekty). Oczywiście takie segmenty, jak wszystkie, mogą być składnikami wyrażen wielocłonowych.

2 Korekta znakowania morfoskładniowego

Podczas korekty znakowania morfoskładniowego powinna być aktywna warstwa IV *Correction*, por. *Praktyczna instrukcja znakowania* sekcja 2.

Korekta znakowania morfoskładniowego nie ma na celu poprawy znakowania morfoskładniowego (choć może się do niego przyczynić), ma jedynie umożliwić poprawne znakowanie leksykalno-semantyczne, a na jego podstawie metaforyczne. Dlatego korekcie podlegają jedynie segmenty, które zamierzamy znakować, oraz segmenty, którym niepoprawnie przypisano część mowy wymagającą znakowania. Tak więc na początku sprawdzamy, czy lemat zapisany w warstwie II i część mowy zapisana w warstwie III są wpisane poprawnie oraz czy forma widniejąca w warstwie I nie zawiera literówek.

Jeśli interpretacja morfoskładniowa segmentu nie jest poprawna, ustalamy wartość znacznika *code* jako

- *lemma_error*, gdy wystąpił błąd lematyzacji, np. *koleja* zamiast *kolej*, *musić* zamiast *musieć*;
- *case* – to przypadek błędnego lematu ograniczony do niepoprawnej kasztowości (typowo dla nazw własnych homonimicznych z wyrazem pospolitym, np. *Mars* – planeta, bóg i *mars* – mina, ale też *Bóg* i *bóg*);
- *pos_error*, gdy przypisana została błędna część mowy (POS), nawet jeśli lemat jest poprawny, np. *piec*, *wieść*, *to* – *to* może być czasownikiem, rzeczownikiem lub partykułą i często jest błędnie znakowane,
- *tag_error*, gdy błąd pisowni/lematyzacji pociągnął za sobą przypisanie błędnej części mowy, np. forma *lata* uznana została formę czasownika *latać* a nie rzeczownika *rok*. W szczególności dotyczy to sytuacji, gdy błąd spowodował przypisanie segmentowi kategorii *ign* (kategoria nieznaną);
- *spell_error*, gdy wystąpiła literówka, np. *tluczenia* (zlematyzowane do *tluc*) zamiast *tlucznia* o lemacie rzeczownikowym *tluczeń*;
- *hyphen* – wyraz wypisywany z myślnikiem, np. *n-i-e ż-y-j-e n-a-p-r-a-w-d-ę*;
- *no_annot* – uproszczona wersja *pos_error* oznaczająca, że wyrazowi błędnie przypisana została część mowy wymagająca lematyzacji, np. *co* we frazie *co roku* jest przyimkiem).

Adekwatnie do typu błędu, należy wpisać skorygowany lemat w pole *lemat* i/lub ustawić część mowy w polu *POS*. Widok segmentów po korekcie błędów różnego typu przedstawiony został na rys. 1. W wypadku segmentów nie wymagających faktycznej anotacji korekta ogranicza się do wpisania kodu *no_annot*.

UWAGA: Nie wprowadzamy korekty błędnej postaci segmentu.

Korekta lematu i części mowy jest bardzo ważna, bo numer znaczenia wybrany podczas właściwego znakowania semantycznego jest przypisywany do pary ⟨lemat, część mowy⟩.

Interesuje nas poprawna, docelowa część mowy. Tak więc korekty części mowy niepodlegającej znakowaniu dokonujemy tylko wówczas, gdy wyraz powinien zostać oznakowany, i w wyniku korekty ustawiamy jedną z części mowy podlegającą znakowaniu. Postępowanie w takim wypadku jest analogiczne do innych powyższych przypadków.

2.1 Skróty i skrótowce

Podobnie jak inne dla innych części mowy, także uznanie segmentu za skrót (brev) może być błędne lub odwrotnie. Na przykład napisy *w im, proc* mogą być skrótami lub formami wyrazów pospolitych. Także skrót oznaczony jako ign wymaga właściwej korekty.

UWAGA: To, że kropka nie jest elementem skrótu (*lek, fot* zamiast *lek., fot.*) jest regułą segmentacji, a nie błędem.

Procedura postępowania jest taka sama, jak opisywana powyżej. Tak więc:

1. pole code wypełniamy standardowo, a do pola POS wpisujemy brev (jeśli cz. mowy była błędnie przypisana),
2. większość skrótów jest nieodmienna, jednak nierzadko musimy zmodyfikować pole lemma (por. rys. obok).
3. W szczególności, jeżeli dany segment jest napisem rozpoczynającym się od liczby zapisanej cyfrowo (np. *12-letni*) lub innym skrótowym zapisem np. (*W-wa* jako *Warszawa*), por. zdanie (1), to:

tag	error	ulica	brev	
	noun			interp
		Ula		.
		ul		.

- wypełniamy pola code oraz POS standardowo, to ostatnie w zależności od docelowej części mowy,
- pole lemma wypełniamy tylko wtedy, gdy w warstwie II brak poprawnego lematu (dla klasy ign lemat jest kopią formy wyrazowej, czyli np. dla *PR-owca* trzeba wpisać *PR-owiec* itp.).

Patrz też rys. 1 (d).

- (1) *W domu została najmłodsza, **12-letnia**, ale na święta zjadą wszyscy. **11-ty** Charytatywny Bal Dziennikarzy to impreza, na której spotykają się nie tylko żurnaliści. W samej **W-wie** strajkuje trzysta trzydzieści zakładów. Pili coca-colę i jedli lody.*
- (2) *Adam naprawdę nie żyje, nie żyje naprawdę, **n-i-e-ż-y-j-e-n-a-p-r-a-w-d-ę**.*

Podobnie postępujemy, gdy pisownia wyrazu rozbita jest na części, por. (2), tyle że pole code ustawiamy na hyphen. W tym wypadku korekta lematu jest rzecz jasna nieodzowna.

UWAGA: Wśród segmentów, których część mowy jest nieznaną (ign) znajdują się także liczby zapisywane jako ciągi cyfr. W żadnym razie nie należy ich korygować ani znakować, nawet jeśli w *Słowsieci* występuje odpowiedni przymiotnik (*trzeci, siedemnasty*) bądź rzeczownik (*trójka, siedemnastka*). Natomiast cyfrowe oznaczenia lematów występują jedynie w nieadekwatnej dla nas części angielskiej.

Nie należy mylić skrótów fraz ze skrótowcami. Skrótowce są wyłącznie skrótowym zapisem nazw własnych, np. *MOSiR – Miejski Ośrodek Sportu i Rekreacji, WOPR – Wodne Ochotnicze Pogotowie Ratunkowe, PTTK, SA – spółka akcyjna* itp. Dla skrótowców właściwą częścią mowy jest noun. Niektóre z nich także są odmienne (np. *PTTK-iem, PTTK-u*), co może prowadzić do konieczności korekty lematu.

2.2 Gerundia i imiesłowy

Istnieje wiele rzeczowników homonimicznych z gerundiami (*podanie, obejście*) oraz przymiotników homonimicznych z imiesłowami (*otwarty, zmęczony, bieżący, interesujący*). Jako że tagowanie takich jednostek różni się wyłącznie kategorią gramatyczną i lematem, błędy takie będą stosunkowo częste. W takim wypadku:

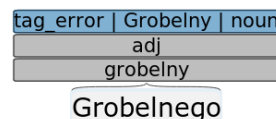
1. ustalamy wartość znacznika code jako `tag_error`;
2. dokonujemy zmiany lematu na lemat odpowiedniego czasownika, rzeczownika lub przymiotnika,
3. dokonujemy adekwatnej zmiany części mowy na `ger`, `pact`, `ppas`, `noun` lub `adj`.

2.3 Nazwy własne

Większość nazw własnych jest nieznaną tagerowi, w związku z czym nie jest interpretowana poprawnie. Jeżeli w warstwie II segmentowi przypisana została klasa `ign` lub inna błędna błędna interpretacja, korygujemy ją zgodnie z regułami opisanymi powyżej.

Dotyczy to w szczególności nazwisk homonimicznych z przymiotnikami, np. gdy nazwiska *Grobelny*, *Mazowiecki* zostały uznane za przymiotniki *grobelny*, *mazowiecki*. Jeśli nazwisko jest znane analizatorowi morfologicznemu, możliwy jest także błąd odwrotny, który rzecz jasna także wymaga korekty.

UWAGA: Nazwiska są zawsze rzeczownikami! (por. rys. obok).



3 Podstawowa anotacja semantyczna

Podczas korekty znakowania morfoskładniowego powinna być aktywna warstwa `V_Lex_Semantics`, por. *Praktyczna instrukcja znakowania* sekcja 2.

Jak napisano we wstępie, anotacji podlegają wszystkie segmenty o kategoriach składniowych odpowiadających rzeczownikom, czasownikom, przymiotnikom oraz przysłówkom, oznaczonym odpowiednio przez `noun`, `verb` (odsłowniki `ger` oraz imiesłowy `ppas` i `pact` to z punktu widzenia anotacji czasowniki, por. sekcja 3.3), `adj` i `adv`. Znaczenia przypisywane są parze \langle lemat, część mowy \rangle uwzględniającej korektę wprowadzoną w warstwie IV (por. punkt 2.2). Jak już pisaliśmy, należy zwracać baczną uwagę na lemat i część mowy, zarówno podlegający znakowaniu, jak i reprezentowany w *Słowsieci*.

Znakowanie segmentów o oznaczeniach innych niż wymienione powyżej omówione zostanie w odrębnych podpunktach.

3.1 Słowsiec

Słowsiec jest jednym z dwóch polskich wordnetów. Wordnet jest przykładem sieci semantycznej zbudowanej wokół idei *pojęcia* reprezentowanego przez zbiór synonimów; stąd powszechnie używany termin *synset*. Dlatego wordnet bywa niejednokrotnie postrzegany jako *leksykalna sieć semantyczna*.

W wypadku *Słowsieci* równie ważnym pojęciem co *synset* jest *jednostka leksykalna*, reprezentowana w postaci pary \langle lemat, numer znaczenia \rangle . Pojęcie to poza wyrazami obejmuje także zleksykalizowane jednostki wielocłonowe.

Słowsiec obejmuje cztery klasy gramatyczne: rzeczowniki, czasowniki, przymiotniki i przysłówki. W pracy korzystamy z wersji *Słowsieci* dostępnej pod adresem: <http://plwordnet4.clarin-pl.eu/>. Sposób wizualizacji danych w *Słowsieci* zorganizowany jest wokół pojęcia jednostki leksykalnej. Na rys. 2 zaprezentowany został standardowy widok słowsieciowy.

Słowność 4.0

← Poprzednie znaczenie Oglądasz 2 z 2 dostępnych znaczeń tego słowa

Hiperonimy

sprawa 1

materia, zespół określonych okoliczności, waż...

← Kolokacyjność

przedyskutować 1

domena: czasowniki mówienia, śpiewania itp.

Synonimy międzyjęzykowe

problem 2

a question raised for consideration or solution

Hiponimy międzyjęzykowe

gut issue 1

an issue that elicits strong emotional reactions

paramount issue 1

an issue whose settlement is more important ...

pocketbook issue 1

an issue whose settlement will affect financial...

quodlibet 1

an issue that is presented for formal disputati...

... Pokaż pozostałe 13 połączeń

Międzyjęzykowe synonimy częściowe

issue 1

an important question that is in dispute and m...

RZECZOWNIK **problem 2**

zagadnienie, zadanie do rozwiązania, kwestia do przemyślenia, podjęcia decyzji

DOMENA związane z myśleniem

PRZYKŁADY To bardzo frapujący problem, spróbujmy się razem zastanowić nad rozwiązaniem.

SYNONIMY

zagadnienie 1
problem, kwestia wymagająca rozstrzygnięcia, rozwiązania, przemyślenia

zagwozdka 1
problem, kwestia, zagadnienie, które trzeba przemyśleć i rozwiązać

kwestia 2
sprawa do wyjaśnienia; zagadnienie, problem

punkt 3
problem, kwestia, przedmiot rozważań

pytanie 4
kwestia, zagadnienie, problem do przemyślenia, rozwiązania

ŹRÓDŁO Słowność

HIPERONIMY [Pokaż ścieżkę do najwyższego hiperonimu](#)

GRAPH [Wyświetl wizualizację graficzną](#)

Hiponimy

koncepcja 2

pojęcie, zagadnienie, idea

problem decyzyjny 2

w teorii obliczeń: pytanie sformułowane w sys...

twardy orzech do zgryzienia 1

skomplikowana sprawa, problem wymagając...

moment Minsky'ego 1

punkt, w którym w wyniku spekulacji załamuj...

... Pokaż pozostałe 17 połączeń

Bliskoznaczne z

trudny orzech do zgryzienia 1

niepoprawna, ale często używana forma zwią...

problemat 1

domena: związane z porozumiewaniem się

Kauzacja

zabijać ćwieka 1

stawiać przed kimś problem, zagwozdkę, tru...

zabić ćwieka 1

postawić przed kimś problem, zagwozdkę, tr...

Derywaty

problematyzacja 1

ujęcie czegoś, opis, omówienie jako problemu...

problemowo 2

w kontekście problemu lub problemowym, w ...

bezproblemowy 4

o treści jakiegoś przekazu, najczęściej estetyc...

Kolokacyjność

czterozakresowy 2

taki, który mieści się w czterech zakresach, w ...

banalność 2

domena: cechy ludzi i zwierząt

wielopiętrowy 2

przenośnie: zawiły, skomplikowany, złożony

nierozwiązywalny 1

domena: przymiotniki jakościowe

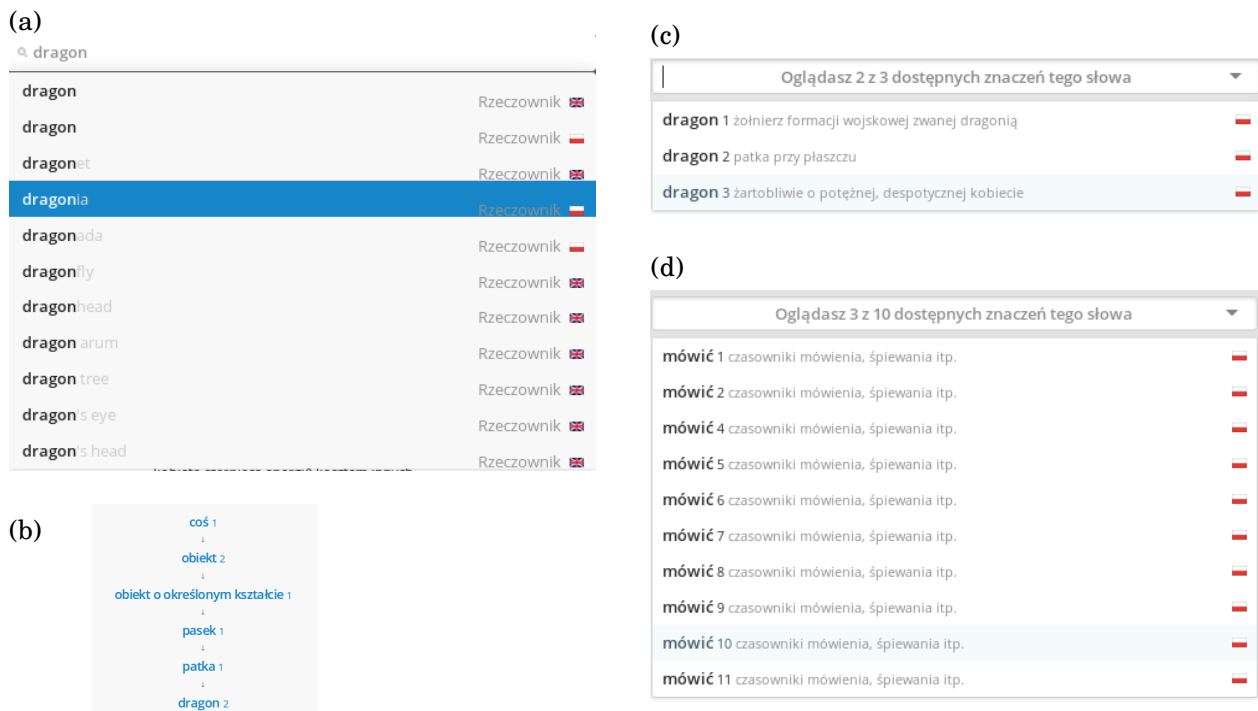
... Pokaż pozostałe 5 połączeń

Holonimy (części)

problematyka 1

domena: grupy ludzi i rzeczy

Rysunek 2: Słowność – ogólny widok jednostki



Rysunek 3: *Słowosiec* – zrzuty ekranów okien pomocniczych

Na początku widok ten jest pusty. Po wpisaniu (początku) lematu pojawia się lista adekwatnych lematów, por. rys. 3(a). Po lewej stronie tabeli widnieją lematy zaczynające się od wybranego napisu, po prawej stronie część mowy oraz flaga wskazująca na język, w jakim lemat ten jest interpretowany². Należy wybrać lemat adekwatny pod wszystkimi tymi kątami. *Słowosiec* nie udostępnia żadnych filtrów, w szczególności nie można ograniczyć się do jednostek o lematach zdanego języka czy danej części mowy.

Górna linia służy do nawigacji i zawiera przyciski Poprzednie znaczenie po lewej oraz Następne znaczenie po prawej. Po środku znajduje się informacja, które z kolei znaczenie oglądamy.

Nie jest to numer znaczenia z punktu widzenia znakowania. Jeśli liczba jednostek jest duża, nie musimy przeglądać ich wszystkich. Kliknięcie na ▼ po lewej stronie pola wyboru powoduje wyświetlenie listy jednostek z podstawowym opisem ich znaczenia (czasem ograniczonego do kategorii semantycznej), por. rys. 3(c,d). Często (choć, jak widać, nie zawsze) pozwala to na wstępny wybór jednostek do dokładniejszej analizy. Zwróćmy uwagę na listę z rys. 3(d). Chociaż wyraz zawiera 10 znaczeń, numery sensów nie są ciągłe, gdyż nie zawierają numeru 3. Dlatego trzecie znaczenie ma numer 4, czwarte znaczenie – numer 5, itd., por. też rys. 3(e). To te numery widoczne przy lematach są istotne i te należy wpisywać w polach nr i nr_close.

W środkowej kolumnie widnieją podstawowe informacje o prezentowanej jednostce leksykalnej. W pierwszym wierszu znajduje się jej część mowy, lemat, numer znaczenia oraz flaga wskazująca na właściwy (bądź nie!) podsłownik. Za każdym razem należy upewnić się, że jest są to dane właściwe (wystarczy przy pierwszej jednostce). Poniżej znajduje się opis znaczenia (bywa pusty).

Każdej jednostce w *Słowosieci* przypisana jest kategoria semantyczna, charakteryzująca w sposób dość ogólny jej znaczenie. Widnieje ona w kolejnym wierszu (oznaczana jest jako DOMENA). Kategoryzacja ta nie do końca jest konsekwentna, mimo to jednak stanowi ważną wskazówkę podczas wyboru jednostki adekwatnie interpretującej wyraz w danym kontekście.

Zarówno jednostki leksykalne, jak i synsety są wiązane przez liczne relacje. Synonimia wyrażana jest przez fakt przynależności do jednego synsetu. Jednostki synonimiczne z prezen-

²Lista ta nie jest skrolowalna i nie musi zawierać wszystkich lematów zaczynających się od wpisanego w oknie napisu.

← Poprzednie znaczenie Oglądasz 3 z 10 dostępnych znaczeń tego słowa Następane znaczenie →

<p>Hiperonimy</p> <p>kontaktować się 1 ■ wchodzić z kimś, czymś w kontakt polegają...</p> <p>← Relacja Initiator (Relacja Walentego)</p> <p>podmiot 3 ■ osoba lub instytucja w terminologii prawni...</p> <p>człowiek 1 ■ jednostka ludzka, osoba</p>	<p>CZASOWNIK mówić 4 ■ kontaktować się z kimś, wymieniać poglądy</p> <p>DOMENA czasowniki mówienia, śpiewania itp.</p> <p>SYNONIMY rozmawiać 1 dialogować 1</p> <p>ŹRÓDŁO Słowność</p> <p>HIPERONIMY Pokaż ścieżkę do najwyższego hiperonimu</p> <p>GRAPH Wyświetl wizualizację graficzną</p>	<p>Hiponimy</p> <p>oftopować 1 ■ oddawać się rozmowom na temat inny niż ...</p> <p>konwersować 5 ■ rozmawiać w języku obcym, najczęściej z os...</p> <p>dziamać 3 ■ głośno, długo i uporczywie mówić</p> <p>gadu-gadu 1 ■ luźno rozmawiać</p> <p>*** Pokaż pozostałe 11 połączeń</p> <p>Bliskoznaczne z</p> <p>gaworzyć 3 ■ domena: czasowniki mówienia, śpiewania i...</p>
---	---	---

Rysunek 4: *Słowność* – ogólny widok jednostki czasownikowej

towaną wyświetlane są poniżej kategorii gramatycznej jako para (lemat, numer znaczenia) – część mowy i język wszystkich elementów synsetu jest taki sam.

Podstawowymi relacjami pomiędzy synsetami są wzajemnie odwrotne hiponimia i hiperonimia oraz holonimia i meronimia. Relacjami pomiędzy jednostkami są m.in. antonimia, konwersja i derywacyjność. Jak widać na rys. 2, po prawej i lewej stronie wyświetlane są nazwy relacji, a pod nimi lista jednostek powiązanych daną relacją z jednostką prezentowaną. U góry po lewej zawsze jest lista bezpośrednich hiperonimów, zaś po prawej – lista bezpośrednich hiponimów (o ile nie są puste). Poniżej wyświetlane są niepuste listy innych relacji. Jeśli lista taka jest długa, wyświetlany jest jej początek. Kliknięcie na Pokaż pozostałe x połączeń powoduje rozwinięcie takiej listy. Natomiast kliknięcie na lemat dowolnej jednostki powoduje przejście do tej jednostki i prezentację danych z jej punktu widzenia.

Najechnie myszą na nazwę relacji powoduje pojawienie się „dymka” z jej opisem.

UWAGA: Część relacji ma charakter międzyjęzykowy. Należy je ignorować, chociaż w akcie desperacji przy niezrozumiałej jednostce można próbować dowiedzieć się czegoś poprzez tłumaczenie.

Pojedyncze znaczenie wyrazu, identyfikowane przez daną jednostkę leksykalną, determinowane jest przez relacje łączące ją z innymi jednostkami. Wszelkie dodatkowe informacje, np. definicje czy glosy, mają jedynie znaczenie uzupełniające. Podstawową rolę pełni tu hierarchia hiperonimii. Możemy ją sobie wyświetlić klikając na Pokaż ścieżkę do najwyższego hiperonimu. Na rys. 3(b) widnieje taka lista dla jednostki *dragon-2* ‘patka przy płaszczu’. Jednak czasami synset nie posiada hiperonimów. Na przykład formy żeńskie wyrazów nie są włączane do tej relacji, tylko wiązane przez relację *żeńskości* (Nazwy żeńskie) z formami męskimi, od których są derywowane. Na przykład dwie jednostki o lemacie *chopinistka* są wiązane z odpowiadającymi im jednostkami o lemacie *chopinista*, z których jeden jest hiponimem *pianisty*, a drugi *muzykologa*. Podobnie *wstążeczka-1* jest połączona relacją *Nacechowanie* (deminutywy) z jednostką *wstążka-1*. W takim wypadku niejednokrotnie dopiero opis wyrazu pierwotnego pozwala ustalić właściwe znaczenie wyrazu pochodnego.

Za każdym razem należy upewnić się, że wybór jednostki jest właściwy, nawet wówczas, gdy mamy do dyspozycji tylko jedną jednostkę. Na przykład *Słowność* 1.8 zawierała jedno znaczenie wyrazu *koncert* będące hiponimem terminu *przedstawienie*, brakowało utworu muzycznego adekwatnego dla zdania *Gramy Beethovena, dziś nawet idą tradycyjnie koncerty fortepianowe*.

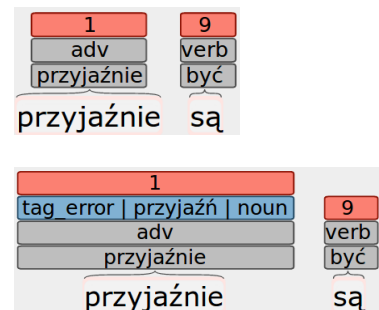
Tak więc w wielu wypadkach w celu jednoznacznego zinterpretowania jednostki konieczne jest posłużenie się pełnym zestawem dostępnych relacji.

3.2 Właściwa procedura

Podstawowa procedura znakowania segmentu przebiega następująco:

1. Jeśli dane wejściowe są poprawne, wyszukujemy w *Słowski* jednostki leksykalne o lemacie znajdującym się w warstwie II lub IV (jeśli nie dokonano korekty, warstwy tej nie ma). Należy przy tym zwrócić uwagę na:
 - zgodność część mowy wpisanej przy formie z częścią mowy z warstwy III (lub IV) ze względu na homonimie lematów (np. *piec*, *wieść*),
 - język oznaczony za pomocą flagi, ze względu na możliwość homonimii wyrazów polskich i angielskich (np. *format*, *plus*, *ton*).
 - Czasem istotna może być też kaszta, jaką pisany jest dany wyraz (np. *Bóg*).
2. Należy przejrzeć wszystkie jednostki o danym lemacie i wybrać te, których opis (głosa, przykłady użycia) oraz relacje (synonimy, hiperonimy, hiponimy i in.) wskazują, że dana jednostka jest adekwatna w znakowanym wypowiedzeniu (tekście).
3. Jeżeli znajdziemy dokładnie jedną taką jednostkę, wpisujemy jej numer w polu nr. Jeżeli jednak jest ich więcej, przeglądamy je powtórnie i:
 - Wybieramy spośród nich najbardziej adekwatną, i jej numer wpisujemy w polu nr.
 - Jeśli z jakichś względów nie jesteśmy w stanie podjąć takiej decyzji, w polu nr wpisujemy jednostkę o najniższym numerze. Numery pozostałych znaczeń, na tyle zbliżonych, że więcej niż jedno okazało się adekwatne w kontekście, wpisujemy w polu nr_close rozdzielając je przecinkami (bez spacji). Nie należy jednak nadużywać tej możliwości.
4. Jeżeli w *Słowski* nie znajdziemy żadnej jednostki adekwatnej dla znakowanego segmentu pospolitego, i nie zachodzi żaden przypadek szczególny opisany w kolejnych sekcjach, w polu a_lack wpisujemy lack.

UWAGA: Proszę o zwracanie uwagi na lemat i część mowy występujące w warstwach II i III. Numer znaczenia wpisany do pola nr jest interpretowany dla tej właśnie pary. Jeśli więc segment *przyjaźnie* ze zdania (3) mamy oznakowany jak na rys. obok u góry, to wpisanie adekwatnego numeru znaczenia rzeczownika *przyjaźń* nie będzie poprawnym znakowaniem, gdyż *de facto* wpisane zostało znaczenie przysłówka *przyjaźnie*. Poprawna interpretacja wymaga wprowadzenia korekty (warstwa IV), por. sekcja 2 s. 3, por. rys. obok u dołu.



(3) [...] *jeśli tak jest, to prawdopodobnie dlatego przyjaźnie są raczej monoplciowe.*

UWAGA: Ustawianie oznaczenia lack nie dotyczy wypowiedzeń metaforycznych, w których widać wyraźny związek z pierwotnym znaczeniem wyrazu. W takim wypadku interpretujemy wszystkie jego składowe w sposób dosłowny.

3.3 Gerundia i imiesłowy przymiotnikowe

W NKJP gerundia (odsłowniki, ger) oraz imiesłowy przymiotnikowe (czynne pact i bierne ppas) lematyzowane są do formy bezokolicznika. Jednak w *Słowski* umieszczono bezpośrednio dużo gerundiów i nieco mniej imiesłowów.

Procedura postępowania jest w ich wypadku bardzo podobna do omówionej powyżej, czyli znajdujemy znaczenie bezokolicznika zgodnie z procedurą opisaną w punkcie 3. Ich wyróżnienie wiąże się z faktem, że podczas dalszego przetwarzania będziemy chcieli wykorzystać

powiązanie w *Słownosieci* jednostek czasownikowych z odpowiadającymi im gerundiami i imiesłowami relacją Synonimii międzyparadygmatycznej.

UWAGA: Nie znakujemy gerundiów jako rzeczowników ani imiesłówów jako przymiotników czy przysłówków. Jeśli taki lemat został ustawiony, wymagana jest korekta morfologiczna.

UWAGA: Szczególnej uwagi wymagają gerundia i imiesłowy homonimiczne z rzeczownikami czy przymiotnikami. Jako rzeczowniki/przymiotniki znakujemy je tylko wtedy, gdy brak wspomnianych powiązań z czasownikiem. W szczególności, rzeczowniki i przymiotniki **nigdy** NIE współwystępują z częstką *się*, por. punkt 6.1. W przeciwnym razie oznakowanie wyrazu jako rzeczownika czy przymiotnika traktujemy jako błąd wymagający korekty, por. sekcja 2.2.

4 Nazwy własne, skrótowce i wtrącenia obcojęzyczne

Jeżeli segment stanowi (poprawnie zinterpretowany morfoskładniowo – dobry lemat i część mowy) składnik nazwy własnej, w polu `a_opt` wpisujemy name niezależnie od tego, czy nazwa ta posiada w *Słownosieci* adekwatne znaczenie czy też nie. Jedyny wyjątek stanowią tu człony wieloczłonowych nazw własnych występujących w *Słownosieci* (nie będące ich centrum); reguły znakowania takich nazw są opisane w sekcji 6.4 s. 17.

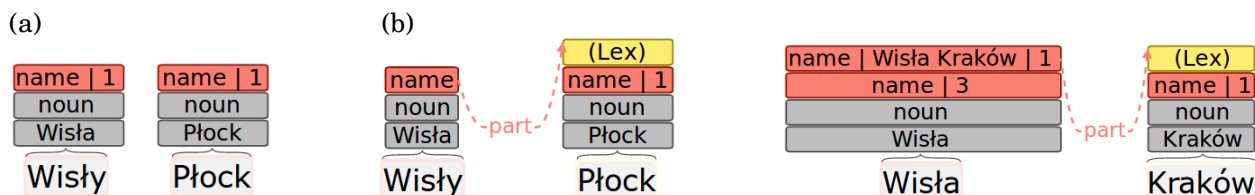
Słownosiec zawiera pewną liczbę nazw własnych, przede wszystkim nazwy geograficzne. W ich wypadku procedura postępowania jest następująca:

- Jeżeli znakowany segment jest nazwą własną (*Płock*, *bochnianin*) lub skrótowcem (*PKS*, *PKB*), to:
 1. W polu `a_opt` wybieramy kod name.
 2. Jeśli nazwa ta została uwzględniona w *Słownosieci*, właściwy numer znaczenia wpisujemy w polu `nr`. Adekwatność nazw należy sprawdzać tak samo jak wypadku wyrazów pospolitych. Na przykład *Mars* to może być planeta, rzymski bóg wojny bądź batonik, jednak w *Słownosieci* uwzględniony jest tylko pierwszy z nich (nie licząc dwóch wyrazów pospolitych, w *Słownosieci* podczas wyświetlania nie jest brana pod uwagę kaszta lematu).
- Jeżeli znakowany segment jest przymiotnikiem pochodzącym od nazwy (*płocki*) postępujemy tak samo, tyle że w polu `a_opt` wpisujemy `name:adj`.

UWAGA: Nie należy wpisywać znaczenia wyrazu pospolitego przy nazwach własnych, np. przy nazwiskach (*Belka*, *Karp*), czy ulicach (*Szeroka*). Nie należy też wpisywać interpretacji nieadekwatnych w kontekście, niezależnie od tego, czy to wyrażenie wieloczłonowe. Na przykład *Wisła Kraków*, *Wisła Płock* czy *Odra Wodzisław* nie są rzekami, tylko klubami sportowymi. Wpisywanie jednostki *Wisła-1* czy *Odra-1* mija się z celem, nawet gdyby wystąpiła sama (*Wisła wygrała...*).

Tak więc znakowanie z rys. 5 po lewej jest niepoprawne, poprawne znakowanie *Wisły Płock* widnieje pośrodku. Przypadkiem klub *Wisła Kraków* (ze skrótową wersją *Wisła*) jest w *Słownosieci* uwzględniona, co prowadzi do znakowania jak na rys. 5 po prawej. To, że *Płock*, *Kraków* nie stanowiące centrum frazy są znakowane jako miasta, jest poprawne, gdyż nie wpływa na interpretację całej frazy.

UWAGA: Nie należy mylić skrótowców ze skrótami fraz. Skrótowce są wyłącznie skrótowym zapisem nazw własnych, np. *MOSiR* – *Miejski Ośrodek Sportu i Rekreacji*, *WOPR* – *Wodne Ochotnicze Pogotowie Ratunkowe*, *PTTK*, *SA* – *spółka akcyjna* itp. Tak więc w wypadku skrótowców wpisujemy `name` w polu `a_opt`, i ewentualnie numer znaczenia w polu `nr`, bo niektóre skrótowce są w *Słownosieci*.



Rysunek 5: Niepoprawne i poprawne znakowanie nazw własnych – widok w programie WebAnno

Większość wtrąceń obcojęzycznych składa się z kilku słów, i ich znakowanie także omówione zostało w sekcji 6.4. Jeśli się jednak trafi pojedynczy wyraz obcy, por. 4, polu `a_opt` wybieramy kod `foreign`. Jeśli przypadkiem *Słownosieć* zawiera adekwatne znaczenie dla takiego lematu, jego numer wpisujemy w polu `nr`.

(4) *Przedstawiony świadek musiał być jednym z sąsiadów (conmarcanus) [...]*

4.1 Skróty

Jeśli rozważany segment jest skrótem (brev), w polu `a_suppl` wstawiamy, w zależności od docelowej części mowy: `brev:adj` (np. *w – wielki*), `brev:adv` (np. *b – bardzo*) lub `brev:noun` (np. *w – wiek*). Jeśli natomiast jest to skrót frazy (np. *br – bieżący rok*, *cdn – ciąg dalszy nastąpi*, *itp. – i tym podobne*), zamiast tego w polu `a_aseman` wybieramy `brev:phrase`, gdyż takie skróty nie są znakowane.

Ponadto (poza tym ostatnim przypadkiem):

1. Jeśli w warstwie II wpisane jest prawidłowe rozwinięcie skrótu (np. *dr – doktor*) bądź też skrót jako taki jest interpretowany w *Słownosieci* (np. *lek.*), korekty lematu dokonywać nie trzeba. W przeciwnym razie w polu `lemma2` wpisujemy pełny lemat, np. *fotografia* dla *fot.* Musi to być lemat adekwatny w kontekście, np. któreś z *wiek*, *wiersz*, *wieś*, *wyspa*, *wielki* dla *w*.

UWAGA: To, że kropka nie jest elementem skrótu (*lek*, *fot* zamiast *lek.*, *fot.*) jest regułą segmentacji, a nie błędem.

2. W szczególności, jeżeli dany segment jest napisem rozpoczynającym się od liczby zapisanej cyfrowo (np. *12-letni*) lub innym skrótowym zapisem np. (*W-wa* jako *Warszawa*), por. (1), to do pola `lemma2` wpisujemy adekwatny pełny lemat, chyba że jakimś cudem postać skrócona jest w *Słownosieci* (wówczas pole `lemma2` pozostawiamy puste).
3. Pole `nr` wypełniamy w sposób właściwy dla aktualnej wartości części mowy i lematu (z uwzględnieniem pola `lemma2`).

5 Dwa poziomy znakowania

Istnieją (na szczęście nieliczne) segmenty, dla których istnieje poprawne (o zbliżonym poziomie adekwatności znakowania) wymagające odmiennej lematyzacji, co uniemożliwia wykorzystanie mechanizmu znaczeń zbliżonych (`nr_close`). Na szczęście wykorzystywane przez nas narzędzie **WebAnno** umożliwia dwukrotne przypisanie znakowania do jednego segmentu.

Możliwość tę wykorzystujemy w sytuacji, gdy wyraz ma dwie możliwe interpretacje pojedynczego segmentu przy odmiennej lematyzacji

- przymiotnik czy przysłówek zarówno w stopniu równym, jak i wyższym/najwyższym (sekcja 5.1),
- rzeczownik w l.poj. i l.mn. (sekcja 5.2),

- jednostki wieloczłonowe (sekcje 6.2 i 6.4).

Z możliwości tej korzystamy jednak jedynie wówczas, gdy adekwatne dla danego segmentu są oba sposoby znakowania.

5.1 Przymiotniki i przysłówki stopniowane nieregularnie

Przymiotniki i przysłówki są lematyzowane do formy stopnia równego niezależnie od tego, czy stopniowanie odbywa się w sposób regularny czy też nie. Jednak w *Słownosieci* znajdują się również formy wyższe i najwyższe niektórych przymiotników i przysłówków (np. *gorszy*, *lepszy*, *najgorszy*, *najlepszy*, *gorzej*, *lepiej* ale już nie *najgorzej*, *najlepiej*).

Procedurę zaczynamy od doboru znaczenia dla lematu w stopniu równym, zgodnie z procedurą opisaną w punkcie 3. Następnie

1. Sprawdzamy, czy w *Słownosieci* istnieje jednostka o lemacie w stopniu właściwym dla formy z warstwy I powiązana z wybraną jednostką dla stopnia równego relacją Stopniowanie (stopień wyższy/najwyższy).
2. Jeżeli znajdziemy taką jednostkę, to, po rozpoczęciu dodatkowego znakowania (pole nr jest już jednak wypełnione), jeśli jednostka o lemacie w stopniu równym została wybrana) wpisujemy w polu *a_suppl* odpowiedni kod *comp/super*, w polu *lemma2* lemat w stopniu wyższym/najwyższym, zaś w polu *nr* wybrany numer znaczenia.
3. Jeśli odpowiedniej jednostki nie znajdziemy, przeglądamy pozostałe jednostki o takim lemacie i sprawdzamy, czy znaczenie któreś z nich wydaje się adekwatne. Jeśli uznamy, że znaczenie to pasuje, postępujemy tak jak powyżej.

UWAGA: Tylko dla przymiotników i przysłówków stopniowanych nieregularnie, lub gdy stopniowanie zmienia semantykę!

Taki sposób działania jest o tyle ważny, że zazwyczaj liczba znaczeń przypisana do lematów w stopniu równym jest znacznie większa niż tych w stopniu wyższym/najwyższym. Na przykład lemat *zły* ma przypisane 11 znaczeń, lemat *dobry* – 17, gdy tymczasem *gorszy*, *najgorszy* mają przypisane po jednym znaczeniu, zaś *lepszy*, *najlepszy* po dwa. Na przykład w zdaniu (5) mamy *dobry* w znaczeniu **dobry-6** i tak samo powinien zostać oznakowany segment *lepszy*.

(5) *To bardzo mocna drużyna, **dobra** w defensywie, ale jeszcze **lepsza** w akcjach ofensywnych.*

Jeśli jednak nie znajdziemy adekwatnego znaczenia ani dla lematu w stopniu równym, ani wyższym/najwyższym, w polu *a_lack* wybieramy kod *lack* bez analizowania, czy brakuje znaczenia w stopniu równym czy wyższym.

5.2 Pluralia tantum

W *Słownosieci* uwzględnione jest znacznie więcej wyrazów o lematach w l.mn. (*pluralia tantum*) niż w analizatorze morfologicznym *Morfeusz* opartym na *Słowniku Gramatycznym Języka Polskiego* (SGJP). Dlatego może się zdarzyć, że segment ma lemat w l.poj., gdy tymczasem adekwatna jednostka ma lemat w l.mn., np. *skrupuły*, *święta*, *względy*, *zabiegi*, por. (6).

(6) *Wrócili pełni **skrupułów** i wątpliwości.*

*Rodzina musi spędzać **święta** razem...*

*Ilu zabiega o moje **względy**?*

*[...]wicewojewoda zadeklarował, że wesprze ich **zabiegi** o zmianę przepisów.*

(7) *Mieliśmy też szereg pomniejszych **świąt** lokalnych.*

*Poddałem się różnym **zabiegom**, masażom, brałem kąpiele.*

*Przyjemne **względy** czyniły mnie graczem, partaczem...*

Jeśli znaczenie dostatecznie bliskie istnieje (np. święto-1), dokonujemy wstępnego znakowania dla tego lematu zgodnie z procedurą opisaną w punkcie 3, a następnie inicjujemy ponowne znakowanie segmentu (por. rys obok). W przeciwnym razie (np. *względy*) znakujemy jedynie lemat l.mn.³



W obu przypadkach:

1. W polu a_suppl wpisujemy kod wyjątku plurale.
2. W polu lemma2 wpisujemy poprawny lemat w l.mn.
3. W polu nr wstawiamy adekwatny numer znaczenia.

6 Lematy wieloczłonowe

W *Słowności* istnieją też lematy wieloczłonowe. Ich znakowanie wymaga przypisania znaczenia ciągowi segmentów. Zdarza się też, że segmenty składające się na takie wyrażenie posiadają odrębne znaczenie i w tenże sposób są znakowane.

Jeśli znakowany jest zarówno pojedynczy segment, jak i wyrażenie wieloczłonowe, w pierwszej wprowadzamy znakowanie proste, a następnie złożone, poprzez ponowne aktywowanie znakowania (por. *Praktyczna instrukcja znakowania* sekcja 4.2.). Znakowanie przypisujemy do segmentu stanowiącego centrum frazy, pozostałe jej składniki przyłączamy za pomocą funkcji *part*, zwaną w *WebAnno* rolą (por. tamże).

UWAGA: Drugi poziom znakowania wdramy jedynie wówczas, gdy centrum frazy posiada niezależną, własną interpretację (ma przypisany numer znaczenia w polu nr).

UWAGA: Znacznik a_lack ustawiamy na `lack:multi` jedynie wtedy, gdy w *Słowności* brak adekwatnego znaczenia zarówno dla całego wyrażenia, jak i dla przynajmniej jednego ze składających się nań segmentów. W takiej sytuacji decydujemy, której interpretacji brakuje. Jeśli uznajemy, że brak jest interpretacji wyrażenia wieloczłonowego jako całości, nie wymagamy interpretacji jego składowych, więc i nie przypisujemy im wartości `lack`. Znacznik `lack:multi` przypisujemy do centrum wyrażenia, a pozostałe człony podłączamy rolą *part* w sposób standardowy. Jeśli natomiast wszystkie człony mają swoje interpretacje, wyrażenie jest kompozycyjne semantycznie i nie wymagamy jego interpretacji jako całości.

Poniżej omówimy znakowanie lematów wieloczłonowych bardziej szczegółowo.

6.1 Współwystępowanie czasowników, gerundiów i imiesłowów z cząstką się

Najprostszym przypadkiem konstrukcji wieloczłonowych jest współwystępowanie czasowników (w różnych postaciach) z cząstką (partykułą, znacznikiem zwrotnym) *się*. W języku polskim współwystępowanie czasowników z tą cząstką tworzy konstrukcje zwrotne. Jednak sposób interpretacji tego znacznika może być różny. Wyróżniamy:

- *się* morfologiczne (*inherentne*) (*bać się, cieszyć się*), gdy wystąpienie cząstki *się* zmienia znaczenie czasownika;
- *się* zwrotne oznaczające czynność wykonywaną na sobie, wymienne na *siebie* (*myć się – myć siebie, akceptować się – akceptować siebie*), por. (8);
(8) *Możesz jednocześnie akceptować się i chcieć się doskonalić.*
- *się* wzajemnościowe (*recyprokalne*) oznaczające czynność wykonywaną na sobie nawzajem przez przynajmniej dwie istoty, por. (9);

- (9) **Łudziliśmy się nawzajem, że nic się nie zmieniło.**
*Gangsterzy **wybijają się** nawzajem.*

- *się bezosobowe* tworzące konstrukcje bezosobowe w czasie teraźniejszym (np. *mówi się, że*), por. (10).

- (10) *Białe kapelusze **czyści się** w zależności od gatunku.*

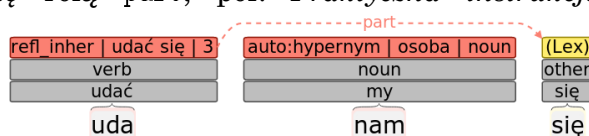
Zazwyczaj kontekst jednoznacznie wskazuje, z jakim rodzajem *się* mamy do czynienia. Może to być związane z doбором znaczenia czasownika, np. *charakteryzować się do występu* – *się zwrotne*, *charakteryzować się zrównoważeniem* – *się morfologiczne*, por. (11).

- (11) *A może jest to jakiś oszust-prestidigitator, który **charakteryzuje się** do każdej wizyty?*
*Podejście rządu do budżetu na rok 1992 **charakteryzowało się** większym realizmem i pragmatyzmem.*

Poszczególne rodzaje *się* traktowane są w różny sposób.

1. *się bezosobowe* jest zjawiskiem czysto składniowym, dlatego znakowaniu podlega „goły” lemat czasownika. Stosujemy procedurę z punktu 3 bez żadnych zmian.
2. W wypadku *się morfologicznego* jedyną sensowną interpretacją jest rozważanie lematów zawierających cząstkę *się*, i dla nich stosujemy procedurę opisaną w punkcie 3. W polu *a_suppl* wpisujemy *refl_inher*, zaś w polu *lemma2* – *<lemat_z_sie>*, np. *bać się*, *charakteryzować się*. Jeżeli adekwatnego znaczenia brakuje, w polu *a_lack* wpisujemy *lack*.
3. W wypadku *się zwrotnego* i *wzajemnościowego* podstawowym znaczeniem jest znaczenie niezwrotne, i dla niego stosujemy procedurę z punktu 3, dopisując w polu *a_suppl* odpowiednio *refl/ recip*. Jeżeli jednak w *Słowsieci* trafi się jednostka, której lemat zawiera *się*, połączona z wybraną jednostką relacją Znaczenia zwrotne/wzajemne⁴, należy zastosować znakowanie uzupełniające, wpisać w nim w polu *lemma2* *<lemat_z_sie>* (np. *atakować się*⁵), zaś w polu *nr* – *<numer znaczenia>*, np. **1**.

W sytuacji, gdy znakujemy segment jednostką o lemacie z *się*, należy połączyć lemat czasownika z właściwym wystąpieniem cząstki *się* rolą *part*, por. *Praktyczna instrukcja znakowania* sekcja 4.2. Sama partykuła *się* nie podlega znakowaniu. Segmenty te nie muszą ze sobą sąsiadować, por. rys. obok.



6.2 Standardowe lematy wieloczłonowe

W *Słowsieci* istnieją też typowe lematy wieloczłonowe. Wyrazowi *szkoła* z (12) może rzecz jasna przypisać znaczenie o numerze **2**. Jednak istnieją bardziej adekwatne jednostki o lematach *szkoła podstawowa*, *szkoła średnia*. Co gorsza, o ile istnieje jednostka (o numerze **1**) nadająca adekwatne znaczenie segmentowi *podstawowy*, to już dla segmentu *średni* takie znaczenie nie istnieje. Natomiast terminy *klatka piersiowa* czy *centrum handlowe* z (13a,b) nie posiadają pasującej interpretacji leksemu *klatka*, *centrum*. W takiej sytuacji znakujemy jednostką o lemacie wieloczłonowym nawet wtedy, gdy w wypowiedzeniu termin nie występuje w całości (13b).

- (12) *Zaczęłam uczyć w **szkole podstawowej**.*
*Ma zakończyć edukację na **szkole średniej**?*

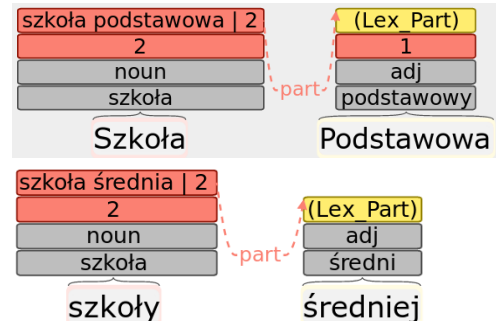
⁴Nie zawsze taka relacja jest wpisana!

⁵W *Słowsieci* jednostka *atakować się-1* nie jest powiązana z żadną jednostką *atakować*. Należy jednak domniemywać, że to *się wzajemnościowe*, bo użycie z *się inherentnym* dla tego czasownika nie istnieje.

- (13) a. *Ostatnio byłam w kinie w jednym z warszawskich **centrów handlowych**.*
 b. *Lubimy zaglądać do takich dużych **centrów**.*
 c. *Poznań stał się jednym z **centrów handlowych** Europy.*

Oczywiście zbitka taka może być przypadkowa. Na przykład w zdaniu (13c) nie chodzi rzecz jasna o centrum handlowe, tylko o ważny ośrodek handlu, dla którego właściwa jest jednostka centrum-3.

Jako że naszym zadaniem jest przede wszystkim znakowanie pojedynczych segmentów, za każdym razem sprawdzamy, czy istnieje jakieś znaczenie, które może stanowić interpretację tego wyrazu, np. szkoła-2 (w polu nr wpisujemy oczywiście tylko numer znaczenia, czyli 2). Jednak w wypadku frazy *klatka piersiowa* żadne z reprezentowanych znaczeń wyrazu *klatka* zastosować się nie daje. Nie znakujemy też wyrazów, dla których powiązanie znaczeniowe jest słabe, np. centrum-4,



czyli ‘duży ośrodek zajmujący się jakąś działalnością’ dla *centrum handlowego*. Podobnie nie znakujemy samego *chwytac* oraz *czy bić*, lecz znakujemy jedynie wyrażenie wieloczłonowe (*chwytac za serce-1, bić pokłony-2*)⁶.

- (14) *Historie z holocaustu wciąż **chwytają za serce**.*
 (15) *Gdzieś w polu stary Kurd **bije pokłony** w stronę Mekki.*

Ogólnie, w takiej sytuacji segment stanowiący centrum (semantyczne) frazy złożonej znakujemy następująco:

1. jeżeli istnieje jednostka o lemacie jednoczłonowym wyrażająca znaczenie segmentu będącego centrum wyrażenia wieloczłonowego (np. szkoła-2), znakujemy ten segment w sposób standardowy, a następnie wdrazamy drugi poziom znakowania;
2. w polu lemma2 wstawiamy lemat wieloczłonowy;
3. w polu nr wpisujemy właściwy numer jednostki o wieloczłonowym lemacie;
4. jeśli dane wyrażenie nie dające się zinterpretować w sposób kompozycyjny (tzn. interpretując każdy składnik oddzielnie) nie jest w *Słownosieci* uwzględnione, w polu a_lack wpisujemy lack.

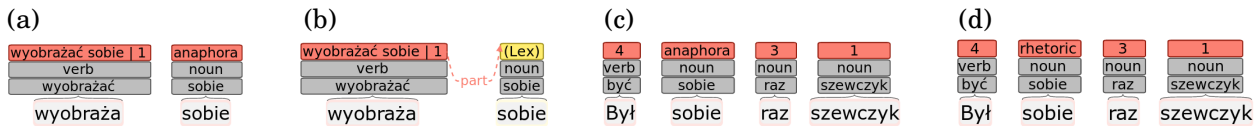
Natomiast pozostałe człony wyrażenia podłączamy do centrum rolą part, a te z nich, które posiadają adekwatną interpretację w *Słownosieci*, znakujemy w sposób standardowy.

6.2.1 Przysłowki zaczynające się od przyimka

W *Słownosieci* występuje też sporo przysłówków tworzonych w sposób regularny z przyimka i przysłówka (*na białe*), z przyimka i przymiotnika w mian. l.poj. r.ż. (*od dawna, z nagła*)⁷ bądź formami resztkowymi (ozn. burk) (*na oścież*). W takim wypadku znakujemy centrum semantyczne frazy w taki sam sposób jak centrum dowolnego innego wyrażenia jednoczłonowego, a następnie wdrazamy dlań drugi poziom znakowania, dopisując lemma2 i podłączając przyimek rolą part. Na przykład *na białe* posiada dwa znaczenia, przy czym *na białe-1* ma synonim *biało-2*, zaś *na białe-2* ma synonim *biało-7*. Z kolei *oścież* nie posiada w *Słownosieci* samodzielnej interpretacji semantycznej, więc mamy jeden poziom znakowania – wieloczłonowy.

⁶Nie zmienia to faktu, że *serce* i *pokłon* posiadają swoją interpretację.

⁷Formy takie są *de facto* archaicznymi postaciami przysłowka, ale znakowane są jako przymiotniki.



Rysunek 6: Niepoprawne i poprawne znakowanie nieanaforycznych wystąpień zaimka *sobie* – widok w programie WebAnno

6.2.2 Cząstka *sobie* w lemacie czasownika

Zaimek *siebie* ma przypisaną część mowy noun, ponadto w automatyczny sposób znakowany jest w polu `a_aseman` jako anaphora (por. sekcja 8.1).

Istnieją jednak sytuacje, w których forma *sobie* jest częścią lematu czasownika, np. *wyobrażać sobie*, *radzić sobie*, *imaginować sobie*, por. (16a), (21). W takich wypadkach traktujemy to jako przypadek lematu wieloczłonowego, czyli

1. Jeżeli istnieje jednostka o lemacie jednoczłonowym wyrażająca znaczenie czasownika (np. *imaginować*), znakujemy zarówno ją, jak i całość (znakowanie dwupoziomowe).
2. W polu `lemma2` wstawiamy lemat wieloczłonowy (np. *radzić sobie*).
3. W polu `nr` wpisujemy właściwy numer jednostki o wieloczłonowym lemacie, a jeśli nie istnieje (np. *imaginować sobie*) w polu `a_lack` wpisujemy `lack`.
4. Podłączamy człon *sobie* rolą `part`.
5. Ponadto dla członu *sobie* w polu `a_aseman` usuwamy kod `anaphora`.

Niepoprawne znakowanie czasownika *wyobrażać sobie* widnieje na rys. 6 (a), zaś skorygowane – obok na rys. (b).

Nie należy takich sytuacji nadużywać. Cząstka *sobie*, często w postaci enklitycznej *se* (podobnie jak *mi* itp.) może występować z wieloma czasownikami, por. (16c), dla których ewidentnie nie jest częścią lematu. Nie jest to też użycie anaforyczne i należy je znakować jako retoryczne `rhetoric`, por. sekcja 7.2, patrz też rys. 6 (c) – pozostawiony błąd oraz (d) – wersja prawidłowa.

- (16) a. *Chyba nie **imaginujesz** pan **sobie**, że wydamy Nirę za pana?*
 [...] dla lepszego efektu zaczęła **sobie imaginować**, że to jej język.
- b. - *Nie **imaginowałem**, że kapitan wytrzyma w solitudzie choćby jedno zdrowaś.*
- c. *Na naszym podwórku biegał sobię pies.*
 - *Jadą się jak bryczkami na ślub.*
Potem zjem sobię resztki wczorajszej pizzy.
Jeszcze trzy lata temu biegała mi po domu, suczka - foksterierka, [...]

6.3 Segmentacja

Może się zdarzyć, że pojedynczy wyraz został rozbity na części, celowo, np. (17a) lub w wyniku błędnej konwersji pliku (17b). W takim wypadku pierwszemu segmentowi przypisujemy lemat całości (np. *gimnastyczny*, *upadek*), ustawiając pole `a_uzup` na `segmentation`, a następnie znakujemy w sposób standardowy. Pozostałe segmenty podłączamy rolą `part`.

UWAGA: Znakowanie pojedynczych segmentów nie ma w tym wypadku sensu. Jeśli przypadkiem segment ma przypisaną automatyczną interpretację (np. *czuję, jak mnie **ocza- mi** rozbierają*), należy ją wyczyścić. Podobnie jest w wypadku wieloczłonowych wtrąceń obcojęzycznych (patrz poniżej), np. *Stasiu, **you must go back***.

- (17) a. *Do **gim- na- stycz- nej** SALI!*

- b. W procesie dziewięciorga oskarżonych o doprowadzenie do **upa- d- ku** Banku Staropolskiego skończył mówić Walenty P.

Charakte ry - zu je się nie za do wa la ją cym sta - nem tech nicz nym, [...]

6.4 Wieloczłonowe nazwy własne i zwroty obcojęzyczne

Część wyrażen wieloczłonowych reprezentowanych w *Słownosieci* to wieloczłonowe nazwy własne, np. geograficzne (*Ameryka Łacińska, Morze Białe, Sucha Beskidzka*) i inne (*Główny Urząd Statystyczny*). Reguły ich znakowania są podobne jak w wypadku innych wyrażen wieloczłonowych.

1. W polu nr wpisujemy adekwatne numery znaczeń, o ile istnieją (np. **1** dla *Ameryka, morze, główny, urząd*). W przeciwnym razie pole to ignorujemy (np. *łaciński, biały, suchy, beskidzki, statystyczny*).
2. Centrum frazy znakujemy (być może dwupoziomowo), przypisując odpowiedni lemat wieloczłonowy i numer znaczenia.
3. Centrum frazy przypisujemy wartość name w polu a_opt, zaś pozostałe człony podłączamy doń rolą part (i nie znakujemy ich samodzielnie).

Wieloczłonowa nazwa własna może zawierać inne wyrażenia wieloczłonowe, w tym nazwy własne. Kolejne poziomy dodajemy tylko wtedy, gdy ma to uzasadnienie – obecność w *Słownosieci*

lub niekompozycyjność semantyczną. W szczególności, znakowanie *Polskiego Związku Piłki Siatkowej* na rys. obok jest zbędne i dodane zostało wyłącznie w celach ilustracyjnych, bo wszystkie segmenty mają swoje znakowanie (*siatkowy* jako człon jednostki *piłka siatkowa-1*), a nazwy tej w *Słownosieci* nie ma (ale *Polski Związek Piłki Nożnej* jest!).

Z kolei większość wieloczłonowych wyrażen obcojęzycznych (np. *chargé d'affaires, high life*) składa się z segmentów nieznanymi analizatorowi morfologicznemu (ign). Nie przeszkadza to w przypisywaniu im znaczenia, tyle że w polu a_opt należy wybrać kod foreign.

Analogicznie postępujemy, gdy wyrażenie nie jest interpretowane (*apage satanas, c'est la vie, n'est-ce pas, panta rhei*), czyli łączymy znakując pierwszy składnik jako foreign – interpretacja słownosieciowa jest w tym wypadku opcjonalna. W szczególności, NIE dokonujemy tłumaczenia takich wyrażen.

UWAGA: Wyrażenie musi znajdować się w polskiej części *Słownosieci*. Część angielską całkowicie ignorujemy. To, że nie istnieją ustawienia interfejsu ograniczające widok do jednego języka jest moim zdaniem błędem.

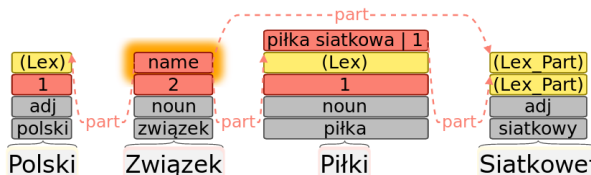
6.4.1 Burkinostki

Segmenty niesamodzielne, występujące w tekście wyłącznie jako składnik wyrażen wieloczłonowych, nazywane są *burkinostkami* (od państwa *Burkina Faso*) i posiadają znacznik części mowy burk.

Występują głównie w nazwach własnych np. *Burkina, Buenos, Sankt*. Zazwyczaj jeden z członów takiej nazwy jest rzeczownikiem. Jeśli tak jest, człon rzeczownikowy uznajemy za centrum nazwy, przypisujemy mu znacznik name, a pozostałe człony podpinamy rolą part, por. *Praktyczna instrukcja znakowania* sekcja 4.2.

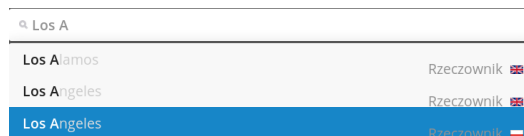
Jeśli wszystkie człony nazwy mają znacznik burk (lub ign), wybieramy pierwszą z nich jako centrum frazy, w polu a_suppl wybieramy znacznik burk:noun, w polu a_opt – znacznik name, a pozostałe człony standardowo podpinamy rolą part.

W obu wypadkach, jeśli nazwa taka występuje w **polskiej części** *Słownosieci*, jej lemat wieloczłonowy wpisujemy w pole lemma2, a numer znaczenia – w pole nr.



UWAGA:

W wypadku takich nazw, w swej większości obcego pochodzenia, trzeba bardzo uważać na symbol języka, wyświetlany w *Słownosieci*. *Burkina Faso*, *Los Angeles*, *Sankt Petersburg* w polskiej *Słownosieci* są, ale *Buenos Aires*, *Los Alamos* – tylko w bazującej na wordnecie prinstońskim jej części angielskiej, por. rys. powyżej.



Oznaczenie burk mają także *wraz*, *wespół* (przymyki *wraz z*, *wespół z*) oraz *dala*, *młodu* (przysłówki *z dala*, *za młodu*) itp.

Nierzeczownikowe składowe przymków kompletnie nas nie interesują (ew. por. sekcja 7.1). Jednak konstrukcje przysłówkowe znakować powinniśmy. Niezależnie od tego, jaka część mowy jest przypisana w warstwie III (np. noun dla *oścież*, other dla *wskroś*) ustawiamy znacznik `a_suppl` na burk:adv, w polu lemma2 wpisujemy lemat (*na oścież*, *na wskroś*, *po ciemku*, *po omacku*, *za młodu*, *z dala* itp.), wybieramy właściwy numer ze *Słownosieci* do pola nr bądź zaznaczamy lack, por. też sekcja 6.2.1.

6.5 Tytuły

W znakowanych tekstach mogą pojawić się tytuły utworów literackich, filmowych, muzycznych itp. Tytuły takie mogą być jednowyrazowe (*Silaczka*), często jednak są to całe wyrażenia (*Nieznosna lekkość bytu*, *Przeminęło z wiatrem*, *Meir Ezofowicz*).

Tytuły mają swoją strukturę wewnętrzną i dlatego wszystkie składające się na nie segmenty (m.in. nazwy własne, np. *Meir Ezofowicz* to imię i nazwisko bohatera) powinny być znakowane zgodnie z regułami opisanymi powyżej. Jednak na zewnątrz funkcjonują jako całości (z założenia rzeczownikowe), jak w zdaniu *Zasadniczo, jeśli się chce zachować wrażenia po przeczytaniu "Przeminęło z wiatrem", to odradzam dalszy ciąg. Z tego względu należy wdrożyć dodatkowy poziom znakowania dla pierwszego segmentu, przypisać mu znacznik title, a pozostałe składniki tytułu powiązać rolą part.*

7 Segmenty niepodlegające znakowaniu

Niektórych segmentów, mimo że należą do czasowników, rzeczowników, przymiotników bądź przysłówków, znakować nie należy.

W szczególności, jeżeli znakowany segment jest całkowitym neologizmem (zazw. ma przypisaną część mowy ign), to zamiast kodu lack w polu `a_lack` wybieramy w polu `a_aseman` kod neolog.

UWAGA: Za neologizmy uznajemy wyłącznie wyrazy, które nie posiadają poprawnej interpretacji morfoskładniowej (są więc rzeczownikami itp. jedynie potencjalnie), i to nie wyłącznie ze względu na homonimie (np. *lata* to forma od *rok* i *latać*), lecz dlatego, że wyraz nie jest jeszcze w pełni utrwalony w języku, i analizator morfologiczny *Morfeusz* go nie znał (np. *fantomowy*). Często jest to utworzony spontanicznie, żartobliwie (np. *oberświniopas*) lub należący do języka fikcyjnego (np. *fantoskopijny*, *fenomenoskop*, *fenomurzyn* itp.)

NIE znakujemy jako neologizmów nazw własnych ani wtrąceń obcojęzycznych – są dla nich przewidziane odrębne znaczniki.

Niektóre inne przypadki zostały omówione poniżej, inne, bardziej szczególne, w punkcie 8.

7.1 Rzeczowniki w wyrażeniach przymkowych

W języku polskim istnieją także tzw. *przymyki złożone*. Może je stanowić ciąg dwóch przymków (*co do*, *na wprost*). Takie ciągi kompletnie nas nie interesują, podobnie jak *wraz/wespół z*. Niejednokrotnie jednak przymyki złożone składają się z właściwego przymka i następującego

po nich rzeczownika (*na temat, w sprawie, z powodu, w ramach, ze strony, na rzecz*), por. (18a). Istnieją też ciągi <przyimek, rzeczownik, przyimek> (*bez względu na, do spółki z*), por. (18b).

- (18) a. *Teraz ty będziesz rozpaczać z **powodu** mojej nieobecności.
Nie wybierał zleceńodawców pod **kątem** ich sympatycznej fizjonomii.
Pod tym **kątem** analizuje się również codzienne zapisy w dzienniczkach uczuć.*
b. *Wróć **bez względu na to**, jaką otrzymam odpowiedź.*
c. *Nad jej czołem prawie pod **kątem** prostym sterczały długie włosy.*

Rzeczowników takich nie ma sensu znakować. Dlatego w polu a_aseman wstawiamy kod prep_element. Lista zidentyfikowanych przez nas przyimków złożonych została dołączona do tej instrukcji. Znow trzeba pamiętać, że zbitki takie nie zawsze oznaczają przyimki złożone, por. (18c).

7.2 Fragmenty retoryczne

Niektóre elementy wypowiedzenia nie odwołują się w żaden sposób do rzeczywistości pozajęzykowej, a jedynie pełnią funkcję wiążącą inne składniki, składniową bądź retoryczną. Często zastępują inne elementy w zdaniu (np. spójniki, partykuły), jak omówione w poprzednim punkcie przyimki złożone.

Na przykład zdanie (19a) różni się od zdania (19b) jedynie natężeniem emocjonalnym. Fraza *do jasnej cholery* stanowi wtręt, wykrzyknik, i nie ma sensu jej znakować, podobnie jak nie znakuje się innych wykrzykników (*Ach! Ha!, No, no...*).

- (19) a. *Powiedziałem, że mam w dupie ich składki i że do jasnej cholery nie życzę sobie...*
b. *Powiedziałem, że mam w dupie ich składki i że nie życzę sobie...*

Słowosiec zawiera jednostkę rzeczownikową wykrzyknik-1, wśród której hiponimów znajduje się m.in. synset {cholera-3, zaraza-4, jasna cholera-1, do jasnej cholery-1}, którego elementy są także rzeczownikami. Nie mamy tu rzecz jasna do czynienia z żadną hiponimią i dlatego prosimy zignorować wszystkie hiponimy tego synsetu.

Podobnie zdanie (20a) może zostać przeformułowane jako (20b). W zasadzie rzeczownik *prawda* zachowuje w nim swoje znaczenie (prawda-1) i należy go oznakować, jednak przysłówek *zawsze* już nie.

- (20) a. *[Pan jest prezesem.] Co prawda \emptyset honorowym, ale zawsze.*
b. *Niby \emptyset honorowym, ale jednak.*

We wszystkich podobnych wypadkach wpisujemy rhetoric w polu a_aseman.

8 Segmenty specjalnej troski

Istnieje pewien zestaw segmentów, które zostały potraktowane w sposób specjalny.

8.1 Zaimki osobowe

Pierwszymi z nich są zaimki osobowe. Zaimki w pierwszej i drugiej osobie *ja, ty, my, wy* oraz rzeczowniki *kto, któż, nikt* zostały automatycznie oznakowane hiperonimem osoba-1. Jeżeli kontekst jest sprzeczny z takim znakowaniem (np. w bajce mogą to być przedmioty), należy usunąć przypisane znaczenie z pola nr oraz kod auto:hipernym z pola a_suppl, a zamiast tego w polu a_aseman wpisać anaphora. W ten sposób znakowany jest automatycznie zaimek trzecioosobowy *on*.

Z wymienionych leksemów jedynie *nikt* oraz *ja* występują w *Słowosieci*. Pierwszy w znaczeniu 'osoba pozbawiona znaczenia', por. (21). Drugi w znaczeniu **1** 'osobowość, zespół cech

człowieka' oraz w znaczeniu 2 'struktura osobowości (ego)', por. (22). W wypadku napotkania takiego nietypowego użycia tych wyrazów należy wpisać numer znaczenia do pola nr oraz wyczyścić pole a_aseman.

- (21) *Kto sobie nie radzi, jest **nikim**, śmieciem, zerem.*
- (22) *Czy moje **ja** ma zniknąć po to, żeby czyjeś inne powiększyło się w dwójnasób?
Pomiędzy moim **Ja** i twoim **Ja** nie ma drabiny, jest przerwa.*

8.2 Zaimki pytajno-zależne i względne

Podobny problem powodują zaimki pytajno-zależne i względne. Zostały one automatycznie oznakowane w polu a_aseman jako interrogative. Jedynie *czyj, gdzie, jaki, który* znajdują się w *Słownosieci*. Znakować należy je jedynie w kontekstach, gdy dają się zamienić na *czyjś, czyjkolwiek, gdzieś, gdziekolwiek, jakiś, któryś, którykolwiek*, por. (23a), wpisując numer znaczenia w polu nr opróżniając pole a_aseman. Niestety, taka sytuacja może dotyczyć pozostałych zaimków tego typu, por. (23b). Wówczas także należy wyczyścić pole a_aseman i zamiast tego ustawić lack w polu a_lack.

- (23) a. *Jednak **czyj** by nie był, powinien on również służyć mieszkańcom wsi. (I)
Czy też zamordował ją **gdzie** indziej i tu czymś ją przywiózł?
Rozglądała się wokoło, czy **gdzie** nie dojrzy któregoś z chłopców, [...]]
Mało **który** policjant jest rozmowny, kiedy ważą się jego losy zawodowe.
Może się **który** poważy kłam mi zadać?*
- b. *Jak będziesz **kiedy** w potrzebie, to przyjdziemy ci z pomocą.*

Szczególną uwagę należy zwrócić na przymiotnik *który*, który (nomen omen) znakujemy w polu a_aseman jako:

- anaphora, gdy posiada wyraźny poprzednik w zdaniu nadrzędnym (standardowe użycie względne, por. (24)),
 - interrogative, gdy mamy do czynienia z konstrukcją pytajno-zależną, por. (25).
- (24) *To kolejna rzecz, **która** mnie wyjątkowo wkurza.
Jak inaczej rozumieć jej pochwały tanecznych wyczynów Patera, o **których** jego była żona miała bardzo marne zdanie?
Ale oto turkot słyszę, **który** ranną mąci ciszę...*
- (25) - *Skąd miałem wiedzieć, **które** zdrowe, które zepsute?
- A mnie trzeba mówić, **które** śruby odkręcać.
Policjanci z daleka widzieli, **który** z pojazdów należy zatrzymać.*

8.3 Znakowanie formy to

W języku polskim wyraz *to* ma wiele interpretacji, dlatego nie należy zakładać, że zostanie zinterpretowany poprawnie. Morfoskładniowo może on mieć przypisane następujące znaczniki:

1. partykuła qub, por. (26), oraz spójnik conj, por. (27), oba z przypisaną uproszczoną cz. mowy other;
2. predykatyw pred, notowany w programie jako verb, por. (28);
3. forma przymiotnika (adj) *ten* w mian. i bier. l.poj., por. (29);
4. zaimek rzeczowny (uznawany za rzeczownik subst, w programie oznaczenie noun).

- (26) *Kongres podnosi coraz **to** nowe problemy w stosunkach z Moskwą.
Z przebiegu gry na Bułgarskiej wynika jednak, że **to** poznaniacy byli lepiej przygotowani do gry.
Mógł wchodzić dopiero po zgaszeniu światła i **to** przez okno.*
- (27) *Jeśli mam jechać, **to** nie ma się co ociągać.*
- (28) *- Nasi rodacy **to** urodzeni handlarze.
Może ten czerwony napis **to** był początek takiego zakazu?*
- (29) *Dla pięciu naszych Czytelników mamy wejściówki na **to** muzyczne spotkanie.
Już wrosłem w **to** środowisko i uważam, że potrafię pogodzić swoje sprawy.*

Rzecz jasna ani spójnika *to*, ani partykuły *to* nie znakujemy. Wystąpienia predykatywne *to* powinny być znakowane przez jego jedyne znaczenie czasownikowe w *Słownosieci* o numerze **2**. Podobnie wystąpienia przymiotnikowe należy znakować jedynym znaczeniem przymiotnika – **1**.

Najwięcej problemów budzi znakowanie wystąpień rzeczownikowych. Standardowe wystąpienia, por. (30), znakujemy w polu *a_aseman* przez *anaphora*. Równie typowe, ale nie anaforyczne, wystąpienia to tzw. *korelaty*, służące do wprowadzania zdań podrzędnych, por. (31). Znakujemy je w polu *a_aseman* przez *grammatical*.

Słowosie zawiera dwa znaczenia rzeczownikowe *to*. Pierwsze, *to-1*, oznacza akt seksualny, por. (32), i w ten sposób powinno być znakowane (w polu *nr*). Drugie, *to-1*, oznacza psychoanalityczne *id* i mało jest prawdopodobne takie wystąpienie w tekście.

- (30) *Dotyczy **to** wszystkich kategorii zawodowych.
Wtedy wydawało mi się **to** nadmierną ostrożnością.*
- (31) *Wyglądało na **to**, że także ubierała się bez okularów.
- Wszystko wskazuje na **to**, że jest to amfetamina.
Jeszcze później pokłócili się o **to**, kto i jak będzie rządził, [...]
Nie miał zupełnie pomysłu na **to**, gdzie ją ulokować na skali przyjaciół i wrogów.*
- (32) *Zawsze wyobrażałam sobie, że robię **to** z kimś bardziej doświadczonym.
Czy marzyłaś kiedyś, żeby robić **to** na plaży?*

Jako *grammatical* znakujemy także wystąpienia *to*, co „spójnikowe” (*za to*, *przy czym*) czy „partykułowate” (*przy tym*), por. (33).

- (33) ***Za to** stary Cichocki przywitał się z nimi z wyraźną ulgą.
Są trochę staroświeckie, **to** prawda, ale **za to** takie niezwykle w dzisiejszych czasach...
A **przy tym** utwór ten zaskakuje swoją prostotą [...]
A **przy tym** bez przerwy coś czyta, [...]
Staralem się **przy tym** zrobić na nich wrażenie kogoś nieobliczalnego, bezwzględnego i "gotowego na wszystko".
Również sam premier Paderewski podjął w tej sprawie zabiegi, **przy czym** kontakty ułatwiał mu poseł Herman Lieberman.*
- (34) *Czasami brała się **za to** opieka społeczna, [...]
Może właśnie **za to** nieposłuszeństwo ludzkość została skazana na wieczne cierpienie?
Dopiero jak usiadłam **przy tym** potworze ceramicznym, ujrzałam Jego.
Księdza **przy tym** nie było.
[...] chciałby wiedzieć, jakie są przebiegi takich widywań, o czym Redaktor z Prezydentem gwarzy, co jedzą i **przy czym** siedzą.
Przy czym górnicy odpoczywają?*
- (35) ***Za co** brała się opieka społeczna?
Za jakie nieposłuszeństwo ludzkość została skazana na wieczne cierpienie?*

Przy jakim potworze ceramicznym usiadłam?

Przy czym księdza nie było.

O czym Redaktor z Prezydentem gwarzy, co jedzą i **przy czym** siedzą?

Przy czym górniczy odpoczywają?

(36) ***Za co** stary Cichocki przywitał się z nimi?

***Za co** są takie niezwykle?

***A przy czym** utwór ten zaskakuje swoją prostotą?

***A przy czym** bez przerwy coś czyta?

***Przy czym** Staralem się zrobić na nich wrażenie kogoś nieobliczalnego, bezwzględnego i "gotowego na wszystko"?

?**Przy czym** kontakty ułatwiał mu poseł Herman Lieberman?

Należy je odróżnić od zwykłych fraz przyimkowych, w których *to, co* ma charakter anaforyczny lub względny (zależny). Jako test potraktujmy fakt, że te standardowe użycia dają się zamienić na pytajne, por. (35) w przeciwieństwie do (36).

Taka procedura znakowania jest właściwa w sytuacji, gdy część mowy wpisana w warstwie III jest poprawna. Jednak błędna anotacja jest w tym wypadku wielce prawdopodobna, więc należy wprowadzić odpowiednią korektę w warstwie IV, por. sekcja 2. W wypadku, gdy segment powinien podlegać znakowaniu semantycznemu, wpisujemy numer znaczenia w polu nr (po sprawdzeniu poprawności części mowy!). Jeśli natomiast znakowane *to* jest anaforą bądź korelatem, w polu a_aseman wpisujemy adekwatny kod.

8.4 Znaczenia czasownika *być*

Ostatni problem to czasowniki *być* i *mieć*. W *Słownosieci* 4.0 wyraz *być* ma 10 znaczeń. Poniżej interpretujemy je na podstawie struktury wordnetu oraz informacji z innych słowników języka polskiego. Interpretacje te zostały skonsultowane z zespołem *Słownosieci*.

1. być-1 – znajdować się, przebywać w jakimś miejscu (o dowolnych obiektach fizycznych),
 - być w domu, w pracy, u kogoś na wsi, przy chorym, przy narodzinach dziecka;
 - być w teatrze, na koncercie, na spacerze;
 - urz. Jestem! «zwrot stanowiący potwierdzenie obecności na sali, w klasie itp.»;
 - być na (swoim) miejscu;
 - być z kimś sam na sam.

Por. też (37i).

Także dla użyć metaforycznych, por. ostatnie dwa zdania w (37i).

2. być-2 – mieć miejsce, odbywać się, następować, trwać przez jakiś czas (o wydarzeniach, sytuacjach czasie), por. (37ii).
3. być-3 – znajdować się w jakimś stanie, fizycznym lub psychicznym, ulegać czemuś, doznawać czegoś (o istotach żywych)
 - być pod czyimś urokiem, wrażeniem;
 - być pod wpływem kogoś, jakiejś idei, substancji;
 - być w dobrym, złym humorze, w nastroju do zabawy, w strachu, w rozpacz, w doskonalszej formie;
 - być w ciąży, w sytuacji bez wyjścia.

Por. też (37iii).

4. być-4 – istnieć w jakimś czasie, żyć (o istotach żywych, gatunkach, także o abstraktach), por. (37iv).

5. być-5 – brać w czymś udział, uczestniczyć (o ludziach)

- być na weselu, na zabawie, na zebraniu;
- być na wojnie, w powstaniu.

Por. też (37v).

6. być-6 – nie jest jasne, czego dokładnie dotyczy to znaczenie. Na pewno istot żywych. Proszę na razie nie używać.

7. być-7 – sięgać, dostawać do jakiejś wysokości, por. (37vii).

8. być-8 – uczęszczać gdzieś, korzystać z czegoś, uczyć się, studiować

- być w liceum, na studiach, na uniwersytecie;
- być na prawie, na medycynie, na trzecim roku prawa;
- być na kursie maszynopisania.

Por. też (37viii).

9. być-9 – posiadać jakąś cechę, por. (37ix).

10. być-10 – synonim predykatywnego *to*; być utożsamianym z czymś innym, por. (37x).

Zauważmy, że być-6 jest hiponimem być-3, które z kolei jest hiponimem być-9, podobnie jak być-7. Tak więc być-9 jest pojęciem bardzo ogólnym i powinno być stosowane jedynie wtedy, gdy jego hiponimy nie pasują.

- (37) i. - *Panie, **byłem** na Syberii, **byłem** w Kazachstanie i widziałem mahometan, **byłem** i w Mongolii, widziałem buddystów.
O szarej godzinie wszyscy **są** przed domem i siedzą tak do późnej nocy.
Czy wierzy pan jeszcze, że kiedyś **byliśmy** pod Moskwą?
Wszystko **było** pod ręką.
Na górze **był** korytarz długi jak w hotelu.
Na skórcie **był** siwy nalot.
Jeden gumowy but **był** na swoim miejscu, drugi spadł.
Moim zdaniem tam **było** pierwotne źródło konfliktu.
Tu **było** centrum spisku.*
- ii. *Bałagan **był** pod każdym względem, a to **było** przed rejestracją.
Był piękny poranek wczesnej kaszubskiej jesieni, [...]]
Na dworze **był** mróz, za silny mróz, gdyż również utrudniał oddech.
Nieraz, kiedy **była** ładna pogoda, wchodziłem do wody już w marcu, [...]]
Czym w Pani życiu **była** Akademia Sztuk Pięknych w Krakowie?
Był luty, ale śniegu w dolinach leżało niewiele.
Może **być** problem przy jakiejś, odpukać, kontroli.
- Mój Boże, to na pewno **był** list do mnie - szepnął Tomek.*
- iii. *Zecer jakoś się tam usadowił, przytulnie musiało mu **być** pod płachtą i względnie sucho, [...]]
Poza tym **byliśmy** pod wpływem Majakowskiego, [...]]
On **był** pod wpływem wielkiego stresu, który rozładował w taki sposób.
[...]] właściciel woreczka **jest** pod wpływem substancji odurzającej.*
- iv. *Awangardzistów **było** bardzo wielu.
A oni **byli**, **są** i **będą**, podobnie rudzi, blondyni, zielonoocy i inni.
Niewątpliwie przed potopem **było** gatunków więcej, ale wyginęły one [...]]*
- v. *Ale **byli** pod strażą i negować tego faktu nie było można.
Biły naraz w kilkanaście maszyn i naraz cała lewa kolumna wyprawy **była** pod ich ostrzałem.*

- vii. *Kostiumy są białoczarne do kolan z koronką i falbankami.
W niektórych miejscach woda była po kostki.
[...] śniegu było za kolana i mróz dokumentnie zaszył rzekę, [...]*
- viii. *niedawno byłem na szkoleniu z toksykologii.*
- ix. *Przymus jest silniejszy od rozsądku.
Popołudnie jest parne i duszne.
Pokój był duży, chociaż nie taki, jak ogromne hangary, w jakich nowa władza lubiła
wydawać zarządzenia i dekryty.*
- x. *Coś jednak się zmieniło w jego domu - od ponad tygodnia był posiadaczem magneto-
widu.
A może jest aniołem?
Jestem dziennikarzem gazety "Zew Robotnika".*

Jednak wyraz ten pełni także funkcje czysto składniowe. W wypadku, gdy forma przyszła tego czasownika użyta jest do tworzenia czasu przyszłego (*będzie biegł, będzie biec*), w polu a_aseman wybieramy kod grammatical. Podobnie postępujemy, gdy czasownik *być* służy do tworzenia strony biernej, chociaż w tym wypadku adekwatne jest znaczenie 9. Jednak np. w zdaniu (38) nie chodzi tak naprawdę o cechę mieszkania, tylko o fakt, że w danym momencie oświetlało je słońce.

Dotyczy to także konstrukcji przypuszczających, por. (39) oraz analitycznych konstrukcji czasu przeszłego i przyszłego z predykatywami. Rzadszy przypadek to archaiczny czas zaprzęszły, por. (41).

- (38) *Mieszkanie Mielnika było oświetlone przez wschodzące słońce.
[...]słodkie mleko może być zastąpione kefirem, serami, jogurtem, w których laktoza zo-
stała rozłożona.
Z Janowskim czy bez, z jego następcą czy bez niego - wniosek będzie głosowany.*
- (39) *I byłby wydał wspólników.
I tak by był umarł!*
- (40) *Na zdjęciach nie było widać ciał ofiar tragedii.
Co tam dokładnie jest, będzie wiadomo po oczyszczeniu stropu.
Zbankrutować będzie można tylko raz.*
- (41) *Właściwie uroczystości już się były rozpoczęły przed kilku dniami wspaniałym wjazdem
najaśniejszej pary do stolicy.*

W wypadku, gdy konstrukcje powyższego typu są skoordynowane z konstrukcjami znakowanymi, przypisujemy wymagane znaczenie. Na przykład dla *być* w zdaniu (42) adekwatne jest znaczenie 9.

- (42) *Teraz wszystko było skoszone, zebrane i martwe.*

8.5 Znaczenia czasownika *mieć*

W *Słownosieci* 4.0 wyraz *mieć* ma 5 znaczeń. Poniżej interpretujemy je na podstawie struktury wordnetu oraz informacji z innych słowników języka polskiego.

- mieć-1 – posiadać pewną własność, cechę:
 - składać się z czegoś, zawierać coś, być wyposażonym w coś; także: mierzyć, liczyć ileś czegoś;
 - być wyposażonym w coś, co odznacza się jakąś cechą;
 - mieć talent, zdrowie, rozsądek, cierpliwość, ambicję, silny charakter, pogodne usposobienie, jakiś gust, wady, zalety, szerokie zainteresowania, wąskie horyzonty;

- mieć jakiś kolor, kształt, smak;
- mieć coś po kimś «dziedziczyć po kimś jakieś cechy»;
- mieć jakieś wykształcenie, przygotowanie do czegoś.

Por. też (43i).

2. mieć-2 – być zmuszonym do czegoś, mieć obowiązek to wykonać por. (43ii).
3. mieć-3 – znajdować się w posiadaniu czegoś, być właścicielem, użytkownikiem lub dysponentem czegoś
 - mieć udziały w czymś;
 - mieć na coś, na kogoś «być w posiadaniu pieniędzy, środków na jakieś wydatki, móc ponosić koszty związane z czymś, z kimś».

Por. też (43iii).

4. mieć-4 – nie jest jasne, czego dokładnie dotyczy to znaczenie. Proszę na razie nie używać.
5. mieć-5 – żywić jakieś uczucia (do kogoś), por. (43v).

- (43) i. *Wysokie drzwi gabinetu **miały** dwa skrzydła [...]*
*Wszystkie domy **mają** gaz i wodę, [...]*
*Czapla siwa **ma** długość do 110 cm , a rozpiętość skrzydeł do 170 cm.*
*Moja siostra Basia **ma** lat 12 i chodzi do piątej klasy.*
*Ich skrzydła pod światło **mają** kolor prawie czarny.*
*"W ciemnej stronie Wenus" **miałam** postrzępione włosy, podarte ciuchy z lumpeksu.*
- ii. **Mają** dać znać, gdy coś wyniknie.
*O której **mają** przyjść?*
Mam coś do zrobienia.
*To wyłącznie ty **masz** dług do spłacenia, [...]*
*Do czegoś **miał** się drukarz przyznawać?*
- iii. *Jedna **ma** nawet suknię całą niby ze złota.*
*Wyprzedzała się, żeby **mieć** na chleb.*
*A tak **miałem** co jeść.*
- v. **Ma** żal do dziennikarzy, że o tym wypadku w ogóle piszą, że pokazali jej syna w telewizji.
Mam żal o to, że nie było tej rozmowy...
*A potem jeszcze do Maxa **mają** pretensje, że pali.*
*Ci ludzie **mają** radość życia, ich temperament mnie pociąga.*
*Pewno też **ma** dosyć wszystkiego.*
*Wielu **ma** im za złe tę służbę.*
Ma wypisane na czole, że tu zostanie.

Wyraz *mieć* także może pełnić funkcje czysto gramatyczne, niesamodzielne.

1. użycie dla wyrażenia przyszłości w czasie przeszłym, por. (44)
2. użycie z imiesłowem biernym w funkcji czasu przeszłego, por. (45).

UWAGA: Podobną strukturę mają też „zwykłe” użycia *mieć*, por. (46). Różnica jest taka, że wówczas podstawowy związek semantyczny występuje pomiędzy *mieć* a rzeczownikiem (*mieć kurtkę, rozpiszę*).

- (44) *Był on nieodwołalnie związany z tym, co **miało** się dzisiaj wydarzyć.*
*Tak **miało** się stać, scenariusz opublikowano drukiem, ale wskazówki zegara stanęły w miejscu.*

Miał zobaczyć Warszawę, zanim pójdzie na studia, i przywieźć ojcu potwierdzenie wspomnień.

Najbliższa przyszłość **miała** rozstrzygnąć o ich realności.

- (45) W liście Czyński pisał, że **ma** całe wakacje zatrute, i to przez nią, [...] [...] zobaczył, że całe nogi **ma** podrapane, pokiereszowane przez gałęzie krzaków i zadry pędów lian.
- (46) Poza tym będziesz **miał** rozpiszę koncertów sporządzoną przez jego menedżera, [...] Ja **miałem** kurtkę uszytą przez Jadwigę [...]

W takich wypadkach, podobnie jak dla *być*, w polu a_aseman wpisujemy kod grammatical. W konstrukcjach zanegowanych może przyjmować znaczenia *być*:

- być-1 (- Nie **ma** go ani w domu, ani pod adresem podanym przez jego żonę.);
- być-2 (- Nie **ma** mrozu, nie jest zimno.);
- być-4 (Takich facetów nie **ma** wielu, szczególnie w naszym roczniku.);
- być-5 (Nie **ma** go na pogrzebie.);
- być-8 – sztuczne, ale możliwe.

W takim wypadku w polu nr wpisujemy adekwatny numer znaczenia *być*, z tym że w polu a_suppl wstawiamy kod wyjątku negation-byc. Nie jest to bowiem błąd ani brak, tylko specyficzna cecha języka.

Istnieją też dwa niekontrowersyjne znaczenia dla *mieć się*.

W *Słownosieci* jest wiele frazeologizmów z *mieć*. Dotyczy to w szczególności konstrukcji lekkich oraz:

- *mieć na sobie* coś, por. (47);
- *mieć przy sobie / obok siebie / przy (swoim) boku* kogoś: przebywać z kimś lub mieszkać z kimś;
- *mieć pod sobą / nad sobą* kogoś: być czyimś podwładnym lub zwierzchnikiem;
- *mieć za sobą* kogoś: być popieranym przez kogoś;
- *mieć za sobą* coś (lata pracy, życie, spotkanie): zakończyć jakąś sytuację, etap życia;
- *mieć przed sobą* coś (przyszłość, spotkanie, perspektywę): czeka nas jakaś sytuacja dla nas ważna;
- *mieć do siebie to, że*, por. (48);
- *mieć w sobie* coś «być interesującym, pociągającym, mieć swoisty urok, wdzięk»;
- *mieć w sobie* coś z kogoś, z czegoś «przejawiać cechy charakterystyczne dla kogoś, czegoś».

Wyrażenia takie znakujemy w podobny sposób jak inne pojęcia o lematach wieloczłonowych, por. (6.2). Znakowanie zależy rzecz jasna od faktu, czy odpowiedni lemat wieloczłonowy został w *Słownosieci* uwzględniony.

- (47) Bo **mam** na sobie tę swoją budrysówkę i granatowe spodnie, co w nich chodzę do budy, i buty już dość zniszczone.
- (48) Wielkie rewolucje **mają** to do siebie, że są niszczyielskie, barbarzyńskie, dzikie.

Poniżej wyliczamy konstrukcje, w wypadku których nie mamy jasności, w jaki sposób znakować wystąpienia *mieć*. W chwili obecnej należy wpisywać ustawiać pole a_lack na lack.

Dotyczy to wypowiedzeń, w których bezokolicznik nie oznacza obowiązku, tylko pewną możliwość zajścia sytuacji (czasem znaczy to, że coś spodziewanego ostatecznie nie zaszło):

- (49) *Nie tak **miał** wyglądać nasz powrót do Warszawy, [...]*
*Jak ja to **mam** rozumieć?*
*I to **ma** być wielka mi wolność.*
*Za to muzyka, która **miała** ilustrować degrengoladę USA, rozbrzmiewała znakomita.*
*Ale skąd Mikołaj Rej z Nagłowic **miałby** o tym wiedzieć, [...]*
*Lecz komu **miał** o tym powiedzieć?*
***Miał** powiedzieć, że przywódcami mas są najczęściej ludzie na wpół obłąkani, [...]*
*Cichowski nie **miałby** po co dzwonić.*
*– Ja **miałbym** panu grozić?*

Nie wiadomo też, jak interpretować *mieć* w następujących zdaniach:

- (50) *Rowiński **ma** syna z pierwszego małżeństwa, wiesz?*
*Przed Niemcem chcesz zataić, że **masz** ojca generała!*
*A kogo **masz** za swych braci?*
- (51) - ***Mam!** - wpadł na pomysł.*
- (52) - *A **masz** ci za mój rozbity nos! a **masz!** a **masz!***
***Masz** go, cholera, dziany facet, fura musiała majątek kosztować, a taka sknera.⁸*
- (53) - ***Masz** tu trzydzieści złotych, idź do miasta i kup sobie buty i koszulę.*
***Masz** tu swoją legitymację i zmykaj.*
*- **Masz**, weź to ode mnie.*
*No, **macie**, zakładajcie.*
- (54) *Wiedział, że nie **miała** mężczyzny.*

W zdaniach (53) nie jest to w zasadzie *mieć-1*, lecz rozkaznikowe *bierz(cie)*, *weź(cie)*.

Brak znaczenia odpowiadającego «znajdować się w jakiejś sytuacji, żyć, przebywać w jakichś warunkach, uczestniczyć w czymś»:

- *mieć* dobrą pogodę, udane wakacje, fatalny rok;
- *mieć* egzamin, spotkanie;
- *mieć* gdzie mieszkać, co czytać, z kim porozmawiać;
- *mieć* rację w czymś, prawo do czegoś (czy to nie lekkie?);
- *mieć* do pogadania (z kimś), do przemyślenia (problem) itp.

Wystąpienie w zwrotach typu *Masz ci los!* *Masz babo placek* nie interpretujemy, tylko oznaczamy jako retoryczne w polu *a_aseman*.

⁸Ale to podobne do zwrotów *Patrzcie go, jaki obraźliwi!* - *Znalazła się wrażliwa, widzicie ją!*.