

****Kodeks standaryzacji****

Zasady lematyzacji wyrażen metaforycznych

****Uwaga: ****Niniejsze zasady mają charakter ogólny i mogą wymagać dodatkowych wytycznych w zależności od konkretnego zbioru danych (np. korpus metafor odzwierzęcych może wymagać osobnych reguł dotyczących nazw ras, fikcyjnych stworzeń itp.). Anotacja ma na celu dać ogólny obraz wyrażenia metaforycznego, który może pomijać pewne niuanse wpływające na interpretację (np. konkretna rasa psa niesie inne konotacje niż ogólne pojęcie *pies*, ale forma ustandaryzowana oddaje szerszą kategorię).

****Część 1: Ogólne zasady standaryzacji****

****A. Forma gramatyczna****

- ****Rzeczowniki: ****Standaryzujemy do formy w ****mianowniku liczby pojedynczej****.
- *nosorożce (l. mn.)* → *****nosorożec*****
- ****Czasowniki: ****Standaryzujemy do ****bezokolicznika****, zachowując prefiksację i aspekt. Jeżeli słownik/zasób nie rozpoznaje formy z prefiksem lub aspektem, usuwamy aspekt.
- *wylatywać* → *****wylatywać***** (prefiks i aspekt zachowane)
- Jeśli słownik nie rozpoznaje formy, cofamy się do czasownika bazowego.
- ****Przymiotniki relacyjne użyte jako metafora: ****Zamieniamy na odpowiadający ****rzeczownik****.
- *szczurzy (przymiotnik)* → *****szczur*****

****B. Zdrobnienia****

- ****Zdrobnienia pieszczotliwe/stylistyczne: ****Standaryzujemy do formy bazowej (niezdrobniałej).
- *misiu / miś (pieszczotliwie)* → *****niedźwiedź*****
- *motylek* → *****motyl*****

****Uwaga: Zdrobniałą formę należy pamiętać przy interpretacji metafory – wymiar afektywny może być istotny dla analizy.***

- ****Zdrobnienia oznaczające młode zwierzęta: ****Nie**** standaryzujemy do formy dorosłego zwierzęcia. Badacz musi na podstawie kontekstu rozstrzygnąć, czy zdrobnienie odnosi się do formy pieszczotliwej dorosłego, czy do młodego/innego bytu.
- *kotek* (w znaczeniu 'młody kot') ≠ *kot*. Pozostaje *****kotek*****.
- ****Zdrobnienia o odrębnym znaczeniu (np. nazwy potraw): ****Nie**** standaryzujemy do formy bazowej zwierzęcia/przedmiotu. To osobne jednostki leksykalne.
- *gołąbek* (potrawa) ≠ *gołąb*. Pozostaje *****gołąbek*****.
- *kopytka* (potrawa) ≠ *kopyto*. Pozostaje *****kopytka*****.

****C. Rodzaj gramatyczny****

- ****Zachowujemy rodzaj gramatyczny/biologiczny. ****Nie neutralizujemy form rodzajowych do formy domyślnej.

- *łania* (samica jelenia) ≠ *jelen* . Pozostaje ***łania***.

Uzasadnienie: Formy rodzajowe mogą być istotne w analizach związanych z płcią (gender-related studies).

D. Nazwy ras, odmian i podkategorii

- Standaryzujemy **nazwy konkretnych ras/odmian do kategorii ogólnej**, którą reprezentują.

- *owczarek niemiecki* → ***pies***

Uwaga: Konkretna rasa może nieść odmienne konotacje (np. owczarek niemiecki vs. pudel), ale forma ustandaryzowana oddaje szerszą kategorię na potrzeby systematycznej anotacji.

E. Znaki towarowe i nazwy marek (eponimy)

- Standaryzujemy **nazwy marek i eponimy do nazwy produktu ogólnego**, który reprezentują.

- *Pampers* → ***pielucha***

- *adidasy (potoczny eponim)* → ***buty sportowe***

F. Postaci fikcyjne i nazwy własne

- **Postaci fikcyjne: Standaryzujemy do najbardziej wyrazistej cechy (most salient feature)** postaci, która jest istotna w kontekście metaforycznego użycia.

- *Król Lew* → ***lew***

- *Świnka Peppa* → ***świnia***

- *Kosogłos* → ***symbol – metafora odnosi się do roli symbolu buntu, nie do ptaka samego w sobie***

- Jeśli postaci nie da się sprowadzić do jednej wyrazistej cechy, **zachowujemy oryginalną nazwę** i wyjaśniamy w komentarzu.

G. Wyrażenia wielowyrazowe (złożenia i kolokacje)

- **Jeśli wyrażenie wielowyrazowe jest hasłem słownikowym, zachowujemy całą frazę** i bazujemy anotację na znaczeniu słownikowym.

- *pies ogrodnika* – ma hasło w słowniku (odnosi się do człowieka); zachowujemy jako ***pies ogrodnika***.

- Jeśli frazy **nie ma w słowniku**, standaryzujemy do **najbardziej wyrazistej cechy** (zazwyczaj rzeczownika głównego; por. Fuoli 2022 nt. premodyfikatorów).

- *zmokła kura* – nie jest stałym hasłem słownikowym → ***kura***

- *bezpieński pies* – nie jest stałym hasłem słownikowym → ***pies***

H. Definicje: potoczne vs. naukowe

- Przy wyborze definicji dla formy ustandaryzowanej zawsze wybieramy **potoczne, codzienne rozumienie** słowa, a nie definicję naukową czy taksonomiczną.

- *ogórek*: używamy potocznego znaczenia 'warzywo', a nie definicji botanicznej ('roślina z rodziny dyniowatych...').

Uzasadnienie: Metafory opierają się na konwencjonalnej, ucieleśnionej wiedzy (por. poznanie ucieleśnione w językoznawstwie kognitywnym). Znaczenie bazowe (BM) powinno odzwierciedlać doświadczenie zmysłowe i codzienne.

I. Świadomość językowa (dane wielojęzyczne)

- Pracując z danymi dwu- lub wielojęzycznymi, należy upewnić się, że definicje i formy ustandaryzowane odpowiadają **właściwemu językowi**. Nie należy mylić terminów z różnych języków.

- Polskie *puzzle* vs. angielskie *jigsaw puzzle* – mogą się różnić zakresem i użyciem.

Jest to szczególnie ważne w zautomatyzowanych procesach pozyskiwania definicji z zasobów wielojęzycznych.

Część 2: Dodatkowe uwagi dla użytkowników Słownosieci (plWordNet)

Poniższe uwagi dotyczą badaczy korzystających ze Słownosieci (plWordNet) do pozyskiwania definicji i łańcuchów hiperonimów.

A. Rzeczowniki występujące tylko w liczbie mnogiej

- *śmieci* – Słownosiec nie rozpoznaje formy pojedynczej *śmieć* jako rzeczownika; traktuje ją jako czasownik (o znaczeniu 'mieć czelność'). W takich przypadkach używamy formy mnogiej, jaką podaje Słownosiec, i odnotowujemy rozbieżność.

B. Usuwanie formy ustandaryzowanej z łańcucha hiperonimów

- Konstruując łańcuch hiperonimów ze Słownosieci, **usuwamy samą formę ustandaryzowaną**, aby uniknąć redundancji. Łańcuch powinien zaczynać się od pierwszego hiperonimu powyżej formy ustandaryzowanej.

C. Znaczenie bazowe (BM) a luki w hiperonimach

- Niektóre definicje w Słownosieci uznawane za znaczenie bazowe (BM) – tj. najbardziej podstawowe, zmysłowo ugruntowane znaczenie słowa (zgodnie z poznanie ucieleśnione) – mogą zawierać sformułowania, które **nie pojawiają się w łańcuchu hiperonimów**. Oznacza to, że pewne informacje pojęciowe z definicji BM mogą nie być widoczne w anotacji opartej wyłącznie na hiperonimach. Badacze powinni być świadomi tej luki i konsultować definicje BM równolegle z łańcuchami hiperonimów.

D. Brak definicji w Słownosieci

- Gdy słowo nie ma definicji w Słownosieci, sprawdzamy **synonim międzyjęzykowy (interlingual link)** w celu znalezienia angielskiego odpowiednika i korzystamy z definicji angielskiej.

- Jeśli ani polska, ani angielska definicja nie jest dostępna w Słownosieci, sięgamy po **słownik zewnętrzny**: PWN (Słownik Języka Polskiego) dla polskiego, Merriam-Webster dla angielskiego. Wybór konkretnego słownika może zależeć od potrzeb badacza.

E. Hasła potoczne / nieformalne

- Słownosiec może zawierać hasła potoczne lub nieformalne. Traktujemy je zgodnie z tymi samymi zasadami standaryzacji, ale należy mieć świadomość, że ich łańcuchy hiperonimów lub definicje mogą być mniej kompletne.

F. Automatyczne pozyskiwanie danych

- Korzystając z narzędzi automatycznych do pobierania definicji ze Słownosieci, należy upewnić się, że narzędzie pobiera definicje w ****odpowiednim języku**** zgodnym ze zbiorem danych. Zasoby wielojęzyczne mogą zwracać angielskie glosy dla zapytań po polsku lub odwrotnie.