

Magdalena Wismont   
Uniwersytet im. Adama Mickiewicza, Poznań  
magwis7@amu.edu.pl

Ewa Nowak-Pasterska   
Uniwersytet im. Adama Mickiewicza, Poznań  
ewa.n@amu.edu.pl

Kinga Zalejarsz   
Uniwersytet im. Adama Mickiewicza, Poznań  
kinga.zalejarsz@amu.edu.pl

Dorota Rojszczak-Robińska   
Uniwersytet im. Adama Mickiewicza, Poznań  
dorroj@amu.edu.pl

## SPECYFIKA STAROPOLSZCZYZNY A ANOTACJA GRAMATYCZNA O LEMATYZACJI TEKSTU STAROPOLSKIEGO<sup>1</sup>

**Słowa kluczowe:** staropolskie apokryfy Nowego Testamentu, średniowieczna polszczyzna, anotacja gramatyczna, lemat, lematyzacja, forma słownikowa

**Keywords:** Old Polish New Testament Apocrypha, medieval Polish language, grammatical annotation, lemma, lemmatization, dictionary form

Dzięki projektowi pod tytułem „Początki języka polskiego i kultury religijnej w świetle średniowiecznych apokryfów Nowego Testamentu. Uniwersalne narzędzie do badań polskich tekstów apokryficznych”, realizowanemu w ramach grantu Narodowego Centrum Nauki na Wydziale Filologii Polskiej i Klasycznej UAM od marca 2018 r., ułatwiony został wieloaspektowy ogląd dziewięciu staropolskich apokryfów, czyli siedmiu rękopisów: *Rozmyślanie przemyskie* (RP), *Sprawa chędogo o męce Pana Chrystusowej* (SCh), *Rozmyślania dominikańskie* (RD), *Karta Rogawskiego* (KR), *List Lentulusa* (LL), *Ewangelia Nikodema* (EN) i *Historija Trzech*

---

<sup>1</sup> Artykuł realizowany w ramach grantu Narodowego Centrum Nauki (projekt nr 2017/26/E/HS2/00083) pt. „Początki języka polskiego i kultury religijnej w świetle średniowiecznych apokryfów Nowego Testamentu. Uniwersalne narzędzie do badań polskich tekstów apokryficznych” pod kierownictwem D. Rojszczak-Robińskiej. Udział współautorów jest równy i wynosi po 25%.

*Kroli* (HTK), oraz dwóch druków: *Żywot Pana Jezu Krysta* (ŻPJK) i *Żywot świętej Anny* (ŻSA). W ramach projektu stworzono narzędzie dające możliwość wieloaspektowej i interdyscyplinarnej eksploracji materiału staropolskiego. Będzie ono przydatne przede wszystkim dla językoznawców i literaturoznawców, ale także teologów czy osób zainteresowanych ogólnie ujętą kulturą dawną. Narzędzie udostępnione jest na stronie *Staropolskie apokryfy Nowego Testamentu* (dalej: Apocrypha). Znajdziemy tam teksty wszystkich dziewięciu apokryfów, ich transliteracje i transkrypcje przygotowane według rozwiązań edytorskich opracowanych przez zespół grantowy, zestawienie źródeł i wcześniejszych transkrypcji, wątków oraz miejsc paralelnych. Oprócz tego na stronie Apocrypha możemy korzystać z komentarza gramatycznego oraz wyszukiwarki gramatycznej i leksykalnej, co jest możliwe dzięki dokonanej anotacji materiału staropolskiego. Każdy wyraz zatem został zakwalifikowany do klasy gramatycznej, następnie przypisano mu odpowiednie wartości kategorii gramatycznych, a także nadano mu lemat. Narzędzia te opisaliśmy w innym artykule zatytułowanym *Anotacja gramatyczna tekstu staropolskiego – problemy i wyzwania* (Deskur, Nowak-Pasterska, Rojszczak-Robińska et al. 2022).

Ręczna anotacja tekstów wymusiła dostosowanie tworzonych funkcjonalności do materiału staropolskiego. Specyfika staropolszczyzny wpłynęła na to, że nie był to łatwy proces. O wielu wyzwaniach pisaliśmy już w poprzednich artykułach. Jedną z problematycznych kwestii było stworzenie odpowiedniej siatki klas i kategorii (zob. *ibid.*). Trudności wpłynęły na powstanie innowacyjnych rozwiązań, którymi były nowe wartości kategorii, czyli o i X, oznaczające: o – nie dotyczy, X – nie można określić (Rojszczak-Robińska, Wismont, Zalejarz et al. 2023). Z kolei sposoby konstruowania form czasu i trybu wpłynęły na decyzję o wprowadzeniu drugiego poziomu anotacji, co też już omówiliśmy w innym artykule z cyklu (Stelmach, Wismont, Zalejarz et al. 2023). Celem niniejszego artykułu zaś jest wskazanie problemów pojawiających się przy lematyzacji tekstów staropolskich oraz przybliżenie reguł, którymi się kierowano.

Lematyzacja to nadanie każdemu wyrazowi lematu, a więc pewnej etykiety wspólnej dla wszystkich form tego wyrazu. We współcześnie tworzonych korpusach tekstów lematyzacja nazywana jest także hasłowaniem. Przeważnie w nadaniu lematu bierze się pod uwagę formę podstawową wyrazu. Taki sposób zastosowano w NKJP, gdzie lematyzacja to sprowadzenie wyrazu do formy podstawowej, nieodmiennej (Przepiórkowski, Bańko, Górski et al. red. 2012: 214), która wskazuje na odpowiedni leksem i stanowi część znacznika, czyli tagu (*ibid.*: 59). W KorBa lematem również stała się forma podstawowa wyrazu, a anotacji dokonano automatycznie za pomocą dwóch narzędzi (Concraft i Toygger), działających niezależnie od siebie. Z tego względu istnieją dwie wersje korpusu z różnicami w zakresie lematyzacji (Gruszczyński, Bronikowska 2024: 7). W narzędziu Apocrypha natomiast anotacji

dokonano ręcznie, a lematem dla każdego wyrazu uczyniono jego formę słownikową, nie zaś formę podstawową<sup>2</sup>.

Podczas ręcznej anotacji gramatycznej zespół grantowy dokonał lematyzacji wszystkich wyrazów według ustalonych reguł (szczegółowo opisanych dalej). I tak na przykład dla wszystkich rzeczowników lematem jest forma mianownika liczby pojedynczej (np. formy *rozum*, *rozumu*, *rozumowi*, *rozumem* mają lemat ROZUM) lub forma mianownika liczby mnogiej (np. lemat LUDZIE, USTA); dla przymiotników – mianownik liczby pojedynczej rodzaju męskiego (np. lemat CHWALEBNY, RADY); dla form czasownikowych i imiesłowowych – bezokolicznik (np. formy *czyńnię*, *czyńcie*, *czyńił*, *czyńiąc*, *czyńiący*, *czyńiony* mają lemat CZYNIĆ). Decyzje dotyczące lematyzacji zostały podjęte w taki sposób, aby oddać specyfikę staropolszczyzny, ukazać różnorodność materiału leksykalnego i uwzględnić wszystkie jednostki leksykalne, a jednocześnie ułatwić przeszukiwanie materiału. Należy podkreślić, że lemat w narzędziu Apocrypha ma postać staropolską, na przykład BOG, EJIPT, JESTLI, LEPAK. Poszczególne rozstrzygnięcia zostaną opisane w dalszej części artykułu. Zbiór lematów stanowi zatem swoisty indeks wyrazów pojawiających się w tekstach apokryficznych, który uwzględnia warianty, wszystkie postaci wyrazu niezależnie od znaczenia. Jak już wspomniano, w narzędziu dostępnym na stronie Apocrypha lemat jest formą słownikową, a nie formą podstawową danego słowa. Wynika to z konieczności ułatwienia wyszukiwania wyrazów. Tak pojęta lematyzacja pomaga bowiem w uporządkowaniu leksemów bez grupowania ich ze względu na znaczenie czy funkcję w zdaniu i co za tym idzie – odnalezieniu ich w udostępnionych wyszukiwarkach: leksykalnej i gramatycznej. Po etykietce właśnie znajdujemy wszystkie formy gramatyczne danego wyrazu, które pojawiły się w materiale apokryficznym. Badacz nie jest skazany na decyzje zespołu anotującego, gdyż może wyszukać na przykład wszystkie użycia słowa *drugi*, niezależnie, czy uznano je za zaimek przymiotny czy liczebnik porządkowy. Tak jest w wyszukiwarce leksykalnej, dostępnej w zakładce *Słowa* (Apocrypha), o której już pisaliśmy (Deskur, Nowak-Pasterska, Rojszczak-Robińska et al. 2022: 175).

Praca z wyszukiwarką po słowach jest intuicyjna. W pasku wyszukiwania należy wpisać interesujący nas leksem. Po zatwierdzeniu wyświetlą się wyniki i ich liczba. W okrągłym nawiasie prezentowana jest frekwencja wyrazu. Otrzymujemy zatem listę wszystkich użyc danego wyrazu w każdej formie gramatycznej i w każdym tekście. Istnieje możliwość sortowania wyników (po lokalizacji lub formach gramatycznych) bądź ich zapisania na dysku (w tym celu należy kliknąć *Eksportuj*).

---

2 Przez formę słownikową rozumiemy pewną etykietę dla zbioru wszystkich form danego leksemu, a przez formę podstawową – formę, od której tworzone są formy fleksyjne (co najwyraźniej widać na przykładzie czasownika i imiesłowów, dla których bezokolicznik jest tylko jedną z form, a nie formą podstawową). W różnych projektach przyjmuje się różne rozwiązania, np. w SPXVI tak rozumiana forma podstawowa jest formą słownikową.

Po wpisaniu trzech pierwszych liter w pasku wyszukiwania pojawiają się podpowiedzi lematów (np. po wpisaniu ciągu liter *roz* otrzymamy podpowiedzi: ROZCIGAĆ, ROZDZIELAĆ, ROZDZIARAĆ, ROZEJŚĆ, ROZERWAĆ, ROZGRZESZENIE, ROZKRZYCZEĆ, ROZKURCZYĆ, ROZKWITNAĆ, ROZLANIE itd.). Warto podkreślić, że użytkownik nie musi znać wszystkich lematów ani wyszukiwać ich w opracowaniach leksykograficznych. Nad listą użyć wybranego lematu pojawią się też jego warianty (np. przy wynikach lematu JEZUS pojawia się odnośnik do lematu JESUS; wariantami lematu NATYCHMIAST będą: NATYMIAST, NATECHMIAST, NATEMIAST, NATEMMIAST). Dzięki temu możliwe było zachowanie różnorodności leksyki staropolskiej bez utrudniania wyszukiwania. Z listy otrzymanych wyników można przejść do pola pracy z tekstem. Poza tym istnieje możliwość dopasowania wyników do bieżących potrzeb użytkownika dzięki zastosowaniu odpowiednich filtrów, na przykład filtra *Tekst* lub *Klasa*. W wyszukiwarce leksykalnej możemy odnajdywać nie tylko całe wyrazy, ale również poszczególne jego części, co jest niezmiernie przydatne w badaniach słowotwórczych (Apocrypha: *O projekcie, Instrukcje*).

### 1. Zasady lematyzacji poszczególnych klas<sup>3</sup>

Zespół grantowy dokonał lematyzacji poszczególnych wyrazów według następujących zasad ustalonych dla każdej z klas:

1. Dla rzeczownika, niezależnie od znaczenia i słowotwórczego pochodzenia (również odczasownikowego), lematem jest zawsze mianownik liczby pojedynczej (np. *anjoła, anjołami* – ANJOŁ, *bojowania, bojowaniu* – BOJOWANIE), z kolei dla *plurale tantum* – mianownik liczby mnogiej (np. *drzwi, drzwiami, drzwiom* – DRZWI). Rzeczowniki dwurodzajowe, nieposiadające wariantów, mają ten sam lemat (np. *ten książkę, to książkę* – lemat KSIĄŻĘ, *ten sługa, ta sługa* – lemat SŁUGA).
2. Lematem przymiotników, również tych zsubiektywizowanych, a także w pełni zadiektywizowanych imiesłów na *-ły*, jest mianownik liczby pojedynczej rodzaju męskiego w stopniu równym (np. CHUDY, UBOGI, DOJRZAŁY). Bez względu na typ odmiany przymiotnika lemat to zawsze forma w odmianie złożonej, na przykład *pełn/pełen* – PEŁNY, *Dawidow/Dawidowy* – DAWIDOWY. W wypadku większości form zależnych występujących w materiale rozstrzygnięcie, czy mamy do czynienia z formą prostą czy złożoną, nie było możliwe.
3. Dla wszystkich przysłówków lematem jest forma w stopniu równym. Choć wiemy, że przysłówki się nie odmieniają, ze względów technicznych uznaliśmy, że takie rozwiązanie ułatwi przeszukiwanie materiału, na przykład forma *barziej*

3 Dokładny opis klas i kategorii zob. Deskur, Nowak-Pasterska, Rojszczak-Robińska et al. 2022.

ma lemat BARZO. Zleksykalizowanym wyrażeniem przyimkowym, na przykład *po lekku*, przypisano lemat nieciągły (PO LEKKU).

4. Zaimki zostały podzielone na kilka odrębnych klas, odpowiadających pod względem gramatycznym poszczególnym częściom mowy. Przyjęto dla nich następujące lematy:

- zaimek rzeczowny – mianownik liczby pojedynczej (np. ON, KTO, COŻ),
- zaimek przymiotny – mianownik liczby pojedynczej rodzaju męskiego (np. TEN, KTORY, JENŻE),
- zaimek liczebny – mianownik (np. TYLE, ILE),
- zaimek przysłowny – forma tekstowa (np. GDZIE, DOKĄD, INDZIE, KĘDY).

Lematyzacja zaimków okazała się pod pewnymi względami trudna. Rozwiązania szczegółowe przedstawimy w dalszej części artykułu.

5. Liczebniki również zostały podzielone na odrębne klasy, wyznaczone na podstawie kryterium semantyczno-formalnego. Przyjęte rozwiązania miały wpływ na przypisanie kategorii oraz lematów, które są następujące:

- liczebnik główny, liczebnik zbiorowy – mianownik (np. PIĘĆ, PIĘCIOR),
- liczebnik porządkowy, liczebnik wieloraki – mianownik liczby pojedynczej rodzaju męskiego (np. PIĄTY, SIEDMIORAKI),
- liczebnik ułamkowy – mianownik liczby pojedynczej rodzaju nijakiego (np. POŁ, POŁCZWARA),
- liczebnik krotny i liczebnik zespolony odmienne – mianownik liczby pojedynczej rodzaju męskiego (np. STOKROTNY, SAMOTRZECI),
- liczebnik krotny nieodmienny – lemat stanowi wyraz tekstowy (np. SIEDMKROĆ).

6. Dla form czasownikowych i imiesłowowych przyjęto następujące rozwiązania:

- forma terażniejsza, forma przyszła prosta, wykładnik czasu przyszłego BYĆ, wykładnik czasu zaprzeszłego BYĆ, aoryst, rozkaźnik, bezokolicznik, aglutynant BYĆ, aglutynant aorystu BYĆ – lematem jest bezokolicznik,
- formy *winien/powinien* – lemat WINIEN/POWINIEN,
- predykatyw – lemat to wyraz tekstowy (np. TRZEBA).

Forma bezokolicznika jest również lematem dla pseudoimiesłowu, a także dla imiesłowu czasu terażniejszego czynnego nieodmiennego, imiesłowu czasu terażniejszego czynnego odmiennego, imiesłowu czasu przeszłego czynnego, imiesłowu biernego, w tym też tak zwanego bezosobnika (np. *szosty a siodmy wiek pospołu poczqł odpoczywajacem* RP177/3–4<sup>4</sup> – lemat imiesłowu czasu terażniejszego czynnego odmiennego *odpoczywajacem*: ODPOCZYWAĆ; *Dziewica Maryja, matuchna jego, kiedy to usłyszała, wszytka, zarzewniwszy sie dla silnego smętku, obumarła, tak iż przez dlugi czas mowić nie mogła* ŻPJK47v/17–19 – lemat imiesłowu czasu przeszłego

4 Wszystkie przykłady cytujemy za stroną Apocrypha. Teksty lokalizujemy w następujący sposób: po skrócie tekstu następuje numer karty, a po ukośniku numer wersu/wersów.

czynnego *zarzewniwszy*: ZARZEWNIC). W wypadku tych klas najwyraźniej widać różnicę w traktowaniu lematu – nie jest on formą podstawową, a jedynie jednostką słownikową.

7. Dla przyimka, spójnika, partykuły, interiekcji, zgramatyzowanego SIE i aglutynantu trybu przypuszczającego BY przypisany lemat to forma tekstowa.
8. W zgromadzonym materiale staropolskim pojawiło się również wiele wyrazów obcych, które stanowią odrębną klasę w narzędziu Apocrypha. Są to przede wszystkim nazwy własne niezaadaptowane fleksyjnie do polszczyzny (np. *Apokalypsim*) oraz wyrazy obcojęzyczne (np. *cetera*, *beth*). Klasa ta nie ma przypisanych kategorii gramatycznych, jedynie lemat równy formie tekstowej. Z tego powodu na przykład *Deus* i *Deo* to osobne lematy.

## 2. Lematyzacja form dyskusyjnych – problemy i rozwiązania

O tym, że (mimo ustalonych zasad i – wydawałoby się – jasnych kryteriów) lematyzacja tekstów dawnych nie jest zadaniem łatwym, świadczą wątpliwości autorów największego leksykalnego przedsięwzięcia związanego ze staropolszczyzną, to jest SStp. Autorzy *Indeksów do Słownika staropolskiego* zauważyli, że przez lata powstawania SStp zmieniło się podejście do wyboru form hasłowych (Eder, Twardzik 2007: VII–XIV). W projekcie, jeśli to tylko było możliwe, o przypisaniu lematu decydowała forma fleksyjna (np. o *zwadzie braciej* RP111/19 – lemat BRACIA, rodzaj żeński) lub otoczenie składniowe (np. *Bracia moja namilejsza* RP27/12–13 – lemat BRACIA; *O bracia moi Żydowie* ŻPJK94v/9–10 – lemat BRAT). W naszej pracy decyzje często były podejmowane już na etapie tworzenia transkrypcji, a nie anotacji gramatycznej (np. zachowanie w transkrypcji wariantów *cztery* i *czterzy* skutkowało przyjęciem dwóch osobnych lematów CZTERY i CZTERZY). Problem z wyborem lematu zauważamy natomiast przede wszystkim wtedy, gdy jedna konkretna forma tekstowa może pochodzić od dwóch (lub więcej) różnych lematów (najczęściej od różnych wariantów fonetycznych). O ile wyrazy takie jak *namiestkom/namiestkom* nie stanowią problemu (tu każda forma gramatyczna będzie wprost wskazywać na podstawę – NAMIASTEK lub NAMIESTEK), o tyle w wypadku na przykład formy *przeszli* nie wiemy, czy przypisać lemat PRZEJŚĆ, PRZEJĆ czy PRZEĆ. Wówczas kierujemy się kilkoma przyjętymi regułami, które postaramy się jasno wskazać.

Pierwszą wskazówką jest frekwencja w tekście lub w tekstach. Widać to bardzo wyraźnie przy lematyzacji przysłówków w stopniu porównawczym. Na przykład dla formy tekstowej *narychlej* (poświadczonej dwukrotnie: *Przez ty rzeczy człowiek **narychlej** może dostąpić serdecznej czystości* ŻPJK23r/16–17; *Abowiem Joachim, wierząc powiedaniu anjelskiemu i ukazując wdzięczność Panu Bogu przez ofiary nabożne, co **narychlej**, oglądawszy swoje bydło, kwapił sie z s<w>oimi do Jerusalem i do imienia*



swego ŹSA<sub>11v/23–28</sub>) lematem mogłoby być RYCHŁO lub RYCHLE. Podobnie jest w wypadku formy tekstowej *rychlej* (poświadczonej 66 razy, np. *Co masz czynić, to czyń rychlej* RP<sub>552/20–21</sub>). Forma stopnia równego *rychło* pojawia się w apokryfach 70 razy, w tym 25 razy w RP, 20 razy w ŹPJK, 10 razy w SCh, 6 razy w HTK, 5 razy w ŹSA i po 2 razy w RD i EN. Forma stopnia równego *rychle* pojawia się tylko raz w RD. To pozwala przyjąć, że lematem dla formy tekstowej *narychlej* i *rychlej* jest RYCHŁO.

Inaczej jest na przykład z formą *pewniej*. Wystąpiła w tekstach czterokrotnie, dwa razy w RP (RP<sub>671/23–672/1</sub>, RP<sub>766/6–18</sub>), raz w HTK (HTK<sub>255v/13–256r/4</sub>), raz w SCh (SCh<sub>126v/3–6</sub>). Forma stopnia równego *pewnie* wystąpiła w tekstach 16 razy (7 razy w SCh, 4 razy w RP, 3 razy w ŹPJK, raz w ŹSA i raz w RD). Forma stopnia równego *pewno* wystąpiła w tekstach 10 razy, w tym 9 razy w RP i raz w ŹPJK. Z tego względu, mimo że forma *pewnie* jest częstsza, dla RP przyjmujemy lemat PEWNO (frekwencja 9 do 4). Dla SCh jedyną poświadczoną formą jest *pewnie*, więc taki lemat przyjmujemy dla stopnia porównawczego. Natomiast w HTK w ogóle nie poświadczono formy stopnia równego. W takich sytuacjach kierujemy się frekwencją we wszystkich apokryfach.

O repartycji lematów decyduje zatem to, czy dana forma podstawowa w ogóle pojawiła się w tekście lub jej frekwencja (w danym tekście lub we wszystkich apokryfach). I tak na przykład dla form tekstowych *krześcijani* (*My, krześcijani, jesteśmy w jego Kościele z osobności łaski jego* ŹPJK<sub>35r/7</sub>) i *krześcijanie* (*A przeto krześcijanie łacińskim językiem, ktorzy sie mają podług obyczaju rzymskiej Cyrekwie, działają i sprawiąją Boskie urzędy* HTK<sub>228v/11–13</sub>) przyjęliśmy osobne lematy (KRZEŚCIJAN i KRZEŚCIJANIN) wobec poświadczenia obu form w mianowniku liczby pojedynczej (*Tu naucza miły Jesus, jako ma każdy krześcijan dostojny jeden drugiego miłować* RP<sub>270/9–11</sub>; *Nie tylko ten przy Krysta, jen przy, aby on był krystus, ale i ten sie krześcijanin być przy* RP<sub>689/11–15</sub>). Inaczej jednak postąpiliśmy w wypadku analogicznych formacji słotwórczych *poganin*, *mieszczanin*. Tych apokryfy nie notują w krótszej formie mianownika liczby pojedynczej \**pogan*, \**mieszczan* (formy te wystąpiły wyłącznie jako formy dopełniacza liczby mnogiej, por. <od> *żydowskiej ziemie do pogan* RP<sub>463/11–12</sub>; *przeżegnanie wszystkich dziatek, sług i dzieweczek, i mieszczan, i zwierząt* HTK<sub>217v/5–7</sub>), zatem ich lematy przybrały postać POGANIN, MIESZCZANIN – w wypadku pierwszego wyrazu poświadczoną także licznymi formami mianownika liczby pojedynczej, na przykład *iz był Piłat poganin* RD<sub>164/7</sub>; *aczkołe był poganin i przez zakonu* HTK<sub>252r/12–13</sub>.

W sytuacjach nierozstrzygniętych (brak formy podstawowej w tekście) decydująca może też być trwałość formy – wybieramy leksem funkcjonujący współcześnie, by ułatwić wyszukiwanie. Tak będzie na przykład z formą tekstową *przejdą* (*Zawierne to wam powiedam, iże jawni grzesznicy i grzeszne żony przejdą was w krolewstwie niebieskim* RP<sub>454/19–21</sub>), gdy lematem mogłoby być PRZEJŚĆ lub PRZEJĆ, wybieramy współczesne PRZEJŚĆ. Dla imiesłowu *bojąc* wybierzemy lemat BAĆ, nie BOJEĆ

czy BOJAĆ. Tu zresztą pomaga też frekwencja występowania formy bezokolicznika w tekstach (*bać* – 13 razy, *bojeć* – raz, *bojać* – raz).

W sytuacji nakładania się na siebie tych dwóch kryteriów rozstrzygające jest dla nas kryterium frekwencji. Tak postąpiliśmy na przykład z wyborem lematów CUD, CUDO. Teksty apokryfów notują w mianowniku i bierniku niemal wyłącznie formę *cudo*<sup>5</sup> (*nowe sie cudo stało* RP65/2; *to cudo o wkrzeszeniu Łazarzowem* RP173/4; *by im ukazał swe cudo* RP30/16). Dlatego przyjęliśmy lemat CUDO, z wyjątkiem nielicznych użyć, w których forma zależna sugeruje deklinację męską *cud* (por. *aż nowina tego cudu do krola przyszła* ŻSA30r/19; *zjawić temu światu cudy* RP207/9–10, dla obu przykładów lemat CUD). Nietrudno również było rozstrzygnąć wybór spośród lematów DZIW, DZIWO. Mimo iż SStp notuje wariant *dziwo*, badane teksty apokryficzne nie poświadczają użycia, które mogłoby sugerować mianownik rodzaju nijakiego *dziwo*, zaś liczne wyraziste przykłady (*a taki dziw nasilnie ustrasz* RP93/11–12; *Ryby morskie, dziwy przeliczne w morzu* RP22/5; *wiele dziwow uczynił* RP43/1) nakazywały przyjęcie lematu DZIW<sup>6</sup>.

Niektóre wyrazy staropolskie nie były w ogóle notowane przez SStp i SPXVI, w tym warianty leksemów, formy gramatyczne czy znaczenia, co ma związek z faktem, że nie wszystkie apokryfy wykorzystano jako źródła wymienionych prac leksykograficznych. W takich sytuacjach, aby ustalić lemat, potrzebna była decyzja odgórna, poparta badaniami leksykograficznymi, etymologicznymi, źródłowymi, na przykład *a na siedlisku onego mieśca, przed oną paciepią, chleb i insze rzeczy przedawany* HTK168r/3–11 – lemat PACIEPIA; *przy padzie, na kogokoli była [włó] włożona, zategoż sie stał zdrow* HTK158r/11–158v/1 – lemat PAD; *gdy ręce Jesukrystowy, ktoreż wszytek świat stworzyły, dary rozdawały, goźdzmi były przybite, krwią okropnie, co sie stanie nad grzesznymi?* SCh104v/2–7 – lemat KRWIĆ; *ale już byli bardzo starzy a skluczeni albo zgarbieni* HTK210v/19–211r/10 – lemat SKLUCZYĆ. Prace nad lematyzacją apokryfów poskutkowały wskazaniem uzupełnień hasłowych do słowników historycznych języka polskiego (Nowak-Pasterska, Rojszczak-Robińska, Stelmach et al. 2023).

Nie tylko wobec nieobecności hasła w słownikach potrzebna była decyzja odgórna (spis takich decyzji po zamknięciu anotacji czytelnik znajdzie w zakładce *Materiały dodatkowe*). Gdy spełnione są wyżej wskazane warunki, staramy się różnicować znaczenia. Tak jest w wypadku czasownika CZYŚĆ ('czytać') i CZCIĆ ('czytać', ale też 'wielbić'). Bezokolicznik *czyść* występuje 9 razy (w ŻPJK, SCh, RP i HTK), bezokolicznik *czcić* 11 razy, ale tylko dwukrotnie w znaczeniu 'czytać' (jedynie w RP). Z tego względu dla form tekstowych na przykład *czcili* w znaczeniu 'czytali' przyjęliśmy lemat CZYŚĆ, a w znaczeniu 'wielbili' – CZCIĆ.

5 Forma mianownika liczby pojedynczej *cud* nie jest poświadczona.

6 SPXVI (VI: 531) notuje DZIWO wyłącznie za SStp.



### 3. Lematyzacja wybranych klas – zastosowane rozwiązania

Pomimo przyjęcia dla każdej z klas jasnych zasad lematyzacji i ustalenia hierarchii kryteriów opisanych powyżej niektóre klasy okazały się nieco kłopotliwe. Poniżej prezentujemy te, w których specyfika staropolszczyzny wymusiła pewne odgórne rozwiązania.

#### 3.1. Zaimki rzeczowne

W wypadku zaimków rzeczownych trzecioosobowych zastanawialiśmy się nad przyjęciem jednego lematu ON dla zaimków *on, ona, ono, oni, one*. Ponieważ pod względem gramatycznym zaimek rzeczowny odpowiada rzeczownikowi, czyli ma rodzaj, a nie odmienia się przez rodzaj, przyjęliśmy osobne lematy ON, ONA, ONO, ONI, ONE.

Pewną innowacją w toku wypracowywania przez nas zasad lematyzacji okazało się przyjęcie lematu **ONY** dla form zaimka osobowego w liczbie mnogiej rodzaju męskiego, poświadczonych w bierniku w formie nietypowej dla paradygmatu męskoosobowego, na przykład *Dziewica matka miłościwa żałobnie weźrawszy na nie i rzekła jem* (RP506/5–7). Dopełnienie wyrażono formą biernika *nie (weźrawszy na nie)* i celownika *jem (rzekła jem)*, które odnoszą się do wcześniej wskazanych apostołów (w zdaniu poprzednim: *A ku ktoremu płaczu dziewice Maryje apostołowie poszli* RP505/21–22). Trudno było je zaanotować jako rodzaj niemęski, skoro zaimki rzeczowne, podobnie jak rzeczowniki, mają rodzaj (np. rzeczownik *dziewica* funkcjonuje w rodzaju męskim i żeńskim), a paradygmaty zaimkowe dopiero się kształtują (słowniki nie wskazują, które formy należą do paradygmatu odmiany danego zaimka, łącząc wszystkie osobowe dla 3. osoby w jednym haśle). Mając zatem to na uwadze, w anotacji tego typu form zdecydowaliśmy się na pewnego rodzaju „kompromisowy” lemat ONY, będący odpowiednikiem wyraźnie męskoosobowego zaimka *oni*, lecz realizowany nierzadko jeszcze w postaci niemęskich form *je, nie*. Obok tego typu charakterystycznych użyć lemat ONY funkcjonuje także jako regularny lemat dla form niemęskiego zaimka osobowego *one*, poświadczonego w formie mianownika w postaci *ony*, na przykład *Tedy ony, tamo będąc, tkały pawłoki* RP16/10–11 (zaimek *ony* ma lemat ONY).

Nieco problematyczne w przypisaniu jednoznacznego lematu w procesie anotacji okazały się formy zaimka *się, siebie, sobie, sobą*<sup>7</sup>. Jednym z naczelnych kryteriów rządzących wyborem odpowiedniego lematu dla danego wyrazu było przyjęcie formy mianownika jako lematu. Tradycyjnie w opracowaniach historycznogramatycznych

7 Odrębną klasę tworzy tzw. zgramatyzalizowane SIE, które może być morfemem słowotwórczym albo wykładnikiem zwrotności lub bezosobowości, stojące przy czasownikach, niemające już charakteru zaimkowego, a w anotacji opatrzone lematem SIE, SIE lub SI.

nie wymienia się formy mianownika w modelu odmiany zaimka *się*. Stojąc wobec defektywności paradygmatu i jednocześnie konieczności nadania lematu każdemu anotowanemu wyrazowi, zdecydowaliśmy się na przypisanie etykiety JA, identycznej z lematem form pierwszej osoby liczby pojedynczej zaimka rzeczownego, na przykład *anjeli ich ciała na **sie** wzięli* RP368/19; *jął wielkie udręczenie mieć nie telko sam o **sie**, ale o swą oblubienicę Maryją* RP54/8–10, gdzie *sie* jako forma biernika liczby pojedynczej zaimka rzeczownego ma lemat JA.

### 3.2. Zaimki przymiotne

Zgodnie z przyjętymi przez nas zasadami anotacji gramatycznej wszystkie zaimki funkcyjnie dzierżawcze zostały zaszerogowane do klasy zaimków przymiotnych (por. Deskur, Nowak-Pasterska, Rojszczak-Robińska et al. 2022: 167), co ma swoje odzwierciedlenie także w lemacie każdego zaimka. Zaimki odmienne dla pierwszej i drugiej osoby liczby pojedynczej i mnogiej nie stanowiły problemu, przyjęły one lemat w rodzaju męskim, zgodnie z zasadą przypisywania lematu wyrazom, dla których rodzaj jest kategorią gramatyczną (np. klasa przymiotników, zaimków przymiotnych, liczebników porządkowych), na przykład *I wszyscy, którzy w **naszej** mocy i w mękach byli* EN145r/3–4 (zaimek w formie miejscownika liczby pojedynczej rodzaju żeńskiego *naszej* ma lemat NASZ). Natomiast zaimki dzierżawcze trzeciej osoby liczby pojedynczej i mnogiej *jego, jej, ich* obsługiwane przez formy dopełniacza zaimków osobowych zachowują lematy JEGO, JEJ, ICH, na przykład *jako ja chował każń Boga, ojca mego, i jestem w **jego** miłości* RP568/2–4 (zaimek przymiotny *jego* ma lemat JEGO), *aż łotry przywiązano ku **ich** krzyżom* RD198/23 (zaimek przymiotny *ich* ma lemat ICH). Decyzję tę opisywaliśmy dokładnie w artykule o potrzebie zastosowania nowych wartości kategorii (Rojszczak-Robińska, Wismont, Zalejarz et al. 2023: 107).

### 3.3. Rzeczowniki

W wypadku rzeczowników problematyczne okazały się rzeczowniki dwurodzajowe oraz te, które zmieniły rodzaj. Słowniki polszczyzny dawnej notują na przykład różne postaci słowa *ciern* (por. SStp I: 297–298), współcześnie utrwalonego jako wyraz rodzaju męskiego. Teksty apokryfów poświadczają liczne formy zależne, synkretyczne dla rodzaju męskiego i nijakiego, na przykład *była uczyniona z **ciernia** morskiego* RP829/1–2; *bo takiego ostrego **cirznia** nie było* RD152/12–13; *Azali będziecie zbierać z **cirnia** winne jagody* RP283/13–14; *ten, który siał między **cirniem*** RP338/7. Choć nie sposób odnaleźć w badanych apokryfach formy mianownika *ciern*, *cirń* czy *cirzn*, zdecydowaliśmy się na wybór lematu w rodzaju męskim, zgodny z formą współczesną CIERNŃ (uwzględniając także dawne warianty fonetyczne CIRŃ bądź

CIRZŃ), a tylko w sytuacji gdy dana forma wyraźnie wskazywała na mianownik lub biernik rodzaju nijakiego, dopuściliśmy lematy: CIRZNIE (np. *gwałtem szło **cirznie** w głowę Jezusowę* RP154/12), CIRNIE (np. *a **cirnie** zrostszy i zagłuszyło nasienie* RP335/18–19). Podobnie kierowaliśmy się zasadą zgodności ze współczesną formą wyrazu, decydując o postaci lematów: KAMIEN, KAMIENIE. Dla wszystkich form niejednoznacznych przyjęto etykietę KAMIEN (np. *złotem, drogim **kamieniem** i perłami jesteś okraszony* ŻPJK109v/7–8), a wyłącznie formom rodzaju nijakiego, które wystąpiły w mianowniku (*rzeczy, aby to **kamienie** było chlebem* ŻPJK23r/30) przypisano lemat KAMIENIE.

Dla form supletywnych rzeczowników zachowaliśmy odrębne lematy. I tak na przykład rzeczowniki *człowiek* i *ludzie* mają odpowiednio lematy CZŁOWIEK i LUDZIE (np. *A Jozef jako prawdziwy **człowiek** natychmiast był posłuszen* RP32/14–16 – lemat CZŁOWIEK; *O niewdzięczni **ludzie**, ślepi i szaleni* ŻPJK83v/1–2 – lemat LUDZIE). Dla form *rok*, *lata* przyjęliśmy lematy ROK i LATO (np. *a tak cały **rok** tajemnie nauczał* ŻPJK24v/6 – lemat ROK; *i trzy **lata** już była popełniła* RP13/9 – lemat LATO). Notabene lemat ROK poświadczony został także w liczbie mnogiej (*a ta dwa, czuż Annasz a Kajfasz, na **roki** trzymali biskupstwo* RP667/8–10).

#### 4. Granice i formy lematu

W tekstach apokryficznych nie brak rozmaitych wyrazów zrosniętych z różnymi pojedynczymi morfemami – partykułami, aglutynantami czy szczątkowymi postaciami zaimków osobowych. Chcąc dokonać rzetelnej anotacji, w której nie pomija się takiej zlepkowej struktury niektórych elementów języka, zdecydowaliśmy o wydzieleniu tych ostatnich i opatrzeniu ich osobnymi lematami. Dzięki temu procesowi nie zatraciliśmy w anotacji licznych części „sklejonych” z różnymi klasami, między innymi partykuł *li*, *ż*, *ci*, *ć*, na przykład *O błogosławiona dziewico, **użrzęli** ja tve lice* RP48/18–19 (forma *użrzęli* to dwa „wyrazy” i dwa różne lematy: forma przyszła prosta pierwszej osoby liczby pojedynczej *użrzę* mająca lemat UŻRZEĆ i partykuła *li* mająca lemat LI); *przez **jegoż** moc wierzę czystotę moję przez wszego pokalania, dokąd jestem żywa, chować* RP61/14–16 (analogicznie do poprzedniego przykładu – zaimek *jegoż* to dwa wyrazy i dwa lematy: zaimek przymiotny *jego* mający lemat JEGO oraz partykuła *ż* mająca lemat Ż); *Tać będzie poświęcona w żywocie matki swej* RP3/23–24 (zaimek *Tać* to dwa różne wyrazy i dwa lematy: zaimek przymiotny *ta* mający lemat TEN oraz partykuła *ć* mająca lemat Ć); *Tenci nie widzi sie nam, by to twój syn był, boć on był nad słońce świetlejszy, a **tenci** trędownaty, ze wszech naszkaradszy* ŻPJK90r/17–19 (zaimek *tenci* wzmocniony partykułą *to* dwa odrębne lematy: TEN, CI)<sup>8</sup>.

8 Zważywszy na fakt, że lemat jest jednym z kluczowych narzędzi do przeszukiwania tekstów na stronie Apocrypha po określonych słowach, należy dodać, że możliwy jest także ogląd tego typu niesamodzielných wyrazów za pomocą odpowiedniej wyszukiwarki.

Dwoma odrębnymi lematami zostały przez nas opatrzone również przyimki połączone z *-ń* będącym śladem dawnej formy zaimka osobowego w bierniku *ji* z poprzedzającym go *n* epentetycznym, powstałym jeszcze w prasłowiańszczyźnie z wyrażen przyimkowych z zaimkami typu *\*wǫn jimь*, w obrębie których dokonały się procesy przesunięć morfologicznych – np. *Niewiasta, usłyszyc to, padła przedeń i poznała, iże jej niemoc albo wrzod był uzdrowion* RP223/12–15 (forma *przedeń* po podziale *przedeń* posiada dwa lematy: lemat przyimka PRZEDE oraz lemat zaimka osobowego ON); *sercem pewnie wierzy weń* Sch106v/16 (podobnie – *weń* ma odpowiednie lematy: WE, ON).

Problemem okazało się rozstrzygnięcie, kiedy dana partykuła jest wciąż samodzielny morfemem, mogący rozmaicie doklejać się do różnych części mowy, a kiedy częścią zaabsorbowaną morfologicznie na przykład przez spójnik. W podejmowaniu takich decyzji kierowaliśmy się zasobem SStp i SPXVI. Dla przykładu spójnik *boć* (por. przywołany powyżej przykład: *Tenci nie widzi sie nam, by to twój syn był, boć on był nad słońce świetlejszy*) według słowników może już być w pełni zintegrowany z towarzyszącą mu pierwotnie partykułą, ma zatem lemat BOĆ. Tak samo traktujemy często występujące spójniki, na przykład *gdyż*, *kiedyż*, i zaimki *ktoryż*, *jenże*, *coż*.

Stojąc wobec trudności w wyznaczeniu prawidłowej granicy między morfemami i tym samym określeniu postaci lematu, musieliśmy przyjąć pewne kryteria podziału wyrazu. Jednym z nich była frekwencja danej formy w tekstach apokryfów. Kryterium to pozwoliło na wydzielenie uzasadnionych granic podziału dla wahających się w staropolszczyźnie form spójników, na przykład *iż*, *iże*, *aż*, *aże*, występujących w połączeniu z aglutynantami. I tak na przykład repartycja *iż* (163), *iże* (1060) w RP, z decydującą przewagą na korzyść drugiego, przesądziła o granicy podziału morfemów w następujący sposób: *A ja tobie mówię, iżeś ty Piotr* RP363/5 – *iżeś*, lematy: IŻE, BYĆ; *abyście poznali a wierzyli, iżem ja w mem Ocsu a Ociec we mnie* RP430/4–5 – *iże|m*, lematy: IŻE, BYĆ. Tymczasem granica podziału morfemów w RD, wykazujących odwrotną repartycję tych spójników, ze znaczącą przewagą *iż* (188) nad *iże* (14), zdecydowała o tym, że w przykładach takich jak: *A iżem dobrze rzekł, a przeto i czemu mie bijesz* RD86/17–18; *Oto cię oskarżono, iżeś sie czynił krolem* RD166/11–12 podział przebiega w następujących miejscach: *iż|em*, *iż|eś*, zaś lematy mają postać: IŻ, BYĆ. Znacząca przewaga *aż* (99) nad *aże* (6) w apokryfach z kodeksu Wawrzyńca z Łaska (EN, HTK, SCh) również zaważyła na decyzji o wydzieleniu lematu AŻ dla form, w których spójnik jest połączony z wykładnikiem osoby, na przykład *a widząc, jestem sie silno ulękl, ażem wezdrzał, ażem upadł na ziemię* EN278r/16–17 – *aż|em*, lematy: AŻ, BYĆ.

W określonych, specyficznych sytuacjach przyjęliśmy lematy nieciągłe. Rozwiązanie to okazało się konieczne dla wyrażen przyimkowych złożonych z przyimka i przysłówka zachowującego szczątkową odmianę. Poszczególne składniki nie mogły zostać zanotowane samodzielnie z uwagi na trudność w przypisaniu im określonych

wartości kategorii gramatycznych, a także w nadaniu jednoznacznej słownikowej etykiety – lematu, na przykład w *tamtych wszystkich stronach zamorskich w mieściech i we wsiech z stara dawna było i jeszcze jest w obyczaju* HTK167r/1–3. Dla wyrażenia z *stara dawna* przyjęto lemat nieciągły Z STARA DAWNA. Przykłady bez przyimka z podobnie opatrzone są lematem nieciągłym, na przykład *chwalebne miasto, które stara dawna zwano Agrypina* HTK259r/6–7 (wyrażenie *stara dawna* ma lemat STARA DAWNA). Łącznie zaanotowano 13 lematów nieciągłych, na przykład *na każdym z osobna gorzkie łzy wylewała* ŻPK54r/30–31 (lemat Z OSOBNA); *uźrał Abrama z daleka a Łazarza na jego łonie siedząc* RP395/15–16 (lemat Z DALEKA); *Potym pożywając rozmowy spolnej z dawna pożądaney o zwiastowaniu anjelskim, weselili sie społem* ŻSA12v/5–7 (lemat Z DAWNA); *Ale Jesus nie tako, po lekku, wyznał mowić* RP99/16–17 (lemat PO LEKKU).

Inne tego typu rozbudowane okoliczniki wyrażone różnymi klasami, gdy tylko możliwe było zaanotowanie każdego składnika wyrażenia, podlegały typowej anotacji, nie wymagały zatem wprowadzania lematu nieciągłego, na przykład *A otworzywszy zamek i znamiona odpieczętowali w zarań rano* EN129r/2–4 (w anotacji wyrażenia *w zarań* przyimek opatrzone jest lematem W, a przysłówek *zarań* lematem ZARAŃ); *A dlatego z mała srebra, które <z> sobą przyniosł był, klenoty, pasy srebrne, kubki i inne rzeczy przedawszy, dał sobie wyryć obraz świętej Anny* ŻSA28r/6–9 (elementy wyrażenia z *mała* mają lematy: Z, MAŁA); *nazwano jest po grecku Likostratus, a po żydowsku Gabota* SCh91r/3–4 (wyrażenia *po grecku*, *po żydowsku* mają odrębne lematy: PO, GRECKI, PO, ŻYDOWSKI).

## 5. Podsumowanie

Lematyzacja tekstów dawnych nie jest sprawą łatwą. W artykule wskazaliśmy przede wszystkim te klasy lub formy, które mogą być dyskusyjne, trudne. Lematyzacja zawsze wiąże się z podjęciem konkretnych, sprecyzowanych decyzji, rozstrzygnięciem zaobserwowanych problemów. Przygotowanie tak istotnych dzieł jak SStp i SPXVI potwierdza jej niełatwy i złożony charakter. W narzędziu Apocrypha z jednej strony starano się zachować właściwości anotowanych tekstów i ich języka (między innymi uwzględniono warianty wyrazów, specyfikę form aglutynowanych, wyrażenia wymagające wprowadzenia lematów nieciągłych), a z drugiej nie utrudniać przeszukiwania materiału (między innymi dzięki odpowiednim rozwiązaniom technicznym, jak np. wyświetlanie podpowiedzi wariantów).

## Literatura

- APOCRYPHA: D. Rojszczak-Robińska, O. Ziółkowska, D. Maslej et al. (red.), *Staropolskie apokryfy Nowego Testamentu*, [on-line:] [apocrypha.amu.edu.pl](http://apocrypha.amu.edu.pl).
- DESKUR A., NOWAK-PASTERKA E., ROJSZCZAK-ROBIŃSKA D. et al., 2022, *Anotacja gramatyczna tekstu staropolskiego – problemy i wyzwania*, „LingVaria” XVII, nr 2 (34), s. 163–180. DOI: 10.12797/LV.17.2022.34.11.
- EDER M., TWARDZIK W., 2007, *Indeksy do Słownika staropolskiego*, Kraków.
- GRUSZCZYŃSKI W., BRONIKOWSKA R., 2024, *Instrukcja korzystania z wyszukiwarki do Elektronicznego Korpusu Tekstów Polskich z XVII i XVIII wieku*, [on-line:] <https://korba.edu.pl/manual>.
- NOWAK-PASTERKA E., ROJSZCZAK-ROBIŃSKA D., STELMACH W. et al., 2023, *Kodeks Wawrzyńca z Łaska jako źródło słowników historycznych*, [w:] D. Rojszczak-Robińska, A. Deskur, W. Stelmach (red.), *Język staropolskich apokryfów. Pytania, problemy, perspektywy*, Poznań, s. 159–194.
- PRZEPIÓRKOWSKI A., BAŃKO M., GÓRSKI R.L. et al. (red.), 2012, *Narodowy Korpus Języka Polskiego*, Warszawa.
- ROJSZCZAK-ROBIŃSKA D., WISMONT M., ZALEJARZ K. et al., 2023, *Specyfika staropolszczyzny a anotacja gramatyczna. O potrzebie nowych wartości kategorii*, „LingVaria” XVIII, nr 1 (35), s. 99–111. DOI: 10.12797/LV.18.2023.35.07.
- SPXVI: *Słownik polszczyzny XVI wieku*, t. I–IV, red. komitet redakcyjny, t. V–XVII, red. M.R. Mayenowa, t. XVIII–XXXIV, red. F. Peplowski, t. XXXV–XXXVI, red. K. Mrowcewicz, P. Potoniec, Wrocław – Warszawa – Kraków 1966–2012.
- SStp: S. Urbańczyk (red.), *Słownik staropolski*, t. I–XI, Wrocław 1953–2002.
- STELMACH W., WISMONT M., ZALEJARZ K. et al., 2023, *Specyfika staropolszczyzny a anotacja gramatyczna. Kategorie czasu i trybu*, „LingVaria” XVIII, nr 2 (36), s. 133–149. DOI: 10.12797/LV.18.2023.36.09.

### Specific Features of Old Polish Language and Grammatical Annotation: Lemmatization of Old Polish Texts Abstract

The article is dedicated to grammatical annotation of Old Polish texts. It discusses the problem with the lemmatization of a medieval text. In the *Apocrypha* tool, a lemma is not the same as the basic form of a word. It is only a dictionary form, or a kind of a label. The author of the article presents the reasons for this decision and the adopted rules for assigning lemmas to different grammatical classes. The article also shows the criteria of the repartition of lemmas. These criteria are the frequency of the basic form in the text or texts, the persistence of the form or the differentiation of meanings. Third person pronouns turned out to be particularly difficult for lemmatization. According to the accepted rules, they have a gender, but do not conjugate with a gender (which forced, for example, the adoption of the masculine lemma ONY). The other examples are forms such as *siebie*, *sobie* [*self*], for which the lemma JA was adopted. The article also describes the issues of determining the lemma boundary when annotating words conjugated with different single morphemes: particles, agglutinants or residual forms of personal pronouns. All of the solutions adopted in the tool and presented in the article aimed at respecting the Old Polish language specific features while facilitating the analysis of the material.