


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

Aleksandra Deskur   
Uniwersytet im. Adama Mickiewicza, Poznań  
aleksandra.deskur@amu.edu.pl

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

## SPECYFIKA STAROPOLSZCZYZNY A ANOTACJA GRAMATYCZNA. O POTRZEBIE NOWYCH WARTOŚCI KATEGORII<sup>1</sup>

**Słowa kluczowe:** anotacja gramatyczna, kategorie gramatyczne, średniowieczna polszczyzna, staropolskie apokryfy Nowego Testamentu

**Keywords:** grammatical annotation, grammatical categories, medieval Polish language, Old Polish New Testament apocrypha

Badania nad dziewięcioma polskimi apokryfami Nowego Testamentu<sup>2</sup> oraz stworzenie narzędzia ułatwiającego ich wieloaspektową analizę to główne założenia

- 
- 1 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 Doroty Rojszczak-Robińskiej. Udział współautorów jest równy i wynosi po 20%.
  - 2 Staropolskie apokryfy Nowego Testamentu to siedem tekstów zachowanych w rękopisach, tj. *Rozmyślanie przemyskie* (dalej: RP), *Sprawa chędogo o męce Pana Chrystusowej* (dalej: SCh),

projektu „Początki języka polskiego i kultury religijnej w świetle średniowiecznych apokryfów Nowego Testamentu. Uniwersalne narzędzie do badań polskich tekstów apokryficznych”. Jest on realizowany w ramach grantu Narodowego Centrum Nauki na Wydziale Filologii Polskiej i Klasycznej UAM od marca 2018 r. Kompleksowe narzędzie internetowe zaprojektowano z myślą o przedstawicielach różnych dyscyplin zainteresowanych nie tylko średniowieczną polszczyzną, lecz także duchowością czy kulturą dawną.

Na stronie *Staropolskie apokryfy Nowego Testamentu* (apocrypha.amu.edu.pl, dalej też *Apocrypha*) znajdują się teksty wspomnianych apokryfów w sporządzonych na potrzeby projektu transliteracjach i transkrypcjach, przygotowanych według wypracowanych przez zespół rozwiązań edytorskich<sup>3</sup>. Różne funkcjonalności – np. zestawienie miejsc paralelnych, wątków, innych transkrypcji, źródeł, anotacja gramatyczna – umożliwiają wszechstronny i multidyscyplinarny ogląd tych niezwykle ważnych dla dziejów polszczyzny tekstów. Podkreślić należy, że transkrybowane teksty apokryfów (ponad 2000 kart) poddano ręcznej anotacji gramatycznej. Każdy wyraz tekstowy został zinterpretowany, tj. przydzielony do konkretnej klasy gramatycznej, a następnie przypisano mu wartości kategorii gramatycznych właściwych danej klasie oraz lemat. Użytkownik ma możliwość odszukania konkretnej formy gramatycznej we wszystkich dziewięciu apokryfach, niezależnie od semantyki wyrazu.

Funkcjonalności narzędzia zaplanowane dla językoznawców przybliżono w artykule *Anotacja gramatyczna tekstu staropolskiego – problemy i wyzwania* (Deskur et al. 2022). Przedstawiono tam również siatkę klas i kategorii z uwzględnieniem rozmaitych problemów dotyczących anotacji tekstu dawnego. Celem niniejszej publikacji jest omówienie jednego z nich, tj. potrzeby wprowadzenia nowych wartości kategorii (0 i X) przy anotacji tekstu staropolskiego.

Anotacja tekstów dawnych wymagała nie tylko zaprojektowania siatki klas i kategorii, ale także uwzględnienia sytuacji, kiedy dana kategoria nie dotyczy anowanej formy lub nie można jej ustalić z powodu różnych czynników wewnętrznych i zewnętrznych, np. uszkodzenie tekstu, forma niepozwalająca na określenie, brak kontekstu. W związku z tym zdecydowano się przyjąć rozwiązanie polegające na wprowadzeniu kategorii 0 i X, które oznaczają odpowiednio: 0 – nie dotyczy, X – nie można określić. Zaplanowanie siatki klas i kategorii ma bezpośrednie przełożenie na precyzyjność filtrów w wyszukiwarce gramatycznej i leksykalnej. Dzięki nim użytkownik narzędzia zyskuje możliwość znalezienia określonej części mowy,

---

*Rozmyślenia dominikańskie* (dalej: RD), *Karta Rogawskiego* (dalej: KR), *List Lentulusa* (dalej: LL), *Ewangelia Nikodema* (dalej: EN) i *Historija Trzech Kroli* (dalej: HTK), oraz dwa druki, tj. *Żywot Pana Jezusa Krysta Baltazara Opeca* (dalej: ŻPJK) i *Żywot świętej Anny* (dalej: ŻSA). Więcej zob. apocrypha.amu.edu.pl/elabo.

3 Zob. *O projekcie, Przyjęte rozwiązania edytorskie* na stronie apocrypha.amu.edu.pl.

na przykład rzeczowników, może też wyszukać konkretną formę gramatyczną, na przykład przymiotniki liczby pojedynczej rodzaju żeńskiego w stopniu równym.

Zastosowanie wartości kategorii 0 i X było rezultatem podjętych na etapie planowania siatki klas i kategorii decyzji o rezygnacji z kompatybilności z innymi korpusowymi projektami i dopasowaniu tej siatki do specyfiki materiału staropolskiego<sup>4</sup>.

Konieczność przyjęcia takiego rozwiązania wynikała z kilku czynników różnego typu – wewnątrzjęzykowych i czysto technicznych. Przede wszystkim decydował o tym kształt materiału. Mamy do czynienia z tekstami z etapu intensywnego rozwoju polszczyzny, z przełomu XV i XVI w. Anotujemy teksty wielowarstwowe, w których zachowane są zarówno formy dawne, dużo starsze (zob. Twardzik, Rzepka 1996a, 1996b, 1997a, 1997b, 1997c, 1999), jak i nowsze. W tym czasie pewne kategorie gramatyczne dopiero się kształtują (jak osobowość, zob. Kępińska 2006), inne zmieniają (jak redukcja kategorii liczby przy liczebniku, zob. Siuciak 2008, Słoboda 2012).

Czynnikiem, który wpływa na konieczność zastosowania wartości kategorii 0 i X, jest np. jednoczesne funkcjonowanie dawnej i nowej odmiany w niektórych klasach. Tak jest z liczebnikiem, o którym Agnieszka Słoboda pisała: „Liczebniki główne w średniowieczu wciąż stanowią zróżnicowaną klasę wyrażań, w której stopniowo dochodzi do zmian semantycznych i składniowych” (ibid.: 231). W okresie, z którego pochodzą analizowane teksty, jeszcze nie zakończył się proces przejścia z dawnej odmiany liczebnika w nową, na przykład liczebnik główny w nowej odmianie nie ma już kategorii liczby i rodzaju (anotujemy wartość jako 0), a stara odmiana ma rodzaj żeński. Dla mianownika i biernika nie ma możliwości, by różnić, z jaką odmianą mamy do czynienia. Wówczas wartość kategorii anotowana jest jako X (więcej zob. część *Zastosowanie wartości 0, X w anotacji wybranych klas*).

Innym czynnikiem wpływającym na zastosowanie wartości X jest np. niemożliwość przypisania wartości rodzaju zaimka w określonych przypadkach, wynikająca ze zmian w zakresie rodzaju jako kategorii (zob. część *Zastosowanie wartości 0, X w anotacji wybranych kategorii*).

Zastosowanie wartości 0 i X wynika także z tego, że zrezygnowano z posługiwania się kategorią fleksemów, jak w Elektronicznym Korpusie Tekstów Polskich z XVII i XVIII w. (do 1772 r.), a niektóre klasy zostały wyodrębnione na podstawie kryterium semantycznego (zob. Deskur et al. 2022). W związku z tym na przykład liczebnik krotny czy wieloraki to klasa zróżnicowana, obejmująca zarówno elementy odmienne (*stokrotny, pięćioraki*), jak i nieodmienne (*trzykroć, czworako*). W Elektronicznym Korpusie Tekstów Polskich z XVII i XVIII w. (do 1772 r.) odpowiadają

4 Różnice wynikają przede wszystkim z modyfikacji klas gramatycznych. Na przykład w Elektronicznym Korpusie Tekstów Polskich z XVII i XVIII w. (do 1772 r.) wyróżniono zaimek nieterciosobowy (*ja, ty, my, wy*) i zaimek trzeciosobowy (*on*), co pozwala na przypisanie rodzaju temu ostatniemu. W narzędziu przygotowywanym przez zespół grantowy zaimki rzeczowne stanowią jedną klasę, co wiąże się z koniecznością nadania wartości 0 w kategorii rodzaju zaimkom pierwszej i drugiej osoby, podczas gdy zaimki trzeciosobowe mają określony rodzaj.

im liczebnik przymiotnikowy i przysłówkowy. W narzędziu *Apocrypha* kategorii właściwe są dla całej klasy, konieczne jest zatem przypisanie nieodmiennym leksemom wartości 0. Podobnie jest z zaimkami (więcej zob. *ibid.*).

Innym czynnikiem wymuszającym zastosowanie wartości kategorii 0 i X jest postać materiału. Mamy do czynienia z rękopisami, nierzadko uszkodzonymi, w wypadku druku ŹSA – z jedynym zachowanym egzemplarzem (nie ma z czym porównać fragmentów uszkodzonych). Gdy brakuje kontekstu składniowego, czasami nie możemy określić wartości danej kategorii. Tak jest zarówno w miejscach z uszkodzeniami mechanicznymi (jak np. w KRAa/7, gdzie widać tylko zapis *tych* i brakuje jakiegokolwiek kontekstu, który pomógłby określić przypadek i rodzaj), jak i tam, gdzie w rękopisie niczego nie brakuje, tekst nie jest urwany, a jednak musimy założyć jakąś pomyłkę kopisty, jak tu:

W tym baczycie, iże Annasz z Kajfaszem byli pospołu, kiedy Jesukrysta przywiedziono, bo inych biskupow <...>. (RP713/23–26)

Mimo że rękopis nie jest w tym miejscu uszkodzony, widać, że zdanie podrzędne wprowadzone spójnikiem *bo* nie zostało dokończony. Z tego względu nie ma możliwości pełnej anotacji gramatycznej grupy imiennej *inych biskupow* (nie można określić przypadku).

Jeszcze innym czynnikiem technicznym jest powiązanie anotacji gramatycznej z leksykalną. W trakcie anotacji gramatycznej wyrazom tekstowym lub ich częściom przypisywane są lematy, które nie funkcjonują jak formy podstawowe, a raczej jak wyrazy słownikowe (na przykład dla imiesłówów lematem są formy bezokolicznika)<sup>5</sup>. W sytuacji, gdy tekst jest uszkodzony tak, że nie sposób zdecydować o tym, do której klasy wyraz należy, ale można wskazać lemat, wstawiamy wartość X na poziomie klasy, co blokuje dalsze anotowanie. Nadal możliwe jest wyszukanie wyrazu w indeksie leksykalnym. Taka wartość X nie jest widoczna dla użytkownika.

### Zastosowanie wartości 0, X w anotacji wybranych klas

Poniżej omawiamy te klasy, dla których wartości kategorii 0 i X okazały się szczególnie istotne, zarówno ze względów merytorycznych, jak i narzędziowych. Nie wskazujemy wszystkich możliwych sytuacji, w których dany leksem może na poziomie klasy lub kategorii mieć przypisaną jedną z tych wartości. Jest to bowiem uwarunkowane różnymi czynnikami zarysowanymi w poprzedniej części artykułu.

5 Zob. część Co i jak anotujemy w opracowaniu Deskur et al. 2022: 166–168.

## 1. Liczebnik<sup>6</sup>

Liczebniki zostały wydzielone w procesie anotacji jako odrębne klasy na podstawie przyjętego przez nas kryterium semantyczno-formalnego. O podziale liczebników na określone klasy zdecydowało kryterium semantyczne, zaś zróżnicowanie wewnętrzne poszczególnych liczebników sprawiło, że nie wskazaliśmy dla każdego typu liczebnika zbioru kategorii gramatycznych<sup>7</sup>. Klasa liczebnika okazała się tą, dla której wprowadzenie wartości kategorii 0, X było konieczne.

Znane jest „zróżnicowanie leksemów liczebnikowych pod względem morfologicznym i semantycznym w średniowieczu” (Słoboda 2012: 21), a „opis właściwości semantyczno-składniowych liczebnika jako części mowy, która dopiero w średniowieczu zaczyna kształtować się jako odrębna klasa gramatyczna, musi z tego powodu uwzględniać różne perspektywy oglądu zjawisk językowych” (ibid.: 5). Ze względu na to dla wszystkich klas liczebników zaplanowano te same kategorie, tj. liczby, przypadka, rodzaju.

W anotacji naszego materiału najtrudniejsze pod względem merytorycznym i narzędziowym okazały się kategorie liczby i rodzaju. Przypadek określany jest najczęściej na podstawie formy gramatycznej lub/i kontekstu składniowego.

Dla niektórych klas liczebników przypisanie wartości 0 w kategorii liczby wynika ze specyfiki tej części mowy, w której – jak pisze A. Słoboda – „przemiany zachodzące w semantyce i składni [...] są przejawem procesów wyrównawczych. Towarzyszą im zmiany fleksyjne, takie jak redukcja kategorii liczby” (ibid.).

### 1.1. Liczebnik główny

Liczebniki główne od *jeden* do *cztery* mają kategorię rodzaju przejętą z wchodzących z nimi w związek rzeczowników. Nie mają kategorii liczby. W procesie ich anotacji kategorii liczby przypisana jest bezwyjątkowo wartość 0, np. *trzy dni* – num\_card nom:m TRZY (RP3o4/13), *trzy nocy* – num\_card nom:nm TRZY (RP3o4/13), *trzem mężom* – num\_card dat:m TRZY (EN131r/13)<sup>8</sup>.

Nieco trudniejsza w przypisywaniu określonych wartości kategorii liczby i rodzaju okazała się anotacja liczebników głównych od *pięciu* wzwyż, a to z uwagi na długotrwały w staropolszczyźnie proces przechodzenia tych liczebników z tzw. dawnej odmiany w nową. Zrodziło to potrzebę i konieczność przyjęcia określonych rozwiązań, w wyniku których liczebniki główne od *pięć* do *dziewięć* oraz liczeb-

6 Za pomoc w ustaleniu sposobu klasyfikacji, analizy i anotacji liczebników staropolskich serdecznie dziękujemy profesor A. Słobodzie (UAM).

7 Zob. *Opracowania, Materiały dodatkowe, Opisy klas i kategorii, Liczebnik* na stronie apocrypha.amu.edu.pl (zawartość w przygotowaniu), zob. też Deskur et al. 2022: 167, 170.

8 Wszystkie przykłady cytujemy za stroną *Apocrypha*.

nik *dziesięć* przyjmują wartość liczby i rodzaju w zależności od tego, czy wystąpiły w tekście w nowej czy w starej odmianie:

- w odmianie dawnej liczebniki te przyjmują liczbę pojedynczą i rodzaj żeński, np. *nad pięćmi miast* (RP408/17), *pięćmi pacierzy* (ŻSA39v/2), *po pięćmi miesiący* (ŻSA9v/22), *siedmi krolow* (RP416/16–17), *po dziesięćmi lat* (RP100/10–11), *nad dziesięćmi miast* (RP408/14);
- w odmianie nowej liczebniki te nie posiadają ani liczby, ani rodzaju, w efekcie czego przyjmują wartość kategorii liczby 0 oraz wartość kategorii rodzaju 0, np. *i odszedł mało dalej od onych ośmi* (RD56/31–32), *nie ostawił dziewięćdziesiąt i dziewięć na puszczy* (ŻPJK36v/12–13).

Gdy liczebniki główne od *pięciu* wzwyż występowały w mianowniku lub bierniku, a kontekst nie pozwalał stwierdzić, że forma odmieniała się według starego typu<sup>9</sup>, nie mogliśmy i nie chcieliśmy wskazywać jednoznacznie typu odmiany, zdecydowaliśmy się więc na przypisanie X w miejscu wartości kategorii liczby i rodzaju (np. *mam pięć bratow* – num\_card acc PIĘĆ (RP397/2), *święto trwało przez ośm dni* – num\_card acc OŚM (ŻPJK20r/15), *w żywocie mojem dziewięć miesięcy przebywał* – num\_card nom DZIEWIĘĆ (RD28/6–7), *dziesięć lat krolował* – num\_card nom DZIESIĘĆ (RP175/8)). Zdecydowaliśmy, że niewłaściwe byłoby uznanie wszystkich tego typu poświadczeń za ślad odmiany nowej, zważywszy na specyfikę materiału staropolskiego, w którym stara i nowa odmiana długo współwystępowały obok siebie. Jak twierdzi A. Słoboda przy oglądzie czasownika „nie jest możliwe stwierdzenie, w którym momencie jego forma jest jeszcze konotowana przez liczebnik, a w którym uniezależnia się od grupy liczebnikowo-rzeczownikowej” (Słoboda 2012: 65).

W podobny sposób traktujemy liczebniki z członem *-naście* i *-dziesięcia*, *-dziesięci*, *-dziesiąt*, mimo że te drugie „genetycznie będące złożeniami, odzwierciedlają odmienny niż przy liczebnikach o wartości 11–19 proces myślowy, mianowicie multiplikację (mnożenie)” (ibid.: 85). Przyjęliśmy, że czym innym jest semantyka liczenia, a czym innym są formalne wykładniki liczebnika jako klasy. W zakresie anotacji wartości kategorii i rodzaju postępujemy z nimi tak samo jak z liczebnikami od *pięciu* wzwyż. Również w tej grupie występują liczebniki dające się wyraźnie zinterpretować jako przykłady starej (wciąż żywej) odmiany, przyjmującej liczbę pojedynczą i rodzaj żeński, np. *począwszy od trzydziści lat* (ŻPJK22r/33), *aż do sześćdziesiąt lat* (ŻSA43r/3), *ona trzydziści pieniędzy* (HTK203r/8–9), *tę trzydziści pieniędzy złotych* (HTK203r/20), *coż by uczyniła dwanaście ciem anjołow* (RP642/2–3). W związku z tym, że najczęściej jednak nie daje się zaobserwować typowych dla starej odmiany relacji składniowych, liczebniki te traktujemy jako skostniałe leksemy, co z kolei uzasadnia wybór wartości 0 dla kategorii liczby i rodzaju (np. *aż do czterdziści dniow*

9 W przeciwieństwie do sytuacji, w których kontekst gramatyczny sugeruje odmianę starego typu, np.: *drugą pięć* RP489/12 – forma *drugą* wskazuje na sg f, czyli starą odmianę.

(EN134v/17), *pięćdziesiąt braciej* (RP173/19), *o tych trzedzieści pieniędzy srebrnych* (HTK207r/15–16)). Daje się zauważyć, że ścieranie się starej i nowej odmiany widoczne jest w większym stopniu w obrębie liczebników głównych z dziesiątkami niż z członem *-naście*.

Niejednoznaczne formy mianownikowe i biernikowe zmuszają nas jednak także do wyboru wartości X wobec niemożności rozstrzygnięcia, czy mamy do czynienia ze starą czy nową odmianą tych liczebników (np. *nalazł w żydowskich księgach pięćnaście znamion* num\_card acc PIĘĆNAŚCIE (RP478/10–11)).

Dla liczebników głównych oznaczających setki przyjęliśmy podobne rozstrzygnięcia. W anotacji określamy, czy forma liczebnika lub jego kontekst wskazują starą bądź nową odmianę.

Szczególną grupą w obrębie omawianej klasy, w której wartości 0 i X znajdują częste zastosowanie, są liczebniki główne nieokreślone, takie jak: *wiele*, *dużo*, *mało*, *niewiele*, które nie przyjmują kategorii liczby ani rodzaju (przypisujemy im zawsze wartość 0, np. *wiele luda*, *wiele rodziny*, *wiele młodzieńców*, *wiele lwow*).

## 1.2. Liczebnik zbiorowy

Liczebniki zbiorowe anotowane są przez nas tak jak liczebniki główne. O tym, czy dany liczebnik przyjmuje wartość kategorii liczby i rodzaju czy jej nie ma, decyduje również dawny bądź nowy model odmiany. Składnia liczebnika zbiorowego i towarzyszącego mu rzeczownika ewoluowała w stronę składni rządu (np. *pięcioro ksiąg*, *siedmioro chleba/chlebów*, *dwoje święceń*), pozbawiając tym samym tę klasę kategorii liczby i rodzaju (w tych wypadkach otrzymują one wartości 0), w przeciwieństwie do dawniejszej składni zgody (np. *pięciora księga*, *siedmior chleb*, *dwoje święcenie*), pozwalającej na wyraźne przypisanie wartości liczby czy rodzaju omawianym jednostkom. Gdy jednak brakuje kontekstów rozstrzygających, liczebniki zbiorowe mają przypisane konsekwentnie wartości 0 dla kategorii liczby i rodzaju.

## 1.3. Liczebniki krotne i wielorakie

Jak wskazaliśmy wcześniej, klasy te są wewnętrznie zróżnicowane. Formy typu *pięciodroć*, *siedmiorako* nie przyjmują kategorii liczby, rodzaju ani przypadku (mając tym samym w tych miejscach 0), w przeciwieństwie do form typu *pięciodrotny*, *siedmioraki*, które mają określone wszystkie wartości kategorii gramatycznych.

## 2. Zaimki

Przyjęty przez zespół podział zaimków na zaimki rzeczowne, przymiotne, przysłówne i liczebne uwzględnia zarówno kryterium semantyczne, jak i formalne – zaimki zostały podzielone na kilka odrębnych klas, pod względem gramatycznym odpowiadających poszczególnym częściom mowy. Nie uwzględniono natomiast podziału funkcjonalnego zaimków (np. zaimki dzierżawcze), podobnie nie wydzielono zaimków nieokreślonych – zaimki nieokreślone trafiły do poszczególnych klas na podstawie kryterium formalnego<sup>10</sup>.

### 2.1. Zaimki rzeczowne

Zaimki *co*, *kto*, *coż*, *ktoż*, *nic*, *nikt*, *cokoli*, *ktokoli* na etapie ręcznej anotacji gramatycznej zostają przypisane do właściwej sobie klasy, nie otrzymują natomiast określonych wartości kategorii liczby i rodzaju (mają wartość 0). Przypisana jest im jedynie wartość przypadku. Inaczej traktujemy zaimek rzeczowny *to*, który otrzymuje wartość liczby pojedynczej oraz wartość rodzaju – nijaki. Jest to związane z jego genezą – inaczej niż *co* i *kto*, nie jest to pierwotny zaimek pytajny<sup>11</sup>.

Ponieważ w przyjętej siatce klas i kategorii gramatycznych nie uwzględniamy funkcjonalnego podziału zaimków, zaimek *co* występujący w funkcji względnej anotowany jest w ten sam sposób (zaimek rzeczowny, wartość liczby – 0, wartość rodzaju – 0, odpowiednia wartość przypadku).

Zaimki rzeczowne osobowe dla pierwszej i drugiej osoby w liczbie pojedynczej (*ja*, *ty*) i mnogiej (*my*, *wy*) otrzymują podczas anotacji, po przypisaniu im właściwej klasy, wartości liczby – pojedynczej lub mnogiej, nie otrzymują zaś wartości rodzaju (przypisujemy w tym miejscu 0). W ten sam sposób są traktowane formy tekstowe *się*, interpretowane jako biernik zaimka rzeczownego z lematem JA<sup>12</sup> (np. *weźmi na sie odzienie włosiane* ŻPK71r/25)<sup>13</sup>. Zaimki trzecioosobowe przyjmują liczbę

10 Zob. *Opracowania, Materiały dodatkowe, Opisy klas i kategorii, Zaimek* na stronie apocrypha.amu.edu.pl (zawartość w przygotowaniu), zob. też Deskur et al. 2022: 167, 170.

11 Odrębną kwestią jest zaimek przymiotny *to* określający rzeczowniki liczby pojedynczej rodzaju nijakiego.

12 Konsekwencją decyzji o formalnym podziale zaimków na rzeczowne, przymiotne itd. jest nieobecność klasy zaimka zwrotnego. Przyjęcie lematu JA było decyzją arbitralną wynikającą m.in. z tego, że lematem dla zaimków rzeczownych jest forma mianownika. Nie mogliśmy pozostawić pustego miejsca w lemacie, aby wszystkie formy *się*, *siebie*, *sobie* itd. były wyszukiwalne. Lemat, o czym pisaliśmy w poprzednim artykule (por. Deskur et al. 2022: 166), jest tylko etykietą służącą przede wszystkim wyszukiwarce leksykalnej, nie gramatycznej. Lematyzacja tekstu dawnego będzie przedmiotem osobnego artykułu.

13 Leksem *się* jest w anotacji traktowany dwojako. Może być on również morfemem słowotwórczym, wykładnikiem zwrotności, bierności lub bezosobowości, wówczas należy do odrębnej klasy (gram\_SIE), zob. Deskur et al. 2022: 171.



pojedynczą lub mnogą oraz odpowiedni rodzaj gramatyczny<sup>14</sup>: męski (*on*), żeński (*ona*), nijaki (*ono*) w liczbie pojedynczej bądź męski (*oni*, *ony*) lub niemęski (*one*, *ony*) w liczbie mnogiej<sup>15</sup>.

## 2.2. Zaimki przymiotne

Zaimki dzierżawcze, ze względu na przyjęty system klas i kategorii, zostały sklasyfikowane jako zaimki przymiotne. Rozkład wartości 0 lub X w kategoriach gramatycznych poszczególnych zaimków tej klasy przebiega różnie. Zaimki pierwszej i drugiej osoby liczby pojedynczej i mnogiej (*moj*, *twój*, *nasz*, *wasz*) są w pełni odmiennie i nie wymagają przypisywania im 0 lub X.

Inaczej jest w anotacji zaimków dzierżawczych trzeciej osoby liczby pojedynczej i mnogiej (*jego*, *jej*, *ich*). W przykładach typu *Tys to stworzył wielkie i bardzo szerokie morze, a na jego głębokości założyłeś uć<w>irdzenie ziemie* (RP5/19–21) można było przyjąć różne rozwiązania:

- anotować *jego*, *jej*, *ich* jako dopełniacz zaimka rzeczownego – stracilibyśmy wówczas dzierżawczą funkcję tych zaimków (*jego* – pron\_subst sg:gen:n ONO),
- anotować *jego*, *jej*, *ich* jako zaimek przymiotny, a rodzaj i liczbę przejąć z rzeczownika, który mógł być oddalony, ale z którym pozostawały w związku (*szerokie morze, a na jego głębokości* – *jego* rodzaj nijaki).

Z uwagi na ich skostniałość przyjęliśmy, że zaimki *jego*, *jej*, *ich* przyjmują wartość 0 w kategorii liczby i rodzaju. Synkretyzm form deklinacyjnych skłonił nas również do przypisania wartości 0 w kategorii przypadku.

## 2.3. Zaimki przysłowne i liczebne

Zaimki przysłowne, odpowiadające pod względem gramatycznym nieodmiennym częściom mowy, nie odmieniają się przez uwzględnione w anotatorze kategorie gramatyczne. Otrzymują zatem na poziomie liczby, rodzaju i przypadku wartości 0.

Wartość kategorii 0 sprawdza się także w anotacji zaimków liczebnych, do których – w związku z przyjętą przez nas klasyfikacją – nie należy wiele jednostek

14 Jak pisaliśmy we wcześniejszym artykule, „nie zdecydowaliśmy się na wprowadzenie kategorii męskoosobowości, ponieważ taka kategoria w staropolszczyźnie dopiero się kształtowała. Zamiast tego wprowadzamy wartość kategorii rodzaj niemęski. Ma to na celu oddanie zróżnicowania form rodzajowych bez jednoczesnego przypisywania staropolszczyźnie kategorii wykształconej dużo później” (Deskur et al. 2022: 173).

15 Zaimek *on* może wystąpić w funkcji wskazującej, klasyfikujemy go wówczas jako zaimek przymiotny, przypisując mu właściwe wartości kategorii w anotatorze (określoną wartość liczby, rodzaju, przypadku). Wpływ na interpretację leksemu *on* jako zaimka wskazującego ma kontekst składniowy i treściowy, czasem też źródło łacińskie.

(podkreślić trzeba, że wyrazy takie jak *wiele*, *dużo*, *mało*, *niewiele* anotujemy jako liczebniki główne bądź przysłówki). Zaimki takie jak *ile*, *ilekroć*, *tyle*, *kilko* nie przyjmują wartości w kategoriach liczby ani rodzaju (jest im przypisane 0).

## Zastosowanie wartości 0, X w anotacji wybranych kategorii

### 1. Stopień

Dla klasy przymiotników przewidziano w anotacji kategorię stopnia, w której wyróżniamy stopień równy i stopień porównawczy (por. Kleszczowa 2004), co zostało dokładnie omówione w przygotowanym przez zespół artykule *Anotacja gramatyczna tekstu staropolskiego – problemy i wyzwania* (Deskur et al. 2022). Jednak w polszczyźnie istnieją przymiotniki, które charakteryzują się niestopniowalnością. W narzędziu w kategorii stopnia zyskują one wartość 0. Do takiej grupy należą szczególnie przymiotniki dzierżawcze, np. *Jesusowy*, *Abramowy*, *aleksandryjski*, *jeruzolimski*, czy też przymiotniki typu *drewniany*, *drzewny*, *wszechmogący*. To, czy dany przymiotnik się stopniuje, określane jest na podstawie poświadczeń form w stopniu porównawczym w analizowanym materiale lub/i słownikach historycznych (SStp, SPXVI). Stopniowalność lub jej brak jest na stałe przypisana do danego leksemu. Nie uwzględniamy znaczenia wynikającego z kontekstu (np. *duchowny*<sup>16</sup> w staropolskich apokryfach występuje prawie wyłącznie w znaczeniu ‘kapłan’, natomiast SPXVI podaje formę stopniowalną *duchowniejszy* – z tego względu w anotacji wyraz przyjmuje wartość stopnia pos).

Te same zasady stosujemy w anotacji przysłówków. Ze względu na opisane rozstrzygnięcia zarówno w wypadku przymiotników, jak i przysłówków nie było potrzeby przypisywania wartości X w kategorii stopnia.

## Wartości 0 i X z perspektywy użytkownika narzędzia

Dzięki zastosowaniu wartości 0 i X możliwa była pełna anotacja gramatyczna materiału. Użytkownik narzędzia widzi jednak tylko część z przypisanych przez anotatora informacji, w zależności od wyboru funkcjonalności oraz sposobu wyświetlania poszczególnych danych<sup>17</sup>.

16 W anotowaniu klas kierujemy się kryterium formalnym. Zsubstantywowane przymiotniki anotujemy jako przymiotniki. Jeśli dany przymiotnik może się stopniować, anotowany jest jako pos lub comp, niezależnie od znaczenia kontekstowego.

17 Więcej na temat funkcjonalności dostępnych na stronie *Apocrypha* zob. część *Funkcjonalności przydatne w badaniach językoznawczych* w: Deskur et al. 2022: 175–176).

W wyszukiwarkach gramatycznych (zakładka *Formy gramatyczne*) umożliwiono wyszukiwanie wszystkich klas oraz kategorii i ich wartości przypisywanych w anotacji, również 0 i X. Po lewej stronie ekranu znajduje się lista tekstów, klas i kategorii. Po wyborze odpowiedniego filtra użytkownik otrzyma interesujące go wyniki. W tym trybie zatem, gdy zaznaczymy w klasach filtr „Przymiotnik”, a w stopniu filtr „0”, po kliknięciu w „Wyszukaj” uzyskamy listę wszystkich niestopniowalnych przymiotników. Rozróżnienie wartości takich jak 0 i X nie ginie dla użytkownika narzędzia, a ich uwzględnienie w wyszukiwarce gramatycznej ułatwia badanie języka staropolskiego z poszanowaniem jego specyfiki. Użytkownik nie musi dokonywać własnej, niekiedy żmudnej i czasochłonnej, ekscerpcji materiału źródłowego, by wyodrębnić te elementy języka, które zamierza przeanalizować. Z tego punktu widzenia udostępnienie użytkownikowi w wyszukiwarce gramatycznej wartości 0 i X jest pomocne w badaniach nad staropolszczyzną.

Z wynikami anotacji można się także zapoznać, włączając tryb komentarza gramatycznego. Użytkownik otrzyma wówczas informację o klasach i kategoriach przypisanych poszczególnym wyrazom. W tym trybie jednak wartości 0 i X nie wyświetlają się – nie są one *de facto* informacją o wartości kategorii (choć roboczo tak je nazywamy), a informacją o niemożliwości przypisania wartości danej kategorii z uwzględnieniem wewnątrz- i zewnątrzjęzykowych przyczyn tej niemożliwości. Jeśli w procesie anotacji w poszczególnej kategorii dla danego elementu przypisano wartość 0 lub X, wówczas w komentarzu gramatycznym nie są one wyświetlane (kolejność elementów komentarza gramatycznego jest zawsze stała, więc użytkownik może zobaczyć, w którym miejscu komentarza nie ma żadnego tagu). Jeśli zatem mamy do czynienia np. z niestopniowalnym przymiotnikiem, to w komentarzu gramatycznym znajdziemy informację o przypisanej klasie (adj) oraz określonych kategoriach (a więc o przypadku, liczbie i rodzaju), natomiast nie pojawi się żadna informacja o stopniowaniu, co świadczy o tym, że ten przymiotnik się nie stopniuje, czyli w kategorii stopnia przypisano mu wartość 0. W takim świetle brak informacji jest znaczący.

Takiego sposobu udostępniania informacji na temat wartości 0 i X nie wypracowano od razu. Początkowo zaplanowano jedynie oznaczenie „Pomiń”, które pozwalało pominąć w anotacji określone wyrazy (np. te w całości wstawione przez wydawcę). Miało ono być tylko wewnętrznym rozwiązaniem, które umożliwiłoby powrót do anotacji pominiętych wyrazów (np. w sytuacji odnalezienia źródła, innej redakcji itp.). Z czasem okazało się, że zaplanowana siatka klas i kategorii wymaga zastosowania dodatkowego oznaczenia dla tych wyrazów, które nie przyjmują danej kategorii właściwej dla swojej klasy (np. stopień przy przymiotnikach dzierżawczych). Wtedy wprowadzono wewnętrzne rozróżnienie na 0 – nie dotyczy i X – nie można określić. Wciąż jednak informacje te nie miały być widoczne dla użytkownika.

W trakcie anotacji znaleziono znacznie więcej miejsc, w których informacja „wartość kategorii = 0” mogłaby być istotna dla użytkownika narzędzia. Dzięki

umożliwieniu wyszukiwania także przez „wartość kategorii = 0” użytkownik może znaleźć np. formy przysłówkowe liczebników wielorakich. Zdecydowano wówczas o udostępnieniu tej informacji użytkownikom.

Dalsza praca pokazała, że również informacja „wartość kategorii = X” nie dotyczy wyłącznie sytuacji technicznych (uszkodzeń tekstu). Pozwala ona na przykład odnaleźć te miejsca, w których nie da się stwierdzić, czy mamy do czynienia ze starą czy z nową odmianą liczebnika. Z tego względu obie te wartości znalazły się w wyszukiwarkach, a wszystkie informacje przypisane przez anotatora są dostępne użytkownikowi.

## Podsumowanie

Wprowadzenie wartości 0 i X okazało się niezbędne w anotacji tekstów staropolskich, a szczegółowe rozwiązania wymagały wzięcia pod uwagę zarówno specyfiki tekstów dawnych, jak i działania narzędzia. Sposób postępowania z miejscami, których z różnych powodów nie można w pełni zaanotować, ewoluował od rozstrzygnięć narzędziowych, funkcjonalnych z perspektywy anotatora, do rozstrzygnięć merytorycznych. Uznano, że informacja o braku wartości niektórych kategorii jest istotna również dla badaczy korzystających z narzędzia, ponieważ odzwierciedla ona charakter języka staropolskiego, „w którym granice między klasami gramatycznymi były płynne, niektóre kategorie się kształtowały, inne zanikały, a paradygmaty poszczególnych leksemów dopiero się stabilizowały” (Deskur et al. 2022: 176). Dzięki zastosowaniu wartości 0 i X badacz ma możliwość uchwycenia procesów zachodzących w staropolszczyźnie. Narzędzie nie podsuwa mu sztucznych granic pomiędzy poszczególnymi etapami rozwojowymi języka. Jest to rozwiązanie nowatorskie, pozwalające uniknąć rzutowania współczesnej perspektywy językowej na materiał dawny.

## Skróty

gram\_SIE – zgramatyzalizowane *się*, num\_card acc – liczebnik główny w bierniku, num\_card dat:m – liczebnik główny w celowniku w rodzaju męskim, num\_card nom – liczebnik główny w mianowniku, num\_card nom:m – liczebnik główny w mianowniku w rodzaju męskim, num\_card nom:nm – liczebnik główny w mianowniku w rodzaju niemęskim, pron\_subst sg:gen:n – zaimek rzeczowny w liczbie pojedynczej w dopełniaczu w rodzaju nijakim

## Literatura

- DESKUR A., NOWAK-PASTERKA E., ROJSZCZAK-ROBIŃSKA D., STELMACH W., WISMONT M., ZALEJARZ K., ZIÓŁKOWSKA O., 2022, *Anotacja gramatyczna tekstu staropolskiego – problemy i wyzwania*, „LingVaria” nr 2 (34), s. 163–180, <https://doi.org/10.12797/LV.17.2022.34.11>.
- Elektroniczny Korpus Tekstów Polskich z XVII i XVIII w. (do 1772 r.), [on-line:] <https://korba.edu.pl/>.
- KĘPIŃSKA A., 2006, *Kształtowanie się polskiej kategorii męsko- i niemęskoosobowości. Język wobec płci*, Warszawa.
- KLESZCZOWA K., 2004, *Kategoria stopnia staropolskich przymiotników*, [w:] H. Kurek, J. Labocha (red.), *Studia linguistica Danutae Wesołowska oblata*, Kraków, s. 85–92.
- RZEPKA W.R., TWARDZIK W.B., 1996a, *Archaizmy fleksyjne w Rozmyślaniu przemyskim*, cz. 1, „Język Polski” LXXVI, s. 97–102.
- RZEPKA W.R., TWARDZIK W.B., 1996b, *Archaizmy fleksyjne w Rozmyślaniu przemyskim*, cz. 2, „Język Polski” LXXVI, s. 321–327.
- RZEPKA W.R., TWARDZIK W.B., 1997a, *Archaizmy fleksyjne w Rozmyślaniu przemyskim*, cz. 3, „Język Polski” LXXVII, s. 41–46.
- RZEPKA W.R., TWARDZIK W.B., 1997b, *Archaizmy fleksyjne w Rozmyślaniu przemyskim*, cz. 4, „Język Polski” LXXVII, s. 102–111.
- RZEPKA W.R., TWARDZIK W.B., 1997c, *Archaizmy fleksyjne w Rozmyślaniu przemyskim*, cz. 5, „Język Polski” LXXVII, s. 298–302.
- RZEPKA W.R., TWARDZIK W.B., 1999, *Archaizmy fleksyjne w Rozmyślaniu przemyskim*, cz. 6, „Język Polski” LXXIX, s. 35–46.
- SIUCIAK M., 2008, *Kształtowanie się kategorii gramatycznej liczebnika w języku polskim*, „Prace Naukowe Uniwersytetu Śląskiego w Katowicach”, nr 2585, Katowice.
- SŁOBODA A., 2012, *Liczebnik w grupie nominalnej średniowiecznej polszczyzny. Semantyka i składnia*, Poznań.
- 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, t. XXXVII, red. P. Potoniec, t. XXXVIII, red. K. Mrowcewicz, Wrocław – Warszawa – Kraków 1966–2020.
- SSTP: S. Urbańczyk (red.), *Słownik staropolski*, t. I–XI, Wrocław – Warszawa – Kraków 1953–2002.

### Old Polish Language Specificity and Grammatical Annotation: On the Need for New Category Values

#### Abstract

The paper is focused on the grammatical annotation of Old Polish texts. The authors discuss the necessity to use unusual grammatical category values: 0 and X. It occurs when any of the values is impossible to determine (e.g. due to damage to the manuscript or insufficient syntactic/semantic context) or a word takes no value in a given category (e.g. an unstated adjective). The article presents factors influencing the choice of 0 and X values. It shows those grammatical classes and categories in whose annotation the use of these values has been highly significant.