

<http://dx.doi.org/10.17651/POLON.36.1>

## ZYGMUNT SALONI

Uniwersytet Warszawski, Warszawa  
saloni@uw.edu.pl

### **Systematyzacja wzorów fleksyjnych dla „Słownika gramatycznego języka polskiego”<sup>1</sup>**

Celem „Słownika gramatycznego języka polskiego” (<http://sgjp.pl/>) w zakresie fleksji jest dokładne podanie paradygmatów leksemów odmiennych polszczyzny współczesnej w wersji pisanej. Został on zasadniczo osiągnięty już w wersji elektronicznej SGJP, dwukrotnie wydanej na płycie CD-ROM (SGJP 2007, SGJP 2012). Zostały opracowane typy odmiany leksemów oraz realizujące je wzory. Leksemom odmiennym zostały przyporządkowane wzory, które mają poprawnie generować formy leksemów. Niestety symbole przypisane poszczególnym wzorom nie układały się w przejrzysty system.

Celem ostatniego etapu pracy, którego wynikiem jest wersja internetowa słownika, było opracowanie jednolitego i w miarę możliwości przejrzystego systemu symboli dla wszystkich wzorów fleksyjnych użytych w słowniku. Ze względu na dużą liczbę wzorów (ok. 1000, w tym prawie 800 rzeczownikowych) to zadanie trudne. Konieczna była dokładna analiza usystematyzowanego materiału – w zestawieniu z dotychczasowymi doświadczeniami jego implementacji w obu opublikowanych wersjach słownika.

Przy tworzeniu systemu wzorów fleksyjnych za podstawę wzięto systematyzację Jana Tokarskiego zawartą w SJPDor. (oraz słownikach wywodzących się od niego). Jej zalety to logiczność, przejrzystość i, co najważniejsze, znaczne rozpowszechnienie. Operuje nią chyba niewielu użytkowników, ale faktycznie istnieje ona w milionach papierowych kopii. I jest wykorzystana, bezpośrednio lub pośrednio, zapewne

<sup>1</sup> Praca powstała w ramach projektu, który został sfinansowany ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2011/01/B/HS2/04695.

Artykuł niniejszy jest napisany przez jednego autora, od którego pochodzi również systematyzacja i numeracja wzorów. Została ona jednak włączona do słownika, którego całościowa organizacja jest dziełem wspólnym zespołu autorskiego SGJP. Dlatego też w narracji użyte są bądź formy liczby mnogiej, bądź formy nieosobowe, dla podkreślenia, że referowane sądy i decyzje są wypracowane wspólnie.

we wszystkich programach komputerowych operujących fleksją polską. Nawiązanie do klasyfikacji Tokarskiego powinno stanowić ułatwienie dla kogoś, kto będzie posługiwał się systemem oznaczeń zastosowanych w SGJP.

## 1. Podstawy teoretyczne

Jest rzeczą oczywistą, że przedmiotem opisu jest w naszej sytuacji pisana wersja polszczyzny. Opozycje fleksyjne języka polskiego, regularnie występujące morfemy gramatyczne czy alternacje muszą być analizowane w kontekście napisów, tego, jak zapisywane są słowa – z uwzględnieniem zarówno najbardziej podstawowych cech polskiego systemu graficznego (np. skomplikowanych zasad oddawania na piśmie spółgłoskowych fonemów miękkich), jak i szczegółowych zwyczajów ortograficznych, czasem powszechnie respektowanych i skodyfikowanych w zasadach pisowni.

Zestawianie regularnie skontrastowanych serii napisów pozwoliło słowa występujące w polskich tekstach zinterpretować jako formy wyrazowe (fleksyjne) i pogrupować w leksemy. Charakterystyczną część wspólną słów (wyodrębnianą graficznie), będących wykładnikami form tego samego leksemu, nazywamy **tronem (fleksyjnym)**<sup>2</sup>, pozostałą – następującą po nim część słowa – **zakończeniem**<sup>3</sup>.

Leksemy układają się w klasy, odpowiadające z grubsza tradycyjnie wyodrębnianym częściom mowy. Leksemy należące do jednej klasy charakteryzują się zbliżonymi cechami fleksyjnymi, mają paradygmaty zbudowane według podobnych zasad. Te właściwości są istotne przy tworzeniu wzorów fleksyjnych. Ten sam wzór fleksyjny reprezentują leksemy, które mają identyczne zakończenia we wszystkich formach wyrazowych, tj. różnią się tylko tronem fleksyjnym, charakterystycznym dla leksemu.

Naczelnym naszym zadaniem jest zdyscyplinowany opis fleksji. Doświadczenia, zarówno samego opisu, jak i – tym bardziej – komputerowej jego implementacji, pokazują konieczność dwustopniowej realizacji tego zadania. Należy bowiem oddzielić opis funkcji form fleksyjnych (fleksja głęboka, ściśle powiązana ze składnią) i czysto zewnętrzny opis poszczególnych form fleksyjnych oraz ich wzajemnych stosunków (fleksja powierzchniowa).

Rozróżnienie to widać bardzo wyraźnie w odmianie leksemów liczebnikowych, w których niewiele słów (np. *pięćset*, *pięciuset* i rzadkie *pięciomaset*) obsługuje skomplikowany paradygmat zbudowany na opozycjach (przypadka i rodzaju), mających wiele wartości, czyli służy do wypełnienia wielu klatek paradygmatu.

W podręcznikach gramatyki pokazuje się z reguły **paradygmaty pełne**, w których poszczególne wykładniki tekstowe (najczęściej – słowa) są interpretowane jako

<sup>2</sup> Cząstka ta, którą w poprzednich wydaniach SGJP nazywaliśmy tematem fleksyjnym, nie musi być najdłuższym ciągiem liter dla wszystkich form danego leksemu, np. dla leksemu *plywak* 'przedmiot' został wyodrębniony tron *plywa*. Kwestia ta zostanie objaśniona niżej.

<sup>3</sup> Obie te cząstki mogą być zerowe, np. dla leksemu *sen* (zerowy tron: *0+sen*, *0+śnie*) albo *dom* (stały temat, zerowe zakończenie w mianowniku i bierniku lp.).

formy leksemów (wyrazów w sensie słownikowym), mające określone wartości kategorii fleksyjnych. Z punktu widzenia wzajemnego stosunku form wygodnie jest formy danego leksemu przedstawić w postaci **paradygmatu uproszczonego**, zawierającego tylko **formy bazowe**, wystarczające do utworzenia wszystkich form za pomocą prostych zabiegów (dopisywanie zakończeń lub ich części). Przy układaniu wzorów posługiwaliśmy się właśnie takimi paradygmatami. Wzory zbudowane zgodnie z określonym schematem paradygmatu **uproszczonego** tworzą dany **typ** wzoru. Nadrzędne względem typów są **nadtypy** (w internetowej wersji SGJP zwane roboczo „**typami odmiany**”), grupujące typy stosowane dla określonej klasy leksemów.

W niniejszym omówieniu, inaczej niż w całym słowniku i instrukcji do niego, gdzie ograniczamy się do technicznego podziału formy wyrazowej na trzon fleksyjny i zakończenie, posługujemy się również terminami wziętymi z tradycyjnej analizy morfologicznej: **temat** danego leksemu jest to wyabstrahowany zbiór ciągów liter charakterystycznych dla danego leksemu (alternantów tematu), między którymi zachodzą relacje uznane za typowe dla języka polskiego, **końcówka** – to ciąg liter występujący w formie wyrazowej po alternancie tematu, charakteryzujący daną formę leksemu<sup>4</sup>. Zwróćmy uwagę na to, że nie można powiedzieć, że występuje ona w końcowej części słowa, może bowiem po niej iść jeszcze **postfiks** (np. +ś, +kolwiek); zdarza się też, że między tematem a końcówką występuje znak graficzny (łącznik lub apostrof), którego tradycyjnie nie włącza się do żadnej z tych części. Z kolei **prefiks** to początkowa część słów będących wykładnikami danego leksemu, odróżniająca je od słów pokrewnych słowotwórczo; leksemy różniące się prefiksem odmieniają się na ogół podobnie.

## 2. Wzory rzeczownikowe (ABCDE)

Usystematyzowanie wzorów rzeczownikowych było najtrudniejsze. Członkowie zespołu SGJP wykonali wiele prób systematyzacji wzorów deklinacyjnych i wprowadzili wiele pojęć pomocniczych (Gruszczyński 1989, Gruszczyński 2001, Gruszczyński, Saloni 2005, Skowrońska 2011). Zasadniczym efektem tych prac było oddzielenie wyróżnianych typów wzorów od klas rodzajowych, wprowadzone już w książce Gruszczyńskiego (1989). Widać to jaskrawo w wypadku podrodzajów rodzaju męskiego: paradygmat pełny każdego z trzech rzeczowników **plywak** może być uznany za szczególną realizację tego samego wzoru, różnice obejmują bowiem tylko cztery klatki paradygmatu, przy czym w dwu z nich (bierniku obu liczb) występuje forma przeniesiona z innej klatki tegoż paradygmatu, a trzecia i czwarta klatka (forma niedeprecjatywna mianownika liczby mnogiej i synkretycznego z nim wołacza) jest specyficzna tylko dla jednego podrodzaju – *m1* (męskoosobowego). Tradycyjnie wyróżnia się też rzeczowniki odmieniane według wzorów przymiotnikowych (w SJPDor. „*odm. jak przym.*”), a także rzeczowniki męskie „odmienne jak żeńskie” (np. w SJPDor.: „**wojewoda**

<sup>4</sup> Ewentualna podzielność końcówki na morfemy (np. w czasownikowych formach aglutynacyjnych, typu *pis+a-l-y-by-śmy*) nie jest istotna dla naszego problemu.

*m odm. jak ż IV*”). W pierwszym wypadku rzeczownik ma wszystkie formy takie jak „homonimiczny” przymiotnik, tylko ograniczone do jednego rodzaju, np. nazwisko **Topolski** ma – po ucięciu dwu pierwszych liter – formy takie jak **polski** w rodzaju *m1*, a **Topolska** – takie jak **polski** w rodzaju *f*. W drugim wypadku różnice są większe, np. rzeczownik **wojewoda** ma – po ucięciu początkowych liter – formy takie jak rzeczownik **woda**, ale z dwoma wyjątkami: dopełniacza (i systemowo synkretycznego z nim biernika) liczby mnogiej (w którym występuje końcówka *-ów*, charakterystyczna dla rodzaju męskiego) oraz specjalnej formy niedeprecjatywnej mianownika (i systemowo synkretycznego z nim wołacza) liczby mnogiej; ponadto w jego paradygmacie nie występuje alternant tematu z *ó* (*wód*).

Obserwacje te były punktem wyjścia do pomysłu, żeby wzór obejmował rzeczowniki różnych rodzajów, jeśli różnice w wybranych kilku kłatkach są uwarunkowane rodzajowo. Prowadziło to do umieszczenia w schemacie wzoru danego typu większej liczby klatek niż obejmuje paradygmat rzeczownika<sup>5</sup>. Natomiast zbędne było dwukrotne umieszczanie we wzorze form objętych synkretyzmem uniwersalnym dla danego typu (np. celownika i miejscownika wzorów typu żeńskiego). Schemat wzoru danego typu (paradygmat uproszczony) obejmuje zatem tylko formy bazowe (zob. SGJP 2012:45–6). Paradygmat konkretnego rzeczownika uzyskuje się z kilku wskaźników: nazwy leksemu, oznaczenia rodzaju symbolu wzoru i typu.

Wzory rzeczownikowe dzielą się na typy – w zależności od zasobu form bazowych wprowadzonych w paradygmacie uproszczonym: męski, żeński, nijaki i zerowy (obejmujący rzeczowniki nieodmienne). Rzeczowniki *plurale tantum* są zaliczane do odpowiedniego typu (i należącego do niego wzoru; są one traktowane jako rzeczowniki defektywne – nie mające, ze względu na rodzaj *p*, form liczby pojedynczej). Na użytek numeracji wzorów wprowadzono jeszcze pojęcie paratypu. Ze zbioru leksemów poszczególnych typów rodzajowych wyodrębniono leksemy odmieniające się „jak przymiotniki”, tj. mające wszystkie formy z zakończeniami przymiotnikowymi, tylko ograniczone do jednego rodzaju – utworzono z nich paratyp przymiotnikowy. Odpowiednio zmodyfikowano też wyjściowe typy: paratyp rodzaju X obejmuje rzeczowniki typu X z wyłączeniem tych, które zaliczono do paratypu przymiotnikowego.

Wyróżniono pięć paratypów, którym odpowiadają następujące znaki, występujące jako pierwsze w symbolach wzorów:

- A – odmiana paratypu przymiotnikowego
- B – odmiana paratypu męskiego
- C – odmiana paratypu nijakiego
- D – odmiana paratypu żeńskiego
- E – odmiana nietyпова

Symbol wzoru rzeczownikowego, podobnie jak czasownikowego, składa się z dwu segmentów rozdzielonych separatorem + (drugi segment jest opcjonalny; wy-

<sup>5</sup> Na przykład wiele wzorów typu żeńskiego (D) ma po znaku + symbol (ów), oznaczający wyjątkową końcówkę dopełniacza liczby mnogiej. Rzecz jasna, końcówka +ów nie występuje w rzeczownikach żeńskich (np. **fala**), tylko w męskich (np. **cieśla**) – oba odmieniają się według tego samego wzoru: D1+(ów).

stępuje tylko wtedy, gdy zachodzi potrzeba)<sup>6</sup>. Segment pierwszy podaje podstawową klasyfikację fleksyjną i najczęściej zawiera litery którejs z form leksemu (często podstawowej) uczestniczące w alternacjach.

Pierwszy segment zaczyna się od znaku A, B, C, D lub E, wskazującego paratyp (i – pośrednio – rzeczownikowość). Na drugim miejscu w symbolu wzoru rzeczownikowego występuje (z wyjątkiem E) liczba z przedziału 1–5, wskazująca na sposób odmiany, przede wszystkim uwarunkowany charakterem końcowej części tematu:

- 1 – temat na spółgłoskę miękką
- 2 – temat na spółgłoskę stwardniałą
- 3 – temat na spółgłoskę tylnojęzykową (fonologicznie – /k/ i /g/, a także – z wyjątkami – /x/ i /h/ – ze względu na reguły użycia *i* zmiękczonego)
- 4 – temat na spółgłoskę twardą (oprócz /k/, /g/ oraz czasem /x/ i /h/)
- 5 – temat niejednolity (tylko dla rzeczowników nijakich, najczęściej: wszystkie formy w sg zneutralizowane)

Podział ten jest – w zasadniczej koncepcji – przejęty z SJPDor., gdzie dla rzeczowników męskich i żeńskich wyróżniono grupy I, II, III i IV, prawie takie same, jak zdefiniowano wyżej: różnica sprowadza się do tego, że do grupy III zaliczono tam również wszystkie rzeczowniki o temacie zakończonym (fonologicznie) /x/ i /h/, kierując się zasobem końcówek. Dla nas istotniejsza była inna cecha: fonemom tym odpowiada w zapisie (c)h, po którym w funkcji znaku zmiękczenia występuje (lub nie występuje) *i* według podobnych zasad, jak po literach odpowiadających fonemom twardym innym niż /k/ i /g/. Ponadto tę samą zasadę rozciągnięto na rzeczowniki nijakie, dla których w SJPDor. zastosowano „uproszczony podział” (tematy na spółgłoskę miękką i stwardniałą razem w grupie I). Dla rzeczowników z grupy *nVI* z SJPDor., jak *muzeum*, utworzono wzór C5.

Podział tematów nie jest stosowany rygorystycznie. Leksemy przyporządkowane poszczególnym wzorom nie zawsze mają zakończenie tematu zgodne z podaną tu listą. Czasem, zwłaszcza w wypadku, gdy temat jest stały, nie alternujący, zaliczane są do wzoru o innym zakończeniu. Zdarza się bowiem, że identyczny zestaw końcówek występuje u rzeczowników po temacie należącym do różnych grup, jak jest na przykład w wypadku leksemów **Montaigne** (temat na miękką) i **George** (temat na stwardniałą). – oba odmieniają się według tego samego wzoru: B1+’w(ów).

Na miejscach trzecim i dalszych symbolu wzoru (przed separatorem +) występują litery uczestniczące w alternacjach, np.:

- B3g – litera *g* (*szpieg*)
- B3k – litera *k* (*rybak*)
- B3q – litera *q* (*Haq*)
- D2(se) – ciąg liter *se* (*berceuse*)
- B1i – litera *i* (*modrzew(ia)*)

<sup>6</sup> Człon symbolu wzoru po znaku + niekoniecznie musi odpowiadać faktycznym właściwościom leksemu, który się według tego wzoru odmienia, bo może się odnosić do innych leksemów – zob. wyżej przypis 4.

Ich funkcją jest uszczegółowienie znaku drugiego (liczbowego). Wystąpienie liter w nawiasie wskazuje, że powinny być one traktowane jako jednolity blok, uczestniczący w alternacjach na prawach jednego symbolu (w podanym przykładzie D2(se) określone fonemowi odpowiada tylko pierwsza litera, por. wymowa [bersöz]). Kolejne litery (niekoniecznie występujące w formie podstawowej) wpisywane są w symbolu wzoru bez nawiasu w porządku *a tergo*, np.:

D4hc (np. *mucha*)  
 B3ke (np. *Turek*)  
 B3gró (np. *mórg*)  
 D4re (np. *kołdra*)  
 D4rei (np. *gra*)

Po ciągu liter (*a tergo*) może wystąpić nietypowa (również: niema albo czytana niestandardowo) końcówka sg:nom, a także 0 – w wypadku, gdy w rzeczowniku żeńskim w sg:nom występuje końcówka zerowa, np.:

D3(ca) (np. *carioca* – por. [karioka])  
 D1we0 (np. *marchew*)  
 D1wei0 (np. *cerkiew*)  
 D2(o) (np. *Trapszo*)

Znaki po separatorze + wskazują odstępstwa od schematu wprowadzonego systematycznie, w szczególności nietypowe końcówki, np.:

B4+u (np. *skarb*)  
 B3keic+w (np. *Maciek*)

Zestaw tych znaków jest obecnie następujący:

- – łącznik występujący między tematem a końcówką (np. *NIK-u*)
- ' – apostrof występujący między tematem a końcówką (np. *Chomsky'ego*)
- (o) – nietypowa końcówka sg:nom +o (np. *Trapszo*)
- u – wyjątkowa końcówka sg:gen +u (np. *filmu*)
- (owi) – wyjątkowa końcówka sg:dat +owi (np. *koniskowi*)
- (u) – wyjątkowa końcówka sg:dat +u (np. *kotu, Berdyczowu*)
- (u?) – wyjątkowa końcówka sg:dat +u – jako alternatywna (np. *Lwu – Lwowi*)
- ! – wyjątkowa końcówka sg:voc, np. zerowa (*Marlowe*), +sie (od: *Smith*)
- !?! – wyjątkowa końcówka sg:voc obok regularnej (np. *mołodcze – mołodcu*)
- e – wyjątkowa końcówka pl:nom +e (np. *szanse, kwadranse, woźnice*)
- w – wyjątkowa końcówka pl:nom:m1 +owie (np. *zięciowie*)
- a – wyjątkowa końcówka pl:nom +a (np. *grunta, bracia*)
- i – wyjątkowa końcówka pl:nom:m1/f/m3 +i (np. *ortodoksi, Kaszubi, nici, śmieci*)
- (ie) – wyjątkowa końcówka pl:nom:m1 +ie (np. *Cyganie*)
- y – wyjątkowa końcówka pl:nom:(m1) +y (np. *rzeczy, myszy, statuy, zdrajcy*)
- c – wyjątkowa końcówka pl:nom:f +ce (np. *ręce*)
- (ów) – wyjątkowa końcówka pl:gen +ów (np. *papierosisków*)
- (y) – wyjątkowa końcówka pl:gen lub sg:gen +y (np. *naręczy, statuy*)

- (i) – wyjątkowa końcówka pl:gen +i (np. *osiedli*)  
 0 – wyjątkowa końcówka zerowa pl:gen (np. *procent, krakowian, służ*)  
 (mi) – wyjątkowa końcówka pl:inst +mi (np. *liśćmi*)  
 (y?) – wyjątkowa końcówka alternatywna pl:inst +y (np. *czasy, słowy*)  
 (ech) – wyjątkowa końcówka pl:loc +ech (np. *Włoszech*)  
 (sz) – nietypowy alternant sz (dla h) (np. *yamasze*)  
 (dz) – nietypowy alternant dz (dla h) (np. *Pradze*)  
 (ż) – nietypowy alternant ż (dla h) (np. *braże*)  
 (yj) – końcówka +yj w pl:gen:fnuni lub sg:nom (np. *Szeptyćkyj*)  
 (-x) – brak formy podanej po minusie (defektywność): (-pl:gen), (-fchar), (-fneut) (np. *Brda*)  
 (pl) – odmiana wyjątkowa w liczbie mnogiej. Tylko A1+(pl) – *sędzia*  
 (sg) – odmiana wyjątkowa w liczbie pojedynczej. Tylko A1+(sg) – *paparazzo*  
 (num.) – odmiana wyjątkowa w specjalnych użyciach. Tylko B4+(num.) – *raz*  
 Pozostałe ciągi znaków zapisywane w nawiasie – odpowiednie litery w którejś z form, najczęściej w zakończeniu, np. (ęt), (ęc), (en), (on), (in), (es) (por. wyżej)  
 M – paradygmat (przymiotnikowy) ograniczony do typu męskiego (*m1* lub *m2*) – w wypadku braku symbolu dla typu A wartość ta jest domyślna  
 N – paradygmat (przymiotnikowy) ograniczony do rodzaju nijakiego  
 Ź – paradygmat (przymiotnikowy) ograniczony do rodzaju żeńskiego  
 P – paradygmat (przymiotnikowy) ograniczony do rodzaju przymnogiego osobowego (np. *wszyscy*)

Ponadto po separatorze występują dłuższe ciągi znaków dla odmiany nietypowej, supletywnej lub dwuczłonowej, np.:

- człowiek – wyjątek leksykalny (dla kilku leksemów)
- ystok – wyjątek leksykalny (dla *Białystok*)
- agłowa – wyjątek leksykalny (dla *białagłowa*)
- latka – wyjątek leksykalny (dla *roczek*)
- lata – wyjątek leksykalny (dla *rok*)
- ystaw – wyjątek leksykalny (dla *Krasnystaw*)
- (rzecz) – wyjątek leksykalny (dla *rzeczpospolita*)

Do paratypu E należą wzory nie pasujące do żadnego z paratypów A–D. Oprócz wzoru E0, obejmującego rzeczowniki nieodmienne, dzielą się one na dwie grupy.

Do pierwszej należą dwa wzory dla rzeczowników *plurale tantum*, odległe od wszystkich wzorów podstawowych typów – zostały oznaczone przez bezpośrednie wskazanie swoich reprezentantów (E+Niemcy, E+stwo)<sup>7</sup>.

Do drugiej grupy wchodzi wzory dla rzeczowników *singulare tantum* (są to jedynne takie rzeczowniki w języku polskim), o wykładnikach odmiany niepodobnych do

<sup>7</sup> Rzeczownik *plurale tantum* **Niemcy** nie pasuje ani do żeńskiego typu D (ze względu na pl:nom), ani do męskiego B (ze względu na pl:gen); rzeczowniki *plurale tantum* wzoru E+stwo mają we wszystkich formach końcówki charakterystyczne dla liczby pojedynczej.

rzeczowników wskazanych wyżej grup, nazywane tradycyjnie zaimkami rzeczownikowymi. Oznaczenia tych wzorów, podobnie jak innych wzorów rzeczownikowych, składają się z dwu segmentów rozdzielonych separatorem +. Druga część jest opcjonalna: podany jest w niej *explicite* postfiks (np. Eco+śkolwiek) lub prefiks (np. Ekt+ni); w pierwszej części, po klasyfikatorze E, podany jest podstawowy alternant tematu (np. Ekto+kolwiek, Ewasze).

### 3. Wzory zaimkowe (Z)

Stosownie do przyjętych w SGJP-online założeń jako odrębną grupę leksemów potraktowano zaimki osobowe. Każdy z sześciu należących do niej leksemów odmienia się według innego wzoru. Wzory te, reprezentujące specjalny „typ odmiany” (ppron), układają się w cztery typy: a (Zmy, Zwy), b (Zja, Zty), b' (Zsię), c (Zon).

### 4. Wzory przymiotnikowe (P)

Analizę deklinacji przymiotnikowej (przymiotnikowego typu odmiany) prowadzono, operując paradygmatem uproszczonym, obejmującym – w wariancie podstawowym – jednaście słów, którymi zostaje wypełniony paradygmat pełny.

Numeracja wzorów przymiotnikowych opiera się na tych samych zasadach co rzeczownikowych. Na pierwszym miejscu litera P wskazuje na przymiotnik.

Na drugim miejscu w symbolu wzoru przymiotnikowego występuje liczba z przedziału 0–4, wskazująca na sposób odmiany, przede wszystkim uwarunkowany charakterem końcowej części tematu (zob. Saloni 1992):

0 – leksem nieodmienny (np. *bordo*)

1 – temat na spółgłoskę miękką (np. *bielusi*)

2 – temat na spółgłoskę stwardniałą (np. *gorący*)

3 – temat na spółgłoskę tylnojęzykową (fonologicznie – /k/ i /g/, np. *egipski*)

4 – temat na spółgłoskę twardą (oprócz /k/ i /g/, np. *sławny*) [Tu zaliczono też ch – ze względu na reguły użycia *i* zmiękczonego.]

Po numerze grupy czasami podane jest zakończenie tematu charakterystyczne dla danego wzoru – podobnie jak w wypadku rzeczowników, zapisane jest ono w kolejności znaków *a tergo* (kolejność zwykła jest zastosowana w ujętych w nawias postfikach, np. *kolwiek*<sup>8</sup>). Występują tutaj (oprócz innych liter, nie wymagających komentarza) następujące znaki:

oę – w paradygmacie wyjątkowe formy o końcówkach +o, +ę (np. *to, tę*)

iV – litera *i* występująca po literze samogłoskowej (we wzorze PliV – *białozyi, burżu*)

Po podaniu informacji wskazujących na typ odmiany i zapis zakończenia omawianych wyrazów mogą pojawić się jeszcze następujące znaki:

<sup>8</sup> Wszystkie formy z postfiksem odpowiadają formom innego leksemu bez postfiksu.



- – brak w mianowniku rodzaju męskiego (l. poj.) formy zakończonej literą samogłoskową (np. *żaden*)
- + – istnienie formy spółgłoskowej w mianowniku męskich (l. poj., np. *łaskaw* obok *łaskawy*)
- ++ – istnienie formy spółgłoskowej w mianowniku podrodzajów męskich i bierniku podrodzaju *m3* (l. poj., np. *mamin*)
- ! – forma hasłowa zakończona na spółgłoskę (choć za poprawną uznajemy też formę zakończoną samogłoską, np. *wart* obok *warty*)
- \$ – występowanie form krótkich (np. *mego*) obok pełnych (np. *mojego*)
- ? – dana cecha wątpliwa (np. istnienie żeńskiej biernikowej formy l. poj. *tamtę* w paradymacie leksemu *tamten*)

We właściwej funkcji w końcowej części symbolu wzoru występują jeszcze dwie litery, które jednak odnoszą się do form specjalnych:

- o – końcówka formy bazowej<sup>9</sup> 5' (np. *samo* – w odróżnieniu od formy 5'': *same*)
- ę – końcówka formy bazowej 8' (np. *tę* – w odróżnieniu od formy 8'': *tą*)

Ponadto:

- def – wskazuje na paradygmat defektywny (*rad* i *kontent*).
- K – oznacza literę *k*, zachowaną w formie bazowej 9: *ki* (*ki* (*diabli*) – zamiast spodziewanego *cy*, por. *jacy* – od *jaki*)
- (en) – zakończenie formy spółgłoskowej ze spółgłoską nietematyczną (np. *ten*)
- (lz) – alternatywną formę 9 (*zmarzli* obok *zmarzli*)

Wzory przymiotnikowe układają się w większą liczbę typów: ich zróżnicowanie opiera się na występowaniu w pewnych wypadkach niestandardowego zestawu form.

## 5. Wzory liczebnikowe (L)

System numeracji wzorów liczebnikowych („typu odmiany” liczebnikowej) jest oparty na zasadach podobnych jak dla rzeczowników i przymiotników – z tą różnicą, że zamiast liter tematycznych wykorzystane są skojarzenia znaczeniowe.

Po pierwszym znaku L (liczebnik), następuje litera określająca typy wzorów liczebnikowych:

- a – typy liczebników nieodmiennych (wyodrębnione na podstawie różnego statusu formy niesamodzielnej, np. *pół*)
- b – odmienny tylko przez rodzaj (np. *półtora*)
- c – wzbogacony typ odmiany liczebnikowej (jak *dwa*)
- d – zasadniczy typ odmiany liczebnikowej (jak *pięć*), podzielony na podtypy ze względu na szczegóły paradygmatu pełnego

Do tworzenia nazw wzorów dla rzeczowników głównych zostały użyte liczby naturalne. Ściślej – w nazwie takiego wzoru na miejscach trzecim i dalszych stoi naj-

<sup>9</sup> Koncepcja form bazowych wraz z odpowiadającymi im oznaczeniami jest opisana we wstępie teoretycznym do SGJP.

niższa liczba naturalna, której nazwa odmienia się według tego wzoru, np.: Lc3, Ld5, Ld11, Ld20, Ld200 itd.

Nieco inną konwencję zastosowano do liczebników nazywających liczby w sposób przybliżony, np.: Ld1P (*paręnaście*), LdW0 (*wieledziesiąt*); nazwy czysto konwencjonalne to tylko: LdX (*ile*), LdT (*tyle*), LdW (*wiele*). Jeśli w formach leksemu występuje postfiks, jest on zapisywany po znaku +, np.: LdX+ś, Ld2+kroć (*dwakroć*).

Przy numerowaniu wzorów dla liczebników dwu pierwszych typów zastosowano najprostsze techniki mnemotechniczne, np.: Lapół, Lbpółtora.

## 6. Wzory czasownikowe (V)

Najbardziej skomplikowane jest przedstawienie wzorów czasownikowych, choć zasadnicza analiza wykładników koniugacyjnych została przeprowadzona wcześniej (przede wszystkim od dawna operowano konsekwentnie paradygmatami uproszczonymi, zawierającymi tylko formy bazowe), a obecnie ograniczono się do wprowadzenia nowych oznaczeń, dostosowanych do całego systemu. Wykorzystano tu doświadczenia Zygmunta Saloniego, który nad fleksją werbalną pracował przez 35 lat – od momentu, gdy w roku 1981 przejął od Jana Tokarskiego rękopis „Schematycznego indeksu polskich form wyrazowych” (Tokarski 1993).

Stwierdził wtedy, że pod względem adekwatności wobec materiału lepsza od późniejszej klasyfikacji Tokarskiego, z SJPDor. (1958), wyróżniającej tylko 18 grup i podgrup koniugacyjnych (nie licząc wyjątków), jest wcześniejsza – z książki „Czasowniki polskie” (1951), wyróżniająca ich 49. Tą też operował on roboczo w dalszych pracach. W celach popularyzatorskich jednak lepiej było odwoływać się do klasyfikacji drugiej jako lepiej znanej (zob. np. Saloni 2001). Właściwa analiza Saloniego szła w kierunku uściślenia systematyzacji Tokarskiego i doprowadziła do wyróżnienia 217 wzorów koniugacyjnych (zob. Saloni 2008). Symbole tych wzorów, stanowiące uszczegółowienie symboli Tokarskiego, były używane w wydaniu I i II SGJP. System ten miał tę wadę, że był bardzo trudny do przyswojenia. Autor przyznaje, że gdy widział symbol, rozumiał tylko jego początkową część, reszty – bez przykładu – nie pamiętał.

Z tych powodów dla internetowej wersji SGJP postanowiono opracować system symboli nawiązujących do oznaczeń z SJPDor. Jednak konieczne było dokonanie czysto formalnych zmian. Niewygodne było mianowicie użycie liczb rzymskich zastosowanych w SJPDor. i późniejszych słownikach – występują w nich bowiem litery alfabetu łacińskiego, co utrudnia automatyczne sortowanie (standardowe jest układanie napisów według kolejności alfabetycznej). Dlatego też trzy pierwsze znaki symbolu stanowią (oprócz czterech wyjątków podanych na końcu niniejszego punktu):

V – ogólny symbol wzorów czasownikowych (tj. opartych o paradygmat według trybów i czasów z syntetycznym bezokolicznikiem; symbol ten nie jest stosowany do czasowników niewłaściwych odmienianych analitycznie, jak *można*, ani do wtórnych czasowników odprzymiotnikowych, jak *powinien*)

xx – dwucyfrowa liczba z przedziału 01–19 (liczby mniejsze od 10 zapisywane są z zerem na początku)

Oznaczenia 01–11 odpowiadają dużym grupom I–XI z SJPDor. (pominięte są modyfikujące litery przy grupach V, VI, VII, VIII i X). Pozostałe numery zostały wykorzystane dla czasowników „bez cyfry [tj. numeru] grupy” w SJPDor., oczywiście odpowiednio pogrupowanych. Jednak pewne czasowniki, których odmiana niewiele odbiega od któregoś z wzorów I–XI, mają przyporządkowany odpowiedni niższy numer dwucyfrowy; specyfika tego wzoru podawana jest w dalszej części oznaczenia (od czwartego znaku).

Ta część oznaczenia wzoru zawiera charakterystykę alternacji występującej w należących do niego czasownikach, podanych – w miarę możliwości – za pomocą liter występujących w poszczególnych formach. Rozpada się ona na dwa segmenty rozdzielone separatorem + (drugi segment jest opcjonalny; występuje tylko wtedy, gdy zachodzi potrzeba).

W pierwszym segmencie (obligatoryjnym) notowany jest przyrostek tematowy (według Jana Tokarskiego), na ogół w dwu alternantach, przedzielonych kreską pionową: praes:sg:3p (tzw. temat czasu teraźniejszego, np. (*rys*)uje) i praet:sg:m:3p (tzw. temat czasu przeszłego, np. (*rys*)owa(*t*)); jeśli oba alternanty są identyczne, notuje się tylko jeden ciąg liter, np. symbol *V05nie|nq* odpowiada podgrupie *Va* w SJPDor., do której należy wiele leksemów (np. *ciągnąć*). Zdarza się, że segment ten jest redundantny w stosunku do wcześniejszej części oznaczenia, np. *V04uje|owa* – oznaczenie *IV* w SJPDor. jednoznacznie wskazuje grupę, która nie obejmuje żadnych odmianek czy podgrup, alternacja w niej jest uniwersalna (np. *rachować*). Zdarza się, że przyrostek tematowy podany w pierwszym segmencie (ze względów poglądowych, dydaktycznych) z punktu widzenia programu SGJP należy do trzonu fleksyjnego (tj. części powtarzającej się we wszystkich formach leksemu), np. *V01a* – litera *a* jest obecna we wszystkich formach czasowników grupy *I* (np. *chować*). Zdarza się też, że w praet:sg:m:3p występuje zerowy alternant przyrostka tematowego: w wypadkach takich po kresce | nie notuje się żadnej litery, np. *V05nie|* – symbol ten odpowiada głównemu wzorowi podgrupy *Vc* z SJPDor. (np. *chudnąć*). Na zasadzie wyjątku zastosowano – po kresce | – symbol 0 (zero) – dla czasowników defektywnych, nie mających formy praet:sg:m:3p (i innych form od tematu czasu przeszłego (*V03eje|0* : *goreć* – *gore*).

Drugi segment (po separatorze +) drugiej części wzoru czasownikowego notowany jest w wypadkach, gdy wcześniejsze jego elementy nie wystarczają do jego jednoznacznej identyfikacji. Notowany jest wtedy jeszcze jeden alternant przyrostka tematowego lub, szerzej, tematu czasownika, uznany za najbardziej charakterystyczny, np. symbol *V05nie|nq+ń* odpowiada podgrupie *Vb* w SJPDor., różniącej się od *Va* formą rozkaznika (*plyń* w opozycji do *ciągnij*). Starano się o to, aby forma notowana po + była najbardziej charakterystyczna dla alternacji występujących w danym wzorze, np. *V06si+szę* (np. *kusić*).

Systemu numerowania grup nie zastosowano do czterech wzorów (ośmiu leksemów) nie mających typowych wykładników koniugacyjnych, lecz przejęte z odmiany przymiotnikowej. Reprezentują one osobny typ *p*: odpowiadające im wzory składają

się z dwu segmentów: pierwszego VA i – po znaku + – formy męskiej danego leksemu: VA+winien, VA+rad, VA+gotów, VA+kontent.

Osobny typ (*b*) reprezentuje czasownik *być*, który jako jedyny w języku polskim ma równocześnie formy czasu przyszłego prostego *będzie* (i odpowiednio formy innych liczb i osób) oraz teraźniejszego *jest* (analogicznie). Jego odmiana od czasowników o wzorze V19*będzie/był* reprezentujących podstawowy typ czasownikowy [v] (np. *nabyć*) różni się tylko naddatkiem – form teraźniejszych; dlatego przypisano mu wzór V19*będzie/był+jest*.

Wszystkie pozostałe czasowniki (oprócz nieodmiennych syntetycznie, zwanych często predykatywami) reprezentują w istocie ten sam typ [v], choć w zależności od swych cech (aspekt, właściwość) mają różny zestaw form w paradygmacie. Wyodrębnienie osobnego typu 67 ma uzasadnienie techniczne (występowanie różnych alternantów w formie męskiej czasu przeszłego w zależności od pozycji cząstki aglutynacyjnej, np. *niosłem* – +*m niósł*).

## 7. Predykatywy (V0)

Inaczej niż w poprzednich wersjach SGJP jako osobny typ odmiany *pred* potraktowano predykatywy (np. *trzeba*). W artykułach hasłowych opisano je jednak jako czasowniki niewłaściwe i oznaczono – paralelnie do innych typów leksemów czasownikowych – symbolem V0.

## 8. Jednostki nieodmienne

System typów zastosowano także do jednostek uwzględnionych w słowniku i występujących w jednej tylko postaci tekstowej (z istotnym wyjątkiem wymienionym niżej). Uwzględniono nawet dwa typy odmiany tych jednostek: zwykłych leksemów nieodmiennych oraz skrótów, które trudno nawet uważać za leksemy języka polskiego: zamiast paradygmatu jest przy nich podawany sposób rozwinięcia.

Z kolei typ nieodmienny (ndm), obejmujący wszystkie leksemy nie należące do sześciu „typów odmiany” wymienionych w punktach 2–7, zawiera wzory dwóch typów. Oprócz typu [ndm] jest wprowadzony typ e, do którego należą przyimki mające zróżnicowane formy w zależności od kontekstu, np. *z/ze*. Zostały one w słowniku oznaczone symbolem Re, podobnie jak przedrostki czasownikowe typu *pod+/*pode+**.

## 9. Perspektywy

Zasadniczym celem przedstawionego wyżej systemu oznaczeń było zbudowanie hierarchii, która ułatwi zastosowania praktyczne. Możliwe powinno być zarówno znalezienie właściwego oznaczenia dla leksemu nieuwzględnionego w słowniku, jak i stworzenie nowego oznaczenia dopasowanego do całości (do pewnego momentu zawierającego elementy wprowadzone we wzorach istniejących).

Mamy bowiem przeświadczenie, że SGJP może być z powodzeniem rozwijany dla celów zarówno uczenia języka polskiego, jak i przetwarzania tekstów w języku polskim.

## 10. Pozycje cytowane

- Gruszczyński W., 1989, *Fleksja rzeczowników pospolitych we współczesnej polszczyźnie pisanej*, Wrocław.
- Gruszczyński W., 2001, *Rzeczowniki w słowniku gramatycznym współczesnego języka polskiego*, [w:] *Nie bez znaczenia... Prace ofiarowane Profesorowi Zygmuntowi Saloniemu z okazji jubileuszu 15 000 dni pracy naukowej*, red. W. Gruszczyński, U. Andrejewicz, M. Bańko, D. Kopcińska, Białystok, s. 99–116.
- Gruszczyński W., Saloni Z., 2006, *Notowanie informacji o odmianie rzeczowników w projektowanym „Słowniku gramatycznym języka polskiego”*, [w:] *Od fonemu do zdania. Prace dedykowane Profesorowi Romanowi Laskowskiemu*, red. I. Bobrowski, K. Kowalik, Kraków, s. 203–213.
- Saloni Z., 1992, *Rygorystyczny opis polskiej deklinacji przymiotnikowej*, [w:] *Zeszyty Naukowe Wydziału Humanistycznego Uniwersytetu Gdańskiego. Prace Językoznawcze*, t. 16, Gdańsk, s. 215–228.
- Saloni Z., 2001, *Czasownik polski*, Warszawa.
- Saloni Z., 2008, *Les modèles de conjugaison des verbes polonais*, [w:] *Questions de linguistique slave. Études offertes à Marguerite Guiraud-Weber, réunies par R. Roudet et Ch. Zaremba*. Aix-en-Provence; s. 279–292.
- SGJP 2007: Saloni Z., Gruszczyński W., Woliński M., Wołosz R., *Słownik gramatyczny języka polskiego*. Warszawa. [płyta CD + broszura, s. 177].
- SGJP 2012: Saloni Z., Woliński M., Wołosz R., Gruszczyński W., Skowrońska D., *Słownik gramatyczny języka polskiego*, wyd. 2, poprawione i uzupełnione, Warszawa. [płyta CD + broszura, s.~179, 2 nlb.].
- SJPDor. 1958–1969: *Słownik języka polskiego*, red. W. Doroszewski, t. 1–11, Warszawa.
- Skowrońska D., 2011: *Makowska D., Formy deprecjatywne w „Słowniku gramatycznym języka polskiego”*. Stan obecny i postulaty, *Polonica XXIX*, s. 205–226.
- Tokarski J., 1951, *Czasowniki polskie*, Warszawa.
- Tokarski J., 1958, *Formy fleksyjne*, [w:] *SJPDor.*, t. 1, s. XLIX–LXXIV.
- Tokarski J., 1973, *Fleksja polska*, Warszawa.
- Tokarski J., 1993, *Schematyczny indeks a tergo polskich form wyrazowych*, red. Z. Saloni, wyd. 2 – 2002, Warszawa.

## SUMMARY

### Systematization of inflectional patterns in the Grammatical Dictionary of Polish

Keywords: Inflectional patterns, Polish grammar, Grammatical Dictionary of Polish.

Słowa kluczowe: wzory fleksyjne, gramatyka polska, SGJP.

The purpose of the Grammatical Dictionary of Polish (sgjp.pl) in regard to inflection is to work out exact paradigms of all inflected lexemes of modern written Polish. These have been established for the most part in the previous versions of the Dictionary. The aim of the online edition of the Dictionary, its latest iteration, was to set up clear and easy-to-understand symbols for the inflectional patterns.

Laying the groundwork, we analyzed a number of regularly contrasted written forms, which we then grouped into lexemes. A wordform consists of a core, common to all wordforms belonging to a lexeme, and an end (these parts are fixed graphically and do not correspond to traditional terms: stem and ending). Lexemes are grouped into classes (these correspond roughly to traditional parts of speech). Lexemes in a class share similarities in their inflection, and thus paradigms. Classes are categorized into inflectional patterns. Lexemes categorized under the same inflectional pattern have identical end in all wordforms.

There are 6 general types of inflectional patterns: nominal, pronominal, adjectival, numeral, verbal and prepositional. Each inflectional pattern has its unique symbol. Each symbol consists of three parts. The first part is a letter, which denotes the general type of the inflectional pattern. The second part of the symbol is a sub-type of an inflectional pattern. The third part of a symbol shows alternations and non-standard phenomena in the paradigm of a given lexeme.