

Ida Kurcz Andrzej Lewicki Jadwiga Sambor
Krzysztof Szafran Jerzy Woronczak

Słownik frekwencyjny polszczyzny współczesnej. Wstęp*

Przedmowa

Oddając do rąk czytelników niniejszy tom, stanowiący pierwszy słownik frekwencyjny języka polskiego, czujemy się w obowiązku przedstawić pokrótce historię prac nad nim, tym bardziej że od ich rozpoczęcia upłynął bardzo długi okres. Był on po części wykorzystany na przeprowadzenie żmudnych i pracochłonnych czynności poszczególnych etapów realizacji zadania, ale po części składały się nań dosyć długie przestoje.

Słownik nasz jest rezultatem współpracy wielu osób i wielu instytucji naukowych, należących do różnych pionów organizacyjnych nauki polskiej czy nawet szerzej — polskiego życia intelektualnego. Współpraca ta czasem była zinstytucjonalizowana, czasem zaś miała charakter nieformalny. Słownik nie jest więc w istocie efektem podstawowej pracy określonej instytucji naukowej, lecz stanowi wspólne osiągnięcie środowiska, w którym został opracowany.

W pracach nad słownikiem w różnych ich fazach uczestniczyło wiele osób, wielu kolegów w różnym zakresie współpracowało z autorami, wielu kierowników placówek naukowych pomogło nam materialnie, udostępniając nieodpłatnie sprzęt komputerowy dla naszych zadań lub przyznając odpowiednie subwencje, konieczne do opłacenia prac technicznych i zakupu materiałów. Pragnęlibyśmy im wszystkim w tym miejscu gorąco podziękować, tym bardziej że omawiając dalej historię prac nad słownikiem, wymieniamy tylko niektórych. Przez tak wiele lat nie możemy nawet pamiętać nazwisk wszystkich naszych, czasem wysoko kwalifikowanych, współpracowników technicznych czy konsultantów, którzy istotnie przyczynili się do tego, że prezentujemy nasz tom w obecnej postaci.

Rozpoczęcie prac nad słownikiem frekwencyjnym datuje się na rok 1967. Wtedy właśnie stenograf śp. Władysław Masłowski zwrócił się do Jadwigi Sambor i Jerzego Woronczaka z prośbą o przeprowadzenie analizy frekwencji słownictwa w próbie polskich tekstów o długości 100 000 słów — dla potrzeb Komisji Systemowej Stowarzyszenia Stenografów i Maszynistek, pracującej nad optymalizacją polskiego systemu stenograficznego. W wyniku tej inicjatywy opracowano — przy aktywnym udziale projektodawcy — słownik dla polskiego stylu publicystycznego, bo teksty reprezentujące ten właśnie styl są najczęściej stenografowane. Już w momencie, gdy przystępowano do realizacji tego zadania, zdecydowano się potraktować analizowaną próbę jako fragment przyszłego słownika frekwencyjnego. Wyniki analizy opublikowano w roku 1972 dzięki pomocy Instytutu Badań Literackich PAN w tomie „Słownictwo współczesnej publicystyki polskiej”, traktując tę edycję jako zeszyt próbny przyszłego słownika frekwencyjnego.

Życzliwe przyjęcie naszego projektu oraz włączenie tematyki frekwencyjnej do problemów badawczych Instytutu Języka Polskiego PAN pozwoliły na opracowanie i wydanie list frekwencyjnych, obejmujących słownictwo korpusu o długości 500 000 słów, który miał stanowić podstawę przygotowywanego słownika. Listy te wydano w latach 1974–1977 w serii pięciu tomów (w jedenastu wolu-

* [Elektroniczna wersja wstępu do pracy:] Kurcz, Ida; Lewicki, Andrzej; Sambor, Jadwiga; Szafran, Krzysztof; Woronczak, Jerzy. *Słownik frekwencyjny polszczyzny współczesnej*. Kraków, 1990. Instytut Języka Polskiego PAN.

minach) pt. „Słownictwo współczesnego języka polskiego. Listy frekwencyjne” (dalej oznaczane jako *SWJP*) dzięki pomocy ówczesnego dyrektora IJP PAN, prof. Stanisława Urbańczyka, któremu zawdzięczamy edycję tych tomów. W finansowaniu prac pomagały również dorywczo następujące instytucje: Wyższa Szkoła Pedagogiczna w Olsztynie, Komitet Językoznawstwa Polskiej Akademii Nauk, Instytut Psychologii Uniwersytetu Warszawskiego.

Czynności, które doprowadziły do wydania *SWJP*, możemy umownie nazwać pierwszym etapem prac. Zostały one przeprowadzone przez zespół autorski w składzie: Ida Kurcz, Andrzej Lewicki, Jadwiga Sambor i Jerzy Woronczak. Opracowali oni koncepcję słownika, czuwali nad przebiegiem prac technicznych, sami wykonując wiele czynności pomocniczych, przeprowadzali korekty merytoryczne (również w następnych fazach pracy). Koncepcja statystycznej prezentacji materiału pochodzi od Jadwigi Sambor, koncepcja jego przetwarzania maszynowego — od Jerzego Woronczaka, który również był autorem użytych na tym etapie programów (istotną rolę odegrała tu konsultacja Jerzego Szczepkowicza). Od początku bowiem zakładano, że zasadnicze prace nad słownikiem zostaną wykonane z wykorzystaniem komputerów, i od początku materiał językowy był przetwarzany maszynowo. Prace techniczne na dalekopisach i innych urządzeniach peryferyjnych prowadzono w utworzonym specjalnie w tym celu laboratorium w Instytucie Filologii Polskiej Uniwersytetu Wrocławskiego (częściowo w Instytucie Badań Literackich PAN), zasadnicze prace obliczeniowe na komputerach — w Centrum Obliczeniowym Katedry Metod Numerycznych Uniwersytetu Wrocławskiego. Czujemy się tu w obowiązku wymienić nazwiska przynajmniej dwóch świetnych pracowników technicznych, przez cały czas realizacji pierwszego etapu związanych z naszym przedsięwzięciem: maszynistki Stefanii Bator i operatora Stanisława Rybarczyka.

Ogromny rozwój techniki komputerowej w czasie, gdy trwały prace nad słownikiem, stał się dla nas, autorów, przyczyną impasu w jego realizacji. Można tu powiedzieć, że nie nadążaliśmy za techniką.

Zakładaliśmy bowiem, że niniejszy słownik zostanie opracowany na komputerze, a następnie przeniesiony na sterowany automatycznie system poligraficzny (fotokład). W trakcie realizacji pierwszego etapu prac nad naszym słownikiem systemy takie zaczynały być stosowane w Polsce. Niestety, koncepcja przeniesienia materiałów na system fotokładu nie została wdrożona, głównie z powodu komplikacji i rozmiarów zadania oraz trudności w nawiązaniu współpracy z odpowiednimi placówkami przemysłowymi. Po roku 1977 prace zostały więc właściwie zawieszono. Wkrótce potem nastąpił rozwój innej dziedziny informatyki: mianowicie zaczęto redagować teksty na komputerach osobistych. Do samodzielnego opracowania słownika tą metodą dotychczasowy zespół autorski nie był przygotowany. Zakodowany (przede wszystkim na taśmach papierowych, tj. nośnikach informacji, które tymczasem wyszły z użycia, a dzisiaj są już całkiem przestarzałe) materiał został jednak w należyтым porządku przechowany. Był on też podstawą prac o charakterze pomocniczym, które okazały się w pewnym stopniu użyteczne w drugim zasadniczym etapie prac nad słownikiem frekwencyjnym. Pod kierunkiem Jerzego Woronczaka opracowano np. listę zbiorczą haseł zawartych w tomach *SWJP*.

Nasze tomy materiałowe — *SWJP* — zdobyły bowiem znaczną popularność w środowisku językoznawczym, zarówno ze względu na przydatność informacyjną, jak i walory metodologiczne. Stanowiły one pierwsze opracowanie statystyczne polskiego materiału językowego o dużej objętości, dążyły do konsekwentnego i maksymalnie zdyscyplinowanego jego przedstawienia. Te cechy słownika były widoczne dla wielu naszych kolegów, którzy zachęcali nas do kontynuowania i zakończenia prac nad polskim słownikiem frekwencyjnym.

Istotnie pomogli nam pod tym względem Zygmunt Saloni i Janusz Bień, którzy — widząc, że prace znalazły się w impasie — aktywnie szukali dróg jego przezwyciężenia. Zakodowane materiały przejął wtedy częściowo Instytut Informatyki Uniwersytetu Warszawskiego, w którym podjęto też próby przeniesienia ich na współczesne komputery i udostępnienia. Jednym z pierwszych elementów przezwyciężenia trudności było odczytanie, zinterpretowanie, a na-

stepnie wydrukowanie przez Mirosławę Miłkowską listy zbiorczej haseł zawartych w tomach *SWJP*, sporządzonej i zapisanej na komputerze, który w momencie odczytywania taśmy nie był już używany. Posiadanie owej listy na papierze umożliwiło nam przyjęcie ustaleń dla dalszej obróbki zgromadzonych materiałów i wstępne opracowanie planu dalszego działania.

Do właściwego opracowania niniejszego tomu, które możemy uważać za drugi etap prac nad słownikiem, przystąpiono w roku 1985. Wtedy właśnie wykonania wszystkich koniecznych prac komputerowych, które jeszcze pozostały, podjął się Krzysztof Szafran. Prace te wykonano w latach 1985–1990 w zasadzie w całości w Instytucie Informatyki Uniwersytetu Warszawskiego na sprzęcie stanowiącym jego własność (wyjątek stanowi drukarka laserowa, którą uprzejmie udostępnił nieodpłatnie Instytut Filozofii Uniwersytetu Warszawskiego). Prace te polegały na przeniesieniu materiałów poszczególnych tomów *SWJP* na współczesne nośniki informacji i udostępnieniu ich na aktualnie używanych komputerach, scaleniu list głównych, wyliczeniu i dodaniu parametrów liczbowych oraz na opracowaniu ostatecznej postaci listy głównej i sporządzeniu list dodatkowych (rangowych). Wypada dodać, że do niniejszego tomu włączono też materiał, którego początkowo nie przewidywano — dane o występowaniu haseł z listy głównej naszego słownika w korpusie tekstów polszczyzny mówionej. Uzyskała je na podstawie własnych badań i dostarczyła na dyskietce Halina Zgólkowa — pięknie jej za to dziękujemy.

Dodatkowym efektem przyjętego trybu pracy jest to, że zarówno cały słownik, jak i niektóre materiały pomocnicze i dodatkowe są zakodowane na nośniku maszynowym i mogą stanowić podstawę dalszych badań prowadzonych metodami komputerowymi.

Niniejszy tom pod względem edytorskim opracował również Krzysztof Szafran, posługując się pakietem programów edytorskich $\text{T}_{\text{E}}\text{X}$. Na ostateczny kształt wydawniczy tomu istotny wpływ wywarli pracownicy Instytutu Informatyki UW: Janusz Bień oraz Hanna Kołodziejka, służąc radą i doświadczeniem oraz proponując rozwiązania trudniejszych problemów typograficznych. Obowiązków redaktora podjął się Zygmunt Saloni. Wydanie tego tomu stało się możliwe dzięki specjalnej subwencji przyznanej nam przez Ministerstwo Edukacji Narodowej. Ze względu na częściowe finansowanie naszych prac w ostatnich latach przez Instytut Języka Polskiego PAN, a także tradycję sięgającą czasów wydania tomów *SWJP* Instytut ten podjął się roli wydawcy tomu dzięki życzliwości dyrekcji Instytutu w osobach profesorów Władysława Lubasia i Kazimierza Rymuta.

Tryb pracy nad słownikiem był też niewątpliwym powodem i tego, że — oddając go do rąk czytelników — widzimy w nim sporo niedoskonałości. Część z nich jest naturalną konsekwencją przyjętych przed z górą dwudziestu laty założeń, których czasem w dalszych fazach pracy nie można było zmienić. Czasem pożądana zmiana, teoretycznie możliwa do wprowadzenia, musiałaby pociągnąć za sobą nakłady pracy, czasu i kosztów zbyt wielkie w stosunku do spodziewanych rezultatów, tak że z żalem musieliśmy z niej zrezygnować.

Na jedną wadę naszego słownika zwróci z pewnością uwagę wielu jego użytkowników. Nasz materiał pochodzi mianowicie z lat 1963–1967 i w dwóch piątych zaczerpnięty jest z tekstów prasowych. Określając w ten sposób zasadniczy korpus materiałowy słownika, chcieliśmy osiągnąć jego jednorodność i reprezentatywność. Osiągnęliśmy ją, ale w stosunku do okresu, który w roku 1990 należy już całkowicie do przeszłości. I to jest pogładowa ilustracja ograniczoności metod statystycznych, przede wszystkim — problemu reprezentatywności próby. Mimo że nasz materiał wyraźnie stracił pod pewnym względem aktualność (nikomu np. nie radzilibyśmy na podstawie naszego słownika układać dla cudzoziemców ćwiczeń związanych tematycznie z życiem politycznym w dzisiejszej Polsce), postanowiliśmy opublikować go we wcześniej zaplanowanej postaci. Opracowanie tak dużego materiału w sposób jednolity i ścisły kosztowało wiele wysiłku, pieniędzy i czasu (przy najlepszej organizacji pracy i sprzyjających okolicznościach musiałoby ono pochłonąć kilka lat), wybranie materiałów współczesnych do opracowania nie zagwarantowałoby wcale, że otrzymany rezultat spełniałby oczekiwania czytelników. Opracowanie polskiego słownika frekwencyjnego

opartego na aktualniejszych materiałach to na pewno ciekawe zadanie o dużych rozmiarach — zadanie dla następnego zespołu autorskiego.

Mamy nadzieję, że mimo usterek tom, który oddajemy do rąk czytelnika, jako pierwszy ogólny frekwencyjny słownik języka polskiego wypełni dotkliwą lukę w polskiej leksykografii.

1. Dane ogólne

Przedstawiany słownik należy do typu słowników morfologiczno-leksykalnych, tzn. podaje informacje o częstościach nie tylko danego leksemu (hasła), ale i jego form fleksyjnych (słowoforn). Według zamierzeń autorów słownik ten winien reprezentować słownictwo ogólne, znane przeciętnemu Polakowi z wykształceniem średnim. Wzorowany jest na serii słowników frekwencyjnych języków romańskich¹; słowniki te, wydane w postaci frekwencyjnych list leksykalno-morfologicznych, stanowiły przełom w leksykografii statystycznej, także ze względu na syntetyczną charakterystykę liczbową haseł za pomocą określonych wskaźników.

Tom niniejszy, który poprzedziły wcześniejsze publikacje podające dane statystyczne na temat polskiego słownictwa, stanowi właściwy słownik frekwencyjny, a więc zawiera dane syntetyczne, dotyczące frekwencji słownictwa w całym badanym korpusie, obejmującym 500 000 słów. Korpus ten składa się z pięciu równych części, z których każda w pierwszej fazie pracy została zanalizowana osobno (zasady obliczania długości próby, czyli liczenia słów zostaną przedstawione w punkcie 3 niniejszego wstępu²). Wyniki tych analiz zostały opublikowane w latach 1974–1977 w serii pięciu tomów pt. *Słownictwo współczesnego języka polskiego. Listy frekwencyjne* (dalej — *SWJP*)³.

Zasady prezentacji materiału w naszym tomie pozostały w ogólnych zarysach takie same, jak w *SWJP*. Są one omówione w dalszych punktach niniejszego wstępu. Są tam również wskazane punkty, w których zasady te zostały zmienione (skutkiem dalszych korekt zmieniły się nieznacznie nawet dane liczbowe dotyczące częstości niektórych haseł i form fleksyjnych).

2. Dobór materiału

Wielkość przyjętej przez nas próby — 500 000 słów — była, historycznie rzecz biorąc, podyktowana przede wszystkim względami techniczno-finansowymi: w latach 1967–68 była to górna granica wielkości próby, jaką mogliśmy opracować maszynowo. Próby dokładnie tej samej długości były też podstawą statystycznych słowników Juillanda. Jak się okazało w trakcie pracy, ta wielkość próby była najdogodniejsza i ze względów ogólniejszej natury: była ona bowiem zbliżona do górnej granicy próby, którą w naszych warunkach zespół kilkuosobowy był zdolny starannie opracować.

Ponieważ zasadniczym celem naszych badań było podanie informacji o polskim języku pisanym, zgodnie ze standardami panującymi we współczesnej lek-

¹ A. Juilland, E. Chang-Rodriguez, *Frequency Dictionary of Spanish Words*. Mouton: The Hague – Paris 1964; A. Juilland, P. M. Edwards, I. Juilland, *Frequency Dictionary of Rumanian Words*. Mouton: The Hague – Paris 1965; A. Juilland, D. Brodin, C. Davidovitch, *Frequency Dictionary of French Words*. Mouton: The Hague – Paris 1970; A. Juilland, V. Traversa, *Frequency Dictionary of Italian Words*. Mouton: The Hague – Paris 1973.

² W wypadku tekstu tych rozmiarów obliczenie jego długości z dokładnością do jednostki nie jest możliwe, ponieważ na wynik mają wpływ rozstrzygnięcia i korekty, wykonane już po zamknięciu badanego korpusu. Na przykład według obliczeń Ireny Kamińskiej długość tekstu wprowadzonego do maszyny wynosiła odpowiednio dla poszczególnych części naszego korpusu: I — 100 265 słów, II — 100 267, III — 100 180, IV — 99 953, V — 98 631, a zatem dla całego korpusu — 499 296 (zob. I. Kamińska-Szmaj, Części mowy w słowniku i tekście pięciu stylów funkcjonalnych polszczyzny pisanej. *Biuletyn Polskiego Towarzystwa Językoznawczego*, z. XLI, 1988, s. 127–136).

³ Tom I. Teksty popularnonaukowe. [Autorzy:] Ida Kurcz, Andrzej Lewicki, Jadwiga Sambor, Jerzy Woronczak. Warszawa 1974, s. 4 nlb., 428 + 2 nlb., 429–858; Tom II. Drobne wiadomości prasowe. [Autorzy:] Ida Kurcz, Andrzej Lewicki, Jadwiga Sambor, Jerzy Woronczak. Warszawa 1974, s. 2 nlb., 398 + 2 nlb., 399–792; Tom III. Publicystyka. [Autorzy:] Andrzej Lewicki, Władysław Masłowski, Jadwiga Sambor, Jerzy Woronczak. Warszawa 1975, s. 4 nlb., 340 + 2 nlb., 341–684; Tom IV. Proza artystyczna. [Autorzy:] Ida Kurcz, Andrzej Lewicki, Jadwiga Sambor, Jerzy Woronczak. Warszawa 1976, s. 282 + 2 nlb., 283–557 + 1 nlb., 558–885; Tom V. Dramat artystyczny. [Autorzy:] Ida Kurcz, Andrzej Lewicki, Jadwiga Sambor, Jerzy Woronczak. Warszawa 1977, s. 320 + 2 nlb., 321–631. Trzeci spośród wymienionych tomów opublikowany został wcześniej jako osobna książka: Andrzej Lewicki, Władysław Masłowski, Jadwiga Sambor, Jerzy Woronczak, *Słownictwo współczesnej publicystyki polskiej. Listy frekwencyjne*. Warszawa 1971, s. 4 nlb., 583.

sykografii statystycznej, kwestią kluczową było dla nas wyróżnienie typowych stylów dzisiejszej polszczyzny pisanej. Ostatecznie — zgodnie z określoną koncepcją teoretyczną⁴). wyróżniliśmy pięć stylów funkcjonalnych, z których każdy reprezentowany jest w słowniku przez próbę tej samej długości 100 000 słów.

Za podstawowe uznaliśmy zróżnicowanie zdyscyplinowanych wypowiedzi pisanych na artystyczne, prasowe i naukowe, jednak w obrębie każdego z dwóch pierwszych typów wydzieliliśmy po dwa style funkcjonalne. Pod względem językowym dość wyraźnie różni się bowiem styl prozy artystycznej i dramatu artystycznego (wykorzystujący środki mówionego dialogu), istotne jest też zróżnicowanie stylu prasowej wypowiedzi informacyjnej i komentującej (publicystycznej). Ostatecznie przyjęty podział stylistyczny tekstów wyróżnia 5 grup tekstów, przynależnych do pięciu stylów funkcjonalnych języka pisanego:

- styl I: teksty popularnonaukowe,
- styl II: drobne wiadomości prasowe,
- styl III: publicystyka,
- styl IV: proza artystyczna,
- styl V: dramat artystyczny.

Teksty wchodzące w skład próby poddanej badaniom (zwanej dalej próbą roboczą) powinny bez żadnych zastrzeżeń reprezentować współczesną polszczyzną pisaną (różniąc się dostatecznie między sobą, by odzwierciedlać jej zróżnicowanie stylistyczne). W słowniku nie powinny też znaleźć wyraźnego odbicia indywidualne cechy tekstów (przede wszystkim ich tematyka). Próba robocza została dobrana metodą losowania z większej grupy tekstów, które będziemy nazywać kanonem podstawowym.

Aby nasza próba tekstów spełniała warunek jednorodności chronologicznej, terytorialnej i językowej, do kanonu podstawowego wybrano teksty spełniające następujące warunki:

1. teksty, których pierwodruki ukazały się w latach 1963–1967;
2. teksty wydrukowane w Polsce, których autorzy — rodzimi użytkownicy polszczyzny — zamieszkiwali stale w kraju (w wypadku, gdy trudno było to stwierdzić z pewnością, np. w wypadku anonimowych tekstów prasowych, przyjmowaliśmy takie założenie *a priori* z niewielkim ryzykiem pomyłki);
3. teksty napisane oryginalnie w obrębie systemu językowego współczesnej polszczyzny pisanej.

Ostatni warunek prowadził do wyeliminowania jako ewentualnej podstawy badań kilku kategorii tekstów: tekstów tłumaczonych z języków obcych, tekstów wierszowanych, tekstów stylizowanych (gwarowo, potocznie, poetycko itp.), tekstów przeznaczonych dla dzieci.

Aby uzyskać jak najbardziej neutralną próbę tekstów (zminimalizować wpływ ich tematyki oraz indywidualnych cech stylu autorów), zdecydowano przyjęć schemat próby oparty na bardzo licznym kanonie możliwie wielu tekstów. Za podstawowy odcinek tekstu poddany badaniu uznano fragment ciągłego tekstu o długości ok. 50 słów, czyli tzw. próbkę gronową. Przybliżone określenie długości próbki gronowej bierze się stąd, że starano się badać zdania w całości, nie zaś ich części (zdanie rozumie się czysto graficznie jako ciąg słów zawarty między dwiema kolejnymi kropkami lub znakami przestankowymi równoważnymi kropce).

Należało więc w obrębie pięciu stylów wyznaczyć próby po 100 000 słów. Otrzymano je, losując dla każdego stylu z bardzo dużej liczby tekstów, sta-

⁴ Jeśli chodzi o zróżnicowanie stylistyczne tekstów, stanowiących podstawę naszego opracowania, mogliśmy także wykorzystać doświadczenia autorów wielu wcześniej opracowanych słowników frekwencyjnych; por. serię słowników romańskich Juillanda (por. przypis 1.), słownik amerykańskiej angielszczyzny (H. Kučera and F. Nelson *Computational analysis of present-day American English*. Brown University Press: Providence, Rhode Island 1967), słownik czeski (J. Jelinek, J. Bečka, M. Tešitelova, *Frekvence slov, slovních druhů a tvarů v českém jazyce*. Praha 1961), słowacki (J. Mistrik, *Frekvencia slov v slovenscine*. Bratislava 1969; s. 726), słowniki rosyjskie (E. A. Steinfeldt, *Russian word count*. Moscow: Progress, b.r.w.; L. N. Zazorina (red.), *Czastotny slovar russkogo jazyka*. Moskwa: Russkij jazyk 1977.

nowiących kanon jego źródeł, 2 000 próbek gronowych — przeciętnej długości ok. 50 słów. Schematy losowania próbek gronowych są zróżnicowane w zależności od stylu i omówione pokrótce niżej w punktach 2.1–2.5.

Niżej omówimy ogólnie zasady wyboru tekstów do kanonu podstawowego i próby roboczej dla poszczególnych stylów. Bardziej szczegółowe informacje na temat sposobów losowania oraz wykaz źródeł (wraz z liczbą próbek losowych wziętych z każdego z nich) znajdują zainteresowani czytelnicy w odpowiednich tomach *SWJP*.

2.1. Teksty popularnonaukowe (styl I)

W zamierzeniach autorów nie znalazło się badanie stylu naukowego i terminologii specjalistycznej. Uznano natomiast za celowe uwzględnienie w słowniku stylu zbliżonego w znacznym stopniu do naukowego. Przyjęto mianowicie, że słownik ma odzwierciedlać słownictwo ogólne osób ze średnim wykształceniem, i postanowiono objąć badaniem teksty popularnonaukowe. Za najbardziej reprezentatywne dla prozy popularnonaukowej uznano wydawnictwa książkowe, przeznaczone dla szerokiego kręgu czytelników. Ustalono arbitralnie, że kanon źródeł nie będzie obejmować czasopism, podręczników, monografii przeznaczonych dla wąskiego grona specjalistów, encyklopedii, leksykonów, materiałów instruktażowych i skryptów oraz materiałów propagandowych.

Aby ustalić kanon tekstów popularnonaukowych, z którego zamierzano dokonać losowania 100 000 słów, na początku zestawiono na podstawie katalogów wydawniczych wykaz 602 książek odpowiadających przyjętemu warunkom. Trzeba było zadbać o należyte proporcje między tekstami reprezentującymi poszczególne typy nauk. Dla uproszczenia przyjęto podział odpowiadający strukturze wydziałów PAN z roku 1967. Wyodrębniono więc cztery grupy nauk: I. nauki społeczne, II. nauki biologiczne (biologia, medycyna, rolnictwo, hodowla), III. nauki matematyczno-fizyczne (matematyka, fizyka, chemia, astronomia, nauki o ziemi), IV. nauki techniczne. Ponieważ wyodrębniony z katalogów zbiór tytułów przeliczonych na liczbę stron wykazał ogromną przewagę tekstów reprezentujących nauki społeczne (stylistycznie najbliższych tekstom publicystyki), zdecydowano przyjąć arbitralnie w wylosowanej próbie jednakową reprezentację każdej z 4 wyodrębnionych w ten sposób grup nauk, losując odpowiednio po 25% próbek (co stanowiło 25 000 słów) tekstów należących do nauk społecznych, biologicznych, technicznych i matematyczno-fizycznych.

2.2. Drobne wiadomości prasowe (styl II)

Styl drobnych wiadomości prasowych uważa się za jeden ze stylów charakterystycznych dla współczesnej polszczyzny pisanej. Za reprezentatywne dla niego uznano teksty mające postać obiektywnych komunikatów, które występują we wszystkich działach gazet i czasopism: zagranicznym, krajowym, sportowym, kulturalnym itp. Wykluczono natomiast z kanonu źródeł wszelkie teksty zawierające jakiegokolwiek elementy komentarza autorów, a więc reportaże, komentarze, artykuły publicystyczne, a także nekrologi i ogłoszenia handlowe (mające schematyczną strukturę zdań i słownictwo).

Za najlepsze źródło tekstów drobnych wiadomości prasowych uznano dzienniki. Za kanon podstawowy przyjęto wszystkie teksty drobnych wiadomości prasowych opublikowane w pięcioleciu od 1 stycznia 1963 r. do 31 grudnia 1967 r. we wszystkich dziennikach (37 tytułów) wychodzących w tym okresie w Polsce. Jego badanie i dobranie próby przeprowadzono w sposób dość skomplikowany (szczegółowo omówiono je w II tomie *SWJP*). Na tym miejscu ograniczymy się do stwierdzenia, że starano się dokładnie odróżnić drobne wiadomości prasowe od innych materiałów, oszacowując ich wzajemną proporcję w pojedynczym numerze gazety, oraz — niezależnie — oszacowano długość tekstów zawierających wiadomości w danym tytule w badanym pięcioleciu. Na tej podstawie poszczególnym tytułom przypisano współczynniki określające liczbę numerów losowanych do próby. Dopiero po tych czynnościach przeprowadzono właściwe

dwustopniowe losowanie: numerów oraz próbek gronowych (po 5 na każdy wylosowany numer).

2.3. Publicystyka (styl III)

Za teksty najbardziej reprezentatywne dla stylu publicystyki uznano ogłaszane w tygodnikach i miesięcznikach artykuły o tematyce politycznej, gospodarczej, społecznej, kulturalnej, mające na celu wyrażanie i kształtowanie opinii publicznej i ujmujące problemy w sposób syntetyczny, a także stenogramy z posiedzeń organów państwowych i partyjnych (na decyzję tę istotny wpływ miał fakt, że nasze badania statystyczne nad słownictwem podjęliśmy z inicjatywy działacza Stowarzyszenia Stenografów i Maszynistek i miały one — w pierwszej kolejności — służyć optymalizacji polskiego systemu stenograficznego). Przy doborze tekstów stenogramów do badania zwracano uwagę na to, aby reprezentowały one pisaną odmianę polszczyzny i nie odzwierciedlały wyraźnych cech stylistycznych języka mówionego. Przy losowaniu tekstów stylu publicystycznego pominięto zarówno teksty o charakterze czysto informacyjnym, jak i felietony oraz reportaże, a więc teksty o ambicjach literackich, które brano pod uwagę przy ustalaniu kanonu podstawowego dla stylu prozy artystycznej.

Najogólniej rzecz ujmując, 600 próbek gronowych (50-wyrazowych) wylusowano z tekstów stenogramów, a 1 400 — z 13 czasopism, które mogły być uznane za najbardziej typowe dla okresu, w którym przeprowadzono badania. Zastosowano tu procedurę podobną do przyjętej dla stylu drobnych wiadomości prasowych: ustalenia liczby pozycji dla każdej z badanych części kanonu podstawowego (tytułu prasowego, grupy stenogramów), a następnie dwustopniowego losowania (pozycji — wewnątrz grupy, próbek gronowych — wewnątrz pozycji).

2.4. Proza artystyczna (styl IV)

Do kanonu podstawowego tekstów reprezentujących styl prozy artystycznej — obok typowych tekstów beletrystycznych o tematyce współczesnej (w tym powieści kryminalnych), historycznej i fantastyczno-naukowej — zaliczono reportaże, eseje i felietony. Nie włączono do niego tekstów dla dzieci, pamiętników, listów i tekstów publicystycznych. Zgodnie z zasadami ogólnymi pominięto też teksty poetyckie i wszelkie utwory stylizowane (archaizowane, dialektyzowane itp.). Za podstawowe źródła tekstów prozy artystycznej uznano wydawnictwa książkowe.

Dla opracowania kanonu podstawowego posłużono się *Adnotowanym Rocznikiem Bibliograficznym „Literatura Piękna”*, wydawanym przez Instytut Bibliografii Biblioteki Narodowej i Stowarzyszenie Bibliotekarzy Polskich, uwzględniając tomy obejmujące okres 1963–67⁵. Na jego podstawie sporządzono wykaz 1 070 książek spełniających nasze warunki ogólne (sformułowane wobec wszystkich tekstów zaliczanych do korpusu podstawowego) i szczegółowe (odnoszące się do tekstów zaliczanych do korpusu podstawowego stylu prozy artystycznej) — wraz z informacją o liczbie stron każdej pozycji. Książki te zawierały w sumie ok. 230 000 stron druku. Cały otrzymany w ten sposób tekst podzielono na 25-stronicowe odcinki, które stały się podstawą dwustopniowego losowania (odcinków, a następnie — próbek gronowych wewnątrz odcinków).

W próbie roboczej znalazły się ostatecznie fragmenty 751 książek. Na 2 000 próbek gronowych ok. 1 400 pochodzi z tekstów o tematyce współczesnej, ok. 250 próbek — z tekstów o tematyce historycznej, ok. 250 próbek — z reportaży i esejów, ok. 100 próbek — z tekstów o tematyce fantastyczno-naukowej. Ponieważ przestrzegaliśmy zasad metodologicznych losowania, mających zapewnić równomierny rozkład jednostek losowanych, możemy założyć, że stosunki te odzwierciedlają w jakiejś mierze strukturę tematyczną współczesnej polskiej prozy artystycznej.

⁵ Tomy te zostały wydane w latach 1965–1970.

2.5. Dramat artystyczny (styl V)

Za teksty reprezentujące styl dramatu artystycznego uznano teksty oparte na dialogu i przeznaczone do realizacji scenicznej. Zgodnie z ogólnymi założeniami słownika do badań wzięto wyłącznie dramaty napisane prozą i nie stylizowane (tj. takie, które można było uznać za reprezentujące współczesną polszczyznę ogólną).

Do korpusu podstawowego dramatu artystycznego zaliczono teksty opublikowane (w formie drukowanej lub wygłoszonej) po raz pierwszy w latach 1963–67 należące do następujących 4 grup: 1. teksty drukowane w czasopiśmie „Dialog”, 2. teksty dramatyczne drukowane w badanym okresie w książkach (oddzielnie i w wydaniach zbiorowych), 3. teksty teatralnych słuchowisk radiowych, 4. teksty dwu powieści radiowych: „Matysiakowie” i „W Jezioranach”.

Choć łączna objętość tekstów zaliczonych do poszczególnych grup była różna (najmniejsza była długość tekstów z „Dialogu”, największa — tekstów słuchowisk radiowych), wszystkie te grupy uznaliśmy za równie ważne przy analizie języka współczesnego polskiego dramatu artystycznego i przyznaliśmy każdej równy udział w próbie roboczej.

Skomplikowana struktura korpusu podstawowego tekstów dramatu artystycznego nastęrczała duże trudności przy opracowywaniu odpowiednich schematów losowania dla każdej z wymienionych grup tekstów. Jednostką losowania w grupach tekstów drukowanych był odcinek tekstu o określonej liczbie stron, w grupach tekstów radiowych — odcinek tekstu o długości określonej liczbą minut (czas trwania audycji na antenie w minutach był rejestrowany w katalogu audycji radiowych, udostępnionym nam przez Naczelną Redakcję Programów Literackich Polskiego Radia). Losowania samych próbek gronowych dokonano już z maszynopisów słuchowisk.

2.6. Polszczyzna mówiona

Materiał porównawczy dla polszczyzny mówionej włączono do naszego słownika na podstawie słownika frekwencyjnego Haliny Zgólkowej⁶ — na odpowiedzialność jego autorki. Jest to w tej chwili najbardziej wiarygodne źródło danych statystycznych na temat polszczyzny mówionej, opracowane w dodatku w pewnej mierze na wzór *SWJP*. Najistotniejsze jest tu to, że wielkość jego próby roboczej jest taka sama jak wielkość próby każdego z badanych przez nas pięciu stylów. Zasady doboru materiału, losowania próbek, opracowania danych itp. różnią się wprawdzie wyraźnie od przyjętych w *SWJP*, ale dla potrzeb niniejszego tomu autorka przystosowała kryteria wyróżnienia jednostek do zasad w nim obowiązujących. Dokładna charakterystyka materiału zawarta jest we wstępie do słownika Zgólkowej.

3. Jednostki badania

Jednostką wyjściową w naszych badaniach jest *słowo* — podstawowa jednostka tekstu rozumianego unilateralnie (tj. niemal jako obiekt fizyczny, bez uwzględniania znaczenia). Rozumienie techniczne terminu *słowo* pokrywa się przy tym z jego rozumieniem potocznym — słowem jest taki ciąg liter, które pisze się razem, bez odstępów. W sposób mechaniczny słowa są identyfikowane na podstawie swej postaci graficznej. W tekście są z reguły wyraźnie wzajemnie pooddzielane za pomocą separatorów słów (odstęp międzywyrazowy, znaki przestankowe i pomocnicze, jak nawias czy cudzysłów). Za separatory słów nie uważa się łącznika (np. jako jedno słowo traktujemy napis *kulturalno-oświatowy*) ani apostrofu (np. napis *Kennedy'ego* traktowaliśmy jako jedno słowo).

⁶ Halina Zgólkowa, *Słownictwo współczesnej polszczyzny mówionej. Lista frekwencyjna i rangowa*. Poznań: Uniwersytet im. Adama Mickiewicza w Poznaniu, 1983; 367 s.

3.1. Operacje wstępne

Założenie, że w kanonie podstawowym mamy do czynienia z tekstami składającymi się tylko i wyłącznie ze słów, było pewnym uproszczeniem. W istocie bowiem próbki gronowe wylosowane do badań zawierały fragmenty, które trudno było — bez skomplikowanych operacji definicyjnych — uznać za słowa czy ciągi słów współczesnego języka polskiego. Były to przede wszystkim napisy nieliterowe oraz mieszane — złożone z liter i symboli nieliterowych. Przed wprowadzeniem tekstu do maszyny należało go doprowadzić do postaci takiej, by był on rzeczywiście ciągiem słów polskich. W tym celu konieczne było pozbycie się elementów takich, jak wskazane wyżej. Traktowano je różnie w zależności od ich charakteru.

W trakcie przepisywania tekstu na nośnik informacji wprowadzano więc następujące modyfikacje:

1. Liczby pisane cyframi arabskimi i rzymskimi przepisywano słownie, używając formy wyrazów odmiennej wynikającej z kontekstu (np. odpowiedniego przypadku liczebnika), np.: „sformułował 53 twierdzenia” → *sformułował pięćdziesiąt trzy twierdzenia*, „po XX zjeździe” → *po dwudziestym zjeździe*.
2. Część cyfrową napisu złożonego z cyfr i liter przepisywano słownie, np.: „50-lecie” → *pięćdziesięciolecie*, „3-dniowy” → *trzydniowy*.
3. Konwencjonalne skróty nazw jednostek miar rozwiązywano, stosując pełny zapis formy wynikającej z kontekstu, np.: „kg” → *kilogram*, „cm” → *centymetr*, „godz.” → *godzina*, „o godz. 9” → *o godzinie dziewiątej*, „w 1966 r.” → *w tysiąc dziewięćset sześćdziesiątym szóstym roku*.
4. Ogólnie stosowane skróty wyrazów i wyrażeń języka polskiego o charakterze nienazwowym rozwiązywano, stosując — jeśli skracany był wyraz odmienny (np. rzeczownik) — formę wynikającą z kontekstu, np.: „prof.” → *profesor*, „dla dra Kowalskiego” → *dla doktora Kowalskiego*, „np.” → *na przykład*, „itd.” → *i tak dalej*, „br.” → *bieżącego roku*, „itp.” → *i tym podobne*.

Wprowadzone w ten sposób do korpusu elementy traktowano dalej jako normalne słowa lub ciągi słów języka polskiego i poddawano dalszym badaniom na tych samych zasadach, co słowa wprowadzone do korpusu przez zwykłe przepisanie zdania zawartego w próbce gronowej. Taka decyzja miała niewątpliwie wpływ na uzyskane rezultaty liczbowe (por. w słowniku hasła **tysiąc** albo **dziewięćset** — zwłaszcza w zestawieniu z **osiemset** i podobnymi). Oczywiście postać słowa w korpusie zależała w wielu wypadkach omawianych w punktach 1–4. od osoby przepisującej tekst, np.:

„Taka liczba słów zajmuje ogółem **ok. 8 000** stron.” →

*Taka liczba słów zajmuje ogółem **około ośmiu tysięcy** stron.*

albo

*Taka liczba słów zajmuje ogółem **około osiem tysięcy** stron.*

Każdy z tych wariantów rozwiązania napisów skrótowych i liczbowych pociąga za sobą inną interpretację nie tylko słowa odpowiadającego napisowi „8”, lecz także inną interpretację słowa *około*.

W odróżnieniu od skrótów wymienionych wyżej w punktach 3. i 4. skróty będące nazwami własnymi i wszelkie tzw. skrótowce (literowce, głoskowce, sylabowce) traktowano jako słowa języka polskiego. Nie rozwiązywano ich zatem i wprowadzano do korpusu bezpośrednio, np.: „bombardowanie DRW” → *bombardowanie DRW*, „rocznica PKWN” → *rocznica PKWN*, „prezydium PAN” → *prezydium PAN*, „w PAN-ie” → *w PAN-ie*, „w PGR-ach” → *w PGR-ach*, „Pafawagiem” → *Pafawagiem*.

Pozostałe napisy niestandardowe z punktu widzenia zasobu wyrazowego języka polskiego, zwłaszcza wzory i symbole, np.: „ H_2O ”, „odcinek AB” itp., usunięto z tekstu badanego, skreślając je z próbek.

3.2. Słowo i słowoforma

W opracowaniach językoznawczych (zarówno na poziomie elementarnym, jak i bardzo zaawansowanym) słowa są tylko punktem wyjścia do definiowania i badania jednostek, które uważa się powszechnie za bardziej istotne elementy języka naturalnego. Słowa nie odpowiadają im dokładnie. Powszechnie więc jako różne traktuje się na przykład jednostki językowe reprezentowane przez słowo *lata* w zdaniach:

- (1) *Ptaszek lata po ulicy.*
- (2) *Od zdarzeń tych minęły długie lata.*

W zdaniu (1) *lata* to oczywiście forma czasownika, a w zdaniu (2) — forma rzeczownika, ich znaczenie jest w sposób oczywisty całkiem różne. Aby przeprowadzać obliczenia interesujące z punktu widzenia języka polskiego, obie te jednostki należy więc rozróżnić. Tymczasem komputer, operując tekstem złożonym ze znaków, niemal w tej postaci, w jakiej występują oryginalnie w tekstach, nie miałby podstaw, aby obie te jednostki rozróżnić, ponieważ składają się one dokładnie z tych samych liter.

Ponieważ postać zewnętrzna nie stwarza podstaw do ich rozróżnienia, należało je odróżnić przy pomocy dodatkowych elementów. W słowniku naszym wprowadzono — zgodnie z praktyką ścisłego opisu językoznawczego, skanonizowaną w pracach strukturalistów — jednostkę językową bardziej skomplikowaną — bilateralną (dwustronną), która stała się punktem wyjścia do dalszych badań. Jednostka ta, zwana najczęściej formą wyrazową, jest definiowana jako uporządkowana para $\langle f, s \rangle$, której pierwszy element f jest nazywany formą, a drugi element s — znaczeniem. W tekstach jako obiekt materialny występuje tylko element f , w naszym wypadku z reguły ma on postać słowa. Element s zawiera te treści, które są komunikowane za pomocą elementu f . Owe treści muszą być rozumiane szeroko: należą do nich nie tylko odesłania do rzeczywistości pozajęzykowej (treści semantyczne), lecz również elementy decydujące o funkcji danej jednostki w tekście (treści gramatyczne): w naszym opracowaniu przede wszystkim te drugie będą uwzględniane przy przechodzeniu od jednostek unilateralnych do bilateralnych. Mówimy, że jednostka unilateralna f jest *wykładnikiem tekstowym* (lub krótko, *wykładnikiem*) jednostki bilateralnej $\langle f, s \rangle$.

W badaniach językoznawczych, także w badaniach statystycznych nad leksyką języka naturalnego, znacznie istotniejsze niż informacje o jednostkach unilateralnych są informacje o jednostkach bilateralnych, dlatego też we wszystkich naszych pracach one były przedmiotem operacji i obliczeń. Podstawowe jednostki bilateralne wprowadzane do pamięci komputera na podstawie tekstu próby roboczej nazywamy *słowoformami*. Słowoformy stanowią przybliżenia form wyrazowych i są od nich odróżniane z powodów technicznych — ze względu na elementy konwencjonalności przy ich określaniu. Słowoforma składa się z dwóch części: przeniesionego bezpośrednio z tekstu słowa oraz jego interpretacji, mającej najczęściej postać symbolu gramatycznego. Ten drugi element słowoformy (s) był przy tym dopisywany do słowa w postaci konwencjonalnej i bardzo uproszczonej — może to być uznane za drugorzędny szczegół techniczny. Istotne jest to, że do danego słowa dopisuje się elementy dodatkowe. Dzięki dopisywaniu tych elementów wypracowano właściwą procedurę postępowania w wypadkach, gdy to samo słowo raz powinniśmy interpretować jako wykładnik jednej słowoformy, a drugim razem — jako wykładnik innej słowoformy.

Takie, mające co najmniej dwie interpretacje, słowo nazywamy *słowem homonimicznym*, a słowoformy, których wykładnikiem tekstowym jest to samo słowo — *homonimami*.

Aby rozróżnić dwie słowoformy reprezentowane w tekście przez słowo *lata*, wystarczyłoby w zasadzie dopisać dowolne symbole odróżniające, np.: $lata_1$ dla *lata* ze zdania (1) i $lata_2$ dla *lata* ze zdania (2). Nie byłoby to jednak wygodne, ponieważ prowadziłyby do traktowania każdego wypadku indywidualnie. Tymczasem procedura interpretowania słów występujących w naszym korpusie głównym musiała być systematyczna i metodyczna.

Zgodnie z tradycją i powszechną praktyką współczesną nie opisujemy w na-

szym słowniku bezpośrednio słowoform, lecz jednostki wyższego rzędu, które mogą być rozumiane jako zbiory słowoform. Są to wyrazy w sensie słownikowym (leksemy). W prawie żadnym realnie istniejącym rejestrze (słowniku) wyrazów polskich nie jest zanotowany bezpośrednio wyraz *psu*, choć takie słowo występuje niewątpliwie w tekstach polskich. Nikt, kto ma pewne doświadczenie w posługiwaniu się słownikami języka polskiego, nie powie jednak na tej podstawie, że tego wyrazu w owym rejestrze nie ma: z reguły jest w nim bowiem zanotowany wyraz **pies**, a *psu* jest traktowane jako regularna forma wyrazowa (celownik liczby pojedynczej) wyrazu słownikowego (leksemu) **pies** — występuje zatem w słowniku pośrednio (w uwikłaniu). Leksemy języka naturalnego (jednostki jego słownika) mogą być traktowane jako zbiory form wyrazowych.

Słowoformy, będące w naszym słowniku przedmiotem czynności obliczeniowych, traktowaliśmy właśnie jako należące do określonych jednostek stanowiących przybliżenia leksemów języka polskiego. Zostały więc one pogrupowane w artykuły hasłowe opisujące jednostki wyższego rzędu. Ponieważ ich wyodrębnienie odbiega niekiedy od postulatów językoznawczych stawianych przy wyodrębnianiu leksemów, owe zasadnicze jednostki opisu, którym poświęcone są artykuły naszego słownika, nazywamy *hasłami*. Hasła w naszym słowniku są zbiorami słowoform. Zasadniczą kwestią przy opracowaniu słownika było zinterpretowanie każdego słowa występującego w materiale źródłowym jako słowoformy należącej do odpowiedniego hasła — czynność tę w praktyce leksykograficznej nazywa się hasłowaniem.

Symbole gramatyczne, dopisywane przy interpretowaniu słów jako słowoform, miały na celu nie tylko umożliwienie rozbicia słów homonimicznych na homonimy, lecz również ułatwienie hasłowania, a także wskazanie funkcji gramatycznej badanej słowoformy. Na przykład przy wprowadzaniu materiału do maszyny słowo *zła* było opatrywane symbolami w sposób następujący: *zła 211* i *zła 121* — w zależności od tego, czy występowało w kontekście typu:

(3) *To jest zła wola.* → *To jest zła 211 wola.*

czy typu:

(4) *Wyrządził im wiele zła.* → *Wyrządził im wiele zła 121.*

Cyfra 2 na pierwszej pozycji symbolu gramatycznego oznacza przymiotnik, cyfra 1 — rzeczownik (cyfra na drugiej pozycji oznacza przypadek, cyfra na trzeciej pozycji — liczbę). Na tej podstawie słowoforma *zła 211* była zaliczana do hasła **zły 2**, a słowoforma *zła 121* — do hasła **zło 1**. Przed wprowadzeniem materiału do maszyny został opracowany ogólny system oznaczeń, umożliwiający systematyczne poklasyfikowanie materiału. Pierwotnie używane symboliczne oznaczenia cyfrowe zostały w niniejszym tomie zastąpione skrótami literowymi o charakterze mnemotechnicznym.

Można to nawet zilustrować pierwszym z wprowadzonych wyżej przykładów. W obecnym słowniku słowo *lata* ze zdania (2) jest wprowadzone w postaci *lata Npl* w hasle **lata S** (pierwotna postać, występująca też w *SWJP*: *lata 112*), a słowo *lata* ze zdania (1) — w postaci *lata* [bez żadnego dodatku] w hasle **latać** (pierwotna postać: *lata 5*). Jak widać, zdarza się, że drugi element słowoformy (symbol gramatyczny) jest w zapisie pominięty; jednak *implicite* jest on zawsze obecny: słowoformy są zawsze traktowane jako należące do określonych hasel.

3.3. Decyzje segmentacyjne

Zasady otrzymywania bilateralnych słowoform z unilateralnych słów w trakcie przygotowywania przez nas materiału do wprowadzenia do maszyny oraz szkielet koncepcji gramatycznej, na której się one opierają, przedstawimy dalej (punkt 4). Zanim można było przystąpić do interpretacji słów, należało bowiem podjąć decyzje segmentacyjne, dotyczące wyjątków od podziału na słowa danego w tekstach próby roboczej. Nie przewidywaliśmy sytuacji, by jako słowoformę interpretować część właściwą słowa (niektóre konsekwencje tej decyzji pokazemy niżej w punkcie 3.3.2).

Dopuszczaliśmy sytuację odwrotną. W niektórych wypadkach uznawaliśmy — zgodnie z praktyką powszechną w pracach językoznawczych — że słowoforma występuje w tekście nie w postaci jednolitego słowa graficznego (zawartego między dwoma separatorami), lecz obiektu unilateralnego dłuższego niż słowo. Jako jednolite formy wyrazowe w pracach tych traktuje się bowiem często wielowyrzowe nazwy pojęć (np. *biały kruk*, *maszyna do pisania*, *Stany Zjednoczone*), w tym tzw. liczebniki złożone (np. *dwadzieścia pięć*, *dwudziesty piąty* czy *sto dwudziesty piąty*), tzw. formy analityczne czy złożone (np. w języku polskim: formy czasu przyszłego czasowników niedokonanych, np. *będzie jechał* i *będzie jechać* albo formy trybu warunkowego typu *byłbym jechał* i *bym jechał*) oraz liczne ustabilizowane połączenia funkcjonujące jako nierozdzielne całości (np. *na pewno*, *w ogóle*) itp. Nasze decyzje przedstawimy niżej w punktach.

3.3.1. Zestawienia

Przygotowując materiał do badań, starano się — ze względu na przyjęte założenie maksymalnej ścisłości i precyzji — w możliwie najmniejszej liczbie wypadków odstępować od danej z zewnątrz segmentacji na słowa. Jako ciągi słów potraktowano przede wszystkim te ustabilizowane wyrażenia, których przynajmniej jeden człon wchodzi w opozycje fleksyjne z innymi jednostkami. Tak więc przy przyjętych przez nas założeniach *biały kruk* to 2 słowa (bo *biały kruk* — *białego kruka*), a *sto dwudziesty piąty* — 3 (bo *sto dwudziesty piąty* — *sto dwudziestego piątego*). Nieliczne wyjątki od tej zasady wymienione są w punkcie 3.3.4.

3.3.2. Formy czasownikowe

Dlatego też człony złożonych form czasowników (np. *będzie* i *pisać* w zdaniu *Będzie pisać listy* albo *bym* i *napisał* w zdaniu *Wczoraj bym napisał list*) traktowano jako samodzielne jednostki — słowa, które poddawano odrębnej interpretacji, uznając je za formy wyrazowe (w tym wypadku jednak sygnalizowano za pomocą symbolu gramatycznego ich specjalną funkcję: części złożonej formy czasownikowej — por. niżej). Tak samo — jako złożone z dwóch słowoform — potraktowano analityczne formy trybu rozkazującego, typu *niech pisze*, do słowa *pisze* dopisywano jednak odpowiedni symbol gramatyczny.

Nieco inną procedurę stosowano wobec nieciągłych form czasu przeszłego i trybu warunkowego o pierwszym członie przyłączonym ortograficznie do innej jednostki tekstowej, np. w zdaniach: *Kiedyśmy przyjechali?* albo *Bobym się obraził!* W wypadkach takich przy przygotowaniu danych słowo *przyjechali* albo *obraził* było opatrywane specjalnym symbolem gramatycznym, który sygnalizował, że w tekście występowało ono razem z tzw. ruchomą końcówką czasownikową. Ową pomocniczą cząstkę formy czasownikowej („końcówkę”) wpisywano do pamięci maszyny wraz ze słowem, do którego była przyłączona ortograficznie. Konsekwencje tej decyzji są widoczne w opracowanych materiałach, włącznie z niniejszym tomem. W skład hasła **kiedy** *Adv* wchodzi mianowicie słowoformy: *kiedy*, *kiedym* i *kiedyśmy*. Podobnie w skład hasła **aby** *Cnj* wchodzi słowoformy: *aby*, *abym*, *abyś*, *abyście* i *abyśmy*. Niekiedy, gdy dołączenie cząstki do słowa powodowało homonimie, dla ujednoznaczenia poprzedzona jest ona łącznikiem, np. *co-s A* w hasle **co** *Prn* (w zdaniu: *Powiedz, coś widział!*), *jaka-s* w hasle **jaki** *Prn* (w zdaniu: *Jakaś ty dziwna!*).

W trakcie tych czynności, dopisując symbol gramatyczny do formy czasownikowej, nie uwzględniano informacji o jakości cząstki dołączanej. Symbol *fut* dopisywano więc do *pisać* jako składowej czasu przyszłego bez względu na to, czy łączyło się z *będzie*, czy z *będziemy*, czy z *będq*, czy z jeszcze innym słowem posiłkowym, a słowoforma zanotowana w słowniku *przyjechali mob* odpowiada złożonej formie czasu przeszłego — bez względu na to, czy to osoba 1., czy 2., a więc zarówno *+śmy przyjechali*, jak i *+ście przyjechali*.

Jako połączenia dwóch słowoform traktowano też tzw. stronę bierną, np. *jest obserwowana*, *został posłany*. I w tym wypadku słowo posiłkowe (w przykładach: *jest*, *został*) było opatrywane symbolem gramatycznym (*aux*).

3.3.3. Cząstka *się*

Cząstkę *się* uznawano na ogół za osobne słowo.

Wyjątkiem były sytuacje, w których *się* występowało w połączeniu z formą imiesłowu przymiotnikowego, typu *znajdującą się*, albo regularnego rzeczownika odsłownego, typu *pojawienie się* — połączenia takie traktowano jako wykładniki jednolitych form wyrazowych. Formy takie zaliczono do odpowiednich hasel, w rozważanych wypadkach: **znajdujący się** i **pojawienie się**. Występujące w składzie odpowiednich słowoform *się* nie zostało wliczone do hasła **się**.

Pozostałe wystąpienia słowa *się* zostały uznane za słowoformy, należące do hasła **się** *Prn* lub — w rzadkich wypadkach — do hasła **siebie**. Wszelkie połączenia form czasownikowych (nie imiesłowowo-przymiotnikowych i nie gerundialno-rzeczownikowych!) uznano zatem za dwa słowa, np. nie tylko *bili się* → *bili i się*, *bić się* → *bić i się* oraz *śpi się* → *śpi i się*, ale także *bali się* → *bali i się*, *bać się* → *bać i się*.

Rozróżniono przy tym dwie sytuacje.

Pierwsza z nich, rzadsza i mniej typowa, dotyczy wypadków, gdy **się** ma funkcje bezosobową: wyklucza ono wtedy wystąpienie w zdaniu podmiotu mianownikowego i może się pojawić nawet przy czasownikach nieprzechodnich, np. *śpi się*, por.:

(5) *Najlepiej jej się śpi nad ranem.*

Takie *się* zostało uznane za pełniące samodzielną funkcję i niezależne od formy czasownikowej, z którą się wiąże. Zostało ono wprowadzone do hasła **się** *Prn* jako odrębna słowoforma *się* *N*. Formy czasownikowe, z którymi się ona wiąże, uznano za niezależne od niego i nie oznaczano w żaden sposób tego, że w tekście były one z nim powiązane.

Inaczej postąpiono w drugiej sytuacji, obejmującej czasownikowe formy zwrotne (w szerokim sensie). Zaliczono do nich wszystkie połączenia form czasownika z *się* nie pełniącym funkcji bezosobowej i nie blokującym pozycji mianownikowej w zdaniu, np.:

(6) *Chłopcy ciągle się bili.*

(7) *Chłopcy musieli się bić.*

(8) *Chłopcy ciągle się bali.*

(9) *Chłopcy musieli się bać.*

Takie *się* zostało uznane za wykładnik innej słowoformy, wprowadzanej do hasła **się** *Prn* w postaci: *się* (bez symbolu gramatycznego). Formy czasownikowe, przy których występowało takie *się*, uznawano za specjalne słowoformy i oznaczono symbolem zwrotności (w tekście wpisywanym do maszyny: cyfra 1 na trzecim miejscu w liczbie, rozpoczynającej się od cyfry 5, oznaczającej czasownik). W niniejszym tomie symbole gramatyczne owych form zastąpił symbolem: [*się*]. Wszystkie słowoformy czasownikowe zawierające symbol zwrotności, a zróżnicowane fleksyjnie, zgrupowano w jednym hasle, w którego nazwie figuruje element **się**. Na przykład słowoforma *bali* [*się*] została zaliczona do hasła **bać się** *V*, a słowoforma *bili* [*się*] — do hasła **bić się** *V* — w odróżnieniu od słowoformy *bili* (występującej np. w zdaniu *Chłopcy bili dziewczynki*), zaliczonej do hasła **bić** *V*. Tak więc w słowniku naszym znajduje się wiele par hasel czasownikowych, których nazwy różnią się obecnością cząstki *się*.

Jako słowoformę hasła **siebie** potraktowano *się* powiązane z przyimkiem o rządzie biernikowym, np.:

(10) *Sam przez się nie ma wartości.*

3.3.4. Wyjątki

Od zasad ortograficznej segmentacji tekstu poczyniono pewną liczbę wyjątków, gdy wbrew ortografii jako wykładnik jednej słowoformy traktowano napis zawierający wewnętrzny odstęp (wtedy słowoforma składa się z większej licz-

by słów niż jedno)⁷. Jako jednolite słowoformy traktowano więc następujące wyrażenia:

1. połączenia imiesłowów przymiotnikowych i regularnych rzeczowników od-słownych z cząstką *się*, np. *odnoszące się, uchylaniem się* (por. wyżej — punkt 3.3.3.);
2. nazwiska zawierające pisaną osobno cząstkę typu *de, von*, np.: *de Gaulle, von Hassel*;
3. połączenia cząstek *co* i *jak* z przysłówkami i przymiotnikami w stopniu naj-wyższym (które — przy tym ujęciu — trzeba by raczej nazwać słowami homonimicznymi z przysłówkami i przymiotnikami w stopniu najwyższym), np.: *co najmniej, co najwyżej, co najgorsze, co najważniejsze, jak najbardziej, jak najlepiej, jak najprędzej, jak najszybciej, jak najlepszy, jak największy, jak największej* (takie słowoformy są opatrzone odpowiednimi symbolami gramatycznymi i zaliczone do jednolitych haseł, np.: **co najmniej** *Prt*, **jak najbardziej** *Adv*, **jak największy** *Adj*; część z nich to hasła przymiotni-kowe — o zróżnicowanych słowoformach);
4. skostniałe połączenia przyimków (czy też słów homonimicznych z przyimka-mi) z formą o nieokreślonym przypadku (na ogół słowem homonimicznym z przysłówkiem lub partykułą), a mianowicie:
 - a) *do*, np.: *do dzisiaj, do dziś, do wewnątrz*;
 - b) *na*, np.: *na długo, na dzisiaj, na dziś, na krótko, na lewo, na nowo, na pewno, na pół, na prawo, na próżno, na wół, na wprost, na zawsze, na zewnątrz*;
 - c) *od*, np.: *od dzisiaj, od kiedy, od wczoraj*;
 - d) *w*, np.: *w lewo, w pół, w prawo*;
 - e) *z*, np.: *z wewnątrz*;
 - f) *za*, np.: *za bardzo, za długo, za późno, za wcześniej, za dużo, za mało, za darmo*;
5. niektóre połączenia przyimków (czy też słów homonimicznych z przyimkami) z przymiotnikami (elementami *o postaci przymiotnikowej*), a mianowicie:
6. *za* z dowolną formą przypadkową, np.: *za duży, za mały, za młoda, za słabym* (takie słowoformy są opatrzone odpowiednimi symbolami gramatycznymi i zaliczone do haseł typu: **za duży** *Adj*, **za mały** *Adj*, **za młody** *Adj*, **za słaby** *Adj*, a więc haseł przymiotnikowych — o zróżnicowanych słowoformach);
 - b) *po* z archaiczną (nie występującą poza tym połączeniem) formą celownika na *-u*, np.: *po ludzku, po polsku, po prostu*;
 - c) *po* z biernikową formą tzw. liczebnika porządkowego (poza tymi połącze-niami liczebniki takie zaliczaliśmy do zbioru przymiotników — por. niżej punkt 4.2.3.), np.: *po pierwsze, po drugie, po trzecie*;
 - d) inne przyimki (zwłaszcza *z*) w połączeniach archaicznych, nie realizują-cych rządu przypadkowego, obowiązującego we współczesnym języku pol-skim, np.: *z bliska, z daleka, z wolna, z grubszą; do niedawna; od dawna; za młodu*;
7. inne wyrażenia, złożone z dwóch lub większej liczby słów, które są po-wszechnie uznawane za jednolite jednostki, zwłaszcza gdy przynajmniej je-den z członów utracił samodzielność, jest archaizmem leksykalnym i poza wyrażeniami złożonymi nie jest używany we współczesnym języku polskim, np.: *gdzie indziej, kiedy indziej; na poły, na przemian, na bakier; nie lada; od dawien dawna; po ciemku, po omacku; w zamian; z dala*.

Zasada wymieniona w punkcie 6. pozwoliła potraktować jako całości wyraże-nia mające wyraźnie jednolitą funkcję semantyczną i odbierane jako niepodziel-ne, ale pisane rozdzielnie, np. wyrażenie *nie ma* (w funkcji zaprzeczonego *jest*) albo tzw. złożone spójniki: *jak gdyby, jako że, mimo iż, mimo że, nie tyle, o ile, o tyle, podczas gdy, tyle że*. Trzeba dodać, że dopuszczano wyłącznie interpretację

⁷ Niektóre z przyjętych przez nas zasad budzą obecnie wątpliwości w samych autorach. Niestety rewizja podjętych przed laty decyzji byłaby na tyle trudna, żeśmy z niej zrezygnowali.

jako jednolitych słowoform ciągów słów bezpośrednio po sobie następujących, a nie dopuszczano sytuacji, aby części jednej słowoformy były przedzielone innymi wyrażeniami. Dlatego wyrażenie *o tyle ... o ile* interpretowano jako dwie słowoformy: *o tyle* i *o ile*.

Zasada ta prowadziła też do uznania za jednolite słowoformy pewnej liczby wyrażen, których wszystkie człony mogą występować w tekstach samodzielnie, a które traktowane są jako utarte połączenia. W niniejszym tomie jako słowoformy zostały zinterpretowane — oprócz wymienionych wyżej — następujące wyrażenia nieciągłe: *bądź co bądź, co prawda, coś niecoś, między innymi; na dobre, na stałe, na lepsze; na odwrót, na ogół, na raz, na razie, na tyle, na co dzień; nie sposób, o wiele, od razu, przede wszystkim, raz dwa, raz po raz, tak samo, w ogóle, w pobliżu, z powrotem, za wiele*.

W wypadkach wątpliwych dotyczących postaci hasłowej — podobnie jak w innych sytuacjach nasuwających wątpliwości — przyjmowano rozstrzygnięcia zawarte w *Słowniku języka polskiego* pod red. W. Doroszewskiego (tj. w rozważanym wypadku za całościowym traktowaniem wyrażenia złożonego przemawiało umieszczenie go jako hasła i objaśnienie jako jednolitej całości).

3.4. Dobór materiału do tomu zbiorczego

Przed dobraniem materiału do niniejszego tomu — zasadniczego słownika — opracowany już materiał podległ dalszej selekcji. Usunięto mianowicie zeń wszystkie nazwy własne, skróty oraz wtręty i cytaty obce (były one na ogół oznaczane specjalnymi symbolami odróżniającymi). Zgodnie z powszechną opinią autorzy uznali ten materiał za mało charakterystyczny dla języka traktowanego jako system (w materiale naszym hasła takie, jak np. *Gomulka* albo *DRW*, miały frekwencję na tyle dużą, że mogłyby się znaleźć w tomie zbiorczym).

3.4.1. Nazwy własne

Jako nazwy własne traktowano na ogół — zgodnie z powszechnym rozumieniem tego pojęcia, niezbędnym przy opanowaniu zasad pisowni — wyrazy pisane ze względów ortograficznych wielką literą. Gdy przygotowywano materiał, symbolem odróżniającym imienia własnego (znak *+*) oznaczano więc: nazwy indywidualne ludzi, zwierząt, przedmiotów (np. pojazdów), instytucji, wytworów kultury (czyli tytuły książek, czasopism, filmów itp.) oraz nazwy geograficzne i topograficzne indywidualnych obiektów. W szczególności za imiona własne uważano nazwy indywidualne homonimiczne z rzeczownikami pospolitymi, np.: (*jamnik*) *Puzon*, (*restauracja*) *Bazyliшек*. Zasadę tę rozszerzono także na konwencjonalne literowe symbole przedmiotów i pojęć, stosowane w odniesieniu do klas czy kategorii (np. *klasa wojewódzka A*) albo witamin.

W wypadku nazw (zwłaszcza tytułów) wielowyrzowych symbolem imienia własnego oznaczano wszystkie jego człony; decyzja ta powodowała, że w tomach *SWJP* figurują jako osobne hasła ze znakiem *+* i niebagatelną frekwencją takie jednostki, jak *w* (różne od regularnego przyimka *w*, bez znaku *+*) czy *i* (różne od regularnego spójnika *i*): jednostki te zostały opuszczone przy przygotowywaniu niniejszego tomu.

Wbrew zasadom ortograficznym, opracowując materiał, nie zaliczono do nazw własnych następujących dwóch grup jednostek:

1. pełnych nieumownych nazw instytucji, typu *spółdzielnia pracy szewców i cholewkarzy, centrala handlowa przemysłu mleczarskiego*;
2. nazw mieszkańców części świata, krajów, prowincji oraz nazw narodów, ras i szczepów, np. danego regionu typu *amerykanin, polak, poznaniak, mazur*.

Jak widać z przykładów, w konsekwencji tej decyzji w stosunku do szeregu wyrazów stosowano zapis niezgodny z regułami ortografii. Było to spowodowane uwarunkowaniami technicznymi. Alfabet dalekopisu, którym posługiwano się w pierwszej fazie pracy, nie rozróżniał wielkich i małych liter. Wprowadzonym przez nas symbolem imienia własnego (znak *+*) oznaczano jednostki, które planowaliśmy — ze względu na znaczenie — wyeliminować w dalszej fazie pracy

z tomu zbiorczego, nie zaś wyrazy pisane ze względów ortograficznych wielką literą. Informacja o pisowni wielką literą została więc w fazie wprowadzania materiału do maszyny zgubiona. Automatyczne przywrócenie jej w trakcie przeprowadzanych później prac komputerowych nie było możliwe, zrezygnowano zaś z wtórnego wprowadzenia wielkich liter przy korekcie, zwłaszcza że skutkiem tej decyzji zlikwidowane zostało rozróżnienie ortograficzne jednostek w kilku parach typu: *Francuz* i *francuz*, *Polka* i *polka*.

Specjalnie podkreślić wypada, że zgodnie z ortografią i praktyką leksyko-graficzną nie zaliczono do nazw własnych przymiotników utworzonych od imion własnych (np.: *bliskowschodni*, *amerykański*, *warszawski*, ani nazw firmowych wyrobów przemysłowych: samochodów, papierosów, słodyczy itp. (np. *wołga*, *warszawa*, *sporty*, *giewonty*, *katarzynki*).

3.4.2. Skróty

Z materiału usunięte zostały również skróty, w pamięci maszyny oznaczone specjalnym symbolem odróżniającym (znak *.*). W ogromnej większości wypadków skróty te miały charakter nazw własnych, np.: *PZPR*, *PRL*. Decyzja ta mogła się przyczynić do wyeliminowania pewnej liczby jednostek nie mających takiego charakteru, np. *PGR*.

3.4.3. Wtręty i cytaty obce

W odniesieniu do wtrętów i cytatów obcych nie były niestety stosowane przy opracowaniu materiału jednolite kryteria. Przy przygotowaniu materiału do tomu III *SWJP* (publicystyki), opracowanego w pierwszej kolejności, wtręty i cytaty skreślono z próbek przed wprowadzeniem danych do pamięci maszyny, gdyż stanowiły materiał nieprzydatny z punktu widzenia potrzeb sponsora (Stowarzyszenia Stenografów i Maszynistek). Podczas pracy nad pozostałymi tomami skreślano tylko obcojęzyczne wtręty zdaniowe, a cytaty leksykalne pozostawiano (np. *grand prix*, *sir*). Zostały one usunięte przy przygotowaniu ostatecznego materiału niniejszego tomu.

3.4.4. Jednostki o niskiej frekwencji

Do tomu niniejszego, czyli zasadniczego słownika, zostały włączone tylko te hasła, których frekwencja w całej próbie jest nie niższa niż 4. Hasła o frekwencji niższej zostały w słowniku pominięte.

Trzeba tu jednak podkreślić, że jednostki zaliczone do materiału, a nie uwzględnione w niniejszym tomie mają istotny wpływ na liczebność całości materiału. Wszystkie podawane w niniejszym tomie liczby wystąpień muszą być traktowane jako liczniki ułamków: o mianowniku 100 000 — dla danych z jednego stylu, o mianowniku 500 000 — dla danych z pełnego materiału. Podstawowe informacje o liczebności jednostek zaliczonych do materiału, a pominiętych w słowniku, podamy w punkcie 7 niniejszego wstępu.

3.4.5. Kwestie ortograficzne i poprawnościowe

Przyjęto zasadę, że formy zapisane w materiale zasadniczym podajemy w postaci oryginalnej. Jeżeli tekst zawierał formę niezgodną z normą współczesnego języka polskiego, została ona w tej postaci — po ewentualnym uzupełnieniu symbolami — wpisana do pamięci maszyny i zamieszczona w słowniku pod odpowiednim poprawnie sformułowanym hasłem (postać hasłowa była przecież dotwarzana), np. słowoforma *roków* w hasło **rok** *S*, słowoforma *tę* *Asg* w hasło **ten** *Adj*.

4. Gramatyczna charakterystyka materiału

4.1. Podstawy teoretyczne

Jak powiedzieliśmy już w poprzednim rozdziale, badaniom w naszym słowniku podlegają obiekty dwu typów — słowoformy (zinterpretowane słowa) i hasła (wyrazy w sensie słownikowym). Z takimi obiektami styka się zresztą każdy w trakcie nauki w szkole podstawowej. Dowiaduje się on na przykład, że *myślami* to narzędnik liczby mnogiej (jedna z form wyrazowych) rzeczownika (leksemu) **myśl**, który jako wyraz odmienny (leksem, czyli wyraz w sensie słownikowym) ma również inne formy. Jedną z tych form to mianownik liczby pojedynczej *myśl*, który jest również formą podstawową leksemu, a więc formą reprezentującą cały leksem: w słowniku, podręczniku gramatyki czy rozważaniach na temat języka. To samo słowo *myśl* jest również tekstowym wykładnikiem 2. osoby trybu rozkazującego czasownika (leksemu), którego formą hasłową jest **myśleć**. Tak więc słowo *myśl* powinno być w naszym opracowaniu — w zależności od funkcji, którą pełni w tekście — potraktowane jako forma dwóch różnych leksemów. W istocie słowo to interpretujemy jako trzy różne formy wyrazowe, jedna z nich należy do leksemu czasownikowego **myśleć**, a dwie — do leksemu rzeczownikowego **myśl**: oprócz mianownika jeszcze biernik liczby pojedynczej. O tym, że to dwie różne formy, łatwo się przekonać, podstawiając zamiast nich w tekście formy rzeczownika **idea** — zamiast pierwszej można podstawić *idea*, zamiast drugiej — *ideę*. O tym, że mianownik i biernik rzeczownika to dwie różne jego formy, także wie każdy. Trzeba tu jednak zwrócić uwagę, że wyodrębnienie form wyrazowych, podobnie jak leksemów, zawiera elementy konwencjonalności. Przyjeliśmy tu także pewne założenia, o których będzie mowa w dalszych rozdziałach wstępu.

Postawową jednostką opisu w naszym słowniku jest *hasło* — jednostka zbliżona do leksemu; jej też opisowi poświęcono artykuł hasłowy w głównej — alfabetycznej — części słownika (tylko do niej będziemy się odwoływać w tym punkcie wstępu). Każde hasło wprowadzone jest do słownika, podobnie jak w innych kompendiach leksykograficznych i encyklopedycznych, za pomocą *nazwy hasła*, tj. napisu, który przeciwstawia je innemu hasłom (nazwy hasła są porządkowane, np. według alfabetu). Nazwa hasła w naszym słowniku jest jednostką bilateralną: wchodzi do niej — oprócz *słowa hasłowego*⁸, będącego tradycyjnie formą podstawową danego leksemu — oznaczenie części mowy i, jeśli zachodzi taka potrzeba, dodatkowy numer podany w nawiasie kwadratowym. Pierwsza część nazwy hasła (słowo hasłowe) wybrana jest zgodnie z tradycją: bezokolicznik dla leksemów odmiennych przez przypadek (oraz, być może, liczbę i, ewentualnie, rodzaj). W nazwie hasła nieodmiennego występuje słowo, będące wykładnikiem jego jedynej słowoformy. Rzadkie wypadki, gdy słowo hasłowe wybrane jest inaczej, zostaną zasygnalizowane w niniejszym wstępie.

W skład hasła wchodzi *słowoformy*. Słowoformy wchodzące w skład jednego hasła różnią się swoimi wykładnikami tekstowymi albo wartościami uwzględnionych w opisie kategorii gramatycznych.

Zasadniczy system oznaczeń zastosowany w słowniku opiera się na określonym schemacie opisu gramatycznego współczesnego języka polskiego. Omówimy go teraz, ponieważ w sposób istotny będziemy się doń odwoływać w dalszych punktach niniejszego wstępu. Podstawowe znaczenie ma dla nas klasyfikacja hasła, nawiązująca do tradycyjnego rozróżnienia części mowy, ale wnosząca do niego istotne modyfikacje.

⁸ W tym punkcie określenie nazwy hasła zawiera uproszczenie. Trzeba by, ściśle rzecz ujmując, wprowadzić jeszcze jedno pojęcie, np. *napis hasłowy*, są bowiem w naszym słowniku hasła, w których ten element składa się z kilku słów.

4.2. Typy haseł (części mowy)

Typy haseł (części mowy) wyróżniono w zasadzie na podstawie kryteriów formalnych, w pierwszej kolejności — fleksyjnych, w drugiej — składniowych. Niżej wymienimy wszystkie typy haseł, podając w nawiasie oznaczenia stosowane w odniesieniu do nich w naszym słowniku. W treści poszczególnych punktów podamy oznaczenia, stosowane przy należących do nich słowoformach. Wszystkie oznaczenia objaśnione są ponadto w wykazie skrótów⁹.

4.2.1. Rzeczowniki (*S*)

Do rzeczowników zaliczono hasła odmieniające się przez przypadek, a nie odmieniające się przez rodzaj, jeśli nie zostały zaliczone do liczebników (klasa 4.2.3) ani zaimków (klasa 4.2.4). Do klasy rzeczowników nie włączono nazw własnych ani skrótów, które — jak już powiedzieliśmy wyżej (por. punkt 3.4) — przy opracowywaniu materiału opatrywano specjalnymi symbolami (obie te klasy jednostek nie zostały włączone do niniejszego tomu). Jako osobne hasła opracowano serie form liczby mnogiej supletywne w stosunku do pojedynczej: **ludzie i lata**.

W opisie gramatycznym słowoform rzeczownikowych określano wartości przypadków i liczb (słowoformy rzeczowników *plurale tantum* oznaczono symbolem liczby mnogiej).

Zastosowano następujące oznaczenia wartości przypadków:

- N* — mianownik (*nominativus*)
- G* — dopełniacz (*genetivus*)
- D* — celownik (*dativus*)
- A* — biernik (*accusativus*)
- I* — narzędnik (*instrumentalis*)
- L* — miejscownik (*locativus*)
- V* — wołacz (*vocativus*)

Wartości kategorii liczby są następujące:

- sg* — liczba pojedyncza (*singularis*)
- pl* — liczba mnoga (*pluralis*)

Zrezygnowano z oznaczania rodzaju rzeczowników.

4.2.2. Przymiotniki (*Adj*)

Do przymiotników zaliczono hasła odmieniające się przez przypadek, rodzaj, liczbę i — niekiedy — stopień; oprócz hasła **on** (zaliczonego do zaimków — por. 4.2.3).

Zgodnie z tą regułą za przymiotniki uznano wszystkie hasła odmieniające się według przypadkowo-liczbowo-rodzajowego schematu deklinacji przymiotnikowej, także te, które tradycyjnie nazywane są liczebnikami (porządkowymi i in., np.: **pierwszy**, a także hasło **jeden**), zaimkami przymiotnymi (np. **ten**) oraz imiesłowami przymiotnikowymi (np. **walczący, wykonany**). Imiesłowy przymiotnikowe są traktowane jako samodzielne jednostki leksykalne.

W opisie gramatycznym słowoform przymiotnikowych określono wartości przypadków i liczb: ich wartości są takie same, jak dla słowoform rzeczownikowych, ale wśród przypadków nie uwzględnia się wołacza. Nie oznaczono wartości rodzajowych, głównie z powodu bardzo skomplikowanej homonimii między różnymi rodzajami przy ustalonych wartościach przypadkowo-liczbowych. Nie oznaczono też wartości stopnia. Jako osobne hasła opracowano serie form stopnia wyższego supletywne w stosunku do równego: **większy, mniejszy, lepszy** i **gorszy** (do haseł tych włączono także formy stopnia najwyższego).

Indywidualne rozstrzygnięcie zastosowano w hasle **jeden**. Podobnie jak w hasłach liczebnikowych (por. punkt 4.2.3.), obok słowoform *jeden Nsg* i *jeden Asg*, oznaczonych wartościami przypadków, znajduje się słowoforma *jeden*, nie

⁹ Omówionemu tu systemowi oznaczeń skrótowych odpowiadał w tomach *SWJP* system oznaczeń cyfrowych. Nie powtarzamy go tu w sposób systematyczny, ponieważ jest należycie omówiony we wstępie do każdego z tomów *SWJP*.

mająca żadnego symbolu gramatycznego. W ten sposób zaznaczano wypadki, gdy słowo to występowało w nazwach liczb wielocyfrowych w postaci nie dostosowanej do wymagań przypadkowych narzucanych przez kontekst (np. *w dwudziestu **jeden** krajach, delegaci trzydziestu **jeden** państw*), albo miało funkcję nieodmiennej przydawki ilościowej (np. *dlugość **jeden** metr trzydzieści centymetrów, w czasie dwie godziny dwadzieścia **jeden** minut, w latach sześćdziesiąt **jeden** — sześćdziesiąt dwa*).

4.2.3. Liczebniki (*Num*)

Do liczebników zaliczono hasła odmieniające się przez przypadek, a nie odmieniające się przez liczbę, i oznaczające stosunki liczbowe (czyli tradycyjnie zaliczane do klasy liczebników). W niniejszym tomie znalazły się następujące hasła liczebnikowe (dla większej informatywności przytaczamy nazwy hasel wraz z ogólną frekwencją hasła):

dwa (936), **trzy** (568), **cztery** (373), **pięć** (431), **sześć** (240), **siedem** (164), **osiem** (221), **dziewięć** (135), **dziesięć** (202), **jedenaste** (42), **dwanaście** (87), **trzynaście** (40), **czternaście** (46), **piętnaście** (109), **szesnaście** (39), **siedemnaście** (42), **osiemnaście** (38), **dziewiętnaście** (16), **dwadzieścia** (400), **trzydzieści** (266), **czterdzieści** (185), **pięćdziesiąt** (262), **sześćdziesiąt** (229), **siedemdziesiąt** (123), **osiemdziesiąt** (124), **dziewięćdziesiąt** (79), **sto** (330), **dwieście** (161), **trzysta** (101), **czterysta** (65), **pięćset** (102), **sześćset** (67), **siedemset** (51), **osiemset** (85), **dziewięćset** (572), **oba** (170), **obydwa** (23), **dwoje** (19), **oboje** (15), **troje** (8), **czworo** (6), **pół** (84), **półtora** (22), **kilka** (400), **kilkanaście** (39), **kilkadziesiąt** (37), **kilkaset** (22), **parę** (96), **ile** (51), **ileż** (4), **tyle** (83), **wiele** (424), **niewiele** (14).

Takie określenie klasy spowodowało, że zaliczono do niej hasła odpowiadające tylko tym leksemom, które zwykle nazywa się liczebnikami głównymi i zbiorowymi (leksemy odmienne przez liczbę i rodzaj, a zwane tradycyjnie liczebnikami, np. **drugi**, **jeden**, zaliczono do przymiotników; a leksemy nieodmienne, zwane tradycyjnie liczebnikami, np. **dwukrotnie** — do przysłówków), oraz leksemom nietypowym (por. ostatnia grupa, często nazywanym zaimkami liczebnymi). Należy specjalnie podkreślić, że jako liczebniki nie zostały zakwalifikowane hasła **tysiąc** i **milion** (zaliczyliśmy je do rzeczowników).

W opisie gramatycznym słowoforn liczebnikowych określono wartości przypadków: ich wartości są takie same, jak dla słowoforn przymiotnikowych. Zgodnie z przyjętymi założeniami nie oznaczono wartości rodzajowych.

W wielu hasłach liczebnikowych obok słowoforn oznaczonych wartościami przypadków znajduje się słowoforna nie mająca żadnego symbolu gramatycznego. W ten sposób zaznaczano takie słowo albo występujące w nazwach liczb wielocyfrowych w postaci nie dostosowanej do wymagań przypadkowych narzucanych przez kontekst (np. *w **sto** dwudziestym pułku, od **tysiąc dziewięćset piątego** roku*), albo mające funkcję nieodmiennej przydawki ilościowej (np. *zwyciężył Kowalski w czasie **pięć** godzin **dziewięć** minut **piętnaście** sekund, ze stopą wzrostu **osiem** procent, w latach **tysiąc dziewięćset sześćdziesiąt** — **sześćdziesiąt trzy**, zwyciężyli Belgowie w stosunku **pięć-cztery***).

4.2.4. Zaimki (*Prn*)

Do zaimków zaliczono następujące hasła, które na podstawie ogólnych zasad znalazły się w tym tomie (przytaczamy nazwy hasel wraz z ogólną frekwencją hasła):

co (2224), **cokolwiek** (9), **coś** (487), **cóż** (90), **ja** (3028), **jedno** (44), **kto** (388), **ktokolwiek** (11), **któs** (226), **któż** (13), **nic** (597), **nikt** (247), **my** (875), **on**

(6650), **samo** (70), **siebie** (1139), **się** (9302), **tamto** (6), **to** (5113), **ty** (1292), **wszystko** (525), **wy** (279).

Do grupy tej zostały zaliczone hasła odmieniające się przez przypadek i nie odmieniające się — oprócz hasła **on** — przez rodzaj. Arbitralnie wybrano słowo hasłowe dla hasel **siebie** i **się**, nie mających mianownika. Słowo *się* było interpretowane jako słowoforma należąca do któregoś z tych leksemów: jeżeli było powiązane ze słowoformą czasownikową — do leksemu **się**, jeżeli — ze słowoformą przyimkową — do leksemu **siebie** (por. wyżej punkt 3.3.3.). Formy supletywne zaimków nie są wyłączone w osobne hasło: np. słowoformy *jego*, *ich* i *mu* są zaliczone do hasła **on**.

W opisie gramatycznym słowoform należących do hasła **on** określono — podobnie jak dla słowoform przymiotnikowych — wartości przypadków i liczb, a nie oznaczono wartości rodzajowych. W opisie pozostałych słowoform zaimkowych określono tylko wartości przypadków.

Zaimki w naszym opisie obejmują więc, jak widać, leksemy nazywane tradycyjnie zaimkami rzeczownikowymi. Inne grupy tradycyjnie rozumianych zaimków zaliczono do innych typów hasel: przymiotne — do przymiotników, liczebne — do liczebników, przysłowne — do przysłówków.

Dla słowoform zaimkowych charakterystyczne jest bardzo silne uwikłanie w homonimie — zostanie ono omówione w punkcie 4.3.

4.2.5. Czasowniki (V)

Do czasowników zaliczono hasła odpowiadające leksemom, których podstawowe słowoformy pełnią w zdaniu funkcję orzekającą. Na tej zasadzie za czasowniki uznano leksemy mające kategorie fleksyjne trybu i czasu, a być może również — osoby, liczby i rodzaju. Do kategorii fleksyjnych nie zaliczono aspektu: odrębnymi hasłami są więc **krzyczeć** i **krzyknąć** itp. Podobnie nie uznano za fleksyjną kategorii strony; wiąże się to z inną decyzją: wszystkie imiesłowy przymiotnikowe stanowią w naszym słowniku samodzielne hasła przymiotnikowe (por. punkt 4.2.2.).

W skład typowego hasła czasownikowego wchodzi słowoformy odpowiadające tzw. formom osobowym, a ponadto bezokolicznik, imiesłów przysłówkowy i bezosobnik (słowoforma zakończona na *-no* lub *-to*). Jeśli dany leksem czasownikowy zawiera formy supletywne, nie są one wyłączone w osobne hasło: np. słowoformy *szedł* i *szła* zaliczone są do hasła **iść**; słowoforma *znajdzie* — do hasła **znaleźć**.

Indywidualną postać mają czasownikowe hasła **powinien** i **winien**, tworzące formy proste w sposób wysoce nietypowy: są one zróżnicowane na podstawie kategorii liczby, osoby i rodzaju; odmianę przez tryby i czasy uzyskuje się poprzez tworzenie form analitycznych. Za ich słowo hasłowe, zgodnie z praktyką leksykograficzną, wybrano 3. osobę liczby pojedynczej rodzaju męskiego. Specjalnym hasłem czasownikowym jest **nie ma**, zawierające tylko jedną słowoformę.

Do klasy czasowników zaliczono również predykatywy, tj. hasła reprezentowane w tekście przez słowa: *brak*, *dość*, *dosyć*, *można*, *nie sposób*, *potrzeba*, *słychać*, *szkoda*, *trzeba*, *warto*, *wiadomo*, *widać*, *wolno*, *wstyd*, *żal*. Odpowiednie leksemy tworzą analitycznie formy trybów (innych niż oznajmujący) i czasów (innych niż teraźniejszy). W naszym słowniku w skład odpowiednich hasel (np. **trzeba**), oprócz słowoformy nie opatrzonej symbolem gramatycznym (np. *trzeba*), wchodzi słowoforma trybu warunkowego; nie zaznaczano natomiast zróżnicowania według kategorii czasu. Dla słowoform należących do omawianych leksemów charakterystyczne jest bardzo silne uwikłanie w homonimie — zostanie ono omówione w punkcie 4.3.

W opisie gramatycznym słowoform czasownikowych nie określano prawie wcale wartości kategorii fleksyjnych. Przyjęto bowiem założenie, że poszczególne formy czasownikowe mają z reguły jednofunkcyjne wykładniki (np. bezokolicznik — *-ć* lub *c*, 1. os. cz. ter. — *-ę* lub *-m* itd.), więc zebrane w hasło

czasownikowym słowa zinterpretuje bez trudu człowiek, a także komputer na podstawie prostego programu.

Oznaczono natomiast specjalne funkcje poszczególnych słów, które dają podstawy do interpretowania ich jako odrębnych słowoform. W hasłach czasownikowych wyróżniono osobnymi symbolami następujące słowoformy:

- *fut* — bezokolicznik albo słowoforma zakończona na *-ł, -ła, -ło, -li* lub *-ły*, użyta jako część składowa czasu przyszłego złożonego (np. *będzie pisać, będzie pisał*);
- *mob* — słowoforma zakończona na *-ł, -ła, -ło, -li* lub *-ły*, użyta jako część składowa czasu przeszłego z cząstką *+m, +ś, +śmy, +ście* (*ruchomą końcówką*) (np. *kiedyście przyjechali*);
- *cnd* — słowoforma zakończona na *-ł, -ła, -ło, -li* lub *-ły* albo słowoforma predykatywna (np. *można, trzeba*), użyta jako część składowa trybu warunkowego z cząstką *+by, +bym, +byś, +byśmy, +byście* (*ruchomą końcówką*), pisana osobno albo łącznie z poprzednim słowem;
- *imp* — słowoforma homonimiczna z formą czasu teraźniejszego, użyta jako część składowa trybu rozkazującego 3. osoby ze słowem *niech* (np. *niech pisze*);
- *pass* — słowoforma wchodząca do leksemu **być, bywać, zostać** lub **zostawać**, użyta jako część składowa strony biernej (np. *był napisany, został napisany* itp.);
- *aux* — słowoforma wchodząca do leksemu **być**, użyta jako część składowa czasów złożonych: przyszłego lub zaprzeczonego (np. *będzie pisał, napisał był*);
- [się] — słowoforma czasownikowa użyta w połączeniu z *się* nie mającym funkcji bezosobowej (np. w zdaniu *Muję się* albo *Deski łatwo się piłują*, a nie w zdaniu *Deski łatwo się piłuje*).

Nie zaznaczano wypadków powiązania z *by* bezosobnika (np. *wykonano by*). Nie odnotowano połączeń predykatywów z *było* i *będzie*, mających funkcję czasu przeszłego i przyszłego (np. *można było, można będzie*).

4.2.6. Przyimki (*Prp*)

Za przyimki uznano hasła, których słowoformy — nie wchodząc do fleksyjnych opozycji przypadkowych ani trybowo-czasowych — mają wymaganie składowe: łączą się z członem o ściśle określonej wartości przypadkowej. W opisie gramatycznym słowoform przyimkowych określono wartości wymaganych przypadków (są one takie same, jak wartości przypadkowe rzeczowników). W niniejszym tomie znalazły się następujące hasła przyimkowe (w nawiasie — ogólna frekwencja hasła):

bez (459), **blisko** (18), **celem** (17), **co** (77), **dla** (1361), **do** (5854), **dokoła** (4), **dookoła** (6), **drogą** (11), **dzięki** (143), **jako** (447), **koło** (59), **ku** (98), **między** (350), **mimo** (126), **na** (8600), **nad** (498), **naprzeciw** (18), **naprzeciwko** (4), **niby** (47), **niczym** (5), **niedaleko** (9), **o** (3072), **obok** (104), **od** (1780), **odnośnie** (7), **około** (103), **oprócz** (46), **po** (1728), **pod** (668), **podczas** (117), **pośród** (63), **pomimo** (12), **ponad** (40), **poniżej** (11), **poprzez** (53), **pośrodku** (5), **pośród** (10), **powyżej** (10), **poza** (144), **prócz** (10), **przeciw** (68), **przeciwko** (117), **przed** (682), **przez** (1464), **przy** (995), **spod** (45), **spośród** (55), **spoza** (5), **sprzed** (20), **temu** (44), **u** (399), **w** (16316), **wbrew** (21), **wedle** (11), **według** (144), **wewnątrz** (20), **wobec** (172), **wokół** (46), **wskutek** (49), **wśród** (185), **wzdłuż** (36), **względem** (17), **z** (8310), **za** (1336), **zamiast** (37), **znad** (8), **zza** (22).

Słowoformy mające taki sam wykładnik tekstowy i różne symbole gramatyczne interpretowano jako należące do tego samego hasła. Czasem do hasła zaliczano też nietypowe słowoformy, występujące w tekstach przy specjalnych uwarunkowaniach (np. *we, przeze*) albo stanowiące tzw. kontrakcje (np. *zań*). Zdarzają się hasła przyimkowe zawierające kilka słowoform, np. **o** *Prp* (*o A* i *o L*), **w** *Prp* (*w A, w L, we A, we L, weń*).

4.2.7. Wykrzykniki (*Int*)

Za wykrzykniki uznano słowa, których słowoformy nie są przeciwstawione wzajemnie na podstawie żadnej kategorii fleksyjnej i nie wchodzą do struktury składniowej zdania. W niniejszym tomie znalazły się następujące hasła wykrzyknikowe (w nawiasie podano ogólną frekwencję):

a (210), **ach** (540), **aha** (210), **ano** (140), **ba** (100), **beu** (6), **brawo** (10), **cha** (18), **dobranoc** (11), **e** (46), **ech** (10), **ej** (4), **ha** (34), **halo** (19), **he** (15), **hej** (8), **hm** (24), **ho** (15), **hop** (6), **mmm** (5), **o** (163), **och** (42), **oho** (4), **oj** (25), **ot** (26), **psiakrew** (7).

Każde hasło wykrzyknikowe zawiera tylko jedną słowoformę.

4.2.8. Spójniki (*Cnj*)

Za spójniki uznano hasła, których słowoformy — nie wchodząc do fleksyjnych opozycji przypadkowych ani trybowo-czasowych — pełnią funkcję łączącą i nie mają wymagania wiązania się z członem o ściśle określonej wartości przypadkowej. W niniejszym tomie znalazły się następujące hasła spójnikowe (w nawiasie ogólna frekwencja):

a (3226), **aby** (430), **aczkolwiek** (7), **albo** (263), **albowiem** (13), **ale** (1963), **ależ** (35), **ani** (333), **aniżeli** (21), **aż** (213), **ażeby** (20), **байд** (55), **bo** (736), **bowiem** (220), **by** (148), **byle** (25), **choć** (113), **choćby** (17), **choć** (123), **choćby** (66), **co** (134), **dopóki** (28), **gdy** (389), **gdyby** (223), **gdyż** (129), **i** (12385), **ile** (8), **im** (35), **iż** (197), **jak** (2251), **jakkolwiek** (23), **jako** (5), **jako że** (10), **jednakże** (52), **jeśli** (325), **jeśliby** (4), **jeżeli** (242), **lecz** (242), **lub** (433), **mimo iż** (11), **mimo że** (31), **natomiast** (191), **ni** (17), **nie tyle** (14), **nim** (9), **niż** (361), **o ile** (38), **o tyle** (16), **oraz** (757), **podczas gdy** (29), **ponieważ** (135), **póki** (19), **przeto** (8), **skoro** (58), **tedy** (5), **to** (784), **toteż** (37), **tyle że** (8), **tym** (103), **więc** (692), **zamiast** (39), **zanim** (64), **zarówno** (138), **zaś** (178), **zatem** (66), **że** (4314), **żeby** (536).

W opisie gramatycznym słowoform spójnikowych nie są używane żadne symbole gramatyczne. Niektóre spójniki mają jednak po kilka słowoform, np. **jak** *Cnj* (*jak*, *jakem*, *jakeś* i *jakeśmy* por. punkt 3.3.2.).

4.2.9. Partykuły (*Prt*)

W tomach *SWJP* nie rozróżniano partykuł i przysłówków, przyjmując jako jeden typ haseł partykuło-przysłówki, oznaczane tym samym symbolem gramatycznym. Była to klasa resztkowa, do której zaliczono jednostki nie wchodzące do żadnej z klas wyróżnionych na podstawie kryteriów pozytywnych. Przed opracowaniem niniejszego tomu symbol partykuło-przysłówka zastąpiono na liście dwoma symbolami: partykuły i przysłówka. Nie posługiwano się przy tym ścisłymi kryteriami, lecz tradycyjnym rozumieniem obu pojęć. Ponieważ partykuł jest dużo mniej, definiujemy je przez listę (tak jak i w poprzednich wypadkach ograniczamy się do jednostek, które wystąpiły w niniejszym tomie):

aby (11), **akurat** (37), **ale** (11), **ano** (5), **байд** (6), **байд co байд** (8), **bodaj** (10), **by** (349), **bynajmniej** (37), **chyba** (293), **co** (51), **co najmniej** (43), **co najwyżej** (13), **co prawda** (30), **coraz** (193), **coś** (18), **cóż** (5), **czego** (20), **czemu** (29), **czy** (1119), **czyli** (120), **czyż** (15), **czyżby** (18), **dłaczego** (265), **dlatego** (278), **dopiero** (229), **doprawdy** (17), **ino** (4), **jak** (225), **jak gdyby** (39), **jakby** (258), **jakoby** (10), **jakoś** (88), **jakże** (54), **jednak** (668), **jedynie** (139), **jeno** (6), **jeszcze** (1014), **lada** (10), **ledwie** (14), **ledwo** (16), **mianowicie** (68), **między innymi** (191), **mniejsza** (5), **może** (469), **możliwie** (27), **nader** (11), **nadto** (7), **naprawdę** (124), **nareszcie** (30), **następnie** (95), **naturalnie** (34), **nawet** (637), **nie** (8341), **niech** (300), **niechaj** (5), **niechże** (12), **niejako** (18), **niemal** (78), **niemniej** (34), **nieomal** (8), **niestety** (81), **niewątpliwie** (43), **no** (600), **oczywiście** (249), **omal** (6), **oto**

(78), otóż (48), owszem (50), pewnie (74), pewno (20), podobno (40), ponadto (65), poniekąd (7), prawdopodobnie (50), przecie (23), przecież (471), przeciwnie (29), przede wszystkim (192), przeważnie (41), przynajmniej (76), przypuszczalnie (8), raczej (89), również (572), rzeczywiście (92), rzekomo (4), słusznie (28), specjalnie (35), szczególnie (108), tak (1903), także (433), tam (105), też (613), to (367), toć (5), tylko (1340), tymczasem (57), widać (17), widocznie (31), właściwie (119), właśnie (384), wprawdzie (58), wreszcie (145), wszak (6), wszakże (6), wyłącznie (54), względnie (26), z (34), za (68), zaledwie (67), zapewne (36), zarazem (35), zasadniczo (10), zazwyczaj (24), zgoła (10), zresztą (170), zwłaszcza (144), że (7).

W nawiasie podano ogólną frekwencję.

W opisie gramatycznym słowoform partykułowych nie są używane żadne symbole gramatyczne. Niektóre partykuły mają jednak po kilka słowoform, np. **jakby** *Prt* (*jakby, jakbym, jakbyś, jakbyście i jakbyśmy* por. punkt 3.3.2.).

4.2.10. Przysłówki (*Adv*)

Wobec decyzji podjętej w punkcie poprzednim, za przysłówki uznajemy hasła, które nie zostały zaliczone do żadnej z klas wymienionych wyżej.

W opisie gramatycznym słowoform przysłówkowych nie wprowadzano żadnych symboli gramatycznych. Niemniej zdarzają się przysłówki mające kilka słowoform, czasem przeciwstawionych sobie według kategorii stopnia. Jako osobne hasła opracowano postacie supletywne stopnia wyższego (włączono do nich także formy stopnia najwyższego): **więcej, mniej, lepiej i gorzej**.

4.3. Homonimia słowoform

Różne słowoformy, których wykładnikiem tekstowym jest to samo słowo, nazywamy, jak pamiętamy, *słowoformami homonimicznymi* lub, krótko, *homonimami*. Słowoformy homonimiczne mogą należeć do jednego hasła lub kilku haseł. Słowo, które jest wykładnikiem więcej niż jednej słowoformy, nazywamy *słowem homonimicznym*. Wypada podkreślić, że nasze użycie wymienionych terminów ma niewiele wspólnego z tradycyjnym rozumieniem tej sfery pojęć, za homonimy bowiem uważa się na ogół jednostki językowe, które — mając te same wykładniki tekstowe — różnią się znaczeniowo lub genetycznie. W naszym wypadku najważniejsze przy rozróżnianiu homonimów były ich odmienne funkcje gramatyczne.

Rozróżnienie słowoform homonimicznych (czyli rozbiecie słów homonimicznych na homonimy) to podstawowy problem techniczny niniejszego opracowania słownikowego. Musiało ono w swym głównym zřębie nastąpić już w trakcie wprowadzania materiału do komputera przed wykonaniem zasadniczych obliczeń. Rozróżnienie to przeprowadzono, dopisując do każdego słowa, co do którego zachodziły podejrzenia, że jest homonimiczne, *symbol gramatyczny* (zwany niekiedy w poprzednim rozdziale symbolem odróżniającym).

Zdarza się niekiedy, że znajdującemu się w korpusie słowu odpowiadała jednoznacznie — z punktu widzenia przyjętych założeń — jedna słowoforma. Przyjęliśmy założenie, że w wypadkach takich symbole cyfrowe dla jej oznaczenia w materiale nie muszą być wprowadzane. Gdy wynikało to z potrzeb dalszego opracowania, oznaczenie mogło zostać dopisane w dalszej fazie pracy. Nie było to jednak regułą i wiele słowoform zanotowanych w niniejszym tomie jest pozbawionych oznaczenia gramatycznego¹⁰. Wypadki takie należy interpretować w ten sposób, że pozycją w artykule hasłowym jest słowoforma

¹⁰ Zdarza się nawet, że zapisy różnych słowoform niczym się nie różnią. Na przykład dwie identycznie zapisane słowoformy *może* [bez dodanych symboli gramatycznych] są zanotowane w dwóch artykułach hasłowych: **może** *Prt* i **móc** *V*, reprezentacja obu tych jednostek w materiale wprowadzonym do komputera była oczywiście różna, odpowiednio: *może 8* (partykuła lub przysłówek) i *może 5* (czasownik).

o domyślnych wartościach gramatycznych. Zachodzi to szczególnie często w wypadku słowoform czasownikowych, przy których oznaczenia gramatyczne były z reguły pomijane.

Oznaczenia gramatyczne przy słowoformach czasownikowych wprowadzano w trakcie wprowadzania materiału do komputera tylko wtedy, gdy istniały podstawy do podejrzewania trudności przy gramatycznej kwalifikacji materiału. Oznaczenia te były uproszczone, skrótowe i miały na celu wyłącznie ułatwienie hasłowania. Potem w niniejszym tomie zostały pominięte, jeśli wynikało to z ustaleń ogólnych.

Na przykład przy wprowadzaniu materiału do komputera została wpisana słowoforma *myśl* 5, która według konwencji przyjętych w niniejszym tomie powinna przybrać postać *myśl* V; oznaczenie gramatyczne służyło temu, by została ona zaliczona do hasła **myśleć** V (w odróżnieniu od słowoform *myśl* Nsg i *myśl* Asg, należących do hasła **myśl** S). Jednak autorzy przyjęli zasadę pomijania oznaczeń gramatycznych w wypadku słowoform, dla których w tym samym hasle systemowo nie istnieje inna słowoforma homonimiczna; jednoznaczna charakterystyka gramatyczna omawianej słowoformy leksemu **myśleć** V (2. osoba trybu rozkazującego) nie została wprowadzona, a skrótowe oznaczenie części mowy jest pominięte, więc w naszym tomie figuruje ona w postaci *myśl*, co oczywiście powinno być traktowane jako zapis uproszczony.

Opatrzenie charakterystyką gramatyczną słowa występującego w tekście sprawiło niekiedy niemalże trudności. Sprowadzały się one przede wszystkim do określenia typów haseł. Granice pomiędzy poszczególnymi częściami mowy nie są bowiem zarysowane ostro. Przy określaniu przynależności danego słowa do klasy gramatycznej niejednokrotnie występują wątpliwości. Dotyczy to zarówno słów, które mogą pełnić funkcję form różnych części mowy w wyniku określonego rozwoju historycznego, gdy dana jednostka językowa nabierała nowych cech syntaktycznych i stawała się podstawą nowej, mającej inną charakterystykę gramatyczną, jak i słów, które skutkiem swych indywidualnych własności trudno poddają się klasyfikacji, stając jakby na pograniczu między kilkoma częściami mowy. W obu tych grupach wypadków dane słowo traktowano jako homonimiczne — stanowiące wykładnik tekstowy dwóch słowoform, należących do haseł różnych typów.

W celu rozróżniania takich trudnych do wykrycia homonimów autorzy wypracowali system kryteriów o charakterze głównie syntaktycznym. Omówimy je niżej. Kryteria te odnoszą się tylko do tych wypadków, które faktycznie były dyskutowane w trakcie opracowywania materiału. Na przykład w punkcie 3 nie rozpatrujemy homonimii słowa *można*, które może wystąpić w zdaniu albo jako orzeczenie, albo też jako mianownik rodzaju żeńskiego przymiotnika **możny**. Tego wypadku homonimii nie braliśmy w ogóle pod uwagę; opracowując materiał, mogliśmy — na podstawie kompetencji językowej — przypuszczać, że w kanonie podstawowym nie wystąpi słowoforma przymiotnikowa *można* — i tak też było w istocie. Tym bardziej w kryteriach nie są uwzględnione wypadki zupełnie przypadkowej i wykrywalnej bez trudu homonimii między poszczególnymi słowoformami odmiennych haseł różnych typów, np. *droga* (słowoforma rzeczownikowa lub przymiotnikowa) albo *pierze* (słowoforma rzeczownikowa lub czasownikowa). Nie rozważamy w nich również sytuacji, gdy słowo jest częścią wykładnika słowoformy, której wykładnik składa się z kilku słów (np. rozpatrujemy tylko interpretację słowa *tyle* jako wykładnika słowoformy liczebnikowej i słowoformy przysłówkowej, nie zaś jako części wyrażenia *nie tyle*, *o tyle*, *tyle że*, interpretowanych jako jednolite słowoformy) — sytuacje takie są szczegółowo omówione wyżej w punkcie 3.3.4.

Omawiając nasze kryteria, poszczególne typy wątpliwości opatrzymy skrótowymi tytułami, wskazującymi istotę problemu.

4.3.1. Rzeczownik czy przymiotnik

Bardzo typowe wątpliwości dotyczyły w tym wypadku stopnia substancywizacji pierwotnych przymiotników. Za kryterium substancywizacji uznano brak

pełnego paradygmatu rodzajowego. Za słowoformy rzeczownikowe uznawano te jednostki, które można uważać nie za wchodzące do opozycji rodzajowej obejmującej w mianowniku liczby pojedynczej rodzaj nijaki obok męskiego i żeńskiego — przy zachowaniu tego samego znaczenia. Stosując to kryterium, uznano nie tylko istnienie hasel rzeczownikowych takich, jak np.: **leśniczy** (rzeczownik rodzaju męskiego), **wytyczna** (rzeczownik rodzaju żeńskiego), lecz również — ze względu na niemożność uzyskania pełnej opozycji rodzajowej przy odpowiednich znaczeniach: **pośpieszny** (rzeczownik rodzaju męskiego), **przewodniczący** (rzeczownik rodzaju męskiego), **przewodnicząca** (rzeczownik rodzaju żeńskiego), **dane** (rzeczownik *plurale tantum* rodzaju niemęskoosobowego). Uznawano więc istnienie leksemów rzeczownikowych o ustalonym rodzaju obok leksemów przymiotnikowych odmiennych przez rodzaj, np. leksemu rzeczownikowego **średnia** rodzaju żeńskiego obok leksemu przymiotnikowego **średni**, odmiennego przez rodzaj; leksemów rzeczownikowych **chory** i **chora** — odpowiednio rodzajów męskiego i żeńskiego obok leksemu przymiotnikowego **chory**, odmiennego przez rodzaj. W opracowywanym do słownika materiale znalazły się więc np. hasła: **średni Adj** i **średnia S**, a także **chora S** (ze względu na niską frekwencję nie wszedł on do niniejszego tomu), **chory S** oraz **chory Adj**. O tym, z którym z nich wiązać wystąpienie odpowiedniego słowa w tekście — jakie przypisywać mu oznaczenie gramatyczne, decydował kontekst: typowy dla formy przymiotnika lub typowy dla formy rzeczownika, np.:

Średnia głębokość Bałtyku nie jest duża. → **średni Adj**
Średnia krajowa w zakresie spożycia alkoholu wzrosła ostatnio wyraźnie.
 → **średnia S**

Uznano też istnienie leksemów rzeczownikowych: **dziesiąta**, **setna** itp., oznaczających wskazane części całości — obok odpowiednich leksemów przymiotnikowych.

4.3.2. Rzeczownik czy czasownik

Trudności nastęczała kwalifikacja słów: *potrzeba*, *brak*, *szkoda*, *żał*, *wstyd*, *strach*. Uważano je za słowoformy czasownikowe, jeżeli w danym zdaniu występowało lub dawało się wstawić wyrażenie (*było*) *nam*, a w przeciwnym wypadku — za słowoformy rzeczownikowe, np.:

Brak ten odczuwano bardzo boleśnie. → **brak S**
W dyskusji brak [nam było] ładu. → **brak V**

4.3.3. Rzeczownik czy przyimek

Problem odnosi się do słów *celem*, *drogą*, *koło*, *skutkiem*, *względem*. Te pierwotne formy przypadkowe rzeczowników we współczesnym języku polskim pełniły wyraźnie i drugą funkcję — zleksykalizowały się jako przyimki. Każde z nich nasuwa wątpliwości interpretacyjne tylko wtedy, gdy występuje w ściśle określonym kontekście, a mianowicie wtedy, gdy jest pierwszym członem wyrażenia, zawierającego jako drugi człon frazę nominalną w dopełniaczu (takie wyrażenie będziemy roboczo nazywali wyrażeniem testowanym), i nie ma własnego określenia przymiotnikowego (to decyduje o jego charakterze rzeczownikowym). Aby w naszym słowniku rozróżnić obie funkcje wymienionych słów, wypracowano system roboczych kryteriów, opartych na możliwości substytucji innymi wyrażeniami lub zastosowania określonych przekształceń (transformacji). Kryteria te są następujące:

a) Słowo *celem* kwalifikowano jako słowoformę przyimkową, jeżeli całe wyrażenie testowane można było w tekście — bez zmiany sensu zdania — zastąpić zdaniem podrzędnym okolicznikowym celu, rozpoczynającym się od słowa *aby*; jeżeli warunek ten nie był spełniony, słowo *celem* było kwalifikowane jako słowoforma rzeczownikowa; np.:

*Rada Bezpieczeństwa zebrała się **celem** przedyskutowania sytuacji politycznej.* (≈ ... *aby przedyskutować sytuację polityczną*) → **celem Prp**

Celem zebrania Rady Bezpieczeństwa było przedyskutowanie sytuacji politycznej. → **cel** *S*

b) Słowo *drogą* kwalifikowano jako słowoformę przyimkową, jeżeli całe wyrażenie testowane można było w tekście — bez zmiany sensu zdania — zastąpić wyrażeniem rozpoczynającym się od słowa *przez*; w przeciwnym wypadku kwalifikowano je jako słowoformę rzeczownikową; np.:

Ustępstwa osiągnięto drogą pertraktacji. (\simeq ... przez pertraktacje.)
→ **drogą** *Prp*
Szli wyboistą drogą zwątpień. → **droga** *S*
Szli drogą zwątpień. → **droga** *S*

c) Słowo *skutkiem* kwalifikowano jako słowoformę przyimkową, jeżeli całe wyrażenie testowane można było w tekście — bez zmiany sensu zdania — zastąpić wyrażeniem rozpoczynającym się od słowa *wskutek*; w przeciwnym wypadku kwalifikowano je jako słowoformę rzeczownikową; np.:

Skutkiem nieostrożności spowodowali pożar. (\simeq Wskutek nieostrożności...) → **skutkiem** *Prp*
Pożar był skutkiem nieostrożności. → **skutek** *S*

d) Słowo *względem* kwalifikowano jako słowoformę przyimkową, jeżeli całe wyrażenie testowane można było w tekście — bez zmiany sensu zdania — zastąpić wyrażeniem rozpoczynającym się od słowa *wobec*; w przeciwnym wypadku kwalifikowano je jako słowoformę rzeczownikową; np.:

Miał obowiązki względem dzieci. (\simeq ... wobec dzieci.) → **względem** *Prp*
Pod pewnym względem poprawił się. → **wzгляд** *S*

e) Słowo *koło* kwalifikowano jako słowoformę przyimkową, jeżeli całe wyrażenie testowane można było w tekście — bez zmiany sensu zdania — zastąpić wyrażeniem rozpoczynającym się od słowa *obok*; w przeciwnym wypadku kwalifikowano je jako słowoformę rzeczownikową; np.:

Koło wozu pasły się konie. (\simeq Obok wozu ...) → **koło** *Prp*
Pękło nam tylne koło wozu. → **koło** *S*

4.3.4. Rzeczownik czy przysłówek

Wątpliwości powstawały przy interpretacji słów *raz* i *razem*. Pierwsze z nich kwalifikowano przysłówkowo, jeśli miało znaczenie ‘jednokrotnie’; drugie — jeśli można je było w tekście bez zmiany sensu zdania zastąpić słowem *wspólnie* (a rzeczownikowo — w pozostałych wypadkach), np.:

Olimpiady urządza się raz na cztery lata. → **raz** *Adv*
Jeszcze raz spróbujmy. → **raz** *Adv*

Odkrycia tego dokonali Rosjanie razem z Amerykanami. → **razem** *Adv*
Chodźmy razem do teatru. → **razem** *Adv*

4.3.5. Czasownik czy przymiotnik

Wątpliwości budziły słowa *winien*, *winna*, *winni* i *winny*, które mogły być interpretowane jako słowoformy haseł **winny** *Adj* albo **winien** *V* (w słowniku naszym jest też hasło **winny** *S*, ale żadna z należących do niego słowoform nie może być interpretowana jako czasownikowa). Słowo budzące wątpliwości było kwalifikowane jako słowoforma czasownikowa, jeżeli w zdaniu towarzyszył mu lub dawał się wstawić bezokolicznik, a w przeciwnym wypadku — jako słowoforma przymiotnikowa, np.:

Spis winien objąć wszystkich członków. → **winien** *V*
Jest mi winien sto złotych. → **winny** *Adj*
Rodzice winni pilnować swych dzieci. → **winien** *V*
Winni są rodzice, nie dzieci. → **winny** *Adj*

4.3.6. Czasownik czy przysłówek

Słowami homonimicznymi, następującym tę wątpliwość są: *dosyć*, *dość*, *wolno* oraz *przeszło*. Nie było możliwe sformułowanie kryterium wspólnego dla wszystkich. Wypracowano więc reguły szczególne:

a) Słowa *dosyć* i *dość* kwalifikuje się jako słowoformy przysłówkowe, jeżeli można je w tekście zastąpić słowami *dużo* lub *wystarczająco*; a jako słowoformy czasownikowe — wtedy, gdy w badanym zdaniu występowało lub dawało się wstawić wyrażenie (*było*) *nam*, np.:

Był on dość wysoki. → **dość** Adv

Dość było spojrzeć, by rzecz odgadnąć. → **dość** V

b) Słowo *wolno* kwalifikuje się jako słowoformę przysłówkową, jeżeli można je w tekście zastąpić słowem *pomału* lub *powoli*, a jako słowoformę czasownikową — wtedy, gdy w badanym zdaniu występowało lub dawało się wstawić wyrażenie (*było*) *nam*, np.:

Jechał wolno. → **wolno** Adv

Nie wolno nam palić. → **wolno** V

c) Słowo *przeszło* kwalifikuje się jako słowoformę przysłówkową, jeżeli można je w tekście zastąpić wyrażeniem *więcej niż*; a w przeciwnym wypadku — jako słowoformę czasownikową, np.:

Od tego wydarzenia minęło przeszło sto lat. → **przeszło** Adv

Od tego wydarzenia przeszło sto lat. → **przejsć** V

4.3.7. Czasownik czy partykuła

Trudności nastęczała kwalifikacja słowa *widać*. Uznano je za odpowiadające jedynej słowoformie leksemu partykułowego, jeżeli można je w tekście zastąpić słowem *widocznie*, a formie czasownikowej — jeżeli w danym zdaniu występowało lub dawało się wstawić słowo *było*, np.:

Nie przyjechał, widać nie mógł. → **widać** Prt

W dali widać było wieże miasta. → **widać** V

4.3.8. Zaimek czy przymiotnik

Trudności nastęczała kwalifikacja słów, stanowiących wykładniki tekstowe wszystkich słowoform hasel: **to** Prn, **tamto** Prn, **wszystko** Prn, **jedno** Prn, **samo** Prn. Słowoformy te są homonimiczne z formami rodzaju nijakiego liczby pojedynczej hasel przymiotnikowych, odpowiednio: **ten** Adj, **tamten** Adj, **wszystek** Adj, **jeden** Adj, **sam** Adj (słowoformy hasła **to** są uwikłane w homonimie znacznie silniej — omówimy je osobno niżej w następnych punktach). W wypadku słów, będących wykładnikami słowoform należących do wymienionych hasel przymiotnikowych, lecz mających inne wartości niż rodzaj nijaki liczby pojedynczej, np. słów mających postać wyraźnie wskazującą na liczbę mnogą lub na rodzaj żeński w liczbie pojedynczej, nie powstają trudności interpretacyjne, np.:

Zaproś tych, których znasz. → **ten** Adj

Dokonała tego sama, bez niczyjej pomocy. → **sam** Adj

W wypadku wystąpienia słowa nasuwającego wątpliwości, interpretowano je jako słowoformę należącą do hasła przymiotnikowego, jeśli występowało ono w pozycji typowo przymiotnikowej (w funkcji przydawki przymiotnej), a jako słowoformę należącą do hasła zaimkowego, jeśli występowało ono w pozycji typowo rzeczownikowej (w funkcji podmiotu, dopełnienia, okolicznika lub przydawki dopełniaczowej), np.:

Dokonała tego wyczynu sama, bez niczyjej pomocy. → **ten** Adj

Dokonała tego sama, bez niczyjej pomocy. → **to** Prn

Dążył do tego celu, aby się zemścić. → **ten** Adj

Dążył do tego, aby się zemścić. → **to** Prn

Skutki tego wydarzenia są nieobliczalne. → **ten** Adj

Skutki tego są nieobliczalne. → **to** Prn

Jest to niebagatelny problem. → **to** *Prn*

To niemowlę choruje, a tamto jest zdrowe. → **ten** *Adj*, **tamten** *Adj*
Zapytali go o to i tamto. → **to** *Prn*, **tamto** *Prn*

Kot wypił wszystko mleko. → **wszystek** *Adj*
Wszystko go martwiło. → **wszystko** *Prn*

Jedno twierdzenie jest dla mnie oczywiste. → **jeden** *Adj*
Jedno jest dla mnie oczywiste: że się mylisz. → **jedno** *Prn*

Błąd tkwił już w samym założeniu. → **sam** *Adj*
Mówili o tym samym. → **samo** *Prn*

W połączeniach *to wszystko*, *to jedno*, *to samo* traktuje się umownie *to* jako członek określający (przydawkę); a *samo*, *wszystko*, *jedno* — jako członek określany, np.:

To samo chciałem powiedzieć. → **ten** *Adj*, **samo** *Prn*
Mieliśmy szczęście w tym wszystkim. → **ten** *Adj*, **wszystko** *Prn*
To jedno warto zapamiętać. → **ten** *Adj*, **jedno** *Prn*

4.3.9. Zaimek czy spójnik

Kwalifikację jako słowoforma zaimkowa lub spójnikowa może mieć słowo *nim*. Rozróżnienie ich nie sprawia trudności, np.:

Opowiadała o nim. → **on** *Prn*
Stracił wszystko, nim się obejrzał. → **nim** *Cnj*

4.3.10. Zaimek, przymiotnik czy spójnik

Wątpliwości dotyczą słowa *tym*. Jest ono kwalifikowane jako słowoforma zaimkowa lub przymiotnikowa, jeśli spełnia warunki wymienione wyżej (por. punkt 4.3.7); jako słowoforma spójnikowa — tylko wtedy, gdy występuje jako drugi członek wyrażenia (nieciągłego) *im ... tym* albo gdy łączy się bezpośrednio z formą stopnia wyższego przysłówków lub przymiotników, np.:

Im dalej w las, tym więcej drzew. → **tym** *Cnj*
Była to tym większa radość, że niespodziewana. → **tym** *Cnj*

4.3.11. Zaimek czy partykuła

Podwójną klasyfikację tego typu mogą mieć słowa: *czego*, *czemu*, *coś*, *cóż*, *niczego* i *nie*. Kwalifikacja ostatniego z nich nie następuje trudności, pięć pierwszych uznawano za słowoformy partykułowe w kontekstach typu:

Czego się patrzysz? (≈ *Dlaczego ...*) → **czego** *Prt*
Czemu to zrobisz? (≈ *Dlaczego ...*) → **czemu** *Prt*
Coś nietego stary wygląda. → **coś** *Prt*
Cóż tak patrzysz? → **cóż** *Prt*
Panny sobie niczego! → **niczego** *Prt*

4.3.12. Zaimek, przymiotnik czy partykuła

Wątpliwości budziło słowo *tego*. Jego interpretacje zaimkowe i przymiotnikowe zostały omówione wyżej w punkcie 4.3.8. Było ono kwalifikowane jako słowoforma partykułowa w kontekstach typu:

Jakby tego ... → **tego** *Prt*

4.3.13. Zaimek, przymiotnik, spójnik czy partykuła

Wątpliwości budziło słowo *to*. Jego interpretacje zaimkowe i przymiotnikowe zostały omówione wyżej w punkcie 4.3.8. Było ono kwalifikowane jako słowoforma spójnikowa, jeśli występowało jako drugi członek wyrażenia nieciągłych: *jeżeli ...*, *to*; *choć ...*, *to*; *skoro ...*, *to* albo którykolwiek z członów wyrażenia nieciągłego *to ...*, *to*.

Słowo *to* kwalifikowano jako partykułę (*Prt*), jeżeli można je było opuścić bez zmiany sensu wypowiedzenia, np.:

Coraz **to** chłodniej! → **to** Prt
Gdyby **to** chodziło o was! → **to** Prt
Kiedy **to** wreszcie skończą się kłopoty? → **to** Prt
Brat zwrócił długi, i **to** przed terminem. → **to** Prt

4.3.14. Zaimek czy przyimek

Wątpliwości budzi słowo *niczym*. Było ono interpretowane jako słowoforma przyimkowa (wymagająca mianownika), jeżeli można je było w tekście bez zmiany sensu zdania zastąpić słowem *niby*, np.:

Była to dziewczyna **niczym** kwiat. → **niczym** Prp

4.3.15. Zaimek, przymiotnik czy przyimek

Budzące takie wątpliwości słowo *temu* było interpretowane jako słowoforma przyimkowa (wymagająca biernika i postpozycyjna), jeżeli wiązało się z określeniem czasu, np.:

Zdarzyło się to kilka lat **temu**. → **temu** Prp

4.3.16. Zaimek, przyimek, spójnik czy partykuła

Skomplikowana była interpretacja słowa *co*. Jako słowoformy zaimkowe (mianownikowe) traktowano, oprócz oczywistych użyc w pozycji rzeczownikowej, także wystąpienia *co* przy stopniu wyższym przysłówka (*co gorzej, co więcej*), np.:

Co więcej, mylisz się. → **co** Prn (*co* N)

Jako słowoformy przyimkowe traktowano *co* wprowadzające określenia czasu i odległości w mianowniku, bierniku lub dopełniaczu (*co godzina, co godzinę, co roku* itd.), np.:

Co roku siejemy i zbieramy. → **co** Prp

Funkcję słowa *co*, zwaną tradycyjnie funkcją zaimka względnego, uznano umownie za spójnikową:

To ten, **co** nie miał szczęścia. → **co** Cnj

Natomiast za słowoformę partykułową uznano *co* w kontekstach typu:

Dopiero **co** przyszedł. → **co** Prt
Było tu już wielu **co** ambitniejszych. → **co** Prt
Nie ma się **co** dziwić. → **co** Prt
Co się tak patrzysz? → **co** Prt

4.3.17. Zaimek czy przysłówek

Słowo *nic* było konwencjonalnie kwalifikowane jako słowoforma przysłówkowa w kontekstach typu:

Nic nie spałem. → **nic** Adv

4.3.18. Liczebnik czy przysłówek

Problem powstawał przy interpretacji słów *wiele* i *tyle*. Były one kwalifikowane jako słowoformy liczebnikowe, jeżeli łączyły się z rzeczownikiem; jako słowoformy przysłówkowe — jeżeli łączyły się z czasownikiem lub wyrażeniem przyimkowym; np.:

Miał **wiele** przygód. → **wiele** Num
Wiele w życiu widział. → **wiele** Adv
To rozwiązanie pozostawia **wiele** do życzenia. → **wiele** Adv
Miał **tyle** przygód, że dech zapierało. → **tyle** Num
Tyle w życiu widział, że dech zapierało. → **tyle** Adv
Miał jeszcze **tyle** do zrobienia! → **tyle** Adv

Inne słowa, będące wykładnikami słowoform leksemów liczebnikowych **tyle** i **wiele** (*tylu, tyloma, wielu, wieloma*) nie sprawiały trudności interpretacyjnych.

4.3.19. Liczebnik, przysłówek czy spójnik

Takie same jak w poprzednim punkcie są zasady liczebnikowej i przysłówkowej interpretacji słowa *ile*, ale dochodzi do nich interpretacja spójnikowa: gdy jest ono drugim członem wyrażenia *nie tyle ..., ile*. Jako przysłówkowe traktuje się też *ile* w wyrażeniu *tyle ..., ile*, np.:

*Było **tyle** wody, **ile** się jej w wiadrze zmieści.* → **tyle Adv, ile Adv**

4.3.20. Przyimek czy przysłówek

Rozróżnienie tytułowe objęło takie słowa, jak *blisko, bliżej, dokoła, dookoła, obok, około, ponad, poniżej, powyżej, wokoło, wokół*. W praktyce nie było ono łatwe. Warto wspomnieć, że w gramatykach i słownikach języka polskiego nie rozgranicza się tych funkcji wymienionych słów. Badane słowo uznawano za słowoformę przyimkową, jeśli można było uznać, że przypadek następującego po niej członu rzeczownika jest przez nią uzależniony; za słowoformę przysłówkową, jeśli warunek ten nie był spełniony (dodatkowym wskaźnikiem przysłówkowości była możliwość zastąpienia badanego słowa elementem o niewątpliwej interpretacji przysłówkowej); np.:

*Byli już **blisko** celu [G] swej wędrówki.* → **blisko Prp**
*Wydobyto **blisko** [≈ prawie] trzy tysiące ton węgla.* → **blisko Adv**
***Blisko** [≈ prawie] połowę załogi stanowili kolorowi.* → **blisko Adv**
*Trzeba go znać bardzo **blisko** [≈ wnikliwie], aby móc go ocenić.*
→ **blisko Adv**

*Byli już **bliżej** celu [G], niż myśleli.* → **blisko Prp**
*Podeszli **bliżej**.* → **blisko Adv**

***Okolo** południa [G] zaczął padać deszcz.* → **około Prp**
*Utworki te miały **około** dwustu [G] nagrań.* → **około Prp**
*Drużyna wróciła do kraju z **około** [≈ mniej więcej ≈ prawie] dwustu punktami.* → **około Adv**
*Artykuły przemysłowe stanowią **około** [≈ mniej więcej ≈ prawie] czterdzieści procent obrotów.* → **około Adv**

*Istnieją obecnie próby wznoszenia komunikacji **ponad** ziemię [A].*
→ **ponad Prp**
***Ponad** miastem [I] unosiły się ciężkie chmury.* → **ponad Prp**
*Drużyna wróciła do kraju z **ponad** [≈ więcej niż] dwustu punktami.*
→ **ponad Adv**
*Na eksport przeznaczają się **ponad** [≈ więcej niż] dziesięć procent produkcji.* → **ponad Adv**
*Przeciętnie gospodarstwa te posiadają **ponad** [≈ więcej niż] sto hektarów.* → **ponad Adv**

*Dyskusja rozwinęła się **wokół** zagadnienia [G] poruszonego w referacie.*
→ **wokół Prp**
*Nie wiedział, co się **wokół** [≈ w sąsiedztwie] dzieje.* → **wokół Adv**

Analogiczne kryteria dotyczą leksemów *dokoła, dookoła, obok, poniżej, powyżej, wokoło*.

4.3.21. Przyimek czy spójnik

Słowo *jako* kwalifikowano jako słowoformę przyimka **jako Prp** (łączyącego się z różnymi przypadkami), gdy łączyło się ono z członem rzeczownikowym; jako słowoformę spójnikową — w pozostałych wypadkach; np.:

*Wypowiedział się **jako** znawca [N].* → **jako Prp**
*Zapytano go **jako** znawcę [A].* → **jako Prp**
***Jako** rzecze Zaratustra.* → **jako Cnj**

4.3.22. Spójnik czy partykuła

Taka alternatywa powstawała przy interpretacji słów: *by*, *aby*, *ale*, *jak*, *że*. W wypadku pierwszych trzech funkcje spójnika łączącego człony i partykuły modyfikującej zdanie są wyraźnie rozgraniczone. Pokazują to np. następujące konteksty:

*Pojechał tam, **by** spotkać się ze znajomym.* → **by** Cnj

*On **by** to zrobił.* → **by** Prt

*Trzeba **by** się na to zdecydować.* → **by** Prt

*Proszono ich, **aby** nie odchodzili.* → **aby** Cnj

*Czy **aby** nie za późno?* → **aby** Prt

*Był stary, **ale** dziarski.* → **ale** Cnj

***Ale** go trzasnął!* → **ale** Prt

Słowo *jak* interpretowano jako słowoformę spójnikową, gdy miało znaczenie porównawcze, rozpoczynało zdanie niepytajne albo wprowadzało przysłówki; a jako słowoformę partykułową — gdy miało znaczenie pytajne; np.:

*Zrobiła się biała **jak** śnieg.* → **jak** Cnj

***Jak** zdobędę skierowanie, przyjadę.* → **jak** Cnj

***Jak** dotychczas, mam spokój.* → **jak** Cnj

***Jak** to zmierzyć? Z **jak** dużą dokładnością?* → **jak** Prt [dwa razy]

*Udało mi się, i to **jak**!* → **jak** Prt

Słowo *że* interpretowano jako słowoformę partykułową w kontekstach typu:

*Prawie **że** nie miał ogona.* → **że** Prt

4.3.23. Wykrzyknik czy spójnik lub przyimek

Rozróżnienie funkcji wykrzyknikowej i spójnikowej słowa *a* oraz wykrzyknikowej i przyimkowej słowa *o* nie wymaga specjalnych testów.

4.3.24. Przyimek czy partykuła

Umownie jako partykuły oznaczano słowa *z*, *ze* i *za* w kontekstach typu:

*Było tego **z** dziesięć sztuk.* → **z** Prt

*Tańczyło **ze** sto par.* → **z** Prt

*Co **za** pech!* → **za** Prt

4.3.25. Przysówek czy partykuła

Rozróżnienie tych funkcji wprowadzono tylko dla słowa *tam*, uznając za partykułowe tylko jego użycia w kontekstach typu:

*E **tam**! Co mi **tam** będziesz gadał!* → **tam** Prt [dwa razy]

4.4. Homonimia słów hasłowych różnych haseł

Homonimiczne mogą być również nazwy kilku haseł, tzn. może się zdarzyć, że w skład nazw różnych haseł wchodzi to samo słowo hasłowe. Jeśli zachodzi to dla haseł mających różną charakterystykę gramatyczną (odpowiadających leksemom różnych typów), rozróżnienie nazw artykułów hasłowych następuje automatycznie — przez dopisanie symbolu części mowy, np.: **wieść** *S* i **wieść** *V*, **tyle** *Num* i **tyle** *Adv*. Jeśli sytuacja taka obejmuje dwa hasła interpretowane jako ta sama część mowy, to przyjęte ogólnie rozwiązania nie wystarczają: do tytułu hasła zostają wówczas dopisane dodatkowe numery rozróżniające, np.: **paść**[1] *V* (hasło, do którego wchodzi m. in. słowoformy o wykładnikach tekstowych: *padnie* i *padł*) i **paść**[2] *V* (hasło, do którego wchodzi m. in. *pasie* i *pał*). Właśnie w takich wypadkach do nazwy hasła wprowadzono w nawiasie kwadratowym numer odróżniający.

Wszystkie te pary haseł o nazwach homonimicznych, których przynajmniej jeden człon znalazł się w niniejszym tomie, przytoczymy niżej w dwu tabelach.

Pierwsza obejmuje hasła, w których występuje przynajmniej jedna forma różnicująca oba hasła z pary. Przytaczamy pełne nazwy tych hasel i dla każdego podajemy jedną słowoformę, odróżniającą dane hasło od drugiego hasła z pary. Słowoformę tę opatrujemy oznaczeniem gramatycznym: jeśli przytoczona jest ona w postaci, w której występuje lub występowałaby w słowniku (dotyczy to hasel przytoczonych w odpowiednich parach dla kontrastu), podana jest bez nawiasu; jeśli oznaczenie jest dodane przy słowie, które — ze względu na reguły ogólne — nie było w słowniku opatrywane nim, ujęte jest w nawias kwadratowy.

Tabela 1

Lp.	Forma hasłowa	Forma różnicująca	Częstość
1.	bal [1] S	balu Gsg	8
	bal [2] S	bala Gsg	1
2.	boleć [1] V	boli [3sg]	30
	boleć [2] V	boleje [3sg]	1
3.	dojrzeć [1] V	dojrzy [3sg]	8
	dojrzeć [2] V	dojrzeje [3sg]	6
4.	dopaść [1] V	dopadnie [3sg]	7
	dopaść [2] V	dopasie [3sg]	brak
5.	dostać [1] V	dostanie [3sg]	102
	dostać [2] V	dostoi [3sg]	brak
6.	głęb [1] S	głębi Gsg	32
	głęb [2] S	głęba Gsg	1
7.	lew [1] S	lwa Gsg	12
	lew [2] S	lewa Gsg	brak
8.	łyk [1] S	łyku Gsg	6
	łyk [2] S	łyka Gsg	brak
9.	mars [1] S	marsa Gsg	5
	mars [2] S	marsu Gsg	brak
10.	metr [1] S	metra Gsg	167
	metr [2] S	metru Gsg	brak
11.	muł [1] S	muła Gsg	1
	muł [2] S	mułu Gsg	6
12.	nastać [1] V	nastanie [3sg]	8
	nastać [2] V	nastoi [3sg]	brak
13.	opaść [1] V	opadnie [3sg]	9
	opaść [2] V	opasie [3sg]	brak
14.	pas [1] S	pasa Gsg	33
	pas [2] S	pasu Gsg	2
15.	paść [1] V	padnie [3sg]	36
	paść [2] V	pasie [3sg]	1

Lp.	Forma hasłowa	Forma różnicująca	Częstość
16.	plastyk[1] S	plastyka Gsg	7
	plastyk[2] S	plastyku Gsg	3
17.	pośłać[1] V	pośle [3sg]	7
	pośłać[2] V	pościele [3sg]	brak
18.	przepaść[1] V	przepadnie [3sg]	10
	przepaść[2] V	przepasie [3sg]	brak
19.	prześłać[1] V	prześle [3sg]	16
	prześłać[2] V	prześciele [3sg]	brak
20.	przestać[1] V	przestanie [3sg]	78
	przestać[2] V	przestoi [3sg]	brak
21.	rozpadać się[1] V [ndk]	rozpadając się [im.przysł.współ.]	5
	rozpadać się[2] V [dk]	rozpadawszy się [im.przysł.uprz.]	brak
22.	rozpaść się[1] V	rozpadnie się [3sg]	4
	rozpaść się[2] V	rozpasie się [3sg]	brak
23.	rząd[1] S	rządu Gsg	48
	rząd[2] S	rzędu Gsg	232
24.	spaść[1] V	spadnie [3sg]	30
	spaść[2] V	spasie [3sg]	brak
25.	upaść[1] V	upadnie [3sg]	11
	upaść[2] V	upasie [3sg]	brak
26.	ustać[1] V	ustanie [3sg]	7
	ustać[2] V	ustoi [3sg]	brak
27.	wypaść[1] V	wypadnie [3sg]	28
	wypaść[2] V	wypasie [3sg]	brak
28.	wysłać[1] V	wyściele [3sg]	40
	wysłać[2] V	wyśle [3sg]	brak
29.	wywrzeć[1] V	wywarł [masc.praet]	5
	wywrzeć[2] V	wywrzał [masc.praet]	brak
30.	zamek[1] S	zamka Gsg	13
	zamek[2] S	zamku Gsg	15
31.	zapaść[1] V	zapadnie [3sg]	9
	zapaść[2] V	zapasie [3sg]	brak
32.	zawrzeć[1] V	zawarł [masc.praet]	13
	zawrzeć[2] V	zawrzał [masc.praet]	2
33.	zbieg[1] S	zbiegu Gsg	9
	zbieg[2] S	zbiega Gsg	3

Przyjęto zasadę, że jednostki homonimiczne są rozróżniane tylko na podstawie cech gramatycznych, nie zaś znaczeniowych. Dlatego też w słowniku prawie nigdy nie rozróżniano haseł, w których skład wchodzi dokładnie takie same

słowoformy (tj. realizowane przez te same słowa i mające takie same charakterystyki gramatyczne). Jedyny wyjątek stanowi niewielka (obejmująca 6 par) grupa tzw. homonimów słowotwórczych, np.: **ranny[1]** *Adj* (od **rana**) i **ranny[2]** *Adj* (od **rano**). Wszystkie one są przymiotnikami odrzeczownikowymi. Oto pełna lista par takich haseł, których przynajmniej jeden człón znalazł się w niniejszym tomie:

Tabela 2

Lp.	Forma hasłowa	Podstawa słowotwórcza	Częstość
1.	gazowy[1] <i>Adj</i>	gaz <i>S</i>	16
	gazowy[2] <i>Adj</i>	gaza <i>S</i>	brak
2.	piątkowy[1] <i>Adj</i>	piątek <i>S</i>	5
	piątkowy[2] <i>Adj</i>	piątka <i>S</i>	brak
3.	plastyczny[1] <i>Adj</i>	plastyka <i>S</i>	15
	plastyczny[2] <i>Adj</i>	plastyk[2] <i>S</i>	5
4.	pocztowy[1] <i>Adj</i>	poczta <i>S</i>	12
	pocztowy[2] <i>Adj</i>	poczet <i>S</i>	brak
5.	ranny[1] <i>Adj</i>	rana <i>S</i>	22
	ranny[2] <i>Adj</i>	rano <i>S</i>	5
6.	szachowy[1] <i>Adj</i>	szachy <i>S</i>	6
	szachowy[2] <i>Adj</i>	szach <i>S</i>	brak

Natomiast nie rozróżniano homonimów przez odwoływanie się wyłącznie do ich znaczenia (np. potencjalne jednostki *para*₁ [*matżeńska*] i *para*₂ [*wodna*] zostały w słowniku ujęte w tym samym haśle **para** *S*).

5. Wskaźniki statystyczne w słowniku

W skład naszego słownika wchodzi cztery listy — lista główna oraz trzy listy rangowe: według wskaźnika *F* (częstość absolutna), wskaźnika *U* (częstość względna) oraz wskaźnika *D* (wskaźnik równomierności rozkładu w stylach).

Częstość absolutna *F* jest to podstawowy wskaźnik liczbowy otrzymany drogą zwykłego sumowania wystąpień poszczególnych jednostek wprowadzonych do słownika. W naszym wypadku wskaźnik *F* otrzymano przez zsumowanie liczb f_i — wystąpień hasła w poszczególnych stylach i ($i = 1, 2, 3, 4, 5$) (dla każdego stylu liczba ta została otrzymana przez zsumowanie liczby wystąpień słowoform, wchodzących w skład hasła). Częstość absolutna *F* oznacza częstość danego leksemu w całej badanej próbie 500 000 słów:

$$(1) \quad F = \sum_{i=1}^5 f_i$$

Drugi — obok częstości — podstawowy wskaźnik liczbowy, podawany w słownikach frekwencyjnych, jeśli struktura próby jest wewnętrznie zróżnicowana, charakteryzuje rozproszenie jednostki w poszczególnych jej podzbiorach. Materiał zasadniczy naszego słownika rozpada się, rzecz jasna, na pięć podzbiorów odpowiadających poszczególnym stylom. W słowniku podawany jest wskaźnik równomierności rozkładu *D*, zwany również dyspersją złożoną (w odróżnieniu od stosowanej niekiedy tzw. dyspersji prostej, równającej się po prostu liczbie podzbiorów korpusu zasadniczego, w których występuje element badany).

Oblicza się go w sposób dość skomplikowany. Dlatego też nie ograniczymy się tylko do podania wzoru, lecz wprowadzimy go stopniowo, wyjaśniając sens poszczególnych operacji.

Niech n oznacza liczbę stylów funkcjonalnych, wyróżnionych w danym słowniku frekwencyjnym, zaś f_i — częstość danego hasła w stylu i ($i = 1, \dots, n$). Na ich podstawie liczymy średnią częstość danej jednostki w stylu:

$$(2) \quad \bar{f} = \frac{\sum_{i=1}^n f_i}{n}$$

W naszym przypadku średnia częstość wynosi:

$$(3) \quad \bar{f} = \frac{F}{5}$$

Dokładniejszą miarą stopnia rozproszenia częstości jednostki w poszczególnych częściach korpusu jest odchylenie standardowe s , określające przeciętne odchylenie częstości badanej jednostki od częstości średniej:

$$(4) \quad s = \sqrt{\frac{\sum_{i=1}^n (f_i - \bar{f})^2}{n - 1}}$$

Wadą odchylenia standardowego jako wskaźnika jest to, że jego wartość jest mocno uzależniona od wartości średniej badanej jednostki (przy wysokiej średniej odchylenie standardowe ma duże wartości, przy małej — niewielkie). Relatywne odchylenia częstości badanej jednostki w stosunku do częstości średniej podaje współczynnik zmienności:

$$(5) \quad v = \frac{s}{\bar{f}}$$

Zarówno odchylenie standardowe, jak i współczynnik zmienności osiągają duże wartości, gdy częstości badanej jednostki f_i w poszczególnych stylach są zróżnicowane, a mają wartość 0, gdy wszystkie te częstości mają wartość identyczną ($f_1 = \dots = f_n$). Dlatego słownik nasz wprowadza wskaźnik dyspersji złożonej, której zależność od zróżnicowania tych wartości jest odwrotna:

$$(6) \quad D = 100 \cdot \left(1 - \frac{v}{\sqrt{n}}\right) = 100 \cdot \left(1 - \frac{\sqrt{\sum_{i=1}^n (f_i - \bar{f})^2}}{\bar{f} \sqrt{n(n-1)}}\right)$$

W naszym przypadku:

$$(7) \quad D = 100 \cdot \left(1 - \frac{v}{\sqrt{5}}\right) = 100 \cdot \left(1 - \frac{\sqrt{\sum_{i=1}^5 (f_i - \bar{f})^2}}{2\sqrt{5} \cdot \bar{f}}\right)$$

Ten właśnie wskaźnik podawany jest w naszym słowniku. Bez trudu możemy dostrzec, że jego wartości wahają się między 0 (przy maksymalnie skoncentrowanym rozkładzie danej jednostki — w jednym tylko stylu) i 100 (przy całkowicie równomiernym rozkładzie danej jednostki we wszystkich częściach korpusu).

Trzeci podstawowy wskaźnik liczbowy, podawany w słownikach frekwencyjnych, częstość relacyjna U powstaje przez przemnożenie wskaźników F i D , relatywizując w ten sposób dane o częstości. Częstość absolutna jest tu mianowicie pomniejszona proporcjonalnie do stopnia nierównomierności rozkładu w poszczególnych stylach:

$$(8) \quad U = \frac{F \cdot D}{100}$$

W wypadku maksymalnie równomiernego rozkładu częstości (gdy $D = 100$) częstość relacyjna U pokrywa się z częstością absolutną F .

6. Struktura słownika

Na słownik nasz składa się lista główna oraz trzy listy rangowe: według wskaźników F, U i D. Na wszystkich czterech listach znalazły się te hasła z zasadniczego korpusu o częstości nie mniejszej niż 4, które nie zostały pominięte — na podstawie ogólnych zasad (por. wyżej punkt 3.4). Wszystkie cztery listy zawierają więc ten sam materiał; zasadnicza różnica między listą główną (alfabetyczną) a listami rangowymi jest taka, że lista główna podaje dane dotyczące zarówno haseł, jak i ich słowoform, podczas gdy listy rangowe ograniczają się do informacji na temat haseł¹¹.

6.1. Struktura listy głównej (alfabetycznej) i budowa artykułu hasłowego

Lista główna podaje zasadnicze dane liczbowe o hasłach i ich słowoformach, uzyskane w trakcie prac nad słownikiem. Lista ta gromadzi wszystkie hasła wytypowane do tomu na podstawie wypracowanej procedury.

O miejscu na liście głównej decyduje nazwa hasła. Nazwy haseł są uporządkowane alfabetycznie według słów hasłowych w kolejności *słowo po słowie*. Zgodnie z nią spacja (odstęp międzywyrazowy) traktowana jest jako osobny znak — znak ten w alfabecie poprzedza wszystkie litery. Jeśli słowa hasłowe dwu haseł są identyczne, o kolejności decyduje następny element nazwy hasła: numer odróżniający lub symbol części mowy. Jeśli zachodzi potrzeba porządkowania według symboli części mowy, nie są one układane według alfabetu, lecz zgodnie z numeracją części mowy, przyjętą przy wprowadzaniu materiału do pamięci komputera (w naszym tomie podane są one w tym porządku w wykazie skrótów). Tak więc na liście głównej hasła występują np. w następującej kolejności: **o** *Prp*, **o** *Int*, **o** *ile Cnj*, **o** *tyle Cnj*, **o** *wiele Adv*, **oba** *Num*, **obalenie** *S*.

Artykuł hasłowy na liście głównej składa się z szeregu wierszy. W pierwszym wierszu podana jest nazwa hasła, w drugim — dane dotyczące hasła w całości, w następnych — dane dotyczące poszczególnych słowoform.

Poczynając od wiersza drugiego, artykuł hasłowy podzielony jest na siedem kolumn. W ostatniej kolumnie wiersza trzeciego i dalszych wypisane są słowoformy, które wystąpiły w materiale. W pięciu pierwszych kolumnach występują dane liczbowe wskazujące liczbę wystąpień badanego elementu (dla wiersza drugiego — hasła, dla dalszych wierszy — słowoformy, której poświęcony jest ten wiersz) w pięciu stylach objętych słownikiem (I — teksty popularnonaukowe, II — drobne wiadomości prasowe, III — publicystyka, IV — proza artystyczna, V — dramat artystyczny); w szóstej kolumnie — dane liczbowe wskazujące łączną liczbę wystąpień tego elementu we wszystkich stylach. Tak więc liczba podana w drugim wierszu stanowi sumę liczb podanych w danym artykule hasłowym pod nią w tej samej kolumnie, liczba podana w szóstej kolumnie — sumę liczb podanych w danym artykule hasłowym przed nią w tym samym wierszu. Liczba podana w drugim wierszu w szóstej kolumnie określa wartość wskaźnika F dla danego hasła.

Na siódmym miejscu w wierszu drugim artykułu hasłowego podane są trzy liczby. Pierwsza określa wartość wskaźnika D dla danego hasła, druga — wartość wskaźnika U dla tego hasła. Liczba trzecia, przytoczona w nawiasie, podaje liczbę wystąpień danego hasła w tekstach polszczyzny mówionej według H. Zgółkowej.

6.2. Struktura list rangowych

Budowa wszystkich list rangowych jest identyczna; różnią się one tym, że względu na jaki wskaźnik uporządkowane są zamieszczone w nich hasła: F (czę-

¹¹ Dla porządku odnotujmy, że przy przytaczaniu danych w naszym słowniku przyjęliśmy konwencję nie stosowaną na ogół w drukach polskich: część ułamkowa liczby jest mianowicie oddzielana od części całkowitej za pomocą kropki, a nie przecinka.

stość absolutna), U (częstość względna) i D (wskaźnik równomierności rozkładu w stylach). Na wszystkich listach hasła uporządkowano według danego wskaźnika w porządku malejącym. Na pozycję listy rangowej składają się trzy elementy: numer kolejny (ranga), nazwa hasła oraz wartość wskaźnika. W wypadku, gdy konkretna wartość wskaźnika przyporządkowana jest kilku hasłom, o ich kolejności (i związanych z nią numerach na liście) decyduje porządek alfabetyczny.

7. Kilka uwag o interpretacji materiału

7.1. Dokładność i wiarygodność oszacowań

Prezentując niniejszy słownik, zdecydowano się na pominięcie hasel o bardzo małej liczbie wystąpień, dla których wskaźnik częstości absolutnej F jest mniejszy od 4, oraz hasel o specjalnym charakterze (nazw własnych, skrótów). Ostatecznie słownik nasz zawiera 10 355 hasel spośród 38 469 znajdujących się na liście zbiorczej hasel zawartych w tomach *SWJP*¹².

Jeśli idzie o materiał do słownika zaliczony, to posługując się klasycznymi metodami statystycznymi, możemy obliczyć dolną granicę wartości, począwszy od których hasła słownika, traktowane jako elementy populacji generalnej, podawane są w nim z ustaloną dokładnością (nie mniejszą niż założony z góry dopuszczalny błąd). W tym celu posłużymy się wzorem na dopuszczalny błąd szacowania δ , którego uwzględnienie jest konieczne przy wnioskowaniu statystycznym:

$$(9) \quad \delta = \frac{u_\alpha}{\sqrt{N \cdot p}}$$

Zadaniem naszym jest wyznaczenie wielkości p , oznaczającej prawdopodobieństwo wystąpienia badanej jednostki (powinno to być prawdopodobieństwo teoretyczne p_{teor} , jednak zamiast niego — ze względów praktycznych — będziemy posługiwać się prawdopodobieństwem empirycznym wystąpienia tej jednostki, które obliczamy, dzieląc liczbę jej wystąpień w próbie przez wielkość próby; w naszym wypadku dla danego hasła h : $p_{emp}(h) = \frac{F(h)}{500\,000}$). We wzorze (9) N oznacza wielkość badanej próby, a współczynnik u_α jest braną z tablic rozkładu Gaussa wartością odpowiadającą ustalonemu współczynnikowi ufności α . Wielkość $(1 - \alpha)$ oznacza prawdopodobieństwo, z jakim parametr p zawiera się w określonym przedziale, zwanym przedziałem ufności.

Przyjmijmy współczynnik ufności $\alpha = 0,05$. Odpowiada mu w tablicach rozkładu Gaussa wartość $u_\alpha = 1,96$. Przeprowadźmy obliczenie dla dopuszczalnego błędu szacowania 0,3. Po przekształceniu wzoru (10) i podstawieniu doń wartości liczbowych, otrzymujemy:

$$(10) \quad p = \frac{1,96^2}{0,3^2 \cdot 500000} \approx 0,00008$$

Z dokładnością 0,3 są zatem oszacowane przez nas w słowniku jednostki, które wystąpiły w całej naszej próbie z częstością nie mniejszą niż 8 jednostek na 100 000 (czyli — z nieco większą dokładnością — nie mniej niż 43 razy w całej próbie o objętości 500 000 słów).

Obliczenia dla innych wartości dopuszczalnego błędu szacowania oraz współczynnika ufności pozostawiamy czytelnikowi¹³. Wspomnijmy tylko, że wiarygodność otrzymanych wyników osłabiają jeszcze dwa czynniki, nie dające się ująć

¹² Prawie wszystkie podstawowe dane liczbowe w niniejszym punkcie przytoczone są na podstawie tej listy, która daje znacznie łatwiejszy dostęp do całości opracowanego materiału niż tekst przygotowany do niniejszego wydania. Ze względu na wprowadzone później w korektach poprawki dane te mogą nie w pełni odpowiadać zawartości tego tomu. Różnice są jednak statystycznie nieistotne.

¹³ Bliższe informacje na temat teorii szacowania parametru p oraz odpowiednie przykłady znaleźć można w pracy: R. Hammerl, J. Sambor, *Statystyka dla językoznawców*. Wydawnictwa Uniwersytetu Warszawskiego: Warszawa 1990, s. 210-216.

w liczby i wzory: niezbyt wyraźne określenie populacji generalnej, dla której przeprowadza się badania, oraz fakt, że zbiór leksemów danego języka etnicznego jest potencjalnie nieskończony.

Podane wyżej obliczenia wskazują, że żądania statystycznej wiarygodności spełniają jednostki o stosunkowo wysokiej częstości. Mimo to zdecydowano się podać w słowniku jednostki o dużo niższych częstościach, ponieważ w swej masie dają jednak wyobrażenie o rozkładzie polskiego słownictwa w tekstach o znacznej objętości. Słownik nasz daje przy tym pierwszy zestaw polskiego słownictwa o znacznej objętości, opracowany ścisłymi metodami.

Dla zorientowania czytelnika w objętości materiału przytoczymy niżej kilka danych na temat haseł pominiętych przy opracowaniu całości:

	liczba haseł	liczba wystąpień
nazwy własne	7 472	18 879
skrót	407	1 941
inne hasła o frekwencji 1	12 941	12 941
inne hasła o frekwencji 2	4 773	9 546
inne hasła o frekwencji 3	2 475	7 425

Jeśli idzie o hasła uwzględnione w słowniku, to mają następujące częstości:

Tabela 3

Częstość	Liczba haseł	Liczba wystąpień
4	1 636	7 544
5	1 116	5 580
6	832	4 992
7	670	4 690
8	548	4 384
9	439	3 951
10	364	3 640
11 – 20	1 909	27 896
21 – 30	816	20 370
31 – 40	497	17 513
41 – 50	299	13 555
51 – 60	194	10 680
61 – 70	167	10 947
71 – 80	114	8 608
81 – 90	94	7 996
91 – 100	78	7 414
101 – 200	354	28 035
201 – 300	104	25 485
301 – 400	52	18 206
401 – 500	30	13 409
501 – 600	12	6 713
601 – 700	11	7 203
701 – 800	11	8 308
801 – 900	9	7 883
901 – 1 000	5	4 826
ponad 1 000	40	148 711

Procent tekstu zajmowany przez wystąpienia haseł o poszczególnych częstościach obliczy czytelnik samodzielnie.

7.2. Strefy leksyki

Wskaźniki F, D i U pozwalają na wydzielenie — z pewnym przybliżeniem — określonych stref leksyki. W zależności od wartości tych wskaźników można

wyróżnić strefę słownictwa gramatycznego, podstawowego, charakterystycznego i rzadkiego.

Do strefy *słownictwa gramatycznego* należą hasła o największych częstościach (wskaźniki F i U) oraz równomierności rozkładu bliskiej 100 (możemy ją konwencjonalnie określić $D > 80$). Do strefy tej wpadają podstawowe proste przyimki (*w, z, za*), spójniki (*że, i*), zaimki (*on*), czasowniki posiłkowe (*być*).

Strefa *słownictwa podstawowego* zawiera hasła, odpowiadające leksemom autosemantycznym, mające bardzo wysoką częstość absolutną i względną oraz dość dużą równomierność rozkładu częstości (np. $D > 50$). W strefie tej znajdują się takie hasła, jak rzeczowniki *człowiek, związek, część*, przymiotniki *wielki, główny, inny*, czasowniki *znać, istnieć*, przysłówki *bardzo, często, daleko* i in.

W strefie *słownictwa charakterystycznego*, a więc nacechowanego stylistycznie znajdują się hasła częste, ale ograniczone w swym występowaniu do jednego lub dwóch stylów funkcjonalnych. Ze względu na niską równomierność rozkładu słownictwo to ma, mimo wysokiej częstości absolutnej F, bardzo małą, a nawet bliską zera częstość relacyjną U. Można uznać, że w naszym słowniku do strefy tej należą hasła, dla których spełnione są warunki: $F \geq 10$, $D \leq 50$.

Słownictwo charakterystyczne dla stylu naukowego to przede wszystkim terminologia (*atom, częstotliwość, elektron*), dla stylu wiadomości prasowych — słownictwo związane z tematyką tekstów (*bramka, ambasada, festiwal*), natomiast w stylach artystycznych do strefy tej należą wyrazy codzienne, potoczne (np. *bać się, gniewać się, babcia, gęba* itp.).

Strefę *słownictwa rzadkiego* wyznacza tylko wskaźnik częstości absolutnej, dla którego możemy przyjąć umownie wartość $F \leq 5$. Wartość wskaźnika D może być dowolna: od maksymalnej (por. hasła o wartości $D = 100$ na liście rangowej) do minimalnej (hasła skupione w jednym stylu mają wartość $D = 0$, por. *antygen, baryczny, bąknąć, blednąć*).

”Latwo zauważyć, że słownictwo rzadkie o skupionych rozkładach częstości odznacza się dużym podobieństwem do słownictwa charakterystycznego. Nacechowanie stylistyczne wielu leksemów rzadkich, nawet o częstości 4 i 5, spowodowane jest wysoce atomistyczną strukturą naszego korpusu, gdyż w zbiorze 2 000 próbek 50-wyrazowych danego stylu już np. czterokrotne pojawienie się leksemu *antygen* jest znaczące. Z tych właśnie względów zdecydowano, by lista frekwencyjna leksykalno-morfologiczna obejmowała słownictwo o częstości nie mniejszej niż 4.