

Opis kodu korpusu

W każdym tomie zdającym sprawę z prac nad listami frekwencyjnymi [*Kurcz i in.*], we

wstępie zamieszczono instrukcje wpisywania tekstów na fiszki oraz „Opis gramatyczny”. Jest

to cenne źródło informacji o kodzie, użytym w korpusie do odróżnienia form homonimicznych. Tekst

drukowany w latach siedemdziesiątych wydaje się jednak mało czytelny

¹. Poniżej przytaczam fragmenty wstępu trzeciego

tomu, z wprowadzonymi nawiasami w miejsce pojedynczych ukośników i pojedynczych ukośników w

miejsce ukośników podwójnych. Podkreślenia wyrazów linią przerywaną, odróżniające w tekście

przykłady analizowanych form, zastąpiłam przez pochylony krój czcionki. Używam polskiego

cudzysłowu. Cyfry stanowiące oznaczenia gramatyczne, jeżeli zawierają tzw. cyfrę na pierwszej

pozycji (oznaczenie części mowy), podaję w nawiasach kwadratowych (tak, jak w komputerowej wersji próbek).

Wstęp

1.1 Jednostka badania

Ze względu na leksykalno-morfologiczny typ podjętych badań frekwencyjnych podstawową jednostką analizy w tekście jest słowoforma, czyli forma fleksyjna

wyrazu. Wyraz definiujemy — z pewnymi wyjątkami — według definicji graficznej, określając go jako ciąg liter zawartych między dwoma kolejnymi

pauzami. Takie ujęcie było konieczne przy maszynowym opracowaniu danych. Z tego

też względu na formy czasów złożonych typu *będę jechał*, *pojechali*

¹Autorzy książki

nie mieli możliwości pisania pewnych znaków, np. nie mogli używać nawiasów — w ich miejsce wpisywali

ukośnik; posługiwali się także podwójnym ukośnikiem.

byli traktuje się jako składające się z dwóch odrębnych słowoform; funkcje tych słowoform jako składowych form czasów złożonych wskazuje jedynie specjalny kod liczbowy (por. Kod fleksyjny str. 40). Podobnie — zgodnie z

zasadą analityczności — wyróżnia się dwie odrębne słowoformy w zestawieniach

nominalnych typu *Stany Zjednoczone*; jako ciągi osobnych słowoform traktujemy także zestawienia liczebników porządkowych i głównych typu *sześćdziesiąty piąty, tysiąc dwieście czternaście* itp.

Od reguły tej stosowano jednakże wyjątki, ujmując niekiedy jako jedną słowoformę segment tekstu składający się z dwóch formalnie różnych członów.

Dotyczyło to następujących wypadków:

1. Dwa formalnie odrębne człony stanowią jednostkę zleksykalizowaną, przy czym przynajmniej jeden z członów utracił samodzielność, stanowiąc archaizm leksykalny bądź fleksyjny — są to wyrażenia typu *po prostu, za młodu, po omacku, na bakier* itp.
2. Dwa formalnie odrębne człony nie dopuszczają wstawienia między nie innego członu, np. *w ogóle, przede wszystkim, na razie* itp.
Tę zasadę stosowano tylko do niektórych leksemów wchodzących w skład utartych połączeń.
3. Częstki *de, von* stojące przed obcymi nazwiskami stanowią wraz z nazwiskiem jedną słowoformę (*de Gaulle, von Hassel*).

Jako jedną słowoformę traktuje się także:

1. imiesłowy przymiotnikowe i rzeczowniki odczasownikowe z zaimkiem *się* (*skradający się, skradanie się*);
2. częstki *co* i *jak* stojące przed przymiotnikami i przysłówkami w stopniu najwyższym wraz z danymi częściami mowy (*co naj-, jak naj-*);
3. częstkę *do* + przysłówek (typ *do dziś*);

4. cząstkę *na* + przysłówkę (typ *na gorąco*, *na bieżąco*,
na nowo, *na zawsze*);
5. cząstkę *od* + przysłówkę (typ *od wczoraj*, *od dziś*);
6. cząstkę *z* + przysłówkę (typ *z zewnątrz*);
7. cząstkę *za* + przysłówkę/przymiotnik (typ *za dużo*,
za mało, *za duży*, *za mały*).

Jeszcze inaczej potraktowano czasowniki zwrotne, zaliczając do nich z punktu

widzenia formalnego wszystkie czasowniki z zaimkiem *się*. Ze względu na definicję graficzną wyrazu oraz biorąc pod uwagę fakt, że informacja o częstości zaimka *się* jako odrębnej słowoformy jest ważna dla językoznawców, interpretowano czasownik zwrotny jako składający się z dwóch

słowoform (typ *biją się*). Aby jednakże nie utracić informacji językowej o zwrotności danego czasownika, umieszczano przy słowoformie czasownika symbol

501 jako znak kodu fleksyjnego wskazujący na zwrotność (*bili* 501 *się*). Stąd też wszystkie słowoformy czasowników zwrotnych (z wyjątkiem imiesłówów przymiotnikowych traktowanych jako osobne hasła) opatrzone są

symbolem 501. Program hasłowania pozwalał w toku opracowania na przekształcenie

symbolu 501 na zaimek *się* przy hasle, stąd „pełna” postać haseł czasowników zwrotnych (por. hasło *bić się*).

W wielu wypadkach wątpliwych dotyczących postaci hasła przyjmowano rozstrzygnięcia zawarte w „Słowniku Języka Polskiego” pod red. W. Doroszewskiego.

2 Instrukcja I

2.1 Rozpisywanie tekstów

1. Skróty rozwiązywano, traktując je jako normalne słowoformy:

prof. = *profesor*, dla *dra Kowalskiego* =

dla *doktora Kowalskiego*, podobnie *mgr*, *doc.*,

tow., *ob.* i wszelkie inne — takie jak: *np.* = *na*

*przykład, itd. = i tak dalej, br. =
bieżącego roku, itp. = i tym podobne.*

2. Jednostki miar pisane skrótami rozwiązywano: *kg = kilogram, cm = centymetr, godz. = godzina.*
3. Liczby pisane cyframi arabskimi i rzymskimi przepisywano słownie:

*po XX zjeździe = po dwudziestym zjeździe,
o godz. 9 = o godzinie dziewiątej,
w 1966 r. = w tysiąc dziewięćset
sześćdziesiątym
szóstym roku.*

4. Złożenia z członem cyfrowym przepisywano słownie:
50-lecie = pięćdziesięciolecie, 3-dniowy = trzydniowy.
5. Skrótownice (literowce, głoskowce, sylabowce) nie były rozwiązywane i traktowano je jako osobne słowoformy: *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.*
6. Pominięto w tekście wszelkie wzory oraz symbole typu H_2O , odcinek AB itp.

3 Instrukcja II

3.1 Imiona własne

Imię własne znaczymy znakiem +.

1. Za imię własne uważa się:
 - (a) imiona i nazwiska ludzi i zwierząt (*Jan + Kowalski +, Azor +*); także, jeśli są homonimiczne z appelatywami (np. *Puzon +, Gołąb +*);

- (b) nazwy umowne instytucji, pojazdów lądowych i wodnych (*kawiarnia Sarenka +*,
statek Orzeł +, *ekspres Błękitna + Fala +*);
- (c) nazwy geograficzne, np. *Wielka + Brytania +*, *Związek + Radziecki +*, *Stany + Zjednoczone +*, *Morze + Czarne +*,
Bagno + (dzielnica Warszawy), *Bliski + Wschód +*;
- (d) tytuły książek, czasopism, filmów, referatów (*Ogniem + i + mieczem +*,
Życie + Gospodarcze +, *Tam + gdzie + rosną + poziomki +*, *O + aktualnej + sytuacji + politycznej + w + świecie +*).

2. Nie są imionami własnymi:

- (a) pełne, nieumowne nazwy instytucji typu *Spółdzielnia pracy szewców*
i cholewkarzy, *Centrala Handlowa Przemysłu Mleczarskiego*;
- (b) przymiotniki utworzone od imion własnych, np. *bliskowschodni*,
amerykański,
warszawski;
- (c) nazwy mieszkańców danego regionu typu *poznanian*, *Amerykanin*,
Chińczyk;
- (d) nazwy firmowe samochodów, papierosów, słodczy, np. *moskwicz*,
wołga,
warszawa, *sporty*, *giewonty*, *katarzynki* itp.

3. Znakiem + znaczone także umowne symbole klas sportowych (*drużyna zakwalifikowała się do klasy a +*) oraz witamin (*witamina b +*).

4 Opis gramatyczny

4.1 Kod fleksyjny

Leksykalno-morfologiczne dane frekwencyjne z naszego opracowania dotyczą

częstości zarówno haseł, jak i form fleksyjnych. Przy ustalaniu haseł na podstawie słowoform występujących w izolacji napotykamy na różne typy form

homonimicznych. Jak wiadomo, istnieją 3 typy homonimii:

1. homonimia leksykalna typu *pokój, zamek*;
2. homonimia morfologiczna dotycząca form fleksyjnych; jest nią w obrębie nomen
homonimia przypadków i liczb (*sprawności* jako gen., dat., loc., sg. i gen. pl.,
dobrym jako instr., loc. sg. oraz dat. pl.);
3. homonimia syntaktyczna dotycząca poszczególnych części mowy, polegająca na tym, że
dany wyraz może pełnić funkcje syntaktyczne rzeczownika bądź czasownika
(*brak, potrzeba*), rzeczownika bądź przysłówka (*czasem*),
rzeczownika bądź przymiotnika (*chory, pośpieszny*), czasownika bądź przysłówka (*wolno*), przysłówka bądź przyimka (*blisko, obok, wokół*) itd.

W opracowaniu naszym pominięto homonimię leksykalną, przede wszystkim ze

względu na trudności teoretyczne w odróżnianiu homonimii leksykalnej od

polisemii. Homonimię morfologiczną i syntaktyczną usuwano, różnicując formy homonimiczne przez dopisywanie umownych symboli cyfrowych.

Ze względu na

ograniczoną pojemność pamięci maszyny liczba symboli gramatycznych została

ograniczona do 63; ułożony kod jest kodem pozycyjnym.

Zwracamy uwagę, że w opracowywanym materiale leksykalnym symbolizacja

odpowiednich cech gramatycznych dotyczy tylko słowoform i haseł homonimicznych,

nie obejmuje więc ona całości badanego słownictwa. Słowoformy i hasła nie

kodowane oznaczają więc formy nie będące homonimami, których funkcje

morfologiczno-syntaktyczne łatwo odczytać z samej postaci wyrazu (por. *domami*).

Tylko w dwóch wypadkach — przy rzeczownikach i przy liczebnikach brak kodu ma

jeszcze inne, odrębne znaczenie. Przy rzeczownikach oznacza on także nieodmienność członu, pełniącego funkcję przydawki rzeczownikowej (przykłady

typu *w hucie Pokój +, do spółdzielni Jedność +*), a także występującego przy odmianie tytułów i nazwisk żeńskich (*z doktor Kowalską +, nagrodę przyznano Janinie + Bator +*).

Przykłady dotyczące liczebników wymagają szerszego omówienia i podane są niżej (por. str. 50).

Homonimie syntaktyczną usuwamy, umieszczając przy homonimicznym haśle symbol

części mowy. Są to cyfry od 1 do 9 występujące *w pierwszej pozycji* po słowoformie. Interpretuje się je następująco:

1. „rzeczownik” — symbol przypisywany rzeczownikom i substantywizowanym przymiotnikom;
2. „przymiotnik” — symbol przypisywany przymiotnikom, liczebnikom i zaimkom przymiotnikowym oraz imiesłowom przymiotnikowym;
3. „liczebnik” — symbol przypisywany liczebnikom głównym i zbiorowym oraz zaimkom liczebnym (*ile, tyle, wiele*);
4. „zaimek” — symbol oznaczający zaimki rzeczownikowe;
5. „czasownik” — symbol oznaczający formy czasownika i prädicativa (*potrzeba, brak*);
6. „przyimek” — symbol oznaczający przyimki pierwotne i wtórne (*blisko, obok*);
7. „wykrzyknik” — symbol odróżniający wykrzykniki od spójników (*a, o*);
8. „przysłówek” lub „partykuła” — symbol podawany przy słowoformach nieodmiennych nie będących przyimkami ani spójnikami;

9. „spójnik”.

Pozycja II występuje jedynie wówczas, gdy w pozycji I użyto symboli 1–6. W obrębie rzeczowników [1] i przymiotników [2] odróżniamy homonimie przypadków (symbole 1–6 na II pozycji) i liczb (symbole 1–2 na III pozycji). Nie uwzględniono natomiast homonimii w obrębie kategorii rodzaju, gdyż pomnożyłoby to znacznie liczbę symboli. Przy czasowniku [5] II pozycja, na której występują cyfry od 1 do 4, pozwala odróżnić użycie bezokolicznika w funkcji składnika form czasu przyszłego od innych jego użyć, oraz trzy różne funkcje formy na -ł (składnik form czasu przyszłego, czas przeszły z ruchomą końcówką i tryb warunkowy z ruchomą partykułą). Symbol 5 na II pozycji wyróżnia użycie form czasu teraźniejszego w funkcji trybu rozkazującego. Z kolei symbole 6–7 na II pozycji zarezerwowano dla wyróżnienia niektórych funkcji gramatycznych czasowników *być*, *bywać*, *zostać* (człony składowe w czasach złożonych i w formach strony biernej). Na III pozycji występuje cyfra 1 przy czasownikach zwrotnych lub 0 (czyli brak symbolu cyfrowego) przy czasownikach niezwrotnych.

4.2 Homonimia syntaktyczna

Ze względu na brak wypracowanych ścisłych kryteriów pozwalających na odróżnienie poszczególnych części mowy, przyjęto przy kodowaniu pewne sprawdziany o charakterze roboczym. Starano się ustalać kryteria wyłącznie syntaktyczne, abstrahowano całkowicie od kryteriów semantycznych. Szczegółowy opis przyjętych przez nas zasad rozróżniania części mowy został opublikowany osobno (Biuletyn PTJ nr XXIX, 1971), tu ograniczymy się do skrótowego omówienia tychże zasad.

Przyjęte kryteria będziemy podawać, omawiając kolejno parami części mowy,

których funkcje mogą pełnić opisywane homonimy syntaktyczne. Dla jasności

oznaczamy określony rodzaj homonimii syntaktycznej symbolami (SV = subst./verbum = 1/5, SAdv, VA itp.).

4.2.1 Rzeczownik czy czasownik (typ SV = 1/5)

Homonimia ta dotyczyła wyrazów *potrzeba*, *brak*, *szkoda*, *strach*, *żal*, *wstyd*.

Kryterium: kwalifikator czasownika [5], jeżeli w danym zdaniu występuje względnie można wstawić człon (*było*) *nam* = zaimek osobowy w celowniku przy nieobecności innego członu predykatywnego.

S: *brak*[1] *ten odczuwano bardzo boleśnie*;

V: *w dyskusji brak*[5] *nam było ładu*.

4.2.2 Przymiotnik czy czasownik (typ AV = 2/5)

Homonimiami syntaktycznymi tego typu są hasła *winien* albo *winny*.

Hasło *winien*/*winny* kwalifikowano:

1. jako czasownik [5], jeżeli w zdaniu towarzyszy mu bezokolicznik: *weryfikacja*

winna [5] *objąć wszystkich członków*;

2. jako przymiotnik [2], jeżeli pełni funkcję przydawki lub orzecznika:

człowiek winny [2] *zbrodni*; *winni* [2] *są*

rodzice,

nie dzieci; także *jest mi winien* [2] *sto*

złotych.

4.2.3 Przysłówek czy czasownik (typ AdvV=8/5)

Homonimia ta dotyczy wyrazów *wolno* oraz *przeszło*.

Wolno kwalifikujemy:

1. jako przysłówek [8], jeżeli jest możliwa substytucja wyrazów *pomału*, *powoli*: *jechał wolno* [8];

2. jako czasownik [5], jeżeli w danym zdaniu występuje względnie można wstawić człon

(*było*) *nam* = zaimek osobowy w celowniku: *nie wolno*

(*nam palić*).

Przeszło kwalifikujemy jako przysłówek [8], jeżeli jest możliwa substytucja wyrazów *więcej niż*, w pozostałych wypadkach kwalifikujemy jako czasownik [5]: *wydobyto przeszło* [8] *tysiąc ton węgla*.

4.2.4 Rzeczownik czy przymiotnik (typ SA = 1/2)

Homonimia ta dotyczy substantywizowanych przymiotników. Ze względu na różne

stopnie tej substantywizacji przyjęto sformułowane przez S. Jodłowskiego formalne kryterium, które wskazuje przy przymiotnikach substantywizowanych na

ograniczenie form rodzaju do dwóch lub nawet do jednego rodzaju. Np. *chory* ma tylko 2 rodzaje w funkcji rzeczownika [1]. Używając tego kryterium kwalifikowano jako rzeczowniki [1] takie wyrazy jak *przewodniczący, biały, młody* itp. A także słownictwo typu *wytyczna, średnia, dane*, również ułamki (*dziesiąta, setna*). Przykłady kontekstów: *otrzymano nowe wytyczne*[1], *średnia*[1] *krajowa w zakresie spożycia, dane*[1] *liczbowe* itp.

4.2.5 Rzeczownik czy przyimek (typ SPrp = 1/6)

Homonimia ta dotyczy niektórych rzeczowników, które w pewnych wypadkach mogą

pełnić funkcje przyimków. W badanych tekstach były to rzeczowniki *celem, drogą, koło, skutkiem, względem*.

Jako kryteria pozwalające uznać dany człon za przyimek [6] przyjęto:

1. możliwość zastosowania określonych transformacji lub substytucji,
2. relację danego członu z innymi członami ciągu syntaktycznego (kontekst).

Celem + gen. nom. actionis jest przyimkiem [6], jeżeli jest możliwa transformacja

tego wyrażenia w podrzędne zdanie okolicznikowe celu, tzn. gdy zachodzi relacja:

celem + gen. nom. act. \longrightarrow *aby* + infinitivus/verbum finitum.

Przykład: *Rada Bezpieczeństwa zebrała się celem* [6]

przedyskutowania sytuacji politycznej \longrightarrow ...

aby przedyskutować sytuację polityczną. Jeżeli podobna

transformacja nie jest możliwa, *celem* uzyskuje kwalifikator rzeczownika [1].

Drogą + gen. nom. actionis jest przyimkiem [6], jeżeli jest możliwa substytucja

wyrażenia przyimkowego *przez* + acc. nom. act.: *drogą* [6] *pertraktacji*

uzyskano ustępstwa \longrightarrow *przez pertraktacje*...

Natomiast *drogą* jest rzeczownikiem [1], jeżeli:

1. jest członem określającym,
2. nie jest możliwa powyższa substytucja.

Skutkiem + gen. oraz *względem* + gen. kwalifikujemy jako przyimki [6], jeżeli

są możliwe odpowiednie substytucje: dla *skutkiem* substytucja *wskutek*,

dla *względem* substytucja *wobec*:

skutkiem [6] *nieostrożności spowodowano pożar*

\longrightarrow *wskutek nieostrożności*..., *mieć*

względem [6] *kogoś zobowiązania* \longrightarrow

mieć wobec kogoś...

Jeżeli zaś substytucje te nie są możliwe, kwalifikujemy oba te człony jako rzeczowniki [1].

Koło + gen. kwalifikujemy jako przyimek [6], jeżeli jest możliwa substytucja

wyrazu *obok*: *koło* [6] *lasu* \longrightarrow *obok*

lasu. W wypadku dwuznaczności (*koło wozu*) o możliwości lub

niemożliwości tej substytucji rozstrzyga kontekst, a więc relacja z innymi członami

ciągu syntaktycznego: *koło* [6] *wozu pały się*

konie, ale *pękło nam tylne koło* [1] *wozu*.

4.2.6 Zaimki

Zaimki klasyfikujemy wyłącznie na podstawie kryterium syntaktycznego,

wyodrębniając tylko dwie klasy zaimków: zaimki o odmianie przymiotnikowej,

mające homonimiczne formy przypadka i liczby (*których, jakie*) oraz zaimki o odmianie rzeczownikowej, mające homonimiczne formy przypadku (*kogo, czym*). Zaimki przymiotnikowe otrzymują symbol [2], zaimki rzeczownikowe — symbol [4]. Tak więc nie uwzględniamy — z bardzo małymi wyjątkami (por. „Hasła kodowane umownie”, str. 54) klasyfikacji semantycznej zaimków (zaimki względne, pytajne, nieokreślone itp.). Rozróżnienie między zaimkami przymiotnikowymi i rzeczownikami nie jest kłopotliwe, ulega ono jedynie zmianie w odniesieniu do zaimków *to, wszystko, samo, jedno*. Mianowicie:

1. Zaimki *ten, ta, to; wszystek, wszystka, wszystko; sam, sama, samo; jeden, jedna, jedno* są traktowane jako zaimki przymiotne [2], jeżeli występują :
 2. w liczbie mnogiej,
 3. w liczbie pojedynczej w funkcji przydawki.
- Przykłady: *ci ludzie; zaproś tych, których znasz; zastanawia mnie to zagadnienie; wszyscy ludzie; wszystkim się zdawało, że Wojski wciąż gra jeszcze; kot wypił wszystko mleko; w zebraniu uczestniczyli sami cudzoziemcy; dokonali tego zupełnie same, bez niczyjej pomocy; błąd tkwił już w samym założeniu; jedne pary szły na prawo, drugie na lewo; jedni chcą tak, inni inaczej; jedna jaskółka wiosny nie czyni;*