

## ZASADY BUDOWY HASŁA SŁOWNIKOWEGO DLA KOŁOKACJI WERBO-NOMINALNYCH

GRAŻYNA VETULANI

### 0. WSTĘP

Budowa hasła słownikowego podlega ograniczeniom formalnym w zależności od zamierzonego użycia słownika. Jeżeli adresatem informacji przyhasłowej jest człowiek, to wymagania formalne w opisie jednostki wyrazowej są zwykle mniejsze z uwagi na kompetencję językową użytkownika. Jeżeli zaś słownik ma być “widziany” przez np. program komputerowy czy wręcz ma stanowić jego część, wymagania dotyczące formy, jednorodności, precyzji itp. będą bardzo rygorystyczne, a znalezienie właściwego formatu stanie się samodzielnym zagadnieniem badawczym.

Rozwój leksykologii i semantyki leksykalnej wiązał się dotąd przede wszystkim z badaniem słownictwa tradycyjnie pojmowanego jako pojedyncze wyrazy. Taki też stan badań jest odzwierciedlony w większości słowników. W niniejszej pracy przedmiotem analizy nie będą jednak pojedyncze wyrazy, lecz znaki nieciągłe, czyli swoistego rodzaju złożenia<sup>1</sup>, które w gramatyce języka polskiego nazywane są *zestawieniami*. Zestawienia to konstrukcje dwu- lub kilkuwyrazowe, stanowiące wykładniki jednego pojęcia. Ten typ jednostek od pewnego już czasu znajduje się w centrum zainteresowań leksykografów, głównie ze względu na ich przydatność zarówno w dydaktyce jak i w tłumaczeniu na język obcy.

---

<sup>1</sup> W tradycji gramatycznej języka polskiego wyróżnia się *zrosty* (np. *psubrat*, *Białystok*, *Wielkanoc*), *złożenia*, w których obok tematów leksykalnych występuje formant słowotwórczy *-o*, *-i* lub *-y* (np. *grzybobrание*, *łamiłówwka*, *męczydusza*) oraz *z e s t a w i e n i a* (np. *woda sodowa*, *dom spokojnej starości*).

## 1. KOŁOKACJE WERBO-NOMINALNE

Do zestawień można w szczególności zaliczyć *kolokacje werbo-nominalne* (znane z literatury jako *zwroty* czy *orzeczenia peryfrastyczne* lub też jako *analityczne czasowniki funkcjonalne*). O tego typu związkach jako o zestawieniach pisał dość wcześnie Szober w pracy pt. *Na straży języka* (1937:220). Później różne typy zestawień omawiali także inni badacze, uwzględniając na przykład całe frazy (zob. Cockiewicz i Śliwiński 1982:153). Terminu *kolokacja* używa się w literaturze zarówno dla określenia rzadkich związków typu idiomatycznego, jak i połączeń luźniejszych, do których należą m.in. *związki łączliwe* (Skorupka 1967:7), będące przedmiotem opisu w niniejszych rozważaniach. Przyjmijmy, iż w kolokacjach, o których mowa (a więc w połączeniach o strukturze: czasownik + rzeczownik abstrakcyjny) zachodzi "zdolność leksemów do tworzenia mniej lub bardziej stałych związków semantyczno-syntagmatycznych z określonymi leksemami" (Szulc 1994:108). Są to więc związki wyrazowe powstałe na zasadzie wzajemnej determinacji, w których spełnione są nie tylko wymogi formy (np. zgodność końcówek), lecz w których zachodzi semantyczna łączliwość wyrazów, determinowana nie tyle przez kontekst sytuacyjny, ile przez system leksykalny języka. Karolak mówi w tym kontekście o "właściwościach wewnątrzrelacyjnych zdań", w których ważną rolę odgrywają między innymi "semantyczno-relacyjne lub inaczej semantyczno-syntaktyczne właściwości składników struktury sensu". Te ostatnie autor rozumie jako "zdolność do wzajemnego współwystępowania" na płaszczyźnie semantycznej, która niekiedy jest określana jako "zgodność semantyczna" (Karolak 1984:14). Związki te wchodzą do języka na zasadzie konwencji mającej uzasadnienie w historii języka (stąd też nazywa się je również *syntagmami konwencjonalnymi* lub *asocjacjami leksykalnymi*).

Kolokacje werbo-nominalne są formami pośrednimi pomiędzy całkowicie potencjalnymi, złożonymi formami morfologicznymi a konstrukcjami frazeologicznymi. W zdaniu pełnią zawsze funkcję orzeczenia. Należą do nich takie formy, jak *być w błędzie*, *popęlić błąd*, *mieć szacunek (dla kogoś)*, *pałać kazanie (komuś)*, *podjąć / powziąć decyzję (jakaś)*, *prowadzić wojnę (z kimś)*, *przypuścić szturm (na coś)*, *snuć domysły*, *spełniać warunki*, *stawić diagnozę (jakaś, komuś)*, *uciąć sobie drzemkę*, *uprawiać sport*, *ustalić cenę (na coś)*, *wymierzyć policzek (komuś)*, *wysunąć sugestię*, *ulec namowicie*, *zapłonąć miłością (do kogoś)*, *wymierzyć / dać / wlepić mandat (komuś, za coś)* itd.<sup>2</sup>

O kolokacjach werbo-nominalnych pisano przy różnych okazjach. Z początku jako o odpowiednikach form syntetycznych (por. *mieć respekt – respektować*, *wymierzyć policzek – spoliczkować*), występujących głównie w tekstach prasowych i naukowych, później jako o jednostkach specyficznych, odrębnych, choć pod pewnymi względami bliskich klasie zwykłych czasowników, występujących we wszystkich odmianach współczesnego języka polskiego.<sup>3</sup>

<sup>2</sup> Orzeczenia peryfrastyczne były ostatnio przedmiotem rozprawy pt. *Rzeczowniki predykatywne języka polskiego. W kierunku syntaktycznego słownika rzeczowników predykatywnych*. Podaje się tam kilka tysięcy podobnych struktur (G. V e t u l a n i 2000b).

<sup>3</sup> W początkowych latach XX-go wieku pisano o tego typu związkach głównie na łamach *Poradnika Językowego* jako o analitycznych (opisowych) konstrukcjach czasownikowych, które

Związki tego typu uważa się za specyficzne jednostki języka, ponieważ:

- 1) często nie mają one odpowiedników syntetycznych (*ponieść klęskę*, *mieć scysję* z...)
- 2) mają inną walencję niż ich syntetyczne odpowiedniki czasownikowe
- 3) "ich udział w systemie i w tekstach jest większy i bardziej zróżnicowany niż się powszechnie sądzi" (Jędrzejko 1998:21); podobnie Bogusławki (1996:357) mówi, że "większość jednostek to nie pojedyncze wyrazy"; szacuje się, iż w języku naturalnym istnieje 3-4 razy więcej jednostek złożonych (zestawień wszystkich typów) niż prostych a ponadto liczba ich nieustannie wzrasta (G. Gross 1989).
- 4) można ujawnić unifikujące kryteria deskrypcyjne dla klasy rzeczowników, które w te związki wchodzi, jak np. współwystępowanie w strukturze z określonym typem czasownika, którego podstawową funkcją jest funkcja gramatyczna (dostarczanie informacji na temat liczby, rodzaju, czasu i aspektu); rzeczowniki (zawsze w użyciu abstrakcyjnym) są w tych związkach predykatami właściwymi.

## 2. BUDOWA HASŁA SŁOWNIKOWEGO DLA KOŁOKACJI WERBO-NOMINALNYCH

Mimo iż kolokacje werbo-nominalne są ważnymi dla zastosowań, odrębnymi jednostkami języka, nie doczekały się dotychczas systematycznego i kompletnego opracowania.<sup>4</sup> Z brakiem takiego opracowania idzie w parze ich słaba obecność w aktualnych słownikach. W tej sytuacji, do czasu powstania nowoczesnego oraz pełnego słownika języka polskiego, należy postulować tworzenie systematycznych, odrębnych słowników explicite wyliczających typowe związki wyrazowe. Wszelkiego rodzaju zestawienia należy wprowadzać więc do słowników jako jednostki samodzielne, a nie przy okazji objaśnień (hasel) słów prostych. W obecnych warunkach pozyskanie tego typu jednostek (dopuszczalnych analitycznych orzeczeń) jest jednak zadaniem dość trudnym z uwagi na problemy związane z:

- 1) doborem jednostek,
- 2) zjawiskiem zmiany językowej,
- 3) wieloznacznością wyrazów,
- 4) źródłami i sposobem pozyskiwania danych językowych.

zastępują formy proste (głównie lata: 1903, 1905, 1907, 1910, 1929, 1936, później: 1950-53, 1955, 1957, 1963-68, 1978).

<sup>4</sup> W roku 1998 ukazał się *Słownik polskich zwrotów werbo-nominalnych. Zeszyt próbny* (pod red. E. Ję d r z e j k o), który prezentuje rozmaite użycia wybranych pięćdziesięciu rzeczowników predykatywnych języka polskiego.

## 2.1. DOBÓR JEDNOSTEK

### 2.1.1. Kolokacje werbo-nominalne jako predykaty

Z uwagi na to, że jednym z głównych zadań lingwistyki jest systematyczny i wyczerpujący opis jednostek leksykalnych, ważną kwestią jest, jakie jednostki będą wprowadzane do słowników i jak będą one opisywane. Dobrą propozycją wydaje się być opis, który ma charakter leksykonu gramatycznego, zakładającego wprowadzanie do słownika jednostki leksykalnej wraz z informacją o jej użyciu w strukturach zdaniowych.<sup>5</sup> W przypadku jednostek wieloznacznych ilość znaczeń musiałaby być zilustrowana odpowiednią ilością struktur w myśl zasady, że jedno użycie (jedna struktura zdaniowa z wszystkimi walencyjnymi dopełnieniami) to jeden sens. W takim podejściu minimalną jednostką analizy nie jest więc pojedyncza jednostka leksykalna, lecz zdanie proste.

Wygodnym punktem wyjścia w budowie słowników jest pozyskanie (i opis) w pierwszej kolejności wyrażań predykatywnych, gdyż te dają się zdefiniować poprzez selekcionowane przez siebie argumenty w podstawowej strukturze użycia, czyli zdaniu prostym. Inaczej mówiąc, można wyznaczyć sens predykatu poprzez dokładny opis podmiotu i wszystkich jego dopełnień.

Predykatami mogą być różne kategorie gramatyczne, w szczególności funkcję tę mogą pełnić rzeczowniki abstrakcyjne, które wraz czasownikami tworzą orzeczenia peryfrastyczne.<sup>6</sup> Ważną rzeczą jest jednak umiejętność odróżniania orzeczeń peryfrastycznych od innych (też licznych) konstrukcji analitycznych, w które te rzeczowniki wchodzi. Spośród wszystkich związków wybrać należy takie, jak np. *ponieść śmierć* (a nie np. *zadać śmierć*, gdzie mamy do czynienia z użyciem kauzatywnym), *odczuwać wstręt* (a nie np. *wzbudzać w kims wstręt*, które jest również kauzatywne) itp. W przypadku wątpliwości przydatnym testem gramatycznym sprawdzającym użycie predykatywne mógłby być schemat grupy imiennej (otrzymanej po nominalizacji zdania wyjściowego), w której członem głównym byłby dany rzeczownik abstrakcyjny (predykat odnoszący się do czynności), po którym następowałoby odwołanie do podmiotu (wykonującego czynność), por. *Jan poniósł śmierć – śmierć Jana, Jan odczuwał wstręt na widok... – wstręt Jana na widok...*

<sup>5</sup> Dobrym przykładem leksykonu gramatycznego w zakresie czasowników języka polskiego może być np. *Słownik syntaktyczno-generatywny czasowników polskich* (P o l a ń s k i 1980).

<sup>6</sup> Ścisłej rzecz biorąc w tego typu orzeczeniach predykatem właściwym (a więc elementem wyrażającym jakąś własność, stan, cechę podmiotu lub relację pomiędzy podmiotem a dopełnieniami) jest rzeczownik abstrakcyjny. Czasownik pełni w tych zwrotach rolę podpory, wskazując gramatycznie na wykonawcę czynności, a zatem dostarczając informacji na temat liczby, rodzaju, a także czasu i aspektu. Niekiedy czasownik będzie wyszukany stylistycznie, wtedy wskaże wyraźnie na sposób, w jaki dana czynność przebiega lub na rejestr użycia.

## 2.1.2. Wprowadzanie nowych kolokacji werbo-nominalnych do słowników

### 2.1.2.1. Nowe związki wyrazowe wynikiem wariantowości języka

Często przyjmuje się, że związki kolokacyjne mają charakter utrwalony. Bliższa analiza pokazuje jednak, że również one, tak jak pozostałe elementy języka, łatwo ulegają zmianom. Na marginesie można powtórzyć tu za Lewickim (1976:18), że: "związków niewymiennocłonowych po prostu nie ma. Dla żartu, w celu zwiększenia ekspresji czy z innych powodów zawsze istnieje możliwość wymiany." I tak kolokacje werbo-nominalne to związki bardzo produktywne. Bardzo łatwo generują się nowe warianty o odmiennej wartości komunikatywnej. Powstają całe serie (rodziny) wyrażań o różnej wartości ekspresywnej w zależności od sytuacji emocjonalno-kontekstowej i rejestru języka. Budowanie rodziny związków odbywa się poprzez "frakcjonowanie" (Lewicki 1976:64), czyli wymiany na pozycji czasownika. "Wszystkie frazeologizmy należące do tej samej rodziny mają inwariant semantyczny, najczęściej identyczny ze znaczeniem jednego z frazeologizmów i łączy je współwystępowanie w każdym tego samego komponentu nominalnego" (Lewicki 1996:13). Kolokacje werbo-nominalne dopuszczają wymiany komponentu werbalnego, oczywiście tam, gdzie jest to możliwe, ponieważ zmianie postaci związku musi towarzyszyć zachowanie podstawowego znaczenia, por.

*mówić / gadać / opowiadać / pleść głupstwa*  
*pałać głupstwo*

W powyższej serii widać wyraźnie, iż wymiana jest uwarunkowana jedynie ekspresywnością czasowników. Potwierdza się więc obserwacja Pajdzińskiej, że: "wariancja leksykalna to wybór jednego z członów uzualnie utrwalonego szeregu substytutów, których funkcje znaczeniowe i identyfikujące są takie same, a możliwe są tylko różnice w planie ekspresywno-emocjonalnym. [...] Zwykle komponenty fakultatywne są swoistymi aktualizatorami znaczenia prymarnego pozostałych komponentów związku, konkretyzują je..." (Pajdzińska 1988:101-102).

Skoro tworzy się całe serie zwrotów, pożądanym byłoby umieszczanie w słowniku jak największej ich liczby po to, by móc oddać najróżniejsze odcienie znaczeniowe. Ten stan rzeczy nie znajduje jednak odbicia w aktualnych opracowaniach. Na przykład w polu *apatia* podaje się *pogrążyć się w apatii* i *popaść w apatię*<sup>7</sup>, tymczasem powinny były znaleźć się w słowniku również takie związki, jak:

*być w stanie apatii / bywać w stanie apatii*  
*popaść / popadać w apatię*  
*ulec / ulegać apatii*

<sup>7</sup> Por. np. hasło słownikowe predykatu *apatia* ze *Słownika języka polskiego* (S z y m c z a k, red. 1983): *apatia* *ż* *I*, *DCMs.* -iii, *blm* <niezdolność do odczuwania wzruszeń; brak zainteresowań; nieczułość, zobojeźnienie, stan odrętwienia>; *Beznadziejna, głęboka apatia. Stan apatii. Pogrążyć się w apatii. Popaść w apatię. Otrząsnąć się z apatii.* <gr.>

*poddać się / poddawać się apatii*  
*pogrążyć się / pogrążyć się w apatii*

przy założeniu, że możliwe są użycia w aspekcie niedokonanym (w słownikach brakuje informacji na temat ograniczeń w zakresie wyrażania kategorii gramatycznych, w tym aspektu).

Podobnie można by wykazać produktywność związków, które łączą predykat *tempo*, por.

*mieć tempo (jakieś)*  
*brać / wziąć tempo*  
*trzymać tempo*  
*utrzymać / utrzymywać tempo*

*Słownik Języka Polskiego* (dalej: SJP) pod red. Szymczaka (1983) rejestruje w tym polu (i w tym znaczeniu) jedynie dwa związki, mianowicie: *wziąć, trzymać dobre tempo*.

Predykat *miłość* tworzyłby np. dwie rodziny derywacyjne, por.

1. *czuć / poczuć miłość (do kogoś)*  
*odczuwać miłość (do kogoś)*  
*darzyć / obdarzyć miłością (kogoś)*  
*palać / zapalać miłością (do kogoś)*  
*plonąć / zaplonąć miłością (do kogoś)* itd.
2. *uprawiać miłość*

Przytoczone przykłady pokazują, iż na pozycji czasownika pojawiają się w języku polskim zarówno formy neutralne (np. *mieć, być w*), jak i wyszukane (np. *plonąć, palać*). W grę wchodzi wszystkie typy czasowników (intensywa, iteratywa, inchoatywa, konwersy itd.). Jędrzejko pisze (1998:50), że: "istotą peryfrastyki werbo-nominalnej jest to, że funkcję synsemantycznego werbalizatora wtórnie pełni wiele czasowników pełnoznacznych (teoretycznie – każdy)".

Zestawianie związków wyrazowych w rodziny może być bardzo cenne, gdyż w jasny sposób pokazuje bogactwo analitycznych form, wewnętrzne zróżnicowanie i wzajemne związki oraz cechy gramatyczne i pragmatyczno-stylistyczne. Taka prezentacja spełnia więc wymogi przejrzystości, chociaż – głównie ze względu na ilość zajmowanego miejsca – byłaby z pewnością niewygodna do prezentacji w tradycyjnych słownikach drukowanych na papierze. W słownikach elektronicznych natomiast wyliczanie jednostek w postaci nawet bardzo długich list nie stanowi przeszkody, gdyż przy obecnych możliwościach technicznych, można wprowadzać praktycznie nieograniczone ilości danych językowych. Należy podkreślić jednak, że zasadniczym kryterium budowy hasła słownikowego opartego na jednej formie rzeczownika są względy semantyczne, choć należy przyznać, że przy takim rozwiązaniu znalazłyby się obok siebie związki bliskoznaczne, niekiedy trudne do zróżnicowania, jak np. pary: *zrobić adiustację tekstu* i *wykonać adiustację tekstu*, itp.

Proponowany układ hasel ma jednak tę zaletę, iż obrazuje szeroko rozmaite sposoby wyrażania się. Zwraca uwagę nie tylko na wyrażenia dobrze utrwalone w języku (znane z literatury), ale również na zwroty luźniejsze, które dopiero zaczynają stawać się wykładnikami jednego pojęcia.

Reasumując, zestawianie predykatów w listy podstawowych struktur, to dobra ilustracja możliwości dopuszczania pewnych elementów na określone pozycje. Wykazuje się w ten sposób wariantywność języka w zakresie analizy opartych na jednym predykanie rzeczownikowym.<sup>8</sup>

Pozyskanie kolokacji werbo-nominalnych może być bardzo ważne nie tylko dla badań leksykograficznych, gdzie chodzi o opis czy klasyfikację jednostek językowych, ale również dla innych dziedzin aplikacyjnych, takich jak na przykład dydaktyka języka polskiego (głównie jako języka obcego, ale też ojczystego). Trudność w opanowaniu danej kolokacji w języku obcym wynika z jej nieprzewidywalności i z faktu, że pomiędzy poszczególnymi językami występują znaczne różnice w doborze elementów składowych (choć zdarzają się też izoleksykalizmy). Umiejętność rozpoznawania i stosowania kolokacji w języku obcym stanowi nieodzowny warunek kompetencji językowej. Stąd też już na bardzo wczesnym etapie nauczania należy uświadamiać uczącym się, że relacje istniejące pomiędzy jednostkami leksykalnymi nie są przypadkowe, i że związki leksykalne mają ściśle określone właściwości strukturalno-semantyczne.

Słowniki kolokacyjne mogą być ponadto z powodzeniem wykorzystywane w tłumaczeniu na język obcy. Szczególnie w tłumaczeniu wspomaganym maszynowo, gdzie stosuje się narzędzia informatyczne do uproszczenia i przyspieszenia przekładu. W tym procesie człowiek pozostaje ostatecznym ekspertem decydującym o wyborze właściwego dla danego kontekstu związku wyrazowego, system wspomagający podpowiada mu tylko, nadające się do wykorzystania, krótkie listy kolokacji, spośród których wybiera on taką, która najlepiej harmonizuje się z całością tekstu. Zauważmy, że użytkownik słownika wybierałby już tylko spośród struktur poprawnych, a to z kolei mogłoby w znacznym stopniu wpłynąć na jakość tłumaczenia. Tym bardziej wydaje się więc istotne dostarczanie możliwie kompletnej listy dopuszczalnych leksykalnie związków wyrazowych (celem ułatwienia możliwie najbardziej trafnego wyboru).

### 2.1.2.2. Zjawiska związane z mechanizmem zmiany językowej

Patrząc na język diachronicznie, dają się zauważyć liczne przesunięcia w leksyce, również w typowych kontekstach. Znaczący to, że pewne połączenia są wypierane przez inne. Buttler, analizując przemiany w polszczyźnie na przełomie XIX-go i XX-go wieku, pokazuje swoistą degradację czasowników oraz dekompletowanie się zespołu rzeczowników, z którymi mogą one tworzyć konstrukcje peryfrastyczne. Podaje na

<sup>8</sup> Rzecz jasna nie chodzi o przytaczanie wszystkich możliwych konstrukcji, a więc również tych, które wydają się nieakceptowalne, lecz które w specyficznych sytuacjach są uzasadnione (np. w użyciach żartobliwych, w sytuacjach wyimaginowanych). Wymagałoby to za każdym razem indywidualnego uzasadnienia tego użycia w opisie hasła. Teoretycznie rzecz biorąc, związków wyrazowych może być bardzo dużo i nie ma potrzeby odnotowywać wszystkich w słowniku.

przykład, że “czasownik *nabyć / nabywać*, kiedyś łączący się z rzeczownikami różnych kategorii znaczeniowych [...], obecnie konotuje tylko nazwy rezultatów” (1988:78).

### 2.1.2.3. Problem akceptowalności nowych związków wyrazowych

Duża produktywność kolokacji werbo-nominalnych (liczne wymiany na pozycji czasownika w zależności od rejestru języka i warunków emocjonalno-kontekstowych a także zmiany diachroniczne) powoduje, że trudno jest czasami podjąć decyzję co do akceptowalności niektórych z nich, a zatem co do możliwości wprowadzania ich do opracowań leksykograficznych. Chodziłoby na przykład o rozstrzygnięcie kwestii, czy związki luźniejsze (np. pochodzące z rejestru języka potocznego) można uznać za poprawne, por.

? *zrobić sobie drzemkę*  
 ? *zrobić (komuś) krzywdę*  
 ? *zrobić (z czegoś) użytek*  
 ? *być w transie*

SJP podaje odpowiednio: *uciąć sobie drzemkę, wyrządzić krzywdę, czynić użytek, wpaść w trans*. Ważna byłaby tu odpowiedź na pytanie, jak traktować związki nie utrwalone, których słowniki jeszcze nie rejestrują, gdyż nie znalazły się one wcześniej w zasobie środków językowych będących podstawą do opracowania leksykograficznego. Różnego rodzaju słowniki, gramatyki a także poradniki poprawnościowe prezentują jedynie użycia znormalizowane, tzw. skodyfikowaną normę językową, tymczasem można tu powtórzyć za Walczakiem, iż każdy z nas “produkuje” innowacje i że “innowacją może być nowy wyraz lub połączenie wyrazów”. I choć, jak autor mówi, większość tych innowacji pozostaje zjawiskiem indywidualnym i efemerycznym, to te, które mają coraz większy zasięg z powodu powszechnego oddziaływania środków masowego przekazu, są szybko podchwytywane i w związku z tym zaczynają się upowszechniać (Walczak 1999: 13-14). Językoznawcy polscy (zob. Buttler 1976, Lewicki 1976, 1977, 1982) wielokrotnie podnosili problem dopuszczalnych wymian. Okazuje się, że zmiany się dokonują, często wbrew normie poprawnościowej. Nowe związki wchodzą stopniowo do regularnego użycia. Przy leksykalizacji zwrotu liczy się czas i ilość powtórzeń. Lewicki (1982) mówi w tym kontekście o tzw. “współczynniku powtarzalności”. Dzisiaj wiele przykładów takich zwrotów mniej utrwalonych znajdujemy w tekstach, które w dalszej kolejności uznaje się za podstawę normalizacji. Zresztą związki słabo utrwalone, a więc rzadkie w uzusie, bywają z tego właśnie tytułu, jako specyficznie nacechowane, stosowane przez autorów. Trudno je zatem pomijać, np. w tłumaczeniu.

## 2.2. WIELOZNACZNOŚĆ PREDYKATÓW

Wymagania składniowe jednostek są pochodną ich właściwości semantycznych. Dobrą reprezentacją pojedynczego znaczenia jednostki językowej może być model podstawowej struktury, w której jest ona użyta (czyli zdanie proste, obrazujące w jaki sposób dana

jednostka wytwarza specyficzny system łączliwości z pozostałymi elementami). W modelu odzwierciedlającym budowę zdania należy uwzględnić wszystkie elementy konieczne, tzn. ustalić obowiązkowe argumenty dla danego użycia, nawet jeśli możliwe jest opuszczenie niektórych z nich w konkretnych realizacjach językowych (konieczne nie znaczy nieopuszczalne)<sup>9</sup>.

Całkowite modele struktur stają się przydatne w wyszukiwaniu elementów, które mogą zdać sprawę z różnych użyć, a więc różnych znaczeń predykatów. W przypadku rzeczowników predykatywnych elementami walencyjnymi są: predykatywnie pusty czasownik<sup>10</sup>, tworzący wraz z predykatem właściwym orzeczenie peryfrastyczne oraz argumenty (czyli podmiot i znaczące dopełnienia). Porównajmy modele struktur ilustrujące poszczególne znaczenia predykatów *chód* i *komentarz*:

<i>chód</i> : <i>mieć chód (jakiś)</i>	(np. <i>chwiejny, szybki, żołnierski...</i> )
<i>mieć chody (lmi)</i>	(‘szybko chodzić’: np. <i>Ależ on ma chody!</i> )
<i>mieć chody (lmi) + argument</i>	(np. <i>mieć chody w ministerstwie</i> )
<i>być na chodzie</i>	(‘działać’: np. <i>Winda jest na chodzie</i> )
<i>komentarz</i> : <i>dać / dawać / robić / zrobić komentarz do czegoś</i>	(np. <i>filmu</i> )
<i>robić komentarze (jakiś)</i>	(np. <i>złośliwe</i> )

W powyższym zestawieniu użycia są rozłączne, a zatem nie stanowią problemu interpretacyjnego: dla predykatu *komentarz* elementem różnicującym znaczenia jest między innymi liczbą. Co prawda można użyć predykat *komentarz* z argumentem wprowadzonym przez przyimek *do* również w liczbie mnogiej, por. *Jan robi / daje komentarze do filmu*. Jednym z możliwych odczytań byłoby wtedy dodatkowo odczytanie, że ‘jest to stałe zajęcie, praca’. Przy takiej interpretacji jednak predykat *komentarz* narzuci czasownikowi aspekt niedokonany, a więc *dawać* a nie *dać*, *robić* a nie *zrobić*. Najlepszą prezentacją znaczenia jest więc całkowity model struktury z informacją gramatyczną odnośnie ograniczeń na poszczególnych pozycjach. Zauważmy na przykład, że grupa imienna *komentarze Jana* mogłaby być wyprowadzona z dwóch różnych zdań, por. *Jan robi komentarze do filmu* oraz *Jan stale robi złośliwe komentarze*. Dopiero dodatkowe argumenty podane explicite mogą tę wieloznaczność rozwiązać, por. *komentarze Jana do filmu* i *złośliwe komentarze Jana*.

Tak więc pożądanym jest umieszczenie w słowniku całego modelu struktury, a nie, jak to się często dzieje obecnie, prezentowanie znaczeń za pomocą zredukowanych grup imiennych. Gdyby chcieć opisać różne znaczenia np. predykatu *komenda*, okazałoby się, że jedynie odwołanie się do czasownika i wszystkich dopełnień pozwala wskazać, iż w jednym przypadku chodzi o ‘wydanie rozkazu’, a w drugim o ‘objęcie dowództwa’, por. dwa zdania:

<sup>9</sup> Np. predykat czasownikowy *czytać* może wystąpić w konkretnej realizacji językowej tylko z jednym dopełnieniem, por. *Janek czyta Marysi* lub *Janek czyta bajkę* albo tylko z samym podmiotem, por. *Janek czyta*, choć wszystkie te elementy łącznie go definiują.

<sup>10</sup> Czasownik ten nazywany jest też *czasownikiem funkcyjnym, podporowym, synsematycznym* lub *werbalizatorem*.

*Dowódca wydał komendę do ataku,  
Dowódca objął komendę nad wojskiem*

Zatem już w treści haseł słownikowych powinno umieszczać się informacje na temat zakresu łączliwości wyrazów. Tylko takie podejście może być bowiem w pełni przydatne w praktycznych zastosowaniach. W przypadku predykatów rzeczownikowych istotne jest podanie przede wszystkim informacji o współwystępujących z nimi czasownikach, gdyż te często nie są przewidywalne. Polański pisze, iż związki wyrazowe "zależą od zasad łączliwości wyrazowej, które często mają charakter konwencjonalny. Żadne racjonalne względy na przykład nie uzasadniają faktu, że rzeczownik *zwycięstwo* łączy się z czasownikiem *odnieść*, natomiast *kłeska* z *ponieść*, *spotkanie* z *odbyć*, zaś *badania*, *doświadczenia* z *przeprowadzić* itp." (Polański 1980:5). Ponadto właśnie na tej pozycji zachodzą liczne wymiany formalne, zaś poszczególne warianty czasownikowe są czynnikiem uściślającym znaczenie predykatu oraz wskazują na rejestr użycia (neutralny, potoczny, uroczysty itd.). Obecnie czasowniki te są systematycznie pomijane, zarówno w słownikach polskich, francuskich jak i niemieckich (zob. na ten temat G. Gross 1989:178). Zresztą, nawet całkowita przewidywalność czasownika nie umniejsza przydatności podania go explicite. Szczególnie jest to ważne dla zastosowań komputerowych, gdyż zwalnia z konieczności programowania skomplikowanych często reguł.

Tym samym łatwo można zgodzić się z Salonim i Świdzińskim, którzy w swoim komentarzu do uaktualnionego wydania *Składni Współczesnego Języka Polskiego* (1998) piszą, że chodzi o "dostarczenie czytelnikowi kompendium gramatycznego na miarę ery rewolucji komputerowej". Potrzeby coraz szerszych badaczy wykorzystujących techniki automatycznego przetwarzania tekstów "wymagają opisu lingwistycznego różnego od tych, jakie wdraża się w edukacji szkolnej".

Reasumując, hasłem słownikowym nie powinna być wyizolowana jednostka leksykalna, lecz zestaw informacji gramatycznych pozwalający zdać sprawę z jej funkcjonowania.

### 2.3. ŹRÓDŁA I SPOSOBY POZYSKIWANIA JEDNOSTEK JĘZYKOWYCH

#### 2.3.1. Kolokacje werbo-nominalne w słownikach języka polskiego

W aktualnych słownikach języka polskiego informacje na temat kolokacji werbo-nominalnych są zazwyczaj niekompletne, a jeżeli się pojawiają, to przeważnie umieszczone są przy czasowniku, od którego tradycyjnie rozpoczyna się analizę składniową zdań. Do rzadkości należy podawanie orzeczeń peryfrastycznych przy hasle rzeczownika abstrakcyjnego, który jest głównym elementem tych orzeczeń. Znaczenie hasła objaśniane jest zazwyczaj opisowo, a informacje przyhasłowe dotyczą najczęściej rozróżnienia semantycznego na poziomie podstawowym, tzn. rozróżnienia pomiędzy użyciem konkretnym i przenośnym. Rzadko kontrastuje się różnorodne znaczenia wyrazu nacechowanego za każdym razem tak samo (np. [+Abstr]), a mimo to niosącego inny sens w zależności od konstrukcji gramatycznej. Jednostki językowe wchodzą w struktury większe, na ogół brakuje jednak prezentacji użyc. Nie chodzi bowiem o ilustrację poprzez

przytaczanie przykładów literackich (te w niektórych słownikach są podawane), lecz o prezentację, która mogłaby być przydatna w dalszych zastosowaniach, wymagających dużej precyzji. Przykłady konkretnych realizacji językowych pomijają często wiele istotnych elementów ze względu na fakt, że zostały użyte w konkretnych warunkach kontekstowo-sytuacyjnych.

W szczególności napotkane problemy można pogrupować, jak następuje:

- 1) Podaje się tylko jedną z możliwych struktur zdaniowych. Trudno jest zorientować się wtedy co do motywacji wyboru (brak struktury w ogóle czy tylko pominięcie jej w słowniku). Jeśli nawet są to wybory spowodowane np. większą częstością występowania w języku, to nie ma przy nich informacji co do ograniczeń w użyciu, np. *spór toczy się o...*, *? ktoś toczy spór o...*, *nie mieć umiaru* (tylko w negacji), *? ktoś ma umiar, nie mieć trosk* (w negacji i w liczbie mnogiej), *? mieć troski / troskę* itd.
- 2) Brakuje rozwiniętych konstrukcji imiennych i czasowników wchodzących w związki z rzeczownikami abstrakcyjnymi (w informacji przyhasłowej rzeczownika). Jeśli taka informacja się pojawia, to najczęściej w związkach dwuwyrazowych, zwykle w postaci grupy imiennej zbudowanej z rzeczownika abstrakcyjnego z jednym elementem – nie koniecznie walencyjnym. Podaje się np. *troska o rodzinę/o pacjenta/o byt*, lecz taka prezentacja jest niewystarczająca, by móc odtworzyć strukturę, z której te grupy pochodzą, por. *? wykazywać / odczuwać troskę*.
- 3) Najczęściej ogranicza się prezentację do konstrukcji dwuwyrazowych, z jedną z możliwych fraz nominalnych, z jakimś możliwym argumentem, przy czym czytelnikowi trudno jest wywnioskować, czy jest to jedyny dopuszczalny związek, czy też możliwe są związki z innymi elementami (inne, podane przy hasle czasowniki, stosują się bowiem do innych użyć).
- 4) Brak jest użyć typowych, o dużej frekwencji użycia w języku polskim, jak np. *mieć romans z...*, *być w stadium...*, *zachowywać spokój*, *mieć sposób na...*, *wziąć rozbieg*, *dokonać rozbioru...*, *być w rezerwie*, *zawrzeć układ o...* (jest tylko: *podpisać układ o...*), *mieć temperament*, *mieć tendencję do...*, *podjąć temat*, *być w transie* (jest tylko: *wpaść w trans*), *mieć świadomość, że...*, *obrać strategię*, *być w szoku*, *żyć w symbiozie z...*, *mieć stosunek do...* (*mieć lekceważący / poważny... stosunek do życia / kolegów...*) itd.
- 5) Czasem podaje się tylko jeden czasownik, gdy tymczasem predykat dopuszcza ich więcej (*odczuwać satysfakcję z...*, brak jest: *mieć satysfakcję z...*). Jest rzeczą oczywistą, iż wszystkich dopuszczalnych związków wypisać nie sposób, należałoby jednak tak budować hasła, by czytelnik mógł się zorientować w możliwościach składniowych predykatu.
- 6) W wielu przypadkach brakuje również argumentów przy związkach wyrazowych, por. *podnieść raban / narobić rabanu* (tu akurat łatwo można odtworzyć jeden z nich, choćby z często występującego zdania: *O co tyle rabanu?*)
- 7) Prawie systematycznie występują braki przy predykatach, które tworzą konwersy, por. *nadać / otrzymać przywilej*, *wywierać wpływ na kogoś (na coś) / ktoś (coś)*

ulega wpływowi, przeprowadzać obróbkę metali/metale ulegają obróbce. Zwykle podaje się tylko jeden związek, np. dla predykatu *rekojmia* znajdujemy *dać: producent daje rekojmie*, brak jest *wyrób podlega rekojmii*, bez trudu znajdujemy *dać rekompensatę*, brakuje *otrzymać rekompensatę*, itp. Taki stan rzeczy powoduje niemożność szybkiej i dokładnej analizy konstrukcji zdaniowych, z których pochodzą grupy imienne, często wieloznaczne, por. *rehabilitacja pacjenta (Lekarz poddaje pacjenta rehabilitacji i Pacjent poddaje się rehabilitacji)*, *obróbka metali (Hutnicy poddają metale obróbce i Metale ulegają / poddają się obróbce)*, *wybór Jana (Jan dokonał wyboru i Ktoś wybrał Jana)*.

- 8) Występują systematyczne braki w danych dotyczących możliwości wyrażania pewnych kategorii gramatycznych (np. liczby przy predykcji i aspektu przy czasowniku). Występując w związkach, wyrazy mogą nie mieć tych samych możliwości wyrażania kategorii gramatycznych co w użyciu samodzielnym, por. *toczyć dyskusję / rozmowę / sprzeczkę* (i brak możliwości użycia w aspekcie dokonanym, por. *\*stoczyć dyskusję...*, choć sam czasownik tworzy parę aspektową *toczyć / stoczyć*). Ogólnie rzecz biorąc chodziłoby o informację co do możliwości występowania form dokonanych i niedokonanych, np. *przewodzić / przeprowadzić badania, podejmować / podjąć badania, wpadać / wpaść w szal* i np. tylko: *toczyć dyskusję / rozmowę / sprzeczkę* lub *narobić świństw / głupstw*.
- 9) Ogólnie można powiedzieć, że brakuje konsekwencji w konstrukcji hasel. Podaje się np. *być w rozkładzie*, a nie ma np. informacji przy predykcji *rozpad*. Podobnie podaje się *dokonać transakcji*, a nie podaje się związku przy predykcji *transfer*, który opisany jest jako 'transakcja'. Podaje się np. *popelnić przestępstwo / dopuścić się przestępstwa*, a nie podaje się struktur przy predykatkach *ojcobójstwo* czy *łapownictwo* itd.
- 10) Słowniki rejestrują zwroty dobrze utrwalone, nie uwzględniają natomiast konstrukcji nowszych, oryginalnych – też poprawnych – będących dopiero w drodze utrwalania się lub pochodzących z rejestru języka mówionego (zawsze poprawnego, choć mniej wyszukanego niż znormalizowany język literacki). Można zatem postulować wprowadzanie do słowników "konkatenacji syntaktycznie poprawnych" (Bogusławski 1979:19).<sup>11</sup>

Wyżej wymienione braki występują nagminnie w słownikach polszczyzny. Przynajmniej część z nich można wyjaśnić tym, że słowniki języka polskiego w zasadzie nie były adresowane do użytkownika obcojęzycznego, jak to ma miejsce w przypadku języków

międzynarodowych, lecz do rodzimego użytkownika języka. Nie były też tworzone z myślą o wykorzystaniu informatycznym. Informacje, które autorom wydawały się oczywiste, były więc często z tego tytułu pomijane.

### 2.3.2. Postulaty w kierunku skutecznego pozyskiwania danych językowych

Dobór (ilość i jakość) materiału do badań jest zwykle podyktowany możliwościami technicznymi, jakie istnieją w momencie ich podejmowania. W ostatnich latach w wyniku powszechnej komputeryzacji obserwujemy niezwykle rozwój tych możliwości, a to wiąże się z kolei z nowymi metodami i narzędziami analizy lingwistycznej.

#### 2.3.2.1. Korpusy tekstowe jako przedmiot badań i jako narzędzie

Fakt, że obecnie niemal każdy tekst w jakiejś fazie powstawania znajduje się na nośniku elektronicznym, umożliwia gromadzenie materiału obserwowanego w postaci korpusów z możliwością łatwego i wielokrotnego dostępu do danych. Pojęcie korpusu, mimo że intuicyjnie zrozumiałe, nie jest ostre. W szerokim znaczeniu może odnosić się do zbioru wszelkich danych językowych będących tekstem, w nieco węższym do zbioru tekstów ciągłych lub wręcz zdań, wreszcie zakłada specjalny sposób selekcji i organizacji danych. Dalej należy rozumieć, że: "korpus jest kolekcją danych językowych wybranych i zorganizowanych według jawnych kryteriów lingwistycznych w celu ustanowienia próby języka" (Sinclair 1996:4). Definicja ta kładzie nacisk na stosowanie kryteriów lingwistycznych w selekcji danych oraz na chęć znalezienia empirycznej bazy do badań nad językiem, a więc chęć zebrania materiału reprezentatywnego dla badanych zjawisk. Przy tym reprezentatywność winna oznaczać pełność (tj. zapewnienie, że wszystkie istotne zjawiska są reprezentowane), a także to że korpus danych odzwierciedla relacje ilościowe pomiędzy zjawiskami (tj. że zjawiska rzadko występujące w języku (uzusie) występują też rzadko w korpusie). W praktyce najczęściej stosowaną techniką zmierzającą do zapewnienia reprezentatywności jest odpowiednio duży rozmiar korpusu lub zestawianie korpusów średniej wielkości (lub wręcz małych), lecz odpowiednio dobranych.

Zwykły zbiór tekstów nie stanowi jeszcze korpusu językowego nadającego się do analizy komputerowej. Gromadzenie tekstów w postaci elektronicznej bazy danych językowych umożliwia wzbogacanie go dodatkową informacją przy wyrazach. Aktualnie istniejące korpusy języków angielskiego, francuskiego, niemieckiego, włoskiego i in. to więc nie same tylko ciągi słów, lecz ciągi słów opatrzone informacjami morfosyntaktycznymi, semantycznymi, prozodycznymi, krytycznymi itd., w postaci różnego typu adnotacji. Warto odnotować, że dla utworzonego już korpusu jakość i ilość tych adnotacji może wzrastać niezależnie od tego, czy zwiększa się sam korpus. Jednocześnie wraz z rozwojem technologii i co za tym idzie z powiększaniem się rozmiarów korpusów zmienia się ocena ich wielkości.<sup>12</sup>

<sup>11</sup> Na przykład w SJP, przy wyrazie *rozsyпка* znajdujemy informację o tym, że występuje on *j e d y n i e* w zwrotach, por. *rozsyпка z III. CMS. ~pce; Im D. ~pek, przestarz.* «rozproszenie, rozbiegnięcie się; pośpieszna, beładna ucieczka» dziś tylko w zwrotach: *Biec, uciekać* itp. w *rozsyпce* «biec, uciekać oddzielnie, nie w skupieniu, w rozproszeniu» Δ *Pójsz, isz* w *rozsyпkę* «o grupie osób, najczęściej o wojsku: przestać tworzyć zwarty szyk, zwartą grupę, rozbiec się; ulec dezorganizacji, rozbiću, rozproszeniu». Tymczasem wydaje się, iż w pewnych okolicznościach z łatwością można zaakceptować także związek, np. *być w rozsyпce*, por. *Wszystkie twoje rzeczy są w rozsyпce*.

<sup>12</sup> Jeszcze 10 lat temu korpus *Brown (Brown University, USA)*, ze swoim milionem słów, uchodził za duży. Obecnie wiele ośrodków badawczych dysponuje danymi tekstowymi liczącymi miliony, a nawet miliardy słów. *British National Corpus (BNC)* obejmuje na przykład 10 milionów słów

### 2.3.2.2. Słowniki elektroniczne a korpusy

Aczkolwiek nie ma związku koniecznego pomiędzy słownikami a korpusami, to coraz częściej okazuje się, że te dwie kategorie zasobów językowych są ze sobą powiązane. Po pierwsze, duże korpusy są coraz częściej wykorzystywane jako podstawa do zestawiania słowników. Dzięki temu słowniki niejako "dziedziczą" z korpusów reprezentatywność (albo jej brak). Argument reprezentatywności jest szczególnie ważny dla słowników średniej wielkości, gdzie autor jest zmuszony do dokonania selekcji haseł. Z kolei słowniki małe, podręczne oraz średnie są szczególnie ważne do celów praktycznych, a więc i poszukiwane na rynku. Ta uwaga dotyczy przede wszystkim tradycyjnych wydań papierowych, dla których fizyczna objętość jest czynnikiem istotnym. Niezależnie od selekcji haseł słownikowych korpusy w miarę ich reprezentatywności mogą stanowić jedno z głównych źródeł informacji zawartej w treści haseł i dostarczać materiał ilustracyjny (przykłady o udokumentowanym pochodzeniu). Po drugie, słowniki – a dokładniej oparte na nich narzędzia informatyczne – mają zasadnicze znaczenie dla poprawiania jakości korpusów. Dotyczy to zwłaszcza korpusów średniej wielkości, dużych lub bardzo dużych, dla których ręczne indeksowanie jest praktycznie niewykonalne. Indeksowanie (tagowanie) polega na jawnym powiązaniu z elementami tekstu informacji o tych elementach oraz ułatwia efektywne wykorzystanie korpusu w programach komputerowych.

Słowniki elektroniczne są przeznaczone prawie wyłącznie do użytku informatycznego i mają ściśle określoną formę, łatwo dzięki temu interpretowaną przez programy komputerowe. Można je uważać za uporządkowane zbiory danych, gdzie jednostką informacyjną jest hasło słownikowe, a identyfikatorem (kluczem) do hasła jest słowo (lub jakaś jego forma). Przede wszystkim konieczny jest w nich taki system opisu, który w sposób jednoznaczny przypisywałby danemu wyrazowi hasłowemu ściśle określone cechy, jak np. jednoznacznie określony paradygmat odmiany w przypadku słownika morfologicznego. Niezależnie od tego, że od słowników elektronicznych wymagane jest bardzo precyzyjne przestrzeganie formatów, precyzja i szczegółowość (pomijane często w słownikach tradycyjnych adresowanych w zasadzie tylko do człowieka), atrybuty słownika tradycyjnego, takie jak czytelność oznaczeń symbolicznych, są również cechami pożądanymi. Cechy te ułatwiają z jednej strony bieżącą konserwację słowników przez lingwistów i informatyków oraz pisanie programów komputerowych wykorzystujących słowniki przez informatyków współpracujących z lingwistami nie-informatykami – z drugiej strony. Słowniki elektroniczne mogą różnić się znacznie pod względem struktury formalnej. Mogą np. być funkcjonalnie równoważne tablicom ze sformalizowanymi danymi przypisanymi formie hasłowej. Formalizacja danych umożliwia pisanie programów komputerowych, które zapewniają dostęp do informacji. Programy te winne być, w miarę możliwości, ogólne, tj. winny obsługiwać bez wyjątku wszystkie przypadki dla rozważanej

opatrzonych adnotacjami (zob. na ten temat: Habert & Nazarenko & Salem 1997). Tendencja do powiększania korpusów jest zrozumiała, gdyż natura zjawisk językowych wymaga często danych szczególnie bogatych. Okazało się na przykład, że korpus liczący milion słów był niewystarczający, by właściwie zilustrować elastyczność wyrażań frazeologicznych, a korpus dwudziestomilionowy nie pozwolił na odnalezienie wystarczającej liczby wystąpień wszystkich wyrażań idiomatycznych.

klasy. Stwarza to wymóg dokładności nie stosowany w słownikach tradycyjnych. Na przykład, aby otrzymać wszystkie odmienione formy dowolnej jednostki fleksyjnej znajdującej się w słowniku w swojej formie podstawowej, trzeba dysponować algorytmem odmiany dla wszystkich klas bez wyjątku. Ten stan rzeczy powoduje, że w słownikach elektronicznych umieszcza się informacje kompletne i szczegółowe, które w zwykłych słownikach są pomijane, ponieważ są ewidentne dla człowieka. Rygorystyczne przestrzeganie zasady bezwyjątkowości powoduje, że jednostki grupują się w zbiory rozłączne, a to z kolei powoduje, że klasyfikacje jednostek językowych znacznie się różnią od tych, do których jesteśmy przyzwyczajeni w tradycyjnych opracowaniach.<sup>13</sup>

Słowniki elektroniczne konstruowane są w różnych celach. Wspólne dla nich jest to, że konsultowane są przez programy komputerowe albo wręcz stanowią część innych programów (zob. też na ten temat M. Gross 1988). Typowym zastosowaniem jest wykorzystanie ich przez procedury rozpoznawania jednostek językowych. W programach tych, w fazie początkowej, tekst zostaje poddany procedurze segmentacji na jednostki graficzne (wyrazy), po czym konsultowany jest słownik w celu określenia cech każdego wyrazu występującego w tekście. Słownik elektroniczny przeznaczony do analizy tekstów będzie często zawierał zakodowane informacje gramatyczne przeznaczone do użycia przez analizator syntaktyczny. Aktualnie występuje tendencja do umieszczania w słowniku możliwie kompletnej informacji o związkach składniowych, w które wchodzi dany wyraz. Tym samym słowniki przybierają formę leksykonów gramatycznych.

### 2.3.2.3. Cele stosowania narzędzi informatycznych

Programy komputerowe pracujące na dużych korpusach tekstowych są dzisiaj z powodzeniem stosowane w analizie języka. Stosowanie ich nie jest jednak celem samym w sobie. Odpowiadając na pytanie, czemu przetwarzanie komputerowe ma służyć, można przede wszystkim wymienić:

- 1) szybki dostęp do informacji językowej zawartej w autentycznych kontekstach, a nie tylko w słownikach normatywnych,
- 2) szybkie i dokładne przetworzenie bardzo dużych korpusów tekstowych, a więc jak najobszerniejszej reprezentacji języka, w celu:
  - a) potwierdzenia danych słownikowych,
  - b) znalezienia nowych (innych) użyc w stosunku do przykładów słownikowych,
- 3) możliwość pozyskania w sposób reprezentatywny i szybki konkretnych danych językowych, jak np. złożeń, związków peryfrastycznych, kolokacji leksykalnych itp.

<sup>13</sup> Na przykład klasyfikacja rzeczownika polskiego opisana w formacie POLEX uwzględniła 202 klasy paradygmatyczne, wśród których 55 to klasy jednoelementowe, tradycyjnie odnotowywane jako wyjątki (Vetulani, Z., Walczak, B., Obrębski, T., Vetulani, G. 1998)



### 2.3.2.4. Postulat utworzenia narodowego korpusu języka polskiego

Język naturalny jest tworem niezwykle złożonym, w którym następują nieustanne zmiany i przesunięcia. Powoduje to zasadnicze trudności w uzyskaniu adekwatnego opisu. W pewnej mierze trudności te mógłby złagodzić dostęp do bardzo dużego i stale uaktualnianego korpusu tekstowego (ciągle nieistniejącego dla języka polskiego). W tym wypadku chodzi o korpus reprezentatywny, czyli o utworzenie empirycznej bazy do badań nad językiem, która byłaby udostępniana do celów naukowych, a zatem o:

- 1) otwartą i stale uzupełnianą listę dzieł będących reprezentatywną próbą polszczyzny pisanej
- 2) dzieła korpusu zapisane w formie elektronicznej (tak, aby możliwe było wyszukiwanie dowolnych form, kontekstów itd.)
- 3) teksty korpusu jednoznacznie zindeksowane (tak, aby konkretny cytat można było oznaczyć indeksem umożliwiającym natychmiastowe automatyczne wyszukiwanie).

Uzyskanie takiego korpusu jest z różnych powodów rzeczą trudną, przy tym koszta nie są jedyną przeszkodą. Ważnym czynnikiem jest problem praw autorskich do tekstu i niechęć udostępniania postaci elektronicznej. W świetle doświadczeń związanych z realizacją projektów, w których brałam udział (COPERNICUS 621, 1032) wydaje się, że problem wymaga, poza środkami finansowymi, uregulowania ustawowego.

### 2.3.3. Korzyści płynące z wykorzystywania narzędzi informatycznych w analizie językoznawczej

Dysponując zasobami w postaci dużych baz danych językowych i przy użyciu właściwych programów informatycznych (wydajnych systemów wyszukiwania), można – dzięki zgromadzonej w nich informacji – skutecznie badać materiał językowy pod kątem różnorodnych problemów. Podajmy tytułem przykładu, bez ambicji wyczerpania listy, że istnienie postulowanych tu zasobów w znacznym stopniu ułatwia:

- 1) rozwijanie gramatyk opisowych oraz słowników (słów prostych, złożonych, kolokacji, określonych typów struktur zdaniowych, itd.),
- 2) sprawdzanie hipotez stawianych intuicyjnie lub po analizie wąskiego materiału,
- 3) konfrontowanie postulowanych modeli języka z rzeczywistymi faktami językowymi,
- 4) wykorzystywanie uzyskanych danych do tworzenia opisu ogólnego (modelu) języka,
- 5) wykrywanie powiązania między zjawiskami językowymi,
- 6) ustalanie norm,
- 7) prowadzenie typologii tekstów i dokumentów,
- 8) wyszukiwanie informacji (np. na temat frekwencji wyrazów, kolokacji itd.),
- 9) przetwarzanie dokumentów,

- 10) indeksowanie jednostek językowych,
- 11) prowadzenie statystyki faktów językowych dotyczącej na przykład:
  - a) klasyfikacji jednostek językowych,
  - b) występowania w danym stylu (w jakiejś epoce, u danego autora...) różnorodnych (tych samych) wyrażań.

### Bibliografia

- Bogusławski, A. (1978):** „Jednostki języka a produkty językowe. Problem tzw. orzeczeń peryfrastycznych”, *Z zagadnień słownictwa współczesnego języka polskiego, Prace Językoznawcze* 91, PAN, Ossolineum, s. 17-30;
- (1987): „Obiekty leksykograficzne a jednostki języka”, *Studia z polskiej leksykografii współczesnej*, II. Red. Z. Saloni, s. 13-34;
- (1988): *Język w słowniku*, Ossolineum, Wrocław – Warszawa – Kraków – Gdańsk – Łódź;
- (1996): „O zasadach rejestracji jednostek języka”, *Poradnik Językowy*, 8:356-364.
- Buttler, D. (1976):** *Innowacje składniowe współczesnej polszczyzny. Walencja wyrazów*. Warszawa;
- (1988): „Leksykalne czasowniki posiłkowe w konstrukcjach peryfrastycznych typu *wywrzeć wpływ* na przełomie XIX i XX wieku”, *Z problemów frazeologii polskiej i słowiańskiej*, IV: 71-82.
- Cockiewicz, W. & Śliwiński, W. (1982):** „O pewnej metodzie badań ilościowych frazeologii współczesnej polszczyzny mówionej w telewizji”, *Stalość i zmienność związków frazeologicznych*, Red. A. Lewicki, Lublin, s. 137-162.
- Doroszewski, W. (1952):** „O pracy nad układem hasel w słowniku”, *Poradnik Językowy*, 10:1-5; 9
- Gross, G. (1989):** „Le dictionnaire et l'enseignement de la langue maternelle”, *Wörterbücher Dictionaries Dictionnaires*, Walter de Gruyter, Berlin – New York;
- Gross, G. & Vivès, R. (red.) (1986):** *Syntaxe des noms*. Langue Française, 69.
- Gross, M. (1975):** *Méthodes en syntaxe*, Hermann, Paris;
- (1981): „Les bases empiriques de la notion de prédicat sémantique”, *Langages*, 63:7-52;
- (1988): *Rapport Scientifique du LADL (1986-1988)*;
- Habert, B. & Nazarenko, A. & Salem, A. (1997):** *Les linguistiques de corpus*, Armand Colin, Paris.
- Hearn, P.M. & Button, D.F. (red.) (1994):** *Language Industries Atlas*, IOS Press, Amsterdam, Oxford, Washington, Tokyo.
- Ibrahim, A. H. (red.) (1996):** *Les supports*. Langages, 121, Paris;
- Jędrzejko, E. (1998):** *Słownik polskich zwrotów werbo-nominalnych. Zeszyt próbny*, Energeia, Warszawa.
- Karolak, S. (1984):** „Składnia wyrażań predykatywnych”, *Gramatyka współczesnego języka polskiego. Składnia*. Red. Z. Topolińska, Warszawa, s. 11-210;
- (1995b): „Walencja”, *Encyklopedia językoznawstwa ogólnego*. Red. K. Polański, Ossolineum, Wrocław – Warszawa – Kraków, s. 580-581.
- Lewicki, A. M. (1976):** *Wprowadzenie do frazeologii syntaktycznej. Teoria zwrotu frazeologicznego*, Katowice;
- (1977): „Zwroty frazeologiczne, czyli predykaty w formie składników nieciągłych”, *Studia Gramatyczne*, I:135-143;
- (1982) (red.): *Stalość i zmienność związków frazeologicznych*, Lublin;
- (1982): „Problemy metodologiczne wariantowości związków frazeologicznych”, *Stalość i zmienność związków frazeologicznych*. Red. A. M. Lewicki, Lublin, s. 37-46;
- (1996) (red.): *Problemy frazeologii europejskiej*, I, Energeia, Warszawa;
- (1996): „Relacyjna siatka derywacyjna jako czynnik onomazjologicznego opisu frazeologicznego”, *Problemy frazeologii europejskiej*, I, Energeia, Warszawa, s. 9-14.
- Pajdzińska, A. (1988):** *Związki frazeologiczne nazywające akt mowy. Semantyka i składnia*, UMCS, Lublin.
- Piotrowski, T. (1994):** *Z zagadnień leksykografii*, PWN, Warszawa.
- Polański, K. (red.) (1980):** *Syntaktyczno-generatywny słownik czasowników polskich*, Ossolineum, Wrocław, Warszawa, Kraków, Gdańsk;
- Saloni, Z. & Świdziński, M. (1998):** *Składnia współczesnego języka polskiego*, PWN, Warszawa.

- Sinclair, J. (1996):** *Preliminary recommendations on Corpus Typology*. Rap. tech., EAGLES (Expert Advisory Group on Language Engineering Standards), may 1996, CEE
- Skorupka, S. (1967):** *Słownik frazeologiczny języka polskiego*, tom 1, Warszawa;
- (1968): *Słownik frazeologiczny języka polskiego*, tom 2, Warszawa;
- (1976): „O wieloznaczności związków frazeologicznych”, *Poradnik Językowy*, 5;
- Szober, S. (1937):** *Na straży języka. Szkice z zakresu poprawności i kultury języka polskiego*, Nasza Księgarnia, Warszawa.
- Szulc, A. (1994):** *Słownik dydaktyki języków obcych*, PWN, Warszawa.
- Szymczak, M. (red.) (1978):** *Słownik języka polskiego*. Warszawa.
- Topolińska, Z. (1984) (red.):** *Gramatyka współczesnego języka polskiego. Składnia*. Warszawa;
- (1984): „Składnia grupy imiennej”, *Gramatyka współczesnego języka polskiego. Składnia*. Red. Z. Topolińska, Warszawa, s. 301-393.
- Vetulani, G. (1995):** „Rzeczowniki predykatywne w języku polskim”, *Studia Romanica Posnaniensia*, XX:177-197, UAM, Poznań;
- (2000a): „Quelques exemples d'analyse des corpus en vue de la traduction”, *Studia Romanica Posnaniensia*, XXV, UAM, Poznań
- (2000b): *Rzeczowniki predykatywne języka polskiego. W kierunku syntaktycznego słownika rzeczowników predykatywnych*. Wydawnictwo Naukowe UAM, Poznań
- Vetulani, Z. & Walczak, B. & Obrębski, T. & Vetulani, G. (1998):** *Unambiguous coding of the inflection of Polish nouns and its application in electronic dictionaries – format POLEX. Jednoznaczne kodowanie fleksji rzeczownika polskiego i jego zastosowanie w słownikach elektronicznych – format POLEX*, UAM, Poznań.
- Vetulani, Z. & Martinek, J. & Obrębski, T. & Vetulani, G. (1998):** *Dictionary Based Methods and Tools for Language Engineering*, UAM, Poznań.
- Walczak, B. (1999):** *Zarys dziejów języka polskiego*, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław.