LINGUISTICA COPERNICANA
Nr 14 / 2017

http://dx.doi.org/10.12775/LinCop.2017.008

Grażyna Vetulani

Uniwersytet im. Adama Mickiewicza w Poznaniu
Instytut Filologii Romańskiej
Uniwersytet Mikołaja Kopernika
Katedra Filologii Romańskiej

Próby formalizacji zdań opartych na predykatach rzeczownikowych języka polskiego

Słowa klucze: język polski; predykacja rzeczownikowa; formalizacja; kodowanie informacji; słowniki elektroniczne

Keywords: Polish language; predicative nouns; formalization; information encoding; electronic dictionaries

0. Wstęp

W niniejszym artykule chcemy zwrócić uwagę na funkcjonowanie w języku polskim rzeczowników, które przyjmują w zdaniu funkcję predykatów, a także przedstawić zasady przyjęte podczas budowy haseł słownikowych dla tych jednostek. Zasady te wpisują się w ogólną koncepcję leksykonu gramatycznego, która zakłada prezentację jednostki leksykalnej wraz z informacją o jej łączliwości z pozostałymi elementami w zdaniu elementarnym. Format słownika ma formę zakodowaną, ponieważ jest przygotowywany na potrzeby zastosowań informatycznych.

1. Rola i znaczenie formalizmów w opisie językoznawczym

Ze względu na nieustanny postęp technologiczny, a w tym na konieczność przetwarzania języka naturalnego, obecnie przywiązuje się coraz większą wagę do kodowania opisu językoznawczego. Bez zastosowania różnego rodzaju formalizmów trudno byłoby sobie dziś wyobrazić funkcjonowanie systemów informatycznych wspomagających tłumaczenie, wytwarzanie za pomocą technik komputerowych określonych dokumentów, automatyczną korektę tekstów czy np. korzystanie z wyszukiwarek umożliwiających natychmiastowy dostęp do informacji zawartych w źródłach internetowych. Z całą pewnością można stwierdzić, iż szybkość dostępu do potrzebnych danych zależy od sposobu formalizacji i organizacji opisu jednostek językowych zgromadzonych w zinformatyzowanych bazach leksykalnych.

Stosowanie metod formalnych w opisie języka sięga końca lat 40. XX wieku. Językoznawcy, zainspirowani modelami matematycznymi i logicznymi, zaczęli wprowadzać do lingwistyki pełne symboli opisy sformalizowane, mając na uwadze ich wykorzystanie w tłumaczeniu automatycznym. Choć ze względu na wielkość, złożoność i elastyczność systemu językowego wykorzystywanie metod formalnych napotyka liczne przeszkody, to niektóre z nich weszły na stałe do dziedziny, wzbogacając jej aparat pojęciowy. Do najbardziej rozpowszechnionych metodologii językoznawczych opartych na sformalizowanych regułach należy chyba gramatyka generatywno-transformacyjna wraz z jej terminologią. Polański pisze, iż: „Pojęcie gramatyki generatywnej wiąże się z wprowadzeniem do językoznawstwa metody modelowania. [...] Do znanych należą także gramatyka kategorialna i gramatyka (semantyka) Montague" (Polański, 1995: 186–187). Innym szeroko stosowanym formalizmem w analizie semantycznej zdań jest *rachunek predykatów.* W logice predykat jest rozumiany jako część zdania (sądu), w której orzeka się pewną własność o podmiocie (tak jest w konstrukcjach czasownikowych nieprzechodnich oraz konstrukcjach atrybutywnych, w których predykaty są jednoargumentowe) lub wyraża relację pomiędzy dwoma (lub więcej) argumentami. Jest to element centralny w strukturze zdaniowej (w nim tkwi sens). Z punktu widzenia gramatyki predykat otwiera pozycje dla pozostałych elementów. Jeśli chodzi o jego formę dyskursywną, może wystąpić pod postacią różnych kategorii gramatycznych (Karolak, 1984: 89). W dalszej części pracy koncentrujemy się na predykatach rzeczownikowych języka polskiego.

1.1. Typy formalizacji

1.1.1. Od pojęcia do formy. Różne realizacje jednego podstawowego sensu

Jeśli przez *formalizację* (*strukturalizację*) należy rozumieć „przyporządkowanie strukturom semantycznym właściwych im wykładników", to zgodnie z tym stwierdzeniem w „wyniku strukturalizacji powstają ciągi należące do płaszczyzny formalnosyntaktycznej i stanowiące tzw. struktury powierzchniowe" (Karolak 1995: 520). Tym samym „jednemu sensowi odpowiada kilka wyrażeń należących do różnych kategorii i subkategorii morfologicznych różniących się między sobą rozmaitymi właściwościami formalnosyntaktycznymi" (Karolak, *ibidem*). O synonimii składniowej pisała też E. Jędrzejko: „W składni oznacza to istnienie różnych wzorców syntaktycznych do wyrażania tej samej struktury semantycznej, traktowanej jako jednostka 'najgłębszego' poziomu. Możliwe jest zatem tworzenie konstrukcji formalnie odmiennych, które przekazują tę samą treść propozycjonalną […]. Za semantycznie równoważne uznajemy składniki wypowiedzenia, którym na poziomie głębokim można przypisać tę samą charakterystykę funkcjonalną predykatu lub argumentu o określonej roli […]" (Jędrzejko 1993: 38, 44). Typowe w dyskursie są zatem sytuacje, kiedy jednemu głębokiemu sensowi odpowiadają różne struktury, np. werbalne / nominalne: *badać* i *prowadzić badania*, *panikować* i *wpadać w panikę*, nominalne / adiektywne: *mieć ambicje* i *być ambitnym, wykazywać konsekwencję* i *być konsekwentnym*, werbalne / nominalne / adiektywne: *złościć się / wpadać w złość / być złym* itp.), pod warunkiem jednak, że skorelowane formy predykatywne mają taką samą walencję.

1.1.2. Formalizacja informacji przyhasłowej

Każda jednostka, prosta lub złożona, jest umieszczana i opisywana w słowniku. Te z kolei buduje się w zależności od zapotrzebowań użytkowników. Kiedy słowniki są przeznaczone dla czytelnika-człowieka, informacja przyhasłowa przybiera zazwyczaj tradycyjną formę opisową, natomiast gdy ich odbiorcą jest program komputerowy, wymagania dotyczące formatu stają się niezwykle rygorystyczne. Znalezienie odpowiedniego zapisu, tj. sformalizowanie hasła, stanowi często odrębne zadanie badawcze, gdyż trzeba

opracować kod, który, z jednej strony, będzie uwzględniał specyfikę systemu językowego i charakterystykę użycia opisywanej formy (np. polisemię wyrazową, ograniczenia w użyciu itd.), a z drugiej, który będzie gotowy do wykorzystania w aplikacjach informatycznych pozbawionych intuicji językowej. Obecnie już istnieją albo są w budowie dla różnych języków coraz pełniejsze i doskonalsze *słowniki elektroniczne*[1], które różnią się od siebie pod względem struktury formalnej oraz przeznaczenia, ale w każdym rygorystycznie przestrzega się kodowania informacji. Dzięki prawie nieograniczonym zasobom pamięci oraz stosowaniu technik kompresji wielkość takich słowników nie stanowi dziś bariery technologicznej. Według Bogackiego (1997) słownik 140 000 leksemów generuje 2 400 000 form fleksyjnych języka polskiego, a więc tyleż linii kodu. Więcej danych na temat zasad kodowania jednostek leksykalnych języka polskiego można znaleźć np. w Vetulani, Walczak, Obrębski, Vetulani (1998).

Dla języka polskiego dość wcześnie, bo jeszcze w okresie, kiedy komputery nie były w powszechnym użyciu, proponowano rozwiązania przydatne do przetwarzania automatycznego. Jako przykład warto przywołać tutaj opis (choć nie uzyskał on formy cyfrowej) zastosowany w pionierskim, biorąc pod uwagę formalny opis czasownika, *Słowniku syntaktyczno-generatywnym czasowników polskich*. Poniżej sformalizowane hasło dla czasownika *brać* (Polański (red.) 1980: 7):

BRAĆ

I. 'ujmować, chwytać, wydobywać co skąd'

$$1.\ NP_N - NP_{Acc} + \left\{ \begin{array}{c} NP_O + NP_{Abl} \\ od \cap NP_G \end{array} \right\}$$

II. 'czerpać, korzystać z czego'

$$NP_N - NP_{Acc} + z \cap NP_G$$

[1] Nie chodzi o e-słowniki, które są *de facto* słownikami tradycyjnymi, tyle tylko, że są dostępne na nośnikach elektronicznych zamiast na papierze, ale o takie, których odbiorcą jest program komputerowy.

Jak pisze Polański (*ibidem*), „Różnica między znaczeniem oznaczonym wyżej jako I. a znaczeniem oznaczonym jako II. wiąże się ściśle z odrębnością struktury syntaktycznej symbolizowanej przez odpowiednie schematy zdaniowe". Widać wyraźnie, iż w tym zapisie chodzi o odzwierciedlenie różnych struktur i zarazem różnych znaczeń tej samej formy wyrazowej. Zapis formalny umożliwia zatem dostęp (np. systemom informatycznym) do potrzebnej informacji lingwistycznej.

Istnieje wiele sposobów formalizacji i organizacji opisu haseł słownikowych. Jeśli chodzi o predykaty rzeczownikowe, wiadomo, iż w sposób naturalny łączą się one z czasownikami, tworząc swoiste związki wyrazowe. Ponieważ łączliwość z czasownikiem oraz z pozostałymi elementami w zdaniu jest typowym przykładem informacji o wymaganiach składniowych formy predykatywnej (i jednocześnie wyznacznikiem jej znaczenia, choć nie jedynym), istotne jest, by tę informację podać w opisie, tj. w haśle słownikowym (lub zakodować na potrzeby zastosowań komputerowych). Lingwiści stosują w takich przypadkach różne rozwiązania. Na przykład A. Lewicki, przy okazji badań nad strukturami z komponentem nominalnym (niemającymi wówczas nic wspólnego z lingwistyką komputerową), pokazał, iż układają się one w *rodziny frazeologiczne*. Pisał, iż „wszystkie frazeologizmy należące do danej rodziny mają inwariant semantyczny, najczęściej identyczny ze znaczeniem jednego z frazeologizmów i łączy je współwystępowanie w każdym tego samego komponentu nominalnego". Zgodnie z tym stwierdzeniem wygodnym rozwiązaniem do zastosowania w słownikach, ale także dlatego, że w ten sposób można zdać sprawę z wariantywności systemu i bogactwa analitycznych form, byłaby prezentacja predykatu rzeczownikowego wraz ze współwystępującymi z nim czasownikami. Jedną rodzinę wyrażeń można utworzyć wokół jednego podstawowego sensu predykatu. Poniżej dwie rodziny (dwie *siatki derywacyjne*, Lewicki 1996: 13) oparte na formie *relacja*, każda wokół innego znaczenia:

I
nadawać / nadać relację z czegoś,
prowadzić / przeprowadzić relację z...,
przedstawiać / przedstawić relację z...,
przekazywać / przekazać relację z...,
składać / złożyć relację z...,
zdawać / zdać relację z...

II
być w relacjach z kimś,
nawiązywać / nawiązać relacje z...,
podtrzymać / podtrzymywać relacje z...,
pozostać / pozostawać w relacjach z...,
pozostawać / pozostać w relacji jakiejś do kogoś,
układać / ułożyć relacje z...,
utrzymywać / utrzymać relacje z...,
wchodzić / wejść w relacje z...,
zachowywać / zachować relacje z...

Rzecz jasna, przedstawione powyżej rozwiązanie nie nadaje się jeszcze do odczytu przez system komputerowy, podobnie jak nie nadają się do tego słowniki tradycyjne wydane w formie książkowej, choć zawierają cenne dane. Opisy tradycyjne są nieprzydatne do przetwarzania informatycznego, gdyż tylko to, co jest sformalizowane może być zaprogramowane. Łatwo można jednak sobie wyobrazić wykorzystanie zaprezentowanego rozwiązania, tzn. przeformatowanie hasła i zakodowanie informacji składniowej dla każdego pojedynczego wyrażenia, pod warunkiem, że wzięłoby się dodatkowo pod uwagę wszystkie ograniczenia gramatyczne związane z użyciem konkretnego zwrotu. A chodzi o takie kategorie selektywne, jak aspekt gramatyczny współwystępującego z rzeczownikiem czasownika (niedokonany / dokonany), liczbę pojedynczą lub mnogą samego predykatu (*nawiązać / nawiązywać relację / relacje*), przyimki wprowadzające argumenty: (*pozostać w relacji z kimś / pozostać w relacji jakiejś do kogoś*), naturę argumentów itd.

Innym przykładem opisu predykatów rzeczownikowych języka polskiego jest format zaproponowany przez Żmigrodzkiego[2], który zastosował rozwiązanie polegające na umiejscowieniu w wejściu słownikowym nie tyle predykatu, ile łączącego się z nim czasownika, czyli *werbalizatora* (strukturę analityczną *czasownik* + *predykat rzeczownikowy* odnajdujemy na końcu linii):

[2] Przykład zaczerpnięty z książki E. Jędrzejko (2002), w której autorka przytacza opis stosowany przez P. Żmigrodzkiego w pracy pt. *Właściwości składniowe analitycznych konstrukcji werbo-nominalnych w języku polskim* (Żmigrodzki 2000).

DAWAĆ // DAĆ$_1$ ((Acc) FBAZ [N → czynność]: *dawać // dać koncert, występ, recital, pokaz*

DAWAĆ // DAĆ$_2$ (Acc + Dat) FBAZ [N → akt intelekt.-woliatyw]: *dawać // dać komu zgodę, pozwolenie, obietnicę*

Powyższe, a także inne, liczne już dzisiaj, rozwiązania mają charakter leksykonów gramatycznych. Ich cechą charakterystyczną jest równoczesne wprowadzanie do słownika jednostki leksykalnej i informacji składniowej z nią związanej. Warto wspomnieć w tym miejscu o istniejących leksykonach gramatycznych opracowanych dla innych języków, pośród których chyba najpełniejszym jest zbiór *tablic składniowych* (*tables syntaxiques*) opracowany dla języka francuskiego. Od lat 70. XX wieku, w zespole L.A.D.L.[3] pod kierownictwem M. Gross'a, konsekwentnie budowano ten rodzaj słowników[4]. W rezultacie rozpoczętych wówczas prac powstała olbrzymia, zinformatyzowana baza językowa zawierająca podstawowe jednostki predykatywne leksyki francuskiej (obejmująca wszystkie kategorie gramatyczne, w tym predykaty rzeczownikowe) wraz z ich gramatyką (zob. Vetulani 2012: 82–85).

2. W poszukiwaniu odpowiedniego formalizmu dla zdań opartych na predykatach rzeczownikowych języka polskiego – na podstawie prac własnych[5]

Jak zostało powiedziane już wyżej, opis predykatów rzeczownikowych języka polskiego jest zgodny z zasadami leksykonu gramatycznego. Kontekstem składniowym, który służy za podstawę do rozróżnienia znaczeń pojedynczej formy jest zdanie elementarne, tj. takie, w którym występuje tylko jeden predykat wraz z implikowanymi przez siebie argumentami.

Od samego początku budowania słownika opis otrzymał formę zakodowaną, gdyż był przygotowywany z myślą o przetwarzaniu komputerowym. Każdemu predykatowi jest przypisany przynajmniej jeden model, który jest

[3] Laboratoire d'Automatique Documentaire et Linguistique (Université Paris 7).

[4] Pełna bibliografia prac powstałych w zespole L.A.D. L. na Uniwersytecie Paris 7 (do roku 1998) w: Lamiroy (red.) 1998.

[5] W tej części artykułu korzystamy częściowo z treści zawartych w artykułach: Vetulani 2013, Vetulani, w druku (zob. bibliografia).

odbiciem jego autentycznego użycia w dyskursie. W przypadku polisemii formy predykatywnej, liczba modeli odpowiada liczbie znaczeń.

2.1. Początki budowania słownika – wczesne lata 90.

2.1.1. Wybór źródła i gromadzenie danych

W pierwszej kolejności należało wyznaczyć zbiór jednostek, tj. rzeczowników, które mogą przyjąć na siebie funkcję predykatu. Kryterium decydującym było użycie rzeczownika w sensie abstrakcyjnym, bowiem tylko tego typu forma ma zdolność orzekania o obiekcie lub o tym, co zachodzi. Wyszukiwanie danych odbywało się wyłącznie metodą tradycyjną, a za korpus obserwacyjny posłużyły informacje przyhasłowe zawarte w *Słowniku Języka Polskiego* (SJP) pod red. M. Szymczaka (1978–1981). Systematyczny ogląd zawartych w tym opracowaniu danych szybko doprowadził do zwiększenia liczby predykatów w stosunku do pierwotnie pobranych z tego dzieła form rzeczownikowych (ok. 8 000), ponieważ jedna forma wchodzi w różne struktury zdaniowe, a każda z nich odpowiada jednemu sensowi. Zaznaczmy przy okazji, iż obserwacja wymagań składniowych musiała się odbywać w sposób wysoce skrupulatny, gdyż należało wziąć pod uwagę wszystkie współwystępujące z predykatem elementy, tj. towarzyszący mu czasownik (często kilka, z których każdy musiał być przeanalizowany pod kątem ograniczeń na poziomie aspektu), liczbę i naturę implikowanych argumentów, sposób wprowadzania argumentów (z przyimkiem lub bez), konieczność pojawienia się w strukturze jakiegoś modyfikatora (przeważnie przymiotnika), bez którego zdanie byłoby niepoprawne gramatycznie itd. Metoda budowania schematów zdaniowych odpowiadała temu, co zostało powiedziane w *Składni języka polskiego*, gdzie S. Karolak (1984: 14) pisał o: „właściwościach wewnątrzrelacyjnych zdań", w których ważną rolę odgrywają reguły „semantyczno-relacyjne lub inaczej semantyczno-syntaktyczne właściwości składników struktury sensu". Te ostatnie autor rozumiał jako „zdolność do wzajemnego współwystępowania", a zatem łączliwość na płaszczyźnie semantycznej, którą określił jako „zgodność semantyczną". Podobnie wypowiadała się Z. Topolińska, choć zwracała uwagę tylko na towarzyszące predykatowi argumenty: „charakter i ilość wymaganych wyrażeń argumentowych uznaliśmy za jedno z najważniejszych kryteriów w procesie charakterystyki i klasyfikacji wyrażeń predykatywnych." (Topolińska 1984: 301).

Mimo licznych trudności wynikających z niskiej jakości źródła (SJP), tzn. braku wielu informacji koniecznych do odtworzenia całkowitego modelu strukturalnego zdania, udało się opracować pierwszą wersję słownika predykatów rzeczownikowych języka polskiego (Vetulani 2000), choć trzeba przyznać, że liczba zaprezentowanych modeli była wówczas dość skromna (wynikało to bezpośrednio z obranej na początku metody). Tym niemniej, po zakończeniu prac pierwszego etapu, można było potwierdzić stosowalność obranego podejścia do języka polskiego.

Kryterium składniowe oraz analiza odniesień semantycznych badanych jednostek doprowadziły do wyodrębnienia 5 klas predykatów (możliwa jest ich dalsza klasyfikacja przy zastosowaniu tych samych kryteriów). Klasa I, niejednorodna, obejmuje rzeczowniki odnoszące się do rozmaitych czynności, zachowań, technik, operacji, analiz, metod, stanów, procesów itd., Klasa II to rzeczowniki będące nazwami cech, Klasa III zawiera nazwy chorób, Klasa IV z kolei to nazwy profesji, a Klasa V obejmuje jednostki, które występują z tzw. *czasownikiem okolicznościowym* (fr. *verbe support d'occurrence*) typu: *mieć miejce, zachodzić, odbywać się*. Ze względu na polisemię form wspomniane klasy nie są rozłączne (por. *opowiedzieć dowcip* – gdzie predykat *dowcip* odsyła do aktu (pojedynczego zachowania) i dlatego przynależy do Klasy I, i *mieć cięty dowcip* – gdzie chodzi o cechę i dlatego jednostka ta jest również elementem Klasy II). Mimo że wstępnie wyznaczono pięć różnych klas predykatów w opracowaniu z 2000 roku zostały opisane wymagania składniowe jedynie dla jednostek Klasy I.

2.1.2. Format opisu

Po pierwszym etapie prac powstał tzw. *zasób początkowy*, który był następnie rozwijany na innym materiale językowym oraz z wykorzystaniem wyspecjalizowanych narzędzi informatycznych (patrz dalej, rozdział 2.2.). Wszystkie schematy z zasobu początkowego podpadają pod ten sam model ogólny: *N0 Vsup (Prep) (MOD) Npred (Prep) N1 (Prep) N2....*, w którym *N0* odnosi się do argumentu-podmiotu (w słowniku systematycznie go pomijamy ze względu na oczywistość użycia), *Vsup* to symbol czasownika, który towarzyszy predykatowi w użyciu, *Npred* reprezentuje sam predykat, a *N1, N2* to kolejne argumenty. *(Prep)* oznacza przyimek, a *(MOD)* reprezentuje obowiązkowy modyfikator występujący najczęściej pod postacią przymiotnika

(por. *Jan ma ciężki / żołnierski / kaczy / ... chód* i **Jan ma chód*). W rzeczywistości więc na słownik składają się schematy (modele) odpowiadające zdaniom elementarnym zgodnie z zaświadczeniami, które wystąpiły w korpusie. Jeden schemat odpowiada jednemu sensowi formy predykatywnej. Tym samym został zniesiony problem jej polisemii.

W pierwszej wersji słownika wejściem słownikowym jest predykat, po którym następują, oddzielone przecinkami, schematy jego użyć: jeden lub kilka, w zależności od odnalezionych w słowniku znaczeń, będących pochodną łączliwości predykatu z właściwymi mu czasownikami. Po czasowniku, w nawiasie, podany jest również (symbolicznie) przypadek gramatyczny predykatu (*D* – Dopełniacz, *B* – Biernik itd.) oraz informacja na temat liczby (zaznacza się jedynie występowanie predykatu obowiązkowo w liczbie mnogiej). Por. (Vetulani 2000a: 158, 164):

aforyzm, m/ układać(B,lmn)
...
głupstwo, n/gadać(B,lmn), pleść(B,lmn), palnąć(B), mówić(B,lmn), opowiadać(B,lmn), robić(B), popełnić(B), narobić(D,lmn)

Explicite podaje się przyimek wprowadzający każdy kolejny argument, a po przyimku, w nawiasie, symbol przypadka gramatycznego argumentu, np.:

agitacja, ż/ przeprowadzać(B)/N1wśród(D)/N2za(N), ulec/N1za(N)
...
agresja, ż/czuć(B)/N1do(D);wobec(D), odczuwać(B)/N1do(D);wobec(D), przejawiać(B)/N1wobec(D), kierować(B)/N1przeciw(C), dokonać(D)/N1na(B), dokonać aktu-(D)/N1na(B), popełnić(B)/N1wobec(D)

Konieczność wystąpienia dodatkowego elementu zaznacza się symbolem *MOD*:

egzystencja, ż/ wieść(B)/MOD

2.2. Wspomagane komputerowo rozwijanie słownika

Podstawowym zadaniem w drugim etapie badań było rozszerzenie słownika. Prace dotyczyły w dalszym ciągu jednostek Klasy I, lecz zostały przeprowadzone nową metodą. Dysponując już zinformatyzowanym korpusem języka polskiego (a właściwie jego fragmentem[6]), oraz odpowiednimi programami do obróbki komputerowej udostępnionego materiału (pakietem słowników elektronicznych języka polskiego, programami indeksującymi wyrazy oraz systemami generującymi konkordancje[7]), przystąpiono do prac nad pozyskaniem nowych struktur w stosunku do tych, które zostały zaprezentowane w pierwszej wersji słownika. Podczas tych prac wykorzystano stworzony ręcznie kod dla zasobu początkowego (opisany wyżej w 2.1.2.), a następnie uzyskano zbiór automatycznie wygenerowanych konkordancji utworzonych dla par *potencjalny czasownik* + *predykat rzeczownikowy* (z predefiniowanej listy)[8]. Dokładny opis zastosowanego w tej fazie algorytmu wraz z opisem technicznym poszczególnych kroków można znaleźć w Vetulani 2012. W tym miejscu ograniczymy się do stwierdzeń, że: Krok 1. (automatyczny) miał na celu filtrowanie korpusu, Krok 2. był etapem analizy ręcznej przez leksykografów, w Kroku 3. (automatycznym) zebrano w tabele konkordancji wszystkie pary *czasownik* + *rzeczownik*, które odpowiadały wzorcowi konstrukcji predykatywnej, w Kroku 4. (ręcznym) ponownie przystąpiono do czynności sprawdzających oraz do opisu syntaktycznego zachowanych konstrukcji, a Krok 5 (ręczny) był etapem weryfikacji danych przez głównego leksykografa. Zastosowana metoda okazała się skuteczna, ponieważ pozwoliła w dość szybkim tempie uzyskać znaczny przyrost danych słownikowych.

---

[6] IPI PAN Korpus (Przepiórkowski 2004). Udostępniona wersja liczyła 80 milonów słów i była nieotagowana.

[7] Wszystkie narzędzia zostały wytworzone w Zakładzie Lingwistyki Komputerowej i Sztucznej Inteligencji na Wydziale Matematyki i Informatyki UAM (kier. Z. Vetulani) w ramach projektów: KBN, 1994–1996: POLEX – POLSKA LEKSYKALNA BAZA DANYCH, projekty KE: CEGLEX (CPERNICUS 1032, 1995–1996) oraz GRAMLEX (COPERNICUS 621, 1996–1998).

[8] Prace prowadzili: G. Vetulani, T. Obrębski, A. Kaliska, M. Nkollo. W zasadniczej części były one finanansowane przez MNiSW (Nr R00 02802), tytuł projektu: „Technologie przetwarzania tekstu polskiego zorientowane na potrzeby bezpieczeństwa publicznego; komunikacja człowieka z systemem informatycznym w warunkach kryzysowych przy użyciu języka naturalnego" (od 15.12.2006 r. do 28.02.2010 r.).

2.3. Nowy format hasła słownikowego

Formalna konieczność wspólnego wystąpienia czasownika i predykatu rzeczownikowego przy jednoczesnej stopniowej leksykalizacji tych elementów powoduje, że można spojrzeć na ten typ predykacji jak na orzekanie w formie znaku nieciągłego. W dalszym ciągu prac nad rozwijaniem słownika nastąpiło zatem skoncentrowanie się na zwrotach (kolokacjach) werbo-nominalnych jako na podstawowych i samodzielnych jednostkach języka. Fakt ten znalazł odbicie w budowie hasła słownikowego.

Ze względu na przeznaczenie komputerowe słownika format poszczególnych haseł otrzymał, podobnie jak w zasobie początkowym, formę zakodowaną. Każdej kolokacji został przypisany jeden schemat strukturalny, odpowiadający wzorcowi elementarnego zdania, zgodnie z przykładami odnalezionymi w analizowanym materiale. W sumie jest ich ponad 14 600. Słownik został opracowany w wersji elektronicznej i dołączony do monografii (Vetulani 2012). Nie licząc drobnych szczegółów, notacja zawarta w haśle pozostała zasadniczo niezmieniona w stosunku do wersji z 2000 roku. Nowością było wprowadzenie do słownika, w celu ilustracji, przykładowych zaświadczeń w formie autentycznych kontekstów pobranych z korpusu. Poniżej wyciąg ze słownika dla predykatu *agresja*:

=>agresja, ż
czuć agresję/ czuć(B)/N1do(D);wobec(D);w stosunku do(D),
dokonać agresji/ dokonać(D)/N1na(Ms),
dokonać agresji/ dokonać aktu(D)/N1na(Ms),
+ dokonywać agresji/ dokonywać(D)/N1na(Ms),
dopuścić się agresji/ dopuścić się(D)/N1na(Ms),
+ dopuszczać się agresji/ dopuszczać się(D)/N1na(Ms),
doświadczać agresji/ doświadczać(D)/N1ze strony(D),
doświadczyć agresji/ doświadczyć(D)/N1ze strony(D),
kierować agresję/ kierować(B)/N1przeciw(C),
odczuwać agresję/ odczuwać(B)/N1do(D);wobec(D);w stosunku do(D),
popełnić agresję/ popełnić(B)/N1wobec(D),
+ przejawiać agresję/ przejawiać(B)/N1wobec(D),
+ przejawić agresję/ przejawić(B)/N1wobec(D),
reagować agresją/ reagować(N)/N1wobec(D),

zareagować agresją/ zareagować(N)/N1wobec(D),
skierować agresję/ skierować(B)/N1przeciw(C),
wybuchać agresją/ wybuchać(N)/N1wobec(C),
+ wybuchnąć agresją/ wybuchnąć(N)/N1wobec(C),
wykazać agresję/ wykazać(B)/N1wobec(D),
wykazywać agresję/ wykazywać(B)/N1wobec(D),
+ zareagować agresją/ zareagować(N)/N1wobec(D),

***** dokonać**	
po tym jak	**[* dokonało_ono_agresji *]** na Kuwejt, podobnie jak swego
ostrzegając: jeżeli ktoś	**[* dokonałby_agresji *]** na Polskę w czasie, gdy
Wprowadzając stan wojenny,	**[* dokonano_agresji *]** w brutalny, bo siłowy sposób
***** dopuścić**	
dzieckiem ojcu, który	**[* dopuścił_się_agresji *]** stosował przemoc wobec matki
chcieli przecież nie	**[* dopuścić_do_takiej_agresji *]?**
***** doświadczać**	
w swej historii	**[* doświadczała_obcej_agresji *].**
***** doświadczyć**	
w swej historii	**[* doświadczyła_obcej_agresji *] .**
***** reagować**	
kontrolowało czynności,	**[* reagowało_agresją *]** i krzykiem na próby nawiązania
***** skierować**	
w Hucie Jedność	**[* skierowali_swoją_agresję *]** przeciwko prezydentowi miasta, przeciwko o
***** wybuchać**	
niezadowolona z siebie,	**[* wybuchała_agresją *].**
***** wykazać**	
To nie policja	**[* wykazała_agresję *].**
To nie policja	**[* wykazał_agresję *]**, to związkowcy zastosowali bezprawne

*** wykazywać	
obserwowany, to znaczy	**[* wykazuje_dużo_agresji *]**, brutalności wobec osoby słab
uczestników zgromadzenia, którzy	**[* wykazywali_szczególną_agresję*] .**
przez okno albo	**[* wykazuje_agresję *]** wobec innego dziecka.

2.4. Dalsze prace z uwzględnieniem nazw cech i właściwości

W związku z pracami mającymi na celu włączenie do słownika predykatów, które są nazwami cech i właściwości[9] (jednostki Klasy II. w monografii z 2000 roku) należało w pierwszej kolejności zaproponować format opisu dla zwrotów opartych na tych jednostkach. Realizacja zadań polegała na analizie istniejącego już wzoru (opracowanego dla jednostek Klasy I.) pod kątem możliwości jego wykorzystania dla nazw, które nie zawsze wykazują ten sam typ łączliwości z elementami w zdaniu. Nadmieńmy, iż pożądane było wykorzystanie tego samego formatu, ponieważ stosowanie jednolitej i spójnej metodologii jest zawsze korzystne w informatyce (poszukuje się maksymalnie jednorodnego formatu dla zróżnicowanych elementów).

Podczas analizy wygenerowanych automatycznie konkordancji zawierających wystąpienia nazw cech i właściwości zarejestrowano dużą liczbę realizacji, w których były one użyte w mianowniku (typowych w języku polskim). W tych przypadkach występowały one najczęściej z czasownikami neutralnymi typu: *cechować* lub *charakteryzować* (por. *rzecznika cechuje obiektywizm, projekt ustawy charakteryzuje nowoczesność i otwartość*), choć zdarzały się także wystąpienia z czasownikami bardziej nacechowanymi, jak: *napadać* (por.: *napadała go wściekłość*), *ogarnąć* (por. *ogarnął go gniew*) itp. Odnotowano także liczne przypadki, dla których można było stosować bez zmian istniejący już format opisu semantyczno-gramatycznego (tzn. taki, jak dla Klasy I.). W rezultacie został zaproponowany wzbogacony wzór hasła, choć – być może – będzie on wymagał jeszcze uściślenia. Jako

[9] Prace toczyły się dzięki uzyskaniu finansowania projektu pt. *Rozbudowa zasobów cyfrowych języka polskiego w zakresie słowników walencyjnych w kierunku leksykonu-gramatyki zorientowana na potrzeby zastosowań informatycznych w humanistyce*, realizowanego w ramach NPRH (MNiSW Nr 0022/FNiTP/H11/80/2011) od 1.02.2012 do 30.04.2015 pod kierownictwem G. Vetulani.

przykład zamieszczamy poniżej schematy strukturalne dla predykatu *arogancja* wraz z przykładami użyć pobranymi z korpusu:

=>arogancja, ż

I
arogancja cechuje/ arogancja cechuje/N0(B),

II
okazywać arogancję/ okazywać(B)/N1w stosunku do(D);wobec(D),
pokazać arogancję/ pokazać(B),
prezentować arogancję/ prezentować(B),
wykazać się arogancją/ wykazać się(N),

*** **cechować**	
politykę w stosunku do Śląska	[* **cechowała głupota i arogancja** *].
W dalszym ciągu	[* **cechuje go arogancja** *] i demonstracja siły.
*** **okazywać**	
zamierzeniach, wreszcie że	[* **okazywał arogancję** *] w stosunku do związków zawodowych
Rokita Wyłącznie, żeby nie	[* **okazywać jakiejś arogancji** *] wobec pani, przeszedłem do tej
*** **pokazać**	
policji chce jutro odwołać, nie	[* **pokazał takiej arogancji** *]...
*** **prezentować**	
Panie ministrze,	[* **prezentuje pan olbrzymią arogancję** *].
*** **wykazać się**	
Pan minister Piechota	[* **wykazał się arogancją** *], gdyż nie pofatygował się w trakcie
*** **wykazywać**	
obserwowany, to znaczy	[* **wykazuje dużo agresji** *], brutalności wobec osoby
uczestników zgromadzenia, którzy	[* **wykazywali szczególną agresję** *].
przez okno albo	[* **wykazuje agresję** *] albo chce wyskoczyć z okna

Słownik predykatów będących nazwami cech i właściwości jest aktualnie w opracowaniu. Planuje się też jego uzupełnienie o jednostki pozostałych klas.

3. Uwagi końcowe

W dobie komunikacji komputerowej, cyfryzacji bibliotek, tłumaczenia wspomaganego komputerowo formalizacja danych językowych wydaje się być wymogiem oczywistym w językoznawstwie. Jednak ze względu na specyfikę konkretnego systemu dobranie formatu do opisu jednostek leksykalnych nie jest już zadaniem oczywistym.

Bibliografia

Bogacki K., 1997, *POLLEX – un dictionnaire électronique morphologique du polonais*, Grenoble: BULLAG, s. 2–9.

Jędrzejko E., 1993, *Nominalizacje w systemie i w tekstach współczesnej polszczyzny*, Katowice: Uniwersytet Śląski.

Jędrzejko E., 1996, Z zagadnień walencji rzeczownika, *Folia Philologica Macedono-Polonica 4*, s. 13–19.

Karolak S., 1984, Składnia wyrażeń predykatywnych, w: Z. Topolińska (red.), *Gramatyka współczesnego języka polskiego. Składnia*, Warszawa: Państwowe Wydawnictwo Naukowe, s. 11–210.

Karolak S., 1995, hasło: *Strukturalizacja*, w: Polański K. (red.), *Encyklopedia językoznawstwa ogólnego*, Wrocław–Warszawa–Kraków: Zakład Narodowy im. Ossolińskich, s. 520.

Lamiroy B. (red.), 1998, Le lexique-grammaire, *Travaux de Linguistique 37*, p. 7–23.

Lewicki A.M., 1996, Relacyjna siatka derywacyjna jako czynnik onomazjologicznego opisu frazeologicznego, w: A. M. Lewicki (red.), *Problemy frazeologii europejskiej*, t. 1, Warszawa: Wydawnictwo Energeia, s. 9–14.

Polański K. (red.), 1995, *Encyklopedia językoznawstwa ogólnego*, Wrocław–Warszawa–Kraków: Zakład Narodowy im. Ossolińskich.

Przepiórkowski A., 2004, *Korpus IPI PAN*, Warszawa: Instytut Podstaw Informatyki PAN.

Szymczak M. (red.), 1978–1981, *Słownik języka polskiego*, Warszawa: Państwowe Wydawnictwo Naukowe.

Topolińska Z. (red.), 1984, *Gramatyka współczesnego języka polskiego. Składnia*, Warszawa: Państwowe Wydawnictwo Naukowe.

Vetulani, G., 2000a, *Rzeczowniki predykatywne języka polskiego. W kierunku syntaktycznego słownika rzeczowników predykatywnych*, Poznań: Wydawnictwo Naukowe UAM.

Vetulani G., 2000b, Zasady budowy hasła słownikowego dla kolokacji werbo-nominalnych, *Scripta Neophilologica Posnaniensia II*, str. 173–190.

Vetulani, G., 2012, *Kolokacje werbo-nominalne jako samodzielne jednostki języka. Syntaktyczny słownik kolokacji werbo-nominalnych języka polskiego na potrzeby zastosowań informatycznych. Część I*, Poznań: Wydawnictwo Naukowe UAM.

Vetulani G., 2013, Budowa syntaktycznego słownika rzeczowników predykatywnych języka polskiego na potrzeby zastosowań informatycznych w dobie aktualnych wyzwań dla językoznawstwa, w: S. Puppel, T. Tomaszkewicz (red.), *Scripta manent – res novae*, Poznań: Wydawnictwo Naukowe UAM, s. 487–498.

Vetulani G., 2017, Problemy z pozyskiwaniem i opisem nazw cech i właściwości w języku polskim, *Kwartalnik Językoznawczy*, 2015/1-2 (21-22), s. 49–61, http://www.kwartjez.amu.edu.pl.

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*, Poznań: Wydawnictwo Naukowe UAM.

Żmigrodzki P., 2000, *Właściwości składniowe analitycznych konstrukcji werbo-nominalnych w języku polskim*, Katowice: Wydawnictwo UŚ.

Attempts of sentences formalization based on the noun predicates in Polish

(s u m m a r y)

In view of the needs related to automatic language processing, we emphasize the necessity of applying formal methods in linguistics in general, and particularly in creation of dictionary entries. We present some of our research achievements pertaining to the creation of a dictionary of Polish predicative nouns. As the dictionary is intended to be computer consulted, its entries take into account polysemy and contain information about the syntactic relations between elements of an elementary sentence. The dictionary is a kind of the lexicon grammar.