ADAM PRZEPIÓRKOWSKI*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA; UNIWERSYTET WARSZAWSKI

ELŻBIETA HAJNICZ, ANNA ANDRZEJCZUK, AGNIESZKA PATEJUK, MARCIN WOLIŃSKI**

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

Walenty: gruntowny składniowo-semantyczny słownik walencyjny języka polskiego

Słowa kluczowe: walencja, składnia, frazeologia, semantyka.

1. Wstęp

Do niedawna największym polskim słownikiem walencyjnym był *Słownik syntaktyczno-generatywny czasowników polskich* (SSG) pod redakcją Kazimierza Polańskiego, wydany w pięciu tomach w latach 1980–1992. Słownik ten zawiera około 27,5 tysiąca schematów walencyjnych dla prawie 11 tysięcy lematów (form hasłowych) czasownikowych. Dużym słownikiem, zawierającym jednak znacznie mniej szczegółowe informacje walencyjne, jest *Inny słownik języka polskiego* (ISJP) pod redakcją Mirosława Bańki, w którym około 30 tysięcy prostych schematów walencyjnych zostało przypisanych leksemom reprezentowanym przez około 17 tysięcy lematów, przy czym około 30 proc. stanowią leksemy rzeczownikowe o często regularnych wymaganiach składniowych (tj. łączące się z opcjonalnymi przydawkami, zwykle dopełniaczowymi). Informacje walencyjne zawiera także *Wielki słownik języka polskiego PAN* (Żmigrodzki 2015).

Celem autorów niniejszego artykułu jest prezentacja nowego słownika walencyjnego opracowanego przez Zespół Inżynierii Lingwistycznej w Instytucie Podstaw Informatyki PAN. Słownik ten, Walenty, jest obecnie – w momencie przygotowywania niniejszego tekstu do druku na początku grudnia 2016 roku – nadal rozwijany, lecz już teraz liczy prawie 87 tysięcy składniowych schematów walencyjnych dla leksemów reprezentowanych przez ponad 15 tysięcy lematów, w tym ponad 12 tysięcy lematów czasownikowych, ponad 2 tysiące rzeczownikowych, niecały tysiąc przymiotnikowych i około 200 przysłówkowych, przy czym w wypadku części mowy innych niż czasowniki notowane są tylko leksemy o nieoczywistych wymaganiach składniowych, a więc na przykład nie rzeczowniki o regularnej łączliwości z przymiotnikami i podrzędnikami dopełniaczowymi. Walenty jest nie tylko istotnie większy od SSG i co najmniej porównywalny z ISJP, lecz także zawiera znacznie bardziej szczegółowe niż te słowniki opisy składniowe, semantyczne i frazeologiczne; wielkość i stopień szczegółowości słownika usprawiedliwiają użycie przymiotnika GRUNTOWNY w tytule niniejszego artykułu. Omawiany

* adam.przepiorkowski@ipipan.waw.pl

** elzbieta.hajnicz@ipipan.waw.pl, anna.andrzejczuk@ipipan.waw.pl, agnieszka.patejuk@ipipan.waw.pl, marcin.wolinski@ipipan.waw.pl

słownik jest publicznie dostępny – w postaci wygodnej zarówno do przeglądania i przeszukiwania za pomocą przeglądarki internetowej, jak i do pobrania w kilku formatach.

2. Walencja i argumenty

Słowniki walencyjne opierają się na założeniu, że typy podrzędników danego predykatu można podzielić na jego argumenty i jego modyfikatory. Na przykład w poniższych zdaniach *Janek* i *gazetę* są argumentami odpowiednich czasowników, *wczoraj* jest ich modyfikatorem, natomiast fraza *na kanapie* jest argumentem czasownika POŁOŻYĆ, lecz modyfikatorem czasownika PRZECZYTAĆ.

(1) Janek położył wczoraj gazetę na kanapie.
(2) Janek przeczytał wczoraj gazetę na kanapie.

Rozróżnienie to, choć na pierwszy rzut oka intuicyjne, opiera się na nieostrych i trudnych do operacyjnego doprecyzowania intuicjach elipsy i semantycznej obligatoryjności, próżno więc szukać w większości słowników walencyjnych ścisłych definicji walencji czy testów na odróżnienie argumentów od modyfikatorów[1]. Nie inaczej jest w słowniku Walenty, w którym decyzja o tym, jakie typy podrzędników uznać za argumenty warte opisania w schemacie walencyjnym, pozostawiona jest intuicji leksykografów. Jedyną wskazówką jest tu zalecenie, by w wypadkach wątpliwych dany typ podrzędnika uznawać za argument, z czego wynika zapewne trochę większa niż oczekiwana średnia liczba argumentów na schemat leksemu czasownikowego w Walentym: 2,74 (w porównaniu z ok. 2,5 w SSG i poniżej 2,0 w ISJP). Oczywiście nie zakłada się, że wszystkie argumenty są obligatoryjne – wręcz przeciwnie, za obligatoryjne uznaje się tylko podrzędniki będące częścią frazeologizmu.

Słowniki walencyjne zwykle nie eksplikują także pojęcia pozycji argumentowej, niezbędnego do podjęcia decyzji o liczbie schematów walencyjnych i argumentów w takich schematach. Rozważmy czasownik CHCIEĆ i takie jego użycia: *chcieć czegoś* (z dopełnieniem nominalnym), *chcieć coś zrobić* (z dopełnieniem bezokolicznikowym), *chcieć, żeby coś się stało* (z dopełnieniem zdaniowym) i *chcieć dobrze* (z dopełnieniem przysłówkowym). Czy te cztery możliwości odpowiadają czterem różnym schematom walencyjnym? A może jednemu schematowi z czterema różnymi pozycjami argumentowymi? W omawianym słowniku odpowiedź na oba pytania jest przecząca, gdyż pozycja argumentowa rozumiana jest tutaj zgodnie z testem koordynacji dla pozycji składniowej rozważanym m.in. przez Marię Szupryczyńską (1996: 137–138): jeśli dwa różne morfoskładniowo typy podrzędników danego predykatu można skoordynować, to wypełniają one tę samą pozycję składniową. Ponieważ możliwość takiej morfoskładniowo niejednorodnej koordynacji nadal bywa kwestionowana, w Walentym jest ona zawsze poparta autentycznymi przykładami, zwykle z Narodowego Korpusu Języka Polskiego

1 Do wyjątków należą czeskie słowniki walencyjne VALLEX (Lopatková 2003) i PDT-Vallex (Hajič i in. 2003), oparte na praskim opisie funkcjonalno-generatywnym (Sgall i in. 1986). Krytykę przyjętych tam kryteriów odróżniania argumentów od modyfikatorów znaleźć można w artykule: Przepiórkowski 2016a, szerzej zaś o problemie odróżnienia argumentów od modyfikatorów z perspektywy polskiej „składni semantycznej" oraz w pewnym ujęciu generatywnym traktują artykuły: Przepiórkowski 2016b, 2017.

(NKJP; Przepiórkowski i in. (red.) 2012; http://nkjp.pl/), zaakceptowanymi przez leksykografów jako poprawne. Poniższe przykłady korpusowe ilustrują kolejno możliwość koordynacji dopełnienia nominalnego i zdaniowego, bezokolicznikowego i zdaniowego oraz przysłówkowego i zdaniowego, a więc wszystkie cztery typy podrzędnika należy uznać za realizację jednej pozycji argumentowej[2].

(3) Wy byście wszystko chcieli naraz i żeby było hop-siup.
(4) Dzidzia i tak już od wielu lat chciała mieć na imię Edyta i żeby się to zdrobniało Dita...
(5) Wiem, że chcecie dobrze i żeby zapanował pokój...

Oczywiście testu koordynacji nie należy stosować całkowicie mechanicznie, gdyż występuje w języku polskim konstrukcja zwana koordynacją leksykalno-semantyczną (Sannikow 1979, 1980; Kallas 1993), jak w poniższych przykładach z NKJP, w których połączone współrzędnie są podrzędniki reprezentujące niewątpliwie różne pozycje składniowe – *kto* i *komu* oraz *nic* i *nigdy*:

(6) Kto i komu dał prawo do wydziedziczenia tego narodu?
(7) Zresztą nic i nigdy nie osiągnie ideału.

Zjawisko to jest jednak ograniczone do koordynacji członów pewnego typu: w przybliżeniu zaimków i wyrażeń kwantyfikatorowych (Przepiórkowski, Patejuk 2014; Patejuk 2015). Zakładamy więc, że możliwe jest odróżnienie koordynacji leksykalno-semantycznej od zwykłej koordynacji różnych realizacji tej samej pozycji składniowej.

3. Warstwa składniowa

Projekt warstwy składniowej słownika został szczegółowo opisany w artykule Przepiórkowski i in. 2014 i zaktualizowany w artykule Hajnicz i in. 2016b; tutaj przedstawimy tylko najważniejsze aspekty poziomu składniowego Walentego na tle wcześniejszych polskich słowników zawierających informacje walencyjne.

Składniowe schematy walencyjne podane są w postaci odpowiedniej dla czynnych form danego leksemu, przy czym jawnie oznaczone są te argumenty, które w stronie czynnej wypełniają pozycję podmiotu (zob. `subj` w (8) poniżej), oraz te, które przechodzą na tę pozycję w stronie biernej (`obj`), np.:

(8) BAJDURZYĆ: `subj{np(str)} + obj{np(str)} + {np(dat)}`

W takim zapisie pozycje argumentowe są zbiorami realizacji morfoskładniowych oznaczonymi nawiasami klamrowymi i oddzielonymi plusem. W wypadku omawianego powyżej czasownika CHCIEĆ jedna z pozycji byłaby zbiorem co najmniej czteroelementowym, zawierającym realizacje: nominalną, bezokolicznikową, zdaniową i przysłówkową. W schemacie (8)

2 Warto tu także wspomnieć o może mniej naturalnym przykładzie Kallas (1993: 123): *Chce pić i papierosa*, w którym współrzędnie złożone są podrzędniki bezokolicznikowy i nominalny.

trzy pozycje składniowe są zbiorami jednoelementowymi, zawierającymi odpowiednio realizacje: `np(str)`, `np(str)` i `np(dat)`, a więc frazy nominalne (`np`) o przypadku gramatycznym podanym w nawiasach okrągłych. W najprostszym razie jest to konkretny przypadek morfologiczny, np. celownik w wypadku `np(dat)`, ale słownik zawiera także subtelniejsze specyfikacje przypadka, gdy nie jest on jednoznacznie ustalony.

W szczególności `str` oznacza przypadek strukturalny, to jest taki, którego wartość morfologiczna zależy nie tylko od danego leksemu, ale także od jego formy lub otoczenia składniowego. Gdy jako strukturalny określony jest przypadek podmiotu, jak w powyższym przykładzie, oznacza on zwykle mianownik, ale w wypadku formy gerundialnej danego czasownika[3] taki argument realizowany jest jako fraza dzierżawcza (zob. *Ciągłe bajdurzenie Piotra zaczyna mnie wkurzać*) lub fraza przyimkowa z przyimkiem PRZEZ i biernikiem (np. w *Pisanie wierszy przez Piotra*...). Oznaczenie takie jest także zgodne z tymi teoriami, według których typowe podmioty liczebnikowe występują w bierniku, a nie w mianowniku (Przepiórkowski 2004). Natomiast gdy jako strukturalny określony jest przypadek argumentu innego niż podmiot, argument występuje w bierniku lub – w odpowiednich kontekstach negatywnych czy też przy gerundium – w dopełniaczu (*Ktoś bajdurzy coś komuś* vs *Ktoś nie bajdurzy czegoś komuś* i *bajdurzenie czegoś*). Jak zostanie to pokazane w kolejnym punkcie, nie można założyć, że dopełniacz negacji dotyczy każdego argumentu biernikowego, gdyż pewne – lecz nie wszystkie – biernikowe argumenty zleksykalizowane muszą być realizowane w bierniku, nawet gdy zanegowany jest ich bezpośredni nadrzędnik.

Inaczej niż we wcześniejszych polskich słownikach zawierających informacje walencyjne (w tym SSG i ISJP), ale podobnie jak np. w czeskim słowniku VALLEX (Lopatková 2003) Walenty uwzględnia informację o kontroli składniowej (ang. *control*), a więc o tym, że w poniższych zdaniach niewyrażony bezpośrednio podmiot czasownika PRZECZYTAĆ rozumiany jest jako tożsamy z podmiotem czasownika osobowego w (9), lecz jako tożsamy z celownikowym dopełnieniem czasownika osobowego w (10):

(9) Janek obiecał Marysi przeczytać gazetę.
(10) Janek kazał Marysi przeczytać gazetę.

Jak wiadomo z literatury generatywistycznej (zob. np. Landau 2013), zjawisko kontroli jest zjawiskiem przynajmniej częściowo składniowym, o czym świadczy choćby uzgodnienie przymiotnika predykatywnego z takim niewyrażonym podmiotem co do rodzaju (i liczby):

(11) Janek obiecał Marysi być miłym / *miłą.
(12) Janek kazał Marysi być miłą / *miłym.

3 Zgodnie z podejściem *Słownika gramatycznego języka polskiego* regularne derywaty odczasownikowe, a więc przede wszystkim formy gerundialne zakończone na *-nie*/*-cie* oraz imiesłowy przymiotnikowe, włączamy do rozszerzonego paradygmatu czasownikowego (Saloni i in. 2012: 106–108).

Aby oddać różnicę zilustrowaną powyżej czasownikami OBIECAĆ i KAZAĆ, ten argument czasownika, który jest rozumiany jako podmiot podrzędnego argumentu bezokolicznikowego, oznaczany jest jako `controller`, a tak „kontrolowany" argument – jako `controllee`:

(13) OBIECAĆ: `subj,controller{np(str)} + {np(dat)} + controllee{infp(_)}`
(14) KAZAĆ: `subj{np(str)} + controller{np(dat)} + controllee{infp(_)}`

Symbol `infp(_)` oznacza wymaganie frazy bezokolicznikowej o dowolnej wartości aspektu. Znaczniki `controller` i `controllee` są też wykorzystywane między innymi do sygnalizowania – pominiętego w innych słownikach – uzgodnienia pomiędzy różnymi argumentami danego predykatu, jak w tym schemacie walencyjnym czasownika UDAWAĆ, w którym jednym z argumentów jest fraza przymiotnikowa uzgadniająca się co do liczby i rodzaju – lecz nie przypadka – z podmiotem (por. *Maria udaje miłą / *miłego / *miłe* oraz podobnie *Maria nie udaje miłej / *miłego / *miłych*):

(15) UDAWAĆ: `subj,controller{np(str)} + controllee{adjp(str)}`

Nieco bardziej skomplikowana jest kwestia uzgodnienia w konstrukcjach łącznikowych (Przepiórkowski 1999; Świdziński 2015):

(16) Ten chłopak jest miły / *miłego.
(17) Pięciu chłopaków jest miłych / *mili.
(18) Pięć dziewczyn jest miłe / miłych.

Wcześniejsze słowniki nie uwzględniają takich zależności wartości przypadka od formy leksemu lub kontekstu składniowego i niesłusznie (bo niezgodnie choćby z przykładem (17)) określają argument przymiotnikowy jako mianownikowy (SSG 1: 55; ISJP 1: 142). W Walentym przypadek przymiotników predykatywnych w takich konstrukcjach określany jest jako `pred`, a więc taki, który w specyficzny sposób uzgadnia się z podmiotem w zależności od morfoskładniowej realizacji tego podmiotu (fraza liczebnikowa określonego typu vs inna fraza nominalna).

(19) BYĆ: `subj,controller{np(str)} + controllee{adjp(pred)}`

Związane z kontrolą składniową jest zjawisko podnoszenia (ang. *raising*), które można zaobserwować w wypadku czasowników takich jak WYDAWAĆ SIĘ, ZACZĄĆ czy SKOŃCZYĆ. Czasowniki takie, gdy łączą się z frazą bezokolicznikową, „dziedziczą" specyfikację podmiotu od podrzędnego bezokolicznika. Skoro więc formy czasownika DZIWIĆ mogą łączyć się z podmiotem nominalnym lub zdaniowym (Świdziński 1992), to i ZACZĄĆ może się

łączyć z takimi podmiotami, gdy podrzędnym czasownikiem jest DZIWIĆ, jak w następujących przykładach z NKJP:

(20) Taki stan ducha jego samego zaczął dziwić.
(21) I dopiero potem zaczęło mnie dziwić, że ludzie czytali Radio Armageddon...

Podobnie, skoro w zdaniu typu *Opłacało się komuś coś zrobić* podmiotem czasownika OPŁACAĆ SIĘ jest fraza bezokolicznikowa, to może ona też być podmiotem w konstrukcji typu *zaczęło się komuś opłacać coś zrobić*, a skoro GRZMIEĆ może wystąpić bez podmiotu (nawet domyślnego), to możliwe są konstrukcje typu *Zaczęło grzmieć*. W Walentym takie czasowniki podnoszenia oznaczone są następująco (z uwzględnieniem ograniczenia na aspekt argumentu bezokolicznikowego):

(22) ZACZĄĆ: `subj,controller{E}+controllee{infp(imperf)}`

Wcześniejsze słowniki walencyjne albo w ogóle nie zdawały sprawy z faktu, że takie czasowniki mogą łączyć się z podmiotami innymi niż nominalne mianownikowe, albo – jak w wypadku elektronicznego słownika Marka Świdzińskiego (1998) – podawały dwa różne hasła dla takich czasowników, z wymaganiem podmiotu i bez wymagania podmiotu, co i tak nie pozwalało na odpowiednie opisanie podmiotów nienominalnych w tych wypadkach, w których realizacje takie dopuszczał podrzędny bezokolicznik (Świdziński 1993; Przepiórkowski 1997).

Wśród innych cech reprezentacji składniowej Walentego nieobecnych we wcześniejszych słownikach wspomnieć można także o rozróżnieniu pomiędzy trzema typami SIĘ, tj. pomiędzy inherentnym SIĘ będącym częścią lematu (np. BAĆ SIĘ), zwrotnym SIĘ zwykle wymiennym na SIEBIE (np. BARYKADOWAĆ, jak w: *barykadować się gdzieś*) oraz wzajemnościowym SIĘ współwystępującym z formami typu *nawzajem* lub *wzajemnie* (np. DARZYĆ, jak w: *darzyć się nawzajem szacunkiem* itp.), jak w następujących (tu uproszczonych) schematach:

(23) BAĆ SIĘ: `subj{np(str)}+{np(gen)}`
(24) BARYKADOWAĆ: `subj{np(str)}+{np(inst)}+{xp(locat)}+{refl}`
(25) DARZYĆ: `subj{np(str)}+{np(inst)}+{recip}`

Oczywiście należy podkreślić, że projekt warstwy składniowej omawianego słownika nie powstawał w próżni, lecz były brane pod uwagę rozwiązania z wcześniejszych słowników: punktem wyjścia był wspomniany już słownik Świdzińskiego (1998), z SSG zaczerpnięta została choćby specyfikacja pewnych argumentów z odniesieniem do ich semantyki: lokatywnej (zob. `xp(locat)` w (24)), adlatywnej, ablatywnej, temporalnej itp., wspólne z ISJP jest m.in. określanie pewnych pozycji jako dzierżawczych (realizowanych przez frazę nominalną w dopełniaczu lub przez zaimki dzierżawcze) oraz oznaczanie pewnych schematów walencyjnych przymiotników jako realizowanych tylko przez formy w pozycji predykatywnej.

4. Frazeologia

Wcześniejsze słowniki walencyjne uwzględniały frazeologię w stosunkowo ograniczonym zakresie[4]. W słowniku Świdzińskiego (1998) pewne schematy oznaczone są jako właściwe (także lub jedynie) frazeologizmom, lecz bez informacji, o jakie frazeologizmy chodzi. Natomiast w słownikach SSG i ISJP znajduje się informacja o konkretnych frazeologizmach, lecz brak informacji o wewnętrznej strukturze podrzędników będących częścią frazeologizmu. Na przykład frazeologizm (*budować*) *zamki na lodzie* wspomniany jest w obu słownikach (SSG 1: 49; ISJP 1: 778[5]), z zaznaczeniem w ISJP, że chodzi o liczbę mnogą leksemu ZAMEK, ale bez informacji, że w kontekście czasowników takich jak BUDOWAĆ ciąg *zamki na lodzie* to faktycznie dwie frazy, które mogą występować rozłącznie, jak w (26)[6], i że w kontekstach zanegowanych występuje dopełniaczowa forma *zamków*, jak w (27)[7]:

(26) Stary Dobrzański wraz z innymi – to typ szowinisty, który liczy na Francuza i zamki buduje na lodzie.
(27) Bądź realistą, nie można budować zamków na lodzie...

W Walentym takie informacje są podane jawnie:

(28) BUDOWAĆ: `subj{np(str)}` + `obj{lex(np(str),pl,'zamek',natr)}`
+ `{lex(prepnp(na,loc),sg,'lód',natr)}`

Podrzędniki czasownika będące fragmentami frazeologizmu oznaczone są jako argumenty zleksykalizowane (zob. dwa wystąpienia symbolu `lex` w (28)). Każdy taki argument opisany jest morfoskładniowo (tutaj jako fraza nominalna o przypadku strukturalnym: `np(str)` i jako fraza miejscownikowa przyimkowo-nominalna z przyimkiem NA: `prepnp(na,loc)`), przy czym podany jest leksem będący centrum danego argumentu (odpowiednio ZAMEK i LÓD) i ewentualne ograniczenia na liczbę gramatyczną (tutaj odpowiednio mnoga: `pl` i pojedyncza: `sg`). Ponieważ kolejność argumentów w schemacie walencyjnym nie musi odpowiadać kolejności ich występowania w zdaniu, schemat (28) opisuje także użycie (26), natomiast specyfikacja przypadka jako `str` zapewnia, że – m.in. w zależności od obecności negacji w otoczeniu składniowym – będzie to biernik lub dopełniacz, więc także użycie (27) jest w tym opisie uwzględnione.

Warto zwrócić uwagę na to, że nie każdy biernikowy podrzędnik czasownika podlega dopełniaczowi negacji; nie dzieje się tak choćby w wypadku frazeologizmu *mieć przechlapane*,

4 Istnieją także słowniki frazeologiczne podające – także w ograniczonym zakresie – informacje walencyjne, np. Stanisław Bąba i Jarosław Liberek (2003: 995, 169–170) charakteryzują składniowo omawiany poniżej frazeologizm (*budować*) *zamki na lodzie* jako „kto + buduje zamki na lodzie", zaś frazeologizm *bić na głowę* jako „kto + bije na głowę + kogo; kto + bije na głowę + kogo + czym; co + bije na głowę + co; co + bije na głowę + co + czym".

5 W ISJP opisany został frazeologizm *zamki na lodzie*.

6 http://www.naleczow.com.pl/prus/oprusie/oprusie.php?id=1887breza&pr=indeks_p.

7 http://sjp.pwn.pl/slowniki/Buduje-zamki-na-lodzie.html.

gdzie biernikowe *przechlapane* raczej nie przechodzi na dopełniacz (por. *nie mieć przechlapane* vs ?**nie mieć przechlapanego*). A zatem w odpowiednim – tutaj bardzo uproszczonym[8] – schemacie dla czasownika MIEĆ podrzędnik oparty na imiesłowie biernym od PRZECHLAPAĆ będzie oznaczony jako `ppasp(acc)`, a nie `ppasp(str)`:

(29) MIEĆ: `subj{np(str)} + {xp(locat)}`
`+ {lex(ppasp(acc),sg,n,aff,'przechlapać',natr)}`

Z wyjątkiem ISJP wcześniejsze słowniki nie podawały informacji o otwartości argumentów zleksykalizowanych na przyjmowanie dalszych podrzędników, a w ISJP ta informacja jest bardzo ograniczona, np. do zasygnalizowania, że powinna wystąpić dodatkowa przydawka, ale bez informacji, którego elementu frazeologizmu przydawka ta jest podrzędnikiem. W Walentym takie informacje podane są jawnie, jak w poniższych trzech uproszczonych schematach dla frazeologizmów *ktoś/coś bije kogoś/coś na głowę*, *ktoś domaga się czyjejś głowy* oraz *ktoś wali głową w mur*:

(30) BIĆ: `subj{np(str)} + obj{np(str)}`
`+ {lex(prepnp(na,acc),sg,'głowa',natr)}`
(31) DOMAGAĆ SIĘ: `subj{np(str)}`
`+ {lex(np(gen),_,'głowa',ratr1({possp}))}`
(32) WALIĆ: `subj{np(str)}`
`+ {lex(np(inst),sg,'głowa',atr({adjp(agr)}))}`
`+ {lex(prepnp(w,acc),sg,'mur',atr)}`

W wypadku schematu dla BIĆ w (30), podobnie jak wcześniej w obu zleksykalizowanych frazach w schemacie (28), dalsza rozbudowa argumentów zleksykalizowanych jest niemożliwa (por. **ktoś bije kogoś na swoją/całą głowę*), co zasygnalizowane jest symbolem `natr`. W wypadku schematu (31) dla DOMAGAĆ SIĘ, w którym brakuje ograniczenia co do wartości liczby gramatycznej (por. *ktoś domaga się jego głowy / ich głów*), wymagane jest dokładnie jedno (`ratr1`) wystąpienie frazy dzierżawczej (`possp`; por. niefrazeologiczne *ktoś domaga się głowy*). Natomiast w wypadku schematu (32) dla WALIĆ forma leksemu GŁOWA może, lecz nie musi (`atr`) być rozbudowana o dalsze podrzędniki przymiotnikowe (`adjp`) uzgadniające się z tą formą (`agr`). Ponadto rozbudowana może być fraza *w mur*; brak w tym wypadku dalszych specyfikacji przy `atr` oznacza, że dodane mogą być dowolne podrzędniki zgodne z gramatyką języka polskiego, np. fraza nominalna w dopełniaczu, jak w następującym przykładzie z NKJP:

(33) Co rusz walisz głową w mur paragrafów.

8 Pełny schemat doprecyzowuje możliwe typy podrzędnika `xp(locat)` oraz pozwala na dalszą rozbudowę podrzędnika imiesłowowego o odpowiednie frazy przysłówkowe.

Jak wiadomo (Nunberg i in. 1994), frazeologizmy czasownikowe różnią się między innymi pod względem możliwości występowania w stronie biernej, ze zleksykalizowanym niepodmiotowym argumentem przechodzącym na pozycję podmiotu. Różnice te są w omawianym słowniku modelowane: w (28) argument nominalny zbudowany na formie leksemu ZAMEK oznaczony jest jako `obj`, tj. przechodzący na pozycję podmiotu w stronie biernej, jak w następującym przykładzie z NKJP:

(34) I z powodu tego trwającego dziesiątki tysięcy lat oszukiwania samego siebie budowane są zamki na lodzie.

Natomiast frazeologizm *chylić głowę wobec kogoś/czegoś* nie wydaje się występować w stronie biernej, stąd analogiczny zleksykalizowany argument nominalny nie jest oznaczony jako `obj` w odpowiednim schemacie czasownika CHYLIĆ:

(35) CHYLIĆ: `subj{np(str)} + {prepnp(wobec,gen)}`
`+ {lex(np(str),_,'głowa',atr({adjp(agr)}))}`

Spośród znanych nam słowników walencyjnych, nie tylko polskich, Walenty zdecydowanie najbardziej szczegółowo opisuje frazeologizmy, a formalizm, który do tego służy, ma zdecydowanie największą siłę wyrazu. Nie oznacza to jednak, że wszystkie aspekty takich frazeologizmów mogą być w tym formalizmie opisane; nie jest na przykład jasne, jak opisywać ustalony szyk argumentów zleksykalizowanych (np. we frazeologizmie *odsyłać od Annasza do Kajfasza*), ograniczenia paradygmatyczne dotyczące głównego czasownika (np. brak czasu przeszłego frazeologizmu *urwać komuś głowę*; Kosek 2013) czy też frazeologizmy, których centrum stanowi koordynacja form czasownikowych (np. *ktoś chucha i dmucha na coś/kogoś* lub *coś kogoś ani ziębi, ani grzeje*). Dyskusję na temat takich problemów – wraz z propozycją pewnych modyfikacji omawianego formalizmu – znaleźć można w artykule Przepiórkowski i in. 2017.

5. Warstwa semantyczna

Z wcześniejszych słowników walencyjnych wymienionych powyżej jedynie SSG koduje pewne informacje semantyczne, a mianowicie preferencje selekcyjne, zwane tam „cechami semantycznymi" (SSG 1: 9). Na przykład przy czasowniku APORTOWAĆ znajduje się informacja, że podmiot powinien oznaczać psa, a dopełnienie – zwierzę lub obiekt nieożywiony. Choć we wstępie znaleźć można definicje kilkunastu możliwych wartości takich cech semantycznych, np. [+Hum] (osobowość) czy [Liqu] (płyn), to w hasłach oprócz nich pojawiają się zupełnie dowolne wartości, np. (SSG 1: 16–17): [leki] (przy APLIKOWAĆ), [pies] (przy wspomnianym APORTOWAĆ), [utwór muzyczny] (przy ARANŻOWAĆ), [powierzchnie] (przy ASFALTOWAĆ), [kwoty pieniężne] (przy ASYGNOWAĆ) czy [subst. chem.] (przy ASYMILOWAĆ).

W Walentym takie informacje podane są w sposób bardziej precyzyjny, a mianowicie z odnośnikami do opracowanego na Politechnice Wrocławskiej słownika semantycznego

Słowosieć (Piasecki i in. 2009, 2016; Maziarz i in. 2015; http://plwordnet.pwr.wroc.pl/). Na przykład przy czasowniku APORTOWAĆ znajduje się informacja, że podmiot to zwykle PIES-2, a drugi argument to zwykle RZECZ-4, gdzie PIES-2 to jedno z siedmiu znaczeń jednostki PIES w Słowosieci, a mianowicie „pies domowy – popularne zwierzę domowe, przyjaciel człowieka" (w odróżnieniu choćby od ogólniejszego „ssaka z rodziny psowatych", a więc także lisa, czy też od „policjanta"), RZECZ-4 zaś to „przedmiot, jakiś nieożywiony obiekt istniejący w naturze bądź wytworzony przez człowieka" (w odróżnieniu np. od RZECZ-5: „przedmiot materialny stanowiący czyjąś własność")[9]. Dzięki wykorzystaniu powiązań semantycznych między jednostkami Słowosieci, w tym struktury hiperonimicznej, możliwe jest precyzyjne wyznaczenie zbioru jednostek spełniających dane preferencje selekcyjne, np. jako wszystkich hiponimów odpowiedniego znaczenia leksemu PIES (KUNDEL, SUKA, WYŻEŁ itd.) oraz ich nacechowanych lub bliskoznacznych odpowiedników (PSISKO, KUNDELEK, SOBAKA, CWAJNOS itd.).

Istotną nowością w stosunku do wcześniejszych polskich słowników walencyjnych jest powiązanie schematów składniowych z semantycznymi ramami walencyjnymi, w których argumenty semantyczne opisane są także za pomocą tzw. ról semantycznych – np. dwa argumenty czasownika APORTOWAĆ mają przypisane role: Initiator i Theme. Pełna rama semantyczna jest zatem listą par: ⟨rola semantyczna, preferencje selekcyjne⟩, np.:

(36) APORTOWAĆ: ⟨Initiator, PIES-2⟩ + ⟨Theme, RZECZ-4⟩

Czasownik APORTOWAĆ ma także zdefiniowane dwa składniowe schematy walencyjne (dla *aportować coś* i *aportować za czymś*), przy czym odpowiednie składniowe pozycje argumentowe są w słowniku skojarzone z odpowiednimi argumentami semantycznymi; w wypadku omawianego czasownika pełną informację składniowo-semantyczną oddaje następująca tabela, w której pierwszy wiersz zawiera role semantyczne, drugi wiersz – odpowiadające im preferencje selekcyjne, a kolejne wiersze przedstawiają schematy walencyjne z pozycjami argumentowymi przypisanymi do ról semantycznych:

Initiator	Theme
PIES-2	RZECZ-4
`subj{np(str)}`	`obj{np(str)}`
`subj{np(str)}`	`{prepnp(za,inst)}`

Opracowany na potrzeby Walentego zestaw ról semantycznych – opisany w artykule Hajnicz i in. 2016a oraz pełniej w przygotowywanym artykule: Andrzejczuk, Hajnicz 2016

9 Specyfikacja preferencji selekcyjnych drugiego argumentu aportować jako RZECZ-4 jest zapewne zbyt mało precyzyjna i może zostać w kolejnych wersjach słownika uszczegółowiona.

– przedstawiony jest w poniższej tabeli. Z jednej strony role dzielą się na *role podstawowe*, służące do opisu obligatoryjnych uczestników sytuacji, oraz *role uzupełniające*, które służą do reprezentacji charakteru okoliczności sytuacji. W ten sposób zaproponowane role mogą zostać wykorzystane w reprezentacji semantycznej całego wypowiedzenia, a nie tylko wymaganych argumentów predykatów. Z drugiej strony role zostały podzielone na trzy grupy (inicjującą, towarzyszącą i kończącą), które określają, na jakim etapie akcji wskazywanej przez predykat pojawiają się uczestnicy akcji lub towarzyszące jej okoliczności. Zestaw ten został zainspirowany między innymi rolami wykorzystywanymi w słowniku VerbNet[10].

	GRUPA INICJUJĄCA	GRUPA TOWARZYSZĄCA	GRUPA ZAMYKAJĄCA
PODSTAWOWE	Initiator Stimulus	Theme Experiencer Factor Instrument	Recipient Result
UZUPEŁNIAJĄCE	Condition	Attribute Manner Measure Location Path Time Duration	Purpose
ATRYBUTY	Source	Foreground Background	Goal

Zupełnie nową koncepcją wprowadzoną w Walentym jest natomiast dwupoziomowa reprezentacja ról semantycznych, polegająca na możliwości opatrywania ich atrybutami, gdy dwa różne argumenty zdają się pełnić tę samą funkcję, jak w zdaniach typu *Janek zamienił się z Marysią jabłkiem na gruszkę*, gdzie podrzędniki *Janek* i *Marysia* powinny być opatrzone rolą Initiator, a *jabłko* i *gruszka* – rolą Theme. W wypadku czasowników tego typu jeden z argumentów z pary często odczuwany jest jako pierwszoplanowy (tutaj: Initiator *Janek* i Theme *jabłko*), a drugi – jako pozostający w tle (odpowiednio: *Marysia* i *gruszka*). Pierwszoplanowym realizacjom danej roli przypisywany jest wtedy atrybut Foreground, a drugoplanowym – Background. A zatem przypisanie ról semantycznych do odpowiednich pozycji składniowych będzie w wypadku czasownika ZAMIENIĆ SIĘ wyglądać następująco (pomijamy tutaj preferencje selekcyjne):

10 http://verbs.colorado.edu/~mpalmer/projects/verbnet.html.

Initiator		Theme	
Foreground	Background	Foreground	Background
subj{np(str)}	{prepnp(z,inst)}	{np(inst)}	{prepnp(na,acc)}

Różne schematy walencyjne danego leksemu mogą realizować różne podzbiory argumentów semantycznych. Ilustruje to następująca tabela dla czasownika CZESAĆ (SIĘ), w której pierwszy schemat odpowiada przykładowi *Janek czesze Marysi włosy szczotką*, a drugi, ze zwrotnym SIĘ, odpowiada zdaniu *Janek czesze się szczotką* (z niezrealizowanym argumentem Theme Foreground). Ponadto tylko trzeci schemat, z SIĘ wzajemnościowym, realizuje zarówno rolę Initiator Foreground, jak i rolę Initiator Background, jak w przykładzie: *Ania miała zwyczaj czesać się nawzajem ze swoją przyjaciółką jedną szczotką*. Ostatni (siódmy) wiersz tabeli, odpowiadający zdaniu *Janek czesze Marysię szczotką*, ilustruje natomiast zjawisko autoalternacji, w którym pozycje jednego schematu walencyjnego mogą być przypisane do różnych ról semantycznych: jest to bowiem ten sam schemat, który występuje w tabeli jako pierwszy (tj. w czwartym wierszu), lecz z niezrealizowaną pozycją {np(dat)} i z pozycją obj{np(str)} odpowiadającą roli Theme Background (*Marysia*), a nie Theme Foreground (*włosy*).

Initiator		Theme		Instrument
Foreground	Background	Foreground	Background	
LUDZIE	LUDZIE	WŁOSY-1	ISTOTY	GRZEBIEŃ-1 SZCZOTKA-1
subj{np(str)}		obj{np(str)}	{np(dat)}	{np(inst)}
subj{np(str)}			{refl}	{np(inst)}
subj{np(str)}	{prepnp(z,inst)}		{recip}	{np(inst)}
subj{np(str)}			obj{np(str)}	{np(inst)}

Powyższa tabela ilustruje także fakt, że preferencje selekcyjne nie muszą wskazywać na jedną tylko jednostkę Słowosieci; rola Instrument jest ograniczona do zbioru dwóch takich jednostek: GRZEBIEŃ-1 i SZCZOTKA-1. Pewne takie zbiory są wykorzystywane w wielu ramach walencyjnych, tak jest np. ze zbiorem składającym się z jednostek: OSOBA-1 i GRUPA LUDZI-1, czy też ze zbiorem składającym się z jednostek: OSOBA-1, ISTOTA ŻYWA-1 i GRUPA ISTOT-1. Dla zwiększenia czytelności słownika i ułatwienia pracy leksykografom zbiory takie zostały nazwane, w tym wypadku: LUDZIE i ISTOTY (por. [+Hum] i [+Anim] w SSG), co także ilustruje powyższa tabela.

Drugą poza Foreground i Background parą atrybutów jest Source i Goal, jak w następujących danych walencyjnych czasownika ABSORBOWAĆ (por. *Organizm absorbuje przez skórę wilgoć z powietrza*)[11]:

Theme		Path	Location
Goal	Source		Source
CIAŁO-3 SUBSTANCJA-1	SUBSTANCJA-1	POWIERZCHNIA-2 meronimia(Theme Goal)	SUBSTANCJA-1
`subj{np(str)}`	`obj{np(str)}`	`{prepnp(przez,acc)}`	`{xp(abl)}`

Powyższa tabela ilustruje także jeszcze jedną możliwość określenia preferencji selekcyjnych, a mianowicie za pomocą relacji leksykalno-semantycznej pomiędzy danym i pewnym innym argumentem opisywanego predykatu. W tym wypadku argument o roli Path może wyrażać „zewnętrzną stronę czegoś" (tak zdefiniowana jest w Słowosieci POWIERZCHNIA-2), lecz także część organizmu wyrażonego za pomocą argumentu o roli Theme Goal (zob. „meronimia(Theme Goal)").

Należy podkreślić, że semantyczne ramy walencyjne i ich powiązania ze schematami składniowymi określone są dla konkretnych znaczeń lematów zdefiniowanych w Słowosieci. W szczególności informacje podane w dwóch powyższych tabelach odpowiadają znaczeniom: CZESAĆ-2, czyli przygładzać włosy lub sierść grzebieniem lub szczotką, a nie np. CZESAĆ-3, czyli oczyszczać włókno z resztek paździerzy, oraz ABSORBOWAĆ-2, czyli pochłaniać w wyniku kontaktu fizycznego, a nie np. ABSORBOWAĆ-1 (czasownik myślenia).

Wróćmy na zakończenie tego punktu do schematów frazeologicznych. Opisane w nich argumenty zleksykalizowane mogą być dwojakiego rodzaju: semantyczne, będące realizacją pewnej roli semantycznej, lub asemantyczne, nieodpowiadające żadnej roli. Pierwszą sytuację ilustruje frazeologizm *witać z otwartymi ramionami*, gdzie zleksykalizowana fraza *z otwartymi ramionami* jest jedną z możliwych realizacji roli semantycznej Manner. Drugą ilustruje wyrażenie *wywinąć orła*, które ma tylko jeden argument semantyczny, Theme, lecz dwa wymagane podrzędniki składniowe: podmiot odpowiadający argumentowi semantycznemu i podrzędnik zleksykalizowany *orła*:

Theme	
ISTOTY	
`subj{np(str)}`	`{lex(np(str),_,'orzeł',atr({adjp(agr)}))}`

11 W wypadku pewnych ról, w tym oznaczających lokalizacje w przestrzeni (Location), atrybuty takie nie muszą występować w parach.

Przykład ten pokazuje, że w Walentym frazeologizmy są na poziomie semantycznym rozumiane tradycyjnie jako pewne całości (tutaj: *wywinąć orła*)[12], które mogą otwierać miejsca walencyjne dla argumentów semantycznych (tutaj: argument Theme realizowany jako podmiot). Mniej tradycyjny – ale dobrze umotywowany – jest natomiast opis składniowy schematów frazeologicznych, w którym to sam czasownik jest podstawową jednostką otwierającą składniowe pozycje walencyjne dla swoich podrzędników: zarówno tych będących częścią frazeologizmu (tutaj: *orła*), jak i tych wymaganych przez frazeologizm jako całość. A zatem w prezentowanym tu ujęciu elementy składni wewnętrznej i zewnętrznej frazeologizmów (Lewicki 1976, 1986) traktowane są łącznie.

6. Dostępność, wykorzystanie i dalszy rozwój

W momencie oddawania niniejszego artykułu do druku opisem semantycznym objętych jest niecałe 55 proc. jednostek słownika, więc – jako niepełna – warstwa ta jest obecnie publicznie dostępna w sposób ograniczony[13]. W pełni dostępna jest natomiast warstwa składniowa słownika wraz z informacjami frazeologicznymi (Hajnicz i in. 2015; Nitoń i in. 2015). Wersja słownika odzwierciedlająca zawsze aktualny stan prac znajduje się pod adresem http://walenty.ipipan.waw.pl/ – zarówno w postaci do przeglądania i przeszukiwania haseł za pomocą dowolnej przeglądarki internetowej, jak i w uaktualnianej co tydzień postaci tekstowej do pobrania. Ponadto co jakiś czas na oficjalnej stronie Walentego, http://zil.ipipan.waw.pl/Walenty, udostępniane są wersje: XML (do dalszego automatycznego przetwarzania przez komputer) oraz PDF (do czytania przez człowieka), przy czym ostatnia dostępna wersja PDF, z kwietnia 2016 roku, liczy ok. 44 300 stron i zajmuje ok. 90 MB miejsca na dysku. Wszystkie postaci słownika dostępne są za darmo na licencji swobodnej (Creative Commons BY-SA 4.0, zgodnej także z GNU GPLv3).

W odróżnieniu od ISJP, który jest słownikiem z założenia popularnym (choć opartym na metodach naukowych), Walenty – podobnie jak SSG – jest słownikiem przede wszystkim naukowym, ale przeznaczonym także do wykorzystania przez komputerowe systemy automatycznego przetwarzania języka polskiego. Omawiany słownik jest obecnie źródłem danych dla dwóch parserów (analizatorów składniowych) opracowanych w Instytucie Podstaw Informatyki PAN: POLFIE (Patejuk, Przepiórkowski 2014, 2015; Patejuk 2016) oraz Świgra (Woliński 2015).

Od lipca 2016 roku realizowany jest kolejny dwuletni projekt związany z europejską inicjatywą CLARIN ERIC, w którego ramach słownik jest dalej rozbudowywany. W planach jest między innymi uzupełnienie warstwy semantycznej, dodanie haseł odpowiadających kolejnym 3,5 tysiącom lematów oraz opracowanie powiązań derywacyjnych pomiędzy hasłami.

12 Tak też rozumiane są frazeologizmy w Słowosieci, gdzie jednostką odpowiadającą powyższej ramie jest WYWINĄĆ ORŁA-1.

13 Jedynie za pomocą wspomnianego poniżej interfejsu sieciowego, a ponadto – obecnie tylko w wersji z końca kwietnia 2016 r. – do pobrania w formacie XML z repozytorium projektu CLARIN-PL, pod adresem http://hdl.handle.net/11321/251.

7. Podsumowanie

Prace Zespołu Inżynierii Lingwistycznej IPI PAN sytuują się na granicy językoznawstwa teoretycznego, lingwistyki korpusowej, leksykografii elektronicznej i – oczywiście – przetwarzania języka naturalnego. W drugiej połowie pierwszej dekady XXI wieku zespół podejmował prace nad automatycznym wydobywaniem informacji walencyjnych z jedynie morfoskładniowo znakowanych korpusów (Przepiórkowski, Fast 2005; Dębowski 2009; Przepiórkowski 2009; Hajnicz 2011). Prace doprowadziły do powstania nowych algorytmów i narzędzi, lecz wydobywane informacje walencyjne obarczone były zbyt dużym błędem, by stać się podstawą gruntownego słownika walencyjnego wysokiej jakości. Dlatego też od 2011 roku, najpierw w ramach europejskiego projektu CESAR[14], a następnie przede wszystkim w ramach infrastruktury europejskiej CLARIN[15], prowadzone były prace nad wspomaganym komputerowo procesem ręcznego – przez współpracujących z zespołem leksykografów i lingwistów – tworzenia słownika walencyjnego, których rezultatem jest obecna postać Walentego.

Prezentowany słownik jest obecnie najbogatszym słownikiem walencyjnym języka polskiego i jedynym, który w pełni integruje informacje składniowe, frazeologiczne i semantyczne[16]. Wydaje się, że także dla innych języków niewiele jest słowników, które dorównywałyby Walentemu zarówno pod względem zakresu empirycznego i liczby objętych spójnym opisem poziomów reprezentacji, jak i pod względem precyzji oraz szczegółowości opisu czy też dostępności. Żywimy nadzieję, że słownik ten zostanie dostrzeżony przez polskie środowisko lingwistyczne i leksykograficzne.

Podziękowania

Za komentarze do wcześniejszej wersji artykułu dziękujemy Mirosławowi Bańce, Jadwidze Linde-Usiekniewicz, Annie Pajdzińskiej oraz dwóm anonimowym recenzentom „Języka Polskiego".

Pracami nad słownikiem kieruje Elżbieta Hajnicz, do której należy przesyłać ewentualną korespondencję związaną z niniejszym artykułem. Na wczesnym etapie rozwoju słownika pracami kierował Adam Przepiórkowski, który jest też głównym autorem treści niniejszego artykułu. Koncepcja warstwy składniowej Walentego została opracowana przede wszystkim przez Agnieszkę Patejuk, Adama Przepiórkowskiego, Marcina Wolińskiego i Elżbietę Hajnicz, z kluczowym udziałem Filipa Skwarskiego w początkowej fazie prac. Autorkami koncepcji warstwy semantycznej słownika są Anna Andrzejczuk i Elżbieta Hajnicz. Pierwszy projekt warstwy frazeologicznej zaproponowany został przez Adama Przepiórkowskiego na podstawie dyskusji z Marcinem Wolińskim, Agnieszką Patejuk i Elżbietą Hajnicz; aktualny format

14 http://clip.ipipan.waw.pl/CESAR.

15 http://clip.ipipan.waw.pl/CLARIN-PL.

16 Wspomnieć jednak należy o zrealizowanym na polonistyce UW projekcie RAMKI (Derwojedowa i in. 2010), który doprowadził do powstania niewielkiego składniowo-semantycznego zasobu zgodnego z metodologią projektu FrameNet (Fillmore i in. 2003).

tej warstwy, znacznie rozbudowany względem pierwotnego, ukształtował się w wyniku dalszych dyskusji w tym gronie.

Praca finansowana w ramach wkładu krajowego na rzecz udziału we wspólnym międzynarodowym przedsięwzięciu „CLARIN ERIC: Wspólne zasoby językowe i infrastruktura technologiczna".

Bibliografia

Andrzejczuk A., Hajnicz E. 2016: *Poziom semantyczny słownika walencyjnego* Walenty, w przygotowaniu.

Bąba S., Liberek J. 2003: *Słownik frazeologiczny współczesnej polszczyzny*, Wydawnictwo Naukowe PWN, Warszawa.

Derwojedowa M., Linde-Usiekniewicz J., Zawisławska M. 2010: *Projekt RAMKI: rygorystyczna aplikacja metodologii kognitywno-interpretacyjnej (ram interpretacyjnych) do opisu polszczyzny*, [w:] M. Zawisławska (red.), *Rygorystyczna aplikacja metodologii kognitywno interpretacyjnej*, Elipsa, Warszawa, s. 7–16.

Dębowski Ł. 2009: *Valence extraction using the EM selection and co-occurrence matrices*, „Language Resources and Evaluation" 43, s. 301–327.

Fillmore C.J., Johnson C.R., Petruck M.R. 2003: *Background to FrameNet*, „International Journal of Lexicography" 16 (3), s. 235–250.

Hajič J., Panevová J., Urešová Z., Bémová A., Kolářová V., Pajas P. 2003: *PDT-VALLEX. Creating a large-coverage valency lexicon for treebank annotation*, [w:] J. Nivre, E. Hinrichs (red.), *Proceedings of the Second Workshop on Treebanks and Linguistic Theories (TLT2003)*, Växjö, Sweden, s. 57–68.

Hajnicz E. 2011: *Automatyczne tworzenie semantycznego słownika walencyjnego*, Problemy Współczesnej Nauki. Teoria i Zastosowania. Inżynieria Lingwistyczna, Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Hajnicz E., Andrzejczuk A., Bartosiak T. 2016a: *Semantic layer of the valence dictionary of Polish* Walenty, [w:] N. Calzolari, K. Choukri, T. Declerck, M. Grobelnik, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC 2016*, ELRA, European Language Resources Association (ELRA), Portorož, Slovenia, s. 2625–2632.

Hajnicz E., Nitoń B., Patejuk A., Przepiórkowski A., Woliński M. 2015: *Internetowy słownik walencyjny języka polskiego oparty na danych korpusowych*, „Prace Filologiczne" LXV, s. 95–110.

Hajnicz E., Patejuk A., Przepiórkowski A., Woliński M. 2016b: *Walenty: słownik walencyjny języka polskiego z bogatym komponentem frazeologicznym*, [w:] K. Skwarska, E. Kaczmarska (red.), *Výzkum slovesné valence ve slovanských zemích*, Slovanský ústav AV ČR, Praga.

Kallas K. 1993: *Składnia współczesnych polskich konstrukcji współrzędnych*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.

Kosek I. 2013: *Paradygmaty zwrotów frazeologicznych – problemy opisu leksykograficznego*, [w:] G. Dziamska-Lenart, J. Liberek (red.), *Perspektywy współczesnej frazeologii polskiej. Między teorią a praktyką leksykograficzną*, Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, Poznań, s. 51–61.

Landau I. 2013: *Control in Generative Grammar. A Research Companion*, Cambridge University Press, Cambridge.

Lewicki A.M. 1976: *Wprowadzenie do frazeologii syntaktycznej. Teoria zwrotu frazeologicznego*, Wydawnictwo Uniwersytetu Śląskiego, Katowice.

Lewicki A.M. 1986: *Składnia związków frazeologicznych*, „Biuletyn Polskiego Towarzystwa Językoznawczego" XL, s. 75–83.

Lopatková M. 2003: *Valency in the Prague Dependency Treebank. Building the valency lexicon*, „The Prague Bulletin of Mathematical Linguistics" 79–80, s. 37–60.

Maziarz M., Piasecki M., Rudnicka E. 2015: *Słowosieć – polski wordnet. Proces tworzenia tezaurusa*, „Polonica" XXXIV, s. 79–97.

Nitoń B., Bartosiak T., Hajnicz E., Patejuk A., Przepiórkowski A., Woliński M. 2015: *DEMO: Access to a valence dictionary of Polish Walenty via Internet browser*, [w:] Z. Vetulani, J. Mariani (red.), *Proceedings of the 7th Language & Technology Conference. Human Language Technologies as a Challenge for Computer Science and Linguistics*, Wydawnictwo Poznańskie, Fundacja Uniwersytetu im. Adama Mickiewicza, Poznań, s. 270–272.

Nunberg G., Sag I.A., Wasow T. 1994: *Idioms*, „Language" 70 (3), s. 491–538.

Patejuk A. 2015: *Unlike coordination in Polish. An LFG account*, rozprawa doktorska, Instytut Języka Polskiego PAN, Kraków.

Patejuk A. 2016: *Integrating a rich external valency dictionary with an implemented XLE/LFG grammar*, [w:] D. Arnold, M. Butt, B. Crysmann, T.H. King, S. Müller (red.), *The Proceedings of the HeadLex16 Conference*, CSLI Publications, Stanford, CA.

Patejuk A., Przepiórkowski A. 2014: *Synergistic development of grammatical resources. A valence dictionary, an LFG grammar, and an LFG structure bank for Polish*, [w:] V. Henrich, E. Hinrichs, D. de Kok, P. Osenova, A. Przepiórkowski (red.), *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13)*, Department of Linguistics (SfS), University of Tübingen, Tübingen, s. 113–126.

Patejuk A., Przepiórkowski A. 2015: *Parallel development of linguistic resources. Towards a structure bank of Polish*, „Prace Filologiczne" LXV, s. 255–270.

Piasecki M., Szpakowicz S., Broda B. 2009: *A Wordnet from the ground up*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.

Piasecki M., Szpakowicz S., Maziarz M., Rudnicka E. 2016: *plWordNet 3.0 – almost there*, [w:] V.B. Mititelu, C. Forăscu, C. Fellbaum, P. Vossen (red.), *Proceedings of the 8th Global Wordnet Conference*, Global Wordnet Association, s. 290–299.

Przepiórkowski A. 1997: *Transmisja wymagań składniowych*, „Polonica" XVIII, s. 29–50.

Przepiórkowski A. 1999: *Case Assignment and the Complement-Adjunct Dichotomy. A Non-Configurational Constraint--Based Approach*, rozprawa doktorska, Universität Tübingen, Tübingen.

Przepiórkowski A. 2004: *O wartości przypadka podmiotów liczebnikowych*, „Biuletyn Polskiego Towarzystwa Językoznawczego" LX, s. 133–143.

Przepiórkowski A. 2009: *Towards the automatic acquisition of a valence dictionary for Polish*, [w:] M. Marciniak, A. Mykowiecka (red.), *Aspects of natural language processing. Essays dedicated to Leonard Bolc on the occasion of his 75th Birthday*, Lecture Notes in Computer Science, vol. 5070, Springer, Berlin, s. 191–210.

Przepiórkowski A. 2016a: *Against the argument-adjunct distinction in Functional Generative Description*, „The Prague Bulletin of Mathematical Linguistics" 106, s. 5–20.

Przepiórkowski A. 2016b: *How* not *to distinguish arguments from adjuncts in LFG*, [w:] D. Arnold, M. Butt, B. Crysmann, T.H. King, S. Müller (red.), *The Proceedings of the HeadLex16 Conference*, CSLI Publications, Stanford, CA.

Przepiórkowski A. 2017: *On the argument-adjunct distinction in the Polish semantic syntax tradition*, „Cognitive Studies / Études Cognitives" 17, przyjęte do publikacji.

Przepiórkowski A., Bańko M., Górski R.L., Lewandowska-Tomaszczyk B. (red.) 2012: *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa.

Przepiórkowski A., Fast J. 2005: *Baseline experiments in the extraction of Polish valence frames*, [w:] M.A. Kłopotek, S.T. Wierzchoń, K. Trojanowski (red.), *Intelligent Information Processing and Web Mining*, Advances in Soft Computing, Springer, Berlin, s. 511–520.

Przepiórkowski A., Hajič J., Hajnicz E., Urešová Z. 2017: *Phraseology in two Slavic valency dictionaries. Limitations and perspectives*, „International Journal of Lexicography" 30 (1), s. 1–38.

Przepiórkowski A., Patejuk A. 2014: *Koordynacja leksykalno-semantyczna w systemie współczesnej polszczyzny (na materiale Narodowego Korpusu Języka Polskiego)*, „Język Polski" XCIV, s. 104–115.

Przepiórkowski A., Skwarski F., Hajnicz E., Patejuk A., Świdziński M., Woliński M. 2014: *Modelowanie własności składniowych czasowników w nowym słowniku walencyjnym języka polskiego*, „Polonica" XXXIII, s. 159–178.

Saloni Z., Woliński M., Wołosz R., Gruszczyński W., Skowrońska D. 2012: *Słownik gramatyczny języka polskiego*, wyd. 2, Warszawa.

Sannikow W.Z. 1979: *Soczitielnyje i srawnitielnyje konstrukcii: ich blizost', ich sintaksiczeskoje priedstawlenije I*, „Wiener Slawistischer Almanach" 4, s. 413–432.

Sannikow W.Z. 1979: *Soczitielnyje i srawnitielnyje konstrukcii: ich blizost', ich sintaksiczeskoje priedstawlenije II*, „Wiener Slawistischer Almanach" 5, s. 211–242.

Sgall P., Hajičová E., Panevová J. 1986: *The Meaning of the sentence in its semantic and pragmatic aspects*, Reidel, Dordrecht.

Szupryczyńska M. 1996: *Problem pozycji składniowej*, [w:] K. Kallas (red.), *Polonistyka toruńska Uniwersytetowi w 50. rocznicę utworzenia UMK. Językoznawstwo*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń, s. 135–144.

Świdziński M. 1992: *Realizacje zdaniowe podmiotu-mianownika, czyli o strukturalnych ograniczeniach selekcyjnych*, [w:] A. Markowski (red.), *Opisać słowa*, Elipsa, Warszawa, s. 188–201.

Świdziński M. 1993: *Dalsze kłopoty z bezokolicznikiem*, [w:] J. Sambor, J. Linde-Usiekniewicz, R. Huszcza (red.), *Językoznawstwo synchroniczne i diachroniczne*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa, s. 303–314.

Świdziński M. 1998: *Syntactic dictionary of Polish verbs. Version 3a*, maszynopis, Uniwersytet Warszawski, Warszawa.

Świdziński M. 2015: *O frazie nominalnej liczebnikowej w pozycji podmiotu po raz kolejny*, „LingVaria" 19, s. 81–97.

Woliński M. 2015: *Deploying the new valency dictionary Walenty in a* DCG *parser of Polish*, [w:] M. Dickinson, E. Hinrichs, A. Patejuk, A. Przepiórkowski (red.), *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT14)*, Instytut Podstaw Informatyki PAN, Warszawa, s. 221–229.

Żmigrodzki P. 2015: Wielki słownik języka polskiego PAN – *historia, stan obecny i perspektywy rozwoju po 2018 roku*, „Biuletyn Polskiego Towarzystwa Językoznawczego" LXVII, s. 177–178.

Rozwiązanie skrótów

ISJP: *Inny słownik języka polskiego*, red. M. Bańko, t. 1–2, Wydawnictwo Naukowe PWN, Warszawa 2000. – SSG: *Słownik syntaktyczno-generatywny czasowników polskich*, red. K. Polański, t. 1–5, Zakład Narodowy im. Ossolińskich, Wydawnictwo PAN, Wrocław-Warszawa-Kraków-Gdańsk 1980–1992. – WSJP PAN: *Wielki słownik języka polskiego PAN*, red. P. Żmigrodzki (online: http://www.wsjp.pl).

Summary

Walenty: a thorough syntactico-semantic valency dictionary of Polish

Keywords: Walenty, valency dictionary.

The paper presents a new valency dictionary of Polish, developed by the Linguistic Engineering Group at the Polish Academy of Sciences. The dictionary, called Walenty, while still under development, is already (as of late 2016) the biggest and most detailed valency dictionary of Polish, the only one which fully integrates syntax, semantics (not only selectional preferences, but also semantic roles) and phraseology. The size and the degree of precision of Walenty justify the use of the adjective 'thorough' (Pol. 'gruntowny') in the title of this paper. The dictionary is freely available – searchable via any browser and downloadable in a variety of formats.