

ELŻBIETA HAJNICZ  
BARTŁOMIEJ NITON  
AGNIESZKA PATEJUK  
ADAM PRZEPIÓRKOWSKI  
MARCIN WOLIŃSKI  
Instytut Podstaw Informatyki  
Polskiej Akademii Nauk  
ul. Jana Kazimierza 5  
01-248 Warszawa  
tel. +48 223 800 500  
hajnicz@ipipan.waw.pl  
aep@ipipan.waw.pl  
adamp@ipipan.waw.pl  
wolinski@ipipan.waw.pl  
bartek.niton@gmail.com

## INTERNETOWY SŁOWNIK WALENCYJNY JĘZYKA POLSKIEGO OPARTY NA DANYCH KORPUSOWYCH

---

SŁOWA KLUCZOWE: schematy walencyjne, słownik elektroniczny, polszczyzna

KEYWORDS: valence schemata, electronic dictionary, Polish

---

### 1. Wprowadzenie

Słowniki walencyjne opisują sposób, w jaki określone jednostki językowe wiążą się ze swoimi podrzędnikami. Taka też jest funkcja *Walentego*, nowego słownika walencyjnego predykatów polskich.

Koncepcja *Walentego* zakłada rygorystyczne przestrzeganie pewnych założeń, przyjętych już podczas wstępnych prac projektowych. Po pierwsze, hasła słownika mają ściśle zdefiniowaną strukturę formalną, która umożliwia jego wykorzystywanie przez automatyczne analizatory składniowe języka (w tym wypadku polskiego), tak zwane *parsery*. W szczególności słownik jest wykorzystywany przez

dwa parsery obecnie rozwijane w Instytucie Podstaw Informatyki Polskiej Akademii Nauk (IPI PAN): oparty na gramatyce Świdzińskiego (1992) parser *Świ-gra* (Woliński 2004, Świdziński i Woliński 2010) oraz oparty na formalizmie LFG parser *POLFIE* (Patejuk i Przepiórkowski 2012). Jednocześnie jednak słownik ten ma mieć postać czytelną dla użytkownika. Po drugie, reprezentowane zjawiska (składniowe i semantyczne) mają mieć udokumentowane pokrycie w rzeczywistości językowej. Jako punkt odniesienia przyjęliśmy *Narodowy Korpus Języka Polskiego* (NKJP; Przepiórkowski i in. 2012). Wszystkie inne źródła, w tym Internet i literatura językoznawcza, miały charakter pomocniczy. Zależało nam przy tym na jak najpełniejszej reprezentacji zjawisk językowych mających charakter walencyjny, w tym frazeologizmów oraz zjawisk nietypowych, takich jak koordynacja fraz o odmiennym charakterze składniowym (Sag i in. 1985, Kallas 1993). Po trzecie wreszcie, struktura słownika ma w sposób elastyczny umożliwiać różne jego zastosowania. Dotyczy to zarówno rozmaitych formatów, w jakich słownik jest udostępniany (format tekstowy, XML-owy, PDF), jak i możliwości tworzenia podsłowników (np. frazeologicznego).

Powyższych założeń nie udało się zrealizować, gdyby *Walenty* tworzony był sposobem tradycyjnym. Omawiany słownik ma jednak postać bazy danych (o dość skomplikowanej strukturze wewnętrznej), obsługiwanej przez dedykowane narzędzie o nazwie *Slowal*, służące do opracowywania, przeglądania i przetwarzania danych słownikowych, co zapewnia niezbędną elastyczność.

Praca nad słownikiem realizowana jest zdalnie, za pośrednictwem przeglądarki internetowej. Hasła *Walentego* opracowywane są przez zespół wykwalifikowanych leksykografów, mających sprecyzowane uprawnienia dotyczące zakresu edycji haseł. Natomiast przeglądanie haseł słownikowych bez możliwości dokonywania jakichkolwiek zmian czy wpisów dostępne jest dla każdego, kto wejdzie na stronę <http://walenty.ipipan.waw.pl/>.

Nie zamierzamy tutaj po raz kolejny szczegółowo opisywać formalizmu *Walentego*. Zainteresowanego Czytelnika odsyłamy do pracy Przepiórkowski i in. 2014c. Celem niniejszego artykułu jest prezentacja różnych formatów zapisu słownika wraz ze sposobem dostępu do niego przez Internet. Poniższe cztery punkty – *Podstawowa struktura słownika*, *Widok haseł*, *Filtrowanie haseł* oraz *Rozwinięcia typów fraz* – omawiają postać internetową słownika, zawierającą aktualny roboczy stan *Walentego*. Na stronie <http://zil.ipipan.waw.pl/Walenty> znaleźć można także „migawki” z różnych etapów tworzenia słownika – obecnie są to wersje *Walentego* ze stycznia i września 2013, marca i września 2014 oraz maja 2015. Dostępne są one w formatach opisanych w punktach *Postać tekstowa słownika* oraz *Słownik w formacie PDF*.

## 2. Podstawowa struktura słownika

*Walenty* zawiera informację walencyjną dla czasowników oraz (na razie nie-licznych) rzeczowników, przymiotników i przysłówków. Tak więc *de facto* składa się z czterech podsłowników. Pojedyncze hasło słownikowe jest identyfikowane przez lemat (np. AFISZOWAĆ, AFIRMACJA, BAĆ, CHĘTNY, OMIESZKAĆ, SPIESZYĆ). Lemat nie zawiera cząstki *się* nawet wówczas, gdy jest dla czasownika obowiązkowa (np. BAĆ). Na tym poziomie nie rozróżniamy też aspektu (np. dla dwuaspektowego SPIESZYĆ). Rozróżnienie dotyczące występowania z cząstką *się*, zanegowania oraz aspektu występuje na poziomie podhasła. W wypadku czasowników takich jak BAĆ SIĘ czy SPIESZYĆ SIĘ, cząstka *się* jest częścią lematu podhasła; następnie występuje informacja o zanegowaniu oraz aspekcie. Tak więc hasło BAĆ ma jedno podhasło: **bać się** (┌, imperf). Podkreślenie '┌' oznacza, że wartość zanegowania jest dowolna, tj. że schematy walencyjne dla tego podhasła nie zależą od tego, czy predykat jest zanegowany, czy nie. Inaczej jest w wypadku czasownika OMIESZKAĆ, które ze schematami podanymi w podhasle **omieszkąć (neg, perf)** występuje tylko w postaci zanegowanej (np. w następującym zdaniu z NKJP: *Erazm nie omieszkął zapewnić profesorskiego gremium paryskiej uczelni*). Wartości **imperf** i **perf** oznaczają natomiast aspekt, odpowiednio: niedokonany i dokonany. Oczywiście jednemu lematowi może odpowiadać większa liczba podhasła – tak jest w wypadku czasownika SPIESZYĆ, który ma aż trzy podhasła: **spieszyć się** (┌, imperf), **spieszyć** (┌, imperf) oraz **spieszyć** (┌, perf).

Każde podhasło stanowi listę schematów składniowych, które mogą być realizowane przez dany predykat (przy określonym aspekcie, zanegowaniu itd.). Każdy schemat to z kolei lista pozycji składniowych, które mogą być razem wymagane przez opisywany predykat. Reprezentowane są wyłącznie schematy o największej liczbie pozycji mogących współwystępować w jednym wypowiedzeniu.

Dwie pozycje są traktowane w sposób szczególny: pozycja podmiotu (subj) oraz pozycja dopełnienia, które podlega pasywizacji (obj). Kolejną informacją uwidacznianą na poziomie pozycji jest szeroko rozumiana kontrola składniowa (Rosenbaum 1967, Landau 2013), gdzie pozycja kontrolująca oznaczona jest jako controller, zaś kontrolowana jako controllee. Ilustruje to rysunek 1 dla jedyne podhasła czasownika OMIESZKAĆ: schemat walencyjny składa się tu z dwóch pozycji, z których pierwsza jest pozycją kontrolującego podmiotu (*Erazm* w cytowanym powyżej zdaniu z NKJP), a druga zawiera kontrolowaną frazę bezokolicznikową (*zapewnić profesorskiego gremium...*). Semantyczna istota kontroli, mająca także swoje objawy składniowe, związane z uzgodnieniem (Przepiórkowski i Rosen 2005), polega na tym, że niewyrażony lokalnie podmiot frazy kontrolowanej (*zapewnić...*) musi być rozumiany jako tożsamy z podmiotem (*Erazm*) opisywanego czasownika (*nie omieszkął*).

Funkcja:	subj.controller	controllee
Typy fraz:	np(str)	infp(␣)

Rys. 1. Schemat walencyjny dla omieszkać (neg, perf)

Każda pozycja reprezentowana jest jako lista dopuszczalnych jej realizacji, tj. specyfikacji kształtów morfoskładniowych podrzędników. Po pierwsze określamy więc, jakiego typu fraza może realizować daną pozycję (nominalna np, przymkowo-nominalna prepnp, bezokolicznikowa infp itp.). Po drugie wskazujemy ograniczenia na morfoskładniową postać frazy, np. przypadek fraz nominalnych i przymiotnikowych (adjp(inst)), przyimek i przypadek dla fraz przymykowych (prepnp(do,gen)) itd.

Warto zauważyć, że symbol podkreślenia ‘\_’ oznacza dowolną wartość kategorii gramatycznej, np. aspektu w infp(␣). Na rysunku 1 tak właśnie opisany jest typ frazy realizujący drugą pozycję, gdyż bezokolicznikowy podrzędnik czasownika OMIESZKAĆ może być zarówno dokonany (np. *zapewnić...* w powyższym przykładzie), jak i niedokonany (*Władysław... nie omieszkał płatać figłów...*; NKJP). Na pierwszej pozycji pojawia się natomiast str jako wartość przypadku. Wartość ta oznacza przypadek strukturalny, tj. taki, którego postać morfologiczna zależy od otoczenia składniowego i formy czasownika. Na przykład wartość przypadku strukturalnego w wypadku frazy, która nie znajduje się na pozycji podmiotu, zwykle oznacza biernik (np. w *Janek czyta książkę*), ale w wypadku zanegowania czasownika lub jego wystąpienia w formie gerundium oznacza ona dopełniacz (np. w *Janek nie czyta książki czy czytanie książki*). Także zwykle mianownikowe podmioty mogą występować w przygerundialnym dopełniaczu, stąd między innymi określenie podmiotu na rysunku 1 jako np(str) (choć akurat w wypadku tego czasownika trudno jest znaleźć takie wystąpienia w formie gerundium).

Pełniejszy i ciekawszy przykład schematu walencyjnego pokazuje rysunek 2. Zawiera on między innymi ocenę danego schematu – jest to schemat *pewny*. Inne możliwe wartości oceny to: *wątpliwy*, *zły*, *archaiczny*, *potoczny* i *wulgarny*. Ocena dotyczy najszerszego rejestru, dla którego dany schemat ma zastosowanie. Oceny *wątpliwy* i *zły* pozwalają na notowanie schematów, co do których poprawności leksykografowie mają wątpliwości, a nawet częstych błędów składniowych bez włączania ich w obręb normy.

Schemat ten dobrze ilustruje wspomnianą powyżej możliwość podania listy dopuszczalnych realizacji morfoskładniowych danej pozycji: specyfikacją jednej z trzech pozycji czasownika POWIEDZIEĆ, pozycji dopełnienia, jest siedmioelementowa lista typów fraz zawierająca między innymi prepnp(o,loc) (jak np. w zdaniu *Powiedziałem jej o matce*) i cp(że) (faza zdaniowa wprowadzona spójnikiem

podrzednym typu że, np. w zdaniu *Powiedziałem jej, że on ją...*). Wymienione tu typy fraz zostały uznane za możliwe realizacje jednej pozycji w schemacie walencyjnym, gdyż możliwa jest ich koordynacja, o czym świadczy jeden z udokumentowanych przykładów użycia tego schematu: *Zanim się odezwała, powiedziałem jej o matce i że on ją...*

Schemat:	pewny [14737] 		
Funkcja:	subj	obj	
Typy fraz:	np(str)	prepnp(o.loc)	np(dat)
		cp(int)	
		cp(że)	
		cp(żeby)	
		prepncp(o.loc.int)	
		prepncp(o.loc.że)	
		prepncp(o.loc.żeby)	

149338	Oczywiście, najpierw musielibyśmy ich przeszkolić, powiedzieć im o bobrach i o tym, jak można je upolować.	pełny NKJP (1800M segmentów)	dobry
149342	Powiedział mi, że do końca pierwszej połowy zostało 3 minuty i żebym wytrzymał do przerwy.	pełny NKJP (1800M segmentów)	dobry
149345	Zanim się odezwała, powiedziałam jej o matce i że on ją...	pełny NKJP (1800M segmentów)	dobry
149351	Na tym szczeblu nie potrafię powiedzieć o planach awaryjnych i czy istnieje możliwość, w razie odcięcia dopływu gazu z Rosji, skorzystania z własnych dostaw.	pełny NKJP (1800M segmentów)	wątpliwy

**Rys. 2. Schemat walencyjny i przykłady dla powiedzień (└, perf)**

Powyższe przykłady ilustrują tylko niektóre własności słownika i możliwe typy realizacji; ich prawie pełną listę znaleźć można we wspomnianym artykule Przepiórkowski i in. 2014c, zaś nowszą koncepcję reprezentacji informacji frazeologicznych – w artykule Przepiórkowski i in. 2014a (oba dostępne na stronie <http://zil.ipipan.waw.pl/Walenty>). Na koniec niniejszego punktu zasygnalizujemy jedynie istnienie złożonych typów fraz „okolicznikopodobnych”, o określonej semantyce i różnorodnych realizacjach morfoskładniowych. We wcześniejszych słownikach walencyjnych, np. słowniku Świdzińskiego 1994, takie typy fraz oznaczane były często jako frazy przysłówkowe, choć mogły być też realizowane przez

inne typy fraz, np. frazy przyimkowe czy nawet zdaniowe. W *Walentym* oznaczone są one jako xp. Istnieje dziewięć typów takich fraz: xp(mod), xp(locat), xp(perl), xp(abl), xp(adl), xp(dur), xp(temp), xp(caus) oraz xp(instr). Większość symboli typu abl (ablatywny) czy dur (duratywny, oznaczający ciągłość w czasie) powinna być jasna. Wyjątkiem może być mod – realizacje typu xp(mod) służą do wyrażenia sposobu wykonania czynności, jak w schematach walencyjnych czasowników ZACHOWYWAĆ SIĘ (*ktoś zachowuje się w jakiś sposób*) czy TRAKTOWAĆ (*ktoś traktuje kogoś w jakiś sposób*); o możliwości różnorodnych realizacji morfoskładniowych takich złożonych typów fraz niech świadczą przykłady uzupełnień ciągu *Janek zachowywał się*: fraza przysłówkowa – *właściwie*, fraza przyimkowa – *bez zarzutu*, fraza zdaniowa – *jakby nic się nie stało*.

Podstawową decyzją przy tworzeniu słownika walencyjnego jest określenie, które podrzędniki mają być w nim notowane. W *Walentym* przyjęte zostało podejście permissywne. Mianowicie ustalono, że poza podrzędnikami wymaganymi w sposób bezwzględny, których opuszczenie prowadzi do konstrukcji eliptycznych, uwzględniane są podrzędniki typowe dla danego czasownika. Podejście to jest podobne do zaproponowanego w słowniku walencyjnym czeskiego Vallex (Žabokrtský i Lopatková 2007: s. 5), z tym że my nie dzielimy pozycji na obligatoryjne, fakultatywne i typowe.

### 3. Widok haseł

Na górnym pasku strony <http://walenty.ipipan.waw.pl/> widnieją zakładki **Hasła**, **Administracja**, **Rozwinięcia typów fraz** oraz **Pobierz słownik**. Pierwsza zakładka zapewnia dostęp do haseł; po lewej stronie widnieje lista haseł wraz z ich statusem<sup>1</sup>, po prawej stronie – graficzna reprezentacja podhaseł aktualnego hasła, którą można przewijać. Fragment hasła SPIESZYĆ z dwoma podhasłami widnieje na rysunku 3. Podstrona zawiera dwie zakładki: **Schematy** (z liczbą schematów hasła) oraz **Przykłady**, zawierająca przykłady nietypowe, nie pasujące do żadnego schematu. Po prawej widnieją informacje o frekwencji w podkorpusach NKJP ręcznie anotowanym (1M) i zrównoważonym (300M).

Pojedynczy schemat reprezentowany jest w postaci tabeli. Poszczególne kolumny tabeli reprezentują pozycje. Pierwszy wiersz tabeli zawiera ocenę schematu, zaś drugi wiersz zawiera informacje dotyczące całej pozycji, czyli funkcje gramatyczne (podmiot, dopełnienie) oraz informacje dotyczące kontroli. Kolejne wiersze zawierają konkretne typy fraz mogące realizować pozycje, i dla każdej pozycji

<sup>1</sup> Słownik jest cały czas w fazie tworzenia, a statusy pokazują, w jakiej fazie przetwarzania jest dane hasło. Niezależnie od statusu, dowolne hasło może ulec korekcie.

wypisywane są niezależnie dla każdej kolumny (tzn. współwystępowanie w wierszu nie ma znaczenia).

Kliknięcie na pole z oceną schematu powoduje wyświetlenie listy przykładowych zdań powiązanych z danym schematem, por. rysunek 2 (dla **powie-dzieć** (**\_**, **perf**)). Z kolei kliknięcie na pojedynczy przykład powoduje wyróżnienie w tabeli typów fraz realizowanych w tym konkretnym przykładzie (informacja ta jest wpisywana przez leksykografów). Obok zdania widnieje źródło, z którego schemat pochodzi (konkretny podkorpus NKJP, literatura, przykład własny). Jako przykłady własne traktowane są także przykłady pochodzące z Internetu. W przeciwieństwie do przykładów pochodzących z korpusu, z których można jedynie obcinać początek lub koniec, przykłady pochodzące z internetu leksykografowie mogli dość dowolnie modyfikować (skracać, poprawiać literówki, interpunkcję itp.).

W celu wyszukania określonego hasła należy wpisać jego lemat (wystarczy po-czątek) w pole tekstowe nad listą haseł i nacisnąć ENTER.

The screenshot shows the Słowa.pl interface. At the top, there are navigation tabs: 'Hasła', 'Administracja', 'Rozwinięcia typów fraz', and 'Pobierz słownik'. A search bar contains 'spieszyc'. Below the search bar is a table of search results. The entry 'spieszyc' is highlighted. To the right, a detailed view for 'spieszyc (1M=25,300M-4932)' is shown. It lists three grammatical schemas with their respective functions and phrase types.

Hasło	Status
spieć	(F) gotowe
spić	(F) sprawdzone
spiec	gotowe
spiekać	gotowe
spiekły	sprawdzone
spienić	(F) gotowe
spieniężyć	gotowe
spieprzyć	gotowe
spierać	(F) gotowe
spierdalać	(F) gotowe
spierzchny	sprawdzone
spierzchnąć	(F) sprawdzone
spierzchnięty	sprawdzone
spieszno	gotowe
<b>spieszyc</b>	(F) gotowe
spiętrzać	(F) gotowe
spiętrzyć	(F) gotowe
spijać	(F) sprawdzone
spiknąć	(F) gotowe
spilować	gotowe
spinać	sprawdzone
spiorunować	gotowe
spisać	(F) sprawdzone
spiskować	(F) gotowe

Schemat	archaiczny [261]
Funkcja:	subj   obj
Typy fraz:	np(str)   np(str)

Schemat	archaiczny [532]
Funkcja:	subj
Typy fraz:	np(str)   refl


Schemat	pewny [105]
Funkcja:	subj_controller   controllee
Typy fraz:	np(str)   infp(_)

Rys. 3. Zrzut ekranu programu *Słowa* z fragmentem hasła **SPIESZYĆ**

#### 4. Filtrowanie haseł

Jedną z zalet tego, że *Walenty* jest słownikiem elektronicznym, jest dostępność w *Słowalu* funkcji filtrowania haseł. Umożliwia to użytkownikowi skupienie się na hasłach reprezentujących interesujące go zjawiska.

Rys. 4. Okno definicji filtru w programie *Slowal*

Aby dokonać filtrowania haseł, należy kliknąć przycisk  znajdujący się po lewej stronie nad listą haseł, obok pola tekstowego. Wyświetli się wówczas formularz z kategoriami filtrowania. Filtrować można po ogólnych cechach podhasła, takich jak aspekt, zanegowanie i zwrotność. Informacje te wybierane są z listy. Ponadto istnieje możliwość filtrowania po poszczególnych typach fraz oraz po całych pozycjach. Odpowiednią informację wpisujemy w formacie tekstowym słownika; program podpowiada listę typów fraz czy też pozycji zgodnych z już wpisaną specyfikacją. I tak wpisanie typu frazy np(str) spowoduje wybór haseł zawierających przynajmniej jeden schemat zawierający frazę nominalną w przypadku strukturalnym – wszystko jedno, na jakiej pozycji (`{np(str)}`, `subj{np(str)}`, `obj{np(str)}`, `koordynacja z np(str)`). Natomiast wpisanie pozycji `obj{np(inst)}` spowoduje wybór haseł posiadających przynajmniej jeden schemat z dopełnieniem narzędnikowym podlegającym pasywizacji, por. rysunek 5.

Wpisywane ograniczenia na postać typów fraz i pozycji mają charakter wyrażeń regularnych (bez możliwości nawiasowania). W praktyce najbardziej użyteczne są symbole `*` oznaczające dowolny ciąg znaków (także pusty). Na przykład zapytanie `prepnp(na,*)` oznacza frazę przyimkową z przyimkiem `na` w dowolnym przypadku; zapytanie `prepnp(*,acc)` oznacza frazę przyimkową z dowolnym przyimkiem wymagającym biernika; zaś `fixed(*)` oznacza frazę dowolnego typu zleksykalizowaną do zamrożonej postaci (por. Przepiórkowski i in. 2014a). Z kolei pozycja `subj{.*}` oznacza schematy posiadające podmiot o dowolnej realizacji,



Hasła    Administracja    Rozwinięcia typów fraz    Pobierz słownik

Hasło    Status

administrować	(F) sprawdzone
bujać	(F) sprawdzone
<b>ciskać</b>	<b>(F) sprawdzone</b>
cisnąć	(F) gotowe
dorzucac	(F) sprawdzone
dorzucić	(F) sprawdzone
dowodzić	(F) sprawdzone
gibać	(F) sprawdzone
gibnąć	gotowe
kopcić	gotowe
manipulować	(F) sprawdzone
miotac	(F) w obróbie
nawigować	(F) sprawdzone
owładnać	(F) sprawdzone
pobujac	sprawdzone
podrzucac	sprawdzone
podrzucić	sprawdzone
pomiatć	(F) gotowe
poniewierac	(F) sprawdzone
rozbujać	gotowe
rzucac	sprawdzone
rzucić	(F) gotowe
sterować	(F) gotowe
szamotać	(F) sprawdzone
szargać	sprawdzone
szarpać	sprawdzone
szarpnąć	sprawdzone
szurnać	(F) sprawdzone
targać	sprawdzone
targnąć	sprawdzone
władać	(F) sprawdzone
współzadzić	(F) sprawdzone
wstrząsać	sprawdzone
wstrząsnąć	sprawdzone
zachwiać	(F) sprawdzone
zawidywać	(F) sprawdzone

Schematy [31]    Przykłady [1]    **ciskać [1]**

**ciskać ( „imperj) :**

Schemat:	pewny [24157]			
Funkcja:	subj	obj		
Typy fraz:	np(str)	np(inst)	lex{prepnp(w.acc),sg,'kąf',natr}	

Schemat:	pewny [7241]				
Funkcja:	subj	obj			
Typy fraz:	np(str)	np(inst)	np(dat)	prepnp(w.acc)	

Schemat:	pewny [7242]			
Funkcja:	subj	obj		
Typy fraz:	np(str)	np(inst)	prepnp(na.acc)	

Schemat:	pewny [7246]			
Funkcja:	subj	obj		
Typy fraz:	np(str)	np(inst)	prepnp(o.acc)	

Przegląd przykładów    Filtrowanie


Identyfikator:	Przykład:	Źródło:	Ocena:
75013	Tu nikt nie ciska ci w twarz toporną ideologią.	pełny NKJP (1800M segmentów)	dobry

**Rys. 5. Lista hasel z dopełnieniem narzędnikowym podlegającym pasywizacji wraz z adekwatnymi schematami czasownika CISKAĆ**

natomiast  $\{.*;.*\}$  oznacza schematy posiadające dowolną pozycję skoordynowaną. Ponadto symbol  $|$  oznacza alternatywę, zaś symbol  $\&$  – koniunkcję warunków. Tak więc ograniczenie na typ frazy  $np(dat)|prepnp(do,gen)$  filtruje hasła z przynajmniej jednym schematem zawierającym którykolwiek z wymienionych typów fraz;  $obj\{.*np(str);.*\}|obj\{.*;np(str).*\}$  oznacza schematy, w których jednym z typów fraz podlegających koordynacji na pozycji dopełnienia jest fraza nominalna w przypadku strukturalnym; zaś ograniczenie na pozycję  $subj\{.*\}\&obj\{.*\}$  filtruje hasła z przynajmniej jednym schematem zawierającym zarówno pozycję podmiotu, jak i dopełnienia.

Kliknięcie na przycisk  spowoduje, że lista hasel zostanie ograniczona do hasel posiadających schematy spełniające łącznie wszystkie kryteria filtrowania

wpisane w polach formularza. Wybranie pola Odfiltruj niepasujące schematy spowoduje, że dla oglądanego hasła wyświetlą się wyłącznie schematy spełniające warunki filtrowania.

Istnieje też uproszczony tryb filtrowania, dotyczący filtrowania schematów aktualnie oglądanego hasła. W tym celu należy kliknąć przycisk **Filtrowanie schematów** po prawej stronie nad listą przykładów. W celu cofnięcia filtrowania należy kliknąć przycisk .

## 5. Rozwinięcia typów fraz

Możliwe realizacje pewnych typów fraz zostały w *Walentym* celowo oddzielone od haseł. Są one udostępniane osobno, poprzez zakładkę **Rozwinięcia typów fraz**. Poza wspomnianymi wcześniej frazami typu xp są to przyimki złożone comprepnp oraz przysłowki advp, grupowane w podobny sposób, co frazy xp. W zestawie tym uwzględnione zostały także dwa bardziej specyficzne typy fraz, czyli distrp służący do opisu dystrybucyjnego po oraz possp służący do opisu fraz dzierżawczych (stosowany głównie w opisie walencji rzeczowników).

Ten sposób reprezentacji umożliwia nam z jednej strony uproszczenie postaci słownika (bardzo przydatne przy jego opracowywaniu) i zwiększenie jego czytelności, a z drugiej strony zachowanie jego spójności. Zmiany zbioru realizacji wprowadzane są w jednym miejscu, bez konieczności sprawdzania, których haseł dotyczą. W celu obejrzenia realizacji konkretnego typu frazy wystarczy wybrać go z rozwijanych list.

## 6. Postać tekstowa słownika

Powyższe przykłady nawiązywały do postaci *Walentego* dostępnej bezpośrednio przez Internet. Generowanie aktualnej wersji *Walentego* w różnych postaciach wymaga obecnie specjalnych uprawnień, gdyż jest procesem czasochłonnym. Co około pół roku postać tekstowa i PDF są jednak generowane przez twórców słownika i udostępniane na stronie <http://zil.ipipan.waw.pl/Walenty>. Najprostszym formatem słownika jest postać tekstowa. Jej głównym przeznaczeniem jest automatyczne przetwarzanie przez komputery. Użytkownikowi w pierwszym momencie może wydawać się nieczytelna i korzystanie z niej wymaga pewnej wprawy. Jednak fakt, że można ją łatwo i szybko przeglądać i przeszukiwać za pomocą dowolnego edytora powoduje, że w niektórych sytuacjach jest to najwygodniejszy sposób korzystania z *Walentego*. Postać tekstową słownika można też w dowolnym momencie pobrać ze *Słowala*, wybierając zakładkę **Pobierz słownik**.

Postać tekstowa reprezentowana jest w postaci listy schematów, gdzie każdy schemat zapisany jest w osobnym wierszu. Jeden z wielu schematów czasownika SPIESZYĆ podany jest poniżej (wiersz został złamany z powodów typograficznych):

```
spieszyć: _: imperf:  
subj,controller{np(str)} + controllee{infp(_)}
```

Informacje dotyczące podhasła powtarzane są przy każdym schemacie i oddzielane dwukropkami – w powyższym przykładzie jest to informacja o formie hasłowej (spieszyć), o tym, że schemat nie jest wrażliwy na występowanie negacji (–) i o aspekcie niedokonanym (imperf). Poszczególne pozycje zapisywane są w nawiasach klamrowych (sygnalizujących możliwość wielu realizacji morfoskładniowych) i oddzielane symbolem ‘+’. Informacje dotyczące całej pozycji (subj, obj, controller, controllee) zapisywane są na początku pozycji i są rozdzielone przecinkami, np. controllee{infp(\_)} czy subj,controller{np(str)}. Natomiast realizacje wymagań składające się na daną pozycję oddzielane są średnikami – ta ostatnia konwencja zilustrowana jest na rysunku 6 zawierającym listę wszystkich schematów czasownika SPIESZYĆ zakodowanych w omawianej postaci. Zauważmy, że postać tekstowa nie zawiera przykładów użycia poszczególnych schematów.

Powyższe zasady tekstowej reprezentacji schematów nie są zbyt złożone, jednak znacznie się komplikują w wypadku zapisu frazeologizmów (Przepiórkowski i in. 2014a). Wynika to z faktu, że *Walenty* umożliwia bardzo szczegółowy opis fraz zleksykalizowanych, z dowolnym poziomem zagnieżdżenia, na przykład pozwala na opisanie faktu, że jednym z typowych podrzędników czasownika WITAĆ może być fraza przymikowo-nominalna o przymioku z i narzędnikowej frazie rzeczownikowej o centrum *ramionami*, przy czym centrum to *musi* być modyfikowane przez frazę przymiotnikową o centrum *otwartymi* (nie można frazeologicznie powiedzieć: *Witali nas z ramionami*), przymiotnik ten *może* być z kolei modyfikowany przez przysłówek typu *szeroko* (*Witali nas z szeroko otwartymi ramionami*), ale przysłówek ten raczej *nie może* być dalej modyfikowany (np. niefrazeologiczne *Witali nas z niezwykle szeroko otwartymi ramionami*<sup>2</sup>).

## 7. Słownik w formacie PDF

Format „tekstowy” omówiony powyżej oznacza czysty tekst przeznaczony do automatycznego przetwarzania oraz przeszukiwania, format PDF zaś przeznaczony jest wyłącznie do studiowania przez użytkownika. Ma on układ zwiększający czytelność i zawiera wiele informacji pominiętych w formacie tekstowym.

<sup>2</sup> Nie opisujemy okazjonalnie występujących ciągów z *bardzo szeroko otwartymi ramionami*.

spieszyć: \_: imperf: subj,controller{np(str)} +  
           controllee{infp(\_)}  
 spieszyć: \_: imperf: subj{np(str)} + {np(dat)} + {np(str)}  
 spieszyć: \_: imperf: subj{np(str)} + {np(dat)} + {prepnp(na,acc)}  
 spieszyć: \_: imperf: subj{np(str)} + {np(dat)} + {prepnp(o,acc)}  
 spieszyć: \_: imperf: subj{np(str)} + {np(dat)} + {prepnp(z,inst)}  
 spieszyć: \_: imperf: subj{np(str)} +  
           {prepnp(do,gen); prepnp(ku,dat); prepnp(na,acc); xp(adl)} +  
           {xp(perl)}  
 spieszyć: \_: perf: subj{np(str)} + obj{np(str)}  
 spieszyć: \_: perf: subj{np(str)} + {refl}  
 spieszyć się: \_: imperf: {np(dat)} +  
           {prepnp(do,gen); prepnp(na,acc); prepnpcp(do,gen,żeby); xp(adl)}  
 spieszyć się: \_: imperf: subj,controller{np(str)} +  
           controllee{infp(\_)}  
 spieszyć się: \_: imperf: subj{np(str)} + {np(dat)} + {np(str)}  
 spieszyć się: \_: imperf: subj{np(str)} + {np(dat)} +  
           {prepnp(o,acc)}  
 spieszyć się: \_: imperf: subj{np(str)} +  
           {prepnp(do,gen); prepnp(na,acc); prepnpcp(do,gen,żeby); xp(adl)}  
 spieszyć się: \_: imperf: subj{np(str)} +  
           {prepnp(z,inst); prepnpcp(z,inst,żeby)}

**Rys. 6. Lista schematów czasownika SPIESZYĆ w formie tekstowym (niektóre wiersze złamane z powodów typograficznych)**

Słownika w tym formacie raczej nie należy drukować w całości: choć zawiera on tylko hasła sprawdzone przez leksykografów nadzorujących proces powstawania słownika, to – w wersji z maja 2015 zawierającej 11334 predykatów – liczy on ponad 30 000 stron. Fragment strony 31 545 tej wersji przedstawiony jest na rysunku 7.

Przedstawiony jest tu schemat walencyjny pierwszego podhasła rzeczownika ŻĄDANIE. Trzy pozycje tego rzeczownika oznaczone są jako **(A)**, **(B)** i **(C)**. Dwie ostatnie mają po jednej możliwej realizacji morfoskładniowej, zaś pierwsza posiada trzy takie realizacje: np(gen), prepnp(o,acc) oraz ncp(gen,żeby). Znaczenie symboli np i prepnp zostało zasygnalizowane powyżej; symbol ncp oznacza natomiast frazę zdaniową (tutaj wprowadzoną przez spójnik podrzędny typu ŻEBY) z korelatem (tutaj w dopełniaczu), np. w ciągu: *żądanie tego, żeby wprowadzić pisownię „tak jak się mówi”*.

Kolejne typy fraz realizujące poszczególne pozycje oznaczone są kolejnymi liczbami w kwadratach – oznaczenia te wykorzystywane są w przykładach dla

zasygnalizowania, które realizacje których pozycji w tych przykładach występują. I tak w obu przykładach występuje tylko pozycja A, choć w drugim z nich jako koordynacja typów fraz [1] (tj. np(gen)) i [3] (tj. ncp(gen,żeby)).

(1) **żądanie** („) — pewny [16618]

(A)	(B)	(C)
[1] np(gen)	[4] possp	[5] prenp(od,gen)
[2] prenp(o,acc)		
[3] ncp(gen,żeby)		

---

1. Niestety - wyrobienie sobie bezstronnego poglądu (jeżeli taki w ogóle istnieje) wymaga pracowitego szperania i zapoznania się z wieloma źródłami, a nie żądania żeby w jednym miejscu ktoś podał gotowe danie.  
A { [3] } pełny NKJP (1800M segmentów) — dobry [184643]

2. Poza tym żądanie odwołania się do zwyczaju i tego, żeby wprowadzić pisownię „tak jak się mówi”, rodzi następny problem: jak kto mówi: mieszkaniec Warszawy czy mieszkaniec Poznania?  
A { [1] ; [3] } pełny NKJP (1800M segmentów) — dobry [184645]

Rys. 7. Fragment strony 31 545 wersji PDF *Walentego* z maja 2015

## 8. Podsumowanie

Celem niniejszego artykułu było przedstawienie powstającego w IPI PAN słownika *Walenty*, już obecnie bardziej zaawansowanego niż inne słowniki walencyjne. Szczegółowy format zapisu schematów walencyjnych nadal ewoluuje i został już parokrotnie opisany (Przepiórkowski i in. 2014c, b, a). Tutaj skupiamy się raczej na pokazaniu, że już teraz zasób ten może być z powodzeniem wykorzystywany przez społeczność lingwistyczną. Różne postacie słownika mają zwiększyć jego użyteczność: wersja dostępna bezpośrednio w Internecie zawiera zawsze aktualne dane i pozwala na zaawansowane filtrowanie haseł, postać tekstowa umożliwia szybkie wyszukiwanie najróżniejszych rodzajów informacji, zaś plik PDF jest chyba najbardziej czytelną postacią dla tych użytkowników, którzy nie mają zwyczaju korzystania z zasobów elektronicznych.

## Literatura

- Kallas, K. (1993). *Składnia współczesnych polskich konstrukcji współrzędnych*. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika.
- Landau, I. (2013). *Control in Generative Grammar: A Research Companion*. Cambridge: Cambridge University Press.

- Patejuk, A., Przepiórkowski, A. (2012). Towards an LFG parser for Polish: An exercise in parasitic grammar development. W: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC-2012)*. Istanbul, Turcja: ELRA, (3849–3852).
- Przepiórkowski, A., Bańko, M., Górski, R.L., Lewandowska-Tomaszczyk, B. (red.) (2012). *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Przepiórkowski, A., Hajnicz, E., Patejuk, A., Woliński, M. (2014a). Extended phraseological information in a valence dictionary for NLP applications. W: *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014)*. Dublin, Irlandia, (83–91). URL <http://www.aclweb.org/anthology/siglex.html>.
- Przepiórkowski, A., Hajnicz, E., Patejuk, A., Woliński, M., Skwarski, F., Świdziński, M. (2014b). Walenty: Towards a comprehensive valence dictionary of Polish. W: N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk, S. Piperidis (red.), *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2010)*. Reykjavík, Islandia: ELRA. ISBN 978-2-9517408-8-4, (2785–2792). URL <http://www.lrec-conf.org/proceedings/lrec2014/index.html>.
- Przepiórkowski, A., Rosen, A. (2005). Czech and Polish raising/control with or without structure sharing. *Research in Language*, 3, 33–66. URL <http://nlp.ipipan.waw.pl/~adamp/Papers/2004-control-czech-polish/>.
- Przepiórkowski, A., Skwarski, F., Hajnicz, E., Patejuk, A., Świdziński, M., Woliński, M. (2014c). Modelowanie własności składniowych czasowników w nowym słowniku walencyjnym języka polskiego. *Polonica*, XXXIII, 159–178.
- Rosenbaum, P. (1967). *The Grammar of English Predicate Complement Constructions*. Cambridge, MA: The MIT Press.
- Sag, I.A., Gazdar, G., Wasow, T., Weisler, S. (1985). Coordination and how to distinguish categories. *Natural Language and Linguistic Theory*, 3, 117–171.
- Świdziński, M. (1992). *Gramatyka formalna języka polskiego*. Rozprawy Uniwersytetu Warszawskiego. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Świdziński, M. (1994). *Syntactic Dictionary of Polish Verbs*. Uniwersytet Warszawski / Universiteit van Amsterdam.
- Świdziński, M., Woliński, M. (2010). Towards a bank of constituent parse trees for Polish. W: P. Sojka, A. Horák, I. Kopeček, K. Pala (red.), *Text, Speech and Dialogue: 13th International Conference, TSD 2010, Brno, Czech Republic, Lecture Notes in Artificial Intelligence*, tom 6231. Heidelberg: Springer-Verlag, (197–204).
- Woliński, M. (2004). *Komputerowa weryfikacja gramatyki Świdzińskiego*. Rozprawa doktorska, Instytut Podstaw Informatyki, Polska Akademia Nauk, Warszawa.
- Žabokrtský, Z., Lopatková, M. (2007). Valency information in VALLEX 2.0: Logical structure of the lexicon. *The Prague Bulletin of Mathematical Linguistics*, 87, 41–60.

## *An on-line corpus-based valence dictionary of Polish*

### s u m m a r y

The paper presents an electronic valence dictionary of Polish empirically based mainly on the National Corpus of Polish. The syntax of lexical entries is only sketched here. The main focus of the paper is the presentation of the lexical data on-line, i.e. via Internet browsers, as well as in two textual formats: in plain text and as a PDF. The emphasis is put here on the functionalities offered by the Internet access to *Walenty*, including the presentation of examples of particular realisations of valence schemata, as well as rich filtering mechanisms for entries and schemata.