

Adam Przepiórkowski, Filip Skwarski,
Elżbieta Hajnicz, Agnieszka Patejuk,
Marek Świdziński, Marcin Woliński

Modelowanie własności składniowych czasowników w nowym słowniku walencyjnym języka polskiego

29 września 2013 (godz. 11:14)

(złożone do Poloników, komentarze mile widziane)

1. Wprowadzenie

Niniejszy artykuł przedstawia zakres zjawisk składniowych modelowanych w elektronicznym słowniku walencyjnym – roboczo nazwanym *Walenty* (<http://clip.ipipan.waw.pl/Walenty>) – opracowywanym w Instytucie Podstaw Informatyki PAN (IPI PAN). Słownik ten ma być wykorzystywany nie tylko przez ludzi, ale także przez programy znajdujące rozbiory składniowe wypowiedzeń polskich, tzw. parsery, w tym dwa parsery obecnie rozwijane w IPI PAN: Świgr (Woliński 2004, Hajnicz i Woliński 2009, Świdziński i Woliński 2010; <http://zil.ipipan.waw.pl/Sk%C5%82adnica>), czyli parser oparty na gramatyce Świdzińskiego (1992a), oraz POLFIE (Patejuk i Przepiórkowski 2012b; <http://zil.ipipan.waw.pl/LFG>), parser oparty na formalizmie LFG (Bresnan 1982, Dalrymple 2001).

Punktem wyjścia dla nowego słownika był niepublikowany elektroniczny słownik walencyjny Marka Świdzińskiego (Świdziński 1994, 1998), wzbogacony na potrzeby

parsera Świgr i dostępny w dystrybucji systemu pod adresem podanym powyżej. Poniższe przykłady ram walencyjnych ilustrują format tego słownika wyjściowego:

wydawać V np(bier)+np(ceł)

wydawać V np(bier)+prepn('za', bier)

wydawać V prepn('na', bier)+np(bier)

Należy przede wszystkim zauważyć, że jest to słownik powierzchniowoskładniowy, nie rozróżniający sensów słów: pierwsza z powyższych ram może być użyta między innymi w znaczeniu wydawania reszty, wydawania polecenia czy wydawania książki. W słowniku jednostki umieszczone są w oddzielnych wierszach. Każdy wiersz składa się z trzech części: z czasownika w formie hasłowej, z informacji o tym, czy czasownik jest właściwy (V), czyli wymagający podmiotu nominalnego, czy też niewłaściwy (Q), czyli niewymagający podmiotu w mianowniku, oraz z maksymalnego schematu walencyjnego. Jeśli danemu czasownikowi można przypisać więcej niż jeden schemat, wówczas każdy ze schematów znajduje się w oddzielnym wierszu, a pozostałe informacje są powtórzone. Jeśli czasownik nie ma żadnych wymagań walencyjnych (nie licząc podmiotu czasowników właściwych), to pole ze schematami pozostaje puste.

Schematy składają się z fraz przedzielonych znakiem +. Możliwe typy fraz to:

- np(PRZYPADEK) – fraza nominalna w danym przypadku (możliwe wartości to: mian, dop, ceł, bier, narz), na przykład: *Jaś ma samochód* (np(bier));
- prepn('PRZYIMEK', PRZYPADEK) – fraza przyimkowa z konkretnym przyimkiem i wymaganym przez niego przypadkiem (dodatkowa możliwa wartość to *miej*), na przykład: *Jaś idzie do szkoły* (prepn('do', dop));
- adjp(PRZYPADEK) – fraza przymiotnikowa w danym przypadku, na przykład: *Jaś pamięta Marysię młodą* (adjp(narz));
- prepadjp('PRZYIMEK', PRZYPADEK) – fraza przyimkowo-przymiotnikowa z konkretnym przyimkiem i wymaganym przez niego przypadkiem, na przykład: *Tę inicjatywę uważam za nieudaną*. – np(bier)+prepadjp(za, bier). W wypadku frazy zbudowanej z przyimka i poprzyimkowej formy specjalnej przymiotnika używamy specjalnej wartości *pop* przypadku, na przykład: *Jaś mówi po polsku* (prepadjp('po', pop));
- sentp('TYP') – fraza zdaniowa konkretnego typu, na przykład pytajnozależna (pz), podrzędna z *że* (*że*), z *BY/ABY/ŻEBY* (*żeby*), z *JAKOBY* (*jakoby*) itd., na przykład: *Jaś wie, że dziś jest wtorek*;

- sentp('PRZYIMEK', PRZYPADEK, 'TYP') – fraza zdaniowa z korelatem z przyimkiem – tak samo jak w wypadku frazy zdaniowej, ale dodatkowo oznaczony jest przypadek, w którym występuje korelat, oraz przyimek, który tego przypadku wymaga, na przykład: *Jaś wiedział o tym, że dziś jest wtorek* (sentp('o', miej, 'że'));
- sentp(PRZYPADEK, 'TYP') – fraza zdaniowa z korelatem bez przyimka, na przykład: *Jaś w końcu uświadomił sobie to, że dziś jest wtorek* (sentp(bier, 'że'));
- advp – fraza przysłówkowa, na przykład: *Jaś ma się dobrze*;
- infp(ASPEKT) – fraza bezokolicznikowa, w nielicznych wypadkach wymaga oznaczenia aspektu czasownika w bezokoliczniku (nd lub dk), na przykład: *Jaś zaczyna pisać* (infp(nd)), jeśli zaś aspekt jest dowolny, to w słowniku zostaje to oznaczone symbolem _, czyli infp(_);
- or – mowa niezależna, na przykład: *Jaś powiedział „mam dość!” i poszedł*;
- refl – użycie zwrotne czasownika np. *Napoleon ogłosił się cesarzem* (użycie zwrotne) wobec *Biskup ogłosił Napoleona cesarzem* (użycie zwykłe) – por. §7.1.

Celem prac prowadzonych w IPI PAN, m.in. w ramach projektów CESAR, NEKST i CLARIN-PL (zob. <http://clip.ipipan.waw.pl/Projects>), jest stworzenie swobodnie dostępnego (ang. *open source*) słownika walencyjnego znacznie większego niż liczący około 2100 haseł¹ słownik wyjściowy i zawierającego – oprócz informacji morfoskładniowych – także informacje semantyczne i frekwencyjne. W niniejszym artykule omawiamy wyniki prowadzonego w ramach projektu CESAR pierwszego etapu tych prac, którego wymiernym rezultatem było przede wszystkim opracowanie nowego formalizmu zapisu powierzchniowskładniowych informacji walencyjnych, lecz także istotne zwiększenie rozmiaru słownika oraz przypisanie ramom przykładów ich realizacji, zwykle pochodzących z Narodowego Korpusu Języka Polskiego (NKJP; <http://nkjp.pl/>; Przepiórkowski i in. 2012). Nowy formalizm uwzględnia zjawiska pominięte w poprzednich słownikach walencyjnych, w tym w wykorzystywanych w niniejszych pracach słownikach Polańskiego (1980–1992) i Bańki (2000), lub uwzględnione w nich w stopniu niewystarczającym, przede wszystkim:

- możliwość koordynacji fraz różnych typów w obrębie danej pozycji składniowej,
- niestabilność przypadkową argumentów nominalnych w niektórych pozycjach,

¹ Haseł w słowniku wyjściowym jest około 1440, jeżeli nie oddzielać lematów z *SIĘ* i *BEZ SIĘ*, tj. jeżeli na przykład jako jedno hasło traktować *BRAĆ* (np. *coś od kogoś*) i *BRAĆ SIĘ* (np. *za coś*).

- podmioty nienominalne,
- ograniczenia możliwości utworzenia strony biernej (pasywizacji),
- relację kontroli składniowej,
- zróżnicowanie typów fraz przysłówkowych,
- konstrukcje frazeologiczne.

Elementy formalizmu opisu schematów walencyjnych uwzględniającego te zjawiska omówione są w kolejnych punktach niniejszego artykułu. Dodatek do artykułu zawiera pełniejszy opis formatu postaci tekstowej (w odróżnieniu od wewnętrznej, bazodanowej) słownika.

2. Formalizm opisu – przykład

Celem niniejszego punktu jest zilustrowanie pierwszych czterech z siedmiu wymienionych powyżej czynników wziętych pod uwagę przy projektowaniu nowego formalizmu opisu ram walencyjnych. Przyjęte rozwiązania wprowadzamy stopniowo, jako kolejne przybliżenia jednej z ram czasownika *TLUMACZYĆ*. Dopiero ostatnie przybliżenia zgodne są z docelowym formatem słownika – wcześniejsze mają charakter pogładowy.

W zdaniach typu *Ktoś tłumaczy coś komuś* mamy do czynienia z ramą, którą w najprostszy sposób można wyrazić tak²:

np(nom) + np(acc) + np(dat)

Argumenty mianownikowe często – lecz nie zawsze – są podmiotami. Do wyjątków należy na przykład drugi argument czasownika *NAZYWAĆ SIĘ*. Z drugiej strony nie każdy podmiot jest mianownikiem. Uznajemy na przykład, że czasownik *DZIWIĆ* może się łączyć z podmiotami zdaniowymi, jak w zdaniu *Jana dziwi, że Maria wybiera Piotra* (Świdziński 1992b). Dlatego też podmioty w ramach walencyjnych oznaczane są *explicite* (por. §3).

Podobne rozważania prowadzą do jawnego oznaczania argumentów podlegających pasywizacji, tj. przechodzących na pozycję podmiotu w wyniku utworzenia strony biernej: są to zwykle argumenty biernikowe, ale nie tylko i nie wszystkie. Na przykład nie można utworzyć strony biernej od łączącego się z biernikiem czasownika *OBCHODZIĆ*

² Przyjmujemy tu terminologię anglosaską: np oznacza *nominal phrase* (faza nominalna), nom – *nominative* (mianownik), acc – *accusative* (biernik), dat – *dative* (celownik) itd.

użytego w znaczeniu *coś obchodzi kogoś*, można natomiast utworzyć stronę bierną od niektórych czasowników łączących się z narzędnikiem, np. RZĄDZIĆ, KIEROWAĆ, DOWODZIĆ czy ZARZĄDZAĆ, z dopełniaczem, np. DOWIEŚĆ, POSZUKIWAĆ, WYSŁUCHAĆ, a może też od niektórych czasowników łączących się z celownikiem, np. ZAGROZIĆ (Zabrocki 1981: 124–125). Kolejnym przybliżeniem rozważanej ramy będzie więc:

subj{np(nom)} + obj{np(acc)} + {np(dat)}

Zauważmy, że argument oznaczony jako biernikowy nie zawsze realizowany jest w tym przypadku. Może on być realizowany jako dopełniacz: obligatoryjnie, gdy jest argumentem zanegowanej formy czasownika TŁUMACZYĆ, i fakultatywnie, gdy negacja znajduje się wyżej w strukturze składniowej zdania, jak np. w: *Nie mogłem ci przecież kazać tłumaczyć mu swoje obawy / swoich obaw* (Przepiórkowski 2000). W literaturze generatywnej przypadki, których morfologiczna realizacja rozstrzyga się w zależności od kontekstu składniowego, nazywane są strukturalnymi (Chomsky 1986: 193), dlatego też odpowiedni argument będzie tutaj oznaczany jako obj{np(str)}, a za jego morfologiczną realizację odpowiedzialne będą reguły gramatyczne³.

Mniej oczywiste jest, że także podmiot należy oznaczyć jako występujący w przypadku strukturalnym. Związane jest to z kontrowersją dotyczącą wartości przypadku typowych fraz liczebnikowych w podmiocie (Przepiórkowski 1999, 2004, Saloni 2005, Miechowicz-Mathiasen i Witkoś 2007). Także wspomniane we Wprowadzeniu parsery języka polskiego przyjmują tutaj różne stanowiska: w Świgrze takie frazy są analizowane jako mianownikowe, zaś w gramatyce POLFIE – jako biernikowe. Stąd decyzja o oznaczaniu takich podmiotów jako subj{np(str)}, co prowadzi do następującego kolejnego przybliżenia rozważanej ramy walencyjnej:

subj{np(str)} + obj{np(str)} + {np(dat)}

Czasownik TŁUMACZYĆ może też występować z argumentem zdaniowym, np.: *Tłumaczyłem im, co mają zrobić*. Argumenty o formie zdań pytajnych, tzw. *pytajnozależne*,

³ Takie reguły są zdefiniowane na przykład w formalizmie Head-driven Phrase Structure Grammar (HPSG; Pollard i Sag 1987, 1994, Przepiórkowski i in. 2002) w pracy Przepiórkowski 1999 oraz w formalizmie LFG w pracy Patejuk i Przepiórkowski 2012a.

oznaczane są w słowniku jako {cp(int)}⁴. W tym wypadku mamy do czynienia z dopełnieniem pasywiwalnym (*Było im tłumaczone, co mają zrobić*), wydaje się więc, że oprócz ramy powyższej potrzebna jest też rama:

subj{np(str)} + obj{cp(int)} + {np(dat)}

Stosując jednak test koordynacji, otrzymujemy, że obie realizacje dopełnienia bliższego – nominalna i zdaniowa – wypełniają tę samą pozycję składniową: *Musiąłem im tłumaczyć najprostsze zasady i dlatego trzeba je stosować*. Powyższe dwie ramy walencyjne połączymy zatem w jedną, w której jedna z pozycji może mieć różne realizacje morfosyntaktyczne:

subj{np(str)} + obj{np(str); cp(int)} + {np(dat)}

W tym momencie jasna powinna być już wprowadzona wcześniej konwencja oznaczania poszczególnych pozycji składniowych nawiasami klamrowymi: pozycje takie traktowane są jako zbiory potencjalnie wielu specyfikacji możliwych realizacji morfoskładniowych. Pozycja ta może zresztą być także wypełniona frazą zdaniową z korelatem, jak w następującym zdaniu z Narodowego Korpusu Języka Polskiego, w którym taka fraza – *to, że pieniędzy tych nikt nie kradnie* – skoordynowana jest z frazą pytajnozależną: *Musiąłem tłumaczyć, jaka jest sytuacja finansowa placówki oraz to, że pieniędzy tych nikt nie kradnie*.

Pełna rama walencyjna (jedna z kilku) dla czasownika TŁUMACZYĆ w obecnej wersji słownika podana jest poniżej, z rozbiciem pozycji składniowych na osobne wiersze:

subj{np(str)} +
obj{np(str); cp(int); cp(że); cp(żeby); ncp(str,int); ncp(str,że)} +
{np(dat)} +
{np(inst); ncp(inst,że)}

Rama ta wyszczególnia dodatkowe realizacje dopełnienia bliższego: jako zdanie podrzędne wprowadzone przez spójnik ŻE (cp(że)) lub ŻEBY (cp(żeby)) oraz jako zdanie podrzędne z korelatem w przypadku strukturalnym (a więc w bierniku lub dopełniaczu): pytajnozależne (ncp(str, int)) lub wprowadzone przez ŻE (ncp(str, że)). Dziwić może

⁴ Skrót cp oznacza *complementizer phrase*, czyli zdanie wprowadzone spójnikiem podrzędnym, zaimek pytajnym lub partykułą pytajną *czy*, int precyzuje zaś, że chodzi tu właśnie o zdania pytajne (*interrogative*).

brak zdań z korelatem i spójnikiem *ŻEBY*, lecz przykłady typu **Tłumaczyłem to, żeby poszedł* wydają się nieakceptowalne przy założonej interpretacji (są akceptowalne, gdy *żeby poszedł* traktować jako osobny okolicznik zdaniowy).

Rama powyższa wprowadza także czwarty argument, realizowany jako narzędnikowa fraza nominalna lub fraza zdaniowa wprowadzona przez spójnik *ŻE* i korelat w narzędniku, a więc także jako koordynacja takich realizacji: *Rodzinie swoją nieobecność na urodzinach córki tłumaczył korkiem i tym, że nie zdążył na czas do domu*.

Podobnie jak w słowniku wyjściowym słownik zawiera wyłącznie maksymalne ramy walencyjne. Zakładamy, że każdy element ramy jest pomijalny, a więc dopuszczamy konstrukcje w różnym stopniu eliptyczne (z wyjątkami opisanymi w punkcie 6).

3. Podmioty i dopełnienia bliższe

Jak wspomniano powyżej, dopuszczamy w słowniku podmioty inne niż uzgadniające się z czasownikiem podmioty mianownikowe, przy czym głównym kryterium dla uznania takich argumentów nieakomodujących za podmioty jest możliwość koordynacji⁵ z niekontrowersyjnymi podmiotami mianownikowymi. Tak jest na przykład w wypadku czasownika *DZIWIĆ*, którego podmiot zdaniowy jest zaświadczony m.in. takim przykładem z NKJP: *...wtedy nie dziwi ta pozorna dualność praw i że nie można dojść do jedności...* Oprócz zilustrowanej tutaj możliwości wypełnienia pozycji podmiotu frazą nominalną w przypadku strukturalnym (mianowniku) i frazą zdaniową ze spójnikiem *ŻE*, możliwe są też realizacje z korelatem, stąd pełniejsza postać odpowiedniej ramy:

subj{np(str); cp(że); ncp(str,int); ncp(str,że)} +
{np(str)}

Zwróćmy uwagę na występowanie w tej ramie niepodmiotowego argumentu nominalnego o przypadku strukturalnym, {np(str)}, który jednak nie jest oznaczony jako dopełnienie bliższe, gdyż nie podlega on pasywizacji: **Maria była dziwiona pozorną dualnością praw / przez pozorną dualność praw*.

⁵ Nie bierzemy tu pod uwagę tzw. koordynacji leksykalno-semantycznej, jak w *Kto i kogo zaprosił?*; zob. Kallas 1993 oraz Przepiórkowski i Patejuk 2013.

4. Kontrola składniowa

Pojęcie kontroli składniowej⁶ łatwo wyjaśnić na przykładach: *Janek obiecał Markowi przyjść* i *Janek kazał Markowi przyjść*. Struktura obu jest podobna: czasowniki OBIECAĆ i KAZAĆ łączą się tutaj z trzema argumentami: podmiotem w mianowniku, *Janek*, argumentem celownikowym, *Markowi*, oraz frazą bezokolicznikową, *przyjść*. Jednak ich schemat semantyczny jest inny: pierwsze zdanie implikuje, że to Janek ma przyjść, a drugie, że przyjść ma Marek. Mówiąc żargonem generatywnym, różne argumenty form finitywnych *kontrolują* głęboki (nie wyrażony na powierzchni) podmiot czasownika PRZYJŚĆ – jest to podmiot w wypadku czasownika OBIECAĆ, lecz argument celownikowy w wypadku KAZAĆ.

Takie zależności oddane są w słowniku walencyjnym następująco:

OBIECAĆ:

```
subj, controller{np(str)} +  
{np(dat)} +  
controllee{infp( )}
```

KAZAĆ:

```
subj{np(str)} +  
controller{np(dat)} +  
controllee{infp( )}
```

Oprócz zapisów *controller* i *controllee*, oznaczających odpowiednio argument *kontrolujący* i argument, którego głęboki podmiot jest *kontrolowany*, w powyższych ramach pojawia się także specyfikacja *infp()*. Oznacza ona frazę bezokolicznikową o dowolnej wartości aspektu, w odróżnieniu od *infp(imperf)* oznaczającego wymaganie niedokonanej frazy bezokolicznikowej, jak w wypadku czasownika *dokończyć* (*Ktoś dokończył coś robić*, ale nie **Ktoś dokończył coś zrobić*):

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

⁶ Poczynając od prac Rosenbauma (1967, 1970), zjawisko to jest szeroko analizowane w literaturze generatywnej; zob. np. Landau (2013), a w kontekście języka polskiego np. Bondaruk (2004) i Witkoś i in. (2011a, b).

Zauważmy, że ze zjawiskiem kontroli są związane pewne fakty dotyczące uzgodnienia. W szczególności, gdy podrzędnym czasownikiem jest BYĆ, WYDAWAĆ SIĘ itp., to jego dopełnienie przymiotnikowe uzgadnia rodzaj i liczbę z niewyrażonym podmiotem tego czasownika, a więc z kontrolującym ten podmiot argumentem nadrzędnego czasownika: *Janek obiecał Marii być grzeczny(m) ale Janek kazał Marii być grzeczną*. Dlatego też opisaną tu reprezentację rozciągamy na inne przypadki uzgodnienia pomiędzy elementami ramy walencyjnej, jak w zdaniu *Piotrek uznał kwestię za zamkniętą*, gdzie dopełnienie bliższe *kwestię* uzgadnia rodzaj i liczbę z nominalnym składnikiem frazy przyimkowej *za zamkniętą*, oraz na konstrukcje do nich analogiczne, nawet jeżeli takie uzgodnienie nie występuje, jak w *Piotrek uznał dzieci za stratę czasu*. Częściowa rama odpowiadająca wystąpieniu czasownika UZNAĆ w *Piotrek uznał kwestię za zamkniętą* wygląda następująco (faktycznie dopełnienie bliższe może być także realizowane jako fraza zdaniowa z korelatem):

subj{np(str)} +
 obj, controller{np(str)} +
 controlee{prepadjp(za, acc)}

Wyjątkowo może się zdarzyć, że w schemacie jednego czasownika występują dwie instancje kontroli składniowej. W takim wypadku dla oznaczenia drugiej instancji stosowane są symbole controller2 i controlee2, jak w poniższej ramie dla czasownika UZNAĆ:

subj, controller{np(str)} +
 controlee, controller2{infp(_)} +
 controlee2{prepadjp(za, acc)}

Przykładem realizacji tego schematu może być zdanie *Prokuratura uznała za stosowne zająć się tym problemem*, w którym fraza bezokolicznikowa *zająć się tym problemem* jest kontrolowana przez podmiot *Prokuratura*, jednocześnie kontrolując dopełnienie przyimkowo-przymiotnikowe *za stosowne* (por. wcześniejsze zdanie *Piotrek uznał kwestię za zamkniętą*, w którym takie dopełnienie, *za zamkniętą*, jest kontrolowane przez dopełnienie bliższe *kwestię*).

Prezentowany tutaj formalizm opisu schematów walencyjnych uwzględnia także znaną dychotomię, którą można zilustrować czasownikami UWIELBIAĆ (*Janek uwielbia*

śpiewać) i ZACZYNAĆ (*Janek zaczyna śpiewać*). Choć w obu wypadkach podmiot czasownika (*Janek*) jest rozumiany jako semantycznie tożsamy z podmiotem dopełnienia okolicznikowego (*śpiewać*), to tylko czasownik ZACZYNAĆ może się łączyć z bezokolicznikami, które wymagają podmiotów innych niż np(str). Podmiotem ZACZYNAĆ nie musi być zatem fraza nominalna typu np(str), jak w wypadku czasownika UWIELBIAĆ, lecz – w zależności od czasownika w bezokoliczniku – może być to fraza zdaniowa typu wymaganego przez ten czasownik (*Jana zaczynało dziwić, że Maria wybiera Piotra*, ale nie **Jana uwielbiało dziwić, że Maria wybiera Piotra*; podobnie *Jana zaczynało kusić, żeby to zrobić*, ale nie **Jana uwielbiało kusić, żeby to zrobić*) czy też fraza bezokolicznikowa (*Być sumiennym zaczęło znaczyć być nieefektywnym*, ale nie **Być sumiennym uwielbiało znaczyć być nieefektywnym*). Podmiotem czasowników typu ZACZYNAĆ może też być frazeologizm wprowadzony przez czasownik w bezokoliczniku (*Serce zaczyna mu krwawić z powodu tego zdarzenia*, ale nie **Serce uwielbia mu krwawić z powodu tego zdarzenia*). W końcu, o ile argument bezokolicznikowy nie wymaga podmiotu, bez podmiotu może wystąpić także czasownik typu ZACZYNAĆ, ale nie zwykły czasownik kontroli (*Zaczyna mżyć*, ale nie **Uwielbia mżyć*).

W literaturze generatywistycznej czasowniki typu ZACZYNAĆ (także PRZESTAWAĆ, WYDAWAĆ SIĘ, MUSIEĆ i in.) określa się mianem *czasowników podnoszących* (ang. *raising verbs*)⁷. W proponowanym tu formalizmie podmioty takich czasowników oznaczane są jako subj, controller{E}⁸, np. dla ZACZYNAĆ:

subj, controller{E} + controllee{infp(imperf)}

w odróżnieniu od analogicznej ramy UWIELBIAĆ:

subj, controller{np(str)} + controllee{infp(imperf)}

⁷ W literaturze polonistycznej problem i możliwe rozwiązanie (we wspomnianym w przypisie 3 formalizmie HPSG) zostały opisane w artykułach Świdziński 1993 i Przepiórkowski 1997.

⁸ W obecnej wersji słownika walencyjnego pojawia się także zapis subj{E} oznaczający niewyrażony składniowo lecz obecny semantycznie podmiot czasowników typu PADAĆ i GRZMIEĆ, jak w zaświadczonym przykładzie (interpunkcja oryginalna): *W nocy mocno padało zacinając od strony północnej* (<http://forum.murator-dom.pl/archive/index.php/t-63469.html?s=1614428e01b591f52ea7ef3960f724f4>, sprawdzone 26 września 2013). Kwestia utrzymania takiego typu podmiotów w słowniku jest obecnie przedmiotem ożywionej dyskusji, nie będziemy więc ich tutaj opisywać, ani też nie uwzględniamy ich w dodatku.

5. Frazy tzw. przysłówkowe

Jeden z typów argumentów w słownikach walencyjnych, które były podstawą dla opisywanego tu słownika *Walenty*, to advp, czyli fraza przysłówkowa. Realizacjami takiego argumentu są na przykład przysłówki w następujących przykładach: *znajdować się daleko*, *stać się wczoraj*, *trwać dość długo* czy *zachowywać się zupełnie karygodnie*. Pozycje wypełnione przez takie argumenty mogą być zwykle realizowane także jako frazy przyimkowe pewnych typów (np. *znajdować się w Warszawie*, *stać się po obiedzie*), a czasami także jako frazy nominalne (*czekać dwie godziny*) czy nawet zdaniowe (*zachowywać się jakby kogoś coś ugryzło*, *zachowywać się jak Pan Bóg przykazał*). Możliwe realizacje morfoskładniowe danych argumentów „przysłówkowych” zależą przy tym od ich typów semantycznych: argumenty lokatywne mogą być realizowane jako frazy przyimkowe wprowadzone przez *w*, *na*, *przy* itp., ale nie *przez* czy *do*, które mogą natomiast wprowadzać argumenty perlatywne (*biec przez wieś*) czy adlatywne (*przywieźć do Zabrza*).

Dlatego też dla zwiększenia szczegółowości opisu w słowniku *Walenty* wprowadzone zostały bardziej precyzyjne oznaczenia tego typu argumentów, a mianowicie jako xp (TYP). Źródłem przyjętej tutaj nazwy xp jest użycie w składni generatywnej symbolu XP na określenie frazy dowolnego typu – w ten sposób uniknąć można sugestii, że omawiane pozycje realizowane muszą być przez przysłówki. Warto zwrócić też uwagę na to, że analogiczne argumenty „przysłówkowe” określone są mianem „okolicznikowych” w *Innym słowniku języka polskiego PWN* pod redakcją Mirosława Bańki (2000) i „nominalnych” w *Słowniku syntaktyczno-generatywnym języka polskiego* pod redakcją Kazimierza Polańskiego (1980–1992).

TYP może obecnie przyjmować jedną z wymienianych wartości:

- locat — miejsce, np. *być w sklepie*, *znajdować się przy drzwiach*,
- abl — punkt początkowy, np. *wyprowadzić ze strefy wojny*,
- adl — punkt końcowy, np. *przywieźć do Zabrza*,
- perl — trasa, np. *biec przez wieś*,
- temp — czas (punkt), np. *mieć miejsce wczoraj*,
- dur — długość trwania, np. *trwać dwie godziny*,
- mod — sposób, np. *traktować źle*, *zachowywać się jak dziecko*.

Podział zbliżony jest do klasyfikacji semantycznej fraz „nominalnych” w słowniku Polańskiego, choć podstawowym przyjętym tutaj kryterium podziału jest istnienie sto-

sunkowo spójnego zestawu realizacji składniowych każdej z klas. Poniżej podajemy kilka przykładów ram walencyjnych zawierających takie „przysłówkowe” specyfikacje argumentów:

BAWIĆ (*ktoś bawi gdzieś*):

subj{np(str)} + {xp(locat)}

BRAĆ SIĘ (*coś bierze się skądś*):

subj{np(str)} + {xp(abl)}

CIĄGNAĆ (*coś ciągnie kogoś gdzieś / do czegoś*):

subj{np(str)} + obj{np(str)} + {xp(adl)}

PRZEWOZIĆ (*ktoś przewozi coś czymś przez coś*):

subj{np(str)} + obj{np(str)} + {np(inst)} + {xp(perl)}

NASTĄPIĆ (*coś nastąpiło kiedyś*):

subj{np(str)} + {xp(temp)}

CIĄGNAĆ SIĘ (*coś ciągnęło się przez jakiś czas*):

subj{np(str)} + {xp(dur)}

OBCHODZIĆ SIĘ (*ktoś obchodzi się z kimś jakoś*)⁹:

subj{np(str)} + controller{prepnp(z,inst)} + controllee{xp(mod)}

W ostatnim przykładzie pojawia się typ argumentu znany już ze słownika wyjściowego, *prepnp(z, inst)*, czyli fraza przyimkowa z przyimkiem *z* (o formach *z* i *ze*) wymagającym narzędnika.

⁹ Istnienie elementu kontrolującego i kontrolowanego w poniższej ramie związane jest z możliwością realizacji frazy „przysłówkowej” sposobu przez frazę z formą *jak* – w takim wypadku wymagane jest ustalenie źródła uzgodnienia: *obchodzić się z kimś jak z dzieckiem/*jak dziecko, traktować kogoś jak pięcioletka/*pięcioletek* itp.

6. Konstrukcje frazeologiczne

Wśród typów argumentów przewidziano także frazeologizmy rzeczownikowe i przyimkowe, przy czym schematy zleksykalizowane były uwzględniane jedynie wówczas, gdy dana leksykalizacja nie była obejmowana przez schemat ogólny. Na przykład następujący schemat dla TRZYMAĆ, realizowany choćby w zdaniu *Cały czas trzymał ją za rękę*, obejmuje także leksykalizacje *trzymać kogoś za słowo* oraz *trzymać za kogoś kciuki*, które w związku z tym nie zostały w słowniku uwzględnione *explicite*.

subj{np(str)} + obj{np(str)} + {prepnp(za,acc)}

A zatem z walencyjnego punktu widzenia frazeologizm to konstrukcja składniowa, która nie jest opisana przez żaden ze schematów systemowo typowych, czyli taka, dla której trzeba wprowadzić odrębny schemat (por. Świdziński 1999). Na przykład dla wyrażenia *Dał nam znać, że wszystko w porządku* należałoby wprowadzić schemat czteroargumentowy, różny od trójargumentowego dla *Dał nam usnąć*. Schemat ten nakłada ograniczenia leksykalne na jeden z argumentów (*znać*). W przyszłości można by i dla zdania *Dał nam popalić*, realizującego wszak schemat systemowo typowy, zapisać specjalny schemat z leksykalnym ograniczeniem selekcyjnym na argument bezokolicznikowy.

Frazeologizmy rzeczownikowe zapisywane są według schematu lexnp(PRZYPADEK, LICZBA, 'LEMAT', MODYFIKACJA), gdzie MODYFIKACJA to jeden z symboli:

- atr — możliwa modyfikacja,
- natr — brak możliwości modyfikacji (*no attribute*),
- ratr — wymagany atrybut (dzierżawczy) (*required attribute*),
- batr — wymagany atrybut swój lub własny (*bound attribute*).

Na przykład jedna z ram czasownika ODBIJAĆ SIĘ wygląda następująco:

subj{np(str)} + {prepnp(w,loc)} + {lexnp(inst,sg,'echo',atr)}

Ramę tę ilustruje następujący przykład z NKJP: *Jubileuszowe uroczystości ku czci Elizy Orzeszkowej długo jeszcze będą odbijać się echem w opowieściach uczestników, w prasowych sprawozdaniach*. Wydawałoby się, że – zamiast podawać formę hasłową i informację o wartościach przypadku i liczby – odpowiedni argument można wyspecyfikować bezpośrednio jako ciąg *echem*. Skoro jednak realizacją takiego argumentu może być fraza zawierająca uzgadniające się z *echem* modyfikatory, np. *szerokim echem*, należy

jawnie podać wartości odpowiednich kategorii, co do których uzgodnienie następuje¹⁰. O możliwości pojawienia się takiego modyfikatora informuje ostatni element specyfikacji argumentu, atr.

Oczywiście nie zawsze taka modyfikacja jest możliwa. Na przykład w wyrażeniu *ktoś daje głowę, że...* czasownik DAWAĆ realizuje ramę, w której argument frazeologiczny opisany jest jako niemodyfikowalny (natr):

subj{np(str)} + {cp(że)} + {lexnp(str,sg,'głowa',natr)}

Przykład ten ilustruje jeszcze jedną cechę ram frazeologicznych: elementów leksykalnych nie można pominąć. Powyższej ramy nie można zredukować do subj{np(str)} + {cp(że)}, co pozwalałoby zaakceptować niepoprawne zdanie **Ktoś daje, że...*

Dwie pozostałe wartości atrybutu MODYFIKACJA realizowane są przede wszystkim we frazeologicznych argumentach mających postać frazy przyimkowej, zapisanych za pomocą schematu preplexnp(PRZYIMEK, PRZYPADEK, LICZBA, 'LEMAT', MODYFIKACJA). Na przykład jedna z ram czasownika OSZALEĆ, odpowiadająca wyrażeniu *Ktoś oszalał na czymś punkcie / na punkcie czegoś*, ma postać następującą:

subj{np(str)} + {preplexnp(na,loc,sg,'punkt',ratr)}

Modyfikacja jest tu oznaczona jako wymagana (ratr), gdyż nie jest poprawne wyrażenie **Ktoś oszalał na punkcie*. W tego typu frazeologizmach wymagany jest zwykle element dzierżawczy (*pójść w czyjeś ślady, świadczyć coś na rzecz kogoś, tracić w czyichś oczach*), ale nie zawsze tak jest, o czym świadczy choćby wyrażenie *Ktoś pokazał się (komuś) z jakiejś strony* (np. *ze złej strony*). Odpowiednia rama czasownika POKAZAĆ SIĘ podana jest poniżej, a więc obecna wersja słownika nie odróżnia wymagania modyfikacji dzierżawczej w argumencie frazeologicznym od wymagania dowolnej modyfikacji w takim argumencie:

subj{np(str)} + {np(dat)} + {preplexnp(z,gen,sg,'strona',ratr)}

Pewnym szczególnym przypadkiem wymagania modyfikacji dzierżawczej, wyszczególnianym w słowniku, jest wymaganie wystąpienia formy jednego z dwóch przymiotników: SWÓJ lub WŁASNY. Można na przykład *przyjąć kogoś pod swój/własny dach*, ale raczej

¹⁰ Rodzaj nie jest podawany jawnie, gdyż wynika z formy hasłowej. Dotychczas w słowniku nie pojawiły się frazeologizmy, w których w pozycji centralnej występowałby rzeczownik rozchwany rodzajowo.

nie *pod dach Janka*, można *coś słyszeć na własne uszy*, ale nie *na uszy czyjeś* itp. Odpowiednia rama walencyjna czasownika SŁYSZEĆ realizowana w takiej konstrukcji to:

subj{np(str)} + obj{np(str)} + {prelexnp(na, acc, pl, 'ucho', batr)}

Formalizm przewiduje także możliwość podania dosłownego brzmienia argumentu, np. *na kwaśne jabłko* w następującej ramie czasownika ZBIĆ:

subj{np(str)} + obj{np(str)} + {fixed('na kwaśne jabłko')}

7. Inne modelowane zjawiska

W powyższych punktach omówiono najważniejsze naszym zdaniem nowe elementy formalizmu opisu ram walencyjnych. Różnic w stosunku do słownika wyjściowego jest jednak więcej.

7.1. Zwrotność

Inaczej niż w słowniku Świdzińskiego 1994, 1998, ale podobnie jak w wersji wykorzystywanej przez parser Świgrę, oprócz wyróżnienia czasowników zwrotnych typu BAĆ SIĘ, zawarta jest także informacja o możliwych zwrotnych realizacjach argumentów pozostałych czasowników, np. czasownika MYĆ. Decyzja o tym, czy potraktować czasownik jako zwrotny, tj. hasłowany z SIĘ, czy jako zwykły o możliwych zwrotnych realizacjach argumentu biernikowego (oznaczanych refl), podejmowana jest na podstawie testu podstawienia: jeżeli możliwe jest zastąpienie elementu *się* frazą typu SIEBIE SAMEGO, czasownik nie jest oznaczany jako zwrotny. Zwrotny jest zatem BAĆ SIĘ, bo nie można powiedzieć **On boi siebie samego*, ale nie MYĆ, bo poprawne jest zdanie *On myje siebie samego*.

7.2. Realizacje zaimkowe

W nowym formalizmie opisu ram walencyjnych wyróżnione zostały dwie sytuacje, w których czasownik zwykle łączący się z frazą zdaniową realizuje to wymaganie za pomocą frazy zaimkowej.

Po pierwsze, czasowniki takie jak DECYDOWAĆ, MAWIAĆ, SĄDZIĆ, TWIERDZIĆ i UWAŻAĆ mogą łączyć się z zaimkami przysłówkowymi typu TAK i JAK, np. *Jak uważasz?*, *Tak*

sądzę. Możliwość taka zasygnalizowana jest za pomocą symbolu *advp(pron)*, np. dla czasownika TWIERDZIĆ:

subj{np(str)} + {advp(pron)}

Zauważmy, że wymaganie to występuje w osobnej ramie niż wymaganie frazy zdaniowej, gdyż wątpliwa jest akceptowalność zdań typu *?*Sądzę tak i że nie przyjdzie*.

Po drugie, większa (w obecnej wersji słownika) grupa czasowników dopuszcza jako realizacje wymagania frazy zdaniowej zaimki rzeczowne typu *CO, COŚ, NIC, TO (SAMO)* itp., np. *Co myślisz?, Nic nie odpowiedział, Pomyślałem to samo* itp. Argumenty takie – za artykułem Postala 2004, który analizował podobne zjawisko w języku angielskim – nazywamy niechromatycznymi (ang. *nonchromatic*) i oznaczamy jako *nonch*, jak w poniższej ramie dla czasownika POMYŚLEĆ:

subj{np(str)} + {prepn(o,loc)} + {nonch}

7.3. Przyimki złożone

Opisywany format słownika przewiduje także argumenty wprowadzone przez przyimki złożone typu *CO DO, NA TEMAT, Z POWODU* itp. W odróżnieniu od specyfikacji zwykłych fraz przyimkowych, w których podana jest zarówno forma hasłowa przyimka, jak i rządony przez niego przypadek, np. *prepn(z, gen)* czy *prepn(z, inst)*, w opisach fraz wprowadzonych przez zaimki złożone występuje tylko ich forma, gdyż nie zaobserwowano możliwości różnych realizacji przypadkowych argumentów przyimków złożonych o tej samej formie – wszystkie wprowadzone dotychczas wymagają dopełniacza. Przykładem takiego opisu jest *comprepn(z powodu)* w następującej ramie czasownika ROZPACZAĆ:

subj{np(str)} + {comprepn(z powodu)}

7.4. Inne oznaczenia przypadku

Powyżej wprowadzone zostało oznaczenie *str* na tzw. wymaganie przypadku strukturalnego, tj. takiego, którego realizacja morfologiczna zależy od kontekstu składniowego. Podobnie wprowadzamy oznaczenie *part* dla przypadku partytywnego, realizowanego jako biernik lub dopełniacz. Tak określone jest wymaganie przypadkowe

czasownika DAĆ wobec jego dopełnienia bliższego w zdaniu *Dajcie wina i całą świnię!* (Przepiórkowski 1999: 175):

subj{np(str)} + obj{np(part)} + {np(dat)}

Taki jest też przypadek formy leksemu WODA we frazeologicznej ramie czasownika NABRAĆ:

subj{np(str)} +
{lexnp(part, sg, 'woda', natr)} +
{prelexnp(w, acc, pl, 'usta', natr)}

Związany z przypadkiem strukturalnym jest przypadek predykatywny, pred, czyli przypadek pewnych wymaganych fraz przymiotnikowych, np. przy czasownikach czuć SIĘ, PAMIĘTAĆ CZY WYDAWAĆ SIĘ. Może on być realizowany jako narzędnik lub przypadek uzgadniający, np. *Pamiętam go młodym* lub *Pamiętam go młodego*, przy czym w wypadku uzgodnienia z podmiotem liczebnikowym przypadek może być rozchwiany: *Pięć osób zostało rannych*, *Kolejnych jedenaście zarzutów było podobnych* (dopełniacz; przykłady z NKJP) wobec *... następne kilkadziesiąt metrów było czyste* (ten sam przypadek strukturalny co *kilkadziesiąt*; przykład z NKJP), obok narzędnika (*Wielu proroków było bezżennymi...*; NKJP). Argument uzgadniający się z taką frazą przymiotnikową oznaczony jest za pomocą opisanego wcześniej schematu związanego ze zjawiskiem kontroli, jako controller. Odpowiednia rama dla czasownika WYDAWAĆ SIĘ będzie więc miała postać następującą:

subj, controller{np(str); ncp(str, int); ncp(str, że); ncp(str, żeby)} +
controllee{adjp(pred)} +
{np(dat)}

Rama ta uwzględnia realizację podmiotu przez frazę zdaniową z korelatem oraz wprowadza trzeci argument – celownikowy (np. *To, że umie grać na banjo, wydawało mi się mało istotne*).

Ostatnią specyfikacją przypadku nie odwołującą się bezpośrednio do jednego z siedmiu przypadków morfologicznych w języku polskim jest postp, czyli „przypadek po-przymiokowy” form przymiotnikowych we frazach po-przymiokowo-przymiokowych (prepadjp) wprowadzonych przez po, np. w *po polsku* czy *po męsku*. Takie frazy wymagane są przez czasowniki typu CZYTAĆ (*po angielsku*), POROZUMIEWAĆ SIĘ (*po polsku*) itp., np. w następującej ramie czasownika UMIEĆ (*po chińsku*):

subj{np(str)} + {prepadjp(po,postp)}

8. Dane liczbowe

Do opisanego tutaj słownika walencyjnego informacje dodawane są dwuetapowo: każde opracowane przez leksykografów hasło podlega weryfikacji przez jednego z (obecnie) dwóch „superleksykografów” panujących nad spójnością i jakością całego słownika. Dlatego też w każdym etapie rozwoju słownika pełne hasła mają jeden z dwóch statusów: gotowe (tj. opracowane przez leksykografa) lub sprawdzone (przez „superleksykografa”).

Hasła w Walentym odpowiadają formie bezokolicznika bez uwzględniania zwrotności i aspektu. Na przykład hasło NARZUCAĆ obejmuje cztery podhasła: niedokonane NARZUCAĆ (*komuś własne zdanie itp.*), niedokonane NARZUCAĆ SIĘ (*komuś*), dokonane NARZUCAĆ (*liści na stertę*) oraz dokonane NARZUCAĆ SIĘ (*piłką na boisku*). Obecnie, na początku września 2013, słownik liczy 6965 takich podhaseł (w tym 3924 sprawdzone przez „superleksykografów”) pogrupowanych w 4811 haseł (w tym 2683 zweryfikowane¹¹). Ram walencyjnych jest 24 199 (14 874 sprawdzonych), a więc ok. 3,5 ramy na jedno podhasło (3,8 w części zweryfikowanej). Dalsze dane liczbowe podane poniżej dotyczą wszystkich haseł oznaczonych jako gotowe, a nie tylko tych ostatecznie zweryfikowanych.

Ramy walencyjne oznaczane są jako pewne, potoczne, archaiczne i wątpliwe. Liczba ram w rozbiciu na te kategorie przedstawia się następująco¹²:

- pewne: 21 931,
- potoczne: 1111,
- archaiczne: 220,
- wątpliwe: 865.

Pozycji składniowych w całym słowniku jest 68 874, a więc ok. 2,8 pozycji na ramę walencyjną. Należy jednak pamiętać, że jedna pozycja może zawierać – oddzielone średnikiem – różne specyfikacje morfoskładniowe. Liczba ram zawierających pozycje z kilkoma specyfikacjami morfoskładniowymi to 3148, a więc ok. 13%. Ta liczba pokazuje,

¹¹ Por. wspomniane we Wprowadzeniu 1440 haseł w słowniku wyjściowym.

¹² Niewielka liczba ram, obecnie około 70, oznaczana jest kwalifikatorem „złe”, oznaczającym pewne użycia nienormatywne.

że możliwość różnych morfoskładniowo realizacji danej pozycji składniowej nie jest zjawiskiem marginalnym¹³.

9. Zakończenie

Skrócony opis słownika walencyjnego i kolejne jego wersje, w tym wersja opisana liczbowo w poprzednim punkcie, znajdują się pod adresem <http://clip.ipipan.waw.pl/Walenty>.

Choć obecnie słownik ma charakter powierzchniowskładniowy, już w takiej postaci może on być użyteczny w automatycznej analizie składniowej zdań polskich. Na przykład dzięki powierzchniowym informacjom walencyjnym możliwe jest stwierdzenie, że zaimek *ją* jest argumentem czasownika finitywnego *uczyłem* w zdaniu *Uczyłem ją pisać*, zaś w zdaniu *Chciałem ją poznać*, nie różniącym się interpretacjami morfoskładniowymi kolejnych słów – argumentem bezokolicznika *poznać*.

Podobnie możliwe jest stwierdzenie, czy dany podrzędnik jest raczej okolicznikiem, jak w zdaniu *Czekałem 2 godziny*, czy też argumentem, jak w zdaniu *Przeczekałem 2 godziny*.

Słownik jest intensywnie rozwijany w ramach projektów i zadań realizowanych w IPI PAN. Pierwsza wersja została udostępniona w wyniku projektu europejskiego CESAR (<http://clip.ipipan.waw.pl/CESAR>), kolejne powstały w trakcie realizacji projektu POIG NEKST (<http://zil.ipipan.waw.pl/NEKST>), obecnie zaś trwają prace związane z projektem infrastrukturalnym CLARIN-PL, którego realizacja rozpoczęła się w połowie roku 2013. W ramach tego projektu słownik ma kilkakrotnie zwiększyć swoją objętość, a także – co bardziej istotne w kontekście niniejszego artykułu – dodane zostaną różnorodne informacje semantyczne. Planowane jest przede wszystkim powiązanie informacji morfoskładniowych z głęboką strukturą argumentów, co z jednej strony będzie wymagać podziału ram według znaczeń, a z drugiej – wiązania ze sobą ram stanowiących różne realizacje powierzchniowe tej samej struktury głębokiej. W projekcie przewidziane jest także przypisanie argumentom preferencji selekcyjnych (czyli na przykład informacji

¹³ Z drugiej strony liczba ta obejmuje także stosunkowo regularne sytuacje, gdy dana pozycja może być realizowana przez frazę nominalną lub przyimkową i przez odpowiadające jej frazy zdaniowe z korelatem. Gdy takich sytuacji nie uwzględniać, ram z pozycjami o kilku idiosynkratycznie różnych specyfikacjach morfoskładniowych jest 585, a więc 2,4%.

o tym, że podmiotem czasownika RZĘĆ jest zwykle fraza oznaczająca konia). Badania związane z frazeologicznymi ramami walencyjnymi planowane są także w sieci naukowej PARSEME (<http://zil.ipipan.waw.pl/PARSEME>), zaś aspekty informatyczne probabilistycznego przetwarzania tekstów z wykorzystaniem ram walencyjnych stanowią temat powstających w IPI PAN rozpraw doktorskich.

Podziękowania

Jak wspomniano we Wprowadzeniu, słownik Walenty powstaje w ramach projektów CESAR (projekt europejski typu ICT-PSP PB Pilot Type B, „Central and South-east EuropeAn Resources”, 01.02.2011 – 31.01.2013, <http://cesar.nytud.hu/>), NEKST (projekt typu PO IG 1.1.2, „Adaptacyjny system wspomagający rozwiązywanie problemów w oparciu o analizę treści dostępnych źródeł elektronicznych”, 01.04.2009 – 25.02.2014, <http://www.ipipan.waw.pl/nekst/>) i CLARIN-PL („Polska część infrastruktury naukowej CLARIN ERIC: Wspólne zasoby językowe i infrastruktura technologiczna”, 01.04.2013 – 31.12.2015, zob. <http://www.clarin.eu/>) i podziękowania należą się przede wszystkim odpowiednim instytucjom finansującym niniejsze badania.

Za obsługę informatyczną projektu odpowiedzialny jest Bartłomiej Nitoń, autor narzędzi do tworzenia i przeszukiwania słownika walencyjnego. Funkcję „superleksykografów” pełnią obecnie Jakub Szymczak i Sebastian Żurowski, którym jesteśmy wdzięczni za dyskusje i komentarze do wcześniejszej wersji artykułu, zaś wśród pozostałych leksykografów wymienić należy przede wszystkim następujących: Piotr Batko, Joanna Filipczak, Anna Grzeszak, Marta Kaluźna, Marcin Opacki, Paulina Rosalska, Maciej Zagondek. Wszystkim powyższym i innym osobom zaangażowanym w rozwój słownika serdecznie dziękujemy.

Literatura

- Bańko M., red., 2000, *Inny słownik języka polskiego*, Wydawnictwo Naukowe PWN, Warszawa.
- Bondaruk A., 2004, *PRO and Control in English, Irish and Polish — A Minimalist Analysis*, Wydawnictwo KUL, Lublin.

- Bresnan J., red., 1982, *The Mental Representation of Grammatical Relations*, The MIT Press, Cambridge, MA.
- Chomsky N., 1986, *Knowledge of Language: Its Nature, Origin, and Use*, Praeger, Westport.
- Dalrymple M., 2001, *Lexical-Functional Grammar*, Academic Press, San Diego, CA.
- Hajnicz E., Woliński M., 2009, How valence information influences parsing Polish with *Świgr*, [w:] *Recent Advances in Intelligent Information Systems*, red. M. A. Kłopotek, A. Przepiórkowski, S. T. Wierzchoń, K. Trojanowski, Akademicka Oficyna Wydawnicza EXIT, Warszawa, s. 193–206.
- Kallas K., 1993, *Składnia współczesnych polskich konstrukcji współrzędnych*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- Landau I., 2013, *Control in Generative Grammar: A Research Companion*, Cambridge University Press, Cambridge.
- Miechowicz-Mathiasen K., Witkoś J., 2007, O podmiotach i ich przypadkach (w nawiązaniu do artykułów Zygmunta Saloniego i Adama Przepiórkowskiego), *Biuletyn Polskiego Towarzystwa Językoznawczego* LXIII, s. 101–114.
- Patejuk A., Przepiórkowski A., 2012a, A comprehensive analysis of constituent coordination for grammar engineering, [w:] *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, Mumbai, India, s. 2191–2207.
- Patejuk A., Przepiórkowski A., 2012b, Towards an LFG parser for Polish: An exercise in parasitic grammar development, [w:] *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, ELRA*, Istanbul, Turkey, s. 3849–3852.
- Polański K., red., 1980–1992, *Słownik syntaktyczno-generatywny czasowników polskich*, Zakład Narodowy im. Ossolińskich / Instytut Języka Polskiego PAN, Wrocław / Kraków.
- Pollard C., Sag I. A., 1987, *Information-Based Syntax and Semantics, Volume 1: Fundamentals*, CSLI Publications, Stanford, CA.
- Pollard C., Sag I. A., 1994, *Head-driven Phrase Structure Grammar*, Chicago University Press / CSLI Publications, Chicago, IL.
- Postal P. M., 2004, Chromaticity: An overlooked English grammatical category distinction, [w:] *Skeptical Linguistic Essays*, red. P. M. Postal, Oxford University Press,

- Oxford, s. 138–158.
- 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, Tybinga.
- Przepiórkowski A., 2000, Long distance genitive of negation in Polish, *Journal of Slavic Linguistics* 8, s. 151–189.
- Przepiórkowski A., 2004, O wartości przypadku podmiotów liczebnikowych, *Biuletyn Polskiego Towarzystwa Językoznawczego LX*, s. 133–143.
- Przepiórkowski A., Kupść A., Marciniak M., Mykowiecka A., 2002, Formalny opis języka polskiego: Teoria i implementacja, Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- Przepiórkowski A., Bańko M., Górski R. L., Lewandowska-Tomaszczyk B., red., 2012, *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa.
- Przepiórkowski A., Patejuk A., 2013, Koordynacja leksykalno-semantyczna w systemie współczesnej polszczyzny. Artykuł prezentowany w ramach konferencji *Bogactwo współczesnej polszczyzny*, Kraków, 12–13 kwietnia 2013, złożony do druku.
- Rosenbaum P., 1967, *The Grammar of English Predicate Complement Constructions*, The MIT Press, Cambridge, MA.
- Rosenbaum P., 1970, A principle governing deletion in English sentential complementations, [w:] *Readings in English Transformational Grammar*, red. R. Jacobs P. Rosenbaum, Ginn, Waltham, MA, s. 20–29.
- Saloni Z., 2005, O przypadkach w języku polskim (na marginesie artykułu Adama Przepiórkowskiego), *Biuletyn Polskiego Towarzystwa Językoznawczego LXI*, s. 27–48.
- Świdziński M., 1992a, *Gramatyka formalna języka polskiego*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.
- Świdziński M., 1992b, Realizacje zdaniowe podmiotu-mianownika, czyli o strukturalnych ograniczeniach selekcyjnych, [w:] *Opisać słowa*, red. A. Markowski, Dom Wydawniczy Elipsa, s. 188–201.

- Świdziński M., 1993, Dalsze kłopoty z bezokolicznikiem, [w:] *Językoznawstwo synchroniczne i diachroniczne*, red. J. Sambor, J. Linde-Usiekniewicz, R. Huszcza, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa, s. 303–314.
- Świdziński M., 1994, *Syntactic dictionary of Polish verbs*. Maszynopis, Uniwersytet Warszawski i Universiteit van Amsterdam.
- Świdziński M., 1998, *Syntactic dictionary of Polish verbs*. Version 3a. Maszynopis, Uniwersytet Warszawski.
- Świdziński M., 1999, O ograniczeniach aparatu pojęciowego schematów zdaniowych, [w:] *Studia Lingwistyczne Ofiarowane Profesorowi Kazimierzowi Polańskiemu na 70-lecie Jego Urodzin*, red. W. Banyś, L. Bednarczuk, S. Karolak, tom 1775 serii *Prace Naukowe Uniwersytetu Śląskiego w Katowicach*, Wydawnictwo Uniwersytetu Śląskiego, Katowice, s. 187–197.
- Świdziński M., Woliński M., 2010, Towards a bank of constituent parse trees for Polish, [w:] *Text, Speech and Dialogue: 13th International Conference, TSD 2010*, Brno, Czech Republic, red. P. Sojka, A. Horák, I. Kopeček, K. Pala, tom 6231 serii *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Heidelberg, s. 197–204.
- Witkoś J., Cegłowski P., Snarska A., Żychliński S., 2011a, *Minimalist Facets of Control. An English-Polish Comparative Overview of Gerunds and Infinitives*, Wydawnictwo Naukowe UAM, Poznań.
- Witkoś J., Cegłowski P., Snarska A., Żychliński S., 2011b, *Wybrane aspekty derywacji i interpretacji wyrażen bezokolicznikowych. Minimalistyczne studium porównawcze polsko-angielskie*, Wydawnictwo Naukowe UAM, Poznań.
- Woliński M., 2004, *Komputerowa weryfikacja gramatyki Świdzińskiego*, rozprawa doktorska, Instytut Podstaw Informatyki PAN, Warszawa.
- Zabrocki T., 1981, *Lexical Rules of Semantic Interpretation: Control and NP Movement in English and Polish*, Uniwersytet im. Adama Mickiewicza, Poznań.

Dodatek: Specyfikacja formatu słownika Walenty

Niniejszy dodatek opisuje format postaci tekstowej Walentego obowiązujący na początku września 2013.

Każdy wiersz zawiera trzy pola oddzielone dwukropkiem i spacją, zawierające formę hasłową, informację o aspekcie i ramę walencyjną, na przykład¹⁴:

```
cisnąć się: imperf: subj{np(str)} + {xp(locat)}  
cisnąć się: perf: subj{np(str)} + {xp(adl)}  
czuć: imperf: subj{np(str)} + {cp(jak); cp(że)}  
darować: _: subj{np(str)} + {np(dat)} + {cp(że)}  
żywić: imperf: subj{np(str)} + obj{np(str)} + {np(inst)}  
żywić: imperf: subj{np(str)} + obj{np(str)} + {prepn(wobec, gen)}  
żywić się: imperf: subj{np(str)} + {np(inst)}
```

W wypadku czasowników posiadających więcej niż jedną ramę walencyjną każda taka rama jest wypisywana w osobnym wierszu, z powtórzeniem formy hasłowej i specyfikacji aspektu, jak dla żywić powyżej.

Wartością pola aspektowego może być imperf (niedokonany), perf (dokonany) lub symbol _ (czasownik dwuaspektowy z tą samą ramą w obu aspektach; zob. DAROWAĆ powyżej).

Trzecie pole zawiera specyfikację 1 ramy i jest ciągiem specyfikacji pozycji składniowych oddzielonych plusem +.

Każda specyfikacja pozycji składniowej to zbiór możliwych realizacji morfoskładniowych (typów frazy) danej pozycji zawarty w nawiasach klamrowych { i }. Przed nawiasami klamrowymi może wystąpić oznaczenie podmiotu (subj; najwyżej jeden w ramie), dopełnienia bliższego (obj; najwyżej jedno), argumentu kontrolującego (controller; najwyżej jeden) i argumentu kontrolowanego (controllee; zwykle nie więcej niż jeden, w wyjątkowych wypadkach dwa). Jeżeli argument jednocześnie odpowiada wyróżnionej funkcji gramatycznej i bierze udział w kontroli, to odpowiednie symbole przed nawiasami klamrowymi oddzielone są przecinkami, np.: subj , controller{np(str)}. W wyjątkowych wypadkach może się także pojawić oznaczenie drugiego argumentu kontrolującego (controller2; najwyżej jeden) i argumentu przezeń kontrolowanego

¹⁴ Dla każdego z czasowników podane zostały tylko wybrane ramy walencyjne.

(controllee2; najwyżej jeden), zob. §4, przy czym argument kontrolowany w jednej instancji kontroli może jednocześnie być elementem kontrolującym w drugiej instancji, np.: obj, controllee, controller2{infp(_)}.

Możliwe typy fraz to:

- np (PRZYPADEK) – fraza nominalna w danym przypadku (możliwe wartości to: nom, gen, dat, acc, inst, str, part);
- prepnp (PRZYIMEK, PRZYPADEK) – fraza przyimkowa z konkretnym przyimkiem i wymaganym przez niego przypadkiem (dodatkowa możliwa wartość to loc, ale nie part);
- adjp (PRZYPADEK) – fraza przymiotnikowa w danym przypadku (także pred, ale nie part);
- prepadjp (PRZYIMEK, PRZYPADEK) – fraza przyimkowo-przymiotnikowa z konkretnym przyimkiem i wymaganym przez niego przypadkiem. W wypadku frazy zbudowanej z przyimka po i poprzyimkowej formy specjalnej przymiotnika (np. *polsku*) używamy specjalnej wartości postp przypadku;
- comprepnp (PRZYIMEK_ZŁOŻONY) – fraza przyimkowa z danym przyimkiem złożonym (zob. §7.3);
- lexnp (PRZYPADEK, LICZBA, 'LEMAT', MODYFIKACJA) – frazeologizm rzeczownikowy (zob. §6), z zadaniem przypadkiem i liczbą (sg, pl lub dowolna _) rzeczownika określonego lematem; możliwe wartości modyfikacji to atr, natr, batr i ratr;
- preplexnp (PRZYIMEK, PRZYPADEK, LICZBA, 'LEMAT', MODYFIKACJA) – frazeologizm przyimkowy (zob. §6);
- fixed('FRAZEOLOGIZM') – frazeologizm o ustalonym brzmieniu (zob. §6);
- cp (TYP) – fraza zdaniowa konkretnego typu: że, żeby, żeby2 (dla fraz zdaniowych ze spójnikiem że, ale także ze spójnikiem ŻEBY, gdy czasownik jest zaprzeczony), gdy (realizowane przez GDYBY, gdy czasownik jest w trybie przypuszczającym), jak, kiedy, jeśli, aż, jakby, jakoby, int, czy;
- prepncp (PRZYIMEK, PRZYPADEK, TYP) – fraza zdaniowa z korelatem z przyimkiem – tak samo jak w wypadku frazy zdaniowej, ale dodatkowo oznaczony jest przypadek, w którym występuje korelat oraz przyimek, który tego przypadku wymaga;
- ncp (PRZYPADEK, TYP) – fraza zdaniowa z korelatem bez przyimka;
- infp (ASPEKT) – fraza bezokolicznikowa, w nielicznych wypadkach wymaga oznacze-

nia aspektu czasownika w bezokoliczniku (imperf lub perf); jeśli aspekt jest dowolny, to w słowniku zostaje to oznaczone symbolem *_*, czyli *infp(_)*;

- *xp(SEM)* – fraza „przysłówkowa” o semantyce danego typu (zob. §5): *locat, abl, adl, perl, temp, dur, mod*;
- *advp(misc)* – fraza przysłówkowa (nie dopuszczająca realizacji przyimkowo-nominalnych itp.);
- *advp(pron)* – zaimek przysłówkowy (zob. §7.2);
- *nonch* – fraza niechromatyczna (zob. §7.2);
- *or* – mowa niezależna;
- *refl* – partykuła *się* realizująca zwrotne użycie czasownika (zob. §7.1);
- *E* – podmiot *podniesiony*, tj. w pełni współdzielony z podmiotem (lub jego brakiem) argumentu bezokolicznikowego (zob. §4); tylko w kontekście *subj, controller{E}* (ale por. przypis 8 na s. 10).