

ADAM PRZEPIÓRKOWSKI, FILIP SKWARSKI, ELŻBIETA HAJNICZ,
AGNIESZKA PATEJUK, MAREK ŚWIDZIŃSKI, MARCIN WOLIŃSKI

Modelowanie własności składniowych czasowników w nowym słowniku walencyjnym języka polskiego

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: Świga (Woliński 2004, Hajnicz i Woliński 2009, Świdziński i Woliński 2010; <http://zil.ipipan.waw.pl/Skladnica>), 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 Świga i dostępny w dystrybucji systemu pod adresem podanym powyżej.

Poniższe przykłady schematów walencyjnych ilustrują format tego słownika wyjściowego:

wydawać V np (bier)+np (cel)
wydawać V np (bier)+prepnp (' za' ,bier)
wydawać V prepnp (' na' ,bier)+np (bier)

Należy przede wszystkim zauważyć, że jest to słownik powierzchniowoskładniowy, nie-rozróżniający sensów słów: pierwszy z powyższych schematów może być użyty 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 formy hasłowej czasownika, 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, cel, bier, narz), na przykład: *Jaś ma samochód* (np(bier));
- prepnp(' PRZYIMEK' , PRZYPADEK) — fraza przyimkowa z konkretnym przyimkiem i wymaganym przez niego przypadkiem (dodatkowa możliwa wartość to miej), na przykład: *Jaś idzie do szkoły* (prepnp(' 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(zabier). W przypadku 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/ZEBY (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 przypadku 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, na przykład: *Napoleon ogłosił się cesarzem* (użycie zwrotne) wobec *Biskup ogłosił Napoleona cesarzem* (użycie zwykłe) — por. pkt § 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

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

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 powierzchniowo-składniowych informacji walencyjnych, lecz także istotne zwiększenie rozmiaru słownika oraz przypisanie schematom 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 przypadkowa argumentów nominalnych w niektórych pozycjach,
- podmioty nienominalne,
- ograniczenia możliwości utworzenia strony biernej (pasywizacji),
- relacje 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 słownika (w odróżnieniu od wewnętrznej, bazodanowej).

2. Formalizm opisu — przykład

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

W zdaniach typu *Ktoś tłumaczy coś komuś* mamy do czynienia ze schematem, który w najprostszym sposobie można wyrazić tak²:

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

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

Podobne rozważania prowadzą do jawnego oznaczania argumentów podlegających pasywizacji, tj. przechodzących na pozycje podmiotu w wyniku utworzenia strony biernej:

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

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 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, WYSLUCHAĆ; a może też od niektórych czasowników łączących się z celownikiem, np. ZAGROZIĆ (Zabrocki 1981: 124–125). Kolejnym przybliżeniem rozważanego schematu 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 TLUMACZYĆ, 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 oznaczony 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 rozważanego schematu walencyjnego:

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

Czasownik TLUMACZYĆ może też występować z argumentem zdaniowym, np.: *Tłumaczyłem im, co mają zrobić*. Argumenty o formie zdań pytajnych, tzw. pytajnozależne, oznaczane są w słowniku jako { cp (int) }⁴. W tym przypadku mamy do czynienia z dopełnieniem pasywizowalnym (*Było im tłumaczone, co mają zrobić*), wydaje się więc, że oprócz powyższego schematu potrzebny jest też schemat:

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

³ Takie reguły są zdefiniowane np. 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.

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

Stosując jednak test koordynacji, przekonujemy się, że obie realizacje dopełnienia bliższego — nominalna i zdaniowa — wypełniają tę samą pozycję składniową: *Musiałem im tłumaczyć najprostsze zasady i dlatego trzeba je stosować*. Powyższe dwa schematy walencyjne połączymy zatem w jeden, w którym jedna z pozycji może mieć różne realizacje morfosyntaktyczne:

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

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 pytajnozęlezną: *Musiałam tłumaczyć, jaka jest sytuacja finansowa placówki oraz to, że pieniędzy tych nikt nie kradnie*.

Pełny schemat walencyjny (jeden z kilku) dla czasownika TŁUMACZYĆ w obecnej wersji słownika podany jest poniżej, z rozbiem pozycji składniowych na osobne wiersze:

$$\begin{aligned} &\text{subj}\{\text{np}(\text{str})\} + \\ &\text{obj}\{\text{np}(\text{str}); \text{cp}(\text{int}); \text{cp}(\text{że}); \text{cp}(\text{żeby}); \text{npc}(\text{str}, \text{int}); \text{npc}(\text{str}, \text{że})\} + \\ &\{\text{np}(\text{dat})\} + \\ &\{\text{np}(\text{inst}); \text{npc}(\text{inst}, \text{że})\} \end{aligned}$$

Schemat ten wyszczególnia dodatkowe realizacje dopełnienia bliższego: jako zdanie podrzędne wprowadzone przez spójnik ŻE ($\text{cp}(\text{że})$) lub ŻEBY ($\text{cp}(\text{żeby})$) oraz jako zdanie podrzędne z korelatem w przypadku strukturalnym (a więc w bierniku lub dopełniaczu): pytajnozęleżne ($\text{npc}(\text{str}, \text{int})$) lub wprowadzone przez ŻE ($\text{npc}(\text{str}, \text{że})$). Dziwić może 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* traktuje się jako osobny okolicznik zdaniowy).

Schemat powyższy 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: *Rodzinnie swoją nieobecność na urodzinach córki tłumaczył korkiem i tym, że nie zdążył na czas do domu*.

Podobnie jak w słowniku wyjściowym notowane są wyłącznie maksymalne schematy walencyjne. Zakładamy, że każdy element schematu jest pomijalny, a więc dopuszczamy konstrukcje w różnym stopniu eliptyczne (z wyjątkami opisanymi w pkt § 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

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

z niekontrowersyjnymi podmiotami mianownikowymi. Tak jest np. w przypadku 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 odpowiedniego schematu:

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

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

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 (niewyrażony 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 schematach pojawia się także specyfikacja *infp (_)*. Oznacza ona frazę bezokolicznikową o do-

⁶ Począwszy 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, 2011b).

wolnej wartości aspektu, w odróżnieniu od *infp* (*imperf*) oznaczającego wymaganie niedokonanej frazy bezokolicznikowej, jak w wypadku czasownika DOKOŃCZYĆ (*Ktoś dokończył coś zrobić*, ale nie **Ktoś dokończył coś zrobić*):

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

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 schematu walencyjnego, jak w zdaniu *Piotrek uznał kwestię za zamkniętą*, gdzie dopełnienie bliższe *kwestię* uzgadnia rodzaj i liczbę z nominalnym składnikiem frazy przyimkowej *za zamkniętą*. Częściowy schemat odpowiadający 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)} +  
controllee{preadjp(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 *controllee2*, jak w poniższym schemacie dla czasownika UZNAĆ:

```
subj, controller{np(str)} +  
controllee, controller2{infp(_)} +  
controllee2{preadjp(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 bezokolicznikowego (*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 analogicznego schematu UWIELBIAĆ:

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

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 np. 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 ktoś coś ugryzło, zachowywać się jak Pan Bóg przykazał*). Możliwe realizacje morfolożniowe 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*).

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 niewyraźny 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>, dostęp: 26.09.2013 r.). 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.

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 można uniknąć 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 Banki (2000) i „nominalnych” w *Słowniku syntaktyczno-generatywnym języka polskiego* pod redakcją Kazimierza Polańskiego (1980–1992).

TYP może obecnie przyjmować jedną z następujących 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 podstawowe przyjęte tutaj kryterium podziału stanowi istnienie stosunkowo spójnego zestawu realizacji składniowych każdej z klas. Poniżej podajemy kilka przykładów schematów 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) }

CIAGNAĆ (*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) }

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

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

OBCHODZIĆ SIĘ (*ktos 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.

6. Konstrukcje frazeologiczne

Wśród typów argumentów przewidziano także frazeologizmy rzeczownikowe i przyimkowe, przy czym schematy zleksykalizowane były uwzględnianie 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 leksykalizację *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 jeden ze schematów czasownika ODBIJAĆ SIĘ wygląda następująco:

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

Schemat ten 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 wyspecy-

⁹ Istnienie elementu kontrolującego i kontrolowanego w poniższym schemacie 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 pięciolatka/*pięciolatek* itp.

fikować 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 schemat, w którym argument frazeologiczny opisany jest jako niemodyfikowalny (*natr*):

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

Przykład ten ilustruje jeszcze jedną cechę schematów frazeologicznych: elementów leksykalnych nie można pominać. Powyższego schematu 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 jeden ze schematów czasownika OSZALEĆ, odpowiadający wyrażeniu *Ktoś oszalał na czymś punkcie/na punkcie czegoś*, ma następującą postać:

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*). Odpowiedni schemat czasownika POKAZAĆ SIĘ podany jest poniżej, a więc obecna wersja słownika nie odróżnia wymagania modyfikacji dzierżawczej w argumentie frazeologicznym od wymagania dowolnej modyfikacji w takim argumentie:

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 np. *przyjąć kogoś pod swój/własny dach*, ale raczej nie *pod dach Janka*, można *coś słyszeć na własne uszy*, ale nie *na uszy czyjeś* itp. Odpowiedni schemat walencyjny czasownika SŁYSZEĆ realizowany w takiej konstrukcji to:

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

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

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

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 rozchwyany rodzajowo.

7. Inne modelowane zjawiska

W powyższych punktach omówiono najważniejsze, naszym zdaniem, nowe elementy formalizmu opisu schematów 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 schematów 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 UWAZAĆ 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 osobnym schemacie 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 realizację 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 poniższym schemacie dla czasownika POMYSLEĆ:

`subj { np (str) } + { prepnp (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. *prepnp (z, gen)* czy *prepnp (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 `comprepnp(z powodu)` w następującym schemacie czasownika ROZPACZAĆ:

```
subj{np(str)} + {comprepnp(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 frazeologicznym schemacie 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 bezzennymi...*; przykład z NKJP). Argument uzgadniający się z taką frazą przymiotnikową oznaczony jest za pomocą opisanego wcześniej schematu związanego ze zjawiskiem kontroli, jako `controller`. Odpowiedni schemat dla czasownika WYDAWAĆ SIĘ będzie więc miał następującą postać:

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

Schemat ten 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 neodwołującą się bezpośrednio do jednego z siedmiu przypadków morfologicznych w języku polskim jest `postp`, czyli „przypadek poprzyimkowy” form przymiotnikowych we frazach przyimkowo-przymiotnikowych (`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 następującym schemacie 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*). Na początku września 2013 roku słownik liczył 6965 takich podhasel (w tym 3924 sprawdzone przez „superleksykografów”) pogrupowanych w 4811 hasel (w tym 2683 zweryfikowane¹¹). Schematów walencyjnych jest 24199 (14874 sprawdzone), a więc około 3,5 schematu na jedno podhasło (3,8 w części zweryfikowanej). Dalsze dane liczbowe podane poniżej dotyczą wszystkich hasel oznaczonych jako gotowe, a nie tylko tych ostatecznie zweryfikowanych.

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

- pewne: 21931,
- potoczne: 1111,
- archaiczne: 220,
- wątpliwe: 865.

Pozycji składniowych w całym słowniku jest 68874, a więc około 2,8 pozycji na schemat walencyjny. Należy jednak pamiętać, że jedna pozycja może zawierać — oddzielone średnikiem — różne specyfikacje morfoskładniowe. Liczba schematów zawierających pozycje z kilkoma specyfikacjami morfoskładniowymi to 3148, a więc około 13 procent. Ta liczba pokazuje, że możliwość różnych morfoskładniowo realizacji danej pozycji składniowej nie jest zjawiskiem marginalnym¹³.

¹¹ Por. wspomniane we wprowadzeniu 1440 hasel w słowniku wyjściowym.

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

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

9. Zakończenie

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

Choć obecnie słownik ma charakter powierzchniowski, 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*, nieróżniącym 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 2013 roku. W ramach tego projektu słownik ma kilkukrotnie 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 schematów według znaczeń, a z drugiej — wiązania ze sobą schematów stanowiących różne realizacje powierzchniowe tej samej struktury głębokiej. W projekcie przewidziane jest także przypisanie argumentom preferencji selekcyjnych (czyli np. informacji o tym, że podmiotem czasownika *RZEĆ* jest zwykle fraza oznaczająca konia). Badania związane z frazeologicznymi schematami walencyjnymi planowane są także w sieci naukowej PARSEME (<http://zil.ipipan.waw.pl/PARSEME>), zaś aspekty informatyczne probabilistycznego przetwarzania tekstów z wykorzystaniem schematów 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 Europe-An Resources”, 1.02.2011–31.01.2013, <http://cesar.nytud.hu/>), NEKST (projekt typu POIG 1.1.2, „Adaptacyjny system wspomagający rozwiązywanie problemów w oparciu o analizę treści dostępnych źródeł elektronicznych”, 1.04.2009–25.02.2014, <http://www.ipipan.waw.pl/nekst/>) i CLARIN-PL („Polska część infrastruktury naukowej CLARIN ERIC: Wspólne zasoby językowe i infrastruktura technologiczna”, 1.04.2013–31.12.2015, 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. Funkcje „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 Zgondek. Wszystkim powyższym i innym osobom zaangażowanym w rozwój słownika serdecznie dziękujemy.

Bibliografia

- Bańko M., red., 2000, *Inny słownik języka polskiego*, Warszawa.
- Bondaruk A., 2004, *PRO and Control in English, Irish and Polish — A Minimalist Analysis*, 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. Wierchoń, K. Trojanowski, Warszawa, s. 193–206.
- Kallas K., 1993, *Składnia współczesnych polskich konstrukcji współrzędnych*, Toruń.
- Landau I., 2013, *Control in Generative Grammar: A Research Companion*, 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 (COLING2012)*, 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.
- 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.
- 1999, Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach, rozprawa doktorska, Universität Tübingen, Tybinga.
- 2000, Long distance genitive of negation in Polish, *Journal of Slavic Linguistics* 8, s. 151–189.

- 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*, Warszawa.
- Przepiórkowski A., Bańko M., Górski R.L., Lewandowska-Tomaszczyk B., red., 2012, *Narodowy Korpus Języka Polskiego*, 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.
- 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*, Warszawa.
- 1992b, Realizacje zdaniowe podmiotu-mianownika, czyli o strukturalnych ograniczeniach selekcyjnych, [w:] *Opisać słowa*, red. A. Markowski, Warszawa, s. 188–201.
- 1993, Dalsze kłopoty z bezokolicznikiem, [w:] *Językoznawstwo synchroniczne i diachroniczne*, red. J. Sambor, J. Linde-Usiekiewicz, R. Huszcza, Warszawa, s. 303–314.
- 1994, *Syntactic dictionary of Polish verbs*. Maszynopis, Uniwersytet Warszawski i Universiteit van Amsterdam.
- 1998, *Syntactic dictionary of Polish verbs*. Version 3a. Maszynopis, Uniwersytet Warszawski.
- 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*, 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, Poznań.
- Witkoś J., Cegłowski P., Snarska A., Żychliński S., 2011b, Wybrane aspekty de-rywacji i interpretacji wyrażen bezokolicznikowych. Minimalistyczne studium porównawcze polsko-angielskie, Poznań.
- Woliński M., 2004, *Komputerowa weryfikacja gramatyki Świdzińskiego*, rozprawa doktorska, Warszawa.
- Zabrocki T., 1981, *Lexical Rules of Semantic Interpretation: Control and NP Movement in English and Polish*, Poznań.

Dodatek: Specyfikacja formatu słownika *Walenty*

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

Każdy wiersz zawiera trzy pola oddzielone dwukropkiem i spacją, zawierające formę hasłową, informacje o aspekcie i schemat walencyjny, np.¹⁴:

```
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)}
żyć: imperf: subj{np(str)} + obj{np(str)} + {np(inst)}
żyć: imperf: subj{np(str)} + obj{np(str)} + {prepn(wobec, gen)}
żyć się: imperf: subj{np(str)} + {np(inst)}
```

W wypadku czasowników posiadających więcej niż jeden schemat walencyjny każdy taki schemat jest wypisywany w osobnym wierszu, z powtórzeniem formy hasłowej i specyfikacji aspektu, jak dla *żyć* powyżej.

Wartością pola aspektowego może być imperf (niedokonany), perf (dokonany) lub symbol _ (czasownik dwuaspektowy z tym samym schematem w obu aspektach; zob. *DAROWAĆ* powyżej). Trzecie pole zawiera specyfikację I schematu 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 schemacie), 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 (controllee2; najwyżej jeden), por. pkt § 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);
- prepn (PRZYIMEK, PRZYPADEK) — fraza przyimkowa z konkretnym przyimkiem i wymaganym przez niego przypadkiem (dodatkowa możliwa wartość to loc, ale nie part);

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

- *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;
- *comprepn* (PRZYIMEK_ZŁOŻONY) — fraza przyimkowa z danym przyimkiem złożonym (por. pkt § 7.3);
- *lexnp* (PRZYPADEK, LICZBA, 'LEMAT', MODYFIKACJA) — frazeologizm rzeczownikowy (por. pkt § 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 (por. pkt § 6);
- *fixed* ('FRAZEOLOGIZM') — frazeologizm o ustalonym brzmieniu (por. pkt § 6);
- *cp* (TYP) — fraza zdaniowa konkretnego typu: *że*, *żeby*, *żeby2* (dla fraz zdaniowych ze spójnikiem *ZE*, 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*;
- *prepn* (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 oznaczenia 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 (por. pkt § 5): *locat*, *abl*, *adl*, *perl*, *temp*, *dur*, *mod*;
- *advp* (*misc*) — fraza przysłówkowa (niedopuszczająca realizacji przyimkowo-nominalnych itp.);
- *advp* (*pron*) — zaimek przysłówkowy (por. pkt § 7.2);
- *nonch* — fraza niechromatyczna (por. pkt § 7.2);
- *or* — mowa niezależna;
- *refl* — partykuła *się* realizująca zwrotne użycie czasownika (por. pkt § 7.1);
- *E* — podmiot *podniesiony*, tj. w pełni współdzielony z podmiotem (lub jego brakiem) argumentu bezokolicznikowego (por. pkt 4); tylko w kontekście *subj*, *controller{E}* (ale por. przypis 8).

SUMMARY

Modelling syntactic properties of verbs in a new valence dictionary of Polish

The paper describes a formalism used to encode syntactic properties of Polish verbs in a new valence dictionary, currently developed at the Institute of Computer Science, Polish Academy of Sciences. The coordination test — supported by attested examples from the National Corpus of Polish — is used to identify diverse morphosyntactic realisations, e.g., nominal and sentential, of the same syntactic argument of a verb. The following phenomena are among those usually absent in previous dictionaries of Polish: structural case assignment, non-nominal subjects, passivisation, syntactic control, various types of adverbial arguments, phraseology.