XCVII 1

INDEKS NR 362050 | NAKŁAD 550 EGZ. | ISSN 0021-6941

JĘZYK POLSKI

PISMO TOWARZYSTWA MIŁOŚNIKÓW JĘZYKA POLSKIEGO

MARZEC 2017 | CENA ZŁ 20 (5% VAT)

MARZEC 2017 | XCVII 1

JĘZYK POLSKI

PISMO TOWARZYSTWA MIŁOŚNIKÓW JĘZYKA POLSKIEGO

Adiustacja, korekta i łamanie: Wydawnictwo JAK, www.wydawnictwojak.pl
Projekt szaty graficznej: Joanna Tyborowska, joannatyb@gmail.com
Witryna internetowa czasopisma: http://jezyk-polski.pl

Czasopismo wydawane przy współudziale Instytutu Języka Polskiego Polskiej Akademii Nauk

Dofinansowano ze środków Ministra Kultury i Dziedzictwa Narodowego

Adres redakcji: al. Mickiewicza 31, 31–120 Kraków
e-mail: jezykpolski@ijp-pan.krakow.pl
Wersją pierwotną (referencyjną) czasopisma jest wersja papierowa.

ISSN 0021–6941

WARUNKI PRENUMERATY

Prenumerata realizowana przez RUCH S.A.: Zamówienia na prenumeratę w wersji papierowej można składać bezpośrednio na stronie www.prenumerata.ruch.com.pl. Ewentualne pytania prosimy kierować na adres e-mailowy: prenumerata@ruch.com.pl lub kontaktując się z Telefonicznym Biurem Obsługi Klienta pod numerem: 801 800 803 lub 22 717 59 59 – czynne w godzinach 7–18. Koszt połączenia według taryfy operatora.

Odbiorcy krajowi i zagraniczni mogą również zamawiać prenumeratę „Języka Polskiego" w firmach kolporterskich:

a) Kolporter – informacje: http://kolporter.com.pl lub pod nr. tel. 801 205 555,

b) Garmond Press – informacje: http://www.garmondpress.pl.

Koszt prenumeraty dla członków Towarzystwa Miłośników Języka Polskiego na rok 2017 wynosi wraz z roczną składką członkowską 85,00 zł. Aby zostać członkiem Towarzystwa (lub przedłużyć swoją przynależność członkowską na 2017 rok), należy przekazać wymienioną kwotę na konto Towarzystwa w PKO BP SA I O Kraków 51 1020 2892 0000 5402 0150 5460 w terminie do 31 stycznia 2017 roku i wypełnić deklarację członkowską.

Pojedyncze zeszyty czasopisma, bieżące i archiwalne, można nabywać w Biurze Zarządu Głównego (al. Mickiewicza 31, 31–120 Kraków, e-mail: tmjp@ijp-pan.krakow.pl).

ARTYKUŁY I ROZPRAWY

Lingwistyka komputerowa – czy po prostu lingwistyka?

Gdyśmy publikowali (w z. 1–2/2015 „Języka Polskiego") zbiór artykułów podsumowujących naukową refleksję nad polszczyzną w latach 1989–2014, byliśmy świadomi, że nie obejmuje on pełnego spektrum zainteresowań naszych językoznawców okresu najnowszego. Jedną z najpoważniejszych luk stała się nieobecność materiału dotyczącego lingwistyki komputerowej – dyscypliny, która, choć dostrzegalna już w okresie wcześniejszym, rozwinęła się w pełni dopiero po roku 1989, a przyczyna tego jest całkiem prosta. Dopiero wtedy technika komputerowa wydoskonaliła się i upowszechniła (zwłaszcza w Polsce) w stopniu odpowiednim, aby badania z tej dziedziny nabrały rozmachu i przestały być tylko domeną wąskiej elity mającej dostęp do niebywale kosztownego sprzętu. Brak ów uzupełniamy w niniejszym zeszycie „Języka Polskiego", który otwiera przeglądowy artykuł Magdaleny Derwojedowej, poświęcony lingwistyce komputerowej lat minionych, a po nim zamieszczamy kilka prac opisujących badania z tego zakresu obecnie prowadzone lub niedawno ukończone. Oczywiście, nie pokazują one całości dokonań polskiej lingwistyki komputerowej i inżynierii językowej. Prace takie są dziś prowadzone w tylu ośrodkach i przez tylu uczonych, że dopiero w przyszłości może się uda je syntetycznie ująć i opisać.

Gdy myślę dziś o historii polskiej lingwistyki komputerowej, przypomina mi się jedno znamienne zdarzenie. W 2005 roku odbył się w Warszawie zjazd Polskiego Towarzystwa Językoznawczego na temat *Media elektroniczne w języku i lingwistyce*. Zamiarem jego głównego organizatora, prof. Janusza Bienia, notabene jednego z pionierów lingwistyki komputerowej w Polsce, było m.in. upowszechnienie wśród większej grupy naszych językoznawców wiedzy o możliwościach i osiągnięciach komputerowej analizy języka. Zjazd ten okazał się wszak chyba jednym z najbardziej kameralnych w historii Towarzystwa; spotkali się na nim właściwie tylko przedstawiciele środowisk lingwistów komputerowych i ich nieliczni sympatycy. Szersze kręgi badaczy najwyraźniej uznały wydarzenie za nieinteresujące dla siebie. Zarejestrowanym gościom zjazdu rozdawano płytę CD ze zdigitalizowaną wersją słownika Knapiusza w formacie djvu (który wkrótce potem stał się standardem dla rozwijających się bibliotek cyfrowych). Traf chciał, że bezpośrednio z tego zjazdu udałem się na inną konferencję, poświęconą w głównej mierze historii języka polskiego, i tam pochwaliłem się ową płytą. Zaskoczenie niektórych z obecnych diachronistów było trudne do opisania. Nie wiedzieli bowiem ani o fakcie, że taka wersja najważniejszego siedemnastowiecznego słownika z językiem polskim istnieje, ani że można ją było otrzymać gratis na owym zjeździe, ani tym bardziej, że już kilka lat temu została opublikowana na CD w innej formie i była rozprowadzana płatnie. Doskonale to odzwierciedlało ówczesną sytuację, w której lingwiści i informatycy zajmujący się komputerową analizą języka oraz pozostali językoznawcy stanowili w zasadzie dwa całkiem odrębne środowiska, słabo się komunikujące ze sobą.

Minęło od tamtego czasu dwanaście lat. Upowszechniły się biblioteki cyfrowe, elektroniczne wersje słowników, powstał Narodowy Korpus Języka Polskiego, którego roli w rozwoju

warsztatu badań nad polszczyzną nie sposób przecenić i z którego korzystają niemal wszyscy językoznawcy, a wciąż trudno się oprzeć wrażeniu, że relacje między przedstawicielami lingwistyki komputerowej czy inżynierii językowej a pozostałą częścią środowiska dalekie są od ideału. Dominujące wydają się dwie postawy. Niektórzy językoznawcy, nazwijmy ich, tradycyjni, skłonni są uważać językoznawców, nazwijmy ich, informatycznych, za kogoś w rodzaju „panów od komputera", których zwykli wzywać na pomoc w wypadku awarii sprzętu, którym ewentualnie można zlecić pewne czynności pomocnicze w procesie badania naukowego. Najwyraźniej nie mają świadomości, że w środowisku tym powstają również prace o charakterze teoretycznym, a nawet nowe modele opisu polszczyzny, adaptujące najnowsze osiągnięcia lingwistyki światowej. Inni zaś przeciwnie, uważają ich za posiadaczy „wiedzy tajemnej", czarnoksiężników, którzy za pomocą jakichś zaklęć wydobywają z czeluści Internetu informacje, w rzeczywistości możliwe do uzyskania przez każdego, kto wie, gdzie szukać i jaką komendę do programu wpisać. Z kolei przedstawiciele lingwistyki komputerowej wydają się czasem zapominać, że znaczna część ich pracy opiera się (ciągle) na wiedzy wypracowanej przez lingwistów, i to nieraz w czasach, kiedy o komputerze mało komu się śniło. O ileż mniej moglibyśmy znaleźć w NKJP, gdyby w zamierzchłej dla wielu z dziś żyjących lingwistów przeszłości Jan Tokarski nie zaczął prac nad systematyzacją polskiej fleksji, gdyby tych prac potem nie kontynuował Zygmunt Saloni, sam i z zespołem, przy współpracy wielu osób, co w efekcie doprowadziło do uzyskania wiedzy morfologicznej i syntaktycznej zawartej w tzw. tagsecie korpusu. Czym byłaby *Słowosieć*, i czy byłaby w ogóle, gdyby nie tezaurus Rogeta, wydany w połowie XIX wieku, a przede wszystkim *Dobór wyrazów* Romana Zawilińskiego, jeden z najbardziej niedocenionych (zresztą przez samych językoznawców i leksykografów) polskich słowników ubiegłego stulecia? W czasach mody na „łączenie otwartych danych", na „crowdsourcing" i inne podobne działania informatycy językowi wydają się czasami nie rozumieć, jaka jest różnica między słownikiem elektronicznym skompilowanym z innych a tym opracowanym merytorycznie od początku przez językoznawców czy leksykografów, opartym na analizie aktualnego materiału językowego, że nie każdy musi się zgodzić na włączenie wypracowanej przez siebie wiedzy jako składnika, cegiełki do produktu, który firmować będzie kto inny, a właściwy autor zostanie tylko wspomniany w atrybucji bibliograficznej. Szkodzi to wszystko wzajemnemu zaufaniu.

Ponieważ jednak wszystko wskazuje na to, że jakiekolwiek badanie języka w XXI wieku bez odwoływania się do metod czy technik związanych z lingwistyką komputerową, z drugiej strony jakikolwiek rozwój lingwistyki komputerowej i inżynierii językowej bez korzystania z wiedzy uzyskanej w drodze analiz czysto językoznawczych, nie będą, poza nielicznymi wyjątkami, możliwe, a w każdym razie nie będą wystarczająco skuteczne i płodne poznawczo, oba środowiska – lingwistów komputerowych i pozostałych – powinny się nastawić na współpracę i takie ułożenie wzajemnych stosunków, które by u żadnej ze stron nie wywoływało uczucia dyskomfortu. W dalszej przyszłości zaś środowiska te pewnie powinny się zintegrować w jedno. Środowisko badaczy języka – czyli językoznawców.

Piotr Żmigrodzki

ARTYKUŁY I ROZPRAWY

MAGDALENA DERWOJEDOWA* | UNIWERSYTET WARSZAWSKI

Językoznawstwo komputerowe i inżynieria lingwistyczna po 1989 roku

Słowa kluczowe: językoznawstwo komputerowe, inżynieria lingwistyczna, przetwarzanie języka naturalnego.

Lingwistyka komputerowa (informatyczna) to dziedzina językoznawstwa, która wykorzystuje metody komputerowe (statystyczne lub algorytmiczne) do badania i modelowania języka naturalnego, natomiast inżynieria lingwistyczna jest nakierowana na tworzenie aplikacji (programów), które przetwarzają dane językowe (Mykowiecka 2007: 9). Rozgraniczenia te są w dużym stopniu umowne i opierają się na przeciwstawieniu podstaw teoretycznych zastosowaniom praktycznym; często wymiennie używa się też terminu ogólniejszego – przetwarzanie języka naturalnego: (Natural Language Processing (NLP), por. Przepiórkowski 2008: 3–6; Hajnicz 2011: 1).

Dystans dzielący polską technologię językową od rozwiązań, którymi mogą się cieszyć przetwarzający teksty w bardziej zaawansowanych w tej dziedzinie językach, ciągle istnieje, jednak wiele zaległości i braków zostało nadrobionych: występują zasoby korpusowe i leksykalne, a także zestawy narzędzi lingwistycznych. W wielu wypadkach nie była to prosta adaptacja, dostosowanie do fleksyjnej specyfiki polszczyzny wymagało bowiem niemałej inwencji, niektóre przedsięwzięcia to dzieła całkowicie oryginalne. Zasobami, które udało się uzupełnić, są przede wszystkim korpusy polszczyzny (Hebal-Jezierska 2014). Zasobami tego rodzaju są też analizatory morfologiczne, słownik walencyjny czy wordnet, które wraz z innymi zasobami niekorpusowymi omówię w pierwszej części. Druga jest poświęcona narzędziom do przetwarzania języka naturalnego.

Słowosieć (http://plwordnet.pwr.wroc.pl/wordnet/) jest polskim wordnetem, czyli siecią wyrazów poklasyfikowanych ze względu na łączące je relacje znaczeniowe (Derwojedowa i in. 2007b; Maziarz i in. 2012). Relacja synonimii grupuje jednostki w tak zwane *synsety* (*sets of synonyms*, zbiory synonimów), natomiast hierarchię znaczeń tworzy hiperonimia. Siatka powiązań jednostek Słowosieci jest jednak bardziej złożona, uwzględnia inne relacje leksykalne oraz powiązania słowotwórcze; dodatkowo podano ekwiwalenty angielskie z wordnetu princetońskiego (Fellbaum 1998). Słowosieć jest rozległym zasobem leksykalnym, pod względem liczby haseł (170 tys.) największym obecnie na świecie. Ze względu na bogate polskie słowotwórstwo do konstruowania sieci zależności leksykalnych posłużono się wzorcami

* derwojed@uw.edu.pl

czeskimi z wielojęzycznego programu EuroWordNet (por. Derwojedowa, Zawisławska 2007). Niewątpliwym osiągnięciem polskiego wordnetu jest zastosowanie wspomagania maszynowego i eksperymenty w zakresie automatycznego wykrywania relacji znaczeniowych (Piasecki i in. 2009; Maziarz i in. 2012). Dzięki temu możliwe było szybkie powiększenie słownika, prace te stały się też zaczynem ciekawych eksperymentów w zakresie łączliwości wyrazów – na stronie projektu znajduje się także wyszukiwarka funkcji podobieństwa znaczeniowego podająca, jakie wyrazy występują najczęściej w tych samych kontekstach (Broda i in. 2008; Broda, Piasecki 2011). Słowosieć jest zasobem nieustannie rozwijanym przez zespół kierowany przez Macieja Piaseckiego, co pozwala żywić nadzieję, że nie tylko będzie rozległa, ale zostanie też rygorystycznie zweryfikowana. Drugim polskim wordnetem jest PolNet, wykorzystywany na przykład w systemie wspomagania w sytuacjach kryzysowych (Vetulani i in. 2010a,b; Bouvry i in. 2012).

Zasobami leksykalnymi i narzędziami do ich opracowania są programy do tworzenia słowników nazw wielosegmentowych, odmieniania takich jednostek oraz ich wyszukiwania w tekście. Są to na przykład słownik nazw własnych, słownik nazw geograficznych i słownik gramatyczny frazeologizmów oraz narzędzia do wykrywania jednostek nazewniczych (*named entities*; Graliński i in. 2010; Marciniak, Rabiega-Wiśniewska 2010; Czerepowicka, Kosek 2011; Marciniak i in. 2011; Savary, Waszczuk 2012; Marcińczuk i in. 2013). Inną grupę stanowią słowniki walencyjne – niewielki (1 tys. jednostek), ale bardzo szczegółowy Marka Świdzińskiego (1994), zaadaptowany na potrzeby przetwarzania maszynowego *Słownik syntaktyczno-generatywny* Kazimierza Polańskiego (Polański (red.) 1980–1992; Grund 2000), słownik CEGLEX wykorzystany następnie w narzędziach GRAMLEX (Vetulani i in. 1998) oraz najnowszy z nich, a zarazem najbardziej rozbudowany, rozwijany od kilku lat Walenty (Przepiórkowski i in. 2014a). Słownik ten powstał na podstawie NKJP (Przepiórkowski i in. (red.) 2012) – zawarte w nim informacje to te wymagania, które można wydobyć, analizując dane korpusowe. Jest on ponadto zapisany w formacie umożliwiającym współpracę z innymi programami, np. wykorzystanie w parserach. Choć z rozstrzygnięciami autorów można by dyskutować (w szczególności co do zestawu notowanych fraz czy interpretacji pewnych zjawisk), to jest to największe i najszczegółowiej opracowane polskie źródło walencyjne, zawierające informacje powierzchniowo- i głębinowoskładniowe, do tego dostępne dla ogółu zainteresowanych (Przepiórkowski 2004b; Przepiórkowski i in. 2014a,b; Hajnicz i in. 2015). W wypadku Walentego wykorzystano metody automatyczne do identyfikacji fraz zależnych; szczegółowo procedurę identyfikacji składników zdaniowych potencjalnie pełniących funkcje poszczególnych argumentów opisuje Elżbieta Hajnicz, rozważając też możliwe alternacje, powiązania struktury powierzchniowo- i głębinowoskładniowej (mapowanie argumentów i fraz). Konstruuje ona odpowiednie testy oraz omawia przeprowadzone eksperymenty, które poprzedziły powstanie Walentego (Hajnicz 2011). Słownik jest rozbudowywany, uzupełniany i ulepszany.

Inna pod względem założeń i celów, choć również zawierająca informację o wymaganiach składniowych, jest nieduża, bo licząca około 200 haseł, baza RAMKI. Wykorzystuje ona teorię ram interpretacyjnych Ch.J. Fillmore'a (1982). Ten szczupły zasób charakteryzuje się olbrzymią homonimią jednostek, przyjęto bowiem, że o wydzieleniu osobnego hasła decyduje różnica w którymkolwiek z kryteriów: znaczenia, własności składniowych (realizowanych schematów),

własności ramowych (przywoływanych ram) oraz odpowiedniości ról głębokich i funkcji powierzchniowych. Każdej jednostce przyporządkowano ramę, zestaw ról, scenariusz, schemat składniowy oraz zilustrowano ją przykładami z korpusu, na których zostały oznaczone podstawowe funkcje powierzchniowe oraz elementy ramy (Derwojedowa i in. 2010).

Wśród programów do przetwarzania języka naturalnego można wyróżnić parsery składniowe, analizatory morfologiczne, dyzambiguatory i tagery oraz narzędzia bardziej wyspecjalizowane, np. do wydobywania z tekstu jednostek określonego typu. Szczególną pozycję w tej grupie zajmuje monografia M. Świdzińskiego zawierająca wyczerpujący formalny powierzchniowy opis składniowy polszczyzny (Świdziński 1992). Są w niej zdefiniowane w języku Prolog (Colmerauer 1978) reguły analizy struktur składniowych polszczyzny, od wypowiedzenia do konstrukcji najniższego rzędu. Autor sceptycznie odnosił się do możliwości jej implementacji. Tymczasem już po kilku latach podjęto takie próby (Bień 1996; Bień i in. 2000), a w 2004 roku powstał program Świgra (od: Świdzińskiego gramatyka) Marcina Wolińskiego. W latach 2009–2013 gramatyka była przez obu badaczy rozwijana, obecnie M. Woliński prowadzi prace nad znaczącym ulepszeniem parsera. Bezpośrednim wynikiem tych badań jest bank drzew rozbiorów gramatycznych Składnica (Świdziński, Woliński 2009, 2010; Woliński i in. 2011). Gramatyka doczekała się też istotnego rozszerzenia przez Macieja Ogrodniczuka (2006) o składnię konstrukcji liczebnikowych.

Zarówno kompletnością, jak i subtelnością opisu różnią się od gramatyki Świdzińskiego inne opisy formalne, jak choćby gramatyka Tomasza Obrębskiego (2002). Jest ona pierwszym formalnym ujęciem zależnościowym, podstawą działającego programu. I chociaż jej zakres jest ograniczony, to gramatyka ta zawiera kilka ciekawych rozwiązań, np. w zakresie opisu interpunkcji czy uzgodnień. Drzewa zależnościowe zawiera też wspomniana już Składnica. Powstały one w wyniku konwersji pierwotnych drzew składnikowych według reguł zaproponowanych przez Alinę Wróblewską (2012). Innym oryginalnym na gruncie polskim rozwiązaniem jest gramatyka opracowana w formalizmie HPSG, a następnie zaimplementowana przez zespół z IPI PAN. Opis ten nie jest kompletny, ale prezentuje główne cechy formalizmu i (niekiedy dyskusyjnych) rozstrzygnięć gramatyki (Przepiórkowski i in. 2001). Kompletna i wyczerpująca ma być rozwijana obecnie gramatyka wykorzystująca teorię LFG (Lexical Functional Grammar; Patejuk, Przepiórkowski 2015). Jej istotnym *novum* jest uwzględnienie semantyki, dotąd w analizie składniowej w zasadzie pomijanej. Analiza semantyczna jest też celem prac nad gramatyką kategorialną języka polskiego (Marczewski 2012; Jaworski 2013). W gramatyce została podjęta próba rozwiązania problemów, jakich nastręcza formalizm w odniesieniu do języka polskiego, a więc swobodnej linearyzacji, odróżnienia składników luźnych i wymaganych oraz bogatej morfologii.

Warunkiem wstępnym analizy składniowej jest analiza morfologiczna. Jest ona również niezbędna do bardziej zaawansowanego wyszukiwania w korpusach. Z wielu analizatorów powstałych na przestrzeni ostatnich 25 lat tylko nieliczne przetrwały próbę czasu i rozwoju, przede wszystkim aktualizacji słownika. Sytuacja analizy fleksyjnej jest przy tym wyjątkowa – została ona bowiem formalnie opisana i usystematyzowana wcześniej (Tokarski 1951, 1961a,b, 1962, 1963a,b,c,d, 1973, 1993; Gruszczyński 1989; Saloni 1992, 2000, 2001; Saloni i in. 2007/2012). Ze *Schematycznego indeksu* a tergo *polskich form wyrazowych* Jana Tokarskiego (1993) korzystały programy SAM-96

i udoskonalony SAM-98 Krzysztofa Szafrana (1997) oraz Morfeusz M. Wolińskiego (2006, 2014). Analizą *a fronte* posłużyli się twórcy analizatora Amor (Rabiega-Wiśniewska, Rudolf 2002; Rudolf 2004). Uniwersalny mechanizm opracowany przez Gábora Prószékyego z oryginalnym polskim słownikiem jest wykorzystywany przez analizator PoMor Roberta Wołosza (2005). Program ten wyróżnia się tym, że korzysta z dwóch podsłowników, z których jeden umożliwia włączenie do analizy form dawnych, regionalnych i przestarzałych. Na potrzeby tłumaczenia maszynowego powstał moduł analizy morfologicznej LEX (Graliński i in. 2013), a z programu do korekty pisowni – analizator Morfologik (Woliński i in. 2012).

Narzędzia te były i są w różnym stopniu używane w przetwarzaniu języka naturalnego, np. SAM znalazł zastosowanie w systemach tłumaczenia maszynowego (Fabian i in. 1999; Graliński 2000, 2001), natomiast Morfeusz jest podstawowym narzędziem analizy w zestawie narzędzi korpusowych i parserów IPI PAN, w tym kolejnych programów do znakowania korpusów – tagerów (Przepiórkowski 2004a, 2008; Przepiórkowski i in. (red.) 2012). Morfeusz umożliwia też użytkownikom samodzielną, dość wygodną analizę poszczególnych słów i całych tekstów, jest także wyposażony w moduł syntezy. Od pewnego czasu trwają prace nad rozszerzeniem słownika programu o możliwość analizowania tekstów dawniejszych (z lat 1600–1750 i 1830–1918; por. Derwojedowa i in. 2014; Gruszczyński i in. 2014).

Blisko tych programów lokują się programy do odmiany wyrażeń wielosegmentowych. Praktycznym potrzebom szerszego odbiorcy (np. redaktora, ucznia czy studenta) odpowiadają automaty Zbigniewa Bronka do odmiany grup liczebnikowych, nazwisk i nazw geograficznych (http://nlp.actaforte.pl/), tworzące wszystkie poprawne formy odmiany. Dla każdej klasy wyrażeń autor skonstruował niewielką gramatykę, wypełniając programami pewną lukę – analizatory zwykle koncentrują się na rzeczownikach pospolitych i wyrażeniach jednosegmentowych. W wypadku automatu składającego i rozkładającego grupy liczebnikowe powstało narzędzie o dużej szczegółowości – prezentowane są też warianty skrajnie rzadkie, zresztą jako takie oznaczone.

Programom do analizy fleksji nie towarzyszą w równej obfitości opisy słowotwórcze. Istnieje co prawda teoretyczny opis Joanny Rabiegi-Wiśniewskiej (2008), jednak powiązania słowotwórcze znajdują miejsce co najwyżej w zasobach leksykalnych, np. w Słowosieci i SGJP3 (Derwojedowa i in. 2007a,b; Broda i in. 2008; Piasecki i in. 2008; Maziarz i in. 2012; Saloni i in. 2015) oraz – w niewielkim zakresie – w Morfeuszu i PoMorze (Wołosz 2005; Woliński 2014). Można więc mieć nadzieję, że i w tej mierze nastąpi znaczący postęp.

Dane otrzymane z analizy morfologicznej mogą być wieloznaczne ze względu na homonimię. Zadaniem dyzambiguatora jest wybranie interpretacji poprawnej w danym kontekście. Problem ten jest rozwiązywany przez adaptację do specyfiki polszczyzny rozwiązań dla innych języków. Stosowane metody można podzielić na metody bazujące na regułach opisujących ograniczenia danego języka i na metody statystyczne. Oba te sposoby bywają też łączone. Ograniczenia są wykrywane automatycznie w tekście lub formułowane przez eksperta na podstawie jego wiedzy językowej. Przykładem reguły może być sąsiedztwo prawostronne wyrazu *przeszło* – jako operator adnumeratywny poprzedza ono liczebnik (Grochowski 1984), jako czasownik zajmuje pozycje ze znacznie mniejszymi ograniczeniami.

Metodami statystycznymi posłużył się Łukasz Dębowski w pionierskim na gruncie polskim trigramowym dyzambiguatorze opracowanym na potrzeby Korpusu IPI (Dębowski 2003). Do oznakowania finalnej wersji tego korpusu użyto natomiast regułowego tagera TaKIPI (Piasecki, Godlewski 2006), który wykorzystuje drzewa decyzyjne oraz reguły ujednoznaczniania. W NKJP zastosowano jeszcze inne podejście, korpus został oznakowany za pomocą programu Pantera, czyli adaptowanego przez Szymona Acedańskiego tagera Brilla. W module PELCRA NKJP zastosowano ujednoznacznianie wykorzystujące maksimum entropii, program został opracowany przez Piotra Pęzika (Acedański 2010; Przepiórkowski i in. (red.) 2012). Powstały też programy wykorzystujące inne techniki: warstwowe, pamięciowe czy hybrydowe (Radziszewski, Śniatowski 2011; Radziszewski 2012, 2013; Kobyliński 2014). Wyniki działania poszczególnych tagerów wahają się zależnie od tego, jakie czynniki są brane pod uwagę (poprawna kwalifikacja jednostek znanych i nieznanych, błędne przyporządkowanie między klasami wyrazów, błędne przyporządkowanie form synkretycznych itd.), w podstawowym zakresie są jednak dość zbliżone (Mykowiecka 2007; Kobyliński 2014). Mimo takiej mnogości programów samodzielne ujednoznacznienie zebranego korpusu jest zadaniem trudnym technicznie – uruchomienie programów wymaga znacznych umiejętności. Innym problemem są zestawy znaczników (tagsety) nie zawsze odpowiadające przyzwyczajeniom czy potrzebom użytkownika. Program w jakimś stopniu konfigurowalny, przyjaźniejszy użytkownikom, a być może też podający informację o homonimiach niemożliwych do rozwiązania, jest ciągle wyzwaniem.

W ostatnim dwudziestopięcioleciu powstały też interesujące narzędzia o węższej specjalizacji, np. analizator tekstów medycznych (Marciniak, Mykowiecka 2014), systemy tłumaczeniowe (Jassem 2001) czy system wspomagający działania w sytuacjach kryzysowych (Vetulani i in. 2010b). Wśród osiągnięć tego typu warto wymienić program odpowiadający na pytania w języku naturalnym na podstawie polskiej Wikipedii (Przybyła 2015) czy programy do wydobywania informacji z tekstów sumeryjskich (Jaworski 2008) oraz prace w zakresie automatycznej lub wspomaganej komputerowo klasyfikacji tekstów (Borchmann i in. 2015). Znacznie większym przedsięwzięciem jest program Thetos, tłumaczący komunikaty polskie na język migany (system językowo-migowy, SJM). Zawiera on moduł analizy i syntezy morfologicznej, analizuje teksty również semantycznie (Szmal, Suszczańska 2002). Innym wyspecjalizowanym programem przewidzianym do szerszego nawet niż Thetos stosowania jest sprawdzający trudność tekstów Jasnopis. Pod uwagę brane są cechy leksykalne (wyrazy trudne), morfologiczne i składniowe, a stopień trudności jest wyliczany z uwzględnieniem różnych metod statystycznych. Ocena trudności jest przedstawiana użytkownikowi na obrazowej skali jako liczba klas, które trzeba ukończyć, żeby dany tekst zrozumieć. Wskazywane są też te cechy tekstu, które utrudniają zrozumienie (Gruszczyński, Ogrodniczuk (red.) 2015). Do badań szczegółowych, choć bardziej nastawionych na zastosowanie w innych aplikacjach niż jako pomoc użytkownikowi, należy automatyczne wykrywanie koreferencji (Ogrodniczuk i in. 2015). W szerokim rozumieniu jest to nie tylko powierzchniowa anafora, ale również inne typy odniesień – metaforyczne czy wyrażone za pomocą relacji leksykalnych – hipo- i hiperonimii czy meronimii). Wymaga to nie tylko uchwycenia językowo specyficznych wykładników nawiązania, czyli

analizy składniowej, ale też rozstrzygania o identyczności referentów, a także odróżnienia koreferencji od innych zjawisk tekstowych, np. współekstensji, czyli przynależności do tego samego pola znaczeniowego.

Uproszczona analiza konstrukcji składniowych, a więc taka, w której zakłada się, że wynikiem nie muszą być wszystkie możliwe interpretacje lub interpretacje te mogą być niedookreślone, to tzw. płytkie parsowanie. Jest ono stosowane na przykład do wyszukiwania słów kluczowych, wydobywania informacji z tekstów, automatycznego odpowiadania na pytania, a także automatycznego wykrywania kolokacji czy schematów (ram) walencyjnych. Gramatyką napisaną z myślą o płytkim parsowaniu polszczyzny jest Spejd, a wśród przykładów zastosowania jest ujednoznacznianie składniowe, rozpoznawanie wyrażeń wielosegmentowych, nazw własnych, słów pisanych z dywizem, interpretowanie skrótów i skrótowców (w tym pisowni z dywizem i końcówką fleksyjną), liczb zapisanych cyframi jako mających wartości kategorii gramatycznych, dat. Gramatyka definiuje proste konstrukcje składniowe, takie jak rozmaite grupy nominalne czy zdaniowe (Przepiórkowski 2008).

Na koniec trzeba wspomnieć o badaniach ilościowych cech tekstu na użytek filologii, czyli o stylometrii, oraz opracowanym w zespole filologów i informatyków pakiecie *stylo*, który ułatwia takie badania, można go na przykład wykorzystać do analiz wersologicznych czy afiliacyjnych (Pawłowski i in. 2010; Eder 2011; Eder i in. 2013).

We współczesnych badaniach językoznawczych metody i wyniki językoznawstwa komputerowego są wykorzystywane w innych obszarach badań, takich jak wydobywanie danych z korpusów i ich opracowywanie na potrzeby badań różnego typu, a także w leksykografii.

Bibliografia

Acedański S. 2010: *A morphosyntactic Brill tagger for inflectional languages*, [w:] H. Loftsson, E. Rögnvaldsson, S. Helgadótir (red.), *Advances in Natural Language Processing*, 7th International Conference on NLP, IceTAL 2010, Reykjavik, Iceland, August 16–18, Springer, s. 3–14.

Bień J. 1996: *Komputerowa weryfikacja formalnej gramatyki Świdzińskiego*, „Biuletyn Polskiego Towarzystwa Językoznawczego" LII, s. 147–164.

Bień J., Szafran K., Woliński M. 2000: *Experimental parsers of Polish*, [w:] *3. Europäische Konferenz „Formale Beschreibung slavischer Sprachen, Leipzig 1999"*, „Linguistische Arbeitsberichte" 75, Institut für Linguistik, Universität Leipzig, Leipzig, s. 185–190.

Borchmann Ł., Graliński F., Jaworski R., Wierzchoń P. 2015: *A semi-automatic method for thematic classification of documents in a large text corpus*, [w:] F. Mambrini, M. Passarotti, C. Sporleder (red.), *Proceedings of the Workshop on Corpus-Based Research in the Humanities (CRH)*, Instytut Podstaw Informatyki PAN, Warszawa, s. 13–21.

Bouvry P., Kłopotek M.A., Leprevost F.,Marciniak M., Mykowiecka A., Rybiński H. (red.) 2012: *Security and Intelligent Information Systems: International Joint Conference, SIIS 2011, Warsaw, Poland, June 13–14, 2011, Revised Selected Papers*, Lecture Notes in Computer Science, vol. 7053, Springer (online: http://www.springer.com).

Broda B., Derwojedowa M., Piasecki M., Szpakowicz S. 2008: *Corpus-based semantic relatedness for the construction of Polish Word Net*, [w:] *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, European Language Resources Association (ELRA), Marrakech, Morocco, s. 1800–1807 (online: http://www.lrec-conf.org/proceedings/lrec2008/pdf/459_paper.pdf, dostęp: 20 października 2015).

Broda B., Piasecki M. 2011: *Parallel, massive processing in SuperMatrix – a general tool for distributional semantic analysis of corpora*, „International Journal of Data Mining, Modelling and Management" 5 (1), s. 1–19 (online: doi: 10.1504/IJDMMM.2013.051924).

Calzolari N., Choukri K., Declerck T., Loftsson H., Maegaard B., Mariani J., Moreno A., Odijk J., Piperidis S. (red.) 2014: *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, ELRA, Reykjavík, Iceland (online: http://www.lrec-conf.org/proceedings/lrec2014/index.html, dostęp: 30 października 2015).

Colmerauer A. 1978: *Metamorphosis grammars*, [w:] L. Bolc (red.), *Natural Language Communications with Computers*, Lecture Notes in Computer Science, vol. 63, Springer, Berlin, s. 133–189.

Czerepowicka M., Kosek I. 2011: *Problemy opisu związków frazeologicznych w formalizmie „Multifleks" (na przykładzie rodzaju wyrażeń frazeologicznych)*, [w:] M. Bańko, D. Kopcińska (red.), *Różne formy, różne treści*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa, s. 117–126.

Derwojedowa M., Kieraś W., Skowrońska D., Wołosz R. 2014: *Współczesne narzędzia leksykograficzne a analiza tekstów dawniejszych*, „Polonica" XXXIV s. 21–27.

Derwojedowa M., Linde-Usiekniewicz J., Zawisławska M. 2010: *Rygorystyczna aplikacja metodologii kognitywno-interpretacyjnej*, Elipsa, Warszawa.

Derwojedowa M., Piasecki M., Szpakowicz S., Zawisławska M. 2007a: *Polish WordNet on a shoestring*, [w:] G. Rehm, A. Witt, L. Lemnitzer (red.), *Proceedings of Biannual Conference of the Society for Computational Linguistics and Language Technology, Tübingen, April 11–13 2007*, Universität Tübingen, Tübingen, s. 169–178.

Derwojedowa M., Piasecki M., Szpakowicz S., Zawisławska M. 2007b: *Relacje w polskim WordNecie*, Raporty Seria PREPRINTY 1, Instytut Informatyki Stosowanej Politechniki Wrocławskiej, Wrocław.

Derwojedowa M., Zawisławska M. 2007: *Relacje leksykalne w polskiej i czeskiej bazie WordNet (Lexical relations in Polish and Czech WordNet)*, [w:] Z. Rudnik-Karwatowa (red.), Z Polskich Studiów Slawistycznych. *Prace na XVI Międzynarodowy Kongres Slawistów, Ochryda 9–16 września 2008*, *XI Językoznawstwo*, Instytut Slawistyki PAN, Warszawa, s. 15–23.

Dębowski Ł. 2003: *A reconfigurable stochastic tagger for languages with complex tag structure*, [w:] *Proceedings of the Workshop on Morphological Processing of Slavic Languages. 10th Conference of the European Chapter of Association for Computational Linguistics*, Budapest, s. 63–70.

Eder M. 2011: *Style-markers in authorship attribution. A cross-language study of the authorial fingerprint*, „Studies in Polish Linguistics" 6, s. 99–114.

Eder M., Kestemont M., Rybicki J. 2013: *Stylometry with R: A suite of tools*, [w:] *Digital Humanities 2013. Conference Abstracts*, University of Nebraska-Lincoln, Lincoln, s. 487–489 (online: http://dh2013.unl.edu/abstracts/, dostęp: 20 października 2015).

Fabian P., Migas A., Suszczańska N. 1999: *Zastosowania analizy morfologicznej i składniowej w procesie rozpoznawania mowy*, „Speech and Language Technology. Technologia Mowy i Języka", t. 3, s. 155–165.

Fellbaum C. 1998: *WordNet. An Electronic Lexical Database*, MIT Press, Cambridge, Mass.

Fillmore C.J. 1982: *Frame semantics*, [w:] *Linguistics in the Morning Calm*, The Linguistic Society of Korea, Seoul, s. 110–137.

Graliński F. 2000: *Hasłowanie korpusu polskich tekstów informatycznych (1,2 mln słów) – raport*, „Speech and Language Technology. Technologia Mowy i Języka", t. 4, cz. 2, s. 147–153.

Graliński F. 2001: *Komputerowa analiza tekstu polskiego. Zastosowanie w systemie POLENG*, praca magisterska, Wydział Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza, Poznań.

Graliński F., Jassem K., Junczys-Dowmunt M. 2013: *PSI-Toolkit: A natural language processing pipeline*, [w:] A. Przepiórkowski, M. Piasecki, K. Jassem, P. Fuglewicz (red.), *Computational Linguistics*, Studies in Computational Intelligence, vol. 458, Springer, Berlin–Heidelberg, s. 27–39 (online: doi: 10.1007/978-3-642-34399-5_2).

Graliński F., Savary A., Czerepowicka M., Makowiecki F. 2010: *Computational Lexicography of Multi-Word Units: How Efficient Can It Be?*, [w:] E. Laporte, P. Nakov, C. Ramisch, A. Villavicencio (red.), *Proceedings of Multiword Expressions: from Theory to Applications (MWE 2010)*, Workshop at COLING 2010, August 28, Beijing, China.

Grochowski M. 1984: *Projekt klasyfikacji syntaktycznej polskich leksemów nieodmiennych*, „Polonica" X, s. 73–97.

Grund D. 2000: *Komputerowa implementacja słownika syntaktyczno-generatywnego czasowników polskich*, „Studia Informatica" 21 (3), s. 243–256.

Gruszczyński W. 1989: *Fleksja rzeczowników pospolitych we współczesnej polszczyźnie pisanej*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Gruszczyński W., Adamiec D., Ogrodniczuk M. 2014: *Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do 1772 r.) – prezentacja projektu badawczego*, „Polonica" XXXIII, s. 309–316.

Gruszczyński W., Ogrodniczuk M. (red.) 2015: *Jasnopis, czyli mierzenie zrozumiałości polskich tekstów użytkowych*, Wydawnictwo ASPRA-JR, Warszawa.

Hajnicz E. 2011: *Automatyczne tworzenie semantycznych słowników walencyjnych*, Problemy Współczesnej Nauki. Teoria i Zastosowania. Inżynieria Lingwistyczna, Akademicka Oficyna Wydawnicza EXIT, Warszawa.

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

Hebal-Jezierska M. 2014: *Praktyczny przewodnik po korpusach języków słowiańskich*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa (online: http://www.iszip.uw.edu.pl/files/pdf/praktyczny_przewodnik.pdf, dostęp: 30 maja 2016).

Jassem K. 2001: *POLENG: System tłumaczenia automatycznego z języka polskiego na język angielski*, [w:] S. Gawiejnowicz (red.), *Wybrane zastosowania współczesnej informatyki*, Poznańskie Towarzystwo Przyjaciół Nauk, Poznań.

Jaworski W. 2008: *Contents modelling of Neo-Sumerian Ur III economic text corpus*, [w:] D. Scott, H. Uszkoreit (red.), *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, Coling 2008 Organizing Committee, Manchester, UK, s. 369–376.

Jaworski W. 2013: *Przetwarzanie języka polskiego za pomocą kategorialnej gramatyki logicznej* (online: http://fit2013.mat.umk.pl/slides/FIT-2013-Jaworski.pdf, dostęp: 15 lutego 2016).

Kłopotek M.A., Wierzchoń S.T., Trojanowski K. (red.) 2006: *Intelligent Information Processing and Web Mining: Proceedings of the International IIS: IIPWM'06 Conference held in Ustroń, Poland, June 19–22, 2006*, Advances in Soft Computing, Springer, Berlin.

Kobyliński Ł. 2014: *PoliTa*: *A multitagger for Polish*, [w:] N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, ELRA, Reykjavík, Iceland, s. 2949–2954 (online: http://www.lrec-conf.org/proceedings/lrec2014/index.html, dostęp: 30 października 2015).

Marciniak M., Mykowiecka A. (red.) 2009: *Aspects of natural language processing. Essays dedicated to Leonard Bolc on the Occasion of His 75th Birthday*, Lecture Notes in Computer Science, vol. 5070, Springer, Berlin.

Marciniak M., Mykowiecka A. 2014: *Terminology extraction from medical texts in Polish*, „Journal of Biomedical Semantics" 5 (online: doi: 10.1186/2041-1480-5-24).

Marciniak M., Rabiega-Wiśniewska J. 2010: *Elektroniczny słownik fleksyjny nazw Warszawy*, „Prace Filologiczne" LVIII, s. 271–282.

Marciniak M., Savary A., Sikora P., Woliński M. 2011: *Toposław – A Lexicographic Framework for Multi-word Units*, [w:] Z. Vetulani (red.), *Human Language Technology. Challenges for Computer Science and Linguistics. 4th Language and Technology Conference, LTC 2009, Poznań, Poland, November 6–8, 2009, Revised Selected Papers*, Lecture Notes in Artificial Intelligence, vol. 6562, Springer, Berlin, s. 139–150.

Marcińczuk M., Kocoń J., Janicki M. 2013: *Liner2 – A Customizable Framework for Proper Names Recognition for Polish*, [w:] R. Bembenik, L. Skonieczny, H. Rybiński, M. Kryszkiewicz, M. Niezgódka (red.), *Intelligent Tools for Building a Scientific Information Platform. Advanced Architectures and Solutions*, Springer, Berlin–Heidelberg, s. 231–253, (online: doi: 10.1007/978-3-642-35647-6_17).

Marczewski P. 2012: *Building a lexicon for a categorial grammar of the Polish language*, praca magisterska, Wydział Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego, Warszawa.

Maziarz M., Piasecki M., Szpakowicz S. 2012: *Approaching plWordNet 2.0*, [w:] *Proceedings of the 6th Global Wordnet Conference*, Matsue, Japan (online: http://www.plwordnet.pwr.wroc.pl/ltg/files/publications/paper%2042.pdf, dostęp: 20 października 2015).

Mykowiecka A. 2007: *Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym*, Wydawnictwo Polsko-Japońskiej Wyższej Szkoły Technik Komputerowych, Warszawa.

Obrębski T. 2002: *Automatyczna analiza składniowa języka polskiego z wykorzystaniem gramatyki zależnościowej*, rozprawa doktorska, Instytut Podstaw Informatyki PAN, Warszawa.

Ogrodniczuk M. 2006: *Weryfikacja korpusu wypowiedników polskich (z wykorzystaniem gramatyki formalnej Świdzińskiego)*, praca doktorska, Wydział Neofilologii Uniwersytetu Warszawskiego, Warszawa (online: http://bc.klf.uw.edu.pl/30/, dostęp: 20 października 2015).

Ogrodniczuk M., Głowińska K., Kopeć M., Savary A., Zawisławska M. 2015: *Coreference in Polish. Annotation, Resolution and Evaluation*, Walter De Gruyter, Berlin–Boston.

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

Pawłowski A., Krajewski M., Eder M. 2010: *Time series modelling in the analysis of Homeric verse*, „Eos" XCVII, s. 79–100.

Piasecki M., Godlewski G. 2006: *Reductionistic, Tree and Rule Based Tagger for Polish*, [w:] M.A. Kłopotek, S.T. Wierzchoń, K. Trojanowski (red.), *Intelligent Information Processing and Web Mining. Proceedings of the International IIS: IIPWM'06 Conference held in Ustroń, Poland, June 19–22, 2006*, Advances in Soft Computing, Springer, Berlin, s. 531–540.

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

Piasecki M., Szpakowicz S., Derwojedowa M., Zawisławska M. 2008: *Lexical units as the centrepiece of a wordnet*, [w:] *Conference on Intelligent Information Systems 2008*, Challenging Problems in Science: Computer Science, s. 351–357.

Polański K. (red.) 1980–1992: *Słownik syntaktyczno-generatywny czasowników polskich*, Zakład Narodowy im. Ossolińskich, Wrocław.

Przepiórkowski A. 2004a: *Korpus Instytutu Podstaw Informatyki PAN. Wersja wstępna*, Instytut Podstaw Informatyki PAN, Warszawa.

Przepiórkowski A. 2004b: *Towards the design of a syntactico-semantic lexicon for Polish*, [w:] *Intelligent Information Processing and Web Mining. Proceedings of the International IIS: IIPWM'06 Conference held in Ustroń, Poland, June 19–22, 2006*, „Advances in Soft Computing", Springer, Berlin, s. 237–246.

Przepiórkowski A. 2007: *Towards a partial grammar of Polish for valence extraction*, [w:] *Selected contributions from the conference Grammar and Corpora, Sept. 25–27, Liblice*, Academia, Prague, s. 127–133.

Przepiórkowski A. 2008: *Powierzchniowe przetwarzanie języka polskiego*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Przepiórkowski A., Bańko M., Górski R., Lewandowska-Tomaszczyk B. (red.) 2012: *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa (online: http://nkjp.pl/settings/papers/NKJP_ksiazka.pdf, dostęp: 30 października 2015).

Przepiórkowski A., Hajnicz E., Patejuk A., Woliński M., Skwarski F., Świdziński M. 2014a: *Walenty: Towards a comprehensive valence dictionary of Polish*, [w:] N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, ELRA, Reykjavík, Iceland, s. 2785–2792 (online: http://www.lrec-conf.org/proceedings/lrec2014/index.html, dostęp: 30 października 2015).

Przepiórkowski A., Kupść A., Marciniak M., Mykowiecka A. 2001: *Formalny opis języka polskiego. Teoria i implementacja*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.

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

Przybyła P. 2015: *Odpowiadanie na pytania w języku polskim z użyciem głębokiego rozpoznawania nazw*, rozprawa doktorska, Instytut Podstaw Informatyki PAN, Warszawa.

Rabiega-Wiśniewska J. 2008: *Formalny opis derywacji w języku polskim. Rzeczowniki i przymiotniki*, Wydawnictwo Wydziału Polonistyki Uniwersytetu Warszawskiego, Warszawa.

Rabiega-Wiśniewska J., Rudolf M. 2002: *Amor – program automatycznej analizy fleksyjnej tekstu polskiego*, „Biuletyn Polskiego Towarzystwa Językoznawczego" LVIII, s. 175–186.

Radziszewski A. 2012: *Metody znakowania morfosyntaktycznego i automatycznej płytkiej analizy składniowej języka polskiego*, rozprawa doktorska, Wydział Informatyki i Zarządzania Politechniki Wrocławskiej, Wrocław.

Radziszewski A. 2013: *A tiered CRF tagger for Polish*, [w:] R. Bembenik, L. Skonieczny, H. Rybiński, M. Kryszkiewicz, M. Niezgódka (red.), *Intelligent tools for building a scientific information platform. Advanced architectures and solutions*, Springer, Berlin, s. 215–230.

Radziszewski A., Śniatowski T. 2011: *A Memory-Based Tagger for Polish*, 4th Language and Technology Conference, LCT 2009, Poznań, November 6-8, 2009 (online: https://github.com/mariana-scorp/lt-project/blob/master/papers_on_pos-tagging/Radziszewski%20-%202009%20-%20A%20Memory-Based%20Tagger%20for%20Polish.pdf).

Rudolf M. 2004: *Metody automatycznej analizy korpusu tekstów polskich. Pozyskiwanie, wzbogacanie i przetwarzanie informacji lingwistycznych*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa.

Saloni Z. 1992: *Rygorystyczny opis polskiej deklinacji przymiotnikowej*, „Zeszyty Naukowe Wydziału Humanistycznego Uniwersytetu Gdańskiego. Prace Językoznawcze" 16, Gdańsk, s. 215–228.

Saloni Z. 2000: *Wstęp do koniugacji polskiej*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn.

Saloni Z. 2001: *Czasownik polski*, Wiedza Powszechna, Warszawa.

Saloni Z., Woliński M., Wołosz R., Gruszczyński W., Skowrońska D. 2007/2012: *Słownik gramatyczny języka polskiego* (dokument elektroniczny), wyd. 1, 2, Warszawa.

Saloni Z., Woliński M., Wołosz R., Gruszczyński W., Skowrońska D. 2015: *Słownik gramatyczny języka polskiego*, wyd. 3 (online: http://sgjp.pl/).

Savary A., Waszczuk J. 2012: *Narzędzia do anotacji jednostek nazewniczych*, [w:] A. Przepiórkowski, M. Bańko, R. Górski, B. Lewandowska-Tomaszczyk (red.), *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa (online: http://nkjp.pl/settings/papers/NKJP_ksiazka.pdf, dostęp: 30 października 2015), s. 225–252.

Szafran K. 1997: *Automatyczna analiza fleksyjna tekstu polskiego (na podstawie „Schematycznego indeksu* a tergo*" Jana Tokarskiego)*, rozprawa doktorska, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa.

Szmal P., Suszczańska N. 2002: *Tłumaczenie tekstów na język migowy w systemie TGT-1: zasady i realizacja*, „Speech and Language Technology. Technologia Mowy i Języka", t. 6, s. 113–124.

Świdziński M. 1992: *Gramatyka formalna języka polskiego*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Świdziński M. 1994: *Syntactic Dictionary of Polish Verbs*, Uniwersytet Warszawski, Universiteit van Amsterdam, Warszawa–Amsterdam.

Świdziński M., Woliński M. 2009: *A new formal definition of Polish nominal phrases*, [w:] M. Marciniak, A. Mykowiecka (red.), *Aspects of natural language processing. Essays dedicated to Leonard Bolc on the occasion of his 75th Birthday*, Lecture Notes in Computer Science, vol. 5070, Springer, Berlin, s. 143–162.

Świdziński M., Woliński M. 2010: *Towards a bank of constituent parse trees for Polish*, [w:] P. Sojka, A. Horák, I. Kopeček, K. Pala (red.), *Text, Speech and Dialogue. 13th International Conference, TSD 2010*, Brno, Czech Republic, Lecture Notes in Artificial Intelligence, vol. 6231, Springer, Heidelberg, s. 197–204.

Tokarski J. 1951: *Czasowniki polskie. Formy, typy, wyjątki. Słownik*, Wydawnictwo S. Arcta, Warszawa.

Tokarski J. 1961a: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Deklinacja rzeczownikowa i inne nieprzymiotnikowe*, „Poradnik Językowy", nr 4, s. 156–163, nr 8, s. 343–352.

Tokarski J. 1961b: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Założenia ogólne. Deklinacja przymiotnikowa*, „Poradnik Językowy", nr 3, s. 97–111.

Tokarski J. 1962: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Koniugacja*, „Poradnik Językowy", nr 4, s. 145–158.

Tokarski J. 1963a: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Rozpoznawanie form deklinacji przymiotnikowej*, „Poradnik Językowy", nr 1, s. 2–21.

Tokarski J. 1963b: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Rozpoznawanie form deklinacji rzeczownikowej*, „Poradnik Językowy", nr 2, s. 55–77.

Tokarski J. 1963c: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Rozpoznawanie form osobowych w koniugacji*, „Poradnik Językowy", nr 3–4, s. 111–130.

Tokarski J. 1963d: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Rozpoznawanie koniugacyjnych form nieosobowych*, „Poradnik Językowy", nr 5–6, s. 173–184.

Tokarski J. 1973: *Fleksja polska*, Państwowe Wydawnictwo Naukowe, Warszawa.

Tokarski J. 1993: *Schematyczny indeks* a tergo *polskich form wyrazowych*, red. i oprac. Z. Saloni, Wydawnictwo Naukowe PWN, Warszawa.

Vetulani Z. (red.) 2011: *Human Language Technology. Challenges for Computer Science and Linguistics: 4th Language and Technology Conference, LTC 2009, Poznań, November 6–8, 2009, Revised Selected Papers*, Lecture Notes in Artificial Intelligence, vol. 6562, Springer, Berlin.

Vetulani Z., Kubis M., Obrębski T. 2010a: *PolNet – Polish WordNet: Data and tools*, [w:] N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (red.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta, s. 3793–3797 (online: http://www.lrec-conf.org/proceedings/lrec2010/, dostęp: 20 października 2015).

Vetulani Z., Marcinak J., Obrębski T., Vetulani G., Dąbrowski A., Kubis M., Osiński J., Walkowska J., Kubacki P., Witalewski K. 2010b: *Zasoby językowe i technologie przetwarzania tekstu. POLINT-112-SMS jako przykład aplikacji z zakresu bezpieczeństwa publicznego*, Wydawnictwo Uniwersytetu im. Adama Mickiewicza, Poznań.

Vetulani Z., Martinek J., Obrębski T., Vetulani G. 2008: *Dictionary based methods and tools for language engineering*, Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, Poznań.

Woliński M. 2004: *Komputerowa weryfikacja gramatyki Świdzińskiego*, rozprawa doktorska, Instytut Podstaw Informatyki PAN (online: http://www.ipipan.waw.pl/ wolinski/publ/mw-phd.pdf, dostęp: 5 listopada 2015).

Woliński M. 2006: *Morfeusz – a practical tool for the morphological analysis of Polish*, [w:] M.A. Kłopotek, S.T. Wierzchoń, K. Trojanowski (red.), *Intelligent Information Processing and Web Mining. Proceedings of the International IIS: IIPWM'06 Conference held in Ustroń, Poland, June 19–22, 2006*, Advances in Soft Computing, Springer, Berlin, s. 503–512.

Woliński M. 2014: *Morfeusz reloaded*, [w:] N. Calzolari i in. (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, ELRA, Reykjavík, Iceland (online: http://www.lrec-conf.org/proceedings/lrec2014/index.html, dostęp: 30 października 2015), s. 1106–1111.

Woliński M., Głowińska K., Świdziński M. 2011: *A preliminary version of Składnica – a treebank of Polish*, [w:] Z. Vetulani (red.), *Human Language Technology. Challenges for Computer Science and Linguistics: 4th Language and Technology Conference, LTC 2009, Poznań, Poland, November 6–8, 2009, Revised Selected Papers*, Lecture Notes in Artificial Intelligence, vol. 6562, Springer, Berlin, s. 299–303.

Woliński M., Miłkowski M., Ogrodniczuk M., Przepiórkowski A., Szałkiewicz Ł. 2012: *PoliMorf: a (not so) new open morphological dictionary for Polish*, [w:] N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, LREC 2012, ELRA, Istanbul, Turkey, s. 860–864 (online: http://www.lrec-conf.org/proceedings/lrec2012/index.html, dostęp: 20 października 2015).

Wołosz R. 2005: *Efektywna metoda analizy i syntezy morfologicznej w języku polskim*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Wróblewska A. 2012: *Polish dependency bank*, „Linguistic Issues in Language Technology" 7 (1) (online: http://elanguage.net/journals/index.php/lilt/article/view/2684).

Summary

Computational linguistics and language engineering after 1989

Keywords: computational linguistics, language engineering, natural language processing.

The paper presents research on NLP in Poland in the last 25 years. In this review tools and resources for syntactical and morphological research with the exception of corpora and dictionaries have been presented. Formal grammars, parsers, taggers, morphological analysers, valence dictionaries, wordnets, multi-word dictionaries and other lexical resources of this kind are briefly discussed.

MACIEJ OGRODNICZUK*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

Lingwistyka komputerowa dla języka polskiego: dziś i jutro

Słowa kluczowe: lingwistyka komputerowa, lingwistyka informatyczna, przetwarzanie języka naturalnego, technologie językowe, narzędzia i zasoby językowe dla polszczyzny.

1. Wprowadzenie

W 2006 roku na łamach „LingVariów" ukazał się entuzjastyczny tekst Marka Świdzińskiego na temat zdobyczy polskiej lingwistyki korpusowej, w którym autor przedstawiał korzyści, jakie językoznawstwu przyniosła, także w Polsce, rewolucja informatyczna. Sposób, w jaki środowisko językoznawcze przyjęło te zmiany, dość dobrze opisują z kolei w publicystycznych artykułach Piotr Żmigrodzki (2015a) i Adam Przepiórkowski (2015). W niniejszym tekście chciałbym spojrzeć na sytuację lingwistyki informatycznej w Polsce oraz na dalsze kierunki prac nad komputerowym przetwarzaniem polszczyzny z perspektywy o dziesięć lat późniejszej oraz z punktu widzenia przedstawiciela tej dyscypliny, często postrzeganej wyłącznie jako usługowa dla środowiska językoznawczego.

Zacznijmy zatem od kilku słów na temat przedmiotu naszej pracy i trudności, przed jakimi stoimy. Lingwistyka komputerowa (lingwistyka informatyczna, inżynieria lingwistyczna) to dyscyplina z pogranicza obu dziedzin – informatyki i językoznawstwa, której celem jest badanie języka naturalnego z punktu widzenia potrzeb jego przetwarzania i modelowania metodami komputerowymi oraz tworzenie elektronicznych zasobów i narzędzi przetwarzania języka na potrzeby usługowe. Powyższa definicja, łącząca istniejące dotąd definicje (w szczególności dwie: Janusza S. Bienia i Bonnie Webber[1]), zwraca uwagę na dwa aspekty problemu: jest to jednocześnie dziedzina teoretyczna i stosowana. O pierwszym zdają się zapominać językoznawcy, o drugim – sami informatycy. Nie znam szczegółów tego rozstrzygnięcia, ale o stosunku polskiego środowiska językoznawczego do lingwistyki komputerowej jako dziedziny badań w obszarze nauk humanistycznych najlepiej świadczy niedawne usunięcie jej z panelu HS 2.6 w konkursach Narodowego Centrum Nauki, chociaż w projektach prowadzonych w moim rodzimym Zespole Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN często więcej jest pracy ściśle językoznawczej (wykonywanej przez zaprzyjaźnionych lingwistów) niż informatycznej. Taka jest zresztą specyfika większości polskich zespołów zajmujących się przetwarzaniem języka: naszą domeną jest właśnie lingwistyka informatyczna, a nie

* maciej.ogrodniczuk@ipipan.waw.pl

1 Bień 1988; Webber 2001, zob. też Piasecki 2008.

informatyka lingwistyczna, z konsekwencjami w postaci rezygnacji z wyścigu w rozwoju technologii anglojęzycznej na rzecz świadomej lingwistycznie technologii dla polszczyzny; stąd też nasza częsta obecność w najważniejszych polskich pismach językoznawczych.

Po stronie informatycznej grzechem głównym jest właśnie trudność z przyznawaniem się do częściowo usługowego wymiaru lingwistyki komputerowej i kłopoty komunikacyjne obu światów. Być może wynikają one z faktu postrzegania informatyków jako wykonawców, a nie partnerów w pracy badawczej. Usługowość rozumiem zatem nie jako gotowość nauczenia lingwistów języka zapytań do wyszukiwarki czy pomoc w stworzeniu własnego korpusu, ale jako rzeczywistą współpracę w badaniach.

Mimo wspomnianych trudności mój punkt widzenia na stan lingwistyki informatycznej w Polsce jest optymistyczny: jesteśmy świadkami bardzo intensywnego rozwoju komputerowych zasobów i narzędzi dla języka polskiego[2], polskie środowisko lingwistyczno-informatyczne bierze udział w licznych grantach polskich i europejskich, uczestniczy w dużych infrastrukturach badawczych CLARIN i DARIAH, łączących nauki techniczne i humanistyczne, a sami użytkownicy humaniści coraz częściej sięgają po metody komputerowe naszego autorstwa. Trwa owocna współpraca między polskimi ośrodkami badawczymi, a komputerowy opis polszczyzny obejmuje coraz głębsze poziomy analizy. W dalszej części tekstu postaram się pokazać, w którym miejscu jesteśmy, i zaproponować kilka kierunków badań na najbliższe lata, jeśli nie dla całego środowiska, to przynajmniej dla naszego warszawskiego zespołu.

2. Stan technologii językowej dla polszczyzny okiem lingwisty komputerowego

2.1. Raport językowy

W roku 2011 warszawski Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN brał udział w międzynarodowym projekcie CESAR, którego celem (wraz z innymi inicjatywami podobnego typu prowadzonymi pod wspólnym szyldem META-NET-u) było zebranie, scalenie oraz zrównoleglenie elektronicznych zasobów i narzędzi językowych dla trzydziestu języków europejskich, a następnie udostępnienie ich we wspólnym repozytorium. Miało to stworzyć nowe możliwości badawcze oraz technologiczne w dziedzinie przetwarzania języków naszego kontynentu. Projekt zakończył się, moim zdaniem, sporym sukcesem: doprowadził do powstania wielu ważnych zasobów, jak np. *Słownik gramatyczny języka polskiego* (Saloni i in. 2007/2015), doprowadził do unormowania kwestii licencyjnych oraz odświeżył stronę internetową CLIP (Ogrodniczuk, Przepiórkowski 2013)[3] z rejestrem dostępnych publicznie narzędzi dla polszczyzny oraz polskich zespołów badawczych zajmujących się przetwarzaniem języka i prowadzonych przez nie projektów.

Innym ważnym wynikiem prac był tzw. raport językowy (Miłkowski 2012) upowszechniający wiedzę nt. technologii językowych dla polszczyzny i ich potencjalnych zastosowań. Jego ważnym zadaniem było też przedstawienie obrazu bieżącej sytuacji w dziedzinie przetwarzania

2 Por. http://clip.ipipan.waw.pl/LRT – strona zawierająca listę dostępnych w sieci narzędzi i zasobów dla polszczyzny.

3 Computational Linguistics in Poland ('Lingwistyka komputerowa w Polsce') – por. http://clip.ipipan.waw.pl (strona w języku angielskim).

języka w Polsce. Na podstawie oceny eksperckiej w skali od 0 (ocena bardzo niska) do 6 (bardzo wysoka) przedstawiał dane nt. dostępności, jakości i elastyczności technologii oraz zasobów w orientacyjnie zarysowanych dziedzinach. Wyniki tej oceny przytaczam w tabeli 1.

Tabela 1. Stan dostępnych technologii językowych dla języka polskiego (przedruk, s. 31 raportu)

	Liczba	Dostępność	Jakość	Zakres	Dojrzałość	Trwałość	Elastyczność
Technologie językowe (narzędzia, technologie, aplikacje)							
Rozpoznawanie mowy	1	2	3	4	3	2	4
Synteza mowy	4	3	6	5	4	4	3
Analiza gramatyczna	4	4,5	4,5	4,5	4	4	3
Analiza semantyczna	1	1	3	1	1	2	2
Generowanie tekstu	1	1	1	1	1	1	2
Tłumaczenie maszynowe	3	4	3	3	3	4	3
Zasoby językowe (zasoby, dane, bazy wiedzy)							
Korpusy tekstów	3	2	4	4	5	5	3
Korpusy równoległe	3	1	4	4	5	5	5
Korpusy języka mówionego	1	0	3	3	2	2	2
Zasoby leksykalne	3	3	4	4	4	4	3
Gramatyki	3	2	4	4	3	2	2

Wykaz ten dobrze obrazuje stan prac sprzed pięciu lat: już wtedy doskonale radziliśmy sobie z syntezą mowy[4], nieźle z analizą morfoskładniową (tu oznaczoną jako gramatyczna), jako dobrą ocenialiśmy dostępność jednojęzycznych korpusów tekstów[5], jako gorszą – korpusów równoległych; z dostępnością korpusów mówionych było gorzej niż źle. Raport zawiera więcej podobnych wniosków, głównie pesymistycznych: wskazuje na brak standaryzacji zasobów oraz na konieczność intensyfikacji prac nad głęboką analizą tekstu, zasobami ontologicznymi i elektronicznymi słownikami walencyjnymi.

4 Był to moment triumfu polskiego systemu Ivona w wielu międzynarodowych konkursach syntezy mowy, po którym firma będąca jego autorem została przejęta przez globalną firmę Amazon.

5 Ze względu na moment zakończenia prac nad Narodowym Korpusem Języka Polskiego (Przepiórkowski i in. (red.) 2012).

Ważnych informacji dostarcza także porównanie stanu technologii dla języka polskiego i innych języków europejskich; tabela 2 przedstawia jego wyniki dla zadania analizy tekstu.

Tabela 2. Stan technologii językowych dostępnych dla 30 języków europejskich dla zadania analizy tekstu (przedruk, s. 33 raportu)

Doskonała jakość	Bardzo dobra jakość	Dobra jakość	Średnia jakość	Słaba/zerowa jakość
	angielski	francuski hiszpański niderlandzki niemiecki włoski	baskijski bułgarski czeski duński fiński galisyjski grecki kataloński norweski **polski** portugalski rumuński słowacki słoweński szwedzki węgierski	chorwacki estoński irlandzki islandzki litewski łotewski maltański serbski

2.2. Sonda środowiskowa

Z perspektywy czasu widać, że zarówno usytuowanie polszczyzny w rodzinie języków europejskich, jak i wnioski na temat stanu rozwoju technologii były właściwe; kierunki najintensywniejszych prac prowadzonych w ciągu ostatnich lat pokrywają się z powyższymi postulatami. Powstaje duży semantyczny słownik walencyjny Walenty (Przepiórkowski i in. 2014), z którego zaczynają korzystać parsery, czyli narzędzia do głębokiej analizy tekstu (Patejuk, Przepiórkowski 2012; Jaworski, Przepiórkowski 2014; Woliński 2015), od wielu lat dynamicznie rozwija się Słowosieć (Piasecki i in. 2014), trwają prace nad poprawą jakości i dostępności korpusów równoległych i mówionych (Pęzik 2015, 2016).

Wiarygodną ocenę stanu obecnego najlepiej jednak przeprowadzić metodą konsultacji ze środowiskiem, postanowiłem zatem powtórzyć ocenę metodą prostej sondy. Na etapie tworzenia pytań na podstawie tabeli z raportu językowego okazało się, że określenia rodzajów narzędzi i zasobów są zbyt ogólne (stąd zapewne dobra ocena jakości „analizy gramatycznej", który to termin byłby dziś raczej kojarzony z analizą składniową, a nie morfoskładniową), postanowiłem zatem przejąć listę bardziej szczegółowych kategorii z istniejących badań zagranicznych (np. Strik i in. 2002) i dostosować ją do specyfiki polskiego środowiska. Sonda w części dotyczącej obecnego stanu prac składała się z dwóch pytań:

1. Jak ocenia Pan/Pani dostępność narzędzi i zasobów dla polszczyzny?
   Proszę o podanie Państwa opinii nt. stopnia możliwości wykorzystania istniejących narzędzi i zasobów dla języka polskiego.
   a. Niedostępne
   b. Dostępne w ograniczonym stopniu (np. za opłatą, w ograniczonym zakresie)
   c. Dostępne w wystarczającym zakresie
   d. Nie wiem
2. Jak ocenia Pan/Pani jakość technologii językowych dla polszczyzny?
   Proszę o podanie Państwa opinii nt. ogólnej jakości dostępnej obecnie technologii (darmowej lub komercyjnej) dla języka polskiego, biorąc pod uwagę liczbę i wielkość zasobów, różnorodność i zakres dziedzinowy itp.
   a. Brak lub słaba jakość
   b. Średnia jakość
   c. Dobra jakość
   d. Doskonała jakość
   e. Nie wiem

Lista ocenianych narzędzi i zasobów liczyła 29 pozycji:

1. Rozpoznawanie mowy
2. Synteza mowy
3. Segmentacja tekstu
4. Analiza morfologiczna
5. Synteza morfologiczna
6. Lematyzacja
7. Tagowanie (ujednoznacznianie morfoskładniowe)
8. Parsowanie zależnościowe
9. Parsowanie powierzchniowe
10. Parsowanie głębokie
11. Rozpoznawanie nazw własnych
12. Ujednoznacznianie sensu słów
13. Analiza sentymentu/opinii
14. Głęboka analiza semantyczna
15. Analiza referencji
16. Analiza pragmatyczna
17. Analiza dyskursu
18. Korekta pisowni
19. Generowanie tekstu
20. Tłumaczenie maszynowe
21. Streszczanie tekstu
22. Ekstrakcja informacji
23. Korpusy języka pisanego
24. Korpusy języka mówionego
25. Korpusy równoległe
26. Słowniki elektroniczne
27. Gramatyki formalne
28. Tezaurusy
29. Wordnety

Ankieta została rozesłana do dwustu osób związanych ze środowiskiem lingwistyczno-informatycznym w Polsce, zebrano 23 odpowiedzi z 14 jednostek (uczelni, instytutów badawczych i firm). Tabela 3 przedstawia wyniki sondy ograniczone do kategorii użytych w raporcie z 2011 roku i przeskalowane w sposób umożliwiający porównanie wyników obecnych z tymi sprzed pięciu lat.

Tabela 3. Stan dostępnych technologii językowych dla języka polskiego na podstawie sondy przeprowadzonej wśród przedstawicieli środowiska lingwistyczno-informatycznego

	Liczba	Dostępność	Jakość	Zakres	Dojrzałość	Trwałość	Elastyczność
Technologie językowe (narzędzia, technologie, aplikacje)							
Rozpoznawanie mowy	1	3	3	4	3	2	4
Synteza mowy	4	4	5	5	4	4	3
Analiza gramatyczna	4	4,5	5	4,5	4	4	3
Analiza semantyczna	1	1	2	1	1	2	2
Generowanie tekstu	1	3	2	1	1	1	2
Tłumaczenie maszynowe	3	3	2	3	3	4	3
Zasoby językowe (zasoby, dane, bazy wiedzy)							
Korpusy tekstów	3	4	4	4	5	5	3
Korpusy równoległe	3	3	3	4	5	5	5
Korpusy języka mówionego	1	3	3	3	2	2	2
Zasoby leksykalne	3	3	4	4	4	4	3
Gramatyki	3	3	3	4	3	2	2

Oznaczenia jaśniejszym kolorem odpowiadają poprawie oceny, ciemniejszym – jej pogorszeniu. Wynik negatywny należy raczej interpretować jako zmianę w sposobie postrzegania danej dziedziny (w porównaniu z technologiami podobnego typu albo dostępnymi dla innych języków) niż faktyczny spadek jakości danego narzędzia czy zasobu. Ze względu na obszerniejszą listę kategorii warto przyjrzeć się także wynikom dla technologii, które nie były uwzględnione na liście z 2011 r.: wśród nich za najbardziej dostępne (o dostępności w wystarczającym zakresie) uznano oprócz wskazanych w tabeli także mechanizmy segmentacji tekstu, tagowania i wordnety. Za niedostępne uznano z kolei narzędzia do głębokiej analizy semantycznej i pragmatycznej oraz analizy dyskursu. Najlepszą jakość przypisano natomiast technologii syntezy mowy, segmentacji tekstu, analizy i syntezy morfologicznej oraz korekty pisowni (doskonała jakość). Spośród technologii dostępnych najniżej oceniono jakość technologii do streszczania tekstu, a niewiele lepiej do parsowania głębokiego, ujednoznaczniania sensu słów, analizy sentymentu/opinii, generowania tekstu i, co już zostało pokazane w tabeli, tłumaczenia maszynowego. W mojej opinii oceny te odpowiadają obecnemu stanowi rozwoju wymienionych technologii.

3. Zadania dla lingwistyki komputerowej w Polsce

Postulaty, które teraz przedstawię, należy odczytać jako w dużej mierze subiektywną prezentację kierunków rozwoju lingwistyki komputerowej w Polsce. Mam nadzieję, że moje propozycje będą atrakcyjne dla obu środowisk naukowych – humanistycznego i informatycznego, pomogą odpowiedzieć na ich potrzeby i pozwolą im się jeszcze bardziej zbliżyć. W pewnym stopniu poruszają one problemy zasygnalizowane w odpowiedziach ankietowych, ale przedstawiam je jednak z perspektywy planów (i aspiracji) warszawskiego zespołu.

3.1. Korpus narodowy

Pierwsza wersja Narodowego Korpusu Języka Polskiego była organizacyjnym i badawczym przełomem, który jeszcze długo będzie wpływał na środowisko humanistyczne i lingwistyczno-informatyczne w Polsce. Prace prowadzone w ramach projektu rozwojowego MNiSW doprowadziły do powstania nie tylko samego zasobu korpusowego, ale także wielu narzędzi informatycznych do przetwarzania języka polskiego. Warto jednak zwrócić uwagę, że tryb projektowy doprowadził do powstania zbioru, który od zakończenia prac trwa w postaci zamrożonej – ostatnie teksty pochodzą z roku 2011, a automatyczne analizy lingwistyczne zostały wytworzone narzędziami o kilka generacji wcześniejszymi niż używane obecnie. Zapewnienie aktualności i stałego monitoringu NKJP, który stanowi przecież jeden z podstawowych zasobów referencyjnych dla rzeszy użytkowników, wydaje się koniecznością. A potrzeb jest dużo więcej: równie ważne wydaje się wielokierunkowe rozszerzanie bazy materiałowej, opracowanie metod reprezentacji i wizualizacji danych lingwistycznych, rozwój narzędzi analitycznych i eksploracyjnych. Próbę dokładniejszego zdefiniowania potrzeb w tym zakresie podjęliśmy już w ramach grupy roboczej „Korpusowa infrastruktura badawcza dla polszczyzny" powołanej w ramach konsorcjum DARIAH-PL, warto zatem powtórzyć kierunki, w których widzimy rozwój korpusu narodowego:

1) Rozszerzenie bazy materiałowej Narodowego Korpusu Języka Polskiego (o podkorpusy diachroniczne, podkorpusy równoległe, specjalistyczne, w tym gwarowy, dane mówione, podkorpus synchroniczny tworzony na bazie bibliotek cyfrowych i Internetu).
2) Opracowanie korpusowych standardów znakowania dla języka polskiego:
   a) opracowanie formatu znakowania wszystkich tekstów (także historycznych i gwarowych) rozszerzonymi metadanymi i znacznikami morfosyntaktycznymi,
   b) opracowanie formatu znakowania tekstów współczesnych znacznikami składniowymi, semantycznymi i dyskursywnymi.
3) Opracowanie metod reprezentacji danych multimedialnych:
   a) opracowanie metod zrównoleglania tekstów ze źródłami (dane mówione, wideo, skany),
   b) opracowanie metod wyszukiwania w plikach źródłowych.
4) Opracowanie interfejsu użytkownika korpusu uwzględniającego proste formułowanie zapytań oraz atrakcyjną i czytelną wizualizację wyników:
   a) opracowanie intuicyjnych nakładek graficznych i interakcyjnych,

b) opracowanie metod wizualizacji zaawansowanych struktur lingwistycznych (drzewa składniowe, koreferencje itp.).
5) Rozwój narzędzi do samodzielnego tworzenia korpusów.
6) Rozwój narzędzi do eksploracji i analizy danych korpusowych.
7) Monitoring źródeł internetowych na potrzeby korpusowe i leksykograficzne.

O istotności prac w każdym z tych kierunków świadczy fakt uruchomienia wielu projektów finansowanych ze środków Narodowego Programu Rozwoju Humanistyki oraz Narodowego Centrum Nauki, w których ramach powstają obecnie korpusy diachroniczne: XVI, XVII–XVIII i XIX wieku, korpus polskich tekstów prasowych (aktualnie z lat 1945–1954) czy korpus gwarowy. Jednocześnie powstają również reprezentacje zaawansowanych struktur lingwistycznych w rodzaju relacji referencyjnych czy metafor wraz z narzędziami analitycznymi. Projekty te, nawet jeśli wykorzystują dane korpusu narodowego, są jednak prowadzone niezależnie, bez intencji jego wzbogacenia, co wydaje mi się dużą stratą. A przecież podobnie do *Wielkiego słownika języka polskiego* PAN (Żmigrodzki 2015b) wielki korpus polszczyzny jest podstawowym zasobem językowym, który powinien istnieć w sposób stały i stanowić rodzaj „instytucji narodowej". Prace nad nim nie mogą się ograniczać do jednego projektu – wymagają nadzoru środowiska humanistycznego, doradztwa naukowego na najwyższym poziomie i, co też trzeba powiedzieć, stałego finansowania. Szczególnie ten ostatni postulat wydaje się trudny do spełnienia, jak może się wydawać po lekturze długiej listy potrzeb, jednak najpilniejsza kwestia rozpoczęcia prac nad utrzymaniem korpusu nie wydaje się aż tak mało realistyczna i w moim przekonaniu wystarczyłoby do niej jednoczesne osiągnięcie trzech celów:
1) zapewnienie ciągłości życia NKJP, co stanowiłoby spełnienie części postulatów autorów monografii korpusu (Przepiórkowski i in. (red.) 2012) jego ustawicznego monitorowania oraz uzupełniania i ulepszania, w fazie wstępnej ograniczonego do dostosowania opisów lingwistycznych do najnowszych narzędzi analitycznych; tego rodzaju działania wymagałyby pracy zaledwie kilku osób,
2) zapewnienie nadzoru nad długofalowymi kierunkami rozwoju NKJP przez radę programową złożoną z wybitnych językoznawców,
3) opracowanie metody konstruowania grantów w sposób nagradzający włączanie ich wyników do zasobu korpusu narodowego i zapewniający finansowe wsparcie tego procesu.

Powyższa formuła jest oczywiście luźną propozycją, której celem jest rozpoczęcie dyskusji nad stworzeniem zasad funkcjonowania i rozwoju korpusu narodowego. Na jej bazie chciałbym niniejszym rozpocząć lobbing na rzecz tego ważnego zasobu i poprosić czytelników o poparcie tej inicjatywy.

3.2. Nowe tematy – i mozolnie naprzód

Drugim istotnym kierunkiem prac jest, moim zdaniem, formalny opis polszczyzny w zakresie od dawna podejmowanym dla języków zachodnich, a u nas wciąż traktowanym jako pieśń

przyszłości – chodzi przede wszystkim o kwestie związane z analizą dyskursu rozumianą jako generowanie struktur wypowiedzi wykraczających poza poziom zdania. Temat ten wpisuje się w prowadzone obecnie prace nad komputerowym opisem takich zjawisk, jak referencja, metafora oraz argumentacja w języku.

Oczywiście mój subiektywny punkt widzenia może nie odzwierciedlać istotności tematów dla pozostałej części środowiska lingwistyczno-informatycznego, sięgnę zatem raz jeszcze do ankiety, w której zadałem też trzecie pytanie, prosząc o podanie informacji o planach rozwoju danej technologii w zespole respondenta, ze wspomnianą wyżej szczegółową listą 29 technologii, i otrzymałem następujące odpowiedzi:

1. Jakość technologii jest wystarczająca, nie ma potrzeby prowadzenia dalszych prac.
2. Obecnie prowadzimy prace w tej dziedzinie.
3. Zamierzamy rozwijać narzędzia/zasoby z tej dziedziny w przyszłości.
4. Temat jest ważny, ale nie planujemy prowadzić prac w tej dziedzinie.
5. Nie uważamy tego tematu za istotny.
6. Trudno powiedzieć.

Intencją pytania było sprawdzenie, które problemy uważamy jako środowisko za rozwiązane, które są dla nas ważne (bo prowadzimy już prace w danej dziedzinie, zamierzamy je prowadzić lub uważamy temat za istotny), a które nie są warte badawczej uwagi. Ciekawym wynikiem sondy okazało się stwierdzenie, że praktycznie nie ma dziedzin, które zostałyby uznane za nieważne; za kwestie rozwiązane najwięcej ankietowanych uznało korektę pisowni (5 odpowiedzi z 23 udzielonych), dostępność korpusów języka pisanego (4), narzędzi do segmentacji tekstu, lematyzacji, analizy morfologicznej oraz takich zasobów jak korpusy równoległe i wordnety (3). Za tematy ważne uznano tłumaczenie maszynowe (19), rozpoznawanie mowy (17), syntezę mowy, tagowanie, analizę sentymentu/opinii i ekstrakcję informacji (16), a także streszczanie tekstu, rozwój korpusów języka pisanego, słowników elektronicznych i gramatyk formalnych (15). Wskazania te nie zmieniły się znacząco przy ograniczeniu odpowiedzi do technologii „przyszłościowych", czyli odjęciu odpowiedzi uwzględniających prowadzone prace. W kontekście poprzedniego rozdziału komentarza wymaga obecność na obu listach korpusów języka pisanego; ocena ich dobrej dostępności może wynikać z zaufania do NKJP z jednoczesnym wskazaniem na potrzebę rozwoju korpusów dla polszczyzny – w tym również korpusu narodowego.

Oczywiście przy okazji kreślenia planów na przyszłość nie można pominąć wątku stałego poprawiania jakości dostępnych narzędzi i zasobów. Niedościgły w niektórych dziedzinach poziom 90-procentowej dokładności jest dla wielu celów niewystarczający, gdyż nawet jedno błędne wskazanie na każde dziesięć to za dużo, zwłaszcza gdy wynik jednego narzędzia (np. ujednoznacznienie części mowy w przetwarzanym tekście) jest źródłem danych dla kolejnego (np. analizatora składniowego wykorzystującego informację o częściach mowy). Poprawa narzędzi działających na różnych poziomach wiedzy lingwistycznej – analizatorów składniowych i semantycznych, mechanizmów do wykrywania nazw własnych, koreferencji – choć mało spektakularna, jest niezwykle ważna z perspektywy wykorzystania ich wyników w praktyce.

Oprócz dziedziny *stricte* badawczej nie można też zapomnieć o nie mniej ważnej części „usługowej". Szlak jest już przetarty; środowisko informatyczne bierze udział w dwóch ważnych inicjatywach dotyczących humanistyki cyfrowej – infrastrukturach badawczych CLARIN-PL i DARIAH-PL. CLARIN-PL jest inicjatywą środowiska informatycznego wspierającą projekty badawcze w naukach humanistycznych i społecznych poprzez umożliwienie korzystania z dostępnych zasobów i narzędzi w zadaniach związanych z przetwarzaniem języka. Inicjatorem DARIAH-PL jest natomiast środowisko humanistyczne działające na rzecz tworzenia, rozwoju i utrzymania infrastruktury humanistyki cyfrowej w Polsce. Co ważne, instytucje reprezentujące główne polskie centra badawcze zajmujące się lingwistyką informatyczną[6] są członkami obu inicjatyw, włączając się w badania humanistyczne w zakresie własnych zainteresowań i dostępnych środków.

Mimo wielu kłopotów właściwych polskiej nauce polskie środowisko lingwistyki informatycznej działa bardzo prężnie, a trzy główne tematy: składnia, semantyka i dyskurs, wydają się wyznaczać kierunki prac na najbliższe lata. Muszę jednak po raz kolejny zastrzec, że powyższy tekst przedstawia przede wszystkim moje osobiste poglądy, a w drugiej kolejności (na podstawie odpowiedzi z sondy) punkt widzenia informatyków zajmujących się przetwarzaniem języka, może zatem nie odzwierciedlać stanowiska użytkowników, których priorytety mogą być zgoła inne. W związku z tym zachęcam humanistów do współpracy, dzielenia się swoimi problemami, a także wynikami własnych prac. To najlepsza metoda na poprawianie jakości narzędzi dla polszczyzny, z których oba środowiska będą mogły korzystać.

Bibliografia

Bień J.S. 1988: *Komputery a język naturalny*, „Biuletyn Polskiego Towarzystwa Informatycznego" VI (2–3), s. 6–7.

Jaworski W., Przepiórkowski A. 2014: *Syntactic approximation of semantic roles*, [w:] A. Przepiórkowski, M. Ogrodniczuk (red.), *Advances in Natural Language Processing*, Lecture Notes in Artificial Intelligence, vol. 8686, Springer, Heidelberg, s. 193–201.

Miłkowski M. 2012: *Język polski w erze cyfrowej*, Springer, Berlin–Heidelberg (online: http://doi.org/10.1007/978-3-642-30811-6).

Ogrodniczuk M., Przepiórkowski A. 2013: *CLIP – portal internetowy łączący projekty, narzędzia, zespoły związane z komputerowym przetwarzaniem języka polskiego*, „Język Polski" XCIII, s. 242.

Patejuk A., Przepiórkowski A. 2012: *Towards an LFG parser for Polish. An exercise in parasitic grammar development*, [w:] N. Calzolari, K. Choukri, T. Declerck, M.U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (LREC 2012), ELRA, Stambuł, s. 3849–3852.

Pęzik P. 2015: *Spokes – a search and exploration service for conversational corpus data*, [w:] *Selected Papers from CLARIN 2014*, Linköping University Electronic Press, Linköpings universitet, Linköping, s. 99–109.

Pęzik P. 2016: *Exploring phraseological equivalence with paralela*, [w:] E. Gruszczyńska, A. Leńko-Szymańska (red.), *Polskojęzyczne korpusy równoległe. Polish-language Parallel Corpora*, Instytut Lingwistyki Stosowanej Uniwersytetu Warszawskiego, Warszawa, s. 67–81.

Piasecki M. 2008: *Cele i zadania lingwistyki informatycznej*, [w:] P. Stalmaszczyk (red.), *Metodologie językoznawstwa. Współczesne tendencje i kontrowersje*, Lexis, Kraków.

6 Zob. http://clip.ipipan.waw.pl/Centers.

Piasecki M., Maziarz M., Szpakowicz S., Rudnicka E. 2014: *PlWordNet as the Cornerstone of a Toolkit of Lexico-semantic Resources*, [w:] H. Orav, C. Fellbaum, P. Vossen (red.), *Proceedings of the Seventh International Global Wordnet Conference (GWC 2014)*, University of Tartu Press, Tartu, s. 304–312.

Przepiórkowski A. 2015: *Inżynieria lingwistyczna a obecna sytuacja językoznawstwa polskiego*, „LingVaria" X, nr 2, s. 135–145.

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

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

Saloni Z., Woliński M., Wołosz R., Gruszczyński W., Skowrońska D. 2007/2015: *Słownik gramatyczny języka polskiego*, wyd. 1: Warszawa 2007 (dokument elektroniczny), wyd. 2: Warszawa 2012 (dokument elektroniczny), wyd. 3: 2015 (online: http:// sgjp.pl).

Strik H., Daelemans W., Binnenpoorte D., Sturm J., de Vriend F., Cucchiarini C. 2002: *Dutch HLT resources. From BLARK to priority lists*, [w:] J. Hansen, B. Pellom (red.), *Proceedings of the Seventh International Conference on Spoken Language Processing (ICLSP 2002)*, Denver, USA, s. 1549–1552.

Świdziński M. 2006: *Lingwistyka korpusowa w Polsce – źródła, stan, perspektywy*, „LingVaria" I, nr 1, s. 23–32.

Webber B.L. 2001: *Computational perspectives on discourse and dialogue*, [w:] D. Schirin, D. Tannen, H. Hamilton (red.), *The Handbook of Discourse Analysis*, Blackwell Publishers Ltd.

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

Żmigrodzki P. 2015a: Wielki słownik języka polskiego PAN *a obecna sytuacja językoznawstwa polskiego*, „LingVaria" X, nr 1, s. 109–119.

Żmigrodzki P. 2015b: Wielki słownik języka polskiego PAN – *historia, stan obecny i perspektywy rozwoju po 2018 roku*, „Biuletyn Polskiego Towarzystwa Językoznawczego" LXXI, s. 177–187.

Summary

Natural language processing for Polish: today and tomorrow

Keywords: computational linguistics, natural language processing, language technology, language resources and tools for Polish.

The article attempts at framing directions for future work on computational processing of Polish in the face of recent intensive development of electronic tools and resources and close co-operation between Polish research centres involved in computational linguistics. The author regards renewing the work on the National Corpus of Polish as the most important topic, naming it the basic resource for Polish linguistics and listing the most urgent objectives: extension of the sources and linguistic representation as well as inclusion of diachronic, dialectical and parallel corpora. With respect to language technology, the author calls for enrichment of formal description of Polish with syntactic, semantic and discourse-feature representation and constant improvement of quality of tools and resources by means of co-operation between linguists and computer scientists.

ADAM PRZEPIÓRKOWSKI*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA; UNIWERSYTET WARSZAWSKI

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

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

Walenty: gruntowny składniowo-semantyczny słownik walencyjny języka polskiego

Słowa kluczowe: walencja, składnia, frazeologia, semantyka.

1. Wstęp

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

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

* adam.przepiorkowski@ipipan.waw.pl

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

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

2. Walencja i argumenty

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

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

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

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

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

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

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

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

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

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

3. Warstwa składniowa

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

4. Frazeologia

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

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

W Walentym takie informacje są podane jawnie:

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

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

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

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

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

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

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

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

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

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

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

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

(33) Co rusz walisz głową w mur paragrafów.

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

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

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

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

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

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

5. Warstwa semantyczna

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

6. Dostępność, wykorzystanie i dalszy rozwój

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

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

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

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

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

7. Podsumowanie

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

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

Podziękowania

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

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

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

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

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

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

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

Bibliografia

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rozwiązanie skrótów

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

Summary

Walenty: a thorough syntactico-semantic valency dictionary of Polish

Keywords: Walenty, valency dictionary.

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

AGNIESZKA PATEJUK*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

ADAM PRZEPIÓRKOWSKI**

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA; UNIWERSYTET WARSZAWSKI

POLFIE: współczesna gramatyka formalna języka polskiego

Słowa kluczowe: gramatyka, LFG, polski, składnia.

1. Wstęp

Niniejszy artykuł przedstawia nową gramatykę formalną języka polskiego – POLFIE, która jest rozwijana od 2010 roku w ramach projektów realizowanych w Zespole Inżynierii Lingwistycznej IPI PAN (m.in. NEKST[1] i CLARIN-PL[2]). POLFIE łączy prace w nurcie teoretycznym – gramatyka opiera się na jednej z najważniejszych teorii lingwistycznych, oraz implementacyjnym – gramatyka stanowi podstawę automatycznego systemu analizy składniowej tekstów polskich. Artykuł zawiera krótki opis teorii lingwistycznej leżącej u podstaw gramatyki (§2) oraz prezentację wybranych zjawisk lingwistycznych uwzględnionych w gramatyce POLFIE (§3), a także wspomina o komputerowej implementacji i dostępności gramatyki (§4).

2. Teoria LFG

Lexical Functional Grammar (LFG; Bresnan (red.) 1982; Dalrymple 2001; Bresnan i in. 2015) jest teorią lingwistyczną aktywnie rozwijaną od drugiej połowy lat siedemdziesiątych XX wieku. Chociaż LFG jest teorią generatywną w tym sensie, że ma solidne podstawy formalne, nie należy do chomskiańskich gramatyk transformacyjnych. LFG jest też teorią silnie zleksykalizowaną – wiele zjawisk składniowych można opisać bezpośrednio w leksykonie (zamiast w ogólnych regułach składniowych), co często pozwala na subtelniejszy opis.

Teoria LFG oferuje bogatą reprezentację dzięki zastosowaniu wielu równoległych, powiązanych ze sobą poziomów reprezentacji – dwa podstawowe poziomy to c-struktura i f-struktura. C-struktura to struktura składnikowa (ang. *c-structure* od *constituent structure*) w formie drzewa, powstała na podstawie kategorii składniowych, takich jak fraza nominalna (NP) czy fraza czasownikowa (IP), natomiast f-struktura to struktura funkcyjna (ang. *f-structure* od *functional structure*) w formie struktury atrybutów (ang. *attribute-value matrix*, AVM),

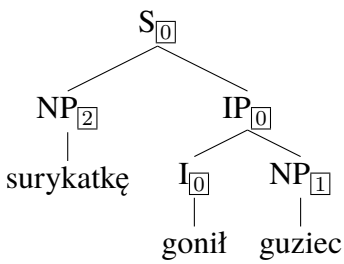
* agnieszka.patejuk@ipipan.waw.pl
** adam.przepiorkowski@ipipan.waw.pl

1 http://zil.ipipan.waw.pl/NEKST.

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

powstała na podstawie takich funkcji gramatycznych jak podmiot (SUBJ) czy dopełnienie bliższe (OBJ), ale zawierająca też inne atrybuty. Przykładowe struktury dla zdania *Surykatkę gonił guziec* podane są poniżej. Powiązania między strukturami oddane są za pomocą zmiennych takich jak 1, która w tym wypadku ustala związek pomiędzy frazą nominalną (NP) *guziec* w c-strukturze w (1) i podmiotem (SUBJ) w f-strukturze w (2).

(1)



(2)

$$\left[\begin{array}{lll} \textsc{pred} & \text{`GONIĆ}\langle\boxed{1},\boxed{2}\rangle\text{'} & \\ \textsc{subj} & \boxed{1} & \left[\begin{array}{ll}\textsc{pred} & \text{`GUZIEC'}\\ \textsc{case} & \textsc{nom}\\ \textsc{gend} & \textsc{m2}\end{array}\right] \\ \textsc{obj} & \boxed{2} & \left[\begin{array}{ll}\textsc{pred} & \text{`SURYKATKA'}\\ \textsc{case} & \textsc{acc}\\ \textsc{gend} & \textsc{f}\end{array}\right] \\ \textsc{tense} & \textsc{past} & \end{array}\right]_{\boxed{0}}$$

Funkcje gramatyczne są podstawowym pojęciem teorii LFG – w przeciwieństwie do teorii chomskiańskich funkcje gramatyczne nie są pochodną pozycji w drzewie, co jest istotne dla języków o swobodnym szyku, takich jak język polski. Funkcje gramatyczne nie są również bezpośrednio zależne od kategorii składniowych: podmiot nie musi być nominalny, ponieważ znane jest zjawisko podmiotów niekanonicznych, np. zdaniowych (*Jana dziwiło, że Maria wybiera Piotra*), które mogą być koordynowane z podmiotami nominalnymi (*Jana dziwiło, że Maria wybiera Piotra, i jej brak gustu*; por. Świdziński 1992b, 1993). Funkcje gramatyczne nie są też zależne od roli semantycznej, co jest widoczne na przykładzie pasywizacji, gdzie dopełnienie strony czynnej (*Surykatkę gonił guziec*) staje się podmiotem strony biernej (*Surykatka była goniona (przez guźca)*) mimo zachowania tej samej roli semantycznej. Funkcje gramatyczne są zatem pojęciem składniowym, które określa relację między nadrzędnikiem i jego podrzędnikiem – mogą być w pewnym stopniu zależne od pozycji w drzewie (szczególnie w językach z ustalonym szykiem), kategorii składniowej czy roli semantycznej, ale nie mogą być do żadnego z tych pojęć sprowadzone.

Nadrzędnik w relacji opisanej daną funkcją gramatyczną reprezentowany jest za pomocą atrybutu PRED. Wartość tego złożonego atrybutu zawiera semantyczną charakterystykę nadrzędnika, na którą składają się: wyrażony tym nadrzędnikiem predykat (reprezentowany jako jego lemat, np. GONIĆ) oraz argumenty tego predykatu, zwykle podane jako zmienne odwołujące się do wartości odpowiednich funkcji gramatycznych (zob. 1 i 2 w (2)). Argumenty predykatu dzielą się na dwa rodzaje: argumenty semantyczne, które są elementami listy zamkniętej w nawiasach kątowych (<...>), i ewentualne argumenty niesemantyczne (zob. §3.2 poniżej), które podane są poza tą listą (zob. 1 w (10) i 2 w (11)). Istnieją też predykaty, których lista argumentów jest pusta: np. rzeczowniki GUZIEC i SURYKATKA w (2) – ich atrybut PRED ma wartość składającą się wyłącznie z lematu.

Poza funkcjami gramatycznymi i atrybutem PRED f-struktura zawiera również inne atrybuty, m.in. morfoskładniowe, takie jak CASE i GEND w (2)–(3), które określają wartość przypadka i rodzaju gramatycznego[3]. Możliwe jest wprowadzenie dowolnych atrybutów, które są potrzebne do analizy, np. atrybutu NEG, który umożliwia opisanie zależności przypadka dopełnienia bliższego od negacji: w zdaniu *Guziec nie gonił surykatki*, któremu odpowiada f-struktura w (3), obecność negacji (NEG) powoduje wystąpienie dopełniacza negacji, natomiast w zdaniu bez negacji, *Guziec gonił surykatkę*, z f-strukturą jak w (2), wymagany jest biernik.

(3)

$$\boxed{0}\begin{bmatrix} \text{PRED} & \text{'GONIĆ}\langle\boxed{1},\boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{'GUZIEC'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M2} \end{bmatrix} \\ \text{OBJ} & \boxed{2}\begin{bmatrix} \text{PRED} & \text{'SURYKATKA'} \\ \text{CASE} & \text{GEN} \\ \text{GEND} & \text{F} \end{bmatrix} \\ \text{NEG} & + \\ \text{TENSE} & \text{PAST} \end{bmatrix}$$

3. Analizy wybranych zjawisk

Niniejszy punkt przedstawia zwięźle analizy niektórych zjawisk objętych gramatyką POLFIE; pełniejsze opisy znaleźć można w literaturze cytowanej w poszczególnych podpunktach.

3.1. Kontrola i podnoszenie

Formalizm LFG umożliwia reprezentację znanych z literatury generatywnej zjawisk kontroli składniowej i podnoszenia[4]. Niektóre predykaty wymagają bezokolicznika (*Antek kazał/obiecał Erykowi zaśpiewać*), którego podmiot nie może być wyrażony leksykalnie, natomiast jego rozumienie jest ściśle określone przez predykat wymagający bezokolicznika – w wypadku czasownika KAZAĆ jest to jego argument celownikowy (*Erykowi*), natomiast w wypadku OBIECAĆ jest to jego podmiot (*Antek*). Teoria LFG obejmuje takie zjawiska opisem, dając możliwość określenia takich zależności między argumentami predykatów (argumentem czasownika głównego a podmiotem bezokolicznika) za pomocą relacji kontroli, która zapisywana jest bezpośrednio w słowniku, we wpisach leksykalnych odpowiednich predykatów: wpis dla KAZAĆ w (4) wskazuje dopełnienie dalsze (OBJ_θ) jako podmiot bezokolicznika

3 Dla oszczędności miejsca pominięte zostały inne atrybuty morfoskładniowe, w tym NUM oznaczający liczbę gramatyczną.

4 Ze stosunkowo niedawnych prac wspomnieć można monografię Landau 2013, a także – w ujęciu minimalistycznym (Chomsky 1995) i z odniesieniem do polszczyzny – prace: Witkoś i in. 2011ab.

(XCOMP SUBJ), natomiast wpis dla OBIECAĆ wskazuje podmiot (SUBJ) jako podmiot bezokolicznika (XCOMP SUBJ).

(4) (↑ PRED) = 'KAZAĆ<(↑ SUBJ)(↑ OBJ_θ)(↑ XCOMP)>'
(↑ OBJ_θ) = (↑ XCOMP SUBJ)

(5) (↑ PRED) = 'OBIECAĆ<(↑ SUBJ)(↑ OBJ_θ)(↑ XCOMP)>'
(↑ SUBJ) = (↑ XCOMP SUBJ)

Różnica w kontroli między predykatami KAZAĆ i OBIECAĆ wynikająca z odpowiednich wpisów w leksykonie jest odzwierciedlona w f-strukturach: (6) odpowiada zdaniu *Antek kazał Erykowi zaśpiewać*, natomiast (7) jest reprezentacją zdania *Antek obiecał Erykowi zaśpiewać*.

(6)
$$\begin{bmatrix} \text{PRED} & \text{'KAZAĆ}\langle \boxed{1},\boxed{2},\boxed{3}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{bmatrix} \\ \text{OBJ}_\theta & \boxed{2}\begin{bmatrix} \text{PRED} & \text{'ERYK'} \\ \text{CASE} & \text{DAT} \\ \text{GEND} & \text{M1} \end{bmatrix} \\ \text{XCOMP} & \boxed{3}\begin{bmatrix} \text{PRED} & \text{'ZAŚPIEWAĆ}\langle \boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{2} \end{bmatrix} \\ \text{TENSE} & \text{PAST} \end{bmatrix}$$

(7)
$$\begin{bmatrix} \text{PRED} & \text{'OBIECAĆ}\langle \boxed{1},\boxed{2},\boxed{3}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{bmatrix} \\ \text{OBJ}_\theta & \boxed{2}\begin{bmatrix} \text{PRED} & \text{'ERYK'} \\ \text{CASE} & \text{DAT} \\ \text{GEND} & \text{M1} \end{bmatrix} \\ \text{XCOMP} & \boxed{3}\begin{bmatrix} \text{PRED} & \text{'ZAŚPIEWAĆ}\langle \boxed{1}\rangle\text{'} \\ \text{SUBJ} & \boxed{1} \end{bmatrix} \\ \text{TENSE} & \text{PAST} \end{bmatrix}$$

Jak wspomniano w §2, w f-strukturach do oznaczenia argumentów danego predykatu użyte są zmienne; w obu omawianych przykładach zmienna [1] odnosi się do podmiotu czasownika osobowego (*Antek*), zaś [2] – do argumentu celownikowego (*Erykowi*) analizowanego tutaj jako dopełnienie dalsze (OBJ_θ). Trzecim argumentem obydwu predykatów jest sygnalizowany zmienną [3] argument bezokolicznikowy, XCOMP. W przeciwieństwie do innych omawianych dotychczas typów argumentów XCOMP należy do argumentów otwartych – jego podmiot musi być kontrolowany przez argument wyższego predykatu i nie może być zrealizowany leksykalnie[5]. W f-strukturze (6) dla KAZAĆ podmiotem predykatu ZAŚPIEWAĆ jest [2], czyli argument celownikowy predykatu KAZAĆ; natomiast w f-strukturze (7) dla OBIECAĆ podmiotem predykatu ZAŚPIEWAĆ jest [1] – podmiot predykatu OBIECAĆ.

Należy podkreślić, że reprezentacja kontroli składniowej ma nie tylko znaczenie semantyczne; następny podpunkt pokazuje, że ma również wpływ na takie zjawiska składniowe jak uzgodnienie.

5 Relacja kontroli zawarta we wpisie leksykalnym danego czasownika (np. (↑ OBJ_θ) = (↑ XCOMP SUBJ) w (4)) określa, który argument czasownika kontroli jest tożsamy z podmiotem kontrolowanego bezokolicznika – z tego powodu nie jest możliwe wypełnienie pozycji podmiotu kontrolowanego bezokolicznika w inny sposób.

3.2. Argumenty predykatywne

Teoria LFG umożliwia również adekwatną reprezentację argumentów predykatywnych w zdaniach takich jak *Marysia wydaje się Antkowi miła* czy *Marysia uważa Antka za miłego* – w obu argumentem predykatywnym jest forma przymiotnika MIŁY (w drugim zdaniu zagnieżdżona we frazie przyimkowej), ale odnosi się ona do różnych argumentów czasowników: podmiotu w pierwszym zdaniu i dopełnienia bliższego w drugim. Reprezentacja tych zależności możliwa jest przy użyciu mechanizmu kontroli opisanego w §3.1 – odpowiednie wpisy w leksykonie wskazują, do czego odnosi się argument predykatywny (XCOMP-PRED): w wypadku WYDAWAĆ SIĘ wpis leksykalny (8) wskazuje, że argument predykatywny odnosi się do podmiotu (SUBJ), natomiast według wpisu (9) dla czasownika UWAŻAĆ argument taki odnosi się do dopełnienia bliższego (OBJ).

(8) (↑ PRED) = 'WYDAWAĆ SIĘ<(↑ OBJ_θ)(↑ XCOMP)>(↑ SUBJ)'
(↑ SUBJ) = (↑ XCOMP-PRED SUBJ)

(9) (↑ PRED) = 'UWAŻAĆ<(↑ SUBJ)(↑ XCOMP)>(↑ OBJ)'
(↑ OBJ) = (↑ XCOMP-PRED SUBJ)

W wyniku powyższych specyfikacji powstają następujące f-struktury: (10) odpowiada zdaniu *Marysia wydaje się Antkowi miła*, natomiast (11) reprezentuje zdanie *Marysia uważa Antka za miłego*. Specyfikacja w (8) sprawia, że w (10) wartość funkcji SUBJ dopełnienia predykatywnego XCOMP-PRED jest tożsama z wartością funkcji SUBJ czasownika WYDAWAĆ SIĘ, co zasygnalizowane jest wielokrotnym wystąpieniem zmiennej 1. Z kolei specyfikacja w (9) sprawia, że w (11) podmiotem dopełnienia predykatywnego jest dopełnienie bliższe czasownika UWAŻAĆ, co zasygnalizowane jest za pomocą zmiennej 2.

(10)
$$\begin{bmatrix}
\text{PRED} & \text{'WYDAWAĆ SIĘ}\langle 2,3\rangle 1\text{'} \\
\text{SUBJ} & 1\begin{bmatrix} \text{PRED} & \text{'MARYSIA'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{F} \end{bmatrix} \\
\text{OBJ}_\theta & 2\begin{bmatrix} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{DAT} \\ \text{GEND} & \text{M1} \end{bmatrix} \\
\text{XCOMP-PRED} & 3\begin{bmatrix} \text{PRED} & \text{'MIŁY}\langle 1\rangle\text{'} \\ \text{SUBJ} & 1 \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{F} \end{bmatrix} \\
\text{TENSE} & \text{PRES}
\end{bmatrix}$$

(11)
$$\begin{bmatrix}
\text{PRED} & \text{'UWAŻAĆ}\langle 1,3\rangle 2\text{'} \\
\text{SUBJ} & 1\begin{bmatrix} \text{PRED} & \text{'MARYSIA'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{F} \end{bmatrix} \\
\text{OBJ} & 2\begin{bmatrix} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{ACC} \\ \text{GEND} & \text{M1} \end{bmatrix} \\
\text{XCOMP-PRED} & 3\begin{bmatrix} \text{PRED} & \text{'MIŁY}\langle 2\rangle\text{'} \\ \text{SUBJ} & 2 \\ \text{CASE} & \text{ACC} \\ \text{GEND} & \text{M1} \\ \text{PFORM} & \text{ZA} \end{bmatrix} \\
\text{TENSE} & \text{PRES}
\end{bmatrix}$$

Analiza, według której elementy predykatywne mają podmiot (SUBJ), wynika m.in. z tego, że w niektórych językach, w tym w języku polskim czy rosyjskim, mogą one wystąpić bez łącznika, np. *Antek miły*. W takim wypadku – ponieważ w teoriach typu LFG z zasady unika się wprowadzania elementów, które nie występują na powierzchni (w rozważanym przykładzie byłby to niewyrażony łącznik) – powstaje f-struktura w (12), w której nadrzędnym predykatem jest przymiotnik predykatywny MIŁY[6].

(12)

$$\begin{bmatrix} \text{PRED} & \text{`MIŁY}\langle \boxed{1} \rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{bmatrix} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{bmatrix}$$

Jawne wskazanie elementu, do którego odnosi się dopełnienie predykatywne, pozwala na zapewnienie odpowiedniego uzgodnienia między tymi elementami – rodzaj i liczba przymiotnika predykatywnego muszą się uzgadniać z odpowiadającymi atrybutami argumentu, do którego ten przymiotnik się odnosi (por. niegramatyczność zdania **Marysia wydaje się Antkowi miły* czy też **Marysia uważa Antka za miłych*). Zapewnienie takiego uzgodnienia wymaga odpowiedniej reprezentacji zależności między tymi argumentami – teoria LFG to umożliwia.

Inaczej jest z przypadkiem gramatycznym przymiotnika predykatywnego: podczas gdy w wypadku WYDAWAĆ SIĘ podlega on uzgodnieniu z podmiotem czasownika, przy UWAŻAĆ jest on narzucany przez przyimek ZA, a w wypadku predykatu UDAWAĆ jest to tzw. przypadek strukturalny, czyli taki, którego wartość morfologiczna zależy nie tylko od nadrzędnego predykatu, ale także od innych cech otoczenia składniowego, w tym od negacji (Patejuk, Przepiórkowski 2014c; *Marysia udaje miłą*, *Marysia nie udaje miłej*).

F-struktura dla zdania *Antek obiecał Marysi próbować być miły* w (13) pokazuje, jak łatwo można poradzić sobie z dalekim uzgodnieniem predykatywnym w formalizmie LFG. W zdaniu tym *Antek* jest podmiotem czasownika OBIECAĆ, którego podrzędnikiem bezokolicznikowym jest forma bezokolicznikowa czasownika PRÓBOWAĆ, który z kolei ma własne dopełnienie bezokolicznikowe w postaci formy czasownika BYĆ wymagającego dopełnienia predykatywnego (forma przymiotnika MIŁY). Zatem element, z którym uzgadnia się dopełnienie predykatywne, *Antek*, i samo dopełnienie predykatywne, *miły*, znajdują się na różnych poziomach struktury składniowej. Opisanie takiego dalekiego uzgodnienia predykatywnego wymaga odpowiedniej reprezentacji składniowej, którą formalizm LFG oferuje – w wyniku relacji kontroli składniowej podmiot OBIECAĆ przypisany do zmiennej $\boxed{1}$ jest współdzielony z podmiotem PRÓBOWAĆ, a następnie z podmiotem BYĆ – ten ostatni predykat współdzieli swój

6 Omówienie innych możliwych analiz takich konstrukcji znaleźć można w artykule Patejuk, Przepiórkowski 2014b.

podmiot z dopełnieniem predykatywnym (XCOMP-PRED). W rezultacie w f-strukturze argumentu predykatywnego informacja o atrybutach podmiotu predykatu OBIECAĆ (przypadek, rodzaj, liczba) dostępna jest lokalnie, dzięki czemu możliwe[7] jest zapewnienie uzgodnienia z elementem predykatywnym. W ten sposób zjawisko pozornie nielokalne sprowadzone jest do sekwencji trzech lokalnych oddziaływań składniowych: dwa razy kontroli i raz uzgodnienia argumentu predykatywnego z podmiotem.

(13)

$$\begin{bmatrix} \text{PRED} & \text{`OBIECAĆ}\langle \boxed{1},\boxed{2},\boxed{3}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{bmatrix} \\ \text{OBJ}_\theta & \boxed{2}\begin{bmatrix} \text{PRED} & \text{`MARYSIA'} \end{bmatrix} \\ \text{XCOMP} & \boxed{3}\begin{bmatrix} \text{PRED} & \text{`PRÓBOWAĆ}\langle \boxed{4}\rangle\boxed{1}\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{XCOMP} & \boxed{4}\begin{bmatrix} \text{PRED} & \text{`BYĆ}\langle \boxed{5}\rangle\boxed{1}\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{XCOMP-PRED} & \boxed{5}\begin{bmatrix} \text{PRED} & \text{`MIŁY}\langle \boxed{1}\rangle\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Szczegółową analizę takich i bardziej skomplikowanych konstrukcji z argumentami predykatywnymi w języku polskim znaleźć można w artykułach: Patejuk, Przepiórkowski 2014b oraz Przepiórkowski, Patejuk 2012b.

3.3. Haplologia SIĘ

Teoria LFG umożliwia również adekwatne opisanie zjawiska haplologii SIĘ w języku polskim. Istnieją predykaty, które wymagają wystąpienia elementu SIĘ, np. *Antek boi *(się)*[8] czy *Antek śmieje *(się)*; w obu wypadkach występuje tzw. SIĘ inherentne (bez kontrybucji semantycznej). Możliwe jednak jest opuszczenie SIĘ, jeżeli występuje ono już przy wyższym strukturalnie czasowniku, np. *Antek bał się śmiać*. Co więcej, w łańcuchu czasowników (ciągu następujących po sobie bezokoliczników) między predykatami wymagającymi SIĘ może wystąpić predykat, z którym SIĘ nie występuje, np. w zdaniu *Antek bał się próbować śmiać*.

7 Podczas gdy w omawianym przykładzie występuje uzgodnienie przypadka, nie jest ono możliwe, kiedy kontrolę sprawuje dopełnienie, jak np. w zdaniu *Antek kazał Marysi próbować być miłą/*miłej* – zamiast uzgodnienia wymagany jest tzw. narzędnik predykacji (opisany w POLFIE), możliwy również w przykładzie z OBIECAĆ (*Antek obiecał Marysi próbować być miłym*).

8 Zapis taki oznacza, że zdanie jest gramatyczne, gdy element w nawiasie występuje, ale niegramatyczne bez tego elementu; oznaczająca niegramatyczność gwiazdka odnosi się więc do opcjonalności sygnalizowanej nawiasami.

F-struktura w (14) odpowiada obu wersjom zdania *Antek bał się śmiać* (*się*) (tj. z obecnym lub opuszczonym drugim *się*), natomiast (15) odpowiada obu wersjom zdania *Antek bał się próbować śmiać* (*się*). Teoria LFG umożliwia odnoszenie się do wyższych i niższych warstw f-struktury: za pomocą ograniczenia w (16) możliwe jest sprawdzenie, czy SIĘ występuje lokalnie (jak w *Antek śmieje się*) lub w wyższym predykacie w łańcuchu czasowników (jak w *Antek boi się* (*próbować*) *śmiać*). Ograniczenie to działa następująco: operator * przy funkcji XCOMP odpowiadającej dopełnieniu bezokolicznikowemu oznacza dowolnie wiele (w tym zero) wystąpień tej funkcji po kolei – zatem (XCOMP*↑) pozwala wejść dowolnie daleko w górę, do wyższej f-struktury w łańcuchu czasowników (z ŚMIAĆ SIĘ do PRÓBOWAĆ i dalej do BAĆ SIĘ). Następnie sprawdzane ($=_c$) jest, czy wartość atrybutu SIE w tym fragmencie f-struktury to +, co oznacza wystąpienie SIĘ.

(14)
$$\begin{bmatrix} \text{PRED} & \text{`BAĆ SIĘ}\langle \boxed{1},\boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{bmatrix} \\ \text{XCOMP} & \boxed{2}\begin{bmatrix} \text{PRED} & \text{`ŚMIAĆ SIĘ}\langle \boxed{1}\rangle\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{SIE} & + \end{bmatrix} \\ \text{SIE} & + \end{bmatrix}$$

(15)
$$\begin{bmatrix} \text{PRED} & \text{`BAĆ SIĘ}\langle \boxed{1},\boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{bmatrix} \\ \text{XCOMP} & \boxed{2}\begin{bmatrix} \text{PRED} & \text{`PRÓBOWAĆ}\langle \boxed{3}\rangle\boxed{1}\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{XCOMP} & \boxed{3}\begin{bmatrix} \text{PRED} & \text{`ŚMIAĆ SIĘ}\langle \boxed{1}\rangle\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{SIE} & + \end{bmatrix} \end{bmatrix} \\ \text{SIE} & + \end{bmatrix}$$

(16) $((\text{XCOMP}^*\uparrow)\ \text{SIE}) =_c +$

Przedstawiona powyżej analiza jest uproszczona, ponieważ nie odróżnia leksykalnego wystąpienia SIĘ (*Antek bał się śmiać się*) od haplologicznego (*Antek bał się śmiać*). Nie uwzględnia

też reprezentacji różnych typów SIĘ oraz wielofunkcyjności SIĘ, jak w jednym z odczytań zdania *Śmiało się z kłopotów*, gdzie SIĘ jest jednocześnie inherentne (wymagane przez czasownik ŚMIAĆ SIĘ) oraz bezosobowe (konstrukcyjne; por. *Dobrze czytało się ten artykuł*). Obydwa aspekty są objęte pełną analizą przedstawioną w artykule Patejuk i Przepiórkowski 2015a.

3.4. Zależności nielokalne

Teoria LFG umożliwia także opis zależności nielokalnych (ang. *unbounded dependencies*) – pod tym pojęciem kryje się m.in. zjawisko zwane czasami ekstrakcją (ang. *wh-extraction*), widoczne w takich zdaniach jak: *Kogo powiedziałeś, że chcesz próbować zaprosić?*. *Kogo* znajduje się tu w linearnym otoczeniu formy predykatu POWIEDZIEĆ, chociaż nie może być analizowane jako jego argument; jest natomiast argumentem predykatu ZAPROSIĆ, który jest dopełnieniem bezokolicznikowym (XCOMP) predykatu PRÓBOWAĆ, a ten z kolei jest dopełnieniem bezokolicznikowym predykatu CHCIEĆ, który w końcu jest dopełnieniem zdaniowym (COMP) predykatu POWIEDZIEĆ. Zatem zaimek *Kogo* znajduje się we frazie – odpowiadającej czasownikowi POWIEDZIEĆ – oddalonej o trzy poziomy zagnieżdżenia (COMP XCOMP XCOMP) od predykatu ZAPROSIĆ, który wymaga tego zaimka jako dopełnienia (OBJ) i warunkuje wartość jego przypadka gramatycznego. W omawianej gramatyce zdaniu takiemu odpowiada f-struktura w (17).

(17)

$$\begin{bmatrix}
\textsc{pred} & \text{`POWIEDZIEĆ}\langle \boxed{1},\boxed{2}\rangle\text{'} \\
\textsc{udf} & \left\{ \boxed{6} \begin{bmatrix} \textsc{pred} & \text{`KTO'} \\ \textsc{case} & \textsc{acc} \end{bmatrix} \right\} \\
\textsc{subj} & \boxed{1}\begin{bmatrix} \textsc{pred} & \text{`PRO'} \end{bmatrix} \\
\textsc{comp} & \boxed{2}\begin{bmatrix}
\textsc{pred} & \text{`CHCIEĆ}\langle \boxed{3},\boxed{4}\rangle\text{'} \\
\textsc{subj} & \boxed{3}\begin{bmatrix} \textsc{pred} & \text{`PRO'} \end{bmatrix} \\
\textsc{xcomp} & \boxed{4}\begin{bmatrix}
\textsc{pred} & \text{`PRÓBOWAĆ}\langle \boxed{5}\rangle\boxed{3}\text{'} \\
\textsc{subj} & \boxed{3} \\
\textsc{xcomp} & \boxed{5}\begin{bmatrix} \textsc{pred} & \text{`ZAPROSIĆ}\langle \boxed{3},\boxed{6}\rangle\text{'} \\ \textsc{subj} & \boxed{3} \\ \textsc{obj} & \boxed{6} \end{bmatrix}
\end{bmatrix} \\
\textsc{comp-form} & \textsc{że}
\end{bmatrix}
\end{bmatrix}$$

Powyższa f-struktura jest wynikiem interakcji odpowiednich wpisów leksykalnych i reguł składniowych. W szczególności reguła składniowa odpowiedzialna za realizację zaimka pytajnego przy czasowniku osobowym przypisuje temu zaimkowi specyfikacje w (18); strzałka w dół, ↓, oznacza tu sam zaimek, sygnalizowany w (17) zmienną 6, zaś strzałka w górę, ↑, oznacza

f-strukturę zdania. Specyfikacja ↓ ∈ (↑ UDF) powoduje zatem dodanie zaimka pytajnego do zbioru UDF, w którym znajdują się elementy podlegające ekstrakcji, zaś specyfikacja ↓ = (↑ COMP XCOMP* OBJ) powoduje, że *Kogo* jest jednocześnie dopełnieniem (OBJ) jednego z ciągu bezokoliczników (XCOMP*) w zagnieżdżonej frazie zdaniowej (COMP) – tutaj ciągu bezokoliczników *próbować zaprosić*. Zaimek pytajny jest zatem argumentem predykatu ZAPROSIĆ, który wpływa na przypadek gramatyczny tego zaimka – wobec braku negacji jest to biernik.

(18) ↓ ∈ (↑ UDF)
↓ = (↑ COMP XCOMP* OBJ)

3.5. Koordynacja

Formalizm LFG pozwala także na objęcie opisem rozmaitych zjawisk składniowych związanych z koordynacją – zostaną tu przedstawione tylko ciekawsze i trudniejsze zjawiska.

Podczas gdy łatwe jest opisanie współdzielenia podrzędnika przy koordynacji, gdy podrzędnik znajduje się na lewym lub prawym krańcu frazy skoordynowanej (*Antek szedł i śpiewał* czy *Kupił i od razu przeczytał dwie książki*), trudniejsze są przypadki, gdy podrzędnik jest „wpleciony", tj. znajduje się wewnątrz jednej ze skoordynowanych fraz (*Szedł Antek i śpiewał*). W ramach teorii LFG możliwa jest analiza, według której współdzielony podrzędnik *Antek*, wpleciony w jeden z członów koordynacji, reprezentowany jest analogicznie do podrzędnika na krańcu frazy skoordynowanej (Patejuk i Przepiórkowski 2015a). Wspólna f-struktura dla obu powyższych zdań – *Antek szedł i śpiewał* oraz *Szedł Antek i śpiewał* – dana jest w (19), gdzie podmiot obu predykatów oznaczony jest zmienną 1. Gramatyka natomiast oczywiście przypisze tym zdaniom różne c-struktury, co oddaje różnicę w szyku.

(19)
$$\left[\begin{array}{ll}\left\{\left(\begin{array}{ll}\left[\begin{array}{ll}\text{PRED} & \text{'IŚĆ}\langle\boxed{1}\rangle\text{'}\\ \text{SUBJ} & \boxed{1}\left[\begin{array}{ll}\text{PRED} & \text{'ANTEK'}\\ \text{CASE} & \text{NOM}\\ \text{GEND} & \text{M1}\\ \text{NUM} & \text{SG}\end{array}\right]\\ \text{TENSE} & \text{PAST}\end{array}\right], \left[\begin{array}{ll}\text{PRED} & \text{'ŚPIEWAĆ}\langle\boxed{1}\rangle\text{'}\\ \text{SUBJ} & \boxed{1}\\ \text{TENSE} & \text{PAST}\end{array}\right]\end{array}\right)\right\}\\ \text{COORD-FORM} \quad \text{I}\end{array}\right]$$

Teoria LFG umożliwia także sformalizowanie zjawiska koordynacji różnych kategorii (Przepiórkowski, Patejuk 2012a; Patejuk, Przepiórkowski 2014a; Patejuk 2015), w którego ramach koordynowane są frazy o odmiennej kategorii składniowej, odpowiadające jednak jednej pozycji składniowej, a zatem mające wspólną funkcję gramatyczną – f-struktura w (20) odpowiada zdaniu *Chcę pić i papierosa* (Kallas 1993: 123), gdzie dopełnienie (OBJ) jest frazą skoordynowaną, która składa się z bezokolicznika (*pić*) i rzeczownika (*papierosa*).

(20)
$$\begin{bmatrix} \text{PRED} & \text{`CHCIEĆ}\langle\boxed{1},\boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{`PRO'}\end{bmatrix} \\ \text{OBJ} & \boxed{2}\left\{\begin{bmatrix}\text{PRED} & \text{`PIĆ}\langle\boxed{1}\rangle\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{CONTROLLER} & \boxed{1}\end{bmatrix}, \begin{bmatrix}\text{PRED} & \text{`PAPIEROS'} \\ \text{CASE} & \text{ACC}\end{bmatrix}\right\} \\ & \begin{bmatrix}\text{COORD-FORM} & \text{I} \\ \text{CONTROLLER} & \boxed{1}\end{bmatrix} \end{bmatrix}$$

Reprezentacja LFG oddaje fakt, że predykat CHCIEĆ jest czasownikiem wymagającym kontroli bezokolicznika przez podmiot (podmiot domyślny CHCIEĆ przypisany do zmiennej [1] jest też podmiotem PIĆ), jednak kontrola dotyczy jedynie bezokolicznikowego członu koordynacji, a nie rzeczownika (*papierosa*). Ponadto wzięte jest pod uwagę nadawanie przypadka strukturalnego: dotyczy ono z kolei jedynie rzeczownikowego członu koordynacji (*papierosa*), natomiast nie ma zastosowania w wypadku członu bezokolicznikowego. Nadawanie przypadka jest lepiej widoczne w przykładzie, w którym nie występuje forma o biernikowo-dopełniaczowym synkretyzmie przypadka: *Chcę pić i kanapkę*, gdzie bez negacji KANAPKA występuje w bierniku, natomiast w obecności negacji, *Nie chcę pić ani kanapki*, występuje w dopełniaczu (przykład analogiczny do *Nie chciał pić ani kanapki*; Kallas 1993: 92).

Ostatnim wspomnianym tutaj interesującym zjawiskiem dotyczącym koordynacji, które zostało uwzględnione w omawianej gramatyce, a może przysparzać problemów innym teoriom, jest koordynacja różnych funkcji gramatycznych (Patejuk, Przepiórkowski 2012; Przepiórkowski, Patejuk 2014; Patejuk 2015), znana również pod nazwą koordynacji leksykalno-semantycznej. Jest ona możliwa pod warunkiem, że poszczególne elementy koordynacji należą do tego samego typu semantycznego (np. zaimki pytajne, negatywne, upowszechniające, typu *-kolwiek* itd.) – jak w zdaniu *Jakie i skąd zdobywał informacje?* (Kallas 1993: 141), gdzie skoordynowane są zaimki pytajne. W tym przykładzie koordynacja łączy formę *Jakie* – przymiotnik będący modyfikatorem rzeczownika *informacje*, który jest dopełnieniem bliższym (OBJ) czasownika – i formę *skąd*, która jest z kolei modyfikatorem samego czasownika. Skoordynowane są tu zatem frazy pochodzące z różnych poziomów f-struktury (modyfikator czasownika i modyfikator dopełnienia tego czasownika).

Teoria LFG umożliwia analizę formalną tego zjawiska i stworzenie f-struktury, która oddaje zarówno fakt koordynacji danych elementów, jak i to, że odpowiadają one innym funkcjom gramatycznym, a ponadto mogą być podrzędnikami innych nadrzędników – reprezentacją omawianego zdania jest f-struktura w (21). Zaproponowana analiza umieszcza skoordynowane elementy w atrybucie UDF i, analogicznie jak przy ekstrakcji omówionej w §3.4, współdzieli je z odpowiednimi elementami f-struktury: zaimek przymiotny *Jakie* jest przypisany do zmiennej [3] i odpowiada modyfikatorowi (ADJ) dopełnienia (OBJ) formy czasownika ZDOBYWAĆ, a więc modyfikatorowi rzeczownika *informacje*, natomiast zaimek przysłówkowy *skąd*

przypisany jest do zmiennej 4 i odpowiada modyfikatorowi (ADJ) czasownika. Taka analiza pozwala na zapewnienie właściwego uzgodnienia między dopełnieniem (*informacje*) a jego modyfikatorem (*Jakie*) mimo tego, że są od siebie znacznie oddalone linearnie. Pełną analizę tego zjawiska przedstawia rozprawa doktorska Patejuk 2015.

(21)

$$\begin{bmatrix}
\text{PRED} & \text{'ZDOBYWAĆ}\langle \boxed{1}, \boxed{2} \rangle\text{'} \\
\text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{'PRO'}\end{bmatrix} \\
\text{OBJ} & \boxed{2}\begin{bmatrix}\text{PRED} & \text{'INFORMACJA'} \\ \text{CASE} & \text{ACC} \\ \text{ADJ} & \left\{\boxed{3}\right\}\end{bmatrix} \\
\text{ADJ} & \left\{\boxed{4}\right\} \\
\text{UDF} & \left\{\left[\left\{\left[\boxed{3}\begin{bmatrix}\text{PRED} & \text{'JAKI'} \\ \text{CASE} & \text{ACC} \\ \text{TYPE} & \text{INT}\end{bmatrix}, \boxed{4}\begin{bmatrix}\text{PRED} & \text{'SKĄD'} \\ \text{TYPE} & \text{INT}\end{bmatrix}\right\}\right]\right\} \\
 & \begin{bmatrix}\text{COORD-FORM} & \text{I}\end{bmatrix}
\end{bmatrix}$$

4. Implementacja i dostępność

LFG jest na tyle sformalizowaną teorią lingwistyczną, że możliwe jest zapisanie gramatyki w postaci umożliwiającej automatyczną analizę składniową za pomocą dedykowanej tej teorii platformy implementacyjnej XLE stworzonej przez Xerox PARC[9]. Istnienie opartego na gramatyce analizatora składniowego, tzw. parsera, ma fundamentalne znaczenie dla rozwoju gramatyki, istotnie ułatwia bowiem weryfikację przyjętych w gramatyce rozwiązań i umożliwia prześledzenie interakcji pomiędzy modułami gramatyki opisującymi różne, ale czasem współwystępujące zjawiska. Prezentowana tu gramatyka POLFIE została w pełni zaimplementowana w środowisku XLE. Co więcej, gramatyka jest rozwijana w ramach międzynarodowego projektu PARGRAM (Parallel Grammar Project; Butt i in. 2002; http://pargram.b.uib.no/), czego efektem jest nacisk na stosowanie w gramatyce mechanizmów i analiz, o których wiadomo, że są wspólne dla większej liczby języków, a nie arbitralnych rozwiązań, działających w wypadku polszczyzny, ale niekoniecznie dających się zastosować do opisu podobnych zjawisk w innych językach. Gramatyka POLFIE jest członkiem projektu PARGRAM od roku 2011 i jako jedyna reprezentuje w projekcie języki słowiańskie; rozwijane są również gramatyki następujących języków: angielskiego, greckiego (nowożytnego), indonezyjskiego, niemieckiego, norweskiego (bokmål i nynorsk), urdu, węgierskiego i wolof.

9 Xerox Linguistic Environment (http://www2.parc.com/isl/groups/nltt/xle/).

Implementacja gramatyki POLFIE maksymalizuje korzystanie z istniejących zasobów językowych dla języka polskiego: gramatyka powstała na bazie dwóch wcześniejszych zaimplementowanych gramatyk języka polskiego, a mianowicie GFJP (Gramatyki Formalnej Języka Polskiego; gramatyki DCG wykorzystywanej przez parser Świgra; Świdziński 1992a; Woliński 2004), która stała się inspiracją dla reguł tworzących drzewa POLFIE, oraz FOJP (Formalnego Opisu Języka Polskiego; gramatyki HPSG towarzyszącej publikacji: Przepiórkowski i in. 2002), która była inspiracją dla struktur funkcyjnych POLFIE. Ponadto POLFIE korzysta z dwóch zasobów do tworzenia leksykonu: z analizatora Morfeusz2 (Woliński 2014) jako źródła informacji morfoskładniowej i ze słownika Walenty (Przepiórkowski i in. 2014, 2017; Hajnicz i in. 2016) jako źródła informacji o wymaganiach walencyjnych predykatów (Patejuk 2016).

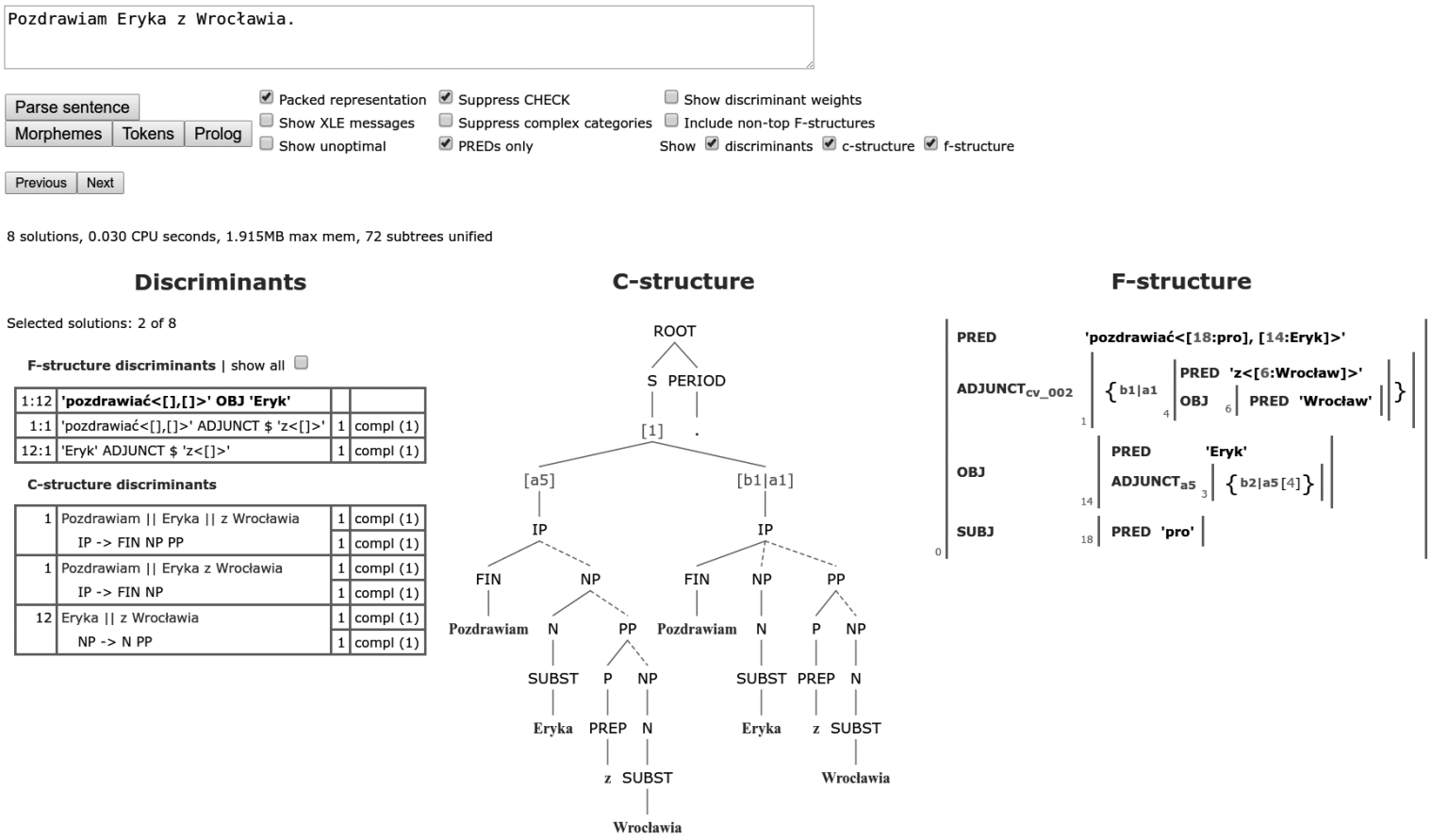
Choć początkowe etapy rozwoju gramatyki POLFIE były zainspirowane wspomnianymi wcześniejszymi gramatykami formalnymi, to obecna postać POLFIE istotnie się od nich różni, zachowując ich cechy pozytywne. Podobnie jak FOJP, ale inaczej niż GFJP, gramatyka POLFIE oparta jest na stabilnej i szeroko znanej teorii lingwistycznej, od dekad rozwijanej przez lingwistów na całym świecie, nieograniczającej się do składni, lecz obejmującej swoim opisem także inne poziomy lingwistyczne, w tym przede wszystkim semantykę. Poszczególne analizy zaimplementowane w gramatyce POLFIE prezentowane są środowisku lingwistycznemu, przede wszystkim – ale nie tylko – od 2012 roku w ramach kolejnych międzynarodowych konferencji LFG. Podobnie natomiast jak GFJP, ale inaczej niż FOJP, gramatyka POLFIE ma szeroki zasięg empiryczny, co jest możliwe między innymi dzięki wykorzystaniu wspomnianych powyżej zewnętrznych zasobów leksykalnych, a także dzięki względnej efektywności platformy implementacyjnej. Analizy gramatyki POLFIE weryfikowane są na ręcznie znakowanym zrównoważonym podzbiorze Narodowego Korpusu Języka Polskiego (NKJP; Przepiórkowski i in. (red.) 2012; http://nkjp.pl/), czyli na tzw. korpusie NKJP1M (Degórski, Przepiórkowski 2012; Szałkiewicz, Przepiórkowski 2012); na podstawie gramatyki i NKJP1M powstał – i nadal jest rozwijany – korpus języka polskiego zawierający ręcznie ujednoznacznione głębokie rozkłady składniowe (Patejuk, Przepiórkowski 2015b), liczący obecnie około 11 000 zdań[10].

Zarówno sama gramatyka, jak i powstały na jej podstawie bank struktur składniowych są publicznie dostępne (na swobodnej licencji GNU GPLv3). Wersje źródłowe gramatyki znaleźć można na poświęconej jej stronie http://zil.ipipan.waw.pl/LFG/, najwygodniej jednak jest testować gramatykę za pomocą interfejsu sieciowego pod adresem http://iness.mozart.ipipan.waw.pl/iness/xle-web, opartego na opracowanej na Uniwersytecie w Bergen platformie INESS służącej do – między innymi – wizualizacji wyników gramatyk XLE i do ręcznego ujednoznaczniania analiz (Rosén i in. 2012; http://iness.uib.no/).

Platforma ta, dzięki niezwykłej łatwości znajdowania poprawnych analiz wśród tych oferowanych przez gramatykę, została także wykorzystana do budowy wspomnianego banku struktur, dostępnego bezpośrednio ze strony http://iness.uib.no/11. W wypadku zdania o więcej niż

10 Dla porównania: oparty na GFJP bank drzew Składnica (http://zil.ipipan.waw.pl/Składnica/) liczy obecnie prawie 10 700 zdań, nie zawiera jednak informacji o funkcjach gramatycznych itp.; istotnie większy jest natomiast wspomniany podkorpus NKJP1M, w którym znakowanych składniowo jest ponad 70 tys. zdań, zawiera on jednak tylko płytkie informacje składniowe, bez pełnych rozbiorów zdań.

11 Po wejściu na stronę należy kliknąć na „Treebank Selection", następnie wybrać „CLARIN-PL".



Rysunek 1. Częściowo ujednoznaczniona analiza zdania *Pozdrawiam Eryka z Wrocławia* (uproszczona: tylko atrybut PRED)

jednym rozkładzie składniowym INESS prezentuje wszystkie c-struktury w postaci jednego grafu (jak na środku rys. 1) i wszystkie f-struktury w podobnie upakowanej postaci (jak w prawej części rys. 1). Jednocześnie jednak system analizuje różnice między rozbiorami i prezentuje je w znacznie czytelniejszej postaci tzw. wyróżników (jak w lewej części rys. 1). W zdaniu z rys. 1, *Pozdrawiam Eryka z Wrocławia*, podstawowa niejednoznaczność polega na tym, że fraza *z Wrocławia* może być modyfikatorem czasownika (*pozdrawiam z Wrocławia*) lub rzeczownika (*Eryk z Wrocławia*). Faktycznie liczba rozbiorów jest większa – jest ich osiem – co wynika z tego, że predykat POZDRAWIAĆ może mieć jeden, dwa lub trzy argumenty, a forma rzeczownika ERYK może być analizowana jako dopełnienie bliższe (OBJ) lub jako biernikowy modyfikator czasownika (przez analogię do formy rzeczownika GODZINA w zdaniu *Godzinę pozdrawiałem* (*kogoś*) *z Wrocławia*). Rys. 1 przedstawia sytuację, gdy zaakceptowany już został (pogrubiony na zrzucie ekranu) wyróżnik 'pozdrawiać<[],[]>' OBJ 'Eryk' mówiący, że predykat POZDRAWIAĆ ma dwa argumenty, a ERYK jest dopełnieniem bliższym tego predykatu. Interpretację, w której fraza *z Wrocławia* modyfikuje czasownik, można następnie wybrać poprzez zaakceptowanie wyróżnika 'pozdrawiać<[],[]>' ADJUNCT $ 'z<[]>' (według którego fraza *z Wrocławia* jest modyfikatorem, ADJUNCT, predykatu dwuargumentowego POZDRAWIAĆ) lub poprzez odrzucenie wyróżnika 'Eryk' ADJUNCT $ 'z<[]>' (który mówi, że fraza *z Wrocławia* jest modyfikatorem predykatu ERYK) – obydwie drogi prowadzą do wyboru tej samej f-struktury, w której fraza *z Wrocławia* jest modyfikatorem czasownika.

5. Zakończenie

Gramatyka POLFIE jest nie tylko największą polską gramatyką formalną opartą na ustabilizowanej teorii lingwistycznej, ale także jedną z największych gramatyk rozwijanych w ramach wspomnianego powyżej projektu PARGRAM. W niniejszym artykule naszkicowaliśmy analizy pewnych zjawisk, których inne gramatyki języka polskiego nie obejmują i które – przynajmniej w niektórych wypadkach – byłyby trudne do lingwistycznie adekwatnego opisania w innych podejściach. Gramatyka i powstały na jej podstawie korpus składniowy są nadal rozwijane – obecnie w ramach realizowanego od lipca 2016 roku dwuletniego projektu związanego z europejską inicjatywą CLARIN ERIC.

Bibliografia

Bresnan J. (red.) 1982: *The mental representation of grammatical relations*, *MIT Press Series on Cognitive Theory and Mental Representation*, The MIT Press, Cambridge, MA.

Bresnan J., Asudeh A., Toivonen I., Wechsler S. 2015: *Lexical-functional syntax*, wyd. 2, Blackwell Textbooks in Linguistics, Wiley-Blackwell.

Butt M., King T.H., Masuichi H., Rohrer C. 2002: *The parallel grammar project*, [w:] N.J. Carroll, R. Sutcliffe (red.), *Proceedings of the workshop on grammar engineering and valuation*, s. 1–7.

Chomsky N. 1995: *The Minimalist Program*, The MIT Press, Cambridge, MA.

Dalrymple M. 2001: *Lexical Functional Grammar*, Academic Press, San Diego.

Degórski Ł., Przepiórkowski A. 2012: *Ręcznie znakowany milionowy podkorpus NKJP*, [w:] A. Przepiórkowski, M. Bańko, R.L. Górski, B. Lewandowska-Tomaszczyk (red.), *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa, s. 51–58.

Hajnicz E., Patejuk A., Przepiórkowski A., Woliński M. 2016: *Walenty: słownik walencyjny języka polskiego z bogatym komponentem frazeologicznym*, [w:] K. Skwarska, E. Kaczmarska (red.), *Výzkum slovesné valence ve slovanských zemích*, Slovanský ústav AV ČR, Praga, s. 71–102.

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.

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

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

Patejuk A., Przepiórkowski A. 2012: *Lexico-semantic coordination in Polish*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'12 Conference*, CSLI Publications, Stanford, CA, s. 461–478.

Patejuk A., Przepiórkowski A. 2014a: *Control into selected conjuncts*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'14 Conference*, CSLI Publications, Stanford, CA, s. 448–460.

Patejuk A., Przepiórkowski A. 2014b: *In favour of the raising analysis of passivisation*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'14 Conference*, CSLI Publications, Stanford, CA, s. 461–481.

Patejuk A., Przepiórkowski A. 2014c: *Structural case assignment to objects in Polish*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'14 Conference*, CSLI Publications, Stanford, CA, s. 429–447.

Patejuk A., Przepiórkowski A. 2015a: *An LFG analysis of the so-called reflexive marker in Polish*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'15 Conference*, CSLI Publications, Stanford, CA, s. 270–288.

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

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

Przepiórkowski A., Hajnicz E., Andrzejczuk A., Patejuk A., Woliński M. 2017: *Walenty: gruntowny składniowo-semantyczny słownik walencyjny języka polskiego)*, „Język Polski" XCVII, z. 1, s. 30–47.

Przepiórkowski A., Kupść A., Marciniak M., Mykowiecka A. 2002: *Formalny opis języka polskiego. Teoria i implementacja*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Przepiórkowski A., Patejuk A. 2012a: *On case assignment and the coordination of unlikes. The limits of distributive features*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'12 Conference*, CSLI Publications, Stanford, CA, s. 479–489.

Przepiórkowski A., Patejuk A. 2012b: *The puzzle of case agreement between numeral phrases and predicative adjectives in Polish*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'12 Conference*, CSLI Publications, Stanford, CA, s. 490–502.

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

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

Rosén V., De Smedt K., Meurer P., Dyvik H. 2012: *An open infrastructure for advanced treebanking*, [w:] J. Hajič, K. De Smedt, M. Tadić, A. Branco (red.), *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, Istanbul, s. 22–29.

Szałkiewicz Ł., Przepiórkowski A. 2012: *Anotacja morfoskładniowa*, [w:] A. Przepiórkowski, M. Bańko, R.L. Górski, B. Lewandowska-Tomaszczyk (red.), *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa, s. 59–96.

Świdziński M. 1992a: *Gramatyka formalna języka polskiego*, Rozprawy Uniwersytetu Warszawskiego, t. 34, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

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

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

Witkoś J., Cegłowski P., Snarska A., Żychliński S. 2011a: *Minimalist facets of control. An English-Polish comparative overview of gerunds and infinitives*, Wydawnictwo Naukowe Uniwersytetu im. A. Mickiewicza, Poznań.

Witkoś J., Cegłowski P., Snarska A., Żychliński S. 2011b: *Wybrane aspekty derywacji i interpretacji wyrażeń bezokolicznikowych. Minimalistyczne studium porównawcze polsko-angielskie*, Wydawnictwo Naukowe Uniwersytetu im. A. Mickiewicza, Poznań.

Woliński M. 2004: *Komputerowa weryfikacja gramatyki Świdzińskiego*, praca doktorska, Instytut Podstaw Informatyki Polskiej Akademii Nauk, Warszawa.

Woliński M. 2014: *Morfeusz reloaded*, [w:] N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, ELRA, Reykjavík, Iceland, s. 1106–1111.

Summary

POLFIE: a contemporary formal grammar of Polish

Keywords: grammar, LFG, Polish, syntax.

This paper presents the new formal grammar of Polish, POLFIE, couched in Lexical Functional Grammar. The paper introduces the basics of LFG, discusses the way of representing syntactic structures in LFG and demonstrates the power of the LFG formalism on the basis of the analysis of selected phenomena of Polish in POLFIE. The paper shows that the LFG theory makes it possible to provide a formalised, adequate and linguistically deep description of syntactic phenomena such as: syntactic control, predicative arguments, the haplology of *się*, unbounded dependencies and coordination; it also takes into account interactions between these. POLFIE is accompanied by a computer implementation in XLE.

MAŁGORZATA MARCINIAK, AGNIESZKA MYKOWIECKA, PIOTR RYCHLIK*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

Automatyczne wydobywanie terminologii dziedzinowej z korpusów tekstowych

Słowa kluczowe: terminologia dziedzinowa, korpusy tekstowe.

1. Wprowadzenie

Wydobywanie terminologii z korpusów dziedzinowych jest zadaniem polegającym na identyfikacji fraz, zwanych terminami, typowych dla dziedziny, której dotyczą zgromadzone teksty. Pojęcie terminologii dziedzinowej jest powszechnie stosowane, opracowywane są słowniki terminów, np. ekonomicznych (Gęsicki, Gęsicki 1996), literackich (Głowiński i in. 2010), sztuk pięknych (Kozankiewicz (red.) 1976). Słowniki te składają się zwykle z haseł będących terminami dziedzinowymi wraz z opisem ich znaczenia. Problemy związane z tworzeniem, opracowaniem, normalizacją terminologii dotyczącej jakiejś dziedziny stanowią przedmiot zainteresowania naukowców wielu specjalności. Temat ten przewija się w pismach bibliotekoznawczych, np. „Przeglądzie Bibliotecznym" (2003, nr 1/2), „Zagadnieniach Informacji Naukowej" oraz „Praktyce i Teorii Informacji Naukowej i Technicznej". Jest on interesujący dla językoznawców i tłumaczy, a także dla naukowców zajmujących się daną dziedziną.

Pomimo wielu publikacji dotyczących terminologii nie ma precyzyjnej definicji mogącej posłużyć do automatyzacji procesu wyboru haseł, które powinny znaleźć się w takich słownikach, a także do oceny uzyskanych wyników. Przyjrzyjmy się, jak hasła *termin* oraz *terminologia* opisane są w słownikach języka polskiego.

W *Słowniku języka polskiego* pod redakcją Witolda Doroszewskiego (SJPDor) termin to 'wyraz albo połączenie wyrazowe o specjalnym, konwencjonalnie ustalonym znaczeniu naukowym lub technicznym', natomiast terminologia to 'ogół terminów, którymi posługuje się dana dziedzina wiedzy, techniki; mianownictwo'. W *Słowniku języka polskiego PWN* (Drabik i in. 2012) te same hasła definiuje się następująco: termin to 'wyraz lub wyrażenie o specjalnym znaczeniu w jakiejś dziedzinie', natomiast terminologię stanowi 'ogół terminów i słownictwa używanego w danej dziedzinie wiedzy, zwłaszcza nauki lub techniki'.

Powyższe definicje słownikowe nie wskazują jednak drogi postępowania w wypadku wykonania zadania w programie komputerowym. Dlatego też na potrzeby automatycznej ekstrakcji terminologii dziedzinowej najpierw zdefiniujemy, co będziemy rozumieli przez pojęcie terminu.

Większość słowników zawierających terminologię dziedzinową składa się z fraz rzeczownikowych. Wprawdzie niektórzy badacze (np. Savova i in. 2003) biorą pod uwagę frazy

* Malgorzata.Marciniak@ipipan.waw.pl, Agnieszka.Mykowiecka@ipipan.waw.pl, Piotr.Rychlik@ipipan.waw.pl

czasownikowe, jednak zwykle terminami są formy nominalne, czyli *recytacja wiersza*, a frazy czasownikowe *recytowanie wiersza* czy *recytować wiersz* nie są uwzględniane. Na potrzeby niniejszego zadania ograniczymy więc pojęcie wyrażenia dziedzinowego do frazy rzeczownikowej. Jednak nie każda napotkana w tekście fraza rzeczownikowa jest terminem dziedzinowym. Przez termin dziedzinowy będziemy rozumieli frazę rzeczownikową, która w tekstach dziedzinowych występuje dostatecznie często, by przypuszczać, że opisuje pojęcie istotne dla dziedziny i której częstość występowania jest jednocześnie niższa w tekstach spoza dziedziny.

Zadanie ekstrakcji terminologii z tekstów dziedzinowych może być wykonywane z myślą o wielu zastosowaniach, począwszy od tworzenia słowników i tezaurusów terminologii dziedzinowej czy tworzenia słowników wielojęzycznych wykorzystywanych do automatycznego tłumaczenia, przez indeksowanie tekstów lub powiązaną z nim ekstrakcję informacji, aż po tworzenie ontologii dziedziny, której dotyczą zgromadzone teksty. W zależności od planowanego zastosowania konstrukcja identyfikowanych fraz może się nieco zmieniać, tak by rezultaty były najlepiej dopasowane do planowanego celu.

2. Automatyczna ekstrakcja terminologii dziedzinowej

W literaturze opisanych jest wiele rozwiązań służących realizacji zadania ekstrakcji terminologii z tekstów dziedzinowych (zob. Pazienza i in. 2004; Korkontzelos i in. 2008). Wszystkie te metody wymagają zgromadzenia korpusu tekstów reprezentatywnych dla danej dziedziny. Zawarte w nim teksty poddawane są wstępnej analizie językowej przypisującej poszczególnym wyrazom nazwy części mowy (POS), a w językach fleksyjnych również ich pełną interpretację morfologiczną. Niezależnie od przyjętych szczegółowych rozwiązań proces ekstrakcji terminologii składa się z kilku etapów omówionych w dalszej części tego rozdziału.

2.1. Identyfikacja fraz

Do identyfikacji fraz będących kandydatami na terminy dziedzinowe korzysta się z wiedzy lingwistycznej dotyczącej ich konstrukcji gramatycznej w rozważanym języku. Takie podejście opisane jest w pracach Johna Justesona i Slavy Katza (1995) oraz Kateriny Frantzi i in. (2000). W tym celu stosuje się zwykle płytkie parsowanie (ang. *shallow parsing*). Istnieją wprawdzie rozwiązania, w których używana jest pełna, głęboka analiza syntaktyczna (ang. *deep parsing*) całych zdań (zob. Savova i in. 2003), ale takie podejście wydaje się nieefektywne. Należy też wspomnieć, że są także rozwiązania, w których całkowicie pomijana jest analiza syntaktyczna. Na przykład w pracy Joachima Wermtera i Udo Hahna (2005) kandydatami na terminy jest n kolejnych słów (n-gramy). Wybór gramatycznie poprawnych fraz dokonywany jest wówczas na późniejszym etapie.

W języku polskim terminami mogą być następujące frazy:
– rzeczowniki, akronimy lub skróty pochodzące od rzeczownika,
– rzeczowniki z sekwencją przymiotników (występujących przed rzeczownikiem lub po nim),
– powyższe sekwencje rzeczownikowe modyfikowane tego samego typu frazą w dopełniaczu,
– frazy rzeczownikowe modyfikowane frazami przyimkowymi,
– koordynacja wymienionych powyżej fraz.

Jednak nie każda fraza zbudowana zgodnie z powyższymi regułami jest terminem dziedzinowym. Juan C. Sager w książce *A practical course in terminology processing* (1990) wskazuje, jakie kryteria powinny spełniać frazy terminologiczne. Jednym z nich jest możliwość ustalenia znaczenia frazy bez rozpatrywania kontekstu, w którym ona wystąpiła. Dlatego też frazy będące terminami dziedzinowymi nie powinny zawierać pewnych słów, np. *inny*, *niektóry*, *jakiś*, *pewien*, *poprzedni*. Fraza *inna faktura materiału* nie tworzy terminu, podczas gdy tak samo zbudowana fraza *bogata faktura materiału* jest dobrą kandydatką na termin. Inną grupą słów, które nie powinny występować w wybranych frazach, są określenia czasu – uniwersalne dla każdej dziedziny. Są to między innymi takie słowa, jak: *dzień*, *miesiąc*, *godzina*, oraz konkretne nazwy – *poniedziałek* lub *grudzień*. Powszechnie stosowanym rozwiązaniem w automatycznej ekstrakcji terminologii jest opracowanie listy słów, które nie powinny występować w terminach.

W języku polskim grupą wyrażeń prowadzących do tworzenia fraz niebędących właściwymi kandydatami na terminy są przyimki złożone. Jeśli rozważymy frazę *na gruncie muzyki współczesnej*, to bez wykluczenia przyimka złożonego *na gruncie* z możliwości tworzenia fraz terminologicznych przy próbie identyfikowania wszystkich fraz rzeczownikowych otrzymamy frazę *grunt muzyki współczesnej*. Proponujemy więc wykluczenie przyimków złożonych (i ich fragmentów) z fraz kandydujących do miana terminu.

2.2. Szeregowanie fraz

Wyselekcjonowane wstępnie frazy rzeczownikowe różnią się między sobą potencjalną przydatnością przy tworzeniu terminologii dziedzinowej. Przykładowo, teksty dotyczące muzyki zawierać mogą frazy *orkiestra symfoniczna*, *drugi utwór* czy *kardiolog*. Ta pierwsza jest niewątpliwie terminem związanym z wykonywaniem muzyki, druga raczej nie zostanie zaklasyfikowana jako termin, natomiast trzecia pochodzi z innej dziedziny i mogła pojawić się na przykład w kontekście opisywania pozamuzycznego wykształcenia muzyka czy kompozytora. Aby w tysiącach fraz rzeczownikowych zidentyfikować te najściślej związane z konkretną dziedziną, potrzebna jest metoda takiego ich szeregowania, by na początku wyznaczonej listy znajdowały się najbardziej charakterystyczne frazy. Porządek ten może wyznaczać miara liczbowa uwzględniająca cechy, które uznamy za ważne dla wyróżnienia terminów dziedzinowych. Najważniejszą informacją, z jakiej możemy skorzystać przy konstrukcji takiej miary, jest liczba wystąpień frazy w tekście. Im częściej fraza występuje, tym większa szansa, że jest ona związana z daną dziedziną. Biorąc pod uwagę jedynie częstość występowania fraz, nie uwzględnilibyśmy jednak co najmniej dwóch ważnych obserwacji. Po pierwsze, frazy wielowyrazowe występują rzadziej niż pojedyncze wyrazy, tak więc kryterium częstości powinno być różne dla fraz o różnej długości. Po drugie, część terminów dziedzinowych pojawia się bardzo często wewnątrz innych fraz, na przykład fraza *orkiestra symfoniczna* pojawia się w takich dłuższych określeniach, jak *nowojorska orkiestra symfoniczna*, *filadelfijska orkiestra symfoniczna*, *narodowa orkiestra symfoniczna*, i wtedy częstość występowania terminu, który stanowi część dłuższej frazy (tutaj: *orkiestra symfoniczna*), byłaby zaniżona. Rozwiązaniem tego problemu jest uwzględnianie wystąpień fraz wewnętrznych. Rodzi to jednak kolejne

problemy, gdyż nie każda fraza wewnętrzna jest terminem dziedzinowym. Jeśli jakaś fraza występuje zawsze w otoczeniu tych samych słów, najbardziej prawdopodobne jest to, że to cała dłuższa sekwencja, a nie jej fragment, tworzy termin dziedzinowy. Natomiast duża zmienność kontekstów często wskazuje na to, że rozważany fragment jest terminem. Przykładowo, wyszukując frazy wewnętrzne w terminie *Uniwersytet Kardynała Stefana Wyszyńskiego*, zidentyfikujemy frazę *Uniwersytet Kardynała Stefana*. Jeśli występuje ona tylko w kontekście podanej frazy dłuższej, to powinna zostać uznana za termin. Z kolei duża liczba kontekstów dla frazy *kościół parafialny* (m.in. *stary*, *dawny*, *rzymskokatolicki*, *zakrystia*, *wieża*) jest argumentem przemawiającym za uznaniem tej frazy za termin dziedzinowy. Propozycją miary istotności frazy kandydującej na termin dziedzinowy, która uwzględnia poruszone wyżej problemy, jest współczynnik *C-value* (Frantzi i in. 2000):

$$C-value(f)=\begin{cases} l(f)^{*}\, freq(f), \text{ jeżeli fraza } f \text{ nigdzie nie występuje jako zagnieżdżona} \\ l(f)^{*}\left(freq(f)-\dfrac{\sum_{b\in T_f} freq(b)}{P(T_f)} \right), \text{ w przeciwnym wypadku} \end{cases}$$

Przez T_f oznaczmy zbiór fraz zawierających badaną frazę *f*. W zacytowanym wzorze liczba wystąpień danej frazy ($freq(f)$) mnożona jest przez wartość $l(f)$ zwiększającą końcową wartość dla fraz dłuższych (typowym rozwiązaniem jest przyjęcie, że $l(f)$ równe jest logarytmowi z długości frazy *f* zwiększonemu o 1 lub jest to logarytm z długości frazy dla fraz wielowyrazowych i wybrana stała dla pojedynczych wyrazów). Składnik odejmowany obniża wagę fraz wewnętrznych o wartość stanowiącą wynik podzielenia liczby wszystkich wystąpień badanej frazy (zarówno tych samodzielnych, jak i wewnętrznych) przez liczbę różnych typów kontekstów ($P(T_f)$), w których te wystąpienia mają miejsce. Wartość ta jest tym większa, im mniej kontekstów różnego rodzaju ma fraza dla tej samej ogólnej ich liczby. Jeśli fraza nigdy nie występuje samodzielnie, a tylko w jednym kontekście innej frazy rzeczownikowej, to końcowa wartość współczynnika równa jest zeru.

Przykładowe wyniki uzyskane dla zbioru około jednego miliona słów tekstów dotyczących muzyki (głównie jazzu), wykazują, że współczynnik *C-value* ma zbliżoną wartość (181,9) dla pojedynczego słowa (*kompozycja*), które wystąpiło ponad 1800 razy w prawie 500 różnych kontekstach, i dla frazy dwuwyrazowej (*muzyka jazzowa*), która wystąpiła około 180 razy w prawie 60 różnych kontekstach (179,4). Natomiast pojedyncze słowo *charakter*, które wystąpiło 185 razy w 96 różnych kontekstach, ma współczynnik *C-value* równy 18,4.

Współczynnik *C-value* wykorzystywany jest do porządkowania fraz zidentyfikowanych w tym samym zbiorze tekstów. Porównywanie współczynników uzyskanych dla różnych zbiorów wymaga ich normalizacji.

2.3. Uproszczona forma podstawowa frazy

Dla języka polskiego (podobnie jak dla innych języków fleksyjnych) zliczanie częstości fraz, zarówno tych, które wystąpiły samodzielnie w tekście, jak i tych zagnieżdżonych w innych

frazach, nie jest czynnością polegającą na prostym porównywaniu napisów. Następującą frazę w narzędniku: *renowacją nawy głównej* należy zidentyfikować jako wystąpienie frazy podstawowej: *renowacja nawy głównej*. W jej wnętrzu wystąpiły cztery zagnieżdżone frazy o następujących formach podstawowych: *renowacja nawy*, *nawa główna*, *renowacja* oraz *nawa*. Żadnej z tych form nie odnajdziemy w rozważanej frazie w narzędniku, a tylko dwie z nich dają się zidentyfikować przez proste porównywanie napisów w formie podstawowej. Jeżeli chcemy więc ustalić częstość wystąpienia frazy o podanej formie podstawowej, to musimy znaleźć w tekście wystąpienia form tej frazy we wszystkich przypadkach. Problem zliczania częstości fraz może być rozwiązany na kilka sposobów. Pierwszą nasuwającą się metodą jest sprowadzenie wszystkich wystąpień fraz rzeczownikowych w tekście do form podstawowych. Zadanie to wymaga jednak jednoznacznej i poprawnej lematyzacji fraz, co dla języka polskiego nie zawsze jest zadaniem prostym do wykonania w programie komputerowym. Rozważmy przykładową frazę w dopełniaczu *ołtarza głównego kościoła parafialnego*, w której wypadku decyzja o tym, czy przymiotnik *główny* opisuje *ołtarz* czy *kościół*, wymaga wiedzy o świecie. Powyższej frazie możemy przypisać następujące lematy: *ołtarz główny kościoła parafialnego* i *ołtarz głównego kościoła parafialnego.* Zdecydowanie o tym, jakiego stopnia przymiotnika należy użyć, lematyzując np. frazę *wyższej szkoły*, wiąże się z trudnościami. Nie wiemy bowiem, czy powinna to być *wyższa szkoła* czy też *wysoka szkoła*.

Innym rozwiązaniem jest zastosowanie uproszczonej formy podstawowej, którą uzyskujemy, lematyzując poszczególne elementy frazy. Dla wyżej rozważanej frazy jest to forma *renowacja nawa główny*, a dla zagnieżdżonych w niej fraz: *renowacja nawa*, *nawa główny*, *renowacja* oraz *nawa*. Ich wystąpienia w oznaczonym morfologicznie tekście, z lematami przypisanymi do poszczególnych słów, dają się zidentyfikować za pomocą prostego porównywania napisów. Rozwiązanie to również nie jest doskonałe, gdyż należy pamiętać, że istnieją frazy, które mają różne właściwe formy podstawowe, a których uproszczone formy podstawowe są takie same. Przewagą tego rozwiązania jest jednak prostota zastosowania go w omawianym zadaniu. Poniżej podajemy, kiedy takie rozwiązanie prowadzi do błędnego połączenia różnych znaczeniowo fraz.

– Frazy, w których element główny jest modyfikowany frazami rzeczownikowymi różniącymi się liczbą, np. *remont nawy bocznej* oraz *remont naw bocznych*. Jeśli zadaniem jest opracowanie słownika terminologicznego, to połączenie tych fraz uproszczoną frazą podstawową jest korzystne. Należy podkreślić, że w pracy nad słownikiem terminologicznym niezbędne jest ustalenie zasad, według których dokonuje się wyboru odpowiedniej formy podstawowej terminu podanego w słowniku. Jednak jeśli terminologia ma posłużyć do ekstrakcji informacji z tekstów, istotne może być rozróżnienie obu fraz i ustalenie, czy remont dotyczył obu naw bocznych czy tylko jednej.

– Frazy zawierające przymiotniki użyte w różnym stopniu, np. *lewa nawa wysoka* oraz *lewa nawa wyższa*. Obie frazy mają taką samą formę podstawową *lewy nawa wysoki*. Pierwsza fraza jest stwierdzeniem faktu, że nawa jest wysoka, natomiast druga fraza mówi o względnej wielkości lewej nawy, zapewne w porównaniu z prawą nawą.

– Frazy zawierające imiesłowy przymiotnikowe w formie zanegowanej i niezanegowanej mają te same uproszczone formy podstawowe, np. *niewydany tomik poezji* oraz *wydany tomik*

poezji mają identyczne formy podstawowe *wydać tomik poezja*, gdyż dla imiesłowów przymiotnikowych informacja o zanegowaniu jest jedną z kategorii gramatycznych.

– Frazy zawierające słowa należące do różnych klas gramatycznych (części mowy), mających jednak takie same uproszczone formy podstawowe. Sytuacja taka zachodzi dla fraz zawierających rzeczowniki odsłowne (gerundia) oraz imiesłowy przymiotnikowe, które mają takie same formy podstawowe będące bezokolicznikami. Istnieją więc frazy, które pomimo różnych konstrukcji syntaktycznych są sprowadzone do takiej samej uproszczonej formy podstawowej. Przykładem takich fraz są *uzgodnienie*$_{ger}$ *terminu renowacji* oraz *uzgodniony*$_{ppas}$ *termin renowacji*, obie mają uproszczoną formę podstawową *uzgodnić termin renowacja*. W pierwszej frazie elementem głównym jest rzeczownik odsłowny *uzgodnienie*, który jest modyfikowany frazą rzeczownikową w dopełniaczu *terminu renowacji*. Elementem głównym drugiej frazy jest natomiast rzeczownik *termin*, który modyfikowany jest imiesłowem przymiotnikowym *uzgodniony*.

Opisane powyżej sytuacje występują w danych sporadycznie. Niektórych z nich można uniknąć, dodając do formy podstawowej informację o klasie gramatycznej słowa lub kategorii gramatycznej prowadzącej do konfliktu.

2.4. Wyodrębnianie zagnieżdżonych fraz

Jedną z istotnych cech stosowanej przez nas metody *C-value* jest zwrócenie uwagi na istnienie fraz terminologicznych, które jako maksymalne frazy występują w tekstach albo bardzo rzadko, albo wcale. Cecha ta jest o tyle ważna, że korpusy dziedzinowe zwykle nie są bardzo duże, wobec tego uwzględnienie w procesie ekstrakcji terminologii jedynie fraz maksymalnych może prowadzić do pominięcia istotnych terminów. W metodzie tej przyjęto więc, by we frazach maksymalnych rozpoznawać wszystkie poprawne gramatycznie frazy wewnętrzne. Podejście takie prowadzi jednak do tworzenia fraz semantycznie niepoprawnych, które powstają w rezultacie obcięcia istotnego elementu frazy. Przykładowo, we frazie *druga wojna światowa* możemy wyróżnić trzy poprawne gramatycznie terminy. Są to: *druga wojna*, *wojna* oraz *wojna światowa*. O ile dwie ostatnie frazy są poprawnymi terminami, o tyle pierwsza z nich nie powinna być rozważana jako termin. W pracy Marciniak, Mykowiecka 2015 zaproponowana została metoda pozwalająca na wyeliminowanie lub znaczne obniżenie rangi takich „urwanych" fraz. Metoda ta dzieli każdą maksymalną frazę na najwyżej dwie frazy zagnieżdżone. Jako miejsce podziału wybiera się najsłabsze ogniwo w danej frazie. Wybór dokonywany jest na podstawie siły powiązań poszczególnych par słów (bigramów) liczonych dla całego korpusu na podstawie lematów słów. Siła wiązania jest ustalana za pomocą współczynnika NPMI (ang. Normalised Pointwise Mutual Information, znormalizowana punktowa informacja wzajemna), zaproponowanego przez Gerlofa Boumę (2009). Dla bigramu „*x y*", współczynnik ten jest liczony według poniższego wzoru, gdzie *x* i *y* są lematami kolejnych słów w tekście; $p(x)$ i $p(y)$ to częstość wystąpienia lematów w rozważanym korpusie, a $p(x,y)$ to częstość bigramu „*x y*".

$$NPMI(x,y) = \ln \frac{p(x,y)}{p(x) * p(y)} / -\ln(p(x,y))$$

Współczynnik NPMI jest tym większy, im częściej słowa *x* i *y* występują obok siebie. Jeśli słowa te zawsze ze sobą sąsiadują, cały współczynnik równy jest 1. Zaproponowana metoda pozwala na wyeliminowanie fraz *druga wojna* oraz *pierwsza wojna* z ogólnodostępnego (http://zil.ipipan.waw.pl/Korpus%20plWikiEcono) korpusu tekstów ekonomicznych, gdyż wartość NPMI dla powyższych bigramów wynosi odpowiednio 0,36 oraz 0,25, podczas gdy dla bigramu *wojna światowa* jest to 0,79.

2.5. Analiza kontrastywna

Opisana powyżej metoda szeregowania fraz pozwala na selekcję z tekstu częstych fraz o strukturze odpowiadającej terminom dziedzinowym. Aby jednak otrzymany zbiór terminów był wystarczająco obszerny, trzeba przetworzyć stosunkowo dużo danych, a im większy jest zbiór, tym większa szansa, że składające się na niego teksty zawierają istotną liczbę wystąpień terminów z różnych innych dziedzin. W analizowanym zbiorze tekstów dotyczących muzyki takimi terminami są przykładowo *silna osobowość*, *stary kontynent*, *wielkie szczęście*. Metodą pozwalającą na odrzucenie terminów należących do innych dziedzin tematycznych jest porównanie list terminów otrzymanych dla różnych kolekcji tekstów. Dobór tych kolekcji ma istotny wpływ na wynik porównania. Jeżeli jako korpus kontrastywny wybierzemy zbiór tekstów ogólnych, mamy szanse na wyeliminowanie terminów typu *własny sposób* czy *łatwe zadanie*. Natomiast jeżeli chcemy wyeliminować terminy specjalistyczne z innej pokrewnej dziedziny, powinniśmy analizować zbiory tekstów z tego właśnie zakresu. Samo porównanie częstości występowania terminu w różnych zbiorach danych nie daje pożądanych rezultatów, gdyż zbiory te mają z reguły różną wielkość. W literaturze opisanych jest kilka strategii porównywania terminów pochodzących z analizy różnych zbiorów tekstów. Każda z tych metod opiera się na innej mierze różnic w rozkładzie terminów w porównywanych tekstach. W przygotowanym programie zaimplementowano trzy z nich. W metodzie pierwszej (Rayson, Garside 2000) wykorzystuje się współczynnik LL (logarytm wiarygodności, ang. Log-Likelihood) mówiący, w jakim stopniu się różni częstość występowania konkretnego terminu w dwóch porównywanych korpusach. Druga metoda (Bonin i in. 2010) łączy częstość występowania w korpusie dziedzinowym z odwrotną częstością występowania w korpusie ogólnym (liczoną jako stosunek wielkości korpusu do częstości badanego terminu).

Trzeci sposób porównywania (CSmw) (Basili i in. 2001), przeznaczony dla terminów wielowyrazowych, uwzględnia nie tylko częstość występowania pełnych terminów, ale też częstość występowania słów stanowiących element główny badanej frazy. W tym współczynniku częstość terminu jest tak wygładzona, by zmniejszyć wariancje wyników dla terminów rzadkich. Niezależnie od wielości podejmowanych prób brakuje jednak uniwersalnej, skutecznej metody rozróżniania terminów występujących rzadko od terminów pochodzących z innej dziedziny.

3. Program TermoPL

TermoPL[1] jest programem do automatycznego wyszukiwania terminów ze zbioru dokumentów dotyczących wybranej dziedziny wiedzy. Przed uruchomieniem programu należy

1 Program TermoPL można pobrać ze strony http://zil.ipipan.waw.pl/TermoPL.

oznakować teksty dokumentów tagerem dla języka polskiego[2], który podzieli wejściowy tekst na tzw. tokeny, czyli słowa, znaki interpunkcyjne oraz inne ciągi znaków, przypisując im ujednoznacznioną interpretację morfologiczną. Wyniki działania tagera powinny być zapisane w pliku tekstowym, w którym zastosowano system kodowania znaków UTF-8. Program akceptuje trzy formaty plików wejściowych: format przyjęty w Narodowym Korpusie Języka Polskiego (NKJP) (Przepiórkowski i in. (red.) 2012), XCES oraz następujący prosty format:

```
forma #lemat #tag#,
```

w którym każdy token opisany jest przez jedną linię tekstu składającą się z jego formy ortograficznej, lematu oraz interptetacji morfologicznej opisanej za pomocą znaczników przyjętych w NKJP. W formacie tym poszczególne zdania oddzielone są specjalną linią tekstu:

```
& #& #interp#.
```

TermoPL przetwarza pliki wejściowe zdanie po zdaniu, wyszukując w nich najdłuższe ciągi tokenów zgodne z zadaną gramatyką. Użytkownik może wybrać standardową gramatykę wbudowaną w program lub zdefiniować swoją własną. Zbiór reguł określających gramatykę standardową jest następujący:

```
NPP : $NAP NAP_GEN*;
NAP[agreement] : AP* N AP*;
NAP_GEN[case = gen] : NAP;
AP : ADJ | PPAS | ADJA DASH ADJ;
N[pos = subst, ger];
ADJ[pos = adj];
ADJA[pos = adja];
PPAS[pos = ppas];
DASH[form = "-"];
```

Symbole `NAP` i `NAP_GEN` oznaczają frazy rzeczownikowe, z tą jednak różnicą, że `NAP_GEN` jest frazą rzeczownikową w dopełniaczu (`[case = gen]`). Zakłada się również, że tokeny dopasowane do wzorca `NAP` (a także do `NAP_GEN`) zgadzają się między sobą co do liczby, rodzaju i przypadka (`[agreement]`). Standardowa gramatyka rozpoznaje frazy składające się z rzeczownika (`N[pos = subst, ger]`), przed którym i po którym może wystąpić dowolna liczba modyfikatorów przymiotnikowych (`AP`), np. *miejska biblioteka publiczna*. Frazy te z kolei mogą być modyfikowane przez dowolną liczbę tego samego typu fraz w dopełniaczu, np. *Sekretariat Naczelnika Wydziału Działalności Gospodarczej Komendy Powiatowej Policji*.

Znalezione w tekście frazy pasujące do tego opisu po przekształceniu na uproszczoną formę podstawową stają się początkową listą kandydatów na terminy. W następnym kroku program

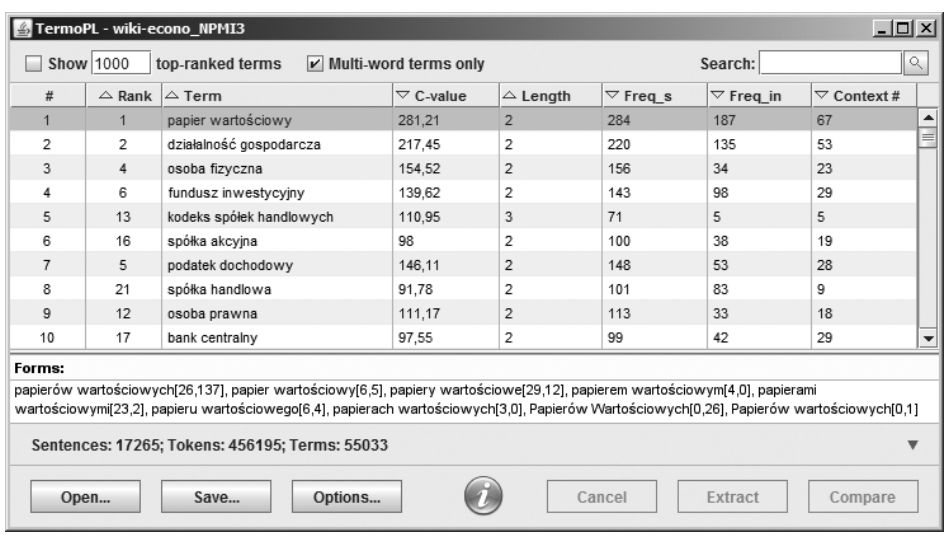
2 Można skorzystać z serwisu demo.clarin-pl.eu/demo/tager.html#.

odnajduje w rozpoznanych wcześniej frazach inne podfrazy zgodne z przyjętą gramatyką, stosując rekurencyjnie jedną z dostępnych metod selekcji fraz zagnieżdżonych, o których mowa była w rozdziale 2.4. Wygenerowane w ten sposób frazy dołączane są do listy kandydatów.

Użytkownik może samodzielnie zdefiniować listy słów i fraz nieuwzględnionych przez program. Mogą to być frazy zawierające określone słowa (np. *ten*, *taki*, *jaki*), przyimki złożone (np. *na modłę*, *pod płaszczykiem*, *u podstaw*) lub będące terminami ogólnymi (np. *aktualny stan*, *druga połowa*, *mała zmiana*).

Zidentyfikowane w tekście terminy mogą zostać przekształcone z formy uproszczonej na formę podstawową. Przykładowo, sekwencja *prawo karny* zamieniona zostaje na prawidłową formę *prawo karne*. Użytkownik ma wpływ na to, które tokeny zostaną finalnie przekształcone na formę podstawową. W standardowej gramatyce są to tylko tokeny dopasowane do symbolu `NAP`, o czym informuje program znak `$` przed tym symbolem. Na ogół forma podstawowa terminu generowana przez program jest frazą w liczbie pojedynczej, chyba że wszystkie formy tego terminu w badanym korpusie są w liczbie mnogiej (np. *drogi oddechowe*).

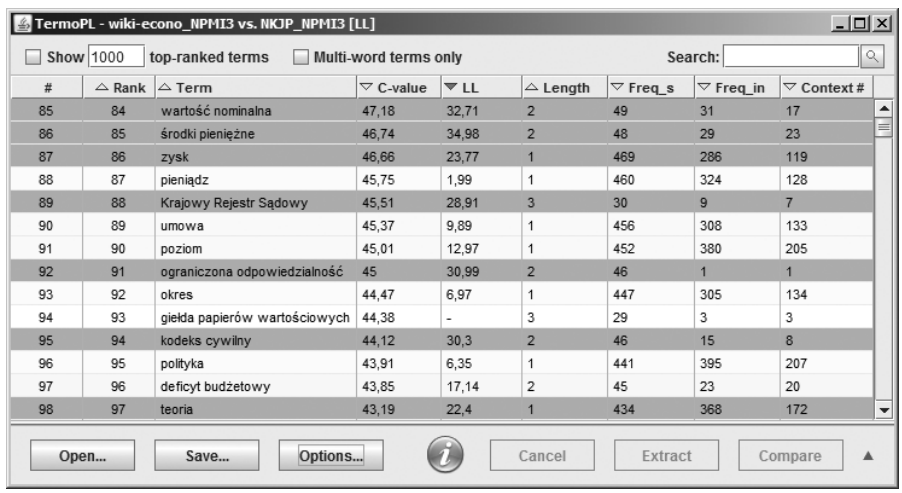
Po skompletowaniu pełnej listy potencjalnych terminów program szereguje je, obliczając dla każdego z nich współczynnik *C-value*. Posortowana lista teminów wyświetlana jest na ekranie monitora w postaci tabeli, która oprócz formy uproszczonej lub podstawowej (`Term`) zawiera: kolejny numer terminu na liście (`#`), jego ocenę (`Rank`), *C-value* (`C-value`), długość (`Length`), całkowitą liczbę wystąpień (`Freq_s`), liczbę wystąpień w kontekście innych fraz (`Freq_in`) oraz liczbę tych kontekstów (`Context #`). Przykładowe rezultaty przedstawione są na rysunku 1. Zawierają one tylko formy wielowyrazowe, gdyż wybrana została opcja *Multi-word terms only*.



Rysunek 1. Fragment listy terminów dla korpusu wiki-ekono

Wyodrębniony zbiór terminów można zapamiętać w pliku tekstowym i użyć go do porównania z innym zbiorem terminów, o czym była mowa w podrozdziale 2.5. W tabeli rezultatów

pojawi się wówczas nowa kolumna zawierająca obliczony współczynnik istotności terminu w korpusie. Poszczególne wiersze tabeli zostaną także oznaczone różnymi odcieniami kolorów w zależności od tego, w jakim stopniu dany termin jest reprezentatywny w badanych korpusach. Im ciemniejszy (bardziej nasycony) jest kolor, tym bardziej dany termin jest reprezentatywny dla jednego z dwóch porównywanych korpusów. Różne odcienie koloru żółtego wskazują na to, że termin jest relatywnie częściej używany w aktualnie badanym korpusie. Na rysunku poniżej kolor ciemniejszy oznacza sytuację odwrotną, tzn. że dany termin jest bardziej charakterystyczny dla korpusu porównawczego. Przykładowe wyniki porównania za pomocą metody LL pokazane są na rysunku 2.



#	Rank	Term	C-value	LL	Length	Freq_s	Freq_in	Context #
85	84	wartość nominalna	47,18	32,71	2	49	31	17
86	85	środki pieniężne	46,74	34,98	2	48	29	23
87	86	zysk	46,66	23,77	1	469	286	119
88	87	pieniądz	45,75	1,99	1	460	324	128
89	88	Krajowy Rejestr Sądowy	45,51	28,91	3	30	9	7
90	89	umowa	45,37	9,89	1	456	308	133
91	90	poziom	45,01	12,97	1	452	380	205
92	91	ograniczona odpowiedzialność	45	30,99	2	46	1	1
93	92	okres	44,47	6,97	1	447	305	134
94	93	giełda papierów wartościowych	44,38	-	3	29	3	3
95	94	kodeks cywilny	44,12	30,3	2	46	15	8
96	95	polityka	43,91	6,35	1	441	395	207
97	96	deficyt budżetowy	43,85	17,14	2	45	23	20
98	97	teoria	43,19	22,4	1	434	368	172

Rysunek 2. Fragment listy terminów z korpusu wiki-ekono porównanego z korpusem 1M NKJP

4. Podsumowanie

Każda dziedzina wiedzy i forma komunikacji mają własne, charakterystyczne dla siebie, słownictwo. W artykule omówiliśmy problemy związane z ekstrakcją terminologii z tekstów dziedzinowych w języku polskim oraz przedstawiliśmy program TermoPL służący do realizacji tego zadania[3]. Wykorzystaliśmy w nim ogólnie uznaną metodę *C-value*, która wymagała dostosowania do fleksyjnego charakteru języka polskiego. Wprowadziliśmy również modyfikację w procesie rozpoznawania fraz zagnieżdżonych, która ogranicza rozpoznawanie fraz „urwanych". Omówiony program powstał w ramach realizacji projektu ClarinPL[4].

3 Program zaprezentowany w niniejszym artykule zawiera implementację metod opracowywanych w ciągu paru ostatnich lat wspólnie przez Małgorzatę Marciniak i Agnieszkę Mykowiecką. M. Marciniak jest główną autorką rozdziałów 1, 2.1, 2.3 i 2.4. A. Mykowiecka jest główną autorką rozdziałów 2.2 i 2.5. Piotr Rychlik zaimplementował prezentowaną wersję programu i jest głównym autorem rozdziału 3. Wszyscy autorzy pracowali wspólnie nad ostateczną redakcją całości tekstu.

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

Bibliografia

Basili R., Moschitti A., Pazienza M.T., Zanzotto F.M. 2001: *A contrastive approach to term extraction*, [w:] *Terminologie et intelligence artificielle. Rencontres*, Inist-CNRS, Vandoeuvre lès Nancy, France, s. 119–128.

Bonin F., Dell'Orletta F., Venturi G., Montemagni S. 2010: *A contrastive approach to multi-word term extraction from domain corpora*, [w:] *Proceedings of the 7th International Conference on Language Resources and Evaluation*, Malta, s. 19–21.

Bouma G. 2009: *Normalized (pointwise) mutual information in collocation extraction*, [w:] *Proceedings of the Biennial GSCL Conference*, Gunter Narr Verlag, Tübingen, s. 31–40.

Drabik L., Kubiak-Sokół A., Sobol E. 2012: *Słownik języka polskiego*, Wydawnictwo Naukowe PWN, Warszawa.

Frantzi K., Ananiadou S., Mima H. 2000: *Automatic recognition of multi-word terms: the C-value/NC-value method*,„International Journal on Digital" 3 (2), s. 115–130.

Gęsicki Ł., Gęsicki M. 1996: *Słownik terminów ekonomiczno-prawnych*, InterFart, Łódź.

Głowiński M., Kostkiewiczowa T., Okopień-Słowińska A. 2010: *Słownik terminów literackich*, Zakład Narodowy im. Ossolińskich, Wrocław.

Justeson J.S., Katz S.M. 1995: *Technical terminology. Some linguistic properties and an algorithm for identification in text*, „Natural Language Engineering" 1, s. 9–27.

Korkontzelos I., Klapaftis I.P., Manandhar S. 2008: *Reviewing and evaluating automatic term recognition techniques*, *Advances in natural language processing*, Lecture Notes in Artificial Intelligence, vol. 5221, s. 248–259.

Kozankiewicz S. (red.) 1976: *Słownik terminologiczny sztuk pięknych*, Państwowe Wydawnictwo Naukowe, Warszawa.

Marciniak M., Mykowiecka A. 2015: *Nested term recognition driven by word connection strength*, „Terminology" 21 (2), s. 180–204.

Pazienza M.T., Pennacchiotti M., Zanzotto F.M. 2004: *Terminology extraction. An analysis of linguistic and statistical approaches*, [w:] *Knowledge Mining, Proceedings of the NEMIS 2004 Final Conference*, Studies in Fuzziness and Soft Computing, vol. 185, Springer, Berlin, s. 255–279.

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

Rayson P., Garside R. 2000: *Comparing corpora using frequency profiling*, [w:] *Proceedings of the Workshop on Comparing Corpora*, vol. 9, Association for Computational Linguistics, Stroudsburg, PA, s. 1–6.

Sager J.C. 1990: *A practical course in terminology processing*, John Benjamins, Amsterdam, Philadelphia.

Savova G.K., Harris M., Johnson T., Pakhomov S.V., Chute C.G. 2003: *A data-driven approach for extracting „the most specific term" for ontology development*, [w:] *Proceedings of AMIA, Annual Symposium*, American Medical Informatics Association, Bethesda, MD, s. 579–583.

SJPDor: *Słownik języka polskiego*, red. W. Doroszewski, t. 1–4, Państwowe Wydawnictwo „Wiedza Powszechna", Warszawa 1958–1962, t. 5–11, Państwowe Wydawnictwo Naukowe, Warszawa 1963–1969.

Wermter J., Hahn U. 2005: *Massive biomedical term discovery*, Lecture Notes in Computer Science, vol. 3735, s. 281–293.

Summary

Automatic extraction of domain terminology from text corpora

Keywords: domain terminology, text corpora.

Every knowledge domain or form of communication has its own characteristic vocabulary. In the traditional approach, dictionaries containing words and multi-word terms identifying important concepts and their lexical equivalents were created by specialists in a subject area. This method, however, is very time-consuming and therefore inadequate, especially for rapidly changing domains. In this paper we present a computer program allowing for automatic identification of noun phrases being potential domain terms and their ranking according to some measure of significance.

WITOLD KIERAŚ, MARCIN WOLIŃSKI*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

Morfeusz 2 – analizator i generator fleksyjny dla języka polskiego[1]

Słowa kluczowe: analiza i synteza morfologiczna, fleksja, przetwarzanie języka naturalnego.

1. Wprowadzenie

Analiza morfologiczna (fleksyjna), której celem jest przypisanie słowom tekstowym wszystkich możliwych interpretacji fleksyjnych, jest podstawowym zadaniem z zakresu automatycznego przetwarzania języka. Analizator nie dopasuje poprawnej interpretacji fleksyjnej do kontekstu składniowego czy semantycznego – tym zajmują się inne wyspecjalizowane programy[2] lub anotatorzy. Podstawowym celem analizatora jest zatem dostarczenie możliwie kompletnej informacji fleksyjnej anotatorom lub programom znakującym automatycznie zbiór tekstów w postaci elektronicznej.

Morfeusz oczywiście nie jest pierwszym programem przeznaczonym do analizy morfologicznej. Inne aplikacje tego typu powstawały co najmniej od połowy lat 90., czyli od ponad dwudziestu lat. Wśród nich warto wymienić m.in. warszawskie analizatory SAM autorstwa Krzysztofa Szafrana (1996) i AMOR Joanny Rabiegi-Wiśniewskiej i Michała Rudolfa (2002), poznańskie LEM (Vetulani i in. 1998) i UAM Text Tools (Obrębski, Stolarski 2005) czy też wciąż rozwijany na Węgrzech przez Roberta Wołosza PoMor (Wołosz 2005). Wszystkie te programy zostały szczegółowo omówione w przeglądowej pracy Elżbiety Hajnicz i Anny Kupść (2001). Do późniejszych aplikacji tego typu zaliczyć należy także analizator Morfologik (Miłkowski 2010), jak również obie wersje Morfeusza (pierwsza z nich została udostępniona w 2001 roku).

Od strony technicznej Morfeusz 2 to moduł programistyczny, który można wbudować w inne programy komputerowe lub wykorzystywać jako samodzielną aplikację. Jest on publicznie dostępny na bardzo liberalnej licencji BSD, umożliwiającej wykorzystanie programu w projektach zarówno naukowych, jak i komercyjnych. Program jest przygotowywany w postaci gotowej do użycia na popularnych platformach sprzętowych. Morfeusz jest zaawansowanym narzędziem informatycznym, wymagającym od użytkownika odpowiednich umiejętności

* wkieras@ipipan.waw.pl, wolinski@ipipan.waw.pl

1 Praca finansowana w ramach wkładu krajowego na rzecz udziału we wspólnym międzynarodowym przedsięwzięciu „CLARIN ERIC: Wspólne zasoby językowe i infrastruktura technologiczna". Współautorstwo obejmuje wszystkie etapy pracy nad artykułem. Głównym twórcą koncepcji Morfeusza jest Marcin Woliński, głównym wykonawcą (programistą) Michał Lenart. Witold Kieraś jest autorem zestawu reguł segmentacyjnych analizatora.

2 Moduł ujednoznaczniający kontekstowo interpretacje fleksyjne nazywa się żargonowo tagerem. Z wykorzystaniem Morfeusza opracowano kilka tagerów języka polskiego, w szczególności TAKIPI zastosowany do znakowania Korpusu IPI PAN oraz tager Pantera zastosowany do znakowania Narodowego Korpusu Języka Polskiego.

i spełniającym ściśle określone wymagania w zakresie wąsko określonej dziedziny, niemniej jednak dla słabiej zorientowanych technicznie użytkowników przygotowano również interfejs graficzny, dzięki któremu można nie tylko korzystać z samego analizatora, ale też tworzyć jego własne zmodyfikowane wersje. Z działaniem Morfeusza 2 można także zapoznać się dzięki internetowej wersji demonstracyjnej pod adresem http://sgjp.pl/morfeusz/demo/.

2. Segmentacja

Morfeusz przetwarza tekst w postaci elektronicznej, będący ciągiem znaków[3], i dokonuje segmentacji, czyli podziału ciągów znaków na takie podciągi (segmenty, słowa tekstowe), którym można przypisać jakąś interpretację fleksyjną, a następnie przypisuje im wszystkie możliwe interpretacje. Interpretacją fleksyjną nazywamy trójkę, w której skład wchodzą: 1) segment, 2) forma hasłowa leksemu (lemat) i 3) znacznik zawierający informację na temat klasyfikacji gramatycznej leksemu oraz wartości kategorii gramatycznych. Klasyfikacja gramatyczna w warstwie językoznawczej opiera się w dużej mierze na pracach Zygmunta Saloniego i Marka Świdzińskiego (Saloni 1974; Saloni, Świdziński 2001), dodatkowo jednak używa również pojęcia fleksemu, stworzonego przez Janusza Bienia (1991) i wykorzystywanego później m.in. przy tworzeniu systemu znaczników fleksyjnych w korpusie IPI PAN (Woliński 2003) oraz – w nieco zmienionej postaci – w Narodowym Korpusie Języka Polskiego (Przepiórkowski i in. (red.) 2012). Nie będziemy tu opisywać szczegółów technicznych i teoretycznych związanych z segmentacją ponad to, co jest niezbędne do zaprezentowania możliwości analizatora Morfeusz 2. Czytelnika zainteresowanego szczegółami odsyłamy do wyżej wymienionych publikacji.

Najprostszy sposób segmentowania wypowiedzenia polega na podzieleniu go na słowa tekstowe od spacji do spacji. Jest to jednak rozwiązanie niesatysfakcjonujące. Co prawda przyjmujemy założenie, że interpretowane segmenty nie przekraczają granicy słowa tekstowego, czyli maksymalny ich zakres wyznaczają otaczające je spacje, z pewnością jednak nie jest to podział odpowiednio drobny. Oczywistym i banalnym przykładem mogą być znaki interpunkcyjne, które – zgodnie z zasadami polskiej interpunkcji – w wielu wypadkach nie są oddzielone od słowa, po którym występują. Nieco poważniejszy problem stanowią formy czasu przeszłego i trybu przypuszczającego czasowników. Rozważmy przykładowe zdania (1) i (2), które zasadniczo wyrażają to samo i są równie poprawne – w pierwszym jednak człon *ście* wchodzi w skład niepodzielnej formy czasownikowej, w drugim zaś – jest zapisany łącznie ze spójnikiem podrzędnym. Niekiedy tego typu przesunięcie jest obowiązkowe i powoduje, że zdanie (4) jest gramatyczne, a zdanie (3) – nie.

(1) Powiedziała, że to czytaliście.
(2) Powiedziała, żeście to czytali.
(3) *Powiedziała, żeby to czytaliście.
(4) Powiedziała, żebyście to czytali.

3 W całym tekście terminu *znak* używamy w znaczeniu technicznym. Nie chodzi więc o znaki językowe, ale takie, które są przedmiotem np. standardu Unicode.

Z tego powodu pierwsza wersja Morfeusza traktowała wszystkie formy typu *czytaliście* jako dwa segmenty, nazywane odpowiednio pseudoimiesłowem przeszłym i aglutynantem. Jako osobne segmenty traktowane były również połączenia spójników podrzędnych i aglutynantów. Analogicznie opisywano także formy trybu przypuszczającego (z uwzględnieniem trzeciego segmentu *by*). Jest to opis językoznawczo spójny i uzasadniony, jednak w niektórych zastosowaniach nadmiernie skomplikowany, dlatego nowy Morfeusz umożliwia użytkownikowi wybranie interpretacji, w której formy czasu przeszłego i trybu przypuszczającego są interpretowane jako pojedyncze segmenty, jeśli tylko wszystkie człony tych złożonych form są zapisane łącznie.

Przesunięcie członu formy czasownikowej, będącego wykładnikiem osoby i liczby, nie ogranicza się jednak tylko do połączeń ze spójnikami podrzędnymi. Poniższe przykłady obrazują podobne tego typu połączenia z formami przymiotników (5), przysłówków (6), tradycyjnych zaimków rzeczownych (7) i rzeczowników (8). Takich połączeń analizator wcześniejszej wersji nie rozpoznawał, a obecnie jest to możliwe dzięki rozbudowanemu systemowi konstruowania reguł segmentacyjnych. Jak wszystko jednak, również analizowanie takich połączeń ma swoją cenę – zwiększa niejednoznaczność analiz, w skrajnych wypadkach jest źródłem systematycznej wieloznaczności, jak w połączeniach form męskich przymiotników w mianowniku (i bierniku – w wypadku podrodzaju nieżywotnego) liczby pojedynczej z aglutynantem *-m*, będącym wykładnikiem pierwszej osoby liczby pojedynczej. Wtedy słowa tekstowe typu *ładnym*, *prostym* miałyby zawsze niejednoznaczne fleksyjnie interpretacje, ponieważ mogą być to zarówno formy narzędnika liczby pojedynczej rodzaju męskiego i nijakiego, jak i formy przymiotnikowe z aglutynantem *-m*, przy czym ta druga interpretacja, choć możliwa, jest nieporównanie mniej prawdopodobna. Z tego powodu zatem możliwość analizowania takich połączeń jak w zdaniach (5)–(8) jest w Morfeuszu 2 fakultatywna i uruchamiana na życzenie użytkownika przez wybór odpowiedniej opcji w ustawieniach.

Co ciekawe, aglutynant może być również współdzielony przez dwie formy czasownikowe, jak w przykładzie (9), a to jest dodatkowym argumentem za uznaniem go za osobny segment.

(5) Możemy odpowiedzieć: a weźcie go sobie sami, skoro s a m i ś c i e go tu nam sprowadzili. (W. Jagielski, *Modlitwa o deszcz*)

(6) S t r a s z n i e ś m y się samym sobie podobali. (S. Aleksijewicz, *Czasy secondhand*[4])

(7) Tam wszystkiego spróbowaliśmy, w s z y s t k o ś m y poznali. (S. Aleksijewicz, *Cynkowi chłopcy*)

(8) No cóż, p r o f e s o r a m i ś m y, ale mamy i przezwiska komilitońskie jeszcze, burszowskie czasy pamiętające... (S. Lem, *Pamiętnik znaleziony w wannie*)

(9) W ciągu całej historii nie ż y l i ś m y, tylko s t a r a l i się przeżyć. (S. Aleksijewicz, *Czasy secondhand*)

3. Lematyzacja, czyli hasłowanie

Analizator segmentuje jednostki tekstowe w taki sposób, by mogły zostać zinterpretowane – każdemu z wyodrębnionych segmentów musi zatem być w stanie przypisać jakąś interpretację

4 Ten i kolejne cytaty z książek Swietłany Aleksijewicz w tłumaczeniu Jerzego Czecha.

fleksyjną. Interpretowanie polega na przypisaniu odpowiedniego lematu oraz znacznika fleksyjnego. Samo przypisywanie formy hasłowej nazywane jest lematyzacją lub hasłowaniem.

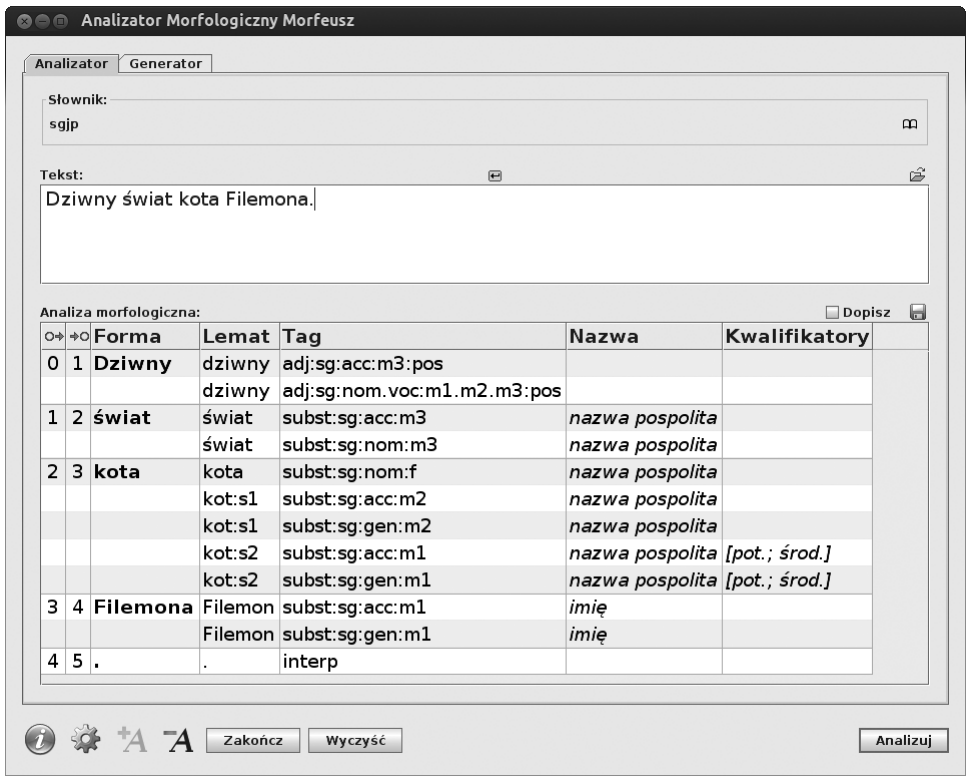
Lemat jest identyfikatorem leksemu, swego rodzaju konwencjonalną i arbitralnie dobraną etykietą, która jednoznacznie wyróżnia formy fleksyjne należące do paradygmatu odmiany tego samego leksemu. Tradycyjnie lemat jest wykładnikiem konwencjonalnie ustalonej formy leksemu, zwanej formą hasłową, np. dla czasowników jest to bezokolicznik. W wypadku jednak, gdy istnieje wiele leksemów o tej samej formie hasłowej, potrzebna jest dodatkowa informacja identyfikująca leksem. Taka sytuacja zachodzi np. dla dwóch leksemów o formie hasłowej *piec*: PIEC[1] (czasownik) i PIEC[2] (rzeczownik). Lematy tych dwu leksemów zostają w analizach Morfeusza rozszerzone dodatkowo o poprzedzoną dwukropkiem etykietę pozwalającą rozróżnić te dwie jednostki: *v* dla czasownika i *s* dla rzeczownika. Jeśli homonimiczne jednostki reprezentują tę samą kategorię gramatyczną (czy ściślej: są opisane przez ten sam fleksem), przy etykiecie pojawi się również indeks, np. rzeczowniki ADMIRAŁ[1] 'stopień oficerski w marynarce wojennej' i ADMIRAŁ[2] 'gatunek motyla', różniące się podrodzajem męskim, otrzymają odpowiednio formy hasłowe: *admirał:s1* i *admirał:s2*.

4. Przykład analizy

Rysunek 1 przedstawia przykładową analizę ciągu *Dziwny świat kota Filemona.* W kolumnie *Forma* widać kolejne wydzielone w wypowiedzeniu segmenty – w tym wypadku segmentacja jest prosta, jedynym segmentem wydzielonym nie na podstawie spacji między słowami jest końcowa kropka. W kolejnych wierszach znajdą się interpretacje fleksyjne dla poszczególnych słów tekstowych. Słowo *Dziwny* ma siedem interpretacji – we wszystkich wypadkach są to formy tego samego przymiotnika DZIWNY – które dzięki pewnej konwencji notacyjnej zostały skrócone do dwóch wierszy. Pierwszy wiersz mówi, że słowo może być formą biernika liczby pojedynczej rodzaju m3 (tzw. męskiego nieżywotnego), drugi zaś – mianownika lub wołacza dowolnego z trzech rodzajów męskich. Kropka w zapisie typu *m1.m2.m3* jest alternatywą i pozwala skrócić trzy oddzielne interpretacje fleksyjne do jednego wiersza wynikowego.

Kolejny segment, czyli *świat*, ma dwie interpretacje rzeczownikowe: biernikową i mianownikową. Trzecie słowo (*kota*) może być formą jednego z trzech leksemów rzeczownikowych: mianownikiem żeńskiego rzeczownika KOTA, biernikiem lub dopełniaczem męskiego żywotnego rzeczownika KOT[1] 'zwierzę' albo biernikiem lub dopełniaczem męskiego osobowego rzeczownika KOT[2] 'początkujący żołnierz, uczeń 1. klasy, student 1. roku'. W kolumnie *Kwalifikatory* widać, że interpretacje związane z rzeczownikiem KOT[2] są stylistycznie nacechowane, zostały oznaczone kwalifikatorami *pot.* (potoczne) i *środ.* (środowiskowe).

Słowo *Filemona* zostało rozpoznane jako biernik lub dopełniacz od imienia FILEMON. Ostatnim segmentem w analizie jest kropka, opisana po prostu jako znak interpunkcyjny. Ze względu na swoją wielofunkcyjność znaki interpunkcyjne w opisie fleksyjnym stosowanym przez Morfeusza na zasadzie konwencji są traktowane jako osobne segmenty (podobnie jak np. liczby zapisane cyframi rzymskimi lub arabskimi). Nie postulujemy oczywiście istnienia leksemów reprezentowanych przez znaki interpunkcyjne.



		Forma	Lemat	Tag	Nazwa	Kwalifikatory
0	1	**Dziwny**	dziwny	adj:sg:acc:m3:pos		
			dziwny	adj:sg:nom.voc:m1.m2.m3:pos		
1	2	**świat**	świat	subst:sg:acc:m3	*nazwa pospolita*	
			świat	subst:sg:nom:m3	*nazwa pospolita*	
2	3	**kota**	kota	subst:sg:nom:f	*nazwa pospolita*	
			kot:s1	subst:sg:acc:m2	*nazwa pospolita*	
			kot:s1	subst:sg:gen:m2	*nazwa pospolita*	
			kot:s2	subst:sg:acc:m1	*nazwa pospolita*	*[pot.; środ.]*
			kot:s2	subst:sg:gen:m1	*nazwa pospolita*	*[pot.; środ.]*
3	4	**Filemona**	Filemon	subst:sg:acc:m1	*imię*	
			Filemon	subst:sg:gen:m1	*imię*	
4	5	**.**	.	interp		

Rysunek 1. Przykładowa analiza fleksyjna wypowiedzenia *Dziwny świat kota Filemona*

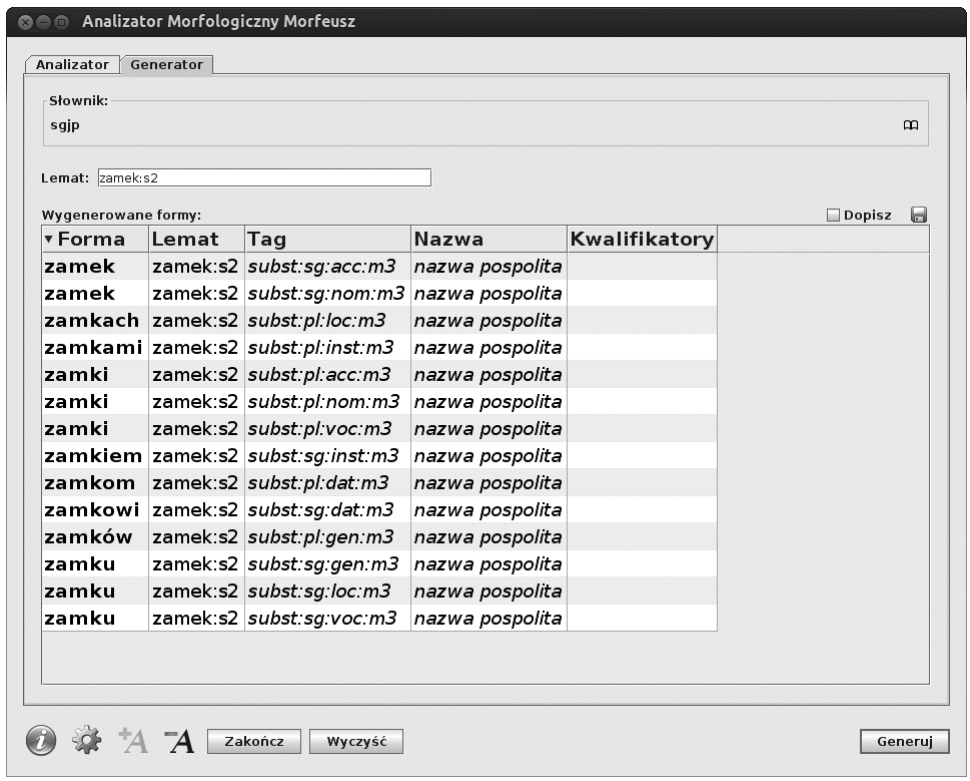
Jak widać, w kolumnie *Nazwa* wszystkie interpretacje rzeczownikowe mają przypisaną albo wartość *nazwa pospolita*, albo typ nazwy własnej, np. *imię*.

5. Generator

Zupełnie nową funkcją Morfeusza 2 jest możliwość generowania wszystkich form fleksyjnych leksemu. Do syntezy pełnych paradygmatów program używa tych samych danych słownikowych, które wykorzystuje do analizy.

W celu wygenerowania wszystkich form danego leksemu należy programowi podać jego formę hasłową. W wypadku form niejednoznacznych Morfeusz zwróci formy wszystkich leksemów o podanej formie hasłowej. Na przykład podanie programowi formy *zamek* spowoduje wygenerowanie dwu osobnych paradygmatów dla dwu leksemów: ZAMEK[1] 'urządzenie' i ZAMEK[2] 'budowla'. Oba paradygmaty różnią się jedynie formą dopełniacza liczby pojedynczej, ale ponieważ w parze z różnicą w odmianie idzie również różnica znaczenia, stanowią one osobne hasła SGJP i – w konsekwencji – również w Morfeuszu. Ze względu jednak na homonimię form hasłowych będą one opatrzone identyfikatorami rozróżniającymi te leksemy. Jeśli użytkownik ma świadomość istnienia w słowniku więcej niż jednego leksemu o określonej formie hasłowej i chce w generatorze uzyskać formy tylko jednego z nich, może

programowi podać formę hasłową wraz z identyfikatorem, np. dla lematu *zamek:s2* generator zwróci wszystkie formy leksemu ZAMEK[2]. Przykład syntezy form tego właśnie rzeczownika zaprezentowano na rysunku 2.



Forma	Lemat	Tag	Nazwa	Kwalifikatory
zamek	zamek:s2	*subst:sg:acc:m3*	*nazwa pospolita*	
zamek	zamek:s2	*subst:sg:nom:m3*	*nazwa pospolita*	
zamkach	zamek:s2	*subst:pl:loc:m3*	*nazwa pospolita*	
zamkami	zamek:s2	*subst:pl:inst:m3*	*nazwa pospolita*	
zamki	zamek:s2	*subst:pl:acc:m3*	*nazwa pospolita*	
zamki	zamek:s2	*subst:pl:nom:m3*	*nazwa pospolita*	
zamki	zamek:s2	*subst:pl:voc:m3*	*nazwa pospolita*	
zamkiem	zamek:s2	*subst:sg:inst:m3*	*nazwa pospolita*	
zamkom	zamek:s2	*subst:pl:dat:m3*	*nazwa pospolita*	
zamkowi	zamek:s2	*subst:sg:dat:m3*	*nazwa pospolita*	
zamków	zamek:s2	*subst:pl:gen:m3*	*nazwa pospolita*	
zamku	zamek:s2	*subst:sg:gen:m3*	*nazwa pospolita*	
zamku	zamek:s2	*subst:sg:loc:m3*	*nazwa pospolita*	
zamku	zamek:s2	*subst:sg:voc:m3*	*nazwa pospolita*	

Rysunek 2. Przykładowa synteza wszystkich form jednego z dwu homonimicznych rzeczowników ZAMEK

6. Słowniki i segmentacja

Morfeusz dostępny jest z dwoma słownikami: aktualnymi danymi internetowej wersji *Słownika gramatycznego języka polskiego* (Saloni i in. 2015, por. artykuł w tym samym numerze) oraz słownikiem *Polimorf* (Woliński i in. 2012). Kolejne wydania Morfeusza są generowane automatycznie przez system Kuźnia (Szejko 2014) zarządzający pracą nad oboma słownikami. W przyszłości słowników może być więcej, ponieważ Morfeusz zaczyna być wykorzystywany również do analizy tekstów dawniejszych i może być rozszerzany o dane tworzone w innych projektach słownikowych (np. słownik fleksyjny polszczyzny XIX wieku po 1830 roku (Derwojedowa i in. 2014) czy słownik XVII i XVIII w. (Bronikowska i in. 2016)).

Możliwe jest również tworzenie różnych innych wyspecjalizowanych słowników fleksyjnych o konkretnym przeznaczeniu, które w specyficznych zastosowaniach albo całkowicie zastąpią standardowe słowniki, albo będą stanowić jedynie ich rozszerzenie. Przykładowo, można

sobie wyobrazić słownik fleksyjny obejmujący słownictwo typowe dla żargonu polskich internautów, stworzony z myślą o analizie korpusowej forów i czatów internetowych. Stanowiłby on rozszerzenie i uzupełnienie jednego z dwu standardowych słowników.

W wypadku Morfeusza słownik to po prostu duży plik tekstowy zawierający pięć kolumn. W każdym wierszu zapisana jest informacja o jednej formie fleksyjnej – w kolumnach znajdują się kolejno: forma tekstowa, lemat, znacznik fleksyjny oraz informacja o tym, czy jest to forma reprezentująca nazwę własną, i informacja o stylistycznych i zakresowych ograniczeniach użycia zapisana w postaci kwalifikatorów. Dwie ostatnie kolumny mogą być również puste.

Oprócz słownika do poprawnego działania Morfeusz potrzebuje jeszcze definicji tagsetu i reguł segmentacji. Tagset to zbiór wszystkich możliwych znaczników fleksyjnych, które analizator może przypisywać analizowanym segmentom. Z kolei w zbiorze reguł segmentacji – w pewnym uproszczeniu – definiuje się sposoby, w jakie ciąg znaków wejściowych może być podzielony na podciągi, by w efekcie powstała poprawna analiza. Wszystkie te trzy elementy, czyli słownik, tagset i reguły segmentacyjne, w nowej wersji Morfeusza można modyfikować i dostosowywać do potrzeb użytkownika. Najbardziej znacząca jest jednak możliwość tworzenia reguł segmentacji, otwierająca m.in. pole do wykorzystania pewnego rodzaju informacji słownikowej, która dotąd nie była wykorzystywana w analizatorze. *Słownik gramatyczny języka polskiego*, czyli podstawowe źródło danych lingwistycznych dla Morfeusza, notuje obok regularnych form deklinacyjnych przymiotników i liczebników również tzw. formy złożeniowe, czyli niesamodzielne jednostki występujące jedynie w określonych konstrukcjach słowotwórczych. Na przykład formy złożeniowe *popularno* i *pięcio*, będące formami odpowiednio przymiotnika POPULARNY i liczebnika PIĘĆ, występują w takich wyrazach jak POPULARNONAUKOWY, POPULARNOHISTORYCZNY, POPULARNOROZRYWKOWY; PIĘCIOCENTYMETROWY, PIĘCIOPIĘTROWY, PIĘCIORUBLÓWKA. Nie wszystkie z tych przykładowych wyrazów muszą się znajdować w słowniku analizatora, w szczególności rzadkie, choć poprawne, połączenia tego typu najpewniej w słowniku się nie znajdą, nie sposób bowiem wszystkich ich przewidzieć. Reguły segmentacyjne pozwalają jednak przekazać mechanizmowi analizatora informację, że jeśli dany ciąg znaków da się rozłożyć na dwa, z których pierwszy będzie formą złożeniową liczebnika lub przymiotnika, drugi zaś formą przymiotnika (czy w wypadku leksemu PIĘCIORUBLÓWKA – rzeczownika), to całość można zanalizować jako formę wyrazową o interpretacji fleksyjnej identycznej z interpretacją drugiego członu złożenia i formą hasłową, która składa się z formy hasłowej drugiego członu poprzedzonej odpowiednią formą złożeniową. Istnieje także możliwość takiego przekształcenia reguł segmentacyjnych, by powyższe przykładowe wyrazy były analizowane nie jako jeden segment, ale jako dwa oddzielne. Z możliwości tej korzysta się przy złożonych formach przymiotnikowych zapisywanych z dywizem typu *biało-czerwony*.

Tego samego mechanizmu używa się w Morfeuszu również w wypadku licznych przedrostków, takich jak *pseudo-*, *anty-*, *euro-*, *ekstra-*, *eks-*, które wchodzą w bardzo produktywne połączenia z rzeczownikami i przymiotnikami, tworząc wyrazy o łatwym do przewidzenia znaczeniu. Dzięki temu analizator jest w stanie poprawnie zinterpretować również rzadkie i nietypowe połączenia utworzone za pomocą przedrostków tego typu. W efekcie zatem Morfeusz

jest w stanie efektywnie analizować pewne określone produktywne konstrukcje słowotwórcze, choć oczywiście tylko nieliczne.

Podobnie jak w wypadku słowników, łatwo też można sobie wyobrazić różne zestawy reguł segmentacyjnych stosowane do różnych celów. Segmentacja we współczesnych tekstach literackich czy prasowych nie musi się pokrywać z segmentacją używaną do analizy tekstów dawniejszych, choćby ze względu na inne zasady pisowni łącznej i rozdzielnej. Na przykład dla słowników polszczyzny dawniejszej można sformułować regułę pozwalającą na analizowanie partykuły NIE zapisanej łącznie z formą czasownikową.

7. Podsumowanie

W artykule staraliśmy się w sposób przystępny przedstawić cele i zastosowania nowej wersji analizatora morfologicznego Morfeusz. Wszystkie opisane wersje analizatora dostępne są do pobrania na stronie http://sgjp.pl/morfeusz/. Nowe wersje są generowane raz w tygodniu, o ile w słowniku dokonano jakichś zmian. Dzięki temu dane Morfeusza są w zasadzie na bieżąco aktualizowane z internetową wersją SGJP.

Bibliografia

Bień J.S. 1991: *Koncepcja słownikowej informacji morfologicznej i jej komputerowej weryfikacji*, Rozprawy Uniwersytetu Warszawskiego, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Bień J.S., Saloni Z. 1982: *Pojęcie wyrazu morfologicznego i jego zastosowanie do opisu fleksji polskiej (wersja wstępna)*, „Prace Filologiczne" XXXI, s. 31–45.

Bronikowska R., Gruszczyński W., Ogrodniczuk M., Woliński M. 2016: *The use of electronic historical dictionary data in corpus design*, „Studies in Polish Linguistics", 11(2), s. 47–56, doi:10.4467/23005920SPL.16.003.4818.

Derwojedowa M., Kieraś W., Skowrońska D., Wołosz R. 2014: *Współczesne narzędzia leksykograficzne a analiza tekstów dawniejszych*, „Polonica" XXXIV, s. 21–27.

Hajnicz E., Kupść A. 2001: *Przegląd analizatorów morfologicznych dla języka polskiego*, IPI PAN Research Report 937, Institute of Computer Science, Polish Academy of Sciences, Warsaw.

Miłkowski M. 2010: *Developing an open-source, rule-based proofreading tool*, „Software: Practice and Experience", 40 (7), s. 543–66.

Obrębski T., Stolarski M. 2005: *UAM Text Tools – A text processing toolkit for Polish*, [w:] Z. Vetulani (red.), *Proceedings of 2nd Language & Technology Conference*, Wydawnictwo Poznańskie, Fundacja Uniwersytetu im. A. Mickiewicza, Poznań, s. 301–304.

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

Rabiega-Wiśniewska J., Rudolf M. 2002: *Amor – program automatycznej analizy fleksyjnej tekstu polskiego*, „Biuletyn Polskiego Towarzystwa Językoznawczego" LVIII, s. 175–186.

Saloni Z. 1974: *Klasyfikacja gramatyczna leksemów polskich*, „Język Polski" LIV, s. 3–31, 93–101.

Saloni Z., Gruszczyński W., Woliński M., Wołosz R., Skowrońska D. 2015: *Słownik gramatyczny języka polskiego*, wyd. 3 (online: http://sgjp.pl).

Saloni Z., Świdziński M. 2001: *Składnia współczesnego języka polskiego*, Państwowe Wydawnictwo Naukowe, Warszawa.

Szafran K. 1996: *Analizator morfologiczny SAM-95, opis użytkowy*, Rap. tech. TR 96-05 (226), Instytut Informatyki Uniwersytetu Warszawskiego, Warszawa.

Szejko J. 2014: *Aplikacja webowa do opracowywania słowników fleksyjnych*, praca magisterska, Uniwersytet Warszawski, Warszawa.

Vetulani Z., Martinek J., Obrębski T., Vetulani G. 1998: *Dictionary based methods and tools for language engineering*, Uniwersytet im. Adama Mickiewicza, Poznań.

Woliński M. 2003: *System znaczników morfosyntaktycznych w korpusie IPI PAN*, „Polonica" XXII–XXIII, s. 39–55.
Woliński M. 2006: *Morfeusz – a practical tool for the morphological analysis of Polish*, [w:] M.A. Kłopotek, S.T. Wierzchoń, K. Trojanowski (red.), *Intelligent Information Processing and Web Mining*, Advances in Soft Computing, Springer, Berlin, s. 503–512.
Woliński M. 2014: *Morfeusz reloaded*, [w:] N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, ELRA, Reykjavík, Iceland, s. 1106–1111.
Woliński M., Miłkowski M., Ogrodniczuk M., Przepiórkowski A., Szałkiewicz Ł. 2012: *PoliMorf: a (not so) new open morphological dictionary for Polish*, [w:] N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012*, ELRA, Istanbul, Turkey, s. 860–864.
Wołosz R. 2005: *Efektywna metoda analizy i syntezy morfologicznej w języku polskim*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Summary

Morfeusz 2 – an inflectional analyser and generator for Polish

Keywords: morphological analysis and synthesis, inflection, natural language processing.

Morfeusz is a well known application, used in Polish computational linguistics for over a decade. In the paper we present a new version of the program, focusing on new features which have been introduced since the previous version.

WITOLD KIERAŚ, MARCIN WOLIŃSKI*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

Słownik gramatyczny języka polskiego – wersja internetowa[1]

Słowa kluczowe: fleksja, paradygmatyka, słownik.

W artykule przedstawimy nowe wydanie *Słownika gramatycznego języka polskiego* (SGJP). Inaczej niż jego dwa poprzednie wydania (Saloni i in. 2007b, 2012), które były udostępniane w postaci programu instalowanego w komputerze (zob. Saloni i in. 2007a), teraz słownik jest w całości dostępny w Internecie, a jego funkcje użytkowe zostały rozszerzone. Skupimy się przede wszystkim na widocznych dla użytkownika zmianach w stosunku do poprzednich wersji.

Koncepcja stworzenia słownika gramatycznego języka polskiego powstała co najmniej 35 lat temu, a bezpośrednią inspiracją dla takiego pomysłu był niewątpliwie *Słownik gramatyczny języka rosyjskiego* A. Zalizniaka (1977/2003). Prace przygotowawcze mające doprowadzić do jego powstania obejmowały stopniowo formalny opis deklinacji polskich rzeczowników pospolitych (Gruszczyński 1989), koniugacji (Saloni 2001), aż wreszcie zaowocowały pierwszym wydaniem SGJP w 2007 roku. W przeciwieństwie do wcześniejszych prac SGJP od początku funkcjonował jako słownik elektroniczny, a nie papierowy. Przeniesienie go do Internetu rozszerza grono jego potencjalnych użytkowników i stwarza możliwość szybkiej i łatwej aktualizacji jego danych.

Autorami najnowszego wydania słownika są Zygmunt Saloni, Włodzimierz Gruszczyński, Robert Wołosz, Marcin Woliński i Danuta Skowrońska. Głównym wykonawcą aplikacji webowej prezentującej SGJP jest Jan Szejko.

1. Materiał słownika

SGJP dysponuje rozległą podstawą leksykalną. Zasadnicza jego część składa się z listy haseł SJPDor, którego zasób słownictwa sięgał nawet ostatnich dziesięcioleci XVIII wieku, w związku z tym słownik zawiera sporo słownictwa przestarzałego, dziś nieużywanego nawet w języku literackim. Z drugiej strony jednak materiał SGJP był uzupełniany przez jego autorów na różnych etapach prac, w tym również o słownictwo najnowsze. W efekcie zatem materiał SGJP jest obecnie znacznie obszerniejszy niż SJPDor i obejmuje ponad 330 000 haseł zawierających w swoich paradygmatach łącznie blisko 4 300 000 form wyrazowych generowanych przez

* wkieras@ipipan.waw.pl, wolinski@ipipan.waw.pl

1 Współautorstwo obejmuje wszystkie etapy pracy nad artykułem. Autorzy dziękują prof. Zygmuntowi Saloniemu za cenne uwagi redakcyjne i pomoc w doborze przykładów.

1114 ręcznie tworzonych wzorów odmiany. Szczegółowe różnice liczbowe między pierwszym wydaniem SGJP z 2007 roku a obecnym stanem w wersji online słownika z podziałem na klasy gramatyczne prezentuje tabela 1.

Tabela 1. Szczegółowe porównanie ilościowe zmian w SGJP między pierwszym wydaniem a obecną wersją internetową

	1. WYDANIE		3. WYDANIE (ONLINE)	
	HASEŁ	WZORÓW	HASEŁ	WZORÓW
w sumie	**244 669**	**1 095**	**334 845**	**1 114**
prefiksy	81	2	112	2
leksemy	244 588	1 095	334 733	1 116
o odmianie rzeczownikowej	135 529	762	172 178	767
zaimki osobowe	6	6	6	6
odsłowniki	29 590	2	29 526	2
odprzym. nazwy cech na *-ość*	28 980	1	62 445	1
pozostałe rzeczowniki	76 953		80 201	
nazwy własne	8 782		10 710	
pospolite	68 171		69 491	
o odmianie przymiotnikowej	65 671	71	103 761	77
imiesłowy	34 301		36 304	
czynne	13 931	1	13 877	1
bierne	20 370		22 427	4
przymiotniki	31 370	71	67 457	77
stopnia wyższego	950	1	1 065	1
stopnia równego	30 420	70	66 392	77
przysłówki odprzymiotnikowe	11 146	1	26 577	1
stopnia wyższego	1 106	1	1 243	1
stopnia równego	10 040	1	25 512	1
liczebniki	98	45	117	47
czasowniki	29 532	215	29 955	222
predykatywy	35	1	34	1
odmienne	29 497	214	29 925	221
pozostałe	2 612	2	1 967	2
pozostałe przysłówki	491	1	502	1
partykuły	193	1	201	1
przyimki	113	2	118	2
spójniki	121	1	121	1
wykrzykniki i zbliżone	458	1	490	1
skróty	1 117	1	392	17
inne	119	1	143	1

Inaczej niż większość innych słowników SGJP nie objaśnia znaczeń wyrazów. Co prawda w części haseł pojawiają się różnego rodzaju informacje sygnalizujące znaczenie lub przybliżające je, ale spełniają one jedynie zadanie pomocnicze. Konstytutywną cechą hasła słownikowego w SGJP jest postać paradygmatu odmiany i wypełnienie jego klatek formami gramatycznymi – w konsekwencji dwa różne leksemy o identycznej odmianie będą opisane w jednym haśle. Przykładowo, dwa leksemy MATKA o znaczeniach ‘kobieta mająca dziecko’ oraz ‘mała mata’, słowotwórczo zupełnie niepowiązane, są opisane w jednym haśle, ponieważ ich paradygmaty niczym się od siebie nie różnią. Z drugiej strony istnieją w SGJP trzy leksemy TYTAN: 1. ‘przen. osoba bardzo silna’, 2. ‘rodzaj chrząszcza’, 3. ‘pierwiastek chemiczny’, różniące się rodzajem gramatycznym i w konsekwencji również postacią paradygmatu – układem synkretyzmów form biernikowych, a także obecnością form męskoosobowych w mianowniku i wołaczu liczby mnogiej (w wypadku pierwszego z nich) i zakończeniem dopełniacza liczby pojedynczej (*-u* w trzecim znaczeniu, *-a* zaś w pozostałych).

W przeciwieństwie do słowników ogólnych lub specjalistycznych zawierających informacje o odmianie, które podają zwykle tylko wybrane formy lub zakończenia niektórych charakterystycznych dla leksemu form, SGJP prezentuje wszystkie formy fleksyjne w postaci osobnych tabel dla każdego leksemu. Jest to główne zadanie słownika – wszelkie inne informacje, np. na temat nacechowania stylistycznego, zakresu tematycznego, znaczenia czy pochodzenia, są wyłącznie pomocnicze i nie mają charakteru systematycznego.

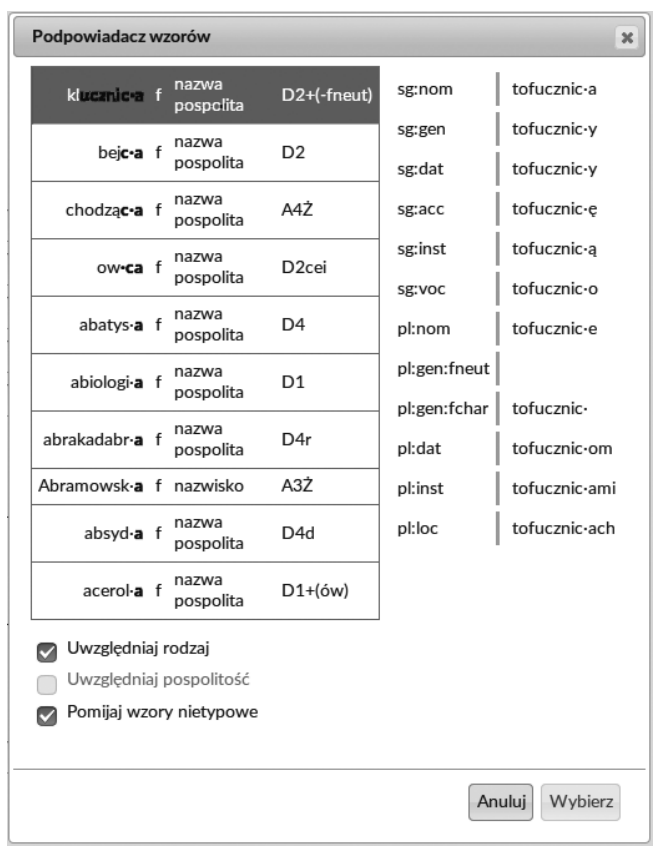
2. Kuźnia

W trakcie prac nad pierwszym i drugim wydaniem słownika autorzy posługiwali się sześcioma bazami danych programu Microsoft Access, z których każda miała inną budowę, dostosowaną do specyfiki wzorów poszczególnych typów. W efekcie każda baza miała swojego opiekuna i tylko on dokonywał w niej zmian. Wtedy, gdy te bazy powstawały, nie wszyscy autorzy dysponowali stałym szerokopasmowym dostępem do Internetu i taki tryb pracy wydawał się najpraktyczniejszy. Z czasem jednak widać było również zasadnicze ograniczenia tego sposobu pracy, które hamowały możliwość szybkiej aktualizacji haseł lub poprawiania błędów dostrzeżonych przez użytkowników. Przeniesienie SGJP do Internetu, przy jednoczesnym założeniu, że słownik będzie rozwijany, wymagało stworzenia zupełnie nowych narzędzi nie tylko do prezentacji danych, ale i do ich redagowania, jak również scalenia osobnych baz dla różnych części mowy w jedną dużą bazę danych. W efekcie powstała aplikacja webowa o nazwie Kuźnia (Szejko 2014), dzięki której redaktorzy dysponujący odpowiednimi uprawnieniami mogą dokonywać zmian w całym materiale słownika, a efekt ich pracy jest widoczny od razu. Kuźnia stanowi zatem interfejs edycyjny dla nowej bazy danych, jak również interfejs czytelnika. W tym drugim wypadku jest to oczywiście poziom dostępu ograniczony jedynie do oglądania zawartości słownika (czyli haseł i wzorów), nie umożliwia zaś jego edycji.

Nowy interfejs edycyjny zasadniczo zmienił sposób pracy jego redaktorów. W poprzedniej wersji SGJP redaktor tworzący nowe hasło musiał utworzyć rekord w bazie danych, nie mógł jednak od razu zobaczyć bezpośredniego efektu swojej pracy w postaci tabeli odmiany danego leksemu, a co za tym idzie – nie mógł też sprawdzić poprawności paradygmatu. W wersji online autor nie tylko może zobaczyć, w jaki sposób zmienia się paradygmat odmiany leksemu pod

wpływem zmiany wzoru, ale może też liczyć na podpowiedź, jaki wzór wybrać. Narzędzie podpowiadające wzory odmiany dla wpisanej formy hasłowej proponuje je na podstawie zakończenia tej formy oraz informacji o tym, jak odmieniają się hasła obecne już w bazie o takich samych zakończeniach oraz – w wypadku rzeczowników – o takim samym rodzaju gramatycznym. Jest to szczególnie przydatne przy wprowadzaniu do słownika słownictwa najnowszego, bazującego na produktywnych formantach. Przykładowo, słowu TOFUCZNICA ‘danie z tofu naśladujące jajecznicę’, któremu leksykograf przypisał część mowy i rodzaj, mechanizm podpowiadania wzorów w pierwszej kolejności przypisał taki sam wzór, jaki w słowniku ma słowo KLUCZNICA (a także m.in. JAJECZNICA). Jest to zatem podpowiedź poprawna, choć – jak widać na rysunku 1 – dostępne są również inne propozycje, w tym m.in. dwie o odmianie przymiotnikowej. Nie jest to oczywiście mechanizm niezawodny, a jedynie pomoc dla redaktora ograniczająca liczbę błędów związanych z przypisaniem wzoru odmiany.

Kuźnia jest aplikacją, która może być używana nie tylko do rozwijania SGJP, ale też do wszelkich innych słowników fleksyjnych polszczyzny współczesnej i historycznej zgodnych z modelem SGJP. Wymaga oczywiście dostarczenia odpowiednich danych leksykalnych oraz bazy wzorów odmiany.



Rysunek 1. Podpowiadanie wzorów

3. Filtry

Dla użytkowników SGJP prawdopodobnie najistotniejszą nowością jest znacznie bardziej rozbudowany system filtrów. We wcześniejszych wersjach słownika użytkownik mógł wyszukiwać formy hasłowe lub gramatyczne w całym słowniku, ewentualnie ograniczając wyszukania do klas gramatycznych (czasowników, rzeczowników itd.). W wersji internetowej możliwości wyszukiwania i filtrowania wyników są znacznie większe. Użytkownik może wyszukać leksem ze względu na kształt jego formy hasłowej lub dowolnej formy w jego paradygmacie, w szczególności zaś ze względu na zawieranie odpowiednich podciągów znaków czy odpowiadanie wzorcowi zadanemu przez wyrażenie regularne[2]. Może również wyszukiwać leksemy ze względu na to, według jakich wzorów lub typów wzorów się odmieniają. Wreszcie, może wyszukiwać leksemy ze względu na ich cechy gramatyczne – przynależność do odpowiedniej klasy gramatycznej, wartość kategorii rodzaju (dla rzeczowników) lub aspektu (dla czasowników i ich derywatów), przechodniość.

Możliwość tworzenia kryteriów filtrowania dotyczy również kwalifikatorów i informacji o tym, czy dany rzeczownik jest nazwą własną, a jeśli tak, to jakiego typu. Należy przy tym pamiętać, że kwalifikatory stylistyczne i zakresowe oraz typologia nazw własnych w słowniku stanowią jedynie informację pomocniczą, mającą na celu ułatwienie rozszyfrowania znaczeń, zwłaszcza w wypadku haseł homonimicznych, nie stanowią jednak systematycznej i w pełni uporządkowanej informacji słownikowej.

Różne kryteria filtrowania można ze sobą łączyć, definiując w jednym zapytaniu kilka warunków połączonych za pomocą koniunkcji lub alternatywy, czyli w taki sposób, że każdy wyszukany leksem musi spełniać każdy warunek albo każdy leksem musi spełniać przynajmniej jeden z warunków.

Załóżmy, że chcemy odnaleźć w słowniku wszystkie rzeczowniki żeńskie zakończone na *-yni* typu JĘZYKOZNAWCZYNI, ZDOBYWCZYNI, będące derywatami od odpowiednich rzeczowników męskich. W tym celu można nałożyć filtr na klasę gramatyczną (rzeczownik) i rodzaj (f). W tym konkretnym wyszukaniu określenie klasy gramatycznej jest zresztą nadmiarowe, ponieważ jedyną – oprócz właściwych rzeczowników – wyodrębnioną w SGJP klasą leksemów, które mają przypisany rodzaj gramatyczny, są odsłowniki, nie mogą one jednak mieć rodzaju żeńskiego i nie charakteryzują się wymienionym wyżej zakończeniem. Następnie można dodać warunek na postać formy hasłowej: *Hasło kończy się na yni*. Użytkownik powinien uzyskać w ten sposób listę 130 rzeczowników żeńskich o poszukiwanym zakończeniu – ich lista pojawi się w oknie haseł po lewej stronie, a dokładna ich liczba zostanie wyświetlona na górnym pasku (w tytule karty) przeglądarki internetowej. Jeśli chcemy rozszerzyć zapytanie tak, by lista objęła również rzeczowniki zakończone na *-ini* typu CZŁONKINI, BOGINI, nie uzyskamy pożądanego efektu przez dodanie kolejnego warunku opartego na postaci zakończenia formy hasłowej, ponieważ te dwa warunki wykluczają się wzajemnie (żaden wyraz nie może mieć jednocześnie zakończeń *-yni* i *-ini*). Nie można też zmienić sposobu łączenia warunków

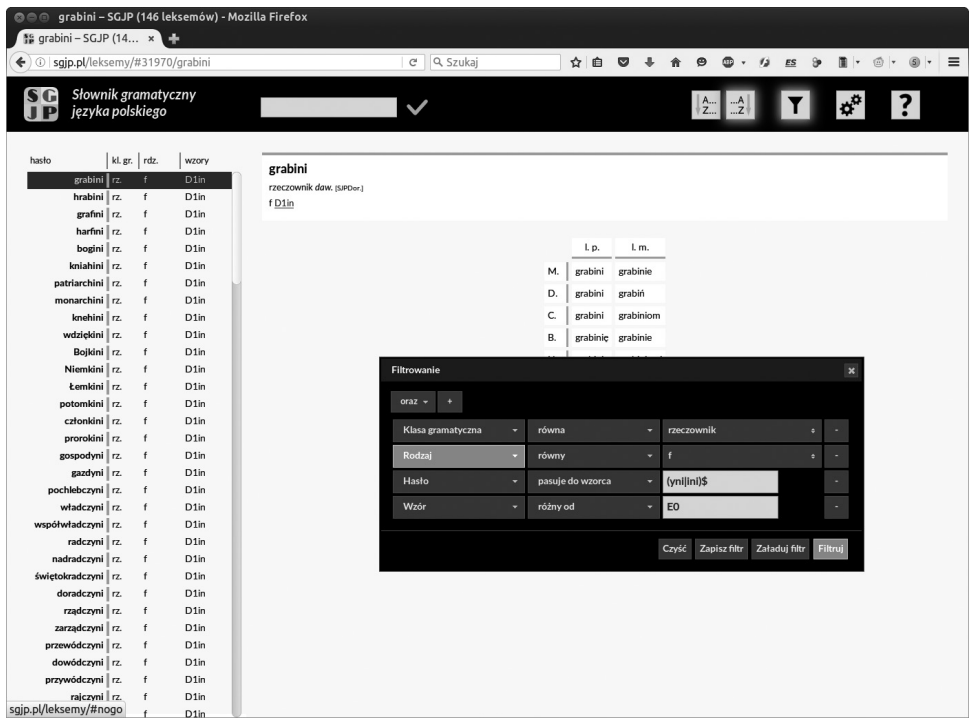
2 Wyrażenia regularne to wzorce, zapisane w zwięzły, formalny sposób, opisujące ciągi znaków pewnego typu.

z koniunkcji na alternatywę (nawet po pominięciu nadmiarowego warunku opartego na kategorii gramatycznej), ponieważ do zbioru wynikowego wpadłyby zarówno wszystkie rzeczowniki żeńskie, jak i wszystkie leksemy o formie hasłowej zakończonej na *-yni* (np. BURSZTYNI, DZIEWCZYNI) lub *-ini* (np. MARTINI, BIKINI, PAGANINI). Możemy oczywiście wywołać dwa oddzielne zapytania (każde dla jednego z dwu zakończeń) i obejrzeć wyniki obu wyszukiwań, co jednak może okazać się mało wygodne. Dlatego ciekawszą możliwością będzie zmodyfikowanie zapytania w taki sposób, by połączyć w jednym warunku dwa typy zakończeń. W tym celu należy skorzystać z opcji *Hasło pasuje do wzorca...* i jako wzorzec podać: `(yni|ini)$`. Wzorzec jest wyrażeniem regularnym, zapisanym według powszechnie stosowanej konwencji opisanej w wielu miejscach, w tym również w polskiej Wikipedii. Ujęte w nawiasy okrągłe wyrażenie składa się z dwu zakończeń oddzielonych znakiem pionowej kreski oznaczającej alternatywę oraz znaku `$` mówiącego, że szukane leksemy mają mieć formy hasłowe o zakończeniu określonym bezpośrednio przed tym znakiem. Dzięki tak zmodyfikowanemu filtrowi lista wyszukanych leksemów powiększy się do 147, ale pojawi się na niej jeden niepożądany element: rzeczownik MINI o podwójnej kwalifikacji rodzajowej (żeńsko-nijakiej) – spełnia on wszystkie warunki formalne filtru, lecz nie jest to derywat żeński i w związku z tym nie pasuje do grupy rzeczowników, których szukamy. Przy konstruowaniu nieco bardziej złożonych filtrów dość często zdarza się, że na liście wyników pojawiają się tego typu elementy niepożądane, które można jednak eliminować na różne sposoby. W tym wypadku można np. dodać warunek wykluczający leksemy, którym przypisano jakikolwiek inny rodzaj niż żeński (filtr: *Rodzaj różny od...*), lub wykluczający rzeczowniki formalnie nieodmienne, które zawsze mają przypisany wzór E0 (*Wzór różny od E0*). Dzięki temu na liście wyników uzyskamy 146 leksemów spełniających warunki określone na początku (por. rysunek 2).

Użytkownik łatwo zauważy, że wszystkie leksemy na liście odmieniają się według tego samego wzoru: D1in, czyli wzoru odmiany typu żeńskiego o temacie zakończonym na spółgłoskę miękką, charakteryzującym się w zapisie wymianą końcowego *ni* w formie podstawowej na *ń* (w dopełniaczu liczby mnogiej). Można zatem spróbować sprawdzić, czy wśród derywatów żeńskich tego typu znajdą się takie, które mają zakończenie *-ni*, ale nie *-yni* lub *-ini*. W tym celu można odfiltrować wszystkie leksemy żeńskie, które odmieniają się według tego wzoru – powinno ich być 153, czyli o siedem więcej niż w poprzednim wyszukaniu. Istnieje zatem siedem leksemów o takiej cesze, można je znaleźć, rozszerzając filtr o warunek niepasowania do wzorca `(yni–ini)$`. Na liście powinny znaleźć się rzeczowniki: ACANI, ASANI, KSIENI, ŁANI, MAHARANI, WACANI, WASANI. Pięć z nich to faktycznie derywaty żeńskie odpowiednio od rzeczowników: ACAN, ASAN, MAHARADŻA, WACAN i WASAN. Pozostałe dwa zaś – KSIENI i ŁANI – nie mają swoich męskich odpowiedników, choć reprezentują ten sam typ odmiany.

W podobny sposób można wyszukać np. wszystkie rzeczowniki męskie, które w wołaczu mają (jako oboczną lub jako jedyną) nietypową formę na *-cze*, np. GŁUPIEC (*głupcze*), SZEWC (*szewcze*). W tym wypadku wystarczy zdefiniować filtr ograniczający wyniki do rzeczowników o rodzaju innym niż żeński, nijaki lub przymnogi (czyli w praktyce w jednym z trzech rodzajów męskich), których forma hasłowa kończy się na *-c* i które mają w swoim paradygmacie formę kończącą się na *-cze*. Na liście powinno znaleźć się ponad sześćset takich leksemów. W tym

wypadku użytkownik może zauważyć, że wszystkie leksemy na liście będą miały przypisany wzór zawierający znak wykrzyknika (!). Znak ten informuje o tym, że wzór uwzględnia nietypową końcówkę wołacza liczby pojedynczej (nie tylko końcówkę *-cze*, ale też np. końcówkę zerową; zob. Saloni 2016). Oczywiście nie wszystkie zjawiska powierzchniowofleksyjne są opisane za pomocą odpowiednich symboli we wzorze, ale w wielu wypadkach mogą się one okazać przydatne przy wyszukiwaniu obszernych list leksemów o konkretnych cechach.



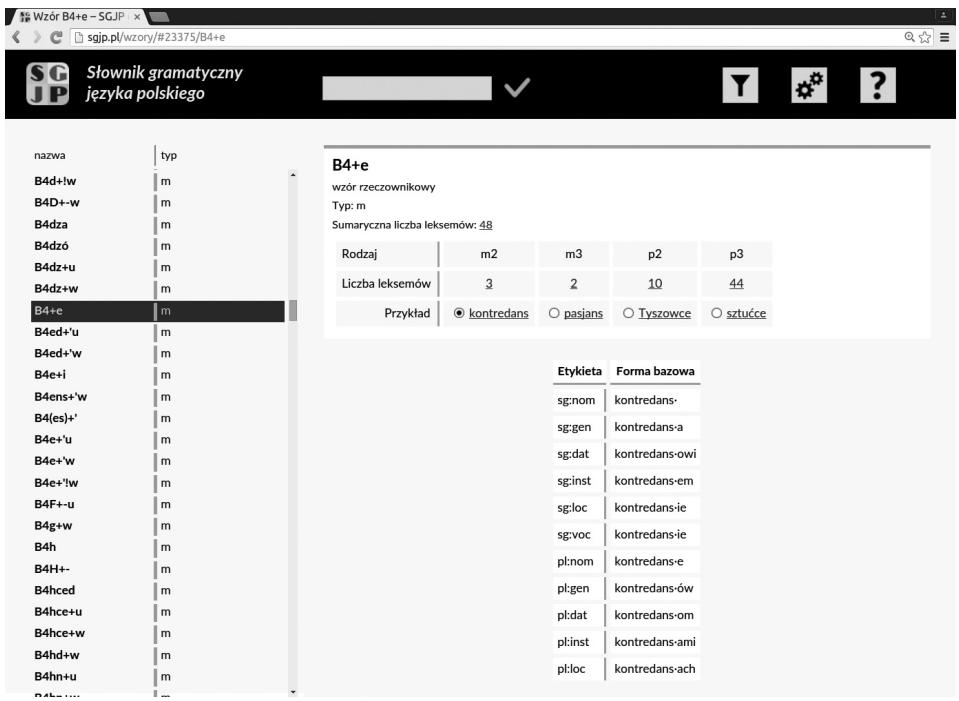
Rysunek 2. Przykład filtru składającego się z czterech warunków, z których trzeci ilustruje możliwość wykorzystania wyrażeń regularnych do przeszukiwania zawartości słownika

Filtry w postaci dostępnej w nowej wersji SGJP dają użytkownikowi znacznie większe możliwości niż w wersjach wcześniejszych. Oczywiście wraz z siłą wyrazu zwiększa się też stopień skomplikowania struktury filtrów. Niektóre decyzje związane z filtrami były również podyktowane względami praktycznymi, przede wszystkim szybkością przeszukiwania obszernej bazy danych.

4. Wzory

Ważną zmianę w stosunku do poprzednich wydań SGJP stanowi prezentacja wzorów odmiany jako integralnej i pełnoprawnej części słownika. O ile wcześniej wzory odmiany z punktu widzenia użytkownika były jedynie enigmatycznymi etykietami przypisanymi do haseł,

których szczegółowe znaczenie dostrzegali jedynie najbardziej wnikliwi czytelnicy, o tyle teraz stanowią one komponent, który użytkownik może przeglądać i wykorzystywać przy wyszukiwaniach (pokazano to wyżej). Same oznaczenia wzorów również uległy zmianie, a system nazewniczy wzorów został opisany i udostępniony na stronie słownika, jak również w postaci osobnego artykułu (por. Saloni 2016). Nowy sposób etykietowania wzorów nawiązuje do oznaczeń grup deklinacyjnych i koniugacyjnych Jana Tokarskiego w postaci znanej z SJPDor i powielanej w wielu późniejszych słownikach, powinien więc być stosunkowo łatwy do zrozumienia dla czytelników obytych z tym tradycyjnym opisem odmiany w polszczyźnie. Wzory SGJP są oczywiście znacznie bardziej szczegółowe i jest ich dużo więcej niż grup Tokarskiego, ponieważ muszą uwzględniać wszystkie możliwe alternacje, często bardzo rzadkie i specyficzne, które w wypadku grup Tokarskiego były opisywane w haśle słownikowym.



Rysunek 3. Widok wzorów w słowniku. Na ilustracji widać wzór `B4+e`, czyli wzór odmiany męskiej o zakończeniu tematu na spółgłoskę twardą, charakteryzujący się dodatkowo specyficzną końcówką *-e* w mianowniku liczby mnogiej. Wzór jest używany przez 48 różnych leksemów o czterech różnych rodzajach: dwóch męskich i dwóch przymnogich

Opisany wcześniej system filtrów jest dostępny również w widoku wzorów, choć oczywiście kryteria ich budowania są znacznie skromniejsze – ograniczają się głównie do nazwy wzoru i do typu odmiany. Dzięki nim można np. sprawdzić, ile jest w całym słowniku wzorów czasownikowych, czyli takich, których nazwa zaczyna się od litery V, lub ile wariantów w SGJP ma konkretna grupa koniugacyjna Tokarskiego (np. grupa V została rozbita aż na

27 wzorów, bardzo regularna grupa IV zaś nie wymagała tworzenia żadnych dodatkowych wariantów). Przejrzysty sposób tworzenia etykiet wzorów pozwala też na ustalenie słownikowej frekwencji poszczególnych grup deklinacyjnych i koniugacyjnych w widoku leksemów. Przykładowo, czasowników ze wspomnianej grupy V w SGJP jest ponad cztery razy mniej niż czasowników z grupy IV.

5. Przyszłość słownika

Specyfika słownika elektronicznego funkcjonującego w medium internetowym powoduje, że powinien się on znajdować niemal w ciągłym rozwoju. Dotyczy to zresztą wszelkiego rodzaju innych zasobów i narzędzi lingwistycznych: np. korpusów, gramatyk formalnych, parserów. Gdy przestają być rozwijane i poprawiane, szybko się dezaktualizują lub powiązane z nimi oprogramowanie przestaje działać. Słownik elektroniczny wymaga zatem stałej opieki, która z kolei wymaga nakładów czasu i pracy ze strony jego redaktorów. Bodźcem do pracy jest nie tylko powiększone grono użytkowników publicznie dostępnego słownika internetowego, ale także zainteresowanie ze strony innych projektów, w których jego dane mogą okazać się przydatne. Dane SGJP są wykorzystywane przynajmniej w dwóch takich projektach: stanowią źródło informacji gramatycznej dla *Wielkiego słownika języka polskiego* PAN (Żmigrodzki 2015) oraz są podstawą słownikową analizatora morfologicznego Morfeusz w jednej z jego wersji (Woliński 2014), używanego przez wiele innych programów. W obu tych istotnych zastosowaniach SGJP powinien być wykorzystywany również w perspektywie co najmniej najbliższych kilku lat. Ponadto słownik jest też coraz częściej przytaczany jako źródło informacji gramatycznej w innych zasobach internetowych, zwłaszcza od czasu udostępnienia wersji online. Przykładowo, w Wikisłowniku (https://pl.wiktionary.org) SGJP jest przywoływany co najmniej kilkaset razy.

Pomimo swojej rozległej podstawy leksykalnej SGJP ma też pewne obszerne luki, wśród nich są przede wszystkim nazwy własne, które nigdy nie były w nim notowane w sposób systematyczny. Niewielka próbka nazw własnych obecnych w SGJP obejmuje około 11 tysięcy haseł, z czego ponad 4300 to nazwiska i ich człony, blisko 3100 to imiona, a 2700 to różnego rodzaju nazwy geograficzne i ich człony (głównie nazwy miejscowości). Nazwy własne stanowią zatem niespełna 3,3 proc. wszystkich jednostek zanotowanych w słowniku, co nie odzwierciedla faktycznego udziału tego typu słownictwa w tekstach polskich. Uzupełnienie tej luki jest możliwe, wymaga jednak zaplanowania i zrealizowania poważniejszych prac badawczych i leksykograficznych.

Bibliografia

Gruszczyński W. 1989: *Fleksja rzeczowników pospolitych we współczesnej polszczyźnie pisanej*, Zakład Narodowy im. Ossolińskich, Warszawa–Wrocław.

Saloni Z. 2001: *Czasownik polski*, Wiedza Powszechna, Warszawa.

Saloni Z. 2016: *Systematyzacja wzorów fleksyjnych dla „Słownika gramatycznego języka polskiego"*, „Polonica" XXXVI, s. 5–18.

Saloni Z., Gruszczyński W., Woliński M., Wołosz R. 2007*a*: *Grammatical dictionary of Polish. Presentation by the authors*, „Studies in Polish Linguistics" 4, s. 5–25.

Saloni Z., Gruszczyński W., Woliński M., Wołosz R. 2007*b*: *Słownik gramatyczny języka polskiego*, Wiedza Powszechna, Warszawa.

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

SJPDor: *Słownik języka polskiego*, red. W. Doroszewski, t. 1–4, Wiedza Powszechna, Warszawa 1958–1962, t. 5–11, Państwowe Wydawnictwo Naukowe, Warszawa 1963–1969.

Szejko J. 2014: *Aplikacja webowa do opracowywania słowników fleksyjnych*, praca magisterska, Uniwersytet Warszawski, Warszawa.

Woliński M. 2014: *Morfeusz reloaded*, [w:] N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, ELRA, Reykjavík, Iceland, s. 1106–1111 (online: http://www.lrec-conf.org/proceedings/lrec2014/index.html).

Zalizniak A.A. 2003: *Grammaticzeskij słowar' russkogo jazyka*, wyd. 4, Russkij jazyk, Moskwa.

Żmigrodzki P. 2015: *Wielki słownik języka polskiego PAN – historia, stan obecny i perspektywy rozwoju po 2018 roku*, „Biuletyn Polskiego Towarzystwa Językoznawczego" LXXI, s. 177–187.

Summary

Grammatical Dictionary of Polish – an online version

Keywords: inflection, surface morphology, dictionary.

In the paper we present a new online version of the *Grammatical Dictionary of Polish*. We focus on new features which have been introduced since the previous editions of the dictionary which were available as a desktop application. The most interesting changes include a full view of inflectional patterns together with their new systematization and significantly extended searching tools allowing to filter entries of the dictionary based on various criteria. Some practical examples of using these new features are presented.

MONIKA CZEREPOWICKA* | UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

ZYGMUNT SALONI** | UNIWERSYTET WARSZAWSKI

O odmianie czasowników o podstawach *mleć* i *pleć*[1]

Słowa kluczowe: czasowniki, odmiana, fleksja, paradygmat.

Z czasownikami o podstawach ***mleć*** i ***pleć*** (tzn. z grupą czasowników różniących się od wskazanych tylko przedrostkiem) mamy w języku polskim tradycyjnie kłopoty. Są to kłopoty praktyczne, bo słowniki zgodnie, i to od kilku pokoleń, podają zestaw form, które są przyporządkowane wskazanym bezokolicznikom:

1	*miele*	*piele*
2	*mielę*	*pielę*
3	*mielą*	*pielą*
4	*miel*	*piel*
5	*mleć*	*pleć*
6	*mełł*	*pełł*
7	*mełłem*	*pełłem*
8	*mełło*	*pełło*
9	*mełli*	*pełli*
10	*mielono*	*pielono*
11	*mielenie*	*pielenie*
12	*mieleni*	*pieleni*

Jest to zbiór form bazowych (pozwalających utworzyć wszystkie formy za pomocą prostych automatycznych zabiegów) ze *Słownika gramatycznego języka polskiego* (SGJP): zestaw jedenastu form wprowadzony przez Jana Tokarskiego (1951) i powszechnie używany w polskich słownikach, poczynając od SJPDor (Tokarski 1958), wzbogacony o formę dwunastą (nigdy nie mającą indywidualnego alternanta, ale zsynchronizowaną z formą jedenastą, jak w przykładzie wyżej, albo dziesiątą, jak dla *chciani*). Te formy są zhierarchizowane. Z punktu

* czerepowicka@gmail.com

** zasaloni@uw.edu.pl

1 Współautorstwo obejmuje wszystkie etapy pracy nad artykułem.

widzenia ogólnego obrazu koniugacji polskiej (także – co ma dla nas mniejszą wagę – w perspektywie historycznej) najważniejszy jest podział na trzy podzbiory: (A) formy od tematu czasu teraźniejszego (1–4), (B) formy od tematu czasu przeszłego (6–9) oraz (C) bezosobnik, imiesłów przymiotnikowy bierny, odsłownik (10–12). W całej koniugacji formy C wyodrębniają się wyraźnie, jednak na ogół są zbliżone do form B – typu przeszłego; natomiast dla rozważanej grupy jej alternant tematu wzięty jest od form typu A. Bezokolicznik (5) tworzy się na ogół, ale z wyjątkami, od tematu czasu przeszłego. Dla rozważanej grupy ma on własny alternant tematu.

Czasowniki o podstawach ***mleć*** i ***pleć*** mają odmianę indywidualną i nietypową. W gramatykach omawia się je jako osobną grupkę. J. Tokarski w *Czasownikach polskich* (1951) utworzył dla nich specjalną podgrupę XIIIe (typ *miele – mełł*). Po uproszczeniu przez autora jego systematyzacji koniugacji zostały one włączone do grupy XI, formalnie nie zawierającej podgrup (formy podawane przy haśle pozwalają na utworzenie pełnego paradygmatu: takiego jak podany na początku niniejszego artykułu).

Taki paradygmat przedstawiają niemal wszystkie współczesne podręczniki i poradniki językowe. Wyjątkiem jest tu *Nowy słownik poprawnej polszczyzny* (NSPP)[2], podający w haśle **mleć**: „(*nie*: mlić) ndk XI, mielę, mielisz, miele, mełł (*nie*: mleł, *nie*: miołł, *nie*: mlił), mełła (*nie*: mleła, *nie*: miołła, *nie*: mliła), mełliśmy (*nie*: miełliśmy), mielony (*nie*: mełty, *nie*: mlety)". Zaskakujące jest tu zestawienie obok siebie form 2. i 3. osoby liczby pojedynczej czasu teraźniejszego nie powiązanych wzajemnie, należących do różnych serii (formy z obu tych serii występują oczywiście w tekstach), naruszające związki między formami, jak by się wydawało we fleksji polskiej bezwyjątkowe (w formach tych powinien wystąpić ten sam alternant tematu: w 2. osobie + *isz* wtedy i tylko wtedy, gdy w 3. osobie końcówka + *i*). Ta decyzja NSPP jest najwyraźniej pomyłką techniczną, a może interpretacyjną[3]. Nie ma analogicznej formy w haśle **pleć**, w którym jest też odesłanie bez oceny poprawnościowej („por." – i tak samo w drugą stronę) do **pielić**. Na marginesie odnotujmy, że słowniki przyjmują w wypadku analizowanych czasowników postawę wyraźnie normatywną, widoczną nawet w doborze haseł. Datuje się ona co najmniej od SJPDor, w którym jest hasło **pielić** z kwalifikatorem „*rzad.*", natomiast hasła **mielić** nie ma.

Omawiane czasowniki są mieszane – jak wskazują już odsyłacze w SJPDor – z czasownikami grupy VI, o odmianie regularnej ze stałym przyrostkiem tematowym *-i-* w formach podzbiorów (A) i (B) oraz *-on-/-en-* w formach podzbioru (C). Dla czasownika należącego do tej grupy, np. ***mielić***, wszystkie formy tworzy się jednoznacznie. W Narodowym Korpusie Języka Polskiego (NKJP; podkorpus zrównoważony) liczba form przyporządkowanych do poszczególnych leksemów przedstawia się następująco:

2 Odnotujmy (na życzenie recenzenta), że taki zapis powtarza *Wielki słownik poprawnej polszczyzny* (WSPP).

3 Tak też oceniła tę sytuację p. Alicja Witorska, która była redaktorem koordynującym tego słownika. Taka pomyłka może być brzemienna w skutki, może być bowiem traktowana jako obowiązujące zalecenie normatywne – to zaś doprowadzi do wprowadzenia dodatkowego (i zbędnego) rozgraniczenia form w paradygmacie.

Tabela 1. Frekwencja form leksemów *mleć* i *pleć* w NKJP – podkorpus zrównoważony

LEKSEM	LICZBA FORM	LEKSEM	LICZBA FORM
mleć	849	mielić	116
pleć[4]	323	pielić	69

Wyszukiwanie słów odpowiadających formom tych leksemów daje wynik następujący:

Tabela 2. Zestawienie form na podstawie podkorpusu zrównoważonego NKJP. Liczby podane bezpośrednio po słowie odpowiadają liczbie jego wystąpień w badanym korpusie, liczby w nawiasie – liczbie jego wystąpień zapisanych wielką literą albo wielkimi literami lub też poprzedzonych prefiksem *nie+*.
Oznaczenie .* zastępuje dowolny ciąg liter

MIELIĆ	MLEĆ	PIELIĆ	PLEĆ
mieli: 1	miele: 65 (+14 +1)	pieli: 0	piele: 6
mielisz: 1	mielesz: 0	pielisz: 0	pielesz: 1
mielimy: 26 (+3)	mielemy: 5	pielimy: 1	pielemy: 0
mielicie: 1	mielecie: 2	pielicie: 0	pielecie: 0
mielę: 0	mielę: 7	pielę: 0	pielę: 5 (+26)
mielą: 0	mielą: 81 (+7)	pielą: 0	pielą: 11 (+2)
mieląc.*: 0	mieląc.*: 58 (+3)	pieląc.*: 0	pieląc.*: 15 (+1)
miel: 0	miel: 8	piel: 0	piel: 5 (+1+79)
mielmy: 0	mielmy: 0	pielmy: 0	pielmy: 0
mielcie: 0	mielcie: 1	pielcie: 0	pielcie: 0
mielił.*: 52 (+2)	mełł.*: 85 (+7)	pielił.*: 27 (+2)	pełł.*: 8
w tym:			
mielił: 26	mełł: 32 (+5)	pielił: 9	pełł: 2
mieliła: 13 (+2)	mełła: 19 (+2)	pieliła: 12 (+2)	pełła: 3
mieliło: 2	mełło: 9	pieliło: 1	pełło: 1
mieliły: 11	mełły: 25	pieliły: 5	pełły: 2
mielili: 5	mełli: 3 (+1)	pielili: 2 (+1)	pełli: 0
mielić: 25 (+1)	mleć: 42 (+1+1)	pielić: 25	pleć: 66 (+5)

4 W SGJP (za SJPDor) są dwa leksemy o wyrazie hasłowym *pleć* – czasownik i rzeczownik. Rzadki rzeczownik ***pleć*** ‘plecenie, splot’ ma we wszystkich formach *e* w temacie; homonimiczna jest więc tylko forma hasłowa – z bezokolicznikiem czasownika. Tymczasem w NKJP wśród 73 wystąpień słowa *pleć* i *Pleć* tylko 2 są uznane za formę rozkazującą czasownika; pozostałe 71 zostały otagowane jako mianownik rzeczownika – błędnie. Roboczo przyjęliśmy więc, że wszystkie przyporządkowania danemu słowu wyrazu hasłowego *pleć* odnoszą się do form czasownikowego leksemu ***pleć***.

mielon.*: 0	mielon.*: 194 (+13+1)	pielon.*: 0	pielon.*: 3 (+4)
w tym:			
mielono: 0	mielono: 18	pielono: 0	pielono: 1
mielony: 0	mielony: 3	pielony: 0	pielony: 0

mielen.*: 0	mielen.*: 235 (+13+1)	pielen.*: 0	pielen.*: 79 (+2)
w tym:			
mielenie: 0	mielenie: 21	pielenie: 0	pielenie: 33

Na dane te wyraźny wpływ ma powiązanie form z określonym hasłem identyfikującym leksem (czyli *hasłowanie* albo *lematyzowanie*; obecnie technicznie czynność tę nazywa się *tagowaniem*). Postaraliśmy się więc zweryfikować te dane – przez bardziej złożone wyszukiwania, lekturę przykładów i ich analizę.

Zaczniemy od pary czasowników ***pleć*** i ***pielić***. Musimy jednak wyjść poza dane przedstawione w tabeli. Notuje ona bowiem brak form czasu teraźniejszego *pieli*; tymczasem w korpusie jest 14 wystąpień segmentu *pieli*, wszystkie otagowane jako formy leksemu ***piać***, choć tylko 3 powinny być potraktowane w ten sposób, a 9 – odpowiada formie [pielić:fin:sg:ter:imperf][5] (ponadto jest jeszcze jedna, niepoprawna, forma czasownika ***piąć*** **(*się*)** i jedna pomyłka literowa). A w całym podkorpusie zrównoważonym jest tylko 6 wystąpień segmentu *piele* zinterpretowanych – poprawnie – jako odpowiadające [pleć:fin:sg:ter:imperf]. Weźmy pod uwagę jeszcze te słowa, dla których występuje homonimia form od czasowników obu grup: *pielę* – 5 wystąpień, *pielą* – 12, formy imiesłowów współczesnych (przysłówkowego i przymiotnikowego) – 17, formy rozkaźnika – 2, formy bezosobnika, imiesłowu przymiotnikowego biernego i odsłownika (traktowanych w NKJP jako formy leksemu czasownikowego) – 88, czyli 124 wystąpienia. Po odjęciu tej liczby od ogólnej liczby wystąpień form leksemu ***pleć*** (323) otrzymamy 199 wystąpień słów, które możemy (z pewnym przybliżeniem) uważać za jednoznaczne wykładniki form tego leksemu. Gdybyśmy hasłowali odwrotnie, niż zrobiono w NKJP, tj. wszystkie formy homonimiczne potraktowali jako wystąpienia leksemu ***pielić***, a ponadto doliczyli do niego wystąpienia formy czasu teraźniejszego *pieli* (9), to byłaby równowaga w stosunku do ***pleć*** – 198 : 199, niekorzystna dla ***pielić*** (bo wliczono tu wystąpienia słów homonimicznych). Ale liczba wystąpień leksemu ***pleć*** (który automatycznie traktujemy jako czasownik, choć jest i rzeczownik mający taką samą formę hasłową) jest znacznie zawyżona. Zastanawia bardzo duża liczba wystąpień bezokolicznika, a ściślej, wystąpień słowa *pleć*, oznaczonych w NKJP jako formy leksemu ***pleć*** (z reguły – rzeczownika, co potraktowaliśmy jako mechaniczną pomyłkę). Rzeczywiście, to opis błędny. Najbardziej typowy przykład otagowania tego słowa wygląda następująco: „Nie **pleć** [pleć:subst:sg:nom:f] bzdur!". Przeczytanie

5 Etykieta (tag) znakowania morfosyntaktycznego z NKJP zawiera następujące informacje gramatyczne: nazwę leksemu (*pielić*), typ fleksemu (*fin* – forma nieprzeszła), wartość kategorii gramatycznej liczby (pojedyncza), wartość kategorii osoby (trzecia), wartość kategorii aspektu (niedokonany). Szczegółowe zasady anotacji w korpusie zostały omówione w Przepiórkowski i in. (red.) 2012.

71 przykładów użycia *pleć* i *Pleć* pozwoliło uznać, że tylko 2 to użycia bezokolicznika (43 to rozkaźnik, a 26 – błędy literowe, zamiast *płeć*). Zatem powyższy stosunek form leksemu ***pielić*** do ***pleć*** wypada skorygować – 198 : 130. Jednak stosunek form jednoznacznie zaliczających się do któregoś leksemu wynosi 78 : 130. Uznalibyśmy więc przewagę form leksemu ***pleć***, ale umiarkowaną. Zastanawia jednak dalej duża liczba wystąpień przypisana leksemowi, znacznie przewyższająca sumę liczb wystąpień poszczególnych form. Tu w naszych obliczeniach nastąpiła niespodzianka, którą można było zauważyć dopiero, sprawdzając napisy wielką literą. Okazało się, że jest jedno wystąpienie rozkaźnika o postaci *Piel* (nazwisko) i 79 o postaci *PIEL* (jest to skrótowy podpis w „Słowie Polskim. Gazecie Wrocławskiej"). I teraz mielibyśmy stosunek ***pielić*** do ***pleć*** – 78 : 50 form jednoznacznie zakwalifikowanych do danego leksemu. Ale jeszcze druga z tych liczb jest za duża. Okazało się bowiem, że pośród 26 wystąpień słowa *Pielę* tylko 3 powinny być zakwalifikowane jako formy czasownika (pozostałe to biernik nazwiska; w okresie gromadzenia korpusu była w Sejmie posłanka Piela); oba wystąpienia słowa *Pielą* też powinny być uznane za formy rzeczownika. I mamy już stosunek ***pielić*** do ***pleć*** – 78 : 25. Jeśli do obu leksemów dodamy formy homonimiczne, stosunek ten będzie wynosić 162 : 109. Przewaga w tekstach ***pielić*** jest zatem niewątpliwa.

Jednak liczbę przypisaną leksemowi ***pielić*** też musimy trochę zmniejszyć: spośród 69 uznanych wystąpień jego form 10 trzeba uznać za wyhasłowane błędnie (to *Pieli*, odpowiadające formom nazwiska ***Piela***), pozostałych 59 to niewątpliwe formy tego leksemu, rozpoczynające się od *pieli*. Zmniejsza to nieco wskazany wyżej stosunek do 68 : 25 (148 : 109 z doliczeniem form homonimicznych). Mamy więc przewagę form ***pielić*** nad ***pleć***.

Podobne oszacowania dla leksemów ***mielić*** i ***mleć*** są trudniejsze. Można na pewno wskazać znaczną liczbę form wspólnych dla obu, które zostały jednoznacznie oznaczone jako należące do leksemu ***mleć***: *mielę* – 7 wystąpień, *mielą* – 88, formy imiesłowów współczesnych – 61, 8 wystąpień słowa *miel* (z których notabene tylko 2 odpowiadają rozkaźnikowi czasownika, a pozostałe to albo pomyłki, albo zapisy tekstów obcojęzycznych), jedno wystąpienie słowa *mielcie* (słowo *mielmy* w badanym korpusie nie wystąpiło), formy bezosobnika, imiesłowu przymiotnikowego biernego (czyli form wyszukiwanych jako *mieleni.**. lub *mielon.**) i odsłownika – 457, czyli ogółem 622 wystąpienia. A więc wystąpień form jednoznacznie należących do ***mleć*** jest 227 (na ogólną liczbę wyszukanych 840). Według wyszukiwarki IPI PAN Poliqarp zrównoważony NKJP ma zawierać 116 form leksemu ***mielić***, a więc połowę tej liczby. Jest to jednak wyraźne niedoszacowanie. Przeczytaliśmy wszystkie przykłady wyhasłowane dla ***mielić*:** jest tam kilka błędów tagowania, przede wszystkim będących skutkiem pomyłek w tekstach źródłowych. Niewątpliwy i zaskakujący błąd w przypisaniu tagu odnosi się do jedynego wystąpienia segmentu *mieli*: „wcielił ich do wojsk francuskich, z którymi odtąd równych zaszczytów, praw i korzyści używać mieli[mielić:fin:sg:ter:imperf]". Jednak jest zupełnie nieprawdopodobne, aby w całym korpusie nie było ani jednego segmentu *mieli* jako wykładnika czasu teraźniejszego. Wiarygodność tego wyniku podważa zarówno stosunek 6 : 9 podstawowych form czasu teraźniejszego *piele* i *pieli* (analogiczne formy *miele* i *mieli* miałyby występować w stosunku 65 : 1 czy nawet 65 : 0 !?), jak i stosunek liczby wystąpień indywidualnych form czasu teraźniejszego od ***mleć*** i ***mielić*** (*+sz*, *+my*, *+cie*) – 7 : 31. Niestety, olbrzymia liczba wystąpień

segmentu *mieli* w badanym korpusie (ponad 1000, regularnie interpretowanych jako formy czasownika ***mieć***) uniemożliwia wyłowienie[6] tych jego wystąpień, które powinny być interpretowane jako [mielić:fin:sg:ter:imperf]. Możemy jednak przyjąć, że liczba wystąpień form czasownika ***mielić*** nie jest mniejsza od liczby wystąpień form czasownika ***mleć***.

Dużo łatwiejsze do zinterpretowania są wyniki dla leksemów utworzonych od rozpatrywanych wyżej przez derywację przedrostkową:

LEKSEM	LICZBA FORM	LEKSEM	LICZBA FORM
zemleć	41	*zmielić*	476
rozemleć	0	*rozmielić*	2
przemleć	83	*przemielić*	28
namleć	0	*namielić*	0
wymleć	1 (wymielony)	*wymielić*	0
umleć	1 (umielą)	*umielić*	0
omleć	1 (omielony)	*omielić*	0
wypleć	21	*wypielić*	14
przepleć	0	*przepielić*	0
opleć	6 (–3)	*opielić*	14 (–10)
obepleć	0	*obpielić*	0
podepleć	0	*podpielić*	0

Liczby podane w nawiasie po znaku minusa odnoszą się do wystąpień odpowiednich słów pisanych wielką literą jako formy nazwiska *Opiela*. Interpretacji szczegółowej wymagają tylko trzy wiersze spośród podanych wyżej.

Tak więc wśród 28 odnotowanych form czasownika ***przemielić*** wszystkie są charakterystyczne, a wśród 83 słów zakwalifikowanych jako formy ***przemleć*** tylko 7 jednoznacznie do niego należy (72 słowa odpowiadają formom bezosobnika, odsłownika i imiesłowu biernego, 3 razy jest *przemielą* i raz *przemiel*): *przemleć* 3, *przemełło* 1, *przemiele* 2, *przemielesz* 1. A więc formy czasownika ***przemielić*** są czterokrotnie częstsze niż ***przemleć***. Dziesięciokrotna przewaga form ***zmielić*** nad ***zemleć*** jest jednak wyolbrzymiona – to skutek wyhasłowania form bezosobnika, odsłownika i imiesłowu biernego pod ***zmielić***, prawdopodobnie dlatego, że w przedrostku w leksemie ***(ze)mleć*** jest wstawne *e*, które nie występuje w tych formach (i we wszystkich formach czasownika ***zmielić***). A to są dokładne odpowiedniki aspektowe ***mielić*** i ***mleć***.

Podobnie jest w odpowiednikach aspektowych ***pielić*** i ***pleć***. Z zanotowanych 21 wyhasłowanych pod ***wypleć*** tylko 6 jest charakterystycznych (*wypiele* 1, *wypleć* 3, *wypełła* 2), a wszystkich 14 słów potraktowanych jako formy ***wypielić*** jest zinterpretowanych poprawnie.

Przystąpmy do analizy użycia poszczególnych form obu serii.

6 Jego częstość w mniejszych korpusach 1 : 100 000 wskazywałaby na to, że w rozważanym korpusie jest ok. 3000 jego wystąpień.

Z punktu widzenia systematyzacji paradygmatu wahania w podzbiorze form A sprawiają najmniej trudności: sprowadzają się one do zamiany samogłoski tematycznej *e* na *i* (oczywiście są też formy, w których występuje neutralizacja opozycji). Wydaje się, że jest tu równowaga. Formy z *-e-* (archaicznego wariantu) występują w tekstach często, na co wpływa też czasem ich związanie frazeologiczne (por. *cielę – ogonem miele*). Formy z *-i-*, charakterystyczne dla grupy VI, były dawniej i są obecnie bez zastrzeżeń akceptowane we współczesnej polszczyźnie.

Znacznie większe trudności występują w formach podzbioru B (tj. w formach bazowych 6-9). Można je znaleźć w internecie, śledząc warianty piosenki dziecięcej *Gdzieżeś ty bywał, czarny baranie?* W odpowiedzi na pytanie: *Cóżeś tam robił, czarny baranie?*, padają różne formy przeszłe omawianego czasownika, najczęściej – akceptowane tradycyjnie *mełł* (*mączkę – mełł mączkę, mościwy panie*), ale także – *meł*, *mleł*, *mlął*, *miął*, *miołł* (w zapisie gwary warmińskiej)[7]. Nie ma wśród nich skądinąd częstej formy *mielił*, mającej jedną sylabę więcej, a więc nie pasującej rytmicznie do wiersza. Do kwestii tej wrócimy niżej.

Zupełnie identyczne są w czasownikach obu serii (***mleć*** i ***mielić***) formy podzbioru C. Jest to jednak jednolitość pozorna. Formy obecnie używane (*mielony*, *mielenie*) są przecież charakterystyczne dla grupy tematowej VI i są w niej panujące. A w czasownikach o podstawach ***mleć*** i ***pleć*** są warianty. Po pierwsze – w tekstach nowopolskich – taki temat, jak w dzisiejszym archaicznym bezokoliczniku, może też wystąpić w formach podzbioru C, które jako imienne nie zajmują pozycji centralnej w całym paradygmacie. W SJPDor w cytowanym już haśle ***mleć*** mamy także: „mielenie (*daw.* także *mlenie*)" z cytatem ze Staszica („Młyny [...] do m l e n i a całego miejskiego zboża"), a w haśle ***rozemleć*** – „rozmielony (*daw.* też: *rozemlony*)" z cytatem z Krasińskiego powtórzonym za SW („na piasek r o z e m l o n e gruzy" – z IV części *Irydiona*). Po drugie, może się w nich pojawić też specjalny alternant: *.ełt.*|*.ełc.*, przed którym – jak widzieliśmy – przestrzega NSPP. W hasłach SJPDor występuje tylko dwa razy: w haśle **pleć** – „pielony (*rzad.* pełty)" i **przepleć** „przepielony (*rzad.* przepełty)", i to bez ilustracji przykładowej. Nie ma go w hasłach: **opleć** i **wypleć**, **mleć**, **namleć**, **zemleć**, **rozemleć**, **przemleć**, **omleć**, **umleć**, **wymleć**.

Tymczasem właśnie ten alternant jest historycznie uzasadniony.

Interesujące nas czasowniki wywodzą się z prasłowiańskiej koniugacji *-e-*/*-o-* tematowej (I grupa wg A. Leskiena, por. Tokarski 2001: 227). W rozważanych rdzeniach czasownikowych było sonantyczne *ļ′*, po którym w formach występował wykładnik fleksyjny, np. **mļ′lъ* > *mełł*, **pļ′lъ* > *pełł* (zob. Stieber 1958: 56). Z tego rdzenia wywodzą się formy podzbioru C, które Aleksander Brückner w swym słowniku etymologicznym wyprowadza następująco:

> **mleć**, *mielę*, *mełł*, *mełty* (wedle *mielę*, *miele* urabiają *mielony*, *zmlej* obok *zmiel*, a dalej i *zmielił*, *zmielić*) (Brückner 1927: 339).
>
> **pleć**: dziś *plewić*, nowa postać, podobna jak *mielić*, p. *mleć*; oba czasowniki są nieco zgodnej odmiany: *pielę*, *pielesz*, *pełł*, *pełty*, chociaż *mleć* z **miel-ti*, ale *pleć* z **plew-ti* poszło (Brückner 1927: 418).

7 W Słowniku warszawskim (SW) jest aż dziewięć form męskich czasu przeszłego: *mieł*, *mlił*, *mełłył*, *mleł*, *mlął*, *myłł*, *mołł*, *meł*, *mluł* (oraz trzy oznaczone jako przestarzałe: *mioł*, *miął*, *mlął*).

Podobnie interpretuje te formy Andrzej Bańkowski, zamieszcza nawet hasła odsyłaczowe do **mełcie**, **mełty**, **pełty**, **pełcie**. W ostatnim jest nota: „też **przepełcie** 1582 (Grzegorz z Żarnowca: *Kościół potrzebuje p r z e p e ł c i a kąkolu i motłochu*), od **przepleć**" (Bańkowski 2000: 525). Ten sam cytat podany jest w haśle ***przepleć*** (***przepielę, przepełty***) (Bańkowski 2000: 888). Jest zresztą w L, SW i SPXVI.

Dziś formy z *.ełt.*|*.ełc.* są niewątpliwie rzadkie, nawet bardzo rzadkie czy wyjątkowe. Uczciwie odnotujmy, że podkorpus zrównoważony Narodowego Korpusu Języka Polskiego notuje tylko dwie formy z tym alternantem:

> Pszenicka n a m e ł t o, moskoliki napiecone, grule w piwnicy, świnka i kurki cekajom noza („Tygodnik Podhalański" 1999).

> Natomiast z młynarską profesją związana jest historia polskiego herbu szlacheckiego Paprzyca, którego wygląd tak został w herbarzu opisany: kamień młyński do góry postawiony, szarego koloru, z żeleźcem w środku, jak do m e ł c i a, w polu białem, na hełmie ośm szczeniąt („Co Tydzień Jaworzno" 2005).

Znalezienie większej liczby poświadczeń form podzbioru C (imiesłowu biernego, odsłownika i bezosobnika) w starszych tekstach nie sprawia jednak większych trudności. W *Słowniku polszczyzny XVI wieku* (SPXVI) znajdujemy bowiem na przykład:

> Z bobu m e ł t e g o mąką wołom często w karmią mieszána/ tuczy ie bárzo rychło (P. Crescentius, *Księgi o gospodarstwie*, wyd. 2).

> W wodźie dwoiáko bywa wárzon/ álbo m e ł t y kámieniem/ álbo w całym źiárniu (P. Crescentius, *Księgi o gospodarstwie*, wyd. 2).

> Dał k temu wolne łowienie [ry]b w stawie i w młynie m e l c i e wolne przez miar (*Lustracje województwa rawskiego*).

> Klucznik około mąki m e ł ć i a ma doyźrzeć (A. Gostomski, *Gospodarstwo*).

Nic dziwnego, że *mełcie* notowane jest w *M. Arcta słowniku staropolskim* (SSA), wydanym w 1920 roku.

Elektroniczny słownik języka polskiego XVII i XVIII wieku (SPXVII) notuje M. i D. liczby pojedynczej: *pełcie*, *pełcia*. Tego typu formy znajdujemy także w tekstach literackich z XIX wieku:

> Zamiast do winnicy, wszedł pomiędzy krzaki i skały; zamiast myśleć o p e ł c i u czy podgartywaniu latorośli, przypomniał sobie dawniejsze „ona czeka", – pośpieszał i nie zatrzymał się, aż na pochyłości przeciwległej, z której mu się odsłonił widok na dolinę dubrowacką (T.T. Jeż, *Lat temu dwieście*).

Twoja i moja niewiasta pozostaną w domu na m e ł c i e językami, ty zaś i ja pójdziemy na robotę (T.T. Jeż, *Słowiański Hercog*).

Muzułmanie za kratę spojrzenia ciekawe zwracali, prawo atoli dozwalało im patrzeć jeno na komin, na którym od rana do wieczora węgle się żarzyły i którego okap ugarnirowany był w tacki, w dżezmy (imbryczki) i filiżanki, w puszki drewniane na kawę m e ł t ą i w butelek kilka, których szyjki przyłożone z wierzchu fajkami glinianymi i okręcone cybuchami skórzanymi znamionowały naczynia, narghilami zwane (T.T. Jeż, *W zaraniu*).

Wyłamywać się z nich jest śmiesznością, w tym młynie co padnie pod kamień nie wstrzyma obrotu a z m e ł t e m zostanie (J.I. Kraszewski, *My i oni*).

Świniarze jedne, nie wiedzą tego, że z m e ł t y słód można i trzy miesiące trzymać! (A. Dygasiński, *Gorzałka*).

Świeże kwiaty wiosenne, starannie o p e ł t e i podlewane, jaśniały, jakby w najweselszym dziewiczym ogródku (J.I. Kraszewski, *Jasełka*).

Zamiast mleć jak dawniej m e ł t o, wzięto się do stawiania sztucznych młynów na wzór zagranicznych, a tymczasem i żytko i mielników czarci porwą! (J. Zacharyasiewicz, *Noc królewska*).

Pszenica też przyjedzie w całości z m e ł t a (G. Zapolska, *Listy z lat 1881–1901*).

Zresztą takie formy znajdujemy także współcześnie (tj. w XX wieku), choć pełny korpus NKJP nie wnosi nic nowego do materiału (nazwisko Zmełty i kilka tekstów kompletnie niespójnych), np.:

Niby pszenicy twarde ziarna,
Mimo że żywa moc w nas drzemie,
P r z e m e ł c i twardym kołem żarna
Schodzimy w niebo, w piekło, w ziemię (J. Iwaszkiewicz, *Lato 1932*).

Cielesny trud oddało się zatem w arendę różnym samobiegom, których można było i przeciw sąsiadom używać, pozostała jednak uciążliwa praca umysłowa, wykoncypowaliśmy więc zmyślny za nas przemysł, jako to myślniczki i myślnice, mielące świat na cyfry, a w z m e ł t y m odcedzające ziarno od plew (S. Lem, *Edukacja Cyfrania*).

W internecie ponadto znaleźliśmy:

Elektryczny młynek do ziarna i maszynka do płatków CombiStar Waldner Biotech ustawić na stabilną poziomą płaszczyznę i nastawić grubość m e ł c i a (zwolnić śrubę i okręcaniem włazu ustawić wymaganą grubość, następnie znów zabezpieczyć śrubą).

Wydajność i wynik m e ł c i a ziaren jest zależny od właściwości mielonego surowca. Bardziej suchy surowiec ma większą wydajność i delikatniejsze m e ł c i e (http://www.atranet.pl/

mlynek_elektryczny_do_zboz_maszynka_do_platkow_waldner_biotech_combi-star,89,83.html, dostęp 20.12.2015).

O pielęgnacji buraków tak pisano: „buraki wymagają w czasie wzrastania, a osobliwie z początku kilkakrotnego p e ł c i a (pielenia), które...[...] (J. Malec, *Burak cukrowy – historia, aktualne problemy, przyszłość*, Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin, Poznań 2004).

Historyczne i dzisiejsze warianty w paradygmacie są skutkiem mocno skomplikowanego rozwoju sonantu *ļ'* w polszczyźnie (zob. Stieber 1958: 56–58). Widać je także w wahaniach form podzbiorów A, a zwłaszcza B. Oprócz zupełnie regularnego pojawienia się samogłoski [*e*] przed [*l*] trzeba zwrócić uwagę na stwardnienie spółgłoski przed sonantem w formach podzbiorów B i C, skutkiem czego jest dzisiejsza wymiana [m]/[m'], [p]/[p'] w paradygmatach odpowiednich leksemów. Spółgłoska wargowa mogła też być twarda w formach podzbioru A, por. cytaty z L:

W tym kraju roli nie orzą, nasienia nie sieją, kąkolu nie p e l ą (Marcin Bielski, *Kronika polska*).

Corkę swoję zszedł August z kobietami, a one p i e l ą jej włosy z siwizny (*Apoftegmata* Plutarcha w tłumaczeniu Bieniasza Budnego).

Oba cytaty podane są dokładnie – por. SPXVI. Słownik ten – w główce hasła **pleć** – zwraca uwagę na oboczność *piel-/pel-*.

Ogromna wariantywność form podzbioru B jest odnotowana w podstawowych słownikach XX-wiecznych. Formę *mleł*, której podstawą jest bezokolicznik, zauważył nawet SJPDor, który podaje w haśle ***mleć***: „mełł (*daw*. także *mleł*)" z cytatem z *Pana Tadeusza*:

Zerwał się mówić, pierwsze słowo niewyraźnie
M l e ł w ustach, aż przez zęby wyleciało: „Błaźnie!" [...] (księga V).

Co ciekawe, w haśle ***pleć*** jest forma analogiczna w cytacie z tego samego źródła, ale nie wybita w główce:

[On jeszcze! ileż razy na parkanie siadał,
By ją dójrzeć przez okna;] w konopie się wkradał,
Żeby patrzeć, jak ona p l e ł a swe ogródki,
Rwała ogórki albo karmiła kogutki (księga XII).

Taka forma regionalna występuje też między innymi u Józefa I. Kraszewskiego w postaci *mléł*. Było to więc wymawiane węziej, co można było również zapisać *mlił*, jak zanotował Samuel B. Linde formę przyhasłową w haśle ***mleć*** (L). Znajdująca się w tym słowniku egzemplifikacja

pokazuje również wahania samogłoskowe wewnątrz rdzenia *e* : *o*, obecne w XVI-wiecznej polszczyźnie, np. „Samson w łańcuchach m i o ł w ciemnicy" (*Biblia* w tłumaczeniu B. Budnego, Księga Sędziów 16,21; 1572). Tu cytat jest niedokładny, ale słowo ***mioł*** zostało zapisane dobrze (podobnie w SPXVI).

Możemy zatem podsumować.

Omawiana grupa czasowników od pradawnych czasów jest nieustabilizowana: wahania w odmianie widać i w staropolszczyźnie, i w innych językach słowiańskich. Ograniczyliśmy się jednak do obserwacji tekstów polskich. Prowadziła ona do wyraźnych wniosków.

Od długiego czasu omawiane czasowniki (o podstawach *mleć*, *pleć*), nietypowe morfologicznie, znajdują się pod silnym wpływem regularnej koniugacji według grupy VI (o podstawach *mielić*, *pielić*[8]). Formy tworzone według tego wzorca wchodziły do polskiego systemu stopniowo. Początkowo zmianie uległy formy podzbioru C (z *mełty*, *mełcie* na *mielony*, *mielenie*). Oddziaływanie było na tyle silne, że formy nowe uznano za jedynie poprawne i wprowadzono do paradygmatów podawanych w gramatykach i słownikach – obok tradycyjnych form nieregularnych podzbiorów A, B oraz bezokolicznika. Otrzymano w konsekwencji paradygmat hybrydalny.

Dziś używane są tradycyjne formy z podzbioru A, np. *miele*, *piele*, choć wyraźnie rzadziej[9]. Ale formy nowe typu *mieli* (z grupy VI), które przed 50 laty wywoływały negatywną reakcję w gronie specjalistów od kultury języka polskiego, są dziś co najmniej tak samo popularne, jeśli nie częstsze (por. dane korpusowe).

Natomiast użycie form podzbioru B jest mocno ograniczone (mówi się *mielił* zamiast *mełł*). Chociaż wykształceni Polacy znają formy przeszłe typu *mełł*, *pełł* ze szkoły, nie używają ich spontanicznie. Wiedzą, że należy tak mówić, ale tak nie mówią. Formy przeszłe trafiają się, i to w sporej liczbie, w materiałach pisanych, lecz w żywej mowie zostały one całkiem wyparte przez regularne formy z przyrostkiem tematowym *-i-* (tak jak dawniej przestały być używane formy z podzbioru C[10]).

Teoretycy polskiej kultury języka powinni przyjąć do wiadomości taki stan rzeczy[11].

8 Oprócz omawianego w artykule mieszania systematycznego trafia się też używanie czasowników o podobieństwie dalszym. W polszczyźnie XVII wieku **pleć** było wypierane przez regularnie odmieniany (według grupy VI) czasownik **plewić**, co widać już w tytule głównego zbioru fraszek Wacława Potockiego: *Ogród, ale nie plewiony* oraz przywołanym cytacie ze słownika A. Brücknera.

9 Potwierdzenie tych tez znajdujemy także w danych współczesnej polszczyzny mówionej – w odpowiedniej bazie cyfrowej nie odnaleźliśmy form czasowników o podstawach **pleć** i **mleć** (por. http://spokes.clarin-pl.eu/ oraz Pęzik 2015).

10 Dziś znane są chyba tylko entuzjastom gier słownych (zob. np. http://test.scrabblemania.pl/). W wyszukiwarkach stron internetowych zawierających zbiór słów używanych w Scrabble za prawidłowe uznano następujące alternanty *.ełt.*|*.ełc.*: *mełty – mełci*, *nametły – namełci*, *przemełty – przemełci*, *wymełty – wymełci* oraz *pełty – pełci*, *opełty – opełci*, *przepełty – przepełci*, *wypełty – wypełci*.

11 Z satysfakcją – już po napisaniu tego tekstu – odnotowaliśmy, że nastąpiła wyraźna zmiana w ocenie użycia form omawianych czasowników. Postulowaną przez nas wskazówkę podał udostępniony niedawno w internecie *Dobry słownik*, początkowo tylko w hasłach *mielić* i pochodnych (hasła dla *pielić* i pochodnych zostały wprowadzone – po naszym zapytaniu – w maju 2016 r.). Niestety, dostęp do tego słownika jest ograniczony, wymagana słona opłata, także do użytku w celach badawczych. Dziękujemy p. Aleksandrowi Siwickiemu, który – opłaciwszy abonament trzymiesięczny – przeprowadził również potrzebne wyszukiwania.

Przyjęli go do wiadomości autorzy SGJP i podali dla wszystkich czasowników o podstawach **mleć** i **pleć** paralelne leksemy o podstawach **mielić** i **pielić**. Dla **mleć** i **pleć** w III, internetowym, wydaniu tego słownika zanotowano odmianę konsekwentną z *mełto* i *mełcie*, opatrując jednak odpowiednie leksemy kwalifikatorem *przest.*, wskazującym, że życie ich form jest dziś wyraźnie ograniczone, nacechowane; neutralne, szeroko używane są formy od **mielić** i **pielić**. Formy typu *mełto*, a także pochodne leksemy, np. ***mełty***, ***mełcie***, mają kwalifikator silniejszy: *daw.*

Wykaz źródeł wraz ze stosowanymi w pracy skrótami

Bańkowski A. 2000: *Etymologiczny słownik języka polskiego*, t. 2, Wydawnictwo Naukowe PWN, Warszawa.

Brückner A. 1927: *Słownik etymologiczny języka polskiego*, Krakowska Spółka Wydawnicza, Kraków.

Czesak A., Szałkiewicz Ł., Żurowski S., *Dobry słownik* (online: www.dobryslownik.pl).

L: S.B. Linde, *Słownik języka polskiego*, wydanie drugie poprawne i pomnożone, t. 1–6, Zakład Narodowy im. Ossolińskich, Lwów 1854–1860.

NSPP: *Nowy słownik poprawnej polszczyzny PWN*, red. A. Markowski, Wydawnictwo Naukowe PWN, Warszawa 1999.

SJPDor: *Słownik języka polskiego*, red. W. Doroszewski, t. 1–4, Wiedza Powszechna, Warszawa 1958–1962, t. 5–11, Państwowe Wydawnictwo Naukowe, Warszawa 1963–1969.

SGJP: Z. Saloni, M. Woliński, R. Wołosz, W. Gruszczyński, D. Skowrońska, *Słownik gramatyczny języka polskiego*, wyd. 3 (online www.sgjp.pl).

SPXVI: *Słownik polszczyzny XVI wieku*, red. M.R. Mayenowa (t. 1–34), K. Mrowcewicz (t. 35–36), t. 1–22, Zakład Narodowy im. Ossolińskich, Wydawnictwo Polskiej Akademii Nauk, Wrocław–Warszawa–Kraków 1966–1994, t. 23–36, Instytut Badań Literackich PAN, Warszawa 1995–2012.

SPXVII: *Elektroniczny słownik języka polskiego XVII i XVIII wieku*, red. W. Gruszczyński (online: http://sxvii.pl).

SSA: *M. Arcta słownik staropolski*, oprac. A. Krasnowolski, W. Niedźwiedzki, Wydawnictwo M. Arcta, Warszawa 1920.

SW: *Słownik języka polskiego*, red. J. Karłowicz, A. Kryński, W. Niedźwiedzki, t. 1–8, nakładem prenumeratorów i Kasy im. Mianowskiego, Warszawa 1900–1927 (Słownik warszawski).

WSPP: *Wielki słownik poprawnej polszczyzny PWN*, red. A. Markowski. Wydawnictwo Naukowe PWN, Warszawa 2004.

Bibliografia

Pęzik P. 2015: *Spokes – a search and exploration service for conversational corpus data*, [w:] *Selected Papers from the CLARIN 2014 Conference*, October 24–25, 2014, Soesterberg, The Netherlands, 99–109. Linköping Electronic Conference Proceedings. Linköping University Electronic Press, Linköpings Universitet.

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

Saloni Z. 2000: *Wstęp do koniugacji polskiej*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn.

Stieber Z. 1958: *Rozwój fonologiczny języka polskiego*, Państwowe Wydawnictwo Naukowe, Warszawa.

Tokarski J. 1951: *Czasowniki polskie*, Wydawnictwo S. Arcta, Warszawa.

Tokarski J. 1958: *Czasownik*, [w:] *Słownik języka polskiego*, red. W. Doroszewski, t. 1, Państwowe Wydawnictwo „Wiedza Powszechna", Warszawa, s. LXIII–LXXXIV.

Tokarski J. 2001: *Fleksja polska*, wyd. 2 uzupełnione, Wydawnictwo Naukowe PWN, Warszawa.

Summary

On the inflection of the verbs with the stems *mleć* and *pleć*

Keywords: verbs, inflection, paradigm.

The article discusses the contemporary situation of the lexical variants of two Polish verbs: *mleć* : *mielić*, *pleć* : *pielić* (and their prefixed derivatives). As a rule Polish dictionaries give the hybrid paradigms with the majority forms built from the traditional variants *mleć* and *pleć*, but the participle and the gerund are taken from the new variants *mielić* and *pielić*, belonging to the productive conjugational pattern (*mielony*, *mielenie*; *pielony*, *pielenie* instead of the old forms *mlety*, *mlecie*; *plety*, *plecie*, now completely obsolete). In Polish text corpora the old variants are seemingly prevalent, but this is the result of automatic tagging, favoring the variants accepted by traditional normative dictionaries. In fact, new variants are more frequent. In conclusion, the authors propose to divide the paradigms according to their history and to recognize *mielić* and *pielić* as neutral variants. This solution has been introduced in the new on-line edition of the *Grammatical Dictionary of Polish*.

MARCIN KUCZOK* | UNIWERSYTET ŚLĄSKI, KATOWICE

Rodzaje motywacji metonimicznej eufemizmów polskich

Słowa kluczowe: eufemizm, metonimia formalna, metonimia referencyjna, metonimia zdaniowa, metonimia illokucyjna.

1. Wstęp

Celem niniejszego artykułu jest przedstawienie czterech różnych rodzajów metonimii wpływających na proces tworzenia eufemizmów w języku polskim. Dynamiczny rozwój językoznawstwa kognitywnego w ostatnich dekadach, zwłaszcza wśród językoznawców z krajów anglosaskich, doprowadził do zwiększonego zainteresowania zjawiskiem metonimii, którą postrzega się jako istotne narzędzie poznawcze, motywujące nasz sposób używania języka. Jak pokazują badania, metonimia jest narzędziem często wykorzystywanym przez użytkowników języka przy tworzeniu eufemizmów (Engelking 1984; Allan, Burridge 1991, 2006; Dąbrowska 1992, 1994; Bierwiaczonek 2013; Duda 2014; Kuczok 2015). Niemniej jednak brakuje w literaturze przedmiotu pogłębionej analizy eufemizmów polskich, dokonywanej z perspektywy paradygmatu lingwistyki kognitywnej.

W pierwszej części zostanie przedstawione pojęcie eufemizmu z uwzględnieniem jego charakteru kulturowego. Kolejna sekcja poświęcona będzie zagadnieniu metonimii obecnej w języku, z uwzględnieniem rodzajów metonimii wyliczanych przez Bogusława Bierwiaczonka (2013). Następnie zostaną podane przykłady eufemizmów, w których można dopatrzyć się kolejno metonimii formalnej, referencyjnej, zdaniowej oraz illokucyjnej. Przykłady eufemizmów podawane w niniejszym opracowaniu pochodzą w większości z trzeciego wydania *Słownika eufemizmów polskich* (SEP) Anny Dąbrowskiej, z monografii *Eufemizmy współczesnego języka polskiego* (1994) oraz artykułu *Eufemizmy mowy potocznej* (1992) tejże autorki, a także sporadycznie z Narodowego Korpusu Języka Polskiego (NKJP) dostępnego w zasobach Internetu.

2. Eufemizm jako zjawisko językowo-kulturowe

Słowo *eufemizm* pochodzi od greckiego słowa *euphemia* (εὐφημία), oznaczającego tyle, co ‘użycie słów jako dobrej wróżby’, które z kolei jest złożeniem *eu* (εὖ), czyli ‘dobry’, oraz *pheme* (φήμη), czyli ‘pochwała, pochlebstwo’ (Dąbrowska 1994: 33; Allan, Burridge 2006: 29). Jak zaznacza A. Dąbrowska (1994: 51), eufemizmy są ściśle związane z tabu, którego obecność w kulturze stanowi motywację i uzasadnienie tworzenia eufemizmów. Ich podstawową funkcją jest łagodzenie negatywnej konotacji słów naznaczonych tabu. Anna Engelking (1984: 127) z kolei zwraca uwagę na to, że eufemizmy winny cechować się świeżością skojarzeniową, która wyklucza używanie form utartych i zleksykalizowanych, a także ekspresywnością emocjonalną,

* marcin.kuczok@us.edu.pl

która może być podniosła i uroczysta bądź żartobliwa, humorystyczna lub nawet pogardliwa wobec denotatum.

Eufemizmy są ściśle związane z zasadami uprzejmości w komunikacji. Zasady te, według Penelope Brown i Stephena Levinsona (1987: 61), opierają się na pojęciu *twarzy*, rozumianej jako 'publiczny obraz samego siebie'. Uprzejmość polega na łagodzeniu takich sposobów wypowiadania się, które zagrażają twarzy osoby mówiącego lub też słuchacza. Na podstawie takiego założenia eufemizm można zdefiniować jako alternatywę dla niepożądanego wyrażenia stosowaną w celu uniknięcia możliwej utraty twarzy mówiącego, słuchaczy lub osób trzecich (Allan, Burridge 1991: 11). Używamy zatem eufemizmów, by, po pierwsze, uniknąć negatywnego odbioru siebie przez innych ludzi, po drugie, by móc wyrażać swoje myśli bez ograniczeń, i po trzecie, by zachować stosowną delikatność oraz nie obrażać swoich rozmówców (Allan, Burridge 2006: 29–31).

Potrzeba zwracania uwagi na uprzejmość w komunikacji i – co za tym idzie – stosowania eufemizmów wyrasta ze zjawiska kulturowego określanego jako *tabu*. Samo słowo *tabu* pochodzi z języka tonga, w którym *tapu* lub *tabu* oznacza zachowanie lub czynność zakazane, niedozwolone w kulturze rdzennych mieszkańców Polinezji. Słowo to Europa zawdzięcza podróżom kapitana Jamesa Cooka (1728–1779), angielskiego żeglarza i odkrywcy, który podczas swych wypraw odwiedził Polinezję, w tym wyspy Tonga (Dąbrowska 1994: 14–15; Allan, Burridge 2006: 2–4).

Według A. Dąbrowskiej, należącej do nielicznej grupy językoznawców zajmujących się tematyką eufemizmów we współczesnym języku polskim, tabu w polszczyźnie obejmuje następujące sfery życia:

1) wierzenia religijne, w tym nazwy Boga, Matki Bożej i diabła,
2) nazwy niebezpiecznych zwierząt,
3) choroby, w tym śmiertelne, trudno uleczalne i weneryczne,
4) śmierć i zjawiska z nią związane,
5) wady, nałogi i przywary ludzkie, w tym defekty moralne i umysłowe,
6) cechy fizyczne człowieka, w tym wiek, brzydotę, łysinę, chudość, grubość, zbyt niski lub wysoki wzrost, chwilowo niekorzystny wygląd, kalectwo i brud,
7) status finansowy i nazwy pieniędzy,
8) etykietę językową, w tym formy zwracania się do drugich, oznaczanie *ja*, tryb rozkazujący i negację,
9) przewinienia, przekroczenia i kary,
10) połajanki, przekleństwa i wyzwiska,
11) nazwy części ciała, w tym pośladki, piersi kobiece oraz narządy płciowe męskie i żeńskie,
12) nazwy części garderoby,
13) nagość,
14) czynności fizjologiczne, w tym przemianę materii, zapachy ciała i stany fizjologiczne kobiety,
15) życie seksualne,

16) politykę i dyplomację, w tym wojnę i okupację, problemy związane z rasą oraz gospodarką (Dąbrowska 1994: 69–72).

Jak zwraca uwagę ta sama autorka, tabu ma charakter ambiwalentny. Z jednej strony może bowiem odnosić się do zjawisk świętych, do sfery religijnej życia człowieka, a z drugiej obejmuje to, co uważane jest za nieczyste, nieprzyzwoite i budzące odrazę (Dąbrowska 1994: 16–17). O tym, jak bardzo trudnym i delikatnym zarazem tematem jest omawianie eufemizmów, może świadczyć fakt, że – jak zauważa A. Dąbrowska we wstępie do swojego *Słownika eufemizmów polskich* (SEP: 12–13) – także w różnego typu pracach językoznawczych i leksykograficznych unika się podawania wprost verbum proprium, a więc słowa objętego tabu, które dany eufemizm oznacza. Być może wynika to z obawy, że nawet w tekście naukowym, traktującym, wydawałoby się, omawiane zagadnienie w sposób poważny i obiektywny, mogą one zniesmaczyć czy wręcz urazić czytelnika. Pomimo to w niniejszym artykule zdecydowano się podawać verbum proprium każdego z analizowanych eufemizmów wprost, o ile to tylko jest możliwe i uzasadnione, z założeniem, że tego wymaga od nas uczciwe przedstawienie analizowanego materiału językowego.

3. Rodzaje metonimii w języku

Jak pisze A. Dąbrowska, metonimia, nazywana również zamiennią, tradycyjnie w retoryce rozumiana była przede wszystkim jako używanie słowa pozostającego w realnym związku skojarzeniowym z zamierzoną treścią znaczeniową w miejsce verbum proprium (Dąbrowska 1994: 333). Wzrost zainteresowania zjawiskiem metonimii pojawił się pod koniec XX wieku wraz z narodzinami i rozpowszechnieniem się paradygmatu językoznawstwa kognitywnego, które kładzie nacisk na związek form językowych z procesami poznawczymi oraz mechanizmami pojęciowymi zachodzącymi w umyśle człowieka. Również takie zjawiska jak metafora czy właśnie metonimia mają według kognitywistów przede wszystkich charakter pojęciowy, co więcej, są raczej kwestią spontanicznego sposobu myślenia aniżeli samej tylko manipulacji strukturami języka, na przykład w celach retorycznych lub estetycznych. Według ogólnej definicji zaproponowanej przez George'a Lakoffa i Marka Johnsona (1988: 58) z *metonimią pojęciową* mamy do czynienia wtedy, gdy używamy „pewnej rzeczy, realnie istniejącej, by mówić o innej rzeczy, która jest z nią związana". Elżbieta Tabakowska z kolei określa metonimię pojęciową jako zjawisko, „dzięki któremu znaczenie podstawowe słowa służy do określania części elementów pozostających w jakimś związku z desygnatem lub na odwrót – cały desygnat określany jest przez część elementów z nim związanych" (Tabakowska (red.) 2001: 54). Innymi słowy, relacje semantyczne w metonimii oparte są na związku przyległości, która polega na styczności określonych elementów. Przykładowo, zdanie *Wypił całą butelkę* oznacza, że ktoś wypił całą *zawartość* tej butelki.

Vyvyan Evans (2009: 74–75) ujmuje metonimię pojęciową jako operację, w której należy wyróżnić element *nośnika* oraz element *tematu*. W podanym powyżej przykładzie *butelka* będzie stanowić nośnik, natomiast *zawartość butelki* to temat, pomiędzy nimi dochodzi zaś do rzutowania typu POJEMNIK ZA ZAWARTOŚĆ. Należy także zwrócić uwagę, że w metonimii

rzutowanie zawsze ma miejsce w obrębie tej samej domeny pojęciowej, w odróżnieniu od pokrewnego mechanizmu *metafory pojęciowej*, w której wypadku dochodzi do rzutowania pomiędzy dwiema różnymi domenami. Inne przykłady metonimii to między innymi PRODUCENT ZA PRODUKT (*Podaj mi Szekspira z tamtej półki* w znaczeniu 'podaj mi książkę Szekspira z tamtej półki'), CZĘŚĆ ZA CAŁOŚĆ (*Podaj pomocną dłoń* w znaczeniu 'pomóż mi'), CAŁOŚĆ ZA CZĘŚĆ (*Popsuł mi się samochód* w znaczeniu 'popsuł się silnik samochodu') czy też SKUTEK ZA PRZYCZYNĘ (*Jej twarz rozpromieniła się* w znaczeniu 'rozradowała się').

Korzystając z propozycji językoznawców kognitywnych, głównie angielsko- i polskojęzycznych, B. Bierwiaczonek (2013: 27) wyróżnia cztery rodzaje metonimii obecnej w języku: metonimię formalną, referencyjną, zdaniową i illokucyjną.

1) *Metonimia formalna* jest oparta na formalnych relacjach pomiędzy nośnikiem a tematem. O metonimii na poziomie form językowych pisze również hiszpański anglista Antonio Barcelona (2005), dopatrując się takiej metonimii w rzutowaniach typu ISTOTNA CZĘŚĆ FORMY ZA CAŁĄ FORMĘ. Przykładami występowania metonimii formalnej w polszczyźnie mogą być skrótowce (Powszechny Zakład Ubezpieczeń → PZU) i ucięcia (*specjalista* → *spec*). B. Bierwiaczonek podaje także przykłady motywowane metonimią na poziomie zestawień: *schabowy* zamiast *kotlet schabowy*, *pomidorowa* zamiast *zupa pomidorowa*, *skrzydłowy* zamiast *gracz skrzydłowy*, oraz określeń: *dzielnicowy* zamiast *policjant dzielnicowy*, *zdrowy* zamiast *zdrowy człowiek* czy też *pijany* zamiast *pijany mężczyzna* (Bierwiaczonek 2013: 77). W sposobie konstruowania zdań metonimia formalna widoczna jest na przykład w pomijaniu dopełnienia niektórych czasowników, takich jak *pić*, *jeść*, *palić*, *prać*, *prowadzić*, ale też w zdaniach eliptycznych typu *Ona już nie bierze* (narkotyków) lub *Przepraszam, czy tu biją* (zwykłych ludzi)*?*

2) *Metonimia referencyjna* zachodzi, gdy pewna rzecz, osoba lub zjawisko odnosi się do innej rzeczy, osoby lub zjawiska. Klasyczne przykłady to chociażby przedstawione powyżej rzutowania typu CZĘŚĆ ZA CAŁOŚĆ, CAŁOŚĆ ZA CZĘŚĆ, PRODUCENT ZA PRODUKT czy też SKUTEK ZA PRZYCZYNĘ.

3) *Metonimia zdaniowa* to metonimia, w której dane wypowiedzenie jest używane zamiast innego wypowiedzenia w obrębie tego samego wyidealizowanego modelu kognitywnego, to znaczy umysłowej reprezentacji teorii na temat jakiegoś aspektu rzeczywistości, na której tle funkcjonują jednostki języka (Evans 2009: 178). Metonimia zdaniowa może odnosić się do całego zdania, wraz z podmiotem i grupą orzeczenia, lub tylko do samego orzeczenia. Na przykład udzielając odpowiedzi na pytanie: *Jak spędziliście weekend?* w formie: *Marysia dostała darmowe bilety do kina*, posługujemy się metonimią, w której spędzanie całego weekendu jest konceptualizowane i opisywane przez inne zdanie, odnoszące się do jednego centralnego wydarzenia całego weekendu, jakim było wyjście do kina. Oba te zdania dotyczą tego samego wyidealizowanego modelu kognitywnego, który moglibyśmy określić jako *spędzanie weekendu*. Z kolei stwierdzając: *Dała radę ukończyć studia*, tak naprawdę mamy na myśli tylko inne orzeczenie, mianowicie: *Skończyła studia*, podczas gdy podmiot zdania pozostaje ten sam. Tutaj również oba zdania dotyczą tego samego wyidealizowanego modelu kognitywnego – tym razem jest nim *ukończenie studiów*.

4) *Metonimia illokucyjna* oznacza, że jeden illokucyjny akt mowy użyty jest zamiast innego illokucyjnego aktu mowy, ale oba akty składają się na ten sam scenariusz aktu mowy. Według

teorii aktów mowy Johna L. Austina *akty illokucyjne* zawierają się w tym, co chcemy osiągnąć przez wypowiedzenie danego wyrażenia – może to być przykładowo prośba, rozkaz, życzenie lub oznajmienie czegoś (Austin 1993: 640–691). Niekiedy jednak stosujemy tak zwane *pośrednie akty mowy*, o których pisał John Searle (1987) – ma to miejsce wówczas, gdy zamierzona moc illokucyjna różni się od mocy illokucyjnej formy systemowej wyrażenia użytego przez mówiącego i może zostać odczytana tylko w kontekście danej sytuacji (Grzegorczykowa 2001: 36–37, 78; Levinson 2010: 306–321). Na przykład życzenie wyrażone w zdaniu (1a) w zasadzie wyraża prośbę, która została sformułowana wprost w zdaniu (1b).

(1a) Chciałbym, żebyś zamknął okno.
(1b) Proszę, zamknij okno.

W ujęciu omawianej klasyfikacji metonimii pośrednie akty mowy można zatem alternatywnie opisać jako wypowiedzenia motywowane metonimicznie. Ze względu na to, że mamy tutaj do czynienia z rzutowaniem metonimicznym pomiędzy różnymi aktami illokucyjnymi, mówimy o *metonimii illokucyjnej*.

W kolejnych częściach niniejszego artykułu dokonamy przeglądu eufemizmów polskich na podstawie motywującego ich powstanie rodzaju metonimii, kolejno: formalnej, referencyjnej, zdaniowej i illokucyjnej. Ze względu na obszerne rozmiary analizowanego materiału w poszczególnych sekcjach zaprezentowane zostaną jedynie wybrane przykłady eufemizmów.

4. Eufemizmy motywowane metonimią formalną

Pierwsza grupa eufemizmów, motywowanych metonimią formalną, zasadza się na różnego rodzaju skróceniach denotatum. Należą do nich ucięcia wyrazów. Przykładowo, *O kur...* stanowi eufemistyczne ucięcie nacechowanego negatywnie słowa *kurwa* (SEP: 216), a *hera* i *koka* są używane zamiast nazw *heroina* i *kokaina* (SEP: 174). Przykładem zastosowania literowca jako eufemizmu może być określenie *choroba w*, które metonimicznie zastępuje wyrażenie *choroba weneryczna* (SEP: 129). Z kolei użycie takich form jak *CKM* zamiast wyrażenia *ciężko kapująca mózgownica* (SEP: 159) oraz *opeer* (*OPR*) w miejsce wyrazu *opierdol* (SEP: 224) ilustruje metonimię formalną, funkcjonującą w akronimach. Metonimii formalnej można także dopatrzyć się w eufemizmach stworzonych dzięki mechanizmowi elipsy. Na przykład w wykrzyknieniach *O rany!*, które jest skróconą wersją *O rany boskie!* (SEP: 234), lub *Do jasnej!*, które zastępuje *Do jasnej cholery!* (Dąbrowska 1994: 185).

Kolejna grupa eufemizmów, w których można dopatrzyć się motywacji metonimicznej, to słowa i wyrażenia powstałe w wyniku manipulacji fonetycznej. Eufemizmy tworzone w ten sposób B. Bierwiaczonek (2013: 70) klasyfikuje jako przykłady ilustrujące metonimię formalną. Na przykład wyrazy *furwa*, *kurna*, *kuźwa*, *kurde*, *kuchnia*, *kurcze* stanowią zmodyfikowane fonetycznie formy słowa *kurwa* bądź to przez wymianę pierwszej głoski (*furwa*), bądź też zmianę drugiej sylaby wyrazu *kurwa* (SEP: 216–218). Podobnie jest z wyrazami *jejku*, *jeżu*, *jeny*, *jerum*, które pochodzą od wykrzyknienia *Jezu!* i zachowują jego pierwszą sylabę *je-* (SEP: 233–234). Formą wykorzystania brzmienia negatywnie nacechowanego związku frazeologicznego jest tworzenie rymujących się wyrażeń, często o zabawnym charakterze, na

przykład eufemistyczne wyrażenie *do jasnej-ciasnej*, używane zamiast wykrzyknienia *Do jasnej cholery!* (SEP: 229). Można tutaj dodać także zjawisko gry półsłówek, polegające na przestawieniu sylab, jak w przykładzie *słynna praczka*, który wziął się z wyrażenia *płynna sraczka* i oznacza 'rozwolnienie' (SEP: 69).

W innej grupie eufemizmów dochodzi do całkowitego zastąpienia nacechowanego negatywnie wyrazu innym słowem. W zjawisku *substytucji* mogą to być słowa zupełnie niezwiązane ani semantycznie, ani formalnie z verbum proprium eufemizmu. Na przykład wyrażenie *rany Julek* funkcjonuje jako eufemizm wyrażenia *rany boskie* (Dąbrowska 1994: 282), a zwrot *pocałuj mnie w gębę* zastępuje *pocałuj mnie w dupę* (SEP: 42). Niekiedy stosuje się także zapożyczenia z innych języków, które choć oznaczają to samo co nacechowany negatywnie polski wyraz lub zwrot, formalnie się różnią i mogą pełnić funkcję eufemizującą dzięki zamaskowaniu swojej negatywnej konotacji. Przykładem może być wykorzystanie w miejsce polskiego wyrazu *gówno* (SEP: 231) angielskiego wykrzyknienia *Shit!*, które u polskiego odbiorcy nie musi wywoływać pejoratywnych skojarzeń. Szczególnymi zapożyczeniami używanymi w roli eufemizmów są profesjonalizmy, bardzo często pochodzące z łaciny, które są stosowane na przykład w dyskursie medycznym do nazywania naznaczonych tabu zjawisk fizjologicznych i chorób. Ilustracją tego zjawiska może być użycie słowa *defekacja* na wypróżnianie się lub określanie wymiotów wyrażeniem *torsje* (Dąbrowska 1992: 156). Z zapożyczeniami wiąże się także inna strategia tworzenia eufemizmów, oparta na celowej stylizacji brzmienia polskich wyrazów nacechowanych negatywnie na wyrazy obcojęzyczne, jak na przykład w wypadku wyrazu *djupa*, który jest stylizacją fonetyczną na język angielski polskiego słowa *dupa* (Dąbrowska 1994: 195).

Innym mechanizmem, który można powiązać z metonimią formalną, jest używanie zaimków wskazujących, takich jak *to* lub *tamto*, które mogą odnosić się do wielu różnych rzeczy objętych tabu w danym kontekście komunikacyjnym, na przykład do choroby nowotworowej, morderstwa, śmierci, płodów ludzkich po usunięciu ciąży, stosunku seksualnego albo do korzystania z toalety (SEP: 67, 84, 102, 132, 143, 151). Również zaimki osobowe zastępują słowa nieprzyzwoite, na przykład w odniesieniu do męskiego członka czasem stosuje się zaimki *on*, *go*, *jego*, *niego* (SEP: 54). Te same zaimki mogą odnosić się do innych rzeczy lub osób objętych tabu, na przykład do diabła (SEP: 236). Ostatnią strategią tworzenia eufemizmów motywowaną metonimią formalną jest stosowanie zdrobnień – zastąpienie wyrazu nacechowanego negatywnie formą zdrobniałą łagodzi jego pejoratywną i obraźliwą konotację. Przykładem może być użycie słowa *głuptasek* zamiast *głupiec* lub *głuptas* (SEP: 160) bądź słowa *pieniążki* zamiast *pieniądze* (SEP: 241). Niekiedy w tworzeniu eufemizmów mamy do czynienia z jednoczesnym zastosowaniem kilku mechanizmów językowych opartych na metonimii formalnej. Dla przykładu określenie męskich jąder słowem *dżonderka* jest zarówno zdrobnieniem, jak i stylizacją w wymowie na język angielski właściwej formy zdrobniałej w języku polskim, która winna brzmieć *jąderka* (SEP: 49).

5. Eufemizmy motywowane metonimią referencyjną

W analizowanych przykładach eufemizmów można dopatrzyć się wielu różnych rzutowań z rodzaju metonimii referencyjnej. Po pierwsze, domenę źródłową dla metonimii stanowią

odniesienia do ludzi jako użytkowników pewnych rzeczy, posiadaczy lub wykonawców pewnych czynności. Eufemizmy takie jak *damskie części* na oznaczenie żeńskich narządów płciowych (Dąbrowska 1992: 140) lub *kobieca przypadłość*, oznaczająca menstruację (SEP: 79), są motywowane rzutowaniem typu POSIADACZ ZA WŁASNOŚĆ. Z kolei wyrażenie *strony dla dorosłych*, użyte na określenie internetowych stron pornograficznych (NKJP), kryje w sobie metonimię typu UŻYTKOWNIK ZA UŻYTKOWANY PRZEDMIOT. Natomiast w wyrażeniu *mężowski obowiązek*, oznaczającym stosunek płciowy (SEP: 94), można dopatrzyć się motywacji metonimicznej opartej na rzutowaniu typu WYKONAWCA CZYNNOŚCI ZA CZYNNOŚĆ.

Inna metonimia referencyjna motywująca tworzenie eufemizmów w języku polskim to CECHA CHARAKTERYSTYCZNA ZA RZECZ LUB OSOBĘ, która może zostać zilustrowana określeniami *zły* lub *czarny* oznaczającymi szatana (SEP: 235–237) bądź też wyrażeniem *błękitni chłopcy*, użytym w odniesieniu do policjantów (Dąbrowska 1994: 173). Oprócz tego funkcja lub rola może odnosić się metonimicznie do osoby lub rzeczy, na przykład w eufemizmach *Pan władza* na określenie policjanta (SEP: 209) lub *kusiciel* jako nazwa diabła (SEP: 235). Z kolei w innych eufemizmach źródłem metonimii mogą być pewne właściwości danej rzeczy, na przykład materiał użyty do produkcji, jak w eufemizmach *guma* lub *gumka*, oznaczających prezerwatywę (SEP: 81). Co więcej, także nazwa producenta lub marki produktu bywa domeną źródłową dla metonimii referencyjnej w eufemizmach, na przykład w wyrazie *Leszek*, który jest zdrobnieniem od słowa *Lech*, nazwy jednej z polskich marek piwa (SEP: 180), lub *wyborowa*, które to słowo jest jednocześnie nazwą jednej z marek wódki produkowanej w Polsce (SEP: 183). Warto też wspomnieć takie wyrażenia jak *ćwiartka* lub *połówka*, które mogą być eufemizmami wódki (SEP: 178, 182), zgodnie z rzutowaniem TYPOWA ILOŚĆ LUB OBJĘTOŚĆ PRODUKTU ZA PRODUKT.

Bywają w języku polskim eufemizmy, w których można dopatrzyć się motywacji metonimicznej opartej na rzutowaniu typu KONKRETNY ZA OGÓLNY. Przykładem może być wyrażenie *domek z serduszkiem* na określenie ubikacji (SEP: 75). Są także takie eufemizmy, które wydają się opierać na przeciwnym rzutowaniu metonimicznym: OGÓLNY ZA KONKRETNY, jak w wyrazie *narządy*, stosowanym na oznaczenie narządów płciowych (SEP: 47), lub w wyrażeniu *brzydkie słowo*, używanym w odniesieniu do przekleństwa (SEP: 212). Nieco inną metonimię referencyjną, choć również polegającą na swego rodzaju rozszerzaniu znaczenia eufemizowanego wypowiedzenia, opartą jednak tym razem na rzutowaniu typu WIĘCEJ ZA MNIEJ, można zaobserwować w zawołaniach typu *Panie kierowniku!*, *Szefie!*, *Szefowo!*, wypowiadanych pod adresem ludzi, którzy nie zajmują tak wysokich stanowisk, po to, by zmaksymalizować pozytywny wizerunek rozmówcy (Dąbrowska 1992: 145). Metonimicznego rozszerzenia znaczenia można się także dopatrywać w rzutowaniu typu CAŁOŚĆ ZA CZĘŚĆ, na przykład w wyrażeniu *wysokie czoło*, które jest eufemizmem łysiny (SEP: 36), lub też w określeniach *tyłek*, *tył* i *tylna część ciała* (SEP: 45–46), użytych w odniesieniu do pośladków.

Kolejna metonimia motywująca eufemizmy to SKUTEK ZA PRZYCZYNĘ, gdy mówi się o kimś *chory na nogi* lub *zawiany*, mając na myśli, że jest pijany (SEP: 189, 193). Pewne okoliczności związane z wydarzeniem lub czynnością stanowiącą tabu mogą funkcjonować jako domeny źródłowe dla metonimii. Przykładowo, słowo *okres* odnosi się do częstotliwości występowania

menstruacji (Dąbrowska 1994: 231), a wyrażenia *cichacz* (SEP: 73) lub *tajniak* (Dąbrowska 1994: 229), określające ciche puszczanie wiatrów, odnoszą się do sposobu wykonywania czynności, który zastępuje czynność.

6. Eufemizmy motywowane metonimią zdaniową

Jedną z najczęstszych metonimii zdaniowych motywujących tworzenie eufemizmów jest rzutowanie typu CZYNNOŚĆ ZA INNĄ CZYNNOŚĆ WYKONYWANĄ W TYM SAMYM MIEJSCU lub też w innym ujęciu CZYNNOŚĆ ZA KONSEKUTYWNĄ CZYNNOŚĆ (Duda 2014: 56). Ilustrację tejże metonimii mogą stanowić przykłady (2a), (3a) i (4a), które funkcjonują jako eufemizmy zdań oznaczonych odpowiednio (2b), (3b) i (4b).

(2a) Zaglądał do kieliszka. (SEP: 186)
(2b) Pił alkohol.
(3a) Spaliście ze sobą? (SEP: 100)
(3b) Uprawialiście seks?
(4a) Idę na posiedzenie. (SEP: 68)
(4b) Idę oddać stolec.

Inną częstą metonimią zdaniową stosowaną w tworzeniu eufemizmów jest rzutowanie typu MIEJSCE ZA CZYNNOŚĆ, którą można także sprecyzować jako chodzenie do danego miejsca za charakterystyczną czynność wykonywaną w tym miejscu (Bierwiaczonek 2013: 158). Przykładowo, predykat *iść z kimś do łóżka* może eufemistycznie oznaczać ‘uprawiać z kimś seks’ (SEP: 93), natomiast zwrot *iść do łazienki* może znaczyć ‘oddawać mocz lub stolec’ (Dąbrowska 1994: 223).

Metonimii tego rodzaju można także dopatrzyć się w zdaniach takich jak *Oszczędnie operowała prawdą* lub *Rozminęła się z prawdą*, które oznaczają po prostu ‘kłamała’ (SEP: 165–167). Ilustrują one rzutowanie typu WIĘCEJ ZA MNIEJ. Podobnie rzecz się ma z powiedzeniem *robić z kimś te rzeczy*, które może eufemistycznie oznaczać ‘uprawiać z kimś seks’ (SEP: 101–102). A. Dąbrowska (1994: 337–339) kwalifikuje tego typu wypowiedzi – opierające się na formie opisowej i podawaniu większej ilości informacji, niż potrzeba w danej sytuacji – jako *peryfrazy*.

Istnieje także grupa eufemizmów polskich, które wydają się motywowane metonimią zdaniową typu SKUTEK ZA PRZYCZYNĘ czynności, należącej do sfery tabu. Zdania (5a–7a) to eufemizmy zdań (5b–7b) motywowane tą metonimią.

(5a) Zamknęła oczy. (SEP: 146)
(5b) Umarła.
(6a) Ma gruby/chudy portfel. (SEP: 243)
(6b) Jest bogaty/biedny.
(7a) Ukrywa brzuszek. (SEP: 81)
(7b) Jest w ciąży.

Z kolei zdania (8a), (9a) i (10a) ilustrują przeciwstawną metonimię zdaniową, motywującą powstanie eufemistycznych odniesień do zdań (8b), (9b) i (10b), to jest rzutowanie typu PRZYCZYNA ZA SKUTEK. Trzeba zauważyć, że owa przyczyna stanowiąca nośnik rzutowania metonimicznego może być zarówno dosłowna, jak w przykładach (8a) i (9a), jak i metaforyczna, co pokazuje zdanie (10a).

(8a) Lubi dobrze zjeść. (SEP: 37)
(8b) Jest gruba.
(9a) Odczuwam takie mdłości. (SEP: 71)
(9b) Chcę wymiotować.
(10a) Upadł na głowę. (Dąbrowska 1994: 113)
(10b) Jest głupi.

7. Eufemizmy motywowane metonimią illokucyjną

Przykładami metonimii illokucyjnej, motywującej proces tworzenia wyrażeń eufemistycznych, są akty zaprzeczania, prośby, nakazywania i pytania, które mogą zastępować akty stwierdzania czegoś, co odnosi się do sfery tabu. Przykładowo, zdania (11a) i (12a) zawierają zaprzeczenia, które mają moc illokucyjną zdań twierdzących podanych w (11b) i (12b). Mamy tutaj do czynienia ze zjawiskiem retorycznym nazywanym *litota*, polegającym na zaprzeczeniu przeciwieństwa pojęcia podstawowego (Dąbrowska 1994: 356).

(11a) Nie jest zdrowy. (SEP: 131)
(11b) Jest chory.
(12a) To nie jest prawda. (SEP: 166)
(12b) To kłamstwo.

Stwierdzenie czegoś, co jest nacechowane negatywnie, może być zastąpione eufemistycznie aktem prośby, jak podano w przykładzie (13a), zapytania, jak w przykładzie (14a), czy też nakazu, co pokazuje zdanie (15a). Te trzy akty są eufemizmami aktów podanych w przykładach (13b), (14b) oraz (15b).

(13a) Proszę o woreczek! (SEP: 73)
(13b) Będę wymiotować.
(14a) Czyś ty się dzisiaj dobrze wyspała? (Dąbrowska 1994: 143)
(14b) Źle wyglądasz.
(15a) Wytrzeźwiej, umyj się, nabierz chlujności, a potem spytaj jeszcze raz. (NKJP)
(15b) Śmierdzisz/jesteś odrażający.

Innym przykładem metonimii illokucyjnej w eufemizmach może być zastępowanie aktu stwierdzania czegoś czy też zgody z kimś aktem warunkowego zaprzeczania. Przykład taki

zawiera dialog ujęty w (16a). Odpowiedź uczestnika B kryje w sobie siłę illokucyjną zdania wyrażonego w (16b).

(16a) A: Ludzie mówią, że jestem gruby i brzydki.
B: Nie, nie jest pan brzydki i gruby, tylko dobrze i zdrowo pan wygląda. (Dąbrowska 1992: 146)
(16b) Zgadzam się, że jest pan brzydki i gruby.

8. Podsumowanie

Jak wynika z przeprowadzonej powyżej analizy, wśród eufemizmów polskich można odnaleźć przykłady ilustrujące wszystkie cztery rodzaje metonimii, o których piszą B. Bierwiaczonek (2013) oraz inni językoznawcy kognitywni. Eufemizmy, za którymi stoi metonimia formalna, obejmują formy językowe tworzone poprzez różnego rodzaju ucięcia, akronimizację, elipsę, jak również zapożyczenia słów o tym samym znaczeniu z języków obcych lub z gwar zawodowych. Wśród podanych przykładów znalazły się także modyfikacje słów objętych tabu, takie jak zmiany fonetyczne pojedynczych głosek i całych sylab, upodobnienie wymowy do języka obcego lub zastosowanie zdrobnień. Niekiedy dochodzi także do zastąpienia negatywnie nacechowanego słowa w danym wyrażeniu zupełnie innym niezwiązanym znaczeniowo z pierwotnym słowem.

Eufemizmy motywowane metonimią referencyjną wykorzystują jako nośniki rzutowania metonimicznego wykonawcę czynności, użytkownika lub posiadacza przedmiotu, cechę charakterystyczną bądź też funkcję lub rolę osoby lub przedmiotu, materiał użyty do wytworzenia przedmiotu, producenta lub markę konkretnego produktu, typową ilość używanego produktu, skutek stanu rzeczy, czas i sposób wykonywania czynności, jak również uogólnienia lub ukonkretnienia pewnych zjawisk, czynności, miejsc bądź rzeczy. Z kolei w metonimii zdaniowej motywacją do tworzenia eufemizmów są takie nośniki, jak miejsce wykonywania pewnych objętych tabu czynności czy też chodzenie do tych miejsc, inne czynności wykonywane w tym samym miejscu, skutki pewnych czynności, jak również ich przyczyny oraz powiedzenie więcej, niż wynika to z charakteru opisywanego zjawiska. Co się zaś tyczy metonimii illokucyjnej, w analizowanych eufemizmach akty stwierdzania faktu objętego tabu mogą zostać zastąpione aktami zaprzeczania, także warunkowego, pytania, prośby i nakazywania.

Przeprowadzona analiza pokazuje, że obecność metonimii jako czynnika motywującego tworzenie licznych eufemizmów w języku polskim pozwala na skuteczne unikanie bezpośrednich odniesień do sfery tabu zarówno na poziomie form językowych, skojarzeń znaczeniowych, jak również całych zdań oraz aktów mowy. Nie ulega wątpliwości, że metonimia rozumiana przez kognitywistów jako zjawisko pojęciowe odgrywa w tworzeniu eufemizmów o wiele większą rolę, niż nadawano jej we wcześniejszym, tradycyjnym ujęciu. Jak wykazaliśmy w niniejszym opracowaniu, jej wpływu można dopatrzyć się w wielu różnych językowych strategiach tworzenia eufemizmów, także w takich, których zazwyczaj nie łączyło się ze zjawiskiem metonimii.

Bibliografia

Allan K., Burridge K. 1991: *Euphemism and dysphemism. Language used as shield and weapon*, Oxford University Press, Oxford.

Allan K., Burridge K. 2006: *Forbidden words. Taboo and the censoring of language*, Cambridge University Press, Cambridge.

Austin J.L. 1993: *Mówienie i poznawanie. Rozprawy i wykłady filozoficzne*, przeł. B. Chwedeńczuk, Wydawnictwo Naukowe PWN, Warszawa.

Barcelona A. 2005: *The multilevel operations of metonymy in grammar and discourse, with particular attention to metonymic chains*, [w:] F. Ruiz de Mendoza, S. Peña Cervel (red.), *Cognitive linguistics. Internal dynamics and interdisciplinary interaction*, Mouton de Gruyter, Berlin, s. 313–352.

Bierwiaczonek B. 2013: *Metonymy in language, thought and brain*, Equinox, Sheffield.

Brown P., Levinson S.C. 1987: *Politeness. Some universals in language usage*, Cambridge University Press, Cambridge.

Dąbrowska A. 1992: *Eufemizmy mowy potocznej*, Język a Kultura, t. 5, s. 119–178.

Dąbrowska A. 1994: *Eufemizmy współczesnego języka polskiego*, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław.

Duda B. 2014: *The synonyms of* fallen woman *in the history of the English language*, Peter Lang, Frankfurt am Main.

Engelking A. 1984: *Istota i ewolucja eufemizmów (na przykładzie zastępczych określeń śmierci)*, „Przegląd Humanistyczny" 4, s. 115–129.

Evans V. 2009: *Leksykon językoznawstwa kognitywnego*, przeł. M. Buchta, M. Cierpisz, J. Podhorodecka, A. Gicala, J. Winiarska, Universitas, Kraków.

Grzegorczykowa R. 2001: *Wprowadzenie do semantyki językoznawczej*, wyd. 3 popr. i rozszerz., Wydawnictwo Naukowe PWN, Warszawa.

Kuczok M. 2015: *Adult drinks and hanky-panky. Types of metonymic motivation in English X-phemisms*, „Linguistica Silesiana" 36, s. 111–125.

Lakoff G., Johnson M. 1988: *Metafory w naszym życiu*, przeł. T.P. Krzeszowski, Państwowy Instytut Wydawniczy, Warszawa.

Levinson S.C. 2010: *Pragmatyka*, przeł. T. Ciecierski, K. Stachowicz, Wydawnictwo Naukowe PWN, Warszawa.

NKJP: Narodowy Korpus Języka Polskiego, 2008–2012 (online: http://nkjp.pl, dostęp: 24 lutego 2016).

Searle J.R. 1987: *Czynności mowy. Rozważania z filozofii języka*, przeł. B. Chwedeńczuk, Instytut Wydawniczy Pax, Warszawa.

SEP: A. Dąbrowska, *Słownik eufemizmów polskich, czyli w rzeczy mocno, w sposobie łagodnie*, wyd. 3, Wydawnictwo Naukowe PWN, Warszawa 2009.

Tabakowska E. (red.) 2001: *Kognitywne podstawy języka i językoznawstwa*, Universitas, Kraków.

Summary

Types of metonymic motivation in Polish euphemisms

Keywords: euphemism, formal metonymy, referential metonymy, propositional metonymy, illocutionary metonymy.

The article aims at presenting the metonymic motivation behind euphemisms in the Polish language. Euphemisms are used by speakers as tools for coping with the negatively charged sphere of taboo that is present in culture. They allow both the speaker and the hearer to save their faces, and to follow the principles of politeness in language. The study is based on works in the field of cognitive linguistics, in which metonymy is perceived as a cognitive mechanism that is conceptual in nature. Among the analyzed examples four groups of Polish euphemisms are distinguished, according to the four types of metonymy motivating their creation: formal metonymy, referential metonymy, propositional metonymy, and illocutionary metonymy.

JOANNA DUSKA, DOROTA MIKA*

INSTYTUT JĘZYKA POLSKIEGO POLSKIEJ AKADEMII NAUK, KRAKÓW

Piętnastowieczne *hejstać* a współczesne *hejtować* – o agresji werbalnej w polszczyźnie[1]

Słowa kluczowe: agresja werbalna, semantyka, motywacja semantyczna.

Hejstać i *hejtować* to dwa wyrazy, które pojawiły się niezależnie od siebie w dwóch różnych okresach rozwoju polszczyzny. Łączy je to, że należą do słownictwa z zakresu agresji werbalnej o bardzo silnym, negatywnym ładunku, wykazują też pewne podobieństwo w brzmieniu. Celem naszej pracy jest ustalenie ich znaczeń, zbadanie ich etymologii i określenie, czy możemy mówić także o ich wspólnym pochodzeniu.

1. Agresja werbalna w pracach językoznawczych

Spośród większych prac poświęconych agresji można wymienić dwa studia: Marii Peisert *Formy i funkcje agresji werbalnej. Próba typologii* (2004) oraz Bożeny Taras *Agresja. Studium semantyczno-pragmatyczne* (2013). W obu książkach znajduje się obszerna bibliografia z tego zakresu. W roku 1999 odbyło się ogólnopolskie konwersatorium w Karpaczu *Język a kultura*, którego wynikiem był tom *Życzliwość i agresja w języku i kulturze* (Dąbrowska, Nowakowska (red.) 2005). Problem agresji werbalnej pojawił się jednak wcześniej w polskich pracach językoznawczych w artykułach Krystyny Pisarkowej dotyczących obrazy (1976) oraz honoru (1978). W następnych latach Kazimierz Ożóg (1981) badał wyrazy obraźliwe, Maciej Grochowski (1982) dokonał wstępnej analizy jednostek wyrażających negatywne relacje osobowe (kpiny, zniewagi, upokorzenia), Renata Grzegorczykowa w artykule *Obelga jako akt mowy* (1991) odwołała się do etymologii i historii wyrazów z tego zakresu.

Czym zatem jest agresja werbalna? Jadwiga Puzynina określa język agresji jako zachowanie będące wyrazem złych emocji w stosunku do odbiorcy lub kogoś, o kim mowa, które zakłada negatywne wartościowanie (Puzynina 1997: 77), a B. Taras podkreśla również, że zachowania agresywne mają na celu zdeprecjonowanie odbiorcy (Taras 2013: 52). Na podstawie przeglądu stanowisk (Arnolda H. Bussa, Elliota Aronsona, Michaela Argyle'a, Bernarda Fine'a, Burnesa E. Moore'a, Leonarda Berkovitza i Henryka Pietrzaka) zawartego w publikacji M. Peisert

* joanna.duska@ijp-pan.krakow.pl, dorota.mika@ijp-pan.krakow.pl

1 Wkład obu autorek w powstanie niniejszego artykułu jest równy. Wspólnie opracowałyśmy koncepcję i założenia tekstu, wspólnie przeprowadziłyśmy analizy i nadałyśmy pracy ostateczny kształt.

(2004: 23–24) można przyjąć, że agresja werbalna, która jest przejawem agresji psychicznej, to reakcja słowna posługująca się groźbą i odrzuceniem; przeciwnika określa się jako złego, niepotrzebnego, umniejsza się jego wartość, jej celem jest osiągnięcie przewagi nad innymi, wyrządzenie szkody, przykrości, krzywdy.

2. Wyraz *hejstać* w staropolszczyźnie

Jak dotąd nikt nie zajmował się dokładniejszym opracowaniem czasownika *hejstać*. Losy zabytku, z którego pochodzi, zostały przedstawione w *Opisie źródeł Słownika staropolskiego* (Twardzik (red.) 2005: 206). Był to datowany na połowę XV wieku rękopis przechowywany w Cesarskiej Bibliotece Publicznej w Petersburgu (sygn. Lat. I O 32). Według opisu w *Przeglądzie językowych zabytków staropolskich do r. 1543*:

> kodeks oprawny w drzewo i skórę, bez początku; zawiera wypisy z proroków, traktaty o pokucie, spowiedzi, objaśnienie przykazań Boskich, dekalog rymowany polski na k. 27b, wypisy zdań i sentencyj moralnych w porządku alfabetycznym, według pierwszych wyrazów; w ustępie p.t. „parvulorum peccata" ciekawe wzmianki o przekleństwach i kradzieżach chłopców wiejskich z wyrazami polskiemi; traktat o cnotach, „casus reservati", „autoritates regis Salomonis"; k. 207 z napisem z boku: Johannes de Radomske, przepisy lekarskie np. o puszczaniu krwi, formuły czarodziejskie (sator, arepo, tenet, opera, rotas k. 241b), objaśnienie nazw miesięcy ze znaczeniami polskiemi, objaśnienie znaków zodyaku, ustępy „de symonia", „de hereticis" i innych wiele drobnych rzeczy. Na okładce przepisy lekarskie z wyrazami polskiemi i data 1454 (Łoś 1915: 515–516).

Rękopis wrócił wprawdzie do Biblioteki Narodowej w Warszawie na mocy traktatu ryskiego (1921), ale niestety został spalony przez Niemców w czasie drugiej wojny światowej. Informację o jego zawartości w zakresie staropolszczyzny zawdzięczamy Hieronimowi Łopacińskiemu (1860–1906), który zajmował się nim jeszcze w Petersburgu i sporządził stosowne notatki. Badacz ten był w stałym kontakcie z warszawskim środowiskiem językoznawczym, brał udział w pracach nad Słownikiem warszawskim i w tym też słowniku wyraz *hejstać* został odnotowany po raz pierwszy w znaczeniu ‘znieważać’. Na jego podstawie hasło to w takiej samej postaci pojawiło się w *Michała Arcta słowniku staropolskim*, opracowanym przez Antoniego Krasnowolskiego, wydanym w 1914 roku w Warszawie. Słownik ten, jak wskazuje karta tytułowa, zawiera „26.000 wyrazów i wyrażeń używanych w dawnej mowie polskiej", a określenie „staropolski", w odróżnieniu od ram przyjętych dla SStp pod redakcją Stanisława Urbańczyka, było rozumiane znacznie szerzej. Równolegle do wspomnianych prac słownikowych ukazał się *Przegląd językowych zabytków staropolskich do r. 1543* opracowany przez Jana Łosia i wydany w Krakowie w roku 1915 – również w nim wykorzystano materiał zebrany przez Hieronima Łopacińskiego. *Przegląd zabytków* jest jednym ze źródeł SStp i i na jego podstawie utworzono hasło *hejstać*. Zostało ono objaśnione definicją ‘bluźnić, złorzeczyć, maledicere, convicia dicere’ i jest poświadczone jednym tylko cytatem jako hapax legomenum. Na podstawie zawartości tak zwanej kartki materiałowej podano cytat w postaci: *Heystacz blasfemare XV*

med. Zab 516. Szerszy kontekst we wspomnianym powyżej *Przeglądzie językowych zabytków staropolskich* Jana Łosia wygląda następująco:

> K. 138 blasfemare – *heystacz*... Nam pro eodem peccato (tj. blasphemia) dum filiorum Israel pueri Elizeum prophetam equitantem in asino inclammassent illudentes et dicentes *geszdzy lyszku*[2], aput nos enim dicitur *hele hele* velut *kobila kobila* (Łoś 1915: 516).

Tłumaczenie:

> K. 138 lżyć – *hejstać*... Bowiem, odnośnie do tego samego grzechu (tj. zniewagi), gdy dzieci żydowskie lżyły proroka Elizeusza jadącego na ośle, wyszydzając go i mówiąc: *jedź łysku*, u nas mówi się [na to] *hele hele* podobnie *kobyła, kobyła*[3].

Fragment ten nawiązuje do biblijnej historii wyszydzenia proroka Elizeusza przez małych chłopców, w rękopisie poprzedza on bezpośrednio rozdział *Parvulorum peccata* (Grzechy dzieci) (por. Łoś 1915: 515). W powyższym cytacie użyto dwóch czasowników z zakresu agresji werbalnej, z których jeden wskazuje na wyśmiewanie (*illudentes*), drugi na lżenie (*inclammasent*) proroka – zachowania te są postrzegane w kategoriach grzechu. Jest to parafraza następującego fragmentu *Wulgaty*:

> Ascendit autem inde in Bethel. Cumque ascenderet per viam, pueri parvi egressi sunt de civitate et illudebant ei dicentes: Ascende, calve; ascende, calve! (IV Reg 2,23)

Illudo według *Słownika łaciny średniowiecznej* oznaczało ‘wyśmiewać się, kpić, szydzić, naigrawać się’ (t. V, 1978–1984: 68). Nie zachowało się polskie XV-wieczne tłumaczenie tego fragmentu Starego Testamentu (w *Biblii Królowej Zofii* brakuje w tym miejscu dziewięciu kart), natomiast XVI-wieczne tłumaczenia konsekwentnie oddają łacińskie *illudebant* jako *naśmiewały się* (por. *Biblia Leopolity* 1561, *Biblia Radziwiłłowska* 1563, *Biblia Budnego* 1572, *Biblia Wujka* 1599). Tłumaczenie tego fragmentu w *Biblii Tysiąclecia* wygląda następująco:

> Stamtąd poszedł do Betel. Kiedy zaś postępował drogą, mali chłopcy wybiegli z miasta i naśmiewali się zeń wzgardliwie, mówiąc do niego: «Przyjdź no, łysku! Przyjdź no, łysku!» (IV Krl 2,23).

Współczesne znaczenie czasownika *naśmiewać się* ma być może słabszy wydźwięk niż w okresie staropolskim (SStp: ‘wyśmiewać, drwić, szydzić’; SPXVI: ‘wyśmiewać się, drwić; urągać, szydzić, bluźnić’), skoro dla oddania rozmiaru czynu popełnionego przez dzieci tłumacze

2 W SStp wyraz *łysek* został zdefiniowany jako ‘człowiek wyłysiały, łysy, bez włosów na głowie, homo calvus’; ponieważ znaczenie to utworzono na podstawie jedynego, omawianego tu cytatu z *Zabytków* Łosia, należałoby dodać, że jest to określenie pogardliwe, będące wyzwiskiem.

3 Tłumaczenie Joanna Duska, konsultacja Jagoda Marszałek.

Biblii Tysiąclecia posłużyli się bardziej rozbudowaną konstrukcją składniową, stąd mowa o tym, że chłopcy „naśmiewali się zeń wzgardliwie".

Takie szersze spojrzenie na omawiany fragment: uwzględnienie jego pozycji w rękopisie, czyli powiązanie tematyczne z rozdziałem *Peccata parvulorum* oraz uwzględnienie kontekstu biblijnego, do którego bezpośrednio nawiązuje, pozwala na prawidłowe zinterpretowanie zapisu *blasfemare – hejstać*. Jest to prawdopodobnie przykład łacińsko-polskiej glosy marginalnej do przytoczonego fragmentu karty 138 omawianego rękopisu. *Słownik łaciny średniowiecznej w Polsce* podaje dwa znaczenia dla *blasfemare*: 1) 'bluźnić, znieważać Boga lub religię'; 2) 'lżyć, znieważać' (Plezia (red.) 1953: 1118). Dla omawianego cytatu kluczowe jest drugie znaczenie. Co w kontekście tych rozważań można powiedzieć o znaczeniu staropolskiego czasownika *hejstać*? Wyraz ten oznaczał naganne czy wręcz agresywne zachowanie werbalne. Autorzy SW zdefiniowali *hejstać* jako 'znieważać', autorzy SStp jako 'bluźnić, złorzeczyć'. Na podstawie przedstawionych wyżej informacji można stwierdzić, że znaczenie tego wyrazu należy oddać jako 'lżyć, znieważać', co zarówno znajduje potwierdzenie w kontekście, w którym się on pojawia, jak i pozostaje w zgodzie ze średniowiecznym znaczeniem łacińskiego czasownika *blasfemare*, z którym został zestawiony. Tłumaczenie *hejstać* jako 'bluźnić' zaproponowane przez autorów SStp nie wydaje się odpowiednie, gdyż we współczesnym języku polskim, w którym formułowane są objaśnienia wyrazów hasłowych, *bluźnić* oznacza wyłącznie uwłaczanie czemuś, co przez religię uznawane jest za święte, omawiany zaś kontekst wskazuje, że nie chodziło tu o uwłaczanie Bogu czy religii.

Na początku XXI wieku w polszczyźnie zaczęła się bardzo szybko rozrastać rodzina wyrazowa innego podobnie brzmiącego czasownika z zakresu agresji werbalnej, którym jest *hejtować*.

3. Hejt i *hejtować* jako przykład najmłodszej leksyki z zakresu agresji werbalnej

Pod koniec XX wieku stary angielski rzeczownik *hate* poszerzył swoje znaczenie. Pod wpływem międzynarodowych regulacji prawnych związanych z ochroną praw człowieka rozpowszechniły się połączenia typu: *hate crime* 'przestępstwo motywowane nienawiścią' i *hate speech* 'mowa nienawiści'. W języku ogólnym doszło do rozszerzenia na sferę internetową zakresu znaczeniowego frazeologizmu *hate mail* 'obraźliwe listy, listy z pogróżkami'. Wszystkie te konstrukcje wskazują na „nienawiść czynną", czyli realizującą się w działaniu. Pod wpływem nowych połączeń sam wyraz *hate* zaczął być używany także w takim znaczeniu[4] i w nim właśnie został zapożyczony do języka polskiego.

Wyrazu *hejtować* nie uwzględniono we wspomnianych wcześniej publikacjach M. Peisert (2004) i B. Taras (2013). Po raz pierwszy odnotowano go, zgodnie z chronologią podaną w WSJP, w roku 2007[5]. Jest to nowe zapożyczenie angielskie w polszczyźnie, które bardzo szybko się upowszechnia. Anna Niepytalska-Osiecka w roku 2014 wymieniała następujące wyrazy należące do

4 Przejście od znaczenia: 'uczucie nienawiści' do znaczenia 'działanie powodowane nienawiścią' jest widoczne także w wypadku derywatów rzeczownika *hate*, por. ang. *hater*: pierwotnie 'osoba, która nienawidzi', później także 'osoba, która wypowiada lub pisze nieprzyjemne rzeczy o kimś lub krytykuje jego osiągnięcia, szczególnie w Internecie' (por. Niepytalska-Osiecka 2014: 349).

5 W publikacji Piotra Flicińskiego i Stanisława Wójtowicza *Hip-hop słownik* z roku 2007.

tej rodziny: *hejt*, *hejterstwo*, *hejtować*, *hejcić*, *hejterski*, *hejtowy*, *zhejtować*, *zhejcić*, *zahejtować* (Niepytalska-Osiecka 2014: 343–352). Wciąż powstają nowe derywaty, powyższą listę można uzupełnić o takie wyrazy, jak *hejterada*, *hejterka* czy *hejtowanie*. Wyraz *hejt* w języku polskim poszerzył swoje znaczenie – z nienawistnego komentarza zamieszczonego w Internecie stał się określeniem wszelkich przejawów agresji werbalnej. W przeciwieństwie do zapożyczeń z języka angielskiego powodowanych wyłącznie modą jest to wyraz, który nazywa nowe zjawisko, pierwotnie związane z komunikacją internetową, polegające na prześladowaniu, nękaniu psychicznym tego, kto jest obiektem hejtu. Celowo użyto tu sformułowania *nowe zjawisko*. Samo zjawisko agresji nie jest nowe w przeciwieństwie do agresywnych zachowań, które mogą się różnić motywacją i formą, gdyż jak zauważa M. Peisert:

> Zdolność do agresywnych zachowań jest u człowieka, podobnie jak mowa, wrodzoną zdolnością, składnikiem wyposażenia genetycznego, ale formy agresywnych zachowań są wyuczone i wchodzą w skład kompetencji kulturowo-językowej członków konkretnej grupy ludzi (Peisert 2004: 25).

Naszym zdaniem hejt najbliższy jest takim znanym formom agresywnych zachowań językowych, jak inwektywa ‘wypowiedź będąca napastliwym wystąpieniem przeciw jakiejś osobie lub instytucji, operująca określeniami obraźliwymi i zniesławiającymi’, pamflet ‘utwór publicystyczny lub literacki, często anonimowy, mający charakter demaskatorskiej i zwykle ośmieszającej krytyki znanej osoby, środowiska społecznego czy instytucji’ oraz paszkwil ‘utwór, zwykle anonimowy lub publikowany pod pseudonimem, wymierzony przeciw konkretnej osobie, ośmieszający ją złośliwie, w sposób insynuacyjny, obelżywy i zniesławiający, aby skompromitować ją w oczach opinii publicznej’[6]. Inwektywa była wypowiedzią wprost. Pamflet i paszkwil były to formy literackie o napastliwym charakterze, anonimowe lub pisane pod pseudonimem. Drukowano je lub odczytywano publicznie. Powstawały w okresach ożywienia życia społecznego, ich celem była przede wszystkim krytyka, która miała służyć zmianie sytuacji politycznej albo społecznej. Przedmiotem krytyki były osoby znane – zwykle warstwa rządząca.

Czym różni się od nich *hejt*? WSJP definiuje go jako: 1) ‘ogół negatywnych emocji i ocen, które wyrażają się we wrogich i krzywdzących wypowiedziach pojawiających się masowo’; 2) ‘wypowiedź, napis lub obrazek o wymowie wrogiej lub krzywdzącej kogoś’. Jest to więc jednocześnie nazwa zjawiska zarówno z zakresu komunikacji społecznej, jak i gatunku wypowiedzi. Wprowadzenie tego wyrazu oraz powstanie całej jego rodziny mają źródło w specyficznym typie komunikacji, jakim jest komunikacja internetowa, która umożliwia formułowanie wypowiedzi publicznych, pozwalając zachować anonimowość. W środowisku internautów nastąpił gwałtowny wzrost negatywnych komunikatów służących obrażaniu i dyskredytowaniu, wymierzonych w osoby, grupy społeczne, zjawiska czy światopogląd. Ten akt werbalny ma na celu wyrażenie negatywnej opinii, krytykę, wytykanie wad i błędów bez podawania

6 Definicje na podstawie *Słownika terminów literackich* (Sławiński (red.) 2000): *inwektywa* (s. 220), *pamflet* (s. 368), *paszkwil* (s. 378).

argumentów merytorycznych; bywa powodowany zazdrością. Jak piszą autorzy projektu „Najnowsze słownictwo polskie", prowadzonego w ramach Obserwatorium Językowego Uniwersytetu Warszawskiego – w skrajnej postaci przyjmuje on formę bezwzględnej słownej napaści, obraźliwego, agresywnego, prowokacyjnego i skrajnie krytycznego komentarza[7]. Wypowiedzi jednostkowe mogą nabierać charakteru masowego, stąd możemy mówić o *fali hejtu* czy *hejteradzie*. Wyraz *hejt* oraz jego derywaty pierwotnie pojawiały się w internetowych wypowiedziach, ale jak pisze A. Niepytalska-Osiecka: „Z wypowiedzi na forach internetowych zaczynają przenikać do innych typów tekstów, a także do potocznego wariantu polszczyzny mówionej" (Niepytalska-Osiecka 2014: 351). Grono odbiorców jest zatem bardzo szerokie, są to przede wszystkim użytkownicy portali społecznościowych i forów internetowych, a po oderwaniu się hejtu od rzeczywistości wirtualnej także użytkownicy innych mediów (prasy, telewizji, radia). Co warto podkreślić, autorem hejterskiej wypowiedzi, podobnie jak obiektem hejtu, może być każdy.

4. Pochodzenie *hejstać* i *hejtować*

Zacznijmy od czasownika *hejstać*. W SW pojawia się adnotacja wskazująca na niemiecką proweniencję – autorzy wywodzą go od *hetzen*. Jak wskazuje Gerhard Köbler (1995), czasownik *hetzen* w języku średnio-wysoko-niemieckim (1100–1500, a więc wtedy, kiedy mogło nastąpić zapożyczenie do języka polskiego), znaczył 'hetzen, jagen, antreiben' ('szczuć, polować, zapędzić'), rozwinął się prawdopodobnie z zachodniogermańskiego **hatjan* 'verfolgen, hetzen' ('ścigać, szczuć'). Autor słownika odsyła do wyrazu *Haβ* 'nienawiść', który wiąże się z pie. **kād- /*kəd-* 'seelische Verstimmung, Kummer, Haβ, Sorge, Leid' ('cierpienie psychiczne, żal, nienawiść, zmartwienie, krzywda'). Za związkiem *hetzen* i *Hass* przemawia stosunek kauzatywny, na który wskazuje *Słownik etymologiczny języka niemieckiego* pod redakcją Wolfganga Pfeifera (1989) – formą kauzatywną czasownika *hassen* 'Haβ empfinden' ('odczuwać nienawiść') jest *hetzen* 'hassen machen, zur Verfolgung bringen' ('odnosić się do kogoś z nienawiścią; prześladować kogoś'). Staropolski czasownik *hejstać* mógłby być bezpośrednim zapożyczeniem z języka średnio-wysoko-niemieckiego. Należałoby jednak rozważyć także sytuację, w której źródłosłów wyrazu *hejstać* byłby niemiecki, ale wyraz ten przeszedłby do języka polskiego w sposób pośredni, np. z języka czeskiego[8]. Internetowy słownik języka staroczeskiego[9] potwierdza istnienie wyrazu *hesovati* (wraz z pojedynczą odmianką *hejsovati*) w znaczeniu: 1) 'co (zvěř) štvát, honit; (psa) pobízet k lovu'; 2) jur. 'koho kam hnát, pohánět, předvolávat k soudu' (1. 'kogo, co (zwierzę) gonić, ścigać; (psa) podjudzać do ścigania'; 2. prawn.

7 Online: http://nowewyrazy.uw.edu.pl/haslo/hejt.html (dostęp: 20 maja 2016).

8 Wśród wyrazów zgromadzonych w słowniku zapożyczeń niemieckich w języku polskim, opracowanym przez zespół badaczy z Uniwersytetu w Oldenburgu, nie odnotowano wprawdzie wyrazu *hejstać*, lecz pojawia się inny wyraz, który może tłumaczyć taki sposób adaptacji. W okresie staropolskim pojawił się zapożyczony bezpośrednio z języka niemieckiego wyraz *hetman*, w tym samym okresie pojawiła się także jego forma oboczna – *hejtman*, która jest notowana przez SStp oraz w zabytkach z I poł. XVI w. w SPXVI. Polscy etymolodzy (Brückner, Sławski, Bańkowski) są zgodni, że druga z nich przeszła do polszczyzny za pośrednictwem języka czeskiego. Także *hejstać*, podobnie jak *hejtman*, mogło ukształtować się pod wpływem języka czeskiego.

9 *Elektronický slovník staré češtiny* (online: http://vokabular.ujc.cas.cz/listovani.aspx, dostęp: 18 maja 2016).

‘kogoś gdzieś ścigać, pozywać przed sąd’). Badacze czescy wywodzą go z niemieckiego *hetzen*, o czym wspomina m.in. Kateřina Volekova (2015: 317). W średniowieczu czeszczyzna silnie oddziaływała na język polski, hipoteza o jej pośrednictwie także w wypadku *hejstać* wydaje się możliwa, gdyż istniał staroczeski wyraz *hesovati* (*hejsovati*), który mógł być podstawą polskiego *hejstać*, o znaczeniu kontynuującym niemieckie *hetzen* ‘szczuć’. Trzeba jednak zauważyć, że znaczenia czeskiego *hesovati* i polskiego *hejstać* różnią się, co może wskazywać na to, że wyrazy te zostały zapożyczone równolegle i niezależnie od siebie. Kwestia ta wymaga szerszych badań porównawczych.

Przejdźmy teraz do etymologii wyrazu *hejtować*. Według słownika etymologicznego języka angielskiego (Skeat 1965: 234) czasownik *to hate* ‘nienawidzić’, z którego wywodzi się współczesne *hejtować*, również pochodzi z zachodniogermańskiego **hatjan*, w języku staroangielskim czasownik *hatian* oznaczał ‘odnosić się wrogo, z niechęcią’, współcześnie *to hate* znaczy ‘nienawidzić’. Tak więc źródeł staropolskiego *hejstać* i współczesnego *hejtować* należy szukać w rekonstruowanym **hatjan* (z pie. **kād- /*kəd-*).

5. Motywacje semantyczne *hejstać* i *hejtować*

Renata Grzegorczykowa w artykule dotyczącym obelgi jako aktu mowy prezentuje etymologię i historię kilku wyrazów z zakresu agresji werbalnej. Są to takie czasowniki (lub ich rodziny), jak: *urazić*, *lżyć*, *znieważać*, *ubliżyć*, *uwłóczyć*, *zniesławić*, *wyzywać*, *urągać*, *łajać* (Grzegorczykowa 1991). Autorka zalicza je do następujących kategorii: „czynności fizyczne”, „umniejszanie”, „zniesławianie”, „wydawanie dźwięków”. Mariola Jakubowicz, wskazując podstawy motywacyjne kilku wybranych wyrazów z kręgu agresji (*agresywny*, *napastliwy*, *prowokacyjny*), pisze, że pogłębiona analiza większej liczby leksemów może przynieść wyraźniejszy obraz zarysowujących się prawidłowości. Badacze wskazują na tendencje do konceptualizacji agresji przede wszystkim za pomocą leksemów z zakresu ruchu (Jakubowicz 2005: 67) i nazw pewnych czynności fizycznych (Grzegorczykowa 1991: 198). W której grupie znalazłyby się leksemy *hejstać* i *hejtować*?

Analiza leksyki staropolskiej z zakresu agresji werbalnej, przeprowadzona przez nas pod kątem ustalenia podstaw motywacyjnych wyrazów, wykazała, że dla tego typu słownictwa najczęściej są nimi: 1) RUCH (np. *naigrawać* – w okresie staropolskim ‘naśmiewać się, szydzić’, z praindoeuropejskiego **aiǧ-* ‘ruszać (się) gwałtownie, wywijać, trząść’); 2) KONTAKT FIZYCZNY WPŁYWAJĄCY NA OBIEKT (np. *gabać* w staropolszczyźnie ‘prześladować, dręczyć, napastować’, od pie. **gʰabʰ-* ‘chwytać, brać’); 3) KONTAKT WERBALNY, w tym: a) mówienie (np. *przeklinać* – w SStp ‘źle życzyć, słowami ściągać nieszczęście, potępiać, ubliżać, wyjmować spod prawa’, pie. **kel-* ‘wołać, krzyczeć, hałasować, rozbrzmiewać’); b) onomatopeje (np. *sapać* – w SStp ‘dyszeć gniewem, rzucać pogróżki’, pie. **sụ̯ep-* ‘dyszeć, sapać’); 4. EMOCJE, w tym: a) wstręt, obrzydzenie, niechęć (np. *hydanie* – w SStp ‘obmawianie, obmowa, oszczerstwo’ z pie. **gʷ ōu̯-* / **gʷū-* ‘gnój, kał, odchody, coś wstrętnego’, tu także *mierziączka* w SStp ‘obelga, zniewaga, obraza’; *szkarady* ‘uwłaczający czci, obraźliwy’; *żadanie* ‘odraza, wstręt, obrzydzenie, też to, co budzi odrazę, wstręt, obrzydzenie’); b) cierpienie (np. *sromocenie* – w SStp ‘zniesławienie, zniewaga, obelga’ z pie. **ḱor-mo-* ‘cierpienie, ból, hańba’, *hejstać* ‘lżyć, znieważać’

(zgodnie z przytoczonymi wyżej ustaleniami) z pie. **kād- / *kəd-* ‘cierpienie psychiczne, żal, nienawiść’); c) tutaj także śmiech jako wyraz emocji (np. *naśmiewać się* – w SStp: ‘wyśmiewać, drwić, szydzić’, od pie. **smeį-* ‘śmiać się’)[10]. Gdyby sporządzić analogiczny wykaz dla współczesnych wyrazów z zakresu agresji werbalnej, *hejtować* znalazłoby się także w grupie motywowanej emocjami.

6. Podsumowanie

Porównując średniowieczne *hejstać* ze współczesnym *hejtować*, można stwierdzić, że w dwóch różnych okresach rozwoju polszczyzny, które dzieli ponad pięćset lat, do języka polskiego zostały niezależnie od siebie zapożyczone obco brzmiące wyrazy związane z nienawiścią. Odnoszą się one do sfery relacji społecznych, podstawa ich motywacji semantycznej, związana z praindoeuropejskim rdzeniem **kād- / *kəd-*, wskazuje na cierpienie psychiczne, poczucie żalu i krzywdy, mogące się przerodzić w niechęć czy nienawiść. Znaczenia polskich zapożyczeń *hejstać* i *hejtować* oderwały się od pierwotnego podłoża emocjonalnego i nazywają działania, których celem nie jest wyrażenie emocji, lecz spowodowanie u kogoś negatywnego stanu psychicznego wskutek agresywnych zachowań werbalnych. W języku polskim pojęcie nienawiści zakłada perspektywę agensa, a więc tego, kto ją odczuwa, a patiens, czyli podmiot nienawiści, ma charakter drugorzędny. W wypadku wyrazów *hejstać* i *hejtować* perspektywa zostaje odwrócona, uwaga nie jest skierowana na agensa, jest on mało istotny, jego rola ogranicza się jedynie do wykonania czynności, ważny jest natomiast patiens, czyli ten, w kogo wymierzone są agresywne działania.

Przedstawiłyśmy tutaj losy dwóch zapożyczeń opartych na tym samym rdzeniu, które pojawiły się w różnych okresach rozwoju systemu leksykalnego języka polskiego, a wcześniej rozwijały się w różnych kręgach kulturowych. Niezależnie od drogi, jaką przebyły te wyrazy, zanim pojawiły się w języku polskim, kontynuują one znaczenie tej samej podstawy etymologicznej. W ten sposób należy uznać, że ich podobieństwo brzmieniowe i znaczeniowe nie jest przypadkowe.

Bibliografia

Bańko M., Czeszewski M., Burzyński J. (red.): *Najnowsze słownictwo polskie*, Obserwatorium Językowe Uniwersytetu Warszawskiego (online: http://nowewyrazy.uw.edu.pl/haslo/hejt.html, dostęp: 20 maja 2016).

Bibliorum sacrorum iuxta Vulgatam Clementinam. Nova editio 1951: curavit A. Gramatica, Typis polyglottis Vaticanis, Vaticanum.

Boryś W. 2005: *Słownik etymologiczny języka polskiego*, Wydawnictwo Literackie, Kraków.

Brückner A. 1927: *Słownik etymologiczny języka polskiego*, nakładem Krakowskiej Spółki Wydawniczej, Kraków.

Dąbrowska A., Nowakowska A. (red.) 2005: *Życzliwość i agresja w języku i kulturze*, Język a Kultura, t. 17, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław.

Grochowski M. 1982: *Zarys analizy semantycznej grupy jednostek wyrażających negatywne etycznie relacje osobowe („kpina", „zniewaga", „upokorzenie")*, „Polonica" VIII, s. 57–72.

10 Etymologia wyrazów na podstawie haseł: ze *Słownika etymologicznego języka polskiego* (Boryś 2005): *igrać* (s. 197), *nagabnąć* (s. 347–348), *kląć* (s. 233), *sapać* (s. 538–539), *ohyda* (s. 385), *mierzić* (s. 325), *szkaradny* (s. 603), *srom* (s. 573), *śmiać się* (s. 617) oraz ze *Słownika etymologicznego języka polskiego* (Brückner 1927): *żadać się* (s. 660).

Grzegorczykowa R. 1991: *Obelga jako akt mowy*, „Poradnik Językowy", nr 5–6, s. 196–201.
Jakubowicz M. 2005: *Źródła motywacji semantycznej nazw z kręgu życzliwości i agresji*, [w:] A. Dąbrowska, A. Nowakowska (red.), *Życzliwość i agresja w języku i kulturze*, Język a Kultura, t. 17, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław, s. 61–68.
Köbler G. 1995: *Deutsches etymologisches Wörterbuch* (online: http://www.koeblergerhard.de/derwbhin.html, dostęp: 20 maja 2016).
Krasnowolski A., Niedźwiedzki W. (oprac.) 1914: *M. Arcta słownik staropolski*, t. 1, Wydawnictwo M. Arcta, Warszawa.
Łoś J. 1915: *Przegląd językowych zabytków staropolskich do roku 1543*, nakładem Akademii Umiejętności, Kraków.
Niepytalska-Osiecka A. 2014: *O* fejku, lajku *i* hejcie *w polszczyźnie internetowej*, „Język Polski" XCV, s. 343–352.
Ożóg K. 1981: *O współczesnych polskich wyrazach obraźliwych*, „Język Polski" LXI, s. 179–187.
Peisert M. 2004: *Formy i funkcje agresji werbalnej. Próba typologii*, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław.
Pfeifer W. 1989: *Etymologisches Wörterbuch des Deutschen*, Akademie-Verlag Berlin, Berlin.
Pisarkowa K. 1976: *Pragmatyczne spojrzenie na akty mowy*, „Polonica" II, Wrocław, s. 265–279.
Pisarkowa K. 1978: *Hasło* honor *jako przedmiot analizy pragmatyczno-językowej*, „Polonica" IV, s. 126–127.
Pismo Święte Starego i Nowego Testamentu w przekładzie z języków oryginalnych, wyd. 1, Pallottinum, Poznań 1965.
Plezia M. (red.) 1953: *Słownik łaciny średniowiecznej w Polsce*, t. 1, Zakład Narodowy im. Ossolińskich, Wrocław–Kraków–Warszawa.
Puzynina J. 1997: *Słowo – wartość – kultura*, Towarzystwo Naukowe Katolickiego Uniwersytetu Lubelskiego, Lublin.
Skeat W.W. 1965: *A Concise Etymological Dictionary of the English Language*, Clarendon Press, Oxford.
Sławiński J. (red.) 2000: *Słownik terminów literackich*, Zakład Narodowy im. Ossolińskich, Wrocław.
Taras B. 2013: Agresja. *Studium semantyczno-pragmatyczne*, Wydawnictwo Uniwersytetu Rzeszowskiego, Rzeszów.
Twardzik W. (red.) 2005: *Opis źródeł Słownika staropolskiego*, Lexis, Kraków.
Voleková K. 2015: *Česká lexikografie 15. století*, Academia, Praga.

Rozwiązanie skrótów

SPXVI: *Słownik polszczyzny XVI wieku*, red. M.R. Mayenowa, t. 16, Zakład Narodowy im. Ossolińskich, Wrocław 1985. – SStp: Słownik staropolski, red. S. Urbańczyk, t. 1–9, Zakład Narodowy im. Ossolińskich, Wydawnictwo PAN, Wrocław–Warszawa–Kraków 1953–1987, t. 10–11, Instytut Języka Polskiego PAN, Kraków 1988–2002. – SW: *Słownik języka polskiego*, red. J. Karłowicz, A. Kryński, W. Niedźwiedzki, t. 1–8, nakładem prenumeratorów i Kasy im. Mianowskiego, Warszawa 1900–1927. – WSJP: *Wielki słownik języka polskiego PAN*, red. P. Żmigrodzki (online: http://www.wsjp.pl/, dostęp: 19 maja 2016).

Summary

The fifteenth-century *hejstać* and contemporary *hejtować* – about verbal aggression

Keywords: verbal aggression, semantics, semantic motivation.

The title of the article contains two words in the field of verbal aggression: the Old Polish verb *hejstać* – unique in the light of preserved materials, dating back to the fifteenth century – and the contemporary Polish verb *hejtować* – a very popular borrowing from English. The authors clarify their meanings, refer to their etymology and identify their common origin from **hatjan*. Finally, authors compare them with other words in this field, based on their semantic motivation.

MONIKA BUŁAWA* | INSTYTUT JĘZYKA POLSKIEGO POLSKIEJ AKADEMII NAUK, KRAKÓW

O nowych użyciach czasownika *masakrować* we współczesnej polszczyźnie

Słowa kluczowe: współczesna polszczyzna, leksyka, rozwój semantyczny wyrazów.

W jednym z niedawnych zeszytów „Języka Polskiego" czytelnicy mogli zapoznać się z tekstem Roberta Słabczyńskiego poświęconym rozwojowi semantycznemu rzeczownika *masakra* oraz modzie na ten wyraz w najnowszej polszczyźnie (Słabczyński 2015). W ostatnim czasie popularność zyskuje również inny leksem należący do tej samej rodziny słowotwórczej – czasownik *masakrować*, co zostało już dostrzeżone przez publicystów (*„Korwin masakruje...", czyli o politycznej karierze pewnego słowa*[1]), doczekało się oceny normatywnej[2], a nawet stało się przedmiotem refleksji o charakterze religijno-moralnym[3]. Warto zatem analizę poświęconą rzeczownikowi *masakra* uzupełnić o opis funkcjonowania we współczesnym języku polskim wspomnianego czasownika, tym bardziej że zwracające uwagę użycia typu *Korwin masakruje lewaka* to tylko jeden z etapów jego o wiele bardziej złożonego rozwoju semantycznego, który dotychczas w niewielkim stopniu został uwzględniony w opisach leksykograficznych tego leksemu.

Czasownik *masakrować*, będący – tak jak i *masakra* – zapożyczeniem z języka francuskiego (fr. *massacre* 'masakra', *massacrer* 'masakrować') (Bochnakowa (red.) 2012: 203), po raz pierwszy odnotowany został w SW II: 892 z definicją 'tłuc na miazgę, miażdżyć, mordować, znosić bez miłosierdzia, rozbijać na proch'[4]. Ten sam zakres semantyczny omawianego leksemu uwzględnia hasło w SJPDor IV: 488 'urządzać masakrę, bić zadając rany, mordować krwawo, bez miłosierdzia; rozbijać, tłuc na miazgę'.

Definicje zamieszczone w późniejszych słownikach (SJPSzym, SWJP, PSWP, ISJP, USJP) ograniczają użycie czasownika *masakrować* do nazywania jedynie takich działań, których obiektem są ludzie (ewentualnie zwierzęta). W podstawowym znaczeniu czasownik *masakrować* definiowany jest jako 'mordować w okrutny, bestialski sposób' (USJP). W SJPSzym i SWJP definicje zawierają element implicytnie informujący, że czasownika tego używa się zwłaszcza w odniesieniu do sytuacji, gdy zabijana jest większa liczba osób (por. *masakrować* 'urządzać masakrę',

* monika.bulawa@ijp-pan.krakow.pl

1 Felieton autorstwa Michała Fala: http://natemat.pl/103155,korwin-masakruje-czyli-o-politycznej-karierze-pewnego-slowa (komentarze pod tekstem pochodzą z maja 2014 r., co wskazuje na datę opublikowania tego felietonu).

2 Ewa Kołodziejek, *Zmasakrować i zaorać*, felieton w „Kurierze Szczecińskim", w Internecie opublikowany w lutym 2016 r. (http://24kurier.pl/blogi/ewa-kolodziejek/zmasakrowac-i-zaorac/).

3 O. Kasper Kaproń, *Potęga słowa*, felieton opublikowany w styczniu 2016 r. (http://www.deon.pl/religia/kosciol-i-swiat/komentarze/art,2287,potega-slowa.html).

4 *Masakrować* występuje w SW także w haśle *Miażdżyć* jako jeden z elementów definicji ('gnieść, tłuc, kruszyć na miazgę, masakrować; gruchotać').

masakra ‘zabijanie, mordowanie, zwłaszcza masowe’ SJPSzym; *masakrować* ‘dokonywać masakry’, *masakra* ‘zadawanie ran, ciosów, śmierci (zwykle większej liczbie ludzi lub zwierząt’ SWJP[5]). Wszystkie sprawdzane słowniki rejestrują ponadto drugie znaczenie[6] czasownika *masakrować*, które na przykład w PSWP zostało zdefiniowane jako ‘okaleczać ciało, powodować silne obrażenia’.

Nie notują natomiast wymienione wyżej słowniki znaczenia ‘rozbijać, tłuc na miazgę’, wyodrębnionego w SJPDor (po średniku)[7] i zilustrowanego przykładem: „[Niemieckie lotnictwo] masakrowało splątane wężowiska pociągów na stacjach”, mimo że jak świadczą dane z NKJP, jest ono obecne we współczesnej polszczyźnie, zob. np.:

(1) Wykroczyli przeciw temu Niemcy [...], *masakrując* żydowskie cmentarze i używając kamiennych macew do remontu dróg (NKJP: J. Chrzan, *Po Zaduszkach*, „Dziennik Bałtycki”, 2009-11-06).

(2) – Jeżdżące na placu autobusy *masakrują* ulice – tłumaczy kierownik Urbański (NKJP: (awu), *Wyprowadzą autobusy z rynku?*, „Dziennik Łódzki”, 2006-06-09).

(3) Ornitolodzy mówią: precz z angielskimi trawnikami. Z przerażeniem patrzą, jak służby miejskie niefachowo opiekują się zielenią i *masakrują* poszycie w parkach (NKJP: G. Zdrojewska, *Ptaki nad Łodzią*, „Express Ilustrowany”, 2003-04-12).

Nie we wszystkich powyższych cytatach chodzi o podane w przywołanej definicji „rozbijanie, tłuczenie na miazgę” (zob. (3)); istotne jest jednak to, że przedmiotem działania nazywanego czasownikiem *masakrować* są w tym wypadku nie ludzie, lecz obiekty materialne. Proces rozwoju semantycznego omawianej jednostki idzie jednak dalej i podobnie jak wiele innych leksemów, których pierwotne znaczenie dotyczy zjawisk fizycznych, wtórnie zaczyna ona służyć do nazywania zjawisk z innych dziedzin ludzkiego doświadczenia. Materiał zawarty w NKJP (uzupełniony o materiały pozyskane z Internetu) pozwala na wyodrębnienie kilku przenośnych znaczeń czasownika *masakrować* (a także jego odpowiednika aspektowego *zmasakrować*[8]), nieuwzględnianych przez słowniki języka ogólnego[9]:

1. ‘powodować, że coś przestaje istnieć’, np.:

(4) Trwa proces świętowania powrotu Violetty Villas, kompletnie *masakrujący* legendę artystki. W programie „Jak oni śpiewają” skończył się jej tlen, pamięć, a wcześniej dykcja (Kuba Wojewódzki, *Mea pulpa, czyli kronika popkulturalna Kuby Wojewódzkiego*, „Polityka”, 2008-11-22).

5 Nie przytaczam całości definicji, a jedynie te ich fragmenty, które są tu istotne.

6 Formalnie rzecz biorąc, tylko w PSWP zostało ono wyróżnione jako osobne znaczenie, opatrzone odrębnym numerem. W pozostałych słownikach ten element semantyczny został podany w obrębie jednej definicji (ISJP) lub też wprowadzony po średniku (SJPSzym) albo jednostce *także* (SWJP, USJP).

7 Podaje je także SWO PWN.

8 Nie podaję cytatów z tą formą, aby nie pomnażać nadmiernie liczby przykładów; uwzględniam natomiast w poniższych przykładach formę gerundium *masakrowanie*.

9 Sprawdzone zostały: SJPSzym, SWJP, PSWP, ISJP, USJP, a ponadto SSmołk oraz SPLP. W tym ostatnim czasownik *masakrować* odnotowany został w znaczeniu ‘bić, miażdżyć, niszczyć’, m.in. z przykładem: *masakrowane upałem miasto* (cytat skrócony). W WSJP na razie brak interesującego mnie tu hasła.

(5) Złośliwość tych drobnoustrojów polega na tym, że gdy cały organizm próbuje cieszyć się wiosną, radość ta skutecznie jest *masakrowana* przez nadmierną i uciążliwą pracę śluzówki oraz przez salwy gruźliczego kaszlu (NKJP: slonkoq, *G jak grypa*, „Pinezka.pl", 2005-05).

2. 'powodować, że coś zostaje zniekształcone albo też w inny sposób znacznie zmienione na gorsze' (w tym znaczeniu często używany w odniesieniu do niewłaściwie wykonywanego lub interpretowanego utworu literackiego, muzycznego itp.), np.:

(6) [...] Januszowski tekst samowolnie poprawiał, kilka wierszy opuścił, w dwóch [...] sens zmienił, a niejednokrotnie go zaprzepaścił. Dla przykładu, jak bardzo *masakrował* Januszowski tekst Kochanowskiego, przytaczam za Budką trzy wiersze według wydania z roku 1585 (W. Weintraub, *Jakiego Kochanowskiego znamy?*, „Język Polski" XV, 1930: 66–67[10]).

(7) Nawet partyjni redaktorzy pism dawali do zrozumienia, że „są z nami". Cenzor, który *masakrował* filmy Kieślowskiemu, mówił po cichu: tak naprawdę jestem z panem! (T. Sobolewski, *Twarze Kieślowskiego*, „Gazeta Wyborcza", 1997-03-08).

(8) Hoffman w podobnym *masakrowaniu* klasyków polskiej literatury ma wprawę – zrobił przecież takie „Ogniem i mieczem", w której Chmielnicki jest bohaterem pozytywnym, Wiśniowiecki krwawym psychopatą, a bohaterowie wyobraźni całych pokoleń Polaków zmienili się w zapitego menela, głupawego kurdupla i tępego dewota (NKJP: Internet).

(9) Jak już musisz pisać głupoty, to chociaż nie *masakruj* swojego tekstu piekielną interpunkcją (NKJP: Internet).

(10) Ta postać jest bardzo ostra. To potwornie *masakruje* moją psyche. Staram się zrozumieć tok myślenia prostytutki i jej klienta. Nie jestem mechanicznym aktorem (NKJP: P. Rąpalski, *Biała Dama wystąpi w filmie Almodovara*, „Gazeta Krakowska", 2008-08-07).

(11) Partia ma dość kłopotów na poziomie państwa, radny z wyrokiem *masakruje* jej wizerunek na poziomie miasta i powiatu [...] (NKJP: A. Pleś, *Rezygnacja z przytupem*, „Gazeta Krakowska", 2007-07-20).

3. 'działać w sposób bardzo niekorzystny dla kogoś':

(12) Działający w sąsiedztwie supermarketu „Max" mówią, że konkurencja ich *masakruje*, bo ich utarg spadł aż o 70 proc. – informuje urzędnik (NKJP: J. Cyrus, *Markety atakują!*, „Dziennik Zachodni", 2001-02-14).

(13) Widziałem wspaniałych ludzi. Ale urzędnicy ich po prostu *masakrowali*. [...] Nie mieli naturalnego odruchu sięgania po rewolwer, kiedy urzędnik powoływał się na byle jaki przepis, żeby zamknąć im usta (NKJP: Cz. Kielecki, *Scenarzysta*, 2009).

10 Drugie zdanie cytatu zostało przytoczone w SJPDor s.v.; zgodnie z praktyką tego słownika cytat podany po kwalifikatorze *przen.*, bez definicji.

4. ‘osiągać dużą przewagę w rywalizacji sportowej lub innego rodzaju’ (czasownik *masakrować* w tym znaczeniu ma swój odpowiednik rzeczownikowy – *masakra* ‘wysoka porażka jednej z rywalizujących ze sobą (np. w zawodach sportowych) osób lub drużyn’ (Słabczyński 2015: 245)), np.:

(14) Wisła *masakruje* Górnik pod Wawelem![11] (http://www.futbol.pl/news/wisla-masakruje-gornik-pod-wawelem).

(15) Ponoć Estonia produkuje jakieś 3 filmy rocznie, z których wszystkie *masakrują* konkurencję na festiwalach (NKJP: Internet).

(16) Nowy sondaż. PiS *masakruje* PO (http://www.se.pl/wiadomosci/polityka/nowy-sondaz-pis-masakruje-po_635958.html, dostęp: 11 kwietnia 2016 r.).

5. ‘bardzo ostro kogoś lub coś krytykować’[12], np.:

(17) Ostatnie sztuki Tennessee Williamsa były przez krytyków w Nowym Jorku *masakrowane* (K. Bielas, J. Szczerba, *Nic tak dobrze nie robi pisarzowi jak upokorzenie*, „Gazeta Wyborcza”, 1998-10-09).

(18) Formuła programu jest niesprawiedliwa, wiem. Na tym polega jego sukces. Zaprasza się dwóch autorów do rozmowy o ich dziełach, później pierwszego *się masakruje*, zaś drugiego koronuje (B. Jędrasik, *Gorączka. Opowiadania wyuzdane*, 2007).

(19) Ze szczególną przyjemnością Wheen *masakruje* współczesnych filozofów francuskich uprawiających coś, co nazywają refleksją postmodernistyczną, a jego zdaniem jest zwykłym bełkotem i zdradą wielkiego dziedzictwa Rousseau, Condorceta, Woltera (E. Bendyk, *Tryumf głupoty*, „Polityka”, 2006-01-07).

(20) Podobną wizję historii tworzyli „maleńcy uczeni” „profesora” Majewskiego z „Syzyfowych prac”, *masakrując* w swych wypracowaniach nieszczęsną Polskę jako „gniazdo rozbestwionej szlachty mordującej lud ruski przy akompaniamencie okrzyków psiakrew i psiadusza” (NKJP: Internet).

W pokazanych wyżej znaczeniach czasownik *masakrować* konotuje grupę nominalną w funkcji dopełnienia. W takiej konstrukcji składniowej, typowej dla niego także w znaczeniu podstawowym, używany jest on najczęściej. Bez dopełnienia występuje we frazie *cena masakruje* / *ceny masakrują*, odnotowanej w wypowiedziach internautów, np.:

11 W cytatach pochodzących z Internetu poprawiona została interpunkcja oraz ortografia.

12 Por. następujące znaczenie rzeczownika *masakra* wyróżnione przez R. Słabczyńskiego (2015: 245): ‘wypowiedź, w której nadawca negatywnie ocenia jakieś osoby albo zjawiska, przy tym wyrażona „w sposób arogancki, bezwzględny i dosadny” (np. *najlepsze masakry Korwina*).

(21) Może i produkt dobry, ale *cena masakruje* (https://imagazine.pl/2013/06/28/mophie-juice-pack-plus-120-wiecej/).

(22) Menu wygląda całkiem dobrze i faktycznie *ceny nie masakrują* (www.maltaigozo.pl/bad-bull-bar-b-que-house-najlepsza-baza-bugibba-square/).

Frazę tę, mającą swój odpowiednik w wyrażeniu *masakra cenowa*, odnoszącym się do wysokiej ceny jakiegoś produktu (Słabczyński 2015: 243), można interpretować jako konstrukcję eliptyczną w stosunku do kolokacji *cena masakruje* (*czyjąś*) *kieszeń* / *portfel* (także poświadczonej w wypowiedziach internautów). Druga możliwa interpretacja wiąże się z funkcjonowaniem czasowników *masakrować* i *zmasakrować* jako jednostek wyrażających ocenę poprzez informowanie o reakcji podmiotu wartościującego na dany obiekt, np.:

(23) Jestem *zmasakrowany* cenami, szczególnie ogrzewanie centralne mnie rozwaliło mentalnie (forum.muratordom.pl).

(24) Nie miałem pojęcia, jaka to płytka kobieta. Jestem *zmasakrowany* jej wypowiedziami (kuba.tvn.pl/aktualnosci,836.../paranienormalni-i-luxuria-astaroth-u-kuby,116261.htm).

(25) Jestem *zmasakrowany* łatwością robienia poprawnych zdjęć (interaktywnie.com/forum/grupy-dyskusyjne/slajdy-1534361.html).

(26) Obejrzałem pierwszy odcinek i jestem *zmasakrowany*. Dziękuję tym, dzięki którym się o nim dowiedziałem (www.wykop.pl/tag/utopia/archiwum/2015-01).

(27) Zupa wręcz *masakruje* swym apetycznym wyglądem (veganfairytale.blogspot.com/2012/12/tyrolska-zupa-ziemniaczana.html).

(28) Na totalnym spontanie wyszło coś naprawdę fajnego! Trzepaliśmy ostatnio kuchnię czeską i wyszło nam coś zupełnie zwariowanego, ale kolizja smaków w tym zestawie po prostu *masakruje*! (czosnekichilli.pl/author/czosnekchilli/).

W powyższych wypowiedziach za pomocą omawianych czasowników nadawcy komunikują, że dany obiekt wywarł ((23)–(26)) lub wywiera ((27)–(28)) na nich silne wrażenie: negatywne ((23)–(24)) lub pozytywne ((25)–(28)), w ten sposób wyrażając nie wprost ocenę tego obiektu[13] (por. użycie w takiej samej funkcji czasowników *rozwalać* (zob. WSJP) oraz *zabić*[14]). Powiązanie z nacechowaniem pejoratywnym uznałabym za pierwotne, biorąc pod uwagę podstawowe znaczenie tych czasowników. Ich ambiwalencja aksjologiczna w przedstawionej funkcji analogiczna jest do tej, jaka cechuje rzeczownik *masakra*, który także oprócz informowania o negatywnych emocjach

13 W zdaniach (27)–(28), w których grupa nominalna nazywająca obiekt oceny pełni funkcję podmiotu, nie zostało eksplicytnie wyrażone, na kim wywiera on pozytywne wrażenie. Można jednak przyjąć, że chodzi tu o nadawcę, ewentualnie także o inne osoby, które z danym obiektem by się zetknęły.

14 „Malarskość tych prac i kolorystyka zabija! Gratulacje!" (max3d.pl/forum/threads/50187-Teczka-2d-Waldek!/page11).

(np. *Masakra, ale tu bałagan*) bywa używany do wyrażania oceny pozytywnej (np. *Masakra, co za koncert!*) (Słabczyński 2015: 241, 244).

Z czasownikiem *masakrować* oraz jego odpowiednikiem aspektowym *zmasakrować* można w ostatnim czasie często spotkać się w wypowiedziach dotyczących tematyki politycznej. Korzystając z wyszukiwarki Google[15], liczne przykłady tego rodzaju użyć można wyeksercerpować z tekstów publikowanych w Internecie. Omawiane jednostki występują w nich przeważnie w połączeniu z grupą nominalną, w której członem konstytutywnym jest rzeczownik nazywający osobę, ewentualnie grupę osób (partię polityczną itp.), np.: *Janusz Korwin-Mikke masakruje młodego lewaka*; *Przemysław Babiarz masakruje dziennikarkę „Newsweeka"*; *Grzegorz Braun masakruje dziennikarkę TVP1*; *Robert Winnicki masakruje u Lisa Andrzeja Celińskiego*; *Narodowiec masakruje prof. Andrzeja Zolla*; *prof. Gliński masakruje red. Żakowskiego*; *Kaczyński masakruje Neumanna*; *Kwaśniewski ostro masakruje Tomasza Lisa*; *młody poseł PiS masakruje Nowoczesną*; *profesor psychologii mocno masakruje młodego coacha*[16]; *Jarosław Gowin zmasakrował Balcerowicza*; *Winnicki zmasakrował obrażającego go posła*; *Ziemkiewicz zmasakrował Młynarskiego*; *Wojciech Cejrowski zmasakrował posła z Nowoczesnej*; zdarzają się również połączenia z grupami nominalnymi innego rodzaju, np. *R. Wójcikowski (Kukiz'15) masakruje projekt Swetru*[17] *dotyczący finansowania partii*. Choć zdecydowanie przeważają użycia, w których czasownik konotuje grupę nominalną, możliwe jest także jej pominięcie, np. *Korwin-Mikke masakruje w rosyjskiej telewizji*, a także rozbudowywanie podstawowej konstrukcji *ktoś masakruje / zmasakrował kogoś* o dodatkowe elementy, np. *Internauta masakruje Betlejewskiego w temacie uchodźców*; *Rafał Ziemkiewicz masakruje lewaka o imigrantach*[18]; [Grzegorz Braun] *zmasakrował prowadzącą program dziennikarkę swą dosadną argumentacją*. Frazy z omawianym czasownikiem występują często jako tytuły artykułów w portalach internetowych lub nagrań udostępnianych na kanale YouTube, mając na celu – jako tak wyeksponowany element komunikatu – przyciągnąć uwagę odbiorcy, a zarazem stanowiąc zwięzłe ujęcie przedstawionych w nim treści.

We wszystkich przytoczonych frazach czasownik *masakrować* jest jednostką służącą nadawcy do referowania wypowiedzi osoby trzeciej. W tego rodzaju użyciach może on występować w znaczeniu zdefiniowanym wyżej jako 'bardzo ostro kogoś lub coś krytykować'. Dotyczy to nie tylko takich połączeń jak *masakrować projekt*, ale także niektórych kolokacji, w których omawiany wyraz konotuje grupę nominalną odnoszącą się do osoby lub grupy osób, np.

> (29) Cejrowski *masakruje* ekipę Kopacz: „Ten rząd sprowadza na nas bezpośrednie zagrożenie życia!" [nagłówek]. Wojciech Cejrowski jak zwykle nie owija w bawełnę! Przestrzega, że Europa przeżywa inwazję islamu – drugą fazę „dżihadu". A rząd Ewy Kopacz, który aprobował forsowaną przez Berlin politykę i zgodził się na przyjęcie fali emigrantów, nazywa władzą wyrodną, działającą na szkodę państwa i obywateli (http://wpolityce.pl/polityka/266376-cejrowski).

15 Kwerenda przeprowadzona w kwietniu–czerwcu 2016 r.

16 Przykład niezwiązany z tematyką polityczną.

17 Chodzi o Ryszarda Petru, lidera partii Nowoczesna.

18 Wyrażenie *w temacie czegoś* uznawane jest za niepoprawne (WSPP). Pozytywnej oceny normatywnej nie zyskałaby też z pewnością konstrukcja *masakrować o czymś*.

Jako osobną grupę należy wyodrębnić te użycia czasownika *masakrować*, w których służy on do referowania wypowiedzi formułowanych w trakcie dyskusji. Poniżej przedstawiam odpowiednie przykłady, przytaczając jednak nie tylko zdania, których omawiana jednostka jest komponentem, ale także wypowiedzi, które za jej pomocą były referowane – są to wprawdzie konteksty dość obszerne, jednak potrzebne do ukazania semantyki analizowanej jednostki: skoro służy ona do referowania pewnych treści, konieczne jest pokazanie, czym treści te się charakteryzują.

(30) Janusz Korwin-Mikke *masakruje* młodego lewaka [tytuł filmu na kanale Youtube]. Korwin-Mikke [odpowiadając dyskutantowi stawiającemu za wzór kraje, w których obywatele najmniej boją się o swoje finanse]: Proszę pana, spośród dwóch armii, z których jedna wygrywa wojnę, a druga siedzi w pieleszach, w tej pierwszej oni się bardziej boją o swoje życie, ale oni zwyciężają. Bo widzi pan, celem życia nie jest przeżycie. Dyskutant: A co? Korwin-Mikke: To ja przepraszam, nie będę zastępował pańskich nauczycieli. Ludzie mają swoje cele, mają swoje ambicje, mają swoje ideały i za to walczą, i umierają, a pan chce tylko żyć! To niech pan sobie żre i żyje dalej... (https://www.youtube.com/watch?v=9P2QTdS9ItU).

(31) Przemysław Babiarz *masakruje* dziennikarkę „Newsweeka" [tytuł artykułu]. [...] – To krótka pamięć – ten sposób przedstawiania sytuacji. Znaczy... urodziliśmy się wczoraj wszyscy, tak? Nie pamiętamy, jak wyrzucano tych samych dziennikarzy parę lat temu, nie pamiętamy, jak manipulowano TK [Trybunałem Konstytucyjnym] w maju... Wtedy Komisja Europejska i Europa nie pochylała się z troską, a dziś się pochyla... – mówił Przemysław Babiarz.
Dziennikarka „Newsweeka" próbowała jeszcze polemizować. [...] Riposta Przemysława Babiarza pozbawiła dziennikarkę „Newsweeka" argumentacji.
– Różnica jest taka, co próbował powiedzieć Cezary. Różnica nie polega na tym, że mamy spór, powiedzmy, niemiecko-polski, tylko różnica w zależności od tego, czy my mamy poglądy lewicowo-liberalne czy konserwatywne – mówił Przemysław Babiarz.
– Nie – odparła Renata Kim.
– Ale właśnie o to chodzi. Proszę zwrócić uwagę, że politycy niemieccy z konserwatywnej opcji bardzo podobnie do polskich polityków PiS widzą ten problem. To jest różnica ideologiczna – zakończył Przemysław Babiarz (http://niezalezna.pl/75302-przemyslaw-babiarz-masakruje-dziennikarke-newsweeka-co-jej-powiedzial).

(32) Robert Winnicki *masakruje* u Lisa Andrzeja Celińskiego [tytuł filmu na Youtube]. Celiński: Kim jest prezydent najważniejszego w NATO kraju, o najsilniejszej armii, największych zdolnościach obronnych? Czasem to nie jest Barack Obama? Winnicki: A co to ma do rzeczy? Celiński: Taki etniczny Amerykanin? Prawda? To ma [słowo niezrozumiałe] do rzeczy, że wtedy, kiedy rozumie się europejski korzeń kulturowy... Winnicki: Etniczni Amerykanie to Indianie, panie ministrze. Dawno nie było prezydenta Indianina (https://www.youtube.com/watch?v=RyNbAnOvsRw).

(33) Narodowiec *masakruje* prof. Andrzeja Zolla [tytuł filmu na YouTube]. Narodowiec: Panie profesorze, czy chciałby pan żyć w kraju, który nie jest w Unii Europejskiej, w którym nie ma

Trybunału Konstytucyjnego, w którym jest powszechny dostęp do broni? Zoll: Nie. Narodowiec: Czyli uważa pan, że demokracja w Szwajcarii jest zagrożona, bo te wszystkie rzeczy są w Szwajcarii, a jakoś tam [końcówka niezrozumiała] (https://www.youtube.com/watch?v=RZ8nSofijK8).

(34) Jarosław Gowin *zmasakrował* Balcerowicza – tytuł informacji w portalu Fronda, w której zostały zacytowane poniższe wpisy z Twittera:
– W ciągu paru miesięcy Polska spadła z 18 na 47 miejsce w rankingu Wolności Prasy. Czy @Jaroslaw_Gowin to zauważył? – napisał na Twitterze Leszek Balcerowicz.
– @LBalcerowicz to podsumowanie 2015. Nowy rząd istniał przez 1/10 roku. Czyli my odpowiadamy za spadek o 3 miejsca, a PO-PSL o 26. Zgadza się? – odpisał Jarosław Gowin. (http://www.fronda.pl/a/jaroslaw-gowin-zmasakrowal-balcerowicza-to-podsumowanie-2015-nowy-rzad-istnial-przez-110-roku,70631.html).

(35) Kwaśniewski ostro *masakruje* Tomasza Lisa!!! [tytuł filmu na YouTube]. Lis: [...] nie za wcześnie Pan przestał być prezydentem. Pięćdziesiąt parę lat. Kwaśniewski: Jakie parę? Lis: No parę. Kwaśniewski: Jakie parę? Lis: No pan jest czterdziesty szósty? Kwaśniewski: Co czterdziesty szósty? Lis: Rocznik. Nie, zaraz, w momencie wyboru miał pan 43, czyli plus 10, to jest 53, no 54, 5 lat. Kwaśniewski: Panie redaktorze, a ja miałem o panu takie dobre zdanie. [...] Ja jestem urodzony w listopadzie pięćdziesiątego czwartego roku. Będąc wybrany [...], miałem 41 lat. Zakończyłem kadencję po 10 latach, mając lat 51 (https://www.youtube.com/watch?v=mZtbusqdYhA).

(36) Ja lubię, jak pan Wiklina inteligentnie *masakruje* oponenta (NKJP: Internet).

(37) Jak szybko *zmasakrować* rozmówcę [historyjka obrazkowa zakończona radą: „Bezsensowny, wymyślony na poczekaniu cytat to najlepszy sposób na wygranie kłótni"] (http://www.obrazki.jeja.pl/94364,jak-szybko-zmasakrowac-rozmowce.html).

Wypowiedzi referowane za pomocą omawianych jednostek formułowane są najczęściej w trakcie rozmowy, mogą jednak również stanowić element polemiki prowadzonej w innej sytuacji komunikacyjnej, np. na Twitterze (zob. (34)) lub na sali sejmowej, kiedy posłowie nie prowadzą ze sobą bezpośrednio dyskusji, ale mogą w swoich wystąpieniach odnosić się do innych przemówień wygłoszonych podczas posiedzeń. W takim kontekście osadzone jest na przykład zdanie *Młody poseł PiS masakruje Nowoczesną*, podsumowujące sejmowe wystąpienie, w którym podważony zostaje zarzut – sformułowany przez partię Nowoczesna wobec PiS – o niezgłaszanie projektów ustaw przez to ugrupowanie. „Młody poseł PiS" odpiera zarzut, podając liczbę projektów wysuniętych przez jego partię od początku kadencji Sejmu oraz zaznaczając, że Nowoczesna sama nie wykazała się w tym czasie ani jedną inicjatywą ustawodawczą.

W przedstawionych frazach znaczenie czasownika *masakrować*, choć bliskie znaczenia 5., jest w stosunku do niego zmodyfikowane ze względu na charakter referowanych wypowiedzi.

Sytuacja komunikacyjna, w jakiej wypowiedzi te funkcjonują, sprawia, że za ich wspólną i najważniejszą cechę należy uznać cel, w jakim są formułowane, a którym jest wykazanie przez nadawcę niezgodności z prawdą podanych przez dyskutanta faktów lub też niesłuszności wyrażanych przez niego poglądów. Przegląd wypowiedzi referowanych za pomocą czasownika *masakrować* pokazuje, że użytkownicy języka sięgają po tę jednostkę przede wszystkim w odniesieniu do wypowiedzi, w których krytyka oponenta prowadzi do jego ośmieszenia, które zwracają uwagę polemiczną sprawnością wypowiadającej się osoby, lub też takich, które można określić jako dosadne, ostre. Nie muszą natomiast być to wypowiedzi brutalne, agresywne czy też z innego powodu naruszające normy prowadzenia dyskusji; chociaż niektóre takie bywają, nie jest to ich cecha dystynktywna. Podobnie też – choć w odniesieniu do wielu przykładów czasownik *masakrować* można uznać za synonim jednostek *ośmieszać* czy *kompromitować* – narażania kogoś na wstyd albo śmieszność nie uważałabym za element niezbędny, aby można było mówić o *masakrowaniu*. Dla analizowanych tu użyć jednostki *masakrować / zmasakrować kogoś* proponowałabym zatem następującą definicję: ‘wykazywać / wykazać nieprawdziwość informacji podawanych przez dyskutanta lub niesłuszność wyrażanych przez niego poglądów, często w sposób dosadny lub prowadzący do jego ośmieszenia’. Znaczenie to interpretowałabym jako swego rodzaju kontaminację wyróżnionych wcześniej znaczeń 5. (element krytyki, negatywnej oceny) oraz 4. (*zmasakrować kogoś* to ‘odnieść zwycięstwo nad dyskutantem’[19], traktowanym w kategoriach przeciwnika do pokonania).

Biorąc pod uwagę fakt, że w wielu wypadkach nacechowanie aksjologiczne leksemu w znaczeniu podstawowym zachowywane jest także wówczas, gdy występuje on w znaczeniach wtórnych, podkreślić trzeba, że w przytoczonych użyciach czasownik *masakrować* nie wyrażał negatywnego wartościowania czynności, którą nazywał. Były to użycia, które można interpretować jako neutralne, opisowe, albo też stanowiły one element tekstów, w których wyrażano pozytywną ocenę referowanej wypowiedzi, por. np.

(34) Patryk Jaki tak *masakruje* Kijowskiego, że ten musi ratować się łykiem wody! [tytuł] Ależ uderzenie! Wiceminister sprawiedliwości znany z ciętego języka, dobitnie udowadnia liderowi KOD Mateuszowi Kijowskiemu, gdzie jest jego miejsce, bo z pewnością nie w roli zbawcy narodu. [...] Mateusz Kijowski kolejny raz udowodnił, że nie posiada elementarnej wiedzy o Polsce i temacie, o którym będzie rozmawiał w mediach (http://www.obalamyglupote.pl/2016/02/patryk-jaki-tak-masakruje-kijowskiego.html).

Również w takich kontekstach istnieje jednak powiązanie z negatywnym wartościowaniem: czynność nazywana przez czasownik *masakrować* nie jest wprawdzie oceniana negatywnie przez nadawcę, ale jest ona niekorzystna (zła) dla tego, kto jej podlega (dla osoby nazywanej przez grupę nominalną pełniącą funkcję pacjensa).

Omawiane jednostki mogą występować także w konstrukcjach, w których wykonawca czynności jest jednocześnie odbiorcą jej skutków – *ktoś masakruje / zmasakrował sam siebie /*

19 Ten aspekt wyraźnie pokazuje przykład (37), w którym *zmasakrować* oznacza ‘wygrać kłótnię’.

samego siebie (np. *Petru masakruje sam siebie*, *Kukiz masakruje samego siebie*). Do wyrażenia podobnych treści służą także formy zwrotne, przy czym różnica pomiędzy nimi a formami niezwrotnymi wykracza w tym wypadku poza to, co wynika z samego faktu wystąpienia czasownika w stronie zwrotnej. W zebranych przeze mnie przykładach formy niezwrotne nazywały jedynie zachowania werbalne, te drugie nie podlegały zaś takim ograniczeniom (por. np. *masakrować się swoim życiorysem* (40)), co sprawia, że można uznać je za synonimy czasowników *ośmieszać/ośmieszyć się*, *kompromitować/skompromitować się*, por.:

(38) Piotr Semka *masakruje się* sam na antenie Polskiego Radia. Wg tego pana burdy kiboli były wynikiem działań policji – tymczasem pierwsza ofiara została zabrana przez karetkę o 15:50 na skutek bójki pomiędzy uczestnikami marszu bez najmniejszego udziału policji (http://www.wykop.pl/link/2239688).

(39) Obejrzyjcie i posłuchajcie, jak poseł Winnicki *zmasakrował się* z moją pomocą [wpis Joanny Senyszyn na Twitterze] (https://twitter.com/joanna_senyszyn/status/726347324998467584).

(40) Petru *masakruje się* sam swoim życiorysem [komentarz pod filmem zatytułowanym *Korwin masakruje Petru*] (https://www.youtube.com/watch?v=6b6kWlK3wf4).

(41) Grzegorz Braun *zmasakrował się* sam własną głupotą życiową (http://f.kafeteria.pl/temat/f1/grzegorz-braun-zmasakrowal-sie-sam-wlasna-glupota-zyciowa-p_6224256).

Przedstawione w artykule nowe znaczenia czasownika *masakrować* bazują na mechanizmie metafory. Najbliższe znaczenia podstawowego są znaczenia wyróżnione wyżej jako 1. i 2. Ich wspólnym elementem jest rezultat działań nazywanych za pomocą czasownika *masakrować*: to, że jakiś obiekt przestaje istnieć, zostaje pozbawiony swego pierwotnego kształtu lub w jakiś inny sposób zmieniony na gorsze, przy czym w domenie źródłowej odnosi się to do sytuacji zabijania i okaleczania ludzi, w domenie docelowej – do zjawisk mających niematerialny charakter. Przedmiotem działań nazywanych omawianym czasownikiem bywają również przedmioty materialne (zob. (1)–(3)); takie użycia (np. *masakrować chodniki*) można traktować jako stadium pośrednie między połączeniami typu *masakrować ludzi* a kolokacjami takimi jak *masakrować legendę*. Jeśli chodzi o inne znaczenia, ich wspólnym elementem semantycznym, łączącym je ze znaczeniem podstawowym, jest to, że nazywają one działania dla kogoś lub dla czegoś niekorzystne[20]. Nie dotyczy to znaczenia ‘wywierać na kimś silne wrażenie: negatywne lub pozytywne’, którego elementem przejętym ze znaczenia pierwotnego jest fakt intensywnego oddziaływania – w tym wypadku nie na ciało człowieka, lecz na jego myśli i uczucia.

Duża skala działań czy też intensywność zjawisk nazywanych za pomocą omawianej jednostki istotna jest także w odniesieniu do pozostałych znaczeń. W wypadku znaczenia 4. chodzi

20 W odniesieniu do użyć z dziedziny sportu źródłem skojarzenia mógłby być boks jako dyscyplina, w której wygrana łączy się czasem ze zmasakrowaniem przeciwnika w sensie fizycznym. W polszczyźnie znaczenie ‘osiągać dużą przewagę w rywalizacji sportowej lub innego rodzaju’ może stanowić zapożyczenie semantyczne z języka angielskiego (zob. dalej).

o wygraną ze znaczną przewagą, nie o samą wygraną, znaczenia 5. zaś – o bardzo ostrą krytykę, a nie jedynie o wyrażenie negatywnej oceny. W wypadku znaczenia 2. wspomniany element semantyczny wyraźnie uwidacznia się przez porównanie z synonimicznymi jednostkami, np. *niszczyć*, *psuć*, *szkodzić*, które takiego komponentu nie zawierają (por. *niszczy*, *psuje*, *szkodzi w niewielkim stopniu* vs. **masakruje w niewielkim stopniu*).

W odniesieniu do leksemu *zmasakrować* warto jeszcze wspomnieć, że nie tylko wytworzył on nowe znaczenia czasownikowe, ale także w formie imiesłowu przymiotnikowego biernego uległ procesowi adiektywizacji, nabierając znaczenia ‘bardzo zmęczony’, np.:

(42) Jestem *zmasakrowany* po 14 godzinach pracy (https://www.facebook.com/PrzeciwnicyElzbietyHasko).

(43) Jestem *zmasakrowany* i padnięty, a muszę jeszcze się spakować i przygotować parę rzeczy (www.wykop.pl/wpis/18687729/).

Rozwój znaczeniowy podobny do tego, jaki został tutaj opisany w odniesieniu do polskiego czasownika, można także obserwować w wypadku fr. *massacrer* oraz ang. *massacre* (również zapożyczenia z języka francuskiego (OED IX: 436)). Stwierdzenie jednak (formułowane w kategoriach nie tyle pewności, ile prawdopodobieństwa), czy mamy tu do czynienia z niezależnym od siebie, analogicznym rozwojem, czy też z wpływem tych języków na polszczyznę, wymagałoby analiz nieograniczających się jedynie do zestawienia haseł słownikowych, ale także biorących pod uwagę uporządkowane chronologicznie i porównywalne dane korpusowe. Język francuski przywołuję tutaj jako język, z którego wywodzi się etymon polskiej jednostki; o jego wpływie można oczywiście mówić jedynie w odniesieniu do starszych, dawniej odnotowanych znaczeń (por. przykład (6) z 1930 roku). W wypadku znaczeń nowszych nie można natomiast zignorować faktu silnego oddziaływania języka angielskiego na współczesną polszczyznę, w tym na słownictwo sportowe.

Poniżej przedstawiam te znaczenia jednostek francuskich i angielskich niezwiązane z zabijaniem i okaleczaniem ludzi, które mają odpowiedniki we wtórnych znaczeniach czasownika polskiego[21]:

– fr. *massacrer*: ‘abîmer, détruire avec brutalité, avec violence’[22] (TLF); ‘mettre (une chose) en très mauvais état’[23] (Le Grand Robert); ang. *massacre*: ‘mangle, mutilate’[24]: *some people who ordinarily massacre grammar*[25] (Webster’s Dictionary);

21 W zakresie języka francuskiego zostały sprawdzone następujące słowniki: TLF (słownictwo z okresu 1789–1960), Le Grand Robert, Grand Larousse, Le Petit Robert; w zakresie języka angielskiego – Webster’s Dictionary, OED, RHUD, LDCE oraz słowniki w wersji internetowej: www.merriam-webster.com/dictionary; http://www.dictionary.com/browse/massacre?s=t. Dane znaczenie cytuję za tym słownikiem, który pokazuje je jako pierwszy.

22 ‘brutalnie, gwałtownie uszkadzać, niszczyć’.

23 ‘powodować, że jakaś rzecz znajduje się w bardzo złym stanie’.

24 Podane synonimy odsyłają do innych haseł słownika (*mutilate* ‘okaleczać (przen. ‘zniekształcać, cenzurować)’; *mangle* ‘pokiereszować, poharatać (przen. ‘zniekształcić, np. artykuł, tekst’), za: *Nowy słownik Fundacji Kościuszkowskiej angielsko-polski*, red. J. Fisiak, Universitas, Kraków 2003).

25 Ludzie, którzy zwykle masakrują gramatykę.

– fr. *massacrer*: ‘gâter de façon irréparable’: *Cette vie est trop courte pour que nous la massacrions de vains regrets*[26] (TLF);

– fr. *massacrer*: ‘gâter involontairement par un travail maladroit ou brutal, par une interprétation ou une exécution très mauvaise’: *massacrer un rôle*[27] (TLF); ang. *massacre*: ‘to do (something) very badly, to ruin (something) because of lack of skill’: *He really massacred that song*[28] (www.merriam-webster.com/dictionary);

– ang. *massacre*: informal. ‘to defeat decisively, esp. in sports’[29] (RHUD);

– fr. *massacrer*: ‘critiquer avec violence’: *Les critiques ont massacré son roman*[30] (Le Petit Robert).

Pokazując jako materiał porównawczy powyższe zbieżności pomiędzy znaczeniami polskiej jednostki a znaczeniami jej francuskiego i angielskiego odpowiednika, należy jednak zaznaczyć, że przedstawiony w artykule rozwój semantyczny czasownika *masakrować* możliwy jest także do wytłumaczenia na gruncie rodzimym.

Bibliografia wraz ze stosowanymi w pracy skrótami

Bochnakowa (red.) 2012: *Wyrazy francuskiego pochodzenia we współczesnym języku polskim*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.

Grand Larousse: *Grand Larousse de la langue française*, sous la direction L. Guilbert, R. Lagane, G. Niobey, vol. 4, Larousse, Paris 1989.

ISJP: *Inny słownik języka polskiego*, red. M. Bańko, t. 1–2, Wydawnictwo Naukowe PWN, Warszawa 2000.

LDCE: *Longman Dictionary of Contemporary English*, third edition with the New Words supplement, Longman, Harlow 2001.

Le Grand Robert: *Le Grande Robert de la langue française. Dictionnaire alphabétique et analogique de la langue française*, deuxième édition entièrement revue et enrichie par A. Rey, vol. 6, Le Robert, Paris 1989.

Le Petit Robert: *Le Petit Robert. Dictionnaire alphabétique et analogique de la langue française*, nouvelle édition du Petit Robert de Paul Robert, sous la direction de J. Rey-Debove, A. Rey, Paris 2013.

NKJP: *Narodowy Korpus Języka Polskiego* (online: http://nkjp.pl/).

OED: *Oxford English Dictionary*, second edition, prepared by J.A. Simpson, E.S.C. Weiner, vol. 9, Clarendon Press, Oxford 1989.

PSWP: *Praktyczny słownik współczesnej polszczyzny*, red. H. Zgółkowa, t. 1–50, Kurpisz, Poznań 1994–2005.

RHUD: *Random House Unabridged Dictionary*, second edition, ed. S.B. Flexner, Random House, New York 1993.

SJPDor: *Słownik języka polskiego*, red. W. Doroszewski, t. 1–11, Państwowe Wydawnictwo Naukowe, Warszawa 1958–1969.

SJPSzym: *Słownik języka polskiego*, red. M. Szymczak, t. 1–3, wyd. 8, Wydawnictwo Naukowe PWN, Warszawa 1993.

Słabczyński R. 2015: Masakra – *wyraz modny współczesnej polszczyzny*, „Język Polski” XCV, s. 237–249.

SPLP: *Słownik polskich leksemów potocznych*, red. W. Lubaś, t. 1–8, Lexis, Kraków 2001–2015.

SSmołk: *Słowa, słowa... Czy je znasz?*, red. T. Smółkowa, Instytut Języka Polskiego PAN, Kraków 2013.

SW: *Słownik języka polskiego*, red. J. Karłowicz, A. Kryński, W. Niedźwiedzki, t. 1–8, nakładem prenumeratorów i Kasy im. Mianowskiego, Warszawa 1900–1927.

SWJP: *Słownik współczesnego języka polskiego*, red. B. Dunaj, Wilga, Warszawa 1996.

SWO PWN: *Słownik wyrazów obcych PWN*, red. J. Tokarski, Państwowe Wydawnictwo Naukowe, Warszawa 1980.

26 ‘psuć w nieodwracalny sposób’: To życie jest zbyt krótkie, żebyśmy je masakrowali próżnymi żalami.

27 ‘psuć niezamierzenie przez niezręczną pracę, przez bardzo złe wykonanie lub interpretację’: masakrować rolę.

28 ‘robić coś bardzo źle, niszczyć coś z powodu braku umiejętności’: Naprawdę zmasakrował tę piosenkę.

29 pot. ‘zdecydowanie pokonać, zwłaszcza w sporcie’; znaczenie to notują także wymienione wyżej słowniki internetowe.

30 ‘gwałtownie krytykować’: Krytycy zmasakrowali jego powieść.

TLF: *Trésor de la langue française. Dictionnaire de la langue du XIX^e et XX^e siècle (1789–1960)*, vol. 1–16, Gallimard, Paris 1971–1994.
USJP: *Uniwersalny słownik języka polskiego*, red. S. Dubisz, t. 1–4, Wydawnictwo Naukowe PWN, Warszawa 2003.
Webster's Dictionary: *Webster's Third New International Dictionary of the English Language Unabridged*, ed. Ph.B. Gove, vol. 2, Encyclopaedia Britannica, William Benton, Chicago [etc.] 1971.
WSJP PAN: *Wielki słownik języka polskiego PAN*, red. P. Żmigrodzki (online: http://www.wsjp.pl/).
WSPP: *Wielki słownik poprawnej polszczyzny*, red. A. Markowski, Wydawnictwo Naukowe PWN, Warszawa 2005.

Summary

New uses of the verb *masakrować* in contemporary Polish

Keywords: contemporary Polish, lexis, semantic development of words.

The article presents the semantic development of the verb *masakrować* in contemporary Polish. Meanings of this lexeme not recorded in dictionaries of Polish have been discussed. Particular attention has been paid to the popularity it is gaining in texts regarding political topics where it is used to report words of a third party (in sentences like *Korwin masakruje lewaka*).

ZE ZJAWISK WSPÓŁCZESNEGO JĘZYKA

BOGUSŁAW DUNAJ* | PAŃSTWOWA WYŻSZA SZKOŁA ZAWODOWA W TARNOWIE

Odmiana i rodzaj gramatyczny rzeczowników *show*, *talk-show* i podobnych

Zapożyczane z różnych języków do polszczyzny wyrazy podlegają adaptacji fonetycznej, graficznej oraz morfologicznej. Stopień i szybkość adaptacji w zakresie poszczególnych podsystemów zależy od różnych czynników. Najliczniejsze obecnie w języku polskim zapożyczenia angielskie są na ogół dość łatwo przystosowywane do polskiego systemu fleksyjnego. Dotyczy to przede wszystkim rzeczowników zakończonych na spółgłoski. Przejmowane do polszczyzny rzeczowniki angielskie o wygłosie spółgłoskowym są od razu włączane do odpowiednich paradygmatów fleksyjnych i kategorii rodzajowych, przy czym angielska postać graficzna takich form bywa najczęściej zachowana (przynajmniej przez pewien czas). Przykładowo można przytoczyć formy takie, jak: *deal*, *dealu*, *deale*, *dealów*; *hashtag*/*hasztag*, *hasztagów*, *hasztagami*; *hejt*, *hejtu*, *hejcie*, *hejtów*; *self-publishing*, *self-publishingu*; *spread*, *spready*, *spreadów*; *start-up*, *start-upy*; *target*, *targetu*, *targetów*; *think tank*, *think tanki*, *think tankami*; *tweet*, *tweety*.

Z tego prostego schematu wyłamuje się niewielka grupa rzeczowników zakończonych w języku angielskim na literę *w*, której w wymowie odpowiada spółgłoska *u̯*. Wyrazy te w języku polskim zachowują zazwyczaj w zapisie przejętą z pierwowzoru literę *w*, której w wymowie odpowiada najczęściej głoska *u̯*, rzadziej *v*. Do pierwszej grupy należą następujące przykłady: *chow-chow*, *know-how*, *peep-show*, *reality show*, *show* i *talk-show*. Drugą reprezentują zaledwie dwa wyrazy: *bungalow* i *ludlow* ‘rodzaj maszyny drukarskiej’. Ten ostatni przykład należy do terminologii specjalistycznej. Wymowa końcowej spółgłoski, zgodna z zapisem graficznym, sprawia, że dwa ostatnie wyrazy są przyporządkowane odpowiednim paradygmatom męskim i oczywiście odmieniane.

Inaczej rzecz przedstawia się z rzeczownikami, w których końcowe *w* wymawia się jako *u̯*. Mimo że w języku polskim istnieje spora grupa rzeczowników zakończonych na spółgłoskę *u̯*, będącą odpowiednikiem litery *ł* (np. *zapał*, *wał*, *anioł*, *dzięcioł*, *gruczoł*, *kół*, *stół*, *wół*, *artykuł*, *brokuł*), to zapewne rozbieżność pomiędzy wymową a pisownią w wyrazach takich jak *show*, *talk-show* w stosunku do polskich zwyczajów graficzno-fonetycznych powoduje, że są one zasadniczo nieodmienne. Nieodmienne są wyrazy *chow-chow*, *know-how*, *peep-show*, *reality show* i *show*. Również wyraz *talk-show* jest zwykle nieodmieniany. Tylko sporadycznie pojawiały się przykłady odmiany tego wyrazu, por. np. *w talk-showach* (SWJP II: 414). Takie przykłady skłoniły zapewne autorów cytowanego słownika, a także *Nowego słownika poprawnej polszczyzny* (NSPP: 1031) do dopuszczenia (jako rzadkiej) odmiany tego wyrazu. W późniejszych słownikach (np. ISJP II: 805; USJP IV: 16; WSO: 1136) traktuje się *talk-show* jako

* bogdun@gmail.com

formę nieodmienną. Formę odmienną spotyka się wyjątkowo współcześnie, np. *talk showu* (YouTube). Częstsza jest ona w zapisach ze spolszczoną grafią, np. *tok szoły* (www.tvshow.com.pl), *w tok szole* (www.fronda.pl; f.kafeteria.pl).

Zakończenie na spółgłoskę twardą omawianych wyrazów powinno jednoznacznie decydować o przyporządkowaniu do rodzaju męskiego. I rzeczywiście większość przykładów pojawiających się w tekstach ma taką charakterystykę rodzajową. Na uwagę zasługuje jednak fakt, że notuje się również zapisy, w których *show* i *talk-show* mają rodzaj nijaki, por.: *na takie show* (DzP 2016, nr 28); *Show zgromadziło najbardziej znanych wykonawców* (DzP 2016, nr 47); *interesujące show* (DzP 2016, nr 90); *to show* (DzP 2016, nr 113); *ostatnie, wspólne, krótkie show* (ŻnG 2016, nr 25); *show, które sprawi* (Tel 2016, nr 27); *To polityczne show* (NCz 2016, nr 11–12); *telewizyjne show* (GPC 2016, nr 105); *największe charytatywno-medialne show* (PN 2016, nr 29); *podniebne show mazuryairshow*; *nakręcić ośmieszające* [...] *tok szoł* (GP 2016, nr 15). Formę *to show* spotkałem nawet w artykule pewnego językoznawcy (w wersji przedłożonej do recenzji wydawniczej).

W tym miejscu nasuwa się pytanie, co powoduje, że wyrazy te bywają włączane do grupy rzeczowników nijakich. Wiązać to należy zapewne z faktem, że synonimiczna forma *widowisko* jest rodzaju nijakiego. I to ten czynnik semantyczny jest źródłem pewnego rozchwiania rodzajowego omawianych wyrazów.

Jeśli chodzi o aspekt poprawnościowy, to sytuacja w tym względzie jest zupełnie jednoznaczna. Za jedynie uprawnioną formę rodzajową należy uznać rodzaj męski. Takie rozstrzygnięcie znajdujemy w *Nowym słowniku poprawnej polszczyzny* (NSPP: 912, 1031) i w innych słownikach. Zapisy, w których występuje rodzaj nijaki, musi się traktować jako błędy językowe.

Źródła

DzP – Dziennik Polski; GP – Gazeta Polska; GPC – Gazeta Polska Codziennie; NCz – Najwyższy Czas!; PN – Polska Niepodległa; Tel – Tele Tydzień; ŻnG – Życie na Gorąco.

Słowniki

ISJP: *Inny słownik języka polskiego*, red. M. Bańko, t. 1–2, Wydawnictwo Naukowe PWN, Warszawa 2000.

NSPP: *Nowy słownik poprawnej polszczyzny*, red. A. Markowski, Wydawnictwo Naukowe PWN, Warszawa 1999.

SWJP: *Słownik współczesnego języka polskiego*, red. B. Dunaj, Wilga, Warszawa 1996, wyd. 2, t. 1–2, Wydawnictwo Przegląd Reader's Digest, Warszawa 2001.

USJP: *Uniwersalny słownik języka polskiego*, red. S. Dubisz, t. 1–4, Wydawnictwo Naukowe PWN, Warszawa 2003.

WSO: *Wielki słownik ortograficzny*, red. E. Polański, Wydawnictwo Naukowe PWN, Warszawa 2016.

RECENZJE

BARBARA ŚCIGAŁA-STILLER* | INSTYTUT JĘZYKA POLSKIEGO POLSKIEJ AKADEMII NAUK, KRAKÓW

Agnieszka Sieradzka-Mruk, *„Radość i nadzieja, smutek i trwoga" w nabożeństwie drogi krzyżowej. Wybrane aspekty ewolucji dyskursu religijnego w XX wieku na przykładzie leksyki dotyczącej uczuć*

KSIĘGARNIA AKADEMICKA, KRAKÓW 2016, S. 255

Katolicki dyskurs religijny od dawna cieszy się szczególnym zainteresowaniem wielu badaczy. Na tle dwudziestowiecznych wydarzeń historycznych (m.in. obu wojen światowych, komunizmu i jego upadku, II soboru watykańskiego) oraz stopniowych przemian socjokulturowych (m.in. rozpowszechnienia Internetu, rozwoju mediów) przeszedł on ewolucję na płaszczyźnie językowej i gatunkowej.

Teksty nabożeństwa drogi krzyżowej nie zostały jak dotąd poddane szczegółowej analizie językoznawczej. Monografia Agnieszki Sieradzkiej-Mruk doskonale wypełnia tę lukę, wzbogacając badania dotyczące problematyki komunikowania i wyrażania uczuć w języku, a także badania nad samym gatunkiem drogi krzyżowej.

Przedmiotem zainteresowania Autorki stały się konkretne teksty drogi krzyżowej pokazane w kontekście kulturowym i sytuacyjnym, będące przykładem realizacji gatunku drogi krzyżowej.

Badaczka poddała analizie ok. stu tekstów dróg krzyżowych wydanych przed II soborem watykańskim i ok. stu tekstów współczesnych, posoborowych. Praca składa się z trzynastu rozdziałów – rozdział pierwszy, o charakterze wprowadzającym, przybliża problematykę, prezentuje stan badań i metodę badawczą; rozdział drugi przedstawia najwyższy poziom pola leksykalno-semantycznego uczuć (tj. hiperonim innych nazw uczuć – *uczucie* i *czuć*/*poczuć*) oraz hiponimy wyodrębniające się ze względu na charakterystykę czasową i intensywność (tj. wyrazy *nastrój*, *wzruszenie*, *afekt*, *uniesienie*, *patetyczny*, *przejmować się*, *poruszyć*, *rozrzewnienie*, *rozczulać się*, *namiętność*); kolejne rozdziały (3–11) zawierają językową analizę materiału (pole radości, pole sympatii, pole miłości, pola dobroci, łaski i miłosierdzia, pole smutku, pole gniewu i złości, pole strachu, pole zdziwienia wraz z rozwijającymi je polami uczuć czy postaw pokrewnych); rozdział ostatni, podsumowujący, przedstawia wyniki przeprowadzonych badań. Monografię kończą wykaz źródeł cytowanych oraz indeks omawianych leksemów.

Jak podkreśla Autorka, ze względu na szczególną tematykę (męka, ukrzyżowanie i śmierć Jezusa Chrystusa) oraz bogactwo opisu przeżyć emocjonalnych w badanym materiale niezwykle ważną rolę odgrywa leksyka dotycząca uczuć, dlatego głównym przedmiotem swojej

* barbara_maria@poczta.fm

pracy czyni obserwację zmian zachodzących w systemie leksykalnym oraz semantyce nazw uczuć, starając się odpowiedzieć m.in. na pytania o nowe leksemy, sposoby wzbogacania zasobu leksykalnego, przesunięcia znaczeniowe z innych pól semantycznych, zapożyczenia z innych języków, zawężanie i poszerzanie znaczeń, nacechowanie emocjonalne i wartościujące analizowanych leksemów.

Badaczka, porównując nazwy uczuć w przedsoborowych i posoborowych tekstach nabożeństwa drogi krzyżowej, odwołuje się do koncepcji pól leksykalno-semantycznych oraz do analizy składnikowej (komponencjalnej, semowej), zakładającej podzielność znaczenia danego leksemu na prostsze składniki, atomy znaczeniowe. A. Sieradzka-Mruk nawiązuje także do teorii prototypu i podobieństwa rodzinnego.

Punktem wyjścia analizy zgromadzonego materiału jest założenie, że leksyka dotycząca nazw uczuć tworzy rozbudowaną siatkę, strukturę podporządkowaną głównemu hiperonimowi. Leksemy dotyczące uczuć można podzielić na dwa pola leksykalno-semantyczne: uczuć pozytywnych i negatywnych, w których obrębie wyróżniamy kolejne podpola semantyczne. Na peryferiach tego dychotomicznego podziału znajdują się leksemy, których nie można jednoznacznie sklasyfikować, na przykład *namiętność* może dotyczyć miłości lub nienawiści, natomiast *wzruszenie* może mieć charakter smutny lub radosny. Autorka każdorazowo wskazuje centralne pola oraz podpola semantyczne położone na ich peryferiach.

I tak w rozdziale 3 analizie zostały poddane nazwy uczuć pozytywnych niewymagających obiektu, które tworzą pole radości i uczuć pokrewnych. Centrum pola stanowi pole radości reprezentowane przez leksemy *radość*, *radosny*, *radośnie*, *radować się*, *cieszyć się*, *uciecha*, *pociecha*, *weselić się*, *wesele*, *wesoły*, *szczęście*, *szczęśliwy* oraz – częściowo nakładające się na pole radości – pole przyjemności reprezentowane przez leksemy *przyjemność*/*przyjemny*. Na peryferiach pola semantycznego radości znajdują się podpole ulgi i pociechy, nadziei i pokrewnych stanów psychicznych oraz pole dumy zawierające także pojęcia pokory, pychy i podobne.

W rozdziale 4 autorka bada nazwy uczuć pozytywnych skierowanych na obiekt, tzn. pole sympatii i uczuć pokrewnych. W centrum tego pola znajdują się leksemy *lubić* (*lubienie*) i *sympatia*, natomiast peryferie wypełniają leksemy z pola czułości i serdeczności, upodobania, podziwu i zachwytu, szacunku czy wdzięczności.

W rozdziale 5 badaczka przedstawia bogatą grupę leksemów z pola miłość, reprezentowaną przez leksemy: *miłość*, *kochać*, *ukochać*, *miłować*, *umiłować*, *miłowanie*, *kochanie*, *kochany*, *ukochany*, *miłosny*, *miłośny*, *miłośnie*, *miłośnik*, *rozkochać się*, *zakochać się*, *rozmiłować się*, *zamiłowanie*, *afekt*, a również bogate podpola leżące na peryferiach pola miłości, mianowicie pole uwielbienia i uczuć pokrewnych, pole przyjaźni i pole bliskości.

Rozdział 6 dotyczy grupy leksemów, które znajdują się na obrzeżu pola semantycznego sympatii i miłości i są związane z pozytywnymi uczuciami skierowanymi na jakiś obiekt oraz z postawami wynikającymi z uczuć, a przejawiającymi się w zachowaniu – *dobroć*, *łaska*, *miłosierdzie*.

Rozdział 7 został poświęcony uczuciom negatywnym, nieskierowanym na żaden obiekt. Centrum pola wypełniają leksemy związane ze smutkiem – *smutek*, *smutny*, *smutno*, *smutnie*, *smucić się*, *zasmucić*/*zasmucać*, natomiast na peryferiach leżą pola związane z żalem, goryczą

i rozczarowaniem, zniechęceniem i załamaniem, a także pole rozpaczy i beznadziejności, pole zmartwienia, przykrości oraz pole wstydu.

Na rozdział 8 składa się analiza leksemów dotyczących przeżyć względnie gwałtownych, krótkotrwałych, negatywnych, skłaniających do działania na szkodę obiektu wywołującego dane uczucie. Centrum pola semantycznego stanowią leksemy *gniew* i *złość*, z kolei na peryferiach znajdują się leksemy należące do pola wściekłości, zajadłości, irytacji i oburzenia.

Polu nienawiści, wrogości i niechęci, a więc uczuciom negatywnym wymagającym obiektu, mającym charakter długotrwałych postaw emocjonalnych, został poświęcony rozdział 9. Na peryferiach tego pola znajdują się leksemy związane z zazdrością i zawiścią, pogardą, wstrętem i uczuciami pokrewnymi.

Uczuciom negatywnym ukierunkowanym na obiekt poświęcono rozdział 10. W centrum pola znajdują się tutaj leksemy takie jak *strach*, *bać się*, *lęk*, *lękać się*, z kolei na peryferiach nazwy uczuć intensywnych z pola strach (*trwoga*, *trwożyć się*, *przerażenie*, *przerazić się*, *groza*, *zgroza*), ponadto nazwy uczuć mniej intensywnych z tegoż pola (*obawa*, *obawiać się*, *niepokój*, *niepokoić się*) oraz nazwy postaw – skłonności do odczuwania strachu (*bojaźliwy*, *tchórzostwo*, *tchórzliwy*), a także nazwy oznaczające brak strachu lub przezwyciężenie tego uczucia (m.in. *odwaga*, *śmiałość*, *męstwo*, *dzielność*, *bohaterstwo*).

Całość rozważań dopełnia rozdział 11 dotyczący leksemu *zdziwienie*, który oznacza reakcję na zdarzenie sprzeczne z naszymi oczekiwaniami. W centrum znajdują się leksemy takie jak *zdziwienie*, *dziwić się*, *dziwić*, *zdziwić się*, *dziwny* (*najdziwniejszy*), na peryferiach leżą wyrazy nacechowane stylistycznie, wskazujące na intensywność uczucia (*przedziwny*, *zadziwia*, *zdumienie*, *zaskoczenie*, *szok*, *szokować*, *wstrząs*, *podziw*, *rozczarowanie*).

Dzięki przeprowadzonej analizie, którą ilustruje bardzo bogaty materiał językowy, można zaobserwować przemiany językowo-stylistyczne, jakim uległ dyskurs drogi krzyżowej, zwłaszcza na płaszczyźnie semantycznej. Do najważniejszych zjawisk należy zawężanie znaczeń leksemów dotyczących uczuć (przechodzenie leksemów z ogólnego poziomu uczuć do konkretnego pola uczuć pozytywnych bądź negatywnych, na przykład w wypadku leksemów *czułość*, *serdeczność*, *współczucie* czy *szok*), przesunięcia znaczeniowe (np. z poziomu wyższego do niższego). Niekiedy zawężenie znaczenia (specjalizacja znaczeniowa) prowadzi do zaniku pewnych leksemów w analizowanych tekstach na przestrzeni lat, np. leksem *namiętność* zawęził swoje znaczenie do miłości erotycznej i nie jest już używany w stosunku do Jezusa, podobnie jak wyraz *miłośnik*, dziś używany jedynie z nazwą bytu nieosobowego w roli obiektu. Kolejnym zjawiskiem jest poszerzanie znaczeń, przechodzenie z leksyki wyspecjalizowanej do języka ogólnego (np. wyraz *sadyzm*, który z nazwy dewiacji seksualnej stał się synonimem okrucieństwa jako czerpania przyjemności z zadawania komuś bólu), a także nabieranie przez niektóre wyrazy negatywnego nacechowania emocjonalnego (np. *litość*, *politowanie*, *żałosny*, *ubolewać*). Na wzbogacenie pól leksykalno-semantycznych w najnowszym okresie wpływa wiele zjawisk, między innymi poszerzenie obszaru zainteresowań oraz potrzeba wyrażania znaczeń specyficznych. Dla ewolucji dyskursu drogi krzyżowej charakterystyczny staje się wzrost liczby nazw uczuć słabych przy jednoczesnym zubożeniu zasobu uczuć o wielkiej intensywności, na co niewątpliwie wpływają przemiany estetyczne, odejście od średniowiecznego doloryzmu

i barokowego upodobania do kontrastów. Przemiany dotyczą ponadto wartościowania niektórych uczuć (np. *rozczulanie*, *wzruszanie się* oceniane negatywnie czy uczucie rozpaczy traktowane coraz częściej jako grzech).

Zmiany nastąpiły również na płaszczyźnie gatunkowej, mianowicie intencją komunikacyjną badanych tekstów stała się zachęta do czynnej postawy wobec rzeczywistości (większa rola funkcji perswazyjnej), do szukania analogii między sytuacjami drogi krzyżowej a sytuacjami życiowymi, a nie, jak wcześniej, wzbudzanie poczucia winy (cierpienia, poniżenia, strachu, wstydu) prowadzącego do nawrócenia (mniejsza rola funkcji emotywnej). Ponadto pojawiają się nowe modele uczuć (radość, nadzieja zmartwychwstania, a także celowe pomijanie leksemów konotujących skojarzenia seksualne, takich jak *rozkosz*, *namiętność*, *oziębłość*), zmienia się również interpretacja symbolicznych wydarzeń i zachowań pasyjnych (np. Weronika przedstawiana jako symbol odwagi, a nie litości). Współczesny dyskurs przełamuje więc tradycjonalizm dominujący we wcześniejszych tekstach. Obrazy uczuć i zachowań oprawców stają się obecnie bardziej stonowane, natomiast – co warto podkreślić – przedmiotem uczuć negatywnych stają się osoby spoza świata przedstawionego, na przykład współcześni dziennikarze.

Ewolucja dyskursu drogi krzyżowej nastąpiła na tle zróżnicowanych zjawisk językowych i przemian kulturowych. Monografia A. Sieradzkiej-Mruk, ukazując bogactwo zmian semantycznych na przykładzie leksyki dotyczącej nazw uczuć, stanowi interesujące uzupełnienie i rozwinięcie dotychczasowych badań z zakresu dyskursu religijnego, a dokładnie pasyjnego, i otwiera tym samym nowe pole badań językoznawczych.

DYSKUSJE I POLEMIKI

JANUSZ S. BIEŃ* | KATEDRA LINGWISTYKI FORMALNEJ, UNIWERSYTET WARSZAWSKI

Status prawny uchwał ortograficznych

W jubileuszowym artykule Przewodniczącego Rady Języka Polskiego PAN czytamy:

> Istotne jest to, że w art. 13.1. [ustawy o języku polskim] wyposażono RJP w moc stanowiącą w zakresie ortografii i interpunkcji. Oznacza to, że zasady dotyczące pisowni mają moc prawnie obowiązującą, a nie są tylko zbiorem przepisów ustalanych przez gremia językoznawcze (Markowski 2016: 7).

Takie z pewnością były intencje projektodawców ustawy (w szczególności Walerego Pisarka), ale nie znaczy to, że uchwalona w 1999 roku ustawa rzeczywiście wprowadziła możliwość stanowienia prawa przez Radę Języka Polskiego. Treść przywołanego przez Andrzeja Markowskiego punktu ustawy jest następująca: na wniosek jednego z pięciu wyliczonych organów (w tym ministrów kultury, oświaty i szkolnictwa wyższego) lub z własnej inicjatywy Rada wyraża w formie uchwały (procedurę określa regulamin Rady – niezbędne jest formalne głosowanie i kwalifikowana większość głosów) opinie o używaniu języka polskiego:

1) w działalności publicznej,
2) w obrocie na terytorium Rzeczypospolitej Polskiej z udziałem konsumentów,
3) przy wykonywaniu na terytorium Rzeczypospolitej Polskiej przepisów z zakresu prawa pracy

oraz ustala zasady ortografii i interpunkcji języka polskiego.

Jak widać, nie ma w ustawie mowy o „prawnie obowiązującej mocy" uchwał.
Paweł Czarnecki w książce *Ustawa o języku polskim. Komentarz* pisze:

> W doktrynie zwraca się uwagę na zewnętrzny charakter uchwał w zakresie ortografii i interpunkcji, a zatem uchwały ortograficzne mają być stosowane w stosunku do wszystkich, którzy posługują się tym językiem. Nie sposób jednak stwierdzić, że uchwały mają wiążącą moc w stosunku do adresatów. Zarówno regulamin, jak też przepisy ustawy o języku polskim nie ustalają obowiązku publikacyjnego uchwał Rady, zatem niezależnie od zamkniętego katalogu źródeł prawa w rozumieniu art. 87 Konstytucji RP trudno wymagać, aby obywatele mieli obowiązek prawny się do nich stosować (wyróżnienia J.S.B.) (Czarnecki 2014: 160–161).

* jsbien@uw.edu.pl

Cytowany art. 87 Konstytucji pochodzi z rozdziału III *Źródła prawa* i stwierdza, że:

> Źródłami powszechnie obowiązującego prawa Rzeczypospolitej Polskiej są: Konstytucja, ustawy, ratyfikowane umowy międzynarodowe oraz rozporządzenia. Źródłami powszechnie obowiązującego prawa Rzeczypospolitej Polskiej są na obszarze działania organów, które je ustanowiły, akty prawa miejscowego.

Warto przypomnieć również art. 88:

> Warunkiem wejścia w życie ustaw, rozporządzeń oraz aktów prawa miejscowego jest ich ogłoszenie. Zasady i tryb ogłaszania aktów normatywnych określa ustawa.

P. Czarnecki w swych poglądach nie jest odosobniony. Radosław Mędrzycki pisał wcześniej:

> Stosowanie wytycznych i opinii wydawanych w formie uchwały opiera się raczej na autorytecie i powadze Rady; nie znajdziemy zatem podstawy prawnej uznającej opinie Rady za powszechnie wiążące (wyróżnienia J.S.B.) (Mędrzycki 2009: 193).

Nie jest to również pogląd nowy. Maciej Zieliński, profesor prawa, autor wielokrotnie wznawianego podręcznika *Wykładnia prawa. Zasady, reguły, wskazówki* (2017), członek Rady niemal od początku jej istnienia i od 2011 roku jej wiceprzewodniczący, na VII posiedzeniu plenarnym Rady w dniu 7 października 1999 roku mówił:

> Rada Języka Polskiego jest tylko instytucją opiniodawczo-doradczą, co oznacza, iż na wniosek zainteresowanych instytucji powinna ona wyrażać opinie i udzielać porad co do używania języka polskiego, lecz nikt nie ma obowiązku przestrzegania jej zaleceń (wyróżnienia J.S.B.) (Rada Języka Polskiego 1999).

Ten pogląd potwierdzają wyroki sądów, np. Wojewódzki Sąd Administracyjny w Lublinie stwierdził:

> Zalecenia Rady Języka Polskiego Komitetu Językoznawstwa [sic!] PAN nie są w żadnym rozumieniu aktem normatywnym. Stanowią jedynie szczególną formę opinii biegłego, a zatem nie mogą stanowić prawnej podstawy rozstrzygnięcia (por. art. 87 i następne Konstytucji) (wyróżnienia J.S.B.) (Wojewódzki Sąd Administracyjny w Lublinie 2004).

Ten i kilka innych podobnych wyroków cytuje R. Mędrzycki w swoim artykule z 2009 roku, od tego czasu mogły pojawić się nowe.

Swoją drogą Katarzyna Kłosińska, sekretarz Rady od 2003 roku do dzisiaj, pisała:

> Nie mamy – na szczęście – władzy nad językiem ani nad społeczeństwem; wydajemy jedynie opinie w sprawach, z którymi zwracają się do nas instytucje, firmy i osoby prywatne (wyróżnienia J.S.B.) (Sekretarz RJP 2003, s. 40).

Na koniec warto podkreślić, że ewentualne wątpliwości dotyczące prawnego statusu uchwał Rady w żaden sposób nie dotyczą uchwały ortograficznej nr 1, która została podjęta przed wejściem w życie ustawy o języku polskim i w konsekwencji jest po prostu zwykłą opinią językoznawców.

Bibliografia

Czarnecki P. 2014: *Ustawa o języku polskim. Komentarz. Stan prawny na 15 lutego 2014 roku*, Wydawnictwo Prawnicze LexisNexis, Warszawa.

Markowski A. 2016: *Dwudziestolecie Rady Języka Polskiego przy Prezydium PAN*, „Język Polski" XCVI, nr 4, s. 5–14.

Mędrzycki R. 2009: *Uchwały ortograficzne Rady Języka Polskiego w systemie tak zwanych niezorganizowanych źródeł prawa administracyjnego*, „Administracja. Teoria. Dydaktyka. Praktyka" 3 (16), s. 157–170.

Rada Języka Polskiego 1999: *Obowiązki, prawa i potrzeby Rady Języka Polskiego w świetle Ustawy o języku polskim z dn. 7 października 1999 r.* (online: http://web.archive.org/web/20030804052346/http: //www.rjp.pl/inf_dzial_obowiazki.html).

Sekretarz RJP 2003: *Zagrożenia polszczyzny; obcojęzyczne szyldy*, [w:] *Komunikaty Rady Języka Polskiego przy Prezydium Polskiej Akademii Nauk*, nr 2 (13).

Wojewódzki Sąd Administracyjny w Lublinie 2004: II SA/Lu 1310/03, Wyrok WSA w Lublinie z 11 maja 2004 (http://www.orzeczenia-nsa.pl/wyrok/ii-sa-lu-1310-03/ ewidencja_ludnosci_dowody_tozsamosci_akty_stanu_cywilnego_imiona_i_nazwisko_obywatelstwo_paszporty/d207f1.html).

Zieliński M. 2017: *Wykładnia prawa. Zasady, reguły, wskazówki*, wyd. 7 uzup., Wolters Kluwer, Warszawa.

Summary

Legal status of orthographic resolutions

Keywords: Council for the Polish Language, Act on the Polish language, Constitution.

Markowski claims that the resolutions of the Council for the Polish Language constitute legal regulations. This does not seem to be the case.

ANDRZEJ MARKOWSKI* | UNIWERSYTET WARSZAWSKI

Odpowiedź na polemikę prof. Janusza Bienia

W tekście *Status prawny uchwał ortograficznych* prof. Janusz Bień przedstawił uwagi na temat podejmowanych przez Radę Języka Polskiego uchwał w zakresie zasad dotyczących pisowni, będące polemicznym odniesieniem się do jednego ze zdań artykułu mojego autorstwa.

W artykule tym użyłem sformułowania „moc prawnie obowiązująca" w odniesieniu do uchwał ortograficznych Rady Języka Polskiego. Określenie to zostało jednak użyte w pewnym kontekście, w którym powinno być także odczytane. Jest to kontekst powołanego w moim artykule przepisu art. 13 ust. 1 ustawy o języku polskim oraz kontekst całego artykułu, w którym cytowane określenie zostało użyte.

W powołanym przepisie przyznana została RJP kompetencja do podejmowania uchwał wyrażających opinie w zakresie tam wskazanym oraz co do ustalania zasad ortografii i interpunkcji języka polskiego. Działania Rady znalazły tym samym swoje umocowanie prawne w wymienionym wyżej zakresie, którego przedtem nie miał żaden inny organ tego typu. Konsekwencje prawne takiego rozwiązania dotyczą tego, że RJP uzyskała podstawę prawną do podejmowania uchwał, o których mowa; jest bowiem przez przepis ustawy wskazana jako organ do tego kompetentny. Na podstawie uchwał RJP ustalających zasady ortografii i interpunkcji powstają normy powszechnie obowiązujące w tym zakresie (normy poprawności językowej), które nie są normami prawnymi (normami prawnie obowiązującymi), ale upoważnienie do ich ustanowienia jest upoważnieniem zawartym w przepisach prawa. Poza tym podmioty, na których ciąży prawny obowiązek ochrony języka polskiego, muszą, realizując ten obowiązek, przestrzegać norm zawartych w omawianych uchwałach RJP. Biorąc powyższe pod uwagę, nie ma wątpliwości co do d o n i o s ł o ś c i p r a w n e j uchwał ustalających zasady ortografii i interpunkcji języka polskiego. Uchwały te nie mają z prawnego punktu widzenia mocy prawnie obowiązującej jak normy prawa karnego czy cywilnego, ale też nie w znaczeniu prawniczym użyte zostało określenie o „mocy prawnie obowiązującej" omawianych uchwał. W dalszej części mojego artykułu przypominam bowiem, że „czasami wzywa się Radę do interpretacji ustawy o języku polskim, a także do egzekwowania przepisów tej ustawy, a przecież nie leży to w kompetencjach RJP. Rada jest instytucją «opiniodawczo-doradczą», a nie «organem egzekwującym». Może wyrażać opinie, nieraz bardzo jednoznacznie negatywne, o faktach językowych, ale nie ma bezpośredniego wpływu na praktykę językową w przestrzeni publicznej".

Biorąc pod uwagę kontekst całego artykułu *Dwudziestolecie Rady Języka Polskiego przy Prezydium PAN*, użyte w nim sformułowanie, będące przedmiotem uwag polemicznych ze strony prof. Bienia na temat statusu prawnego uchwał ortograficznych, może być rozumiane właśnie jako „d o n i o s ł o ś ć p r a w n a", a nie jako obowiązujące prawo. Taka też była moja intencja jako autora artykułu.

* * *

Na tym dyskusję w tej sprawie zamykamy. Redakcja

* a.markowski@uw.edu.pl

KRONIKA

PRZEMYSŁAW WIATROWSKI* | UNIWERSYTET IM. ADAMA MICKIEWICZA, POZNAŃ

V sympozjum naukowe z cyklu *Perspektywy współczesnej frazeologii polskiej* pt. *Frazeologia w stylach i gatunkach mowy*, Poznań, 7 listopada 2016

7 listopada 2016 roku w budynku Collegium Maius Uniwersytetu im. Adama Mickiewicza w Poznaniu odbyło się sympozjum naukowe *Frazeologia w stylach i gatunkach mowy*, zorganizowane przez Zakład Frazeologii i Kultury Języka Polskiego oraz Pracownię Leksykograficzną Instytutu Filologii Polskiej UAM. Było to już piąte spotkanie z cyklu *Perspektywy współczesnej frazeologii polskiej*. W konferencji wzięło udział dziesięcioro badaczy z ośrodków naukowych w Krakowie, Olsztynie, Opolu, Poznaniu, Szczecinie, Warszawie i we Wrocławiu.

Uczestników sympozjum powitała profesor Uniwersytetu im. Adama Mickiewicza w Poznaniu dr hab. Anna Piotrowicz, kierownik Zakładu Frazeologii i Kultury Języka Polskiego, która przypomniała problematykę wcześniejszych spotkań. Konferencję otworzył dr Krzysztof Skibski, wicedyrektor Instytutu Filologii Polskiej.

W pierwszym wystąpieniu zatytułowanym *Frazeologia gatunkowa a leksykografia przekładowa. Wybrane problemy i przykłady* Wojciech Chlebda (Uniwersytet Opolski) przywołał klasyczną koncepcję gatunków mowy Michała Bachtina, wedle której o przynależności różnych tekstów do jednego gatunku przesądzają trzy czynniki: zbieżność tematyczna, jednolitość kompozycyjna i wspólnota stylistyczna, czyli występowanie tych samych lub podobnych środków językowych, w tym wielowyrazowych jednostek języka. Problemem dla frazeologa i frazeografa jest fakt, że nie dysponujemy pełnym wykazem gatunków mowy. Brak takiego katalogu utrudnia ich wyczerpujący podział na gatunki „twarde", zrytualizowane, i gatunki „miękkie", o strukturze bardziej płynnej. W gatunkach zrytualizowanych (w tekstach prawnych i prawniczych, liturgicznych i in.) frazogeniczne środki językowe mają charakter obligatoryjny i konstytutywny, czyli odznaczają się potencjałem gatunkotwórczym. Inne problemy, z którymi przychodzi się zmierzyć frazeologowi i leksykografowi, dotyczą: sposobów ekscerpcji wielowyrazowych jednostek języka z tekstów zaliczanych do jednego gatunku, teoretycznego pytania o to, co się będzie uważać za wielowyrazową jednostkę języka, metod ekscerpcji tego, co badacz za jednostkę języka uzna, także wyboru z repertuaru zgromadzonych elementów tylko tych, które są frazemami gatunkowymi. Wyzwaniem dla frazeologii przekładowej jest z kolei nadanie wybranym frazemom gatunkowym postaci haseł słownikowych i przypisanie im ekwiwalentów w wybranym języku docelowym. Referent podzielił się także ze słuchaczami

* przemek@amu.edu.pl

spostrzeżeniami poczynionymi podczas pracy nad frazeologią gatunkową w *Podręcznym idiomatykonie polsko-rosyjskim*.

Drugie wystąpienie: *Związki frazeologiczne w interpretacjach dzieci w wieku przedszkolnym i wczesnoszkolnym* wygłosiła Kinga Kuszak (Uniwersytet im. Adama Mickiewicza w Poznaniu). Zadaniem, które postawiła sobie autorka, było opisanie, w jaki sposób dzieci interpretują, wyjaśniają, definiują wybrane związki frazeologiczne. Do analizy K. Kuszak wybrała kilka grup połączeń wyrazowych, z którymi dzieci stykają się w różnych sytuacjach komunikacyjnych oraz w literaturze dla nich przeznaczonej. Są to frazeologizmy odnoszące się do: rodziny (np. *głowa rodziny*), humoru (np. *zrywać boki ze śmiechu*), szkoły i uczenia (np. *dostać od kogoś szkołę*), świata (np. *ósmy cud świata*). Autorka szczegółowo przebadała dziecięce interpretacje związków: *być odciętym od świata*, *coś jest stare jak świat*, *coś jest nie z tego świata*, *zapomnieć o całym świecie*. Obserwacje pokazały, że częściej na pytania o znaczenie związków frazeologicznych odpowiadały dzieci przedszkolne. Dzieci z klas I–III czyniły to rzadziej, co wynika z ich doświadczeń edukacyjnych, ostrożności w formułowaniu wypowiedzi wiążącej się z oczekiwaniami nauczyciela (dorosłego). Skojarzenia odległe częściej pojawiają się także w grupie dzieci w wieku wczesnoszkolnym.

Następnie głos zabrała Jolanta Ignatowicz-Skowrońska (Uniwersytet Szczeciński). W wystąpieniu *Kwestie dyskusyjne w opisie leksykograficznym nowych związków frazeologicznych* referentka omówiła trudności dotyczące rejestracji w opracowaniach leksykograficznych nowego w polszczyźnie połączenia wyrazowego *walka buldogów pod dywanem*. Wskazane wyrażenie wciąż jest na etapie frazeologizacji, a zgromadzone przez badaczkę konteksty pozwalają śledzić proces jego utrwalania się w języku polskim: od posługiwania się nim jako cytatem czy parafrazowanym cytatem ze wskazaniem źródła, aż po moment, kiedy geneza związku przestaje być znana, a cytat przekształca się w skrzydlate wyrażenie, jednostkę języka, i staje się własnością ogółu. W aktualizacjach analizowanego połączenia realizują się dwa komponenty semantyczne: ‘walka, rywalizacja o władzę, wpływy, dominację w obrębie jakiejś grupy’ oraz ‘walka, zazwyczaj ukryta, niejawna’. W *Wielkim słowniku języka polskiego* połączenie opatrzone jest kwalifikatorem *książkowy*. Zdaniem autorki nie wydaje się on uzasadniony. Analiza źródeł przekonuje, że połączenie jest typowe dla stylu dziennikarskiego oraz swobodnych wypowiedzi zamieszczanych na interaktywnych i nieinteraktywnych stronach internetowych. Nie ma charakteru oficjalnego, nie podwyższa rejestru wypowiedzi, w której się pojawia, nie przydaje jej też walorów erudycyjnych.

Po krótkiej przerwie uczestnicy spotkania wysłuchali referatu *Frazeologizmy w katalogach reklamowych biur podróży* przygotowanego przez Alicję Nowakowską (Uniwersytet Wrocławski), która tematem wystąpienia uczyniła różnorodne frazemy zawarte w tytułach nadawanych imprezom turystycznym przez ich organizatorów, czyli przez biura podróży, i zamieszczanych w katalogach tychże biur. Autorka wyodrębniła blisko osiemdziesiąt nazw zawierających reprodukty językowe. Są wśród nich frazeologizmy (w postaci kanonicznej bądź zmodyfikowanej, np.: *Austriackie gadanie*, *Chorwacki strzał w dziesiątkę*, *Grecja w małym palcu*, *Portugalia od deski do deski*, *Zamki na piasku*, *Narty jak w banku*) i skrzydlate słowa: tytuły filmów, piosenek,

mitologizmy i biblizmy. Część omawianych konstrukcji podlega modyfikacjom. Dochodzi na przykład do wymiany komponentów: *Przyczajony tygrys, ukryty wok*, innowacji skracającej: *W służbie Jej Królewskiej Mości*, kontaminacji: *Szkocja – wzgórza walecznych serc*, innowacji składniowej: *W poszukiwaniu zaginionej arki*. Źródłem nominacji są ponadto tytuły utworów muzycznych (np. *Podróż do krainy uśmiechu*), programów radiowych i telewizyjnych (np. *Taniec z Khmerami*), nazwy produktów, z których słynie dany kraj (np. *Tam, gdzie dojrzewa śliwowica*), historyczne i mityczne nazwy miejsc (np. *Ziemia obiecana*). Częsta intertekstualność badanych tytułów wiąże się z celami reklamowymi katalogów. Funkcje analizowanych nazw są zbieżne z funkcjami nagłówków prasowych i sloganów reklamowych.

Piąta referentka, Ewa Młynarczyk (Uniwersytet Pedagogiczny im. KEN w Krakowie), zaprezentowała wybrane związki frazeologiczne funkcjonujące w nagłówkach informacji internetowych (*Frazeologia nagłówków informacji internetowych*). Swoistą cechą tego typu nagłówków jest ich interakcyjność. Przedmiotem zainteresowania badaczki były nagłówki umieszczone na pierwszych stronach portali internetowych. Cel wystąpienia sprowadziła do próby odpowiedzi na pytanie, czy można mówić o swoistej frazeologii nagłówków internetowych bądź – szerzej – medialnych. Materiał egzemplifikacyjny został zaczerpnięty z portalu Onet.pl. Autorka ograniczyła prezentację do kilku frazemów, które łączy podobny typ obrazowania. Ich domenę źródłową stanowi polowanie (np. *polowanie na kogoś/coś*, *urządzić polowanie na kogoś/coś*, *polowanie z nagonką*, *medialna nagonka*, *urządzić obławę*, *poczuć krew*, także szereg wariantów, których osią strukturalno-semantyczną jest wyrażenie *na celowniku*: *być/znaleźć się na celowniku kogoś/czegoś*, *mieć kogoś na celowniku*, *brać na celownik kogoś/coś*). Pojawiają się one nie tylko w tekstach dotyczących spraw społeczno-politycznych, ale także rywalizacji sportowej czy realiów ekonomicznych. Na podstawie zebranego materiału E. Młynarczyk doszła do wniosku, iż można mówić o tworzeniu się swoistej warstwy frazeologii nagłówków internetowych (być może – ogólniej – medialnych).

Następny referat: *Frazeologia a terminologia. Uwagi na podstawie badań frazeograficznych* wygłosiła Gabriela Dziamska-Lenart (Uniwersytet im. Adama Mickiewicza w Poznaniu). Autorka – na przykładzie konstrukcji odnotowanych przede wszystkim w *Nowym słowniku frazeologicznym* Renardy Lebdy – podjęła problem relacji między frazeologią a terminologią. Wymieniony słownik zamieszcza szereg połączeń o niejasnym statusie, nienotowanych w innych opracowaniach leksykograficznych, np.: *chemia organiczna*, *czasy stanisławowskie*, *historia nowożytna*, *ciało pedagogiczne*, *fizyka jądrowa*, *funkcja liniowa*, *kamień filozoficzny*, *kwadrans akademicki*, *miłośnik książek*, *liczba doskonała*. Wśród wyekscerpowanych przez badaczkę konstrukcji znalazły się frazeologizmy, terminy oraz połączenia leksykalne. Stanowią one nominalne grupy syntaktyczne odznaczające się zestawem następujących cech ogólnych: można im przypisać określoną charakterystykę semantyczną, pełnią funkcję nominatywną, są to połączenia syntagmatyczne, łączą się z innymi klasami wyrażeń, stanowią składniki zdań, pełnią funkcję rzeczowników. Referentka podkreśliła, że podobieństwo strukturalne między terminami złożonymi i frazeologizmami nie jest warunkiem wystarczającym, by terminy włączyć w zakres badawczy frazeologii. Trudności wynikają z istotnych różnic pomiędzy

frazeologizmami i terminami, z ich usytuowania w systemie języka oraz dotyczą odmienności stylistycznych. Istnieje jednak pewien zbiór terminów złożonych stanowiących obszar peryferyjny frazeologii.

W popołudniowej części obrad pierwsza zabrała głos Monika Czerepowicka (Uniwersytet Warmińsko-Mazurski w Olsztynie), która we wstępie wystąpienia *Frazeologia gatunków mowy a język polityki na przykładzie satyry politycznej Janusza „Szpota" Szpotańskiego* przedstawiła zestaw najważniejszych cech satyry politycznej oraz zaprezentowała podstawę materiałową referatu (cykl czterech utworów, które łączy postać towarzysza Szmaciaka). Efekt prześmiewczy uzyskuje twórca dzięki operacjom na związkach frazeologicznych i skrzydlatych słowach, czasem terminach. Konstrukcje te pojawiają się w nowych, zaskakujących kontekstach. Operacje na frazeologizmach, polegające głównie na wymianie członów, ujęciu jednego z komponentów czy połączeniu dwóch różnych związków wyrazowych, służą czterem celom: obnażają fałsz oficjalnych wyjaśnień, sloganów, rytualnych formuł językowych, budują dystans między ja lirycznym a światem przedstawionym, służą karykaturze postaci oraz odświeżają skostniałe struktury słowne. Operacje na skrzydlatych słowach, pozwalające – podobnie jak przekształcenia frazeologizmów – obnażyć fałsz, ośmieszyć, nadto zbudować silne relacje z czytelnikiem poprzez włączenie go w krąg osób wtajemniczonych, rozumiejących aluzyjne zabiegi autora, mają w badanych utworach charakter przywołań różnych tekstów kultury.

Z kolei Piotr Müldner-Nieckowski (Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie) w wystąpieniu zatytułowanym *Frazeologia w prozie pisarzy współczesnych publikujących w polskiej prasie literackiej. Rekonesans* zaprezentował własne przemyślenia na temat opracowywania materiału frazeologicznego wynotowanego z prozy artystycznej. W kręgu zainteresowań autora znalazło się zagadnienie związane z rodzajami źródeł, z których należy korzystać, gdy bada się literaturę pod kątem występującej w niej frazeologii. Referent odniósł się do istniejących korpusów tekstowych i przedstawił ich wady oraz zalety. Omówił także różne aspekty korzystania ze źródeł bezpośrednich. Badacz zaprezentował wykaz tytułów czasopism literackich, które zostaną poddane ekscerpcji frazeologicznej, a także walory gromadzenia danych z tego typu materiału. Referent sformułował kilka problemów badawczych wymagających rozpatrzenia (np. czy frazeologia obecna w prozie jest tworzywem tego samego typu co w poezji? Czy na podstawie ilościowej analizy użycia frazeologizmów można mówić o różnych stylach pisarskich i idiolektach? Czy sfrazeologizowanie tekstu literackiego jest cechą stylistyczną czy tylko formą narracyjną?).

Krzysztof Skibski wygłosił referat *Ewokacja frazeologizmów w odbiorze poezji współczesnej – czego nie widać*. Przedmiotem wystąpienia autor uczynił zjawisko ewokowania frazeologizmu oraz kategorię śladu frazeologicznego pojmowanego jako każda obecność frazeologizmu w tekście literackim – niezależnie od formy tekstowej i szczegółowych procesów semantycznych wywoływanych w trakcie odczytania. Ślad frazeologiczny jest zjawiskiem tekstowym ewokującym w lekturze frazeologizm wirtualny, czyli taki, który zyskał swoją stabilną postać formalno-semantyczną w uzusie lub normie. Wskazanie śladu frazeologicznego w tekście poetyckim jest uznaniem specjalnej relacji frazeologizmu wirtualnego i wiersza jako formy

tekstu. Relacja ta polega na dynamicznym związku między typowymi parametrami frazeologizmu wirtualnego i wynikających z nich sugestii interpretacyjnych a poetyckimi parametrami śladu frazeologicznego w konkretnym tekście poetyckim. Śladem frazeologicznym badacz nazwał zarówno wystąpienie elementów jednostki frazeologicznej w tekście, jak i brak takich elementów, realizuje się on bowiem jako przyczyna ewokacji dzięki dostrzeżeniu w odbiorze różnych czynników tekstowych, które motywują przywołanie frazeologizmu wirtualnego. Poruszane przez referenta zagadnienia zyskały stosowną egzemplifikację tekstową.

Ostatnie wystąpienie zatytułowane *Frazeologizmy w prozie Olgi Tokarczuk* przygotował Jan Zgrzywa (Uniwersytet im. Adama Mickiewicza w Poznaniu/Collegium Polonicum w Słubicach). W centrum zainteresowania autora znalazły się modyfikacje związków wyrazowych występujące we wszystkich utworach prozatorskich Olgi Tokarczuk. Referent utożsamił pojęcie modyfikacji frazeologicznej z pojęciem innowacji frazeologicznej rozumianej jako każde odchylenie od normy frazeologicznej. W toku dalszego wywodu referent poszukiwał odpowiedzi na pytanie, jakie skutki semantyczne pociągają za sobą zmiany struktury frazeologizmu w tekście artystycznym. W tym celu konfrontował konkretną modyfikację z postacią kanoniczną związku, która stanowiła podstawę tej modyfikacji. Zgromadzony materiał pozwolił autorowi dojść do wniosku, że przekształcenia frazeologizmów odgrywają w prozie Tokarczuk istotną rolę, mimo że są środkiem stylistycznym stosowanym przez pisarkę sporadycznie. Modulują one wypowiedź, nadając jej znamiona indywidualności i niepowtarzalności, wyrażają humor narratora i bohaterów, prowokują odbiorcę, zapraszając go do udziału w zabawie językowej. Są komentarzami metajęzykowymi, gdyż wyrażają dystans twórcy do stałych, umownych związków wyrazowych i często tę umowność demaskują.

Z wygłoszonymi podczas spotkania referatami będzie się można zapoznać w publikacji książkowej zapowiedzianej przez organizatorów.

VARIA

NOWOŚCI WYDAWNICZE

Bibliografia onomastyki polskiej od roku 2001 do roku 2005 włącznie, oprac. Iwona Nobis, przy współpracy Rozalii Przybytek, Instytut Języka Polskiego PAN, Kraków 2016, s. 358.

Czesak Artur, *Współczesne teksty śląskie na tle procesów językotwórczych i standaryzacyjnych współczesnej słowiańszczyzny*, Księgarnia Akademicka, Kraków 2015, s. 366.

Dawne z nowym łącząc... In memoriam Mariani Kucała, red. Jolanta Klimek-Grądzka, Małgorzata Nowak, Towarzystwo Naukowe KUL, Katolicki Uniwersytet Lubelski Jana Pawła II, Lublin 2016, s. 439.

Dialog z Tradycją, t. 5: *Językowe dziedzictwo kultury materialnej*, red. Ewa Młynarczyk, Ewa Horyń, Collegium Columbinum, Kraków 2016, s. 532.

Dubisz Stanisław, *Językoznawcze studia polonistyczne (pisma wybrane, uzupełnione, zmienione)*, t. 4: *Historia języka polskiego*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2016, s. 359.

Dyskurs w aspekcie porównawczym, red. Andrzej Charciarek, Anna Zych, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2016, s. 246.

Dzięgiel Ewa, Czarnecka Katarzyna, Kowalska Dorota A., Yanushevska Lyudmyla, *Polskojęzyczna prasa na Ukrainie sowieckiej w latach 1918–1939. Przegląd tytułów i treści*, Instytut Języka Polskiego PAN, Kraków 2016, s. 495.

Jakubczyk Marcin, *Leksykografia polsko-francuska XVIII wieku w perspektywie metaleksykograficznej*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2016, s. 260.

Kaszewski Marek, *Studia z leksykografii historycznej*, wyd. 2 rozszerzone, Universitas, Kraków 2016, s. 193.

Kita Małgorzata, Coming out *w polskiej przestrzeni dyskursywnej*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2016, s. 200.

Lingua et gaudium. Księga jubileuszowa ofiarowana profesorowi Janowi Miodkowi, red. Monika Zaśko-Zielińska, Małgorzata Misiak, Jan Kamieniecki, Tomasz Piekot, Oficyna Wydawnicza ATUT, Wrocław 2016, s. 600.

*Miasto. Przestrzeń zróżnicowana językowo, kulturowo i społecznie***,** t. 6, red. Małgorzata Święcicka, Monika Peplińska**,** Wydawnictwo Uniwersytetu Kazimierza Wielkiego, Bydgoszcz 2016, s. 447.

*Mówi się, czyli o wymowie i wymowności Polaków. Materiały IX Forum Kultury Słowa, Szczecin, 9–11 października 2013***,** red. Ewa Kołodziejek, Agnieszka Choduń**,** Volumina.pl, Daniel Krzanowski, Szczecin 2016, s. 288.

Nobis Iwona, *Rozwój fleksji nazw miejscowych w języku polskim*, Instytut Języka Polskiego PAN, Kraków 2016, s. 538.

Podhajecka Mirosława, *A history of Polish-English/English-Polish bilingual lexicography (1788–1947)*, Wydawnictwo Uniwersytetu Opolskiego, Opole 2016, s. 629.

Polskojęzyczne korpusy równoległe / Polish-language Parallel Corpora, red. Ewa Gruszczyńska, Agnieszka Leńko-Szymańska, Uniwersytet Warszawski, Instytut Lingwistyki Stosowanej, Warszawa 2016, s. 280.

Rozmyślanie przemyskie. Świadectwo średniowiecznej kultury religijnej, red. Jerzy Bartmiński, Artur Timofiejew, Państwowa Wyższa Szkoła Wschodnioeuropejska w Przemyślu, Przemyśl 2016, s. 221.

Sławkowa Ewa, *Tekst literacki w kręgu językoznawstwa*, t. 2, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2016, s. 206.

System – tekst – człowiek. Studia nad dawnymi i współczesnymi językami słowiańskimi, red. Małgorzata Gębka-Wolak, Joanna Kamper-Warejko, Iwona Kaproń-Charzyńska, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń 2016, s. 290.

Święcicka Małgorzata, *Młoda mowa. Studia nad polszczyzną dzieci i młodzieży*, Wydawnictwo Uniwersytetu Kazimierza Wielkiego, Bydgoszcz 2016, s. 512.

Zestawiła Monika Buława

Zarząd Główny Towarzystwa Miłośników Języka Polskiego i Redakcja „Języka Polskiego” z żalem zawiadamiają o śmierci

Profesor Antoniny Grybosiowej

absolwentki Uniwersytetu Jagiellońskiego, uczennicy Witolda Taszyckiego i Zenona Klemensiewicza, badaczki i miłośniczki języka polskiego, od 1960 roku związanej ze śląskim środowiskiem polonistycznym.

Profesor Antonina Grybosiowa pracowała na Uniwersytecie Śląskim, a później w Wyższej Szkole Umiejętności Społecznych. Po doktoracie z dziedziny historii polskiej składni skupiła swoje zainteresowania na polszczyźnie współczesnej, którą opisywała i komentowała w licznych artykułach, ogłaszanych na łamach czasopism nie tylko językoznawczych. Ogromne znaczenie miała Jej działalność organizacyjna, popularyzatorska i pedagogicznojęzykowa w Kole (później Oddziale) Katowickim (w latach 1976–1991 – Sosnowieckim) TMJP. Pełniła funkcje we władzach Oddziału, m.in. w latach 1977–1980 i 1988–1991 była przewodniczącą jego zarządu. Wygłaszała liczne odczyty na zebraniach środowiskowych, występowała w lokalnych rozgłośniach radiowych, publikowała felietony w miejscowej prasie, współorganizowała telefoniczną poradnię językową TMJP i Instytutu Języka Polskiego UŚ, była autorką tekstu i członkinią jury pierwszego Ogólnopolskiego Dyktanda, zorganizowanego w 1987 roku. Działalność swoją traktowała jako misję, sprawowaną z oddaniem i odwagą.

Żegnamy w Jej osobie aktywną działaczkę naszego Towarzystwa, niestrudzoną orędowniczkę troski o zachowanie polszczyzny pięknej, starannej i poprawnej.

OD REDAKCJI

„Język Polski" ukazuje się od 1913 roku. W roku 1921 stał się organem nowo powołanego Towarzystwa Miłośników Języka Polskiego. Twórcą formuły czasopisma jest Kazimierz Nitsch, który pełnił funkcję jego współredaktora (wraz z Romanem Zawilińskim, Janem Łosiem i Janem Rozwadowskim), a później redaktora. Po jego śmierci kierownictwo redakcji obejmowali kolejno: Zenon Klemensiewicz, Jan Safarewicz, Stanisław Urbańczyk, Marian Kucała, Krystyna Pisarkowa. Obecnie redaktorem naczelnym jest Piotr Żmigrodzki. Choć przez miejsce wydawania i osoby redaktorów związane z Krakowem, jest to pismo całego środowiska językoznawców i miłośników polszczyzny, które od przeszło wieku towarzyszy rozwojowi badań naukowych i edukacji językowej, służy publikowaniu artykułów naukowych, dyskusji, polemik i recenzji oraz dokumentowaniu lingwistycznego życia naukowego w Polsce.