

SEMANTYKA A KONFRONTACJA JĘZYKOWA

tom 4

POLSKA AKADEMIA NAUK
INSTYTUT SŁAWISTYKI

SEMANTYKA A KONFRONTACJA JEZYKOWA

tom 4

Pod redakcją
Violetty Koseskiej-Toszewej
i Romana Roszki



Warszawa 2008

Tom opiniowali do druku
ZBIGNIEW GREŃ i MAŁGORZATA KORYTKOWSKA

**Pozycja dofinansowana
przez ????????????????????**

Skład i łamanie
Danuta Roszko

© Copyright by Sławistyczny Ośrodek Wydawniczy
Printed in Poland

ISBN ????????????????????

Sławistyczny Ośrodek Wydawniczy
Instytut Sławistyki PAN
Pałac Staszica, ul. Nowy Świat 72, 00-330 Warszawa
tel./fax [+48] 22 826 76 88 tel. [+48] 22 827 17 41
sow-ispaw@wp.pl www.ispaw.waw.pl

SPIS TREŚCI

Wstęp — V. Koseska-Toszewa, R. Roszko	7
Małgorzata Korytkowska (Uniwersytet Łódzki, Łódź), Violetta Koseska-Toszewa (Instytut Slawistyki PAN, Warszawa), <i>Polsko-bułgarska gramatyka konfrontatywna</i>	??
Viara Maldjewa (Instytut Slawistyki PAN, Warszawa), Kategorie słowotwórcze a kategorie nazw w badaniu konfrontatywnym	??
Лилия Крумова-Цветкова (Институт за български език, БАН, София), Диана Благоева (Институт за български език, БАН, София), <i>За кавантитативността, изразявана от вида и времето на глагола в българския език (в съпоставка с полски)</i>	??
Roman Roszko (Instytut Slawistyki PAN, Warszawa), <i>Egzystencjalna fraza nominalna w badaniu konfrontatywnym</i>	??
Natalia Kotsyba (Instytut Slawistyki PAN, Warszawa), <i>Ingresywność i inchoatywność jako początkowe rodzaje czynności a kwantyfikacja predykatów w zdaniu (na przykładzie języka ukraińskiego)</i>	??
Joanna Satola-Staškowiak (Instytut Slawistyki PAN, Warszawa), <i>Zdania z zawieszoną asercją na podstawie polskich i bułgarskich zdań praeterytalnych</i>	??
Володимир А. Широков (Український мовно-інформаційний фонд Національної академії наук України), <i>Квантова лінгвістика</i>	??
Elżbieta Hajnicz (Instytut Podstaw Informatyki PAN, Warszawa), <i>Od tekstu do znaczenia: przyczynki do automatycznego tworzenia słownika semantycznego czasowników polskich</i>	??
Daniela Sorea (University of Bucharest, Bucharest), <i>Blends of Seasons and Mind Styles in Plath's <Spinster></i>	??
Daniela Sorea (University of Bucharest, Bucharest), <i>Hyperbole and Irony as Sitcom Props</i>	??

WSTĘP

Tom 4. studiów *Semantyka a konfrontacja językowa* porusza różnorodne zjawiska semantyczne głównie opisywane na podstawie konfrontowanych języków.

Tom 4. zawiera również artykuły z modnymi obecnie zagadnieniami z lingwistyki kwantowej, lingwistyki korpusowej oraz kognitywnej. Różnorodność zagadnień poruszanych przez autorów powinna zainteresować szeroki krąg lingwistów, zwłaszcza zajmujących się teorią i metodologią języka.

Violetta Koseska-Toszewa
Roman Roszko

ELŻBIETA HAJNICZ
Instytut Podstaw Informatyki PAN
Warszawa

**OD TEKSTU DO ZNACZENIA:
PRZYCZYNKI DO AUTOMATYCZNEGO TWORZENIA
SŁOWNIKA SEMANTYCZNEGO CZASOWNIKÓW
POLSKICH**

1. Wprowadzenie

Jest truizmem stwierdzenie, że w procesie kształtowania się języka zestaw pojęć zawsze wyprzedza opisujące je formy językowe. Proces ten trwa do dzisiaj, czego sztandarowym przykładem jest powstawanie rozmaitych slangów grupowych (zawodowych i środowiskowych), i ich przenikanie do języka ogólnego.

Patrząc jednak na przyswajanie języka z punktu widzenia jednostki, proces poznawania świata przebiega równoległe wraz z uczeniem się jego opisu. (*Patrz, kochanie, kaczka. Kaczka płynie do nas. Daj kaczce chlebka.*), np. [Gleitman, 1990]. Nie znaczy to co prawda, że mózg przechowuje wiedzę o świecie i wiedzę o języku w tym samym miejscu i w ten sam sposób, ale niewątpliwie informacje te są ściśle ze sobą powiązane.

Istnienie języka daje nam jednak inną wspaniałą możliwość: tak przekaz ustny jak pisany pozwala nam dowiadywać się wielu rzeczy, z którymi nie zetknęliśmy się i być może nigdy nie zetkniemy w naszym osobistym doświadczeniu. Potrafimy więc wyrobić sobie jakiś pogląd na rzeczywistość wyłącznie w oparciu o opis językowy, oczywiście jedynie dzięki odwołaniom do naszego doświadczenia i wiedzy o świecie. Im bardziej odległe są te rzeczywistości, tym powstałe wyobrażenia bywa bardziej mylące; stąd wielkie powodzenie rycin, ilustracji, a współcześnie fotografii.

Z oczywistych względów technicznych proces akwizycji wiedzy przez komputery nie może przebiegać w sposób analogiczny do ludzkiego. Istnieją co

prawda skomplikowane systemy rozpoznawania obrazu (identyfikacji obiektów); istnieją roboty wyposażone w dwie kamery w celu uzyskania obrazu stereograficznego, wyposażone w czujniki dotykowe. Jednak integracja tych technologii z rozwiązaniami z zakresu *Przetwarzania języka naturalnego* (ang. *Natural Language Processing; NLP*) jest obecnie czystą fantazją.

Niezależnie od tego jednak, czy mamy do czynienia z komunikacją ustną czy za pośrednictwem tekstu pisanego, w odbiorze komunikatu najważniejszy jest proces interpretacji przekazu językowego. Niejednokrotnie przyjmuje się, że proces ten przebiega od formy (pisanej bądź ustnej)¹, poprzez interpretację jej struktury (morfoskładniowej i składniowej) do reprezentacji pojęciowej.

Efektywne przetwarzanie informacji językowej (tekstowej) przez komputery wymaga przeprowadzenia analogicznego procesu, a uzyskanie adekwatnej reprezentacji pojęciowej (semantycznej) jest immanentnym składnikiem tego procesu, bez którego termin *Przetwarzanie języka naturalnego* można nieomalże uznać za nadużycie.

NLP jest dziedziną rozwijającą się już od lat. Obecnie istnieją rozbudowane techniki i zasoby i służące do analizy morfoskładniowej [Hajnicz i Kupś, 2001; Woliński, 2006] i składniowej, także dla języka polskiego [Obrębski, 2002; Woliński, 2004]. Z drugiej strony istnieją też złożone formalizmy, głównie logiczne, przeznaczone do reprezentacji semantyki tekstów. Wśród nich należy wyróżnić podejścia uwzględniające tematykę czasu i aspektu, takie jak *Event Semantics* [Davidson, 1967; Landman, 2000; Persons, 1990], DRT [Kamp i Reyle, 1993]. Jednak możliwość ich praktycznego wykorzystania wciąż jest ograniczona ze względu na lukę spowodowaną przez brak zasobów komputerowych z zakresu semantyki leksykalnej, opisujących poszczególne słowa i zależności między nimi.

Oczywiście tematyka ta także cały czas jest rozwijana. Na świecie powstaje wiele zasobów takich jak WordNet (Fellbaum, 1998; Miller i in., 1990; Vossen, 1998). Od jakiegoś czasu podejmowane są też próby tworzenia zasobów zawierających bardziej złożoną informację reprezentującą zależności semantyczne pomiędzy pojęciami i reprezentującymi je jednostkami leksykalnymi (poszczególnymi frazami występującymi w zdaniu) dla możliwie szerokiego podzbioru języka, wspólnego dla różnych sfer życia. Chodzi tu o strukturę predykatywno-argumentową wyrazów, przede wszystkim czasowników, wzbogaconą o informację semantyczną, co jest tematem niniejszego opracowania. W większości przypadków zasoby te są tworzone ręcznie, przy zaangażowaniu dużej ilości sił ludzkich, czasu i środków. Najbardziej znane

¹Ignorujemy tu poziom odcodowania sygnału dźwiękowego czy pisma.

z nich to FrameNet [Baker i in., 1998; 2003; Fillmore i in., 2001; 2002; 2003], VerbNet [Dang i in., 1998; 2000; Kipper i in., 2000; 2004] i PropBank [Kingsbury i Palmer, 2002; Kingsbury i in., 2002]. Ręczne tworzenie takich zasobów umożliwia nie tylko dużą precyzję, ale i stwarza szansę uwzględnienia dodatkowych informacji, np. wiedzy o świecie, nie mniej jednak stanowi barierę ilościową. Dlatego coraz częściej próbuje się realizować zadania z zakresu semantyki leksykalnej półautomatycznie i automatycznie [Lapata i Brew, 2004; McCarthy, 2000; Merlo i Stevenson, 2001; Schulte im Walde, 2003], nie tylko ze względu na ograniczenie zaangażowanych sił i środków, ale i większą elastyczność: metody automatyczne można zastosować do różnych rodzajów tekstów, uzyskując wyniki adekwatne dla pojedynczych zastosowań; zmniejsza to także wrażliwość na upływ czasu (łatwiej jest modyfikując posiadane już zasoby). Prace takie są prowadzone także dla języka polskiego. Podejście, którego ideę zamierzamy zaprezentować w niniejszym opracowaniu, wpisuje się w ten nurt.

Na czym to jednak, najogólniej rzecz ujmując, ma polegać? Otóż determinując znaczenie wyrazu (w szczególności w wypadku wystąpienia wieloznaczności bądź homonimii) należy uwzględnić dwa czynniki: strukturę wypowiedzenia, w którym wyraz ten się znajduje, oraz znaczenie innych wyrazów w kontekście. Rzecz jasna sama struktura składniowa jest niewystarczająca: sformułowanie poprawnie skonstruowanego acz pozbawionego treści zdania nie jest trudne (por. *Kot rozpuszcza się w błyskawicy o skorupie ponad uprzejmość z Cukier rozpuszcza się w wodzie o temperaturze ponad 10°C*). Z drugiej strony nieustrukturalizowany ciąg całkiem poprawnie zgrupowanych wyrazów też jest trudny, a często wręcz niemożliwy do interpretacji, to struktura bowiem wskazuje nam łączące wyrazy zależności (por. *Skrzydła w lot machać ptak z Ptak macha w locie skrzydłami*).

W wypadku przetwarzania komputerowego nakłada się na to element masy. Możemy bowiem sprawdzić, jak często pojawia się dany schemat (np. składniowy), jakie wyrazy i jak często w nim występują. Komputery nie są bowiem równie elastyczne jak ludzkie mózgi, jednak bez trudu precyzyjnie przetworzą, przeliczą i zapamiętają duże zbiory danych. Zresztą człowiek także potrafi się uczyć w ten sposób. Czytając tekst w obcym języku, potrafimy pojąć nieznaną nam słowo, jeśli tylko wystarczająco często pojawi się w wystarczająco zróżnicowanym kontekście [Gillette i in., 1999; Wittek, 2002].

Celem tego opracowania nie jest jednak omówienie metod statystycznego uczenia się semantyki leksykalnej. Raczej chcemy pokazać zjawiska, jakie można w ten sposób wykrywać. Skupimy się przy tym na najbardziej złożonym i chyba najciekawszym z nich, zjawisku *alternacji*.

2. Źródła danych do tworzenia słownika semantycznego

Jak już wspominaliśmy, naszym celem jest automatyczne tworzenie słownika semantycznej subkategoryzacji czasowników polskich. Podstawowym założeniem, na jakim opieramy swoje plany, jest przekonanie, że wiedzę semantyczną ukrytą wewnątrz struktury syntaktycznej tekstu da się wydobyć poprzez wykrycie regularnych (tzn. dostatecznie częstych) zależności pomiędzy tworzącymi ją wyrazami. Aby móc to uczynić, musimy wprawdzie mieć dostęp do takiej wiedzy o tekście.

2.1. Korpus IPI PAN

Podstawowym zasobem mającym umożliwić nam realizację tego zadania jest tworzony w Zespole Inżynierii Lingwistycznej IPI PAN *Korpus IPI PAN* (dalej KIPi) — największy dostępny publicznie zbiór tekstów w języku polskim, ujednolicony pod względem kodu (utf-8) i formatu (pliki XML, a dokładniej XCES, czyli xml-owej wersji *Corpus Encoding Standard* opracowanego w ramach TEI — *Text Encoding Initiative*). Co ważne, powstały w ramach grantu KBN 7 T11C043 20 [Przepiórkowski, 2004; Przepiórkowski i in., 2003] KIPi jest anotowany morfosyntaktycznie (za pomocą analizatora morfologicznego *Morfeusz*, [Woliński, 2006]). Dane leksykograficzne dla *Morfeusza* były wprowadzane ręcznie, lecz już proces dezambiguacji (ujednoznaczniania) anotacji był dokonywany automatycznie, za pomocą tagera wytrenowanego na tzw. *próbce uczącej* [Dębowski, 2004; Piasecki, 2006; 2007]. Obecnie korpus ten stanowi podstawę do tworzenia dwóch kolejnych zasobów: słownika walencyjnego² języka polskiego (opracowywanego w ramach grantu 3 T11C 00328 kierowanego przez dra A. Przepiórkowskiego: [Dębowski i Woliński, 2007; Fast i Przepiórkowski, 2005; Przepiórkowski i Fast, 2005; Przepiórkowski, 2006a; 2007a]) oraz polskiej wersji wordnetu (patrz poniżej), które to zasoby powstają na bazie informacji występujących w korpusie. Prowadzone są też prace nad oznaczeniem granic fraz w zdaniach składających się na teksty korpusu [Przepiórkowski, 2006b; 2007b, Przepiórkowski i Buczyński, 2007].

2.2. Słowność, czyli polski wordnet

Słowność dla języka polskiego (ang. *plWordNet*) jest to sieć relacji semantycznych (np. synonimii, hiperonimii i meronimii) dla wybranych najczęstszych jednostek leksykalnych (jedno- lub wielowyrazowych). Sieć ta stanowić ma tezaurus elektroniczny o strukturze zbliżonej do tezaury WordNet dla języka angielskiego oraz tezaurusów EuroWordNet skonstruowanych dla innych języków [Derwojedowa i in., 2007, 2008]. Jest ona tworzona w ra-

²Słownik walencyjny jest to słownik zawierający informacje o wymaganiach składniowych czasowników, czyli opisujący jakie frazy mogą być ich argumentami na poszczególnych pozycjach składniowych.

mach grantu 3 T11C01829 pt. *Automatyczne metody konstrukcji sieci semantycznej leksemów polskich na potrzeby przetwarzania języka naturalnego* kierowanego przez dra M. Piaseckiego z Instytutu Informatyki Stosowanej Politechniki Wrocławskiej. W ramach prac nad Słowosiecią dr Magdalena Zawisławska stworzyła wstępną klasyfikację na grupy znaczeniowe (oparte na klasyfikacji oryginalnej), zwane dalej *kategoriami semantycznymi*.

Gwoli ciekawości warto zaznaczyć, że część składających się na Słowosieć zależności także ma być wydobywana automatycznie (na bazie tekstów zgromadzonych w KIPi).

W pracach nad automatycznym tworzeniem słownika subkategoryzacji semantycznej czasowników zamierzamy wykorzystać informacje zgromadzone w Słowosieci w sposób jak najbardziej precyzyjny, lecz początkowo zamierzamy ograniczyć się do wspomnianych *kategorii semantycznych*.

2.3. Słownik syntaktyczno-semantyczny

Jak wspomnieliśmy powyżej, słownik walencyjny dopiero powstaje. W szczególności, nie jest znany format, w jakim słownik ten ma być przechowywany. Jednak na użytek niniejszych rozważań wystarczy wiedzieć, że każdemu czasownikowi przypisany będzie zbiór *schematów (wzorców) walencyjnych*, stanowiących listę wymagań składniowych danego czasownika, tzn. informacji o typie frazy (rzeczownikowa, przyimkowa, bezokolicznikowa itp.) i narzucanych na nią ograniczeniach (np. przypadek dla fraz rzeczownikowych i przyimkowych).

Taka struktura rozbudowana zostaje następnie o informację semantyczną dotyczącą poszczególnych argumentów, czyli pochodzące ze Słowosieci kategorie semantyczne wyrazów (rzeczowników, czasowników, przysłówków) będących realizacjami tych argumentów w tekście. Oznacza to, że schematy walencyjne zostaną rozbite na drobniejsze. Na przykład zdania *Piotr przejechał małego psa samochodem* oraz *Piotr przejechał cały kraj samochodem* mają identyczną strukturę syntaktyczną, lecz różne znaczenie. Jesteśmy w stanie to wykryć, gdyż *pies* ma kategorię semantyczną *zwierzę*, zaś *kraj* ma kategorię *miejsce*.

W rezultacie otrzymujemy postać pośrednią słownika, która w przybliżeniu wygląda następująco:

- każdemu czasownikowi przypisany jest zestaw schematów składniowo-semantycznych;
- każdemu schematowi przypisana jest (ponumerowana) lista argumentów;
- każdy argument posiada swą charakterystykę syntaktyczną (identyczną jak w schemacie czysto składniowym) wraz z informacją, czy jest

to argument wymagany czy opcjonalny. Przypisany mu też jest zbiór (pochodzących ze Słowosieci) kategorii semantycznych.

Przypisanie argumentowi całego zbioru kategorii semantycznych w miejsce pojedynczego elementu wynika z faktu, że zmiana kategorii semantycznej argumentu nie musi koniecznie oznaczać zmiany znaczenia czasownika. Na przykład znaczenie czasownika *kupić* nie zależy od faktu, czy kupujemy *krzesło*, *psa* czy *ziemię*.

W ostatecznej postaci słownika część syntaktyczna i semantyczna będą rozdzielone. W części semantycznej wymienione będą wszystkie potencjalne argumenty czasownika wraz z kategoriami semantycznymi. Syntaktyczny słownik walencyjny rozbudowany będzie o odnośniki do realizowanych przez dany schemat składniowy wzorców semantycznych z wyszczególnieniem powiązań między argumentami.

3. Alternacje

Funkcjonuje dość powszechnie przyjmowany pogląd, że istnieje zależność pomiędzy leksykalnie semantyczną reprezentacją czasownika a syntaktyczną realizacją jego argumentów, wyrażony w postaci *Hipotezy ogólnego wyrównania* (ang. *Universal Alignment Hypothesis*) [Perlmutter, 1978; Perlmutter i Postal, 1984]. Hipoteza ta pozwala na przyjęcie bardziej rygorystycznego założenia, że czasowniki charakteryzujące się analogicznymi wymaganiami składniowymi tworzą spójne semantycznie klasy [Levin, 1993; Levin i Rapaport Hovav, 2005].

Aby móc mówić o „analogicznych wymaganiach składniowych” czasowników nie wystarczy jednak rozważać każdą ich walencję osobno. Określeniu, jakie wzorce syntaktyczne są realizacjami jakich schematów znaczeniowych służy wykrywanie zjawiska zwanego *alternacją* bądź nieco wężiej *diatezą*.

W literaturze anglojęzycznej definiuje się *alternację* jako *regularną wariantywność w syntaktycznym uzewnętrznieniu argumentów czasownikowych* [McCarthy i Korhonen, 1998]. Dwa zdania tworzą alternację, jeśli:

- ich znaczenie jest zbliżone (najlepiej identyczne): taka sama liczba argumentów;
- znaczenie jednego jest uszczegółowieniem znaczenia drugiego: różna liczba argumentów.

W literaturze polskojęzycznej [Szupryczyńska, 1973] mówi się o zjawisku:

- alternacji, gdy zmianie ulega jedynie postać syntaktyczna poszczególnych argumentów;
- transformacji, gdy zmianie ulega struktura zdania.

Poniżej wykorzystamy wygodne rozróżnienie na *alternacje* i *transformacje* stosowane w literaturze polskiej. Ponieważ jednak nie istnieje wyczerpu-

jąca semantyczna klasyfikacja omawianego zjawiska w języku polskim, będziemy posługiwać się terminologią angielską.

W poniższym omówieniu zamierzamy powiązać zjawisko alternacji i transformacji struktury syntaktycznej wypowiedzeń (argumentów czasowników) z ich semantyczną charakterystyką. Jak zobaczymy, nie jest to powiązanie jednoznaczne.

Należy także pamiętać, że naszym dalekosiężnym celem jest automatyczne wykrywanie omawianego zjawiska. Dlatego w przykładach prezentowanych w dalszej części tekstu będziemy prezentować wyłącznie takie czasowniki, których frekwencja w korpusie daje szansę wykrycia poszukiwanych zależności metodami statystycznymi (choć bynajmniej nie gwarantuje tego ze względu na nierównomierną dystrybucję alternujących wariantów).

3.1. Alternacje pojedynczego argumentu

W większości wypadków zjawisko alternacji zachodzi dla pojedynczego argumentu czasownika.

- alternacja **biernik–dopełniacz**
 - bronić: dziecko/dziecka, kobietę/kobiety (ale: *rację/racji)
 - dodawać: olej/oleju, pomidory/pomidorów
 - kupić: chleb/chleba, ogórki/ogórków, cukier/cukru
(ale: samochód/*samochodu, pralkę/*pralki)
- alternacja **biernik–narzędnik**
 - rzucić: kamień/kamieniem, piłkę/piłką
- alternacja **biernik–fraza przyimkowa**
 - dochodzić: granicy / do granicy, celu / do celu, prawdy / do prawdy
(ale: *przystanku / do przystanku)
 - przechodzić: ulicę / **przez** ulicę
- alternacja **dopełniacz–fraza przyimkowa**:
 - bać się burzy / **z powodu** burzy, nauczyciela / **przez** nauczyciela
- alternacja pomiędzy **wyrażeniami przyimkowymi**:
 - zapaść się **w** śnieg / **w** śniegu

Jak widać, w wypadku niektórych czasowników alternacja zachodzi jedynie dla wybranych argumentów. Zazwyczaj warunkowane to jest ich charakterystyką semantyczną. Warunkiem może być *partytywność* argumentu (*chleb* dla czasownika *kupić*). Innym warunkiem jest *żywotność* argumentu (*kobieta* dla czasownika *bronić*). Może to też być jego *abstrakcyjność* (*cel* dla *dochodzić*).

Warto zauważyć, że dla części czasowników zachodzenie alternacji zależy od ich aspektu: *ukroić tort/tortu* ale *kroić tort/*tortu*, *liznąć krem/kremu* ale *lizać krem/*kremu*. Niestety, nie są to czasowniki dostatecznie częste

w korpusie, by mogły zostać wybrane do automatycznej analizy, nie mniej jednak warto o tym fakcie pamiętać na przyszłość.

Powyższe alternacje, choć do pewnego stopnia różnicowały kategorie argumentów, nie posiadały głębszego znaczenia semantycznego. Jednak poniższe dwie alternacje, *datywna* i *benefaktywna*, są bardzo często wspomniane w literaturze przedmiotu [Levin i Rappaport Hovav, 2005], także w zakresie analizy automatycznej [Joanis i Stevenson, 2003, Lapata i Brew, 2004]. Zauważmy przy okazji, że w tym wypadku alternacja dotyczy drugiego wymagania czasownika (pierwsze, biernikowe, nie ulega zmianie).

- alternacja **celownik–fraza przyimkowa**:
 - **do** + **D** (alternacja **datywna**)
 - Jan posłał/odniósł książkę Piotrowi / do Piotra*
 - Jan zgłaszał uwagi dyrektorowi / do dyrektora*
 - Jan zgłaszał uwagi zarządowi / do zarządu*
 - Problem
 - Fabryka wysłała szafę klientowi / do klienta*
 - Fabryka wysłała szafę firmie / do firmy Piotra*
 - Fabryka wysłała szafę do magazynu / *magazynowi*
 - Problem
 - Jan dał książkę Piotrowi / *do Piotra*
 - Jan dał książkę do introligatora / (–)introligatorowi*
 - Jan dał książkę do introligatora / introligatorowi do oprawy*
 - **dla** + **D** (alternacja **benefaktywna**)
 - Piotr zostawił/przekazał wiadomość Annie / dla Anny*
 - Maria kupiła zabawki dziecku / dla dziecka*
 - Piotr zrobił dziecku zabawkę / zabawkę dla dziecka*
 - Autor przygotował scenariusz reżyserowi / dla reżysera*
 - Problem
 - Jan posłał książkę dla Piotra (przez jego siostrę)*
 - Jan posłał/dał książkę Piotrowi dla Marii*

Na początek należy zauważyć, że jeden ze schematów walencyjnych uczestniczących w obu tych zjawiskach wymaga na pozycji dopełnienia frazy rzeczownikowej w celowniku. Jednak wymaganiem alternatywnym jest raz fraza przyimkowa **do**, a za drugim razem — **dla**. Różnica ta ma poważne konsekwencje semantyczne: w pierwszym wypadku argument pełni rolę semantyczną odbiorcy, a w drugiej — beneficjenta.

Tym ważniejsza jest obserwacja, że wykrycie, która z tych dwóch alternacji zachodzi dla danego czasownika wcale nie jest proste. Stwierdzenie, że odbiorcą akcji datywnych może być wyłącznie człowiek, zwierzę lub grupa ludzi (w tym instytucja) jest dość oczywiste³.

³W wypadku czasowników benefaktywnych w ogóle nie da się użyć innego typu rze-

Pierwszy problem stanowi czasownik *dać* zdający się sztandarowym czasownikiem datywnym, a jego dopełnieniem jest *odbiorca*. Jednak (w przeciwieństwie do angielskiego *give*) podleganie tego czasownika alternacji datywnej jest wątpliwe. Przede wszystkim, typowy ożywiony odbiorca akcji może być wyrażony wyłącznie w (bardziej zapewne naturalnym) schemacie celownikowym. W tym znaczeniu *dać* jest bliskoznaczne z *ofiarować*: sugeruje, że odbiorca jest odbiorcą docelowym, „ostatecznym”. Podmiana tych czasowników w zdaniach dotyczących introligatora uwidacznia różnicę znaczeń⁴: *Jan ofiarował książkę introligatorowi *do oprawy*. Można coś ofiarować introligatorowi jak każdemu, jednak nie można ofiarować tego w jakimś celu. Czasownik *dać* ma więc dwa znaczenia: jedno w schemacie celownikowym, drugie w schemacie przyimkowym. Zachodząca tu alternacja jest więc jedynie pozorna. Jednak automatyczne wykrycie tego faktu wymaga subtelного podejścia. Podejście oparte na opisanych w rozdziale 2.2 kategoriach semantycznych wykryje jedynie dysproporcję w występowaniu czasownika *dać* w obu schematach. Nie jest to jednak dostateczny sygnał do odrzucenia takiej informacji, gdyż rozkład realizacji podlegającym alternacji schematów składniowych czasownika w korpusie wcale nie musi być równomierny [Merlo i Stevenson, 2001]. Zależy on od typu alternacji, więc istotne może być porównanie zachowania całego zbioru czasowników potencjalnie podlegających alternacji danego typu.

Najważniejszym jednak problemem jest fakt, że czasowniki podlegające alternacji datywnej mogą występować również z frazą przyimkową **dla**. Co więcej, na tej pozycji mogą znajdować się analogiczne argumenty jak w wypadku fraz celownikowej i **do** (łącznie z instytucjami). Także w tym wypadku występowanie alternacji jest iluzoryczne: fraza **dla** zawsze wyznacza beneficjenta. Jediną sugestią (poza oczywistą, bardzo silną dysproporcją wystąpień) jest istnienie dla tych czasowników schematu, który zawiera dwa wymagania: fraza rzeczownikowa w celowniku (lub przyimkowa **do**) oraz przyimkowa **dla**. Poza tym można arbitralnie narzucić wyższy priorytet alternacji datywnej niż benefaktywnej.

Specyficznym i dość rzadkim przykładem alternacji **celownik–fraz przyimkowa** jest:

• **między** + **D**

Mikołaj rozdał prezenty dzieciom / między dzieci

W tym wypadku odbiorca (beneficjent?) musi stanowić zbiorowość.

czowników; por. *Piotr zostawił/przekazał krzesło *magazynowi / *dla magazynu*.

⁴ *Ofiarować* w ogóle nie posiada schematu z **do**.

3.2. Alternacje dwóch argumentów

Istnieją jednak sytuacje, w których mamy do czynienia z równoczesną alternacją dwóch argumentów czasownika. Co ważniejsze, nie są to zjawiska niezależne, lecz warunkujące się nawzajem. Podstawowym przykładem takiego zjawiska jest alternacja **lokatywna**:

- alternacja **⟨biernik, narzędnik⟩** — **⟨fraza przyimkowa, biernik⟩**
Piotr (za)ładował wóz/worek ziemniakami
Piotr (za)ładował ziemniaki na wóz/ do worka
- alternacja **⟨biernik, fraza przyimkowa⟩** — **⟨fraza przyimkowa, biernik⟩**
Piotr posprzątał bałagan w pokoju
Piotr posprzątał pokój z bałaganu
Maria wytarła (kurz) (z półki/ na półce)
Maria wytarła półkę (z kurzu)
- transformacja **⟨podmiot, dopełnienie (narzędnik)⟩** — **⟨fraza przyimkowa, podmiot⟩**
Ogród/Pole pachnie kwiatami (od kwiatów)
Kwiaty pachną w ogrodzie/ na polu
(W ogrodzie/ Na polu pachnie kwiatami/ od kwiatów)

Warto zwrócić uwagę, że w ostatnim wypadku mamy do czynienia z *transformacją* struktury zdania, gdyż za taką należy uznać przesunięcie dopełnienia na pozycję podmiotu (i *vice versa*).

Ponadto dla czasownika *załadować* i innych podobnych konstrukcji narzędnikowych mamy do czynienia jedynie z przybliżoną parafrazą, gdyż oznacza on zapełnienie *wozu/worka*, gdy tymczasem postać zdania z frazą przyimkową takiego wymagania nie stawia. Pełna parafraza pojawia się jedynie dla czasownika niedokonanego (*ładować*).

Zauważmy, że we wszystkich powyższych przykładach przyimek determinujący miejsce (*na wóz/ do worka, na polu/ w ogrodzie*) zależy od rzeczownika, nie od czasownika. Ponadto musi to być przyimek określający to miejsce, a nie położenie względem niego: użycie fraz *pod wóz, koło worka, za ogrodem, przy polu, obok pokoju* w powyższych zdaniach nie prowadziłyby do powstania parafrazy względem zdań zawierających odpowiednie rzeczowniki na pozycji podmiotu, a więc nie mielibyśmy do czynienia z alternacją.

Niestety, żaden z czasowników podlegających alternacji lokatywnej nie jest dostatecznie częsty w KIPI, by móc brać pod uwagę jej automatyczne wykrywanie przy wykorzystaniu obecnie dostępnych zasobów.

3.3. Transformacja zwrotna

Chociaż powyżej uznaliśmy alternację lokatywną za przypadek transforma-

cji, klasyczne pojęcie *transformacji* ma związek ze zmianą strony czasownika. Typowym przykładem transformacji jest więc *transformacja bierna*. Ponieważ jednak jest to przekształcenie o charakterze standardowym, zachodzące dla wszystkich czasowników przechodnich, nie zamierzamy się nim w niniejszym omówieniu zajmować. Skupimy się na znacznie ciekawszym zjawisku *transformacji zwrotnej*, związanej z przejściem czasownika ze strony czynnej do strony zwrotnej. Z transformacją zwrotną mamy do czynienia w następujących przypadkach:

1. Nie zachodzą zmiany składniowe:

a. *się* zwrotne

bronić żony (przed atakiem) – *bronić się (przed atakiem)*
zapisać dziecko do szkoły – *zapisać się do szkoły*

Konstrukcje takie charakteryzuje możliwość zamiany partykuły *się* na zaimek zwrotny *siebie*.

b. *się* dostawione⁵

pytać o drogę – *pytać się o drogę*

2. Czasownik zwrotny zmienia strukturę połączenia (wymaganie składniowe dopełnienia):

a. zmienia przypadek rzeczownika:

• transformacja **biernik** → **dopełniacz**

dzielić wrażenia – *dzielić się wrażeniami*
podjąć pracę – *podjąć się pracy*
trzymać poręcz – *trzymać się poręczy*

• transformacja **biernik** → **narzędnik**

wykazać odwagę – *wykazać się odwagą*
wymienić adresy – *wymienić się adresami*

b. biernik przechodzi w wyrażenie przyimkowe:

• transformacja **B** → **do+D (na+B)**

przygotowywać wystąpienie – *przygotowywać się do wystąpienia*
organizować akcję – *organizować się do akcji (na akcję)*

• transformacja **B** → **w+Msc**

kochać dziewczynę – *kochać się w dziewczynie*
ćwiczyć biegi – *ćwiczyć się w biegach*
poprawiać wyniki – *poprawiać się w wynikach*

⁵W wypadku czasownika *prosić* dostawienie *się* zmienia znaczenie wypowiedzenia na pejoratywne. W rezultacie są one używane w nieco innych kontekstach, np. *prosić o pomoc* i *prosić się o kłopoty*. Jednak rozróżnienie tych kontekstów w ramach analizy automatycznej wymaga dokładnych danych ze Słowosieci.

- transformacja $\mathbf{B} \rightarrow \mathbf{z} + \mathbf{N}$

<i>kryć swoje nałogi</i>	–	<i>kryć się ze swoimi nałogami</i>
<i>deklarować pomoc</i>	–	<i>deklarować się z pomocą</i>
<i>narzucać swoje towarzystwo</i>	–	<i>narzucać się ze swoim towarzystwem</i>
<i>spotkać znajomego</i>	–	<i>spotkać się ze znajomym</i>
<i>bawić dziecko</i>	–	<i>bawić się z dzieckiem</i>

Warto zauważyć, że ostatnia transformacja ma charakter dwojaki. W początkowych trzech zdaniach mamy *się zwrotne*, w kolejnych dwóch — *się wzajemne (recyprokalne)*, występujące wyłącznie w liczbie mnogiej. Różnica ta objawia się także żywotnością dopełnienia. Czasowniki z *się* recyprokalnym charakteryzuje potencjalna łączliwość z partykulą *nawzajem* ew. z frazą *ze sobą*.

Przyjrzyjmy się innym bardziej konkretnym przykładom ilustrującym to zjawisko:

- Jan spotyka Piotra*
- Jan spotyka się z Piotrem*
- Jan i Piotr spotykają się (ze sobą)*
- Jan lubi Piotra*
- Jan lubi się z Piotrem*
- Jan z Piotrem lubią się (nawzajem/ze sobą)*
- Jan i Piotr lubią się (nawzajem/ze sobą)*

Zjawisko to także może dotyczyć czasowników z *się* morfologicznym:

- Jan różni się od Piotra*
- Jan z Piotrem różnią się (między sobą)*
- Jan i Piotr różnią się (między sobą)*

We wszystkich tych przypadkach (poza ostatnim) parafraza jest dość ograniczona, gdyż wypowiedzenie w stronie czynnej sugeruje bierność Piotra, która jest obiektem (a dokładniej mówiąc kontragenssem), a nie agensem akcji. Pokazują to następujące przykłady:

- Jan pocałował zdjęcie Marii*
- *Jan i zdjęcie Marii pocałowali się*
- Samochód zderzył się z ciężarówką*
- Samochód i ciężarówka zderzyły się*
- Samochód zderzył się ze słupem*
- *Samochód i słup zderzyły się*

Oczywiste jest, że *zdjęcie* i *słup* nie mogą pełnić roli agensa, choć doskonale mieszczą się w roli obiektu.

Istnieją także przypadki transformacji zwrotnej tego typu, gdy dochodzi do modyfikacji znaczenia wypowiedzenia. Doskonałym przykładem takiego zjawiska jest:

W drugiej grupie czasowników podmiot czynny stanowią ludzie oraz zatrudniające ich instytucje. Charakteryzują się wymianą na frazę przyimkową **u+D** dla podmiotu żywotnego oraz **na/w+MSc** dla podmiotu nieżywotnego.⁶ Zmiana agensa jest wyraźna (dla obu argumentów żywotnych, w wypadku podmiotu nieżywotnego możemy mieć do czynienia z przyczyną (*nóż*)).

- żywotny podmiot czynny

<i>Publiczność wypełniła teatr</i>	⇔	<i>Teatr wypełnił się publicznością</i>
<i>Ludzie przyjęli zwyczaj</i>	⇔	<i>Zwyczaj przyjął się wśród ludzi / między ludźmi</i>
<i>Spółeczność przyjęła zwyczaj</i>	⇔	<i>Zwyczaj przyjął się w społeczności</i>
<i>Robotnicy rozpoczęli strajk (w fabryce)</i>	⇔	<i>Strajk rozpoczął się (w fabryce)</i>

- Oba podmioty nieżywotne

<i>Nauki biologiczne przyjęły teorię ewolucji</i>	⇔	<i>Teoria ewolucji przyjęła się w naukach biologicznych</i>
<i>Woda wypełniła szklankę</i>	⇔	<i>Szklanka wypełniła się wodą</i>
<i>Owacje zakończył przedstawienie</i>	⇔	<i>Przedstawienie zakończyło się owacjami</i>
<i>Wiatr otwiera drzwi</i>	⇔	<i>Drzwi otwierają się od wiatru</i>

- b. podmiot czynny ⇒ celownik lub wyrażenie przyimkowe
 dopełnienie (biernikowe) ⇒ podmiot zwrotny
- | | | |
|---|---|---|
| <i>Hrabina urodziła syna</i> | ⇔ | <i>Syn urodził się hrabinie / u hrabiny</i> |
| <i>Piotr zamknął oczy</i> | ⇔ | <i>Oczy zamknęły się Piotrowi</i> |
| <i>Piotr otworzył usta (ze zdiwienia)</i> | ⇔ | <i>Usta otworzyły się Piotrowi (ze zdiwienia)</i> |

Także w tym wypadku strona zwrotna używana jest w znaczeniu biernym.

c. alternacja **kauzatywna**

podmiot czynny	⇒	znika (ew. wyr. przyimkowe)
dopełnienie (biernikowe)	⇒	podmiot zwrotny
<i>Prezes zakończył spotkanie</i>	⇔	<i>Spotkanie zakończyło się</i>
<i>Piotr otworzył drzwi (kluczem)</i>	⇔	<i>Drzwi otworzyły się</i>

Alternacja kauzatywna jest najczęściej w literaturze omawianym przypadkiem alternacji [McCarthy, 2000; Merlo i Stevenson, 2001; Rappaport i Hoav, 2005], warto więc chyba poświęcić jej trochę więcej uwagi. Niestety, klasyczne przykłady omawiane w literaturze dotyczą czasowników, które nie występują w KIPi zbyt często. Co prawda niektóre ze wcześniej omawianych

⁶Takie zróżnicowanie występuje także dla innych czasowników niepodlegających transformacji czynnej, np. *kupować chleb u/od piekarza / w piekarni*, *złatwiać sprawy u urzędnika / w urzędzie*, *zamówić lustro u szklarza / w zakładzie szklarskim*. Jednak kwestie te wykraczają poza zakres bieżącej analizy.

przykładów można uznać za przypadki alternacji kazuatywnej, np. *Nóż zranił Jana* czy *Wiatr otwiera drzwi*, jednak na pewno nie są to przykłady standardowe. Dlatego pozwolimy sobie omówić kilka przykładów, mimo że w ich wypadku nie możemy liczyć na automatyczne wykrywanie tej alternacji.

Piotr zbił szybę (piłką)
Piłka zbiła szybę
Szyba zbiła się
Kelner stłukł całą tacę szklanek
Cała taca szklanek stłukła się (kelnerowi)
Kucharz stopił masło (w garnku / na ogniu / ?ogniem)
Ogień/ Garnek stopił masło*
Masło stopiło się
Ciotka upiekła ciasto (w piekarniku)
Ciasto upiekło się (ciotce) (w piekarniku)
Kelner napętnia kieliszki (winem)
Kieliszki napętniają się (winem)
(Wino napętnia kieliszki)

Powyższe przykłady alternacji kazuatywnej dotyczyły czasowników, dla których w schemacie nieprzechodnim podmiot stanowi obiekt akcji. Alternacja taka zachodzi także dla czasowników, dla których w schemacie nieprzechodnim podmiot stanowi agens akcji (odbiorca w schemacie przechodnim czasownika):

Jeździec popędził/pognał konia (przez las)
Koń popędził/pognał (przez las)
Piotr powłókł Marię do kina
Maria powłokła się (z Piotrem) do kina
Nadzorca zaharował/zamęczył niewolników (na śmierć)
Niewolnicy zaharowali/zamęczyli się (na śmierć)

Alternacja kazuatywna często zachodzi nie dla wszystkich wystąpień czasownika (o tej samej walencji), co zazwyczaj wskazuje na różnice znaczeniowe między tymi wystąpieniami:

Jan zbił psa (kijem)
**Kij zbił psa*
**Pies zbił się*

Co ciekawsze, w obu wypadkach argumenty czasownika *zbić* pełnią taką samą rolę semantyczną: agensa, obiektu i narzędzia. Z drugiej strony możemy mieć do czynienia z pozorną alternacją, jak w przykładzie:

Matka myje dziecko (gąbką)
Dziecko myje się (gąbką)
**Gąbka myje dziecko*

Tym razem nie mamy do czynienia z alternacją kauzatywną, gdyż w jej przypadku to narzędzie, a nie obiekt, ma preferencję do realizacji na pozycji podmiotu. Wydaje się ponadto, że alternacja taka nie powinna być dopuszczalna, gdy obiekt należy do tej samej klasy semantycznej co agens (wyznaczonej dla tego czasownika), tzn. może wystąpić na pozycji podmiotu w zdaniu bardziej szczegółowym (*Dziecko myje siostrzyczkę*). Jednak tej obserwacji przeczą wspomniane powyżej (rzadkie co prawda) przypadki czasowników *powlec*, *zaharować*, *zamęczyć*. Do zupełnej konsternacji prowadzi jeszcze kolejny przypadek:

Piotr zranił Jana (nożem)

Jan zranił się (nożem)

Nóż zranił Jana

W tym wypadku, podobnie jak dla czasownika *myć*, pierwsze dwa zdania nie stanowią przypadku alternacji, mimo że narzędzie (*nóż*) może przesunąć się na pozycję podmiotu... Klasyfikacja tych przypadków i znalezienie kryteriów umożliwiających ich rozróżnienie, zwłaszcza metodami automatycznymi, to zadanie do dalszych badań.

4. Podsumowanie

W niniejszym omówieniu zaprezentowaliśmy przegląd czasowników podlegających zjawisku alternacji wymagań składniowych bądź transformacji struktury składniowej. Naszym dalekosiężnym celem jest automatyczne wykrywanie tych zjawisk. Dlatego skupiliśmy się na czasownikach, które występują w KIPi na tyle często, by automatyczna analiza miała szansę powodzenia. Prezentacja ta pokazuje jednak, jak trudne to będzie zadanie. Przede wszystkim, różne przypadki alternacji, a zwłaszcza transformacji, wykazują odmienne zachowania, przede wszystkim pod względem dystrybucji argumentów pomiędzy alternujące wypowiedzenia. Zróznicowanie zbiorów tych argumentów (i ich kategorii semantycznych) może być wynikiem zarówno niejednoznaczności czasownika (*przejechać most / przez most* vs. *przejechać psa / *przez psa*) jak i różnic w znaczeniu samego argumentu (*dać książkę do introligatora / introligatorowi do oprawy / Piotrowi / *do Piotra*). Alternacje mogą mieć też pewne priorytety zachodzenia, np. alternacja *datywna* nad *benefaktywną*.

Może więc okazać się, że nie da się uzyskać jednolitej metody automatycznego wykrywania różnych typów alternacji/transformacji, i konieczne będzie przyjęcie założeń właściwych dla poszczególnych zjawisk. Niniejsze opracowanie wskazuje nam, na jakie pułapki należy zwracać uwagę.

Bibliografia

- Baker, C. F., Fillmore, C. J. i Lowe, J. B. (1998)** The Berkeley FrameNet Project, *Proceedings of COLING-ACL'98 Conference*, Montreal, Canada: 86–90.
- Baker, C. F., Fillmore, C. J. i Cronin, B. (2003)** The structure of the FrameNet database, *International Journal of Lexicography* **16** 3: 281–296.
- Dang, H.T., Kipper, K. i Palmer, M. (2000)** Integrating compositional semantics into a verb lexicon, *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, Saarbrücken, Germany.
- Dang, H.T., Kipper, K., Palmer, M. i Rosenzweig, J. (1998)** Investigating regular sense extensions based on intersective Levin classes, *Proceedings of COLING-ACL'98 Conference*, Montreal, Canada: 293–299.
- Davidson, D. (1967)** The logical form of action sentences, w W. N. Rescher (red.) *The logic of decisions and actions*, University of Pittsburgh Press, Pittsburgh, PA: 81–120.
- Derwojedowa, M., Piasecki, M., Szpakowicz, St. i Zawisławska, M. (2007)** Polish WordNet on a Shoestring, *Data Structures for Linguistic Resources and Applications: Proceedings of the GLDV 2007 Biannual Conference of the Society for Computational Linguistics and Language Technology*, Universität Tübingen.
- Derwojedowa, M. Zawisławska, M., Bilińska J. i Głąbska M. (2008)** Nouns in the WordNet for Polish, *Études Cognitives* 8, SOW, Warszawa, Poland: 211–222.
- Dębowski, Ł. (2004)** Trigram morphosyntactic tagger for Polish, *Proceedings of the IIS:IIPWM'04 Conference, Intelligent Information Processing and Web Mining*, Springer: 409–413.
- Dębowski, Ł. i Woliński, M. (2007)** Argument co-occurrence matrix as a description of verb valence, *Proceedings of the 3rd Language & Technology Conference*, Poznań, Poland.
- Fast, J. Przepiórkowski, A. (2005)** Automatic Extraction of Polish Verb Subcategorization: An Evaluation of Common Statistics, *Proceedings of the 2nd Language & Technology Conference*, Poznań, Poland: 191–195.
- Fellbaum, C., red. (1998)** *WordNet — an electronic lexical database*, MIT Press, Cambridge, MA.
- Fillmore, C. J., Baker, C. F. i Sato, H. (2002)** The FrameNet database and software tools, *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC)*, Las Palmas, Spain: 1157–1160.
- Fillmore, C. J., Johnson, C. R. i Petruck, M. R. L. (2003)** Background to FrameNet, *International Journal of Lexicography* **16** 3: 235–250.

- Gillette, J., Gleitman, H., Gleitman, L. i Lederer, A. (1999)** Human simulation of vocabulary learning, *Cognition* **73**: 135–176.
- Gleitman, L. (1990)** The structural sources of verb meanings, *Language Acquisition: A Journal of Developmental Linguistics* **1** (1): 3–55.
- Hajnicz, E. i Kupść, A. (2001)** *Przegląd analizatorów morfologicznych dla języka polskiego*, Prace IPI PAN 937, Warszawa.
- Joanis, E. i Stevenson, S. (2003)** A general feature space for automatic verb classification, *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2003)*, Budapest, Hungary: 163–170.
- Kamp, H. i Reyle, U. (1993)** *From discourse to logic*, Kluwer, Dordrecht.
- Kingsbury P. i Palmer, M. (2002)** From TreeBank to PropBank *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC)*, Las Palmas, Spain: 1989–1993.
- Kingsbury P. Palmer, M., i Marcus, M. (2002)** Adding semantic annotation to the Penn TreeBank, *Proceedings of the Human Language Technology Conference*, San Diego, CA: 252–256.
- Kipper, K., Dang, H. T. i Palmer, M. (2000)** Class-based construction of a verb lexicon, *Proceedings of the 17th National Conference on Artificial Intelligence AAAI-2000*: 691–696.
- Kipper, K., Snyder, B. i Palmer, M. (2004)** Using prepositions to extend a verb lexicon, *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics*, Boston, MA: 23–29.
- Landman, F. (2000)** *Events and plurality*, Kluwer, Dordrecht, Holland.
- Lapata, M. i Brew, C. (2004)** Verb class disambiguation using informative priors, *Computational Linguistics* **30** (1): 45–73.
- Levin, B. (1993)** *English verb classes and alternation: a preliminary investigation*, University of Chicago Press.
- Levin, B. i Rappaport Hovav, M. (2005)** *Argument Realisation*, Cambridge University Press.
- McCarthy, D. (2000)** Using semantic preferences to identify verbal participation in role switching alternations, *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-2000)*, Seattle WA: 256–263.
- McCarthy, D. i Korhonen, A. (1998)** Detecting verbal participation in diathesis alternations, *Proceedings of the COLING-ACL'98 Conference*, Montreal, Canada: 1493–1495.

- Merlo, P. i Stevenson, S. (2001)** Automatic verb classification based on statistical distributions of argument structure, *Computational Linguistics* **27** (3): 373–408.
- Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D. i Miller, K. (1990)** Introduction to WordNet: an on-line lexical database, *International Journal of Lexicography* **3** (4): 235–244.
- Obrębski, T. (2002)** *Automatyczna analiza składniowa języka polskiego z wykorzystaniem gramatyki zależnościowej*, praca doktorska, IPI PAN, Warszawa.
- Parsons, T. (1990)** *Events in the semantics of English*, MIT Press, Cambridge, MA.
- Perlmutter, D.M. (1978)** Impersonal passives and the unaccusative hypothesis, *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society*, Berkeley Linguistics Society, Berkeley, CA: 157–189.
- Perlmutter, D.M. i Postal, P.M. (1984)** The 1-advancement exclusiveness law, w D.M. Perlmutter i C. Rosen (red.) *Studies in Relational Grammar 2*, University of Chicago Press, Chicago, IL: 81–125.
- Piasecki, M. (2006)** Hand-written and Automatically Extracted Rules for Polish Tagger, w Petr Sojka, Ivan Kopecek i Karel Oliva (red.), Brno 2006 *Proceedings of Text, Speech and Dialog*, LNCS, Springer, Berlin.
- Piasecki, M. i Godlewski, G. (2006)** Effective Architecture of the Polish Tagger, w Petr Sojka, Ivan Kopecek i Karel Oliva (red.), Brno 2006 *Proceedings of Text, Speech and Dialog*, LNCS, Springer, Berlin.
- Przepiórkowski, A. (2004)** *Korpus IPI PAN. Wersja wstępna*, Instytut Podstaw Informatyki PAN, Warszawa.
- Przepiórkowski, A. (2006a)** What to acquire from corpora in automatic valence acquisition, w Violetta Koseska-Toszewa and Roman Roszko, (red.) *Semantyka i konfrontacja językowa*, **3**.
- Przepiórkowski, A. (2006b)** On Heads and Coordination in a Partial Treebank, *Proceedings of the TLT 2006*: 163–174.
- Przepiórkowski, A. (2007a)** On Heads and Coordination in Valence Acquisition, *Proceedings of the Computational Linguistics and Intelligent Text Processing (CICLing 2007)*, Springer Verlag, LNCS series: 50–61.
- Przepiórkowski, A. (2007b)** A Preliminary Formalism for Simultaneous Rule-Based Tagging and Partial Parsing, *Data Structures for Linguistic Resources and Applications: Proceedings of the GLDV 2007 Biannual Conference of the Society for Computational Linguistics and Language Technology*, Universität Tübingen, Gunter Narr Verlag: 81–90.
- Przepiórkowski, A., Bański, P., Dębowski, Ł., Hajnicz, E. i Woliński, M. (2003)** Konstrukcja korpusu IPI PAN, *Polonica XXII–XXII*: 33–38.

- Przepiórkowski, A. i Buczyński, A. (2007)** Shallow Parsing and Disambiguation Engine, *Proceedings of the 3rd Language & Technology Conference*, Poznań, Poland.
- Przepiórkowski, A. i Fast, J. (2005)** Baseline Experiments in the Extraction of Polish Valence Frames, *Proceedings of the IIS:IIPWM'05 Conference, New Trends in Intelligent Information Processing and Web Mining*, Springer: 511-520.
- Schulte im Walde, S. (2000)** Clustering verb semantically according to their alternation behaviour, *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, Saarbrücken, Germany: 747-753.
- Vossen, P., red. (1998)** *EuroWordNet: a multilingual database with lexical semantic network*, Kluwer Academic Publishers, Dordrecht, Holland.
- Wittek, A. (2002)** *Learning the meaning of change-of-state verbs: a case study of german child language*, Mouton de Gruyter, Berlin, New York.
- Woliński, M. (2004)** *Komputerowa weryfikacja gramatyki Świdzińskiego*, praca doktorska, IPI PAN, Warszawa.
- Woliński, M. (2006)** Morfeusz — a Practical Tool for the Morphological Analysis of Polish, *Proceedings of the IIS:IIPWM'06 Conference, Intelligent Information Processing and Web Mining*, Springer: 503-512.