

Zlecniodawca: IPI PAN
Zleceniobiorca: Jakub Kozakoszczak

Dzieło do umowy z dnia: 28.11.2016
Termin zakończenia pracy: 15.12.2016

Reprezentacja semantyczna przyimków i ich modyfikatorów dla parsera ENIAM

Wojciech Jaworski, Jakub Kozakoszczak

Zrealizowane w ramach projektu: „CLARIN – Polskie wspólne zasoby językowe i infrastruktura technologiczna”

Tytuł pracy zamówionej: „Koncepcja reprezentacji semantycznej przyimków uwzględniająca ich modyfikowalność”

Adres dzieła: http://wiki.nlp.ipipan.waw.pl/clarin/Parser%20kategorialny?action=AttachFile&do=view&target=JK_reprezentacja_przyimkow.pdf

Opracowanie dokumentu: Jakub Kozakoszczak

1	Koncepcja	2
1.1	Założenia ogólne	2
1.2	Znaczenie przyimka semantycznego	2
1.3	Człony relacji oznaczanej przez przyimek	3
1.4	Role tematyczne wyrażen przyimkowych	3
2	Przyimki polskie	5
2.1	Wybrane częste sensy i hasła leksykalne	5
3	Indeks przyimków	12

1 Koncepcja

1.1 Założenia ogólne

Proponowana reprezentacja rozwiązuje problem modyfikowalności przyimków przez przypisywanie im zmiennej odniesieniowej i nadanie roli tematycznej. Z uwagi na występowanie modyfikatorów dla przyimków semantycznych (np. „dość głęboko w szafie” , „5 km od domu” (dodatkowy argument „5 km”) „w szczególnym związku z” (modyfikacja przyimka złożonego), przyimki te nie mogą być reprezentowane jako predykaty wiążące identyfikator odniesienia swojego podrzędnika z identyfikatorem odniesienia nadrzędnika. Dlatego z każdym przyimkiem semantycznym wiążemy zmienną reprezentującą jego własne odniesienie i wiążemy ją z zmienną reprezentującą odniesienie nadrzędnika jedną z relacji oznaczających role tematyczne.

Drugim problemem jest wieloznaczność przyimków. Dla każdego przyimka identyfikujemy jego sensy, każdy traktujemy jak oddzielny przyimek (oddzielną jednostkę leksykalną), zaliczamy każdy do klasy podobnych sensów i wyczerpująco opisujemy uwzględniając dla każdego sensu rolę tematyczną, typ semantyczny całej frazy przyimkowej i preferencje selekcyjne przyimka. Preferencje selekcyjne przyimka semantycznego dotyczą nie tylko frazy rzeczownikowej zależnej, ale są też nakładane na jego nadrzędnik, zwykle czasownik lub rzeczownik.

Tak, jak fizyce korzysta się z modeli, o których wiadomo, że nie opisują wszystkiego i są na nie kontrprzykłady, podobnie przedstawiony tu model semantyki języka nie opisuje wszystkich zjawisk, jakie występują w języku, a w założeniu ma opisywać to, co powszechnie występuje w tekstach języka polskiego. Model ten ma dwa poziomy. Jeden z nich wskazuje podział przyimków na sensy i przypisuje im role tematyczne oraz preferencje selekcyjne. Drugi z nich stanowi uzasadnienie poprawności podziału w pierwszym poprzez formalne zdefiniowanie sensów przyimków w założonej „ontologii”. Pierwszy poziom jest bardziej użytkowy a drugi podaje model rzeczywistości.

1.2 Znaczenie przyimka semantycznego

Zgodnie z założeniami reprezentacji wypracowanej w ramach projektu CLARIN 1 reprezentujemy znaczenie przyimka jako jego algorytm znaczeniowy i odniesienie. Odniesieniami przyimków są relacje, np. jedna z możliwych relacji dla przyimka „blisko” jest zdefiniowana jako

$$v(\text{blisko}) = \{(X, Y) : \text{dist}(X, Y) \leq e\} \quad (1)$$

gdzie $e=0,01$ a „dist” jest to polimorficzna metryka zdefiniowana jak poniżej:

1. gdy X i Y są punktami w przestrzeni liniowej, jest to metryka Euklidesowa,
2. gdy X i Y są podzbiorami przestrzeni liniowej, jest to metryka Hausdorffa nad metryką Euklidesową. W szczególności, gdy X i Y są liczbami, wtedy

$$\text{dist}(X, Y) = |X - Y|, \quad (2)$$

3. gdy X i Y są parametryzowane przez czas, czyli traktujemy je jak funkcje takie, że $X(t)$ jest punktem lub podzbiorem przestrzeni liniowej, wtedy

$$\text{dist}(X, Y) = \sup_t (\text{dist}(X(t), Y(t))) \quad (3)$$

Algorytm znaczeniowy „blisko” zwraca `true` dla tych i tylko tych obiektów, które są relacją definiowalną w postaci

$$R = \{(X, Y) : \text{dist}(X, Y) \leq e\} \quad (4)$$

gdzie „dist” jest metryką, a e liczbą rzeczywistą większą od 0.

1.3 Człony relacji oznaczanej przez przyimek

- Przez B oznaczmy odniesienie **podrzędnika przyimka**.
- Przez P oznaczmy odniesienie **wyrażenia przyimkowego przestrzennego**.
- Przez R oznaczmy odniesienie **przyimka**.
- Przez A oznaczmy odniesienie **nadrzędnika wyrażenia przyimkowego**.
- W przypadku określeń czasu nadrzędnikiem przyimka jest zazwyczaj nazwa zdarzenia lub stanu.
- W przypadku określeń czasu podrzędnikiem przyimka jest zazwyczaj nazwa przedziału czasowego albo nazwa zdarzenia.
- Odniesieniami **nazw przedziałów czasowych** są interwały.
- Odniesieniami **nazw zdarzeń** są zdarzenia.
- Odniesieniami **wyrażeń przyimkowych będących określeniami czasu** są interwały.
- Przez *.beg* będziemy oznaczać początek interwału, przez *.end* koniec, a przez *.len* długość.
- Przez *.time* będziemy oznaczać czas zdarzenia/stanu, przez *.loc* będziemy oznaczać parametryzowane przez czas miejsce zdarzenia/stanu/przedmiotu.

1.4 Role tematyczne wyrażeń przyimkowych

Role tematyczne dla przyimków definiujemy jako trójargumentowe relacje, których pierwszym argumentem jest odniesienie nadrzędnika przyimka, drugim jest odniesienie przyimka, a trzecim odniesienie podrzędnika przyimka.

We wzorach poniżej relacja $X \sqsubset Y$ jest uogólnieniem zawierania zbiorów zdefiniowanym jak poniżej:

- gdy X i Y są punktami w przestrzeni liniowej, wtedy

$$X \sqsubset Y \text{ wtw } X \text{ jest podzbiorem otoczki wypukłej } Y \quad (5)$$

- gdy X i Y są parametryzowane przez czas, wtedy

$$X \sqsubset Y \text{ wtw } \forall t : X(t) \text{ jest podzbiorem otoczki wypukłej } Y(t) \quad (6)$$

Moim zdaniem wszędzie poniżej $R(X,Y)=1$ powinno być bez tego "=1", bo litera predykatowa jest w FOL funktorem zdania- a nie nazwotwórczym.

Definicje ról tematycznych

1. Loc

$$\text{Loc}(A, R, B) \equiv \exists P : \forall t \in A : A.loc(t) \sqsubset P(t) \wedge R(P, B.loc) = 1 \quad (7)$$

2. LocGoal

$$\begin{aligned} \text{LocGoal}(A, R, B) \equiv \exists P : A(A.end) \sqsubset P(A.end) \\ \wedge A(A.beg) \not\sqsubset P(A.beg) \wedge R(P, B.loc) = 1 \end{aligned} \quad (8)$$

3. LocSrc

$$\begin{aligned} \text{LocSrc}(A, R, B) \equiv \exists P : A(A.beg) \sqsubset P(A.beg) \\ \wedge A(A.end) \not\sqsubset P(A.end) \wedge R(P, B.loc) = 1 \end{aligned} \quad (9)$$

4. Time

$$\text{Time}(A, R, B) \equiv R(A.time, B) = 1 \quad (10)$$

5. Time

$$\text{Time}(A, R, B) \equiv R(A.time, B.time) = 1 \quad (11)$$

6. Count

$$\text{Count}(A, R, B) \equiv R(A.count, B) = 1 \quad (12)$$

7. Order

$$\begin{aligned} \text{Order}(A, R, B) \equiv \exists P : P \text{ jest porządkiem indukowanym} \\ \text{przez algorytmy znaczeniowe } A \text{ i } B \wedge R(A, B) = 1 \end{aligned} \quad (13)$$

8. Cond

$$\begin{aligned} \text{Cond}(A, R, B) \equiv \exists P : P \text{ jest porządkiem oddającym} \\ \text{zależności przyczynowo-skutkowe} \wedge R(A, B) = 1 \end{aligned} \quad (14)$$

Np. dla zdania „Strzał padł koło południa” mamy $\text{type}(A, \text{paść})$, $\text{type}(R, \text{koło}_1)$, $\text{type}(B, \text{południe})$ oraz wiążącą je rolę tematyczną $\text{Time}(A, R, B)$. Zadanie opisanie semantyki przyimków sprowadza się teraz do wskazania ich (przyimków)

1. lematu i schematu składniowego,
2. preferencji selekcyjnych nałożonych na nadrzędnik,
3. preferencji selekcyjnych nałożonych na podrzędnik,
4. roli tematycznej,
5. sensu przyimka – nazwy algorytmu znaczeniowego,
6. ewentualnego numeru sensu w istniejących słownikach języka polskiego,
7. przykładu

Oprócz tego potrzebujemy wskazać sensy identyczne oraz zdefiniować poszczególne algorytmy znaczeniowe.

Przyjmujemy, że przyimki mogą mieć wiele sensów. Oznaczamy je poprzez dodawanie do lematów przyimków kolejnych liczb naturalnych. Sensy różnych przyimków mogą być identyczne. Mają wtedy ten sam algorytm znaczeniowy, jak np. w przypadku blisko-
 znacznych przyimków przestrzennych $\text{koło}_1 = \text{blisko}_1 = \text{około}_1 = \text{niedaleko}_1 = \text{opoda}_1 = \text{pod}_2 = \text{obok}_1$. Algorytmy znaczeniowe przyimków definiujemy tak, by były niezależne od ról tematycznych. Przykładowa definicja algorytmu znaczeniowego przyimka „blisko” podana jest powyżej (definicja 4), kolejne dwie dla „w₁” (zarazem „do₁”, „z₁” i „podczas₁”) i „pod₁” podane są poniżej:

$$w_1(R) = \text{true} \text{ wtw } R \text{ jest relacją definiowalną w postaci } \forall R\{(X, Y) : X \sqsubset Y\} \quad (15)$$

$$\begin{aligned} \text{pod}_1(R) = \text{true} \text{ wtw } R \text{ jest relacją definiowalną w postaci:} \\ \{(X, Y) \in \text{MIEJSCE} \times \text{MIEJSCE} : X.\text{height} < Y.\text{height}\} \\ \cup \{(X, Y) \in \mathbb{R} \times \mathbb{R} : X < Y\} \cup \{(X, Y) : X \prec Y\} \end{aligned} \quad (16)$$

gdzie \prec to zdefiniowany kontekstowo ostry porządek.
 Jeśli X i Y są parametryzowane przez czas, to

$$R = \{(X, Y) : \forall t : (X(t), Y(t)) \in R\} \quad (17)$$

2 Przyimki polskie

2.1 Wybrane częste sensy i hasła leksykalne

Przyimki niesemantyczne (argumenty)

dla-9 przyimek wyznaczający obiekt odniesienia komunikowanego sądu, najczęściej osobę, z punktu widzenia której sąd ten jest wydawany

$$\text{Był to dla mnie rok pełen sukcesów.} \quad (18)$$

lemat i skład- nia	pref. sel. na nadrz.	pref. sel. na podrz.	rola tema- tyczna	sens (algo- rytm)	numer SJP	przykład
koło:gen	?	CZAS	Time	koło ₁	2	Zadzwoń! koło szóstej.
około:gen	?	CZAS	Time	około ₁		
blisko:gen	?	PRZED. FIZ.	Loc	blisko ₁	a	Mieszka blisko stacji metra.
koło:gen	?	PRZED. FIZ.	Loc	koło ₁		1
niedaleko:gen	?	PRZED. FIZ.	Loc	niedaleko ₁		Mieszkali niedaleko parku.
opodal:gen	?	PRZED. FIZ.	Loc	opodal ₁		Konie pasły się opodal lasu.
pod:inst	?	PRZED. FIZ.	Loc	pod ₂		3
obok:gen	?	PRZED. FIZ.	Loc	obok ₁		1
pod:acc		PRZED. FIZ.	LocGoal	pod ₂		
koło:gen	?	LICZBA	Count	koło ₁		Ma koło pięciu lat.
blisko:gen	?	?	?	blisko ₁		Obracał się blisko prezesa
podczas:gen	?	CZAS	Time	podczas ₁		
w:loc	?	PRZED. FIZ.	Loc	w ₁	1	
w:acc	?	PRZED. FIZ.	LocGoal	w ₁	2,4	
do:gen	?	PRZED. FIZ.	LocGoal	do ₁	3	
z:gen	?	PRZED. FIZ.	LocSrc	z ₁	1b	
po:loc	?	ZDARZENIE	Time	po ₁	6	
po:loc	?	ZDARZENIE	Time	po ₁	6	
po:loc	?	CZAS	Time	po ₁		
po:loc	?	PRZED. FIZ.	Loc	po ₁	3	
pod:acc	?	?	LocGoal	pod ₁	1,2	Kot spał pod biurkiem.
pod:acc	?	?	Count	pod ₁	14	Pod wiceministrem było pięciu dyrektorów departa- mentu.
pod:inst	?	?	Loc	pod ₁	1	Drzwi otworzyły się pod naporem tłumu. Kazano mu
pod:inst	?	?	Order	pod ₁	16	opuścić lokal pod groźbą eksmisji.
pod:inst	?	?	Cond	pod ₁	8,17	

do-12 przyimek komunikujący cel czynności

Oddał buty do naprawy. (19)

do-17 przyimek wprowadzający składnik zdania rządzony przez czasownik, przymiotnik, rzeczownik lub określający je

Przywiązał się do dziadka. Był podobny do siostry. (20)

do-18 przyimek wprowadzający samodzielne wypowiedzenia nieoznajmujące, m.in. przekleństwa, formuły grzecznościowe, odezwy

Do diabła! Do zobaczenia! Do broni! (21)

dokoła-21, dookoła-22 przyimek określający główny temat tego, o czym mowa w zdaniu

Wszystkie jego myśli krążyły dokoła rodzinnego domu. (22)

ku-35 przyimek wraz z przyłączanym rzeczownikiem komunikujący, że zbliża się dana pora roku lub dnia

Dni są coraz dłuższe i ma się ku wiosnie. (23)

ku-36 przyimek występujący w zdaniu komunikującym, że ktoś jest zainteresowany inną osobą pod względem erotycznym

Nasze dzieci mają się ku sobie. (24)

między-40 przyimek komunikujący, iż zależność, o której mowa, dotyczy wymienionych w zdaniu osób, zjawisk, przedmiotów lub faktów

Pokrewieństwo między nimi było bardzo dalekie. (25)

między-42 przyimek wprowadzający określenie przeznaczenia tego, co stanowi rezultat dokonanego podziału

Podzielił majątek między synów. (26)

między-44 przyimek komunikujący, iż wybór, przed którym stoi dana osoba, dotyczy wymienionych przedmiotów lub osób

Wybieraj między mną a nim. (27)

między-45 przyimek wprowadzający określenie zbioru, do którego należą lub spośród którego się wyróżniają charakteryzowane w zdaniu przedmioty lub osoby

Słynęła z urody między rówieśnikami. (28)

na-50 przyimek tworzący z nazwami czynności wyrażenia określające cel, skutek lub sposób odbywania się tych czynności

Składać na pół. (29)

na-51 przyimek łączący z nadrzędnymi czasownikami rzeczowniki stanowiące ich uzupełnienie

Liczyć na pomoc. (30)

na-52 z nadrzędnymi rzeczownikami wyrazy stanowiące ich uzupełnienie

Skrzynka na listy. (31)

Beneficient

dla-7 przyimek komunikujący o przeznaczeniu obiektu

Film dla młodzieży. (32)

Comparison

aniżeli-1 przyimek łączący podstawę porównania z wyrażeniem, które komunikuje zwykle wyższy stopień nasilenia cechy

Była ładniejsza aniżeli jej koleżanka. (33)

co-6 przyimek używany w połączeniach, komunikujących, że przedmioty pod względem danej cechy są tożsame

Krawat tego koloru co koszula (34)

jak-24 przyimek wprowadzający wyrażenie porównawcze

Nos miała jak kartofelek. (35)

Condition

dzięki-23 przyimek komunikujący o cechach, faktach lub osobach, które spowodowały, że dana sytuacja jest dla kogoś pomyślna

Uratował się dzięki przytomności umysłu. (36)

mimo-38 przyimek oznaczający rozbieżność między tym, co się dzieje, a tym, czego się należało spodziewać

Mimo późnej pory wyszedł na spacer. (37)

na_skutek-361

pod_wpływem-400

w_wyniku-479

z_powodu-504

Duration

na-48 przyimek tworzący wyrażenia oznaczające trwanie czegoś, okres, termin lub porę dziania się czegoś

Wyjechać na weekend. (38)

Location

blisko-3 przyimek komunikujący o małej odległości między przedmiotami

Mieszka blisko stacji metra. (39)

dokoła-20 przyimek komunikujący, że to, o czym mowa w zdaniu, ma miejsce ze wszystkich stron czegoś znajdującego się w środku

Dokoła klombów posadzono tulipany. (40)

koło-29 przyimek komunikujący o małej odległości między przedmiotami

Usiądź koło mnie. (41)

koło-31 przyimek przyłączający nazwę przedmiotu będącego obiektem czynności, o której mowa w zdaniu

Dłubał coś koło zegara. (42)

między-39 przyimek komunikujący, że to, o czym mowa, jest z obu stron otoczone czymś

Ulica ciągnęła się między domami. (43)

między-43 przyimek ograniczający dwustronnie zbiór cech, które można przypisać nietypowemu obiektowi, o którym mowa

Bluzka koloru między żółtym a brązowym. (44)

na-46 przyimek tworzący wyrażenia oznaczające miejsce dziania się lub znajdowania się czegoś

Pracował na polu. (45)

Location_goal

do-13 przyimek komunikujący kierunek ruchu

Skoczył do wody. (46)

do-15 przyimek wyznaczający górną lub dolną granicę czegoś

Inflacja spadła do dziesięciu procent. (47)

do-16 przyimek komunikujący o rezultacie procesu lub charakteryzujący stopień nasilenia tego procesu

Blacha rozpałała się do czerwoności. (48)

ku-32 przyimek przyłączający określenie miejsca, w kierunku którego ktoś zmierza lub coś prowadzi

Odwrócił się nagle i ruszył ku drzwiom. (49)

na-47 miejsce lub kierunek, będące celem ruchu, czynności

Pojechali na Mazury. (50)

Extra_result

ku-34 przyimek przyłączający określenie czyjegoś stanu wewnętrznego, który wytworzył się w następstwie zdarzenia lub zjawiska, o którym mowa w zdaniu

Ku radości maluchów spadł śnieg. (51)

Measure

na-49 przyimek tworzący wyrażenia oznaczające miarę, ocenę wielkości oraz zakres ich zastosowania

Gruby na palec. (52)

Member

bez-2 przyimek komunikujący zwykle brak, nieobecność czegoś lub kogoś

Las bez grzybów. Sukienka bez rękawów. Odejść bez pożegnania. (53)

Math

do-19 przyimek między liczebnikami oznaczający potęgowanie

Dwa do piętej. (54)

Purpose

dla-10 przyimek wprowadzający przyczynę jakichś działań

Na dziki polowano dla mięsa. (55)

dla-8 przyimek komunikujący cel czynności

Tupali nogami dla rozgrzewki. (56)

do-11 przyimek komunikujący o przeznaczeniu obiektu

filiżanka do kawy (57)

ku-33 przyimek przyłączający określenie celu czynności, o której mowa w zdaniu

Nakręcono ten film ku przestrodze młodzieży. (58)

Quantifier

co-5 przyimek używany w połączeniach charakteryzujących częstość, z jaką coś się powtarza

Przystawał co krok. (59)

jako-28 przyimek przyłączający liczebnik porządkowy, który wskazuje na kolejność charakteryzowanego zdarzenia względem innych

Sekretarz komisji przyszedł na zebranie jako pierwszy. (60)

Role

jako-25 przyimek wprowadzający określenie roli, jaką pełni jakaś osoba w danych okolicznościach, a także określenie stosunku tej osoby do innych osób

Jako gospodarz spotkania starał się być dla wszystkich życzliwy i miły. (61)

jako-27 przyimek wprowadzający określenie sposobu potraktowania lub oceny danego obiektu lub stanu rzeczy

Takie stany są określane jako depresyjne. (62)

Time

blisko-4 lub między punktami czasowymi

Było już blisko daty wyjazdu. (63)

jako-26 przyimek wraz z przyłączanym rzeczownikiem wskazujący na okres życia osoby, o której mowa w zdaniu

Jako dziecko przebyła wszystkie choroby zakaźne. (64)

koło-30 przyimek poprzedzający określenie liczby, ilości, miary lub czasu, który komunikuje, że jest ono przybliżone

Zadzwoniła koło szóstej. (65)

lada-37 przyimek łączący się z nazwą jednostki czasu, komunikujący, iż coś stanie się w najbliższym czasie

Przyjadą lada godzina. (66)

między-41 przyimek wprowadzający określenia granic czasowych, w których miało miejsce dane zdarzenie

Wróćę między drugą a trzecią. (67)

Time_goal

do-14 przyimek określający czas zakończenia zdarzenia

Spał do południa. (68)

3 Indeks przyimków

aniżeli-1	do-16	koło-31	na-46
bez-2	do-17	ku-32	na-47
blisko-3	do-18	ku-33	na-48
blisko-4	do-19	ku-34	na-49
co-5	dokoła-20	ku-35	na-50
co-6	dokoła-21	ku-36	na-51
dla-7	dokoła-22	lada-37	na-52
dla-8	dzięki-23	mimo-38	na_skutek-361
dla-9	jak-24	między-39	pod_wplywem-400
dla-10	jako-25	między-40	w_wyniku-479
do-11	jako-26	między-41	z_powodu-504
do-12	jako-27	między-42	
do-13	jako-28	między-43	
do-14	koło-29	między-44	
do-15	koło-30	między-45	