

Lista zjawisk semantycznych przeznaczonych do uwzględnienia w reprezentacji semantycznej parsera ENIAM

Wojciech Jaworski, Jakub Kozakoszczak

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

Tytuł pracy zamówionej: „Opracowanie listy zjawisk semantycznych przeznaczonych do uwzględnienia w reprezentacji semantycznej generowanej przez parser kategoriálny”

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

Opracowanie dokumentu: Jakub Kozakoszczak

1	Wstęp	
	Zakres i przeznaczenie listy	3
2	Lokalność i nielokalność składniowo-semantyczna	
	Opis zjawiska	3
2.1	Zjawisko wyciągania kwantyfikatorów poza zakres operatora spowodowane ko-referencją elementów argumentów operatora	4
3	Niejawne argumenty semantyczne	
	Opis zjawiska	4
3.1	Lista niejawnych argumentów semantycznych	4
4	Znaczenie relacji składniowych	
4.1	Uogólniona intersektywność	5
4.2	Rozbiór składniowy w podejściu semantycznym	6
4.3	Zależność roli tematycznej od znaczenia rzędników	6
5	Nieintersektywność	

6	Relacyjność	
	Opis zjawiska	7
7	Przyrematyczność	
	Opis zjawiska	8
8	Operatory nieprawdomówne (<i>nonveridical</i>)	
	Opis zjawiska	9
8.1	Niefaktywność	9
9	Cechy składniowo-semantyczne jednostek nieopisanych w zasobach	
	Opis problemu	10
9.1	Lista cech składniowo-semantycznych	10
10	Oznaczanie miary i liczebności referenta	
	Opis zjawiska	10
10.1	Użycie pojemnikowe	11
10.2	Podział liczebników ze względu na dokładność	11
10.3	Miara referentów niepoliczalnych	11
11	Niedenotatywne funkcje języka	
	Opis zjawiska	11
11.1	Metatekstowość	11
12	Podmiot epistemiczny	

1 Wstęp

Zakres i przeznaczenie listy

Lista zbiera zjawiska semantyczne, których spójne opisanie w ramach jednolitej implementowalnej reprezentacji znaczenia o szerokim pokryciu dla języka polskiego wymaga szczegółowych, głębszych badań. Są to przede wszystkim zjawiska polegające na wprowadzaniu przez wyrażenia elementów znaczenia do zdań w sposób inny podstawowy, czyli przez zawężającą modyfikację znaczenia bezpośredniego nadrzędnika składniowego. Ten podstawowy sposób nazywamy roboczo *lokalnym wkładem znaczeniowym*. Lista jest wstępnym rozpoznaniem, które posłuży jako mapa zadań w następnych etapach rozwoju kategoryjnego parsera składniowo-semantycznego ENIAM.

2 Lokalność i nielokalność składniowo-semantyczna

Opis zjawiska

Podzbiorem logiki pierwszego rzędu, który wystarcza do wyrażenia znakomitej większości wypowiedzi w języku naturalnym, jest logika egzystencjalno-koniunkcyjna (logika $\exists\wedge$). Język reprezentacji znaczenia parsera ENIAM jest bardziej złożony, bo wzbogacony między innymi o operator DSCR wiążący zmienną z formułą oraz pozalogiczne predykaty relacyjne, jak TYPE i COUNT i role tematyczne, jak AGENT i THEME. Istnieją jednak jednostki języka naturalnego, wśród nich kwantyfikatory, operatory logiczne i modyfikatory nieintersektywne, których znaczenie leksykalne (w sensie wkładu w znaczenie zdania) jest niewyraźne w logice $\exists\wedge$ z operatorem DSCR, np. dla języka polskiego

- każdy, prawie każdy, każdy z powyższych, co dziesiąty, każdy ... ma inny ..., nawzajem, razem, codziennie, nigdy, zawsze, tylko, jakiś, pewien, jedyny, oba, inny nie tylko ... lecz także

Jednostki te nazywać będziemy *jednostkami o nielokalnym wkładzie składniowo-semantycznym* lub po prostu *operatorami nielokalnymi*. Pozostałe jednostki mają *lokalny wkład składniowo-semantyczny*.

Interesują nas następujące cechy operatorów nielokalnych:

- konstrukcje składniowe, w których występują,
- semantyka, czyli wprowadzane zmienne i liczba argumentów,
- odwzorowanie konstrukcji składniowych na semantykę.

Nie będziemy opatrywać operatorów nielokalnych pełnymi hasłami leksykalnymi, a jedynie zaznaczać ich wystąpienia w formach logicznych w sposób taki, by potencjalne dodanie znaczenia leksykalnego nie wymagało dalszego zmieniania formy logicznej zdania.

Z uwagi na to, że operatory nielocalne wchodzą ze sobą w interakcje w zdaniu (np. mogą stać w różnej kolejności), zadane przez nas odwzorowanie konstrukcji składniowych na formę logiczną nie będzie stanowić przepisu na jednoznaczny semantykę, w szczególności nie będziemy na tym etapie rozwiązywać problemu niejednoznaczności zakresu operatorów nielokalnych.

Osobnym problemem jest reprezentacja nieciągłych operatorów nielokalnych, np. *Każda strona ma inny kolor* i interakcji między operatorami nielokalnymi innych niż zagnieżdżanie się zakresów np. w przypadku kwantyfikatorów rozgałęzionych w zdaniu Hintikki.

Negację traktujemy jako operator nielokalny mający jeden argument i nie wprowadzający zmiennej.

Spójniki „**jeśli ... , to ...**” oraz „**jeśli ... , ...**” użyte w znaczeniu logicznym uznajemy za operatory nielocalne.

2.1 Zjawisko wyciągania kwantyfikatorów poza zakres operatora spowodowane koreferencją elementów argumentów operatora

Jeśli rolnik ma krowę, to ją doi (1)

W powyższym przykładzie „jeśli” traktujemy jako logiczną implikację, „rolnik” i „krowa” są kwantyfikowane egzystencjalnie, a „ją” i „pro” (podmiot domyślny czasownika „doi”) są kwantyfikowane koreferencyjnie (ι). Aby rozwiązać koreferencję trzeba przenieść powyższe 4 kwantyfikatory ponad implikację co powoduje zamianę kwantyfikatorów egzystencjalnych przy „rolniku” i „krowie” na uniwersalne.

Analogiczne zjawisko będzie występować przy wszystkich operatorach nielokalnych mających więcej niż jeden argument.

3 Niejawne argumenty semantyczne

Opis zjawiska

Niezbędne jest rozpoznanie i oznaczenie argumentów niejawnych w leksemach i wypracowanie skutecznej koncepcji ich opisu i przetwarzania.

Za niejawne argumenty semantyczne uznajemy te argumenty, które są niezbędne do ustalenia znaczenia wyrazu mającego ten argument. Przykłady występowania wyrażeń z różnymi argumentami niejawnymi:

- większy – większy niż słoń
Znaczenia wyrazu „większy” nie da się obliczyć bez argumentu „niż słoń”.
- Biegnie. – On biegnie. – Jaś biegnie.
W pierwszym zdaniu inicjatora (INITIATOR) reprezentujemy oznaczając go niemym zaimkiem (PRO), w drugim jest on wyrażony za pomocą zaimka „on”, w trzecim jest wskazany jawnie jako „Jaś”. Aby obliczyć znaczenie PRO i „on” potrzebna jest wartość koreferencyjnego lub deiktycznego argumentu.
- Karol chce. – Karol chce pracować.
Brakującego argumentu składniowego czasownika „chcieć” nie traktujemy jako niejawnego argumentu semantycznego.

3.1 Lista niejawnych argumentów semantycznych

Po wstępnej analizie przyjmujemy istnienie sześciu typów niejawnych argumentów semantycznych:

1. Okazjonalne
Np. *tu*.
2. Deiktyczne
Np. *taki*.
3. Porównawcze
Leksemy z tym argumentem mogą przyjmować porównawczy argument przyimkowy.
4. Porządkowe
Np. *kolejny*.
5. Koreferencyjne
Np. *on*.
6. Relacyjne
Aby określić znaczenie przymiotników i liczebników z argumentem relacyjnym (*duży, sporo*) trzeba wiedzieć, jaki predykat jest modyfikowany przez dane słowo.

4 Znaczenie relacji składniowych

4.1 Uogólniona intersektywność

Ważną cechą semantyczną, która przysługuje nie wszystkim leksemom, a pozwala na reprezentowanie ich wkładu znaczeniowego jako zawężającego znaczenie szerszej frazy, jest *intersektywność*. Tradycyjnie rozumiana intersektywność przypisywana jest przymiotnikom i oznacza, że denotacja frazy rzeczownikowej z przymiotnikiem intersektywnym jest podzbiorem denotacji frazy bez niego.

Widzę czarnego kota. \rightarrow Widzę kota. (2)

Rozumienie to można uogólnić na przysłowki i mówić o przysłówkach intersektywnych. Zbadania wymaga, w jakich innych częściach mowy lub pozycjach składniowych da się mówić o podziale na wyrażenia nieintersektywne i intersektywne w sensie uogólnionym – wyrażenia takie powinny spełniać większość lub wszystkie testy:

Test 1. Zaczynając od korzenia drzewa rozbioru składniowego dodajemy kolejne węzły i sprawdzamy, czy ze znaczenia drzewa rozszerzonego wynika znaczenie drzewa mniejszego. Jeśli tak, dodawane węzły są intersektywne w sensie uogólnionym.

Test 2. Zaczynamy od liści i ucinamy je po jednym, sprawdzając, czy znaczenie zdania staje się bardziej ogólne. Jeśli tak, ucinane węzły są intersektywne w sensie uogólnionym.

Test 3. Porównujemy typy semantyczne wyrażenia złożonego i tego samego wyrażenia bez jednego z podrzędników. Jeżeli pierwszy z tych typów jest podtypem drugiego, podrzędnik jest intersektywny w sensie uogólnionym.

Jaś biegnie. \rightarrow Biegnie. (3)

$$\text{cztery jabłka} \subset \text{jabłka} \quad (4)$$

$$\text{Mieszkam nieopodal fontanny.} \stackrel{?}{\rightarrow} \text{Mieszkam nieopodal.} \quad (5)$$

Przykłady 3 i 5 ilustrują wynikanie zdań, przykład 4 mówi o zawieraniu się denotacji. Ostatni przykład pokazuje, że rozstrzygnięcia wymaga problem relacji znaczenia jednobrzmiących leksemów przysłówkowych i przyimkowych. Pożądaną opcją wymagającą zweryfikowania jest identyczność znaczeniowa wszystkich lub większości takich par *modulo* argument niejawny przysłówka.

4.2 Rozbiór składniowy w podejściu semantycznym

Do wyliczenia reprezentacji semantycznej zdania potrzebne są dane wejściowe – rozbiór składniowy w podejściu semantycznym. Otrzymanie takiego rozbioru wymaga przekształcenia bardziej niskopoziomowego rozbioru w podejściu morfoskładniowym z użyciem informacji o znaczeniu wzajemnych rzędników. Przekształcenie jest potrzebne, gdy bezpośredni podrzędnik węzła n nie spełnia jego ograniczeń selekcyjnych, ale podrzędnik następnego rzędu (wnuczek n) je spełnia. Spełnianie ograniczeń selekcyjnych przez leksem polega na tym, że jego typ semantyczny jest podtypem jednego z typów wskazanych w ograniczeniach selekcyjnych. W takim przypadku podrzędnik pierwszego i drugiego rzędu są zamieniane miejscami w drzewie rozbioru.

Typowym przykładem zadania, w którym opisana wyżej procedura jest nietrywialna, jest rozpoznanie, czy wystąpienie przyimka przewidzianego w ramie składniowej nadrzędnego czasownika jest realizacją tej ramy, jak „w głowie”, czy jest to przyimek semantyczny, jak w wyrażeniu „w samolocie” w przykładzie:

$$\text{W samolocie wielu osobom kręci się w głowie.} \quad (6)$$

W rozwijanej dla parsera ENIAM odmianie podejścia semantycznego o tym, czy bezpośrednim podrzędnikiem czasownika jest przyimek czy rzeczownik, decyduje ostatecznie to, czy preferencje selekcyjne argumentu ramy uzgadniają się z typem semantycznym rzeczownika czy frazy przyimkowej. Podobny problem dotyczy liczebników i rzeczowników w użyciu pojemnikowym (patrz sekcja 10.1)

4.3 Zależność roli tematycznej od znaczenia rzędników

Podrzedniki dzielimy na argumenty i modyfikatory rozumiane semantycznie. Rola tematyczna podrzędnika zależy za każdym razem od jego typu zależności składniowej, a ponadto

- w przypadku argumentu – od znaczenia nadrzędnika. Przykładowo, odniesienie podmiotu czasownika *czuć* to EXPERIENCER zdarzenia, czasownika *myśleć* – to AGENT zdarzenia, a jego dopełnienia – to THEME zdarzenia.
- w przypadku modyfikatora – od jego znaczenia (podrzednika). Przykładem jest modyfikator rzeczownikowy w narzędniku. Taki modyfikator może wprowadzać role INSTRUMENT, TIME lub MANNER w zależności od jego znaczenia:

Naprawiła szafę **młotkiem**. (7)

Naprawiła szafę **wieczorem**. (8)

Naprawiła szafę **własnym sposobem**. (9)

Oddzielnej uwagi wymagają też rzeczownikowe frazy wolne w celowniku, przymiotniki, które mają rolę tematyczną inną niż ATTRIBUTE i przysłówki, które mają rolę tematyczną inną niż MANNER.

5 Nieintersektywność

Modyfikatory nieintersektywne mają nielokalny wkład znaczeniowy. Przykładami modyfikatorów nieintersektywnych są

• „były”
były prezydent (10)

• „fałszywy”
fałszywy brylant (11)

• „co najmniej”
co najmniej cztery dni (12)

Były i *fałszywy* są funktorami, które tworzą wraz ze znaczeniem nadrzędnika pojęcie na tyle zmodyfikowane, że jego denotacja nie musi lub nie może mieścić się w przecięciu oryginalnych denotacji. Funktor *co najmniej* bierze za argument nazwę liczby bądź miary i zwraca nowe określenie liczby bądź miary, które nie jest zawężeniem określenia oryginalnego.

6 Relacyjność

Opis zjawiska

Przez relacyjność modyfikatora rozumiemy tradycyjnie tę jego cechę, że jego znaczenie zależy w pewnym ustalonym zakresie od znaczenia jego nadrzędnika. Dogodnym sposobem jej reprezentowania jest relacyjny argument niejawny (punkt 6 na liście niejawnych argumentów w części 3.1).

• „ledwo” w znaczeniu ‘prawie nie’
Ledwo zdał egzamin. (13)

Mamy tu funktor który modyfikuje pojęcie będące jego składniowym nadrzędnikiem (analogicznie jak leksem „były”). Funktor *ledwo* ma domyślny argument – referencyjny porządek, zależny od nadrzędnika.

• niedużo
niedużo mleka (14)

- dużo
dużo złota (15)

Te typowe przykłady przymiotników relacyjnych mają znaczenie jednoznacznie do-determinowane znaczeniem nadrzędnika. „Dużo złota” to inna ilość niż „dużo węgla”.

- już
już za daleko (16)

- jeszcze
jeszcze we wtorek (17)

- aż
aż cztery (18)

Te funktory mają domyślny argument wskazujący skalę jako czasową, przestrzenną itp. ale niedostarczony przez sam nadrzędnik. Nadrzędnik ogranicza ich znaczenie, niekoniecznie je determinując. (Ew. skala jest jednoznacznie przypisana do znaczenia leksemu a ich wielość wynika z wielości znaczeń leksemu)

7 Przyrematyczność

Opis zjawiska

Przyrematyczne są te partykuły, które obejmują swoim zakresem zawsze remat zdania, choć może to być niewidoczne na poziomie morfoskładni i szyku, często swobodnego.

- chyba
Oslo jest chyba stolicą Norwegii. (19)

Oslo jest chyba stolicą **Norwegii**. (20)

- także
Pójdziemy także do kina. (21)

- również
Pojedziemy również do Stanów. (22)

- też
Polecimy też do Berlina. (23)

- nawet
Popłyniemy nawet do Oslo. (24)

8 Operatory nieprawdomówne (*nonveridical*)

Opis zjawiska

Szereg wyrażeń języka można definiować jako funktory od argumentów zdaniowych, w tym przysłówki, partykuły i czasowniki przyjmujące dopełnienie zdaniowe. Kluczowym podziałem przy reprezentowaniu ich znaczenia jest to, czy prawdziwość frazy pociąga prawdziwość argumentu (*veridicality*) czy nie (*nonveridicality*). Przykładem jest większość partykuł i wyrażeń przyimkowych wyrażających modalność epistemiczną, jak

- „chyba”
Chyba idzie zima. (25)

- „pewnie”
Pewnie spadnie śnieg. (26)

- „może”
Może będzie przymrozek. (27)

- „być może”
Być może spadnie grad. (28)

- „na pewno”
Na pewno spadnie grad. (29)

- „oczywiście ”
Spadnie oczywiście grad. (30)

- „prawdopodobnie”
Prawdopodobnie będzie ślisko. (31)

które stanowią ocenę stopnia uzasadnienia sądu. *Prawdopodobnie* może też być partykułą modalną aletyczną wyrażającą możliwość i w tym znaczeniu też jest operatorem nieprawdomównym.

Tryb rozkazujący także jest nieprawdomówny, choć zależy to dodatkowo od znaczenia czasownika: Ze zdania „Podaj mi sól” wynika pragmatycznie, że teraz odbiorca nie podaje soli, ale ze zdania „Bądź punktualny” nie wynika, że odbiorca punktualny nie jest.

8.1 Niefaktywność

Najlepiej znanym podtypem nieprawdomówności jest niefaktywność, tradycyjnie przypisywana leksemom czasownikowym. W naszej analizie faktywność i niefaktywność przysługuje pozycji w ramie walencyjnej, co uwzględnia odmienne zachowanie podrzędników tego samego leksemu wprowadzanych przez różne spójniki podrzędne.

- „roić sobie, że _”
Anna roi sobie, że pada. (32)

- „chcieć, żeby _”
Anna chce, żeby nie padało. (33)

Oddzielnej analizy wymaga prawdomówność spójników leksykalnych i faktywność pozycji wypełnionych przez frazy czasownikowe niefinitywne – bezokolicznikowe, imiesłowowe, odsłownikowe, bezosobnikowe, którym nie przysługuje np. czas gramatyczny.

9 Cechy składniowo-semantyczne jednostek nieopisanych w zasobach

Opis problemu

Słownik *Walenty*, który jest podstawowym źródłem sformalizowanej informacji o cechach składniowo-semantycznych jednostek współczesnego języka polskiego, nie obejmuje niektórych klas gramatycznych lub obejmuje je w niedostatecznym stopniu. Według wstępnego rozeznania wiele z brakujących leksemów rzeczownikowych, przymiotnikowych, przysłówkowych i przyimkowych wnosi wkład znaczeniowy lokalnie, a więc w sposób najbardziej podstawowy. Niezbędny jest jednak ich opis składniowo-semantyczny pozwalający przypisać im odpowiednią rolę tematyczną. Dalsze potrzeby zasobowe obejmują typ semantyczny leksemów nieopisanych w Słownosieci.

9.1 Lista cech składniowo-semantycznych

Docelowy opis cech składniowo-semantycznych obejmuje następujące cechy:

1. Rola tematyczna łącząca wyraz z nadrzędnikiem, jeśli leksem podrzędnika ma ją do siebie przypisaną oraz jest uogólnienie intersektywny
2. Struktura składniowa argumentów leksemu (rama składniowa), np. dla przyimka będzie to
 - (a) Liczba argumentów jawnych przyimka (por. *między_a*, *między_i*)
 - (b) Okalającość przyimka (por. *między_a*)
 - (c) Pozycyjność przyimka (por. postpozycyjne *(100 lat) temu*)
 - (d) Przypadek wymagany przez przyimek
3. Preferencje selekcyjne i role tematyczne argumentów intersektywnych w sposób uogólniony
4. Typ semantyczny leksemu, jeśli leksem nie występuje w Słownosieci. Dla przyimków i spójników podrzędnych jest to typ semantyczny frazy, których są elementem głównym.

10 Oznaczanie miary i liczebności referenta

Opis zjawiska

Szczegółnej uwagi wymagają wyrażenia, których typem semantycznym jest ILOŚĆ dla wyrażeń policzalnych lub MIARA dla wyrażeń niepoliczalnych. Ich rolę tematyczną jest odpowiednio COUNT lub MEASURE.

10.1 Użycie pojemnikowe

Użycie pojemnikowe rzeczownika to takie, w którym rzeczownik odnosi do miary referenta swojego podrzędnika rzeczownikowego w dopełniaczu w morfoskładniowym (niesemantycznym) podejściu do składni. Przykładem jest „łyżka” w wyrażeniu „łyżka cukru”. Żeby otrzymać rozkład zdania zawierającego rzeczowniki w użyciach pojemnikowych w składni semantycznej, jak Universal Dependencies, traktujemy pojemniki jak podrzędniki swoich tradycyjnych składniowych nadrzędników. Do tego celu niezbędne jest oznaczenie pojemników w języku polskim.

10.2 Podział liczebników ze względu na dokładność

Liczebniki niewłaściwe (*trochę, sporo*) i większość rzeczowników w użyciu oznaczającym miarę nie podaje dokładnej wartości wielkości, którą opisuje. Żeby reprezentować ich znaczenia, trzeba je podzielić ze względu na dokładność, opisać zakres niedokładności i zaproponować spójną koncepcję reprezentacji znaczeń niedokładnych.

10.3 Miara referentów niepoliczalnych

Do zdefiniowania, które leksemy wprowadzają w języku polskim relację MEASURE (np. *niedużo*), konieczne jest ich oznaczenie w zasobie typu słownikowego.

11 Nienotatywne funkcje języka

Opis zjawiska

Szczegółnej uwagi wymagają wszystkie leksemy i konstrukcje pełniące funkcję fatyczną (*halo*), konatywną (np. tryb rozkazujący), emotywną (np. ekspresywizmy) lub metatekstową (np. *że tak powiem*). Do opisu ostatniej należą złożone zagadnienia znaczenia spójników, wyrażen przyrematycznych i przytematycznych.

Przykłady wyrażen pełniących pierwotnie funkcję nienotatywną:

- „halo”
- „brawo”
- „br”
- „miau”
- „że tak powiem”

11.1 Metatekstowość

Wyrażenia metatekstowe, przede wszystkim partykuły i spójniki, odnoszą do fragmentu wypowiedzenia lub wyrażają komentarz na jego temat. Przykładem jest

- „więc”

Padło, więc wyszliśmy tańczyć. (34)

Spójnik *więc* informuje, że zachodzi związek quasi-przyczynowy między odniesieniami koordynowanych członów.

12 Podmiot epistemiczny

Niektóre wyrażenia i kategorie gramatyczne odnoszą do stanów mentalnych osoby niekończeniowo tożsamej z nadawcą. Należą tu

- partykuły modalne

– „chyba”

Piotr mówi, że chyba skończyli już układać. (35)

– „pewnie”

Piotr mówi, że pewnie skończyli już układać. (36)

W obu przykładach osoba referencyjna potwierdziłaby modyfikowany sąd, gdyby wiedziała, że pewna brakująca przesłanka jest prawdziwa, ale osoba tą nie jest nadawca zdania.

- partykuły miriatywne

– „nawet”

Anna narzekała, że nie przyszedł nawet przywitać się z rodzeństwem. (37)

Modyfikowany sąd dziwi podmiot epistemiczny, który ponownie nie musi być tożsamy z nadawcą.

- tryb rozkazujący, ponieważ wskazuje na intencje nadawcy
- spójniki przeciwstawne i wprowadzające okoliczniki przyzwolenia

– „choć”

Maria powiedziała, że wyszli tańczyć, chociaż padał deszcz. (38)

– „ale”

Maria powiedziała, że padał deszcz, ale oni tańczyli. (39)

– „a”

Maria powiedziała, że padał deszcz, a oni tańczyli. (40)