
Wstępna lista zjawisk semantycznych przeznaczonych do szczegółowej analizy

Wojciech Jaworski, Jakub Kozakoszczak

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

Tytuł pracy zamówionej: „Opracowanie wstępnej listy zjawisk semantycznych, które podlegać będą szczegółowej analizie”

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

Opracowanie dokumentu: Jakub Kozakoszczak

1	Zakres i przeznaczenie listy	3
2	Operatorowość	3
2.1	Zjawisko wyciągania kwantyfikatorów poza zakres operatora spowodowane koreferencją pomiędzy argumentami elementami argumentów operatora	4
3	Niejawne argumenty semantyczne	4
4	Znaczenie relacji składniowych	5
4.1	Uogólniona intersektywność	5
4.2	Rozbiór składniowy w podejściu semantycznym	6
4.3	Zależność roli tematycznej od cech pozaskładniowych	6
5	Cechy semantyczne wyrażen nieintresektywnych	7
5.1	Funktorowość zmieniająca pojęcie	7
5.2	Funktorowość wyluskująca cechę pojęcia	7
5.3	Zależność znaczenia pojęcia od typu semantycznego nadrzędnika - relacyjność	8
5.4	Przyrematyczność	8
5.5	Faktywność	8
5.6	Nieweredyczność	8
5.7	Metajęzykowość	9

5.7.1	Odwołanie do osoby referencyjnej, subiektywność	9
6	Cechy składniowo-semantyczne jednostek nieopisanych w zasobach	9
7	Oznaczanie miary i liczebności referenta	10
7.1	Użycie pojemnikowe	10
7.2	Podział liczebników ze względu na dokładność	10
7.3	Miara referentów niepoliczalnych	10
8	Niedenotatywne funkcje języka	10
9	Zdarzenia, procesy (zmiany stanu rzeczy w czasie)	10
10	Relacje czasowe i przestrzenne	10

1 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 kategorialnego parsera składniowo-semantycznego ENIAM.

2 Operatorowość

Takie leksemy, czy konstrukcje, do wyrażenia których w logice nie wystarcza koniunkcja, predykaty relacyjne, dscr i kwantyfikatory egzystencjalny, tzn. nie są wyrażalne w logice egzystencjalno-koniunkcyjnej rozszerzonej o dscr.

Tzn. kwantyfikatory, albo operatory nielokalnie zmieniające strukturę zdania. Zwane dalej kwantyfikatory.

Przykłady: każdy, prawie każdy, każdy z powyższych, co dziesiąty, każda strona ma inny kolor, nawzajem, razem, codziennie, nigdy, zawsze, tylko, jakiś, pewien, jedyny, oba, inny nie tylko ...lecz także

Interesują nas następujące cechy kwantyfikatorów:

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

Nie będziemy definiować kwantyfikatorów a jedynie zaznaczać ich wystąpienia w formach logicznych w sposób taki, by po zdefiniowaniu kwantyfikatora znaczenie formy logicznej było zgodne ze znaczeniem zdania.

Z uwagi na to, że kwantyfikatory będące w zdaniu wchodziły ze sobą w interakcje (np. mogą stać w różnej kolejności) zadane przez nas odwzorowanie konstrukcji składniowych na semantykę nie będzie stanowić przepisu na jednoznaczność semantykę, w szczególności nie będziemy na tym etapie rozwiązywać problemu niejednoznaczności zakresu kwantyfikatorów.

Osobnym problemem jest reprezentacja kwantyfikatorów wprowadzonych przez kilka leksemów znajdujących się w różnych częściach zdania (np. Każda strona ma inny kolor.) i interakcji między kwantyfikatorami innej niż wpływ na swój zakres np. kwantyfikatory rozgałęzione w zdaniu Hintikki.

Negację traktujemy jako kwantyfikator mający jeden argument i nie wprowadzający zmiennej.

“jeśli ..., to ...” oraz “jeśli ..., ...” użyte w znaczeniu logicznym uznajemy za kwantyfikatory. Uwaga zgodnie z formalizmem wprowadzonym w Walentym “jeśli” wiąże zdanie podrzędne z orzeczeniem zdania nadrzędnego za pomocą roli tematycznej Condition.

Trzeba będzie zbadać użycia roli Condition i określić czy są one uogólnienie intersektywne. Jeśli nie są trzeba zbadać w jaki sposób należy je przekształcić do postaci formuły logicznej.

Rolę Manner też można traktować jako nielokalną.

2.1 Zjawisko wyciągania kwantyfikatorów poza zakres operatora spowodowane koreferencją pomiędzy argumentami elementami 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" (niemy podmiot "doi") są kwantyfikowane koreferencjnie (jota?). 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 kwantyfikatorach mających więcej niż jeden argument.

3 Niejawne argumenty semantyczne

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.

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 wyrażen 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.

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 przysłówki 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)

cztery jabłka \subset jabłka (4)

Mieszkam nieopodal fontanny. $\overset{?}{\rightarrow}$ Mieszkam nieopodal. (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 lekssem 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:

W samolocie wielu osobom kręci się w głowie. (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 7.1)

4.3 Zależność roli tematycznej od cech pozaskładniowych

Podrzędniki dzielimy na argumenty i modyfikatory rozumiane semantycznie, czyli tak, że rola tematyczna argumentu zależy przede wszystkim od znaczenia nadrzędnika i typu zależności składniowej między nim a podrzędnikiem.

były prezydent (7)

Przykładowo, odniesienie podmiotu *myśleć* to AGENT zdarzenia, odniesienie dopełnienia to jego THEME, a podmiot *czuć* to EXPERIENCER. W przypadku modyfikatorów, ich rola tematyczna nie zależy od znaczenia nadrzędnika, a od typu zależności i znaczenia podrzędnika. Przykładem jest modyfikator rzeczownikowy w narzędniku. Taki modyfikator może wprowadzać role INSTRUMENT, TIME lub MANNER.

Naprawiła szafę **młotkiem**. (8)

Naprawiła szafę **wieczorem**. (9)

Naprawiła szafę **własnym sposobem**. (10)

Podobnie oddzielnej uwagi wymagają rzeczownikowe frazy wolne w celowniku, przyimotniki, które mają rolę tematyczną inną niż ATTRIBUTE i przysłówki, które mają rolę tematyczną inną niż MANNER.

5 Cechy semantyczne wyrażeń nieintersektywnych

5.1 Funktorowość zmieniająca pojęcie

- “były” traktujemy jako funktor, który modyfikuje pojęcie składniowego nadrzędnika, tworząc nowe.

były prezydent (11)

- “ledwo” ma dwa znaczenia: 2. określające położenie pojęcia na jakiejś skali porównawczej, w jakimś porządku np.

Ledwo zdał egzamin (12)

Mamy tu funktor który modyfikuje pojęcie będące jego składniowym nadrzędnikiem (analogicznie jak leksem “były”). Funktor ten ma też domyślny argument (położenie w jakimś porządku) zależny od nadrzędnika

- co najmniej: funktor biorący określenie liczby bądź miary i generujący nowe określenie liczby bądź miary.

5.2 Funktorowość wyłuskująca cechę pojęcia

Modyfikuje aspekt znaczenia swojego podrzędnika, generując obiekt innego typu semantycznego niż podrzędnik. To jaki to jest aspekt jest zdeterminowane znaczeniem funktora. Leksem funktora może mieć kilka znaczeń dotyczących różnych aspektów.

- “ledwo” ma dwa znaczenia: 1. czasowe

Słońce ledwo weszło (13)

Ledwo zabrał się do pracy, zadzwonił telefon. (14)

“ledwo” traktujemy tu jak funktor biorący zdarzenie i generujący czas chwilę po nim. W funkcji spójnika podrzędnego, czas ten określa czas zdarzenia ze zdania nadrzędnego, w funkcji przysłówka czas ten określa czas zdarzenia wynikającego z kontekstu, co można wyrazić jako pro-zdarzenie.

- “gdy” czasowe bierze sytuację (zdarzenie) i generuje jej czas.

Wszedł, gdy go wpuścili. (15)

- pod

pod stołem (16)

Funktor biorący obiekt i zwracający miejsce, wskazujący, że jest to miejsce, gdzie znajduje się nadrzędnik.

pod stół (17)

Funktor biorący obiekt i zwracający miejsce, wskazujący, że jest to miejsce, do którego zmierza się nadrzędnik.

- ponieważ: funktor biorący sytuację i generujący CZEMU

5.3 Zależność znaczenia pojęcia od typu semantycznego nadrzędnika - relacyjność

- “ledwo” ma dwa znaczenia: 2. określające położenie pojęcia na jakiejś skali porównawczej, w jakimś porządku np.

Ledwo zdał egzamin (18)

Mamy tu funktor który modyfikuje pojęcie będące jego składniowym nadrzędnikiem (analogicznie jak leksem “były”). Funktor ten ma też domyślny argument (położenie w jakimś porządku) zależny od nadrzędnika

- dużo, niedużo: aż mają domyślny argument zależny od nadrzędnika “dużo złota” to inna ilość niż “dużo węgla”.
- już, jeszcze, aż mają domyślny argument wskazujący skalę jako czasową, przestrzenną itp. nadrzędnik ogranicza znaczenie, niekoniecznie je determinując. (ew. skala jest jednoznacznie przypisana do znaczenia leksemu a ich wielość wynika z wielości znaczeń leksemu)

5.4 Przyrematyczność

- także, również, też: są przyrematyczne, stwierdzają, że istnieją inne obiekty oprócz wskazanego, które spełniają resztę drzewa
- nawet - sąd dziwi osobę referencyjną, przyrematyczny, miriatywny

5.5 Faktywność

(napisać czym jest, czemu przysługuje, jakie pytania sobie z nią związane zadajemy)

podrzędniki zdaniowe rzeczowników i czasowników z Walentego - faktywność przysługuje pozycji w ramie walencyjnej

spójniki zdaniowe - faktywność przysługuje spójnikowi (zdaniu podrzędnemu).

faktywność fraz niefinitywnych (bezokolicznikowych, praet, ger,imps, ...), jak poradzić sobie z brakiem czasu gramatycznego

5.6 Nieweredyczność

- “prawdopodobnie” modalność aletryczna. (ew. jest to funktor, który wyciąga z sytuacji jej odległość od świata aktualnego)
- chyba, pewnie, prawdopodobnie, może, być może, na pewno, oczywiście stanowią ocenę stopnia uzasadnienia sądu, modalność epistemiczna
- tryb rozkazujący Ze zdania “Bądź bardziej punktualny” wynika pragmatycznie, że teraz słuchacz punktualny nie jest.

5.7 Metajęzykowość

Metajęzykowość jest to orzekanie czegoś o sądzie.

- więc: zachodzi związek quasiprzyczynowy między koniunktami (sytuacjami)

5.7.1 Odwołanie do osoby referencyjnej, subiektywność

- nawet - sąd dziwi osobę referencyjną, przyreumatyczny, miriatywny
- chyba, pewnie: osoba referencyjna potwierdziłaby modyfikowany sąd, gdyby wiedziała, że pewna kontekstualna przesłanka jest prawdziwa
- tryb rozkazujący wskazuje intencje mówcy
- chociaż, ale, a są miriatywne, poza tym są zwykłą koniunkcją faktów

6 Cechy składniowo-semantyczne jednostek nieopisanych w zasobach

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. 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.

7 Oznaczanie miary i liczebności referenta

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

7.1 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.

7.2 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.

7.3 Miara referentów niepoliczalnych

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

8 Nietenotatywne funkcje języka

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

halo, brawo, brr, miau, "że tak powiem"

9 Zdarzenia, procesy (zmiany stanu rzeczy w czasie)

czym się różni zdarzenie od sytuacji i od sądu koreferencyjność zdarzeń: "Zosia ukryła piłkę wkładając ją do szafy" - dwa czasowniki opisujące to samo zdarzenie.

10 Relacje czasowe i przestrzenne

Relacje czasowe (role tematyczne Time i Duration, pytanie kiedy?) i przestrzenne (role tematyczne Location i Path, pytania gdzie?, skąd?, dokąd?, którędy?)

250 10. Relacje przyczynowe, celowe, warunkowe i przyzwoleniowe (role Purpose i Condi-
251 tion)

252 Czy rozróżnienie pomiędzy Purpose i Condition to rozróżnienie pomiędzy intencjo-
253 nalnymi i nieintencjonalnymi okolicznościami wpływającymi na podjęcie akcji.

254 _____
255 ————

256 Dodać w części wyraźnie oddzielonej możliwość przebadania też zjawisk elipsy i me-
257 tonimii i podobnych - patrz. rozdz. „Typy znaczeń wariantowych” w R. Grzegorzczkowi
258 „Wprowadzenie do semantyki językoznawczej” ss. 44-47. Ta część raportu musi być opa-
259 trzona zastrzeżeniem, że będzie wykonana tylko w razie, gdy zostanie czasu i środków.