

Wykład 8. Cechy semantyczne wyrazów (słownik Wordnet). Opis struktury semantycznej zdania.

Cechy semantyczne wyrazów (podejście składnikowe)

Cechy semantyczne opisują własności semantyczne wyrazów. Można je definiować dla wszystkich wyrazów z treścią semantyczną (np. w języku angielskim wyróżnia się dwie kategorie wyrazów: **Content Words** - wyrazy z treścią semantyczną oraz **Structural Words** - wyrazy strukturalne). Do pierwszej grupy zalicza się: rzeczowniki, czasowniki, przymiotniki oraz niektóre przysłówki.

Jednym ze sposobów opisu właściwości semantycznych wyrazów jest zastosowanie zestawu cech elementarnych. Podejście takie zastosowali autorzy "Słownika syntaktyczno-generatywnego czasowników polskich" do opisu własności fraz, jakie mogą wystąpić z danym czasownikiem. W pracy używany jest następujący zestaw cech semantycznych:

[+ Abstr] – abstrakcyjność, [- Abstr] – konkretność,
[+ Anim] – żywotność, [- Anim] – nieżywotność,
[+ Hum] – osobowość, [- Hum] – nieosobowość,
[Coll] – zbiorowość, [Elm] – żywioł, [Fl] – roślina, [Instit] – instytucja, [Instr] – narzędzie,
[Liqu] – płyn, [Mach] – maszyna, [Mat] – materiał, [Pers] – część

Cechy te mogą tworzyć różne kombinacje np.:

[- Abstr, - Anim] – obiekt materialny nieożywiony np. kamień, szkło
[+ Hum, Pars] – część ciała ludzkiego np. ręka, głowa, ząb

darować – 1. dawać na własność, podarować, ofiarować

$NP_N \text{ — } NP_D + NP_{Acc}$

NP_N – grupa rzeczownikowa w mianowniku (kto? co?)

NP_D – grupa rzeczownikowa w celowniku (komu? czemu?)

NP_{Acc} – grupa rzeczownikowa w bierniku (kogo? co?)

Oprócz wymagań składniowych słownik określa również wymagania semantyczne dla poszczególnych fraz:

$NP_N \rightarrow [+HUM]$

$NP_D \rightarrow [+HUM]$

$NP_{Acc} \rightarrow \begin{bmatrix} +ANIM \\ -HUM \end{bmatrix} \begin{bmatrix} -ABSTR \\ -ANIM \end{bmatrix}$

Do klasyfikacji wyrazów języka angielskiego, a tym samym do określenia ich cech semantycznych można użyć słownika WORDNET, który został opracowany na uniwersytecie w Princeton. Szczegółowe informacje na temat słownika znajdują się na stronie:

<http://wordnet.princeton.edu/>

Słownik ten jest dostępny w wersji elektronicznej (w roku 2006 zawierał około 150000 wyrazów).

Słownik zorganizowany jest w postaci pewnej sieci, która odzwierciedla zależności znaczeniowe między wyrazami. Podstawowe relacje występujące w słowniku to:

Relation	Definition	Example
Hyperym	From concepts to superordinates	<i>breakfast → meal</i>
Hyponym	From concepts to subtypes	<i>meal → lunch</i>
Has-Member	From groups to their members.	<i>faculty → professor</i>
Member-Of	From members to their groups.	<i>copilot → crew</i>
Has-Stuff	From things to what they're made of.	→
Stuff-Of	From stuff to what it makes up.	→
Has-Part	From wholes to parts	<i>table → leg</i>
Part-Of	From parts to wholes.	<i>course → meal</i>
Antonym	Opposites	<i>leader → follower</i>

Figure 16.5 Noun Relations in WordNet.

hiponimia – relacja łącząca wyraz o szerszym znaczeniu z wyrazem o węższym znaczeniu np. kwiat – róża, poruszać się – iść,

hiperonimia – relacja odwrotna do hiponimii

Uwaga !. Wyraz o szerszym znaczeniu nazywamy hiperonimem, wyraz o węższym znaczeniu hiponimem np. wyraz „kwiat” jest hiperonimem dla wyrazów: „róża”, „tulipan” itd.

synonimia – relacja między wyrazami o identycznym znaczeniu np. barwa – kolor

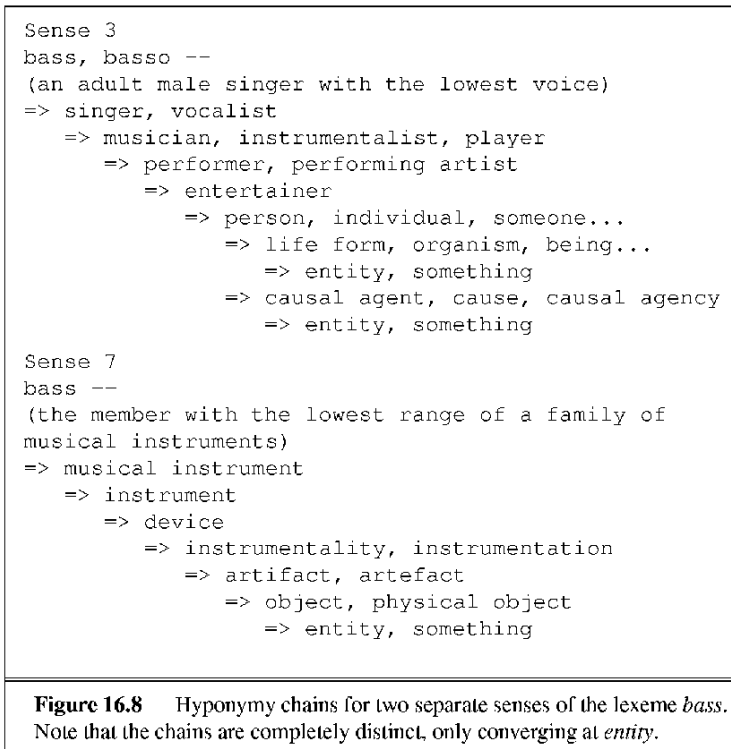
antonimia – relacja między wyrazami o przeciwnym znaczeniu np. duży – mały, ruszać się – stać

meronimia – relacja między pewnym obiektem (pojęciem), a jego częścią składową np. koło – samochód, palec – ręka

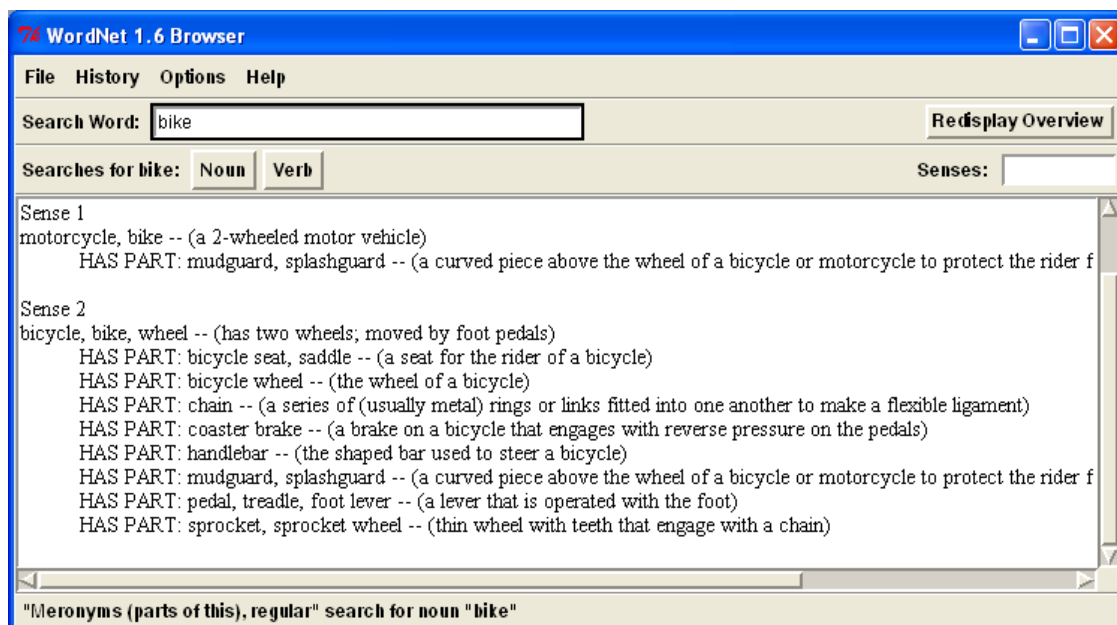
holonimia – relacja odwrotna do meronimii.

Uwaga ! Wyraz, który jest częścią składową lub fragmentem większej całości nazywa się meronimem np. wyraz „palec” jest meronimem wyrazu „ręka”, i odwrotnie” „ręka” jest homonimem wyrazu „palec”.

Relacja hiponimii dla rzeczownika „bass” (w znaczeniu „śpiewak” oraz „instrument muzyczny”):



Relacja meronimii dla rzeczownika „bike” (w znaczeniu „motorower” i „rower”):



Ponieważ między wyrazami o zupełnie różnym znaczeniu nie ma żadnych relacji słownik WORDNET jest faktycznie połączeniem wielu niezależnych sieci wyrazów (tzw. synsetów). W ten sposób można wyróżnić np. 15 kategorii (klas) czasowników oraz 25 kategorii rzeczowników.

Klasyfikacja rzeczowników

Symbol cechy	cecha semantyczna	Przykład
b _{n1}	<i>act, action, activity</i>	<i>walk, run, work, play, dance</i>
b _{n2}	<i>animal, fauna</i>	<i>mammal, fish, bird, dog, cat</i>
b _{n3}	<i>artefact</i>	<i>table, chair, car, train, house</i>
b _{n4}	<i>attribute, property</i>	<i>age, size, colour, consistency, taste</i>
b _{n5}	<i>body, corpus</i>	<i>head, arm, leg, eye, hand, finger</i>
b _{n6}	<i>cognition, knowledge</i>	<i>mind, perception, history, chemistry</i>
b _{n7}	<i>communication</i>	<i>language, voice, talk, letter, massage</i>
b _{n8}	<i>event, happening</i>	<i>accident, party, scandal, start</i>
b _{n9}	<i>feeling, emotion</i>	<i>love, hate, hope, happiness, apathy</i>
b _{n10}	<i>food</i>	<i>meat, egg, milk, juice, apple, lunch</i>
b _{n11}	<i>group, collection</i>	<i>class, family, herd, shoal, array</i>
b _{n12}	<i>location, place</i>	<i>room, garden, London, district</i>
b _{n13}	<i>motive</i>	<i>impulse, occasion, passion, reason</i>
b _{n14}	<i>natural object</i>	<i>sun, moon, sky, river, lake, mountain</i>
b _{n15}	<i>natural phenomenon</i>	<i>rain, snow, wind, fermentation</i>
b _{n16}	<i>person, human beings</i>	<i>man, woman, child, adult, doctor</i>
b _{n17}	<i>plant, flora</i>	<i>tree, flower, grass, corn, oak</i>
b _{n18}	<i>possession</i>	<i>present, donation, inheritance, salary</i>
b _{n19}	<i>process</i>	<i>synthesis, reaction, evolution, erosion</i>
b _{n20}	<i>quantity, amount</i>	<i>unit, number, byte, millimetre, degree</i>
b _{n21}	<i>relation</i>	<i>identity, parity, part, chronology</i>
b _{n22}	<i>shape</i>	<i>point, line, circle, column, solid</i>
b _{n23}	<i>state, condition</i>	<i>life, death, health, employ, freedom</i>
b _{n24}	<i>substance</i>	<i>atom, mixture, water, wood, oxygen</i>
b _{n25}	<i>time</i>	<i>hour, day, week, year, Monday</i>

Klasyfikacja czasowników

Symbol cechy	kategoria czasownika	Przykład
b _{v1}	<i>verbs of bodily functions</i>	<i>sleep, ache, tire, snore</i>
b _{v2}	<i>verbs of change</i>	<i>convert, increase, correct</i>
b _{v3}	<i>verbs of communication</i>	<i>order, thank, teach</i>
b _{v4}	<i>competition verbs</i>	<i>fight, race, arm, team</i>
b _{v5}	<i>consumption verbs</i>	<i>drink, eat, smoke, use</i>
b _{v6}	<i>contact verbs</i>	<i>rub, wipe, grip, slice, paint</i>
b _{v7}	<i>cognition verbs</i>	<i>deduce, induce, infer</i>
b _{v8}	<i>creation verbs</i>	<i>create, make, write</i>
b _{v9}	<i>motion verbs</i>	<i>run, go, swim, fly</i>
b _{v10}	<i>emotion or psyche verbs</i>	<i>love, hate, hope, fear</i>
b _{v11}	<i>stative verbs</i>	<i>be, live, exist, occur, have</i>
b _{v12}	<i>perception verbs</i>	<i>see, hear, smell, taste</i>
b _{v13}	<i>verbs of possession</i>	<i>have, give, receive, rob</i>
b _{v14}	<i>verbs of social interaction</i>	<i>impeach, educate, vote, try</i>
b _{v15}	<i>weather verbs</i>	<i>rain, snow, thunder, hail</i>

Ostatnio pojawiła się “polska” wersja słownika Wordnet. Szczegółowe informacje na stronie:

<http://plwordnet.pwr.wroc.pl/browser/index.jsp>

Należy zauważyć, że własności semantyczne posiadają również przymiotniki i przysłówki. Przymiotniki najczęściej określają rzeczowniki. Przymiotniki dzielone są na dwie główne grupy: jakościowe (opisowe) oraz relacyjne. Przymiotniki jakościowe opisują cechy (atrybuty) szeroko rozumianego obiektu reprezentowanego przez rzeczownik np. wielkość (*small, large*), wysokość (*low, high*), ciężar (*light, heavy*), kolor (*white, black*), wiek (*young, old*), temperaturę (*cold, hot*), ocenę (*bad, good*) itd. Przymiotniki relacyjne, które tworzone są przeważnie od rzeczowników opisują pewną relację do określanego rzeczownika np. "*musical* instrument".

Przysłówki są najczęściej określnikami czasownika, przymiotnika lub innego przysłówka. Przysłówki obejmują wiele podklas różniących się składniowo i semantycznie. Przysłówki dzielone są najczęściej na następujące grupy: sposobu (*well, fast*), miejsca (*here, out*), czasu (*today, early, soon*), stopnia (*enough, much, quite*), częstotliwości (*always, never, often*). Oddzielną grupę stanowią przysłówki zdaniowe, które modyfikują całe zdanie lub frazę wyrażając opinię mówiącego np. *surely, perhaps, probably* itd.

Analiza semantyczna – wybrane przykłady

W tym miejscu zostaną przedstawione wybrane metody analizy semantycznej zdań w celu uzyskania reprezentacji ich znaczenia. Zakres i dokładność uzyskanej reprezentacji znaczenia są uzależnione w znacznym stopniu od posiadania różnego rodzaju wiedzy oraz zastosowanych metod wnioskowania. W procesie analizy semantycznej wykorzystuje się następujące rodzaje wiedzy:

- wiedzę o znaczeniu wyrazów,
- wiedzę o powiązaniach między znaczeniem a strukturą gramatyczną zdania,
- wiedzę dotyczącą struktury wypowiedzi i jej kontekstu (dyskurs – wypowiedź dłuższa niż jedno zdanie),
- wiedzę „zdroworozsądkową” dotyczącą określonej dziedziny.

Uzyskana reprezentacja znaczenia danej wypowiedzi np. zdania, tekstu może być następnie wykorzystana np. w systemach typu pytanie-odpowieź, systemach wydobywania informacji z tekstów czy do maszynowego tłumaczenia.

Jedną z używanych metod jest analiza semantyczna bazująca na wynikach analizy syntaktycznej (ang. syntax-driven semantic analysis). Schemat tej metody przedstawia następujący rysunek:

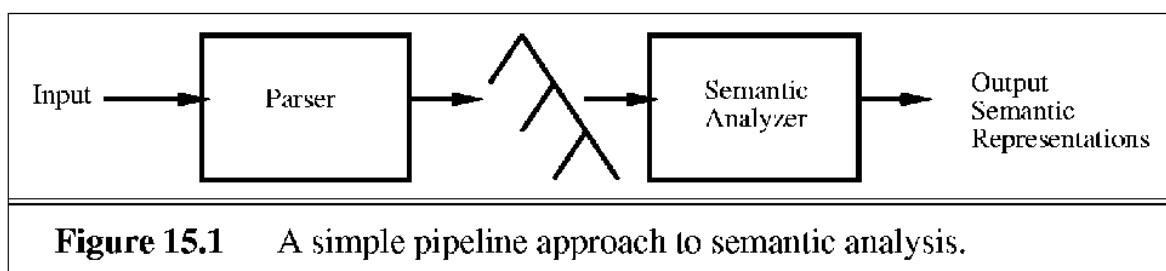


Figure 15.1 A simple pipeline approach to semantic analysis.

Przebieg analizy zdania „AyCaramba serves meat” przedstawia rysunek:

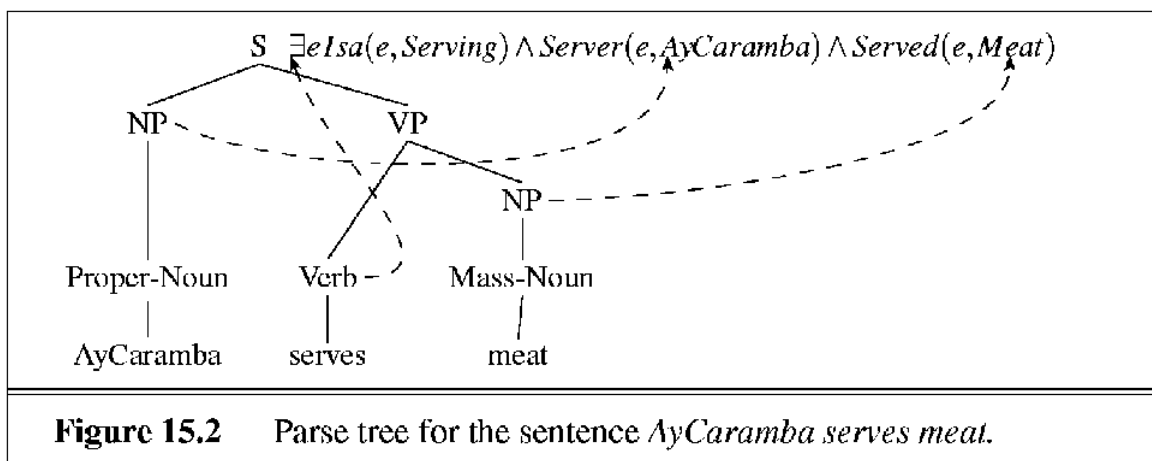


Figure 15.2 Parse tree for the sentence *AyCaramba serves meat*.

Określanie kategorii semantycznych grup słów

Prezentowany model reprezentacji semantyki zdań języka naturalnego oparty jest na gramatyce przypadków. Semantykę zdania określa zestaw kategorii semantycznych (np. ACTION, AGENT, OBJECT) oraz cech semantycznych (np. *motion_verb*, *contact_verb*, *person*, *animal*) przypisanych frazom występującym w zdaniu.

Podstawę określenia struktury semantycznej zdania stanowi podział zdania na grupy słów uzyskany w wyniku analizy syntaktycznej. Poszczególne grupy reprezentowane są przez odpowiednie symbole a_j . W rezultacie otrzymujemy:

$$\tilde{\mathbf{a}}(s_x) = a_{x1} \ a_{x2} \ \dots \ a_{xn} \quad , \text{ gdzie } n - \text{ilość grup w zdaniu}$$

Dla każdej grupy należy określić rdzeń, który stanowi w grupie podstawowy element semantyczny. Najczęściej rdzeń semantyczny pokrywa się z rdzeniem syntaktycznym grupy np. dla grupy werbalnej jest to czasownik leksykalny. Proces określenia rdzenia grupy jest zróżnicowany. W wielu przypadkach wystarczy znaleźć w grupie wyraz o określonej kategorii leksykalnej. Problemem może być jednak znalezienie właściwego rdzenia w grupie nominalnej (ze względu na jej złożoność). W języku angielskim grupa nominalna składa się z rdzenia, któremu towarzyszą przydawki przed rdzeniem (premodyfikacja) lub po rdzeniu (postmodyfikacja):

an expensive transistor radio

the title of the book

the man who lives next door

Problemy ze znalezieniem rdzenia grupy mogą być rozwiązane przez powiązanie tego procesu z analizą syntaktyczną zdania. W przypadku zastosowania gramatyki linków do podziału zdania na grupy słów, można określić również rdzenie poszczególnych grup. W tym celu można wykorzystać fakt, że grupa werbalna łączy się za pomocą linków strukturalnych z rdzeniami pozostałych grup.

Cechy semantyczne wyrazów przechowywane są w słowniku. Na podstawie słownika każdemu rdzeniowi grupy zostaje przypisana odpowiednia cecha semantyczna. Cecha ta jest uogólniana na całą grupę (grupa otrzymuje cechę semantyczną rdzenia). Cechy semantyczne odgrywają istotną rolę podczas określania kategorii semantycznej grupy.

Ostatecznie strukturę semantyczną zdania określają: ciąg kategorii semantycznych $\tilde{\mathbf{u}}$ oraz ciąg cech semantycznych $\tilde{\mathbf{b}}$ przypisanych rdzeniom grup:

$$\tilde{\mathbf{a}}(s_x) = a_{x1} \ a_{x2} \ \dots \ a_{xn} \quad - \text{ciąg grup słów}$$

$$\tilde{\mathbf{b}}(s_x) = b_{x1} \ b_{x2} \ \dots \ b_{xn} \quad - \text{ciąg cech semantycznych}$$

$$\tilde{\mathbf{u}}(s_x) = u_{x1} \ u_{x2} \ \dots \ u_{xn} \quad - \text{ciąg kategorii semantycznych}$$

Określanie kategorii semantycznych grup słów jest procesem złożonym. Jako podstawowe czynniki decydujące o kategorii semantycznej grupy przyjęto:

- cechę semantyczną rdzenia grupy,
- cechę semantyczną czasownika pełniącego w zdaniu funkcję orzeczenia,
- rodzaj grupy (rzeczownikowa, przyimkowa, przysłówkowa),
- rodzaj zdania (zdanie w stronie czynnej lub biernej) i pozycja grupy w zdaniu,
- w przypadku grupy przyimkowej rodzaj użytego przyimka.

Cecha semantyczna grupy jest jednym z najważniejszych czynników decydujących o kategorii semantycznej grupy. Wynika to z faktu, że cecha semantyczna w znacznym stopniu narzuca (lub wyklucza) konkretne kategorie semantyczne. Na przykład, grupie posiadającej cechę semantyczną *"person"* można przypisać jedną z kategorii: AGENT, CO-AGENT, EXPERIENCER, BENEFICIARY. Cechy semantyczne *"artefact"* lub *"natural_object"* narzucają natomiast kategorie semantyczne OBJECT lub INSTRUMENT, wykluczają natomiast kategorię AGENT. Analogicznie grupy o cechach semantycznych *"attribute"*, *"place"* oraz *"time"* powinny posiadać odpowiednio następujące kategorie: FEATURE, LOCATION i TIME.

Czasownik jako główny element zdania decyduje w znacznym stopniu o strukturze całego zdania. Cecha semantyczna czasownika można decydować również o kategorii semantycznej innej grupy. Rozpatrzmy następujące zdania:

s_1 – *John broke the window with a hammer*

s_2 – *John saw the girl with the long hair*

Bez znajomości cechy semantycznej czasownika nie można nadać właściwej kategorii semantycznej pierwszej grupie w obu zdaniach. Grupa *"John"*, posiadająca cechę semantyczną *"person"* otrzyma w zdaniu s_1 kategorię AGENT, w zdaniu s_2 kategorię EXPERIENCER. O właściwej kategorii decyduje cecha semantyczna czasownika: czasownik *"break"* należy do grupy *contact_verb* i podmiot zdania otrzyma kategorię AGENT, natomiast czasownik *"see"* do grupy *perception_verb* i podmiot zdania otrzyma kategorię EXPERIENCER.

Kategorię semantyczną grupy można określić również na podstawie jej kategorii leksykalnej. Najlepiej o tym świadczy przykład grupy werbalnej - w zdaniu pojedynczym posiada ona kategorię semantyczną ACTION. Grupa przysłówkowa może posiadać jedną z następujących kategorii: METHOD, LOCATION, TIME (o właściwej kategorii decyduje cecha semantyczna przysłówka np. grupa z przysłówkiem czasu lub częstotliwości otrzyma kategorię TIME).

O kategorii semantycznej grupy może decydować typ zdania (w stronie czynnej lub biernej) oraz pozycji grupy w zdaniu. Rozpatrzmy następujące zdanie:

s₃ - John gave Mary a book

Przypisanie właściwej kategorii semantycznej grupom "*John*" i "*Mary*" następuje na podstawie ich pozycji w zdaniu ("*John*" - AGENT, "*Mary*" - BENEFICIARY).

Kategoria semantyczna przypisana grupie może również zależeć od występującego w niej przyimka. Rozpatrzmy następujące zdania:

s₄ - A letter is written by John

s₅ - A book was given to John

W zdaniu *s₄* wyraz "*John*" otrzyma kategorię AGENT, natomiast w zdaniu *s₅* kategorię BENEFICIARY. Zdania *s₄* i *s₅* zawierają czasowniki o różnych cechach semantycznych: "*write*" - "*creation_verb*" oraz "*give*" - "*verb_of_possession*". O wyborze kategorii zadecydowała zarówno cecha semantyczna czasownika, jak i przyimek występujący w grupie.

Najważniejszym z wymienionych czynników wydaje się być cecha semantyczna grupy. W związku z tym dla każdej kategorii semantycznej u_i określono zbiór $\mathbf{B}(u_i)$, który zawiera zestaw oczekiwanych cech semantycznych:

$\mathbf{B}(\text{AGENT}) = \{\textit{person}, \textit{animal}\},$

$\mathbf{B}(\text{OBJECT}) = \{\textit{artefact}, \textit{body}, \textit{collection}, \textit{natural_object}, \textit{plant}\},$

$\mathbf{B}(\text{INSTRUMENT}) = \{\textit{artefact}, \textit{body}, \textit{natural_object}\},$

$\mathbf{B}(\text{LOCATION}) = \{\textit{place}, \textit{artefact}, \textit{natural_object}, \textit{event}\}$ itd.

Algorytm nadawania kategorii semantycznych grupom słów składa się z następujących kroków:

KROK 1 – znalezienie grupy werbalnej (otrzymuje ona kategorię ACTION),

KROK 2 – określenie typu zdania (rozkazujące, oznajmujące - w stronie czynnej lub biernej),

KROK 3 – określenie kategorii semantycznej pierwszej grupy w zdaniu, jeżeli zdanie jest oznajmujące,

KROK 4 – określenie kategorii semantycznych pozostałych grup w zdaniu.

Znalezienie grupy werbalnej polega na sprawdzeniu kategorii leksykalnych rdzeni poszczególnych grup. W przypadku zdania rozkazującego grupa werbalna jest pierwszą grupą w zdaniu, w zdaniu oznajmującym występuje najczęściej na drugiej pozycji. Analiza grupy werbalnej pozwala również stwierdzić, czy w zdaniu użyto strony czynnej lub biernej.

Kategoria semantyczna pierwszej grupy w zdaniu oznajmującym zależy od: cechy semantycznej grupy, cechy semantycznej czasownika oraz strony zdania. Ilustrują to następujące reguły:

AGENT – cecha grupy: *person*; cecha czasownika: *creation_verb*, *motion_verb*, *contact_verb*; strona czynna,

EXPERIENCER – cecha grupy: *person*; cecha czasownika: *emotion_verb*; strona czynna,

THEME – cecha grupy: *flora*, *natural_object*; cecha czasownika: *stative_verb*; strona czynna

INSTRUMENT – cecha grupy: *artefact*, *body*; cecha czasownika: *creation_verb*, *motion_verb*, *contact_verb*; strona czynna,

BENEFICIARY – cecha grupy: *person*; cecha czasownika: *contact_verb*; strona bierna,

OBJECT – cecha grupy: *artefact*, *natural_object*, *plant*; cecha czasownika: *creation_verb*, *verb_of_possession*; strona bierna.

W celu określenia kategorii semantycznych pozostałych grup w zdaniu opracowane zostały specjalizowane funkcje (oddzielnie dla każdej kategorii semantycznej u_i). Każda funkcja stanowi zbiór reguł opisujących warunki syntaktyczne i semantyczne, które powinna spełniać grupa, aby otrzymać daną kategorię semantyczną u_i . Pierwszym krokiem działania każdej funkcji jest sprawdzenie, czy cecha semantyczna b_r analizowanej grupy należy do zbioru $\mathbf{B}(u_i)$ związanego z daną kategorią semantyczną. Jeżeli $b_r \notin \mathbf{B}(u_i)$ działanie danej funkcji zostaje zakończone. Występujące w funkcjach reguły mają najczęściej postać koniunkcji kilku warunków. Poniżej przedstawiono przykłady reguł dla kategorii semantycznych OBJECT, LOCATION i METHOD.

Aby grupie przypisać kategorię OBJECT musi być spełniony jeden z warunków:

- grupa rzeczownikowa, cecha grupy: *artefact*, *natural_object*, *plant* lub *collection*,
- grupa rzeczownikowa, cecha grupy: *communication*, cecha czasownika: *creation_verb*,
- grupa rzeczownikowa, cecha grupy: *quantity*, cecha czasownika: *cognition_verb*.

Aby grupie przypisać kategorię LOCATION musi być spełniony jeden z warunków:

- grupa rzeczownikowa, cecha grupy: *place*,
- grupa przysłówkowa, rdzeń grupy: przysłówek miejsca (np. *here*),
- grupa przyimkowa z przyimkiem określającym miejsce (np. *in*, *to*, *on* itd.), cecha grupy: *place*, *artefact*, *natural_object*, *event*.

Aby grupie przypisać kategorię METHOD musi być spełniony jeden z warunków:

- grupa przysłówkowa, rdzeń grupy: przysłówek sposobu,
- grupa przyimkowa z przyimkiem "*in*", rdzeń grupy: rzeczownik "*order*", "*way*" itp.

Przykład 1

s_x – *Mary wrote John a letter*

a_1 – Mary	<i>person</i>	AGENT
a_2 – wrote	<i>creation</i>	ACTION
a_3 – John	<i>person</i>	BENEFICIARY
a_4 – a <u>letter</u>	<i>communication</i>	OBJECT

Przykład 2

s_x – *John was reading aloud in his room*

a_1 – John	<i>person</i>	AGENT
a_2 – was <u>reading</u>	<i>cognition</i>	ACTION
a_3 – aloud	<i>manner</i>	MANNER
a_4 – in his <u>room</u>	<i>artefact</i>	LOCATION

Podstawowe problemy, które wystąpiły w trakcie tworzenia algorytmu:

- syntaktyczny podział na grupy słów nie zawsze odzwierciedla strukturę semantyczną zdania

s_1 – *George (is in love)*

s_2 – *The Browns (will pay a visit) to Jane – The Browns will visit Jane*

s_3 – *Calculate the sum of numbers – Add the numbers*

- trudności z wyborem właściwej kategorii semantycznej

s_4 – *John should tell (the true) to his wife*

s_5 – *John spends (a lot of money) (on books)*

- zbyt duża ogólność niektórych cech semantycznych np. *artefact*

Ewolucja w stronę systemu uczącego się

Przyjęto następującą postać rekordu opisującego reguły w systemie:

- *poz* – pozycja grupy w zdaniu,
- *sent_voice* – strona zdania (czynna, bierna),
- *group_feature* – cecha semantyczna grupy,
- *verb_feature* – cecha semantyczna czasownika,
- *prep_feature* – rodzaj przyimka (lub konkretny przyimek)
- *group_cat* – kategoria semantyczna grupy
- *prev_group* – kategoria semantyczna poprzedniej grupy