Metodologiczne założenia semantyki komputerowej

Agnieszka Kułacka

King's College London

1. Wstęp

Semantyka komputerowa jest stosunkowo młodym działem językoznawstwa, który rozwinął się na bazie semantyki formalnej oraz wykorzystania narzędzi informatycznych do analizy znaczenia zdań. W ramach semantyki formalnej buduje się modele teoretyczne do opisu fragmentów języka naturalnego i założenia zastosowane dla niej przyjęły się na gruncie semantyki komputerowej. Semantyka formalna skupia w sobie podejścia reprezentowane przez semantykę prawdziwościową, semantykę teoriomodelową oraz gramatykę automatową zwaną też formalną, gramatykę kategorialną i gramatykę Montague'a.

Semantyka prawdziwościowa oraz semantyka teoriomodelowa korzystają szeroko z pojęcia możliwych światów. Świat możliwy opisuje całość stanów rzeczy należących do jednego płata Uniwersum, który obejmuje wszystko, co istnieje lub co można sobie wyobrazić. Co składa się na ten płat, czyli świat możliwy, zależy od decyzji definiującego. Metodologia badań semantyki prawdziwościowej opiera się na założeniu, że do określenia znaczenia zdania potrzebne jest zdefiniowanie fragmentu świata możliwego, w którym zdanie to opisywać będzie prawdziwy stan rzeczy. Natomiast w ramach semantyki teoriomodelowej buduje się matematyczne modele wartości semantycznych (znaczeń) wyrażeń językowych, opisujących stan rzeczy w świecie możliwym. Teoria gramatyk formalnych pełni rolę służebną w stosunku do semantyki teoriomodelowej, dostarczając jej teoretycznych narzędzi do budowania tych matematycznych modeli. Gramatyka formalna to uporządkowana czwórka $G = \langle V, \Sigma, \sigma, P \rangle$, gdzie V , Σ to alfabet pomocniczy i podstawowy, σ to wyróżniony element alfabetu pomocniczego, a P jest zbiorem reguł wyprowadzania. Reguły wyprowadzania posługują się rachunkiem zwanym gramatyką kategorialną. Wersja gramatyki kategorialnej, stworzona przez Ajdukiewicza w 1935 roku, definiuje kategorie syntaktyczne indukcyjnie. Dwie podstawowe kategorie

to: N (nazwa) i S (zdanie). Jeśli P i Q są kategoriami syntaktycznymi to P/Q jest również kategorią syntaktyczną. Richard Montague zakładał, że języki naturalne można opisać w taki sam sposób jak języki formalne i jego gramatyka kategorialna fragmentu języka angielskiego miała taki cel spełniać.

Wspólnym mianownikiem pokrótce omówionych wyżej podejść jest założenie o kompozycyjności i dyskretności znaczenia, statyczność analizy zdań, systematyczność ich interpretacji, bezkontekstowość lub ściśle zdefiniowany kontekst. W procesie rozumienia lub budowy znaczenia zdania następuje ścisły rozdział między semantyką a składnią. Analizator składniowy, zwany też parserem, dokonuje wstępnej analizy wypowiedzenia i podejmuje decyzję o jego gramatyczności. Dopiero wtedy następuje analiza semantyczna zdania. Proces ten jest automatyczny i zakłada unarność interpretacji znaczenia przez wszystkich użytkowników danego języka.

2. Semantyka formalna

Semantyka formalna jest owocem pracy logików i filozofów języka. Centralnym pojęciem w tej dziedzinie językoznawstwa jest znaczenie. W przeciwieństwie do innych podejść, takich jak psychologiczne budowanie konceptu w umyśle użytkownika języka, w semantyce formalnej znaczenie odnosi się do obiektywnej rzeczywistości i opisuje przedmioty istniejące poza umysłem jako znaczenie tych pojęć. Relacje istniejące między tymi przedmiotami stanowiącymi element obiektywnej rzeczywistości istnieją również poza umysłem jednostki i są postrzegane obiektywnie, a więc unarnie przez wszystkich użytkowników języka. Ważnymi założeniami dla budowania znaczenia większych całości z mniejszych jednostek są kompozycyjność i dyskretność atomów znaczeń, czyli jednostek znaczenia (Porter, Partee 2002; Lappin 1996). W niniejszym tekście omówię pokrótce najważniejsze podejścia do analizy znaczenia, które składają się na semantykę formalną.

2.1. Semantyka prawdziwościowa

W semantyce prawdziwościowej znaczenie zdania twierdzącego, a tylko takie się rozważa, może być poznane poprzez określenie warunków koniecznych i dostatecznych na to, by było ono uznane za prawdziwe. Innymi słowy, należy podać, jak powinien wyglądać świat możliwy, w którym zdanie to jest prawdziwe. Związek między znaczeniem zdania a możliwym światem oddaje jedną

z własności języka, a mianowicie jego kreatywność w sposobie opisywania stanów rzeczy (Dowty i in. 1981).

Rozważmy przykład: ‘Siedzę przy komputerze’. By zdanie to było prawdziwe, w rozważanym możliwym świecie muszą istnieć osoba, w tym wypadku – ja, jakiś komputer oraz między tymi elementami świata możliwego musi zachodzić pewien związek: osoba występująca w zdaniu musi siedzieć przed opisaną rzeczą¹. Znaczenie zdania ‘Siedzę przy komputerze’ to opisany stan rzeczy, zawarty w jednym z możliwych światów. Należy zwrócić uwagę, że samo zdanie jest fragmentem języka, natomiast jego znaczenie jest elementem świata możliwego. Zdanie ‘Siedzę przy komputerze’ jest wyrażone językiem przedmiotowym, natomiast opisane wyżej jego znaczenie wyrażone zostało metajęzykiem², jednym z systemów, który służyć ma temu właśnie celowi, a więc opisowi języka przedmiotowego. Innym systemem, który może służyć temu celowi, jest sztuka. Znaczenie zdania może oddać stosowny obraz zawierający wyżej opisane elementy i relacje między nimi. Przyjrzymy się prostemu programowi, napisanemu w języku Prolog:

```
szczęśliwy(jan).
słuchaMuzyki(kasia).
```

W tym możliwym świecie istnieją dwa obiekty: Jan i Kasia. O pierwszym wiemy, że jest szczęśliwy, a o drugim, że słucha muzyki. Jeśli więc wypowiemy zdanie: ‘Jan jest szczęśliwy’, w tym świecie jest ono prawdziwe i ma znaczenie *szczęśliwy(jan)*; znaczenie podane jest w metajęzyku. Jeśli wypowiemy zdanie ‘Kasia jest szczęśliwa’, zdanie to będzie fałszywe, ale możemy rozbudować świat możliwy, by włączyć *szczęśliwy(kasia)*. Jeśli natomiast będziemy chcieli powiedzieć ‘Okno jest szczęśliwe’, trudno będzie nam stworzyć świat, w którym *szczęśliwy(okno)* jest zdaniem prawdziwym³. Dzieje się tak dlatego, że szczęśliwy jest przymiotem rzeczy ożywionych.

2.2. Semantyka teoriomodelowa

Znaczenie w semantyce teoriomodelowej jest definiowane jako relacja między symbolami języka a pewnymi bytami, które są niezależne od tego języka. W skrajnych przypadkach relacja ta jest relacją równoważności między wyrażeniem a jego referentem w rzeczywistości. Referencja i prawda są podstawowymi

¹ Pominę tutaj dodatkowe znaczenie ‘siedzieć przy komputerze’, czyli pracować na komputerze.

² Terminy te zawdzięczamy Alfredowi Tarskiemu, uznanemu na świecie logikowi.

³ Pominę tutaj zjawisko metafory.

pojęciami, jakimi semantyka teoriomodelowa operuje i dzięki nim możemy opisywać znaczenie na jej gruncie (Gamut 1991).

Semantyka teoriomodelowa jest postrzegana jako metoda opisu języka za pomocą konstruowania matematycznych modeli, które oddawałyby znaczenie wyrażeń języka, posługując się tylko referencją i prawdą. Jednym z ważnych założeń, do którego wrócę w dalszej części niniejszego tekstu, jest założenie o kompozycyjności znaczenia wyrażeń złożonych. Znaczenie wyrażeń buduje się systematycznie ze znaczeń wyrażeń prostszych, posługując się pewnymi regułami języka. Modele buduje się indukcyjnie. Ich konstrukcję rozpoczyna się od określenia zbioru obiektów w rozważanym możliwym świecie oraz nadania interpretacji stałym języka w kontekście obiektów tego świata. Za pomocą wybranego sposobu przypisywania nadaje się interpretacje występującym w modelu zmiennym. Następnie na ich podstawie definiuje się wyrażenia złożone (Dowty i in. 1981).

Przyjrzyjmy się następującemu przykładowi programu języka Prolog. Wyrażenia pisane małą literą oznaczają stałe, wyrażenia zaczynające się od dużej litery oznaczają zmienne.

wlasnieZjadl(komar,krewJana).

wlasnieZjadl(zaba,komar).

wlasnieZjadl(bocian,zaba).

konsumuje(X,Y) :- vlasnieZjadl(X,Y).

konsumuje(X,Y) :- vlasnieZjadl(X,Z), konsumuje(Z,Y).

Widzimy, że w tym świecie możliwym komar właśnie wypił (a dosłownie zjadł) trochę krwi Jana, żaba właśnie zjadła komara oraz bocian właśnie zjadł żabę. Stałe w tym fragmencie języka to: *wlasnieZjadl*, *komar*, *krewJana*, *zaba*, *bocian* oraz *konsumuje*, które omówię poniżej. Zmienne to *X* i *Y*, za które można podstawić nazwy obiektów (nazwy odnoszące się do referentów) w tym świecie, czyli *krewJana*, *komar*, *zaba*, *bocian*, bo tylko takie występują w zdaniu utworzonym za pomocą predykatu *wlasnieZjadl* oraz *konsumuje*. Wyrażenia złożone to zdania⁴, np. *wlasnieZjadl(komar,krewJana)*, czyli komar właśnie wypił trochę krwi Johna. Zdanie złożone typu *A :- B* oznacza, że zdanie *A* jest prawdziwe, jeśli zdanie *B* jest prawdziwe, natomiast przecinek w ostatniej linii programu zastępuje spójnik 'i'. Dzięki językowi Prolog utworzyliśmy indukcyjnie model fragmentu rzeczywistości, posługując się pojęciami referencji oraz prawdy. Ten opis matematyczny świata możliwego zastępuje wypisanie wszystkich występujących w danym świecie możliwym obiektów oraz relacji między nimi.

⁴ Zdania prawdziwe w programie Prolog to wyrażenia zakończone kropką.

2.3. Gramatyka formalna

W badaniach semantyki komputerowej wykorzystuje się również gramatykę formalną. Gramatyka formalna to uporządkowana czwórka $G = \langle V, \Sigma, \sigma, P \rangle$, gdzie V , Σ to alfabet pomocniczy i podstawowy, σ to wyróżniony element alfabetu pomocniczego, a P jest zbiorem reguł wyprowadzania. Reguły wyprowadzania mają następującą formę: $x \rightarrow y$, gdzie x jest elementem $(V \cup \Sigma)^+$, a y jest elementem $(V \cup \Sigma)^*$, natomiast $(V \cup \Sigma)^+$ oznacza dowolny niepusty ciąg znaków jednego lub obu alfabetów, z kolei $(V \cup \Sigma)^*$ oznacza dowolny ciąg znaków jednego lub obu alfabetów; ciąg ten może być również pusty. Wyrażenia języka, w , utworzone są z alfabetu podstawowego i należą do zbioru $\{w \in \Sigma^* : \sigma \xRightarrow{*} w\}$, gdzie $\xRightarrow{*}$ to dowolny, również pusty, ciąg zastosowanych reguł wyprowadzania (Gross, Lentin 1970; Linz 2006).

Sprawdzanie, czy ciąg znaków należy do języka gramatyki, przebiega następująco:

- (1) Napisz wyróżniony element, σ .
- (2) Zastąp element alfabetu pomocniczego, stosując odpowiednią regułę ze zbioru P .
- (3) Powtarzaj krok (2) aż do wyeliminowania wszystkich zmiennych.

Rozważmy następujący przykład gramatyki formalnej:

$$G = \langle \{\langle s \rangle, \langle np \rangle, \langle pp \rangle, \langle n \rangle, \langle v \rangle\}, \Sigma, \langle s \rangle, P \rangle$$

gdzie P jest zbiorem następujących reguł:

$$\begin{aligned} \langle s \rangle &\rightarrow \langle np \rangle \langle v \rangle \mid \langle v \rangle \\ \langle np \rangle &\rightarrow \langle n \rangle \mid \langle n \rangle \langle pp \rangle \\ \langle pp \rangle &\rightarrow w \text{ bieli} \mid w \text{ czerni} \\ \langle n \rangle &\rightarrow \text{mężczyzna} \mid \text{kobieta} \\ \langle v \rangle &\rightarrow \text{strzela} \end{aligned}$$

Alfabet podstawowy to zbiór liter oraz znaku spacji. Symbol ‘|’ oznacza wybór, np. zapis $\langle n \rangle \rightarrow \text{mężczyzna} \mid \text{kobieta}$ oznacza, że rzeczownikiem jest słowo ‘mężczyzna’ albo ‘kobieta’. Zauważmy, że znak spacji jest ostatnim znakiem wyrazu.

Wyprowadzimy ciąg ‘kobieta w bieli strzela’. Za pomocą reguł wyprowadzania sprawdzimy, że nasz ciąg należy do języka gramatyki.

$$\begin{aligned} \langle s \rangle &\Rightarrow \langle np \rangle \langle v \rangle \Rightarrow \langle n \rangle \langle pp \rangle \langle v \rangle \Rightarrow \text{kobieta} \langle pp \rangle \langle v \rangle \Rightarrow \text{kobieta w bieli} \langle v \rangle \\ &\Rightarrow \text{kobieta w bieli strzela} \end{aligned}$$

Elementem wyróżnionym tej gramatyki jest $\langle s \rangle$. Z niego wyprowadzamy za pomocą pierwszej reguły wyprowadzania $\langle np \rangle \langle v \rangle$, a następnie stosując drugą regułę wyprowadzania, otrzymamy $\langle n \rangle \langle pp \rangle \langle v \rangle$. Z $\langle n \rangle$ wyprowadzimy ‘kobieta’, z $\langle pp \rangle$ – ‘w bieli’, a z $\langle v \rangle$ – ‘strzela’. A więc ‘kobieta w bieli strzela’ należy do zbioru ciągów języka $\{w \in \Sigma^* : \sigma \xrightarrow{*} w\}$, ponieważ $\langle s \rangle \xrightarrow{*} \text{kobieta w bieli strzela}$, jak pokazaliśmy powyżej. Natomiast ciąg ‘kobieta’ nie należy do tej gramatyki.

$$\langle s \rangle \Rightarrow \langle np \rangle \langle v \rangle \Rightarrow \langle n \rangle \langle v \rangle \Rightarrow \text{kobieta} \langle v \rangle$$

Jak widzimy, nie wyeliminowaliśmy wszystkich zmiennych. Ciąg „kobieta $\langle v \rangle$ ” nie należy do Σ^* , gdyż zawiera elementy, które nie należą do Σ , tj. $\langle v \rangle$.

2.4. Gramatyka kategoryalna

Pojęcie gramatyki kategoryalnej zostało wprowadzone przez Yehoshui Bar-Hillela w 1964 roku jako termin, odnoszący się zarówno do jego prac, jak i wcześniejszych publikacji dwóch polskich filozofów – Stanisława Leśniewskiego i Kazimierza Ajdukiewicza. W 1935 roku Ajdukiewicz zaproponował analizę struktur gramatycznych za pomocą kategorii. Każdy leksem może należeć do jednej lub więcej kategorii w zależności od konotacyjnych cech semantycznych znaczenia wyrazów. Ze zdefiniowanych wcześniej kategorii podstawowych można utworzyć zbiór wszystkich kategorii rekursywnie. Zbiór kategorii CAT to najmniejszy zbiór, spełniający następujące warunki: (1) jeśli A jest kategorią podstawową, to A należy do CAT, (2) jeśli A i B należą do CAT, to A/B również należy do CAT. Zbiór leksemów o kategorii A oznacza się przez B(A), a zbiór wszystkich wyrazów o kategorii A oznaczamy P(A). Ten ostatni definiuje się indukcyjnie w następujący sposób: (1) jeśli A jest kategorią i x należy do B(A), to x należy również do P(A), (2) jeśli x należy do P(A) i y należy do P(C/A), to y(x) (funkcja y określona dla argumentów x) należy do P(C) (Oehrle i in. 1988).

Przyjrzyjmy się przykładowi gramatyki, którą podałam w poprzednim podrozdziale. Rozróżniłam następujące kategorie: $\langle s \rangle$, $\langle np \rangle$, $\langle pp \rangle$, $\langle n \rangle$, $\langle v \rangle$. Niektóre z nich mogą zdefiniować jako kategorie podstawowe, a pozostałe jako pochodne, utworzone kompozycyjnie z podstawowych. Zbiór podstawowych kategorii będzie się składać z $\langle s \rangle$, $\langle n \rangle$. Fraza przyimkowa $\langle pp \rangle$ jest o kategorii $\langle pp \rangle =_{\text{def}} \langle np \rangle / \langle n \rangle$ co oznacza, że jest funkcją, która jako argument bierze wyrażenie o kategorii $\langle n \rangle$ i zwraca wartość kategorii $\langle np \rangle$. Kategoria $\langle v \rangle =_{\text{def}} \langle s \rangle / \langle np \rangle$ oznacza zbiór czasowników, które jako argument biorą wyrażenia o kategorii $\langle np \rangle$, a zwracają wyrażenia o kategorii $\langle s \rangle$.

Przeanalizujemy rozważane w poprzednim podrozdziale zdanie ‘kobieta w bieli strzela’.

Leksem ‘kobieta’ kategorii $\langle n \rangle$ łączy się z wyrażeniem ‘w bieli’ kategorii $\langle np \rangle / \langle n \rangle$, by utworzyć wyrażenie złożone kategorii $\langle np \rangle$. Leksem ‘strzela’ jest kategorii $\langle v \rangle$, czyli $\langle s \rangle$, $\langle np \rangle$, i łączy się z wyrażeniem o kategorii $\langle np \rangle$, tworząc wyrażenie o kategorii $\langle s \rangle$. Należy zauważyć, że jest wiele gramatyk kategorialnych, różniących się bazą kategorii podstawowych, zdefiniowanych w zależności od potrzeb opisywanego fragmentu języka.

2.5. Gramatyka Montague’a

Gramatyka Montague’a jest modelem, którego celem jest opisanie znaczenia wyrażeń języka naturalnego językiem logiki. Powstała w latach siedemdziesiątych, jako praktyczne rozwiązanie problemu samozwrotności języka, cechy, która powoduje semantyczne paradoksy, takie jak:

- (1) Zdanie (1) jest fałszywe.

Jeśli uznamy, że zdanie (1) jest prawdziwe, to znaczy, że jest fałszywe. Jeśli zdanie (1) jest fałszywe, to mówi prawdę. Samozwrotność języka powoduje, że jest on semantycznie zamknięty. Jako odpowiedź na ten paradoks, Alfred Tarski wprowadził dwa pojęcia: język przedmiotowy, czyli ten, który jest przedmiotem opisu, oraz metajęzyk, czyli język służący do opisu języka przedmiotowego. Polski logik zaproponował również stworzenie systemu warstw języka, gdzie warstwa n jest językiem przedmiotowym, a warstwa $n + 1$ jest metajęzykiem. Jest to rozwiązanie w przypadku, gdy warstwy należą do tego samego języka (Dowty i in. 1981; Gamut 1991).

Montague sformułował kilka modeli fragmentów języka angielskiego, z których najbardziej znany jest model PTQ (*proper treatment of quantifiers*;

Thomason, red. 1974). Wyrażenia językowe tłumaczone są najpierw na język logiczny, a powstałe wyrażenia logiczne interpretowane są w danym modelu. Pierwszym etapem budowania tego modelu semantycznego jest podział leksykonu danego fragmentu języka na podzbiory o określonych kategoriach. Dzięki ustaleniom gramatyki kategoryjnej oraz zasadom składniowym gramatyki Montague’a wiemy, które wyrażenia językowe mogą łączyć się z którymi, by powstawały wyrażenia bardziej złożone semantycznie.

Weźmy przykład zdania z poprzedniego podrozdziału: ‘kobieta w bieli strzela’. By była możliwa interpretacja znaczenia tego zdania w języku logicznym, konieczne jest zaproponowanie reguł tłumaczenia. W klasycznej gramatyce Montague’a reguły pozwalają znaczeniom zdań przyjmować postać zdań logicznych i znaczenie zdania ‘kobieta w bieli strzela’ to:

$$(2) \exists x ((W_BIELI) (KOBIEITA) (x) \wedge STRZELA (x))$$

Wyrazy pisane drukowanymi literami oznaczają leksemy, które zostały przetłumaczone na stałe logiczne danej gramatyki Montague’a. Wyrażenie to powstało po podstawieniu stałych logicznych do wyrażenia rachunku lambda:

$$(3) \lambda X. \lambda Y. \lambda Z. \exists x (XY (x) \wedge Z (x))$$

Wyrażenie to zredukowane do postaci (2) za pomocą redukcji beta:

$$(4) (\lambda X. \lambda Y. \lambda Z. \exists x (XY (x) \wedge Z (x))) (W_BIELI) (KOBIEITA) (STRZELA)$$

Zauważmy, że wyrażenie przyimkowe ‘w bieli’ modyfikuje rzeczownik ‘kobieta’. W_BIELI jest funkcją, która jako argument bierze znaczenie słowa ‘kobieta’, a jej wartością jest znaczenie frazy rzeczownikowej ‘kobieta w bieli’. Argumentem funkcji $(W_BIELI) (KOBIEITA)$ oraz $STRZELA$ jest element świata możliwego, a wartościami stałe boolowskie: prawda, fałsz.

W przypadku semantyki komputerowej wyrażenia językowe muszą zostać zapisane w takiej postaci, by były odczytywane i interpretowane przez program komputerowy. Należy jednak pamiętać, że forma zdania zostanie wybrana przez programistę.

3. Założenia semantyki formalnej

Omówiłam różne podejścia, wspomagające analizę znaczenia wyrażen językowych. Przyjrzymy się teraz ich wspólnemu mianownikowi, czyli założeniom, jakie wszystkie przyjmują na początku swoich badań. Jednym z najważniejszych jest kompozycyjność i dyskretność znaczenia, gdyż one pozwalają na jednoznaczne budowanie z wyrażen prostszych bardziej skomplikowane.

3.1. Kompozycyjność i dyskretność znaczenia

Zasada kompozycyjności znaczenia jest nazywana również zasadą Glottloba Frege'a, filozofa, który zauważył, że interpretacja wyrażen złożonych jest funkcją interpretacji wyrażen prostszych. Umożliwia ona też interpretację wyrażen nowych, wcześniej nieanalizowanych, jeśli elementy znaczenia są już wcześniej znane oraz konstrukcja, w jakiej są zawarte, została utworzona na podstawie wcześniej przyswojonych reguł tworzenia znaczeń złożonych ze znaczeń prostszych. Zasadzie tej podlega duża część wyrażen języka naturalnego, nie podlegają jej wyrażenia idiomatyczne oraz te, których dodatkowe znaczenie nadaje kontekst wypowiedzi.

W parze z zasadą kompozycyjności języka występuje dyskretność znaczenia, a więc jednostki znaczenia w wyrażeniu złożonym dają się jednoznacznie od siebie oddzielić, a ich współwystępowanie w tym wyrażeniu nie nadaje mu dodatkowego znaczenia. W semantyce formalnej odrzucane są wszelkie konotacje związane z całością wypowiedzi, a także zabarwienie kulturowe. Nie jest możliwa interpretacja metafor oraz przenośni, gdyż wymagają one analizy całości wyrażenia i jego znaczenia w danym kontekście na podstawie znajomości i analizy tej wypowiedzi w kontekście wyjściowym.

3.2. Unarność interpretacji znaczenia

W badaniach semantyki formalnej kolejnym istotnym założeniem jest jednakość interpretacji znaczenia przez wszystkich użytkowników języka. Semantyka formalna nie rozważa znaczenia wyrażen idiolektycznych, relacjonemów oraz innych zwrotów używanych przez pewien krąg osób. Zakłada się istnienie abstrakcyjnego użytkownika języka, który rozumie te same wyrażenia zawsze w taki sam sposób. Unarność interpretacji znaczenia związana jest z następnymi założeniami istniejącymi w semantyce formalnej, czyli statycznością i systematycznością analizy znaczenia wyrażen językowych.

3.3. Statyczność i systematyczność analizy zdań

Semantyka formalna odrzuca istnienie diachronii albo raczej zauważa znikomy jej wpływ na analizę znaczenia wyrażenia. Zarówno znaczenia leksemów, jak i reguł rządzących łączeniem znaczeń wyrażen niższych w wyższe są statyczne, nie podlegają zmianom ani przejściowym dyktowanym kontekstem, ani mającym charakter trwały.

Wyrażenia językowe interpretowane są systematycznie; istnieje zestaw reguł budowania znaczenia wyrażen złożonych ze znaczenia wyrażen prostszych i interpretacja jest prowadzona w sposób uporządkowany, według tego właśnie systemu reguł. Nie jest możliwe zastosowanie zasady redundancji, czyli nadmiarowości informacji, dzięki której człowiek jest w stanie zrozumieć częściowo zniszczoną wypowiedź. Taka wypowiedź jest niemożliwa do interpretacji w ramach semantyki formalnej.

3.4. Automatyczność procesu analizy zdań

Abstrakcyjny użytkownik języka, stojący za interpretacją znaczenia wyrażen językowych w semantyce formalnej, dokonuje jej w sposób automatyczny. Interpretacja dokonywana jest bez udziału świadomości i woli, co jest konsekwencją charakteru tego użytkownika, nie jest nim człowiek, oraz biegłości używania języka przez niego; ma on dostęp do wszystkich reguł jednocześnie. W czasie analizy zdania przez program wiemy, że reguły nie są dostępne jednocześnie, ale kolejno. Idealna kolejność, przy założeniu bezkontekstowości wypowiedzi to taka, która odzwierciedla frekwencję stosowania danych reguł przez użytkowników.

3.5. Kontekst

Jak już wspomniałam przy omawianiu poprzednich założeń, kontekst nie jest brany pod uwagę przy interpretacji znaczenia wypowiedzi. Kontekst rozumiany jest tutaj jako lokalizacja wypowiedzi w czasoprzestrzeni oraz w jednym z możliwych światów. Pierwszy element lokalizacji jest fizyczny i ten nie jest częścią interpretacji wypowiedzi w semantyce formalnej, natomiast drugi element, przestrzeń logiczna, pozostaje jedynym śladem kontekstu podczas tej interpretacji. Znaczenia wypowiedzi językowych rozważa się w danym możliwym świecie tak jak zostało to przedstawione w podrozdziale 2.1.

4. Semantyka komputerowa

Omówione w niniejszym tekście założenia, stanowiące tło badań w semantyce formalnej, w każdej jej postaci przyjęły się na gruncie badań nad znaczeniem w semantyce komputerowej. Należy jednak dodać, że proces interpretacji wypowiedzenia jest linearnie uporządkowany; interpretuje się kolejno wyrazy ciągu oraz z nich inkrementalnie buduje się znaczenie jednostek wyższych. Reguły w programie uporządkowane są frekwencyjnie, te częściej stosowane w wypowiedziach stoją wyżej niż te używane rzadziej. Jeśli na jakimś etapie analizy znaczenia okaże się, że zastosowano błędną regułę, program wraca do punktu podjęcia błędnej decyzji i sięga po następną w rzędzie, odpowiadającą za tę część struktury.

Przyjrzymy się następującemu programowi napisanemu w języku Prolog:

$s(A,B) :- np(A,Y),v(Y,B).$ (1)

$s(A,B) :- v(A,B).$ (2)

$np(A,B) :- n(A,B).$ (3)

$np(A,B) :- n(A,Y),pp(Y,B).$ (4)

$n([Word|W],W) :- lex(Word,n).$ (5)

$v([Word|W],W) :- lex(Word,v).$ (6)

$pp([Word|W],W) :- lex(Word,pp).$ (7)

$lex(kobieta,n).$ (8)

$lex(mezczyzna,n).$ (9)

$lex(strzela,v).$ (10)

$lex(w_bieli,pp).$ (11)

$lex(w_czerni,pp).$ (12)

Pierwsze cztery reguły to reguły tworzące zdania i frazy nominalne. Kolejne linijki programu przypisują leksemom ich kategorie, a ostatnia część programu to leksykon. Analizując prawdziwość zdania ‘kobieta w bieli strzela’, program zastosuje regułę (1), a następnie będzie próbował zastosować (3). Analizowanie drugiego członu wypowiedzi spowoduje, że okaże się to niemożliwe; niepowodzeniem skończy się stosowanie kolejno reguł (6) i (10). Podjęta będzie próba zastosowania reguły (4), co uruchomi ciąg stosowania reguł (5), (8), a następnie (7) i (11).

Widzimy, że w programie reguły wyprowadzania, które podałam jako przykład w gramatyce bezkontekstowej, zostały przetłumaczone na język komputerowy. Posłużyłam się kategoriami przy budowaniu leksykonu w tym programie. W innych programach pozostałe podejścia znalazły swoje zastosowanie, jak wykazałam przy ich omawianiu.

5. Kierunek dalszych badań

Dynamiczne zastosowanie reguł przez parsera oraz linearne rozumienie wypowiedzi są założeniami, które różnią semantykę komputerową od semantyki formalnej. One jednak sprawiają, że powstały model lepiej oddaje proces rozumienia języka. Następnym krokiem w rozwoju semantyki komputerowej będzie próba eliminacji założeń, które uproszczają model rozumienia języka, ale przez to powodują, że staje się on zbyt odległy od rzeczywistych procesów, zachodzących podczas rozumienia wypowiedzi. Uwzględnienie kontekstu jest tutaj największym wyzwaniem. Dzieje się tak dlatego, że podejmowane są dopiero próby formalnego opisu kontekstu za pomocą systemu indeksów, który nie został jeszcze w pełni opracowany.

Bibliografia

- Ajdukiewicz, K. [1935] 1960. „O spójności syntaktycznej”, [w:] *Język i poznanie* 1, Warszawa: PWN, 222–242.
- Dowty, D. R., Wall, R. E., Peters, S. 1981. *Introduction to Montague Semantics*, London: Reidel.
- Gamut, L. T. E. 1991. *Logic, Language and Meaning*, Chicago: University of Chicago Press.
- Gross, M. i A. Lentin. 1970. *Introduction to Formal Grammars*, Berlin: Springer.
- Lappin, S. (red.). 1996. *The Handbook of Contemporary Semantic Theory*, Oxford: Blackwell.
- Linz, P. 2006. *An Introduction to Formal Languages and Automata*, Boston: Jones and Bartlett.
- Ohehrle, R. T., Bach, E., Wheeler, D. (red.), 1988. *Categorial Grammars and Natural Language Structures*, Dordrecht: Reidel.
- Porter, P., Partee, B. H. (red.), 2002. *Formal Semantics, The Essential Readings*, Oxford: Blackwell.
- Thomason, R. H. (red.), 1974. *Formal Philosophy. Selected Papers of Richard Montague*, Forge Village: Murray.