

Refleksje nad refleksjami

Maciek Godek

11 stycznia 2009

Autorefleksja

Gdybyśmy uparli się, żeby w niniejszym tekście nie używać pojęć, których znaczeń wcześniej nie zdefiniowaliśmy, mielibyśmy duży kłopot: musielibyśmy bowiem wyjaśnić w pierwszej kolejności, co to znaczy „upierać się”, co to znaczy „używać”, co to znaczy „mieć” czy wreszcie — co to znaczy „znaczyć”; musielibyśmy podać definicję definicji i wyjaśnić znaczenie pojęcia pojęcia — i to (w założeniu) osobie, która nie ma żadnych intuicji związanych z tymi słowami. Trudno oczekiwać spełnienia zwłaszcza tego ostatniego postulatu: w najlepszym razie taka osoba widziałaby w tych słowach jedynie bezsensowne ciągi liter, a w najgorszym nawet nie przyszłoby jej do głowy, że tutaj jest coś napisane.

Czytelnik zawsze podchodzi do tekstu z pewnym zestawem pojęć nabytych wcześniej — trudno by było wyobrazić sobie artykuł całkowicie autoteliczny, wyjaśniający sam siebie od początku do końca, nie wymagający od czytelnika absolutnie żadnej wcześniejszej wiedzy do zrozumienia go. Nawet pusta kartka nie spełniałaby tego kryterium, bo każdy rozumiałby ten „symbol” na swój własny sposób (jednocześnie być może rozumiejąc jego wieloznaczność), ewentualnie mógłby nawet nie zwrócić na nią uwagi. Widać stąd, że zrozumienie ma charakter całkowicie podmiotowy — sens nie może się zmieścić w żadnym artykule, książce, filmie, obrazie jakiegokolwiek innej wypowiedzi — za każdym razem właściwym „dawcą” sensu jest podmiot usiłujący pojąć zastaną rzeczywistość (czy też: zrozumieć intencje autora danej wypowiedzi). Podobnie, jedna osoba nie może w dosłownym sensie przekazać jakichś treści innej osobie: może ją co najwyżej zainspirować do zrekonstruowania tych treści w swoim umyśle.

Powiemy zatem, że jedna osoba (rozumiejąca) rozumie drugą (rozumianą), jeżeli potrafi w swojej głowie zrekonstruować jej myśli na podstawie pewnych komunikatów. (Przez komunikat rozumiemy wszystkie obserwowalne zachowania rozumianej osoby; nie ma przy tym znaczenia, czy komunikaty są wysyłane świadomie, czy nieświadomie).

Od razu zauważamy, że nigdy nie możemy być całkowicie pewni, czy ktoś rozumie nas właściwie (bo nie dysponujemy bezpośrednim wglądem w cudzy umysł). O tym, czy ktoś nas zrozumiał, możemy wnioskować również wyłącznie na podstawie komunikatów zwrotnych, które nam ów ktoś wysyła. Innymi słowy, nie istnieje żadne obiektywne kryterium, które mogłoby orzec, że ktoś kogoś właściwie zrozumiał. Nie oznacza to oczywiście, że jedna osoba nie może zrozumieć drugiej — oznacza to jedynie tyle, że nie można całkowicie stwierdzić, że jedna osoba rozumiała drugą zgodnie z jej intencją.

Można zapytać, czy zrozumienie ogranicza się zawsze do rozumienia jednych ludzi przez drugich (do tej pory rozważaliśmy bowiem tylko ten przypadek). Czy człowiek jest w stanie zrozumieć coś, co nie jest drugim człowiekiem? Z drugiej strony, czy coś, co nie jest człowiekiem, może zrozumieć człowieka?

Najpierw rozważymy pierwszy przypadek. Często mówi się, że, ktoś coś rozumie, na przykład — że uczeń rozumie matematykę, że mechanik rozumie zasadę działania silnika, że naukowiec rozumie prawa fizyki. Czy mamy wówczas do czynienia z takim samym rozumieniem, jak w przypadku rozumienia człowieka przez człowieka?

Ktoś mógłby oczywiście powiedzieć, że uczeń, który rozumie matematykę, to po prostu uczeń rozumiejący swojego nauczyciela bądź rozumiejący autora podręcznika, z którego się uczy. Naturalnie jest w tym trochę racji, ale czy kiedy mówimy o rozumieniu matematyki, naprawdę mamy na myśli rozumienie nauczyciela matematyki? Wydaje się, że raczej nie — że osoba rozumiejąca matematykę tym się charakteryzuje, że potrafi przy pomocy jej narzędzi rozwiązywać różne abstrakcyjne problemy. Podobnie z mechanikiem — mówimy, że rozumie zasady działania silnika, kiedy ma w swojej głowie pewien model tego silnika i potrafi używać tego modelu do diagnozowania stanu prawdziwego silnika i podejmowania odpowiednich działań naprawczych (tj. usuwania wszystkich czynników mogących powodować anomalia w jego działaniu). Wreszcie o naukowcu powiemy, że rozumie prawa fizyki, jeśli potrafi z tych praw wyciągać odpowiednie wnioski, których prawidłowość może następnie zweryfikować przy pomocy odpowiednich metod.

Szczególnie dwa ostatnie przypadki trudno by było zinterpretować jako „rozumienie kogoś”. Ani silnik, ani prawa fizyki, nie są osobami. Nie jest jednak wykluczone, że do ich rozumienia używamy tych samych mechanizmów, przy pomocy których udaje nam się rozumieć innych ludzi.

Pozostaje nam do rozważenia jeszcze jeden przypadek: coś, co nie jest człowiekiem, ale jest zdolne do rozumienia ludzi. Oczywiście, nie możemy oczekiwać, że silnik albo prawo fizyki zrozumie człowieka — to byłoby niedorzeczne, głównie z tego powodu, że nie mielibyśmy żadnych przesłanek pozwalających stwierdzić, że rzeczywiście ma miejsce jakieś rozumienie. (Chociaż można wyobrazić sobie sytuację, w której silnik przez długi czas nie chce się uruchomić, by w końcu — ładnie poproszony przez kierowcę — odpalić. Wcale nie można wykluczyć, że w takim przypadku rzeczywiście mogłoby zajść zrozumienie człowieka przez silnik).

Filozofowie jednak często zastanawiają się nad możliwością myślenia maszyn, a w szczególności komputerów. Pytają: czy możliwe jest stworzenie myślących urządzeń w oparciu o deterministyczne prawa logiki? Czy istnieje jakiś sposób pozwalający sprawdzić, czy dane urządzenie jest świadome?

Jeden z pionierów informatyki i sztucznej inteligencji, Alan Turing, sformułował pewne pragmatyczne kryterium mające na celu rozwiązanie tego zagadnienia. Nazywa się ono testem Turinga i stwierdza tyle, że o myślącej maszynie możemy mówić wtedy, gdy uda się jej przekonać swojego rozmówcę-człowieka, że ma do czynienia z istotą świadomą swojego istnienia. Widać wyraźnie, że to kryterium ma zdecydowanie charakter operacyjny i w żaden sposób nie wyjaśnia, na czym polega myślenie czy bycie świadomym bycia.

Tymczasem można oczekiwać, że zdolność do stworzenia myślącej maszyny pozwoli nam lepiej zrozumieć samych siebie i nasze możliwości, a także — w jakimś sensie — rozszerzyć naszą naturę, niejako wykraczając ponad to, czym jesteśmy teraz.

Refleksja pierwsza: o modelach

Gdy opowiadaliśmy sobie o mechaniku rozumiejącym działanie silnika, wspomnieliśmy o mentalnym modelu tego silnika w głowie mechanika. Teraz postaramy się skupić na pewnych właściwościach i charakterystycznych cechach tego modelu. Oczywiście nie będziemy wchodzić w szczegóły techniczne związane z liczbą zaworów w cylindrach, bo to nie o to tutaj chodzi. Chcemy raczej pokazać, na czym polega model w ogóle, co jest jego istotą i jakie czynniki go konstytuują.

Przede wszystkim, silnik jest bytem samodzielnym. Chociaż przeważnie stanowi element jakiegoś większego urządzenia, na przykład samochodu, nie musi być rozważany w tym kontekście. Ważne jest to, że silnik jest źródłem ruchu (najczęściej obrotowego), a w praktyce jest urządzeniem przetwarzającym jakieś paliwo na ruch (operując abstrakcjami fizycznymi powiedzielibyśmy, że silnik jest urządzeniem przetwarzającym jeden rodzaj energii w inny).

Jednak to nie wszystko. Szybkością obrotu wału silnika da się sterować poprzez regulację dopływu paliwa do cylindrów. Ponadto pracujący silnik posiada w każdej chwili pewien stan wewnętrzny, który wprawdzie jest dość skomplikowany, ale można go dość dobrze przybliżyć poprzez podanie szybkości obrazania się wału silnika — wiadomo, że gdybyśmy nagle odcięli dopływ paliwa, wał obracałby się jeszcze chwilę siłą inercji.

Mechanik posiada jeszcze w swojej głowie ideę, w oparciu o którą silnik działa: eksplozja paliwa wewnątrz cylindra spowodowana iskrą ze świecy zapłonowej powoduje wypchnięcie tłoka, który wprawia w ruch wał silnika. Fakt, że iskra powoduje eksplozję benzyny, jest w tym modelu swego rodzaju aksjomatem — tak po prostu jest; mechanik raczej nie będzie się zastanawiał, co sprawia, że paliwo jest takie wybuchowe (dociekaniem tego typu zajmują się chemicy).

Zostawmy na razie naszego przyjaciela mechanika sam na sam z jego silnikiem i rozważmy nieco bardziej ludzki przykład. Zastanówmy się nad jakąś sztuką wystawianą w teatrze. Dramat najpierw jest czytany przez reżysera, który stara się zrozumieć główną wymowę dzieła oraz motywacje, jakie rządzą poszczególnymi postaciami. Następnie reżyser przydziela role poszczególnym aktorom i każdy aktor próbuje dokładnie zrozumieć i poczuć swoją postać — zarówno od strony przyczynowej (skąd biorą się intencje, czego ta postać tak naprawdę chce), jak emocjonalnej (co czuje dana postać, na jakie bodźce jest wrażliwa). Ostatnim ogniwem tego łańcucha jest widz, który ogląda spektakl — ten zaś próbuje się zidentyfikować z najbliższymi mu postaciami na scenie w różnych sytuacjach i wyciąga wnioski dotyczące swojego własnego życia.

Przykład może jest dość odległy od codziennych relacji międzyludzkich, ale wydaje mi się, że dobrze odślania pewne właściwości człowieka, które w normalnych sytuacjach są ledwie zauważalne. Przede wszystkim, człowiek posiada zdolność przeżywania życia innego, niż swoje własne — najdobitniej ukazują to aktorzy, którzy wręcz na tym zarabiają. Ale również reżyser i widz dokonują tej samej czynności, choć może na nieco innej zasadzie.

Zdolność do wyobrażania sobie siebie w nierzeczywistych sytuacjach jest absolutnym fundamentem dla zrozumienia drugiego człowieka. Mogę próbować przyjąć punkt widzenia swojego sąsiada i wyciągnąć z tego odpowiednie wnioski, na przykład odnośnie tego, jakie powinno być moje zachowanie.

Bynajmniej nie oznacza to, że kiedy próbuję przyjąć czyjś punkt widzenia, robię to w sposób poprawny — że w mojej głowie pojawia się chociaż jedna myśl, która jest obecna w głowie tej osoby. Jeżeli chcę się upewnić, czy mi się to udało, mogę zapytać tej osoby, ale ponownie podkreślam, że nigdy nie będę mógł mieć całkowitej pewności.

Powyższe rozważania mają stanowić punkt wyjścia dla następującego sformułowania: istotą zrozumienia czegoś jest dla człowieka wytworzenie w swoim umyśle modelu tego czegoś. Mechanik zajmuje się modelem silnika, logik — modelami logiki formalnej (ale nie tylko w sensie definiowanym przez teorię modeli), fizyk rozważa modele świata materialnego, neurofizjolog — modele ludzkiego mózgu. Dobry pisarz ma w głowie model czytelnika, reżyser — model widza. Aktor ma model siebie na scenie oraz odpowiednie modele dla każdej z granych przez siebie postaci. To oczywiście pewne uproszczenie — tak naprawdę każdy człowiek posiada bardzo wiele różnych modeli. Czy jest coś, co by je wszystkie łączyło?

Naturalnie, kiedy przyjrzymy się im bliżej, dostrzeżemy pewne cechy wspólne. W każdym przypadku mamy do czynienia z pewnymi działaniami. W przypadku silnika działaniami będą: wstrzyknięcie paliwa do cylindra oraz wystrzelenie iskry przez świecę zapłonową (skutkujące eksplozją oparów benzyny). To są działania wewnątrzmodelowe. Możemy wyobrazić sobie również ingerencję zewnętrzną w model: na przykład otwarcie albo przedziurawienie cylindra. Jeżeli spojrzymy na silnik z zewnątrz, nie analizując jego szczegółowego działania, dojdziemy do wniosku, że na pracujący silnik możemy wpływać poprzez regulację dopływu paliwa, silnik zaś będzie odpowiadał na to zmianą prędkości obrotowej wału.

Model, który posiada logik przeprowadzający dowód założeniowy, również zawiera pewien stan wewnętrzny (aktualny wiersz dowodowy) oraz zbiór dopuszczalnych działań (reguł wnioskowania), przy pomocy których można przekształcać formułę.

Gdy rozważymy model człowieka, który posiada, dajmy na to, jakiś dramatopisarz, to dostrzeżemy, że w ogólności człowiek zawsze funkcjonuje w jakimś otoczeniu, zawsze reaguje na nie w odpowiedni dla siebie sposób, zawsze coś sobie myśli i czuje i zawsze może jakoś oddziaływać na otaczającą go rzeczywistość. Model człowieka jest w pewnym sensie najpełniejszy ze wszystkich i dlatego to przede wszystkim na nim się skupimy w dalszej części tekstu.

Chcąc uogólnić wnioski, posłużę się następującą hipotezą odnośnie modeli ludzkich w umysłach: w modelu każdego obiektu czy przedmiocie możemy wyróżnić następujące elementy:

- środowisko, w jakim rozważamy dany obiekt bądź zbiór obiektów
- stan wewnętrzny, w jakim znajduje się obiekt
- możliwości oddziaływania środowiska na stan obiektu
- możliwości oddziaływania obiektu na środowisko
- reguły, w oparciu o które stan obiektu ewoluuje

W cybernetyce funkcjonuje pojęcie agenta, czyli czegoś, co posiada swój stan wewnętrzny, percepcję, możliwość działania oraz funkcję ewolucji. W dalszej części będziemy posługiwać się tym pojęciem zamiennie z obiektem bądź przedmiotem modelu. Postaramy się również sformalizować wyniki naszego wywodu, ale o tym później. Na razie wróćmy do naszego modelu umysłu ludzkiego.

Pytanie „kim jest człowiek?” uchodzi zazwyczaj za wielkie (i w gruncie rzeczy nierozstrzygalne) zagadnienie metafizyki. W historii ludzkości różnie próbowano na nie odpowiadać, za każdym razem kładąc akcent na różnych właściwościach istoty ludzkiej, i tak czasem nazywano człowieka „istotą dwunożną nieopierzoną”, innym razem „zwierzęciem rozumnym”, a jeszcze innym — odbiorcą końcowym produktów systemu ekonomicznego. Immanuel Kant twierdził (chyba nieco idealistycznie), że człowiek nigdy nie powinien być środkiem do celu, a zawsze powinien być celem. Znanych jest wiele cząstkowych odpowiedzi na to pytanie, zawsze ujmujących człowieka w różnych perspektywach jego istnienia.

My również nie będziemy się łudzić, że uda się nam udzielić pełnej odpowiedzi na to pytanie. Skupimy się jednak na pewnych aktach nieodrodnie związanych z ludzką umysłowością, ustarakając się je ująć w sposób możliwie najogólniejszy. Zaczniemy od tego, że człowiek z reguły posiada poczucie bycia sobą, jakieś swoje własne „ja”. Trudno by było dokładnie wymienić elementy, które wchodzą w zakres tego „ja”, ale wydaje się, że są to między innymi myśli i odczucia, które mają wpływ na zachowanie tego kogoś, a także — w pewnej mierze — sposób patrzenia na siebie w kontekście innych bytów (w szczególności zaś w kontekście społecznym). Można stwierdzić, że tak, jak człowiek potrafi wytworzyć w swoim umyśle modele innych ludzi, wytwarza również model siebie samego. Ma on, rzecz jasna, nieco odmienną naturę od pozostałych modeli, ale posiada również wiele cech wspólnych.

Samo poczucie „ja” oczywiście nie wystarczy. Elementem niezbędnym do wprawienia człowieka w ruch jest wola, chęć osiągnięcia jakiegoś stanu. Naturalnie sama wola nie wystarczy — niezbędna jest jeszcze możliwość podejmowania jakichś działań, przy pomocy których można ten stan osiągnąć.

Inżynier mógłby potraktować motywy ludzkiego postępowania jako próbę minimalizacji różnicy pomiędzy stanem pożądanym a aktualnym ze względu na działania, które trzeba wykonać (oczywiście, każde działanie oraz uchyb stanu należałoby wówczas opatrzyć odpowiednimi normami, czyli funkcjami odwzorowującymi je w zbiór liczb rzeczywistych). Wówczas wszystko, co przybliży nas do celu, możemy uznać za dobre (w tym ściśle określonym sensie), a wszystko, co nas od tego celu oddala — za złe. W taki sposób traktuje się człowieka na przykład w modelu ekonomicznym, gdzie jednak przyjmuje się dość arbitralne założenia odnośnie woli. Przedstawianie ludzkich czynności jako zadania optymalizacyjnego ma pewne oczywiste zalety — przede wszystkim matematyczną ścisłość, z jaką możemy opisać ludzkie życie. Musimy jednak pamiętać o tym, że nawet najlepszy opis matematyczny nie ujmie istoty tego, czym jest człowiek.

Do tej pory wyróżniliśmy sobie trzy czynniki wchodzące w skład ludzkiego umysłu: poczucie „ja”, wolę oraz działania. Zauważyliśmy, że w „ja” możemy wyróżnić „moje myśli” i „to, co czuję”, a także „jaki jestem”. Określając tę ostatnią przypadłość, odnieśliśmy się do świata zewnętrznego („tego wszystkiego, co nie jest mną”), a w szczególności do świata innych ludzi. Wcześniej zwróciliśmy jeszcze uwagę na to, że potrafimy wyobrażać sobie sytuacje, które nie mają miejsca w rzeczywistości — zdolność tę można uznać za elementarną operację umysłu. Analizując ludzkie działania z punktu widzenia optymalizacji sformulowaliśmy sobie pojęcia „dobry” i „zły”, zwracając przy tym uwagę, że mają wymiar relatywny i są związane z danym kryterium optymalizacji.

Wydaje się, że w powyższym wywodzie ujęliśmy skrótowo przynajmniej niektóre najważniejsze pojęcia związane z ludzkim umysłem. Wrócimy do tego tematu później, a póki co spróbujemy sobie odpowiedzieć na pytanie, czym jest pojęcie.

Refleksja druga: o pojęciach elementarnych

Na początku tej pracy postulowałem całkowitą wyjaśnialność kolejnych używanych w niej pojęć przez poprzednie. Doszedłem do wniosku, że jest to niemożliwe, a to z tego względu, że żadne pojęcie nie jest i nie może być zawarte w tekście: pojęcie jest czymś, co może się znajdować tylko i wyłącznie w umyśle; tekst może stanowić co najwyżej inspirację dla umysłu do wytworzenia jakichś pojęć. Następnie zacząłem pisać ten artykuł, używając wielu pojęć o różnej złożoności. Zasugerowałem między innymi, że należałoby podać definicję definicji. Teraz chciałbym przyjrzeć się temu pomysłowi nieco bliżej.

Proces definiowania polega na wyrażaniu pojęć złożonych przy pomocy pojęć prostszych. Brzmi to świetnie, ale skąd wiemy, że jedno pojęcie jest prostsze od drugiego? Wątpliwości może też budzić zagadnienie, czy możemy sobie definiować wszystkie pojęcia w nieskończoność, czy w końcu natrafimy na jakieś pojęcia niedefiniowalne?

Wszystko się wyjaśnia, jeśli dopowiemy sobie, że zawsze definiujemy w oparciu o jakiś system niedefiniowalnych pojęć elementarnych. Wobec tego dane pojęcie może być w jednym systemie pojęć prostsze od innego, ale w innym może już być bardziej złożone, a w jeszcze innym może być uznane za elementarne, czyli najprostsze. Można zapytać, czy istnieje jakiś system pojęć wyróżniony względem innych pod względem przyswajalności przez ludzki umysł.

Istnieją w tej kwestii stanowiska pesymistyczne. Na przykład doktor Rafał Urbaniak stwierdził, że „ogólnie twierdzenia o definiowalności, nieredukowalności i intuicyjnej jasności w odniesieniu do [języka] naturalnego są mocno ryzykowne i trudno je uzasadnić. Z listą [pojęć elementarnych] mamy też problem arbitralności”¹.

Innego zdania jest Anna Wierzbicka z Australian National University. Jest ona autorką i apologetką koncepcji Naturalnego Metajęzyka Semantycznego (NMS), tworzonego i weryfikowanego w oparciu o szczegółowe badania języków i kultur z całego świata. Wyróżniła listę około 60 pojęć języka naturalnego, które jej zdaniem powinny być zrozumiałe dla każdego człowieka i przy pomocy których można zdefiniować wszystkie inne pojęcia. Sama Wierzbicka zastrzega, że Naturalny Metajęzyk Semantyczny wymaga jeszcze wiele pracy, a jego koncepcja nie jest wolna od zastrzeżeń:

Nie twierdzę, że Naturalny Metajęzyk Semantyczny jest „językiem idealnym” w tym sensie, iż miałby stanowić ostateczną odpowiedź na poszukiwania wyrażanych językowo uniwersaliów pojęciowych. Jest to raczej pewne przybliżenie do ideału, wymagające poprawek, gdyż ów język tworzy się metodą prób i błędów; lepiej jest jednak operować próbnym i niedoskonałym zestawem indefinibiliów, niż żadnym. [10]

Metajęzyk Wierzbickiej znalazł szereg zastosowań w wielu dziedzinach badań językowych i kulturowych, ale może też stanowić doskonałą inspirację dla teoretyków umysłu: można bowiem mieć nadzieję, że stworzenie modelu umysłu potrafiącego w odpowiedni sposób interpretować i przetwarzać te pojęcia wystarczy do zaprojektowania systemu

¹korespondencja własna

rzeczywiście rozumiejącego ludzki język (a na pewno skuteczniejszego niż współczesne tłumacze kontekstowe oparte na bazach danych). Jedyne pytanie brzmi: co to znaczy „odpowiedni sposób”?

Myślę, że warto przytoczyć listę pojęć uznanych w tym języku za podstawowe. Zostały one w dość arbitralny sposób podzielone na kilkanaście kategorii (zarówno podział, jak i lista pojęć podlegają nieustannym modyfikacjom; poniższa lista pochodzi z [10]):

<i>elementy nominalne</i>	JA, TY, KTOŚ, COŚ (RZECZ), LUDZIE, CIAŁO
<i>predykaty mentalne</i>	MYŚLEĆ, WIEDZIEĆ, CHCIEĆ, CZUĆ, WIDZIEĆ, SŁYSZEĆ
<i>kwantyfikatory</i>	JEDEN, DWA, DUŻO, WSZYSTKO (WSZYSTKIE, WSZYSCY), NIEKTÓRE
<i>działania, zdarzenia i ruch</i>	ROBIĆ, DZIAĆ SIĘ (ZDARZAĆ SIĘ), RUSZAĆ SIĘ
<i>partonomia i taksonomia</i>	CZĘŚĆ, RODZAJ
<i>istnienie i posiadanie</i>	BYĆ, MIEĆ
<i>określniki</i>	TEN (TA, TO), TEN SAM, INNY
<i>życie i śmierć</i>	ŻYĆ (BYĆ ŻYWYM), UMRZEĆ
<i>mowa</i>	MÓWIĆ, SŁOWO, PRAWDA
<i>ewaluatory</i>	DOBRY, ZŁY
<i>deskryptory</i>	DUŻY, MAŁY
<i>czas</i>	KIEDY, PRZED, PO, DŁUGO (DŁUGI CZAS), KRÓTKO (KRÓTKI CZAS), TERAZ
<i>przestrzeń</i>	GDZIE, POD, NAD, DALEKO, BLISKO, WEWNĄTRZ, TUTAJ
<i>koncepty logiczne</i>	NIE, MÓC, JEŻELI, BO (PONIEWAŻ, Z POWODU), BYĆ MOŻE
<i>intensyfikatory</i>	BARDZO, WIĘCEJ
<i>podobieństwo</i>	(TAK, TAKI) JAK

Wiele osób może być zdziwionych pojęciami zawartymi na tej liście: o każdym z nich można by napisać niejedną rozprawę doktorską. Z drugiej strony, większość osób nie powinna mieć problemu ze zrozumieniem zdań używających powyższych pojęć (co więcej, podobno są one przekładalnie niemalże jeden do jednego na wszystkie języki świata).

NMS nie składa się jednak wyłącznie powyższych pojęć. Oprócz nich opracowywane są reguły ich łączenia (nazywane gramatyką uniwersalną) oraz metodologia przeprowadzania definicji pojęć złożonych. Akcentując kulturowy aspekt języka, Wierzbicka unika wielu problemów związanych z czysto logicznym ujmowaniem języka. Nie chciałbym jednak wchodzić w szczegóły, wszystkich zainteresowanych odsyłam do [10].

Tym, na co chcę zwrócić szczególną uwagę, jest fakt, że część indefinibiliów Metajęzyka pokrywa się z wyróżnionymi przez nas pojęciami podstawowymi pochodzącymi z prostej analizy umysłu (JA, CHCIEĆ, ROBIĆ, DZIAĆ SIĘ, CZUĆ, MYŚLEĆ, DOBRY, ZŁY i LUDZIE). Niektóre z pojęć NMS są związane specyficznie z myśleniem (w szczególności pojęcia zakwalifikowane jako predykaty logiczne NIE, MÓC, JEŻELI, BO, BYĆ MOŻE, kwantyfikatory JEDEN, DWA, DUŻO, WSZYSTKO, NIEKTÓRE, a także określniki TEN, TEN SAM, INNY, podstawowe elementy nominalne KTOŚ i COŚ, określenia istnienia i relacji BYĆ I MIEĆ, oraz szczególnie ciekawy i bogaty w możliwości łącznik określający relację podobieństwa TAK(I) JAK).

Powyższy zbiór elementarnych jednostek semantycznych wydaje się czymś w rodzaju układanki — właściwa teoria umysłu powinna połączyć je w odpowiedni sposób (być może dodając jakieś własne elementy). Podstawowa wątpliwość może dotyczyć języka, w którym zamierzamy układać tę układankę. Językiem tym nie będzie oczywiście NMS, bo jest on zbyt skomplikowany. Tym, czego poszukujemy, jest względnie prosty system formalny, którego model semantyczny jest kompatybilny z Uniwersalną Maszyną Turinga, na przykład jakiś uniwersalny język programowania komputerów albo rachunek λ . (Nie chcę niepotrzebnie zużywać miejsca na opis koncepcji algorytmu, rachunku λ czy maszyn Turinga/von Neumanna, gdyż te tematy zostały już wstarczająco dobrze opisane w innych publikacjach. Całkiem dobry i przystępny opis tych wynalazków Czytelnik znajdzie w [8] i innych publikacjach Penrose’a wydanych w Polsce; osobom dobrze znającym język angielski i posiadającym ciocię w Ameryce polecam również [7], klasyczną pozycję traktującą o myśleniu maszyn).

Istota naszego opisu ma się zasadzać na dwóch elementach: po pierwsze, na określeniu jakiegoś stanu (który rozumiemy tu dość ogólnie jako pewien zbiór elementów — przez element zaś możemy rozumieć na przykład dowolne wyrażenie rachunku λ) oraz na pojęciu zmiany. Oczekujemy, że właśnie w takim języku — w języku zmiany stanu — uda się nam adekwatnie opisać działanie umysłu. Jeżeli przyjmiemy takie założenie, to automatycznie zawieramy w naszym modelu kategorie czasu (wynikające z faktu zaistnienia zmiany) i — być może — przestrzeni, a zarazem w jakimś sensie dziedziczymy z samych założeń pojęcia TERAZ i TUTAJ. Nie są one wprawdzie bezpośrednio dostępne świadomości naszego dynamicznego modelu umysłu, ale powinny się łatwo dać wydobyć.

Osobne miejsce należy poświęcić pozornie prostemu łącznikowi TAK(I) JAK, określającemu podobieństwo dwóch rzeczy. Na rynku idei są dostępne dość wyszukane teorie opisujące skomplikowane mechanizmy tworzenia analogii i metafor. Obiecującym tropem wydaje mi się opracowywana przez Gillesa Fauconnier teorię *przestrzeni mentalnych* ([4], [5], [6]), ale niestety nie poznałem jej jeszcze w stopniu pozwalającym na wyciągnięcie z niej odpowiednich

wniosek. Ważne jest to, żeby uświadomić sobie, jak wielka złożoność może czyhać nawet za takimi pozornie prostymi pojęciami.

Teraz, kiedy w przybliżeniu określiliśmy sobie narzędzia, przy pomocy których chcemy uprawiać teorię umysłu, wypadałoby się zastanowić nad szkicowym projektem stworzenia sztucznego umysłu: od czego zacząć (i na czym skończyć). Wydaje się, że wygodnie byłoby w tym celu dysponować względnie bogatym komputerowym modelem sztucznego świata. Dobrze by było, żeby nasz umysł mógł się poruszać w trójwymiarowej przestrzeni tak, jak robią to ludzie i sterować ciałem podobnym do ludzkiego. Dzięki temu będziemy mogli łatwiej tworzyć konceptualizacje zbliżone do tych, które posiada człowiek. Oczywiście niezbędne są także pewne uproszczenia względem rzeczywistego świata. Przede wszystkim, żebyśmy mogli uniknąć problemów z rozpoznawaniem obrazów, dźwięków i zapachów, powinny one trafiać do modelu umysłu bezpośrednio od modelu świata, z pominięciem symulowania medium, w którym te sygnały się rozchodzą. W samym ciele kontrolowanym przez umysł nogi ludzkie można na początku zastąpić kołami, żeby ominąć złożony problem chodzenia. Są to bowiem zagadnienia po pierwsze trudne, a po wtóre — wtórne.

Kiedy już będziemy dysponowali modelem świata, możemy przystąpić do tworzenia modelu umysłu (przy okazji chcę zwrócić uwagę, że powinno się symulować nie jeden umysł, ale kilka na raz, bo większość ciekawych fenomenów językowych można obserwować dopiero w procesie komunikacji). Należałoby zacząć od modelu rudymenarnego, który można opisać w następujących słowach: mogę się czuć dobrze lub źle; niektóre rzeczy sprawiają, że czuję się lepiej, a inne — że czuję się gorzej; unikam rzeczy, które sprawiają, że czuję się gorzej i chcę doznawać rzeczy, dzięki którym czuję się lepiej. Zachowania, których możemy się spodziewać w takim modelu, to na przykład unikanie złego pokarmu i przyjmowanie pokarmu dobrego.

Kolejnym etapem jest dodanie do tego modelu pamięci, dzięki czemu będzie on w stanie nabierać nawyków. W dalszej kolejności należy rozwijać funkcję samopoczucia w taki sposób, żeby uwzględniała coraz więcej czynników (oprócz potrzeb fizjologicznych na przykład potrzeby społeczne). Najtrudniejszym etapem wydaje się dodanie możliwości tworzenia hipotez i modelowania umysłów innych istot. Docelowo chcielibyśmy wytworzyć coś, co jest zdolne do tworzenia modeli, które są zdolne do tworzenia modeli, które są zdolne do tworzenia modeli i tak dalej. Wydaje się to zadaniem szalenie skomplikowanym, ale nie niemożliwym.

Refleksja trzecia: o Bogu i świadomości

Częstą trudnością, na którą napotykają filozofowie umysłu, szczególnie ci wywodzący się z nurtów fizykalistycznych, jest problem świadomości. Pytają oni: „jak to jest możliwe, żeby z samej materii nagle wyłoniła się świadomość”. Proponują też różne rozwiązania: niektórzy szukają świadomości w procesach kwantowych [8], a inni twierdzą, że odczuwanie ma charakter epifenomenalny — gdyby go nie było, wszystko inne byłoby takie samo [1] (ale nie tłumaczą przy tym, na czym mogłaby polegać różnica pomiędzy umysłem, który rzeczywiście odczuwa, a tym, który tylko udaje).

Nie zwracają przy tym uwagi na fakt, że samo poznanie materii jest aktem wtórnym względem ich własnego doświadczania świata, na co szczególną uwagę zwrócił Kartezjusz ([3], [2]). Nie możemy mieć całkowitej pewności nie tylko co do tego, czy komputer potrafi odczuwać, ale również co do tego, że inni ludzie nie są bezdusznymi robotami udającymi tylko, że coś przeżywają. Co więcej, nie mamy nawet pewności, że wszystko, czego sami doznajemy, nie jest tylko jakimś snem, że naprawdę życie które uważamy za swoje nie jest projekcją jakiegoś mnicha buddyjskiego medytującego gdzieś w Tybecie (i że „tak naprawdę” nie jesteśmy tym mnichem). A skoro tak, to prawdziwość materii jest w pewnym sensie mniejsza, niż prawdziwość myśli i stwierdzenie jakoby świadomość miała wynikać z materii, zawsze będzie tylko nieweryfikowalną hipotezą.

My sami również jesteśmy zdolni do przeżywania życia innego, niż „nasze własne”, chociażby w kinie, teatrze czy grze komputerowej. Rozumiemy innych ludzi dlatego, że potrafimy z nimi współodczuwać. Również możemy programować komputery właśnie dzięki owej cudownej właściwości wyobraźni — potrafimy ograniczyć naszą percepcję i niejako „transponować” naszą świadomość na mikroprocesor. I w tym sensie możemy powiedzieć, że procesor czy program komputerowy ma świadomość — to my ją mu nadajemy poprzez naszą interpretację.

Z drugiej strony, wszystko, czego istnienie możemy stwierdzić, stwierdzamy poprzez świadomość właśnie — to jest absolutnie nieuniknione. O ile nie można stwierdzić, że jeżeli czegoś nie doświadczamy, to to nie istnieje (jak chciałby Berkeley [9]), prawdą jest, że wszystko, o czym możemy mówić, w jakiś sposób musi istnieć w naszej świadomości.

W taki właśnie sposób rozumiem Boga — jako istotę, która wyobraża sobie, że jest mną i Tobą, jako część wspólną wszystkich „ja”, czyste odczuwanie, jako czystego Ducha (o naturze zupełnie innej, niż materia i zupełnie do niej nieredukowalnej). Wydaje mi się, że rozsądniej jest przyjąć taką wersję, niż doszukiwać się świadomości w „nieodkrytych jeszcze przez fizyków procesach kwantowych”.

Literatura

- [1] Andrzej Chmielecki. *Między mózgiem a świadomością*. wydawnictwo Instytutu Filozofii i Socjologii PAN, 2001.
- [2] René Descartes. *Medytacje o pierwszej filozofii*. wydawnictwo Antyk, Kęty, 2001.
- [3] René Descartes. *Rozprawa o metodzie poprawnego kierowania rozumem i dociekania prawdy w naukach*. wydawnictwo Antyk, Kęty, 2002.
- [4] Gilles Fauconnier. *Mental Spaces*. Cambridge University Press, 1994.
- [5] Gilles Fauconnier. *Mappings in Thought and Language*. Cambridge University Press, 2005.
- [6] Gilles Fauconnier and Mark Turner. *The Way We Think*. Basic Books, New York, 2003.
- [7] Douglas Hofstadter. *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books, New York, 1979.
- [8] Roger Penrose. *Nowy umysł cesarza: o komputerach, umyśle i prawach fizyki*. PWN, Warszawa, 1995.
- [9] Władysław Stróżewski. *Ontologia*. Aureus – Znak, Kraków, 2006.
- [10] Anna Wierzbicka. *Semantyka. Jednostki elementarne i uniwersalne*. wydawnictwo UMCS, Lublin, 2006.